Quantification of Individual Rugby Player Performance through Multivariate Analysis and Data Mining

A thesis presented for the fulfilment
of the requirements for the degree of
Doctor of Philosophy
at Massey University, Albany,
New Zealand.

Paul J. Bracewell B.Sc M.Appl.Stat(Hons)

2003

OGS 10

MASSEY UNIVERSITY
APPLICATION FOR APPROVAL OF REQUEST TO EMBARGO A THESIS
(Pursuant to AC 98/168 (Revised 2), Approved by Academic Board 16.02.99)

Name of Candidate: Paul J. Bracewell ID Number: 95052126

Degree: PhD Dept/Institute/School: Statistics/IIMS

Thesis Title: Quantification of Individual Rugby Player Performance Through Multivariate Analysis and Data Mining

Name of Chief Supervisor: Denny H. Meyer Telephone Extn: 9495

As author of the above named thesis, I request that my thesis be embargoed from public access until (date) 28/02/05 for the following reasons:

- [x] Thesis contains commercially sensitive information.
- [ ] Thesis contains information which is personal or private and/or which was given on the basis that it not be disclosed.
- [ ] Immediate disclosure of thesis contents would not allow the author a reasonable opportunity to publish all or part of the thesis.
- [ ] Other (specify): ______

Please explain here why you think this request is justified:

This thesis details the construction of a commercial rating system and discusses the business processes that contribute to the calculation of such a rating system. Public dissemination of this information would remove the competitive edge the sponsoring company would otherwise possess in the commercial sector.

Signed (Candidate): [signature] Date: 22/2/02

Endorsed (Chief Supervisor): D. H. Meyer Date: 22/2/02

Approved/~~Not Approved~~ (Representative of VC): [signature] Date: 24/2/02

*Note: Copies of this form, once approved by the representative of the Vice-Chancellor, must be bound into **every** copy of the thesis.*

Abstract

This doctoral thesis examines the multivariate nature of performance to develop a contextual rating system for individual rugby players on a match-by-match basis.

The data, provided by Eagle Sports, is a summary of the physical tasks completed by the individual in a match, such as the number of tackles, metres run and number of kicks made. More than 130 variables were available for analysis. Assuming that the successful completion of observed tasks are an expression of ability enables the extraction of the latent dimensionality of the data, or key performance indicators (KPI), which are the core components of an individual's skill-set.

Multivariate techniques (factor analysis) and data mining techniques (self-organising maps and self-supervising feed-forward neural networks) are employed to reduce the dimensionality of match performance data and create KPI's. For this rating system to be meaningful, the underlying model must use suitable data, and the end model itself must be transparent, contextual and robust.

The half-moon statistic was developed to promote transparency, understanding and interpretation of dimension reduction neural networks. This novel non-parametric multivariate method is a tool for determining the strength of a relationship between input variables and a single output variable, whilst not requiring prior knowledge of the relationship between the input and output variables. This resolves the issue of transparency, which is necessary to ensure the rating system is contextual.

A hybrid methodology is developed to combine the most appropriate KPI's into a contextual, robust and transparent univariate measure for individual performance. The KPI's are collapsed to a single performance measure using an adaptation of quality control ideology where observations are compared with perfection rather than the average to suit the circumstances presented in sport.

The use of this performance rating and the underlying key performance indicators is demonstrated in a coaching setting. Individual performance is monitored with the use of control charts enabling changes in form to be identified. This enables the detection of strengths/weakness in the individual's underlying skill-set (KPI's) and skills.

This process is not restricted to rugby or sports data and is applicable in any field where a summary of multivariate data is required to understand performance.

Acknowledgements

During the course of this thesis I have received support, encouragement and advice from my supervisors, academics, sports-people, fellow postgraduate students and business associates. I thank all these people whole-heartedly for their assistance, understanding and contributions allowing this thesis to be completed.

I am especially grateful to the input provided by my supervisor, Associate Professor Denny Meyer, who inspired my pursuit of statistics, nurtured my passion for sport statistics, encouraged me to chase my ideas and provided me with the necessary direction and support required to complete this research. I cannot speak highly enough of Denny's wonderful influence upon my research and development as a student. Additionally, her assistance with defining the theoretical variance associated with the univariate parametric half-moon statistic was most helpful.

The encouragement and guidance provided by my co-supervisor, Dr Siva Ganesh, is also greatly appreciated. Additionally, I wish to thank Professor Jeff Hunter and Dr Paul Cowpertwait for their constructive comments for improving the heuristic test of independence introduced in Chapter Five

The input of business mentor, Mr Chris Lines, of the Eagle Technology Group was also extremely valuable. His assistance was instrumental in obtaining funding via the Graduate in Industry Fellowship from Technology New Zealand. Further, Chris provided excellent support relating to the feasibility of the models created. Additional to the academic and business contribution was the beneficial assistance provided by the numerous top-level coaches, selectors and players who made themselves available for discussion and freely voiced their opinions.

I am also grateful to the support provided by my family. Some of the philosophies pursued in this thesis were sown at a young age by listening to my father and his brothers debating their theories on sports coaching and performance – all with first class playing experience, in cricket and/or rugby. Given that they are still involved in high level coaching, their ideas proved to be relevant, challenging and thought provoking.

Table of Contents

Embargo iii

Abstract v

Acknowledgements vii

Table of Contents ix

List of Illustrations xiii

List of Tables xiv

1 Introduction 1

1.1 Prologue 1
1.2 Ideology 3
1.3 Framework 4
1.3.1 Previous Ideals *4*
1.3.2 Relevance of Research *6*
1.4 Research Questions 7
1.5 Process 9
1.6 Summary of Methodologies 11
1.6.1 Methods *11*
1.6.2 Philosophies *12*
1.7 Summary of Results 14
1.8 Layout 16

Part One: Introducing The Eagle Rating

2 Literature Review – Statistics In Sport 21

2.1 Statistics and Complex Team Sport 21
2.2 Definition of Sporting Ability and Performance 24
2.3 Quantification of Sporting Performance in Team Sports 28
2.3.1 Data Collection *29*
2.3.2 Selection of Numerical Measures *30*
2.3.3 Variety of Measures Between and Within Positions *32*
2.3.4 Combined Measures *34*
2.3.5 Validation of Measures *36*
2.3.6 Rating Systems for Individuals in Team Sports *37*
2.3.7 Commercial Systems *40*
2.4 Quality Control as Individual Performance Monitoring in Sport 42

3 Multivariate Analysis, Quality Control, and the Birth of the Eagle Rating........ 45

3.1 Overview 45
3.2 Data 47
3.3 Methods for Data Analysis 54
3.3.1 Cluster Analysis 54
3.3.2 Factor Analysis 56
3.4 Cluster Analysis 59
3.4.1 Clustering Philosophy 60
3.4.3 Coaching Perspective 63
3.4.4 Analysis 69
3.4.5 Application 70
3.4.6 Discussion 76
3.5 Factor Analysis 78
3.5.1 Underlying Philosophy 78
3.5.2 Applying Factor Analysis 80
3.5.3 Results 83
3.5.4 Discussion 86
3.6 Multivariate Quality Control 87
3.6.1 Standard Quality Control and Performance Monitoring in Sport 88
3.6.2 Existing Multivariate Control Methodology and the Perfection Method ... 90
3.7 Summary of the Eagle Rating 97
3.8 Implementation 100
3.9 Limitations 106
3.10 Conclusion 111

Part Two: Improving The Eagle Rating

4 Literature Review – Data Mining and Neural Networks 115

4.1 Data Mining 116
4.2 Neural Networks 121
4.2.1 Relation to Biological Networks 121
4.2.2 Brief Historical Overview 122
4.2.3 Applications 128
4.2.4 Cautions 129
4.3 Supervised Neural Networks 136
4.3.1 Types of Learning 137
4.3.2 Supervised Neural Networks Structure 138
4.3.3 Neural Network Architecture 142
4.3.4 Training Algorithm for Neural Networks 144
4.3.5 Activation Functions 147
4.4 Self-Supervised Neural Networks in Dimension Reduction 149
4.5 Self-Organising Maps in Dimension Reduction 151
4.5.1 Architecture 152
4.5.2 SOM Mechanical Properties 155
4.6 Interpreting Networks 157
4.7 Neural Networks and Sport 158

5 Investigation and Development of a Heuristic Test for Independence 163

5.1 The Introduction of a Heuristic Test for Independence 163
5.1.1 Covariance, Correlation and the Coefficient of Determination 165
5.1.2 Other Tests of Independence 167
5.2 Underlying Philosophy for the Half-Moon Statistic 169
5.2.1 Mathematical Background for Bivariate Case 170
5.3 Parametric Test for Bivariate Case 172
5.3.1 Disadvantages of Parametric Test 174
5.4 Non Parametric Test for Bivariate Case... 175
5.4.1 Effect of Sample Size 177
5.4.2 Effect of Skewness and Kurtosis 178
5.4.3 Examining Non-Linear Functions 183
5.5 Expansion to the Multivariate Case (Multiple Inputs) 189
5.6 Implementation of MHM to Calculate Relative Influence 192
5.7 Application 194
5.8 Conclusion 196

6 The Rebirth of the Eagle Rating Through Neural Networks 199

6.1 Data Pre-Processing 200
6.1.1 Sub-Group Identification 200
6.1.2 Data Limitations 202
6.1.3 Data 206
6.1.4 Defence/Attack Philosophy 210
6.2 Dimension Reduction Neural Network 212
6.3 Factor Analysis 229
6.3.1 Condensed Factor Analysis Module using Key Attributes 229
6.3.2 Player Performance Indices 235
6.3.3 Standard Factor Analysis using Summary Variables 235
6.4 Dimension Reduction Self Organising Map 238
6.5 Quartile Method Based Method for Discrete Ranking 245
6.6 Summary of Potential Rugby Ratings 248

7 Comparison, Selection and Implementation of Suitable Methods 251

7.1 Methods for Comparison 252
7.1.1 Original Eagle Rating 254
7.1.2 Factor Analysis for Condensed Data 255
7.1.3 Neural Network 257
7.1.4 1 × 10 Self Organising Map 258
7.1.5 4 × 4 Self Organising Map 260
7.1.6 Paired Self Organising Map 261
7.1.7 Augmented Indices 262
7.1.8 Quartile Model 264
7.2 Comparison of Methods 265
7.2.1 Model Context 266
7.2.3 Model Validity 269

7.2.3 Model Robustness *271*
7.3 Discussion 273
7.4 Implementation of Quartile Model 275
7.4.1 Implementation in the Media *275*
7.4.2 Implementation in the Coaching Environment *276*
7.5 Implementing Statistics in a Diagnostic Coaching Structure 276
7.5.1 Basic Philosophy *276*
7.5.2 Performance Applicability *279*
7.5.3 Creation of Stable Statistics *281*
7.5.4 Incorporating 'Quality Control' Into a Diagnostic Structure *282*
7.5.5 Application of Control Charting Procedures *283*
7.6 Conclusion 285

8 Conclusion 287

8.1 Overview 287
8.2.1 Research Questions *288*
8.2.2 Research Answers *288*
8.3 Implications 292
8.3.1 Limitations *293*
8.4 Contributions 294
8.5 Future Research 295

Appendix A: Data Description 299

Appendix B: Results of Factor Analysis 307

Appendix C: Job Requirements 317

Appendix D: Commercial Applications 321

Appendix E: Glossary 327

Appendix F: Half-Moon Related Methods 331

References 333

List of Illustrations

2.1: Schematic Representation Of Performance 25
3.1: Boxplot Showing Openside Flanker Domination of Tackle Counts 69
3.2: Dendogram Indicating Potential Positional Clusters 72
3.3: Boxplot for Metres Kicked Per Game by Cluster 74
3.4: Two Dimensional Plot Of Perfection Method Explanation 92
3.5: Relationship Between The Magnitude Of Perfection And Normality 93
3.6: Histogram Of The Eagle Rating 96
3.7: Normal Probability Plot For Eagle Ratings From The 2000 Super 12 96
3.8: Comparison Of The Eagle Rating And Wynne Gray 102
3.9: Shewhart Control Chart Monitoring Cullen's Super 12 Performance 105
4.1: Mcculloch-Pitts Neuron For The Logical And Function 124
4.2: A Simple Neuron 139
4.3: Multilayer Neural Network For Output K 140
4.4: Self-Supervised Learning 150
4.5: Self Organising Map Architecture For A 2-D SOM 153
4.6: Output Layer For A 1×10 Self Organising Map 154
5.1: Geometrical Representation Of The Half-Moon Statistic 170
5.2: Boxplot Of Parametric HM Statistic For Varied Radii 171
5.3: Histogram Of Parametric HM 173
5.4: Plot Of Parametric HM Statistic Against The Coefficient Of Determination 174
5.5: Histogram Of Non-Parametric HM 175
5.6: Plot Illustrating The Effect of m on the SD For The HM Statistic 176
5.7: Relationship Between Sample Size And Half Moon Variability 178
5.8: Relationship Between The HM Statistic And Skewness 181
5.9: Relationship Between The HM Statistic And Kurtosis 181
5.10: 3D Plot Demonstrating Relationship Between HM, Skewness And Kurtosis 181
5.11: HM Statistic Versus R-Sq For Polynomial Functions 184
5.12: Type II Errors For Non-Parametric Hm Statistic By Varied Sample Size 184
5.13: Mean DHM Statistic Versus R-Sq For Polynomial Functions 186
5.14: Scaled Chi-Square Statistic Versus R-Sq For Polynomial Functions 186
5.15: Mean DHM Statistic Versus Scaled Chi-Square Statistic 186
5.16: DHM Statistic Versus Observed R-Sq For Linear Functions 188
5.17: Scaled Chi-Square Statistic Versus Observed R-Sq For Linear Functions 188
5.18: Boxplot Illustrating Unscaled Z Distances For n=1,2,3,10 And 20 191
5.19: Boxplot Illustrating Scaled Z Components For n=1,2,3,10 And 20 191
6.1: Draftsman Plot Of First Defensive 'Output' Node And Input Variables 217
6.2: Draftsman Plot Of Second Defensive 'Output' Node And Input Variables 217
6.3: Draftsman Plot Of First Attack 'Output' Node And Input Variables 218
6.4: Draftsman Plot Of Second Attack 'Output' Node And Input Variables 218
6.5: Implication Of Principal Planes And Surfaces 219
6.6: 1×10 Self Organising Map for Midfield Backs 243
6.7: 4×4 SOM Referenced Against Attack And Defence Indices For Midfield Backs 243
6.8: Boxplot Of Attack Measure By SOM Column For Midfield Backs 244
6.9: Box Plot Of Defensive Measure By SOM Row For Midfield Backs 244
6.10: One-Dimensional 4×4 SOM Referenced Against Attack And Defence Indices 245
6.11: Structure Of Forced Operator Model Scaling For Attack And Defence 247
7.1: Flow Chart Applying An Individualistic Rating System As A Diagnostic Tool 278
7.2: Shewhart Chart Of Kelleher's Overall Rating By Match In The Super 12, 2000 284
7.3: EWMA Of Overall Rating For Kelleher By Match In The Super 12, 2000 285

List of Tables

3.1: Summary Of Eagle Sports Variables 50
3.2: Correlation Matrix Indicating Correlation Between Selected Variables 51
3.3: Perceived Positional Clusters....67
3.4: Cross Tabulation Indicating Relationship Between Position And Cluster 73
3.5: Cross Tabulation Indicating Relationships Between Clusters And Position 75
3.6: Summary Of Factor Labels By Positional Cluster....84
3.7: Varimax Rotated Factor Loadings And Communalities For Midfield Backs 85
3.8a: Correlation Matrix of Ratings With Different Magnitudes Of Perfection 94
3.8b: Correlation Matrix of Ratings With Different Magnitudes Of Perfection....94
3.9: Top 22 Performers In The 2000 Super 12 100
3.10: Allocation Of Tri-Nation Representatives Ranked By The Eagle Rating 101
3.11: Anova Differentiating Between Internationals And Non-Internationals 104
3.12: Anova Demonstrating Differences Between Winners And Losers 104
5.1: Skewness And Kurtosis Data....182
5.2: MHM Comparable Radii 190
5.3: Calculation Of Key MHM Components Using A Stepwise Procedure 193
5.4: HM Procedure Output For Attack KPI By Summarised Key Inputs....195
6.1: Summary Variables For Modelling 207
6.2: Correlation Matrix Of Summarised Variables....209
6.3: Optimal Architecture For Midfield Defence Neural Networks 215
6.4: Optimal Architecture For Midfield Attack Neural Networks....215
6.5: Numerical Measures For Defensive Parameter Importance 220
6.6: Numerical Measures For Attacking Parameter Importance....224
6.8: Summary Of Self-Supervising Neural Results For Attack....226
6.9: Summary Of Self-Supervising Neural Results For Defence 226
6.10: Summary Of Important Variables In Neural Network Module 228
6.11: Rotated Factor Loadings And Communalities For Midfield Defence Variables....230
6.12: Rotated Factor Loadings And Communalities For Midfield Attack Variables 230
6.13: Coefficients For Midfield Attacking Variables 231
6.14: Factor Analysis Module Summary 234
6.15: Appropriate Methods For Positional Cluster Performance Indices 235
6.16: Rotated Factor Loadings And Communalities For Midfield Backs 236
6.17: Summary Of Factor Analysis Results....237
6.18: Differences In Variable Importance Between SOM Topologies....241
6.19: 4×4 Som Co-Ordinates And Output ID 243
7.1: 4×4 Som Topology 261
7.2: Correlations Between Input Variables And Obtained Ratings 266
7.3: Correlations Between Match Outcome And Meta-Data....268
7.4: Inter-Rating Correlations 269
7.5: Top 22 Performers in the 2000 Super 12 According to the Augmented Indices 270
7.6: Comparative Performance Measures Of Ratings....271
A.1: Summary Of The Eagle Sports Coding System....300
B.1: Labels And Codes For Factor Analysis 307

Chapter One

Introduction

New Zealanders are passionate about *rugby union.* Hours can be spent discussing the comparative merits of various individuals in any number of rugby competitions, from the Super 12 Tournament to grass roots club rugby. When the All Blacks are named for the first time each year, this news dominates the front page of the major newspapers around the country. Most rugby observers have an opinion regarding potential selections.

1.1 Prologue

Invariably, when discussing team selections, conversation revolves around ability and performance, or the perception of ability and performance. Many factors can introduce bias to the perception of ability and performance, none bigger than provincial allegiance. An hour spent listening to "Sports Talk" on the radio will confirm this. Waikato has their Dwayne Monkley; Otago their John Leslie, provincial stalwarts who never wore the All Black jersey. Were they good enough? Whilst the question is no longer relevant, the underlying theme is important. Is it possible to compare players objectively?

The fundamental philosophy adopted in this thesis continues the research explored at the masterate level by Bracewell (1999). This research focused on quantifying the performance of individual first class cricket players based upon contribution to the team, for use in a selection context. Cricket lends itself easily to statistics as numerical output can be generated every ball.

Originally, cricket data was to be assessed in this thesis. However, talks with the Auckland Cricket Association and New Zealand Cricket proved unsuccessful. The Auckland Cricket Association indicated interest in the research, but they were not prepared to fund the programme. Eagle Sports were approached at the same time. Unlike the Auckland Cricket Association, Eagle Sport, a division of Eagle Technology, were prepared to fund the research and assist in sourcing external funding, which was provided by Technology New Zealand in the form of a Graduate Research in Industry Fellowship (GRIF). More importantly, Eagle Sports had a data set and a specific problem. Consumers (Sky Television and Ultimate Rugby) wanted an objective measure that reflected an individual rugby player's performance on a match-by-match basis.

During the research programme, a number of key events disrupted the commercial progress of the Eagle Rating. In December of 2000, the owner of Eagle Technology, Trevor Eagle passed away. Additionally, the Hong Kong based company that funded *Ultimate Rugby* did not pay out all the $250,000 prize money for the fantasy competition that was based on the work from Chapter Three. Initially Eagle Sports' core business was selling univariate match statistics to New Zealand first class teams and coaches. However, Eagle Sports' main competitor in the sale of univariate rugby statistics, AnalySports, with funding from the New Zealand Rugby Football Union, started providing match statistics to New Zealand first class teams and coaches for free. Consequently data collection by Eagle Sports ceased after the third round of the Super 12 competition, 2001. This is why only the 2000 Super 12 data is used in the second part of this thesis and a number of the variables detailed in Appendix A are not included. It is hoped that the results will advance the game of rugby in New Zealand as well as the production of other sporting ratings for commercial purposes. Apart from Eagle Sports considerable interest has been shown by companies from very diverse fields (Information Technology, Manufacturers and Sport Science).

1.2 Ideology

Using objective match statistics obtained from Eagle Sports, this thesis seeks to quantify individual rugby player performance. Quantifying performance enables comparisons to be made between players and for individual players between matches. Techniques from multivariate analysis, quality control and data mining are employed to liberate the story contained by objective match statistics, providing more fuel for "Monkley and Leslie" type debates.

However, first performance must be quantified. There is a distinction between ability and performance as is explained in Chapter Two. However, before statistical techniques can be applied, the justification for implementing such a procedure must be proven. A large pool of literature supporting the use of statistics for assessing sports performance is described in Chapter Two.

It is also in Chapter Two that the importance of sport statistics in rugby is introduced. The impact statistics has in a rugby environment is readily identifiable in coaching publications issued by the New Zealand Rugby Football Union (NZRFU), *Principles of Rugby Coaching* and *Advanced Coaching Course.* Reference to the antiquated *Coaching Accreditation Manual: Level 3* also provides useful insight. This publication has been replaced by the *Advanced Coaching Course* and this is where the focus will rest as this is targeted at the top-level coaches. Consequently, the statistical package that is produced based upon this work is tailored with this target audience in mind. This has immediate impact on the structure of this thesis. Whilst the primary aim of this thesis is to show that individual rugby player performance can be quantified using multivariate techniques, the use of such ratings in team profiling and ranking of individuals (NZRFU, 2000) in a selection procedure has had a powerful influence. In particular the latter stages of Chapter Seven demonstrate how such ratings can be incorporated into a diagnostic coaching structure.

1.3 Framework

To quantify individual rugby player performance, a variety of statistical techniques capable of dimension reduction are used to first extract individual player performance (using the principles of Spearman (Manly, 1994)), and these techniques are compared. The relative strengths and weaknesses of the applicable techniques are discussed throughout the body of this thesis specifically in Chapters Three and Six, but centred in Chapter Seven.

1.3.1 Previous Ideals

In order to extract a meaningful performance rating, the approach adopted is similar to Spearman's work relating to the quantification of intelligence (Weiten, 1995; Manly, 1994). Spearman concluded that the cognitive abilities of individuals, determined from many specific mental tests, shared an important core factor relating to general mental ability (Weiten, 1995, p. 362). Factor analysis was developed by Spearman to enable the extraction of the latent construct relating to mental ability (Manly, 1994). Similarly, we wish to identify the characteristics of performance expressed from match data. This is based on the assumption that match performance is the manifestation of ability and that performance is measurable based on the observance of physical tasks performed in match conditions.

Ultimately this is a form of data scoring, a term often referred to in the data-mining environment, which means filling in the outputs (Berry & Linoff, 1997). Scoring systems have been used for the past 30 years in a range of applications such as identifying the risks of individuals defaulting on home loans and companies succeeding. A wide range of statistical and data-mining techniques are used to enable scoring to occur (Kitts, Freed & Vrieze, 2000; Langley, 1997; Berry & Linoff, 1997; Grady, Schryver & Leuze, 1999). Essentially characteristics of the data set are extracted to provide interested parties suitable ratings from which decisions and predictions can be made (Berry & Linoff, 2000), so scoring is generally associated with dimension reduction.

There is the danger of losing crucial information in the dimension reduction process. However, given that it is very difficult for a human observer to detect patterns in a multivariate data set (Grady *et al.*, 1999), it is often necessary to condense a vast number of variables into a more manageable number. This enables patterns and trends to be identified visually, or in tables in the case of meta-data. The danger of information loss can be minimised by using effective dimension reduction techniques that eliminate the redundancies in the data by identifying the true dimensionality (key performance indicators) of the input data.

Most scoring systems need to produce accurate scores which are robust in the sense that they are applicable to everyone (or everything) in the target population. The risks associated with scoring a particular model are dependent on the intended use of the obtained information. The information extracted from the analyses in this thesis forms part of a tool intended to supplement the experiences of the coaching staff and selectors, as well as contributing to the entertainment experience provided by professional rugby. The risks are therefore much less than would be the case in identifying the likelihood of a company defaulting on a loan when an inaccurate score may have dire financial implications. But, the accuracy and robustness of the rugby ratings are only two of their requirements. If these ratings fail to make sense to the people who should be using them they will never be used. Transparency and contextuality are therefore also essential. Indeed, misinterpretation of the ratings is the greatest threat to the success of the rating system, provided the key issues are resolved.

The ability of models to score values correctly erodes over time (Berry & Linoff, 1997). This is especially true in a dynamic game such as rugby union where rule changes, game theories and new fads will impact on the nature of the game and the subsequent models. Thus the model must continually be up-dated to ensure that model context and appropriateness is retained.

Whilst scoring is a very broad term, fundamentally, it is Spearman's approach that is the most closely related to the quantification of individual rugby player performance, and this provides the necessary framework. Accordingly, factor analysis and other suitable dimension reduction techniques are used to extract the necessary information relating to individual performance.

1.3.2 Relevance of Research

The need for rugby ratings was described in the commercial context by Chris Lines (Lines & Bracewell, 2000, p.2) in the application for the funding which was obtained for this research. Importantly, the economic significance of the research was stressed, due to the novelty and desire for a univariate rugby rating obtained from match performances.

> "As the quality and usefulness of statistical information is improved, more coaches, selectors, and sport analysts will avail themselves to the service. The ability to provide a single statistic (The Eagle Rating) for player ability (performance) is significant to fans, therefore significant to media outlets in general. Eagle Sports expects to generate additional revenue of (specified income) per annum from the sale of the enhanced statistical service."

Additionally, Eagle Sports were approached directly by the media (Sky Television and Ultimate Rugby) with the request for a single statistic for individual rugby player performance. The intended uses are mainly entertainment driven, with match ratings applicable in match broadcasts, reviews, build-ups and other marketing promotions such as collector cards. This highlights the need for statistics in the commercial/media sector.

Additionally as mentioned in section 1.2, the NZRFU coaching manuals are punctuated with reference to quantitative assessment. Furthermore, the NZRFU coaching resources and organisational behaviour texts explain how certain social phenomena, such as social loafing can be reduced by the use of measurement.

AnalySports do provide KPI, but these have no statistical basis, created using human experience. A major benefit of using statistically based KPI's is objectivity. This objectivity was beneficial in using the rating as the core component of a skill-based fantasy game as described in a report presented to the Lotteries Commission explaining why the *Ultimate Rugby* was skill-based and not chance-based in nature.

Sport statistics are a core component of the entertainment industry generated by professional sports. Furthermore, statistics are being used in the sporting environment, both by the media and coaching staff, therefore it is important to illustrate and advertise the correct implementation of statistical procedures in this field.

The growing use of scoring applications (often using off-the-shelf software) means that the approaches developed in this thesis also have relevance outside of the sporting environment. Using the wrong type of score is a common occurrence because not everyone has the ability to develop their own scoring system. Purchased systems are difficult to assess because transparency is usually absent (as software suppliers want to maintain their competitive advantage/secrets). This thesis develops an important new method, which can be used to introduce transparency to any scoring system.

1.4 Research Questions

With the increase of professionalism in sport and the role sport plays in the entertainment industry, the desire to understand the performance of an individual is increasing. The performance of an individual in a sporting context must be viewed as multifaceted. Thus previous attempts by both coaches and the media to describe performance on a univariate level are insufficient. Presently vast arrays of univariate measures are provided by the media (AnalySports, Eagle Sports, The New Zealand Herald and Sky Television to name a few) to describe performance with the assumption that these values are a reflection of performance. With the availability of computers and techniques that enable rapid analysis of multivariate data, it makes sense to usher sport statistics into a new era.

No objective rating systems exist for rugby or for any other complex team sport. Consequently, we must ask the question "*is it possible to measure individual rugby player performance in a team environment?*" For this to be possible, performance must be defined and then operationalised. The necessary data and associated systems were provided by Eagle Sports, providing the ideal platform for assessment.

Whilst every individual performance will differ on a match-by-match basis, it is hypothesised that some elementary traits (performance measures) exist within the data, identifiable by position and defined role within a team environment. What methods are appropriate for extracting these elementary traits? Given that we are trying to identify the key constructs (Key Performance Indicators), or the latent dimensionality of expressed performance, it is suggested that the most suitable techniques will involve dimension reduction. Subsequently, it is proposed that the appropriate methods are factor analysis, self-supervised feed-forward neural networks, and self-organising maps.

If Key Performance Indicators (KPI) can be identified, for each individual in each match, "*how can these separate components be combined to provide an effective overview of performance?*" Additionally, depending on the method used, are the KPI's interpretable, and how important is interpretability on the final product? Credibility is a major issue associated with a rating system, and this can be partly attributed to contextuality. A problem with feed-forward neural networks is their perceived lack of interpretability, which is necessary to ensure that contextuality is achieved. Is it possible then, to create a method enabling interpretation of these networks? Furthermore, what are the potential problems associated with these analyses? The questions can be summarised as follows: "*How can the system of ratings be constructed in order to achieve transparency, contextuality and robustness?*"

Provided that a rating system can be established, additional questions are broached, impacting on the usefulness of the proposed rating system in the rugby environment. Most importantly, "*how can a measure of individual rugby player performance be employed appropriately by interested parties, in particular coaches?*"

The final question that must be asked concerns the relevance of this work for other rating systems, in particular, for rating systems in other sports.

1.5 Process

The research process was driven by the needs of Eagle Sports to ensure their survival as a viable commercial entity. Mr Chris Lines, Business Unit Manager of Eagle Sports provided invaluable feedback regarding potential models from a commercial and computer science perspective. Due to the commercial involvement, the quality of the data available was unparalleled, with input from some of New Zealand's top rugby brains, such as former All Black first five eighths and first class coaches, Grant Fox and Wayne Smith, in creation of the data collection process and associated systems. Furthermore, the commercial involvement lead to direct discussions with other top level coaches, as well as current and former All Blacks and first class players. This was necessary to ensure the resultant statistics were suitable for the target market. These meetings proved crucial in establishing the justification for the adoption of the proposed statistical rating system and in establishing how such a system fits into the rugby environment.

In creating the initial Eagle Rating, the time pressure was immense, with the official launch date of the fantasy game *Ultimate Rugby*, in late February 2000, almost two months after commencing analyses of the Eagle Sport data. The feedback obtained from *Ultimate Rugby* provided the necessary insight relating to the requirements of a rating system aimed at the sporting public. The defence of player performance ratings from the *Ultimate Rugby* module provided the inspiration for interpreting statistics in a diagnostic coaching environment. This initial work acted as a pilot study for the work to follow.

The problems of the pilot study were identified and we sought to improve the Eagle Rating in subsequent analyses. The data collection procedures were also expanded to cover gaps in the coding process, and these were also supplemented with additional information relating to subjective measures such as judgement and execution (which had an objective component incorporated to ensure inter-coder reliability). However, due to the business reasons stated earlier, the new information was not collected.

Due to the specialised nature of the job requirements of individual rugby players, it was necessary to identify groupings that shared the same core skills or were expected to undertake similar physical tasks. Identifying positional clusters from the data using cluster analysis proved unsuccessful due to the constraints of the data, requiring groupings of similar positions to be identified based on expert opinion.

To make full use of the potential promised by neural networks, it was necessary to develop a method for interpreting these highly flexible models. The half-moon statistic was developed and proved useful in establishing transparency and context on a particular model.

This enabled the final modelling process to begin. Using several different procedures models were tested and compared with a hybrid approach eventually chosen that depended on position.

Essentially, we sought to create an expert system (Mena, 1999) capable of mimicking the opinions of unforgetful, human experts from match statistics obtained from on-field activities, using our models. However, our 'expert' typically had one season's worth of data to be trained with, compared with a 'human expert', who typically has lived and breathed the game of rugby union for much longer. As a result, throughout the research process caution was required to ensure that the models attained were sufficiently contextual, transparent and robust.

At this point the bilateral theme that dominates this thesis becomes evident. To fully understand the entirety of this work a background in both rugby and statistics is required. However, as this work is based predominately on statistical modelling, a brief glossary is included (Appendix E) which describes some of the rugby terms encountered. The two separate themes run parallel throughout, with the emphasis shifting continuously. Thus, the material within this thesis is rather compartmentalised. Specific sections deal predominately with statistics and others with rugby philosophy. The implementation chapters provide the common ground and show how to utilise effectively the information generated from both a rugby and statistical perspective. However, it is impossible to fully separate the two components as they are tightly linked.

1.6 Summary of Methodologies

To successfully quantify individual rugby player performance, the methods used and the underlying philosophy that shaped the use of these methods form an important partnership. In this section the key methods used are summarised before the associated philosophies that affected these analyses are briefly discussed.

1.6.1 Methods

In order to quantify individual rugby player performance using objective multivariate data, several techniques were used.

Cluster Analysis

Initially cluster analysis was used to identify positional clusters within a team, explored primarily in Chapter Three. Clusters were identified (nine) using a hierarchical procedure and these clusters were then refined using an agglomerative k-means procedure. Ward's method was used for the linkage of the hierarchical clusters. Expert opinion (coaches, trainers and journalists) was also used to define positional clusters.

Dimension Reduction

Three core methods were used for dimension reduction in Chapters Three and Six of this thesis, detailed below.

Firstly, factor analysis was employed using a subset of the 93 physical task variables, employing a varimax rotation. The perfection method was used to collapse the factor variables (five) into a single performance measure.

This was followed by the implementation of a self-supervised feed-forward neural network with a bottleneck, trained using backpropagation and with early stopping to improve generalisation. This was a four-layered network – with a hyperbolic tangent activation function in the first and third layers and a linear activation function in the second and fourth layers. This was applied to each of the key attributes (attack and defence) for each positional cluster (16 networks were trained each using seven summarised variables). The bottleneck nodes were combined to create KPI's using the perfection method before the relevant KPI's were combined using the perfection method.

The final method explored was a self-organising map, trained with a batch algorithm and utilising all of the 19 summarised variables. Initially a one-dimensional (1×10) topological mapping was used before this was expanded to a two-dimensional (4×4) mapping for each positional cluster.

1.6.2 Philosophies

To allow the research questions proposed in this thesis to be fully explored, a number of different approaches were required to ensure suitable information was extracted from the data set using the methods detailed previously. These are described briefly below.

Expert System Development

The ultimate goal of this thesis is to produce a statistical model that mimics the opinion of an unbiased expert human observer. Thus the philosophies discussed in this section are techniques to help in the attainment of this goal.

Specialisation/Generalisation

In the identification of positional clusters, the number of clusters adopted needed to consider the trade-off between specialisation (too many clusters) and generalisation (too few clusters).

Use of Winning Teams

In the desire to extract the latent components of performance, only the winning teams were explored in the pilot study. If latent features could not be identified from a specific subset of performances, then it was likely that these latent features did not exist. Further, at first class level it is expected that more desirable characteristics of performance would be expressed in winning teams.

Perfection Method

To successfully condense KPI's to a single measure of individual performance, a suitable philosophy needed to be developed. Modifying control chart theory to compare observations with perfection rather than the average created a method suitable for use in the sporting environment.

Separation of Attack and Defence

Due to the small number of observations per adjustable parameter in the neural network, the number of variables involved in each analysis was reduced to avoid over-fitting. This was achieved by identifying distinct core attributes of performance (attack and defence).

Influence versus Formulation

In creating a suitable model it is the overall influence that is of interest rather than the actual model formula. Transparency and context are crucial for promoting the rating system and for this to occur, the general influence of the variables involved is more important than the specific co-efficients required to calculate the rating.

Optimal Solution?

An optimal solution is not required. A good solution that makes sense to a rugby observer is sufficient. The accuracy of the ratings is not of vital significance because ratings generally will be expressed as whole numbers.

Contextual Models versus Robust Models

To ensure models are suitable for use as a rating system, there is a trade off between contextuality and robustness (over-sensitivity to changes in match involvement) that needs to be considered. This is necessary to ensure that comparisons can be made, between-individuals, within-individuals, between positions and within-positions.

1.7 Summary of Results

This thesis shows that individual rugby player performance can be quantified and that interested parties can use the resultant statistics in a meaningful way.

Expert opinion was used to define positional clusters that needed to be identified due to the different demands placed on different positions in a game situation. Cluster analysis failed to provide convincing results from the Eagle Sport database; however, due to match volatility and non-performance this was understandable.

The extensive Eagle Sport database was used to reveal the latent dimensionality of the physical task data. Underlying a single measure of individual performance in a match is the existence of separate components, or Key Performance Indicators (KPI's). Dimension reduction techniques (factor analysis, self-supervised feed-forward neural networks and self-organising maps) indicated the existence of statistically viable (contextual) KPI's. The KPI's were then collapsed to a single performance measure via the perfection method. This is an adaptation of quality control ideology and observations are compared with absolute perfection rather than the average to suit the circumstances presented in sport, producing a one-dimensional value that is reflective of an individual's performance in a single match.

However, due to violations of context and robustness in some components of the positional cluster models, none of the methods mentioned above were suitable as a generic method for all positional clusters. Consequently a hybrid method was developed that created augmented indices using key attributes constructed using the output from factor analysis and self-supervising feed-forward neural networks. This ensured that the models obtained for each positional cluster were transparent and contextual. These are crucial issues in a sports rating system. The third problem area, robustness, was solved by adopting a quartile-based method for each of the key attributes, prior to combination via the perfection method. Thus a contextual, robust and transparent univariate measure for individual performance had been successfully created.

Finally, the use of a performance rating and the underlying key performance indicators is demonstrated, by simply constructing and then deconstructing the rating system. After construction, individual performance is monitored with the use of control charts enabling changes in form to be identified. This enables the detection of strengths/weakness in the individual's underlying skill set (KPI's) and skills, through the deconstruction of the performance rating(s).

In the above development major problems were encountered with ensuring transparency in the case of the neural networks, commonly referred to as black boxes in the literature (Ripley, 1994). A new statistical tool has been developed to solve this problem.

The half-moon statistic was developed to promote transparency, understanding and interpretation of dimension reduction neural networks. This novel non-parametric multivariate method is a tool for determining the strength of a relationship between input variables and an output variable, whilst not requiring prior knowledge of the relationship between the input and output variables. This resolves the issues of transparency and interpretability, which are necessary to ensure the rating system is contextual. Additionally, this method shows a great deal of promise as a diagnostic tool for use with multivariate data.

1.8 Layout

This thesis is deliberately constructed in two parts. The first part introduces the Eagle Rating, the product of a pilot study. The second part then seeks to improve on the initial rating system by applying artificial neural networks. The basis for this structure is to communicate many of the underlying arguments and assumptions inherent in the Eagle Rating in an uncomplicated manner by focusing on the rugby issues first. Essentially, Part One deals with many of the rugby themes, issues and assumptions that must be considered in statistical analyses. This provides useful background to both rugby and statistical ideologies. This structure is imposed in order to create flow. A number of appendices are included to aid the flow and concentrate specifically on either rugby or statistical ideology. From a statistical perspective, the use of the rugby data is merely an illustration of how the multivariate methods contained are applied. Conversely, from a rugby perspective, the statistical methods involved are just tools needed to calculate individual performance measures. Importantly, it is the underlying philosophy of using multivariate techniques to extract meaningful performance measures for an individual athlete that is the dominant theme throughout.

Before analyses can occur performance must be quantified. Chapter Two examines the distinction between ability and performance. The literature supporting the use of statistics for assessing sports performance and the relevance in rugby is also described in this chapter.

Chapter Three focuses on creating a rating system that quantifies individual performance using traditional techniques for multivariate analysis. The handling of issues specific to rugby is confronted within the statistical framework. A core component is the identification of different groupings based on position; due to the different expectations placed on these positional clusters imposed in a match context. Then implementing factor analysis as a technique for summarising match data allows a condensed data set to be assessed that retains the majority of the information from the original variables. Highlighting the bilateral theme in Chapter Three is the inclusion of a section relating to multivariate quality control. This area of statistics is 'borrowed', enabling a reduction of the results from the factor analysis to a single measure of match performance. This provides a statistic that is relatively stable from which inferences about ability and performance can be made. However, an adjustment in the mathematics is required as quality control focuses on the average, whereas in the sporting environment the 'best' is expected. Thus, the principle of Hotelling's T^2 control statistic is modified using a Mahalanobis distance to compare observations to perfection. The results of Chapter Three indicate that individual rugby player performance can be quantified answering the first research question. This rating system is dubbed the Eagle Rating in recognition of the commercial backing supplied by Eagle Sports. From a commercial perspective the use of the Eagle Rating in the fantasy game *Ultimate Rugby*, and in determining the *Sky Man of the Match*, illustrates the applicability of this research to the wider community. To conclude the chapter the limitations of this rating are explored.

Using an improved data set and seeking to improve upon the Eagle Rating, Part Two commences with a review of data mining techniques in Chapter Four. Incorporating the lessons learnt from Part One of this thesis, applicable data mining methods are explored from rugby and statistical perspectives. Key issues influencing the methodology for quantifying individual rugby player performance – Robustness, Context, Transparency, and Implementation – shape the techniques explored.
A potential problem with the use of data mining techniques is the lack of interpretability, specifically regarding neural networks. Interpretability is crucial for contextual models to be created. Due to the deficiencies of conventional sensitivity analysis for understanding the influence of input variables on the outputs in neural networks, Chapter Five introduces the half moon statistic as a method capable of interpreting the holistic impact of input variables upon the subsequent outputs. This method is developed from a univariate parametric test for independence to a non-parametric, multivariate statistic that can be thought of as a functionally independent, multivariate coefficient of determination.

This new methodology is then incorporated into the analyses of Chapter Six where two data mining methods are used in attempts to quantify performance. Self-organising maps and self-supervised feed-forward backpropagation neural networks are the fundamental methods assessed. Additionally, factor analysis is revisited so that the comparisons engaged in Chapter Seven are relative. To supplement these techniques a quartile model is also introduced which is a hybrid method that attempts to incorporate the strengths of factor analysis, self-organising maps and neural networks, into a simple 1-10 rating, answering the second research question.

The results from Chapter Six are explored in Chapter Seven, identifying the most suitable techniques and models. A number of issues need to be considered, most importantly context, as this leads to credibility. Robustness is also a key component when adopting the final model due to the dynamic and volatile conditions that comprise rugby. The best method is identified and implemented in contextual applications, indicating the potential use of the new rating system and answering the third research question. The appropriate use of the new rating is demonstrated in a diagnostic coaching environment, placing the relevance of the obtained statistics in context and answering the fourth research question.

Chapter Eight concludes the thesis with a detailed deliberation over the relevance and appropriateness of the methods and data involved, the work remaining, and answering the final research question, posed above, namely "*What is the relevance of this work for rating systems, in particular for rating systems in other sports?*"

Part One

Introducing the Eagle Rating

Chapter Two

Literature Review – Statistics in Sport

In order to bring together information from vastly different sources adequately, the literature review is divided into two parts. The first part is given in this chapter, which deals with the contribution of statistics to sport, in terms of quantifying performance through both univariate and multivariate techniques. An overview of the applicability of quality control theory to sport concludes this portion of the literature review. Chapter Four describes the literature regarding neural networks and the suitability of such methods to this thesis.

2.1 Statistics and Complex Team Sport

In dealing with sport, this research stays confined to complex team games such as rugby union and other forms of football (Rugby Union, Rugby League, Soccer, American Football, Gaelic Football and Australian Rules). Whilst the structure and rules of these games may differ, common skill based components of the football codes are given as follows; passing, evasion, catching, contact, kicking and unpredictability (Douge, 1987). Examples from the varying codes of football are of interest due to the dependencies of the interactions between competitors in a complex and chaotic environment. The literal context of the terms complexity and chaos is sufficient for the purposes of this thesis. However, the scientific definitions of these terms and the relevance to the sporting environment provide an interesting tangent and a potential avenue for further research. A key element of the definition of chaos is the sensitive dependency on initial conditions (O'Hare, 1996; Townend, 1996; Isham, 1993; Jensen, 1993). This is clearly exhibited in the football codes where options at the disposal of an individual are governed by the actions of all individuals and contributing factors in that match (Townend, 1996).

Recording the options taken by an individual in a team sport along with the subsequent outcomes would add interesting information to a potential database. The resultant statistics could be used in a performance-monitoring scheme. There are several key reasons for measuring and evaluating performance in team sport. A review of the NZRFU Advanced Coaching Guide demonstrates a number of uses for statistics in rugby such as compiling a team profile, selecting a team, designing training strategies, game planning and preparation, team strategies and team dynamics. Further reasoning for use of statistics in rugby is outlined more precisely in the out-dated NZRFU Coaching Accreditation Manual: Level Three (1991). Exploring these two coaching guides provides, to some extent, the justification for collecting and assessing statistical data. The second module of the Coaching Accreditation Manual, *Teaching Skills and Skills Analyses,* discusses the benefits of statistics in rugby coaching.

Whilst the coaching certification associated with the Level Three course has been superseded by the Advanced Coaching Course, the themes discussed are still relevant. Qualitative analysis always has a role to play because there will always be aspects/influences that cannot be quantified adequately, such as technical, biomechanical and decision making processes. Quantitative assessment is also needed in order to provide more objective information, as assessment based only upon qualitative analysis can be distorted by specific events such as; injuries, the score board, player's decisions, refereeing decisions and the most recent 'disaster' (NZRFU, 1991). Furthermore human limitations make qualitative analysis difficult in a sporting context. Recall of events is impaired by a number of factors as described by the NZRFU (1991), detailed as follows:

- It is impossible for coaches to absorb all the events of a game
- Individuals have limited recall of what has occurred
- Events that happen only a few times may not be recalled
- Tension, emotion and personal bias significantly affect the retention of relevant data

Again referring to the NZRFU (1991), the importance of statistics as a coaching tool is highlighted by the fact that improvements in performance are related to the quality of feedback given to players after a game. This feedback is most effective if given as soon as possible. However, performance may not be improved if there are flaws in this feedback. This is particularly applicable to peripheral play away from the ball which involves most of the players most of the time. Any qualitative feedback the coach gives to players is based on partial information because it is impossible to see all players at once and not be distracted by play with the ball.

The improved development and subsequent reliance upon video analyses (incorporating digital footage and software enabling instant recall of events) to review events can remedy a number of the problems associated with qualitative analysis in post match analysis. However, statistics are needed to an element of objectivity. This will be pursued further in Chapter Seven with respect to incorporating statistics in a diagnostic coaching structure.

Essentially, the purpose of sport statistics is to aid the creation of a high performing team. Briefly, in order to build a high performance team, appropriate performance measures are required (Greenberg and Baron, 1997). Reviewing team dynamics presents the best illustration of how statistics can be beneficial in a team environment. Specifically, statistics can expose the social phenomenon of social loafing, which is the result of diminished responsibility in a group environment and results in decreased energy expenditure on group tasks (Greenberg *et al.*, 1997). Social loafing usually occurs when individual contributions are not identifiable (Chu, 2000). "Social loafing is reduced when individual inputs are directly monitored (e.g. individual statistics; lap times; assists; tackles; goals etc) (Chu, 2000, p5)."

Having shown that quantitative analysis, and hence statistics, have a part to play in evaluating individual performance in a team sport, we must consider what is being quantified. The purpose of this thesis is to evaluate the performance of individual rugby players. Naively, performance is directly attributable to ability. More specifically, the properties of performance are a result of the hypothetical construct referred to as ability.

2.2 Definition of Sporting Ability and Performance

Before getting too far into the literature review, it is necessary to define sporting ability and performance. In sport, skill must be perceived as sporting performance which tends to increase the probability of success (Thomas, 1970). Further, sports skill can be defined as "any behaviour which tends to improve performance in sport (Thomas, 1970, p.126)". According to the Oxford Senior Dictionary (Hawkins, 1990), the combination of skills gives rise to ability. This definition fits nicely with rugby terminology and the continual reference to a player's skill set by coaching staff determining player value (de Lacy & Fox, 2000; Stewart, 1987). Thus sporting ability is defined as the combination of skills that tend to improve the performance in sport. Performance can then be viewed as the manifestation of ability.

It is possible to measure skill as outlined in the definition of ability. However, it must be noted that what is measured may not be what is inferred from this measurement. This is an important distinction and essentially a critical assumption that underwrites the development of the Eagle Rating introduced in Chapter Three and modified in Chapter Six. For this thesis, skill is inferred from performed tasks. Tasks are directly measurable (countable) and to complete the task successfully, the skill to perform such a task must exist. Mental and physiological tasks that contribute to the performance of a physical task, such as passing the ball, are not directly measurable from a game and are inferred from the performance on physical tasks. Thus skill comprises the set of mental, physical and physiological tasks that are required to increase the probability of success. Consequently, performance is the observed expression of ability in a match situation, which is established based on the completion of operationalised physical tasks, or skills. This relationship is illustrated in Figure 2.1.

To successfully perform a physical task, such as kicking the ball in a game situation, the other factors that comprise skill must also be successfully performed. Therefore by measuring physical tasks, the existence of a particular skill can be inferred as the physical task cannot be completed unless the full set of tasks that comprise the skill (physiological, physical and psychological) are completed. Fitness and mental skills are core components of rugby ability. They may not be directly measurable from a rugby match, but the level of these components can be inferred from other measurable skills. When a skill cannot be replicated in a game environment due to either a lack of mental application or fitness, the absence of such data will suggest potential weakness. Essentially, what is the point of a player being able to wrestle the ball off any opposition player if that individual can never get to the breakdown? This example acknowledges that confounding variables, such as fitness and mental attributes, do not pose a problem. The confounding effects are attached directly to the skill-set and observed physical task in a match situation, which relates to perceived performance and the consequent inferences regarding ability. Consequently, quantified skill is inferred from quantified tasks. As ability is the combination of skills, by combining the measurement on tasks, the expression of an individual's ability (performance) is quantified.

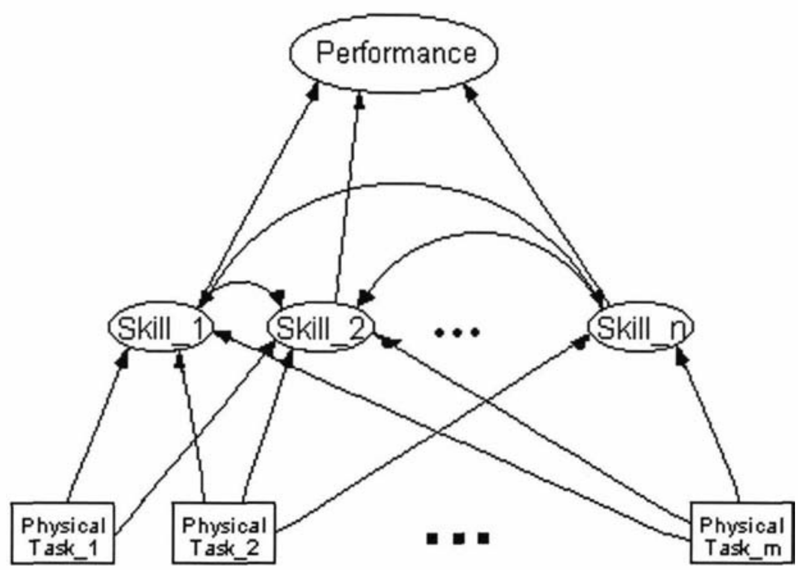


Figure 2.1: Schematic Representation of Performance for a Single Match Incorporating Physical Tasks

Figure 2.1 is important in conceptualising the steps required to create a rugby rating. It is important to deviate slightly at this stage and briefly examine potential methods for combining the measurements on tasks. Whilst individual ratings are an important component of the sporting entertainment package, use and development of measuring performance is more natural in the education sector due to the ease of data operationalisation and collection. The simplest way of assessing student learning performance is to have them sit a test or exam. The marks that are obtained can be used to extract inferences regarding certain abilities of the student.

The desire to quantify cognitive ability, or general mental ability, (*g*) in humans is based on the hotly disputed premise that this is reflective of an individual's capacity for learning and predictive of future job-performance (Segall, 2001). Segall cites a number of references that have found the best predictor of future job performance and learning capacity for inexperienced individuals is general mental ability. Consequently, research has focused on identifying and refining estimates of cognitive ability.

The situation presented in the initial part of this thesis is parallel to the situation that confronted Charles Spearman (1904) in the quest to quantify intelligence. Spearman's (1904, 1923) major work focused on the assessment and study of intelligence leading to the early development of factor analysis (Manly, 1994). Through factor analysis, Spearman concluded that all cognitive abilities share an important core factor, labelled *g* for general mental ability (Weiten, 1995). Monitoring different tests for intelligence and other forms of academic assessment, Spearman sought to identify intelligence, or more aptly, academic ability. The parallels are evident immediately; the testing arena for this thesis is a series of rugby games, from which we seek to identify rugby ability and performance, utilising pre-specified numerical measures, or benchmarks, relating to on-field performance.

Although Spearman's two-factor theory (which identifies general ability and test specific ability) has been surpassed by hierarchical models of intelligence, the identification of *g* is widely assumed by cognitive theorists to be at the apex of all mental measurements (Segall, 2001). Consequently, as this thesis seeks to establish an overall measure of performance, which is viewed as a product of ability, Spearman's philosophy provides an ideal starting point. The prevailing approach for identifying cognitive abilities remains consistent with Spearman's original formulation (Segall, 2001). Consequently, the literature regarding testing abilities in the educational and psychological sector is not pursued, other than to examine briefly Spearman's approach to quantifying cognitive ability. This is due to the differences between quantifying sporting and academic abilities, namely in the areas of operationalisation, variable involvement, changeable match conditions and constraints, and confounded data.

Underlying the functionality of factor analysis is the following premise; if the measurements of a number of variables are highly correlated, this suggests that a single factor is influencing all of them (Weiten, 1995). The early factor models proposed by Spearman for intelligence focused on one core factor, *g*. However, it is clear that intelligence is made up of more than one core attribute as is rugby. Based upon this idea Spearman developed his two-factor theory of mental tests. This theory was later modified to consider several common factors, plus a part specific to the variable in question. Thurstone (1938, 1955) emphasised that intelligence involves multiple abilities and identified seven different factors, via factor analysis, called primary mental abilities. As a consequence, in order to produce a single rating for an individual rugby player, the skill components that represent performance must first be identified, just as Thurstone did in identifying the primary mental abilities of intelligence. This forms the basis of statistically sound Key Performance Indicators (KPI's). The applicability of Spearman's general approach is further expanded upon, with respect to the rugby context, in section 3.5.1 of Chapter Three.

It could be argued that the dimension reduction techniques employed in this thesis only serve to identify variables that correlate well. However, provided the combination of these correlated variables can be considered to represent some core skill, as shown in the second layer of Figure 2.1, then this approach is justified.

Returning to the main argument, having discussed potential methodologies and the underlying philosophy for combining measures of performance, it is important to examine how the events in a match are operationalised. To measure sporting ability, we need to measure quantitatively the sporting skills of an individual. This is achieved by measuring the performance on physical tasks. In terms of the problem presented in this thesis, the definition of ability is workable with the given data structure. Quite simply, one can infer the level of an individual's skill from the univariate data relating to physical tasks in our possession. The skill set of an individual to tackle, catch, kick, and evade defenders is evident within the data set. A note of caution must be included at this point: It would be naive to assume that an individual's ability to play the game of football could be completely described by a set of numerical measures. However, a large portion of an individual's talent is expressed, exemplified by such statistics. The statistics that are detailed in this thesis are associated with interactions between an individual and the ball while co-operating with team members, but there is a much wider area involved, associated with off-the-ball play as shown in the following section.

Quantification of performance in individual sports is not considered in this thesis, as the interactions between co-operating individuals are non-existent, making the task of evaluation noticeably easier. Performance by an individual in individual sports can be extrapolated from the outcome of the event, but the interaction between co-operative individuals in the competing environment that is team sport provides a more interesting challenge. This is explored in the next section.

2.3 Quantification of Sporting Performance in Team Sports

A number of issues are involved with the quantification of performance in team sports. This section looks at the key areas of data collection, selection of numerical measures, the variety of measures between and within positions, validation of measures, measurement categories and types, and commercial systems.

"One major characteristic of the past two decades has been the emergence of science as an important factor in football coaching (Kuhn, 1995)." As a result "a considerable amount of research has been devoted to establishing the need for objective forms of analysis and their importance in the coaching process (e.g. Franks & Miller, 1986; MacDonald, 1984). They have clearly established the difficulties facing any single individual attempting to analyse and remember objectively the events occurring in complex team games (Hughes, 1991)." These difficulties include the decision of which information is important in a game context (Patrick & McKenna, 1987). The second problem lies in the efficient recording, storage and general handling of generated data, however with the availability of computer technology this is off set somewhat (Franks & Miller, 1986). The first problem, objective quantification, is avoided through the consultation with appropriate experts. Consultation with experts enables the identification of key attributes that define performance. For example, Eagle Sports call upon former All Blacks and coaches of first class teams, Grant Fox (Auckland Assistant Coach) and Wayne Smith (Former All Black Coach) in this consultation process. This process is highly subjective and is shaped by the experiences of the expert involved. Once these attributes have been identified, it is then the role of the statistician to define how the information is extracted and utilised in a meaningful way. In particular, statistical analyses will enable the most important attributes to be identified. This thesis uses factor analysis, self-supervised networks and self-organising maps for this purpose.

2.3.1 Data Collection

Having established the need for quantitative analysis it is important to investigate how such numerical measures of sporting performance are selected and collected, in order to understand the strengths and weaknesses of such an approach. Notational analysis, video analysis and data recording for the collection of such data are a huge area of sports science (Hughes, 1987; Treadwell, 1987; McKenna, Patrick, Sandstrom & Chennels, 1987; Patrick & McKenna, 1987).

Coaches, scouts and managers have adopted, designed and developed systems for gathering information for many years (Hughes, 1991). A review of the conference proceedings from the World Congress of Science and Football illustrates this, with an entire section devoted to this area in 1987 and again in 1991. The 1995 conference also featured articles implementing these systems, indicating the acceptance of these methods. This vast area of research will not be addressed; instead the focus will be on the potential output of notational analysis and video analysis. "Analysis of performance is central to the coaching process and has evolved over the past decade into a discipline known as notational analysis (Townend, 1996, p7)." The need for objective forms of analyses is mentioned by Townend with reference to a study by Hughes (1996) which found that coaches recalled incorrectly up to 70 per cent of what happened in games when their accounts were compared with videotapes. This led to the conclusion that match statistics resulting from a notational analysis are more reliable than an 'expert witness'.

Briefly, video analysis provides a reusable input for assessing performance, enabling every player to be assessed (Rattue, 2000; McKenna *et al.*, 1987). Notational analysis provides the conversion of match play, either live or from recorded footage, to raw statistical data (Treadwell, 1987). There is nothing particularly spectacular about either of these methods. Video analysis essentially requires repeated viewing of the footage. However, to be of use the "viewer" needs to be an expert such that relative information, albeit subjective, can be extracted. Notational analysis can be more objective, and typically involves the collation of information with regard to prespecified criteria, such as tackles made, or passes given. Typically only physical tasks are examined. Some information is lost through this process, namely off the ball activities. Specifically, video footage tends to focus on the ball, thus ignoring the contributions of individuals not in the visual frame. "Watching a player at the ground still gives you the best indication because you can see all the off-the-ball stuff (Wayne Smith in Rattue, 2000)." This limitation must be acknowledged.

Notational analysis systems are readily developed, modified, adapted and implemented to meet the requirements of a given coaching regime. The number of coaches, senior level and above, who have members of the squad collecting specific statistics on the sidelines of a match in progress is a testament to the adaptability and applicability of notational analysis. The key principles underlying notational analysis are objectivity and adaptability. Objectivity is crucial because subjectivity may cloud the truth, as shown in Hughes' study mentioned by Townend (1996) and outlined by the NZRFU (1991). The adaptability of notation analysis allows widespread application. For example, a coach may be concerned about the work rates of his tight forwards. By recording the number of rucks and mauls that targeted individuals attended, objective evidence is obtained. The creation of a notational analysis system is essentially very simple and based on two criteria. Simply, the *information* required is restricted by the *resources* available, that is codes could be created for almost every conceivable occurrence (tackle made from the left hand side using the left shoulder enveloping the ball) but to do so requires a great deal of time and effort. This balance requires careful selection of numerical measures in order to maximise the value of the information whilst minimising the cost of the collection procedure.

2.3.2 Selection of Numerical Measures

A review of the literature relating to the selection of numerical measures for individual rugby playing ability and performances involves a precursory glance at the potential of objective quantification of performance, rather than demonstrating how this area has been advanced. More often than not the researchers mention the lack of literature in the area of performance measures. The data chosen represent three distinct key areas; physiological, psychological and technical skill, and can be either objective or subjective. All three of these key areas measure important components of an individual's perceived skill set, which characterises performance and generates the impression of ability. The relative importance of these categories is dependent on the sport, level and intended inferences.

Arnold *et al.* (1980) took 14 anatomical and physiological measurements of 56 scholarship players of American football in 1976 with the intent of comparing these physiological measures with perceived ability. The perceived ability recorded is based upon the subjective measurements of player ability as determined by coaching staff. Polynomial regression, step-wise multiple regression and correlations were used to identify six measures as the best predictors of football ability.

Turning to soccer, Franks *et al.* (1999) used six physical and physiological measurements of 64 international youths and an ANOVA to investigate any differences between successful and non-successful individuals. A successful individual was defined as one who had been offered a 'professional' contract. No significant differences were identified.

Similarly, and more applicable, was a study by Pienaar *et al.* (1998) identifying and developing rugby talent among ten year old boys in South Africa. The methodology applied by these researchers was identical to that used by investigators of individual sports. The ability of an individual was assessed subjectively by the coaches and selectors. In total, 14 physical and motor tests combined with 14 anthropometric measurements for each individual were used to establish the significant indicators of rugby talent. A stepwise discriminant analysis identified eight variables that discriminated maximally between talented and non-talented individuals. The specification of non-talented and talented was based upon whether or not the individual was in the top three teams out of 22 participating in a school-boy rugby tournament. A canonical analysis was then undertaken using the identified variables and the players were ranked according to the first canonical component. These results were then compared with the representative side comprising the 'best' players, selected from the same pool of players from all 22 teams, with a success rate of 88%. There is a danger in attempting to predict rugby talent at such a young age due to the physical growth and the impact that growth has on how an individual creates and applies their potential skill set. However, at a later age group, once physical development has levelled off, physical attributes tied to specific objective measures are potentially more useful.

Patrick and McKenna (1987) developed a computer system specifically designed to allow quantitative and qualitative data to be entered directly via a special keyboard for

Australian Rules football. Only univariate summary measures were produced by this system. However, the authors demonstrate the usefulness of such statistics by pinpointing areas of game strategy that needed to be improved to facilitate winning. Of one game analysed it was noted the opposition team gained a lot of possession uncontested, with the home side convincingly beaten. Fine tuning tactics to place more pressure on the opposition resulted in a much closer result in the return match. A natural progression is to look at the key components of winning and losing from a large number of games. From this study the authors "were unable to differentiate between successful and unsuccessful teams on the basis of a simple summation of match statistics. "Obviously a complex interaction of many variables affects the game (Patrick and McKenna, p. 273, 1987)." This indicates a need for multivariate analysis as an extended layer of match analysis.

At the same conference McKenna and Patrick teamed up with Sandstom and Chennels (1987) to deliver a paper analysing activity patterns in Australian Rules football. Videotape analysis linked into a computer system was used to collate the data. The purpose of this study was to identify the game demands on Four Victorian Football League players. In this instance game demands refers to the physiological effort required and no attention was paid to the technical skills of the individuals. Data reduction was used, but in this case only to sum relevant categories and produce a ratio of work to rest. Following the analysis recommendations were made in terms of structuring the fitness training required for Australian Rules football.

2.3.3 Variety of Measures Between and Within Positions

More importantly from the paper by Pienaar *et al.* (1998) is the mention of a lack of performance measures. The researchers bemoan the lack of research into performance criteria and cite "Salmela and Regnier (1983) who noted the isolation of performance criteria is more crucial for team sports, in that there is an even greater variety of tasks (compared to individual sport); the problem becomes even more complex in team sports where mini-performances must be conceived of within a team context, considered against the strengths and weaknesses of other team-mates and the given demands of each position (p. 697)." This statement has important implications on the pre-processing of the data set presented in this study. Further, in complex team sports where a co-operative contribution is required to ensure the success of the team, performance of an individual may be obscured as that individual compensates for the strengths or weaknesses of other individuals. Moreover it accentuates the fact that the specific sub-groups of a rugby team need to be acknowledged, such that the typical jobs of an individual can be identified.

In a study that did differentiate between positions, albeit very generally identifying just backs and forwards, Deutsch *et al.* (1999) compared the work rates between top club and Super 12 rugby union players. Their opening statement drove their research, "The work rates characteristic of professional rugby union players have yet to be determined. Without information regarding the demands of play at these standards, and the difference between playing positions, specificity in testing and training is unlikely to be achieved (p 810)." The basis of this study was not to assess the skill set but to determine the distribution of exercise and rest periods throughout the match. However, it is important to note how the skill sets and hence performance measures of different positions will differ, just as the physiological demands differ. Video data of 67 players, from either Dunedin first grade Club competition or the Otago Highlanders Super 12 team, was collated over two seasons. Using an ANOVA it was found that the work rates of backs and forwards are significantly different at both levels. Not surprisingly the work rate of forwards exceeded that of the backs. The only difference between club and Super 12 was the workrate of backs, which was lower for Super 12 participants than in club rugby. Important is the reference to work rate, as this infers involvement. Ultimately it is the quality of this involvement that we are seeking to quantify.

Treadwell (1987) undertook research investigating physiological demands in sport, but included rugby with the intent of demonstrating methods and procedures used in data collection. One game played at Cardiff Arms Park was analysed. The author identified groupings of individual positions as follows: inside backs, midfield backs, outside backs, tight five forwards and loose forwards. These natural groupings are important when making comparisons with the cluster analysis performed in Chapter Three as part of the data cleaning process. More examples of subjective clustering by interested parties are detailed in Chapter Three as part of the debate regarding positional clusters.

2.3.4 Combined Measures

Factor analysis provides a means of combining several performance measures into a single more reliable and more powerful rating of effectiveness. An example of the use of factor analysis in rugby appears at the Fourth World Congress of Science and Football. Quarrie and Williams (1999) used factor analysis to identify pre-season attributes of 258 rugby players. Two factors were identified, endurance and speed/power, obtained from variables consisting of a battery of anthropometric and physical performance tests. The variable from these factors with the highest loading was then used in a regression analysis as the dependent variable against a number of categorical variables. This included self-rated playing ability which was obtained from a questionnaire eliciting information regarding rugby playing and injury experience, training patterns, health and lifestyle patterns. This was then used to identify the significant components of pre-season fitness.

At the same conference Hodge (1999) also employed factor analysis, as well as cluster analysis to examine the motivational profiles of individuals in rugby union. 257 adult males were assessed on a number of psychological variables, based on three key areas, goal orientations, perceived ability and self-concept of physical ability. From the factor analysis, cluster analysis was used to identify groupings, relating to scores based on task and ego subscales.

A comparison between the application of sport science in professional soccer in Germany and Australian Rules (AFL) was the basis of research performed by Kuhn (1995). Most important was the illustration of the demand profiles of each sport. The emphasis in the AFL was placed on skill. However, in soccer higher regard was attributed to physiological attributes. This in some way clarifies the slant of the literature relating to soccer performance. The continual emphasis on physiological measures is undoubtedly due to their perceived importance in determining an individual's performance. This in some way may have affected other similar sports, such as rugby, where skill would normally be afforded the greatest respect. Within this article there is reference to evaluation of performance through subjective/objective diagnosis, laboratory/field tests and the use of videotape. Again the emphasis employed for the AFL was upon game evaluation, which was neglected to some extent in the above German soccer league analysis.

Chervenjakov (1987) investigated the playing effectiveness of soccer players in Bulgaria after developing methodology whilst coaching in Japan. Player activity was recorded using 34 codes, inclusive of skill based parameters. The summations of these codes from a game were then combined into six scales. The weights of the 34 codes within each of these scales were based on the opinions of top-rated Bulgarian club soccer coaches. The scales were developed with the intent of reducing the amount of data required to illustrate the full effectiveness of player. This demonstrates the need for dimension reduction techniques in a sport science context.

Pollard, Reep and Hartley (1987) applied dimension reduction in producing a paper incorporating factor analysis to enable an objective comparison between team playing styles. Six general style-of-play variables were collated from 32 World Cup (1982) matches and 42 English First Division (1984-85) matches. Teams that played only once were excluded, leaving 32 teams for analysis. The general style of play variables were calculated by noting the occurrence of the following ball orientated events: long forward passes, long goal clearances, centres (number of attacks reaching opposition half), regaining possession in attack, possession in defence, and multi-pass movements. The interpretation of the six variables is determined by a number of strict guidelines. It is not necessary to include these here. Factor analysis was used to identify teams with a similar playing style. Using a varimax rotation three factors were obtained explaining 92.5% (60.0%, 18.8% and 13.8% respectively) of the variation in the data matrix. The analysis produced some clear differences, thus enabling coaches to modify subjective interpretations of opposition teams and develop tactics to combat the relevant style profile. The authors commented that "The analysis described can be considered as a first step towards a comparison of playing styles based on objective quantitative measures instead of the more usual subjective assessments (p. 314)." Again the potential of dimension reduction techniques in interpreting player skill sets, manifested and projected in a team situation in this case, is illustrated.

As mentioned earlier, performance is the physical manifestation of ability, which is a combination of separate skills, identifiable from the completion of physical tasks. Further physical task-based data collection must also consider the fact that the variables are likely to be correlated and thus the data set will be exposed to redundancies. As shown by Pollard *et al.* (1987), the power of the data can be improved by implementing methods such as factor analysis, which in effect humanise the data and impending results. That is the data becomes easy to interpret, either visually or through appropriate analysis, making it more useful for individuals at whom the data is aimed, namely coaches, selectors and other interested parties. Alternatively, ability can be viewed as a latent variable, which is impossible to measure directly because it can only be observed, indirectly, through various physical task-based variables.

Continually the same themes occur in research that deals with ability in complex team sports such as the football codes; the comparison of numerical measurements (physiological, technical skill or psychologically based) with the perceived ability of an individual. However because ability has not been quantified objectively, researchers have had to make do with subjective and self-perception measures. Further, it is demonstrated that the methods applied in part one of this thesis, factor analysis and cluster analysis, are used already in the rugby research environment.

Interpretation is a key facet of statistical analyses. The need for interpretation necessitates the use of dimension reduction techniques in a team sport environment. Clearly, performance in a team sport environment is multifaceted, highlighted by the discussions in this subsection. From a contextual perspective, rugby can be broken down into key areas, relating to job description (Appendix C). Recurrent sub-groups of skill sets such as attack, defence and set-piece play dominate the impression of an individual's performance and perceived ability. Each of these themes can be expressed via a collection of statistics, such as players beaten, tackles broken, metres run and the number of runs per game. Intuitively, aspects of each facet of play will be related. That is if a player is good defensively, he will make a number of tackles, tackle assists, turnover tackles and tackles over the advantage line. Chapter Three discusses the multivariate nature of the data used in this thesis with respect to the relevant methodologies adopted.

2.3.5 Validation of Measures

The validity of the data measures used has a direct bearing on the quality, reliability and generaliseability of the results. Discussions of validity are most appropriate when contextual. The description of the data in Chapter Three introduces the data used in this thesis and the procedures that ensure its validity. However, previous research, which is introduced next, also emphasises the importance of validity.

Tumilty *et al.* (1999) provided an interesting study looking at the correlation between the subjective measures of player ability as determined by two experienced coaches based on six game-related qualities and a variety of game-related laboratory and field tests. Game-related situations refer to events and scenarios that are likely to be encountered in a match situation. Only nine elite youth players were examined and no significant correlation was obtained between the coaches' opinions and the objective testing measures. It is of importance to note that the data generated by the laboratory testing was not game-related, only representing a small component of the individual's skill set and not applied to a game situation. The authors acknowledge this limited realism. However, this study hints at a useful check for any models using expert opinion. An obvious frame of reference for validation is between coaching staff opinion and objectively created statistical models. The comparison with 'expert' opinion is part of the Eagle Rating validation process.

Based on the need for accurate and objective feedback, Hughes, Robertson and Nicholson (1987) developed a system that enabled rapid and objective analysis of soccer. This was derived from 24 skill-based actions. Data was collected from the 1986 Soccer World Cup with the results from successful and unsuccessful teams compared. Successful teams were defined as those that reached the semi-finals, whereas unsuccessful teams were eliminated after the first round. Comparisons were made using ANOVA's and provided only simplistic recipes for victory, for example longer retention of possession is an important component of winning. It is likely that a multivariate analysis of these data would have produced more useful results.

2.3.6 Rating Systems for Individuals in Team Sports

It is evident that a multivariate approach to quantifying sporting performance is a relatively rare approach. Fundamentally, this is due to the design and requirements of distinct rating systems. This thesis examines performance from an individual level, as required by Eagle Sports due to the requirements of coaches, media and other interested parties. There are a number of studies that do not examine the components (individual athletes) that contribute towards the collective ability of the team. Stefani (1998) reviews a number of different rating systems applied in a variety of sports from three broad categories – combat sports, independent sports and object sports. As rugby is an object sport - "each competitor tries to control an object in direct competition with the opposition (e.g. soccer and chess) (Stefani, 1998, p249)" - only this category is applicable as the aims and measuring systems relating to combat and independent sports are incompatible with rugby. Further, the categories are rather broad. Object sports such as chess and rugby have little in common. The overriding theme of the rating systems reviewed by Stefani is the dependence on outcome. As mentioned earlier, outcomes in rugby are not easily defined making it impossible for Stefani's system to be applied on an individual level. Instead the performance on physical tasks (incorporating psychological, physiological and technical skill successes) provides the basis for the Eagle Rating system, hence the need for multivariate techniques.

More applicable to this thesis is the underlying design of rating systems and performance measures. Sports rating systems for team sports are based on one of two premises, either focused on individualistic contribution or team contribution. The distinction is quite pronounced. Individualistic rating systems, such as the batting average[1] in cricket, depend entirely on what an individual has accomplished – with no regard to the team cause. This is a fundamental oversight, as in team sports at the highest level the pursuit of victory is important. Essentially, individualistic systems assume that because an individual has performed well, their team has benefited accordingly. Stern (1998) emphasises these aspects by questioning who should be rewarded for a touchdown pass thrown by the quarterback in gridiron – the quarterback for the pass, or the receiver for the catch and run? Clearly the same issues evident in gridiron are also present in rugby union, namely that each member of the team must contribute to the team by performing appropriate tasks when required. For instance, for a wing to score a try after a flowing back movement, almost every individual will have been involved. Described simply, the process originates from the forward pack, who ensure that the ball was secured and presented to the half back cleanly; to the inside backs who provided quick accurate service to the midfield backs, who ensured the direction of attack was straight enough to commit the opposition defence; and finally the outside backs who use pace and evasion to cross the try line.

[1] An individual's batting average is calculated by dividing the total number of runs scored by the total number of dismissals.

However, gridiron focuses on differing univariate aspects of performance relative to the individual's primary role, as this is sufficient to describe performance. That is, the task of the receiver is to catch the ball and run. Therefore, his performance can be determined by how far he ran with the ball in hand.

Conversely, rating systems based on team contribution consider what an individual has done to aid the team's pursuit of victory, or at the very least how the individual's performance has contributed to the team's performance. This is the basis of Bracewell's (1999) performance indices for both batting and bowling in cricket. Further, the more computational intense Coopers-Lybrand ratings (formerly Deloitte Ratings) (Berkmann, 1991) consider individual performance relative to the team in cricket with consideration given to extrinsic variables such as pitch conditions. Similarly, baseball's player game percentage (PGP) estimates the value of baseball player performance based on the player's contribution to victory (Bennet, 1992). However, sports like baseball and cricket have clearly defined (or definable) outcomes for each interaction. That is each time a ball is delivered a finite number of events occur, such as batsman misses or hits, is dismissed or not and so forth. This is not established easily in games like rugby.

Before a team contribution rating system can be developed for rugby, an individualistic rating system must be created. This allows the structure of performance to be understood. Understanding performance leads to inferences about ability. This is the cornerstone of the rating problem. Without suitable performance measures, suitable rating systems cannot be constructed. Also, the media demand a rating for each player in each match. Therefore, quantifying individual rugby player performance on an individual level is a necessary component of a contribution system.

Additionally, a team rating system must also be developed. Due to the nature of rugby, given that at any one time on the field there are at most 15 co-operating individuals on any one team, it is necessary to understand the interactions between these individuals. These interactions relate to how individual performances combine to produce the team performance. This thesis tackles the initial problem of quantifying rugby performance on an individualistic level. Once this has been done, the same techniques can then be applied to team data. Theoretically, this would then enable player interactions within different sub-units (NZRFU, 1998) of a team to be assessed.

Rating systems for teams in various sports are described by Stefani (1998). However this is a different problem to that proposed in this thesis, although the same techniques could be implemented at a team level to quantify team performance. As mentioned previously, this thesis focuses on the individual level. Accordingly, a review of team rating systems from anything other than a multivariate approach is redundant. Due to the applicability of such systems to a commercial audience, a review of commercial systems is more applicable.

2.3.7 Commercial Systems

Eagle Sports has provided the commercial backing for this thesis in return for the development of performance measures. An overview of the data collection system employed by Eagle Sports is given in the data description section in Chapter Three. The Eagle Sports database is based on notational analysis incorporating some 130 physical task-based measurements, two subjective measures relating to judgement and execution and spatial data relating to territory. A number of other time-dependent statistics, referred to as second generation statistics, can be extracted from the database and related to the outcomes of events. The database includes over 1000 first class games and was continually updated prior to the death of Trevor Eagle in December 2000. Aspects of the database can be viewed by the public via the Eagle Sport website (www.eaglesports.co.nz). The objective of the statistics collected is to give interested parties an objective analysis of events that take place on the rugby field. As the Eagle Sports Statistical System is a commercial operation, full details cannot be given but the following methods are akin to those utilised by Eagle Sports.

An approach similar to that employed by Eagle Sports was described by Partridge, Mosher and Franks (1991), where a list of 38 key skill-based events was recorded during actual match play. A total of 52 games, covering the World Cup and World Collegiate soccer championship, televised by the Canadian Sports Television Network (TSN) were examined. The goal of this study was to determine if there were any differences in playing level for these two competitions. A Hotelling T^2 test was used to identify any significant differences. Results indicated several significant differences with the conclusion that the lower ranked collegiate teams need to be selective when attempting to replicate styles of play presented in the World Cup. This effectively mentions that the generality of models developed from World Cup soccer will not carry over to collegiate soccer. This is a notable caution that can be extended to other sports.

A major competitor of Eagle Sports in the provision of sport statistics is a New Zealand based company, AnalySports, based at the New Zealand Institute of Rugby at Massey University's Palmerston North Campus. Statistics are provided at no cost to New Zealand's 27 Provincial unions although the package is designed for televised games and is subsequently aimed at the 10 first division teams. Video analysis is converted to raw data through notational analysis. A notable inclusion in the data collection sequence is subjective data relating to quality of performance. Through the AnalyRugby system, match footage is converted to quantitative data. The 'statistical' events are linked directly to the match footage through digital technology. That is using the computer interface provided, users can access and view the relevant task associated with a particular statistic. For example, a coach can view all the missed tackles made by a certain player in a certain game simply by clicking upon the relevant window displayed on screen (AnalySports, 2000).

Commercial systems have the benefit of greater resources, enabling the collection of more in-depth information. Consequently, it is not necessary to delve too deeply into the notation systems of non-commercial systems. The level of information collected by non-commercial systems generally lacks the sophistication or refinement of commercial systems, or is irrelevant due to assessment of sports other than rugby union. Secondly, information regarding successful systems employed by coaches or commercial bodies is not readily available, as a vital edge over competitors may exist within these systems and those who use the technology, namely the coaches, do not partake in formal sharing of information or subject ideas to peer-review. But, in general, quantifying performance objectively through multivariate techniques has not been attempted at the level required due to the lack of data. This highlights the novelty of the Eagle Sports process in the area of quantifying ability and emphasises the importance of its database. Importantly, as will be demonstrated in Chapter Seven, the use of statistical systems must be kept in perspective, so the end user is aware of what is being measured and the capabilities of the employed system in order to extract the best information. Consequently, the processes for implementing statistical systems are as important as the statistical systems themselves.

The intent of this section was to firstly justify the use of statistics to quantify performance. This was achieved with reference to coaching documents prepared by the NZRFU. Secondly, this section describes how data is collected using notation analysis, a simplistic, adaptable and effective method, but dependent upon two criteria, level of information desired and available resources. The numerical measures involved generally relate to the skills that the researcher is trying to measure. In this thesis the emphasis is on skill based interactions with the ball and as a result statistics relating to the performance on physical tasks are collated. This utilises the assumptions defined in section 2.2, that to complete a physical task successfully, the physiological and psychological tasks must also be successfully completed. The variety of measures between and within positions on these physical tasks creates the need to look at clusters of positions that may be evident within the team. Methods of combining statistics, utilising techniques such as factor analysis, are discussed and the section concludes with a summary of relevant commercial type activities.

The next section considers the use of performance measures for monitoring performance and ability. Reviewing possible analyses, the first question is what does the data represent? The idea of monitoring an individual's performance can be likened to quality control (Mosteller, 1997).

2.4 Quality Control as Individual Performance Monitoring in Sport

As indicated above, there is very little literature on the topic of performance or ability measures for sports, particularly team sports. Even less literature exists detailing the potential use of quality control theory as a tool for monitoring the performance of athletes. Mosteller (1952) utilised unbiased estimation for quality control developed during World War II to find the probability that the better team in the World Series would win (Mosteller, 1997). The better team is determined from prior performance. Using a theorem developed by Girshick, Mosteller and Savage (1946) to obtain the unique unbiased estimate for $\hat{p}$ for a binomial process, Mosteller estimated the probability of a team winning the best of seven series. However, establishing the probability of success from historical result data ignores the underlying contribution and interaction of individuals. A more sound and rigorous proposition is to evaluate individual performance, then establish how the individuals combine to form a team performance.

This thesis only considers evaluating individual performance. Only Bracewell (1999) has applied quality control to sports data with the specific purpose of monitoring the performance of individual sportsmen. This research focused on the performance outputs in cricket. Utilising the assumption that an individual's natural ability is expressed via performance outputs, this research sought to describe and understand the underlying statistical processes of player performances. After determining that the assumptions of control charting procedures were met, that is uncorrelated, random and approximately normally distributed data, the data were used to monitor the estimate of natural ability via widely accepted control methods, such as Shewhart control charts, CUSUM, EWMA and multivariate versions of these procedures.

Quality control is applicable in a sporting sense as "the individual athlete is viewed as an ongoing process to be improved (Roberts, 1992, p 27)." Consequently, quality control techniques are applicable for monitoring this process. It is the use of multivariate quality control statistics that provide a very useful application in this thesis due to the multifaceted nature of ability. However, a major problem is encountered due to the basic structure of these charting procedures. Bracewell (1999) avoided such a problem by dealing with mostly two variables, so that a simple bivariate plot could be used to show inferior or superior performance. Given that standard quality control procedures consider the difference from the average (Montgomery, 1997), as they stand they are clearly insufficient for the scenario presented by sports data when the number of variables considered exceeds two. This is due to the difficulty in identifying possible causes for an alarm. Given one or two variables, graphs can demonstrate the reasoning for an alarm, due to either superior or inferior performance. "A unique situation arises with the study of sports data in relation to quality control. For any given standard it is preferable to be better than the average (Bracewell, p. 44, 1999)." Thus control charting procedures need to be adapted to cater for this alteration in thinking. Chapter Three looks at this in greater depth. The relationship between the factor analysis and multivariate control charting as described in Chapter Three is again an extension of quality control theory. Jackson (1980) recommends the use of control charts based upon principal components, which are linear combinations of the original variables.

Further expansion of the ideas relating to performance monitoring with relevance to sport and the importance of using data summarised via dimension reduction techniques is explored further in Chapters Three and Six. This enables the themes to be expressed fully with the introduction and discussion of appropriate methodology.

This chapter, while reviewing the limited literature available on the area of quantifying individual sporting ability and performance, serves foremost as a further indication of how this research is shaped. Throughout this thesis, contextual debate discussing the relevance of methods and results is presented. This was especially relevant in the first two sections, where justification for the use of statistics and what the statistics represent for performance appraisal in rugby was discussed. The philosophical justification for collection of statistics relating to performance attributes was outlined in the initial section. The second section defined ability and performance. Using these definitions it was implied that performance could be quantified using the methods, notably notational analysis, described in the third section. The final section touched on the implication that sport performance is a concept parallel to industry process performance, encouraging the use of quality control techniques for monitoring performance. This discussion leads into the following chapter. Chapter Three uses cluster analysis, factor analysis and quality control theory mentioned in this chapter to create the first version of the Eagle Rating, a measure of individual performance for rugby players.

Chapter Three

Multivariate Analysis, Quality Control, and the Birth of the Eagle Rating

3.1 Overview

This chapter provides the first of two applied components in this study. The purpose of this chapter is to demonstrate that complex individual sporting performance in a game such as rugby can be quantified using standard techniques for multivariate analysis. The emphasis is placed on skill based interactions with the ball. Chapter Two showed that this is a novel approach. The second applied component in this thesis, implementing non-linear capable dimension reduction neural network techniques, is covered in Chapter Six.

The objective of this chapter is to show that a point estimate for individual rugby player performance can be represented by a single numerical value calculated using multivariate analysis. As ability is derived from performance, performance must first be quantified. Having established a suitable method for extracting performance measures from match data, it is then applicable to use a series of performance measures to estimate an individual's ability. Essentially, this chapter chronicles the early development of the Eagle Rating, thereby providing a reference for techniques applied in later chapters. To achieve this new rugby statistic, three distinct steps are required.

Cluster analysis, employing Ward's linkage and a squared Euclidean distance measure, is initially used as part of the data cleaning process to identify groupings of positions. Identifying distinct groups of positions is crucial due to the different demands and specifications placed on each position during a rugby game. The argument presented depends largely on a rugby perspective. Contrasting coaching viewpoints on potential positional clusters are coupled with statistical techniques to identify the most suitable way of grouping positions.

Secondly, factor analysis using iterated principal component analysis and a varimax rotation is used for dimension reduction. A generic template is adopted for each of the positional clusters to aid the marketing of the product. Idealistically, the factors obtained will relate directly to specific aspects of match play such as attack, defence and so forth. This hypothetical structure is suggested by the schematic representation of performance in Figure 2.1. Consequently, the factors could then be interpreted as Key Performance Indicators (KPI's). KPI's are recognised and widely used in the coaching environment as measures of performance. The identification of models congruent with rugby thinking ensures that a relevant and marketable statistical package is produced.

Finally, multivariate quality control in its guise as performance monitoring is developed and used to produce a single statistic, the Eagle Rating, for each individual performance per match. Hotelling's T^2 control statistic is modified by placing the emphasis on perfection rather than the average. This produces a multivariate control technique based on a Mahalanobis distance which is suitable for application in a sporting environment.

Essentially, this chapter is the pilot study for the Eagle Rating. It is experimental and was performed under pressure to deliver a saleable product before the start of the Super 12 competition in February, 2000. A performance rating had to be extracted to satisfy the demands of Ultimate Rugby (see Appendix D) and this chapter reviews the steps taken to create a working prototype.

Work relating to non-performance discussed in this chapter has been published in *Research Letters in the Information and Mathematical Sciences* (Bracewell, 2001). Additionally, research relating to the perfection method has been accepted for publication in the September issue of the *Journal of Sports Sciences* (Bracewell, 2003).

3.2 Data

The variables and coding structure used to quantify performance in this chapter are described in greater detail in Appendix A. 34 Test matches played during 1999 provided 762 observations from individuals who participated in the entire game (80 minutes). This covered 267 different individuals. The data was focused on matches played by four of rugby's world powers: New Zealand, Australia, South Africa and England, and also included their opponents. This data set provided the basis for the introduction of the Eagle Rating in the fantasy game, Ultimate Rugby during the 2000 Super 12 competition (Appendix D). This segment gives a brief overview without compromising the commercial aspects of the business. Sportsnet Limited (Eagle Sports), a division of Eagle Technology Group Limited, has collected and collated raw statistics related to on-field performances of players and teams in the game of Rugby Union since March 1996. To date more than 1000 first-class games have been recorded. The data collected includes games played at first-class (see Appendix E) level world-wide, covering major tournaments such as the World Cup, Tri-Nations, Five/Six Nations, Super 12, and the Currie Cup. Collection is limited to televised games. Only men's rugby is analysed due to the limited televised coverage of women's rugby. The emphasis for data collection is placed on televised first-class matches involving New Zealand teams or major New Zealand opponents. Over 120 separate measures of univariate performance have been collected in raw form (Appendix A), including measures such as the number of tackles made, tackles missed and passes given. These measures are distributed as reports to coaches (including the All Blacks), media and fans for a subscription fee (www.eaglesports.co.nz).

For the purpose of this thesis, an observation relates to the on-field participation of an individual for an entire game. The variables are individual totals for each activity for a game. This quantifies performance on a match by match basis.

Quality assurance is built into the data collection process ensuring that recorded data is accurate. Several steps in the data collection process maintain the quality of the data. These are described next.

Selection of Coders

The most important stage of the collection process is the selection of the coders. Individuals employed to record the data must have a full understanding of rugby – the rules, structure and flow – enabling them to distinguish between a ruck and a maul, identify the advantage line and so forth. In the initial training phase there is a high turnover of prospective employees as individuals with inadequate rugby knowledge are eliminated.

Clearly Defined Codes

The codes that are used are clearly defined and structured so that no ambiguity exists, leading to high between-coder reliability. The structure of the codes is governed by the revolutionary database design developed under the direction of the late Trevor Eagle (Potter, 2000).

Audits

Regular examination of the work submitted by the coders enables both within-coder and between-coder reliability to be monitored. Testing is carried out where a specified segment of a game is coded independently by all coders. Inter-coder reliability is expected to exceed 95 percent. Repeated testing using the same segment of play enables intra-coder reliability to be assessed.

On-going Training and Evaluation

Coders are required to attend regular training meetings and undergo testing to ensure the quality of the data is maintained. Further, coding rehearsals are encouraged. Due to the nature of the close-knit, team orientated working environment, debate and discussion regarding coding issues and interpretation are promoted, leading to a thorough understanding of the coding process.

Discussion

Rugby is a dynamic game. Unusual occurrences of events that occur outside of the defined codes may require special handling. Working in a team environment promotes discussion, which leads to clarification, prompting between-coder reliability and a greater understanding of the codes and structure of rugby.

Sequencing

After the game has been coded, it is sequenced by an individual who did not code the game. This involves identifying passages of play, analogous to phases (see Appendix E) except sequences are cumulative. This involves "reading" the game from the written transcript and serves as an additional check for coding accuracy. As the code syntax is well established and sequencers are familiar with the "structure" of a game, anomalies are identified and rectified. An example of a coded segment of play is provided in Appendix A.

Data Entry

The final step in the data collection process is entering the data into the system. A specialised programme, tailored to the coding system, prevents any distortion of codes as they are entered. As the code syntax is well established and data entry personnel are familiar with the "structure" of a game of rugby, anomalies are readily identified. The data is then stored in a Microsoft Access database, with an interface enabling specific data to be accessed via the internet.

A selection of 31 variables is used for the initial quantification of performance, covering the main aspects of play (infringements counts as one variable). These variables can be broken up into a number of key areas as shown in Table 3.1 on the next page.

A number of other variables can be generated from the variables listed in Table 3.1, notably tackling precision and frequency ratios such as breaks per game. Whilst attack and possession variables measure essentially the same activities, there is a fundamental difference in the interpretation. Attack variables relate specifically to back type play. This generally occurs with more space available to participants and ordinarily has more flow. Conversely, possession variables relate to typical forward play, specifically the hard grind that occurs close to a breakdown and general in-close physical contact with team members and the opposition. Backs can participate in possession based activities just as forwards can engage in attack based variables. For instance, when All Black number eight Scott Robertson enters the backline outside the second five eighth and encounters the opposition defence he is credited with attack metres. However, when Robertson receives a pass close in to a ruck and tries to bust through the opposition defence, possession coding is used.

Attack

- Metres Run
- Runs per game
- Metres per break
- Players beaten
- Tackles broken
- Tries scored
- Line breaks

Possession

- Metres Run
- Runs per game
- Metres per break
- Defence beaten
- Hit ups
- Tries scored

Defence

- Tackles
- Advantage line tackles
- Turnover tackles
- Tackle assists

Lineouts

- Jumps won
- Jumps won against the throw

Loose play

- First three players to a breakdown
- Loose ball gained

Infringements

Kicking

- Kicking metres
- Metres per kick
- Kicks per game
- Kick offs
- Kick off retained
- Goal kicking accuracy

Handling

- Passes to hand
- Passes in tackle
- Turnovers

Table 3.1: *Summary of Eagle Sports Variables*

The data is significantly correlated. Table 3.2 shows the Pearson correlations for a selection of variables from the data set described earlier. The variables selected represent key areas of rugby (running, kicking, errors and defence). The correlations are calculated using the 762 observations from the 34 test matches played in 1999 as described earlier. The substantial number of correlated variables justifies the implementation of multivariate techniques.

Variable	Attack Metres	Possession Metres	Defence Beaten	Kicking Metres	Total Fouls	Total Turnovers	Ruck Impact
Possession Metres	-0.39						
Defence Beaten	0.64	-0.25					
Kicking Metres	0.57	-0.36	0.21				
Total Fouls	-0.24	0.23	-0.22	-0.14			
Total Turnovers	0.38	-0.04	0.23	0.29	-0.08		
Ruck Impact	-0.23	0.38	-0.21	-0.25	0.28	0.02	
Tackles	0.04	0.13	-0.06	0.06	0.23	0.07	0.34

Table 3.2: *Correlation Matrix Indicating Correlation Between Selected Variables*

The data described is for individual players in individual matches. However, the cluster analysis of the data is intended to produce groupings of positions with similar roles and skill-sets. Consequently, this assumes that players in the same position have presented similar skill-sets in all matches. This assumption is false due to variable match conditions and constraints which may not require an individual to use certain components of their skill-set. However, it is anticipated that there may be some general traits that are identifiable within the data set, due to generic responsibilities of individual positions; for example first five eighths kicking for touch, and loose forwards arriving in the first few to the break down. But, the expert opinions of coaches and media provide a useful check of the cluster analysis results.

A major issue is how the data is managed and structured to provide a suitable indication of match involvement. Key aspects of data management relate to the handling of outliers and the scaling of involvement to recognise contribution to the team. As the handling of the data impacts on the suitability of methods implemented it is important to broach this subject before discussing potential methodologies.

It is well known that outliers can influence estimates of individual performance by tending to inflate the mean. However, it is desirable to retain the impact of outliers associated with positive events as these provide an indication of the true capabilities of the individual (Bracewell, 1999). Further, in a sporting sense the impact of outliers is necessary as this is an indication of the competitive nature of top-level sport. Athletes either exploit opposition weaknesses and co-operative strengths, or are exploited themselves. Extreme values are important indicator measures. In most instances, counts, such as number of tackles made, are dealt with in the data used in this thesis thereby providing an indication of general involvement.

Scaling individual performances with respect to either team mates or match involvement is unpractical because the game of rugby is vulnerable to differing conditions and influences. Also, outlying performances vary from match to match and raw counts are an important indication of involvement. This highlights the non-traditional nature of the data. Retaining the uniqueness of individual performance in the initial stages of quantifying performance is preferred. This ensures that outlying performances are not distorted, or overly distort the data of others. Additionally the use of the raw measures makes the data easier to handle and reduces the computational expectations placed upon a system. Coaches are interested in assessing the counts of events, as this is more indicative of work-rate on an individual level. When percentages are involved, the coach must also consider the relative work rates of other individuals.

Contribution based measures (such as percentages) place emphasis on game structure, where simple counts and summed counts focus on pure involvement. While counts are also dependent upon game structure they minimise the potential impact of rare events, such as turnover tackles. However, knowledge of the distributional properties (distribution family, mean and standard deviation) could be used to scale the data and emphasise generic positional traits.

Another possible scaling measure is to standardise performances with respect to each match. However, as outlined previously, the effect of outliers is necessary in a raw state because outliers may have undue influence on the scaled data of team mates, distorting the distinction between positions. This also influences the type of transformation adopted and this can impact on the outliers. In this chapter a transformation is applied when approximate normality (or symmetry) can be introduced to the data for that variable; otherwise the data is left untransformed. This is detailed later.

Both approaches, scaling and transformation, are fraught with potential problems. However, the most representative measures of performance need to be adopted. If scaling is required, an obvious solution is to assess the amount of game-play as this provides a situation where between match comparisons can be made. This standardisation could be achieved by dividing the summed counts by a time-related measure, such as minutes or number of sequences (see Appendix E).

The number of sequences in each match obviously varies game to game dependent on many game specific constraints, such as weather conditions, refereeing flexibility and skill level. The number of sequences per match at the Super 12 level is approximately normally distributed with a mean of 151.88 and a standard deviation of 20.23, obtained from all 69 Super 12 matches played in 2000 (fail to reject null, $p = 0.828$). An understanding of the number of sequences per game is potentially useful. This is because to give an indication of overall contribution and involvement the number of sequences could be used for scaling. However, depending on the flow of the game, there may be similar amounts of player involvement in a match featuring 130 sequences as there is in a match with 170 sequences. In both instances the amount of time allotted is 80 minutes, indicating that participants had equivalent time periods to be involved. The 130-sequence match may feature longer sequences than the 170-sequence match, such as longer strings of uninterrupted back play. Or the 170-sequence match may have a higher penalty count or have a higher error rate requiring more set piece play.

Having briefly described the data and examined some of the problems and limitations associated with the data set, the methods used to quantify performance are described next.

3.3 Methods for Data Analysis

The computer software packages SAS, Minitab and Excel were used for the work described in this section. Standard methodology for both cluster and multivariate analysis was applied to the data described briefly in the previous section and outlined more thoroughly in Appendix A. The use of the primary multivariate methods in this section, namely cluster analysis and factor analysis, are well documented as demonstrated by the large number of textbooks that describe in detail the use of these techniques (Sharma, 1996; Hair, Anderson, Tatham & Black, 1995; Manly, 1994; Krzanowski & Marriott, 1994; Morrison, 1990; Krzanowski, 1988; Cliff, 1987; Chatfield & Collins, 1980). The emphasis detailed in the following sections relates to the situational considerations for using cluster analysis and factor analysis with the rugby data, rather than focusing on the mathematics of the techniques.

3.3.1 Cluster Analysis

Science and art converge as the results obtained from analyses and expert opinions need to be carefully balanced to provide positional clusters that are contextual from a rugby perspective and also reduce the within-cluster variability. In a coaching sense, certain positions of players can be grouped together based upon the similarity of roles, required skill set, common tasks and so forth.

A danger of performing a cluster analysis on player performances is that potentially there are no clear-cut groupings using positions. This is because that even within positions roles may differ depending on the skill set of the individual. That is one of the strengths of team sport; strengths and deficiencies of individuals are combined to form a team. A prime example is the position of hooker. Roughly, players in this position can be considered as two types, tight or loose. Tight hookers take on the role of the basic front row forward, adopting a more physical approach. On the contrary, loose hookers tend to be more in the mould of the back row, ranging wide and becoming involved in more open and flowing match play. Classification of player types by identifying styles of positional play within player positions is an obvious area for further research.

Hair *et al.* (1995) suggest two steps are required to obtain clusters. Initially, clusters are identified using a hierarchical procedure and then these clusters are refined using an agglomerative k-means procedure. To identify clusters, a linkage method and distance measure must be chosen.

Ward's linkage is the hierarchical cluster procedure adopted as it combines clusters of the greatest similarity at each stage. This is opposed to other linkage methods which tend to link items (or groups of items) based on how similar the individual items (groups) are. Fundamentally, Ward's method forms clusters by maximising within-cluster homogeneity as opposed to computing distances between clusters (Sharma, 1996). For this application, comparison is suited to within-positional clusters more than between positional clusters. As positional clusters will possibly exhibit an overlap between tasks, single linkage (as well as complete linkage and average linkage) may not produce a clear distinction between intuitively different positions, such as loose forwards and midfield backs. Therefore it is easier to look at performance on similar tasks rather than performance on different tasks. Consequently, identifying similar traits (maximising the within-cluster homogeneity) should provide clusters of similar positions. Although Ward's linkage tends to combine clusters so that they are approximately equal in size, by placing similar featured clusters together this approach is more suited to the rugby data.

For quantifying the distances between clusters and/or observations Euclidean distances are used rather than Manhattan distances. This is implemented to recognise the importance of outliers in defining sport performance, as outlined in Section 3.2. These outliers may provide "beacons", allowing certain positions to be readily identified, such as kicking metres gained by the inside backs. This supports the use of Euclidean distances which tend to accentuate outlier results, whilst Manhattan distances based on absolute distances do not. Provided the data used is representative of typical first class play, the outlying performances are a necessary component and indication of performance, as discussed previously; this aspect is crucial and is discussed later, where atypical outliers are identified and removed. Further, Hair *et al.* (1995) recommend a squared Euclidean distance as the distance measure for Ward's linkage.

The data is standardised so that all variables have an equivalent scale. This standardisation occurs across all observations within the group to be assessed in the given analysis, whether it is cluster analysis or factor analysis. For cluster analysis standardisation occurs with respect to all observations, whereas in the factor analysis standardisation occurs with respect to the relevant positional cluster. Standardisation is not performed on an individual match level due to the variability of match circumstances.

Natural groupings of positions within the data need to be identified in order that separate factor analyses can be performed on each positional cluster so that the specific traits of each positional cluster can be identified.

3.3.2 Factor Analysis

A similar situation is encountered with factor analysis, where a rugby perspective provides the intuitive framework and relative frame of reference for accepting or rejecting potential analyses. This is the most important aspect of creating the performance measure. As the rating system is designed to be used in a public arena, the context and interpretability of the performance measure are the overriding considerations. This is required to create a product that is both statistically viable and marketable to the consumer. From a dimension reduction perspective, any natural groupings of variables to be found in the data should fit with traditional rugby based thought as shown in Figure 2.1. That is attack based variables should compose one factor, defensive attributes should comprise another and so forth. The relative importance of these attributes will differ between positional clusters, emphasising the need to segregate the groupings prior to the factor analyses.

It is important to note the reasoning behind the use of factor analysis rather than principal component analysis, as these are the two basic methods available for obtaining dimension reduction (factor solutions). According to Hair *et al.* (1995), the choice between the two methods is based on two criteria: (1) the research objective and (2) prior knowledge regarding the variance in the variables. The remainder of this section refers to Hair *et al.* (1995).

Hair *et al.* (1995) state that the overall variance consists of three types (common, specific and error). Common variance is "variance which is shared with all other variables". "Specific (or unique) variance is unique to that variable and not explained or associated with any other variables". Thirdly the variance associated with error is the variance of a variable due to errors in data collection or measurement (Hair *et al.*, 1995, pp. 365-366).

In this thesis little prior knowledge regarding the contribution of the unique variance to the overall variance is known but error variance is expected to be large due to changeable match constraints. This provides a useful indication as to which method is most appropriate, as explained below. Essentially, principal component analysis is appropriate when the researcher is primarily concerned about prediction, or identifying the minimum number of factors while maximising the amount of variation explained. Factor analysis is the most suitable method when the latent dimensions or constructs represented by the original variables are of interest and limited knowledge regarding the unique and error variance is available (Hair *et al.*, 1995). That is factor analysis focuses on the shared variance rather than the unique variance. In this thesis, this places the emphasis on identifying the variance associated with the skills that govern physical tasks. More importantly, Krzanowski (1988) implies that when latent variables are being extracted, such as match performance and key performance indicators, factor analysis is most suitable. Due to the changeable structure of rugby, the interest in this thesis is with the specific traces of underlying performance, which are determined by the underlying latent dimensions of performance and are identifiable by the shared variance. Consequently, to quantify performance using dimension reduction techniques the latent dimensions of performance need to be extracted, anticipating that these constructs will relate to key rugby ideas of attack, defence and so forth.

Rotation of the extracted factors is a useful tool for interpretation (Hair *et al.*, 1995). Two major types of rotation exist, orthogonal and oblique. With orthogonal rotation, factors are rotated whilst maintaining a 90 degree angle between the reference axes. That is each factor remains uncorrelated. The aim of rotation is to simplify the rows and columns of the factor matrix, thereby simplifying the interpretation process. Varimax is one of the most popular methods for orthogonal factor rotation (Hair *et al.*, 1995).

With this orthogonal rotation technique, the emphasis is on column simplification by rotating axes such that some loadings tend to be –1, 1 or 0, because the nature of the relationship is easiest to interpret given these types of factor loadings. Varimax rotation provides a clearer separation of the factors and has proven very successful as an analytical approach to obtaining rotation of factors.

Of the other two major orthogonal rotation techniques, quartimax seeks to simplify rows and consequently tends to create a large general factor. Given that there are definite distinct aspects of rugby play, such as attack and defence, this approach is not suitable. The third major rotation technique is the equimax, which attempts to create a compromise between the quartimax and varimax approaches. This method is used infrequently and has not gained widespread acceptance (Hair *et al.*, 1995, p. 384).

Hair *et al.* (1995) state that there exist no specific rules for adopting a particular rotational technique. However, some broad guidelines are identified, stating that if the goal is to reduce the number of original variables regardless of the need for meaningful factors, then orthogonal rotations are favourable. However, if the task is to identify several theoretically meaningful factors then an oblique solution is more appropriate. This reasoning is based on the premise that, realistically, few variables are uncorrelated, as suggested in an orthogonal rotation. An oblique rotation is most favourable in this thesis because the interest is to extract meaningful factors. However, given the variability in match conditions, and exposure of only partial individual skill sets due to limited involvement, major categories such as attack and defence can be thought of as being independent. Essentially, the combination of many different skill-sets from many different individuals in similar positions exposed to many different match conditions produces factors that are effectively independent. Again, the impact of skill sets upon the factors can be shaped due to the influences of the three types of variation described previously. The interest is sorting through all the associated variability associated with match play and associating the key characteristics of performance, which can be assumed to be independent. Therefore an orthogonal rotation with a varimax rotation is deemed appropriate for the factor analysis in this thesis.

3.4 Cluster Analysis

A series of steps need to be completed to quantify individual rugby player performance. The first task is to identify similar groupings of on-field positions. This can be achieved utilising the statistical technique, cluster analysis. As explained earlier, identification of positional clusters is necessitated by the fact that within a rugby team the job requirements (see Appendix C) for each position differs. This section shows that due to the complex nature of rugby, the results presented from a cluster analysis are not sufficiently refined for our purposes. This is not an uncommon finding. "Cluster analysis is not a rigorous and sharp statistical tool, and should be applied with care and with the assistance of any other information about the sampling units (Morrison, 1990, p385)." Therefore in an attempt to provide this extra information, this section starts by providing a contextual debate for positional groupings based on the opinions of rugby observers. A more objective grouping is then obtained from a cluster analysis of the data set and finally the coaching perspective and cluster analysis are used together to establish the optimum number of positional clusters.

In a game such as rugby it is possible to identify examples that confirm a direction of thought in terms of player groupings as will be detailed later in Section 3.4.2. However, building upon the foundation of statistics it is important to assess recurrent themes and trends that typify rugby, such as principles of attack and defence. This will produce an approach that is both objective and realistic. Initially, the slant is towards statistical evidence, with the opinions of rugby observers providing justification for the course of action. However, an appropriate balance between the rugby observer and statistical evidence needs to be adopted to ensure that the most appropriate positional groupings are found.

Firstly a discourse based on rugby literature identifies potential groupings from a coaching perspective. Secondly a cluster analysis is performed and compared with the groupings discussed in the first section. Thirdly the evidence is combined to produce positional clusters to be used as the foundation for the factor analysis and to develop appropriate performance measures.

3.4.1 Clustering Philosophy

Before commencing a dimension reduction of the multivariate data, the data cleaning process requires that any groups within the data which have fundamentally different skill sets be identified. Two benefits result from this. Firstly, modelling errors are reduced as factor models fit more adequately the more tightly defined the group. Secondly, sensible answers that can be related to the contextual situation are obtained. In a contextual sense, it is inappropriate to consider a model for a prop based on kicking ability, as this is not a relevant skill in playing that position successfully. To be identified as homogenous groups, items within each cluster of player performances must exhibit the same multivariate skill/task profile.

"The demands of each game and of each position are quite different (Hazeldine & McNab, 1991, p13)." Although there are specific roles on a rugby field, covered by broad positional definitions, the expectation upon each position differs with particular reference to a number of key areas:

- state of the match
- state of the competition
- game plan
- weather
- time
- individual strengths/weaknesses
- team strengths/weaknesses
- opposition strengths/weaknesses
- performance of individuals in neighbouring positions

This list is not exhaustive, but it highlights the variability of conditions and constraints that impact on how rugby is played. The above variability causes a high level of measurement error in the performance measures, reinforcing the comments from the previous section relating to the unsuitability of principal components as a means for extracting factors.

The broad positional definitions may differ depending on the above key areas. This has potential implications for both the cluster and factor analyses affecting the number of positional clusters deemed appropriate. The clusters adopted must provide a good balance between generalisation and specialisation. These terms are defined below.

Generalisation

Generalisation allows the overlap of positions. That is during the course of the game generalisation allows the possibility of players moving from one position to another. Generalisation recognises that certain positions have equivalent expectations regarding a potential role and also acknowledges that there is significant skill set overlap between positions; that is players are able to play in more than one position. For example, All Black Taine Randall has started test matches in all loose forward positions (number eight, blindside flanker and openside flanker). Alama Ieremia has featured in All Black line-ups at both second five eighth and centre.

Specialisation

But, many positions have special key tasks that need to be performed due to the job description. This highlights the potential skill set of an individual and determines where an individual is capable of playing. Andrew Mehrtens, Steven Bachop, Grant Fox, Simon Culhane are All Black first five eighths of the 1990's who would not be considered as starting prospects in any other position due to certain components of their skill set. These individual are not renowned for using physically aggressive defence to intimidate opposition. Pairing their defence qualities with the ability to direct play affirms this statement and provides an obvious example of these components. Similarly the halfback role can be pinpointed as being specialised, with quick precision passing off the ground and in close-quarter situations a prerequisite of the position. This emphasises the specialised core skill set required to play in such positions.

Continuing this specialist line of reasoning is potentially counter productive as the specialist differences between other positions could be highlighted. For example: tight-head and loose-head props, locks who jump at two in the lineout as opposed to four and so forth. Essentially the thrust of the argument presented here is that within each potential positional cluster, props for example, there is considerable skill-set overlap that should be generalised. The important point is that some positions have a unique skill-set, half back for instance, which discourages the formation of positional clusters that include these positions.

A selection of broad arbitrary groups could be specified but in order for dimension reduction techniques to be successful (produce useful performance measures) the preceding criteria involving specialisation and generalisation must be noted. Potentially there will be a minimum of one cluster and a maximum of fifteen clusters representing the fifteen positions. The problem here is to choose an appropriate number of clusters.

Substitution rules suggest that fifteen clusters are inappropriate. There are only seven named reserves to cover fifteen positions, emphasising the necessary skill-set overlap within a team. Reserves are absorbed into the position vacated when they join the game. A convenient by-product of the cluster analysis based on seven clusters would be that a reserve could be named to cover each cluster. However, in a named playing squad of 22, five individuals must be capable of playing in the front row, placing immediate restrictions on potential clusters.

In choosing the number of clusters a balance between generalisation and specialisation needs to be applied. If too many clusters are chosen potential problems arise. Group membership can be small resulting in a relative lack of data. This means that specialisation is emphasised at the expense of generalisation which narrows the skill-set scope. The effect of a reduction in the perceived skill-set scope can be profound if the nature of a game places demands on the individual outside of the expected skill-set. Conversely, when only a small number of clusters are identified, important positional requirements are lost. That is generalisation is emphasised at the expense of specialisation. This causes positional biases to occur in the resultant factor models, whereby one position may be unable to match the performance of another position in a cluster group.

For example, it is inappropriate to group all loose forwards together when evidence suggests (see Figure 3.3) that openside flankers have a significantly higher tackle count. In a situation where all loose forwards were grouped together, openside flankers would tend to dominate lists with reference to the number of tackles made, when by the very nature of the game openside flankers are expected to make more tackles than either the number eight or blindside flanker. To confound matters, South African teams tend to play the openside flanker in the number 6 jersey, rather than number 7, which is the general convention.

The positional clusters need to be large enough to provide sufficient generalisation and yet small enough to cater for specialisation. Gathering evidence, statistical, experimental and anecdotal from a coaching perspective, provides an ideal compromise for establishing the clusters necessary to proceed with the dimension reduction of performance data and eventual quantification of individual performance. The coaching perspective considers attributes that are not necessarily immediately evident from a statistical database. However, in some shape or form, different properties with respect to a player can be inferred provided that enough sampling situations are presented with enough players to give a valid comparison. This essentially parallels the central limit theorem, whereby given a large enough sample an impression of the individual's true ability (mean) is obtained.

Coaching based clusters focus on similar positional definitions based on physical, mental and skill expectations. These groupings are inherent within the structure of the rugby data. A team is a group of differently skilled players are combined to increase the likelihood of winning. This is essentially the basis of rugby. Players specialise in a certain position, due to physiological attributes, technical skill, and mental attributes. Evidence of the importance placed on each position, or unit, is outlined in Appendix C, reproduced with the permission of the NZRFU.

Team sport provides a situation where the skills of co-operating individuals are combined in a favourable manner to form a team. That is strengths are heightened and weaknesses lessened. "Most teams aim at having the game played in a way that they feel gives them an advantage over the opposition by playing to your own strength and against their weaknesses (Whineray, 1978, p13)." This is a direct consequence of a team's desire to play their best, an objective which is not questioned at first class level.

3.4.2 Coaching Perspective

By definition the coaching perspective on positional groupings is subjective in comparison with a cluster analysis grouping. But each position requires a core skill-set and has specific tasks and the suitability of an individual to a position is determined by the relevant skill-set coupled with the physical body type required to undergo the rigors of a certain position. Thus most coaches will put together similarly specified positions.

From this perspective, cluster analysis is performed to confirm what is already suspected. Furthermore, if the existing clustering relationship is fuzzy, then the opinions of the coaches can be used refine the clusters.

This section identifies clusters of positions based on information from coaching specifications and is assisted by a cluster analysis. The glossary in Appendix E contains an overview of brief positional definitions as specified by the NZRFU (2000). Jersey numbers (1, 2,..., 15) represent positions as outlined in Appendix E. From a coaching perspective, the following authors divide the starting 15 positions into various positional clusters. The information relating to the specification of these groupings consider a wide range of factors; physiology, skill level and role within the team as perceived by the coach, trainer and observer. Authors that specify only Backs and Forwards have not been included. This list is not exhaustive, but serves to highlight key themes that are congruent with rugby thinking. Listed with the author is the background in which the divisions are made (coaching, fitness, journalist or kinetics).

Stewart (1987) (Coaching)

Front Row (1, 2, 3), Tight Five (1, 2, 3, 4, 5) implying the second row is a separate unit, Flankers & Number Eight (6, 7, 8), Inside Backs (9, 10), Midfield Backs (12, 13) and Outside Backs (11, 14, 15).

Rigg & Reilley (1987) (Fitness)

Front Row (1, 2, 3), Second Row (4, 5), Back Row (6, 7, 8), Halfbacks (9, 10) and Backs (11, 12, 13, 14, 15).

Treadwell (1987) (Kinetics)

Front Five (1, 2, 3, 4, 5), Backrow (6, 7, 8), Halfbacks (9, 10), Centres (12, 13) and Outside Backs (11, 14, 15).

Bell, Cobner, Cooper & Phillips (1991) (Fitness)

Props and Locks (1, 3, 4, 5), Hookers and Back Row (2, 6, 7, 8), and Backs (9, 10, 11, 12, 13, 14, 15).

Hazeldine & McNab (1991) (Fitness)

Props and Locks (1, 3, 4, 5), Hookers and Back Row (2, 6, 7, 8), Halfbacks (9) and Backs (10, 11, 12, 13, 14, 15).

NZRFU (1998) (Coaching)

Props (1, 3), Hooker (2), Locks (4, 5), Loose Forwards (6, 7, 8), Inside Backs (9, 10), Midfield (12, 13) and Outside Backs (11, 14, 15).

Hutchinson (2000) (Coaching)

Tight Four (1, 3, 4, 5), Flankers (6, 7), Hooker and Number Eight (2, 8), Interior Backs (9, 10, 12, 13) and Outside Backs (11, 14, 15).

Cameron (1980) and Akers and Miller (2000) (Journalist)

Props (1, 3), Hookers (2), Locks (4, 5), Flankers (6, 7), Number Eights (8), Halfbacks (9), Five-Eighths (10, 12), Three-quarters (11, 13, 14) and Fullbacks (15).

McLean (1979) (Journalist)

Props (1, 3), Hookers (2), Locks (4, 5), Flankers (6, 7), Number Eights (8), Halfbacks (9), Five-Eighths (10, 12), Centre-Three-quarters (13), Wing-Three-quarters (11, 14) and Fullbacks (15).

Palenski, Chester and McMillan (2000) (Journalist)

Props (1, 3), Hookers (2), Locks (4, 5), Flankers (6, 7), Number Eights (8), Halfbacks (9), First Five-Eighths (10), Second Five-Eighths (12), Centre (13), Wings (11, 14) and Fullbacks (15).

The source of the clusters highlights the reasoning underlying the divisions. McLean, Cameron, Akers, Miller, Palenski, Chester and McMillan are sports journalists who need to describe the composition of a touring squad, or comment on individual records, with as much detail as possible. Hazeldine *et al.* focus on a physiologically basis which places the emphasis on the fitness attributes required. Consequently, the number of clusters is minimised to reduce the workload of administering generic training programmes. The coaching based clusters, which centre on the required skill, provide the most applicable results in terms of performance measures due to the focus on skill.

However, the effect of fitness expectations and physiological requirements can not be ignored completely, as these too comprise the skill-set of a player. Further evidence of this is demonstrated in Appendix C with reference to positional responsibilities.

Prior to applying a dimension reduction technique, any natural groupings in the rugby data must be identified. But to be of any use in a commercial sense, these groupings will be associated with the defined positions of rugby union.

The job description of an individual will alter depending on patterns of play, or strategies. Different mini-units work together in different scenarios. Consider set piece play, specifically scrummaging. As an example, the following mini-units can be defined (NZRFU, 2000):

- Front Five
- Front Row and Halfback
- Loose Forwards
- Loose Forwards and Halfback

Each mini-unit is responsible for a specific section of play. That is, the Front Five's role is to ensure that a solid base is set at scrum time, with the purpose of disrupting opposition ball or ensuring clean ball for the halfback or number eight of their own team. Likewise, the Front Row and Halfback need to work closely to ensure the ball is hooked effectively at scrum time. Mini-units can also be non-position specific, such as amongst the forward pack during phase play (NZRFU, 2000), listed below:

- Ball Carrier and Loose Forwards
- Support Players

From this discourse it becomes clear that the skill-set required within the mini-units to complete the tasks mentioned above supports positionally specified clusters. Consider the mini-unit comprising the Front Row and the Halfback at scrum time. Simply, the Halfback feeds the ball into the scrum; the props provide a stable base from which the hooker can strike at the ball. Within this mini-unit each of the positions has a set job as required by their job description necessitating a specific group of skills from their skill-set.

In proposing clusters, emphasis is given to coaching based opinions, as skill-sets provide a large component of the basis for division. Table 3.3 displays a numeric measure using the author groupings indicated previously. The first two letters of the authors' names head the column that houses their opinion. These opinions are converted to a numerical scale using a 'distance from action' measure (DfA). This is based loosely on the jersey numbers of individuals. For each author the average 'distance from action' measure is used for each grouping. The distance from action uses the jersey numbers of the players due to the general positional layout (see Appendix E) of players during the course of a match. For example, Bell *et al.* group the hooker with the loose forwards. The DfA for the positions in this grouping is calculated as follows: (2+6+7+8)/4, which is 5.75. Each position within this grouping is given this value such that for each column, where positions have the same DfA, this means that those positions belong to the same group. Importantly, tight-head and loose-head props are both given a distance from action of one due to their involvement in the front row. This allows the presence of the hooker in the front row cluster to be distinguished. The left and right wingers are also given identical distance measures of 13.5. This reflects the similarity of these two positions.

Position	*DfA*	*Be*	*Ha*	*Ri*	*Tr*	*Hu*	*St*	*NZ*	*Ca*	*Mc*	*Pa*
1	1	2.75	2.75	1.33	2.6	2.75	1.33	1	1	1	1
2	2	5.75	5.75	1.33	2.6	5	1.33	2	2	2	2
3	1	2.75	2.75	1.33	2.6	2.75	1.33	1	1	1	1
4	4	2.75	2.75	4.5	2.6	2.75	4.5	4.5	4.5	4.5	4.5
5	5	2.75	2.75	4.5	2.6	2.75	4.5	4.5	4.5	4.5	4.5
6	6	5.75	5.75	7	7	6.5	7	7	6.5	6.5	6.5
7	7	5.75	5.75	7	7	6.5	7	7	6.5	6.5	6.5
8	8	5.75	5.75	7	7	5	7	7	8	8	8
9	9	12	9	9.5	9.5	10.5	9.5	9.5	9	9	9
10	10	12	12.5	9.5	9.5	10.5	9.5	9.5	10.5	10.5	10
11	13.5	12	12.5	13	14	14	14	14	13	13.5	13.5
12	11	12	12.5	13	12.5	10.5	11.5	11.5	10.5	10.5	11
13	12	12	12.5	13	12.5	10.5	11.5	11.5	13	12	12
14	13.5	12	12.5	13	14	14	14	14	13	13.5	13.5
15	15	12	12.5	13	14	14	14	14	15	15	15
# Groups		3	4	5	5	5	6	7	9	10	11

Table 3.3: *Perceived Positional Clusters and Distance from Action*

A number of groupings are evident from the previous discussions and are linked persistently in the above table. The following pairs of positions are often paired.

1. Props (1, 3)
2. Locks (4, 5)
3. Flankers (6, 7)

From a coaching perspective, the following units are identified.

4. Outside Backs (11,14,15)
5. Midfield Backs (12,13)

This leaves four positions that are treated very differently:

6. Hooker (2)
7. Number Eight (8)
8. Halfback (9)
9. First Five Eighth (10)

This suggests nine cluster groupings. The opinions specified above will be revisited later in this section. Of the coaching opinions, only Hutchinson (2000) separates the Number Eight from the Flankers. Of all the positional clusters, the back row gives cause for the most contention. Anecdotal evidence suggests that the role of openside flanker is very much different to that of his fellow loose forwards. It is also suggested that the openside plays an entirely different style of game. The openside flanker performs a similar role to that of the wing forward from the early part of the 20^{th} century when a two man front row was employed as part of a 2-3-2 scrum formation. This is opposed to the modern convention of 3-4-1 required by modern rugby law.

A comparison of boxplots for each position (denoted by jersey number) shows that the tackle counts tend to be higher for openside flankers (7) than for any other position. The tackle counts were transformed by a square root transformation (after adding one to reduces the impact of zero counts) to introduce symmetry and stabilise the variance within each position. Only observations where the individual participated in more than 60 minutes of the match for the 34 test matches played in 1999 featuring Australia, New Zealand, South Africa and England were used. This provided 617 observations for Figure 3.1.

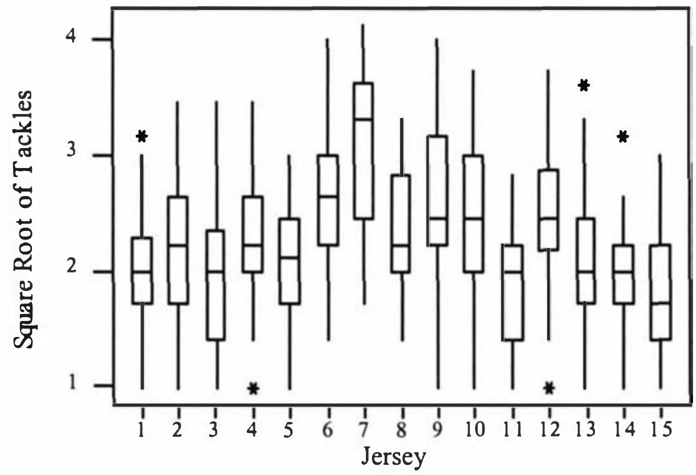


Figure 3.1: *Boxplot Showing Openside Flanker Dominance of Tackle Counts*

The boxplot in Figure 3.1 clearly demonstrates the dominance of an openside flanker in tackle counts. This then poses the question of how to group the loose forwards. It is inappropriate to group all three positions together as the tackle count for these positions differs. There is obviously some overlap between these three positions, with authors grouping all three loose forwards together or just the two flankers together. A potential grouping puts the blindside flanker and number eight together and having the openside flanker as a separate grouping. Physiological differences hold crucial evidence to support this choice. Stewart (1987) discussed the loose forwards and their relevant lineout positioning. In general the blindside flanker and number eight jump in the 5 and 6 spots in the lineout (not necessarily in that order) with the openside flanker jumping at the back in the seventh spot. Customarily the openside flanker is an extremely quick reactor and mobile player. In order to offset any disadvantage at lineout time the other two of the loose trio will consequently need to be heavier and taller players (Stewart, 1987). A cluster analysis is useful at this time to see if the data can present any natural resolution to this debate.

3.4.3 Analysis

A wealth of literature relating to cluster analysis, both underlying theory, applications and methodology is readily available (for example: Sharma, 1996; Hair, Anderson, Tatham & Black, 1995; Manly, 1994; Krzanowski & Marriott, 1994; Everitt & Dunn, 1991; Morrison, 1990; Krzanowski, 1988; Digby & Kempton, 1987; Romesburg, 1984; Gordon, 1981; Chatfield & Collins, 1980; Hartigan, 1975.).

Cluster analysis is an established technique for grouping individuals or items so that items in a cluster are more similar with other items in that cluster than items in other clusters (Hair *et al.*, 1995). A mathematical foundation ensures that the clusters specified are of objective. Applicable in situations where a 'natural' structure is sought among the observations based on a multivariate profile, cluster analysis is the most commonly used technique for identifying homogenous groupings (Hair *et al.*, 1995).

The desired result consists of clusters of items that exhibit high internal homogeneity and high between cluster heterogeneity (Hair *et al.*, 1995). However, players of a completely different nature can fill certain positions. That is within a position the manner in which an individual participates may differ completely from others. This causes the difference in the coaching-based groupings discussed previously and poses a problem in terms of validating clusters. However, this does not suggest a problem in the clustering procedure itself because player performances for individual matches are being clustered, not positions. But it may create problems in terms of creating a saleable product, because the market demands a product to which it can relate, such as positional based clusters.

It is possible for individuals to "personalise" the position, making use of certain attributes within their skill-set that may be additional or far superior to the skill-set demanded in a particular position. The previous section highlighted a variety of potential positional clusters based on different experiences of the sport. For example a hooker can be of either a tight five or loose forward mould, or capable of playing both styles dependent on match situation and game strategy. Not only may an individual differ within the position, the individual may differ within themselves, given match constraints. However this all contributes to the beauty of rugby and the adaptability required to perform at the highest level.

3.4.4 Application

Ward's method is used for the linkage of hierarchical clusters. Primarily there are two broad groups readily accepted, back and forwards. As rugby is a team game, all individuals are focused on the same goal, victory, or at the very least maximising performance. In order to achieve this goal, each individual must perform a prespecified role. It is the overlap of roles and the similarity between positions, not the difference between positions which is important. This makes Ward's method an appropriate method because similar traits are sought as discussed in Section 3.3.1.

The recommended distance measure for Ward's method is the squared Euclidean distance (Hair *et al.*, 1995). After performing the initial cluster identification on the standardised data, the clusters are refined using the k-means clustering procedure, again using standardised data. Standardisation is important, as the scaling of the variables is different and standardisation ensures that all variables contribute equally to the analysis. This is a crucial step as the likelihood of occurrences differs greatly among variables, positions and games. The initial clusters function as a starting point for the k-means procedure. As this method is non-hierarchical observations can be transferred between clusters once placed, if necessary. That is observations can be moved if its multivariate profile is more congruent with another cluster grouping than the previously specified cluster.

The data are not square-root transformed prior to the cluster analysis so that the inter-variable relationships are not distorted. Transformation would typically reduce the effect of extreme values in the case of very skew distributions. This does not pose a problem, as the standard statistical assumptions of normality and homoscedasticity are not relevant in either cluster analysis or factor analysis, only with post-analysis statistical testing of the output from the analyses.

The initial target market for this thesis is Super 12 franchises and first division provincial unions, with the focus placed on the Super 12 and NPC competitions. As the contested fixtures of these competitions are relatively even, it is desirable that the data set used is reflective of the type and style of rugby played. Analysing data from these specific competitions would be more desirable, but this was not available due to the cost and lack of time to create suitable data to meet the deadlines Ultimate Rugby deadlines. To replicate the high standards and nature of the target competitions, several games needed to be removed. The aim of the data cleaning was to provide a data set covering top teams in relatively close encounters, as this is what is expected in the target competitions. The 34 test matches in the preliminary database included rugby minnows, Uruguay, Fiji, Spain, Tonga, USA and Romania alongside the giants of international rugby such as, New Zealand, South Africa, Australia and England. The disparity between the nations is obvious when the 1999 World Cup results are considered. For example England demolished Tonga 101:10 and New Zealand destroyed Italy 101:3. These heavy defeats highlight the difference between the top nations and lower ranked countries. Thus in order to get an appropriate profile of how the game is played, individuals were ignored from teams participating in a match where the difference in final score exceeded 30 points. This reasoning is based on the balance of the rugby played. A team beaten by a large margin will have spent a disproportionate time defending; likewise the winning side will have spent a disproportionate time on attack. Realistically, one-sided matches do not produce performance results that are reliable or representative of the first class rugby from the target market. Further, the outliers from these matches would distort the outliers from tighter matches, which are necessary to highlight areas of exploitation. Further, only individuals who played more than 60 minutes where included in the data set for clustering. Individuals are expected to have made a recognisable contribution after this time. By taking a large portion of time, the ebb and flow of a game is largely covered. A total of 361 observations were obtained for analysis in this manner. The 31 variables described in Table 3.1 are used (infringements are included as one variable).

A dendogram from the initial cluster analysis using Ward's Linkage and squared Euclidean distance suggested two or three groups, as shown in Figure 3.1 below. The y-axis measures the distance (squared Euclidean) between clusters (Minitab, 1996).

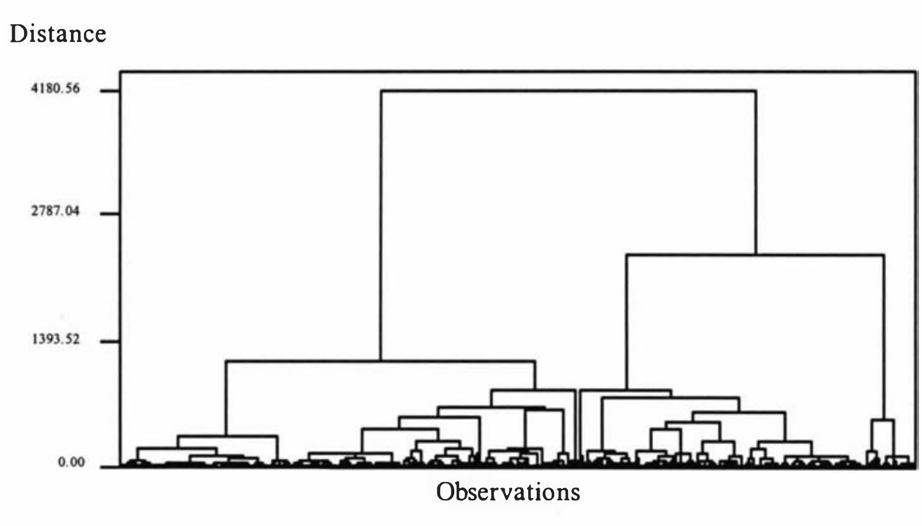


Figure 3.2: *Dendogram indicating Potential Positional Clusters*

The dendogram above is a cross tabulation between position and the three clusters indicates a contextual basis to Table 3.4. To avoid confusion clusters are denoted by letters (which are assigned arbitrarily), whilst positions retain the referenced jersey numbers.

	CLUSTER		
POSITION	A	B	C
1	95	5	--
2	100	--	--
3	100	--	--
4	100	--	--
5	100	--	--
6	92	8	--
7	100	--	--
8	100	--	--
9	--	100	--
10	--	22	78
11	--	96	4
12	16	80	4
13	--	100	--
14	4	96	--
15	--	77	23

Table 3.4: *Cross Tabulation of Row Percentages indicating Relationship Between Position and Cluster*

The cross tabulation (Table 3.4) indicating the relationship between the clusters obtained from the cluster analysis and position is conclusive. The row percentage indicates the relative dispersal of the positions and highlights the natural groupings. Three clusters are clearly defined (obtained by cutting the dendogram at a squared Euclidean distance of approximately 1400), which can be thought of as forwards, backs and first five eighths.

These clusters are given the arbitrary labels of A, B and C in the table above. If two clusters are examined (by dissecting the dendogram at approximately 2800) as opposed to the three specified in the above table, clusters B and C merge, creating two very distinct, natural and contextual clusters (forwards and backs). There is more justification for selecting just two clusters, based on the distances between clusters, as shown by the dendogram in Figure 3.2. However, the third smaller cluster, which is more closely aligned with the backs than the forwards, is no less an important cluster. The percentage of positions distributed in each cluster emphasises the accuracy of classification. The process was repeated using single linkage with unconvincing results.

An important anomaly is that almost one quarter of Fullbacks (23.08%) are included in the First Five Eighth Cluster (Cluster C). Intuitively, this is due to the Fullbacks kicking game, contrasting the returning of received position via tactical kicking as opposed to counter attacking by running the ball back. This is clearly seen in the boxplot below which compares the square root of the total metres kicked for the three clusters. This difference is pronounced and is the key to the separation of the two groupings amongst the backs.

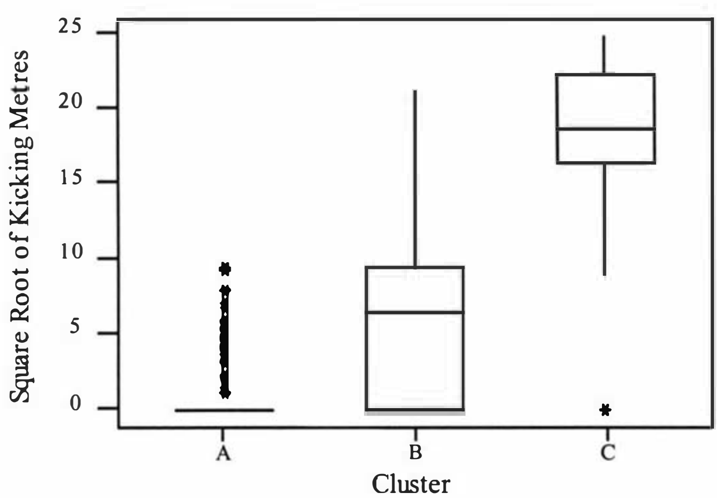


Figure 3.3: *Boxplot Showing Difference between Clusters for the Square Root of Metres Kicked per Game*

The difference between clusters is pronounced, with individuals in cluster C averaging 323.8 kicking metres per game, compared to just over thirty metres per game for the other backs and almost zero for the forwards.

The result specifying two or three clusters is too generalised. As no other divisions are particularly clear, the clustering procedure is repeated specifying 15 clusters, one for each position on the field. The clusters are given arbitrary labels (A-O). The resultant cross tabulation, shown in Table 3.5, between cluster and position confirms the blurred division between positions. The numerical values are row percentages.

Several clusters are indicated along a positional structure in Table 3.5. Halfbacks are strongly represented in cluster F with 84%. The midfield pairing of second five eighth and centre are accounted for in Cluster H. The first five eighth position is split between two clusters, G and O. Props are found lumped together in Cluster L.

Group	*Cluster*														
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*
Position *1*	11	--	16	--	--	--	--	5	--	--	--	63	--	5	--
2	29	8	38	13	--	--	--	--	--	--	--	12	--	--	--
3	--	9	30	--	--	--	--	--	--	--	--	44	--	17	--
4	4	--	8	33	38	--	--	--	--	--	--	4	--	13	--
5	8	4	25	8	21	--	--	--	--	--	--	17	--	17	--
6	33	--	4	21	25	--	--	--	--	--	--	13	--	4	--
7	15	15	7	30	15	--	--	--	7	--	--	--	--	11	--
8	35	22	9	30	--	--	--	--	--	--	--	4	--	--	--
9	--	--	--	--	--	84	--	--	4	8	4	--	--	--	--
10	--	--	--	--	--	--	43	--	--	4	17	--	--	--	35
11	--	--	--	--	--	--	--	16	12	24	16	--	28	--	4
12	--	4	--	8	--	--	--	40	8	4	20	--	8	4	4
13	--	---	--	--	--	--	--	44	16	12	8	--	20	--	--
14	--	--	--	--	--	--	--	17	21	8	21	--	29	4	--
15	--	--	--	--	--	--	19	--	4	4	61	--	8	--	4

Table 3.5: *Row Percentages Indicating Relationships between Clusters and Position*

Whilst the above table produces very few definitive clusters, some basic groupings are evident. Cluster A includes the major groupings of blindside flankers and number eights. The hookers also have strong representation (29%) but feature more strongly in Cluster C, mixed in with the tight forwards, namely tight head prop, and right lock (5). This validates the argument that hookers are either of the tight mould or loose mould. A notable exclusion from this grouping is the openside flanker, with only 15% found in this cluster. This evidence suggests that of the loose forwards, the blindside flanker and number eight should be grouped together, with the openside flanker treated separately.

The other clusters can be interpreted loosely as follows. The wings are spread roughly between three clusters. Bringing these smaller clusters together (I, J, K) also brings the fullback into this grouping (predominately featured in K). It is not surprising the outside backs are spread across a number of different clusters. The involvement of these individuals is dependent on the flow and structure of the game. For example, it is possible for wings to be starved of the ball, not through any fault of their own, but perhaps due to the opposition loose forwards preventing the inside backs from distributing the ball. The outside back positions are split across several clusters with various involvement levels relating to the relative performance of the team in each match.

An obvious feature of the cross tabulation is the considerable overlap of positional clusters. Whilst the clusters are somewhat artificially defined, having forced 15 as opposed to the two or three clusters that were evident in the dendogram (Figure 3.2), the evidence provided by this objective analysis will be compared to the analysis of the subjective evidence described previously in order to find a more definite number of clusters, as well as definitions for these positional clusters.

3.4.5 Discussion

A return to the subjective groupings based on rugby orientated opinion on page 67 provides the opportunity to run another cluster analysis. The average 'distance from action' measure for each author given in Table 3.3 is used to cluster the 15 positions. Ward's clustering procedure provided a number of interesting results which helped specify cluster numbers. From the batch of coaches reviewed, the NZRFU coaching resources are the superior source, as they are used to coach the coaches. The NZRFU specified seven clusters, providing an ideal starting point. Below are the cluster groupings that are produced when a particular number of clusters are specified (7, 8, or 9), where positions are represented by jersey numbers.

Seven Clusters: (1, 3) (2) (4, 5) (6, 7, 8) (9, 10) (12, 13) (11, 14, 15)

Eight Clusters: (1, 3) (2) (4, 5) (6, 7) (8) (9, 10) (12, 13) (11, 14, 15)

Nine Clusters: (1, 3) (2) (4, 5) (6, 7) (8) (9) (10) (12, 13) (11, 14, 15)

The cluster analysis that used the objective data indicated that both the halfback and first five eighth are unique statistically. As a result they must be grouped separately. The number of clusters specified was increased sequentially until these positions were separated. The halfback and first five eighth are separated when nine clusters are specified as this does not cloud the distinction of roles for the specific positions. Positive increments were used as a reduction would have lead to generalisation.

At this stage it is better to err on the side of specialisation, than generalisation. Nine clusters provide an appropriate balance between specialisation and generalisation.

Having established nine clusters as a suitable number of positional clusters, the next task is to define these groupings based on the evidence provided. The objective cluster analysis provided the four following definite groups: Props, Halfback, First Five Eighth and the Midfield. The hooker needs to be treated as a separate cluster, given the specialist task of lineout throwing, as this skill obviously impacts on the overall perception of general ability. This leaves four groups to organise. Clearly locks should be grouped together, based on the opinion of coaches. Three groupings need to be created from six positions (6, 7, 8, 11, 14, 15). Two immediate groupings become clear, loose forwards and outside backs. Three realistic options are available for splitting these groups. Given the subjective evidence, the flankers could be grouped together. However, based on the loose forward argument presented earlier and the objective evidence provided the blindside flanker and number eight could be grouped together. Alternatively, the fullback could be separated from the wings. However, the subjective evidence suggests that these positions be treated similarly. Based on the belief that the openside flanker role is entirely different to any other position on the field provides sufficient evidence to separate this position from the loose forwards and treat the remaining positions in the back three as a unit. Thus the nine positional clusters to be used in the factor analysis are specified as follows:

1. Props (1, 3)
2. Hooker (2)
3. Locks (4, 5)
4. Blindside Flanker and Number Eight (6, 8)
5. Openside Flanker (7)
6. Halfback (9)
7. First Five Eighth (10)
8. Midfield (12, 13)
9. Outside Backs (11, 14, 15)

These clusters will be re-examined in Chapter Six when neural networks are explored. At this stage the nine positional clusters are deemed to be sufficiently similar and will be used to separate the data in the factor analysis. For each positional cluster, factor analysis will be used to produce specialised performance ratings.

3.5 Factor Analysis

Having identified positional clusters from the cluster analysis, it is desirable to condense the vast array of information presented into a much smaller set of variables that can be easily interpreted by interested parties. In this section the implications of engaging a dimension reduction technique are discussed. As with the cluster analysis, a rugby perspective guides the creation of relevant factor models. The data structure is unknown, but it is clear from the cluster analysis that there exists some multivariate profile that resulted in the clear differentiation between backs and forwards, and, to some extent, first five eighths. Broad headings alluding to the important features in a rugby player in each position are easily established, as are the key tasks. Evidence of such is provided in coaching guides, such as the Advanced Rugby Coaching Course compiled by the NZRFU. This provides an ideal checkpoint to which the resultant factor models can be compared and supports the decision to have at least seven clusters.

What ensues in this section is a discussion relating to the applicability of factor analysis as a dimension reduction technique in forming the basis of a numerical measure for individual rugby player performance. To successfully reduce the multivariate data to a lower dimensionality, a series of steps are undertaken. The later sections discuss the impact of highly correlated data, transformations and cleaning before the actual analysis is completed. A summary of the results is produced in this section, with the more sensitive information contained in Appendix B.

3.5.1 Underlying Philosophy

As described earlier, the desire to quantify rugby player performance is parallel to the situation that confronted Charles Spearman (1904) in the quest to quantify intelligence. The applicability of factor analysis (and other dimension reduction techniques) to create meaningful estimates of performance is based on the hypothetical construct of performance defined in Figure 2.1.

A numerical value is sought by Eagle Sport that is reflective of general rugby playing performance. Instead of quantifying performance through structured tests, the search for rugby performance must focus on unstructured rugby match participation, with the assumption that over time the nature of an individual's ability will create his career statistics based upon match performances. Due to the combative nature of rugby, non-measurable hypothetical constructs are not accounted for in structured tests. These non-measurables include ideals such as 'internal fortitude' – the willingness to continue playing when the game becomes tough. Each time an individual participates in a game of rugby this participation can be viewed as a sampling opportunity. Essentially the game becomes a forum by which individuals express their inherent ability. In line with the definition of ability featured at the start of Chapter Two, collection of match data referring to individual skill-sets is a primary starting point for quantifying performance. Reference to an individual's skill-set highlights the ability to reduce and record aspects of performance. Multivariate analysis provides the techniques to reveal the structure of ability via the multivariate profile that is expressed as performances from many match situations. Identification of the core components that constitute performance enables the capabilities of an individual to be examined for all these component skills. Every time a statistic is recorded from a match type situation, the implications of such a skill can be reasoned and established. The greater an individual's ability to play the game, the greater their potential involvement in the game. Work rate is especially obvious in the statistics that are generated. The more often a player is involved, the higher their work rate. Obviously, the higher the successful involvement, the greater the individual's inferred ability. However, as player involvement is being measured, non-performance also needs to be considered. Non-performance does not always relate to an individual's ability. Three key areas for non-performance can be isolated:

1. Player inferior
2. Player superior
3. Game structure

The first situation is of no concern as it is expected that the successful contribution of inferior individuals to the on-field performance of a team will be minimal. This means that obtained measures of performance will correctly incorporate this form of non-performance. However, the two remaining areas are of concern. A superior player may find that he is targeted or over marked in a match context. From a team perspective this can be favourable by creating space and opportunities for other team members. However, from an individual perspective, involvement will appear less than expected.

Similarly the structure of the game may not allow an individual to become involved to the extent expected. This may be due to tactics, weather or even the performances of other individuals. However, by considering many sampling situations, it is expected that an individual will have encountered many different rugby situations enabling a reliable estimate of performance to be obtained for any one player. It is assumed that performance will override performance inhibitors in the long-run.

3.5.2 Applying Factor Analysis

Having identified the positional clusters, quantifying individual rugby player performance becomes a straightforward process. Exploratory factor analyses were performed because there were no preconceived thoughts regarding the actual structure of the data. Utilising a standard statistical software package such as MINITAB, factor analysis is performed relatively quickly for each positional cluster. An iterated principal component factor analysis with a varimax rotation is employed for reasons outlined in Section 3.3.2. Whilst it is reasonable to expect some correlation between the factors on an individual level, from a group perspective (all individual performances in a positional cluster) it is just as likely that key attributes will be independent due to the mixture of different match conditions/constraints and individual skill-sets/capabilities. Orthogonality will be addressed again in Chapter Six.

The task becomes one of interpretation and ensuring the specified factors are reflective of the game of rugby. Contextual results are of the utmost importance, especially considering the statistics generated will be used in a public arena. A marketing perspective is essential, as the target population is a rugby public not familiar with multivariate statistical techniques. For this reason the marketing influence has perhaps the greatest sway on selecting the final factors. A process of cleaning and manipulation was required to ensure that firstly the factor modules were contextual, and secondly that the factors meaning could be easily communicated.

The assumptions underlying factor analysis are more conceptual than statistical (Hair *et al.*, 1995). The primary source of concern is the correlation amongst variables. Significant linear correlations must exist within the data matrix to justify the use of factor analysis. Remedies to improve the validity of the general statistical assumptions of linearity, normality and constant variance can potentially lower the correlation between variables, which may impact on the factor analysis and its applicability. However, if statistical tests (such as quality control and process monitoring) are to be performed post-analysis, transformation may be required to satisfy the assumptions of normality, linearity and homoscedasticity. A square root transformation was applied when it introduced symmetry to the data, otherwise no transformations were performed. Details of which variables were transformed are provided for each positional cluster in Appendix B.

Hair *et al.* (1995) suggests that a factor analysis is only appropriate if a substantial number of variables have an absolute correlation exceeding 0.3. As shown in Table 3.2 in section 3.2 describing the data where a partial correlation matrix is displayed, almost one third of pairings have an absolute correlation greater than 0.3 (8/28).

Further support for the use of dimension reduction techniques in developing a rating system for individual rugby player performance is provided by Cronbach's alpha coefficient. Using the same 361 observations used in the cluster analysis – where individuals had played more than 60 minutes and the difference in final scores did not exceed 30 points from the 34 test matches played during 1999 – and the 41 variables mentioned in section 3.2, a Cronbach alpha of 0.79 was obtained for the standardised data. Cronbach's "alpha is a measure of the level of mean intercorrelation weighted by variance, or a measure of mean intercorrelation for standardized data (Garson, 2001)". Consequently, the obtained Cronbach's alpha (0.79) suggests that the data is highly inter-related, therefore making a factor analysis (or other dimension reduction technique) worthwhile, as it is highly likely that a lower dimensionality exists within the data.

According to Garson (2001), the applicability of a set of variables for use as a scale can be based upon Cronbach's alpha. Given the expected variation in match conditions and circumstances, the looser structure employed in a social science setting provides relevant protocols for the handling of our data. In this instance, Cronbach's alpha should exceed 0.7 to support the use of the variables in a rating scale, which is confirmed in this instance.

Further, Garson (2001) also notes that alpha will be higher with 1) more consistent within-subject responses and 2) greater variability between subjects in the sample. In our data, 1) refers to consistency in style of play which is due to ability and skill-set, whilst 2) refers to the differing skill-sets of individuals, which is heightened in this preliminary example as all individuals from all positions were included in the initial run. With the relatively high alpha obtained (0.7912), the intuitive expectation relating to the data structure is confirmed and supports the use of dimension reduction techniques to extract measures of performance for the rugby data.

Despite the high alpha, for the purposes of modelling, each positional cluster is treated separately. Two commercial constraints direct the selection of the factor models. The first requires that all models must be comprised of the same number of factors. This induces a uniform template, allowing a standard structure to be marketed to the public. Secondly, each of the factors must have a relationship to a key facet of rugby, whether it is attack, defence, kicking or set piece play. This ensures that the models are meaningful and as a result marketable and would not be possible if a single factor analysis was performed for all positions simultaneously.

The percentage of variance criterion is based on the cumulative variance extracted from successive factors. The number of factors selected is usually dependent on the cumulative variance breaching a specified threshold. For social science applications, where information is often less precise, a threshold of 60 percent is suggested (Hair *et al.*, 1997). Due to the complexities of rugby, especially the variation between match strategies, this threshold is a relevant starting point.

When the correlation matrix is not positive definite (Smith & Teo, 1989), the factor analysis procedure will not function. This is caused by too high a correlation between variables so the removal of certain variables from the analysis remedies this situation. Variables that induced a non-positive definite correlation matrix were identified by trial and error. Constant columns also need to be deleted from a factor analysis. Through the modelling procedure it became clear that in some positional clusters, factor models produced confusing and nonsensical loadings. To rectify this situation, performances of individuals from winning teams were examined to see if contextual factor models could be obtained. This approach is justified in that the level of rugby being assessed the ultimate goal is winning. Coaches are interested in what it takes to win, or maximise the chance of winning. Over simplifying the situation, to win a team must defend and/or attack better than the opposition. Provided the matches are not overly dominated by one team, then a relative indication of overall participation is obtained. This process was performed purely to provide contextual results under pressure. Furthermore the attainment of contextual results justified the continuation of the research path.

As a consequence players are assessed according to the standards set by participants from winning teams. The 503 observations from starting individuals (jerseys 1-15) who were not in 'one-sided' matches (absolute difference in score greater than 30) and played more than one half (40 minutes) are included in the analyses. The number of observations in each positional cluster is given in Table 3.6. At this point it may seem useful to redo the cluster analysis using the same refined data set. However, for reasons that will be expanded upon in Section 3.9 and again in Chapter Six, this is not necessary. Having established limits and guidelines for the modelling procedure a structured factor analysis can be performed.

3.5.3 Results

Factor analyses were performed on each of the nine positional clusters. Five factors were required to explain more than 60% of the variation for five of the nine positional clusters and for the remaining positional clusters five factors explained close to 60% of the variation. For this reason, five factors were considered sufficient.

For each factor, a variable is considered significant if its factor loading exceeds 0.5 in absolute value (Manly, 1994). However, this 'cut-off' is a rule-of-thumb and not based on any mathematical proposition. The literature contains references to the use of values ranging from 0.25 (Chatfield *et al.*, 1989), 0.3 (Cliff, 1987) to 0.6 (Morrison, 1996; Sharma, 1996). Sharma reinforces this statement, stating that the cut-of-range specified by authors fluctuates between 0.4 and 0.6. However, Hair *et al.* (1995, p. 385) place the use of the loadings in an applied perspective, "factor loadings greater than ±.30 are considered to meet the minimal level; loadings of ±.40 are considered more important;

and if the loadings are ±.50 or greater, they are considered practically significant." Based on the remarks by Manly and Hair *et al.*, a cut-off of ±.5 will be adopted.

A summary of the results for all positional clusters is in Table 3.6. The labels of each factor for each positional cluster are also given. This table includes the variance explained by the five factors (var) and the number of observations used (k). Broad descriptive labels for the factors are used to produce a marketable product. Simple labels are useful for introducing and explaining the rating system to the rugby public. Similarly labelled factors have similar interpretations due to the combination of similar physical task variables. Consequently, the combination of similar physical task variables (whilst not identical) provides a useful estimate of the relevant skill for that particular cluster. However, it would be unwise to consider the 'Open Attack' factor for a prop to be identical to the 'Open Attack' factor for an outside back, due to the involvement these individuals have in a game. However, the factors do represent the same attributes.

Cluster	*Factors*					*var*	*k*
	1	2	3	4	5		
Props	Fringe Attck	Defence	Open Attack	P. I.	Handling	68.9%	67
Hookers	Open Attack	Fringe Attck	L. T.	Defence	Errors	63.5%	32
Locks	Handling	Fringe Attck	Open Attack	P. I.	Defence	67.0%	68
6 & 8	Fringe Attck	Open Attack	Errors	P. I.	Defence	54.5%	66
Openside	Fringe Attck	Open Attack	Handling	P. I.	Defence	61.4%	33
Halfback	Open Attack	Fringe Attck	Kicking	Handling	Defence	56.5%	34
1st 5/8	General 5/8	Fringe Attck	Kicking	Defence	Open Attack	70.6%	35
Midfield	Open Attack	Fringe Attck	Kicking	Errors	Defence	57.4%	66
Outsides	Open Attack	Kicking	Fringe Attck	Defence	Handling	59.0%	102

P.I. = Physical Involvement/Confrontational Impact, L.T. = Lineout Throwing

Table 3.6: *Summary of Factor Labels by Positional Cluster*

General 5/8 refers to the general skills of first five eighths with the ball in hand; namely passing, running and kicking. The other variable names are self-explanatory. Lineout throwing is only for hookers, as the occurrence of quick throw-ins is minimal. Physical involvement refers to work done at ruck and maul time. Specifically, this refers to early arrival at the breakdown. Depending on the positional cluster, this factor may also include possession and tackling variables. Although the weightings of these variables in the similarly named factors are not the same, they represent the same aspect of play.

Table 3.7 displays the rotated varimax loadings for the midfield back cluster. Appendix B contains further results for the other positional clusters from the factor analysis, which is kept brief to protect the commercial interests of Eagle Sports.

Variable	Open Attack	Fringe Attack	Kicking	Errors	Defence	Comm-unality
Attack Metres	**0.93**	0.12	-0.05	0.05	0.16	0.90
Attack Metres per Break	0.44	0.27	0.23	-0.05	0.40	0.48
Attack Runs per Game	**0.83**	-0.07	-0.23	0.08	-0.09	0.76
Defence Beaten	**0.73**	0.06	-0.06	0.03	0.03	0.55
Attack Passes to Hand	**0.77**	-0.09	-0.13	0.04	-0.08	0.62
Possession Metres	0.01	**-0.90**	-0.05	0.04	0.05	0.82
Poss. Metres per Break	-0.02	**-0.82**	-0.05	-0.01	0.03	0.68
Poss. Runs per Game	0.02	**-0.93**	-0.01	-0.01	0.01	0.87
Laybacks	0.10	**-0.85**	-0.03	-0.14	-0.12	0.76
Poss. Passes to Hand	0.05	-0.37	0.07	0.18	0.06	0.18
Tackles	-0.01	-0.26	-0.10	0.12	**0.67**	0.54
Missed Tackles	-0.05	-0.23	0.07	0.42	-0.28	0.31
Tackling Percentage	0.13	0.01	-0.11	-0.29	**0.71**	0.61
Loose Ball Gained	0.34	-0.06	-0.38	0.28	0.22	0.38
Kicking Metres	0.06	0.03	**-0.91**	-0.04	0.04	0.84
Metres per Kick	0.13	-0.03	**-0.83**	-0.07	-0.08	0.71
Kicks per Game	0.16	0.11	**-0.86**	-0.03	0.14	0.80
Total Fouls	-0.38	0.08	-0.24	0.18	0.43	0.42
Total Turnovers	0.15	0.02	0.02	**0.90**	0.04	0.83
Ruck Impact	0.10	-0.36	0.08	0.20	0.16	0.21
Possession Lost	0.15	0.02	0.02	**0.90**	0.04	0.83
Advantage Line Tackles	-0.03	-0.47	0.09	0.13	0.42	0.42
Turnover Tackles	0.04	-0.06	0.03	0.10	**0.62**	0.40
Tackle Assists	-0.19	-0.24	-0.02	0.40	0.21	0.30
Total Metres	**0.93**	0.03	-0.06	0.06	0.17	0.90
Tackles Broken	**0.67**	-0.04	0.03	0.09	0.04	0.46
Players Beaten	0.42	0.03	-0.24	-0.10	0.01	0.24
Passes in Tackle	**0.51**	-0.23	0.17	0.10	-0.16	0.37
Line Breaks	**0.77**	-0.20	0.09	-0.01	0.07	0.65
Handling	**0.74**	-0.16	-0.43	0.07	0.07	0.77
Tries	**0.63**	0.14	-0.17	-0.14	-0.05	0.20
VARIANCE	*6.3*	*4.0*	*2.9*	*2.3*	*2.2*	*17.8*
%VARIANCE	*20.3*	*12.9*	*9.5*	*7.5*	*7.1*	*57.4*

Table 3.7: *Varimax Rotated Factor Loadings and Communalities for Midfield Backs*

3.5.4 Discussion

The resultant factor models were deemed a success due to the fulfilment of the two main criteria, structure and context. Almost all of the communalities are greater than 0.5, indicating that the majority of the variance of the variables involved is explained by the first five factors. A convenient by-product was that nine generic labels could describe the factors for all the positional clusters. A minimal amount of manipulation was required to ensure the models where indicative of performance. This meant that some factors needed to be fully reversed. That is the loadings were weighted negatively and thus a good performance would relate to a highly negative score. By multiplying the selected factor coefficients by negative one, this problem was easily rectified ensuring continuity across all factors. An overview of the positional clusters reveals four of the factor models had a communal variance of less than 60 percent. However, as Table 3.6 indicates, the communal variance is only slightly less, with a minimum of 54 percent. Further, in most instances the sixth factor produced nonsensical solutions, either with high loadings on less than three variables, contrasting equally favourable variables (tries against defenders beaten), or combining variables that represented opposing levels of performance (missed tackles and loose ball gained). Whilst it is possible to argue the reasoning for the behaviour of such factors, this is not desirable for a rating system to be used by rugby observers. Persisting with five factors described the majority of the communal variance and most importantly provided contextual information.

The coefficients from the factor analysis are used to create the factor scores. This enables an automated process within the Eagle database. For each match in which an individual participates, a vector containing the five factors can be calculated. As a single rating is sought, a method of combining the five factor scores that constitute the performance vector must be identified. As a direct result of the factor analysis and the use of orthogonal rotation, the five factor scores are uncorrelated. Multivariate quality control is explored in the next section to assess if this combination is possible.

3.6 Multivariate Quality Control

The background theme of this thesis is derived from Shewhart's (1939) three steps in quality control and the corresponding three senses of statistical control.

> "Broadly speaking, there are three steps in a quality control process: the *specification* of what is wanted, the *production* of things to satisfy the specification, and the *inspection* of the things produced to see if they satisfy the specification" (Shewhart, 1939, p1).

Immediate parallels can be drawn to the rugby environment from these key steps. In a rugby sense, specification is the identification of the job requirements for an individual, determined by team structure, game plan and other factors discussed previously in this chapter. The multivariate techniques of cluster analysis and factor analysis provide the statistical means to quantify specification. This produces a "statistical state constituting a statistical limit to which one may hope to go in improving the uniformity of quality (Shewhart, 1939, p. 1)." Essentially, quantification of performance provides an insight into the expectation placed on players and the opportunity for rating players realistically and objectively. Uniformity of quality refers to the consistency of a player's performance. That is, not only is it desired for players to produce their best, it is expected that their best is replicated every time they take the field. Production is the on-field performance of an individual, the domain of the coach and individual. Finally, inspection is the judgement as to how well an individual has performed. Thus the first and last steps are assessed in this thesis; what is performance (specification), and how does an individual's performance relate to other competing individuals (inspection)? Multivariate quality control provides the last step in specification, by creating a single value that is representative of individual performance. However, traditional control charting theory must be modified to accommodate the relative informal setting that is presented by sports data. The suitability of control procedures is discussed initially in order to establish applicability.

Given that the measurements of the 'product' are reflective of quality, function, or performance, the nature of the 'product' has no bearing on the general applicability of control charts (Montgomery, 1997).

An important assumption associated with the application of control charts to sports data is that of normality. It is reasonable to assume performance is random over time. Cook (1977) found that elements of World Series Baseball were random. Additionally, Bracewell (1999) found that individual performance measures of New Zealand first class cricketers were uncorrelated over time. Whilst baseball, cricket and rugby are unrelated sports; at the first class level the same objectives are pursued, to win or at least maximise performance. This suggests that first class rugby performances will also exhibit no autocorrelation.

Intuitively, individual performance can be viewed as being random, given that it is expected that an individual's performance at the highest level should not be dependent on past performances. Further, "Uncertainty plays a large role in sports, and one can argue that the uncertainty associated with sports outcomes is one reason that sports are so popular (Stern, 1997, p.19)."

Additionally, given the variability in match conditions and match constraints, we will assume that, on an individual level, with respect to inherent abilities, performance is random. This assumption then enables the detection of changes in the process, or changes in the performances of an individual, which can relate to changes in performance.

However, this is not the main focus of this thesis. The mathematics of statistical quality control is used to calculate a single measure of individual rugby performance. Further, given that the assumption of independence holds, a natural and applicable by-product is the use of this measure in standard control charts.

3.6.1 Standard Quality Control and Performance Monitoring in Sport

When multivariate quality control is applied to the sport scene it is for the purpose of performance monitoring. Athletic (or sporting) performance is not mechanistic and applying standard quality control procedures would infer this characteristic. It is not appropriate to observe individual performance through purely objective statistical measures such as control charts. Utilisation of such techniques need to be balanced with qualitative measures and subjective assessment performed by the coaching staff.

The label, "*performance monitoring*", implies a less regimented approach than "*quality control*". This phrasing has important implications on how a performance measure is perceived and used.

Naming of the procedure aside, quality control is the ideal tool for assessing individual sporting performance. Monitoring quantified measures of sporting ability is very similar to the statistical area of quality control (Mosteller, 1997). The theory of quality control is almost parallel to sport statistics in that the nature of a process is quantified, such that any deviation from normal is quickly identified. Sport statistics seek to quantify sporting ability (the 'process' in this case). Control charting procedures provide the ideal medium to allow coaches, selectors and other interested parties to quickly evaluate the form of an individual relative to their expected performance level.

In the sporting situation it is more desirable to be 'out-of-control' on the upper side rather than being average. Consequently, a shift in thinking is required to accommodate the change from the attainment of mediocrity to brilliance. In fact, the interest is with the deviation from what an individual is expected and capable of accomplishing, given parameters attained from past performance. In order to achieve this, comparison is made with perfection. Successful implementation of quality control using objective information must be balanced carefully with qualitative information due to the exposure of individuals to changeable match constraints. Chapter Seven indicates how quality control should be used to monitor performance.

Current multivariate quality control is limited in the sporting context due to comparison to the average. Sports people aim for perfection, not mediocrity, thus to apply control theory to sports data the focus needs to change. Hotelling's T^2 chart can be modified to produce a "perfection method" whereby observations are compared to "unattainable perfection", rather than the average. Essentially, this performance measure is a Mahalanobis distance.

3.6.2 Existing Multivariate Control Methodology and the Perfection Method

There exist multivariate control charts capable of reducing both correlated and uncorrelated multidimensional data to a single statistic (Lowry & Montgomery, 1995; Lowry, Woodall, Champ & Rigdon, 1992; Tracey, Young & Mason, 1992; Quesenberry, 1991; Crosier, 1988; Hotelling, 1947). However, there are problems in applying this existing methodology to a situation involving sport. Bracewell (1999) showed how multivariate control charts could be used to identify out-of-control player performance in cricket. However, this approach identified only "out-of-control" performance, which obviously mixes the very good and the very bad. But, due to the low dimensionality of the data, scatter plots enabled the cause of the alarm to be identified visually. A major problem with multivariate control charts is the inability to detect the cause of an alarm quickly from the numerical value of the control statistic. Further compounding this problem is the fact that positivity or negativity of the alarm can be difficult to establish when squared measures are used. In the drive to establish a single value relating to the merit of individual performance, it is desired that a low score will relate to poor performance and vice versa. By establishing a point of "unattainable perfection" a reference point that reflects the desired context is given. The better an individual scores on a given factor or variable, the closer to "unattainable perfection". As the referenced perfection is fixed it remains constant for all, ensuring that a fair and comparative situation for rating players is employed.

Essentially a Hotelling's T^2 control chart for individual observations uses a measure of distance from a given standard when the data involved are independent. As our interest is with the highest scores possible, it makes sense to fix a point in space that is indicative of high performance and measure the distance from this. This point must be set far enough away so the probability of the largest vector component being greater than that of the fixed point on the corresponding axis is zero. The fixed point is described as unattainable perfection and is arbitrarily set at (50, 50, 50, 50, 50) for the five factors in each positional cluster. The choice of 50 as the magnitude of perfection is defended at the end of this section. The formula used to calculate the Eagle Rating, which is a scaled version of the above measure is described first.

A demonstration of the reasoning behind the development of the perfection method is best displayed in a two-dimensional plot.

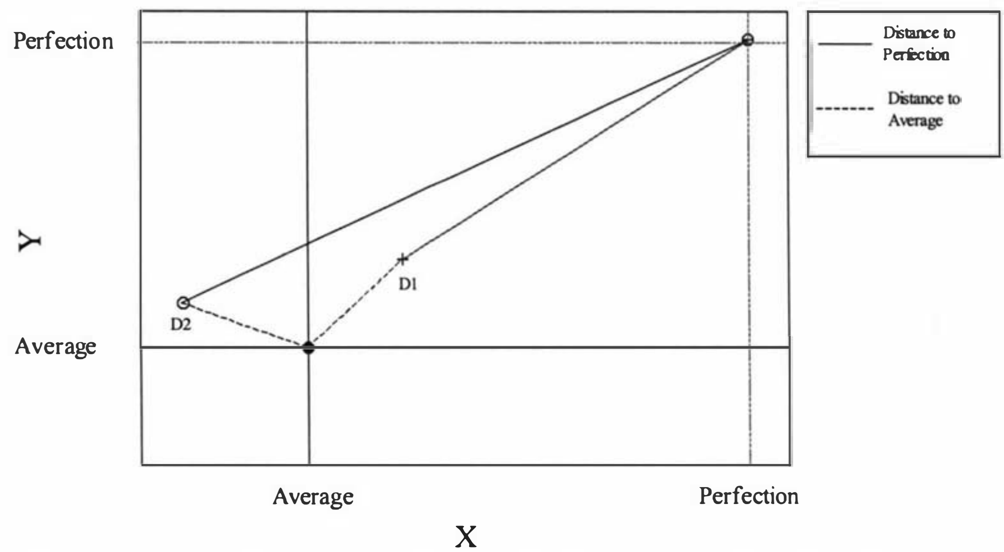


Figure 3.4: *Two Dimensional Plot of Perfection Method Explanation*

Figure 3.4 illustrates in two dimensions the justification for use of the perfection method as a quality control statistic. Instead of comparing the distance from an individual observation to the average, the point is compared with a point in space that can be described as unattainable perfection, using Euclidean distance measures. Whilst D1 and D2 are the same distance away from the average, Figure 3.4 clearly indicates that D1 is a superior performance in a sporting context as it is closer to perfection. The mathematical explanation of the above graphic can be represented in the form of Hotelling's T^2 Statistic.

Mathematical Expansion for the Comparison with Perfection Control Statistic

A modification of the Hotelling's T^2 Control chart (1947) to cater for the case presented above is achieved by comparing the input observations with "perfection" rather than the average.

Let the ith individual observation on p variables be:

$$X_i' = (x_{i1}, \ldots, x_{ip})$$

The estimated mean vector from a random sample of m observations is then given as

$$\overline{X}_m' = (\overline{x}_1, \ldots, \overline{x}_p) \text{ where } \overline{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$$

The estimated covariance matrix is described by the following

$$S_m = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X}_m)(X_i - \overline{X}_m)'$$

The control statistic for the observation X_i is then defined as

$$B_i = (50 - (X_i - \overline{X}_m))' S_m^{-1} (50 - (X_i - \overline{X}_m)).$$

The Eagle Rating uses this method to reduce the five factors to a single value that is reflective of an individual player's overall performance.

Parameterisation of the Eagle Rating

To allow the Eagle Rating to be scaled appropriately so that it is presented to the consumer in a suitable format, it is necessary to establish the properties of the control statistic that results from the perfection method.

As the Factor Scores are independent and approximately normally distributed with mean, 0, and standard deviation, 1, the distributional properties of each of the components in the control statistic can be assessed simply through basic properties of expectation and variance. Replacing X_i by $F_i = (F_{i1},\ldots, F_{i5})$ and substituting means of zero and an identity covariance matrix, the following control statistic is obtained:

$$\begin{aligned} B_i &= (50 - F_i)'(50 - F_i) \\ &= \sum_{j=1}^{5}(50 - F_{ij})^2 \\ &= 12500 - 100\sum_{j=1}^{5} F_{ij} + \sum_{j=1}^{5} F_{ij}^{\,2} \end{aligned}$$

In this expression, the second term has a normal distribution while the third has a $\chi^2_{(5)}$ distribution. However, the impact of the chi-square distribution on this statistic is minimal, because of the relative size of its coefficient. As a result the control statistic is approximately normal with mean, 12500, and variance 50000. Further, the distribution of the above rating is a non-central chi-squared variable with degrees of freedom equal to the number of independent factors (five) and a non-centrality parameter equal to 12500 (Johnson & Kotz, 1970). The non-centrality parameter is the number of factors multiplied by the squared magnitude of perfection. This information can be used to scale the above control statistic appropriately in order to obtain the Eagle Rating. A core component of the rating system is the defined magnitude of perfection.

The magnitude of perfection used for comparison is rather arbitrary, although it must satisfy three conditions; 1) perfection must never be exceeded by any standardised observation on any dimension, 2) the magnitude of perfection must be large enough so that the distances between observed performances and perfection (the rating) is approximately normal, and 3) perfection must remain fixed so that valid comparisons can be made.

It is necessary for the rating distribution to approximate normality to validate the use of standard control charting procedures. Using the five factor scores for all observations (m=2687) from the Super 12, 2000 competition provides the ideal data set to illustrate the effect of the magnitude of perfection and normality as shown in Figure 3.5. Whilst the approximation to normality is always rejected (p=0.000), the Anderson-Darling statistic, an empirical cumulative distribution function based test for normality, is used to demonstrate that the distribution of the obtained Eagle Ratings tend to normality as the magnitude of perfection increases (MINITAB, 1996). This is due to the fact that the influence of the chi-square component of the mixed distribution is reduced as the magnitude of perfection increases.

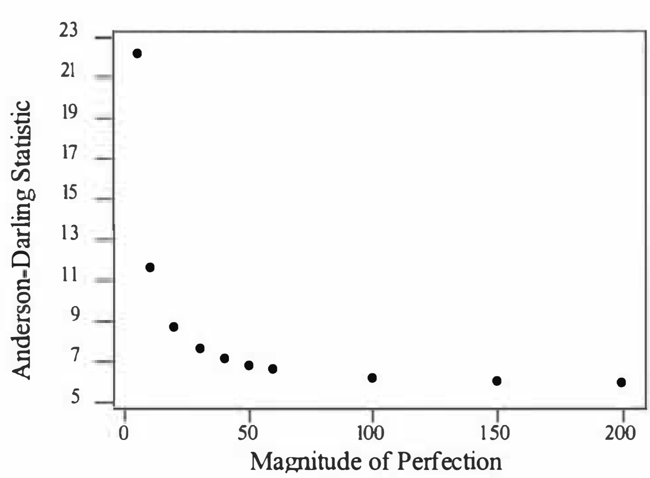


Figure 3.5: *Relationship Between the Magnitude of Perfection and Normality*

From Figure 3.5 is can be seen that the relationship between the magnitude of perfection and normality becomes relatively stable when perfection is set above 50. Additionally, the difference in ratings becomes minimal as the magnitude of perfection increases, which can be illustrated by comparing the relationship between values computed with different magnitudes of perfection, which is illustrated in Tables 3.8a and 3.8b using the correlations obtained from the data set described earlier.

Perfection	*50*	*40*	*30*	*20*	*10*
40	1.000				
30	0.999	1.000			
20	0.997	0.998	0.999		
10	0.976	0.979	0.984	0.991	
5	0.886	0.892	0.902	0.921	0.965

Table 3.8a: *Correlation Matrix Indicating Correlations Between Ratings with Different Magnitudes of Perfection*

Perfection	*50*	*60*	*100*	*150*
60	1.000			
100	1.000	1.000		
150	0.999	1.000	1.000	
200	0.999	1.000	1.000	1.000

Table 3.8b: *Correlation Matrix Indicating Correlations Between Ratings with Different Magnitudes of Perfection*

Table 3.8a and Table 3.8b indicate that the acquired ratings calculated with different magnitudes of perfection are approximately equivalent, based on the highly significant correlations ($p=0.000$). This indicates that the selection of perfection is arbitrary. The correlations for the ratings of different magnitudes of perfection are so high suggesting that inferences about performance are unlikely to be affected by the magnitude of perfection.

Finalised Mathematical Model of the Eagle Rating

It is desired for marketing purposes that the Eagle Rating is centred on 50 and has a standard deviation of 12.5, so that the probability of the value falling outside the region [0,100] is 6×10^{-5}. This is achieved by defining the Eagle Rating (ER) as follows:

$$B_i = \sum_{j=1}^{5}(50 - F_{ij})^2$$

$$ER_i = \frac{E(B_i) - B_i}{\sqrt{Var(B_i)}} \times 12.5 + 50$$

$$= \frac{12500 - B_i}{\sqrt{50000}} \times 12.5 + 50$$

with higher scores relating to better performances by the individual.

In summary the Eagle Rating is a single value rating system for first class rugby players. It is a "quality control" statistic that can be used to rate players on a scale of approximately 0-100; with zero indicating poor performance and 100 representing excellence. The Eagle Rating combines the five factors obtained from the factor analysis in a meaningful way that reflects individual performance.

Figures 3.6 and 3.7 over page indicate the resultant statistic for all individual performances during the 2000 Super 12 Competition is approximately normal in distribution. The Histogram (Figure 3.6) reveals a bell-shaped curve and the normal probability plot (Figure 3.7) is approximately linear supporting the presence of normality. The skewness and kurtosis coefficients are -0.2066 and 3.5327 respectively (skewness = 0 and kurtosis = 3 are expected for a normal distribution (Smith, 1993)). However, the Anderson-Darling statistic (6.797) and its associated p-value (0.000) (MINITAB, 1996) suggest that normality must be rejected. This is probably due to the impact of individuals who did not play a full 80 minutes of rugby because there are more low Eagle Ratings than expected in a normal distribution.

The approximation to normality is necessary for the application of standard univariate quality control methodology, such as the Shewhart control chart for individual observations. But, as far as the rugby enthusiast is concerned, the following issue is far more important.

An important consideration is that each of the five factors is given equal weighting in the calculation of the Eagle Rating. This is a debatable point. In almost all positional clusters, attack or possession based attributes dominated the first and most important factor as determined by the percentage of variance explained. However, defence is an equally important component relating to an individual's performance. This is a potential limitation of the Eagle Rating and this will be expanded upon in section 3.9.

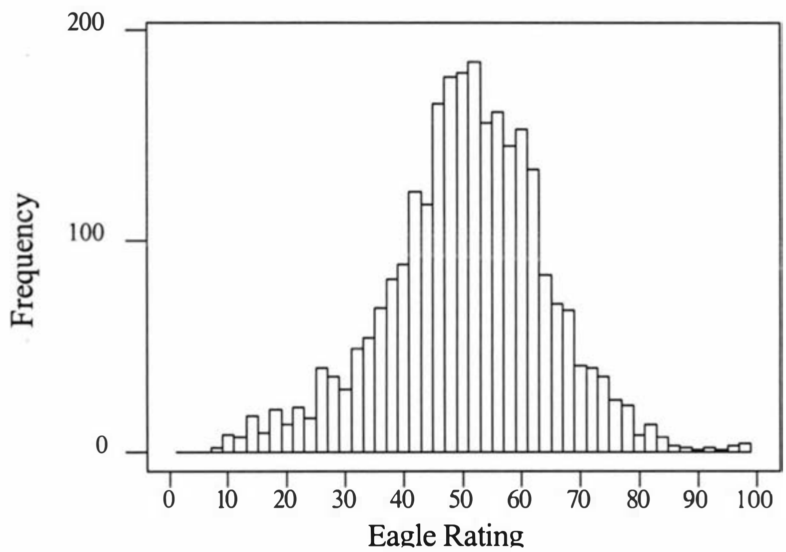


Figure 3.6: *Histogram of the Eagle Rating Showing all Performances from the 2000 Super 12*

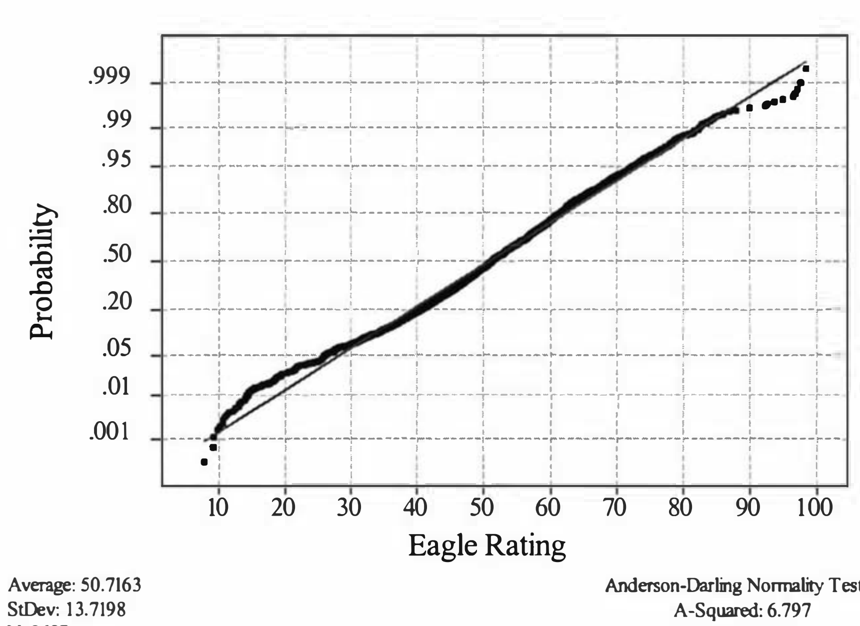


Figure 3.7: *Normal Probability Plot for Eagle Ratings from the 2000 Super 12 (Perfection Magnitude = 50)*

Standard quality control theory involves comparison to the average. Shifting the point of comparison from the average to 'perfection' produces a multivariate quality control statistic that is suitable for use in the sporting context. More importantly, the resultant control statistic can easily be associated with good or poor performance on a 0-100 scale. Thus it is suitable for a commercial sports rating system targeted at rugby observers such as coaches, spectators and the media. Additionally as the Eagle statistic is approximately normal, traditional univariate quality control procedures can be applied to monitor player performance over time. Naively, the Eagle Rating is a point estimate of individual rugby player ability. However, given that an individual's performance can be affected by game circumstance, the rating is a measure of performance. Once a single match rating is obtained, typical univariate control procedures are applicable, as the comparison to the expected overall performance is the avenue for performance monitoring. In particular, the interest surrounds the issue of whether or not the individual is performing to the expected level by assessing the variability in performance.

With the process for calculating the Eagle Rating now defined, an overview of what the Eagle Rating represents follows.

3.7 Summary of the Eagle Rating

Initially traditional techniques such as cluster analysis and factor analysis are implemented to quantify performance. Factor Analysis produced a vector of five orthogonal elements representing the key elements of performance for each individual based on the nine positional units of a rugby team identified through a combination of cluster analysis and coaching discussion.

Cluster Analysis plays an important role in the pre-processing of the data. As the role of each position differs, the natural structure of the data for each position will differ. Segregating the data by position increases the common variance explained by factor analysis due to the specificity of positional play, ultimately delivering models of a more precise nature. For this study cluster analysis used a squared Euclidean distance measure and Ward's Linkage. A number of the clusters were not clearly as defined as hoped, but hinted at groupings consistent with rugby thinking. This is because for some positions, individual's can, and do, approach their required tasks in a different manner, dependent on their skill-set. This is not of major concern as it is all part of how the game is played.

Following the identification of positional clusters, the data from each unit were rechecked and if necessary transformed to approximate normality, or at the least introduce symmetry. However because of the nature of sports data, some of the input data is highly skewed, discrete data. As a result, a square root transformation does little to remedy the skewness and these variables were left unchanged. This does not pose a problem, as the assumptions of normality and homoscedasticity are not relevant in either cluster analysis or factor analysis, only in post-analysis statistical testing.

Further data cleaning was required in the Factor Analysis. For a number of positions variables existed that were constant and hence needed removal. The presence of other variables meant that the covariance matrix was not positive definite, and thus unprocessable. A heuristic approach allowed these variables to be identified and removed. Some of the models produced confusing and nonsensical loadings. In some instances the data were reduced to look only at the performances of individuals from non-losing teams. This approach is justified in that for the level of rugby being assessed the ultimate goal is winning. Players are assessed according to the standards set by participants from winning teams. As this was a pilot study, if contextual results were not obtained for a specific subset, then the ability to obtain contextual results from the entire population would be doubtful. Almost all of the communalities were greater than 0.5, indicating that the majority of the variance was explained by the first five factors for almost all variables.

As a result of the Factor Analysis, a vector containing five factor scores was obtained for each individual, with each component being from an approximately standard normal distribution. This vector was then reduced to a one-dimensional control statistic, using an adaptation of Hotelling's T^2 statistic for individual observations, giving rise to the perfection philosophy described in this section. In normal quality control settings, any departure from the given standards, whether higher or lower, gives cause for worry.

However, in a sporting context we are not interested in deviation from the average when reducing the multivariate data to a lower dimensionality, because the attainment of mediocrity is almost as undesirable as poor play. Thus a shift in thinking was required. The introduction of the perfection method enabled the calculation of the Eagle Rating, a relevant sport statistic for assessing the general performance of individual rugby players.

Key Points in Applying the Eagle Rating

1) General Measure of Performance:

The Eagle Rating is a measure of general performance. That is, the overall aspects of an individual rugby player's performance are measured relative to those individuals in a comparative position. If a player is deficient in just one area, their rating will be less. All Black first five eighth Andrew Mehrtens is an ideal hypothetical example; he may have had a fantastic game in terms of setting up play, creating space for his outsides and gaining field position; but he may not have been involved in the game as much defensively as other first five eighths. Thus the calculated Eagle Rating will suffer. However, because the Eagle Rating is a measure of general performance, coaches should look at each of the five factor scores for specific performance measures separately, identifying if the specified tasks were performed satisfactorily.

2) Relative Measure of Performance:

The Eagle Rating scores are calculated with respect to the given match level (Test Match, Super 12, or NPC for example). Players are compared across their position, Props with Props and so forth. Obviously the Attacking requirement of a prop is different to that of a wing, and this is reflected in the composition of the models. Scores are scaled (mean 50, standard deviation 12.5) such that a maximum of 100 and a minimum of 0 provide the approximate boundaries.

3) Totally Objective:

The Eagle rating is a result of a number of mathematical calculations performed on data collected from defined codes. Human emotion does not influence the output.

Having summarised the development of the Eagle Rating, it is important to demonstrate the effectiveness of this new rugby statistic.

3.8 Implementation

Following the acquisition of the Eagle Rating, a demonstration of its effectiveness is required. The competence of this new rugby statistic can be illustrated by two methods. The first method compares the ratings with expert opinion. Secondly implementing the method in its role as a performance monitoring tool illustrates the usefulness of this new statistic. The following examples use the mean rating for the 285 individuals who played in four or more Super 12 matches during the 2000 season.

The most influential experts in estimating player ability are international selectors. Players who are perceived to be the best are chosen to represent their country. Table 3.9 ranks individuals in decreasing order according to the average Eagle Rating obtained from all matches participated in during the 2000 Super 12 competition and has a column which indicates if international honours were achieved that year. The final column shows that the table is dominated by outside backs (9). Super 12 rugby is entertainment and the results below imply that this rugby tournament favours outside backs more than test rugby does. This suggests that test match rugby should not be generalised to encompass other levels of rugby. In Chapter Six a more suitable data set is used.

Rank	*Name*	*Mean Eagle Rating*	*N*	*Franchise*	*International?*	*Position*
1	Joe Roff	68.98	12	ACT	Australia	Outside
2	Christian Cullen	68.15	11	Hurricanes	New Zealand	Outside
3	Pieter Rossouw	66.57	11	Stormers	South Africa	Outside
4	Chris Latham	66.19	10	Reds	Australia	Outside
5	Jonah Lomu	64.79	6	Hurricanes	New Zealand	Outside
6	Steve Larkham	63.45	13	Brumbies	Australia	5 Eighth
7	Rod Kafer	62.04	12	Brumbies	Australia	Midfield
8	Todd Miller	61.75	6	Chiefs	---	Outside
9	Thinus Delport	61.20	12	Cats	South Africa	Outside
10	Craig Dowd	60.84	11	Blues	New Zealand	Prop
11	Willie Meyer	60.05	12	Cats	South Africa	Prop
12	Paul Thomson	59.94	11	Blues	---	Prop
13	Troy Flavell	59.91	10	Blues	New Zealand	Lock
14	Gordon Slater	59.78	10	Hurricanes	New Zealand	Prop
15	Doug Howlett	59.77	10	Blues	New Zealand	Outside
16	Hendrick Gerber	59.16	11	Stormers	---	6 and 8
17	Kees Meeuws	58.75	9	Highlanders	New Zealand	Prop
18	Daniel Herbert	58.61	10	Reds	Australia	Midfield
19	Pita Alatini	58.27	12	Highlanders	New Zealand	Midfield
20	Robert Markram	58.11	6	Stormers	---	Outsides
21	Scott Robertson	58.10	13	Crusaders	New Zealand	6 and 8
22	Byron Kelleher	58.07	12	Highlanders	New Zealand	Halfback

Table 3.9: *Top 22 Performers in the 2000 Super 12*

The previous table, stacked with international representatives, indicates the congruency between the opinions of international selectors and the Eagle Rating. However, the entire population needs to be examined in this way.

The following table (Table 3.10) indicates the number of players (internationals) selected to represent their country for the Tri-Nations series of 2000 in each sequential group of 15 individuals. Only those who participated in four or more matches are included. The cumulative percentage indicates the proportion of internationals accounted for in the rankings. The rankings are in terms of the average Eagle Ratings obtained during the Super 12, 2000.

Ranking	*Number of Internationals*	*Cumulative Percentage*
1-15	12	16.00%
16-30	8	26.67%
31-45	6	34.67%
46-60	8	45.33%
61-75	5	52.00%
76-90	5	58.67%
91-105	7	68.00%
106-120	4	73.33%
121-135	1	74.67%
136-150	4	80.00%
151-165	0	80.00%
166-180	4	85.33%
181-195	1	86.67%
196-210	3	90.67%
211-225	3	94.67%
226-240	1	96.00%
241-255	3	100.00%
256-270	0	100.00%
271-285	0	100.00%

Table 3.10: *Allocation of Tri-Nation Representatives Ranked by the Eagle Rating*

Clearly, Tri-Nations representatives dominate the upper echelons of the ranked Eagle Ratings. This supports the Eagle Ratings use as a point estimate for individual rugby player performance and ability. The failure of the Eagle Rating to predict international selection in some instances can be explained as follows. Individuals are not selected solely on the basis of the performances expressed during this tournament. Prior performance is also considered. This accounts for a number of the internationals at the lower end. Also when an international is playing in a side that is significantly weaker, dominance and involvement in the game will suffer, and hence the Eagle Rating will also be diminished.

Another apparent "expert" is the media. Following each All Black test, Wynne Gray, sports writer for the New Zealand Herald gives each individual a match rating, based on his subjective opinion. The following graph compares the Wynne Gray opinion with the objective Eagle Rating from the three All Black tests played on the 2000 Northern Hemisphere Tour. The discrete measurement scales for the Wynne Gray rating is confined to only five points in this case.

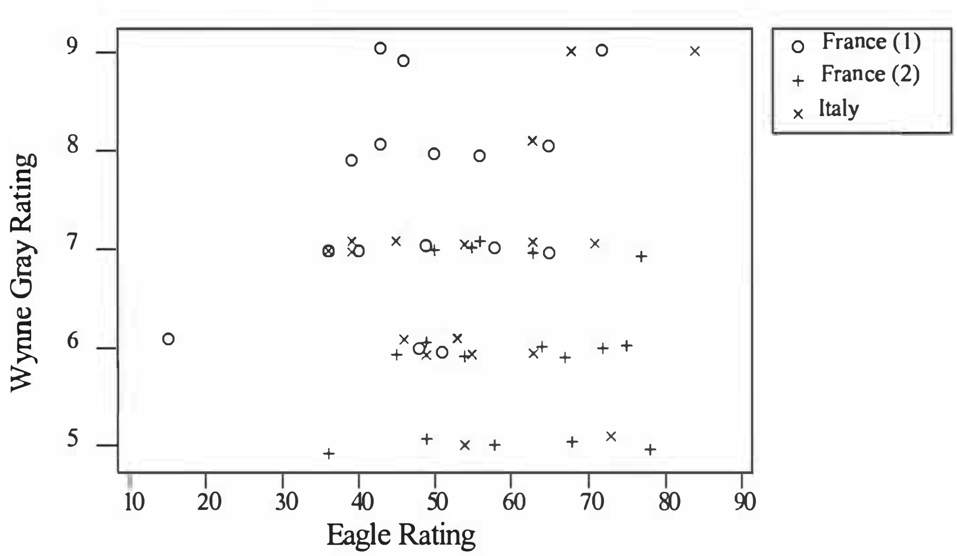


Figure 3.8: *Comparison of the Eagle Rating and Wynne Gray*

Figure 3.8 suggests there is no clear relationship between the Eagle Rating and the Wynne Gray ratings. The Wynne Gray rating is significantly higher (ANOVA p-value = 0.000) for the first French test won by the All Blacks than the subsequent second test in which the All Blacks were defeated. This is indicative of personal biases that may enter a structured rating scheme. The Eagle Rating gave higher ratings for the second test (ANOVA p-value = 0.017). This is easily accounted for, as in the first test there were 179 sequences compared to the 207 sequences in the sequel. Obviously as more rugby was played, individuals participated in the match more and consequently were rewarded with higher ratings; approximately 16 percent higher, coinciding with the 16 percent increase in sequences.

Importantly, it must be considered that Gray's perception of the game is based only on real time footage, whereas the Eagle coders have the benefit of repeated slow motion replays. Intuitively, the ratings suggested by Gray are dependent on the perceived structure of the match. In the French tests, where New Zealand was expected to avenge the 1999 World Cup semi-final loss, the emphasis of Gray's ratings appears to have been placed on traditional test match characteristics, forward domination and tactical kicking. Indeed, this is supported by correlations extracted from the factor scores that comprise the Eagle Rating and a standardised Wynne Gray rating. The Wynne Gray rating was standardised with respect to each match, which removes match by match bias in an attempt to remove between match variability. Interestingly, there is a highly significant correlation between the Physical Impact factor score for positional clusters which have this key performance indicator and the Wynne Gray rating in the two tests against France ($r = 0.576$), Further the score on the Kicking factor for positional clusters with this attribute also had a strong linear relationship with the Wynne Gray rating with a correlation of 0.527.

However, in the much looser Italian game where the All Blacks were victorious by a comfortable margin (56:19), physical involvement was not considered as crucial with a correlation of –0.365. Thus it appears Gray's ratings do have a degree of objective merit, albeit in a narrowed scope. Or conversely, elements of the Eagle Rating match the subjective view of media "experts".

A valid expectation is that the combined ability of the selected teams will exceed that of the non-selected individuals. Data relating to the previous example in Table 3.10 is employed. The following ANOVA indicates that this expectation yields a significant difference. Table 3.11 reveals a significant collective difference between those considered for international duties during the 2000 Tri-Nations tournament, and those not selected, supporting the use of the Eagle Rating a performance measure that is reflective of individual ability.

Another expectation is that, in general, victory would correspond with higher Eagle Ratings, as to win teams collectively need to perform better than the opposition. Another ANOVA is performed to test this conjecture using data obtained from individual's who played full games (80 minutes) in the 2000 Super 12.

```
One-way Analysis of Variance
Analysis of Variance for the Average Eagle Rating
Source       DF        SS         MS        F        P
Team          3    2086.1      695.4    17.68    0.000
Error       281   11049.0       39.3
Total       284   13135.0
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level         N      Mean      StDev  ---+---------+---------+---------+---
Others      211    48.012      6.421  (--*--)
NZ           27    55.358      4.876                    (-------*------)
Australia    26    54.084      5.962                (-------*-------)
SA           21    52.149      6.647         (--------*--------)
                                      ---+---------+---------+---------+---
Pooled StDev =      6.271             48.0      51.0      54.0      57.0
```

Table 3.11: *ANOVA differentiating between Internationals and Non-Internationals*

```
Analysis of Variance for Eagle Rating by Result
Source       DF        SS         MS        F        P
result        1      4462       4462    30.00    0.000
Error       572     85069        149
Total       573     89531
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level         N      Mean      StDev  ----------+---------+---------+------
Loss        245     46.77      12.14   (-----*-----)
Win         329     52.41      12.24                           (-----*----)
                                      ----------+---------+---------+------
Pooled StDev =      12.20                    47.5      50.0      52.5
```

Table 3.12: *ANOVA demonstrating Differences Between Winners and Losers according to the Eagle Rating*

Once more the result is conclusive, indicating that individuals in winning teams tend to have significantly higher Eagle Ratings (p=0.000). To win a game, it is expected that one must play better than the opposition, and the ANOVA in Table 3.12 supports this verisimilitude. The previous examples illustrate that the Eagle Rating produces results that are expected given that the rating is representative of general playing performance.

The following example utilises the Eagle Rating as initially intended. In this example, shown in Figure 3.9, the performance of All Black Fullback, Christian Cullen, is monitored throughout the 2000 Super 12. A stand out feature of Figure 3.9 is that the four "lows" of Cullen's Super 12 season which correspond with losses incurred by the Wellington Hurricanes. Cause and effect cannot be established from this graph.
That is, did the Hurricanes lose because Cullen played poorly (by his standards) or was he shut out of the game by the opposition? Perhaps Cullen was used in a decoy mode or maybe even weather conditions hampered his involvement?

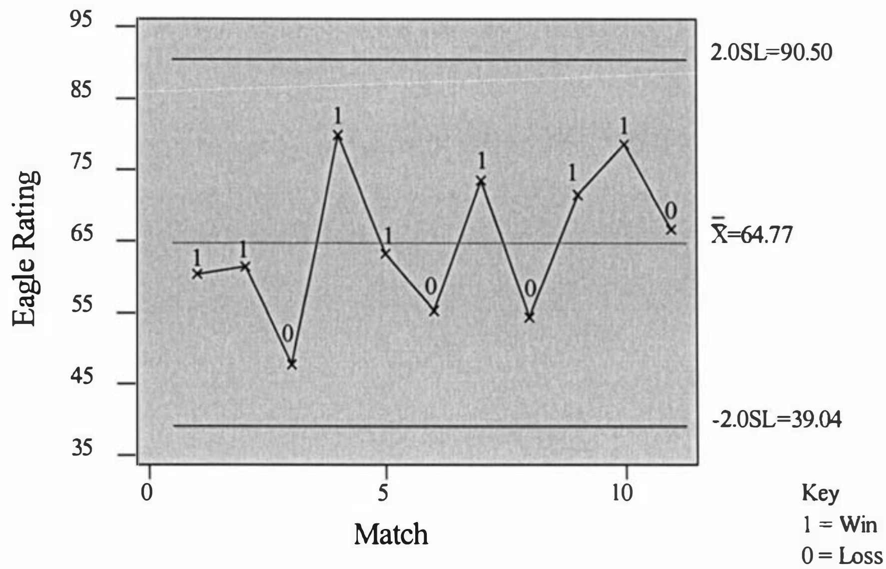


Figure 3.9: *Shewhart Control Chart Monitoring Cullen's Super 12 Performance*

The answers to such questions can be identified by assessing the factor scores, raw match statistics and video footage, enabling game strategies to be developed that maximise the potential for victory. However, the fact that no alarms were signalled for Cullen suggests that he should not be blamed for the losses.

Upper and lower control limits are set two standard deviations away from the centre line as these tighter limits are more applicable for application in sport (Bracewell, 1999). This is due to the limited sampling opportunities. In the case of the Super 12 Tournament, there will be at most 13 matches played by any given individual, assuming their team makes it through to the play-offs. This means that a control chart has little power (opportunity) to detect "out of control" situations. Tight upper and lower control limits set two standard deviations away from the centre line are therefore appropriate. In an additional attempt to increase the power all eight run rules are applied to identify any changes performance.

In addition all eight run rules are applied (Montgomery, 1997) in order to detect gradual trends and other patterns as well as extreme performances. As no alarms are signalled, this suggests that Cullen performed consistently to his expected level of ability. Considering that the population mean is 50, we can see that Cullen's reputation as a world class Full Back is well founded given his mean Eagle Rating of almost 65.

3.9 Limitations

This section will focus on the limitations of the modelling process, specifically the mathematical techniques involved and not the overall ideology behind the calculation of the rating, which will be broached in Chapters Seven and Eight as the thesis concludes.

The main limitation relating to the first publicised version of the Eagle Rating is the dependency on factor analysis and the inherent assumption of linearity. The success of reducing the dimensionality of a large number of variables is dependent on the dimension reduction technique applied (assuming that some latent structure exists within the data provided). A major problem is the linear dependency of factor analysis, which was exposed during the Sky Man-of-the-Match Competition, is that as the amount of play increased, so did the ratings. Additionally, significant non-linear relationships may exist. The non-linearity of relationships will be investigated through the use of neural networks and self-organising maps as dimension reduction tools.

Other potential limitations are loosely associated with the linear dependence of factor analysis. These need to be identified before attempting to produce an improved rating.

The nature of any model is restricted by the data included. A wide variety of measures are specified in the factor models. However, there are many situations that are not fully specified. The development of the Eagle Sport database is an on-going process, new coding procedures and improvements are researched regularly. Having shown the ability to quantify individual rugby player performance, it is a simple process to include new variables as they are created and reassess the models once sufficient data is collected to enable inter-variable relationships to be estimated. However, there is always the information/resource trade-off that occurs when collating the data. The expense of the data collection process must be compared with the value of the obtained data. The need for new data must consider this trade-off. A situation where new information is required is when the typical structure of the game alters.

It could be argued that coaching staff could impact upon the obtained rating by substituting players at different stages of the game and playing certain players against weaker opposition. However, implementing 'quality control' into a diagnostic coaching structure, introduced in Chapter Seven, puts the use of this methodology in perspective. This allows coaches to identify performance issues and identify the associated physical task grouping. At this point it is the responsibility of the coach to call upon all available information to decide the appropriate action. In general, a coach is provided with a team selected by more than one individual. Additionally, the number of substitutions available to a rugby coach is restricted. This reduces the impact a coach may have upon player ratings through selection and substitution.

The style in which rugby is played is dynamic. Rule changes and techniques employed by successful teams ensure that the manner in which a rugby match is contested is continually evolving. Thus models employed to represent performance must continually evolve to remain relevant. The impact of rule changes on positional clusters cannot be ignored. The effect of the new lineout laws allowing lifting and the reduced emphasis on set piece play could have potential impact on the physiological nature of players selected in the forward pack. As tall, heavy players are harder to lift in the lineout their place may be taken by smaller, more mobile individuals. This could lead directly to a greater skill-set overlap between locks and loose forwards and thus more generalisations between these positions. Already, children are taught that in general play, everyone is a loose forward in that it is everybody's responsibility to retain/gain possession. However, with mauls now able to reconstitute, there is also the need for bulk. Additionally, scrummaging will always be an important component of the game because of the necessary platform it provides in restarting the game after minor infringements (knock-on, accidental off-side and so forth). Regardless of the physical body types, models must be adaptable to cater for changes in the game. An obvious change in the 2001 season is at the kick-offs starting each half. Most teams now place the ball on a high cone, rather than a standard kicking tee, to enable the kicker to gain more height, therefore making the kick more contestable. Ongoing adaptation is explored as part of the neural network learning process in Chapter Six.

Equal weighting was afforded to each of the factors. This approach is potentially unsound as the relative importance of tasks differs. Further some skills are deemed more important than others, such as defence and handling ability. The approach adopted in this chapter is relatively safe, as it does not place any importance on any particular skill. A potential resolution is to create two areas for evaluation, attack and defence, and providing equal weighting to the obtained attack and defensive measures. This is a valid approach as, stated simplistically, a team either has the ball or it does not, consequently the statistics collated can be set to relate to either condition.

Scaling with respect to involvement is another area for consideration. The amount of play in each match fluctuates, as mentioned in section 3.2. The sigmoid activation functions, introduced in the next chapter, prevent the tendency for values to increase as the amount of play increases. However, if a linear approach is adopted, then scaling with respect to the amount of play provides a useful frame of reference.

The handling of partial games is an issue. If the contribution to the team cause is considered, then the overall rating obtained would suffice. However, if a measure for relative individual involvement is required, the rating must be adapted to consider time spent on-field. However, with the prolific use of "impact" players and extensive use of the reserves bench, there are many issues to be considered, involving some of the following:

- If the rating for an incomplete game performance is adjusted to account for on-field time, does this penalise those that participated in the full match?
- Is an individual entering the fray amongst tiring competitors expected to contribute more?
- If the individual was the best available (based on skills and fitness), would they have started the match?

The expectations placed upon an individual will differ based on the game plan, situation, and individual skill-set. Consequently, the initial approach of considering the contribution to the team cause provides the safest and easiest method for rating partial performances. However, Chapters Six and Seven will assess a hybrid of dimension reduction techniques to introduce robustness to the rating, stabilising the impact of varying amounts of match play and partial match play. In this thesis "robustness" is the ability of the performance rating, or the rating component under consideration, to withstand variable match conditions, except where stated (Chapter Five).

Individuals changing from one positional cluster to another mid-match are also of concern. The prototype introduced in this chapter ignores this effect on the basis that neighbouring positional clusters have relevant overlap in the measured skill-sets, providing approximately the same results. This effect is problematic but cannot be resolved in any cost-effective manner. Obviously the ratings could be adjusted, but there are many occasions when individuals move out of position to accommodate the skill-sets of team-mates, such as moving the halfback into the lineout to allow a bulkier individual to participate in a set move off the lineout. Thus the effect of recording all shifts in positional play would be time-consuming and provide little additional information.

Cluster analysis is inappropriate as a method for identifying positional clusters. This conclusion is more conceptual rather than mathematical (Hair *et al.*, 1995). Already there has been given considerable reference to individual skill-sets, positional responsibilities and so forth. Because skill-sets and positional responsibilities overlap the positional clusters defined by expert opinion are not separable (linear or non-linear). This is not a problem, as the positional clusters can be defined based on expert opinion and it is this expert opinion that will provide the basis of the positional clusters in Chapter Six.

The act of clustering is an important step in data pre-processing. From a rugby perspective, prior to the commencement of any analyses, it must be accepted that the roles and requirements of distinct positions differ. Whilst the roles of certain positions are congruent to some extent, dependent on team structure, capabilities of the individual and state of the match in progress, the most realistic approach to using any identified clusters, is to identify clusters by position. This enables ratings to remain contextual and allow for direct comparison between individuals competing for the same position in a selection situation. Thus before we can begin modelling, we must identify relative groupings (Gerber, 1996). This also acts as another form of data cleaning/ preparation. The clustering procedure implemented in Chapter Three is fraught with limitations, most importantly the inability to provide clusters refined to balance the need for specialisation and generalisation. It is for these reasons that the clustering procedure is not implemented in the chapters to follow. The dynamics and exposure to match volatility have been mentioned repeatedly, where the relative participation experienced by an individual is affected by the structure of the match.

These limitations are considered below in more detail. Three naturally forming clusters were identified in section 3.4, which can be attributed to three simple areas, namely, forwards, backs and first five eighths. An overview of the coding system reveals that the tasks measured by attack and possession attributes appear to overlap and this re-labelling of similar tasks reduces the impact that the sum of these variables has in the cluster analysis. Essentially the tasks are the same, except attack attributes are noted when the play involvement is akin to standard back-type play. Similarly possession attributes are awarded when typical forward-type play is engaged. The division between backs and forwards becomes obvious; backs are more likely to engage in typical back-type play, just as forwards display more forward-type behaviour. The effects of this can be most pronounced on the resultant match rating. Consider the All Black European Tour of 2000 where All Black number eight, Ron Cribb played a very wide ranging, loose role. The focus of his play was involvement in the backline, notably entering the backline in midfield. This was highlighted in the resultant match statistics (www.eaglesports.co.nz/skytv/skyrep.asp), where it is obvious that Cribb shirked the workload expected of a traditional loose forward and adopted a style of play expected from a midfielder. Importantly, the division of the clusters into three is shaped by performance on three different groups of variables, attack and possession mentioned above, and goal kicking, mentioned below. This division appears simplistic and accounts for the lack of specialisation that occurred in the cluster analysis.

The third cluster, identifying first five eighths is based on goal kicking attributes. The majority (78%) of first five eighths and a small collection (23%) of fullbacks are found in the third cluster. The majority of first five-eighths goal kick, although not all do as some fullbacks also take on the goal kicking duties (as well as other positions). These duties are not restricted to solely these positions, but generally ability to kick is a core component of the skill-sets of the named positions. Former Australian and Queensland lock John Eales is a prime example of an individual skill-set covering more than the positional specific requirements. New Zealanders will not readily forget Eales converting a last minute penalty to secure the Bledisloe Cup and Tri-Nations championship at Wellington in 2000. This highlights the frailty in depending on one variable to create the positional clusters.

Additionally the effect of non-performance will hamper any clear divisions. Essentially it is difficult to distinguish statistically the difference between a prop that has had an outstanding game with the ball in hand and an outside back that has played poorly as their general involvement statistics will be similar.

The limitations described in this section are offset partially by the major strength of the models developed, which is transparency. Additionally, ignoring the cluster analysis in favour of the coaching clusters offsets the limitations in the three-unit cluster. Consequently, no such problems exist when the factor analysis is performed for each of the nine clusters. Enquiries relating to the scoring system employed in the fantasy game, Ultimate Rugby could quickly be answered with sound statistical evidence, obtained in a similar manner to that explained in Chapter Seven and the application of statistics in a diagnostic coaching structure, working backwards from the single performance measure ultimately to the raw data.

In attempting to create an expert system, we are trying to create a system that can interpret the game of rugby based upon the data presented. Results from this chapter have made a good start, but obviously more work is required. The ratings and KPI's (factor scores) obtained from this thesis are to be compared with expert opinion which is based on life times of playing and participating in the game enabling full understanding of the idiosyncrasies of the game of rugby.

3.10 Conclusion

This chapter outlined the creation of the Eagle Rating as a sport statistic that represents an estimate of individual rugby player performance on a match by match basis. Three steps were required. At each step the philosophical justification of the procedures and rugby context were emphasised. This was crucial in defining a relevant product. Whilst the limitations of linearity are expressed in the preceding section, the success of such a statistic depends on its acceptance in the public forum. The Eagle Rating was immediately adopted in the commercial environment. Ultimate Rugby (see Appendix D) and the Sky Man-of-the-Match competition were based on the Eagle Rating, developed from the work described within this chapter. The implementation section

(3.8) indicated the potential uses for the Eagle Rating and demonstrated its effectiveness in a rugby context. Limitations aside, the primary version of the Eagle Rating is a unique statistic worthy of attention in the rugby arena.

The early stages of this chapter discussed the data collection process and systems in place to ensure quality data are obtained. Based on the large amount of individual rugby player data collected by Eagle Sports on a match by match basis, it was required that a single match rating that described an individual's performance be developed to satisfy commercial interest. As a number of the variables collected are highly correlated, dimension reduction techniques were applied to reduce redundancies in the data in each of nine distinct positional clusters. The positional clusters were identified using expert opinion. The application of factor analysis produced five factors that summarised individual rugby player performance, explaining more than 60% of the original variability in the data set for most of the nine clusters. Due to the commercially sensitive nature of the project, full details of the data and the components of the five key factors cannot be given. Multivariate quality control was then adapted and applied to the five factors scores to reduce them to a single value, the Eagle Rating, using a Mahalanobis distance.

Validity and potential uses for the obtained rating were then briefly examined. Most importantly, the aim of this chapter, to show that individual rugby player performance can be quantified using multivariate analysis has been fulfilled. The remainder of this thesis looks to improve on the primary Eagle Rating by catering for non-linear relationships within the data through the use of neural networks and self-organising maps. Chapters Four and Five review the technical details associated with neural networks before Chapter Six seeks to improve the methods for quantifying individual rugby player performance that were used in this pilot study. Importantly, this chapter has shown that cluster analysis is inappropriate and that expert coaching opinion should be used to identify the positional cluster. Additionally, underlying factors, or key performance indicators, must be contextual. This is essential to provide transparency and defend the structure of the factor models. These lessons will be incorporated into the analyses in Chapter Six.

Part Two

Improving the Eagle Rating

Chapter Four

Literature Review – Data Mining and Neural Networks

This chapter gives a brief overview of data mining then focuses entirely on artificial neural networks, specifically feed-forward neural networks and self-organising maps. This is due to the focal point of this thesis on methods capable of dimension reduction.

In terms of the research questions proposed in this thesis, this chapter is somewhat redundant, intended to provide a general overview of the neural network techniques involved. Whilst the applicability of the techniques is explored with respect to the rugby data, this chapter can be avoided by those who are more interested in the application of the techniques.

Three techniques are investigated in this thesis. The first is factor analysis, as shown in Chapter Three. This is followed by neural networks and self-organising maps which are described in this chapter, developed further in Chapter Five and applied in Chapters Six. All three methods are compared in Chapter Seven. The details of these techniques are available widely in the literature with the texts by the following authors providing an excellent overview (Haykin, 1999; Berry & Linoff, 1997; Bishop, 1995; Fausett, 1994).

Traditional multivariate analysis techniques for dimension reduction, namely principal component analysis and factor analysis, are special linear cases of non-linear capable neural networks with both methods encompassed by the field of data mining (Cheng & Titterington, 1994; Warner & Misra, 1996), with which this chapter commences.

4.1 Data Mining

There is no shortage of literature detailing the usefulness of data mining techniques for a wide range of fields and problems (e.g. Mackinnon & Glick, 1999; Glymour, Madigan, Pregibon & Smyth, 1997; Jorgensen & Gentleman, 1998; Gerber, 1996; Hand, 1998; Moxon, 1996.).

Essentially, "data mining is a set of techniques used in an automated approach to exhaustively explore and bring to the surface complex relationships in very large data sets (Moxon, 1996)." Mackinnon *et al.* (1999) reiterate this definition describing data mining as "the extraction of previously unknown information from data bases that may be large, noisy and have missing data (p. 255)" and further cite the use of this definition by Chatfield (1997) and Fayyad (1997).

The goal of data mining is to extract useful, but previously unknown information, from typically massive collections of non-experimental, sometimes non-traditional data (Mackinnon *et al.*, 1999). The data set presented in this thesis is certainly non-traditional and non-experimental; further, with over 1000 games logged over a five year period, the data set is large, but not massive by data mining standards. Moreover, the information to be extracted (the quantitative performance of an individual) is unknown, but often guessed by sports writers such as Wynne Gray for the New Zealand Herald and Mark Hinton for the Sunday Star Times. Both these sports columnists have published their own quantitative ratings (Scaled 0:10) following All Blacks test matches as discussed in Chapter Three. This is the ultimate aim of the thesis to quantify individual rugby player performance as a single value, based on measures of performance gathered for each individual on a match-by-match basis.

The database used in this thesis was described as unique by Potter (2000) following the death of Trevor Eagle. In fact, as shown in the Chapter Two literature review, very little work has been done with large sports databases in terms of analysing game play data to understand performance in rugby. At the end of this section an example using the data mining of basketball data is provided.

Having briefly described what data mining is, the thrust now turns to illustrating the advantages this methodology and philosophy has over traditional methods. Like any technique, data mining has its proponents and opponents. The key to the main argument against the use of data mining comes from the previous definition by Moxon (1996) and the use of the term 'exhaustively explore'. This phrase has negative connotations and is associated with other terms such as 'data dredging', and 'fishing expeditions' in search of significant p-values (Mackinnon *et al.*, 1999). Chatfield (1995) suggests that data mining is associated with an analyst going to extreme lengths to obtain a good fit. The underlying criticism refers to over-fitting, that is, too many model parameters. Chatfield (1995) also states that neural networks are a class of models that may lead to over-fitting.

Overfitting occurs when a more complicated model appears to have a better fit than a simpler model, but when presented with new data the performance deteriorates. This is due to the forced implication of a given family of models for the solution. Again referring to neural networks, "the large number of parameters (and architectures) which may be tried means that they can usually be made to give a good within-sample fit (Chatfield, p. 428, 1995)." The idea being dealt with here is the specification of a model to fit the data. As a result over-fitting is in part the over-dependence on a particular (complicated) model structure. It is on this note of over-fitting that we see the transition to the positive consequences of data mining. If the model structure selected is incorrect, over-fitting effectively creates a reliance on the parameters within this model that may produce incongruent solutions when presented with new data. Reliance on long ranging behaviour from such models is dangerous (Newbold, Agiakloglou & Miller, 1993). Here is the shift in the argument. Traditional methods usually assume that there exists some (simple) true model (Fildes & Makridakis 1994), usually a linear model.

> "The estimation of model parameters traditionally assumes a model has a prespecified known form and takes no account of possible uncertainty regarding the model structure. This implicitly assumes the existence of a 'true' model, which many would regard as a fiction." (Chatfield, p. 419, 1995)

In contrast data mining allows a wide range of model forms, called a solution space. For any given problem there exists a good model existing in the solution space. Due to computational difficulties or the unattainability of answers due to mathematical bounds, this might not be the 'true' model but there will be models that closely approximate the 'true' model. Some of these models will be more satisfactory than others. How then is the optimal model from the infinite solution space found? Obviously the methods applied need to be capable of covering as much of the solution space as possible. This is where neural networks come into the frame with the ability to represent non-linear models. The important thing to note here is that whilst non-linear modelling can cater for conventional linear models, the reverse is not true. As a result neural networks can produce simple linear models or complex non-linear models, while traditional methods are usually restricted to linear models.

Traditional methods restrict analysts to just one model, or family of models, believed to be the 'true' model. Data mining makes use of the computer led revolution and allows different models to be fitted simultaneously or combined (Chatfield, 1995). Thus not only is a greater region of solution space made available, but more tolerant modelling approaches are applied. It must be remembered that although non-traditional, non-experimental, noisy data is being dealt with, the size of the data sets allow the use of more complex, more realistic models.

This highlights the benefit of techniques such as neural networks, which do not assume any particular model structure, which in turn reduces the potential error of model uncertainty to some extent. Noteworthy in the preceding paragraphs is the need to beware of over-fitting in neural networks. This can be avoided by the use of early stopping procedures, which will be explored later in section 4.4. Early stopping ensures that a model does not fit a set of data so well that it fails when applied to fresh data.

A major advantage of modern techniques is their departure from the following assumption. 'Traditional statistical methods tacitly presume that phenomena or populations are static (Mackinnon, p. 257, 1999)." When presented with new data, or new situations, modern techniques are able to adapt to the new data; that is, they are able to learn (Balakrishnan, Cooper, Jacob & Lewis, 1994). The data presented by rugby is certainly not static, as evidenced by the evolution of game tactics. This is most easily demonstrated through the adoption of a 'Brumbies[1]' style of play by the 2000 All Blacks, coached by Wayne Smith, where teams looked to retain possession regardless of field positioning. This was a change from the previous season under John Hart employing a more traditional style of play, utilising the long kicking abilities of first five-eighth Andrew Mehrtens to first gain field position in the opposition half, then applying pressure to force an error in order to regain the ball. The evolution to a league style of play in the 2001 Super 12 season – where attacking individuals receive the ball close to the breakdown and look to either break the flat, condensed defence or reset – is highlighted by the additional method individuals have adopted to lay the ball back; through the legs as opposed to the traditional side-on plant.

The major disadvantage of data mining, namely over-fitting, is outweighed by the advantages. These are non-specificity of models, enhancement of the solution space and adaptation to dynamic data. An example is given to illustrate the usefulness of data mining in a sporting environment.

The ideal example to show the use of data mining in a sporting environment is presented by Bhandri, Colet, Parker, Pines, Pratap and Ramanujam (1997). It details the use of data mining with National Basketball Association (NBA) data. Utilising a specially developed data mining package, 'Advanced Scout', this paper demonstrated how such applications enabled NBA coaching staff to discover interesting patterns in basketball game data. In the 1995-96 season, the software was distributed to 16 of the 29 teams who indicated that it was a valuable tool. The package was well received by the NBA as it facilitated a higher quality of play, which in turn provided additional value to the fans. The raw data collected includes such information as who took a shot, the type of shot, the outcome, any rebounds and so forth. Each action is associated with a time code, an important step allowing the data and video footage to be aligned such that any outlying or strange occurrences can be witnessed visually. After the data are checked for missing data and impossible events by the software and a domain expert (typically a coach), transformation and reformatting of the raw data occurs, enabling a presentation to be made available to the coaches in the form of a play sheet. With the

1 ACT Brumbies: Champions, 2001 and Runners up in the 2000 Super 12 competition. Tactics revolved around ball retention and decoy runners.

data entered into the system, coaching staff can then instigate a general data mining query, initiating an automatic search for interesting patterns of play in a specified field, such as shooting performance, or possession analysis; events that lead to either favourable, or unfavourable outcomes. The information obtained can then be used to optimise team selection and determine game plays. The major point to derive from this study is not the actual methodology, but the perceived usefulness from the coaching, NBA and fans perspective.

The amount of statistical information required, and potentially generated by basketball, is much less than that for rugby for several reasons. Most importantly, there are more complex interactions involved in rugby (30 players interacting compared to 10). The game objectives are much simpler for basketball (put the ball through the hoop to score points, as opposed to either scoring through one of the four possible avenues: try and conversion, field goal or penalty). However, the article described demonstrates the use of techniques that are designed to successfully extract unknown information from non-traditional sources in a sporting environment. This is encouraging for the use of neural networks, in a rugby environment. The latter stages of Chapter Seven will illustrate how a univariate measure of performance obtained from multivariate data can be used to identify anomalous performances in a coaching environment.

The NZRFU provides the provincial unions with a rudimentary statistics package developed and provided by AnalySports, which is similar to that described above. Events are recorded on DVD enabling instant access to visual events identified using a time stamp. However, the ability to analyse relationships or conduct statistical analysis beyond simple arithmetic is restricted. A number of the measures incorporated into the AnalySports package are a replication of the data collated by Eagle Sports. This means that the multivariate package developed in this thesis could be incorporated into the system provided by AnalySports (or any similar system) to increase the power of statistical analyses available to coaching staff, as outlined in Chapter Seven.

4.2 Neural Networks

To commence this section the origins of artificial neural networks are described. Then a brief history of neural networks is outlined before a selection of examples demonstrating the variety of applications in which neural networks have been tested is provided. A general description of the structure of neural networks concludes this section. The description is deliberately left till last enabling an unimpeded flow into the remaining sections detailing the neural network functions of specific interest, notably networks capable of dimension reduction.

4.2.1 Relation to Biological Networks

Artificial neural networks are models that are designed to emulate a biological neural network inherent in the structure of the human brain (Trippi & Turban, 1993). The desire to mimic the brain is borne out of the fact that whilst computers are quick and accurate at solving simple numerical procedures the human brain still out performs artificial intellectual power on complex tasks (Becker, 1991). "On a practical level the human brain has many features that are desirable in a computer (Warner & Misra, p. 284, 1996)." These include the ability to generalise from abstract ideas, recognise patterns amidst noise, quickly recall memories, and withstand localised damage (Warner *et al.*, 1996). The performance on these tasks is characterised by the ability to withstand noise and ambiguity whilst simultaneously satisfying many constraints (Becker, 1991). All these attributes are desirable in a model that quantifies sporting ability, or match performance. However, one of the most important features is the ability to withstand localised damage. This credits the model with some immunity to violation of assumptions, which will be shown later to be crucial in accepting a particular type of model. Subsequently, the lessons from biology may provide clues to the suitability of a particular model.

The extent to which researchers adhere to a specific biological neural structure for an artificial neural network varies. For some, the parallelisms between nature and artificial contrivance are the primary concern, because these researchers are driven by the desire to better understand how the human brain works. Others find motivation in the ability to perform useful tasks well. For the sake of simplicity, it would be useful to ignore the biological foundations of neural networks, and address only the functionality issues. Further, as application rather than theory is emphasised in this thesis, little attention will be paid to the biological basis of artificial neural networks with the focus switched to the artificial replication of these networks.

However, the biological basis of neural networks must also be used to consider how the system operates as a whole. The interpretation of unsupervised networks from a holistic perspective encourages the interpretation of the hidden layers of a neural network in order to obtain contextual and therefore valid results. Contextual, robust models are most desirable as they underpin the credibility of such a model.

The ability of neural networks to model performance realistically via objective mathematics provides the focal point, in line with the aim of this thesis to quantify individual rugby player performance using dimension reduction techniques. The development of such mathematics provides insight into the most suitable methods for dimension reduction.

4.2.2 Brief Historical Overview

A brief historical overview of neural networks is included at this point. Most generic reviews of neural networks provide historical accounts of the researchers involved (For example: Haykin, 1999; Warner & Misra, 1996; Fausett, 1994; Cowen & Sharp, 1988). Fausett (1994, pp. 22-27) and Haykin (1999, pp. 38-44) provide an excellent overview with much of the work in this section a summary of those texts.

Beginnings

Initially driven by the quest for artificial intelligence, neural networks have existed since the Second World War following the development of the McCulloch-Pitts neuron which is generally regarded as the first neural network (McCulloch & Pitts, 1943). The combination of many simple neurons into neural systems was identified as the source of increased computational power. The weights of these neurons were set such that the neuron performed a simple logic operation such as AND, OR, AND NOT, XOR. The combination of different neurons performing different functions forms a net, capable of producing any output which is representative of the combinations of the logic functions (Fausett, 1994). The threshold is an important feature of the McCulloch-Pitts neuron still applied in many artificial neurons.

The application of the threshold is paralleled in the biological neuron, where a neuron will only fire when the threshold has been exceeded (Weiten, 1995). Consequently, the activation of the McCulloch-Pitts neuron is binary. Essentially it is an all-or-nothing process with the neuron either firing (activation of 1), or not (activation of 0). The architecture is exhibited over the page in Figure 4.1, for a McCulloch-Pitts neuron for the logical AND response, where a threshold of two is needed to illicit a 'true' response from the neuron. This indicates the key features of the McCulloch-Pitts neuron, where the neurons are connected directly by weighted paths. The connection path is excitatory if the weight is positive, inhibitory otherwise. The threshold is fixed and determined by prior analysis. Not illustrated in this example is the effect of inhibitory signals. McCulloch-Pitts neurons are designed such that if any input has an inhibitory response, denoted by a negative connection weight, the neuron will not fire. X1 and X2 denote the input units and Y is the output unit. The activation function for y is specified as follows:

$$f(y\_in) = \begin{cases} 1 & if \;\; y\_in \geq 2 \\ 0 & f \;\; y\_in < 2 \end{cases}$$

where y_in is the sum of the signals generated by the input units, as indicated below in neural network notation.

$$y\_in = \sum_{j=1}^{p} w_j x_j$$

In Figure 4.1 $w_j = 1$ for both connection weights. If both X1 and X2 are 'true', then the sum of the excitatory signals to Y is two, thereby inducing a 'true' response from the output unit Y. More specifically, to elicit a response from the neuron X1 AND X2 must be true.

This basic structure is common to most neural networks. Such a model is potentially useful in the sporting context, instigating a subjective model based on the opinion of coaches with discrete output, as demonstrated in Section 6.5. However, training (optimisation of weights and neural network architecture) would need to be imposed by a domain expert, although the structure of the model developed in Section 6.5 allows the performance of many individuals to specify the thresholds. That is the output statistics divide the population into ranked groups based on thresholds obtained from the data, such as the quartiles. Further, to adequately cover all aspects of play the model would be potentially large, unless key structures, or variables, can be identified within the data set, highlighting the need for dimension reduction techniques.

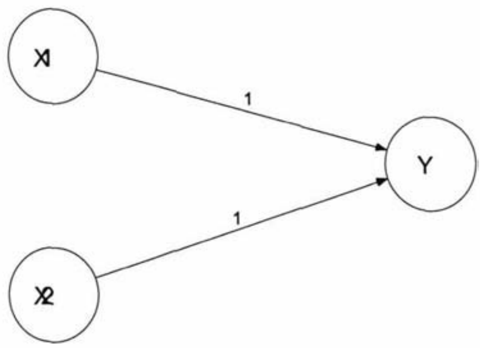


Figure 4.1: *McCulloch-Pitts Neuron for the Logical AND Function*

Learning Laws

The first learning law for artificial neural networks was designed several years later by Hebb (1949). The fundamental concept underpinning this development was that if two neurons were active simultaneously, the connection between them should be increased (Fausett, 1994). As a result learning occurs by modifying the weights between neurons when they are active simultaneously. As neurons are either active, or not, this has a binary basis as featured in the McCulloch-Pitts neuron.

Interest boomed during the next two decades with a number of researchers developing neural net based methodology, specifically a large class of networks called perceptrons (Block, 1962; Minsky & Papert, 1988 (originally published 1969); Rosenblatt, 1958, 1959, 1962). Based on the retina in the visual system, a perceptron is typically comprised of three layers of neurons; sensory units, associator units and a response unit. The response unit employs, like the networks described previously, a threshold output function. The connections between the sensory units and associator units, whilst fixed, are adjustable as a consequence of the perceptron learning rule. This learning rule uses an iterative process more powerful than the Hebb rule. Fundamentally, if an error occurs between the calculated output data and the training target data the weights are altered, otherwise no alteration occurs (Bishop, 1995). Training continues until no error occurs. Under suitable assumptions, this iterative learning procedure will converge to the correct weights, provided the correct weights exist in the defined solution space (Fausett, 1994).

A similar learning rule to that developed for the perceptrons, the least mean squares or delta rule, was created by Widrow and Hoff (1960). This learning rule is based on the minimisation of the squared error between the activation of the calculated output unit and the target training value for each training pattern or observation. These nets are often called ADALINE (ADAptive LInear NEuron). This learning rule, designed for a single layer network, is a precursor of the backpropagation rule for multilayer nets (Fausett, 1994). Due to the early success of the perceptrons, enthusiastic claims where made that were later found to be unjustified by Minsky and Papert (1969) due to the learning procedure being limited to linear separable problems (Warner & Misra, 1996). Linear separability requires that the decision boundary be of a linear nature, enabling linear separation between classes of solutions. Further, single layer perceptrons failed to solve simple problems such as the "exclusive or" function. The "exclusive or" function (XOR), for a two input unit case, requires that one input unit is true and the other is not. Additionally there existed no general method for training a multilayer net (Fausett, 1994).

This led to a reduction in active research during the 1970's. The common thread of modelling memory linked most research during this period (Anderson, 1968, 1970, 1972; Anderson, Silverstein, Ritz & Jones, 1977; Willshaw, Buneman, & Longuet-Higgins, 1969). This involved supervised neural networks in which the output for a set of input values had to be memorised. It was during this period that unsupervised, or self-organising neural networks, involving the clustering of input patterns (observations) began. These networks featured a new type of learning called competitive learning.

Competitive Learning

Competitive learning was approached for the first time by Grossberg (1976) and von der Malsberg (1973). With competitive learning a group of neurons compete to become active in an unsupervised learning procedure in which the weights of the network are adjusted after the winning nodes have been selected (Fausett, 1994). Weights for nodes close to the winning node were increased at the expense of nodes far from the winning node. This process continued until the weights were stable, and each observation was assigned to its winning node. Nodes were then colour-coded according to their popularity, giving a visual impression of clustering.

Grossberg (1983) also collaborated with Carpenter and developed a theory of self - organising neural networks labelled adaptive resonance theory (Carpenter & Grossberg, 1983). Two forms exist for clustering either binary vectors (ART1) or continuous-valued vectors (ART2). The principle of self-organisation for this technique "involves a bottom-up recognition layer and a top-down generative layer of nodes (Haykin, 1999, p. 41)". If the pattern expressed by the two layers matches, then a dynamical state (adaptive resonance) where the effect of the neural activity is amplified and prolonged is created (Haykin, 1999). This is similar to the amplification of coinciding waves with equivalent wavelengths that occurs when the crests are harmonious. As the network is competitively based, the unit with the largest net input wins and is accepted provided the ratio of the norm of the activation function and the norm of the input vector exceeds a value termed the vigilance parameter (Fausett, 1994). The input data is self-organised into categories with the variation allowed in each category specified by a user defined vigilance parameter, which controls the degree of similarity required for patterns to be assigned to the same cluster (Fausett, 1994).

Kohonen (1972) started out during this period examining associative memory neural nets. Associative memory neural nets are reliant on similarities between the presented data and stored training data. Presented information is compared to previously stored patterns, which can be matched provided that the input vector is sufficiently similar to a stored pattern. This is the basis of pattern association, a special form of the mapping problem where inputs are matched to specified outputs, although the desired output is not just "yes" or "no" (Fausett, 1994). Such a clustering process could be implemented in this thesis, although the influence of non-performance and match volatility would distort the use of this methodology as a clustering tool that identifies the patterns associated with each positional cluster. Similar to the visual cortex, where the cells communicate with one another extensively in a rich processing network (Weiten, 1995), Kohonen's (1982) self-organising feature maps (SOMs) implement a topological structure for clustering units in pattern recognition. Originally developed for images and sounds, self-organising maps can cater for data compression, dimension reduction and clustering (Berry *et al.*, 1997). Unlike feed-forward backpropagation networks where input variables are mapped to a defined number of output units, SOM's map the input patterns to a grid-like structure consisting of clusters from which only a solitary "winning" output unit is expressed for each input pattern. SOMs capable of dimension reduction, along with feed-forward neural networks are examined in greater detail later in this chapter.

Appeal of supervised neural networks began to return with the development by Hopfield (1982) of the Hopfield model for content-addressable memory. Content-addressable memory sees information stored, and addressed, based on partial patterns relating to content rather than by location of the information as in traditional computing. More importantly, the back-propagation algorithm was developed in order to solve some of the problems previously encountered when training neural networks by McClelland, Rumelhart, and the PDP Research Group (1986) (Warner & Misra 1996). Backpropagation is not restricted to linear separable problems and can be generalised as a training method for multilayered nets.

The idea of backpropagation was explored earlier (Werbos, 1974) but not fully appreciated at the time (Warner *et al.*, 1996). Backpropagation is a learning algorithm for multilayer neural networks based on the minimisation of the total squared error between the calculated output data and the training target data. A multilayer network consists of the fundamental input and output layers, with the addition of a number of hidden layers, such that there are two or more layers of weighted connection paths. When counting the layers the common practise is to not count the input layer as it does not perform any computation (Warner, *et al.*, 1996). This methodology is suitable for use with the rugby data, provided relevant outputs can be found for the input vectors.

Divergence of Theories

The divergence of different theories becomes quite apparent at this stage with different situations requiring different types of neural network methodology. Boltzmann machines, based on stochastic units were employed for tasks such as pattern completion

(Hinton & Sejnowski, 1986). Generally used for solving constrained optimisation problems, Boltzmann machines have fixed weights representative of the constraints encountered by the problem. The solution is sought by altering the activations (1 or 0) of the units based on probability distributions and the potential effect of change on the energy function or consensus function of the net (Fausett, 1994). The energy function is a function of the weights and activations of the neural network. If this can be found for an iterative process, the process will converge. It is desired that the consensus function be maximised in a constrained optimisation problem.

Time series problems have sought resolution through recurrent networks such as the Elman network (Elman, 1990), the Jordan network (Jordan, 1989) and real-time recurrent learning (Williams & Zipser, 1989). Recurrent networks have closed-loop signal paths from a unit back to it enabling the reception of feedback.

To research further into other types of neural networks and their origins is redundant here because the continual drive to quantify rugby player performance numerically governs the research path. However, the diverse nature and wide applicability of neural networks is continually emphasised as the quest for dimensionality reduction of rugby player performance continues.

4.2.3 Applications

An abundance of literature is available to describe and implement neural networks. Even though neural networks are just one of many techniques falling under the data mining banner, the applications are equally as diverse. Cheng and Titterington (1994) listed a number of interesting and varied applications dealing with pattern classification and pattern recognition. This list included; sexing of faces; identification of cancerous cells; prediction of re-entry trajectories of spacecraft; automatic recognition of hand written characters and so forth. Warner and Misra (1996) also provide an illustration of applications including: identifying underwater sonar contacts (Gorman & Sejnowski, 1988); predicting heart problems in patients (Baxt, 1991); playing backgammon (Tesauro, 1990) and again the list continues. Whilst this list is not exhaustive, and there is no intention of it being so, it highlights the willingness of practitioners to apply neural networks to problems that have previously only had conventional, linear based techniques for resolution. The nature of the specific architectures, learning algorithms and activation functions applied in relevant studies is left to the later sections (4.3-4.5).

Networks for dimension reduction, under the guise of non-linear principal component analysis have assessed diverse examples ranging from instrument calibration (Borgaard, 1995), to simulated examples (Malthouse, 1998; Pal & Eluri, 1998; Jiang, Wang, Chu & Yu, 1996). These networks are unsupervised in that the output nodes carry the same information as the input nodes, while the hidden nodes carry the same information in reduced dimensionality. Jiang *et al.* also examined the chemical components of 26 different tobacco samples, involving 16 measurements to differentiate between two types of curing methods. Ten nodes were specified in the hidden layer, adopted for simplicity and two nodes representing the types of curing method formed the bottleneck layer. This analysis showed that the neural networks outperformed traditional principal component analysis in identifying the two different methods for curing tobacco. These results were confirmed in an example assessing steroidal compounds. Stamkpoulos, Diamantras, Maglaveras and Strintzis (1998) demonstrated how dimension reduction neural networks could be used to detect defective heart rhythms on an individual level, based on data obtained from an ECG. This method emphasised the use of a four-layer bottle necked self-supervising network, which is expanded upon in more detail in section 4.5. The specifics of each network are not entered into, for, as demonstrated in the section relating to the structure of networks, literature suggests that the structure of a network is dependent on the problem (Lotlikar, 2000) and must be identified heuristically (Jaing *et al.*, 1996).

Clearly the ability to apply neural networks for dimension reduction to the rugby data is supported by the diversity of these examples. However, to confidently apply such methodology to this thesis, an understanding of the potential weaknesses must also be examined.

4.2.4 Cautions

Not all researchers portray the use of neural networks in a favourable manner. (Schwarzer, Vach, Schumacher & 2000) summarised the misuses of feed-forward artificial neural networks in prognostic and diagnostic classification in oncology.

They identified seven key problem areas which are outlined as follows:

1. mistakes in estimation of misclassification probabilities
2. fitting of implausible functions
3. incorrectly describing the complexity of a network
4. omitted information regarding network
5. use of inadequate statistical competitors
6. insufficient comparison with statistical method
7. naive application of artificial neural networks to survival data

The misuses described above can be attributed to either user error (1)(3)(4)(5)(6)(7) or system limitations (2). Classification (1) and analysis of survival data (7) are not applicable to this thesis. Whilst a clustering procedure is considered, it is in an unsupervised context, preventing the calculation of misclassification. The fitting of implausible functions (2) is an important concept as it relates to the contextuality of the model. Plausibility is required to satisfy the consumers and this issue is crucial in selecting the models discussed in Chapters Six and Seven. Additionally, details of the network architecture are provided (4) as this indicates the susceptibility to overfitting. The reference to statistical competitors (5) relates to other statistical methods, not networks with competitive layers. The comparison of statistical methods (5)(6) is featured in Chapter Seven which directly compares factor analysis, described in Chapter Three, and the dimension reduction neural networks applied in Chapter Six. The architecture of the networks employed are also described in Chapter Six, thus the caution regarding describing the complexity of networks (3) is timely. Whilst the problems identified areas for concern when applied in a medical type situation, classification of rugby players is not a literal case of 'life and death'. However, the cautions expressed are a reminder that with the application of any statistical technique, the modeller must understand the model context, the implications of modelling error and the need for model reliability and validity.

"Life and Death" does enter the modelling fray in rugby, not in terms of the output, but in terms of the context and the applicability of obtained models. If the models are incongruent with the rugby public's perception of ability and performance, and there is no contextual justification for such a difference, then the rating system will die. Therefore the structure imposed on the models is important.

Schwarzer *et al.* (2000) identify that the uncritical use of artificial neural networks can lead to serious problems such as fitting of implausible models (2). However, this can be overcome by interpreting the hidden layers, or at least identifying the variables that are exerting the most influence on the output units leading to interpretation. This issue is addressed in Chapter Five, where a method is introduced that enables the influence of variables on the output to be measured, leading to interpretation. "Neural Networks are opaque. Even knowing all the weights on all units throughout the network does not give much insight into why the network works (Berry *et al.*, 1997, p. 324)." This is because the holistic implications of the data presented in one step – from input to output – must be considered. That is the interactive effect of input variables combine to produce the resultant output, with any hidden layers viewed as assisting this process. A holistic interpretation is important to establish contextual answers. Berry *et al.* (1997) touts sensitivity analysis as a method for understanding the relative importance of each input variable and this is described further in section 4.6.

Other areas of research, such as computer science, seek also to interpret neural networks. Various rule extraction methods can be used for extracting knowledge regarding the behaviour of a network which employs categorical or non-numerical data (as opposed to continuous data), assisting users with interpretation and understanding the underlying processes (Andrews, Diedrich & Tickle, 1995; Maire, 1999; Garcez, Broda & Gabbay, 2001). In the dimension reduction problem motivating this thesis, numerical data (continuous and discrete) is presented, making sensitivity analysis more appropriate. The half-moon statistic, which is developed and explored in Chapter Five, is also only appropriate for numerical data. It compliments the graphical nature of a sensitivity analysis by giving a statistical test for the significance of an input-output relationship, while a sensitivity analysis provides only graphical support for these relationships. Additionally, the half-moon statistic can be expanded to cater for multiple input variables. Sensitivity analysis can also achieve this by considering the effect of groups of input variables to see if certain combinations have a particular importance with respect to the output variables. However, the half moon statistic considers the impact of all observations at once, not just the quartiles as commonly done in sensitivity analysis. Consequently it is more attune to the global implications of the relationship between input and output. The half-moon statistic gives an indication of not only the presence of a pattern, but the relative strength of the pattern that is imposed by an influential variable.

The ability to interpret the neural networks, or at least understand the underlying processes, creates a similar situation to that described in Chapter Three where the models constructed using factor analysis were adopted only when interpretation indicated that the model was contextual. This is due to the need for model transparency in creating a model that rugby observers will be comfortable with. The flexibility of neural networks were cited as a major concern by Schwarzer *et al.* (2000) as this leads to the fitting of implausible functions in a biological/medical setting. Quite clearly, a case for inflexibility in the model can be argued for in the application of neural networks to medical purposes. As all human bodies share the same essential characteristics, models do not need to change or evolve to match changing circumstances. However, rugby is a dynamic environment where the interaction of variables change, given any number of circumstances referred to in Chapter Three as well as the maligned law changes. Thus a degree of flexibility is required when attempting to quantify performance and infer ability, or, more importantly when attempting to quantify the adaptability of any individual to volatile match conditions. The relative inflexibility of factor analysis, due to its reliance on linear dependencies within the data set, is therefore certainly a hindrance in a modelling process.

With reference to the overview of data mining earlier in this chapter, the goal of data mining is to extract useful information from non-experimental sometimes non-traditional data (Mackinnon *et al.*, 1999). Essentially the purpose of data mining is to uncover latent information that is not easily extracted via traditional 'inflexible' functionally dependent models, such as linear factors. In a medical application it is desired that the model be known so that numerical inferences relating to the cause and possible cures can be made (e.g. ischemia detection (Stamkopoulos *et al.*, 1998)), requiring the relative inflexibility of specified models (Wyatt, 1995). With the quest to quantify rugby performance (where the goal is to produce a number that is representative of individual performance on the rugby field) we are afforded more freedom. The functions underlying the production of this rating value are important only insofar as to extract the latent combined influence of variables over an output. In a successful model the influence of these underlying functions will be congruent with the impression gained by an unbiased and non-forgetful observer. Thus it is the interpretation of the influence – which must be contextual – not the interpretation of the function that is of importance in the rugby context, whereas the medical context requires a more rigorous approach where the function must also be quantified.

The final problem discussed with relevance to this thesis is that of over-fitting. This is discussed in the opening section of this chapter, 4.1. Although Schwarzer *et al.* (2000) do not list over-fitting as one of the seven key problem areas, it is indeed a major problem. Over-fitting is the result of over dependency on the inherent structure specified through the training data. That is, the flexibility of the modelling procedure has been over-exploited. A complex model that fits the training data very well is obtained, which may not model new data well. Ethically, this is the same as adopting a family of models that are inappropriate for the data presented. A rough analogy is applying non-linear regression to data when linear regression is more suitable. Most importantly it affects the ability to generalise the results beyond the training data set. Employing an early stopping procedure or regularisation can alleviate over-fitting.

The simpler of the two methods for preventing over-fitting, early stopping requires the data set to be placed in three subsets (training, validation and testing). The training data is used to define the parameters of the network (biases and weights, which are the model coefficients). When the validation data set is presented the errors of the validation data set are monitored as this provides an indication of when over-fitting sets in. The error from the validation set will normally decrease during the early iterations as the training data becomes attuned to the training data set, by adjusting the parameters to minimise the error of the training data set. However, when over-fitting starts to occur, that is when the model starts to focus on the specificity of the training data, the error in the validation fit will start to increase. Therefore, the parameters associated with the minimum validation error rate are used as this is when the model is at its best in terms of generalisation. The test data set is used to compare different models (with different starting values and different architectures), and is not used in the training process. Due to its simplicity, early stopping can be employed with most training functions (Demuth *et al*., 1998).

Regularisation is mathematically based and provides smoother responses than early stopping, whilst also reducing the susceptibility of a network to over-fitting. This method requires the modification of a performance function which is typically the sum of squares of the network errors. Generalisation is improved by defining the performance function as the mean sum of squares of the network errors (mse) plus the mean sum of squares of the network weights and biases (msw). As outlined in Demuth et al. (1998) the regularised performance function (msereg) is given by

$$\text{msereg} = \gamma\text{mse} + (1 - \gamma)\text{msw},$$

where γ is the performance ratio, $\text{mse} = \frac{1}{N}\sum_{i=1}^{N} e_i^2$, and $\text{msw} = \frac{1}{n}\sum_{j=1}^{n} w_j^2$, with e referring to the error of the network, w associated with the weights and biases, n is the number of weights and biases in the network and N is the number of observations. This performance function causes the network to have smaller weights and biases, inducing smoother functions which are less likely to over-fit (Demuth *et al.*, 1998). However, the performance ratio value, which is difficult to optimise, is crucial in forming the network. If γ is too large, over-fitting will occur and, conversely, if the ratio is too small, then the network will fail to adequately fit the data.

Alternatively, reducing the size of the network can be used to reduce the vulnerability to over-fitting (Demuth *et al.*, 1998). More complex the functions can be created when a network becomes larger due to the relatively large number of adjustable parameters. However, if the network is just large enough to provide an adequate fit, then the problem of over-fitting is reduced. Therefore a careful balancing act is required. Early stopping is the easiest method to employ, and whilst it may not produce smooth functional approximations as with regularisation, this is not a necessary concern with the problem presented in this thesis.

These problems relate to the performance of neural nets. Another potential difficulty is the understanding of how such results were achieved. "No matter how well or how poorly a neural network performs it is extremely difficult to understand *how* it works (Ciampi *et al.*, 1997, p992)." It is the lack of interpretability of neural networks that has led to their perception as 'black boxes' (Ciampi *et al.*, 1997). The interpretability is a core component of this thesis and a remedy is explored in depth in Chapter Five. The interpretability is fundamental due to the transparency, marketing and credibility issues inherent to a system such as the Eagle Rating.

Any lack of interpretability leads to possible uncertainty regarding the solution. Medsker *et al.* (1993) argue that most neural networks cannot guarantee an optimal solution, a completely certain solution or sometimes, even repeatability for the same input data because different starting values produce different solutions. However, given that an optimal solution exists within the potential solution space, a correctly implemented neural network should close in on a good approximation of the optimal solution, provided the network does not become stuck in local minima. Obviously if solutions aren't repeated then the global solution has yet to be uncovered or may not exist within the defined solution space. This usually means that the model is too complex requiring simplification in some form. This accounts for all three qualms expressed by Medsker *et al.* (1993).

The solution uncovered in this thesis may not be a globally optimal solution, nor does it need to be. However, it must be contextual enabling the rating system to be marketed. Provided the solution found is contextual, it will satisfy the three key problem areas relating to applicability namely transparency, credibility and marketability. Firstly, contextual solutions are transparent. This is because the resultant output parallels the thinking of rugby observers. Thus the rating system will match the output of an unbiased, non-forgetful expert. There are limitations obviously, due to the scope of the data recorded, such as the exclusion of positional play (which is defined by spatial attributes). However, by providing the best possible statistical solution that matches human thought processes, we are providing the consumer with an unbiased perspective. This leads into the credibility issue. Provided the rating system adopted is contextual and explained to the rugby public in rugby terminology, it will be understood. If this understanding is congruent with the common belief of how different events should be rated, the system will be accepted and with acceptance comes credibility. A credible product is a marketable product. If the rating system is contextual and transparent, it is easier to explain and sell to rugby observers. However, marketing also plays an important part in generating understanding and credibility for the rating system.

A major issue in this creative process relates to the training times for neural networks. Training times are often cited as a problem in older literature but no longer present a problem due to the continued development of modern computing power and techniques that have cut training times and expanded computational limits (Zhang, 1999). The nature and limitations of the data used in this thesis are such that training times are quite short, typically less than five minutes using SAS Enterprise Miner on a desktop computer with a Pentium II processor and 96 Megabytes of RAM. Armed with this computing power and aware of the potential pitfalls of neural networks (over-fitting, interpretability, and contextuality), an exploration of the relevant neural networks follows in the next section.

4.3 Supervised Neural Networks

The broad properties and origins of artificial neural networks have been covered briefly in the preceding sections. The following sections further examine the techniques applicable to this thesis. Firstly, the generic structure of artificial neural networks are inspected before the self-supervised networks necessary to extract meaningful measures of performance are discussed in Sections 4.4 and 4.5.

The dimension reduction problem in this thesis is not typical. Generally, supervised networks are of two main types, prediction or classification. Prediction provides fitted or forecasted information about events in the future, while classification assigns observations to one of two or more categories (Haykin, 1999). Typically, prediction feed-forward neural networks only have one output node, the forecasted variable; while classification involves classifying an input pattern to a certain category (Fausett, 1994) with one output node assigned to each category and the predicted values from a classification network are associated with the probabilities of category membership. The target values are one for the correct category/node and zero for the incorrect category/node in the case of a classification neural network. For the dimension reduction problem, there is one node for each dimension in the feed-forward neural network. However, the value of the nodes for each dimension corresponds to the value of that input pattern for that particular dimension. Importantly, the broad learning techniques for these networks share the same principles.

4.3.1 Types of Learning

Neural networks have two broad types of learning, supervised and unsupervised, that govern the properties of the network. Supervised learning has a target value associated with each input pattern while unsupervised learning does not. In the case of unsupervised learning the network must learn to group or cluster input patterns based on similarities (Warner *et al.*, 1996). A self-supervised network can be considered as a special type of supervised network which achieves dimension reduction by using the input patterns in the output layer (Kramer, 1991). This seems paradoxical as there are no output data to act as target data, however, the input patterns act as the targets. This works quite well in the dimension reduction setting as explained in Section 4.4.

In supervised learning, training of the network is achieved by presenting a series of input patterns, each with an associated target pattern and adjusting the weights according to a learning algorithm in order to minimise the discrepancy between the observed and fitted values (Fausett, 1994, p. 15). When correct input-output examples are abundant, backpropagation is a suitable method for supervised learning (Hecht-Nielsen, 1988). This method will be described shortly.

With an unsupervised learning scheme, there is no matching target for each input pattern during learning. That is unsupervised learning requires the weights of a neural net to be modified without specifying the desired output for an input pattern, although a self-supervised network can be created (Fausett, 1994). This is exactly the case presented in this thesis. There is no notion of what rugby player performance is, or how to quantify such a value. Unfortunately, many spectators will lay claim to what constitutes a good performance based on their perception of play, shaped by their own learning experiences. However, such subjective perceptions may have no statistical basis, or even contextual basis. This is the challenge pitted against the models developed in this thesis, and highlights the need for contextual resolution of the network employed.

Unsupervised learning, like that performed in SOM, is generally used in self-organising neural networks for clustering data, extracting principal components, or curve fitting (Fausett, 1994), based on competitive learning principles (Haykin, 1999). The network forms internal representations (classifications) of the input patterns based on the features (statistical regularities) present (Haykin, 1999). Application studies have shown that the SOM algorithm is capable of performing dimension reduction when the data has a high dimensionality, but a smaller intrinsic dimensionality (Cherkassky *et al.*, 1998, p.196). These data properties were demonstrated in Chapter Three.

The most popular training algorithm, backpropagation is unsuitable in an unsupervised learning scheme due to the lack of specified output. However, a self-supervising network can be established enabling the backpropagation algorithm to be used, allowing 'off-the-shelf' software to be implemented. A self-supervising network creates a supervised network by using the input data as the target when no output data exists, as demonstrated in Section 4.4, enabling the principal components to be extracted.

4.3.2 Supervised Neural Networks Structure

Before assessing unsupervised neural networks, it is important that the simpler case presented by supervised networks is reviewed. As will be demonstrated in the section associated with unsupervised networks, the components of a supervised network structure forms the building blocks associated with the self-supervising network that is to be adopted. Therefore, it is useful to assess the properties of supervised networks.

"Neural nets are basically mathematical models of information processing (Fausett, 1994, p. 2)." In this thesis the ultimate goal is to convert the on-field activities of an individual rugby player to a single numerical value, or at the very least an easily manageable finite vector that relates to performance in key contextual areas that can be used to infer ability. The first step in this process is the conversion of match play to data, via notational analysis. It is then required that this information be processed to produce a reliable contextual numerical indicator of performance. In Chapter Three traditional multivariate techniques, cluster analysis and factor analysis, were used to process the information generated from a rugby match. This section introduces the underlying structure that allows neural networks to enter the stage as mathematical models for information processing.

A neural network is composed of processing elements, or neurons, that can be organised in different ways, or architectures (Medsker *et al*., 1993). Variations on the processing elements and architecture characterise each network. The underlying structure of a neural network is characterised by three key concepts, described by Fausett (1994) as follows.

1) *Architecture*; the pattern of connections between the neurons.
2) *Training* (or Learning Algorithm); the method of determining the weights on the connections.
3) *Activation Function*; this is applied to the weighted sum of the inputs of a neuron to translate the inputs into an output.

A section dedicated to each of these key areas follows shortly. Firstly, the building blocks of a neural network are introduced.

Each neuron receives a defined finite number of inputs, which are processed before an output is delivered. The processing of the input data involves the weighted summation of all inputs received before the data is transformed via an activation function. A network is formed by a collection of neurons, with the arrangement defined by the architecture. A basic structure is shown in Figure 4.2. The neurons can be connected in a number of ways dependent again on the architecture and the type of network. Each connection is weighted. It is the adjustment of these weights that provides the learning mechanisms and adaptation of a neural network. The learning mechanisms adopted serve to further characterise the network.

Figure 4.2 depicts a simple neuron. Information is received as weighted input into the neuron where it is summed, transformed via an activation function then forwarded on as an output. The description featured above is widely available in the literature.

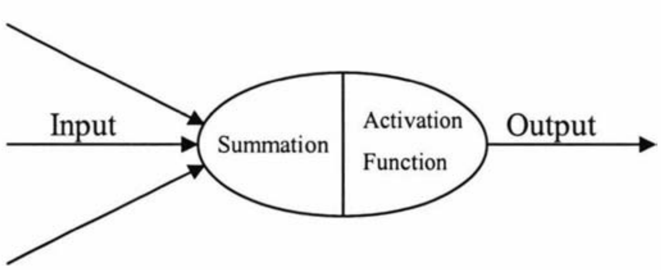


Figure 4.2: *A Simple Neuron*

Most literature reviewing neural nets includes an introductory overview of the schematic representations of a network and underlying elements before launching into the relevant debate (Crick, 1989; Trippi *et al.*, 1993, Hecht-Nielsen, 1988; Lowe, 1997; Hinton, 1992; Fausett, 1994; Ciampi *et al.*, 1997; Cheng *et al.*, 1994; Warner *et al.*, 1996; Schwarzer *et al.*, 2000). The usual structure described by these authors, and most commonly applied, is a two layer feed-forward network, with a varied number of inputs, hidden nodes and outputs. A schematic representation of a two layered neural network, typically used for prediction, with n input nodes, p hidden nodes and a single (k=1) output node (y_k) is featured in Figure 4.3.

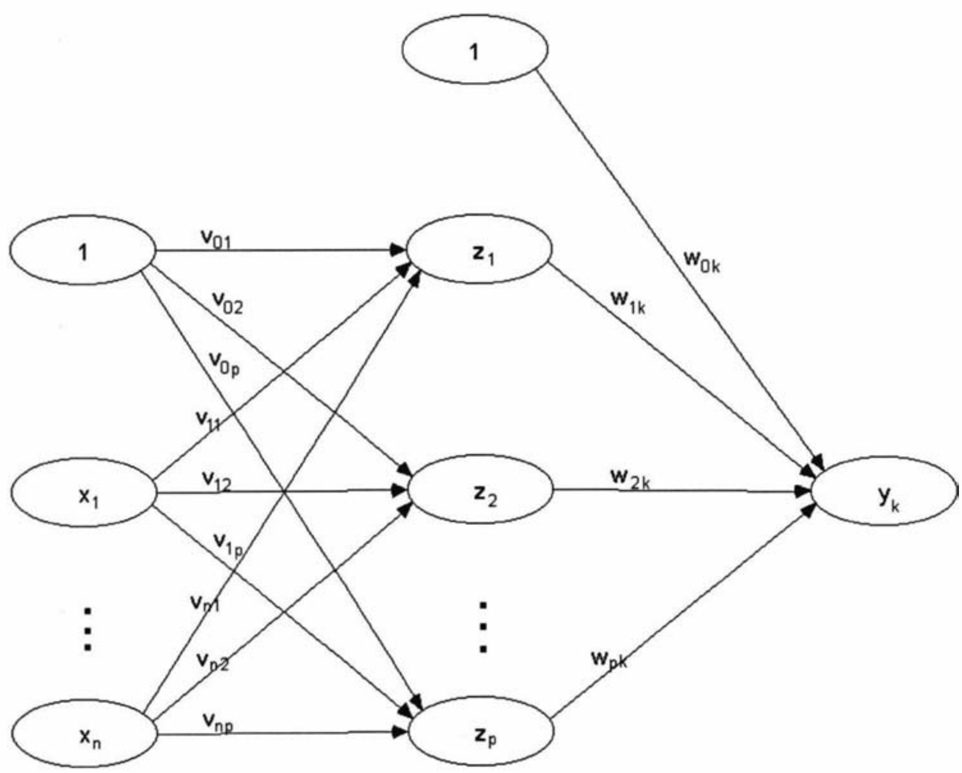


Figure 4.3: *Multilayer Neural Network for Output k*

The nodes in the layer featured by x_i comprise the input layer. These nodes are connected with the nodes in the next layer, the hidden layer, represented by z_j. Finally, the hidden layer is mapped to the kth node in last (output) layer (y_k). Note that nodes are connected only with the immediately preceding layer. There exist no connections within layers in a feed-forward network. Each connection bears a weight v_i and w_j that maps a specific link between nodes. For example the node x_1 is mapped to the node z_2 via the weighted connection path v_{12}.

The outputs for the hidden nodes are defined as

$$z_2 = f_a\left(v_{02} + \sum_{i=1}^{p} v_{i2}x_i\right)$$

where f_a is the activation function, typically some sort of sigmoidal or hyperbolic tangent function. Obviously z is a combination of the values $x=(x_1,\ldots, x_p)$. If an identity activation function was employed in the hidden layer, then this equation is the same as a multiple linear regression equation, where the weights v_j are simply the regression coefficients (Cheng *et al.*, 1994). Finally

$$y_k = f_b(w_{0k} + \sum_j z_j w_{jk})$$

where f_b is the activation function. In the output layer the activation function tends to be linear as this enables arbitrarily smooth functions to be obtained that provide coverage of an extended solution space, as outlined in section 4.4. The sigmoidal activation functions act as data filters by limiting the permissible amplitude range of the output signal to some finite value (Haykin, 1999, p.10). This means that extreme values are mapped to some asymptote in the case of the s-shaped sigmoidal functions, while linear activation functions would map these values towards infinity.

Firstly, we will look at different architectures used for the respective classes of problems. This will give an insight into how to structure the architecture for the main problem encountered in this thesis, namely dimension reduction.

Cybenko and Funahashi have shown that any continuous function can be represented to any degree of accuracy by a neural network with one hidden layer and sigmoid transfer functions (Lowe, 1997). Cybenko (1989) has also shown that any discontinuous function can be represented to any degree of accuracy by a neural network with at most two hidden layers (Lowe, 1997). This implies that neural networks are capable of representing almost any function (Lowe, 1997). However, Ciampi and Lechevallier (1997) are more cautious, "The neural network approach finds its justification in a central theorem that, under certain 'reasonable' conditions (described in the next section) *any* function can be approximated by a feed-forward neural network with one hidden layer (Hecht-Nielsen, 1990; White, 1989, 1992) (Ciampi *et al.*, p 992, 1997)." Feed-forward networks are comprised of units that are only connected to the units in the next layer (Warner *et al.*, 1996). It is this flexibility, and the removal from any prespecified functional form that make neural networks so powerful, especially in a situation like quantifying rugby player performance, where not only is the functional form unknown, but there is no specified output. This endears neural networks to this study.

4.3.3 Neural Network Architecture

The architecture of a neural network refers to the arrangement of neurons, or nodes, and the pattern of connection links between nodes (Fausett, 1994). The choice of activation function also contributes to the definition of the architecture of a neural network (Ciampi *et al*., 1997). In this section only multilayer feed-forward networks are examined due to the basic premise, mentioned previously, that under certain 'reasonable conditions' any function can be approximated by a feed-forward network with one hidden layer (Ciampi *et al*., 1997). The reasonable conditions require that the number of hidden nodes can be arbitrarily large (Stamkopoulos, Diamantaras, Maglaveras & Strintzis, 1998). Demuth and Beale (1998, p. 5-2) are more specific stating that "networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities." The most common structure for a neural network comprises an input layer, one or more hidden layers and an output layer (Lowe, 1997).

The following sections deal with training algorithms and activation functions, leaving this section to look at the architecture of the networks, that is the number of nodes in the input, hidden and output layers. As mentioned previously, any continuous function can be approximated by a two-layered feed-forward neural network which has only one hidden layer. This section addresses the problem of how many nodes are required in each layer. This problem also arises in the case of unsupervised and self-supervised neural networks. In order to prevent over-parameterisation (over-fitting) the number of nodes should be carefully controlled.

This introduces an important step in the modelling process, pruning of nodes and input variables. There are two potential methods, pruning either before or after the network is trained. Due to the relatively high number of variables in this study, conceivably exceeding 130, pruning will occur prior to any analysis. This will have a number of beneficial effects. Firstly, the number of connections is reduced, lessening the time required to train the network because fewer parameters need to be estimated. Secondly, a network of minimum size is less likely to learn the noise present in the training data (Haykin, 1999, p. 218). Expert opinion is used to remove superfluous variables presenting the attributes that are perceived as important based on prior experiences in a selection and ranking context. This has important implications when marketing the model based on contextual attributes, making it easier to explain.

An equally important consideration is the number of nodes in the hidden layer. In the case of a supervised classification network the hidden layer requires more neurons than the number of classes in order to learn correctly (Tagliaferri, Capuano & Gargiulo, 1999). This is because some neurons represent the same class and some neurons are positioned between classes without representing any specific one (Tagliaferri, 1999). Increasing the size of the hidden layer increases the power of the network to classify correctly. However, this exposes the network to the risk of over-fitting (Berry *et al.*, 1997). Therefore, a careful balance is required, maximising the potential power of the network through the flexibility afforded by the hidden layer, whilst minimising over-fitting and, hence, reducing the number of weights and the required computing resources.

The optimal number of nodes adopted in a hidden layer can be difficult to specify (Murata, Yoshizawa & Amari, 1994). In order to combat this difficulty, Murata *et al.*, introduced the Network Information Criterion (NIC) to identify "unfaithful", or unrealisable models, which is a generalisation of Akaike's information criterion (AIC). Essentially, the NIC specifies whether or not more neurons should be added to a network by examining the relationship between training error and validation error with respect to the number of training examples and complexity of the network in terms of the number of hidden nodes. Consequently, the 'best' model should be selected by minimising either the MSE, AIC or NIC.

Instead of mechanising the process, Jiang, Wang, Xu, and Yu (1996) recommend adapting the number of hidden nodes to suit the particular problem under consideration. Generally, the number of hidden neurons is determined from a few preliminary trials (Lotlikar, 2000). For example, Cooper (1999) implemented a supervised feed forward network with four nodes in the input layer, a hidden layer and one output node. A range of nodes in the hidden layer, between four and twelve, were considered before settling on eight as the optimal model. Thus some degree of uncertainty and flexibility is afforded the modelling procedure in relation to the number of hidden nodes. An expert must make this choice with a suitable trade off between model context, generalisability and reduced prediction error.

Berry *et al.* (1997) describe heuristic rules for determining the number of nodes in the hidden layer. Firstly, the hidden layer should never be more than twice as large as the input layer. An ideal starting point is to make the hidden layer the same size as the input layer or one more than this, as suggested by Borggard (1995). If the network over-trains (learns the training data too well), then the number of the units should be reduced. If it is not sufficiently accurate the size of the hidden layer should be increased.

Another important consideration is the sample size. For each weight requiring training in the data set, 5-10 observations are required (Berry *et al.*, 1997). This aspect is most likely to cause problems in this thesis due to limited sampling opportunities. With so many raw variables, an unfeasibly large data set is required to adequately train the network. Therefore the raw variables need to be trimmed in a manner that preserves the essence with minimal loss of information.

Schwarzer *et al.* (2000) list a variety of feed forward networks applied in a medical context. Of the 50 architectures described, no apparent rules or structure are apparent, confirming that the network architecture is dependent upon the structure suggested by the data and imposed by the domain expert.

4.3.4 Training Algorithm for Neural Networks

Learning or training is the process by which the values of the weights that connect the intermediate layers are found. The training of a network is generally an exercise in numerical optimisation, where some error function is minimised (Sarle, 2001). The backpropagation algorithm is the most popular method for training supervised (and self-supervised) neural networks (Ciampi *et al.*, 1997; Warner *et al.*, 1996). However, there are many reasonable alternatives for determining the weights from a data set based on general optimisation algorithms such as Quasi-Newton, conjugate gradient (Ciampi *et al*., 1997), Levenberg-Marquadt (Demuth *et al*., 1998) and trust-region (Sarle, 2001). Additionally there are many variations of the backpropagation algorithm, such as resilient backpropagation, that are designed to prevent an overloading of computing resources (Demuth *et al*., 1998). However, for the sake of simplicity and ease of implementation, a batch backpropagation will be employed, due to its popularity and presence in a variety of common statistical packages such as SAS, Matlab, and Clementine. The entire data set is processed before the weights are updated, making this a batch process. Batch-type algorithms are most appropriate in this thesis due to the potential diversity of rugby data patterns caused by the variety of match situations encountered. Non-batch process (i.e. sub-samples) cannot reflect the entire range of feasible data.

An alternative to backpropagation is numerical differentiation where the derivatives inherent to the backpropagation algorithm are approximated. However, when a network is implemented using computer software it is recommended that backpropagation be used rather than numerical differentiation (Bishop, 1995). However, this requires an increase in computing resources.

Backpropagation, a gradient descent algorithm, is a generalisation of the Widrow-Hoff learning rule to multilayer networks and non-linear differentiable transfer functions (Demuth *et al.* 1998). As described by Fausett (1994, p. 295), the error information term (δ_k) is calculated for each output unit (y_k), in terms of the corresponding target pattern (t_k) and the hidden node outputs (z_j) using the following formula

$$\delta_k = (t_k - y_k) f_b'\left(w_{0k} + \sum_j^p z_j w_{jk} \right),$$

in which f_b' is the first derivative of the output activation function with respect to its argument. This term provides the necessary information to calculate the weight correction term (Δw_{jk}) using a learning rate (α) such that

$$\Delta w_{jk} = \alpha \delta_k z_j \text{ (or } \Delta w_{0k} = \alpha \delta_k \text{ for the bias)}$$

which is used to update the weights w_{jk}. The error information term (η_j) for each hidden node output (z_j) is calculated in terms of the corresponding inputs (x_j) by multiplying the error information term (δ_k) from the above (second) layer by the first derivative of the activation function (in the first layer) as follows

$$\eta_j = \sum_{k=1}^{m} \delta_k w_{jk} f_a'\left(v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \right).$$

From this the weight correction term (Δv_{ij}) between the input and hidden layers is computed

$$\Delta v_{ij} = \alpha \eta_j x_i \text{ (or } v_{0j} = \alpha \eta_j \text{ for the bias)}$$

providing the means to update the weights and biases, for the output layer

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk},$$

and for the hidden layer

$$v_{ij}(\text{new}) = v_{ij}(\text{old}) \ \Delta v_{ij}.$$

This process is repeated for either a specified number of training cycles (epochs) or until some minimal error threshold is achieved. Essentially, the network calculates output values upon presentation of a training pattern using the existing weights. The calculated output is compared to the target output, with the subsequent difference between the expected and observed result denoting the error. The error then is fed back through the network and the weights are adjusted to reduce the resultant error (Berry *et al*., 1997).

Backpropagation is most useful when the relationship between input and output is non-linear and training data is abundant (Hinton, 1992). An objection to the use of backpropagation as a model of learning, in a biological sense, is that in most instances people learn without the aid of a teacher (Hinton, 1992). This is based on learning through experiences and understanding the relative consequences of those experiences. However, as mentioned earlier, our interest is purely with the attainment of suitable models, not biological plausibility. Consequently, our focus is with backpropagation as a self-learning mechanism.

As mentioned previously in Section 4.2.4, a potential problem with neural networks is over-fitting. To ensure the robustness of the models adopted in this thesis, early stopping will be employed.

4.3.5 Activation Functions

The activation function, sometimes referred to as the transfer function, is applied to the weighted sum produced from the immediately preceding layer to translate that input into an output. Activation functions are bounded, monotonic and differentiable everywhere (Lowe, 1997). In order to achieve the advantages of multilayer nets, notably flexibility and expanded coverage of the solution space, a non-linear activation function is required (Fausett, 1994). A range of transfer functions are available, with the appropriate selection dependent on the bounds to be placed on the data with respect to the problem being solved. The most simple is the identity or linear function.

Identity (Linear)

$f(x) = x$ Continuous.

In a two-layer non-linear feed-forward network, a sigmoid function is usually adopted as the activation function in the first layer, with a linear activation function in the second layer as this allows the output neurons to take on any value (Demuth, 1998).

A threshold activation function is able to convert continuous input data to binary data. This is particularly useful in classification problems.

Threshold (Binary Step, Hard Limit, Heaviside)

$$f(x) = \begin{cases} 1 & if\ x \geq \vartheta \\ 0 & if\ x < \vartheta \end{cases}$$ where ϑ is the threshold.

Sigmoid functions, or s-shaped curves, are useful as activation functions, especially in nets trained by backpropagation. This is due to the simple form of the derivative causing a reduction in the computational burden during training (Fausett, 1994). The logistic sigmoid creates output that is bounded below by 0 and above by 1. This is useful where the desired output values are either binary or in the interval between 0 and 1, such as a probabilistic function. This function approximates the threshold function in that there is a tendency for it to produce values close to zero or one.

Logistic (Binary Sigmoid)

$$f(x) = \frac{1}{1 + e^{-x}}$$ Bounds [0, 1], Continuous

Borggaard (1995) identified a sigmoid function as the most suitable function for use with dimension reduction-capable networks. The binary sigmoid function can be modified to cover any range desired, centred about any point and to have any desired slope (Fausett, 1994). A commonly used activation function that has a wider range is the sigmoid function, defined as follows:

Sigmoid Function (Bipolar Sigmoid)

$$f(x) = \frac{\left(1 - e^{-x}\right)}{\left(1 + e^{-x}\right)}$$ Bounds [-1, 1], Continuous

The hyperbolic tangent function is closely related to the bipolar sigmoid

Hyperbolic Tangent

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$ Bounds [-1, 1], Continuous

For the context presented by this thesis, the most appropriate activation functions will depend on the layer, but they must produce continuous output. The effect of the bounded range is negated by the presence of a linear activation function in the middle layer allowing almost any value to be adopted by the output units. This is desired because of the handling of performance data. Consequently poor performance must relate to a low output. Thus if a player's contribution is left wanting or they are penalised or cost their team possession, their match rating should be reduced. Allowing for negative values ensures that poor performances have a negative impact on the resultant rating value. Importantly, sigmoid activation functions force outlying values to tend towards a permissible asymptotic bound, such as 1,-1 or 0, depending on the function employed. This reduces the effect of extended match play that proved problematic with the Eagle Rating calculated using the linear-based factor analysis.

A vast range of activation (transfer) functions can be employed including the following: Competitive, Hard limit, Symmetric hard limit, Log sigmoid, Positive linear, Linear Radial basis, Saturating linear, Symmetric saturating linear, Soft max, Hyperbolic tangent sigmoid, Triangular basis, Arctangent, Non-saturating, Log, Gaussian (Demuth *et al.*, 1998; SAS, 2000). The details of each of these activation functions is inconsequential, as sigmoid functions are the most appropriate for the dimension reduction problem specified by this thesis, as supported by Borggard (1995).

The broad ideology behind the neural network techniques relating to dimension reduction which is to be applied in this thesis follows in the next two sections. The philosophy of quantifying performance remains the primary concern as outlined in Chapter One.

4.4 Self-Supervised Neural Networks in Dimension Reduction

A popular form of a self-supervised network involves mapping the inputs to inputs. This type of network can be used as a dimension reduction scheme, in which the inputs also act as the outputs (or targets), and the network flows through a bottleneck as will be demonstrated shortly. The bottleneck summarises the input information in fewer dimensions. The nodes in this layer represent the non-linear principal components for the input data. The literature illustrates the use of self-supervising feed forward networks as a method for non-linear principal component analysis, which produces the desired effect of dimension reduction (Haykin, 1999; Stamkpoulos, Diamantras, Maglaveras & Strintzis, 1998; Berry & Linoff, 1997; Jiang, Wang, Chu & Yu, 1996; Kramer, 1991). Due to the overlap in generic structure and methodology, many of the properties of self-supervised feed-forward neural networks were described in section 4.3, and are not repeated in this section.

Self-supervised learning implements a network with a bottleneck where the network is trained to produce at output the same pattern presented at the input stage (Almeida, 1993). This process is capable of extracting non-linear features (principal components) from the input data, if such features exist. During the training process, the network is required to jointly optimise the features and the reconstruction of the input data. Often for this process, multilayer feed-forward networks trained with backpropagation are employed, utilising the quadratic cost functional for optimisation ($C_{(x,\hat{x})}$) (Almeida, 1993), detailed as follows, $C_{(x,\hat{x})} = \| x - \hat{x} \|^2$, where x is the observed input variable and $\hat{x}$ is the expected input variable. The solution found by self-supervised networks is not unique as there are an infinite number of different combinations of the first principal component (Almeida, 1993). Consequently, if a self-supervising network was adopted, the interpretation of the hidden layers would be of the utmost importance, due to the potential range of solutions.

The following diagram is an adaptation of Almeida's schematic representation of bottle self-supervising network which is also known as Kramer's non-linear principal components analysis (Malthouse, 1998). Similar to the neural network diagram illustrated in Figure 4.4, each layer is clearly identifiable. Following training the dashed component, which is where the attempted replication of the original data occurs, can be discarded (Almeida, 1993).

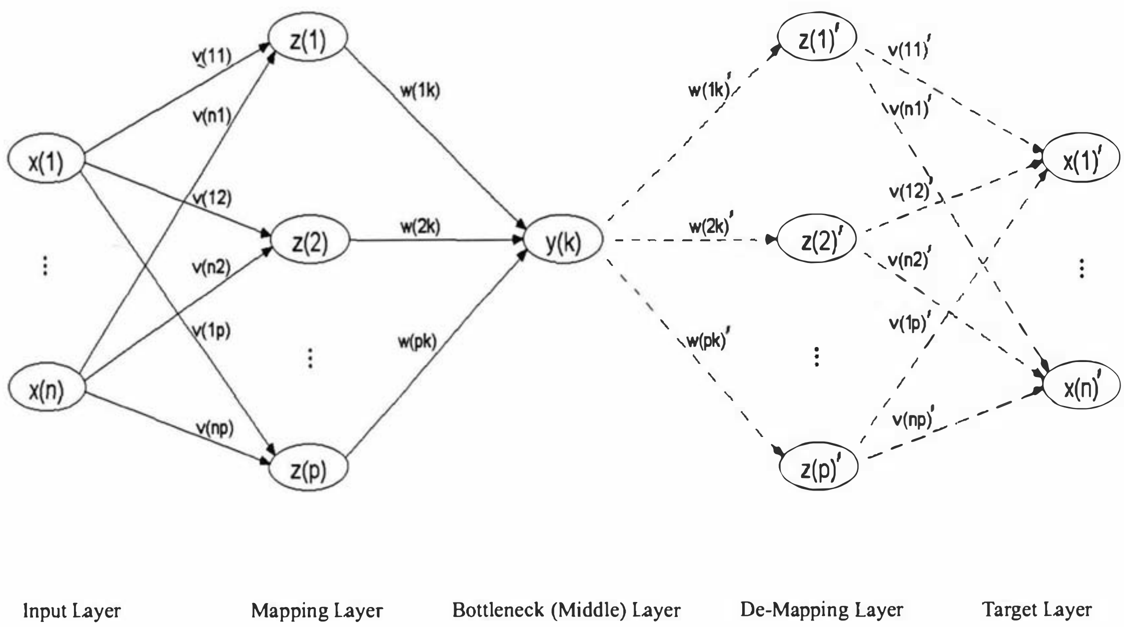


Figure 4.4: *Self-Supervised Learning*

Similar to Figure 4.3 the nodes in the layer featured by x_i (which in this figure is represented as x(i)) comprise the input layer. These nodes are connected with the nodes in the next layer, the hidden layer (mapping layer), represented by z_j. The hidden layer is mapped through the bottleneck (middle) layer via what ultimately becomes the output layer (y_k), which has k nodes which, hypothetically, will be key performance indicators (attack, defence and so forth). The network then attempts to replicate the initial input layer x_i' (target layer). This is achieved by mapping the input layer through a hidden layer z_j' (de-mapping layer). Note that the identical structure has been imposed on the replication pathway and that nodes are connected only with the immediately proceeding layer. There exist no connections within layers in a feed-forward network. Each connection bears a weight v_i and w_j that maps a specific link between nodes. For example the node x_1 is mapped to the node z_2 via the weighted connection path v_{12}.

Biases have been omitted for convenience, but would appear as with Figure 4.3, in the input layer, mapping layer, bottleneck (middle) layer, and de-mapping layer.

"Almost all unsupervised learning procedures can be viewed as methods of minimising the sum of two terms, a code cost and a reconstruction cost (Hinton, 1992, p108A)." In the case of a self-supervised feed-forward backpropagation network, the code cost is associated with the reduction of the variables in the original data to the number of variables specified by the bottleneck and is the number of bits required to explain the activities of the hidden units (Hinton, 1992). Conversely, the reconstruction cost is associated with the conversion from the bottle neck data back to the original data, $C_{(x,\hat{x})}$, described previously.

The following quotation refers to a five layered network, with terminology different to that used in this thesis, as the input layer is counted as the first layer. Sigmoid activation functions are found in layers 2 and 4, with linear functions located in layers 3 and 5. "The layers in 2 and 4 must have non-linear activation functions so that layers 1, 2, and 3 and layers 3, 4, and 5 can represent arbitrarily smooth functions (Malthouse, 1998, p168.)." That is non-linear (sigmoid) activation functions are found in the input-mapping layer and the bottleneck-demapping layer to enable non-linear functions to be replicated.

The availability of "off-the-shelf" software, flexibility and subsequent implementation are the major factor in the decision to adopt a self-supervising network.

4.5 Self-Organising Maps in Dimension Reduction

The self-organising map is a special case of an unsupervised neural network (Berry & Linoff, 1997). Self-organising maps are based on the same underlying components as feed-forward, backpropagation networks (Haykin, 1999). However, the input data is projected directly onto an ordered topological output grid that represents the data using simple geometric relationships (Kohonen, 2001). That is self-organising maps do not have a hidden layer.

4.5.1 Architecture

The architecture of self organising maps is best illustrated graphically. Earlier in this chapter the basic architecture of a neural network was displayed (page 140). The same format is used in Figure 4.5 to display the basic structure of a self-organising map (SOM) which is a replication of the image produced by Berry and Linoff (1997). Only four of the 12 weights connecting the input layer to the output layer are labelled.

Unlike the neural network in Figure 4.3, the weights ($w_{i,j}$) (which are represented as w(i,j) in figure 4.5) relate to the position of the output node (y_j) rather than contributing numerically to a combination of links (a node can be though of as a variable). This means that during training the position of the output nodes change as the network learns where the cluster centres are located.

The SOM assigns the input pattern for each observation to one output node in the topological grid. The input pattern is the collection of corresponding input units from each input variable for each observation (x_{1b}, x_{2b} for the bth observation). The input units are the individual observations that collectively form each input variable (X_1 and X_2) in the input layer. Inputs connect directly with the output layer, where only the "winning" unit is expressed. In a SOM the winning unit is the output unit which most closely matches the input pattern. This decision is based on the minimum of the distance metric used (typically the square of the Euclidean distance) (Fausett, 1994). Consequently, each input pattern is mapped to one output unit (y_{jb}). Output nodes are defined using the mean values of all their assigned observations. The aim of the learning process is to map very similar input patterns to the same output node.

Each node (input variable) in the input layer receives information from one input unit at a time. There is an independent weight associated with each connection from the input nodes to each output node. The output layer is a grid, where each node is connected to all preceding input nodes, but not to other output nodes. This grid-like structure is important for training the SOM. The order in which the input patterns are presented has no effect on the final form of the feature map in a batch SOM algorithm (Haykin, 1999). Further there is no need for the value of the learning rate (η) to change. The variety of conditions presented in rugby require a batch algorithm to model performance.

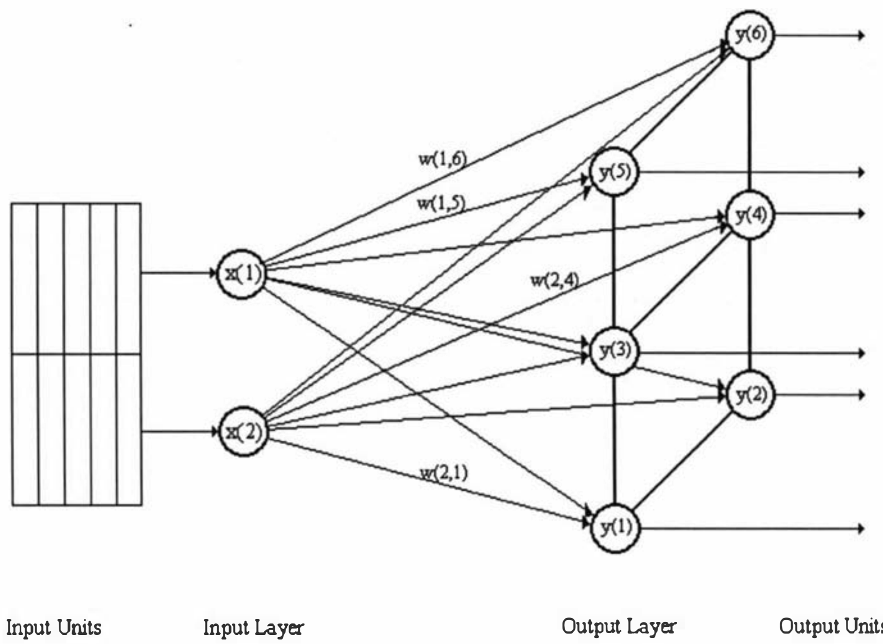


Figure 4.5: *Self Organising Map Architecture for a 2-D SOM*

Each pattern is assigned to an output node. The weights of this node and neighbouring weights are modified using the equation:

$$w_{i,j(new)} = w_{i,j} + \eta(x_i - w_{i,j})$$

where η represents the learning rate. This means the weights leading to the "winning" unit are then adjusted to strengthen the response to a particular input pattern. The weights (or codebook vectors) are adjusted such that similar input patterns correspond to the same output node. Due to neighbourliness, not only are the weights for the victorious unit adjusted, but the units in the immediate vicinity, or neighbourhood, are also adjusted so that there is more similarity between near nodes than distant nodes. This feature is convenient for creating a rating system for individual rugby player performance as nodes will be "ranked" in the output map. The size of the neighbourhood and the strength of adjustment are controlled by the neighbourliness parameter. A hypothetical one-dimensional SOM (1×10) is provided in Figure 4.6. This is appropriate for quantifying performance in this thesis because of the similarities it shares with a likert scale for ranking performances. Further, the intent of the thesis is to extract a single value for each player from each match. It would be convenient if such a value could be extracted in one step. The cluster centres (denoted by y(i)) are expected to represent points on an ordinal performance scale. This is due to the neighbourliness property of SOM's mentioned previously.

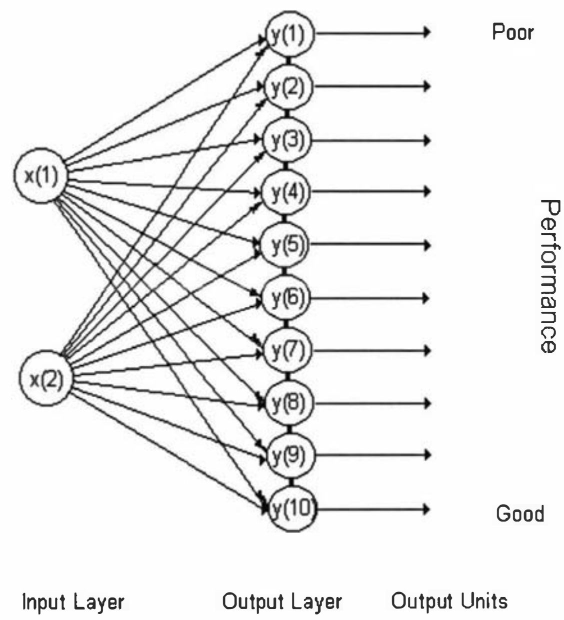


Figure 4.6: *Output Layer for a 1×10 Self Organising Map*

The general two-dimensional grid-like structure of the SOM is the product of earlier research into the recognition of images, composed of a two-dimensional array of pixel values (Berry & Linoff, 1997). Any structure can be imposed on the output layer, with neighbourhoods ranging from polygonal structures to three-dimensional arrays (Berry & Linoff, 1997). However if the data is not intrinsically multidimensional, the imposition of the topographic structure may lead to the sub-optimal placement of input patterns in the output layer (Bishop, 1995). In this thesis we use a one-dimensional SOM in an attempt to produce a single rating straight from the physical task data and a two dimensional SOM in order to separate the rating into two components (attack and defence).

Application studies have shown that the SOM algorithm is capable of performing dimension reduction when the data has a high dimensionality, but a smaller intrinsic dimensionality (Cherkassky *et al.*, 1998, p.196). However, the SOM may not perform well with high-dimensional maps due to the increase of the estimation error as the dimensionality increases, given a fixed sample size (Cherkassky *et al.*, 1998, p.196). Consequently, the number of dimensions in the topographical map should match the true number of latent dimensions to produce optimal results.

4.5.2 SOM Mechanical Properties

The number of dimensions used for the SOM map and the number of nodes for each dimension are obviously crucial. If map size is too small, the non-linearity present in the data will not be accurately reflected (SAS, 2000). Additionally, if the map size is too large, the analysis will become computational intensive leading to increased demand on computing resources. Also, a SOM typically identifies fewer clusters than it has output nodes (Berry and Linoff, 1997). The empty nodes (no assigned input patterns) may produce misleading results (SAS, 2000). For instance this may result in imperfect ordinal scales (e.g. reversed rankings for clusters in Figure 4.6). But it is required that the number of input variables are greater than the dimensionality of the topographic map in order to avoid over-fitting.

The optimum dimension, the size of the map and final neighbourhood size for the modification of weights is generally found via trial and error (SAS, 2000). This iterative process must be considered as part of the modelling process. However, if the SOM is not very large (approximately a few hundred input patterns) then the selection of the process parameters may not be crucial (Kohonen, 2001).

The fundamental difference between the self-supervising feed forward network and the self-organising map lies in the output. A self-supervising feed-forward neural network uses one node to produce a single interval performance measure. The SOM creates a map in which each dimension relates to a KPI. Each node will correspond to a different level of performance. Due to the neighbourliness feature of SOM it is expected that performance in each KPI will be ranked accordingly in each dimension. This is demonstrated in one-dimension for the hypothetical mapping of performance in Figure 4.6 shown previously.

Adjusting the weights of neighbouring output nodes requires some form of smoothing for two main reasons. The first reason is to ensure that nodes on the edge of the topological map have the same degree of variability of input patterns as the nodes from the centre of the map. This means that the obtained rating system is relatively consistent. Secondly, it is desired that the results of the SOM represent the latent structure of the input data. For this to occur, local minima need to be avoided. This may produce performance ratings that do not make sense. Two smoothing methods are used to achieve desirable results. Local-linear (Wand and Jones, 1995) and Nadaraya-Watson (Haykin 1999) smoothing are commonly employed in the training of SOM to obtain the output nodes.

Nadarya-Watson smoothing occurs first and is applied until either the number of specified iterations is reached or the convergence criterion is met. The local-linear smoothing then runs until the same conditions are met. These methods are described in the references given above. Local-linear smoothing eliminates the border effect where seeds (initial y positions) near the border of the grid are compressed in the topological mapping (Sarle, 2000; Wand & Jones, 1995, pp. 129-130). This means that the optimal minimum mean squared error (MSE) is inflated at the boundaries. Wand and Jones (1995) state that whilst this effect is merely an asymptotic argument, the boundary effect is pronounced in practise. Local-linear smoothing returns optimal MSE that are more similar at both the interior and exterior removing the discrepancy, whereas Nadaraya-Watson smoothing does not. This means that the within cluster variation is similar for output nodes at the interior and exterior. Nadaraya-Watson smoothing is applied first to prevent the network becoming stuck in bad configurations (local minima), a condition to which local linear smoothing is prone (SAS, 2000). Thus a fairly stable training process is employed. A kernel function (Wand and Jones, 1995) is built into the Nadarya-Watson smoothing process (Haykin, 1999). The shape of the kernel in the smoothing procedure used to define the neighbourhood of an output node for weighting purposes is not crucial in self-organising maps (SAS, 2000). It is the neighbourhood size that is important. The default process in SAS is to set the neighbourhood size to half the map size and have this decrease monotonically over time as suggested by Kohonen (2001). However, if the neighbourhood reaches zero too quickly, then the SOM becomes vector quantization, losing the spatial relevance of the output grid. Therefore it is important not to let the neighbourhood size become zero during training if topographical mapping is required (SAS, 2000).

The training of the SOM commences with the input patterns mapped to output nodes known as seeds. The location of the seeds are initialised using principal components which are found using data from the input nodes. For a two-dimensional SOM the seeds are set to an evenly spaced grid in the plane of the first two principal components

(SAS, 2000). Obviously, when a one-dimensional map is employed, points are evenly spread along the principal curve (first principal component). This provides the starting point from which the seeds are modified, until the final mapping for performance level is obtained. Consequently, the structure of the final map is shaped by the natural groupings of the input patterns. This choice of seeds provides the SOM with a head start, as the important variable relationships are identified in the initial map. An alternative principle is to initialise the seeds randomly, using randomly chosen input patterns for the output nodes. This was originally used to demonstrate the tendency of self-organisation in SOM's and is no longer considered necessary (Kohonen, 2001).

4.6 Interpreting Networks

As described in Section 4.2.4, neural networks are opaque (Berry & Linoff, 1997). To help understand the nature of these networks, sensitivity analysis can be used to determine how the predicted output of a network is influenced by each input or combination of inputs (Haykin, 1999). In the literature, authors use the term sensitivity analysis to represent different techniques.

Berry and Linoff (1997) describe a process where three manipulated levels (minimum, average and maximum) are presented to the network for each variable. The resultant impact on the output is monitored to gauge the overall influence, allowing the sensitivity of an output variable to input variables to be established. Obviously if the output varies greatly when presented with the three manipulated levels for a specific variable, the output is sensitive to that input variable and so forth.

Another approach is that described by Haykin (1999). He states that sensitivity analysis can be performed by accumulating the error measure generated from the presentation of test data to a network trained using backpropagation. This enables the sensitivity of the output to the inputs to be established over the entire test data set; as there will be a negative relationship between the error and importance of a particular input. However, these methods are somewhat cumbersome and the practical feasibility of sensitivity analysis is severely limited when there are several input variables.

Further, sensitivity analysis can be a graphical method. That is plots of the inputs against the output can be examined to identify relationships. Obviously, if a relationship exists between an input and output, then the output is sensitive to that input.

The aims of sensitivity analysis are similar to the half moon statistic which was introduced in Chapter Five. Like the HM statistic, sensitivity analysis can be expanded to consider the effect of groups of variables to see if certain combinations have a particular importance. This requires the manipulation of groups of variables using the three manipulate variables (minimum, average and maximum) approach described by Berry and Linoff (1997). However, the half moon statistic considers the impact of all observations at once, not just three key points, so is more attune to the global implications of the relationship between input and output. This gives an indication of not only the presence of a pattern, but the relative strength of the pattern that is imposed by an influential variable.

The software package, SAS, extracts rules for SOM using a cluster profile tree. This is calculated from a decision tree. The decision tree lists the percentages and numbers of cases assigned to each cluster and the threshold values of each critical input variable. This is displayed as a hierarchical tree and also lists all of the rules used to create the profile tree (SAS, 2000). To create the tree, splitting rules are chosen to maximise the splitting criterion, which uses the F-test for interval data (SAS, 2000). The extracted rules enable an understanding of the information obtained from the SOM. However, this understanding is only partial because the tree creates each branch based on a threshold for the critical input variables. This means that complicated relationships and important interactions are lost in the rule extraction process, especially during the early stages (top branches).

4.7 Neural Networks and Sport

According to Townend (1996) neural networks have seen limited application in the sporting environment. At the 3rd conference on Mathematics and Computers in Sport, Townend presented three examples utilising neural networks in sport. All three examples dealt briefly with the biomechanical aspects of sport as demonstrated below.

The first example, by Barton and Lees (1997) investigated the complicated motion of the human gait. Such an investigation cannot be performed in terms of single parameters. All variables must be used simultaneously together with any interdependencies. In this investigation a classification neural network exhibited great potential (Townend, 1996). Three conditions of gait were examined (normal, simulated leg-length difference and simulated leg-weight difference). For each of the three conditions hip-knee joint angle diagrams were constructed for eight subjects. The hip-knee joint angle diagrams allow investigation of the changes in the knee joint angle as a function of the hip-joint angle. The fast Fourier transform was applied to the 128 angle values gathered at regular time intervals, to extract the 30 pieces of data that were necessary to define the essential in characteristics for each of the gait conditions. This data were then applied to a four layer, backpropagation net consisting of 30 neurons in the input layer and three neurons in the output layer. The 30 neurons of the input layer corresponded to the 30 pieces of data generated by each of the eight subjects. The three neurons in the output layer referred to the conditions of gait examined. The hidden layers consisted of first five, and then four neurons, respectively. 24 data patterns (eight subjects and three conditions) were divided into training and test data sets. The capability of the neural network was assessed by the assignment ratio, which is the ratio of successful recognitions to the total number of test patterns. Four different training and test subsets were used with each of the test patterns submitted four times, giving a total of 16 neural networks. The four best neural networks were retained. Each of these networks assigned the unknown gait patterns to the correct condition in 83.3% of the cases. This indicated that neural networks could be used in the automated diagnosis of gait disorders.

The second example, again by Barton and Lees (1995), utilised neural networks in assessing different insole materials used in the construction of running shoes. From this study it was found that classification networks were capable of recognising different foot pressure points associated with different insole conditions. Accuracy of the networks was maintained even in the absence of complete data sets. Also, insensitivity to random noise was proven.

The final example given by Townend (1996), described the work of Barton, Lees and Wit (1996) investigating the EMG activity of the four large muscle groups of the leg during a number of steps and the vertical component of the ground reaction force obtained from one of these steps. This last measurement is very difficult to obtain. The EMG values were used as the input to a predictive neural network trained to map these inputs to the corresponding vertical force component. Following training, previously unknown EMG data were presented to the neural network and the results indicated that the proposed method predicted the vertical ground reaction successfully (high correlation coefficient 0.9 ± 0.05 and low residual mean squared error 0.06 ± 0.02). The implications of this study allow data to be collected during unrestrained activity (as opposed to a force platform test) over several strides (rather than just one) and then analysed by the neural network in order to predict the vertical component of a step, making the difficult measurement unnecessary.

All three examples presented above involved supervised learning. Whilst the connection to sport was rather tenuous, the examples applied neural networks in different guises, firstly as a pattern recognition tool in classification (Barton *et al*., 1997, 1995) and secondly as a tool for prediction (Barton *et al*., 1996). In this thesis classification was of potential use in the pre-processing of the data to identify similar groupings of positions. However, as the potential positional clusters are not separable at the level of refinement required, this is no longer relevant. Prediction is also not our purpose. Instead this thesis has produced the first sports example of neural networks for dimension reduction.

Neural networks are applicable in many situations. However, only a portion of these methods are suitable for this thesis. In many practical applications there is no need to estimate high dimensionality in multivariate data because multivariate data generally have a dimensionality lower than that originally proposed (Cherkassky & Mulier, 1998, p.162). That is redundancies exist within the data that can be removed through dimension reduction techniques, allowing data of a lower dimensionality to be assessed. In the case presented by rugby there are clearly a number of redundancies with the high dimensionality of the initial data set. Core components can be identified intuitively, areas which were confirmed via factor analysis in Chapter Three, such as attack, defence and physical involvement.

Whilst there is an abundance of literature detailing the methodology of neural networks in supervised prediction and pattern recognition (i.e. classification), and a reasonable coverage of material relating to unsupervised clustering and pattern recognition, there is only limited information regarding dimension reduction techniques using neural networks. A possible explanation for the dearth of such information could be due to interpretation difficulties for the hidden layers in such a network. Self-supervising networks with a bottleneck layer provide the architectural framework required for the dimension reduction problem presented in this thesis.

There are four potential methods for performing dimension reduction in this thesis due to the absence of suitable target values on an individual level, as follows:

a) supervised using match outcomes as outputs
b) self-supervised using a bottleneck layer and employing inputs as outputs
c) self-supervised using factor scores as outputs
d) unsupervised using self-organising maps

The first approach would require the interactions between individual's to be assessed as discussed in Chapter Two. However, in order to fully appreciate the effect of these individual interactions, subjective quality measures relating to task performance are required as well as spatial information and an interpretation of sports chaotic element. As an extensive data set containing descriptive measures of performance doesn't exist, such an approach cannot be adopted. Further, the resultant output would require team performance to be quantified as well as individual performance leading to inferences regarding performance on both an individual and team level. Using factor scores as outputs in a self-supervised network would not provide any additional information in trying to create a single value rating system. Consequently, the most applicable methods for use with the rugby data, for the problem defined in this thesis are the self-supervised network using a bottleneck layer and employing inputs as outputs, and the unsupervised network using self-organising maps.

In the next chapter, Chapter Five, a new method for interpreting networks is introduced. Chapter Six applies and explores the neural network techniques discussed in this chapter (self-supervising feed-forward neural networks and SOM), with strengths and limitations briefly examined.

Chapter Five

Investigation and Development of a Heuristic Test for Independence

5.1 The Introduction of a Heuristic Test for Independence

This chapter provides an important component of the second part of this thesis. In the quest to improve the quantification of individual rugby player performance featured in Chapter Three, more flexible neural network models are to be applied. As discussed in Chapter Four, neural networks are applicable tools for solving many types of statistical problems (Fausett, 1994). However, a major limitation of the application of supervised and unsupervised neural networks to a data set is the inability to interpret the hidden layers (Medsker, Turban & Trippi, 1993; Cheng & Titterington, 1994; Lowe, 1997; Warner & Misra, 1996; Ciampi & Lechevallier, 1997; Schwarzer, Vach & Schumacher, 2000). This chapter rectifies this limitation by introducing a method for detecting and comparing the importance of input-output relationships.

In this chapter, an *algorithmic* test for independence is constructed. It is intended that this test will indicate the presence of relationships between two or more variables, generally an input-output combination. Firstly the basic philosophy behind the new heuristic test is discussed. This leads to the introduction of the half moon statistic (HM) which is a new method for detecting dependence between variables. The properties of this technique are expanded and then tested against traditional methodology, namely correlation based analysis. A Chi-Square test of independence is also used to validate the new technique.

This chapter will show that the HM statistic is a powerful and versatile tool for use in exploratory data analysis. Primarily, this new statistic is for the detection of non-independence between two variables. The HM measure, like correlation in a linear setting, increases strictly monotonically as the strength of association between variables increases. What makes this statistic so useful is the fact that no model assumptions are necessary. In this chapter 'robustness' is the common statistical definition that concerns the lack of sensitivity to violations of assumptions. Input variables are the same as predictor, independent or explanatory variables. Also, dependent and response are synonyms for output.

The usefulness of the HM statistic is not limited to measuring the association between pairs of variables. It is expanded to cater for the multivariate case (MHM) by considering the overall impact a set of input variables has upon an output variable. This allows the significance of a multiple input-single output association to be calculated.

The advantage that the HM statistic has over conventional methods for determining the strength of association between variables is functional independence. That is, the nature of the function describing the association between the input and output does not need to be known. This is a change from traditional model formulation and related statistical inference which "assumes the existence of a 'true' model" (Chatfield, 1995, p. 419).

In order to construct a much needed flexible method that determines whether a relationship exists, without assuming an underlying model (Cheng & Titterington, 1994), an *algorithmic* approach is adopted (Breiman, 2001). This emphasises the workability of the proposed method using empirical evidence, as adopted in data mining and related disciplines (Hand, 1999), rather than constructing a mathematical proof that the methodology works.

This development is driven by the need to interpret the hidden layers of a neural network. The evolution of the HM statistic is based on the question: what happens when there is a relationship between two variables? In a graphical sense, there is quite clearly a pattern. Whether this pattern is a straight or curved line, significant or not, the strength of this "pattern" can be observed and should therefore be measurable. Section 5.2 examines the HM statistic as a method capable of measuring such patterns.

Current methodologies are functionally dependent and are not flexible enough to detect significant relationships between the layers of neural networks. For example, linear correlation tests for linear dependency whereby points are compared to a straight line (Walpole, 1974). The comparison to a straight line as suggested by linear correlation assumes relationships will be in a linear form or immediately reducible to a linear form. This requires the functional form of the relationship to be known (or approximated) in order for an appropriate transformation to be found. More complex relationships can also be examined, provided the relationship between the input and output is known. The opposite approach, assuming the input and output to be independent, is the idea behind the half moon statistic.

The theory and ideology introduced in this chapter is currently being reviewed for publication by the *Australian and New Zealand Journal of Statistics* (Bracewell).

5.1.1 Covariance, Correlation and the Coefficient of Determination

Typically, covariance, correlation and the coefficient of determination are methods used to quantify the strength of association between variables. These methods are related as discussed in most introductory statistical texts.

The covariance between two random variables (X, Y) is given by:

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

A non-zero value of the covariance is evidence of non-independence (Johnson & Leone, 1964, p.57). The scaling of the X-Y variables by subtracting the respective means focuses the measure on the relative dispersal from the centre of the swarm. Whilst it is convenient to regard the covariance as an index of dependence, this value can be zero even when the variables in question are not independent due to a non-linear relationship (Johnson & Leone, 1964, p. 57). Covariance is not scale invariant in that a linear transformation of the X and Y data will alter the covariance. Additionally, this measure is restricted to only two variables. However, this covariance can be standardised to produce a dimensionless measure (correlation).

The coefficient of linear correlation has scale invariance. This is a ratio obtained by dividing the covariance by the square root of the product of the variances (Johnson & Leone, 1964, p.400), as shown below:

$$\mathrm{Corr}\,(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

This scales the measure such that it has meaningful bounds [-1, 1]. "The correlation between variates is a measure of the departure from circularity of the projection of the swarm (after standardisation) on to the plane formed by the two variate co-ordinate axes" (Krzanowksi, 1988, p51). In other words, it is a "measure of the eccentricity of the ellipse so formed" (Krzanowksi, 1988, p51). As with covariance, calculation of the correlation coefficient compares observations with the centre of the ellipse, or swarm. The measurement of the eccentricity of an ellipse restricts meaningful association measures to linear relationships only (see also Johnson & Leone, 1964).

The coefficient of determination is a measure of dependency for linear and non-linear relationships, defined as follows:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

Subtracting the fitted model value, $\hat{y}_i$, from the ith observed value, y_i gives the residual, e_i, for this observation. The coefficient of determination can be interpreted as the proportion of the total variability that is explained by the model provided there is a suitable base model for comparison (Scott & Wild, 1991). In the case of a linear relationship between two variables (X, Y) it is simply the square of the correlation coefficient – hence the name 'r^2'. It can be used for any model form with any number of predictor variables, and is the most commonly used descriptive measure of the strength of a relationship (Levine, Krehbiel & Berenson, 2000). However, to obtain the residual, e, in the definition above, the fitted values must first be obtained. Importantly, when non-linear models are examined, the coefficient of determination becomes less meaningful (Scott & Wild, 1991). For comparative purposes in this thesis the coefficient of determination is evaluated for non-linear functions, using properties of expectation and variance for independent normally distributed data. This ensures that an adequate base model for comparison is used so that the coefficient of determination is meaningful.

5.1.2 Other Tests of Independence

There also exist other techniques and non-parametric measures of correlation that can be used as tests of independence. These are described briefly in this section.

Spearman's rank correlation (r′) is similar to sample correlation and is computed by replacing the ith observed value of x (x_i) with its associated rank (x_i'), and similarly for y, such that:

$$r' = 1 - \frac{6\sum_{i=1}^{k}(x_i' - y_i')^2}{k(k^2-1)}$$

Here k is the sample size. However, to provide a meaningful measure of dependence, Spearman's rank correlation is restricted to linear relationships, or functions that are either increasing or decreasing (not necessarily strictly).

Kendall's τ is an alternative measure of correlation based on rank order statistics. This measure is defined as:

$$\tau = \frac{4R}{k(k-1)}$$

R is the number of correctly ordered ranked (X, Y) points and k is the sample size (see Johnson & Leone 1964). This test of association is useful only for functions that are either increasing or decreasing due to the comparison of the (X, Y) rank pairings. That is functions with local minima and/or local maxima will not be accurately represented by rank pairings.

More sophisticated tests of independence have been developed. Kallenberg and Ledwina (1999) constructed a data driven rank test for independence. This was done to consider forms of non-linear dependence between variables. The data is used to identify the degree of polynomial describing the relationship between the input and output variable. A stepwise procedure is implemented to account for the correlation of higher order polynomials, with a modification of Schwarz's selection rule used for selecting the best polynomial (see Kallenberg & Ledwina, 1999). Spearman's rank correlation forms the basis of this correlation measure. However, assumptions are made relating to the model family f(x) in that the polynomial order must be known.

Copula functions (also known as grade representation) are becoming increasingly popular as a method of determining multivariate dependence (Kallenberg & Ledwina, 1999). However, this requires the distributional family of the joint distribution of the variables being examined to be known (Joe, 1997). Consequently, copula functions are not examined in this thesis.

Another test for independence is the Chi-Square test which uses categorical data (see Smith, 1993). This test is used in this chapter to validate the new HM statistic. "Categorical" data is created by placing observations into one of four categories using the quartiles for both variables (input and output), thereby creating a four by four contingency table. The data is divided into quartiles for convenience and ease of manipulation. The use of fewer categories gives poorer results. The chi-square statistic, χ^2, is defined as follows:

$$\chi^2 = \sum_{all\ cells} \frac{(f_o - f_e)}{f_e},$$

where f_0 is the observed frequency in a given cell and f_e is the expected frequency of that cell. The expected frequency is obtained by multiplying the corresponding row and column total frequencies and then dividing this product by the overall count (k).

The lack of independence between variables implies some form of association. Intrinsic relationships of a neural network can be extracted graphically using sensitivity analysis. Sensitivity analysis determines how sensitive the predicted output of a network is to each input, or combination of inputs (Haykin, 1999). Three manipulated levels (minimum, average and maximum) are presented to the network for each variable. The resultant impact on the output is monitored to gauge the overall impact. The sensitivity is represented by the relative amount of change in the output as a response to the manipulation of the input variables. Alternatively, sensitivity analysis can be performed by accumulating the error measure for each input variable generated from the presentation of test data using a back-propagation algorithm. This enables the sensitivity to be established over the entire test data set on the basis that there will exist a negative relationship between the error, and the importance of the variable in the model (Berry & Linoff, 1997). However, these methods are somewhat cumbersome. The practical feasibility of sensitivity analysis is severely limited when there are several input variables.

Other areas of research, such as computer science, also seek to interpret neural networks. Various rule extraction methods can be used for extracting knowledge regarding the behaviour of a network which employs categorical or non-numerical data. This is to assist users with the interpretation and understanding of the underlying processes (Andrews, Diedrich & Tickle, 1995; Maire, 1999; Garcez, Broda & Gabbay, 2001). In the dimension reduction problem motivating this thesis, numerical data is presented, making sensitivity analysis more appropriate. The half-moon statistic is also only appropriate for numerical data. It compliments a sensitivity analysis by giving a statistical test for the significance of an input-output relationship, while a sensitivity analysis provides graphical support for these relationships. Additionally, the half-moon statistic can be expanded to cater for multiple input variables.

If the model family of the function describing the association between two variables is non-linear and unknown, then the previously described methods cannot be easily applied to achieve a measure of association. In the following sections it will be demonstrated that by shifting the focus from the centre of the swarm, formed by a population of X-Y observations, to the perimeter of the swarm, a functionally independent measure of association can be derived.

5.2 Underlying Philosophy for the Half-Moon Statistic

The introduction of a functional relationship between two standard normal variables distorts the circular swarm projected onto the plane formed by the two variate axes. This distortion should be measurable, regardless of the functional form of that relationship. This distortion will cause a change in the variance associated with the distances between the points and the circle circumference.

Consider the vertical distance (Z) from a specified point (X, Y) to a point, (X, U), on an enveloping circle of predefined radius, r, as shown in Figure 5.1. Then the variance of Z can be used to measure the strength of the X-Y relationship, as the distribution of distances will differ depending on the strength of the association. It will be shown that the HM statistic is not sensitive to the value of r, for r sufficiently large ($r \geq 10$), provided that the data does not have extreme skewness or kurtosis.

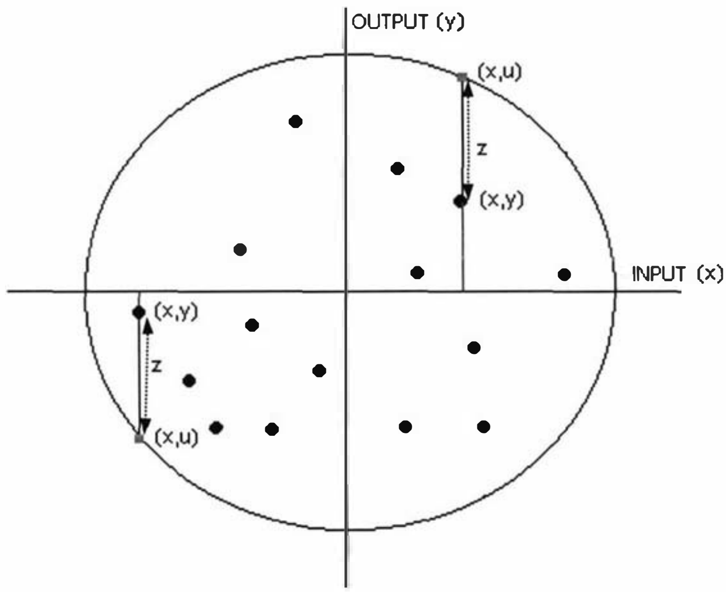


Figure 5.1: *Geometrical Representation of the Half-Moon Statistic*

5.2.1 Mathematical Background for Bivariate Case

Figure 5.1 illustrates the geometrical derivation of the half-moon statistic for the bivariate case. The magnitude of z is calculated for every (x, y) pair by subtracting the absolute value of the output y from the absolute value of u, where u is the y co-ordinate for a point on the circle perimeter with input x. The half-moon statistic is defined using the sample standard deviation for z, calculated using all available (x, y) pairs. Given that the data has been standardised with a mean of zero and a standard deviation of one, under the initial imposition of independence, consider a circle centred at (0, 0) with radius r, where r is chosen so that all the (x, y) points lie inside the circle. Then, if x_j is the jth observation of the input variable and y_j is the jth observation of the output variable,

$$x_j^2 + u_j^2 = r^2 \qquad (1)$$

$$u_j = \sqrt{r^2 - x_j^2}, \qquad \text{where } r \geq |x| \qquad (2)$$

$$z_j = u_j - |y_j| \qquad (3)$$

with sample standard deviation:

$$s_z = \sqrt{\frac{\sum_{j=1}^{j=k} z_j^2 - k\bar{z}^2}{k-1}} \qquad (4)$$

Then the half moon statistic (HM) is defined as

$$HM = s_Z - \sigma_z \qquad (5)$$

where σ_Z is the population standard deviation for Z when X and Y are independent. Z is the vertical distances to the enveloping circle. Using the absolute value for the output y projects the points into half a circle (upper half); hence the name, half-moon statistic.

The distance measure chosen can be calculated in a number of ways. These are described briefly in Appendix F. However the method proposed here produced the most promising results. The asymptotic zero-mean (see Larson, 1982, p.349) feature of this statistic is crucial for creating a robust statistic for the non-parametric and multivariate extensions to follow.

The adoption of a particular radius for the univariate parametric statistic presented here is arbitrary, although it must exceed the magnitude of each input observation (after standardisation). This effect is illustrated in Figure 5.2, with boxplots for seven different radii, calculated using 500 parametric half-moon statistics, each obtained from 500 observations from independent standard normal populations. It can be clearly seen from this boxplot that the variances are equivalent. A radius of 10 is adopted in this thesis.

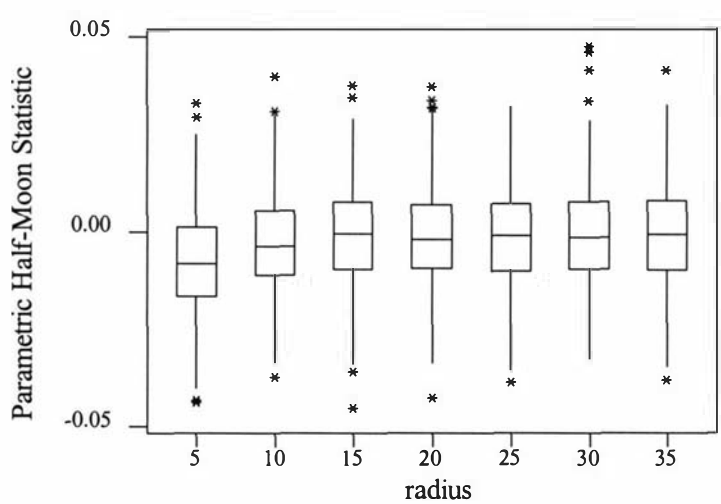


Figure 5.2: *Boxplot of the Parametric Half-Moon Statistic for Varied Radii*

A parametric test for independence is created using the distributional properties of the HM statistic when the input-output variables are independent, which is examined before a more flexible (and suitable) non-parametric measure is developed.

5.3 Parametric Test for Bivariate Case

The half-moon statistic is defined using the standard deviation of Z. When X and Y have independent standard normal distributions the expected value for the HM statistic is easily calculated by using the two independent terms indicated in (3). The probability density function associated with $|Y|$ is a folded standard normal, defined as follows:

$$f_{|Y|}(y) = \frac{2\exp[\frac{-y^2}{2}]}{\sqrt{2\pi}}, \; y>0$$

with $E(|Y|)=2/\sqrt{2\pi}$ and $Var(|Y|)=1-2/\pi$ and the probability density function associated with X^2 is chi-squared with one degree of freedom.

The expected value of $\sqrt{r^2 - X^2}$ is approximated using a Taylor expansion (Anton, 1995). When r^2 is greater than or equal to the largest x^2, then it can be shown that for X ~ N(0, 1) (where the normal distribution of X is truncated on the range [-r, r]),

$$\begin{aligned} Var(\sqrt{r^2 - X^2}) &= E(\sqrt{r^2 - X^2})^2 - E^2(\sqrt{r^2 - X^2}) \\ &= E(r^2 - X^2) - E^2(r - \frac{X^2}{2r} - \frac{X^4}{8r^3} - \ldots). \end{aligned}$$

Because the higher moments of the normal distribution become negligible for a normally distributed random variable, the variance becomes,

$$\begin{aligned} Var(\sqrt{r^2 - X^2}) &\approx (r^2 - 1) - \left(r - \frac{1}{2r} - \frac{3}{8r^3}\right)^2 \\ &\approx \frac{1}{2r^2} - \frac{3}{8r^4} - \frac{9}{64r^6}, \end{aligned}$$

which provides a good approximation (Wand & Jones, 1995). Therefore for two independent, standard normal (X, Y) populations the expected variance is,

$$Var(Z) \approx \left(1 - \frac{2}{\pi}\right) + \left(\frac{1}{2r^2} - \frac{3}{8r^4} - \frac{9}{64r^6}\right). \quad (6)$$

The square root of the expected variance Var(Z) is the population standard deviation, σ_Z, given independence and the bounded conditions described above.

Therefore, the parametric half-moon test statistic for testing independence between two standardised normally distributed variables is then defined as,

$$HM \approx s_z - \sqrt{\left(1 - \frac{2}{\pi}\right) + \left(\frac{1}{2r^2} - \frac{3}{8r^4} - \frac{9}{64r^6}\right)}.$$

A generic HM statistic with a mean of zero is produced by subtracting the population standard deviation under independence, σ_Z, from the sample standard deviation, s_Z, as shown in Figure 5.3.

The rejection region for the hypothesis of independence is in the tails of this distribution because input-output dependence increases the absolute value of the HM statistic.

A simulated distribution of 800 parametric HM statistics under the null hypothesis of independence is shown in Figure 5.3 for a sample size k of 500. The distribution of HM under the assumption of normality appears normally distributed – which is substantiated by the Anderson-Darling test which failed to reject normality (p=0.337) – with a mean of zero, and a standard deviation of 0.0125.

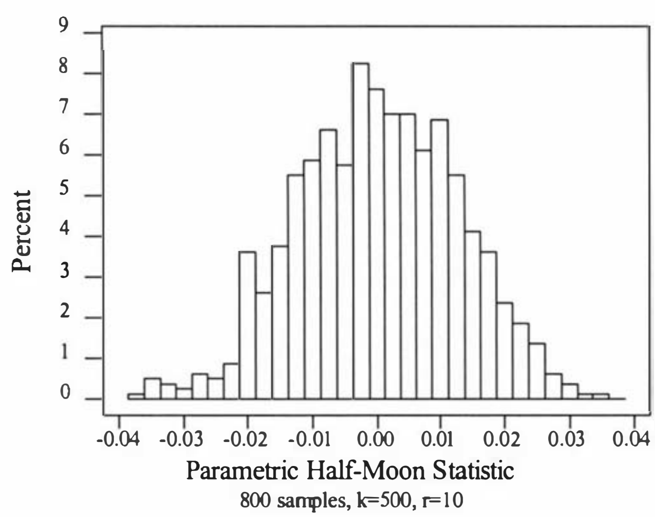


Figure 5.3: *Histogram of Parametric HM under the Assumption of Independence and Normality for X and Y*

This enables appropriate critical values for the parametric test to be established. That is when X and Y are standard normal then for a 5% significance level and a sample size of 500, we must reject the null hypothesis of independence whenever HM lies outside the range [-0.0242, 0.0242] (analogous to a coefficient of determination of 0).

The corresponding critical value for independence at the 5% level of significance for the coefficient of determination is 0.0077 ($\pm 0.0877^2$), obtained using Fisher's approximation (Johnson & Leone, 1964). The effect of k on the critical values for independence is examined later for the non-parametric case.

It is important to validate the HM statistic as a test for independence using accepted measures of association. This is done for linear relationships of the forn y = b.x + c.e, for x ~ N(0, 1), e ~ N(0, 1), b^2 = 0.0, 0.1,...,1.0 and $c^2 = (1-b^2)$. Plotting the parametric HM statistic against the coefficient of determination in Figure 5.4 indicates a strictly monotonically increasing relationship. This supports the use of this statistic as a measure for the strength of association between variables in the linear case, when X and Y are normally distributed. One large sample size (8000) with only one repetition is used to emphasise the strictly monotonic relationship between R-Sq and DHM. Critical values for the HM statistic are examined in the following sections.

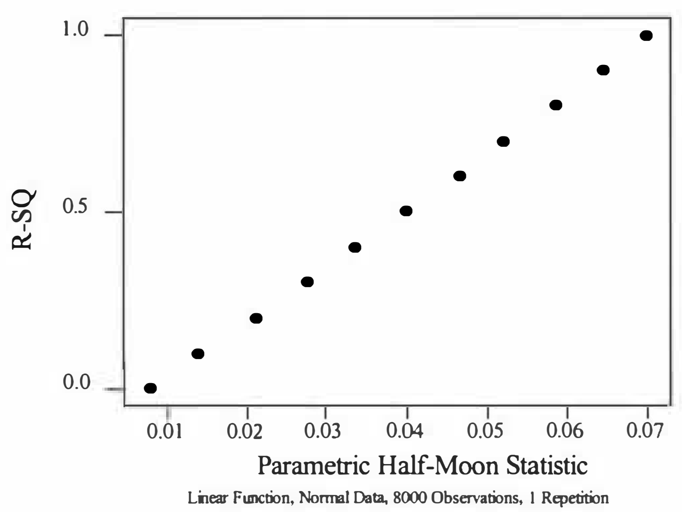


Figure 5.4: *Plot of the Parametric Half-Moon Statistic against the Coefficient of Determination*

The accuracy of the approximations introduced in this section to describe the properties of the parametric HM statistic are not dwelled upon, as the following sections will show that the non-parametric method is a superior test for independence.

5.3.1 Disadvantages of Parametric Test

Whist the previous sections have demonstrated the workability of the parametric half-moon statistic for linear relationships, the dependence on the normal distribution produces an inflexible method that is vulnerable to slight departures from normality and the impact of outliers. Additionally, a change from a linear to a non-linear relationship between X and Y is not explained well due to changes in the variation of the raw parametric statistic. This is a consequence of changes in the expected population standard deviation caused by alterations to the underlying distributions (equation 6).

The parametric test is vastly improved by removing the parametric properties from the calculation of the half-moon statistic. The non-parametric half moon statistic is a highly flexible method, which compensates for changes in the type of function determining the X-Y variable association by considering the unique characteristics of the distribution that result from the association.

5.4 Non Parametric Test for Bivariate Case

When the distributions of X and Y are unknown, a resampling approach is used to define a non-parametric test of independence for the half-moon statistic. The test statistic is defined as

$$HM = s_z - \bar{s}_z$$

where $\bar{s}_Z$ is an estimate for σ_Z. This is calculated using the average of m standard deviations of Z obtained from independent samples selected without replacement from the standardised Y data, with the standardised X data unchanged. m is the non-parametric resampling coefficient. The resampling of the X data is used to ensure that $\bar{s}_z$ is calculated for independent X and Y samples. Figure 5.5 shows the distribution of this non-parametric HM statistic when X and Y are independent standard normal variables when m=3.

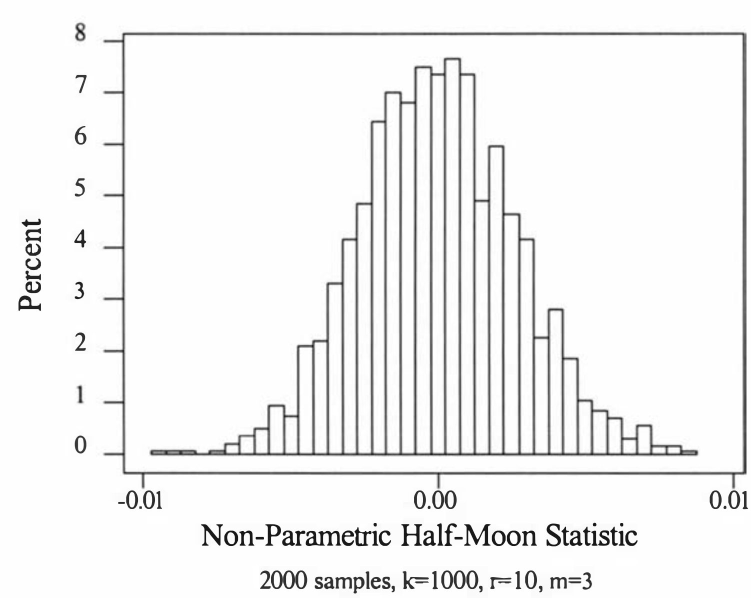


Figure 5.5: *Histogram of Non-Parametric HM under Assumption of Independence and Normality*

A simulated distribution of 2000 non-parametric HM statistics generated from independent X-Y variables, shown above in Figure 5.5, seems normally distributed with mean, 0, and standard deviation, 0.0027 for a sample size of 1000. The Anderson-Darling normality test fails to reject normality for this data (p=0.178). This suggests that critical values of ±0.0053 at the 5% level of significance are appropriate. Due to the resampling approach adopted in the construction of the HM statistic, the standard deviation of this statistic is dependent on sample size.

Due to the sampling process, the choice of m impacts on the standard deviation of the non-parametric HM statistic, the type II error and also the time taken to compute the half moon statistic. That is, as m increases, the variance of the HM statistic decreases and consequently smaller critical values are obtained. Reducing the magnitude of the critical values decreases the likelihood of accepting a null hypothesis of independence when significant dependence is present (type II error). Increasing m requires more calculations to be performed, thereby increasing the computational time. This effect is minimal in the testing of actual data, but it is an issue when exploring the workability of the HM statistic through multiple simulations.

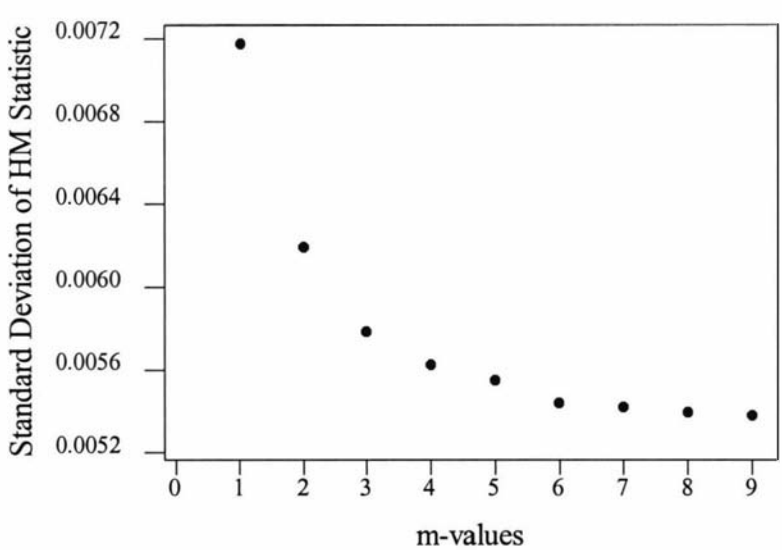


Figure 5.6: *Plot Illustrating the Effect of the Re-Sampling Coefficient on the Sample Standard Deviation for the HM Statistic*

However, as shown in Figure 5.6 for 100 observations from standard normal data, the effect of reduced HM standard deviations as a result of increased m becomes less pronounced for $m \geq 3$.

Figure 5.6 illustrates the effect of decreasing standard deviation for 200 half-moon statistics obtained from 100 observation, given an increase of the re-sampling coefficient, m. To reduce the computational time and minimise the type II error of the half moon statistic, the re-sampling coefficient is set at 3 for this thesis. Altering of m will alter the critical values examined in this thesis.

The non-parametric test is a more precise test for independence than the parametric test because the idiosyncrasies of the data are retained. That is the unique characteristics of the data are incorporated into the calculation of HM, reducing the variability of HM, thereby reducing the probability of Type I and Type II errors occurring. Comparing the parametric and non-parametric versions of the half moon statistic using the same data (standard normal data, $k = 500$, $m = 3$) highlights this effect. Whilst in both cases the mean is zero from 800 examples, the standard deviations of the methods are hugely different. The parametric test has a standard deviation of 0.01253 which is almost 3.5 times larger than the standard deviation for the non-parametric test, 0.00361. Consequently, the parametric test is not explored further.

5.4.1 Effect of Sample Size

Due to the resampling process, the variability of the HM statistic is directly related to the sample size. In this simulation the HM statistic was calculated using standard normal data for the independent input and output variables. Employing 25 different sample sizes (50, 100,..., 1250) with 200 repetitions for each sample size, provided the data for Figure 5.7 and the subsequent regression analysis. Transforming sample size using the power (-1/2) produced the linear relationship evident in the figure below.

As a consequence of this linear relationship, appropriate critical values can be obtained with regard to the sample size. More importantly, because the distribution of the half-moon statistic is approximately $N(0, \psi^2)$, where ψ is determined by the sample size, p-values can be evaluated. From the regression analysis using the data from the above figure;

$$\psi = 0.0818174k^{-1/2},$$

where k is the sample size (p=0.000, R-SQ = 99.2).

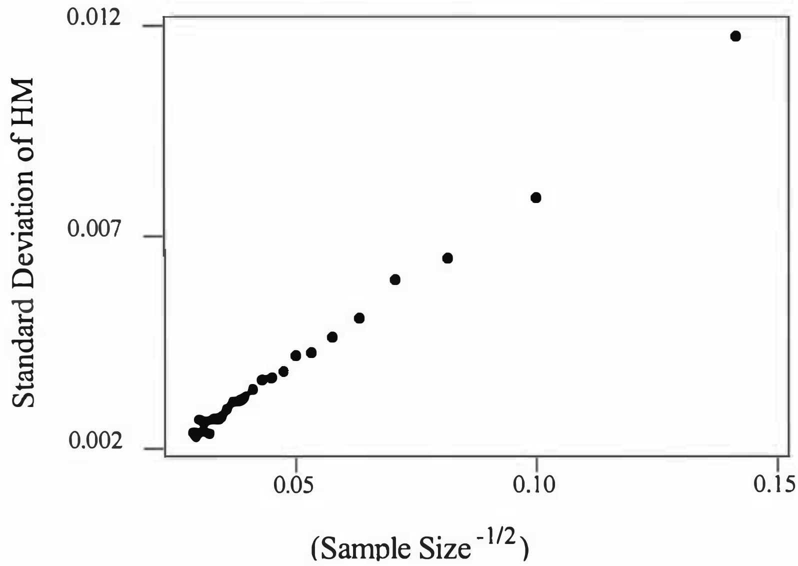


Figure 5.7: *Relationship between Sample Size and Half Moon Variability*

Therefore, for k = 8000, the following critical values apply (10%, ±0.0015; 5%, ±0.0018, 1% ±0.0024). Constructing a two-tailed test for independence at the 5% level of significance using 5000 independent N(0,1) X-Y pairings, with variable sample size (k = 50, 100,..., 1250) returned a misclassification error rate of 4.92% (246/5000).

With a mean, 0, and a standard deviation, ψ, critical values are calculated from standard normal tables at the required significance level. The use of normally distributed simulated data in Figure 5.5 means that those critical values are only appropriate for variables that are approximately normal.

For the HM statistic to be effective, it needs to withstand skewness and kurtosis in the input variables, due to the influence of the encompassing circle. The effect of skewness and kurtosis is established empirically in the following section.

5.4.2 Effect of Skewness and Kurtosis

Preliminary analysis revealed that positively skewed data tended to increase the standard deviation of the HM statistic. As sport data is generally skewed this is an important consideration for this thesis. A number of distributions with different measures of skewness and kurtosis are used as input variables and assessed under independence to provide sufficient information on how the HM statistic is affected by these attributes. Whilst some of these distributions are not likely to be encountered in this thesis (beta, F, weibull and Cauchy for example), they do provide a data set that allows the impact of these attributes to be evaluated statistically. In the figures 5.8 and 5.9 it is shown that there is a relationship between the coefficients of skewness and kurtosis, and the HM statistic. The sample coefficients of skewness, ς_3, and kurtosis, ς_4, are defined as follows (Smith, 1993),

$$\varsigma_3 = \frac{g_3}{\left(\sqrt{g_2}\right)^3},$$

$$\varsigma_4 = \frac{g_4}{\left(\sqrt{g_2}\right)^4},$$

where g represents the qth sample moment of x about the sample mean for k observations as follows,

$$g_q = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^q}{k}$$

such that normally distributed data has a skewness coefficient of 0 and a kurtosis coefficient of 3.

The coefficient of kurtosis was transformed by first taking the natural log and then the square root, to produce Figure 5.9. When the kurtosis of the input distribution exceeds approximately 40 (transformed kurtosis $\approx$ 2) the linear relationship that exists between the transformed kurtosis and the square root of the standard deviation of the HM statistic deteriorates. Consequently the corrected statistic is not suitable for data with high kurtosis. For highly skewed distributions and distributions with excessive kurtosis, standardisation fails to bring all points within the encompassing circle. That is some values of x^2 exceed r^2. Consequently the calculations required to produce the HM statistic are undefined for these observations. This also leads to problems in defining the theoretical moments of some components of the HM statistic. However, simulation allows these moments to be approximated.

The linear relationship evident in figures 5.8 and 5.9 allowed a correction factor for independent X-Y distributions. Figure 5.10 shows the combined effect of kurtosis and skewness upon the standard deviation of the HM statistic from a three-dimensional perspective. The *D*istribution corrected *HM* statistic (DHM) was obtained via regression analysis using the 34 observations detailed in Table 5.1 (R-SQ = 99.0%).

Distributions with high skewness and kurtosis apparent in the previous figures were omitted from the regression analysis and are denoted by italics in Table 5.1. As indicated below,

$$\mathrm{DHM} = \frac{\mathrm{HM}}{\left(0.013391 + 0.0016113\,|\varsigma_3| + 0.057329\sqrt{\ln(\varsigma_4)}\right)^2},$$

where ς_3 is the coefficient of skewness and ς_4 is the coefficient of kurtosis. Similar to the HM statistic, the DHM statistic is affected by sample size. Consequently, this more robust statistic is approximately $N(0, \kappa^2)$ distributed for independent X-Y pairings, where $\kappa = 15.2752k^{-1/2}$ and k is the sample size. As before, the formula for κ was obtained using simulated normal data for 25 sample sizes (R-SQ = 99.2). This information is then used to calculate p-values, indicating the likelihood that the association between X and Y variables differ from the null hypothesis of independence. Importantly, the HM and DHM statistics are equivalent for normally distributed data ($\varsigma_3 = 0$, $\varsigma_4 = 3$), after correction for sample size. That is both become standard normal distributions with means of zero and equivalent standard deviations following the adjustments for sample size.

The DHM statistic was used to create a two-tailed test for independence at the 5% level of significance. Calculated using 5000 independent standard normal X-Y pairings with a randomly chosen sample size (k = 50, 100,..., 1250) returned a misclassification error rate of 4.8% (238/5000). Using the skewed data that produced figures 5.8 and 5.9 produced an error rate of 5.2% (2033/39000) for the DHM statistic. These errors are obviously very close to the required 5% level of significance.

Using a bootstrapping procedure to obtain the test statistic for independence can circumvent the need for sample size and distribution corrections. However, investigation of an extended bootstrapping procedure is left to future research as the HM statistic and its derivatives developed in the previous sections are sufficient for the needs presented in this thesis. That is, the HM statistic behaves in a very predictable way (given different sample sizes, skewness and kurtosis). It makes sense to adjust the statistic (HM) to reduce the variability of the test statistic (DHM) and allow a single generic table to be used regardless of sample size or distribution.

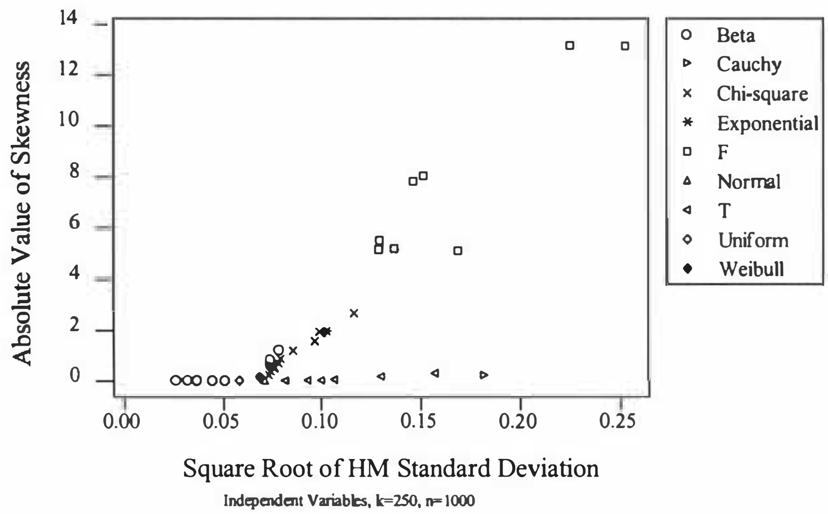


Figure 5.8: *Relationship Between the HM Statistic and Skewness*

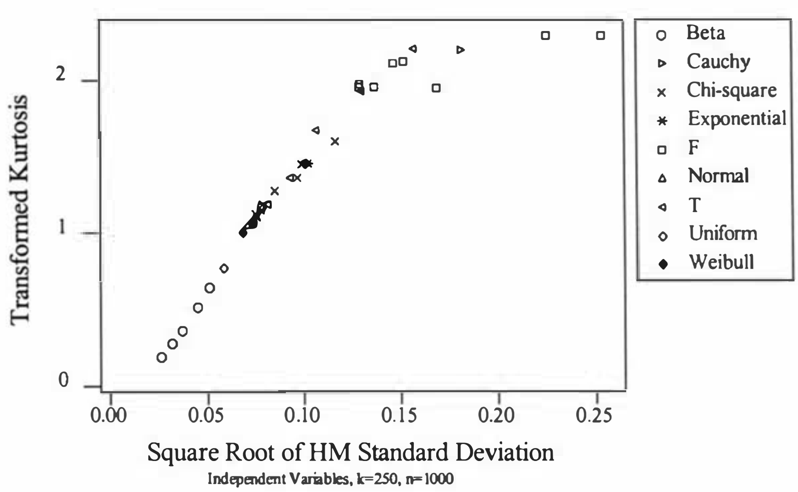


Figure 5.9: *Relationship Between the HM Statistic and Kurtosis*

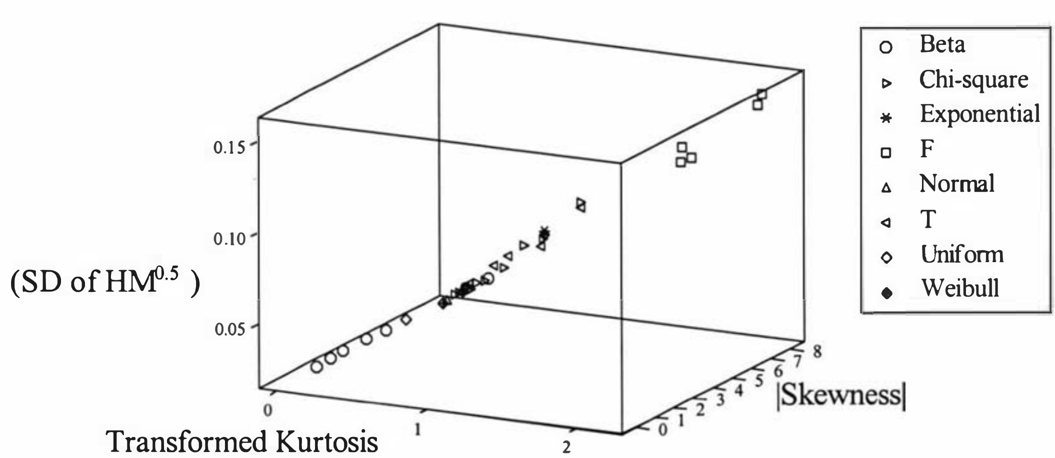


Figure 5.10: *3D Plot Demonstrating Relationship Between HM, Skewness and Kurtosis*

Distribution	*Description*	*Mean (Skewness)*	*Mean (Kurtosis)*	*SD(HM)*
Beta	B(0.01,0.01)	0.0042	1.0344	0.0007
Beta	B(0.05,0.05)	0.0090	1.0780	0.0010
Beta	B(0.1,0.1)	0.0017	1.1352	0.0013
Beta	B(0.25,0.25)	-0.0059	1.2971	0.0020
Beta	B(0.5,0.5)	0.0013	1.5127	0.0026
Beta	B(2,0.5)	-1.2468	3.8374	0.0061
Beta	B(3,1)	-0.8559	3.1009	0.0055
Chi-square	Chi(60)	0.2541	3.0728	0.0054
Chi-square	Chi(30)	0.3664	3.1894	0.0055
Chi-square	Chi(25)	0.4998	3.3446	0.0057
Chi-square	Chi(120)	0.5482	3.4168	0.0058
Chi-square	Chi(20)	0.6028	3.5018	0.0057
Chi-square	Chi(15)	0.7180	3.7151	0.0061
Chi-square	Chi(10)	0.8650	4.0653	0.0063
Chi-square	Chi(5)	1.2131	5.0528	0.0073
Chi-square	Chi(3)	1.5531	6.3498	0.0093
Chi-square	Chi(2)	1.8941	8.0061	0.0098
Chi-square	Chi(1)	2.6573	12.8224	0.0136
Exponential	e(1)	1.9221	8.2008	0.0106
F	F(1,3)	8.0492	89.3096	0.0231
F	F(1,5)	5.5294	48.6389	0.0167
F	F(3,3)	7.8587	86.5426	0.0216
F	F(3,5)	5.1898	45.9162	0.0187
F	F(5,5)	5.1412	45.4764	0.0166
Normal	N(0,1)	0.0033	2.9960	0.0050
T	T(2)	0.1755	40.5614	0.0170
T	T(3)	-0.0574	16.1598	0.0114
T	T(4)	-0.0321	8.3991	0.0100
T	T(5)	-0.0375	6.3354	0.0087
T	T(9)	0.0090	4.0830	0.0067
Uniform	U[0,1]	0.0058	1.8117	0.0034
Weibull	W(1,2)	1.9182	8.2040	0.0103
Weibull	W(2,1)	0.6101	3.1703	0.0054
Weibull	W(3,0.5)	0.1611	2.7139	0.0047
Cauchy	*C(0,1)*	*-0.2239*	*125.3620*	*0.0329*
F	*F(1,1)*	*13.1916*	*188.6945*	*0.0508*
F	*F(3,1)*	*13.1447*	*187.2275*	*0.0642*
F	*F(8,3)*	*5.1173*	*44.8449*	*0.0286*
T	*T(1)*	*-0.2680*	*126.1929*	*0.0248*

Table 5.1: *Skewness and Kurtosis Data*

5.4.3 Examining Non-Linear Functions

The strength of the HM statistic is independence from assumptions about the functional relationship between an input and output variable. That is, measures that are reflective of association are provided, regardless of the nature of that relationship. To illustrate this effect, coefficients of determination are simulated for three polynomial functions (linear, quadratic and cubic). This simulation is carefully controlled so that the previously defined coefficient of determination is an accurate estimate of the association between the input and output. In these instances, the appropriate model for each polynomial at each defined level of association (R-Sq) was evaluated using properties of variance (Cryer, 1986) using independent standard normal data. These coefficients of determination are plotted against the corresponding HM statistics below. For example to obtain a coefficient of determination of 0.7 (70%), the following functions were employed, using 8000 observations from standard normal data for x and e;

Linear: $y = \sqrt{7}.x + \sqrt{3}.e,$

Quadratic: $y = \sqrt{7}.x^2 + \sqrt{6}.e,$

Cubic: $y = \sqrt{7}.x^3 + \sqrt{48}.e,$

where x is the input variable, e is noise and y the output of interest. Both X and Y have to be standardised before the HM statistic is calculated.

One large sample size is used to best illustrate the strictly monotonic relationship between R-Sq and DHM. Figure 5.11 shows clearly that there is a direct relationship between the appropriate coefficient of determination (R-Sq) and the obtained HM statistic. This justifies the use of the HM statistic as a measure of the strength of association between variables for polynomial functions. Importantly, the HM statistic increases monotonically as the coefficient of determination increases for each type of function, suggesting that the HM statistic is probably suitable for use as a test for independence when the input-output function is unknown.

In Figure 5.11, the 95% confidence limits for independence for the coefficient of determination (0.0005) were established using Fisher's approximation (Johnson & Leone, 1964). The bounds for the half moon statistic (0.0018) were obtained previously.

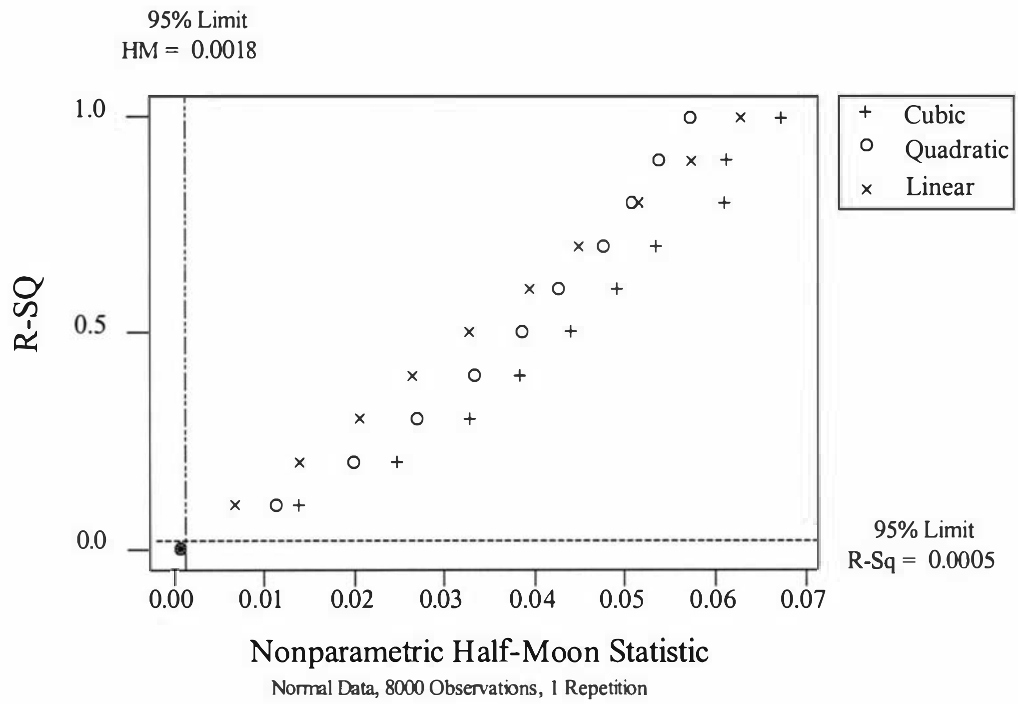


Figure 5.11: *HM Statistic versus R-Sq for Polynomial Functions*

The Type II errors of the HM and DHM statistics – the probability of failing to reject the null hypothesis of independence when the null hypothesis is false – are displayed in Figure 5.12 for both one-tailed (upper-tail) and two-tailed tests.

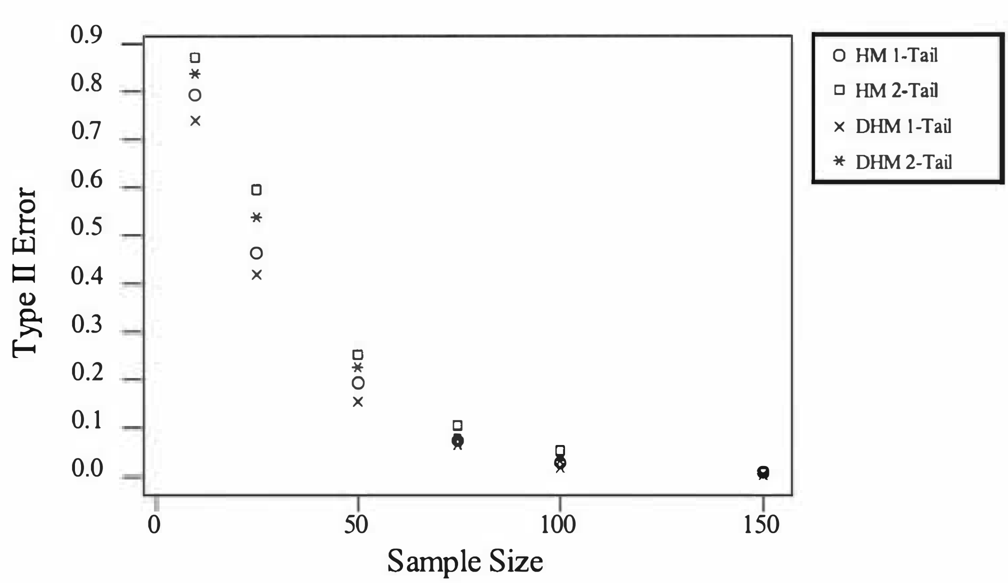


Figure 5.12: *Type II Errors for Non-Parametric HM Statistic by Varied Sample Size*

200 HM and DHM statistics obtained from the X-Y association determined by three polynomial functions (linear, quadratic and cubic) with the coefficient of determination set at 50%, calculated as described earlier in this section. This produced 600 observations for each level of sample size (10, 25, 50, 75, 100, 150), using standard normal data. The type II error was established given a 5% level of significance. Clearly, the type II error decreases as the sample size increases. Consequently, both the HM stand DHM statistics are powerful given large sample sizes.

Another test for independence is the chi-square test (see Smith, 1993) which uses categorical data. The DHM statistic is compared with a chi-square test of independence statistic as further validation (Smith, 1993). Observations are placed into one of four categories using the quartiles for both variables (input and output) to enable this test to be performed (thereby creating a four by four contingency table).

Simulated standard normal data was used to provide the data for figures 5.13-5.17, using three types of functions (linear, quadratic and cubic). For each of these functions ten different samples sizes were used (100, 200,..., 1000). 20 observations were recorded for the eleven different levels of the coefficient of determination (0.0, 0.1,..., 1.0). Relationships were derived using the same format as that described on page 183. That is, each point in the following graphs was obtained from the mean of 200 observations and contains a mixture of sample sizes.

The chi-square test statistics were scaled by dividing the obtained statistic by three times the sample size (3k). This gave the chi-square statistics a minimum of zero and maximum of one for every sample size which enables better comparison. A square root transformation was applied to the scaled data to better illustrate the relationship this data has with the coefficient of determination and the DHM statistic.

Figure 5.13 supports the findings from Figure 5.11, that there is a monotonically increasing relationship between the coefficient of determination and the obtained means for the DHM statistic.

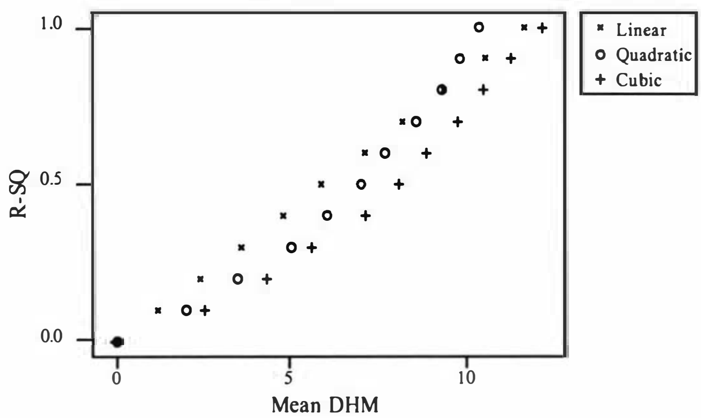


Figure 5.13: *Mean DHM Statistic versus R-Sq for Polynomial Functions*

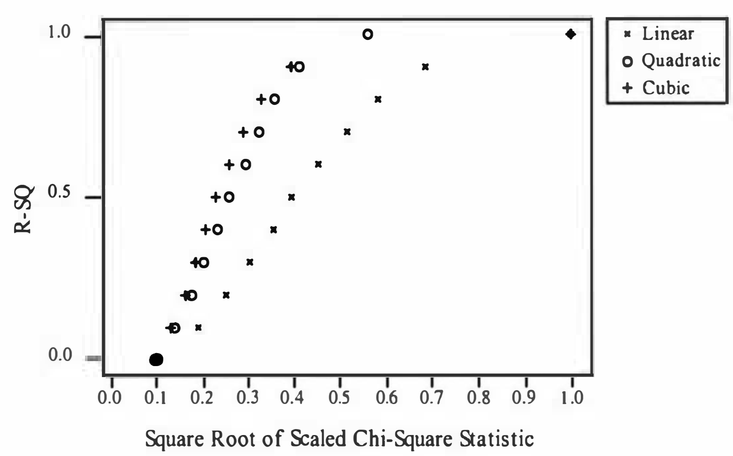


Figure 5.14: *Scaled Chi-Square Statistic versus R-Sq for Polynomial Functions*

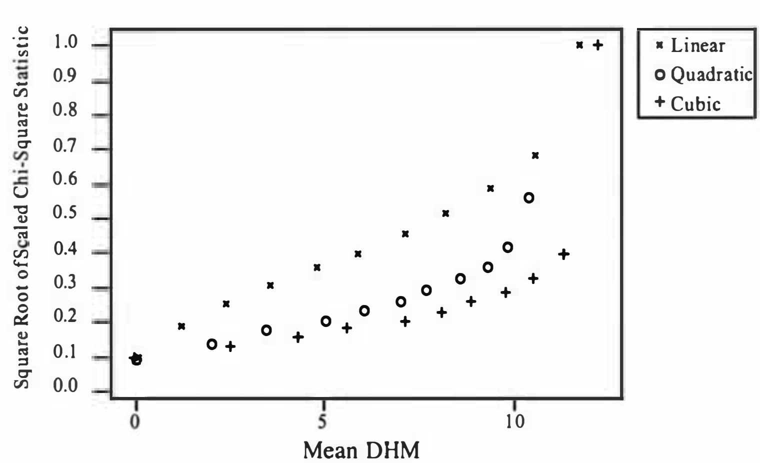


Figure 5.15: *Mean DHM Statistic versus Scaled Chi-Square Statistic for Polynomial Functions*

Figure 5.14 shows a similar relationship between the coefficient of determination and chi-square test of independence. However, there are different relationships for each type of function. This suggests that equivalent results cannot be obtained from the chi-square test of independence statistic unless the type of function is known. If the type of function is known, then the obtained statistic can be adjusted so that inferences can be made regarding the strength of the relationship between the input and output.

There is a monotonic increasing relationship between DHM and the Chi-Square test of independence statistic in Figure 5.15. This supports the use of the new DHM statistic as a test of independence. There is a different relationship visible for each polynomial function. Given that this effect was not evident in Figure 5.14, this suggests that the DHM performs better as a test of association than the Chi-Square test of independence statistic. That is the DHM statistics are not dependent on the type of function determining the relationship between the input and output variables (Figure 5.13).

The previous figures have shown only point estimates. The variability of the DHM and scaled Chi-Square test statistics with respect to the observed coefficient of determination for the linear case is shown in the following two figures. Both graphs use the 2200 observations for the linear case that generated figures 5.13-5.15 described previously.

The figures 5.16 and 5.17 show that in each instance the variability of the test statistics is minimal compared to the coefficient of determination. This means that the obtained test statistics provide good estimates of the strength of a linear input/output relationship, as defined by the observed coefficient of determination. This highlights the applicability of these techniques as tests of independence and as tests of association in the linear case. Unlike the DHM statistic, the Chi-Square test of independence responds differently to other types of polynomial functions. This means that the DHM statistic is more appropriate as a generic test of independence and association.

Polynomial functions are focused upon in this thesis as these are the most plausible relationships for describing performance. However, given that a function is a rule that assigns each input unit to exactly one output element (Fraleigh, 1994); it is assumed that the HM statistic will cope with most dependencies.

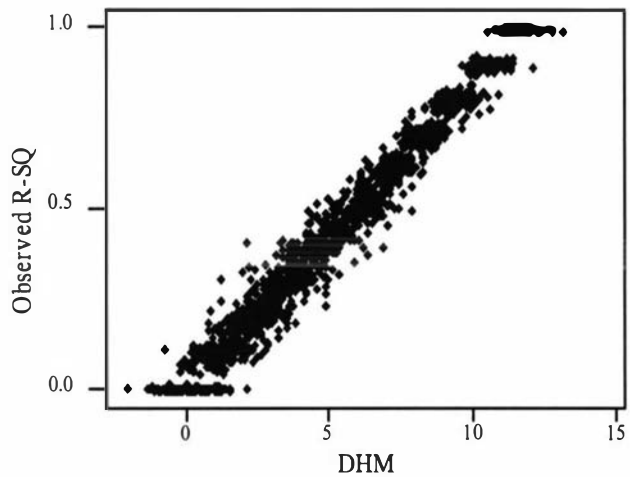


Figure 5.16: *DHM Statistic versus Observed R-Sq for Linear Functions*

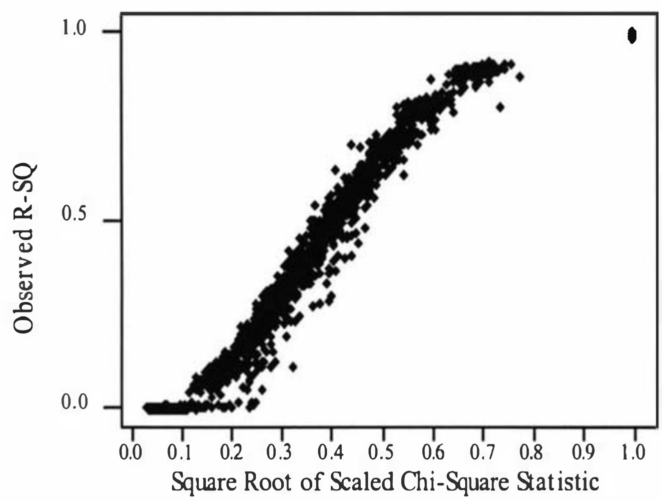


Figure 5.17: *Scaled Chi-Square Statistic versus Observed R-Sq for Linear Functions*

In spite of this, some types of functions perform poorly, due to the folded nature of the 'Half Moon' statistic. This occurs because the absolute value of the output variable is used in the calculation of the HM statistic. In particular, trigonometric functions with a minor element of noise have an increased tendency to be identified as an insignificant relationship. This is because the 'wave' can be lost in the standardisation and subsequent folding of the variables in the calculation of Z.

Polynomial functions tend to have larger sample standard deviations for Z than independence. This results in larger HM statistics. Conversely, circular and wave type functions tend to have smaller HM values when compared to independence.

The univariate non-parametric half moon statistic is a robust measure of association. As a test statistic it has good coverage (accurate probabilities for type I error) and high power. The next section expands the univariate statistic to a multivariate case to test for independence in a system of inputs with respect to an output variable.

5.5 Expansion to the Multivariate Case (Multiple Inputs)

The half-moon statistic is easily extended to the multivariate case where instead of testing for independence between one input and one output variable, several input variables are considered with respect to an output. Let *n* be the number of input variables. Equation (2) is replaced with (7) to enable calculation of the multivariate half-moon statistic (MHM).

$$u_j = \sqrt{r^2 - \sum_{i=1}^{n} x_{ij}^2} \tag{7}$$

for the jth observation and, as before, $z_j = u_j - |y_j|$ and $MHM = s_z - \bar{s}_z$.

Geometrically, instead of comparing points relative to the surface of a circle, comparison is relative to the surface of spheroid-type structures with the relevant dimensionality of the spheroid determined by the number of input variables to be considered. Because distance provides the basis for comparison (shown in Figure 5.1) increased dimensionality has no effect and the non-parametric MHM will still be one-dimensional as defined in (5). Increased dimensionality will affect the population standard deviation calculated under the assumption of independence unless r is adjusted. Expansion to the multivariate case assuming independence, identically distributed normal inputs (X_i) and $r^2 \geq x_i^2$ for all i, yields the following:

$$\mathrm{Var}(u_j) = \mathrm{Var}(\sqrt{r^2 - \sum_{i=1}^{n} X_i^2}) = E(\sqrt{r^2 - \sum_{i=1}^{n} X_i^2})^2 - E^2(\sqrt{r^2 - \sum_{i=1}^{n} X_i^2}).$$

When the inputs are independent, the variance becomes

$$\mathrm{Var}(u_j) = E(r^2 - X^2) - E^2(r - \frac{X^2}{2r} - \frac{X^4}{8r^3} - \ldots)$$

where $X^2 \sim \chi_n^2$ so $E(X^2) = n$, and $E(X^4) = (n^2 + 2n)$. Given that r is sufficiently large, the higher moments become negligible.

Taylor's expansion gives the following approximation for $\mathrm{Var}(u_j)$,

$$\mathrm{Var}(u_j) \approx (r^2 - n) - \left(r - \frac{n}{2r} - \frac{n^2(n+2)^2}{8r^3} \right)^2$$

$$\approx \frac{n}{2r^2} - \frac{n^2(n+2)}{8r^4} - \frac{n^2(n+2)^2}{64r^6}.$$

Such that $\mathrm{Var}(Z) = \mathrm{Var}(u_j) + \mathrm{Var}(|Y|)$. As before, the population standard deviation given independence, σ_Z, is the square root of the expected variance. Therefore, for the multivariate case given independence and normality,

$$\sigma_Z \approx \sqrt{\left(\frac{n}{2r^2} - \frac{n^2(n+2)}{8r^4} - \frac{(n^2+2n)^2}{64r^6} \right) + \left(1 - \frac{2}{\pi} \right)}. \qquad (8)$$

With increased dimensionality, the radius must be correctly defined such that equivalent values for HM and MHM statistics are obtained. This means that the radius r must be determined by the number of input variables, n, so that satisfying equations (6) and (8) provide equal variances for the HM and MHM statistics. This produces comparable HM statistics regardless of the dimensionality. This has important implications in the assessment of the interactions between input variables and combinations of input variables with an output variable. In particular, it means that the same critical values can be used for HM and MHM statistics, because the variability of the HM statistic will be equivalent regardless of n. For an initial (n=1) radius of 10, Table 5.2 lists the comparable radii for more than one input variable. For n inputs, an approximation for the required radius is $\sqrt{10^2 n}$.

n	*Radius*	*n*	*Radius*
1	10.000000	8	28.026990
2	14.124152	9	29.687046
3	17.276279	10	31.250366
4	19.923110	15	38.006570
5	22.245653	20	43.568617
6	24.336842	25	48.340410
7	26.251940	30	52.530719

Table 5.2: *MHM comparable radii*

The effect of the comparable radii is illustrated in Figure 5.18 using boxplots representing Z distances for various numbers of n simulated input variables. The simulated variables are normally distributed, independent and have a sample size of 1000.

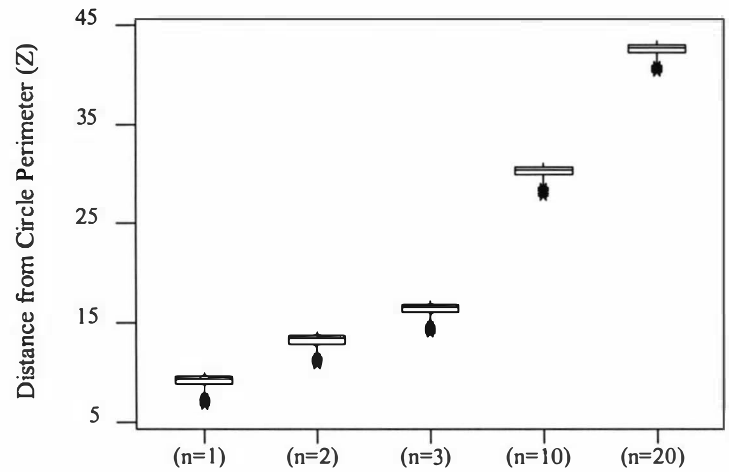


Figure 5.18: *Boxplot Illustrating Unscaled z Distances for n=1, 2, 3, 10 and 20*

Figure 5.18 shows that as the radius increases in accordance with an expansion in the number of input variables (see Table 5.2) the central location of the corresponding distribution increases. However, the central locality of the mean is irrelevant for the half-moon statistic which is defined in terms of σ_Z. Figure 5.19 displays the effect of re-scaling the measures by removing the expected value of Z, as defined earlier.

Figure 5.19 shows that Z is distributed similarly regardless of the number of input variables provided the correct radius is adopted. Figure 5.19 highlights the equivalence of variance, whilst Figure 5.18 shows that the dependency on the radius relates to the expected value of Z and not the variance of Z. This is true regardless of the initial (n=1) value of r, and it highlights the robustness of the HM statistic to the r-value chosen.

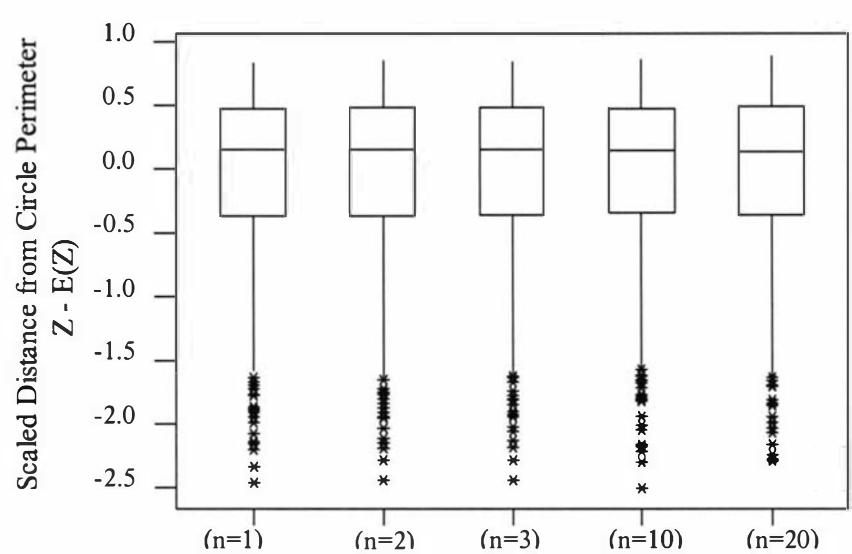


Figure 5.19: *Boxplot Illustrating Scaled z Components for n=1, 2, 3, 10 and 20*

When the distributions of the variables are unknown, the non-parametric procedure can be applied to the MHM statistic. However, further work is required to remedy skewness and kurtosis in the MHM procedure, possibly using a bootstrapping procedure.

Importantly, as the distribution of the MHM statistic is equivalent to that of the HM statistic, the same approximate distributional properties hold ($N(0, \psi^2)$). This enables tests of independence for the system to be conducted. Furthermore, the MHM uses the same adjustments for sample size as the HM, as well as the same critical values.

5.6 Implementation of MHM to Calculate Relative Influence

Obtaining the equivalent variance, σ_z^2, for varied number of input variables is crucial, as this allows combinations and interactions of variables to be assessed by removing variables and adjusting the radius accordingly. This is achieved by adopting comparable radii enabling the influence of input parameters or combinations of input parameters to be evaluated with reference to given standards regardless of the specified dimensionality.

This section will construct a relative influence measure for each variable, which provides two powerful results. Firstly, the use of the relative influence measure promotes parameter parsimony in statistical modelling.

Secondly, as the combined effect of input variables can be assessed, combinations of variables can be chosen to maximise the information explained, whilst minimising the number of variables required. Related to parameter parsimony is the idea of variable redundancy. Examining the combined relative influence for the set of input variables allows the redundancy of that system to be established. Identifying redundancies is another exciting area for development, which is currently being investigated and showing great promise.

The previous section detailed how the combined effect of the input variables on an output could be established by calculating the overall MHM. More importantly, by omitting each variable one at a time, the relative influence an input variable has on the output variable, with respect to the other variables, can be established. This is an important step in establishing the redundancy of the input-output system. Further, interactions of input variables can be assessed by omitting specific combinations of input variables. The calculations of the core components defining the MHM are outlined in Table 5.3. The nature of the formula denoting u_j in the defining calculation column of the table below is specified for computational ease. Once u_j is computed, calculation of the MHM proceeds as usual with $z_j = u_j - |y_j|$ and $MHM = s_z - \bar{s}_z$. Note MHM_{x1} = MHM with x1 omitted from the 'model' for Y and MHM is the MHM when all input variables are included in the model for Y. Also note that the 'model' is never specified because it does not need to be.

The overall MHM provides an indication of the strength of the relationship between the input variables and the output. Given that MHM exceeds a suitable critical value indicates that the collective influence of the input variables is responsible for influencing the output variable. By omitting one or more input variables, those variables, or combination of variables responsible for the disruption can be identified. This procedure is similar to the stepwise procedures used in stepwise regression and best subsets regression. As an illustration of the Table 5.3, given four input variables and an initial radius of ten, r is 19.923100 (see Table 5.1) for the overall MHM. Calculations proceed as before in determining the non-parametric half-moon statistic.

Variables of Interest	*n for relative radii*	*Defining Calculation for u_j*	*Relative Influence Measure*
ALL	n	$\sqrt{r^2 - \sum_{i=1}^{n} x_{ij}^2}$	-
x1	n-1	$\sqrt{r^2 + x_1^2 - \sum_{i=1}^{n} x_{ij}^2}$	$\frac{(MHM - MHM_{x1})}{MHM}$
x2	n-1	$\sqrt{r^2 + x_2^2 - \sum_{i=1}^{n} x_{ij}^2}$	$\frac{(MHM - MHM_{x2})}{MHM}$
Combinations of x1 & x2	n-2	$\sqrt{r^2 + (x_1^2 + x_2^2) - \sum_{i=1}^{n} x_{ij}^2}$	$\frac{(MHM - MHM_{x1 \times x2})}{MHM}$
Combinations of x1, x2 & x3	n-3	$\sqrt{r^2 + (x_1^2 + x_2^2 + x_3^2) - \sum_{i=1}^{n} x_{ij}^2}$	$\frac{(MHM - MHM_{x1 \times x2 \times x3})}{MHM}$

Table 5.3: *Calculation of Key MHM Components using a Stepwise Procedure*

The relative influence explained by a variable is obtained by scaling the omitted MHM by simply subtracting the omitted MHM from the overall MHM and dividing through by the overall MHM as demonstrated in Table 5.3. This measure can be used to rank the input variables in order of importance.

By omitting variables one at a time the impact of the chosen variables can be measured. The bigger the influence of an omitted input, the further the omitted MHM will be from the MHM calculated using all input variables, and the higher will be the relative influence measure. Conversely, the relative influence measure of unimportant variables will tend towards zero. It is possible for these values to be negative, due to the sampling approach and the relationship between the one omitted MHM and the overall MHM. Negative relative influence measures indicate obvious redundancy. This has important implications in the evaluation of redundancies.

This information for the omitted MHM can be coupled with the HM and DHM statistics to gain an understanding of the model properties. An important aspect is the calculation of the relative influence, which is an indication of system redundancy. This is currently being developed.

5.7 Application

To demonstrate the effectiveness of these new tools in an applicable environment, an exploration of the individual rugby player rating system developed in Chapter Three is analysed. In this example an attack KPI obtained using factor analysis for a cluster of individuals (midfield backs) is assessed with respect to 17 key input summary variables as indicated in Table 5.4 (described in Chapter Six). The data contains 279 observations relating to performances by starting midfield backs (second five eighths and centres) from the Super 12 competition, 2000.

Firstly, the univariate HM statistics can be used to interpret the output in the same way that loadings are used to interpret the factors in factor analysis (Hair *et al.*, 1995). Typically, if the absolute value of a loading (r) exceeds 0.5, then the variable is deemed to be having a significant influence upon that factor. In this instance, if a variable and the output are not independent as determined by the HM statistics, then that variable and the output are significantly associated.

As shown in Table 5.4, the data is skewed and has moderate values of kurtosis. Consequently, the HM statistic is unsuitable for use with this data and the DHM must be used instead. The p-values of DHM indicate that *Defence Beaten*, *Errors*, *Breaks*, *Running Metres*, *Kicks Caught*, *Passes* and *Laybacks* are the variables that are significantly associated with the output (Attack KPI). As expected, these variables are predominately attacking based variables, which is congruent with the previous interpretation of this KPI. The use of the RI measure allows the importance of the variables to be ranked. The most important variable is the number of *Breaks* (1) made in each game, followed by the number of *Laybacks* (2).

The multivariate aspects of the HM procedure provide an understanding of how the inputs collectively influence the output. Importantly, there is significant association between this set of inputs and the output (MHM = 0.09, p-value = 0.00). That is the input-output relationship is not independent. This was expected given that the attack KPI is created by the combination of the significant variables mentioned previously.

Variable	*HM*	*HM(p)*	*DHM*	*DHM(p)*	*RI*	*RI Rank*	*r*	*Skewness*	*Kurtosis*
Defence Beaten	0.0324	**0.00**	4.71	**0.00**	0.05	5	**0.63**	1.2	4.0
Errors	0.0235	**0.00**	3.35	**0.00**	0.03		0.21	1.1	4.2
Breaks	0.0933	**0.00**	8.63	**0.00**	0.21	1	**0.86**	1.8	10.4
Kicks	0.0013	0.79	0.12	0.90	-0.03		0.11	2.1	11.1
Running Metres	0.0587	**0.00**	8.44	**0.00**	0.12	3	**0.88**	1.0	4.2
Kicking Metres	-0.0087	0.08	-0.68	0.46	-0.05		0.04	2.8	15.9
Kicks Caught	0.0251	**0.00**	2.45	**0.01**	0.03		0.17	2.3	8.6
Breakdown Impact	0.0125	0.01	1.62	0.08	0.00		0.26	1.3	4.9
Passes	0.0510	**0.00**	5.52	**0.00**	0.09	4	**0.61**	1.5	7.1
Laybacks	0.0727	**0.00**	6.29	**0.00**	0.16	2	**0.76**	2.2	12.2
Tackles	-0.0035	0.47	-0.61	0.50	-0.04		-0.03	0.7	3.2
Missed Tackles	0.0071	0.15	0.91	0.32	-0.01		0.04	1.2	5.1
Tackle Assists	0.0033	0.50	0.37	0.69	-0.02		-0.02	1.5	6.6
Tries	-0.0011	0.82	-0.11	0.90	-0.03		0.11	2.3	7.4
Loose Ball Gained	-0.0047	0.34	-0.58	0.53	-0.04		0.16	1.4	5.4
Harassment	0.0108	**0.03**	0.87	0.34	-0.01		0.02	3.3	13.3
Infringement	0.0141	**0.00**	1.40	0.12	0.01		0.10	2.4	8.1

Table 5.4: *HM Procedure Output for Attack KPI by Summarised Key Inputs*

This application has shown the wealth of information generated by the HM procedure and demonstrated how these statistics may be used. The behaviour of the parametric properties of the raw HM statistic suggests that there is the possibility of understanding the functional relationship between associated variables. However, information relating to the type of function involved can be obtained graphically. Consequently, effective implementation of the HM procedure requires adequate graphical support.

5.8 Conclusion

This chapter has introduced important tools for exploratory data analysis. Measurements for the degree of association between input-output variables (HM, DHM, MHM and RI) were developed and explored. Due to the simple nature of the mathematics involved, a MINITAB macro was developed to automate the implementation of the entire HM procedure.

Initially the philosophy of independence was explored using empirical tools which led to the development of the half moon statistic. This was defined mathematically which promoted the evolution of a parametric test for independence to a non-parametric version.

The non-parametric half moon test statistic (HM) performs as well as the coefficient of determination in measuring the strength of a polynomial relationship between two variables, with the added bonus that a functional form for the input-output relationship does not have to be assumed. As a test for independence using standard normal input data, the statistic is approximately normally distributed (given the null hypothesis of independence) with mean, 0, and a standard deviation, ψ, that is determined by the sample size. Whilst the statistic does not determine the form of the model, it does indicate which input variables are associated with the output variable. This statistic was then modified to handle moderate skewness and kurtosis in the data (DHM). This statistic, assuming independence, is also normally distributed with a mean, 0, and a standard deviation, κ, which is dependent on sample size. This statistic outperforms the chi-square test of independence as a test of association.

The univariate HM was then expanded to consider multiple inputs (MHM), enabling the interactions of input variables to be assessed. This is made possible through the application of a dynamic radius, which is altered depending on the number of input variables included in the model. Consequently, the distribution of the MHM statistic under independence is distributed the same as the HM statistic given the same conditions, allowing the use of the MHM to measure input importance in a stepwise fashion. A relative influence measure (RI) provides an indication of the importance of input variables, or combination of important variables.

These heuristic tools have been shown to be powerful, flexible, and effective. To compliment the increased availability of powerful computing tools and algorithms in statistics there is the need for methods that recognise significant relationships quickly and effectively. Consequently the tools developed in this chapter are highly relevant to aid algorithmic statistical inference.

Much work is required in order to provide a sound mathematical foundation for these methods. In addition much empirical work is needed in order to determine when they are unreliable. This chapter has considered only polynomial functions and a few distributions. The impact of data with extreme kurtosis and skewness needs further exploration, although these effects can be reduced with transformation of the data prior to analyses. The effect of skewness and kurtosis in the MHM also requires attention. However, this research can be circumvented by the development of a bootstrapping procedure to obtain a test statistic for independence. Additionally, the relationship between the one omitted MHM and the MHM needs to be further investigated to provide a versatile heuristic test for redundancy.

Finally, through the adoption of an algorithmic approach, a powerful and versatile heuristic technique for exploratory data analysis has been created that has demonstrated its potential under numerous simulations.

Importantly, sufficient understanding of the half moon statistic has been obtained to proceed in this thesis and extract information relating to the important input variables in a dimension reduction sense, leading to greater transparency in Chapter Six and beyond.

Chapter Six

The Rebirth of the Eagle Rating Through Neural Networks

The purpose of this chapter is similar to that of Chapter Three, where multivariate techniques were applied to the rugby data provided by Eagle Sports in order to develop a measure of individual rugby player performance. In this chapter methods for improving the original Eagle Rating using a more applicable data set and more flexible modelling techniques, namely self-supervised neural networks and self-organising maps, are investigated. The fundamental issues relating to the data remain the same and will be reviewed briefly where required. However, the change to neural networks as a modelling tool brings different issues that need to be resolved and approaches justified to ensure the validity of the final models proposed.

The literature review in Chapter Four examined the relevant background of the methods to be employed. Due to the common implementation of neural networks, the methods adopted are examined widely in the literature. The most important aspect of this chapter, from an academic perspective, is the interpretation of the networks, employing the half moon statistic developed in Chapter Five. This promotes transparency in the networks, which is required to ensure the models obtained are contextual.

The data is pre-processed in the first section. This includes the identification of positional clusters. Then, in the second section, the data for each of the positional clusters is segmented so that for each distinct positional cluster the data can be compressed using dimension reduction networks.

Attack and defence variables tend to be uncorrelated with each other and with positional specific skill components. This enables the Eagle Rating to be evaluated in the third and subsequent sections using self-supervised neural networks, factor analysis, self-organising maps and a quartile method (hybrid) model. A comparison of the results obtained from Chapters Three and Six is examined in Chapter Seven.

Elements of this chapter have been published in the *Artificial Neural Networks and Expert Systems* conference proceedings (Bracewell, Meyer and Ganesh, 2001).

6.1 Data Pre-Processing

6.1.1 Sub-Group Identification

As mentioned in Chapter Three, it is imperative that any natural groupings within the data must be identified as part of the data-cleaning and pre-processing modelling phase. However, as shown at the conclusion of Chapter Three there are a number of limitations experienced employing clustering techniques to rugby data. The most notable is the exposure to match volatility and the effect of non-performance. Consequently neural network based clustering techniques will not be applied; instead clusters will be based on the groupings provided by experts in the early stages of Chapter Three. Thus to avoid the pitfalls associated with clustering procedures identified in Chapter Three, and impose structure to cater for job specificity; eight clusters are adopted based on expert opinion, as follows:

- Props
- Hookers
- Locks
- Loose Forwards
- Half Backs
- First Five Eighths
- Midfield Backs
- Outside Backs

The only difference from the initially adopted clusters is the addition of the openside flanker to the other back row forwards. This stems from comments made by respected All Black Selector, Peter Thorburn (2001) stating that there is no real difference in the loose forwards as there is so much switching around at lineout time. The only potential difference cited was the need for a Number Eight who could take the ball up off the back of the scrum.

Flexibility is touted as a strength of neural networks. However, this attribute could be lost in the modelling process, due to the dependency on the architectures adopted and the data presented. The duality within positions, such as hookers adopting a tight or loose role dependant on various issues relating to match constraints, prevents statistically based clustering procedures from being applied. Like positions need to be grouped together as this is how comparison occurs in the rugby environment, especially in a selection context.

However, because of positional duality and differences in perceived performance, strong natural clusters based on the positional clusters are lacking. Allowances must be made for this fact in the structure of the network. Ideally, the adopted architecture needs to allow for different styles of play to be compared simultaneously within a positional cluster. As a result more nodes may be required to cover the full spectrum of performance components expressed by first class rugby players. Finally, this means that a cautionary approach must be taken when collapsing the obtained performance vector into a single value. Unfortunately, an amount of flexibility is removed from the modelling process due to the comparatively reduced data set. As mentioned previously, there are few sampling opportunities presented in a sporting context. This lack of data imposes initial restrictions upon model flexibility and architecture. These ideas are addressed in section 6.1.2.

Preliminary investigation with a self-organising map revealed four distinct clusters, highlighting the dependence on specific task measures. A four by six array was used as this was deemed sufficiently large to allow clusters to be expressed in the map topology. Two dimensions were used rather than one to provide more flexibility for cluster formation. In general the dimensionality of the map is determined by the underlying structure of the data. In this case the two dimensions could be considered to represent attack and defence. The SOM was performed to confirm the Chapter Three cluster analysis, which did not provide any useful groupings. The default properties of SAS Enterprise Miner were used to obtain the map. The four clusters corresponded to Backs, Forwards, First Five Eighths and Hookers. For the first three clusters the underlying reason for division is identical to the explanation detailed previously in Chapter Three, namely attacking attributes, possession attributes and goal kicking. The attribute defining the hooker cluster is equally as obvious, defined by throwing in at lineout time. Due to the frustratingly obvious clusters proposed these results are unsatisfactory. It is necessary to identify different groupings of similar positions due to the different roles adopted on-field. Performing dimension reduction on each of the distinct positional clusters and scaling the obtained output for each of these clusters provides the basis for an effective rating system. That is, due to standardisation within the clusters, relative comparisons can be made between-positional clusters, within-positional clusters and between matches. Given the similarity of results it seemed redundant to explore the SOM for clustering any further.

Feasibly, three methods could be employed to reduce the n components representing ability obtained from the various methods to a singular value that represents match performance. These methods are factor analysis, dimension reduction neural networks and finally the perfection method developed in Chapter Three. A major limitation of the early Eagle Rating was that each factor was given equal weighting in the final calculation. Potentially this is a problem due to the insistence that some skills are deemed more important than others, such as defence and handling ability. This problem is largely avoided in this chapter as a direct benefit of the limited amount of data.

As each method is applied, it is evident that the method of final reduction is of lessened importance due to the pre-processing of the data, which reduces the latent variables that need collapsing into a single value. Before describing this aspect of the modelling process, the data limitations in a neural network context are discussed.

6.1.2 Data Limitations

As mentioned in Chapter Four, there exist heuristic rules for determining the number of nodes in the hidden layer. These rules also recommend that the size of the training sample, given the network structure, should have at least 5-10 observations per adjustable parameter. For a self-supervising network, where each layer has a bias, the number of adjustable parameters is calculated using the following formulation, $(m+f+1)(M_1+M_2)+m+f$, where m is the number of input variables, f the number of nodes in the bottleneck layer and M_1, M_2 correspond to the number of nodes in the mapping and de-mapping layers (Kramer, 1991). Multiplying this equation by five allows the recommended minimum number of observations in a training set to be established. For example, a potentially suitable network for modelling rugby performance with the architecture 30, 31, 5, 31, 30 with biases in all layers would require 11335 observations in the training set. If the training data set represents 40% of the data, then the total data set requires at least 28338 input patterns. To put this in perspective, assessing the first five eighth cluster for Super 12 rugby would require approximately 200 seasons of data to be considered. Considering the Super 12 started in 1996, the amount of data available is clearly deficient.

Obviously, this is a major problem as the sampling opportunities in sport are limited. In the early stages of this thesis, it was mentioned that the database is reasonably large, covering more than 1000 first class games of rugby. However, each of the different competitions and levels differ in style and intensity, reducing the suitability of some segments of the database. The data set could be expanded by considering a number of different components of the Eagle Sport Database by including:

- Different Years
- Different Competitions

However, this imposes new problems. As mentioned repeatedly, rugby is a dynamic and evolving sport. The style of play adopted in the 1996 Super 12 differs from 2000 due to the emphasis on mauling play. It is more than likely that 2002 will present a different style of play; given that as of June 2001 mauls are allowed five seconds to reconstitute. The art of mauling has been revived as a result, as this reduces the number of opposition players fanning out around the fringes of a maul, which means that the backs have fewer defenders to evade. Thus the emphasis on certain skills differs. Consequently, this limitation must be acknowledged in the modelling process.

The modelling data set could be expanded by incorporating different competitions. The focus is a rating system for top level rugby. Potentially, test rugby and domestic competitions could be added to the Super 12 data. But the intensity of the competitions impacts on the amount of football played, or involvement opportunities. This can be assessed by comparing the total number of sequences in each level of competition. The style of play poses an additional problem, as this shapes the input patterns presented to the model. This means that standardising variables with respect to the various competitions would be of little use.

Ultimately, the target audience and relevant rugby competition impacts on the data set adopted. If the NPC was used to augment the data set, a New Zealand style of play would be imposed on the model structure. This could be offset by including South Africa's Currie Cup. However, Australia has no comparable domestic tournament, due to the limited number of State sides. Essentially, the competitions included define the applicability of the model. A reduced scope is preferable initially. Other competitions can be tested later to assess the generalisablity (i.e. robustness) of the models employed. Realistically, Super 12 and Test data are the most applicable competitions as they have greater applicability to the commercial sector in New Zealand, whilst retaining a narrowed competition scope.

Therefore, the amount of data does not meet the recommended threshold. Using limited data examples exposes the models to over-fitting, through the tendency to fit the stochastic variation in the data, rather than the underlying functions (Kramer, 1991). Intuitively, this is not a problem if the data we have is fully representative of the population and that individual rugby player ability is imposed on match statistics. This is based on the premise that there exists some general, quantifiable, underlying structure for individual rugby player performance. It is the search for this quantifiable structure of performance that forms the backbone of this thesis. It is assumed that there exists an underlying function that governs rugby performance. Then this function will have a greater influence over the variables than the stochastic component. However, if the structure is non-existent or weak, then the stochastic variation in the data may dominate the models. Such domination will be manifested in the form of uncontextual models.

Increasing the amount of data available through re-sampling is not an option. Re-sampling by reselecting individual performances randomly is inappropriate as each match situation presents a unique set of circumstances that ultimately produce a 'unique' multivariate trace, which relates to individual performance given the specific constraints applied in a specified match. As the number of 'unique' input patterns is low, this will still expose the network to over-fitting. This effect will be worsened by re-sampling in that if a rare multivariate profile is repeated this will reinforce the illusion that that input pattern is not an anomaly. Further, this increases the exposure to the stochastic variation in the data set.

Only one option remains, the specification of a different model structure. This is a straightforward process. It involves identifying the key areas to be evaluated and the associated variables. This then reduces dramatically the number of variables involved, adjustable parameters and therefore the size of the required data set.

The above highlights a major deficiency in the application of neural networks to rugby data. Essentially, context is imposed on the modelling process through the use of a condensed set of variables chosen subjectively from a larger set of initial variables. However, this approach is beneficial in some regards. Discernible player involvement in a match is restricted, therefore some particular events occur rarely, such as tackles over the advantage line, tries scored, lineout throws lost by the hooker at the back of the lineout and so forth. The influence of these variables can be swamped by more common variables, such as passes received and given, which will fluctuate anyhow given match constraints and volatility. Thus by combining certain variables in a careful manner, the behaviour of the variables (including the rare variables) is stabilised and their characteristics can be identified in subsequent analyses. This gives the dimension reduction process a 'head-start', thereby reducing the computational time; and in the final stages reduces the cost of implementation, interpretation and, most importantly, assists in ensuring that the models are relatively robust.

Another important consideration is the inclusion of reserves and partial games in the modelling data set. In order to have sufficient applicable data, it is desirable for these data to remain in the same competition. Rather than just considering players who have been on the field a specified period of time, as was done in Chapter Three, the impact of substitutes, or partial games, must be considered. Substitutes play a crucial role in rugby today. Gone are the archaic times when reserves rarely participated. Thus the impact of partial games on perceived rugby performance must be acknowledged and catered for. Consequently individual contribution to the team is considered and thus the time spent on field is ignored. This presents another problem. It is possible for a reserve to come on for the last two or three minutes of a match. In attempting to identify the generic structure of performance it is best if such minimal match involvement is ignored. Also, in the data set used for this thesis it is virtually impossible to identify the position played from reserve jersey numbers. As a result only starting positions are reviewed.

With reference to ability, using partial games is not a problem as "Most players change their behaviour very little during a game (NZRFU, 2000, p. 3.2)." This statement infers that one does not have to watch the entire match to obtain an impression of ability, in a selection context. Or more importantly an individual does not need to participate in a full match for an impression of their ability to be obtained. As a result, both partial and full games are considered. Minimal match involvement is restricted as a player going off in the early stages of a match will be the result of either injury or foul play which is rare. Player positions are clearly identifiable. Finally, a clean data set that is representative of the performances of individuals in the Super 12, 2000 is obtained.

6.1.3 Data

The data used for the modelling of individual performance in this chapter is from the 2000 Super 12. This provided 69 games for analyses (66 matches in pool play, 2 semi-finals and a final). Comparable data from the previous year's competition is available but lacking key variables, such as speed to the breakdown, and subsequently not used.

Transformation of the data is not necessary as neural networks can learn non-linear functions (SAS, 2000). Although data transformation can improve generalisation, speed up training and reduce the undue influence on prediction, transformations can disrupt the relationship between variables, however, in this instance the disruption is minimal. As with the need for transformation in factor analysis, transformations are more a conceptual than a statistical necessity (Hair *et al.*, 1997). Consequently, in the initial stages untransformed data is used. If the modelling process is completely unsuccessful, then it can be revisited using transformed data. Untransformed data is easier for rugby observers to interpret. The rule extraction produced by the self-organising map is a prime example. Raw measures relate directly to match events, making the process easier to understand.

All starting individuals are included. This means that not all individuals sampled played an entire game of football. Since the introduction of the substitution rule in 1997 (Palenski *et al.*, 2000), substitutes, or 'impact players', have become a core component of first-class rugby. This is apparent in teams where there is sufficient depth, such that players are deemed of equal ability, and in teams where certain individuals have key skills that are suitable for key periods of a match.

As described earlier, it is necessary to reduce the large number of variables collated by Eagle Sports. To reduce the number of variables involved in the computation of the Eagle Rating, a set of new summary variables are created involving the following codes as outlined in Appendix A.

New Variable	*Original Variables*
Defence Beaten	AA + PA + AB + PB
Errors	AD + ATD + ATF + ATI + ATK + DCD + DCK + PD + PTD + PTF + PTI + PTK + KCD + KCK + AT + PT + KF + KD
Breaks	Count (AY, AYL, PY, PYL, PYLH)
Kicks	Count (AK, KY)
Running Metres	AY + AYL + PY + PYL + PYLH
Kicking Metres	AK + KY
Breakdown Impact	DB1 + DB2 + DB3
Tries	AS + PFS
Passes	AP + APT + PP + PPT
Layback	AL + PL
Lineouts Won	$JA^{+} + JW^{+}$ (7 + 7)
Lineouts Lost	JT^{+} (7)
Kicks Caught	DC + KC + AC
Tackles	DT + DTL + DOT
Tackle Assists	DA
Tackles Missed	DM
Infringements	I^{+} (23)
Loose Ball Gained	DG
Harassment	DI + DD

Table 6.1: *Summary Variables for Modelling*

$^{+}$Note: These codes represent those original variable codes detailed in Appendix A beginning with the given letter(s). The subscript provides the number of variables for each broad code.

Table 6.1 summarises 94 of the original variables collected by Eagle Sports into just 19, significantly reducing the adjustable parameters necessary in the neural networks, thereby reducing the effect of over-fitting. Additionally, fewer variables require shorter computational times. As the summarised variables describe the main events on the rugby field a minimal loss of information is incurred.

Further, these variables are still highly correlated as shown in the Table 6.2, where the correlation for the majority of the variables exceeds 0.3 (Hair *et al*., 1997), implying that a multivariate analysis is suitable for dimension reduction due to the presence of latent structures within the data set. The correlations listed in Table 6.2 are calculated from all starting individuals from the Super 12, 2000. That is all those with jersey numbers ranging from 1 to 15.

The limitations of neural networks and sports data force a change in the modelling approach applied in this section. Due to insufficient data to meet the suggested criteria of 5-10 observations per adjustable parameter, the data is split to match the key concepts of attack and defence. This continues with the theme that this rating system is a measure of generalised performance and not specialised performance.

Combative object team sports have two key aspects: attack and defence. That is when a team has the ball its' objective is to score points. Conversely when the team lacks possession the goal is to prevent the other team scoring. In rugby, individuals are expected to participate in both aspects, and accordingly, are judged on the performance of both. Therefore the estimate of ability is obtained from the combined performance on both these attributes over a number of matches. As a result this natural division in the variables is used to streamline the modelling process in order to combat the deficiencies of neural networks in relation to lack of data. The key components to be assessed are:

1) Attack
2) Defence

In addition there are

3) Positional Cluster Specific Skill Components

Variables	*Errors*	*Breaks*	*Kicks*	*Running Metres*	*Kicking Metres*	*Breakdown Impact*	*Passes*	*Lay-backs*	*Lineouts Won*	*Lineouts Lost*	*Kicks Caught*	*Tackles*	*Missed Tackles*	*Tackle Assists*	*Tries*	*Loose ball Gained*	*Harassment*	*Infringement*
Dfnc Beat	0.19	0.42	0.17	0.67	0.15	-0.12	0.12	0.23	-0.11	-0.08	0.15	0.00	0.04	-0.05	0.26	0.25	0.06	-0.03
Error		0.34	0.25	0.36	0.24	0.00	0.20	0.15	0.01	-0.05	0.13	0.08	0.13	0.04	0.11	0.26	0.11	0.02
Brks			0.41	0.77	0.39	0.04	0.40	0.56	-0.08	-0.03	0.16	0.17	0.18	0.10	0.19	0.34	0.09	0.07
Kcks				0.37	0.98	-0.20	0.39	-0.05	-0.14	-0.08	0.19	0.01	0.20	-0.02	0.11	0.28	0.09	-0.06
Run Mtr					0.35	-0.12	0.27	0.41	-0.14	-0.09	0.26	0.05	0.12	-0.03	0.37	0.41	0.14	0.01
Kck Mtr						-0.20	0.36	-0.05	-0.14	-0.08	0.19	0.01	0.20	-0.02	0.10	0.27	0.09	-0.06
BDI							-0.18	0.29	0.26	0.08	-0.06	0.34	0.06	0.32	-0.10	-0.02	0.02	0.12
Pass								0.07	-0.05	-0.07	0.10	0.14	0.16	0.05	0.06	0.24	0.07	0.02
Lay									0.21	0.06	0.07	0.25	0.12	0.19	0.04	0.23	0.03	0.11
Line Won										0.10	0.06	0.12	0.02	0.16	-0.05	-0.03	-0.02	0.07
Line Lost											-0.03	0.05	0.00	0.04	-0.04	-0.04	-0.01	0.00
Kick Cght												-0.01	0.07	-0.02	0.12	0.18	0.03	-0.03
Tckl													0.27	0.48	-0.01	0.08	0.06	0.13
Miss Tckl														0.22	0.00	0.09	0.04	0.01
Tckl Ast															-0.07	0.02	-0.01	0.12
Tries																0.13	0.09	-0.01
LB Gain																	0.07	0.04
Hrst																		0.00

Table 6.2: *Correlation Matrix of Summarised Variables*

This base of core components can easily be added to by incorporating information relating to special areas, such as scrummaging, goal kicking, and so forth. Subjective measures describing player judgement and execution of skill can also be added to the base components, to expand the depth of coverage. This can be achieved by using the perfection method to condense all relative variables to a single value. But first it is necessary to establish a suitable foundation. In this thesis analysis will focus on the base components of attack and defence. The other components are to be considered in the future.

Attack can be thought of as tasks performed when an individual has the ball. Conversely, defence is the events that occur without the ball. The positional cluster specific skill components relate to either kicking (backs), lineouts (forwards, excluding props) and scrummaging (props). As mentioned earlier, if statistics are to be used to rate individual performance we must accept what is measured. As it is unrealistic to incorporate technical measures (technique, positional play) into a statistical analysis because of clouded cause and effect (who is responsible for the scrum going backwards?), the rating system must remain fairly general. This is why the primary focus is aimed at attack and defence, as these are the primary roles of individual involvement.

6.1.4 Defence/Attack Philosophy

These key components of performance are comprised of the following summary variables which are described in Table 6.1.

Attack	*Defence*	*General Positional-Specific Components*
Defence Beaten	Breakdown Impact	Kicks (Outside and Midfield Backs)
Breaks	Kicks Caught	Kicking Metres (First Five, Halfback)
Running Metres	Tackles	Lineouts Won (Loose Forwards, Locks)
Tries	Tackle Assists	Lineouts Lost (Hookers)
Passes	Tackles Missed	Scrummaging (Props)$^{+}$
Layback	Loose ball Gained	
Errors	Harassment	

Scrummaging+ does not appear in Table 6.1 as it is a team statistic. Infringements are not included at this stage as this can be considered a 'grey' area. To adequately incorporate infringements into an analysis, the context of the infringement must be considered. That is did the individual infringe because;

1. they did not know the rules properly?
2. were forced to transgress the laws by a member of the opposition?
3. infringed deliberately to prevent the opposition from scoring a try and content to risk a possible three points rather than a possible seven?
4. Retaliated to foul play from the opposition?
5. Other mitigating circumstances?

Sometimes an infringement can be seen as the correct course of action. For instance a back line lacking pace in the outside backs may choose to be off-side in close to the set piece/ruck/maul thereby putting extreme pressure on the opposing inside backs so that the opposing outside backs are not given enough space to defeat their team mates out wide. Or similarly consider an example where an individual is isolated and surrounded by the opposition; the individual throws the ball deliberately into touch to prevent the opposition scoring a try. It is cheating, but it is for the benefit of the team. All these aspects must be considered and put into perspective. Thus it is safer to leave infringement values out. To make the use of infringement data more applicable, the actual cost of the infringement needs to be built into the coding procedure. Such as denoting the total territory lost, or the points conceded as a result.

Generic constraints must be considered. If events can be considered favourable or unfavourable depending on match circumstances (duality), then this impacts on the interpretation of the output. Consequently, the impact of such a variable should be reduced. The example illustrated by infringements incurred supports the occurrence of variable duality. Whilst infringements are fundamentally poor there are occasions where the best option available to an individual is to infringe.

The codes specified for attack and defence meet the criteria of duality. Kicking is not included with the attack variables, or with the defence variables due to the lack of quality measures available in this data set. Included as a positional skill cluster, it can easily be modified to cater for change in game structure or include additional information that highlights the implications and context of a given event.

As the positional cluster specific skill components refer to only one variable for each positional cluster, there is no need to incorporate these aspects into an analysis. Therefore these variables need only to be standardised with respect to the specified positional cluster. The attack, defence and positional variables can be combined into a robust relative rating system.

6.2 Dimension Reduction Neural Network

In this section, only one positional cluster (*midfield back*) is explored to illustrate how the techniques described throughout this thesis are applied, without compromising the confidential aspects of the Eagle Rating system. This provided 279 observations from starting performances midfield backs in the 69 matches played in the Super 12, 2000. The variables used are the summary variables described in section 6.1.4. Firstly neural networks are examined followed by a comparable revision of factor analysis. Self-organising maps provide the last of the pure modelling procedures before quartile method (hybrid) models are explored, taking the best of each of the methods to provide a robust, contextual model that is easy to implement, producing the Eagle Rating.

At this stage kicking variables are not incorporated because this is regarded as a rather complicated specialist skill which needs to have performance duality identified prior to inclusion. For some positions the ability to kick is more important than others and it can be difficult to define what constitutes a good kick. Big is not always best. Accuracy is important. However, the outcome of each kick needs to be assessed, for sometimes punting can just be a reckless way of turning the ball over. At other times a kick provides the opportunity to retain position, gain territory, apply pressure or even create scoring opportunities. Kicking is essentially two highly correlated variables (number of kicks and kicking metres), which will be dealt with separately for relevant positions. As specified by the NZRFU's positional responsibilities which are replicated in Appendix C, the ability to kick is only specified for inside and midfield backs (9, 10, 12 and 13). In the early stages of this section a generic performance rating template is adopted. Specialist skills such as kicking and set-piece play are discussed briefly later, with regard to the limitations of a statistical collection procedure in section 6.5.

Firstly, neural networks are examined. As discussed in Chapter Four, the architecture of a neural network must be determined by trial and error. For both the defensive and attacking variables, seven input variables are available. The generic details described below are applicable to all the networks created for the other positional clusters. Essentially the philosophy described will suffice for quantifying individual performance in many other team sports.

A four layer self-supervising network with an input layer (x), mapping layer (y), bottleneck layer (z), de-mapping layer (y') and input reconstruction layer (target layer) (x') is used. This structure will be given as (x, y, z, y', x') in this chapter. The bottleneck layer will be referred to as the 'output layer', as this is the component of interest. A hyperbolic tangent activation function, which is closely related to the bipolar sigmoid, is employed in the mapping and de-mapping layers. A linear combination function acts in each layer to sum all inputs. All layers have bias, apart from the 'output' layer, which has no activation or bias impacting on the data stream. Bias was omitted on the 'output' layer as it was found that this generally caused an increase in the average error (mean square error) of the network which was used to define the optimal network when it was included. To improve generality, early stopping is implemented.

From the Eagle Sport database, data from 279 starting midfield performances in the 2000 Super 12 was obtained, covering 49 individuals. All these starting performances are included in the primary data set. A 40%/30%/30% random partition is applied to obtain the training/validation/testing data, providing 112 individual performances in the training data set. Using back-propagation the data-set was trained to minimise the average error in relation to the input reconstruction layer. It is the average overall error that is used to select the most suitable architecture for this thesis. The results for defence and attack each obtained from the corresponding seven summary variables outlined in Section 6.1.4 are displayed in Tables 6.3 and 6.4 respectively, with the optimal networks denoted in bold type. The network with the lowest average error is defined as the optimal. The average error is equivalent to the mean square error with the units relative to the data in the bottleneck layer.

Summarising Tables 6.3 and 6.4, the defensive variables a (7, 4, 2, 4, 7) network produced the lowest overall average error for the midfield data. Given this structure, there are 87 adjustable parameters equating to 1.29 observations per adjustable parameter, falling well short of the recommended five to ten observations. Therefore this model is susceptible to over-fitting.

The attack network for midfield backs requires a (7, 7, 2, 7, 7) structure to minimise the overall average error. This translates to 0.76 observations per adjustable parameter, clearly exposing the network to over-fitting.

For each network, a linear based network was also applied, by removing the mapping and de-mapping layers and the non-linear activation functions leaving a two-layered network. The average error produced for these networks gives an interesting insight into the structure of the rugby data. For the midfield defensive attributes a (7, 2, 7) network is the best of the linear networks, producing an overall average error of 1.09, only slightly higher than that of the four-layered network. However, for attack three 'output' nodes in the bottleneck layer are required to approximately match the average error from the four-layered network with only two 'output' nodes. This suggests that there is some non-linearity in the data that can be explained best as a non-linear principal surface rather than a principal plane.

Two 'output' nodes in the bottleneck layer were found to provide the optimal network in terms of average error. Whilst an increase in the number of bottleneck nodes occasionally decreased the average error of the network, expansion beyond two nodes would place increased pressure on the methods required for collapsing the latent data into a single value representative of individual performance. Additionally, extra nodes in the bottleneck layer create more adjustable parameters. Finally, two nodes are viewed as the optimal number of nodes to minimise the information lost in the subsequent steps required for the creation of a single valued rating system.

Having obtained the optimal architectures for each positional attribute network, it is then necessary for interpretation to take place. Two methods are applied, one graphical and the other numerical, for interpreting the obtained 'output'.

Input Layer	*Mapping Layer*	*Bottleneck Layer*	*Average Error*	*Adjustable Parameters*
7	**4**	**2**	**0.9222**	**87**
7	5	2	0.9296	107
7	7	2	1.0312	147
7	-	2	1.0859	35
7	-	3	1.1498	49
7	6	2	1.3924	127
7	-	1	1.8144	21
7	8	2	1.8376	167
7	7	1	1.8413	133
7	8	1	1.9973	151
7	5	1	2.2778	97
7	3	2	2.4466	67
7	3	1	2.8280	61
7	6	1	3.0008	115
7	4	1	3.0604	79

Table 6.3: *Optimal Architecture for Midfield Defence Self-Supervising Neural Networks Based on Average Error*

Input Layer	*Mapping Layer*	*Bottleneck Layer*	*Average Error*	*Adjustable Parameters*
7	**7**	**2**	**4.4593**	**147**
7	-	3	4.6919	49
7	5	2	5.3152	107
7	8	2	6.7703	167
7	3	2	7.1929	67
7	3	1	7.3232	61
7	8	1	8.7255	151
7	7	1	9.3138	133
7	6	1	9.8108	115
7	6	2	9.8384	127
7	4	2	10.167	87
7	5	1	20.009	97
7	4	1	26.277	79
7	-	2	52.8984	35
7	-	1	150.684	21

Table 6.4: *Optimal Architecture for Midfield Attack Self-Supervising Neural Networks Based on Average Error*

The graphical interpretation (Figures 6.1-6.4) reveals the sensitivity of the 'output' to each input variable. The plots show the sensitivity of each variable through the rate of change each variable undergoes as the 'output' changes. Mathematically the rate of change refers to the derivative of the function specifying the relationship between the input variable and the output variable. This is the version of sensitivity analysis specified by Haykin (1999). As a further reinforcement, the coefficient of determination is also included, assuming that any relationships are linear. Clearly, some of the relationships cannot be assumed to be linear as shown between the first defensive node and breakdown impact in Figure 6.1. However, these relationships can be approximated by linear functions. This has a critical impact on how the principal lines/curves and planes/surfaces are constructed. This effect is illustrated conceptually in Figure 6.5

Figures 6.1-6.4 show graphically which input variables have the most influence upon the obtained output. Breakdown impact has the strongest relationship in Figure 6.1. In Figure 6.2, it is the number of tackles made per game that has the greatest impact. The number of breaks per game is the most influential input variable in Figure 6.3, with running metres, passes and laybacks also being highly correlated with the output. Finally, Figure 6.4 shows that the output variable is most sensitive to running metres, with the number of breaks and the number of defenders beaten also having an influence on the output variable.

The following graphs (Figures 6.1-6.4) show clearly which variables the 'output' is most sensitive to. For the first defensive node, breakdown impact has the greatest influence. The number of tackles made determines the second node. This is a good result. Contextually, breakdown impact is important, as it is an indication of how often an individual has become involved at a breakdown. Realistically, if a midfield back has reached a breakdown, they will be in the first few individuals there. This is because it is not their specified role (see Appendix C). Ball security is crucial and, given the opportunity, an individual from any position must act as a loose forward and either secure possession or attempt to steal it from the opposition.

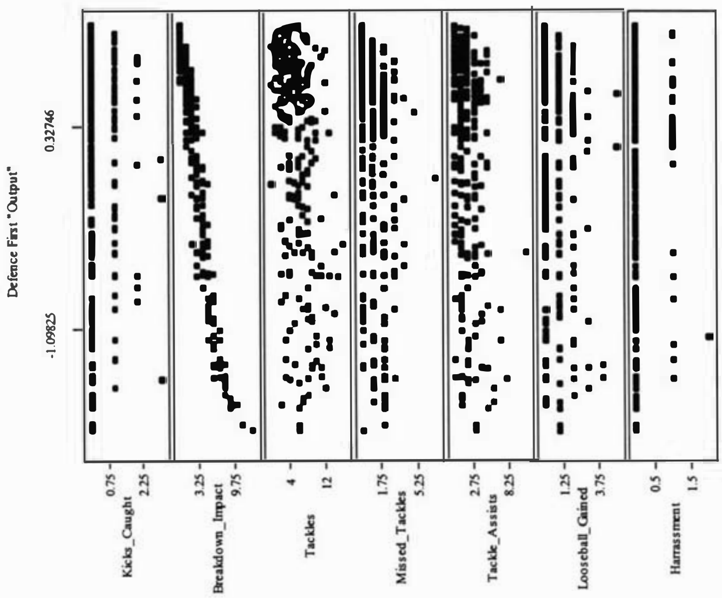


Figure 6.1: *Draftsman Plot of First Defensive 'Output' Node and Input Variables*

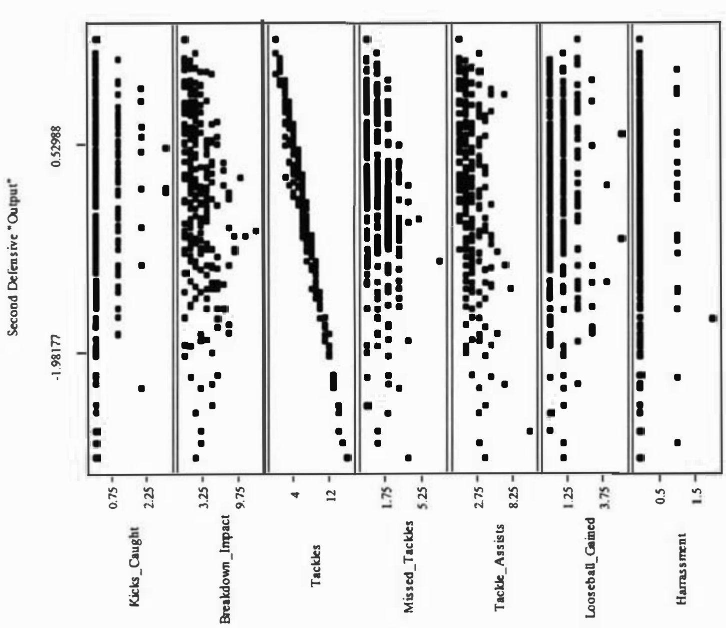


Figure 6.2: *Draftsman Plot of Second Defensive 'Output' Node and Input Variables*

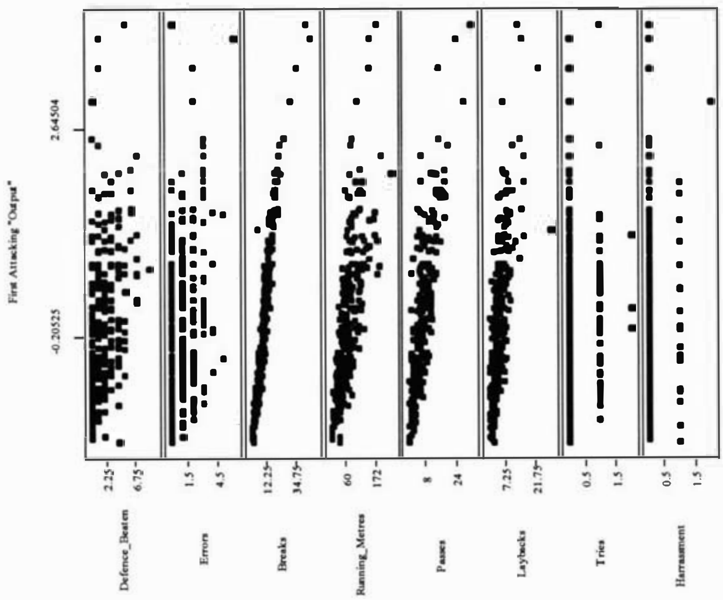


Figure 6.3: *Draftsman Plot of First Attack 'Output' Node and Input Variables*

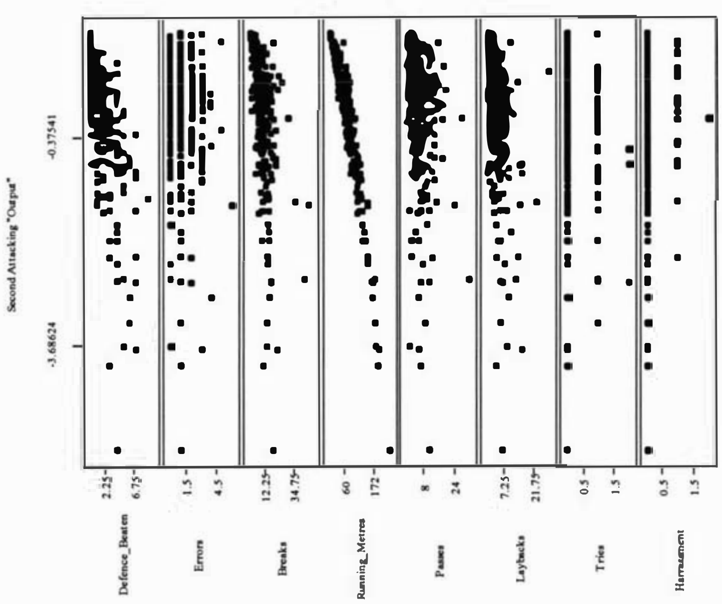
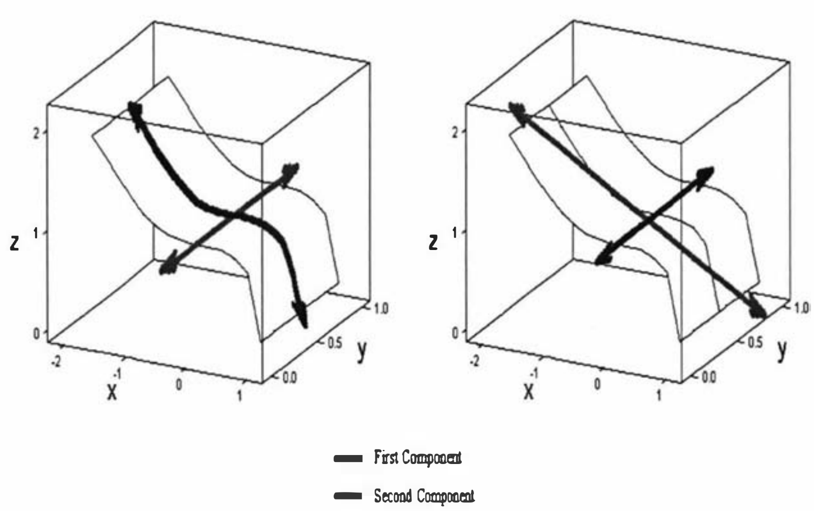


Figure 6.4: *Draftsman Plot of Second Attack 'Output' Node and Input Variables*

Secondly, tackling is important as specified by the positional responsibilities, contributing the most towards shaping the impression of defensive ability. Therefore by adding these nodes together a contextual measure for defensive performance is obtained. For some positions, outside backs for instance, an important component of defensive performance is positional play. This is where they position themselves on the field, to either deter the opposition from adopting a particular option, or to be in a position to interact directly with an adopted option. This cannot be measured statistically, as the generation of such information comes at a prohibitive cost. However, if transmitters of some type could be attached to players, so that their movements could be collated automatically, then such information could be built into a model. Similarly, video capture techniques may also provide useful spatial data.

Figure 6.5 illustrates the conceptual composition of a hypothetical principal surface in three dimensions. This principal surface consists of two attributes, the first (x) is based on a cubic function and the second (y) is linear in nature. Given that the principal surface is dominated by the cubic function, a non-linear dimension reduction method (a) will correctly identify the inherent structure, as shown on the left. Conversely, when a linear technique is applied, the non-linear component must be approximated by a linear function (b). In this hypothetical case, the second component is then deemed the most important because it represents the strongest linear relationship.



a) Non-Linear Case b) Linear Case

Figure 6.5: *Implication of Principal Planes and Surfaces*

Limiting relationships have a large influence in rugby data. That is performance is bounded by the amount of activity an individual can participate in. This can be viewed as the individual's performance ceiling, which is governed by match structure, volatility and the individual's ability.

Having identified graphically the impact of the input variables upon the resultant output, it is useful to compare the numerical interpretation with the graphical interpretation. This presents the first opportunity to apply the half-moon statistic in its specified role. The results are displayed in Table 6.5 which uses the DHM statistic to measure the association between the obtained outputs for the (7, 4, 2, 4, 7) neural network, which was obtained from the midfield backs defensive summary variables, and 17 of the 19 summary variables for Midfield backs (lineouts won and lost are not included as these are inappropriate variables).

Variable	*First 'Output'*			*Second 'Output'*			*Combined*		
	DHM	*p*	R^2	*DHM*	*p*	R^2	*DHM*	*p*	R^2
Defence Beaten	-1.18	0.20	0.02	-1.11	0.23	0.02	-1.04	0.25	0.02
Kicks Caught	-0.77	0.40	0.00	-1.23	0.18	0.00	-0.03	0.97	0.00
Errors	-0.85	0.35	0.00	-0.81	0.38	0.00	-1.29	0.16	0.00
Breaks	-0.74	0.42	0.07	-1.24	0.18	0.01	-0.89	0.33	0.04
Kicks	0.07	0.94	0.00	-1.31	0.15	0.00	-0.44	0.63	0.00
Running Metres	-1.08	0.24	0.00	-1.29	0.16	0.00	-1.31	0.15	0.00
Kicking Metres	-0.34	0.71	0.00	-1.22	0.18	0.00	-0.22	0.81	0.00
Breakdown Impact	8.63	**0.00**	**0.88**	-0.27	0.77	0.09	5.64	**0.00**	**0.57**
Passes	0.54	0.56	0.10	-0.06	0.95	0.04	0.39	0.67	0.09
Laybacks	-0.04	0.96	0.06	-0.46	0.61	0.01	0.09	0.92	0.04
Tackles	1.29	0.16	0.15	11.02	**0.00**	**0.96**	7.33	**0.00**	**0.54**
Missed Tackles	-1.37	0.13	0.05	-0.24	0.79	0.11	-0.50	0.58	0.10
Tackle Assists	1.19	0.20	0.08	3.72	**0.00**	0.15	3.39	**0.00**	0.14
Tries	-0.67	0.47	0.00	0.98	0.28	0.00	0.37	0.69	0.00
Loose Ball Gained	0.42	0.65	0.04	0.13	0.89	0.01	0.08	0.93	0.03
Harassment	0.17	0.85	0.01	1.43	0.12	0.01	0.80	0.38	0.01
Infringement	-0.14	0.88	0.00	0.03	0.97	0.00	0.16	0.86	0.00

Table 6.5: *Numerical Measures for Defensive Parameter Importance*

The coefficient of determination is included assuming that the relationship between the 'output' and the specified input variable is approximately linear. The graphical interpretation of the sensitivity of each 'output' indicates that this is a fair assumption. Statistically significant variables are denoted in bold type. As the data is skewed, the distribution corrected HM statistic (DHM) is used to provide p-values, with a 5% level of significance adopted. The cut-off for the coefficient of determination is 0.25. The graphical interpretation of the patterns between output and defensive variables suggested by the draftsman plots (Figures 6.1-6.4) is reinforced by the numerical interpretation provided by the half-moon statistics in Table 6.5. The first defensive node is clearly determined by breakdown impact and the second defensive node is determined by the number of tackles made per game. The DHM statistic also includes the number of tackle assists as an influential variable in the second output. Tackles made and arrival to the breakdown are indicative of a midfield backs' involvement off the ball.

The nature of the neural network equation specifying the calculation of defensive performance can be written explicitly. However, this involves 87 terms (adjustable parameters, or weights) and is cumbersome. Therefore, the defensive performance of midfield backs is specified by merely adding the first defensive 'output' and the second defensive 'output'. This approach is justified, as it is an approximation of the perfection method. However, to ensure that a high score relates to a good performance, this performance measure needs to be multiplied by negative one. That is, $D_i = -(d_{1i} + d_{2i})$, where d_{1i} is the first output for the ith observation and d_{2i} is the second output for the ith observation.

In the final "Combined" column of Table 6.5 calculated using D_i, tackles and breakdown impact are the most influential variables in determining defensive performance, as determined by the self-supervising network. These results are contextual and although there is the chance that over-fitting occurred due to the lack of input patterns per adjustable parameter, the final result is something which can be sold to the rugby public.

In terms of the positional responsibilities defined by the NZRFU (1991) and replicated in Appendix C, breakdown impact reflects high work rate in support play. That is the individual has arrived early at the breakdown to either support his own ball carrier to ensure possession is retained, or to disrupt possession for the opposition. This indicates an understanding and reading of the game to some extent. That is a midfield back should only commit to the breakdown if their presence is beneficial. Obviously to do so, the back will have to be amongst the first few there; otherwise it is better left to the forwards who are better equipped with the necessary skills.

Additionally it is necessary for a midfield back to be a strong, accurate, intimidating tackler. These adjectives relate to turnover tackles, tackles over the advantage line and missed tackles. All are incorporated into this model – turnover tackles and advantage line tackles are grouped with standard tackles, as the occurrence of such events is relatively rare. Thus no distinction is made other than the quantity of tackles. The number of missed tackles refers to accuracy, but has no impact on the neural network as indicated by the low half-moon statistic. This suggests that this variable has little impact in determining the defensive performance of midfield backs. This is easily argued against, for it is missed tackles that allow the opposition to breach the tackle line, thus giving their attack the chance to get in between the defensive line and the try line. However what the neural network implies is that the number of tackles made, assistance provided in the tackle, and breakdown impact, are more important than the number of tackles missed. Again the rarity of this variable's occurrence and the seemingly random involvement restricts this variable from having an impact. This is where a team contribution statistic is more suitable as it relates individual performance to team performance.

The neural network model for defence is adopted because it is a measure of successful defensive involvement. This makes it appropriate for use in a general rating system for performance, which can be incorporated into inferences about ability.

Further support comes from the selection protocols for midfield backs specified by the NZRFU (1991, p196). The following aspects are the key criteria for which selectors of rugby sides focus on in selecting an individual:

- Passing Skills
- Defensive Ability
- Kicking Skills
- Tactical Understanding (Option Taking)
- Support
- Ability to Make Breaks
- Work Rate

At this point the results obtained from the factor analysis are put into context. Work rate can be established from the combined impact of the attack and defence indices. Defensive ability relates in some part to the positional responsibilities defined by the NZRFU. However, it is a generic defensive measure that is created in this section. Passing skills and the ability to make breaks are incorporated into the attack index. The value of passing skills is inferred as a result of the successful completion of the passing task. As mentioned earlier, kicking skills are not covered at this stage, due to insufficient evidence regarding the performance of the task, and the fact that kicking is described by only two variables. Support refers partially to breakdown impact, but also to the re-involvement of an individual in a given phase. This is not measured, and would require the manual extraction of data from the sequence of play data. Such an extraction would be time consuming and costly. Whilst the coding structure is in place for recording the options taken, the data is not available. Nonetheless, the majority of criteria assessed by rugby selectors are covered by the raw indices described in this chapter, highlighting their relevance.

On a technical note, networks using a logistic activation function yielded average errors that were much larger than those obtained with the hyperbolic tangent activation function. Adding bias to the 'output' nodes did not improve upon the model, and so were omitted.

Progressing from the defensive attributes to the attacking attributes provides the focus for Table 6.6, which explores the association between the obtained outputs from the (7, 7, 2, 7, 7) neural network and the midfield backs attack summary data.

Repeating the process for the attacking variables reveals a number of variables that have significant influence on the output. In the first 'output', the number of breaks made is clearly the strongest, as indicated graphically in Figure 6.3 and numerically in Table 6.6. Passes, Laybacks and Running Metres are the other variables with a strong influence. The results from the half-moon statistic and coefficient of determination produce similar results. Visually, a strong relationship exists between the second 'output' and the total metres run. The quantity of defenders beaten, number of runs (breaks) and the number of passes are also related to the second attack 'output' node. However, this may be due to the correlation between these variables and running metres. The relationship between the number of breaks and the second 'output' node is deemed significant by the half-moon statistic, but not by the coefficient of determination. However, the coefficient of determination is not far below the threshold of 0.25 (0.24), suggesting that there is a non-linear effect occurring. A closer review of this relationship reveals potentially influential points (in a regression model sense) associated with individual's who have made a high number of runs per game that possibly unduly influences the half-moon statistic and the coefficient of determination. Undoubtedly, this effect is due to the dominance of running metres on the second attack 'output' node, and the subsequent correlation between the number of breaks and total metres run (r=0.70).

Variable	*First 'Output'*			*Second 'Output'*			*Combined*		
	DHM	*p*	R^2	*DHM*	*p*	R^2	*DHM*	*p*	R^2
Defence Beaten	1.34	0.14	0.13	4.65	**0.00**	**0.42**	3.84	**0.00**	**0.33**
Kicks Caught	-0.24	0.79	0.00	1.69	0.12	0.02	1.13	0.22	0.01
Errors	0.92	0.31	0.07	-0.43	0.64	0.01	1.26	0.15	0.04
Breaks	9.05	**0.00**	**0.89**	3.08	**0.00**	0.24	5.90	**0.00**	**0.67**
Kicks	1.80	0.06	0.06	-0.13	0.89	0.00	0.72	0.43	0.02
Running Metres	5.76	**0.00**	**0.55**	11.23	**0.00**	**0.89**	10.72	**0.00**	**0.94**
Kicking Metres	0.93	0.31	0.03	-0.47	0.61	0.00	0.20	0.82	0.01
Breakdown Impact	1.65	0.11	0.08	-0.69	0.45	0.00	0.35	0.70	0.02
Passes	8.08	**0.00**	**0.66**	2.13	**0.02**	0.04	4.36	**0.00**	**0.34**
Laybacks	5.00	**0.00**	**0.50**	-0.12	0.90	0.13	2.47	**0.01**	**0.38**
Tackles	-0.36	0.69	0.00	0.46	0.62	0.01	-0.39	0.67	0.00
Missed Tackles	0.48	0.60	0.01	-0.42	0.64	0.00	-0.28	0.76	0.00
Tackle Assists	0.92	0.31	0.00	0.37	0.68	0.01	0.08	0.93	0.00
Tries	-0.43	0.64	0.01	0.89	0.33	0.05	1.08	0.24	0.03
Loose Ball Gained	-0.08	0.93	0.02	-0.95	0.30	0.00	-0.91	0.32	0.01
Harassment	1.14	0.22	0.01	-0.53	0.56	0.00	1.05	0.25	0.00
Infringement	1.64	0.10	0.00	1.45	0.11	0.00	1.66	0.07	0.00

Table 6.6: *Numerical Measures for Attacking Parameter Importance*

Where the first attack node is comprised of four variables involving handling (passing, laybacks) and running quantity (running metres, number of breaks); the second node deals with the quality of running (defence beaten) as well as the running quantity. These results are somewhat satisfactory, reinforced by aspects of Appendix C. However a notable absentee variable is the number of errors incurred. This variable is crucial regarding the perception of performance on attack. As it is excluded there is hesitation in applying a self-supervising network for describing attacking performance.

Additionally, there is a degree of overlap between the two nodes (coefficient of determination = 0.28) which is only a problem due to the poor coverage of the variables.

A relative influence analysis (Table 6.7) provides an indication of the importance of the variables in the attack and defence combinations. As 'donut' type functions are uncontextual in the rugby setting, influence values less than zero can be disregarded and consequently the ranks of the RI measure are suitable for determining importance. The top three variables for each category are emphasised by large bold type.

For the combined defence output, Breakdown Impact, Tackles and Tackles Assists are clearly influential in the analysis, reinforcing the conclusions obtained from Table 6.5.

Conversely, in the combined attack output the obtained MHM statistics do not mirror the findings from Table 6.6 due to the redundancy in the variables. This is due to the high correlation exhibited between variables, such as between the number of breaks and total metres run, mentioned previously. Defence Beaten and Running Metres and the number of Breaks are the influential variables in this analysis.

Variable	*Relative Influence*	
	Defence	*Attack*
Defence Beaten	14	**2**
Kicks Caught	10	4
Errors	15	13
Breaks	16	**3**
Kicks	13	10
Running Metres	17	**1**
Kicking Metres	12	15
Breakdown Impact	**1**	14
Passes	5	5
Laybacks	8	11
Tackles	**2**	9
Missed Tackles	11	12
Tackle Assists	**3**	8
Tries	6	7
Loose ball Gained	9	17
Harassment	4	16
Infringement	7	6

Table 6.7: *Relative Influence Ranks for Attack and Defence Combinations*

The optimal networks for each of the positional clusters are displayed in the following tables (6.8, 6.9). The optimal network was based on the minimisation of the average error (mean squared error). Akaike's Information Criterion is also a suitable method for determining optimal structure (Neath & Cavanaugh, 2000), but not displayed here. It also considers the number of parameters involved and will compensate in order to find model parsimony. Comparison of the MHM for each model may provide additional insight into the most suitable model (maximum MHM). However, this is left for future research. The number of observations per adjustable parameters is noted, and reveals the extent to which these models are exposed to over-fitting. All suitable models have the number of observations per adjustable parameter much less than the lowest recommended value of five. However, the data present is representative of the population as it is a census of all starting players. Finally, each model is assessed in terms of usability. Explanations regarding the usability follow the tables displayed on the following page.

Positional Cluster	*Smpl Size*	*Train Smpl*	*Structure*	*Adjstbl Paramtr*	*Useable ?*	*Obs per Paramtr*	*Ave error*
Props	277	111	7,5,2,5,7	107	No	1.04	2.01
Hookers	139	56	7,5,2,5,7	107	Yes	0.52	1.53
Locks	280	112	7,5,2,5,7	107	Yes	1.05	2.52
Loose Forwards	420	168	7,5,2,5,7	107	No	1.57	5.44
Halfback	140	56	7,5,2,5,7	107	Yes	0.52	27.24
First Five Eighth	140	56	7,2,7	35	No	1.6	4.40
Midfield	279	112	7,7,2,7,7	147	No	0.76	4.46
Outside Backs	420	168	7,7,2,7,7	147	No	1.14	3.94

Table 6.8: *Summary of Self-Supervising Neural Results for Attack*

Positional Cluster	*Smpl Size*	*Train Smpl*	*Structure*	*Adjstbl Paramtr*	*Useable ?*	*Obs per Paramtr*	*Ave error*
Props	277	111	7,6,2,6,7	127	Yes	0.87	0.93
Hookers	139	56	7,4,2,4,7	87	Yes	0.64	1.74
Locks	280	112	7,8,2,8,7	167	No	0.67	1.97
Loose Forwards	420	168	7,8,2,8,7	167	Yes	1.01	1.60
Halfback	140	56	7,2,7	35	Yes	1.60	0.89
First Five Eighth	140	56	7,4,2,4,7	87	Yes	0.64	1.57
Midfield	279	112	7,4,2,4,7	87	Yes	1.29	0.92
Outside Backs	420	168	7,4,2,4,7	87	Yes	1.93	0.87

Table 6.9: *Summary of Self-Supervising Neural Results for Defence*

Certain models are rejected either because they give an uncontextual solution, or there is poor coverage of the variables due to overlap and redundancy between the hidden nodes. Whilst the overlap could be justification for implementing only one node, due to the coverage of events (favourable and unfavourable), such a result is not desirable from a marketing perspective.

Reasons for rejection

1) First Five Eighth: Attack
   First and second components display a high degree of overlap (correlation = 0.73) and poor variable coverage exposing the model to match volatility.
2) Loose Forwards: Attack
   There is a high degree of overlap between the first and second components (correlation = -0.79).
3) Midfield Backs: Attack
   Similarly, an element of redundancy is evident within the two nodes. Additionally the solution is inferior to that provided by factor analysis. This is explored in detail shortly in the next section.
4) Outside Backs: Attack
   Again, the two components exhibit extreme congruence (correlation = 0.96).
5) Props: Attack
   The first attacking component is uncontextual, as it penalises the number of passes made whilst rewarding the number of times the defence is beaten.
6) Locks: Defence
   The second defensive component is uncontextual as it trades breakdown impact against the number of kicks caught, both of which are key tasks relating to the performance of locks.

It must be acknowledged that the number of observations per adjustable parameter is less than recommended. However, the data is representative of the population, for it is a census of all starting players. Secondly, and most importantly, the results accepted are contextual. Admittedly, context is imposed on the model by the manual combination and splitting of variables. User intervention was required to reduce the possible impact of over-fitting. Nonetheless, the structure of the data revealed by the neural networks reflects the expected positional responsibilities proposed by the NZRFU (1991).

The interpretation of each of the networks (attack and defence) which is specified by positional cluster is featured in Table 6.10. The significant variables are displayed for each node of each main attribute. Rejected results are indicated by *italic* type.

In order to compensate for the deficiencies experienced in modelling the key attributes of Midfield Back performance, the models are augmented using factor analysis based on the same variables described for each of the key attributes of attack and defence.

Positional Cluster	Attack		Defence	
	First Output Node	Second Output Node	First Output Node	Second Output Node
Hookers	breaks lay-backs	defence beaten running metres lay-backs	breakdown impact	tackles tackle assists
Props	*defence beaten* *passes*	*breaks* *running metres lay-backs*	breakdown impact	tackles
Locks	breaks passes lay-backs	defence beaten breaks running metres lay-backs	*breakdown impact* *kicks caught*	*breakdown impact* *tackles* *tackle assists*
Loose Forwards	*defence beaten* *running metres* *tries*	*defence beaten* *breaks* *running metres* *passes* *lay-backs*	tackles tackle assists	breakdown impact tackles
Halfbacks	passes	breaks running metres	tackles tackle assists	tackles
First 5/8	*defence beaten* *breaks* *running metres*	*breaks* *running metres* *passes* *lay-backs*	tackles	tackle assists
Midfield	*breaks* *running metres* *passes* *lay-backs*	*defence beaten* *breaks* *running metres*	breakdown impact	tackles tackle assists
Outside Backs	*defence beaten* *breaks* *running metres* *passes*	*defence beaten* *breaks* *running metres* *passes*	kicks caught loose ball gained	tackles

Table 6.10: *Summary of Important Variables in Neural Network Module*

The networks are trained with only one season's data. This is to be pitted against rugby observers who have many years of rugby experience. This is a short-fall of applying statistical techniques to complex process with which individuals have an intuitive understanding.

6.3 Factor Analysis

Having completed the application of self-supervised neural networks to the 2000 Super 12 data, it is time to revisit factor analysis. Initially, factor analysis is performed on the same subset of variables (attack/defence) as specified for use in the neural networks in Table 6.1. This is expanded to include all variables (17 for midfield backs as lineout based variables are excluded). For each iteration of the factor analysis procedure, a varimax rotation is applied to the extracted principal components. Whilst the results from the neural networks suggest that the factors may be oblique rather than orthogonal, oblique rotations (promax) produced almost identical results to the orthogonal rotation (varimax). This congruence of solutions suggested that the underlying factors (KPI's) were independent and consequently, orthogonal based rotation was employed instead. Employing orthogonal factors makes the implementation of the perfection method simpler due to the inclusion of the identity covariance matrix.

Section 6.3.1 uses factor analysis to examine the structure of the attack and defence data (see page 210) separately for each positional cluster using the data described in Table 6.1. Section 6.3.3 repeats the factor analysis including all the summary variables in Table 6.1 for each positional cluster.

6.3.1 Condensed Factor Analysis Module using Key Attributes

Firstly, the defensive variables are examined. To match the neural networks, only two factors are obtained for each subset attribute. As mentioned previously, only two key components are obtained to reduce the potential loss of information in the final reduction. Although some information is lost at this stage, the most influential variables are characterised by the factor analysis. For the 279 starting midfield backs in the 2000 Super 12, the loadings for defence are specified in the following table. The midfield analyses are reported in detail. Only the final results are presented for other positional clusters to protect the commercial content (Table 6.14).

Interesting results are produced by the factor analysis of the defensive variables. In the first factor, three variables have loadings exceeding 0.5, all associated with tackling. However, the good is mixed with the bad. That is missed tackles mimic the behaviour of tackle assists and tackles. This is unsatisfactory, because if this was applied as a rating scheme, individual's would be 'rewarded' for missing tackles! The second factor produces a trade-off situation, pitting the number of kicks caught against the amount of loose ball gained. Again this is unsatisfactory as the resultant factor scores are clouded.

Variable	*Loadings*		*Communality*
	Factor 1	*Factor 2*	
Breakdown Impact	0.49	0.09	0.25
Kicks Caught	0.14	**-0.64**	0.43
Tackles	**0.72**	0.12	0.53
Missed Tackles	**0.59**	-0.22	0.39
Tackle Assists	**0.71**	0.12	0.51
Loose ball Gained	0.10	**0.66**	0.45
Harassment	0.12	0.45	0.22
% Variance Explained	*0.23*	*0.16*	*0.40*

Table 6.11: *Rotated Factor Loadings and Communalities for Midfield Defensive Variables*

Further, only 40% of the variability in the data is explained by these two factors. Additionally, only tackles and tackle assists are explained satisfactorily (communalities exceeding 0.5). Therefore, factor analysis is inappropriate for assessing the defensive attributes of midfield backs. This indicates that the self-supervising network is more applicable.

Variable	*Loadings*		*Communality*
	Factor 1	*Factor 2*	
Defence Beaten	0.70	0.27	0.57
Errors	0.02	-0.78	0.61
Breaks	0.84	-0.41	0.87
Running Metres	0.87	-0.12	0.77
Passes	0.61	-0.42	0.54
Laybacks	0.75	-0.03	0.56
Tries	0.05	-0.44	0.19
% Variance Explained	*0.41*	*0.18*	*0.59*

Table 6.12: *Rotated Factor Loadings and Communalities for Midfield Attacking Variables*

Unlike the factor analysis performed on the defensive data, promising results are produced from the attacking variables with regard to the importance of the 'error' variable. Table 6.12 shows the loadings for each of the specified variables in the two attacking factors. The two factors explain the majority of the variability (59%) inherent within the seven variables. Only the number of tries scored is poorly explained, as implied by the communality of 0.19, well below the accepted threshold of 0.5. The first attacking factor is shaped by the number of defenders beaten, runs or breaks, total running metres and the handling variables – passes and tries. This is essentially an overview of all attacking properties listed. Most importantly, the glaring omission from the self-supervising neural network, errors, largely determines the behaviour of factor two. Thus, the results of this analysis are contextual and provide an approximation of individual attacking performance for midfield backs.

To calculate this estimate, the coefficients are required. These are listed in Table 6.13. Utilising the fact that the factors are orthogonal, and therefore independent, the coefficients are added together, as shown in the combined column. This is essentially an approximation of the perfection method.

Variable	*Coefficients*		
	Factor 1	*Factor 2*	*Combined*
Defence Beaten	0.32	0.37	0.68
Errors	-0.13	-0.69	-0.82
Breaks	0.25	-0.22	0.03
Running Metres	0.31	0.05	0.36
Passes	0.16	-0.27	-0.11
Laybacks	0.28	0.11	0.39
Tries	-0.06	-0.38	-0.44

Table 6.13: *Coefficients for Midfield Attacking Variables*

However, this step provides problems due to the calculated coefficients. As a high score is desired to denote good performance, the coefficients for favourable events should be positive whilst the coefficients for unfavourable events should be negative. However, following the addition of factors one and two, the number of tries scored and passes given become negative. As both are favourable events, this is not desirable. Additionally the number of runs made receive almost no credit (0.03).

Adopting the same generic template for the factor analysis for each positional cluster would be convenient. However, for some positional clusters the data may not have a meaningful latent structure relating to performance.

Further, it is not necessary to adopt two factors. However, the problem of the information loss and the lack of suitable trade-offs between favourable and unfavourable variables means that one factor may be insufficient. Additionally the lack of variable coverage can expose the model to match volatility when only one factor is extracted. It is undesirable to rate an individual's performance on only a few variables, as this may introduce error and bias.

The number of tries scored has a low communality and is removed from the analysis. This is not a problem in defining midfield performance for the following reasons. A major role of midfield backs is to create space and thus scoring opportunities for the outside backs. Further midfield backs seldom score (tries were scored in only 44 instances by the midfield starting sample of 279). Systems such as Fantasy Rugby place too much emphasis on points scored. As rugby is a team game, the process of scoring points is more important than who actually scores them. Many instances can be viewed, such as Leon MacDonald's first test try against Argentina at Christchurch 23 June 2001. Tana Umaga clearly would have scored if he had chosen to go for the try-line himself, however, he opted to pass the ball, unopposed, to MacDonald barely metres from the line. Systems such as Fantasy Rugby (www.fantasyrugby.com) would credit Umaga with an assist. However, the fact that Umaga beat several defenders while scything through the opposition defence before offloading a pass should be more rewarding considering the positional responsibilities.

Rerunning the factor analysis for the midfield back positional cluster with the number of tries scored omitted due to the low communality provides a useful insight into the linear structure of performance. This was repeated twice, in the first instance two factors were extracted, using a varimax rotation; and one factor was extracted in the second instance. Combining the two independent factors by addition (variance explained = 68.2%) provided inconclusive results as no distinction was made between the number of errors made and the other favourable variables. Similarly, the one factor procedure (variance explained = 50.6%) also returned poor results for the same reason.

However, the first factor (variance explained = 42.5%) from the two factor analysis, using a varimax rotation, specified using the six attacking variables (excluding tries) returned a suitable structure, where the necessary contrast between favourable and unfavourable variables was exhibited for the 2000 Super 12 data.

Whilst it only explains 42.5% of the variation in the data, due to the contextual and statistical basis, this is a desirable model to adopt for estimating the attacking performance of midfield backs and it is defined as follows,

$$A_i = 0.455b_i - 0.235e_i + 0.154r_i + 0.361m_i + 0.032p_i + 0.252l_i,$$

where b_i is the number of defenders beaten by the ith individual standardised with respect to midfield backs. Similarly, the following parameters refer to the standardised variables for the ith individual: Errors (e), Breaks (r), Running Metres (m), Passes (p), and Laybacks (l).

The results generated in the preceding two sections imply that general performance for a midfield back can be represented combining the self-supervising network for defence and the factor analysis for attack. However, in some instances the results were not clear and required further treatment to ensure the results were understandable.

The following table summarises the results gathered from the factor analysis for both attack and defence for all the positional modules. Uncontextual aspects of factors are denoted by type. Correct trade-offs are given in *italics*. Combining the information from the neural network section and this abbreviated factor analysis allows two indices to be created, for attack and defence. As indicated in Table 6.10 and 6.14 for some positions, namely midfield backs (attack), locks (defence) and props (attack), both the corresponding neural network output and obtained factor scores were uncontextual. To remedy this problem, the coefficients produced from the factor analyses were massaged to provide contextual results. For both the midfield backs and props this involved taking the maximum coefficient for favourable variables and the minimum coefficient for unfavourable variables. However, for the locks this still provided unusable results as the number of kicks caught were indicated to be non-effectual. The number of kicks caught is an important aspect of a lock's performance. The positional responsibilities of Locks include 'agility in the air and good hands/timing'. Consequently more work is required to adequately quantify the defensive ability of locks. This is achieved by removing the number of kicks caught from the defensive analysis and including this variable with the positional responsibilities for a lock. This then makes the defensive factor analysis suitable.

Positional Cluster	*Attack*		*Defence*	
	First Factor	*Second Factor*	*First Factor*	*Second Factor*
Hookers	defence beaten breaks running metres lay-backs	errors passes *- tries*	tackles missed tackles tackle assists	kicks caught *- loose ball gained*
Props	defence beaten breaks running metres lay-backs	passes *- tries*	tackles missed tackles tackle assists	kicks caught harassment
Locks	breaks running metres lay-backs	defence beaten tries	tackles tackle assists	breakdown impact - kicks caught *- missed tackles*
Loose Forwards	defence beaten breaks running metres passes lay-backs	tries *- errors*	tackles tackle assists	kicks caught loose ball gained
Halfbacks	errors breaks running metres passes	defence beaten breaks running metres tries	tackles missed tackles tackle assists	breakdown impact kicks caught *- loose ball gained* *- harassment*
First 5/8	breaks running metres passes lay-backs	tries *- errors*	tackle assists *- kicks caught*	breakdown impact loose ball gained
Midfield	defence beaten breaks running metres passes lay-backs	errors	tackles missed tackles tackle assists	kicks caught *- loose ball gained*
Outside Backs	breaks running metres passes lay-backs	defence beaten running metres tries	tackles missed tackles tackle assists	kicks caught loose ball gained

Table 6.14: *Factor Analysis Module Summary*

Thus having settled on appropriate measures for attack and defence this provides two comparative measures for reviewing performance that are ideal because of the transparency and contextuality. These indices are reviewed in the next section. Positional indices are ignored in what follows.

6.3.2 Player Performance Indices

Positional Cluster	*Attack*	*Defence*
Hookers	Neural Network	Neural Network
Props	Factor Analysis	Neural Network
Locks	Neural Network	Factor Analysis
Loose Forwards	Factor Analysis	Neural Network
Halfbacks	Neural Network	Factor Analysis
First 5/8	Factor Analysis	Neural Network
Midfield	Factor Analysis	Neural Network
Outside Backs	Factor Analysis	Neural Network

Table 6.15: *Appropriate Methods for Positional Cluster Performance Indices*

It is important to note that the structure of each positional index will differ in basic properties. Thus standardisation within positional clusters of the adopted indices must occur before key attributes can be related across positional clusters such that inter-positional comparisons can be made.

This section has created a meaningful two-dimensional estimate in terms of attack and defence for performance that can be used to compare results generated by other applicable methods, which will be referred to as the augmented neural networks. Essentially, this provides the benchmark with which other methods will be compared. Other methods need to prove superiority, through context, robustness and ease of implementation. The next section revisits factor analysis using all the summarised variables (19) as specified in Table 6.1.

6.3.3 Standard Factor Analysis using Summary Variables

A repeat of the methods applied in Chapter Three using the Table 6.1 summarised data from the Super 12 comprises this section. Using a varimax rotation and adopting the number of factors that are required to explain more than 60% of the variation in the data, the following loadings are obtained for midfield backs.

Variable	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*	*Cmunlty*
Defence Beaten	**0.63**	-0.08	-0.20	-0.20	-0.39	-0.18	0.51
Errors	0.21	-0.05	-0.05	0.07	0.08	**0.67**	0.89
Breaks	**0.87**	0.23	0.06	0.19	0.14	0.18	0.94
Kicks	0.11	**0.96**	0.02	0.02	-0.03	-0.07	0.84
Running Metres	**0.88**	-0.02	-0.09	-0.06	-0.2	0.01	0.94
Kicking Metres	0.04	**0.96**	0.02	-0.08	-0.03	-0.04	0.54
Breakdown Impact	0.26	0.03	0.33	0.13	**0.58**	0.09	0.75
Passes	**0.62**	0.27	0.15	0.48	0.15	0.14	0.67
Lay backs	**0.76**	0.02	0.10	-0.15	0.20	-0.10	0.59
Kicks Caught	0.17	0.17	0.15	**-0.70**	-0.12	-0.06	0.59
Tackles	-0.03	-0.03	**0.75**	0.14	-0.04	0.18	0.59
Missed Tackles	0.04	0.05	**0.62**	-0.12	-0.12	-0.23	0.45
Tackle Assists	-0.02	0.03	**0.66**	-0.00	0.21	0.19	0.54
Tries	0.12	0.08	0.14	0.13	**-0.65**	-0.48	0.52
Loose ball Gained	0.16	0.11	0.06	0.43	-0.07	-0.48	0.47
Harassment	0.02	0.03	0.17	0.46	-0.30	-0.04	0.34
Infringement	0.10	-0.00	-0.01	0.02	0.19	**-0.57**	0.37
% Variance Explained	*0.18*	*0.12*	*0.10*	*0.08*	*0.07*	*0.07*	*0.62*

Table 6.16: *Rotated Factor Loadings and Communalities for Midfield Backs*

Six factors were required to explain more than 60 percent of the variability in the midfield back data. Four variables were inadequately explained by the factor analysis, namely missed tackles, loose ball gained, harassment and infringements as indicated by communalities less than 0.5 in Table 6.16. The highly influential variables (loadings greater than 0.5) in each factor are denoted with bold type.

Interpreting the factor loadings exceeding 0.5 in absolute value allows the factors to be labelled. From Table 6.16, it is evident that factor 1 is associated with attacking variables. The second factor refers to kicking performance. Similar to the reduced factor analysis performed earlier, factor 3 deals with defensive attributes but does not differentiate between tackles made and tackles missed. Factor 4 reflects the number of kicks caught. However for this factor there are a number of variables just below the 0.5 threshold, passes (0.479), loose ball gained (0.432) and harassment (0.459). Essentially, this factor is associated with handling attributes. However, it is a balancing act between kicks caught and the handling variables; passes, loose ball gained and harassment. Obviously as each variable is favourable, this detraction is not desirable. Similarly, the fifth factor compares breakdown impact against the number of tries scored. Finally, factor 6 trades errors against the number of infringements incurred. Clearly, some of these factors are uncontextual and therefore unusable. These factors are 3, 4, 5 and 6.

However, the structure of the factor analysis provides important insight into the performance of midfield backs in the Super 12. Attack is the most important attribute followed by kicking and then defence. Thus it is important to include some measure of kicking with the performance rating for individual midfield backs.

The following table summarises the results obtained by the factor analysis for each of the positional clusters. The most obvious change between this and the chapter three analyses is that the number of factors required has increased for all but the first five eighths.

Positional Cluster	*Sample Size*	*Input Variables*	*Factors Required*	*Illogical Factors*
Props	277	19	7	1
Hookers	139	19	7	3
Locks	280	19	7	1
Loose Forwards	420	19	7	1
Halfback	140	17	6	3
First Five Eighth	140	17	5	2
Midfield	279	17	6	4
Outside Backs	420	17	6	0

Table 6.17: *Summary of Factor Analysis Results*

Additionally, only the outside backs do not possess any illogical factors. The presence of so many illogical factors means that the results obtained from the factor analysis are unusable, as they are uncontextual.

The factors are uncontextual in the sense that the factor scores produced do not represent good or bad performance. From a rugby perspective a number of the factors make sense and deal with the same core attributes. For example a number of the factors relating to defence in a variety of positional clusters are inappropriate purely because as tackles increase so do the number of missed tackles. This makes it difficult to differentiate between good and bad performance as high factor scores could relate to high tackle counts or high missed tackle counts. Other situations that are understandable, but inappropriate, are trade-offs between loose ball gained and kicks caught. Both variables deal with non-direct collection of possession. However, one variable deals with gaining possession in the air, the other from the ground. This could be used to distinguish between aerial and ground-based players. In this case where a high factor score is obtained for a large number of loose balls gained and few kicks caught, and a low factor score relates to a high number of kicks caught and few loose balls gained. However, a factor score in the middle is unclear, is it due to a high number of variables cancelling each other out, or a low number of variables cancelling each other out. This is why factors structured in such a manner are unsuitable.

This suggests that there is an element of non-linearity present in the data set. Rather than manipulate the analyses by adding and removing variables, in an attempt to obtain contextual resolution, factor analysis is left. Already contextual results have been obtained from the augmented neural networks (augmented indices) as defined in Table 6.15, this is deemed sufficient to move onto the next stage, assessing self-organising maps.

The data is not transformed and this has no impact on the analyses. The effect of transforming the data for midfield backs prior to the factor analysis is minimal, with the only apparent difference being that the fifth and six factor components were swapped using data transformed by a square root.

6.4 Dimension Reduction Self Organising Map

A different structure of model is now introduced. Chapter Four showed that a self-organising map is a special type of neural network that performs unsupervised learning. Unlike the networks examined in the previous sections, SOM's are capable of producing only one value, which we will take to represent performance. Two different structures are imposed on the neural networks applied in this section. Firstly a one-dimensional SOM is applied, in an attempt to approximate the main principal curve that represents individual player performance. This is then expanded to a two dimensional array in an attempt to map the principal surface that summarises individual rugby player performance using all 19 variables described in Table 6.1.

The following table describes the array architectures of the self-organising maps applied to the midfield back data. In all instances the maps were computed on data standardised with respect to the relevant positional cluster. The first two columns of Table 6.17 describe the map topology, and then the number of output nodes is given. To obtain a one-dimensional value of performance it is necessary to reduce the SOM topology to a one-dimensional array which is achieved by adding the co-ordinates together and subtracting one (that is the node (1, 2) = 2; (4, 2) = 5). The 'Reduction to 1-D' column lists how many 'output clusters' are available given this reduction procedure, as it is important that the separate dimensions are afforded equal weighting, that is the node (2,1) has the same performance level as (1,2).

Map Topology		*Output Nodes*	*Reduction to 1-D*	*R-Sq*
Rows	*Columns*			
1	10	10	10	0.34
2	5	10	6	0.33
3	3	9	7	0.30
4	4	16	8	0.47
1	16	16	16	0.46
2	8	16	9	0.43
5	5	25	9	0.54
4	6	24	9	0.54
3	8	24	10	0.53
2	12	24	13	0.52
1	24	24	24	0.54

Table 6.17: *Optimal SOM Topology*

It is desirable to have approximately 10 final clusters, as this provides a rating scale of 1-10 which is the general convention adopted by sports reporters in the media to rate performance and a scale that rugby observers are comfortable with.

It is evident from Table 6.17 that as the number of nodes increases so does the co-efficient of determination. This makes sense, as for each typical input pattern a node is required. Obviously the more typical patterns, the more nodes that are needed to explain the variability in the entire data set. An exciting prospect of the self-organising maps is the retention of information. Whilst the Factor Analysis models explain more than 60 percent of the variability in the data, much of this is lost in the final reduction to a singular performance measure. Assuming independence and that each of five factors is equally important in explaining 60 percent of the data, the final model will explain 12 percent of the variation in the original data because in this instance the final summation explains only one fifth of the information from the factor scores.

By comparing this hypothetical minimum of 12 percent of variation explained by the factor analysis to the obtained tabulated values for the coefficient of determination for midfield backs, it is evident that the SOM is superior at information retention.

The self-organising map extracts rules that allow the output to be interpreted. This is displayed in a cluster profile tree, described briefly in Chapter Four. It is at this stage that potential problems are encountered. Whilst it must be acknowledged that the variables are highly correlated, the manner in which some of the nodes are decided creates an element of scepticism about the possible use of the SOM as a rating system. Even though similar clusters are close to each other due to the neighbourhood function, the cluster profile tree creates doubts about the use of the SOM. As an interpretation tool, the cluster profile tree is a cumbersome object to handle. For the 1×10 self-organising map for midfield backs, 10 steps are required to assign all observations to relevant bins. This requires 12 pages of A4 output from SAS. The importance of each variable, defined by the cluster profile tree, as described in section 4.6, is displayed in the following table.

A different structure is adopted for the two different maps but both methods make sense from a rugby perspective, with running metres deemed to be the most important variable in both maps. However, little information is gleaned from Table 6.18, as whilst the importance of each variable is listed, information regarding how each variable impacts on the performance rating is not clear. These findings are best illustrated graphically, making use of the performance indices for attack and defence developed earlier.

Variable	*(1×10)*	*(4×4)*
Defence Beaten	0.43	0
Errors	0	
Breaks	0.26	0.41
Kicks	0.95	0.25
Running Metres	1.00	1.00
Kicking Metres	0	0.77
Breakdown Impact	0.55	0
Passes	0.55	0.21
Lay backs	0.21	0.25
Lineouts Won	0.15	0.15
Lineouts Lost	0	0
Kicks Caught	0.73	0
Tackles	0.54	0.92
Missed Tackles	0	0.15
Tackle Assists	0	0
Tries	0	0
Loose ball Gained	0.21	0
Harassment	0.46	0.69
Infringement	0	0

Table 6.18: *Differences in Variable Importance Between a 1×10 and a 4×4 SOM Topology*

Figures 6.6-6.9 illustrate the underlying contextuality of the SOMs in the case of Midfield Backs. To understand simply how the system operates holistically, the means for each level with respect to the attack and defence (augmented) indices calculated earlier are plotted for each SOM level. For example, the 10 levels produced in a 1×10 SOM are displayed in Figure 6.6 together with the mean attack and defence indices. Somewhat surprisingly a SOM level of 10 appears to be miscast with a relatively low attack index. The reason for this is provided by the cluster profile tree.

What the cluster profile tree exposes is the tendency of certain events to dominate the structure of the rating system. For an individual to score 10 on the one-dimensional self-organising map they must have thrown 15 or more passes. That is the sole criteria. Whilst overall this is understandable due to the correlation between other variables, this is not a contextual description of high performance. It is the holistic behaviour of the system that is of interest. A similar effect is found with the 4×4 SOM in Figure 6.7.

It is preferable that the SOM topology is symmetrical as it is expected to map the principal surface which is denoted by attack and defence as shown in Figures 6.8 and 6.9. However, extending this surface too far (5×5) and the map deteriorates due to the blind acceptance of the kicking attributes. A potential reason for this is found in the intrinsic dimensionality of the rugby data being greater than two because of the positional variable. Consequently, when the topological map of the SOM becomes too large, the SOM attempts to introduce the higher dimensionality into the two-dimensional topological mapping. Another possible option, as suggested in Figures 6.6 and 6.7, is to ignore the SOM ranking the clusters based upon the location of their centres with respect to the attack and defensive indices instead. However, this distorts the symmetry of the map, and is not pursued. Figure 6.10 displays the result of adding the separate topological co-ordinates of the 4×4 SOM together then subtracting one to get a more appropriate ranking, as suggested previously.

The four by four SOM approximates the principal surface inherent within the data structure. What is exciting is that the two dimensions of this surface map the core components that were identified at the start of this chapter, attack and defence. This defends the approach adopted in creating the indices. This congruency between methods suggests that there exists some principal surface that can be approximated to estimate individual rugby player performance.

The following graphs show clearly that the columns of the 4×4 self-organising map relate to attack, and the rows correspond to defence. This implies that the self-organising map has identified a principal surface that is congruent with the manual division explained earlier in this chapter relating to neural networks. This justifies the approach adopted by indicating that these are actually the key components, and that performance can be thought of as a principal surface relating to attack and defence.

		SOM Column Co-Ordinates (Attack)			
SOM Co-Ordinates Row (Defence)		*1*	*2*	*3*	*4*
	1	1	2	3	4
	2	5	6	7	8
	3	9	10	11	12
	4	13	14	15	16

Table 6.19: *4×4 SOM Co-Ordinates and Output ID*

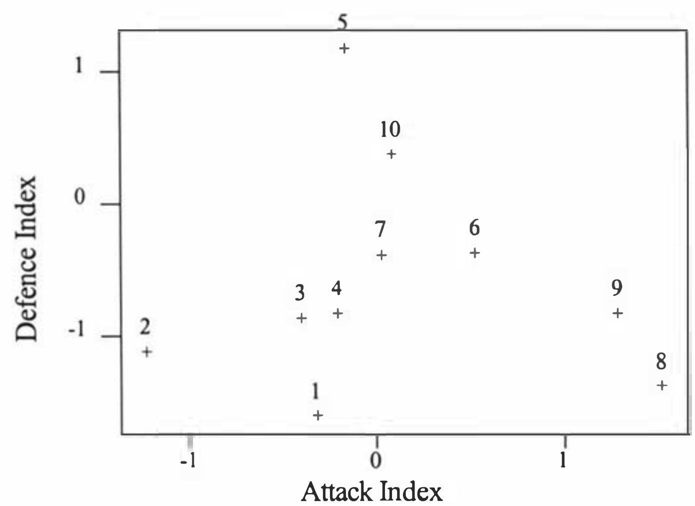


Figure 6.6: 1×10 Self-Organising Map referenced against Attack and Defence Indices for Midfield Backs

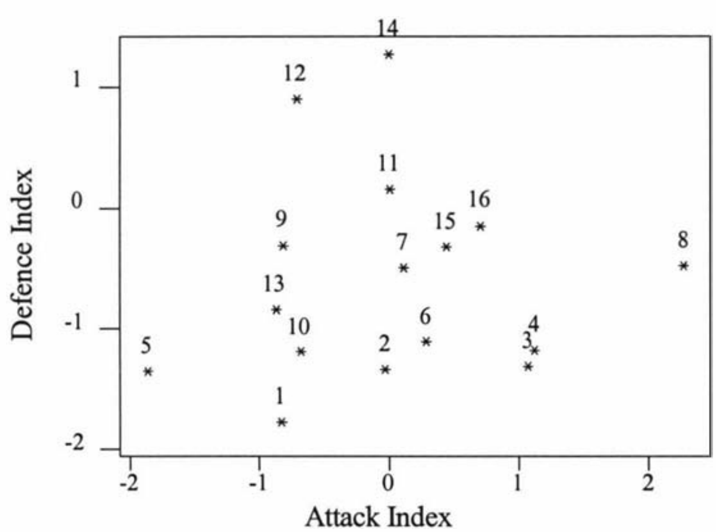


Figure 6.7: *4×4 Self Organising Map referenced against Attack and Defence Indices for Midfield Backs*

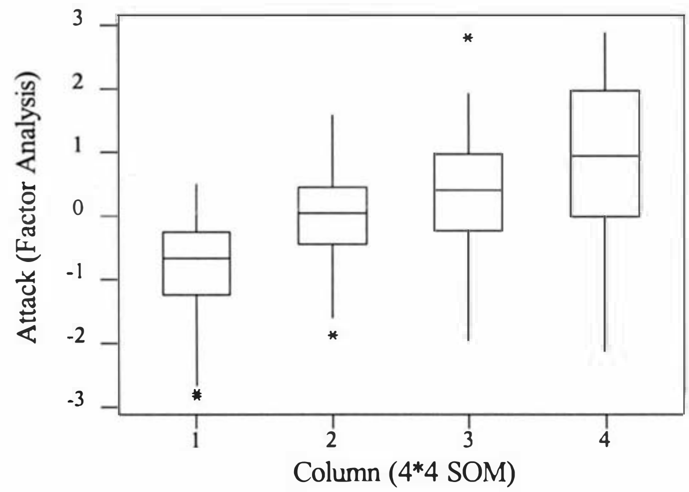


Figure 6.8: *Boxplot of Attack Measure by SOM Column for Midfield Backs*

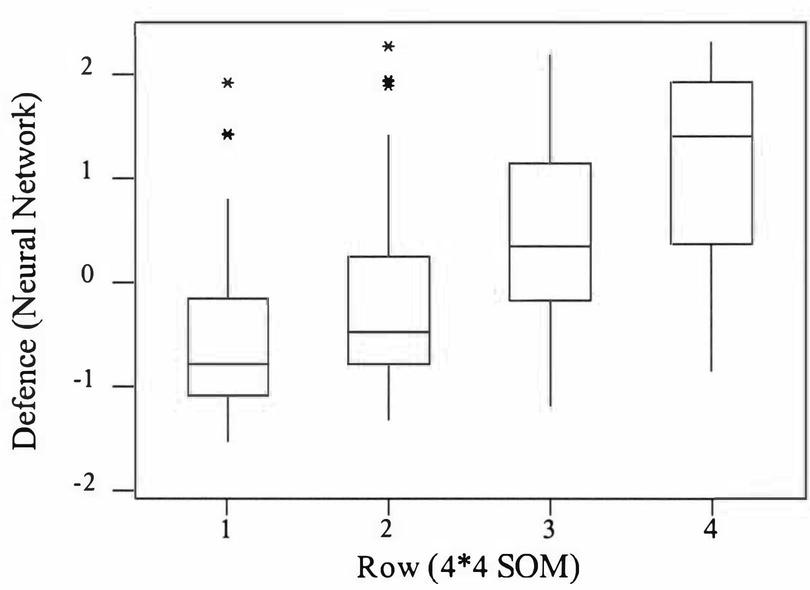


Figure 6.9: *Box plot of Defensive Measure by SOM Row for Midfield Backs*

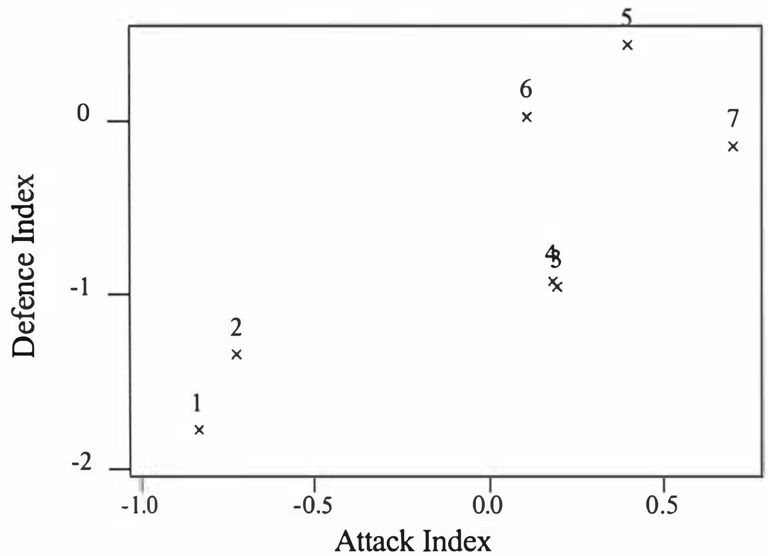


Figure 6.10: *Combined 4×4 Self Organising Map Output Referenced against Attack and Defence Indices for Midfield Backs*

There is one anomaly in this map (Figure 6.10) and that is the mean of cluster 5 is better than the mean of cluster 6 in terms of both attack and defence. However, the positional variable is not addressed above, perhaps explaining this anomaly.

6.5 Quartile Method Based Method for Discrete Ranking

Unfortunately the results expressed by the previous methods present ranking systems that are insufficient for the needs of this project. A one-dimensional ranking that incorporates defence, attack and positional skills is needed. To create a more contextual rating system the operator could intervene and provide a model based on expert opinion. This step is considered last resort as it could introduce bias and reduce the objectivity of the rating. This approach is explored before the KPI developed previously from the different methods are combined to produce a suitable rating using the data quartiles.

The previous section showed that the SOM reduced the data into a suitable form, but the cluster definitions were inappropriate for use in the Eagle Rating due to the discrete nature of the output. However, by employing expert opinion on the attributes required by position as outlined in the NZRFU principles of coaching, reproduced in Appendix

C, this allows a manual decision tree to be created. This is similar to the decision tree used to interpret self-organising maps in SAS. However, it is only for the loose forwards that such methodology can be employed as the majority (7/13) of positional responsibilities are directly measurable which is listed as follows with original codes in parentheses:

- High work rate in support and cover (Consider Tackle Involvement - DT, DTL, DOT, DA; Breakdown Involvement – DB1, DB2, DB3; Attack Involvement – RP, AYM, PYM, PL, AL, AP, PP, APT, PPT)
- Speed to breakdown (DB1, DB2, DB3)
- Good running and linking skills (AY, PY, APT, PPT, AP, PP)
- Strong tackles (DTL, DOT)
- Able to drive ball carrier back around fringes and turn ball over in tackle (DOT)
- Possible lineout option (JW^{+})
- Win ball on ground (DG)

Consequently, this form of analysis is not applied, as a generic model that has a consistent methodology and structure is sought.

As specified by the NZRFU positional responsibilities, the performance for many positions cannot be described entirely by statistical means. Midfield Backs for example must have the following attributes, among others:

- Straightness in attack
- Excellent communication/organisation skills
- Understand/read the game.

This information is not easily extracted or quantified. Therefore a model cannot be created based upon this information. The only insight into performance stems from the work applied previously in this chapter and Chapter Three.

Ensuring the model is robust and contextual is a relatively simple task. Consider the attributes for attack and defence produced by the factor analysis and the neural networks. In both instances the most suitable model has been selected, ensuring that the method adopted was contextual. These methods can be made more robust by considering a non-parametric ranking scale associated with a likert type scale. To do this, the data is split into quartiles. Quartiles are measures of non-central location that are often used to describe data. The ordered data is split into quarters with equal numbers of observations in each quarter (Levine, Krehbiel & Berenson, 2000). Quartiles act as the cut-off for determining which quarter an observation will be placed. Quartiles are adopted for convenience and so that the data may be applied to the quartile control chart developed by Bracewell (1999). Additionally, if three KPI are used, 10 different levels of performance are obtained.

Four stages are established for considering match performance.

- Well above average
- Above average
- Below average
- Well below average

This scale can be incorporated into the existing model structure by dividing the ranked output into four approximately equal sections. The most realistic method for doing so revolves around the quartiles, and thereby introducing an element of robustness into the model by considering a non-parametric approach. For this situation attack and defence are transformed to discrete data by the following method for transforming the summarised attribute scores.

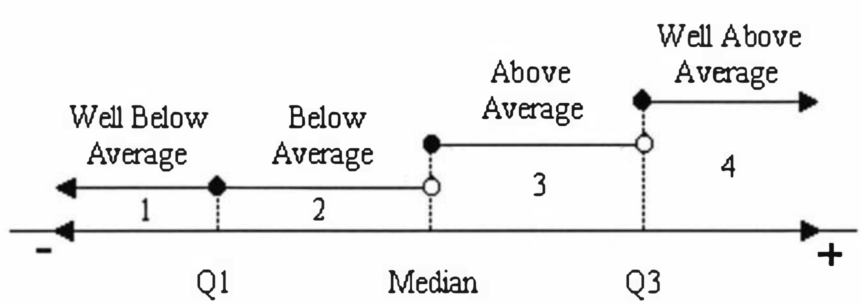


Figure 6.11: *Structure of Quartile Method Scaling for Attack, Defence and Positional Variable*

Output scores relating to performance that are less than the lower quartile (Q1) are denoted well below average and given a score of one. The next quartile covers the output scores extending from the lower quartile to less than the median and is credited with a score of two. Similarly, above average scores relate to output equalling or exceeding the median value but not exceeding the upper quartile (Q3) and receive a score of three. Finally all values greater than or equal to the upper quartile relate to well above average performance and attract a score of four. The resultant discretised values for attack and defence are then added together, providing a minimum of two and a maximum of eight. This raw component system is then augmented using the positional cluster specific skill components. The same scaling applies, however some adjustments need to be made depending on the position involved.

For example, for hookers it is obviously not desirable for them to lose lineout throws. Thus the scaling is inverted, that is above well average performance refers to output scores falling below the lower quartile. To produce a consistent scale, two units are subtracted from the computed score to give the rating a range between one and ten.

This quartile model provides an additional method to assess in Chapter Seven. This model is contextual, robust and easy to implement.

6.6 Summary of Potential Rugby Ratings

Three core methods for dimension reduction have been assessed, namely factor analysis, self-organising maps and self-supervised feed-forward neural networks. The benefits and drawbacks will be discussed with particular reference to the output of the rating system in Chapter Seven. Through the application of these methods three potential models are generated from this chapter – firstly, the augmented indices that calculated the key attributes of performance using summarised variables and the techniques of factor analysis and neural networks; a condensed four by four self organising map; and finally a hybrid-type quartile model that sought to combine the best features of each of the three techniques assessed.

This chapter has shown that in the sporting context, none of the methods were suitable as a stand-alone method for quantifying individual rugby player performance due to limitations ranging from robustness, to the inability to extract meaningful results for all positional clusters using just one technique. However, by taking the best features from each of the methods, the philosophical requirements necessary to create a sport rating system are obtained. This required the highlights of each method to be melded into a hybrid method. Consequently, it is possible to incorporate data mining techniques into the sporting environment.

In Chapter Three a number of issues that impinged on the context of the Eagle Rating were discovered. This alerted us to potential problems in Chapter Six. In the course of this chapter other problem areas surfaced and were dealt with as described below.

The problems associated with the positional clusters were explained in Chapter Three. In this Chapter the eight positional clusters based on expert opinion described in Chapter Three were adopted.

A danger of the rating system was the sensitivity to match involvement. Neural network techniques were employed in an attempt to remedy this problem. The robustness of the system is examined in Section 7.2.3 in the next chapter. Introducing a discrete coding system based on the quartiles of the positional cluster population reduced a similar problem, the exposure to match volatility.

In order to successfully apply neural networks, the number of variables had to be drastically reduced to avoid overfitting. This approach condensed the number of variables from 93 to 19. This required a simplification of the number of key attributes, where three general attributes were identified, attack, defence and positional specific skills, enabling a generic template to be adopted.

In some instances a generic template for the factor analyses could not be adopted. The blanket generic approach produced answers that were uncontextual or too complicated to explain in some instances. Under these circumstances the factor analyses were redone to augment the existing factors/indices.

Each of the methods employed will be discussed in Chapter Seven with attention given to the perceived strengths and weaknesses of each method allowing a model that satisfies the primary concerns of context, robustness and implementation to emerge.

Chapter Seven

Comparison, Selection and Implementation of Suitable Methods

Following the application of the methodologies discussed throughout this thesis, it is necessary to evaluate and compare methods such that the 'best' method is adopted for quantifying performance and the subsequent implementation as the Eagle Rating. The 'best' method is the most contextual and robust model for describing individual rugby player performance in a single match. Whilst the continual reference to context and robustness becomes repetitive, the importance of these terms cannot be stressed enough. This is due to the intended exposure of the rating system to public scrutiny and the commercial implications thereof. The practicalities relating to the implementation of the selected methodology must also be considered when adopting a particular method.

In this chapter the unconventional statistical approach that was adopted in Chapter Six in reaction to the demand for contextual models is expanded upon. Specifically, rating systems that are the result of the split analysis for the key attributes of defence and attack are examined and compared with the conventional techniques. The unconventional hybrid technique formed by combining neural networks and factor analysis to produce the quartile method is also compared. The key aspects of transparency (validity), context and robustness provide the necessary reference for comparison.

The most important requirement of the selected model is context. If the results are not congruent with rugby thinking, then the rating is of little use as it will not be credible.

Robustness is a related concept where models are required to produce contextual results in a variety of situations. Given that the model selected is contextual, easily interpreted, and robust; a framework for marketing the rating can be developed. Robustness relating to varying match conditions is difficult to test. Accordingly, the reference to robustness is generally conceptual.

This chapter commences with an overview of eight potentially suitable methods applied in Chapters Three and Six for quantifying individual rugby player performance. Each of these dimension reduction methods is discussed, briefly detailing the techniques involved and focusing on the perceived strengths and weakness. The second section compares the suitable methods with the key attribute indices identified in the early stages of Chapter Six due to their coherency. In the discussion section, the best method is identified, which is then implemented in subsequent sections. Discussion relating to the implementation of a rating derived from multivariate methods focuses on the coaching environment. This important section demonstrates how a rating derived from dimension reduction techniques can be incorporated into a diagnostic coaching structure, placing the obtained statistics correctly into the sporting environment. The impact of the work provided in this chapter is then summarised in the final section.

The section describing the implementation of statistics in a diagnostic coaching structure has been published in *Research Letters in the Information and Mathematical Sciences* (Bracewell, 2002) and distributed on request to a number of top level coaches. Additionally, the work describing expert systems in this chapter is under review for *Expert Systems* (Bracewell).

7.1 Methods for Comparison

Three methods for dimension reduction have been assessed in this thesis, namely factor analysis, self-organising maps and self-supervised feed-forward neural networks. Each method has advantages and disadvantages which are either mathematical or conceptual. Different variations of these methods have been applied to maximise the benefit from the methodological strengths and minimise the weaknesses. In this section the benefits and drawbacks are discussed with particular reference to the output of the rating system.

In the case where the multiple original variables are reduced to multiple components (factor analysis and self-supervising neural networks), the perfection method introduced in Chapter Three is used to further decompose match performance to a singular value for each individual. When the components are independent, the perfection method can be approximated by simply summing the standardised components in question.

Subsequently, the standardised multiple components are summed to produce a singular value that represents individual match performance. Obviously this assumes that each key attribute is of equal importance. Considering just attack and defence this is a valid assumption. Expanding the performance measure to include a positional specific attribute introduces a potentially problematic situation. However, the aspects covered such as lineout ability and kicking ability can be considered to be of equal importance for given positions. In support of this statement, these events are 'more visible', in that they are key tasks which capture viewer attention and the associated media scrutiny.

The three core methods produce eight methods for comparison. All are univariate individualistic-based rating systems for individual match performance. That is the ratings only account for what an individual has done with no regard to the performance of peers. Factor analysis is used in four methods, as described in Chapters Three and Six. The first application of factor analysis was the creation of the original Eagle Rating using the perfection method. This process was repeated in Chapter Six using the condensed data set. Factor analysis is also instrumental in calculating the augmented indices and the subsequent quartile model. The underlying construction of both the augmented indices and quartile methods is identical with each index relating to key performance attributes defined by either factor analysis or a neural network, whichever produced the most suitable results.

Similarly, self-supervising feed-forward neural networks are utilised in three methods; the aforementioned augmented indices, quartile model and as a stand-alone component-based system.

Finally, self-organising maps are featured in three methods, two relating to a different dimensionality. The first SOM is one-dimensional using a 1×10 topology. This is expanded in the second instance to a two-dimensional mapping with four columns and four rows producing 16 output nodes. The third method induces a one-dimensional mapping from the 4×4 SOM by adding the separate locations of the corresponding dimensions for each individual performance to introduce symmetry.

Apart from the original Eagle Rating, which operates on nine positional clusters, all others use the eight positional clusters outlined in Chapter Six (Props, Hookers, Locks, Loose Forwards, Halfbacks, First Five Eighths, Midfield Backs and Outside Backs). The adoption of the eight positional clusters was based on expert opinion discussed in Chapter Three, as outlined in Chapter Six. It was unreasonable to expect numerical methods to identify the blurred boundaries between positions at top level rugby.

7.1.1 Original Eagle Rating

The first method applied, the original Eagle Rating was produced using factor analysis using between 25 and 31 variables for each positional cluster (Appendix B). This indicated that it was possible to quantify individual rugby player performance using statistical dimension reduction techniques. Using a varimax rotation, five factors were required to explain approximately sixty percent of the variability in the data for nine positional clusters. The perfection method was introduced to adequately combine the separate factors into a meaningful one-dimensional performance measure. A drawback of the initial method was the linearity assumption. Further, the rating is vulnerable to the amount of play (action) in a match. This problem was highlighted during the All Blacks European Tour. Whilst there is some justification for increasing ratings with respect to involvement, it became particularly evident that this is not desirable in a rating system for individual rugby players. Essentially, as the total number of sequences in a match increases so does the original Eagle Rating. Intuitively this makes sense as the more positive involvement one has in a match the more that individual should be rewarded. However, this was exposed as a problem when the Eagle Rating was used to select the Sky Television Man-of-the-Match during the All Blacks European Tour, 2000. This highlights the problem of incorporating a linear structure into a model for individual performance.

Additionally, equal weighting was afforded to each of the factors. This approach is unsound as the relative importance of tasks differs. A prime example for a midfield back is the comparison of attack out-wide in traditional back-type play as opposed to attack around the fringes of a breakdown in typical forward-orientated play. Clearly, it is more important for a midfield back to operate better in traditional back-type situations.

Orthogonal rotations (varimax) were used to make the factors easier to interpret. This forced the model to employ more variables in order to withstand match volatility.

However, a major advantage of using factor analysis is the ease of implementation due to the low number of adjustable parameters and the fact that the rating system can be represented as a linear equation. This enables the rating system to be quickly adopted into an automated framework enabling automatic calculation of the rating. The relatively low number of adjustable parameters ensures that it is relatively easy to update the system. Unfortunately, factor analysis is vulnerable to match volatility and therefore lacks robustness. This makes it unsuitable as a stand-alone method for quantifying performance.

7.1.2 Factor Analysis for Condensed Data

A replication of the general methodology applied in creating the initial Eagle Rating, the second iteration involving factor analysis used the condensed data set created to reduce the impact of over-fitting in the more flexible methods generated by neural networks. This data set summarised the available 93 variables into 19. The reduction to 19 broad variables will reduce the cost of data collection considerably. Instead of fixing the number of factors to accommodate uniformity amongst the positional clusters as with the original Eagle Rating, the number of factors adopted was dependent on the combined variability accounting for at least sixty percent of the total variability. As the positional responsibilities differ between positional clusters, in relation to the size of the general skill set, the number of factors required to explain the latent dimensionality will also differ. Additionally, positional clusters that have broad skill sets (loose forwards) may not produce clearly defined latent dimensionality because of task overlap and match volatility.

Once again the rating output is restricted to linear based combinations of variables, again exposing the system to match volatility and level of relative individual involvement. Further a number of the results obtained were non-contextual. The factors were non-contextual in the sense that they were inappropriate for determining player performance.

However, it must be noted that the factor analysis did provide interesting information regarding the type of performance, in some instances providing the contrasts between types of players. An example of this is determining whether a loose forward is more likely to gain possession as an 'on-the-ground scavenger', or as a 'linking, agile, ball-in-the-air type player'. Such information is of obvious use especially in determining 'horses for courses' and cannot be ignored completely. It is therefore clearly desirable to build underlying performance measures, such as key attributes, into a rating system.

The use of continuous data is also advantageous for identifying small differences between comparable players, although it does expose the system to match volatility and involvement, when it has a linear basis. Again, this diminishes the robustness of the system.

Making use of the lessons learnt from the initial Eagle Rating, a rating system for individual performance and/or ability can only ever be a general measure of performance and/or ability, due to the limitations in data collection and off the ball activities that are also indicative of performance and ability. Also, the effect and influence of the underlying key performance indicators and key attributes cannot be ignored, as these add transparency to the system.

For comparison purposes in Section 7.2, a rating based on the condensed factor analysis is created by reducing the obtained factor scores to a single performance measure via the perfection method.

In adopting factor analysis the primary consideration is the ease of implementation and the lack of robustness. Ultimately, the lack of robustness restricts factor analysis as the basis for a rating scheme as this is linked to context, which is crucial in gaining credibility.

7.1.3 Neural Network

The introduction of neural networks as a potential modelling tool served as a reminder that typically sport generates relatively little data. To combat over-fitting potentially arising from the large number of adjustable parameters in a network that is capable of mapping non-linear functions, the number of variables needed to be greatly reduced. System economy was achieved with remarkable ease, 93 variables were condensed to just 19, with minimal information loss. Sensibly, this condensed data set is more suitable for statistical purposes due to the small number of interactions an individual has in a match (for example consider total tackles made rather than the distinct types of events such as turnover tackles made with the left shoulder coming from the outside in).

To further condense the variables, appropriate variables were then put into two major categories, attack or defence. This action is supported by the results obtained from the two-dimensional 4×4 SOM (attack = columns, rows = defence), and has huge implications on how performance is perceived and quantified as it indicates the existence of a principal surface relating to the key concepts of attack and defence.

To quantify performance using the key attributes (attack and defence), a four layer feed-forward self-supervising network, with a bottleneck that represents the dimension reduction "output" was applied to each attribute for each positional cluster. Bias was included on all layers except in the bottleneck "output" layer and a hyperbolic tangent was utilised as the activation function in the mapping and de-mapping layers. To improve the generalisation of the network, early stopping was employed. The network architecture was established heuristically to minimise the average error in the test data set. Whilst the number of nodes in the hidden layers varied, the bottleneck layer was two in all instances. This emphasises the multifaceted nature of performance.

A rating system based purely on neural networks is created by combining the two nodes via the perfection method in each of the key performance attributes (attack and defence). These two key attribute measures are then summed through an approximation of the perfection method. This measure is used for comparisons in section 7.2.

Realistically, despite the effort to reduce the number of variables and consequently the associated number of adjustable parameters, the networks employed were still prone to over-fitting. Typically, there was 1 observation per adjustable parameter, rather than the recommended 5-10 observations.

Additionally not all models were contextual, invalidating the use of neural networks as a stand-alone method. However, for some positional clusters the results were contextual, indicating that the self-supervised feed-forward neural networks could be of potential use.

There were clearly a number of benefits arising from the use of self-supervising feed-forward neural networks. Continuous data is obtained, allowing small differences in performance to be identified. The use of sigmoid activation functions reduces the impact of outliers. The 'S' shaped curves of sigmoid functions dampen the effect of outliers and extreme values reducing the impact of match volatility and involvement. This enables the limiting relationships that are part of rugby performance to be better handled. Graphical evidence suggests that there is a limiting relationship, or an upper ceiling that restricts performance in variables with respect to other variables.

Ultimately the use of neural networks in quantifying performance is limited due to the vast amount of data required to justify the inclusion of adjustable parameters without over-fitting. However, a component based system isolating key measures reduces greatly the number of adjustable parameters.

The half moon statistic was developed to aid the interpretability of the networks. However, influence was easily established using graphical methods associated with sensitivity analysis because so few variables were involved.

7.1.4 1 × 10 Self Organising Map

A one-dimensional self-organising map was established to approximate and explore the first principal curve using the 19 summarised variables. Ten nodes were adopted, in an attempt to induce an automatic rating scale, similar to a likert scale, with one representing inferior performance and ten representing superior performance. A batch

SOM algorithm, with both local-linear and Nadaraya-Watson smoothing was employed on the standardised data. The seeds were initialised using principal components and other default settings of SAS Enterprise Miner. In some regards the resultant output was contextual but unfortunately the SOM tended to trade good attack performances against good defensive performances instead of producing a single performance rating scale. This suggested that a two dimensional map was necessary in order to separate out these key attributes.

Further, rule extraction based on a decision tree enabled the structure of the map to be better understood. However, whilst the rule extraction could be mapped to humanistic perceptions and relevant justifications made, there is cause for apprehension. The cause for apprehension is particularly evident at the top end of the scale, where only one 'decision' is made before categorising an individual's performance. Utilising the midfield example, the structure of the model is such that for a midfield back to score a 10, they need to throw 15 or more passes to hand. Whilst this could be justified through positive involvement, presentation of such interpretations to the general public would prove disastrous, regardless of the underlying structure. Match performance can not be isolated through the performance on just one key task. It is only at the top end of the map that this is a problem, as only one decision is used to determine the components of the first bin.

A major benefit of using a one-dimensional SOM is that the data set is reduced to a singular measure of performance in one step, thereby reducing both information loss and the introduction of bias through incorrect weightings.

Further, the obtained data is discrete incorporating robustness into the system. The rule extraction of the SOM demonstrates this robustness by reducing the impact of match involvement through increased activity, by having 'cut-off' regions (greater than, less than). Unfortunately, the immediate reduction to discrete data makes it virtually impossible to detect small differences in level of performance. Whilst this is not necessary for a system aimed at the general media, as a diagnostic tool for coaches and selectors this is unsatisfactory.

So the required model transparency (interpretation) obtained through rule extraction provides results that cannot be exposed to public scrutiny. Further, the one-dimensional mapping is uncontextual, contrasting attack performance against defensive performance. This renders a one-dimensional SOM unsuitable and it is not sufficiently sensitive for use as a diagnostic tool.

7.1.5 4 × 4 Self Organising Map

As a direct consequence of the unsatisfactory trade-off between attack and defence in the one-dimensional SOMs, the topology of the map was expanded to two dimensions. The best structure was found iteratively, by considering model context. Based on the findings of the one-dimensional SOM, contrasting attack against defence, it was important that balance was incorporated in the two-dimensional SOM by setting the number of rows equal to the number of columns. As a consequence a 4×4 SOM was adopted.

Using the key attribute indices (obtained in Section 6.3.2 utilising factor analysis and self-supervising neural networks) as barometers it was found that one dimension was associated with attack and the other with defence. As the separate dimensions were aligned with the contextually defined indices, this reinforced the decision to identify the key attributes in this way.

The contextuality of the SOM is an exciting development. Unfortunately, the rule extraction proved untrustworthy due to the expression of dependence on only one key task to initiate cluster segmentation. However, the indices can be used to provide a more conceptual understanding of the SOM values.

Another difficulty arising with the use of a two-dimensional SOM can be directly attributed to the two-dimensional map topology. The two key attributes of attack and defence need equal weightings to reflect relative importance. Also the discrete nature of the data and the immediate reduction to a singular measure of performance restricts the applicability of the obtained information to coaching and selection procedures.

Attack				
+	(1,4)	(2,4)	(3,4)	(4,4)
	(1,3)	(2,3)	(3,3)	(4,3)
	(1,2)	(2,2)	(3,2)	(4,2)
–	(1,1)	(2,1)	(3,1)	(4,1)
	–	*Defence*		+

Table 7.1: *4×4 SOM Topology*

Unfortunately the two-dimensional map is itself unusable due to the unequal weightings given to each of the key attributes. Essentially, the output from the 4×4 SOM must be treated as a pair of points to produce an appropriate ranking system. Also the discrete nature of the data and the immediate reduction to a singular measure of performance restricts the applicability of the obtained information to coaching and selection procedures. However, the two-dimensional topology is of some use to the coaches as performance on separate key attributes can be established, although due to the discrete nature of the output such information is rather coarse.

The most important feature of the self-organising map is the crucial insight into the structure of individual performance. The two-dimensional SOM is an approximation of the principal surface in terms of two key dimensions, attack and defence. This supports the selection of the indices as a diagnostic tool for coaches and selectors.

7.1.6 Paired Self Organising Map

As described in the previous segment, symmetry with respect to the attack and defence attributes is required due to exposure to match volatility, involvement and the requirement for equal weightings on these attributes. Symmetry in the two-dimensional mapping infers equal importance. This is achieved by simply adding together the separate co-ordinates that identify performance in the two-dimensional mapping. Addition of the two key attributes is justified with reference to the perfection method. The ensuing rating system consists of only seven discrete values, which is very restrictive. However, the problem of symmetry and associated equal weighting is solved. Additionally, the method is robust and can be shown to be contextual, if the relationship with the key attribute indices is used rather than the rule extraction.

Unfortunately, due to the immediate reduction to a single match rating, the structure of that performance is not evident, due to the lack of key performance indicators. However, the underlying two-dimensional topology can be used to identify elements of performance associated with the key elements of attack and defence. Further, the discrete nature of the output is not suitable for fine analyses.

Seemingly, we appear to be back at the starting point. Summarising the desired values for a rating system for individual rugby player performance that have been identified so far in this section:

1. Context
2. Robustness
3. Continuous Output
4. Singular Performance Measure
5. Summary of Performance

A number of these attributes are contradictory as outlined in the previous subsections. However, the next two methods seek to accommodate these features, by implementing the lessons learned from factor analysis, self-supervising feed-forward neural networks and self-organising maps.

7.1.7 Augmented Indices

The augmented indices combine the key attributes of performance together to from a single rating. As described in Chapter Six, these key attributes are based on the key areas of match performance; attack, defence and a key positional task. The key positional tasks cover position specific tasks such as lineouts for second and back rows, lineout throwing for hookers, kicking for backs and scrummaging for props.

The attack and defensive indices are established using either factor analysis or neural networks with a diminished number of key variables. The choice of methodology for each attribute, for each position, from either factor analysis or neural networks is determined by the context of the results. Consequently, the indices are contextual and cover the main aspects of individual performance as described by match statistics.

Additionally, this provides continuous data which is necessary for advanced users (coaches and selectors) whilst retaining context and transparency in the model structure.

The third key attribute relating to the key positional task is determined similarly to the attack and defensive indices. However, in several cases, the measure of performance is univariate, enabling the standardised variable to be combined with the other key attributes to produce a rating of match performance. However, the data for two broad groups of positions is deficient – kicking for backs and scrummaging for prop forwards. The kicking attribute is easily established by expanding current data collection procedures to cover the underlying reasons for kicking, namely retention of possession, gain of territory and application of pressure. Thus kicking performance is manifested by simple codes describing if possession was retained or not, if the kick was contested or not, and the net territory gained as a result of the kick. These measures can be combined by using either a neural network or factor analysis to produce the most contextual result.

Scrummaging, is a far more difficult task to measure quantitatively, as mentioned previously. This is an area of concern. The simplest handling involves rewarding props for each scrum they attended, assuming that their involvement in other aspects of play is influenced by the number of scrums. It is unrealistic to quantify scrummaging ability without the use of costly measuring equipment combined with knowledge of the intended action of both teams. Consequently, the rating system must be referred to as a measure of general performance.

However, considering match involvement, credit needs to be given to the amount of time forwards spend in activities such as scrummaging and mauling as this reduces their potential involvement in attacking and defensive aspects of play. However, due to limitations in data collection and the difficulty in measuring technical aspects of such activities these cannot be identified adequately. As mentioned earlier it is for this reason that systems such as the Eagle Rating can only be a general measure of performance. In the Eagle Rating performances are quantified relative to the respective positional cluster membership, this means that within-position and between-positions can be made based on the premise that the obtained ratings are based on successful involvement, which is a function of performance which is determined by an individual's ability. Thus, acknowledging that the Eagle Rating is a general measure of performance and does not possess specific technical performance measures, we are able to proceed.

Having established that on a global level attack and defence are unrelated, as there is no significant correlation between these key attributes (Midfield Backs: $r = -0.030$, $p = 0.614$) (This point will be revisited in Section 7.2). Consequently, the perfection method can be approximated by simply adding the independent variables together; the Augmented Index is calculated by simply summing the three standardised key attributes. This approximation allows faster processing of the rating. Consequently, the data is continuous and approximately normally distributed. The underlying key attributes provide the level of data that is necessary for summing up individual performance in a match. However, it is vulnerable to match volatility and involvement. This problem is off set by incorporating a discrete coding into the final calculation, thereby introducing robustness to the model.

7.1.8 Quartile Model

The quartile model (QM) uses the output from the Augmented Index to create a discrete performance measure. By making the data discrete, the model is made more robust, by reducing the influence of volatile match conditions and the outlying performances associated with increased match involvement. The underlying defence, attack and positional indices provide the knowledge of performance that is necessary to make inferences regarding ability and match performance in a coaching and selection sense.

Using the lessons gleaned from all previous methods allows the quartile model to combine the best features. Robustness is introduced by making the data discrete in a similar fashion proposed by the rule extraction of self-organising maps.

The QM involves comparing performance to a set of standards established by all comparable individuals. Four levels are adopted, making use of the observed quartiles (Well Below Average, Below Average, Above Average, Well Above Average). The terminology is intentionally loose, as it is aimed at rugby personnel. Units are accumulated for incremental increases in performance. Specifically, 'Well Below Average' performance on an attribute yields one unit, through to 'Well Above Average' which nets four units. Using three attributes with the same point structure yields a minimum score of three units and a maximum of twelve units. By subtracting two units from the final rating a rating scale extending from one to ten is obtained.

Combining the information from the augmented indices with the QM rating provides continuous data that can be used to identify small differences, whilst the conversion to discrete data about the quartiles introduces the robustness of the self-organising map. Throughout the system, context remains. That is the successful involvement of a player must be high to be rewarded with a high rating.

This section provided the necessary overview of the methods involved, detailing strengths and weaknesses and how, and why, different hybrid methods were created. The next section compares the models directly looking at the key areas of context, robustness and model validity.

7.2 Comparison of Methods

To compare the methods discussed throughout this thesis the midfield back positional cluster is used. Already, these methods have been discussed in Chapters Six and Seven with emphasis given to the subjects of context and robustness. This section uses numerical comparisons to determine which method performs 'best' based on the concepts of context and robustness. Additionally, the validity of the ratings system is examined briefly.

The methods used are as described in this thesis and summarised in the previous section. The augmented indices and QM incorporate three key attributes for this section; attack, defence and kicking (kicking metres) for the 279 performances surveyed in the previous sections. Prior to this section the augmented indices have been developed as the benchmark for the performance of the rating systems explored in this thesis. It is accepted that the indices are contextual, based on the emphasis on this aspect in their creation. Thus, it is expected that a suitable rating system will behave in a similar manner.

Firstly, the context of each of the methods is examined with respect to the input variables before validity and robustness are investigated.

7.2.1 Model Context

The most important component in compiling a rating system for quantifying individual rugby player performance is ensuring that the obtained models and subsequent output is contextual. Table 7.2 compares the correlation of each input variable with the rating value generated from each of the eight methods described previously.

	QM	*Augment Indices*	*Original Rating*	*Neural Network*	*Factor Analysis*	*SOM (1×10)*	*SOM (4×4)*	*Paired SOM*
Defence Beaten	0.31	0.33	0.14	0.29	0.46	0.56	-0.18	0.18
Errors	-0.31	-0.33	-0.08	0.14	-0.29	0.13	-0.03	0.02
Breaks	0.41	0.45	0.43	0.70	0.26	0.70	0.32	0.60
Kicks	0.46	0.63	0.54	0.13	0.47	0.26	0.51	0.51
Running Metres	0.35	0.35	0.26	0.66	0.43	0.71	0.02	0.41
Kicking Metres	0.38	0.54	0.55	0.09	0.48	0.19	0.48	0.46
Kicks Caught	0.19	0.17	0.08	0.11	0.54	0.15	0.18	0.23
Breakdown Impact	0.44	0.45	0.10	0.62	-0.07	0.24	0.40	0.41
Passes	0.32	0.41	0.39	0.61	0.11	0.57	0.41	0.59
Laybacks	0.45	0.47	0.24	0.55	0.38	0.57	0.22	0.47
Tackles	0.41	0.39	0.25	0.48	0.26	0.11	0.46	0.37
Missed Tackles	0.15	0.17	-0.38	0.24	0.31	0.11	0.26	0.17
Tackle Assists	0.25	0.25	0.08	0.22	0.28	0.10	0.44	0.32
Tries	-0.12	-0.11	0.02	0.13	0.27	0.16	-0.08	0.04
Loose Ball Gained	0.22	0.23	0.16	0.18	0.18	0.16	0.26	0.27
Harassment	0.07	0.07	0.06	0.10	0.04	0.18	0.08	0.15
Infringement	0.12	0.10	0.06	0.07	0.18	0.07	0.10	0.09

Table 7.2: *Correlations between Input Variables and Obtained Ratings*

Fundamentally, all the models behave in a manner that is expected. That is for the majority of rating systems, the favourable input variables are positively correlated with the univariate rating measure. Correlations are used given the expectation that the relationships between the variables and the output can be expected to be approximately linear, based on the desired property that the rating will increase for favourable variables and decrease for unfavourable variables.

The important distinction between the methods is those that are negatively correlated with unfavourable events (missed tackles, errors) and those that are not. The original rating, based on factor analysis is the only method that has negative correlations between all the unfavourable variables and the rating system. However, the correlation between the original rating and the number of errors is not significant ($p = 0.209$).

Three rating schemes are significantly negatively correlated with the number of errors committed ($p < 0.05$) (QM, Augmented Indices and the reworked Factor Analysis). Generally, the models are contextual as favourable variables are positively correlated with the rating scales. However, the relationship between the unfavourable variables and the rating systems is not as clearly defined. Four methods have significant negative correlations with unfavourable variables. It is from these four methods that the most contextual model will be obtained. The quartile model (QM) and the related augmented indices are contextual as this property was imposed during their creation. Additionally the original factor analysis (original Eagle Rating) created contextual models as discussed in Chapter Three. The general behaviour of the reworked factor analysis suggests that it is indeed contextual. However, it lacks the necessary key performance indicators and associated transparency for it to be suitable as a rating method that can be defended in the public arena. Whilst the QM and augmented indices appear to be negatively correlated ($p = 0.047$ and 0.067 respectively) with the number of tries scored, this is not important for the reason outlined in the creation of the indices, due to the lack of importance of the Tries in determining individual performance.

Essentially, the QM and augmented indices are part of the same proposed package. Consequently, the better of the two rating schemes needs to be decided upon – either the original factor analysis model developed in Chapter Three or the model based on the QM/augmented indices which were developed in Chapter Six in response to the weaknesses of the methods applied in that chapter.

It has been demonstrated that the 'best' ratings are significantly correlated with key events. Favourable events are positively correlated and unfavourable events are negatively correlated implying that there is some degree of contextuality within the ratings. It is reasonable to expect that contextual individual ratings will be positively correlated with match outcome. There is basis for an argument that winning teams are not always the best performing teams due to the impact of chance events and spoiling play. However, the general expectation is that the collective effort of the team must be superior to that of the opposition in order to win.

The collective effort of the team is examined to demonstrate the contextuality of the "best" models. The quartile method is not examined as this is simply a discretised form of the augmented indices. The augmented indices combined via the perfection method and the original factor analysis rating from Chapter Three are assessed. Match outcome is defined by the difference in team scores (own team score - opposition team score). This data is approximately normal (p=0.07) with mean 0 and standard deviation 19.

As rugby is a team game, individuals work together in this collective to achieve victory. Consequently, selected meta-data is compared with match outcome. The meta-data is obtained by considering all starting individuals for each team in each match to produce seven different statistics as detailed in the table below. The correlation between the meta-data per team from each of the 69 Super 12 matches played in 2000 and the match outcome are examined in the table below.

Meta Data	*Augmented Indices*	*Original Rating*
Median	0.224	0.247
Mean	0.312	0.343
Q1	0.260	0.314
Q3	0.289	0.223
Min	0.182	0.185
Max	0.355	0.187
Sum	0.314	0.208

Table 7.3: *Correlations between Match Outcome and Meta-Data from Augmented Indices and the Original Eagle Rating*

Table 7.3 shows that there is significant positive correlation between match outcome and the meta-data (from the individuals per team in each game). Although this correlation is weak it supports the contextuality of these two models. The positive correlation implies that generally teams that play collectively better tend to have larger winning margins. This is what arm chair critics would expect.

Another important consideration relating to these systems is the apparent validity. This section has shown that all the methods produce respectable output. Consequently as all the methods are trying to estimate individual rugby player ability and performance, it is reasonable to expect that there is congruence between these methods. This is examined in the next section.

7.2.3 Model Validity

As each of the methods described previously is an estimate of individual rugby player performance, it is sensible to assume that methods will be highly correlated. The correlation between the different methods provides the data for Table 7.4

	QM	Augment Indices	Original Rating	Neural Network	Factor Analysis	SOM (1*10)	SOM (4*4)
Augmented Indices	0.90						
Original Rating	0.47	0.52					
Neural Network	0.65	0.66	0.33				
Factor Analysis	0.61	0.66	0.32	0.40			
SOM (1*10)	0.48	0.54	0.31	0.69	0.48		
SOM (4*4)	0.55	0.62	0.42	0.49	0.39	0.29	
Paired SOM	0.66	0.73	0.52	0.70	0.53	0.63	0.86

Table 7.4: *Inter-Rating Correlations*

All the systems are significantly correlated ($p < 0.05$), suggesting that the essential elements of these systems are shared. As the context of the results has been shown in the previous section and through the process of constructing the rating output, the information overlap exhibited by the methods implies that the obtained results are valid. Essentially, this thesis has sought to quantify the principal curve/surface/hyper surface that governs player performance. The correlation exhibited between the methods suggests that such a principal curve/surface/hyper surface exists. This defends the use of multivariate dimension reduction techniques to extract measures of individual rugby player performance.

Table 3.11 validated the approach adopted in Chapter Three by comparing the 22 individuals with the highest average Eagle Rating with those selected to participate in the Tri-Nations tournament of 2000. International selectors must be considered the foremost authorities in judging player talent. That approach is repeated in Table 7.5 using the augmented indices. In order to qualify for ratings individuals needed to participate in three or more matches in the Super 12 tournament 2000.

Rank	*Name*	*Mean AI Rating*	*N*	*International?*	*Positional Cluster*	*Mean Eagle Rating*	*Eagle Ranking*
1	Rod Kafer	72.94	12	Australia	Midfield	62.04	7
2	Stirling Mortlock	70.25	13	Australia	Outside	54.81	
3	Ron Cribb	69.41	13	New Zealand	Loosies	53.48	
4	Todd Miller	68.97	6	---	Outside	61.75	8
5	Steve Larkham	68.33	13	Australia	5 Eighth	63.45	6
6	Joe Roff	67.51	12	Australia	Outsides	68.98	1
7	Christian Cullen	67.21	11	New Zealand	Outsides	68.15	2
8	Willie Meyer	66.66	12	South Africa	Prop	60.05	11
9	Selbourne Boome	66.24	5	South Africa	Lock	51.98	
10	Pieter Rossouw	65.29	11	South Africa	Outside	66.67	3
11	Jonah Lomu	65.18	6	New Zealand	Outside	64.79	5
12	Chris Latham	64.64	10	Australia	Outside	66.19	4
13	Todd Blackadder	64.18	14	New Zealand	Lock	55.79	
14	Mark Mayerhofler	63.58	9	---	Midfield	55.68	
15	B. Van Straaten	63.32	10	South Africa	5 Eighth	46.06	
16	Byron Kelleher	63.24	12	New Zealand	Halfback	58.07	22
17	Kees Meeuws	63.21	9	New Zealand	Prop	58.75	17
18	Taine Randell	63.18	8	New Zealand	Loosies	55.61	
19	Charles Marais	63.17	10	South Africa	Hooker	48.65	
20	Elton Flatley	62.55	9	Australia	5 Eighth	50.17	
21	Paul Tito	62.47	10	---	Lock	45.83	
22	Mark Hammett	62.23	13	New Zealand	Hooker	49.71	

Table 7.5: *Top 22 Performers in the 2000 Super 12 According to Augmented Indices*

Nineteen of the top 22 performers as judged by the augmented indices were selected as internationals. This agreement suggests that there is an overlap between the selectors' perceptions of performance and the ratings.

A core component of a contextual system is the inherent robustness. This is investigated in the next section along with other properties that are desirable in a rating system.

7.2.3 Model Robustness

As mentioned repeatedly throughout this thesis, the rugby match environment is highly changeable. Individuals are exposed to many different conditions and constraints that need to be satisfied adequately in order to maximise the team's chances of a favourable outcome. These different conditions and constraints do affect individual involvement and subsequently impact on the generated match statistics. With robustness comes a loss of information. This cannot be helped and is part of the process, ensuring a credible platform is obtained. Thus for a rating system to be robust it must not be unduly influenced by changes in match requirements. However, this match volatility is very hard to measure with any ease. Comparing the rating output with the time spent on field is the easiest way of determining vulnerability to changeable match involvement. This is displayed in the final two columns of Table 7.4.

Method	*MHM Value*	*Normality (p-value)*	*Correlation with Time*	
			r	*p-value*
QM	0.021838	0.000	0.185	0.002
Augmented Indices	0.066354	0.191	0.190	0.002
Original Rating	0.035633	0.105	0.169	0.006
Neural Network	0.062126	0.007	0.277	0.000
Factor Analysis	0.06542	0.001	0.094	0.122
SOM (1*10)	0.038321	0.000	0.222	0.000
SOM (4*4)	0.04289	0.000	0.165	0.006
Paired SOM	0.056078	0.000	0.229	0.000

Table 7.6: *Comparative Performance Measures of Ratings*

The rating system is independent if the multivariate half moon statistic is less than the critical value of 0.0112 at the 5% level of significance (k=250). All methods demonstrate significant association ($p < 0.05$) between all the variables and the estimate of performance, as shown in the 'MHM Value' column of Table 7.6.

Additionally, it is desirable that the obtained rating output is approximately normally distributed such that it can be applied in a performance monitoring scheme using control charts that are dependant on the assumption of normality. Table 7.6 indicates that only two methods produce approximately normal ratings, the augmented indices ($p = 0.191$) and the original factor analysis rating ($p = 0.105$). Coincidentally, these methods were found to be the most contextual from section 7.2.1.

Almost all the measures are significantly correlated with time ($p < 0.05$) although this relationship is weak ($0.165 < r < 0.277$). The reworked factor analysis is not correlated with time, suggesting that this is the most robust method. However, the composition of this method was uncontextual due to inappropriate contrasts between favourable variables. It is the uncontextual basis that has resulted in the analysis being uncorrelated with time. That is the effect of increased involvement is negated as for some factors favourable variables have a negative weighting. So, a little correlation is good, but not too much, it seems. The handling of time is highly debatable and numerous first class coaches and players have expressed contrary opinions. A conservative approach has been adopted in this thesis where ratings are awarded on the basis of an individual's actual contribution. As more data becomes available in each relevant competition, the impact of partial match play can be studied in more detail.

Essentially, the involvement will tend to increase based on involvement, and this is by our design. The loss of robustness is due to the desire to create contextual output, where ratings increase with positive involvement and decrease with negative involvement. Essentially, the effect of ratings increasing with positive involvement cannot be avoided. The effect of increasing positive involvement can only be dampened by using methods such as the quartile model where non-parametric techniques induce discrete data based on a series of benchmarks.

The desire to induce contextuality in the model has come at the expense of robustness. Robustness of the system can be viewed as a function of context. However, contextuality is more important than robustness when creating the model. Importantly, the correlation with time of the contextual models was weak and therefore variation for time played is not too much cause for concern. Attempts to address the involvement issue by correcting with a time factor [Raw Rating × (80 minutes / Time On-field)], provided worse results (Time corrected QM $r = -0.367$). Further, if the general ratings that are obtained are viewed as an indication of the contribution to the team, then the weak relationship with time is acceptable. Obviously, an individual cannot contribute to the same extent in a shortened time frame. Accordingly, why should an individual be rewarded on a comparative level for a partial performance as an individual who has participated in the entire match? This reinforces that the weak relationship with time is acceptable.

Given the conservative approach adopted there will be some correlation with time. However, it does imply that as involvement increases the ratings will tend to increase. Therefore of the two methods that have shown contextual, valid results (QM/augmented indices and original factor analysis), the QM/augmented indices are most appropriate due to the introduction of robustness by protecting against increasing match involvement by the use of quartile coding, so that reasonable limits are placed on what an individual can score.

7.3 Discussion

The major differences between the forms of artificial neural networks explored in this thesis are the underlying mechanics and resultant output. By comparing these basic elements the strengths and weaknesses of each model can be identified.

Factor analysis produces orthogonal elements that summarise the key components of the data source presented, such that the interpretation of each factor leads to the identification of Key Performance Indicators. An additional step (perfection method) is required to induce a single output variable that relates to a measure of individual performance in each match, as demanded by the consumer. The nature of the performance measure is continuous, formed by a series of weighted linear equations.

Similarly, the output produced by a self-supervising feed-forward back-propagation neural network is essentially the same, although it is not restricted to linear mappings of the data. The KPI's are obtained from the bottleneck node of a self-supervising network, with the additional step of collapsing the bottleneck data to a solitary output, via the perfection method, required to satisfy demand for a univariate measure of performance. Again, this output is continuous, although the derivation is more complicated than that expressed by the factor analysis. Through the dimension reduction process data are transformed by an activation function, which, as discussed earlier, gives neural networks the ability to mimic non-linear functions and squashes extreme values, dampening the impact of outliers. The nature of the obtained performance measure is continuous which allows small differences in player performance to be detected.

With both the factor analysis and self-supervising feed-forward backpropagation neural network a further loss of information occurs when the defined KPI's are reduced to the single performance measure, unlike the self-organising map where only one output per input pattern is produced. Essentially each performed task is given a weighting, depending on the perceived importance. That is each time a task is performed it results in the addition, or subtraction, of the cumulative sub-rating points that comprise the final rating measure. The resultant output is vulnerable to match volatility, removing the required robustness from the model. However, provided enough variables are incorporated into the model, an element of stability is introduced to the rating system due to the expectation of relative involvement by the individual. The amount of match play provides the key source of match volatility, and is most extreme in the linear case presented by factor analysis. Consequently, the tendency for values to increase based on positive match involvement is not as problematic as claimed. Provided that there is adequate marketing support and explanation, the nature of the rating system can easily be explained to the rugby public.

The self-organising map induces a solitary discrete output. Consequently no KPI's are obtained. However, in producing the one output from the input pattern, a higher portion of the data is explained, reducing the loss of information experienced by the back propagation networks. The KPI's are important for advanced users and this will be shown in Section 7.5.

The best methods identified were the original factor analysis and the QM/augmented indices. The QM/augmented indices model was deemed best because of the protection against strictly increasing match involvement using the quartile coding based on the quartiles. Additionally, the summarised variables used lead to a simplified data collection process which in turn reduces the time taken to collect the data.

The previous section illustrated that the robustness of the ratings was lost due to the emphasis on contextuality. Through the development and implementation of the methods described in this thesis careful consideration was given to these important aspects of model structure. However, for sports data this is a difficult balancing act that has not been adequately fulfilled by the methods employed. However, as the loss of robustness is due to the increase of positive match involvement is not as problematic as first thought and the resultant output can be defended based upon the transparency of the models, which is due to contextuality.

7.4 Implementation of Quartile Model

The augmented indices that are the foundation for the QM are used to extract estimates of individual rugby player performance for use in the following sections. Reducing raw match footage to a singular numerical measure of performance is useful for gauging an individual's impact upon a match. This information can be adopted by the media to support the value of particular individuals. However, the power of this novel statistic is increased when the latent variables used in calculating the Eagle Rating are produced in analysis of match data. This increased power is necessary for coaches and selectors to better understand the nature of a performance.

7.4.1 Implementation in the Media

A primary audience for a rating system is the public. The best way to reach observers of rugby is through the mass media represented by print and television. The demands for a rating scheme aimed at the spectators and presented by the media are less than those required by coaches and selectors. Nonetheless, the requirements are important and are a subset of the needs of coaches and selectors. The key concepts of context and robustness have been well established by now. Implementation is another important aspect when dealing with the media. The processes involved in computing a rating system must be such that deadlines for the media can be met. There is little use in a rating system that cannot be exposed to the target audience when required.

For ease of understanding, the maximal information must be imparted in the simplest manner. For the needs of the media a solitary performance measure for each individual in each match is all that is necessary. Support provided in the form of interpretation resulting from the underlying key performance indicators is also necessary to assist in explanation and define the context of the ratings, so that the rugby public understand the information presented. Examples of media applications are provided in Appendix D.

7.4.2 Implementation in the Coaching Environment

The challenges confronting the coaches and selectors are different to the media and an entire section is dedicated to this important aspect.

The amount of data a coach or a selector needs to review to understand individual performance can be greatly reduced by implementing statistical procedures. The flowchart on page 278 illustrates the process underlying the calculation of a rugby performance rating for inferring individual rugby player ability.

Through the dimension reduction process relatively stable data is obtained relating to the performance level of individuals. This stability is due to the involvement of many variables and consequently reduces the impact of match volatility. Accordingly, statistics such as the Eagle Rating can be employed in a diagnostic coaching structure as shown in the following section.

7.5 Implementing Statistics in a Diagnostic Coaching Structure

Statistics are a natural by-product of competitive sport, for in many instances this information is used to determine match result (runs, goals, points, time). Encapsulated within the resultant match data are gems relating to match performance. Extracting this information can lead to valuable insight into the performance of individuals and their relative capacity or ability. This section will explore the use of statistics relating to individual performance from a statistical perspective, relating sound theory and methodology to enhance the information obtained, enabling 'real' problem areas to be identified. The basic approach applied is applicable to many sports; however, in this instance rugby will be used as an example, with reference to the commercially developed Eagle Rating.

7.5.1 Basic Philosophy

The philosophy adopted is simple. Instead of monitoring univariate, or single task orientated variables such as tackle counts, a measure of overall performance - in this case the Eagle Rating - is calculated and monitored. The measure is comprised of Key

Performance Indicators (KPI's) which are effectively a summary of the single task variables for each match. The match performance measure is monitored, over a player's career, and this provides insight regarding the potential of the player. Given that a problem is identified, the cause is found by establishing the abnormal KPI and from this the deficient skill is detected. Further, the actual event can be established from the skill grouping in many cases. It is then up to the coach to decide whether or not the problem requires remedial intervention. The basic structure required to implement statistics in a diagnostic coaching structure, as described, is outlined in Figure 7.5.

Due to match volatility, issues associated with non-performance and fluctuating individual involvement, stable statistics must be obtained, before any useful trend analysis or diagnostic interpretation can occur. Statistically sound data is achieved through considering holistic performance over a series of matches; that is the overall involvement an individual has had in a match provides the starting point for analyses.

For each match a stable indicator of performance is created from the separate task variables with minimal information loss. This forms the calculation stage, shown in Figure 7.5. From a coaching perspective it is the performance monitoring stages and beyond that are of interest in Figure 7.5.

In order to implement statistics in a coaching diagnostic structure four steps must be satisfied prior to the calculation of statistics that provide the basis for any reliable inference:

1. Define Performance
2. Operationalise Performance Definitions
3. Quantify Match Involvement via Operationalised Measures
4. Calculate Performance Measures - a) KPI's,
   b) Overall Performance

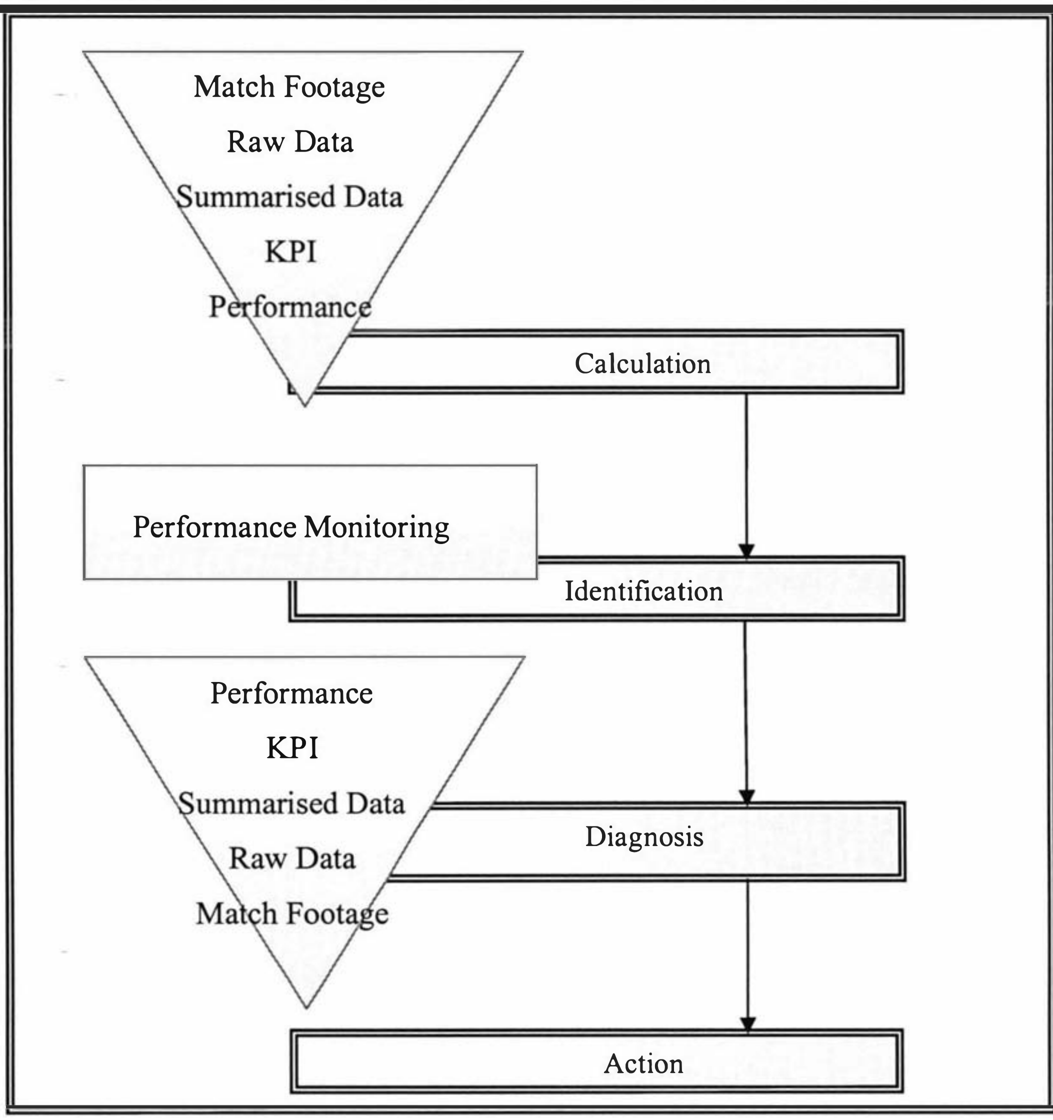


Figure 7.1: *Flow Chart Displaying the Four Steps in Applying an Individualistic Rating System as a Diagnostic Tool*

These steps are necessary to provide a stable foundation from which performance can be assessed and monitored. Because of the differing circumstances confronting an individual in a sporting contest, there will obviously be differences in performance. The previous four steps allow performance to be quantified. This information, coupled with information obtained from prior performances allows the match behaviour of an individual to be monitored using statistical techniques such as control charting procedures. These procedures are designed to detect real (significant) changes in performance over time and are demonstrated in Section 7.5.5.

Given an overall measure of performance on a match-by-match basis, sequential points can be plotted graphically using EWMA or Shewhart control charts (Montgomery, 1997). The expected capability of an athlete, defined from prior performance, is used to set process limits for these charts. From this graphical display of performance, changes are evident if the process limits (expected capability of an athlete) are exceeded or additionally in the case of the Shewhart chart one or more of the run rules are violated. If no significant changes in performances are identified this indicates that an individual is playing to his/her expected performance level. Given this performance level is acceptable, action required is minimal. However, if a change of performance is evident, this can be due to one of three scenarios: player's performance has improved, player's performance has worsened, performance is stable and unstable match conditions have caused a false alarm. Given the available chain of evidence (Performance Measure, KPI's, Task Data, Specific Event) it is then up to the coach to decide on the appropriate course of action. A hypothetical example is used to illustrate this point. A deficient performance measure is detected, and investigation reveals that this is due to an inferior defensive performance. The defensive attributes are examined, and it is clear that tackling was the problem; specifically too many tackles were missed. Assessing the missed tackles it is found that the missed tackles occurred when the player was attempting to use the left shoulder to drive in and make the tackle. The coach's task is to decide whether or not the individual needs remedial coaching to correct this deficiency, or if extenuating circumstances caused the lapse, and act accordingly.

This essentially summarises the processes involved with incorporating statistics into a diagnostic coaching structure. The reasoning behind the adoption of such methodology is expanded upon later. Having outlined briefly the basic philosophy and goals of performance monitoring, it is important to establish the applicability of statistics in a coaching environment.

7.5.2 Performance Applicability

Statistics are applicable in the rugby environment, as outlined by the New Zealand Rugby Football Union in the Coaching Accreditation Manual: Level 3 (NZRFU, 1991). The second module of this publication, Teaching Skills and Skills Analyses: Analysis of Performance (pp. 9-15), discusses the benefits of statistics in a rugby coaching environment which are summarised in the following two paragraphs.

Assessment based upon only qualitative analysis can be distorted by specific events such as: injuries, current score, player's decisions, referee rulings and the most recent "calamity" or accomplishment, based upon real time and live input collection. In addition, recall of events is impaired by a number of factors, described briefly as follows. Firstly, it is impossible for coaches to absorb all the events of a game. Further individuals have limited recall of what has occurred and may struggle to put events into context, and rare events may not be recalled at all. Finally tension, emotion and personal bias significantly affect the retention of relevant data.

The need for use of more reliable quantitative analysis is highlighted by the fact that "improvements in performance are related to the quality of feedback given to players after a game; this is most effective if given as soon as possible". Consequently, "performance may not be improved if there are flaws in this feedback. This is particularly applicable to peripheral play away from the ball which involves most of the players most of the time. Any feedback the coach gives to players is based on partial information because it is impossible to see all players at once and not be distracted by play with the ball (NZRFU, 1991, p. 10)." Further, social phenomenon such as social loafing can probably only be rectified through the use of statistics that identify and attribute performance to individual players (Greenberg & Baron, 1997).

The improved development and subsequent reliance upon video analyses (incorporating digital footage and software enabling instant recall of events) to recall events can remedy a number of these problem areas in post match analysis, and statistical analyses can help coaches to pinpoint those portions of game play that need consideration.

Assessing the capabilities of an individual with respect to prior performance identifies the aberrant behaviour. Statistical techniques define performance parameters, based on prior performance of many similar individuals (defined by positional responsibilities). Match performance is then monitored with respect to these parameters, allowing trends in performance to be identified. The keys steps in creating such a process are crucial in developing a structurally sound platform on which to base inferences regarding player performance.

7.5.3 Creation of Stable Statistics

Statistically, an individual's ability cannot be inferred from a single match. The performance of an individual must be monitored over several matches, depending on the level of significance required. From a series of performances, trends in player performance can be established as well as the associated strengths and weaknesses. More importantly, given a series of matches ability can be inferred. As mentioned previously, individual "ability is inferred from performance, which is inferred from the combination of the successful and unsuccessful completion of physical tasks." In order to complete a successful physical action in a game situation, the associated mental tasks must also be completed successfully. This is an important assumption associated with applying statistics to rate individual performance. Earlier chapters explored the definition of performance and ability (Chapter Two) and the subsequent quantification in greater depth than is necessary here.

Sports performance is multifaceted. That is performance on many relevant, and often related, tasks is required to compete effectively. Consequently, relative performance on each attribute must be considered. However, in a team situation, or even in a direct contest with opposition, certain tasks are performed rarely due to match constraints. Match conditions, constraints and the subsequent impact on individual involvement are crucial in statistical analyses. In every match, individuals are exposed to volatile match conditions. Given the rugby example, an individual may not make many tackles because his/her team has dominated possession. Consequently it is inappropriate to determine level of performance through univariate statistics such as tackle counts because tackle counts may not give a full impression of defensive performance or overall performance, given the match structure.

In order to combat match volatility, multivariate statistical techniques provide the ideal tool in summarising performance. Given that performance on separate related tasks is due to some core ability (attack, defence, kicking and so forth) inferences regarding performance relating to that core ability can be extracted using multivariate statistical techniques. The most suitable statistical techniques involve dimension reduction, where multiple variables are reduced to a smaller number of variables based on the overlap of information between variables. The vulnerability of solitary variables to match volatility is reduced by the combination of several related variables. Dimension reduction techniques such as factor analysis, neural networks and self-organising maps can be used to extract statistically sound KPI's, based upon the inherent structure of individual performance created from many individuals in similar positions over many games at a relative level. There is the potential for each level of rugby (international, provincial, club, schoolboy) to have a different set of KPI's, due to the structure of the game and skill level at each level.

Finally, the separate core skill components or key performance indicators for a single player in one match can be combined to produce a solitary measure of performance using multivariate process control procedures, and averaging over several matches provides a measure of player performance.

7.5.4 Incorporating 'Quality Control' Into a Diagnostic Structure

At this stage it must be stressed that observance of individual performance through purely objective statistical measures is just a tool for identifying level of performance. Utilisation of such a technique needs to be balanced with qualitative measures and subjective assessment performed by the coaching staff. The label, 'performance monitoring', implies a less regimented approach than 'quality control'. This phrasing has important implications on how a resultant statistics package is perceived and therefore implemented. Naming of the procedure aside, quality control is the ideal tool for assessing individual sporting performance as discussed in Section 3.6.

Control charting procedures are an ideal tool that allows coaches, selectors and other interested parties to quickly evaluate the form of an individual. However, in the sporting situation it is more desirable to be 'out-of-control' on the upper side than being average. Consequently, a shift in thinking is required to accommodate the change from the attainment of mediocrity to brilliance. In fact, the interest is with the deviation from what an individual is expected and capable of accomplishing, given parameters attained from past performance. In order to achieve this, comparison is made with perfection.

The control charts associated with quality control provide an excellent medium for displaying performance and changes in performance visually. Not only can performance monitoring be used to indicate potential changes and causes in individual performance, control charts also enable an athlete's response to changes in training structure, coaching style and so forth to be monitored. Assessing the team as a collective using the average performance rating enables coaches to be made more accountable, as changes in the collective unit can be attributed to the involvement of the coaching staff and related programmes.

However, over dependency on statistics can promote coaching in hindsight whereby awareness of potential problems is identified after the event (B. Bracewell, 2001). Thus it is important to keep the use of statistics in perspective. Through the summary of match footage, information is lost, regarding certain technical, biomechanical and decision making aspects (human elements). Consequently, statistics can never be more than just a tool in the sporting environment. The strength of this tool can often be the weakness due to the objectively defined nature, and is thus devoid of emotion. Rugby is chaotic given that actions are defined by some initial conditions, match context must be taken into account when assessing performance.

Intuitive coaches are able to define performance, however due to human limitations recall of all events is impossible. Correctly defined and implemented statistical process provides the necessary "notes" to assist coaches in the recall of events and diagnosing of faults. Incorporating performance monitoring into a coaching structure allows certain aspects of play to be highlighted.

Importantly, sport statistics can be used to reinforce the beliefs of an intuitive coach or augment deficient areas of awareness. Implemented in the correct manner, statistics can suggest where a coach should look, to remedy certain weaknesses or provide ideas on how to maximise strengths.

7.5.5 Application of Control Charting Procedures

As an illustration of the control charting procedure, a Shewhart control chart illustrating the form of Otago Highlander halfback Byron Kelleher from the Super 12 Tournament, 2000 is provided in Figure 7.6. Control limits are set two standard deviations away from the mean due to reduced sampling opportunities as discussed in Chapter Three.

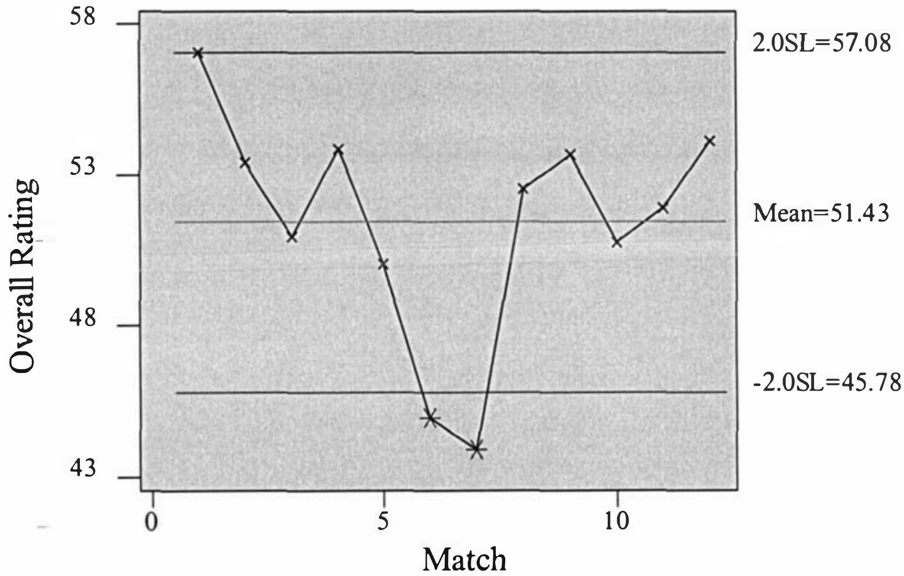


Figure 7.2: *Shewhart Chart of Kelleher's Overall Rating by Match in the Super 12, 2000*

For this tournament, Kelleher provided two performances that are statistically abnormal. Firstly, a well below expected rating in match 6 is obtained from the 34-15 loss to the ACT Brumbies; and secondly another less than expected performance is produced in match 7 where the Highlanders were beaten by the Auckland Blues 26-16. Assessing the underlying KPI's indicates that the problem in both these matches is due to less than expected performances associated with a measure of handling. Further investigation reveals that this deficiency was directly associated with the number of successful passes delivered. Evidence suggests that this is not a technical problem, but is related to the relative involvement of Kelleher. In both matches, the Otago Highlanders were dominated by the Brumbies and the Blues. As a consequence Otago received very little possession, leaving Kelleher with little ball to deliver to his backs. Because the overall rating considers overall performance and Otago lacked possession, this causes reason to question current All Black Kelleher's defensive involvement. It is reasonable to expect that if the attacking opportunities are reduced then the defensive opportunities are increased. However, considering team statistics it is found that a team-mate is responsible for Kelleher's reduced ranking. Former All Black and openside flanker Josh Kronfeld carried Otago's defensive effort in these games. Consequently, there is just cause underlying the abnormal performances presented by Byron Kelleher that is not related to technical deficiencies or reduced work-rate.

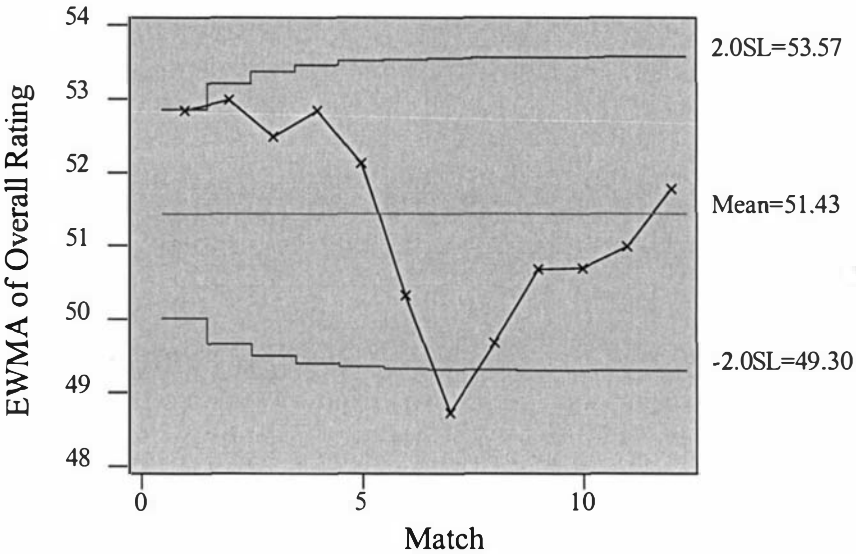


Figure 7.3: *EWMA of Overall Rating For Kelleher by Match in the Super 12, 2000*

Similarly, using an EWMA control chart with $\lambda = 0.25$ (Bracewell, 1999) allows trends in form (performance) to be quickly identified. One alarm is signalled, associated with match 7. Investigation of the reasoning underlying the cause of this alarm is due to the circumstances described in the previous paragraphs. Apart from the aberrant observation associated with lack of possession, the EWMA suggests that Kelleher performed to his expected performance level.

7.6 Conclusion

Three core methods for dimension reduction have been assessed, namely factor analysis, self-organising maps and self-supervised feed-forward neural networks. The benefits and drawbacks were discussed with particular reference to the output of the rating system. In the sporting context, none of the methods were suitable as a stand-alone method for quantifying individual rugby player performance. However, by taking the best features from each of the methods, the philosophical requirements necessary to create a sport rating system were obtained. This required the highlights of each method to be melded into a hybrid method, the quartile model. Consequently, it is possible to incorporate data mining techniques into the sporting environment.

Comparing the methods it was found that generally the different models produced contextual results on a global level. Further, the congruence between the methods implied that the same core performance surface was estimated. This indicated that the results were valid based on the high correlation between methods.

Contextuality came at the expense of robustness. The construction of the methods is such that this cannot be escaped. Such an effect can only be dampened by selecting a method that has mechanisms for coping with the positive increase in match involvement incorporated. Consequently, the QM/augmented indices component system for quantifying individual rugby player performance was adopted. However, the issue of increasing ratings with the direct increase can be defended with appropriate marketing and education.

Finally, it was illustrated how the estimates of performance can be employed in a coaching environment. As sport presents many variable conditions, which affect match statistics, the techniques employed to analyse match statistics must take into consideration the natural variability presented by different matches. Consequently, the Eagle Rating was developed as a relatively stable measure of match performance. The underlying philosophy is taken from statistical process control, which was developed for ensuring manufactured goods meet specifications. However, with slight modifications, the control charts and techniques for ensuring processes are under control can be applied to monitor match performance and subsequently identify any changes in form, and associated strengths and weaknesses. Because athletes strive for perfection, not the average, standard process control techniques must be adjusted to compare performance to perfection, rather than the average. We do not expect athletes to perform 'in-control'. However, we are interested in changes in performance. Realistically, performances are going to differ from match to match due to different match constraints and conditions. Consequently, our interest is in identifying statistically significant changes in performance and identifying the associated causes. The changes may be either extrinsic or intrinsic. The coach can readily identify this by isolating the component (KPI) that caused the change in the overall measure of performance and identifying the cause of that change in the actual match footage. Whilst statistics have an important part to play in sport, their use must be contextual, and methodologies employed must be statistically sound and firmly supported by qualitative analysis. Given correct usage statistics can provide valuable insight into individual performance enabling strengths and weaknesses to be diagnosed.

Chapter Eight

Conclusion

This thesis has sought to quantify individual rugby player performance. An individual's performance is a sample of an individual's ability; however, it is dependent on a number of match constraints and variations. This study was driven by commercial interest and required that a single match rating be developed that described an individual's performance using data collected by Eagle Sports on a match-by-match basis. However, it has achieved much more than this. It has provided a procedure for developing rating systems using cutting edge technology, and a new tool that allows these ratings to be interpreted has been created. The importance of the end user in a rating system is emphasised, and the issues associated with interpretation and implementation are explored at length. This study has shown how, in principle, rating systems should be designed.

8.1 Overview

The early stages of this thesis outlined the applicability of statistics to sport. Due to the limitations of the human observer, statistics are an important tool in providing an unbiased, objective overview of a match. In order to create a suitable numerical summary of match events the right data needed to be collected. The data collection process and systems needed to ensure quality data are obtained were discussed only briefly, to protect the commercial aspects of Eagle Sports' core business.

8.2.1 Research Questions

No previous attempts have been made to quantify individual sporting performance using multivariate data. So the thesis has broken new ground. In particular this thesis answers the following research questions:

- Is it possible to measure individual rugby player performance in a team environment?
- How can the separate components (KPI's) be combined to provide an effective overview of performance?
- How can the system of ratings be constructed in order to achieve transparency, contextuality and robustness?
- How can a measure of individual rugby player performance be employed appropriately by interested parties?
- What is the relevance of this work for other rating systems, in particular, for rating systems in other sports?

8.2.2 Research Answers

Due to the commercial involvement and application, the first performance rating that was developed is called the Eagle Rating. This novel sport statistic represents an estimate of individual rugby player performance on a match-by-match basis, calculated by reducing the dimensionality of multivariate data. The estimate of performance is defined by the observed physical tasks (or lack thereof). From the successful completion of a physical task it is inferred that the other tasks (such as physiological and psychological) required to complete the physical task have also been successfully completed.

Importantly, the roles of individuals in the game of rugby differ greatly, based on their position, skill set and game situation. To account for this between-individual variability, individuals were separated into relevant positional clusters. This division was based on expert opinion after it was found that clustering techniques were unable to separate meaningful positional clusters. Due to the overlap of skill sets and job requirements, there is no strictly definable boundary expressed in the match statistics, therefore making clustering techniques inappropriate.

The initial pilot study used a subset of variables from 93 match variables for each player. As the data was highly correlated, dimension reduction techniques were applied to reduce redundancies in the data in each of nine distinct positional clusters to identify the latent factors. Factor analysis using an orthogonal varimax rotation produced five factors that summarised individual rugby player performance, explaining more than 60% of the original variability in the data set for most of the nine clusters. Orthogonal rotations were favoured, as this tended to increase the coverage of the variables, which reduced the potential impact of match volatility. The five factors scores were reduced to a single value, the Eagle Rating, using a Mahalanobis distance. The components of individual performance (factor scores) were compared with perfection, rather than the average as typically specified in quality control. The overall individual performance value was immediately adopted by the fantasy game *Ultimate Rugby* as the scoring mechanism for which individuals accrued points. Later, the Eagle Rating was used to determine the man of the match in the Sky Television Man of the Match competition. The relative lack of robustness possessed by the linear-based factor analysis was exposed at this point with ratings tending to increase with respect to match involvement. This initial rating system also had some problems in terms of contextuality and transparency. The work that followed addressed all these problems, while retaining the ease of implementation.

The original Eagle Rating based on factor analysis provided the necessary comparison with which to compare other techniques for dimension reduction. Two types of neural networks were explored, self-organising maps and self-supervised feed-forward neural networks.

In order for the neural networks to produce a transparent, contextual rating system, it was necessary to develop a method for interpreting the underlying actions of these networks. A new heuristic, the half moon statistic, was therefore developed for relationship detection. This statistic was defined mathematically allowing the evolution of first a parametric, and then a non-parametric test for independence. This was later expanded to the multivariate case. The distribution of the test statistic under the assumption of independence was approximately normal enabling p-values for signifying influential relationships to be obtained, which enabled the new statistic to be implemented as a stand-alone technique for determining significant input-output relationships.

This information allowed the association between input and output variables to be tested for significance, leading to a greater understanding of the holistic activity of the networks. This contributed to the attainment of a contextual rating in Chapter Six.

The shift to methods capable of non-linear dimension reduction showed that the seemingly large Eagle Sport database was too small to employ neural networks without the risk of over-fitting. Consequently, the data set was summarised from 93 variables to just 19, increasing the ratio of input patterns to adjustable parameters to almost one, which was still short of the recommended five to ten observations per observational parameter. However, because contextual results were obtained from the networks in Chapter Six, this suggested that the performance component of the data was approximated, not the stochastic component, indicating that overfitting had not occurred.

Three core methods for dimension reduction were assessed; namely factor analysis, self-organising maps and self-supervised feed-forward neural networks. Generalisability was introduced to the self-supervised feed-forward networks by employing early stopping. Backpropagation was used to train the four layer networks with the number of nodes in the hidden layers specified by trial and error. The SOM performed on standardised data explored both one and two dimensional topologies with the desire to create an ordinal rating.

The benefits and drawbacks of these methods were considered with particular reference to the output of the rating system and this lead to the creation of a hybrid methodology. It was shown that in the sporting context, none of the methods were suitable as a stand-alone method for quantifying individual rugby player performance due to limitations ranging from meaningless key performance indicators to lack of robustness. However, by taking the best features from each of the methods, the philosophical requirements necessary to create a sport rating system were obtained. This was achieved through the use of a quartile model in conjunction with the most contextual results from the factor analysis and neural networks, for the key areas of attack and defence. The use of the hybrid method indicated that it was possible to use data mining techniques in the sporting environment to create a transparent, contextual and robust rating system for individual rugby player performance answering the first three research questions.

Importantly, it was illustrated how the estimates of performance can be employed in a coaching environment. As sport presents many variable conditions, which affect match statistics, the techniques employed to analyse match statistics must take into consideration the natural variability presented by different matches. Consequently, the Eagle Rating (MARK II) was developed as a relatively stable measure of match performance. The underlying philosophy is taken from statistical process control, which is commonly used for ensuring manufactured goods meet specifications. Whilst athletes are not expected to perform in control, control charts provide an ideal medium for viewing and identifying trends in player performance. Performances differ from match to match due to different match constraints and conditions. Consequently, it is desired that statistically significant changes in performance and the associated causes are identified. These changes may be either extrinsic or intrinsic. The coach can identify the significant disturbances in form by isolating the component (KPI) that caused the change in the overall measure of performance and following the data trail to identify potential causes in the actual match footage. Whilst statistics have an important part to play in sport, their use must be contextual, and methodologies employed must be statistically sound and firmly supported by qualitative analysis. Given correct usage, statistics can provide valuable insight into individual performance enabling strengths and weaknesses to be diagnosed. This answers the fourth research question.

This thesis has outlined the necessary steps to create a successful rating system using multivariate data. Additionally, the problems and remedies, many of which are not rugby specific, are also explored. Finally, this thesis is relevant to other rating systems as the concepts of transparency, context and robustness are required to enable the end user to understand the 'expert system', answering the final research question. Importantly, the need for the audience to understand the output is crucial. Creators of a rating system cannot hide behind the mechanics of the system and must be able to defend/justify the output.

Furthermore, significant commercial interest in the methodologies has been expressed by companies from a diverse range of fields (Manufacturing, Information Technology and Sport Science) highlighting the relevance of this research. Due to confidentiality issues specific details cannot be provided, however, each company seeks a rating system that will give them a competitive technological edge in their relevant commercial sector.

8.3 Implications

The results of this thesis are exciting given the potential use by both coaches and the media. Contextual data (KPI's) with a lower dimensionality can easily be displayed, enabling visual comparisons (within-individuals and between-individuals) to be made rapidly. This is useful to the media for communicating vast amounts of information in the simplest possible way aiding the entertainment package of professional sport which is supported by media coverage.

The use of statistics in a sporting culture was also discussed. Properly implemented statistical procedures will improve the quality of the secondary information available to coaches and selectors. The ratings and KPI's that comprise the secondary information are important because they 'smooth' the inherent match volatility to which univariate statistics are vulnerable. The problems associated with univariate statistics due to match volatility were discussed, which highlighted the need for 'stable' data, which is provided in the form of the overall performance measure obtained via multivariate techniques. Furthermore the time spent exploring a database is minimised by pinpointing the possible areas of anomalous behaviour.

Adages such as 'form' and 'consistency' are referred to based on humanistic experience and observance of events, which may be flawed as discussed in Chapter Two. Consequently, a statistically sound objective rating system can be used as supplemental evidence to either confirm or rebuke claims regarding trends in performance. This is useful for settling the 'Monkley/Leslie' type debates mentioned in Chapter One.

As individuals are assessed in the competitive environment, confounding variables that may hamper skill testing are incorporated into the performance measure. Thus more meaningful analyses may be obtained from field testing than laboratory testing.

The statistics created in this thesis are a tool, for use by coaches and the media. For experienced, intuitive coaches, these statistics should merely serve as confirmation of what they already know. Ideally, then, these tools will enable inexperienced or unintuitive coaches gain greater insight into the mechanics of match play. Importantly, the techniques explored and developed can be extended to other subjects.

The development of the half-moon statistic enables greater use of neural networks as a decision making tool and in rating systems. This is due to the greater understanding that comes with interpretability of a system.

8.3.1 Limitations

The limits of a statistical rating system are governed by the data available. In this thesis, the data available was restricted due to commercial circumstances. Furthermore, there was no information relating to the spatial, technical or team aspects of performance. Data relating to spatial and technical aspects of performance are costly and difficult to obtain. Information relating to team play is very difficult to define or collect. Because of these limitations, the rating system developed in this thesis can only be viewed as a general measure of individual performance.

Additionally, sport generates relatively small data sets for data mining. Thus the dimension reduction networks utilised in this thesis may be prone to overfitting when retraining occurs. This will be manifested in the contextuality of the models.

When a rating system is viewed as supplemental information, then the negative aspects relating to the application of the rating system are dispelled. Incorrectly and overused statistics create the possibility of coaching in hindsight, where action is taken well after the event (B. Bracewell, 2001). As a result coaching becomes reactive rather than proactive, a situation which is undesirable at the first class level. Additionally, as sport statistics are outcome focused, the processes (such as technical and mental) may be overlooked. Consequently, provided the Eagle Rating (and other measures of sporting performance in complex team sports) is viewed as a general measure of performance that is a useful tool for detecting trends in performance and communicating information to the public via the media, no problems will arise.

The KPI's are given equal weightings in the calculation of the Eagle Rating (I & II). The modified rating (II) has equal weightings for the attacking and defensive attributes which is sensible. However, when additional components are added (judgement/execution, technical, other skills and spatial) careful consideration must be given to the weightings so that the rating system does not become distorted, leading to an uncontextual overall rating.

A commercial rating system must have sufficient marketing and technical support; otherwise it will struggle to reach the intended users. Transparency and contextuality facilitate the interpretation necessary to defend and sell the rating. Thus in the early stages of a ratings evolution, the necessary support can be costly.

8.4 Contributions

The major contributions of this thesis are listed below:

- Provision of a mechanism for establishing the relative influence of inputs upon an output in a neural network, from both a univariate and multivariate perspective
- Calculation of an objective performance rating for individual rugby players on a match-by-match basis, for use in the media, using multivariate analysis and data mining techniques
- Demonstrating the use of this rating in the coaching environment
- Identification of the key problems in a sport rating system for commercial use
- Describing the fundamental philosophy for statistically based ratings systems (Transparency, Context, Robustness)

As outlined in the final research question, the processes described above are relevant for any rating system. Essentially, the rugby data has proved to be a useful and challenging example which was used to demonstrate how dimension reduction techniques can be incorporated into a rating system, based on the philosophy of Spearman's desire to quantify intelligence.

8.5 Future Research

The quantification of individual rugby player performance is an important and exciting step in uncovering more sophisticated patterns and trends that maybe evident within the Eagle Sport database. This thesis has illustrated that individual rugby player performance can be quantified using multivariate analysis and data mining techniques. The presence of this new statistic creates a vast array of research. As mentioned in Chapter Two, the Eagle Rating developed is individualistic by nature. This needs to be adapted to consider the interaction with other team-members. Having quantified individual rugby performance, it is necessary to quantify team performance. Having both sets of information will allow a supervised network to be constructed that can be used to refine and explore individual performances relative to the team, other competing individuals and so forth. The knowledge of how individuals interact creates the potential for optimising team selection by optimising individual combinations in the various sub-units of a rugby team.

The rating system needs to be expanded further by expanding the base of core components (attack, defence and positional tasks). These can easily be added to by incorporating information relating to special areas, such as scrummaging, goal kicking, and so forth. Subjective measures describing player judgement and execution of skill can also be added to the base components, to expand the depth of coverage. This can be achieved by using the perfection method to condense all relative variables to a single value. Additionally, the expansion of the models to consider other competitions, both domestic and international is desirable to gain a fuller understanding of the differences between the styles of play.

The existence of statistically sound performance measures are potentially useful in longitudinal studies of coaching programmes, looking at the impact on the relative level of player performance.

Ultimately, performance measures need to be related to spatial measures due to their importance in determining player performance. Harking back to Chapter Two where it was mentioned that in complex team sports, decisions made are sensitive to initial conditions, such as field position and the placement of other interacting individuals. For this type of analysis to be engaged, this spatial data needs to be collected.

Given that individuals are rated upon measures of performance extracted from actual match play, this has important implications in other sporting codes. Sport testing is a large area, where individuals are tested in a laboratory environment to establish changes in performances. However, when tested in the actual environment, the confounding variables that affect performance (competitiveness, technique, and mental application) are incorporated into the performance, giving a better indication of potential performance in the competitive arena.

This research is not restricted to rugby. Already, there has been significant interest in applying the general philosophy in this thesis to assessing the performance of rowers at an Olympic level. Obviously rowing is much simpler to quantify than rugby, with only three core variables required for analysis; strength, speed and endurance. Additionally, there is a defined outcome for the event, time.

The procedures developed in this thesis are also relevant outside the sports arena. Business rating systems can be developed, given suitable multivariate data, and this can be used to examine and compare the performances of products, sales representatives, marketing promotions and so forth.

This thesis has laid the foundation for the exciting prospect of using data mining techniques to explore multivariate databases in the sports environment. However, the limitations must be acknowledged. A performance-rating scheme is only as good as the data collected. Information is lost, as described in this thesis due to the information/resource trade-off. However accepting these limitations, rugby observers are left with a powerful tool for exploring individual sporting performance and ability.

However, from a statistical point of view the most exciting aspect of this thesis is the half moon statistic. This statistic shows promise as a means of identifying significant relationships where other methods fail (e.g. correlation based analysis). Mathematically and statistically, there is a need for much research associated with the half moon statistic developed in Chapter Five. The effect of different distributions, functions and sample size need to be explored in greater depth in order to establish relative cut-offs for independence and significant association via a bootstrapping procedure. Additionally, by researching the effect of different input-output functions the raw half moon statistic (prior to re-scaling), it may be possible to identify the families of functions acting upon the inputs. This would lead to greater interpretability of neural networks. There also exists the potential to incorporate the half moon statistic in a learning algorithm, by adjusting network parameters in an attempt to maximise the multivariate half moon statistic.

Appendix A

Data Description

This section contains confidential details of the Eagle Sport coding system. The information described within is the property of Eagle Sports (Sportsnet Ltd). Inclusion is intended to provide clarification of the brief details presented in Chapters Three and Six. Firstly a table summarising the codes is featured. Then notes in the table relating to specific codes are defined. A demonstration of the coding structure follows. Finally a brief description of the operation concludes this confidential section.

The following table has five columns; Code, Description, Players, Quantity and Notes. The Code column refers to the defined label for the event summarised in the Description column. 'Players' refers to the inclusion of a player for the described activity; Y for yes, N for no and O for optional. Similarly, the Quantity column indicates the requirement of additional information; again Y for yes and N for no. The final column, Notes, indicates the type of quantity required, whether a count of the number of players beaten, the distance of a kick in metres, the angle of a shot at goal in degrees, or some other specified measure. The other specified measures are outlined in further detail immediately following the table. All distances are expressed relative to the touchline, except field goal metres and goal metres. These two codes measure the distance to the goal posts. The angle of the kick, expressed in degrees with a range of 0°-180°, refers to the angle to the posts. For example, a kick from right in front is 90°, a kick on the 22 metre line and the left hand touch line is approximately 30°, whilst the same kick from the opposite side of the field has an angle of 150°. The data described within this section relates only to that applied in this study.

Code	Description	Players	Quantity	Notes
AA	Players Beaten	Y	Y	Count
AB	Tackles Broken	Y	Y	Count
AC	Catch by Chase	Y	N	
AD	Attack Ball Dropped and Regained	Y	N	
AK	Attack Kick	Y	Y	Distance
AL	Attack Layback	O	N	
AP	Pass to Hand	Y	N	
APT	Attack Pass in Tackle	Y	N	
AS	Tries Scored	Y	N	
AT	Attack Turnover	O	N	
ATD	Dropped Ball	O	N	
ATF	Forward Pass	O	N	
ATI	Attack Intercept	O	N	
ATK	Attack Knock On	O	N	
AY	Attack Metres	Y	Y	Distance
AYL	Advantage Line Crossed on Attack	Y	Y	Distance
AYM	Metres per Break	N	N	
DA	Tackle Assists	Y	N	
DB1	1st to Breakdown	Y	N	
DB2	2nd to Breakdown	Y	N	
DB3	3rd to Breakdown	Y	N	
DC	Catch from Kick	Y	N	
DCD	Catch Dropped	Y	N	
DCK	Catch Knock On	Y	N	
DCM	Mark	Y	N	
DD	Charge Down	Y	Y	1,2,3
DG	Loose Ball Gained	Y	N	
DI	Defence Intercept	Y	N	
DM	Tackles Missed	O	N	
DOT	Turnover Tackle	Y	N	
DP	Ball forced dead	Y	N	
DS	Snaking	Y	N	
DT	Tackles Made	Y	N	
DTL	Tackles Over the Advantage Line	Y	N	
GA	Goal Angle	Y	Y	Degrees
GD	Goal Pressure	Y	Y	1,2,3
GFP	Field Goal Points	Y	Y	0,3
GFY	Field Goal Metres	Y	Y	Distance
GP	Goal Points	Y	Y	0,2,3
GY	Goal Metres	Y	Y	Distance
IF	Foul Play	O	N	
IFD	Foul Play Deliberate	O	N	
IH	Head High	O	N	
IHD	Head High Deliberate	O	N	
IJ	Jump Foul	O	N	
IJD	Jump Foul Deliberate	O	N	
IK	Kick Off Foul Deliberate	O	N	
IKD	Kick Off Foul	O	N	
IM	Maul Foul	O	N	
IMD	Maul Foul Deliberate	O	N	
IO	Off Side	O	N	

IOD	Off Side Deliberate	O	N	
IP	General Foul	O	N	
IPD	General Foul Deliberate	O	N	
IR	Ruck Foul	O	N	
IRD	Ruck Foul Deliberate	O	N	
IS	Scrum Foul	O	N	
ISB	Sin Binned	Y	N	
ISD	Scrum Deliberate	O	N	
ISO	Sent Off	Y	N	
IST	Penalty Try	N	N	
IT	Throw-in Foul	O	N	
ITD	Throw-in Foul Deliberate	O	N	
JA	Jumps Won Against Throw	O	N	
JAB	JA Back	O	N	
JABS	JA Back Short	O	N	
JAF	JA Front	O	N	
JAFS	JA Front Short	O	N	
JAM	JA Middle	O	N	
JAMS	JA Middle Short	O	N	
JL1	Front Lifter	Y	N	
JL2	Back Lifter	Y	N	
JT	Jumps Lost Throw	O	N	
JTB	JT Back	O	N	
JTBS	JT Back Short	O	N	
JTF	JT Front	O	N	
JTFS	JT Front Short	O	N	
JTM	JT Middle	O	N	
JTMS	JT Middle Short	O	N	
JW	Jumps Won	O	N	
JWB	JW Back	O	N	
JWBS	JW Back Short	O	N	
JWF	JW Front	O	N	
JWFS	JW Front Short	O	N	
JWM	JW Middle	O	N	
JWMS	JW Middle Short	O	N	
JXC	Throw in – contested	Y	N	
JXN	Throw in – noncontested	Y	N	
KC	Catch Kick Off	Y	N	
KCD	Dropped Kick Off	O	N	
KCK	Knock On Kick Off	O	N	
KD	Charged Down	Y	N	
KF	Kick Out on Full	Y	N	
KO	Kick Off	Y	N	
KOW	Kick Off Retained	Y	N	
KP	Ball kicked dead	Y	N	
KS	Supporting/Lifter at Kick Off	Y	N	
KY	Kicking Metres	Y	Y	Distance
KYM	Metres per Kick	N	N	
PA	Players Beaten	Y	Y	Count
PB	Tackles Broken	Y	Y	Count
PD	Possession Ball Dropped and Regained	Y	N	
PFS	Tries Scored	Y	N	

PG	Ripped Ball	O	N	
PL	Possession Layback	O	N	
PM	Maul Won	N	N	
PML	Losing Side Maul	N	N	
PMY	Maul Metres Won	N	Y	Distance
PMYL	Maul Metres Lost	N	Y	Distance
PP	Possession Passes to Hand	Y	N	
PPT	Possession Pass in Tackle	Y	N	
PR	Ruck Won	N	N	
PRL	Losing Side Ruck	N	N	
PS	Scrum Won	N	N	
PSBL	Scrum Won Blind Left	N	N	
PSBR	Scrum Won Blind Right	N	N	
PSL	Scrum Lost on Put In	N	N	
PSN	Scrum Feed Turnover	N	N	
PSOL	Scrum Won Open Left	N	N	
PSOR	Scrum Won Open Right	N	N	
PST	Tighthead	N	N	
PSY	Scrum Metres Won	N	Y	Distance
PT	Possession Turnover	O	N	
PTD	Possession Dropped Ball	O	N	
PTF	Possession Forward Pass	O	N	
PTI	Possession Intercept	O	N	
PTK	Possession Knock On	O	N	
PY	Possession Metres	Y	Y	Distance
PYL	Advantage Line Crossed in Possession	Y	Y	Distance
PYLH	Possession Hit Up	Y	Y	Distance
PYM	Metres per Break	N	N	
RG	Gathered Ball	Y	N	
RP	Received Pass	Y	N	
SUB	Substitute	Y	N	
WA	Advantage	N	N	
WE	Advantage Over	N	N	
WR	Wrong Referee	N	N	
X	Judgement	Y	Y	1,2,3,4,5
Y	Execution	Y	Y	1,2,3,4,5

Table A.1: *Summary of the Eagle Sports Coding System*

Notes Defined

DD *Charge Down*

1 = Complete

2 = Partial

3 = Ineffective

GD *Goal Pressure*

1 = Difference between scores more than 7 points

2 = Difference between scores less than 7 points

3 = Difference between scores less than 7 points, less than 10 minutes to play

GFP/GP *Field Goal Points/Goal Points*

0 = Missed attempt

2 = Successful conversion

3 = Successful dropped goal/penalty.

X *Judgement* (Guidelines)

1 = Well Below Expectation	Destroys opportunity or option leads to turnover creating definite scoring opportunity for opposition
2 = Below Expectation	Ruins opportunity or option leads to turnover
3 = Expected (Default)	Normal flow of game continued
4 = Above Expectation	Creates Opportunity
5 = Well Above Expectation	Creates Definite Scoring Opportunity

Y *Execution*

1 = Well Below Expectation	Poor Technique - Precision / Halts team match flow
2 = Below Expectation	/ Disrupts match flow
3 = Expected (Default)	/ Continues flow
4 = Above Expectation	/ Aids match flow
5 = Well Above Expectation	Excellent Technique - Precision / Sparks team match flow

Execution Skill Examples:

Passing

1 = pass bounces, no recipient

2 = pass on hip, or below knee, above head

3 = pass in front

4 = pass in front, chest high

5 = pass in front, chest high, person given option of taking gap

Touch Finding

1 = Fails to find touch, opposition return kick with interest

2 = Touch not found

3 = Finds touch

4 = Finds touch, more than 70% of total kicking distance

5 = Finds touch, more than 85% of total kicking distance

Judgement and execution codes are designed to incorporate a subjective perspective into the coding structure, whilst maintaining an element of objectivity. The objectivity is required to enhance both within-coder and between-coder reliability.

Chapter Three described the fundamental differences between attack-based variables and possession-based variables. The overlap of codes is evident from the above table, for example AA and PA both measure players beaten, but relate to different situations.

Coding Structure

The following is the opening segment of play between the New Zealand Colts and the South African Colts in the Under 21 Final played in 2000. The home side, New Zealand is on top.

							1
NZ	10 KO		15 DG AY3 AP	8 RP PY10 PL	6 DB2 1DB3	PR	
SA		9 KC AY2 KY36		14 DT	1 DB1		

The above translates as follows. The NZ First five eighth kicked off, the ball caught by the SA halfback who ran two metres before kicking the ball 36 metres downfield. The NZ fullback gathered the ball, ran three metres before giving a pass to his fellow number eight, who ran 10 metres in heavy traffic before being tackled by the right wing for SA. The NZ number eight laid the ball back. The SA loose head prop was the first to the breakdown followed by the NZ blindside flanker and loose head prop. The breakdown then turned into a ruck won by NZ, denoting the end of the first sequence.

Recording Details

Specific details are not necessary, as only the data and its quality are of relevance in this thesis. A brief description of the processes involved is sufficient for this purpose.

All coding is performed at the Eagle Sports office. Match footage is recorded using three video recorders linked to Sky Digital decoders and archived. Approximately 10 hours in total are required to log a match. Each match is divided into three segments (0:30), (30:60), (60:end) and coded concurrently by three coders. This highlights the need for high inter-coder reliability to ensure coding continuity throughout the match. Dividing the match into thirds ensures the coders are working short spells resulting in higher quality work. The match is then sequenced, before proceeding to the data entry stage, which takes three to four hours, depending on the pace of the game. Generally a high paced game has more action, and thus more coding is required, leading to an increased workload. The final step is the publication of the data to the Eagle Sport database, which also involves an up-date of the on-line data (www.eaglesports.co.nz).

Appendix B

Results of Factor Analysis

The varimax factor loadings obtained for each of the nine positional clusters from the factor analysis of Chapter Three are given in this section. Significant loadings are in **bold** for each model. Underlined variables had a square root transformation applied for that cluster. The following descriptive codes are used. Next to each descriptive label is the calculation involving the relevant codes from Appendix A. The superscript * indicates that the code covers a number of variables as shown in Appendix A. For example JA* represents the codes – JA, JAB, JABS, JAF, JAFS, JAM and JAMS.

Label	*Code*	*Label*	*Code*
Mtr1	AY	KCKOff	KO
MtrprBrk	AYM	KCKORtnd	KOW
RnsprGm	AY/AYM	%kkortnd	KOW/KO
DfncBtn	AB	Tfouls	I*
Ps+Hnd	AP	TtlTnovr	Turnover
Trs	AS	Goal%	%Goal
Mtr2	PY	ruckimpa	DB1+DB2+DB3
MtrprBrk2	PYM	+lntckls	DTL
RnsprGm2	PY/PYM	tnvrtckl	DOT
TcklsBkn	PA+PB	tcklsast	DA
Trs2	PFS	Tm	AY+PY+KY
Lybck	PL	tcklsbkn	AB+PB
Ps+Hnd2	PP	psntckl	APT+PPT
Tckls	DT	lnbrk	PYL+AYL
msTckls	DM	htup	PYLH
%tkls	DM/(DM+DT+DOT+ DTL)	Handling	AC+DC+DCM+DG+JA*+JW*+KC+PG+RG+RP
LsblGnd	DG	Scrumswo	PS*
JWATH	JA*	infringe	IS
JW	JW*	throws w	JW*
KMtrs	KY	Total th	JXC+JXN
MtrprKCK	KYM	Throwing	JW*/(JXC+JXN)
KckprGM	KY/KYM		

Table B.1: *Labels and Codes for Factor Analysis*

Prop Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	0.088	-0.039	**0.965**	0.015	0.035	0.943
MtrprBrk	0.023	-0.158	**0.928**	-0.010	0.049	0.889
RnsprGm	0.058	0.100	**0.936**	0.002	-0.112	0.901
<u>Mtr2</u>	**0.989**	0.024	-0.014	-0.041	0.028	0.982
<u>MtrprBrk2</u>	**0.745**	0.096	0.100	-0.130	0.071	0.596
<u>RnsprGm2</u>	**0.894**	-0.084	-0.026	0.048	-0.116	0.824
<u>Lybck</u>	**0.681**	-0.137	-0.091	0.311	-0.300	0.678
Ps+Hnd2	0.279	-0.150	-0.059	-0.076	**-0.535**	0.396
<u>Tckls</u>	-0.023	**0.654**	-0.225	0.442	-0.256	0.740
msTckls	-0.106	**-0.731**	-0.060	-0.068	-0.301	0.645
%tkls	0.094	**0.729**	-0.269	0.219	-0.193	0.698
LsblGnd	-0.006	-0.191	0.060	0.012	**-0.758**	0.615
Tfouls	-0.014	0.421	0.268	0.384	-0.430	0.582
TtlTnovr	0.179	**-0.827**	-0.082	0.134	-0.166	0.768
<u>ruckimpa</u>	-0.053	0.215	-0.175	**0.505**	-0.399	0.494
+lntckls	0.164	0.176	0.310	**0.766**	0.184	0.774
tcklsast	-0.086	0.137	0.025	-0.020	**-0.669**	0.475
<u>TM</u>	**0.975**	-0.001	0.166	-0.042	0.033	0.981
tcklsbkn	**0.801**	-0.010	-0.065	0.065	0.075	0.656
psntckl	0.116	-0.109	0.186	0.229	0.200	0.643
lnbrk	**0.776**	-0.051	0.297	0.237	-0.028	0.750
htup	**0.674**	-0.115	-0.012	0.055	-0.067	0.476
<u>Handling</u>	**0.588**	-0.239	-0.020	0.177	**-0.593**	0.787
lost	0.007	0.058	0.099	**-0.895**	-0.170	0.843
scrumswo	0.160	-0.300	0.014	**0.543**	-0.160	0.436
infringe	-0.041	-0.373	-0.201	-0.133	-0.277	0.276
Variance	6.0456	4.0744	3.2400	2.6591	2.5957	18.6148
% Var	0.224	0.151	0.120	0.098	0.096	0.689

Lock Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	0.064	0.088	**-0.912**	-0.279	-0.001	0.922
MtrprBrk	0.048	0.050	**-0.910**	-0.046	0.105	0.845
RnsprGm	0.292	0.018	**-0.510**	**-0.649**	-0.171	0.795
Mtr2	0.107	**-0.823**	-0.083	-0.358	-0.242	0.883
MtrprBrk2	0.171	**-0.785**	0.074	-0.154	-0.097	0.684
RnsprGm2	0.159	**-0.755**	-0.106	-0.340	-0.211	0.766
Lybck	**0.698**	**-0.505**	-0.028	-0.318	0.112	0.857
Ps+Hnd2	**0.783**	-0.161	0.082	-0.151	-0.434	0.856
Tckls	0.368	-0.071	0.237	-0.209	**-0.508**	0.498
msTckls	0.172	0.101	-0.002	-0.026	**-0.538**	0.330
%tkls	0.412	0.044	0.489	-0.493	-0.048	0.655
LsblGnd	0.238	**-0.551**	0.336	0.139	0.006	0.493
JWATH	**0.712**	-0.093	0.099	0.407	-0.131	0.709
JW	**0.871**	-0.164	-0.009	-0.136	-0.040	0.805
TtlTnovr	0.255	-0.257	-0.137	-0.356	**-0.509**	0.535
ruckimpa	0.212	-0.086	0.232	**-0.692**	-0.103	0.595
tcklsast	-0.282	0.348	0.425	-0.326	-0.258	0.554
TM	0.446	**-0.533**	-0.132	**-0.590**	-0.237	0.905
tcklsbkn	-0.072	0.496	-0.021	-0.292	-0.248	0.399
psntckl	-0.075	-0.167	-0.160	-0.487	-0.500	0.545
lnbrk	-0.042	-0.206	-0.031	**-0.601**	0.068	0.411
htup	-0.294	-0.045	**-0.794**	0.286	0.101	0.811
Handling	**0.866**	-0.399	0.011	-0.152	-0.113	0.945
Tries	0.441	0.140	-0.142	-0.314	0.459	0.544
scrumswo	-0.017	-0.101	0.151	0.055	**-0.629**	0.432
Variance	4.3463	3.4963	3.3395	3.3332	2.2592	16.7746
% Var	0.174	0.140	0.134	0.133	0.090	0.671

Hooker Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	**0.962**	0.096	0.004	-0.070	0.001	0.939
MtrprBrk	**0.599**	0.147	-0.031	-0.299	0.113	0.484
RnsprGm	**0.955**	0.072	0.025	0.007	0.031	0.919
Mtr2	0.033	**-0.866**	0.150	0.229	-0.031	0.827
MtrprBrk2	0.326	-0.385	0.166	0.218	0.052	0.333
RnsprGm2	-0.085	**-0.875**	-0.006	0.108	-0.033	0.785
Lybck	0.040	**-0.782**	-0.054	0.245	0.101	0.686
Ps+Hnd2	0.322	-0.375	0.399	0.237	0.156	0.484
Tckls	-0.070	-0.171	-0.297	**0.663**	0.005	0.561
msTckls	0.107	-0.284	**-0.593**	-0.082	-0.079	0.456
%tkls	-0.035	0.014	-0.161	0.441	-0.092	0.231
LsblGnd	0.129	-0.323	0.314	0.381	0.154	0.389
Tfouls	-0.016	-0.428	-0.186	-0.124	-0.018	0.233
TtlTnovr	-0.020	0.171	0.078	0.066	**-0.945**	0.933
ruckimpa	-0.142	-0.044	0.300	0.551	-0.145	0.436
+lntckls	0.048	-0.130	-0.133	**0.662**	0.171	0.504
tcklsast	0.181	-0.128	0.037	0.430	-0.287	0.318
TM	**0.801**	-0.485	0.099	0.059	0.035	0.891
psntckl	0.133	0.198	0.119	0.531	0.532	0.636
lnbrk	**0.581**	**-0.552**	0.027	-0.061	0.257	0.712
htup	-0.093	**-0.743**	-0.202	-0.066	0.220	0.654
Handling	**0.622**	**-0.528**	0.223	0.365	0.115	0.862
Tries	**0.800**	0.033	-0.017	0.161	-0.219	0.715
THROWS W	0.035	-0.057	**0.865**	-0.337	-0.104	0.878
TOTAL TH	-0.044	-0.083	**0.808**	-0.392	-0.028	0.817
Throwing	0.300	0.129	**0.588**	0.092	-0.258	0.527
Variance	4.6303	4.3809	2.7972	2.7869	2.5469	17.1422
% Var	0.171	0.162	0.104	0.103	0.094	0.635

Blindside and No.8 Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	0.119	**0.883**	-0.216	-0.071	0.048	0.848
MtrprBrk	0.175	0.487	0.073	-0.101	-0.055	0.287
RnsprGm	0.090	**0.793**	-0.257	0.029	0.072	0.709
Mtr2	**0.954**	0.094	-0.191	-0.070	-0.013	0.960
MtrprBrk2	0.484	-0.023	-0.365	-0.186	0.047	0.405
RnsprGm2	**0.919**	0.112	-0.047	0.036	-0.085	0.868
Lybck	**0.814**	0.036	0.058	0.145	0.131	0.705
Ps+Hnd2	**0.606**	0.384	0.076	0.106	0.221	0.580
Tckls	-0.100	-0.049	-0.103	**0.543**	0.117	0.332
msTckls	-0.095	0.070	0.227	0.220	**-0.837**	0.815
%tkls	0.044	-0.048	-0.201	0.025	**0.851**	0.770
LsblGnd	0.016	0.312	0.163	0.371	0.029	0.263
JWATH	0.146	0.233	**0.515**	0.359	0.217	0.517
JW	0.041	0.100	0.229	0.209	**0.556**	0.417
Tfouls	0.010	-0.331	0.230	0.172	-0.168	0.220
TtlTnovr	0.048	0.202	**-0.809**	0.283	0.004	0.778
ruckimpa	0.031	-0.074	-0.098	**0.706**	-0.037	0.516
+lntckls	-0.018	0.024	-0.201	**0.649**	-0.084	0.470
tcklsast	0.220	-0.340	0.051	0.471	0.076	0.395
TM	**0.695**	**0.630**	-0.248	-0.055	0.011	0.944
tcklsbkn	0.103	0.119	-0.411	-0.016	0.219	0.242
psntckl	0.107	**0.516**	-0.066	-0.007	0.006	0.282
lnbrk	0.374	0.273	-0.453	0.265	0.030	0.490
htup	0.071	-0.098	0.021	0.240	-0.162	0.099
Handling	**0.642**	0.497	0.166	0.240	0.363	0.877
Tries	0.192	0.181	0.121	-0.075	0.240	0.148
Variance	4.2681	3.3409	2.6422	2.2704	2.1901	14.7116
% Var	0.158	0.124	0.098	0.084	0.081	0.545

Openside Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	0.018	**0.882**	0.270	-0.126	-0.096	0.876
MtrprBrk	0.051	0.302	**0.557**	-0.252	-0.004	0.467
RnsprGm	-0.064	**0.909**	0.123	-0.002	-0.089	0.853
Ps+Hnd	-0.069	**0.696**	0.056	0.227	-0.039	0.545
Mtr2	**0.954**	0.045	0.174	-0.034	-0.022	0.945
MtrprBrk2	**0.794**	-0.141	0.033	0.186	0.081	0.692
RnsprGm2	**0.882**	0.153	0.215	-0.082	-0.085	0.862
Lybck	**0.838**	0.032	0.099	0.024	-0.203	0.755
Ps+Hnd2	0.300	-0.008	**0.861**	0.112	-0.115	0.857
Tckls	-0.147	-0.000	-0.024	**0.772**	-0.148	0.641
msTckls	-0.145	-0.030	-0.008	**0.650**	0.424	0.623
%tkls	0.064	0.151	0.014	-0.049	**-0.709**	0.532
LsblGnd	0.228	0.433	-0.208	0.023	-0.468	0.502
JWATH	0.105	0.066	0.125	0.400	-0.377	0.333
JW	0.163	-0.220	**0.837**	-0.135	0.124	0.809
Tfouls	0.129	-0.059	-0.309	0.498	-0.078	0.370
TtlTnovr	0.453	0.323	0.260	0.389	0.073	0.534
ruckimpa	0.088	-0.085	0.001	**0.620**	-0.010	0.399
+lntckls	-0.028	0.016	0.373	0.377	-0.281	0.361
tnvrtckl	-0.111	0.229	0.381	0.356	0.053	0.339
tcklsast	0.018	0.023	0.144	0.103	**-0.680**	0.495
TM	**0.651**	**0.628**	0.215	-0.121	-0.039	0.881
tcklsbkn	**0.552**	0.206	-0.025	-0.124	-0.084	0.371
psntckl	0.171	0.061	**0.823**	0.035	-0.166	0.739
lnbrk	0.319	**0.833**	-0.090	0.081	-0.181	0.843
htup	0.257	0.476	-0.184	-0.010	-0.213	0.372
Handling	**0.574**	0.364	**0.584**	0.099	-0.241	0.870
Tries	0.117	0.512	0.012	-0.218	0.049	0.326
Variance	4.7709	4.4154	3.5653	2.5440	1.8981	17.1937
% Var	0.170	0.158	0.127	0.091	0.068	0.614

Halfback Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	**0.791**	0.207	0.395	-0.180	0.122	0.872
MtrprBrk	**0.787**	0.302	0.017	0.146	-0.067	0.737
RnsprGm	0.325	-0.017	**0.567**	-0.400	0.304	0.681
DfncBtn	**0.877**	-0.040	0.008	-0.053	-0.081	0.779
Ps+Hnd	0.119	-0.101	0.041	**-0.910**	0.104	0.865
Mtr2	0.072	**0.941**	0.054	0.016	-0.006	0.895
MtrprBrk2	0.040	**0.740**	0.059	0.199	-0.142	0.613
RnsprGm2	0.016	**0.929**	0.020	-0.084	0.083	0.877
Lybck	0.082	**0.748**	0.031	0.164	0.046	0.596
Ps+Hnd2	0.357	**0.608**	-0.166	-0.256	0.117	0.604
Tckls	-0.164	0.133	0.098	0.130	**-0.528**	0.349
msTckls	-0.406	0.090	0.142	0.174	0.319	0.325
%tkls	0.350	0.033	-0.097	0.071	**-0.693**	0.618
LsblGnd	0.070	0.122	0.427	-0.047	-0.205	0.246
KMtrs	0.044	0.013	**0.779**	-0.134	-0.258	0.694
MtrprKCK	0.223	0.041	0.411	-0.216	**-0.541**	0.560
KckprGM	-0.083	-0.021	**0.824**	-0.023	0.111	0.699
Tfouls	0.052	-0.020	-0.072	-0.201	**-0.559**	0.361
TtlTnovr	0.187	0.284	-0.141	-0.104	-0.063	0.150
ruckimpa	0.083	-0.094	0.416	0.246	0.022	0.250
+lntckls	-0.032	0.037	0.114	-0.441	-0.313	0.308
tnvrtckl	-0.075	-0.270	-0.254	-0.390	-0.197	0.334
tcklsast	-0.067	-0.067	0.195	0.107	-0.456	0.266
TM	**0.765**	0.329	0.383	-0.172	0.115	0.884
tcklsbkn	**0.827**	0.182	0.032	-0.098	-0.081	0.734
Plyrsbtn	**0.656**	-0.204	-0.057	0.125	0.051	0.494
psntckl	0.405	0.030	0.403	-0.199	-0.292	0.452
lnbrk	**0.517**	0.387	0.430	-0.267	-0.078	0.679
htup	-0.190	**0.535**	0.367	-0.061	-0.352	0.585
Handling	0.174	0.091	0.182	**-0.891**	0.107	0.876
Tries	0.221	0.171	0.102	-0.083	0.177	0.126
Variance	4.9528	4.2423	3.1823	2.7444	2.3885	17.5103
% Var	0.160	0.137	0.103	0.089	0.077	0.565

First Eighth Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	0.313	0.043	**0.629**	0.356	0.166	0.649
MtrprBrk	0.181	-0.351	-0.174	**0.644**	-0.284	0.682
RnsprGm	0.150	0.315	**0.766**	-0.129	0.356	0.851
Ps+Hnd	-0.122	0.130	**0.746**	-0.160	-0.047	0.616
Mtr2	**-0.964**	0.003	-0.064	-0.126	-0.032	0.950
MtrprBrk2	**-0.893**	-0.045	-0.155	-0.111	0.014	0.836
RnsprGm2	**-0.951**	0.042	-0.179	-0.095	0.018	0.948
Lybck	**-0.800**	-0.231	0.110	0.229	-0.074	0.764
Ps+Hnd2	**-0.819**	0.262	-0.214	-0.018	0.084	0.793
Tckls	-0.202	**-0.514**	-0.001	**-0.647**	0.176	0.754
msTckls	-0.075	0.336	**-0.552**	0.161	0.108	0.462
%tkls	-0.031	**-0.469**	0.344	**-0.708**	0.080	0.847
LsblGnd	0.237	0.261	**0.770**	-0.061	-0.069	0.726
KMtrs	0.122	0.919	0.065	0.057	0.065	0.871
MtrprKCK	-0.047	0.160	-0.382	-0.134	**-0.645**	0.608
KckprGM	0.236	**0.805**	0.226	0.081	0.313	0.860
%kkortnd	0.029	0.269	0.035	**0.723**	0.038	0.599
Tfouls	0.212	0.417	0.304	-0.442	-0.240	0.564
TtlTnovr	-0.107	0.063	-0.430	-0.134	0.279	0.296
Goal%	-0.128	**0.811**	0.062	-0.109	-0.186	0.725
ruckimpa	0.058	0.219	-0.071	**-0.649**	-0.046	0.480
tcklsast	0.286	0.171	-0.255	-0.137	**0.692**	0.674
lnbrk	0.215	-0.467	-0.060	-0.064	-0.037	0.274
Handling	0.042	**0.613**	**0.745**	-0.057	0.102	0.946
Tries	0.216	-0.024	0.065	0.053	**-0.912**	0.885
Variance	4.5411	4.1329	3.8194	2.8482	2.3190	17.6606
% Var	0.182	0.165	0.153	0.114	0.093	0.706

Midfield Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	**0.927**	0.115	-0.053	0.054	0.158	0.903
MtrprBrk	0.441	0.267	0.234	-0.054	0.396	0.481
RnsprGm	**0.827**	-0.069	-0.229	0.083	-0.085	0.755
DfncBtn	**0.734**	0.059	-0.062	0.026	0.027	0.548
Ps+Hnd	**0.768**	-0.094	-0.127	0.040	-0.075	0.622
Mtr2	0.007	**-0.903**	-0.045	0.036	0.048	0.820
MtrprBrk2	-0.017	**-0.822**	-0.047	-0.001	0.033	0.679
RnsprGm2	0.015	**-0.930**	-0.013	-0.008	0.010	0.865
Lybck	0.098	**-0.845**	-0.029	-0.136	-0.121	0.757
Ps+Hnd2	0.048	-0.367	0.072	0.178	0.063	0.177
Tckls	-0.006	-0.258	-0.097	0.123	**0.668**	0.537
msTckls	-0.051	-0.228	0.066	0.416	-0.279	0.309
%tkls	0.126	0.013	-0.105	-0.290	**0.709**	0.613
LsblGnd	0.338	-0.064	-0.379	0.275	0.216	0.384
KMtrs	0.061	0.025	**-0.911**	-0.038	0.043	0.838
MtrprKCK	0.125	-0.031	**-0.825**	-0.065	-0.075	0.706
KckprGM	0.161	0.105	**-0.861**	-0.033	0.143	0.800
Tfouls	-0.376	0.075	-0.239	0.184	0.425	0.419
TtlTnovr	0.146	0.021	0.018	**0.896**	0.044	0.827
ruckimpa	0.096	-0.364	0.081	0.198	0.156	0.212
+lntckls	-0.025	-0.471	0.087	0.126	0.416	0.419
tnvrtckl	0.039	-0.055	0.027	0.098	**0.620**	0.399
tcklsast	-0.186	-0.243	-0.019	0.400	0.208	0.298
TM	**0.929**	0.033	-0.060	0.060	0.165	0.899
tcklsbkn	**0.666**	-0.041	0.029	0.086	0.038	0.455
Plyrsbtn	0.421	0.025	-0.236	-0.096	0.008	0.243
psntckl	**0.509**	-0.231	0.165	0.097	-0.155	0.373
lnbrk	**0.770**	-0.199	0.093	-0.014	0.070	0.647
Handling	**0.742**	-0.159	-0.431	0.068	0.065	0.771
Tries	0.362	0.143	-0.168	-0.138	-0.054	0.201
Variance	6.3053	4.0050	2.9432	2.3397	2.1924	17.7857
% Var	0.203	0.129	0.095	0.075	0.071	0.574

Outside Back Model

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Mtr1	**0.830**	-0.388	0.020	0.053	-0.196	0.881
MtrprBrk	**0.762**	0.177	-0.016	0.020	0.146	0.634
RnsprGm	0.423	**-0.731**	0.035	0.074	-0.371	0.857
DfncBtn	**0.870**	0.113	0.070	0.057	-0.006	0.778
Ps+Hnd	0.342	-0.271	0.090	0.053	**-0.771**	0.795
Mtr2	0.025	0.024	**0.954**	0.000	0.017	0.912
MtrprBrk2	0.083	0.050	**0.870**	0.033	0.034	0.769
RnsprGm2	0.021	0.049	**0.947**	-0.064	-0.009	0.904
Lybck	-0.048	0.043	**0.795**	-0.125	0.015	0.652
Tckls	-0.088	-0.104	0.142	**-0.854**	0.078	0.773
msTckls	-0.064	-0.139	0.098	0.027	0.416	0.207
%tkls	-0.087	-0.006	0.054	**-0.755**	-0.134	0.599
LsblGnd	-0.130	**-0.549**	-0.164	0.131	-0.257	0.428
KMtrs	-0.102	**-0.905**	-0.063	0.067	0.106	0.849
MtrprKCK	-0.109	**-0.543**	0.029	-0.043	-0.014	0.310
KckprGM	-0.072	**-0.886**	-0.099	0.057	0.170	0.832
Tfouls	-0.023	-0.043	0.153	0.103	0.293	0.122
TtlTnovr	0.039	**-0.517**	0.038	-0.015	0.072	0.275
Goal%	-0.104	**-0.512**	-0.111	0.037	0.324	0.392
ruckimpa	-0.126	-0.214	-0.001	-0.016	-0.258	0.128
+lntckls	0.093	-0.007	-0.023	**-0.603**	-0.111	0.385
tnvrtckl	0.213	0.220	-0.054	**-0.521**	0.160	0.393
tcklsast	-0.151	0.102	0.034	-0.457	-0.177	0.274
TM	**0.860**	-0.320	0.066	0.042	-0.192	0.885
tcklsbkn	**0.752**	0.068	0.176	0.129	0.072	0.623
Plyrsbtn	**0.666**	0.149	-0.081	-0.121	-0.131	0.505
psntckl	0.189	0.024	0.249	-0.161	**-0.577**	0.457
lnbrk	**0.810**	0.012	0.022	-0.061	-0.286	0.743
Handling	0.232	**-0.790**	0.059	0.064	-0.436	0.876
Tries	**0.662**	0.118	-0.090	-0.029	-0.094	0.469
Variance	5.4343	4.4438	3.4431	2.2856	2.1006	17.7073
% Var	0.181	0.148	0.115	0.076	0.070	0.590

Appendix C

Job Requirements

This supplementary section provides a broad description of the job requirements of positional clusters, or units, as defined by the NZRFU (1998). The following information is adapted from *Principles of Rugby Coaching*, Stage Two, Module 19 – The Selection Process, pages 192 and 193 with the permission of the NZRFU. The purpose is to inform the reader of the attributes that are deemed most important within each unit. Careful reading will indicate the skill set overlap between positions. For example, the ability to stand in a tackle is considered an important attribute for both outside backs and the midfield.

Positional Responsibilities

Outside Backs

- Safe under high ball
- Sound tackler
- Acceleration and top end speed
- Understand role in defence
- Communication / organisation within unit
- High work rate in cover defence/ counter attack / backline entry
- Vision / running lines / timing for backline penetration (creating extra man)
- Strength in tackle –ability to stay on feet
- Provide attacking threat on blindside

Mid Field

- Accurate passing skills
- Vision – able to put outsides into space
- Strong, accurate, intimidating tackler
- Strength for go-forward and able to stand in tackle
- Straightness in attack
- Able to confront / commit opponent and get ball way under pressure
- Able to kick accurately off either foot
- Excellent communication / organisation skills
- Understand the game
- Speed to make outside break
- High work rate in support play

Inside Backs

- Understand game tactically
- General awareness and alertness
- Excellent vision
- Good communication / organisation skills
- Able to "call the shots"
- Decisive / Authoritative
- Accurate passer and kicker off either hand and foot
- Good tackler (doesn't shirk confrontation)
- Acceleration to make initial break
- Able to threaten defence (commit opponents)
- High work rate in support and cover defence
- Organise defence
- Able to withstand both physical and mental pressure

Loose Forwards

- High work rate in support and cover
- Speed to breakdown and speed endurance
- Good running and linking skills
- Strong tackles and an understanding of role in defensive pattern
- Able to drive ball carrier back around fringes and turn ball over in tackle
- Know running lines in attack and defence
- Possible lineout option
- Understand laws, particularly tackle ball, ruck and maul
- Win ball on ground
- Vision and awareness in tackle situation
- Exceptional fitness levels
- Relentless – not intimidated
- Good all round skills

Locks

- Win own lineout ball
- Contest opposition ball at lineout time when required
- Strong and durable, with physical presence
- Agile in air, good hands and timing
- Strength in scrum, ruck and maul
- Low body position in contact and pushing or grouped phases
- Able to concentrate on set phase first then get quickly to the breakdown and in support
- Strong in tackle
- Able to “pick and go” through middle of ruck
- Relentless – never give in

Props

- Strong scrummager, technically correct
- Strong at supporting / lifting in lineouts and kick-offs
- Mobile / quick to breakdown
- Low body position in contact, scrum, ruck and maul
- Very strong in upper body and legs
- Durable / mentally tough
- Agile / explosive for "pick and go" through rucks
- Good tackier and high work rate
- Not intimidated

Hooker

- Accurate thrower to lineout
- Strong scrummager
- General awareness as lineout and second phases – know when to stay out / go in
- Strong tackler
- Exceptional fitness levels
- Able to apply pressure physically and psychologically
- Strong in upper body and legs
- Able to catch, pass and run effectively with the ball
- Accurate in ball retention

Appendix D

Commercial Applications

The linear based Eagle Rating described in Chapter Three was implemented immediately in the public arena (February 2000). The first two commercial applications of the Eagle Rating are presented briefly in this section.

Firstly, the Eagle Rating was introduced as the basis for Ultimate Rugby, a fantasy sports game where participants selected a team, limited by a salary cap, and earned points depending on the performance of selected individuals, judged according to match statistics (www.UltimateRugby.com). Points awarded and accumulated were based simply on the Eagle Rating. The game description as it appeared in the February issue (2000) of *Sky Watch* is shown on pages 323-326.

Following Ultimate Rugby, Sky Television used the Eagle Rating as the foundation for their Sky "Man of the Match" competition, where participants selected in their opinion, the best player. The subjective perspective was compared with the Eagle Rating and a winner selected from the correct entries (www.eaglesports.co.nz/skytv/SKYFilter.asp). An example of this competition is appended on the next page, before the Ultimate Rugby insert.

'Man of the Match' Competition Results

Eagle Sports Player Ratings

All Blacks vs France 11th November 2000

Congratulations to Chrissie Middleton of Gisborne
Winner of this weeks Rodd & Gunn Oil Skin Jacket.

Player	Rating	Factor1	Factor2	Factor3	Factor4	Factor5
Anton Oliver (Man of the Match)	72	1.5376	-0.8722	1.3555	4.4401	-2.2502
Christian Cullen	65	0.6486	0.3898	0.1716	0.9488	0.5964
Tana Umaga	65	0.2976	-0.1687	-2.3826	0.3932	4.8653
Jonah Lomu	58	0.5463	-0.8193	3.6105	-1.5408	-0.2397
Justin Marshall	56	0.5961	-0.9456	0.3118	0.1556	0.9240
Carlos Spencer	54	-2.0011	0.3770	1.0380	0.9632	0.4841
Greg Somerville	51	-0.0163	0.1209	-0.1326	-0.0730	0.2175
Reuben Thorne	50	-1.1106	0.1966	0.1124	0.5459	0.2486
Norman Maxwell	49	-1.3454	0.6492	1.5816	-2.0106	1.1048
Scott Robertson	48	-0.7854	0.0549	0.2481	-0.5210	0.5956
Andrew Mehrtens	46	1.1594	-2.4307	2.4910	-0.5305	-1.2985
Ron Cribb	43	-1.6272	0.8877	-0.8862	-0.2132	0.7073
Todd Blackadder	43	0.8250	-1.0411	-0.9663	0.7267	-0.8143
Douglas Howlett	40	-0.6120	-0.1242	0.8095	-1.4389	-0.3786
Troy Flavell	39	0.7914	-0.7089	-0.9259	0.6462	-1.6721
Daryl Gibson	36	-0.7813	0.1194	-0.7238	-0.5198	-0.5867
Gordon Slater	27	-1.8408	-1.9461	4.6821	-2.5038	-2.0480
Greg Feek	15	-1.9859	-1.9772	1.5938	-1.8801	-1.8413

THE EAGLE RATING

Eagle Sports has maintained a database on 5000 first class rugby players around the world for the last 4 years. Eagle spends an average of 12 hours analysing games for 93 main statistical attributes and has developed a unique set of attributes for each position in the rugby team.

Eagle Sports provides accurate statistics to national and premier division teams in the UK, South Africa, Australian, New Zealand and Argentina. Eagle statistics provide a competitive advantage to coaches around the world.

Now the ultimate indicator of a players season and game by game performance has arrived, it is called the Eagle rating. Arm chair sports critics can now get a score which is an accurate reflection of a player's performance, game by game and across the season.

These are just some of the factors which make up Eagle Ratings on rugby positions:

• Players Beaten • Tackles Broken • Chase By Chase • Attack Ball Dropped and Regained • Attack Kick • Attack Pass in Tackle • Tries Scored • Attack Turnover • Dropped Ball • Forward Pass • Attack Intercept • Attack Knock On • Attack Metres • Metres per Break • Adv. Line Crossed on Attack • Tackle Assists • 1st to Breakdown • 2nd to Breakdown • 3rd to Breakdown • Catch from Kick • Catch Dropped • Catch Knock On • Charge Down • Loose Ball Gained • Defence Intercept • Tackles Missed • Turnover Tackle • Tackles Made • Tackle over Adv. Line • Goal Angle • Goal Pressure • Field Goal Points • Field Goal Yards • Goal Left Field • Goal Points • Goal Right Field • Goal Yards • Foul Play • Foul Deliberate • Head High • Head High Deliberate • Jump Foul • Kick Off Foul • Kick Off Deliberate • Maul Foul • Offside • Offside Deliberate • General Foul • General Deliberate • Ruck Foul • Ruck Deliberate • Scrum Foul • Sin Binned • Scrum Deliberate • Sent Off • Penalty Try • Throw-in Fowl • Throw Deliberate

Eagle Ratings - Super 12, 1999

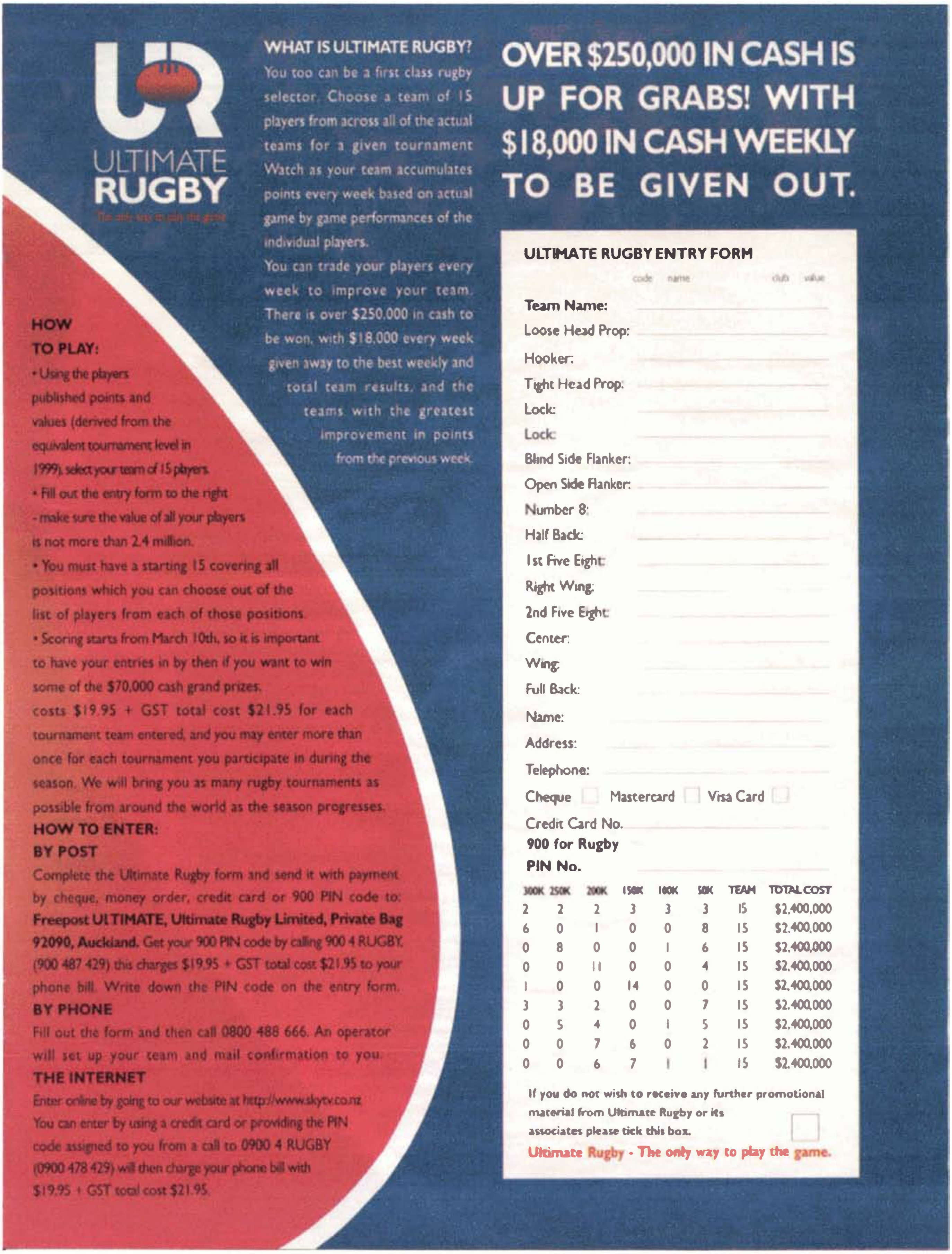
CODE	CHRISTIAN	SURNAME	RATING	VALUE
LOOSE HEAD PROPS				
5222	Tana	Turing	59	250,000
5110	Bill	Young	53	200,000
4232	Cameron	Blades	47	150,000
3177	Carl	Hoeft	53	200,000
3684	Dan	Crowley	54	200,000
4990	David	Hewitt	84	300,000
3421	Ortan	Grunwald	45	100,000
3069	Greg	Feek	50	200,000
4459	Greg	Somerville	50	200,000
5302	Joe	McDonnell	47	150,000
5416	Marc	Dos	36	50,000
3231	Michael	Collins	53	200,000
3224	Mike	Edwards	53	200,000
3224	Mike	Edwards	53	200,000
5328	Ollie	Le Roux	60	250,000
3331	Paul	Thomas	56	250,000
3715	Richard	Harry	53	300,000
3975	Rob	Kempson	48	150,000
5094	Rod	Moore	48	150,000
3569	Toks	Van der Linde	48	150,000
4665	David	Briggs	0	50,000
4768	Kevin	Yates	0	50,000
3531	Marius	Mostert	0	50,000
4265	Simon	Kerr	0	50,000
HOOKERS				
3125	Shane	Carter	2	50,000
3486	Leon	Boshoff	0	50,000
3957	Keven	Mealamu	0	50,000
3197	Anton	Oliver	52	200,000
3423	Charles	Marsh	43	100,000
3424	Chris	Rossouw	54	250,000
3102	Dave	Hayes	49	150,000
3232	Greg	Smith	52	200,000

Continued over...

Eagle Ratings - Super 12, 1999

CODE	CHRISTIAN	SURNAME	RATING	VALUE
4616	Henry	Kemp	72	300,000
5060	Jeremy	Paul	52	200,000
3946	Johan	Van Wyk	45	100,000
3721	Mark	Cock	46	150,000
3357	Mark	Hammett	41	100,000
3151	Mike	Sexton	41	100,000
3685	Michael	Foley	49	150,000
3571	Murray	Visser	54	200,000
3413	Naka	Drotské	58	250,000
3107	Norman	Hewitt	51	200,000
3290	Paul	Mitchell	47	150,000
4014	Tommy	Webb	43	100,000
3680	Timote	Tavalea	38	50,000
TIGHT HEAD PROPS				
5503	Fletcher	Dyson	0	50,000
5508	Ben	Darwin	0	50,000
3719	Andrew	Heath	53	200,000
3169	Con	Barrell	47	150,000
3314	Craig	Dowd	50	200,000
4140	Glen	Pancho	53	200,000
4150	Gordon	Slater	43	100,000
5586	Jaco	Espag	55	250,000
4404	John	Smit	50	150,000
3343	Kees	Meeuws	50	200,000
5072	Kobus	Visagie	48	150,000
3665	Patrico	Noriega	57	250,000
3199	Paul	Martin	48	150,000
4202	Piet	Boer	44	100,000
3454	Willie	Meyer	55	250,000
3955	Carl	Hayman	0	50,000
5074	Eugene	Meyer	0	50,000
5811	Johan	Human	0	50,000
5336	Simon	Rossi	0	50,000
LOCKS				
5326	Tom	Willis	0	50,000
3644	Simon	Maling	0	50,000
5629	Marius	Bosman	0	50,000
3600	Johan	Ackermann	0	50,000
3078	Dallas	Seymour	0	50,000
5605	Albert	Van den Berg	42	100,000
3345	Brendon	Timmins	47	150,000
5404	Charles	Rechelmann	37	50,000
3361	George	Leaupepe	81	250,000
3130	Chresten	Davis	35	50,000
3662	David	Giffin	58	250,000
3292	Dean	Waller	47	150,000
3040	Glenn	Taylor	44	100,000
3577	Hottie	Louw	40	100,000
3207	Inoke	Afeaki	43	100,000
3682	John	Eales	89	300,000
3524	John	Slade	48	150,000
3711	John	Welborn	48	150,000
3178	John	Botha	44	100,000
4161	Justin	Harrison	53	200,000
3599	Krynauw	Otto	47	150,000
3522	Mark	Andrews	44	100,000
3679	Mark	Connors	63	300,000
5121	Nathan	Sharpe	48	150,000
3041	Norman	Maxwell	48	150,000
3658	Owen	Finegan	80	300,000
3627	Philip	Smit	60	250,000
4367	Paul	Tito	40	100,000
3314	Robin	Brooke	44	100,000
4212	Royce	Willis	54	200,000
4807	Selborn	Boome	54	200,000
3154	Steve	Lancaster	33	50,000
3147	Todd	Blackadder	63	300,000
3724	Tom	Bowman	54	250,000
4452	Troy	Flavell	55	250,000
3713	Warwick	Waugh	30	50,000
4496	Victor	Matfield	30	50,000
3739	Wium	Basson	49	150,000
3428	Bruce	Thorne	0	50,000
5413	Campbell	Feather	0	50,000
5266	Chris	Jack	0	50,000
BLIND SIDE FLANKER				
4159	Stuart	Pinkerton	44	100,000
4389	Ralf	Schroeder	0	50,000
3317	Andrew	Blowers	59	250,000
3407	Charl	Van Rensburg	61	250,000
3574	Corne	Krige	69	300,000
3329	Dylan	Mika	49	150,000
5062	Jim	Williams	41	100,000
3406	Johan	Erasmus	58	250,000
3192	Kalvin	Middleton	18	50,000
3597	Ruben	Kruger	27	50,000
4139	Matt	Cockbain	60	250,000
3708	Michael	Brial	44	100,000
4124	Pieter	Krause	31	50,000
4297	Reuben	Thorne	56	250,000
5064	Troy	Jaques	46	150,000
OPEN SIDE FLANKER				
3661	Ipolito	Fenukitau	0	50,000
4149	Fosi	Maka	0	50,000
3710	Daniel	Manu	0	50,000
3400	Andre	Venter	61	250,000
3859	Brett	Robinson	45	100,000
3680	David	Wilson	46	150,000
4250	Duncan	Blaikie	75	300,000
3626	Gideon	Wiese	38	50,000
4042	Glen	Marsh	42	100,000
3612	Jamie	Brookes	38	50,000
3198	Joshua	Kronfeld	61	250,000
3047	Justin	Collins	35	50,000
3726	Keith	Gleeson	31	50,000
3181	Kupu	Vanisi	55	250,000
5487	Mark	Murray	50	200,000
3565	Rabbie	Brink	55	250,000
3157	Scott	Robertson	47	150,000
3519	Wayne	Fyvie	57	250,000
3609	Piet	Arnold	0	50,000
NUMBER 8's				
3193	Taine	Randell	57	250,000
5421	Ron	Cribb	0	50,000
3965	Harry	Makin	0	50,000
3459	Andre	Vos	53	200,000
5590	Anton	Leonard	54	200,000
3573	Bobby	Skinstad	62	300,000
3210	Brad	McLeod-Henderson	75	300,000
3230	Dean	Muir	45	100,000
3208	Filo	Tiatia	50	200,000
5493	Gordon	Falcon	34	50,000
3596	Schutte	Bekker	58	250,000
5083	Toutai	Sirimaa	45	100,000
3681	Toutai	Kefu	50	200,000
4201	Warren	Brosnihan	46	150,000
4108	Xavier	Rush	35	50,000
4258	Koulo	Takino	0	50,000
4621	Piet	Joubert	0	50,000
3538	Russell	Nelson	0	50,000
HALF BACK				
5108	Travis	Hall	0	50,000
3704	Sam	Payne	0	50,000
4499	Jimmy	Powell	0	50,000
3488	Jacob	Raulini	61	250,000
5826	George	Smith	0	50,000
5796	Craig	Davidson	0	50,000
3993	Conrad	Breytenbach	0	50,000
3172	Aaron	Flynn	23	50,000
4469	Byron	Kelleher	70	300,000
3460	Chad	Alcock	51	200,000
4234	Chris	Whitaker	67	300,000
3077	Danny	Lee	30	50,000
4622	Dan	Van Zyl	57	200,000
5606	David	Von Hoesslin	51	200,000
5584	Deon	De Kock	52	200,000
3656	George	Gregan	62	250,000
3561	Joggie	Viljoen	37	50,000
3280	Mark	Robinson	44	100,000
3728	Sam	Cordingley	41	100,000
5111	Steve	Devine	48	150,000
3405	Werner	Swanepoel	64	300,000
4530	Jason	Spice	53	200,000
3246	Rhys	Duggan	48	150,000
5401	Brett	McCormack	0	50,000
5694	Dean	Hall	0	50,000
1ST FIVE EIGHT'S				
5473	Tanner	Vili	84	250,000
4137	Shane	Drahm	0	50,000
4262	Orene	Ai'i	0	100,000
5795	Nathan	Williams	0	50,000
5333	Aaron	Mauger	0	50,000
3170	Andrew	Mehrtens	42	100,000
3665	Braam	Van Straaten	35	50,000
3330	Carlos	Spencer	55	250,000
3034	David	Holwell	50	200,000
5194	Gaffie	Du Toit	73	300,000
4563	Glenn	Jackson	57	250,000
5085	Christian	Warner	42	100,000
5063	Craig	Mcmullen	35	50,000
4147	Jannie	de Beer	58	250,000
5545	Leon	MacDonald	61	250,000
3581	Louis	Koen	47	150,000
5077	Manuel	Edmonds	47	150,000
3546	Peter	O'Neil	32	50,000
3729	Nathan	Spooner	46	150,000
3648	Rod	Kafer	79	300,000
3652	Steve	Larkham	61	250,000
3186	Tony	Brown	52	200,000
3677	Elton	Flatley	0	50,000
RIGHT WING				
5023	Tai	Tabai	0	50,000
3244	Bruce	Reihana	58	250,000
4203	Casper	Steyn	33	50,000
5068	Chester	Williams	54	250,000
3244	Bruce	Reihana	58	250,000
4203	Casper	Steyn	33	50,000
5068	Chester	Williams	54	250,000
3687	Damian	Smith	38	50,000
3962	Douglas	Howlett	65	300,000
3962	Douglas	Howlett	65	300,000
3555	Jason	Smart	44	100,000
3962	Douglas	Howlett	65	300,000
3462	Henry	Pedro	40	100,000
5613	Rene	Tuivaiti	62	300,000
3324	James	Kerr	39	50,000
3475	Jason	Little	65	300,000
3650	Joe	Roff	47	150,000
2987	Jonah	Lomu	70	300,000
5601	Marne	Du Toit	53	200,000
3552	Pieter	Rossouw	73	300,000
5082	Scott	Staniforth	44	150,000
3545	Stefan	Terblanche	38	50,000
3417	Stephen	Brink	59	250,000
3741	Kaya	Malotana	0	50,000
2ND FIVE EIGHT				
3449	Tiaan	Joubert	0	50,000
3403	Brendon	Venter	63	250,000
2991	Alama	Ieremia	68	300,000
3588	Andre	Snyman	46	150,000
5058	Craig	Innes	81	300,000
3587	Dave	Van Schalkwyk	48	150,00
3161	Daryl	Gibson	67	300,000
3653	James	Holbeck	41	100,000
3446	Japie	Mulder	53	200,000
3213	Jason	O'Halloran	67	300,000
5538	Joe	Gillingham	50	200,000
3635	Lourens	Venter	70	300,000
3162	Mark	Mayerhofler	69	300,000
3731	Nathan	Grey	66	300,000
4167	Pieter	Muller	54	250,000
3269	Pita	Alatini	74	300,000
4167	Pieter	Muller	54	250,000
3269	Pita	Alatini	74	300,000
2993	Walter	Little	64	300,000
5615	Wayne	Julies	43	100,000
CENTERS				
5795	Trevor	Halstead	0	50,000
5350	Steve	Kefu	0	50,000
4902	Rico	Gear	0	50,000
3670	Adam	Magro	53	200,000
3742	Chris	Kruger	42	100,000
2985	Christian	Cullen	70	300,000
3674	Daniel	Herbert	65	300,000
5504	Darren	Junee	32	50,000
5067	Deon	Kayser	57	250,000
3323	Eroni	Clarke	73	300,000
3674	Daniel	Herbert	65	300,000
5504	Darren	Junee	32	50,000
5387	Keith	Lowen	52	200,000
5067	Deon	Kayser	57	250,000
3323	Eroni	Clarke	73	300,000
3608	Grant	Esterhuizen	32	50,000
5603	Gus	Theron	19	50,000
3114	Mark	Ranby	39	50,000
3031	Norman	Berryman	50	200,000
3336	Ross	Kopas	58	250,000
4380	Rua	Tipoki	68	300,000
2992	Scott	McLeod	48	150,000
5059	Sterling	Mortlock	79	50,000
4351	Robbie	Fleck	57	250,000
WING				
4126	Wickus	Venter	0	50,000
3215	Tana	Umaga	54	250,000
4235	Aisea	Sosaia	38	50,000
3676	Ben	Tune	42	100,000
3160	Bradley	Fleming	32	50,000
3189	Brendon	Laney	54	300,000
3578	Breyton	Paulse	43	100,000
3436	Jannie	Van der Walt	40	100,000
3264	Joel	Vidiri	42	100,000
5802	John	Daniels	27	50,000
5076	Mathew	Dowling	32	50,000
3649	Mitch	Hardy	33	50,000
5076	Mathew	Dowling	32	50,000
3649	Mitch	Hardy	33	50,000
5540	Wim	Meyer	23	50,000
5436	Dan	Parkinson	0	50,000
4076	Marcus	Parkinson	0	50,000
5123	Rich	Nalatu	0	50,000
FULL BACK				
3733	Chris	Latham	75	300,000
3053	Daryl	Lilley	48	150,000
5505	Duncan	McCrae	75	300,000
3053	Daryl	Lilley	48	150,000
3383	Jorrie	Kruger	38	50,000
2985	Adrian	Cashmore	14	50,000
5505	Duncan	McCrae	75	300,000
2986	Jeff	Wilson	73	300,000
3618	Kobus	Engelbrecht	29	50,000
3541	Marius	Goosen	75	300,000
3694	Matthew	Burke	83	300,000
3553	Percy	Montgomery	70	300,000
5073	Thinus	Delport	62	250,000
4514	Todd	Miller	61	250,000
4144	Graeme	Bond	0	50,000
4523	Luke	Crichton	0	50,000
4456	Paul	Miller	0	50,000
4436	Russell	Bennett	0	50,000
POSITIONS TO BE CONFIRMED				
5830	Jason	Chandler	0	50,000
3732	Marc	Stcherbina	0	50,000
3050	Nick	White	0	50,000
5581	Sean	Hardman	0	50,000
5592	Sias	Wagner	0	50,000
3642	Vorster	Venter	0	50,000

Players with a '0' rating have not played rugby at a comparable level of Super 12 or World Cup and have therefore been assigned a value of $50,000



ULTIMATE RUGBY

WHAT IS ULTIMATE RUGBY?

You too can be a first class rugby selector. Choose a team of 15 players from across all of the actual teams for a given tournament. Watch as your team accumulates points every week based on actual game by game performances of the individual players.

You can trade your players every week to improve your team. There is over $250,000 in cash to be won, with $18,000 every week given away to the best weekly and total team results, and the teams with the greatest improvement in points from the previous week.

OVER $250,000 IN CASH IS UP FOR GRABS! WITH $18,000 IN CASH WEEKLY TO BE GIVEN OUT.

HOW TO PLAY:

- Using the players published points and values (derived from the equivalent tournament level in 1999), select your team of 15 players.
- Fill out the entry form to the right - make sure the value of all your players is not more than 2.4 million.
- You must have a starting 15 covering all positions which you can choose out of the list of players from each of those positions.
- Scoring starts from March 10th, so it is important to have your entries in by then if you want to win some of the $70,000 cash grand prizes.

costs $19.95 + GST total cost $21.95 for each tournament team entered, and you may enter more than once for each tournament you participate in during the season. We will bring you as many rugby tournaments as possible from around the world as the season progresses.

HOW TO ENTER:

BY POST

Complete the Ultimate Rugby form and send it with payment by cheque, money order, credit card or 900 PIN code to: **Freepost ULTIMATE, Ultimate Rugby Limited, Private Bag 92090, Auckland.** Get your 900 PIN code by calling 900 4 RUGBY (900 487 429) this charges $19.95 + GST total cost $21.95 to your phone bill. Write down the PIN code on the entry form.

BY PHONE

Fill out the form and then call 0800 488 666. An operator will set up your team and mail confirmation to you.

THE INTERNET

Enter online by going to our website at http://www.skytv.co.nz You can enter by using a credit card or providing the PIN code assigned to you from a call to 0900 4 RUGBY (0900 478 429) will then charge your phone bill with $19.95 + GST total cost $21.95.

ULTIMATE RUGBY ENTRY FORM

code name club value

Team Name:

Loose Head Prop:

Hooker:

Tight Head Prop:

Lock:

Lock:

Blind Side Flanker:

Open Side Flanker:

Number 8:

Half Back:

1st Five Eight:

Right Wing:

2nd Five Eight:

Center:

Wing:

Full Back:

Name:

Address:

Telephone:

Cheque ☐ Mastercard ☐ Visa Card ☐

Credit Card No.

900 for Rugby PIN No.

300K	250K	200K	150K	100K	50K	TEAM	TOTAL COST
2	2	2	3	3	3	15	$2,400,000
6	0	1	0	0	8	15	$2,400,000
0	8	0	0	1	6	15	$2,400,000
0	0	11	0	0	4	15	$2,400,000
1	0	0	14	0	0	15	$2,400,000
3	3	2	0	0	7	15	$2,400,000
0	5	4	0	1	5	15	$2,400,000
0	0	7	6	0	2	15	$2,400,000
0	0	6	7	1	1	15	$2,400,000

If you do not wish to receive any further promotional material from Ultimate Rugby or its associates please tick this box. ☐

Ultimate Rugby - The only way to play the game.

RULES & REGULATIONS

PICKING A TEAM:

Based on their performance in 1999 first class tournaments all players have been given starting salary values between $50,000 and $300,000. There are 6 levels:

300,000	200,000	100,000
250,000	150,000	50,000

You, as the manager, have to choose 15 players from across all the actual teams participating in the tournament. Below we have provided you with some possibilities as team combinations. A small tip - look for the players that do not have Eagle Ratings, these could be bargains that have been placed in the lower salary class, use your knowledge and skills to select those up and coming players for your team.

300K	250K	200K	150K	100K	50K	TEAM	TOTAL COST
2	2	2	3	3	3	15	$2,400,000
6	0	1	0	0	8	15	$2,400,000
0	8	0	0	1	6	15	$2,400,000
0	0	11	0	0	4	15	$2,400,000
1	0	0	14	0	0	15	$2,400,000
3	3	2	0	0	7	15	$2,400,000
0	5	4	0	1	5	15	$2,400,000
0	0	7	6	0	2	15	$2,400,000
0	0	6	7	1	1	15	$2,400,000

SCORING

- Ultimate Rugby uses Eagle ratings based on 12 hours analysis of every first class NZ rugby match over the last 4 yrs.
- Each position has 5 factors which determine the players value over the rugby game. Factors of performance include tries, conversions, sin bins, penalties, meters on attack and defense, jumping and coverage and many others.
- Eagle Sports provides the Eagle rating of each players weekly performance with a score between 0 - 100. All players weekly performance ratings in a team are added together to give the team manager their weekly score. This is added to their season total.
- Players can play out of position during the season because ratings are comparable between positions.
- The weekly total of points is added to the total during the season to give a points total at the end of the tournament.

TEAM MANAGER PACKS

Confirmation of your selected team and your allocated PIN number will be posted to you upon receipt of your form or once your telephone or internet entry has been processed.

TRADES

With $250,000 in prizes at stake, team managers should trade their players to improve the chances of their team doing well. If your team is not doing well, you can trade players and the next week your team may generate lots of points from the performance of all the players. This could make you eligible for hundreds of prizes that we give away to team owners that make the largest weekly points totals and who also make the largest jumps up the ladder.

- Team managers can trade their players from 10th of March, 2000 onwards. Each trade costs $3.00.
- Team managers who keep their players for the entire season can also include them in their team next season at the price they paid this year, even if that player's value has risen a lot.
- The salary value to trade each player, changes every week according to their performance in that weeks game.
- Trades received or completed after 5pm before the first weekly game (Friday or Thursday) will apply to the following weeks games.
- "Team managers" can trade players at current market value through the season, which means the value of their team can exceed the salary cap during the season. You can trade your players by:
  • Call 0800 588 444 and get one of the operators to help you trade players. You can pay for this in multiples of $3.00 by: • Credit card with the operator • Charge to your telephone bill. Call 0900 3 TRADE, 0900 4 TRADE, 0900 5 TRADE up to 0900 9 TRADE. This will charge between $9.00 for 3 trades and $27.00 for 9 trades to your phone bill in multiples of $3.00. Call the operator with the PIN number to make your trades. You do not have to use all trades immediately • Payment into our bank account with receipt details for the operator.

STATISTICS

Statistics on the previous weeks games will be up dated every Monday by 5pm, so your teams results and standings will change every week and be available for viewing after 5.30pm each Monday.

PRIZES

All prize winners will be notified directly of their wins. Managers who have used the phone billing system to pay for their team managers packs and trades must have a current phone bill. Where there is more than one winner with the same score, the prize money for the prize positions involved will be added together and split evenly between the winners. For example if 2nd and 3rd have the same score, $1000 and $500 will be added and each will receive $750.

Three Separate Prize Pools Weekly:

For the highest total cumulative points achieved by teams for that week.

1st **$2,500**, 2nd **$1,000**, 3rd **$500**, 4th - 10th **$100**, **$50**, 26th - 50th **$25**

Prizes for the teams with the largest weekly increase in points relative to the previous week.

1st **$2,500**, 2nd **$1,000**, 3rd **$500**, 4th - 10th **$100**, 10th - 25th **$50**, 26th - 50th **$25**

For the highest team $ values each week.

1st **$2,500**, 2nd **$1,000**, 3rd **$500**, 4th - 10th **$100**, 10th - 25th **$50**, 26th - 50th **$25**

A Total Prize Pool of over $250,000.

[illegible]

Appendix E

Glossary

First Class Match:

A first class match is defined by the New Zealand Rugby Football Union (NZRFU) by a number of set criteria. Official NZRFU matches are afforded first class status. For any match other than an official NZRFU fixture, first-class status is only granted if the following three conditions are met:

1. The match is listed in the NZRFU fixtures list
2. Both teams are recognised "Senior A" selection.
3. Team Sheets from both unions are received by the NZRFU within 7 days. Formal application for first-class status must be made for any match played under the following circumstances:

1. The match is not included in the NZRFU fixtures list.
2. One or both teams is not a recognised "Senior A" selection.
3. One or both teams contain invitation players.
4. Any provincial team playing an overseas state or national team which is not an official NZRFU tour. Applications received for first-class status will only be considered under the following conditions:

1. Application is made by the host union.
2. The application includes correctly completed team sheets of both teams involved
3. Application is received by the NZRFU within 21 days of the match being played.

Lineout Positioning

Below is a generic lineout structure. Not all lineouts are structured as such, but there are key conventions employed. The hooker 2 throws to the lineout. Locks (4, 5) occupy jump at 2 and 4 in the lineout. 5 tends to be the larger of the two locks. Props occupy spots 1 and 3, and act as lifters, with the prop occupying the front of the lineout also providing a ball-winning option. 6 tends to be the taller or more likely lineout option jumping at 5, where the lock (jumping at 4) and number eight can lift in the lineout.

2 1 4 3 5 6 8 7

Phase

A phase is a period of undisrupted of play. The end of a phase is encountered when either a ruck or maul forms. Phases accumulate until the referee stops play.

Positional Layout

The positional layout presented below is a typical "scrum" formation, with the forwards (1-8) committed to the scrum, and the backs line up on attack.

Blindside							Midfield				Openside
		1		2		3					
	6		4		5		7				
				8							
					9						
							10				
								12			
	11								13		
										14	
							15				

Forwards

1=Tighthead Prop
2=Hooker
3=Loosehead Prop
4=Lock
5=Lock
6=Blindside Flanker
7=Openside Flanker
8=Number Eight

Backs

9=Halfback
10=First Five Eighth
11=Left Wing
12= Second Five Eight
13=Centre
14=Right Wing
15=Fullback

Note: In South Africa 6=Openside Flanker and 7=Blindside Flanker

Sequence

Sequences are the accumulated count of phases in a match. For example, the first set piece of a game may have three phases and the second set piece may have four phases. At such a point there have been seven sequences of play. This accumulation occurs to till the end of the match.

Set Piece

Set piece play is the means by which play is recommenced following the stoppage of play by the referee. Primarily scrums and lineouts are set pieces but occasionally kick-offs and penalties can be viewed as set piece.

Useful Websites for Further Information

www.rugbyheaven.com.au/
www.rugbyresource.co.uk/
www.rugbytactics.com/
www.coachingrugby.com/
www.nzrugby.com/

Appendix F

Half-Moon Related Methods

In the process of creating the half moon statistic, a number of similar methods were briefly investigated. As each of these methods provided ineffectual results they were not pursued. Relating to equation 3, if the absolute value of y was not taken, this gives rise to the full moon method where output values less than the mean output would have a higher z_{jk} value than those above the output mean. That is the range of the data is potentially the radius doubled (2r). This thins out the data representing z_{jk}. This is demonstrated below in a modification of equation 3

$$z_{jk} = \sqrt{r^2 - x_{jk}^2} - y_j \tag{9}$$

Taking the absolute value of the input as well as the absolute value of the output is redundant due to the squaring of the input x_{jk} in the early stages of the calculation.

The data can be condensed further by initially comparing the input and output together. Setting the minimum value as ν and the maximum as η presents the quarter method, which condenses the data for z_{jk} into a quarter circle, with the ν axis representing the minimum values and the η axis the maximum values. Expressed in a similar form to that of the equations (2-3), with x_j the input and y_j the output, the quarter moon method is:

$$\nu_{jk} = \min(|x_{jk}|, |y_j|) \tag{10}$$

$$\eta_{jk} = \max(|x_{jk}|, |y_j|) \tag{11}$$

and incorporated into the final stages of condensation as:

$$u_{jk} = \sqrt{r^2 - \nu_{jk}^2}$$

$$z_{jk} = u_{jk} - |\eta_j|$$

before being applied to equation 5 in the final calculation. In early trials, the moments representing the full moon and quarter methods did not produce any strong relationships with linear correlation when the linear case was presented. As a result these methods were not pursued. It is reasonable to expect a method that is to be used for detecting patterns will, when just the linear case is presented, map the correlation with a certain degree of accuracy.

Further methods explored included investigating the closest distance a point is to the circle, again with a set radius. This was found by a two step process firstly using the Pythagorean theorem to obtain the Euclidean distance between the point given by the co-ordinates representing the input and output and the origin, which, as mentioned earlier, is just the mean input and mean output (0,0). Secondly this value is subtracted from the radius. Equation 12 demonstrates this procedure as one step.

$$c_j = \sqrt{r^2 - (x_j^2 + y_j^2)} \tag{12}$$

However, the subtraction of the distance to the origin from the radius is a superfluous step, as this produces numerical values that are equivalent to comparison with the origin. Again this type of method produced unsatisfactory results.

Logically, the half moon statistic is generated retaining respect of the initial parameters. That is differences are calculated with respect to the y-axis. This relative frame of reference is lost in both the quarter-moon and absolute-distance methods. Intuitively, results are blurred in the full moon method as negative input values will always have a greater magnitude, z, than positive input values. Wave functions have the potential to cause problems for the half moon statistic, due to the folded result of considering the absolute value of the output. However, such functions are unlikely to occur in the context presented by this thesis.

References

Akers, C., & Miller, G. (Eds.). (2000). *The Sky Television Rugby Almanack of New Zealand* (66th ed.). Auckland: Hodder Moa Beckett.

Almedia, L. B. (1993). Neural Preprocessing Methods. In Cherkassky, V., Friedman J. H., & H. Wechsler (Eds.), *From Statistics to Neural Networks: Theory and Pattern Recognition Applications* (Vol. 136, pp. 213-225). Berlin: Springer-Verlag.

Amin, R.W, & Reynolds, M.R. &, Bakir, S, (1995). Non-parametric Quality Control Charts Based on the Sign Statistic. *Communications in Statistics – Theory and Methods.* 24(6), p.1597.

AnalySports. (2000). *User Manual: AnalyRugby*. Palmerston North: AnalySports.

Anderson, J.A. & Rosenfeld, E. (Eds.)(1988). *Neurocomputing: Foundations of Research.* Cambridge, MA: MIT Press.

Anderson, J.A. (1968). A Memory Storage Model Utilizing Spatial Correlation Functions. *Kybernetics.* 5, pp. 113-119.

Anderson, J.A. (1970). Two Models For Memory Organization. *Mathematical Biosciences.* 8, pp. 137-160.

Anderson, J.A. (1972). A Simple Neural Network Generating An Interactive Memory. *Mathematical Biosciences.* 14, pp. 197-220. Reprinted in Anderson & Rosenfeld (1988), pp. 181-192.

Anderson, J.A., Silverstein, J.W., Ritz, S.A. & Jones, R.S. (1977). Distinctive Features, Categorical Perception, And Probability Learning: Some Applications Of A Neural Model. *Psychological Review.* 84, pp 413-451. Reprinted in Anderson & Rosenfeld (1988), pp. 287-326.

Andrews, R., Diederich, J., & A. B. Tickle, (1995) Survey and Critique of Techniques for Extracting Rules from Trained Artificial Neural Networks. *Knowledge-Based Systems.* 8, pp. 373-389.

Andrews, R., Diederich, J., & Giles, L. (1995). Rule Extraction for Neural Networks. Available: www.aic.nrl.navy.mil/~aha/research/machine-learning.html [2001, 16/8/01].

Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and Critique of Techniques for Extracting Rules from Trained Artificial Neural Networks. *Knowledge-Based Systems*, 8(6), pp. 373-389.

Anton, H. (1995). *Calculus With Analytic Geometry.* (5th ed.). New York: John Wiley & Sons.

Arnold, J.A., Brown, B., Micheli, R.P., & Coker, T.P. (1980). Anatomical And Physiologic Characteristics To Predict Football Ability. *The American Journal Of Sports Medicine.* 8 (2), pp. 119-122.

Azuaje, F., Dubitzky, W., Lopes, P., Black, N., Adamson, K., Wu, X., & White, J. A. (1999). Predicting Coronary Disease Risk Based on Short-Term RR Interval Measurements: A Neural Network Approach. *Artificial Intelligence in Medicine*, 15, pp. 275-297.

Bakir, S.T, & Reynolds, M.R.Jr, pp. (1979). A Non-Parametric Procedure for Process Control Based on Within-Group Ranking. *Technometrics.* 21(2), pp. 175-183.

Balakrishan, P. V. S., Cooper, M. C., Jacob, V. S., & Lewis, P. A. (1994). A Study of Classification Capabilities of Neural Networks Using Unsupervised Learning: A Comparison with K-Means Clustering. *Psychometrika*, 59(4), pp. 509-525.

Barton J.G. & Lees, A. (1995). The Feasibility Of A Neural Network Approach For Assessing Insole Materials: A Preliminary Study. XV Congress of the International Society of Biomechanics, Jyvaskyla, Finland, (July, 1995).

Barton J.G. & Lees, A. (1997). An Application Of Neural Networks For Distinguishing Gait Patterns On The Basis Of Hip-Knee Joint Angle Diagrams. *Gait and Posture.* 5 (1), pp. 28-33.

Barton J.G., Lees, A., & Wit, A. (1996). The Use Of Neural Networks To Estimate Ground Reaction Forces From Electromyographic Activity. 2nd National Conference of the IPEMB and 3rd Biennial Radiotherapy Physics Meeting, Leeds, UK, (September, 1996).

Bate, R. (1987). Football Chance: Tactics And Strategy. Communications To The 1st World Congress Of Science And Football. pp. 293-301 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football*. E & Fn Spon: London.

Baxt, W.G. (1991). Use Of An Artificial Neural Network For The Diagnosis Of Myocardial Infarction. *Annals of Internal Medicine*. 115, pp. 843-848.

Becker, S. (1991). Unsupervised Learning Procedures for Neural Networks. *International Journal of Neural Systems*, 2(1 & 2), pp. 17-33.

Begin, J., & Proulx, R. (1996). Categorization in Unsupervised Neural Networks: The Eidos Model. *IEEE Transactions on Neural Networks*, 7(1), pp. 147-154.

Bell, W., Cobner, D., Cooper, S.M., & Phillips S.J. (1991). Anareobic Performance And Body Composition Of International Rugby Players. In Reilley, T., Lees, A., Davids, K., & Murphy, W.J. (Eds.). *Science and Football II*. London: E. & F.N. Spon

Bennett, J. M. (1998). Baseball. In J. M. Bennett (Ed.), Statistics in Sport (pp. 25-64). New York: Arnold.

Bennett, J. M. (1992). Did Shoeless Joe Jackson Throw the 1919 World Series? Proceedings of the Section on Statistics in Sports of the American Statistical Association, pp. 33-38.

Berkmann, J. (1990). *The Complete Guide To Test Cricket In The Eighties*. Partridge Press: London

Berry, M. J. A., & Linoff, G.S. (2000). *Mastering Data Mining*. New York: John Wiley & Sons.

Berry, M. J. A., & Linoff, G.S. (1997). *Data Mining Techniques*. New York: John Wiley & Sons.

Bhandri, I., Colet, E., Parker, J., Pines, Z., Pratap, R. & Ramanujam, K. (1997). Advanced Scout: Data Mining And Knowledge Discovery In NBA data. *Data Mining and Knowledge Discovery*. 1 (1), pp. 121-125.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford.

Block, H.D. (1962). The Perceptron: A Model For Brain Functioning, I. *Reviews of Modern Physics*. 34, pp 123-135. Reprinted in Anderson & Rosenfeld (1988), pp. 138-150.

Borggaard, C. (1995). Modeling Non-linear Data Using Neural Networks Regression in Connection with PLS or PCA. *Frontiers in Analytical Spectroscopy* (163), pp. 209-217.

Bourquin, J., Schmidli, H., Hoogevest, P. v., & Leuenberger, H. (1998). Advantages of Artificial Neural Networks (ANNs) as Alternative Modeling Technique for Data Sets Showing Non-Linear Relationships using Data from a Galenical Study on a Solid Dosage Form. *European Journal of Pharmaceutical Sciences*, 7,pp. 4-16.

Bracewell, B. P. (2001). *Coaching - The 'Contract Syndrome' Disease*. Available: www.cricketnz.co.nz/page19.html (2001, 18/9/01).

Bracewell, P.J. (2003). Monitoring Meaningful Rugby Ratings. *Journal of Sport Sciences*. (To Appear).

Bracewell, P. J. (2002). Implementing Statistics in a Diagnostic Coaching Structure. *Research Letters in the Information and Mathematical Sciences*, 3, pp. 79-84.

Bracewell, P. J. (2001). Perception of Individual Rugby Player Performance and the Impact of Non-Performance on Statistical Analyses. *Research Letters in the Information and Mathematical Sciences*, 2, pp. 19-22.

Bracewell, P.J. (2000). *The Eagle Rating: An Explanation for Spectators*. Available: www.eaglesports.co.nz. (2000, 29/3/00)

Bracewell, P.J. (1999). *Statistical Methods For Cricket Team Selection*. Unpublished Masters Thesis, Massey University, Albany Campus, New Zealand.

Bracewell, P.J. A Versatile Heuristic Test for Independence. *Australian and New Zealand Journal of Statistics*. Manuscript under Review.

Bracewell, P.J. An Expert System Framework for Assisting Diagnostic Rugby Coaching of Individuals using Multivariate Data. *Expert Systems*. Manuscript under Review.

Bracewell, P.J., Meyer, D.H., & Ganesh, S. (2001). Extracting Measures of Performance of Individual Rugby Players: Data Mining in Sport. *Proceedings of Artificial Neural Networks and Expert Systems Conference*. pp. 38-42. Otago University: Dunedin.

Breiman, L. (2001). Statistical Modelling: The Two Cultures. *Statistical Science*. 16 (3) pp. 199-231.

Cameron, D. (1980). *Retreat from Glory*. Auckland: Hodder and Stoughton.

Capobianco, E. (2000). Neural Networks and Statistical Inference: Seeking Robust and Efficient Learning. *Computational Statistics and Data Analysis*, 32, pp. 443-454.

Carpenter, G.A. & Grossberg, S. (1983), A Massive Parallel Architecture For A Self-Organizing Neural Pattern Recognition Machine. *Computer Vision, Graphics and Image Processing*. 37, pp. 54-115.

Casella, G., & Berger, R.L., (1990). *Statistical Inference*. California: Duxbury Press.

Casti, J. (2001). Would-be Worlds: The Science and the Surprise of Artificial Worlds. Presentation at Massey University, Albany Campus (9/3/01).

Chakraborty, G., & Chakraborty, B. (2000). A Novel Normalization Technique for Unsupervised Learning in ANN. *IEEE Transactions on Neural Networks*, 11(1), pp. 253-257.

Champ, C.W., & Woodall, W.H. (1987). Exact results for Shewhart Control Charts with Supplementary Runs Rules. *Technometrics*. 29(4), pp. 393-399.

Chatfield, C. & Collins, A.J. (1980). *Introduction To Multivariate Analysis*. Cambridge: Chapman & Hall.

Chatfield, C. (1989). *The Analysis Of Time Series, An Introduction*. (4th ed). London: Chapman & Hall.

Chatfield, C. (1997). Data mining. *Royal Statistical Society News*. 25, pp. 1-2.

Chatfield, C. (1995). Model Uncertainty, Data Mining and Statistical Inference. *Journal of the Royal Statistical Society, Series A*. 158 (3), pp. 419-466.

Cheng, B. & Titterington, D.M. (1994). Neural Networks: A Review Form A Statistical Perspective. *Statistical Science*. 9 (1), pp. 2-54.

Cherkassky, V., & Mulier, F. (1998). *Learning From Data: Concepts, Theory and Methods*. New York: John Wiley & Sons.

Chervenjakov, M. (1987). Assessment Of The Playing Effectiveness Of Soccer Players. Communications To The 1st World Congress Of Science And Football. pp. 286-292 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football*. E & Fn Spon: London.

Chu, M. (2000). Team Dynamics. pp. 1-11. In NZRFU. *Advanced Coaching Course*. Pukekohe: NZRFU.

Chung, R.-I., & Hsiao, P.-Y. (1997). Unsupervised Query-Based Learning of Neural Networks Using Selective-Attention and Self-Regulation. *IEEE Transactions on Neural Networks*, 8(2), pp. 205-217.

Ciampi, A. & Lechevallier, Y. (1997). Statistical Models As Building Blocks Of Neural Networks. *Communications in Statistics – Theory and Methods*. 26 (4), pp. 991-1009.

Cliff, N. (1987). *Analyzing Multivariate Data*. Orlando: Harcourt Brace Jovanovich.

Collins, D. R., & Hodges, P. B. (2001). *A Comprehensive Guide to Sports Skills Tests and Measurement* (2nd ed.). Lanham, Maryland: The Scarecrow Press, Inc.

Cook, E. (1977). An Analysis Of Baseball As A Game Of Chance By The Monte Carlo Method. pp. 50-54 In Ladany S.P., & Machol R.E. (Eds.). *Optimal Strategies in Sport*. Amsterdam: North Holland.

Cooper, J. S. B. (1999). Artificial neural networks versus multivariate statistics: an application from economics. *Journal of Applied Statistics*, 26(8), pp. 909-921.

Cowen, J.D. & Sharp, D.H. (1988). Neural Nets. *Quarterly Reviews of Biophysics*. 21, pp. 365-427.

Crick, F. (1989). The Recent Excitement About Neural Networks. *Nature*. 337 January, pp. 129-132.

Crosier, R.B. (1988). Multivariate Generalizations Of Cumulative Sum Quality Control Schemes. *Technometrics*. 30 (3), pp. 291-303.

Cryer, J.D. (1986) *Introductory Time Series Analysis*. Boston: PWS-Kent Publishing Company.

Cybenko, G. (1989). Approximation By Superpositions Of A Sigmoidal Function. *Mathematics of Control, Signals, and Systems*. 2, pp. 303-314.

de Lacy, H., & Fox, G. (2000). *Think & Play Winning Rugby*. Auckland: Harper Sports.

Delicado, P. (2001). Another Look at Principal Curves and Surfaces. *Journal of Multivariate Analysis*, 77, pp. 84-116.

Demuth, H., & Beale, M. (1998). *Neural Network Toolbox User's Guide*. Natick: MathWorks.

Deustch, M, Kearney, G.A. & Rehrer, N.J. (1999). A Comparison Of Competition Work Rates In Elite Club And 'Super 12' Rugby Union. Communications To The 4th World Congress Of Science And Football. Journal Of Sports Sciences. 17, pp. 809-810

Diederich, J. (1992). Explanation and Artificial Neural Networks. *International Journal of Man-Machine Studies*, 37, pp. 335-357.

Digby, P.G.N., & Kempton, R.A. (1987). *Multivariate Analysis of Ecological Communities London*. New York : Chapman and Hall.

Doganaksoy, N, & Faltin, F.W, pp. Tucker, W.T. (1991). Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment. *Communication in Statistics*, pp. Theory and Methods. 20(9), pp. 2775-2790.

Doherty, S. K., Gomm, J. B., & Williams, D. (1997). Experiment Design Considerations for Non-Linear System Identification Using Neural Networks. *Computers and Chemical Engineering*, 21(3), pp. 327-346.

Douge, B. (1987). Football: The Common Threads Between The Games. Communications To The 1st World Congress Of Science And Football. pp. 3-6 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football*. E & Fn Spon: London.

Duh, M.-S., Walker, A. M., Pagano, M., & Kronlund, K. (1998). Prediction and Cross-Validation of Neural Networks Versus Logistic Regression: Using Hepatic Disorders as an Example. *American Journal of Epidemiology*, 147(4), pp. 407-413.

Edelman, G. M., & Reeke, G. N. J. (1982). Selective Networks Capable of Representative Transformations, Limited Generalizations, and Associative Memory. Proceedings of the National Academy of Sciences, USA, 79, pp. 2091-2095.

Elman, J.L. (1990). Finding Structure In Time. *Cognitive Science*. 14, pp. 179-211.

Everitt, B.S., & Dunn, G. (1991). *Applied Multivariate Data Analysis*. New York: Halsted Press.

Fan, K. (1996). Approximate Dimension Reduction for Filtered Markov Chains. *Communications in Statistics: Stochastic Models*, 12(1), pp. 1-16.

Faraggi, D., & Simon, R. (1995). The Maximum Likelihood Neural Network as a Statistical classification Model. *Journal of Statistical Planning and Inference*, 46(1), pp. 93-104.

Fausett, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. New Jersey: Prentice Hall.

Fayyad, U.M. (1997). Editorial. *Data Mining and Knowledge Discovery*. 1 (1), pp. 5-10.

Feraud, R. (1997). PCA, *Neural Networks and Estimation for Face Detection*. pp. 424-432.

Fernandez, J. M. V., & Manteiga, W. G. (2000). Resampling for Checking Linear Regression Models Via Non-Parametric Regression Estimation. *Computational Statistics and Data Analysis*, 35, pp. 211-231.

Fildes, R. & Makridas, S. (1994). *The Impact Of Empirical Accuracy Studies On Time Series Analysis And Forecasting*. Discussion Paper. Management School, Lancaster University, Lancaster.

Forrest, D., & Simmons, R. (2000). Making up the Results: The Work on Football. *The Statistician*, 49(2), pp. 253-260.

Fraleigh, J. B. (1994). *A First Course in Abstract Algebra* (5th ed.). New York: Addison-Wesley.

Frangos, C.C. (1984). *The Jackknife, The Bootstrap and Other Non-parametric Methods in Statistics: Survey and Applications*. Johannesburg: University of the Witwatersrand.

Franks, A.M., Williams, A.M., Reilly, T. & Nevill, A. (1999). Talent Identification In Elite Youth Soccer Players: Physical And Physiological Characteristics. Communications To The 4th World Congress Of Science And Football. *Journal Of Sports Sciences*. 17, p. 812.

Franks, B.D., & Deutsch, H. (1973). *Evaluating Performance In Physical Education*. New York: Academic Press.

Franks, I.M. & Miller, G. (1986). Eyewitness Testimony In Sport. *Journal Of Sport Behaviour*. 9, pp. 39-45.

Freund, J.E, pp. (1992). *Mathematical Statistics*. (5th ed). New Jersey: Prentice-Hall International, Inc.

Fu, L. (1994). Rule Generation From Neural Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 4(8), pp. 1114-1124.

Garcez, A. S. d. A., Broda, K., & Gabbay, D. M. (2001). Symbolic Knowledge Extraction From Trained Neural Networks: A Sound Approach. *Artificial Intelligence*, 125, pp. 155-207.

Garson, G.D. (2001). *PA 765 Statnotes: An Online Textbook*. Available http://www2.chass.ncsu.edu/garson/pa765/statnote.htm (2001, 3/10/01).

Gerber, C. (1996). Excavate Your Data. *Datamation*. 42 (9), pp. 40-43.

Girolami, M., Cichocki, A., & Amari, S.-I. (1998). A Common Neural-Network Model for Unsupervised Explanatory Data Analysis and Independent Component Analysis. *IEEE Transactions on Neural Networks*, 9(6), 1pp. 495-1501.

Girshick, M.A., Mosteller, F.A., & Savage, L.J. (1946). Unbiased Estimates for Certain Binomial Sampling Problems with Applications. *Annals of Mathematical Statistics*. 17, pp. 13-23.

Glymour, C., Madigan, D., Pregibon, D. & Smyth, P. (1997). Statistical Themes And Lessons For Datamining. *Data Mining and Knowledge Discovery*. 1 (1), pp. 11-28.

Glymour C., Madigan D., Pregibon D., & Smyth P. (1996). *Statistical Inference And Data Mining*. Communications of the ACM, 39(11), pp. 35-41.

Gordon, A.D. (1981). *Classification: Methods For The Exploratory Analysis Of Multivariate Data*. New York: Chapman and Hall.

Gorman, R. P., & Sejnowski, T. J. (1988). Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets. *Neural Networks*, 1, pp. 75-89.

Grady, N.W., Schryver, J.C., & Leuze, M.R. (1999). *Mining for Personal Profiles*. Available: http://citeseer.nj.nec.com/482614.html (2002, 18/2/02).

Greenberg, J., & Baron, R.A. (1997). *Behavior in Organizations*. (6th ed.). New Jersey: Prentice-Hall.

Grossberg, S. (1976). Adaptive Pattern Classification And Universal Recoding. I: Parallel Development And Coding Of Neural Feature Detectors. *Biological Cybernetics*. 23, pp. 121-134.

Haines, L. M. (1997). Optimal Design for Neural Networks. Paper presented at the Proceedings of AMS-IMS-SIAM, Seattle.

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). *Multivariate Data Analysis with Readings*. (4th ed.). Prentice Hall: Englewood Cliffs, New Jersey.

Hall, M. A., & Holmes, G. (2000). *Benchmarking Attribute Selection Techniques for Data Mining*. Hamilton: Waikato University.

Hand, D. J., Blunt, G., Kelly, M. G., & Adams, N. M. (2000). Data Mining for Fun and Profit. *Statistical Science*, 15(2), pp. 111-131.

Hand, D.J. (1999). Statistics and Data Mining: Intersecting Disciplines. *SIGKDD*. 1, pp. 16-19.

Hand, D.J. (1998). Data Mining: Statistics And More? *The American Statistician*. 52 (2), pp. 112-118.

Hartigan, J.A. (1975). *Clustering Algorithms*. New York: Wiley.

Hawkins, J.M. (Ed)(1990). *Oxford Senior Dictionary*. (7th ed.). Oxford University Press: London.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation* (2nd ed.). New Jersey: Prentice & Hall.

Hayter, A.J, pp. Tsui, K-L. (1994). Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology*. 26(3), pp. 197-208.

Hazeldine, R., & McNab, T. (1991). *Fit for Rugby*. London: Kingswood Press.

Hebb, D.O. (1949). *The Organisation Of Behaviour*. New York: John Wiley & Sons. Introduction and Chapter 4 reprinted in Anderson & Rosenfeld (1988), pp. 45-56.

Hecht-Nielsen, R. (1988). Neurocomputing: Picking The Human Brain. *IEE Spectrum*. March, pp. 36-41.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley: Reading, Mass.

Hinton, G. E. (1992). How Neural Networks Learn From Experience. *Scientific American*, pp. 104-109.

Hinton, G.E. & Sejnowski, T.J. (1986). Learning And Relearning In Boltzmann Machines. In *Parallel Distributed Processing* 1 (7).

Hodge, K. (1999) Motivational Profiles In Rugby Union. Communications To The 4th World Congress Of Science And Football. *Journal Of Sports Sciences*. 17, p. 827.

Hodge, K., & McKenzie, A. (1999). *Thinking Rugby*. Auckland: Reed.

Honeybourne, J., Hill, M., & Moors, H. (1996). *Advanced Physical Education and Sport*. Cheltenham: Stanley Thornes Ltd.

Hopfeld, J.J. (1982). Neural Networks And Physical Systems With Emergent Collective Computational Abilities. Proceeding Of The National Academy of Scientists. 79, pp. 2554-2558. Reprinted in Anderson & Rosenfeld (1988), pp. 579-584.

Hotelling, H. (1947). *Multivariate Quality Control. Techniques Of Statistical Analysis*. Eisenhart, Hastay & Wallis (Eds.). New York: Mcgraw-Hill.

Hughes, M. (1991). Notation Analysis In Football. Communications To The 2nd World Congress Of Science And Football. pp. 151-159 In Reilly, T., Clarys, J., & Stibbe, A. (Eds.) *Science And Football II*. E & Fn Spon: Ipswich.

Hughes, M., Robertson, K. & Nicholson, A. (1987). Comparison Of Patterns Of Play Of Successful And Unsuccessful Teams In The 1986 World Cup For Soccer. Communications To The 1st World Congress Of Science And Football. pp. 363-367. In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football*. E & Fn Spon: London.

Hayward, L. (2000). Available: http://www.homestead.com/rugbycoach/ (2000, 21/5/00).

Hwang, J. T. G., & Ding, A. A. (1997). Prediction Intervals for Artificial Neural Networks. *Journal of the American Statistical Association*, 92(438), pp. 748-757.

Isham, V., (1993). Statistical Aspects of Chaos: A Review. pp. 124-200. In Barndorff-Nielsen, O.E., Jensen, J.L. & Kendall. W.S. (Eds.). *Networks and Chaos - Statistical and Probabilistic Aspects*, Chapman & Hall: London.

Jackson, J.E. (1980). Principal Components And Factor Analysis: Part I – Principal Components. *Journal Of Quality Technology*. 12.

Jensen, J.L., (1993). Chaotic Dynamical Systems with a View Towards Statistics: A Review. pp. 201-250. In Barndorff-Nielsen, O.E., Jensen, J.L. & Kendall. W.S. (Eds.). *Networks and Chaos - Statistical and Probabilistic Aspects*, Chapman & Hall: London.

Jiang, J., Wang, J., Chu, X., & Yu, R. (1996). Neural Network Learning To Non-Linear Principal Component Analysis. *Analytica Chimica Acta* (336), pp. 209-222.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Johnson, N.L. & Leone, F.C. (1964). *Statistics and Experimental Design: In Engineering and the Physical Sciences*. Volume 1. New York: Wiley.

Johnson, N.L. & Kotz, S. (1970). *Continuous Univariate Distributions, Volume 2.* Boston: Houghton Mifflin.

Jones, B. J., Wells, L. J., Peters, R. E., & Johnson, D. J. (1993). *Guide to Effective Coaching Principles and Practice* (3 ed.). Boston: WCB McGraw-Hill.

Jones, L. K. (2000). Local Greedy Approximation for Nonlinear Regression and Neural Network Training. *The Annals of Statistics*, 28(5), pp. 1379-1389.

Jordan, M.I. (1989). Serial Order: A Parallel, Distributed Processing Approach. In *Advances in Connectionist Theory: Speech* (eds. J. Elman & D. Rummelhart. Hillsdale, NJ: Erlbaum.).

Jorgensen, M & Gentleman, R. (1998). Data Mining. *Chance*. 11 (2), pp. 34-39.

Kallenberg, W.C.M., & Ledwina, T. (1999). Data Driven Rank Tests for Independence. *Journal of the American Statistical Association*. 94. pp. 285-301.

Kambhatla, N., & Leen, T. K. (1994). Fast Nonlinear Dimension Reduction. *Advances in Neural Information Processing Systems*, 6, pp. 152-159.

Kano, Y., & Harada, A. (2000). Stepwise Variable Selection in Factor Analysis. *Psychometrika*.

Kaplan, R. M. (1980). How do Fans and Oddsmakers Differ in their Judgements of Football Teams. *Personality and Social Psychology Bulletin*, 6(2), pp. 287-292.

Karayiannis, N. B., & Mi, G. W. (1997). Growing Radial Based Neural Networks: Merging Supervised and Unsupervised Learning with Neural Growth Techniques. *IEEE Transactions on Neural Networks*, 8(6), pp. 1492-1506.

Kitts, B., Freed, D., & Vrieze, M. (2000). *Cross-Sell: A Fast Promotion-Tunable Customer-Item Recommendation Method Based On Conditional Independent Probabilities.* In Proceedings of ACM SIGKDD International Conference, pp. 437-446.

Knorr-Held, L. (2000). Dynamic Rating of Sports Teams. *The Statistician*, 49(2), pp. 261-276.

Kohonen, T. (1972). Correlation Matrix Memories. *IEEE Transactions on Computers*. C-21, pp. 353-359.

Kohonen, T. (1982). Self-Organized Formation Of Topologically Correct Feature Maps. *Biological Cybernetics*. 43, pp. 59-69.

Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE*, 78(9), pp. 1464-1480.

Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Berlin: Springer.

Kosko, B. (1991). Structural Stability of Unsupervised Learning in Feedback Neural Networks. *IEEE Transactions on Automatic Control*, 36(7), pp. 785-792.

Kramer, M.A. (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AIChE*. 37(2): pp. 233-243.

Kruschke, J. K., & Movellan, J. R. (1991). Benefits of Gain: Speeded Learning and Minimal Hidden Layers in Back-Propagation Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(1), pp. 273-280.

Krzanowski, W.J. & Marriott, F.H.C. (1994). *Multivariate Analysis, Part I*. London: Edward Arnold.

Krzanowski, W.J. (1988). *Principles Of Multivariate Analysis*. Oxford: Clarendon Press.

Kuamr, V. P., & Manolakos, E. S. (1997). Unsupervised Statistical Neural Networks for Model-Based Object Recognition. *IEEE Transactions on Signal Processing*, 45(11), pp. 2709-2718.

Kuhn, W. (1995) Comparison Of Australian Rules Football And German Professional Soccer With Regard To Planning And Execution Of Training And Competition. Communications To The 3rd World Congress Of Science And Football. *Journal Of Sports Sciences*. 13, pp. 514-515.

Lahav, O. (1994). Artificial Neural Networks as Non-Linear Extensions of Statistical Methods in Astronomy. *Vistas in Astronomy*, 38, pp. 251-256.

Langley, P. (1997). *Machine Learning For Adaptive User Interfaces.* Proceedings of the 21st German Annual Conference on Artificial Intelligence. pp. 53-62. Freiburg, Germany: Springer.

Larson, H.J. (1982). *Introduction to Probability Theory and Statistical Inference.* (3rd ed.). Toronto: Wiley.

Levine, D.M., Krehbiel, T.C., & Berenson, M.L. (2000). *Business Statistics.* (2nd ed.). Prentice Hall: Upper Saddle River, New Jersey.

Lindgren, B.W. (1993). *Statistical Theory.* (4th ed.). New York: Chapman & Hall.

Lines, B.C., & Bracewell, P.J. (2000). Quantification of Individual Rugby Player Ability Through Multivariate Analysis & Data Mining. *Graduates in Industry Fellowship Application.* Submitted to Technology New Zealand 23/6/00.

Liu, R.Y., & Tang, J. (1996). Control Charts For Dependent And Independent Measurements Based On Bootstrap Methods. *Journal of the American Statistical Association* 91, pp. 1694-1700.

Lotlikar, R., & Kothari, R. (2000). Bayes-Optimality Motivated Linear and Multilayered Perceptron-Based Dimensionality Reduction. *IEEE Transactions on Neural Networks*, 11(2), pp. 425-463.

Lowe, J. (1997). Neural Networks. *The Actuary*. October, pp. 20-21.

Lowry, C.A, & Montgomery, D.C. (1995). A Review of Multivariate Control Charts. *IIE Transactions*. 27, pp. 800-810.

Lowry, C.A., Woodall, W.H., Champ, C.W., & Rigdon, S.E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*. 34(1), pp. 46-53.

Luijtens, K., Symons, F., & Vuylsteke-Wauters, M. (1994). Linear and Non-linear Canonical Correlation Analysis: An Exploratory Tool for the Analysis of Group-Structured Data. *Journal of Applied Statistics*, 21(3), pp. 43-61.

Macdonald, N. (1984). Avoiding The Pitfalls In Player Selection. *Coaching Science Update*. pp.41-45.

Mackinnon, M.J., & Glick, N. (1999). Data Mining And Knowledge Discovery In Databases – An Overview. *Australian & New Zealand Journal Of Statistics*. 41 (3), pp. 255-276.

Maindonald, J. H. (1999). *New Approaches to Using Scientific-Data Statistics: Data Mining & Related Technologies in Research & Research Training*. Available: http:/www.anu.edu.au/academia/graduate/papers/gs98_2.html (1999, 26/10/1999].

Maire, F. (1999). Rule-Extraction by Backpropagation of Polyhedra. *Neural Networks*, 12, pp. 717-725.

Malouche, Z., & Macchi, O. (1998). Adaptive Unsupervised Extraction of One Component of a Linear Mixture with a Single Neuron. *IEEE Transactions on Neural Networks*, 9(1), pp. 123-138.

Malthouse, E.C. (1998). Limitations of Nonlinear PCA as Performed with Generic Neural Networks. *IEEE Transactions on Neural Networks*. 9(1): pp. 165-173.

Manly, B. F. J. (1994). *Multivariate Statistical Methods: A Primer* (2nd ed.). London: Chapman & Hall.

Markham, I. S., & Ragsdale, C. T. (1995). Combing Neural Networks and Statistical Predictions to Solve the Classification Problem in Discriminant Analysis. *Decision Sciences*, 26(2), pp. 229-242.

McClelland, J.L., Rumelhart, D.E. & The PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.

McCulloch, W.S. & Pitts, W. (1943). A Logical Calculus Of The Ideas Immanent In Nervous Activity. *Bulletin Of Mathematical Biophysics*. 5, pp. 115-133. Reprinted in Anderson & Rosenfeld (1988), pp. 18-28.

McGilchrist, C.A, & Woodyer, K.D, pp. (1975). Note on a Distribution-Free Cusum Technique. *Technometrics*. 17(3), pp. 321-325.

McKenna, M.J., Patrick, J.D., Sandstom, E.R. & Chennels, M.H.D. (1987). Computer-Video Analysis Of Activity Patterns In Australian Rules Football. Communications To The 1st World Congress Of Science And Football. pp. 274-281 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football*. E & Fn Spon: London.

McKenzie, A., Hodge, K., & Sleivert, G. (Eds.). (2000). *Smart Training for Rugby*. Auckland: Reed.

McLean, T. (1979). *Mourie's All Blacks*. Auckland: Hodder and Stoughton.

Medsker, L., Turban, E., & Trippi, R. R. (1993). Neural Network Fundamentals for Financial Analysts. In R. R. Trippi & E. Turban (Eds.), *Neural Networks in Finance and Investing* (pp. 3-25). Chicago: Probus.

Mena, J. (1999). *Data Mining Your Website*. Woburn, MA: Butterworth-Heinemann.

Meyer, D. H. (1995). Disadvantages of Moving Averages. *The New Zealand Statistician*, 30(1), pp. 9-22.

MINITAB (1996). *MINITAB reference manual*. Release 11. USA

Minsky, M.L. & Papert, S.A. (1988). *Perceptrons, Expanded Edition*. Cambridge, MA: MIT Press. Original edition, 1969.

Montgomery, D.C. (1997). *Introduction to Statistical Quality Control*. (3rd ed.). New York: John Wiley & Sons.

Morrison, D.F. (1990). *Multivariate Statistical Methods* (3rd Ed.). Singapore: Mcgraw-Hill.

Morton, H. (2000). A Sports Statistician Looks at the Tri-Nations Series. *NZ Rugby World* (37), pp. 85-87.

Mosteller, F. (1952). The World Series Competition. *Journal Of The American Statistical Association*. 65, pp. 322-380.

Mosteller, F. (1997) Lessons from Sport Statistics. *American Statistical Association*. 51 (4), pp. 305-310.

Moxon, B. (1996). Defining Data Mining. *DBMS Data Warehouse Supplement*. August.

Murata, N., Yoshizawa, S., & Amari, S.-i. (1994). Network Information Criterion - Determining the Number of Hidden Units for an Artificial Neural Network. *IEEE Transactions on Neural Networks*, 5(6), pp. 865-872.

Murnion, S. D. (1996). Spatial Analysis using Unsupervised Neural Networks. *Computers and Geosciences*, 22(9), pp. 1027-1031.

Neath, A. A., & Cavanaugh, J. E. (2000). A Regression Model Selection Criterion Based on Bootstrap Bumping for Use With Resistant Fitting. *Computational Statistics and Data Analysis*, 35, pp. 155-169.

Newbold, P., Agiakloglou, C. & Miller, J. (1993). Long-Term Inference Based On Short-Term Forecasting Models. In *Developments in Time Series Analysis.* Subba Rao, T. (Ed), pp. 9-25. London: Chapman and Hall.

Nicole, S. (2000). Feedforward Neural Networks For Principal Components Extraction. *Computational Statistics and Data Analysis* (33), pp. 425-437.

NZRFU. (1991). *Coaching Accreditation Manual: Level 3.* Wellington: NZRFU.

NZRFU (1998). *Principles of Rugby Coaching.* Crawford, E. (ed.). Wellington: NZRFU.

NZRFU. (2000) *Advanced Coaching Course.* Paper presented at the NZRFU Northern Region Advanced Coaching Course, Pukekohe Stadium 9/6/00-11/6/00).

NZRFU. (2001) *Ensuring the Future Success of New Zealand Rugby*. Paper presented at the Media Presentation (13/01/2000)..

O'Hare, M. (1996). Chaos Pitch. *New Scientist.* June. pp. 24-28.

Orre, R., Lansner, A., Bate, A., & Lindquist, M. (2000). Bayesian Neural Networks with Confidence Estimations Applied to Data Mining. *Computational Statistics and Data Analysis*, 34, pp. 473-493.

Ott, L., Mendenhall, W., (1985). *Understanding Statistics.* (4th ed.). Boston: Duxbury Press.

Owen, A. B. (2000). Assessing Linearity in High Dimensions. *The Annals of Statistics*, 28(1), pp. 1-19.

Pal, N. R., & Eluri, V. K. (1998). Two Efficient Connectionist Schemes for Structure Preserving Dimensionality Reduction. *IEEE Transactions on Neural Networks*, 9(6), pp. 1142-1154.

Palenski, R., McMillan, N., & Chester., R. (2000). *Men in Black: 20th Century Commemorative Edition.* Hodder Moa Beckett: Auckland.

Partridge, D., Mosher, R.E. & Franks, I.M. (1991). A Computed Assisted Analysis Of Technical Performance – A Comparison Of The 1990 World Cup And Intercollegiate Soccer. Communications To The 2nd World Congress Of Science And Football. pp. 221-231 In Reilly, T., Clarys, J., & Stibbe, A. (Eds.) *Science And Football II.* E & Fn Spon: London.

Patrick, J.D., Mckenna, M.J. (1987). The Caber Computer System: A Review Of Its Application To The Analysis Of Australian Rules Football. Communications To The 1st World Congress Of Science And Football. P267-273 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football.* E & Fn Spon: London.

Pedrycz, W., & Waletzky, J. (1997). Neural-Network Front Ends in Unsupervised Learning. *IEEE Transactions on Neural Networks*, 8(2), pp. 390-401.

Pienaar, A. E., Spammer, M. J. & Steyn, H. S. Jr. (1998). Identifying And Developing Rugby Talent Among 10-Year-Old Boys: A Practical Model. *Journal Of Sports Sciences*. 16, pp. 691-699.

Poff, N. L., Toka, S., & Johnson, P. (1996). Steam Hydrological and Ecological Responses to Climate Change Assessed with an Artificial Neural Network. *Limnology and Oceanography*, 41(5), pp. 857-863.

Pollard, R., Reep, C. & Hartley, S. (1987). The Quantitative Comparison Of Playing Styles In Soccer. Communications To The 1st World Congress Of Science And Football. pp. 309-315 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football.* E & Fn Spon: London.

Potter, T. (2000). Eagle Flew High In NZ Business. *Sunday Star-Times.* Business Section, p. 3 (17/12/2000).

Quarrie, K.L. & Williams, S. (1999) Factors Associated With Pre-Season Fitness Attributes Of Rugby Players. Communications To The 4th World Congress Of Science And Football. *Journal Of Sports Sciences*. 17, p. 817.

Quesenberry, C.P. (1991). SPC Q Charts For Start-Up Processes And Short Or Long Runs. *Journal Of Quality Technology*. 23 (3), pp. 213-224.

Rattue, C. (2000). Rugby Sweat, Files and Videotape. *Weekend Herald*, p. 4. (11-12/3/01)

Reep, C., Benjamin, B., & Pollard, R. (1971). Sport and the Negative Binomial Distribution. pp. 188-195. In Ladany S.P., & Machol R.E. (Eds), *Optimal Strategies in Sport.* Amsterdam: North Holland.

Rigg, P., & Reilley, T. (1987). *A Fitness Profile And Anthropometric Analysis Of First And Second Class Rugby Union Players.* In Reilley, T., Lees, A., Davids, K., & Murphy, W.J. (Eds.). *Science and Football.* London: E. & F.N. Spon

Ripley, B. D. (1994). Neural Networks and Related Methods for Classification. *Journal of The Royal Statistical Society*, Series B, 56(3), pp. 409-456.

Roberts, H.V. (1992). Can TQM Improve Athletic Performance? Proceedings of the Section on Statistics in sports of the American Statistical Association.

Robertson, B., & Osborne, B. (1984). *Rugby Coaching the New Zealand Way*. Auckland: Hutchinson Group.

Romesburg, H.C., (1984). *Cluster Analysis for Researchers*. Belmont, California: Lifetime Learning Publications.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. *Psychological Review*. 65, pp. 386-408. Reprinted in Anderson & Rosenfeld (1988), pp. 92-114.

Rosenblatt, F. (1959). Two Theorems Of Statistical Separability In The Perceptron. Mechanization of Thought Process: Proceeding of a Symposium Held at the National Physical Laboratory, November 1958. London: HM Stationary Office, pp. 421-456.

Rosenblatt, F. (1962). *Principles of Neurodynamics*. New York: Spartan.

Rutherford, D. (2001). NZRFU Submission to the Gaming Review. Gaming Review Team. Wellington: NZRFU.

Sackrowitz, H. (2000). Refining the Point(s) - After-Touchdown Decision. *Chance*, 13(3), pp. 29-34.

Salmela, J.H. & Regnier, G. (1983). A Model For Sport Talent Identification. *Science Periodical On Research And Technology In Sport*. October, pp.1-8.

Sanger, T. D. (1994). Optimal Unsupervised Motor Learning for Dimensionality reduction of Nonlinear Control Systems. *IEEE Transactions on Neural Networks*, 5(6), pp. 965-973.

Sarle, W. S.(2001) *What does Unsupervised Learning Learn*. Available: ftp://ftp.sas.com/pub/neural/FAQ.html [2001, 27/04/01].

SAS (2000). *Enterprise Miner 4.0 User Interface Help*. SAS Institute.

Schwarzer, G., Vach, W., & Schumacher, M. (2000). On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Statistics in Medicine*. 19(4), pp. 541-561.

Scott, A., & Wild, C. (1991). Transformations and R^2. *The American Statistician*. 45(2), pp. 127-129.

Segall, D. O. (2001). General Ability Measurement: An Application of Multidimensional Item Response Theory. *Psychometrika*, 66(1), pp. 79-97.

Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons.

Shewhart, W.A. (1939). *Statistical Method From The Viewpoint Of Quality Control*. Washington D.C.: Graduate School of the Dept. of Agriculture.

Smith, D. J., & Teo, K.L. (1995). *Linear Algebra* (5th ed.). Wellington: New Zealand Mathematical Society.

Smith, P.J. (1993). *Into Statistics*. Melbourne: Nelson

Smith, W. (2001). *Suzuki Sportzone*. (presenters. Quinn, K. & MacBeth, J.). TV ONE.

South, M. C., Wetherill, G. B., & Tham, M. T. (1993). Hitch-Hikers Guide to Genetic Algorithms. *Journal of Applied Statistics*, 20(1), pp. 153-175.

Spearman, C. (1923). *The Nature of "Intelligence"*. London: MacMillan.

Spearman. C. (1904). "General Intelligence" Determined and Measured. *American Journal of Psychology*. 15, pp. 201-293.

Stamkopoulos, T., Diamantaras, K., Maglaveras, N., & Strintzis, M. (1998). ECG Analysis Using Nonlinear PCA Neural Networks for Ischemia Detection. *IEEE Transactions on Signal Processing*, 46(11), pp. 3058-3067.

Stefani, R. (1998). Predicting Outcomes. In J. M. Bennett (Ed.), *Statistics in Sport* pp. 249-276. New York: Arnold.

Stern, H. S. (1996). Neural Networks in Applied Statistics. *Technometrics*, 38(3), pp. 205-220.

Stern, H. S. (1998). American Football. In J. M. Bennett (Ed.), *Statistics in Sport* pp. 3-24. New York: Arnold.

Stern, H.S. (1997). A Statistician Reads the Sports Pages. *Chance* 10, pp. 19-23.

Stewart, J. J. (1997). *Rugby: Developments in the Field of Play*. Palmerston north: Dunmore Press.

Stewart, J.J. (1987). *Rugby: A Tactical Appreciation*. Auckland: Rugby Press.

Tagliaferri, R., Capuano, N., & Gargiulo, G. (1999). Automated Labeling for Unsupervised Neural Networks: A Hierarchical Approach. *IEEE Transactions on Neural Networks*, 10(1), pp. 199-203.

Taguichi, G., & Rajesh, J. (2000). New Trends in Multivariate Diagnosis. *Sankhya: The Indian Journal of Statistics*, 62, pp. 233-248.

Tambouratzis, G. (1993). Comparison of Supervised and Unsupervised Discriminator-Based Logic Neural Networks. *Electronics Letters*, 30(3), pp. 248-249.

Tesauro, G. (1990). Neurogammon Wins Computer Olympiad. *Neural Computation*. 1, pp. 321-323.

Thomas, V. (1970). *Science and Sport: The Measurement and Improvement of Performance*. Faber & Faber: London.

Thorburn, P. (2001). *Suzuki Sportzone*. (presenters. Quinn, K. & MacBeth, J.). TV ONE.

Thrun, S. (1996). *Explanation-Based Neural Network Learning: A Lifelong Learning Approach*. Boston: Kluwer Academic Publishers.

Thurstone, L.L. (1955). *The Differential Growth of Mental Abilities.* Psychometric Report Number 14. Chapel Hill: University of North Carolina.

Thurstone, L.L. (1955). *Primary Mental Abilities.* Psychometric Monographs Number 1. Chicago: University of Chicago Press.

Townend, M.S. (1996). Some Recent Applications of Advanced Mathematics in Sport. Conference Proceedings from 3rd Conference on Mathematics and Computers in Sport, Bond University, Queensland (September, 1996).

Tracey, N.D., Young, J.C., & Mason, R.L., (1992). Multivariate Control Charts for Individual Observations. *Journal of Quality Technology* 24, pp. 88-95.

Treadwell, P.J. (1987). Computer-Aided Match Analysis Of Selected Ball Games (Soccer And Rugby Union). Communications To The 1st World Congress Of Science And Football. pp. 282-287 In Reilly, T., Lees, A, Davids, K., & Murphy, W.J. (Eds.). *Science And Football.* E & Fn Spon: London.

Trippi, R.R., & Turban, E. (1993). *Neural Networks In Finance And Investing: Using Artificial Intelligence To Improve Real-World Performance.* Chicago, Illinois: Probus Publishing Company.

Tumilty, D., Burke., L. & Martin, D. (1999) Relationship Between The Assessment Of Individual Soccer Coaches On Players' Capabilities, And Between Assessment And Test Results. Communications To The 4th World Congress Of Science And Football. *Journal Of Sports Sciences.* 17, p. 817.

Unwin, A., & Unwin, D. (1998). Exploratory Spatial Analysis with Local Statistics. *The Statistician*, 47(3), pp. 415-421.

van Camp, D. (1992). Neurons for Computers. *Scientific American*, pp. 125-127.

von der Malsburg, C. (1973). Self-Organizing Of Orientation Sensitive Cells In The Striate Cortex. *Kybernetik.* 14, pp. 85-100.

Waller, N. G., Kaiser, H. A., Illian, J. B., & Manry, M. (1998). A Comparison of the Classification Capabilities of the 1-Dimensional Kohonen Neural Network with Two Partitioning and Three Hierarchical Cluster Analysis Algorithms. *Psychometrika*, 63(1), pp. 5-22.

Walpole, R.E. (1974). *Introduction to Statistics.* (2nd ed.). New York: Macmillan.

Walsh, P. (1984). *Winning Rugby.* Auckland: Novalit Books.

Wand, M. P., & Jones, M. C. (1995). *Kernel Smoothing.* London: Chapman & Hall.

Warner, B., & Misra, M. (1996). Understanding Neural Networks as Statistical Tools. *The American Statistician*, 50(4), pp. 284-293.

Weiten, W. (1995). *Psychology: Themes and Variations* (3rd ed.). Pacific Grove: Brooks/Cole.

Werbos, P. (1974). *Beyond Regression: New Tools For Prediction and Analysis in the Behavioral Sciences* (PhD Thesis). Cambridge, MA: Havard University Committee on Applied Mathematics.

Whineray, W. (1978). *The Highest Honour.* In Palenski, R., McMillan, N., & Chester., R. (2000). *Men in Black: 20th Century Commemorative Edition.* Hodder Moa Beckett: Auckland.

White, H. (1989). Some Asymptotic Results For Learning In Single Hidden Layer Feedforward Networks. *Journal Of The American Statistical Association.* 84, pp. 1008-1013.

White, H. (1992). *Artificial Neural Networks: Approximation And Learning Theory.* Oxford: Basil Blackwell

Widrow, B. & Hoff, M.E. Jnr. (1960). Adaptive Switching Circuits. IRE WESCON Convention Record. part 4, pp. 96-104. Reprinted in Anderson & Rosenfeld (1988), pp. 126-134.

Wilhelm, A., & Steck, R. (1998). Exploring Spatial Data by Using Interactive Graphics and Local Statistics. *The Statistician*, 47(3), pp. 423-430.

Williams, B., & Williams, D. (1998). Performance Indices for Multivariate Ice Hockey Statistics. In. Bennett, J. M (Ed.), *Statistics in Sport* pp. 154-156. New York: Arnold.

Williams, R.J. & Zipser, D. (1989). A Learning Algorithm For Continually Running Fully Recurrent Neural Networks. *Neural Computation.* 1, pp. 270-280.

Willshaw, D.J., Buneman, O.P. & Longuet-Higgins, H.C. (1969). Non-Holographic Associative Memory. *Nature.* 222, pp. 960-962.

Wise, B. M., Holt, B. R., Gallagher, N. B., & Lee, S. (1995). A Comparison of Neural Networks, Non-Linear Biased Regression and a Genetic Algorithm for Dynamic Model Identification. *Chemometrics and Intelligent Laboratory Systems*, 30, pp. 81-89.

Woodward, A. M., Jones, A., Zhang, X.-Z., Rowland, J., & Kell, D. B. (1996). Rapid and Non-Invasive Quantification of Metabolic Substances in Biological Cell Suspensions using Non-Linear Dielectric Spectroscopy with Multivariate Calibration and Artificial Neural Networks. Principles and Applications. *Biochemistry and Bioenergetics*, 40, pp. 99-132.

Wyatt, J. (1995). Nervous About Artificial Neural Networks. *The Lancet*, 346(8984), pp. 1175-1177.

Zhang, J. (1999). Developing Robust Non-Linear Models Through Bootstrap Aggregated Neural Networks. *Neurocomputing*. 25, pp. 93-113.

61.325 Study Guide (1997). *Statistical methods for quality improvement*. Massey University.

61.725 Study Guide (1998). *Statistical quality control*. Massey University.