Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Dealing with Sparsity in Genotype×Environment Analyses

A thesis presented in partial
fulfilment of the requirements
for the degree of
Doctor of Philosophy
in Statistics

at Massey University,
Palmerston North, New Zealand.

A. Jonathan R. Godfrey
2004

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

SUPERVISOR'S DECLARATION

This is to certify that the research carried out for the Doctoral thesis entitled "Dealing with Sparsity in Genotype-by-Environment Analyses" was done by Anthony Jonathan Royce Godfrey in the Institute of Information Sciences and Technology, Massey University, Palmerston North, New Zealand. The thesis material has not been used in part or in whole for any other qualification, and I confirm that the candidate has pursued the course of study in accordance with the requirements of the Massey University regulations.

Supervisor's Name: Associate Professor Chin-Diew Lai

Signature: C. D. Lai

Date: 2 - 8 - 04

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

CERTIFICATE OF REGULATORY COMPLIANCE

This is to certify that the research carried out in the Doctoral Thesis entitled **"Dealing with Sparsity in Genotype-by-Environment Analyses"** in the Institute of Information Sciences and Technology, Massey University, New Zealand:

(a) is the original work of the candidate, except as indicated by appropriate attribution in the text and/or in the acknowledgements;

(b) that the text, excluding appendices/annexes, does not exceed 100,000 words;

(c) all the ethical requirements applicable to this study have been complied with as required by Massey University, other organisations and/or committees which had particular association with this study, and relevant legislation.

Please insert Ethical Authorisation code(s) here: **N.A.**

Candidate's Name: Anthony Jonathan Royce Godfrey

Signature: **Date:**

Supervisor's Name: Chin-Diew Lai

Signature: C. D. Lai **Date:** 2/8/04

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

CANDIDATE'S DECLARATION

This is to certify that the research carried out for my Doctoral thesis entitled "*Dealing with Sparsity in Genotype-by-Environment Analyses*" in the Institute of Information Sciences and Technology, Massey University, Palmerston North, New Zealand, is my own work and that the thesis material has not been used in part or in whole for any other qualification.

Candidate's Name Anthony Jonathan Godfrey

Signature

Date 02/08/2004

Abstract

Researchers are frequently faced with the problem of analyzing incomplete and often unbalanced genotype-by-environment (G×E) matrices which arise as a trials programme progresses over seasons. The principal data for this investigation, arising from a ten year programme of onion trials, has less than 2,300 of the 49,200 combinations from the 400 genotypes and 123 environments. This 'sparsity' renders standard G×E methodology inapplicable. Analysis of this data to identify onion varieties that suit the shorter, hotter days of tropical and subtropical locations therefore presented a unique challenge.

Removal of some data to form a complete G×E matrix wastes information and is consequently undesirable. An incomplete G×E matrix can be analyzed using the additive main effects and multiplicative interaction (AMMI) model in conjunction with the EM algorithm but proved unsatisfactory in this instance.

Cluster analysis has been commonly used in G×E analyses, but current methods are inadequate when the data matrix is incomplete. If clustering is to be applied to incomplete data sets, one of two routes needs to be taken: either the clustering procedure must be modified to handle the missing data, or the missing entries must be imputed so that standard cluster analysis can be performed.

A new clustering method capable of handling incomplete data has been developed. 'Two-stage clustering', as it has been named, relies on a partitioning of squared Euclidean distance into two independent components, the G×E interaction and the genotype main effect. These components are used in the first and second stages of clustering respectively.

Two-stage clustering forms the basis for imputing missing values in a G×E matrix, so that a more complete data array is available for other G×E analyses. 'Two-stage imputation' estimates unobserved G×E yields using inter-genotype similarities to adjust observed yield data in the environment in which the yield is missing. This new imputation method is transferrable to any two-way data situation where all observations are measured on the same scale and the two factors are expected to have significant interaction. This simple, but effective, imputation method is shown to improve on an existing method that confounds the G×E interaction and the genotype main effect. Future development of two-stage imputation will use a parameterization of two-stage clustering in a multiple imputation process.

Varieties recommended for use in a certain environment would normally be chosen using results from similar environments. Differing cluster analysis approaches were applied, but led to inconsistent environment clusterings. A graphical summary tool, created to ease the difficulty in identifying the differences between pairs of clusterings, proved especially useful when the number of clusters and clustered observations were high. 'Cluster influence diagrams' were also used to investigate the effects the new imputation method had on the qualitative structure of the data.

A consequence of the principal data's sparsity was that imputed values were found to be dependent on the existence of observable inter-genotype relationships, rather than the strength of these observable relationships. As a result of this investigation, practical recommendations are provided for limiting the detrimental effects of sparsity. Applying these recommendations will enhance the future ability of two-stage imputation to identify those onion varieties that suit tropical and subtropical locations.

Acknowledgements

I wish to acknowledge the efforts of many people that have culminated in the submission of this work. First, and foremost, let me thank my supervisors. Professor Graham Wood gave me the opportunity to start a PhD. I will remain eternally grateful for this, the endless supply of advice, kind words, and the occasional kick in the pants that all students need to achieve the best they can. Thanks to Dr Mike Nichols and Dr Ganes Ganesalingam also for their contributions. Other Massey staff have also played a part; Greg Arnold for S-PLUS debugging and advice, Mark Bebbington and Geoff Jones for LaTeXcommands and debugging, and everyone that came to my seminars and provided feedback over the years.

This work would not have been considered if the real world problem had not existed. Dr Lesley Currah has contributed on numerous occasions with the knowledge surrounding the trials programme. Her patience must at times have worn thin, but it is hoped that she will relish the opportunity that now exists to send our findings to the many collaborators around the world. E-mail has served us well, and the time difference has allowed us each to send well considered questions and responses. Thank you Lesley for everything, including the opportunities for publications, your assistance with proof reading, and generally acting as another supervisor.

There are in fact very many people that need to be remembered for their contributions. I hope each of them receives some joy at the news of the submission of my thesis.

Special mention needs to be paid to those individuals that have assisted in the reading of many papers, document preparation, and all things visually dependent; Zoë Wood, Vince Pegg (for weekends and weeknights), Allister Campbell, Will Samuel, Judy Cann, Graeme Robinson, Michele Bisset, and others.

How can we forget those that ensured that I reached the start of the PhD, and then sustained me through the years. My parents and family have provided the breaks away from Palmerston North that revitalize the spirit; Mum for finishing her PhD first, so that she had time for proof reading. My many flatmates from 1998 to 2002 who put up with my occasional rantings and experimental cooking. The All Blacks and the Black Caps for their various interruptions. Radio Sport for commentaries and the slow paced progress of work over the summer months. The staff at the MUSA shop are yet to vanquish my appetite for apple and cinnamon muffins, let alone the raspberry twists.

Inanimate items serve frequently without thanks. Praise be to the Concise Oxford English Dictionary and various sources of LaTeXnotes available on the Internet. I have now learned more about the English language and how to present it than I expected at the outset. I thought my command of the language and sentence construction was better than the constant proofing showed me. Sorry to the purists, but I boldly went, and here is the fruit of my labour.

Contents

Abstract i

Acknowledgements iii

Contents v

List of Figures ix

List of Tables xiii

I Background to the Problem 1

1 Setting the scene 3

1.1 Introduction 3
1.2 Some key definitions 4
1.3 Issues relevant to G×E research 4
1.4 The path to a solution 7
1.5 Significance of this investigation 10
1.6 Summary 11

2 Fundamentals of G×E analysis 12

2.1 Introduction 12
2.2 Establishing the significance of G×E interaction 13
2.3 Joint regression 17
2.4 Variations on the joint regression model 22
2.5 Multiplicative models 26
2.6 Cluster analysis 33
2.7 Stability measures 39
2.8 Other models and methods 46
2.9 Summary 51

3 The principal data 52
3.1 Introduction 52
3.2 The trials programme 52
3.3 The problem of missing data and sparsity 57
3.4 Initial data analysis. 59
3.5 Initial modelling using regression analysis 72
3.6 Finding a suitable subset of data 77
3.7 Adapting current methodology to allow for incomplete data 80
3.8 Analysis with EM-AMMI 89
3.9 Summary 91

II Development of a Solution 93

4 Distance measures 95
4.1 Introduction 95
4.2 Main effect and interaction distance 96
4.3 A partition of Euclidean distance 100
4.4 Computation of main effect and interaction distances 103
4.5 The metric nature of main effect and interaction distances 104
4.6 Estimating unobserved distances 107
4.7 A survey of distance measures 109
4.8 Comparison of main effect and Euclidean distances 114
4.9 Summary 119

5 Two-stage clustering 121
5.1 Introduction 121
5.2 Description of two-stage clustering 124
5.3 An example of two-stage clustering 125
5.4 Application of two-stage clustering to the trials programme data 129
5.5 Expressing clustering in a parametric model framework 138
5.6 Summary 142

6 Two-stage imputation 143
6.1 Introduction to imputation of missing data 143
6.2 Imputing missing G×E data 147
6.3 An example of two-stage imputation 150
6.4 Comprehensive testing of two-stage imputation 153
6.5 Application of two-stage imputation to the trials programme data 164
6.6 Further ideas for two-stage imputation 171
6.7 Summary 172

7 Determining mega-environments 173
7.1 Introduction 173
7.2 Use of available yield data to cluster environments 175
7.3 Use of fully imputed yield data to cluster environments 184
7.4 Use of covariates to cluster environments 186
7.5 Summary 197

8 Comparing cluster analyses 207
8.1 The need to compare cluster analyses 207
8.2 Cluster influence diagrams 209
8.3 Gauging strength of the relationship between two cluster analyses 215
8.4 The consistency of available data 218
8.5 The consistency of imputed data 223
8.6 The ability of yield data to reflect covariate information 230
8.7 The effect of imputation on the G×E structure 235
8.8 The dependence of imputations on commonality of test environments 245
8.9 Summary 248

III The Solution: Results and Implications 251

9 Genotype selections for a new environment 253
9.1 Introduction 253
9.2 Starting with graphical approaches 255
9.3 Genotype selections using stability measures 262
9.4 Genotype selections using geographically similar environments 271
9.5 Genotype selections using mega-environments 273
9.6 Results using partial imputation 275
9.7 Summary 278

10 Advice for future trials programme designers 280
10.1 Introduction 280
10.2 Selection of environments 281
10.3 Covariate information 285
10.4 Experimental design for individual trials 287
10.5 Modelling of planting density 291
10.6 Selection of genotypes to include in each environment 292
10.7 Enhancing the connectedness of the G×E matrix 298
10.8 Summary 306

11 Conclusion **309**
11.1 Introduction 309
11.2 Background to the investigation 309
11.3 New theoretical developments 312
11.4 Implications and recommendations arising from the investigation 315
11.5 Summary 317

A Data from the Onion Trials Programme **318**

B References **330**

List of Figures

1.1 Flow chart depicting the solution path 8

2.1 Alternative representation of Finlay and Wilkinson (1963) plot of genotype stability and mean yield 21

3.1 Genotype means, standard deviations, and the number of times each was used plotted against one another 60
3.2 Environment means, standard deviations, and the number of times each was used plotted against one another 61
3.3 Normal probability plots for yield 63
3.4 Genotype means, standard deviations, and the number of times each was used plotted against one another (using square roots of yield) 64
3.5 Environment means, standard deviations, and the number of times each was used plotted against one another (using square roots of yield) 65
3.6 Diagnostic plots of environment altitudes and latitudes 66
3.7 Onion yields plotted against altitude and latitude 67
3.8 Square root of yield plotted against altitude and latitude 68
3.9 Environmental means and standard deviations of square roots of yields plotted against latitudes and altitudes 69
3.10 Yields of Red Synthetic HZ plotted against environment altitude and latitude 70
3.11 Relative yields of Red Synthetic HZ plotted against environment altitude and latitude 71
3.12 Range of growing periods plotted against median of growing periods for environments 72
3.13 Medians and ranges of growing periods plotted against latitude 73
3.14 Diagnostic plots for the best regression model for yield 76
3.15 Comparison of adjusted coefficients of variation for 87 genotypes 82
3.16 Histograms of adjusted superiority scores for Onion Data I and II 86
3.17 Plots of adjusted superiority measures versus genotype usage 88

4.1 Two pairs of genotypes showing the difference between shape and level similarity 98

4.2 Graphical representation of the partition of Euclidean distance for two genotypes in two environments . . . 102
4.3 Diagram illustrating shortest path concept . . . 108
4.4 Expected value of main effect and Euclidean distance . . . 117
4.5 Mean squared error of main effect and Euclidean distance . . . 118

5.1 An example of first stage clustering of complete data . . . 126
5.2 An example of first stage clustering of incomplete data . . . 127
5.3 An example of second stage clustering of incomplete data . . . 128
5.4 Effects of using within-environment standardization on Onion Data I . . . 130
5.5 Effects of using within-environment standardization on Onion Data II . . . 131
5.6 Effects of using within-environment standardization on the square roots of Onion Data I yields . . . 132
5.7 Effects of using within-environment standardization on the square roots of Onion Data II yields . . . 133
5.8 First stage clustering of genotypes from Onion Data I . . . 134
5.9 First stage clustering of genotypes from Onion Data II . . . 136
5.10 Two-stage clustering of a first stage cluster from Onion Data I . . . 139
5.11 Two-stage clustering of a first stage cluster from Onion Data II . . . 140

6.1 Yields of three genotypes plotted over six environments to illustrate imputation strategies . . . 149
6.2 Clustering of 58 genotypes using Euclidean distance of Ouyang *et al.* (1995) 152
6.3 Yields of three similar genotypes plotted against an environmental index to illustrate imputation results . . . 152
6.4 Boxplots of yields versus environments for Fox and Rathjen (1981) data . . . 155
6.5 Histograms of range standardized imputed yields for Onion Data I and II . 165
6.6 Within environment correlation of fully imputed Onion Data I and II plotted against environment usage . . . 167
6.7 Dendrogram of Onion Data I genotypes, clustered using interaction distance applied to two-stage imputed data . . . 167
6.8 Dendrogram of Onion Data II genotypes, clustered using interaction distance applied to two-stage imputed data . . . 168
6.9 Dendrogram of Onion Data I genotypes, clustered using Euclidean distance applied to incomplete data . . . 169
6.10 Dendrogram of Onion Data II genotypes, clustered using Euclidean distance applied to incomplete data . . . 169
6.11 Dendrogram of Onion Data I genotypes, clustered using Euclidean distance applied to nearest cluster imputed data . . . 170

6.12 Dendrogram of Onion Data II genotypes, clustered using Euclidean distance applied to nearest cluster imputed data 170

7.1 Histograms of within-genotype standardized yields. 176
7.2 Histograms of within-genotype standardized square roots of yields. 178
7.3 Dendrogram of all 123 environments of the Onion Trials Programme clustered using incomplete data 179
7.4 Dendrogram of 109 Onion Data I environments clustered using incomplete data 179
7.5 Dendrogram of the 98 Onion Data II environments clustered using incomplete data 181
7.6 Genotype standard deviations plotted against means for Onion Data I and II after two-stage imputation 184
7.7 Effects on within-environment standard deviations and means of the within-genotype standardization 185
7.8 Dendrogram of 109 Onion Data I environments based on two-stage imputed yields 186
7.9 Dendrogram of 98 Onion Data II environments based on two-stage imputed yields 189
7.10 Heat unit variables plotted against one another 193
7.11 Photoperiod variables plotted against one another 194
7.12 Normal probability plots and boxplots of transformed proxy variables . . . 198
7.13 Dendrogram of 101 environments clustered using proxy variables 199
7.14 Normal probability plots and boxplots of transformed covariates 201
7.15 Normal probability plots and boxplots of transformed covariates 202
7.16 Dendrogram of 89 environments clustered using proxy variables 203
7.17 Dendrogram of 79 environments clustered using constructed covariates . . . 205

8.1 Cluster influence diagram for dendrograms of Figures 7.16 and 7.17 210
8.2 Cluster influence diagram for dendrograms in Figures 7.13 and 7.16 212
8.3 Cluster influence diagram for dendrograms in Figures 7.13 and 7.17 213
8.4 Cluster influence diagram for dendrograms in Figures 7.3 and 7.4 219
8.5 Cluster influence diagram for dendrograms in Figures 7.4 and 7.5 220
8.6 Cluster influence diagram for dendrograms in Figures 7.3 and 7.5 221
8.7 Cluster influence diagram for dendrograms in Figures 7.4 and 7.8 225
8.8 Cluster influence diagram for dendrograms in Figures 7.5 and 7.9 226
8.9 Cluster influence diagram for dendrograms in Figures 7.8 and 7.9 228
8.10 Cluster influence diagram for dendrograms in Figures 6.7 and 6.8 229
8.11 Cluster influence diagram for dendrograms in Figures 7.3 and 7.13 231
8.12 Cluster influence diagram for dendrograms in Figures 7.4 and 7.16 233

8.13 Cluster influence diagram for dendrograms in Figures 7.5 and 7.17 234
8.14 Cluster influence diagram for dendrograms in Figures 7.8 and 7.16 236
8.15 Cluster influence diagram for dendrograms in Figures 7.9 and 7.17 237
8.16 Effects of nearest cluster imputation on clustering of genotypes in Onion Data I 239
8.17 Effects of nearest cluster imputation on clustering of genotypes in Onion Data II 240
8.18 Effects of two-stage imputation on clustering of genotypes in Onion Data I 242
8.19 Effects of two-stage imputation on clustering of genotypes in Onion Data II 243

9.1 Correlation of latitude and expected yield plotted against mean yield 256
9.2 Biplot of first and second interaction axes for Onion Data I 257
9.3 Biplot of first and second interaction axes for Onion Data II 258
9.4 Scree plots for principal component contributions to Onion Data I and II . 260
9.5 Histograms of Wricke's ecovalences for Onion Data I and II 266
9.6 Histograms of adjusted Lin and Binns superiority scores for Onion Data I and II 270
9.7 Average relative performances of genotypes over geographically similar environments 273
9.8 Average relative performances of genotypes over environments within a mega-environment 275

10.1 Inter-genotype connectedness of Onion Data II 300
10.2 Inter-environment connectedness of Onion Data II 301
10.3 Flow chart showing the suggested method for selecting genotypes for a new trial 305

List of Tables

2.1 Summary of stability statistics 40
2.2 A second summary of stability statistics 41

3.1 Models used to explain yield performance 74
3.2 Summary of various regression models fitted to the sparse Onion Trials Programme data 75
3.3 ANOVA for the best regression model for yield 76
3.4 Partial ANOVA specific to the most commonly tested genotype of the Onion Trials Programme 77
3.5 Effects of data deletion on principal data. 78
3.6 Genotypes with a high adjusted coefficient of variation 81
3.7 Genotypes with high contribution to G×E interaction 84
3.8 Genotypes selected using the adjusted superiority score 85
3.9 Six onion varieties that were successful in the single Trial where they were tested 87
3.10 Correlations of fitted values from a series of EM-AMMI models 89
3.11 Number of iterations taken to fit EM-AMMI models, and correlation of results between data sets 89
3.12 Unrealistic yields found using various EM-AMMI models 90
3.13 Effects of using an alternative set of starting values for EM-AMMI models . 90

5.1 Fifteen G×E combinations randomly deleted to create an incomplete set of data. 127
5.2 Comparison of cluster memberships for complete and incomplete data . . . 127
5.3 Cluster memberships for the first stage clustering of Onion Data I genotypes 135
5.4 Cluster memberships for the first stage clustering of Onion Data II genotypes 137

6.1 Imputed values found for fifteen deleted observations 151
6.2 Summary details of the data sets used in simulation testing 154
6.3 Comparison of all pairs of imputation methods (raw data). 157
6.3 *continued* 158

6.4 Comparison of all pairs of imputation methods (within-environment standardized data). 159
6.4 *continued* 160
6.5 Correlations of imputed values and deleted values (raw data) 162
6.6 Correlations of imputed values and deleted values (within-environment stan standardized data) 163
6.7 Details of Onion Data I and II imputed values trimmed during imputation . 165

7.1 Cluster memberships for the dendrogram presented in Figure 7.3 180
7.1 *continued* 181
7.2 Cluster memberships for the dendrogram presented in Figure 7.4 182
7.3 Cluster memberships for the dendrogram presented in Figure 7.5 183
7.4 Cluster memberships for the dendrogram presented in Figure 7.8 187
7.5 Cluster memberships for the dendrogram presented in Figure 7.9 188
7.6 Correlations of proxy variables for environments 192
7.7 Summary of regression models for proxy variables using latitude and altitude as explanatory variables 195
7.8 Summary of best subsets regression of latitude and altitude on proxy variables196
7.9 Cluster memberships for the dendrogram presented in Figure 7.13 200
7.10 Cluster memberships for the dendrogram presented in Figure 7.16 204
7.11 Cluster memberships for the dendrogram presented in Figure 7.17 206

8.1 Numericl summary statistics for Figures 8.1, 8.2, and 8.3 217
8.2 Numerical summary statistics for Figures 8.4, 8.5, and 8.6 222
8.3 Numericl summary statistics for Figures 8.7 and 8.8 224
8.4 Numerical summary statistics for Figures 8.9 and 8.10 227
8.5 Numerical summary statistics for Figures 8.11 to 8.15 232
8.6 Summary statistics that quantify effects of imputation methods on clustering of genotypes 241
8.7 Number of genotype clusters found using sparse and imputed data 244
8.8 Complete table of numerical summary statistics to support all cluster influence diagrams of Chapter 8 246
8.9 λ and U coefficients for clustering of genotypes based on imputed data and commonality of test environments 247

9.1 Details of twelve environments with covariate data, but not used in imputations 254
9.2 Genotypes with high specific adaptation to low latitude environments . . . 256
9.3 Genotype codes used in the biplots of Figures 9.2 and 9.3 259

9.4 Partial ANOVA for various AMMI models applied to imputed yields of Onion Data I 261
9.5 Partial ANOVA for various AMMI models applied to imputed yields of Onion Data II 262
9.6 Genotype selections chosen using the adjusted coefficient of variation 264
9.7 Genotype selections chosen using Wricke's ecovalence 265
9.8 Genotype selections chosen using adjusted superiority scores 267
9.9 Check variety selections for Onion Data I using two nonparametric stability measures 268
9.10 Check variety selections for Onion Data II using two nonparametric stability measures 268
9.11 Check variety selections for Onion Data I using two more nonparametric stability measures 269
9.12 Check variety selections for Onion Data II using two more nonparametric stability measures 269
9.13 Correlations of wide adaptation stability measures 271
9.14 Genotypes selected on their success in geographically similar environments . 272
9.15 Genotypes selected on their success in environments within the same mega-environment 274
9.16 Genotypes not given estimated performances in the new trial when using partial imputation 276
9.17 Genotype selections for the new Yemeni trial, based on partially imputed Onion Data I and II G×E matrices 277

10.1 Covariate genotype information recommended for collection 285
10.2 Covariate environment information recommended for collection 286
10.3 Additional covariate information recommended for collection during new trials in the Onion Trials Programme 286
10.4 Genotype pairs whose linkage would be enhanced by the suggested new trial 302
10.5 A portion of the concurrence matrix for Onion Data I 304

A.1 Names of varieties used in core data sets 319
A.1 *continued* 320
A.2 Genotypes not included in Onion Data I or II 321
A.2 *continued* 322
A.3 Codes for trials included in Onion Data I and II 323
A.4 Summary information for environments 324
A.4 *continued* 325
A.4 *continued* 326
A.5 Artificial covariates used in environmental clustering 327

A.5 *continued* 328
A.5 *continued* 329