Analysis of spectral response patterns of Kiwifruit orchards using satellite imagery to predict orchard characteristics of commercial value before harvest

A thesis presented in fulfilment of the requirements for the degree of

PhD Prod Tech

in the

School of Engineering and Advanced Technology , Massey University

Linda Yvonne Mills

2015

ABSTRACT

Several characteristics of kiwifruit determine its value to the kiwifruit marketing company, Zespri Ltd, and to the grower. The foremost of these is the dry matter content. Much effort is expended in predicting the final dry matter content of the fruit as early in the season as possible so that the optimal dry matter content can be achieved. Dry matter content is currently measured through a destructive 90-fruit protocol that may be repeated several times in a season on each maturity block.

Remote sensing data available from modern satellites can provide four-colour (red, green, blue and near-infrared) data with resolution down to 1-2m, less than the size of one kiwifruit vine. Many indices can be created from these and correlated to the characteristics of plants with indifferent results.

This thesis presents the development of an index wherein the four colours are used to create a three-dimensional unit colour vector that is largely independent of light level. This transform was used to allow the direct visualisation of data from a number of satellite images of the Te Puke kiwifruit growing area in New Zealand over five years, for which dry matter content values were available from the 90-fruit protocol.

An attenuation model was chosen to correct the top-of-atmosphere light intensities recorded by the satellite cameras to those at ground level. The method of Hall *et al*., (1991) was found to reduce the variation of fiduciary pixels by the largest amount and was used.

The visualisation revealed that there was an axis along which dry matter was ordered by magnitude. A regression line of best fit was applied to this data producing an R^2 value of 0.51 with a standard mean-square error of 0.76. This is significantly lower than the average mean-square error of 1.05 for the 90-fruit protocol.

Comparison of the predictive power of other indices, based on one image, showed a range of R^2 values of 0.008 to 0.49. The method developed in this thesis produced an R^2 of 0.70 for the same data.

CERTIFICATE OF AUTHORSHIP

I, Linda Mills, hereby declare that this submission is my own work and that to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made.

………………………………
Linda Mills

ACKNOWLEDGEMENTS

Supervisors:
Dr Rory Flemmer, School of Engineering and Advanced Technology, Massey University
Dr Huub Bakker, School of Engineering and Advanced Technology, Massey University

Funding and Data support:
Zespri International Limited, Mount Maunganui, New Zealand
AgFirst, Katikati , New Zealand
Trevelyans Pack and Cool, Te Puke, New Zealand
DMS Progrowers Ltd, Tauranga, New Zealand
Andrew Jonkers, Brisbane, Australia

Editing and Sanity support:
My daughters-in-law Carly and Ruth and my husband Doug.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND DEFINITIONS

Actinidia chinensis	*Actinidia chinensis* is a fruiting vine native to China. The first commercialised cultivar in New Zealand was *Hort16A,* licensed by Zespri International Limited.) *Hort16A* is also referred to as Gold in this study.
Actinidia deliciosa	*Actinidia deliciosa* is a fruiting vine native to China. The predominant cultivar grown in New Zealand for export is *Hayward* which was commercialised in the 1940's. *Hayward* is also referred to as Green in this study.
AgFirst BOP	AgFirst Bay of Plenty is an independent company based in Katikati, New Zealand that provides services to the New Zealand Kiwifruit Industry. www.agfirstbop.co.nz
Brix	Brix is the soluble sugar content (SSC) of a fruit as measured by a refractometer. A refractometer uses light refraction to measure different sugar concentrations providing a value with the unit of measure of degrees brix.
DM	Dry Matter is what is left of a slice of fruit after it has been dehydrated, weighed in grams.
DN	A Digital Number represents the intensity of the signal received by the satellite sensor as reflected or emitted by a given area of the earth's surface.
ISO Week	International Organisation for Standardisation leap week calendar system that is part of the ISO 8601 date and time standard with the week with the year's first Thursday in it being ISO Week 1.
MDM	Measured Dry Matter

MSE	Mean Squared Error, measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated
%DM	Percentage Dry Matter, the dry weight of a material, in this case kiwifruit, expressed as a percentage of the fresh weight.
PDM	Predicted Dry Matter
Psa	The bacterial kiwifruit vine disease Pseudomonas syringae pv actinidiae.
R	Correlation coefficient
R^2	Coefficients of determination
RMSE	Root Mean Square Error is the square root of MSE or the standard deviation
SE	Standard Error
TE	Tray Equivalent, unit of measure of kiwifruit quantity being the number of pieces of fruit of a certain size that fit into a standard tray based on weight
TZG	Taste Zespri Grade
Yield	TE per hectare
Zespri	Zespri International Limited.

CHAPTER 1 INTRODUCTION

1.1 Introduction

The purpose of this study is to add value to the New Zealand kiwifruit industry by offering a new technique that informs growers of the potential fruit quality their vines are producing. This technique provides the information sufficiently early in the season so that the growers can make orchard management decisions to improve the value of their crops. The New Zealand kiwifruit industry formed the basis of this analysis because the industry is mature and fairly sophisticated with respect to the different fruit attributes by which the growers are paid for their crops. It is therefore necessary to capture the economic imperatives – which is possible in a specific locale.

It is expected that growers will use the results of this work because there are clearly established links between vine management and fruit quality. This technique provides the information sufficiently early in the season so that the growers can make orchard management decisions to improve the value of their crops. The growers are provided with a whole orchard view and definition of areas of homogeneity within the orchard. Such knowledge is important to growers as their orchard income is dependent on the quality of the fruit produced. In addition, their orchard costs are reduced when vine management actions do not have to be applied across all areas of the orchard.

This chapter outlines:

- a summary of the New Zealand Kiwifruit Industry
- the importance of fruit quality
- how fruit quality contributes to grower income
- the links between vine management and fruit quality
- the remote sensing basis of the new technique
- the working hypothesis that led to the development of the technique

1.2 The New Zealand Kiwifruit Industry

The New Zealand Kiwifruit Industry is a billion dollar export industry. Zespri, as the single-desk marketer of kiwifruit for New Zealand is vitally concerned to maintain a regime of continuous product improvement. All of the kiwifruit exported by Zespri is eaten fresh and consequently food safety and fruit quality are of great importance. Fruit quality encompasses both the external and internal attributes of the fruit.

The important crop characteristics that contribute to the value of kiwifruit are:

- visual appearance of the fruit
- taste of the fruit (approximated by the measurement of dry matter)
- size of the fruit
- storage of the fruit
- quantity of fruit per hectare (yield)

Each of these five crop characteristics carries differentiated payments and growers endeavour to produce the right combination to maximise their returns. The visual appearance of the kiwifruit determines whether it is graded against Zespri standards into Class 1 category (first grade) or Class 2 category (second grade) at the time of packing. Within Class 1, the remaining four characteristics compete with each other for share of return (refer Figure 1).

In addition, there are minimum standards set by Zespri for both the taste and the size of the kiwifruit. Fruit must achieve this minimum before it is acceptable to be purchased by Zespri as standard supply. During storage, under cool-storage conditions, fruit deteriorates over time and again there are standards that the fruit must still achieve before shipping for export, the point at which Zespri takes ownership of the fruit.

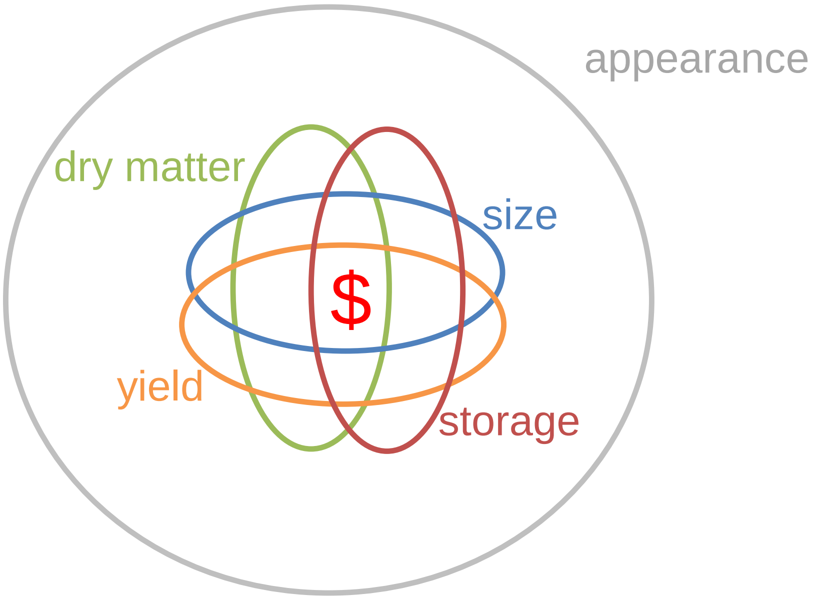


Figure 1: Four competing crop characteristics with differentiated financial returns

These four competing crop characteristics are discussed in depth in this chapter as they are the major influencers of financial returns within the New Zealand kiwifruit industry.

Orchard management techniques can influence the outcome of these crop characteristics, despite the environmental conditions of each season. Zespri uses commercial means to signal the quality of kiwifruit required in order to satisfy the market. In turn, increased market satisfaction results in increased revenue for growers.

The current methods for sampling the orchards to predict these characteristics are manual and are labour intensive. They also require destruction of the sampled fruit. From these sample results, it is attempted to extrapolate information across broader orchard areas, but these are not necessarily homogenous in nature.

There are approximately 12,000 hectares of kiwifruit being grown in New Zealand. 79% of these 12,000 hectares are produced in the Bay of Plenty with 40% of the national production specifically centred around the town of Te Puke. This area was chosen as the location for this research because of the high density production. During the period of study, 2005 – 2009, there were two kiwifruit varieties grown in any significant volume in the region – a cultivar of *Actinidia deliciosa* commonly known as Green or *Hayward* and a cultivar of *Actinidia chinensis* commonly known as Gold or *Hort16A*. The area covered in the May 2005 image is shown in Figure 2.

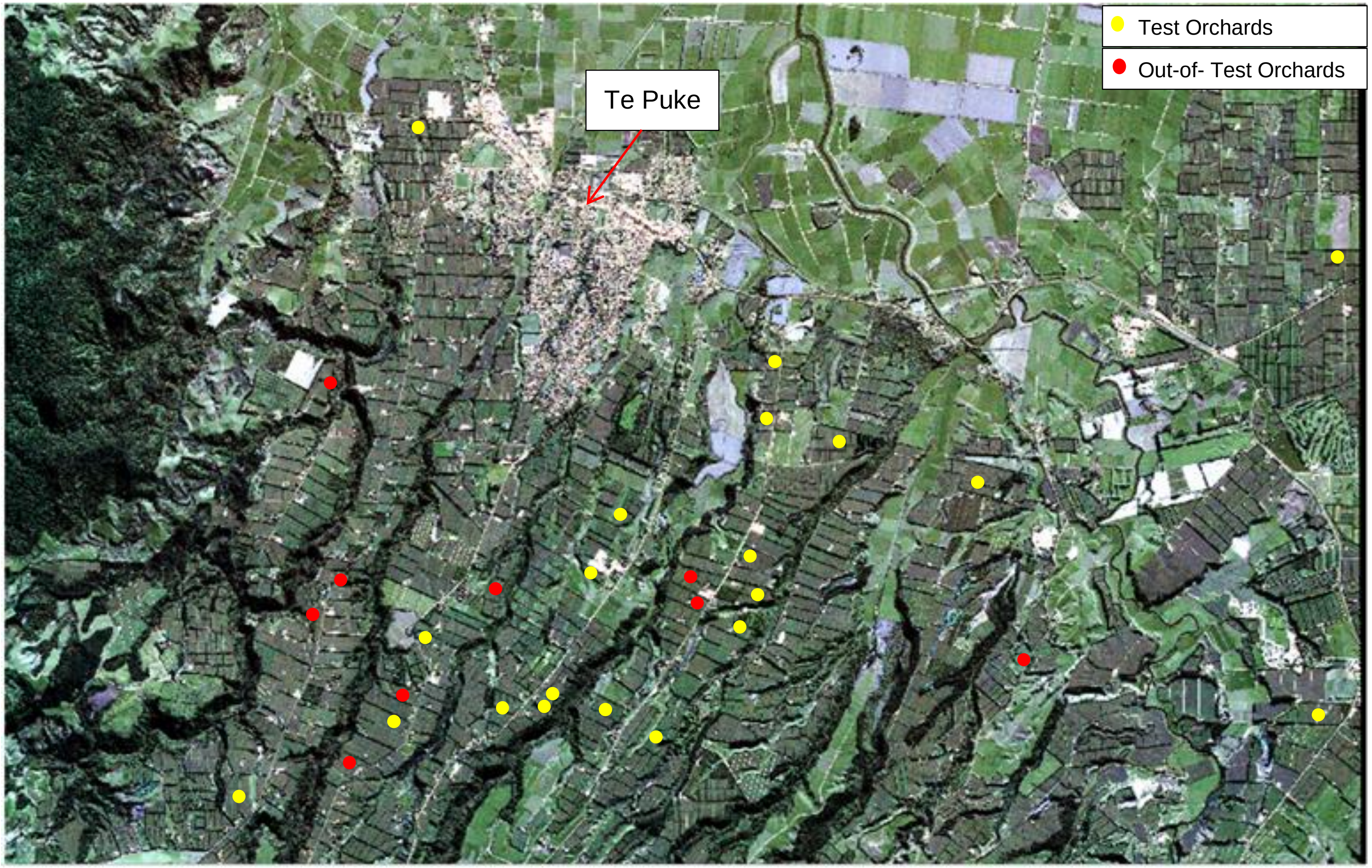


Figure 2: Area of study, Te Puke, Bay of Plenty, New Zealand, May 2005

The opportunity to develop methods for remote monitoring of on-orchard attributes was considered useful because large areas could be assessed in the same time frame, under the same conditions and would not require destructive testing. Remote monitoring has the potential:

- to predict fruit quality resulting from vine management and climate
- to identify areas within the orchards that would benefit from targeted orchard management
- to identify *sui generic* areas of the orchard for harvest segregation
- to identify the potential onset of maturity for harvest planning

There is extensive literature on remote monitoring of vegetation and crops, which is discussed in depth in chapter two. All of the methodologies are based on the correlation of ratios between the energies for different spectral regions and desired crop characteristics and there are more than twenty such correlations defined to date for different crops and crop characteristics. Clearly, it is desirable to provide a direct phenomenological predictor rather than a derived correlative predictor and it is the purpose of this thesis to reach that point.

1.3 Kiwifruit Orchards

Kiwifruit orchards are identified by a Kiwifruit Property Identification Number. This is land based and has a one to one relationship with a Valuation Number, as defined by Quotable Value New Zealand.

Kiwifruit orchards have a hierarchy of management units illustrated in Figure 3:

- a vine, the canopy area of each vine is approximately 30-35 m^2
- a block, which is a group of vines
- a Maturity Area, which is a group of blocks

Figure 3: Kiwifruit orchard, Orchard #2, Te Puke, from May 2005 Satellite image

These management units are defined and detailed on orchard maps (refer Figure 4), many of which have GPS co-ordinates. A Maturity Area is an identifiable area or group of areas in an orchard that contains fruit of a similar quality and maturity. The goal of the Maturity Area selection is to identify areas within an orchard, producing fruit of a uniform maturity and dry matter content that can be harvested and packed as discrete units.

Block	Males	Females	Rows	Bays	Area
1A	30	182	9	212	0.54 ha
1B	24	189	9	213	0.52 ha
2A	21	167	9	188	0.47 ha
2B	22	169	9	191	0.49 ha
3A	29	108	9	137	0.33 ha
3B	39.5	135	9	174.5	0.42 ha
4	54	219	9	273	0.65 ha
5	42	161	9	203	0.47 ha
6	36	268	20	304	0.76 ha

Figure 4: Orchard map for kiwifruit orchard, Orchard #2, Te Puke

Maturity Areas can be redefined each year and are registered with AgFirst BOP, the Zespri contracted independent laboratory. This registration is required because the kiwifruit within a Maturity Area must meet certain maturity criteria before they are cleared for harvest and AgFirst BOP is responsible for collecting and assessing samples from all of the Maturity Areas. All Maturity Areas must have a dry matter measurement recorded (by AgFirst, BOP) which dictates what payments are made to the growers.

The maturity criteria define the state of physiological development which must be reached (as measured by a 90-fruit random sample) before fruit from a maturity area may be removed and yet is able to continue to develop (ripen) until it is suitable for consumption. For each cultivar the tests for estimating maturity are different, but may include brix (level of soluble sugar in the fruit), dry matter, various measures of flesh colour and the percentage of black seeds in the fruit.

1.3.1 Crop Characteristics and Revenue

The net market returns for various kiwifruit products are separated into revenue pools depending on the variety, class (grade standard based on appearance) and growing method (organic or conventional). Within each pool the differentiated payments are calculated and paid from the pool first and then the remaining net income from sales (after costs) is paid out on a per Tray Equivalent basis across all trays in the pool (refer to 1.3.1.3 for an understanding of Tray Equivalent).

Simply put, growers with crops that are above average in one or more of the commercially differentiated characteristics will earn more than those growers with below average crops. This payment mechanism establishes a competitive commercial base from which to encourage growers to continuously improve the quality of their crops.

1.3.1.1 Appearance

Over the years it has been recognised that the external appearance of fruit has an important appeal to consumers when purchasing a premium product. Zespri has established grade standards for these appearance factors against which the fruit is graded during packing.

The main appearance factors are:

- shape
- consistent skin colour

- consistent size in the boxes
- lack of blemish or marks caused by fruit rubbing against each other during growth
- no punctures or other physical damage
- clean skin surface (no dirt, bird lime etc.)
- no pests on or pest damage to the fruit

The grade of the fruit produced is the component within a variety / growing method (conventional or organic) combination that determines the revenue pool from which the fruit will share net income. This segregation into different pools occurs before any other differentiated payment is made.

1.3.1.2 Taste

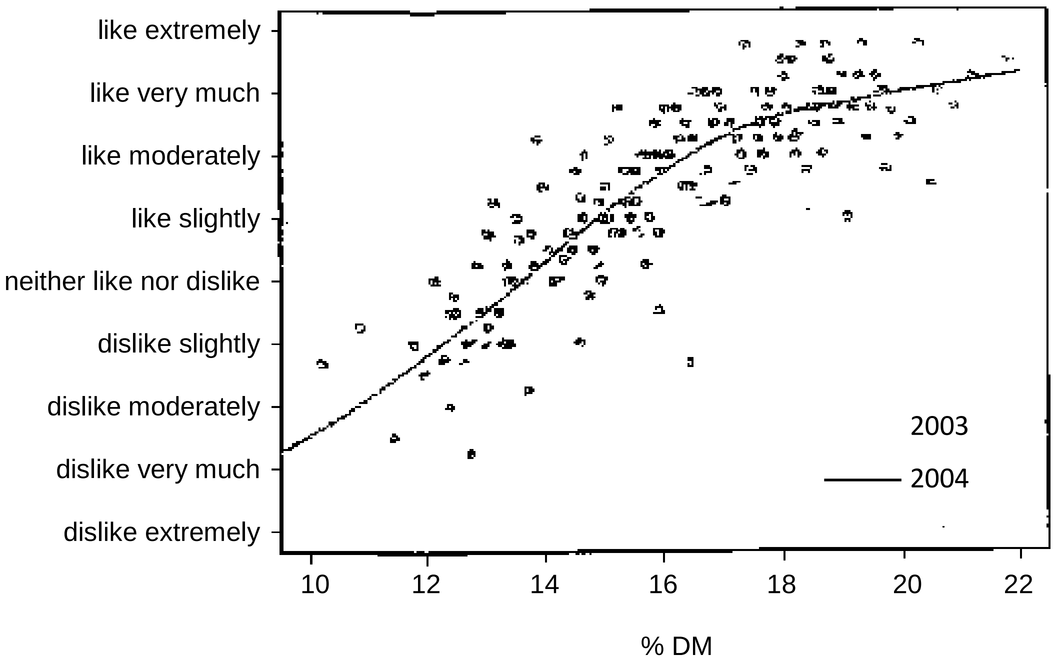
The taste of the kiwifruit is crucial to consumer satisfaction and therefore a significant attribute of differentiation for Zespri and growers. Several consumer studies have consistently shown the positive influence of taste on consumer purchase intent. The key findings from a joint research project in 2004, involving Zespri Innovation, HortResearch, Auckland University and TNS Global market research company showed that taste (as estimated through the measurement of Dry Matter [DM]) was a dominant influence on purchase intent when compared to the price of the fruit and the size of the fruit (refer Table 1).

Table 1: Market survey results of purchase decisions for kiwifruit

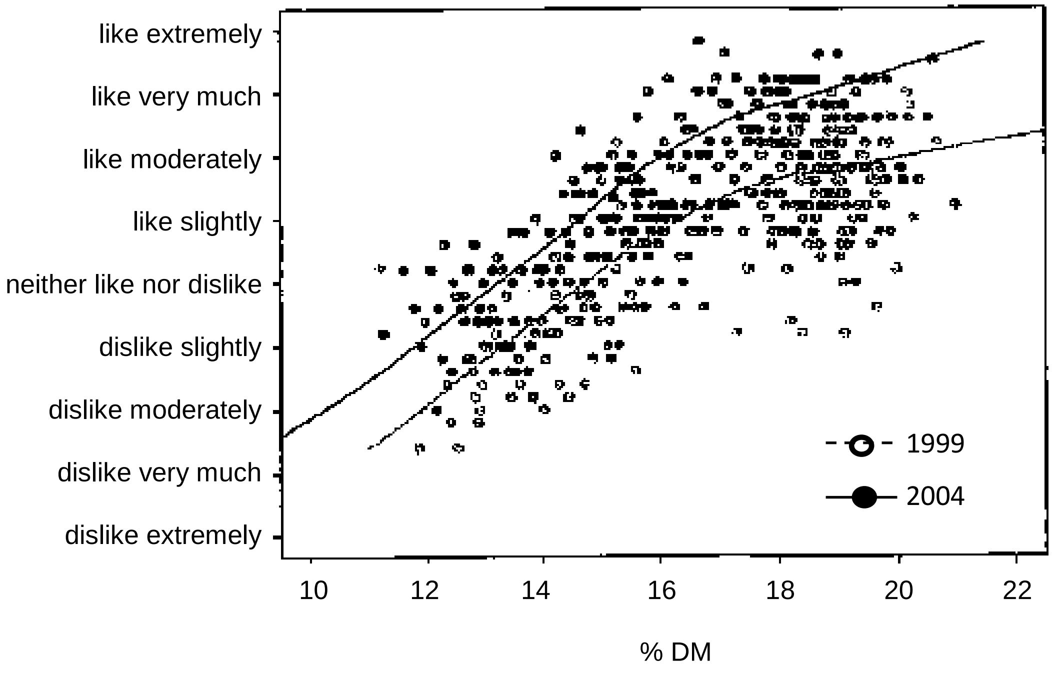
	Taste (DM)	**Price**	**Size**
Gold Kiwifruit	58%	37%	4%
Green Kiwifruit	64%	31%	5%

Further consumer research by Zespri has shown that taste is a very important internal attribute (Burdon, *et al*., 2004; Harker, *et al*., 2009), especially in key high value markets such as Japan. Higher dry matter levels of kiwifruit (as a measure of better taste) were also shown to correlate strongly with the increased likelihood of repurchase of kiwifruit by consumers as shown in Figure 5 from the 2003 and 2004 studies.

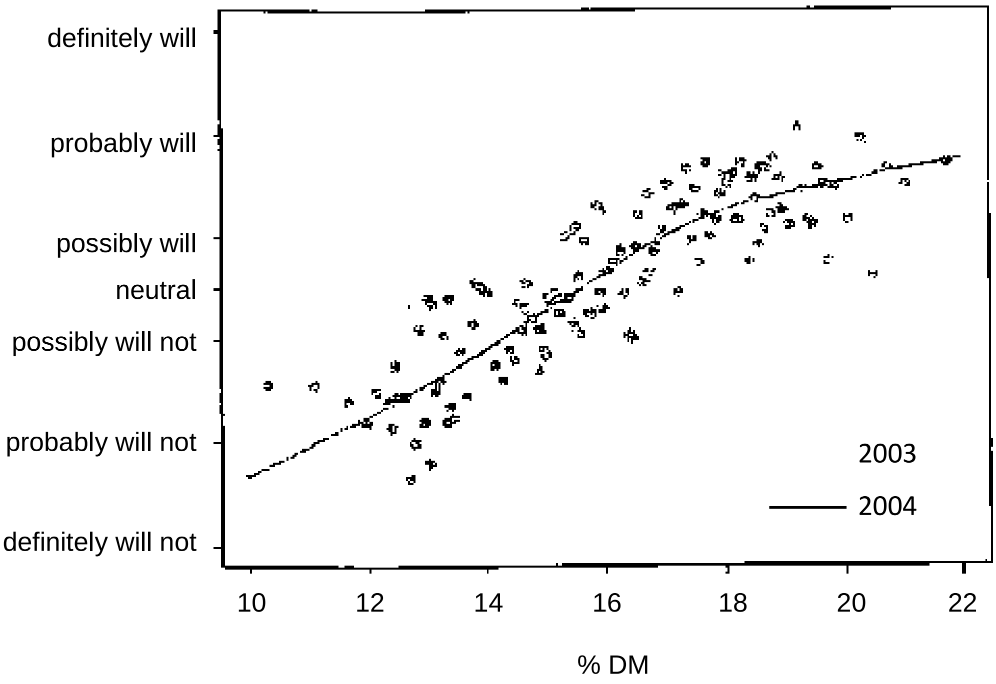
(A) Gold Liking Data



(B) Green Liking Data



(C) Gold Purchase Data



(D) Green Purchase Data

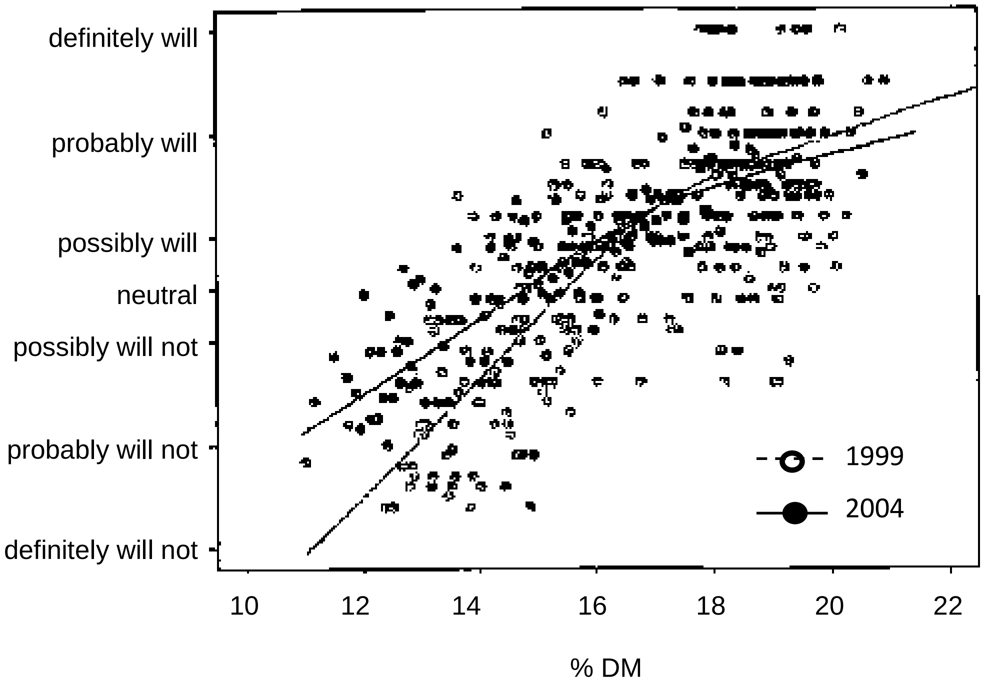


Figure 5 (A)-(D): Consumer preference and repurchase intent for kiwifruit presented at different dry matter levels

Research has shown that the dry matter of the kiwifruit can be used as an indicator of the taste of the kiwifruit when it is ripened and ready to eat. The dry matter content of fruit is related to the total carbohydrate concentration in the fruit which is comprised of both starch and sugar concentrations (Hopkirk, 1991). The dry matter of the fruit accumulates over time as the fruit grows and matures. During maturation, several changes occur.

Initially there is an increase in fresh weight consisting of water, acids and dry matter content, which is mainly starch at that time.

As the fruit matures, starch is converted to sugar. This conversion results in a decrease in starch concentrations, increase in soluble sugar (and total sugar concentrations) and a decrease in fruit firmness as the fruit approaches full maturity. The level of fruit maturity at time of consumption contributes to the flavor and taste of kiwifruit (Beever *et al.*, 1990 . The conversion to sugar will continue after the fruit has been harvested and is in storage before consumption. The dry matter content of unripe kiwifruit is therefore considered to be a predictor of the final eating quality of the kiwifruit when ripe.

In recognition of the importance of taste, New Zealand kiwifruit growers are paid a premium for kiwifruit that have a higher percentage of dry matter. These premiums are published annually by Zespri, based on a specifically devised formula. To calculate these premiums, a Maturity Area has a randomly selected 90 fruit sample collected and measured for dry matter content. The percentage of dry matter determined for the sample is converted to a Taste Zespri Grade (TZG) value using a proprietary Zespri formula.

A Maximum Taste Payment (MTP) is defined as a percentage of the Total Fruit Payment that is being assigned for the characteristic of taste. Green MTP is 40% of the Total Fruit Payment and Gold MTP is 70% of the Total Fruit Payment in a given year. This MTP is multiplied by the TZG of the Maturity Area to calculate the value per Tray Equivalent (TE) to be paid for fruit harvested and exported from the Maturity Area. Table 2 details the TZG levels from 2005 to 2009. As an example in 2009 the Total Fruit Payment was $5.30 and at 40% the MTP for Green was $2.12 therefore based on the average Green TZG of 0.52 the average Green taste payment would be $1.10.

Table 2: Historical Taste Zespri Grades (TZG)

	2005	2006	2007	2008	2009
Green	0.37	0.43	0.50	0.52	0.52
Gold	0.49	0.50	0.53	0.54	0.58

On average, approximately 20% of the total Green revenue and 29% of the total Gold revenue is paid to growers based solely on taste characteristics. This differentiated value for taste is the driver behind this study.

1.3.1.3 Size

New Zealand kiwifruit can be grown to a larger size without use of chemicals, purely by orchard management techniques, climatic conditions and soil type. This capability is a point of differentiation that provides a competitive advantage to New Zealand growers as larger sized kiwifruit can command higher premiums in market. These premiums are recognised by differentiated size payments to growers.

Size is determined by the weight of individual pieces of kiwifruit and the New Zealand kiwifruit industry has a standard unit of measure called a Tray Equivalent (TE). The kiwifruit are graded in the pack-house by weight and any kiwifruit within a band of weights is designated as a specific size. The number of pieces of kiwifruit within that band that can fit into a standard tray determines the size designation and is known as a TE for that size. Larger fruit is worth more per TE than smaller fruit.

1.3.1.4 Storage

Kiwifruit is harvested in late-March to mid-June in New Zealand and shipped to markets from late-March to early-November. Between harvest and shipping the kiwifruit are stored in cool-stores with some deterioration in quality and therefore a resultant loss of fruit over time. The kiwifruit is checked and if necessary repacked prior to shipping and it is during this repack that unsatisfactory fruit are removed.

The loss of this kiwifruit is only compensated at an average industry level which means that fruit that stores above average is worth an additional premium to the growers and fruit that stores below average will not be fully compensated (a redistribution of funds between growers). These storage incentives, gained or lost, are a significant area of competition for the post-harvest companies (those that pack and store kiwifruit prior to export) within the industry.

1.3.1.5 Yield

Yield is the quantity of kiwifruit per hectare and is measured as the gross number of Class 1 (premium export grade) TE per hectare packed. All trays that form part of a revenue pool, share the residual net income on a per TE basis after all other differentiated payments have been paid. Per hectare income less per hectare management costs, defines a grower’s net profitability and yield is an important variable contributor to these profits.

High yield production methods can compromise taste, size and storage characteristics. Many growers aim to achieve combinations of high yield, in addition to being above average for one or more of the other characteristics that earn differentiated payments thus improving their total annual per hectare returns.

1.3.2 Kiwifruit Canopy

Kiwifruit is a deciduous vine that is grown with a horizontal canopy over pergola styled structures; the vines left to their own devices grow rampantly. Pruning to manage canopy vigour is carried out in winter and summer in New Zealand. Mowat and Maquire ((Mowat & Maguire, 2006)) found that light open canopies like that pictured in Figure 6 (a) are more likely to produce higher dry matter fruit than dark closed canopies such as those pictured in Figure 6 (b).

Figure 6 (a)-(b): Kiwifruit canopy examples; (a) light, open canopy, (b) dark, closed kiwifruit canopy

Differentiated taste payments were first introduced in the 2004 kiwifruit season. One of the biggest concerns growers faced at this time was how to manage the canopies in order to improve the kiwifruit taste characteristic and thus maximise their payment. Research undertaken and compiled by Zespri was shared with New Zealand kiwifruit growers. This research showed that canopy management can have a positive influence on dry matter production, despite factors such as climate, soil type and vine age generally being outside of the growers control. These findings are summarised in Table 3 and show the direct impact of orchard management on dry matter accumulation, highlighting the competing nature of the various crop characteristics.

This information is derived from numerous research studies in New Zealand and overseas. The level of impact varies with each study as dry matter accumulation is influenced by the many factors mentioned above, therefore the effects described in the Table 3 should not be considered cumulative.

Table 3: Orchard management practises and the influence they have on the level of dry matter in kiwifruit (as provided in a handout to growers by Zespri, 2004)

Orchard Management Practise	**Influence on Dry Matter %**
Climate	⬆ ⬇
Region	⬆ ⬇
Site (soil/aspect/shelter)	⬆ ⬇
Fruit Sizing Tools e.g. Benefit Kiwi™, NAA Gel (considered at the time)	⬇ ?
Canopy Management	
Single layered canopy	⇧
Good leaf health	⇧
Wider cane spacing	⇧
High leaf to fruit ratio	⇧
Good cane quality	⇧
Position of fruit on vine	⇧
Crop Load	⇧ ⇩
Length of Growing Season	⇧ ⇩
Trunk Girdling	⬆
Repeat Cane Girdling	⇧
Good Pollination	⇧
Vine Age	⬆ ⬇
Deficit Irrigation	⇧
Nutrition	?

Increase DM: ⇧ <0.5% DM change ⬆ ≥0.5% DM change

Decrease DM: ⇩ <0.5% DM change ⬇ ≥0.5% DM change

? Unknown or current research trying to quantify

Girdling (the equivalent of ring-barking) either the trunks or the canes of the kiwifruit vines occurs during spring and/or summer in New Zealand. Girdling and the other orchard management methods in Table 3 are employed with little knowledge of orchard homogeneity and usually before any fruit quality predictions can be made, due to timing of application versus timing of dry matter development in kiwifruit for testing.

1.4 Remote Sensing

Remote sensing and its application to agriculture is the cornerstone of this study. Application of remote sensing to kiwifruit analysis was considered to be suitable because the horizontal nature of kiwifruit canopies provides good representative reflectance from the vines. In addition, the commercial drivers within the New Zealand kiwifruit industry and the ability for canopy management to influence the resultant attributes means that remote sensing could add value to New Zealand kiwifruit growers.

1.4.1 Precision Agriculture

Precision agriculture assesses a large area of vegetation and classifies specific homogeneous sub-areas within the larger area. Improved prediction of yield and quality is based on summing the sub-areas, rather than extrapolating samples over a larger (assumed homogeneous) area and this results in more accurate predictions and financial forecasting. The modern era of precision agriculture has developed rapidly over the last forty years and now frequently includes the use of remote sensing.

Remote sensing in this study is the acquisition and analysis of images of the surface of the earth using above ground sensors which are usually airborne or spaceborne. Images are now so good, and comparatively inexpensive, that there is a trend in precision agriculture to attempt to use such images remotely to assess the condition of vegetation. Such images, from the latest generation of sensors, usually provide the numeric data for each pixel that is representative of the energy in red, green, blue and infrared wavebands. Pixels tend to be of the order of a 2 to4 metres squared on the ground.

In addition, the assessment information derived through precision agriculture can be used for targeted management of sub-areas. This is of particular value for kiwifruit orchards as there are orchard management techniques, such as girdling, that can be used to specifically improve the dry matter of kiwifruit and could be targeted to sub-areas that have been identified as having lower dry matter.

1.4.2 Vegetation Analysis

Precision agriculture, and more specifically remote sensing, studies vegetation within images and seeks to identify a change of state. In this research a base assumption is that canopy management methods will result in a change of kiwifruit canopy state and the resultant canopy state is representative of the quality of the kiwifruit produced.

Biologists ask three levels of questions when studying the environment and in particular vegetation (Brase, 2006):

- what is the vegetation?
- how much vegetation is there?
- what is the condition of that vegetation?

At a superficial level, humans can judge that some plants are luxuriant and healthy – based purely on their colour and degree of foliation. It is hoped that remote examination of foliage will provide a quantifiable measurement of colour and foliage variability that is representative of the innate attributes of the vegetation being considered, which in this study is the dry matter content of the fruit produced by kiwifruit vines.

All plants, including kiwifruit vines, capture resources (such as light, water and nutrients) from their environment and also encounter stress (such as water, nutrients, pests and climate). The crop canopy is an expression of the resultant combination of resource capture and stress, and provides an indication of plant health and yield capability (Wiegand & Richardson, 1984).

Many biochemical processes in vegetation, such as photosynthesis, respiration, evapotranspiration and decomposition, are related to the foliar concentrations of bio-chemicals such as chlorophyll, water, nitrogen, lignin and cellulose (Curran, 2001). Remote sensing of vegetation is possible because it detects gross spectral features and changes resulting from changes in pigment absorption, within-leaf scattering, water absorption, and foliar biochemical absorption.

Green plants, including kiwifruit vines, reflect light within the range of green wavelengths and absorb light within the ranges of the blue and red wavelengths, hence the green leaf colour seen by the human eye (Curran, 2001). Chlorophyll is the green pigment present in green plants and is essential for photosynthesis. Photosynthesis transforms the absorbed light energy to chemical energy for carbohydrate production (including the starch content) which is a major component of dry matter.

The Chlorophyll absorption spectrum in the visible light range is depicted in Figure 7 (Vernier, 2009).

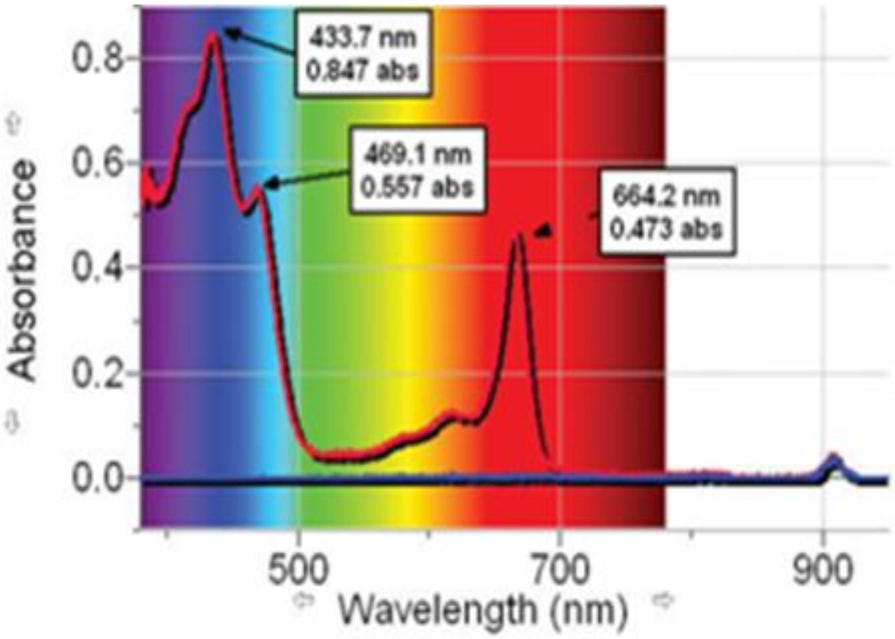


Figure 7: The absorption spectrum of chlorophyll

A specific spectral signature can be determined for a specific object when reflectance is measured over a wide range of wavelengths, and can be used to identify vegetation, as seen in Figure 8 (Brase, 2006).

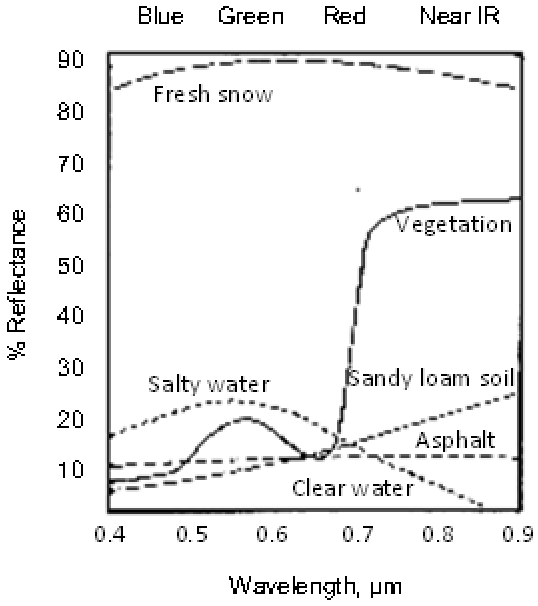


Figure 8: The reflectance spectrum of vegetation compared to other features

The curve depicting foliar reflection of different plants (Brase, 2006) is shown in Figure 9. Kiwifruit vines are deciduous and would be expected to have a reflectance profile similar to that for deciduous plants.

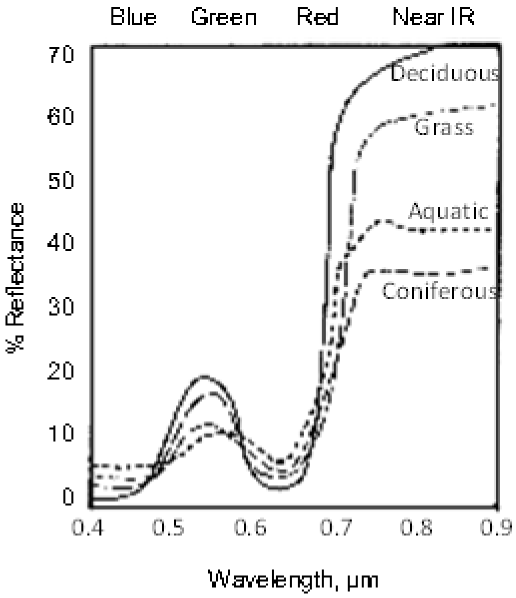


Figure 9: The reflectance spectrum of various types of vegetation

Unhealthy plants do not absorb red and blue light to the same degree as healthy plants. (Figure 10). The degree of reflected red and blue light combined with the reflected green light results in the yellow/brown appearance of unhealthy plants (Brase, 2006).

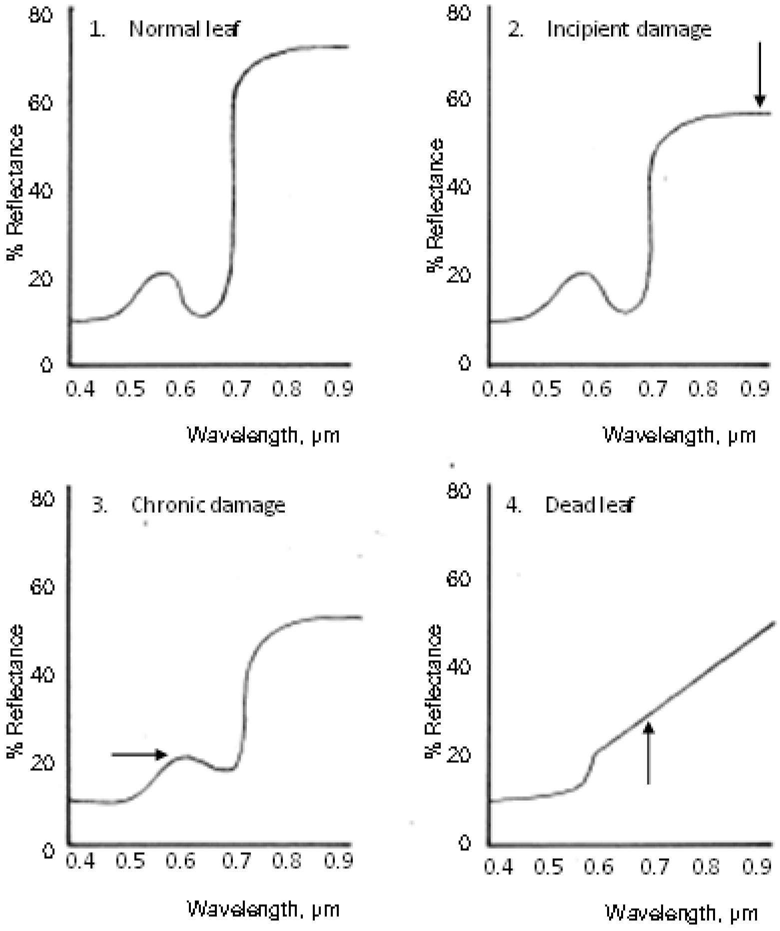


Figure 10: Reflectance spectrum variance depending on leaf health (Brase, 2006)

Based on the variations of spectral signatures reflecting from vegetation resulting from morphology and photosynthetic function it is conjectured that potential plant performance could be measured as a function of the spectral signature. The requirement of this study is to be able to differentiate kiwifruit from the other vegetation in the area and to determine the quantum and condition of the kiwifruit canopy as a predictor of fruit attributes. Multispectral satellite images of the reflected light from a kiwifruit orchard may therefore provide a precise measuring instrument – given an adequate descriptive model.

1.4.3 Kiwifruit Photosynthesis

The photosynthetic capacity of kiwifruit leaves, grown in New Zealand, has been studied by (Buwalda *et al*.,1991) and grown in the Slovak Republic by (Olah & Masarovicova, 1995) . Buwalda *et al*., (1991) found that the net CO_2 assimilation rate increased over the first 3 to 5 months following bud burst before peaking and then starting to decline. This process begins in New Zealand around the end of September.

Buwalda *et al*., (1991) found that the leaves continue to emerge over a 4 month period and therefore the canopy during the growing season contains leaves at varying stages of photosynthetic capacity. The first leaves emerging would be reaching full capacity in late February and the last leaves emerging could be reaching full capacity in late June.

There are contradictory reports on when kiwifruit vines reach full photosynthetic capacity. Other studies reported by Greer (1999) found that kiwifruit could behave like other plants, with full photosynthetic capacity being achieved at full leaf expansion. Buwalda *et al*., (1991) agreed that the leaf chlorophyll content increased and peaked at the same time as photosynthetic capacity and the specific leaf area declined with leaf age while the carbon content remained constant. The leaf nitrogen content was found to have a strong influence on radiation and CO_2 assimilation and therefore photosynthetic capacity.

Olah and Masarovicova (1995) studied the photosynthetic response of kiwifruit leaves to investigate the effect of irradiance, ambient CO_2 concentration, chlorophyll content, specific leaf area (SLA), specific leaf mass (SLM) and photosynthetic productivity (PP). The study was undertaken in the Slovak Republic (Northern Hemisphere) on *Actinidia deliciosa* with fully expanded leaves in June-July (the equivalent of December-January in New Zealand).

This study compared both male and female leaves and found little significant difference except that female leaves had greater photosynthetic capacity (refer Table 4). It is the female vines that produce the fruit and comprise the majority of the canopy.

Table 4: Average values of male and female kiwifruit leaf attributes, at full leaf expansion (Olah and Masarovicova, 1995)

	Male Leaf	**Female Leaf**
Average specific leaf area	1.282 dm^2/g	1.029 dm^2/g
Average specific leaf mass	0.805 g/dm	0.983 g/dm
Average Chlorophyll a content	5.821 mg/g (dry mass)	5.286 mg/g (dry mass)
Average Chlorophyll b content	1.976 mg/g (dry mass)	1.857 mg/g (dry mass)
Average Chlorophyll a+b	7.797 mg/g (dry mass)	7.143 mg/g (dry mass)
Average photosynthetic productivity	0.215 mg (dry mass) / m^2per s	0.349 mg (dry mass) / m^2per s

Leaf photosynthesis is driven by external factors such as light and temperature and internal factors such as leaf ontogeny and development, and carbohydrate supply and demand (Greer, 1999). The productivity of fruiting plants is roughly related to the total radiation intercepted over the growing season (Greer, 1999).

Fruit growth depends on the ability of fruit to compete with vegetative growth for a supply of assimilates from the vine leaves. Assimilates flow from a source (leaves) to the closest sink (fruit) along a gradient. In kiwifruit, assimilates can travel to fruit greater than 2m from the nearest leaf and leaves on non-fruit bearing branches can support fruit on other branches (Seager *et al*., 1995). Leaves can provide assimilates until they senesce and abscise (become old and fall off).

The leaves closest to where the fruit is hanging are a major source of carbon for fruit growth. Factors influencing photosynthesis within the canopy, especially for those leaves providing carbon sources to the fruit, would contribute to the heterogeneity in fruit size distribution found within kiwifruit canopies. Variation in both the time and the space of the interception of radiation within a plant canopy, is one of the factors influencing photosynthesis and can cause a variation in supply of assimilated carbon, this in turn affects the distribution of fruit size on the vines (Seager, *et al*., 1995).

Pruning (a 5:1 leaf-to-fruit) and girdling of laterals were found significantly to increase the carbohydrate (starch and sugar) concentrations of the fruit. A 2% increase in dry matter content of the fruit from vines with a 5:1 leaf-to-fruit ratio was achieved compared to the control (1:1 leaf-to-fruit ratio). The soluble sugar concentrations also increased more rapidly over time in the 5:1 leaf-to-fruit ratio vines (a sign of faster maturation) (Seager, *et al*., 1995).

Zhang *et al*., (2004) modelled the effect of sunshine hours on a theoretical orchard. These resulted from various combinations of orchard shelter orientations, location of the orchard and time of year. The purpose of their study was to determine the optimum orientation of shelter on an orchard to minimise wind effect and maximise sunshine quantity and uniformity. The model was based on an orchard in the Bay of Plenty (approximately 37^{o} latitude) which would be exposed to south to south-west prevailing winds.

Shelter orientation at 0^{o} (North to South) provides lower maximum sunshine hours but greater uniformity of light received. 90^{o} orientation (East to West) provided the maximum sunshine hours but the sunlight received was less uniform so that large areas received little or no sunshine. A change to a 10^{o} to 40^{o} orientation could provide better wind protection and not much reduction in quality of sunshine received compared to a 0^{o} orientation. It is presumed (but not shown in this research) that a more uniform block of kiwifruit vines will produce more uniform fruit with more consistent quality including dry matter (Zhang *et al*., 2004).

1.5 Working Hypotheses

At the time this study was commenced dry matter was a significant component of grower payments and a grower’s ability to manage their orchard to produce higher dry matter fruit provided a competitive advantage and an ability to earn an above average income. It was understood that canopy management and other orchard practises contributed to the health of the vine and therefore the quality of the fruit produced. The quality of the fruit could only be determined through destructive testing. Kiwifruit vines are grown in a horizontal aspect with the leaves above the fruit. An aerial view of the canopy could provide insights into the quality of the fruit produced. In addition, a satellite view that covered a significant portion of many orchards at the same time would provide comparative information under similar conditions. . A literature review provided some insights into the increasing use and value of remote sensing in crop management. This provided the basis on which to postulate this study. Several assumptions were proposed:

- Leaf function of kiwifruit vines contributes to the accumulation of dry matter into fruit.
- Multiple factors including weather and orchard management practices influence the kiwifruit leaf canopy and a vine’s ability to accumulate dry matter into its fruit.
- Remote sensing data determines the state of the leaf canopy regardless of what factors contribute to that state and can therefore be used to predict the dry matter potential of the kiwifruit produced.

Based on these assumptions the following questions were posed to be answered:

- Is it possible to determine from canopy characteristics the taste and dry matter components of the kiwifruit that will be obtained at harvest?
- How early can predictions be made before harvest?
- How can the predictive information be disseminated to growers for use to improve orchard management practises and to increase their returns?

CHAPTER 2 REVIEW OF LITERATURE

2.1 Introduction

A new technique to predict kiwifruit dry matter has been developed in this study. The technique developed applies remote sensing analysis, from spectral measurements captured by satellite imaging, to predict kiwifruit dry matter. This chapter is a review of the studies that have used remote sensing methodologies with agricultural applications to predict crop attributes.

2.2 Remote Sensing Options

Moran *et al*., (1997) concluded that image based remote sensing could provide useful information for precision crop management of time-constrained activities. They evaluated both aircraft and satellite based systems and found that although both systems had advantages, further work was required to deliver useful remote sensing in agriculture.

Aircraft imaging systems were commonly used because they could provide multispectral imagery with a spatial resolution of less than 1 metre up to a few metres per pixel. Satellite based sensors were limited in both imagery and spatial resolution, with fixed spectral bands and coarse spatial resolutions of greater than 10 metres. In addition they had limited repeat coverage having only one data set obtained per infrequent overhead pass and time delays between image acquisition and delivery of data to users. Whilst aircraft based sensors did not have these limitations they were in turn difficult to calibrate and hard to register to map co-ordinates over large areas (Moran *et al*., 1997). Satellite based sensors could provide the opportunity to capture data over wide areas at exactly the same time under the same conditions (synoptic viewing). This synoptic viewing capability was the reason a satellite option was chosen for this study.

The launch of the IKONOS and QuickBird satellites (in 1999 and 2001 respectively), with their higher spatial resolution (0.6 to 4 metres) and higher frequency of overhead pass, opened up new opportunities to explore the value of remote sensing using satellites. The specifications of this generation of satellites are detailed in Appendix 1.

Curran (2001), in his review, noted a rapid transition from airborne to spaceborne imaging spectrometry following the launch of these more advanced satellite sensors. Geological research was the main driver behind the development of imaging spectrometry during the

1990’s and this was the basis from which ecological remote sensing was developed. The transition to spaceborne platforms offered new tools and opportunities for ecological studies at both local and global levels.

Lillesand *et al*, (2004) summarised the characteristics of an ideal remote sensing system compared to the reality at that time (Appendix 2). Apart from issues surrounding data processing and multiple data users there is also the complexity of the energy source, atmosphere, energy-matter interactions and the sensor itself. The effect of the atmosphere varies depending on the type of sensor and platform used. The further away the sensor platform is from the earth’s surface the greater the impact of the atmosphere. The requirements for atmospheric attenuation are covered later in this chapter.

Dorigo *et al*., (2007) stated three conclusions related to remote sensing using reflectance measurements:

1. The main value was that it could provide information on the state of vegetative canopies and identify variances in those canopies. This was because remote sensing enabled study at a level of coverage, and at spatial and temporal resolutions which could not be achieved by ground level observation.
2. Much of the data being analysed in studies at that time was limited by being single-sensor, mono-directional and mono-temporal.
3. Canopy reflectance is a result of many interacting factors and different combinations of those factors can result in an ambiguous reflected signal which makes correlation impossible or complex.

The capability of remote sensing systems has improved over time, in step with improved technological capability. Studies using the higher sub-metre spatial resolution imagery from the QuickBird satellite launched in 2001 (DigitalGlobe, 2009) were now being undertaken (Bausch *et al.*, 2008; Yang *et al.*, 2006). Results from these studies showed that spaceborne imagery analysis was now achieving results comparative to those previously achieved from airborne imagery and were just as useful for within-field crop management. The research from these particular studies is covered in greater detail in Section 2.4.1.

Researchers are now investigating the use of unmanned aerial systems as an alternative platform for spectral data collection. Unmanned aerial systems have a potential lower operational cost, improved temporal and spatial resolution and greater flexibility in image acquisition programming (Zhang and Kovacs, 2012); (Jannoura *et al.*, 2015). However, improvements in platform design, standardisation of geo-referencing and mosaic compilation and information extraction are still required to increase the reliability of results.

2.3 Vegetative Remote Sensing

More than a quarter of the global land surface area is covered by agro-ecosystems (human-induced ecosystems orientated towards the production of food and or other valuable goods such as timber, fibre or fuel). Models that can describe the response of agro-ecosystems to changes in management strategies and/or to changes in climatological and environmental conditions are desirable tools.

Numerous studies conducted over the last twenty years have developed vegetative indices and spectral variables from multispectral images. These can be correlated with yield data, leaf area index (LAI), photosynthetic potential (chlorophyll content) or biomass (Hatfield *et al.*, 2008); (Cammarano *et al*., 2014).

Multispectral imagery is created when a sensor records several reflectance values into bands. The most common bands are blue, green, red and near-infrared. The captured reflectance by vegetation in these four bands can be used to develop correlations against characteristics such as yield, LAI, photosynthetic potential and biomass.

Hyperspectral images are multispectral with a fine wavelength resolution. Vegetation indices from hyperspectral images are related to very specific wavelengths and have been used in many leaf-level studies focused on predicting crop physiological status through the extraction of leaf biochemical parameters such as concentration of chlorophyll a + b, carotenoids, water and dry matter. However for application on a larger scale it is more useful to predict crop physiological status from broader measurements, usually over four bands.

2.3.1 Reflectance, chlorophyll and photosynthesis

Chlorophyll content is important because it is related to nitrogen content and therefore to photosynthesis and is a guide to the general physiological status of the vegetation (Haboudane *et al.*, 2008). In addition, plant photosynthesis produces ninety percent of the plant dry matter and is the basis for dry matter accumulation and yield in crops (Tian *et al.*, 2005).

Conditions that influence photosynthesis therefore have an effect on the quality and quantity of crops. The impact of water and nitrogen stress on photosynthesis is one of the major effects on yield and quality variation. Chlorophyll content of green leaves is often used as an indicator of plant photosynthetic productivity and the spectral reflectance in visible wavelengths of plants is mainly due to chlorophyll content (Al-Abbas et al., 1974). Refer also to Figure 7 in Chapter 1 - the absorption spectrum of chlorophyll.

A change in the reflectance/absorbance of leaves is greatest in the visible part of the electromagnetic spectrum during leaf expansion as leaf pigments develop. (Jacquemoud, *et al*., 1996) found that there was little change in the NIR spectral region during leaf development. The leaf optical properties in the NIR spectral region are driven by the mesophyll structure, dry matter and water content of leaves as depicted in Figure 11 (uncited internet image, (LED Digital Visuals, 2009).

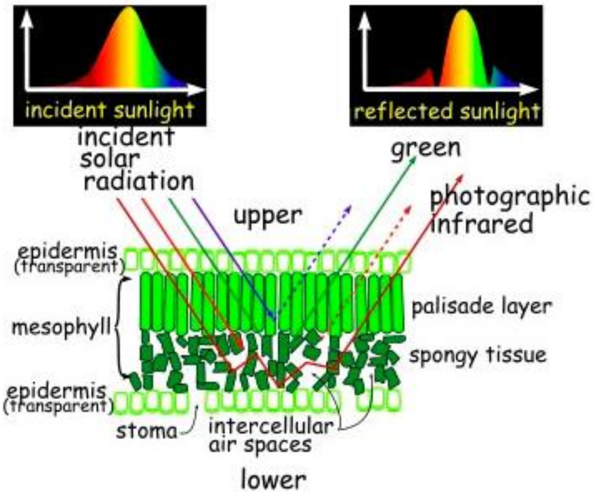


Figure 11: Leaf structure and utilisation of incident sunlight, (LED Digital Visuals, 2009)

(Gitelson and Merzlyak, 1997) studied leaves from 6 different plant species (tobacco, chestnut, cotoneaster, fig, maple and vine) in a laboratory setting to determine the reflectance sensitivity, at different wavelengths, to varying levels of chlorophyll. The study found that there was a high level of reflectance sensitivity across a wide range of chlorophyll concentrations (0.3 to >60 µg/cm2), across a wide radiance band 530-630nm and again at a narrow band near 700nm with correlation r^2 values of 0.91 to 0.98 across all 6 species tested.

Reflectance outside of these two wavelength zones, in particular in the near-infrared region (750-900nm) were considered to be insensitive to change in chlorophyll content.

Two ratios (ρ750nm/ρ550nm) and (ρ750nm/ρ700nm), where ρ=reflectance, were developed and analysed for predictive capability of chlorophyll content. The validation between predicted and measured chlorophyll content had an r^2 value of 0.94 for both ratios.

2.3.2 Leaf Area and Photosynthetic Capacity

Numerous studies have investigated leaf area correlations using Leaf Mass Area (LMA), Specific Leaf Area and Leaf Area Index (LAI). They found that the Normalised Difference Vegetation Index (NDVI) most consistently correlates with the amount of vegetation present and LAI in particular. This is defined below.

Within species, photosynthetic capacity generally has a positive correlation with leaf dry mass per unit area (LMA) (Niinemets, 1999) but correlates differently with LMA and nitrogen content per unit area between species and within species grown in different environments.

Specific leaf area ((SLA), leaf area per unit dry mass) is considered to indicate the expected return (yield, quality, fruit production, plant health) on those resources captured during photosynthesis. High SLA leaves are productive, short lived, vulnerable to herbivores and work best in a resource-rich environment (Wilson *et al*., 1999).

Leaf Area Index ((LAI), the quantity of leaves in the canopy i.e. area of leaf per area of ground) is an input parameter in many biophysical models and is used to measure growth, predict yield and monitor change in canopy structure caused for example through levels of irrigation, drought or disease. In capital-intensive orchard systems, LAI estimates can assist with better management decisions e.g. pruning and precision farming. LAI measurements are time-consuming, expensive and often not feasible for remote areas. There are two groups of methods used to estimate LAI from remote sensing data – the model-based approach and the empirical approach. The empirical approach is based on the use of vegetation indices. The advantage of the empirical approach over the model-based approach is that it requires no previous site-specific information and the computation time is shorter (Delalieux, *et al*., 2008).

Jordon (1969) determined a correlation between the red to near-infrared (NIR) spectral ratio and LAI. This correlation has been defined as the Normalised Difference Vegetation Index (NDVI) and is widely studied and used to predict LAI.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})}$$

Large scale application of reflectance measures at canopy level is difficult because LAI and chlorophyll act simultaneously on the spectrum. NDVI saturates when applied to dense canopies and, secondly, LAI and chlorophyll have similar effects from the green region (~500nm) to the red-edge (~750nm). Thus vegetative indices based on red and NIR will overestimate LAI for canopies with high chlorophyll content and underestimate canopies with low chlorophyll content (Delalieux, *et al*., 2008).

Many studies have tried to improve the NDVI performance. Gitelson (2004) proposed a modification of the NDVI to create the wide dynamic range difference index (WDRVI). A coefficient “a” ranging between 0.05 and 0.20 was applied to the NIR value.

$$\mathrm{WDRVI} = \frac{(\mathrm{a} \times \mathrm{NIR} - \mathrm{Red})}{(\mathrm{a} \times \mathrm{NIR} + \mathrm{Red})}$$

This modified index resulted in a linear correlation against the fraction of vegetation across the full range of growth (Gitelson, 2004).

(Lofton, *et al*., 2012) considered the impact of in-season estimation of yield potential of sugarcane and whether the NDVI values could be adjusted using factors such as cumulative growing degree days (CGDD) and days from the beginning of the year (DFY). They found that adjusting the NDVI readings using CGDD and DFY increased the accuracy of estimating yield potential but only CGDD adjustment improved the relationship between NDVI and cane tonnage.

2.4 Indices in Use

Appendix 3 gives a summary of indices and their uses. This is not an exhaustive summary. There have been many other indices defined and more recent reviews of indices that evaluate their performance under different conditions and comment on the ability to use such indices across a wider range of vegetative types and using data from different collection sources (Rembold *et al.*, 2013); (Anatoly and Gitelson, 2013); (Hunt, *et al.*, 2013); (Huang *et al.*, 2013), (Xiao *et al.*, 2013); (Wang *et al.*, 2014); (Son *et al.*, 2014).

Vegetative indices, such as NDVI, derived using empirical approaches are useful for heterogeneous vegetative systems and where differences in atmospheric and solar conditions between images need to be minimized. They do however have difficulty in separating the multi-spectral characteristics of biomass from leaf quality (Delalieux, et al., 2008). Many new or improved indices are being developed.

2.4.1 QuickBird Satellite Research

Beeri *et al*., (2005) used data captured from several different satellites (Landsat 5, SPOT 5, Quick-Bird 2, Ikonos 2) and linear regression to develop correlations for biomass (at peak season) and leaf quality (just prior to harvest) for sugar beet grown in North Dakota, USA. Leaf quality was determined by measurements of nitrogen (N) and carbon (C) content and the C:N ratio. Leaves with a high C:N ratio are considered low quality as leaves are not easily mineralized (broken down and releasing minerals back into the soil).

Digital numbers (DN), a value representing the relative level of reflectance captured by the satellite sensor, from the acquired images were corrected with ground reflectance using the following equation:

$$\mathrm{R_x} = \frac{(\pi * \mathrm{d}^2 * (\mathrm{gain_x} * (\mathrm{DN_x} - \mathrm{min_x})))}{(\mathrm{E} - \mathrm{SUN_x} - * \sin(\theta))}$$

where R is the ground reflectance for each band (x), d is the Earth-Sun distance in astronomical units for the image date, gain is the band specific rescaling factor, DN is the digital number in the raw image, min is the lower DN in the specific band, E-SUN is the mean solar exo-atmospheric irradiance and θ is the sun elevation angle (Beeri *et al.*, 2005 .

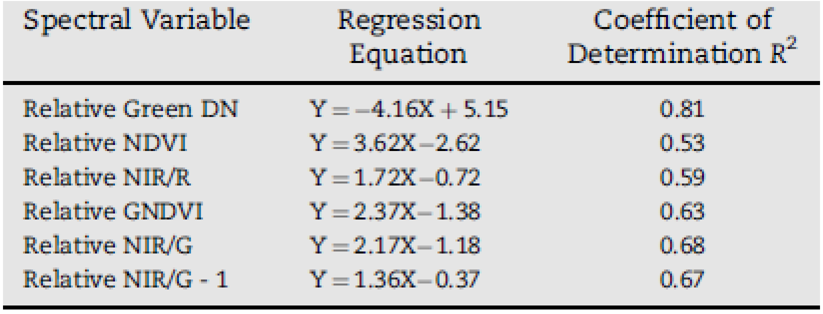
Leaf level analysis showed that sugar beet had optimum C and C:N ratio detection in the green and blue spectral regions. Satellite spectral models were able to map N-credit per unit area and also provide qualitative canopy assessment where leaf C:N ratio models could be used to delineate areas of very poor leaf quality that would potentially have lower organic nitrogen available the following season. Biomass of homogenous sugar beet crops in a single image correlated well with NIR band data alone and the use of indices such as NDVI and GNDVI did not increase accuracy Beeri *et al*., (2005).

Yang *et al*., (2006) studied the yield of grain sorghum in Texas, USA. They found that 3-band (Green, Red and NIR) or 4-band (Blue, Green Red and NIR) regressions derived from QuickBird satellite imagery worked just as well as the best vegetative index (NG

(NIR/Green)). R^2 values ranged up to 0.81 for the various combinations and resolutions reported.

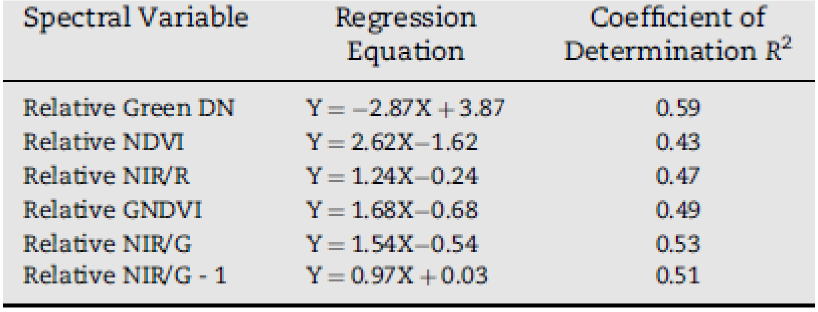
Bausch *et al*., (2008) studied maize crops grown in Colorado, USA. They looked for correlations between leaf nitrogen content, yield (grain only) and biomass (total vegetation including grain) against spectral data from QuickBird imagery and from ground based spectral measurements. The strongest correlation for all three characteristics was with the normalized green waveband (R^2 values of 0.59 to 0.81). The vegetative index also provided reasonable correlations (R^2 values of 0.43 to 0.72) and indices based on the green waveband had stronger correlations than indices based on the red waveband. Comparison of QuickBird data with ground-based reflectance data captured at the same time indicated that the chlorophyll index (NIR/Green-1), normalised by the mean value of the index from the green waveband, might not require atmospheric correction for comparisons within an image. Table 5 and Figure 12 shows the results for the various indices tested for correlation to biomass and yield (Bausch, *et al*., 2008).

Table 5 (a)-(b): Relationships between relative spectral variable (X) and relative grain yield (Y) in (a) or relative total biomass (Y) in (b)



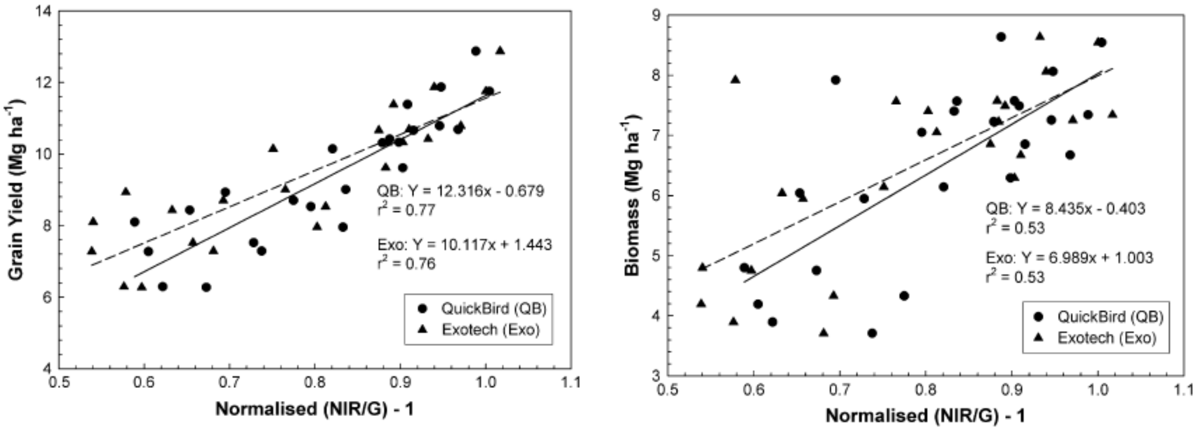
Spectral Variable	Regression Equation	Coefficient of Determination R^2
Relative Green DN	$Y = -4.16X + 5.15$	0.81
Relative NDVI	$Y = 3.62X - 2.62$	0.53
Relative NIR/R	$Y = 1.72X - 0.72$	0.59
Relative GNDVI	$Y = 2.37X - 1.38$	0.63
Relative NIR/G	$Y = 2.17X - 1.18$	0.68
Relative NIR/G - 1	$Y = 1.36X - 0.37$	0.67

(a) Grain Yield



Spectral Variable	Regression Equation	Coefficient of Determination R^2
Relative Green DN	$Y = -2.87X + 3.87$	0.59
Relative NDVI	$Y = 2.62X - 1.62$	0.43
Relative NIR/R	$Y = 1.24X - 0.24$	0.47
Relative GNDVI	$Y = 1.68X - 0.68$	0.49
Relative NIR/G	$Y = 1.54X - 0.54$	0.53
Relative NIR/G - 1	$Y = 0.97X + 0.03$	0.51

(b) Biomass



(a) Grain Yield (b) Biomass

Figure 12 (a)-(b): Relationship between normalized (NIR/G)-1 (X) and relative grain yield (Y) in (a) or relative total biomass (Y) in (b) calculated from QuickBird DN and Exotech ground-based reflectance values

2.4.2 Research using other Satellite Systems

Tian *et al*., (2005) studied the leaf photosynthetic characteristics of rice in China under different water and nitrogen conditions in a controlled environment using ground based equipment. This study showed a linear relationship (with an R^2 value of 0.69) between canopy leaf photosynthetic potential and a NIR/Red ratio index of wavelengths (810nm/680nm) as seen in Figure 13. The near infrared band is insensitive to change in photosynthesis and the red band is more sensitive to stress manifested through a change in photosynthesis. The ratio between the two could therefore correct for any influence from other factors such as irradiance variance, leaf orientation, irradiance angles and shading.

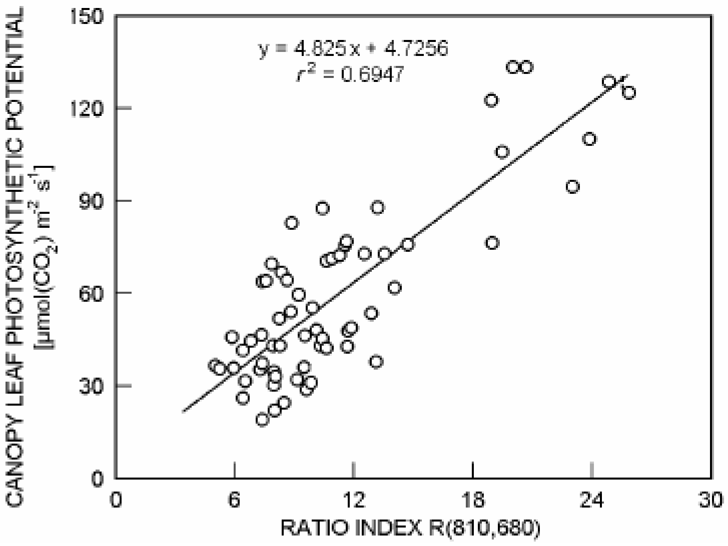


Figure 13: Relationship between canopy leaf photosynthetic potential and ratio index R(810,680)

Hunt *et al*., (2005) used model aircraft based sensors to capture data to determine the nutrient status of corn and the biomass of corn, alfalfa and soybean crops. This study found a linear correlation between a normalized Green Red Difference index (NGRDI) and biomass.

$$\mathrm{NGRDI} = \frac{(\mathrm{Green} - \mathrm{Red})}{(\mathrm{Green} + \mathrm{Red})}$$

Linearity was maintained until biomass level of <150 g m^2 was reached (before full canopy closure) and then plateaued off as shown in Figure 14. NGRDI was found not to have any consistent correlation to nutrient status being unable to detect variations in chlorophyll level (Hunt *et al*., 2005). They concluded that the model aircraft system used in their study has potential but needed better sensors.

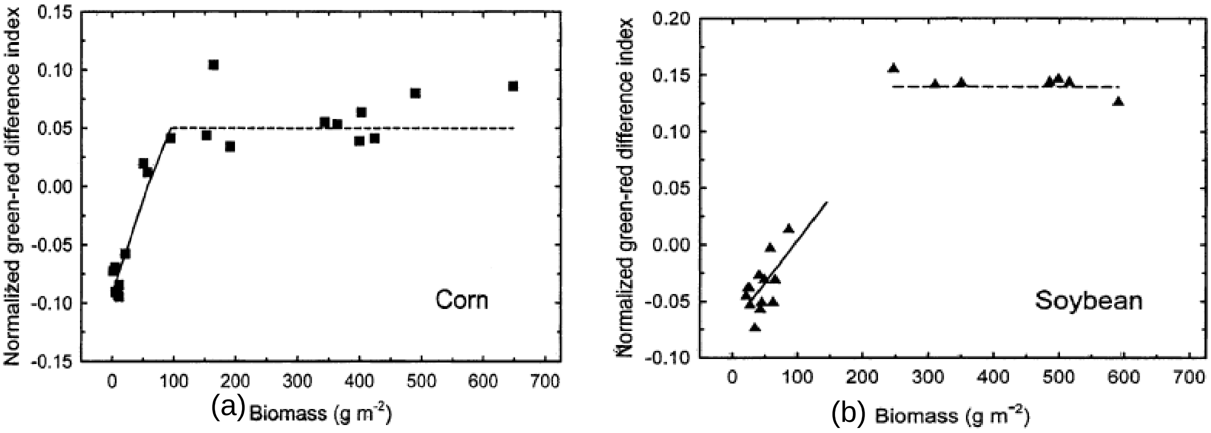


Figure 14 (a)-(b): NGRDI versus corn biomass (a) and soybean biomass (b)

Balaghi *et al*., (2008) studied wheat yield in Morocco. Early wheat yield predictive information enables management of wheat import requirements and drought risk management. Official estimates of yield were based on costly field surveys and the information was often available too late to be of use. This study focused on the Normalised Difference Vegetative Index (NDVI) as a predictor of yield because it is especially useful in semiarid regions with high year-to-year variations in vegetation. NDVI imagery (10-daily over several months) registered by the earth observation system NOAA-AVHRR (National Oceanic and Atmospheric Administration, 2009) was used for years 1990 through 2003.

NDVI data was summed over periods of time, together with rainfall and temperature data to develop provincial and national forecasting models using empirical, ordinary-least-squares-regression analysis. Figure 15 details the predicted yield compared to the observed yield using near-harvest satellite imagery for 1990 through 2003. NDVI was the most important parameter and was able to explain most of the yield variation, although the model achieved greater accuracy incorporating rainfall and temperature for arid, high-rainfall areas. Accurate forecasts could also be achieved starting with data as early as 2 months prior to harvest.

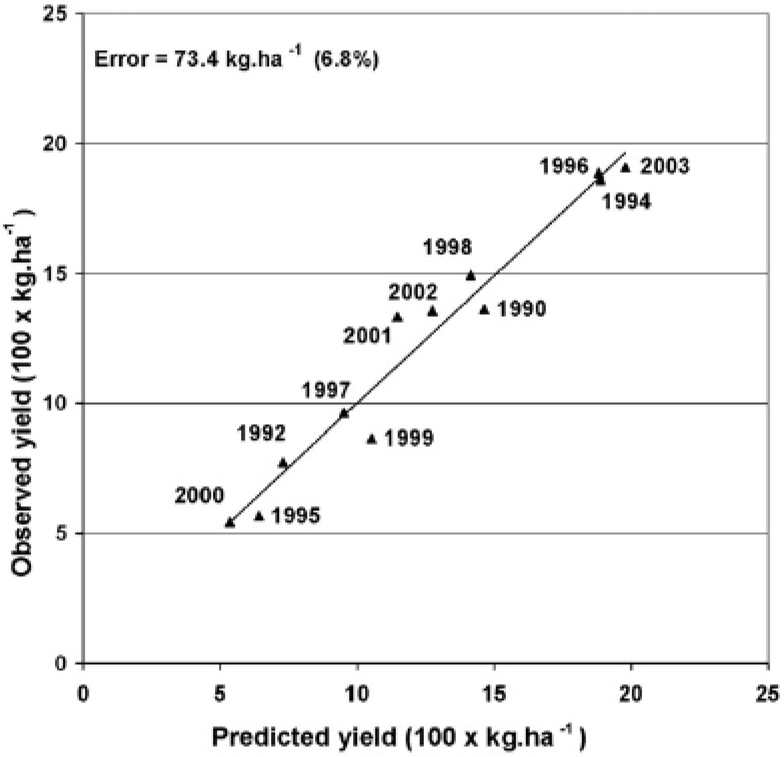


Figure 15: Relationship between observed and predicted wheat yields at a national level, Morocco, 1990 through 2003

Bégué *et al*., (2008) studied time-series data to evaluate the dynamics of within-field variability of sugarcane over the French West Indies. Crop growth variability is related to many factors all of which can be interactive. These include:

- Time-independent factors associated with the substrate such as topography, soil type and depth or previous land use
- Annual factors such as differences in planting and emergence or weather conditions.
- Seasonal factors such as plant disease, weed development, severe climatic events or irrigation malfunction.

NDVI data was derived from SPOT images over French West Indies captured at different times and over a 3-year period. To get comparable physical values between images, the digital counts were converted to Top-of-atmosphere (TOA) reflectance values using the following equation:

$$\rho = \frac{\mathrm{DC}.\pi}{\mathrm{G}.\cos(\theta_s).\mathrm{E}_s}$$

where ρ is the TOA reflectance, DC is the digital count, G is the sensor absolute calibration gain (contained in the metadata associated with the image), θ_s is the solar zenith angle and E_s is the solar radiation in the appropriate wavelength. No atmospheric correction was used because the study only used clear sky images and the image processing (classification and correlation) was done on a daily basis. The standard NDVI formula was used with the ρ value calculated for the red and near infrared bands:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}}$$

This study found that the best correlation between NDVI and yield were obtained from images acquired around peak vegetative development but before onset of senescence. The study concluded that:

- NDVI values are linked with topography and plant water status
- knowledge of the crop phenology (growth stages) is required to interpret the spectral pattern
- single-data images may be insufficient to diagnose crop condition or for prediction
- the pattern of vigour within the fields can help diagnose growth anomalies that can be verified by focused ground inspection and sampling
- satellite remote sensing has the potential to support crop-management decision making

Hyperspectral studies of citrus in Japan found that a partial-least-squares regression of factors, in addition to reflectance, provided greater prediction of fruit yield (Ye *et al.*, 2008). 72 visible and near-infrared wavelengths from 407nm to 898nm, from an Airborne Spectrometer for Applications (AISA) Eagle system were used in the study. Yield-prediction models tested in this study used five techniques:

- several vegetative indices (VIs)
- key wavelengths determined by simple-correlation analysis (SCA)
- principal components based on principal-component regression (PCR)
- partial-least-squares factors (PLS)
- important wavelengths determined by B-matrix based on PLS regression

Vegetative indices did not correlate and the principal components extracted from the data were found to be weak predictors of yield. Key and important wavelength models showed improved correlation but the best yield prediction was from models using factors based on canopy features such as canopy size (Ye, *et al.*, 2008).

Ye and Sakai (2012) did further research using multispectral imaging which focused on NDVI as a predictor of citrus fruit yield. This later study confirmed the interrelationships between canopy features and the fruit yield of citrus crops (Ye & Sakai, 2012).

2.5 Atmosphere

Aside from spatial resolution and sensor limitations, airborne and satellite-based sensor platforms suffer from atmospheric interference.

2.5.1 Atmospheric Interference

Atmospheric interference is the result of scattering and absorption of reflectance signals by gases and aerosols in the atmosphere. Atmospheric scattering and absorption depend on wavelength (Gaut *et al.*, 1975).Atmospheric scattering:

- Does not change wavelength
- Reduces the quantity of incoming radiation reaching the earth's surface
- Is dependent on the wavelength of incoming radiation and the size of the scattering particles
- The earth's atmosphere has a large number of particles of about 0.5 microns. These preferentially scatter short wavelengths.

Atmospheric absorption:

- Gases or particles in the atmosphere capture solar radiation and emit as long waves
- Most reflection in the atmosphere is caused by clouds
- Reflectivity of a cloud can range from 40% to 90%

Most of the radiation from the sun is in the visible band (400-700nm) centred on 500nm. Approximately 30% of the total incoming sunlight is reflected by either the earth's atmosphere or the earth's surface.

Wavelengths shorter than 200nm are absorbed in the atmosphere by oxygen and nitrogen, 200-300nm wavelengths are absorbed by ozone and wavelengths greater than 700nm (infrared) are absorbed by carbon dioxide, ozone and water (present in the earth's atmosphere in both liquid and vapour form). See also Appendix 4.

2.5.2 Atmospheric Attenuation

Atmospheric attenuation is not considered necessary when analyzing data from a single image provided that the training data and classification data are in the same relative scale (Lillesand *et al.*, 2004, Song *et al.*, 2001). However, for multi-temporal analysis, atmospheric attenuation is essential unless the images are rectified and the digital numbers (DN) of all images are placed in a single data set so that training data is derived from the combined data set before classification (Song, *et al*., 2001).

Song *et al*., (2001) studied multi-temporal Landsat TM images taken over the Pearl River Delta, China and compared the impact of different atmospheric attenuation methods on classification and change detection. They determined that simple dark object subtraction, with or without the Rayleigh atmosphere attenuation, or relative atmospheric attenuation was most effective and that the more complex attenuation algorithms did not improve performance. Rayleigh scattering is the scattering of light by particles smaller than the wavelength of the light.

Unless accurate measurements of atmospheric optical properties are recorded at the time of image capture, radiative transfer codes (RTC's) (Song, *et al*., 2001), also known as radiative transfer models (RTM's) (Wu *et al.*, 2005); (Arnis *et al.*, 2014), cannot be used. Collection of atmospheric optical properties is impractical if not impossible to obtain and the procedures required are too expensive to use operationally. Thus image-based attenuation algorithms are used instead. These can be absolute or relative. Absolute attenuation converts the DN values to surface reflectance. Relative attenuation adjusts the DN values so that the same DN values in the corrected images represent the same reflectance regardless of what the actual ground reflectance value may be.

Dark object subtraction (DOS) is one of the most widely used and simple of the methods. Dark objects have zero or small surface reflectance. This method assumes the image contains dark objects and that there is a horizontally homogeneous atmosphere. Therefore the minimum DN value in the histogram of the entire image is considered to be the effect of atmosphere and is subtracted from all other pixels (Chavez, 1989). Areas of water or stable areas of dense dark vegetation within an image can be used for DOS.

Hall *et al*., (1991) described a radiometric rectification method for multi-temporal images based on using one image as a reference and radiometric control sets of objects within the images that would not be expected to change over time. The control-set values were used to determine a calculation to transform each image with respect to the reference image, so that the control sets would appear the same in all images. This method was checked against actual atmospheric, optical-depth data collected at the same time as the reflectance values and it was found that this method could correct the reflectance for the effects of relative atmospheric differences to within 1% for visible and near infra-red bands. The radiometric control sets used were for bright objects and dark objects (Hall, *et al*., 1991).

The rectification transform calculated is as follows:

$$\text{Transformed } x_i = m_i x_i + b_i$$

where

$$m_i = \frac{(B_{Ri} - D_{Ri})}{(B_{Si} - D_{Si})}$$

$$b_i = \frac{(D_{Ri}B_{Si} - D_{Si}B_{Ri})}{(B_{Si} - D_{Si})}$$

and

x = digital number of reflectance; i = band

B=bright radiometric control set; D=dark radiometric control set;

R=reference image; S=subject image

The assumption of a linear relationship between image bands across time is the basis for relative atmospheric attenuation. Measurements of reflectance from pseudo-invariant features (PIF's) within the image are used to determine the linear relationship. PIF's are spatially well defined and spectrally radiometrically stable objects such as water, large areas of asphalt or concrete and areas of mature evergreen forest. Song *et al*., (2001) developed a variation on this approach which they called the Ridge Method.

The Ridge Method uses all of the information in the image and does not require the identification of PIF's. A density plot was created by plotting the DN value for each pixel at one wavelength in one image taken at time 1 against the DN value for each pixel at the same wavelength in the second image taken at time 2. Reflectance from stable areas in the images creates a ridge through the centre of the density plot and the formula of the line passing through the ridge defines the relationship between the two images. Attenuations need to be calculated for all pixels at each wavelength relative to the base image. In a multi-temporal series one image is used as the base image and all of the rest of the images in the series are corrected back to the same base image.

Kaufman and Tanré (1996) developed an indirect attenuation method for the NDVI that makes it more resistant to atmosphere. The Atmospheric Resistant Vegetation Index (ARVI) is defined as:

$$\mathrm{ARVI} = \frac{(\mathrm{NIR} - (2\mathrm{Red} - \mathrm{Blue})}{(\mathrm{NIR} + (2\mathrm{Red} - \mathrm{Blue})}$$

They (Kaufman & Tanré, 1996) found that ARVI was on average four times less sensitive to atmospheric effects than the NDVI when correlating it as a function of vegetation cover.

NDVI and simple ratio indices can be greatly affected by atmosphere (McDonald *et al.*, 1998). The NDVI formula should be rewritten to take into account additive atmospheric effects:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red}) - (\mathrm{ANIR} - \mathrm{ARed})}{(\mathrm{NIR} + \mathrm{Red}) - (\mathrm{ANIR} + \mathrm{ARed})}$$

where ANIR is the additive atmospheric effect of the NIR band and ARed is the additive atmospheric effect of the Red band (Song, *et al*., 2001).

(Chen, *et al*., 2004) developed a simple method to obtain high-quality NDVI time-series data based on using the Savitzky-Golay filter to reduce effects caused by cloud contamination and atmospheric variability. Savitzy and Golay, (1964), proposed a simplified least-squares-fit convolution for smoothing and computing derivatives of a set of consecutive values. The general equation of the simplified least-squares convolution for NDVI time-series smoothing is as follows:

$$\mathrm{Y}_{\mathrm{j}}^{*} = \frac{\sum_{\mathrm{i}=-\mathrm{m}}^{\mathrm{i}=\mathrm{m}} \mathrm{C}_{\mathrm{i}}\, \mathrm{Y}_{\mathrm{j}+\mathrm{i}}}{\mathrm{N}}$$

where Y is the original NDVI value, Y* is the resultant NDVI value, C_i is the coefficient for the *i*th NDVI value of the filter (smoothing window), and N is the number of convoluting integers and is equal to the smoothing window size (2m+1). The index j is the running index of the original ordinate data table. The smoothing array (filter size) consists of 2m+1 points, where *m* is the half-width of the smoothing window. This study concluded that the new method developed was more effective than the BISE (best index slope extraction) algorithm and Fourier-based fitting method.

Wu *et al*., (2005) also evaluated atmospheric attenuation methods for humid environments. They studied corn and potato canopies in Minnesota using QuickBird imagery. The algorithms evaluated were dark object subtraction (DOS), cosine approximation model (COST) and the apparent reflectance model (AR) compared to ground-based measurements. Their results showed that a combined DOS-COST model worked best to correct the visible wavelengths but did not work with NIR wavelength. However, they also devised a separate NIR modification which only required air temperature and relative humidity readings.

The DOS-COST model used for the visible wavelength atmospheric attenuation is:

$$\rho_\lambda(x,y) = \frac{\pi}{T_\lambda^v} \frac{L_\lambda^s(x,y) - L_\lambda^p}{T_\lambda^s D E_\lambda^0 \cos(\theta^3) + E_\lambda^d}$$

$\rho_\lambda(x,y)$ is the spectral reflectance at-surface for an image pixel at column x and row y; E_λ^o (W/m^2 µm) is the extraterrestrial solar spectral irradiance for wavelength λ (µm) at the mean Earth-Sun distance; D is the correction factor for the Earth-Sun distance; E_λ^d (W/m^2 µm) is down-welling, atmosphere-scattered, solar spectral irradiance; L_λ^s (W/m^2 sr µm) is total spectral radiance received at-sensor within the sensor's field of view; L_λ^p (W/m^2 sr µm) is path radiance or upwelling atmosphere-scattered spectral radiance within the sensor's field of view; T_λ^s and T_λ^v are atmospheric spectral transmittances at λ in solar path with a solar zenith angle θ^s and in view path with a satellite view angle θ^v, respectively.

E_λ^0 was obtained from the revised solar spectrum for each band of QuickBird data

D was represented by a Fourier series

It was assumed $E_\lambda^d = 0$

$$L_\lambda^s(x,y) = \frac{K_\lambda DN_\lambda(x,y)}{\Delta_\lambda}$$

where K_λ (W/m^2 sr count) is the absolute radiometric calibration factor for spectrum λ with an effective bandwidth of Δ_λ (µm); K_λ values are derived from the image metadata files.

L_λ^p was estimated with the relative spectral scattering DOS model (λ^{-2}) under clear atmospheric conditions. Clear deep lakes were used as dark objects and one-percent DOS techniques were applied to estimate L_λ^p in the blue band as the radiance from these dark objects minus the radiance from surfaces with 1% reflectance. The estimated L_λ^p in the blue band was used as the starting values for the computation of L_λ^p for other bands of each image with the relative spectral scattering model.

$T_\lambda^s = \cos(\theta^s)$; the value of θ^s was found in the image metadata files

$T_\lambda^v = \cos(\theta^v)$; the value of θ^v was found in the image metadata files

The NIR band atmospheric transmittance modification is:

$T_\lambda^\theta = T_\lambda^m T_\lambda^a T_\lambda^w T_\lambda^o$

T_λ^θ varies with meteorological conditions and atmospheric gas profiles. Under clear sky conditions, the content of atmospheric permanent gases (nitrogen, oxygen and argon) and ozone amount are relatively constant over time and horizontally homogeneous. It was also found that aerosol loading is relatively stable for the study region around noon on clear summer days and atmospheric pressures only vary slightly. Therefore, for a fixed solar zenith angle or view angle, the effects of Rayleigh scattering (T_λ^m), aerosol extinction (T_λ^a), and ozone absorption (T_λ^o) tend to be relatively invariable. Therefore the only variable for NIR is water vapour (T_λ^w).

2.6 Summary

Research has shown that the state of vegetation at any given time is a result of multiple interacting influences (climate, nutrients, and pests). The health, state and photosynthetic capability of the vegetation can be determined from studying the reflectance of radiant energy (sunlight) by the vegetation. Leaf photosynthesis (the source) produces carbohydrate (sugar and starch) and contributes to the quality of fruit (the sink).

Most crop production research has looked at the influence of specific factors (fertilizer, spray or irrigation requirements), chlorophyll or other biochemical content, growth over time, leaf area and leaf mass in isolation. Considering the research and advancements in technology it should be possible and practical, especially in a specific vegetative rich area such as kiwifruit in Te Puke, to derive reflectance information from a canopy that would provide useful insights into the quality of fruit being produced without having to know exactly the contributing factor/s.

In the kiwifruit industry this would still provide information that could add to the overall value of the crop through harvest and post-harvest management decisions. Also depending on the timing of being able to determine the crop status it might still be possible to use some techniques such as girdling or thinning to further influence the fruit status prior to harvest.

With improved spatial resolution and frequency of overpass, satellite-based remote sensing has been shown to predict crop production as well as other airborne or in-field systems. With spatial resolutions of less than 4m and the canopy area of individual kiwifruit vines being 25-35 m^2, satellite-based remote sensing should also be able to provide insight into the combined effects of within-vine and between-vine variations. Satellite-based remote sensing also has the advantage of synoptic viewing, which allows for comparative assessments over a large area at the same point in time. This could provide insight into maturity timing and harvest phasing.

As a result of the technical specifications of the QuickBird satellite and the benefits of remote sensing using satellite imagery, it was decided to use QuickBird satellite-based remote sensing to study kiwifruit orchards in the Te Puke region to develop tools to assist in pre-harvest decision-making. Two atmospheric attenuation methods were tested based on the method of Hall *et al*., (1991) and the method of Song *et al*., (2001). This is discussed further in Chapter 3.

CHAPTER 3 MATERIALS AND DATA PREPARATION

3.1 Introduction

The materials used in this thesis are the processed satellite images, the digital data from these images, the ground-based sample data measured directly from kiwifruit orchards and the software used for the data analysis.

This chapter details the sources of data, software programmes and the methodology used to extract and prepare the data for analysis.

3.2 Materials

3.2.1 Satellite Images

Satellite image capture was chosen because it was contemplated that there would be a need to consider the relativity of canopy between orchards of different management structure and of different dry matter outputs and that the data for the orchards being compared was best captured under like conditions at the same time to minimise data normalisation.

Several options for satellite image capture were considered. Cost, availability and sample size were all considered as well as the spatial resolution. Images needed to be multispectral so that differences in the core reflectance wavebands of vegetation, R, G, B and NIR, could be determined. Spatial resolution needed to be such that it was meaningful in the context of vine size and because multi-temporal images were contemplated as well as a requirement to minimise weather impact a reasonably high frequency of overpass was necessary. Considering the options (summarised in Appendix 1) the QuickBird satellite option was chosen because of its multispectral resolution, in addition to being available and affordable for the area covered.

Multispectral satellite images were booked and purchased through external agents Terralink International, New Zealand (Terralink) and SKM Consulting, New Zealand and Australia (SKM). A set of co-ordinates centred in the proposed area of study (chosen because of the density of kiwifruit orchards and the availability of ground based data) was provided to the image service providers. The images, centred on coordinates near Te Puke, New Zealand and covering 17 square km, were captured by the QuickBird 2 satellite owned and operated by DigitalGlobe, U.S. The specifications of QuickBird 2 are detailed in Appendix 1. The image service providers ensured all requirements were in place to support processing and geo-referencing of the images, being their area of expertise. The same process was used for each image to ensure consistency. Full details were provided with the images. The captured images were pre-processed and geo-referenced by either Terralink or SKM before being provided. QuickBird 2 products are delivered as radiometrically corrected image pixels. The radiometric correction includes a dark offset subtraction and a non-uniformity correction (detector-to-detector relative gain).

When booking an image, a range of dates had to be provided and there was no guarantee that an image would be captured. Images were requested year on year in the same date ranges however not all requests were met. This meant that the available images and the number of images were generally outside of the control of the author and the study had to use what was available even if less preferred.

In addition, whilst the cost was affordable funds were not unlimited and the study had a finite time frame for completion. The images covered an area representing 1,700 hectares of which approximately 30% was kiwifruit canopy. There were hundreds of individual orchards represented in the area and the 20 test orchards were representative of different locations and orchard management. The locations of the 20 test orchards and the 9 out-of-test orchards are shown on Figure 2.

Thirteen images were purchased of which eight were selected as the core basis of this study. The eight core images were classified as being cloud free. Table 6 provides the technical image details, Table 7 the quick reference details and Table 8 details the spectral characteristics of the QuickBird multispectral wavebands. Figure 2 shows the raw image for 2005 and Appendix 5 shows the raw image for 2009. The area of each image capture was not the same each time, however all images were geo-referenced and could be aligned for orchard location and data extraction.

Table 6: Satellite image technical details

Image*	Acquisition Date and Time **	Spatial Resolution (m per pixel)	# of x pixels	# of y pixels
A	2005-05-12T22:23:09.634357Z	2.730	5021	3178
B	2006-03-14T22:37:51.566397Z	2.560	5354	3379
C	2006-04-14T22:33:34.714991Z	2.891	4660	2943
D	2006-05-15T22:29:59.008835Z	2.917	4609	2951
E	2007-01-17T22:32:47.799212Z	2.907	4992	3470
F	2008-01-25T22:34:18.408284Z	2.718	5440	3735
G	2008-04-11T22:39:21.804878Z	2.756	2880	3948
H	2009-05-02T22:28:17.405598Z	2.532	4672	4090
* All images captured by QuickBird-2; Images A-E procured through Terralink; Images F-H procured through SKM				
**Z (Zulu) time was formerly Greenwich Mean Time (GMT) and is now Coordinated Universal Time (UTC); UTC+12 hours = New Zealand Standard Time (NZST)				

The images captured on different days, months and years were also referenced by the number of days from the beginning of each year (Table 7).

Table 7: Satellite image quick reference details

Image	Days through Year	Month-Year
A	132	May-05
B	73	Mar-06
C	104	Apr-06
D	135	May-06
E	17	Jan-07
F	25	Jan-08
G	102	Apr-08
H	122	May-08

Table 8: Multispectral characteristics of QuickBird-2 satellite channels

Channel	Wavelength (nm)	Effective Bandwidth (nm)
Blue	450-520	68
Green	520-600	99
Red	630-690	71
Near Infrared	760-890	114

3.2.2 Digital Numbers (DNs)

QuickBird 2 has four colour channels and hence, 4 Digital Numbers (DN). Each DN represents the intensity of the signal received by the satellite pixel. QuickBird 2 collects data using an 11-bit dynamic range. This allows 2^{11} or 2048 possible intensity values for each pixel. DigitalGlobe delivered the data in a 16-bit integer format which requires that the original 11 bits of data are simply stored in a 16-bit file. Placeholders are added to account for the 5-bit data difference, but no stretching is performed and no information is lost or gained.

3.2.3 Software

- Geomatica V10 software from PCI Geomatics, Canada was used to visualise the images and cross-reference the data as well as to classify the different features in the area of interest. This software is a commercially available geospatial imaging software suite.

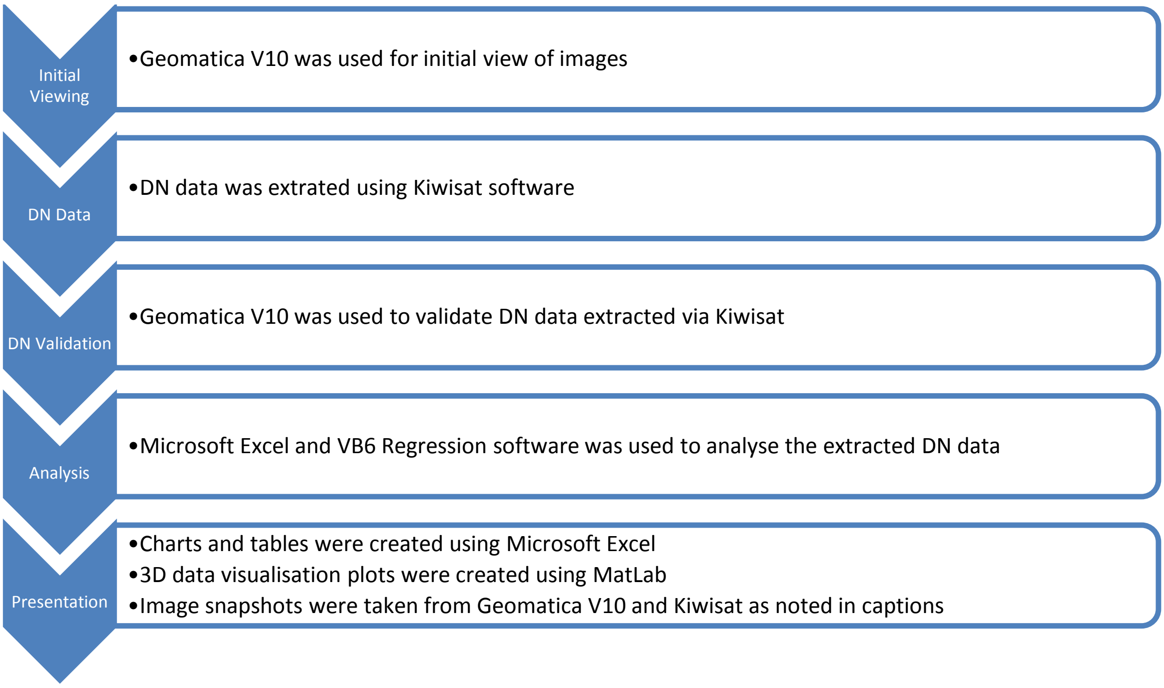
- Software using Microsoft Visual Studio Solution was written to the author’s specifications. This software (Kiwisat) allowed the selection of specific orchard areas and the extraction of the relevant associated DNs for these from each of the images. This software was created because bulk data extracts were not possible from Geomatica V10, although data could be exported for a maximum 198 pixel groups centred on a selected pixel. Kiwisat software was validated against Geomatica V10 extracts to ensure correct alignment of pixels and values before it was extensively used.

- Regression software already written in VB6 was provided as a base by the Supervisor and was used to perform the regression analysis.

- Microsoft Excel was used to analyse, present and graph results. This software is a commercially available

- MatLab software, developed by MathWorks was used to prepare and present the 3 dimensional graphic images. This software is a commercially available.

Figure 16 provides and overview of how the various software solutions were utilised in this study.



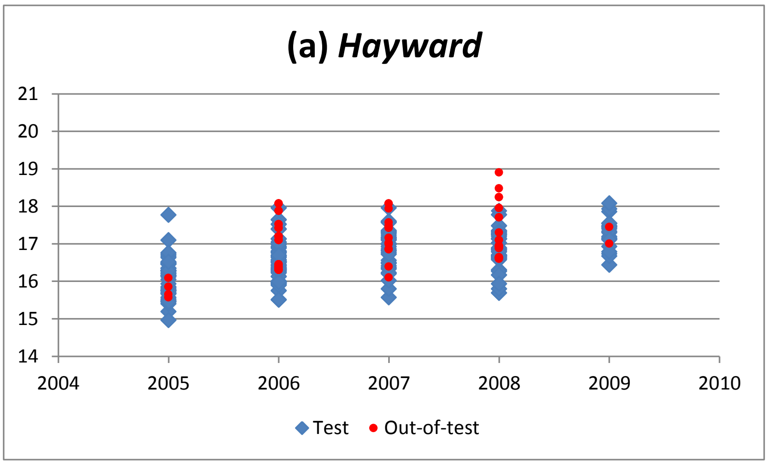
- Geomatica V10 was used for initial view of images
- DN data was extrated using Kiwisat software
- Geomatica V10 was used to validate DN data extracted via Kiwisat
- Microsoft Excel and VB6 Regression software was used to analyse the extracted DN data
- Charts and tables were created using Microsoft Excel
- 3D data visualisation plots were created using MatLab
- Image snapshots were taken from Geomatica V10 and Kiwisat as noted in captions

Figure 16: Overview of software utilisation

3.2.4 Test Orchards

A core group of 123 blocks from 20 test orchards that were common to each of the eight images were selected for study and used to develop predictive models. The criteria to select these orchards were that they covered both varieties of kiwifruit, they covered as wide a range of measured dry matter values as possible and they were physically distributed across the study area. Individual blocks, or combinations thereof were defined by growers and submitted as Maturity Areas in any given year.

Orchards that were used to create a predictive model could not also be used to test the predictive capability of the model when applied in general. This required a second group of orchards that could be used for out-of-sample testing. An additional 9 orchards, which were not used in the model creation, were used to test the predictive capability of the model. These 9 orchards were chosen using the same criteria as the 20 test orchards. The location of each of the test and out-of-test orchards is shown in Figure 2. The range of dry matters for the test and out-of-test orchards for 2005-2009 by variety is shown in Figure 17.



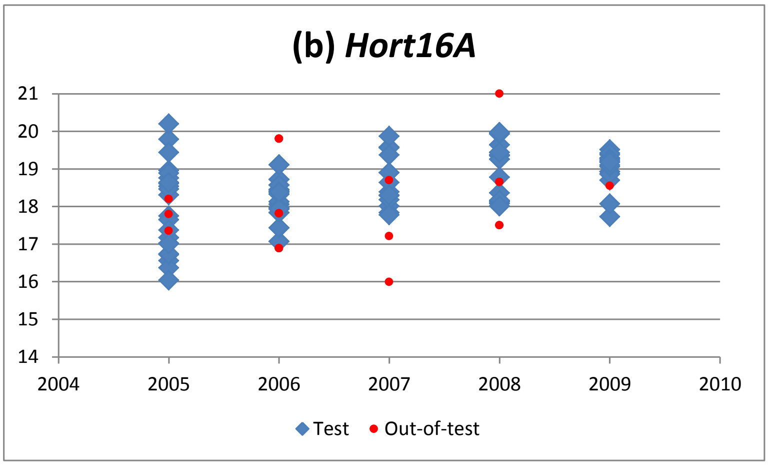


Figure 17: Range of dry matter by variety for test and out-of-test orchards

3.2.4.1 Orchard Maps

Orchards are divided into blocks and blocks are grouped together as a Maturity Area. Each orchard has a map that shows the location, size (hectares), layout and orientation of the orchard including the designation of blocks. Refer Figure 4, Chapter 1 for a sample of an orchard map.

Once a set of test orchards was selected based on the criteria of variety, dry matter range and location orchard maps for the test orchards were obtained from AgFirst BOP, New Zealand. These maps were used to facilitate the location of specific orchards within an image and provided detail on block and Maturity Area level designations so that ground-based sample values could be aligned with associated pixel DN values. Many of the orchards have been GPS mapped but the coordinates were not known. The orchards were located in the images from the physical address of the orchard, cross-referenced using google maps. The blocks and Maturity Areas were then identified from the layouts shown on the orchard map.

Long term the GPS coordinates of the canopy corners for each block could be used provided the geo-referencing was confirmed between the image and the co-ordinate mapping system.

3.2.4.2 Ground Based Orchard Data

3.2.4.2.1 Full Bloom Date

Full Bloom date is the date when 90% of flowers are open or have opened. *Hort16A* kiwifruit vines flower earlier than *Hayward* kiwifruit vines. Full Bloom date is used as a common reference point for sample comparison with respect to maturity development. Table 9 shows the average full bloom date for each variety of kiwifruit in each of the years of study for the test orchards used in this study.

Table 9: Average Full Bloom Date of Test Orchards, 2005 to 2009

	Average Full Bloom Date of Test Orchards	
Harvest Season	***Hayward***	***Hort16A***
2005	26.11.04	29.10.04
2006	23.11.05	20.10.05
2007	22.11.06	18.10.06
2008	24.11.07	22.10.07
2009	25.11.08	20.10.08

Hayward kiwifruit are considered to be fully mature approximately 160 days after Full Bloom and *Hort16A* kiwifruit are considered to be fully mature approximately 200 days after Full Bloom. There will be some variation year on year depending on climatic conditions. Full maturity means the fruit is now capable of natural ripening, the fresh weight of the kiwifruit stops increasing, the seeds are black and the starch starts converting to sugars.

3.2.4.2.2 Average Size of Fruit

As the crop on the vine matures, samples of fruit are taken, weighed and tracked to assist in the prediction of the final size of the fruit at harvest. Fresh weight (fruit growth) accumulates over time in a reasonably consistent manner, although the rate of growth is very climate dependent. Industry-standard fresh-weight-accumulation curves are used to determine the final full-maturity fruit weight based on the weight of the sample and the days after Full Bloom that the sample was taken.

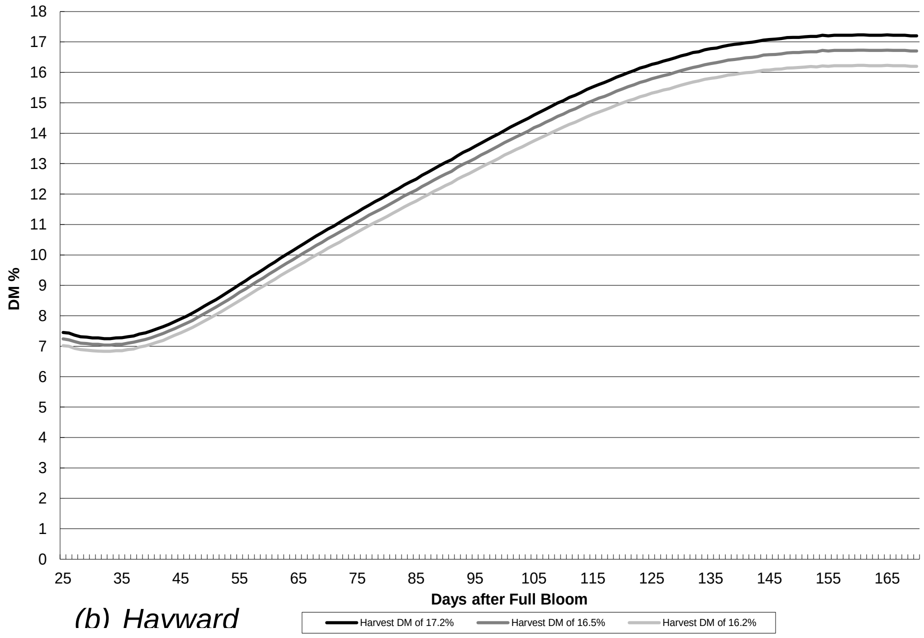
The actual fruit size is only known after it has been graded through the pack-house. Part of the grading process is to weigh the fruit and determine its size as discussed previously (in Section 1.3.1.3 of Chapter 1). The average size of the fruit produced from a Maturity Area is calculated as the volume-weighted average of the sizes packed from the whole Maturity Area.

3.2.4.2.3 Dry Matter Data

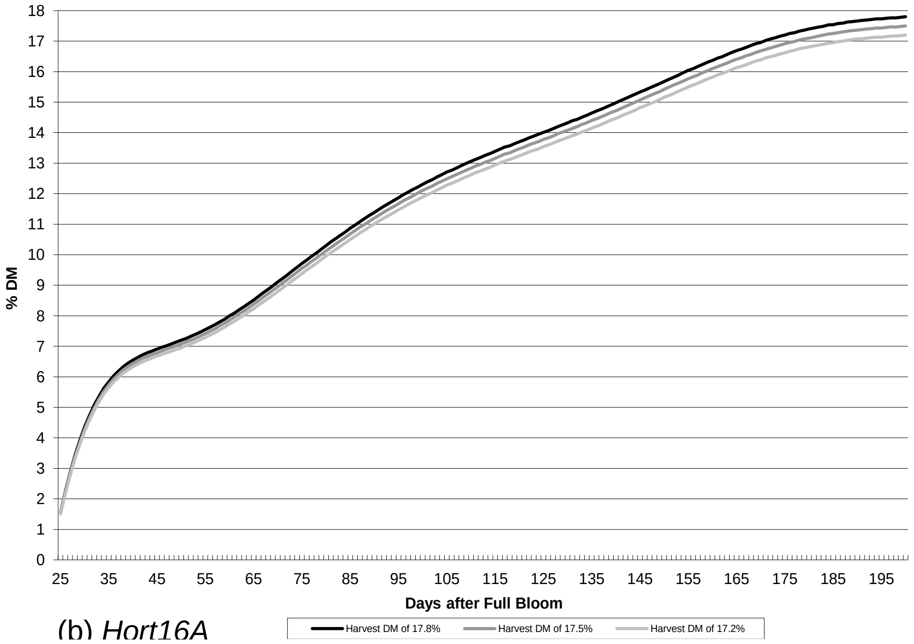
Each Maturity Area has at least one average dry-matter value associated with it, determined from a 90-fruit sample prior to harvest. Several 90-fruit samples may be tested over time so that a grower can monitor dry-matter accumulation and make a decision as to when to harvest. The collection date and the days from Full Bloom are known for each sample.

There is a standardised protocol for selecting a 90-fruit sample from a maturity area. AgFirst BOP, New Zealand is an independent contractor to the industry and is responsible for sampling all maturity areas and calculating the average dry matter value of each sample. Dry-matter data was obtained from AgFirst BOP in an electronic format. This data was associated with each of the specific orchard maturity areas being analysed. The standard deviation of these samples ranges from 0.52 to 1.83 with an average of 1.05 for a dry-matter range of 14 to 20. This imprecision is significant and important in view of our later predictions. The current 90-fruit test is the basis for the dry matter payments to growers. It is known that the more fruit sampled will increase the accuracy of the test but a 90-fruit sample was deemed an appropriate compromise between the practicalities of the test (physical collection, destructive testing and manual process) and the accuracy of the results, with all growers exposed to the same level of accuracy. Many growers will often have more than 1 test taken and payment was made on the highest test achieved. The standard deviation spread reflects the degree of variability of dry matter percentage of fruit within a vine, between vines and within a designated maturity area.

Dry matter accumulates over time in a reasonably consistent manner. Industry dry-matter-accumulation curves were obtained from AgFirst BOP. Figure 18a and 18b shows the curves for *Hayward* and *Hort16A* respectively for 2005. These curves are used to determine the full-maturity dry matter of the fruit based on the dry matter of samples at the time the sample was taken at a known number of days after Full Bloom.



(b) Hayward



(b) *Hort16A*

Figure 18 (a)-(b): 2005 dry-matter-accumulation curves for kiwifruit; (a) Green, (b) Gold

The harvest criteria for a limited volume of fruit may be set at a lower level than if it was fully mature (but still at such a level that it will continue to mature and ripen off the vine), therefore some fruit may be harvested before the fruit is fully mature and so has not yet finished accumulating dry matter.

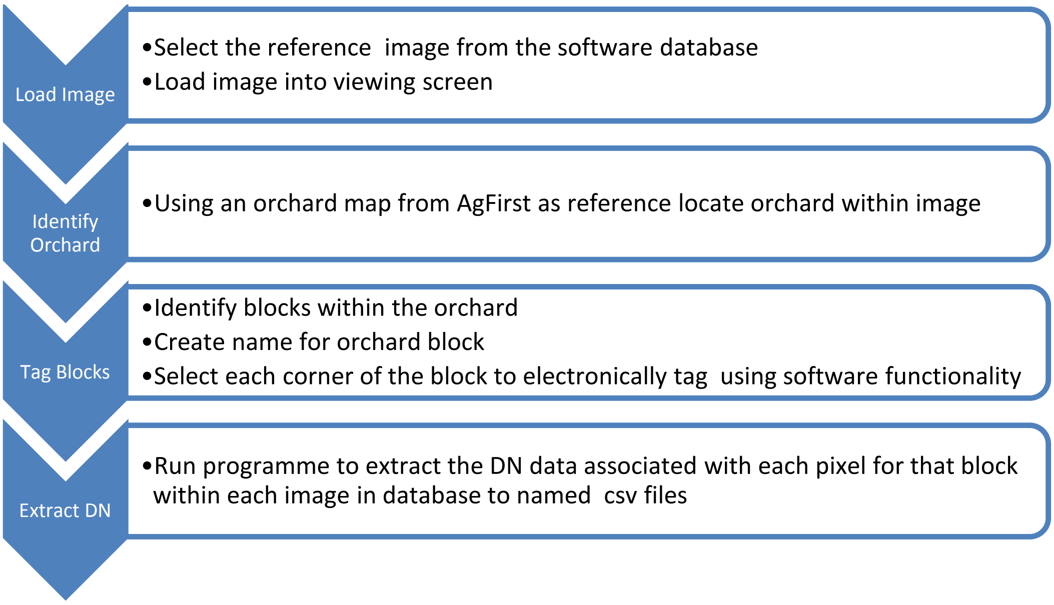
Images are not necessarily captured at the same time as a dry-matter sample is measured and a dry-matter sample is not necessarily taken at full maturity. When an area is in full bloom with the majority of flowers fully open, it is considered to be the start of dry-matter accumulation. The final accumulated or ‘full-maturity’ dry matter for each test Maturity Areas was calculated from the rate of accumulation determined from the dry-matter-accumulation curves (Figure 18).

3.3 Methodology for DN Extraction from images

A Maturity Area is unique and has multiple pixels that can be associated with one average ground-based sample. The 20 test orchards were visually identified in the first image using the orchard maps. Orchard block designations do not change from year to year but the assignment of blocks to Maturity Areas does. All DN data was extracted from the images at block level so that blocks and associated DNs could be recombined as required to match the Maturity Area designation and associated ground based sample data. (Refer to Figure 3 of Chapter 1).

3.3.1 Kiwisat DN Extraction

The flow chart in Figure 19 provides an overview of the DN data extraction process.



- Select the reference image from the software database
- Load image into viewing screen

- Using an orchard map from AgFirst as reference locate orchard within image

- Identify blocks within the orchard
- Create name for orchard block
- Select each corner of the block to electronically tag using software functionality

- Run programme to extract the DN data associated with each pixel for that block within each image in database to named csv files

Figure 19: Process flow for DN data extraction using Kiwisat software

Using the Kiwisat software the corner of each block for the 20 test orchards was electronically tagged and named. A sample of tagged blocks is shown in Figure 20, with the brightness altered to enhance the corner tags.

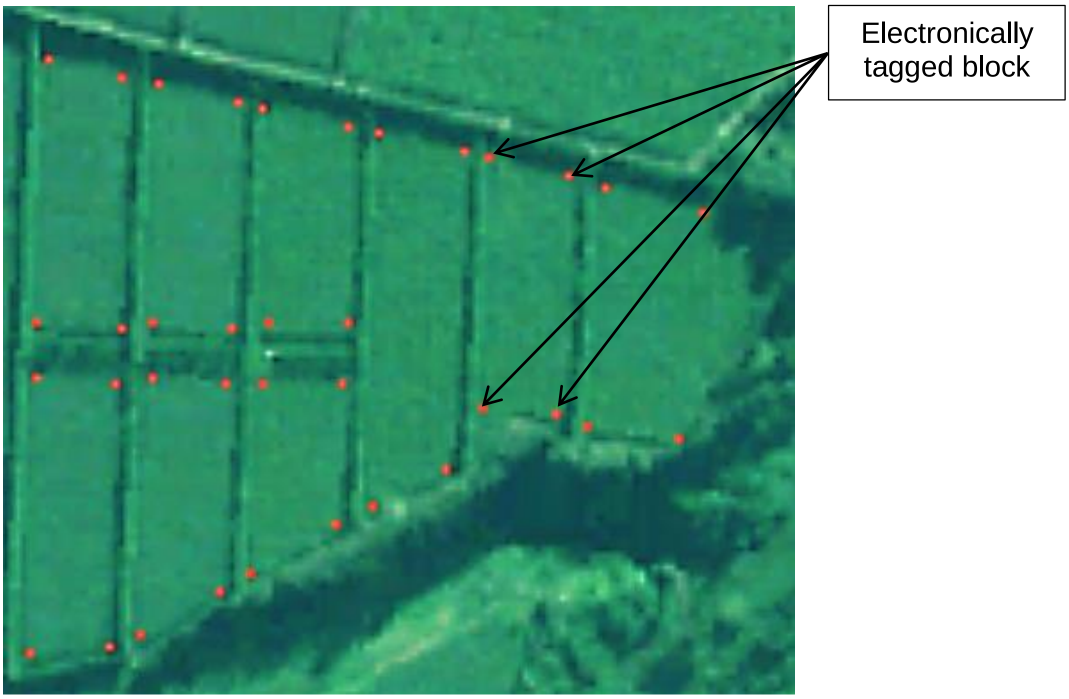


Figure 20: Electronically tagged kiwifruit orchard blocks using Kiwisat software, Orchard #2, May 2005 image, block 5 highlighted

The corners of the blocks were tagged as close to the edges of the canopy as possible. As noted in Section 1.3, one or more blocks can be designated to define a Maturity Area and a Maturity Area is such that all of the fruit within the area will reach maturity and be able to be harvested at the same time. Subject to the results of maturity testing the Maturity Area could be redefined. The 90-fruit sample is taken randomly across the Maturity Area, (but not from the edge rows of a block as other research studies have shown fruit characteristics are more variable in these regions), regardless of how many blocks are included and the total size of the Maturity Area itself (although there is a maximum size allowed for a Maturity Area). This 90-fruit sample is used to determine the maturity of the fruit and the dry matter of the fruit within that area. As noted by the standard deviation range for dry matter values in Section 3.2.4.2.3 there is a level of dry matter variability within a Maturity Area.

Checks were done on several blocks to determine the impact on the average DN values if the outer edge pixel values were excluded from the data set. It was found that the average was changed by less than 1DN.for R, G and B bands and less than 20DN for the NIR band. In addition when tagging the blocks it was erred on the side of caution to be slightly inside the outer canopy edge rather than outside.

An algorithm was then developed that extracted the DN data for all the pixels within each tagged block into an Excel spreadsheet. Even though Maturity Areas are supposedly as homogeneous as possible it is known there is variability within the canopy cover. Outlier values (greater than 3.5 deviations) were excluded from any calculations. Data from all blocks designated to form part of a Maturity Area in any given year were combined to calculate the average DN values associated with each Maturity Area. It is expected that this study which considers the total canopy of the Maturity Area could detect variability within the canopy which could then be used to re-define Maturity Areas to be more homogenous in nature.

A sample of the extract for the block shown in Figure 20 is presented in Table 10.

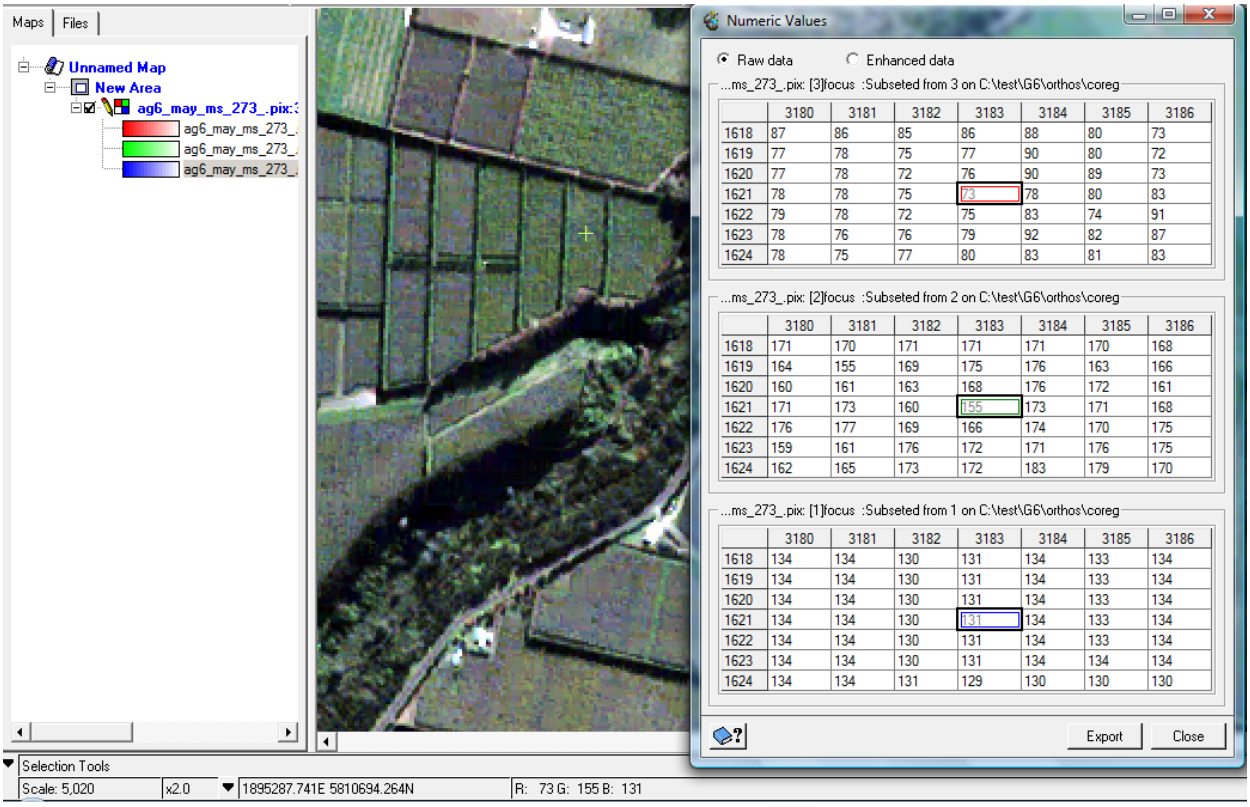
Table 10: Raw DN values extracted from kiwifruit block 5, Orchard #2 using Kiwisat software, May 2005 image

Block	Image	Pixel x	Pixel y	Red	Green	Blue	NIR
Orchard #2, block 5	ag6_may_ms_273	3183	1618	86	171	131	566
Orchard #2, block 5	ag6_may_ms_273	3183	1619	77	175	131	554
Orchard #2, block 5	ag6_may_ms_273	3183	1620	76	168	131	567
Orchard #2, block 5	ag6_may_ms_273	3183	1621	73	155	131	527
Orchard #2, block 5	ag6_may_ms_273	3183	1622	75	166	131	548
Orchard #2, block 5	ag6_may_ms_273	3183	1623	79	172	131	579
Orchard #2, block 5	ag6_may_ms_273	3183	1624	80	172	129	574

Once the blocks were tagged in the first image, the software extracted the data for each block in each of the other images (correcting for spatial resolution variances and any offsets between images).

3.3.2 Geomatica validation

Geomatica software was used to validate that the Kiwisat software was extracting and storing the data correctly as represented by the data in Figure 21 matching that in Table 10.



● Raw data ○ Enhanced data

...ms_273_.pix: [3]focus :Subseted from 3 on C:\test\G6\orthos\coreg

	3180	3181	3182	3183	3184	3185	3186
1618	87	86	85	86	88	80	73
1619	77	78	75	77	90	80	72
1620	77	78	72	76	90	89	73
1621	78	78	75	73	78	80	83
1622	79	78	72	75	83	74	91
1623	78	76	76	79	92	82	87
1624	78	75	77	80	83	81	83

...ms_273_.pix: [2]focus :Subseted from 2 on C:\test\G6\orthos\coreg

	3180	3181	3182	3183	3184	3185	3186
1618	171	170	171	171	171	170	168
1619	164	155	169	175	176	163	166
1620	160	161	163	168	176	172	161
1621	171	173	160	155	173	171	168
1622	176	177	169	166	174	170	175
1623	159	161	176	172	171	176	175
1624	162	165	173	172	183	179	170

...ms_273_.pix: [1]focus :Subseted from 1 on C:\test\G6\orthos\coreg

	3180	3181	3182	3183	3184	3185	3186
1618	134	134	130	131	134	133	134
1619	134	134	130	131	134	133	134
1620	134	134	130	131	134	133	134
1621	134	134	130	131	134	133	134
1622	134	134	130	131	134	133	134
1623	134	134	130	131	134	134	134
1624	134	134	131	129	130	130	130

Scale: 5,020 | x2.0 | 1895287.741E 5810694.264N | R: 73 G: 155 B: 131

Figure 21: Validation of Kiwisat DN extraction using Geomatica software

3.4 Data Preparation

A key issue when analysing multi-temporal images is the correction of atmospheric variance between images. The images being analysed in this study are technically cloud free although that does not mean the atmospheric conditions are identical.

3.4.1 Top-of-Atmosphere DNs

The QuickBird 2 data has been radiometrically corrected, including a dark offset subtraction and a non-uniformity correction (detector-to-detector relative gain). QuickBird 2 image files contain the absolute radiometric calibration factor (absCalFactor) [W-m^{-2}-sr^{-1}-count^{-1}] for a given band. This absCalFactor was used to calculate the top-of-atmosphere band-integrated radiance image pixels ($\mathrm{L_{Pixel,Band}}$) [W-m^{-2}-sr^{-1}].

$$\mathrm{L_{Pixel,Band}} = \mathrm{absCalFactor_{Band}} \times \mathrm{q_{Pixel,Band}}$$

where:

$\mathrm{L_{Pixel,Band}}$ are top-of-atmosphere band-integrated radiance image pixels [W-m^{-2}-sr^{-1}].

$\mathrm{absCalFactor_{Band}}$ is the absolute radiometric calibration factor [W-m^{-2}-sr^{-1}-count^{-1}] for a given band.

$\mathrm{q_{Pixel,Band}}$ are the radiometrically corrected image pixels (DN). The radiometric correction includes a dark offset subtraction and a non-uniformity correction (detector-to-detector relative gain), and:

The absCalFactors for QuickBird are as follows:

$\mathrm{Blue} = 0.01604120$ [W-m^{-2}-sr^{-1}-count^{-1}]

$\mathrm{Green} = 0.01438470$ [W-m^{-2}-sr^{-1}-count^{-1}]

$\mathrm{Red} = 0.01267350$ [W-m^{-2}-sr^{-1}-count^{-1}]

$\mathrm{NIR} = 0.01542420$ [W-m^{-2}-sr^{-1}-count^{-1}]

The Red (R), Green (G), Blue (B) and Near-infrared (NIR) DN's for each pixel were multiplied by the respective R, G, B and NIR absCalFactors to calculate the top-of-atmosphere band-integrated radiance image pixels($L_{\mathrm{Pixel,Band}}$).These top-of-atmosphere adjusted pixel DN values were used for all further analysis.

3.4.2 Atmospheric Attenuation

As shown in Figure 10, and as used for the basis of vegetative remote sensing described in Chapter 2, the reflectance characteristics of vegetation changes under different conditions such as canopy density and leaf health. Change in reflectance between images could be due to many factors. Section 2.5 recognises that atmospheric interference is a significant issue in remote sensing, especially when using multi-temporal data. It is important that objects with minimal change in reflectance over time are used to correct for atmospheric interference. There are objects other than plants that are less likely to change their reflectance characteristics because the composition of the objects does not change. Such objects, known as fiduciary objects, can therefore be used to detect changes caused by factors other than the changing composition of the object, such as atmospheric differences. Roads and rooves were chosen to be used in this study as these are easily identifiable in the images.

Ten objects common to each image (Table 11) were used as fiduciary markers for atmospheric attenuation. The reflectance characteristics of these fiduciary markers are not expected to change over time; that is, they should have the same colour vector regardless of when the image was captured and any variance would be due to atmospheric interference, light source intensity and angle of light source relative to satellite sensor angle. Given the multi-spectral aspect of the study rooves of different colours were also used to enable comparisons between methods of atmospheric attenuation.

Table 11: Fiducial marker details

Marker Number	Fiduciary Marker	Marker Colour
1	Roof 1	Blue
2	Roof 2	Blue
3	Roof 3	Orange
4	Roof 4	White
5	Roof 5	White
6	Roof 6	White
7	Road 1	Grey
8	Road 2	Grey
9	Road 3	Grey
10	Road 4	Grey

The intensity of the light source, and thus the brightness of the image, do need to be considered as multi-temporal images are being compared. For each image solar radiation of varying brightness is attenuated to an unknown, wavelength-dependant, degree as the light passes through the atmosphere before striking the canopy. Light originating from the canopy, and of a different colour is then attenuated as it passes up through the atmosphere, to be captured by the satellite.

We elect to categorise this attenuation by four coefficients – one for each of our four wavelengths. The reading from each pixel of the camera retina is represented by a vector in four-space, (Red (R), Green (G), Blue (B) and Near-infrared (N)). It is important to note that the length of each such vector reflects the number of photons captured by the pixel and is determined merely by the amount of light, not by its colour.

Accordingly, it is convenient to reduce the four dimensional vector to a unit vector. By doing this, we can specify the complete vector with only three coordinates, say R, G, B. The fourth component, NIR, can then be computed from these three.

The length of the vector, r, is defined by the following equation:

$$\mathrm{r} = \sqrt{(\mathrm{R}^2 + \mathrm{G}^2 + \mathrm{B}^2 + \mathrm{N}^2)}$$

The unit vector is defined by:

$$\mathrm{u} = \frac{\mathrm{R}}{\mathrm{r}}, \frac{\mathrm{G}}{\mathrm{r}}, \frac{\mathrm{B}}{\mathrm{r}}, \frac{\mathrm{N}}{\mathrm{r}}$$

And to check that u is in fact a unit vector the following equation should equal 1.

$$\left(\frac{\mathrm{R}}{\mathrm{r}}\right)^2 + \left(\frac{\mathrm{G}}{\mathrm{r}}\right)^2 + \left(\frac{\mathrm{B}}{\mathrm{r}}\right)^2 + \left(\frac{\mathrm{N}}{\mathrm{r}}\right)^2 = 1$$

The un-attenuated, average, top-of-atmosphere pixel value for each fiducial marker in each image was converted to a unit vector and plotted for each band, Figure 22.

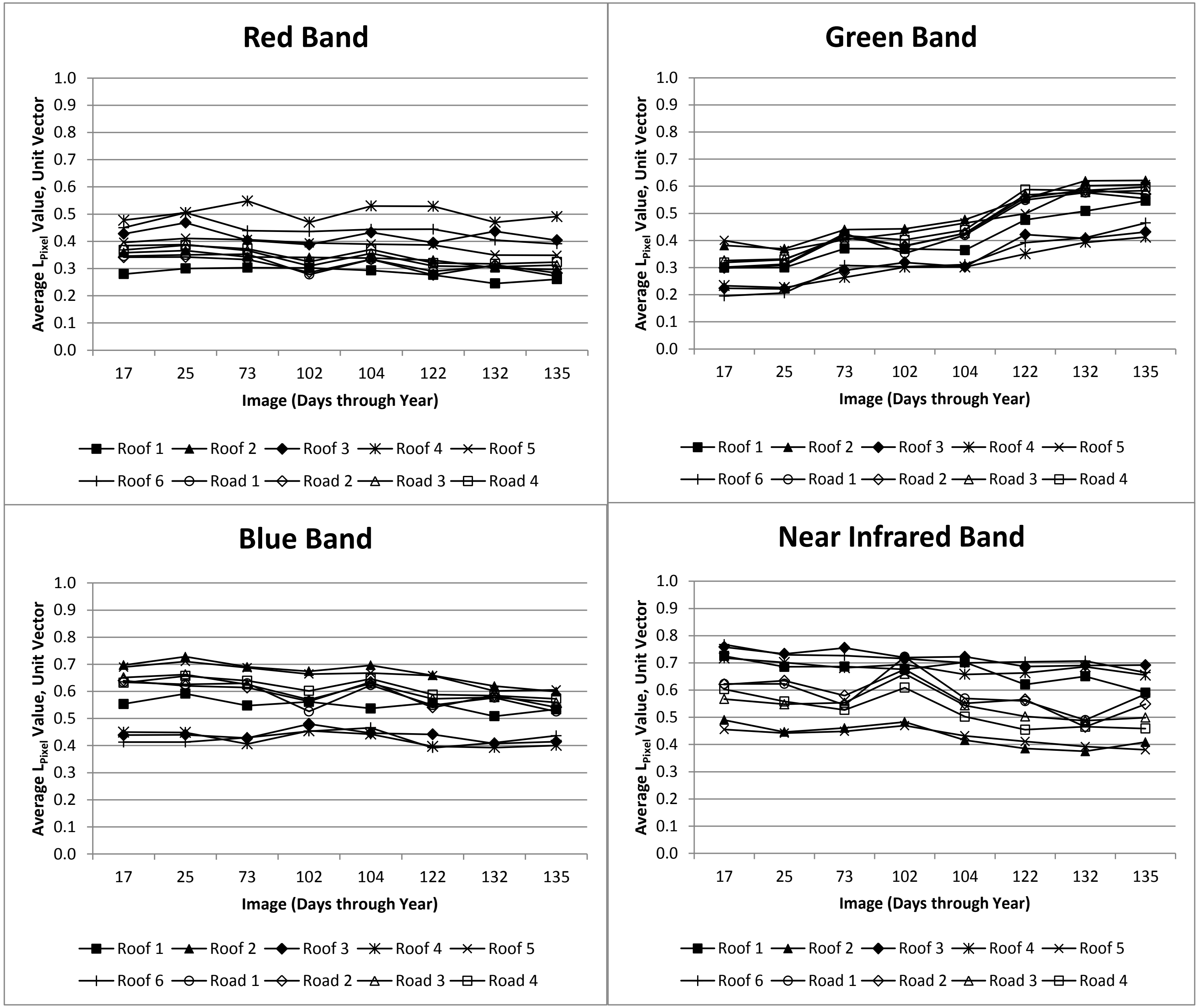


Figure 22: Un-attenuated average $L_{Pixel,Band}$ top-of atmosphere unit vector values for each fiducial marker for each image

It is not practical, or always possible, to obtain all of the appropriate reference readings to calculate the atmospheric attenuation of each of the images. Such readings were not available for this study. However, given that the study needs to compare the relativity of information within and between images, attenuation methods that normalize the images relative to each other were investigated.

Figure 22, the un-attenuated values, shows the variation between the different fiducial objects in different images captured at different times of the year and over multiple years. The DN data was extracted from an area of each fiducial object that was comprised of many pixels and it is the average of these pixels for each object that has been plotted. The different fiducial objects were of different colours captured under different atmospheric conditions and with varying levels of light intensity. Plotting as unit vectors normalises for the brightness effects from varying light intensity. The intention of atmospheric attenuation is to smooth the values for each fiduciary object over time and across images to minimise the variability due to factors not related to the actual state of the object. Each object would not end up with the same value but would have a value based on its reflective state, and this value would be expected to be the same for the same object in any given image on any given day.

Two methods of relative attenuation were tested:

> Method 1 followed the methodology of Hall *et al*., (1991), to transform each image relative to one image designated the reference image.
>
> Method 2 assumes a linear relationship between the LPixel,Band top-of-atmosphere values for each object in each image relative to one image designated the reference image, similar to the method of Song *et al*., (2001).

3.4.2.1 Method 1: Atmospheric Attenuation per Hall *et al*., 1991

The radiometric rectification method of Hall *et al*., 1991 was followed. This method compensates for sensor calibration, atmospheric differences and illumination differences between images using a linear transformation with coefficients calculated to equate the individual band means of the radiometric control sets in each image. This method uses two radiometric control sets, one bright and one dark. Fiducial marker 5 (Roof 5, White) was used as the bright control set and fiducial marker 8 (Road 2, Grey) was used as the dark control set. Image A (May-05, 132 Days through Year), was used as the reference image.

The top-of-atmosphere pixel values for Fiducial marker 5 and Fiducial marker 8 were transformed for each sample image relative to the reference image, based on the following formula:

$$\text{Transformed } x_i = m_i x_i + b_i$$

where

$$m_i = \frac{(B_{Ri} - D_{Ri})}{(B_{Si} - D_{Si})}$$

$$b_i = \frac{(D_{Ri} B_{Si} - D_{Si} B_{Ri})}{(B_{Si} - D_{Si})}$$

and

x = digital number of reflectance; i = band

And:

B=bright fiducial marker set 5; D=dark fiducial marker set 8; R=reference Image A; S=subject Image B-H respectively.

The coefficients m_i and b_i, which were determined from the transformations for each band and each image, refer Table 12, were then used to transform the L_{Pixel} values for the remaining eight fiducial markers that were not used to calculate the coefficients. Successful attenuation should result in the average $L_{Pixel,Band}$ values for each fiducial marker in all images being the same and equal to the respective average $L_{Pixel,Band}$ values in the reference image. The corrected average $L_{Pixel,Band}$ top-of-atmosphere values were converted to unit vector values for each marker in each image and were plotted for each band, Figure 23. As expected the average $L_{Pixel,Band}$ values for fiducial markers Roof 5 and Road 2, respectively, are equal in all images and equal to the respective values of the reference image.

Table 12: Coefficients determined for atmospheric attenuation Method 1

	Coefficient m				**Coefficient b**			
Image	**R**	**G**	**B**	**N**	**R**	**G**	**B**	**N**
A	1.0000	1.0000	1.0000	1.0000	-	-	-	-
B	0.5242	0.5264	0.5440	0.5871	0.4749	0.9157	0.7602	0.0350
C	0.6601	0.6665	0.6851	0.7294	0.2465	0.3851	0.2258	- 0.1974
D	0.9428	0.9304	0.9279	1.2847	0.3893	0.7065	0.5073	- 1.1743
E	0.6811	0.6740	0.6908	0.9392	- 0.3000	- 0.5215	- 0.4497	- 2.7008
F	0.5467	0.5286	0.5420	0.7957	0.1688	0.5216	0.4217	- 1.8646
G	0.5581	0.5827	0.6206	0.7547	0.5816	0.8419	0.5460	- 1.4110
H	0.6147	0.6154	0.6573	0.7295	0.9648	1.7234	1.2750	0.5393

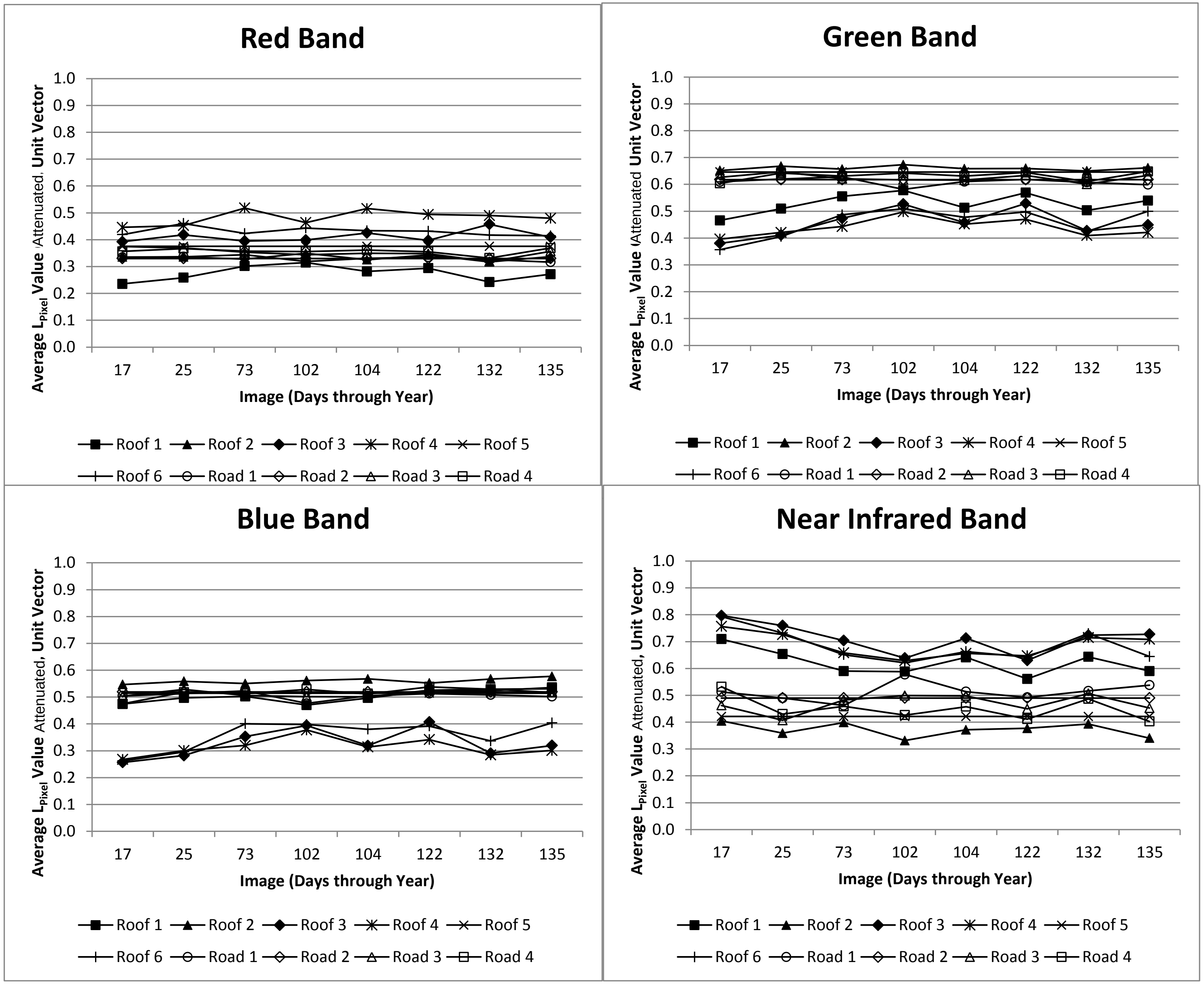


Figure 23: Attenuated average $L_{Pixel,Band}$ top-of-atmosphere unit vector values for each fiducial marker for each image using Method 1

3.4.2.2 Method 2: Linear Attenuation

Image A was used as the reference image against which all others were attenuated in a relative manner. This method is similar to that proposed by Song *et al*., 2001.

The average top-of-atmosphere pixel values for each fiducial marker for each band were plotted for Images B, C, D, E, F, G, H against Image A respectively and the coefficients were determined from the linear regression trend-line equation (Table 13).

Table 13: Attenuation coefficients determined from linear regression trend-lines

		Image							
		May-05	**Mar-06**	**Apr-06**	**May-06**	**Jan-07**	**Jan-08**	**Apr-08**	**May-09**
	Coefficient	**A (132)**	**B (73)**	**C (104)**	**D (135)**	**E (17)**	**F (25)**	**G (102)**	**H (122)**
RED	α	1.0000	0.6420	0.7907	1.0773	0.7370	0.5783	0.7789	0.7645
	β	-	- 0.0132	- 0.4318	- 0.0543	- 1.1634	- 0.5101	- 0.3049	0.4201
GREEN	α	1.0000	0.5313	0.7118	0.9689	0.6789	0.5191	0.6283	0.6506
	β	-	0.7659	- 0.1871	0.2636	- 1.2884	- 0.1231	0.2083	1.1461
BLUE	α	1.0000	0.5580	0.7346	0.9348	0.6074	0.4856	0.6914	0.7009
	β	-	0.6594	- 0.1864	0.2679	- 0.4117	0.2868	0.0148	0.8089
NIR	α	1.0000	0.8584	0.9411	1.2903	0.6106	0.5873	1.1387	0.9967
	β	-	- 1.3310	- 1.4025	- 1.0882	- 1.0771	- 0.7248	- 3.7433	- 0.3596

All average $L_{Pixel,Band}$ values for all objects, in each image and each band were then attenuated using the formula:

$$y = \alpha x + ß$$

Where

y is the attenuated L_{Pixel} value
x is the L_{Pixel} value before attenuation
α is the slope coefficient
ß is the intercept coefficient

The results of the attenuation using Method 2 are shown in Figure 24.

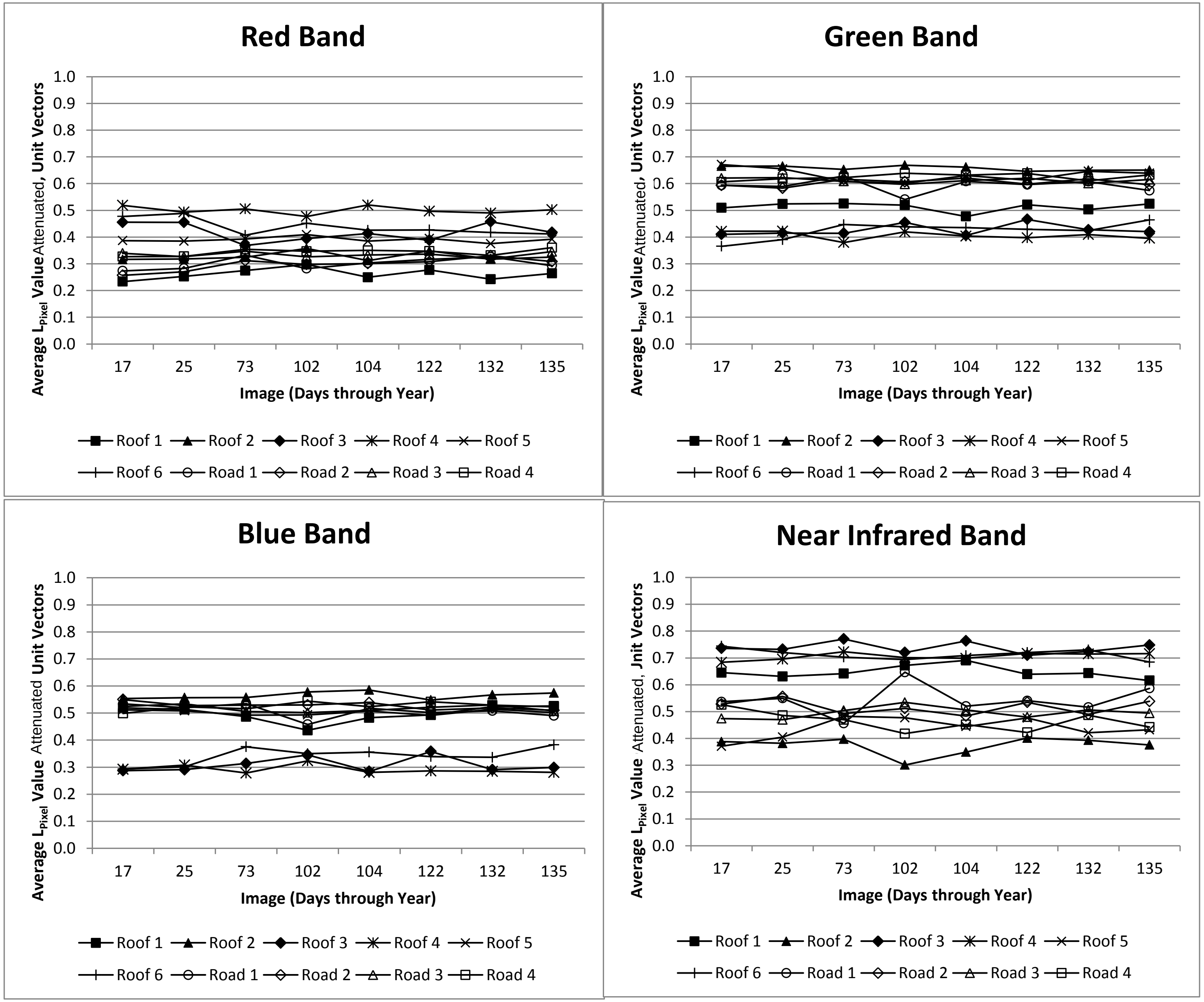


Figure 24: Attenuated average $\mathbf{L}_{\mathbf{Pixel,Band}}$ top-of-atmosphere unit vector values for each fiduciary marker for each image using Method 2

3.4.3 Evaluation of Attenuation Methods

The average and standard deviation of the top-of-atmosphere unit vector values for all of the fiduciary objects for each wave band across all images were calculated to evaluate the attenuation methods tested. The results of these calculations are shown in Table 14.

Table 14: Comparison of attenuation methods by evaluating standard deviations

		R	**G**	**B**	**N**
Un-attenuated	**Average**	0.3902	0.4352	0.4785	0.6211
	Standard Deviation	0.0759	0.1186	0.1087	0.1117
Method 1	**Average**	0.3678	0.5670	0.4634	0.5421
	Standard Deviation	0.0592	0.0882	0.0918	0.1237
Method 2	**Average**	0.3613	0.5494	0.4585	0.5637
	Standard Deviation	0.0733	0.0950	0.0994	0.1264

The expectation is that atmospheric attenuation of the images would result in least variance of each fiduciary object for each wave band across all images. This would result in the lowest standard deviation. On this basis Method 1 provided the best attenuation except for the near-infrared band and was used in this study.

CHAPTER 4 DATA MODEL

4.1 Introduction

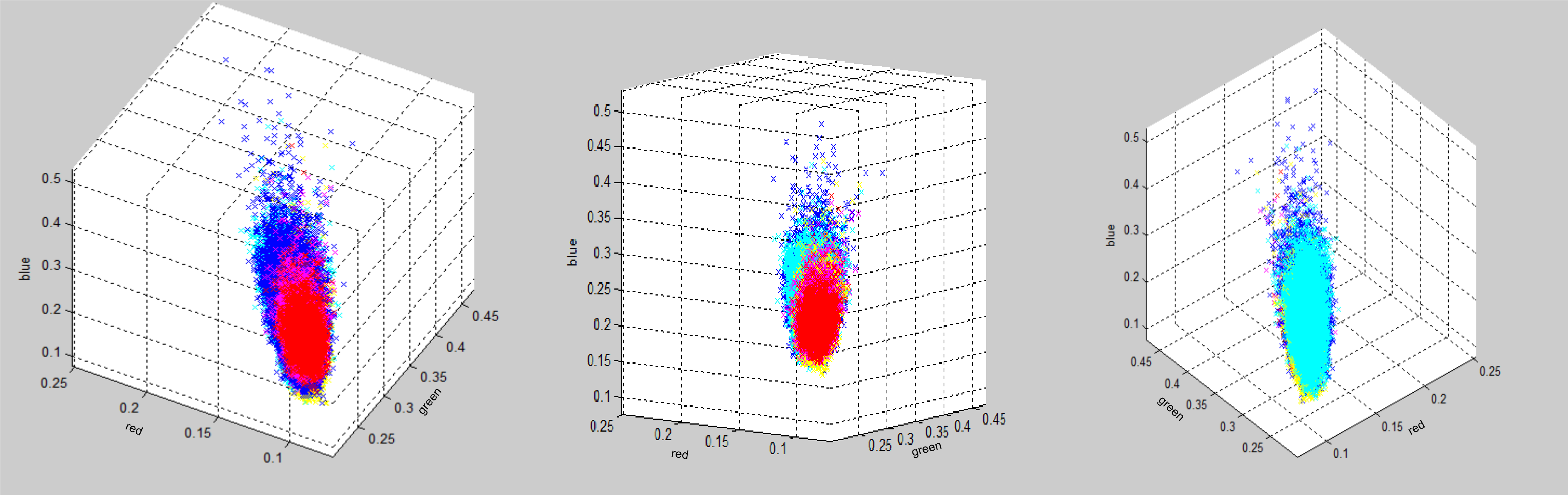
This study seeks to find a process that will lead to a predictive method for DM based on the red, green, blue and near infrared (R, G, B, N) values for each pixel. These colour emissions should be affected by the physiology of the leaves as the plant develops through the growing season. The question is how best to correlate the colour values to dry matter. The problem is even more complex because this study aims to predict the dry matter of the mature fruit before the kiwifruit is of a size to measure using current methods.

In trying to predict crop qualities from remotely sensed images, researchers have come up with many correlations, twenty of which are laid out in Appendix 3. It seems that the methodology of developing these correlations is somewhat serendipitous, as it is hard to imagine any systematic process that might have led to them. Many such developments of correlative techniques are informed by known functional relationships between the quantities measured. Without this, it seems that the logical development of a multi-dimensional correlation is best done on the basis of visualization. Certainly techniques such as principal component analysis, dimensional analysis and regression techniques can be done ‘blind’ but, as we consider Appendix 3, it seems that a visualization technique is desirable. This visualization approach is offered in this study.

As dry matter correlates with four-dimensional qualities, R, G, B and NIR, we cannot directly visualize the data that has been collected. Corresponding to each pixel vector there is a measured value for dry matter. However we have observed that it is the ‘colour’ of the light reflected from the leaves that is of interest. The brightness depends on the number of incident photons but we can declare that we require a unit vector of the four components to represent the ‘colour’ and that this vector is independent of the number of photons striking the canopy and being reflected from it. The ‘colour’ of this unit vector will be affected by atmospheric attenuation but steps have been taken to remove this effect from the measurements. This was described in Section 3.4.2.

Given the length of a unit vector is 1 then the simplification that any three of the coordinates will define the fourth cordinate in a unit vector can also be proposed. Therefore, the data can be plotted in three-space, for example R, G, B, and no information will be lost. If each point is plotted according to a colour scheme reflecting the dry matter, it should be seen that the cloud of points has shape and that there is systematic distribution of the dry matter in this shape. This may not be the case and it might require that each axis should be subject to some mathematical transformation in order to bring out such an underlying structure. Historically, many such analyses have postulated functional forms and then assumed that the functions are unity, this seems the prudent first step.

The core Maturity Areas for Image A (n=65) were grouped by their average Measured dry-matter values. All of the pixels of each group, totalling 93,277 data points across the maturity areas, were plotted in 3D. When such a plot is considered, Figure 25, it seems that there is indeed a structure and systematic distribution.



Symbol	X	X	X	X	X
Measured Dry Matter	>14-16	>16-17	>17-18	>18-19	>19-21
Number of Maturity Areas	18	28	8	8	3
Number of Pixels	21,656	53,576	6,978	7,077	3,990

Figure 25: Three views of a 3D plot of the unit vectors of pixels for the core Maturity Areas for Image A, May 2005

The results displayed in Figure 25 reveal clusters of data points that formed layers within a cloud-like shape that appeared to have a relationship with dry matter level. This shape did in fact seem to be apparent without further mathematical transformation. What was now required was a means by which to describe the shape. The shape could be rotated within the MatLab software but the sensitivity of the rotation did not enable better images than those represented in Figure 25.

The data representation by dry matter in Figure 25, depicted by symbol colour, looked to be stacked 'planes' of symbols of same colour depiction associated with a level of dry matter. This is stylised in Figure 26. The relationship also looked as though it might be linear.

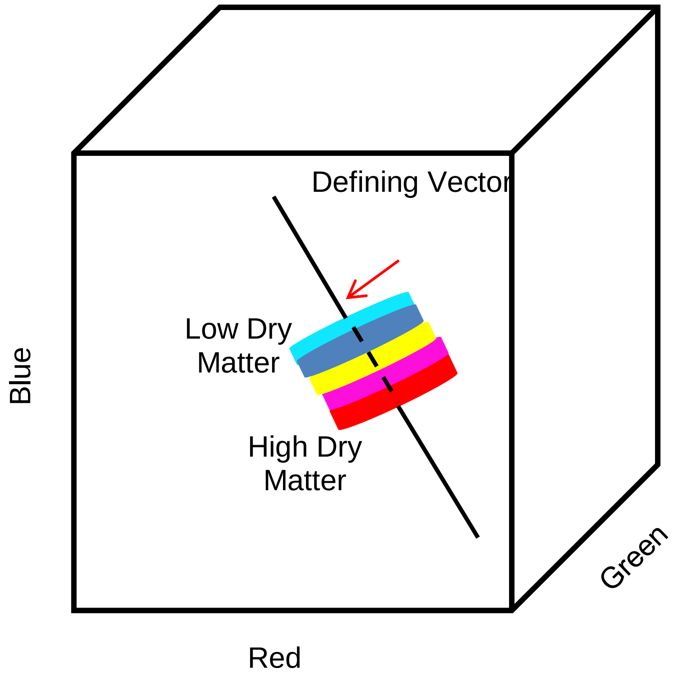


Figure 26: Representation of the data structure in 3D

As a starting point to describe the data structure it could be seen from Figure 25 and Figure 26 that there could exist a vector that defines the axis of the cloud, such that distance along this defining vector correlates with dry matter. Imagine the data structure as a cylindrical loaf of sliced bread with each slice being the layer of data points associated with a particular level of dry matter where there is a defining vector that passes along the axis of this loaf. The first slice corresponds to data points having the lowest dry matter and the last slice corresponds to data points having the highest dry matter. The ideal outcome would be that the defining vector does exist and prediction of dry matter is possible using the visualization theory. In practice, the actual measured dry matter has an average standard deviation of 1.05, and as the basis from which to determine the predictive correlation, it is very noisy.

As the simplest possibility, and seemingly supported by the data representation in Figure 25, let us imagine that $DM = \alpha X + \beta$, where X is the distance along the defining vector from the start of the defining vector (as a reference point). A quadratic relationship can be considered in due course if required.

Two planes, (0, G, B) and (1, G, B), are defined. The defining vector, anchored on the axis of the loaf, must pass through each of these planes. The intersection with the first plane will be at S, the start of the defining vector, with coordinates $(0, G_1, B_1)$. The intersection with the second plane will be at E, the end of the defining vector, with coordinates $(1, G_2, B_2)$. This leaves four unknown quantities (G_1, B_1, G_2, B_2) that define the vector that need to be determined.

An objective function can be described that can be used in a simplex regression with the problem presenting itself as a forward calculation. This calculation requires that the four values associated with the defining vector (G_1, B_1, G_2, B_2) are known, as well as the (R, G, B) coordinates for each data point, P, in order to determine X, the distance along the defining vector.

Firstly, a Point Q, with coordinates (R_3, G_3, B_3) must be sought on the defining vector corresponding to its closest approach to Point P with coordinates (R, G, B). That is to say, the line joining this point Q to point P is orthogonal to the vector. We then define X as the distance from point Q to point S with coordinates $(0, G_1, B_1)$ which we recall is the point where the defining vector intersects the first plane. The calculation also requires the two coefficients (α, β) to be known to allow dry matter to be predicted from the linear formula $DM = \alpha X + \beta$.

The measured value of dry matter is known and can be used to determine the sum-of-squares error between Measured dry matter and Predicted dry matter for the cohort of data points. This is the objective function. In the nature of simplex regressions, the six parameters $(G_1, B_1, G_2, B_2, \alpha, \beta)$ are systematically adjusted to provide the lowest possible objective function.

4.2 Regression Analysis Formulae

Figure 27 illustrates the various points and coordinates associated with the regression calculations. The shortest distance to the line is where the line formed from data point P (R, G, B) intersects with the defining vector at a 90° angle at point Q (R_3, G_3, B_3) and the two lines are orthogonal by default. The 3-dimensional point-to-line distance formula from Wolfram MathWorld (http://mathworld.wolfram.com/Point-LineDistance3-Dimensional.html) was used in this study as the basis for the regression.

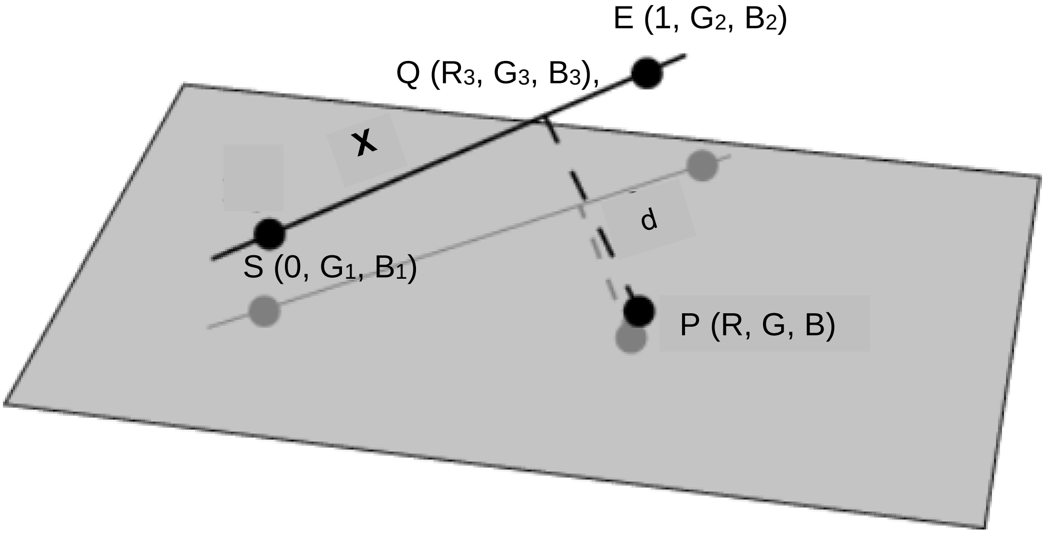


Figure 27: Illustration of a Point to Line Distance in 3D, based on Wolfram MathWorld

The formulae used in the calculations are set out below, remembering that the only known variables are the Measured DRY MATTER values and the data points P with coordinates (R,G.B) for each image. In addition two coordinates of the defining vector are also fixed, $R_1 = 0$ and $R_2 = 1$.

To check line QP is orthogonal to line SE then:

$$((R, G, B) - (R_3, G_3\ B_3)) \cdot ((1, G_2, B_2) - (0, G_1, B_1)) = 0 \qquad \text{Formula (1)}$$

Where $R_1 = 0$ and $R_2 = 1$ and **.** is the dot product

A vector along a line SE is given by:

$$v = \begin{bmatrix} R_1 + (R_2 - R_1)t \\ G_1 + (G_2 - G_1)t \\ B_1 + (B_2 - B_1)t \end{bmatrix} \qquad \text{Formula (2)}$$

The squared distance, d, between a point on the line with parameter t and a point Q with the coordinates (R_3, G_3, B_3) is therefore:

$$d^2 = [(R_1 - R_3) + (R_2 - R_1)t]^2 + [(G_1 - G_3) + (G_2 - G_1)t]^2 + [(B_1 - B_3) + (B_2 - B_1)t]^2$$

Formula (3)

To minimise the distance, set $\frac{d(d^2)}{dt} = 0$ and solve for t to obtain:

$$t = -\frac{(R_1 - R_3)\cdot(R_2 - R_1)}{|R_2 - R_1|^2}$$ where · denotes the dot product. Formula (4)

Using the 3D distance formula based on applying the Pythagoras theorem twice:

$$\text{distance} = \sqrt{\text{width}^2 + \text{length}^2 + \text{height}^2}$$

$$\text{distance}, X = \sqrt{(B_3 - B_1)^2 + (R_3 - R_1)^2 + (G_3 - G_1)^2}$$ Formula (5)

Distance X has a linear relationship with dry matter so that:

Predicted DM = $\alpha X + \beta$. Formula (6)

The R, G, B and N values were converted to top-of-atmosphere values and attenuated for atmosphere using Method 1 described in Section 3.4.2.1. Unit Vectors for R, G, B and N were then calculated for each data point before regression analysis commenced.

4.2.1 Calculating the objective function for the regression

The regression calculation:

R_1 was set at 0 and R_2 was set at 1.

Formula (4) was used to determine t for each point P.

Formula (2) was used to find the coordinates of Q = (R_3, G_3, B_3).

Formula (5) was used to find distance X.

Formula (6) was used to predict DM.

The regression calculation followed these steps to solve G_1, B_1, G_2, B_2, α and ß to minimise the sum-of-squares error between Predicted DM and Measured DM.

Table 15 is a composite representation of the data used for the regression. It shows the number of data points for each image as well as the range of dry matter. The results of the regression analysis are described in Section 4.3.

Table 15: Summarised data for the core Maturity Areas used for the regression analysis

		Number of Maturity Areas	**Minimum DM**	**Maximum DM**	**DM Range**
Image A	Gold	22	16.04	20.20	4.16
	Green	43	14.96	17.77	2.81
	All	**65**	**14.96**	**20.20**	**5.24**
Image B	Gold	15	17.07	19.11	2.04
	Green	42	15.51	17.96	2.45
	All	**57**	**15.51**	**19.11**	**3.60**
Image C	Gold	15	17.07	19.11	2.04
	Green	42	15.51	17.96	2.45
	All	**57**	**15.51**	**19.11**	**3.60**
Image D	Gold	15	17.07	19.11	2.04
	Green	40	15.51	17.96	2.45
	All	**55**	**15.51**	**19.11**	**3.60**
Image E	Gold	14	17.78	19.87	2.09
	Green	34	15.57	17.96	2.39
	All	**48**	**15.57**	**19.87**	**4.30**
Image F	Gold	14	18.01	19.97	1.96
	Green	34	15.69	17.88	2.19
	All	**48**	**15.69**	**19.97**	**4.28**
Image G	Gold	11	18.01	19.97	1.96
	Green	24	15.69	17.88	2.19
	All	**35**	**15.69**	**19.97**	**4.28**
Image H	Gold	18	17.73	19.51	1.78
	Green	20	16.44	18.08	1.64
	All	**38**	**16.44**	**19.51**	**3.07**
All Images	Gold	124	16.04	20.20	4.16
	Green	279	14.96	18.08	3.12
	All	**403**	**14.96**	**20.20**	**5.24**

4.3 Regression Results for Dry Matter

Regression runs were carried out in order to predict dry matter and the accuracy of such predictions was determined. The accuracy of the proposed solution to predict dry matter was determined by calculating the sum-of-squares error and R-squared analysis. On this basis three investigations were performed:

The first investigation sought to predict dry matter by regressing on all available data (Section 4.3.1). Once this was determined it was then investigated whether the variety of kiwifruit impacted on the predictive capability (Section 4.3.2) and whether the combined solution retained its predictive capability at an individual image level (Section 4.3.3). In addition, following the investigation at an individual image level, it was considered whether predictive capability varied over time, dependant on when each image was captured (Section 4.3.4).

4.3.1 Multi-temporal, Multi-image Regression

The data for the core reference Maturity Areas in all images (403 data points), as detailed in Table 15, Section 4.2.1, was used as a combined data set for a full multi-image, multi-temporal regression. The average R, G, B values for each of the 403 Maturity Areas and the corresponding Measured DM values were inputs for the regression. The results of this regression are shown in Figure 28. Table 16 details the coordinate and coefficient values that were determined by this regression.

Table 16: Regression output, coordinates and coefficients

G_1	B_1	G_2	B_2	α	β
11.8048	10.5001	26.7072	- 7.1169	119.2304	- 30.5309

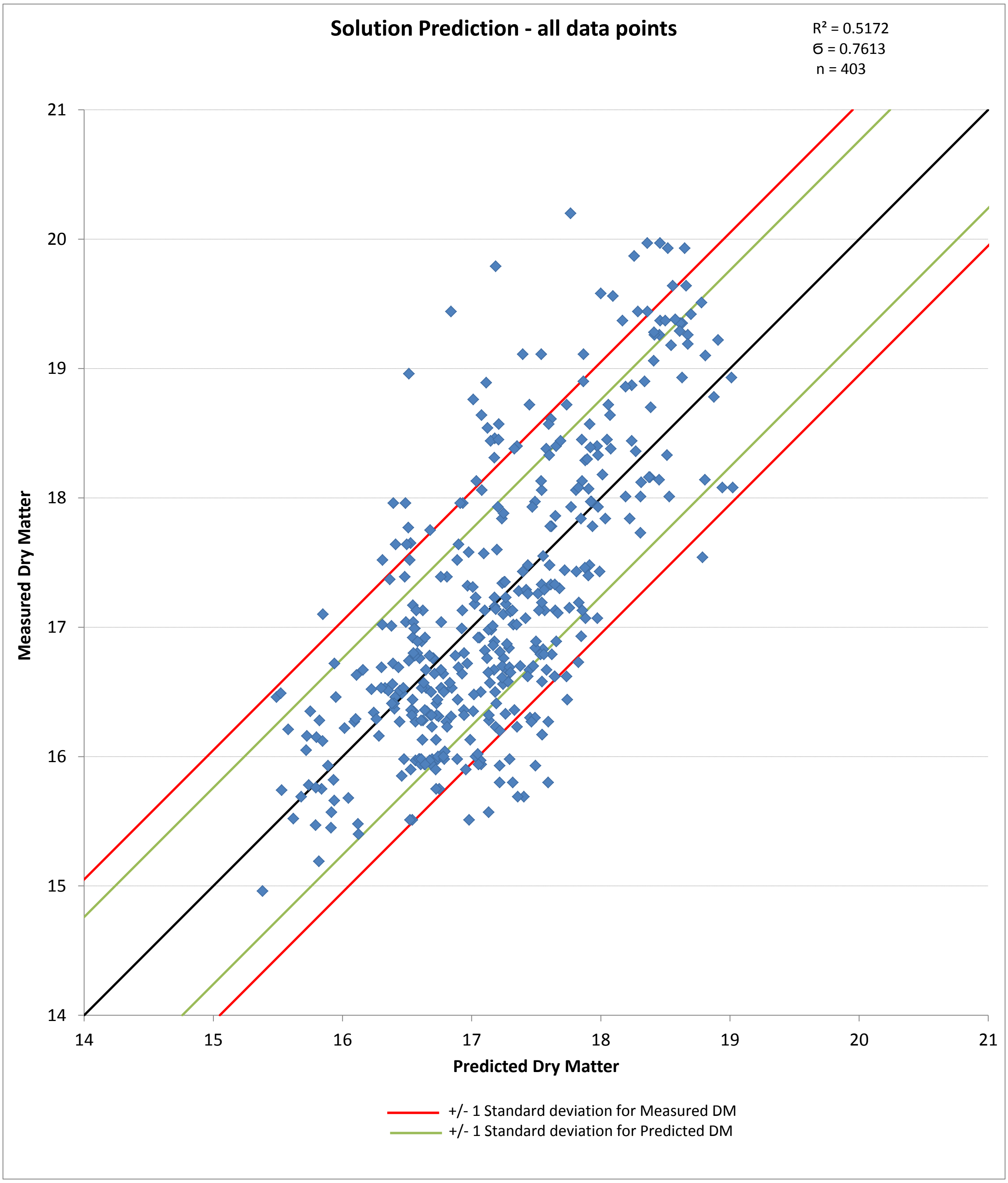


Figure 28: Multi-temporal, multi-image regression for the combined data set of all images with error bars showing one standard deviation of measured and predicted dry matter

Figure 28 shows a strong correlation given the standard deviations of the Measured DM values. The error bars show one standard deviation of the dry matter data as measured or predicted. The correlation was also strongest using linear regression with no improvement in standard deviation when quadratic relationships were tested.

4.3.2 Kiwifruit variety and prediction from multi-image regression

There are two varieties of kiwifruit, *Hayward* and *Hort16A*, of the same genus *Actinidia*. The *Hort16A* variety has, on average, a higher dry matter than the *Hayward* variety; refer to Table 15, Section 4.2.1.

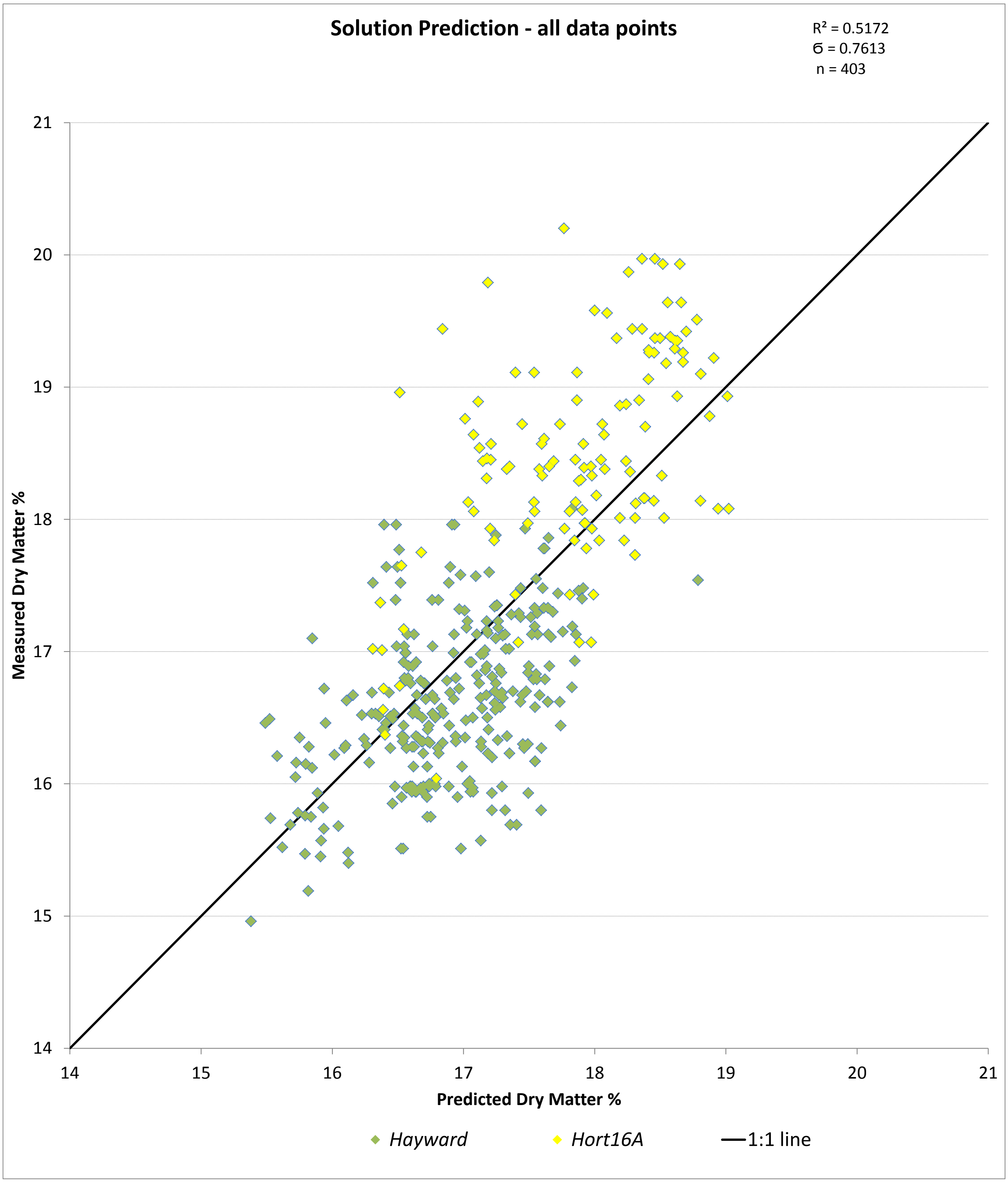


Figure 29: Multi-temporal, multi-image regression for the combined data set of all images showing the data points designated as *Hayward* or *Hort16A*

Considering Figure 29, it appears that the predictive formula for the fully combined regression provides similar predictive capability for *Hayward* and *Hort16A* kiwifruit. The data points for both are falling along the same line of best fit when Predicted DM is plotted against Measured DM. The sum-of-squares error for each of *Hayward* and *Hort16A* resulting from the combined regression is 0.6062 and 0.7248 respectively compared to the combined σ of 0.7613. These σ values confirm that the predictive accuracy is similar for *Hayward* and *Hort16A* and that the same regression results can be used to predict dry matter for both varieties.

4.3.3 Individual image prediction from multi-image regression

The Predicted DM values were plotted against the Measured DM values for each image. The Predicted DM values were those calculated using the coordinates and coefficients in Table 16, Section 4.3.1. Figure 30 shows these results and that the formula determined from the multi-temporal multi-image regression in Section 4.3.1 does provide good predictive capability at the individual image level. The predictive accuracy across the individual images had σ values ranging from 0.61 to 0.80 and R^2 values ranging from 0.49 to 0.71 compared to the σ of 0.76 and R^2 of 0.52 of all images together. The fact that there does not seem to be significant variation from image to image implies that the atmospheric attenuation strategy is adequate.

4.3.4 Time dependence

Two seasons, 2006 (Images B-D) and 2008 (Images F-G), had time series data. In addition two images were captured as early in the season as January (Image E 2007 and Image G 2008), refer Table 7, Section 3.2.1. Considering the predictive accuracy for these images in Figure 30 it is seen that dry matter can be predicted with similar accuracy early in the season, and as early as January. This means that this technology can be utilised early in the growing season when dry matter cannot otherwise be easily measured or well correlated to at-harvest dry matter because the fruit are not of sufficient size.

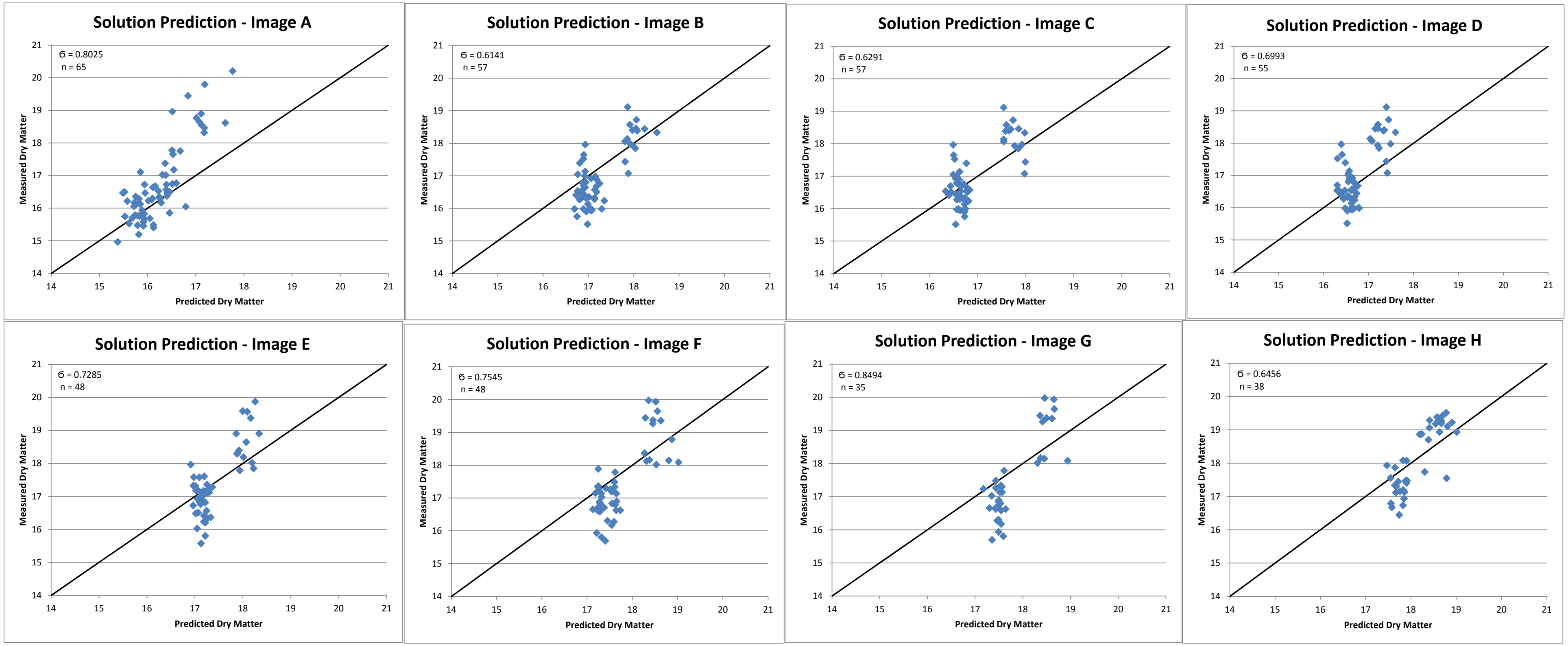


Figure 30: Predicted Dry Matter versus Measured Dry Matter for all reference data points shown by individual image; using the solution formula to predict the dry matter

4.4 Applying a standard formula to predict dry matter

In summary, the method proposed to predict dry matter has several steps. These steps are applicable for any image.

Calculation steps:

1. Extract raw R, G, B and N Digital Numbers for all pixels associated with a measured dry matter value from a given image.
2. Calculate average R, G, B and N Digital Numbers of all pixels associated with a measured dry matter value.
3. Adjust average R, G, B and N Digital Numbers using absCalFactors to top-of-atmosphere.
4. Attenuate for atmosphere using R, G, B and N coefficients from Method 1 relevant to the image from which the raw R, G, B and N Digital Numbers were extracted.
5. Calculate unit vectors for each data point.
6. Calculate the distance value using the 0, G1, B1 and 1, G2, B2 coordinates determined from the combined regression of all images. These coordinates are detailed in Table 16, Section 4.3.1.
7. Calculate the predicted dry matter using the α and β coefficients determined from the combined regression of all images. These coefficients are detailed in Table 16, Section 4.3.1.

4.5 Visualising the solution – a worked example

Following the calculation of Predicted DM (using the regression analysis formulae steps in Section 4.2.1 and the solution application steps in Section 4.4) for all reference data points from all images, a subset of twelve maturity areas were selected to demonstrate the solution as a worked example. The tables in this section provide the relevant values for the twelve maturity areas required by the formulae to test the maths. This subset of twelve maturity areas was chosen because they represented a separation of dry matter values, came from various images and had low variance between Predicted DM and Measured DM so that they could be used to clearly demonstrate the solution concept when plotted (Figure 31) – the number of data points used in the solution otherwise made it impossible to clearly visualise the concept. The final solution used the full set of data and this subset from within the full data set was not separated or used in any other way for the solution therefore did not influence the outcome.

The four unknown coordinates of the defining vector starting at S with the coordinates $(0, \mathrm{G}_1, \mathrm{B}_1)$ and ending at E with the coordinates $(1, \mathrm{G}_2, \mathrm{B}_2)$ were outputs from the regression (refer Table 16). The defining vector coordinates are shown shaded in grey in Table 17. Refer also Figure 27, Section 4.2.

The coordinates of the defining vector and those of each reference data point (R, G, B), were used to calculate the value of t for each reference data point using Formula (4), Section 4.2. Knowing the value of t and using Formula (2), Section 4.2, the vector coordinates for the point along the line SE known as $\mathrm{R}_{3,}\mathrm{G}_{3,}\mathrm{B}_{3,}$ for each reference data point were then calculated. Refer Table 17.

Table 17: Coordinates for line of best fit and shortest lines from each reference data point

	R_1	G_1	B_1		R_2	G_2	B_2	
S	-	11.80478	10.50006	E	1.00000	26.70715	- 7.11692	
	start				end			
	R_3	G_3	B_3		R	G	B	t
Q_1	0.01678	12.05486	10.20443	P_1	0.12912	0.28746	0.25665	0.01678
Q_2	0.01680	12.05517	10.20405	P_2	0.12212	0.28141	0.25049	0.01680
Q_3	0.01699	12.05801	10.20070	P_3	0.12023	0.27016	0.23511	0.01699
Q_4	0.01706	12.05908	10.19944	P_4	0.10927	0.25003	0.21529	0.01706
Q_5	0.01713	12.06008	10.19826	P_5	0.12114	0.27240	0.23286	0.01713
Q_6	0.01723	12.06159	10.19647	P_6	0.13745	0.29819	0.25253	0.01723
Q_7	0.01730	12.06255	10.19534	P_7	0.07259	0.17850	0.14565	0.01730
Q_8	0.01746	12.06500	10.19244	P_8	0.19055	0.38136	0.31897	0.01746
Q_9	0.01760	12.06702	10.19005	P_9	0.13744	0.30364	0.24610	0.01760
Q_{10}	0.01772	12.06888	10.18786	P_{10}	0.09445	0.22508	0.17343	0.01772
Q_{11}	0.01799	12.07289	10.18311	P_{11}	0.19131	0.38749	0.30816	0.01799
Q_{12}	0.01795	12.07233	10.18378	P_{12}	0.20998	0.42139	0.33904	0.01795

Formula (1), Section 4.2, was used to check that the shortest lines for each reference data point were orthogonal to the line of best fit. Refer Table 18.

Table 18: Shortest lines orthogonality check

$R_2 - R_1$	$G_2 - G_1$	$B_2 - B_1$	
1.00000	14.90237	- 17.61698	
orthogonality test			
$R_i - R_{3,i}$	$G_i - G_{3,i}$	$B_i - B_{3,i}$	dot product
0.11234	- 11.76739	- 9.94777	000.0E+0
0.10531	- 11.77377	- 9.95356	000.0E+0
0.10324	- 11.78785	- 9.96559	000.0E+0
0.09221	- 11.80905	- 9.98415	000.0E+0
0.10401	- 11.78767	- 9.96540	000.0E+0
0.12022	- 11.76339	- 9.94394	000.0E+0
0.05530	- 11.88405	- 10.04969	000.0E+0
0.17309	- 11.68363	- 9.87347	000.0E+0
0.11984	- 11.76338	- 9.94395	000.0E+0
0.07672	- 11.84380	- 10.01443	000.0E+0
0.17332	- 11.68540	- 9.87495	000.0E+0
0.19203	- 11.65094	- 9.84474	000.0E+0

Distance X was calculated using Formula (5), Section 4.2. Distance X was then used to predict dry matter using the coefficients α and β output by the regression analysis in Table 16, Section 4.3.1 and using Formula (6), Section 4.2. The X values, α and β coefficients, Predicted DM, Measured DM and variance values are shown in Table 19.

Table 19: Predicted dry matter using solution coefficients

$x_2 - x_1$	$y_2 - y_1$	$z_2 - z_1$
13.90237	- 44.32413	7.11692

	predicting DM				Measured DM (MDM)	Variance (MDM-PDM)
i	X	α	β	PDM$_i$		
1	0.38758	119.23040	- 30.53093	15.68	15.69	0.0096
2	0.38807	119.23040	- 30.53093	15.74	15.78	0.0412
3	0.39246	119.23040	- 30.53093	16.26	16.29	0.0275
4	0.39413	119.23040	- 30.53093	16.46	16.49	0.0292
5	0.39567	119.23040	- 30.53093	16.65	16.67	0.0249
6	0.39801	119.23040	- 30.53093	16.92	16.99	0.0662
7	0.39950	119.23040	- 30.53093	17.10	17.13	0.0286
8	0.40329	119.23040	- 30.53093	17.55	17.55	- 0.0038
9	0.40643	119.23040	- 30.53093	17.93	17.97	0.0422
10	0.40931	119.23040	- 30.53093	18.27	18.36	0.0891
11	0.41553	119.23040	- 30.53093	19.01	18.93	- 0.0823
12	0.41465	119.23040	- 30.53093	18.91	19.22	0.3116

The R, G, B unit vector values for each of the twelve maturity areas were plotted in 3D (refer Figure 31). The black line in Figure 31 is the defining vector determined by the regression analysis with reference to Figure 26, Section 4.1, and this line is SE with reference to Figure 27, Section 4.2. The coloured lines are the shortest lines from the data point at the average of each maturity area to the black line; with reference to Figure 27, Section 4.2 these lines are P_iQ_i. Note that the intersections of these coloured lines with SE are ordered according to the dry matter content of the maturity areas. The regression has therefore correctly captured the relationship between the dry matter content of these maturity areas and their RGB vectors.

As discussed in Section 4.1 the data points of similar dry matter fall in a plane as plotted in Figure 25 and schematically represented in Figure 26. This is a result of data points being sourced from different images. Despite being sourced from different images the data points are all still orthogonal to the same defining vector as described by the solution. However they follow different directions to the point where they meet on the defining vector, subject to the image they are sourced from. This can be more clearly seen in Figure 31 where points Q_7 and Q_{10} come from distinctly different directions. Table 20 details the image source for each of the 12 points used to visualise the solution.

Table 20: Image details of the visualisation subset

Data Point	Image Source
Q_1	A
Q_2	A
Q_3	A
Q_4	A
Q_5	C
Q_6	B
Q_7	E
Q_8	H
Q_9	B
Q_{10}	F
Q_{11}	H
Q_{12}	H

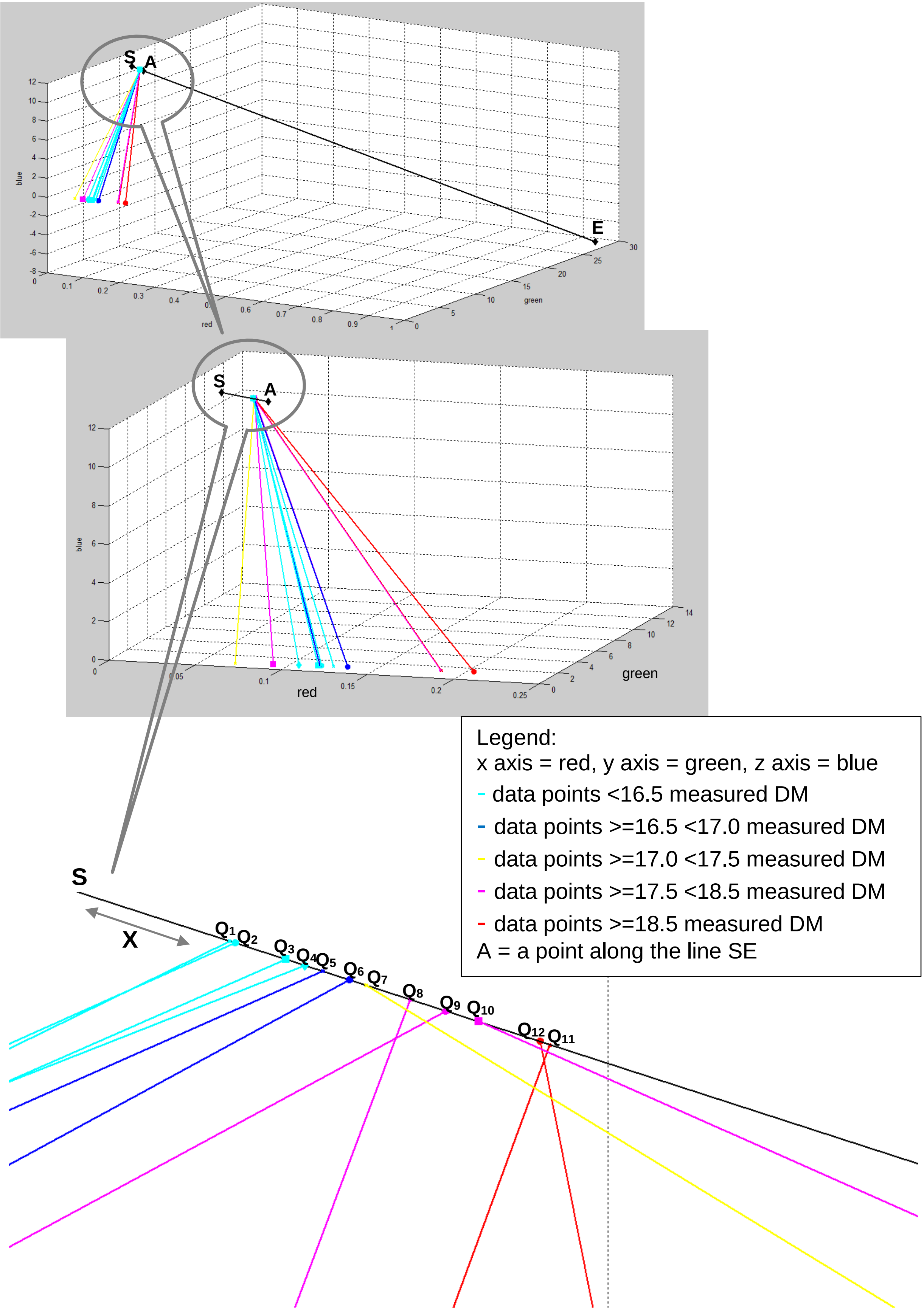


Figure 31: 3D plot of the solution using selected data points

4.6 Testing the method on data that was not used in the teaching set

In order to test the predictive capability of the solution out-of-sample tests were performed. The Predicted DM for an additional 99 data points from nine orchards, and from within all images, was calculated using the process detailed in Section 4.4, Chapter 4. None of these 99 data points were used in the original regression and on average there were 12 data points from each image (ranging from 3 - 15). Twenty one of the 99 data points were from the Gold variety. The results are presented in Figure 32.

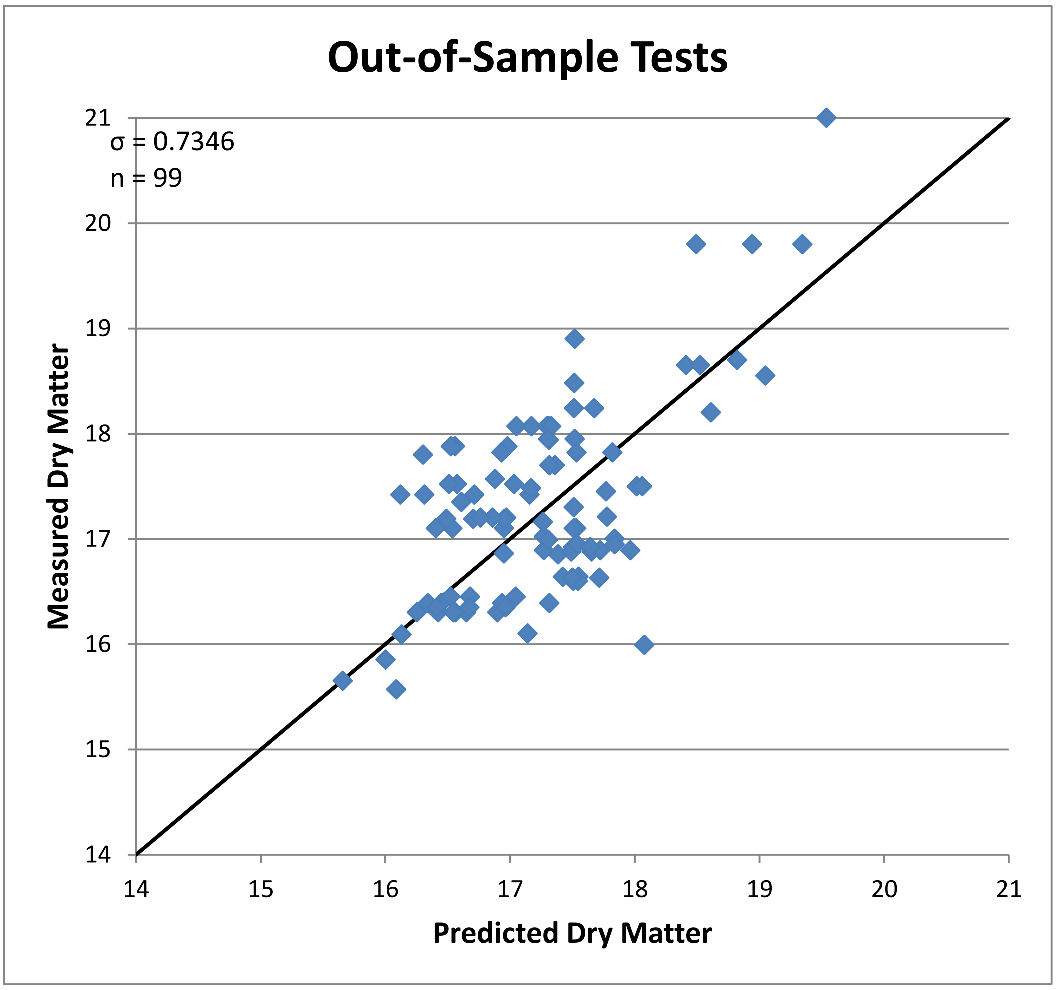


Figure 32: Out-of-sample test results for the prediction of dry matter

The predicted dry matter for data from the same orchard and maturity area was compared for each of the three images captured in 2006 (Images B - March 2006, Image C - April 2006 and Image D – May 2006) as shown in Table 21.

Table 21: Comparison of predicted dry matter from same orchard and maturity area data across the three images captured through the 2006 season

				Predicted DM				
Test Orchard Number	**Maturity Area Number**	**Variety**	**Measured DM**	**Mar-06**	**Apr-06**	**May-06**	**Average**	**σ**
1	a	*Hayward*	17.42	16.71	16.31	16.12	16.38	0.3022
2	a	*Hort16A*	17.82	17.82	17.54	16.93	17.43	0.4545
3	a	*Hort16A*	19.80	19.34	18.94	18.49	18.93	0.4254
4	a	*Hayward*	16.45	17.05	16.68	16.52	16.75	0.2688
4	b	*Hayward*	16.30	16.65	16.25	16.42	16.44	0.1979
4	c	*Hayward*	16.30	16.90	16.56	16.55	16.67	0.1979
5	a	*Hayward*	17.10	16.95	16.54	16.41	16.63	0.2855
6	a	*Hayward*	16.39	16.94	16.45	16.34	16.58	0.3175
6	b	*Hayward*	16.35	16.96	16.68	16.41	16.68	0.2757
7	a	*Hayward*	17.88	16.98	16.56	16.52	16.69	0.2534
7	b	*Hayward*	17.52	17.03	16.57	16.51	16.71	0.2858
8	a	*Hayward*	17.20	16.97	16.86	16.76	16.86	0.1040
8	b	*Hayward*	17.19	16.95	16.71	16.49	16.72	0.2324
8	c	*Hayward*	18.07	17.17	17.31	17.30	17.26	0.0759
9	a	*Hort16A*	16.89	17.97	17.72	17.27	17.65	0.3520

Table 21 shows a consistent prediction of dry matter at full maturity with low variability across the 2006 multi-temporal image series. This supports the conclusion in Section 4.3.4, Chapter 4, that the method proposed by this study can predict dry matter early in the season.

4.7 Comparison of this study to existing indices

The indices found during the literature search (refer Appendix 3) were the results of other studies to determine correlations to various crop qualities. These indices were tested as to how well they might predict dry matter in kiwifruit. The tests used the top-of-atmosphere R, G, B and N values as applicable to the indices being tested. The results of Measured DM plotted against twelve of these indices are represented in Figure 34 and summarised in Table 22.

Table 22: Summary of Indices Results

Index as Correlated against Dry Matter	R^2 Value
Red	0.0084
Green	0.4283
Blue	0.2655
NIR	0.1601
NIR/Red	0.1107
NIR/Green	0.0462
NIR/Blue	0.2211
NG-1	0.0462
NDVI	0.1154
BNDVI	0.2157
GNDVI	0.0515
NGRDI	0.2469
Green/Blue	0.3993
This Work	0.5172

The best correlations against Measured DM were achieved with Green values alone. Blue alone also showed an interesting reverse trend to green, although not as strong. There was no index that was as capable in prediction for dry matter as the proposed solution in this study. Whilst no mention of Green as a ratio of Blue was found referenced during the literature search it was decided to consider Green as a ratio of Blue given the trends apparent. This is consistent with the regression technique used in this study that varied Green and Blue to solve the coordinates of the defining vector.

The Green/Blue ratio for all 403 data points used in the multi-temporal multi-image regression (unit vector values following atmospheric attenuation using Method 1) were plotted against Measured DM and the formula for the line of best fit was used to predict dry matter. Refer Figure 33 (a) for the determination of the formula for line of best fit. Figure 33 (b) shows Predicted DM plotted against Measured DM. Whilst the Green/Blue ratio has some predictive capability the predictive capability is not as accurate as that achieved by the solution proposed in this study.

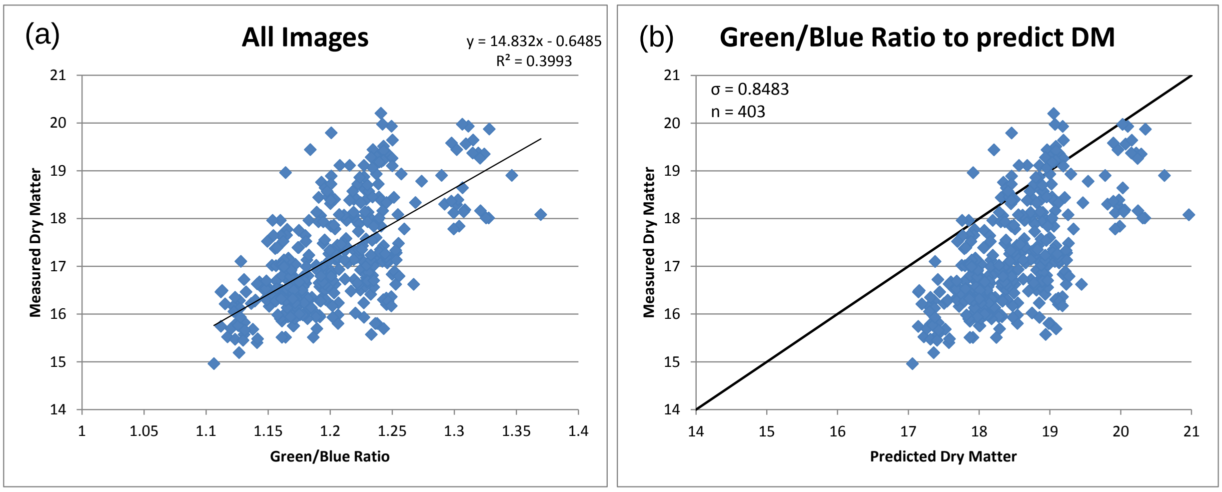


Figure 33: Green to Blue ratio as a predictor of dry matter (a) line of best fit formula (b) Measured Dry Matter versus Predicted Dry Matter

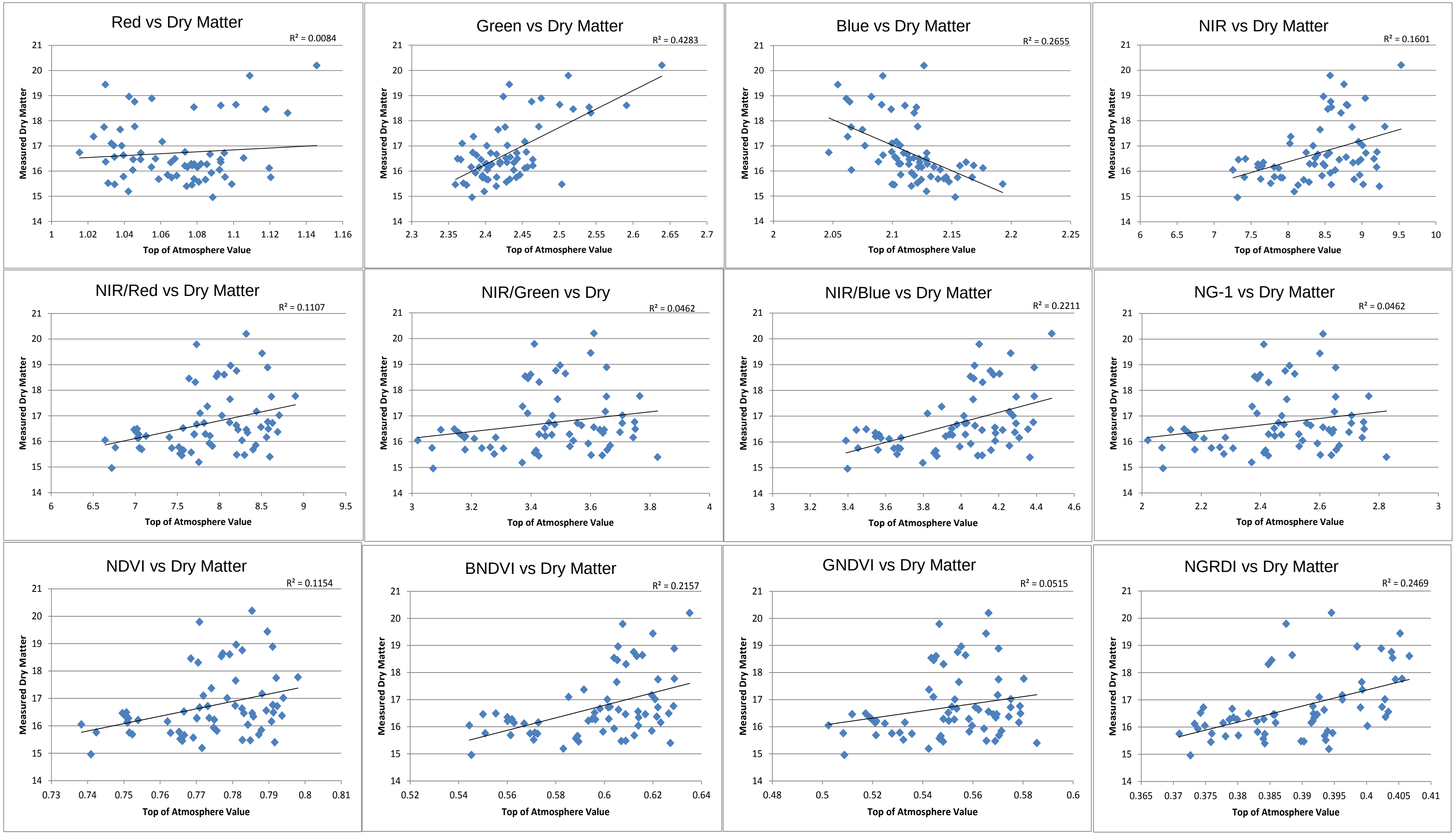


Figure 34: Correlation of various indices against measured dry matter using top-of-atmosphere values for Image A

4.8 Comparison of this study with current dry matter sampling system

The current method of determining the dry matter content of kiwifruit involves the destructive testing of 90 kiwifruit. Refer Section 3.2.4.2.3. This method has an average standard deviation ranging from 0.52 to 1.80 with an average of 1.05. It will be used a number of times through the growing session. It is a direct measure of dry matter as it dehydrates the fruit to determine the weight of the residue, the dry matter.

By comparison, the method developed here is a non-contact and non-destructive method. It does take not long to implement but it is reliant on the satellite capturing images on cloudless days and as much as a fortnight thereafter before the images are available. The average standard deviation is 0.76, despite the method using the colour of the kiwifruit leaves as a proxy for dry matter. We note that the data this works with are noisy but that the method produces less error in its predictions.

Since we regressed against this noisy data and obtained a smaller variance than the actual data, we are uncertain about the accuracy of the method but judge that it will add value to the kiwifruit industry. If several images are taken of the same areas during the season, the aggregate prediction for each maturity area is bound to be better than the existing measurement.

The cost implications of using this method over the current industry-standard 90-fruit protocol are significant. It should be further evaluated to test its robustness over larger data sets, with a view to being made available to growers in the future.

Being a remote sensing method, it would be as simple to measure the dry matter content of orchards in such places as Chile as it would be those near Te Puke. This has implications in terms of identifying dry matter content in distant orchards as well as those of competitors.

4.9 Other Predictive Functions Considered

During the course of the study consideration was also given as to the potential to predict other fruit characteristics.

- Soluble Sugar or brix is not practical to consider as it continues to develop after harvest and is also a function of dry matter however timing of maturity measured by a combination of internal characteristics such as seed colour, brix and dry matter is a possibility but unless all orchards had ground based data tested for these at the same time it is not possible to develop a prediction.

- Yield and average fruit weight was considered however insufficient ground based data was available – this requires knowledge of total harvested kilograms of fruit (not just export volumes) and volume of under and over-sized fruit (not just volume of export size).
- Orchard maps detailing the predicted dry matter distribution across an orchard is possible as shown as an example in Figure 35. Such maps would enable growers to decide if certain areas require additional attention e.g. be targeted for girdling, or which rows could be combined for harvest to optimise income.

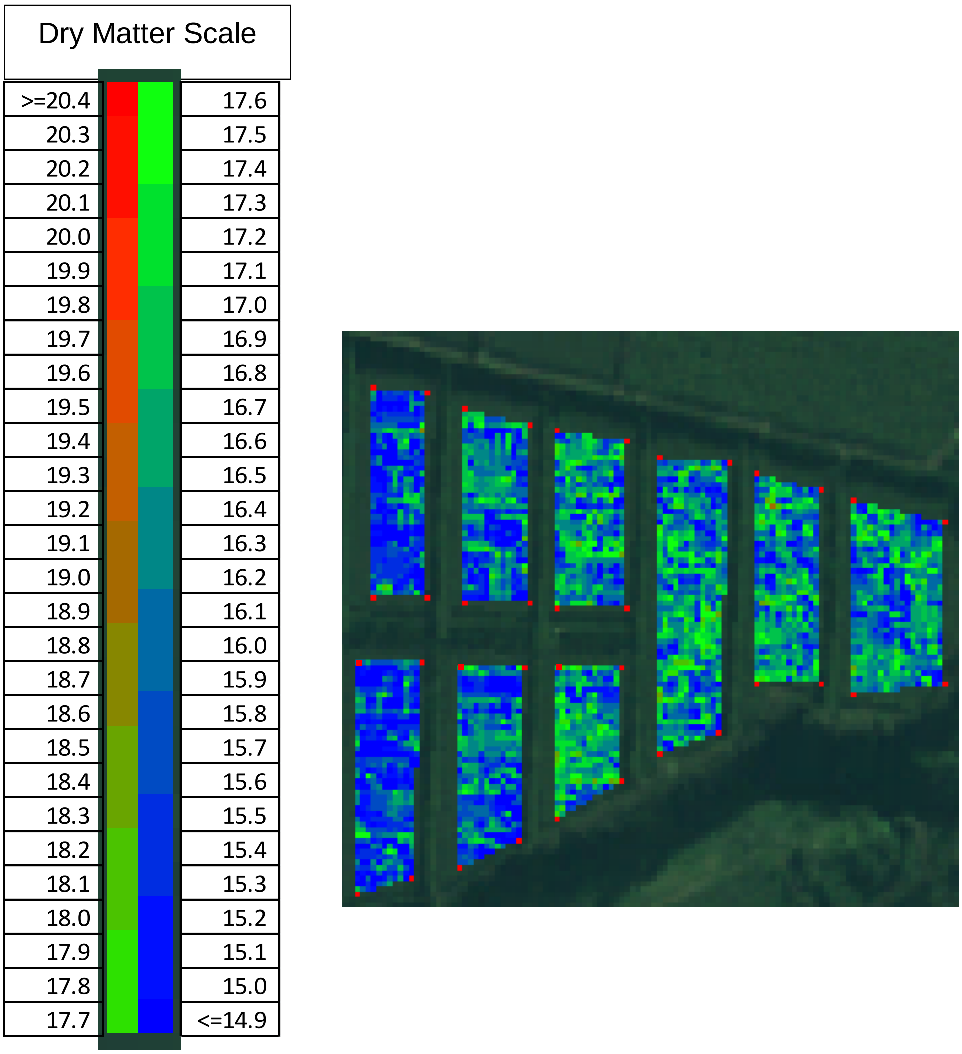


Figure 35: Example of a predicted dry matter plot for Orchard #2, Te Puke, May 2005

4.10 Conclusions

Analyses of current methods for dry matter prediction, and tests of the performance of a new regression method, have reached the following conclusions.

- The key to understanding the complex variation of dry matter with the energy in each of the four spectral bands is a suitable visualization. The method presented reduces this four-dimensional problem to a three-dimensional problem – thus permitting visualization.

- The methods for predicting crop characteristics from spectral data that are reported in literature are empirical only. Of those that could predict dry matter from image A, the best method had an R^2 value of 0.42 and the worst, 0.008 (Figure 30).

- Following visualization, an analytical method was developed (refer Section 4.2) to predict dry matter over 403 maturity areas, ranging over 5 years. The results of this multi-temporal prediction had an R^2 of 0.51. The R^2 for image A was 0.70.

- An atmospheric attenuation method, based on Hall *et al*., (1991) was chosen, as it provided the largest reduction in the variability of fiduciary pixels from the un-attenuated images, refer Section 3.4.3.

- The industry-standard method of determining dry matter is a time-intensive and destructive test with an average standard deviation of 1.05. By comparison, the method developed here is a non-contact and non-destructive method, which requires little time to implement but is reliant on the capture of images on cloudless days and may then take several weeks to be released. The solution presented has an average standard deviation of 0.76.

- The method presented here could be made available to growers. Further evaluation could be undertaken to:
  a. Investigate methods of image capture other than the use of satellite-based sensor platforms such as aircraft or drone based platforms. Options and technology have advanced during the time of this study.
  b. Apply the solution in a real example with data capture, maturity area segmentation, specified areas of ground data capture

- The method seemed equally competent on *Hort16A* and *Hayward* varieties (Section 4.3.2) and, surprisingly, was entirely capable of predicting dry matter from images so early in the season that the kiwifruit are not of sufficient size to be measured (Section 4.3.4). No consistent variation with time was observed (Section 4.3.3).

- The method can be applied to predict the quality of kiwifruit grown in origins other than New Zealand such as Chile, Italy or China very early in the growing season. This has significant commercial value to Zespri.

- The method developed in this thesis has the potential to add significant value to Zespri and kiwifruit growers. The method needs further validation and development to assess its robustness with a view to making it available to growers. It is yet to be seen if it can be further improved and to determine other kiwifruit characteristics such as crop yield or maturity timing.

REFERENCE LIST

Al-Abbas, A. H., Barr, R., Hall, J. D., Crane, F. L., & Baumgardner, M. F. (1974). Spectra of normal and nutrient-deficient maize leaves. Agronomy Journal, 66, 16-20.

Anatoly, A., & Gitelson, A. A. (2013). Remote estimation of crop fractional vegetation cover: the use of noise equivalent as an inidcator of performance of vegetation indices. International Journal of Remote Sensing, 34(17), 6054-6066.

Arnis, A., Manan, W. N. A., Ahmad, N. N., & Mansor, S. B. (2014). The combined technique of regional atmospheric correction. Paper presented at the Aerospace Electronics and Remote Sensing Technology Conference, Yogyakarta, 216-222

Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10(4), 438-452.

Bausch, W. C., Halvorson, A. D., & Cipra, J. (2008). Quickbird satellite and ground-based multispectral data correlations with agronomic parameters of irrigated maize grown in small plots. Biosystems Engineering, 101(3), 306-315.

Beeri, O., Phillips, R., Carson, P., & Liebig, M. (2005). Alternate satellite models for estimation of sugar beet residue nitrogen credit. Agriculture, Ecosystems & Environment, 107(1), 21-35.

Beever, D. J., Hopkirk, G., Warrington, I. J., & Weston, G. C. (1990). Fruit development and fruit physiology Kiwifruit: Science and Management.

Auckland, New Zealand: Ray Richards.

Bégué, A., Todoroff, P., & Pater, J. (2008). Multi-time scale analysis of sugarcane within-field variability: improved crop diagnosis using satellite time series? Precision Agriculture, 9(3), 161-171.

Brase, T. (2006). Introduction. In S. Gomoll (Ed.), Precision Agriculture (pp. 1-28). New York, U.S.: Thomson Delmar Learning.

Burdon, J., McLeod, D., Lallu, N., Gamble, J., Petley, M., & Gunson, A. (2004). Consumer evaluation of "Hayward" kiwifruit of different at-harvest dry matter contents. [Article]. Postharvest Biology and Technology, 34(3), 245-255.

Buwalda, J. G., Meekings, J. S., & Smith, G. S. (1991). Seasonal changes in photosynthetic capacity of leaves of kiwifruit (Actinidia deliciosa) vines. Physiologia Plantarum, 83, 93-98.

Cammarano, D., Fitzgerald, G. J., Casa, R., & Basso, B. (2014). Assessing the robustness of vegetation indices to estimate wheat N in Mediterranean environments. Remote Sensing 6, 2827-2844.

Chavez, P. S., Jr. (1989). Radiometric calibration of Landsat Thematic Mapper multispectral images. Photogrammetric Engineering and Remote Sensing, 55, 1285-1294.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sensing of Environment, 91(3-4), 332-344.

Curran, P. J. (2001). Imaging spectrometry for ecological applications. International Journal of Applied Earth Observation and Geoinformation, 3(4), 305-312.

Delalieux, S., Somers, B., Hereijgers, S., Verstraeten, W. W., Keulemans, W., & Coppin, P. (2008). A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. Remote Sensing of Environment, 112(10), 3762-3772.

DigitalGlobe (2009). Basic Imagery and Standard Imagery. www.digitalglobe.com, 10.4.09.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R., & Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation, 9*(2), 165-193.

Gaut, N. E., Reifenstein III, E. C., & Seivering, H. (1975). Interaction mechanisms - within the atmosphere. In R. G. Reeves (Ed.), Manual of remote sensing; American Society of Photogrammetry (First ed.). Falls Church, VA.: Keuffel and Esser Company.

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of Plant Physiology, 161(2), 165-173.

Gitelson, A. A., & Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plan leaves. International Journal of Remote Sensing, 18(12), 2691-2697.

Greer, D. H. (1999). Seasonal and daily changes in carbon acquisition of kiwifruti leaves with and without axillary fruit. New Zealand Journal of Crop and Horticultural Science, 27, 23-31.

Haboudane, D., Tremblay, N., Miller, J. R., & Vigneault, P. (2008). Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. IEEE Transactions on Geosciences and remote sensing, 46, 423-437.

Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S. J. (1991). Radiometric Rectification: Toward a common radiometric response among multidate, multisensor images. Remote Sensing of Environment(35), 11-27.

Harker, F. R., Carr, B. T., Lenjo, M., MacRae, E. A., Wismer, W. V., Marsh, K. B., et al. (2009). Consumer liking for kiwifruit flavour: a meta-analysis of five studies on fruit quality. Food Quality and Preference, 20(1), 30-41.

Hatfield, J. L., Gitelson, A. A., Schepers, J. S., & Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. Agronomy Journal Centennial Supplement, S, 117-131.

Hopkirk, G. (1991). Dry matter - an alternative maturity test? New Zealand Kiwifruit Journal, 75, 30.

Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (2013). Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. PLOS One, 8(8).

Hunt, E. R., Cavigelli, M., Daughtry, C. S. T., McMurtrey, J. E., & Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. Precision Agriculture, 6(4), 359-378.

Hunt, E. R. J., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S. T., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. International Journal of Applied Earth Observation and Geoinformation, 21, 103-112.

Jacquemoud, S., Ustin, S. L., Verdebout, J., Schmuck, G., Andreoli, G., & Hosgood, B. (1996). Estimating leaf biochemistry using the PROSPECT leaf optical properties model. Remote Sensing of Environment, 56(3), 194-202.

Jannoura, R., Brinkmann, K., Uteau, D., Bruns, C., & Joergensen, R. G. (2015). Monitoring of crop biomass using true colour aerial photographs taken from a remote controlled hexacopter. Biosystems Engineering, 129, 341-351.

Jiang, D., Wang, N.-b., Yang, X.-h., & Wang, J.-h. (2003). Study on the interaction between NDVI profile and the growing status of crops. Chinese Geographical Science, 13(1), 62-65.

Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. Ecology, 50(4), 663-666.

Kaufman, Y. J., & Tanré, D. (1996). Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. Remote Sensing of Environment, 55(1), 65-79.

LED Digital Visuals (2009). Google Search Leaf Structure. http://leddv.com/E_xp15_Memo_3.jpg.

Leon, C. T., Shaw, D. R., Cox, M. S., Abshire, M. J., Ward, B., Wardlaw, M. C., et al. (2003). Utility of remote sensing in predicting crop and soil characteristics. Precision Agriculture, 4(4), 359-384.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2004). Concepts and foundations of remote sensing Remote sensing and image interpretation (fifth ed., pp. 1-57). United States of America: John Wiley & Sons, Inc.

Lofton, J., Tubana, B., Kanke, Y., Teboh, J., Viator, H., & Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalised difference vegetative index. Sensors, 12, 7529-7547.

McDonald, A. J., Gemmell, F. M., & Lewis, P. E. (1998). Investigation of the utility of spectral vegetation indices for determining information on coniferous forests. Remote Sensing of Environment, 66(3), 250-272.

Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sensing of Environment, 61(3), 319-346.

Mowat, A., & Maguire, K. (2006, Feb 20-24). Canopy management and dry matter of 'Hayward' kiwifruit. Paper presented at the 6th International Symposium on Kiwifruit, Rotorua, NEW ZEALAND.

Niinemets, U. (1999). Research review: components of leaf dry mass per area - thickness and density - alter leaf photosynthetic capacity in reverse directions in woody plants. New Phytologist, 144(1), 35-47.

Olah, R., & Masarovicova, E. (1995). Some photosynthetic characteristics in the leaves of kiwifruit. New Zealand Journal of Crop and Horticultural Science, 23, 155-161.

Pettersson, C.-G., Söderström, M., & Eckersten, H. (2006). Canopy reflectance, thermal stress, and apparent soil electrical conductivity as predictors of within-field variability in grain yield and grain protein of malting barley. Precision Agriculture, 7(5), 343-359.

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sensing, 5, 1704-1733.

Ren, J., Chen, Z., Zhou, Q., & Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. International Journal of Applied Earth Observation and Geoinformation, 10(4), 403-413.

Savitzy, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 36, 1627-1639.

Seager, N. G., Hewett, E. W., & Warrington, I. J. (1995). Manipulation of carbohydrate concentrations in kiwifruit. New Zealand Journal of Crop and Horticultural Science, 23, 213-218.

Son, N. T., Chen, C. F., Chen, C. R., Minh, V. Q., & Trung, N. H. (2014). A comparative analysis of multi-temporal MODIS EVI and NDVI data for large-scale rice yield estimation. Agricultural and Forest Meteorology, 197, 52-64.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? Remote Sensing of Environment, 75(2), 230-244.

Tian, Y., Zhu, Y., & Cao, W. (2005). Monitoring leaf photosynthesis with canopy spectral reflectance in rice. Photosynthetica, 43(4), 481-489.

Vernier (2009). Google Search for Chlorophyll; Viewed 18 April 2009. http://www.vernier.com/innovate/innovativeuse63.html

Wang, L., Tian, Y., Yao, X., Zhu, Y., & Cao, W. (2014). Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crops Research, 164, 178-188.

Wiegand, C. L., & Richardson, A. J. (1984). Leaf area, light interception, and yield estimates from spectral components analysis. Agronomy Journal, 76, 543-548.

Wilson, P. J., Thompson, K., & Hodgson, J. G. (1999). Specific leaf area and leaf dry matter content as alternative predictors of plant strategies. New Phytologist, 143(1), 155-162.

Wu, J., Wang, D., & Bauer, M. E. (2005). Image-based atmospheric correction of QuickBird imagery of Minnesota cropland. Remote Sensing of Environment, 99(3), 315-325.

Xiao, Y., Zhao, W., Zhou, D., & Gong, H. (2013). Sensitivity analysis of vegetation reflectance to biochemical and biophysical variables at leaf, canopy, and regional scales. Geoscience and Remote Sensing 52(7), 4014-4024.

Yang, C., Everitt, J., & Bradford, J. (2006). Comparison of QuickBird satellite imagery and airborne imagery for mapping grain sorghum yield patterns. Precision Agriculture, 7(1), 33-44.

Ye, X., & Sakai, K. (2012). Fruit yield estimation through multispectral imaging In A. K. Srivastava (Eds.), Advances in Citrus Nutrition, Section 30, page 453, google e book

Ye, X., Sakai, K., Sasao, A., & Asada, S. (2008). Potential of airborne hyperspectral imagery to estimate fruit yield in citrus. Chemometrics and Intelligent Laboratory Systems, 90(2), 132-144.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. Precision Agriculture, 13(6), 693-712.

Zhang, J., Scarrow, S., & Barnes, H. T. (2004). Model of the effects of a range of shelter orientations on shading patterns in kiwifruit (Actinidia sp.) blocks. New Zealand Journal of Crop and Horticultural Science, 32, 407-417.

Appendix 1: Summary of Satellite Specifications

Satellite	**QuickBird**	**Landsat**	SPOT 2,3,4 / 5	**IKONOS**	**EROS**
Company	DIgitalGolbe	NASA/US Geological Survey	**SPOT Image**	Space Imaging	ImageSat International
Resolution	High	Medium	Medium	High	High
Scene Size	16.5 x 16.5 km	180 x 180 km	60 x 60 km	11 x 11 km	13.5 x 13.5 km
Bands	5	7	5	5	1
- multispectral	2.44 m	30.0 m	20 m / 10 m	4 m	Not available
- panchromatic	0.61 m	15.0 m	10 m / 5 m	1 m	1 m, 1.8 m, 4 m
- fused	0.61 m	15.0 m	/ 5 m	1 m	Not available
Stereo pairs	Available	Not available	Available	Available	Not available
Revisit frequency	1-3.5 days	16 days	2-3 days	1-3 days	2-4 days
Orbit altitude	450 km	705 km	832 km	681 km	460 & 600 km
Launch date	2001	1972-1999	1986-2002	1999	2000
Applications	Mapping, urban planning, agriculture and forestry, military, weather research, land management	Mapping, agriculture and forestry, military, change detection, disaster management, oceanography, water quality research, geology, vegetation mapping	Mapping, agriculture and forestry. Military, change detection, disaster management, land management	Mapping, agriculture and forestry, military, change detection, disaster management, vegetation mapping	Mapping, urban planning, military, change detection, disaster management, oceanography, geology

Appendix 2: Summary of the Characteristics of Ideal and Real Remote Sensing Systems (Lillesand, *et al*., 2004)

	Ideal Remote Sensing System	**Real Remote Sensing System**
Energy Source	• Energy over all wavelengths • Constant, known, high level of output regardless of time and place	• Passive remote sensing systems rely on reflected and/or emitted energy from earths surface • Spectral distribution of reflected sunlight and self-emitted energy is not uniform • Solar energy levels vary with time and location • Different earth surfaces/materials emit energy with varying degrees of efficiency • Need to calibrate for source characteristics case by case or deal with relative energy units at any given time or location
Atmosphere	• Atmosphere that does not modify the energy from the source – to or from the earths surface • Non interfering regardless of wavelength, time, place, and sensing altitude	• Compounds energy source variation, modifying strength and spectral distribution of the energy received by the sensor • Effect varies with wavelength, time and place • Calibration to eliminate or compensate for atmospheric effects is required especially for repetitive observations of the same area

Energy-matter Interactions	• Unique energy-matter interactions at the earths surface • Interactions that generate reflected and/or emitted signals • Interactions selective with respect to wavelength • Interactions known, invariant and unique to each and every earths surface feature type and subtype	• The earths surface/materials do not reflect and/or emit energy in a unique, known way • Spectral response patterns can be ambiguous with very different material types having spectral similarity • Current understanding of energy-matter interactions is not well developed
Supersensor	• A sensor highly sensitive to all wavelengths • Sensor that yields spatially detailed data on the absolute brightness (radiance) from a scene as a function of wavelength throughout the spectrum • Simple reliable, requiring minimal power and space, accurate and economical to operate	• An ideal sensor does not exist, no single sensor is sensitive to all wavelengths • Sensors have fixed limits of both spectral sensitivity and spatial resolution • Choice of sensor involves trade-offs between spectral sensitivity and spatial resolution • Sensors can have restrictive power, space and stability requirements which dictate the platform from which the sensor is operated and can make acquisition of remote sensing data expensive •
Data Processing	• Radiance-versus-wavelength response generated, transmitted to ground, geometrically and radiometrically corrected and processed into interpretable formats in real time	• Capability to generate data exceeds capacity to handle the data • Data processing requires thought, hardware, time experience and reference data

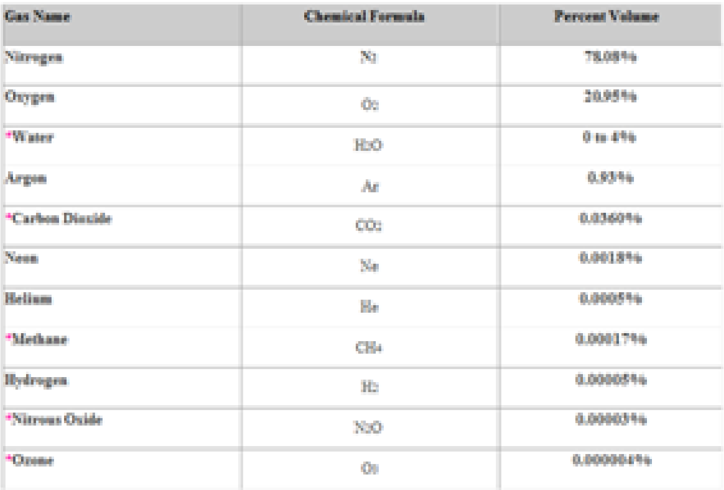
		- Usually not real time and not always for the exact area and time span of interest
Multiple Data Users	- Knowledgeable users in different disciplines - Same data in multiple formats for different users	- Data becomes information only if user knows how it was generated, how to interpret it and how best to use it - An understanding of problem at hand is necessary for the productive application of remote sensing methodology - No single combination of data acquisition and analysis procedures will satisfy the needs of all data users

Appendix 3: Summary of Indices Currently Used in Remote Sensing Research

Vegetation Index -Spectral Variable - Chlorophyll Index	**Correlation Characteristic**	**Reference**
Blue (B)	Yield	(Yang, *et al*., 2006)
Green (G)	Nitrogen content, yield, biomass Yield	(Bausch, *et al*., 2008) (Yang, *et al*., 2006)
Red (R)	Yield Leaf photosynthetic characteristics	(Yang, *et al*., 2006) (Tian, *et al*., 2005)
Near-infrared (NIR)	Yield Leaf quality, biomass	(Yang, *et al*., 2006) (Beeri, *et al*., 2005)
NIR/Blue (NB)	Yield	(Yang, *et al*., 2006)
NIR/Green (NG)	Nitrogen content, yield, biomass Yield	(Bausch, *et al*., 2008) (Yang, *et al*., 2006)
NIR/Red (NR)	Nitrogen content, yield, biomass Yield Leaf area index (LAI) Leaf photosynthetic characteristics	(Bausch, *et al*., 2008) (Yang, *et al*., 2006) (Jordan, 1969) (Tian, *et al*., 2005)
Normalised Difference Vegetation Index (NDVI) = (NIR-Red)/(NIR+Red)	Nitrogen content, yield, biomass Yield Yield Within-field spatial variability Leaf quality, biomass Yield Yield	(Bausch, *et al*., 2008) (Yang, *et al*., 2006) (Balaghi *et al*., 2008) (Bégué *et al*., 2008) (Beeri, *et al*., 2005) (Ren *et al*., 2008) (Jiang *et al*., 2003)
Integral NDVI	Yield	(Jiang, *et al*., 2003)
BNDVI = (NIR – Blue)/(NIR + Blue)	Yield	(Yang, *et al*., 2006)

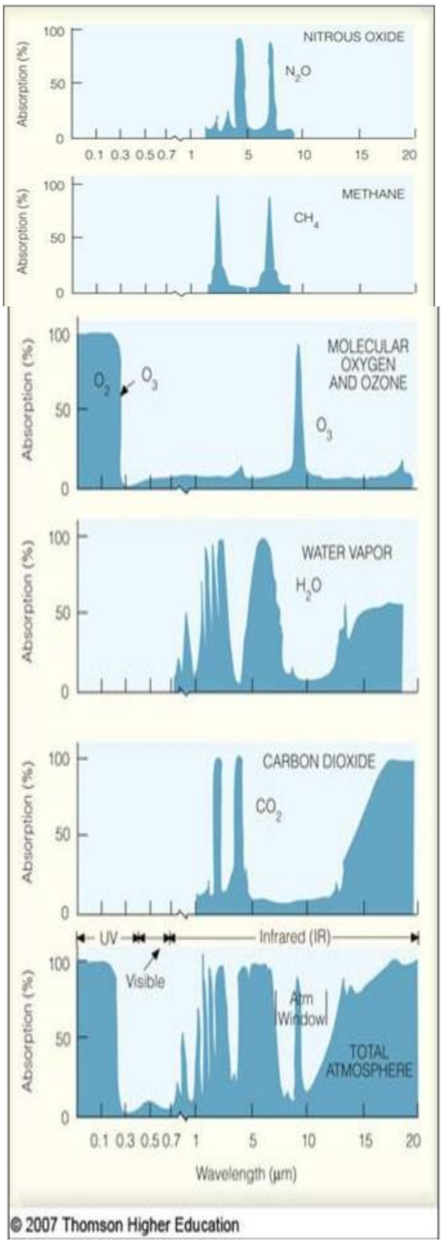
GNDVI = (NIR - Green)/(NIR + Green)	Nitrogen content, yield, biomass Yield Leaf quality, biomass	(Bausch, *et al*., 2008) (Yang, *et al*., 2006) (Beeri, *et al*., 2005)
Wide Dynamic Range Vegetation Index WDRVI = (a*NIR - Red)/(a*NIR + Red)	Vegetation Fraction (VF)	(Gitelson, 2004)
Normalised Green Red Difference Index NGRDI = (Green – Red)/(Green + Red)	Biomass, nutrient status Yield	(Hunt, *et al*., 2005) (Leon, *et al*., 2003)
Chlorophyll Index NG-1 = ((NIR/Green)-1)	Nitrogen content, yield, biomass	(Bausch, *et al*., 2008)
Band depth normalized to centre $BNC = (1-(R/R_i))/(1-(R_c/R_{ic}))$	Foliar biochemicals	(Curran, 2001)
Band depth normalized to area $BNA = (1-(R/R_i))/A)$	Foliar biochemicals	(Curran, 2001)
Transformed chlorophyll absorption ratio index $TCARI=3[(R_{700}-R_{670}) - 0.2 (R_{700} - R_{550}) * (R_{700} / R_{670})]$	Yield, crude protein	(Pettersson *et al.*, 2006)
Partial least squares factors PLS	Yield	(Ye, *et al*., 2008)
Optimised soil adjusted vegetation index $OSAVI = (1+0.16)(R_{800} - R_{670}) / (R_{800}+R_{670}+0.16)$	Yield, crude protein	(Pettersson, *et al*., 2006)

Appendix 4: Average composition of the atmosphere up to an altitude of 25km and the absorption effect by the atmosphere of radiation at different wavelengths.



Gas Name	Chemical Formula	Percent Volume
Nitrogen	N_2	78.08%
Oxygen	O_2	20.95%
*Water	H_2O	0 to 4%
Argon	Ar	0.93%
*Carbon Dioxide	CO_2	0.0360%
Neon	Ne	0.0018%
Helium	He	0.0005%
*Methane	CH_4	0.00017%
Hydrogen	H_2	0.00005%
*Nitrous Oxide	N_2O	0.00003%
*Ozone	O_3	0.000004%

* variable gases



© 2007 Thomson Higher Education

Appendix 5: Raw image of study area captured May 2009