Feasibility of Remote Sensing to Detect Psa Infected Vines or Vines at Risk of Psa Infection

Report on Project V11261 for Zespri Group Ltd.

James A. Taylor and Brett M. Whelan
Precision Agriculture Laboratory, Faculty of Agriculture and Environment, The University of Sydney

Submitted 14th December, 2012

EXECUTIVE SUMMARY

V11261- Feasibility of Remote Sensing to Detect Psa Infected Vines or Vines at Risk of Psa Infection

This project was a preliminary assessment of the ability of remotely sensed (satellite) imagery to detect vine stress related to *Pseudomonas syringae* pv. *actinidiae* infection in kiwifruit. A historical archived time-series of 5 images from the RapidEye satellite was obtained for the 2010-11 season. The imagery covered the period from early October 2010 to early February 2011. A sixth image from March 2011 was obtained but found to be too affected by cloud cover to be useful in the project.

The images were processed to correctly geo-reference them to orchard boundaries, remove any cloud artefacts, remove any potential edge-effects and to calculate a series of vegetative indices (VIs) from the satellite band information. In total there were 12 (VIs) calculated using various combinations of the 5 satellite bands (Blue, Green, Red, Red-edge and Near-InfraRed). The mean responses of all the VIs for each date were derived for each block within the orchards (KPINS). Information on orchard infection (at the KPIN level) was obtained from KVH Inc. and related to the block-level mean VI responses (i.e. in an infected orchard all blocks were considered to be infected).

A regression tree analysis was used initially as a data-mining exercise before stepwise binomial logistic regression was used to model the disease response (Infected or Non-Infected) using the VIs at the 5 different dates (early October to early February). The December image (26/12/2010) was the best at discriminating between dead vines, infected (stressed) but not dead vines and healthy vines using the regression tree analysis. However, in the proper jack-knifed modelling exercise, the early October image (02/10/2010) was the best indicator of infected vines in the 2010/11 season. This image was obtained at a time when the canopy was actively growing but had not yet reached full closure. In contrast, a later October image (21/10/2010) taken around full closure and flowering, poorly discriminated between infected and non-infected vines. The early season growth rate appears to be very indicative of infection, however, the vine physiology (and canopy response) at flowering may masking the disease response. The VIs from the images obtained later in the season (15/01/2011 and 01/02/2011) were not particularly useful in the modelling and may reflect either a variable rate of disease progression and expression during the season or within season management effects on the canopy response.

Further work is needed to better understand how the canopy response early in the season is affected by infection; however, early season images acquired before canopy closure appear to be a useful tool in identifying orchards where the disease expression is likely in the coming season. The utility of an early season image in disease modelling will assist in targeted sampling and proactive management of orchards within the current season.

Many different vegetative indices were derived and used in this work. VIs that incorporated the green, red-edge and/or NIR bands appeared to be the best performed. Imagery with these bands should be preferred in future acquisitions. The temporal response of the VIs was also derived on a bock-level basis; however this information was not useful in identifying infected orchards, probably for the same reasons that the late season images were also poor predictors.

There were some analysis issues identified in the project. Infection was recorded by the date of the positive laboratory result, rather than the date that the material was obtained from the orchard (or date that the disease was observed). Infection was also recorded at the KPIN level, but it was unclear if all bocks associated with a KPIN were infected at that time. At the start of the project the spatial data provided were poorly organised. We understand this situation now is much better. The project was also beset by contractual issues with the University of Sydney.

TABLE OF CONTENTS

LIST OF TABLES

Table 5: The VIs and associated date that were selected in the BLR jack-knifing. Only the results for the 1-4 parameter models are presented (The 5 and 6 were not superior and have a wide range of variables selected).

LIST OF FIGURES

Final Report: V11261- Feasibility of Remote Sensing to Detect Psa-V Infected Vines or Vines at Risk of Psa-V Infection

James Taylor and Brett Whelan, Precision Agriculture Laboratory, FAE, The University of Sydney

INTRODUCTION

Remote detection of *Pseudomonas syringae* pv. *actinidiae* (Psa) infection in kiwifruit vines and/or vines at risk of infection could assist in the management and control of this destructive plant pathogen. Multispectral remote sensing of plant pathogens, such as rice sheath blight has been previously demonstrated (Qin and Zhang, 2005). Remotely sensed imagery has also been effective in mapping and managing industry-wide pathogens, such as the Phylloxera threat to the grapevine industry of southern Australia (see http://www.healthyvines.com.au/Information.aspx). In these cases, remotely sensed images have been able to provide early detection of an infestation, which has promoted the rapid management/treatment of the disease. In New Zealand, the mapping of frost occurrence from satellite data has been used to determine sites where early autumn or late spring frosts are more likely to damage stonefruit trees, and subsequently predispose the trees to stone fruit blast (*Pseudomonas syringae*).

Remote sensing could be used in several ways to assist with management of Psa.

a) Identifying vines (orchards) that are 'at risk' of infection before they are infected. If plant stress affects the susceptibility of kiwifruit vines to infection then it is possible that the spectral signature from the canopy can identify vines/orchards that are under stress and more susceptible to infection (Jones and Schofield, 2008). This stress would be caused by an external factor (water stress, nutrient, disease pressure etc.) not directly associated with the Psa-V threat. This information could also be paired with other spatial information on existing sites of Psa-V infection and preferential distribution patterns for Psa-V (e.g. the direction of prevailing winds, overland water flow/runoff etc.) to assess the risk of infection.
b) Identifying vines that are in the early stage of infection but asymptomatic to the human eye. The reflectance from leaf tissue in the red-edge and NIR parts of the electromagnetic spectrum (EMS) is much more sensitive to changes in plant health (compared to reflectance in the visible part of the EMS that humans can see). Therefore images that incorporate information in the red-edge and NIR and beyond are able to identify early stages of plant stress. If the stress is caused by Psa-V then it may be possible to identify the infection before it is fully expressed and ensure that the full range of management options is available.

The potential for the spectral signature of the canopy to be affected by multiple environmental/physiological effects may pose some problems in direct identification of Psa-V infection. Multi-spectral (3-10 band) imagery in general has the ability to detect differences in canopy response, but lacks the resolution in the number of bands and width of the bands to directly determine the cause of the stress without some ground-truthing or *a priori* information. Hyperspectral imagery (20-200 bands) may be able to identify both the stress and the cause, however, this technology is currently considerably more expensive, not historically archived and computational more complex for analysis.

If the spatial and spectral band resolution of satellite imagery proves insufficient to detect Psa-V in the early stages of infection then ground-based sensing techniques may be better suited to solving this problem. Sankaran et al. (2010) provide a recent review on the advanced techniques available for detecting plant diseases that can be used in the field or on field collected samples.

The current project will use existing archived satellite sensory imagery collected from Te Puke to assess the feasibility of multispectral remote sensing to detect Psa-V infected vines, or vines at risk of infection. Several satellite monitoring systems are currently deployed to provide images of the Earth at various resolutions (ranging from >100 m to <2 m pixels). Imagery is routinely collected even if it is not ordered, creating large archives of data. The quality of this archived data is determined by the revisit time of the satellite(s) to a location and of the absence/presence of cloud cover. Of the commercially available satellite sensors, the Rapideye sensor has the best available archived time-series of high-resolution imagery for the Te Puke region in the 2010-11 growing season. (Other sensors investigated included SPOT, Landsat, ASTER, Ikonos/GeoEYE and Quickbird/WorldView.).

MATERIALS AND METHODS:

Data

Imagery

Multispectral (5-band Visible-NIR) imagery from the Rapideye satellite was acquired from AAM for October 2nd 2010, October 21st 2010, December 26th 2010, January 15th 2011, February 1st 2011 and March 15th 2011. These were archived 16-bit images. The images were orthorectified and delivered at a 5 m pixel resolution. The details of the 5-bands in the multi-spectral images are provided in Table 1.

The October 21st, 2010 image was chosen as a base layer (least cloud cover over the target area – Fig. 1) and the other images were geo-rectificed to this image. The 16-bit geo-rectified images were mosaicked (if necessary) and converted into 8-bit imagery (resolution of the radiometric band information). There were three dates (October 21st, December 26th and March 15th) that spanned the greater Te Puke area to the west of the target survey area. For the other three dates the imagery was centred only on Pukehina (see Fig. 1). The late season March 15th image contained a large amount of cloud cover. It was hoped that there would be enough orchard response between the cloud cover to make this image useful. However, during the analysis process this was found not to be the case, and the results involving the March 15th image are not included in the report. (NB: All imagery, shapefiles and data tables - raw and corrected – are appended in digital form).

Table 1: Details of the band information associated with the Rapideye imagery.

Band No.	Region	Band Range (nm)
Band 1	Blue	440-510
Band 2	Green	520-590
Band3	Red	630-685
Band4	Red-edge	690-730
Band5	Near Infrared	760-850

Change Detection Analysis.

A change detection analysis was performed by comparing the October 2nd, December 26th, January 15th, February 1st and March 15th image with the October 21st image. This looked for gross differences between the images on a pixel by pixel basis. Effectively this identified pixels with cloud effects in the images. (NB. there are other ways to do this, e.g. using a supervised classification, but the difference between the cloud and ground response makes this a very simple way of identifying the cloud affected pixels).

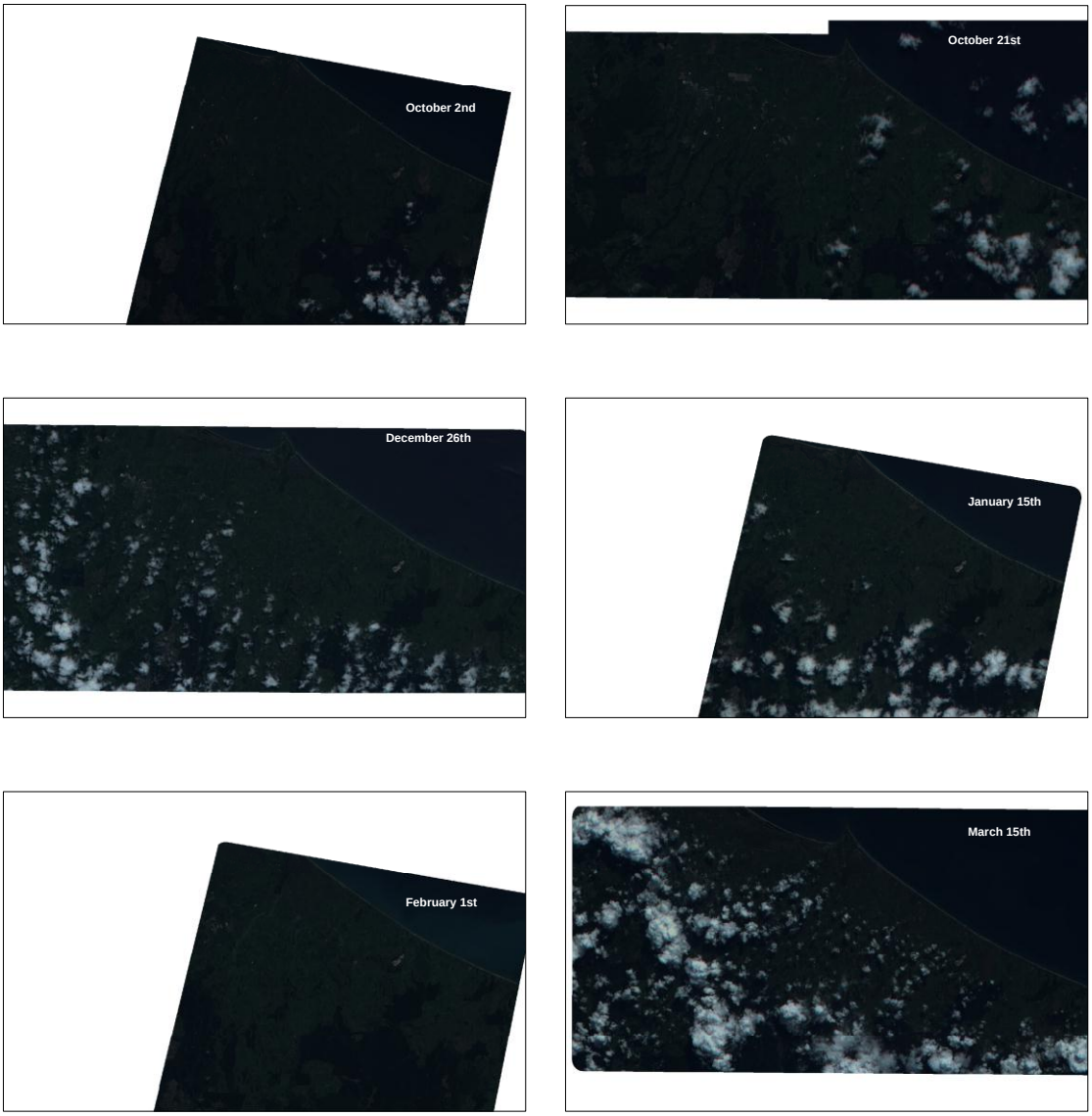


Figure 1. True colour images of the Rapideye imagery available over TePuke-Pukehina during the 2010-2011 growing season.

Vector data

Two vector layers (shapefiles) were received. A point file with information pertaining to the presence (or absence) of Psa-V in orchards was made available by Kiwifruit Vine Health Inc.. The Psa-V information was linked to a KPIN identifier, a single centroid coordinate (x,y) location and the date of Psa-V confirmation (i.e. date of a positive test result, not the date of collection or observation). A polygon file of orchard block boundaries was also received from AgFirst (Brad Stevens). Polygons were identified by a KPIN identifier and, in most cases, a block ID as well. The polygon file had two issues – many overlain and redundant polygons and generally poor rectification. The polygon file was edited to try and remove the redundant excess polygons (generally polygons labelled 'trelwyn' or 'draw' were kept rather than the 'import_zespri'). Polygons were also moved if needed to align more closely with the imagery. Where relevant, the individual polygons were assigned a new identifier consisting of the KPIN and Block ID (KPIN Block ID or KBI).

Data manipulation

Reclassing the Psa-V Response

The Psa-V data was provided with an approximate date of confirmation (not date of collection). Table 2 shows the classification generated by KVH Inc. based on the timing of a positive Psa-V test and the total number of KBIs and KPINs that were recorded. For the modelling these groups were reclassified and simplified. Groups A, B and C were merged to form an 'Early' (2010-11) positive Psa-V group. This corresponded with the period over which the imagery was collected. The D group may contain infections from 2011 but also from the new 2012-13 growing season. For this reason it was not included in the 'Early' group but designated as a 'Mid' period of infection. The E and F Groups would appear to be the result of 2012-13 infections and were designated as a 'Late' period of infection. Group G was retained as the 'Control' group with no infection.

Table 2: Infection detection periods with the original classification from AgFirst and the reclassification for this analysis. The total number of KPINs and KBIs associated with the data are also displayed[†].

Grouping	Date Range	No of KPINs with positive PSa-V	No of KBIs with positive PSa-V	ReClassed Groupings
A	Nov 10 - Mar 11	1	4	Early
B	Apr - June 2011	2	4	Early
C	July - Sept 2011	66	288	Early
D	Oct - Dec 2011	207	915	Mid
E	Jan - Mar 2012	41	124	Late
F	Apr - June 2012	8	35	Late
G	Control (No Psa-V)	66	217	No

† This is the total number of KPINs associated with the KVH Inc data which includes orchards outside of the image extent. These numbers will decrease when the Psa-V data is trimmed to the image extents.

ISSUE: Date of collection would have been a much more useful measure than date of a positive test. Work was contracted initially in August 2011, with imagery time-series from the 2010-11 growing seasons. However, only 8 KBIs (3 KPINs of Grouping A and B) were actually formally identified in this time period.

Derivation of vegetative indices

A 5 m grid was generated over the entire survey area. This is the same resolution as the pixel imagery i.e. a grid point = a pixel. The area-wide grid file was trimmed to points that lay within the corrected orchard polygons (from Agfirst). A 12 m buffer was then applied to remove points that were located close to an orchard boundary (within 2.5 pixels). This should ensure that only points over the kiwifruit canopy are included in the final 5 m grid (point) file. The intent is to avoid imagery band information that may contain information associated with errors in the polygon alignment, geo-rectification and orchard effects such as roads or windbreaks. Figures 2 and 3 show examples of the points trimmed to the orchard boundaries and then with the buffer applied. Some small blocks have very little data left after this, e.g. some of the blocks in the yellow outlined orchard (KPIN) in Fig. 3. These were trimmed later in the data manipulation process.

Figure 2: 5m grid cut to polygon boundaries

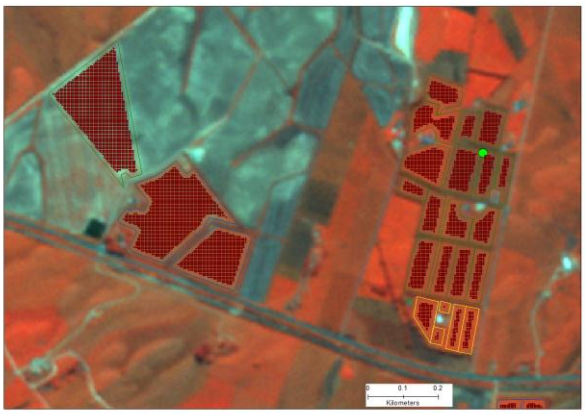


Figure 3: 12 m buffer applied to give final grid

The band information from each image was then extracted to the trimmed, buffered 5 m point file. Each grid point was given the attributes of the image pixels that it intersected with. Similarly, the attributes of the corrected orchard polygon file, including the KPIN and KBI identifiers, were assigned to each point based on the polygon that each point (pixel) was located within. Finally information relating to whether the pixel was affected by cloud cover was extracted to the points. The final point file therefore had coordinates (x,y), KPIN ID and KBI ID, band information from each of the 6 images and cloud cover for the 6 images. For some of the images (Oct 2nd, Jan 15th and Feb 1st) the point file contained points outside the imagery boundary.

The 5 m grid spreadsheet of the band information was imported into R. A script was run to generate 12 commonly used vegetative indices (VIs). These are summarised, with their formula, in Table 3. The VIs used various combinations of the 5 bands available from the Rapid-eye imagery, including the red-edge band (Band 4). Some of the VIs were originally developed to be derived from narrow band imagery. These are generally chlorophyll-based indices. They have been applied here to the broader bands of the Rapid-eye imagery to test if there is relevant information in these VIs for Psa-V detection. Recent work (Prof. A.A. Gitelson, University of Nebraka, *pers. comm*. Article submitted to the Agronomy Journal) has shown that these indices can be used effectively with broad-band multispectral imagery in broadacre crops.

The resulting spreadsheet contained the x, y coordinates of the trimmed and buffered pixel, KPIN and KBI ID, 12 VIs from each image (6 image dates) and whether or not the pixel was affected by cloud cover or not.

Extraction of mean response at the KBI level

The mean VI responses for each date and mean cloud cover for each KBI were calculated. KBIs with less than 9 points (pixels) were omitted at this stage. The data was then subset into a 'Season' and 'Spring' response. The 'Season' data set contained KBIs that were common to all the images from October 2nd, 2010 through to February 1st, 2011. This restricted the study area to Pukehina. The 'Spring' data set contained KBIs that were common to the October 21st and December 26th images, which covered both Te Puke and Pukehina. (NB. The data from Pukehina in the Oct 21st and Dec 26th images is therefore common to both datasets). Blocks (KBIs) with more than 5% cloud cover were removed from both the Spring and Season data sets. Finally the Psa-V infection associated with the relevant KPIN was assigned to the mean KBI responses in both the Spring and Season spreadsheets. These formed the final data sets for analysis. Both datasets are provided as associated text files.

Table 3: Vegetative indices (VIs) derived from the extracted image band information. The VIs were calculated at a pixel-scale (5 m pixel).

Name	Abbrev.	Formula	Reference
Simple Ratio	SR	NIR/Red	Rouse et al., 1973
Normalised Difference Vegetative Index	NDVI	(NIR - Red)/(NIR + Red)	Rouse et al., 1973
Green NDVI	GNDVI	(NIR - Green)/(NIR + Green)	Gitelson and Merzlyak, 1998
Red-edge NDVI	RENDVI	(NIR - Red-Edge)/(NIR + Red-Edge)	Gitelson and Merzlyak, 1994
Modified Red-Edge NDVI	modRENDVI	(NIR – Red-Edge)/(NIR + Red-Edge – (2 * Blue))	Datt, 1999
Enhanced Vegetative Index	EVI	(2.5 *(NIR - Red))/(NIR + (6 * Red) - (7.5 * Blue)+1))	Huete, et al., 1994
Enhanced Vegetative Index 2	EVI2	(2.5 *(NIR - Red))/(NIR + (2.4 * Red) + 1)	Huete, et al., 1997
Photosynthetic Vigour Ratio	PVR	Green/Red	SpecTerra Systems, 1999
Green Chlorophyll Index	GCI	(NIR/Green) - 1	Gitelson et al., 2003
Red-edge Chlorophyll Index	RECI	(NIR/Red-Edge) - 1	Gitelson et al., 2003
Triangular Vegetative Index	TVI	0.5*((120*(Red - Green) – 200 * (Red-Green)))	Broge and Leblanc, 2000
MERIS Terrestrial Chlorophyll Index	MTCI	(NIR – Red-Edge)/(Red-Edge-Red)	Dash and Curran, 2004

ISSUES: There is a large discrepancy between the spatial resolution of the Psa-V data and the VI data. The Psa-V is only recorded at a KPIN level; however the majority of KPINs describe multiple blocks. There was no information available on whether a positive Psa-V result for a KPIN was attributable to all or only a subset of blocks (KBIs) within a KPIN. Some KPINs are associated with large areas (34 of the 391 KPINs had more than 10 blocks associated with them). The modelling could be done at the KPIN or KBI (Block) level. The block level was chosen. This increases the number of data points available for modelling, especially in images that had a smaller spatial extent (less KPINs captured) and/or cloud effects. Modelling at the KPIN level was performed initially. However, there seemed to be a mixed response in some KPINS (affected and unaffected blocks) which created problems when modelling with a smaller sampling size. By modelling at the KBI level, there will be some blocks assigned with a positive Psa-V KPIN (due to a positive result in another block) that may not have been affected at that time. Thus false positives are expected. The converse, a positive Psa-V in a KBI but a negative KPIN Psa-V is less likely to have been recorded in the data, i.e. false negatives should be less prevalent than false positives. Some KPIN data (and Psa results) were also missing geolocations.

Data Analysis

Analysis was performed separately on the 'Season' and 'Spring' data sets. Analysis was confined to KBIs (KPINs) with either an 'Early' (Infection) or 'Control' (Non-Infection) response i.e. data are confined to KBIs that were likely to have infection in the season when the imagery was collected and sites with no known infection as of the end of the 2012 season. The 'Mid' and 'Late' responses were

not used as there is uncertainty, particularly with the 'Mid' response, on the actual state of infection during the imagery acquisition period.

Data-mining: Regression tree analysis

As a first analysis, a regression tree partitioning was performed on an individual image basis. The 12 VIs from each image were used to partition the variance between the 'Non-Infection' ('Control') class and the 'Infection' ('Early') class. The analysis was performed on a per date basis.

This is a first look at the data to gain an understanding of the data structure. Although several splits (nodes) were applied to each date, only the first split is recorded and shown (Table 4). Of interest was how the different VIs were selected and chosen and how well the regression tree 'fit' the categorical response. An example regression tree for the best fit is shown in the Appendix.

Modelling: Binomial Logistic Regression

Binomial Logistic Regression (BLR) was used to model the categorical 'Infection' (Early) and 'Non-Infection' (Control) Psa-V classes against the mean VI response within the KBIs. BLR was preferred to Discriminant Analysis as it more flexible in its assumptions, predominantly that predictors in the model (VIs) do not need to be normally distributed, linearly related or of equal variance (Tabachnick and Fidell, 1996). The BLR was performed in R using a jack-knifing approach for prediction and validation.

In both the Season and Spring data sets the ratio of 'Infection':'Non-Infection' KBIs was skewed (~2:1). The 'Infection' data were subset to the same size as the 'Non-Infection' data (n = 133 or 380 for the Season and Spring data sets). NB: This was done for each jack-knifing such that the Non-Infection sites were common to each iteration but the chosen 'Infection' sites differed. The combined data (either 266 KBIs for the Season Data or 760 KBIs for the Spring Data) was then randomly split into training and test subsets. The test subset was ~20% of the original data. A stepwise approach was then used to determine the best model for a 1-, 2-, 3-, 4-, 5- and 6-parameter BLR model on the training subset before the model was applied to the test subset. A maximum of 6 parameters was chosen to avoid over-fitting and to constrain the analysis to a method that would be sensible for industry application. The Chi^2 value, percentage of positive predictions (%PP) (Infection as Infection and Non-Infection as Non-Infection), percentage of False Positive (%FP) (Non-Infection predicted as Infection) and percentage of False Negatives (%FN) (Infection predicted as Non-Infection) were recorded for each iteration based on the fit to the test data. The % of times each VI (including date) was selected in the models was also recorded.

In evaluating the success of the imagery for monitoring Psa-V infection, the main interest is in the number of correct decisions but also in the False Negative percentage (%FN). The FN% indicates KBIs that were infected but were modelled as not-infected, i.e. escaped detection. The False Positive % (FP%) are the inverse, where stress symptoms are observed but not from Psa-V infection.

Modelling: Random Forest Regression

A Random Forest modelling approach was also applied to the same test and training data sets derived for the BLR (i.e. applied 100 times). This was trialled to test if there were non-linear combinations of the VIs that were useful for prediction. The BLR is a linear modelling process. Since the intent is to predict from the Regression Tree models, the Random Forest methodology was used (Breiman, 2001). Random Forests analysis is a more robust adaptation of Regression Trees analysis, thus more suited to deriving robust models for prediction. A classical Regression Tree analysis is more suited to data mining than data modelling.

The Random Forests models were restricted to a maximum of 5 splits and the end-nodes to a minimum size of 17. As for the BLR, the Chi^2, %PP, %FN and %FP were calculated from the jack-knifed results. The important predictors within the Random Forest modelling were also extracted. This analysis was performed in R using the randomForests package (Liaw and Wiener, 2002). The restriction to 5 splits was again chosen to avoid over-fitting and to constrain the Random Forest analysis to approximately the same parameterisation as the highest-order BLR models for comparison.

Temporal statistics associated with the Vegetative Indices

During the initial discussions on this project it was hypothesised that the temporal trend in the canopy response over the season may be indicative of Psa-V infection (or non-infection). For this reason, an effort was made to incorporate some information from the time-series, rather than just the individual 'snap-shot' imagery.

For each image, the calculated VIs were converted into a relative response (0-1) for each pixel. The relative KBI response was then extracted for each KBI for each image used in the 'season' data (Oct 2^{nd} to Feb 1^{st}). A linear model was fitted to each VI data set to model the response over time (Oct 2^{nd} considered Day 1). The gradient (β) and fit (r^2) were recorded for each VI in each KBI. The mean (μ) and standard deviation (σ) of each of the VI responses in each KBI over the 5 images was also extracted. These statistics (β, r^2, μ and σ) were analysed using the exploratory regression tree analysis and initially incorporated into the BLR and RF models described above. However, the data mining and modelling of the Psa-V data did not show any response to these statistics. Consequently, these varaibles have been omitted for the results and discussion section. A brief discussion on the possible reason for this result is given in the general discussion.

RESULTS AND DISCUSSION

Data Mining – Regression tree (using the Season Data)

There was no clear pattern in the preferred VIs over the imagery dates. However, indices that incorporated information within the green band (520-590 nm) and NIR appeared to be preferred. For example, the GNDVI was preferred to the RENDVI or conventional NDVI for the December 26^{th} image. The PVR, TVI and GCI ratios that were selected at the first split on other dates also include the green response. The December 26^{th} image was the best fit in sthe Regression Tree Analysis (actual tree shown in the Appendices). Interrogating this response, it appears that the infected vines at this stage of the season have reduced or absent canopy vigour. The KBIs in the tail (< 0.65) probably contain either dead or dying vines (Fig 4). This could be verified by checking if the identified KPINs in the histogram tail were actual orchards with very early infections/mortality (KPINs identified in the caption of Fig. 4).

Table 4: Splits associated with a regression tree analysis that used information from only ONE date. The main intent is to see if there is a pattern to the variables selected and the % of misclassified. NB. This is a data mining exercise and results are not always directly transferable to a data modelling process.

Image Date	First Split VI	Rule for Early	Early-Early	Early-NoPsa	NoPsa-NoPsa	NoPsa-Early	Misclassed %
Oct02	PVR	PVR < 1.49	130	6	93	84	28.75
Oct21	TVI	TVI < 5743	119	19	80	95	36.42
Dec26	GNDVI	GNDVI < 0.746	183	18	81	31	15.65
Jan15	EVI	EVI < 2.6423	140	9	90	74	26.52
Feb01	GCI	GCI < 5.979	127	5	94	87	29.39

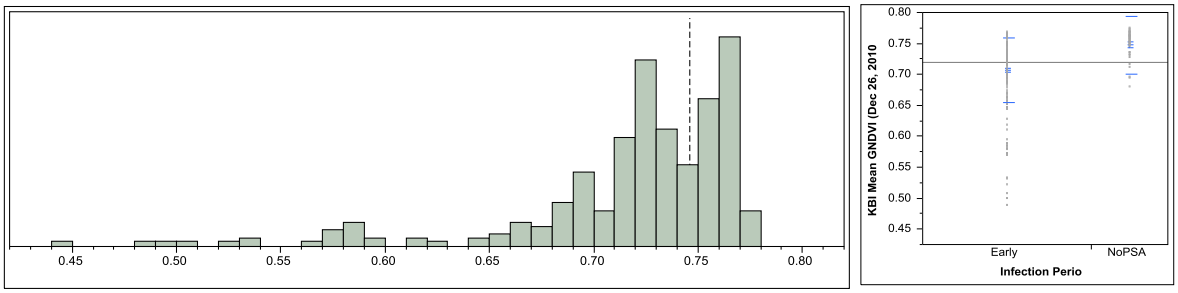


Figure 4: Left: Histogram of the GNDVI response from the Dec 26th image for the KBIs associated with the Early and Non-Infected groupings (Table 2). The dotted line indicates the splitting point from the regression tree analysis. Right: The data split into the two categorical responses, showing that the tail is associated with early infected orchards. The KPINs associated with these points (KBIs) (GNDVI < 0.65) are 1190, 1825, 1879, 1883, 2067, 2357, 2754, 3211, 6322, 6689, 7078, 8286, 9510

There is some indication of a bimodal population in the GNDVI response > 0.70. However, some of the blocks identified as infected are still exhibiting strong canopy vigour at this stage. This may have to do with the type and stage of infection (but neither was recorded in the data supplied). Likewise, some non-infected blocks have lower canopy vigour (health) (GNDVI < 0.75), which is possible due to other stresses or management not related to Psa-V.

The VIs from the later season images (Jan 15th and Feb 1st) or earlier season images (Oct 2nd and Oct 21st) were not as effective as partitioning the variance in the Infection response as the Dec 26th image. For the later season imagery, this may be due to additional 'noise' being introduced from mid-season management and/or the possibility that orchards identified as infected early in the season have been removed or are dead. This may cause them to exhibit a strong plant response from the renewal of ground cover in the orchard that is visible from an aerial/space platform. The imaging system is not able to distinguish between different species – only if there is a vigorous or poor plant response in the pixel. The early season imagery may be difficult to interpret due to differential early season growth rates and phenology stages (e.g. flowering) making the VI signal more difficult to interpret.

The strong difference between infected (and affected) vines midseason is expected, especially if the disease is causing vine mortality. From a management perspective, confirming with remote sensing what can be seen visually is of little value. It is preferable to be able to model an early (non-visible) change in plant response. The bimodal distribution in the GNDVI above 0.70 (Fig. 4) indicates that this may be possible.

Data Modelling – Binomial Logistic Regression (BLR) and Random Forests (RF)

The following results relate to the fits of the actual vs. predicted response from the jack-knife modelling of both BLR and RF. The BLR is modelled with 1-6 parameters (variables) while the RF is modelled with a maximum of 5 splits. Results are presented from both the the Seasonal and Spring datasets.

Season Data

The data was subset to produce evenly weighted positive (Infection) and negative (Non-Infection) responses. Under these conditions a random allocation of predictions should generate on average 50% of predictions to be correct. The results from the modelling (Fig. 5) show that the BLR correctly predicts between 75-79% of the time on average in the Season dataset. The incorporation of the

imagery VIs therefore improves the prediction of infected and non-infected orchards. Increasing the number of parameters in the BLR from 1 to 2 increased the quality of prediction, however further increases to 3-6 parameters did not significantly change the fit (Chi^2) or the number of correct predictions made. For management, the most important parameter is perhaps the percentage of False Negatives (%FN). This statistic indicates how often the model predicts an infected orchard as being non-infected. In this case, an infected orchard would not be identified as a problem. The inverse error is less important if the precautionary principle is applied. The percentage of False Positives (%FP) indicates how often a non-infected orchard is predicted as being infected. As indicated earlier, there are other factors (other disease/pest pressure, environmental pressure, management effects etc.) that produce a decrease in canopy health (vigour), which will mirror the response of Psa-V on the canopy. There is also a potential recording error from the recording of infection at a KPIN and not a block (KBI) level, which may increase the %FP. Therefore, False Positives are expected and the model results show that there is a higher proportion of FP predicted as hypothesised. Sampling in these orchards will verify if the depressed canopy response is due to Psa-V or an alternate pressure/error. With the precautionary principle applied it is better to verify that these orchards are not Psa-V orchards. The 2-parameter BLR model produced the lowest %FN. However, the quality in this prediction was at a cost in the %FP made by the 2-parameter BLR model.

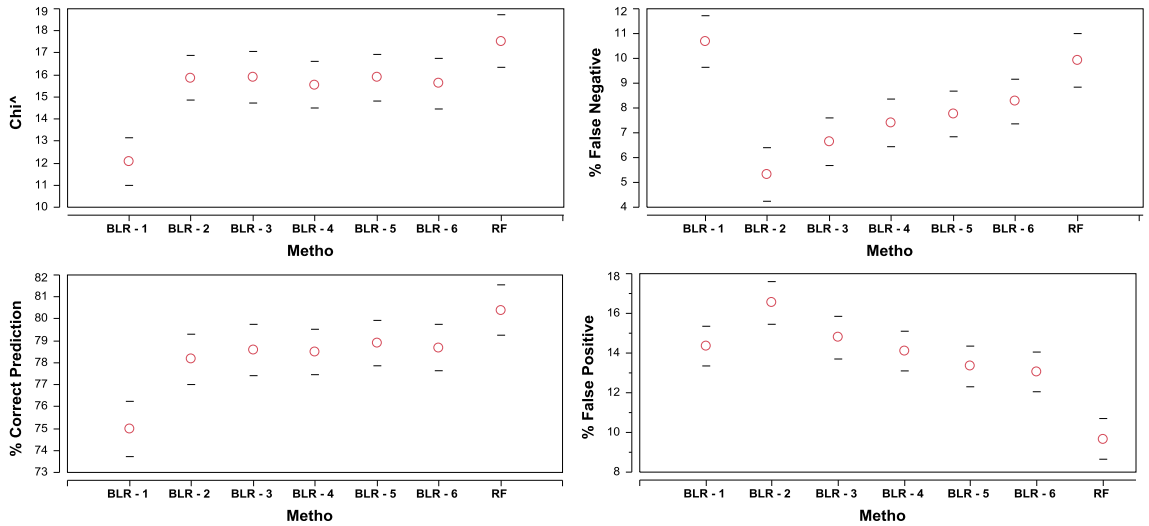


Figure 5: Plots of the Chi^2, percentage of False Negatives, False Positives and Correct Predictions from the Binomial Logistic Regression and Random Forests modelling on the **Season** dataset (all available images). The circle indicates the mean response across 100 jack-knifed iterations while the bars indicate the upper and lower 95% confidence levels from the jack-knifed results.

The random forest approach was used to test if there was a non-linearity involved that should be modelled. The RF was restricted to 5 splits (i.e. parameterised similar to the 5- and 6-parameter BLR). The overall predictive power was slightly better than the BLR but this was due to better predictions of correct outcomes and fewer FP. The FN outcome from the RF was worse than all but the 1-parameter BLR. This is arguably the statistic that should be minimised. On these results it appears that there is no advantage to pursuing a non-linear modelling approach.

Selected predictors in the BLR jack-knifing.

The 2-parameter BLR appeared to be the best model from the jack-knifing metrics. Table 5 shows which variables (VIs) were most commonly selected in the 1 – 4 parameter BLR models. The variables within the BLR-2 model were dominated by VIs derived from the early season (Oct 2nd) image. The first selection was nearly always the PVR layer. Even when the PVR from Oct 2nd was

chosen as the first variable in the stepwise the process, the second variable was usually another VI from Oct 2nd.

Table 5: The VIs and associated date that were selected in the BLR jack-knifing. Only the results for the 1-4 parameter models are presented (The 5 and 6 were not superior and have a wide range of variables selected).

Parameters	Variable and percentage of time it was selected at that level
1	PVR Oct 2nd (92%); GCI Feb 1st (7%); 1 others at 1%
2	RENDVI Oct 2nd (32%); RECI Oct 2nd (20%); GNDVI Oct 2nd (18%); Mod RENDVI Oct 2nd (13%); MTCI Oct 2nd (8%); 5 others at <= 4%
3	PVR Feb 1st (39 %); GCI Feb 1st (28%); EVI Dec 26th (23%); 5 others at <= 5%
4	EVI Dec 26th (47%); TVI Feb 1st (20%); GCI Feb 1st (10%); 12 others at <= 5%

The VIs from the Dec 26th image were not selected. As discussed previously, the December response may be spread across 3 different responses – dead/dying, infected and healthy. The 'tail' in the histogram (Fig. 4) may be causing problems with the modelling. The main VIs chosen within the RF models are given in the Appendices. In the RF model the December 26th VIs are more dominant, perhaps because the RF algorithm is better able to handle the tri-modal response.

The apparent value of the October 2nd image indicates that infection can be identified with a reasonable accuracy from the early season response, which is a very promising finding. In early October the vine canopy is growing rapidly and approaching full closure but is not at full closure. Infected vines would appear to be either retarded in their cane/leaf development and/or in the health (chlorophyll content) of the leaves, leading to a lower PVR (VI) response. The histogram of the PVR data from Oct 2nd that is used in the modelling is shown in Figure 6. Again a bimodal distribution is evident, with lower PVR values generally indicative of infected orchards. The 'tail' evident in the GNDVI histogram from Dec 26 is not evident, probably because the badly affected orchards are still developing at this stage and have yet to 'collapse'.

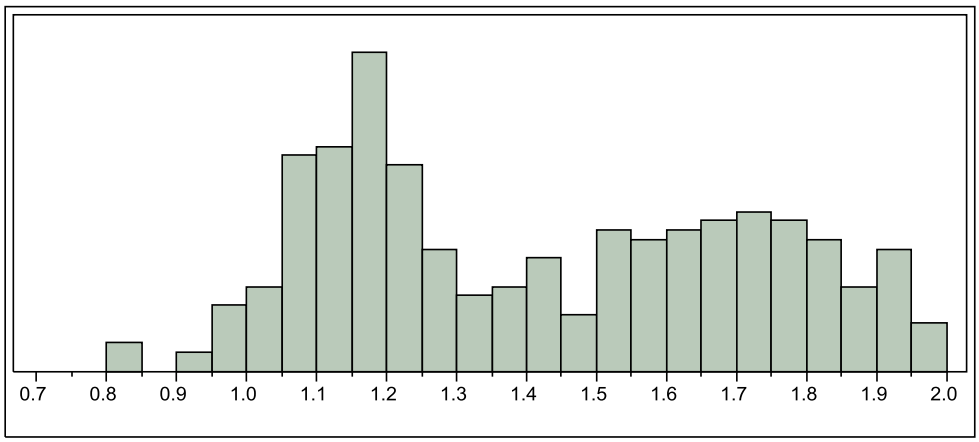


Figure 6: Histogram of the PVR response within KBIs from the October 2nd Image - restricted to the Early and Non-Infected data (Table 2). The bimodal distribution is clearly illustrated. Infected vines are typically located in the range of 0.9 – 1.4.

Spring Data (Oct 21st and Dec 26th only)

This is the same analysis as the previous section applied only to the October 21st and December 26th image extended over the Te Puke as well as the Pukehina orchards.

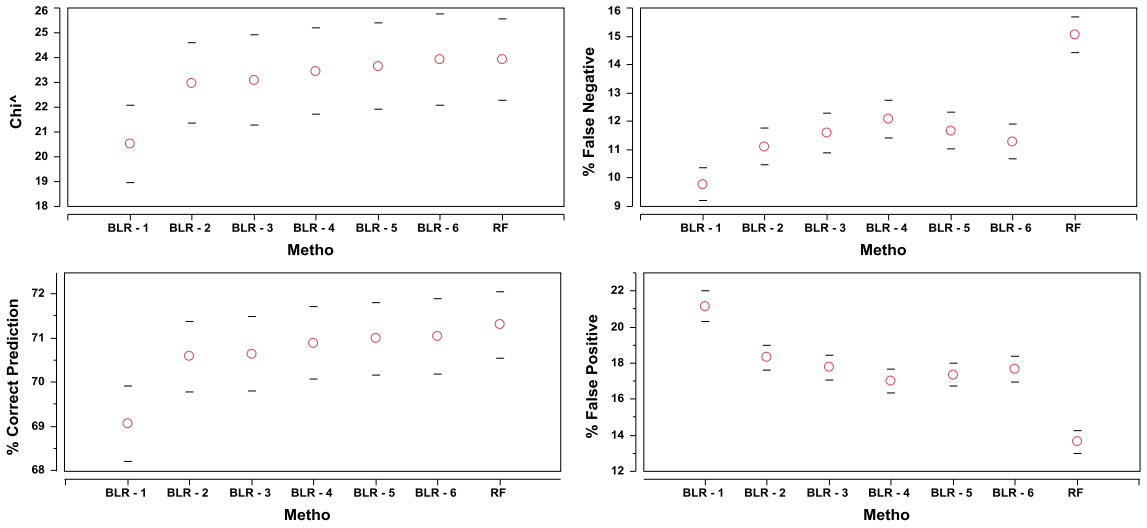


Figure 7: Plots of the Chi2, percentage of False Negatives, False Positives and Correct Predictions from the Binomial Logistic Regression and Random Forest modelling of the **Spring** dataset (Oct 21st and Dec 26th). The circle indicates the mean response across 100 jack-knifed iterations while the bars indicate the upper and lower 95% confidence levels from the jack-knifed results.

These results show a similar trend as the 'Season' results. The 2-parameter BLR improves the model fit (Chi^2) but adding more parameters beyond this does not indicate any improved prediction power. The 1-parameter model has the lowest percentage of False Negatives (note this is a higher percentage than those reported earlier), but the highest False Positives (analogous to the 2-parameter BLR model for the Season data). The main VIs selected in the 1-parameter BLR were GCI Dec 26th (87%) and RECI Dec 26th (13%). All variables selected in the jack-knifing process for the 1, 2 and 3-parameter BLR models were from the December 26th image. This again indicates the lack of value of the late October (21st) image. Since the models are only relying on the December response it is not surprising that there is little value in the higher-parameter models. During data-mining, the regression tree (on all data) identified the GNDVI ratio as dominant (Section 1). The BLR indicates that the GCI or RECI ratios are the best predictors. These are similar indices and again indicates a potential benefit to having the green and red-edge response (as well as the NIR) when imaging the canopy.

General Discussion

The Psa-V infection in kiwifruit can occur either superficially or systemically. The former relates to infection on the foliage, which is characterised by leaf spotting as the vine attempts to limit infection and eventually a systemic infection. The latter relates to infection within the vascular tissue, which leads to vine death. The result is that the progression of the disease may proceed in different forms and at different rates; this differential development makes it hard to separate out the response as the season develops. However, the early season response to the disease appears to be more stable and therefore more useful.

The October 21st image results

The imagery acquire soon after canopy closure and around flowering (Oct 21st image) was less useful than the images acquired before (Oct 2nd) or after (Dec 26th) this date (for both the Season and Spring datasets). It is difficult in this preliminary study to know if this is a result of full canopy closure (and the difference in Oct 2nd is due to a differential rate of canopy development) or a physiological response associated with flowering that alters the canopy (leaf) reflectance. Studies on how different stages of development affect canopy and canopy reflectance would be useful (for

understanding the response relate to the disease and for general understanding of how to deploy canopy sensors in kiwifruit).

Time-series analysis

A time-series of images was obtained for two reasons: i) to identify if there is a preferred time (stage of phenology) and ii) to determine if there the trend in the temporal canopy response was useful in identifying the disease. The temporal statistics were not useful in the modelling process and have not been presented. The hypothesis is that the expression of the disease follows no pattern with vine growth or climatic conditions, and may occur at different stages of growth. This makes it difficult to identify a stable temporal pattern in the evolution of the disease. The temporal vegetation response is also affected by within-season management and the potential for a background vegetation response in affected orchards.

CONCLUSIONS

The early season development of the vine canopy was a strong indicator of Psa-V infection. Consequently, low vigour responses in imagery around the start of October could be used to identify and target sampling and verification of disease spread. Later in the season, interpreting the canopy response is more complex, in part due to management and also a differential rate of expression in disease vines. The reason for the bimodal canopy response early in the season is likely to be Psa-V related, however, ground-truthing is required. Vegetative indices that incorporate the reflectance in the green portion of the electromagnetic spectrum appeared to work the best. The Red-edge response also was useful and may be preferable to the collection of the blue response.

Analysis and interpretation of remote-sensed imagery would be facilitated by having orchard data collected with better spatial and temporal information. In particular, the collection of information pertaining to blocks (or maturity areas) rather than enterprise (orchard) level data and data on the date of infection observed (visual or sample taken) rather than the date of a positive laboratory test. Information on the severity and/or type of infection when identified (internal (vascular), external (leaf) etc.) would help with the interpretation of the imagery and possible allow better modelling of the differential disease response.

A preliminary time-series analysis indicated no value to this approach, however this was preliminary and with better temporal (and spatial data) this may be worth revisiting.

REFERENCES

Broge, N. H. and Leblanc, E. 2000. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sensing of the Environment, 76, 156–172.

Dash, J. and Curran, P.J. 2004 The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing Volume 25(23), 2004

Datt, B., 1999. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests Using Eucalyptus Leaves. Journal of Plant Physiology 154:30-36.

Gitelson, A.A. and Merzlyak, M.N. 1994. *Spectral Reflectance Changes Associated with Autumn Senescence of* Aesculus Hippocastanum *L. and* Acer Platanoides *L. Leaves*. *Spectral Features and Relation to Chlorophyll Estimation*. Journal of Plant Physiology 143:286-292.

Gitelson, A.A., and Merzlyak, M.N. 1998. Remote sensing of chlorophyll concentration in higher plant leaves. Advances in Space Research 22:689-692.

Gitelson, A.A., Gritz, Y. and Merzlyak, M.N. 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithm for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology 160:271-282.

Huete, A., Justice, C. and Liu, H. 1994. Development of vegetation and soil indices for MODIS-EOS. Remote Sensing of Environment 29, 224–234.

Huete, A.R., H. Liu, K. Batchily, and W. van Leeuwen 1997. A Comparison of Vegetation Indices Over a Global Set of TM Images for EOS-MODIS. Remote Sensing of Environment 59(3):440-451

Jones, H.G. and Schofield, P. 2008. Thermal and other remote sensing of plant stress. General Applied Plant Physiology, Special Issue 34(1-2), p19-32

Liaw A. and Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), p18-22

Qin, Z. and Zhang, M. 2005 Detection of rice sheath blight for in-season disease management using multispectral remote sensing. International Journal of Applied Earth Observation and Geoinformation 7(2), p115-128.

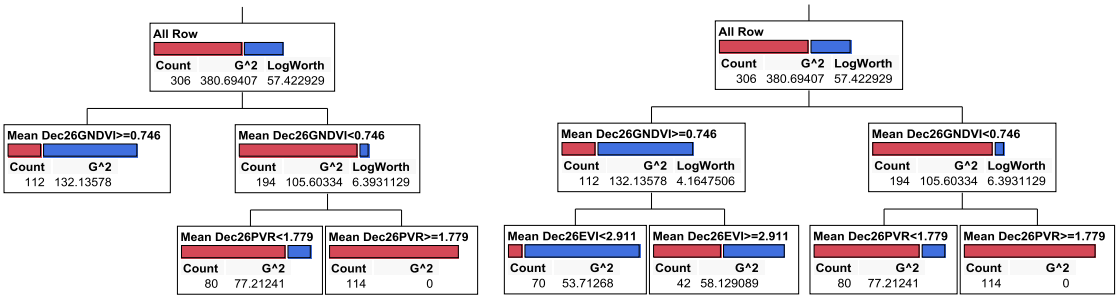
Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering, 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351 I: 309-317.

Sankaran, S., Mishra, A., Ehsani, R. and Davis, C. 2010. A review of advanced techniques for detecting plant diseases. Computers and Elctronics in Agriculture. 72, 1-13.

SpecTerra Systems (1999) Presentation and Analysis of Data, SpecTerra Systems Pty Ltd, Leederville, Western Australia, http://www.specterra.com.au/dmsv_data_frame.html

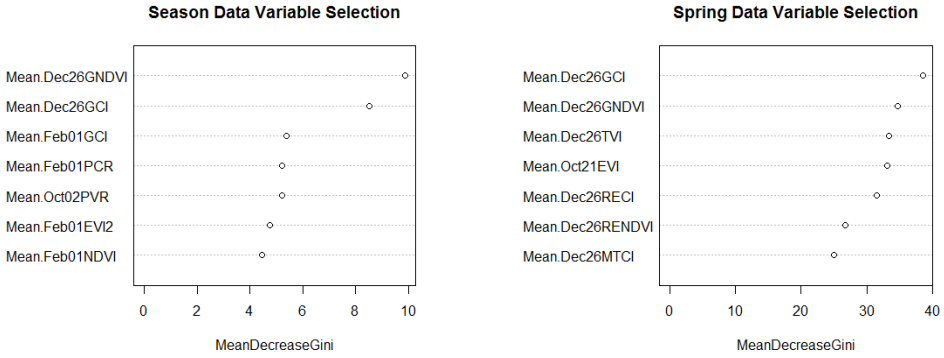
Tabachnick, B. and Fidell, L.1996. Using Multivariate Statistics, Third edition. Harper Collins

APPENDICES:



An example of regression tree analysis applied to the Early and No Infected Group. The categorical data is partitioned according to the mean response within each KBI in the December 26th image. The primary split was made based on the GNDVI response. A 2 (left) and 3 (right) split solution is shown here (very similar results for both explaining ~ 50% of the variance)

Data Mining - Importance of variables from the Random Forestmodelling.



Selected predictors in the BLR jack-knifing.

Season:

1: PVR Oct 2nd (92%); GCI Feb 1st (7%); 1 others at 1%

2: RENDVI Oct 2nd (32%); RECI Oct 2nd (20%); GNDVI Oct 2nd (18%); Mod RENDVI Oct 2nd (13%); MTCI Oct 2nd (8%); 5 others at <= 4%

3: PVR Feb 1st (39 %); GCI Feb 1st (28%); EVI Dec 26th (23%); 5 others at <= 5%

4: EVI Dec 26th (47%); TVI Feb 1st (20%); GCI Feb 1st (10%); 12 others at <= 5%

Spring Data (Oct 21st – Dec 26th only):

1: GCI Dec 26th (87%); RECI Dec 26th (13%)

2: PVR Dec 26th (77%); NDVI Dec 26th (20%); 2 others at <= 2%

3: GNDVI Dec 26th (41%); EVI2 Dec 26th (22%); RENDVI Dec 26th (12%); Mod RENDVI Dec 26th (9%); 6 others at <= 4%

4: RENDVI Dec 26th (31%); TVI Dec 26th (25%); PVR Oct 21st (14%); 9 others at <= 9%

Selected nodes in the RF jack-knifing.

Season:

1st Split: PVR Oct 2nd (41%); GCI Dec 26th (22%); GNDVI Dec 26th (21%); GCI Feb 1st (7%); GNDVI Feb 1st (6%)

2nd Split: GNDVI Dec 26th (32%); PVR Oct 2nd (20%); GCI Dec 26th (19%); GCI Feb 1st (14%); 6 others at <= 5%
3rd Split: GNDVI Dec 26th (27%); GCI Dec 26th (22%); GCI Feb 1st (19%); PVR Oct 2nd (14%); 6 others at <= 6%

Spring Data (Oct 21st – Dec 26th only):
1st Split: GCI Dec 26th (82%); GNDVI Dec 26th (18%)
2nd Split: GNDVI Dec 26th (18%); GCI Dec 26th (81%)
3rd Split: RECI Dec 26th (80%); RENDVI Dec 26th (15%)
4th Split: RENDVI Dec 26th (74%); RECI Dec 26th (19%)
(The selected variables are alternating. All but one model had GCI and GNDVI from Dec 26th as the first two nodes in the tree structure. Nodes 3 and 4 were either RECI or RENDVI.)