Modelling hill country pasture production: a decision tree approach

B.S. ZHANG, I. VALENTINE and P.D. KEMP
Institute of Natural Resources, Massey University, PB 11-222, Palmerston North
baisen.zhang.1@uni.massey.ac.nz

Abstract

Decision tree models were applied to predict annual and seasonal pasture production and investigate the interactions between pasture production and environmental and management factors in the North Island hill country. The results showed that spring rainfall was the most important factor influencing annual pasture production, while hill slope was the most important factor influencing spring and winter production. Summer and autumn rainfall were the most important factors influencing summer and autumn production respectively. The decision tree models for annual, spring, summer, autumn and winter pasture production correctly predicted 82%, 71%, 90%, 88% and 90 % of cases in the model validation. By integrating with a geographic information system (GIS), the outputs of these decision tree models can be used as a tool for pasture management in assessing the impacts of alternative phosphorus fertiliser application strategies, or potential climate change, such as summer drought on hill pasture production. This can assist farmers in making decisions such as setting stocking rate and assessing feed supply.

Keywords: data mining, decision tree, GIS, hill slope, rainfall

Introduction

In the last few decades, considerable research has been conducted on hill country pasture in New Zealand; most of it related to pasture production (e.g. Gillingham et al. 1998; Dodd & Ledgard 1999; Moir 2000). There is a large amount of data in the literature and kept by researchers in the form of raw or unpublished data. This data potentially provide a very useful resource to develop models for hill country pasture production.

Traditionally, there are two main approaches to modelling pasture production: using mechanistic (theoretical) models to reveal the causal factors determining pasture production (e.g. Moir *et al.* 2000), and using empirical (statistical) models, usually in one of the regression forms, to simulate pasture production and investigate the interrelationship between pasture and environmental factors (e.g. Lambert *et al.* 1983; Sala *et al.* 1988; Scott 2002). Mechanistic models, because of their strong theoretical bases, tend to be more general and robust than empirical models (Rickert *et al.* 2000). Empirical models, on the other hand, have the advantage of high predictive accuracy over mechanistic models for the areas in which the models are developed, and can also provide insight into the ecosystem processes if the input variables are properly chosen and ecologically meaningful (Guisan & Zimmermann 2000; Rickert *et al.* 2000).

With the development of computer technology, a new empirical modelling method, data mining (a process of posing various queries and extracting useful information, patterns, and trend often previously unknown from large quantities of data possibly stored in databases) (Thuraisingham 1999), has become popular due to its strong ability to predict new cases based on previously known information. Decision tree is one of the data mining methods and has been widely used in the social (Scheffer 2002) and medical sciences (Petitti 2000). It has also had increasing application in environmental modelling with considerable accuracy and ease of interpretation (e.g. Iverson & Prasad 1998; Vayssieres *et al.* 2000; Scheffer 2002; Yang *et al.* 2003). Its advantages are: (1) easily deals with non-linear responses, (2) easily incorporates nominal (such as grazing animals) and categorical (such as soil types) variables in a model, and (3) is able to indicate the relative importance of input variables and the interaction among them.

Decision tree models for annual and seasonal pasture production were developed for North Island hill country, using easily obtained climatic variables such as rainfall, temperature and solar radiation, management variables such as phosphorus (P) and nitrogen (N) fertiliser inputs and topographic variables such as slope, aspect as model inputs. The aim was to evaluate decision tree models for predicting pasture production and the effects of environmental factors and management inputs on pasture production.

Methods

Data

Data for annual and seasonal pasture production, fertiliser management (N, P fertiliser inputs) and topographic features (slope and aspect) were obtained from the literature and from researchers providing raw or unpublished data. Most climatic data (rainfall, temperature and global solar radiation) were obtained from the National Institute of Water & Atmospheric Research (NIWA), New Zealand. There were 21 input variables (Table 1) and 1900 samples in the dataset. These samples were from Whatawhata, Ballantrae, Te Kuiti, Riverside,

Table 1 Variables used in the decision tree analyses.

Model roles	Variable symbol	Units	Variable description
Input	N_fert	kg/ha/y	annual N fertiliser input
	P_fert	kg/ha/y	annual P fertiliser input
	P_fert5	kg/ha	5-year cumulative P fertiliser input
	temp_y	°C	annual mean daily temperature
	temp_sp	°C	spring mean daily temperature
	temp_su	°C	summer mean daily temperature
	temp_au	°C	autumn mean daily temperature
	temp_wi	°C	winter mean daily temperature
	rain_y	mm	annual rainfall
	rain_sp	mm	spring rainfall
	rain_su	mm	summer rainfall
	rain_au	mm	autumn rainfall
	rain_wi	mm	winter rainfall
	rain_warm	mm	sum of spring and summer rainfall
	solar_y	$MJ/m^2/d$	annual mean daily global solar radiation
	solar_sp	$MJ/m^2/d$	spring mean daily global solar radiation
	solar_su	$MJ/m^2/d$	summer mean daily global solar radiation
	solar_au	$MJ/m^2/d$	autumn mean daily global solar radiation
	solar_wi	$MJ/m^2/d$	winter mean daily global solar radiation
	aspect		1 (sunny), 0 (shady)
	slope	degree	hill slope angle
Target	annual production	kg/ha/y	annual aboveground DM per hectare
	spring production	kg/ha/season	spring aboveground DM per hectare
	summer production	kg/ha/season	summer aboveground DM per hectare
	autumn production	kg/ha/season	autumn aboveground DM per hectare
	winter production	kg/ha/season	winter aboveground DM per hectare

Summerlee, Waipawa, Mauriceville, Mikimiki, Gladstone, Whareama, etc.

Model development and validation

The decision tree models for annual, spring, summer, autumn and winter pasture production were analysed in SAS Enterprise Miner, Version 4.1 (SAS Institute 1999). The F test was used as the criterion in splitting the data which is the main procedure for developing a decision tree. In the development of the decision trees, the whole dataset was randomly partitioned into two parts: training data (70% of the total) and validation data (30% of the total). The decision trees were derived using the training data and fit was assessed using the validation data. The "best" tree which had the smallest average squared error (ASE) from validation data was selected.

After the decision tree models for annual and seasonal pasture production were developed, an empirical validation (Mitchell 1997) was implemented using the validation data. This empirical validation first used the decision tree models to predict the annual and seasonal pasture production for the validation data and graphed the predicted against observed means for observations had same prediction. The 95% confidence interval of the observation mean was set as an acceptable error for the prediction.

Results

Decision tree models

The decision tree model for annual pasture production is shown in Figure 1 and restricted models for spring, summer, autumn and winter pasture production are shown in Figure 2.

In each decision tree, the value in the upper-most rectangle is the predicted average annual or seasonal pasture production for the whole study area. The variable and the value in the rectangle below the upper-most rectangle are the most significant variable selected to split the tree and the split-point (value of that variable at which the split is made). Prediction goes to left-side branch if pasture with the splitting variable is less than the split-point, and goes to right-side branch if pasture with the splitting variable is equal to, or more than, the split-point. The predictions of pasture production are

Figure 1 The decision tree model for annual pasture production. Predicted productions are in the unshadowed rectangle (kg DM/ha/yr), splitting variables and split-points are in the shadowed rectangles. Prediction goes to the left-side branch when the splitting variable is less than the split-point, and goes to the right-side branch when the splitting variable is equal to, or more than, the split-point. See Table 1 for variable descriptions and units.

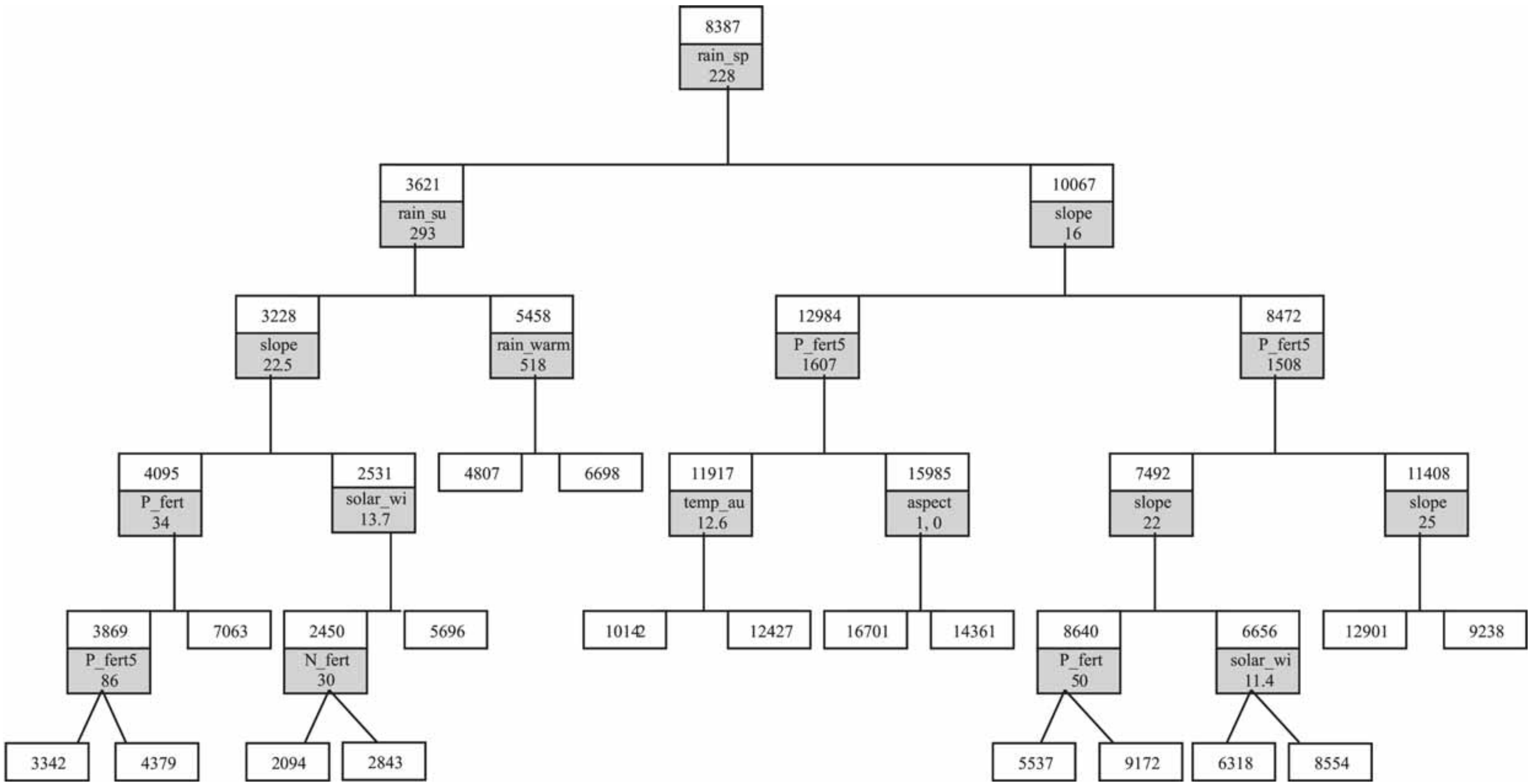


Figure 2 The first three levels of decision tree models for spring, summer, autumn and winter pasture productions. See the caption of Figure 1 for the description of decision tree interpretation.

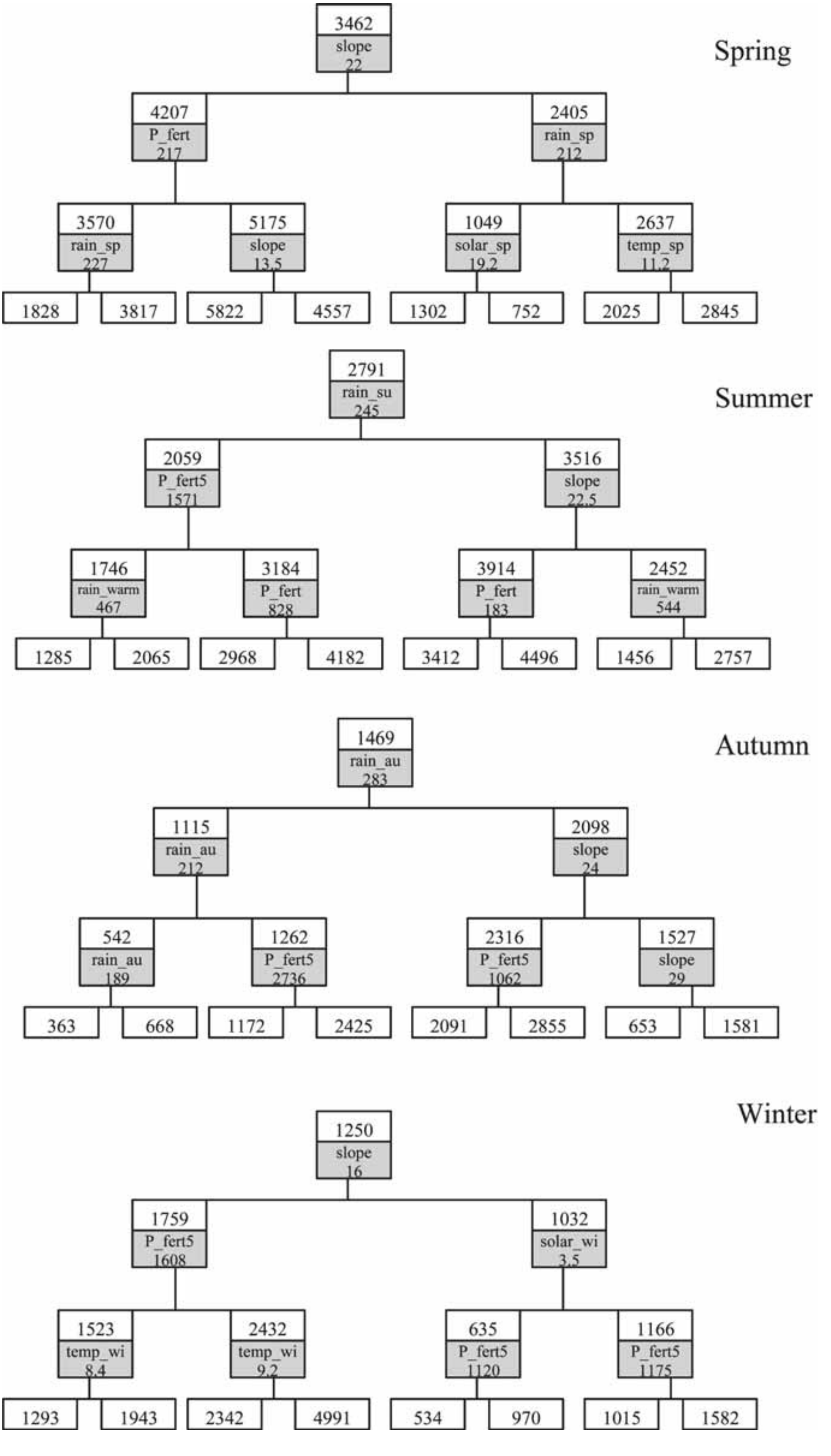
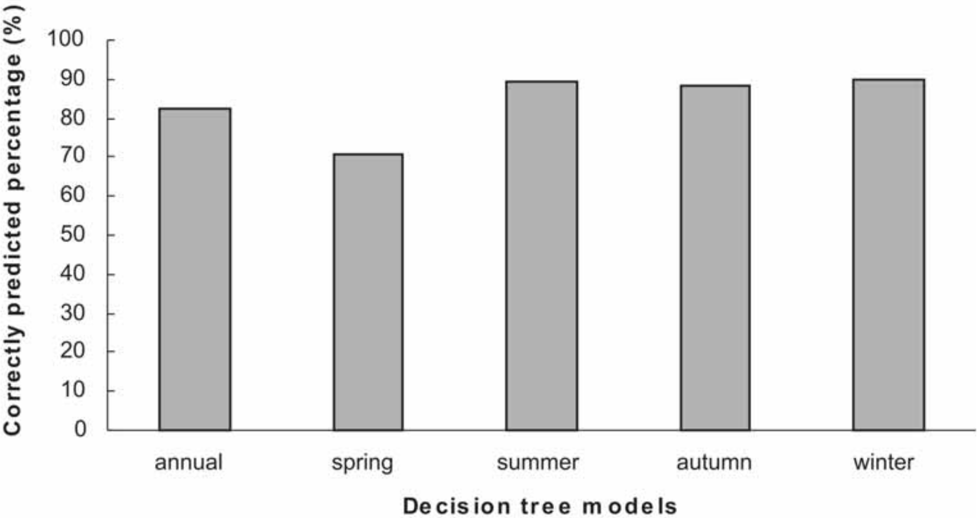


Figure 3 Results of empirical validations of the decision tree models for annual and seasonal pasture production.



made by a series of constraints defined by the input variables and their split-points. For example, for pastures with a spring rainfall equal to, or more than 228 mm, a slope less than 16°, a five-year cumulative P fertiliser input less than 1607 kg/ha, and a autumn mean daily temperature less than 12.6 °C, then the predicted average annual pasture production was 10142 kg dry matter (DM)/ha (Figure 1).

The relative importance of environmental and management variables on pasture production in the decision trees are ranked by the order they were selected in splitting the decision tree. The variable first selected was more influential than those selected after it. Spring rainfall was the most significant variable selected to split the decision tree model for annual production, followed by summer rainfall for those spring rainfall less than 228 mm and slope for those spring rainfall equal to, or more than 228 mm and so on. Slope was the most significant variable selected to split the decision tree models for spring and winter pasture production, and summer and autumn rainfall, respectively, were the most significant variables selected to split the decision tree models for summer and autumn pasture production.

In general, rainfall, slope, annual and five-year cumulative P fertiliser input were the most important variables influencing pasture production in hill country, while N fertiliser input, aspect, solar radiation and temperature also played a significant role in formulating both annual and seasonal pasture production.

Model validation

Empirical validation indicated that the decision tree models for annual, spring, summer, autumn and winter pasture production correctly predicted respectively 82.4%, 70.6%, 89.5%, 88.2% and 90.0 % of cases in the validation data (Figure 3).

Discussion

The high predictive accuracy in the model validation (Figure 3) shows that decision tree performed very well as a modelling approach in predicting pasture production. Other applications of the decision tree in classifying remote sensed vegetation data (Yang *et al.* 2003) and in predicting tree species abundance (Iverson & Prasad 1998) also indicated that the decision tree had very good performance as a modelling approach. The good performance of decision tree with respect of its predictive accuracy suggests that decision tree models can be used as a tool for evaluating pasture management options such as the impact of alternative P fertiliser application strategies, or for examining potential climate change, such as summer droughts, on hill pasture production. Additionally, by integrating the outputs of decision tree models with a geographic information system (GIS) and visualising predictions for a specific farm, the potential production of areas of the farm can be determined. A high resolution digital elevation model (DEM) for the farm or pasture under study is essential for this process and these are becoming widely available with the rapid development of GIS technology.

The hierarchical structure of the decision trees also clearly revealed the relative importance of input variables in influencing hill pasture production. It suggests that pasture production in hill country was firstly regulated by unmanageable climatic and topographic variables and was then influenced by manageable variables such as P and N fertiliser application.

Rainfall has been generally recognised as the key factor influencing the hill-pasture production (Lambert *et al.* 1983; Rickard *et al.* 1985; White 1990). The important role of spring rainfall as the determinant of annual pasture production indicated in our study was not previously fully recognised. Radicliffe & Baars (1987) revealed that spring and summer rainfall accounted for 60% of the variation in annual pasture production, but they did not analyse the effect of spring rainfall alone.

The effect of P fertiliser input on pasture production was greatly influenced by the hill slope. Pasture with a five-year cumulative P fertiliser input equal to, or more than, 1607 kg DM/ha had an average annual production of 15985 kg DM/ha when pasture had a slope less than 16^{o}, but only had an average annual production of 11408 kgDM/ha when pasture had a slope equal to, or more than 16^{o} and a similar P application (Figure 1). A more efficient application of P fertiliser could be applying P fertiliser on pastures with low slope. This is also a good choice for controlling no-point phosphorus pollution in water as the runoff on a high slope is higher than on a low slope (Battany & Grismer 2000).

One limitation of the decision tree models is that the effect of N fertiliser input on pasture production was not fully reflected. This is because the dataset used to develop the models only had limited samples with N fertiliser application. More data that incorporate the response of pasture production to N fertiliser are needed to improve the model performance in predicting pasture production under N fertiliser application.

ACKNOWLEDGEMENTS

The authors would like to thank Greg Lambert, Allan Gillingham, Mike Dodd, Maurice Gray, Phillipa Nicholas, Ignacio Lopez, Jim Moir, Jamie Blennerhassett, and NIWA, New Zealand for their data. Thanks to Greg Lambert and Allan Gillingham again for advice on this work, and the two referees for comments on the early edition of the manuscript.

REFERENCES

Battany, M.C.; Grismer, M.E. 2000. Rainfall runoff and erosion in Napa valley vineyards: effects of slope, cover and surface roughness. *Hydrological Processes 14*: 1289-1304.

Dodd, M.B.; Ledgard, S.F. 1999. Long-term (?) effects of withholding superphosphate application on North Island hill country: a 10-year update. *Proceedings of the New Zealand Grassland Association 61*: 69-73.

Gillingham, A.G.; Gray, M.H.; Smith, D.R. 1998. Pasture responses to phosphorus and nitrogen fertilizers on dry hill country. *Proceedings of the New Zealand Grassland Association 60*: 135-140.

Guisan, A.; Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling 135*: 147-186.

Iverson, L.R.; Prasad, A.M. 1998. Predicting abundance of 80 tree species following climate change in eastern United States. *Ecological Monograph 68*: 465-485.

Lambert, M.G.; Clark, D.A.; Grant, D.A.; Costall, D.A.; Fletcher, R.H. 1983. Influence of fertilizer and grazing management on North Island moist hill country 1. Herbage accumulation. *New Zealand Journal of Agricultural Research 26*: 95-108.

Mitchell, P.L. 1997. Misuse of regression for empirical validation of models. *Agricultural Systems 54*: 313-326.

Moir, J.L.; Scotter, D.R.; Hedley, M.J. 2000. A climate-driven, fertility dependent, pasture production model. *New Zealand Journal of Agricultural Research 43*: 491-511.

Petitti, D. 2000. Meta-analysis, decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine (2nd edition). Oxford University Press, New York, USA.

Radcliffe J.E.; Baars J.A. 1987. Temperate pasture production. pp. 7-17. *In*: Ecosystems of the World 17B. Ed. Snayden, R.W.

Rickard, D.S.; McBride, S.D.; Fitzgerald, P.D. 1985. The effect of soil moisture deficits on pasture yield. *New Zealand Agriculture Science 20*: 7-12.

Rickert, K.G.; Stuth, J.W.; McKeon, G.M. 2000. Modeling pasture and animal production. pp. 29-66. *In*: Field and laboratory methods for grassland and animal production research. Eds. Mannetje, L.; Jones, R.M. CAB Publishing, Wallingford, UK.

Sala, O.E.; Parton, W.J.; Joyce, L.A.; Lauenroth, W.K. 1988. Primary production of the central grassland region of the United States. *Ecology 69*: 40-45.

SAS Institute Inc. 1999-2001. Cary, NC, USA.

Scheffer, J. 2002. Data mining in the survey setting: Why do children go off the rails? *Research Letters in the Information and Mathematical Sciences 3*: 161-189.

Scott, D. 2002. Sustainability of New Zealand high-country pastures under contrasting development inputs. 8. Modelling sheep carrying capacity. *New Zealand Journal of Agricultural Research 45*:151-163.

Thuraisingham B. 1999. Data Mining: Technologies, Techniques, Tools, and Trends. CRC press, New York, USA.

Vayssieres, M.P.; Plant, R.E.; Allen-diaz, B.H. 2000. Classification trees: An alternative non-parametric approach for predicting species distribution. *Journal* *of Vegetation Science 11*: 679-694.

White, J.G.H. 1990. Hill and high country pasture. pp. 299-336. *In*: Pastures: their ecology and management. Ed. Langer, R.H.M., Oxford University Press, Auckland.

Yang, C.C.; Prasher, S.O.; Enright, P.; Madramootoo, C.; Burgess, M.; Goel, P.K.; Callum, I. 2003. Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems 76*: 1101-1117.