Modelling Community Productivity, Species Abundance and Richness in a Naturalised Pasture Ecosystem

Baisen Zhang

2005

Modelling Community Productivity, Species Abundance and Richness in a Naturalised Pasture Ecosystem

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Plant Science
at Massey University, Palmerston North
New Zealand

Baisen Zhang

2005

Abstract

This study focuses on modelling community productivity, species abundance and richness, and the impact of climate change and alternative phosphorous fertiliser application strategies on pasture productivity by integration of decision tree and regression modelling approaches with a geographical information system (GIS) in a naturalised hill-pasture ecosystem in the North Island, New Zealand, using data derived from research conducted on hill-pastures over the last several decades.

The results indicated that the decision tree models had a high predictive capability and clearly revealed the relative importance of environmental and management factors in influencing community productivity, species abundance and richness. Spring rainfall was the most significant factor influencing annual pasture productivity in the North Island, while hill slope was the most significant factor influencing spring and winter pasture productivity. Annual P fertiliser input and autumn rainfall were the most significant factors influencing summer and autumn pasture productivity, respectively. For species functional group abundance, soil Olsen P was the most significant factor influencing the relative abundance of low fertility tolerance grasses (LFTG) and moss, while soil bulk density, slope and annual P fertiliser input were the most significant factors influencing the relative abundance of legume, high fertility response grasses (HFRG) and flatweeds, respectively. Legume abundance was the most significant factor influencing species richness in the hill-pasture. Species richness increased with an increase in legume abundance and showed a tendency for a hump-shaped response. Grazing animal species also had a significant effect on species richness; pasture grazed by sheep had more species than pasture grazed by cattle. Climate change scenarios of temperature increases of 1-2 ^{o}C and rainfall changes of -20% to +20% would have a great impact (-46.2% to +51.9%) on pasture production in the North Island. Pasture in areas with relatively low rainfall had a higher response to increased P fertiliser input than pastures in areas with a relatively high rainfall.

In conclusion, the integration of a GIS with decision tree and regression models in this study provided an approach for effective predictive modelling of community productivity, species abundance and richness in the hill-pasture. This modelling approach can also be used as a tool in pasture management such as in assessing the impact of climate change and alternative fertiliser management on pasture production.

Key words: climate change, community productivity, data mining, decision tree, functional group, hill-pasture, geographical information system, GIS-based modelling, multivariate analysis, pasture production, predictive modelling, regression, relative abundance, species richness.

Acknowledgements

I would like to express my sincere thanks to my supervisors Dr. Ian Valentine and Assoc. Prof. Peter Kemp for their guidance, encouragement, friendship and endless reading of my manuscripts throughout my doctoral study.

My thanks to those who generously provided their unpublished and/or raw data to me for developing my models; without this data, I would definitely have not finished my thesis before the scheduled time. They are Dr. Jamie Blennerhassett, Dr John Brock, Dr. Mike Dodd, Dr. Allan Gillingham, Mr Dave Grant, Mr. Maurice Gray, Dr. Greg Lambert, Dr. Ignacio Lopez, Dr. Jim Moir, and Dr. Phillipa Nicholas. Thanks to National Institute of Water & Atmospheric Research (NIWA) New Zealand, with the assistance of Ms. Elaine Fouhy, for providing climate data. I am very appreciative of their contributions toward my study and thesis.

I would like to acknowledge the support of people who helped me in various ways during my study. Thanks to Prof. Richard Aspinall, Assoc. Prof. Alex Chu, Dr. Allan Gillingham, Dr. David Gray, Assoc. Prof. Mike Hedley, Dr. Greg Lambert, Dr. Kerry Harrington, Prof. John Hodgson, Assoc. Prof. Cory Matthew, Prof. Russ Tilman and Dr. Todd White for discussion and comments on my work and thesis manuscripts. Thanks to Mr. Mike Tuohy for providing the DEM of the North Island, Mr. Matthew Irwin and Mr. Des Costall for their help with GIS analyses, Dr. Siva Garnish and Mr. Judi Scheffer for advice on statistics, and Dr. Guisan Antonia and Dr. Louis Iverson for advice on issues of developing ecological models. Thanks to the administrative staff and computer lab staff in Institute of Natural Resources for their assistance during my study.

My thanks to my fellow students Tehseen Aslam, Bhoj Bahadur, Zulfiqar Butt, Zaker Hussain, Jian He, Xiongzhao He, Mofakkarul Islam, Congqi Liu, Qianhe Liu, Ernest Okorley, Tara Pande, Endang Savitri, Haoran Sun, Zulkefly Sulaiman, Edmundo Viegas, Minhua Xu and others for their friendships and/or helpful discussions on my thesis. Special thanks to Entin Daningsih for frequent discussion on statistics, and Tri Priantoro on GIS.

I would like to express my thanks to the Massey Doctoral Scholarship and the Helen E Akers Ph.D. Scholarship for financial support during this study.

I wish to thank my parents, my sister and brothers for their love, encouragement and support. Very special thanks to my wife Airong and my son Liyu (Oliver) for their love, patience and understanding during my three years' study.

Table of Contents

List of Tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 1. General Introduction

良好的开始是成功的一半。

A good start guarantees half of a success ----- *Chinese aphorism*

Hill-pastures in New Zealand are naturalized vegetation that originated from clearance of native podocarp rainforest, fern or scrub and the over-sowing with introduced pasture species by early European settlers (Hilgendorf 1936). There are about five million hectares of hill-pasture in New Zealand, which account for nearly one third of the total pasture area, mainly located in the North Island (White 1990).

Considerable research has been conducted on these hill-pasture in the last few decades (e.g. Suckling 1975; Radcliffe 1982; Lambert *et al.* 1996; Nicholas 1999; López 2000; Moir 2000; Gillingham 2001; Blennerhassett 2002); most of it related to pasture productivity, species abundance and composition, and species richness. There is a large amount of data either in the literature or kept by researchers in the form of raw or unpublished data.

Though these data were analysed for the purposes they were collected, they still contain much more information than has been revealed due to the research objectives, or due to a meta-analysis not being applied to the dataset. Therefore, the whole dataset provides the potential to develop models that investigate how environmental factors, such as topography and climate, and management factors, such as fertilizer application strategies, influence pasture ecosystem processes over space and time.

With the advances in data analysis technology, a new modelling approach - data mining, has become available and its effectiveness in querying and extracting useful information, patterns and trends, often previously unknown, from large quantities of existing data has been demonstrated (Thuraisingham 1999). Decision tree (also called classification and regression tree) (Breiman *et al.* 1984; Vayssieres *et al.* 2000) is one of these data mining methods and has been widely used in the social (Scheffer 2002) and medical sciences (Petitti 2000). It has also had increasing applications in environmental modelling showing considerable accuracy and effectiveness in investigating interactions between model target (dependent) and input (independent) variables (Iverson & Prasad 1998; Vayssieres *et al.* 2000; Scheffer 2002).

Geographic information system (GIS) is an information technology with the capacity to store, analyze and display both spatial and spatial-related data (Parker 1988). GIS has become an increasingly valuable tool in the computer-based modelling of environmental and ecological processes. The integration of GIS with environmental and ecological models has been widely used to model species distribution a (Franklin 1995; Guisan & Zimmermann 2000), community productivities 1997; Li *et al.* 1998), impacts of climatic changes on vegetations (Ivers 1998; Tan & Shibasaki 2003) and watershed hydrology (Johnston 1998; Wadsworth & Reweek 1999), for example. This integration greatly enhances the spatially investigative and predictive capability of environmental and ecological models.

Although a large number of studies have been conducted on hill-pasture in the last several decades, most of them focused on a single location and/or only lasted for a few years; there is a need to investigate hill-pasture ecosystem processes by explicitly considering spatial and temporal heterogeneity in environment and management. Hence the objectives of this study are to investigate and model community productivity, species abundance and species richness at a large spatial and temporal scale by applications of decision tree and other approaches with integration of GIS. Insights obtained from these models will be helpful in enhancing the understanding of hill-pasture ecosystem processes at landscape or regional scale and in improving pasture management.

The following studies were conducted to achieve these objectives:

1) Modelling the community productivity of hill-pasture in the North Island, New Zealand.

2) GIS-based predictive modelling of community productivity in a hill-pasture ecosystem.

3) Modelling the impacts of climate changes and alternative phosphorus (P) fertiliser management on hill-pasture production in the North Island, New Zealand.

4) Modelling functional group abundance of plant species in a hill-pasture ecosystem.

5) Modelling and investigating species richness patterns and underlying factors in a hill-pasture ecosystem.

These studies are presented in Chapter 3 to 7 and are prepared in the style of a journal paper ready for publication. In Chapter 3, a decision tree is developed to investigate and model pasture community productivity at a regional scale (the North Island, New Zealand) and is compared with a regression model to assess its performance. In Chapter 4, a decision tree model is integrated with a GIS to spatially predict patterns of pasture community productivity in a hill-pasture. The utility of this spatial modelling approach in pasture management is discussed. Chapter 5 assesses the potential impacts of climate changes and alternative P fertiliser management on pasture production for the whole North Island, New Zealand. In Chapter 6, decision tree models are developed to investigate and model functional group abundance of plant species in a hill-pasture and are integrated with a GIS to generate predictions for functional group abundance over space. In Chapter 7 patterns and processes of species richness in a hill-pasture are investigated using decision tree and other approaches.

References

Blennerhassett J.D. (2002) *Pasture growth constraints on dry steep east coast hill country*. Ph.D. thesis, Massey University, New Zealand

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) *Classification and regression trees*. The Wadsworth statistics/probability series. Chapman & Hall. Inc. New York, NY.

Franklin J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499

Gillingham A.G. (2001) Precision management of fertiliser application to pasture. *First Australian Geospatial Information and Agriculture Conference,* Sydney, Australia, 17-19 July 2001, 534-541

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

Hilgendorf F.W. (1936) *Pasture plants and pastures of New Zealand.* Christchurch, New Zealand, Whitcombe and Tombs Limited

Iverson L.R., Martin E.D., Scott C.T. & Prasad A. (1997) A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio (U.S.A.). *Landscape Ecology*, 12, 331-348

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

Johnston C. (1998) *Geographic information system in ecology*. Oxford, Malden, MA: Blackwell Science

Lambert M.G., Barker D.J., Mackay A.D. & Springett J.O. (1996) Biophysical indicators of sustainability of North Island hill pastures. *Proceedings of the New Zealand Grassland Association*, 57, 31-36

Li J., Liang T. & Chen Q. (1998) Estimating grassland yields using remote sensing and GIS technologies in China. *New Zealand Journal of Agricultural Research*, 41, 31-38

López I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment*. Ph.D. thesis, Massey University, Palmerston North, New Zealand

Moir J.L. (2000) *Investigations into the influence of fertilizer history and climate regime on the soil fertility, soil quality and pasture production of Wairarapa hill country soil*. Ph.D. thesis, Massey University, Palmerston North, New Zealand

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems*. Ph.D. thesis, Massey University, Palmerston North, New Zealand

Parker H.D. (1988) The unique qualities of a geographic information system: a commentary. *Photogrammetric Engineering and Remote Sensing*, 54, 1547-1549

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. (1982) Effects of aspect and topography on pasture production in hill country. *New Zealand Journal of Agricultural Research*, 25, 486-496

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

Suckling F.E.T. (1975) Pasture management trials on unploughable hill country at Te Awa. 3. Result for 1959-69. *New Zealand Journal of Agricultural Research*, 3, 351-436

Tan G. & Shibasaki R. (2003) Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. *Ecological Modelling*, 168, 357-370

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

Wadsworth R. & Reweek J. (1999) *GIS for ecology: an introduction*. Harlow: Addison Wesley Longman

White J.G.H. (1990) Hill and high country pasture. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

Chapter 2. Literature Review

孔子曰: “温故而知新，可以为师矣”（《论语・为政》）.

To review what we have known is a key to learn what we don’t know----- *Confucius*

2.1. Introduction

There has been increasing concern over the impacts of environmental changes and intensive land use on the ecosystem processes (Chapin III *et al.* 2000; Tan & Shibasaki 2003). In a pasture ecosystem, these impacts may result in a decrease in productivity and replacement of high feed quality species by poor species (Baars *et al.* 1990; Langer 1990; Campbell 1996). Modelling pasture ecosystem processes is an efficient way to investigate the interrelationship between pasture and environmental and management factors, and to quantify the potential impacts which may result from environmental changes and intensive land use. Insights obtained from the model outputs could also provide implications for better management strategies for sustainable development. In this chapter I aim to review the research outcomes for hill-pasture in New Zealand and the main aspects of theories, technologies and research in environmental and ecological modelling. Specifically, this review covers the following: (1) hill-pasture in New Zealand, (2) plant ecology in relation to species diversity, distribution and abundance, (3) GIS and GIS-based environmental modelling, and (4) data mining and decision tree.

2.2. Hill-pasture in New Zealand

2.2.1. Introduction

Hill-pastures in New Zealand are naturalized vegetations that originated from the clearance of native podocarp rainforest, fern or scrub and the oversowing with introduced pasture species in hill-country by early European settlers (Hilgendorf 1936). There are about 5 million hectares of hill-pasture in New Zealand, which are nearly 1/3 of the total pasture and are mainly located in the North Island. Hill-country farming is a major enterprise and a key aspect of New Zealand economy, accounting for nearly one third of the total export earnings derived from agriculture (Trustrum *et al.* 1983; White 1990).

Up until the 1950's, hill pastoral farming was primarily concerned with maintaining the introduced pasture species such as browntop (*Agrostis capillaris*), chewings fescue (*Festuca rubra*), crested dogstail (*Cynosurus cristatus*) and white clover (*Trifolium repens*) at existing levels of fertility, which were often low (White 1990). There were very few legumes in these pastures and improvement of soil fertility through nitrogen fixation was minimal. Improvement of hill-pastures occurred with the advent of aerial top dressing in the 1940's, when superphosphate and lime were commonly applied. The combination of superphosphate application and oversowing legume seed resulted in vastly improved pastures. The superphosphate fertilizer stimulated legume growth which, in turn, fixed atmospheric nitrogen and encouraged grass growth. This improvement of hill-pastures meant a shift from low feed quality species like chewings fescue, browntop and danthonia, to more desirable species such as perennial ryegrass (*Lolium perenne*) (White 1990; Nicholas 1999).

Pastures on hill-country are diverse both regionally and on a single farm. This diversity is caused by climatic factors such as temperature and rainfall, edaphic factors such as soil moisture and soil fertility, and biotic factors such as grazing behaviour and pasture management (Suckling 1975; Lambert & Roberts 1978, White *et al.* 2004). These factors are further modified by altitude, slope, aspect, and micro-topography within a hill farm (Lambert *et al.*1986; White *et .al.* 2004). Annual rainfall in hill-pasture varies from 300 to over 2000 mm with most hill-country in the 600-1500 mm range. Temperature also varies dramatically from the north to south; it varies not only with latitude and altitude, but also with season and aspect (Radcliffe & Lefever 1981; White 1990).

Topographic variation in slope and aspect in hill-country is a characteristic of all hill-pasture and resulted in the existence of diverse habitats differing in climatic, edaphic, and biotic characteristics. In southern latitudes, hillsides tending towards more northerly aspect receive greater net radiation, are warmer, and generally drier (Radcliffe & Lefever 1981). Depending on season and aspect, land slope also modifies net radiation received and increasing slope usually reduces soil moisture status. By influencing animal grazing and excreta patterns, aspect also influences soil nutrient status. Within a hillside, further variation in soil nutrient levels may occur because of differences in land slope and the associated development of stock tracks and camps that result from animal movement and resting habits. Accumulation of soil nutrients occurs on stock tracks and camps as a result of nutrient transfer patterns. The combined effects of these climatic and soil differences result in contrasting pasture species composition and production patterns (Gillingham 1982;

Sheath & Boom 1985b). In most parts of New Zealand hill-pasture, a seasonal lack of water is a major factor limiting pasture production, and is a primary cause of year-to-year variability in yield (Rickard *et al*. 1985). Although large areas of the North Island receive more than 1000mm rainfall, late spring and summer usually have less rainfall than autumn and winter, resulting in seasonal soil water deficits for plant growth. Also, factors such as shallow soils with low water-holding capacity, sloping soil with high run-off, and sunny faces with above-average solar radiation and evapotranspiration, can also result in water deficits (Kerr *et al*. 1986; Barker & Dymock 1993). Soil moisture levels are always higher on south-than north-facing slopes. On north-facing slopes soil moisture levels were always higher on easy than steep slopes, but on south-facing aspects the opposite was often the case (Gillingham *et al*. 1998).

2.2.2 Species composition and diversity in hill-pasture

Species composition in hill-pastures is more diverse than lowland pastoral systems in New Zealand. Lambert *et al*. (1986) identified approximately 20 species in a survey of a North Island, New Zealand hill-pasture. Some species typical of these pastures were grasses such as ryegrass (*Lolium perenne)*, browntop (*Agrostis capillaris)*, sweet vernal (*Anthoxanthum odoratum)*, crested dogstail (*Cynosurus cristatus)*, Yorkshire fog (*Holcus lanatus)*, poa (*Poa* spp.), red fescue (*Festuca rubra)*, danthonia (*Rytidosperma* spp.), legumes such as white clover (*Trifolium repens)*, suckling clover (*Trifolium dubium)*, and lotus (*Lotus pedunculatus)* and other species such as flatweeds (e.g. *Plantago lanceolata*, *Hypochaeris radicata* and *Leontodon taraxacoides*), moss (*Muscii* spp.), *Centella uniflora*, and *Nertera setulosa* (Lambert *et al*. 1986) (Table 1). This diversity is brought about through both management of the pasture and the wide variety of environmental conditions present on most hill farms (Nicholas 1999).

Based on their morphological traits, responses to fertility, and other features, species in hill-pasture are generally classified into several functional groups: low fertility tolerant grasses (LFTG), such as browntop, sweet vernal, crested dogstail, danthonia, red fescue; high fertility responsive grasses (HFRG), such as ryegrass, Yorkshire fog, poa; legumes, such as white clover, suckling clover, and lotus; flatweeds, such as *Plantago lanceolata*, *Hypochaeris radicata* and *Leontodon taraxacoides*; and moss (Lambert *et al.* 1986; Nicholas 1999) (Table 1).

Table 1. Plant species and functional groups of at Ballantrae (after Lambert *et al.* (1986) and Nicholas *et al.* (1998))

Functional group	Species
HFRG	*Lolium perenne, Holcus lanatus, Poa pratensis, Poa annua and Dactylis glomerata.*
LFTG	*Agrostis capillaris, Anthoxanthum odoratum, Cynosurus cristatus, Rytidosperma* spp., *Festuca rubra.*
Legume	*Trifolium repens, Trifolium dubium, Trifolium subterraneum and Lotus pedunculatus.*
Flatweeds	*Plantago lanceolata, Hypochaeris radicata, and Leontodon taraxacoides.*
Moss	*Musci* spp.

Climate and soil fertility are the main determinants of species composition in hill-pastures. These differ on a regional basis, and between hillsides and micro-sites because of variations in slope and aspect. As a result, diverse plant communities are a common feature of these hill environments (Suckling 1975). As soil fertility and moisture increase, pastures move from being predominantly low fertility tolerant grasses and/or annual legume towards associations dominated by ryegrass and white clover. Within a given climate and soil fertility status, species composition can be modified by different stock classes, stocking rates, and grazing managements (Levy 1970; Clark et al. 1984; Sheath & Boom 1985b). The greatest effect of grazing on composition appears to be mediated by differences in the grazing behaviour of animals. Within sheep-dominated systems, increases in cattle and/or goats reduce some weed species, increase white clover content, and ultimately lead to rapid pasture improvement (Levy 1970; Suckling 1975; Clark *et al.* 1984; Sheath & Boom 1985b). Pasture grazed by goats develop a Yorkshire fog-white clover association with strong white clover growth on all slope classes, unlike sheep grazed pasture, which contains little white clover and is dominated by perennial ryegrass and browntop (Clark *et al.* 1984).

Management factors that affect species composition include stock type, grazing regime, defoliation intensity, stock treading intensity, and pasture improvement techniques such as fertilization and oversowing (Nicholas 1999). Topography was found to have the greatest influence on pasture composition because of its influence on stock management, animal behaviour and nutrient transfer. For example, ryegrass is dominant on camps and tracks, as are other high fertility responsive grasses such as Yorkshire fog, poa and cocksfoot. Their content decreases with increasing slope. Browntop and other low fertility tolerant grasses such as sweet vernal, crested dogstail, chewings fescue, danthonia; legumes such as suckling clover and *lotus pedunculatus*; other species such as catsear (*Hypochaeris radicata* L.), hawkbit (*Leontodon taraxacoides* [Vill.]), plantain, moss and also dead material increase with increased slope in hill-pastures (Lambert & Roberts 1978; Sheath & Boom 1985a; Lambert *et al.* 1986; Liu 1996). Composition differences caused by soil type were not great, except where they appeared to be related to soil moisture characteristics (Grant & Brock 1974). Landscape patterns of vegetation diversity in pasture communities are determined by the local-scale processes, influenced by the availability of nutrients (White 2004).

2.2.3 Pasture productivity in hill-pasture

Pasture growth can be considered to be determined by physiographic factors (climate, topography), soil factors (nutrient, moisture), and pasture factors (species, density). All may operate directly on pasture growth or indirectly through modifying the influence of one or more other factors (Ledgard *et al.* 1982). In hill-pasture, climate factors like temperature and rainfall, soil fertility, especially P and N contents, topographic factors like slope and aspect, and species composition determine the quantity and quality of pasture production, and its seasonal distribution (Lambert *et al.* 1983; Gillingham *et al.* 1998; Moir *et al.* 2000). Pasture production in the North Island is usually in a range of 2000 – 13000 kg/ha, depending on the climate, topography and pasture management (Zhang *et al.* 2004).

Aspect can have a marked influence on pasture productivity. Pasture production was higher on the south aspect than the north aspect (Lambert 1978), but some other work showed that the converse situation was true because of better winter production (Gillingham *et al.* 1998). Usually on the North Island, shady (southerly) aspects have higher productivity than sunny (northerly) aspects during periods of moisture stress, whereas sunny aspects produce more herbage at other times. In drier hill-country, larger differences among aspects in summer might be expected, because of a longer period of moisture stress (Lambert *et al.* 1983).

Hill slope, due to its important role in redistributing available moisture, fertility and solar radiation, greatly influences hill-pasture productivity. It was revealed that hill slope was positive correlated with soil bulk density and negative correlated with soil Olsen P, soil total nitrogen and soil moisture (Lopez 2000; Lopez *et al.* 2003). A strong negative relationship between slope and annual production is usually observed in hill-pasture (Gillingham & During 1973; Lambert & Roberts 1978).

A major factor causing variability in pasture growth rate is the fertility of soil. On hills this effect is further complicated by changes in slope and aspect of the soil surface (Gillingham & During 1973). Increased P content stimulates legume growth which in turn fixes atmospheric nitrogen, which encourages grass growth. Withholding P application could result in 29-35% less annual pasture production and 54-72% less legume production (Roach *et al.* 1996). Soil phosphorus tends to have an interactive effect with soil moisture (rainfall) on pasture production, as rainfall increases, the size of the pasture response to per unit of Olsen P (a measure of soil phosphorous content) also increases (Moir 2000). Pasture production also responds to N fertilizer, indicating that N is a strong limiting factor for hill-pasture productivity, despite the contribution of N from legumes fixation (Luscombe & Grant 1981; Gillingham *et al.* 1998; Blennerhassett 2002).

2.3. Plant ecology in relation to species diversity, distribution and abundance

2.3.1. Biodiversity

Biodiversity is the variety of life, and can be defined at three levels: genetic diversity, species diversity and landscape diversity (Huston 1994). Among these three levels, species diversity is the most commonly used in ecological study, and can be measured by species richness (species number), evenness (species relative abundance) and some diversity indices which combine richness and evenness (Purvis & Hector 2000). Humans are concerned about biodiversity because the world is losing species due to human induced environmental deterioration and habitat fragmentation. The impacts of losing and changing biodiversity in ecosystems are far-reaching on both ecosystem functioning and the services that humans derive from ecosystems (Chapin III *et al.* 2000; Tilman 2000).

2.3.2. Diversity and ecosystem function

Ecosystem functioning refers to the biogeochemical activities of an ecosystem, or the flow of materials (nutrients, water, atmospheric gases) and processing of energy (Ehrlich & Wilson 1991). The effect of biodiversity on ecosystem functioning, especially the diversity and community stability relationship, has long been a controversial topic in ecological study (Huston 1994; McCann 2000; Naeem 2002a). Before 1970, ecologists believed that a more diverse community enhanced ecosystem stability. "Simple community were more easily upset than that of richer ones; that is, more subject to destructive oscillation in populations, and more vulnerable to invasion" [Charles Elton in (Tilman 1996)]. Since 1973, theoretical studies indicated that diversity tends to destabilize community dynamics (May 1973). Recent studies on biodiversity and ecosystem functioning (Hooper & Vitousek 1997; Lonnie 1997; Loreau *et al.* 2001; Naeem 2002a) with evidences from the experiments manipulated to directly control biodiversity tended to agree that diversity within an ecosystem is, on average, correlated positively with community stability (Naeem & Li 1997; McCann 2000; Tilman 2000). These studies also show that biodiversity is positively related to plant community productivity, that is, greater biodiversity leads to greater productivity (Caldeira *et al.* 2001; Loreau *et al.* 2001; Mouquet *et al.* 2002; Naeem 2002b).

Species diversity has functional consequences because the number and kinds of species present determine the traits of organisms that influence ecosystem process (Chapin III *et al.* 2000). Species traits may mediate energy and material fluxes directly or may alter abiotic conditions (for example, limiting resources, disturbance and climate) that regulate process rates (Huston 1994; Schulze & Mooney 1994). The components of species diversity that determine this expression of traits include the number of species (species richness), their relative abundance (species evenness), the particular species present (species composition), the interaction among species, and the temporal and spatial variation in these properties. In addition to its effects on current functioning of ecosystem, species diversity influences the resistance and resilience of ecosystems to environmental changes (Huston 1994; Schulze & Mooney 1994; Chapin III *et al.* 2000). On the other hand, though a large number of traits of plant species may be functional within a community, fewer traits are directly associated with species abundances over a larger landscape (Mabry *et al.* 2000). Limiting environmental factors (e.g., minimum temperature) may "filter" species by constraining their occurrence in a community to those species having evolved particular traits that allow them to overcome the constraints (Grime *et al.* 1997). Diaz et al. (1998) found a strong association between a regional climate gradient and measured plant traits which suggested the effect of environmental filtering. These filters can influence species composition and diversity at multiple and hierarchical scales of time and space along environmental gradients (Poff 1997).

Many hypotheses have been generated to explain the relationship between biodiversity and ecosystem functioning (Huston 1994; Naeem 1998; McCann 2000; Tilman *et al.* 2001). The insurance hypothesis states that biodiversity provides an "insurance" or a buffer, against environmental fluctuations, such that it maintains the stability (resistance or resilience) of the ecosystem. Therefore, increasing diversity increases the odds that at least some species will respond differentially to variable conditions and disturbances, and greater diversity increases the chances that an ecosystem has functional redundancy containing species that are capable of functionally replacing important species (Naeem 1998). The niche complementarity hypothesis says that productivity can be greater in a community with higher diversity because of the inter-specific differences in resource requirements, and differences in spatial and temporal resource and habitat use, or from the positive interaction, that is, more efficient resource exploitation (Tilman *et al.* 2001).

All species in a community are not equal. The loss or addition of species with certain functional traits may have a great impact, and others have little impact, on a particular ecosystem process, but different processes are likely to be affected by different species.

Functional diversity has greater impact on ecosystem processes than species diversity (Tilman *et al.* 1997).

A functional group, which is a common measure of functional diversity, is a set of species that have similar effects on a specific ecosystem-level biogeochemical process, such that substitution among species within a functional group has "minor" impacts on ecosystem processes. Membership in functional groups can vary depending on the biogeochemical process of interest (Hobbie *et al.* 1994; Vitousek & Hooper 1994), and they use the same resource in such a way that they could potentially compete very intensely with one another (Schulze & Mooney 1994).

The functional characteristics of the component species in any ecosystem are likely to be at least as important as the number of species for managing critical ecosystem processes and services (Hooper & Vitousek 1997). Some species have such an important role in ecosystem processes that they are termed keystone species. Keystone species are those species whose activity and abundance determine the integrity of the community and its unaltered persistence through time, that is stability. Experimental removal of a keystone species should result in the loss of some species and replacement by others (Bond 1994).

Most ecosystems are non-additive function of the traits of two or more species, because interaction among species, rather than the presence or absence of species, determines ecosystem characteristics. Species interaction, including mutualism, trophic interaction (predation, parasitism, and herbivory), and competition may affect an ecosystem directly by modifying pathways of energy and material flow, or indirectly by modifying the abundance or traits of species with strong ecosystem effects (de Ruiter *et al.* 1995). The functional characteristics of the component species in any ecosystem are likely to be at least as important as the number of species for managing critical ecosystem processes and services (Hooper & Vitousek 1997).

Contrasting to biodiversity and ecosystem functioning theory, traditional community ecology considers species diversity as a dependent variable controlled by abiotic environmental conditions and ecosystem-level constraints, and primarily focused on dominant species as biotic controllers of ecosystem functioning (Loreau *et al.* 2001). Since most recent studies on biodiversity and ecosystem functioning are conducted by manipulating species diversity as a designed variable and making all other factors constant, they have been criticized, especially the positive biodiversity and productivity relationship, as an experimental artefact (Grime 2002). Beside, many studies in natural ecosystems also lack the evidence to support this relationship (Grime 2002). Whether the positive relationship between biodiversity and ecosystem functioning is an artefact of manipulated experiments or a discovered truth still needs further investigation.

2.3.3. Patterns of species diversity

Species diversity is related to the spatial scale, that is, the area within which species diversity (e.g. richness) is measured. There is an increase in species diversity with increased sample area. The pattern of how diversity increases with increasing sample area is called species/area curve (Huston 1994). Environmental heterogeneity is considered the main reason determining this diversity/area relationship; increasing sample area includes additional habitat types with groups of different species (Judas 1988; Huston 1994). To characterise diversity on different scale, Whittaker (1972) proposed alpha (α), beta (β) and gamma (ν) diversity. Alpha diversity is within-area diversity, measured as the number of species occurring within an area of a given size (Huston 1994). It is usually called local diversity. Beta diversity designates the degree of species change along a given habitat or physiographic gradient. Gamma diversity usually refers to overall diversity within a large region (Cornell 1985; Heywood 1995).

Species diversity is often related to the resource availability in the environment. In environments with very low concentration of resources or nutrients, species richness is generally low because only few species can survive. In such a resource-poor environment, an increased availability of limiting resources will increase growth and survival of several species. Conversely, in an environment with high concentrations of resources, a further increase in the limiting resource results in a decreasing species richness (Huston 1979, 1994). When nutrients or other resources that increase the productivity added to a system, diversity often decreased. The addition of fertilizer to herbaceous plant communities often results in a sharp decrease in species diversity. High concentrations of resource have been suggested to favour species that tend to outcompete others and dominate the ecosystem (Grime 1973).

Species richness is related to community productivity for a broad range of organisms found in different types of ecosystems. Often, biological species richness increases with increasing productivity until it reaches a plateau, beyond which further increases in productivity are associated with a decline in species richness, that is a hump-shaped pattern (Proulx & Mazumder 1998). Diversity tends to decline in high productivity environments is considered due to competitive exclusion by favoured species that became dominant under the condition (Grime 1973; Huston 1994; Osem *et al.* 2002).

Species richness is often observed as a hump-shaped or unimodal response along natural gradients of increasing biomass or resources, but linear (positive or negative) and non-significant responses are also commonly observed (Brown 1973; Goldberg & Miller 1990; Gough *et al.* 1994; Waide 1999). The responses of species richness are scale-dependent (Gross et al. 2000). Chase and Leibold (2002) found that at the local scale, species richness has a hump-shaped relationship with productivity, while at the regional scale it is a positive linear relationship. Species richness is also related to environmental heterogeneity (Huston 1994; Bell et al. 2000; Lundholm & Larson 2003), with a negative or positive relationship being found when samples are within a homogenous habitat and a hump-shaped relationship being found when samples encompasses different habitat types (Guo & Berry 1998). Oksanen (1996) claims that the hump-shaped relationship between diversity and biomass is just an artefact due to the fixed small plot size; as plants at high biomass are bigger, and therefore there are less species in a plot.

Many studies have found that the highest levels of species diversity were maintained at some “intermediate’’ frequency or intensity of disturbance. At high rates of mortality (disturbance or predation), diversity was reduced because some species were unable to recover from mortality. At low rates of mortality, diversity was reduced by competitive exclusion as dominant species eliminated poorer competitors. This phenomenon is described as intermediate disturbance hypothesis (Conell 1978), which is a non-equilibrium hypothesis to explain regional variation of species diversity and claims that if disturbance is frequent, only rapidly colonizing species have time to establish a community before disturbance comes again. As the time between disturbance events increase, or disturbance become less severe, more species can migrate into a community, so the species diversity increases (Conell 1978; Huston 1979; Vujnovic *et al*. 2002).

Herbivores are generally thought to enhance plant diversity by their direct consumption of competitively dominant plant species and indirect effects on plant competition (McNaughton 1985; Belsky 1992). However, other studies suggest that herbivores sometimes have a weak, or even negative, effect on plant diversity. Plant species richness decreases with high grazing in nutrient-poor ecosystems, while it increases with high grazing in nutrient-rich ecosystem (Proulx & Mazumder 1998). Domesticated large grazers managed at low stocking rates on productive grassland increase plant diversity, but high stocking rates can decrease plant diversity. Insect herbivores often have weak or negative effects plant diversity (Hodgson & Illius 1996).

Herbivores can influence species richness at both local and regional scale. Local disturbance and selective grazing can enhance diversity at local scales, but strong selection for grazing-tolerant plant species within the species pool might reduce diversity at larger scale. The effect of herbivores on plant diversity also differs with the environment. Grazing mammals in more productive grassland increase plant diversity. Grazers in arid or very saline environments often do not change or can even decrease diversity (Oiff & Ritchie 1998).

2.3.4. Species distribution and abundance

There has been an increasing interest in studies on species distribution and abundance in the last decade due to the concern over climatic change and the impact of human activities on species distribution and diversity (Prentice 1992; Tappeiner *et al.* 1998; Vujnovic *et al.* 2002). Recent research has focused on the theories related to species niche, competitive exclusion, species response curves along environmental gradients, and vegetation succession to investigate species distribution and abundance in the face of environmental change and anthropogenic influence (Austin *et al*. 1990; Austin 2002; Guisan *et al*. 2002).

Species distribution and abundance along environmental gradients in an ecosystem are determined by species' intrinsic physiological characteristics, the interaction among species themselves, and the abiotic environments through providing resources and imposing constraints on them (Guisan & Zimmermann 2000). Species response shape to resource gradients is described as species' niche, including fundamental niche and realized niche (Franklin 1995; Austin 2002). Hutchinson (1957) defined the species' fundamental niche as the hypervolume defined by environmental dimensions within which that species can survive and reproduce. A species may be excluded from parts of its fundamental niche because of competition and other biotic interactions. The reduced hypervolume is then termed the realized niche (Austin *et al.* 1990; Austin *et al.* 1994). The fundamental niche is primarily a function of physiological performance and ecosystem constraints. The realized niche additionally includes biotic interactions and competitive exclusion (Guisan & Zimmermann 2000).

Ecological (or environmental) gradients can be classified into three categories, namely resource, direct, and indirect gradients. Resource gradients address matter and energy consumed by plants or animals (nutrients, water, light for plants, food, water for animals). Direct gradients are environmental parameters that have physiological importance, but are not consumed (i.e. temperature, pH). Indirect gradients are variables that have no direct physiological relevance for a species' performance (slope, aspect, elevation, topographic position, habitat type, geology). Indirect variables usually replace a combination of different resources and direct gradients in a simple way (Franklin 1995; Guisan & Zimmermann 2000; Austin 2002). Environmental gradients may be either proximal or distal. Proximal and distal refer to the position of the predictor in the chain of processes that link the predictor to its impact on plant. The most proximal gradient will be the causal variable determining plant response. For example, available soluble soil phosphate concentration at the root hair would be a more proximal resource gradient than total soil phosphorus. Indirect gradients are clearly distal variables (Austin 2002).

Species response curve along the ecological gradient is generally assumed to be a Gaussian form, e.g. a unimodal, symmetric bell-shaped curve (Gauch & Whittaker 1972; Whittaker *et al.* 1973; Giller 1984; Krebs 1994). However, these assumptions are argued for lack of evidence. Studies have revealed that only some species may have a unimodal response curve, many species' responses to environmental variables is asymmetric, that is skewed (Minchin 1989; Austin 1999). The expected shape of a species response will vary with the nature of the gradient. Response to an indirect gradient could take any shape (Austin 2002).

Spatial patterns of species abundance are produced by the relative difference of species in their ability to compete or survive over a range of environmental conditions. Competition displaces species toward environmental conditions that they are able to tolerate, but which the species that outcompete them under optimal conditions cannot tolerate. The ecological optimum of a species along a resource gradient is generally constrained on the high resource end primarily by competition, and on the low resource end primarily by physiological limitation. Thus for many species, the ecological optimum is closer to their physiological limit than to their physiological optimum (Huston 1994).

In a hierarchical scheme of environmental controls on the distribution of plant species, physiologically based climatic variables that are related to direct gradients control plant distributions at the largest spatial and temporal scale (Franklin 1995). Whereas topographical and edaphic factors determine the local and regional scale for the distribution of plant species (Tappeiner *et al.* 1998; Gotteried *et al.* 1999).

Species coexist because of interspecific trade-offs (1) between their competitive abilities and their dispersal abilities; (2) between their competitive abilities and their susceptibility to disease, herbivory or predation; (3) between their abilities to live off average conditions and their abilities to exploit resource pulses; or (4) between their abilities to compete for alternative resources in a heterogeneous landscape (Hastings 1980; Tilman 1982). Coexistence was found to rely on the fluctuation in population densities, while community level densities (the summation of the competing population density) varied very little (Tilman 1982). If a habitat is spatially heterogeneous, that is, if different individual plants living in different portions of the habitat experience different resource supply ratios, then this heterogeneity could allow a large number of species to coexist (Tilman 1994).

Competitive exclusion principle states that if two species are competing for exactly the same limiting resources, they can not coexist, that is, one of them will be completely excluded by the other and will become locally extinct (Hardin 1960; Huston 1994).

Effort to explain the biodiversity in the context of competitive exclusion falls into two general classes. One is based on the idea that competitive exclusion does not happen instantaneously, and many factors can slow the approach to equilibrium or alter the relative strength of competitive interaction. The other general explanation for higher diversity was the acknowledgement that the competitive exclusion did occur, but there was sufficient patchiness in the environment that a species that became extinct on one patch would survive on another patch, and thus allow the species diversity to be maintained at a large scale (Huston 1994).

2.4. GIS and GIS-based ecological and environmental modelling

2.4.1. GIS in ecological and environmental study

The utility of GIS in ecological or environmental studies is commonly in storing, managing, and integrating spatially referenced data relating to points (e.g., individual trees), lines (e.g., rivers, roads), and polygons (e.g., forest boundaries, habitat types, territorial ranges), conducting spatial queries (e.g., searching for areas in which a particular species or feature occurs), engaging in geographic analysis (e.g., statistical analysis of relationships between habitat and reproductive success), and displaying data in the form of high-quality maps (Scholten & de Lepper 1991; Dominy & Duncan 2001).

Topography is an important driving variable in many ecological processes because of its influence on insolation, water flow, and organism movement. GIS provides a number of methods for analyzing topography. Digital elevation model (DEM), a raster data format, and triangulated irregular network (TIN), a vector data format, are widely used to represent altitude data in GIS. Topographic analyses in GIS are very powerful functions in environmental and ecological studies. GIS can be used to directly derive map layers from DEM or TIN such as slope (the rate of change in elevation), aspect (the direction which a slope surface faces), inflection (the curvature of a surface in the direction of slope), surface water flow and catchments. Also, by analyzing topographic functions in modifying other environmental features, GIS can also indirectly generate map layers of air temperature, soil moisture, solar radiation, and so on (Iverson *et al.* 1997; Johnston 1998; Stocks & Wise 2000). The ability to analyze digital topographic data has significantly advanced ecological and hydrological modelling (Guisan & Zimmermann 2000).

One of the powerful operations that can be performed by GIS is the vertical intersection, or map overlay, of spatially distributed data. Overlay operations can be performed for the purposes of merging separate spatial databases, (e.g., hydrology layer with a soil layer), for analyzing spatial intersections between data layers, or for analyzing temporal changes. GIS can operate graphical overlay, which directly overlays two data layers; logical overlay, which employs Boolean operators to analyze the spatial coincidence of input data layers, and arithmetic overlay, which performs mathematical operations across multiple data layers (Iverson *et al.* 1997; Johnston 1998; Stocks & Wise 2000). For example, Iverson (1997) predicted forest composition and productivity of Ohio forest in U.S.A. by using an integrated moisture index which was produced by overlaying map layers of a slope-aspect shading index, cumulative flow of water downslope, curvature of the landscape, and water-holding capacity of the soil.

2.4.2. GIS in ecological and environmental modelling

Models provide ecologists with tools for extrapolating field measurements and integrating complex ecological information over space and time. GIS have become increasingly valuable tools in the computer-based modelling of environmental processes. The current generation of environmental models requires large amounts of spatial data as input and produce predictions that can be displayed as a map. GIS are able to produce data required as input to models and excel at displaying spatial predictions. The integration of GIS with environmental models is emerging as a significant new area of GIS development and has been widely used in modelling species distribution and abundance, community productivities, impacts of climatic changes on vegetations, watershed hydrology, and solar radiation (Franklin 1995; Iverson *et al.* 1997; Johnston 1998; Guisan & Zimmermann 2000; Stocks & Wise 2000).

GIS has played various roles in environmental modelling. The commonly accepted roles for GIS in environmental modelling are as a pre- and post processor, preparing data files for input to the model, and displaying the model result as map form for prediction and queries. This procedure is usually carried out by coupling, in which separate environmental models are linked with GIS packages (Stocks & Wise 2000).

Statistical models are based on empirical observations and contain one or more random variables. When the empirical relationships needed for model development are not known, a GIS can be used to assemble spatial data on landscape properties, to derive new data that are syntheses of the originals, and statistically analyze those new data to determine the strength of interaction among them. The resulting empirical relationships can be used to predict gradients of habitats, net primary production, nutrients circles, and other ecosystem properties across the landscape. Most GIS programs have the function of statistical analysis, but for more advanced statistical modelling, GIS generated or derived input data are usually exported to professional statistical software, and trained models are imported to GIS for prediction or further query and analysis. Multivariate statistical techniques coupled with GIS analyses provide the means for quickly compiling data, synthesizing these data, and developing predictive models to relate ecological functions to quantifiable landscape characteristic (Johnston 1998).

Simulation models of population, ecosystem, and landscape have been successfully coupled with GIS, making model prediction more powerful in simulating ecosystem processes. This coupling is most successful with models that predict outcomes of processes such as succession, net primary production, and nutrient cycling, from parameters derived from maps or digital satellite data. Simulation models of population, ecosystem and landscape can be incorporated with spatial elements to duplicate ecological functions via coupled differential equations that describe key ecosystem and landscape processes. In these simulation models, GIS is used to derive input variables, spatially extrapolate results, and display results (Franklin 1995; Johnston 1998).

Expert systems are computer systems that advise on or help solve real-world problems that would normally require a human expert's interpretation. One of the very important and also very influential uses of GIS and GIS-based environmental modelling is to link them with an expert system. This linkage has found wide applications in species conservation management, environmental management, and forest and pasture management (Johnston 1998).

2.4.3. GIS-based predictive modelling of species distribution and abundance

GIS-based predictive modelling of species distribution, a statistical modelling approach carried out in a GIS environment, focuses on the impact of environmental change and intensive land use on species diversity, distribution and abundance, community structure and production, and some other vegetation features in studies related to biogeography, conservation biology, climate change research, species management, ecosystem management and restoration, and agricultural and forestry management (Franklin 1995; Zimmermann & Kienast 1999; Guisan & Zimmermann 2000; Austin 2002).

GIS-based predictive modelling of species distribution starts with the development of some type of model followed by the application the model to a GIS system to produce the predictive map, a realization of the model. Computerized predictive vegetation modelling is made possible by the availability of digital maps of topography and other environmental variables such as soils, geology and climate variables, and geographic information system software for manipulating these data. Especially important to predictive modelling of species distribution and abundance are interpolated climatic variables related to physiological tolerances, and topographic variables, derived from digital elevation grids, related to site energy and moisture balance (Franklin 1995).

Commonly there are three steps in GIS-based predictive modelling of species distribution: 1) database development, 2) model calibration, and 3) model prediction or visualization in GIS. Database development and model visualization usually involve GIS implementation, while model calibration is carried out in a statistic software package (Johnston 1998; Guisan & Zimmermann 2000; Hunsaker 2001; Austin 2002).

There are two types of data generated in database development: those of vegetation data, which may include species diversity, distribution, abundance, and biomass, and those of environmental data, which may include climate, topology, geology, and disturbance. Both of them can be from field sampling or derived from GIS database (such as topology data from DEM), remote sensed data, and other sources. The vegetation data containing the modelling target is the dependent variable, all other factors are the independent variable (indicator variable) in calibrating the statistical modelling.

The variety of statistical techniques used in predictive modelling is growing. Ordinary multiple regression and its generalized form (GLM) are very popular and are often used for modelling species distributions. Other methods include multiple regression, neural networks, ordination, decision tree, Bayesian models, locally weighted approaches (e.g. GAM), environmental envelopes or even combinations of these models (Guisan & Zimmermann 2000). Choosing a proper statistical technique depends on the data characteristics such as whether they are quantitative or qualitative, or ordinal or continuous, and the possible species response curve along the environmental gradients (Austin 2002).

Although GIS are widely used tools in all types of spatially explicit studies, most GIS software still lack important statistical functions for predictive purposes. This is a serious flaw because not all statistically derived models are easy to implement in a GIS environment. As a result, to import the model from a statistic package to a GIS is usually a tedious and time-consuming procedure for some of the above mentioned statistic models, and this is also true for the export of GIS derived data to the statistical software.

2.5. Data mining and decision tree

2.5.1. Data mining

Data mining is the process of posing various queries and extracting useful information, patterns, and trends often previously unknown from large quantities of data already present in databases. The goals of data mining include detecting abnormal patterns and predicting the future based on past experiences and current trends (Thuraisingham 1999). The development of data mining is a result of the natural evolution of information technology. A great effort in the information industry in recent years has been to deal with huge amounts of data and the imminent need for turning such data into useful information and knowledge (Han & Kamber 2001).

There are a series of steps in data mining: data collection and database creation, data management (including data storage and retrieval, database transaction processing) and data analyses and understanding (Han & Kamber 2001).

2.5.2. Technologies used in data mining:

Data mining integrates many technologies in data analysis. Some of them are newly developed approaches with the development of information technology, and some of them have existed for many decades (Thuraisingham 1999; Dunham 2002). The most commonly used technologies are as follows (Han & Kamber 2001):

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).
- Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concept of evolution.
- Nearest neighbour method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset. Sometimes called the k-nearest neighbour technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.

2.5.3. Decision tree

A 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 applications in environmental modelling with considerable accuracy and ease of interpretation (Iverson & Prasad 1998; Vayssieres *et al*. 2000; Scheffer 2002; Yang *et al*. 2003).

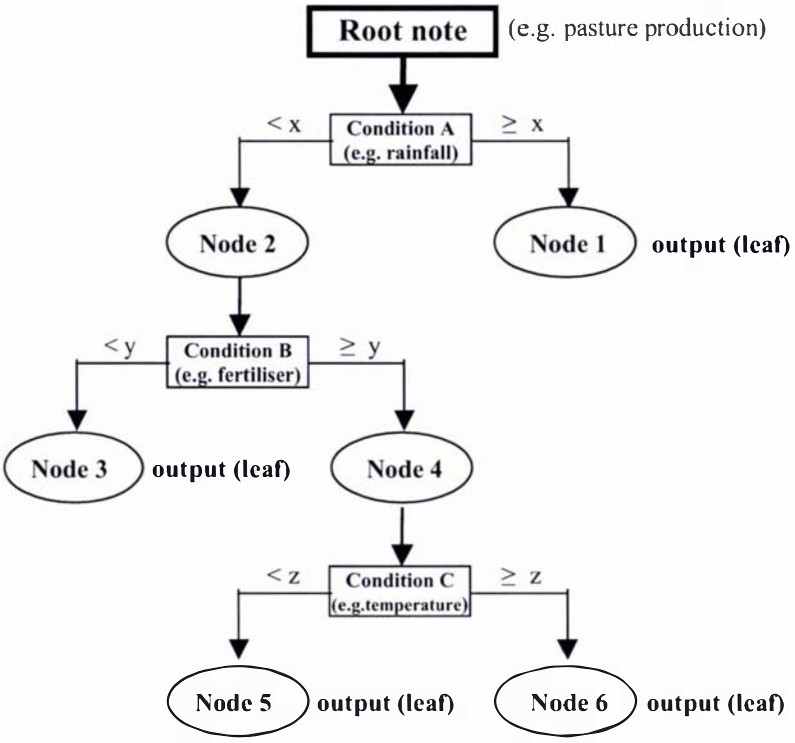


Fig 1. General structure of a decision tree.

The decision tree approach was originally developed by Breiman *et al.* (1984) and was named classification and regression tree (CART). It is a non-parametric modelling approach, which recursively splits the multidimensional space defined by the independent variables into zones that are as homogenous as possible in term of the response of the dependant variable (Vayssieres *et al.* 2000). The result of the analysis is a binary hierarchy structure called a decision tree that contains the rules to predict the new cases (Breiman *et al.* 1984; Dunham 2002).

The main procedure for developing a decision tree model is to split the data for target (dependant) variable based on it response to input (independent) variables. A decision tree model starts from the root node which is the dataset containing all observations of target variables (for example, pasture production) and applies a condition to split the data (Figure 1). This condition is based on one of the input variables (for example, rainfall) and a split-point (say 800 mm). If the input variable in the condition is larger than or equal to the threshold (≥ 800 mm), those observations in the target data corresponding to this are classified to node 1; otherwise, the observations in the target data are classified to node 2 (<800 mm). The same procedure recursively applies to each of node 1 and 2 and the following nodes until the further splitting is stopped by stopping rules. The final nodes (outputs) are called leaves and are the final predictions, which are made by series combinations of input variables and the split-points. For example, suppose the model target is pasture production, conditions A, B and C are rainfall, fertiliser and temperature, respectively and the split-points for these three conditions are 800 mm, 200 kg/ha and 8°C, respectively, when pasture has a rainfall less than 800 mm, a fertiliser input equal to or more than 200 kg/ha and a temperature equal to or more than 8°C, the predicted pasture production is the amount in node 6.

There are different criteria used to split the data, depending on the nature of the target variable. For interval variables like pasture productivity, the split criterion is variance reduction, or F test (Breiman *et al.* 1984; Fernandez 2003). In the case of variance reduction, a complete search is applied to all the input variables and possible split-points to select one variable that ultimately explains the variance of the target variable by splitting the dataset of the target variable into two sub-datasets. Then a same procedure is recursively applied to each of the sub-dataset until the dataset cannot be further split based on defined rules. Suppose D is the dataset containing the target variable and is going to be split into two sub-datasets D_L and D_R. DEV_D is the deviance of dataset D (squared error from the mean):

$$DEV_D = \sum_{obsi} (Yi - Y_{bar}) \qquad (1)$$

Where Yi is a observation and Y_{bar} is the mean of the target variable D. The variance reduction by splitting dataset D into D_L and D_R is expressed as follows:

$$\Delta = DEV_D - (DEV_L + DEV_R) \qquad (2)$$

Where DEV_L and DEV_R are the deviance of D_L and D_R, respectively. An input variable and a possible split-point that maximise Δ, or minimise the sum of ($DEV_L + DEV_R$) is first selected to split the data D, and then a search is implemented for each sub-dataset D_L and D_R and the following sub-datasets recursively (Breiman *et al.* 1984).

In the case of an F test, an input variable and a possible split-point that gives the most significant p-value in the F test associated with variance reduction is selected.

The stopping rules are carried out by setting the maximum tree levels and the minimum observations required for a split search. The maximum tree level defines the size of the tree and prevents the tree becoming too complex, while the minimum observations required for a split search, prevents a few special cases influencing the output of the tree. The significance level of F test assigned for variance reduction is another stopping rule in decision tree model development.

Model assessment is an important part in decision tree development which enables a good model output and prevents overtraining, otherwise the model would only fit the data from which it was trained (Witten & Erank 2000). For models having an interval target (i.e. continuous variable), the measure of average squared error (ASE) is commonly used (Fernandez 2003). ASE is similar to mean squared error (MSE) in a general linear model and is an indicator of model goodness-of-fit; the smaller the ASE, the better the model fits. For the assessment of the decision tree, if there are validation data available, it is assessed by the ASE of the validation data; otherwise, it is assessed by the ASE of the training data. The first option is better. When validation data are available, a “best” tree that has the smallest ASE from validation data is selected (Fernandez 2003).

The decision tree has many advantages over other model approaches such as regression (Thuraisingham 1999; Vayssieres *et al.* 2000). Namely: (1) It has no strict assumption for the distribution of the target variable (dependent variable) about which regression assumes normal distribution. Also, there is no multicollinearity problem when input variables (independent variables) are highly correlated, which is a limitation of multiple regression. (2) Decision tree deals with non-linear models easily without any variable transformation. (3) Decision tree can indicate the relative importance of input variables with respect to their influences on the model target, and can indicate the interactions among input variables. (4) It can easily incorporate ordinal (such as those measured as low, medium and high), nominal (such as those for soil types) and interval (such as those for biomass) variables in the same model.

Naturally, decision tree also has its limitations: it requires a relatively large amount of training data; it cannot express linear relationships in a simple and concise way like regression does; it cannot produce a continuous output due to its binary nature; and it has no unique solution, that is, there is no best solution (Iverson & Prasad 1998; Scheffer 2002).

2.6. Conclusion

Hill-pastures in New Zealand are naturalized vegetations originated from clearance of native podocarp rainforest, fern or scrub and oversowing with introduced pasture species in hill-country by early European settlers. In hill-pasture, species composition and abundance and pasture productivity are strongly influenced by climatic, topographical and soil factors, and the pasture management, such as grazing regimes and fertilizer application.

In an ecosystem, species diversity, distribution and abundance are closely related to the abiotic environmental conditions (e.g. temperature, topography and available resources), the frequency and intensity of disturbances (e.g. fire, grazing), and the biotic factors (e.g. species' physiological traits and the interaction among them). Species diversity plays a very important role in ecosystem functioning, especially in maintaining community stability.

GIS, as a powerful tool in dealing with spatial factors, has been introduced in ecological and environmental study and has shown a very promising application in modelling species distribution and abundance. The application of GIS in ecological and environmental modelling greatly enhanced the capability of obtaining input data for model analysis and presenting model outputs over space.

Decision tree, as one of the data mining modelling approaches, has been widely used in the social and medical sciences. It has also had increasing applications in environmental modelling with considerable accuracy and ease of interpretation. An integration of decision tree with GIS seems to be a feasible combination in modelling and investigating ecosystem processes in pasture ecosystems.

2.7. References

Austin M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. *Oikos*, 86, 170–178

Austin M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101-118

Austin M.P., Nicholls A.O., Doherty M.D. & Meyer J.A. (1994) Determining species response functions to an environmental gradient by means of a B-function. *Journal of Vegetation Science*, 5, 215-228

Austin M.P., Nicholis A.O. & Margules C.R. (1990) Measurement of the realized qualitative niche: environmental niche of five eucalyptus species. *Ecological Monograph*, 60, 161-177

Baars J.A., Radcliffe J.E. & Rollo M.D. (1990) Climatic change effects on seasonal patterns of pasture production in New Zealand. *Proceedings of the New Zealand Grassland Association*, 51, 43-46

Barker D.J. & Dymock N. (1993) Introduction, production, and persistence of five grass species in dry hill country 6. Ballantrae, North Island, New Zealand. *New Zealand Journal of Agricultural Research*, 36, 41-47

Belsky A.J. (1992) Effects of grazing, competition, disturbance and fire on species composition and diversity in grassland. *Journal of Vegetation Science*, 3, 187-200

Blennerhassett J.D. (2002) *Pasture growth constraints on dry steep east coast hill country.* Ph.D. thesis, Massey University, New Zealand

Bond W.J. (1994) Keystone species. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and Ecosystem Function,* Springer-Verlag, 237-250

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) *Classification and regression trees.* The Wadsworth statistics/probability series. Chapman & Hall. Inc. New York, NY.

Brown J.H. (1973) Species diversity of seed-eating desert rodents in sand dune habitats. *Ecology*, 54, 775-787

Caldeira M.C., Ryel R.J., Lawton J.H. & Pereira J.S. (2001) Mechanisms of positive biodiversity-production relationships: insights provided by ^{13}C analysis in experimental Mediterranean grassland plots. *Ecology Letters*, 4, 439-443

Campbell B.D. (1996) Climate change and pastures. *Proceedings 48th Ruakura Farmers' Conference held at Ruakura, New Zealand,* 11 June 1996, 67-74

Chapin III F., Zavaleta E.S., Eviner V., Naylor, R., Vitousek P.M., Reynolds H.L., Hooper D., Lavorel S., Sala, O.E., Hobbie S.E., Mack M.C. & Diaz S. (2000) Consequences of changing biodiversity. *Nature*, 405, 234-242

Clark D.A., Rolston M.P., Lambert M.G. & Budding P.J. (1984) Pasture composition under mixed sheep and goats grazing on hill country. *Proceedings of the New Zealand Grassland Association*, 45, 160-167

Conell J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science*, 199, 1302-1310

Cornell H.V. (1985) Local and regional richness of cynipine gall wasps on California oaks. *Ecology*, 66, 1247-1260

de Ruiter P.C., Neutel A. & Moore J.C. (1995) Energetics, patterns of interaction strengths, and stability in real ecosystem. *Science*, 269, 1257-1260

Diaz S., Cabido M., & Casanoves F. (1998) Plant functional traits and environmental filters at a regional scale. *Journal of Vegetation Science*, 9, 113-122

Dominy N.J. & Duncan B. (2001) GPS and GIS methods in an African rain forest: applications to tropical ecology and conservation. *Conservation Ecology [electronic resource]* http://www.consecol.org/vol5/iss2/art6, 5

Dunham M.H. (2002) *Data Mining: Introduction and Advanced Topics.* Pearson Education INC. Upper Saddle River, New Jersey 07458

Ehrlich P.R. & Wilson E.O. (1991) Biodiversity studies: science and policy. *Science*, 253, 758-762

Fernandez G. (2003) *Data mining using SAS applications.* Chapman & Hall/CRC, Boca Raton, Florida

Franklin J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499

Gauch H.G. & Whittaker R.H. (1972) Coenoline simulation. *Ecology*, 53, 446-451

Giller P.S. (1984) *Community Structure and the Niche.* Chapman and Hall, London, England

Gillingham A.G. (1982) Topographic and management effects on dung distribution by grazing sheep. *Proceedings of the New Zealand Grassland Association*, 43, 161-170

Gillingham A.G. & During C. (1973) Pasture production and transfer of fertility within a long-established hill pasture. *New Zealand Journal of Experimental Agriculture*, 1, 227-232

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

Goldberg D.E. & Miller T.E. (1990) Effects of different resource additions on species diversity in an annual plant community. *Ecology*, 71, 213-225

Gotteried M., Pauli H., Reiter K. & Grabherr G. (1999) A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions*, 5, 241-251

Gough L., Grace J.B. & Taylor K.L. (1994) The relationship between species richness and community biomass: the importance of environmental variables. *Oikos*, 70, 271-279

Grant D.A. & Brock J.L. (1974) A survey of pasture composition in relation to soils and topography on a hill country farm in the southern Ruahine Range, New Zealand. *New Zealand Journal of Experimental Agriculture*, 2

Grime J.P. (1973) Control of species density in herbaceous vegetation. *Journal of Environmental Management*, 1, 151-167

Grime J.P. (1997) Biodiversity and ecosystem function: the debate deepens. *Science*, 277, 1260-1261

Grime J.P. (2002) Declining plant diversity: empty niches or functional shifts? *Journal of Vegetation Science*, 13, 457-460

Guisan A., Edwards Jr T.C. & Hastie T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157, 89-100

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

Han J.W. & Kamber M. (2001) *Data mining: concepts and techniques.* International Thomson, London.

Hardin G. (1960) The competitive exclusion principle. *Science*, 131, 1292-1297

Hastings A. (1980) Disturbance, coexistence, history, and competition for space. *Theoretical Population Biology*, 18, 363-373

Heywood V.H. (1995) *Global biodiversity assessment*, New York: Cambridge University Press, pp89

Hilgendorf F.W. (1936) *Pasture plants and pastures of New Zealand.* Christchurch, New Zealand, Whitcombe and Tombs Limited.

Hobbie S.E., Jensen D.B. & Chapin III F.S. (1994) Resource supply and disturbance as controls over present and future plant diversity. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and Ecosystem Function*, *Springer-Verlag, 385-405*

Hodgson J.G. & Illius A.W. (1996) *The ecology and management of grazing system.* Wallingford : CAB International

Hooper D.U. & Vitousek P.M. (1997) The effects of plant composition and diversity on ecosystem processes. *Science*, 277, 1302-1305

Hunsaker C. (2001) *Spatial uncertainty in ecology: implications for remote sensing and GIS applications.* New York: Springer

Huston M.A. (1979) A general hypothesis of species diversity. *American Naturalist*, 113, 81-101

Huston M.A. (1994) Biological diversity: *The coexistence of species on changing landscapes.* Cambridge University Press. Cambridge, UK

Hutchinson G.E. (1957) Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology*, 22, 415-427

Iverson L.R., Martin E.D., Scott C.T. & Prasad A. (1997) A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio (U.S.A.). *Landscape Ecology*, 12, 331-348

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

Johnston C. (1998) *Geographic information systems in ecology.* Oxford, Malden, MA: Blackwell Science

Judas M. (1988) The species-area relationship of European Lumbricadae (Annelida, Oligochaeta). *Oecologia*, 76, 579-587

Kerr J.P., de Ruiter J.M. & Hall A.Z. (1986) The magnitude and variability of seasonal water deficits for pasture and crop growth. *New Zealand Agriculture Science*, 13, 13-18

Krebs C.J. (1994) *Ecology: The Experimental Analysis of the Distribution and Abundance* (fourth ed.). Harper & Row, New York

Lambert M.G., Clark D.A., Grant D.A. & Costall D.A. (1986) Influence of fertilizer and grazing management on North Island moist hill country 2. Pasture botanical composition. *New Zealand Journal of Agricultural Research*, 29, 1-10

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

Lambert M.G. & Roberts E. (1978) Aspect differences in an unimproved hill country pasture 2. Edaphic and biotic difference. *New Zealand Journal of Agricultural Research*, 21, 255-260

Langer R.H.M.e. (1990) *Pastures : their ecology and management.* Auckland, N.Z. : Oxford University Press

Ledgard S.F., Sheath G.W. & Gillingham A.G. (1982) Influence of some soil and pasture components on the growth of hill pastures 1. Winter and spring production. *New Zealand Journal of Agricultural Research*, 10, 239-244

Levy E.B. (1970) *Grassland of New Zealand.* Wellington, New Zealand. Government Printer.

Liu C. (1996) *Effect of slope class on defoliation frequency and severity of Trifolium repens by sheep in hill pasture.* Master thesis, Massey University, Palmerston North, New Zealand

Lonnie A.W. (1997) High productivity in grassland ecosystem: effected by species diversity or productive species? *Oikos*, 80, 183-184

Lopez I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Lopez I.F., Lambert M.G., Mackay A.D. & Valentine I. (2003) The influence of topography and pasture management on soil characteristics and herbage accumulation in hill pasture in the North Island of New Zealand. *Plant and Soil*, 255, 421-434

Loreau M., Naeem S., Bengtsson J., Grime J.P., Hector A., Hooper D.U., Huston M.A., Raffaelli D., Schmid B., Tilman D. & Wardle D.A. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294, 804-808

Luscombe P.C. & Grant D.A. (1981) Responses of ryegrass, browntop, and an unimproved resident pasture in hill country, to nitrogen, phosphorus, and potassium fertilisers. II. Species composition of resident pastures. *New Zealand Journal of Agricultural Research*, 9, 237-241

Mabry C., Ackerly D. & Gerhardt F. (2000) Landscape and species-level distribution of morphological and life history traits in a temperate woodland flora. *Journal of Vegetation Science*, 11, 213-224

May R.M. (1973) Stability and complexity in model ecosystem. *Princeton University Press*

McCann K.S. (2000) The diversity-stability debate. *Nature*, 405, 228-233

McNaughton S.J. (1985) Ecology of grazing ecosystem: the Serengeti. *Ecological Monograph*, 53, 291-320

Minchin P.R. (1989) Montane vegetation of Mt. Field massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio*, 83, 97-110

Moir J.L. (2000) *Investigations into the influence of fertilizer history and climate regime on the soil fertility, soil quality and pasture production of Wairarapa hill country soil.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

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

Mouquet N., Moore J.L. & Loreau M. (2002) Plant species richness and community productivity: why the mechanism that promotes coexistence matters. *Ecology Letters*, 5, 56-64

Naeem S. (1998) Species redundancy and ecosystem reliability. *Conservation Biology*, 12, 39-45

Naeem S. (2002a) Biodiversity equals instability? *Nature*, 416, 23-24

Naeem S. (2002b) Ecosystem consequences of biodiversity loss: the evolution of a paradigm. *Ecology*, 83, 1537-1552

Nacem S. & Li S. (1997) Biodiversity enhances ecosystem reliability. *Nature*, 390, 507-509

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Nicholas P.K., Kemp P.D. & Barker D.J. (1998) Biodiversity, stability and pasture management - the role of functional groups. *Proceedings of the 9th Australian Agronomy Conference, Wagga wagga, 1998*

Oiff H. & Ritchie M.E. (1998) Effects of herbivores on grassland plant diversity. *Trends in Ecology and Evolution*, 13, 261-265

Osem Y., Perevolotsky A. & Kigel J. (2002) Grazing effect on diversity of annual plant communities in a semi-arid rangeland: interactions with small-scale spatial and temporal variation in primary productivity. *Journal of Ecology*, 90, 936-946

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

Poff N.L. (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*, 16, 391-409

Prentice I.C. (1992) Climatic change and long-term vegetation dynamics. In: Glenn-Lewin, D.C., Peet, R.K., and Veblen, T.T. (Eds.), *Plant Succession-Theory and Prediction*, Chapman & Hall, London, 293-339

Proulx M. & Mazumder A. (1998) Reversal of grazing impact on plant species richness in nutrient-poor vs. nutrient-rich ecosystem. *Ecology*, 79, 2581-2592

Purvis A. & Hector A. (2000) Getting the measure of biodiversity. *Nature*, 405, 212-218

Radcliffe J.E. & Lefever K.R. (1981) Aspect influences on pasture microclimate at Coopers Creek, North Canterbury. *New Zealand Journal of Agricultural Research*, 24, 55-56

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

Roach C.G., Nemaia E.K.K., Ledgard S.F., Brier G.J. & McLay C.D.A. (1996) Effects of long-term difference in fertiliser history on hill country: seasonal pasture production, legume growth and soil phosphorus status. *Proceedings of the New Zealand Grassland Association*, 57, 105-109

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

Scholten H.J. & de Lepper M.J.C. (1991) The benefits of the application of geographical information systems in public and environmental health. *World Health Statistics Quarterly*, 44, 160-170

Schulze E.D. & Mooney H.A. (1994) Ecosystem function of biodiversity: a summary. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and ecosystem function*, Springer-Verlag, 497-509

Sheath G.W. & Boom R.C. (1985a) Effect of November-April grazing pressure on hill country pastures 2. Pasture species composition. *New Zealand Journal of Agricultural Research*, 13, 329-340

Sheath G.W. & Boom R.C. (1985b) Effect of November-April grazing pressure on hill country pastures 3. Interrelationship with soil and pasture variation. *New Zealand Journal of Agricultural Research*, 13, 341-349

Stocks C.E. & Wise S. (2000) The role of GIS in environmental modelling. *Geographical and Environmental Modelling*, 4, 219-235

Suckling F.E.T. (1975) Pasture management trials on unploughable hill country at Te Awa. 3. Result for 1959-69. *New Zealand Journal of Agricultural Research*, 3, 351-436

Tan G. & Shibasaki R. (2003) Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. *Ecological Modelling*, 168, 357-370

Tappeiner U., Tasser E. & Tappeiner G. (1998) Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an Alpine area. *Ecological Modelling*, 113, 225-237

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

Tilman D. (1982) *Resource competition and community structure (Monographs in population biology).* Princeton University Press

Tilman D. (1994) Community diversity and succession: the role of competition. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and Ecosystem Function*, Springer-Verlag, 327-344

Tilman D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, 77, 350-363

Tilman D. (2000) Cause, consequences and ethics of biodiversity. *Nature*, 405, 208-211

Tilman D., Knops J., Wedin D., Reich P., Ritchie M. & Siemann E. (1997) The influence of functional diversity and composition on ecosystem processes. *Science*, 277, 1300-1302

Tilman D., Reich P.B., Knops J., Wedin D., Mielke T. & lehman C. (2001) Diversity and productivity in a long-term grassland experiment. *Science*, 294, 843-845

Trustrum N.A., Lambert M.G. & Thomas V.J. (1983) Erosion, a drop away in production. *Soil and Water*, 19, 11-19

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

Vitousek P.M. & Hooper D.U. (1994) Biological diversity and terrestrial ecosystem biogeochemistry. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and ecosystem function*, Springer-Verlag, 3-14

Vujnovic K., Wein R.W. & Dale M.R.T. (2002) Predicting plant species diversity in response to disturbance magnitude in grassland remnants of central Alberta. *Canadian Journal of Botany*, 80, 504-511

Waide R.B. (1999) The relationship between prodcutivity and species richness. *Annual Review of Ecology and Systematics*, 30, 257-300

White J.G.H. (1990) Hill and high country pasture. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

White T.A., Moore K.J. & Barker D.J. (2004) The importance of local processes to landscape patterns of grassland vegetation diversity. *New Zealand Journal of Agricultural Research*, 47, 199-207

Whittaker R.M. (1972) Evolution and measurement of species diversity. *Taxon*, 21, 213-251

Whittaker R.M., Levin S.A. & Root R.B. (1973) Niche, habitat and ecotope. *American Naturalist*, 107, 321-338

Witten I. & Erank E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers San Francisco, California

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

Zhang B.S., Valentine I. & Kemp P.D. (2004) Modelling hill country pasture production: a decision tree approach. *Proceedings of the New Zealand Grassland Association*, 66, 197-201

Zimmermann N.E. & Kienast F. (1999) Predictive mapping of alpine grassland in Switzerland: Species versus community approach. *Journal of Vegetation Science*, 10, 469-482

Chapter 3. Modelling the Productivity of Naturalised Hill-pasture in the North Island, New Zealand: a Decision Tree Approach

As a new modelling methodology, the decision tree has been widely used in the social and medical sciences. It has also had increasing applications in environmental modelling with considerable accuracy and ease of interpretation. However, to my best knowledge there has been no literature showing that it has been used in modelling community productivity of any vegetation types. In this chapter, decision tree models were developed and were compared with regression models to assess their performance in modelling community productivity in the naturalised hill-pasture ecosystem in the North Island, New Zealand.

---

*A paper derived from this chapter with a title "Modelling the productivity of naturalised pasture in the North Island, New Zealand: a decision tree approach" by B. Zhang, I. Valentine & P.D. Kemp has been published in **Ecological Modelling** (in press, available on-line).*

Abstract. Decision tree, one of the data mining methods, has been widely used as a modelling approach and has shown better predictive ability than traditional approaches (e.g. regression). However, very little is known from the literature about how the decision tree performs in predicting pasture productivity. In this study, decision tree models were developed to investigate and predict the annual and seasonal productivity of naturalised hill-pasture in the North Island, New Zealand, and were compared with regression models with respect to model fit, validation and predictive accuracy. The results indicated that the decision tree models for annual and seasonal pasture productivity all had a smaller average squared error (ASE) and a higher percentage of adequately predicted cases than the corresponding regression models. The decision tree model for annual pasture productivity had an ASE which was only half of that of the regression model, and adequately predicted 90.1% of the cases in the model validation which was 10.8 percentage points higher than that of the regression model. Furthermore, the decision tree models for annual and seasonal pasture productivity also clearly revealed the relative importance of environmental and management variables in influencing pasture productivity, and the interaction among these variables. Spring rainfall was the most significant factor influencing annual pasture productivity, while hill slope was the most significant factor influencing spring and winter pasture productivity, and annual P fertiliser input and autumn rainfall were the most significant factors influencing summer and autumn pasture productivity, respectively. One limitation of using the decision tree to predict pasture productivity was that it did not generate a continuous prediction, and thus could not detect the influence of small changes in environmental and management variables on pasture productivity.

3.1. Introduction

Modelling pasture productivity has long been an interest of agronomists and plant ecologists, either for investigating ecosystem processes or predicting pasture yield for practical purposes (e.g. Dyne 1970; Seligman & Baker 1993; Stuth *et al.* 1993; Sorenson 1998; Moir 2000). Traditionally, there are two main approaches to modelling pasture productivity: using mechanistic models to reveal the causal factors determining pasture productivity (Gilmanov *et al.* 1997; Riedo *et al.* 1998; Foy *et al.* 1999; Moir *et al.* 2000), and using empirical models, usually in one of the regression forms, to simulate pasture productivity and investigate the interrelationship between pasture and environmental factors (Lambert *et al.* 1983; Sala *et al.* 1988; Paruelo & Tomasel 1997; Scott 2002). Mechanistic models, because of their strong theoretical bases, tend to be more general and widely applicable 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 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). An empirical model, therefore, is usually a better choice for a predictive purpose, such as assessing climatic impact and fertiliser effect on pasture productivity, due to its reality and accuracy.

With the development of computer technology, a new empirical modelling method, data mining, has become popular due to its strong ability to predict new cases based on previously known information (Witten & Erank 2000; Dunham 2002). Data mining is a process of querying and extracting useful information, patterns, and trends often previously unknown from large quantities of existing data (Thuraisingham 1999). 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 applications in environmental modelling with considerable accuracy and ease of interpretation (Iverson & Prasad 1998; Vayssieres *et al.* 2000; Scheffer 2002; Yang *et al.* 2003).

Decision tree is a non-parametric modelling approach, which recursively splits the multidimensional space defined by the independent variables into zones that are as homogenous as possible in term of the response of the dependant variable (Vayssieres *et al.* 2000). The result of the analysis is a binary hierarchy structure called a decision tree with branches and leaves that contains the rules to predict the new cases (Breiman *et al.* 1984; Dunham 2002). Decision tree has many advantages over other model approaches such as regression (Thuraisingham 1999; Vayssieres *et al.* 2000). Namely: (1) It has no strict assumption for the distribution of the target variable (dependent variable) about which regression assumes normal distribution. Also, there is no multicollinearity problem when input variables (independent variables) are highly correlated, which is a limitation of multiple regression. (2) Decision tree deals with non-linear models easily without any variable transformation. (3) Decision tree can indicate the relative importance of input variables with respect to their influences on the model target, and can indicate the interactions among input variables. (4) It can easily incorporate ordinal (such as those measured as low, medium and high), nominal (such as those for soil types) and interval (such as those for biomass) variables in the same model.

Naturally, decision tree also has its limitations: it requires a relatively large amount of training data; it cannot express linear relationships in a simple and concise way like regression does; it cannot produce a continuous output due to its binary nature; and it has no unique solution, that is, there is no best solution (Iverson & Prasad 1998; Scheffer 2002).

Little information is available from the literature on the performance of decision tree in predicting pasture productivity. In this study, we focused on the productivity (aboveground biomass) of naturalised hill-pasture in New Zealand as the modelling target, and developed and assessed the decision tree models for annual and seasonal pasture productivity. The main aim of this study was to evaluate the performance of decision tree in predicting pasture productivity and investigating the interrelationship between pasture productivity and environmental and management factors in hill-pasture, New Zealand. A common method to assess a new modelling approach is comparing it with a well-known one with respect to model fit and validation (Rykiel Jr 1996; Mitchell 1997). As the most commonly used modelling approach and the "cornerstone" of empirical models, the regression model provides an intuitive standard of model performance (Scheffer 2002). A comparison of decision tree model with a regression model will give a clear indication of how well it performed. Thus, the second aim of this study was to compare and assess the decision tree with regression for modelling pasture productivity in terms of model fit, validation and predictive accuracy.

3.2. Materials and methods

3.2.1. Study area

This study covers most of the hill-pasture in the North Island, New Zealand (Fig. 1). Hill-pasture in New Zealand are naturalized vegetation originated from clearance of native podocarp rainforest, fern or scrub and over-sown with introduced pasture species by early European settlers (Hilgendorf 1936). There are about five million hectares of hill-pasture in New Zealand, which accounts for nearly one third of the total pasture area in New Zealand, mainly located in the North Island (White 1990).

Climate is very diverse in the hill-pasture of the North Island with mean annual rainfall ranging from 800-1600 mm and mean annual daily temperature ranging from 9 - 15 ^{0}C (Tomlinson & Sansom 1994). Although large areas of hill-pasture receive more than 1000 mm rainfall per year, late spring and summer usually have less rainfall than autumn and winter. This results in a seasonal lack of water for most of the hill-pasture, which is a major factor limiting pasture productivity, and is a primary cause of year-to-year variation of pasture production (Rickard *et al.* 1985; Radcliffe & Barrs 1987; Barker & Dymock 1993).

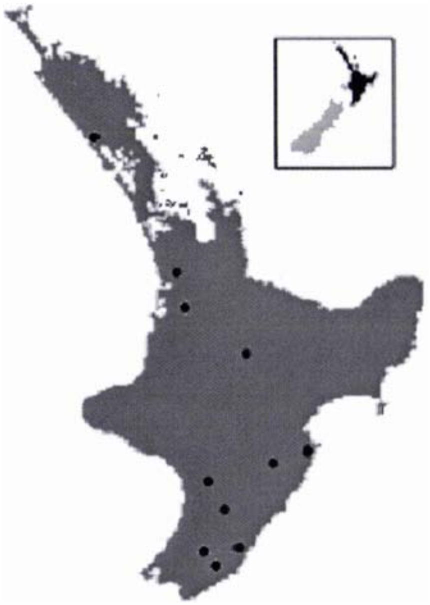


Fig. 1. The study area and the main sampling locations

Topographic features such as slope, aspect, and altitude play an important role in hill-pasture ecosystem processes (Radcliffe 1982). In southern latitudes, the north aspect receives more net radiation, is warmer, and generally drier than the south aspect. Depending on season and aspect, hill slope also modifies solar radiation received (McAneney & Noble 1975), and, usually, increasing slope reduces soil moisture status (Radcliffe & Lefever 1981). Aspect and slope also influence soil nutrient status by influencing animal grazing and excreta patterns (Gillingham 1982; Sheath & Boom 1985).

Over-sowing legume species like white clover (*Trifolium repens*) and lotus (*Lotus pedunculatus*), and top-dressing P fertiliser (mainly single superphosphate, approximately 9% of P) have been the major management practices in hill-pasture. P fertilizer stimulates legume growth which in turn fixes atmospheric nitrogen and encourages grass growth (White 1990). P fertilizer input and application history are the important indicators of soil fertility status in hill-pasture (McCall & Thorroli 1991).

3.2.2. Database setup

Developing a decision tree requires a large amount of training data, and the decision tree works best if sufficient samples are available (Clark & Pregibon 1992; Iverson & Prasad 1998). In the last several decades, considerable research has been conducted on hill-pasture in New Zealand (e.g. Suckling 1975; Radcliffe 1982; Lambert *et al.* 1996; Nicholas 1999; López 2000; Moir 2000; Gillingham 2001; Blennerhassett 2002); most of it related to pasture productivity. There is a large amount of data in the literature and kept by researchers in the form of raw or unpublished data, which provides a valuable resource to develop the decision tree for hill-pasture productivity. In establishing the dataset, an effort was made to collect as many climatic, environmental and management variables as possible, aiming to cover the most important factors influencing pasture productivity. Data for pasture productivity (annual and seasonal productivity), soil property (bulk density, pH, Olsen-P), fertiliser management (N, P fertiliser inputs) and topographic features (slope and aspect) were obtained from the literature and from researchers providing the 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. Potential evapotranspiration (PET) was derived from temperature and solar radiation using Priestley & Taylor's equation (Priestley & Taylor 1972). Since the rainfall data was measured from the horizontal surface and the P fertiliser was commonly aerially applied to hill-pasture with a uniform rate, the actual rainfall and P fertiliser received were adjusted by the cosine of the slope angle. Actual solar radiation on hilly terrain is strongly influenced by topography (Antoni 1998), so it was calculated from observed solar radiation at that location by an empirical method developed for the North Island, New Zealand (McAneney & Noble 1975).

In New Zealand, pasture productivity is commonly measured by cutting re-growth from a trimmed quadrat (usually 1 m^2 in area) with a cutting period of about one to two months depending on the growth rate of plants (i.e. trimming technique) (Lynch & Mountier 1954). Productivity is usually measured for a period of one year, and seasonal productivity is calculated from cuttings covering the season with spring from September to November.

This dataset covers most types of hill-pastures in the North Island (Fig. 1). There were 37 variables, including 32 input variables (independent variables) and five target variables (dependent variables), and 1900 samples in this dataset (Table 1).

3.2.3. Models development

The decision tree models for annual and seasonal pasture productivity were analysed in SAS Enterprise Miner, Version 4.1 (SAS Institute Inc., 1999-2001, Cary, NC, USA). The regression models for annual and seasonal pasture productivity were analysed in SAS 8.2 (SAS Institute Inc., 1999-2001, Cary, NC, USA). Annual and seasonal pasture productivity data were square root transformed to reduce the heterogeneous variances in residual as all random errors have the same variance is a fundamental assumption underlying linear regression (Sen & Srivastava 1990). The transformed data were also used in the decision tree analyses to make the results comparable with those of the regression models, although it is not essential for the decision tree analysis to assume a normal distribution of the target variable (Vayssieres *et al.* 2000). All the outputs of the decision tree models were back-transformed to make them practicable. In the development of the decision trees and regression models for annual and seasonal pasture productivity, the whole dataset was randomly partitioned into two parts: the training data (70% of the total) and the validation data (30% of the total). The decision trees and regression models were trained by the training data and validated by the validation data.

3.2.3.1. Decision tree

The main procedure for developing a decision tree model is to split the data of target variable based on its response to input variables. In SAS Enterpriser Miner, there are different criteria used to split the data, depending on the nature of the target variable.

For interval variables like pasture productivity, the split criterion is variance reduction, or F test. In the case of variance reduction, a complete search is applied to all the input variables and possible split-points to select one variable that ultimately explains the variance of the target variable by splitting the dataset of target variable into two sub-datasets. Then the same procedure is recursively applied to each of the sub-dataset until the dataset cannot be further split based on defined rules. Suppose D is the dataset containing the target variable and is going to be split into two sub-datasets D_L and D_R. DEV_D is the deviance of dataset D (squared error from the mean):

$$DEV_D = \sum_{obsi} (Yi - Y_{bar}) \qquad (1)$$

Where Yi is an observation and Y_{bar} is the mean of the target variable D. The variance reduction by splitting dataset D into D_L and D_R is expressed as follows:

$$\Delta = DEV_D - (DEV_L + DEV_R) \qquad (2)$$

Where DEV_L and DEV_R are the deviance of D_L and D_R, respectively. An input variable and a possible split-point that maximise Δ, or minimise the sum of (DEV_L + DEV_R) is first selected to split the data D, and then the same search is implemented for each sub-dataset D_L and D_R and the following sub-datasets recursively (SAS Online Help: Getting Started with Enterpriser Miner Software).

In the case of F test, an input variable and a possible split-point that gives the most significant p-value in F test associated with variance reduction is selected.

We compared both criteria in developing the decision trees and chose variance reduction as the split criterion since it gave very similar patterns of variance reduction between training data and validation data in the model assessment (see later). However, the decision tree outputs from using these two criteria were very similar.

3.2.3.2. Regression

Regression models were analysed using the same training data used in the decision tree analyses. Since there were too many factors significant in the regression analysis of annual pasture productivity, resulting in a serious multicollinearity problem, a principal component analysis was applied to reduce the dimensions of the input variables. The first nine principal components, which accounted for 89% of the total variance, were chosen as model inputs for the regression analysis.

Table 1. Variables used in the decision tree and regression analyses

Variable symbol	Units	Range	Variable description
input variable			
pH	$-\log_{10}[H^+]$	4.7-5.9	soil pH
BD	g/ cm^3	0.75-1.20	soil bulk density
OlsenP	µg/g	3.8-160.0	soil Olsen P
N_fert	kg/ha	0-180	annual elemental N fertiliser input per ha
P_fert	kg/ha	0-76	annual elemental P fertiliser input per ha
P_fert_5	kg/ha	0-268	5 years cumulative elemental P fertiliser input per ha
P_fert_10	kg/ha	0-429	10 years cumulative elemental P fertiliser input per ha
temp_y	^0C	10.1-15.4	annual mean daily temperature
temp_sp	^0C	10.4-14.4	spring mean daily temperature
temp_su	^0C	14.8-19.5	summer mean daily temperature
temp_au	^0C	12.1-16.2	autumn mean daily temperature
temp_wi	^0C	7.2-11.4	winter mean daily temperature
rain_y	mm	417-1727	annual rainfall
rain_sp	mm	71-501	spring rainfall
rain_su	mm	37-545	summer rainfall
rain_au	mm	149-311	autumn rainfall
rain_wi	mm	152-429	winter rainfall
rain_warm	mm	117-850	sum of spring and summer rainfall
solar_y	MJ/m^2	4.17-19.32	annual mean daily global solar radiation
solar_sp	MJ/m^2	6.74-22.18	spring mean daily global solar radiation
solar_su	MJ/m^2	9.36-24.31	summer mean daily global solar radiation
solar_au	MJ/m^2	1.16-16.59	autumn mean daily global solar radiation
solar_wi	MJ/m^2	0.55-14.90	winter mean daily global solar radiation
PET_y	$mm/m^2/d$	0.34-3.19	annual mean daily evapotranspiration
PET_sp	$mm/m^2/d$	0.82-3.99	spring mean daily evapotranspiration
PET_su	$mm/m^2/d$	1.40-4.40	summer mean daily evapotranspiration
PET_au	$mm/m^2/d$	-0.23-2.76	autumn mean daily evapotranspiration
PET_wi	$mm/m^2/d$	-0.40-2.15	winter mean daily evapotranspiration
aspect			hill slope aspect. eg. N NE S SW NW, etc.
aspect_adj			sunny (NW, N, NE, E), shady (SE, S, SW, W)
slope	degree	2-45	hill slope angle
legume_p	%	0-81.2	The relative abundance of legume in pasture
target variable			
annual productivity	kg/ha/y	793-25763	annual aboveground dry matter per ha
spring productivity	kg/ha/season	117-11275	spring aboveground dry matter per ha
summer productivity	kg/ha/season	39-7699	summer aboveground dry matter per ha
autumn productivity	kg/ha/season	116-4598	autumn aboveground dry matter per ha
winter productivity	kg/ha/season	10-6739	winter aboveground dry matter per ha

Table 2 shows the eigenvectors (CORR) of the input variables for the first nine principal components. Values in bold font indicate the main input variables contributing to the principal components. The main variables contributing to the first principal component were solar radiation and PET. The main variables contributing to the second principal component were temperature and rainfall. The main variables contributing to the third principal component were P fertiliser input and soil pH. For the regression analyses of seasonal pasture productivity, we used the original variables as model inputs since the multicollinearity problem was manageable. A forward stepwise approach was used to select the variables with a significance level $P<0.05$ in analysing the regressions of annual and seasonal pasture productivity. As stepwise approach had limitations in selecting significant variables (e.g. it can select variables that significant but meaningless for the studied system or it can eliminate a variable from model once a significant but meaningless variable is included), backward and forward approaches were also used to facilitate the variable selection. Residuals were checked for identifying outliers and unexplained variance, and the condition index was checked for identifying collinearity. Further improvements of model fit were then made accordingly.

3.2.4. Model assessment and empirical validation

Model assessment is an important step in developing the decision tree, as it enables a good model output, and prevents overtraining, which otherwise would develop a model that only fits the data from which it was trained (Witten & Erank 2000). The SAS Enterpriser Miner has several options to assess the model having an interval target (i.e. continuous variable) (Fernandez 2003). The measure of average squared error (ASE) is a commonly used option. ASE is similar to mean squared error (MSE) in general linear model and is an indicator of model goodness-of-fit; the smaller the ASE, the better the model fits. For the assessment of decision tree, if there are validation data available, it is assessed by the ASE of the validation data; otherwise, it is assessed by the ASE of the training data. The first option is better (SAS Online Help: Getting Started with Enterpriser Miner Software). When validation data are available, a "best" tree that has the smallest ASE from validation data is selected. An example of model assessment using the ASE of the validation data is illustrated in Fig. 2 for the decision tree model for annual pasture productivity.

Table 2. Eigenvectors (CORR) for the first nine principal components. Numbers in bold font indicate the main contributions of input variables to the principal components.

Variables	Princ.1	Princ.2	Princ.3	Princ.4	Princ.5	Princ.6	Princ.7	Princ.8	Princ.9
pH	0.021	0.046	**0.370**	0.046	-0.240	-0.121	-0.130	**0.328**	-0.255
BD	0.094	0.241	-0.071	0.101	-0.054	0.218	0.246	**0.298**	0.161
N_fert	0.024	-0.009	-0.050	0.049	0.044	**-0.482**	-0.115	**0.632**	-0.082
OlsenP	-0.006	0.117	0.190	-0.053	-0.207	0.120	**-0.425**	-0.071	**0.564**
P_fert	-0.017	-0.217	0.110	-0.207	-0.096	**0.349**	-0.101	0.208	-0.156
P_fert_10	-0.030	-0.116	**0.393**	**-0.303**	-0.231	0.104	0.083	0.055	0.043
P_fert_5	-0.039	-0.127	**0.353**	**-0.357**	-0.187	0.112	0.109	0.103	0.084
temp_y	0.030	**0.304**	0.237	0.087	0.151	0.046	-0.034	-0.072	-0.075
temp_sp	0.025	**0.293**	0.219	0.132	0.106	-0.040	-0.008	0.040	0.169
temp_su	0.027	**0.313**	0.214	0.053	0.062	-0.116	0.025	-0.070	-0.129
temp_au	0.005	**0.278**	0.176	-0.039	0.122	0.169	-0.124	-0.142	-0.146
temp_wi	0.042	0.233	0.204	0.221	0.246	0.190	-0.028	0.052	0.076
t_min_w	0.016	0.201	0.291	-0.047	-0.073	**-0.353**	0.232	-0.093	-0.192
rain_y	0.044	-0.272	0.197	**0.336**	0.050	0.075	0.014	-0.054	-0.050
rain_sp	0.027	**-0.290**	0.163	0.131	-0.012	-0.016	0.108	-0.154	-0.229
rain_su	0.031	-0.172	0.111	0.361	0.009	-0.046	0.021	0.255	**0.430**
rain_au	0.021	-0.193	0.213	0.225	-0.041	-0.144	0.294	-0.318	0.150
rain_wi	0.048	-0.127	0.072	0.282	0.138	0.374	**-0.362**	0.093	**-0.372**
rain_warm	0.036	**-0.294**	0.173	0.289	-0.004	-0.036	0.085	0.035	0.079
solar_y	**0.346**	-0.020	0.000	-0.058	0.034	0.012	0.035	0.001	0.014
solar_sp	**0.320**	-0.066	-0.101	-0.086	-0.093	-0.055	-0.141	-0.039	0.024
solar_su	0.044	-0.174	0.099	-0.263	**0.477**	-0.037	0.059	0.052	0.100
solar_au	**0.335**	-0.007	0.007	-0.026	0.120	0.068	0.040	0.084	0.023
solar_wi	**0.316**	0.046	-0.027	0.040	-0.244	-0.036	0.014	-0.088	-0.070
PET_y	**0.346**	0.003	0.018	-0.050	0.044	0.017	0.035	-0.005	0.008
PET_sp	**0.328**	-0.022	-0.068	-0.069	-0.076	-0.062	-0.138	-0.032	0.053
PET_su	0.051	-0.121	0.143	-0.266	**0.507**	-0.062	0.066	0.041	0.078
PET_au	**0.334**	0.007	0.017	-0.028	0.124	0.079	0.035	0.074	0.014
PET_wi	**0.318**	0.052	-0.019	0.048	-0.229	-0.027	0.019	-0.084	-0.065
aspect_A	**0.311**	-0.028	0.022	-0.029	0.040	0.012	0.078	-0.036	-0.014
slope	-0.021	0.135	-0.123	0.044	-0.098	**0.380**	**0.577**	0.247	-0.014

There are also some other ways to prevent over-training of a decision tree. One is to set the maximum tree levels, which defines the size of the tree and prevents the tree becoming too complex. Another one is to set the minimum observations required for a split search, which prevents a few special cases influencing the output of the tree. We set six as the maximum tree levels and 13 observations as the minimum observations required for a split search in developing the decision tree models for annual and seasonal pasture productivity. The recommendation on minimum observations was derived using the program by assessing the size of the whole dataset (Fernandez 2003).

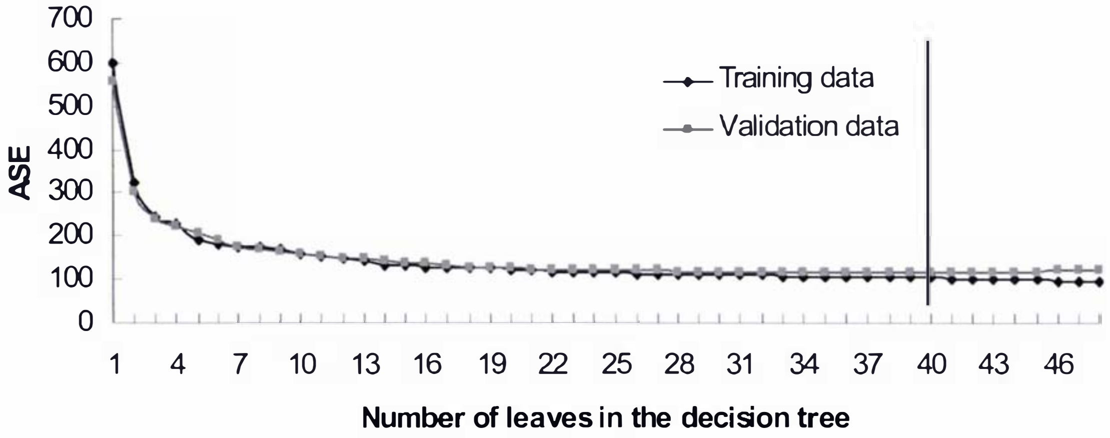


Fig. 2. Model assessment in developing the decision tree model for annual pasture productivity using average squared error (ASE) of the validation data. The vertical line indicates the smallest ASE from the validation data and the number of leaves selected for the final tree.

After the decision trees and regression models were developed, an empirical validation was implemented for all the decision trees and the regression models using the validation data. This empirical validation graphs the deviations (the differences between the predictions and the observations) against observations of the validation data, and checks the percentage of adequately predicted cases out of the whole validation data based on a given acceptable error (Mitchell 1997). This method is a more stringent model validation approach than methods using regression or correlation between model predictions and observations, which were criticised as not appropriate by some modellers (Harrison 1990; Mitchell 1997). Considering the potential variance heterogeneity in pasture productivity resulting from different sampling times, locations and managements, and the sampling error in measuring pasture productivity, we chose

±20% of the observation as an acceptable error of prediction. The percentage of deviations within ±20% of observations is considered as a standard for model predictive accuracy (Mitchell 1997).

3.3. Results

3.3.1. Decision trees and regression models

The decision tree models for annual, spring, summer, autumn and winter pasture productivity are given in Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7.

In each decision tree, the value in the upper most rectangle is the average pasture productivity. 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. Further prediction is made recursively based on the next significant variable in the same way for the first prediction, until the tree leaves, which are the final predictions in the decision tree, are reached. For example, in the decision tree model for annual pasture productivity (Fig. 3), the average pasture productivity for the whole study area was 7789 kg/ha, spring rainfall was the most significant factor selected to split the tree with a split-point of 212 mm. Pastures with a spring rainfall less than 212 mm had a predicted average productivity of 3280 kg/ha, while pastures with a spring rainfall equal to, or more than, 212 mm had a predicted average productivity of 9579 kg/ha. Further prediction was made by next splitting factor (slope) and a split-point (22.5^{o} for pastures with less than 212 mm spring rainfall) after the first prediction, and so on until it reached to the final prediction. The final prediction of pasture productivity was made by a series of constraints defined by the input variables and their split-points. For example, if pastures with a spring rainfall less than 212 mm, a slope less than 22.5^{o}, an annual temperature less than 13.4 ^{o}C, a five-year cumulative P fertiliser input (elemental P) less than 4.4 kg/ha, and a soil bulk density less than 0.85 g/cm^{3}, then the predicted average pasture productivity was 2010 kg/ha.

The relative importance of environmental and management variables on pasture productivity in the decision trees was 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 productivity, whereas annual N fertiliser input and five-year accumulated P fertiliser input were the significant manageable variables. Slope was the most significant variable selected to split the decision tree models for spring and winter pasture productivity, and annual P fertiliser input and autumn rainfall were the most significant variables selected to split the decision tree models for summer and autumn pasture productivity.

The decision tree model for annual pasture productivity had 40 leaves with a predicted average pasture productivity ranging from 1332 kg/ha to 15820 kg/ha (Fig. 3). In comparison, the decision tree models for spring, summer, autumn and winter pasture productivity had only 25, 26, 26, and 28 leaves, respectively. The predicted average annual, spring, summer, autumn and winter pasture productivity were 7789, 3462, 2791, 1469 and 1250 kg/ha, respectively. Spring productivity accounted for about 40% of the total pasture productivity, while summer, autumn and winter productivity accounted for about 30%, 16% and 14% of the annual productivity, respectively.

The regression model for annual pasture productivity is displayed in Table 3. Details of the regression models for seasonal pasture productivity are not shown here for simplicity. A summary of them is given in Table 4. In general, all five regression models were very significant ($P < 0.001$), and had relatively high adjusted R^2 values. For annual pasture productivity, solar radiation, PET, P fertiliser and rainfall showed a positive effect, while temperature showed a negative effect. Regression models for seasonal pasture productivity selected a similar set of variables as in the decision tree models for seasonal pasture productivity. Interaction terms and quadratic terms were not selected in regression models for seasonal productivity as including them in the model did not substantially increase the model fit, but would result in a serious multicollinearity problem. However, as some factors such as P fertiliser and Olsen were correlated and were chosen in a model (Table 4), there was still a multicollinearity problem in each of the regression model for seasonal pasture productivity. I did not remove the entire correlated variable from the model because the existed multicollinearity did not affect the estimation of target variable but the responses of some coefficients of input variables (Freund & Littell 1991), and I intended to maximise the model fit for the purpose of comparing them with the decision tree models to assess the performance of the decision tree models..

Table 3. Regression model for annual pasture productivity

a. Model fit

Mean of Response	91.6109	R-Square	0.6437
Root MSE	14.4702	Adj R-Sq	0.6424

b. Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	7	713933.9795	101990.5685	487.09	<.0001
Error	1887	395111.4314	209.3860	.	.
C Total	1894	1109045.4110	.	.	

c. Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	F Stat	Pr > F
Intercept	1	91.6109	0.3324	275.60	<.0001	.	
PCR1	1	1.1351	0.1163	9.76	<.0001	95.32	<.0001
PCR2	1	-5.5353	0.1237	-44.74	<.0001	2001.44	<.0001
PCR3	1	6.8950	0.1976	34.90	<.0001	1217.73	<.0001
PCR6	1	-0.5933	0.2570	-2.31	0.0211	5.33	0.0211
PCR7	1	-1.5719	0.2990	-5.26	<.0001	27.64	<.0001
PCR8	1	-1.8178	0.3446	-5.27	<.0001	27.82	<.0001
PCR9	1	-2.0561	0.3507	-5.86	<.0001	34.36	<.0001

3.3.2. Model assessment and empirical validation

A comparison of the ASE of the decision trees and the regression models for annual and seasonal pasture productivity is displayed in Fig. 8. The ASE of each decision tree was smaller than that of the corresponding regression model, indicating the better model fit of decision tree than regression. The ASE of the decision tree model for annual pasture productivity was approximately half of that of the regression of annual pasture productivity. However, the difference between ASE of each decision tree model for seasonal pasture productivity and the corresponding regression model was small.

Table 4. Summary of the regression models for annual and seasonal pasture productivity. See Table 1 for variable symbols.

Models	Input variables	F value	Pr>F	Adj R^2
Annual productivity	principal components	487.09	<0.0001	0.642
Spring productivity	BD, N_fert, OlsenP, P_fert, P_fert5, aspect, slope, rain_wi, solar_wi, solar_sp	122.26	<0.0001	0.656
Summer productivity	pH, N_fert, P_fert, P_fert5, aspect, slope, rain_sp, temp_sp, solar_su, legume_p	106.62	<0.0001	0.664
Autumn productivity	BD, N_fert, P_fert, aspect, slope, rain_au, rain_su, temp_au, solar_au	130.34	<0.0001	0.641
Winter productivity	N_fert, OlsenP, P_fert, P_fert5, aspect, slope, solar_ua, rain_wi, temp_wi	91.24	<0.0001	0.587

Empirical validations of the decision tree and the regression of annual pasture productivity indicated that the decision tree had 91% adequately predicted cases based on the validation data, while the regression had 80% (Fig. 9). The deviations of the decision tree model for annual pasture productivity were within ±40, while those of the regression of annual pasture productivity were within ±50.The results of empirical validations of both decision trees and regressions of annual and seasonal pasture productivity were summarised in Fig. 10. In general, the percentage of adequately predicted cases for each decision tree was higher than the corresponding regression model.

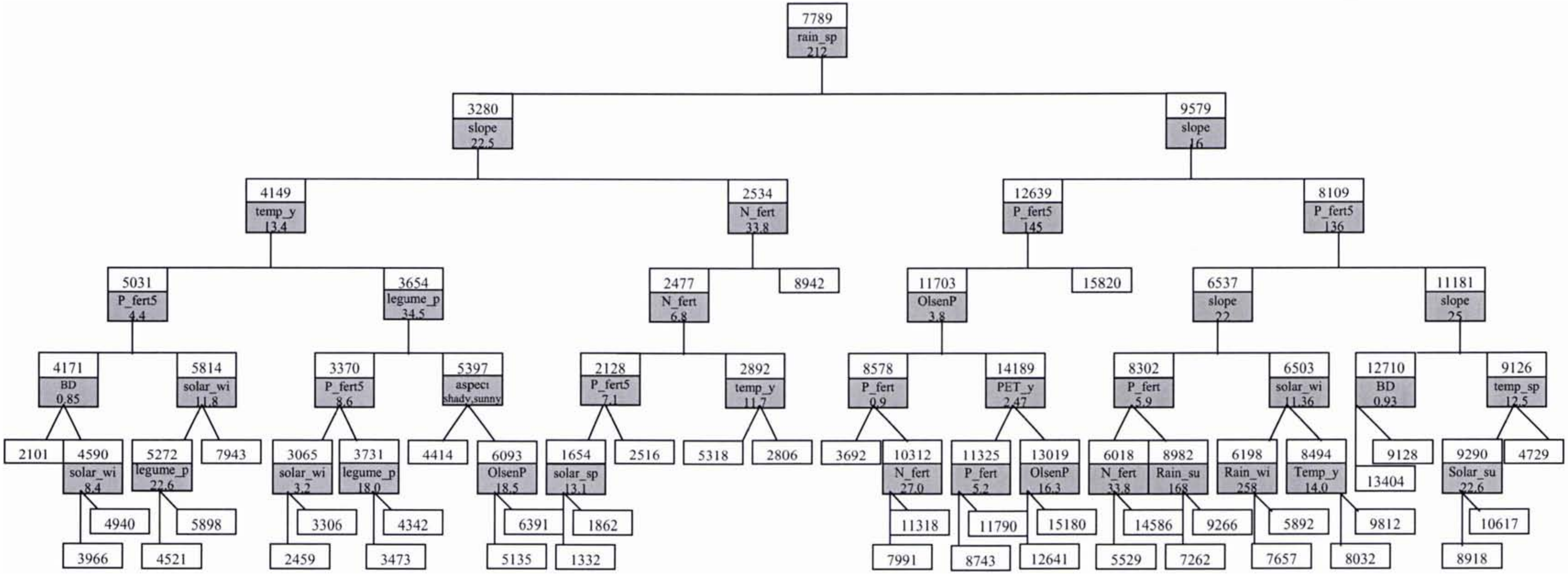


Fig. 3. The decision tree model for annual pasture productivity. Predicted productivities are in the un-shaded rectangles, splitting variables and split-points are in the shaded rectangles. Prediction goes to the left-side branch when the splitting variable is less than the split-point, and goes to the right-side branch when the splitting variable is equal to, or more than, the split-point (in the case of aspect, pasture with aspect set before the comma go to left branch, others go to right branch). See Table 1 for variable descriptions and units.

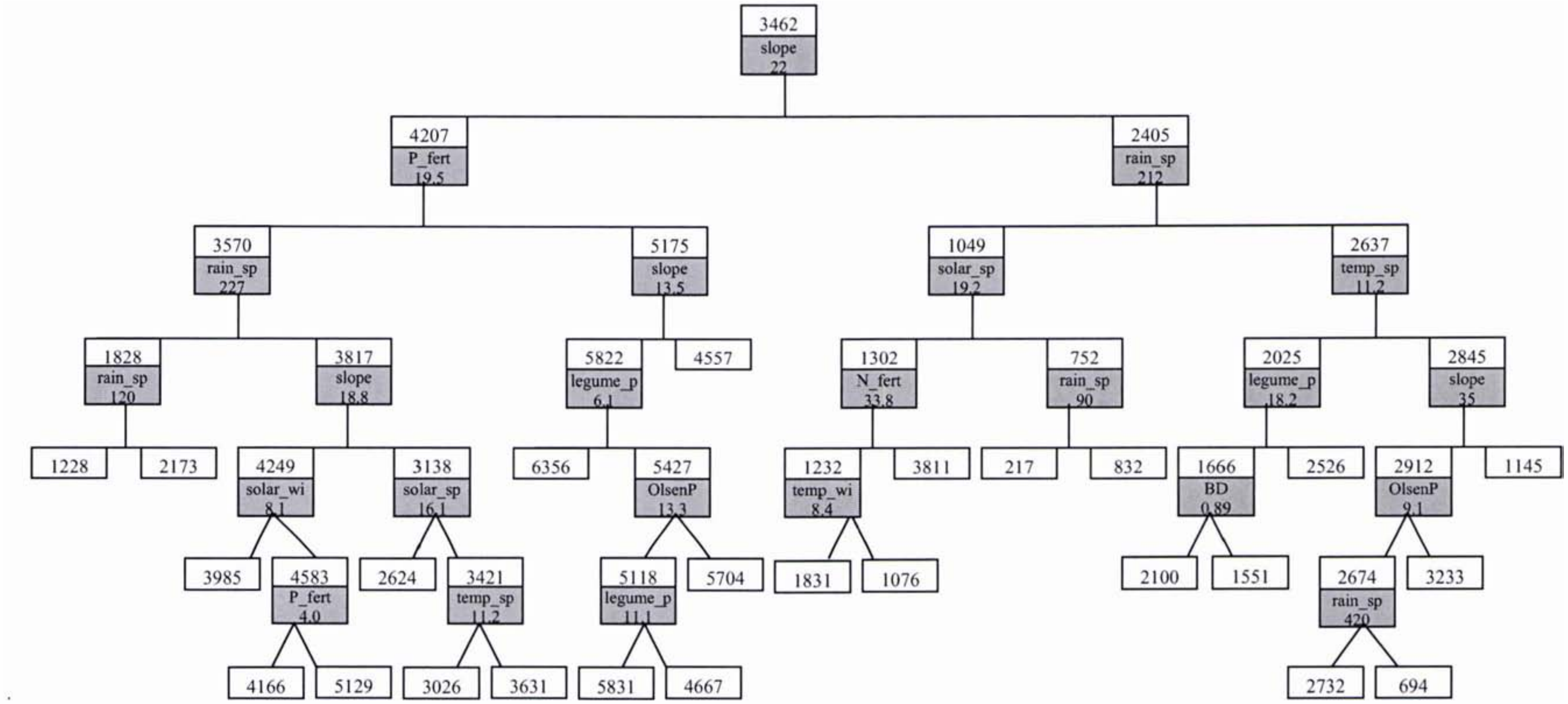


Fig. 4. The decision tree model for spring pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

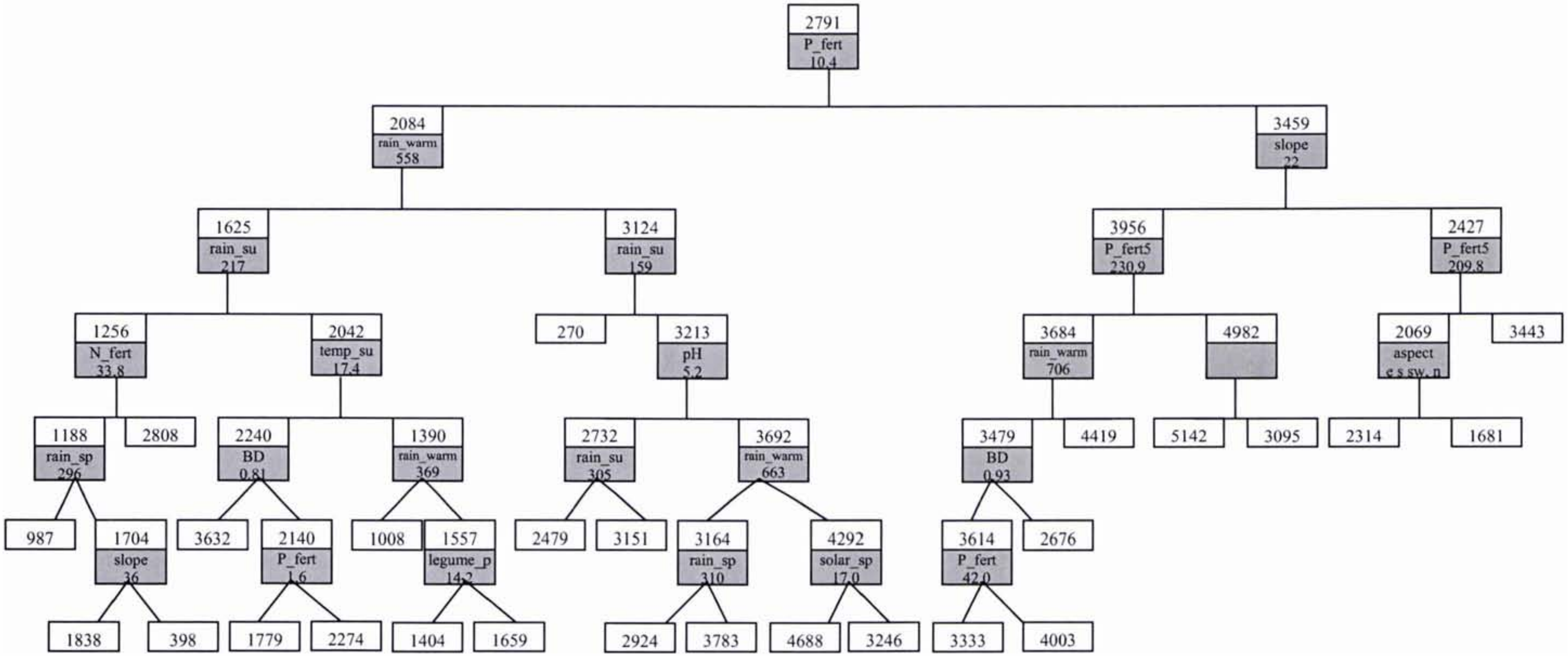


Fig. 5. The decision tree model for summer pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

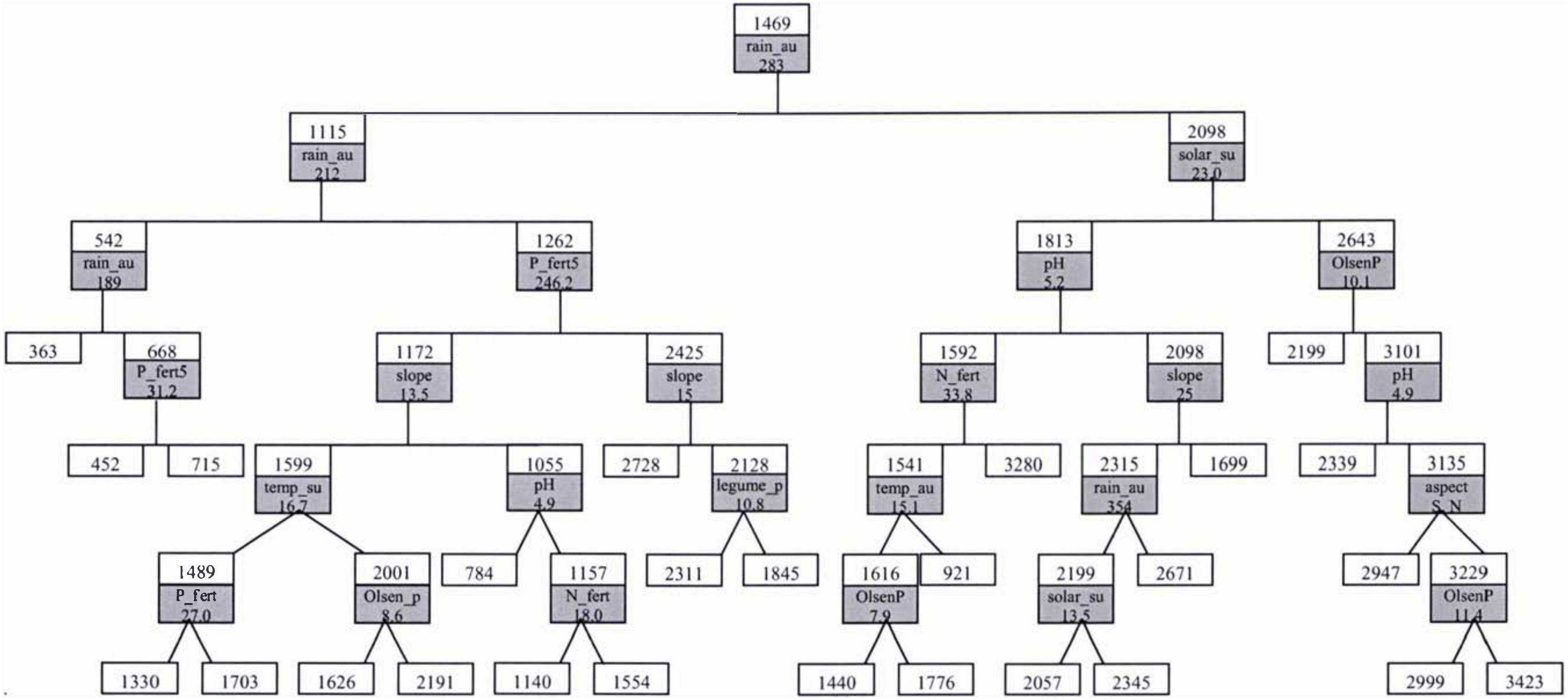


Fig. 6. The decision tree model for autumn pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

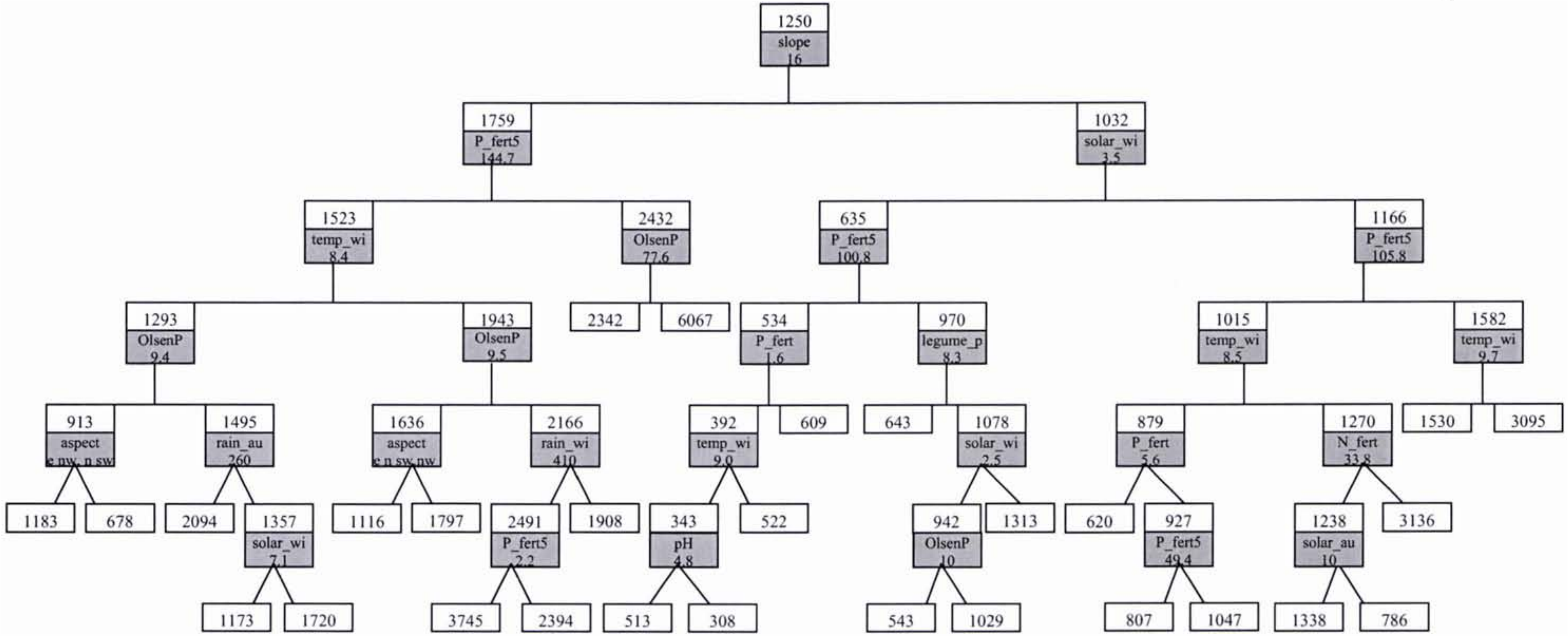


Fig. 7. The decision tree model for winter pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

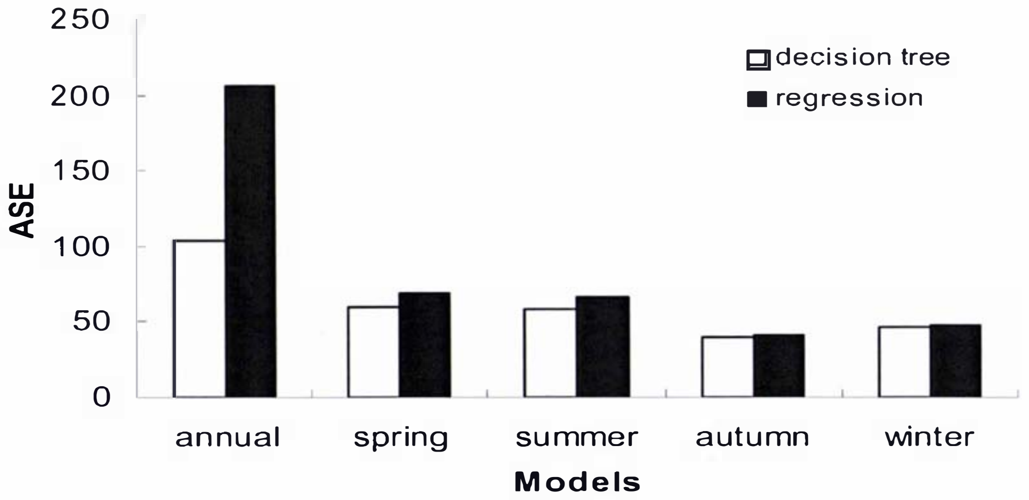


Fig. 8. Comparison of average square error (ASE) between the decision trees and the regression models for annual and seasonal pasture productivity

3.4. Discussion

3.4.1. Performance of model

Both the decision trees and regression models were satisfactory in predicting annual and seasonal pasture productivity, but the decision trees performed better with respect to model fit and predictive accuracy.

The regression models generally met the basic assumptions of regression analysis, and were all highly significant ($P < 0.001$) (Table 4) and explained a relatively high percentage of the variance of model targets considering the large sampling area, and the heterogeneity of the environment and management over the area. They also adequately predicted a considerable proportion of cases in the model empirical validation (Fig. 10). Compared to the regression models, the decision trees all had a smaller ASE (Fig. 8) and a higher percentage of adequately predicted cases (Fig. 10) than the corresponding regression models. For example, the ASE of the decision tree model for annual pasture productivity was only about half of that of the regression model for annual pasture productivity and had 90.1% adequately predicted cases, which was 10.8 percentage points higher than that of the regression model. This smaller ASE and higher percentage of adequately predicted cases for the decision trees indicated that they had better performance than the regression models in predicting hill-pasture productivity.

Similarly, an application of decision tree in predicting plant species distribution (Vayssieres *et al.* 2000) showed that the decision tree performed significantly better than a polynomial logistic regression model for four of the six cases considered, and as well as in the two remaining cases. 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.

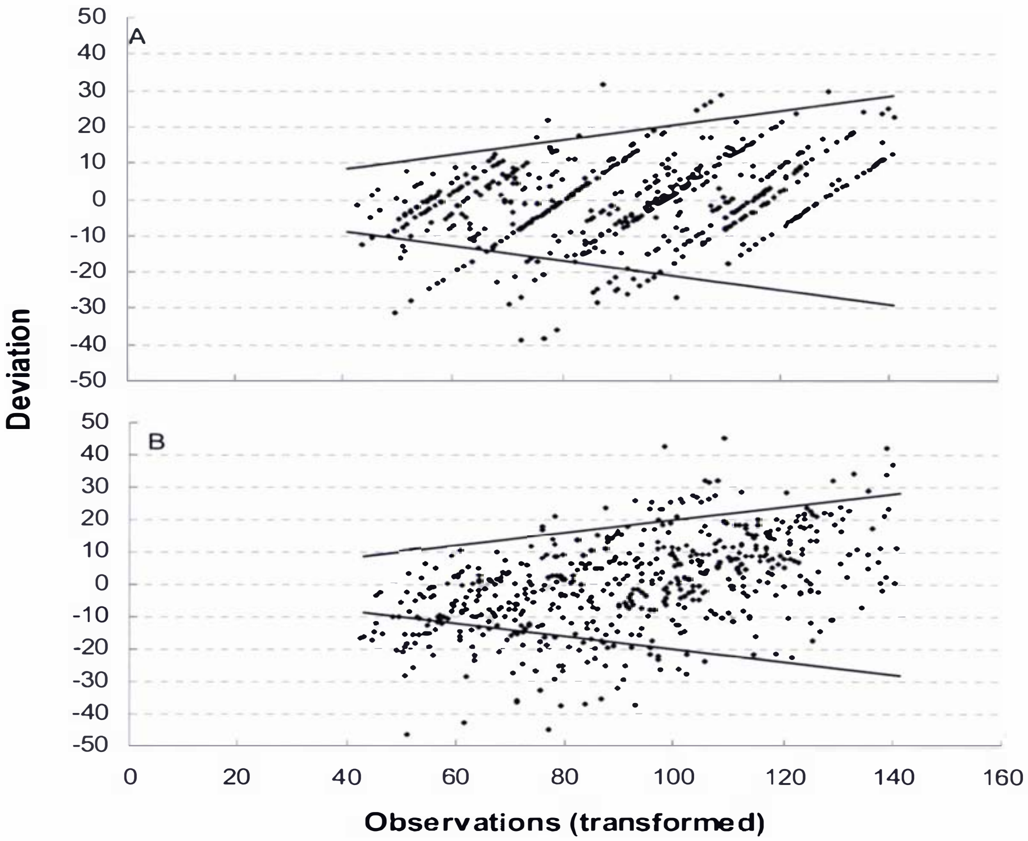


Fig. 9. Empirical validation of the decision tree (A) and the regression model (B) for annual pasture productivity. Deviation is the difference between predicted and observed pasture productivity. The two spreading lines show the ±20% of the observations, and 91% of predictions in the decision tree and 80.1% of predictions in the regression are within the ±20% of the observations of the validation data. Both observation and deviation are transformed data for pasture productivity.

However, the performance of the decision tree models for seasonal pasture productivity was not as good as that of the decision tree model for annual pasture productivity. This may be firstly because the decision tree model performs better for a complicated situation with more significant input variables such as this for the annual productivity. Secondly, the quality of seasonal pasture productivity data was not as good as annual pasture productivity data. Most of the seasonal pasture productivity was calculated from the pasture cuttings covering a season (Lambert et al. 1996), which may allocate the production that occurred in one season to another season, and introduce errors into the seasonal pasture productivity data. For example, a cutting covering a period from early August to late September would be evenly allocated to winter and spring in calculating the seasonal pasture productivity, but spring will usually have more production than winter since it is warmer. Also, carryover effects from season to season (e.g. due to drought) are likely to be more influential compared to carryover effects from year to year. This suggests the quality of a decision tree will ultimately depend on the quality of the training data.

Compared to other pasture and productivity models, the decision tree model for pasture productivity has several significant characteristics. Paruelo and Tomasel (1997) used another data mining approach – artificial neural network (ANN) to predict pasture ecosystem attributes and compared their performances with regression models. They also found that ANN had better predictive ability than regression model. However, unlike the decision tree model, the output of ANN is very difficult to interpret (Scheffer 2002). As an empirical-oriented modelling approach, the decision tree models can only be used in the same or similar areas as their ability to extrapolate beyond the special scope is very limited. On the other hand, mechanistic pasture simulation models such as PaSim (Riedoa *et al.* 2002), SPUR (Foy *et al.* 1999), HP-model (Riedo et al. 1998) and CENTURY (Gilmanov et al. 1997) are more general and widely applicable because of their strong ability in explicitly revealing the mechanisms of the systems they represent (Rickert et al. 2000), but these mechanistic simulation models generally have a poor performance with respect to their predictive ability (Guisan and Zimmermann 2000).

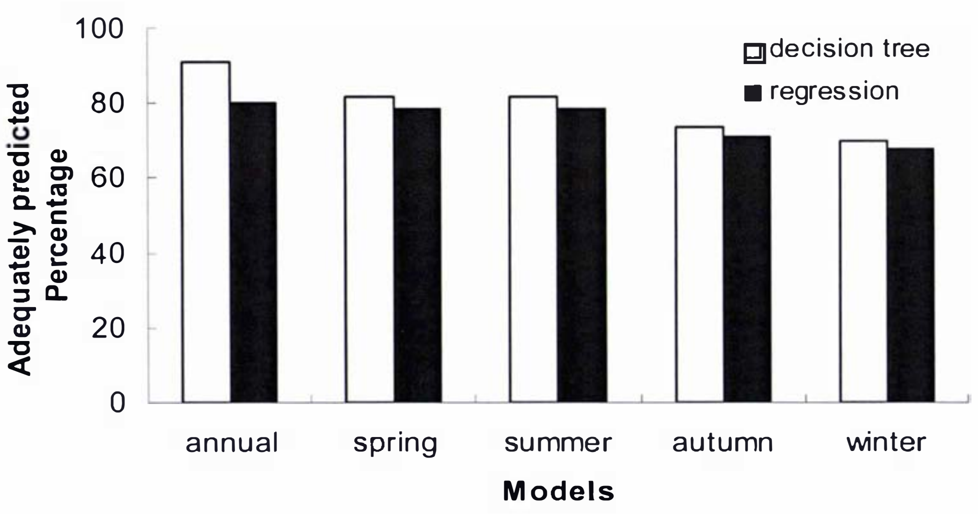


Fig. 10. Percentage of adequately predicted cases of the decision trees and the regression models for annual and seasonal pasture productivity in the model empirical validation.

3.4.2. Insights from the decision trees

3.4.2.1. General pattern

Besides having a better model fit and higher predictive accuracy than the regression models, the hierarchical structure of the decision trees also clearly revealed the relative importance of input variables in influencing pasture productivity. For example the decision tree model for annual pasture productivity showed that available spring rainfall was the most significant factor influencing hill-pasture productivity (Fig. 3). Hill slope was the second most significant variable influencing pasture productivity for pastures with both high and low spring rainfall. Annual mean daily temperature, N and P fertiliser inputs and soil fertility status (Olsen-P) were the significant variables influencing pasture productivity under the specific climatic and topographic conditions.

This hierarchical structure of the decision tree also revealed different response patterns of pasture productivity to the interaction of the environmental and management variables. For example, in the decision tree model for annual pasture productivity, when pastures received less than 212 mm rainfall in spring, those with a hill slope equal to, or more than, 22.5° responded to the management variable N fertiliser, but those with a hill slope less than 22.5° responded to annual daily mean temperature.

3.4.2.2. Effects of environmental and management variables

3.4.2.2.1. Climate

The available rainfall in spring was indicated as the most significant factor influencing pasture productivity in hill-pasture (Fig. 3). Pastures with equal to, or more than, 212 mm rainfall in spring had an average annual productivity which was 6300 kg/ha higher than pastures with less than 212 mm rainfall in spring. Though 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 productivity was not previously fully recognised. This may be because most research was conducted locally and the variation in climate at a local-scale was usually not very significant, and thus failed to detect the response of pasture growth to spring rainfall. In hill-pasture, autumn and winter usually have enough rainfall and summer is usually dry due to unreliable rainfall and high evapotranspiration (Woodward *et al.* 2001). The available spring rainfall, therefore, becomes a key factor influencing pasture productivity with about 40% of pasture production occurring in spring. This is partially in agreement with the result obtained by Radicliffe and Baars (1987) as they 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.

Heavy rainfall in winter can also reduce winter pasture productivity. For example, pastures received equal to, or more than, 410 mm rainfall in winter produced less productivity than those with less than 410 mm rainfall (Fig. 7), this may be because of the constraint of temperature and (or) solar radiation on pasture growth resulted from heavy rainfall.

High temperature increases evapotranspiration and reduces the available soil moisture (Bircham & Gillingham 1986). In the decision tree model for annual pasture productivity, when pastures received less than 212 mm rainfall in spring, those with an annual mean daily temperature less than 13.4 ^{0}C had an annual productivity of 5031 kg/ha, which produced about 1400 kg/ha more than pastures with an annual mean daily temperature more than 13.4 ^{0}C.

3.4.2.2.2. Fertiliser and soil fertility

It was indicated from the decision tree model for annual pasture productivity (Fig. 3) that pasture responded to N fertiliser better when on a high slope. Pastures with high slope generally have less legume than pastures with low slope (Ledgard et al. 1987) and N content in soil is inversely related to slope due to the nutrient transfer (Gillingham and During 1973, Ledgard et al. 1982). These factors result in N deficiency in soils with high slope, and may be the reasons why pastures with high slope respond to N fertiliser better. This better response to N fertiliser on pastures with high slope was also reported by Gillingham et al. (1998).

P fertiliser application is one of the most important managements in improving pasture productivity in hill-pasture. It was indicated from the decision tree model for summer pasture productivity that annual P fertiliser input had marked effect on summer pasture productivity (Fig. 5). Pastures received equal to, or more than, 10.4 kg/ha annual P fertiliser increased summer pasture productivity by 1400 kg/ha comparing with those received less than 10.4 kg/ha. Five-year cumulative P fertiliser input was the most significant factor influencing pasture productivity for both high and low slope when spring rainfall was equal to, or more than, 212 mm in the decision tree model for annual pasture productivity (Fig.3), indicating a strong interaction between P fertiliser input and available soil moisture.

The effect of Olsen-P on pasture productivity was greatly influenced by the available soil moisture. Pastures with more than 3.8 $\mu g/cm^3$ and less than 16.3 $\mu g/cm^3$ Olsen-P in the soil had an average annual productivity of 12641 kg/ha when they received more than 212 mm rainfall in spring, but only had an average annual productivity of 5135 kg/ha when pastures with a similar Olsen-P condition but received less than 212 mm rainfall in spring (Fig. 3). This interaction between moisture and Olsen-P on pasture productivity was reported by Moir *et al.* (2000) and may be one reason why previous research (Sinclair *et al.* 1997) failed to establish a reliable link between pasture productivity and soil Olsen-P content.

3.4.2.2.3. Topography

Hill slope, due to its important role in redistributing available moisture, fertility and solar radiation, greatly influences hill-pasture productivity (Gillingham & During 1973; Lambert & Roberts 1978). The impact of hill slope on pasture productivity is well recognised (Lambert *et al.* 1983; White 1990; Gillingham *et al.* 1998). In the decision tree models for annual and seasonal pasture productivity, the influence of slope on pasture productivity was fully expressed. Hill aspect, unlike hill slope, mainly influenced seasonal pasture productivity in some extent. Generally, shady aspects have higher productivity than sunny aspects during periods of soil moisture stress, whereas sunny aspects produce more production at other times (Lambert *et al.* 1983).

3.4.2.2.4. Legume

Legumes fix atmospheric nitrogen and encourage grass growth. The relative abundance of legume in hill-pasture is usually an indicator of pasture management status (Edmeades *et al.* 1990). High percentage of legume had a marked effect on annual pasture productivity, especially for pastures with less rainfall (Fig. 3). But it was indicated in the decision tree model for spring pasture productivity (Fig. 4) that in spring, high percentage of legume had a negative effect on pasture productivity for pastures with a low slope. For example, when slope was less than 13.5^{o}, pastures with equal to, or more than, 6.1% legume had a spring productivity of 5427 kg/ha, while pastures with less than 6.1% legume had a spring productivity of 6357 kg/ha. Another example was that pastures with equal to, or more than, 11.1% legume had a spring productivity of 4667 kg/ha, while pastures with less than 11.1% legume had a spring productivity of 5831 kg/ha. This may be because legumes especially white clover grow slower during spring than during summer-autumn period (Ledgard *et al.* 1987).

3.4.2.2.5. Bulk density and pH

Soil bulk density was revealed having a negative correlation with soil fertility and a positive correlation with slope in hill-pasture (López 2000). It was indicated from the decision trees that, in general, pastures with higher bulk density had less productivity than pastures with lower bulk density when other factors were same. However, pasture with higher bulk density may have higher productivity (Fig. 3) for pastures with low spring rainfall and low slope, indicating that the relationship between soil bulk density and pasture productivity was complex. For soil pH, the general pattern was that pastures with higher soil pH had higher productivity than pastures with lower soil pH. The responses of pasture productivity on soil bulk density and pH were in agreement with those of the previous research (Lambert *et al.* 1996).

3.4.2.3. Limitations of decision tree

Due to the binary nature of the decision tree, the responses of pasture productivity to the environmental and management variables are not continuous. This make the decision tree not being able to reflect the influence of small changes of input variables on the model target, especially when the input variable has a relatively large range. For example, in the decision tree model for annual pasture productivity (Fig. 3), when spring rainfall is equal to, or more than, 212 mm and slope is less than 16^o, pasture productivity was predicted by five-year cumulative P fertiliser input with a split-point of 144.7 kg/ha. This makes predictions either less or more than the split-point covering a large range of P fertiliser input, and could result in a large variance in the predicted pasture productivity. Some statistic programs have the option to split the tree into more than two groups, which might be useful for better prediction, but will create a tree that is too complex to deal with in practice.

The quality of a decision tree depends on the quality of the training data. If the training data lack the representation of some information, a decision tree would lack the response to that information (Iverson & Prasad 1998; Vayssieres *et al.* 2000). Since the seasonal pasture productivity was not from designed measurement, the decision tree models for seasonal pasture productivity did not perform as well as the decision tree model for annual pasture productivity. More seasonal pasture productivity data from designed measurements need to be gathered to improve the performance of the decision tree models for seasonal pasture productivity.

3.5. Conclusion

Decision tree, as a modelling approach, had better performance in predicting hill-pasture productivity than the regression model with respect to model fit and predictive accuracy. The decision trees clearly revealed the relative importance of environmental and management variables in influencing pasture productivity, and the interaction among these variables. Spring rainfall was indicated as the most significant factor influencing annual pasture productivity, while hill slope was the most significant factor influencing spring and winter pasture productivity, and annual P fertiliser input and autumn rainfall were the most significant factors influencing summer and autumn pasture productivity. One limitation of using decision tree to predict pasture productivity was that it did not generate a continuous prediction, and thus could not detect the influence of a small change in environmental and management variables on pasture productivity.

3.6. References

Antoni O. (1998) Modelling daily topographic solar radiation without site-specific hourly radiation data. *Ecological Modelling*, 113, 31-40

Barker D.J. & Dymock N. (1993) Introduction, production, and persistence of five grass species in dry hill country 6. Ballantrae, North Island, New Zealand. *New Zealand Journal of Agricultural Research*, 36, 41-47

Bircham J.S. & Gillingham A.G. (1986) A soil water balance model for sloping land. *New Zealand Journal of Agricultural Research*, 29, 315-323

Blennerhassett J.D. (2002) *Pasture growth constraints on dry steep east coast hill country.* Ph.D. thesis, Massey University, New Zealand

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) *Classification and regression trees.* The Wadsworth statistics/probability series. Chapman & Hall. Inc. New York, NY.

Clark L.A. & Pregibon D. (1992) Tree-based models. In: Hastie, T.J (ed.), *Statistical Models in S.*, Wadsworth, Pacific Grove, California, USA, 377-419

Dunham M.H. (2002) *Data Mining: Introduction and Advanced Topics.* Pearson Education INC. Upper Saddle River, New Jersey 07458

Dyne G.M.v. (1970) A systems approach to grasslands. *Proceedings of the XI International Grassland Congress, Surfers Paradise, Australia*, A131-A142

Edmeades D.C., Wheeler D.M., Rys G. & Smith N. (1990) Effect of pasture composition on lime and phosphorus responses on a dryland soil. *Proceedings of the New Zealand Grassland Association*, 52, 171-175

Fernandez G. (2003) *Data mining using SAS applications.* Chapman & Hall/CRC, Boca Raton, Florida

Foy J.K., Teague W.R. & Hanson J.D. (1999) Evaluation of the upgraded SPUR model (SPUR2.4). *Ecological Modelling*, 118, 149-165

Freund R.J. & Littell R.C. (1991) *SAS System for Regression* (second edition). *SAS Institute Inc., Cory, NC, USA, 94*

Gillingham A.G. (1982) Topographic and management effects on dung distribution by grazing sheep. *Proceedings of the New Zealand Grassland Association*, 43, 161-170

Gillingham A.G. (2001) Precision management of fertiliser application to pasture. *First Australian Geospatial Information and Agriculture Conference, Sydney, Australia, 17-19 July 2001*, 534-541

Gillingham A.G. & During C. (1973) Pasture production and transfer of fertility within a long-established hill pasture. *New Zealand Journal of Experimental Agriculture*, 1, 227-232

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

Gilmanov T.G., Parton W.J. & Ojima D.S. (1997) Testing the 'CENTURY' ecosystem level model on data sets from eight grassland sites in the former USSR representing a wide climatic/soil gradient. *Ecological Modelling*, 96, 191-210

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

Harrison S.R. (1990) Regression of a model on real-system output: An invalid test of model validity. *Agricultural Systems*, 34, 183-190

Hilgendorf F.W. (1936) *Pasture plants and pastures of New Zealand.* Christchurch, New Zealand, Whitcombe and Tombs Limited.

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., Barker D.J., Mackay A.D. & Springett J.O. (1996) Biophysical indicators of sustainability of North Island hill pastures. *Proceedings of the New Zealand Grassland Association*, 57, 31-36

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

Lambert M.G. & Roberts E. (1978) Aspect differences in an unimproved hill country pasture 2. Edaphic and biotic difference. *New Zealand Journal of Agricultural Research*, 21, 255-260

Ledgard S.F., Brier G.J. & Littler R.A. (1987) Legume production and nitrogen fixation in hill pasture community. *New Zealand Journal of Agricultural Research*, 30, 413-421

López I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Lynch P.B. & Mountier N.S. (1954) Cutting techniques in grassland experiments. *New Zealand Journal of Science and Technology*, 36A, 375-385

McAneney K.J. & Noble P.E. (1975) Estimating solar radiation on sloping surfaces. *New Zealand Journal of Experimental Agriculture*, 4, 195-202

McCall D.G. & Thorroli B.S. (1991) Fertiliser history is a useful predictor of soil fertility status. *Proceedings of the New Zealand Grassland Association*, 53, 197-202

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

Moir J.L. (2000) *Investigations into the influence of fertilizer history and climate regime on the soil fertility, soil quality and pasture production of Wairarapa hill country soil.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

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

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Paruelo J.M. & Tomasel F. (1997) Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecological Modelling*, 98, 173-186

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

Priestley C.H.B. & Taylor R.J. (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100, 81-92

Radcliffe J.E. (1982) Effects of aspect and topography on pasture production in hill country. *New Zealand Journal of Agricultural Research*, 25, 486-496

Radcliffe J.E. & Barrs J.A. (1987) Temperate pasture production. In: Snaydon, R.W. (Ed.) *Ecosystems of the World*, 17B, 7-17

Radcliffe J.E. & Lefever K.R. (1981) Aspect influences on pasture microclimate at Coopers Creek, North Canterbury. *New Zealand Journal of Agricultural Research*, 24, 55-56

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. In: Mannetje, L. and Jones, R.M.(Eds.), *Field and laboratory methods for grassland and animal production research*, CAB Publishing, Wallingford, UK: 2000, 29-66

Riedo M., Grub A., Rosset M. & Fuhrer J. (1998) A pasture simulation model for dry matter production, and fluxes of carbon, nitrogen, water and energy. *Ecological Modelling*, 105, 141-183

Riedoa M., Milforda C., Schmidb M. & Sutton M.A. (2002) Coupling soil–plant–atmosphere exchange of ammonia with ecosystem functioning in grasslands. *Ecological Modelling*, 158, 83-110

Rykiel Jr E. (1996) Testing ecological models: the meaning of validation. *Ecological Modelling*, 90, 229-244

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

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

Seligman N.G. & Baker M.G. (1993) Modelling as a tool for grassland science progress. In: Baker, M.G. (Ed.) *Grassland for Our World*, Sir Publishing, Wellington, New Zealand, 228-233

Sen A.K. & Srivastava M. (1990) *Regression analysis: theory, methods and applications.* New York : Springer-Verlag

Sheath G.W. & Boom R.C. (1985) Effect of November-April grazing pressure on hill country pastures 3. Interrelationship with soil and pasture variation. *New Zealand Journal of Agricultural Research*, 13, 341-349

Sinclair A.G., Johnstone P.D. & Smith L.C. (1997) Relationship between pasture dry matter yield and soil Olsen P from a series of long-term field trials. *New Zealand Journal of Agricultural Research*, 40, 559-567

Sorenson J.T. (1998) Modelling and simulation in applied livestock production science. In: Peart, R.M. and Carry, R.B. (Eds.) *Agricultural Systems: Modelling and Simulation*, Marcel Dekker Inc. New York, 475-494

Stuth J.W., Hamilton W.T., Conner J.C., Sheehy D.P. & Baker M.J. (1993) Decision support systems in the transfer of grassland technology. In: Baker, M.G. (Ed.) *Grassland for Our World*, Sir Publishing, Wellington, New Zealand, 234-242

Suckling F.E.T. (1975) Pasture management trials on unploughable hill country at Te Awa. 3. Result for 1959-69. *New Zealand Journal of Agricultural Research*, 3, 351-436

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

Tomlinson A.I. & Sansom J. (1994) *Temperature normals for New Zealand for the period 1961 to 1990.* Wellington [N.Z.]: NIWA

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. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

Witten I. & Erank E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers San Francisco, California

Woodward S.J.R., Barker D.J. & Zyskowski R.E. (2001) A practical model for predicting soil water deficit in New Zealand. *New Zealand Agriculture Science*, 44, 91-109

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

Chapter 4. Predictive Modelling of Hill-pasture Productivity: Integration of a Decision Tree and a Geographic Information System

The decision tree models were shown to perform better than the regression models in the prediction of annual and seasonal pasture productivity in the previous chapter. In this chapter decision tree models were integrated with a geographic information system (GIS) to develop a predictive modelling approach on hill-pasture productivity, with capabilities of incorporating spatial factors such as slope and aspect and presenting model output over space for areas of interest.

A paper derived from this chapter with a title "Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographic information system" by B. Zhang, I. Valentine, P.D. Kemp & G. Lambert has been published in ***Agricultural Systems*** *(in press, available on-line).*

Abstract. One challenge in predictive modelling of productivity for pastures varying in topography, soils or management is to achieve the prediction over space with acceptable accuracy. As a new modelling approach, the decision tree has been shown to have high predictive accuracy; while geographic information systems (GISs), with their strong ability to deal with spatial factors, have been widely used in environmental modelling. Integration of a decision tree approach with a GIS offers a potential solution in meeting this challenge. In this study, decision tree models were developed for annual and seasonal pasture productivity using environmental and management variables and the outputs of these decision trees were integrated with a GIS to get predictions of pasture productivity in a hill-pasture grazing system. Results showed that the decision tree model for annual pasture productivity was verified in three of four test farmlets. The decision tree models also revealed the relative importance of environmental and management variables and their interaction in influencing pasture productivity. Hill slope, soil Olsen P and annual P fertiliser input were the most significant variables influencing annual pasture productivity, while hill slope, annual P fertiliser input, autumn rainfall and soil Olsen P were the most significant variables influencing spring, summer, autumn and winter pasture productivity, respectively. The successful integration of the decision tree model with a GIS in this study provided a platform to predict pasture productivity for pastures with heterogeneous environmental variables and management features, and to present model predictions over space for further application and investigation. This modelling approach can be used as, or incorporated in, decision support systems to improve pasture management, and to investigate the interrelationship between pasture productivity and environmental and management variables.

4.1. Introduction

Modelling pasture productivity has long been an interest of plant ecologists and agronomists either for investigating ecosystem processes or estimating productivity as a substitute for observation (Dyne 1970; Seligman & Baker 1993; Stuth *et al.* 1993; Sorenson 1998; Moir *et al.* 2000). Predictive modelling of pasture productivity using environmental and management variables has both practical and ecological applications (Rickert *et al.* 2000). For example, it can be used or associated with decision support systems to improve pasture management (Stuth *et al.* 1993), and can also be applied to investigate the interrelationship between pasture productivity and environmental and management variables in a pasture ecosystem (Parton *et al.* 1993).

One challenge of predictive modelling of pasture productivity for pastures with variable management, edaphic or topographic features is to achieve the prediction over space with acceptable accuracy. It requires a model incorporating spatial factors and a way to present the prediction over space (Li *et al.* 1998; Wadsworth & Reweek 1999). This is especially the case for predictive modelling of hill-pasture productivity due to the heterogeneity of topographic and micro-topographic features, e.g. as influenced by elevation, slope and aspect. These topographic features have a marked effect on pasture productivity through influences on available soil moisture and nutrients, solar radiation, temperature and animal behaviour, which in turn affect pasture productivity (Gillingham *et al.* 1998; Lopez 2000; Blennerhassett 2002).

Geographic information systems (GIS) have been widely used in environmental modelling (e.g. Johnston 1990; Antoni 1998; Wadsworth & Reweek 1999; Dominy & Duncan 2001), and have been increasingly used in predictive modelling of vegetation properties such as species distribution and abundance (e.g. Franklin 1995; Iverson *et al.* 1997; Johnston 1998; Guisan & Zimmermann 2000; Stocks & Wise 2000) and community productivity (e.g. Iverson *et al.* 1997; Li *et al.* 1998; Tan & Shibasaki 2003), showing its strong ability in analysing spatial factors, and facilitating prediction over space. Linking a model with high predictive ability with a GIS provides the ability to predict productivity for pastures with heterogeneous distribution of environmental and management factors.

The commonest way of using GIS in environmental modelling is to prepare input data and then visualize the model prediction (Johnston 1998; Stocks & Wise 2000). GIS can also be used to develop models using its internal analysis functions such as map algebra, but this is usually restricted to some cartographic modelling, since most GISs have only limited ability to analyse statistical models. More complex models are usually developed in professional statistics software using input data generated or derived from GIS, and the model output is then imported back into a GIS to get the prediction (Stocks & Wise 2000). The power of GIS becomes apparent when it is coupled with models that predict outcomes of ecosystem processes such as succession, net primary production, and nutrient cycling with parameters derived from GIS (Johnston 1998).

For a predictive model to be coupled with a GIS, two aspects in particular need to be considered: one is its predictive accuracy; the other is the ease with which the model can be imported into the GIS. There are different approaches that have been used in predicting pasture productivity: model-based approaches e.g. empirical models (Sala *et al.* 1988) and mechanistic models (Moir *et al.* 2000), and data mining approaches e.g. artificial neural networks (Franklin 1995; Johnston 1998). The data mining approach performs better than the model-based approach with respect to predictive accuracy (Witten & Erank 2000; Dunham 2002; Scheffer 2002), and thus is a more appropriate method for predictive purposes. Among the data mining approaches, the decision tree approach has been widely used in the social (Scheffer 2002) and medical sciences (Petitti 2000), and also has shown an increasing application in environmental modelling where it has considerable accuracy and the ability to reveal the relative importance of environmental factors. The output of a decision tree is also very easy to import into a GIS (Iverson & Prasad 1998; Vayssieres *et al.* 2000; Scheffer 2002; Yang *et al.* 2003).

The decision tree approach is a non-parametric machine-learning modelling method, which recursively splits the multidimensional space defined by independent variables into zones that are as homogenous as possible in term of the response of the dependent variable (Vayssieres *et al.* 2000). The result of the analysis is a binary hierarchy structure called a decision tree with branches and leaves that contain the rules to predict the new cases (Breiman *et al.* 1984; Dunham 2002). Besides its high predictive accuracy, there are several other advantages that a decision tree has over the model-based approach. The decision tree has no strict assumptions regarding the distribution of the target (dependent) variable (Breiman *et al.* 1984) and it can easily incorporate nominal, ordinal and interval variables in the same model (Thuraisingham 1999; Vayssieres *et al.* 2000).

However, there appear to have been no published studies that use a decision tree to predict pasture productivity despite the approach's wide application elsewhere. In this study, decision tree models for annual and seasonal pasture productivity (aboveground biomass) were developed for hill-pasture in the North Island, New Zealand, and predictions were realised and verified in a GIS. This study aimed at developing a GIS-based predictive modelling approach to simulate hill-pasture productivity, and to assess it performance with respect to model predictive accuracy and the ability to reveal the key factors influencing pasture productivity. The success of this GIS-based predictive modelling approach could find applications in investigating ecosystem processes and improving pasture management.

4.2. Methods

4.2.1. Study area

This study was carried out on the AgResearch Ballantrae Research Station at latitude S40°18′, longitude E175°50′ in the North Island of New Zealand (Fig. 1) with average annual rainfall 1270 mm and average annual daily temperature 12.3 °C. Altitude ranges from 125 to 350m a.s.l..

The hill country in Ballantrae is dissected and very heterogeneous with complex combinations of slope and aspect even within a small area (Lambert & Roberts 1978). This heterogeneity increases the diversity of micro-climate, soil characteristics, species composition and the behaviour of grazing animals, and further increases the complexity of pasture productivity patterns (Lopez 2000). In southern latitudes, north aspects receive more net radiation, and are warmer and generally drier than south aspects. Available soil moisture generally diminishes with increasing slope (Radcliffe & Lefever 1981). By influencing animal grazing and excretal return patterns, aspect and slope also influence soil nutrient status (Gillingham 1982; Sheath & Boom 1985).The pasture at Ballantrae is mainly dominated by browntop *(Agrostis capillaris)* and ryegrass *(Lolium perenne),* and white clover *(Trifolium repens)* is the main pasture legume species (Nicholas 1999).

Data for developing the decision tree models were sampled within the whole area of the Station. The GIS-based model prediction and verification were carried out in a pasture which was within the Station and was about 90 ha in area (Fig. 1). Within this pasture there were four small farmlets (small farm fenced for the same management) with different fertiliser application treatments since the early 1970s: two of them had approximately 120kg and 360kg annual P fertiliser (Single SuperPhosphate, approximately 9% of P) input per year since 1974 (hereafter referred as LL and HH), another two farmlets also had approximately 120kg and 360 kg P fertiliser per year from 1974 to 1982, but had none since (hereafter referred as LN and HN). The farmlets

LL, LN, HN, and HH had 8, 10, 8 and 10 paddocks respectively. Each paddock was approximately eight ha in area. A detailed description of these four farmlets and their management history can be found in Lambert *et al.* (2000).

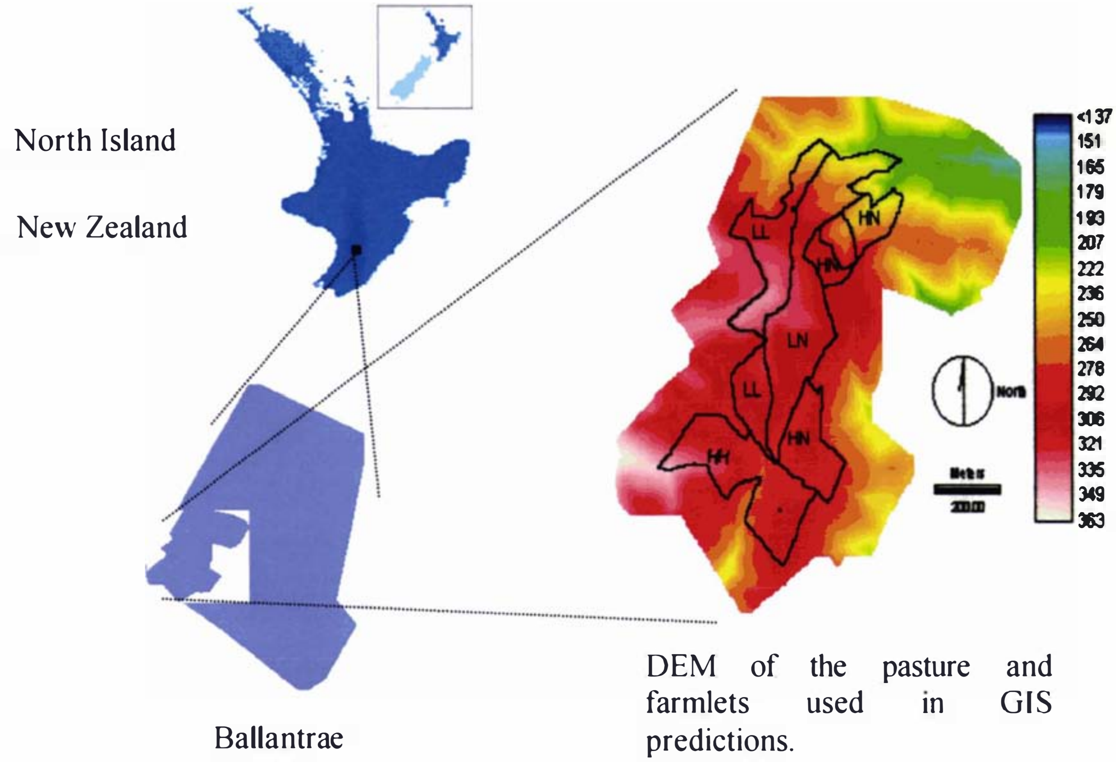


Fig. 1. Study area, digital elevation model (DEM) and the farmlets' used in validating the GIS prediction.

4.2.2. Model development

4.2.2.1. Dataset generation

Data for pasture productivity (annual and seasonal), soil properties (bulk density, pH, Olsen-P), fertiliser management (N, P fertiliser inputs) and topographic features (slope and aspect) were from a long-term experiment (see Lambert *et al.* 1986), unpublished Ph.D. theses and other experiments conducted at Ballantrae during 1972-1998. Most climatic data (rainfall, temperature and global solar radiation) were obtained from the National Institute of Water & Atmospheric Research (NIWA), New Zealand. Potential evapotranspiration (PET) was derived from temperature and solar radiation using Priestley & Taylor's equation (Priestley & Taylor 1972). Since the rainfall data was measured from the horizontal surface and the P fertiliser was commonly aerially applied to hill-pasture with a uniform rate, the actual rainfall and P fertiliser received were adjusted by the cosine of the slope angle. Actual solar radiation on hilly terrain is strongly influenced by topography (Antoni 1998), so it was calculated from observed solar radiation at that location by an empirical method developed for the North Island, New Zealand (McAneney & Noble 1975). Therefore, the variation in rainfall and solar radiation for a sample site included year-to-year variation and the adjustments based on topographic features (slope and aspect).

There were 30 variables, including 25 input variables (independent variables) and five target (dependent) variables, and 1100 samples in the dataset (Table 1). These samples were collected using a stratified random method within the study area by a 1 m^2 quadrat. The climate data and P fertilizer data were calculated instead of observed for each sample.

4.2.2.2. Decision tree development and assessment

The decision tree models for annual, spring, summer, autumn and winter pasture productivity were analysed in SAS Enterprise Miner, Version 4.1 (SAS Institute Inc., 1999-2001, Cary, NC, USA). The main procedure for developing a decision tree model is to split the data of target variable based on it response to input variables. There are different criteria used to split the data, depending on the nature of the target variable. For interval variables (i.e. continuous variables) like pasture productivity, the split is based on variance reduction or F tests. I compared both criteria and chose variance reduction as the split criterion. However, the decision tree outputs from using these two criteria were very similar.

Model assessment is an important step in developing the decision tree, as it enables a good model output, and prevents overtraining which otherwise would develop a model that only fits the data from which it is trained (Witten & Erank 2000). SAS Enterpriser Miner has several options to assess models having interval targets. The measure of average squared error (ASE) is commonly used. ASE is similar to mean squared error (MSE) in general linear models so the smaller the ASE the better the model fit to the data. If there are validation data available, it is assessed by the ASE of validation data; otherwise the ASE of the training data is used. In the development of the decision trees here, 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; and the "best" tree which had the smallest ASE from validation data was selected.

Table 1. Variables used in the decision tree and GIS analyses

Model role	Variable symbol	Units	Variable description
input	pH	$-\log_{10}[H^+]$	soil pH
	BD	g/cm^3	soil bulk density
	OlsenP	µg/g	soil Olsen P
	N_fert	kg/ha/y	annual elemental N fertiliser input
	P_fert	kg/ha/y	annual elemental P fertiliser input
	P_fert5	kg/ha	5-year elementnal accumulative P fertiliser input
	P_fert10	kg/ha	10-year elemental accumulative P fertiliser input
	temp_y	^oC	annual mean daily temperature
	temp_sp	^oC	spring mean daily temperature
	temp_su	^oC	summer mean daily temperature
	temp_au	^oC	autumn mean daily temperature
	temp_wi	^oC	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	annual mean daily global solar radiation
	solar_sp	MJ/m^2	spring mean daily global solar radiation
	solar_su	MJ/m^2	summer mean daily global solar radiation
	solar_au	MJ/m^2	autumn mean daily global solar radiation
	solar_wi	MJ/m^2	winter mean daily global solar radiation
	aspect		hill slope aspect. eg. N NE S SW NW, etc.
	slope	degree	hill slope angle
target	annual productivity	kg/ha/y	annual aboveground dry matter per hectare
	spring productivity	kg/ha/season	spring aboveground dry matter per hectare
	summer productivity	kg/ha/season	summer aboveground dry matter per hectare
	autumn productivity	kg/ha/season	autumn aboveground dry matter per hectare
	winter productivity	kg/ha/season	winter aboveground dry matter per hectare

I set six as the maximum tree levels, and 11 as the minimum observations required for a split search. The recommendation on minimum observations was derived using the program by assessing the size of the whole dataset (SAS Online Help: Getting Started with Enterpriser Miner Software).

4.2.3. GIS-based prediction and model validation

The outputs of the decision tree models for annual and seasonal pasture productivity were imported into a GIS to derive "predictions" of annual and seasonal productivity for the pasture described previously. A 5 × 5 m resolution digital elevation model (DEM) of this pasture was interpolated from a XYZ coordinates file produced by New Zealand Aerial Mapping Ltd. A map layer of the four farmlets was created from the DEM based on an existing fence map. All the GIS map layers of input variables were then generated. Map layers of rainfall and P fertiliser input were adjusted by a cosine of the slope angle, and layers of solar radiation were adjusted by slope and aspect with methods previously mentioned. All the map layers had the same resolution (5 × 5 m).

In 1994, the annual and seasonal pasture productivity was measured on these four farmlets with a "trim technique" involving harvesting pasture regrowth from trimmed plots (1 m^2) protected from sheep grazing (Lambert *et al.* 1996). For each of spring, summer, autumn and winter productivity, and also for annual totals, 24, 30, 24 and 30 observations were obtained for LL, LN, HN and HH, respectively. These observations were taken from systematically designed plots which included major aspect and slope categories within a farmlet. This data was independent from the data used to develop the decision tree models.

GIS-based predictions were made based on the climatic and management variables in 1994. The observed productivity was used to verify the predictions. Spatial predictions for annual and seasonal pasture productivity were implemented by coupling the outputs of the decision trees with GIS using a map overlay function. The prediction outputs were the GIS map layers of annual, spring, summer, autumn and winter pasture productivity. Fig. 2 shows the procedures for generating the model predictions.

The average annual and seasonal productivity estimates for the four farmlets were then extracted from the map layers of predicted productivity by calculating the average productivity of all the pixels (5 m×5 m) within each farmlet (there are 2737, 3689, 2366 and 2769 pixels for the farmlets LL, LN, HN and HH, respectively). The predicted productivity was then compared with the observed productivity for an empirical validation of the model (Mitchell 1997). The 95% confidence interval of the observed productivity mean was set as an acceptable error of the prediction. Since the map layers of predicted productivity covered an area which was larger than the four farmlets, only the area within the four farmlets was used to validate the predictions. This model validation method is acceptable for its purpose. However, a better method is to validate individual prediction in a pixel (contrast to using the average within a farmlet) by comparing it with observation at the corresponding site, but this would require a large number of measurements.

The GIS analyses were implemented in Idrisi 3.2 (Clark Labs, Clark University, Worcester, MA, USA).

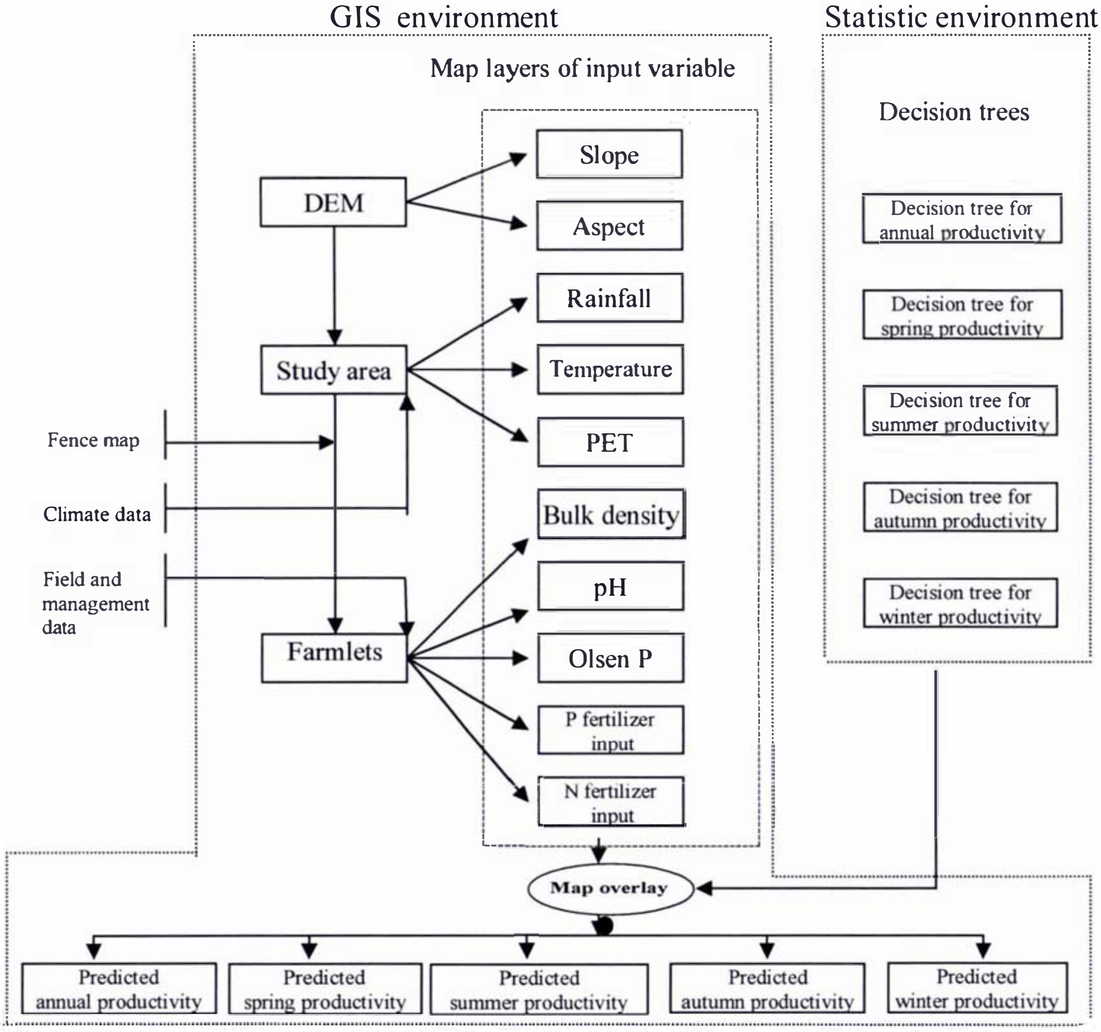


Fig. 2. Diagram of the procedures for generating the model predictions in GIS. The rectangles in the GIS environment area represent the GIS map layers. The arrows indicate the orders and the resources in generating predictions on pasture productivity. Names of input variables and their description are in Table 1.

4.3. Results

4.3.1. Decision trees

Decision tree model for annual, spring, summer, autumn and winter pasture productivity are displayed in Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7. Each decision tree is a hierarchical structure that contains rules of prediction. Pasture productivity was first split into two branches by a variable which best explained the variance. This splitting continued for each of the branches with the same method, and so on until no further splitting could be made. Prediction was made using a series of constraints defined by the input variables and their split-points. For example, in the model for annual pasture productivity (Fig. 3), when pasture with a slope equal to, or more than 22°, and a five-year cumulative P fertiliser (elemental P) input equal to, or more than 101.8 kg and a warm season rainfall equal to, or more than 540 mm, the predicted annual pasture productivity was 10347 kg/ha.

The hierarchical structures of the decision trees can also indicate the relative importance of environmental and management variables in influencing productivity and the interaction among these input variables. The variable first selected in splitting the tree is more influential than those selected after it. In the decision tree model for annual pasture productivity slope was the first variable selected to split the tree, and was the most important factor influencing annual pasture productivity. Olsen P (for slopes less than 22°) and five-year cumulative P fertiliser input (for slopes equal to or more than 22°) were the second most important factors (Fig. 3). The most important variables influencing spring pasture productivity were also slope, Olsen P and five-year cumulative P fertiliser input (the same as the annual pasture productivity) (Fig. 4). For summer pasture productivity, annual P fertiliser input, summer rainfall (for low annual P fertiliser input) and slope (for high annual P fertiliser input) were the most important variables (Fig. 5). Autumn rainfall, ten-year cumulative P fertiliser input (for low autumn rainfall) and five-year cumulative P fertiliser input (for high autumn rainfall) were the most important variables influencing autumn pasture productivity (Fig. 6). Olsen P, five-year cumulative P fertiliser input (for low Olsen P) and winter solar radiation (for high Olsen P) were the most important variables influencing winter pasture productivity (Fig. 7).

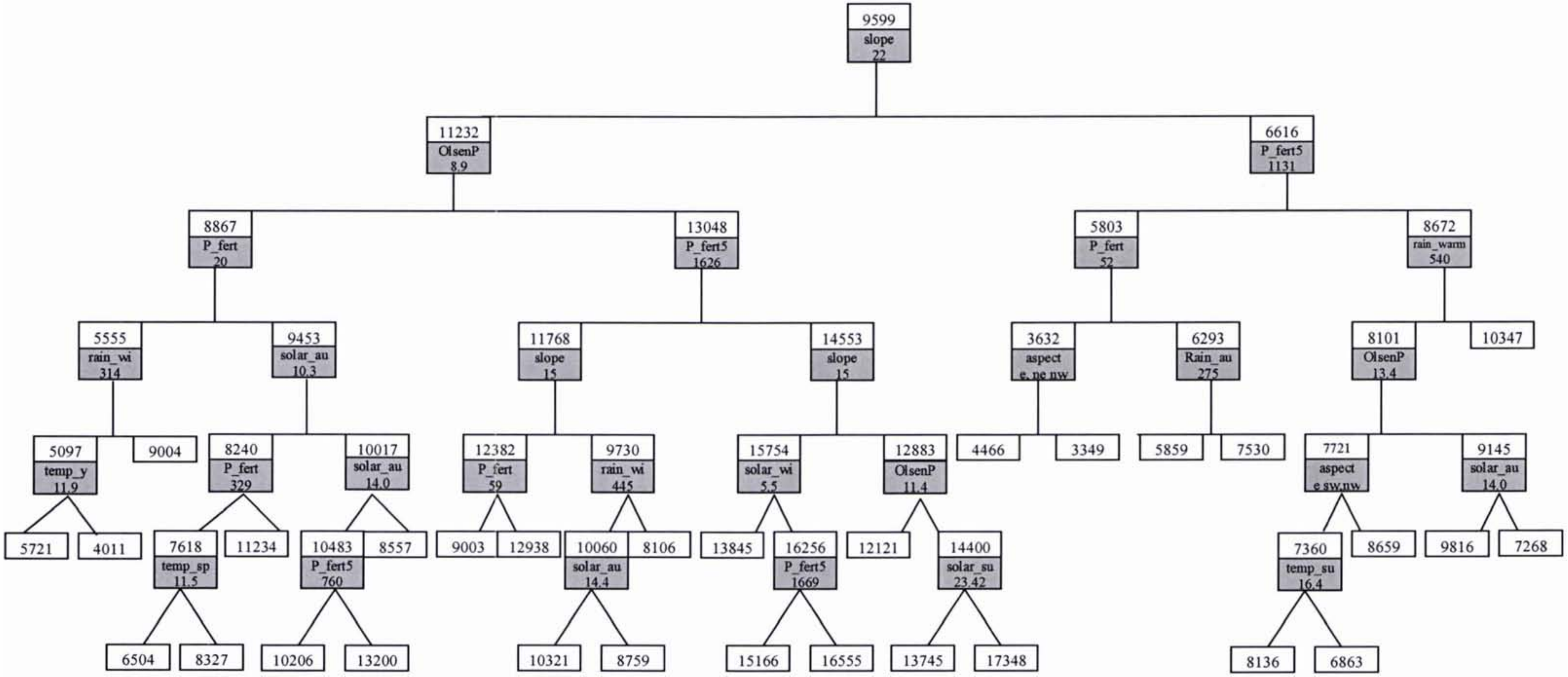


Fig. 3. The decision tree model for annual pasture productivity. Predicted productivity are in the un-shaded rectangles, splitting variables and split-points are in the shaded rectangles. Prediction goes to the left-side branch when the splitting variable less is than the split-point, and goes to the right-side branch when the splitting variable is equal to, or more than, the split-point (in the case of aspect, pasture with aspect set before the comma go to left branch, others go to right branch). See Table 1 for variable symbols and unit descriptions.

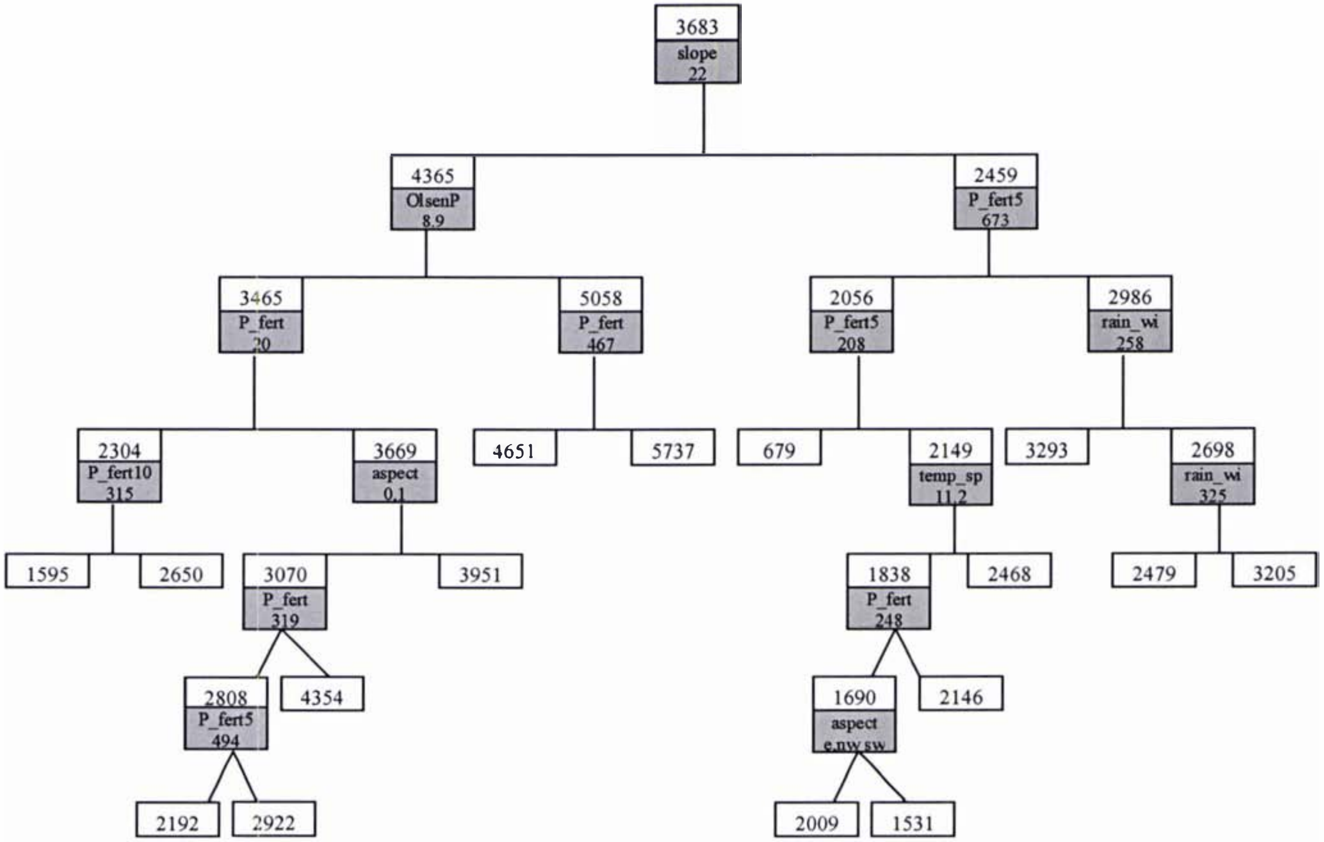


Fig. 4. The decision tree model for spring pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

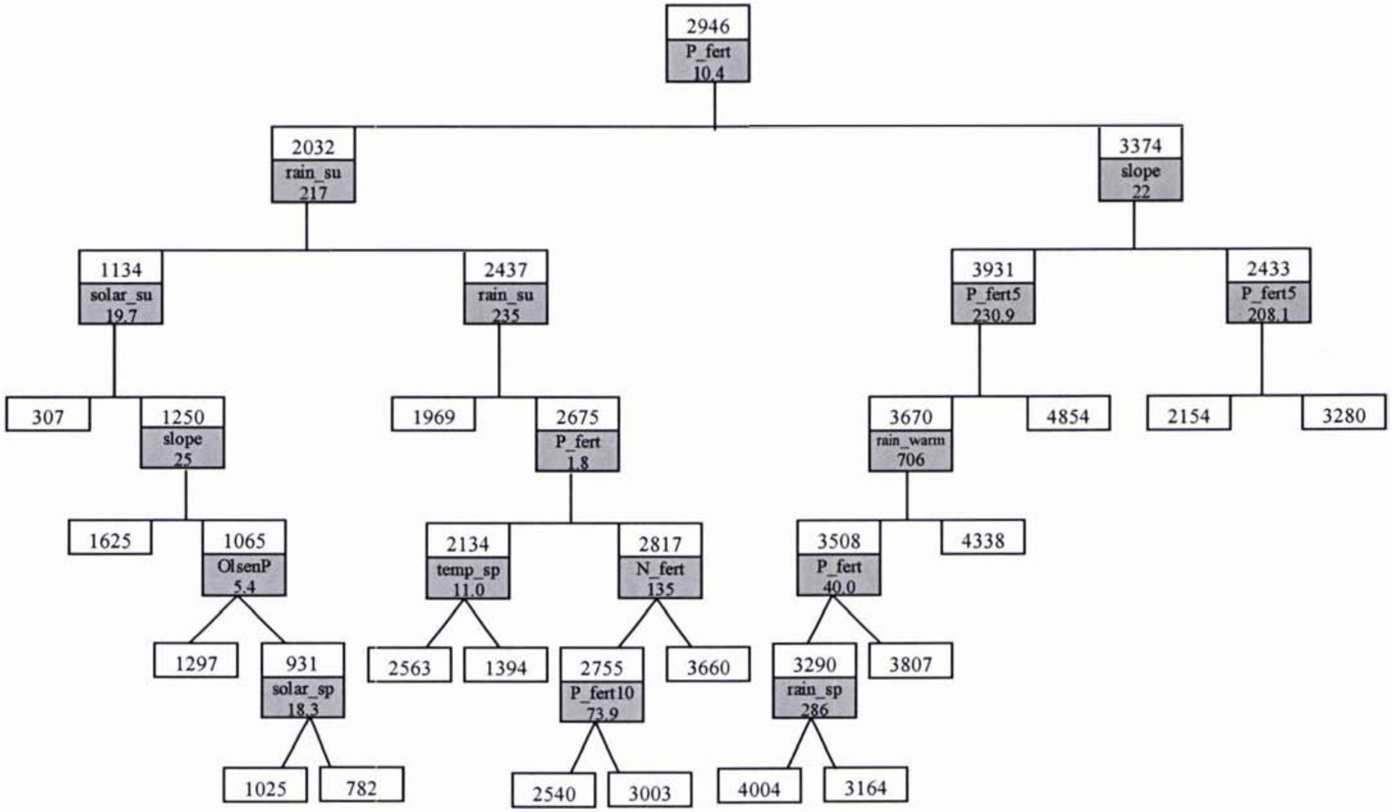


Fig. 5. The decision tree model for summer pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

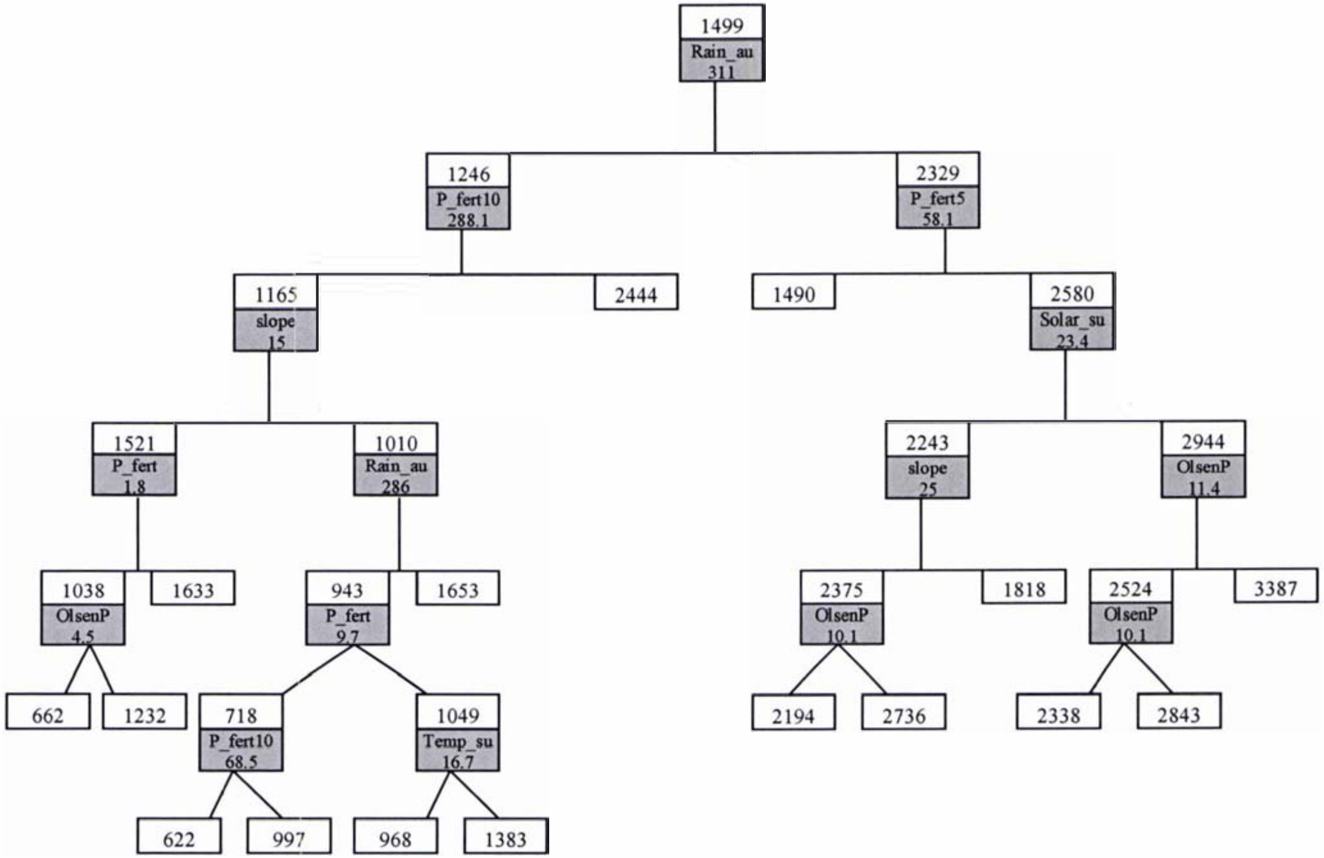


Fig. 6. The decision tree model for autumn pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.

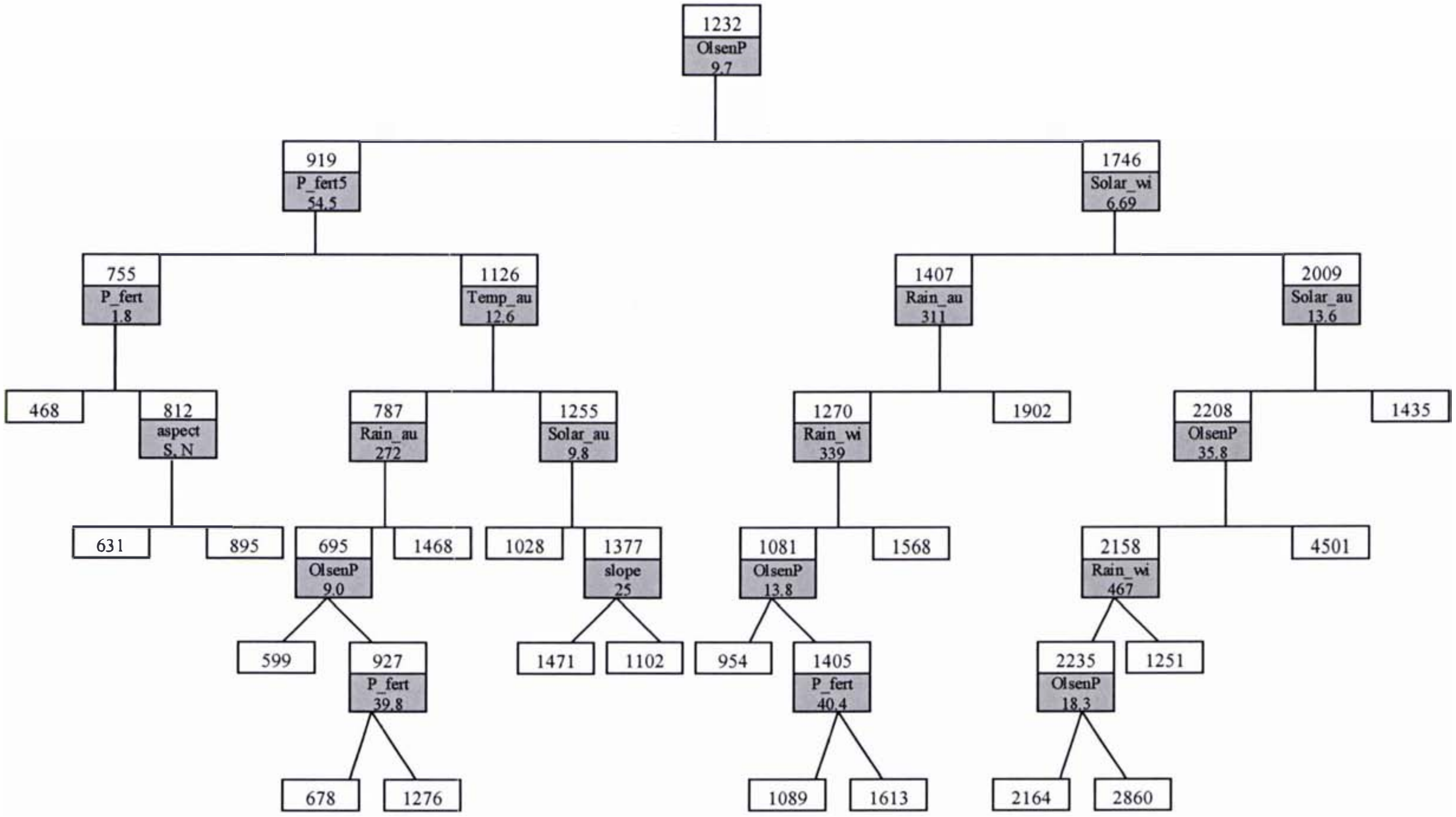
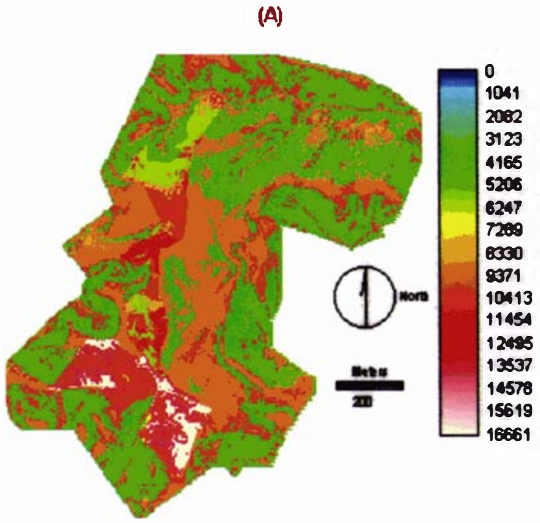
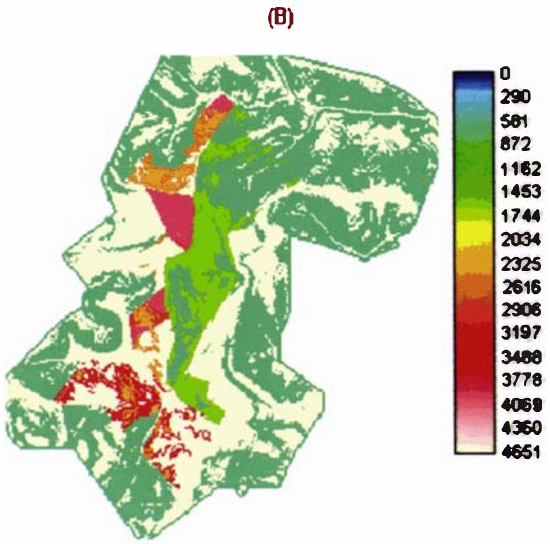
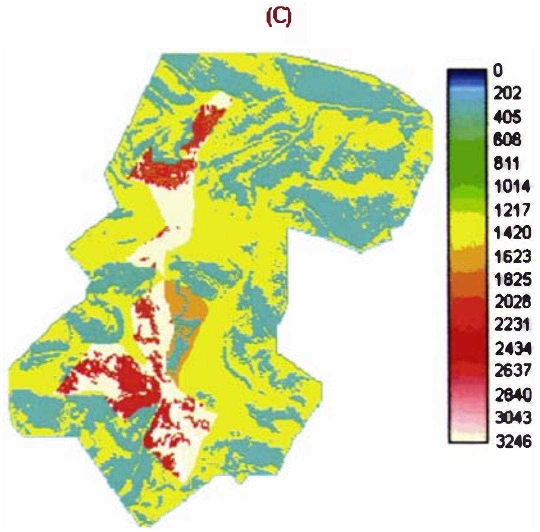
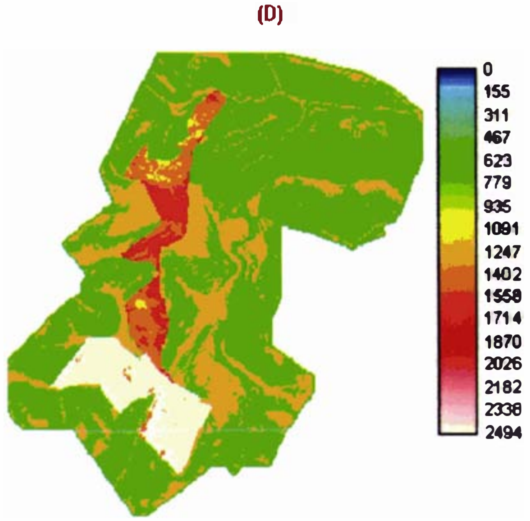


Fig. 7. The decision tree model for winter pasture productivity. See caption of Fig. 3 for the interpretation of decision tree.









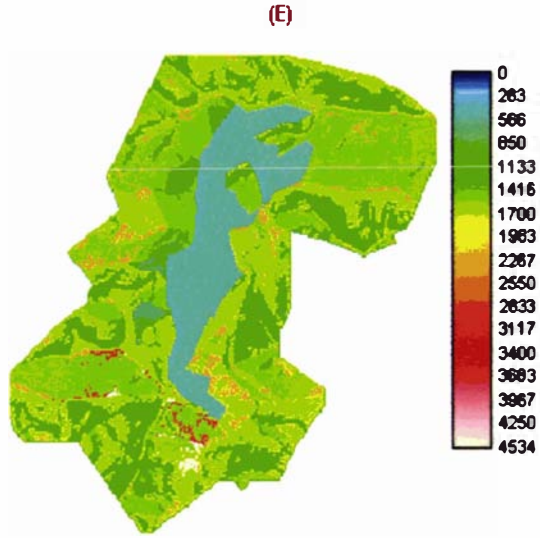


Fig. 8. Map layers of the predicted annual (A), spring (B), summer (C), autumn (D) and winter (E) pasture productivity (kg/ha). Note the different scales for productivity in the map keys.

The interactions of environmental and management variables on pasture productivity can be easily found in the decision trees. For example, there was a interaction between fertility and moisture in the decision tree model for annual pasture productivity (Fig. 3): when pasture received a five-year cumulative P fertiliser input equal to, or more than 101.8 kg, pasture with a warm season rainfall equal to, or more than 540 mm had a mean annual productivity of 10347 kg/ha, while that with a warm season rainfall less than 540 mm only had a mean annual productivity of 8101 kg/ha.

4.3.2. GIS-based prediction and model validation

GIS-based predictions for annual and seasonal pasture productivity are shown in Fig. 8. The predicted average annual and seasonal pasture productivity of the four farmlets LL, LN, HN and HH were graphed against the observed average annual and seasonal productivity in Fig. 9. The predicted average annual and seasonal pasture productivity of the four farmlets LL, LN, HN and HH were derived from 2737, 3689, 2366 and 2769 pixels, respectively.

The prediction for average annual pasture productivity was validated in the farmlets LL, LN and HH. Though the prediction in the farmlet HN was not verified, the difference between the prediction and observation was still within 25% of the observation mean. The predictions for spring and autumn pasture productivity were verified in LL. No prediction for summer productivity was verified among the four farmlets. The prediction for winter productivity was verified in LL and HH.

4.4. Discussion

The results show that the decision tree model for annual pasture productivity performed very well with three farmlets being validated. The difference between prediction and observation for another farmlet was still within 25% of the observation mean (Fig. 9). This indicates that the decision tree is a reliable and accurate approach for predicting annual hill-pasture productivity. An application of the decision tree in predicting plant species distribution (Vayssieres et al. 2000) showed that the decision tree performed significantly better than a polynomial logistic regression model for four of the six cases considered, and as well as in the two remaining cases. Applications of the decision tree in classifying remote sensed vegetation data (Yang et al. 2003) and in predicting tree species abundance (Iverson and Prasad 1998) also indicated that the decision tree had very good performance as a modelling approach.

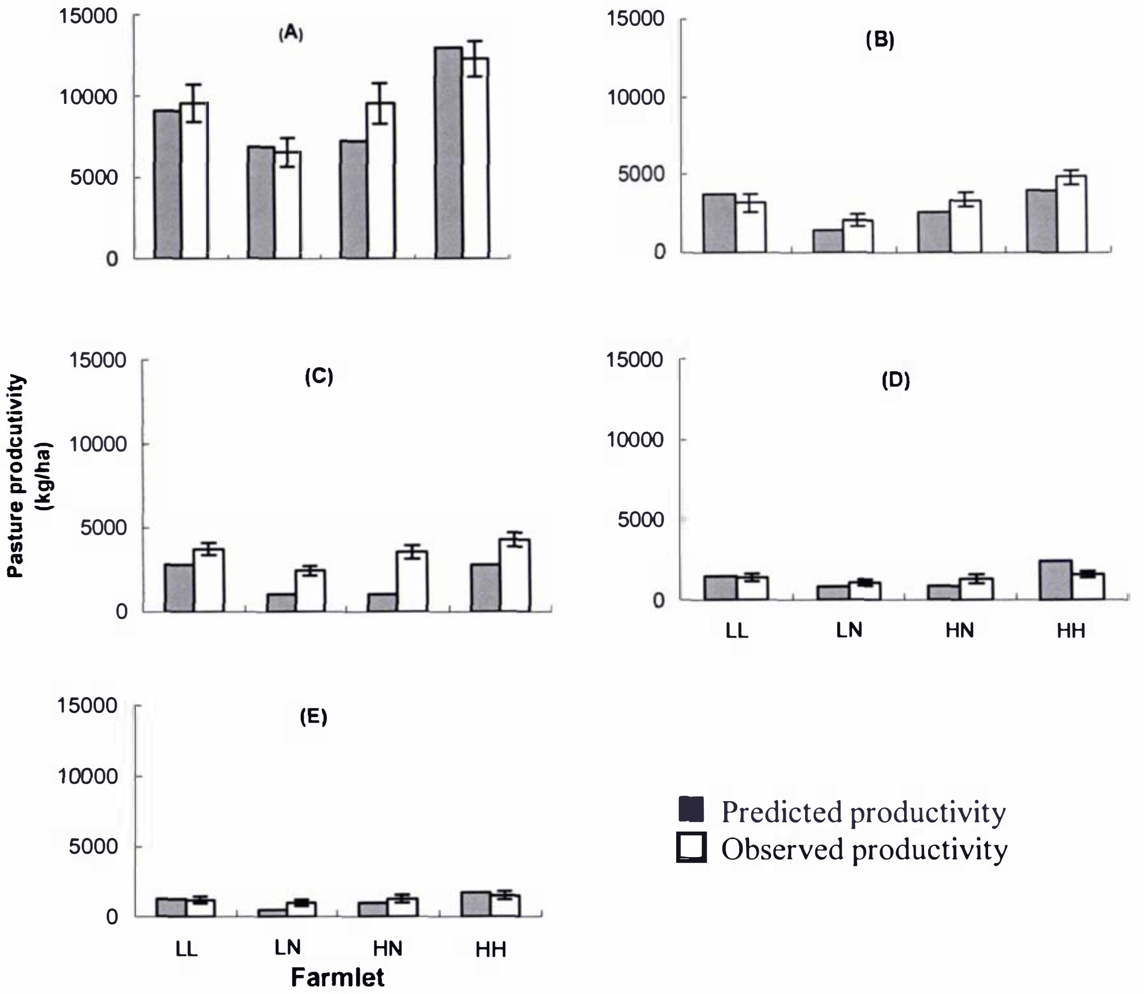


Fig. 9. Comparison of the predicted and the observed annual (A), spring (B), summer (C), autumn (D) and winter (E) pasture productivity of the farmlets LL, LN, HN and HH. Bars indicate the 95% confidence interval of the observed productivity means.

However, the empirical validations of the decision tree models for seasonal pasture productivity were not as good as that for annual pasture productivity. This may be because the quality of seasonal pasture productivity data used in developing the decision tree models for seasonal pasture productivity were not as good as annual pasture productivity data, as they were calculated from the pasture cuttings covering a season (Lambert *et al.* 1996) instead of measuring from a designed experiment, which may allocate the production that occurred in one season to another season, and introduce errors into the seasonal pasture productivity data. Also, carryover effects from season to season (e.g. due to drought) are likely to be more influential compared to carryover effects from year to year. This suggests the quality of a decision tree will ultimately depend on the quality of the training data. Further improvement in the predictive accuracy of the decision tree models for seasonal pasture productivity can be made by gathering more seasonal pasture productivity data in trials designed specifically for the purpose.

Hill slope, due to its important role in influencing available moisture, fertility and solar radiation, greatly influences hill-pasture productivity. The impact of hill slope on pasture productivity is well recognised (White 1990; Gillingham et al. 1998), and was fully indicated in the decision tree models for annual and seasonal pasture productivity. To realise its effect over space GIS, therefore, became an essential part in this modelling process. The successful integration of decision tree with GIS provided a platform to predict pasture productivity for pastures with heterogeneous topographic and management features, and to present model prediction over space for further application and investigation.

A DEM is usually the most important element in a GIS-based modelling process. The quality (resolution and accuracy) of a DEM can greatly influence the model prediction as many environmental factors are derived from it (Iverson *et al.* 1997; Wadsworth & Reweek 1999; Stocks & Wise 2000). A high resolution DEM was used in this study to generate slope and aspect, and the adjustments in rainfall, P fertiliser input and solar radiation were then made based on them. This enabled the model predictions to reflect the small changes of topographic features that influence pasture productivity in hill-pasture with heterogeneous micro-topography.

The advantage of this modelling approach is that it can be used as, or incorporated in, a decision support system to improve pasture management. For example, it can be applied to predict pasture productivity for an area of interest or to assess the effect of alternative fertiliser application strategies on pasture productivity. It also can be used to investigate relationships between pasture productivity and environmental factors. For example, to assess the seasonal climate variability such as warm spring or drought summer on subsequent pasture productivity.

One limitation of this modelling approach is that some soil properties such as soil Olsen P, pH and bulk density were used as input variables to simulate pasture productivity. These soil features may increase the predictive ability of the models, but in practice, they are usually very hard to obtain on a large scale. To develop a decision tree that only incorporates climatic, topographic and management variables would be a practical choice, though this may reduce the predictive ability of the model to some degree.

4.5. Conclusion

The predictive modelling approach described here incorporated two elements: decision tree and GIS. The decision tree approach was found to perform well in predicting annual pasture productivity. It had good predictive accuracy and revealed the relative importance of environmental and management variables and their interaction in influencing hill-pasture productivity in the studied area. Using GIS in this predictive modelling approach facilitated the derivation of topographical variables and visualised model outputs for interpretation and analysis. This GIS-based predictive modelling approach can be used to investigate the interrelationship between pasture productivity and environmental factors and to improve pasture management.

4.6. References

Antoni O. (1998) Modelling daily topographic solar radiation without site-specific hourly radiation data. *Ecological Modelling*, 113, 31-40

Blennerhassett J.D. (2002) *Pasture growth constraints on dry steep east coast hill country.* Ph.D. thesis, Massey University, New Zealand

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) Classification and regression trees. *The Wadsworth statistics/probability series.* Chapman & Hall. Inc. New York, NY.

Dominy N.J. & Duncan B. (2001) GPS and GIS methods in an African rain forest: applications to tropical ecology and conservation. *Conservation Ecology* [electronic resource] http://www.consecol.org/vol5/iss2/art6

Dunham M.H. (2002) *Data Mining: Introduction and Advanced Topics.* Pearson Education INC. Upper Saddle River, New Jersey 07458

Dyne G.M.v. (1970) A systems approach to grasslands. *Proceedings of the XI International Grassland Congress, Surfers Paradise, Australia*, A131-A142

Franklin J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499

Gillingham A.G. (1982) Topographic and management effects on dung distribution by grazing sheep. *Proceedings of the New Zealand Grassland Association*, 43, 161-170

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., Martin E.D., Scott C.T. & Prasad A. (1997) A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio (U.S.A.). *Landscape Ecology*, 12, 331-348

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

Johnston C. (1998) *Geographic information systems in ecology*. Oxford, Malden, MA: Blackwell Science

Johnston L.B. (1990) Analyzing spatial and temporal phenomena using geographical information systems. *Landscape Ecology*, 4, 31-43

Lambert M.G., Barker D.J., Mackay A.D. & Springett J.O. (1996) Biophysical indicators of sustainability of North Island hill pastures. *Proceedings of the New Zealand Grassland Association*, 57, 31-36

Lambert M.G., Clark D.A., Grant D.A. & Costall D.A. (1986) Influence of fertilizer and grazing management on North Island moist hill country 2. Pasture botanical composition. *New Zealand Journal of Agricultural Research*, 29, 1-10

Lambert M.G., Clark D.A., Mackay A.D. & Costall D.A. (2000) Effects of fertiliser application on nutrient status and organic matter content of hill soils. *New Zealand Journal of Agricultural Research*, 43, 127-138

Lambert M.G. & Roberts E. (1978) Aspect differences in an unimproved hill country pasture 2. Edaphic and biotic difference. *New Zealand Journal of Agricultural Research*, 21, 255-260

Li J., Liang T. & Chen Q. (1998) Estimating grassland yields using remote sensing and GIS technologies in China. *New Zealand Journal of Agricultural Research*, 41, 31-38

Lopez I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

McAneney K.J. & Noble P.E. (1975) Estimating solar radiation on sloping surfaces. *New Zealand Journal of Experimental Agriculture*, 4, 195-202

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

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Parton W.J., Scurlock J.M.O., Ojima D.S., Gilmanov T.G., Scholes R.J., Schimel D.S., Kirchner T., Menaut J.-C., Seastedt T., Garcia Moya E., Kamnalrut A. & Kinyamario J.L. (1993) Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles*, 7, 785-809

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

Priestley C.H.B. & Taylor R.J. (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100, 81-92

Radcliffe J.E. & Lefever K.R. (1981) Aspect influences on pasture microclimate at Coopers Creek, North Canterbury. *New Zealand Journal of Agricultural Research*, 24, 55-56

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

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

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

Seligman N.G. & Baker M.G. (1993) Modelling as a tool for grassland science progress. In: Baker, M.G. (Ed.) *Grassland for Our World*, Sir Publishing, Wellington, New Zealand, 228-233

Sheath G.W. & Boom R.C. (1985) Effect of November-April grazing pressure on hill country pastures 3. Interrelationship with soil and pasture variation. *New Zealand Journal of Agricultural Research*, 13, 341-349

Sorenson J.T. (1998) Modelling and simulation in applied livestock production science. In: Peart, R.M. and Carry, R.B. (Eds.) *Agricultural Systems: Modelling and Simulation*, Marcel Dekker Inc. New York, 475-494

Stocks C.E. & Wise S. (2000) The role of GIS in environmental modelling. *Geographical and Environmental modelling*, 4, 219-235

Stuth J.W., Hamilton W.T., Conner J.C., Sheehy D.P. & Baker M.J. (1993) Decision support systems in the transfer of grassland technology. In: Baker, M.G. (Ed.) *Grassland for Our World*, Sir Publishing, Wellington, New Zealand, 234-242

Tan G. & Shibasaki R. (2003) Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. *Ecological Modelling*, 168, 357-370

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

Wadsworth R. & Treweek J. (1999) *GIS for ecology: an introduction.* Harlow: Addison Wesley Longman

White J.G.H. (1990) Hill and high country pasture. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

Witten I. & Erank E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers San Francisco, California

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

Chapter 5. Modelling the Impact of Climate Changes and Alternative Phosphorus Fertiliser Management on Pasture Production in the North Island, New Zealand

Studies in previous chapters indicated that the decision tree model for pasture productivity performed better than the regression model in terms of predictive ability. However, one disadvantage of the decision tree model is that it is not able to reflect the influence of small changes in input variables on the model target. In this study a polynomial regression model was developed and was integrated with a geographical information system (GIS) to evaluate the impacts of climate change and alternative phosphorus (P) fertiliser management on pasture production. Insights obtained from the previous decision tree models provided useful help in developing this regression model.

---

*A paper derived from this chapter with a title "Modelling the impact of climate changes on pasture production in the North Island, New Zealand" by B. Zhang, I. Valentine & P.D. Kemp has been submitted to **Climatic Change** for publication.*

Abstract. To assess the potential impact of climate changes on pasture production in the North Island, New Zealand, eight climate scenarios of increased temperature and increased (or decreased) rainfall were investigated by integrating a polynomial regression model with a geographic information system (GIS). The impact of alternative phosphorus fertiliser application on pasture production was also investigated. The results indicated that the climate change scenarios assuming an increase in temperature by 1-2 oC and a rainfall change by -20% to +20% would have a very significant impact on pasture production with a predicted pasture production variation from -46.2% to +51.9% compared with the normal climate from 1961-1990. However, the variation was in a range between -20.4% to +9.6% when averaged over the climate zones. Increased temperature would generally have a positive effect on pasture production in the South and Southeast of the North Island, and increased rainfall would have a positive effect in the Central, South and Southeast of the North Island and a negative effect in the north of the North Island. The interaction of decreased rainfall and increased temperature would have a negative impact for the whole North Island except some central areas with high rainfall. The result also indicated that pasture with low rainfall had a higher response to increased P fertiliser input than pastures with high rainfall, which likely resulted from the high runoff and associated P loss in the high rainfall area. Relevant management practices for coping with potential climate change and improved strategies for fertiliser use were discussed. One limitation of this study was that the effect of elevated CO_2 concentration on pasture production was not assessed.

5.1. Introduction

Climate is a key driver of pasture ecosystem processes and fundamentally controls the amount of biomass available and its distribution among seasons (Sala *et al.* 1988; Barrett *et al.* 2002; Bai *et al.* 2004). In New Zealand, rainfall is the most important climate factor constraining pasture growth (Chapter 3) with spring and summer rainfall explaining about 60% of variation in pasture production (Radcliffe & Barrs 1987). Temperature is also a key factor influencing pasture growth especially during winter and early spring (Baars & Waller 1979; White 1990).

The world is getting warmer and global temperatures today are about 0.6 ^{o}C higher than they were in the early 1900s due to the greenhouse gases emissions (Ministry for the Environment 2001). This warming, from 1941, has resulted in more days with greater than 30 °C and fewer days with less than 0 °C per year in New Zealand (Plummer *et al.* 1999). Climate models have predicted that global average temperatures in 2100 will be between 1.0 to 5.4 °C higher than those in 1900 depending on the scenarios assumed (e.g. Kattenberg *et al.* 1996; IPCC 2000; Houghton *et al.* 2001), and global average precipitation will increase by 1 to 9% by 2100 and more intensive precipitation events are likely over many areas (Houghton *et al.* 2001). New Zealand is likely to warm by about two-thirds of the global mean temperature change, and is expected to increase faster in the North Island than in the South Island, with rainfall likely to increase in the west of the country and decrease in the east (Ministry for the Environment 2001). Temperature is expected to increase in the range 0.6 to 2.8 °C and the changes in rainfall are expected in the range of -20% to +30% (Ministry for the Environment 2001).

This climate change and variability has a pronounced impact on ecosystems (McCarthy *et al.* 2001), and may result in changes in vegetation structure and composition as well as productivity (Armstrong 1996; Riedo *et al.* 1997; Iverson & Prasad 1998; Joos *et al.* 2001). As a highly climate-driven industry, pastoral agriculture is especially vulnerable to this climate change and variability (Thornley & Cannell 1997; Kenny *et al.* 2000; Halloy & Mark 2003).

Few studies that have modelled the impacts of climate changes on pasture production in New Zealand. These models have, on average, predicted an increase in pasture production. For example, Baars *et al.* (1990) predicted that annual yield will be unchanged in the East Coast of the North Island and will increase by 8% in Waikato, and 20% in the South Island. For lowland pastures, the climate changes may increase annual pasture production per hectare by 10% to 30% (Martin *et al.* 1991). Similarly, Campbell (1996) estimated an average increase in pasture production of 10% to 15% by 2050-2100, but the effect may be negative in some regions and higher than this in some others. Although the projected climate change may generally have a positive effect on pasture production, it may also bring great uncertainties such as those that result from drought and heavy rainfall.

The magnitude of changes in temperature and rainfall likely as the result of climate change and variability are uncertain (Allen *et al.* 2001). Previous studies of climate change impact on pastures in New Zealand (Baars & Waller 1979; Martin *et al.* 1991; Campbell 1996) assumed only one or two climate change scenarios, and model predictions were for point locations. This is a limitation in assessing the climate change impacts over space for an area of interest under alternative climate change scenarios. With advances in the use of geographic information systems (GISs), a GIS-based model can simultaneously assess multiple scenarios for large areas over space. In this study, I developed a polynomial regression model and linked this model with a GIS to predict and assess the impacts of climate changes on pasture production in the North Island under multiple climate change scenarios. This assessment focused on the impact of changes i temperature and rainfall. The effect of increased carbon dioxide concentration was not evaluated. The model was developed from data collected mainly from hill-pasture, but it has been applied across all pastoral land in the North Island.

The productivity of New Zealand's pastures relies heavily on the regular input of phosphorus (P) through phosphate-based fertilisers. The amount of P fertiliser input is strongly influenced by the economic benefit obtained by farmers and Government policies (Gillingham *et al.* 1990). Scenarios were designed and modelled to evaluate the effect of different P fertiliser regimes on pasture production.

5.2. Methods

5.2.1. Study area

This study covered the pasture in the North Island of New Zealand, which accounts for approximately 70% of total land. Forest and other land covers account for the remaining area (Hunter & Blaschke 1986) (Fig. 1). Climate is very diverse in the North Island with mean annual rainfall ranging from about 800 to 1600 mm and mean annual daily temperature ranging from about 9 to 15 °C (Tomlinson & Sansom 1994). Based on the temperature and rainfall normal (long-term average) data for 1961 to 1990, I classified the North Island into five climate zones using the GIS climate surfaces interpolated from these temperature and rainfall normal data. They are high rainfall and medium temperature zone (HR-MT), high rainfall and high temperature zone (HR-HT), medium rainfall and high temperature zone (MR-HT), medium rainfall and medium temperature zone (MR-MT), low rainfall and medium temperature zone (MR-HT) (Table 1). Fig. 1 outlines the regions for these climate zones.

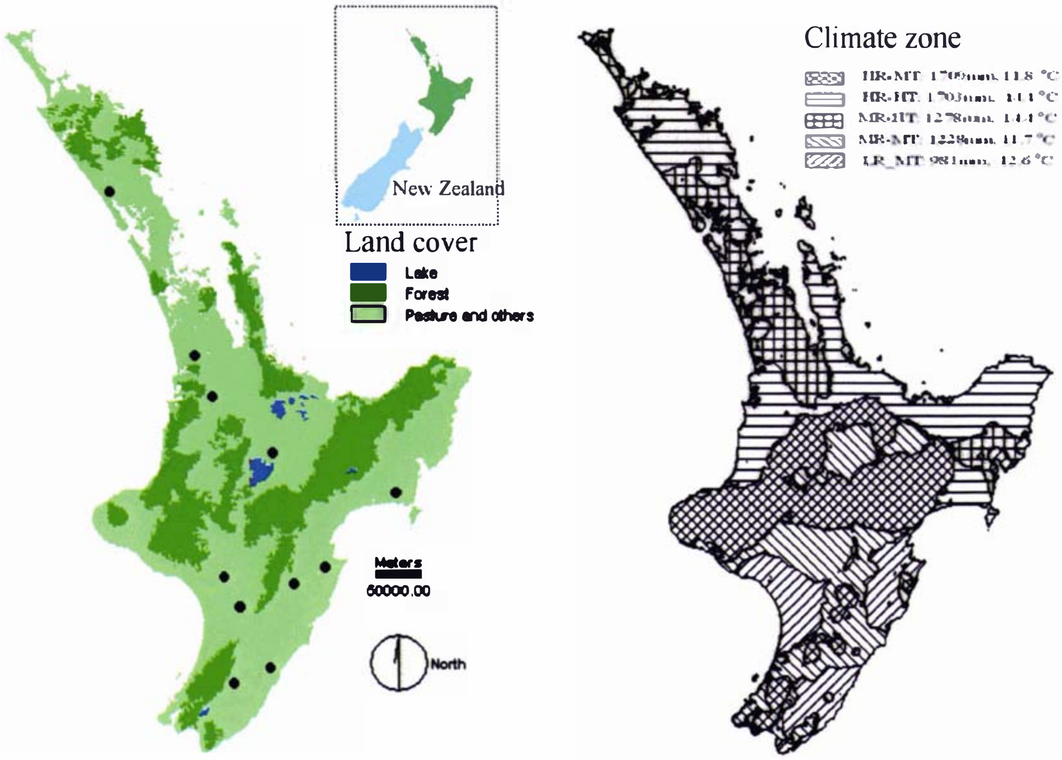


Fig 1. Study area, land cover and climate zone. See table 1 for the climate zone explanations.

5.2.2. Dataset

In last few decades, considerable research has been conducted on pasture in the North Island, New Zealand; most of it related to pasture production (Lambert *et al.* 1983; Gillingham *et al.* 1998; Dodd & Ledgard 1999; Moir *et al.* 2000). There is a large amount of data in the literature and kept by researchers in the form of raw or unpublished data. These data provide a very useful resource to develop models for assessing influence of climate factors on pasture production.

I focused on annual and seasonal rainfall, annual and seasonal daily mean temperature, annual P fertiliser (superphosphate) input and application history, nitrogen (N) fertiliser (urea) inputs, hill slope and soil Olsen P as key input (independent) variables in model analysis. These variables were identified in a previous decision tree model (Chapter 3) as the most significant factors influencing pasture production in the North Island. Data for annual pasture production (aboveground dry matter), fertiliser input, hill slope and soil Olsen P were from the literature and collected raw/unpublished data. Climate data were from the National Institute of Water & Atmospheric Research (NIWA). Since the rainfall data were measured from the horizontal surface and the P fertiliser was commonly aerially applied to hill-pasture with a uniform rate, the actual rainfall and P fertiliser received were adjusted by the cosine of the slope angle. Most pasture production data were measured by cutting re-growth from a trimmed quadrat (mainly 1 m2) with a cutting period of about one to two months depending on the growth rate of plants (Lambert et al. 1983). There were about 1900 samples in this dataset and those were from Whatawhata, Ballantrae, Te Kuiti, Riverside, Summerlee, Waipawa, Mauriceville, Mikimiki, Gladstone and Whareama. Most of these samples were from hill-pasture.

Table 1. Climate zones in the North Island based on normal climate data from 1961-1990

Climate zone	Rainfall mean (mm)	Rainfall range (mm)	Temperature mean (°C)	Temperature range (°C)
HR-MT	1709	1400-5899	11.8	8.5-13.0
HR-HT	1703	1400-2891	14.1	13.0-15.6
MR-HT	1278	1100-1399	14.4	13.0-15.8
MR-MT	1228	859-1399	11.7	8.8-13.0
LR-MT	981	725-1099	12.6	11.0-15.6

5.2.3. DEM and GIS surfaces

A 100 m resolution DEM of the North Island was obtained from the Precision Agricultural Centre, Massey University. GIS map layers of annual and seasonal daily mean temperature, annual and seasonal rainfall were interpolated from the NIWA temperature and rainfall normal data during 1961-1990 with a 100 m resolution. This interpolation generated surfaces of annual and seasonal temperatures rainfalls for the North Island from point data with a distance-weighted procedure. There were 364 points (locations) of rainfall and 101 points of temperature used in interpolating the rainfall and temperature surface, respectively. Hill slope and fertiliser input surfaces were also generated from this DEM. All the GIS analyses were conducted in Idrisi 3.2 (Clark Labs 1997).

5.2.4. Model development and assessment

A polynomial regression model was developed in SAS 8.1 and SAS Enterprise Miner, Version 4.1 (SAS Institute 1999).

Annual pasture production data was square root transformed to reduce the heterogeneous variances as all random errors having the same variance is a fundamental assumption underlying linear regression (Sen & Srivastava 1990). Climate variable (annual and seasonal rainfall and daily mean temperature), N fertiliser input and soil Olsen P were standardised to a range of 0-1 by scaling all the data with the largest observation in the dataset (Johnson & Wichern 1992). The variable of P fertiliser input was a combination of both annual P fertiliser input and P fertiliser application history as the previous decision tree model (Chapter 3) indicated that five-year cumulative P fertiliser input was a very significant variable in influencing pasture production. It was calculated as follows:

$$\text{P fertiliser input} = 0.5 \times \text{input}_{\text{annual}} + 0.5 \times \text{input}_{\text{history}} \qquad (1)$$

Where $\text{input}_{\text{annual}}$ is the annual P fertiliser input, $\text{input}_{\text{history}}$ is the average annual P fertiliser input in the previous five years before the sampling year. The P fertiliser input was also standardised by the method used for N fertiliser input and soil Olsen P. Hill slope was standardised by a cosine transformation to a scale of 0-1. This standardisation made the modelling process easy to conduct, especially in GIS analysis.

As pasture production showed a curvilinear response to temperature and rainfall in a primary analysis and this was also revealed in the decision tree model (Chapter 3), quadratic (squared) terms of temperature and rainfall were included in the model analyses. An interaction term between rainfall and P fertiliser input was also included in the model analyses along with other variables.

A forward stepwise approach was used to select the variables with a significance level of $P < 0.05$ in model analyses. To overcome the limitation of the stepwise approach, backward and forward approaches were also used to facilitate the variable selection (Sen & Srivastava 1990). Residuals were checked to identify outliers, and the condition index was checked to identify multicollinearity. Further improvements of model fit were then made accordingly.

Table 2: Scenarios of climate changes and P fertilizer input.

Scenario	Scenario description
Normal	using normal temperature and rainfall data (1960-1991) and a universal 100 P fertiliser (superphosphate) input, no N fertiliser input
R +10%~T +1 °C	rainfall increased by 10% and temperature increased by 1 °C from the normal scenario
R -10%~T +1 °C	rainfall decreased by 10% and temperature increased by 1 °C from the normal scenario
R +10%~T +2 °C	rainfall increased by 10% and temperature increased by 2 °C from the normal scenario
R -10%~T +2 °C	rainfall decreased by 10% and temperature increased by 2 °C from the normal scenario
R +20%~T +1 °C	rainfall increased by 20% and temperature increased by 1 °C from the normal scenario
R -20%~T +1 °C	rainfall decreased by 20% and temperature increased by 1 °C from the normal scenario
R +20%~T +2 °C	rainfall increased by 20% and temperature increased by 2 °C from the normal scenario
R -20%~T +2 °C	rainfall decreased by 20% and temperature increased by 2 °C from the normal scenario
150 kg/ha P	P fertiliser increased by 50 kg/ha from the normal scenario
200 kg/ha P	P fertiliser increased by 100 kg/ha from the normal scenario

A model assessment was implemented in the model analyses. The whole dataset was first randomly partitioned into two parts: the training data (70% of the total) and the validation data (30% of the total). The regression model was analysed using the training data and model was assessed by the validation data. This assessment compared the average squared error (ASE) of the potential sub models and the corresponding ASE of each sub model in the validation data. ASE is similar to mean squared error (MSE) and is an indicator of model goodness-of-fit; the smaller the ASE, the better the model fits. A "best" model that had the smallest ASE from validation data was selected (Fernandez 2003).

5.2.5. Climate change scenarios and GIS-based model prediction

With regard to the estimation of the changes in temperature (0.6 to 2.8 °C) and rainfall (-20% to +30%) by Ministry for the Environment (2001), I used eight scenarios of climate change which assumed an annual daily mean temperature increase by 1 to 2 °C and an annual total rainfall change by -20% to +20%. A normal scenario, which represented the average temperature and rainfall data for the period of 1961-1990, was used as a check for these scenarios. This normal scenario assumed a universal 100 kg P fertiliser input and no N fertiliser input. Two scenarios of P fertiliser input (150 kg/ha and 200 kg/ha) were also used. Details of these scenarios are in Table 2.

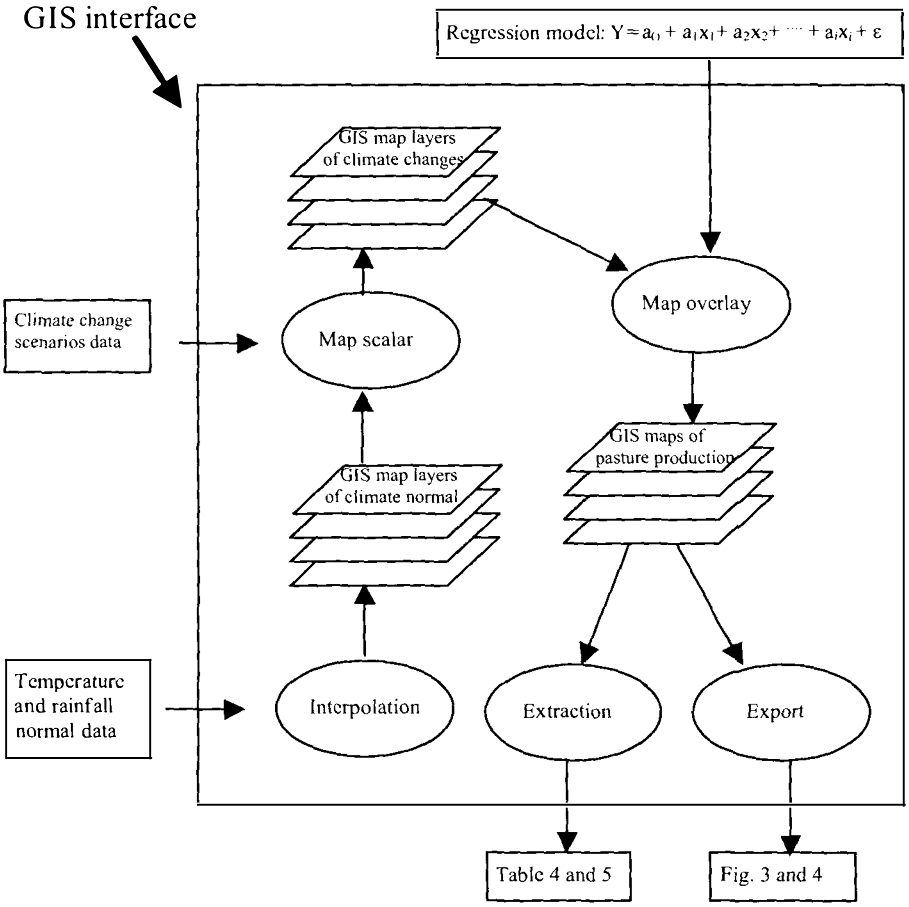


Fig. 2. The procedure for generating GIS-based predictions for pasture production under different climate scenarios. The rectangles represent data files and results, the ellipses represent the GIS functions and the parallelograms represent the GIS map layers.

Map layers of climate variables (temperature and rainfall) under different scenarios were first prepared based on the normal annual and seasonal mean daily temperature and rainfall map layers generated previously and were standardised by using the method previously described. Map layers of P fertiliser input under the two fertiliser input scenarios were also prepared and standardised. The GIS-based predictions for pasture production under different scenarios were then implemented in GIS using the Image Calculator by linking the developed regression model (Stocks & Wise 2000). The predicted pasture production was then back transformed for easy interpretation. Pasture productions in the different climate zones were extracted from the predicted GIS map layers of pasture production and were presented separately. Fig. 2 shows the main procedure for generating GIS-based prediction for pasture production under different climate and P fertiliser input scenarios.

5.3. Results

5.3.1. The regression model

Table 3 shows the results of the regression model for annual pasture production. The model was very significant (F =349.72, P <0.001) and explained 67.8% variance in annual pasture production. The significant input variables selected in the model included hill slope, P fertiliser input, annual daily mean temperature, annual and spring total rainfall, the quadratic terms of annual daily mean temperature and annual total rainfall and an interaction term between annual total rainfall and P fertiliser input (Table 3c). Soil Olsen P was significant in model analyses; however, because it was closely correlated with P fertiliser input and resulted in a high multicollinearity problem, it was not selected in the model as P fertiliser input and application history were indicated as more significant variables than Olsen P (Chapter 3). This was also good from a practical point of view as Olsen P is difficult to measure over a large scale. As this dataset only contained very limited samples of applying N fertiliser, the N fertiliser input was not significant in model analysis.

The effect of P fertiliser input on pasture production was positive, while the effects of temperature and rainfall were positive when they were low and were negative when they were high. The interaction between P fertiliser input and annual total rainfall was negative.

5.3.2. GIS-based model prediction

GIS-based prediction for annual pasture production under the normal scenario is shown in Fig. 3. Table 4 shows the predicted pasture productions (mean ± standard deviation) for the five climate zones under the normal scenario, the climate change scenarios and the two P fertiliser input scenarios, which were extracted from GIS map layers of predicted pasture production (map not shown). Fig. 4 shows the changes of predicted pasture production (mean ± standard deviation, kg/ha) between the climate change scenarios and the normal scenarios, and between the two P fertilizer input scenarios and the normal scenario for the whole study area. Table 5 displays the differences in pasture production in amount (mean ± standard deviation kg/ha) and in percentage (%) between the climate change scenarios and the normal scenarios, and between the two P fertilizer input scenarios and the normal scenario for the whole study area and the five climatic zones, which were extracted from Fig. 4.

Under the current climate situation (the normal scenario), the climate zone of HR-HT had the highest predicted average pasture production (12800 kg/ha); while LR-MT had the lowest (7600 kg/ha). The HR-HT was about 70% higher than the LR-MT.

For the eight climate change scenarios, scenario R +20%~T +2 °C had the highest predicted average pasture production (11320 kg/ha) for the whole study area, which was 9.6% higher than the normal scenario; scenario R -20%~T +2 °C had the lowest predicted average production (8220 kg/ha), which was 20.4% lower than normal scenario. For the two P fertilizer input scenarios, the 150 kg/ha P scenario was 3.6% higher and the 200 kg/ha P scenario was 7.2% higher than the normal scenario for the whole study area, respectively.

The predicted average pasture productions for the climate zone HR-MT all increased by 6.7% to 20.5% under the eight climate change scenarios compared with the normal scenario. Opposite to the climate zone HR-MT, the predicted average pasture productions in the climate zone HR-HT decreased by 30.1% to 7.7% under the eight climate change scenarios. The predicted average pasture production in the climate zone MR-HT decreased for all other climate scenarios by -46.2% to -2.3% except under the scenarios of R +20%~T +1 °C, which increased by 2.4%. For the climate zone MR-MT, the predicted average pasture production increased by 2.2% to 43.6% for all other climate scenarios except under the scenario R -20%~T +1 °C, which decreased by 6.2%. The predicted average pasture production for the climate zone LR-MT increased under the scenario R +10%~T +1 °C (25.4%), R +10%~T +2°C (22.4%), R +20%~T +1 °C (38.0%) and R +20%~T +2 °C (34.8%), and decreased under all other scenarios by -

22.0% to -3.8%. That is the average pasture production increased under all the increased rainfall scenarios and decreased in all the decreased rainfall scenarios.

For the two P fertilizer input scenarios, an increased average pasture production was predicted for all the five climate zones compared with the normal scenario which assumed a 100 kg/ha P fertilizer input. For 150 kg/ha scenario, the highest increase was in the climate zone LR-MT (6.7%) and the lowest was in the climate zone HR-HT (2.1%). For 200 kg/ha scenario, the climate zone LR-MT also had the highest increase (13.6%) and the climate zone HR-HT had the lowest increase (4.2%).

Table 3: Summary of the regression model for pasture production. The variable symbols are: Cos(slope) (cosine slope), P_fert (P fertilizer input, kg/ha), Temp_y (annual daily mean temperature, °C), Rain_sp (spring total rainfall, mm), Rain_y (annual total rainfall, mm).

a. Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	8	539956	67494.0	349.72	<.0001
Error	1319	254562	193.0		
C Total	1327	794517			

b. Model fit

R-square	0.6796	Adjusted R-square	0.6777

c. Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|
Intercept	1	-712.5	88.47	-8.05	<.0001
Cos(slope)	1	150.3	8.04	18.70	<.0001
P_Fert	1	55.9	7.94	7.04	<.0001
Temp_y	1	1216.3	214.80	5.66	<.0001
Rain_sp	1	41.5	4.92	8.44	<.0001
Rain_y	1	229.8	24.54	9.37	<.0001
Rain_y*P_fert	1	-45.7	11.70	-3.91	<.0001
Square (temp_y)	1	-671.1	127.20	-5.28	<.0001
Square (rain_y)	1	-123.6	19.80	-6.24	<.0001

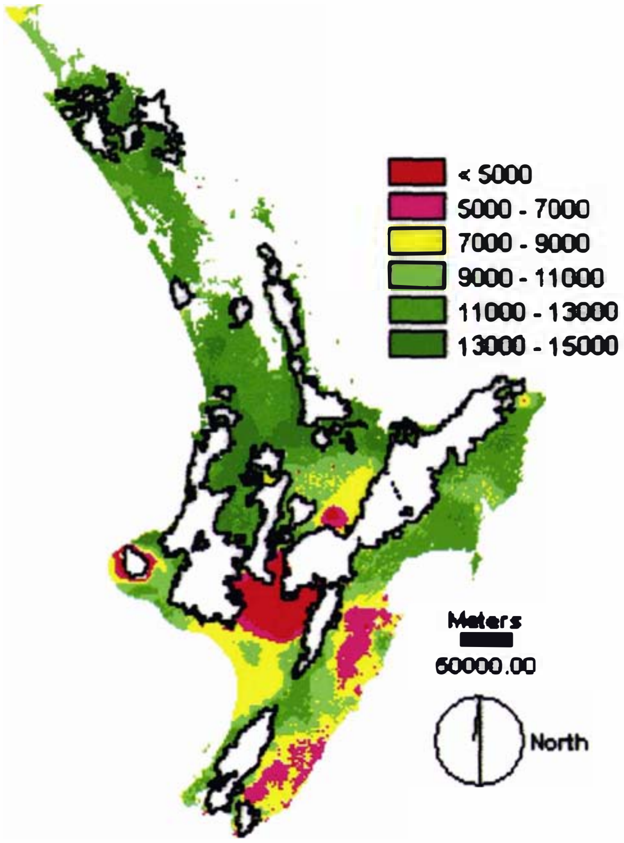


Fig. 3. Predicted pasture production (kg/ha) under normal scenario for the North Island. The white areas on the map are forest.

For scenarios that assumed 1 °C increase in temperature a change in pasture production of -29.8% to 41.2% was predicted when rainfall change was from -20% to +20%; while for scenarios that assumed 2 °C increase in temperature a change in pasture production of -46.2% to +51.9% was predicted when rainfall change was from -20% to +20% compared with the current climate situation (the normal scenario).

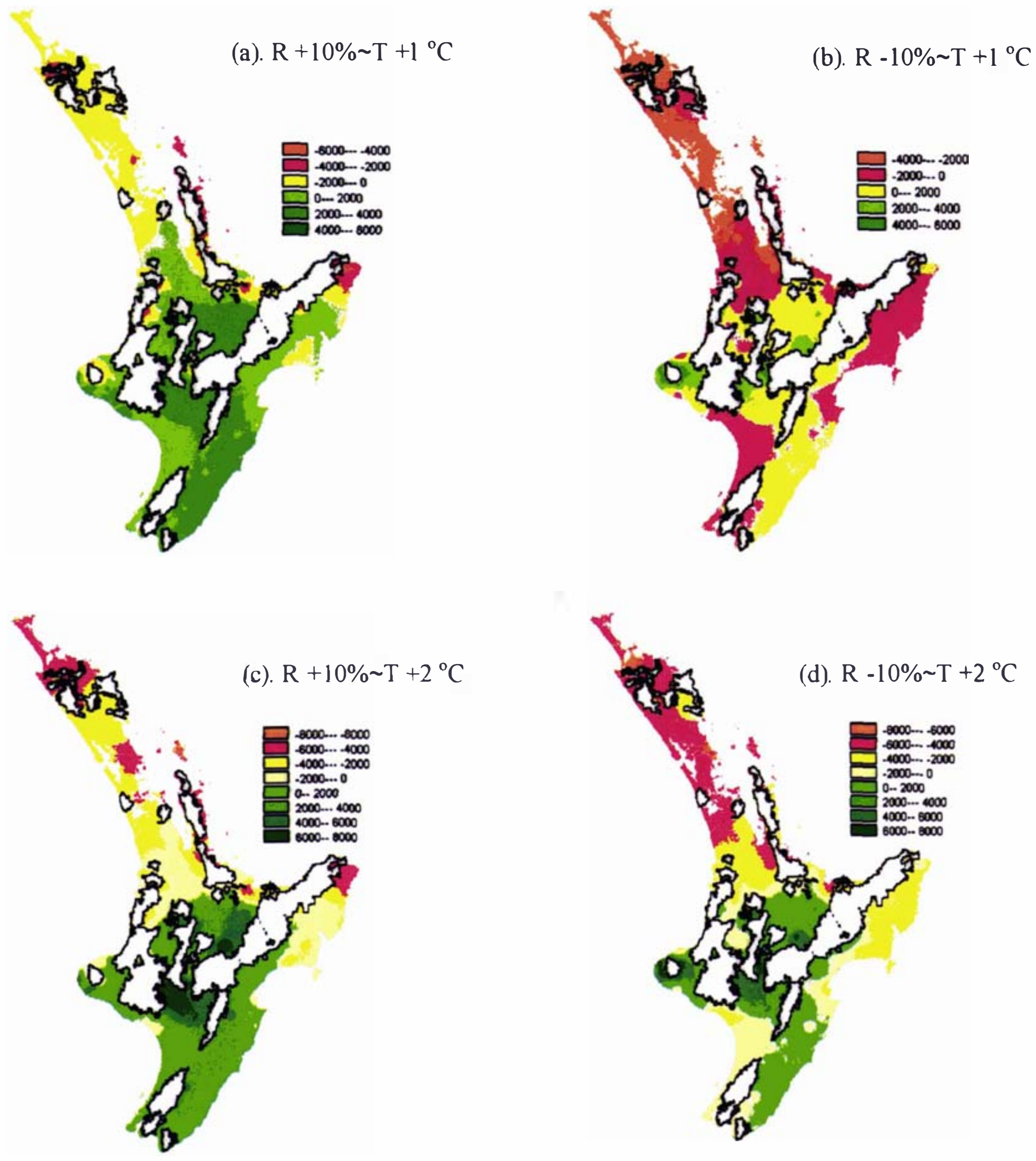
For scenarios that assumed 1°C temperature increase, the highest decrease in pasture production was found in the climate zone MR-HT under scenario R -20%~T +1 °C and the highest increase was found in the climate zone MR-MT under scenario R +20%~T +1 °C. For scenarios that assumed 2 °C temperature increase, the highest decrease in pasture production was also found in the climate zone MR-HT under scenario R -20%~T +2 °C and the highest increase was also found in the climate zone MR-MT under scenario R +20%~T +2 °C.

Table 4: Predicted pasture production (mean ± standard deviation) (kg/ha) for the whole study area and the five climatic zones under different scenarios. See Table 1 for the climate zones and Table 2 for scenarios.

Scenarios	All area	HR-MT	HR-HT	MR-HT	MR-MT	LR-MT
normal	10328 ± 2782	10330 ± 2736	12807 ± 1010	11523 ± 941	8144 ± 3048	7601 ± 1069
R +10%~T +1 °C	11169 ± 2190	12007 ± 2867	11826 ± 1811	11252 ± 1334	10861 ± 2600	9532 ± 901
R -10%~T +1 °C	9934 ± 2366	11802 ± 1931	11640 ± 1426	9360 ± 1288	8894 ± 2496	7316 ± 818
R +10%~T +2 °C	10433 ± 2413	12612 ± 2583	9864 ± 2150	9016 ± 1802	11697 ± 1856	9302 ± 1206
R -10%~T +2 °C	9239 ± 2497	12448 ± 1517	9693 ± 1899	7326 ± 1667	9637 ± 1882	7117 ± 1090
R +20%~T +1 °C	11323 ± 2292	11480 ± 3308	11214 ± 2339	11803 ± 1327	11502 ± 2546	10489 ± 907
R -20%~T +1 °C	8879 ± 2460	11024 ± 1711	10825 ± 1570	8093 ± 1219	7639 ± 2335	6113 ± 744
R +20%~T +2 °C	10587 ± 2552	12048 ± 3127	9312 ± 2498	9511 ± 1832	12369 ± 1741	10247 ± 1227
R -20%~ T +2 °C	8223 ± 2567	11662 ± 1420	8949 ± 1952	6205 ± 1548	8321 ± 1787	5931 ± 993
150 kg/ha P	10697 ± 2772	10573 ± 2814	13074 ± 1049	11977 ± 932	8544 ± 3103	8111 ± 1088
200 kg/ha P	11073 ± 2766	10820 ± 2898	13345 ± 1100	12440 ± 923	8953 ± 3155	8637 ± 1107

Table 5: Changes in pasture production (mean ± standard deviation, kg/ha) in amount (mean ± standard deviation kg/ha) and in percentage (%) between the climate change scenarios and the normal scenarios and between the two P fertilizer input scenarios and the normal scenario for the whole study area and the five climatic zones. See Table 1 for the climate zones and Table 2 for scenarios.

Scenarios	All area		HR-MT		HR-HT		MR-HT		MR-MT		LR-MT	
	Kg/ha	%	Kg/ha	%	Kg/ha	%	Kg/ha	%	Kg/ha	%	Kg/ha	%
R +10%~T +1 °C	841 ± 1714	8.1	1677 ± 1111	16.2	-981 ± 1134	-7.7	-271 ± 815	-2.3	2717 ± 630	33.4	1931 ± 738	25.4
R -10%~T +1 °C	-394 ± 1589	-3.8	1472 ± 1233	14.3	-1167 ± 1003	-9.1	-2162 ± 709	-18.8	749 ± 750	9.2	-285 ± 745	-3.8
R +10%~T +2 °C	105 ± 3134	1.0	2282 ± 1817	22.1	-2943 ± 1550	-23.0	-2507 ± 1394	-21.8	3552 ± 1641	43.6	1701 ± 1489	22.4
R -10%~T +2 °C	-1089 ± 2965	-10.5	2118 ± 2088	20.5	-3114 ± 1500	-24.3	-4197 ± 1224	-36.4	1493 ± 1656	18.3	-484 ± 1408	-6.4
R +20%~T +1 °C	995 ± 2264	9.6	1150 ± 1768	11.1	-1593 ± 1776	-12.4	280 ± 884	2.4	3358 ± 657	41.2	2888 ± 729	38.0
R -20%~T +1 °C	-1449 ± 1843	-14.0	694 ± 1953	6.7	-1982 ± 1389	-15.5	-3430 ± 652	-29.8	-505 ± 878	-6.2	-1488 ± 747	-19.6
R +20%~T +2 °C	259 ± 3496	2.5	1717 ± 2277	16.6	-3495 ± 1945	-27.3	-2012 ± 1467	-17.5	4225 ± 1697	51.9	2646 ± 1513	34.8
R -20%~T +2 °C	-2105 ± 3047	-20.4	1332 ± 2700	12.9	-3858 ± 1711	-30.1	-5318 ± 1116	-46.2	177 ± 1709	2.2	-1670 ± 1352	-22.0
150 kg/ha P	369 ± 161	3.6	243 ± 164	2.4	267 ± 160	2.1	455 ± 83	3.9	399 ± 99	4.9	510 ± 82	6.7
200 kg/ha P	745 ± 291	7.2	490 ± 285	4.7	538 ± 266	4.2	918 ± 112	8.0	808 ± 157	9.9	1036 ± 104	13.6



(a). R +10%~T +1 °C

(b). R -10%~T +1 °C

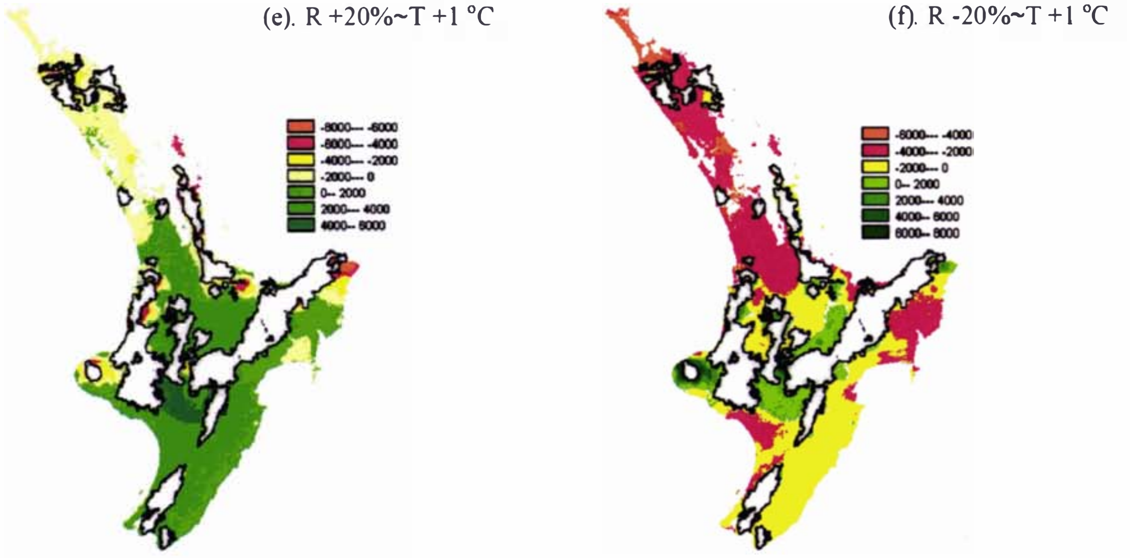
(c). R +10%~T +2 °C

(d). R -10%~T +2 °C

Fig. 4. Changes in pasture production (mean ± standard deviation, kg/ha) under the climate change scenarios (difference between the climate change and the normal scenario) (a-h) and P fertiliser input scenarios (difference between the P fertiliser input scenarios and the normal scenario) (i-j), for the North Island. The white areas on the map are forest. Note the different scales in the map key.

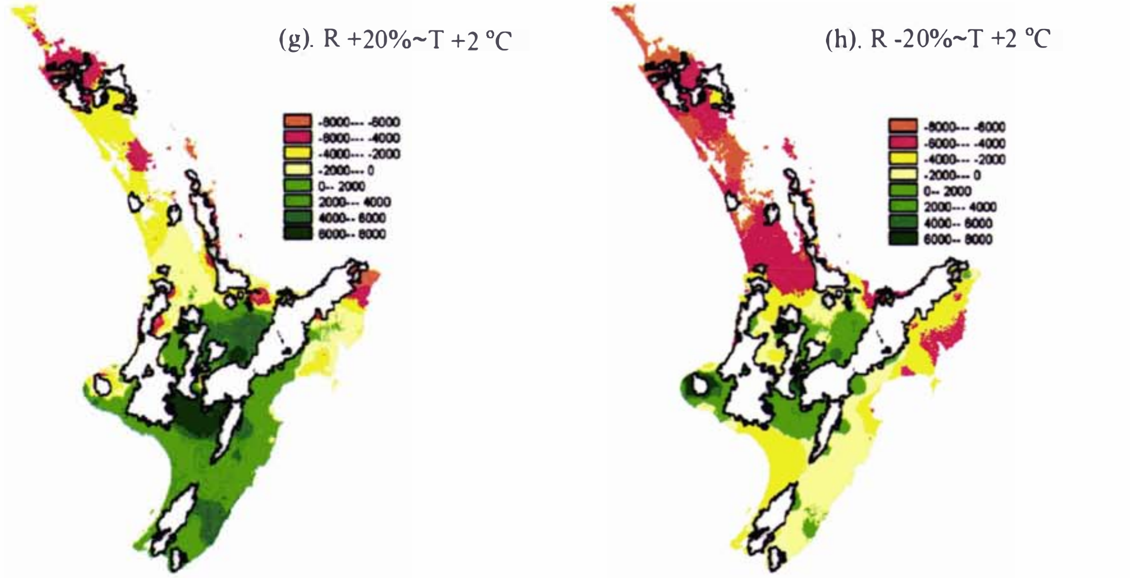
Fig. 4 continued on next page

Fig. 4 continued



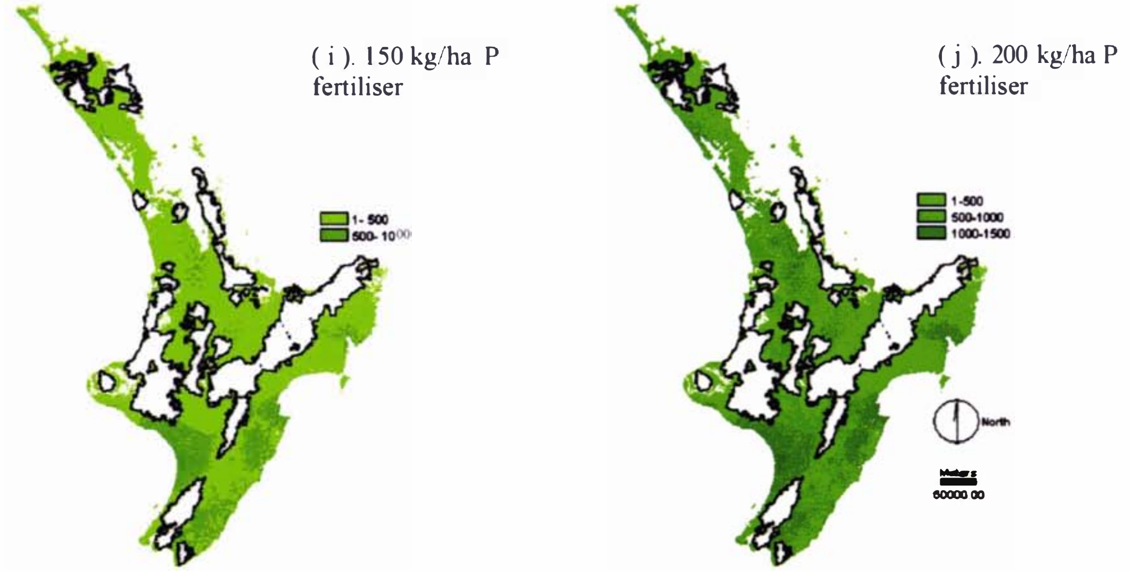
(e). R +20%~T +1 °C

(f). R -20%~T +1 °C



(g). R +20%~T +2 °C

(h). R -20%~T +2 °C



(i). 150 kg/ha P fertiliser

(j). 200 kg/ha P fertiliser

Scenarios assuming a 10% decrease in rainfall predicted a change of pasture production by -24.3% to +20.5%. Scenarios that assumed a 10% increase in rainfall predicted a change in pasture production by -23.0% to +43.6%. Scenarios that assumed a 20% decrease in rainfall predicted a change in pasture production by -46.2% to +12.9%. Scenarios that assumed a 20% increase in rainfall predicted a change in pasture production by -27.3% to +51.9%.

5.4. Discussion

The model predicted a wide range (-46.2% to +51.9%) of pasture production change for the eight climate change scenarios compared with the normal climate from 1961-1990 in the different climate zones (Table 5). This variation in pasture production indicated that the potential impacts of climate change on pasture in the North Island would be very significant. These pasture production changes were in a range of -20.4% to +9.6% under the eight climate scenarios when averaged over all the climate zones, which was in general agreement with the findings by Baars *et al.* (1990), Martin *et al.* (1991) and Campbell (1996). Their predictions were usually in the range of +8% to +30% and a negative effect was also predicted in some regions. However, my model predicted a decrease in pasture production under the four scenarios with decreased rainfall for the whole North Island when averaged over all the climate zones, showing that the interaction of decreased rainfall and increased temperature will have significant negative effect on pasture production.

The climate change report of Ministry for the Environment (2001) estimated that the climate change in rainfall was likely to result in an increase in the West of the New Zealand and a decrease in the East. Under such a climate change situation, the four scenarios that assumed increased rainfall would be the case for the West and the four scenarios that assumed decreased rainfall would be the case for the East of the North Island. Corresponding to these climate situations, pasture production in most of the South-west would, on average, increase by +11.1% to +51.9%, in most of the North and North-west would, on average, decrease by -2.3% to -27.3% and in most of the South-east would, on average, decrease by -3.8% to -22.0%. This indicates that decreased rainfall and increased temperature would have negative effect for areas having low rainfall (e.g. LR-MT) and areas having high temperature (e.g. HR-HT, MR-HT), respectively, with moisture stress being the major limitation on pasture growth. Currently most pastures in New Zealand are ryegrass and white clover dominated and both of them are not drought-tolerant. Developing and wide use of drought-tolerant species in areas where climate changes are likely to have negative impact (e.g. North and South-east of the North Island) would be a good management practice in the face of climate change.

Results from the model predictions for the P fertiliser input scenarios indicated that the effect of increasing P fertiliser input from 100 kg/ha (normal scenario) to 200 kg/ha would increase pasture production by 4.2% to 13.6%, depending on the climate zone. It was also indicated that pasture in areas with relatively low rainfall had a higher response to increased P fertiliser input than pastures in areas with a relatively high rainfall. An interaction of soil Olsen P and rainfall on pasture production was recognised and the effect of Olsen P on pasture production was positively influenced by available soil moisture (Chapter 3 and Moir et al. 2000). The reason why the interaction between increased P fertiliser input and rainfall was negative was not clear. A possible reason is that the high rainfall area may have higher soil P content than low rainfall area, and thus response to P fertiliser was less effective than for the low rainfall area. However, by checking the dataset, soil Olsen P was found lower in high rainfall areas than low rainfall areas, showing that this may not be the reason for the low response of P fertiliser to pasture production in high rainfall areas. A possible explanation is that runoff resulting from rainfall caused the associated P fertiliser loss. As P fertiliser was mainly aerially applied and most fertiliser may have remained on the surface of pastures, the proportion of P fertiliser loss by runoff may be positively related to the rainfall, and thus, the area with high rainfall may have low efficiency in utilising P fertiliser. Lambert *et al.* (1985) also found that proportion of total P in dissolved inorganic form in runoff tended to be higher for catchments applying more P fertiliser than those applying less P fertiliser. This negative interaction between rainfall and P fertiliser input on pasture production implies that increased rainfall that resulted from climate change may reduce the P fertiliser utilisation efficiency in hill-pasture.

The regression model developed in this study had a very good model fit considering that the training data were from different locations and times. This was partially because the most significant input variables and the interactions among these variables in explaining variance in pasture production were revealed from the previously developed decision tree model (Chapter 3). These variables were also recognised as the key drivers in influencing pasture production in some other studies (e.g. Radcliffe 1982; Lambert *et al*. 1983; Gillingham *et al*. 1998; Moir *et al*. 2000). It was indicated by this model that apart from P fertiliser input and hill slope, all other independent variables were temperature, rainfall, and the interaction between rainfall and P fertiliser. The importance of climate and the impact of climate changes on pasture production were well supported by the model.

Naturally, this model also has some limitations. Firstly, it did not assess the effect of the elevated CO_2 concentration on pasture production (Riedo *et al*. 1997), no effect of extreme climatic events such as drought or heavy rainfall, which may result from climate change were assessed (Salinger *et al*. 2000). Secondly, the data used in developing the regression model were mainly from hill-pasture, and therefore, model predictions for lowland pasture may not be as good as for hill-pasture, as lowland pasture was assumed to be the low slope ($< 15^o$) area in hill-pasture. However, as a scenario analysis, the objective of this study was not to accurately predict pasture production under climate changes, but rather, to assess the possible impacts of climate change scenarios on pasture production and management practices which may need to be considered under such an impact.

In summary, the model predictions for climate change scenarios suggested that the potential climate changes would have significant impact on pasture production in the North Island, New Zealand. Increased temperature would generally have a positive effect on pasture production in the South and South-east of the North Island, and increased rainfall would have a positive effect in the Central, South and South-east of the North Island and a negative effect in the North of the North Island. The decreased rainfall would only have a positive effect in some central areas with high rainfall. The interaction of increased temperature and decreased rainfall would, on average, have a negative impact on pasture production for the whole North Island. Management strategies such as developing and use drought tolerant species, need to be considered to cope with the potential drought resulted from reduced rainfall or increased temperature for areas such as North and East coast of the North Island. The model prediction also indicated that pasture in areas with relatively low rainfall had a higher response to increased P fertiliser input than pastures in areas with a relatively high rainfall, probably due to the high runoff and associated P loss in high rainfall area. This suggests that use N fertiliser in areas with high rainfall would be a better choice with respect to both economic benefit and environmental health.

5.5. References

Allen M., Raper S. & Mitchell J. (2001) Uncertainty in the IPCC's Third Assessment Report. *Science*, 293, 430-433

Armstrong A.C. (1996) The impact of climate change on grassland production and utilisation. *Aspects of Applied Biology*, 45, 79-84

Baars J.A., Radcliffe J.E. & Rollo M.D. (1990) Climatic change effects on seasonal patterns of pasture production in New Zealand. *Proceedings of the New Zealand Grassland Association*, 51, 43-46

Baars J.A. & Waller J.E. (1979) Effects of temperature on pasture production. *Proceedings of the Agronomy Society of New Zealand*, 9, 101-104

Bai Y.F., Han X.G., Wu J.G., Chen Z.Z. & Li L.H. (2004) Ecosystem stability and compensatory effects in the Inner Mongolia grassland. *Nature*, 431, 181-184

Barrett J.E., McCulley R.L., Lane D.R., Burke I.C. & Lauenroth W.K. (2002) Influence of climate variability on plant production and N-mineralization in Central US grassland. *Journal of Vegetation Science*, 13, 383-394

Campbell B.D. (1996) Climate change and pastures. *Proceedings 48th Ruakura Farmers' Conference held at Ruakura, New Zealand, 11 June 1996*, 67-74

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

Fernandez G. (2003) *Data mining using SAS applications.* Chapman & Hall/CRC, Boca Raton, Florida

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

Gillingham A.G., Rickard S., Power I.L. & Riley J. (1990) Long-term effects of withholding phosphate application on North Island hill country: Whatawhata

Research Centre. *Proceedings of the New Zealand Grassland Association*, 51, 11-16

Halloy S.R.P. & Mark A.F. (2003) Climate-change effects on alpine plant biodiversity: A New Zealand perspective on quantifying the threat. *Arctic and Alpine Research*, 35, 248-254

Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J., Dai X., Maskell K. & Johnson C.A. (2001) *Climate Change 2001: The scientific basis, intergovernmental panel on climate change Geneva.* Cambridge University Press, Cambridge, U.K., and New York, U.S.A., pp 881

Hunter G.G. & Blaschke P.M. (1986) *The New Zealand land resource inventory vegetation cover classification.* Wellington, NZ: Published for the National Water and Soil Conservation Authority by the Water and Soil Directorate, Ministry of Works and Development, 1986

IPCC (2000) *Special report on emmision scenarios.* Cambridge University Press. Cambridge UK, pp 599

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

Johnson R.A. & Wichern D.W. (1992) *Applied Multivariate Statistical Analysis.* 3rd Edition. Englewood Cliffs, New Jersey: Prentice Hall

Joos F., Prentice I.C., Sitch S., Meyer R., Hooss G., Plattner G.K., Gerber S. & Hasselmann K. (2001) Global warming feedbacks on terrestrial carbon uptake under the Intergovernmental Panel on Climate Change (IPCC) emission scenarios. *Global Biogeochemical Cycles*, 15, 891-907

Kattenberg A., Giorgi F., Grassl H., Meehl G.A., Mitchell J.F.B., Stouffer R.J., Tokioka T., Weaver A.J. & Wigley T.M.L. (1996) Climate models - projections of future climate. pp 285-357. In J. T. Houghton et al., editors. *Climate Change 1995: the science of climate change.* Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Kenny G.J., Warrick R.A., Campbell B.D., Sims G.C., Camilleri M., Jamieson P.D., Mitchell N.D., McPherson H.G. & Salinger M.J. (2000) Investigating climate change impacts and thresholds: An application of the CLIMPACTS integrated assessment model for New Zealand agriculture. *Climatic Change*, 46, 91-113

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

Lambert M.G., Devantier B.P., Nes P. & Penny P.E. (1985) Losses of nitrogen, phosphorus, and sediment in runoff from hill country under different fertiliser and grazing management regimes. *New Zealand Journal of Agricultural Research*, 28, 371-379

Martin R.J., Korte C.J., McCall D.G., Baird D.B., Newton P.C.D. & Barlow N.D. (1991) Impact of potential change in climate and atmospheric concentration of carbon dioxide on pasture and animal production in New Zealand. *Proceedings of the New Zealand Society of Animal Production*, 51, 25-33

McCarthy J.J., Canziani O.F., Learty N.A., Dokken D.J. & White K.S. (2001) *Climate change 2001: impacts, adaptation, and vulnerability,* international panel on climate change, Geneva . Cambridge University Press, Cambridge, U.K. and New York, U.S.A., pp1032

Ministry for the Environment (2001) *Climate change impacts on New Zealand.* Ministry for the Environment, Wellington, New Zealand

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

Plummer N., Salinger M.J., Nicholls N., Suppiah R., Hennessy K.J., Leighton R.M., Trewin B.C., Page C.M. & Lough J.M. (1999) Changes in climate extremes over the Australian region and New Zealand during the twentieth century. *Climatic Change*, 42, 183-202

Radcliffe J.E. (1982) Effects of aspect and topography on pasture production in hill country. *New Zealand Journal of Agricultural Research*, 25, 486-496

Radcliffe J.E. & Barrs J.A. (1987) Temperate pasture production. In: Snaydon, R.W. (Ed.) *Ecosystems of the World*, 17B, 7-17

Riedo M., Gyalistras D., Grub A., Rosset M. & Fuhrer J. (1997) Modelling grassland responses to climate change and elevated CO_2. *Acta Oecologica*, 18, 305-311

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

Salinger M.J., Stigter C.J. & Das H.P. (2000) Agrometeorological adaptation strategies to increasing climate variability and climate change. *Agricultural and Forest Meteorology*, 103, 167–184

Sen A.K. & Srivastava M. (1990) *Regression analysis: theory, methods and applications.* New York : Springer-Verlag

Stocks C.E. & Wise S. (2000) The role of GIS in environmental modelling. *Geographical and Environmental modelling*, 4, 219-235

Thornley J.H.M. & Cannell M.G.R. (1997) Temperate grassland responses to climate change: an analysis using the Hurley Pasture Model. *Annals of Botany*, 80, 205-221

Tomlinson A.I. & Sansom J. (1994) *Temperature normals for New Zealand for the period 1961 to 1990.* Wellington [N.Z.]: NIWA

White J.G.H. (1990) Hill and high country pasture. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

Chapter 6. A Decision Tree Approach to Modelling Functional Group Abundance of Plant Species in a Pasture Ecosystem

Species abundance has been considered to have several distributions and it is not appropriate to use least square (LS) regression. Generalised linear model (GLM) has been successfully used to model species abundance using the measure of discrete cover, but it is not suitable when abundance is an interval variable such as measured by biomass. Decision tree has no strict requirement for variable distribution and can easily incorporate ordinal, nominal and interval variables in a model. It is, therefore, an appropriate approach for modelling species abundance with productivity or biomass as the measure. In this chapter, a combination of decision tree and GIS were used to predict relative abundance of functional groups of plant species in a hill-pasture ecosystem.

A paper derived from this chapter with a title "A decision tree approach modelling functional group abundance in a pasture ecosystem" by B. Zhang, I. Valentine & P.D. Kemp has been published in ***Agriculture, Ecosystem and Environment*** *(in press, available on-line).*

Abstract. Due to concerns over the impact of intensive land-use and climate change on species diversity, there have been increasing interests in the prediction of the distribution (presence/absence) and abundance of plant species. However, few models focus on species abundance, measured as productivity or biomass, which would contribute to ecological studies and environment management. Decision tree, one of the data mining approaches, was used to model the relative abundance of five functional groups of plant species, namely high fertility response grasses (HFRG), low fertility tolerance grasses (LFTG), legume, moss and flatweeds in a New Zealand hill-pasture ecosystem using aboveground biomass. The model outputs were integrated with a geographic information system (GIS) to map and validate the predictions on a pasture. The decision tree models clearly revealed the interactions between the functional groups and environmental and management factors, and also indicated the relative importance of these factors in influencing the functional group abundance. Soil Olsen P was the most significant factor influencing the abundance of LFTG and moss, while soil bulk density, slope and annual P fertiliser input were the most significant factors influencing the abundance of legume, HFRG and flatweeds, respectively. Generally, topographic feature (slope) and soil fertility (Olsen P) were the two key factors underlying the patterns of abundance for these five functional groups. The decision tree models also performed well with respect to the predictive accuracy. For the five functional groups studied, there was an overall predictive accuracy of 75%. Modelling functional group abundance simplified the investigation of the complex interrelationship between species and environment in a pasture ecosystem. The integration of the decision tree with GIS in this study provides a platform to investigate community structure and functional composition for a pasture over space, and thus can be applied as a tool in pasture management. For example, predicting and mapping the HFRG abundance in a pasture could guide more specific phosphorus fertiliser application by outlining the areas that have less HFRG composition.

6.1. Introduction

There have been increased applications of the predictive modelling of species distribution and abundance in plant ecology and vegetation science due to concern over the impact of intensive land-use and climate change on species distribution and diversity, and the related issues in species conservation and environment management (e.g. Palmer & Van Stden 1992; Tappeiner *et al.* 1998; Wadsworth & Reweek 1999; Zimmermann & Kienast 1999; Stocks & Wise 2000; Rouget *et al.* 2001; Guisan *et al.* 2002). Current practices mainly focus on establishing the correlation between plant species and environment by detecting the important variables underlying, or explaining this relationship and quantifying species' realised niche in a multiple space defined by these variables, and, usually, presenting predicted species distribution (presence/absence) over an area of interest in a GIS (Franklin 1995; Guisan & Zimmermann 2000; Austin 2002). The success of this modelling approach to species distribution is that it is founded on species niche theory and gradient analysis (Austin 2002).

One limitation of most of the above models is that they can predict the presence or absence of a species, but not the abundance of the species in the community. Yet, the importance of measuring species abundance in investigating species diversity, species community structure and ecosystem processes has been well recognised (May 1975; Hahel 1990; Camargo 1995; Tilman 1996; Welson *et al.* 1996). Presence or absence of a species is predicted in these models either because it is adequate for the modelling purpose, or, more possibly, because it is difficult to obtain the species abundance data used to develop the models. There are models that predict species abundance using discrete cover (e.g. Gottfried *et al.* 1998; Guisan & Harrell 2000), but cover is often estimated visually in these studies, usually using an interval scale, such as the Braun-Blanquet approach (Braun-Blanquet 1964). In grassland, or pasture ecosystems, discrete cover as a measure of species abundance is usually not adequate in investigating the community structure as species that have the same abundance (discrete cover) may have very different productivity or biomass (Chiarucci *et al.* 1999). Previous research has indicated that productivity, or biomass, is the most appropriate measure of species abundance, especially in a pasture or grassland ecosystem (Whittaker 1965; Welson *et al.* 1996; Guo & Rundel 1997; Chiarucci *et al.* 1999).

The modelling of species distribution and abundance has exploited many approaches, including multiple regression (Fels 1994), generalized linear model (GLM) (Austin *et al.* 1994; Lehmann *et al.* 2003), artificial neural networks (ANN) (Fitzgerald & Lees 1992; Gullison & Bourque 2001), decision trees (Iverson & Prasad 1998; Vayssieres *et al.* 2000), canonical correspondence analysis (CCA) (Ohmann & Spiess 1998; Guisan *et al.*

1999), and generalized additive model (GAM) (Brown 1994; Lehmann 1998). Species abundance has been considered to have several distributions, including Poisson distribution, negative binomial distribution, canonical log-normal distribution, broken-stick distribution and ordinal distribution (see Guisan & Zimmermann 2000). It is, therefore, not appropriate to directly use least square (LS) regression to model species abundance due to the violation of its normal distribution assumption. Though variable transformation can be applied to cope with this problem, it only alleviates the problem to some degree (Freund & Littell 1991). Similarly, CCA also assumes normal distribution though it has been argued that CCA works well when the assumption is not held (ter Braak 1985). GLM has been successfully used to model species abundance using the measure of discrete cover (Guisan & Harrell 2000), but it is not suitable when abundance is an interval variable such as measured by biomass. A model which is free of variable distribution assumptions and has no limitation of variable types would be an appropriate one in modelling species abundance.

Decision tree, one of the data mining approaches, has been used to model species distribution and abundance in several studies and was indicated to have a very high predictive ability, and performed better than GLM (Iverson & Prasad 1998; Vayssieres *et al.* 2000; Rouget *et al.* 2001). It has no strict requirement for variable distribution, and can easily incorporate ordinal, nominal and interval variables in a model (Breiman *et al.* 1984; Iverson & Prasad 1998; Vayssieres *et al.* 2000; Witten & Erank 2000). These advantages enable decision tree to be an appropriate approach in modelling species abundance with productivity or biomass as the measure.

A functional group is defined as a set of species that have similar effects on a specific ecosystem-level biogeochemical process (Vitousek & Hooper 1994). It is commonly used as a vegetation unit in ecological studies for investigating ecosystem processes and simplifying the complex interaction between vegetation and the environment (Korner 1994). The functional group is also an appropriate vegetation unit in assessing pasture management practices (Nicholas *et al.* 1998; Boer & Stafford Smith 2003).

In this study, I develop decision tree models for the relative abundance of functional groups of plant species using the measure of aboveground biomass in a hill-pasture in New Zealand, and present model predictions in a GIS for model validation and further investigation and application. The objectives of this study were: 1) to develop decision tree models for relative abundance of the functional groups; 2) to assess the performance of decision trees in modelling functional group abundance; and 3) to investigate the interrelationship between functional group abundance and environmental and pasture management variables, identifying the key variables driving or explaining these relationships and the implications for pasture management.

6.2. Methods

6.2.1. Study area

This study was carried out on the AgResearch Ballantrae Research Station at latitude S40°18′, longitude E175°50′ in the North Island of New Zealand (Fig. 1) with average annual rainfall 1270 mm and average annual daily temperature 12.3 °C. Altitude ranges from 125 to 350 m a.s.l. The hill country in Ballantrae is dissected and very heterogeneous with complex combinations of slope and aspect even within a small area (Lambert & Roberts 1978). This heterogeneity increases the diversity of micro-climate, soil characteristics, species composition and the behaviour of grazing animals (Lopez 2000). The pasture on Ballantrae is dominated by browntop *(Agrostis capillaris)* when soil fertility is low (or co-dominated by browntop and moss (*Musci* spp.) when soil fertility is low and hill slope is high) and co-dominated by ryegrass *(Lolium perenne)*, browntop and white clover *(Trifolium repens)* when soil fertility is high (Nicholas 1999).

The GIS-based model prediction and verification were carried out for a pasture, which was within the Station and was about 90 ha in area (Fig. 1). Within this pasture there were four small farmlets with different fertiliser application treatments since the early 1970s: two of them had approximately 120kg and 360kg annual P fertiliser (SuperPhosphate – approximately 9% P content) input per year since 1974 (hereafter referred as LL and HH), and the other two farmlets also had approximately 120kg and 360 kg P fertiliser per year from 1974 to 1982, but had none since (hereafter referred as LN and HN). The farmlets LL, LN, HN, and HH had 8, 10, 8 and 10 paddocks respectively. Each paddock was approximately 8 ha in area. A detailed description of these four farmlets and their management history can be found in Lambert *et al* (2000).

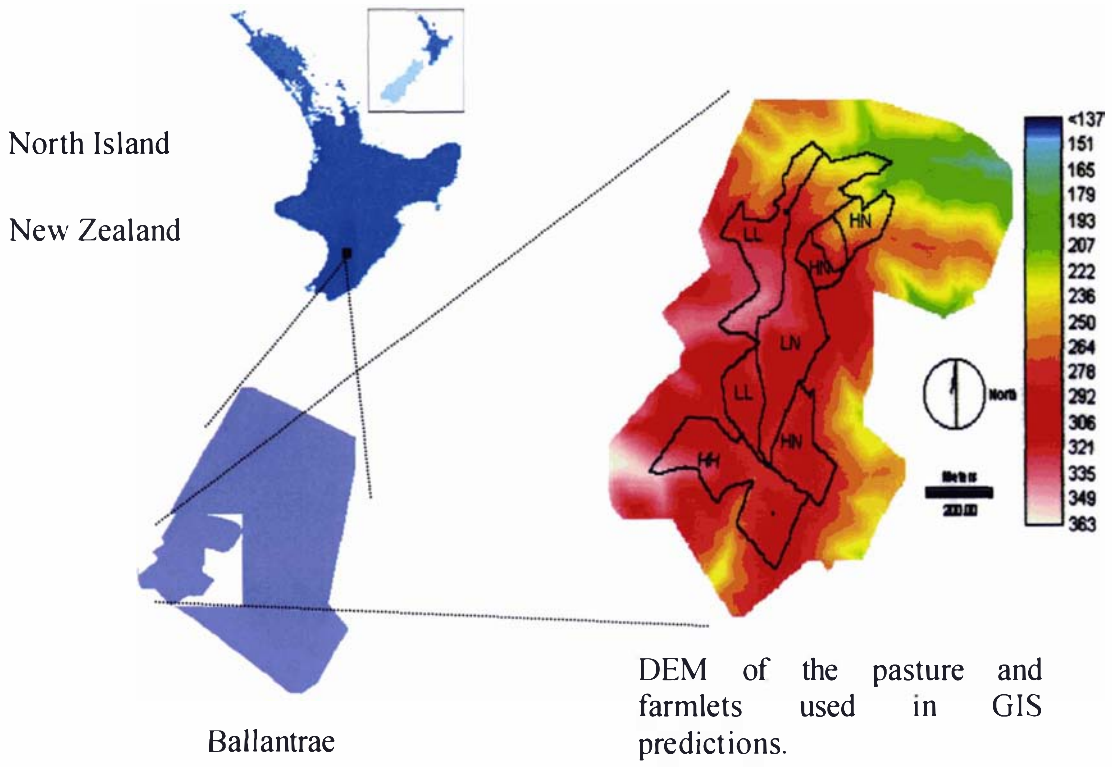


Fig. 1. Study area, digital elevation model (DEM) and the farmlets' used in validating the GIS prediction.

6.2.2. Model development

6.2.2.1. Dataset

Species data are from a long-term experiment (see Lambert et al. 1986), unpublished Ph.D theses (Nicholas 1999; Lopez 2000) and some other experiments conducted at Ballantrae during 1972-1998 within the whole area of the Station. A total of 1742 samples were collected using a stratified random method within the study area using 0.5 m^2 quadrats. Samples were collected during late spring to early summer (October to early December), which included aboveground biomass of different functional groups and soil features (see below). Aboveground biomass was measured by a "trim technique" involving harvesting pasture regrowth from trimmed areas protected from sheep grazing (Lambert *et al.* 1996) with a regrowth period from 30-50 days. Relative abundance of a functional group was calculated as the percentage of its aboveground biomass out of the total aboveground biomass in a sample.

Table 1. Input variables used in the decision tree and the GIS analyses

Variable symbol	Unit	Variable description
pH	$-\log_{10}[H^+]$	soil pH
BD	g/cm^3	soil bulk density
OlsenP	µg/g	soil Olsen P
N_fert	kg/ha/y	annual elemental N fertiliser input
P_fert	kg/ha/y	annual elemental P fertiliser input
P_fert5	kg/ha	5-years cumulative elemental P fertiliser input
P_fert10	kg/ha	10-years cumulative elemental P fertiliser input
temp_y	^0C	annual mean daily temperature
temp_sp	^0C	spring mean daily temperature
temp_su	^0C	summer mean daily temperature
temp_wi	^0C	winter mean daily temperature
rain_y	mm	annual rainfall
rain_sp	mm	spring rainfall
rain_su	mm	summer 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_wi	$MJ/m^2/d$	winter mean daily global solar radiation
aspect		hill slope aspect. e.g. N NE S SW NW, etc.
slope	degree	hill slope angle
G_animal	s, c,	grazing animal species: sheep (s), cattle(c).

Environmental data, including topographic variables (aspect, slope), edaphic variables (soil bulk density, pH, Olsen-P) and pasture management variables (N, P fertiliser application and grazing animal species) were measured or observed from the experiments from which the species abundance was measured. Climatic variables, including rainfall, temperature and global solar radiation, obtained from the National Institute of Water & Atmospheric Research (NIWA), New Zealand, were measured at a meteorology station located in the study area. There were a total of 23 climatic, topographic and management variables in the dataset (Table 1).

Since the rainfall data was measured from the horizontal surface and the P fertiliser was commonly aerially applied to hill-pasture with a uniform rate, the actual rainfall and P fertiliser received were adjusted by the cosine of the slope angle. Actual solar radiation on hilly terrain is strongly influenced by topography (Antoni 1998), so it was calculated from observed solar radiation at that location by an empirical method developed for the North Island, New Zealand (McAneney & Noble 1975).

Table 2. Functional groups of at Ballantrae [after Lambert *et al.* (1986) and Nicholas (1999)]

Functional group	Species
HFRG	*Lolium perenne, Holcus lanatus, Poa pratensis, Poa annua and Dactylis glomerata.*
LFTG	*Agrostis capillaris, Anthoxanthum odoratum, Cynosurus cristatus, Rytidosperma* spp., *Festuca rubra.*
Legume	*Trifolium repens, Trifolium dubium, Trifolium subterraneum and Lotus pedunculatus.*
Flatweeds	*Plantago lanceolata, Hypochaeris radicata, and Leontodon taraxacoides.*
Moss	*Musci* spp.

There were about twenty species existing in the naturalized hill pastures. I classified these species into functional groups primarily based on the commonly defined functional categories (grass, legume, forb), and further classified each of the grass and forb categories into two functional groups considered nutrient response and plant growth form, respectively, by synthesizing previous work (Lambert *et al.* 1983; Nicholas 1999). These five functional groups are: high fertility response grasses (HFRG), low fertility tolerance grasses (LFTG), legume, moss, and flatweeds. Table 2 lists these functional groups and the species allocated to them.

6.2.2.2. Decision tree development and assessment

I developed the decision tree models for the relative abundance of the five functional groups in SAS Enterprise Miner, Version 4.1 (SAS Institute Inc., 1999-2001, Cary, NC, USA). The main procedure for developing a decision tree model is to split the data of target variable based on its response to input variables. There are different criteria used to split the data, depending on the nature of the target variable. For interval variables (i.e. continuous variables) like species abundance, the split criterion is variance reduction or F test. I compared both criteria in developing the decision tree models for relative abundance of the functional groups and chose F test as the split criterion since this was statistically more reliable. A significance level of $P<0.05$ was selected for the F test in developing these decision trees.

Also, there is an interactive training procedure to develop a decision tree by arbitrarily forcing a variable into the decision tree if the variable meets the splitting criterion. I used this interactive training method to investigate the influence of specific variables on the functional groups.

Model assessment is an important step in developing the decision tree, as it enables a good model output, and prevents overtraining, which otherwise would develop a model that only fits the data from which it was trained (Witten & Erank 2000). The SAS Enterpriser Miner has several options to assess the model having an interval target. The measure of average squared error (ASE) is commonly used. ASE is similar to mean squared error (MSE) in a general linear model and is an indicator of model goodness-of-fit; the smaller the ASE, the better the model fits. In the development of the decision tree models, 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; and the "best" tree which had the smallest ASE from validation data was selected. I set six as the maximum tree levels, and 11 as the minimum observations required for a split search. The recommendation on minimum observations was derived using the program by assessing the size of the whole dataset (Fernandez 2003), in developing the decision tree models for species relative abundance of the functional groups.

6.2.2.3. GIS-based prediction and model validation

The outputs of the decision tree models for relative abundance of the functional groups were imported into a GIS to derive the "predictions" of relative abundance of HFRG, LFTG, legume, moss and flatweeds for the pasture described previously.

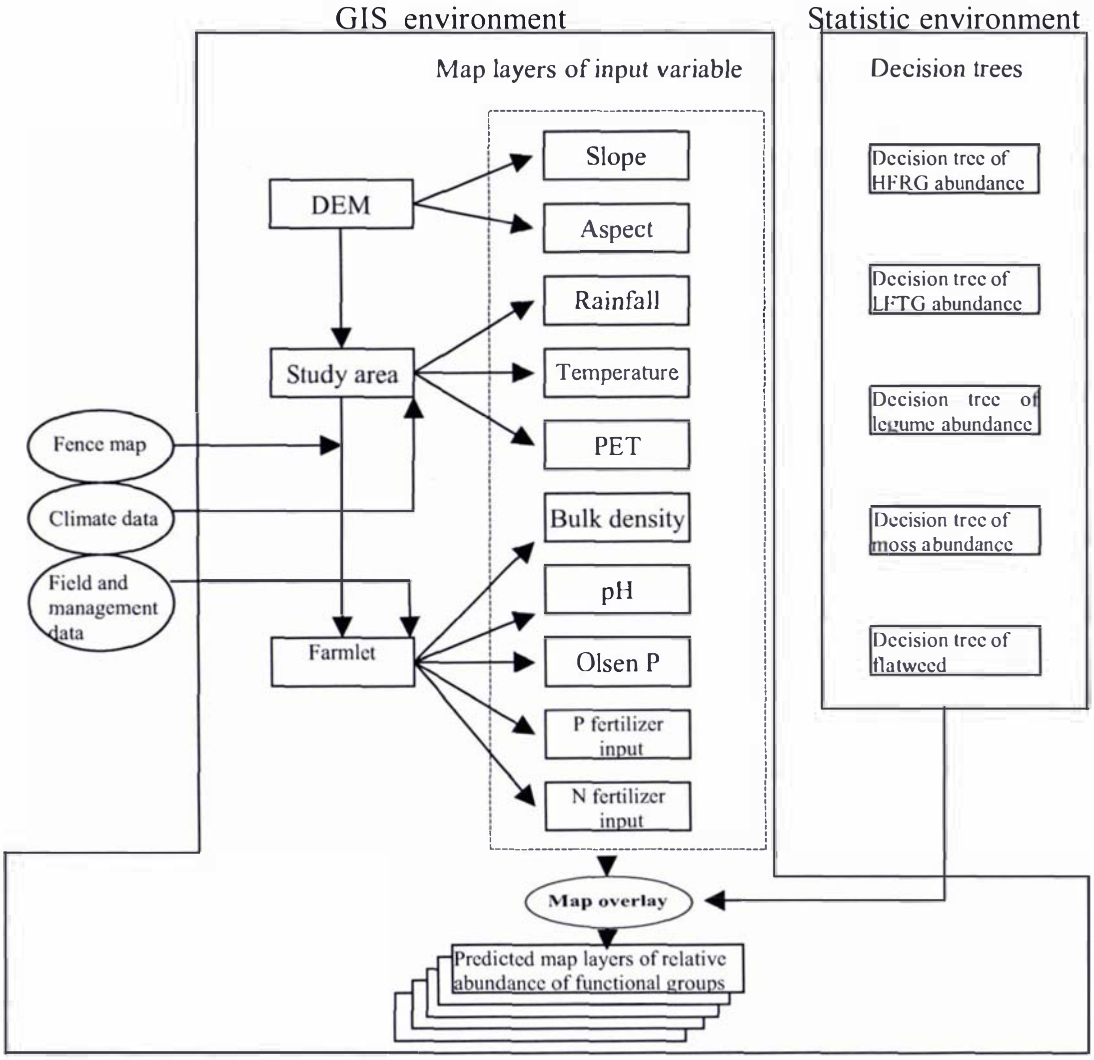


Fig. 2. Diagram of the procedures for generating the model predictions in GIS. The rectangles in the GIS environment area represent the GIS map layers. The arrows indicate the orders and the resources in developing the predictions of relative abundance of the functional groups in GIS. Names of input variables and their description are in Table 1.

A 5 × 5 m resolution digital elevation model (DEM) of this pasture was interpolated from a XYZ coordinate file produced by New Zealand Aerial Mapping Ltd. A map layer of the four farmlets was created based on an existing fence map. All the GIS map layers of input variables were then generated. Map layers of rainfall and P fertiliser input were adjusted by a cosine of the slope angle, and layers of solar radiation were adjusted by slope and aspect with methods previously mentioned. All the map layers had the same 5 × 5 m resolution.

In late spring 1993, the relative abundance of the functional groups was measured using aboveground biomass by cutting re-growth from a trimmed quadrat (0.5 m^2)

(Lambert et al. 1996). Three replicate quadrats, representing different combinations of slope and aspect in each paddock within each farmlet, were used. So there were a total of 24, 30, 24 and 30 observations in LL, LN, HN and HH, respectively. The GIS-based predictions were made based on the climatic and management variables in these four farmlets in 1993. The measured relative abundance was used to validate the model predictions. These data were independent from the data used in developing the decision tree models.

Spatial predictions for the relative abundance of the functional groups were implemented by coupling the outputs of the decision trees with GIS using a map overlay function. The prediction outputs were the map layers of relative abundance of HFRG, LFTG, legume, moss and flatweeds. Fig. 2 shows the procedures for generating these model predictions in GIS.

The average relative abundance of the functional groups in the four farmlets was then extracted by calculating the average of all the pixels (5×5m) within each farmlet. The predicted average abundance was compared with the observed average relative abundance for an empirical validation of the model (Mitchell 1997). The 95% confidence interval of the observed relative abundance mean was set as an acceptable error for the prediction. Because the map layers of predicted relative abundance covered an area which is larger than the four farmlets, only the area within the four farmlets was used to validate the predictions. As the observed relative abundance summed moss and flatweeds together, the sum of the predicted relative abundance of these two functional groups was used to validate their prediction. All the GIS analyses in this study were conducted in Idrisi 3.2 (Clark Labs, Clark University, Worcester, MA, USA).

6.3. Results

6.3.1. Decision trees

Figs 3, 4, 5, 6 & 7 display the decision tree models for relative abundance of functional groups HFRG, LFTG, flatweeds, legume and moss, respectively.

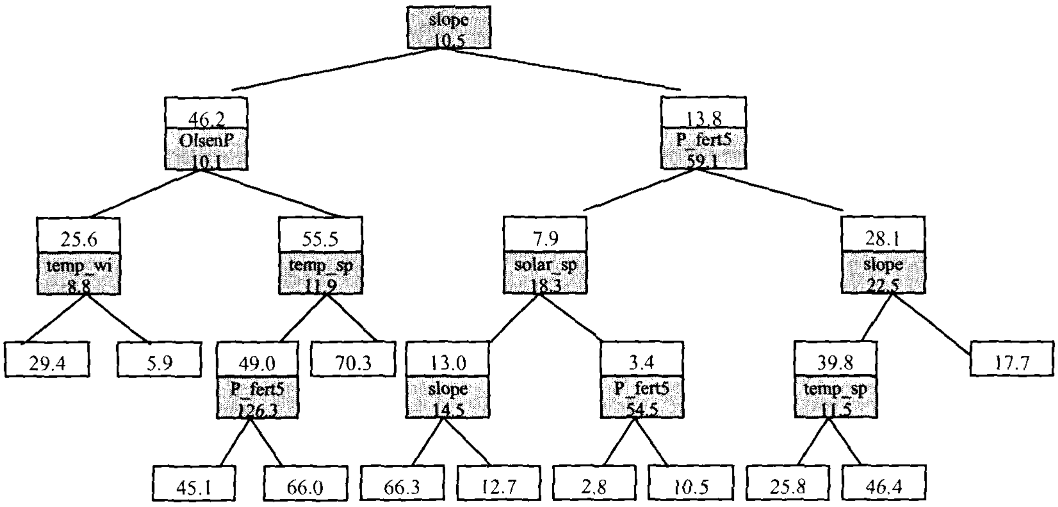


Fig. 3. The decision tree model for relative abundance of high fertility response grasses (HFRG). Predicted relative abundance (in percentage) is in the un-shaded rectangles, while splitting variables and split-points are in the shaded rectangles. Prediction goes to the left-side branch if the splitting variable is less than the split-point, and goes to the right-side branch if the splitting variable is equal to, or more than, the split-point. See Table 1 for variable symbols and unit descriptions.

Each decision tree is a hierarchical structure that contains rules of prediction. Relative abundance of a functional group was first split into two branches by a variable which best explained the variance. This splitting continued for each of the branches with the same method, and so on until no further splitting could be made. Prediction was made by a series of constraints defined by the input variables and their split-points. For example, in the decision tree model for relative abundance of HFRG (Fig. 3), when pasture had a slope less than 10.5^0, and an Olsen P equal to, or more than 10.1 μg/g and a mean spring temperature equal to, or more than 11.9 ^{0}C, the predicted relative abundance was 70.3%.

Slope, Olsen P, five-year cumulative P fertiliser input, spring and winter temperatures, and spring solar radiation were the significant variables influencing HFRG abundance. The lowest abundance of HFRG (2.8%) was found in pasture with high slope ($\geq 10.5^0$), high spring mean daily solar radiation ($\geq$ 18.3 $MJ/m^2/d$) and low five-year cumulative P fertiliser input ($<$ 54.5 kg/ha), while the highest abundance (70.3%) was found in pasture with low slope ($< 10.5^0$), high Olsen P ($\geq$ 10.1 μg/g) and high spring temperature ($\geq$ 11.9 ^{0}C) (Fig.3).

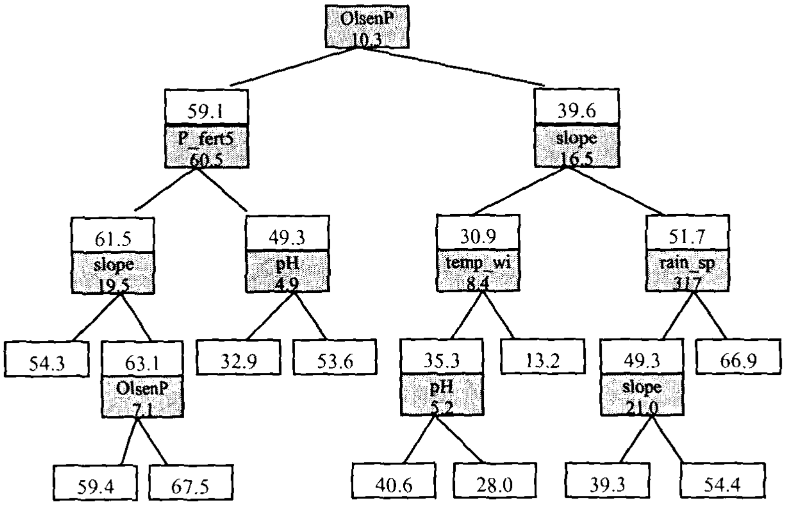


Fig. 4. The decision tree model for relative abundance of low fertility tolerance grasses (LFTG). See the caption of Fig. 3 for the description of decision tree interpretation.

The significant variables influencing LFTG were similar to those influencing HFRG. The lowest abundance (13.2%) was found in pasture with high Olsen P (≥ 10.3 μg/g), low slope (< 16.5^0) and high winter temperature (≥ 8.4 ^{0}C), while the highest abundance of LFTG (67.5%) was found in pasture with low Olsen P (< 10.3 μg/g), high slope (≥ 19.5^0) and low five-year cumulative P fertiliser input (< 60.5 kg/ha) (Fig. 4).

The flatweeds abundance range was from 0.7 to 9.8% with the highest abundance found in pasture with low annual P fertiliser input (< 9.7 kg/ha), sheep grazing and high winter rainfall (≥ 261 mm) (Fig. 5).

Soil bulk density, spring and winter mean daily solar radiation, slope, winter rainfall and five-year cumulative P fertiliser input were the significant variables influencing legume abundance. The highest legume abundance (30.5%) was found in pasture with high soil bulk density (≥ 0.98 g/cm^3), low spring mean daily solar radiation (< 19.5 MJ/m^2/d), medium five-year cumulative P fertiliser input (> 81.8 kg/ha but ≤ 1083 kg/ha), while the lowest legume abundance (2.5%) was found in pasture with low soil bulk density (< 0.98 g/cm^3), high winter mean daily solar radiation (≥ 10.5 MJ/m^2/d) and high slope (≥ 31.5^0) (Fig. 6).

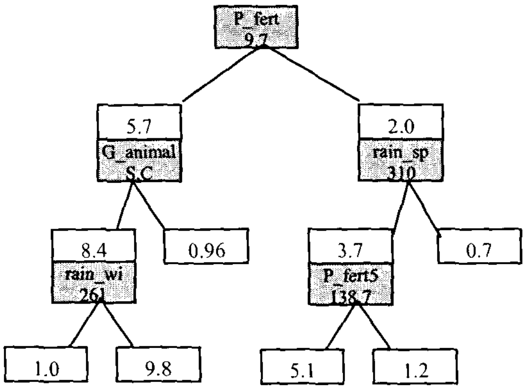


Fig. 5. The decision tree model for relative abundance of flatweeds. See the caption of Fig. 3 for the description of decision tree interpretation.

The lowest abundance of moss (2.3%) was found in pasture with high Olsen P (≥ 9.0 μg/g) and low slope ($< 28.5^0$), while the highest abundance (48.9%) was found in pasture with low Olsen P (< 8.0 μg/g), high winter mean daily solar radiation (≥ 1 0.9 MJ/m^2/d) and sheep grazing (Fig.7).

The relative importance of environmental and management variables in influencing the functional group abundance in a decision tree was indicated by the order they were selected in splitting the tree. The variable selected first was more influential than those selected after it. For LFTG and moss, Olsen P was the most significant variable influencing their relative abundance. Slope, soil bulk density and annual P fertiliser input were the most significant variables influencing HFRG, legume and flatweeds relative abundance.

In general, slope and Olsen P were the two key factors underlying the patterns of abundance for these five functional groups. Climatic factors, pH and grazing animal species also played important, but secondary roles in influencing their abundance. A summary of average responses of the five functional groups to Olsen P and slope is presented in Fig. 8, which was obtained by an interactive training of the decision tree when specifying Olsen P or slope as the splitting variable. An Olsen P of 10μg/g or a slope about 11^0 was found to be the critical point for LFTG and HFRG to change their dominance in pasture. LFTG and HFRG also showed a significant change in relative abundance at a slope of approximately 22^0.

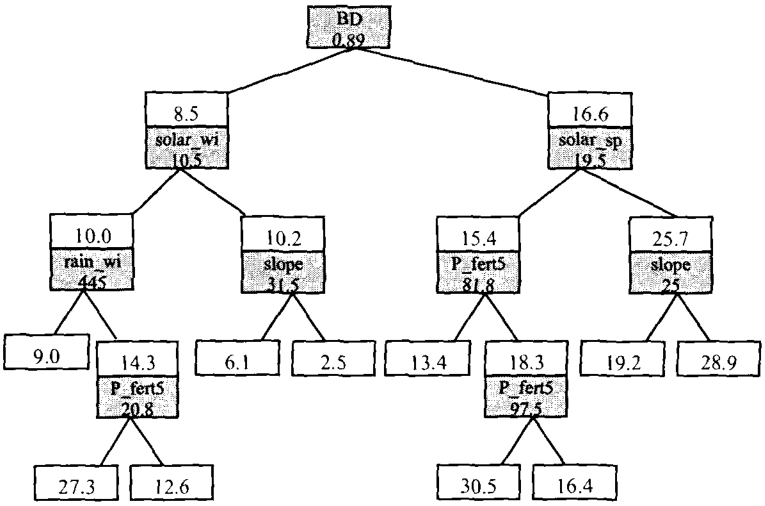


Fig. 6. The decision tree model for relative abundance of legume. See the caption of Fig. 3 for the description of decision tree interpretation.

6.3.2. GIS-based prediction and model validation

GIS-based predictions for relative abundance of HFRG, LFTG, legume, moss and flatweeds are shown in Fig. 9. The extracted average relative abundance of HFRG, LFTG, legume, moss and flatweeds in the four farmlets LL, LN, HN and HH were graphed against the observed average relative abundance of them in Fig. 10. The predictions for legume were in agreement with (i.e. prediction was within the 95% confidence interval of the observed relative abundance mean) the observations in all the four farmlets, while the predictions for HFRG and LFTG were in agreement with the observations in three of the four farmlets, and the predictions for moss and flatweeds were in agreement with the observations in two of the four farmlets. The overall predictive accuracy for the five functional groups in the four farmlets was 75% (12 of the 16 farmlets).

6.4. Discussion

The decision tree models performed very well as a predictive modelling approach in this study. The overall predictive accuracy of 75% is high considering the strict criterion used in the model validation (95% confidence interval of the observation mean as the accepted error). An application of the decision tree in predicting species distribution (Vayssieres *et al.* 2000) showed that the decision tree performed significantly better than a polynomial logistic regression model for four of the six cases considered, and as well as the regression model in the two remaining cases. 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 found that the decision tree had very high predictive ability.

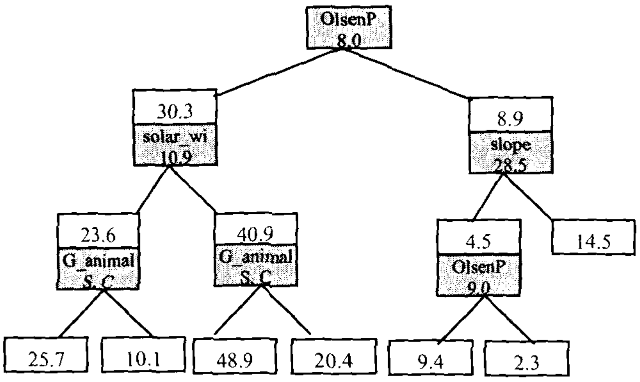


Fig. 7. The decision tree model for relative abundance of moss. See the caption of Fig. 3 for the description of decision tree interpretation.

The hierarchical structure of the decision trees clearly revealed the relative importance of environmental and management variables in influencing relative abundance of the functional groups. Topographic feature (slope) and soil fertility (Olsen P) were indicated as the most important factors driving the abundance of HFRG (Fig. 3), LFTG (Fig. 4) and moss (Fig. 7). With the increased soil Olsen P or decreased slope, pasture becomes more HFRG dominant; while with the decreased soil Olsen P or increased slope, pasture becomes more LFTG and moss dominant (Fig. 8). This general pattern of species dominance was consistent with those reported by Lambert *et al.*(1986) and Nicholas *et al.* (1998). Since soil Olsen P is positively influenced by P fertiliser input in pasture (Gillingham *et al.* 1998; Lambert *et al.* 2000), the amount of fertiliser input, therefore, is one of the key factors influencing the relative abundance of most functional groups in this study area. The effect of slope on species abundance may be from several influences: it may influence species abundance by redistributing and modifying the soil temperature, solar radiation received (McAneney & Noble 1975), and available soil moisture and fertility (Lopez 2000). However, as indicated by Lopez et al. (2000), hill slope was correlated with many factors such as soil total nitrogen, soil Olsen P, and soil water holding capacity in the hill-pasture. The responses of functional groups of plant species to hill slope, as well as to other environment and management factors are complex..

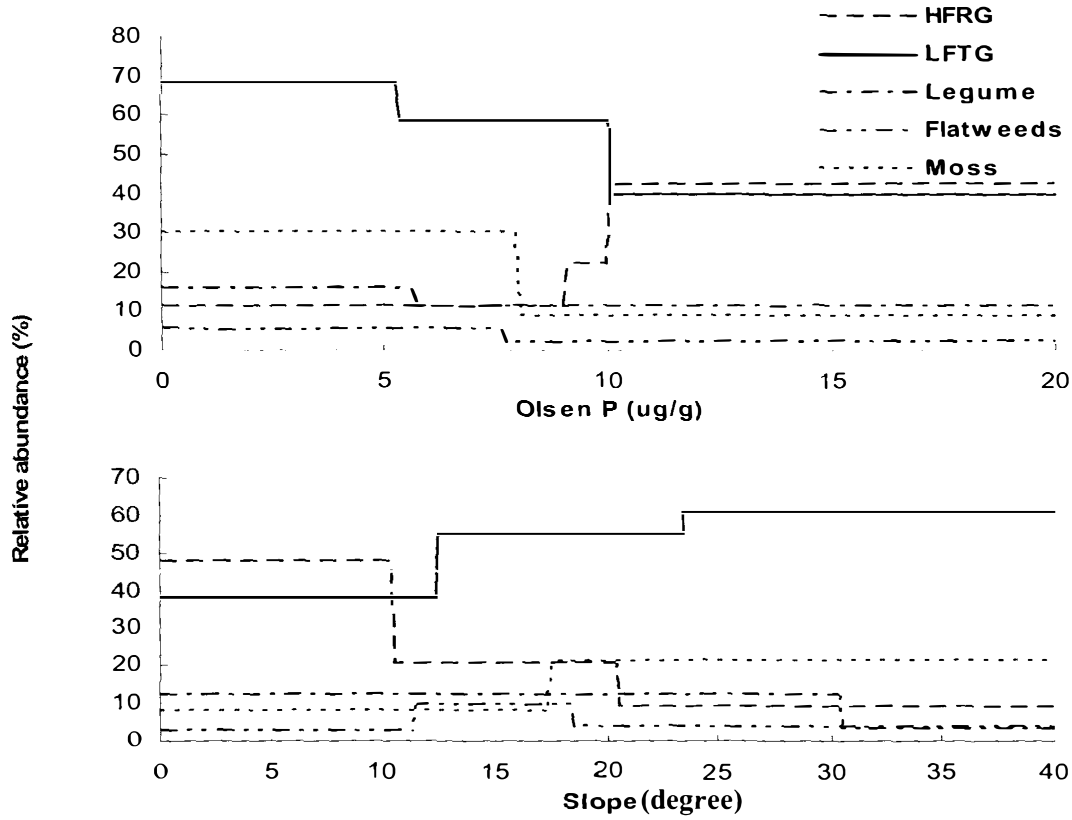


Fig. 8. The responses of functional groups HFRG, LFTG, legume, moss and flatweeds to Olsen P and slope gradients.

Soil bulk density was indicated as the most significant variable influencing legume abundance (Fig. 6); they were positively correlated with a threshold of 0.89 g/cm3. Lopez (2000) revealed that soil bulk density was positive correlated with hill slope and negative correlated with soil Olsen P, soil total nitrogen and soil moisture. However, the influence of soil bulk density on relative abundance of legume has not been previously reported and the interpretation is not clear. This result has an implication that the practice of applying P fertiliser in hill-pasture to encourage legume growth can be compromised in certain localities by the low bulk density in a pasture.

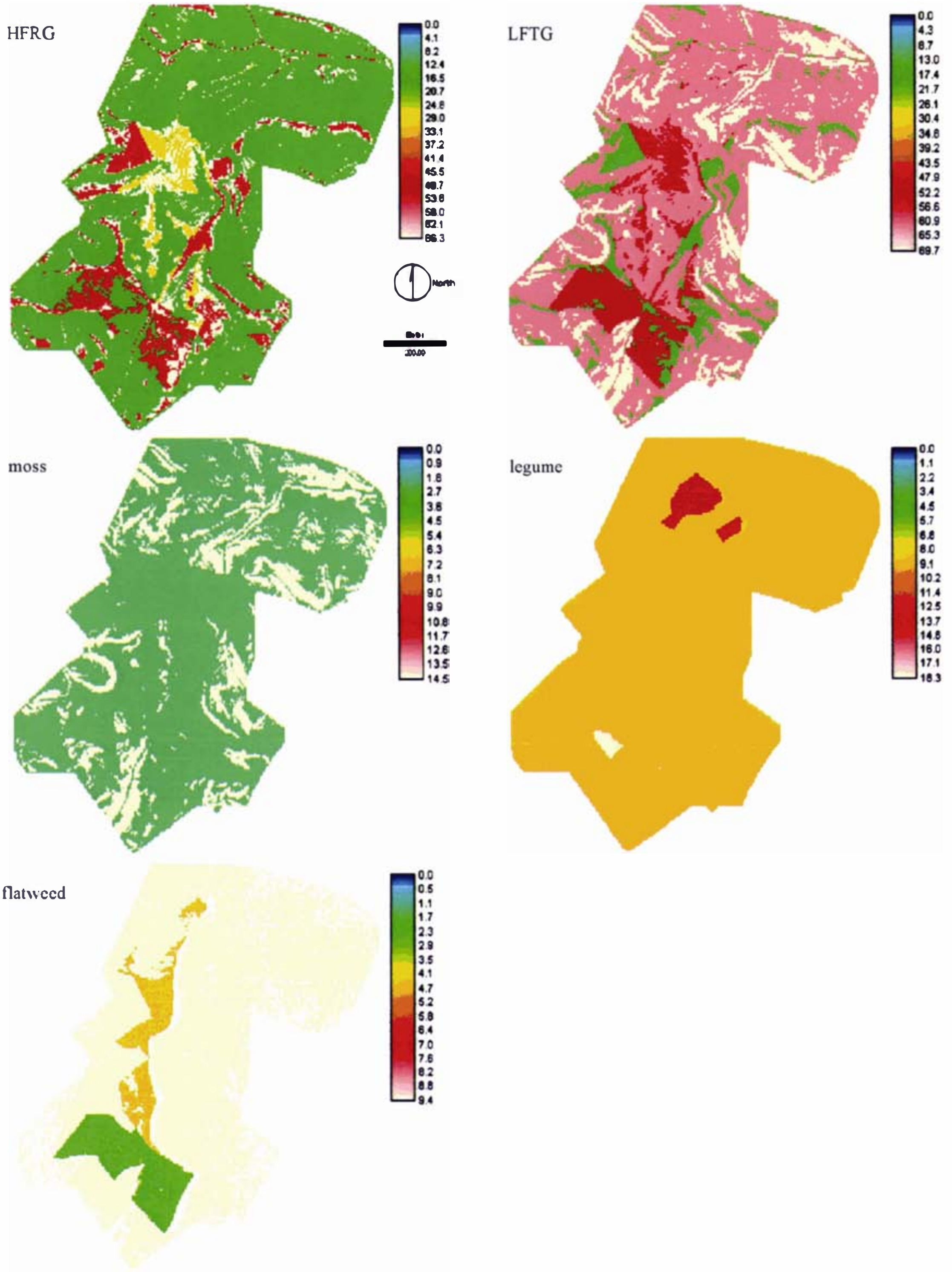


Fig. 9. Map layers of predicted relative abundance (%) for high fertility response grasses (HFRG), low fertility tolerance grasses (LFTG), legume, moss and flatweeds. Note the different scales on map legends.

Grazing animal species influence species abundance by changing the competition pattern in a pasture community through their selective defoliation (McNaughton 1985). HFRG, LFTG and legume showed no significant response to different grazing animal species. However, compared to sheep grazing, cattle grazing significantly reduced the relative abundance of flatweeds (Fig. 5) and moss (Fig 7), indicating that cattle grazing may be a method for weed control in pasture.

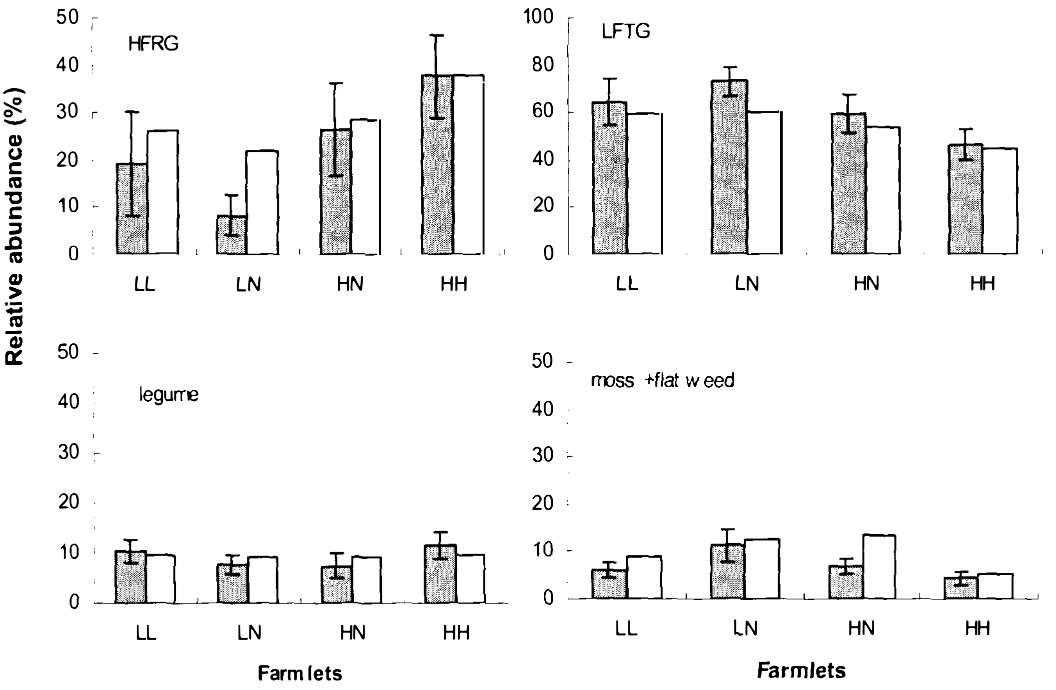


Fig. 10. The predicted☐ and the observed■ relative abundance of the functional groups in the farmlets LL, LN, HN and HH. Bars indicate the 95% confidence interval of the observed relative abundance means. Note the different scale on the y-axis for LFTG.

It was indicated from Fig. 9 that legume abundance in the predicted pasture had less variation than other functional groups. This may reflect the facts that the variation of soil bulk density, spring and winter solar radiation within the pasture was small, which were the most important factors explaining legume abundance.

One characteristic of the decision tree is that it assesses a threshold of an environmental variable that below or above which a species can be found (Moore *et al.* 1991), or the relative abundance of a species or functional group has a significant change. For example, when soil Olsen P was less than 10.0 μg/g, the average relative abundance of LFTG was 58.4% (Fig. 8), and when it was equal to, or more than 10.8 μg/g, the average relative abundance of LFTG was 39.6%. This pattern suggests that the 10.8 μg/g of soil Olsen P is a threshold at which the competition between LFTG and HFRG comes to an "equilibrium": below this point pasture is dominated by LFTG, while above this point, pasture becomes more HFRG dominant. However, the decision tree models for the abundance of functional groups cannot indicate the species response curve along an environmental gradient due to the "stair-step" response (Fig. 8). Response curves are important in testing an ecological hypothesis (Guisan & Zimmermann 2000; Vayssieres *et al.* 2000).

Modelling abundance of functional groups instead of abundance of individual species in this study simplified the investigation on the complex interrelationship between species and environment. The integration of modelling functional abundance and GIS can be used as an important tool in precision agriculture (Gillingham 2001). For example, hill-pasture management in New Zealand usually aims at increasing pasture production by applying P fertiliser. This modelling approach provides a platform to map HFRG abundance in a pasture over space, and thus could guide P fertiliser application to the areas that have more HFRG composition for better economic return.

One limitation of this modelling approach was that no interaction between species or species functional groups was incorporated in the model. This is because I intended to develop a predictive modelling approach only using easily obtained environmental and management variables to predict the functional group abundance for practical purposes. Inclusion of the interaction would make the prediction hard to implement. Besides, the interaction between species or functional group can be partially reflected in the model by environmental and management variables to some extent. For example, the competition between HFRG and LFTG can be reflected by Olsen P content: when Olsen P is low, the competition between them is relatively weak since LFTG is in dominance in a community; when Oslen P is high, the competition between them is strong since the community is co-dominated by both of them (Lambert *et al.* 1983; Lambert *et al.* 1986).

In summary, the decision tree was indicated as an appropriate modelling approach in predicting relative abundance for functional groups of plant species in the hill-pasture ecosystem. It had a high predictive accuracy and revealed the relative importance of environmental and management factors in influencing functional group abundance. By integrating with a GIS, the decision tree models can be used as decision support tools in pasture management such as in facilitating P fertiliser application for pastures with low HFRG composition by outlining the composition of HFRG over space. It was indicated from the decision tree models that the most important factors influencing the relative abundance of functional groups HFRG, LFTG, legume, flatweeds and moss were hill slope, soil Olsen P, soil bulk density, annual P fertiliser input and soil Olsen P, respectively. Generally, slope and soil Olsen P were the two key factors underlying the patterns of abundance for these five functional groups.

6.6. References

Antoni O. (1998) Modelling daily topographic solar radiation without site-specific hourly radiation data. *Ecological Modelling*, 113, 31-40

Austin M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101-118

Austin M.P., Nicholls A.O., Doherty M.D. & Meyer J.A. (1994) Determining species response functions to an environmental gradient by means of a B-function. *Journal of Vegetation Science*, 5, 215-228

Boer M. & Stafford Smith M. (2003) A plant functional approach to the prediction of changes in Australian rangeland vegetation under grazing and fire. *Journal of Vegetation Science*, 14, 333-344

Braun-Blanquet (1964) *Pflanzensoziologie.* 3rd ed. Springer, Wien

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) *Classification and regression trees.* The Wadsworth statistics/probability series. Chapman & Hall. Inc. New York, NY.

Brown D.G. (1994) Predicting vegetation types at treeline using topography and biophysical disturbance variables. *Journal of Vegetation Science*, 5, 642-656

Camargo J.A. (1995) On measuring species evenness and other associated parameters of community structure. *Oikos*, 74, 538-542

Chiarucci A., Wilson J.B., Anderson B.J. & De Dominicis V. (1999) Cover versus biomass as an estimate of species abundance: does it make a difference to the conclusion? *Journal of Vegetation Science*, 10, 35-42

Fels J.E. (1994) *Modeling and mapping potential vegetation using digital terrain data.* Ph.D. dissertation, College of Forest Resources, North Carolina State University

Fernandez G. (2003) *Data mining using SAS applications.* Chapman & Hall/CRC, Boca Raton, Florida

Fitzgerald R.W. & Lees B.G. (1992) The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. *Proceedings of the XVII Congress ISPRS. Bethesda, MD: American Society of Photogrammetry and Remote Sensing*, 570-573

Franklin J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499

Freund R.J. & Littell R.C. (1991) *SAS System for Regression* (second edition). *SAS Institute Inc., Cory, NC, USA, 94*

Gillingham A.G. (2001) Precision management of fertiliser application to pasture. *First Australian Geospatial Information and Agriculture Conference, Sydney, Australia, 17-19 July 2001*, 534-541

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

Gottfried M., Pauli H. & Grabherr G. (1998) Prediction of vegetation patterns at the limits of plant life: a new view of the alpine-nival ecotone. *Arctic and Alpine Research*, 30, 207–221

Guisan A., Edwards Jr T.C. & Hastie T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157, 89-100

Guisan A. & Harrell F.E. (2000) Ordinal response regression models in ecology. *Journal of Vegetation Science*, 11, 617-627

Guisan A., Weiss S.B. & Weiss A.D. (1999) GLMs versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143, 107-122

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

Gullison J.J. & Bourque C.P. (2001) Spatial prediction of tree and shrub succession in a small watershed in Northern Cape Breton Island, Nova Scotia, Canada. *Ecological Modelling*, 137, 181-199

Guo Q. & Rundel P.R. (1997) Measuring dominance and diversity in ecological communities: choosing the right variable. *Journal of Vegetation Science*, 8, 405-408

Hahel F.J. (1990) The hierarchical nature of community persistence: a problem of scale. *American Naturalist*, 136, 328-344

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

Korner C. (1994) Scaling from species to vegetation: The usefulness of functional groups. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and Ecosystem Function, Springer-Verlag*

Lambert M.G., Barker D.J., Mackay A.D. & Springett J.O. (1996) Biophysical indicators of sustainability of North Island hill pastures. *Proceedings of the New Zealand Grassland Association*, 57, 31-36

Lambert M.G., Clark D.A., Grant D.A. & Costall D.A. (1986) Influence of fertilizer and grazing management on North Island moist hill country 2. Pasture botanical composition. *New Zealand Journal of Agricultural Research*, 29, 1-10

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

Lambert M.G., Clark D.A., Mackay A.D. & Costall D.A. (2000) Effects of fertiliser application on nutrient status and organic matter content of hill soils. *New Zealand Journal of Agricultural Research*, 43, 127-138

Lambert M.G. & Roberts E. (1978) Aspect differences in an unimproved hill country pasture 2. Edaphic and biotic difference. *New Zealand Journal of Agricultural Research*, 21, 255-260

Lehmann A. (1998) GIS modeling of submerged macrophyte distribution using generalized additive models. *Plant Ecology*, 139, 113-124

Lehmann A., McC. Overton J. & Leathwick J.R. (2003) GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 160, 165-183

Lopez I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

May R.M. (1975) Ecology and evolution of communities. In: Cody, M.L., and Diamond, J.M. (Eds.), *Ecology and Evolution of Communities.* Harvard University Press. Cambridge, Massachusetts, USA

McAneney K.J. & Noble P.E. (1975) Estimating solar radiation on sloping surfaces. *New Zealand Journal of Experimental Agriculture*, 4, 195-202

McNaughton S.J. (1985) Ecology of grazing ecosystem: the Serengeti. *Ecological Monograph*, 53, 291-320

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

Moore D.E., Lees B.G. & Davey S.M. (1991) A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Journal of Environmental Management*, 15, 59-71

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Nicholas P.K., Kemp P.D. & Barker D.J. (1998) Biodiversity, stability and pasture management - the role of functional groups. *Proceedings of the 9th Australian Agronomy Conference, Wagga wagga, 1998*

Ohmann J.L. & Spiess T.A. (1998) Regional gradient analysis and spatial pattern of wood plant communities. *Ecological Monograph*, 68, 151-182

Palmer A.R. & Van Stden J.M. (1992) Predicting the distribution of plant communities using annual rainfall and elevation: an example from southern Africa. *Journal of Vegetation Science*, 3, 261-266

Rouget M., Richardson D.M., Lavorel S., Vayreda J., Gracia C. & Milton S.J. (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science*, 12, 491-502

Stocks C.E. & Wise S. (2000) The role of GIS in environmental modelling. *Geographical and Environmental modelling*, 4, 219-235

Tappeiner U., Tasser E. & Tappeiner G. (1998) Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an Alpine area. *Ecological Modelling*, 113, 225-237

ter Braak C.J.F. (1985) CANOCO: *A FORTRAN program for canonical correspondence analysis and detrended correspondence analysis.* IWI-TNO, Wangeningen, The Netherlands

Tilman D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, 77, 350-363

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

Vitousek P.M. & Hooper D.U. (1994) Biological diversity and terrestrial ecosystem biogeochemistry. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and ecosystem function*, Springer-Verlag, 3-14

Wadsworth R. & Treweek J. (1999) *GIS for ecology: an introduction.* Harlow: Addison Wesley Longman

Welson J.B., Wells T.C.E., Trueman I.C., Jones G., Atkinson M.D., Crawley M.J., Dodd M.E. & Silverton J. (1996) Are there assembly rules for plant species abundance: An investigation in relation to soil resources and successional trends. *Journal of Ecology*, 84, 527-538

Whittaker R.H. (1965) Dominance and diversity in land plant communities. *Science*, 147, 250-260

Witten I. & Erank E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers San Francisco, California

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

Zimmermann N.E. & Kienast F. (1999) Predictive mapping of alpine grassland in Switzerland: Species versus community approach. *Journal of Vegetation Science*, 10, 469-482

Chapter 7. Modelling and Investigating Species Richness Patterns and Underlying Factors in a Pasture Ecosystem

A combination of decision tree, regression and correlation analyses was used in this chapter to detect the patterns of species richness, and the possible driver underlying the patterns in a hill-pasture ecosystem.

A paper derived from this chapter with a title "Modelling species richness patterns in a naturalised pasture ecosystem" by B. Zhang, I. Valentine & P.D. Kemp has been submitted to ***Journal of Applied Ecology*** *for publication.*

Abstract. Species richness patterns and underlying factors have long been an interest of ecological studies. The mechanisms that regulate species richness in an ecosystem are not yet settled. Decision tree model and other analyses were applied to investigate factors that influenced species richness and responses of species richness to these factors in a naturalised pasture ecosystem. The results indicated that legume abundance played the most important role in influencing species richness. Species richness increased with the increase of legume abundance and showed a tendency of hump-shaped response. The effect of legume abundance on species richness was probably a result of species complementarity resulting from nitrogen fixation of legume species. Grazing animal species also had significant influence on species richness. Pasture grazed by sheep had more species than pasture grazed by cattle probably because sheep grazed lower and more selectively than cattle, and reduced the height of the competitively dominant species, thereby allowing more species to co-exist. The effects of aboveground biomass, soil fertility, climate and pasture management on species richness were also investigated.

7.1. Introduction

Species richness patterns and underlying factors have long been an interest in ecological studies (Grime 1973; Abrams 1995; Zobel 1997; Waide *et al.* 1999; Chase & Leibold 2002). Species richness is often observed as a hump-shaped or unimodal response along natural gradients of increasing biomass or resources (Grime 1973; Hastings 1980; Tilman 1996), while linear (positive or negative) and non-significant responses are also commonly observed (Brown 1973; Goldberg & Miller 1990; Gough *et al.* 1994; Waide *et al.* 1999).

Explanation of the hump-shaped relationship between species richness and resources (or biomass) tends to focus on species interaction, especially competition as the driving factor (Grime 1973; Rajaniemi 2002). When a community is in a low resource environment (indicated by low biomass), species richness is generally low because it is assumed that only few species can survive. With the increased availability of resources, species richness increases as the environment can support more species. A high resource environment (high biomass) favours species that outcompete other species, and eliminate less competitive species from the community (Grime 1973; Rosenzweig 1992; Huston & DeAngelis 1994).

There are also other explanations of the hump-shaped relationship between species richness and resources (or biomass). Species pool models (Zobel 1997; Grace 2001) suggest that the decrease in diversity of the environment with high availability of resources, or biomass, may not be a consequence of the increased competition, but rather of the decreased size of the species pool that may be able to survive in the enriched environment.

The responses of species richness are scale-dependent (Gross *et al.* 2000). Chase and Leibold (2002) found that at the local scale, species richness has a hump-shaped relationship with productivity, while at the regional scale it is a positive linear relationship. Species richness is also related to environmental heterogeneity (e.g. those resulted from micro-topography, soil fertility and climatic variables) (Huston 1994; Bell *et al.* 2000; Lundholm & Larson 2003), with a negative or positive relationship being found when samples are within a homogenous habitat and a hump-shaped relationship being found when samples encompasses different habitat types (Guo & Berry 1998). Oksanen (1996) claims that the hump-shaped relationship between diversity and biomass is just an artefact due to the fixed small plot size; as plants at high biomass are bigger, and therefore there are less species in a plot.

There are many other factors also playing a role in regulating species richness. Environmental factors, such as rainfall, pH and salinity can strongly influence species richness (Garcia *et al.* 1993; Gunnarsson *et al.* 2000; Knapp *et al.* 2002). Their influence may work indirectly by influencing biomass or resource availability, or directly when they do not affect biomass (Maranon & Garcia 1997).

Grazing has a marked influence on species richness (Belsky 1992; Oiff & Ritchie 1998). Grazing animals are generally thought to enhance species richness by their direct consumption of competitively dominant species and indirect effects on plant competition (McNaughton 1985; Belsky 1992). However, some studies suggest that the effect of grazing animal depends on the environmental conditions; species richness decreases with heavy grazing in nutrient-poor ecosystems, while it increases with heavy grazing in nutrient-rich ecosystems (Proulx & Mazumder 1998).

Ecosystems are complex. Factors that influence biomass and resource availability may influence species richness. On the other hand, factors that influence species richness may also influence biomass. To investigate the key factors influencing species richness in an ecosystem and the underlying mechanisms, a simultaneous assessment of these factors is required. In the study reported here, I used decision tree, a data mining approach (Thuraisingham 1999) and other analyses, to investigate the patterns of species richness and the underlying factors in a naturalised pasture ecosystem in the North Island, New Zealand. Decision tree has been widely used in many scientific areas as a modelling approach and has shown a strong ability in detecting the relative importance of input (independent) variables and their interactions on model target (dependent variable) (Iverson & Prasad 1998; Vayssieres *et al.* 2000; Yang *et al.* 2003). Topographic features, soil properties, climatic factors, pasture management factors and relative abundance of dominant plant species and functional groups were included as input variables in this analysis. My aims were to investigate the important factors that influence species richness and its response patterns to these factors, and to explore the mechanisms regulating these response patterns.

7.2. Methods

7.2.1. Study area

This study was on a pasture at Ballantrae, a hill-pasture research station of AgResearch, New Zealand, which is located at latitude S40°18′, longitude E175°50′ in the North Island with an average annual rainfall 1270 mm and an average annual daily temperature 12.3 °C. Altitude ranges from 125 to 350 m a.s.l.. The pasture was about 200 ha in area with ten farmlets differing in phosphorus fertiliser input and grazing animal species (cattle, sheep) since 1970s. Topographic features at Ballantrae are dissected and very heterogeneous (Lambert & Roberts 1978). The pasture was dominated by browntop (*Agrostis capillaris*) and ryegrass (*Lolium perenne*), while white clover (*Trifolium repens*) was the main legume species in the pasture (Nicholas 1999). Appendix 1. provides a list of the species in this study area.

7.2.2. Model development

7.2.2.1. Dataset

Data were from a long-term experiment (see Lambert *et al.* 1986), unpublished Ph.D. theses and some other experiments conducted at Ballantrae over the period of 1972-1998. Samples were collected during late spring to early summer (October to early December) using a 0.5 m^2 quadrat with a stratified random sampling method. There are 700 samples in this dataset Aboveground biomass was measured by a "trim technique" involving harvesting pasture regrowth from trimmed areas protected from sheep grazing (Lambert *et al.* 1996) with a regrowth period from 30-50 days. Biomass was sorted by species, and was standardised for per quadrat per 30 days. Species richness was the species number within a quadrat. Relative abundance of dominant species and functional groups were calculated as the percentage of their biomass out of the total biomass within a quadrat. These species included ryegrass, browntop and white clover. Functional groups of plant species included high fertility response grasses (HFRG), low fertility tolerance grasses (LFTG), legume, moss, and flatweeds.

Environmental data, including topographic features (aspect, slope) and soil properties (bulk density, pH, soil Olsen P, soil total N), and pasture management (phosphorus (P) and nitrogen (N) fertiliser input and grazing animal species (cattle or sheep)) were measured or recoded from the experiments from which the species richness and aboveground biomass were measured. Climatic variables, including rainfall, temperature and global solar radiation, were obtained from the National Institute of Water & Atmospheric Research (NIWA), New Zealand.

All the environmental, pasture management and vegetation variables involved in this study are listed in Table 1.

7.2.2.2. Correlation analysis

Pearson correlations between species richness and all environmental, management and vegetation variables were analysed in SAS 8.2 (SAS Institute Inc., 1999-2001, Cary, NC, USA). Partial correlation between species richness and each variable that had a significant ($P < 0.05$) correlation with species richness were also analysed, using each of the soil Olsen P, aboveground biomass, legume abundance, total N, slope or a combination of some of them as partial variables. This partial correlation analysis intended to detect whether a variable had an independent correlation with species richness when taking away the effect resulting from correlation of this variable with others.

Table 1. Environmental, pasture management and vegetation variables used in model analyses

Variable symbol	Unit	Variable description
pH	$-\log_{10}[H^+]$	soil pH
BD	g/cm^3	soil bulk density
total_N	%	total nitrogen in soil
OlsenP	µg/g	soil Olsen P
N_fert	kg/ha/y	Annual elemental N fertiliser input
P_fert	kg/ha/y	annual elemental P fertiliser input
P_fert5	kg/ha	5-years cumulative elemental P fertiliser input
P_fert10	kg/ha	10-years cumulative elemental P fertiliser input
temp_y	^0C	annual mean daily temperature
temp_sp	^0C	spring mean daily temperature
temp_su	^0C	summer mean daily temperature
temp_wi	^0C	winter mean daily temperature
rain_y	mm	annual rainfall
rain_sp	mm	spring rainfall
rain_su	mm	summer 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_wi	$MJ/m^2/d$	winter mean daily global solar radiation
aspect		hill slope aspect. e.g. N NE S SW NW, etc.
slope	degree	hill slope angle
animal	s, c	grazing animal species. "s" sheep, "c" cattle
biomass	g	aboveground biomass/$0.5m^2$/30 growing days
abundance	%	relative abundance of dominant species and functional groups
species richness		species number in a quadrat (/$0.5m^2$)

7.2.2.3. Decision tree development and assessment.

I first developed a decision tree model for species richness using all available environmental, vegetation and management variables as input variables to detect the most important variables and their relative importance in influencing species richness.

As some of the environmental, management and vegetation variables are highly correlated, I conducted a principal component analysis for environmental and management variables, and legume species abundances which were detected having significant correlation with species richness, to reduce the directions of these variables and then used the first ten components which explained about 90% of total variance of the variables to develop a principal components based decision tree model for species richness. This model aimed to detect the effect of group variables on species richness. Table 2 shows the eigenvectors of the first ten principal components.

Table 2. Eigenvectors (Corr) for the first ten principal components. Numbers in bold font indicate the main contributions of the variables to the principal components.

Variable	PCR1	PCR2	PCR3	PCR4	PCR5	PCR6	PCR7	PCR8	PCR9	PCR10
pH	0.3898	-0.0152	0.4433	-0.2372	0.4268	0.2688	0.0319	-0.0197	0.1111	-0.0353
BD	-0.2574	-0.3939	-0.0811	0.4199	0.2471	0.3715	-0.0078	0.3931	0.1667	0.2576
totalN	0.3065	0.3074	0.3183	**-0.6191**	-0.0771	-0.311	-0.0474	0.0646	0.0858	0.0913
OlsenP	0.3342	0.026	0.452	-0.3973	0.2617	0.0523	-0.0593	0.142	0.2007	0.2092
P_fert	0.2339	-0.1783	0.4505	-0.0991	**0.6258**	0.3587	0.0228	-0.1644	-0.1204	-0.1322
P_fert10	0.142	**-0.6324**	**0.7004**	-0.1145	0.1229	0.05	-0.0178	-0.0977	-0.046	-0.075
P_fert5	0.142	**-0.6324**	**0.7004**	-0.1145	0.1229	0.05	-0.0178	-0.0977	-0.046	-0.075
temp_y	**-0.8506**	0.2217	0.0418	-0.3864	0.2168	0.0605	0.035	0.0332	-0.0365	-0.0178
temp_sp	**-0.6087**	0.4271	-0.0766	-0.3361	0.3629	0.2539	0.0408	-0.0516	-0.0718	-0.0823
temp_su	**-0.8889**	0.1618	0.0014	-0.3447	0.107	-0.0907	0.0315	0.122	0.0051	0.0543
temp_wi	0.6731	-0.4697	0.2961	0.2388	-0.2006	0.0218	-0.042	-0.1489	-0.0161	-0.076
rain_y	0.5086	**0.7176**	-0.3712	-0.0679	0.226	0.1052	0.0286	-0.0204	-0.0179	-0.0025
rain_sp	**0.8996**	0.2116	0.1924	-0.1818	-0.0975	-0.1619	-0.0063	-0.0858	-0.0073	-0.019
rain_su	**0.8001**	0.4109	-0.3338	0.2025	0.0894	0.1127	0.0011	-0.0457	0.003	-0.004
rain_wi	0.3144	**0.7406**	-0.3636	-0.1043	0.3391	0.2478	0.0355	-0.0696	-0.0509	-0.0546
rain_warm	**0.8776**	0.3872	-0.2221	0.1182	0.0479	0.0504	-0.0007	-0.0589	0.0006	-0.0081
solar_y	-0.1833	0.5821	**0.6245**	0.4172	-0.1811	0.0867	0.0201	0.0278	-0.0143	0.0202
solar_sp	-0.0088	**0.6495**	**0.6385**	0.3432	-0.1495	0.1032	0.0303	-0.0117	-0.0309	-0.0092
solar_su	-0.5668	0.4736	0.4717	-0.3903	-0.0304	-0.1706	0.059	0.0688	-0.0197	0.0249
solar_wi	-0.1972	0.5497	0.594	0.4795	-0.1687	0.1264	0.0118	0.0212	-0.0149	0.0176
aspect_A	-0.1561	0.4723	0.5292	0.4596	-0.2256	0.1116	0.0715	0.0553	0.0094	0.0027
slope	-0.4177	-0.3457	-0.3706	**0.6068**	0.0796	0.345	-0.0162	-0.0029	-0.0059	-0.0336
Yield	0.5338	-0.0207	0.1325	0.0893	0.0512	0.1063	-0.062	0.3796	0.2341	0.2987
Trp	0.2554	0.0042	0.0652	0.314	**0.5025**	**-0.5144**	-0.0538	0.1697	-0.2345	0.24
Tpr	0.0811	0.008	0.0667	0.3503	0.2573	-0.2812	0.1299	**0.4491**	-0.0416	**-0.6326**
Tdu	-0.4895	0.0891	0.0383	0.398	0.1834	-0.0694	-0.2383	**-0.4308**	-0.0736	0.1817
Tsu	-0.2359	0.2133	-0.0156	0.0763	0.0998	-0.0997	**-0.6078**	-0.0939	**0.6074**	-0.2662
Lotus	-0.115	-0.072	-0.0295	0.207	0.1546	-0.2028	**0.716**	-0.2715	0.529	0.031
Legume	-0.1658	0.0704	0.075	**0.5921**	**0.5569**	**-0.5095**	-0.0776	-0.1201	-0.0312	0.1079

In order to assess variables which had an independent significant effect on species richness (the variables which had a significant (P <0.05) correlation with species richness both in correlation and partial correlation analyses), or were considered as the important factors (biomass and P fertiliser), I developed decision tree models for species richness only using each of these variables as the input variable to investigate the responses of species richness to these variables.

Decision tree models were developed in SAS Enterprise Miner, Version 4.1 (SAS Institute Inc., 1999-2001, Cary, NC, USA). The main procedure for developing a decision tree model is to split the data of target variable based on its response to input variables. There are different criteria used to split the data, depending on the nature of the target variable. For interval variables (i.e. continuous variable) like species richness, the split criterion is variance reduction or F test. In the case of variance reduction, a complete search is applied to all the input variables and the possible split-points to select one variable that gives a maximum amount of variance reduction for the target data by splitting the data into two groups (that is to select a variable that ultimately explains the variance of the target variable), and the same search is applied to each of the sub-groups and so on recursively. For the criterion of F test, a variable which gives the most significant P value in the F test is selected instead of giving the maximum amount of variance reduction in the case of variance reduction (SAS Online Help: Getting Started with Enterpriser Miner Software). I chose F test as the split criterion since this criterion was statistically more reliable. A significant level of P <0.05 was selected for F test in developing these decision trees.

Model assessment is an important step in developing the decision tree, as it enables a good model output, and prevents overtraining, which otherwise would develop a model that only fits the data from which it was trained (Witten & Erank 2000). SAS Enterpriser Miner has several options to assess the model having an interval target. The measure of average squared error (ASE) is commonly used. ASE is similar to mean squared error (MSE) in a general linear model and is an indicator of model goodness-of-fit; the smaller the ASE, the better the model fits. In developing the decision trees, the whole dataset was randomly partitioned into two: the training data (70% of the total) and the validation data (30% of the total). The decision trees were trained by the training data and assessed by the validation data; a "best" tree with the smallest ASE from the validation data was selected.

I set six as the maximum tree levels and 11 observations as the minimum observations required for a split search. The recommendation on minimum observations was derived using the program by assessing the size of the whole dataset (Fernandez 2003).

7.3. Results

7.3.1. Correlation

Table 3 shows Pearson correlation coefficients and *P* values between species richness and the variables which had a significant (*P* <0.05) correlation with species richness, and partial correlation coefficients and *P* values between species richness and these variables when each of soil Olsen P, aboveground biomass, slope or a combination of them were used as partial variables.

Table 3. Pearson correlation coefficients and significance among input variables that had a significant (P <0.05) correlation with species richness, and the partial correlation coefficient and significance between these variables and species richness when each of the soil Olsen P, aboveground biomass, legume abundance, total N or a combination of them were used as partial variables. See Table 1 for variable descriptions.

Correlation

	biomass	OlsenP	slope	legume	P_fert5	P_fert	pH	rain_sp	solar_sp	total_N
species richness	-0.1152	-0.2718	0.1980	0.2833	-0.1560	-0.0775	-0.1513	-0.1677	-0.0986	-0.2283
	0.0026	*<.0001*	*<.0001*	*<.0001*	*< 0001*	*0.0435*	*<.0001*	*<.0001*	*0.0101*	*<.0001*

Partial correlation

	biomass	OlsenP	slope	legume	P_fert5	P_fert	pH	rain_sp	solar_sp	total_N
species richness	-0.0569	-0.2539	0.0723	0.2446	-0.0597	-0.0162	-0.0122	-0.0510	-0.0671	-0.1572
	0.1392	***<0.0001***	*0.0599*	***<0.0001***	*0.1202*	*0.6745*	*0.7532*	*0.1847*	*0.0810*	***<0.0001***
partial variable	Olsen P	biomass	Olsen P biomass	Olsen P biomass slope	Olsen P	Olsen P	Olsen P	biomass	biomass	biomass

Aboveground biomass, soil Olsen P, soil total N, slope, legume abundance, soil pH, five-year cumulative P fertiliser input and spring rainfall had a very significant (*P* <0.01) correlation with species richness, while annual P fertiliser input and spring daily mean solar radiation had a significant (*P* <0.05) correlation with species richness. However, only soil Olsen P, soil total N and legume abundance had a significant (*P* <0.05) partial correlation with species richness. For aboveground biomass, slope, five-year cumulative P fertiliser input, annual P fertiliser input, pH, spring rainfall and spring solar radiation, their correlation with species richness were not significant when taking off the effects of partial variables.

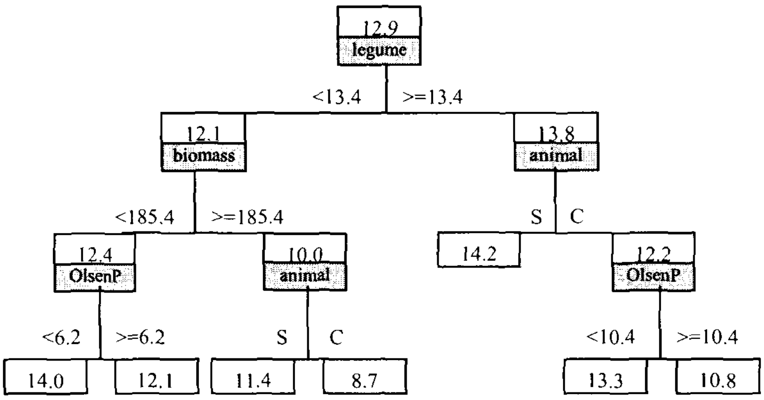


Fig. 1. The decision tree model for species richness using the environmental, management and vegetation variables. Predicted species richness is in the un-shaded rectangles, splitting variables are in the shaded rectangles. Prediction goes to the left-side branch if the splitting variable is less than the split-point, and goes to the right-side branch if the splitting variable is equal to, or more than, the split-point. See Table 1 for variable and unit descriptions.

7.3.2. Decision trees

Fig. 1 displays the decision tree model for species richness using all available environmental, management and vegetation variables. The variables that were selected in the decision tree were significant (*P* <0.05, F test) in explaining the variance in species richness, and the variable first selected was more influential on species richness than the variables selected after it. Relative abundance of legume species was the most significant variable in explaining the variance in species richness, while aboveground biomass, grazing animal species and soil Olsen P were also significant in explaining the variance in species richness. Legume abundance showed a positive effect on species richness, while aboveground biomass and soil Olsen P showed a negative relationship with species richness. Sheep grazed pasture had more species than cattle grazed pasture.

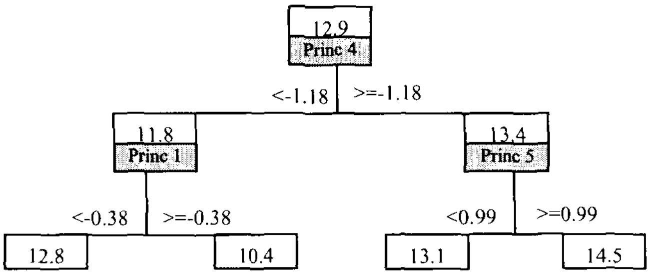


Fig. 2. The decision tree model for species richness using principal components as input variables. Princ 1, Princ 4 and Princ 5 represent principal components 1, 4 and 5, respectively. The splitting points are values of principal components. Princ 1 was mainly composed of annual, spring and summer temperature (all negative), and spring and summer rainfall. Princ 4 was mainly composed of total N (negative), slope and legume abundance. Princ 5 was mainly composed of annual P fertiliser input and legume abundance. See the caption of Fig. 1 for the description of decision tree interpretation.

A decision tree made prediction on the model target by using a series of constraints defined by the input variables and their split-points. For example, when pasture with a legume relative abundance less than 13.4%, and an aboveground biomass less than 185.4 g/0.5m^2 and a soil Olsen P less than 6.2 μg/g, the predicted average species richness was 14.0 (Fig. 1). A decision tree also shows the responses of model target to the input variables. For example, when legume abundance was equal to, or more than, 13.8%, species richness responded to the grazing animal species, when legume abundance was less than 13.8%, species richness responded to aboveground biomass. These different responses of species richness to input variables could indicate the interaction among the input variables.

Fig. 2 shows the decision tree model for species richness using the principal components as the input variables. For the first ten principal components, only the first, fourth and fifth principal components were selected by decision tree as the significant (P <0.05) variable in explaining the variance in species richness. The first principal component was mainly composed of annual, spring and summer temperature (all negative), and spring and summer rainfall. The fourth principal component was mainly composed of total N (negative), slope and legume abundance. The fifth principal component was mainly composed of annual P fertiliser input and legume abundance. It was indicated from this decision tree that soil fertility and legume abundance were the main factors influencing species richness, while climatic factors, mainly temperature and rainfall also played a role in influencing species richness with higher temperature increasing while higher rainfall decreasing species richness.

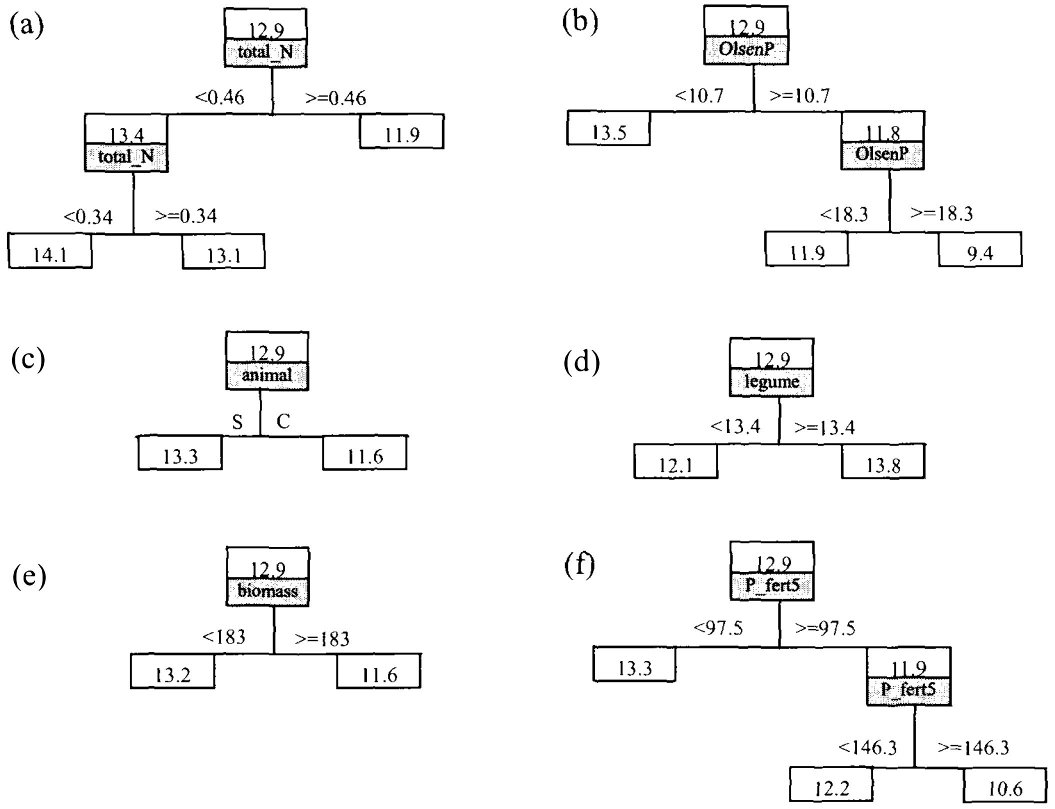


Fig. 3. The decision tree models for species richness using total nitrogen (a), Olsen P (b), grazing animal species (c), legume abundance (d), aboveground biomass (e) and five-year cumulative P fertiliser input (f) as input variables, respectively. See the caption of Fig. 1 for the description of decision tree interpretation.

Fig. 3 shows the decision tree models for species richness only using each of the soil total N, soil Olsen P, grazing animal species, legume abundance, aboveground biomass and five-year cumulative P fertiliser input as input variable. The decision trees show species richness had a negative response to soil total N, while the responses of species richness to the other four variables were in agreement with those indicated in Fig. 2.

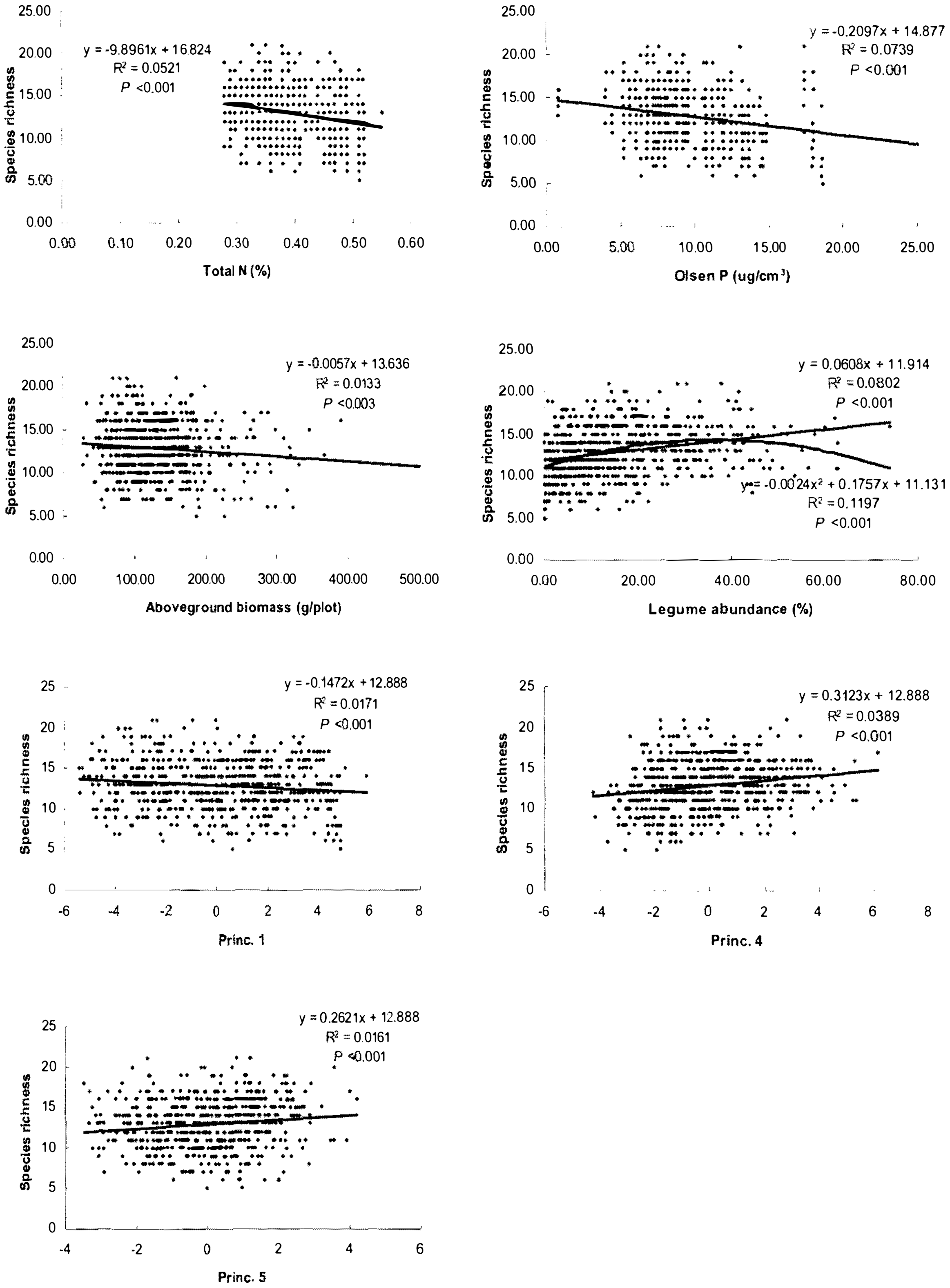


Fig. 4. Responses of species richness to soil total N, soil Olsen P, aboveground biomass, legume abundance, and principal components 1, 4, and 5.

The responses of species richness to soil Olsen P, total N, aboveground biomass, legume abundance and principal components 1, 4 and 5 are plotted in Fig. 4. Regression models best fit each of the responses are also indicated. For legume abundance, a curvilinear model fitted better on the response than a linear model, for all other factors, the best fit was a linear model.

7.4. Discussion

Legume abundance was indicated by the decision tree model (Fig. 1) as the most important variable in explaining the variance in species richness. The correlation analysis (Table 3) also showed that correlation coefficient between legume abundance and species richness was the highest among the variables that had a significant correlation with species richness. The reasons why legume played an important role in explaining the variance in species richness may be because it had a significant correlation with factors that could influence species richness, such as biomass, soil fertility (N and P), fertiliser input and slope, or may be because legume species had an independent influence on species richness. Increased legume abundance may increase aboveground biomass and N content in soil; also increased annual P fertiliser input can stimulate legume growth (Chapter 8 and Lambert *et al.* 1986), but this would lead to decreased species richness as both aboveground biomass, soil N content and annual P fertiliser input had negative correlation with species richness (Table 3). Legume abundance had a positive correlation with slope (Chapter 8 and Lambert *et al.* 1986) and slope had a positive correlation with species richness. This may partially explain why increased legume abundance resulted in increased species richness, but as indicated in Table 3, when including slope aboveground biomass and soil Olsen P as partial variables, the correlation coefficient between legume abundance and species richness only slightly decreased from 0.2833 to 0.2446, and the partial correlation was still very significant ($P < 0.0001$). This suggests that legume abundance may have an independent influence on species richness. However, it is hard to draw a conclusion that legume was a driver of species richness as the result was not from a designed experiment in which legume abundance was a treatment. It was not clear how legume exerted influence on species richness. One possibility is that increased legume abundance may increase the degree of species complementarity and thus resulting in the coexistence of more species in a community as legume species can fix atmosphere N_2 by symbiotic fixation, and do not compete with other species for nitrogen (Spehn *et al.* 2002; van Ruijven & Berendse 2003).

Sheep grazed pasture had more species than cattle grazed pasture (Fig. 1 and Fig. 4c), suggesting that the grazing animal species also had a significant effect on species richness. Herbivores are generally thought to enhance plant diversity by their direct consumption of competitively dominant plant species and indirect effects on plant competition in a nutrient-rich ecosystem (McNaughton 1985; Proulx & Mazumder 1998). Sheep tend to graze lower and more selectively than cattle (Clark *et al.* 1984; Nicol *et al.* 1993; Betteridge *et al.* 1994). This sheep grazing behaviour, comparing to cattle grazing, may reduce the height of dominant species and, therefore, decrease the competition among species, and allow more species to co-exist in a community. Another reason that sheep grazed pasture had more species than cattle grazed pasture may be because cattle treading can seriously damage sone pasture species in wet winters (Pande 2002).

Climatic factors such as temperature, solar radiation and rainfall also showed significant correlation with species richness (Fig. 2), but their influence on species richness mainly worked indirectly by influencing aboveground biomass as when aboveground biomass was included as a partial variable, the partial correlation between species richness and temperature, and between species richness and solar radiation were not significant (Table 3).

Aboveground biomass had a negative effect on species richness (Fig. 1), but as indicated in Table 3, the correlation between biomass and species richness was not significant when soil Olsen P was included as a partial variable. On the other hand, the correlation between soil Olsen P and species richness was significant when biomass was included as a partial variable. This suggests that the effect of biomass on species richness may be a hidden effect of soil nutrient. This is also in agreement with the result in the decision tree model using principal components (Fig. 3) that soil fertility and legume abundance were the main factors influencing species richness.

Species have different response rates to resources; some always respond better than others (Tilman 1982; Valverde *et al.* 1997; Hubbard *et al.* 1999; Nicholas 1999). When resources are high, those species which have the better response (i.e. faster growth) may dominate the community and exclude the less responsive species, but when resources are low, all species may have a similar growth rate and thus all of them can co-exist in a community. Pasture with a high aboveground biomass had low species richness in this grazing ecosystem not because at high biomass the dominant species competed with others by shading or other effects (Rajaniemi 2002), but rather because regular grazing reduced biomass and height of the dominant species, and species which have better response to high resources dominated the community and excluded the less responsive species.

As a hump-shaped response is commonly observed and can be very well explained by competitive exclusion theory (Grime 1973), all other responses are usually analysed for reasons why they are not hump-shaped responses (Gough *et al.* 1994; Grace 1999). The hump-shaped response is usually observed in an environmental gradient that extends from extremely "poor" to extremely "rich" within a local area (Guo & Berry 1998), and a restricted gradient may, therefore, give positive, negative and non-significant responses depending on the ranges of environmental gradient sampled (Gough *et al.* 1994; Guo & Berry 1998). The observed negative response of species richness to aboveground biomass soil Olsen P and total N may be due to the samples only covering a "narrow" gradient.

Though legume abundance had a positive correlation with species richness as indicated in decision tree models and correlation analyses, I found a curvilinear model fit the response better than a linear model (Fig. 4). This indicated that species richness is more like a hump-shaped response to legume abundance; when legume abundance increased to a threshold, further increase would lead to the strong competition of legume species with other species and would result in reduced species richness.

However, the results also indicated that the influences of each of the legume abundance, grazing animal species, soil fertility and other factors on the species richness were weak, though very significantly, suggesting that no one factor was the main element controlling species richness. It was the cumulative effects of many significant factors underlying the patterns of species richness in this pasture ecosystem.

My results did not support the results found by Chase & Leibold (2002) that species richness is a hump-shaped response at the local-scale. Also species richness and composition recorded in this pasture exhibited little change from the 1970s to 1990s (Nicholas 1999; Lopez 2000). That is, the available species capable of growing under this environmental condition were the same during the studied periods, suggesting that species pool may not be a factor regulating species richness pattern.

It is a controversial topic that whether higher diversity is beneficial for an ecosystem (Loreau 2001; Pifisterer & Schmid 2002). From an agricultural point of view, a community which has a small number of productive species is desirable, as more species means more unproductive and/or weed species, which would require more effort in management (McNaughton 1994). However, from an ecological point of view, higher diversity is considered desirable as ecosystem functioning such as stability is, in general, positively correlated to species diversity (Huston 1994; Tilman & Downing 1994). Generally for hill-pasture ecosystems in New Zealand, it is probably better to only have several highly productive species such as perennial ryegrass (*Lolium perenne*) and whiter clover (*Trifolium repens*), but a few studies (e.g. Daly *et al.* 1996) indicated that pastures with multi-species had better performance than ryegrass-white clover pastures, in dry areas as ryegrass and white clover are not very drought-tolerant (Campbell 1996). Therefore, a desirable species pattern in hill-pasture could be having less but mainly high productive species at local scale (e.g. paddock scale) to maintain high production and having higher species richness at a larger scale (e.g. catchment scale) to maintain community stability (Tilman & Downing 1994; Nicholas *et al.* 1998).

In summary, legume abundance, grazing animal species and soil fertility are the most significant variables explaining the variation in species richness in this naturalised pasture ecosystem. Species richness had a positive response to legume abundance, and a negative response to soil fertility and aboveground biomass. Sheep grazed pasture had more species than cattle grazed pasture. Climatic factors, by influencing biomass, also influenced species richness in some degree. The influence of legume abundance on species richness may be a result of niche complementarity as legume species can fix atmosphere N_2 and do not compete with other species for nitrogen. Pasture grazed by sheep had more species than pasture grazed by cattle may be because sheep grazing reduced more height of the dominant species due to their more selectively grazing behaviour and allow more species grow together.

7.6. References

Abrams P.A. (1995) Monotonic or unimodal diversity-productivity gradients: what does competition theory predict? *Ecology*, 76, 2019-2027

Bell G., Lechowicz M.J. & Waterway M.J. (2000) Environmental heterogeneity and species diversity of forest sedges. *Journal of Ecology*, 88, 67-87

Belsky A.J. (1992) Effects of grazing, competition, disturbance and fire on species composition and diversity in grassland. *Journal of Vegetation Science*, 3, 187-200

Betteridge K., Fletcher R.H., Liu Y., Costall D.A. & Devantier B.P. (1994) Rate of removal of grass from mixed pastures by cattle, sheep and goat grazing. *Proceedings of the New Zealand Grassland Association*, 56, 61-65

Brown J.H. (1973) Species diversity of seed-eating desert rodents in sand dune habitats. *Ecology*, 54, 775-787

Campbell B.D. (1996) Climate change and pastures. *Proceedings 48th Ruakura Farmers' Conference held at Ruakura*, New Zealand, 11 June 1996, 67-74

Chase J.M. & Leibold M.A. (2002) Spatial scale dictates the productivity-biodiversity relationship. *Nature*, 416, 427-430

Clark D.A., Rolston M.P., Lambert M.G. & Budding P.J. (1984) Pasture composition under mixed sheep and goats grazing on hill country. *Proceedings of the New Zealand Grassland Association*, 45, 160-167

Daly M.J., Hunter R.M., Green G.N. & Hunt L. (1996) A comparison of muti-species pasture with ryegrass-white clover pasture under dry condition. *Proceedings of the New Zealand Grassland Association*, 58, 53-58

Fernandez G. (2003) *Data mining using SAS applications.* Chapman & Hall/CRC, Boca Raton, Florida

Garcia L.V., Maranon T., Moreno A. & Clemente L. (1993) Above-ground biomass and species richness in a Mediterranean salt marsh. *Journal of Vegetation Science*, 4, 417-424

Goldberg D.E. & Miller T.E. (1990) Effects of different resource additions on species diversity in an annual plant community. *Ecology*, 71, 213-225

Gough L., Grace J.B. & Taylor K.L. (1994) The relationship between species richness and community biomass: the importance of environmental variables. *Oikos*, 70, 271-279

Grace J.B. (1999) The factors controlling species density in herbaceous plant communities: an assessment. *Perspectives in Plant Ecology, Evolution and Systematics*, 2, 1-28

Grace J.B. (2001) The roles of community biomass and species pools in the regulation of plant diversity. *Oikos*, 92, 193-207

Grime J.P. (1973) Competitive exclusion in herbaceous vegetation. *Nature*, 242, 344-347

Grime J.P. (1997) Biodiversity and ecosystem function: the debate deepens. Science, 277, 1260-1261

Gross K.L., Willig M.R., Gough L., Inouye R. & Cox S.B. (2000) Patterns of species density and productivity at different spatial scales in herbaceous plant communities. *Oikos*, 89, 417 -- 427

Gunnarsson U., Rydin H. & Sjors H. (2000) Diversity and pH changes after 50 years on the boreal mire Skattlosbergs Stormosse, Central Sweden. *Journal of Vegetation Science*, 11, 277-286

Guo Q. & Berry W.L. (1998) Species richness and biomass: dissection of the hump-shaped relationships. *Ecology*, 79, 2555-2560

Hastings A. (1980) Disturbance, coexistence, history, and competition for space. *Theoretical Population Biology*, 18, 363-373

Hubbard R.K., Ruter J.M., Newton G.L. & Davis J.G. (1999) Nutrient uptake and growth response of six wetland/riparian plant species receiving swine lagoon effluent. *Transactions of the ASAE*, 42, 1331-1341

Huston M.A. (1994) *Biological diversity: The coexistence of species on changing landscapes.* Cambridge University Press. Cambridge, UK

Huston M.A. & DeAngelis D.L. (1994) Competition and coexistence: the effects of resource transport and supply rates. *American Naturalist*, 144, 954-977

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

Knapp A.K., Fay P.A., Blair J.M., Collins S.L., Smith M.D., Carlisle J.D., Harper C.W., Danner B.T., Lett M.S. & McCarron J.K. (2002) Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science*, 298, 2202-2205

Lambert M.G., Barker D.J., Mackay A.D. & Springett J.O. (1996) Biophysical indicators of sustainability of North Island hill pastures. *Proceedings of the New Zealand Grassland Association*, 57, 31-36

Lambert M.G., Clark D.A., Grant D.A. & Costall D.A. (1986) Influence of fertilizer and grazing management on North Island moist hill country 2. Pasture botanical composition. *New Zealand Journal of Agricultural Research*, 29, 1-10

Lambert M.G. & Roberts E. (1978) Aspect differences in an unimproved hill country pasture 2. Edaphic and biotic difference. *New Zealand Journal of Agricultural Research*, 21, 255-260

Lopez I.F. (2000) *Ecology of pastoral communities in a heterogeneous environment.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Loreau M., Naeem S., Bengtsson J., Grime J.P., Hector A., Hooper D.U., Huston M.A., Raffaelli D., Schmid B., Tilman D. & Wardle D.A. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294, 804-808

Lundholm J.T. & Larson D.W. (2003) Relationships between spatial environmental heterogeneity and plant species diversity on a limestone pavement. *Ecography*, 26, 715-722

Maranon T. & Garcia L.V. (1997) The relationship between diversity and productivity in plant communities: facts and artefacts. *Journal of Ecology,* 85, 95-96

McNaughton S.J. (1985) Ecology of grazing ecosystem: the Serengeti. *Ecological Monograph*, 53, 291-320

McNaughton S.J. (1994) Biodiveristy and function of grazing ecosystem. In: Schulze, E.D, Mooney, H.A. (Eds.), *Biodiversity and Ecosystem Function*, Springer-Verlag, 361-382

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand

Nicholas P.K., Kemp P.D. & Barker D.J. (1998) Biodiversity, stability and pasture management - the role of functional groups. *Proceedings of the 9th Australian Agronomy Conference*, Wagga wagga, 1998

Nicol A.M., Russel A.J.F. & Wright I.A. (1993) Integrated grazing of goats with sheep or cattle on continuously grazed pasture. *Proceedings of the XVII International Grassland Congress*, 1320-1322

Oiff H. & Ritchie M.E. (1998) Effects of herbivores on grassland plant diversity. *Trends in Ecology and Evolution*, 13, 261-265

Oksanen J. (1996) Is the humped relationship between species richness and biomass an artefact due to plot size? *Journal of Ecology*, 84, 293-295

Pande T.N. (2002) *Pasture dynamics under cattle treading.* Ph. D. thesis, Massey University, Palmerston North, New Zealand

Pifisterer A.B. & Schmid B. (2002) Diversity-dependent production can decrease the stability of ecosystem functioning. *Nature*, 416, 84-86

Proulx M. & Mazumder A. (1998) Reversal of grazing impact on plant species richness in nutrient-poor vs. nutrient-rich ecosystem. *Ecology*, 79, 2581-2592

Rajaniemi T.K. (2002) Why does fertilization reduce plant species diversity? Testing three competition-based hypotheses. *Journal of Ecology*, 90, 316-324

Rosenzweig M.L. (1992) Species diversity gradients: we know more and less than we thought. *Journal of Mammalogy*, 73, 715-730

Spehn E.M., Scherer-Lorenzen M., Schmid B., Hector A., Caldeira M.C., Dimitrakopoulos P.G., Finn J.A., Jumpponen A., O'Donnovan G., Pereira J.S., Schulze E.-D., Troumbis A.Y. & Körner C. (2002) The role of legumes as a component of biodiversity in a cross-European study of grassland biomass nitrogen. *Oikos*, 98, 205-218

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

Tilman D. (1982) *Resource competition and community structure (Monographs in population biology).* Princeton University Press

Tilman D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, 77, 350-363

Tilman D. & Downing J.A. (1994) Biodiversity and stability in grassland. *Nature*, 367, 363-365

Valverde T., Pisanty I. & Rincon E. (1997) Growth response of six tropical dune plant species to different nutrient regimes. *Journal of Coastal Research*, J13, 497-505

van Ruijven J. & Berendse F. (2003) Positive effects of plant species diversity on productivity in the absence of legumes. *Ecology Letters*, 6, 170-175

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

Waide R.B., Willig M.R., Steiner C.F., Mittelbach G., Gough L., Dodson S.I., Juday G.P. & Parmenter R. (1999) The relationship between productivity and species richness. *Annual Review of Ecology and Systematics*, 30, 257-300

Witten I. & Erank E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers San Francisco, California

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

Zobel M. (1997) The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence. *Trends in Ecology and Evolution*, 12, 266-269

Chapter 8. General Discussion and Conclusions

一出戏即将落幕，另一出戏又要上演。

It is an end，and it is a new start.

8.1. Introduction

In the studies reported in the previous chapters, community productivity, species functional group abundance and species richness in hill-pasture in the North Island, New Zealand, were investigated by using decision tree, regression and other modelling approaches with integration of a geographic information system (GIS) in most of the cases.

Three main aspects appeared to be very significant. The first is the methodological aspect of applying a decision tree approach in modelling pasture productivity, species abundance and species richness. To the best of my knowledge, this is the first time a decision tree was used to model community productivity in vegetation science. The second aspect is the advantages of using GIS to derive spatial and non-spatial variables as model input and to present model predictions spatially. The third aspect is the insight obtained from these models and their implications in ecological study and pasture management. In this chapter, I will discuss the above three aspects and synthesise the general conclusions obtained from the previous chapters.

8.2. The decision tree approach

The responses of pasture productivity to environmental and management factors are usually non-linear and the distribution of productivity data is often not normal (Moir *et al.* 2000; Rickert *et al.* 2000). Productivity data collected from different times and places also contains high heterogeneity (i.e. highly variable). Also, some of the environmental factors that influence productivity are nominal variables, for example, aspect and soil type. It is, therefore, difficult to use regression models to predict pasture productivity due to the assumptions underlying regression models and their limited ability to incorporate nominal variables (Sen & Srivastava 1990). A decision tree, on the other hand, is a suitable method to deal with these kinds of data: it has no strict requirement on the response and distribution of the target variables, and it is easy to incorporate nominal, categorical and interval variables in a model (Vayssieres *et al.* 2000).

However, because of the discontinuous nature in generating predictions in the decision tree approach, it is mainly used for modelling categorical or ordinal variables

(Breiman *et al.* 1984; Scheffer 2002) and seems not an intuitive choice in modelling pasture productivity. There will be a trade-off between generating continuous prediction and being able to analyse the pasture productivity data that cannot be adequately analysed using regression models.

The applications of the decision tree method in Chapter 3 and Chapter 4 indicated that the sacrifice of not generating continuous predictions did not significantly influence the model output. In fact, the decision tree models had very good performance and had better predictive ability than the regression models. For example, the decision tree model for annual pasture productivity adequately predicted 90.1% of the cases in the model validation which was 10.8 percentage points higher than that of the regression model. The decision tree models also clearly indicated the relative importance of input variables on model target and the interaction among them.

Species abundance has been considered to have several distributions, including Poisson distribution, negative binomial distribution, canonical log-normal distribution, broken-stick distribution and ordinal distribution (see Guisan & Zimmermann 2000). It is, therefore, not appropriate to use least square regression to model species abundance due to the violation of its normal distribution assumption. The applications of the decision tree provided an alternative approach to model species abundance. The results indicated that decision tree was an appropriate method in this case with respect to predictive accuracy and investigating the interrelationship between target variables and input variables. For example, the decision tree model for the five functional groups had an overall predictive accuracy of 75% and indicated the most significant variables which influenced the abundance of the five functional groups.

The decision tree is also a good approach in facilitating the development of a regression model. By using variables revealed as the most significant variables in influencing pasture productivity and the interaction among these variables in the decision tree models in Chapter 3 and 4, the polynomial regression model developed for assessing climate change and alternative phosphorus fertiliser application strategies (Chapter 5) demonstrated a better goodness-of-fit than the regression model developed in Chapter 3 using principal component analysis.

There are several disadvantages of using the decision tree to model community productivity, species abundance and species richness. Firstly, it is not as concise as a regression model in expressing the relationship between model target and input variables. A regression model shows whether an input variable is positively or negatively influencing model target and in what scale by using a mathematic equation. A decision tree model only shows an average response of model target to input variables. Secondly, a decision tree cannot generate a continuous prediction, and thus could not detect the influence of small changes in environmental and management variables on model targets.

8.3. Advantages of applying geographic information system (GIS)

Integration of GIS with environmental and ecological models has greatly enhanced the investigative and predictive capability of environmental and ecological models (Iverson *et al.* 1997; Johnston 1998; Li *et al.* 1998; Wadsworth & Reweek 1999). Here GIS has been linked with the decision tree models in modelling pasture productivity (Chapter 4) and the functional group abundance of plant species (Chapter 6), and has been linked with a regression model in assessing the impact of climate change and alternative phosphorous fertiliser application strategies on pasture production (Chapter 5). The power of using GIS in these modelling processes was demonstrated in deriving input variables for model analyses and generating model predictions over space for the area of interest. As for deriving input variables, GIS not only can be used to generate spatial features such as slope and aspect from a digital elevation model (DEM), but it can also be used to interpolate point data into surface over large areas. For example, in Chapter 5, the climate surfaces for rainfalls and temperatures were developed by interpolating point observations into a continuous surface for the whole North Island for model predictions in the scenario analyses.

The combination of GIS with the decision tree and regression models in previous chapters provided a platform to generate model predictions for specific farm or pasture of interest. Information obtained from analysing these predictions can be used to facilitate pasture management. For example, in Chapter 6 the predicted abundance of HFRG species was displayed in a GIS map and areas with less HFRG composition can be outlined for applying phosphorus fertiliser for encouraging their growth. The decision tree models for annual and seasonal pasture productivity developed in Chapter 3 are currently being used by another research project[1] to assess the response of hill-pasture to P fertiliser for better economic return as a practice of precision agriculture.

Using GIS is especially important in modelling pastures with heterogonous topology and management such as those in hill-country. However there are also limitations of using GIS in these modelling processes, which may include the lack of a high-resolution digital elevation model (DEM) and the requirement for computer hardware and software. However, these limitations will gradually disappear with the development of technologies in computer science and GIS.

8.4. Model insights and implications in ecological study and pasture management

8.4.1. Pasture productivity

The available spring rainfall was indicated as the most significant factor influencing annual productivity in hill-pasture of the North Island, while hill slope was the second most significant variable influencing pasture productivity for pastures with both high and low spring rainfall. N and P fertiliser inputs, soil Olsen-P and annual mean daily temperature were the significant variables influencing pasture productivity after spring rainfall and slope (Chapter 3). This suggests that pasture production in hill country was firstly regulated by unmanageable climatic and topographic variables and then was 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; White 1990; Bai *et al.* 2004). Radicliffe & Baars (1987) revealed that spring and summer rainfall accounted for 60% of the variation in annual pasture production. However, the important role of spring rainfall as the determinant of annual pasture production was not fully recognised before. The significant influence of spring rainfall on pasture productivity was because about 40% of the annual productivity was produced in spring (Chapter 3).

---

[1] Variable Rate Application Technology in the New Zealand Aerial Topdressing Industry.

Both the annual P fertiliser input and five-year cumulative input were significant in influencing pasture productivity. However, five-year cumulative P fertiliser input had a more important effect on pasture productivity than current annual P fertiliser input, indicating that P fertiliser application history made a significant contribution toward soil P fertility. However, current annual P fertiliser input showed a marked effect in increasing summer pasture productivity.

8.4.2. Impact of climate change and alternative P fertiliser application strategies on pasture production

As indicated in Chapter 3 and 4 and by others (Riedo *et al.* 1997; Bai *et al.* 2004), climate is a key driver of pasture production and the projected climate changes for New Zealand will undoubtedly have a significant impact on pasture production (Ministry for the Environment 2001). In this study, the climate change scenarios assumed an increase in temperature by 1-2 °C and a change of rainfall from -20% to +20% relative to climate normal data from 1960-1990 have predicted a wide range (-46.2% to +51.9%) of pasture production change in hill-pasture in the North Island (chapter 5). These climate scenarios, when averaged over the climate zones, predicted a range of -20.4% to +9.6% production change, which was generally in agreement with the predictions in some previous studies (e.g. Baars *et al.* 1990; Martin *et al.* 1991; Campbell *et al.* 1996). It needs to be emphasised that under the scenarios of decreased rainfall and increased temperature, there would be a decrease in pasture production on average for the whole North Island, showing that the interaction of decreased rainfall and increased temperature will impose a negative effect on pasture production. Developing and wide use of drought-tolerant species in areas where climate changes are likely to have negative impact (e.g. North and South-east of the North Island) would be a good management practice in the face of climate change

Pasture in areas with relatively low rainfall had a higher response to increased P fertiliser input than pastures in areas with a relatively high rainfall. This negative interaction between rainfall and P fertiliser input on pasture production has two implications: firstly increased rainfall that resulted from climate change may reduce the P fertiliser utilisation efficiency in hill-pasture, and secondly, using N fertiliser instead of P fertiliser, may be a better choice with respect to reducing non-point source P pollution in surface water, and increasing economic return from fertiliser application in areas with high rainfall, as recommended by Lambert & Clark (1985).

8.4.3. Functional group abundance

Hill-pasture management aims to increase the composition of high fertility response grasses (HFRG) and legumes and to reduce low fertility tolerant grasses (LFTG) and moss by applying phosphorus fertiliser and introducing legume species. Chapter 6 indicates that slope and soil Olsen P were the most significant factors driving the abundance of HFRG, LFTG and moss in hill-pasture. With increased soil fertility or decreased slope, pasture becomes more HFRG dominant; while with the decreased soil fertility or increased slope, pasture becomes more LFTG and moss dominant. Since soil Olsen P is positively influenced by P fertiliser input in hill-pasture (Gillingham A.G., Gray M.H. & Smith D.R.1998; Lambert *et al.* 2000), the amount of fertiliser input, therefore, is one of the key factors influencing the relative abundance of functional group HFRG, LFTG, flatweeds and moss in hill-pasture.

Soil bulk density was indicated as the most significant variable influencing legume abundance; legume abundance was positively correlated with soil bulk density with a threshold of 0.89 g/cm^3. The influence of soil bulk density on relative abundance of legume has not been previously reported and the interpretation is not clear. This result has an implication that the practice of applying P fertiliser in hill-pasture to encourage legume growth can be compromised in certain localities by the low bulk density in a pasture.

The species of grazing animal influence pasture species abundance by changing the competition pattern in a pasture community through their selective defoliation (McNaughton 1985). HFRG, LFTG and legume showed no significant response to different grazing animal species. However, compared to sheep grazing, cattle grazing significantly reduced the relative abundance of flatweeds and moss.

8.4.4. Species richness

Species richness has long been an interest of ecological studies (Grime 1973; Abrams 1995; Waide *et al.* 1999; Chase & Leibold 2002). Species richness is indicated to be closely related to community stability and/or productivity (Tilman *et al.* 2001; Mouquet *et al.* 2002; Naeem 2002). These studies mainly focused on resource availability and species interaction, mainly competition, as the forces driving species richness (Grime 1977; Waide *et al.* 1999; Mouquet *et al.* 2002). In Chapter 7, legume abundance was indicated as the most important variable in explaining the variation in species richness in hill-pastures. The mechanism may be because increased legume abundance increases the degree of species complementarity and thus resulted in the coexistence of more species in a community as legume species can fix atmospheric N_2 by symbiotic fixation, and do not compete with other species for nitrogen (Spehn *et al.* 2002; van Ruijven & Berendse 2003). The result in this case indicated that species complementarity, as well as competition, may also play a role in regulating species richness.

Sheep grazed pasture had more species than cattle grazed pasture, suggesting that the grazing animal species also had a significant effect on species richness. Herbivores are generally thought to enhance plant diversity by their direct consumption of competitively dominant plant species and indirect effects on plant competition in a nutrient-rich ecosystem (McNaughton 1985; Proulx & Mazumder 1998). Sheep tend to graze lower and be more selective than cattle (Clark *et al.* 1984; Nicol *et al.* 1993; Betteridge *et al.* 1994). This sheep grazing behaviour, compared to cattle grazing, may reduce the height of dominant species and, therefore, decrease the competition among species, and allow more species to co-exist in a community.

8.5. Conclusions

1) The decision tree is an appropriate modelling method in predicting pasture community productivity, species functional group abundance and species richness in the hill-pasture ecosystems. It had very high predictive accuracy and clearly indicated the relative importance of input variables on model targets. It can also be used to facilitate regression analysis in selecting the input variables. The limitation of using the decision tree to predict pasture productivity and species abundance and richness was that it did not generate a continuous output, and thus could not detect the influence of small changes in environmental and management variables on model targets.

2) Geographic information system (GIS), with its strong ability in analysing spatial and spatial related variables, provided a powerful tool in modelling

and investigating pasture productivity and species abundance for pastures with heterogeneous topographic, edaphic, or management features. It helped to derive input data such as topology and climate variables and to achieve the prediction for further analyses. Integration of a GIS and statistical models provided a platform to investigate community structure and functional composition for a pasture over space, and thus can be applied as a tool in pasture management.

3) Spring rainfall was the most significant factor influencing annual pasture productivity, while hill slope was the most significant factor influencing spring and winter pasture productivity. Annual P fertiliser input and autumn rainfall were the most significant factors influencing summer and autumn pasture productivity, respectively.

4) Climate change scenarios of increased temperature by 1-2 ^{o}C and rainfall change by -20% to +20% would have a great impact on pasture production in North Island. From these climate scenarios a wide range (-46.2% - +51.9%) of pasture production variation was predicted compared with the current climate situation. Increased temperature would generally have a positive effect in the south and southeast part of the North Island. Increased rainfall would have a positive effect on pasture production in the central, south and southeast parts of the North Island, but a negative effect in the north part of the North Island. The decreased rainfall would only have a positive effect in some central areas with high rainfall.

5) Pasture in areas with relatively low rainfall had a higher response to increased P fertiliser input than pastures in areas with a relatively high rainfall.

6) Soil Olsen P was the most significant factor influencing the relative abundance of low fertility tolerance grasses (LFTG) and moss, while soil bulk density, slope and annual P fertiliser input were the most significant factors influencing the relative abundance of legume, high fertility response grasses (HFRG) and flatweeds, respectively in hill-pasture. Generally, slope and soil Olsen P were the two key factors underlying the patterns of abundance for these five functional groups.

7) Legume abundance was the most significant factor explaining the variation in species richness in hill-pasture. Species richness increased with the increase of legume abundance and showed a tendency of a hump-shaped response. The effect of legume abundance on species richness was possibly a result of species complementarity resulting from nitrogen fixation of legume species. Grazing animal species also have a significant effect on species richness; pasture grazed by sheep had more species than pasture grazed by cattle.

8.6. References

Abrams P.A. (1995) Monotonic or unimodal diversity-productivity gradients: what does competition theory predict? *Ecology*, 76, 2019-2027

Baars J.A., Radcliffe J.E. & Rollo M.D. (1990) Climatic change effects on seasonal patterns of pasture production in New Zealand. *Proceedings of the New Zealand Grassland Association*, 51, 43-46

Bai Y.F., Han X.G., Wu J.G., Chen Z.Z. & Li L.H. (2004) Ecosystem stability and compensatory effects in the Inner Mongolia grassland. *Nature*, 431, 181-184

Betteridge K., Fletcher R.H., Liu Y., Costall D.A. & Devantier B.P. (1994) Rate of removal of grass from mixed pastures by cattle, sheep and goat grazing. *Proceedings of the New Zealand Grassland Association*, 56, 61-65

Breiman L., Friedman J.H., Olshen R.A. & Stone J.C. (1984) *Classification and regression trees*. The Wadsworth statistics/probability series. Chapman & Hall. Inc. New York, NY.

Campbell B.D., McKeon G.M., Gifford R.M., Clark H., Stafford Smith M.S., Newton P.C.D. & Lutze J.L. (1996) Impacts of atmospheric composition and climate change on temperate and tropical pastoral agriculture. In: Bouma, W. J., Pearman, G. I., and Manning, M. R. (eds.), *Greenhouse: Coping with Climate Change*. CSIRO Publishing, Collingwood, Victoria, Australia, pp. 171–189.

Chase J.M. & Leibold M.A. (2002) Spatial scale dictates the productivity-biodiversity relationship. *Nature*, 416, 427-430

Clark D.A., Rolston M.P., Lambert M.G. & Budding P.J. (1984) Pasture composition under mixed sheep and goats grazing on hill country. *Proceedings of the New Zealand Grassland Association*, 45, 160-167

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

Grime J.P. (1973) Competitive exclusion in herbaceous vegetation. *Nature*, 242, 344-347

Grime J.P. (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist*, Vol 111, 1169-1194

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

Iverson L.R., Martin E.D., Scott C.T. & Prasad A. (1997) A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio (U.S.A.). *Landscape Ecology*, 12, 331-348

Johnston C. (1998) *Geographic information system in ecology*. Oxford, Malden, MA: Blackwell Science

Lambert M.G. & Clark D.A. (1985) Increasing productivity of hill country with nitrogen. *Proceedings 20th Technical Conference of New Zealand Fertiliser Manufacturers' Research Association*, 1, 274-283

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

Lambert M.G., Clark D.A., Mackay A.D. & Costall D.A. (2000) Effects of fertiliser application on nutrient status and organic matter content of hill soils. *New Zealand Journal of Agricultural Research*, 43, 127-138

Li J., Liang T. & Chen Q. (1998) Estimating grassland yields using remote sensing and GIS technologies in China. *New Zealand Journal of Agricultural Research*, 41, 31-38

Martin R.J., Korte C.J., McCall D.G., Baird D.B., Newton P.C.D. & Barlow N.D. (1991) Impact of potential change in climate and atmospheric concentration of carbon dioxide on pasture and animal production in New Zealand. *Proceedings of the New Zealand Society of Animal Production*, 51, 25-33

McNaughton S.J. (1985) Ecology of grazing ecosystem: the Serengeti. *Ecological Monograph*, 53, 291-320

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

Mouquet N., Moore J.L. & Loreau M. (2002) Plant species richness and community productivity: why the mechanism that promotes coexistence matters. *Ecology Letters*, 5, 56-64

Naeem S. (2002) Ecosystem consequences of biodiversity loss: the evolution of a paradigm. *Ecology*, 83, 1537-1552

Nicol A.M., Russel A.J.F. & Wright I.A. (1993) Integrated grazing of goats with sheep or cattle on continuously grazed pasture. *Proceedings of the XVII International Grassland Congress*, 1320-1322

Proulx M. & Mazumder A. (1998) Reversal of grazing impact on plant species richness in nutrient-poor vs. nutrient-rich ecosystem. *Ecology*, 79, 2581-2592

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

Riedo M., Gyalistras D., Grub A., Rosset M. & Fuhrer J. (1997) Modelling grassland responses to climate change and elevated CO_2. *Acta Oecologica*, 18, 305-311

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

Sen A.K. & Srivastava M. (1990) *Regression analysis: theory, methods and applications.* New York: Springer-Verlag

Spehn E.M., Scherer-Lorenzen M., Schmid B., Hector A., Caldeira M.C., Dimitrakopoulos P.G., Finn J.A., Jumpponen A., O'Donnovan G., Pereira J.S., Schulze E.-D., Troumbis A.Y. & Körner C. (2002) The role of legumes as a component of biodiversity in a cross-European study of grassland biomass nitrogen. *Oikos*, 98, 205-218

Tilman D., Reich P.B., Knops J., Wedin D., Mielke T. & lehman C. (2001) Diversity and productivity in a long-term grassland experiment. *Science*, 294, 843-845 van Ruijven J. & Berendse F. (2003) Positive effects of plant species diversity on productivity in the absence of legumes. *Ecology Letters*, 6, 170-175

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

Wadsworth R. & Treweek J. (1999) *GIS for ecology: an introduction.* Harlow: Addison Wesley Longman

Waide R.B., Willig M.R., Steiner C.F., Mittelbach G., Gough L., Dodson S.l., Juday G.P. & Parmenter R. (1999) The relationship between productivity and species richness. *Annual Review of Ecology and Systematics*, 30, 257-300

White J.G.H. (1990) Hill and high country pasture. In: R.H.M. Langer (Ed.), *Pastures: Their Ecology and Management*, Oxford University Press, Auckland, N.Z. 1990, 299-336

Appendix 1. List of species in the hill-pasture studied in Chapter 7 [after Lambert et al. (1986) and Nicholas et al. (1998)]

Functional group	Species
High fertility response grasses (HFRG)	*Dactylis glomerata, Holcus lanatus, Lolium perenne, Poa annua and Poa pratensis.*
Low fertility tolerance grasses (LFTG)	*Agrostis capillaris, Anthoxanthum odoratum, Cynosurus cristatus, Festuca rubra, Rytidosperma* spp..
Legume	*Lotus pedunculatus, Trifolium repens, Trifolium dubium and Trifolium subterraneum.*
Flatweeds	*Hypochaeris radicata, Leontodon taraxacoides and Plantago lanceolata.*
Moss	*Musci* spp.
Other species	*Achillea millefolium, Bellis perennis, Carex spp., Centella uniflora, Cirsium arvense, Crepis capillaris, Cymbalaria uralis, Luzula spp., Galium arvense, Gnaphalium spp., Hydrocotyle sp., Linum bienne, Montia verna, Nertera setulosa, Polycarpon tetraphyllum, Rumex acetosella, Sagina procumbens, Silene gallica, Stellaria media, Taraxacum officinale and Veronica persica.*

Lambert M.G., Clark D.A., Grant D.A. & Costall D.A. (1986) Influence of fertilizer and grazing management on North Island moist hill country 2. Pasture botanical composition. *New Zealand Journal of Agricultural Research*, 29, 1-10

Nicholas P.K. (1999) *Environmental and management factors as determinants of pasture diversity and production of North Island, New Zealand hill pasture systems.* Ph.D. thesis, Massey University, Palmerston North, New Zealand