Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Ecological Modelling 186 (2005) 299–311
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Modelling the productivity of naturalised pasture in the North Island, New Zealand: a decision tree approach

Baisen Zhang *, Ian Valentine, Peter Kemp

Institute of Natural Resources, Massey University, Palmerston North, New Zealand

Received 20 January 2004; received in revised form 30 November 2004; accepted 17 December 2004
Available online 19 February 2005

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 correctly 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 correctly predicted 90% 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. 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.

Keywords: Decision tree; Data-mining; Modelling; Naturalised hill-pasture; Productivity; Regression model

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. Van Dyne, 1970; Seligman and Baker, 1993; Stuth et al., 1993; Soren-

* Corresponding author. Tel.: +64 6 3569099x7208; fax: +64 6 3569099x5614.
E-mail addresses: baisen.zhang.1@uni.massey.ac.nz (B. Zhang), i.valentine@massey.ac.nz (I. Valentine), p.kemp@massey.ac.nz (P. Kemp).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.12.016

son, 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 and 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 and 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 and 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 and 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 clearly indicate the relative importance of input variables with respect to their influences on the model target, and can also 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 and 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, 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 performs. 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.