



Modelo de Árbol de Decisión: una herramienta para el manejo de la pradera

Decision tree model: A tool for pasture management

Flores, P.G.^{a-b*}, López, I.F.^{b-c}, Kemp, P.D.^c, Dörner, J.^d, Zhang, B.^e

^a Escuela de Graduados, Facultad de Ciencias Agrarias,
Universidad Austral de Chile, Campus Isla Teja, Valdivia, Chile.

^b Instituto de Producción Animal, Facultad de Ciencias Agrarias, Universidad Austral de Chile.
Casilla 567 Valdivia, Chile.

^c Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand.

^d Instituto de Ingeniería Agraria y Suelos, Facultad de Ciencias Agrarias,
Universidad Austral de Chile, Casilla 567, Valdivia, Chile.

^e Queensland Climate Change Centre of Excellence,
Queensland Department of Environment and Resource Management, Australia.

ARTICLE INFO

Article history:
Received 03.06.2016
Accepted 14.11.2016

Keywords:
Production
Dry matter
Datamining
Prediction

Review Article,
Special Issue: Pastures for
Sustainable Productions Systems

**Corresponding author:*
Paulina G. Flores
E-mail address:
paulina.flores@postgrado.uach.cl

ABSTRACT

The pasture is the main productive resource in the south of Chile. The dry matter productivity is related to the environment and efficient resource management. However the interrelationship between the factors involved in pasture productivity is a difficulty due to the large amount of information required to analyze. One way to determine the relationship between pasture productivity and environment is through decision trees models. These models can predict the potential effect of the environment and management applied. The objective of this review is to introduce the use of decision tree model as a tool for pasture management.

RESUMEN

La pradera es el principal recurso productivo del sur de Chile. La productividad de materia seca está relacionada al medio ambiente y al uso eficiente de los recursos. Sin embargo el impacto de la interrelación de todos los factores involucrados en la productividad de la pradera es una de las dificultades a enfrentar debido a la gran cantidad de información que se requiere manejar. Una forma de determinar la relación existente entre la pradera y medio ambiente es a través de la utilización de modelos de árboles de decisión. Estos modelos pueden predecir el efecto potencial de los cambios en el medio ambiente así como del manejo aplicado. El objetivo de esta revisión es introducir la utilización del modelo de árbol de decisión como una herramienta para el manejo de la pradera.

Palabras clave: Producción, materia seca, datamining, predicción.

INTRODUCCIÓN

Para los sistemas productivos del sur Chile la pradera es el principal recurso alimenticio (Balocchi *et al.*, 2002) siendo además el más económico. La prevalencia de especies de alto valor forrajero (como *Lolium perenne*, *Dactylis glomerata*, *Bromus valdivianus*, *Schedonorus arundinaceus* y de leguminosas como *Trifolium repens*) permite el incremento de la producción de

materia seca, lo que es clave para la mantención de un sistema productivo en base a pastoreo (Flores, 2015). Los cambios que ocurren dentro de una pradera están dados por una serie de factores como temperatura, precipitaciones, factores edáficos como la humedad del suelo y la fertilidad, factores bióticos, como el pastoreo animal y el manejo de la pradera (Guisan y Zimmermann, 2000; Korner, 1994; Nicholas *et al.*, 1998; Zhang *et al.*, 2005).

La visión en conjunto de todo lo que rodea a la pradera es una de las principales dificultades a enfrentar debido a la gran cantidad de información que se requiere manejar (Prost *et al.*, 2012). La utilización de modelos ha sido una de las estrategias implementadas como punto de partida para trabajar con una amplia base de datos ayudando a comprender la realidad de un sistema. El objetivo de esta revisión es describir el desarrollo del modelo de árbol de decisión y su utilización como herramienta en el manejo de la pradera.

UTILIZACIÓN DE MODELOS

Los modelos se han convertido en una importante herramienta de investigación agrícola desde su primera aparición hace unos 50 años (Prost *et al.*, 2012). La historia del modelamiento en sus diferentes formas dentro de la investigación en agricultura ha sido expuesta por diversos autores. Inicialmente se enfocó en la estimación de la intercepción de la luz y fotosíntesis (De Wit, 1965) dando origen a la formación de diversos grupos de investigadores dedicados a la modelización. Un área importante de esta rama es el modelamiento de cultivos en donde existen una amplia variedad de modelos desarrollados alrededor del mundo en donde destacan los modelos dinámicos (Sinclair and Seligman, 1996) los cuales han sido estructurados en tres líneas de investigación: 1) Estados Unidos con proyectos diseñados para transferencia de agrotecnología, 2) Australia con un sistema de simulación de producción agrícola y 3) Holanda con modelamiento de crecimiento de cultivos (Jones *et al.*, 2001).

Existen grandes avances en el análisis de datos, en donde el desarrollo de modelos ha logrado ser efectivo en la identificación de patrones y tendencias antes desconocidas, a partir de una gran cantidad de datos experimentales (Breiman *et al.*, 1984; Vayssieres *et al.*, 2000). En Nueva Zelanda, el modelamiento ha sido introducido en estudios medio ambientales y ecológicos, demostrando una gran aplicabilidad y exactitud en sus predicciones. Por esta razón han sido utilizados para predecir la productividad de la pradera (Wan *et al.*, 2009), para calcular dosis de fertilizante nitrogenado y fósforo (Zhang *et al.*, 2007), y para predecir la abundancia de grupos funcionales de especies (Zhang *et al.*, 2005).

Enfoques de modelamiento empírico y mecanístico

En esta revisión se han considerado dos enfoques en el modelamiento de la información; el mecanístico y empírico (Johnson *et al.*, 2008; Moir *et al.*, 2000). El modelo mecanístico ha sido utilizado mayormente en la productividad de una especie a través de ecuaciones que representan la respuesta fisiológica hacia varia-

bles ambientales. Este tipo de modelamiento requiere una amplia base de datos medio ambiental lo que podría ser una limitante al momento de realizar mediciones sobre todo si se trabaja en áreas geográficas distintas (Estes *et al.*, 2013; Rickert *et al.*, 2000).

El modelo empírico ha logrado correlacionar la distribución de una especie con respecto a factores abióticos, como el clima, siendo altamente predictivo en la evaluación de la sustentabilidad (Rickert *et al.*, 2000). Sin embargo ha sido criticado por no considerar la fisiología de una especie con la interacción de factores bióticos y abióticos (Dormann, 2007). Una práctica utilizada para incrementar la confianza de este tipo de enfoques ha sido la integración de múltiples modelos empíricos o la combinación de modelos mecanísticos y empíricos (Keith *et al.*, 2008; Franklin, 2010; Huntley *et al.*, 2010).

La elaboración de modelos predictivos contempla en general los siguientes pasos: 1) desarrollo de una base de datos, 2) elección de las variables, 3) elaboración del modelo, 4) calibración del modelo, y 5) obtención de las predicciones del modelo (Zhang *et al.*, 2005; Prost *et al.*, 2012). EL desarrollo de una base de datos en una planilla Excel es suficiente para ordenar la información con cada una de las variables independientes en contraposición con la variable dependiente a la cual se desean hacer predicciones. La base de datos puede incluir información sobre composición botánica (kg o %), producción de materia seca anual y estacional (kg MS ha⁻¹ año⁻¹), datos del medio ambiente como precipitaciones, temperatura, tipos de suelo, física-química del suelo, entre otros.

Para la elección de las variables es necesario utilizar un método estadístico que permita diferenciar aquellas variables que muestren diferencias significativas las cuáles serán las que conformarán el modelo. Una forma es utilizar una técnica de pasos hacia adelante (forward) para seleccionar aquellas variables que tuvieran un nivel de significancia <0.05 y así conformar el data set con el que se realizará el modelo predictivo (Fernández, 2003).

Existen una variedad de técnicas estadísticas para desarrollar modelos predictivos como son el modelo lineal general, regresiones múltiples, redes neuronales, modelos bayesianos, modelos de simulación, árboles de decisión, e incluso combinaciones de ellos. Pero la elección del modelo más efectivo en donde el input de variables este mayormente representado dependerá de la característica de los datos, ya sean cualitativos o cuantitativos, ordinales o continuos y de la posible respuesta dentro de la lógica del comportamiento de la variable en el medio ambiente.

Desarrollo de modelos: Data mining y árbol de decisión

Data mining

Data mining es un proceso para consultar y extraer información útil sobre patrones y tendencias desconocidas de los datos. El objetivo del data mining consiste en detectar patrones ocultos y predecibles desde una gran base de datos (Thuraisingham, 1999). Para ello se deben seguir los siguientes pasos: 1) colección de datos; 2) creación de una base de datos; 3) procesamiento de los datos; y 4) análisis y entendimiento (Zhang *et al.*, 2005).

Existen varios métodos de análisis de datos que integran el data mining como las redes neuronales, los algoritmos genéticos, análisis de clúster, regresión lineal, árboles de decisión, entre otros.

Árbol de decisión

Los Modelos de árboles de decisión son un tipo de data mining el cual ha sido ampliamente usado en ciencias sociales y en los últimos años incorporado al modelamiento ambiental con una alta precisión en sus predicciones (Yang *et al.*, 2003; Zhang *et al.*, 2005; Corson *et al.*, 2007; Wan *et al.*, 2009; Duff *et al.*, 2012).

El árbol de decisión fue creado por Breiman *et al.*, (1984) con el nombre de árbol de clasificación y regresión (CART), utilizando un enfoque no paramétrico, es decir, una estadística básica de distribución libre, que no requiere la condición de normalidad (Vayssieres *et al.*, 2000). La multidimensionalidad de la información está definida por una variable independiente que explica en su respuesta a la variable dependiente de la

forma más homogénea posible, dando como resultado una estructura con jerarquía binaria de predicciones para nuevos casos (Vayssieres *et al.*, 2000) tal como se explica en la Figura 1.

El inicio del modelo es una raíz o nodo el cual contiene toda la información del set de datos de la variable dependiente para continuar con el nombre de la variable independiente más influyente en la variable dependiente asignándole un valor numérico (x) que representa el valor en que la variable dependiente responde a la variable independiente (splitpoint). El valor de la variable independiente puede ser mayor o menor que x , por lo tanto se continúa con la lectura según corresponda para obtener el valor de la predicción (output).

Existen diferentes criterios utilizados por el modelo para dividir los valores del set de datos, uno de ellos es la reducción de la varianza o F test (Fernández, 2003). En la reducción de la varianza se realiza una búsqueda en todo el set de variables independientes seleccionando aquella que mejor explique la varianza de la variable dependiente con un valor significativo. Esta selección se va haciendo en orden consecutivo y es considerada como un regla del modelamiento (Fernández, 2003).

La Figura 2 representa un ejemplo de modelamiento realizados en una pradera naturalizada mejorada y sometida a pastoreo ovino intensivo en la ciudad de Valdivia, sur de Chile para predecir la productividad de materia seca anual. Para ello se elaboró un data mining que contenía un set de datos con mediciones realizadas durante tres años de investigación (2010-2013) sobre materia seca, precipitaciones, física y química de suelos, temperatura y humedad del suelo para la variable dependiente "Productividad de materia seca anual". El desarrollo, evaluación y ajuste del modelo se explican a continuación:

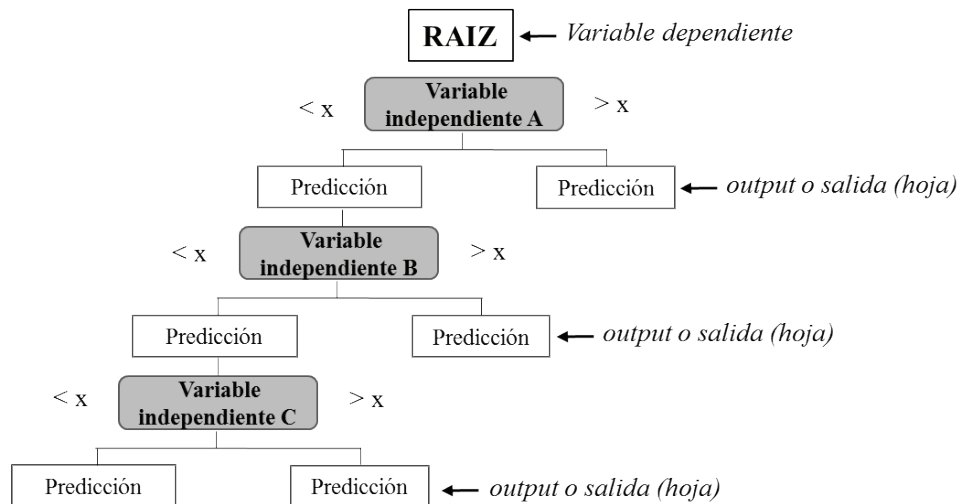


Figura 1. Estructura general de un árbol de decisión (Zhang *et al.*, 2005; Flores, 2015).

Figure 1. General structure of a decision tree (Zhang *et al.*, 2005; Flores, 2015).

1. Modelo de árbol de decisión:

La variable más importante e influyente en la productividad anual de la pradera resultó ser el contenido de nitrógeno mineral del suelo, seguida en segundo orden del porcentaje de saturación de aluminio del suelo y las precipitaciones en primavera. En tercer orden de importancia continúan las temperaturas del suelo de invierno y primavera, seguidas en cuarto orden por el porcentaje de contenido de agua del suelo. El árbol de decisión predice que cuando el contenido de nitrógeno del suelo es mayor de 47.5 mg kg⁻¹ y con precipitaciones en primavera (PP) de 199.6 mm o superiores, podrían alcanzarse 13786 kg MS ha⁻¹ año⁻¹, la cual constituye un excelente rendimiento de la pradera. Cuando las precipitaciones de primavera alcanzan valores menores a 199.6 mm pero con una temperatura del suelo superior a 17.3°C y con un contenido de agua del suelo de 36.1% el rendimiento predicho podría llegar a los 10971 kg MS ha⁻¹ año⁻¹.

En el otro sentido del árbol de decisión, cuando el contenido de nitrógeno es menor de 47.5 mg kg⁻¹, las predicciones indican los rendimientos de materia seca anual más bajos, coincidiendo con un porcentaje de saturación de aluminio del suelo superior a 7.4% y menos de 7.2°C de temperatura del suelo en invierno.

2. Ajuste del modelo

Una parte importante del desarrollo del árbol de decisión es el ajuste del modelo, el cual hace posible obtener una buena salida o predicción. Para ello se compara el modelo de árbol de decisión con la salida de otro modelo. Uno de los parámetros comparativos es el promedio del error cuadrático (ASE) y mientras más bajo este sea, mejor ajuste del modelo (Fernández, 2003).

Para el caso del modelo de productividad anual (Figura 2) se realizó una comparación con un modelo de regresión. El valor ASE registrado para el modelo de árbol de decisión fue 126 y para el modelo de regresión 209 (P = 0.001) lo que significa que el modelo de árbol de decisión tuvo mejor ajuste.

3. Evaluación del modelo

La metodología elegida para evaluar el modelo fue la descrita por Piñeiro *et al.* (2008) denominada regresión OP, en donde se confrontan los valores observados (ubicados en el eje Y) contra los valores predichos (ubicados en el eje X) utilizando un modelo de regresión en donde el intercepto es 0 y la pendiente 1 (Figura 3). En este caso, el coeficiente de determinación (R²) es

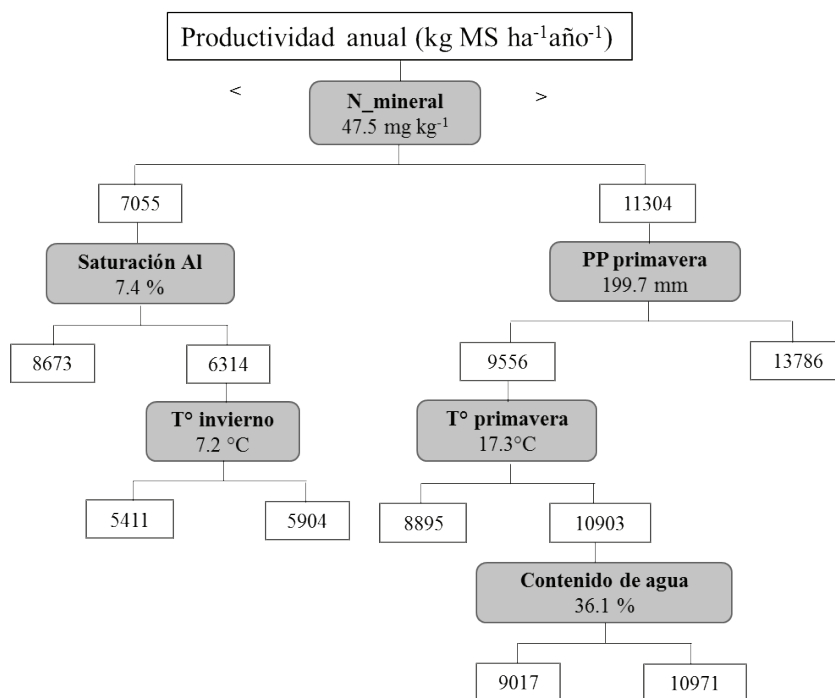


Figura 2. Modelo de árbol de decisión para productividad anual de una pradera naturalizada mejorada bajo pastoreo ovino intensivo. N_mineral, nitrógeno mineral del suelo; Saturación Al, % de saturación de aluminio del suelo; PP primavera, precipitaciones de primavera; T° invierno, temperatura de invierno; T° primavera, temperatura de primavera (Flores, 2015).

Figure 2. Decision tree model for annual productivity of an improved naturalised pasture under sheep grazing intensive. N_mineral, soil nitrogen mineral; Saturación Al, % of soil aluminum saturation; PP primavera, spring rainfall; T° invierno, winter temperature; T° primavera, spring temperature (Flores, 2015).

utilizado como una medida de la proporción de la varianza de los valores observados que fueron explicados por los valores predichos (Gauch *et al.*, 2003; Lobell *et al.*, 2005; Piñeiro *et al.*, 2008). El análisis del coeficiente de determinación junto con la pendiente y el intercepto de la línea de datos entregan elementos para juzgar y construir la performance del modelo (Piñeiro *et al.*, 2008). El valor de un $R^2 = 0.97$ ($P < 0.001$) indicó que el modelo de árbol de decisión es altamente eficiente en sus predicciones. Existen otros parámetros estadísticos que podrían ser utilizados como referente para evaluar este tipo de modelo de regresión como la raíz del error cuadrático medio (RMSE) (White *et al.*, 2000; Lobell *et al.*, 2005), sin embargo, este podría no ser aplicable a regresiones OP porque el riesgo de subestimar el valor real del error entre los valores observados y predichos aumenta (Gauch *et al.*, 2003). Un parámetro que podría utilizarse junto al R^2 es la raíz de la desviación de cuadrática media (RMSD) la cual estima la desviación media del valor predicho respecto del valor observado (White *et al.*, 2000).

4. Análisis del modelo

La salida del modelo ha determinado los factores más influyentes sobre la pradera relacionando las condiciones ambientales y de manejo con el incremento o disminución de la productividad de materia seca. El valor obtenido en el coeficiente de determinación sugieren que el modelo de árbol de decisión puede ser

utilizado como una herramienta para predecir la productividad de una pradera y también respalda la utilización que se le ha dado en otras áreas para predecir grupos funcionales de especies y productividad estacional (Zhang *et al.*, 2005; Wan *et al.*, 2009).

Estos antecedentes demuestran que el modelo de árbol de decisión puede lograr una imagen de la interacción de las variables influyentes sobre las características de una pradera, que incluso podrían ser utilizadas para realizar diagnósticos sobre el estado de un predio determinado y/o anteponerse a situaciones favorables o adversas.

Utilización del modelo de árbol de decisión

La utilización del modelo de árbol de decisión ha sido una elección para trabajar con bases de datos a diferentes escalas sin asumir distribución de normalidad, entregando flexibilidad para manejar relaciones no lineales entre caracteres y clases (Friedl y Brodley, 1997). Por lo mismo es rápido de ejecutar e interpretar (Pal y Mather, 2003). Este modelo también ha sido combinado con técnicas de georreferenciación espacial facilitando el mapeo de superficies para determinar estructuras florísticas (Lees y Ritman, 1991), grupos funcionales de especies (Zhang *et al.*, 2005), implementación de estrategias de manejo de residuos de labranza, distinguiendo entre distintas prácticas con una exactitud de predicción de 89% (Yang *et al.*, 2003). Otros trabajos han reportado una exactitud de predicción entre

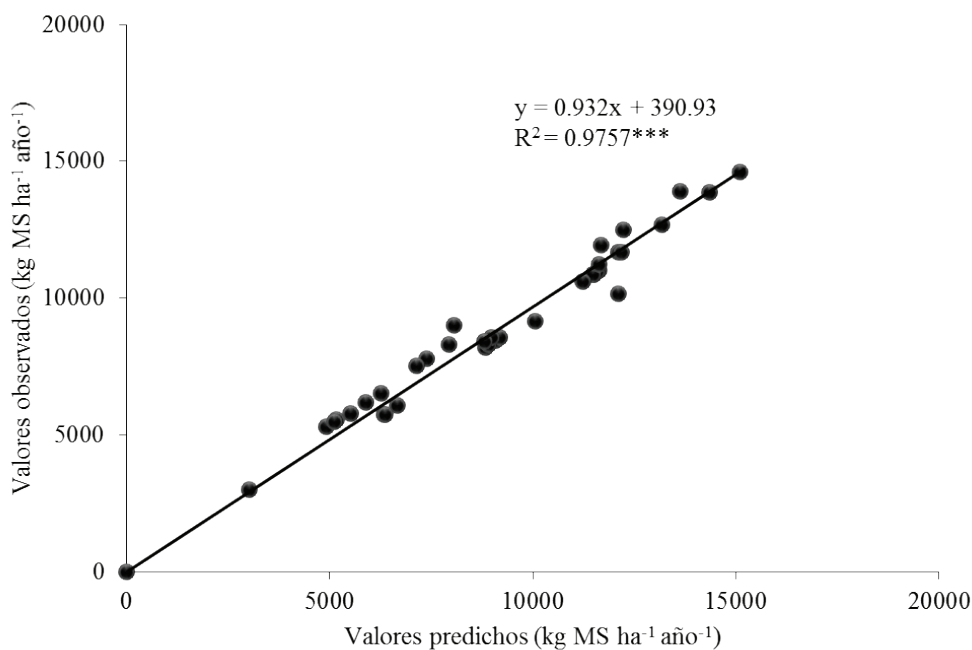


Figura 3. Evaluación del modelo utilizando valores predichos y observado (Flores, 2015).

Figure 3. Model evaluation using predict and observed values (Flores, 2015).

75%-100% en producción de cultivos de maíz (Waheed et al., 2006).

En praderas del Hill country neozelandés Zhang et al., (2006) realizaron un árbol de decisión a partir de un meta-análisis mostrando que las precipitaciones de primavera y la pendiente fueron los factores más influyentes en la producción de materia seca anual. La más alta productividad de materia seca predicha ocurrió cuando las precipitaciones de primavera fueron mayores a 212 mm, la pendiente menor de 16° y el total de fósforo aplicado como fertilizante en 5 años fue mayor de 145 kg. La más baja producción anual predicha ocurrió cuando las precipitaciones de primavera fueron menores de 212 mm, la pendiente mayor a 22.5° y la aplicación de nitrógeno menor de 33.8 mg kg⁻¹.

Desde el punto de vista económico y ambiental el modelo de árbol de decisión ha sido utilizado para determinar la eficiencia de utilización de fertilizante nitrogenado en praderas neozelandesas (Zhang et al., 2007). En este caso el modelo determinó el tiempo de aplicación de fertilizante como la variable más influyente en la eficiencia de utilización de fertilizante nitrogenado siendo los meses de Agosto o Septiembre la mejor época de aplicación. Otras variables como la interacción de las precipitaciones y la temperatura, fósforo Olsen y la pendiente también fueron importantes en la eficiencia de utilización de nitrógeno con una exactitud de la predicción del 69%.

Otras utilidades del modelo de árbol de decisión han sido para determinar cobertura de suelo en bosque tropical (Helmer et al., 2008), erosión del suelo en base a variables medioambientales, tipo de suelo o cobertura (Geissen et al., 2007), simulación espacio temporal del cultivo a bajo costo (Sorel et al., 2010), entre otros.

Ventajas del Modelo de árbol de decisión

Estadísticamente el modelo de árbol de decisión tiene varias ventajas en relación a los enfoques de otros modelos. La principal es que permite manejar variables cualitativas y cuantitativas dentro del mismo modelo (Wu, 2009), no necesita asumir una distribución normal de la variable dependiente con lo cual permite trabajar con una amplia dimensionalidad de la información (DeFries y Chan, 2000; Elnaggar y Noller, 2010; Vayssieres et al., 2000) y es flexible a la falta de información de la base de datos en las variables independientes (Wu, 2009). Tampoco existe problema de multicolinealidad debido a que si existiera una alta correlación entre las variables este se reduce al existir un gran tamaño muestral lo cual es un requisito para desarrollar un árbol de decisión (Vayssieres et al., 2000). El árbol de decisión entrega la importancia de la influencia de una variable independiente y las posibles interacciones con las otras variables revelando interacciones no lineales (DeFries y Chan, 2000).

Desventajas del Modelo de árbol de decisión

La principal desventaja es que los valores de la variable dependiente deben estar completos y en este caso no es flexible con el manejo de la información (Wu, 2009). Al entregar varias alternativas de solución queda al criterio del operario decidir cuál de todas ellas es la mejor (Scheffer, 2002), por lo tanto no existe un árbol de decisión óptimo (Laurance et al., 2004). Por otra parte, los valores predictivos arrojados por el modelo podrían estar afectados por outliers (Friedman, 2001) además de que podría contener variables que estén mejor representadas que otras afectando la salida del modelo con predicciones erróneas (Laurence et al., 2004).

CONCLUSION

La productividad de una pradera está relacionada al medio ambiente, a la utilización eficiente de los recursos y a las decisiones de manejo. La visión en conjunto de todo el sistema es una de las principales dificultades a enfrentar en donde la utilización de modelos ha sido desarrollada como parte de la solución con su primera aparición hace 50 años. La elaboración de un modelo predictivo es a partir de una amplia base de datos en donde el modelo de árbol de decisión surge como una alternativa predictiva dentro del modelamiento ambiental por ser altamente preciso en sus predicciones. El modelo de árbol de decisión ha sido una herramienta útil de fácil utilización e interpretación en el manejo de la pradera y en la utilización de fertilizantes.

AGRADECIMIENTOS

El estudio fue patrocinado por el Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt), Chile, Proyecto No. 1100957. El autor principal agradece a la Beca Doctorado Nacional de la Comisión Nacional de Investigación Científica y Tecnológica (Conicyt), Chile, a la Dirección de Estudios de Postgrado de la Universidad Austral de Chile, y a la Escuela de Graduados de la Facultad de Ciencias Agrarias de la Universidad Austral de Chile.

REFERENCIAS

- Balocchi, L., Pulido, F., Fernández, V., 2002. Comportamiento de vacas lecheras en pastoreo con y sin suplementación con concentrado. *Agricultura Técnica* 62, 87-98. <http://dx.doi.org/10.4067/S0365-28072002000100009>
- 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.
- Corson, M. S., Rotz, C.A., Skinner, R. H., Sanderson, M. A., 2007. Adaptation and evaluation of the Integrated Farm System Model to simulate temperate multiple-species pas-

- tures. *Agricultural Systems* 94, 502-508. <http://dx.doi.org/10.1016/j.agsy.2007.01.003>
- DeFries, R. S., Chan, J. C. W., 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment* 74, 503-515.
- De Wit, C., 1965. Photosynthesis of leaf canopies. In: *Agricultural Research Report 663*. Wageningen, PUDOC, p 57.
- Dormann, C. F., 2007. Promising the future? Global change projections of species distributions. *Basic and applied ecology* 8, 387-397.
- Duff, T. J., Bell, T. L., York, A., 2012. Patterns of plant abundances in natural systems: is there value in modelling both species abundance and distribution?. *Australian Journal of Botany* 59, 719-733. <http://dx.doi.org/10.1071/BT11017>
- Elnaggar, A. A., Noller, J. S., 2009. Application of remote-sensing data and decision-tree analysis to mapping salt-affected soils over large areas. *Remote Sensing* 2, 151-165.
- Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., Turner, W. R., 2013. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. *Global Ecology and Biogeography* 22, 1007-1018.
- Fernandez, G., 2003. *Data Mining Using SAS Applications*, second ed. Chapman & Hall/CRC, Florida.
- Flores, P., 2015. Interacciones ecosistémicas y cambios en las relaciones prateras al mejorar praderas naturalizadas degradadas con pastoreo ovino intensivo. Tesis Doctorado en Ciencias Agrarias, Universidad Austral de Chile. 214 p.
- Franklin, J., 2010. Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions* 16, 321-330.
- Friedl, M. A., Brodley, C. E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61, 399 - 409.
- Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 29, 1189-1232.
- Gauch, H. G., Hwang, J. T., Fick, G. W., 2003. Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal* 95, 1442-1446.
- Geissen, V., Kampichler, C., López-de Llergo-Juárez, J. J., Galindo-Acántara, A., 2007. Superficial and subterranean soil erosion in Tabasco, tropical Mexico: development of a decision tree modeling approach. *Geoderma* 139, 277-287.
- Guisan, A., Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-186. [http://dx.doi.org/10.1016/S0304-3800\(00\)00354-9](http://dx.doi.org/10.1016/S0304-3800(00)00354-9)
- Helmer, E. H., Kennaway, T. A., Pedreros, D. H., Clark, M. L., Marcano-Vega, H., Tieszen, L. L., Carrington, C. S., 2008. Land cover and forest formation distributions for St. Kitts, Nevis, St. Eustatius, Grenada and Barbados from decision tree classification of cloud-cleared satellite imagery. *Caribbean Journal of Science* 44, 175-198.
- Huntley, B., Barnard, P., Altwegg, R., Chambers, L., Coetzee, B. W., Gibson, L., Willis, S. G., 2010. Beyond bioclimatic envelopes: dynamic species' range and abundance modelling in the context of climatic change. *Ecography* 33, 621-626.
- Johnson, I. R., Chapman, D. F., Snow, V. O., Eckard, R. J., Parsons, A. J., Lambert, M. G., Cullen, B. R., 2008. DairyMod and EcoMod: biophysical pasture-simulation models for Australia and New Zealand. *Animal Production Science* 48, 621-631. <http://dx.doi.org/10.1071/EA07133>
- Jones, J., Keating, B., Porter, C., 2001. Approaches to modular model development. *Agriculture Systems* 70, 421-443.
- Keith, D. A., Akçakaya, H. R., Thuiller, W., Midgley, G. F., Pearson, R. G., Phillips, S. J., Rebelo, T. G., 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters* 4, 560-563.
- Körner, 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, Berlin, pp. 117-140.
- Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote sensing of environment* 90, 331-336.
- Lees, B. G., Ritman, K., 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management* 15, 823-831.
- Lobell, D. B., Ortiz-Monasterio, J. I., Asner, G. P., Naylor, R. L., Falcon, W. P., 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. *Agronomy Journal* 97, 241-249.
- 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. <http://www.tandfonline.com/doi/pdf/10.1080/00288233.2000.9513445>
- Nicholas, P. K., Kemp, P. D., Barker, D. J., 1998. Biodiversity, stability and pasture management—the role of functional groups, in: *Proceedings of the 9th Australian Agronomy Conference*, Wagga wagga, Australia.
- Pal, M., Mather, P. M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment* 86, 554-565. [http://dx.doi.org/10.1016/S0034-4257\(03\)00132-9](http://dx.doi.org/10.1016/S0034-4257(03)00132-9)
- Piñeiro, G., Perelman, S., Guerschman, J. P., Paruelo, J. M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed?. *Ecological Modelling* 216, 316-322. <http://dx.doi.org/10.1016/j.ecolmodel.2008.05.006>
- Prost, L., Cerf, M., Jeuffroy, M. H., 2012. Lack of consideration for end-users during the design of agronomic models. A review. *Agronomy for sustainable development* 32, 581-594.
- Rickert, K.G., Stuth, J. W., McKeon, G.M., 2000. Modeling pasture and animal production. In: Mannelje, L. and Iones, R.M.(Eds.), *Field and laboratory methods for grassland and animal production research*, CAB Publishing, Wallingford, pp. 29-66.
- 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.
- Sinclair, T. R., Seligman, N. A. G., 1996. Crop modeling: from infancy to maturity. *Agronomy Journal* 88, 698-704.

- Sorel, L., Viaud, V., Durand, P., Walter, C., 2010. Modeling spatio-temporal crop allocation patterns by a stochastic decision tree method, considering agronomic driving factors. *Agricultural Systems* 103, 647-655.
- Thuraisingham B., 1999. *Data Mining: Technologies, Techniques, Tools, and Trends*. CRC press New York, USA.
- Vayssières, M. P., Plant, R. E., Allen-Diaz, B. H., 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11, 679-694.
- Waheed, T., Bonnell, R. B., Prasher, S. O., Paulet, E., 2006. Measuring performance in precision agriculture: CART—A decision tree approach. *Agricultural water management* 84, 173-185. <http://dx.doi.org/10.1016/j.agwat.2005.12.003>
- Wan, L., Zhang, B., Kemp, P., Li, X., 2009. Modelling the abundance of three key plant species in New Zealand hill-pasture using a decision tree approach. *Ecological Modelling* 220, 1819-1825. <http://dx.doi.org/10.1016/j.ecolmodel.2009.04.031>
- White, J., Hodgson, J., 2000. *New Zealand pasture and crop science*. Oxford University Press.
- Wu, D., 2009. Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications* 36, 9105-9112.
- 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. [http://dx.doi.org/10.1016/S0308-521X\(02\)00051-3](http://dx.doi.org/10.1016/S0308-521X(02)00051-3)
- Zhang, B., Valentine, I., Kemp, P. D., 2005. A decision tree approach modelling functional group abundance in a pasture ecosystem. *Agriculture Ecosystem and Environment* 110, 279-288. <http://dx.doi.org/10.1016/j.agee.2005.04.017>
- Zhang, B., Valentine, I., Kemp, P., Lambert, G., 2006. Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographical information system. *Agricultural Systems* 87, 1-17. <http://dx.doi.org/10.1016/j.agry.2004.10.003>
- Zhang, B., Tillman, R., 2007. A decision tree approach to modelling nitrogen fertiliser use efficiency in New Zealand pastures. *Plant and Soil* 301, 267-278. <http://dx.doi.org/10.1016/j.agee.2005.04.017>