

Patrones de relación entre información biológica e información física y socioeconómica. Cuenca alta del río Otún, Risaralda (Colombia)¹

Relationship Patterns between Biological Information and Physical and Socioeconomic Information: The Otun River Basin in Risaralda (Colombia)²

Padrões de relação entre informação biológica e informação física e socioeconômica. Bacia alta do rio Otún, Risaralda (Colômbia)³

Código SICI: 0123-2126(201206)16:1<265:PRIBIF>2.3.TX;2-5

Yulietb Rossio Coronel-Picón⁴

Nelson Obregón-Neira⁵

Germán Leonardo Jiménez-Romero⁶

¹ Fecha de recepción: 7 de marzo de 2011. Fecha de aceptación: 11 de agosto de 2011. Este artículo se deriva de un proyecto de investigación denominado *Aplicación de modelos de sistemas inteligentes en el estudio de las relaciones y dinámica espacial entre sistemas productivos y sistemas naturales. Caso de estudio para la cuenca media y alta del río Otún-Risaralda-Colombia*, desarrollado por la Pontificia Universidad Javeriana y la Universidad Nacional de Colombia, Bogotá, Colombia.

² Submitted on March 7, 2011. Accepted on August 11, 2011. This article is the result of the research project *Application of Intelligent System Models in the Assessment of the Relations and Spatial Dynamics between Productive Systems and Natural Systems. Case Study for the Middle and Upper Otun River, Risaralda, Colombia*. Developed by Pontificia Universidad Javeriana and Universidad Nacional de Colombia, Bogotá, Colombia.

³ Data de recepção: 7 de março de 2011. Data de aceitação: 11 de agosto de 2011. Este artigo deriva de um projeto de pesquisa denominado *Aplicação de modelos de sistemas inteligentes no estudo das relações e dinâmica espacial entre sistemas produtivos e sistemas naturais. Caso de estudo para a bacia média e alta do rio Otún-Risaralda-Colômbia*, desenvolvido pela Pontificia Universidade Javeriana e a Universidade Nacional da Colômbia, Bogotá, Colômbia.

⁴ Ingeniera civil, Universidad Francisco de Paula Santander, Cúcuta, Colombia. Magíster en Hidrosistemas, Pontificia Universidad Javeriana, Bogotá, Colombia. Consultora especialista del Instituto de Hidrología, Meteorología y Estudios Ambientales (Ideam), Bogotá, Colombia. Correo electrónico: ycoronel@javeriana.edu.co.

⁵ Ingeniero civil, Universidad Francisco de Paula Santander, Cúcuta, Colombia. Magíster en Ingeniería Civil (énfasis Ingeniería de Recursos Hídricos y Ambiental), Universidad de los Andes, Bogotá, Colombia. PhD en Ciencias Hidrológicas, University of California at Davis, Estados Unidos. Profesor titular de la Facultad de Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia. Correo electrónico: nobregon@javeriana.edu.co.

⁶ Biólogo, Universidad de los Andes, Bogotá, Colombia. Magíster en Manejo y Conservación de la Biodiversidad, Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica. Candidato a Doctor en Ciencias Biológicas, Pontificia Universidad Javeriana, Bogotá, Colombia. Profesor e investigador, Departamento de Biología, Facultad de Ciencias, Pontificia Universidad Javeriana. Correo electrónico: german.jimenez@javeriana.edu.co.

Resumen

En este artículo se transforma la información biofísica y socioeconómica de la cuenca del río Otún (Colombia), con el objetivo de identificar patrones de relación entre ambos tipos de información. La identificación de patrones se llevó a cabo mediante sistemas de información geográfica (ArgGis 9.1®), el desarrollo de código para evaluación de relaciones (Matlab 7.1®) y minería de datos basada en árboles de decisión, utilizando el algoritmo J-48, desarrollado en el *Waikato Environment for Knowledge Analysis* (WEKA). La identificación de patrones permitió concluir que, por lo menos, tres variables físicas (altitud, precipitación y temperatura) y una variable socioeconómica (usos del suelo) influyen en la presencia de los mamíferos. Esta información servirá para tomar decisiones sobre manejo y conservación de la fauna en esta cuenca.

Palabras clave

Minería de datos, sistemas de información geográfica, río Otún (Risarlalda, Colombia), árboles de decisión.

Abstract

Biophysical and socioeconomic data were gathered at the basin of the Otun River (Colombia) and then processed with the objective of identifying relationship patterns between the two types of information. This task was carried out by using Geographic Information Systems software (ArgGis 9.1®), the development of a code for the assessment of relationships (Matlab 7.1®), and data mining tools based on trees; using the J-48 algorithm developed at the Waikato Environment for Knowledge Analysis (WEKA). The identification of patterns allowed concluding that, at least, three physical variables (altitude, rainfall and temperature) and a socioeconomic variable (land use) influence the presence of mammals. This information will be employed in other studies related to Decision Support Systems in the realm of management and conservation of wildlife at the basin.

Keywords

Data mining, geographic information systems, Otun river (Risarlalda, Colombia), decision trees.

Resumo

Neste artigo transforma-se a informação biofísica e socioeconômica da Bacia do rio Otún (Colômbia), com o objetivo de identificar padrões de relação entre ambos os tipos de informação. A identificação de padrões foi realizada mediante sistemas de informação geográfica (ArgGis 9.1®), o desenvolvimento de código para avaliação de relações (Matlab 7.1®) e mineração de dados baseada em árvores de decisão, utilizando o algoritmo J-48, desenvolvido no *Waikato Environment for Knowledge Analysis* (WEKA). A identificação de padrões permitiu concluir que, pelo menos, três variáveis físicas (altitude, precipitação e temperatura) e uma variável socioeconômica (usos do solo) influem na presença dos mamíferos. Esta informação servirá para tomar decisões sobre manejo e conservação da fauna nesta Bacia.

Palavras chave

Mineração de dados, sistemas de informação geográfica, rio Otún (Risarlalda, Colômbia), árvores de decisão.

Introducción

Una forma de entender cómo las sociedades humanas deterioran el ambiente natural del cual dependen es tratando de encontrar dónde las actividades socioeconómicas se relacionan con el ambiente biofísico. El modelo conceptual sobre el que se basa la anterior afirmación se origina en el hecho de que las sociedades humanas emplean el conocimiento proveniente de la naturaleza (información biofísica) para diseñar sus sistemas productivos. Estos, a su vez, dependen del conocimiento que surge de modelos económicos y sociales que son transformados a prácticas productivas para el mejoramiento de la calidad de vida (información socioeconómica) (Jiménez, 2010; Primack et ál., 2001; Lui et ál., 2007; Frimpong et ál., 2006).

Esta interacción entre formas de conocimiento lleva al surgimiento de diferentes tipos de relaciones entre naturaleza y humanos, siendo algunas de ellas conflictivas (Liu, 2001; Primack et ál., 2001; Liu et ál., 1999). Estos conflictos se ven reflejados como amenazas a la biodiversidad, que anidan factores causales (relaciones entre sistemas biofísicos y socioeconómicos) y sobre los cuales se actúa por medio de estrategias de manejo y conservación de la biodiversidad (Salafsky et ál., 2002; Primack et ál., 2001; Low et ál., 1999; Soulé, 1991). Ya que las estrategias son la forma de acción, estas precisan de un mejor entendimiento de las relaciones para ser más efectivas.

Dentro de las estrategias se encuentra el ordenamiento para el manejo de cuencas, que para el caso de Colombia se halla reglamentado en el Decreto 1729 del 2002. Este busca promover la planificación y uso sostenible, así como la ejecución de programas y proyectos específicos dirigidos a conservar, preservar, proteger o prevenir el deterioro o restaurar las cuencas, dado que estas son consideradas la unidad ideal de gestión donde se da una dinámica de relaciones entre los sistemas biofísico y socioeconómico.

La fauna silvestre es particularmente sensible a las relaciones mencionadas. Una amplia gama de factores, provenientes de cada uno de los sistemas, influyen

en su presencia, abundancia y distribución, ya que muchos de estos funcionan en diferentes escalas espaciales y temporales (Bolen y Robinson, 2003; Liu et ál., 1999). Dentro de los factores provenientes del sistema biofísico se encuentran la presencia de otras especies, la altitud, la pendiente, la precipitación, la temperatura, los cuerpos de agua y las coberturas naturales presentes (Ricklefs y Miller, 2000; Jiménez, 2003; Mackey y Lindenmayer, 2001).

Dentro de los factores provenientes del sistema socioeconómico se encuentran las formas de uso del suelo, ya sean estas para conservación (áreas protegidas) o para producción (rastrojos, bosque plantado, pasto manejado, cultivos, vivienda y las vías) (Primack et ál., 2001; Jiménez, 2003 y 2000). Hasta aquí se puede decir que, en cuanto la presencia de la fauna se relacione más estrechamente con factores provenientes de uno u otro sistema, estos tendrán efecto en su abundancia y distribución, así como para su persistencia en tiempo y espacio.

Identificar e interpretar de qué manera se dan relaciones dentro de estos sistemas complejos hace necesario usar herramientas que permitan integrar y reconocer patrones que puedan ser utilizados como soporte para la implementación de acciones de conservación más exitosas (Lui et ál., 2007). El creciente desarrollo de la informática, la inteligencia artificial, los sistemas de información, la minería de datos, la adquisición de datos, la gestión/extracción/visualización de datos, entre otros, permiten hoy día la posibilidad de utilizar herramientas que facilitan la identificación y la interpretación de estas relaciones, a partir del empleo de modelos guiados por datos, basados en el uso de herramientas informáticas y matemáticas. En este sentido, es común encontrar aplicaciones de computador desarrolladas desde el punto de vista teórico, pues los conceptos se fundamentan en tecnologías de reconocimiento y patrones, aprendizaje de máquina (*machine learning*), teoría de la información, teoría de control y recuperación de la información.

Con este contexto, la minería de bases de datos permite extraer conocimiento implícito, previamente desconocido, patrones y modelos potencialmente útiles a partir de bases de datos (Kantardzic y Zurada, 2006). Una forma de hacerlo implica el uso de árboles de decisión, que corresponden a un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final que se vaya a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas (Hernández et ál., 2004).

Es un método flexible que puede manejar un gran número de variables y complejas interacciones entre ellas, cuyos resultados son fácilmente interpretables

para cualquier persona (Pérez, 2007). Para elegir qué atributos y en qué orden aparecen en el árbol, se utiliza una función de evaluación llamada *ganancia de información* (reducción de entropía del conjunto al clasificar usando un determinado atributo). Algunos trabajan con atributos nominales únicamente, como el ID3, y otros trabajan con atributos numéricos, como el C4.5 —J48 en *Waikato Environment for Knowledge Analysis* (WEKA)—. Para determinar cuál atributo se ramifica en cada nivel se calcula la información ganada al discriminar con cada atributo y se usa aquel que maximice la ganancia de información. De acuerdo con la teoría de la información, la información se maximiza cuando la entropía se minimiza. Por lo tanto, la ganancia de información podrá calcularse con la disminución de la entropía ($I\{S,at\}$), es decir:

$$I(S, at) = X(S) - X(S, t) \quad (1)$$

Donde $X(S)$ es el valor de la entropía a priori antes de realizar la subdivisión y $X(S, t)$ es el valor de la entropía del sistema de subconjuntos generados. La entropía a priori y para el sistema de subconjuntos es estimada a partir de la siguiente expresión:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2)$$

Donde p_i representa la probabilidad de que un ejemplo tomado al azar pertenezca a la clase i entre un número c de clases probables. p_i se infiere estadísticamente a partir de la frecuencia de los datos de dicha clase. Una manera de evaluar los árboles de decisión es la matriz de confusión. Mayores detalles de estos algoritmos de clasificación pueden encontrarse en Mitchell (1997).

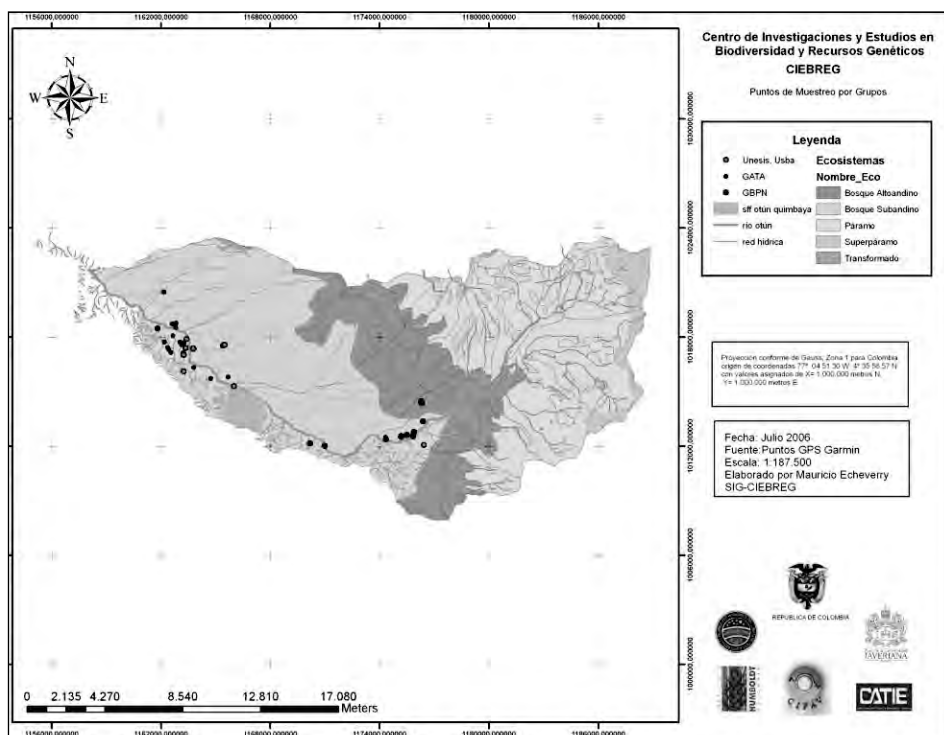
Al considerar las posibilidades para la solución de problemas complejos que la combinación de estas técnicas informáticas y matemáticas ofrecen, en esta investigación se exploró la identificación de las patrones de relación entre la presencia de algunas especies de mamíferos con las características físicas y socioeconómicas de la cuenca del río Otún. Este problema no se ha orientado extensivamente en la literatura especializada y para el caso específico de la región de estudio se espera que los resultados faciliten la concepción e implementación de estrategias de conservación en el marco de los sistemas de soporte de toma de decisiones para el manejo y conservación de la fauna en esta cuenca.

1. Materiales y métodos

1.1. Área de estudio

La cuenca del río Otún se ubica al noroccidente de Colombia, en la cordillera Central de los Andes, al sur del departamento de Risaralda, en las coordenadas 4° 49' latitud norte y 75° 42' longitud oeste. Tiene un área aproximada de 483,13 km² y forma parte de la Gran Cuenca del Río Cauca. En la parte alta de la cuenca se localiza un complejo de áreas protegidas, conformado por el Parque Regional Natural Ucumarí, el Parque Nacional Natural los Nevados, el Parque Municipal Natural Campoalegre y el Santuario de Flora y Fauna Otún Quimbaya. La zona de estudio se localiza en la cuenca media del río Otún entre los 1400 y los 2300 msnm, donde los principales sistemas productivos son la agricultura, la ganadería y la explotación forestal (figura 1).

Figura 1. Cuenca media y alta del río Otún, escala 1:187500



Fuente: SIG-Ciebeg.

1.2. Metodología

Para la identificación de patrones se llevaron a cabo las siguientes tareas: recopilación de la información, definición de variables, integración de la información, procesamiento y análisis de resultados. Una descripción esquemática de la metodología general empleada en la actual investigación se presenta en la figura 2.

1.2.1 Recopilación de información

Se recopiló información biológica, física y socioeconómica de la cuenca del río Otún suministrada por el Centro de Investigaciones y Estudios en Biodiversidad y Recursos Genéticos (Ciebrég), la Pontificia Universidad Javeriana y la Federación Nacional de Cafeteros. La información incluye: 1) curvas de nivel de la cuenca media y alta del río Otún cada 50 m (formato shp); 2) información asociada a temperatura, humedad relativa y precipitación media mensual de las estaciones localizadas en la cuenca del río Otún y cuencas cercanas (periodo 2004-2007); 3) mapas escala 1:25 000 con información sobre usos del suelo, ríos, áreas protegidas, vías para la cuenca media y alta del río Otún de 2005 (formato shp), y 4) información para la cuenca media y alta del río Otún en formato de indicios sobre presencia o ausencia de especies de mamíferos medianos y grandes (huellas y avistamientos de individuos de diferentes especies).

1.2.2. Definición de variables biofísicas y socioeconómicas

Para la definición de las variables se identificaron las relevantes de procesos y sistemas naturales (variables biofísicas); entre tanto, aquellas de naturaleza socioeconómicas fueron aquellas relacionadas con factores del sistema que son producto de acciones antrópicas. Para el ejercicio de integración se adaptaron diversas variables mediante funciones *3D Analyst* y *Spatial Analyst* (ArcGis 9.1[®]) sobre una malla con tamaño de celda de 30 m por 30 m. Adicionalmente, para el preprocesamiento de otro grupo de variables se desarrollaron aplicaciones en el lenguaje de programación Matlab[®] 7.1. Finalmente, los datos sobre fauna silvestre fueron transformados utilizando el estimador Kernel (Zhang y Hyndman, 2004), asociado a los paquetes GenKern y Kern Smooth, en ambiente informático estadístico R.

1.2.3. Integración de la información

Las capas de información generadas en el paso anterior se asociaron en una misma tabla, utilizando la herramienta de análisis para ArgGis 9.1[®]. De esta manera, cada celda del terreno fue representada por una fila de la tabla de datos generada.

1.2.4. Discretización de las variables

A partir de esta tabla de datos generada, de la cual se obtuvieron nueve variables de entrada y una de salida, se definieron tres modelos sometidos a variaciones en cuanto a la discretización de las variables. De estos se eliminaron las filas con valores duplicados. Para todos los modelos se utilizó la siguiente discretización en los usos del suelo: Bosque Natural (BN), Bosque Secundario (BS), Rastrojo (R), Bosque Plantado (BP), Pasto Natural (PN), Pasto Manejado (PM), Cultivos (C), Vivienda (V), Otros (O, lagunas y nieve).

1.2.5. Minería de bases de datos

Para cada uno de los modelos planteados se generaron árboles de decisión utilizando el algoritmo de clasificación *J-48* de la herramienta de minería de datos WEKA. Una vez se ejecutó el algoritmo sobre los datos, los resultados de la evaluación revelaron las variables biofísicas y socioeconómicas más estrechamente relacionadas con la presencia de los mamíferos.

2. Resultados y discusión

2.1. Definición de variables biofísicas y socioeconómicas

Teniendo en cuenta los criterios expuestos en la metodología, se definieron las siguientes variables biofísicas: a) presencia de mamíferos; b) altitud media; c) pendiente; d) precipitación media anual multianual; e) temperatura media anual multianual; f) ríos, y g) usos del suelo como nieve, lagunas, bosque natural, bosque secundario y pasto natural. Igualmente se definieron las siguientes variables socioeconómicas: a) áreas protegidas, b) vías y c) usos del suelo como rastrojo, bosque plantado, pasto manejado, cultivos y vivienda. Estas variables fueron el resultado de las siguientes consideraciones y análisis. La presencia de mamíferos fue tratada para inferir el área donde posiblemente podría existir presencia de estos, a partir de los registros disponibles (densidad de probabilidad) y para que esta área sirviera para los posteriores análisis de relaciones con las demás variables. Para este tratamiento se utilizó el estimador de Kernel o estimador

de núcleo bivariado gaussiano. Para la aplicación del estimador se desarrolló un programa de computador corto (*script*) para dentro del *software* estadístico R, en el que se utilizaron los paquetes funcionales *GenKern* y *Kern Smooth*.

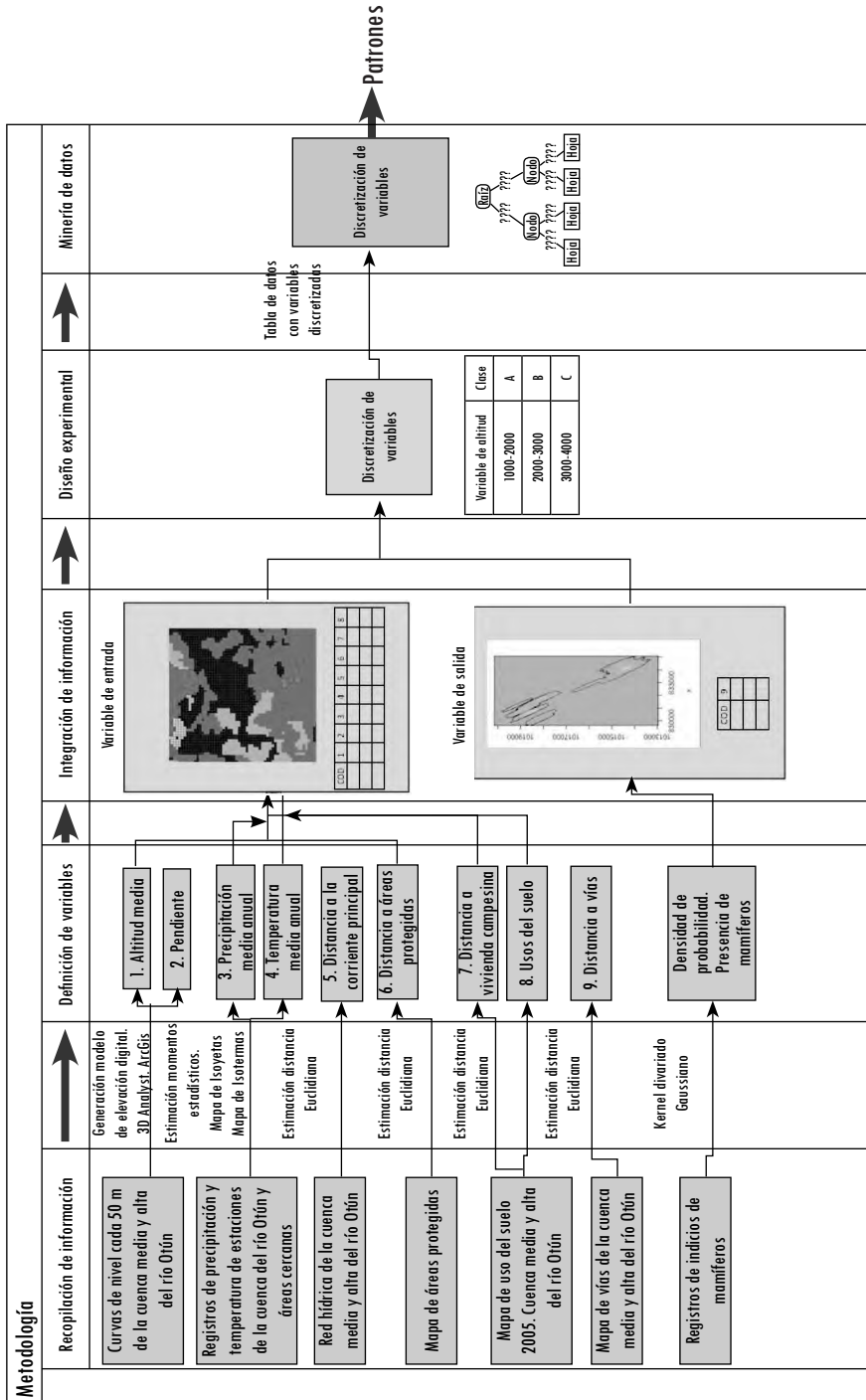
La *altitud media* se obtuvo por medio de un modelo digital del terreno generado a partir de las curvas de nivel de la cuenca media y alta del río Otún. La *pendiente* se estimó, tomando como tamaño de pixel 30 m, a partir del modelo digital de elevación generado; para realizar estas dos tareas se empleó la herramienta *3D Analyst* (ArcGis 9.1[®]). La *precipitación media anual multianual* y la *temperatura media anual multianual* se obtuvieron a través de un proceso de interpolación del valor de estas para cada celda de la grilla generada. En ambos procedimientos se utilizó la herramienta (ArcGis 9.1[®]).

Las capas temáticas de áreas protegidas, ríos y vías se generaron a partir de los mapas disponibles para la cuenca del río Otún. Estos mapas fueron transformados, junto con la variable *uso del suelo para vivienda*, a variables de distancia, utilizando la distancia euclidiana mediada por el desarrollo de un código en Matlab[®] 7.1, el cual permitió estimar la distancia entre la celda objetivo (registro de fauna) y la celda más cercana (área protegida, río, vía y vivienda). *Usos del suelo* se generó a partir de una reclasificación del mapa de coberturas y usos del suelo del 2005 en las categorías: Bosque Natural (BN), Bosque Secundario (BS), Rastrojo (R), Bosque Plantado (BP), Pasto Natural (PN), Pasto Manejado (PM), Cultivos (C), Vivienda (V), Otros (O, por ejemplo: lagunas y nieve).

2.2. Integración de la información

Como resultado de la asociación entre capas de información, se generó una tabla conformada por nueve variables de entrada y una de salida (tabla 1).

Figura 2. Metodología para la identificación de patrones de relación entre variables biológicas, físicas y socioeconómicas



Fuente: presentación propia de los autores.

Tabla 1. Ejemplo de las capas de información generadas (para el caso de BN). Se incluye la asociación entre capas con nueve variables de entrada: altitud (m), pendiente (%), precipitación (mm/año), temperatura (°C), uso_suelo, d_vía, d_río, d_ap, d_vivi, y una variable de salida (Pmamíferos)

Núm.	x	y	Altitud	Pendiente	Precipitación	Temperatura	Uso_suelo	D_vía	D_río	D_ap	D_vivi	Pmamíferos
1	1167000	1014380	1960	58	1630	12	BN	199	365	21	3066	5788543290
2	1166970	1014410	1940	51	1630	12	BN	185	357	21	3033	5740833010
3	1167030	1014350	1955	56	1630	12	BN	202	364	21	3099	5455458930
4	1166940	1014440	1935	60	1630	12	BN	174	347	21	3000	5294260160
5	1167030	1014320	1955	74	1630	12	BN	227	386	21	3103	5171084020
6	1167000	1014350	1970	67	1630	12	BN	223	388	21	3069	5012606070

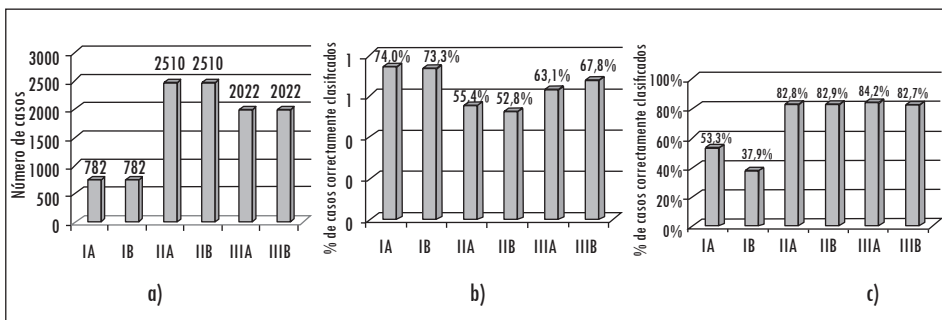
BN: Bosque Natural.

Fuente: presentación propia de los autores.

2.3. Discretización de las variables y minería de bases datos

Se evaluaron tres modelos en función de la discretización de las variables. En el primer modelo, tanto las variables de entrada como las de salida fueron discretizadas en tres clases; en el segundo, en cinco clases, y en el tercero, las variables de entrada fueron discretizadas en cinco clases y las de salida en tres. El número de casos generado fue de 30.383; sin embargo, luego de discretizar las variables y eliminar los valores duplicados, el número de casos utilizados en cada modelo fue de 782 para el modelo I, 2510 para el modelo II y 2022 para el modelo tres III (figura 3a). El porcentaje de casos correctamente clasificados dentro de cada uno de los modelos fue de 74% para el modelo I, 55,4% para el modelo II y 63,10% para el modelo III (figura 3b). Sin embargo, al evaluar el número de casos que son correctamente clasificados dentro de la categoría A de la variable presencia de mamíferos, se observó que el porcentaje de casos correctamente clasificados es superior al 80% para los modelos II y III (figura 3c).

Figura 3. Histograma de patrones empleados en los diferentes casos: a) número total de casos, b) porcentaje de casos correctamente clasificados y c) porcentaje de casos correctamente clasificados contra la variable presencia de mamíferos



Fuente: presentación propia de los autores.

Se utilizaron árboles de decisión en cada uno de los modelos como herramientas de clasificación y, a la vez, como dispositivos para la identificación de leyes de causalidad. En el modelo I fueron incluidas seis de las variables definidas inicialmente: altitud, precipitación, temperatura, distancia a vía, distancia a río y distancia a vivienda. La variable altitud fue incluida en el 92% de los casos correctamente clasificados, en la cual el 57% fue clasificado dentro de la categoría A, que tomó valores entre 1600 y 1835 msnm, y el 45% fue clasificado dentro de la categoría B, cuyos valores variaron entre 1835 y 2070 msnm. La variable precipitación fue incluida en el 100% de los casos correctamente

clasificados y estos se concentraron en la categoría B, que presentó valores de 1680 mm/año a 1770 mm/año.

El 100% de los casos clasificados correctamente incluyeron la variable temperatura, clasificada dentro de la categoría C con un valor de 14 °C. La variable distancia a vía fue incluida en el 63% de los casos correctamente clasificados, el 47% fue clasificado dentro de categoría A (0-500 m), y el 16%, dentro de la categoría B (50-1000 m). La variable distancia a río fue incluida en el 12% de los casos correctamente clasificados dentro de categoría C (1000-1800 m). La variable distancia a vivienda fue incluida en el 53% de los casos correctamente clasificados, el 41% se clasificó dentro de la categoría A (0-2000 m), y el 12%, dentro de la categoría B (2000-4000 m).

En el modelo II fueron incluidas seis de las variables definidas inicialmente. La variable altitud fue incluida en el 5% de los casos correctamente clasificados; presentó un valor entre 1740 y 1880 msnm, que correspondió a la categoría B. La variable precipitación fue incluida en el 94% de los casos correctamente clasificados, el 4% dentro de la clase A (1590-1644 mm/año), el 52% dentro de la clase B (1644-1698 mm/año) y el 37% en la clase C (1698-1752 mm/año). La variable distancia a vía fue incluida en el 38% de los casos correctamente clasificados, el 18 en la clase A (0-400 m) y el 16 % en la clase B (400-800 m).

La variable distancia a río fue incluida en el 4% de los casos dentro de la clase A (0-360 m). La variable distancia a áreas protegidas fue incluida en el 100% de los casos correctamente clasificados dentro de la clase A (0-720 m) y la variable distancia a vivienda fue incluida en el 80% de los casos correctamente clasificados, el 21% dentro de la clase A (0-1200) y el 52% dentro de la clase B (1200-2400).

En el modelo III fueron incluidas seis de las variables definidas al principio. La variable altitud fue incluida en el 6% de los casos correctamente clasificados dentro de la clase B (1740-1880 msnm). La variable precipitación fue incluida en el 91% de los casos correctamente clasificados, el 62% dentro de la clase B (1644-1698 mm/año). La variable distancia a vía fue incluida en el 23% de los casos correctamente clasificados, el 17% en la clase A (0-400 m) y el 6% en la clase B (400-800 m).

La variable distancia a río fue incluida en el 7% de los casos dentro de la clase A (0-360 m). La variable distancia a áreas protegidas fue incluida en el 100% de los casos correctamente clasificados, el 92% dentro de la clase A (0-720 m) y la variable distancia a vivienda fue incluida en el 83% de los casos correctamente clasificados: el 16% dentro de la clase A (0-1200 m) y el 53% dentro de la clase B (1200-2400 m).

A partir de los anteriores resultados se identificó que: a) las variables pendiente y uso del suelo no fueron incluidas dentro de las reglas generadas; b) la variable altitud fue incluida en los tres modelos evaluados y tomó valores entre 1600 y 2070 msnm; c) la variable precipitación fue incluida en los tres modelos evaluados presentando valores entre 1680 mm/año y 1770 mm/año; d) la variable distancia a vía fue incluida en los tres modelos evaluados, el mayor número de casos evaluados presenta valores inferiores a 500 m; e) la variable distancia a río es incluida los tres modelos evaluados, y toma valores menores a 500 m, y f) la variable distancia a áreas protegidas es incluida en dos de los modelos evaluados y en estos es incluida en el 100% de los casos correctamente clasificados.

De lo anterior se puede inferir que, dado que las variables altitud, precipitación, temperatura, distancia a vía, distancia a áreas protegidas y distancia a vivienda son incluidas dentro de las reglas en que se presenta más del 80% de los casos correctamente clasificados, estas se convierten en las variables más relacionadas con la presencia de los mamíferos. Por otro lado, las zonas cercanas a áreas protegidas, de acuerdo con el modelo distancias menores a 720 m, tienden a ser utilizadas por los mamíferos identificados en la cuenca alta del río Otún.

Es de anotar que las áreas protegidas pueden representar para los mamíferos identificados en la cuenca un área de refugio o mejor fuente de recursos. Igualmente, cabe anotar que las variables distancia a vivienda y a vías son consideradas en todos los casos evaluados. Para el caso distancia a vivienda, el mayor porcentaje de casos presenta valores entre 2000 y 4000 m; sin embargo, en los tres modelos se identificaron casos en los que la distancia es inferior a 2000 m. Según esta información, los asentamientos humanos, representados en la vivienda, así como las vías, ejercen algún grado influencia en la presencia de los mamíferos. Este hecho es importante, ya que la presencia de las comunidades humanas y sus obras de infraestructura podrían estar obligando a la fauna a adaptarse y, por consiguiente, a modificar sus patrones de abundancia y distribución, ya que existe evidencia de ello en estos tipos de paisaje (Liu et ál., 1999 y 2007; Bolen y Robinson, 2003; Mackey y Lindenmayer, 2001; Primack et ál., 2001; Jiménez, 2000 y 2003), situación que lleva a pensar que los patrones de relación encontrados aquí podrían ayudar a generalizar modelos de comportamiento de la fauna bajo situaciones de intervención antrópica

3. Conclusiones y comentarios finales

Se ha concebido e implementado una metodología tendiente a identificar patrones de relación para la cuenca del río Otún entre diferentes tipo de información,

tanto socioeconómica como biofísica y con el ánimo de buscar causalidades para la presencia o ausencia de mamíferos. Los resultados sustentan la idea de avanzar hacia la construcción de herramientas que respalden las decisiones en materia de manejo de recursos naturales y de la flora y la fauna. Para esta zona de estudio, por lo menos, tres variables físicas: 1) la altitud, 2) la precipitación y 3) la temperatura; así mismo, una variable socioeconómica (distancia áreas protegidas, distancia vías y distancia a vivienda clasificadas dentro de usos del suelo), influyen en la presencia de los mamíferos y posiblemente en su abundancia y distribución.

Referencias

- BOLEN, E. G. y ROBINSON, W. L. *Wildlife ecology and management*. 5th ed. Englewood Cliffs: Prentice Hall, 2003.
- FRIMPONG, E. A.; ROSS-DAVIS, A. L.; LEE, J. G. y BROUSSARD, S. R. Biophysical and socioeconomic factors explaining the extent of forest cover on private ownerships in a Midwestern (USA) agrarian landscape. *Landscape Ecology*. 2006, vol. 21, pp. 763-776.
- HERNÁNDEZ, J.; RAMÍREZ, M. y FERRIA C. *Introducción a la minería de datos*. Madrid: Pearson Educación, 2004.
- JIMÉNEZ, G. Construcción de un marco conceptual y metodológico para estrategias en manejo y conservación de la biodiversidad en paisajes fragmentados colombianos. 309-329 En: Rodríguez Rodríguez, J. (Ed.). *Elementos para una teoría del desarrollo territorial*. Manizales: Universidad Autónoma de Manizales, 2010.
- JIMÉNEZ, G. Estrategia metodológica para el diseño y evaluación de corredores biológicos: un estudio en Costa Rica. En: Polanco-Ochoa, R. (Ed.). *Manejo de fauna silvestre en Amazonía y Latinoamérica*. Bogotá: Fundación Natura, 2003, pp. 103-107.
- JIMÉNEZ, G. *Propuesta metodológica en el diseño y evaluación de un corredor biológico en la reserva forestal Golfo Dulce, Costa Rica*. Trabajo de posgrado, Centro Agronómico Tropical de Investigación y Enseñanza CATIE, Programa de Educación para el Desarrollo y la Conservación, 2000.
- KANTARDZIC, M. y ZURADA, J. (Eds.). *Next generation of data-mining applications*. s. l.: Chapman and Hall, 2006.
- LIU, J. Integrating ecology with human demography, behavior, and socioeconomics: Needs and approaches. *Ecological Modelling*. 2001, vol. 140, pp. 1-8.
- LIU, J.; DIETZ, T.; CARPENTER, S. R.; ALBERTI, M.; FOLKE, C. et ál. Complexity of coupled human and natural systems. *Science*. 2001, vol. 317, núm. 5844, pp. 1513-1516.
- LIU, J.; OUYANG, Z.; TAYLOR, W. W.; GROOP, R.; TAN, Y. y ZHANG, H. A framework for evaluating the effects of human factors on wildlife habitat: the case of giant pandas. *Conservation Biology*. 1999, vol. 13, núm. 6, pp. 1360-1370.

- LOW, B.; CONSTANZA, R.; OSTROM, E.; WILLSON, J. y SIMON, C. P. Human ecosystem interactions: a dynamic integrated model. *Ecological Economics*. 1999, vol. 34, pp. 227-242.
- MACKEY, B. y LINDENMAYER, D. Towards a hierarchical framework for modeling the spatial distribution of animals. *Journal of Biogeography*. 2001, pp. 1147-1166.
- MITCHELL, T. *Machine learning*. New York: McGraw Hill, 1997.
- PÉREZ, C. *Minería de datos: técnicas y herramientas*. Madrid: International Thomson, 2007.
- PRIMACK, R.; ROZZI, R.; EISINGER, P.; DIRZO, R. y MASSARDO, F. *Fundamentos de conservación biológica: Perspectivas latinoamericanas*. México: Fondo de Cultura Económica de México, 2001.
- RICKLEFS, R. E. y MILLER, G. L. *Ecology*. 4th Ed. New York: W.H. Freeman, 2000.
- ROJAS, C. *Introducción a la gestión integrada de recursos hídricos* [documento en línea]. <http://publicacion05.unipamplona.edu.co/hidroinformatica/portal/home_1/rec/arc_707.pdf> [Consultado: 1-11-2008].
- RUIZ, D. *Heurísticas de selección de atributos para datos de gran dimensionalidad* [documento en línea]. 2006. <<http://www.lsi.us.es/docs/doctorado/tesis/tesisRoberto.pdf>> [Consultado: 20-10-2008].
- SALAFSKY, N.; MARGOLUIS, R.; REDFORD, K. H. y ROBINSON, J.G. Improving the practice of conservation: a conceptual framework and research agenda for conservation science. *Conservation Biology*. 2002, vol. 16, núm. 6, pp. 1469-1479.
- SOULÉ, M. E. Conservation tactics for a constant crisis. *Science*. 1991, vol. 253, pp. 744-750.
- ZHANG, X.; KING, M. L. y HYNDMAN, R. J. *Bandwidth selection for multivariate Kernel density estimation using MCMC* [documento en línea]. 2004. <<http://repec.org/esAUSM04/up.1603.1077410300.pdf>> [Consultado 10-11-2008].