

**UNIVERSIDAD DE CONCEPCIÓN**

**FACULTAD DE INGENIERÍA**

**DEPARTAMENTO DE INGENIERÍA INFORMÁTICA Y  
CIENCIAS DE LA COMPUTACIÓN**



**APLICACIÓN DE MINERÍA DE DATOS ESPACIAL EN UN ESTUDIO DE  
CAMBIO DE USO DE SUELO**

Por

**Guillermo I. Solar Olivares**

Tesis presentada a la  
**ESCUELA DE GRADUADOS  
DE LA UNIVERSIDAD DE CONCEPCIÓN**

Para optar al Grado de  
**MAGÍSTER EN CIENCIAS DE LA COMPUTACIÓN  
CONCEPCIÓN - CHILE**  
2010

## AGRADECIMIENTOS

Agradezco a Dios, por la oportunidad de vivir, por este cuerpo sano y mente capaz de alcanzar esta meta, por brindarme la oportunidad de mejorar cada día. Por guiar mis pasos y mostrarme lo maravilloso de este mundo y poder aportar un granito de arena a su cuidado.

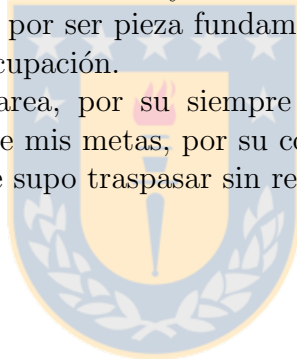
Este estudio se realizó gracias al apoyo del Centro EULA, de la Universidad de Concepción, que desde antes del inicio del programa, apoyó la gestión de recurso humano y material para la investigación. Apoyo brindado por el Dr. Mauricio Aguayo, quien brindara tiempo, guías y perspectiva desde las ciencias ambientales.

Agradezco especialmente a la Dra. Andrea Rodríguez, por su disposición, apoyo, guía y consejos desde mi ingreso al Programa de Magíster en Ciencias de la Computación, así como por su constante incentivo a progresar en mi investigación y formación profesional.

Agradezco muy especialmente a Karla García Bustos, quien desde el pregrado y mi posterior formación de postgrado, ha estado continuamente apoyando y animando a no flaquear en el logro de esta nueva meta, presente en el día a día, aconsejando y aportando con sus conocimientos, acompañando y siempre preocupada, aún así en tiempos difíciles como pareja. Así mismo deseo agradecer a su familia, que para mi fue y será una verdadera segunda familia quienes me brindaron todo su apoyo, cariño y comprensión.

A mis padres Luis Guillermo Solar Berríos y Alicia del Carmen Olivares Olgún, y a mi hermana Macarena Solar Olivares, por ser pieza fundamental en la vida, por todo su apoyo a la distancia, su confianza y preocupación.

Al Dr. Ing. Marcel Szanto Narea, por su siempre buena disposición y tiempo para recibirme y aconsejar en el logro de mis metas, por su confianza y apoyo, por su invaluable experiencia académica que siempre supo traspasar sin restricciones.



## Resumen

El mal manejo del uso de suelo puede amenazar una amplia región geográfica, e incluso externalizar sus efectos. A pesar de que las consecuencias para la vida humana se hacen cada vez más evidentes, el deterioro del suelo no disminuye, incluso en aquellas zonas de naturaleza prístina, y muchas veces alejada de los grandes centros urbanos. Para palear estos efectos, alertas tempranas basadas en cambios históricos de las variables del uso de suelo, sirven como medio para anticipar y prevenir desastres ambientales. Estas alertas se basan en estudios científicos multidisciplinarios, los cuales requieren cada vez de un mayor apoyo de herramientas informáticas para el manejo de información. Ejemplos concretos de apoyo computacional en esta área van desde el uso de sensores remotos, satélites especializados (capaces de captar temperaturas, densidad e imágenes en alta resolución), hasta aplicaciones específicas, algoritmos y procesos automatizados para el análisis de datos ambientales.

El trabajo de esta tesis se enmarca dentro del proyecto FONDECYT Núm. 1070511 (Proyecto “Cambio de uso de suelo en la región nor-patagónica de Chile”) del Centro EULA-Chile (Centro de estudios medioambientales Europa Latinoamérica de la Universidad de Concepción), y tiene como objetivo aplicar técnicas de minado de datos *espacial* que permitan identificar patrones en las variables presentes en un intervalo de tiempo, que describen conforme a sus valores, la dinámica de cambio del uso de suelo en la cuenca del río Aysén. El estudio consiste en descubrir *reglas de asociación espacial* interesantes desde las variables dependientes, determinar el grado de relevancia de estas variables y su asociación respecto a una determinada clase, es decir, cuáles variables promueven cada tipo de uso de suelo (clases objetivos). Además, el estudio contempla la generación de *colocation rules* (en adelante “reglas de colocalidad”) que explotan el concepto de vecindad espacial en la determinación de reglas de asociación que definan el tipo de uso de suelo en un sector en relación a los tipos de uso de suelo de sus vecinos.

Los datos a considerar en el estudio abarcan una ventana de tiempo que va desde 1984 al 2001, e integra información de centros poblados, vías y caminos, curvas de nivel, tipo de uso, entre otros. Esta información proviene de la interpretación de imágenes satelitales, fotografías aéreas y ortofotos, como también, de cartas I.G.M. (Instituto Geográfico Militar) y visitas a terreno. Los resultados del proyecto fueron validados, en cuanto a su grado de interés medioambiental, por un experto de la Unidad de Planificación Territorial del Centro EULA.

Los procesos definidos en esta tesis se separan en los orientados a la obtención de reglas de asociación y en las orientadas a obtener reglas de colocalidad. Para las primeras se trabaja con tablas de minado de datos y un software específico, mientras que para las segundas, se implementó un servidor de bases de datos espaciales, y todos los experimentos fueron ejecutados sobre él.

Como resultados más destacables, se logró determinar un subconjunto de variables que difieren levemente del método usado habitualmente por la Unidad de Planificación Territorial del Centro EULA, de la Universidad de Concepción, ya que se logró incluir en esta muestra significativa, nuevas variables que influyen en la descripción de la dinámica del uso de suelo. Además, se extrajeron reglas de asociación y de colocalidad que fueron validadas por los expertos y que caracterizan la dinámica del uso de suelo.

## Abstract

The bad management of “land cover use” may threaten a large geographic region, and even outsource its effect. Although the consequences for human life are every day more evident, soil degradation is not reduced, even in areas of pristine nature, and often far from major urban centers. To mitigate this effect, early warnings based on historical changes in the predictors of land cover use, these alerts serve like a way to anticipate and prevent environmental disasters. These alerts are based on multidisciplinary scientific studies, which increasingly require greater support in software tools for information management. Specific examples of computer support in this area are the use of remote sensors, specialized satellites (capable of capture temperature, density and high-resolution images) and specific applications, algorithms and automated processes for analyzing environmental data.

The work of this thesis is part of the FONDECYT No. 1070511 project (Project “Land cover use changes in the north-patagonia of Chile”) of the EULA-Chile Center (Europe-Latin American environmental studies Center, Universidad de Concepción, Chile), the objective is apply data mining techniques to identify spatial patterns in the variables presents in a time interval, which describe with their values, the dinamic changes of the land-cover use change in the river Aysen basin. The study try to discover interesting spatial association rules from the dependent variables, determine the relevance of these variables and their association for a particular class (land cover use), ie. what variables promote each type of land use. The study also includes the generation of colocation rules, that rules exploit the concept of neighborhood to determinate spatial association rules that define the type of land use in one sector and its relation to the types of land use from its neighbors.

The data considered in this study is a time lap between 1984 and 2001, and integrates for example: population centers information, ways and roads, level contours, land cover use. This information is obtained with classification techniques of satellite images, aerial photographs and orthophotos, and maps of the Geographic Institute of the Army, and visits to the study area. The project results were validated in their environmental interest by an expert from the EULA center.

The processes defined in this thesis are separated in two parts: the oriented to obtain association rules. And a second part oriented to obtain the colocationrules. For asociation rules, the work use mining tables and a specific software (Weka) for data mining. And for the colocation rules, all the work is based and implemented on a spatial data base system, and all the queries and procces are executed on it.

Some relevant results, it was determined a subset of variables that differ slightly from the method commonly used by the EULA center, like was achieved in this sample include significant new variables that influence the description of the dynamics of land cover use. Furthermore, extracted asociation rules and colocation rules, and were validated by the EULA expert, and that characterize the dynamics of land cover use.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Hipótesis . . . . .	4
1.2. Objetivos . . . . .	4
1.2.1. Objetivo general . . . . .	4
1.2.2. Objetivos específicos . . . . .	5
1.3. Organización de la tesis . . . . .	5
<b>2. Discusión bibliográfica</b>	<b>7</b>
2.1. Bases de datos espaciales . . . . .	7
2.2. Minería de Datos . . . . .	9
2.3. Reglas de asociación espacial . . . . .	12
2.4. Reglas de colocalidad . . . . .	13
2.5. Uso de reglas de asociación para manejo de uso de suelo . . . . .	16
<b>3. Reglas de asociación</b>	<b>19</b>
3.1. Datos y Selección de Variables . . . . .	19
3.1.1. Obtención de datos . . . . .	19
3.1.2. Recopilación y datos de entrada obtenidos . . . . .	20
3.1.3. Preparación de los datos . . . . .	21
3.2. Selección de variables . . . . .	21
3.2.1. Selección de variables según el criterio de condición de cambio del uso de suelo . . . . .	23
3.2.2. Selección de variables según el criterio del tipo de cambio . . . . .	25
3.2.3. Comparación y discusión de la selección de variables . . . . .	26
3.3. Reglas de asociación . . . . .	31
3.3.1. Reglas generadas . . . . .	32
3.3.2. Agrupación de reglas . . . . .	32
3.3.3. Validación de reglas de asociación . . . . .	33
<b>4. Generación de reglas de colocalidad</b>	<b>37</b>
4.1. Obtención de datos . . . . .	39
4.2. Preparación de los datos . . . . .	41
4.2.1. Evaluación de las reglas obtenidas . . . . .	43
4.2.2. Extracción y comparación de reglas . . . . .	50
4.3. Validación de los resultados . . . . .	50

<b>5. Conclusiones y trabajo futuro</b>	<b>57</b>
5.1. Conclusiones . . . . .	57
5.2. Trabajo futuro . . . . .	58
<b>6. ANEXOS</b>	<b>61</b>
6.1. Glosario . . . . .	61
6.2. Abreviaciones de las variables usadas en esta tesis . . . . .	66
6.3. Lista de las Reglas de Asociación obtenidas . . . . .	67
6.4. Códigos Postgres y PostGIS . . . . .	73
6.4.1. Códigos Sql utilizados en creación de tablas de vecindades pares. . . . .	73
6.4.2. Código para el cálculo del porcentaje de adyacencia entre dos polígonos . . . . .	73
6.4.3. Código para generar las distintas tablas, diferenciadas cada una por ser la correspondiente a cada uno de los usos de suelo presente en el área de estudio . . . . .	74
6.5. Códigos para la generación de las reglas de colocalidad . . . . .	75
6.6. Resumen de los valores obtenidos para las métricas $Pr_1$ , $Pr_2$ y $Cp$ . . . . .	77
6.6.1. Razón de participación $Pr_1$ , ecuación 4.1 . . . . .	77
6.6.2. Razón de participación $Pr_2$ , ecuación 4.2 . . . . .	77
6.6.3. Probabilidad condicional $Cp$ , ecuación 4.3 . . . . .	78
6.6.4. Índices de participación (min Pr) para $Pr_2$ , ecuación 4.2 . . . . .	78
6.6.5. Índices de participación (min Pr) para $Pr_1$ , ecuación 4.1 . . . . .	79



# Índice de figuras

1.1. El área de estudio, cuenca del río Aysén. . . . .	2
1.2. Imagen Landsat del área de estudio. . . . .	3
2.1. Las tres formas de elementos espaciales: punto, línea y polígono. . . . .	8
2.2. Ejemplo de datos geográficos almacenados en una BD espacial (fuente: Bogorny 2006) [2]. . . . .	9
2.3. Ejemplos de relaciones espaciales. . . . .	9
2.4. Noción de cluster depende del enfoque del estudio. . . . .	11
2.5. Patrones de colocalidad. . . . .	14
2.6. Modelo concéntrico de eventos. . . . .	15
2.7. Ejemplos de reglas obtenidas desde la figura 2.6 . . . . .	16
2.8. Ejemplo de dos distintas variables usadas en el minado de reglas de asociación. . . . .	17
2.9. Ejemplo de jerarquía de granularidad. . . . .	18
3.1. Mining table definiendo como variable dependiente la condición de cambio o no cambio del uso de suelo. . . . .	21
3.2. Plantaciones cerca de la red vial . . . . .	31
4.1. Modelo concéntrico de puntos y polígonos . . . . .	38
4.2. Vecindad de polígonos con cardinalidad 2 (pares de polígonos) . . . . .	38
4.3. Comparación de dos polígonos, A y B, adyacentes pero donde A es mucho mayor que B. . . . .	39
4.4. Proceso de conversión desde una imagen satelital a raster . . . . .	40
4.5. Un archivo raster convertido en ARCview a shape . . . . .	40
4.6. Muestra de las superficies más pequeñas. . . . .	41
4.7. Ejemplo de un touches $\approx$ 0. . . . .	42
4.8. Participation index para Bosque nativo y Bosque nativo achaparrado . . . . .	44
4.9. Radio de participación para Matorral y Estepa. . . . .	44
4.10. Radio de participación para Praderas y Suelo agrícola. . . . .	44
4.11. Radio de participación para Cuerpos de agua y Nieves y glaciares. . . . .	45
4.12. Radio de participación para Áreas sin vegetación y Humedales. . . . .	45
4.13. Radio de participación para Sombras y Estepa alto andina. . . . .	45
4.14. Radio de participación para Plantaciones y Suelo urbano. . . . .	46

4.15. Índice de correlación lineal. 1: Bosque nativo. 2: Bosque nativo achaparrado. 3: Matorral. 3: Praderas. 5: Praderas. 6: Agrícola. 7: Cuerpos de agua. 8: Nieves y glaciares. 9: áreas sin vegetación. 10: Humedales. 11: Sombras. 12: Estepa alto andina. 13: Plantaciones. 14: Urbano. . . . .	47
4.16. Probabilidad condicional para Bosque nativo y Bosque nativo achaparrado .	47
4.17. Probabilidad condicional para Matorral y Estepa . . . . .	48
4.18. Probabilidad condicional para Praderas y Suelo agrícola . . . . .	48
4.19. Probabilidad condicional para Cuerpos de agua y Nieves y glaciares . . . . .	48
4.20. Probabilidad condicional para Áreas sin vegetación y Humedales. . . . .	49
4.21. Probabilidad condicional para Sombras y Estepa altoandina . . . . .	49
4.22. Probabilidad condicional para Plantaciones y Suelo Urbano . . . . .	50
4.23. Dispersión de las reglas de colocalidad obtenidas. . . . .	54
4.24. Reglas de colocalidad para uso de suelo "1" (bosque nativo): achurado diagonal, y el uso de suelo "2" (bosque nativo achaparrado): achurado con puntos. . . . .	54
4.25. Reglas de colocalidad para uso de suelo "9" (áreas sin vegetación): achurado diagonal, y el uso de suelo "8" (nieves y glaciares): achurado con puntos. . . . .	55
6.1. Resumen de la razón de participación obtenida para $Pr_1$ . . . . .	77
6.2. Resumen de la razón de participación obtenida para $Pr_2$ . . . . .	77
6.3. Resumen de las probabilidades condicionales para las reglas de colocalidad obtenidos . . . . .	78
6.4. Resumen de los índices de participación obtenidos para $Pr_2$ . . . . .	78
6.5. Resumen de los índices de participación obtenidos para $Pr_1$ . . . . .	79





# Índice de tablas

2.1. Ejemplo de support y confidence evaluado en tres reglas . . . . .	11
2.2. Ejemplo de clases de uso de suelo en una jerarquía de tres niveles. . . . .	16
3.1. Variables seleccionadas con Info Gain Attribute Eval, Ranker y Cross valida- tion 2-1, bajo el criterio de la condición de cambio o no-cambio. . . . .	23
3.2. Variables seleccionadas con Info Gain Attribute Eval, Ranker y Full training set, bajo el criterio de la condición de cambio o no-cambio. . . . .	24
3.3. Variables seleccionadas con ChiSquared attribute eval, Ranker y Cross vali- dation 2-1, bajo el criterio de la condición de cambio o no-cambio. . . . .	24
3.4. Variables seleccionadas con ChiSquared attribute eval, Ranker y Full training set, bajo el criterio de la condición de cambio o no-cambio. . . . .	25
3.5. Tipos de cambio de uso de suelo observado. . . . .	25
3.6. Variables seleccionadas con CfsSubsetEval, Best First y Cross validation 2-1, bajo el criterio del tipo de cambio. . . . .	26
3.7. Variables seleccionadas con InfoGain Attribute Eval, Ranker y Cross Valida- tion 2-1, bajo el criterio del tipo de cambio. . . . .	26
3.8. Variables seleccionadas con Info Gain Attribute Eval, Ranker y Full training set, bajo el criterio del tipo de cambio. . . . .	27
3.9. Variables seleccionadas con Chi-squared attribute eval, Ranker y Cross Vali- dation 2-1, bajo el criterio del tipo de cambio. . . . .	28
3.10. Variables seleccionadas con ChiSquared attribute eval, Ranker y Full training set, bajo el criterio del tipo de cambio. . . . .	29
3.11. Variables seleccionadas con el Modelo 1, del centro EULA . . . . .	29
3.12. Variables seleccionadas con el Modelo 2, del centro EULA . . . . .	30
3.13. Variables seleccionadas con el Modelo 3, del centro EULA . . . . .	30
3.14. Usos de suelo presentes en el área de estudio . . . . .	32
3.15. Agrupación de reglas obtenidas con el modelo 2. . . . .	33
3.16. Agrupación de reglas obtenidas con el modelo 3. . . . .	34
3.17. Agrupación de reglas obtenidas con el modelo 4. . . . .	34
3.18. Agrupación de reglas obtenidas con el modelo 5. . . . .	35
4.1. Índice de correlación lineal para $Pr_1$ y $Pr_2$ . . . . .	46
4.2. Índices de participación, por regla y ranking conforme a $P_i$ . . . . .	51
4.3. Seis reglas ordenadas conforme al $P_i$ y con el valor de la $C_p$ . . . . .	51

4.4. Listado de reglas presentadas al experto, valores de cada métricas y la valoración del experto (4= Muy interesante; 3= Interesante; 2=Poco interesante; 1= Interés nulo). . . . .	52
4.5. Correlación entre el experto y las métricas evaluadas. . . . .	53
6.1. Variables presentes en el estudio y su sigla. . . . .	66



# Capítulo 1

## Introducción

El uso de suelo tiene un ciclo definido por la naturaleza, pero la civilización y sus necesidades lo ha interferido y adaptado para su uso personal. Los usos de suelo han ido transformando gran parte de la superficie del globo en los últimos 300 años [19, 13]. Muchos de estos cambios han ido de la mano de la producción humana, es decir, la tala de bosques tropicales y prácticas agrícolas para la subsistencia humana. Todos estos factores han ido alterando de forma drástica el paisaje natural [9, 19]. Incluso se ha llegado a estimar, que sólo en el último siglo, una gran cantidad de ecosistemas a nivel global han sido afectados por el cambio de uso de suelo, siendo el de mayor impacto el que se ha producido en ecosistemas boscosos y praderas naturales, al ser estos convertidos a usos de suelo agrícola, ganadero, forestal, urbano e industrial [21].

En la región nor-patagónica, y para uso de los colonos, entre 1900 y 1950, y autorizados por el gobierno, se procedió a la quema de miles de hectáreas en la zona de Aysén y Coyhaique. El objetivo era convertir dichas zonas en un suelo agropecuario. Debido tanto a condiciones geográficas como climáticas, muchas de estas quemas se descontrolaron, y se estima que la superficie total incendiada alcanzó los 3 millones de hectáreas. Luego, y en el caso puntual del uso como suelo ganadero, el ramoneo dificultó aún más la recuperación del bosque.

Dada la importancia de los cambios en el uso de suelo para la sustentabilidad de los ecosistemas, el estudio de sus *forzantes* y su *dinámica* es prioritario en las ciencias ambientales. Este tipo de estudio, puede basarse tanto en estudios comparativos como evolutivos del paisaje natural. Esta última modalidad considera la variación espacio-temporal del uso de suelo, y permite analizar los cambios físicos que ocurren; identificar los factores *ambientales, sociales y económicos* que determinan estos cambios. Luego, el desarrollo de metodologías de modelación de patrones espaciales y/o temporales, se convierten en herramientas útiles y poderosas en el ámbito de la planificación territorial y evaluación ambiental.

Esta tesis aplica técnicas de minería de datos para modelar la dinámica del cambio de uso de suelo y por medio de esta técnica obtiene reglas que asocian las variables presentes en la región nor-patagónica. La minería de datos es un método utilizado generalmente en grandes bases de datos para extraer información novedosa, no trivial y accionable, la que por métodos de análisis tradicionales no se pueden lograr [11]. Inicialmente, la minería de datos nació como solución para el manejo y aplicación de consultas en grandes bases de datos comerciales. Las bases de datos espaciales nacen como solución para la manipulación de grandes volúmenes de información geográfica, cada vez más compleja y de mayor tamaño.

Así, la minería de datos espaciales surge como respuesta ante la necesidad de poder aplicar métodos de extracción de patrones con un componente espacial. El minado de datos a aplicar en este estudio, apunta a obtener respuestas a la dinámica del cambio de uso de suelo como una alternativa para los estudios ambientales. En particular, se implementa dos tipos de minado de datos que extraen reglas de asociación: *reglas de asociación espacial* y *reglas de colocalidad*.

El trabajo toma como datos de entrada una serie de datos pertenecientes al área de estudio en un tiempo determinado. Estos datos representan los cambios en el uso de suelo desde 1984 al 2001. La zona de estudio se encuentra geográficamente entre los 44°30'- 46°00' de latitud Sur y los 71°30'- 73°00' de longitud Oeste (ver figura 1.1, fuente: Centro EULA).

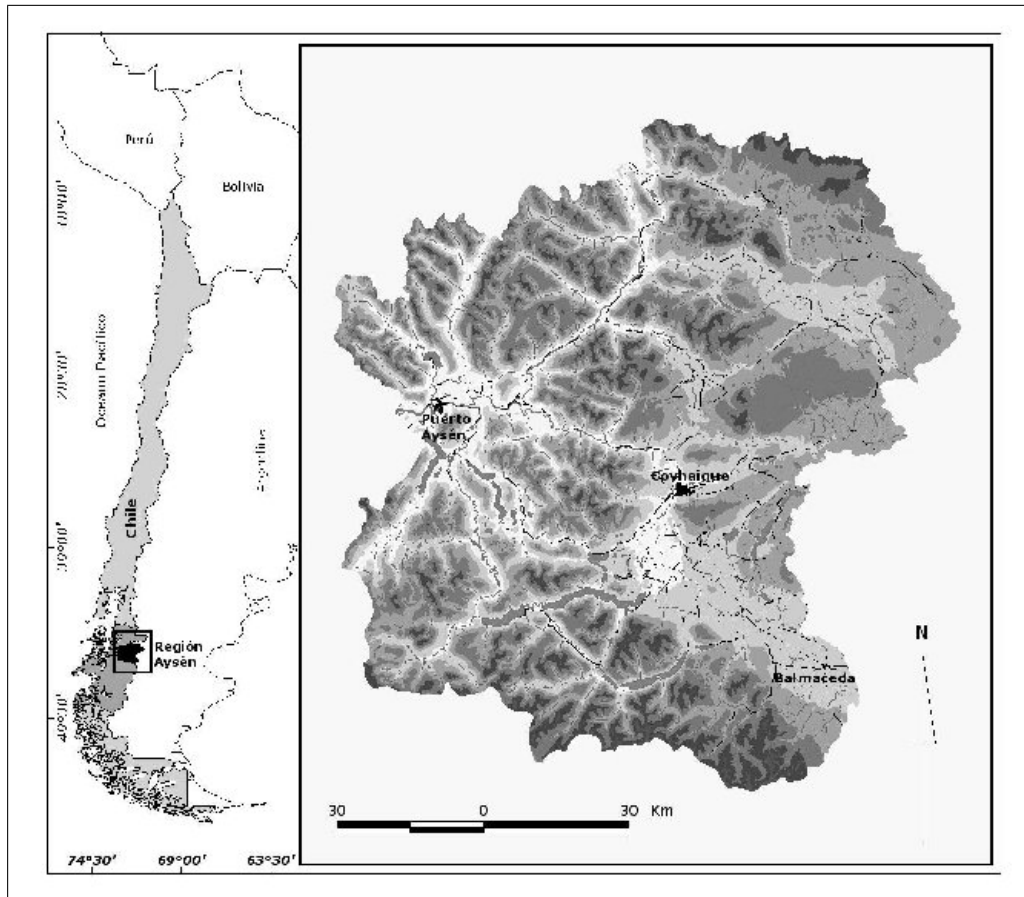


Figura 1.1: El área de estudio, cuenca del río Aysén.

En el proceso de minado de datos se tienen cerca de 50 variables, dentro de las cuales están las vías de accesos, zonas urbanas, tendidos eléctricos, y curvas de nivel, entre otras. Estas variables se evalúan en dos contextos: la dinámica del cambio entre dos tiempos y, la asociación de usos de suelo en un mismo tiempo. Esto, con el objetivo de poder determinar su grado de asociación con el cambio de uso de suelo. Las fuentes de información para este estudio son:

- Imágenes satelitales: Imágenes Landsat<sup>1</sup>(figura. 1.2).

<sup>1</sup>Landsat es un sistema implementado por los E.E.U.U. y consiste en una serie de satélites capaces de

- Fotografías aéreas: Fotografías tomadas del área de estudio, sobrevolando esta misma.
- Cartas IGM.
- Ortofotos.
- Observación directa: Esto permite hacer correcciones de algunos datos interpretados desde los otros medios.

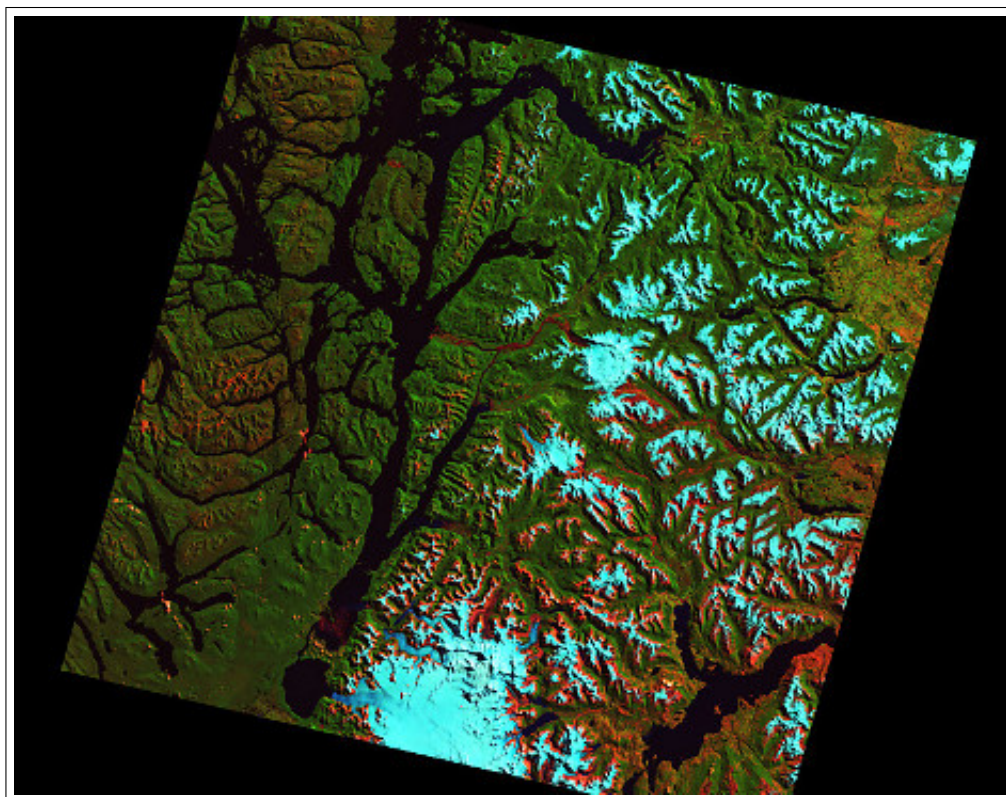


Figura 1.2: Imagen Landsat del área de estudio.

Las imágenes satelitales y fotografías han sido clasificadas<sup>2</sup> de manera de poder identificar elementos geográficos relevantes. Todos los datos están disponibles en formato digital usando una representación raster<sup>3</sup> o vectorial<sup>4</sup>.

El minado de datos, brinda herramientas para poder encontrar patrones en el conjunto de datos extraídos de los medios disponibles. Esta tesis aplica distintos medios de obtener asociaciones entre las variables del estudio.

Buscando la asociación del conjunto total de variables presentes, se realizó un experimento para obtener reglas de asociación. Esto involucra un proceso de preselección de las posibles

---

proporcionar una observación de la Tierra en alta resolución

<sup>2</sup>fotointerpretadas

<sup>3</sup>Archivos GIS con un formato de información geográfica representada a modo de grilla

<sup>4</sup>Archivos GIS con un formato de información geográfica representada a modo de vectores, es decir polígonos, líneas y puntos

variables a utilizar y una selección de variables en base al cual se extraen los patrones como reglas de asociación espacial.

Una segunda metodología aplicada, es la de asociar usos de suelo desde una imagen satelital, esto es, una instantánea de un tiempo específico. Estas reglas de colocalidad, permiten incorporar aspectos de vecindad espacial en las relaciones de adyacencia. Esto quiere decir que permite descubrir relaciones espaciales entre vecinos, ya sea respecto de una distancia entre puntos o regiones, o en el caso de esta tesis, respecto a la adyacencia de regiones.

## 1.1. Hipótesis

La cuenca del Río Aysén presenta características biogeográficas y patrones de ocupación del territorio altamente representativos de los cambios en toda la región nor-patagónica de Chile. El complejo conjunto de procesos que han ido determinando los cambios en el uso de suelo es una característica esencial de las transformaciones sufridas y plantea una serie de interrogantes respecto a los factores que están guiando dichas transformaciones.

Dentro del entorno geográfico, un sin-número de variables se conjugan día a día, interviniendo en mayor o menor grado la dinámica del uso de suelo. Estas variables ambientales, tienen entre ellas mismas mayor o menor asociación. Comprender estas asociaciones, son una herramienta que sirve de base para un trabajo de predicción sobre el estudio de la dinámica del uso de suelo.

Sin embargo, además de este conjunto de las variables ambientales, también es válido el criterio de que los distintos usos de suelo, por sí mismos interactúan entre ellos. Esto es, que los distintos usos de suelo, asociados a sus áreas, perímetros y porcentajes de adyacencia, también se pueden asociar, y por lo mismo, contribuir, al igual que un estudio de variables ambientales, a entender mejor la dinámica del uso de suelo.

Conforme a este contexto, y uniendo tanto las ciencias ambientales como las ciencias de la computación, es que se han planteado las siguientes hipótesis:

- **Existen asociaciones identificables por medio de minado de datos espacial, que permiten entender mejor la dinámica del uso de suelo. Estas asociaciones pueden ser representadas como reglas de asociación, asociando variables medioambientales, sociales y económicas presentes en el área de estudio.**
- **Existen asociaciones espaciales entre los distintos tipos de usos de suelo en el área de estudio, evaluadas conforme a sus relaciones topológicas, y que pueden ser representadas y evaluadas como reglas de colocalidad.**

## 1.2. Objetivos

### 1.2.1. Objetivo general

El objetivo general de esta tesis es asociar factores que de una u otra forma, y aplicando distintos criterios, describen y guían la dinámica del uso de suelo. En este estudio, se plantea identificar reglas de asociación espacial y reglas de colocalidad que permitan estudiar esta

dinámica y sus forzantes. Además, se plantea como objetivo el evaluar los patrones espaciales obtenidos en ambos experimentos.

### 1.2.2. Objetivos específicos

Los objetivos específicos son:

- Implementar una base de datos espaciales con la información obtenida.
- Ejecutar procedimientos de minería de datos para seleccionar un conjunto reducido pero significativo de variables del total de ellas, para comparar con la selección de variables por medio de herramientas estadísticas.
- Definir procedimientos de minería de datos espacial para determinar reglas de asociación entre las variables medioambientales, sociales y económicas presentes en el área de estudio.
- Definir técnicas que conjuguen procedimientos en un servidor de bases de datos espaciales, que permitan establecer reglas de colocalidad entre los distintos usos de suelo presentes en el área de estudio.
- Evaluar los patrones obtenidos por expertos.

## 1.3. Organización de la tesis

La organización de la tesis es la siguiente:

- En el capítulo 2 se presenta la discusión bibliográfica. Documentos sobre trabajos relacionados, así como técnicas de *spatial data mining*.
- En el capítulo 3 se presenta el trabajo de generación de las reglas de asociación, desde la obtención de datos, hasta la validación por el experto de las reglas de asociación obtenidas.
- En el capítulo 4 se describen la generación y los resultados de las reglas de colocalidad obtenidas.
- En el capítulo 5 se exponen las conclusiones del trabajo de tesis, así como el trabajo futuro.
- Los anexos adjuntos incluyen, entre otros, la descripción de todas las variables utilizadas en el estudio y las consultas/algoritmos en SQL para la determinación de las reglas de colocalidad.





# Capítulo 2

## Discusión bibliográfica

Comprender la dinámica del uso de suelo y sus forzantes cada día toma mayor importancia en los distintos estudios medioambientales. En efecto, los cambios de uso y/o cobertura de la superficie del planeta cobran cada vez mayor importancia. Las investigaciones orientadas al análisis de las transformaciones de la ocupación del territorio, se basan generalmente tanto en estudios comparativos como en estudios evolutivos de paisajes naturales. En efecto, los estudios evolutivos, es decir que comparan la variación temporal del uso de suelo, permiten analizar los cambios físicos ocurridos en un territorio. Luego, identificar los factores ambientales, sociales y económicos, así como determinar la relación entre ellos y el cambio de uso de suelo, cobran una necesidad de apoyo de métodos y sistemas especializados. Análisis espaciales se transforman en herramientas de gran utilidad y apoyo en la toma de decisiones en el ámbito de planificación territorial.

La patagonia de Chile presenta un conjunto de características naturales y transformaciones históricas en su paisaje que la convierten en un escenario ideal para la investigación y el desarrollo de metodologías relativas al análisis espacial.

La minería de datos nace a partir de la necesidad del manejo de grandes bases de datos como herramientas para el descubrimiento de nuevo conocimiento. La minería de datos espacial es la aplicación de similares técnicas pero considerando información con una referencia espacial (sistema de coordenadas) y propiedades geométricas. Cabe mencionar que en el espacio geográfico, existe una dependencia generalizada de todos sus componentes, incluso podemos agregar que todos los objetos tienen alguna relación espacial entre sí: de distancia o de orientación, por ejemplo.

### 2.1. Bases de datos espaciales

Un sistema de gestión de bases de datos espacial (BDE) es un sistema que permite almacenar y gestionar datos geográficos, con la principal característica de relacionar los objetos en el espacio [14]. Los objetos en el espacio son representados en administradores actuales de bases de datos usando tres tipos de datos abstractos: puntos, líneas (o polilíneas, que sería una secuencia de varias líneas) y polígonos (conjuntos de líneas cerrados)(ver figura 2.1). En este contexto, estas abstracciones de elementos geográficos podrían hacer referencia a distintos elementos; por ejemplo, un punto podría utilizarse para ubicar en el espacio un punto

de muestreo; una línea podría ser representativa para un camino o un río; y un polígono denotaría una zona con un tipo de uso de suelo específico. Incluso un mismo objeto espacial podría ser representado en diferentes formas, dependiendo de la escala de representación que se utilice.



Figura 2.1: Las tres formas de elementos espaciales: punto, línea y polígono.

La incorporación de modelos geométricos en una base de datos permite representar objetos espaciales a través de atributos temáticos y geométricos. Una base de datos espaciales ofrece *tipos de datos espaciales* y un *lenguaje de consulta y manipulación* para estos tipos de datos. Además, un sistema de bases de datos espaciales provee índices espaciales y métodos de procesamiento de consultas [14], con lo que se logra responder consultas tales como consultas por rango, consultas de combinación espacial y consultas del vecino más cercano, entre otras. En forma resumida, es un sistema que permite almacenar y manipular datos con georeferenciación, al mismo tiempo que determinar relaciones espaciales entre los datos almacenados.

La figura 2.2 muestra un ejemplo de la forma en que los datos geográficos pueden ser almacenados en una base de datos relacional, donde los datos de calles, estaciones de servicio y ríos tienen, no necesariamente, el mismo tipo de representación geométrica (puntos o multi líneas). Además, en estas tablas no existe un atributo explícito (como sería el uso de claves foráneas en un modelo relacional) que relacione, por ejemplo, calles con estaciones de servicio. Sin embargo, una base de datos espacial posee un lenguaje de consulta que permite extraer relaciones espaciales a partir de la representación geométrica de los datos. Por ejemplo, se puede determinar cuáles son las calles que intersectan, o los ríos que están ubicados a menos de 100 m. de una estación de servicio [3].

Todos los datos representados en un sistema de información geográfica y, almacenados en una base de datos espaciales, poseen relaciones entre sí. Los atributos espaciales de los objetos de la figura 2.2 poseen relaciones que no se almacenan necesariamente en forma explícita en la base de datos, pero que se pueden computar con operadores geográficos. Estas relaciones se pueden clasificar básicamente en tres tipos: de distancia, dirección y topología [14](ver figura 2.3):

- **Distancia:** Es una relación métrica que establece una separación entre los objetos.
- **Dirección:** Determina relaciones de posicionamiento en términos de un marco de referencia (ej. norte, sur, este, oeste). Estas son relaciones que son invariantes ante operaciones continuas de escalamiento y traslación.
- **Relaciones topológicas:** Son relaciones que establecen un grado de conectividad entre los objetos (ej. adyacencia, contenimiento, separación, etc.). Estas relaciones son invariantes ante continuas transformaciones de escalamiento, traslación y rotación.

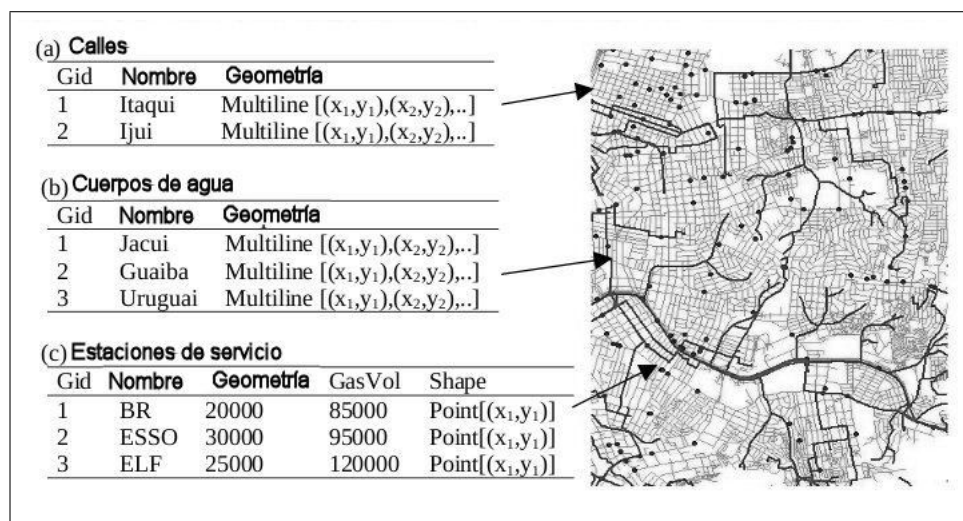


Figura 2.2: Ejemplo de datos geográficos almacenados en una BD espacial (fuente: Bogorny 2006) [2].

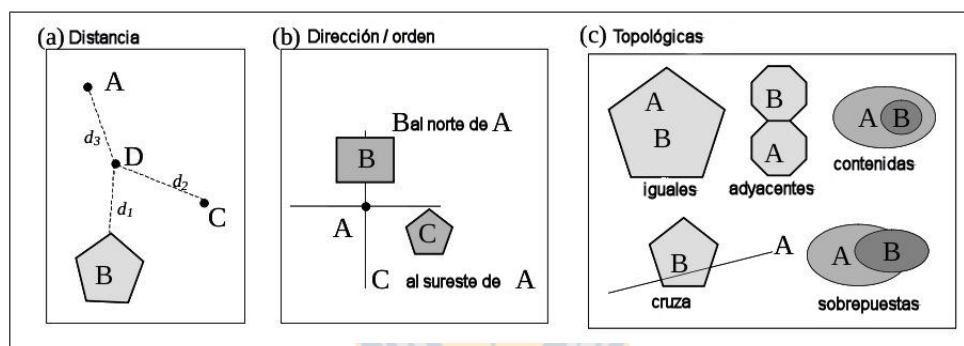


Figura 2.3: Ejemplos de relaciones espaciales.

Por otro lado, en el ámbito geográfico, algunos elementos *siempre* están relacionados con otros, en este caso se habla de *dependencia geográfica*. La dependencia geográfica se define como la *relación espacial entre dos características geográficas, tipo A y tipo B, donde una instancia de A siempre tendrá una relación espacial con al menos una instancia de B* [2]. Estas dependencias geográficas están generalmente definidas en las BDE por los mismos diseñadores mediante reglas de *restricción de integridad semántica* [7], las cuales garantizan la integridad de los datos al imponer condiciones semánticas a las relaciones espaciales.

## 2.2. Minería de Datos

El minado de datos se define como: “**El proceso de extraer conocimiento no-trivial, novedoso, implícito y accionable de grandes bases de datos** [11]”. La propiedad del conocimiento accionable, significa que el conocimiento que puede extraerse desde una BD puede promover una acción [12].

La minería de datos hace posible la exploración, análisis y visualización de datos en

un alto nivel de abstracción sin una hipótesis específica en mente; generalmente aplica sus técnicas considerando igual distribución e *independencia* en los datos [11]. Sin embargo, la independencia de los datos en minería de datos espacial violaría la primera ley de la geografía que dice: *todas las variables, por más alejadas que parezcan geográficamente, tienen alguna relación entre sí* [23]. Esto significa que en una zona de estudio, cualquier cambio observado en una coordenada  $x_1, y_1$  podría producir algún efecto en otra coordenada  $x_2, y_2$ .

El minado de datos es tan solo una parte de lo que se denomina KDD<sup>1</sup>, el cual es un proceso para extraer información interesante y novedosa desde un extenso conjunto de datos [17]; es un proceso interactivo en el que se pueden identificar cinco principales etapas: *selección, preprocesamiento, transformación, minado de datos y evaluación/interpretación* [10]. La selección, preprocesamiento y la transformación de los datos son procesos que preparan los datos para el minado de datos en sí, y conlleva la mayor parte del proceso (entre un 60 % – 80 %).

En minería de datos existen varias técnicas de minado de datos, entre ellas se destacan tres tipos, los que son: clasificación, clustering (o agrupamiento) y reglas de asociación.

La *clasificación* permite *construir estructuras a partir de ejemplos de decisiones pasadas y que podrían usarse en la toma de decisiones sobre casos no vistos, generando modelos predictivos a partir de datos existentes*. La clasificación tiene por objetivo predecir la asignación de una instancia a una clase dada a partir de una muestra con ejemplos positivos y negativos de la misma [8] y se considera un proceso de *aprendizaje supervisado*.

En el ámbito espacial, las técnicas más estudiadas y aplicadas en cuanto a *clasificación* son las que utilizan modelos de regresión, como la regresión lineal y la regresión espacial. Este último tipo de regresión es posible ya que las variables espaciales están georeferenciadas [22].

El *clustering* (Aprendizaje no supervisado) permite *agrupar un conjunto de datos en grupos (clusters) en base a ciertas similitudes*. A diferencia de la clasificación, no existe un conocimiento previo, ya sea del número de clusters, o el significado de los clusters. Por ser un tipo de aprendizaje *no supervisado*, no se requieren ejemplos, y no existen *clases* predefinidas [22]. Dependiendo de la finalidad y los ajustes, el proceso de clustering puede tener distintos resultados (i.e. si tenemos un grupo de 10 polígonos con variados usos de suelo, podría generar cluster dependiendo de su uso de suelo, o por su superficie, o por altura o pendiente, entre otros), como se aprecia en la figura 2.4. La calidad del proceso depende tanto de la medida de similaridad usada, como de la propia implementación.

En el contexto del minado de datos espaciales, el clustering es muy útil para determinados problemas espaciales, ya que trabaja con puntos georeferenciados, tales como los píxeles de una imagen satelital, o puntos de muestreo georeferenciados. Por ejemplo, las regiones de un país podrían agruparse en clusters conforme a su densidad poblacional, ingresos económicos, expectativa de vida y tasa de desempleo, por nombrar algunos ejemplos.

Las distintas técnicas aplicadas en esta tesis, permiten *encontrar patrones*, o sea, regularidades en los datos, en la forma de reglas, como *combinaciones de conjuntos de datos frecuentes*. Formalmente, dado un conjunto de datos de transacciones<sup>2</sup>, se debe encontrar las correlaciones entre los ítems en la forma de reglas *IF-THEN*. En el contexto espacial, al

<sup>1</sup>KDD: Knowledge Discovery in Databases

<sup>2</sup>Una transacción es un conjunto de ítems que coexisten o se cumplen en una base de datos.

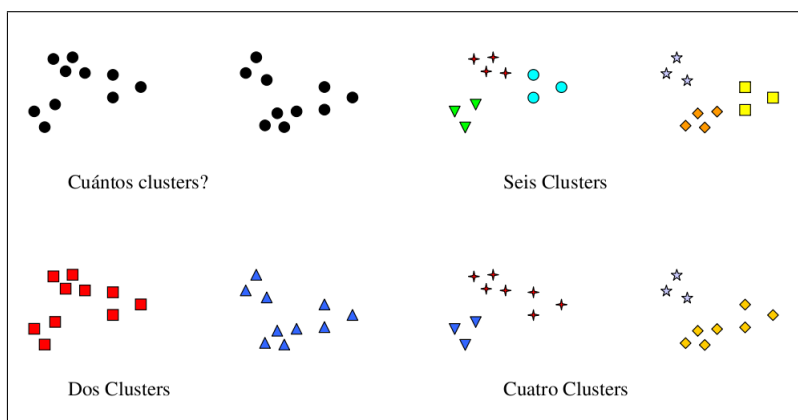


Figura 2.4: Noción de cluster depende del enfoque del estudio.

menos algún componente de la regla debe ser espacial, como por ejemplo, la georeferenciación de una zona de estudio [22]. Las reglas de asociación permiten describir el comportamiento de los datos en cuanto a su asociación.

Debe tenerse especial cuidado en que las reglas de asociación seleccionadas del total de las candidatas, sean efectivamente interesantes y útiles. Por ejemplo, se podría obtener una regla del tipo: *If* la región es de la patagonia chilena *and* la estación es invierno *then* llueve mucho. Esta no es una regla novedosa ni mucho menos interesante, ya que es un hecho natural y cotidiano que no aporta conocimiento relevante ni menos novedoso. Tanto en el minado de datos tradicional, como en el espacial, las reglas de asociación tienen tres componentes: rule, support y confidence. La evaluación de una regla se hace calculando y comparando las últimas medidas: su *confidence* (precisión) y su *support* (frecuencia).

La tabla 2.1 grafica ambas medidas para tres reglas obtenidas de un conjunto de datos  $I$ , denotado como  $I=\{A, B, C, D, E, F\}$  y evaluando un conjunto de transacciones  $T$ , que incluye  $T=\{ABC, ABD, BDE, CEF\}$ .

Tabla 2.1: Ejemplo de support y confidence evaluado en tres reglas

Regla	Support	Confidence
$A \Rightarrow B$	0.50	1.00
$B \Rightarrow C$	0.25	0.33
$F \Rightarrow E$	0.25	1.00

El support de una regla es la probabilidad de que exista un ítem A ( $P(A)$ ), o que ocurra B ( $P(B)$ ). El confidence, es la probabilidad de que, por ejemplo, ocurra B dado que ocurrió A ( $P(B|A)$ ), o que ocurra C dado que ocurrió B ( $P(C|B)$ ).

Uno de los métodos más usados para generar reglas de asociación es el algoritmo *apriori*, el cual explota la búsqueda por medio de evaluaciones del *support*. Si una regla obtiene un alto support, el algoritmo explora todos los sub-conjuntos de datos posibles. El algoritmo funciona buscando primero conjuntos de datos individuales, luego conjuntos de datos de tamaño 2, y así sucesivamente [22].

## 2.3. Reglas de asociación espacial

Una regla de asociación tiene la forma de  $A \rightarrow B$ , donde  $A$  (el antecedente) y  $B$  (el consecuente) son conjuntos de predicados [18, 22].

Las reglas de asociación espacial tienen una característica espacial: *la transacción se instancia con respecto a la característica espacial y los predicados son espaciales*. Formalmente, una regla de asociación de tipo espacial tiene la siguiente forma:

$$P_1 \wedge P_2 \wedge \dots P_n \rightarrow Q_1 \wedge Q_2 \wedge \dots Q_m$$

donde al menos un  $P_i$  o  $Q_j$  debe ser espacial [22].

Ahora bien, en un *contexto espacial* y debido a la correlación espacial que poseen las variables, es que se debe aplicar algún método que considere el factor espacial en la concurrencia de ciertos ítems o hechos.

La *correlación espacial* refleja el efecto que tiene la vecindad (espacial) para una transacción. Las reglas de correlación espacial intentan generalizar las reglas de asociación tradicionales a conjunto de datos que son indexados espacialmente. Las principales diferencias entre reglas de asociación espacial y las no espaciales incluyen:

- Los eventos espaciales son influenciados por eventos adyacentes, o colocalizados<sup>3</sup>.
- Los conjuntos de datos espaciales contienen menos *ítems* que los conjuntos de datos no-espaciales.
- En muchos casos, los elementos de un conjunto de datos espaciales son valores discretos de variables continuas (por ejemplo muestras de  $T^\circ$ ).

Desde este punto de vista, el concepto de transacción debe reemplazarse por el de *adyacencia*, y el *support* por un índice de participación. Una forma fácil de entender este concepto es buscar reglas para distintos habitats de fauna, o en otras palabras, relaciones de colocalidad de un evento y su entorno, como por ejemplo: en el invierno los huemules (*Hippocamelus bisulcus*) se agrupan y habitan las laderas norte de los cerros.

A diferencia del minado de datos tradicional, el minado de datos espacial para generar reglas de asociación se descompone en al menos tres fases:

1. Extraer los predicados o relaciones espaciales: Un predicado espacial es una relación espacial.
2. Encontrar todos los conjuntos de patrones de predicados: Un conjunto de predicados es un patrón frecuente si su *support* es al menos igual a un umbral determinado, llamado *minsup*.
3. Generar reglas de asociación fuertes: Se dice que una regla de asociación es fuerte si alcanza el mínimo *support* y su *confidence* es al menos igual a un umbral determinado, llamado *minconf*.

---

<sup>3</sup>Colocalidad, es un subconjunto del set de datos con valores booleanos de factores espaciales



Existen pocos algoritmos para generar reglas de asociación espacial en comparación a las reglas de asociación tradicionales. Generalmente se basan en un análisis cuantitativo, calculando relaciones de distancias durante el proceso de cómputo. Este tipo de algoritmos trabaja con dos tipos de datos:

1. Geográficos directamente (georeferenciados por coordenadas).
2. Relaciones de distancias y topología: se basan en un análisis cualitativo, considera usualmente relaciones de distancias y topología entre distintos objetos geográficos y un conjunto de variables independientes representados por puntos, líneas y polígonos.

## 2.4. Reglas de colocalidad

El segundo planteamiento de esta tesis, es más puntual que el anterior, ya que trabaja sólo con los usos de suelo, directamente sobre un servidor de bases de datos espaciales. Los archivos son originados desde imágenes satelitales, que han sido clasificadas (fotointerpretadas) por el centro EULA. La motivación de este tipo de reglas de asociación, es identificar los patrones considerando para tal efecto, sólo los usos de suelo, evaluando su superficie, perímetro y porcentaje de adyacencia.

El término *colocalidad* se usa para referirse a un subconjunto de características espaciales booleana (i.e. 1=“adyacente” y 0= “no adyacente”). Una regla de *colocalidad* es de la forma:  $c_1 \Rightarrow c_2(p, cp)$ , donde  $c_1$  y  $c_2$  son *características espaciales*,  $c_1 \cap c_2 = 0$ ,  $p$  es un número que representa la medida de prevalencia (medida que en síntesis define cuánto sobresale esta regla por sobre las otras), y  $cp$  es un número de medición de la probabilidad condicionada.

Una regla de colocalidad es un modelo que busca asociar alguna característica espacial (adyacencia, superficie, perímetro y/o porcentaje de adyacencia) en las vecindades de instancias de otras características espaciales [15]. Para efectos de vecindad, se incorpora el concepto del *proximidad vecina*, ya que con éste se van computando qué áreas son efectivamente vecinas. Este concepto puede definirse en base a las relaciones espaciales (juntos, disjuntos, se topan), relaciones métricas (distancia Euclidiana), o una combinación. El concepto de proximidad vecina puede ser distinto al concepto de vecindad usado en topología, ya que dos objetos que sean próximos en término de una distancia entre ellos pueden no necesariamente ser adyacentes desde un punto de vista topológico.

La figura 2.5 [15], muestra un conjunto que consiste en instancias de varias características booleanas espaciales, cada una representada por un símbolo distinto. En esta figura, se puede observar dos patrones de *colocalidad* (por ejemplo ('+', 'x') y ('0', '\*')).

Huang *et al.* [15] define en su trabajo un modelo concéntrico de eventos, el que se puede apreciar en la figura 2.6, donde  $T.i$  representa las instancias de  $i$  con características del tipo  $T$ , y las líneas entre las instancias representan las relaciones de vecindad.

El procedimiento de Huang y Shekhar [15] determina la cantidad de proximidades vecinas que posee un elemento de otro. Así, se van generando tablas con las posibles combinaciones de vecindad y calculando una métrica de validación de la regla obtenida. Estas reglas quedan almacenadas en la tabla de instancias. Este proceso depende del tamaño ( $k$ ) de la vecindad a considerar. Como ejemplo, para,  $k = 2$  el proceso evalúa primero todos los posibles vecinos para cada tipo de uso de suelo, es decir, comienza calculando las vecindades desde el *evento*

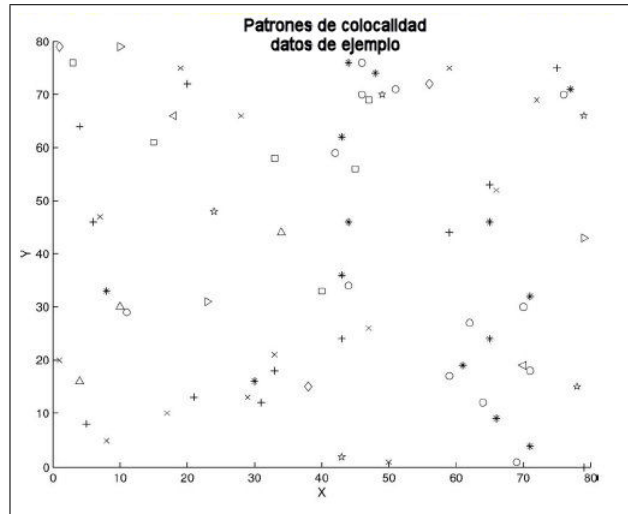


Figura 2.5: Patrones de colocalidad.

A hasta las vecindades del *evento*  $C$ . Así, comienza generando los pares para cada uso de suelo, hasta calcular todos y cada uno de ellos. Dentro del proceso, se van eliminando las duplicaciones de pares. Después de evaluar todas las posibles vecindades, se detiene y deja de generar los conjuntos de reglas de colocalidad de tamaño  $k$ , entregando como salida la unión de estos conjuntos.

La evaluación de las reglas se lleva a cabo utilizando 2 métricas:

1. **Radio de participación (Pr)**: Calcula la participación de la regla encontrada dentro del universo de la muestra y se define como:

$$Pr = \frac{\pi_{f_i}(|table\_instance(c)|)}{|table\_instance(f_i)|}, \quad (2.1)$$

Donde  $\pi$  es la proyección sin duplicación del tipo  $f_i$  sobre las instancias de colocalidad  $c = \{f_1, \dots, f_k\}$  donde participa  $f_i$ .

Para el radio de participación, Huang y Shekhar [15] se utiliza el concepto de *índice de participación*  $pi(c)$ , que es el  $\min_{i=1}^k \{Pr(c, f_i)\}$ , es decir el mínimo valor obtenido de los Pr, ya que el valor de esta métrica depende del sentido en que se evalúe la regla. Este índice es finalmente usado como métrica de validación de un regla de colocalidad porque está relacionado a la función *cross-K* [20], empleada comúnmente para medir la interacción entre pares de figuras espaciales, y porque posee la propiedad antimotónica que puede ser empleada para eficiencia computacional.

2. **Probabilidad condicional o Conditional probability (Cp)**: Es la fracción de instancias de  $c_1$  presente junto a  $c_2$  cuando la regla es evaluada para el  $c_1$ , siendo definida como:

$$Cp = \frac{|\pi_{c_1}(table\_instance(\{c_1 \cup c_2\}))|}{|table\_instance(c_1)|}, \quad (2.2)$$

En la figura 2.6 se aprecia cómo se distribuyen los distintos elementos, y cómo se denotan las colocalidades. En esta figura, A, B y C representan las distintas instancias de eventos



presentes, luego, 1, 2, 3, 4 y 5 indica la ocurrencia de estas instancias. Los nodos representan los eventos presentes en el área de estudio, y los arcos las instancias de colocalidad entre ellos. El proceso de Huan y Shekhar comienza desde patrones de valor 1 (aunque estas no requieren computarse) hasta la mayor que se presente en el experimento. Luego va generando patrones uniendo los distintos puntos de muestreo, pero que no sean del mismo tipo de punto (los tipos A con los tipos B, etc.).

El *radio de participación* es la fracción de instancias de un suceso  $f_i$  dado un suceso como condición, por ejemplo, la fracción de sucesos  $\{A,B\}$  evaluados para el suceso B. Un concepto importante en este modelo de Huang y Shekhar [15] es el de *proximidad de vecinos*, esto ya que a mayor proximidad, mayor fuerza debería de tener un elemento con respecto a otro. El cálculo se hace, para una cobertura o tipo de instancia, desde tamaño 2 (par de puntos), primero se elimina el valor duplicado ( $x,y = y,x$ ) por ser relaciones simétricas. Por ejemplo en la figura 2.6 para las instancias de reglas de colocalidad  $\{A, B\}$  se tienen  $\{A.1, B.1\}$ ,  $\{A.2, B.4\}$ ,  $\{A.3, B.4\}$ , pero solo dos  $\{B.1$  y  $B.4\}$  de las 5 instancias de B participan en la colocalidad  $\{A, B\}$ . Luego, el  $pr(\{A,B\}, B) = 2/5 = 0,4$ . Y el  $pr(\{A, B\}, A) = 0,75$ . De este mismo ejemplo, se tiene que  $Pr(\{A, B\}, B) = 2/5 = 0,4$  y el  $Pr(\{A, B\}, A) = 3/4 = 0,75$ . Luego el índice de participación  $pi(A, B) = \min(0,75, 0,4) = 0,4$ .

Esto se explica mejor con la figura 2.6 [15]. Por ejemplo, la  $C_p$  para la regla  $A \Rightarrow C$ , sería  $2/4$ , es decir 0.5, o en términos de probabilidad un 50%.

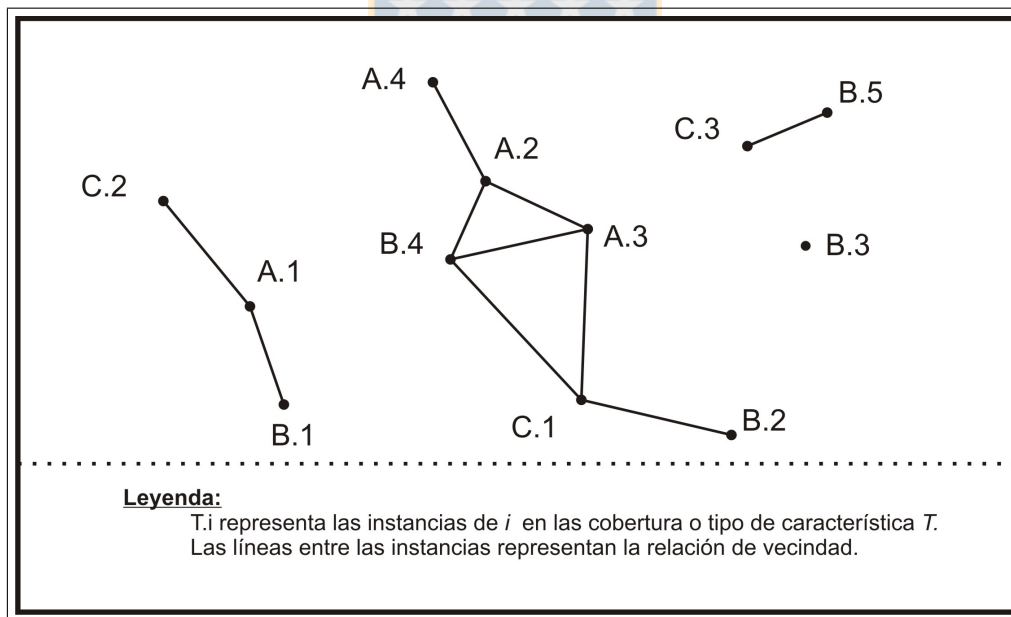


Figura 2.6: Modelo concéntrico de eventos.

En la tabla de la figura 2.7 se muestra un ejemplo válido de reglas de colocalidad tomadas de la figura 2.6. Para  $k=1$  las reglas toman el valor de cada nodo por si solo. Con  $k=2$ , se puede observar que existen dos casos válidos: colocalidad entre  $\{A, C\}$  y entre  $\{B, C\}$ , los cuales se denotan claramente en la figura 2.6 a partir de los arcos que unen las instancias de A con C y de B con C.

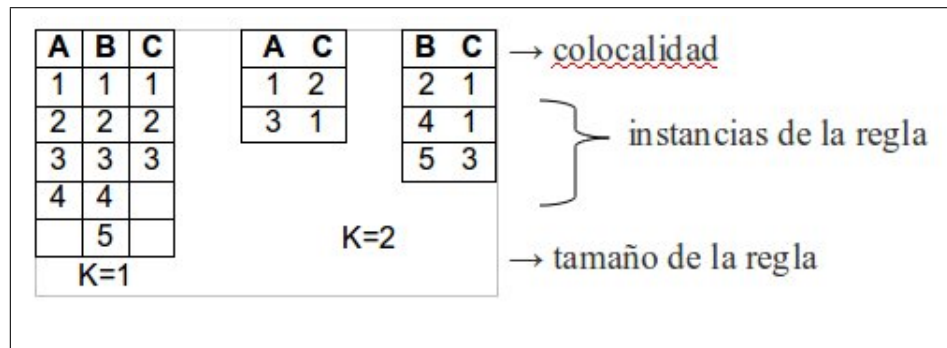


Figura 2.7: Ejemplos de reglas obtenidas desde la figura 2.6

## 2.5. Uso de reglas de asociación para manejo de uso de suelo

Existen varios enfoques para la aplicación de reglas de colocalidad, por ejemplo, Mannis y Liu [18] implementa un algoritmo *a priori* para buscar asociar datos entre el uso de suelo, población minoritaria<sup>4</sup>, pobreza, densidad urbana y la distancia a carreteras. El caso de estudio presentado, corresponde al caso específico de uso de suelo en Denver, Colorado, E.E.U.U., con una ventana temporal desde 1970 hasta 1990. En este trabajo se buscaron patrones desde una base de datos con información socio-económica y cambios en el uso de suelo. La meta era encontrar hipótesis que relacionaran ambos.

El método genera una jerarquía de tres niveles, la que puede apreciarse en la tabla 2.2.

Tabla 2.2: Ejemplo de clases de uso de suelo en una jerarquía de tres niveles.

Nivel 1	Nivel 2	Nivel 3
1 Water	...	...
2 Developed	21 Residential	211 Single-Family residential
		212 Multi-Family residential
	22 Non-Residential	...
3 Bare	...	...
4 Vegetated	41 Woody Vegetation	411 Forested
		412 Shrubland
	42 Herbaceous Vegetation	...

El método calcula relaciones espaciales y temporales por medio de sencillos métodos como el *overlap* de dos cubiertas (i.e. uso de suelo, y caminos). Luego aplicando un índice espacial de *join* se obtienen relaciones de cobertura de suelo y caminos o, por pre-procesamiento de las coincidencias de un punto específico desde el conjunto de datos de caminos y el conjunto de datos de usos de suelo. Esto se hace con un Sistema de Información Geográfica (GIS) para el pre-procesamiento de las relaciones y se crea una tabla sencilla, la que llaman *mining table*. Las relaciones calculadas para cada polígono incluyen:

<sup>4</sup>Población minoritaria: Los no-blancos e hispanos

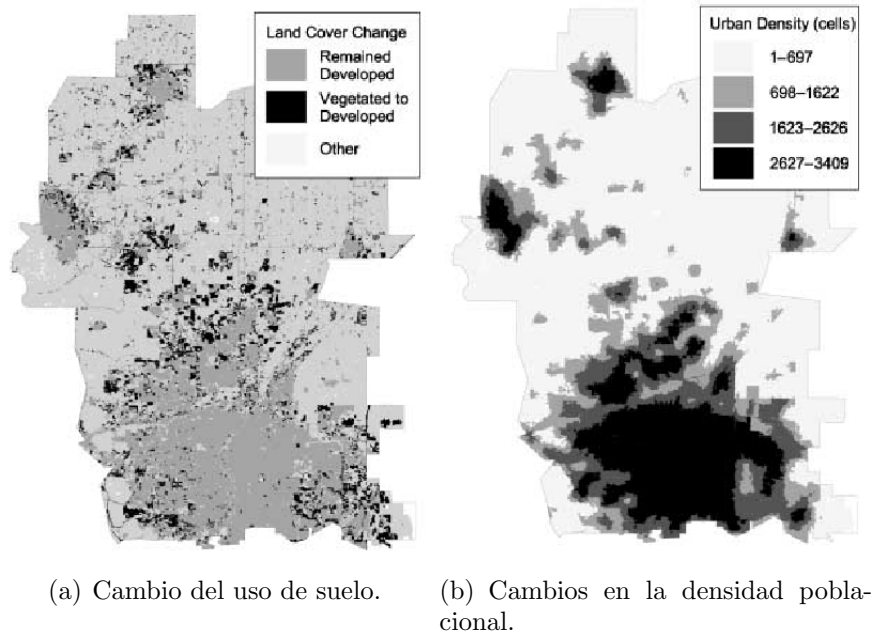
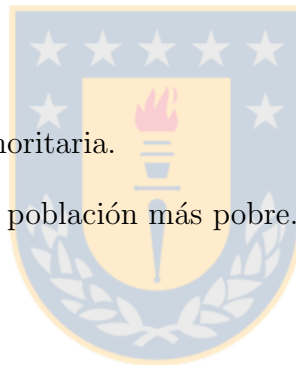


Figura 2.8: Ejemplo de dos distintas variables usadas en el minado de reglas de asociación.

- Cambios de uso de suelo.
- Cambios en la población minoritaria.
- Cambios en el porcentaje de población más pobre.
- Densidad urbana.
- Distancia a autopistas.



Mennis y Liu [18] utilizan el método Classification Based on Associations para generar reglas. Se aplicó primero por los distintos niveles indicados en la tabla 2.2, con lo que se encontraron interesantes reglas entre las variables socio-económicas, y luego agregando el cambio de uso de suelo, densidad urbana y distancia a carreteras. Dentro del proceso, al ir agregando distintas variables, si se encontraba alguna regla que pareciera interesante, se profundizaba usando un nivel jerárquico más fino. Se pueden ver en la figura 2.8 [18], ejemplos de las coberturas mencionadas (uso de suelo y densidad urbana), dos de las variables usadas en el minado de reglas de asociación.

Los resultados de Mennis y Liu [18] generaron una gran cantidad de reglas candidatas, pero finalmente sólo 4 reglas fueron validadas por ellos mismos y por expertos, y sólo dos de ellas relacionadas con el cambio de uso de suelo. Cabe destacar, que por ser un estudio que analizaba la relación del cambio de uso de suelo con los cambios en el nivel de pobreza, en el trabajo sólo se mencionan las reglas que generaron un aumento o una disminución en esta última variable. Además se obtuvieron conclusiones como que aún sigue siendo problemático integrar diversos conjuntos de datos, ya que muchas veces se requiere una interpolación de áreas.

Para Bogorny [2], un foco de atención radica en considerar las relaciones espaciales y las dependencias geográficas. En su trabajo, se extiende el software de minado de datos WEKA<sup>5</sup>. Se incorporó la capacidad de definir relaciones espaciales al cargar una base de datos espaciales. El proceso implementado provee dos niveles de granularidad para las relaciones de distancia y topológicas, y un alto nivel de granularidad para relaciones topológicas (tales como intersecciones), sin un concepto de jerarquía [4].

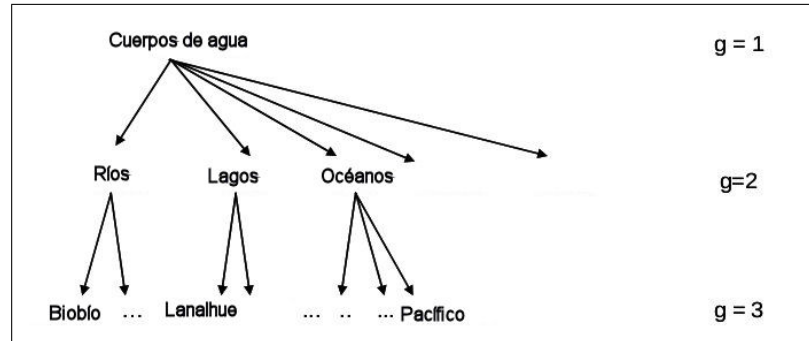


Figura 2.9: Ejemplo de jerarquía de granularidad.

Bogorny [2] desarrolla extensiones que agrega a los estándares del software WEKA, considerando las relaciones espaciales de los conjuntos de datos al momento del minado de datos y la generación de las reglas de asociación. Con este proceso se logra, además, indicar al algoritmo las *dependencias geográficas* y se reduce la cantidad de reglas de asociación generadas, o dicho de otra forma, se logran menos reglas de asociación no interesantes o poco útiles. El preprocesamiento de los datos se realiza con el software WEKA, y la generación de reglas de asociación usando el algoritmo *apriori*.

<sup>5</sup>WEKA es un software del tipo Open Source, pero que en sus primeras versiones no incorpora soporte para Bases de Datos Espaciales

# Capítulo 3

## Reglas de asociación

En esta tesis y para asociar las variables presentes en el área de estudio a modo de reglas de asociación, se consideraron dos conjuntos de datos, que si bien son de distinto formato, obedecen ambos al área de estudio de la cuenca del río Aysén y son extraídos de distintas fuentes de datos. En este capítulo, el conjunto de datos usado para las reglas de asociación, representan la geografía de Aysén y las variables ambientales, sociales y económicas, medidas en dos momentos: 1984 y 2001. Para las reglas de colocalidad, se procede a integrar métodos espaciales en el servidor de bases de datos espaciales, pero considerando como dato, el uso de suelo de cada polígono.

### 3.1. Datos y Selección de Variables

#### 3.1.1. Obtención de datos

La obtención de datos se hace a través de las siguientes fuentes de información:

- Imágenes satelitales Landsat<sup>1</sup>.
- Fotografías aéreas del área de estudio, sobrevolando esta misma.
- Cartas IGM.
- Ortofotos.
- Observación directa.

La clasificación de las imágenes se desarrolla en el Centro EULA-Chile, donde se clasifican las imágenes y se convierten en capas temáticas compatibles con ArcView e Idrisi. Se digitalizan las cartas IGM, así como se digitalizan, georeferencian y se clasifican las imágenes satelitales, ortofotos y fotografías aéreas. Finalmente, se realizan visitas en terreno para corroborar datos entre lo clasificado y lo observado.

---

<sup>1</sup>Landsat es un sistema implementado por los E.E.U.U. y consiste en una serie de satélites capaces de proporcionar una observación de la Tierra en alta resolución

### 3.1.2. Recopilación y datos de entrada obtenidos

Para poder modelar el sistema de minado de datos, se necesita analizar el conjunto de las variables independientes presentes en el área de estudio (accesibilidad a la red de transporte, proximidad de redes de servicios básicos, características demográficas de población, valor del suelo, regulaciones territoriales, límites administrativos, propiedades y datos biofísicos como el tipo de suelo, vegetación, elevación e hidrología, entre otras).

Se han recolectado datos desde el Centro EULA, quienes han realizado un análisis y una cuantificación de los cambios de uso de suelo desde la década de 1940 hasta el año 2007. Este análisis se lleva a cabo a través de un proceso de regresión histórica tomando como base el año 2007. Para el período 1975-2007 se clasificaron cuatro imágenes Landsat que corresponden a los años 1975 (Landsat MSS), 1986 (Landsat TM), 1995 (Landsat TM), (2001 (Landsat ETM+) y 2007 (Imagen Aster).

La información temática (centros poblados, vías de transporte, cuerpos de agua, etc.) fueron extraídos de cartografía I.G.M. a escala 1:25.000 y 1:50.000. Además, se utilizaron ortofotos digitales de propiedades rurales y series de suelo de CIREN-Corfo<sup>2</sup>, a escala 1:20.000. Otras fuentes de información de utilidad, son los catastros de bosque nativo y tramas urbanas digitales. La totalidad de las variables consideradas a priori para el estudio se detallan en anexo 6.2.

La información recopilada, es evaluada y validada mediante el trabajo en terreno. El proyecto contempla la creación de una base de datos espaciales que integre toda la información recolectada para la cuenca del río Aysén. El objetivo de esta base de datos es poder manejar gran cantidad de información espacial (geométrica), relacionada con información temática, de una forma eficiente y explotando las características de los SGBD en cuanto a procesamiento de consultas. Actualmente se tiene implementada una base de datos espaciales utilizando el motor de base de datos PostgreSQL, con la extensión PostGIS.

En la etapa de preprocesamiento de datos, se generan dos *mining tables* que alimentarán independientemente el proceso de minado de datos en su fase de selección de variables. Estas tablas se generan a partir de distintos *joins* ejecutados con ArcView sobre los archivos de cada capa temática<sup>3</sup>. La diferencia de estas dos tablas, radica en el criterio que se aplicará para estudiar la dinámica del cambio del uso de suelo. En esta tesis, se plantea primeramente el criterio sobre si un punto de muestreo cambió o no su uso de suelo del estado inicial al final y, como segundo criterio, se definió el tipo de cambio, es decir, identificar el tipo de cambio. La estructura de la tabla misma, se apega más a una planilla de cálculo (ver figura 3.1), ya que considera todas las variables del estudio, y contiene todos los valores de las tuplas.

Por otra parte, se genera una segunda *mining table*, pero que considera como variable dependiente para la preselección de variables el *tipo de cambio*, es decir, se intenta representar la dinámica del cambio de uso de suelo, o sea, se registra el cambio observado de un tipo de uso de suelo a otro.

---

<sup>2</sup>CIREN, Centro de Investigación de Recursos Naturales

<sup>3</sup>ArcView trabaja con archivos .shp, los que representan *capas temáticas* generadas en la digitalización de la información recopilada



AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC
_AG05	_PR50	_PR25	_PR10	_PR05	_PL50	_PL25	_PL10	_PL05	_USO_1984	_USO_2001	CAMBIO	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	2	2	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0371	0,0097	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0391	0,0117	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0373	0,0025	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	9	9	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0464	0,0163	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0464	0,0388	0,0063	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	8	8	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	9	9	0	
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	9	9	0	
0,0000	0,0167	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	13	1	
0,0000	0,0047	0,0162	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	
0,0000	0,0131	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	13	1	
0,0000	0,0181	0,0051	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	2	13	1	
0,0000	0,0200	0,0428	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1	13	1	
0,0000	0,0228	0,0801	0,0126	0,0000	0,0000	0,0000	0,0000	0,0000	1	1	0	

Figura 3.1: Mining table definiendo como variable dependiente la condición de cambio o no cambio del uso de suelo.

### 3.1.3. Preparación de los datos

La etapa de preparación de datos involucra las tareas de adquisición y la limpieza. Para la preparación de datos espaciales, se apunta a descubrir las relaciones espaciales desde un conjunto de variables que tengan la mayor relevancia con respecto a su relación espacial [4].

Todos los datos son inicialmente procesados en formato Raster por el Centro EULA-Chile que incluye la clasificación de cada capa (una capa por cada variable independiente), lo que permite trabajar visualmente con los datos. Además de esta digitalización, la información clasificada es corroborada por el experto de la Unidad de Planificación Territorial del Centro EULA, en adelante *el experto*, para que este paso se ajuste a la realidad.

Como resultado del procesamiento de archivos del tipo shp<sup>4</sup> se obtuvo una única tabla (mining table) para cada criterio de estudio y para el resto del proceso, la cual contiene un *id* para cada dato (cada dato corresponde al dato de un pixel), sus coordenadas geográficas, y una serie de atributos que a priori, y según la opinión del experto, podrían ser consideradas en el proceso. En esta tabla, los atributos son datos alfanuméricos con dependencia geográfica, como por ejemplo: temperatura, distancias, alturas, entre otros. La tabla final para el proceso, contempla sólo variables que han sido preseleccionadas por sobre el total.

## 3.2. Selección de variables

La selección de variables se efectúa con el objetivo de reducir la cantidad de variables para aplicar el minado de datos. Esto, ya que el costo computacional es exponencial conforme a la cantidad de variables, además, no todas las variables tienen la misma fuerza descriptiva en la dinámica del uso de suelo.

Las selección de variables se desarrolla sobre ambas *mining tables*. Luego se comparan dichas tablas, y se presentan los resultados al experto del centro EULA.

<sup>4</sup>archivos shp: shapes de datos geográficos en formato vectorial

Los procesos de selección de variables se dividen en dos grupos: *filters* (*filtros*) y los *wrappers*:

- **Filtros:** Este proceso selecciona o filtra unas características y se excluyen otras.
- **Wrappers:** Este proceso evalúa la calidad de cada conjunto de atributos en cada momento. Se ejecuta sobre el conjunto de datos de entrenamiento, y se evalúa por medio de alguna medida, como el *accuracy* (*exactitud*).

La desventaja de los *wrappers* es el costo computacional que implica ejecutar el aprendizaje para cada subconjunto de datos. Las aproximaciones que hacen los *filtros* al ignorar los efectos de los subconjuntos involucra una desventaja para éstos.

Para el proceso de selección de variables se aplicaron distintos *filtros*. Un filtro de selección de variables, se basa en las características generales de los datos de entrenamiento y, por ende, selecciona las características independientemente de cualquier algoritmo de aprendizaje, por lo tanto no hereda sesgos de un algoritmo de aprendizaje [5].

Para la preselección de variables, se utilizó el software WEKA, y se experimentó con los siguientes procesos de selección de variables:

- **Attribute sub-set selection:** Evalúa sub-conjuntos de variables que son altamente correlacionados. Las variables se clasifican en tres categorías separadas: muy *redundante*, *poco redundante e irrelevantes*. Una variable es “redundante” si una característica es siempre necesaria para un subconjunto óptimo, no se puede quitar. Es “poco redundante” cuando no siempre es necesaria, pero puede ser necesario para un subconjunto óptimo en ciertas condiciones. Finalmente, se habla de “irrelevante” cuando no es necesaria en absoluto [16, 5].
- **Chi Squared:** Evalúa el valor de una variable mediante el cómputo del valor estadístico *chi-cuadrado*. Los números expresados para cada valor son primero discretizados en varios intervalos usando un método de discretización basado en la entropía [5].
- **Information Gain:** Evalúa la ganancia de información con respecto al conjunto de datos [5, 6].

Para cada experimento, se trabajó con dos modelos de entrenamiento-evaluación:

- **Validación cruzada:** Se divide el conjunto de datos en distintas proporciones. Dado un conjunto de datos en particiones (*folds*). Por cada parte se construye un clasificador con las  $n - 1$  partes restantes y se prueba, por cada una de las  $n$  particiones. Como por ejemplo, si se tiene 9/10, quiere decir que si un conjunto de datos se divide en 10 partes, 9 partes se usarán para entrenamiento y una parte para validación.
- **Conjunto de entrenamiento completo:** Utiliza el conjunto de datos a modo global para todo el proceso, sin división de entrenamiento o validación.

Por otra parte, el orden en que se presentan los datos en el resultado, dependerá de la forma en que se defina que vaya guardando los datos de salida:



- *Best first*: Realiza una búsqueda y ordena los resultados parciales conforme a valores obtenidos. Luego, se eliminan (podan) los otros resultados. Se utiliza frecuentemente en experimentos donde no se cuenta con gran arquitectura de soporte para los cómputos.
- *Ranker*: A diferencia del método anterior, se consideran todos los posibles resultados, y luego los ordena.

Por otro lado, en esta tesis las reglas de asociación espacial se abordan desde dos criterios:

- *Respecto al cambio o no cambio de uso de suelo*: Las variables independientes se evalúan con respecto a la acción de haber cambiado o no el uso de suelo, es decir, se evalúa qué variables se asocian entre sí para describir que haya cambiado o no el uso de suelo.
- *Respecto al tipo de cambio de uso de suelo*: Esta perspectiva intenta asociar las variables independientes con respecto a qué variables independientes se observan.

### 3.2.1. Selección de variables según el criterio de condición de cambio del uso de suelo

Aquí se buscan las variables más relacionadas con la condición de cambio en el uso de suelo presentado. Con el fin de reducir la cantidad de variables pero siempre manteniendo el criterio sobre si el punto de muestreo ha presentado o no cambios en su valor de uso de suelo.

El resultado de selección se encuentra en las tablas de la 3.1 a la 3.4.

Tabla 3.1: Variables seleccionadas con Info Gain Attribute Eval, Ranker y Cross validation 2-1, bajo el criterio de la condición de cambio o no-cambio.

Procedimiento	InfoGain Attribute Eval
Método	Ranker
validación	Cross Validation
VARIABLES	Uso de suelo 2001.
	Distancia a terrenos aptitud agrícola.
	Tamaño propiedad.
	Distancia a caminos asfaltados.
	Distancia a centros urbanos.
	Distancia a plantaciones (1984).
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Densidad de bosque nativo radio 4.500 m.
	Pendiente terreno.
	Precipitación anual.
	Densidad bosque nativo radio 2.250 m.
	Distancia a caminos ripiados.
	Densidad de matorrales radio 4.500 m.
	Elevación.
	Distancia red vial.

Tabla 3.2: Variables seleccionadas con Info Gain Attribute Eval, Ranker y Full training set, bajo el criterio de la condición de cambio o no-cambio.

<b>Procedimiento</b>	<b>InfoGain Attribute Eval</b>
<b>Método</b>	<b>Ranker</b>
<b>validación</b>	<b>Full Training Set</b>
VARIABLES	Uso de suelo 2001.
	Distancia a terrenos aptitud agrícola.
	Distancia a centros urbanos.
	Distancia a plantaciones (1984).
	Distancia a caminos asfaltados.
	Tamaño propiedad.
	Densidad bosque nativo radio 4.500 m.
	Pendiente terreno.
	Precipitación anual.
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Densidad de matorrales densidad 4.500 m.
	Densidad de bosque nativo radio 2.250 m.
	Densidad de praderas radio 4.500 m.
	Distancia caminos ripiados.
	Elevación.

Tabla 3.3: Variables seleccionadas con ChiSquared attribute eval, Ranker y Cross validation 2-1, bajo el criterio de la condición de cambio o no-cambio.

<b>Procedimiento</b>	<b>ChiSquared Attrib Eval</b>
<b>Método</b>	<b>Ranker</b>
<b>validación</b>	<b>Cross validation</b>
VARIABLES	Uso de suelo 2001.
	Distancia a terrenos aptitud agrícola.
	Distancia a caminos asfaltados.
	Tamaño propiedad.
	Distancia a plantaciones (1984).
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Densidad de bosque nativo radio 4.500 m.
	Pendiente terreno.
	Precipitación anual.
	Densidad de bosque nativo radio 2.250 m.
	Densidad de praderas radio 4.500 m.
	Densidad de matorrales radio 4.500 m.
	Distancia a caminos ripiados.
	Elevación.
	Distancia red vial.

Tabla 3.4: Variables seleccionadas con ChiSquared attribute eval, Ranker y Full training set, bajo el criterio de la condición de cambio o no-cambio.

<b>Procedimiento</b>	<b>ChiSquared Attrib Eval</b>
<b>Método</b>	<b>Ranker</b>
<b>validación</b>	<b>Full Training Set</b>
VARIABLES	Uso de suelo 2001.
	Distancia a terrenos aptitud agrícola.
	Distancia a centros urbanos.
	Distancia a plantaciones (1984).
	Distancia a caminos asfaltados.
	Tamaño propiedad.
	Densidad de bosque nativo radio 4.500 m.
	Pendiente terreno.
	Precipitación anual.
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Densidad de praderas radio 4.500 m.
	Densidad de matorrales radio 4.500 m.
	Densidad de bosque nativo radio 2.250 .
	Distancia a caminos ripiados.
	Elevación.

### 3.2.2. Selección de variables según el criterio del tipo de cambio

Aquí se asocian las variables conforme al tipo de cambio observado. Con el fin de reflejar si un uso de suelo cambió o no, se evalúa el tipo de cambio. El tipo de cambio corresponde, por ejemplo, a la dinámica acerca de un polígono cuya propiedad de uso de suelo era 1 (bosque nativo), y luego pasó a 13 (plantación). En esta tesis, se ha focalizado el estudio al cambio desde cualquier estado inicial, pero teniendo como estado final el uso de suelo del tipo “plantación”, ya que es el punto de atención del FONDECYT No. 1070511.

Los tipos de cambios observados se aprecian en la tabla 3.5. Los resultados observados se reflejan las tablas de la 3.6 a la 3.10.

Tabla 3.5: Tipos de cambio de uso de suelo observado.

<b>Tipificación tipo de cambio</b>	<b>cambio observado</b>
1	Sin cambio
2	Bosque nativo a plantación
3	Bosque nativo achaparrado a plantación
4	Matorral a plantación
5	Estepa a plantación
6	Pradera a plantación
7	Cuerpo de agua a plantación
8	Humedal a plantación

Tabla 3.6: Variables seleccionadas con CfsSubsetEval, Best First y Cross validation 2-1, bajo el criterio del tipo de cambio.

<b>Procedimiento</b>	<b>CfsSubsetEval</b>
<b>Método</b>	<b>Best First</b>
<b>validación</b>	<b>Cross Validation</b>
VARIABLES	Densidad e matorrales radio 450 m.
	Densidad de praderas radio 450 m.
	Uso de suelo 1984.
	Uso de suelo 2001.
	Densidad de praderas radio 4.500 m.U_PR50
	U_PR25

Tabla 3.7: Variables seleccionadas con InfoGain Attribute Eval, Ranker y Cross Validation 2-1, bajo el criterio del tipo de cambio.

<b>Procedimiento</b>	<b>InfoGainAttributeEval</b>
<b>Método</b>	<b>Ranker</b>
<b>validación</b>	<b>Cross Validation</b>
VARIABLES	Distancia a centros urbanos.
	Distancia a terrenos aptitud agrícola.
	Distancia a caminos asfaltados.
	Distancia a plantaciones.
	Densidad de bosque nativo radio 4.500 m.
	Densidad de bosque nativo achaparrado densidad 4.500 m.
	Pendiente terreno.
	Precipitación anual.
	Densidad de bosque nativo radio 2.250 m.
	Densidad de matorrales radio 4.500 m.
	Distancia a caminos ripiados.
	Tamaño propiedad.
	Elevación.
	Distancia red vial.
	Densidad de praderas radio 4.500 m.

### 3.2.3. Comparación y discusión de la selección de variables

Los resultados obtenidos se comparan con la selección de variables realizada por el centro EULA-Chile, el cual aplicó modelos estadísticos conforme a diferentes modelos (Análisis de autocorrelación espacial, análisis univariado y análisis multivariado). Los resultados se representan en las tablas 3.11 a la 3.13. Los modelos aplicados se clasificaron según el tipo de variable con respecto al patrón del cambio de uso de suelo (condición de cambio, no-cambio) en cuatro grupos:

- Variables ambientales: denotan presencia o ausencia de factores que limitan o potencian

Tabla 3.8: Variables seleccionadas con Info Gain Attribute Eval, Ranker y Full training set, bajo el criterio del tipo de cambio.

Procedimiento	InfoGain Attribute Eval
Método	Ranker
validación	Full training Set
VARIABLES	Distancia a terrenos de aptitud agrícola.
	Distancia a centros urbanos.
	Distancia a plantaciones (1984).
	Distancia a caminos asfaltados.
	Densidad de bosque nativo radio 4.500 m.
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Pendiente del terreno.
	Precipitación anual.
	Densidad de matorrales radio 4.00 m.
	Densidad de bosque nativo radio 2.250 m.
	Tamaño propiedad.
	Densidad de praderas radio 4.500 m.
	Distancia caminos ripiados.
	Elevación.
	Distancia a red vial.

el cambio de uso de suelo.

- Variables socioeconómicas: denotan valores demográficos y condiciones de producción agrícola.
- Variables de distancia: denota distancia de ciertos elementos que influyen en el patrón espacial del cambio de uso de suelo.
- Variables de entorno: denota densidad de algunos elementos que promueven los cambios sufridos en el área de estudio en el periodo de análisis.

Los análisis estadísticos efectuados por el Centro EULA, fueron:

1. **Análisis de autocorrelación espacial:** La autocorrelación espacial de la variable dependiente, y por motivo de la dependencia geográfica de las variables, es que podría producir pseudoréplicas, produciendo sobreajustes del modelo logrado. A modo de evaluar el efecto de autocorrelación y lograr definir una escala apropiada al muestreo, se efectuó un análisis de autocorrelación espacial sobre las celdas que representan el cambio de uso de suelo. Con este análisis, se logró reducir el efecto de autocorrelación ( $r_p < 0,65$ ) de la variable dependiente usando un muestreo sistemático de tamaño 10x10 celdas.
2. **Análisis univariado:** Para evaluar la capacidad que poseen las variables dependientes de discriminar entre celdas que sufrieron o no sufrieron cambios, en el estudio

Tabla 3.9: Variables seleccionadas con Chi-squared attribute eval, Ranker y Cross Validation 2-1, bajo el criterio del tipo de cambio.

Procedimiento	ChiSquaredAttributeEval
Método	Ranker
validación	Cross Validation 2-1
VARIABLES	Distancia a centros urbanos.
	Distancia a terrenos aptitud agrícola.
	Distancia a caminos asfaltados.
	Distancia a plantaciones.
	Densidad de bosque nativo achaparrado radio 4.500 m.
	Pendiente terreno.
	Densidad de bosque nativo radio 4.500 m.
	Densidad de matorrales radio 4.500 m..
	Precipitación anual.
	Elevación.
	Distancia red vial
	Distancia a caminos ripiados.
	Densidad de bosque nativo radio 2.250 m.
	Densidad de matorrales radio 2.250 m.
	Densidad de praderas radio 4.500 m.

se utilizó el test  $U$  de *Mann-Withney* [1]. Luego, se seleccionaron y consideraron las variables que presentaron un cambio significativo en sus valores medios entre celdas que presentaron cambios y las que no. También se efectuó un análisis de correlación entre las variables dependientes seleccionadas, eligiendo a aquellas que presentaron un *coeficiente de correlación de Spearman* [1]  $< 0,65$  y, en el caso de haber dos variables con correlaciones mayores a 0,65, se eligió a aquella con mayor sentido explicativo en relación al cambio de uso de suelo.

- Análisis multivariado:** Este análisis describe una variable de respuesta binaria de presencia y ausencia de cambios en las celdas, calculando la probabilidad de ocurrencia del evento. Se retienen en el modelo aquellas variables que aportan significativamente a la descripción del patrón espacial del cambio de uso de suelo, mejorando el ajuste global del modelo.

A nota de aclaración, los distintos modelos aplicados por el centro EULA-Chile, difieren en el radio aplicado sobre ciertas variables. Por ejemplo U\_MTxx: densidad de matorrales radio xx, U\_PRxx: densidad de praderas radio xx y U\_BAxx: densidad de bosque nativo achaparrado radio xx, entre otras (donde xx corresponde al radio, el que toma los valores: 05=450m, 10=900m, 25=2250m y 50=4500m).

Las variables seleccionadas fueron analizadas por el experto del Centro EULA-Chile. Si bien el orden o la fuerza de cada variable en los resultados difieren según cada experimento, se observa una clara similitud entre la mayoría, y coincidencia en varias de las variables escogidas por los métodos del Centro EULA-Chile.

Tabla 3.10: Variables seleccionadas con ChiSquared attribute eval, Ranker y Full training set, bajo el criterio del tipo de cambio.

Procedimiento	ChiSquared Attrib Eval
Método	Ranker
validación	Full training set
VARIABLES	Distancia a terrenos aptitud agrícola.
	Distancia a centros urbanos.
	Distancia a caminos asfaltados.
	Distancia a plantaciones.
	Densidad de bosque nativo radio 4.500 m.
	Densidad de matorrales radio 4.500 m.
	Pendiente terreno.
	Densidad bosque nativo achaparrado radio 4.500 m.
	Distancia a red vial.
	Distancia a caminos ripiados.
	Precipitación anual.
	Elevación.
	Densidad de praderas radio 4.500 m.
	Densidad de matorrales radio 2.250 m.
	Densidad de bosque nativo radio 2.250 m.

Tabla 3.11: Variables seleccionadas con el Modelo 1, del centro EULA

Procedimiento	Modelo 1
VARIABLES	Distancia a centros urbanos.
	Distancia a plantaciones.
	Distancia a terrenos aptitud agrícola.
	Temperaturas mínimas.
	Pendiente terreno.
	Tamaño propiedad.
	Densidad de matorrales radio 2.250 m.
	Densidad de praderas radio 2.250 m.
	Densidad de plantaciones radio 2.250 m.

La evaluación del experto, determinó que los modelos más apropiados para cada variable dependiente son:

- Respecto a la condición de *cambio* o *no cambio*:
  - Information Gain Evaluation, con Ranker y Cross Validation (Tabla 3.1);
  - Information Gain Evaluation, con Ranker y Full Training Set (Tabla 3.2).
- Respecto al *tipo de cambio*:
  - Chi Squared Attribute Evaluation, con Ranker y Cross Validation (Tabla 3.9);

Tabla 3.12: Variables seleccionadas con el Modelo 2, del centro EULA

Procedimiento	Modelo 2
VARIABLES	Distancia a centros urbanos.
	Distancia a plantaciones.
	Distancia a terrenos aptitud agrícola.
	Pendiente terreno.
	Tamaño propiedad.
	Densidad de matorrales radio 450 m.
	Densidad de bosque nativo radio 450 m.
	Densidad de bosque nativo achaparrado radio 450 m.
	Densidad de praderas radio 450 m.

Tabla 3.13: Variables seleccionadas con el Modelo 3, del centro EULA

Procedimiento	Modelo 3
VARIABLES	Distancia a centros urbanos.
	Distancia a plantaciones.
	Distancia a terrenos aptitud agrícola.
	Pendiente terreno.
	Tamaño terreno.
	Densidad de matorrales radio 900 m.
	Densidad de praderas radio 900 m.

- ChiSquared Attribute Evaluation, con Ranker y Full Training Set (Tabla 3.10).

De las 52 variables para estudio, 5 de las seleccionadas por los procesos antes mencionados coinciden con la selección hecha en forma independiente por el Centro EULA-Chile. Estas variables seleccionadas son:

- Distancia a terrenos de aptitud agrícola.
- Distancia a centros urbanos.
- Distancia a plantaciones.
- Tamaño propiedad.
- Pendiente terreno.

Cabe destacar que dentro de las variables asociadas con mayor fuerza descriptiva a la dinámica del uso de suelo están las distancias a red vial, así como las pequeñas áreas cercanas a áreas que ya presentan un uso de suelo del tipo *plantación*. Ambas variables, corresponderían efectivamente a la facilidad de acceso a dichas áreas, ya sea por la red vial, o por ser la continuidad de una área que ya presenta un uso de suelo de *plantación*. Además, las variables *densidad de bosque nativo*, *densidad de bosque nativo achaparrado*, *densidad de matorrales*, se repiten pero con un radio diferente (450 m).



A juicio del experto, el grado de interés en los modelos seleccionados obedece a la alta correlación que tienen efectivamente los caminos aledaños a las áreas que presentarían uso de suelo del tipo “plantaciones”(figura 3.2). También los usos de suelo vecinos de usos de suelo que previamente presentan uso de suelo del tipo “plantaciones”, ya que estas áreas son más propensas a que sus propietarios cambien el uso de suelo actual por uno que les brinde mayores ingresos económicos. Otra variable seleccionada por el método planteado en esta tesis, es el tamaño de la propiedad. Pero en el método del Centro EULA, estas quedan fuera. Estas variables, y observando en terreno, si son significativas como por ejemplo: distancia y densidad de caminos, proporción (densidad) de praderas y matorrales y elevación.

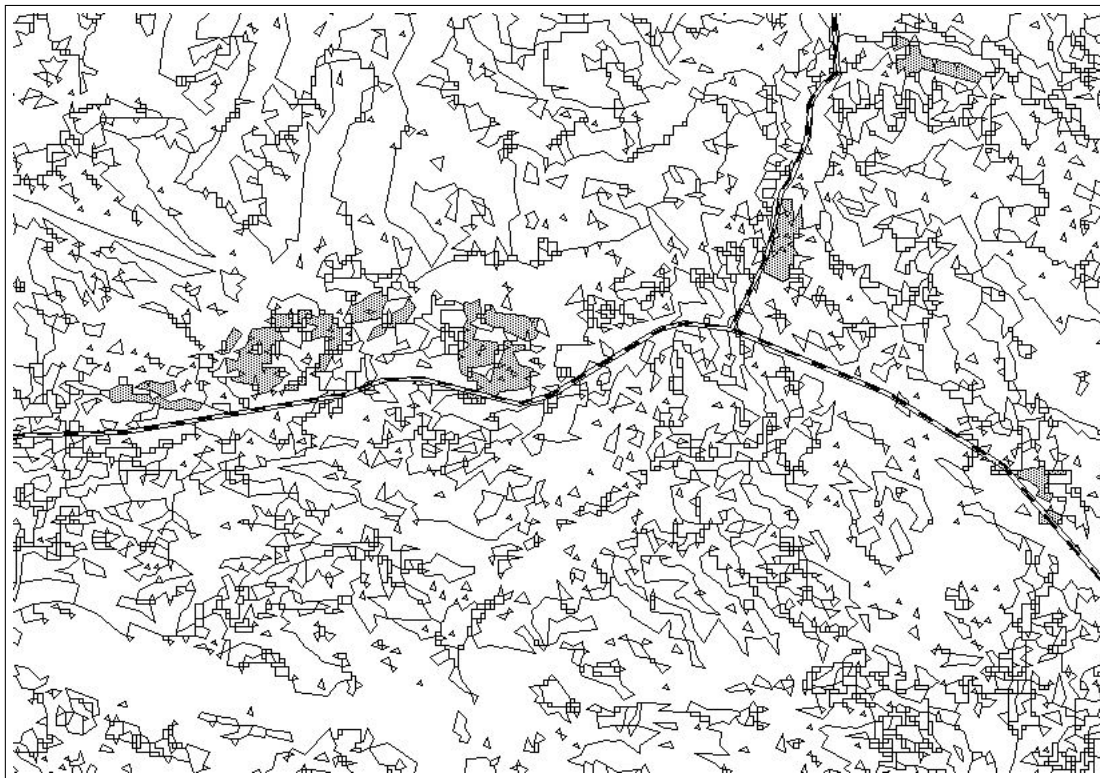


Figura 3.2: Plantaciones cerca de la red vial

### 3.3. Reglas de asociación

La generación de reglas de asociación, permite poder asociar las distintas variables seleccionadas en el área de estudio que describan la dinámica del uso de suelo. En la tabla 3.14 se aprecia la lista de los 14 distintos usos de suelo presentes en el estudio y la cantidad de polígonos y el área total ocupada por cada uso de suelo presentes en el área de estudio.

Para la generación de reglas de asociación, se utilizan sólo las 15 variables con mayor fuerza descriptiva, obtenidas mediante el proceso de selección de variables.

Tabla 3.14: Usos de suelo presentes en el área de estudio

id	Uso de suelo	cantidad de polígonos	superficie (Ha.)
1	Bosque nativo	13.440	671.279
2	Bosque nativo achaparrado	16.556	195.613
3	Matorral	19.574	184.847
4	Estepa	5.237	154.611
5	Pradera	10.998	102.261
6	Agrícola	7.199	75.469
7	Cuerpos de agua	1.112	59.701
8	Nieves y glaciares	3.715	226.810
9	Áreas sin vegetación	9.756	174.238
10	Humedales	3.141	35.167
11	Sombras	6.218	76.349
12	Estepa Altoandina	7.646	60.662
13	Plantaciones	452	19.400
14	Urbano	20	1.269

### 3.3.1. Reglas generadas

Para la generación de las reglas de asociación se aplicó el algoritmo *apriori predictive* [2], considerando la selección de variables validadas por el experto del EULA-Chile. El algoritmo *apriori predictive*, utiliza una combinación del *support* y *confidence*.

Como resultado del proceso, y como se esperaba, se obtuvo una larga lista de reglas de asociación, algunas sin mayor relevancia, las que fueron analizadas por el experto. El listado completo de reglas se expone en el anexo 6.3.

La generación de reglas se realizó según dos criterios: la condición de cambio (cambio o no-cambio), y del tipo de cambio observado. Para los experimentos se utilizaron 15 variables, salvo en el último experimento (modelo 5), que por capacidad de cómputo se usaron 14 variables por cada modelo. Los modelos utilizados son los seleccionados previamente en conjunto con el experto.

Luego estas reglas se agrupan para hacerlos más entendibles y útiles al estudio.

### 3.3.2. Agrupación de reglas

De la primera lista de reglas generadas se descartan las reglas que no reflejan cambios en el uso de suelo. Para el análisis final, se agrupan aquellas reglas que describen una misma dinámica de cambio de uso de suelo:

- *Bajo el criterio de CAMBIO*: En las tablas 3.15 y 3.16, se presentan la agrupación de las reglas obtenidas. Al hablar de distancias, se entenderá en adelante valores expresados en metros, así mismo, al referirse a datos de densidad en radio  $x$ , se entiende que es una zona circular con una radio  $x$  expresado en metros. Este tipo de experimento puede verificarse por medio de SW del tipo SIG, donde se puede recorrer una capa determinada con una herramienta del tipo circular, regulando el radio dentro de un polígono con esas

características de uso de suelo (praderas, bosque nativo, bosque nativo achaparrado, etc.)

Tabla 3.15: Agrupación de reglas obtenidas con el modelo 2.

<b>Modelo 2</b>		
	Regla	Accuracy
1	Tamaño propiedad $\geq 492,4666 - 4854,2798 \implies$ Uso de suelo 2001 = 13(Plantación)	0,87-0,94
2	Tamaño propiedad $\geq 49620,9805 \implies$ Uso de suelo 2001 = 4(Estepa)	0,889
3	Precipitación anual $\geq 4000$ y Tamaño propiedad $\geq 73718,1953 \implies$ Uso de suelo 2001=8(Nieves y glaciares)	0,87
4	Densidad de bosque nativo radio 4500 m $\geq 0,00 \implies$ Densidad de bosque nativo radio 2250 $\geq 0,000$ y Uso de suelo 2001 = 4(Estepa)	0,81
5	Distancia a centros urbanos $\geq 14799,4590 \implies$ Uso de suelo 2001 = 1(Bosque nativo)	0,75

*Nota: La segunda regla expuesta, es re-validada ya que representa un hecho real observado durante el estudio.*

- *Bajo el criterio de TIPO DE CAMBIO:* En las tablas 3.17 y 3.18, se presentan la agrupación de las reglas para la perspectiva del tipo de cambio observado. En este caso, la interpretación que se debe hacer, es que el sujeto de la regla, se asocia al cambio observado y denotado por el predicado de dicha regla. Al igual que con las reglas obtenidas bajo el criterio de cambio o no-cambio, las distancias se expresan en metros, usando un software del tipo SIG.

La mayoría de las reglas de asociación se extraen al considerar como variables independientes las condiciones de *cambio o sin-cambio*, lo que coincide con la perspectiva usada, por experiencia, por el centro EULA.

### 3.3.3. Validación de reglas de asociación

El conjunto final de reglas fue evaluada por el experto, ya que es él quien con su conocimiento previo, puede validar estos resultados. Es decir, determina las reglas que son de interés y de utilidad para el estudio.

El principal cambio en el uso de suelo fue el avance de las plantaciones (uso de suelo 13) sobre praderas(uso de suelo 5) y matorrales(uso de suelo 3). Las reglas validadas son las siguientes:

- Según criterio de la condición de cambio:
  - Modelo 2:
    1. Tamaño propiedad  $\geq 492-4854 \implies$  Uso de suelo 2001 = 13(Plantación) (acc=0,87-0,94)

Tabla 3.16: Agrupación de reglas obtenidas con el modelo 3.

<b>Modelo 3</b>		
	Regla	Accuaracy
1	Tamaño propiedad = 496,4666 - 4854,2798 $\implies$ Uso de suelo 2001 = 13(Plantación)	0,87-0,94
2	Tamaño propiedad $\geq$ 49620,9805 y Densidad de bosque nativo radio 4500 $\geq$ 0,00 $\implies$ Uso de suelo 2001 = 4(Esteba)	0,889-0,890
3	Precipitación anual $\geq$ 4000 y Tamaño propiedad $\geq$ 73718,1953 $\implies$ Densidad de praderas radio 4500 $\geq$ 0,0 y Uso de suelo 2001=8(Nieves y glaciares)	0,87
4	Tamaño propiedad $\geq$ 49620,9805 y Densidad de bosque nativo achaparrado radio 4500 $\geq$ 0,00 $\implies$ Densidad de bosque nativo radio 4500 $\geq$ 0,00 y Uso de suelo 2001=4(Esteba)	0,87
5	Densidad de bosque nativo radio 4500 $\geq$ 0,00 y Densidad de bosque nativo achaparrado radio 4500 $\geq$ 0,00 $\implies$ Tamaño propiedad $\geq$ 49620,9805 y Uso de suelo 2001 = 4(Esteba)	0,874
6	Densidad de bosque nativo radio 2250 $\geq$ 0,00 y Densidad de bosque nativo achaparrado radio 4500 $\geq$ 0,00 $\implies$ Tamaño propiedad $\geq$ 49620,9805 Uso de suelo 2001=4(Esteba)	0,874
7	Densidad de bosque nativo radio 2250 $\geq$ 0,00 $\implies$ Densidad de bosque nativo radio 2250 $\geq$ 0,00 y Uso de suelo 2001 = 4(Esteba)	0,81
8	Distancia a centros urbanos $\geq$ 14799,4590 $\implies$ Uso de suelo 2001 = 1(Bosque nativo)	0,75

Tabla 3.17: Agrupación de reglas obtenidas con el modelo 4.

<b>Modelo 4</b>		
	Regla	Accuaracy
1	Distancia a centros urbanos $\geq$ 14799,459 $\implies$ Uso de suelo 2001=1(Bosque nativo)	0,749-0,750
2	Densidad de bosque nativo radio 4500 m $\geq$ 0,00 $\implies$ Densidad de bosque nativo radio 2250 $\geq$ 0,00 y Uso de suelo 2001=4(Esteba)	0,813

2. Distancia a centros urbanos  $\geq$  14799  $\implies$  Uso de suelo 2001 = 1(Bosque nativo)  
(acc=0,75)

• Modelo 3:

1. Tamaño propiedad  $\approx$  496 - 4854  $\implies$  Uso de suelo 2001 = 13(Plantación)

Tabla 3.18: Agrupación de reglas obtenidas con el modelo 5.

Modelo 5		
	Regla	Accuaracy
1	Distancia a centros urbanos $\geq 14799,459 \implies$ Uso de suelo 2001=1(Bosque nativo)	0,749-0,750
2	Densidad de bosque nativo radio 4500 $\geq 0,00 \implies$ Densidad de bosque nativo radio 2250 $\geq 0,00$ y Uso de suelo 2001=4(Estepa)	0,813
3	Densidad de bosque nativo radio 4500 $\geq 0,00$ y Densidad de bosque nativo achaparrado radio 4500 $\geq 0,00 \implies$ Densidad de bosque nativo radio 2250 $\geq 0,00$ y Densidad de bosque nativo achaparrado radio 4500 $\geq 0,00$	0,875
4	Densidad de bosque nativo radio 2250 $\geq 0,00$ y Densidad de bosque nativo achaparrado radio 4500 $\geq 0,00 \implies$ Densidad de bosque nativo radio 4500 $\geq 0,00$ y Uso de suelo 2001 = 4(Estepa)	0,875

(acc=0,87-0,94)

2. Distancia a centros urbanos  $\geq 14799 \implies$  Uso de suelo 2001 = 1(Bosque nativo)  
(acc=0,75)

- Según el criterio de la condición del tipo de cambio:

- Modelo 4:

1. Distancia a centros urbanos  $\geq 14799 \implies$  Uso de suelo 2001 = 1(Bosque nativo)  
(acc=0,749-0,750)

Se observa que, del total de reglas generadas por el experimento, sólo 5 reglas son finalmente validadas. De hecho, del modelo 5 no se validó alguna regla.

Ahora, para un mejor entendimiento, estas reglas se pueden concentrar y traducir de la forma siguiente:

- Según el criterio de la condición de cambio - no cambio:

1. Si el tamaño de la propiedad es mayor a 492 m<sup>2</sup> y menor que 4854 m<sup>2</sup>, se asocia un uso de suelo del tipo Plantación (uso de suelo 13), con una precisión entre 87 % y 94 %.
2. Si la distancia a centros urbanos es cercana o mayor a los 14799 m, se asocia un uso de suelo del tipo Bosque Nativo (Uso de suelo 1), con una precisión del 75 %.

- Según el criterio de la perspectiva del tipo de cambio observado:

1. Si la distancia a centros urbanos es cercana o mayor a los 14799 m, se asocia un uso de suelo del tipo Bosque Nativo (Uso de suelo 1), con una precisión del 75 %.

Con este último proceso de ajuste, del total de reglas, se validan 3 de ellas, 2 bajo el criterio de la condición de cambio - no cambio, y una para el criterio del tipo de cambio observado.



# Capítulo 4

## Generación de reglas de colocalidad

Otro medio para asociar variables en un entorno geográfico, son las reglas de colocalidad. En esta tesis, se busca asociar los distintos usos de suelo, independiente de las variables ambientales, sociales y económicas, esto es, se decide experimentar con métricas que a diferencia del método del capítulo anterior, evalúan la condición geográfica de los polígonos presentes, esto es, evaluando su grado de adyacencia, superficie y perímetro total.

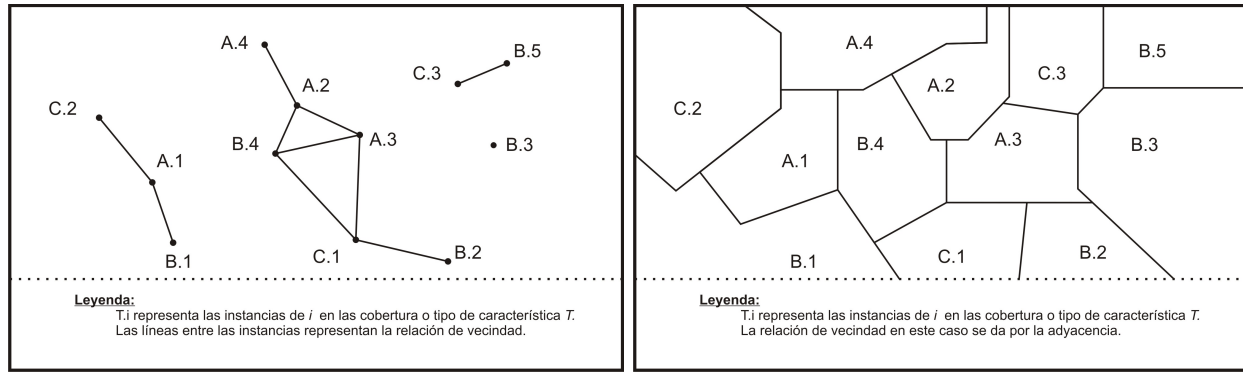
Para las reglas de colocalidad se generan tablas de vecindad(adyacencia) topológica a través de distintos procesos sobre el servidor de base de datos espaciales implementado para la tesis. Para las reglas de colocalidad, se utiliza un conjunto de datos que proviene de la clasificación de imágenes satelitales, convertidas luego a formato raster, y posteriormente, son ingresadas en formato vectorial a la bases de datos espaciales.

Las reglas de colocalidad conocidas utilizan un predicado de proximidad binario, el cual determina si un punto está *cerca o no cerca* de otro punto. La generación de reglas de colocalidad surge a raíz de la necesidad y utilidad que prestan debido a que en ciertos casos, al clasificar una imagen satelital no se puede determinar con precisión qué es lo que aparece realmente en dicha imagen, como consecuencia a sombras o distorsiones rediométricas producto de problemas del sensor o perturbaciones atmosféricas (nubes y otras). En el caso de esta tesis, y el estudio del cambio de uso de suelo en la región norpatagónica de Chile, el aporte es poder asociar polígonos de distintos usos de suelo, considerando para el estudio la totalidad del área de estudio. Con esto, se plantea contribuir a entender la distribución y asociación de los distintos usos de suelo y, como trabajo futuro, poder predecir el uso de suelo de las áreas no identificadas ya sea por sombras o nubes. Las reglas de colocalidad encontradas, son la base de un proceso de clasificación.

Esta tesis trabaja con polígonos de uso de suelo y utiliza el concepto de vecindad o adyacencia entre polígonos realizando un refinamiento métrico que cuantifica el largo de dicha adyacencia para redefinir la métrica de *Pr* (índice de participación). La figura 4.1 ilustra gráficamente la diferencia conceptual entre las reglas de colocalidad entre puntos y entre polígonos.

La diferencia con la reglas de asociación espaciales radica en que no se trata de una noción natural de transacciones espaciales. En el caso particular de este estudio, se utilizó la *vecindad de polígonos*, teniendo en cuenta un radio de participación, que corresponde a los polígonos inmediatamente vecinos a un polígono particular. Este trabajo se enfoca en reglas de colocalidad entre pares de regiones ( $k = 2$ ). Para estas reglas, dos polígonos son vecinos





(a) Modelo concéntrico(puntos).

(b) Adaptación del modelo a polígonos.

Figura 4.1: Modelo concéntrico de puntos y polígonos

si su intersección contiene sólo puntos que caen en la unión de sus bordes. Dos ejemplos que describen la vecindad geográfica se muestran en la figura 4.2.

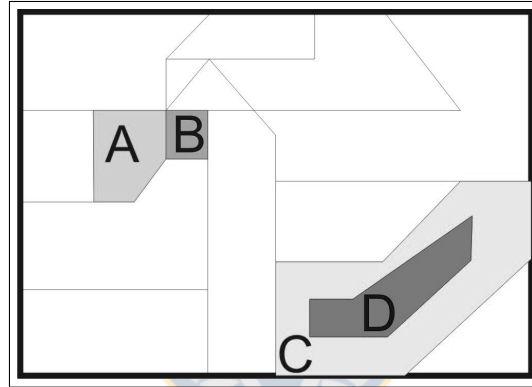


Figura 4.2: Vecindad de polígonos con cardinalidad 2 (pares de polígonos)

Para determinar el grado en que dos polígonos son adyacentes, la relación de adyacencia se cuantifica en términos de la proporción de los bordes que ambos polígonos comparten, respecto al perímetro total de cada polígono.

Basado en [15], las reglas de colocalidad fueron validadas usando dos métricas: razón de participación y probabilidad condicional de la regla. Sin embargo, la razón de participación, se ajusta con un índice de participación. Luego, el Radio de Participación (en adelante razón de participación) es modificado para manejar polígonos, dando lugar a dos tipos de radios de participación:

1.  $Pr_1$ : En vez de usar distancias para determinar si dos puntos son vecinos, se utiliza la longitud de la adyacencia entre dos polígonos. Así, se redefine la razón de participación o *Participation ratio* como un indicador de la fuerza con que un polígono de un tipo de uso particular participa en un regla de colocalidad. Esto se define como:

$$Pr_1(A, B) = \frac{\sum length(intersection(a_i, b_j) | touch(a_i, b_j))}{\sum_{\forall k} perimeter(a_k)}, \quad (4.1)$$

donde  $A$  y  $B$  son dos tipos de suelo,  $length()$  es el largo de la intersección de los bordes de dos instancias  $a_i$  y  $b_j$  de uso de suelo  $A$  y  $B$  que son adyacentes (*touch*) y  $perimeter()$  es el perímetro de una instancia de uso de suelo.

2.  $Pr_2$ : Determina el perímetro total del polígono que participa de dicha adyacencia. Y está definida como:

$$Pr_2(A, B) = \frac{\Sigma perimeter(a_i) | touch(A, B)}{\Sigma_{\forall k} perimeter(a_k)}. \quad (4.2)$$

La segunda métrica (*probabilidad condicional*), no sufre modificaciones, excepto por el hecho de que en vez de evaluar proximidad se evalúa adyacencia.

$$Cp(A, B) = \frac{count(A) | touch(A, B)}{count p(A)}, \quad (4.3)$$

La forma de medir las reglas es asimétrica, esto atendiendo a que se estima a priori que la fuerza de una regla dependerá tanto del tramo de adyacencia de dos polígonos, así como el perímetro de cada polígono involucrado. Así, si el polígono  $A$  tiene un tamaño mayor que el polígono  $B$ , se espera que la regla tenga una fuerza mayor desde el polígono  $A$  al  $B$ , que del polígono  $B$  al  $A$  (ver figura 4.3).

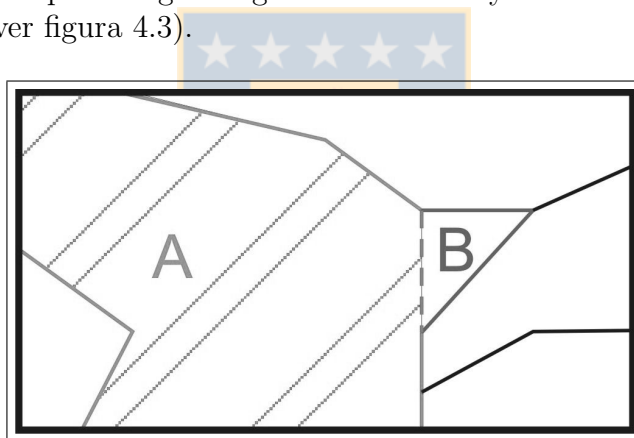


Figura 4.3: Comparación de dos polígonos, A y B, adyacentes pero donde A es mucho mayor que B.

## 4.1. Obtención de datos

El proceso de generación de una regla de colocalidad comienza con la catalogación de las imágenes del área de estudio. Luego, se construye un archivo del tipo *raster* para determinar los usos de suelo que están presentes (figura 4.4). Posteriormente, se convierte a formato *shape* (Archivo vectorial) y se realiza una extracción de propiedades geométricas y topológicas (figura 4.5), y se limpian los datos para eliminar información irrelevante.

Para nuestro estudio, se busca encontrar reglas de colocalidad de cercanía de tipos de uso de suelo dado distintos polígonos adyacentes. Por ejemplo, qué tan asociado está un polígono cuyo uso de suelo sea *bosque nativo* a un polígono cuyo uso de suelo sea *pradera*. Para esto

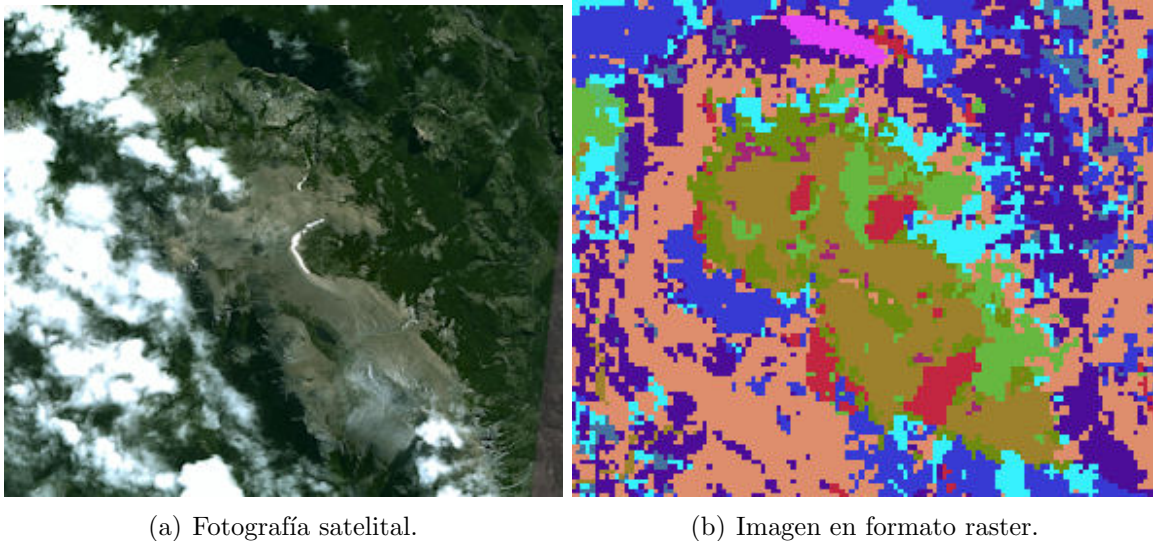


Figura 4.4: Proceso de conversión desde una imagen satelital a raster

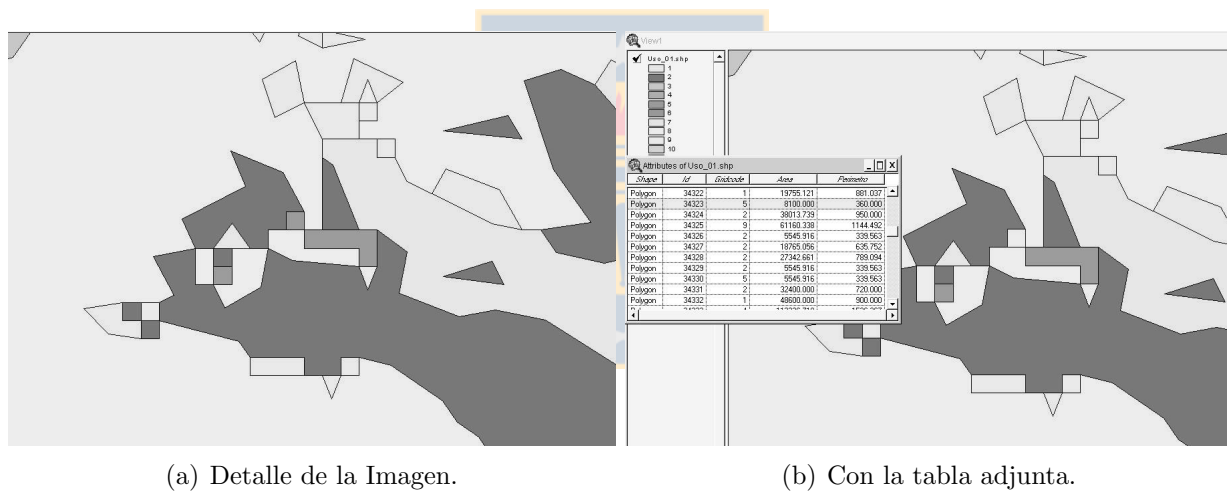


Figura 4.5: Un archivo raster convertido en ARCview a shape

se aplica la métrica sobre el total de reglas obtenidas con  $k = 2$ , donde  $k$  define que se consideran vecindades de dos polígonos. Por ejemplo, con  $Pr_1$  o  $Pr_2$  (ecuaciones 4.1, 4.2) se calcula la fuerza de cada regla, para luego establecer un umbral mínimo para filtrar a priori los resultados. Posteriormente, el experto determina cuáles son efectivamente interesantes. Por ejemplo, si se tiene la regla : Bosque nativo  $\rightarrow$  Bosque nativo achaparrado (0.4 , 0.75), significaría que con una probabilidad del 75 %, y con un radio de participación de 0,4, un polígono de uso de suelo de Bosque Nativo, está asociado a un polígono de Bosque Nativo Achaparrado, dentro del mismo plano.

## 4.2. Preparación de los datos

La extracción de las reglas de colocalidad se realiza sobre una base de datos que contiene regiones correspondientes a diferentes usos de suelo. El origen de estos datos corresponde a una imagen satelital clasificada por el centro EULA, de la cual se reconocen límites de regiones con diferentes usos de suelo. Sin embargo, este proceso automático de clasificación de uso de suelo genera regiones cuyas geometrías y dimensiones distorsionan la extracción de reglas de colocalidad. Consecuentemente, se comienza con una limpieza de los datos que elimina geometrías cuya dimensión imposibilita el manejo adecuado de los cálculos de vecindad.

Esta tarea se realizó utilizando como motor a PostgreSQL<sup>1</sup> con PostGIS. PostGIS añade al sistema de gestión de bases de datos, soporte para archivos geográficos, permite manejar archivos con datos espaciales, vale decir manejar campos de *geometría* (puntos, líneas y polígonos) y, con esta característica, transforma PostgreSQL en una Base de Datos Espaciales que puede ser utilizada en un Sistema de Información Geográfica.

La limpieza de datos permite descartar datos que pudieran distorsionar los resultados del estudio. Así se eliminaron del área de estudio todos aquellos polígonos cuya superficie fuera menor a  $9,000m^2$ . En la figura 4.6, se representan las geometrías de los polígonos con menor superficie presentes en la zona de estudio. De ellas, las dos más pequeñas poseen una superficie menor a  $9,000m^2$ , de las cuales, un triángulo, corresponde al polígono con la superficie más pequeña dentro del área de estudio, y la segunda, un cuadrado, a la segunda superficie más pequeña.

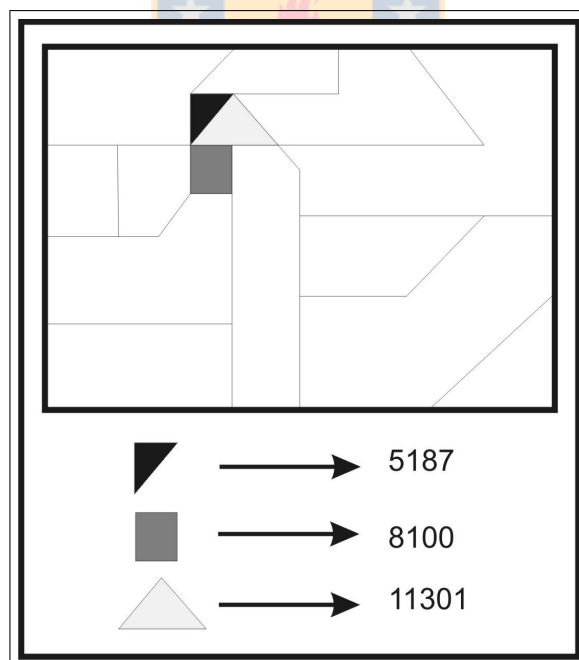


Figura 4.6: Muestra de las superficies más pequeñas.

Los 37.992 polígonos eliminados corresponden al 36 % del total de las tuplas presentes en el área de estudio. Como dato de referencia se puede mencionar que la región más grande

<sup>1</sup>Sistema de Gestión de Bases de Datos relacional, orientada a objetos, bajo licencia BSD (Berkeley Software Distribution, licencia de software libre).

posee una superficie de  $158.697,6Ha$ , la menor  $5.187m^2$ , siendo el promedio de todas las superficies presentes igual a  $193.942 m^2$ . De aquí podemos desprender que las superficies menores a  $9.000m^2$  no afectan a los resultados esperados.

Un aspecto que se tuvo que considerar fue el de las vecindades por esquina, esta vecindad ocurre cuando la intersección de los bordes resulta en un sólo punto de adyacencia o vecindad. En estos casos, el valor de su *adyacencia* tiende a cero y consecuentemente anula los valores calculados de las métricas propuestas (Pr y Cp). Este tipo de casos se aprecia mejor en la figura 4.7. Estos casos de adyacencia fueron descartados para el estudio.

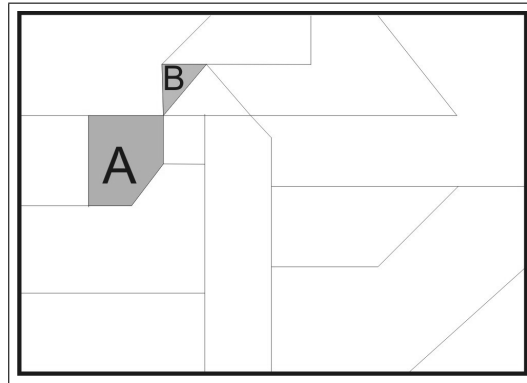


Figura 4.7: Ejemplo de un  $touches \approx 0$ .

Tras esta limpieza de datos, se genera una tabla que contiene toda la información para extraer las reglas de *colocalidad* y las métricas. La estructura de la tabla que se genera contiene los siguientes atributos:

- **Superficie de cada región:** Permite obtener la superficie de cada región. En la tabla generada para el cálculo de las reglas con  $k=2$  se considera dos áreas, correspondiente al área del polígono *A* y del polígono *B*, respectivamente.
- **Perímetro de cada región:** Esto se calculó con ArcGis, y al igual que la superficie, es un dato que se utiliza en la comparación de regiones. En la tabla estos valores corresponden a los atributos *perimetro\_1* y un *perimetro\_2*, correspondiente al perímetro del polígono *A* y el *B*, respectivamente.
- **Adyacencia:** Determina el grado de adyacencia compartida por dos polígonos.

Para la generación de las reglas de colocalidad candidatas se genera una nueva tabla general que contenga todas las vecindades de polígonos con  $k = 2$ . Luego se limpia de la tabla todas aquellas vecindades con  $touches=0$ .

La consulta (en PostGIS) ejecutada para generar esta nueva tabla se puede apreciar en el anexo 6.4.1. Ejecutada la consulta, se obtiene una tabla con todos los datos y todos los posibles vecinos entre dos polígonos dentro del área de estudio. La consulta ejecutada para obtener los porcentajes, se puede apreciar en el anexo 6.4.2.

Luego se generan distintas tablas por cada uso de suelo, y sus distintos vecinos, con el fin de obtener todas las posibles combinaciones de cada polígono de un determinado uso de suelo, con respecto a los otros 13 usos de suelo. Para esto, se ejecuta una nueva consulta (Ver

anexo 6.4.3) que genera una tabla para cada uso de suelo presente en el área de estudio, y todas sus relaciones de vecinos, concentrando además los valores de adyacencia, perímetros, y superficie, entre otros. Ya con estas tablas independientes para cada uso de suelo, se puede decir que se tiene la materia prima para generar las reglas de *colocalidad* candidatas, aplicar las métricas y finalmente, ser presentadas al experto.

### 4.2.1. Evaluación de las reglas obtenidas

Para poder evaluar las reglas de colocalidad se calculan las métricas propuestas para el estudio. Basados en [15], se utilizaron los coeficientes  $Pr_1$ ,  $Pr_2$  y  $Cp$  definidos previamente. A partir de los  $Pr_1$  o  $Pr_2$  se definen los índices de participación respectivos, los cuales finalmente sirven como medida de validación de las reglas de colocalidad, que tiene la fórmula  $C_1 \rightarrow C_2(p, cp)$ . Donde:

- $C_1$  y  $C_2$ , son los tipos de usos que tienen una propiedad de vecindad o colocalidad
- $p$ , es la medida de prevalencia (en este caso el índice de participación definido para  $Pr_1$  o  $Pr_2$ ).
- $cp$ , es la probabilidad condicional de la regla de colocalidad.

Para el cálculo de estas métricas, se desarrolla un script en *php*, el cual se calcula y se guarda en un archivo *csv*<sup>2</sup>, lo que facilita finalmente poder graficar los resultados para una mejor comparación y selección de las reglas de *colocalidad* finales. El código fuente que se desarrolló se encuentra en el anexo 6.5.

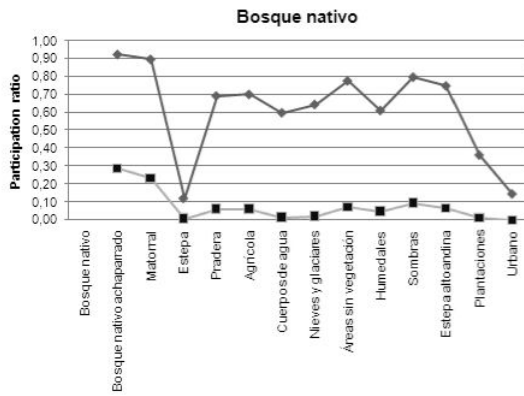
Para comparar los valores obtenidos, todos los resultados se han graficado para un tipo de uso de suelo. Además, el gráfico se ajusta a la regla obtenida para cada uno de los usos de suelo presentes en el estudio (estos gráficos se aprecian en las figuras 4.8 a la 4.14). El objetivo de la *probabilidad condicional* es indicar con qué probabilidad cada regla de colocalidad obtenida se presenta en un caso real. Cabe señalar que los resultados, no son simétricos (la relación del uso de suelo Bosque Nativo con Plantación no es equivalente a la relación del uso de suelo Plantación con Bosque Nativo), ya que la métrica utilizada para la razón de participación, considera los perímetros de ambos usos de suelo, haciendo la diferencia cuando la relación es en una u otra dirección. La probabilidad condicional sufre la modificación respecto su definición original [15] en la definición de proximidad en términos de adyacencia topológica.

Los resultados del  $Pr_1$  y  $Pr_2$  para todos los tipos de uso de suelo considerados en este estudio se presentan en los gráficos de las figuras 4.8 a las 4.14, donde el eje  $x$  identifica cada uno de los 14 tipos de uso de suelo, diferente al usado como referencia en el gráfico, y el eje  $y$  la métrica de validación.

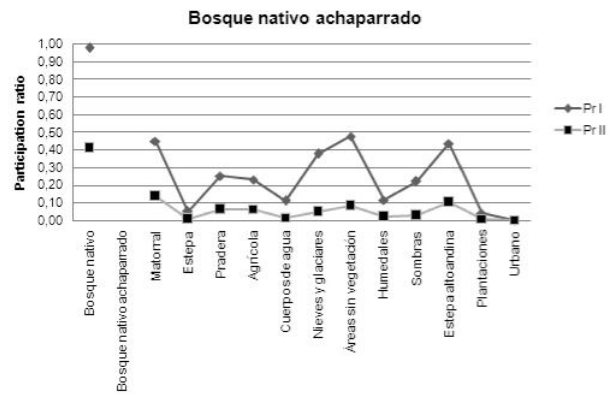
---

<sup>2</sup>csv: comma separated text, texto separado por comas



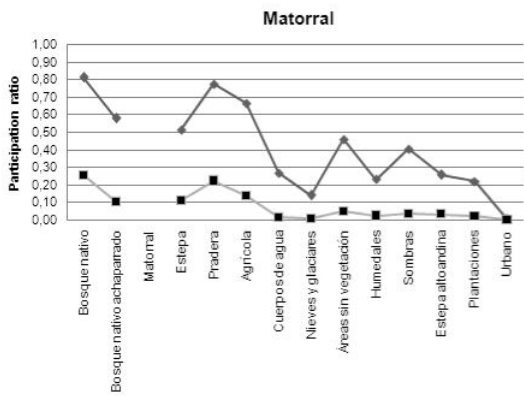


(a) Pr de Bosque nativo.

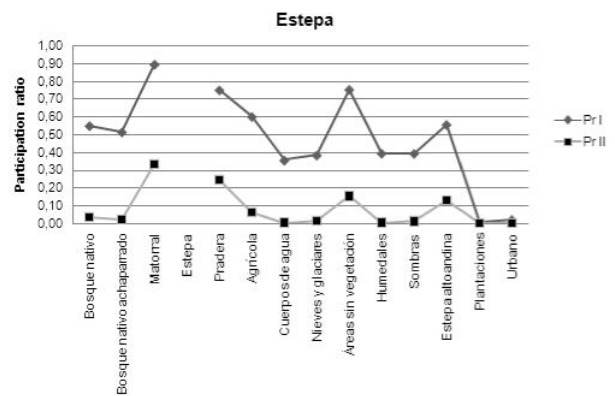


(b) Pr de Bosque nativo achaparrado.

Figura 4.8: Participation index para Bosque nativo y Bosque nativo achaparrado .

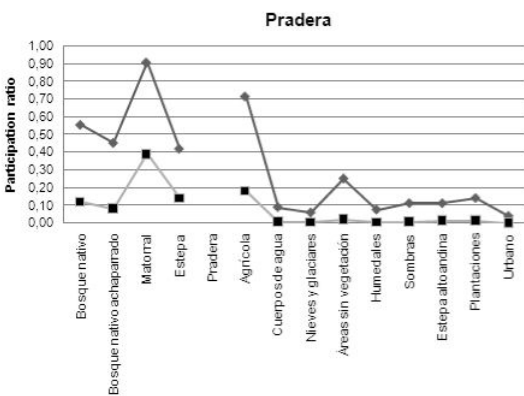


(a) Pr de matorral.

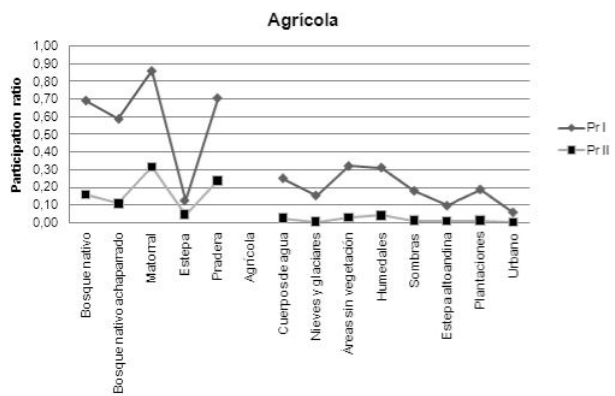


(b) Pr de estepa.

Figura 4.9: Radio de participación para Matorral y Estepa.



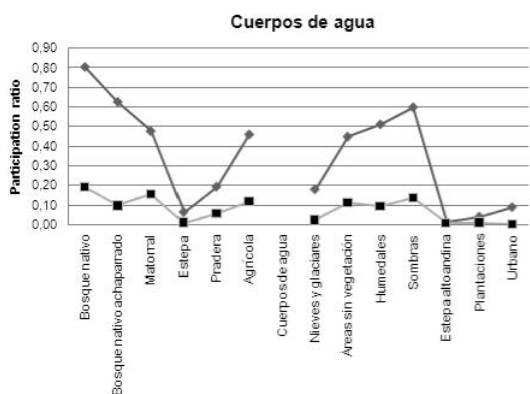
(a) Pr de praderas.



(b) Pr de suelo agrícola.

Figura 4.10: Radio de participación para Praderas y Suelo agrícola.



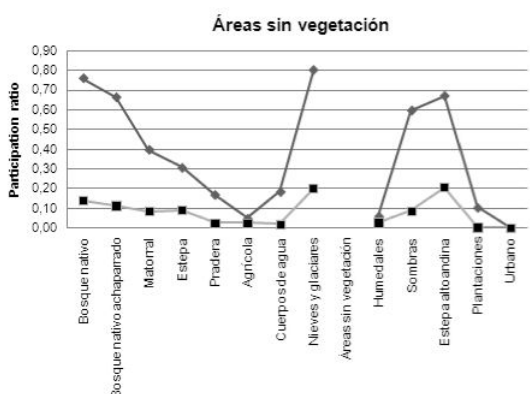


(a) Pr de cuerpos de agua.

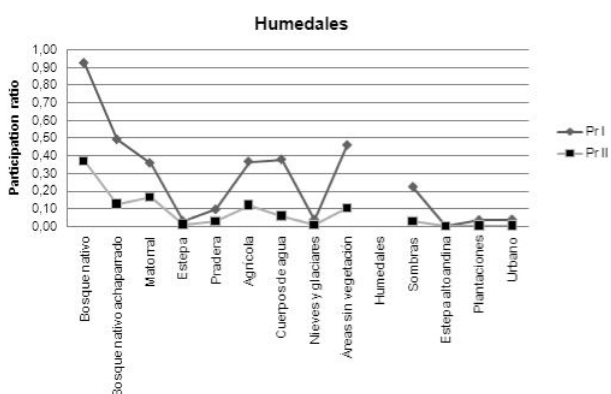


(b) Pr de nieves y glaciares.

Figura 4.11: Radio de participación para Cuerpos de agua y Nieves y glaciares.

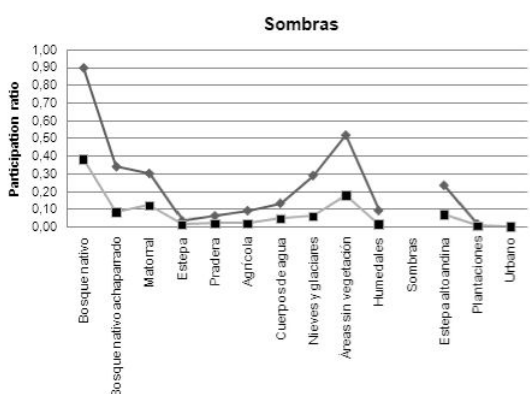


(a) Pr de áreas sin vegetación.

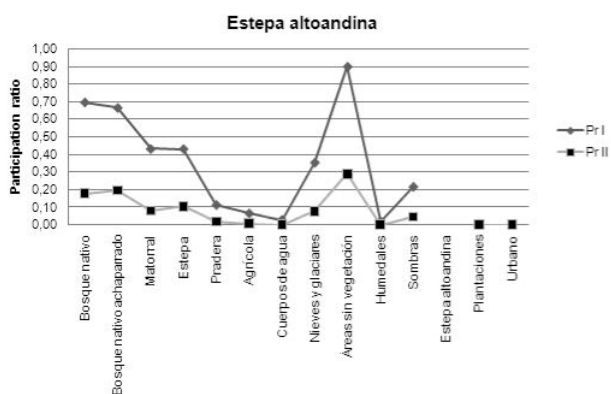


(b) Pr de humedales.

Figura 4.12: Radio de participación para Áreas sin vegetación y Humedales.



(a) Pr de sombras.



(b) Pr de estepa altoandina.

Figura 4.13: Radio de participación para Sombras y Estepa alto andina.

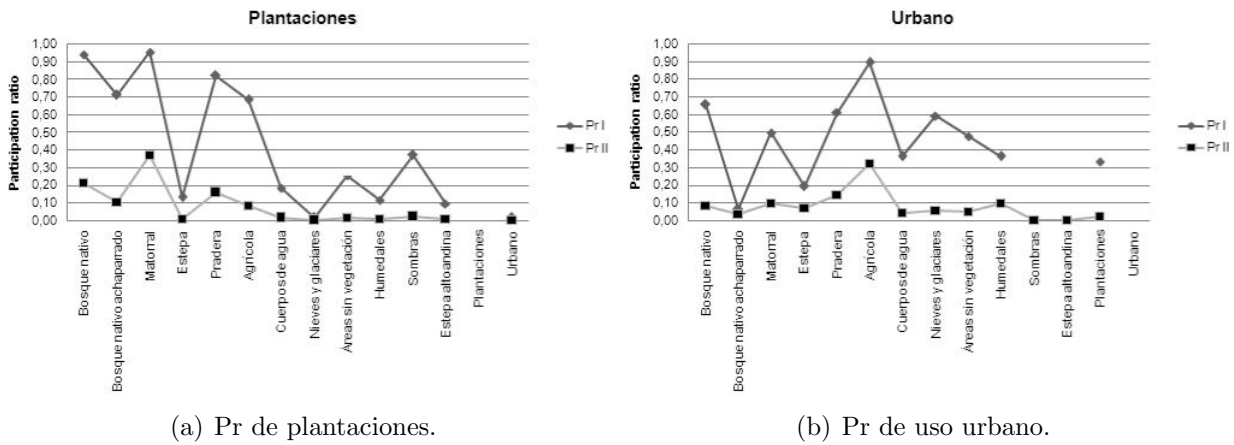


Figura 4.14: Radio de participación para Plantaciones y Suelo urbano.

Una caracterización de las diferencias posibles entre ambas razones de participación se obtiene mediante la correlación de estos valores, tal como se aprecia en la tabla 4.1, y el gráfico 4.15. Se observa una fuerte tendencia de correlación entre ambos índices, existiendo mayores diferencias entre  $Pr_1$  y  $Pr_2$  cuando el tipo de suelo a considerar es *Bosque Nativo*, *Estepa* y *Urbano*.

Tabla 4.1: Índice de correlación lineal para  $Pr_1$  y  $Pr_2$ .

id	Uso de suelo	Coefficiente de correlación
1	Bosque nativo	0,72
2	Bosque nativo achaparrado	0,94
3	Matorral	0,93
4	Estepa	0,79
5	Praderas	0,94
6	Agrícola	0,94
7	Cuerpos de agua	0,94
8	Nieves y glaciares	0,80
9	Áreas sin vegetación	0,91
10	Humedales	0,94
11	Sombras	0,98
12	Estepa alto andina	0,98
13	Plantaciones	0,88
14	Urbano	0,74

Al igual que la razón de participación, se graficó la probabilidad condicional de cada uso de suelo en función de los otros usos de suelo. Tal como se muestra en las figuras 4.16 a la 4.22.

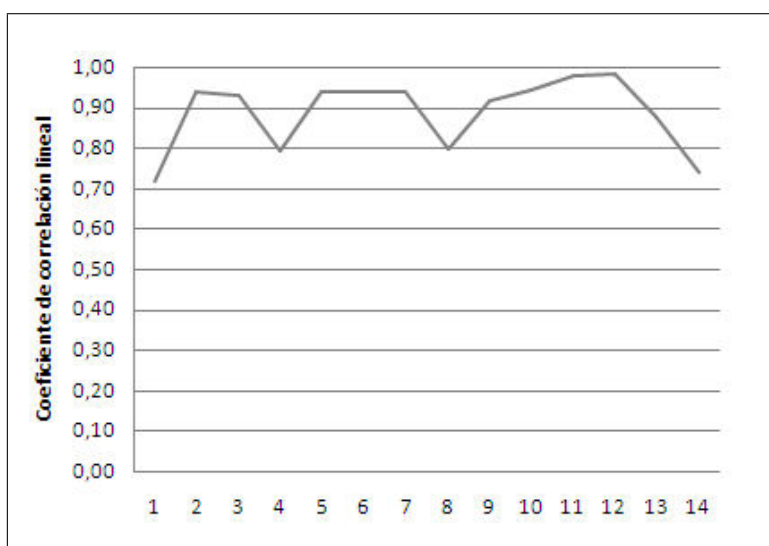
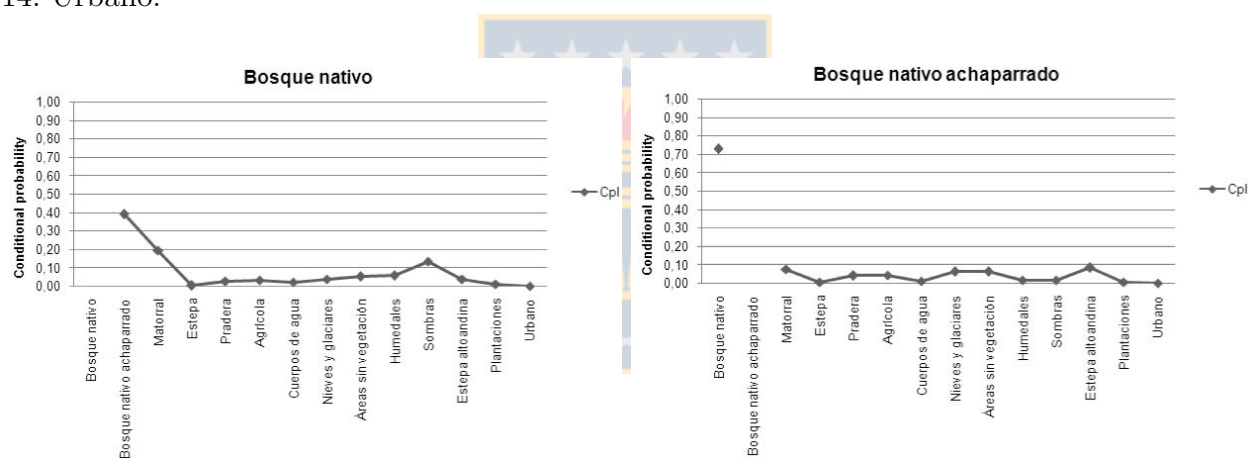


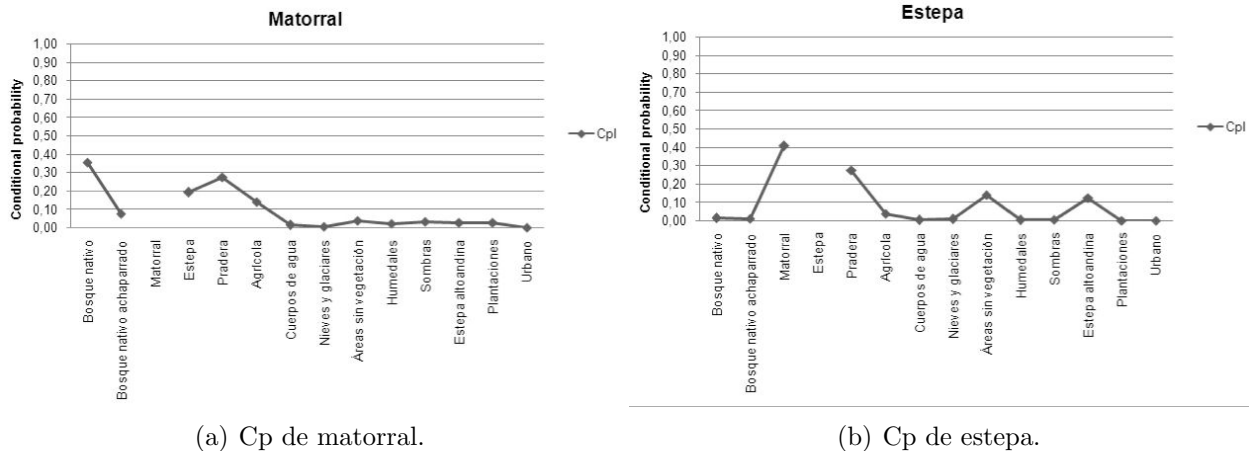
Figura 4.15: Índice de correlación lineal. 1: Bosque nativo. 2: Bosque nativo achaparrado. 3: Matorral. 3: Praderas. 5: Praderas. 6: Agrícola. 7: Cuerpos de agua. 8: Nieves y glaciares. 9: áreas sin vegetación. 10: Humedales. 11: Sombras. 12: Estepa alto andina. 13: Plantaciones. 14: Urbano.



(a) Cpl de Bosque nativo.

(b) Cpl de Bosque nativo achaparrado.

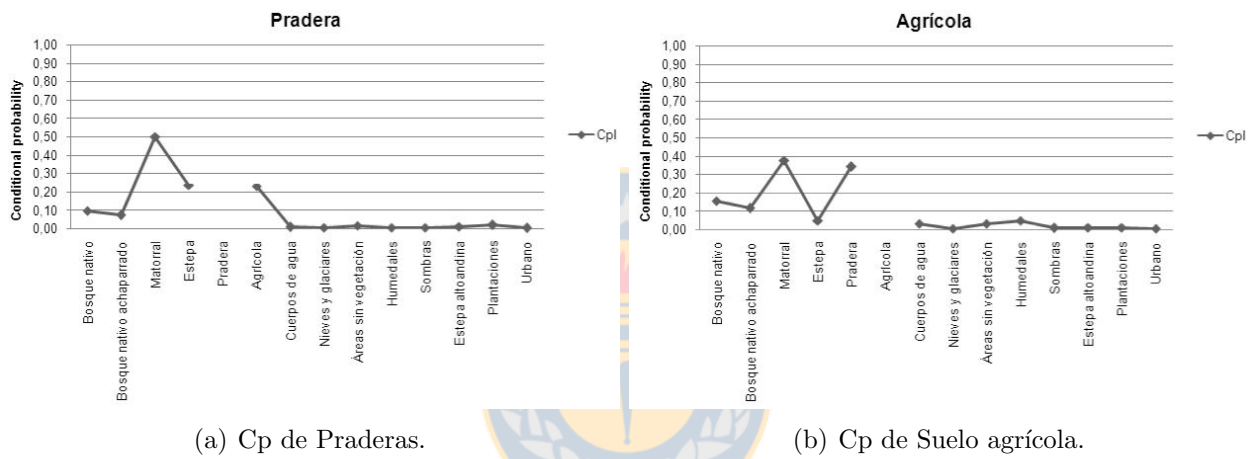
Figura 4.16: Probabilidad condicional para Bosque nativo y Bosque nativo achaparrado



(a) Cp de matorral.

(b) Cp de estepa.

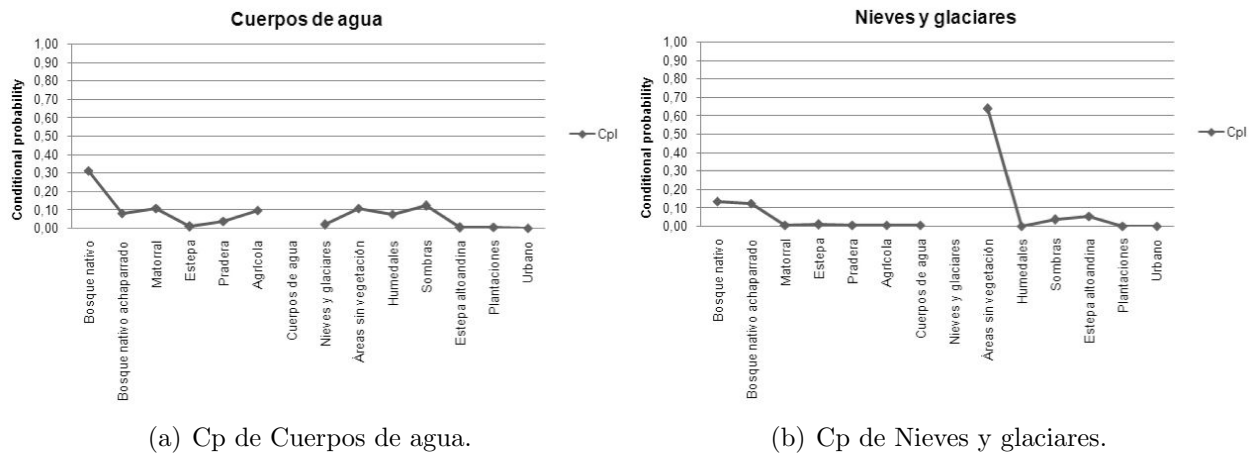
Figura 4.17: Probabilidad condicional para Matorral y Estepa



(a) Cp de Praderas.

(b) Cp de Suelo agrícola.

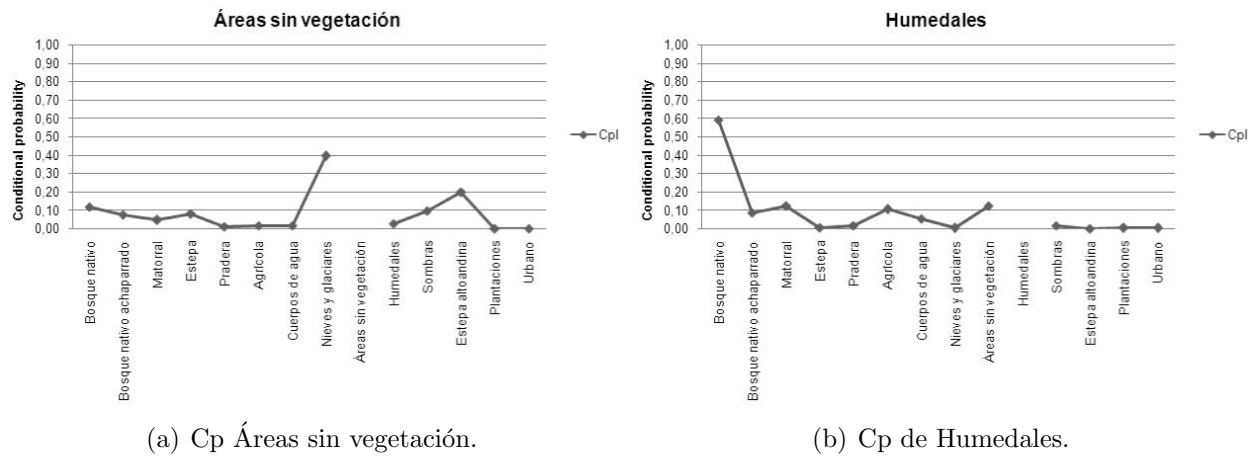
Figura 4.18: Probabilidad condicional para Praderas y Suelo agrícola



(a) Cp de Cuerpos de agua.

(b) Cp de Nieves y glaciares.

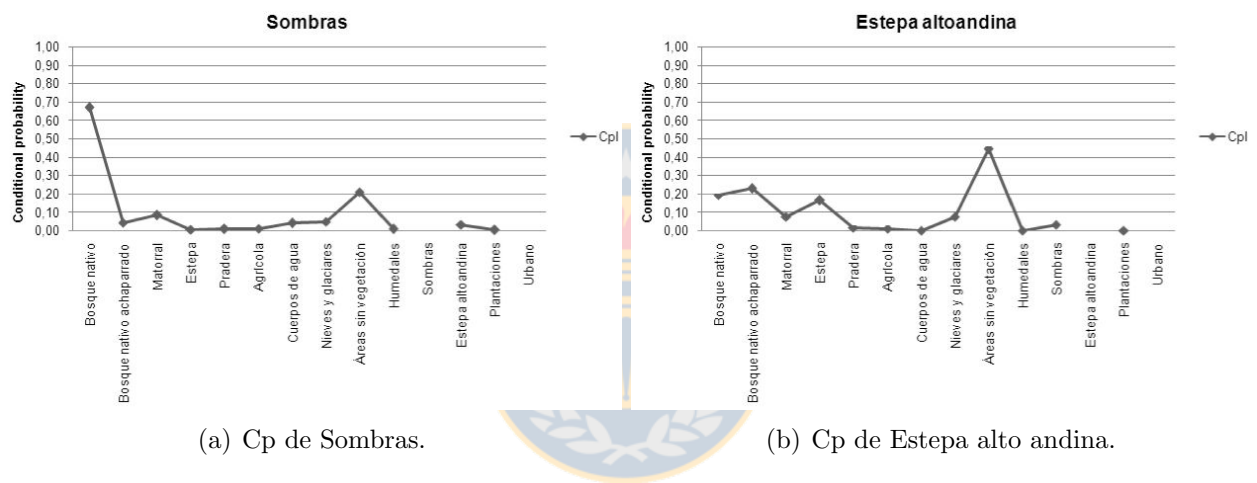
Figura 4.19: Probabilidad condicional para Cuerpos de agua y Nieves y glaciares



(a) Cp Áreas sin vegetación.

(b) Cp de Humedales.

Figura 4.20: Probabilidad condicional para Áreas sin vegetación y Humedales.



(a) Cp de Sombras.

(b) Cp de Estepa alto andina.

Figura 4.21: Probabilidad condicional para Sombras y Estepa altoandina

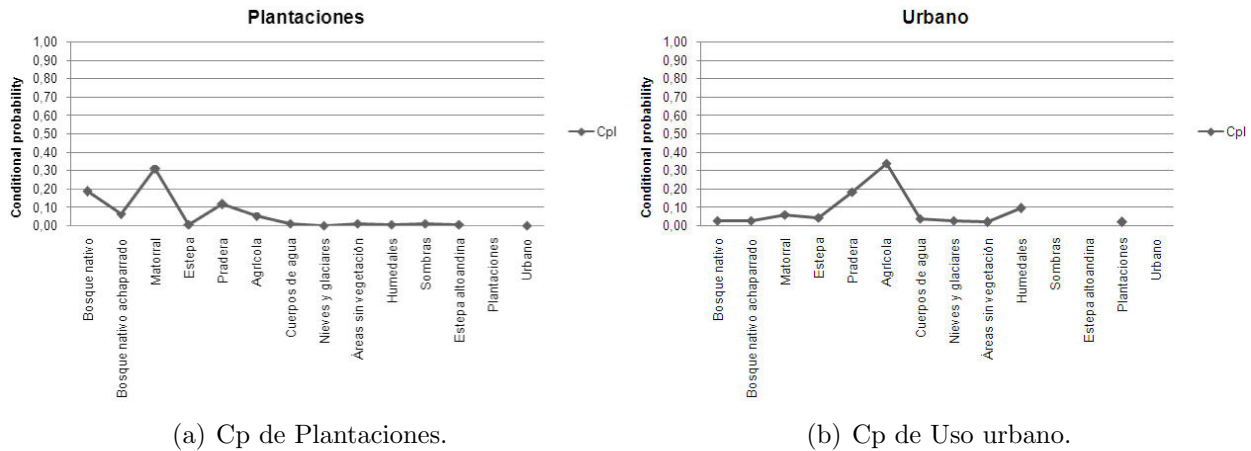


Figura 4.22: Probabilidad condicional para Plantaciones y Suelo Urbano

Luego, y para filtrar este resultado entre  $Pr_2$  y  $Pr_1$ , es que se usa el concepto de *índice de participación*  $pi(c)$ .

#### 4.2.2. Extracción y comparación de reglas

Luego de haber calculado las métricas de validación  $Pr_1$ ,  $Pr_2$  y  $C_p$ , se evalúan los índices de participación  $pi_1$  y  $pi_2$ . En las figuras 6.4 y 6.5 del anexo 6.6 se aprecian los valores de las tablas manteniendo sólo el valor mínimo de  $Pr_1$ , y  $Pr_2$ . Luego, las reglas de colocalidad son ordenadas de mayor a menor según su índice de participación. La tabla 4.2 indica la regla de colocalidad en orden descendente de su índice de participación.

Es importante destacar que los rangos de valores para  $pi_1$  y  $pi_2$  son diferentes, debido a la forma en que son definidos. Mientras  $pi_1$  se basa en un porcentaje de adyacencia entre polígonos,  $pi_2$  considera el polígono completo en la definición de la fuerza de su participación en la regla de colocalidad. Debido a esto, se filtraron las reglas que estuvieran en las primeras 5 posiciones del ordenamiento.

En la tabla 4.3 se indican las reglas que fueron preseleccionadas en función del  $pi_1$  y  $pi_2$  junto con su probabilidad condicional  $C_p$ , y que conforme a la hipótesis, una o más de ellas deberían ser seleccionadas por el experto.

### 4.3. Validación de los resultados

Las reglas obtenidas son presentadas por el experto (E-1) en planificación territorial, quien evalúa el grado de interés de las reglas de colocalidad. Para esto, y a fin de evitar sesgos, se obtienen al azar un sub-conjunto de 30 reglas, del total de 182, aquí se incluyen reglas con valores altos y bajos de las métricas. La tabla y las evaluaciones, tanto de las métricas como del experto, se pueden ver en la tabla 4.4. Los valores posibles para evaluar cada regla son:

1. Nulo interés.

Tabla 4.2: Índices de participación, por regla y ranking conforme a Pi.

Ranking - índices	Pi 1	Pi 2	rank Pi 1	rank Pi 2
Bosque nativo y bosque nativo achaparrado	0,29	0,93	1	1
Matorral y bosque nativo	0,24	0,81	2	2
Matorral y pradera	0,22	0,78	3	5
Áreas sin vegetación y estepa altoandina	0,20	0,67	4	10
Nieves y glaciares y áreas sin vegetación	0,20	0,80	5	4
Pradera y agrícola	0,18	0,71	6	7
Estepa y pradera	0,14	0,42	7	22
Matorral y agrícola	0,14	0,67	8	11
Matorral y estepa	0,11	0,51	9	17
Estepa y estepa altoandina	0,10	0,43	10	21
Bosque nativo achaparrado y estepa altoandina	0,10	0,44	11	20
Bosque nativo achaparrado y matorral	0,10	0,45	12	19
Sombras y bosque nativo	0,11	0,80	13	3
Sombras y Áreas sin vegetación	0,09	0,52	14	29
Estepa y áreas sin vegetación	0,08	0,30	15	16
Áreas sin vegetación y sombras	0,08	0,52	16	18
Bosque nativo achaparrado y áreas sin vegetación	0,08	0,48	17	27

Tabla 4.3: Seis reglas ordenadas conforme al Pi y con el valor de la Cp.

Ranking - índices	Pi 1	Pi 2	rank Pi 1	rank Pi 2	Cp
Bosque nativo achaparrado y Bosque nativo	0,29	0,93	1	1	0,39
Bosque nativo y Matorral	0,24	0,81	2	3	0,35
Matorral y Pradera	0,22	0,78	3	5	0,27
Nieves y glaciares y Áreas sin vegetación	0,20	0,80	4	4	0,40
Áreas sin vegetación y Estepa altoandina	0,20	0,67	5	11	0,20
Sombras y Bosque nativo	0,10	0,90	9	2	0,13

2. Poco interés.

3. Interesante.

4. Muy interesante.



Tabla 4.4: Listado de reglas presentadas al experto, valores de cada métricas y la valoración del experto (4= Muy interesante; 3= Interesante; 2=Poco interesante; 1= Interés nulo).

id	Regla de colocalidad	Pi 1	Pi 2	Exp.
1	Estepa alto andina <i>y</i> plantaciones	0,00	0,00	3
2	Humedales <i>y</i> estepa alto andina	0,00	0,00	2
3	Praderas <i>y</i> cuerpos de agua	0,01	0,09	1
4	Bosque nativo <i>y</i> áreas sin vegetación	0,07	0,76	1
5	Estepa <i>y</i> plantaciones	0,00	0,00	4
6	Praderas <i>y</i> sombras	0,01	06	2
7	Plantaciones <i>y</i> suelo urbano	0,00	0,02	1
8	Bosque nativo <i>y</i> bosque nativo achaparrado	0,29	0,93	3
9	Suelo agrícola <i>y</i> suelo urbano	0,00	0,06	4
10	Áreas sin vegetación <i>y</i> sombras	0,08	0,52	2
11	Humedales <i>y</i> plantaciones	0,00	0,03	3
12	Cuerpos de agua <i>y</i> suelo urbano	0,00	0,09	1
13	Plantaciones <i>y</i> matorral	0,00	0,219	2
14	Bosque nativo <i>y</i> matorral	0,24	0,81	3
15	Estepa <i>y</i> Nieves y glaciares	0,01	0,03	3
16	Nieves y glaciares <i>y</i> suelo urbano	0,00	0,00	1
17	Sombras <i>y</i> plantaciones	0,00	0,03	2
18	Cuerpos de agua <i>y</i> estepa alto andina	0,00	0,01	2
19	Matorral <i>y</i> praderas	0,22	0,78	4
20	Estepa alto andina <i>y</i> bosque nativo	0,07	0,70	3
21	Áreas sin vegetación <i>y</i> estepa alto andina	0,20	0,67	2
22	Matorral <i>y</i> plantaciones	0,02	0,22	4
23	Matorral <i>y</i> nieves y glaciares	0,01	0,03	2
24	Bosque nativo <i>y</i> Suelo agrícola	0,06	0,69	4
25	Sombras <i>y</i> suelo urbano	0,00	—	1
26	Áreas sin vegetación <i>y</i> nieves y glaciares	0,20	0,80	3
27	Bosque nativo <i>y</i> cuerpos de agua	0,02	0,60	1
28	Estepa alto andina <i>y</i> suelo urbano	0,00	—	1
29	Pradera <i>y</i> plantaciones	0,02	0,14	4
30	Bosque nativo <i>y</i> sombras	0,10	0,90	3

La tabla 4.4 muestra, por ejemplo, que tanto el  $Pi_1$  como el  $Pi_2$  para la regla “estepa  $\rightarrow$  plantaciones” obtiene un valor de “nulo interés”, mientras que para la misma regla, el experto la evalúa como “muy interesante”. Esto se debe atribuir, a que las métricas aplicadas en esta tesis, usan sólo reglas basadas en cálculos de relaciones topográficas y en el caso de  $Pi_1$  se considera un refinamiento métrico, pero no consideran el valor del interés o novedad en cuanto a la asociación del uso de suelo, que en el caso del experto, si se considera por su experiencia.

En este sentido, las métricas utilizadas parecen evaluar mejor (valores más altos) las reglas que podrían ser más interesantes. Puesto que no hay una absoluta noción del grado de interés en las asociaciones producidas por las métricas, se aplica un análisis de correlación.

Se aplica el coeficiente de correlación de Spearman <sup>3</sup>. Para ello, el coeficiente de correlación de Spearman fue calculado para todas las evaluaciones, como se muestra en la tabla 4.5. En general, se observó una correlación lineal positiva, que indica que los valores están relacionados entre sí. Sin embargo, esta relación es débil.

Al analizar ambas métricas, se observa que  $Pi_1$ , tiene un coeficiente de correlación mayor que  $Pi_2$ . Esto se verifica al aplicar el "test *T student*", el cual presenta diferencias significativas entre  $Pi_1$  ( $p < 0,01$  ;  $t = 3,823$ ) y la evaluación del experto. Por otra parte, para  $Pi_2$ , no se registraron diferencias significativas ( $p > 0,01$  ;  $t = 1,811$ ) con respecto a la evaluación del experto, lo cual indica una baja correlación entre ambas mediciones.

Tabla 4.5: Correlación entre el experto y las métricas evaluadas.

	métricas	
Experto	$Pi_1$	$Pi_2$
E-1	0,188	0,103

La correlación positiva entre  $Pi_1$  y el experto, puede deberse a que en los procesos efectuados durante el minado de datos, se considera un refinamiento métrico. Es decir, se evaluaron las diferencias dadas por el porcentaje de adyacencia, superficie y perímetro.

Finalmente las reglas validadas por el experto son:

1. Bosque nativo achaparrado y Bosque nativo (0.29, 0.39).
2. Áreas sin vegetación y Nieves y glaciares (0.20, 0.40).

Posteriormente, se generó el gráfico de dispersión, para analizar la agrupación de las reglas obtenidas cercanas a la coordenada 1,1 (figura: 4.23). El resultado esperado coincide con las reglas validadas por el experto.

Las figuras 4.24 y 4.25 muestra la vecindad de usos de suelo obtenidos como reglas de colocalidad.

En términos generales, los resultados sugieren que el enfoque utilizado es efectivo para evaluar el grado de asociación de patrones simples (vecindad de polígonos con  $k=2$ ) descubiertos, en el conjunto de datos.

---

<sup>3</sup>medida de asociación lineal que utiliza los rangos de cada conjunto de datos y compara dichos rangos

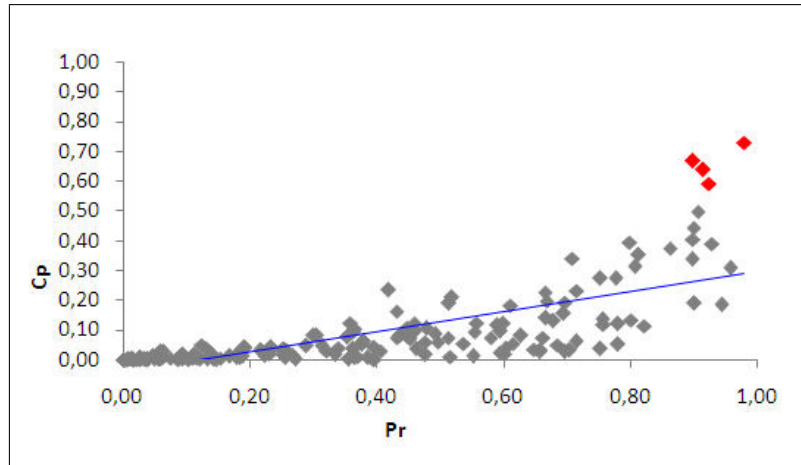


Figura 4.23: Dispersión de las reglas de colocalidad obtenidas.

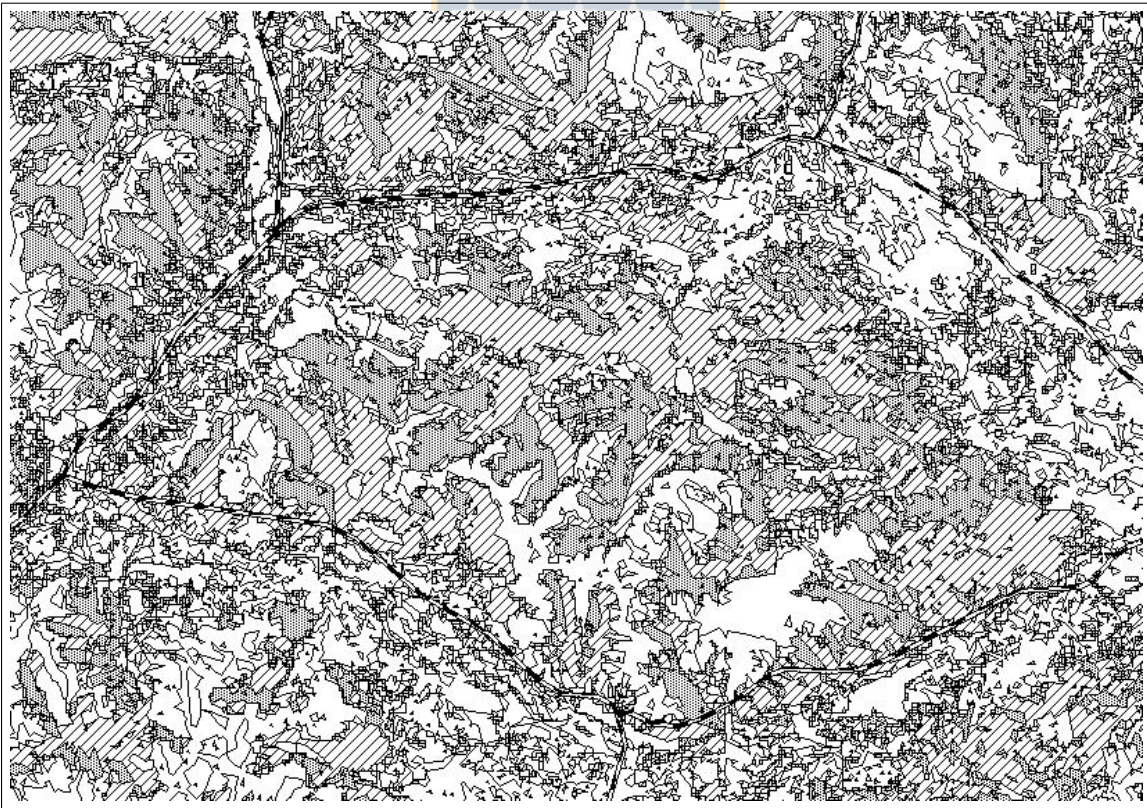


Figura 4.24: Reglas de colocalidad para uso de suelo "1" (bosque nativo): achurado diagonal, y el uso de suelo "2" (bosque nativo achaparrado): achurado con puntos.

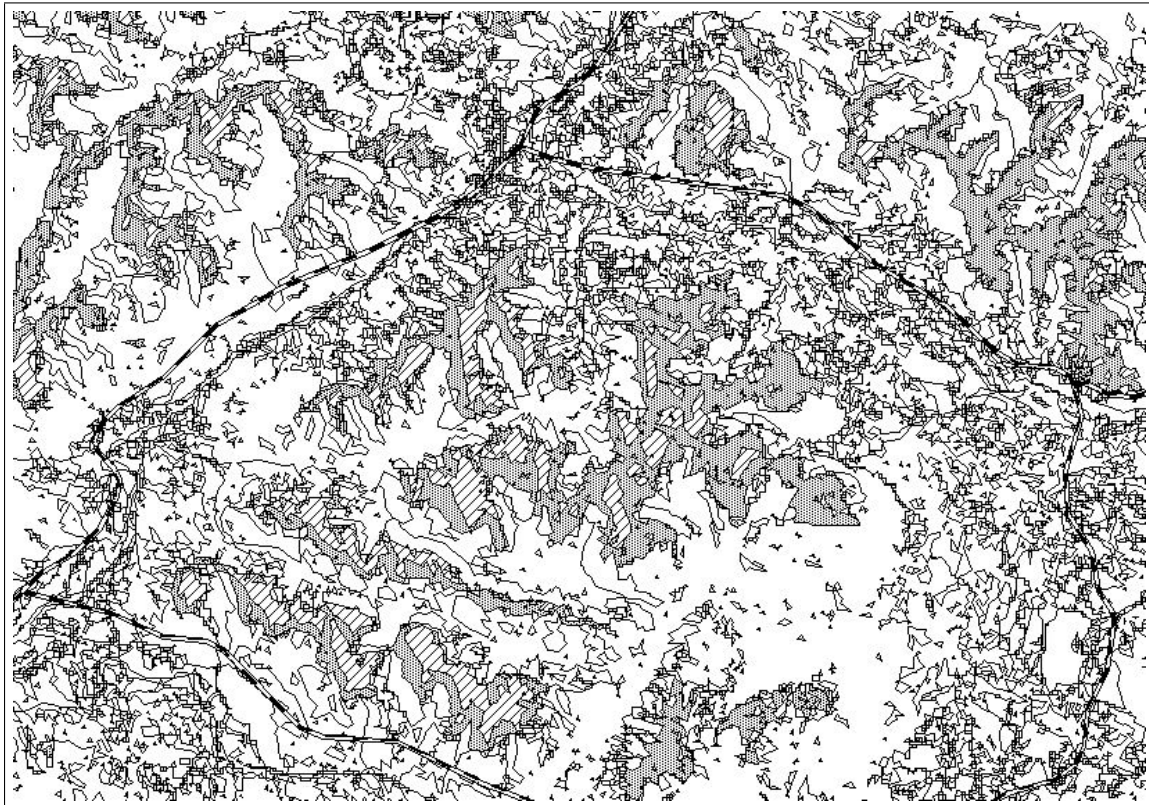


Figura 4.25: Reglas de colocalidad para uso de suelo "9" (áreas sin vegetación): achurado diagonal, y el uso de suelo "8" (nieves y glaciares): achurado con puntos.





# Capítulo 5

## Conclusiones y trabajo futuro

Después de realizar los experimentos propuestos y definidos en esta tesis, se puede decir que los resultados correspondieron a lo esperado y planteado en las Hipótesis de esta tesis.

### 5.1. Conclusiones

Los resultados más relevantes obedecen en primera instancia al haber logrado implementar un método de selección de variables, que si bien sus resultados difieren en las variables de “Distancia y densidad de caminos”, “proporción(densidad) de praderas y matorrales”, y “elevación”, son reconocidas por el experto como variables que en la realidad si son significativas para el estudio.

Los resultados demuestran que:

1. Los métodos de selección de variables por minería de datos espacial, permiten seleccionar variables reales del entorno del área de estudio que presentan mayor asociación con un criterio específico.
2. Existen asociaciones espaciales entre las variables ambientales de un estudio de cambio de uso de suelo, identificables por medio de minado de datos espacial que vienen a ser un real aporte a estudios de la dinámica del uso de suelo, y que se pueden representar como reglas de asociación.
3. Existen 92 reglas de asociación, que describen la dinámica del uso de suelo. Estas reglas se analizaron desde la perspectiva de dos criterios: cambio o no cambio, y desde el criterio del tipo de cambio presentado.
4. Las variables con mayor fuerza descriptiva en la dinámica del uso de suelo incluyen las distancias a la red vial, así como las pequeñas áreas cercanas a áreas que ya presentaban un uso de suelo del tipo plantación.
5. Existen asociaciones entre los distintos tipos de uso de suelo, representables como reglas de colocalidad, y que se pueden generar a través de procesos de minería de datos espacial.

6. Se evaluaron un total de 182 reglas de colocalidad. De las que fueron validadas por el experto sólo dos.
7. Sobre  $P_{i_1}$  y  $P_{i_2}$ , podemos concluir que ambas métricas son medidas de prevalencia en las reglas de colocalidad, validado por el experto al seleccionar las reglas con mayor  $P_i$ .
8. Existe una correlación positiva y fuerte entre las métricas  $P_{i_1}$  y  $P_{i_2}$ .
9. En general, se observó una buena correlación entre la evaluación del experto y  $P_{i_1}$ .
10. Finalmente, los resultados reflejan la diferencia de considerar un refinamiento métrico al evaluar  $P_{i_1}$  y  $P_{i_2}$ , obteniendo para la primera una mejor correlación que la segunda.

## 5.2. Trabajo futuro

En este trabajo ha quedado por explorar en mayor detalle patrones de cambio de uso de suelo con series de tiempo que abarquen más de dos períodos, así como también, extender el estudio en otros sectores de la Patagonia chilena.

Además, se plantea la posibilidad de extender el estudio a evaluar reglas de colocalidad con más de 2 vecindades ( $k \geq 3$ ).





# Bibliografía

- [1] Mauricio Aguayo, T. Wiegand, Gerardo Azocar, K. Wiegand, and Claudia Vega. Revealing the driving forces of mid-cities urban growth patterns using spatial modeling: a case study of los ángeles. *Ecology and Society*, 2007.
- [2] Vania Bogorny. *Enhancing spatial association rule mining in geographic databases*. PhD thesis, Universidade Federal do Rio grande do Soul, 2006.
- [3] Vania Bogorny and Andrey Tietbohl Palma. Extending the weka data mining toolkit to support geographic data preprocessing. Technical Report 354, Univerisdade Federal do Rio Grande do Soul, 2006.
- [4] Vania Bogorny, Andrey Tietbohl Palma, Paulo Engel, and Luis Otavio Alvares. Weka-gdpm: Integrating classical data mining toolkit to geographic information systems. In *WAAMD*, pages 9–16. SBC, 2006.
- [5] You-Shyang Chen and Ching-Hsue Cheng. Forecasting revenue growth rate using fundamental analysis: A feature selection based rough sets approach. In *FSKD (3)*, pages 151–155, 2007.
- [6] Ching-Hsue Cheng and You-Shyang Chen. Extracting rules of initial returns using attribute selection and entropy-based rough sets in electronic firm. In *FSKD (3)*, pages 146–150, 2007.
- [7] Sophie Cockcroft. A taxonomy of spatial data integrity constraints. *GeoInformatica*, 1(4):327–343, 1997.
- [8] Conferencia del CONFIBSIG. *Aplicación de SIG y data-mining para establecer perfiles ecológicos de los humedales en Argentina*, 2007.
- [9] R. S. Defries, L. Bounoua, and G. J. Collatz. Human modification of the landscape and surface climate in the next fifty years. *Global Change Biology*, 8:438–458, May 2002.
- [10] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, pages 1–34. A.I. magazine, 1996.
- [11] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus. Knowledge discovery in databases: An overview. In *Knowledge Discovery in Databases*, pages 1–30. AAAI/MIT Press, 1991.

- [12] Terrance Goan. From data to actionable knowledge: Applying data mining to the problem of intrusion detection, 2000.
- [13] K. K. Goldewijk. Estimating global land use change over the past 300 years: The hyde database. In *Global Biogeochemical Cycles*, volume 15, pages 417–433, 2001.
- [14] Ralf Hartmut Güting. An introduction to spatial database systems. *VLDB J.*, 3(4):357–399, 1994.
- [15] Yan Huang, Shashi Shekhar, and Hui Xiong. Discovering colocation patterns from spatial data sets: A general approach. *IEEE Trans. Knowl. Data Eng.*, 16(12):1472–1485, 2004.
- [16] International Conference on Bioinformatics and computational Biology (BIOCOMP 07). *A comparative study of different machine learning methods on microarray gene expression data*, 2008.
- [17] Survey Paper Krzysztof, Krzysztof Koperski, Junas Adhikary, and Jiawei Han. Spatial data mining: Progress and challenges. In *SIGMOD Workshop on Research Issues on data Mining and Knowledge Discovery (DMKD)*, pages 1–10, 1996.
- [18] Jeremy L. Mennis and Jun Wei Liu. Mining association rules in spatio-temporal data: An analysis of urban socioeconomic and land cover change. *T. GIS*, 9(1):5–17, 2005.
- [19] Navin Ramankutty and Jonathan A. Foley. Estimating historical changes in global land cover: Croplands from 1700 to 1992. In *Global Biogeochemical Cycles*, volume 13, pages 997–1027, 1999.
- [20] B.D. Ripley. Modeling spatial patterns. In *J. Roy Statist. Soc*, volume 39, pages 172–192, 1977.
- [21] Osvaldo E. Sala, F. Stuart III Chapin, Juan J. Armesto, Eric Berlow, Janine Bloomfield, Rodolfo Dirzo, Elisabeth Huber-Sanwald, Laura F. Huenneke, Robert B. Jackson, Ann Kinzig, Rik Leemans, David M. Lodge, Harold A. Mooney, Martijn Oosterheld, N. LeRoy Poff, Martin T. Sykes, Brian H. Walker, Marilyn Walker, and Diana H. Wall. Global biodiversity scenarios for the year 2100. *Science*, 287(5459):1770–1774, 2000.
- [22] Shashi Shekhar and Sanjay Chawla. *Spatial Databases: A tour*. Prentice Hall, 2003.
- [23] Pusheng Zhang, Michael Steinbach, Vipin Kumar, Shashi Shekhar, Pang-Ning Tan, Steven Klooster, and Christopher Potter. Discovery of patterns in earth science data using data mining. *New generations of data-mining applications*, pages 1–24, 2000.

# Capítulo 6

## ANEXOS

### 6.1. Glosario

- **Algoritmo** Secuencia explícita y finita de operaciones que conduce a la solución problema. Si esto lo aplicamos al ambiente GIS (SIG), suele tratarse de un conjunto de operaciones de álgebra de mapas y/o sobre bases de datos espaciales que permiten obtener un resultado mediante la combinación de información espacial y alfanumérica.
- **Atributo** Propiedad o característica de una clase de elementos en una base de datos espacial, por ejemplo, la superficie (tamaño propiedad).
- **Binario** Sistema de numeración basado en dos dígitos: 0 y 1.
- **Bit** Es la menor unidad de información en un sistema informático. Bit proviene de la contracción de *binary digit*. Puede tener solo dos valores: 0 o 1.
- **Byte** Conjunto de 8 bits en el sistema binario. Pueda almacenar un número entero entre 0(00000000) y 256(11111111).
- **Cartografía** Conjunto de técnicas utilizadas para la construcción de mapas.
- **Celda** Elemento básico de información en una estructura *raster* matricial. Representa el valor medio de un área rectangular superpuesta al terreno (podría decirse que es un concepto analógico de *pixel* en una imagen digital).
- **Centroide** Punto interior a un polígono más próximo a su centro geométrico. El centro geométrico de un polígono puede ser exterior de el polígono si es convexo, en ese caso, el centroide se mueve al lugar más próximo posible que cumpla la condición de interioridad.
- **Clasificación** Proceso de agrupamiento de un conjunto de elementos en clases. En el sentido estadístico, una clasificación pretende agrupar los elementos en clases internamente homogéneas pero diferenciables entre ellas por valores de una o varias variables.
- **Colocation** Colocación o colocalidad, un subset de características espaciales booleanas.

- **Coordenada** Cantidad usada para definir una posición en un sistema de referencia, las coordenadas pueden ser lineales (cartesianas) o angulares ( esféricas), según el sistema de referencia.
- **Cuenca hidrológica** Conjunto de puntos del terreno cuyas líneas de flujo convergen en un sumidero, el cual suele hacerse coincidir con un punto singular: una desembocadura o confluencia de ríos.
- **Dato** Hecho verificable de la realidad. Un dato puede ser una medida, una ecuación o cualquier tipo de información que se pueda verificar.
- **Datum** Sistema geométrico de referencia empleado para expresar numéricamente la posición geodésica de un punto sobre el terreno. Cada *datum* se define en función de un elipsoide y por un punto en el que el elipsoide y la Tierra son tangentes.
- **Dominio** En una base de datos espacial se aplica al conjunto de valores posibles de un atributo. Por ejemplo, el conjunto de valores posibles para los usos de suelo de esta tesis, el dominio es "uso\_suelo".
- **Error** Diferencia entre el valor medio medido o estimado y el valor real en un modelo, el error representa la desviación entre lo predicho por el modelo y la realidad, es una estimación de la calidad de la información de un mapa y suele distinguirse del concepto de *precisión*, que hace referencia a la calidad del método de medida utilizado.
- **Estándar** Propiedad que garantiza la uniformidad en los métodos de capturar, representar, almacenar y documentar la información.
- **Georreferenciar** Asignar coordenadas geográficas a un objeto o estructura, el concepto aplicado a una imagen digital implica un conjunto de operaciones geométricas que permiten asignar a cada *pixel* de la imagen un par de coordenadas (x, y) en un determinado sistema de proyección.
- **GPS** Abreviación para *Global Positioning System (Sistema de posicionamiento global)*. Es un sistema satelital mediante el cual es posible estimar las coordenadas geográficas en el momento de un punto en la Tierra. El conjunto de satélites para este efecto reciben el nombre de *constelación GPS*.
- **Imagen digital** Representación gráfica de un objeto mediante una matriz regular que recoge valores de reflectancia, los valores de reflectancia suelen medirse mediante sensores sensibles a ciertos rangos de longitudes de onda de la luz.
- **Incertidumbre** Falta de certeza en un resultado producido por el error en los datos y en los procesos. La incertidumbre va asociada a un valor de probabilidad de que la medida sea o no correcta.
- **Leyenda** Listado ordenado y estructurado de las relaciones símbolo y valores para los datos representados en un mapa geográfico. Su función es poder interpretar los significados gráficos usados, lo que abarca tanto las variables cuantitativas como las cualitativas.

- **Línea** Conjunto ordenado de vectores encadenados, en los archivos del tipo vectorial, las líneas son utilizadas para representar objetos geográficos como carreteras, caminos, tendidos eléctricos, etc. En una estructura topológica, las líneas tienen un sentido y están definidos los lados izquierdo y derecho.
- **Mapa** Modelo geográfico de la superficie terrestre donde se representan objetos espaciales y sus propiedades métricas, topológicas y sus atributos.
- **Metadatos** Información sobre las características de un conjunto de datos, generalmente es información anexa al cuerpo de datos principal. Si hablásemos de un modelo digital de elevaciones, sus metadatos serían sobre extensión geográfica, estadísticas, autoría, calidad de la información, metodología, etc.
- **Modelo** Representación simplificada de un objeto o proceso en la que se representan alguna de sus propiedades. Un modelo no reproduce todas las propiedades del objeto o sistema original. Generalmente los modelos se construyen para conocer las propiedades del objeto real.
- **Modelo de datos** Esquema conceptual que se usa para representar la realidad mediante un modelo. Un modelo de datos intenta en parte solucionar la problemática de como pasar la realidad a un modelo, es decir como representar la realidad adecuada y eficientemente. Un mismo modelo de datos puede expresarse luego en distintas estructuras de datos. Por ejemplo, las siglas GIF, JPG, BMP definen estructuras de datos distintas pero todas ellas se enmarcan en un modelo de datos *raster*, las estructuras de datos pueden diferir en la forma en que codifican y almacenan la información incluso en el mismo esquema conceptual.
- **Modelo digital del terreno** Estructura numérica de datos que representa la distribución espacial de una variable cuantitativa. Es un modelo que representa una propiedad cuantitativa topográfica (*elevación, pendiente*) o no (*temperatura de la superficie del terreno, reflectancia*).
- **Nodo** Vértice inicial o final de una línea. A modo de extensión del término, se aplica a las entidades puntuales que están interconectadas en una *estructura de red*.
- **Ortofoto** Fotografía aérea modificada geoméricamente para ajustarla a un sistema de proyección geográfica. Esta modificación elimina las distorsiones producto de la perspectiva, movimiento de la cámara y el relieve.
- **Pendiente** Ángulo entre la línea normal a la superficie del terreno y la vertical.
- **Proyección** Transformaciones métricas definidas para representar la superficie de la Tierra sobre un plano. Hoy en día hay muchas proyecciones, cada cual posee diferentes propiedades en cuanto a las métricas del objeto real y de su representación plana.
- **Raster** Modelo de datos en el que la realidad se representa mediante teselas elementales que forman un mosaico regular. Cada tesela del mosaico es una unidad de superficie que recoge el valor medio de la variable representada, las teselas pueden ser cuadradas

(celdas) o no (triangulares, hexagonales). Un modelo de datos *raster* esta basado en localizaciones.

- **Sistema de Gestión de Bases de Datos (SGBD)** Sistema informático diseñado para la creación, mantención, modificación, actualización y consulta de bases de datos.
- **Sistema de Información Geográfica (SIG o GIS en inglés)** Sistemas informáticos con herramientas específicas para el manejo de información espacial y sus propiedades. Estos tipos de propiedades son tres: *métricas, topológicas y atributivas*.
- **Sistema de coordenadas** Marco de referencia espacial que permite la definición de localizaciones mediante coordenadas. Estas coordenadas pueden ser lineales (sistemas cartesianos, con ejes ortogonales) o esféricas (donde se usan como coordenadas el acimut y elevación angular).
- **Sistema experto** Conjunto de Bases de Datos y Sistemas de Decisión basado en reglas que genera la respuesta de mayor probabilidad ante un conjunto dado de datos de entrada. Por ejemplo, en un caso de un conjunto de datos de usos de suelo, el sistema genera un resultado con el vecino de uso de suelo más probable.
- **SQL** Acrónimo de *Structured query language, o sistema estructurado de consulta*, un lenguaje estándar para la gestión de bases de datos.
- **Tabla de correspondencia** Tabla donde se establece la relación de una correspondencia entre rangos discretos de una variable continua. Por ejemplo: pendiente 0 → nula; pendiente 1-5 → baja; pendiente 5-10 → media; pendiente 10 y mayor → alta.
- **Tabla de instancias** Tabla que contiene toda la colección de todas las combinaciones de reglas, podría decirse que es como una interpretación de *joins* espaciales.
- **teledetección** Proceso de captura de información sin contacto entre el instrumento de medida y el objeto. Generalmente se aplica en el sentido a las imágenes y datos captados mediante sensores transportados pro aviones o satélites.
- **Tesela** Cada unidad elemental del modelo de datos *raster*. Puede considerarse sinónimo de celda, aunque generalmente la denominación de celda suele usarse en teselas rectangulares.
- **Topografía** Descripción de las formas del terreno.
- **Transformación** Proceso de conversión de coordenadas desde un sistema cartesiano a otro. Habitualmente, la digitalización de un mapa conlleva esta transformación.
- **Vector** Entidad geométrica definida por una magnitud y un sentido. Esta formado por un par de puntos ordenados, el orden define el sentido del vector y la distancia entre punto de origen y punto final la magnitud, de la mano de esto, podemos decir que si la magnitud es nula, el vector es un punto y su sentido queda indefinido.

- **Vectorial** Modelo de datos en el que la realidad se representa mediante vectores o estructuras de vectores. Una estructura vectorial puede ser compleja, por ejemplo: una cadena de vectores forma un arco, una cadena de arcos forma un anillo, uno o varios anillos definen un polígono, se trata de un modelo de datos basado en objetos geométricos frente al modelo *raster*, el que se basa en localizaciones.
- **Vectorizar** Transformación de una estructura raster a una vectorial.





## 6.2. Abreviaciones de las variables usadas en esta tesis

Tabla 6.1: Variables presentes en el estudio y su sigla.

Sigla	Variable
DIS_URB	Distancia a centros urbanos
DIS_PLANT	Distancia a plantaciones 1984
DIS_HIDRO	Distancia a red hídrica
DIS_CAMT	Distancia a red vial
DIS_CAM2	Distancia a caminos ripiados
DIS_CAM1	Distancia a caminos asfaltados
DIS_AGRIC	Distancia a terrenos de aptitud agrícola
DEN_RIOS50	Densidad de ríos radio 4500 m.
DEN_RIOS25	Densidad de ríos radio 2250 m.
DEN_RIOS10	Densidad de ríos radio 900 m.
DEN_RIOS05	Densidad de ríos radio 450 m.
DEN_CAMT50	Densidad de caminos radio 4500 m.
DEN_CAMT25	Densidad de caminos radio 2250 m.
DEN_CAMT10	Densidad de caminos radio 900 m.
DEN_CAMT05	Densidad de caminos radio 450 m.
M_TMIN	Temperaturas mínimas
M_TMED	Temperaturas medias
M_TMAX	Temperaturas máximas
M_SLOPE	Pendiente del terreno
M_PPANUAL	Precipitación anual
M_DEM	Elevación
M_CPUSO	Capacidad usos del suelo
PROP_SUP	Tamaño propiedad
U_MT50	Densidad de matorrales radio 4500 m.
U_MT25	Densidad de matorrales radio 2250 m.
U_MT10	Densidad de matorrales radio 900 m.
U_MT05	Densidad de matorrales radio 450 m.
U_ES50	Densidad de estepa radio 4500 m.
U_ES25	Densidad de estepa radio 2250 m.
U_ES10	Densidad de estepa radio 900 m.
U_ES05	Densidad de estepa radio 450 m.
U_BN50	Densidad de bosque nativo radio 4500 m.
U_BN25	Densidad de bosque nativo radio 2250 m.
U_BN10	Densidad de bosque nativo radio 900 m.
U_BN05	Densidad de bosque nativo radio 450 m.
U_BA50	Densidad de bosque nativo achaparrado radio 4500 m.
U_BA25	Densidad de bosque nativo achaparrado radio 2250 m.
U_BA10	Densidad de bosque nativo achaparrado radio 900 m.
U_AG05	Densidad de bosque nativo achaparrado radio 450 m.
U_AG50	Densidad de agrícola radio 4500 m.
U_AG25	Densidad de agrícola radio 2250 m.
U_AG10	Densidad de agrícola radio 900 m.
U_AG05	Densidad de agrícola radio 450 m.
U_PR50	Densidad de pradera radio 4500 m.
U_PR25	Densidad de pradera radio 2250 m.
U_PR10	Densidad de pradera radio 900 m.
U_PR05	Densidad de pradera radio 450 m.
U_PL50	Densidad de plantaciones radio 4500 m.
U_PL25	Densidad de plantaciones radio 2250 m.
U_PL10	Densidad de plantaciones radio 900 m.
U_PL05	Densidad de plantaciones radio 450 m.
USO_84	Uso de suelo en 1984
USO_01	Uso de suelo en 2000

### 6.3. Lista de las Reglas de Asociación obtenidas

- Conforme a la perspectiva de cambio/no-cambio.
  - Modelo 2, desde donde se generaron las reglas de asociación usando el asociador Predictive Apriori y 15 variables:
    - Tamaño propiedad=2002,0985 15  $\Rightarrow$  Uso de suelo 2001=13(Plantación) 15 acc:(0.94107)
    - Densidad de bosque nativo radio 4500m=0,0000 14  $\Rightarrow$  Densidad de bosque nativo radio de 2250m=0,0000 14 acc:(0.9374)
    - Tamaño propiedad =4854,2798 12  $\Rightarrow$  Uso 2001=13(Plantación) 12 acc:(0.92849)
    - Distancia red vial=180,0000 9  $\Rightarrow$  Distancia a caminos ripiados=180,0000 9 acc:(0.90903)
    - Tamaño propiedad=762,4233 7  $\Rightarrow$  Uso de suelo 2001=13,0000 (Plantación) 7 acc:(0.88884)
    - Distancia a caminos ripiados=180,0000 Uso de suelo 2001=13,0000(Plantación) 7  $\Rightarrow$  Distancia red vial=180,0000 7 acc:(0.88884)
    - Precipitación anual=4000,0000 Tamaño propiedad=124931,2578 7  $\Rightarrow$  Densidad de matorrales radio 4500 m=0,0000 7 acc:(0.88884)
    - Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 4500m=0,0000 7  $\Rightarrow$  Densidad bosque nativo radio 2250 m=0,0000 Uso suelo 2001=4,0000(Estepa) 7 acc:(0.88884)
    - Precipitación anual=4000,0000 15  $\Rightarrow$  Densidad matorrales radio 2500 m=0,0000 14 acc:(0.88247)
    - Distancia red vial=900,0000 6  $\Rightarrow$  Distancia caminos ripiados=900,0000 6 acc:(0.87496)
    - Distancia caminos ripiados=90,0000 6  $\Rightarrow$  Distancia red vial=90,0000 6 acc:(0.87496)
    - Tamaño propiedad=620,5640 6  $\Rightarrow$  Uso suelo 2001=13,0000(Plantación) 6 acc:(0.87496)
    - Tamaño propiedad=496,4666 6  $\Rightarrow$  Uso suelo 2001=13,0000(Plantación) 6 acc:(0.87496)
    - Precipitación anual=4000,0000 Tamaño propiedad=73718,1953 6  $\Rightarrow$  Uso suelo 2001=8,0000(Nieves y glaciares) 6 acc:(0.87496)
    - Tamaño propiedad=49620,9805 Densidad bosque nativo achaparrado radio 4500 m=0,0000 6  $\Rightarrow$  Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
    - Tamaño propiedad=49620,9805 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\Rightarrow$  Densidad de bosque nativo radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)
    - Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\Rightarrow$  Tamaño propiedad=49620,9805 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)

- Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6 ⇒ Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6 ⇒ Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 4500 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6 ⇒ Tamaño propiedad=49620,9805 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 ⇒ Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 4500 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 ⇒ Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Distancia a caminos ripiados=180,0000 10 ⇒ Distancia a red vial=180,0000 9 acc:(0.83341)
- Densidad de bosque nativo radio 4500 m=0,0000 14 ⇒ Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000 12 acc:(0.8125)
- Distancia a centros urbanos=14799,4590 2 ⇒ Uso de suelo 2001=1,0000(Bosque nativo) 2 acc:(0.74999)
- Modelo 3, desde donde se generaron las reglas de asociación usando el asociador Predictive Apriori y 15 variables:
  - Tamaño propiedad=124931,2578 20 ⇒ Densidad de pradera radio 4500 m=0,0000 20 acc:(0.95439)
  - Precipitación anual=4000,0000 15 ⇒ Densidad de pradera radio 4500 m=0,0000 15 acc:(0.94106)
  - Tamaño propiedad=2002,0985 15 ⇒ USO\_2001=13,0000(Plantación) 15 acc:(0.94106)
  - Tamaño propiedad=58048,3750 15 ⇒ Densidad de praderas radio 4500 m=0,0000 15 acc:(0.94106)
  - Densidad bosque nativo radio 4500 m=0,0000 14 ⇒ Densidad de bosque nativo 2250 m=0,0000 14 acc:(0.93739)
  - Tamaño propiedad=4854,2798 12 ⇒ Uso suelo 2001=13,0000(Plantación) 12 acc:(0.92848)
  - Tamaño propiedad=28578,1016 11 ⇒ Densidad de praderas 4500 m=0,0000 11 acc:(0.923)
  - Precipitación anual=1500,0000 8 ⇒ Densidad de praderas radio 4500 m=0,0000 8 acc:(0.89994)
  - Tamaño propiedad=762,4233 7 ⇒ Uso de suelo 2001=13,0000(Plantación) 7 acc:(0.88884)
  - Precipitación anual=4000,0000 Tamaño propiedad=124931,2578 7 ⇒ Densidad de matorral radio 4500 m=0,0000 Densidad de pradera radio 4500 m=0,0000 7 acc:(0.88884)

- Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 4500 m=0,0000 7  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Esteba) 7 acc:(0.88884)
- Precipitación anual=4000,0000 15  $\implies$  Densidad de matorral radio 4500 m=0,0000 Densidad de pradera radio 4500 m=0,0000 14 acc:(0.88247)
- Tamaño propiedad=73718,1953 Uso de suelo 2001=8,0000(Nieves y glaciares) 14  $\implies$  Densidad de pradera radio 4500 m=0,0000 13 acc:(0.87511)
- Tamaño propiedad=55304,5898 22  $\implies$  Densidad de pradera radio 4500 m=0,0000 20 acc:(0.875)
- Tamaño propiedad=620,5640 6  $\implies$  Uso de suelo 2001=13,0000(Plantación) 6 acc:(0.87496)
- Tamaño propiedad=496,4666 6  $\implies$  Uso de suelo 2001=13,0000(plantación) 6 acc:(0.87496)
- Precipitación anual=4000,0000 Tamaño propiedad=73718,1953 6  $\implies$  Densidad de pradera radio 4500 m=0,0000 Uso de suelo 2001=8,0000(Nieves y glaciares) 6 acc:(0.87496)
- Tamaño propiedad=49620,9805 Densidad de bosque nativo achaparrado=0,0000 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Tamaño propiedad=49620,9805 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Esteba) 6 acc:(0.87496)
- Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Tamaño propiedad=49620,9805 Uso de suelo 2001=4,0000(Esteba) 6 acc:(0.87496)
- Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 4500 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Tamaño propiedad=49620,9805 Uso de suelo 2001=4,0000(Esteba) 6 acc:(0.87496)
- Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=8,0000(Nieves y glaciares) 6  $\implies$  Densidad de pradera radio 4500 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Esteba) 6  $\implies$  Tamaño propiedad=49620,9805 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Esteba) 6  $\implies$  Tamaño propiedad=49620,9805 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)

- Densidad de matorral radio 4500 m=0,0000 Uso de suelo 2001=8,0000(Nieves y glaciares) 20  $\implies$  Densidad de pradera radio 4500 m=0,0000 18 acc:(0.86364)
  - Tamaño propiedad=21323,7832 12  $\implies$  Densidad de pradera radio 4500 m=0,0000 11 acc:(0.85724)
  - Tamaño propiedad=13340,3096 11  $\implies$  Densidad de pradera radio 4500 m=0,0000 10 acc:(0.84624)
  - Tamaño propiedad=36993,9883 22  $\implies$  Densidad de pradera radio 4500 m=0,0000 19 acc:(0.83333)
  - Densidad de bosque nativo radio 4500 m=0,0000 14  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 12 acc:(0.8125)
  - Uso de suelo 2001=8,0000(Nieves y glaciares) 95  $\implies$  Densidad de pradera radio 4500 m=0,0000 75 acc:(0.78351)
  - Distancia a centros urbanos=14799,4590 2  $\implies$  Uso de suelo 2001=1,0000(Bosque nativo) 2 acc:(0.74999)
- Conforme a la perspectiva de tipo de cambio.
    - Modelo 4, desde donde se generaron las reglas de asociación usando el asociador Predictive Apriori y 15 variables:
      - Densidad de matorrales radio 4500 m=0,0000 70  $\implies$  Densidad de matorrales radio 2250 m=0,0000 70 acc:(0.98554)
      - Densidad de bosque nativo radio 4500 m=0,0000 14  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 14 acc:(0.93739)
      - Precipitación anual=4000,0000 Densidad de matorrales radio 2250 m=0,0000 14  $\implies$  Densidad de matorrales radio 4500 m=0,0000 14 acc:(0.93739)
      - Distancia red vial=180,0000 9  $\implies$  Distancia a caminos ripiados=180,0000 9 acc:(0.90902)
      - Elevación=732,0000 7  $\implies$  Densidad de matorrales radio 2250 m=0,0000 7 acc:(0.88884)
      - Distancia a caminos ripiados=180,0000 Uso de suelo 2001=13,0000 7  $\implies$  Distancia a red vial=180,0000 7 acc:(0.88884)
      - Precipitación anual=4000,0000 15  $\implies$  Densidad de matorrales radio 4500 m=0,0000 Densidad de matorrales radio 2250 m=0,0000 14 acc:(0.88247)
      - Distancia a red vial=900,0000 6  $\implies$  Distancia a caminos ripiados=900,0000 6 acc:(0.87496)
      - Distancia a caminos ripiados=90,0000 6  $\implies$  Distancia a red vial=90,0000 6 acc:(0.87496)
      - Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)
      - Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)

- Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=8,0000(Nieves y glaciares) 6  $\implies$  Densidad de matorrales radio 2250 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Distancia a caminos ripiados=180,0000 10  $\implies$  Distancia a red vial=180,0000 9 acc:(0.83341)
- Densidad de bosque nativo radio 4500 m=0,0000 14  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 12 acc:(0.8125)
- Precipitación anual=800,0000 Uso de suelo 2001=13,0000(Plantación) 23  $\implies$  Densidad de matorrales radio 2250 m=0,0000 19 acc:(0.8)
- Distancia a centros urbanos=14799,4590 2  $\implies$  Uso de suelo 2001=1,0000(Bosque nativo) 2 acc:(0.74999)
- Modelo 5, desde donde se generaron las reglas de asociación usando el asociador Predictive Apriori y 14 variables, este ajuste en la cantidad de variables para la generación de las reglas de asociación se debe a la capacidad de cómputo del equipo utilizado. La reglas obtenidas son:
  - Densidad de matorrales radio 4500 m=0,0000 70  $\implies$  Densidad de matorrales radio 2250 m=0,0000 70 acc:(0.98554)
  - Densidad de bosque nativo radio 4500 m=0,0000 14  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 14 acc:(0.93739)
  - Precipitación anual=4000,0000 Densidad de matorrales radio 2250 m=0,0000 14  $\implies$  Densidad de matorrales radio 4500 m=0,0000 14 acc:(0.93739)
  - Distancia a red vial=180,0000 9  $\implies$  Distancia a caminos ripiados=180,0000 9 acc:(0.90902)
  - Elevación=732,0000 7  $\implies$  Densidad de matorrales radio 2250 m=0,0000 7 acc:(0.88884)
  - Distancia a caminos ripiados=180,0000 Uso de suelo 2001=13,0000(Plantación) 7  $\implies$  Distancia a red vial=180,0000 7 acc:(0.88884)
  - Precipitación anual=4000,0000 15  $\implies$  Densidad de matorrales radio 4500 m=0,0000 Densidad de matorrales radio 2250 m=0,0000 14 acc:(0.88247)
  - Distancia a red vial=900,0000 6  $\implies$  Distancia a caminos ripiados=900,0000 6 acc:(0.87496)
  - Distancia a caminos ripiados=90,0000 6  $\implies$  Distancia a red vial=90,0000 6 acc:(0.87496)
  - Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)
  - Densidad de bosque nativo radio 2250 m=0,0000 Densidad de bosque nativo achaparrado radio 4500 m=0,0000 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6 acc:(0.87496)

- Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=8,0000(Nieves y glaciares) 6  $\implies$  Densidad de matorrales radio 2250 m=0,0000 6 acc:(0.87496)
- Densidad de bosque nativo achaparrado radio 4500 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 6  $\implies$  Densidad de bosque nativo radio 4500 m=0,0000 Densidad de bosque nativo radio 2250 m=0,0000 6 acc:(0.87496)
- Distancia a caminos ripiados=180,0000 10  $\implies$  Distancia a red vial=180,0000 9 acc:(0.83341)
- Densidad de bosque nativo radio 4500 m=0,0000 14  $\implies$  Densidad de bosque nativo radio 2250 m=0,0000 Uso de suelo 2001=4,0000(Estepa) 12 acc:(0.8125)
- Precipitación anual=800,0000 Uso de suelo 2001=13,0000(Plantación) 23  $\implies$  Densidad de matorrales radio 2250 m=0,0000 19 acc:(0.8)
- Distancia a centros urbanos=14799,4590 2  $\implies$  Uso de suelo 2001=1,0000(Bosque nativo) 2 acc:(0.74999)





## 6.4. Códigos Postgres y PostGIS

A continuación se detallan los códigos desarrollados para la preparación de las tablas desde donde se calculan las reglas de colocalidad.

### 6.4.1. Códigos Sql utilizados en creación de tablas de vecindades pares.

```
CREATE TABLE temp_duos AS(
  SELECT  a.id AS id_1, b.id AS id_2
          a.gridcode AS uso_1, b.gridcode AS uso_2
          a.area AS area_1, b.area AS area_2
          a.perimetro AS perimetro_1, b.perimetro AS perimetro_2
          a.the_geom AS the_geom_1, b.the_geom AS the_geom_2
  FROM uso_01 a, uso_01 b
  WHERE a.gid <> b.gid
  AND a.area > 9000 and b.area > 9000
  AND TOUCHES (a.the_geom, b.the_geom));
```

Esta consulta, crea una nueva tabla con todas las vecindades de pares del área de estudio. Una vez obtenida esta tabla, se calcula el largo de la *adyacencia* entre ambos polígonos por cada tupla, pasando a una nueva tabla previa a la definitiva para trabajar. La consulta que permite la creación de esta tabla es la siguiente:

```
CREATE TABLE temp-2_duos AS(
  SELECT  id_1, id_2
          uso_1, uso_2
          area_1, area_2
          perimetro_1, perimetro_2
  SELECT LENGHT(INTERSECTION(the_geom_1, the_geom_2) AS contacto
  the_geom_1, the_geom_2
  FROM temp_duos);
```

Luego se debe limpiar los *touches* < 1, esto se hace rápidamente con la siguiente instrucción:

```
DELETE FROM temp-2_duos WHERE contacto < 1;.
```

### 6.4.2. Código para el cálculo del porcentaje de adyacencia entre dos polígonos

```
CREATE TABLE duos AS(
  SELECT  id_1 AS id_1, id_2 AS id_2
          uso_1 AS uso_1, uso_2 AS uso_2
          area_1 AS area_1 , area_2 AS area_2
```

```

perimetro_1 AS perimetro_1 , perimetro_2 AS perimetro_2
contacto AS contacto
(contacto*100/perimetro_1) AS p100_de_1, (contacto*100/perimetro_2)
AS p100_de_2
the_geom_1 AS the_geom_1, the_geom_2 AS the_geom_2
FROM temp-2_duos);

```

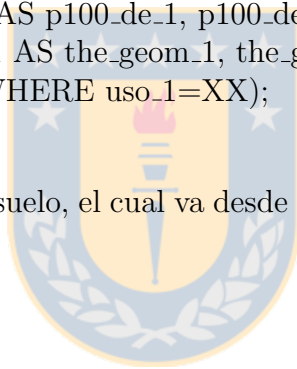
### 6.4.3. Código para generar las distintas tablas, diferenciadas cada una por ser la correspondiente a cada uno de los usos de suelo presente en el área de estudio

```

CREATE TABLE vec_de_XX AS(
  SELECT  id_1 AS id_1, id_2 AS id_2
          uso_1 AS uso_1, uso_2 AS uso_2
          area_1 AS area_1, area_2 AS area_2
          perimetro_1 AS perimetro_1, perimetro_2 AS perimetro_2
          contacto AS contacto
          p100_de_1 AS p100_de_1, p100_de_2 AS p100_de_2
          the_geom_1 AS the_geom_1, the_geom_2 AS the_geom_2
  FROM temp-2_duos WHERE uso_1=XX);

```

donde *XX* representa el uso de suelo, el cual va desde  $n=1$  a  $n=\text{cantidad de usos de suelo}$ .



## 6.5. Códigos para la generación de las reglas de co-localidad

```

<?php
$conn_string = "host=**** port=**** dbname=**** user=**** password=****";
$dbconn = pg_connect($conn_string) or die("No se pudo conectar");
echo conexión exitosa \n";

$archivo = "\ruta\ < archivo.txt >";
if (file_exists($archivo)){
    unlink($archivo);
}

$archivo2 = "\ruta\ < archivo - 2.txt >";
if (file_exists($archivo2)){
    unlink($archivo2);
}

$file2 = fopen($archivo2,'w+');
for($y=1;$y <15;$y++)
{
    $suma = pg_query($dbconn, "SELECT SUM(DISTINCT perimetro)
        FROM uso_01_clean WHERE gridcode=$y");
    while($row = pg_fetch_row($suma)){
    }
    for($x=1;$x<15;$x++)
    {
        $res = pg_query($dbconn,
            "SELECT SUM(DISTINCT perimetro_1)/(SELECT SUM(DISTINCT
            perimetro)FROM uso_01_clean WHERE gridcode=$y) FROM u01_vec_de_$y
            WHERE uso_2=$x");
        if(!$res){
            echo"Error. \n";
            exit;
        } while($row = pg_fetch_row($res)){
            if ($x!=1) fwrite($file, ",");
            fwrite($file, $row[0]); //El Pr para Y vec de X
            if (($y==1)&($x==1)){echo "Procesando Participation Index.";}
            else {echo ".";}
        }
    }
    fwrite($file, " \n");
}
echo "\n";
for($y=1;$y < 15;$y++)

```

```

{
    $suma = pg_query($dbconn, "SELECT SUM(DISTINCT perimetro)
        FROM uso_01_clean WHERE gridcode=$y");
    while($raw = pg_fetch_row($suma)){ }
    for($x=1;$x<15;$x++)
    {
        $res = pg_query($dbconn, "SELECT SUM(contacto)/(SELECT SUM(DISTINCT
            perimetro)FROM uso_01_clean WHERE gridcode=$y) FROM u01_vec_de_$y
            WHERE uso_2=$x");
        if(!$res){
            echo"Error. \n";
            exit;
        }
        while($row = pg_fetch_row($res)){
            if ($x!=1) fwrite($file2, ",");
            fwrite($file2, $row[0]); //El Cp para Y vec de X
            if (($y==1)&($x==1)){echo "Procesando Conditional Probability.";}
            else {echo ".";}
        }
    }
    fwrite($file2, "\n");
} pg_close($dbconn);
fclose($file);
fclose($file2);
echo "\n";
echo "Resultados guardados satisfactoriamente en archivos";
echo "\n";
? >

```



## 6.6. Resumen de los valores obtenidos para las métricas $Pr_1$ , $Pr_2$ y $Cp$

A continuación se exponen mediante tablas, los valores obtenidos para el  $Pr$  y la  $Cp$  de las reglas de colocalidad.

### 6.6.1. Razón de participación $Pr_1$ , ecuación 4.1

Figura 6.1: Resumen de la razón de participación obtenida para  $Pr_1$ .

PR	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,00000000	0,28882435	0,23543829	0,00978333	0,06399324	0,06391086	0,01598286	0,02350058	0,07322046	0,05056434	0,09688581	0,06654721	0,01110150	0,00024716
2	0,41222294	0,00000000	0,13880887	0,00923041	0,06373097	0,06267270	0,01164090	0,04976777	0,08483744	0,02454583	0,02998413	0,10482686	0,00758422	0,00014698
3	0,25194508	0,10407529	0,00000000	0,10949726	0,22146305	0,13684953	0,01386348	0,00522360	0,04725485	0,02386987	0,03279628	0,03206894	0,02078420	0,00030857
4	0,03171106	0,02096268	0,33166433	0,00000000	0,24347420	0,05908272	0,00240336	0,01241738	0,15381534	0,00367181	0,01081514	0,12824621	0,00106816	0,00066760
5	0,12042169	0,08402775	0,38944227	0,14135111	0,00000000	0,18123329	0,00883687	0,00480602	0,02325491	0,00724778	0,00817798	0,01441805	0,01596837	0,00081392
6	0,15845376	0,10886994	0,31706072	0,04519226	0,23877853	0,00000000	0,02445999	0,00551499	0,02936220	0,04034111	0,01108104	0,00791503	0,01057039	0,00240004
7	0,19010289	0,09701127	0,15409113	0,00881921	0,05585497	0,11734444	0,00000000	0,02302793	0,10950514	0,09137678	0,13694268	0,00465458	0,00979912	0,00146987
8	0,10440113	0,15490896	0,02168542	0,01701894	0,01134596	0,00988197	0,00860097	0,00000000	0,47021685	0,00228749	0,06825876	0,12992955	0,00073200	0,00073200
9	0,13755611	0,11167002	0,08295929	0,08915029	0,02321622	0,02224888	0,01729608	0,19884693	0,00000000	0,02619564	0,08508745	0,20434143	0,00116081	0,00027086
10	0,36729503	0,12492519	0,16202873	0,00822861	0,02797726	0,11819270	0,05580491	0,00374028	0,10128665	0,00000000	0,02558348	0,00029922	0,00254339	0,00209455
11	0,37996769	0,08239095	0,12019386	0,01308562	0,01704362	0,01752827	0,04515347	0,06025848	0,17762520	0,01381260	0,00000000	0,06801292	0,00492730	0,00000000
12	0,17670221	0,19502324	0,07957342	0,10505879	0,02034454	0,00847689	0,00103910	0,07765928	0,28881597	0,00010938	0,04604867	0,00000000	0,00114848	0,00000000
13	0,21095890	0,10097847	0,36908023	0,00626223	0,16125245	0,08101761	0,01565558	0,00313112	0,01174168	0,00665362	0,02387476	0,00821918	0,00000000	0,00117417
14	0,08163265	0,03401361	0,09523810	0,06802721	0,14285714	0,31972789	0,04081633	0,05442177	0,04761905	0,09523810	0,00000000	0,00000000	0,02040816	0,00000000

### 6.6.2. Razón de participación $Pr_2$ , ecuación 4.2

Figura 6.2: Resumen de la razón de participación obtenida para  $Pr_2$ .

PR	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		0,92743705	0,90060456	0,12134883	0,69553479	0,70376631	0,60122560	0,64775211	0,77960936	0,61379335	0,80098213	0,75153512	0,36372709	0,14632798
2	0,97904958		0,45048976	0,04969665	0,25218878	0,23345525	0,11160028	0,37831620	0,47677677	0,11406303	0,22242891	0,43644625	0,04018961	0,00088403
3	0,81152179	0,57919933		0,51337244	0,77634857	0,66598817	0,26659540	0,14142683	0,46086842	0,23111852	0,40523410	0,25752094	0,21935456	0,00233321
4	0,55102775	0,51546493	0,89741656		0,75260397	0,60268185	0,35627354	0,38492170	0,75534486	0,39521283	0,39427284	0,55760181	0,00478511	0,02118593
5	0,55389645	0,45157260	0,90648032	0,41838593		0,71467429	0,08828258	0,05772001	0,25023414	0,07378097	0,11044480	0,11036400	0,13969727	0,03764263
6	0,69356465	0,58905979	0,86205516	0,12450656	0,70776261		0,25125028	0,15322823	0,32078832	0,31234251	0,17942126	0,09507133	0,18777842	0,05691452
7	0,80618755	0,62717836	0,47708866	0,06157364	0,19154403	0,46009578		0,17823529	0,44887573	0,51166636	0,59965233	0,01228448	0,03789543	0,08684518
8	0,67835832	0,77938890	0,09418410	0,02645422	0,05633038	0,00768110	0,27080049		0,91462528	0,00674385	0,65673667	0,53666334	0,00016015	0,00042235
9	0,75696923	0,66093746	0,39571114	0,30417280	0,16769305	0,04721896	0,18261111	0,79851918		0,05899829	0,59400199	0,66771785	0,10198714	0,00015219
10	0,92397403	0,49067558	0,35750563	0,02836933	0,09401284	0,36542761	0,37678618	0,03614297	0,45866828		0,22205637	0,00060714	0,03380316	0,03727501
11	0,89803663	0,33871850	0,30070502	0,03239394	0,05979709	0,08920382	0,13129326	0,28845230	0,51767778	0,09056486		0,23313461	0,01491661	
12	0,69632982	0,66508187	0,43114100	0,43067773	0,11222196	0,06389662	0,02675978	0,35308509	0,89941833	0,01670382	0,21516692		0,00468389	
13	0,94285239	0,71369881	0,95695713	0,13218509	0,82083303	0,68416110	0,18280415	0,01553338	0,25437651	0,11632119	0,37007119	0,09386948		0,01766803
14	0,65723066	0,06260837	0,49575378	0,19138797	0,61078203	0,89708386	0,36300591	0,59363804	0,47599203	0,36304661			0,33331152	



### 6.6.5. Índices de participación (min Pr) para $Pr_1$ , ecuación 4.1

Figura 6.5: Resumen de los índices de participación obtenidos para  $Pr_1$

<b>PR 1 min</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>
1 Bosque nativo		0,29	0,24	0,01	0,06	0,06	0,02	0,02	0,07	0,05	0,10	0,07	0,01	0,00
2 Bosque nativo achaparrado				0,01	0,06	0,06	0,01	0,05	0,08	0,02	0,03	0,10	0,01	0,00
3 Matorral		0,10		0,11	0,22	0,14	0,01	0,01	0,05	0,02	0,03	0,02	0,02	0,00
4 Estepa							0,00	0,01		0,00	0,01		0,00	0,00
5 Pradera				0,14		0,18	0,01	0,00	0,02	0,01	0,01	0,01	0,02	0,00
6 Agrícola				0,05			0,02	0,01		0,04	0,01	0,01	0,01	0,00
7 Cuerpos de agua								0,02				0,00	0,01	0,00
8 Nieves y glaciares						0,01	0,01			0,00			0,00	0,00
9 Áreas sin vegetación				0,09	0,02	0,02	0,02	0,20		0,03	0,09	0,20	0,00	0,00
10 Humedales							0,06	0,00				0,00	0,00	0,00
11 Sombras				0,01			0,05	0,06		0,01			0,00	0,00
12 Estepa altoandina				0,11		0,01		0,08		0,00	0,05		0,00	0,00
13 Plantaciones								0,00						0,00
14 Urbano											0,00	0,00		0,00

