



Aplicación de técnicas de clusterización  
para la clasificación de grupos vegetacionales.

*Matías F. Rebolledo G.*

# Índice general

<b>1. Introducción.</b>	<b>4</b>
<b>2. Estado del arte.</b>	<b>5</b>
<b>3. Marco teórico.</b>	<b>12</b>
3.1. Método de k-medias. . . . .	12
3.1.1. Estimación puntual. . . . .	12
3.1.2. Estimación de intervalo. . . . .	12
<b>4. Propuesta.</b>	<b>13</b>
4.1. Selección de variables para identificación de grupos homogéneos. . . . .	13
4.1.1. Descripción de los datos del estudio. . . . .	13
4.2. Desarrollo de un procedimiento automatizado de identificación de grupos. . . . .	14
4.2.1. Pre procesamiento de imágenes. . . . .	15
4.2.2. Transformación de imágenes. . . . .	16
4.2.3. Aplicación de algoritmo de clasificación. . . . .	17
4.2.4. Estadígrafos de evaluación del algoritmo k-medias. . . . .	19
4.2.5. Reconversión de big matrices en imágenes rasters. . . . .	20
4.2.6. Caracterización de agrupamientos. . . . .	20
4.2.7. Evaluación de precisión de clasificación por matriz de confusión. . . . .	22
<b>5. Experimentos.</b>	<b>23</b>
5.1. Resultados de las corridas de kmedias puntuales. . . . .	23
<b>6. Conclusiones.</b>	<b>24</b>
<b>Anexos.</b>	<b>25</b>
Resumen cuadros indicadores k-medias puntuales. . . . .	25
Resumen gráficos indicadores k-medias puntuales. . . . .	33
<b>Referencias.</b>	<b>33</b>
Enlaces consultados. . . . .	33
Medición de precisión en mapas temáticos. . . . .	33
Creación de matrices de distancia de gran tamaño. . . . .	33
Algoritmos usados en la tesis. . . . .	33

Artículos consultados. . . . .	34
--------------------------------	----

## Índice de cuadros

1. Comparación del número de especies amenazadas. . . . .	6
2. Causas, Consecuencias, Impactos y Métodos de estimación de los costos de la desertificación y degradación de las tierras. . . . .	7
3. Valoración económica del impacto sobre la biodiversidad. . . . .	8
4. Relación entre la biodiversidad y los servicios ecosistémicos. . . . .	8
5. Descripción de zonas del estudio. . . . .	13
6. Descripción de las variables del estudio. . . . .	14
7. Tamaños de imágenes satelitales según grupo. . . . .	14
8. Resumen ejecuciones kmedias puntuales para grupo00. . . . .	26
9. Resumen ejecuciones kmedias puntuales para grupo01. . . . .	27
10. Resumen ejecuciones kmedias puntuales para grupo02. . . . .	28
11. Resumen ejecuciones kmedias puntuales para grupo03. . . . .	29
12. Resumen ejecuciones kmedias puntuales para grupo04. . . . .	30
13. Resumen ejecuciones kmedias puntuales para grupo05. . . . .	31
14. Resumen ejecuciones kmedias puntuales para grupo06. . . . .	32

## Índice de gráficos

1. Proceso de homogenización de imágenes satelitales. . . . .	15
2. Proceso de conversión de imágenes satelitales. . . . .	16
3. Proceso de creación de objetos de memoria compartida. . . . .	17
4. Proceso de ejecución de k-medias con big matrices. . . . .	18
5. Proceso de extracción de estadígrafos de evaluación. . . . .	19
6. Proceso de rasterización desde big matrices. . . . .	20
7. Proceso de caracterización de variables según cluster. . . . .	21

# 1. Introducción.

El presente informe resume las labores realizadas durante la implementación de un método de clusterización usando el algoritmo de k-medias para identificar grupos homogéneos de vegetación. Describe los pasos que involucraron, preparación de la información, aplicación del método de k-medias, estimación de métricas de desempeño para obtener el número óptimo de clusters, evaluación de la precisión mediante el método de matriz de confusión y generación de las imágenes resultantes clasificadas.

Tanto en áreas de biodiversidad como en temas medio ambientales diversos, es necesario identificar zonas que posean características homogéneas que puedan ser separadas en sub divisiones para poder implementar planes específicos por cada entidad en virtud de sus características distintivas. Una de las formas es poder hacer uso de imágenes satelitales que representen un conjunto de variables de interés las que a su vez puedan ser agrupadas en grupos más pequeños con características homogéneas entre sí. La imagen satelital además de ser una fuente de datos, representa una zona determinada del espacio y es generada mediante sensores presentes en satelites y vehículos de vuelo especializados. Conforme la tecnología con que el sensor genera la imagen, se tiene que ésta puede contener errores generados por las condiciones atmosféricas como por las mismas propiedades físicas con las que cada sensor genera la información. Por este motivo toda información espacial debe ser cotejada con métodos en terreno para corroborar que la información con la que se cuenta sea lo más exacta posible.

En las últimas décadas, técnicas estadísticas han sido utilizadas para la identificación de grupos homogéneos en diferentes zonas espaciales mediante el uso de información satelital. En nuestro estudio haremos uso de la técnica denominada algoritmo de k-medias o *k-means* en inglés que permite separar grupos homogéneos mediante el criterio de minimización de la suma de cuadrados dentro de cada grupo. A pesar que esta como otras técnicas han ido adoptándose progresivamente para responder preguntas de interés de la biodiversidad y el medio ambiente, hemos sido testigos de un aumento en las capacidades de cómputo con un consecuente abaratamiento en los costos de adquisición de nuevas tecnologías. Esto ha tenido una gran implicancia en términos de viabilidad de implementación de proyectos que requieran mecanismos automatizados que generen información útil a partir de datos que sería considerablemente inviables de conseguir operacionalmente, ya sea en tiempo, costos o mano de obra.

Siguiendo esta línea, esta investigación incorpora un método novedoso para el procesamiento de información satelital que es de por si de un tamaño elevado y que ha restringido muchos estudios en cuanto a recursos computacionales.

## 2. Estado del arte.

En las últimas décadas, las macropresiones se han intensificado tanto cualitativa como cuantitativamente, debido a un mayor crecimiento económico, el cual en una economía extractiva como la chilena, ejerce mayor presión por recursos hídricos, silvícolas, de suelo, mar, alterando los ecosistemas. Por ejemplo la agricultura, que utiliza amplias coberturas espaciales y está inserta en ecosistemas vivos, ha generado presión por agua y suelo, llegando a afectar/amenazar el bosque esclerófilo mediterráneo, que es endémico de cerros y quebradas, debido a cultivos vitivinícolas y frutales que se instalan en ellos. Por parte de la industria forestal, su crecimiento ha afectado al bosque nativo y su biodiversidad y ha significado un deterioro de la capacidad hídrica de las cuencas. Se calcula una pérdida o reducción de bosque nativo en a lo menos 237.126 ha entre 1999 y 2015, tanto por conversión de praderas y matorrales (48 %) como por sustitución por plantaciones forestales exóticas (41 %). El crecimiento de la minería, ha presionado por agua, suelo subsuelo, aire y glaciares, llegando a un punto crítico en que las macrozonas promedian un 2000 % por aprovechamiento consuntivo de aguas subterráneas, mientras en algunas de ellas las estaciones de monitoreo muestran caídas de un 20 % en las precipitaciones. El mayor poder adquisitivo de la población, ha aumentado la presión por agua, ha incrementado el parque automotriz y la contaminación del aire, presentando graves problemas de concentración de MP10 y MP2.5 en todas las regiones con excepción de Magallanes. La presión por políticas públicas medioambientales y fiscalización, será mayor.

La creación del Ministerio del Medio Ambiente, es quizá uno de los mayores cambios en este contexto y destaca también el avance en materia de regulación ambiental con la entrada en vigencia de al menos 11 normas de emisión de contaminantes y cambios en la forma de determinar caudales hidrológicos. La entrada en vigencia el 2008 de la Ley de bosque nativo y fomento forestal, o el 2016 del derecho real de conservación medioambiental Ley 20.930, son asimismo pasos importantes para frenar el avance sostenido de la pérdida de bienes de la naturaleza (suelos, aguas continentales, bosques nativos, biodiversidad del mar y borde costero). Sin embargo, la fiscalización que pueda realizar la Superintendencia del Medio Ambiente y las instituciones sectoriales del Estado es limitada, por ejemplo los brigadistas son escasos, mientras que las extensiones son muy amplias y se encuentran diseminadas por el territorio nacional. En ese escenario las tecnologías de monitoreo hidrológicos, satelital, cobran protagonismo. Contar con más y mejor información es vital. La apertura pública a las fuentes de la información de monitoreo hidrológico, gatillada en parte por la promulgación de la Ley sobre el acceso a la información pública, va en esa dirección. Hoy se cuenta con a lo menos 7 fuentes de datos disponibles en internet, alguna de ellas, como la de la DGA con datos de “monitoreo instantáneos”.

En cuanto a la biodiversidad (tierra y mar), es imprescindible contar con estudios de línea base, se requiere de mucha más investigación, existe una gran brecha al respecto. Por ejemplo, si queremos saber cual es la evolución de la biodiversidad genética en Chile, existen escasos antecedentes para realizar comparaciones. El cuadro 1

compara datos tomados el año 1999 versus el año 2015:

Grupo	1999	2015	%Diferencia
Mamíferos	51	83	39 %
Aves	72	76	5.5 %
Reptiles	45	81	80 %
Anfibios	31	47	16 %
Peces	44	37	-15.9 %
Insectos	s/i	35	-
Otros invertebrados	s/i	65	-
Plantas	69	421	510 %
TOTAL	312	845	171 %

Fuente: Informe país, estado del medio ambiente en Chile, comparación 1999-2015.

Cuadro 1: Comparación del número de especies amenazadas.

Señalar que el aumento de las especies bajo amenaza, se debe además de los factores señalados, al descubrimiento de nuevas especies, gracias a los esfuerzos de más investigaciones.

A las especies bajo amenaza del cuadro 1, se suman los problemas de la erosión. En la categoría “suelos graves y muy graves” en 1979 afectaba a 11.506.000 hectáreas, pasando en el 2010 a 18.112.000 has. Según el estudio “los costos de la inacción de la desertificación y degradación de las tierras” aquellos suelos áridos, semiáridos y subhúmedos secos superan el 60 % del territorio nacional afectado por algún grado de desertificación, donde se concentran los mayores daños a los suelos, la biodiversidad y a la productividad silvoagropecuaria en general. A la falta de estudios sobre estas materias, se suma la dificultad de estimar el costo o valor económico que representa la pérdida de patrimonio natural. En el caso de la desertificación, no existe una masa de estudios que permita comparar resultados y validar metodologías, mientras por otra parte, la información necesaria para llevar a cabo las investigaciones en ocasiones no es suficiente, no está actualizada o no corresponde a lo requerido. Instituciones como el SAG, CNR, CONAF, ODEPA, CIREN y los censos nacionales agropecuarios, proporcionan ciertos datos e información, que mediante algunas metodologías, tales como: “costos de reemplazo”, “costos por erosión” y “estimaciones econométricas de las funciones de producción y rendimientos” se pudo obtener una valorización de dichos costos de inacción. Así por ejemplo, la Región de Coquimbo es la más afectada por el proceso de desertificación, con unos costos de inacción equivalentes al 23,4 % de su PIB agropecuario. Paralelamente en la Región Metropolitana estos costos representan un 9,36 % de su PIB agropecuario regional.

Es importante en un mundo economicista, el traducir a valores económicos el impacto que significa la pérdida de biodiversidad y patrimonio natural. Considerar asimismo que los esfuerzos en el sector público para establecer estrategias y políticas de sustentabilidad no han tenido efecto práctico en la conservación de la biodiversidad. Los reiterados déficits en la aplicación del Sistema de Evaluación de Impacto Ambiental, las tergiversaciones legales en torno a este sistema y la falta de jerarquización que se le da a la biodiversidad en los estudios de

Causas de la desertificación, degradación y sequía (directas, D; asociadas, A)		Consecuencias	On-Off Site	Impactos	Costos directos (D), Indirectos (I)	Metodos de evaluación
Socioeconómicas	Pobreza (d)	Sobreexplotación de recursos vulnerables	On/Off Site	Pérdida de productividad	D	Funciones de producción y de rendimientos, índices de productividad, costos de la migración
	Tenencia y propiedad de la tierra (d)	Cambios en la productividad del predio		Pobreza	D/I	
	Factores demográficos (d)	Migraciones		Migraciones	I	
Físicas	Topografía (d)	Erosión	On Site	Pérdida de productividad	D	Funciones de producción y rendimientos, costos de reemplazo
	Erodabilidad (d)	Erosión	On Site	Pérdida de productividad	D	Costos de reemplazo
	Mal drenaje (d)	Salinización, anegamientos	On/Off Site	Pérdida de productividad	D	Costos de desalinización, costos de reemplazo
	Clima (d/A)	Degradación del suelo	On Site			
Económicas	Deforestación (d), pérdida de cobertura vegetal (d)	Degradación. Pérdida de productividad	On Site	Pérdida de fertilidad por erosión	D/I	Costos de reemplazo de insumos, como fertilizantes
	manejo no sostenible (d)			Degradación por contaminación		Funciones de producción, costos de reemplazo, costos de descontaminar
	Intensificación de la agricultura (d/a)					Costos de días de trabajo perdidos
	Cambio de uso de suelo		On/Off Site			Costos de tierras perdidas
	Desarrollo de infraestructura (p)	Pérdida de tierras agropecuarias	On Site/Off Site			

Fuente: Chile, Los costos de la inacción de la desertificación y degradación de las tierras. PNUD-UE- CEPAL.

Cuadro 2: Causas, Consecuencias, Impactos y Métodos de estimación de los costos de la desertificación y degradación de las tierras.

impacto, han sido también factores de pérdida de especies y deterioro de ecosistemas. A todo lo anterior, se suma la intangibilidad de los atributos que tiene o presenta la biodiversidad tanto para sectores productivos o para el propio ser humano, lo que redundará en más complejidades para determinar el costo/pérdida. Una metodología utilizada para esta finalidad es la de “valoración económica total o VET”. Con esta metodología, por ejemplo se valorizó el impacto que 7 especies exóticas invasoras o EEI, generan al país:

Las EEI tienen tal impacto que el año 2005 se creó el “comité operativo para el control de EEI” (COCEI), el cual vía “resolución exenta 684” del año 2013 se formalizó y desde el año 2015 se amplió a un “plan de acción para la gestión de especies exóticas invasoras”, enmarcado en la estrategia nacional de biodiversidad que se ha planteado Chile. La biomasa chilena (productiva y biodiversidad) requiere de mayor estudio para saber su efecto multiplicador en el medio ambiente y también en las actividades económicas. En esa línea, el impacto de las EEI del cuadro 3 supondría una pérdida país proyectada a 20 años de 2 mil millones de dólares. Si agregamos impacto-costos que aportaría la actividad productiva, el consumo humano, el cambio climático, la pérdida sería más elevada.

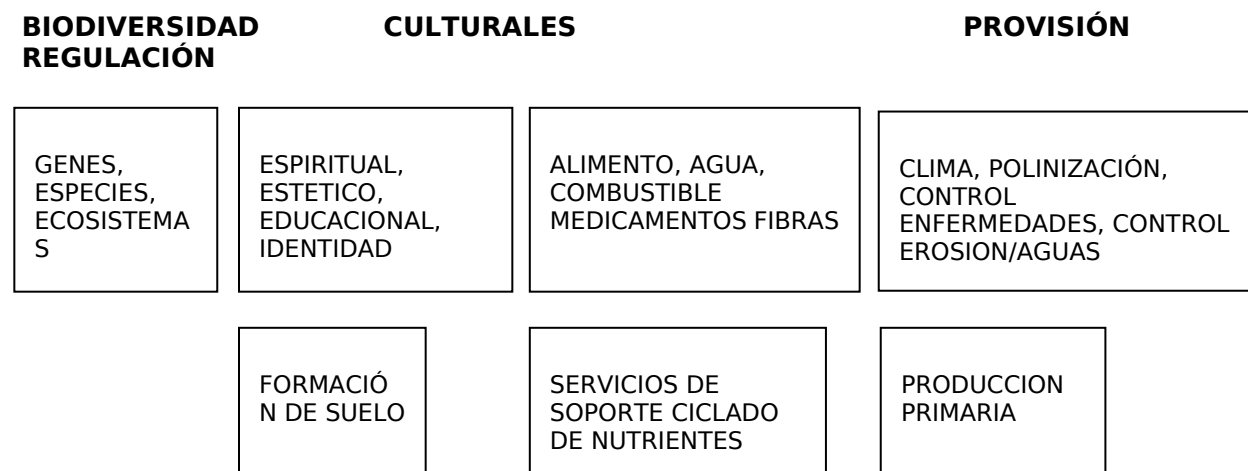
La situación nacional es extrapolable al concierto global, esto se refleja en que la conservación y gestión sustentable de la naturaleza es parte de la nueva “Agenda de Desarrollo Global” para los próximos 15 años. Adoptada en la

ESPECIE	LO QUE CHILE YA HA GASTADO O PERDIDO (COSTO HISTÓRICO)	PÉRDIDA ANUAL	PÉRDIDA PROYECTADA A 20 AÑOS	PÉRDIDA TOTAL PROYECTADA A 20 AÑOS
CASTOR	(-) 64.539.660	(-) 733.094	(-) 5.066.895	(-) 69.606.555
CONEJO	(-) 1.465.754	(-) 3.249.337	(-) 91.962.701	(-) 93.428.455
JABALÍ	(-) 62.321	(-) 38.278.724	(-) 603.249.711	(-) 603.312.032
VISÓN	(-) 608.271	(-) 9.526.620	(-) 416.723.985	(-) 417.332.256
CHAQUETA AMARILLA	(-) 66.766	(-) 21.532.771	(-) 496.970.674	(-) 497.037.440
ZARZAMORA	(-) 145.085.398	(-) 10.580.010	(-) 128.632.996	(-) 273.718.394
ESPINILLO	(-) 54.257	(-) 4.038.771	(-) 49.103.849	(-) 49.158.106
TOTAL	(-) 211.882.427	(-) 87.939.327	(-) 1.791.710.811	(-) 2.003.593.238

Fuente: Valoración económica del impacto de siete especies exóticas invasoras sobre los sectores productivos y la biodiversidad en Chile.

Cuadro 3: Valoración económica del impacto sobre la biodiversidad.

Asamblea General de las Naciones Unidas, en septiembre del 2015 y que refleja el consenso universal existente en torno a la necesidad de cuidar la diversidad biológica de nuestro planeta. De hecho, el objetivo 15 de este nuevo acuerdo explicita la necesidad de proteger, restaurar y promover la utilización sostenible de los ecosistemas terrestres, gestionar de manera sostenible el bosque, combatir la desertificación, detener y revertir la degradación de la tierra, y frenar la pérdida de diversidad biológica. En este sentido la existencia de áreas protegidas, las reservas naturales, son claves para resguardar y proteger esta diversidad biológica.



Fuente: Conservando el patrimonio natural de Chile, el aporte de las áreas protegidas.

Cuadro 4: Relación entre la biodiversidad y los servicios ecosistémicos.

Las Áreas Silvestres Protegidas del Estado (SNASPE) administrado por CONAF cubren 14,3 millones de ha de los cuales 3,9 millones corresponden a bosque nativo, con escaso crecimiento entre los años 1999 y 2015. Tendencia diferente presentaron las áreas protegidas privadas (APP), las cuales crecieron desde 317.000 en 1999 a más de un millón de hectáreas en 2015. Las áreas silvestres nos entregan servicios ecosistémicos de gran valor económico, ambiental y cultural. Por ello, contar con un sistema nacional que regule esta figura en Chile, obedece



a la necesidad de proteger el patrimonio natural existente, configurado por una biodiversidad única en el mundo. El objetivo principal de las áreas protegidas es la conservación de la biodiversidad, pero también contribuyen de manera efectiva al desarrollo sustentable de los territorios en los que están insertas. Los servicios ecosistémicos que proveen las áreas protegidas de Chile demuestran que invertir en conservación de la biodiversidad tiene impactos positivos en el bienestar de la sociedad chilena. Hay que entender que el patrimonio natural es el conjunto de valores naturales que tienen importancia desde un punto de vista estético, cultural, ambiental, científico y productivo. Este patrimonio está constituido por la variedad de paisajes, elementos físicos como cerros, ríos y lagos, así como por la flora y la fauna que habita en un territorio. El patrimonio es aquello reconocido como propio y nuestro, y cambia a lo largo del tiempo. Un patrimonio se va construyendo en y con el desarrollo de la cultura de los pueblos. Además, el patrimonio resulta de una serie de cambios, construcciones, destrucciones y decisiones que se han tomado a lo largo de la historia.

En este sentido el año 1998 la CONAMA expuso explícitamente una declaración política ambiental, siendo el primer hito en esta materia, el año 2008 se incorporaron cambios sugeridos en la institucionalidad ambiental; cambio climático, pérdida de biodiversidad, escasez de agua, la contaminación y la calidad de vida en ciudades y el ingreso de Chile al OCDE que impone un elevado estándar a las políticas públicas y a la calidad del crecimiento. La modificación de la Ley 19.300 sobre medio ambiente. A fines del año 2010 (Ley 20.417), se reformula completamente el Sistema de Evaluación Ambiental y se crea el Ministerio del Medio Ambiente, el Consejo de Ministros para la Sustentabilidad, el Servicio de Evaluación Ambiental, los Consejos Consultivos y la Superintendencia del Medio Ambiente. A través de la Ley 20.600 se establecen tres tribunales ambientales, en Santiago, Valdivia y Antofagasta. El año 2011 el Ejecutivo somete al Congreso Nacional el Proyecto de Ley que crea el Servicio de Biodiversidad y Áreas Protegidas.

La institucionalidad Ambiental vigente el 2015, definida a partir de la Ley 20.147, responde a la creciente demanda por una gestión ambiental con mayores niveles de eficacia y eficiencia, cada vez mayor cobertura. La evaluación del desempeño ambiental realizada por la OCDE con apoyo de la CEPAL respecto del periodo 1990 – 2004, concluía que era “necesario desarrollar y fortalecer las instituciones ambientales en los ámbitos nacional y regional, así como la capacidad de cumplimiento y fiscalización”. Se definen los cuatro instrumentos de gestión ambiental, a saber: a) Resoluciones de Calificación Ambiental, b) Planes de Prevención y/o Descontaminación, c) Normas Ambientales y d) Planes de Manejo. Posteriormente el informe de desempeño ambiental CEPAL/OCDE de 2016, la capacidad humana y técnica de las instituciones ambientales había aumentado significativamente, “pero todavía sin alcanzar el nivel requerido para llevar adelante sus funciones con eficacia”. El informe enfatiza en que “es preciso fortalecer aún más la base de información utilizada para fundamentar el proceso de toma de decisiones relativa al medio ambiente”.

El patrimonio natural de Chile posee características únicas, como por ejemplo, contar con el desierto de Atacama, uno de los sitios con menos precipitación del planeta y a su vez tener áreas con el mayor número de días lluviosos al año. Asimismo, Chile cuenta con una de las reservas de agua dulce más importantes del planeta. Estas características particulares de la naturaleza nacional diferencian su biodiversidad, especialmente la terrestre. A nivel de ecosistemas, Chile contiene algunos reconocidos por su importancia a nivel mundial, tales como el desierto de Atacama, el matorral y bosque esclerófilo, el bosque lluvioso templado valdiviano, la estepa patagónica, los lagos andinos, la puna altoandina y los ecosistemas marinos costeros asociados a la corriente de Humbolt y a zonas de sugerencias. Además, aproximadamente un 40 % del área terrestre, principalmente la parte centro-norte del país, así como varias islas cercanas, incluyendo el archipiélago de Juan Fernández, forman parte del “punto caliente” de los bosques valdivianos lluviosos de invierno de Chile, uno de los 34 puntos críticos de biovidersidad a nivel mundial. Estos puntos se definen como regiones con unas 1.500 especies de plantas endémicas y que han perdido al menos un 70 % de su hábitat original.

Al mismo tiempo, Chile contiene ecosistemas aún poco intervenidos. Nuestro país abarca el bloque más grande de bosques templados en el hemisferio sur, y uno de los tramos más grandes de bosques naturales relativamente sin intervención en el mundo. El país tiene una larga tradición que ha entrelazado nuestro desarrollo con la naturaleza, relacionándose de diversos modos con la biovidersidad. Los pueblos originarios basaban su sistema de vida y costumbres en el uso de diferentes recursos naturales. El sustento de nuestra economía se basa en el uso de los recursos naturales y en la utilización del medio ambiente por los distintos productores, como la minería, la pesca, la industria forestal y acuícola, la agricultura y el turismo. Estas actividades productivas modifican el espacio, al mismo tiempo que las comunidades desarrollan cierta forma de vida que les es propia: domesticar especies, transforman relieve, hacen caminos, canalizan los ríos. En nuestra vida cotidiana recibimos constantemente los beneficios de la naturaleza, el desafío, entonces, es encontrar la forma sustentable para la conservación de este patrimonio diverso. El ambiente juega un rol de proveedor de bienes y servicios como el agua, el aire, el suelo, entre otros y debe protegerse de amenazas, precisamente con el fin de no poner en peligro las potenciales fuentes de desarrollo de nuevas oportunidades para los seres humanos de ahora y el futuro.

A lo largo de los años, el país ha creado leyes y organismos para la protección de la biodiversidad, logrando hasta hoy declarar diversos tipos de áreas protegidas que representan en conjunto el 20,7 % del territorio nacional continental y el 4 % del mar territorial chileno. La naturaleza es parte de nuestra vida cotidiana y está íntimamente ligada con nuestra identidad y desarrollo como país. Es por ello que la conservación de la biodiversidad es una inversión que mantiene viva la base de nuestro desarrollo, el bienestar de los chilenos y las fuentes de oportunidades para las personas. Las tecnologías pueden aportar a producir y procesar grandes volúmenes de datos, de información, que facilite la recopilación de cifras comparables, ayude a las investigaciones

futuras a disponer de mayor y mejor calidad de información, oportuna y actualizada. El seguimiento a factores como el avance de la desertificación, o a las EEI, al reemplazo de bosque nativo por exótico, implica monitoreo en zona norte, centro y sur como en amplias y variadas extensiones, en ese escenario sin lugar a dudas que las tecnologías (satelital, drones, de monitoreo) son una necesidad para generar tanto una línea base como una masa crítica de datos e información para mejorar la toma de decisión y con ello, el diseño de las políticas públicas en materias medio ambientales y de biodiversidad.

### 3. Marco teórico.

#### 3.1. Método de k-medias.

El método de k-medias fue diseñado a mediados de los años cincuenta <sup>1</sup> por el autor (Lloyd 1982) y consiste en un algoritmo iterativo que divide el conjunto de datos en particiones cuya suma de cuadrados dentro de cada partición es mínima.

##### 3.1.1. Estimación puntual.

##### 3.1.2. Estimación de intervalo.

Para cada valor de  $K(1, 2, \dots, K^{MAX})$ :

- Realizar  $M$  corridas ( $>30$ ), cada una con una semilla al azar.
- En cada una de las  $M$  corridas, calcular las medidas de desempeño ( $R^2$ , Elbow, CH), obteniendo así  $M$  valores de cada una.
- Entonces considerar la medida  $x$  (cualquiera de ellas), calcular el promedio  $\bar{x}$  sobre los  $M$  valores y su desviación estándar:

$$s = \sqrt{\frac{\sum_{i=1}^M (x_i - \bar{x})^2}{M - 1}}$$

- Luego el intervalo de confianza para el desempeño promedio de  $x$  se calcula como:

$$\left[ \bar{x} - z^* \cdot \frac{s}{\sqrt{M}}, \bar{x} + z^* \cdot \frac{s}{\sqrt{M}} \right]$$

Para el caso de que el intervalo tenga un 95 % de confianza, el valor de  $z^*$  es 1.96.

---

<sup>1</sup>El artículo fue publicado casi tres décadas después según Wikipedia. Consultado en Septiembre 2019.

## 4. Propuesta.

En esta sección describiremos la estrategia que se usará para poder clasificar grupos vegetaciones y topográficos usando el algoritmo de k-medias. Esta consiste en varios pasos necesarios para poder procesar la información de manera eficiente.

### 4.1. Selección de variables para identificación de grupos homogéneos.

La definición de la muestra consistió en definir un área total para el estudio y también dividir el área total en áreas más pequeñas siguiendo un modelo de clasificación de vegetación previamente definido por los autores (Luebert and Plischoff 2006). Ellos definen trece zonas vegetacionales para la Región Metropolitana y la de Valparaíso con las cuales, definimos seis grupos que constituyen zonas geográficas adyacentes cuya finalidad es disminuir el área total del proyecto en zonas de menor tamaño y por otro lado, mantener homogeneidad dada por los grupos colindantes en cuestión. Adicionalmente se crea un grupo cero que equivale a la sumatoria de las seis zonas menores (ver cuadro 5).

Grupo	Descripción
grupo00	TODOS: Superficie total equivalente a la suma de todos los códigos de área
grupo01	P34: Bosque espinoso mediterráneo costero de <i>Acacia caven</i> - <i>Maytenus boaria</i>
grupo02	P39: Bosque esclerofilo mediterráneo costero de <i>Cryptocarya alba</i> - <i>Peumus boldus</i> P40: Bosque esclerofilo mediterráneo costero de <i>Lithrea caustica</i> - <i>Cryptocarya alba</i>
grupo03	P110: Matorral bajo mediterráneo costero de <i>Chuquiraga oppositifolia</i> - <i>Mulinum spinosum</i> P46: Bosque caducifolio mediterráneo costero de <i>Nothofagus macrocarpa</i> / <i>Ribes punctatum</i>
grupo04	P27: Matorral espinoso mediterráneo interior de <i>Trevoa quinquinervia</i> - <i>Colliguaja odorifera</i> P32: Bosque espinoso mediterráneo interior de <i>Acacia caven</i> - <i>Prosopis chilensis</i> P33: Bosque espinoso mediterráneo andino de <i>Acacia caven</i> / <i>Baccharis paniculata</i> P41: Bosque esclerofilo mediterráneo andino de <i>Quillaja saponaria</i> - <i>Lithrea caustica</i>
grupo05	P111: Matorral bajo mediterráneo andino de <i>Chuquiraga oppositifolia</i> - <i>Nardophyllum lanatum</i> P112: Matorral bajo mediterráneo andino de <i>Laretia acaulis</i> - <i>Berberis empetrifolia</i> P38: Bosque esclerofilo mediterráneo andino de <i>Kageneckia angustifolia</i> / <i>Guindilia trinervis</i>
grupo06	P118: Herbazal mediterráneo andino de <i>Nastanthus spathulatus</i> - <i>Menonvillea spathulata</i> SV: Sin vegetación

Cuadro 5: Descripción de zonas del estudio.

#### 4.1.1. Descripción de los datos del estudio.

Se seleccionaron en inicialmente 6 variables para el estudio (Ver cuadro 6.). Por un lado, 5 de ellas se derivan de una variable llamada modelo de elevación (DEM) y la sexta corresponde al índice de humedad relativo en la superficie foliar (NDMI).

Cada una de las 6 variables fue sometida a un proceso de normalización u homogeneización para lograr que todas

Variable	Descripción
Eje__x	Coordenada en el eje x longitud (metros).
Eje__y	Coordenada en el eje y latitud (metros).
Asp__x	Grado de exposición este oeste (radianes).
Asp__y	Grado de exposición norte sur (radianes).
DEM	Modelo de elevación digital (metros).
NDMI	Índice de humedad de superficie foliar (%).
Slope	Pendiente (grados).
TWI	Índice de rugosidad (%).

Cuadro 6: Descripción de las variables del estudio.

tuvieran un mismo sistema de proyección, tamaño y resolución. Además, se incorporaron las coordenadas latitud y longitud como las 2 primeras variables comunes a todos los grupos para poder realizar un análisis espacial de las variables inicialmente seleccionadas, por lo tanto, pasamos a tener un total de 8 variables por grupo en vez de 6.

Además de la homogenización de las variables fue también necesario estandarizarlas, puesto que ellas se encuentran en unidades de medición y escalas de magnitud diferentes.

Grupo	Número y tamaño de imágenes en píxeles	Total en píxeles
grupo00	6 imágenes de 5.386 alto x 6.814 ancho	36.700.204
grupo01	6 imágenes de 2.190 alto x 2.143 ancho	4.693.170
grupo02	6 imágenes de 4.778 alto x 3.237 ancho	15.466.386
grupo03	6 imágenes de 4.868 alto x 1.598 ancho	7.779.064
grupo04	6 imágenes de 4.097 alto x 3.238 ancho	13.266.086
grupo05	6 imágenes de 5.022 alto x 2.592 ancho	13.017.024
grupo06	6 imágenes de 5.010 alto x 2.327 ancho	11.658.270

Cuadro 7: Tamaños de imágenes satelitales según grupo.

## 4.2. Desarrollo de un procedimiento automatizado de identificación de grupos.

### 4.2.1. Pre procesamiento de imágenes.

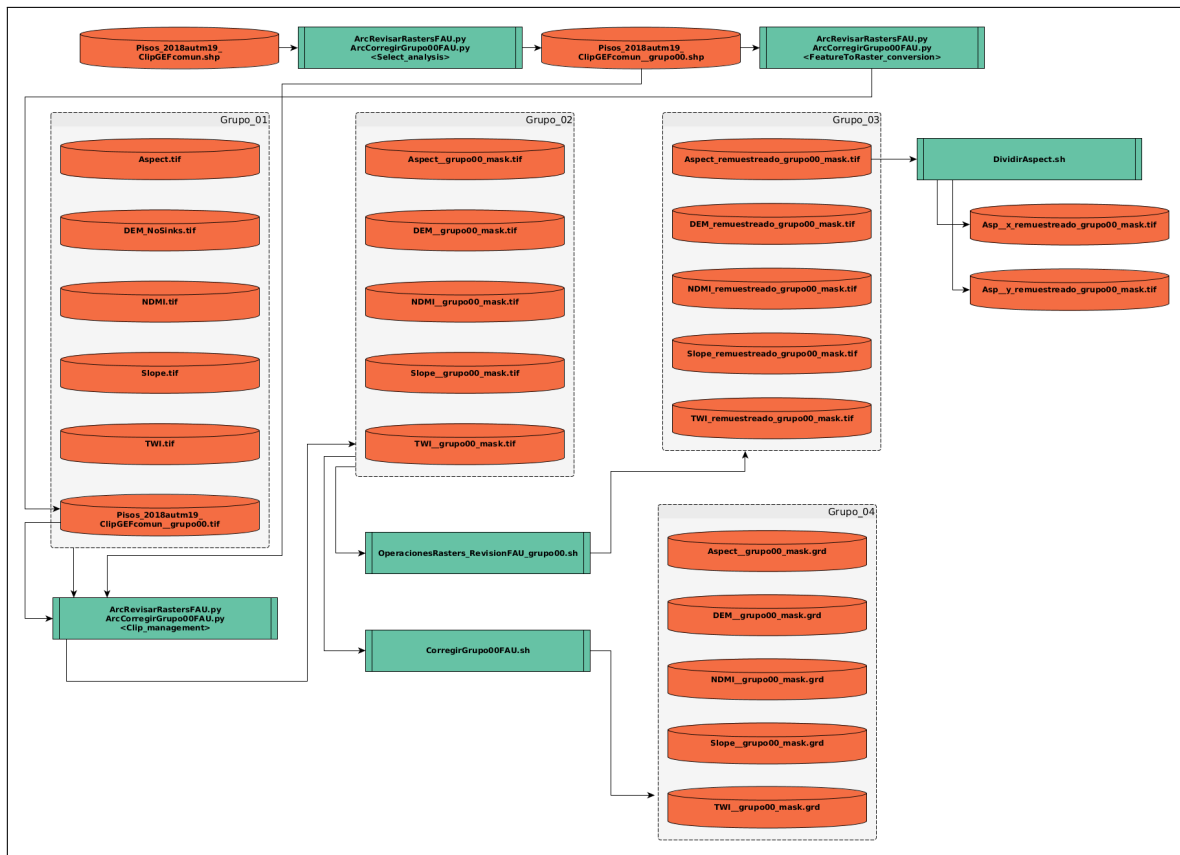


Gráfico 1: Proceso de homogenización de imágenes satelitales.

#### 4.2.2. Transformación de imágenes.

Esta sección consiste principalmente en la conversión de imágenes satelitales en archivos de tipo big.matrix, cuya diferencia principal es que están dispuestos en forma de columnas, mientras que cada imagen está en forma de matriz. El nombre big.matrix es confuso, porque no hace relación a una matriz propiamente tal, sino a que es un data.frame, pero uno que puede ser accedido con una tecnología de memoria compartida (shared memory), permitiendo que en la sesión de trabajo no se consuma memoria RAM en un monto equivalente al tamaño del archivo, sino que continúe la memoria compartida en el disco y tener una sesión completamente dedicada a los procesos que se ejecutarán y no a tener un archivo tan grande levantado en la sesión. Es uno de los pilares con los que este trabajo pudo ser llevado a cabo.

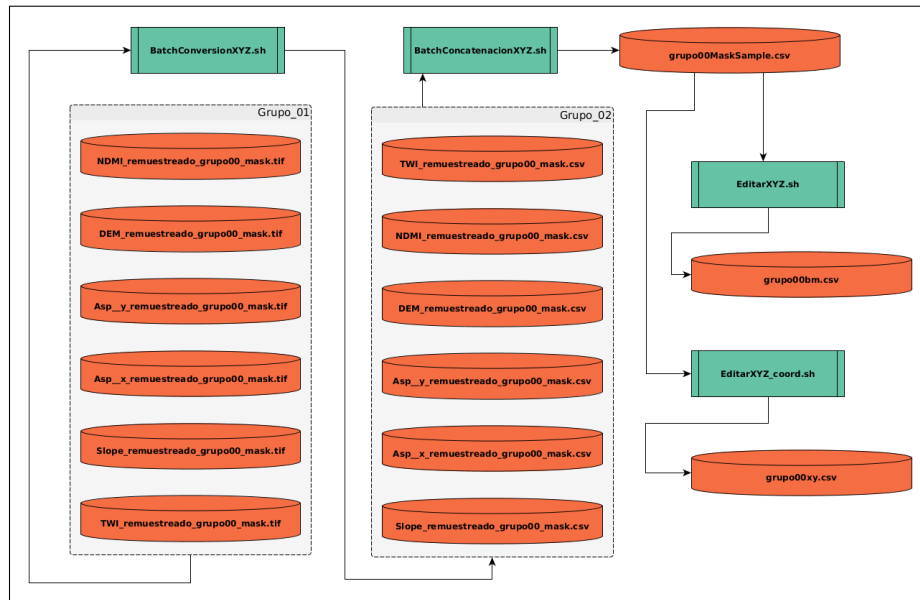


Gráfico 2: Proceso de conversión de imágenes satelitales.



### 4.2.3. Aplicación de algoritmo de clasificación.

En esta sección se utiliza el algoritmo de k-medias que realiza una estimación iterativa en la que calcula la suma cuadrática intra clusters al mismo tiempo que va asignando valores a cada observación como perteneciente a un grupo cluster potencial, vuelve a calcular corrigiendo el valor asignado a algunas observaciones y si obtiene una menor suma cuadrática intra, lo vuelve a re calcular hasta que la diferencia entre cada etapa no es mayor a un umbral de tolerancia, más allá del cual se detiene puesto que converge.

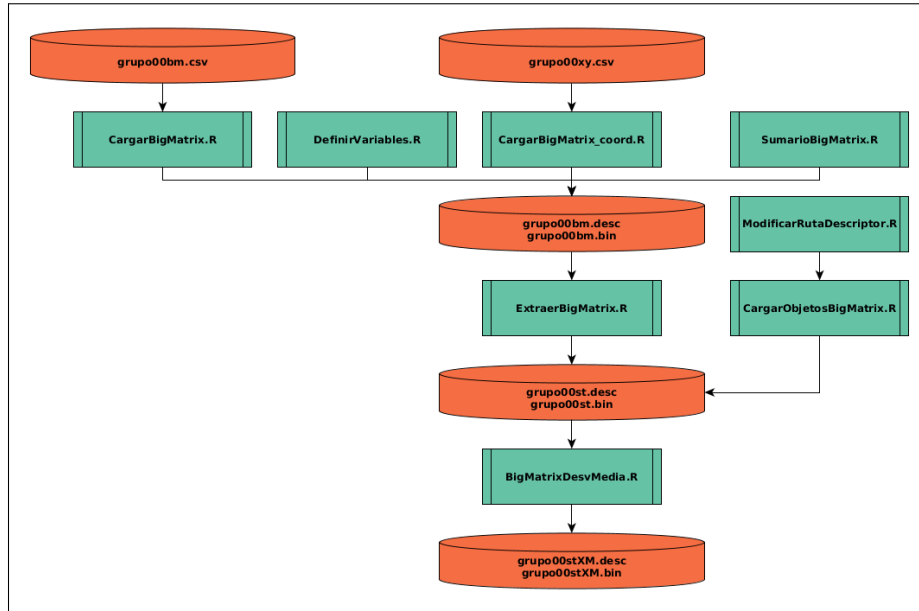


Gráfico 3: Proceso de creación de objetos de memoria compartida.

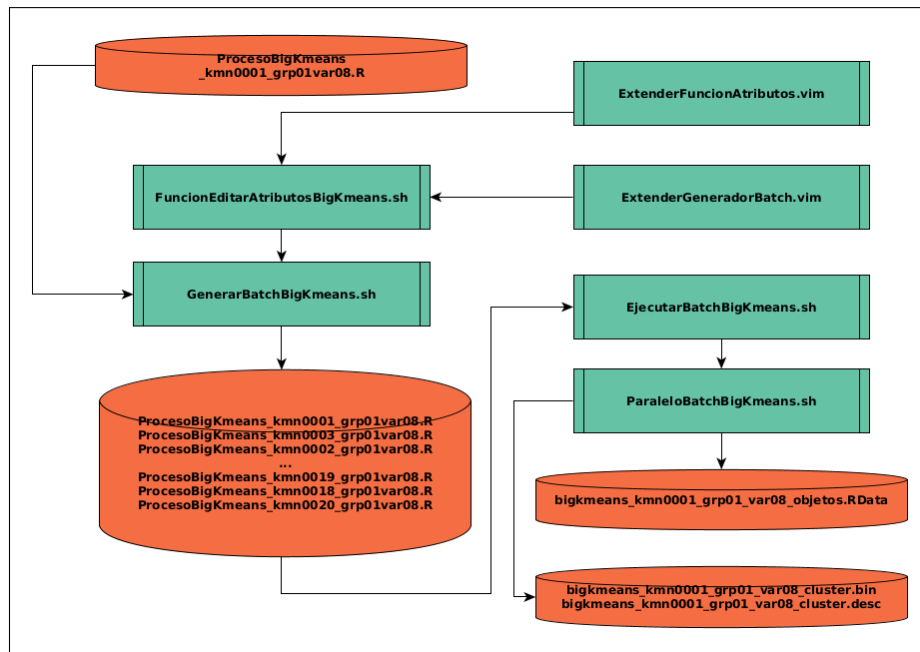


Gráfico 4: Proceso de ejecución de k-medias con big matrices.

#### 4.2.4. Estadígrafos de evaluación del algoritmo k-medias.

Esta sección consiste en calcular a partir de los resultados del procedimiento de clusterización, algunas métricas que permitan identificar el número óptimo de clusters, entre las cuales, tenemos,  $R^2$ , Elbow, Calinski-Harabasz y Silhouette. De las cuatro, se implementaron las tres primeras, mientras que la última tiene una limitación dada por el tamaño muestral lo que conlleva el cálculo de matrices de distancia de tamaño elevado.

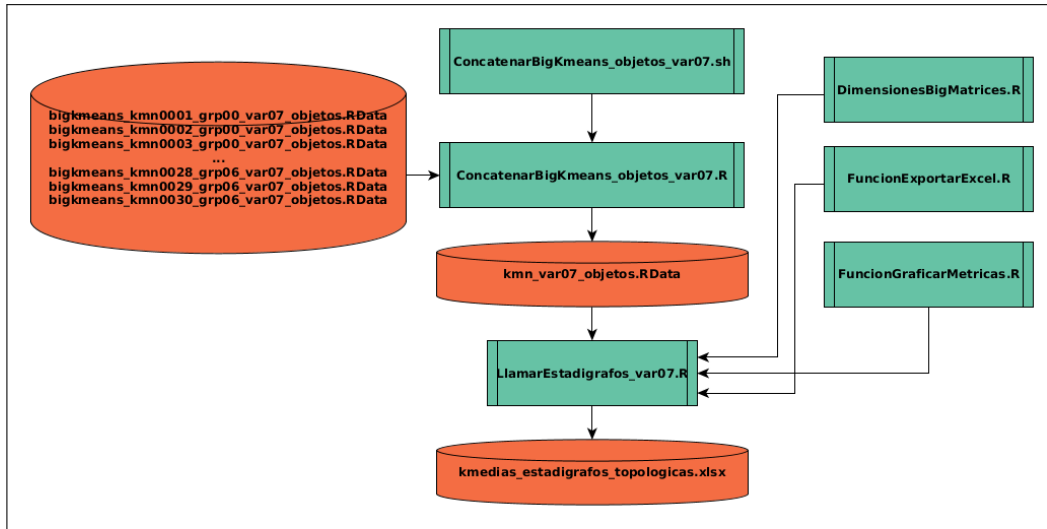


Gráfico 5: Proceso de extracción de estadígrafos de evaluación.

#### 4.2.5. Reconversión de big matrices en imágenes rasters.

Esta sección consiste en la generación de nuevas imágenes satelitales a partir de las columnas con los valores clusters obtenidos tras la aplicación de cada algoritmo. Cada vector columna está en un forma de tipo big.matrix y para convertirlo en imagen raster es necesario realizar varios pasos, de los cuales el más importante, es el de intersección con el vector original de la imagen que contiene las coordenadas espaciales, de manera de poder convertirlo en un objeto Ascii grid y posteriormente en un archivo Geotiff.

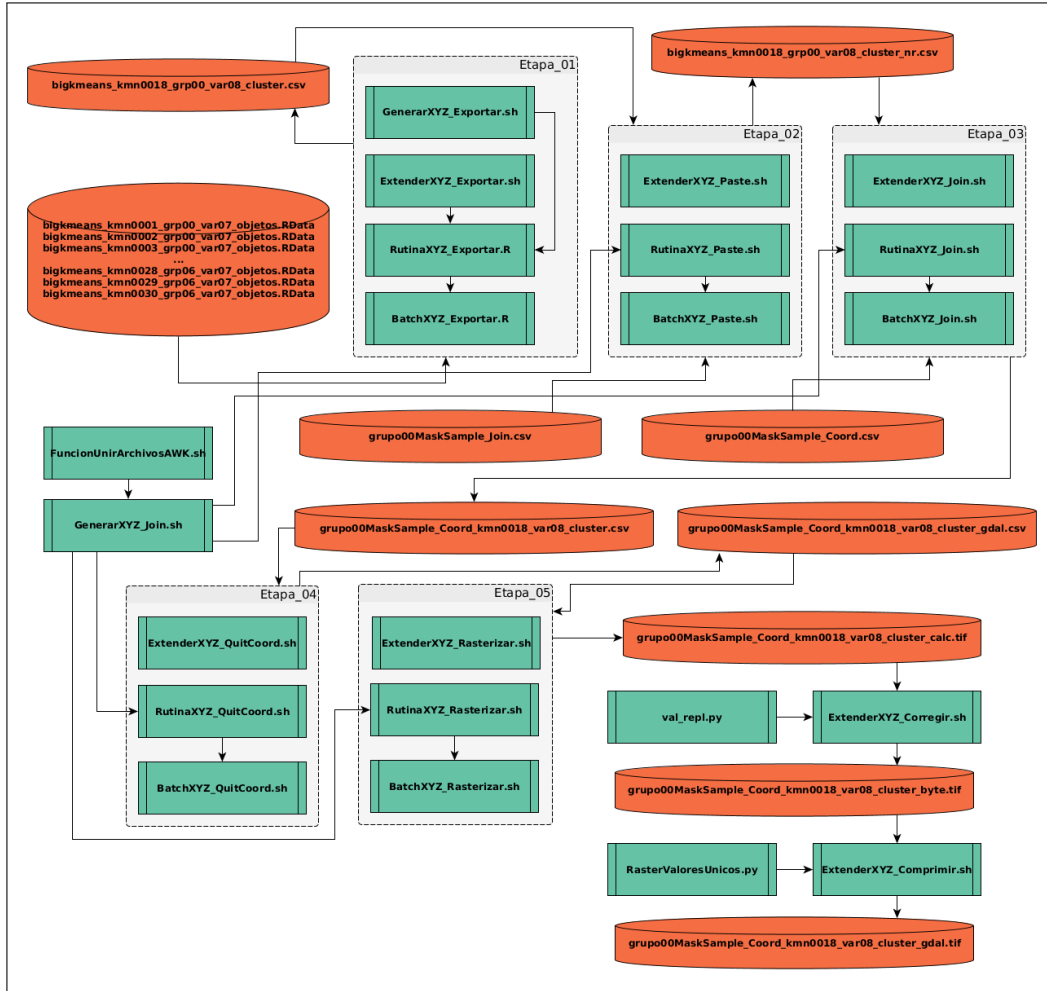


Gráfico 6: Proceso de rasterización desde big matrices.

#### 4.2.6. Caracterización de agrupamientos.

Esta etapa consiste en la caracterización de los clusters seleccionados para asignar nombres a los clusters elegidos. Para ello se calculan indicadores para cada uno de los clusters, tales como, promedio, desviación estándar, error estándar, máximo, mínimo, intervalo de confianza, los que a su vez son graficados para analizar en detalle como caracterizar cada agrupamiento y así ponerles un nombre distintivo a cada uno.

Además de caracterizar cada agrupamiento, tiene otro propósito que es permitir recodificar por tramos los valores de las variables para poder construir una matriz de confusión y poder así evaluar qué tan bien clasificó el algoritmo cluster cada una de las variables originales. En el gráfico 7 está el proceso correspondiente.

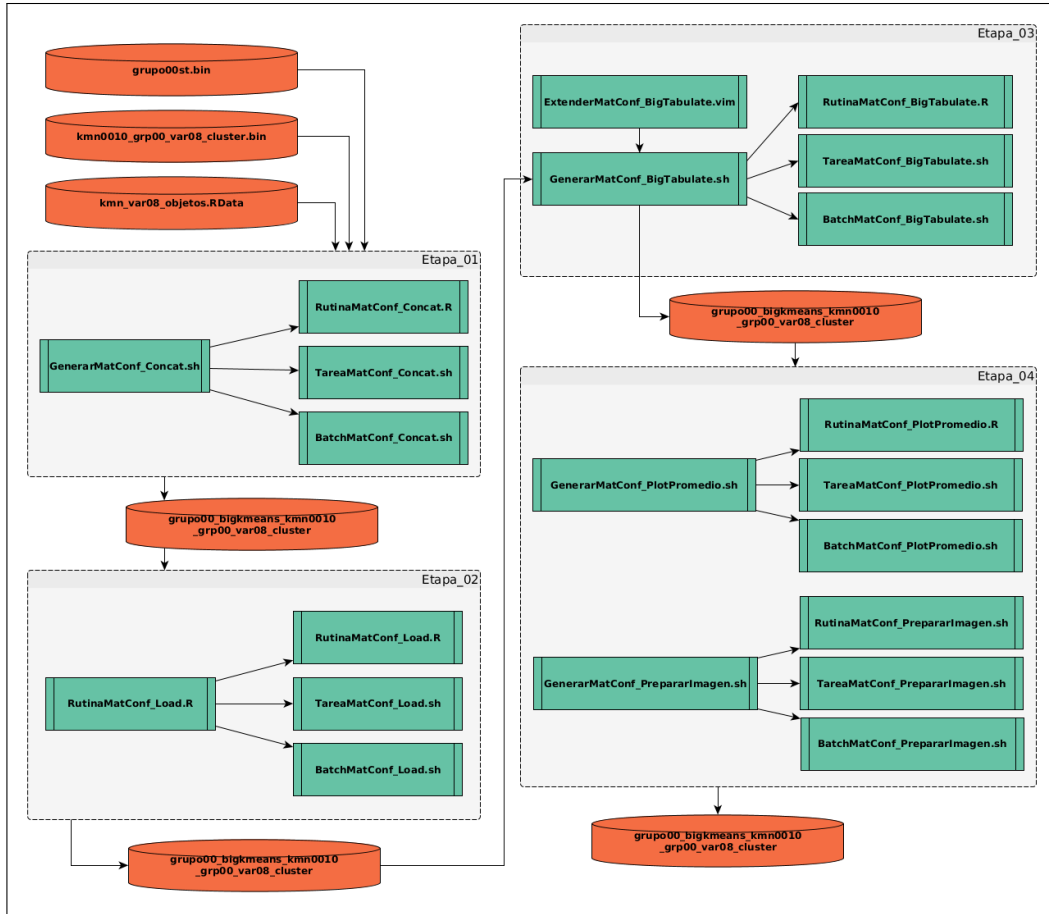


Gráfico 7: Proceso de caracterización de variables según cluster.

#### 4.2.7. Evaluación de precisión de clasificación por matriz de confusión.

Es de interés señalar que no solo para la identificación de zonas homogéneas con intereses de biodiversidad es posible usar técnicas estadísticas. Según los autores (Congalton and Green 2008) existen protocolos que permiten medir que tan exactas son las mediciones que se están haciendo a través del uso de información satelital. Esto forma parte de todo el manejo de información cartográfica espacial que sea medida de manera indirecta y que requiera entre otras cosas visitas en terreno para garantizar la exactitud de las áreas identificadas.

Además de tener una gran variedad de técnicas estadísticas en la actualidad, se mencionó inicialmente que los sensores satelitales, por su diseño, están expuestos a fuentes de perturbación lo que genera errores en los datos medidos. Tanto esto como cualquier procedimiento estadístico debe ser sometido a un protocolo de medición de precisión, lo que conlleva principalmente la puesta en operación de un plan de muestreo en terreno para la verificación de zonas de interés siguiendo también en este caso una metodología estadística como herramienta de planificación.

Si bien es cierto fue de interés para nuestra investigación identificar este conocimiento como una herramienta con gran potencial para nuestros análisis, es necesario tomar en cuenta que esto se enmarca en un problema que es también de orden transversal en muchas áreas de la información. Consiste principalmente en tener a disposición información previamente tipificada. En este caso, es necesario tener a lo menos una muestra de zonas representativas que hayan sido previamente categorizadas a un tipo específico de zona vegetacional o de propiedades topográficas según sea el caso. Como esto es muchas veces inviable en términos de presupuesto y tiempo, no se cuenta con áreas previamente clasificadas para el estudio por lo que no es posible aplicar al menos la herramienta de medición denominada matriz de confusión, puesto que requiere no solo de los grupos identificados mediante el algoritmo de k-medias, sino que además requiere de las áreas nominadas por un protocolo de muestreo en terreno o en última instancia de información secundaria asociada a las propiedades de las zonas de interés, pero no se cuenta con esta información.

Por su parte, el autor (G. F. (. F. Bonham-Carter and Bonham-Carter 1994) señala que para analizar la precisión del algoritmo de k-medias es posible hacer uso de un índice de validez interna, tal como,  $R^2$ , Elbow y el Caliski-Harabasz. El autor (Desgraupes 2018) ha implementado más de cincuenta fórmulas de validación interna y externa para medir la precisión del algoritmo de k-medias. No obstante, existen dos restricciones a la hora de utilizarlos con nuestros datos. Por un lado, no es compatible con el formato *bigmemory* y por el otro, varios de ellos requieren el cálculo de la matriz de distancias, lo que en nuestro estudio es inviable por tamaño de almacenamiento en disco duro.

## 5. Experimentos.

### 5.1. Resultados de las corridas de kmedias puntuales.

El cuadro 8 resume cada una de las 30 corridas de kmedias puntuales realizadas para el grupo del área total, y que a su vez fue calculado para las seis áreas restantes.

## 6. Conclusiones.

Un aspecto importante a considerar antes de analizar cada conclusión dice relación con las cosas que no pudieron ser implementadas en el estudio. Entre ellas la más importante talvez es el índice *Silhouette* que es muy conocido y que permite indentificar con mayor exactitud el número óptimo de clusters. La restricción principal está en la matriz de distancia y en toda librería donde existe una versión de este coeficiente es esta matriz la que se usa como insumo para su cálculo. Otra versión podría diseñarse que no requiera crear en disco duro dicha matriz.

Otro aspecto importante para casos futuros está el uso del algoritmo Batch-k-means e incluso Mini-batch-k-means.



**Anexos.**

**Resumen cuadros indicadores k-medias puntuales.**

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	59,13	0,02	136.874.124	46.397.107	183.271.232	25,32	74,68	7765.579,02	0,00	0,00	0,00	0,00
3	92,37	0,03	120.521.378	62.749.853	183.271.232	34,24	65,76	5963.797,77	-8,92	0,00	-0,12	0,00
4	335,24	0,09	109.496.152	73.775.079	183.271.232	40,25	59,75	5145.100,09	-6,02	2,91	-0,09	-0,23
5	547,98	0,15	100.943.071	82.328.160	183.271.232	44,92	55,08	4671.067,50	-4,67	1,35	-0,08	-0,15
6	657,43	0,18	94.058.475	89.212.756	183.271.232	48,68	51,32	4345.734,96	-3,76	0,91	-0,07	-0,13
7	1.491,45	0,41	88.744.804	94.526.427	183.271.232	51,58	48,42	4066.897,78	-2,90	0,86	-0,06	-0,17
8	1.137,71	0,32	83.507.445	99.763.786	183.271.232	54,44	45,56	3909.793,91	-2,86	0,04	-0,06	0,04
9	3.000,32	0,83	79.970.145	103.301.086	183.271.232	56,37	43,63	3699.058,33	-1,93	0,93	-0,04	-0,28
10	3.168,56	0,88	76.882.749	106.388.482	183.271.232	58,05	41,95	3522.308,09	-1,68	0,25	-0,04	-0,09
11	5.912,70	1,64	74.064.302	109.206.929	183.271.232	59,59	40,41	3377.889,10	-1,54	0,15	-0,04	-0,05
12	6.567,80	1,82	71.600.098	111.671.133	183.271.232	60,93	39,07	3248.169,76	-1,34	0,19	-0,03	-0,09
13	6.169,06	1,71	69.545.365	113.725.866	183.271.232	62,05	37,95	3121.863,41	-1,12	0,22	-0,03	-0,14
14	10.228,11	2,84	67.618.702	115.652.529	183.271.232	63,10	36,90	3014.040,00	-1,05	0,07	-0,03	-0,03
15	12.703,43	3,53	65.744.382	117.526.849	183.271.232	64,13	35,87	2925.192,53	-1,02	0,03	-0,03	0,00
16	16.276,08	4,52	64.170.514	119.100.717	183.271.232	64,99	35,01	2834.598,96	-0,86	0,16	-0,02	-0,14
17	15.806,34	4,39	62.622.222	120.649.009	183.271.232	65,83	34,17	2758.540,14	-0,84	0,01	-0,02	0,01
18	18.015,90	5,00	61.112.328	122.158.903	183.271.232	66,65	33,35	2693.713,26	-0,82	0,02	-0,02	-0,00
19	21.775,12	6,05	59.879.133	123.392.098	183.271.232	67,33	32,67	2622.667,95	-0,67	0,15	-0,02	-0,16
20	24.033,72	6,68	58.628.820	124.642.411	183.271.232	68,01	31,99	2563.333,06	-0,68	-0,01	-0,02	0,03
21	24.584,01	6,83	57.533.547	125.737.684	183.271.232	68,61	31,39	2503.330,81	-0,60	0,08	-0,02	-0,11
22	29.390,69	8,16	56.504.231	126.767.000	183.271.232	69,17	30,83	2447.427,64	-0,56	0,04	-0,02	-0,04
23	36.014,22	10,00	55.506.440	127.764.791	183.271.232	69,71	30,29	2396.895,08	-0,54	0,02	-0,02	-0,01
24	42.001,13	11,67	54.591.176	128.680.055	183.271.232	70,21	29,79	2347.820,14	-0,50	0,05	-0,02	-0,07
25	49.727,50	13,81	53.731.911	129.539.320	183.271.232	70,68	29,32	2301.240,14	-0,47	0,03	-0,02	-0,05
26	56.728,08	15,76	52.830.543	130.440.688	183.271.232	71,17	28,83	2262.516,91	-0,49	-0,02	-0,02	0,07
27	66.040,35	18,34	52.093.430	131.177.801	183.271.232	71,58	28,42	2218.747,41	-0,40	0,09	-0,01	-0,17
28	58.543,44	16,26	51.385.090	131.886.141	183.271.232	71,96	28,04	2177.720,14	-0,39	0,02	-0,01	-0,03
29	72.197,66	20,05	50.681.928	132.589.303	183.271.232	72,35	27,65	2140.430,38	-0,38	0,00	-0,01	0,01
30	62.139,75	17,26	50.009.666	133.261.565	183.271.232	72,71	27,29	2105.022,36	-0,37	0,02	-0,01	-0,03
31	84.855,91	23,57	49.386.987	133.884.244	183.271.232	73,05	26,95	2070.138,63	-0,34	0,03	-0,01	-0,06

Cuadro 8: Resumen ejecuciones kmedias puntuales para grupo00.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	3,23	0,00	13.345.359	2.810.112	16.155.471	17,39	82,61	425.229,05	0,00	0,00	0,00	0,00
3	9,79	0,00	11.875.453	4.280.018	16.155.471	26,49	73,51	363.910,56	-9,10	0,00	-0,11	0,00
4	20,36	0,01	10.857.206	5.298.265	16.155.471	32,80	67,20	328.490,93	-6,30	2,80	-0,09	-0,22
5	54,41	0,02	10.148.547	6.006.924	16.155.471	37,18	62,82	298.825,10	-4,39	1,92	-0,07	-0,24
6	77,58	0,02	9.538.509	6.616.962	16.155.471	40,96	59,04	280.179,76	-3,78	0,61	-0,06	-0,08
7	139,96	0,04	8.975.932	7.179.539	16.155.471	44,44	55,56	269.211,84	-3,48	0,29	-0,06	-0,02
8	122,04	0,03	8.529.050	7.626.421	16.155.471	47,21	52,79	257.958,71	-2,77	0,72	-0,05	-0,16
9	187,33	0,05	8.148.184	8.007.287	16.155.471	49,56	50,44	248.063,33	-2,36	0,41	-0,04	-0,10
10	242,84	0,07	7.808.765	8.346.706	16.155.471	51,66	48,34	239.837,97	-2,10	0,26	-0,04	-0,07
11	253,88	0,07	7.538.004	8.617.467	16.155.471	53,34	46,66	230.861,12	-1,68	0,42	-0,03	-0,17
12	233,29	0,06	7.286.417	8.869.054	16.155.471	54,90	45,10	223.459,02	-1,56	0,12	-0,03	-0,04
13	382,35	0,11	7.090.700	9.064.771	16.155.471	56,11	43,89	215.136,24	-1,21	0,35	-0,03	-0,20
14	449,40	0,12	6.905.016	9.250.455	16.155.471	57,26	42,74	208.104,69	-1,15	0,06	-0,03	-0,03
15	585,15	0,16	6.715.650	9.439.821	16.155.471	58,43	41,57	202.756,26	-1,17	-0,02	-0,03	0,05
16	638,28	0,18	6.565.220	9.590.251	16.155.471	59,36	40,64	196.659,90	-0,93	0,24	-0,02	-0,18
17	905,76	0,25	6.431.604	9.723.867	16.155.471	60,19	39,81	190.820,88	-0,83	0,10	-0,02	-0,09
18	1.140,46	0,32	6.296.888	9.858.583	16.155.471	61,02	38,98	185.979,74	-0,83	-0,01	-0,02	0,03
19	1.167,63	0,32	6.178.152	9.977.319	16.155.471	61,76	38,24	181.179,27	-0,73	0,10	-0,02	-0,10
20	1.355,67	0,38	6.062.287	10.093.184	16.155.471	62,48	37,52	176.955,35	-0,72	0,02	-0,02	-0,01
21	1.509,99	0,42	5.950.751	10.204.720	16.155.471	63,17	36,83	173.150,88	-0,69	0,03	-0,02	-0,02
22	1.619,45	0,45	5.850.722	10.304.749	16.155.471	63,78	36,22	169.368,93	-0,62	0,07	-0,02	-0,09
23	1.579,54	0,44	5.755.139	10.400.332	16.155.471	64,38	35,62	165.879,84	-0,59	0,03	-0,02	-0,03
24	1.869,30	0,52	5.658.455	10.497.016	16.155.471	64,97	35,03	162.878,93	-0,60	-0,01	-0,02	0,03
25	2.324,19	0,65	5.579.656	10.575.815	16.155.471	65,46	34,54	159.484,92	-0,49	0,11	-0,01	-0,17
26	2.474,35	0,69	5.503.089	10.652.382	16.155.471	65,94	34,06	156.359,57	-0,47	0,01	-0,01	-0,01
27	2.611,78	0,73	5.425.359	10.730.112	16.155.471	66,42	33,58	153.612,46	-0,48	-0,01	-0,01	0,03
28	3.065,29	0,85	5.352.610	10.802.861	16.155.471	66,87	33,13	150.950,04	-0,45	0,03	-0,01	-0,05
29	3.642,22	1,01	5.283.808	10.871.663	16.155.471	67,29	32,71	148.393,39	-0,43	0,02	-0,01	-0,04
30	3.840,16	1,07	5.217.662	10.937.809	16.155.471	67,70	32,30	145.975,46	-0,41	0,02	-0,01	-0,03
31	4.290,41	1,19	5.150.659	11.004.812	16.155.471	68,12	31,88	143.820,82	-0,41	-0,01	-0,01	0,03

Cuadro 9: Resumen ejecuciones kmedias puntuales para grupo01.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	30,95	0,01	40.822.653	10.581.274	51.403.927	20,58	79,42	1665.493,74	0,00	0,00	0,00	0,00
3	15,74	0,00	35.882.232	15.521.695	51.403.927	30,20	69,80	1389.747,45	-9,61	0,00	-0,12	0,00
4	78,78	0,02	32.905.775	18.498.152	51.403.927	35,99	64,01	1204.040,47	-5,79	3,82	-0,08	-0,31
5	104,88	0,03	30.566.987	20.836.940	51.403.927	40,54	59,46	1095.033,39	-4,55	1,24	-0,07	-0,14
6	141,53	0,04	28.763.152	22.640.775	51.403.927	44,04	55,96	1011.557,99	-3,51	1,04	-0,06	-0,17
7	388,54	0,11	27.339.640	24.064.287	51.403.927	46,81	53,19	942.616,17	-2,77	0,74	-0,05	-0,16
8	387,63	0,11	26.044.240	25.359.687	51.403.927	49,33	50,67	893.799,31	-2,52	0,25	-0,05	-0,04
9	737,10	0,20	24.948.643	26.455.284	51.403.927	51,47	48,53	851.689,50	-2,13	0,39	-0,04	-0,11
10	652,12	0,18	23.932.479	27.471.448	51.403.927	53,44	46,56	819.515,28	-1,98	0,15	-0,04	-0,03
11	801,96	0,22	22.963.296	28.440.631	51.403.927	55,33	44,67	795.812,29	-1,89	0,09	-0,04	-0,01
12	892,39	0,25	22.227.002	29.176.925	51.403.927	56,76	43,24	766.781,32	-1,43	0,45	-0,03	-0,21
13	1.241,61	0,34	21.556.237	29.847.690	51.403.927	58,06	41,94	741.416,13	-1,30	0,13	-0,03	-0,06
14	1.276,18	0,35	20.945.354	30.458.573	51.403.927	59,25	40,75	718.760,07	-1,19	0,12	-0,03	-0,06
15	1.389,44	0,39	20.369.011	31.034.916	51.403.927	60,37	39,63	699.291,04	-1,12	0,07	-0,03	-0,03
16	2.152,66	0,60	19.928.803	31.475.124	51.403.927	61,23	38,77	676.550,62	-0,86	0,26	-0,02	-0,21
17	3.108,33	0,86	19.490.306	31.913.621	51.403.927	62,08	37,92	657.571,06	-0,85	0,00	-0,02	0,02
18	3.256,57	0,90	19.125.538	32.278.389	51.403.927	62,79	37,21	637.902,69	-0,71	0,14	-0,02	-0,15
19	5.223,07	1,45	18.768.803	32.635.124	51.403.927	63,49	36,51	620.699,34	-0,69	0,02	-0,02	-0,00
20	5.998,33	1,67	18.450.913	32.953.014	51.403.927	64,11	35,89	603.988,57	-0,62	0,08	-0,02	-0,09
21	6.306,23	1,75	18.139.270	33.264.657	51.403.927	64,71	35,29	589.166,71	-0,61	0,01	-0,02	-0,00
22	7.754,21	2,15	17.823.926	33.580.001	51.403.927	65,33	34,67	576.451,72	-0,61	-0,01	-0,02	0,03
23	8.309,48	2,31	17.541.107	33.862.820	51.403.927	65,88	34,12	563.830,12	-0,55	0,06	-0,02	-0,09
24	10.834,12	3,01	17.259.013	34.144.914	51.403.927	66,42	33,58	552.696,84	-0,55	0,00	-0,02	0,01
25	10.118,28	2,81	16.985.255	34.418.672	51.403.927	66,96	33,04	542.519,66	-0,53	0,02	-0,02	-0,01
26	10.431,64	2,90	16.718.737	34.685.190	51.403.927	67,48	32,52	533.218,50	-0,52	0,01	-0,02	-0,01
27	14.681,19	4,08	16.467.795	34.936.132	51.403.927	67,96	32,04	524.288,79	-0,49	0,03	-0,02	-0,04
28	15.440,54	4,29	16.250.750	35.153.177	51.403.927	68,39	31,61	514.792,13	-0,42	0,07	-0,01	-0,12
29	16.641,93	4,62	16.047.355	35.356.572	51.403.927	68,78	31,22	505.606,99	-0,40	0,03	-0,01	-0,05
30	16.886,61	4,69	15.850.214	35.553.713	51.403.927	69,17	30,83	496.999,77	-0,38	0,01	-0,01	-0,02
31	19.011,31	5,28	15.667.095	35.736.832	51.403.927	69,52	30,48	488.551,80	-0,36	0,03	-0,01	-0,06

Cuadro 10: Resumen ejecuciones kmedias puntuales para grupo02.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	3,36	0,00	6.446.626	1.331.149	7.777.776	17,11	82,89	200.751,76	0,00	0,00	0,00	0,00
3	3,17	0,00	5.827.826	1.949.949	7.777.776	25,07	74,93	162.648,92	-7,96	0,00	-0,10	0,00
4	4,60	0,00	5.257.250	2.520.525	7.777.776	32,41	67,59	155.372,88	-7,34	0,62	-0,10	0,02
5	16,99	0,00	4.911.626	2.866.149	7.777.776	36,85	63,15	141.832,97	-4,44	2,89	-0,07	-0,33
6	47,58	0,01	4.607.718	3.170.057	7.777.776	40,76	59,24	133.774,82	-3,91	0,54	-0,06	-0,06
7	40,35	0,01	4.384.897	3.392.878	7.777.776	43,62	56,38	125.377,74	-2,86	1,04	-0,05	-0,22
8	53,66	0,01	4.173.421	3.604.354	7.777.776	46,34	53,66	119.949,78	-2,72	0,15	-0,05	-0,00
9	54,84	0,02	3.993.439	3.784.336	7.777.776	48,66	51,34	115.163,38	-2,31	0,40	-0,04	-0,11
10	75,46	0,02	3.843.918	3.933.857	7.777.776	50,58	49,42	110.551,15	-1,92	0,39	-0,04	-0,13
11	134,94	0,04	3.743.534	4.034.241	7.777.776	51,87	48,13	104.771,01	-1,29	0,63	-0,03	-0,30
12	177,60	0,05	3.644.368	4.133.407	7.777.776	53,14	46,86	100.242,93	-1,27	0,02	-0,03	0,01
13	392,74	0,11	3.553.253	4.224.522	7.777.776	54,32	45,68	96.323,07	-1,17	0,10	-0,03	-0,06
14	364,79	0,10	3.470.622	4.307.153	7.777.776	55,38	44,62	92.810,98	-1,06	0,11	-0,02	-0,07
15	379,33	0,11	3.398.397	4.379.378	7.777.776	56,31	43,69	89.488,96	-0,93	0,13	-0,02	-0,11
16	368,41	0,10	3.331.407	4.446.368	7.777.776	57,17	42,83	86.505,80	-0,86	0,07	-0,02	-0,05
17	545,36	0,15	3.262.106	4.515.669	7.777.776	58,06	41,94	84.112,86	-0,89	-0,03	-0,02	0,06
18	529,70	0,15	3.199.210	4.578.565	7.777.776	58,87	41,13	81.845,64	-0,81	0,08	-0,02	-0,07
19	677,88	0,19	3.135.558	4.642.217	7.777.776	59,69	40,31	79.964,20	-0,82	-0,01	-0,02	0,03
20	809,49	0,22	3.080.145	4.697.630	7.777.776	60,40	39,60	78.038,87	-0,71	0,11	-0,02	-0,11
21	771,17	0,21	3.027.144	4.750.631	7.777.776	61,08	38,92	76.285,98	-0,68	0,03	-0,02	-0,03
22	816,60	0,23	2.976.174	4.801.601	7.777.776	61,73	38,27	74.690,37	-0,66	0,03	-0,02	-0,02
23	1.071,56	0,30	2.930.561	4.847.214	7.777.776	62,32	37,68	73.092,78	-0,59	0,07	-0,02	-0,09
24	930,89	0,26	2.886.859	4.890.916	7.777.776	62,88	37,12	71.613,03	-0,56	0,02	-0,01	-0,03
25	1.013,46	0,28	2.841.021	4.936.754	7.777.776	63,47	36,53	70.389,91	-0,59	-0,03	-0,02	0,06
26	1.178,61	0,33	2.800.637	4.977.138	7.777.776	63,99	36,01	69.109,41	-0,52	0,07	-0,01	-0,10
27	1.296,31	0,36	2.762.211	5.015.564	7.777.776	64,49	35,51	67.895,87	-0,49	0,03	-0,01	-0,03
28	1.686,84	0,47	2.725.309	5.052.466	7.777.776	64,96	35,04	66.753,99	-0,47	0,02	-0,01	-0,03
29	1.684,41	0,47	2.686.754	5.091.021	7.777.776	65,46	34,54	65.791,80	-0,50	-0,02	-0,01	0,06
30	1.611,85	0,45	2.654.175	5.123.600	7.777.776	65,87	34,13	64.714,30	-0,42	0,08	-0,01	-0,14
31	1.937,77	0,54	2.620.357	5.157.418	7.777.776	66,31	33,69	63.782,68	-0,43	-0,02	-0,01	0,05

Cuadro 11: Resumen ejecuciones kmedias puntuales para grupo03.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	28,40	0,01	34.105.768	10.015.727	44.121.495	22,70	77,30	1619.626,40	0,00	0,00	0,00	0,00
3	14,69	0,00	29.706.522	14.414.973	44.121.495	32,67	67,33	1338.110,89	-9,97	0,00	-0,13	0,00
4	34,06	0,01	27.078.234	17.043.261	44.121.495	38,63	61,37	1157.100,59	-5,96	4,01	-0,09	-0,31
5	63,74	0,02	24.954.345	19.167.150	44.121.495	43,44	56,56	1059.037,38	-4,81	1,14	-0,08	-0,11
6	154,96	0,04	23.249.769	20.871.726	44.121.495	47,31	52,69	990.215,15	-3,86	0,95	-0,07	-0,13
7	381,86	0,11	21.917.413	22.204.082	44.121.495	50,32	49,68	931.219,50	-3,02	0,84	-0,06	-0,16
8	513,05	0,14	20.786.424	23.335.071	44.121.495	52,89	47,11	884.486,09	-2,56	0,46	-0,05	-0,10
9	556,47	0,15	19.956.122	24.165.373	44.121.495	54,77	45,23	834.808,72	-1,88	0,68	-0,04	-0,23
10	750,22	0,21	19.102.728	25.018.767	44.121.495	56,70	43,30	802.578,50	-1,93	-0,05	-0,04	0,07
11	807,09	0,22	18.398.506	25.722.989	44.121.495	58,30	41,70	771.078,06	-1,60	0,34	-0,04	-0,14
12	1.018,61	0,28	17.714.355	26.407.140	44.121.495	59,85	40,15	747.416,56	-1,55	0,05	-0,04	0,01
13	1.226,22	0,34	17.142.088	26.979.407	44.121.495	61,15	38,85	723.347,04	-1,30	0,25	-0,03	-0,13
14	1.304,37	0,36	16.649.340	27.472.155	44.121.495	62,26	37,74	700.021,81	-1,12	0,18	-0,03	-0,11
15	2.086,47	0,58	16.240.292	27.881.203	44.121.495	63,19	36,81	676.314,64	-0,93	0,19	-0,02	-0,15
16	2.336,94	0,65	15.823.543	28.297.952	44.121.495	64,14	35,86	657.535,27	-0,94	-0,02	-0,03	0,04
17	3.049,28	0,85	15.474.138	28.647.357	44.121.495	64,93	35,07	638.141,65	-0,79	0,15	-0,02	-0,14
18	2.568,56	0,71	15.145.584	28.975.911	44.121.495	65,67	34,33	620.670,43	-0,74	0,05	-0,02	-0,04
19	3.633,36	1,01	14.848.847	29.272.648	44.121.495	66,35	33,65	604.025,95	-0,67	0,07	-0,02	-0,08
20	4.283,89	1,19	14.563.387	29.558.108	44.121.495	66,99	33,01	589.141,19	-0,65	0,03	-0,02	-0,02
21	4.706,07	1,31	14.293.560	29.827.935	44.121.495	67,60	32,40	575.455,09	-0,61	0,04	-0,02	-0,04
22	6.710,44	1,86	14.010.165	30.111.330	44.121.495	68,25	31,75	564.450,66	-0,64	-0,03	-0,02	0,07
23	6.437,04	1,79	13.773.859	30.347.636	44.121.495	68,78	31,22	552.338,19	-0,54	0,11	-0,02	-0,15
24	7.770,77	2,16	13.553.701	30.567.794	44.121.495	69,28	30,72	540.800,15	-0,50	0,04	-0,02	-0,05
25	8.440,89	2,34	13.357.238	30.764.257	44.121.495	69,73	30,27	529.269,51	-0,45	0,05	-0,01	-0,09
26	8.611,17	2,39	13.154.404	30.967.091	44.121.495	70,19	29,81	519.334,85	-0,46	-0,01	-0,02	0,05
27	9.171,97	2,55	12.963.017	31.158.478	44.121.495	70,62	29,38	509.864,76	-0,43	0,03	-0,01	-0,04
28	11.153,69	3,10	12.786.052	31.335.443	44.121.495	71,02	28,98	500.603,30	-0,40	0,03	-0,01	-0,06
29	9.225,18	2,56	12.606.194	31.515.301	44.121.495	71,43	28,57	492.422,01	-0,41	-0,01	-0,01	0,03
30	11.483,25	3,19	12.445.743	31.675.752	44.121.495	71,79	28,21	484.023,04	-0,36	0,04	-0,01	-0,10
31	10.690,68	2,97	12.288.079	31.833.416	44.121.495	72,15	27,85	476.250,96	-0,36	0,01	-0,01	-0,00

Cuadro 12: Resumen ejecuciones kmedias puntuales para grupo04.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	27,09	0,01	24.893.796	5.047.867	29.941.663	16,86	83,14	758.931,63	0,00	0,00	0,00	0,00
3	14,75	0,00	22.172.786	7.768.877	29.941.663	25,95	74,05	655.682,68	-9,09	0,00	-0,11	0,00
4	57,53	0,02	20.289.460	9.652.203	29.941.663	32,24	67,76	593.499,43	-6,29	2,80	-0,08	-0,22
5	67,52	0,02	18.747.783	11.193.880	29.941.663	37,39	62,61	558.671,15	-5,15	1,14	-0,08	-0,11
6	126,21	0,04	17.751.676	12.189.987	29.941.663	40,71	59,29	514.019,12	-3,33	1,82	-0,05	-0,30
7	193,30	0,05	16.850.040	13.091.623	29.941.663	43,72	56,28	484.648,12	-3,01	0,32	-0,05	-0,04
8	307,88	0,09	16.055.973	13.885.690	29.941.663	46,38	53,62	462.400,13	-2,65	0,36	-0,05	-0,07
9	413,52	0,11	15.493.751	14.447.912	29.941.663	48,25	51,75	436.258,16	-1,88	0,77	-0,04	-0,26
10	574,32	0,16	14.938.198	15.003.465	29.941.663	50,11	49,89	417.672,38	-1,86	0,02	-0,04	0,02
11	973,22	0,27	14.496.724	15.444.939	29.941.663	51,58	48,42	398.750,41	-1,47	0,38	-0,03	-0,18
12	959,38	0,27	14.104.000	15.837.663	29.941.663	52,90	47,10	382.068,10	-1,31	0,16	-0,03	-0,08
13	1.470,05	0,41	13.759.499	16.182.164	29.941.663	54,05	45,95	366.806,71	-1,15	0,16	-0,02	-0,10
14	1.802,34	0,50	13.437.452	16.504.211	29.941.663	55,12	44,88	353.605,45	-1,08	0,07	-0,02	-0,04
15	2.306,21	0,64	13.137.234	16.804.429	29.941.663	56,12	43,88	341.960,62	-1,00	0,07	-0,02	-0,05
16	2.173,81	0,60	12.849.920	17.091.743	29.941.663	57,08	42,92	331.878,32	-0,96	0,04	-0,02	-0,02
17	2.967,93	0,82	12.605.288	17.336.375	29.941.663	57,90	42,10	321.713,76	-0,82	0,14	-0,02	-0,13
18	3.206,39	0,89	12.371.968	17.569.695	29.941.663	58,68	41,32	312.651,48	-0,78	0,04	-0,02	-0,03
19	3.374,30	0,94	12.144.879	17.796.784	29.941.663	59,44	40,56	304.691,02	-0,76	0,02	-0,02	-0,01
20	3.375,50	0,94	11.924.004	18.017.659	29.941.663	60,18	39,82	297.650,35	-0,74	0,02	-0,02	-0,01
21	5.076,26	1,41	11.717.007	18.224.656	29.941.663	60,87	39,13	291.069,22	-0,69	0,05	-0,02	-0,05
22	4.447,37	1,24	11.532.009	18.409.654	29.941.663	61,49	38,51	284.514,81	-0,62	0,07	-0,02	-0,09
23	5.764,16	1,60	11.364.463	18.577.200	29.941.663	62,04	37,96	278.094,28	-0,56	0,06	-0,01	-0,08
24	5.632,68	1,56	11.200.008	18.741.655	29.941.663	62,59	37,41	272.298,35	-0,55	0,01	-0,01	-0,00
25	6.099,68	1,69	11.041.807	18.899.856	29.941.663	63,12	36,88	266.925,62	-0,53	0,02	-0,01	-0,02
26	5.728,03	1,59	10.894.895	19.046.768	29.941.663	63,61	36,39	261.722,63	-0,49	0,04	-0,01	-0,06
27	6.273,79	1,74	10.762.243	19.179.420	29.941.663	64,06	35,94	256.532,42	-0,44	0,05	-0,01	-0,08
28	7.507,48	2,09	10.633.517	19.308.146	29.941.663	64,49	35,51	251.699,69	-0,43	0,01	-0,01	-0,02
29	7.077,44	1,97	10.508.363	19.433.300	29.941.663	64,90	35,10	247.192,99	-0,42	0,01	-0,01	-0,02
30	7.291,42	2,03	10.388.805	19.552.858	29.941.663	65,30	34,70	242.900,95	-0,40	0,02	-0,01	-0,03
31	8.878,99	2,47	10.275.963	19.665.700	29.941.663	65,68	34,32	238.752,59	-0,38	0,02	-0,01	-0,05

Cuadro 13: Resumen ejecuciones kmedias puntuales para grupo05.

k	tiempo.seg	tiempo.hrs	tot.withinss	tot.betweenss	totss	R2	Elbow	CH	dElbow	ddElbow	tElbow	ttElbow
2	39,68	0,01	29.608.525	5.145.866	34.754.392	14,81	85,19	755.024,99	0,00	0,00	0,00	0,00
3	22,43	0,01	26.066.431	8.687.960	34.754.392	25,00	75,00	723.978,67	-10,19	0,00	-0,12	0,00
4	50,77	0,01	23.603.138	11.151.253	34.754.392	32,09	67,91	684.151,18	-7,09	3,10	-0,09	-0,21
5	73,03	0,02	21.845.718	12.908.673	34.754.392	37,14	62,86	641.762,92	-5,06	2,03	-0,07	-0,21
6	144,74	0,04	20.484.565	14.269.826	34.754.392	41,06	58,94	605.258,82	-3,92	1,14	-0,06	-0,16
7	200,52	0,06	19.453.928	15.300.463	34.754.392	44,02	55,98	569.462,49	-2,97	0,95	-0,05	-0,19
8	335,74	0,09	18.720.193	16.034.198	34.754.392	46,14	53,86	531.566,85	-2,11	0,85	-0,04	-0,25
9	414,11	0,12	18.040.564	16.713.827	34.754.392	48,09	51,91	503.100,46	-1,96	0,16	-0,04	-0,04
10	745,88	0,21	17.403.302	17.351.089	34.754.392	49,92	50,08	481.250,72	-1,83	0,12	-0,04	-0,03
11	853,04	0,24	16.888.492	17.865.899	34.754.392	51,41	48,59	459.571,12	-1,48	0,35	-0,03	-0,16
12	1.007,72	0,28	16.412.319	18.342.072	34.754.392	52,78	47,22	441.371,58	-1,37	0,11	-0,03	-0,05
13	1.263,56	0,35	15.961.493	18.792.898	34.754.392	54,07	45,93	426.243,26	-1,30	0,07	-0,03	-0,03
14	1.459,19	0,41	15.564.353	19.190.038	34.754.392	55,22	44,78	412.021,46	-1,14	0,15	-0,02	-0,09
15	1.750,10	0,49	15.161.328	19.593.063	34.754.392	56,38	43,62	401.010,17	-1,16	-0,02	-0,03	0,04
16	2.316,62	0,64	14.795.294	19.959.097	34.754.392	57,43	42,57	390.700,73	-1,05	0,11	-0,02	-0,07
17	2.827,24	0,79	14.457.985	20.296.406	34.754.392	58,40	41,60	381.161,92	-0,97	0,08	-0,02	-0,06
18	3.407,99	0,95	14.154.678	20.599.713	34.754.392	59,27	40,73	371.903,51	-0,87	0,10	-0,02	-0,08
19	3.668,64	1,02	13.894.582	20.859.809	34.754.392	60,02	39,98	362.334,95	-0,75	0,12	-0,02	-0,12
20	4.707,43	1,31	13.640.948	21.113.443	34.754.392	60,75	39,25	353.898,50	-0,73	0,02	-0,02	-0,01
21	5.610,68	1,56	13.398.282	21.356.109	34.754.392	61,45	38,55	346.226,84	-0,70	0,03	-0,02	-0,03
22	7.095,89	1,97	13.186.393	21.567.998	34.754.392	62,06	37,94	338.362,44	-0,61	0,09	-0,02	-0,11
23	6.260,25	1,74	12.984.859	21.769.532	34.754.392	62,64	37,36	331.059,99	-0,58	0,03	-0,02	-0,03
24	7.613,72	2,11	12.797.202	21.957.189	34.754.392	63,18	36,82	324.079,28	-0,54	0,04	-0,01	-0,05
25	8.706,15	2,42	12.617.060	22.137.331	34.754.392	63,70	36,30	317.594,62	-0,52	0,02	-0,01	-0,03
26	8.632,18	2,40	12.454.943	22.299.448	34.754.392	64,16	35,84	311.121,18	-0,47	0,05	-0,01	-0,09
27	11.466,79	3,19	12.304.715	22.449.676	34.754.392	64,60	35,40	304.847,24	-0,43	0,03	-0,01	-0,06
28	13.009,60	3,61	12.147.258	22.607.133	34.754.392	65,05	34,95	299.447,36	-0,45	-0,02	-0,01	0,06
29	12.907,64	3,59	12.001.362	22.753.029	34.754.392	65,47	34,53	294.149,12	-0,42	0,03	-0,01	-0,06
30	14.824,02	4,12	11.859.317	22.895.074	34.754.392	65,88	34,12	289.201,93	-0,41	0,01	-0,01	-0,01
31	15.056,25	4,18	11.719.612	23.034.779	34.754.392	66,28	33,72	284.620,56	-0,40	0,01	-0,01	-0,00

Cuadro 14: Resumen ejecuciones kmedias puntuales para grupo06.



Resumen gráficos indicadores k-medias puntuales.

## Referencias.

Enlaces consultados.

Medición de precisión en mapas temáticos.

- [What is the best approach to evaluate an unsupervised classification](#)
- [Evaluation measures of goodness or validity of clustering without having truth](#)
- [How to measure classification accuracy in the absence of ground truth](#)

Creación de matrices de distancia de gran tamaño.

- [R: distm with Big Memory](#)

Algoritmos usados en la tesis.

- [K-Means tutorial - MLpack](#)

## Artículos consultados.

Bonham-Carter, G. F. (Graeme Francis), and Graeme Bonham-Carter. 1994. *Geographic Information Systems for Geoscientists: Modelling with GIS*. Elsevier.

Congalton, Russell G., and Kass Green. 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition (Mapping Science)*. 2nd ed. Mapping Science. CRC Press. <http://gen.lib.rus.ec/book/index.php?md5=CE1869245D6EBBBC1F67CB64E949347E>.

Desgraupes, Bernard. 2018. *clusterCrit: Clustering Indices*. <https://CRAN.R-project.org/package=clusterCrit>.

Lloyd, S. 1982. "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory* 28 (2): 129–37. doi:[10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).

Luebert, Federico, and Patricio Plischoff. 2006. *Sinopsis Bioclimática Y Vegetacional de Chile*. Editorial Universitaria.