

Reporte práctica profesional: Replicación metodología SDMs de Arclim

Francisca Muñoz Narbona

2025-02-12

Introducción

ARClím, el Atlas de Riesgos Climáticos para Chile, es un proyecto del Ministerio del Medio Ambiente del Gobierno de Chile, desarrollado por el Centro de Investigación del Clima y la Resiliencia (CR2) y el Centro de Cambio Global (CCG) de la Pontificia Universidad Católica de Chile, en colaboración con diversas instituciones nacionales e internacionales. Su objetivo es desarrollar un conjunto de mapas de riesgos asociados al cambio climático en Chile, abarcando múltiples sectores con cobertura nacional y detalle a nivel comunal o puntual, convirtiéndose en una herramienta fundamental para la planificación de políticas públicas y la implementación de medidas de adaptación.

En la actualidad, el mundo enfrenta una triple crisis ambiental: el cambio climático, la contaminación y la pérdida de biodiversidad. En respuesta a este desafío, la Guía de Cambio Climático (2023), elaborada por el Servicio de Evaluación Ambiental (SEA), busca fortalecer el Sistema de Evaluación de Impacto Ambiental (SEIA) para mejorar la identificación y mitigación de los efectos ambientales de proyectos en el territorio nacional. Un aspecto clave dentro del SEIA es la identificación de los objetos de protección ambiental dentro de las áreas de influencia (AI) de los proyectos, lo que implica describir la evolución esperada de los componentes ambientales en ausencia de intervención, considerando escenarios climáticos extremos y posibles aumentos en los riesgos de origen natural.

Un elemento relevante en la caracterización del AI de la biota es el Mapa de Especies de AR-Clím, el cual permite visualizar y descargar modelos de distribución de flora y fauna nativa y endémica de Chile continental. Estos modelos están disponibles para dos períodos: histórico reciente (1980-2010) y futuro cercano (2035-2065) bajo el escenario RCP8.5. Aunque ARClím proporciona información para diversas especies, incluidas arbóreas, arbustivas, herbáceas, anfibios, reptiles, aves e insectos, su cobertura no es completa, dejando sin modelar una parte significativa de la biodiversidad del país.

En este contexto, el objetivo de Cienciambiental es complementar la información disponible en ARClím mediante la replicación de su metodología, con el propósito de generar modelos de distribución para aquellas especies que no han sido incluidas en la plataforma oficial. Para ello, se considera información adicional proporcionada por la empresa y se aplican herramientas de modelamiento que buscan mantener la mayor coherencia posible con los métodos empleados en ARClím.

Metodología

i) ARClím

En la plataforma ARClím se generan modelos de distribución de especies para Chile continental, abarcando 440 especies de flora y 110 especies de fauna, para dos períodos: histórico reciente (1980-2010) y futuro cercano (2035-2065), bajo el escenario RCP8.5. Cada modelo expresa la probabilidad de presencia de una especie en una grilla de 5x5 km², con valores entre 0 (ausencia de la especie) y 1 (máxima probabilidad de presencia), utilizando el modelo Maxent. El desarrollo de estos modelos fue liderado por el equipo del profesor Patricio Pliscoff en la Pontificia Universidad Católica de Chile.

Para la modelación, se emplearon cinco variables climáticas anuales: evapotranspiración promedio, precipitación acumulada, insolación diaria promedio, temperatura mínima diaria promedio y temperatura máxima diaria promedio. Estas variables fueron obtenidas a partir de simulaciones climáticas realizadas para el proyecto ARClím, utilizando 35 modelos globales integrados entre 1970 y 2070. Posteriormente, fueron sometidas a un proceso de escalamiento espacial (downscaling), ajustándolas a una grilla nacional con resolución de 5x5 km².

Por otro lado, las observaciones de ocurrencia georreferenciada de las especies fueron compiladas a partir del proyecto Fondecyt 1181677, dirigido por P. Pliscoff. Estos datos permitieron entrenar los modelos y establecer relaciones entre la presencia de las especies y las condiciones ambientales.

ii) Cienciambiental - Práctica profesional

Con el objetivo de replicar la metodología utilizada en ARClím, Cienciambiental implementó el algoritmo Maxent mediante el paquete dismo en RStudio. Para ello, se emplearon las mismas variables climáticas utilizadas en ARClím, obtenidas a través de su API, mientras que las observaciones de ocurrencia de las especies fueron obtenidas desde la plataforma GBIF.

Los datos de ocurrencia fueron sometidos a un proceso de filtrado para mejorar su calidad y confiabilidad. Se aplicaron los siguientes criterios de selección:

- Eliminación de registros con problemas geoespaciales.
- Consideración únicamente de fuentes confiables.
- Exclusión de datos con incertidumbre en coordenadas superior a 1000 m.
- Uso exclusivo de taxonomía validada.

- Eliminación de registros provenientes de museos, zoológicos y otras fuentes no representativas de la distribución natural de las especies.
- Remoción de registros duplicados.

Tras la ejecución del modelo Maxent, se generaron los mapas de distribución de especies y se realizó una comparación estadística entre los modelos obtenidos por Cienciamambiental y los disponibles en ARClím, para las mismas especies. Para ello, se transformaron los mapas de probabilidad en mapas binarios (considerando como presencia los valores >0.1). A partir de estos mapas, se calcularon:

- Mapa de diferencias entre ambas modelaciones (predicciones hechas por dismo, menos predicciones de ARClím).
- Análisis de correlación entre predicciones.
- Matriz de confusión.
- Índice Kappa para evaluar la similitud de las predicciones.

Los análisis fueron realizados tanto para el período histórico como futuro, considerando únicamente las regiones de Chile donde la especie puede distribuirse. Además, se llevó a cabo una revisión de todas las capas en QGIS para verificar la coherencia espacial de los modelos.

El código desarrollado para este análisis fue almacenado en un [repositorio en GitHub](#) con el fin de garantizar la reproducibilidad y permitir la colaboración con el equipo de datos de la empresa.

El proyecto se estructuró en siete scripts en R, organizados de la siguiente manera:

- (1) Obtención de datos climáticos desde ARClím.
- (2) Descarga de datos de ocurrencia desde GBIF.
- (3) Implementación del modelo Maxent.
- (4) Comparación visual entre los modelos generados con dismo y los modelos de ARClím.
- (5) Recorte de las zonas con probabilidad de presencia y comparación visual entre ambos modelos.
- (6) Análisis estadístico de los modelos, considerando mapas recortados.
- (7) Descarga de datos y capas generadas. (Opcional).

Resultados

Cienciamambiental dispone principalmente de datos de especies de la zona norte de Chile, por lo que una de las especies modeladas en este estudio es *Adesmia atacamensis*, la cual también se encuentra disponible en la plataforma ARClím. A partir de la base de datos GBIF, se recopilaron 86 registros originales de ocurrencia de la especie, los cuales fueron sometidos a un proceso de filtrado, reduciendo el conjunto final a 34 observaciones. En la Figura 1 se presentan los mapas de distribución generados mediante ambas metodologías, tanto para el período histórico como para el período futuro.

En el caso de *Adesmia atacamensis*, se observa que su distribución se concentra principalmente en el norte y centro de Chile, mientras que la zona sur presenta una probabilidad nula de presencia, de acuerdo con los modelos y la bibliografía existente. Para facilitar la comparación estadística entre los modelos, se definió un área de interés y se generaron mapas binarios (considerando presencia con valores >0.1). Este procedimiento evita que la gran cantidad de píxeles con probabilidad cero de presencia sesgue la evaluación de similitud entre los modelos. En la Figura 2 se presenta la transformación de los mapas de probabilidad a mapas binarios.

En la Figura 3, se muestran los mapas de diferencias entre las predicciones generadas por Cienciamambiental y ARClím, para ambos períodos de estudio. Las zonas en azul corresponden a áreas donde ARClím predice presencia de la especie y Cienciamambiental no, mientras que las zonas en rojo representan el caso inverso. Por su parte, las zonas en blanco indican regiones donde ambas metodologías coincidieron en sus predicciones.

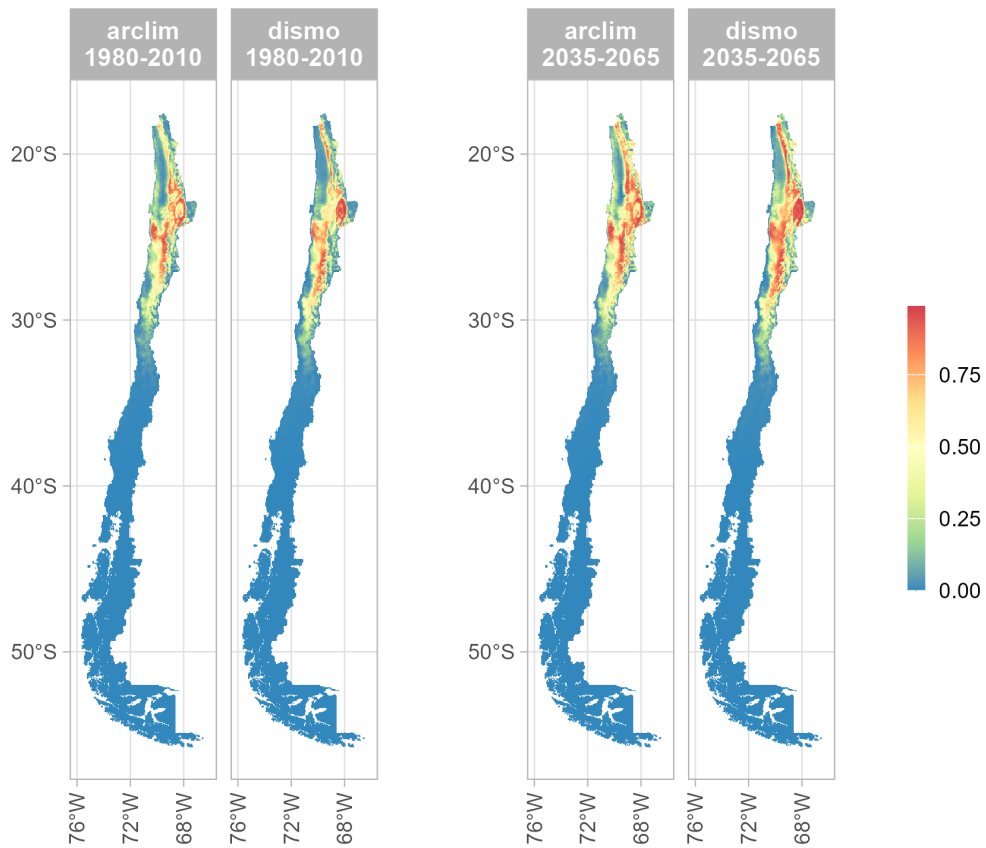


Figure 1: Mapas de distribución de *Adesmia atacamensis*, elaborados por ARClím y Cienciamambiental (dismo). A la izquierda, para el período histórico reciente y a la derecha para el período futuro bajo el esecnario RCP8.5.

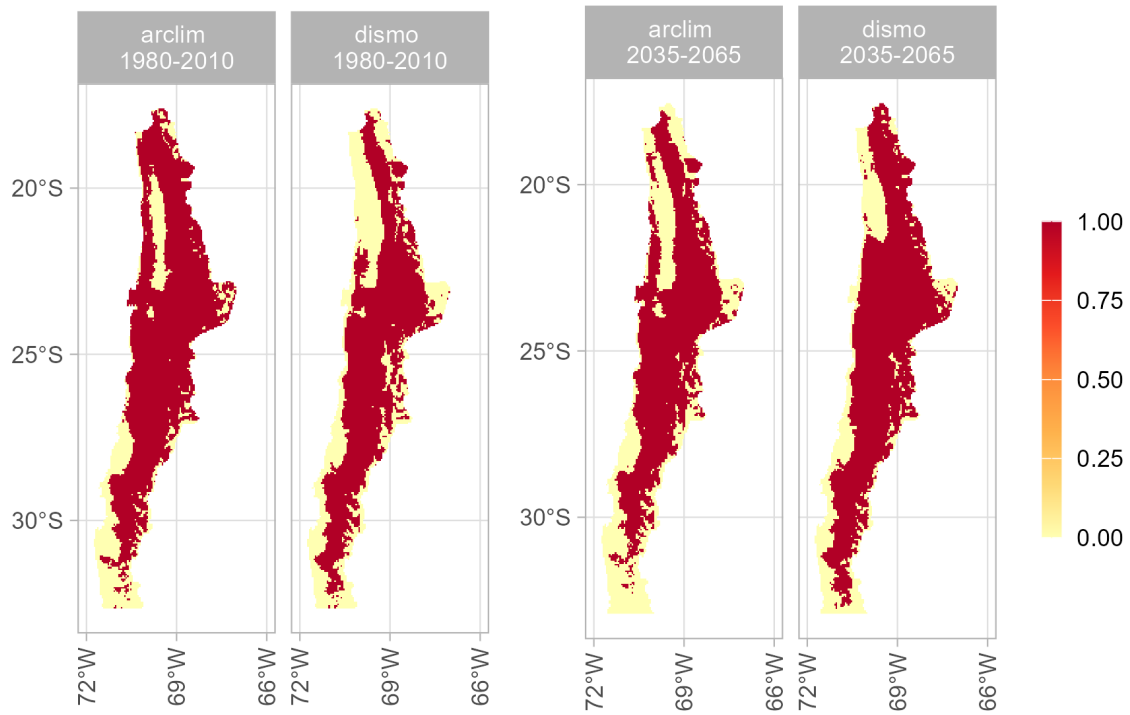


Figure 2: Mapas binarios de distribución de *Adesmia atacamensis* en zona norte y centro de Chile, lugares en donde hay una probabilidad de distribución de la especie mayor a 0.1. A la izquierda, para el período histórico reciente y a la derecha para el período futuro bajo el escenario RCP8.5.

Finalmente, los resultados cuantitativos reflejan una correlación de 0.613 para el período histórico (1980-2010) y 0.663 para el período futuro (2035-2065). La matriz de confusión para cada período se presenta en la Tabla 1 y Tabla 2, donde se evalúa la compatibilidad de predicción a nivel de píxeles en los mapas recortados a la zona de interés. Asimismo, el índice de Kappa, que mide la concordancia entre ambas modelaciones, alcanzó valores de 0.58 para el período histórico y 0.64 para el período futuro, lo que indica una moderada similitud entre ambos enfoques metodológicos.

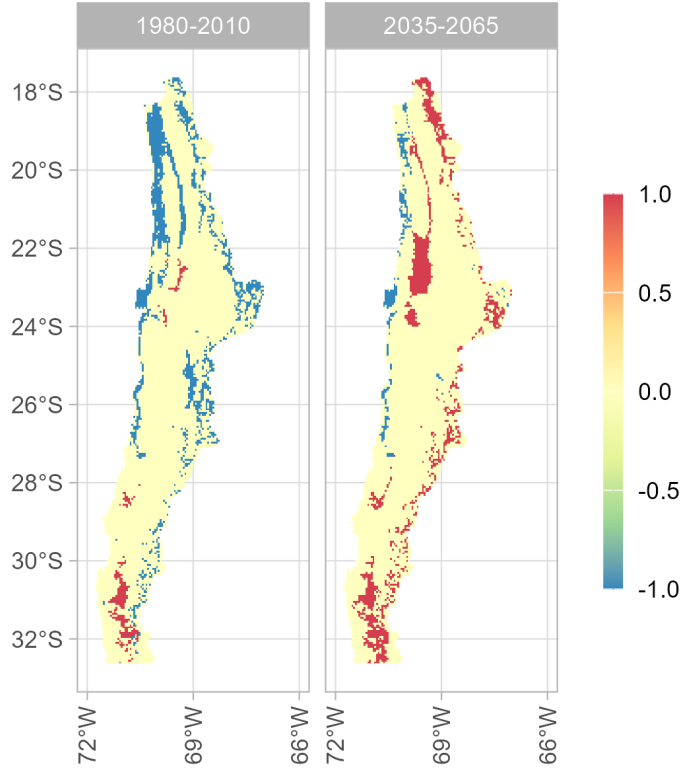


Figure 3: Mapas binarios que indican la diferencia entre predicciones hechas por Cienciamiental y ARClm para ambos períodos de tiempo. Valores de (1) y (-1) indican que hay diferencias entre modelaciones, mientras que valores de (0) indican que son iguales. Un valor de (1) indica que el modelamiento de Cienciamiental sí predice presencia, y el de ARClm no. El valor de (-1) lo contrario.

Table 1: Matriz de Confusión para el Escenario Histórico Reciente (1980-2010).

	FALSE	TRUE
FALSE	2473	1768
TRUE	287	6876

Table 2: Matriz de Confusión para el Escenario Futuro (2035-2065), escenario RCP8.5.

	FALSE	TRUE
FALSE	2718	303
TRUE	1486	7044

Discusión

La replicación de la metodología utilizada en ARClím para la modelación de distribución de *Adesmia atacamensis* en Cienciamambiental permitió evaluar la similitud entre ambos modelos y detectar posibles factores que influyen en las diferencias observadas.

En términos generales, los resultados muestran una correlación de **0.613 para el período histórico (1980-2010) y 0.663 para el período futuro (2035-2065)**. Estos valores indican una correspondencia moderada entre las modelaciones, lo que sugiere que la metodología implementada en Cienciamambiental es capaz de aproximarse en cierta medida a los modelos de ARClím. Sin embargo, la **matriz de confusión y el índice Kappa (0.58 y 0.64, respectivamente)** evidencian que existen discrepancias que deben ser consideradas.

Uno de los aspectos más notorios al analizar las figuras generadas es la presencia de **patrones atípicos en los mapas de distribución**, como una marcada línea horizontal observada en la Figura 1 en el mapa generado con el paquete **dismo**. Esto motivó una revisión detallada de las capas climáticas utilizadas en el modelamiento, identificando que algunas variables, en particular la **precipitación acumulada**, podrían estar introduciendo estas estructuras anómalas. Este hallazgo resalta la necesidad de realizar un análisis más profundo de las capas climáticas antes de su utilización en los modelos, ya que **errores o inconsistencias en sus valores pueden afectar significativamente los resultados**. Sin embargo, estas son las mismas variables empleadas en la fuente oficial de ARClím, por lo que su impacto en las predicciones debería ser consistente en ambos modelos.

Otro factor que podría estar influyendo en las diferencias observadas es la **falta de información detallada sobre la metodología de ARClím**. Si bien se utilizaron las mismas variables climáticas y el mismo modelo Maxent, **no se dispone de los datos de ocurrencia empleados en ARClím**, lo que impide conocer si se aplicaron filtros adicionales o si hubo ajustes específicos en los parámetros del modelo. Además, el uso exclusivo de **cinco variables climáticas** podría ser insuficiente para capturar todos los factores ecológicos que determinan la distribución de la especie. Es posible que otras variables, como **factores edáficos, topográficos o de uso de suelo**, tengan un impacto importante en la distribución de *A. atacamensis* y no estén siendo consideradas en el modelo actual.

Otro aspecto a considerar es la influencia de **la resolución espacial en los resultados**. En ARClím, los modelos emplean una resolución de **5x5 km**, lo que puede ser adecuado para ciertas especies, pero insuficiente para otras con distribuciones más localizadas. En este sentido, sería relevante evaluar si realizar un **downscaling** o incorporar datos de mayor resolución permitiría capturar mejor la distribución de *A. atacamensis* y reducir las diferencias entre los modelos de Cienciamambiental y ARClím.

Además, aunque se realizó una comparación entre los modelos generados y los de ARClím, **no se llevó a cabo una validación independiente de los resultados**. Sería ideal contar con datos de monitoreo en terreno o registros adicionales que permitan contrastar las predicciones con observaciones reales. Esto ayudaría a determinar cuál de los modelos se acerca más a

la distribución real de la especie y evitar depender exclusivamente de la comparación entre modelaciones basadas en metodologías distintas.

Otro aspecto clave que aún no se ha abordado en profundidad es la **incertidumbre en las predicciones**. En este tipo de modelamiento, no solo es relevante conocer la distribución proyectada de una especie, sino también **cuánta confianza se tiene en dichas predicciones**. Cuantificar la incertidumbre permitiría identificar qué variables tienen mayor impacto en el modelo y en qué zonas del mapa las predicciones son menos confiables. Incorporar análisis de sensibilidad y métricas de incertidumbre en estudios futuros permitiría mejorar la interpretación de los resultados y fortalecer la robustez del modelo.

Finalmente, sería pertinente evaluar si **otros enfoques de modelamiento podrían mejorar los resultados**. Si bien MaxEnt es una metodología ampliamente utilizada en la modelación de distribución de especies, otros algoritmos como **Random Forest, GAMs o redes neuronales** podrían proporcionar estimaciones más precisas y robustas. La implementación de estos métodos permitiría comparar su desempeño con el enfoque actual y determinar si ofrecen mejoras en la calidad de las predicciones.

Conclusión

La iniciativa de Cienciambiental de replicar la metodología de ARClím para generar mapas de distribución de especies fue un esfuerzo valioso para comprender el proceso de modelamiento y evaluar su reproducibilidad. Se logró generar mapas comparables y evaluar estadísticamente la similitud entre ambos enfoques. Sin embargo, los resultados indican que existen **diferencias significativas en las predicciones**, lo que resalta la importancia de analizar con mayor profundidad los **datos de entrada**, en particular, la selección y calidad de las capas climáticas y la disponibilidad de datos de ocurrencia.

A pesar de las limitaciones, este estudio representa un **primer paso importante** en la exploración de modelos de distribución de especies dentro de Cienciambiental. Se continuará trabajando en esta línea, pero con enfoques más avanzados que incorporen un **mayor número de variables ambientales** y técnicas de modelamiento más sofisticadas.

En definitiva, replicar la metodología de ARClím es un desafío que requiere no solo una correcta implementación técnica, sino también un acceso detallado a la información utilizada en la modelación original. Para futuras investigaciones, será clave contar con una base de datos más robusta, analizar con mayor detalle los insumos de entrada y explorar la integración de otras metodologías complementarias.

Líneas futuras de trabajo

- **Revisión más detallada de los datos de entrada**, incluyendo datos de ocurrencia y capas climáticas.

- **Evaluación del impacto de la resolución espacial** en la precisión de los modelos.
- **Comparación con otros algoritmos de modelamiento**, como Random Forest, GAMs o redes neuronales.
- **Cuantificación de la incertidumbre** en las predicciones para mejorar la interpretación de los mapas.
- **Exploración de validaciones independientes** con datos de monitoreo en terreno.
- **Aplicaciones prácticas de los modelos**, por ejemplo, en evaluación ambiental y conservación.

Con estas mejoras, Cienciamambiental podrá fortalecer su capacidad para generar modelos de distribución de especies más precisos y confiables, contribuyendo a la toma de decisiones en conservación y planificación ambiental.