

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-Universidad Católica de Chile), con la colaboración de otras instituciones nacionales e internacionales. El objetivo general de ARClím es desarrollar un conjunto de mapas de riesgos relacionados con el cambio climático para Chile empleando un marco conceptual común y una base de datos consistente. Incluye diversos sectores con cobertura nacional y detalle comunal o puntual, convirtiéndose así en una herramienta importante para el diseño de políticas públicas y la implementación de medidas de adaptación (MMA, 2025).

Actualmente el mundo se encuentra enfrentando una triple crisis ambiental: el cambio climático, la contaminación y la pérdida de biodiversidad. Por ello, la Guía de Cambio Climático (2023), desarrollada por el Servicio de Evaluación Ambiental (SEA), busca perfeccionar el Sistema de Evaluación de Impacto Ambiental (SEIA) y fortalecer el cumplimiento de los objetivos para aproximar a un desarrollo sustentable del país. Un paso importante en el SEIA es la descripción de los objetos de protección ambiental, receptores de impactos en las áreas de influencia (AI). De este modo, ha de describirse la posible evolución de los componentes ambientales en una situación sin proyecto de impacto, incluyendo la posibilidad de situaciones climáticas extremas, e investigar el posible aumento de riesgos de origen natural, considerando el escenario más desfavorable.

Un antecedente que se debe considerar para la descripción del AI de la biota es lo señalado por el Mapa de Especies de la plataforma ARClím, que permite visualizar y descargar modelos de distribución de especies para especies de flora y fauna nativas y endémicas presentes en Chile continental, para un período histórico reciente (1980-2010) y un periodo futuro cercano (2035-2065), bajo el escenario RCP8.5. Si bien se encuentran disponibles mapas para especies arbóreas, herbáceas, arbustivas, anfibios, marsupiales, reptiles, aves, insectos y mamíferos, **sólo hay algunas especies modeladas y no todas.**

Es así, que el **objetivo de Cienciambiental es complementar con la información de las especies faltantes de la manera más similar posible a la oficial, replicando la metodología utilizada en la plataforma ARClím**, considerando información adicional proporcionada por la empresa.

Metodología

i) ARClím

En la plataforma ARClím se realizan modelos de distribución de especies para Chile continental para 440 especies de flora y 110 especies de fauna, para un periodo histórico reciente (1980-2010) y un periodo futuro cercano (2035-2065), bajo el peor escenario RCP8.5. En cada caso se muestra la probabilidad de presencia por especie en cada píxel de $5 \times 5 \text{ km}^2$, con valores entre 0 (no hay probabilidad) y 1 (máxima probabilidad) obtenidos a partir del modelo Max-ent. Este modelo emplea un conjunto de variables ambientales y observaciones de ocurrencia georreferenciada, y el modelamiento fue desarrollado por el grupo del profesor Patricio Plischoff de la Pontificia Universidad Católica de Chile.

En el trabajo de ARClím, se emplearon cinco variables climáticas anuales: evapotranspiración promedio, precipitación acumulada, promedio de insolación diaria, promedio de temperatura mínima diaria y promedio de temperatura máxima diaria. Estas variables fueron obtenidas a partir de simulaciones climáticas realizadas para el proyecto ARClím (empleando 35 modelos globales interados entre 1970 y 2070), estas variables fueron sometidas a un proceso de escalamiento (downscaling) llevándolas a una grilla de cobertura nacional, pero con resolución de $5 \times 5 \text{ km}^2$. Por otro lado, los datos de observaciones de ocurrencia georreferencia de las especies, fueron compiladas a partir de los resultados del proyecto Fondecyt 1181677, liderado por P. Plischoff.

ii) Cienciambiental

Para cumplir el objetivo de Cienciambiental y replicar de mejor manera posible la metodología descrita en ARClím para modelos de distribución de especies, se utiliza el paquete Dismo en R para utilizar el modelo Maxent. Respecto a las variables climáticas, se utilizan las mismas proporcionadas en ARClím, obtenidas a través de su API, y sobre los datos de ocurrencia georreferencia, se utiliza como fuente la plataforma GBIF. Ambos datos son obtenidos y manipulados a través de Rstudio. Sólo a los datos de ocurrencia de les aplican filtros para que sean de mejor calidad y confiables, es decir: sin problemas geoespaciales, de fuentes confiables, con incertidumbre de coordenadas menor a 1000m, taxonomía aceptada, se eliminan datos de museos/zoológicos/etc, y sin datos repetidos.

Una vez corrido el modelo Maxent con el paquete Dismo y obtenidos los mapas de distribución, se realiza una comparación estadística entre estos mapas y los disponibles en la plataforma ARClím, para las mismas especies. A partir de la transformación a mapas binarias (probabilidad de presencia mayor a 0.1), se realiza una nuevo mapa que refleja las diferencias de

probabilidades entre ambas fuentes (de Cienciamambiental y ARClím), un análisis de correlación, una matriz de confusión y se obtiene el índice kappa. Todo lo anterior permite evaluar la similitud entre mapas de distribución de especies, se realiza tanto para tiempo histórico como para futuro, y sólo en las regiones de Chile en donde la especie puede distribuirse. Además, se hace un revisión en Qgis de todas las capas obtenidas.

Todo el proyecto es subido a un [repositorio en GitHub](#) para una mejor reproducibilidad del código y compartir el trabajo con el equipo de datos de la empresa. El proyecto cuenta con siete scripts de R:

- (1) Para obtener datos climáticos de ARClím.
- (2) Para obtener datos de ocurrencia de GBIF.
- (3) Realizar el modelo Maxent.
- (4) Comparación visual entre modelos obtenidos con paquete Dismo y los disponibles en ARClím.
- (5) Recorte de las zonas en donde sí hay probabilidad de que distribuya la especie y comparación visual entre ambos modelos.
- (6) Comparación estadística entre modelos, considerando mapas recortados.
- (7) Descarga de datos y capas de mapas obtenidos. *Opcional*.

Resultados

Cienciamambiental cuenta con muchos datos de especies de la zona norte de Chile, así que una de las especies modeladas es *Adesmia atacamensis*, la cual también se encuentra disponible en la plataforma ARClím. Desde GBIF se obtienen 86 registros originales de ocurrencia de la especie, y con los filtros aplicados disminuyen a 34 observaciones.

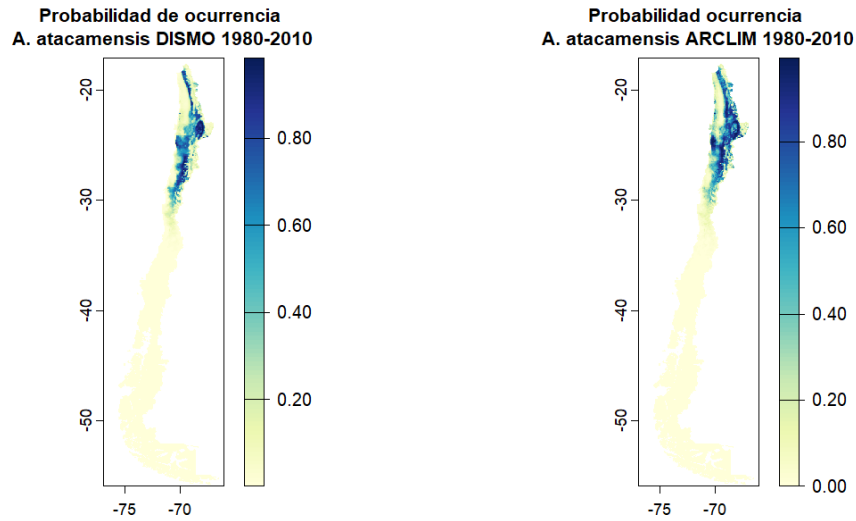


Figure 1: Comparación visual entre mapas de distribución de especies DISMO y ARCLIM.

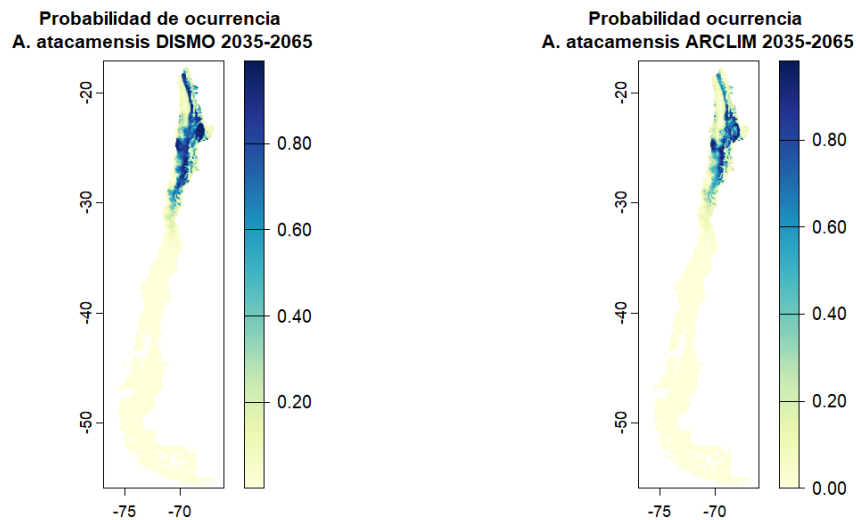


Figure 2: Comparación visual entre mapas de distribución de especies DISMO y ARCLIM.

En el caso de *Adesmia atacamensis* la distribución de la especie es sólo en el centro norte y centro de Chile, siendo la zona sur un área cero probable de distribución (tanto por los modelos como por bibliografía existente). Por ello, se selecciona un área de interés y se transforma el mapa a binario para comparar estadísticamente los mapas hechos por cienciambiental y arclim.

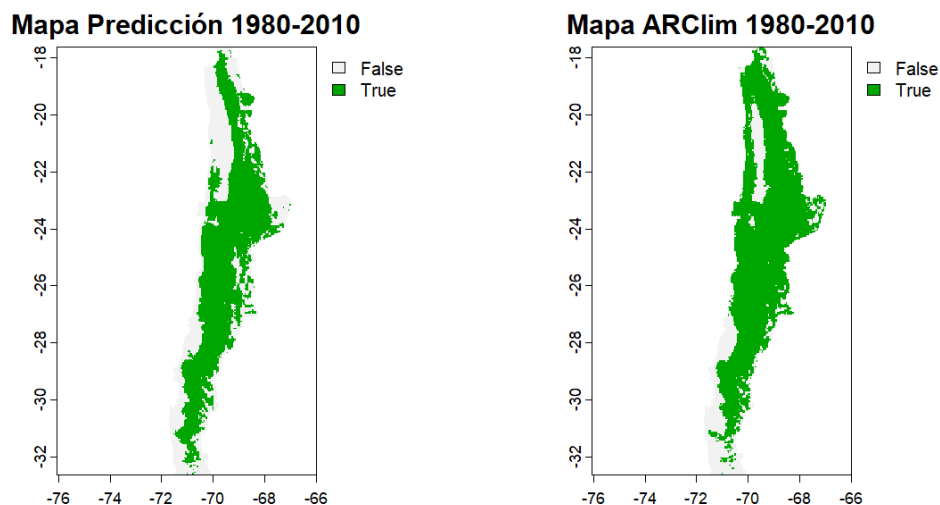


Figure 3: Mapas binarios de distribución de especies en zonas de interés, periodo histórico.

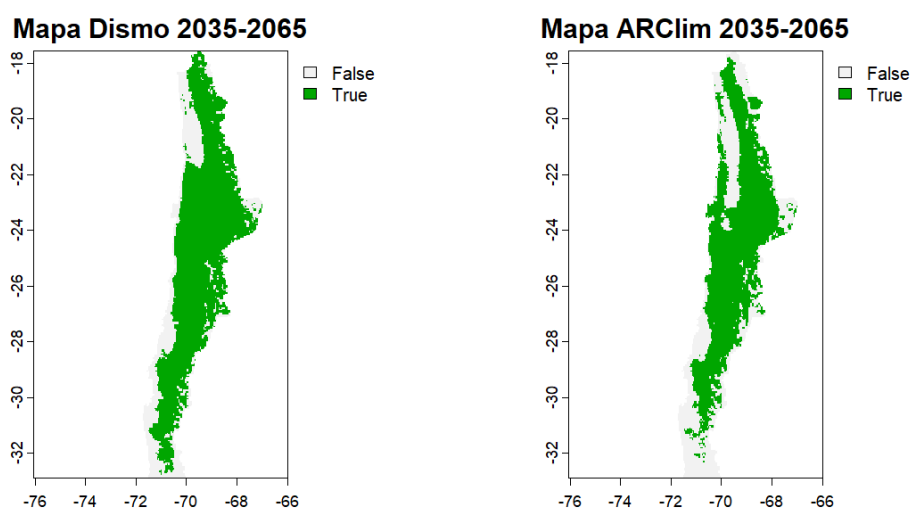


Figure 4: Mapas binarios de distribución de especies en zonas de interés, periodo futuro.

La estadística se trabaja sobre estos mapas binarios. En primer lugar, se obtienen mapas en donde se hace una diferencia entre la modelación hecha por cienciamambiental, menos la modelación de arclim; esto para el periodo histórico como futuro. En la Figura 5, las zonas en azul son donde Arclim predice presencia de la especie y Cienciamambiental no, zonas rojas es lo contrario, y las zonas en blanco es donde hay compatibilidad en las predicciones.

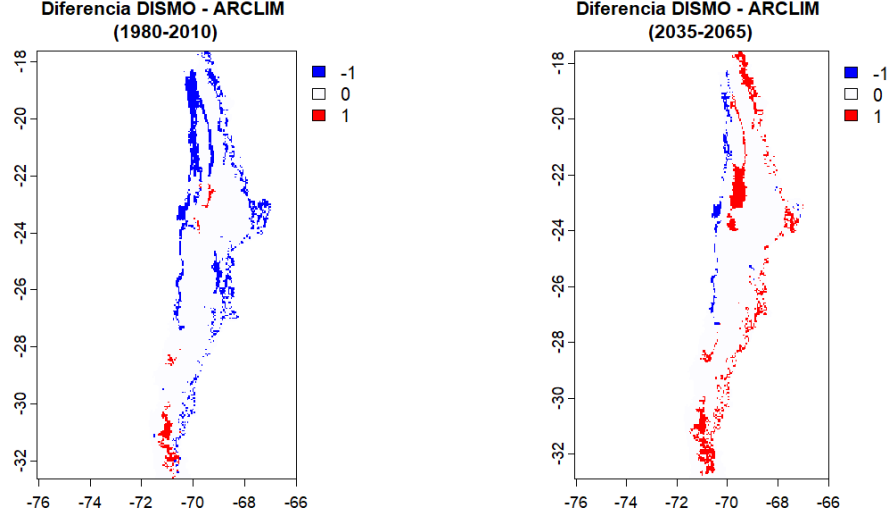


Figure 5: Diferencia de mapas binarios.

La **correlación** es de **0.613** y **0.663**, para (1980-2010) y (2035-2065), respectivamente. La matriz de confusión se presenta en la Tabla 1 y Tabla 2, en donde representa la compatibilidad de predicción de cada píxel entre el mapa elaborado en Cienciamiental y ARCLim. Por último, el **índice de kappa** es de **0.58** y **0.64**, para tiempo histórico y futuro.

Table 1: Matriz de Confusión para el Escenario Histórico Reciente

	FALSE	TRUE
FALSE	2473	1768
TRUE	287	6876

Table 2: Matriz de Confusión para el Escenario Futuro

	FALSE	TRUE
FALSE	2718	303
TRUE	1486	7044

Discusión

La replicación de la metodología utilizada en ARCLim para la modelación de distribución de *Adesmia atacamensis* en Cienciamiental 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 Cienciambiental es capaz de aproximarse en cierta medida a los modelos de ARCLim. 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 extraños en los mapas de distribución**, como una marcada línea horizontal generada en la Figura 1 del mapa generado a partir del paquete dismo, lo que motivó una revisión de las capas climáticas utilizadas en el modelamiento. Se identificó que algunas variables, especialmente la **precipitación acumulada**, podrían estar introduciendo estas estructuras anómalas en los mapas. Esto sugiere la necesidad de un análisis más detallado de las capas climáticas antes de ser utilizadas en los modelos, ya que una variable inadecuada o con errores en su proyección puede afectar significativamente los resultados. De todas formas, son las utilizadas en la fuente oficial de ARCLim.

Otro factor que podría estar influyendo en los resultados es la **falta de información detallada sobre la metodología de ARCLim**. Si bien se utilizaron las mismas variables climáticas y el mismo modelo Maxent, **no se cuenta con los datos de ocurrencia empleados en ARCLim**, lo que impide saber si se aplicaron filtros adicionales o si hubo un ajuste específico de los parámetros del modelo. Además, el uso exclusivo de **cinco variables climáticas** puede ser insuficiente para representar todos los factores ecológicos que determinan la distribución de la especie. Es probable que otras variables, como **factores edáficos, topográficos o de uso de suelo**, tengan un impacto importante en la distribución de la especie y no estén siendo consideradas en el modelo actual.

Otro punto importante a considerar es el impacto de **la resolución espacial en los resultados**. En ARCLim, los modelos utilizan una resolución de **5x5 km**, lo que puede ser adecuado para algunas especies pero insuficiente para otras con distribuciones más locales. En este sentido, sería relevante evaluar si realizar un **downscaling** o usar datos con mayor resolución permitiría capturar mejor la distribución de *Adesmia atacamensis* y reducir las diferencias entre ambos modelos.

Además, aunque se compararon los modelos generados con los de ARCLim, **no se realizó una validación independiente de los resultados**. Sería ideal contar con datos de monitoreo en terreno o registros independientes 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 no depender exclusivamente de la comparación entre mapas generados con 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 importante conocer la distribución proyectada de una especie, sino también **cuánta confianza se tiene en esas predicciones**.

Cuantificar la incertidumbre ayudaría a entender mejor qué variables tienen mayor impacto en el modelo y en qué zonas del mapa las predicciones son menos confiables.

Por último, sería interesante evaluar si **otros enfoques de modelamiento podrían mejorar los resultados**. Si bien MaxEnt es un modelo ampliamente utilizado en este tipo de análisis, otras metodologías como **Random Forest, GAMs o redes neuronales** podrían ser probadas para determinar si permiten obtener predicciones más precisas o robustas.

Si bien la replicación de la metodología fue exitosa en términos de estructura y procesamiento de datos, aún existen diferencias en los resultados que sugieren que **se necesita un mayor control sobre las variables de entrada y una mejor comprensión de los datos utilizados en ARClím**. Esto permitirá optimizar el proceso y mejorar la precisión de los modelos generados.

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.