



UNIVERSIDAD DE CHILE -FACULTAD DE CIENCIAS -ESCUELA DE PREGRADO

“Evaluación de la clasificación ambiental fluvial de Chile
River Environment Classification-Chile (REC-Chile) como
base de datos ambiental para la calibración de modelos de
distribución de especies ícticas en Chile”

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial
de los requisitos para optar al Título de Biólogo Ambiental

Gustavo Ignacio Valenzuela Núñez

Director de Seminario de Título: Prof. Irma Vila Pinto
Codirector de seminario de título: Magister Daniel Zamorano

Marzo 2021
Santiago - Chile



INFORME DE APROBACIÓN SEMINARIO DE TITULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile que el Seminario de Título, presentado por el **Sr. Gustavo Valenzuela Núñez**

“Evaluación de la clasificación ambiental fluvial de Chile
River Environment Classification-Chile (REC-Chile) como base de datos ambiental para la calibración de modelos de distribución de especies ícticas en Chile”

Prof. Irma Vila Caro

Director Seminario de Título: _____

M Cs. Daniel Zamorano Castillo

Director Seminario de Título: _____

Comisión Revisora y Evaluadora

Dr. David Véliz

Presidente Comisión: _____

Dra. Caren Vega

Evaluador: _____

Santiago de Chile, Mayo 2021

Agradecimientos

Tras finalizar esta importante etapa de mi vida me gustaría agradecer a las diferentes personas que me acompañaron en este recorrido y me apoyaron de alguna u otra forma.

En primer lugar, me gustaría agradecer a mi familia que me ha apoyado incondicionalmente durante toda mi vida y en esta etapa no fue la excepción. Gracias a mis padres por su constante apoyo, palabras de aliento, consejos y cariño incondicional que sin duda ayudaron a forjar la persona que soy hoy en día. A mis hermanas me gustaría agradecer por ser un soporte fundamental en mi vida y haberme apoyado siempre que lo necesité.

También me gustaría agradecer a la profesora Irma Vila por su increíble disposición a ayudarme en mi seminario de título, prestándome todo su apoyo desde el inicio y haciendo todo lo posible para que yo pudiera concretar este seminario de forma exitosa. Además, quiero agradecer a Daniel Zamorano, mi cotutor que mostró una dedicación absoluta para ayudarme a en este proceso, enseñándome todo lo que sé acerca del mundo de los modelos.

Por último, me gustaría agradecer a ese grupo de personas que conocí en la Universidad y que me acompañaron día a día. Gracias a Dali, David, Dani, Luciano, Carlitos, Sari, que me acompañaron prácticamente todos los días en las clases, en los almuerzos y en las conversaciones en los pastos, gracias a Constantza y Gabriel, en quienes encontré amigos incondicionales y por supuesto quiero agradecer a quien se terminó por convertir en alguien muy especial, con quien compartí mis éxitos y me ayudó en momentos difíciles brindándome su apoyo, Ailan.

Indice

Resumen	7
Abstract	9
Introducción	11
La ictiofauna chilena y sus características	11
Modelos de distribución de especies	13
SDM en ecosistemas fluviales	16
REC-Chile como clasificación eco hidrológica de los ríos de Chile	19
Objetivos	24
Objetivos específicos	25
Materiales y métodos	26
Registros de ocurrencia	26
Base de datos ambientales	26
Algoritmo de modelación	28
Área de estudio	29
Variables predictoras	31
Análisis estadísticos	33
Evaluación de distribuciones proyectadas	34

Resultados	35
Métricas de desempeño	35
Importancia de las variables	43
Evaluación espacial mediante un mapa de predicción de superposición de predicciones	45
Discusión	52
Desempeño de REC-Chile a diferentes escalas	52
Importancia de variables	57
Distribución predicha para <i>B. microlepidotus</i> en Chile	59
Protocolo y aplicación	61
Conclusión	63
Bibliografía	64

Resumen

La ictiofauna chilena es un grupo de alto valor sistemático y muy amenazado, siendo una gran preocupación su manejo y conservación. En este contexto, modelos de distribución de especies (SDM) se han utilizado para predecir el hábitat de distintas especies, representando una importante herramienta de gestión. Sin embargo, al momento de aplicar dichos modelos en ecosistemas fluviales surgen variados desafíos, tal como escasas de información ambiental y la selección de variables predictoras. Hoy en día existe REC-Chile, un sistema de clasificación eco hidrológica realizada para Chile que cuenta con información ambiental para cada río entre Arica y Puerto Montt. Lamentablemente, los casos de modelación realizados con REC-Chile han sido entrenados para escalas espaciales menores y podría ser necesaria su complementación con variables adicionales, tal como la altitud o la temperatura, que resultarían valiosas para representar características de los hábitats fluviales. De esta forma, este seminario busca evaluar el uso de REC-Chile para calibrar modelos de distribución de especies ícticas, evaluando su desempeño frente a diferentes escalas espaciales y utilizando distintos grupos de variables predictoras. Para realizar este trabajo, se seleccionó a *Basilichthys microlepidotus* Jenyns como especie modelo, cuyos datos de presencia fueron utilizados para entrenar modelos de distribución, utilizando tres conjuntos de variables sobre cuatro superficies distintas haciendo uso de Random Forest como algoritmo de modelado. El primer conjunto de variables incluyó solo las

variables de REC-Chile, el segundo incluyó nuevas clasificaciones obtenidas desde REC-Chile correspondientes a régimen hídrico y caudal, y el tercer conjunto incluyó nuevas variables, como la altitud, la precipitación y la temperatura. De los análisis se obtuvo un buen desempeño de REC-Chile como base ecohidrológica para entrenar SDM ($0,65 < AUC < 0,95$), siendo este desempeño mayor para modelos a escala nacional y con la inclusión de variables que complementen REC-Chile. Para las distintas superficies, los generados con nuevas variables fueron los más consistentes con la distribución conocida para *Basilichthys microlepidotus*, y en general se registró relación entre las variables predictoras más importantes y la auto ecología de la especie. Los modelos generados con los demás conjuntos se vieron afectados en algunos casos de manera negativa por la relación escala-resolución y por contar con variables categóricas como predictores.

Abstract

The Chilean ichthyofauna is a group of high systematic value and very threatened, being a great concern in its management and conservation. In this context, species distribution models (SDM) have been used to predict the habitat of different species, representing an important management tool. However, when applying these models in river ecosystems, various challenges arise, such as scarcity of environmental information and the selection of predictor variables. Today there is REC-Chile, an eco-hydrological classification system carried out for Chile that has environmental information for each river between Arica and Puerto Montt. Unfortunately, the modeling cases carried out with REC-Chile have been trained for smaller spatial scales and it may be necessary to complement them with additional variables, such as altitude or temperature, which would be valuable to represent characteristics of river habitats. In this way, this seminar seeks to evaluate the use of REC-Chile to calibrate distribution models of fish species, evaluating their performance against different spatial scales and using different groups of predictor variables. To carry out this work, *Basilichthys microlepidotus* Jenyns was selected as a model species, whose presence data were used to train distribution models, using three sets of variables on four different surfaces using Random Forest as a modeling algorithm. The first set of variables included only the REC-Chile variables, the second included new classifications obtained from REC-

Chile corresponding to water regime and flow, and the third set included new variables, such as altitude, precipitation and temperature. . From the analyzes, a good performance of REC-Chile was obtained as an ecohydrological basis to train SDM ($0.65 < \text{AUC} < 0.95$), this performance being higher for the national scale and with the inclusion of variables that complement REC-Chile. For the different surfaces, those generated with new variables were the most consistent with the known distribution for *Basilichthys micro-lepidotus*, and in general there was a relationship between the most important predictor variables and the auto-ecology of the species. The models generated with the other sets were affected in some cases negatively by the scale-resolution relationship and by having categorical variables as predictors.

Introducción

La ictiofauna chilena y sus características

La fauna dulceacuícola de Chile se encuentra compuesta por 11 familias, 17 géneros y aproximadamente 45 especies (Habit et al., 2006; Vila et al., 2006; Véliz et al., 2012; Arratia et al., 2017;) ; una baja cantidad si se compara con la de sus países vecinos como Argentina, que cuenta con alrededor de 500 especies (Liotta, 2006), o Perú, con aproximadamente 300 especies (Ortega & Hidalgo, 2008). En la literatura se asevera que las especies dulceacuícolas presentes en Chile poseen un alto valor sistemático, biogeográfico y de conservación, debido a su alto grado de endemismo (Habit et al. 2006; Vila et al. 2006). Estas y otras características, como tener pequeños tamaños corporales (Entre 5 cm y 40 cm; Zamorano et al. 2019) y retener caracteres primitivos que son conferidos por el aislamiento biogeográfico del país (Eigenmann 1927), hacen de la ictiofauna Chilena un interesante objeto de estudio.

Considerando el alto valor del grupo, resulta preocupante los grandes problemas de conservación que este presenta, contando con el 89% de sus especies en estado de conservación vulnerable o de mayor gravedad (aquí falta una referencia). Dentro de las principales causas de su deteriorado estado de conservación, está la modificación del régimen hidrológico debido a canalizaciones, represamiento de los ríos (Vila et al. 1999) y extracción de

agua para riego (Habit et al., 2006), lo que se traduce en la destrucción del hábitat físico de las especies. Además del vertimiento de contaminantes de origen industrial, urbano y agropecuario en las cuencas (Figueroa et al., 2013). Estos factores generan un deterioro progresivo de los ecosistemas acuáticos, cuyos niveles de nutrientes y condiciones físicas cambian alterando las interacciones de las comunidades que habitan dichos ecosistemas (Figueroa et al., 2013). A lo anterior se suma la presencia de especies introducidas, cuyo efecto se consideraría como una causa de extinciones locales de comunidades de peces nativos (Habit et al., 2006; Soto et al., 2006), y la gran falencia en cuanto a conocimiento disponible acerca de la ictiofauna nativa, que dificulta avanzar en políticas o gestión para la conservación y/o rehabilitación de ecosistemas acuáticos (Peredo-Parada et al., 2009). Frente a todas estas problemáticas nace con urgencia la necesidad de investigar y generar herramientas que faciliten la protección y el manejo de la ictiofauna dulceacuícola nacional.

Una herramienta que ha sido usada durante varios años para la protección y el manejo de especies son los modelos de distribución de especies, que permiten la predicción de su nicho potencial, proporcionando una idea aproximada de cuál sería el espacio geográfico habitado por estas y facilitando así el desarrollo de políticas públicas para su conservación (Peredo-Parada et al., 2011). La efectividad de estos modelos se ve reflejada en la gran diversidad de organismos sobre los que se han realizado, entre

los que se pueden encontrar angiospermas (Pouteau et al., 2015), anfibios (Hernández-Gómez et al., 2020; Vasconcelos et al., 2012), hongos (Chamier et al., 1984), himenópteros (Gonzalez et al., 2010), primates (Palminteri et al., 2012), aves (Mi et al., 2017), entre tantos otros.

Si bien, esta herramienta no ha sido utilizada sobre especies acuáticas tan ampliamente como en especies terrestres, sí cuenta con experiencias que han producido excelentes resultados (por ejemplo: Kuemmerlen et al., 2016; McNyset, 2005; Rojas Venegas & Espinoza, 2018; Sherwood et al., 2018). Es por esto que en este trabajo se explorará el uso de modelos de distribución de especies aplicados a la ictiofauna chilena, con el fin de analizar los posibles alcances y aportes que un modelo de estas características puede tener sobre la gestión y manejo de los peces chilenos.

Modelos de distribución de especies

Los modelos de distribución de especies, comúnmente denominados SDM por sus siglas en inglés, fueron definidos por Mateo et al. (2011) como representaciones cartográficas de la idoneidad de un espacio para la presencia de una especie en función de las variables empleadas para generar dicha representación e idoneidad, que no es más que la relación matemática o estadística entre la distribución real conocida y un conjunto de variables independientes que se usan como indicadores. Los SDM consisten básicamente en un proceso de clasificación (Guisan & Zimmermann, 2000), donde

la variable dependiente puede ser dicotómica (presencia / ausencia) o cuantitativa (probabilidad de ocurrencia), mientras que las independientes pueden ser cuantitativas (e.g., temperatura o elevación) o nominales (e.g., litología o uso del suelo) (Mateo et al., 2011).

Si queremos utilizar esta herramienta para el desarrollo de SDM de especies ícticas, según la literatura vigente debemos sortear tres principales desafíos: La obtención y procesamiento de los registros de las especies, la selección de la escala espacial y la resolución a utilizar, y la selección de las variables predictoras (Mertes & Jetz, 2018). Estos desafíos están estrechamente relacionados y hay que abordarlos en conjunto.

En primer lugar, la disponibilidad de registros de presencia es fundamental, ya que a partir de las condiciones ambientales que ocurran en la ubicación de cada presencia se estimará un nicho potencial (Elith & Leathwick, 2009). De esta forma, un mayor número de presencias determinará más información disponible para estimar su distribución potencial. Factores como su cantidad, su distribución espacial o temporal, podrían impactar negativamente los resultados, disminuyendo la calidad de ajuste y sesgando las predicciones espaciales (Fei & Yu, 2016).

En segundo lugar, se considera la escala y resolución a la que se trabajará, la que resulta fundamental definir para evaluar si los registros de presencia son suficientes o no. La escala consiste en el tamaño de la superficie sobre el que se trabaja o proyecta un modelo, mientras que la resolución es la unidad

mínima en la que se divide esta superficie (celdas o píxeles) (Mertes & Jetz, 2018). Estos dos elementos están estrechamente relacionados entre sí, y al mismo tiempo se relacionan con las presencias disponibles, la especie a modelar y el objetivo del estudio. Por ejemplo, para modelar distribuciones de especies a grandes escalas geográficas con bajas resoluciones (píxeles de mayor tamaño), se precisan presencias lo suficientemente espaciadas entre sí, para no quedar agrupados en una pequeña cantidad de píxeles, logrando representar estadísticamente la diversidad ambiental de la especie a modelar (Lauzeral et al., 2013). En otro ejemplo, mientras una especie sésil probablemente requiera de una mejor resolución espacial para ver caracterizado su hábitat (píxeles más pequeños), especies de gran movilidad se relacionarán con el entorno a mayores escalas espaciales, requiriendo para su modelación una resolución más gruesa (píxeles más grandes), ya que ambas se relacionan de manera diferente con las condiciones abióticas del medio (Lauzeral et al., 2013).

El tercer desafío a sortear es la selección de variables predictoras. Este proceso estará estrechamente relacionado con la especie a modelar, la escala geográfica y la resolución en la que se trabajará. Mientras que existen algunas variables que actúan sobre especies en escalas tan pequeñas que se requiere de gran resolución para observar su efecto (en los sistemas lóticos: velocidad de flujo, tipo de sustrato, tipo de meso hábitat, entre otros), otras actúan a escalas más grande y su efecto puede ser observado sin inconvenientes con resoluciones mayores (temperatura media anual, precipitación, entre otros). Finalmente, la

forma en la que estas variables interactúan con las especies está sujeta a características de dichas especies, como su tamaño corporal o dieta, lo que también determinará qué variables predictoras escoger para cada caso.

Cada modelación exitosa ha debido sortear estos tres desafíos (cantidad de registros - resolución y escala – variables predictoras) considerando la especie a modelar y sus objetivos. Por ejemplo, la modelación de angiospermas realizadas por Pouteau et al. (2015), se llevó a cabo sobre la superficie de toda la isla Nueva Caledonia, de unos 24000 km² (casi el doble del área de la Región Metropolitana), usando las variables bioclimáticas con una resolución de aproximadamente 5 km². Mientras que la modelación de nicho de la abeja enana (Gonzalez et al., 2010) proyectó una superficie de aproximadamente 2000 km² de nicho potencial, utilizando variables bioclimáticas con una resolución de 1 km². Otro buen ejemplo sería la modelación de peces realizada de la cuenca del río Mataquito de unos 95 km de longitud (Rojas Venegas & Espinoza, 2018). En este caso, el objetivo del estudio fue establecer la distribución de una especie en un área acotada, por lo que se requirió diferentes variables, como la velocidad, calidad del agua o la disponibilidad de alimento, y la resolución usada fue de pixeles equivalentes a 10 m².

SDM en ecosistemas fluviales

Para llevar a cabo de manera exitosa un SDM en ecosistemas fluviales se debe sortear desafíos como la cantidad de registros, la resolución y

la escala y la selección de variables predictoras, como también las características particulares que se presentan en estos ecosistemas.

Mientras que las áreas estudiadas en ecosistemas terrestres, lacustres o marinos por lo general están constituidas por amplios polígonos uniformes, los ecosistemas fluviales presentan formas alargadas, delgadas y sinuosas, lo que implica una dificultad al momento de representarlas utilizando imágenes satelitales de baja resolución. Esto ha afectado directamente los resultados de los SDM para estos ecosistemas (Lauzeral et al., 2013).

Respecto a la selección de variables ambientales para SDM, en estos ecosistemas se encuentran dos importantes desafíos: (i) Considerar que estos son ecosistemas cuyas variables ambientales se relacionan de forma jerárquica, longitudinal, lateral y vertical, de manera que un segmento de río está afectado por lo que ocurre en el tramo de aguas arriba además del efecto que ejercen estas variables de forma directa en dicho tramo (Domisch et al., 2015). Esto tiene como consecuencia que las variables predictoras necesarias para caracterizar estos ecosistemas sean altamente variadas y se encuentren teóricamente correlacionados entre sí (Elliott, 2010; Leopold, 1969), lo que dificulta su estimación y selección; (ii) Al contrario de los ecosistemas terrestres donde existen bases de datos mundiales públicas de variables ambientales cuyo uso es generalizado (Booth et al., 2014), en los ecosistemas fluviales la cantidad de variables predictoras disponibles para

descargar a escala mundial son relativamente escasas. Esto último, sumado al punto anterior, determina una alta variabilidad de variables predictoras encontradas en literatura, sin existir claridad de un conjunto de variables mínimas a utilizar.

Por ejemplo, entre las variables ambientales utilizadas, podemos encontrar variables topográficas como altitud, pendiente y exposición (Esselman & Allan, 2011; Labay et al., 2011; McNyset, 2005; Phillips et al., 2017); variables climáticas como temperatura media anual, precipitación anual, temperatura máxima del mes más cálido, temperatura mínima del mes más frío, entre otras (Esselman & Allan, 2011; Labay et al., 2011; Taylor et al., 2018; Zamorano et al., 2019); variables bióticas como presencia de especies no nativas y tipos de vegetación (Labay et al., 2011; Taylor et al., 2018); intervención antrópica como presencia de represas o uso de suelo (Taylor et al., 2018); y variables hidrológicas como caudal, velocidad de la corriente, tamaño de drenado desde el lago de origen, sinuosidad, entre otras (Porter et al., 2000; Radinger et al., 2017; Rojas Venegas & Espinoza, 2018). De las variables mencionadas anteriormente, algunas se repiten de forma frecuente en los trabajos de SDM para especies ícticas, como caudal, la altitud, y la temperatura del agua. Sin embargo, los investigadores nunca utilizaron exactamente el mismo conjunto de predictores e, inclusive, en muchos casos no utilizaron el mismo proxy para representar la misma variable ambiental. Por ejemplo, mientras en algunos trabajos el clima fue representado por

una variable como precipitación media anual (Peredo-Parada et al., 2011), en otros fue representado por una serie de variables relativas asociadas a la precipitación y a la temperatura, como temperatura media anual, temperatura estacional, precipitación del mes más húmedo, precipitación del mes más seco, entre otras (Labay et al., 2011). De esta forma, la selección de variables se presenta como un problema particularmente difícil de resolver.

REC-Chile como clasificación eco hidrológica de los ríos de Chile

Como se mencionó anteriormente, la elaboración de un SDM presenta grandes desafíos y, para el caso de Chile, los sistemas fluviales señalan características muy particulares, como altas pendientes, reducidas longitudes, alta sinuosidad, el origen glaciar y volcánico de las cuencas, entre otros (Peredo-Parada, 2010). Es en este contexto que nace REC-Chile, una clasificación eco hidrológica de los ríos de Chile que se crea en base a otra clasificación eco hidrológica llamada REC (River Environment Classification), elaborada por Snelder & Biggs (2002) para los ríos de Nueva Zelanda que se basa en las siguientes premisas:

a) Los patrones ecológicos son dependientes de un conjunto de factores asociados a la escala regional de diversos procesos físicos.

b) Las características ecológicas de los ríos responden a los procesos fluviales, asignando en forma geográficamente independiente las clases de cada tramo de río.

En Peredo-Parada (2011) se fundamentó que gracias a las similitudes eco-hidrológicas entre los ríos de Chile y Nueva Zelanda (régimen hidrológico y geomorfología de los ríos), y también entre las características taxocenóticas de especies de vertebrados y macro invertebrados, es posible adaptar la REC para Chile.

REC-Chile consiste en una clasificación *a priori* creada en los ríos de Chile, ubicada en las cuencas entre los 18°S y 43°S. Está constituida por 6 variables nominales (Clima, Origen de flujo, Geología, Posición relativa en la cuenca, Uso de suelo y Pendiente), que controlan las características fluviales a distintas escalas (Tabla 1). REC-Chile cuenta con una resolución espacial a escala de segmento, caracterizando tramos de río de entre 1 y 10 km de largo, creados en base a la presencia de variables homogéneas presentes en cada uno (Peredo-Parada et al., 2009).

Tabla 1: Variables que contiene REC-Chile con su respectiva escala, respectivas clases y una breve descripción en base a lo descrito en Peredo-Parada (2010). Las escalas se ajustaron siguiendo los parámetros definidos por (Domisch et al., 2015).

Escala	Variable	Clases	Descripción
Continental	Clima	Árido Semiárido Subhúmedo Húmedo Muy húmedo	Está representado como las precipitaciones anuales, determinando características hidrológicas como la magnitud del flujo, junto con la frecuencia de avenidas y de caudales bajos.
Regional	Origen de flujo	Llano Valle Precordillera	Basadas en el porcentaje de precipitación acumulada en cada tramo combinada con

		<p>Cordillera Nieves eternas Lagos Regulaciones Glaciar</p>	<p>la elevación. Se relacionan con patrones hidrológicos en régimen natural. Fue necesario incorporar otra categoría para los tramos donde existen regulaciones antrópicas, por ejemplo embalses; la categoría incorporada es Regulaciones.</p>
	Geología	<p>Depósitos no consolidados Plutónicas e hipoabisales Volcánicas Sedimentarias y mixto sedito-volcánicas Volcánicas fracturadas Carbonatadas Área sin información</p>	<p>Construido en base a las distintas litologías del suelo. Se relaciona, a un nivel de meso escala con las características geoquímicas del agua dominadas por las propiedades geológicas del acuífero.</p>
	Posición relativa de la subcuenca	<p>Cabecera Tramo alto Tramo medio Tramo bajo Endorreica Internacionales</p>	<p>Construida en base al área y posición de los segmentos con respecto a la totalidad del curso de agua.</p>
	Uso de suelo	<p>Áreas urbanas Agricultura Praderas y matorrales Bosques Áreas sin vegetación Cuerpos de agua Humedales Nieves y glaciares Áreas sin información</p>	<p>Construida en base a los usos de suelo que se les han asignado a los tramos por los que pasan los ríos. Se relaciona con procesos como la interceptación inicial y los procesos de evapotranspiración, además controla aspectos fisicoquímicos de la calidad del agua producto de la lixiviación del suelo y arrastre del río</p>
Paisaje	Pendiente del tramo	<p>Alta Media Baja</p>	<p>Describe patrones morfo hidráulicos a un nivel de micro escala, tales como transporte local de sedimento, erosión del cauce, velocidad media del flujo e influencia</p>

			de las condiciones de la ribera.
--	--	--	----------------------------------

REC-Chile se presenta como una útil herramienta al momento de seleccionar variables ambientales para confeccionar un modelo de distribución de especies ícticas en Chile, puesto que entrega *proxys* que representan seis variables que constituyen los ecosistemas fluviales a distintas escalas ecológicas y, además, al combinarse pueden resumir otras variables ambientales. Por ejemplo, en el caso del caudal, variable para la que no existen mapas o bases de datos que registren sus valores para cada tramo de río a nivel nacional, si consideramos las variables de REC-Chile, esta puede ser resumida por la combinación de las variables Clima, origen de flujo y pendiente. Puesto que el clima en conjunto con el origen de flujo nos entrega información del volumen de agua que pasa por un canal y su régimen, y la pendiente funciona como proxy de velocidad, pudiendo estos factores en conjunto servir como proxy para obtener el caudal de un río.

La existencia de bases de datos eco hidrológicas en Chile es escasa, además de REC-Chile existe otra compuesta por 5 variables (Altitud, Geología, Pendiente del fondo del canal y régimen de descarga anual promedio) (Fuster et al., 2015). En este escenario de escasas bases de datos con variables eco hidrológicas, REC-Chile fue seleccionada debido a que cuenta con más variables ambientales para caracterizar el hábitat de las especies,

además que para REC-Chile se encontraron trabajos donde ya se utilizó de forma exitosa esta base de datos en modelación de distribución de especies (Peredo-Parada, 2010; Zamorano et al., 2019), mientras que con la otra base mencionada no se encontraron trabajos de esta índole.

Utilizar REC-Chile para la modelación de especies dulceacuícolas chilenas representaría una gran herramienta de gestión, sin embargo, aún deben resolverse ciertas preguntas para ajustar esta clasificación eco hidrológica a un set de variables predictoras para SDM. En primer lugar, REC-Chile no cuenta con un proxy directo de temperatura, variable de la que depende la concentración de oxígeno disuelto en el agua y, por ende, que también define otros factores como el pH, además de estar directamente relacionada con la energía para la fotosíntesis (Davinson, 1991). Por otro lado, los trabajos de modelización llevados a cabo hasta ahora haciendo uso de REC-Chile (Peredo-Parada, 2010; Zamorano et al., 2019) no han contemplado diferentes escalas geográficas, por lo que se desconoce cuál será el efecto que generará una variación de escalas espaciales en su desempeño, considerando la resolución invariable que esta base de datos presenta y las características propias de cada variable ambiental.

De esta forma, en este seminario se evaluó a REC-Chile como base de datos de variables eco hidrológicas para la elaboración de modelos de distribución de especies ícticas en Chile a distintas escalas espaciales. Para esto, se realizaron modelos de distribución de especies utilizando distintas

escalas (continental, regional y de paisaje) y haciendo uso de diferentes conjuntos variables predictoras presentes en REC-Chile, buscando evaluar el comportamiento de estos SDM ante distintos escenarios de modelación y evaluando el mejor conjunto potencial de variables predictoras ante estos escenarios.

Se seleccionó *Basilichthys microlepidotus* (Jenyns) como especie modelo, debido a la gran cantidad de datos de presencias con las que cuenta, su amplia distribución y por su coherencia con la cobertura actual de REC-Chile. Esta es una especie de hábito fluvial, habita en aguas dulceacuícolas bien oxigenadas, con bajas velocidades, transparentes y con profundidades mínimas de 40 centímetros (Ministerio del Medio Ambiente. 2007, Ministerio del Medio Ambiente. 2013).

Objetivos

El objetivo general de este seminario es evaluar el uso de REC-Chile como base de datos ambiental para calibrar modelos de distribución de especies ícticas en Chile, evaluando su desempeño frente a diferentes escalas espaciales y utilizando distintos grupos de variables predictoras.

Objetivos específicos

1. Desarrollar modelos de distribución de especie para *B. microlepidotus* utilizando distintas escalas espaciales y conjuntos de variables ambientales.
2. Evaluar el desempeño de los modelos realizados y compararlos estadísticamente en función de las distintas superficies y los diferentes conjuntos de variables ambientales utilizados.
3. Comparar las distribuciones proyectadas para *B. microlepidotus* en el territorio para las distintas superficies modeladas considerando su literatura asociada.
4. En función de los resultados, establecer un protocolo de buenas prácticas de modelación de especies ícticas utilizando REC-Chile como base de datos ambientales.

Materiales y métodos

Registros de ocurrencia

Los registros de ocurrencia fueron extraídos de una base de datos creada por el Ministerio del Medio Ambiente que contiene registros geográficos de peces chilenos. Se añadieron además todos los registros de *B. australis*, y *B. microlepidotus* ya que ha surgido evidencia (Véliz et al., 2012) que concluye que ambas corresponden a la misma especie: *B. microlepidotus*, lo que aporta un mayor número de datos de presencia. Por último, con la finalidad de mejorar el resultado para un modelo actual, se dejaron de lado todos los registros anteriores a 1990, teniendo en consideración que registros muy antiguos de lugares donde ya no existe la presencia de la especie podrían producir una sobre-predicción de falsos positivos.

Base de datos ambientales

Tal como se señaló en la introducción, las variables predictoras fueron obtenidas desde REC-Chile, base de datos eco hidrológica comprendida entre los 18°S y 43°S, en la que los ríos fueron separados en 40885 segmentos de longitudes que varían entre los 1 y 10 km con características hidromorfológicas homogéneas (M. Peredo-Parada et al., 2009). Los registros de presencia y las pseudo-ausencias generadas fueron espacialmente asignadas a tramos de REC-Chile, creando un buffer de 500 metros alrededor de estos

y extrayendo los puntos de presencia que se interceptaban con el buffer, para después asociar cada punto de presencia de *B. microlepidotus* a un tramo de REC-Chile, obteniendo de la misma forma su información ambiental. Esto nos dejó con un total de 160 presencias para realizar los modelos de distribución (Fig. 1).

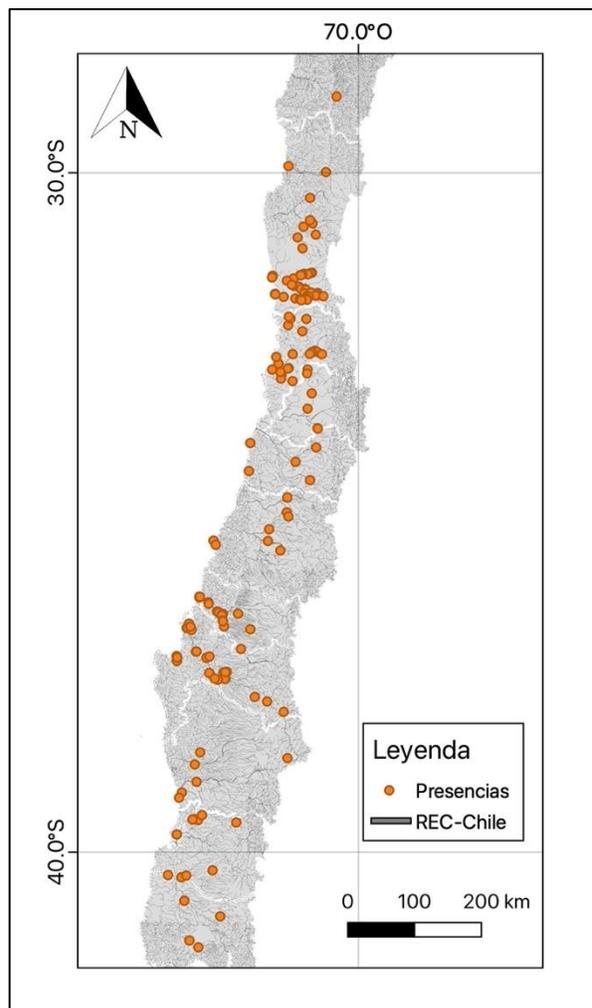


Figura 1. Mapa de Chile que muestra la red de REC-Chile y los registros de presencia una vez filtrados.

Algoritmo de modelación

El método de modelación escogido fue Random Forest (Breiman, 2001), realizado utilizando el software RStudio Team (2016) por medio de la librería CARET (Khun 2016). Este método fue seleccionado por sobre otros puesto que ha demostrado ser una herramienta particularmente buena en modelación eco hidrológica (Peters et al., 2007) y en casos donde existen pocos registros (Mi et al., 2017), como resulta ser el caso de muchas de las especies ícticas en Chile.

Para el entrenamiento de la modelación se utilizó el método de “repeated cross validation” (validación cruzada con repetición, en español). El método de cross-validation consiste en la subdivisión de los datos en k subconjuntos, y mientras $k-1$ conjuntos son utilizados para entrenar el modelo, el conjunto restante es utilizado para validarlo. Este proceso se repite k veces, cambiando en cada iteración el subconjunto de validación, permitiendo que todos los datos sean utilizados para entrenar y evaluar el modelo. Luego, repeated cross validation repite en n ocasiones la validación cruzada, pero cambiando en cada iteración la agrupación de los subconjuntos de datos, obteniendo el promedio de los resultados de todas las repeticiones ponderado por su rendimiento, lo que proporciona una mayor robustez al resultado de entrenamiento. Para realizarlo, se ajustó con 10 repeticiones y segmentando los datos en 8 conjuntos. Dado que no se cuenta con registros de ausencias reales, se utilizaron pseudo-ausencias, correspondientes al doble

de presencias. Siguiendo la metodología recomendada por Barbet-Massin et al. (2012), la selección de pseudo-ausencias se repitió 10 veces, seleccionando el conjunto que arrojó un mejor desempeño en los modelos.

Área de estudio

Para llevar a cabo los modelos, se seleccionaron cuatro escalas espaciales utilizando tres superficies de diferentes tamaños (Fig. 2). Para asociar cada superficie a una escala espacial utilizamos la clasificación de escalas ecohidrológicas propuestas por Domisch et al. (2015) con el fin de contrastar la importancia de las variables con las escalas en las que actúan. Las superficies escogidas fueron cuatro: la primera consiste en toda la extensión de REC-Chile (18°S y 43°S) (Fig. 2a), a la que llamamos “extensión nacional”, escogida por ser una extensión de escala continental, la mayor categoría que se puede encontrar dentro del territorio nacional. La segunda superficie corresponde a un grupo contiguo de 7 regiones correspondiente a la sección sur de REC-Chile (33°S – 43°S) (Fig. 2b), a la que denominamos “zona sur”, seleccionada por representar una escala regional, categoría que sucede a la continental. Además, se seleccionó una segunda superficie de categoría regional correspondiente a cuatro regiones contiguas (36°S – 41°S) (Fig. 2c), a la que llamamos superficie tetra regional, escogida para crear un cambio gradual entre los tamaños de superficie. Por último, se escogió una superficie de escala de paisaje, correspondiente a la región del

BioBio y la región del Ñuble (Fig. 2d, superficie bi regional), puesto que esta es la escala que sucede a regional y representa una modelación a escala de cuenca. La ubicación de cada superficie fue determinada por los registros con los que contaba; 160 registros en la superficie Nacional, 85 para la superficie Sur, 25 en la superficie Tetra regional y 18 en la superficie Bi regional.

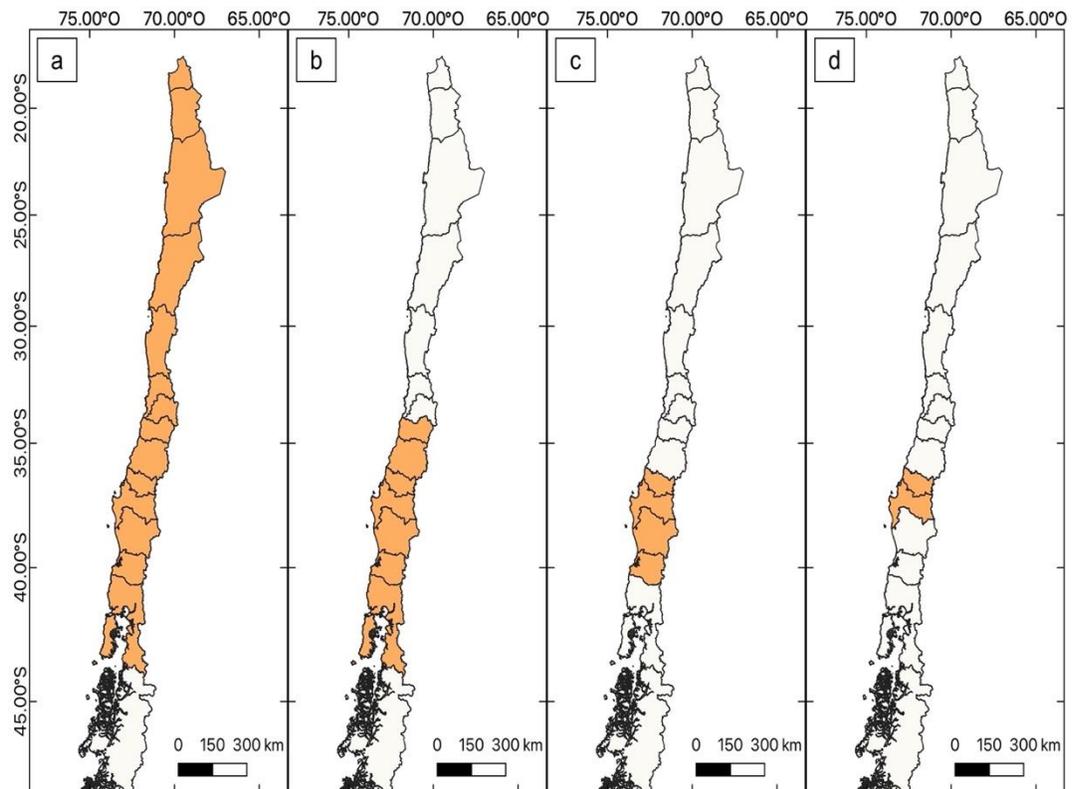


Figura 2. Mapas que muestran las superficies sobre las que se trabajó: superficie nacional (a), superficie sur (b), superficie tetraregional (c) y superficie birregional (d).

Variables predictoras

Para llevar a cabo los modelos se conformaron 3 grupos de variables distintas (Tabla 2). El conjunto Rec, que contiene todas las variables entregadas por REC-Chile de manera independiente (variables en Tabla 1). El conjunto Rec2, que contiene las variables entregadas por la REC-Chile además de dos nuevas variables: régimen hídrico y caudal. La variable régimen hídrico, tiene por objetivo representar el régimen hídrico del río a través de la combinación de las variables “clima” y “origen de flujo”. Puesto que clima entrega información del aporte hídrico recibido por toda la cuenca producto de las precipitaciones y origen de flujo proporciona información de cuál es principal afluente sobre el río, su unión resumiría el régimen hídrico del tramo al considerar un proxy de volumen de agua y sus cambios temporales a escala anual. La otra variable es “caudal”, que consiste en la variable “régimen hídrico” en conjunto con la variable “pendiente”, que está en una escala paisaje, y su aporte da cuenta de cómo el régimen hídrico estaría actuando en cada tramo, interactuando con factores como la velocidad y turbulencia del agua.

El último conjunto de datos es el Mixto, que consiste en las variables de la REC-Chile, además de tres variables numéricas y continuas: Temperatura, Altitud y Precipitación, que fueron extraídas de las capas bioclimáticas proporcionadas por WorldClim (<http://www.worldclim.com/org>). En este grupo de variables, la variable “clima” presente en REC-Chile, que también

clasifica precipitaciones, fue reemplazada dado que incluimos precipitación como variable continua. La variable Temperatura fue obtenida de la capa Bio1 y representa la temperatura media anual, expresada en °C, mientras que la variable Precipitación fue obtenida desde Bio12 y representa la precipitación media anual expresada en milímetros cúbicos. Finalmente, los SDM realizados fueron 12, utilizando 4 superficies x 3 conjuntos de variables predictoras.

La importancia de la participación de cada variable predictiva en cada modelo fue realizado a través del algoritmo varImp de la librería CARET (Khun 2016). En este aspecto, una mayor importancia de variable representa un mejor poder predictivo que otras al ser utilizada como bifurcación superior en los árboles de decisión realizados por Random Forest o, dicho de otra manera, una mayor importancia de variable denota la mejor capacidad de esta frente a las demás para clasificar casos de presencia y ausencia (Kuhn, 2008).

Tabla 2 Tabla con los conjuntos de variables utilizados para llevar a cabo los diferentes modelos, las variables que los componen, y una breve observación respecto a cada uno.

Conjuntos	Variables	Observación
Rec	Clima Origen de flujo Posición relativa el tramo Geología Uso de suelo Pendiente	Este conjunto fue creado con las variables proporcionadas por la REC-Chile de forma independiente.

<p>Rec2</p>	<p>Clima Origen de flujo Posición relativa del tramo Geología Uso de suelo Pendiente Régimen hídrico Caudal</p>	<p>Este conjunto de variables consta de las seis variables de la REC-Chile además de dos variables añadidas: Régimen hídrico, compuesta por la combinación de Clima y Origen de flujo, y Caudal, compuesta por Clima, Origen de flujo y pendiente.</p>
<p>Mixto</p>	<p>Temperatura media anual Precipitación anual Altitud Origen de flujo Posición relativa el tramo Geología Uso de suelo Pendiente</p>	<p>Este conjunto de variables consta de las variables independientes de la REC-Chile, pero clima fue reemplazado por su proxy (precipitación anual), expresado en valores continuos. Además de agregar dos variables adicionales: Temperatura media anual y altitud.</p>

Análisis estadísticos

Para comparar los desempeños de los modelos se utilizó el ROC AUC, parámetro de desempeño utilizado ampliamente en los modelos de distribución de especies y recomendado para Random Forest por sobre otros, como la exactitud o el kappa (Ling et al., 2003). Además, se utilizó la sensibilidad, indicador que mide la capacidad del modelo para predecir de manera correcta las predicciones de presencia, y la especificidad, que se utiliza para medir la capacidad del modelo para predecir de manera correcta las ausencias. Para estos tres indicadores, los criterios de evaluación para un resultado son: desempeño excelente

(0.9–1.0), muy bueno (0.8–0.9), bueno (0.7–0.8), moderado (0.6–0.7), y pobre (0.5–0.6) (Swets, 1988).

Si bien solo se obtuvo 12 modelos finales, en cada repetición de entrenamiento de modelo ($n = 10$) se obtuvieron parámetros de desempeño (AUC, especificidad y sensibilidad), por lo que en total contamos con 120 datos (4 superficies x 3 conjuntos variables x 10 repeticiones), pudiendo abordar estadísticamente las diferencias entre modelos. Para evaluar diferencias significativas se realizó un ANOVA anidado, en donde los desempeños para los modelos con distintos conjuntos de variables fueron anidados dentro de la superficie modelada (~ Superficie/conjunto de variables). Una vez realizado el ANOVA anidado se realizó un ANOVA de una vía entre los modelos con los diferentes conjuntos de variables para cada superficie por separado para explorar de mejor manera los resultados obtenidos. El análisis *a posteriori* se realizó con un test de Tukey usando la función TukeyHSD en RStudio Team (2016).

Evaluación de distribuciones proyectadas

Se realizaron mapas que contrastaron los resultados de las predicciones producidas por los diferentes conjuntos en cada predicción. Para esto, se realizaron tres mapas por superficie; en cada uno se destacaron los tramos según si el tramo era predicho exclusivamente por un conjunto, si era predicho por cualquiera de los otros conjuntos, o si era predicho por los tres conjuntos a la vez. Lo

anterior se realizó con el fin explorar posibles sesgos producidos por los conjuntos de variables y cómo la importancia de estas actúa sobre la predicción. La selección del umbral sobre el cual se asumió una probabilidad de ocurrencia como una presencia se realizó maximizando la suma de la especificidad y sensibilidad (MaxSens+Spec; PresenceAbsence package in R v 3.5.0) (R Core Team., 2017). Para realizar estos mapas se utilizó QGIS Development Team (2018).

Resultados

Métricas de desempeño

Los modelos realizados utilizando el conjunto de variables Mixto fue el que mejor desempeño generó para todas las superficies, obteniendo resultados satisfactorios con valores de AUC entre 0,75 y 0,95, acompañado de altos valores de sensibilidad y especificidad entre 0,56 y 0,97 (Fig. 3). Los demás conjuntos produjeron menores desempeños. Por un lado, el conjunto Rec arrojó modelos con valores de AUC entre 0,75 y 0,89, junto a valores de especificidad y sensibilidad entre 0,35 y 0,94; mientras que el conjunto Rec2 arrojó modelos con valores de AUC entre 0,65 y 0,85 (Fig. 3), junto a valores de especificidad y sensibilidad entre 0,35 y 0,89. Además, la superficie Nacional fue la que mejor desempeño logró para todos los conjuntos, obteniendo resultados satisfactorios para el conjunto Rec y Mixto ($0,83 < AUC < 0,95$), mientras que con el conjunto Rec 2 arrojó resultados menos satisfactorios ($0,68 < AUC < 0,77$) (Fig. 3).

Cuando analizamos los análisis estadísticos para el ROC-AUC, el ANOVA anidado arrojó una diferencia significativa en los resultados de AUC tanto entre superficies como para conjuntos de variables dentro de cada superficie (Tabla 3). En cuanto a la diferencia entre superficies, el test de Tukey reveló que los resultados de AUC tienen diferencias significativas entre el modelo realizado sobre la superficie Nacional y la Birregional, y entre el modelo Nacional y Sur (AUC.: (Nacional=Tetrarregional)>Sur>Birregional)) (Tabla 4). Respecto a los conjuntos de variables, el test de Tukey mostró diferencias significativas entre todos los conjuntos en la superficie nacional (AUC: Mixto > Rec > Rec2), mientras que en el resto de las superficies solo mostró diferencias significativas entre el conjunto Mixto y el resto de los conjuntos (AUC: Mixto > (Rec = Rec2)) (Tabla 5). Cabe destacar el AUC de los modelos Mixto y Rec muestran una tendencia a disminuir a medida que disminuye la escala espacial, no así para los modelos de Rec2, que muestran una tendencia de aumento a medida que aumenta el tamaño de la superficie (Fig. 3).

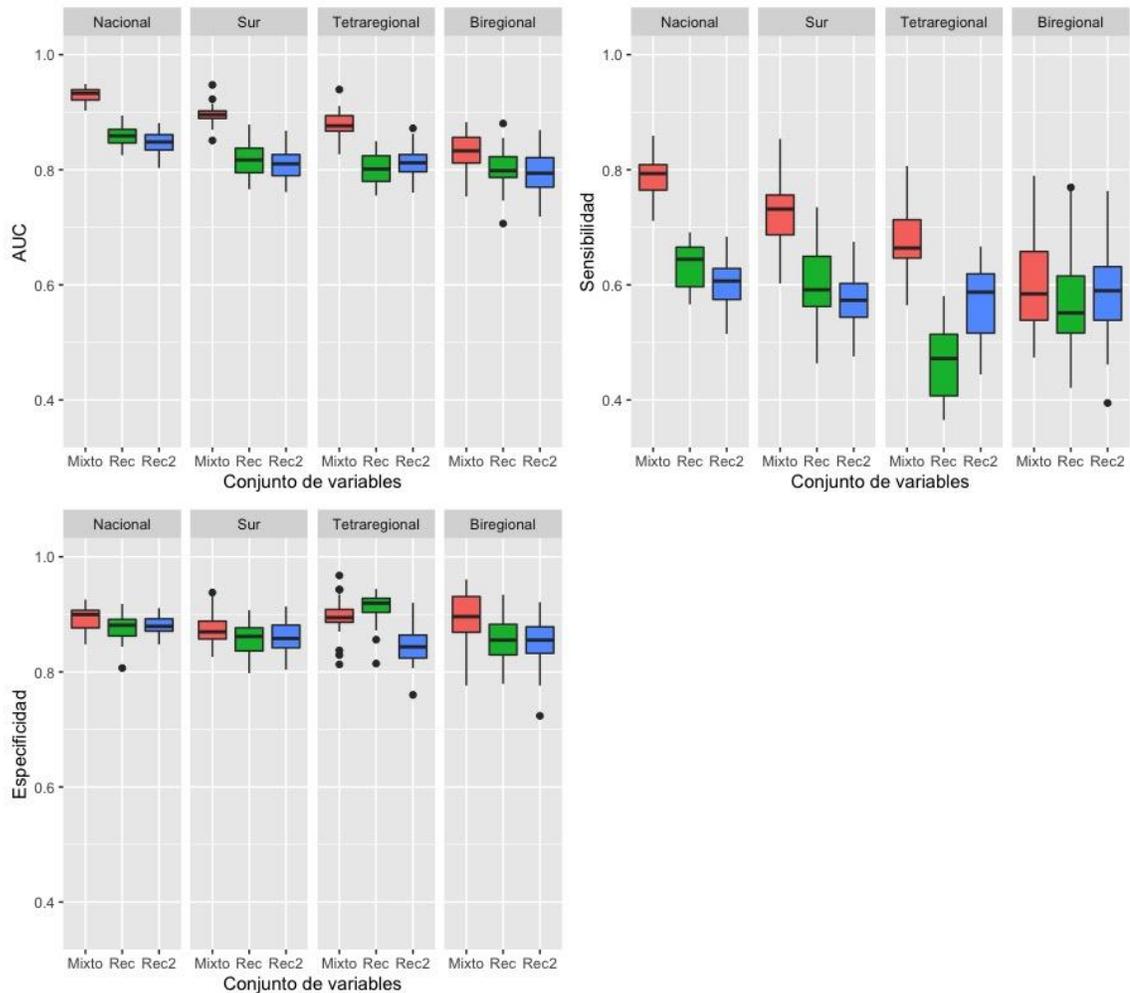


Figura 3. Parámetros de desempeño (AUC, Sensibilidad y Especificidad) para los modelos realizados con cada uno de los conjuntos de variables (Rec, Rec2 y Mixto) sobre las cuatro superficies (Chile, Sur, Tetraregional y Biregional).

La sensibilidad muestra un comportamiento similar al del ROC AUC, disminuyendo en los modelos creados con el conjunto Mixto y Rec a medida que se disminuye la escala espacial, mientras que para los modelos realizados con el

conjunto Rec2 no se muestra una tendencia clara (Fig. 3). El ANOVA anidado arrojó una diferencia significativa en los resultados de sensibilidad tanto entre superficies como para conjuntos de variables dentro de cada superficie (Tabla 3). Sin embargo, el test de Tukey detectó que los resultados de sensibilidad entre superficies no presentan diferencias significativas (Tabla 4). Frente a esto tomaremos en cuenta el resultado del ANOVA anidado (Tabla 3) y consideraremos el resultado con el menor p-valor en el test de Tukey como diferencia significativa que corresponde a la diferencia entre la superficie Tetrarregional y Nacional (Tabla 4). Respecto a los conjuntos de variables, la prueba de Tukey mostró diferencias significativas entre todos los conjuntos en la superficie nacional (Sens: Mixto > Rec > Rec2) (Tabla 5). Para la superficie Sur solo mostró diferencias significativas entre el conjunto Mixto y el resto de los conjuntos (Sens: Mixto > (Rec = Rec2)) (Tabla 5), para la superficie Tetrarregional mostró diferencia significativa entre todos los conjuntos (Sens: Mixto > Rec > Rec2) y en la superficie Birregional no se presentó diferencia significativa entre ningún conjunto (Tabla 5).

Por otro lado, la especificidad arrojó satisfactorios resultados para todos los modelos, posicionándose sobre 0,7 a excepción de un modelo realizado con el conjunto Rec2 en la superficie Tetraregional, y su varianza aumentó conforme disminuía el tamaño de las escalas espaciales (Fig. 3). La prueba de Tukey entre superficies mostró diferencias significativas entre Birregional y Nacional y entre Birregional y Tetrarregional (Espec.: (Nacional = Tetrarregional) > Birregional))

(Tabla 4). Para la comparación entre conjuntos dentro de cada superficie, el análisis de Tukey (Tabla 5) reveló diferencias significativas entre los resultados de especificidad conseguidos con el conjunto Mixto y los demás conjuntos (Espec.: Mixto > (Rec = Rec2)), excepto en la superficie birregional y tetrarregional, donde la Rec2 presente diferencia significativa tanto con el modelo Mixto como con el modelo Rec (Espec.: (Mixto = Rec) > Rec2)).

Tabla 3 Resultados de ANOVA anidada para los conjuntos de variables dentro de las superficies.

Métrica	Variables	DS	Sum sq	Mean sq	F value	P value
AUC	Superficie	3	0.24	0.08	118.09	<2e-16 ***
	Superficie:Conjunto de variables	8	0.39	0.05	71.64	<2e-16 ***
	Residuales	348	0.23	0.00		
Sens	Superficie	3	0.62	0.20	54.75	<2e-16 ***
	Superficie:Conjunto de variables	8	1.70	0.21	56.03	<2e-16 ***
	Residuales	348	1.32	0.00		
Spec	Superficie	3	0.03	0.01	10.44	1.36e-06 ***
	Superficie:Conjunto de variables	8	0.10	0.01	13.38	<2e-16 ***
	Residuales	348	0.35	0.00		

Tabla 4 Resultados de ANOVA y test de Tukey realizado entre los resultados de AUC, Sensibilidad (Sens) y Especificidad (Esp) obtenidos en cada superficie modelada.

Métrica	Nacional-Biregional	Sur-Biregional	Tetraregional-Biregional	Sur-Nacional	Tetraregional-Nacional	Tetraregional-Sur
AUC	0.00**	0.89	0.17	0.00**	0.16	0.54
Sensibilidad	0.31	0.64	0.99	0.94	0.29	0.61
Especificidad	0.00	0.14	0.00**	0.01	0.21	0.53

Tabla 5 Resultados de ANOVA y test de Tukey realizado entre los resultados de AUC, Sensibilidad (Sens) y Especificidad (Esp) obtenidos en cada conjunto de variables modelada.

Métrica	Superficie	Resultados ANOVA		P value test de Tukey		
		F anova	P anova	Rec-Rec2	Rec2-Mixto	Rec-Mixto
AUC	Extensión nacional	217.9	<2e-16 ***	0.03*	0.00***	0.00***
	Zona sur	119.5	<2e-16 ***	0.54	0.00***	0.00***
	Superficie tetra regional	82.53	<2e-16 ***	0.32	0.00***	0.00***
	Superficie bi regional	9.83	0.00***	0.76	0.00**	0.00***

Sens	Extensión nacional	200.6	<2e-16 ***	0.01*	0.00***	0.00***
	Zona sur	68.83	<2e-16 ***	0.06	0.00***	0.00***
	Superficie tetra regional	92.64	<2e-16 ***	0.00***	0.00***	0.00***
	Superficie birregional	1.25	0.29	0.61	0.26	0.8
Spec	Extensión nacional	6.89	<2e-16 ***	0.84	0.00**	0.01*
	Zona sur	3.87	0.00***	0.92	0.03*	0.08
	Superficie tetra regional	35.77	4.6e-16 ***	0.00***	0.32	0.00**
	Superficie birregional	7.60	0.00 ***	0.00**	0.00**	0.82

Importancia de las variables

El rol de las variables predictoras obtenidas por modelo varió entre superficies y conjuntos de variables (Fig. 3). Mientras que para la superficie Nacional las variables más importantes fueron Geología, Clima y Temperatura media anual. En el caso de los modelos Sur y Tetra regional la variable Posición relativa de la sub cuenca fue la más importante para los modelos realizados con Rec y Rec2, lo que varió en el caso del modelo Mixto, donde el modelo Sur presentó la variable Clima como la más importante y en el modelo Tetra regional resultó ser la variable Altitud. El modelo Bi regional arrojó como variable más importante Origen de flujo en Rec y Rec2 y, en el caso del conjunto de variables Mixto la más importante fue Altitud. Vale destacar que las variables complementarias añadidas en el conjunto Mixto presentan una importancia mucho mayor que el resto de las variables de la REC-Chile (Tabla 6).

Tabla 6 Resultados de la importancia de las variables para cada uno de los modelos realizados.

Conjunto de variables	Variables predictoras	Superficie nacional	Zona Sur	Superficie Tetra regional	Superficie Bi regional
Rec	Clima	32.88	21.68	39.07	24.02
	Geología	38.75	31.71	33.29	41.567
	Origen de flujo	35.32	35.70	36.61	47.12
	Pendiente	24.21	31.76	31.28	29.86
	Posición relativa de la sub cuenca	36.90	40.69	53.98	40.62
	Uso de suelo	37.42	35.85	19.24	25.68
	Rec2	Clima	59.48	41.49	36.95
Geología		40.38	35.53	25.15	45.41
Origen de flujo		42.58	38.02	39.00	78.22
Pendiente		44.14	54.18	25.58	43.54
Posición relativa de la sub cuenca		55.14	64.28	59.22	67.047
Uso de suelo		39.06	35.69	16.89	28.47
Caudal		8.96	7.49	8.18	7.73
Mixto	Altitud	91.90	84.88	100	100
	Geología	3.80	5.34	8.89	13.17
	Origen de flujo	5.24	7.42	6.84	18.89
	Pendiente	5.38	8.01	6.0	10.36
	Posición relativa de la sub cuenca	6.24	8.26	9.22	11.14
	Precipitación anual	88.59	80.70	83.52	88.60
	Temperatura media anual	100	100	67.20	79.26
	Uso de suelo	3.20	4.19	2.18	4.96

Evaluación espacial mediante un mapa de predicción de superposición de predicciones

La distribución nacional predicha por el modelo Mixto se ubica entre los ríos Copiapó y Puelo, lo que difiere de las predicciones realizadas por Rec y Rec2, donde la distribución va de Uchusuma (17°32' Sur) a estero Sapos (43°23' Sur) y desde la quebrada de Pintados al río Tictoc, respectivamente, mostrando una mayor área de extensión predicha. Para la predicción realizada con el conjunto Rec (Fig. 4a), se logra apreciar una gran cantidad de tramos con predicción que no coinciden con las predicciones realizadas con los otros conjuntos de variables, sugiriendo una sobre predicción de presencias. Además, se logra ver que los tramos en los que coinciden las predicciones de los tres modelos es en las cuencas de mayor tamaño, como en la del Bio Bío, en la del Choapa o en la del río Calle Calle. Esto se repite en los modelos para todas las superficies.

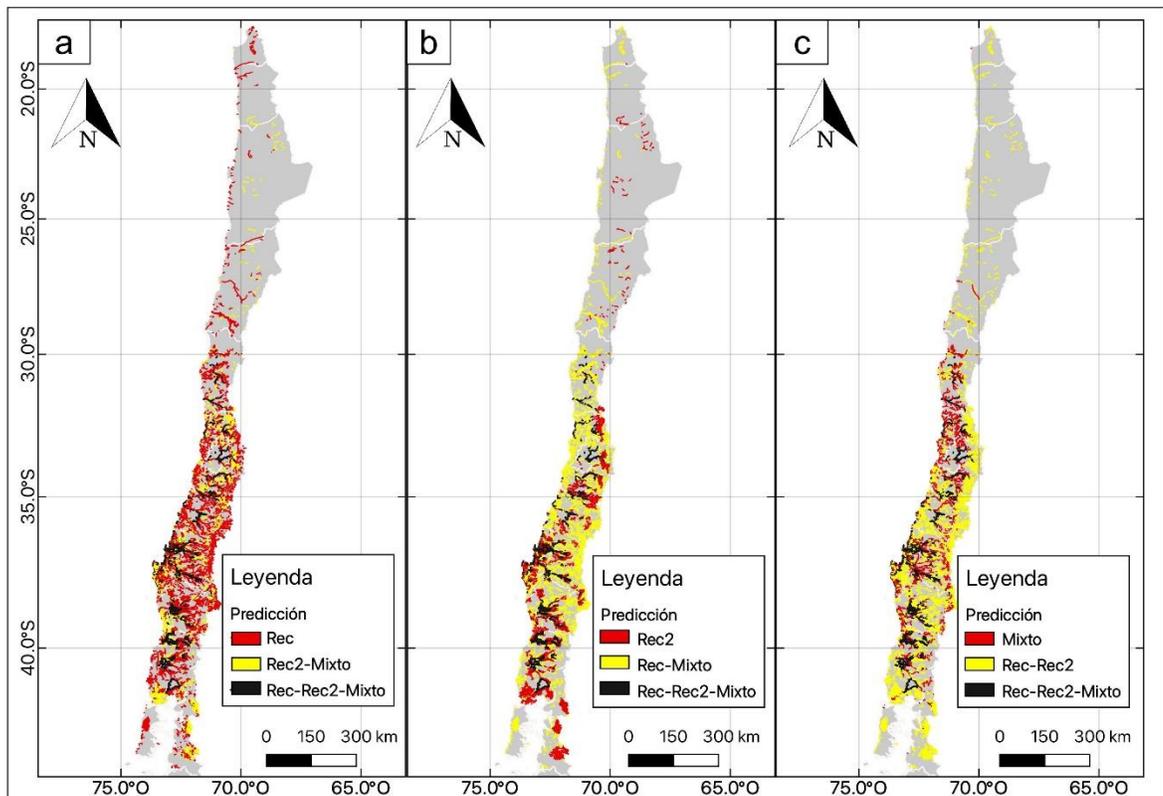


Figura 4. Mapas realizados con las predicciones de los modelos hechos con el conjunto Rec (a), el conjunto Rec2 (b) y el conjunto Mixto (c) para la superficie nacional. En cada uno se muestran los tramos en los que se encontró presencia en rojo, en amarillo las predicciones de los otros, y en negro aparecen los tramos en los que predicen presencia los tres conjuntos.

En la predicción de la zona sur (Fig. 5) encontramos que el conjunto Rec2 predice una menor extensión de ríos como presencia, concentrándose principalmente en tramos de la costa y la cordillera además de las grandes cuencas. En esta zona las predicciones con los conjuntos Rec y Rec2 (Fig. 5a y 5b) son más uniformes.

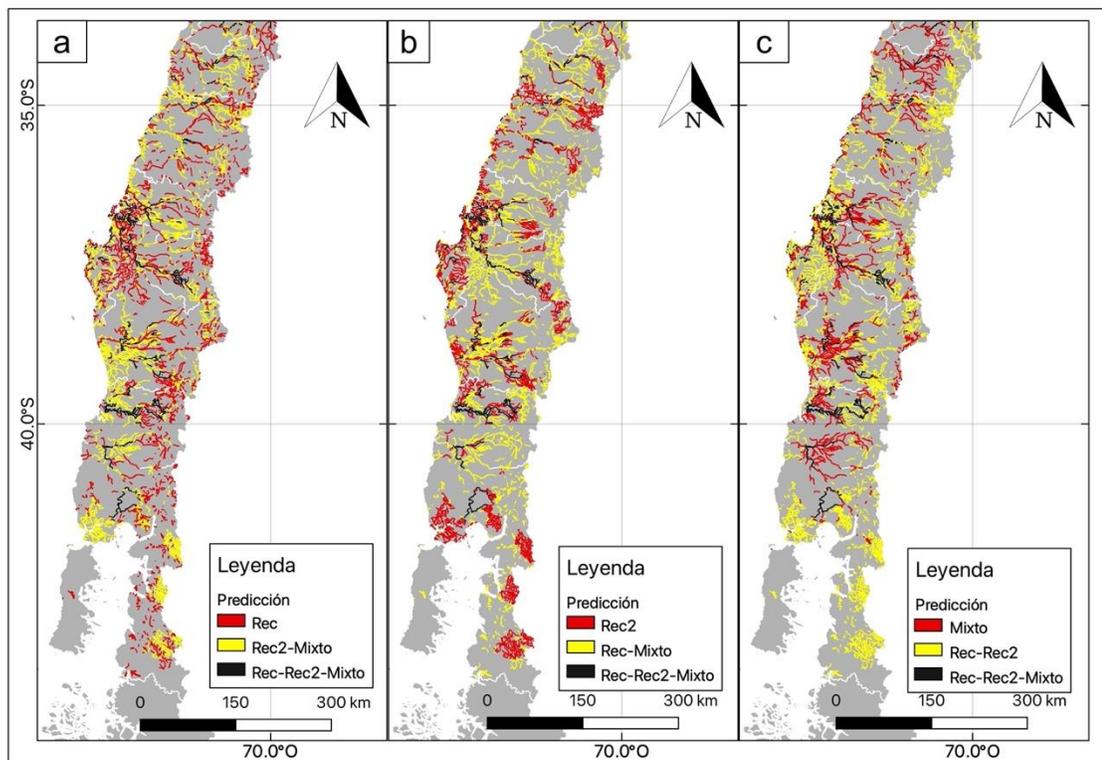


Figura 5. Mapas realizados con las predicciones de los modelos hechos con el conjunto Rec (a), el conjunto Rec2 (b) y el conjunto Mixto (c) para la superficie sur. En

cada uno se muestran los tramos en los que se encontró presencia en rojo y en amarillo las predicciones de los otros. En negro aparecen los tramos en los que predicen presencia los tres conjuntos.

En cuanto a los modelos realizados sobre la superficie tetrarregional (Fig. 7), nos encontramos con que el conjunto Mixto (Fig. 7a) hizo predicciones de presencia en una mayor cantidad de tramos periféricos a las grandes cuencas que los modelos realizados con el conjunto Rec y Rec2, mostrando estas últimas predicciones de presencia principalmente en las grandes cuencas.

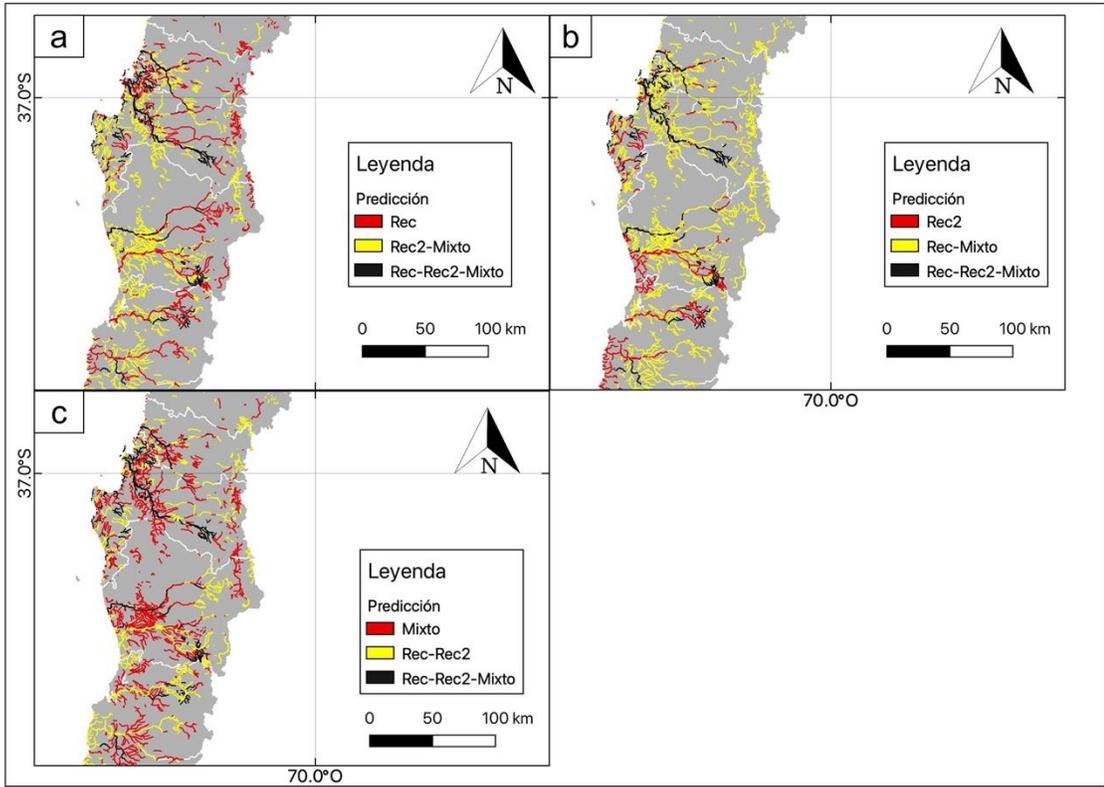


Figura 6. Mapas realizados con las predicciones de los modelos hechos con el conjunto Rec (a), el conjunto Rec2 (b) y el conjunto Mixto (c) para la superficie tetraregional. En cada uno se muestran los tramos en los que se encontró presencia en rojo y en amarillo las predicciones de los otros. En negro aparecen los tramos en los que predicen presencia los tres conjuntos.

Por último, los modelos realizados sobre la superficie birregional (Fig. 7) muestran al conjunto Rec2 nuevamente arroja una mayor cantidad de predicciones de presencia que los otros dos modelos, y al conjunto Rec produciendo una menor cantidad de predicciones de presencia. El conjunto Mixto muestra una predicción de nicho más porosa que los anteriores, es decir, predicciones de presencia menos concentradas y repartidas de manera más homogénea sobre la superficie, obteniendo predicciones de presencia en una mayor cantidad de tramos pertenecientes a pequeñas cuencas en comparación a las predicciones de los otros dos conjuntos.

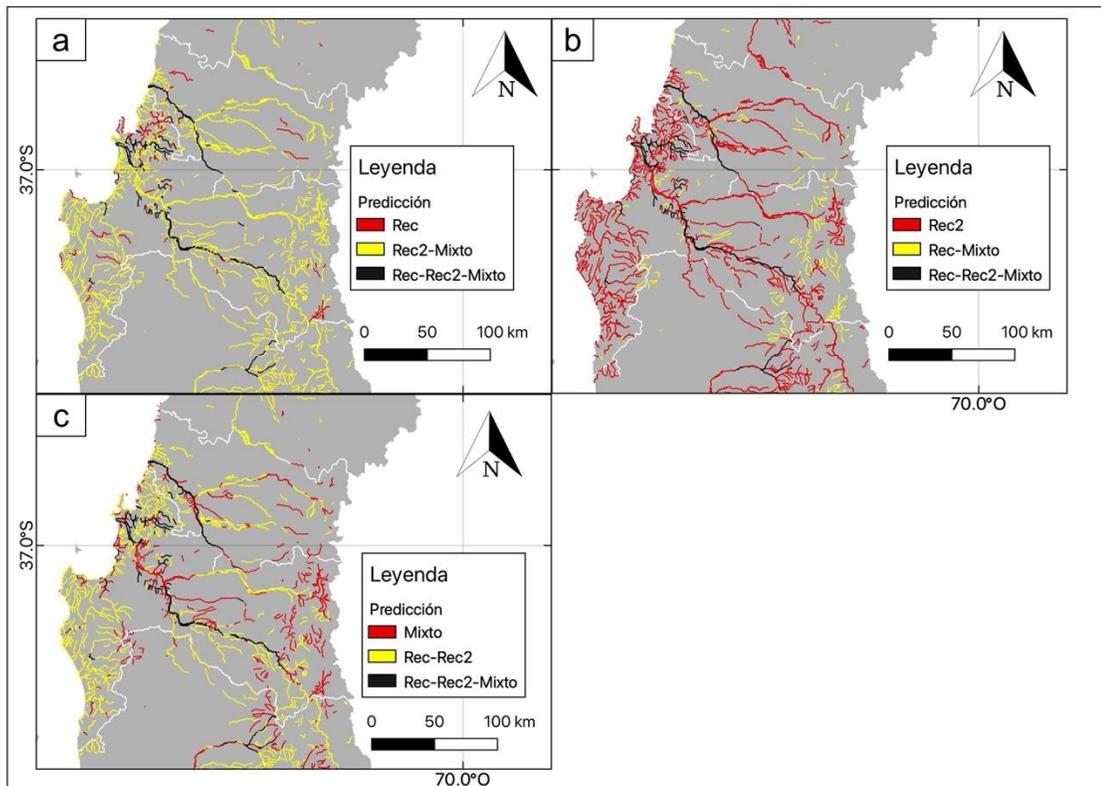


Figura 7. Mapas realizados con las predicciones de los modelos hechos con el conjunto Rec (a), el conjunto Rec2 (b) y el conjunto Mixto (c) para la superficie birregional. En cada uno se muestran los tramos en los que se encontró presencia en rojo y en amarillo las predicciones de los otros. En negro aparecen los tramos en los que predicen presencia los tres conjuntos.

Discusión

Los resultados mostraron que es posible obtener buenos desempeños ($AUC > 0.8$) para SDM de especies ícticas en Chile haciendo uso de la REC-Chile como base de datos de variables ambientales sobre diferentes escalas geográficas, lo que implicaría la posibilidad de utilizar esta base de datos para la modelación de especies con distinto rango de distribución, representando así una herramienta potencialmente valiosa para gestión y manejo de especies ícticas. Al comparar los resultados obtenidos en el presente estudio con los de otros trabajos de SDM para fauna íctica (Por ejemplo: Labay et al., 2011; McNyset, 2005; Radinger et al., 2017), los resultados obtenidos ($0,70 < AUC < 0,93$) son satisfactorios y dentro de los rangos de desempeño obtenidos por otros autores ($0,72 < AUC < 0,98$). Respecto a las variables predictivas, se registró que variables continuas incluidas en los modelos Mixto aumentan su desempeño de manera significativa en todas las escalas evaluadas.

Desempeño de REC-Chile a diferentes escalas

El principal patrón detectado respecto del desempeño de los modelos a distintas escalas espaciales fue la disminución de su capacidad de predicción

conforme se reduce la superficie sobre la que se llevan a cabo, expresado a través de la disminución del AUC y la sensibilidad, mientras que la especificidad se mostró relativamente constante entre las distintas superficies. Este patrón podría ser explicado por la disminución de presencias y pseudo ausencias debido a la disminución de la superficie proyectada (Nacional: 160 presencias, Sur: 85 presencias, Tetra regional: 25 presencias y Bi regional: 18 presencias) y también debido a la relación escala-resolución entre la superficie modelada y la escala de REC-Chile.

Al disminuir la superficie modelada sin variar el tamaño de los tramos, cada modelo a menor escala en nuestro caso cuenta con menor resolución, lo que no es recomendable para obtener resultados consistentes. Si consideramos la propuesta de resolución para distintas escalas de Domisch et al. (2015), tenemos que para la escala continental se recomienda una resolución de 5 a 10 km, mientras que en la escala regional se recomienda una resolución de 1 a 5 km y en la escala paisaje se recomienda una resolución de 100 m a 1 km. En el caso de REC-Chile contamos con tramos de entre 1 a 10 km de longitud, lo que concuerda con las escalas continental y regional planteadas en Domisch et al. (2015), pero podría arrojar problemas a escalas mayores o menores, como es nuestro caso.

Otro factor que podría impactar en nuestros resultados es la relación del rango de hogar de *B. microlepidotus* con la escala de nuestros modelos. Por ejemplo, un artículo realizado en una especie de ciervo (*Cervus elaphus*) registró que ciertas escalas funcionan mejor para determinadas especies según su rango de distribución y que toleran de mejor manera cambios en los tamaños de la muestra (Loe et al., 2012). Lo anterior podría apoyar la idea de que el tamaño de la escala está influyendo en los resultados, más que el tamaño de la muestra y en el caso particular de *B. microlepidotus*, las escalas mayores podrían estar generando mejores resultados debido a su amplia distribución.

El conjunto Rec2 al contrario de los conjuntos Rec y Mixto, produce un aumento en el rendimiento de los modelos según disminuye la superficie. Esto podría estar explicado por la cantidad de categorías en los conjuntos de variables y por la naturaleza de las variables. Mientras el conjunto Rec posee las 6 variables categóricas que poseen entre 3 y 9 categorías cada una, Rec2 posee dos variables más (Régimen hídrico y Caudal), también categóricas, que tienen entre 40 y 120 categorías. Este gran aumento en el número de categorías perjudica la representatividad que pueden llegar a lograr los datos de presencias y pseudoausencias, puesto que al ser tantas, es difícil que queden posicionadas sobre

tramos que contengan todas estas categorías, de esta forma no pudiendo proporcionar la información necesaria para ajustar un modelo predictivo.

Además, un gran número de categorías dificulta la división por parte de los árboles de decisión y es sabido que, si bien, Random Forest muestra mejores desempeños con variables de varias categorías (Strobl et al., 2007), más de 53 de estas afectan su rendimiento, produciendo desempeños no óptimos (Couronné et al., 2018). Así al disminuir las superficies, se comienzan a acotar la cantidad de categorías, dejando disponibles menos categorías en el conjunto Rec2 y esto explicaría, por qué al contrario de Rec, el conjunto Rec2 produce mejores desempeños a menor superficie.

En cuanto al modelo Mixto, podemos ver que en la importancia de variables cómo las variables agregadas del tipo continuas (Precipitación media anual, Temperatura media y Altitud) tienen una importancia considerablemente mayor que las demás (Tabla 6). Esto es debido a que Random Forest considera las categorías de las variables continuas de manera ordinal, de forma que la clasificación resulta en dos conjuntos, uno que consiste en los valores mayores a un número determinado y otro conjunto que consiste en los valores menores a dicho número. De esta forma, las bifurcaciones en el caso de las variables continuas se resumen a mayor o menor que un número y divide los datos de manera más

precisa (Strobl et al., 2007). Es por esto que el modelo Mixto genera mejores resultados, puesto que, contiene variables que le permiten separar y clasificar de mejor manera los datos. Por esto consideramos que en miras de conseguir mejores desempeños en los modelos utilizando REC-Chile, sería mejor optar por la versión continua de sus variables y en caso de no poder obtenerse, quedarse con las categorías originales, puesto que la combinatoria de variables no es favorable para Random Forest. En el caso de querer probar otros algoritmos con la combinación de variables, creando variables con más de 53 categorías, se encontró que una buena opción sería GLM que demostró tener mejor rendimiento que Random Forest para dicho caso (Couronné et al., 2018).

Por último, vale la pena mencionar que los resultados de desempeño obtenidos en este seminario son difíciles de contrastar con la literatura, puesto que existen pocas investigaciones respecto a sistemas fluviales y la exploración del efecto que puedan tener los cambios de resolución o escala en la realización de modelos para estos sistemas. Por esto resulta de gran interés seguir investigando acerca del efecto de las escalas y resoluciones sobre el desempeño de los modelos de distribución de especies ícticas en Chile, además de utilizar otros algoritmos para evaluar si algunos se ajustan mejor a un tipo de escala o a otra.

Importancia de variables

Los modelos desarrollados presentan una importancia de variables en concordancia con aspectos de la ecología de la especie encontrados en la literatura. Por ejemplo, la preferencia de *B. microlepidotus* por aguas bien oxigenadas (Rojas et al., 2016) y la temperatura, que influye directamente sobre la concentración de oxígeno en el agua (Kramer, 1987). Esta variable resultó ser la más importante en el modelo realizado con el conjunto Mixto sobre las superficies Nacional y Sur. Por otro lado, la preferencia por aguas de bajas velocidades y transparentes (Rojas et al., 2016) se relaciona con la Posición relativa de la subcuenca, variable más importante para los modelos realizados con el conjunto Rec y Rec2 sobre las superficies Sur y Tetra regional. Dado que Posición relativa de la subcuenca caracteriza un tramo por su ubicación (Cabecera, Tramo alto, Tramo medio, Tramo bajo, Endorreica), lo que a su vez se encuentra directamente relacionada con la pendiente, disminuyendo conforme nos acercamos a la costa (Mao, 2016), lo que tiene un efecto directo sobre la velocidad y la transparencia del agua (Vannote et al., 1980).

Entre los trabajos de SDM en esta especie, encontramos dos en los que desarrollan un modelo de predicción de nicho a escala de paisaje. Uno de estos trabajos (Rojas Venegas & Espinoza, 2018) plantea una proyección de nicho a

escala de tramo mediante la función de idoneidad multivariada y la función de idoneidad de hábitat ambiental multivariado dentro de la cuenca del río Mataquito. Este trabajo aporta información valiosa referente a cómo se constituye el nicho de *B. microlepidotus* dentro de una cuenca, determinándose que las variables que lo modulan en mayor medida son la velocidad del agua y la temperatura, lo que coincide con nuestros resultados si consideramos que la Posición relativa de la sub cuenca podría representar un buen proxy de ambas variables.

En otro trabajo (Zamorano et al., 2019) realizaron modelos de nicho para distintas especies de peces chilenos mediante el uso de redes neuronales, dentro de los cuales encontramos el modelo de *B. microlepidotus*. En este último trabajo detectó que las variables predictivas más importantes a escala de paisaje para esta especie son la altitud que también coincide con nuestros resultados, y el ancho del canal, que podríamos homologar a la posición relativa de la sub cuenca puesto que el ancho del canal aumenta conforme el un tramo de río se acerca a la desembocadura (Vannote et al., 1980). De esta forma el ancho del canal también coincidiría con nuestros resultados.

Distribución predicha para *B. microlepidotus* en Chile

Cuando evaluamos las superficies proyectadas, encontramos que el modelo Mixto arroja las predicciones de distribución geográficas acotadas dentro de los rangos de distribución histórica tanto latitudinal como longitudinalmente (Ministerio del Medio Ambiente. 2007, Ministerio del Medio Ambiente. 2013) para la superficie Nacional y Sur, tal como se esperaría producto de tener un alto AUC, acompañado de una alta sensibilidad y especificidad. Además, los modelos Mixtos llevados a cabo en la superficie tetra regional y Bi regional también se ajustan bien a la superficie histórica a pesar de tener baja sensibilidad. Esto podría ocurrir porque la superficie histórica está más bien acotada a la zona central de la superficie de Chile y un modelo de alta especificidad y baja sensibilidad debería arrojar una distribución acotada, coincidiendo de esta forma la predicción con la distribución histórica en superficies pequeñas.

Los demás modelos arrojaron resultados visuales con sobre predicciones de presencias o de ausencias, según se esperaría por sus resultados de especificidad y sensibilidad, tal como la sobre predicción de ausencias del modelo con Rec2 en la superficie Nacional y la sobre predicción de ausencias del modelo Rec en la superficie Bi regional. Sin embargo, hubo resultados que llamaron particu-

larmente nuestra atención, dado que produjeron una predicción geográfica contraria a la esperada. Esto se vio en los resultados arrojados por el conjunto Rec para las superficies Nacional y Sur, puesto que, al contar con una alta especificidad y menor sensibilidad, debería mostrar una visualización con distribuciones más bien acotadas a ciertos sectores, como el conjunto Rec2 (Fig. 4b), lo que no ocurrió. Además, el modelo Rec2 con alta especificidad y baja sensibilidad debería mostrar un modelo de predicciones acotadas, sin embargo, resultó ser una predicción que se visualizó como una amplia distribución de la especie. Esto podría estar relacionado con la variabilidad que existe en los tamaños de los tramos, de forma que, si un modelo hace predicciones de presencia en tramos particularmente grandes, el resultado visual será una distribución exagerada en comparación a predicciones de la misma cantidad de presencias, pero en tramos pequeños. Por el contrario, si las predicciones de ausencia están concentradas en los tramos de mayor tamaño, y las presencias se encuentran en los tramos de menor tamaño, se generará un resultado visual en que la presencia esté exageradamente acotada. Estos escenarios podrían generar que los resultados numéricos de métricas como sensibilidad o especificidad no se ajusten a la extensión del mapa predicho, dificultando la interpretación de los resultados.

Protocolo y aplicación

Luego de realizar los modelos haciendo uso de diferentes conjuntos de variables sobre superficies de distinta escala se pueden rescatar algunas recomendaciones a seguir para la realización de SDM de especies chilenas en ecosistemas lóticos, utilizando REC-Chile como base de datos ambientales:

1.- En primer lugar, los registros de presencias en Chile son escasos. La única opción para encontrar una cantidad de registros que permitan realizar un buen modelo de distribución son los registros que posee el Ministerio del Medio Ambiente y la Subsecretaría de Pesca (permisos de pesca), para luego añadir registros que se puedan agregar desde registros personales o de la literatura que no estén contemplados dentro de la base de datos del MMA o Subpesca. De esta forma podemos optar a obtener de manera inmediata una base de registros ya realizada, que reúne tanto registros históricos como de avistamientos más recientes. Si bien especies como *B. microlepidotus* cuentan con mayor cantidad de registros dada su abundancia y rango, especies en categoría de conservación no cuentan con igual cantidad de registros en esta base de datos, pudiendo ser un desafío en casos de modelos con fines de conservación.

2.- En segundo lugar, algunos registros históricos son desde hace más de 100 años atrás, por lo que es recomendable hacer un filtro por fecha de registro. En

el caso de este seminario, se consideraron los registros de los últimos 30 años, con lo que se pudo obtener un modelo de buen desempeño acorde a la realidad. Modelos con datos previos fueron realizados, obteniendo resultados de AUC parecidos a los obtenidos con los datos de los últimos 30 años (0,7-0,85) (datos no publicados). Sin embargo, al proyectarlos, encontramos sectores con predicción de presencia en los que hace muchos años no existen nuevos registros, como es el caso de algunos tramos del río Elqui, mostrando finalmente modelos de alto desempeño, pero representando un nicho que existía en otra época en presencia de condiciones ambientales diferentes.

4.- En cuanto a la selección de variables, este seminario entrega información de interés, mostrando que REC-Chile es una buena herramienta para obtener variables eco hidrológicas. No obstante, el modelo con mejor desempeño fue el que se desarrolló en base a REC-Chile y variables agregadas del tipo continuas. Por esto se recomienda complementar REC-Chile con variables de temperatura y, en caso de ser posible, reemplazar algunas variables de REC-Chile por su versión continua, ya que proporciona mejores desempeños.

5.- Por último, esta metodología debería ser una guía para generar modelos de distribución de fauna íctica de buen desempeño en Chile, pudiendo ser utilizado con fines tanto de conservación como de manejo, aportando información clave a

la hora de crear o modificar áreas protegidas, llevar a cabo mejores labores de fiscalización, servir de soporte a estudios de impacto ambiental, ayudar a focalizar esfuerzos de muestreo en zonas con alta probabilidad de presencia o contrastar resultados de líneas de base con modelos de distribución de dichas especies.

Conclusión

Este seminario ha mostrados que con el uso de Random Forest y REC-Chile es posible realizar buenos modelos de distribución para la ictiofauna chilena a distintas escalas espaciales, generando mejores desempeños los modelos en mayores superficies. Si bien REC-Chile como base de datos de variables ecohidrológicas arroja buenos resultados por sí sola, la complementación con variables climáticas y/o continuas la hace mucho más efectiva. De esta manera se puede concluir que siguiendo las recomendaciones propuestas en este seminario es posible desarrollar buenos modelos de distribución de especies ícticas en Chile.

Bibliografía

- Arratia, G., Vila, I., Lam, N., Guerrero, C. J., & Quezada-Romegialli, C. (2017). Morphological and taxonomic descriptions of a new genus and species of killifishes (Teleostei: Cyprinodontiformes) from the high Andes of northern Chile. *PLOS ONE*, 12(8), e0181989. <https://doi.org/10.1371/journal.pone.0181989>
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, 20(1), 1–9. <https://doi.org/10.1111/ddi.12144>
- Chamier, A.-C., Dixon, P. A., & Archer, S. A. (1984). The spatial distribution of fungi on decomposing alder leaves in a freshwater stream. *Oecologia*, 64(1), 92–103. <https://doi.org/10.1007/BF00377550>

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*, 19(1), 270. <https://doi.org/10.1186/s12859-018-2264-5>

Cruz-Jofré, F., Morales, P., Vila, I., Esquer-Garrigos, Y., Hugueny, B., Gaubert, P., Poulin, E., & Méndez, M. A. (2016). Geographical isolation and genetic differentiation: The case of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), an Andean killifish inhabiting a highland salt pan. *Biological Journal of the Linnean Society*, 117(4), 747–759. <https://doi.org/10.1111/bij.12704>

Davinson, I. (1991). Environmental effects on algal Photosynthesis: Temperature. *Journal of Phycology*, 27, 2–8.

Domisch, S., Jähnig, S. C., Simaika, J. P., Kuemmerlen, M., & Stoll, S. (2015). Application of species distribution models in stream ecosystems: The challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundamental and Applied Limnology / Archiv Für Hydrobiologie*, 186(1), 45–61. <https://doi.org/10.1127/fal/2015/0627>

EIGENMANN, C. 1927. The freshwater fishes of Chile. *Memories of the National Academy of Sciences* 22: 1-63.

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

Elliott, S. (2010). *El río y la Forma. Introducción a la Geomorfología Fluvial*. RiL Editores. Chile.

Esselman, P. C., & Allan, J. D. (2011). Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica: Freshwater protection areas for the fishes of northeastern Mesoamerica. *Freshwater Biology*, 56(1), 71–88. <https://doi.org/10.1111/j.1365-2427.2010.02417.x>

Fei, S., & Yu, F. (2016). Quality of presence data determines species distribution model performance: A novel index to evaluate data quality. *Landscape Ecology*, 31(1), 31–42. <https://doi.org/10.1007/s10980-015-0272-7>

Figuroa, R., Bonada, N., Guevara, M., Pedreros, P., Correa-Araneda, F., Díaz, M. E., & Ruiz, V. H. (2013). Freshwater biodiversity and conservation in mediterranean climate streams of Chile. *Hydrobiologia*, 719(1), 269–289. <https://doi.org/10.1007/s10750-013-1685-4>

Fuster, R., Escobar, C., Lillo, G., & de la Fuente, A. (2015). Construction of a

typology system for rivers in Chile based on the European Water Framework Directive (WFD). *Environmental Earth Sciences*, 73(9), 5255–5268. <https://doi.org/10.1007/s12665-014-3772-x>

Gonzalez, V. H., Koch, J. B., & Griswold, T. (2010). *Anthidium vigintiduopunctatum* Friese (Hymenoptera: Megachilidae): the elusive “dwarf bee” of the Galápagos Archipelago? *Biological Invasions*, 12(8), 2381–2383. <https://doi.org/10.1007/s10530-009-9651-9>

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)

Habit, E., Dyer, B., & Vila, I. (2006). Estado de conocimiento de los peces dulceacuícolas de Chile. *Gayana (Concepción)*, 70(1). <https://doi.org/10.4067/S0717-65382006000100016>

Hernández-Gómez, O., Wuerthner, V., & Hua, J. (2020). Amphibian Host and Skin Microbiota Response to a Common Agricultural Antimicrobial and Internal Parasite. *Microbial Ecology*, 79(1), 175–191. <https://doi.org/10.1007/s00248-019-01351-5>

Kramer, D. L. (1987). Dissolved oxygen and fish behavior. *Environmental Biology of Fishes*, 18(2), 81–92. <https://doi.org/10.1007/BF00002597>

Kuemmerlen, M., Stoll, S., Sundermann, A., & Haase, P. (2016). Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. *Ecological Indicators*, 65, 122–132. <https://doi.org/10.1016/j.ecolind.2015.08.008>

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5). <https://doi.org/10.18637/jss.v028.i05>

Labay, B., Cohen, A. E., Sissel, B., Hendrickson, D. A., Martin, F. D., & Sarkar, S. (2011). Assessing Historical Fish Community Composition Using Surveys, Historical Collection Data, and Species Distribution Models. *PLoS ONE*, 6(9), e25145. <https://doi.org/10.1371/journal.pone.0025145>

Lauzeral, C., Grenouillet, G., & Brosse, S. (2013). Spatial range shape drives the grain size effects in species distribution models. *Ecography*, 36(7), 778–787. <https://doi.org/10.1111/j.1600-0587.2013.07696.x>

Leopold, L. B. (1969). Quantitative comparison of some aesthetic factors among rivers (Vol. 620). US Geological Survey.

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. En Y. Xiang & B. Chaib-draa (Eds.), *Advances in Artificial Intelligence* (Vol. 2671, pp. 329–341). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44886-1_25

Liotta, J. (2006). Distribución geográfica de los peces de aguas continentales de la República Argentina.

Loe, L. E., Bonenfant, C., Meisingset, E. L., & Mysterud, A. (2012). Effects of spatial scale and sample size in GPS-based species distribution models: Are the best models trivial for red deer management? *European Journal of Wildlife Research*, 58(1), 195–203. <https://doi.org/10.1007/s10344-011-0563-5>

Mao, L. (2016). Sedimentos, Hidráulica y Efectos Morfológicos. Ministerio de Energía, Gobierno de Chile. https://cambioglobal.uc.cl/images/proyectos/MesaHidro_Sedimentos_hidr%C3%A1ulica_y_efectos_morfol%C3%B3gicos.pdf

Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. (2011). Modelos de distribución de especies: Una revisión sintética. *Revista Chilena de Historia Natural*, 84(2), 217–240. <https://doi.org/10.4067/S0716-078X2011000200008>

McNyset, K. M. (2005). Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. *Ecology of Freshwater Fish*, 14(3), 243–255. <https://doi.org/10.1111/j.1600-0633.2005.00101.x>

Mertes, K., & Jetz, W. (2018). Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, 41(10), 1604–1615. <https://doi.org/10.1111/ecog.02871>

Mi, C., Huettmann, F., Guo, Y., Han, X., & Wen, L. (2017). Why choose Random

Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, 5, e2849. <https://doi.org/10.7717/peerj.2849>

Ortega, H., & Hidalgo, M. (2008). Freshwater fishes and aquatic habitats in Peru: Current knowledge and conservation. *Aquatic Ecosystem Health & Management*, 11(3), 257–271. <https://doi.org/10.1080/14634980802319135>

Palminteri, S., Powell, G. V. N., Asner, G. P., & Peres, C. A. (2012). LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sensing of Environment*, 127, 98–105. <https://doi.org/10.1016/j.rse.2012.08.014>

Peredo Parada, M. M. (2010). Implementación de una clasificación Eco-Hidrológica de los ríos de Chile y su aplicación a la gestión ambiental [Universitat Politècnica de València]. <https://doi.org/10.4995/Thesis/10251/11515>

Peredo-Parada, M., Martínez-Capel, F., Garófano-Gomez, V., Atenas, M., & Riestra, F. (2009). Base de datos eco-hidrológica de los ríos de Chile: Una herramienta de gestión para los ecosistemas acuáticos. *Gayana (Concepción)*, 73(1), 119–129. <https://doi.org/10.4067/S0717-65382009000100015>

Peredo-Parada, Matías, Martínez-Capel, F., Quevedo, D. I., & Hernández-Mas-

carell, A. B. (2011). Implementation of an eco-hydrological classification in Chilean rivers. http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-65382011000100003&lng=en&nrm=iso&tlng=en

Peters, J., Baets, B. D., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. D., & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, 207(2–4), 304–318. <https://doi.org/10.1016/j.ecolmodel.2007.05.011>

Phillips, N. D., Reid, N., Thys, T., Harrod, C., Payne, N. L., Morgan, C. A., White, H. J., Porter, S., & Houghton, J. D. R. (2017). Applying species distribution modelling to a data poor, pelagic fish complex: The ocean sunfishes. *Journal of Biogeography*, 44(10), 2176–2187. <https://doi.org/10.1111/jbi.13033>

Porter, M. S., Rosenfeld, J., & Parkinson, E. A. (2000). Predictive Models of Fish Species Distribution in the Blackwater Drainage, British Columbia. *North American Journal of Fisheries Management*, 12.

Pouteau, R., Trueba, S., Feild, T. S., & Isnard, S. (2015). New Caledonia: A Pleistocene refugium for rain forest lineages of relict angiosperms. *Journal of Biogeography*, 42(11), 2062–2077. <https://doi.org/10.1111/jbi.12581>

Radinger, J., Essl, F., Hölker, F., Horký, P., Slavík, O., & Wolter, C. (2017). The future distribution of river fish: The complex interplay of climate and land use

changes, species dispersal and movement barriers. *Global Change Biology*, 23(11), 4970–4986. <https://doi.org/10.1111/gcb.13760>

Rojas, P. M., Saavedra, S., & Munoz, C. (2016). Progress in farming of Chilean silverside *Basilichthys microlepidotus* Jenyns, 1841: An alternative for productive diversification. *Latin American Journal of Aquatic Research*, 44(2), 342–354. <https://doi.org/10.3856/vol44-issue2-fulltext-15>

Rojas Venegas, P. M., & Espinoza, C. D. (2018). Habitat suitability for Chilean silverside (*Basilichthys microlepidotus*) in the Mataquito River, Maule Region: A geospatial approach. *Marine and Freshwater Research*, 69(8), 1301. <https://doi.org/10.1071/MF17263>

Sherwood, J. L., Stites, A. J., Dreslik, M. J., & Tiemann, J. S. (2018). Predicting the range of a regionally threatened, benthic fish using species distribution models and field surveys. *Journal of Fish Biology*, 93(5), 972–977. <https://doi.org/10.1111/jfb.13819>

Soto, D., Arismendi, I., González, J., Sanzana, J., Jara, F., Jara, C., Guzman, E., & Lara, A. (2006). Southern Chile, trout and salmon country: Invasion patterns and threats for native species. *Revista Chilena de Historia Natural*, 79(1). <https://doi.org/10.4067/S0716-078X2006000100009>

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285–1293. <https://doi.org/10.1126/science.3287615>

Taylor, A. T., Papeş, M., & Long, J. M. (2018). Incorporating fragmentation and non-native species into distribution models to inform fluvial fish conservation: Distribution Models of a Fluvial Fish. *Conservation Biology*, 32(1), 171–182. <https://doi.org/10.1111/cobi.13024>

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(1), 130–137. <https://doi.org/10.1139/f80-017>

Vasconcelos, T. S., Rodríguez, M. Á., & Hawkins, B. A. (2012). Species distribution modelling as a macroecological tool: A case study using New World amphibians. *Ecography*, 35(6), 539–548. <https://doi.org/10.1111/j.1600-0587.2011.07050.x>

Véliz, D., Catalán, L., Pardo, R., Acuña, P., Díaz, A., Poulin, E., & Vila, I. (2012). The genus *Basilichthys* (Teleostei: Atherinopsidae) revisited along its Chilean distribution range (21° to 40° S) using variation in morphology and mtDNA El género

Basilichthys (Teleostei: Atherinopsidae) analizado a lo largo de su distribución en Chile (21o a 40o S), utilizando rasgos morfológicos y variabilidad del ADN mitocondrial. 12.

Vila, I., Fuentes, L., & Contreras, M. (1999). Peces límnicos de Chile. Boletín del Museo Nacional de Historia Natural, (Boletín 48). 61–75.

Vila, I., Veloso A., Schlatter R. y C. Ramirez. 2006. Macrófitas y vertebrados de los sistemas límnicos de Chile. Editorial Universitaria- 187 pp.

Zamorano, D., Labra, F. A., Villarroel, M., Lacy, S., Mao, L., Olivares, M. A., & Peredo-Parada, M. (2019). Assessing the effect of fish size on species distribution model performance in southern Chilean rivers. PeerJ, 7, e7771. <https://doi.org/10.7717/peerj.7771>