

MULTIVARIATE APPROACHES IN SPECIES DISTRIBUTION MODELLING: APPLICATION TO NATIVE FISH SPECIES IN MEDITERRANEAN RIVERS

Presented by: **Rafael Muñoz Mas**

Supervisor: **Francisco Martínez Capel**

RESUMEN

A nivel mundial, las actividades humanas, la extracción de agua, el almacenamiento, los picos de caudal y el cambio climático están afectando negativamente la biota dulceacuícola. El desarrollo de sofisticadas herramientas científicas para evaluar las amenazas representadas por estos fenómenos ha devenido un área fundamental en la investigación científica. La simulación del hábitat físico ha demostrado ser adecuada para la evaluación de los efectos derivados de distintas alternativas, restauraciones o del cambio climático. Los métodos basados en el hábitat físico, que pueden ser desarrollados a distintas escalas, evalúan la cantidad y calidad (*i.e.* idoneidad) del hábitat físico para diferentes caudales integrando hidrología, hidráulica e información biológica (*i.e.* modelos de idoneidad del hábitat). El desarrollo de los modelos de idoneidad del hábitat puede ser enfocado de muchos modos pero aquellos basados en el uso de técnicas de aprendizaje computarizado han ido ganando adeptos, de este modo son actualmente práctica habitual.

La literatura científica engloba numerosos estudios empleando técnicas de aprendizaje computarizado. No obstante existen aún lagunas en el conocimiento y espacio para testar viejos y nuevos métodos con el objetivo último de explorar la ecología de las especies dulceacuícolas o para la evaluación de caudales ecológicos. Esta tesis se centra en el análisis comprensivo de las capacidades de algunos tipos de Red Neuronal Artificial aún no testados: las Redes Neuronales Probabilísticas (PNN) y los Conjuntos de Perceptrones Multicapa (MLP Ensembles). Los análisis sobre las capacidades de estas técnicas se desarrollaron utilizando la trucha común (*Salmo trutta*; Linnaeus, 1758), la bermejuela (*Achondrostoma arcasii*; Robalo, Almada, Levy & Doadrio, 2006) y el barbo colirrojo (*Barbus haasi*; Mertens, 1925) como especies nativas objetivo. Los análisis se centraron en la capacidad de predicción, la interpretabilidad de los modelos y el efecto del exceso de ceros en las bases de datos de entrenamiento, la así llamada

prevalencia de los datos (*i.e.* la proporción de casos de presencia sobre el conjunto total). Finalmente, el efecto de la escala (micro–escala o escala de microhábitat y meso–escala) en los modelos de idoneidad del hábitat y consecuentemente en la evaluación de caudales ambientales se estudió en el último capítulo.

PNN presentó un buen balance entre precisión y generalización. Dado que el desempeño en el sitio de evaluación fue mayor que el conseguido durante el entrenamiento del modelo. La capacidad clasificatoria no se vio significativamente afectada por la prevalencia. No obstante, este tipo de redes presentó ciertas limitaciones en lo que respecta predicción de probabilidades (*i.e.* presentaron un rango de salida recortado), lo cual podría limitar su aplicación en evaluación de caudales ambientales siempre y cuando la base de datos de entrenamiento presente un gran solapamiento entre categorías.

El paradigma de las MLP Ensembles presentó mejor desempeño que el conseguido con las PNN en problemas de clasificación además de funcionar bien en la regresión de densidades de peces. La selección activa de los MLPs finalmente incluidos en el conjunto (Ensemble) presentó mejor desempeño que incluir el conjunto todos los candidatos entrenados. Eso destaca los beneficios del enfoque de sobreproducción–y–elección, que consiste en iniciar primero un conjunto grande de MLPs, seleccionando a posteriori aquel subconjunto que presenta el mejor desempeño combinado. Se testaron dos métodos para llevar a cabo esta selección; la selección basada en el paso delante (Forward) demostró un desempeño marginalmente superior al método empleando cadenas de bits y un Algoritmo Genético (GA). No obstante, existen diversos métodos para codificar y optimizar MLP Ensembles con GAs. Así el último análisis utilizó el método de Wang and Alhamdoosh. Este método incrementa iterativamente el tamaño del Ensemble con la adición paso a paso de MLP, mientras que en cada iteración el GA busca la mejor combinación de ellos. Los MLP Ensembles optimizados por este método fueron en todo caso precisos y de pequeña dimensión por tanto, a pesar de no existir comparación con los métodos anteriores, se consideró que es un enfoque digno de mención. Los MLP Ensembles demostraron ser capaces de lidiar con bases de datos desbalanceadas, tanto con exceso de ceros como con una baja prevalencia. Consecuentemente, debido a su alta capacidad predictiva y su habilidad para lidiar con bases de datos desbalanceadas el paradigma de los MLP Ensembles debe ser considerado una herramienta prometedora en la ecología.

Por lo que respecta a la idoneidad del hábitat modelada y la ecología de los peces, la trucha común no difirió significativamente del conocimiento previo que se tenía de la especie. De este modo la máxima idoneidad se asignó a

microhábitats relativamente profundos con flujo de caudal lento y sustrato medio o grueso.

Las preferencias de microhábitat de la bermejuela no habían sido estudiadas con anterioridad lo que proporciona información valiosa sobre la idoneidad del hábitat para esta especie. La bermejuela puede ser clasificada como una especie limnofílica orientada hacia el refugio ya que este devino el elemento más importante del modelo. El resto de variables indicaron la máxima idoneidad en microhábitat lentos y profundos, relacionados con la deposición natural de limos en el lecho del río.

La preferencias de hábitat del barbo colirrojo fueron exploradas a meso–escala utilizando la densidad de estos como variable dependiente. El barbo colirrojo prefirió segmentos de río en el tramo medio y alto pero no los más altos y empinados. La importancia de las variables confirmó la preferencia del barbo colirrojo por los hábitats de tipo poza. Finalmente, la anguila y el resto de especies de ciprínidos presente en los sitios de estudio demostraron similares requerimientos de hábitat porque el MLP Ensemble encontró una gran correlación entre sus densidades.

La idoneidad del hábitat (*i.e.* probabilidad de presencia) del barbo colirrojo fue modelada a dos escalas espaciales diferentes (la meso–escala y la micro–escala). El MLP Ensemble a micro–escala mostró alta idoneidad en áreas relativamente profundas con sustrato grueso y corroboró la naturaleza reofílica del barbo colirrojo así como su preferencia por el refugio. El modelo a meso–escala destacó las ventajas de cruzar escalas de trabajo ya que la elevación (una variable de macro–escala) fue seleccionada en el modelo óptimo sugiriendo que el barbo colirrojo elude los tramos más altos muestreados. El barbo colirrojo seleccionó áreas profundas y, a esta escala, el MLP Ensemble parcialmente contradijo su contraparte a micro–escala. En este caso la velocidad tuvo un efecto claramente positivo en la idoneidad del hábitat, lo que nosotros relacionamos con el caudal circulante en el año de muestreo. Finalmente, el barbo colirrojo mostró preferencia por sustratos finos, lo que se puede asociar a los diferentes tipos de sustrato (parches) que aparecen en cada tipo de mesohabitat.

El último elemento analizado fue la influencia de la escala de modelización en la evaluación de caudales ambientales ya que algunos autores han sugerido que la meso–escala deviene en mejores resultados. La legislación española determina que la evaluación de caudales ambientales basada en la simulación del hábitat físico debe llevarse a cabo a la escala de microhábitat. En nuestro caso de estudio el modelo a micro–escala predijo un caudal ambiental mayor. Por tanto, desde un punto de vista legal, esta escala debe ser considerada una elección, al menos, igualmente adecuada que la meso–escala.

Esta tesis compila valiosa información desde un punto de vista metodológico y proporciona conocimiento ecológico relevante que debería priorizar seguimientos adecuados y, en última instancia, acciones de manejo ecológicamente amigables.

