نسخه قبل از انتشار

مروری بر مدلسازی توزیع گونه‌ای آبزیان و مهمترین روش‌های مورد استفاده در آن

فاتح معزی[۱]، هادی پورباقر[۲*]، سهیل ایگدری[۳]، جهانگیر فقهی[۴]

[۱]: گروه شیلات، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: moezifateh@ut.ac.ir

[۲]: گروه شیلات، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: poorbagher@ut.ac.ir

[۳]: گروه شیلات، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: soheil.eagderi@ut.ac.ir

[۴]: گروه مهندسی جنگلداری و اقتصاد جنگل، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: : jfeghhi@ut.ac.ir

[*]: نویسنده مسئول: رایانامه: poorbagher@ut.ac.ir

A review on the aquatics species distribution modelling and its most important methods

Fateh Moëzzi[1], Hadi Poorbagher*[2], Soheil Eagderi[3], Jahangir Feghhi[4]

[1]: Department of Fisheries, Faculty of Natural Resources, University of Tehran, Karaj, Iran. Email: moezifateh@ut.ac.ir

[2]: Department of Fisheries, Faculty of Natural Resources, University of Tehran, Karaj, Iran. Email: poorbagher@ut.ac.ir

[3]: Department of Fisheries, Faculty of Natural Resources, University of Tehran, Karaj, Iran. Email: Soheil.eagderi@ut.ac.ir

[4]: Department of Forestry and Forest Economics, Faculty of Natural Resources, University of Tehran, Karaj, Iran. Email: jfeghhi@ut.ac.ir

[*]: corresponding author: Email: poorbagher@ut.ac.ir

چکیده:

شناخت کافی از الگوهای توزیع گونه‌ای آبزیان در اکوسیستم‌های آبی و تغییرات زمانی و مکانی آن‌ها در گذشته و همچنین پیش‌بینی آن‌ها در آینده یکی از مبانی اصلی مدیریت اکوسیستمی ذخایر این گونه‌ها محسوب می‌گردد. در طول دهه‌های گذشته ابداع و بکارگیری روش‌های مختلف آماری و طبقه‌بندی در مطالعات توزیع گونه‌ای بسیار مورد توجه بوده است. استفاده از رویکردهای مدلسازی در درک پیامدهای تغییرات شرایط محیطی بر نوسانات جمعیتی آبزیان و روابط احتمالی نهفته و همچنین

تشخیص عوامل مهم اثرگذار بر آنها اطلاعات ارزشمندی را برای مدیران شیلاتی فراهم می‌سازد. در این مطالعه، مبانی مدلسازی توزیع گونه‌ای و استفاده از آن در بررسی‌های شیلاتی جمعیت‌های آبزیان به شکلی نسبتاً جامع مرور شده است. مهمترین تکنیک‌های توسعه یافته در این زمینه در قالب چهارچوب‌های روش‌شناختی مربوطه و نیز مزایا و معایب مرتبط با آنها در ارزیابی‌های توزیع گونه‌های ارزشمند آبزی بر اساس مطالعات صورت گرفته تشریح شده است. مباحث مطرح شده می‌تواند در ارتقای دانش کلی در دسترس در ارتباط با مباحث کاربردی مختلف مدلسازی توزیع گونه‌ای با اهمیت شیلاتی مفید باشد.

لغات کلیدی: توزیع گونه‌ای، زیستگاه، عوامل محیطی، مدل‌های آماری کلاسیک، یادگیری ماشین

A review on the aquatics species distribution modelling and its most important methods

Abstract:

Proper recognition of species distribution patterns in aquatic ecosystems and their spatial and temporal changes during the past, and predictions for the future is one of the essential compartments of ecosystem-based fisheries management for aquatics stocks. Over the last decades, innovation and usage of different statistical and classification techniques in species distribution studies have been of high importance. Using modelling approaches has provided critical information for fisheries managers in understanding the impacts of changes in environmental conditions on fluctuations in aquatics populations and finding latent probable relationships between them and the most influencing environmental factors. The present study comprehensively reviewed the basics of species distribution modelling and its usage in fisheries assessment of aquatic species populations. Based on the conducted studies, the most important related developed modelling techniques under their methodological frameworks and their advantages and disadvantages for fisheries distribution modelling analysis are described. The presented topics could improve generally available knowledge about the practical incorporation of different species distribution modelling especially with fisheries importance.

Key words: Species distribution, habitat, environmental factors, classic statistical models, machine learning.

مقدمه

حفاظت و مدیریت یک گونه نیازمند حصول درک مناسب مکانی-زمانی از حضور و بقای آن در اکوسیستم می‌باشد (۱). رویکرد اکوسیستمی در مدیریت منابع شیلاتی مبتنی بر توجه به زیستگاه‌های مطلوب با درنظرگرفتن دینامیک جمعیت‌ها و تصمیم‌گیری‌های متعاقب آن است (۲). برنامه‌ریزی‌های مکانی به عنوان یک جزء ضروری در مدیریت

مؤثر تمامی فعالیت‌های مرتبط با بهره‌برداری‌های انسانی از مناطق آبی شناخته شده است که تحقق بهره‌برداری پایدار از منابع آبزی را تا حد زیادی تضمین می‌نماید (۳ – ۵). موفقیت این نوع برنامه‌ریزی‌ها و طراحی نواحی حفاظت‌شده وابسته به شناخت الگوهای توزیع مکانی گونه‌ها و زیستگاه‌ها و نوسانات آنها در طول زمان است (۶). تشخیص مناطق دارای اولویت با حساسیت بالا که در آن‌ها گونه‌های ساکن از احتمال ماندگاری بالایی برخوردار باشند، مبنای طرحریزی برنامه‌های بزرگ‌مقیاس حفاظتی است (۷). از این رو، اطلاعات صریح مکانی که نهایتاً در چهارچوب نقشه‌های زیستگاهی قابل ارائه باشند، ورودی‌هایی مفید برای سیستم‌های کارکردی موردنیاز جهت اعمال سیاست‌های مدیریتی با مبنای اکولوژیک محسوب می‌شوند (۱).

توزیع گونه‌های مختلف ماهیان در محیط‌های آبی توسط مجموعه پیچیده‌ای از عوامل محیطی و برهم‌کنش‌های آن‌ها کنترل می‌گردد (۸). تعیین عوامل محیطی اثرگذار بر پراکنش مکانی ماهیان، به یکی از مبانی تحقیقات اکولوژیکی در حال مطالعه از طریق رویکردهای ارزیابی زیستگاه تبدیل شده است (۱، ۹). این رویکردها اساساً به بیان روابط بین محیط و حضور گونه می‌پردازند به طوری که توصیف‌کننده میزان تناسب یک زیستگاه برای یک گونه مشخص هستند (۱۰، ۱۱). چنین فرآیندهایی عمدتاً در ارزیابی‌های مکانی و مدیریت گونه‌های تجاری ماهیان به کار گرفته می‌شوند (۱۲– ۱۴). هرچند یافتن نواحی صیادی بهینه محل تجمع ماهیان با توجه به تغییرات مکانی– زمانی پارامترهای محیطی و اثرات متعاقب آن‌ها بر ویژگی‌های زیستی ماهی دشوار می‌باشد، اما پیش‌بینی و تعیین این نقاط و تغییرات زمانی آن‌ها از اهمیت بالایی برای مدیران شیلاتی جهت جلوگیری از هدررفت منابع مالی برخوردار است (۱۵). از این رو دستیابی به اطلاعات قابل اعتماد می‌تواند یک امر چالش‌برانگیز در مدیریت شیلاتی اکوسیستمی جهت تشخیص زیستگاه‌های اپتیمم گونه‌های هدف تلقی شود (۱۶– ۱۹). بنابراین، مطالعه روابط ماهیان، بخصوص گونه‌های دارای اهمیت تجاری یا حفاظتی بالا، با شرایط محیطی حاکم بر محیط‌های زیست آن‌ها و همچنین الگوهای تغییرات این ارتباطات در طول زمان و در گستره‌های مکانی می‌تواند به عنوان یک مبنای علمی بسیار با اهمیت در جهت حفظ و بهره‌برداری مناسب‌تر از منابع آن‌ها مؤثر واقع شود.

مدلسازی توزیع گونه‌ای

مدل‌های توزیع گونه‌ای[1] (SDM) عنوانی کلی برای نام بردن از تکنیک‌ها و روش‌های مدلسازی است که برای بررسی ارتباط حضور گونه‌ها و فراوانی آن‌ها با شرایط محیطی مورد استفاده قرار می‌گیرند. از این مدل‌ها با عنوان مدل‌های مطلوبیت زیستگاه[2]، مدل‌های آشیان اکولوژیکی[3] یا مدل‌های پیش‌بینی زیستگاه[4] نیز نام برده می‌شود (۹). مدل‌های SDM ابزارهایی توانمند در بررسی پیامدهای تغییرات محیطی بر حضور و فراونی گونه‌ها به شمار می‌روند (۲۰) و در ارزیابی مطلوبیت شرایط زیستگاهی و پیش‌بینی توزیع گونه‌ای قابل استفاده می‌باشند (۲۱، ۲۲). این تکنیک‌ها نقش قابل توجهی در برنامه‌های حفاظتی و مدیریتی گونه‌ها و زیستگاه‌ها دارند (۲۳).

از مدل‌های SDM در ارزیابی مکانی و مدیریت گونه‌های تجاری و دارای اهمیت و پیش‌بینی تغییرات مکانی-زمانی قابل وقوع برای آن‌ها در آینده در نتیجه برنامه‌های بهره‌برداری، اثرات تغییرات اقلیمی و یا سایر عوامل استرس‌زای انسانی بر اکوسیستم‌های آبی استفاده می‌شود (۱۲، ۲۴-۲۷). این مدل‌ها با برقراری ارتباط بین حضور گونه، فراوانی آن و دیگر معیارهای مربوط به پارامترهای محیطی و سایر اثرات خودهمبستگی مکانی یا مکانی-زمانی (۲۸) قادرند تا پیش‌بینی‌های آماری قابل قبولی را برای نقاطی که پیش‌تر نمونه‌برداری نشده‌اند، ارائه دهند (۲۹-۳۱). چنین نتایجی در تسهیل شناخت نقاط داغ حضور گونه‌ها از طریق بکارگیری صدک‌های بالایی برآوردهای مکانی بسیار ارزشمند می‌باشند (۱۹، ۳۲). مقایسه اهمیت نسبی متغیرهای زیستگاهی در ساختاردهی جوامع ماهیان نیز با بکارگیری این مدل‌ها امکان‌پذیر است (۱۹، ۳۳-۳۵).

ارزیابی رویکردهای مدلسازی اطلاعات ارزشمندی را در توسعه مدل‌ها در مطالعات فراهم می‌سازد. انتخاب مدل و بکارگیری آن وابسته به نوع داده‌های در دسترس، کیفیت داده‌ها و سطح پیچیدگی مدل‌ها است (۳۶، ۳۷). تعداد قابل توجهی از الگوریتم‌های مختلف در حال حاضر در چهارچوب کلی SDM استفاده شده و روش‌های جدید نیز به صورت پیوسته در حال ابداع و توسعه هستند (۳۸ ، ۳۹). جنبه‌های روش‌شناختی انواع SDM از جهت‌های مختلف مطالعه شده و دلایل تفاوت‌های مشاهده شده در عملکرد آنها مورد بررسی قرار گرفته است (۴۰) از جمله الگوهای نمونه‌برداری (داده‌های حضور، ترکیب داده‌های حضور و عدم حضور، و داده‌های فراوانی یا زیتوده (۴۱، ۴۲) انتخاب

[1] Species distribution models: SDM
[2] Habitat suitability models
[3] Ecological niche models
[4] Habitat estimation models

روش‌های منفرد یا اشتراکی (۴۳)، طول دوره‌های زمانی مورد مطالعه و متغیرهای پیش‌بین مختلف (۴۴)، هم‌خطی[1] (۴۵) و انتخاب سطوح مختلف تفکیک مکانی و زمانی لایه‌های محیطی (۴۶). به طور کلی، دانش موجود در ارتباط با مدلسازی توزیع گونه‌ای بیانگر آن است که هیچ‌گاه نمی‌توان از برتری یک رویکرد مدلسازی منفرد یا کلی در مقایسه با انواع دیگر مدل‌ها سخن گفت (۱۹، ۴۷). بنابراین، این الزام وجود دارد که مطالعات مقایسه‌ای در سطوح مختلف اکوسیستمی جهت تعیین این که چه مدلی مناسب است، انجام گیرد.

رویکردهای مدلسازی توزیع گونه‌ای

رویکردهای مدلسازی توزیع گونه‌ای را می‌توان به طور عمده در سه گروه طبقه‌بندی کرد: (۱) مدل‌های شاخص مطلوبیت زیستگاه[2] (HSI)، (۲) مدل‌های آماری کلاسیک و (۳) روش‌های یادگیری ماشین[3] (ML). انتخاب تکنیک‌های مدلسازی مناسب تا حد زیادی نتیجه‌گیری و استدلال‌های حاصل از ارزیابی‌های توزیع گونه‌ای را تحت تأثیر قرار می‌دهد (۴۸، ۵۰). تعیین یک تکنیک به عنوان بهترین روش با بالاترین توان پیش‌بینی همواره چالش‌برانگیز بوده است؛ زیرا عملکرد پیش‌بینی مدل‌ها با توجه به عوامل متعدد از جمله تفاوت در رویکرد آماری یا الگوریتم‌های مورد استفاده، انتخاب متغیرهای ورودی متفاوت و ویژگی‌های آن‌ها، فرضیات و پیچیدگی مدل‌ها و همچنین ویژگی‌های گونه‌های مورد بررسی متغیر است (۵۱ –۵۳). به همین جهت پیشنهاد شده تا مجموعه‌ای از تکنیک‌های مدلسازی با الگوریتم‌های مختلف جهت دست‌یافتن به مدل بهینه مورد ارزیابی قرار گیرند (۵۴).

مدل شاخص مطلوبیت زیستگاه (HSI)

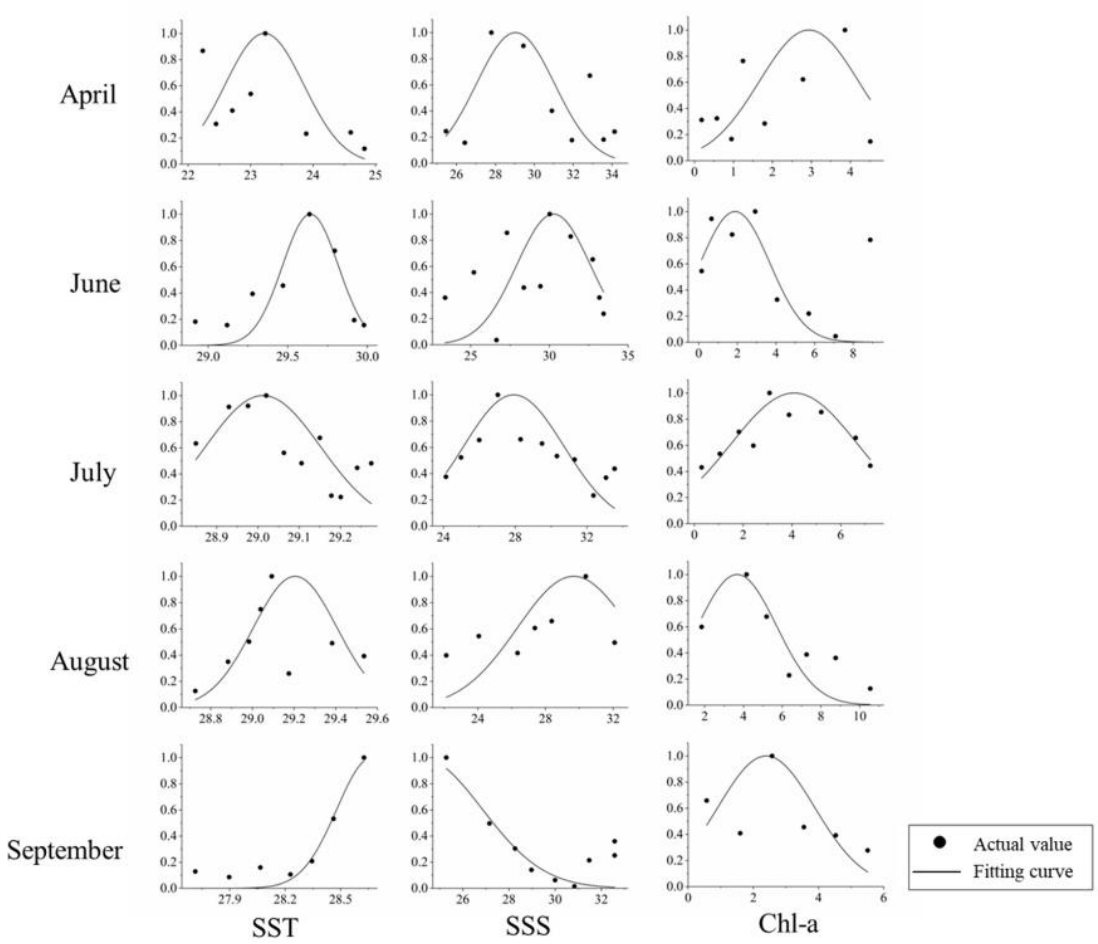
مدل‌های شاخص مطلوبیت زیستگاه (HSI) از اولین و مهمترین روش‌هایی هستند که به شکلی مناسب جهت ارزیابی مطلوبیت زیستگاه و در جهت ارائه راهکارهای حفاظتی در ارتباط با منابع شیلاتی مورد استفاده قرار گرفته‌اند (۴۹، ۵۰). مدل‌های HSI عملکردهای مناسبی را در تعیین توزیع زیستگاهی و حدود مقاومت و تحمل در برابر متغیرهای پیش‌بینی‌کننده محیطی نشان داده‌اند (۹، ۱۹، ۵۷، ۵۸). با این حال، این نوع از مدل‌ها از نظر اطلاعات پایه، مکانیسم‌ها و عملکرد متنوع بوده (۳۶، ۳۷) و از نظر برتری نسبی و اولویت موارد استفاده از آن‌ها، در طیف گسترده موارد کاربردشان قاعده خاصی مورد پیروی قرار نمی‌گیرد (۵۹).

[1] Collinearity
[2] Habitat Suitability Index (HIS) modelling
[3] Machine learning: ML

شاخص مطلوبیت زیستگاه (HSI)، به عنوان معمول‌ترین شاخص زیستگاهی مورد استفاده، یک ابزار محاسباتی جهت بیان ترجیح‌های گونه‌های مختلف در ارتباط با متغیرهای غیر زیستی است (۶۰، ۶۱). این مدل‌های غیرآماری مبتنی بر ترجیح‌های مشاهده شده تک-متغیره برای فاکتورهای زیستگاهی (یا به عبارتی شاخص‌های مطلوبیت[1] (SI)) بوده که نهایتاً به صورت ریاضیاتی با هم ترکیب شده و کیفیت زیستگاه را پیش‌بینی می‌کنند (۱۳، ۱۹، ۶۲). ترجیح یک بازه معین از فاکتورهای فیزیکی بر اساس نسبت درصد استفاده (یا به عبارت دیگر درصد افراد مشاهده شده‌ای که آن بازه مشخص را مورد استفاده قرار می‌دهند) به درصد کل بازه در دسترس از عامل موردنظر قابل تعیین است (۶۱). به طور کلی این شاخص‌ها در بازه مقادیر ۰ و ۱ برای هر یک از متغیرها قابل ارائه هستند به طوری که صفر نشان‌دهنده عدم ترجیح برای آن شرایط زیستگاهی و ۱ به معنای حداکثر ترجیح شرایط موردنظر است (شکل ۱؛ (۶۳)). عموماً زیستگاه فیزیکی وابسته به بیش از یک متغیر است و شاخص‌های مطلوبیت متعددی می‌بایست به منظور تعریف یک شاخص مطلوبیت مرکب، ادغام شوند (۶۱). بدین منظور، روش‌های متفاوتی جهت ترکیب کردن SIهای به دست آمده برای متغیرهای مختلف وجود دارد، از جمله استفاده از میانگین حسابی[2] (AMM)، میانگین هندسی[3] (GMM)، در نظر گرفتن مقادیر حداقل[4] (MINM) و حداکثر[5] (MAXM)، و یا روش CPM[6]. روش میانگین حسابی مبتنی بر این تصور است که شرایط زیستگاهی مناسب برای یک متغیر می‌تواند جبران‌کننده شرایط مناسب زیستگاهی برای متغیرهای دیگر باشد. در روش GMM نیز تا حدی جبران شرایط نامناسب یک متغیر توسط دیگر متغیرها وجود دارد، اما شاخص HSI در آن در صورت برابر صفر بودن یکی از SIها معادل صفر خواهد بود (۶۴). روش CPM بر این فرض استوار است که موجود هر یک از متغیرها را مستقل از دیگر متغیرها انتخاب می‌کند و ضرب کردن SIهای متغیرها متناظر با ضرب‌کردن احتمال‌های مستقل متغیرهای متفاوت است (۶۱، ۶۵). در روش MINM هیچ سطحی از جبران برای شرایط نامناسب یک متغیر توسط سایر متغیرها در نظر گرفته نمی‌شود و بر این فرض مبتنی است که عامل

[1] Suitability index: SI
[2] Arithmetic mean model
[3] Geometric mean model
[4] Minimum model
[5] Maximum model
[6] Continued product model: CPM

دارای بیشترین محدودیت، حد بالای مطلوبیت مجموعه شرایط را تعیین می‌کند (۶۶). در روش MAXM نیز شرایط مطلوب زیستگاهی متناسب با متغیر دارای کمترین محدودیت در محاسبه HSI در نظر گرفته می‌شود.



شکل ۱. نمودارهای شاخص‌های مطلوبیت (SI) برازش یافته مربوط برای پارامترهای محیطی مختلف (SST، SSS و Chl-a) در ارتباط با توزیع ماهی در ماه‌های مختلف. (برگرفته از: Wang et al. (2023) (۶۳)).

به طور کلی دو فرض ضمنی در محاسبه روش‌های فوق در نظر گرفته می‌شود: ۱) تمامی متغیرها از اهمیت یکسانی در رشد و بقای موجود برخوردارند؛ و ۲) تمامی متغیرهای محیطی مستقل بوده و هیچ گونه برهم‌کنشی بین آن‌ها وجود ندارد (۶۰). با استفاده از مدل‌های وزن‌دهی شده نسبی تا حدودی اثرات نامطلوب فرض اول برطرف می‌گردد، هر چند که وزن‌دهی متغیرها بر اساس اثرات واقعی آن‌ها نیست. اما به هر صورت اعمال برهم‌کنش‌های بین متغیرها

در محاسبه HSI امکان‌پذیر نمی‌باشد و در نظر گرفتن انتخاب زیستگاه بر پایه یک فرآیند چند متغیره با در نظر گرفتن اثرات متقابل و همبستگی بین متغیرها میسر نیست (۶۷-۶۹).

روش‌های آماری کلاسیک

روش‌های آماری کلاسیک از جمله مدل‌های خطی (LM)[1]، مدل‌های خطی تعمیم‌یافته[2] (GLM) و مدل‌های جمعی تعمیم‌یافته[3] (GAM)، اساساً مبتنی بر رگرسیون‌های خطی ساده یا چندگانه هستند (۳۳، ۳۵) که به شکل بسیار گسترده در مدلسازی اکولوژیکی مورد استفاده قرار گرفته‌اند.

مدل خطی (LM)

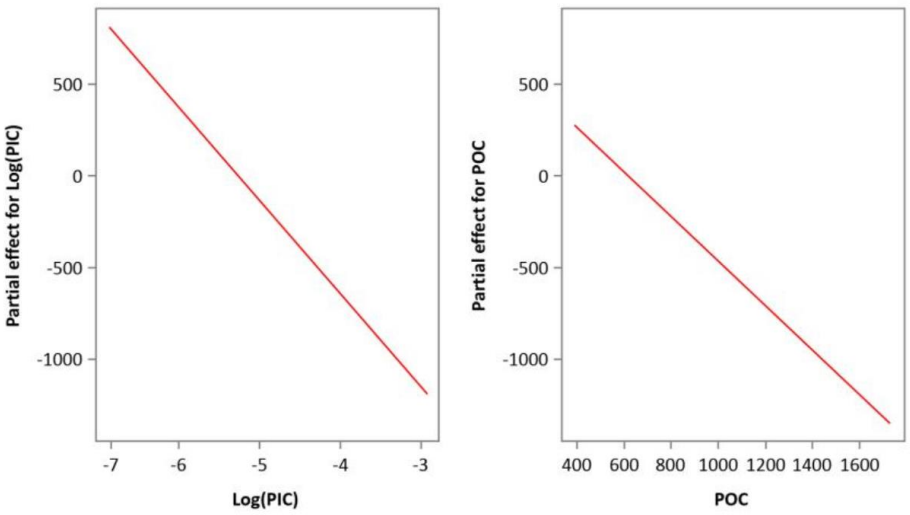
استفاده از این مدل‌ها ساده بوده و اعمال و بکارگیری آن‌ها مستقیماً قابل انجام است (۴۷). مدل‌های رگرسیونی خطی با داشتن سه فرض اصلی جهت استفاده دارای محدودیت‌هایی هستند: ۱) خطاها می‌بایست به صورت یکسان و مستقل توزیع شده باشند یا به عبارت دیگر واریانس متغیر پاسخ برای تمامی مشاهدات ثابت باشد؛ ۲) خطاهای مدل می‌بایست از یک مدل توزیع گاوسی[4] نرمال تبعیت کند؛ و ۳) تابع رگرسیون برای متغیرها خطی باشد. این فرضیات غالباً برای داده‌های واقعی دور از انتظار می‌باشند. داده‌های اکولوژیکی غالباً پیچیده بوده و به ندرت پیش‌فرض‌های آماری این نوع از روش‌ها (از جمله استقلال و نرمال بودن مشاهدات) را تأمین می‌کنند (۷۰- ۷۲). همچنین مدل‌های خطی قادر به پرداختن به اثرات متقابل پیچیده در حالتی که ارتباط بین متغیر وابسته و یک متغیر مستقل تحت تأثیر ترکیبی از دیگر متغیرهای مستقل قرار می‌گیرد، نیستند. مدل‌های خطی به توزیع‌های به شدت اُریب متغیرهای مستقل که عمدتاً در پارامترهای بوم‌شناختی مشاهده می‌شوند، حساس می‌باشند (۷۳). با این حال، مدل‌های رگرسیونی مدرن (از جمله GLM و GAM) تا حدودی قابلیت بیشتری در پرداختن به داده‌های فاقد پیش‌فرض‌های مورد اشاره برای مدل‌های خطی دارند.

مدل خطی تعمیم‌یافته (GLM)

مدل‌های خطی تعمیم‌یافته (GLM) مدل‌های آماری هستند که امروزه نیز در مطالعات اکولوژیکی مورد استفاده قرار می‌گیرند و مبنایی برای دیگر انواع مدل‌ها به حساب می‌آیند (۷۴). مدل‌های GLM، مدل‌هایی منعطف و با کاربرد

[1] Linear Models: LM
[2] Generalized Linear Models: GLM
[3] Generalized Additive Models: GAM
[4] Gaussian distribution

گسترده در مدلسازی توزیع گونه‌ای هستند که از طریق محاسبه حضور گونه‌ها به صورت توابعی پارامتریک از متغیرهای محیطی عمل می‌کنند (۴۹، ۵۰، ۷۵). این نوع مدل‌ها برای داده‌های با تغییرات خطی کاربرد دارند. مدل GLM در واقع یک نوع بسط داده شده از مدل‌های خطی است که امکان توزیع‌های خطای غیرنرمال برای متغیرهای پاسخ برای آن وجود دارد. در واقع فرض کلیدی این مدل‌ها آن است که رابطه مقادیر مورد انتظار از متغیر پاسخ با عبارت پیش‌بین خطی باشد (۷۵) (شکل ۲؛ (۴۲)). در GLM، داده‌ها می‌توانند از توزیع‌های متعدد خانواده نمایی[1] از توزیع‌های احتمال پیروی کنند (از جمله توزیع‌های نرمال، دوجمله‌ای[2]، پواسون[3]، هندسی[4]، دوجمله‌ای منفی[5] و نمایی) (۷۶) که اغلب آن‌ها از تناسب بیشتری با ساختارهای خطای غیرنرمال در داده‌های اکولوژیکی برخوردارند. تمامی مدل‌های خطی تعمیم‌یافته از سه جزء تشکیل شده‌اند: یک توزیع تشخیص‌دهنده متغیر پاسخ، یک عامل پیش‌بینی‌کننده خطی که تعیین‌کننده متغیرهای محیطی مورد استفاده به عنوان متغیرهای پیش‌بین مدل است، و یک تابع ربط[6] که توصیف‌کننده ارتباط بین متغیرهای خطی و مقادیر مورد انتظار از متغیر پاسخ است (۷۷). این تکنیک در مطالعات مختلفی در زمینه ارزیابی ذخایر شیلاتی دریایی مورد استفاده قرار گرفته است (۷۸–۸۰).



شکل ۲. نمودارهای مربوط به اثرات نسبی متغیرهای محیطی (POC و log(PIC)) اثرگذار بر پراکنش ماهی در یک مدل GLM.
(برگرفته از: Moëzzi et al. (2023) (۴۲)).

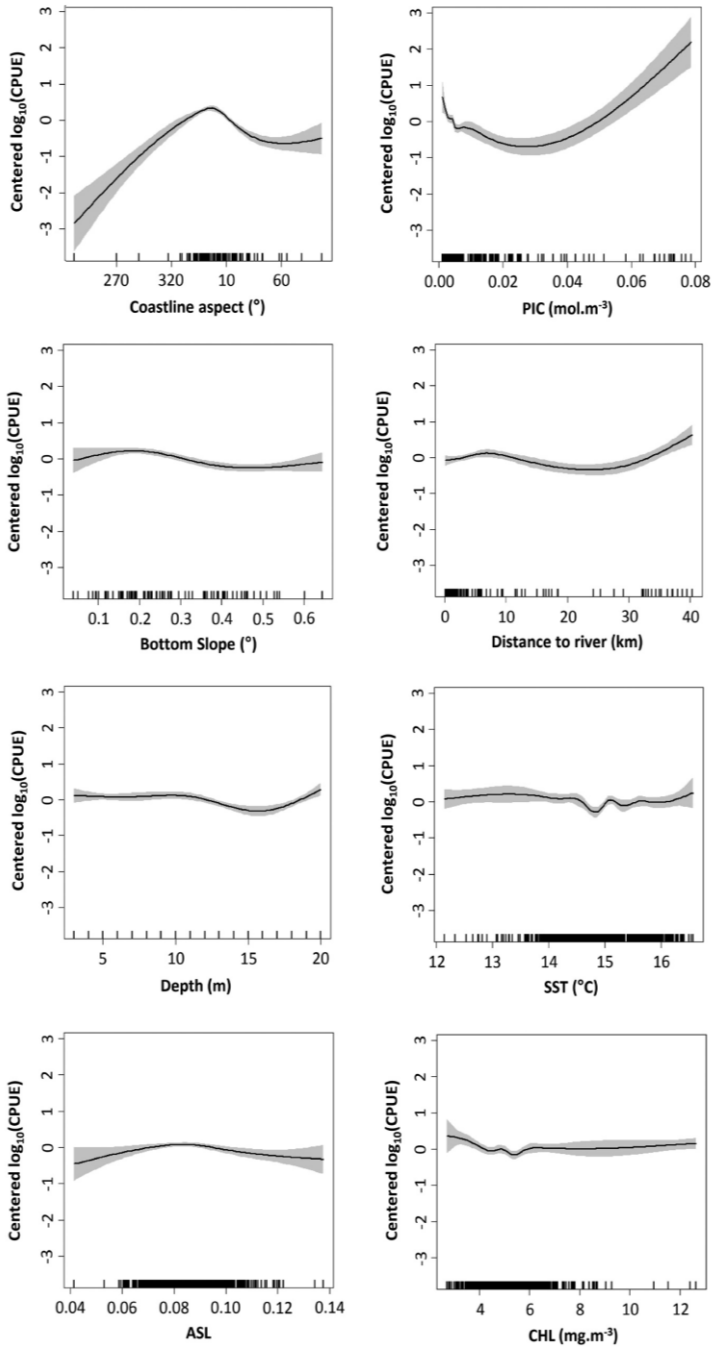
[1] Exponential family
[2] Binomial
[3] Poisson
[4] Geometric
[5] Negative binomial
[6] Link function

مدل جمعی تعمیم‌یافته (GAM)

مدل‌های جمعی تعمیم‌یافته (GAM)، نوع توسعه‌یافته از مدل GLM است و شکل بسیار معمول مدلسازی از نوع رگرسیونی به شمار می‌رود. مدل‌های GAM توابع هموارسازی ناپارامتریک را به منظور توصیف منعطف پاسخ بر مبنای روابط خطی و غیرخطی به‌کار می‌برند (۸۱، ۸۲). در واقع مدل GAM، یک تابع ارتباط را برای ایجاد رابطه بین میانگین‌های متغیر پاسخ و یک تابع هموارکننده از متغیرهای پیش‌بین مورد استفاده قرار می‌دهد. نقطه قوت این مدل‌ها عبارت است از توانایی آنها در پرداختن به روابط غیرخطی و غیریکنواخت بین پاسخ و مجموعه متغیرهای پیش‌بین (۱۹، ۷۷) (شکل ۳؛ (۱۹)). در این مدل‌ها امکان تبدیل‌های مختلف روی متغیرهای مستقل قبل از وارد کردن آن‌ها به مدل وجود دارد (۸۳) و ارتباط خطی معمول بین متغیرهای مستقل و متغیر وابسته با توابع غیرخطی جایگزین می‌شود تا روندهای تغییرات غیرخطی در داده‌ها شناخته و مدلسازی شوند (۸۴). این مورد توانایی کار با داده‌های غیرخطی را بهبود می‌بخشد، با این وجود مدل مذکور قادر به پرداختن به اثرات متقابل بین متغیرها نیست. مدل‌های آماری با کاربرد بسیار گسترده متعلق به این گروه از مدل‌ها، شامل مدل‌های جمعی برای داده‌های گاوسی، مدل‌های لجستیک ناپارامتری برای داده‌های دارای توزیع دوجمله‌ای و مدل‌های خطی لگاریتمی ناپارامتری برای داده‌های دارای توزیع پواسون می‌باشند. با وجود این که در مدل‌های GAM امکان پرداختن به داده‌های غیرخطی فراهم شده است، اما از آنجایی که این مدل‌ها وضعیتی جمعی (افزاینده) دارند، قادر به بررسی تأثیر اثرات متقابل متغیرها نیستند به طوری که این متغیرها می‌بایست به صورت منفرد تعیین شده و به مدل اضافه شوند (۸۳).

مدل‌های جمعی تعمیم‌یافته (GAM) به وفور در ارزیابی روابط غیرخطی بین حضور گونه‌ها و متغیرهای محیطی مورد استفاده قرار می‌گیرند (۶، ۷، ۱۰، ۱۹، ۸۳، ۸۴). مدل‌های GAM جهت توضیح ارتباطات بین فراوانی ماهیان و عوامل محیطی (۸۶، ۸۷) و پیش‌بینی فراوانی‌های نسبی ماهیان (۶، ۷، ۸۸) به کار گرفته شده‌اند. این مدل‌ها همچنین جهت تعریف توابع ترجیح زیستگاهی ماهیان در دیگر چهارچوب‌های مدلسازی زیستگاهی نیز استفاده شده‌اند (۸۹). عملکرد مناسب و توانایی بالای پیش‌بینی این تکنیک در مدلسازی روابط بین زیتوده ماهیان و پارامترهای زیستگاهی در مقایسه با مدل‌های HSI تجربی و روش‌های یادگیری ماشین در مطالعات متعددی گزارش شده است (۶، ۷، ۱۰،

۸۳). عملکرد مناسب مدل‌های جمعی تعمیم‌یافته در فراهم‌سازی برداشت‌های بسیار دقیق و قابل اعتماد در مطالعات متعددی نشان داده شده است (۴۳، ۹۰، ۹۱). هر چند مدل‌های GAM از مدل‌های دارای کاربرد گسترده به حساب می‌آیند و استفاده آنها از توابع هموارکننده غیرپارامتریک امکان توصیف منعطف پاسخ‌های پیچیده گونه‌ای به محیط را فراهم ساخته است، پیچیدگی محاسباتی آنها تولید پیش‌بینی‌هایی برای مجموعه‌های داده مستقل از جمله در سیستم-های اطلاعات جغرافیایی (GIS) را دشوار ساخته است (۹۲).



شکل ۳. نمودارهای برازش یافته هموارکننده‌های مدل GAM مربوط به بیومس ماهی در برابر پارامترهای محیطی مختلف. (برگرفته از Moëzzi et al., (2022) (۱۹)).

روش‌های یادگیری ماشین

یادگیری ماشین به صورت یک زمینه تحقیقاتی تعریف می‌گردد که در آن از برنامه‌ها و الگوریتم‌های کامپیوتری در جهت شناخت و تفسیر روابط موجود در یک مجموعه داده استفاده می‌گردد (۹۳). فرضیه اصلی یادگیری ماشین آن است که یک ماشین (یا به عبارتی یک الگوریتم یا مدل) قادر است تا پیش‌بینی‌هایی را بر اساس داده‌های در دسترس ارائه نماید (۹۴). تکنیک پایه‌ای نهفته در تمامی روش‌های یادگیری ماشین آن است که یک ترکیب تکرارشونده از آماره‌ها و فرآیندها در جهت به حداقل رساندن خطا و به حداکثر رساندن تطابق پیش‌بینی‌ها و مشاهدات، اعمال شده و در درجات و سطوح مختلف با هم ترکیب می‌شوند (۹۴). بسیاری از الگوریتم‌های یادگیری ماشین به صورت تکرارشونده تمام یا بخش زیادی از نتایج و پیامدهای احتمالی را در جهت یافتن بهترین نتیجه آزمون می‌کنند به طوری که بهترین خروجی بر اساس انتظارات کاربر برای مسئله مورد بررسی تعیین می‌گردد.

در طول دهه‌های گذشته استفاده از تکنیک‌های یادگیری ماشین در مطالعات اکولوژیکی بسیار پرطرفدار بوده است؛ زیرا الگوریتم‌های به کار گرفته شده در این تکنیک‌ها امکان مطالعه انواع مختلف داده‌ها، تعداد زیاد متغیرهای پیش‌بین دارای همبستگی، و همچنین مدلسازی روابط سلسله مراتبی پیچیده همراه با سطوح بالای دقت پیش‌بینی را در مقایسه با دیگر روش‌های آماری کلاسیک فراهم نموده‌اند (۴۲، ۵۰، ۹۵–۹۷). برخی از این الگوریتم‌ها به طور مشخص برای پرداختن به چالش‌های موجود در مطالعات اکولوژیک طراحی شده‌اند، از جمله برای شرایطی که تعداد زیادی از متغیرهای پیش‌بین بالقوه با مقادیر کم نمونه در دسترس هستند (۹۸). تکنیک‌های یادگیری ماشین بسیار متعدد و متنوع بوده (۹۹) و در مجموع به دو صورت مورد استفاده قرار می‌گیرند: یادگیری تحت نظارت[1] و یادگیری بدون نظارت[2].

در یادگیری تحت نظارت، یک مدل برای پیش‌بینی یک متغیر پاسخ بر اساس یک مجموعه داده برازش می‌یابد در حالی که در یادگیری بدون نظارت، هیچ متغیر پاسخی تعیین نمی‌شود و الگوریتم‌های مورد استفاده تنها الگوها و گروه‌بندی‌های موجود در داده‌ها را بر اساس ساختار آنها تشخیص می‌دهند. این الگوریتم‌های یادگیری آماری غالباً با

[1] Supervised learning
[2] Unsupervised learning

تمرکز بر دقت پیش‌بینی مورد استفاده قرار می‌گیرند و هرچند معمولاً میزان اهمیت متغیرها در آنها برآورد می‌گردد، اما هیچ گونه مقادیر آماری P جهت تعیین معنی‌داری متغیرها ارائه نمی‌شود (۱۰۰).

در سال‌های اخیر، الگوریتم‌های یادگیری ماشین در مطالعات متعددی در مدلسازی روابط بین جوامع موجودات و محیط آنها مورد استفاده قرار گرفته‌اند. الگوریتم‌های یادگیری ماشین مبتنی بر درخت‌های تصمیم‌گیری[1] (DT) از جمله روش‌های درخت رگرسیونی تقویت شده[2] (BRT) و جنگل تصادفی[3] (RF)، و تکنیک ماشین بردار پشتیبان[4] (SVM) دو گروه از روش‌های بسیار پرکاربرد یادگیری ماشین هستند که در این نوع از مطالعات مورد استفاده قرار گرفته‌اند (۸۱، ۸۵، ۹۶، ۱۰۱). این مدل‌ها توان پرداختن به موضوعات مرتبط با داده‌های گمشده و ارتباطات غیرخطی و همچنین ظرفیت بالای ایجاد مدل‌های با سطح کمتر خطای پیش‌بینی را در بررسی‌های زیستگاهی ماهیان نشان داده‌اند (۸۱، ۹۵، ۹۶، ۱۰۲ –۱۰۴).

مدل‌های درخت–مبنا[5]

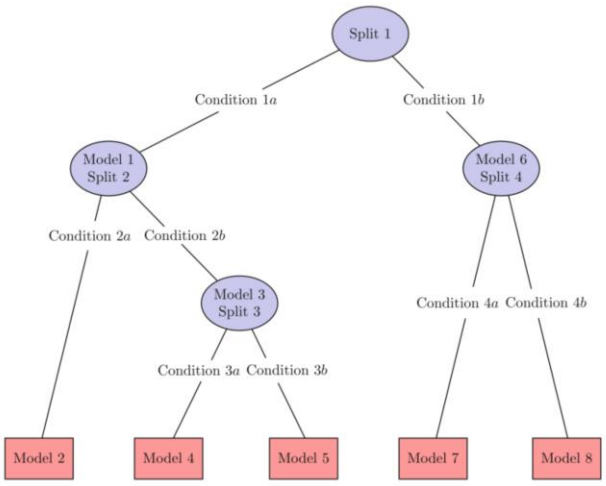
مدل‌های درخت–مبنا یا به عبارتی درخت‌های تصمیم‌گیری از مدل‌های ساده و قابل تفسیر برای انجام رگرسیون و طبقه‌بندی در میان روش‌های یادگیری ماشین به حساب می‌آیند (۱۰۵). این مدل‌ها مقید به فرض خطی بودن داده‌ها نیستند و توان پرداختن به اثرات متقابل بین متغیرها را نیز دارند. این مدل‌ها، تکنیک‌های رگرسیونی ناپارامتریک هستند که درخت‌های تصمیم را بر مبنای الگوریتمی که به طور مکرر مجموعه داده را تقسیم می‌کند، رشد می‌دهند. درخت‌های تصمیم متشکل از یک ساختار منشعب شونده هستند که مجموعه داده را در یک چهارچوب سلسله مراتبی تقسیم می‌کنند (شکل ۴؛ (۹۹)). متغیر پاسخ در درخت تصمیم معمولاً از نوع عددی (درخت‌های رگرسیونی) یا طبقه‌ای (درخت‌های طبقه‌بندی) است و متغیرهای پیش‌بین می‌توانند از هر نوعی باشند (۱۰۶). مراحل ساخت هر درخت عبارت است از: ۱) یک متغیر پیش‌بین انتخاب می‌شود (گره ریشه‌ای) که بر اساس آن مجموعه داده به دو گروه هموژن تقسیم می‌شود؛ ۲) این روند به طور جداگانه برای هر زیرگروه (گره‌های برگ‌ها) اعمال می‌شود تا ساخت درخت متوقف گردد؛ و ۳) درخت کامل بر مبنای قاعده خطای استاندارد جهت اجتناب از بیش‌برازش به

[1] Decision trees
[2] Boosted Regression Tree: BRT
[3] Random Forest: RF
[4] Support vector machine
[5] Tree-based models

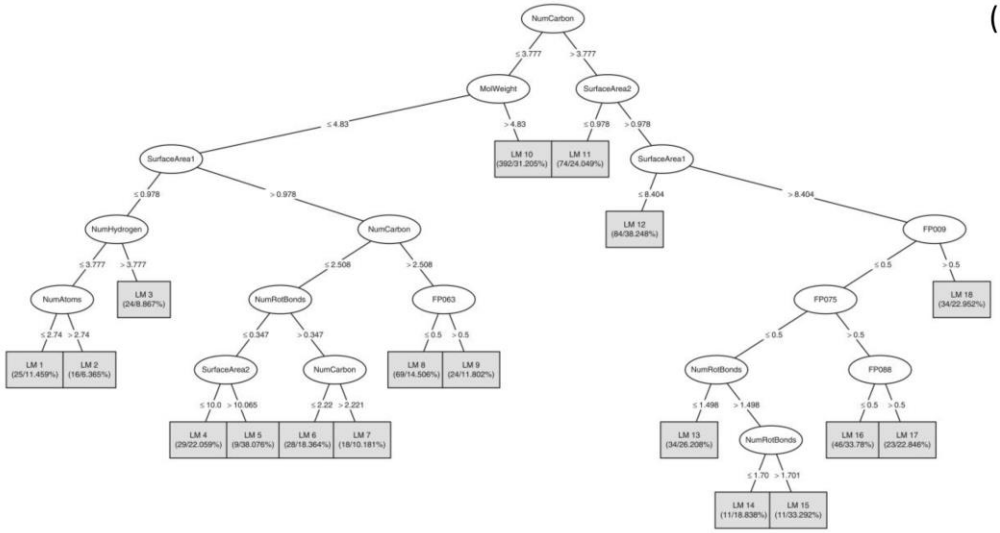
صورت برگشتی تا اندازه مناسب آن پیراسته می‌گردد (۸۱، ۱۰۷، ۱۰۸). تقسیم تا زمانی که یک معیار توقف حاصل نشود یا تا زمانی که یک تقسیم (به عبارت دیگر گره در درخت) تنها شامل یک مشاهده باشد، ادامه می‌یابد. اهمیت متغیرها در مدل‌های درخت–مبنا بر حسب کاهش در خطای پیش‌بینی حاصل از تقسیم در هر گره قابل دستیابی است (۱۰۹).

مدل‌های درخت–مبنا بر اساس ساختار آنها به شکل مؤثری به غیرخطی بودن و عدم پیوستگی (یا اثرات آستانه‌ای) در ساختار داده‌ها پرداخته و به شکل مناسبی مباحث مربوط به اثرات متقابل بین متغیرها را پوشش می‌دهند. با این وجود، این مدل‌ها در مدلسازی روابط خطی ضعیف بوده و نسبت به تغییرات کوچک در داده‌های مورد استفاده در برازش مدل حساس می‌باشند (۷۲، ۱۱۰).

(الف)



(ب)



شکل ۴. نمای شماتیک یک مدل درختی رگرسیونی (الف)؛ یک مدل درخت رگرسیونی نهایی برای یک مجموعه داده (ب). در تصویر اعداد انتهایی در هر گره پایانی نشان‌دهنده نمونه‌ها و درصد پوشش مربوط به آن گره می‌باشد. (برگرفته از: Kuhn and Johnson (2013) (۹۹)).

مدل درخت رگرسیونی تقویت شده (BRT)

مدل‌های درختی رگرسیونی تقویت شده (BRT) روش‌های مجتمعی از میانگین‌گیری مدل هستند که نقاط قوت دو الگوریتم درخت‌های رگرسیونی و Boosting را ترکیب می‌کنند. این تکنیک‌ها در جهت بهبود عملکرد پیش‌بینی نسبتاً ضعیف یک مدل منفرد درختی از طریق برازش تعداد بسیار زیادی درخت ساده و ترکیب آنها جهت پیش‌بینی در یک فرآیند مرحله به مرحله عمل می‌کنند (۹۵). جنبه کلیدی boosting در مدل‌های BRT در مقایسه با دیگر تکنیک‌های مدل‌های تجمیعی (از جمله RF) آن است که هر درخت متوالی بر اساس خطای باقیمانده مجموعه درخت به دست آمده ساخته می‌شود (۸۳). به عبارت دیگر، درخت‌های جدید بر داده‌های با برازش ضعیف از زیرنمونه‌های قبلی متمرکز هستند و از این رو به صورت پیوسته ارتقا می‌یابند. در برازش مدل‌های BRT لازم است تا دو پارامتر برای آن‌ها تعیین گردد: ۱) نرخ یادگیری[1] (lr) که سهم هر درخت در مدل در حال رشد را تعیین می‌کند؛ و ۲) پیچیدگی درخت[2] (tc)، که اندازه (یا تعداد تقسیم‌ها) در هر درخت را کنترل می‌کند (۱۰۶).

مدل‌های BRT مزایای مدل‌های درخت–مبنا را در خود دارند: پرداختن به مقادیر گم‌شده و پرداختن به انواع مختلفی از متغیرهای پیش‌بین، دستکاری خودکار اثرات متقابل بین متغیرهای پیش‌بین و ایمن‌سازی در برابر تغییرشکل داده و داده‌های پرت (۸۵، ۹۵، ۱۱۱). اهمیت نسبی متغیرها از طریق میانگین‌گیری تعداد دفعات انتخاب متغیر برای ایجاد انشعاب در درخت و میزان مربع بهبود در نتایج حاصل از این تقسیم‌ها، قابل محاسبه است (۱۰۹). این مقادیر در مقیاس ۱۰۰ بوده و اعداد بیشتر نشان‌دهنده تأثیر بیشتر متغیر پیش‌بین بر متغیر پاسخ هستند. توانایی مدلسازی اثرات متقابل نیز توسط پارامتر tc کنترل می‌شود به طوری که مقدار این پارامتر تعیین‌کننده تعداد گره‌های هر درخت و متعاقباً توانایی مدلسازی اثرات متقابل است (۸۵). مدل‌های BRT غالباً به دلیل قوت درخت‌های رگرسیونی و تقویت

[1] Learning rate: lr
[2] Tree complexity: tc

میانگین‌گیری مدل‌ها که بهبود عملکرد پیش‌بینی را سبب می‌شود، نسبت به مدل‌های درختی منفرد از عملکرد پیش‌بینی بهتری برخوردارند. هر چند ممکن است حدودی از بیش‌برازش در خروجی‌های آن‌ها رخ دهد، اما از طریق فرآیند اعتبارسنجی متقابل[1] (CV) این میزان به حداقل می‌رسد (۸۵، ۱۱۲). این روش به شکل موفق جهت مدلسازی مطلوبیت زیستگاه تعدادی از گونه‌ها مورد استفاده قرار گرفته است (۸۳، ۸۵، ۱۰۴).

مدل جنگل تصادفی (RF)

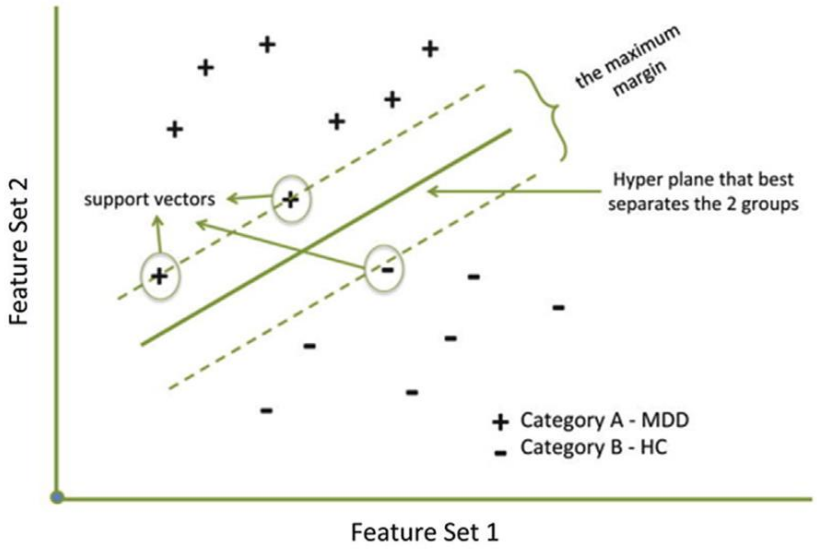
جنگل تصادفی یک رویکرد میانگین‌گیری مدل است که تعداد صدها و هزاران درخت تصادفی را بر اساس مجموعه‌ای از متغیرها و مشاهدات با انتخاب تصادفی ایجاد می‌کند (۷۲، ۹۱). RF متشکل از مجموعه‌ای از درخت‌های طبقه‌بندی یا رگرسیونی به منظور ارائه پیش‌بینی‌هایی فاقد بیش‌برازش می‌باشد (۷۲، ۱۱۳). این رویکرد مجتمع درختی قادر است تا عدم ثبات درخت‌های آغازین را تنظیم نموده و پیش‌بینی‌های قابل اعتمادی را منتج شود (۷۲، ۱۱۴). این مدل غالباً با رویکردی متشکل از چند مرحله مورد استفاده قرار می‌گیرد (۱۱۵): ۱) نمونه‌هایی از داده‌های اولیه و اصلی به صورت بوت‌استرپ[2] تعیین می‌گردند (ntrees)؛ ۲) درخت‌های رگرسیونی بر اساس این نمونه‌ها ایجاد می‌شوند و در هر گره از درختان، زیرمجموعه‌ای از متغیرها به صورت تصادفی برای تعریف انشعاب (mtry) انتخاب می‌شوند، و بهترین انشعاب تعیین می‌گردد؛ و ۳) پیش‌بینی نهایی بر اساس میانگین‌گیری از پیش‌بینی‌های حاصل از تعداد ntrees درخت ایجاد شده، به دست می‌آید. توان طبقه‌بندی و پیش‌بینی بالای این روش در مطالعات متعدد (۴۲، ۹۷) گزارش شده است که تا حد زیادی ناشی از کارآیی آن در شناخت روندهای داده، نداشتن پیش‌فرض‌های اختصاصی در ارتباط با توزیع داده‌ها و قدرت آن در پرداختن به اثرات متقابل بین متغیرها است (۱۱۵). این روش در مطالعات مختلفی در ارتباط با اکوسیستم‌های آبی و مباحث شیلاتی از جمله پیش‌بینی تنوع گونه‌ای و برآورد مطلوبیت زیستگاهی نیز مورد استفاده قرار گرفته است (۴۲، ۵۰، ۸۱).

مدل ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM) از جدیدترین روش‌های یادگیری ماشین است که می‌تواند اثرات غیرخطی و اثرات متقابل بین متغیرها را در فرآیند مدلسازی مورد بررسی قرار دهد. این تکنیک، متغیرهای پیش‌بین را به یک فضای با

[1] Cross-validation
[2] Bootstrapping

ابعاد بیشتر از متغیرها برداشت می‌کند به طوری که در آن مسئله موردنظر برای پیش‌بینی از یک راهکار خطی برخوردار باشد (۱۱۷). در این روش، تعدادی صفحات تفکیک‌کننده جهت طبقه‌بندی و تمایز داده‌های ورودی به کار گرفته می‌شوند (۱۱۸) (شکل ۵؛ (۱۱۹)). در صورتی که توابع مربوط به صفحات تفکیک‌کننده خطی نباشند، داده‌ها در یک فضای با ابعاد بیشتر برداشت می‌شوند به طوری که در این فضا داده‌ها به صورت خطی قابل تفکیک باشند. این برداشت داده‌ها از طریق استفاده از گروهی از توابع با نام کرنل[1] انجام می‌گیرد که تبدیل‌های غیرخطی را برای داده‌های اصلی اعمال می‌کنند (۱۲۰). مبانی و ویژگی‌های محاسباتی این روش پیچیده بوده که در اینجا بدان اشاره نمی‌شود و شرح الگوریتم‌های مورد استفاده در آن به طور کامل در منابع مختلف از جمله Witten *et al.* (2011) (۱۲۱) ارائه شده است.



شکل ۵. نمای شماتیک صفحه جداکننده بردارهای پشتیبان مرتبط با دو گروه از نقاط داده (Category A-MDD و Category B-HC) پیش‌بینی شده. (برگرفته از Pinser and Schnyer, (2020) (۱۱۹)).

جمع‌بندی

در این مطالعه، مبانی مدلسازی توزیع گونه‌ای و جنبه‌های مختلف اهمیت آن در مطالعات تغییرات زمانی و مکانی آبزیان مورد بررسی گردید. همچنین اهمیت بررسی‌های زیستگاهی گونه‌های آبزیان و نقش تکنیک‌های مختلف

[1] Kernels

مدلسازی از جمله روش‌های آماری کلاسیک و روش‌های نوین یادگیری ماشین به عنوان مهمترین روش‌های مورد استفاده به همراه نقاط قوت و ضعف آنها در چهارچوب‌های مختلف توزیع گونه‌ای، تشریح شد. مطالب ارائه شده در این تحقیق می‌تواند به شکل قابل توجهی در ارتقای دانش موجود در ارتباط با مباحث مختلف مطالعات توزیع گونه‌ای آبزیان مفید واقع گردد.

منابع

1. Giannoulaki, M., Iglesias, M., Tugores, M.P., Bonanno, A., Patti, B., De Felice, A., Leonori, I., Bigot, J.L., Tičina, V., Pyrounaki, M.M., & Tsagarakis, K. (2013). Characterizing the potential habitat of European anchovy *Engraulis encrasicolus* in the Mediterranean Sea, at different life stages. *Fisheries Oceanography*. 22(2).
2- Olsen, Z. (2019). Quantifying nursery habitat function: variation in habitat suitability linked to mortality and growth for juvenile Black Drum in a hypersaline estuary. *Marine and Coastal Fisheries*. 11(1), 86-96.
3- Mengerink, K.J., Van Dover, C.L., Ardron, J., Baker, M., Escobar-Briones, E., Gjerde, K., Koslow, J.A., Ramirez-Llodra, E., Lara-Lopez, A., Squires, D. & Sutton, T. (2014). A call for deep-ocean stewardship. *Science*. 344(6185), 696-698.
4- Clark, M.R., Schlacher, T.A., Rowden, A.A., Stocks, K.I., & Consalvey, M. (2012). Science priorities for seamounts: research links to conservation and management. *PloS one*. 7(1), p.e29232.
5- O'Leary, B.C., Brown, R.L., Johnson, D.E., Von Nordheim, H., Ardron, J., Packeiser, T. & Roberts, C.M. (2012). The first network of marine protected areas (MPAs) in the high seas: the process, the challenges and where next. *Marine Policy*. *36(3*), 598-605.
6- Parra, H.E., Pham, C.K., Menezes, G.M., Rosa, A., Tempera, F. & Morato, T. (2017). Predictive modeling of deep-sea fish distribution in the Azores. *Deep Sea Research Part II: Topical Studies in Oceanography*. *145*, 49-60.
7- Schismenou, E., Tsoukali, S., Giannoulaki, M. & Somarakis, S. (2017). Modelling small pelagic fish potential spawning habitats: eggs vs spawners and in situ vs satellite data. *Hydrobiologia*. 788(1), 17-32.
8- Maravelias, C.D., Tsitsika, E.V. & Papaconstantinou, C. (2007). Seasonal dynamics, environmental preferences and habitat selection of John Dory (*Zeus faber*). *Estuarine, Coastal and Shelf Science*. 72(4), 703-710.
9- Guisan, A., & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*. 135(2-3), 147-186.
10- Su, N.J., Chang, C.H., Hu, Y.T., Chiang, W.C. & Tseng, C.T. (2020). Modeling the spatial distribution of swordfish (*Xiphias gladius*) using fishery and remote sensing data: approach and resolution. *Remote Sensing*. 12(6), p.947.
11- Zwolinski, J.P., Emmett, R.L. & Demer, D.A. (2011). Predicting habitat to optimize sampling of Pacific sardine (*Sardinops sagax*). *ICES Journal of marine Science*. 68(5), 867-879.
12- Ramirez-Llodra, E., Tyler, P.A., Baker, M.C., Bergstad, O.A., Clark, M.R., Escobar, E., Levin, L.A., Menot, L., Rowden, A.A., Smith, C.R. & Van Dover, C.L., (2011). Man and the last great wilderness: human impact on the deep sea. *PLoS one*. 6(8), p.e22588.
13- Tian, S., Chen, X., Chen, Y., Xu, L. & Dai, X. (2009). Evaluating habitat suitability indices derived from CPUE and fishing effort data for *Ommatrephes bratramii* in the northwestern Pacific Ocean. *Fisheries Research*. 95(2-3), 181-188.
14- Moëzzi, F., Poorbagher, H., Eagderi, S., Feghhi, J. (2024). Time series modelling for the Caspian Kutum (*Rutilus frisii*) catch using SARIMA model. *Environmental Sciences*, 22(2), In press. (In Persian)

15- Geronimo, R.C., Franklin, E.C., Brainard, R.E., Elvidge, C.D., Santos, M.D., Venegas, R., & Mora, C. (2018). Mapping fishing activities and suitable fishing grounds using nighttime satellite images and maximum entropy modelling. *Remote Sensing*. 10(10), p.1604.
16- Liu, X., Wang, J., Zhang, Y., Yu, H., Xu, B., Zhang, C., Ren, Y. & Xue, Y. (2019). Comparison between two GAMs in quantifying the spatial distribution of *Hexagrammos otakii* in Haizhou Bay, China. *Fisheries Research*. 218, 209-217.
17- Chang, Y.J., Sun, C.L., Chen, Y., Yeh, S.Z. & Dinardo, G. (2012). Habitat suitability analysis and identification of potential fishing grounds for swordfish, *Xiphias gladius*, in the South Atlantic Ocean. *International Journal of Remote Sensing*. 33(23), 7523-7541.
18- Vinagre, C., Fonseca, V., Cabral, H. & Costa, M.J. (2006). Habitat suitability index models for the juvenile soles, *Solea solea* and *Solea senegalensis*, in the Tagus estuary: defining variables for species management. *Fisheries research*. 82(1-3), 140-149.
19- Moëzzi, F., Poorbagher, H., Eagderi, S., Feghhi, J., Dormann, C.F., Khorshidi Nergi, S., Amiri, K. (2022). Modelling habitat preference of Caspian Kutum, *Rutilus Kutum*, using non-linear habitat suitability indices and generalised additive models. *Regional Studies in Marine Science*, 102715.
20- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., & Martin, T.G. (2013). Predicting species distributions for conservation decisions. *Ecology letters*. 16(12), 1424-1435.
21- Zhang, Z., Mammola, S., Xian, W. & Zhang, H. (2020). Modelling the potential impacts of climate change on the distribution of ichthyoplankton in the Yangtze Estuary, China. *Diversity and Distributions*. 26(1), 126-137.
22- Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S., Scroggie, M.P., & Woodford, L. (2011). Using presence-only and presence–absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*. 48(1), 25-34.
23- Robinson, C.L., Proudfoot, B., Rooper, C.N. & Bertram, D.F. (2021). Comparison of spatial distribution models to predict subtidal burying habitat of the forage fish *Ammodytes personatus* in the Strait of Georgia, British Columbia, Canada. *Aquatic Conservation: Marine and Freshwater Ecosystems*. 31(10), 2855-2869.
24- Li, G., Cao, J., Zou, X., Chen, X., & Runnebaum, J. (2016). Modeling habitat suitability index for Chilean jack mackerel (*Trachurus murphyi*) in the South East Pacific. *Fisheries Research*. 178, 47-60.
25- Olden, J.D., Jackson, D.A. & Peres-Neto, P.R. (2002). Predictive models of fish species distributions: a note on proper validation and chance predictions. *Transactions of the American Fisheries Society*. 131(2), 329-336.
26-Heggenes, J.A.N., (1996). Habitat selection by brown trout (*Salmo trutta*) and young Atlantic salmon (*S. salar*) in streams: static and dynamic hydraulic modelling. *Regulated Rivers: Research & Management*. 12(2-3), 155-169.
27- sFausch, K.D., Hawkes, C.L., & Parsons, M.G. (1988). Models that predict standing crop of stream fish from habitat variables: 1950-85. Gen. Tech. Rep. PNW-GTR-213. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 52 p, 213.
28- Costa, T.L., Pennino, M.G., & Mendes, L.F. (2017). Identifying ecological barriers in marine environment: The case study of Dasyatis marianae. *Marine environmental research*. 125, 1-9.
29- Martínez-Minaya, J., Cameletti, M., Conesa, D. & Pennino, M.G. (2018). Species distribution modeling: a statistical review with focus in spatio-temporal issues. *Stochastic environmental research and risk assessment*. 32(11), 3227-3244.
30- Moore, C., Drazen, J.C., Radford, B.T., Kelley, C. & Newman, S.J. (2016). Improving essential fish habitat designation to support sustainable ecosystem-based fisheries management. *Marine Policy*. 69, 32-41.
31- Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.C. (2010). New trends in species distribution modelling. *Ecography*. 33, 985-989.
32- Colloca, F., Bartolino, V., Lasinio, G.J., Maiorano, L., Sartor, P., & Ardizzone, G. (2009). Identifying fish nurseries using density and persistence measures. *Marine Ecology Progress Series.* 381, 287-296.
33- Beger, M., & Possingham, H.P. (2008). Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. *Marine Ecology Progress Series.* 361, 1-13.

34- Purkis, S.J., Graham, N.A.J. & Riegl, B.M. (2008). Predictability of reef fish diversity and abundance using remote sensing data in Diego Garcia (Chagos Archipelago). *Coral reefs*. 27(1), 167-178.
35- Pittman, S.J., Christensen, J.D., Caldow, C., Menza, C. & Monaco, M.E. (2007). Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecological modelling*. 204(1-2), 9-21.
36- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*. 17(1), 43-57.
37- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. *Diversity and distributions*. 14(5), 763-773.
38- Efron, B., & Hastie, T. (2016). Computer Age Statistical Inference. Cambridge University Press, Cambridge, UK
39- Moëzzi, F., Poorbagher, H., Eagderi, H., Feghhi, J. (2020). Spatiotemporal prediction of chlorophyll-a concentration in the Caspian Sea using logistic regression and Markov chain. *Environmental Resources Research*, 8(2), 175-186.
40- Elith, J., & Graham, C.H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*. 32(1), 66-77.
41- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*. 27(4), 437-448.
42- Moëzzi. F., Poorbagher, H., Eagderi, S., Feghhi, J. (2023). Comparing the performance of generalised linear model (GLM) and random forest (RF) models in predicting catch distribution of Caspian Kutum (*Rutilus frisii*). *Journal of Fisheries*, 76(1), 27-38. (In Persian)
43- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions*. 15(1), 59-69.
44-Morán-Ordóñez, A., Lahoz-Monfort, J.J., Elith, J. & Wintle, B.A. (2016). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions?. *Global Ecology and Biogeography*. 26(3), 371-384.
45-Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 36(1), 27-46.
46- Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007). What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological monographs*. 77(4), 615-630.
47- Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K., & Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological modelling*. 221(8), 1119-1130.
48- Bučas, M., Bergström, U., Downie, A.L., Sundblad, G., Gullström, M., Von Numers, M., Šiaulys, A., & Lindegarth, M. (2013). Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: evaluating the scope for predictive mapping using different modelling approaches. *ICES Journal of Marine Science*. 70(6), 1233-1243.
49- Moëzzi, F., Poorbagher, H., Eagderi, H. (2018). Evaluating effects of model type and importance of habitat parameters on prediction of biodiversity indices (a case study with fishes of the Totkabon River in the southern Caspian Sea). *Journal of Fisheries*, 71(3), 225-235.
50- Poorbagher, H., Eagderi, S., Moëzzi, F. (2023). Comparison of the performance of the linear and non-linear models in habitat suitability estimation (case study: Razi scraper, *Capoeta razii*). *Journal of Natural Environment*, Published online. (In Persian)
51- Zhang, Y., Xu, B., Ji, Y., Zhang, C., Ren, Y. & Xue, Y. (2021). Comparison of habitat models in quantifying the spatio-temporal distribution of small yellow croaker (*Larimichthys polyactis*) in Haizhou Bay, China. *Estuarine, Coastal and Shelf Science*. 261, p.107512.
52- Shabani, F., Kumar, L. & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and evolution*. 6(16), 5973-5986.

53- Ward, E.J., Jannot, J.E., Lee, Y.W., Ono, K., Shelton, A.O. & Thorson, J.T. (2015). Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications*. 25(8), 2198-2209.
54- Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E. & Lundquist, C.J., (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*. 4, p.421.
55-Guisan, A., Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*. 8(9), 993-1009.
56- Wootton, J.T. (1992). Indirect effects, prey susceptibility, and habitat selection: impacts of birds on limpets and algae. *Ecology*. 73(3), 981-991.
57- Mountrakis, G., Im, J. & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 66(3), 247-259.
58- Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*. 12(5), 361-371.
59-Li, M., Zhang, C., Xu, B., Xue, Y. & Ren, Y. (2017). Evaluating the approaches of habitat suitability modelling for whites potted conger (*Conger myriaster*). *Fisheries Research*. 195, 230-237.
60- Beecher, H.A., Caldwell, B.A. & DeMond, S.B. (2002). Evaluation of depth and velocity preferences of juvenile coho salmon in Washington streams. *North American Journal of Fisheries Management*. 22(3), 785-795.
61- Vadas Jr, R.L. & Orth, D.J. (2001). Formulation of habitat suitability models for stream fish guilds: do the standard methods work? *Transactions of the American Fisheries Society*. 130(2), 217-235.
62- Chen, X., Tian, S., Chen, Y., & Liu, B. (2010). A modeling approach to identify optimal habitat and suitable fishing grounds for neon flying squid (*Ommostrephes bartramii*) in the Northwest Pacific Ocean. *Fishery Bulletin*. 108(1).
63- Wang, D., Yu, J., Lin, Z. & Chen, P. (2023). Spatial–Temporal Distribution of Fish Larvae in the Pearl River Estuary Based on Habitat Suitability Index Model. *Biology*, 12(4), p.603.
64- Brown, S.K., Buja, K.R., Jury, S.H., Monaco, M.E., & Banner, A. (2000). Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays, Maine. *North American Journal of Fisheries Management*. 20(2), 408-435.
65- Bovee, K.D. (1986). *Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology* (No. 21). National Ecology Center, Division of Wildlife and Contaminant Research, Fish and Wildlife Service, US Department of the Interior.
66- Korman, J., Perrin, C.J., & Lekstrum, T. (1994). A guide for the selection of standard methods for quantifying sportfish habitat capability and suitability in streams and lakes of British Columbia. Limnotek Research and Development Inc., Vancouver, for British Columbia Environment, Fisheries Branch. *Research and Development Section, Vancouver*.
67- Jowett, I.G. (2003). Hydraulic constraints on habitat suitability for benthic invertebrates in gravel-bed rivers. *River Research and Application*. 19(5-6), 495-507.
68- Leclerc, M., Saint-Hilaire, A. & Bechara, J. (2003). State-of-the-art and perspectives of habitat modelling for determining conservation flows. *Canadian Water Resources Journal*. 28(2), 135-151.
69- Nykänen, M., Huusko, A. & Mäki-Petäys, A. (2001). Seasonal changes in the habitat use and movements of adult European grayling in a large subarctic river. *Journal of Fish Biology*. 58(2), 506-519.
70-Xu, F., Du, Y.A., Chen, H. & Zhu, J.M. (2021). Prediction of fish migration caused by ocean warming based on SARIMA model. *Complexity*, *2021*.
71- Potts, J.M., & Elith, J. (2006). Comparing species abundance models. *Ecological modelling*. 199(2), 153-163.
72- Breiman, L. (2001). Random forests. *Machine Learning.* 45, 5–32.
73- Tyler, E.H., Speight, M.R., Henderson, P. & Manica, A. (2009). Evidence for a depth refuge effect in artisanal coral reef fisheries. *Biological Conservation*. 142(3), 652-667.
74- Moisen, G.G. & Frescino, T.S. (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling.* 157(2-3), 209-225.
75- McCullagh, P., & Nelder, J.A. (2019). Generalized linear models. Routledge, New York. P.532.

76- Myers, R.H., Montgomery, D.C., Vining, G.G. & Robinson, T.J. (2012). *Generalized linear models: with applications in engineering and the sciences*. John Wiley & Sons.
77- Ahmadi-Nedushan, B., St-Hilaire, A., Bérubé, M., Robichaud, É., Thiémonge, N., & Bobée, B. (2006). A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications*. 22(5), 503-523.
78- Sguotti, C., Lynam, C.P., García-Carreras, B., Ellis, J.R. & Engelhard, G.H. (2016). Distribution of skates and sharks in the North Sea: 112 years of change. *Global change biology*. 22(8), 2729-2743.
79- Campbell, R.A. (2015). Constructing stock abundance indices from catch and effort data: Some nuts and bolts. *Fisheries Research*. 161, 109-130.
80- Hart, R.A. (2012). Stock assessment of brown shrimp (*Farfantepenaeus aztecus*) in the US Gulf of Mexico for 2011. NOAA technical memorandum NMFS-SEFSC; 638
81- Li, Z., Ye, Z., Wan, R., & Zhang, C. (2015). Model selection between traditional and popular methods for standardizing catch rates of target species: a case study of Japanese Spanish mackerel in the gillnet fishery. *Fisheries Research*. 161, 312-319.
82- Lehmann, A., Overton, J.M., & Leathwick, J.R. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*. 157(2-3), 189-207.
83- Knudby, A., Brenning, A., & LeDrew, E. (2010). New approaches to modelling fish–habitat relationships. *Ecological Modelling*. 221(3), 503-511.
84- Wood, S.N. (2017). Generalized additive models: an introduction with R. CRC Press Taylor and Francis Group.
85- Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*. 321, 267-281.
86- Rezaei, R. & Sengül, H. (2018). Development of generalized additive models (GAMs) for *Salmo rizeensis* endemic to north-eastern streams of Turkey. *Turkish Journal of Fisheries and Aquatic Sciences*. 19(1), 29-39.
87- Murase, H., Nagashima, H., Yonezaki, S., Matsukura, R. & Kitakado, T. (2009). Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. *ICES Journal of Marine Science*. 66(6), 1417-1424.
88- Drexler, M., & Ainsworth, C.H. (2013). Generalized additive models used to predict species abundance in the Gulf of Mexico: an ecosystem modeling tool. *PloS one*. 8(5), p.e64458.
89- Grüss, A., Chagaris, D.D., Babcock, E.A., & Tarnecki, J.H. (2018). Assisting ecosystem-based fisheries management efforts using a comprehensive survey database, a large environmental database, and generalized additive models. *Marine and Coastal Fisheries*. 10(1), 40-70.
90- Lauzeral, C., Grenouillet, G., & Brosse, S. (2013). Spatial range shape drives the grain size effects in species distribution models. *Ecography*. 36(7), 778-787.
91- Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*. 16(4), 1145-1157.
92- Hastie, T., & Tibshirani, R. (1990) Generalized Additive Models. London: Chapman and Hall, 335 pp.
Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*. 83(2), 171-193.
93- Gobeyn, S., Mouton, A.M., Cord, A.F., Kaim, A., Volk, M. & Goethals, P.L. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling*. 392, 179-195.
94-Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*. 1, p.e8621.
95-Elith, J., Leathwick, J.R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology*. 77(4), 802-813.
96- Olden, J.D., Lawler, J.J. & Poff, N.L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*. 83(2), 171-193.
97- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology*. 88(11), 2783-2792.

98- Oppel, S., Strobl, C., & Huettmann, F. (2009). Alternative methods to quantify variable importance in ecology. Ludwig-Maximilians-Universität, München, 1–7
99- Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. Springer. New York.
100- Halsey, L.G. (2019). The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology letters*. 15(5), p.20190174.
101- Pittman, S.J., & Brown, K.A. (2011). Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PloS one.* 6(5), p.e20583.
102- Hua, C., Li, F., Zhu, Q., Zhu, G., & Meng, L. (2020). Habitat suitability of Pacific saury (*Cololabis saira*) based on a yield-density model and weighted analysis. *Fisheries Research*. 221, p.105408.
103- Xue, Y., Guan, L., Tanaka, K., Li, Z., Chen, Y. & Ren, Y. (2017). Evaluating effects of rescaling and weighting data on habitat suitability modeling. *Fisheries Research*. 188, 84-94.
104- Froeschke, J.T. & Froeschke, B.F. (2011). Spatio-temporal predictive model based on environmental factors for juvenile spotted seatrout in Texas estuaries using boosted regression trees. *Fisheries Research*. 111(3), 131-138.
105- Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press.
106- Dai, L., Hodgdon, C., Tian, S., Chen, J., Gao, C., Han, D., Kindong, R., Ma, Q., & Wang, X. (2020). Comparative performance of modelling approaches for predicting fish species richness in the Yangtze River Estuary. *Regional Studies in Marine Science*. 35, p.101161.
107- França, S., & Cabral, H.N. (2015). Predicting fish species richness in estuaries: which modelling technique to use?. *Environmental Modelling & Software*. 66, 17-26.
108- McKenney, D.W. & Pedlar, J.H. (2003). Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. *Forest Ecology and Management*. 175(1-3), 497-507.
109- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*. 1189-1232.
110- Hastie, T., Tibshirani, R., & Friedman, J.H. (2001) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.
111- Froeschke, J.T., & Froeschke, B.F. (2016). Two-stage boosted regression tree model to characterize southern flounder distribution in Texas estuaries at varying population sizes. *Marine and coastal fisheries.* 8(1), 222-231.
112- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., & Li, J. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 29(2), 129-151.
113- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*. 7(1), 1-13.
114-Siroky, D.S. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*. 3, 147-163.
115- Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R news*. 2(3), 18-22.
116- Olaya-Marín, E.J., Martínez-Capel, F. & Vezza, P. (2013). A comparison of artificial neural networks and random forests to predict native fish species richness in Mediterranean rivers. *Knowledge and Management of Aquatic Ecosystems*, (409), p.07.
117- Moguerza, J.M. & Muñoz, A. (2006). Support vector machines with applications. *Statistical Science*. 21(3), 322-336.
118- Vapnik, V. (1995). The nature of statistical learning theory. In: Information science and statistics. Springer-Verlag, New York, NY (USA).
119- Pinser, D.A. & Schnyer, D.M. (2020). Support vector machine. In: Machine learning (pp. 101-121). Academic Press.
120- Munoz-Mas, R., Fukuda, S., Pórtoles, J. & Martinez-Capel, F. (2018). Revisiting probabilistic neural networks: a comparative study with support vector machines and the microhabitat suitability for the Eastern Iberian chub (*Squalius valentinus*). *Ecological Informatics*. 43, 24-37.
121- Witten, I.H., Frank, E., Hall, M.A. (2011). Data Mining, Practical Machine Learning Tools and Techniques, 3rd ed. Morgan Kaufmann, San Francisco, p. 629