

УДК 639.2.053.1:551.46

В.В. Кулик, А.А. Байталюк, О.Н. Катугин, Е.И. Устинова*Тихоокеанский филиал ВНИРО (ТИНРО),
690091, г. Владивосток, пер. Шевченко, 4**МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ УЛОВОВ САЙРЫ
В СВЯЗИ С ФАКТОРАМИ ОКРУЖАЮЩЕЙ СРЕДЫ**

При моделировании использовали ежедневную информацию, полученную из многомерной системы вариационной оценки океана (MOVE), доступную для области между 140 и 159° в.д. В эту область попадает около 95 % всех уловов, предоставленных Комиссии по рыболовству в северной части Тихого океана (NPFC), и 100 % российских уловов в 1994–2017 гг. Позиции судов на лову и в поиске сайры интерполированы с разрешением в 1 км из Отраслевой системы мониторинга, а температура поверхности моря (SST) — из многомасштабного анализа сверхвысокого разрешения. Общее пространственное разрешение для моделей приведено в соответствие самому грубому — в системе MOVE (0,1° по широте и долготе). Сначала мы оценили вклад в вероятность нахождения вида и значение на перестановках 184 предикторов из возможных комбинаций ежедневных продуктов SST и MOVE с задержкой от 0 до 7 дней и окном скользящей средней от 0 до 7 дней в MaxEnt. SST, температура воды (WT) и ее градиент (WTG) на глубине 50 м для текущего дня вылова и до 2 предыдущих дней для SST и от 3 до 7 дней для WT и WTG на глубине 50 м имели самый высокий вклад и значения на перестановках. Затем мы испытали более универсальные методы — обобщенные аддитивные модели (GAM) и случайного леса (Random forest) с этими предикторами. Последний превзошел MaxEnt и GAM по различным показателям прогностической точности. Его точность достигла 0,86 с площадью под кривой ошибок AUC = 0,7. Годовая сумма площадей с условиями, предпочитаемыми сайрой согласно случайному лесу в ИЭЗ, показывает значительную ($p < 0,05$) корреляцию (0,96) с общими уловами сайры в последние экстремальные годы ее вылова (максимальный вылов сайры в ИЭЗ в 2008 г., максимальный вылов сайры всеми участниками NPFC в 2014 г., минимальный вылов сайры в ИЭЗ в 2017 г. и максимальный вылов сайры всеми участниками NPFC в Конвенционном районе в 2018 г.).

Ключевые слова: тихоокеанская сайра, вылов, Тихий океан, модель распределения вида, SST, максимальная энтропия, обобщенная аддитивная модель, случайный лес.

DOI: 10.26428/1606-9919-2019-199-213.

* Кулик Владимир Владимирович, кандидат биологических наук, начальник отдела, e-mail: vladimir.kulik@tinro-center.ru; Байталюк Алексей Анатольевич, кандидат биологических наук, заместитель директора ВНИРО — руководитель Тихоокеанского филиала ВНИРО (ТИНРО), e-mail: aleksei.baitaliuk@tinro-center.ru; Катугин Олег Николаевич, кандидат биологических наук, начальник отдела, e-mail: oleg.katugin@tinro-center.ru; Устинова Елена Ивановна, кандидат географических наук, ведущий научный сотрудник, e-mail: elena.ustinova@tinro-center.ru.

Kulik Vladimir V., Ph.D., head of department, Pacific branch of VNIRO (TINRO), Shevchenko Alley, 4, Vladivostok, 690091, Russia, e-mail: vladimir.kulik@tinro-center.ru; Baitaliuk Aleksey A., Ph.D., deputy head of VNIRO — head of Pacific branch of VNIRO (TINRO), Shevchenko Alley, 4, Vladivostok, 690091, Russia, e-mail: aleksei.baitaliuk@tinro-center.ru; Katugin Oleg N., Ph.D., head of department, Pacific branch of VNIRO (TINRO), Shevchenko Alley, 4, Vladivostok, 690091, Russia, e-mail: oleg.katugin@tinro-center.ru; Ustinova Elena I., Ph.D., leading researcher, Pacific branch of VNIRO (TINRO), Shevchenko Alley, 4, Vladivostok, 690091, Russia, e-mail: elena.ustinova@tinro-center.ru.

Kulik V.V., Baitaliuk A.A., Katugin O.N., Ustinova E.I. Modeling distribution of saury catches in relation with environmental factors // *Izv. TINRO.* — 2019. — Vol. 199. — P. 193–213.

Pacific saury *Cololabis saira* is widely distributed in the North Pacific, with commercial harvesting in the area between 140–172° E. Relationship of its commercial catches distribution with environmental factors is investigated using the daily SST data, the daily data set of multivariate ocean variational estimation system (MOVE) produced by Meteorological Research Institute (Japan) for the area between 140–159° E (about 95 % of all catches and 100 % of the Russian catches of saury were landed in this area in 1994–2017), and the daily set of saury catches position with 1 km resolution collected by the Russian vessel monitoring system. Spatial resolution for all data sets is upscaled to the resolution of MOVE system (0.1 x 0.1 degree). Contribution and permutation importance for the catch distribution are estimated for 184 possible combinations of SST and MOVE products with the lags of 0–7 days and moving average window from 0 to 7 days using the method of maximum entropy (MaxEnt). For synchronic relationships, the best results are found for SST, water temperature at 50 m depth and its spatial gradient, moreover, SST provides high contribution with the lag up to 2 days and the temperature at 50 m and its gradient — with the lag 3–7 days. The same sets of environmental parameters are used for the catches distribution modeling with GAMs and Random Forest techniques; the latter method shows better accuracy and other indices of the confusion matrix. Year-to-year changes of the total area with predicted conditions favorable for the saury fishery within the EEZ of Russia and Japan correlate strongly ($r = 0.96$, $p < 0.05$) with the total annual catch of saury, in particular for the extreme years (high catches in 2008, 2014, and 2018, low catch in 2017).

Key words: pacific saury, catch, North Pacific, SDM software (species distribution modeling), sea surface temperature (SST), method of maximum entropy (MaxEnt), generalized additive model (GAM), Random Forest technique.

Введение

Тихоокеанская сайра *Cololabis saira*, Брев. — короткоцикловая рыба, широко распространенная в 10–30-метровом слое эпипелагиали по всей северной части Тихого океана [Филатов и др., 2011]. Нерест сайры происходит на субстрат, поэтому предполагалось, что этот вид неретический и его ареал разорван в центре северной части Тихого океана (ЦСТО), но И.Б. Бирман сообщил о скоплениях сайры в ЦСТО в 1958 г. после экспедиции на НИС «Витязь» [Парин, 1960]. Позже скопления сайры в ЦСТО были найдены японскими исследователями [Kobayashi et al., 1968]. С 1980-х гг. при проведении экспедиций с весны по осень в ЦСТО регулярно обнаруживаются скопления сайры [Байталюк, Давыдова, 2002; Baitaliuk et al., 2013]. В качестве нерестового субстрата сайра может использовать любые плавающие предметы, к которым икринки способны приклеиться нитями. В настоящее время считается, что ее ареал охватывает обширную акваторию в Тихом океане между 20 и 60° с.ш. [Филатов и др., 2011]. По данным Комиссии по рыболовству в северной части Тихого океана (North Pacific Fisheries Commission — NPFC) с 1994 по 2017 г. ни один участник этой региональной рыбохозяйственной организации не добывал сайру восточнее 172° в.д., а 95 % всего вылова пришлось на район между 140 и 159° в.д. Российские рыбаки получили 99,5 % всего своего улова здесь же. Максимальные уловы крупной сайры по данным экспедиционного промысла Тайваня располагались до 150° в.д., а среднеразмерной — до 155° в.д. [Chang et al., 2019].

Миграция сайры в район восточнее Курильских островов начинается в июле. После захода в российские воды косяки сайры перемещаются в юго-западном направлении, начиная свой первый и для большинства рыб последний нерест, уже в августе, а продолжающие нагул особи смещаются с водами субарктической структуры в юго-западном и южном направлениях [Филатов, 2015]. В июне-августе скопления нагуливающейся сайры образуются преимущественно у фронтов течений Ойасио и Камчатско-Курильского, а также в их циклонических вихрях, которые за сутки могут смещаться до 15 миль, и сайра в них держится, пока не снизятся концентрации кормового зоопланктона [Филатов, 2015]. Статистическая связь вылова сайры японскими рыбаками с близостью фронта Ойасио уже известна и оценена давно [Yasuda, Watanabe, 1994]. Она объясняется временем, необходимым на один поход японского судна к

скоплению сайры и обратно в порт для сдачи свежей рыбы: если фронт близко, поход занимает 2–3 дня, а если удален — неделю.

Формирование благоприятных для нагула и образования сайровых скоплений районов определяется комплексом факторов [Филатов и др., 2011]. Важнейшими из них являются вертикальная термическая структура вод, характеризующаяся определенной толщиной верхнего квазигодородного слоя (не менее 6 м) и вертикальным градиентом сезонного термоклина $\geq 0,19$ °C/м [Филатов, 1984], а также доступность кормового зоопланктона. Сама по себе температура, как и ее аномалии, может быть очень разной в успешные/неуспешные для сайрового промысла годы. Однако «геометрия» поля температуры поверхности океана (SST — Sea Surface Temperature) служит индикатором важных динамических процессов, влияющих и на вертикальную структуру вод, и на пространственное распределение кормового зоопланктона. Связь концентраций сайры с гидрологическими структурами обнаружена еще в начале прошлого века [Uda, 1936].

Оперативное океанографическое обеспечение сайрового промысла традиционно базировалось на судовой и дистанционной информации о поле температуры воды. В современный период эта характеристика также активно используется в оперативной работе. Количественно вклад SST оценивали в помесечных моделях индексов пригодности местообитания (Habitat Suitability Indices — HSI) тихоокеанской сайры. Вклад температуры поверхности океана в оценку HSI максимален в сравнении с косвенным показателем продуктивности вод (Sea Surface Chl-*a* — SSC), вихревой кинетической энергией (Eddy Kinetic Energy — EKE) и аномалиями высоты поверхности моря (Sea-Surface Height — SSH). Максимальный вклад SST сначала оценивался в 93 % в октябре [Syah et al., 2014], но позже эта оценка уменьшилась до 72 %, а максимум составил 73 % в ноябре [Syah et al., 2016]. Оценка вклада SST в оценку HSI производилась в одной из самых распространенных платформ машинного обучения для моделирования распределения видов (Species Distribution Modelling — SDM) методом максимальной энтропии — MaxEnt [Phillips et al., 2006].

Обобщенные аддитивные модели (Generalized Additive Models — GAM) также показывают самое большое снижение информационного критерия Акаике (An Information Criterion — AIC [Sakamoto et al., 1986]) при использовании SST в дополнение к SSC, EKE и аномалиям SSH. Следовательно, и в GAM вклад SST в определении HSI максимален [Syah et al., 2017]. Модели HSI, настроенные по уловам экспедиционного промысла Тайваня в зависимости от SST, SSH, поверхностной солености (Sea Surface Salinity — SSS) и чистой первичной продуктивности (Net Primary Production — NPP), также в помесечном временном масштабе показывают, что SST описывает максимальную долю дисперсии (7 %) [Chang et al., 2019].

Аналогичный вывод следует из документов технической рабочей группы NPFC по определению состояния запаса сайры, в которых SST используется в оптимизированных моделях при стандартизации как помесечных, так и суточных уловов с использованием GAM и других обобщенных линейных моделей (Generalized Linear Model — GLM) [Кулик, 2018].

Посуточное временное разрешение моделирования относится к краткосрочному прогнозированию. Прогноз пространственного распределения потенциально промысловых районов на 3–5 сут вперед позволяет сузить акваторию поиска рыб в 8–10 раз. При посуточном прогнозировании площадь поиска сокращается еще больше, поэтому в ТИНРО разрабатывалась диалоговая система для краткосрочного прогнозирования промысловой обстановки (КППО), в первую очередь для сайры, сардины и др. [Бочаров, 1993]. Тем не менее создание формализованного метода КППО не было полностью осуществлено [Филатов и др., 2011].

Максимальный вылов сайры всеми участниками NPFC приходится на 2014 г. — 630 тыс. т. Россия тогда в исключительной экономической зоне (ИЭЗ) получила последний улов выше среднего — 64 тыс. т — при среднем вылове 37 тыс. т за период с 1995 по 2018 г. Затем уловы России в ИЭЗ только сокращались, составляя 30,0 % в 2015 г., 19,0 в 2016 г., 4,1 в 2017 г. и 3,6 % в 2018 г. от улова России в ИЭЗ в 2014 г. Этот обвал возник

не только в связи с уменьшением биомассы сайры и плотности ее скоплений вплоть до 2017 г., но и с более чем шестикратным сокращением числа судов под российским флагом в ИЭЗ: с 61 судна на лову в 2014 г. до 10 судов в 2018 г., из которых только 7 продержалось большую часть путины. Максимальный вылов сайры в ИЭЗ приходится на 2008 г. — 448 тыс. т, когда вылов России достиг второго результата в ИЭЗ — 78 тыс. т — при 49 судах, а российский максимум вылова сайры в ИЭЗ пришелся на 2007 г., составив 95 тыс. т при 61 судне. Очевидно, что при прочих равных условиях число судов не определяет вылов линейно.

Япония смогла преодолеть исторический минимум уловов в ИЭЗ, который в 2017 г. составил в сумме для всех участников NPFC всего 74 тыс. т, нарастив число судов в Конвенционном районе от 102 до 125. Это позволило Японии увеличить вылов сайры в Конвенционном районе в 3 раза — до 47 тыс. т в 2018 г. В сумме все участники NPFC поставили новый рекорд добычи сайры в этом районе в 2018 г. — 350 тыс. т, но Россия, выставив всего 4 судна в открытых водах, получила минимальный относительно других участников вылов — всего 5,5 тыс. т, уступив Вануату (8,2 тыс. т) с тем же числом судов (4) [<https://www.npfc.int/summary-footprint-pacific-saury-fisheries>].

В условиях малого количества российских судов на промысле сайры как в открытых водах, так в ИЭЗ, когда практически невозможно вести полноценную промысловую разведку и охватывать большие площади поисковыми операциями, возрастает необходимость применения КППО, при котором распределение сайры моделируется с использованием дистанционных данных. Таким образом, цель настоящего исследования — оценить точность методов машинного обучения для КППО сайры.

В отношении тихоокеанской сайры имеются примеры успешного применения как традиционных SDM для определения HSI — MaxEnt, так и более общего метода — GAM. Опыт автоматизации КППО в ТИНРО показывает, что однозначно выбрать оптимальную модель практически невозможно, но обобщенное решение по нескольким моделям зачастую оказывается лучше решений, полученных на основании отдельных моделей. В простейшем случае задача классификации перспективных и неперспективных районов промысла с использованием множества различных алгоритмов классификации может быть решена с использованием процедуры голосования в комитете или ансамбле моделей [Бочаров, 1993]. Этому принципу полностью соответствует метод случайного леса — Random forest [Breiman, 2001], в основе которого лежит использование комитета (ансамбля) деревьев принятия решений. Мы испытаем Random forest (RF) в дополнение к MaxEnt и GAM. Последние методы уже показали большую площадь ($> 0,8$) под кривой ошибок (area under the receiver operating characteristic curve — AUC) в MaxEnt [Syah et al., 2016], что на 30 % выше случайного выбора (когда AUC = 0,5). В GAM объяснялось 36 % дисперсии [Syah et al., 2017]. Прежние GAM и MaxEnt настроены в месячном масштабе и без временных задержек, что неприемлемо для оперативного прогнозирования, когда прогноз требуется на несколько дней вперед, а не за прошедший месяц.

Материалы и методы

Проведены испытания надежности прогнозов, основанных на гидрологических данных из общедоступных источников с временными задержками до одной недели. Более длительные временные задержки могут быть исследованы позднее, если среди оцененных предикторов будут обнаружены такие, у которых сдвиг окончания окна осреднения достигнет 7 дней, и при этом их вклад в HSI будет выше, чем у предикторов, более близких к датам наблюдений. Биологической причиной для задержек является невозможность мгновенного концентрирования зоопланктона, на полях которого может задержаться сайра. Таким образом, после возникновения комплекса благоприятных условий должно пройти несколько дней, а мы должны попытаться учесть эти задержки, так как прямых наблюдений кормового зоопланктона за все дни промысла по всей акватории у нас нет.

В материалах для SDM большое значение имеют реальные «нулевые» наблюдения — это те наблюдения, где точно известно, что искомым признаком классификации событий (например, поимка вида) отсутствовал (но его поиск был). В нашем случае «нуль» не обозначает полное отсутствие сайры, а обозначает лишь то, что в месте поиска по тем или иным причинам капитан не решил остановиться и начать лов, поэтому далее «нули» идут в кавычках, чтобы напоминать о цели исследования — КППО для судов, активно ищущих высокие концентрации сайры. Сайра наблюдается в научных съемках в ИЭЗ РФ Японского, Охотского и Берингова морей, но промысел сайры Россией там не ведется [Филатов и др., 2011].

В предыдущем исследовании SDM сайры [Syah et al., 2016] «нулевые» уловы имитировались: случайно выбирались вокруг мест уловов, так как места безуспешного поиска сайры в наблюдениях отсутствовали. Метод MaxEnt изначально разрабатывался с допущением о возможности его использования в условиях отсутствия реальных «нулевых» наблюдений [Guillera-Arroita et al., 2014], но наш личный [Kulik et al., 2016; Кулик и др., 2017] и мировой опыт показывает, что замена имитированных «нулевых» наблюдений реальными способна значительно улучшить прогностическую способность SDM [Li, Guo, 2013].

Мы использовали все позиции судов, ловивших сайру бортовыми ловушками, сохраненные в Отраслевой системе мониторинга (ОСМ) Центра системы мониторинга рыболовства и связи с 2004 по 2018 г. в ночное время, которое определялось по углу захода солнца за горизонт глубже 6 астрономических градусов по каждой паре координат и всемирному координированному времени по алгоритму [Michalsky, 1988], реализованному со всеми последовавшими исправлениями в пакете *oce* [<https://cran.r-project.org/web/packages/oce>] для языка программирования R [<https://cran.r-project.org/>].

Во время поисковых операций суда двигаются с высокой скоростью, поэтому мы интерполировали дополнительные позиции между соседними через 1 км в расширении системы управления базами данных — PostgreSQL [<http://www.postgresql.org/>] для решения геоинформационных задач — PostGIS [<http://postgis.net/>] — на сфероиде с параметрами из всемирной системы геодезических параметров Земли 1984 г. (WGS 84). Классификация операций на поисковые (реальные «нули» в бинарной классификации), на лову (истинные «единицы») и прочие осуществлена согласно времени каждой операции, соответствующего диапазонам бюджета времени каждого судна в ОСМ. В случае пометки позиции «на лову», но скорости движения к следующей позиции более 3 м/с такая пометка изменялась на «поиск». Прочие операции и позиции, отстоящие от предыдущей позиции дальше, чем судно могло удалиться на максимальной скорости, указанной в реестре судов ОСМ, удалены.

За каждые сутки по Гринвичу, выраженные порядковым номером дня в году, а также до семи предыдущих для каждой позиции (паре координат) каждого судна получены:

1) значения SST с разрешением 0,01 географического градуса из проекта Multi-scale Ultra-high Resolution (MUR) sea surface temperature [Chin et al., 2017];

2) градиенты SST (SSTG), рассчитанные нами с использованием алгоритма Белкина [Belkin, o'Reilly, 2009] в пакете *grec* [<https://cran.r-project.org/web/packages/grec/>] для языка R по SST из проекта MUR;

3) температуры (T) на глубине 50 м с разрешением 0,1 географического градуса из проекта Meteorological Research Institute Multivariate Ocean Variational Estimation (MOVE) system [Usui et al., 2006];

4) градиенты T (TG) на глубине 50 м, рассчитанные нами с использованием алгоритма Белкина в пакете *grec* для языка R по T из проекта MOVE;

5) T на глубине 100 м из проекта MOVE;

6) TG на глубине 100 м, рассчитанные нами с использованием алгоритма Белкина в пакете *grec* для языка R по T из проекта MOVE;

7) скорости течений на поверхности в северном (*v*) направлении из проекта MOVE;

8) скорости течений на поверхности в восточном (*u*) направлении из проекта MOVE;

9) EKE , рассчитанный по формуле (1):

$$EKE = \frac{v^2 + u^2}{2}. \quad (1)$$

Дополнительно получены по 7 колонок по каждому из 9 вышеперечисленных предикторов (задержка во времени до 7 дней). Они использованы далее для расчета скользящего среднего значения для каждой позиции каждого судна за каждую ночь с окном от 1 до 7 дней в прошлое. Итого получены 207 колонок по 9 гидрологическим предикторам.

Самым грубым разрешением (0,1 географического градуса) обладают данные из системы MOVE, поэтому остальные наборы (координаты позиций через 1 км и SST из MUR, их градиенты через 0,01 географического градуса) также усреднены до масштаба в 0,1 географического градуса. Эти же данные (MOVE) послужили границей исследования на севере (49° с.ш.) и востоке (159° в.д.). В нашем случае при использовании данных из ОСМ мы потеряли только 0,5 % от сохраненных там уловов. В итоге из 6,7 млн наблюдений (позиций улова или поиска сайры), интерполированных через 1 км, осталось 216 489 в разрешении 0,1 географического градуса.

Включение температуры воды на горизонтах 50 и 100 м в ряд исследуемых параметров, описывающих состояние среды, основано на следующем. Как правило, в СЗТО она лучше отражает динамику вод и «биологически значимые» крупномасштабные и мезомасштабные фронтальные зоны, чем SST, особенно в летний период и в начале осени. В период максимальных скоростей летнего прогрева вод состояние поверхностного слоя существенно зависит от ветровых условий, так как при штилевой погоде толщина хорошо прогретого верхнего слоя может достигать всего лишь нескольких метров, и здесь из-за высокой стратификации вод и отсутствия перемешивания с нижележащими холодными слоями могут сформироваться значительные положительные аномалии температуры, которые мгновенно разрушатся при первом же шторме. Таким образом, в летний период и в начале осени SST отражает термическое состояние лишь самого тонкого поверхностного прогретого слоя, и, как показали результаты анализа океанографических съемок ТИНПРО, уже на горизонтах 10–30 м знак аномалий может смениться на противоположный. Согласно этим же съемкам еще более репрезентативной, возможно, была бы температура на горизонте 30 м, однако в оперативном доступе таких данных для всего промыслового сезона не существует. По перепаду температуры на поверхности и на горизонте 50 м можно косвенно судить о градиенте в сезонном термоклине. Обычно летом в ЮКР глубина залегания нижней границы сезонного термоклина не превышает 50 м [Истоки Ойясио, 1997] с дальнейшим ее заглублением по мере развития осеннего охлаждения. Осеннее охлаждение достигает этого горизонта в районах промысла сайры в ноябре. Поле температуры воды на горизонте 100 м служит индикатором границ распространения вод, переносимых важнейшими течениями СЗТО — Куроисио и Ойясио [Kawai, 1969, 1972; Ogawa, 1989; Takasugi, Yasuda, 1993; и др.]. Так, согласно работе Yoshida [1992] для Ойясио такой границей является изотерма 5 °C на этом горизонте.

Доступ к базе данных по температуре воды на горизонтах 50 и 100 м Японского метеорологического агентства JMA был получен в рамках международного проекта NEAR-GOOS [<https://ds.data.jma.go.jp/gmd/goos/data/trtdb/jma-pro.html>], в которой подповерхностная температура оценивалась с помощью численной модели усвоения данных об океане [Usui et al., 2006], лежащей в основе многомерной системы вариационного оценивания состояния океана Метеорологического научно-исследовательского института Японии (Meteorological Research Institute Multivariate Ocean Variational Estimation system, MOVE/MRI.COM), которая в настоящее время продолжает совершенствоваться [Usui et al., 2017].

Выбор самых информативных предикторов осуществлен в MaxEnt в пакете для языка R — *dismo* [<https://cran.r-project.org/web/packages/dismo>] сначала по первым восьми предикторам (и их комбинациям со сдвигом различных окон усреднения в прошлое), составив первый набор данных, а затем аналогично по первым шести и EKE ,

составив второй набор данных. Такое деление вызвано тем, что *EKE* детерминирован скоростями течений (формула (1)), поэтому совместное их включение в мультипликативную модель с задержками, где скорости в одной и той же ячейке в соседние дни мало различаются, бессмысленно.

В прежней работе по SDM сайры в MaxEnt [Syah et al., 2016] прогностическая способность была показана через AUC, но в случае имитированных «нулевых» наблюдений использование AUC некорректно, а вместо него лучше применять специальный индекс — F по формуле (2) [Li, Guo, 2013]:

$$F = (1 + \beta^2) \cdot \frac{p \cdot r}{(\beta^2 \cdot p) + r}, \quad (2)$$

где β — положительный действительный коэффициент; p — положительная прогностическая ценность, определяемая по формуле (3); r — чувствительность (sensitivity, или recall), определяемая по формуле (4):

$$p = \frac{A}{A + B}; \quad (3)$$

$$r = \frac{A}{A + C}, \quad (4)$$

где A — число верно определенных (true positive) мест улова (где прогноз HSI преодолел пороговое значение, отделяющее «единицу» от «нуля», и совпал с истинной «единицей»); B — число ложноположительных (false positive) прогнозов (мест, где прогнозировался улов, а на самом деле поиск там был, но улова не было); C — число ложноотрицательных (false negative) прогнозов (мест, где прогнозировалось отсутствие улова, а улов на самом деле был).

В нашем случае использование AUC корректно, но мы также рассчитали F при $\beta = 1$, что соответствует равномерному взвешиванию всех типов ошибок [Li, Guo, 2013], а также и другие характеристики матрицы неточностей — специфичность (*Specificity*) по формуле (5), детектирование доминирования (*Detection Prevalence*) по формуле (6), доминирование (*Prevalence*) по формуле (7), сбалансированную точность (*Balanced Accuracy*) по формуле (8) и общую точность (*Accuracy*) по формуле (9):

$$Specificity = \frac{D}{B + D}; \quad (5)$$

$$Detection\ Prevalence = \frac{A + B}{A + B + C + D}; \quad (6)$$

$$Prevalence = \frac{A + C}{A + B + C + D}; \quad (7)$$

$$Balanced\ Accuracy = \frac{r + Specificity}{2}; \quad (8)$$

$$Accuracy = \frac{A + D}{A + B + C + D}, \quad (9)$$

где D — число верно предсказанных мест отсутствия улова (true negative). С общей точностью связан показатель относительного превосходства над случайным гаданием — индекс Каппа (κ), рассчитываемый по формуле (10) [Cohen, 1960]:

$$\kappa = \frac{Accuracy - p_c}{1 - p_c}, \quad (10)$$

где p_c — вероятность случайного соответствия по формуле (11):

$$p_c = Detection\ Prevalence \cdot Prevalence + \frac{C + D}{A + B + C + D} \cdot \frac{B + D}{A + B + C + D}. \quad (11)$$

Индекс Каппа до сих пор широко используется, но в определенных условиях он недооценивает вероятность корректности классификации [Olofsson et al., 2014],

поэтому некоторые исследователи считают его бесполезным [Pontius, Millones, 2011]. Мы рассчитали индекс Каппа, как и все остальные перечисленные выше, но выбор финальной модели сделали по AUC и положению на графике Тэйлора [Taylor, 2001].

Тестирование MaxEnt, GAM и RF проведено на 20 % случайно отобранных наблюдений, исключенных из обучающей выборки, с сохранением баланса преобладания «нулей», как в исходном наборе (80 %), а также с отношением «нулей» к «единицам» как 2 : 1 и 1 : 1.

В отличие от MaxEnt и RF, в GAM можно явно задавать и оценивать виды сглаживающих функций и их сложность. Мы использовали тонкопленочные сплайны [Wood, 2003] с ограничением числа псевдоузлов до 4 для одномерных предикторов. Тензорное произведение было ограничено в размерности оснований до 5 по каждой из трех осей, как это рекомендуется для использованной функции [Wood, 2017]. Тензорное произведение использовано в моделировании общей схемы миграции по координатам и порядковому номеру дня в году в пакете *mgcv* [<https://cran.r-project.org/web/packages/mgcv>] для языка R. Выбор финальной GAM осуществлен по критерию Шварца-Байеса — BIC [Sakamoto et al., 1986], более строгому, чем AIC, к числу параметров модели, а ее окончательная настройка произведена методом ограниченного максимального правдоподобия, как это рекомендовано разработчиком *mgcv* [Wood, 2011].

Настройка RF произведена с параметрами по умолчанию из пакета для его многопоточной реализации для языка R — *ranger* [Wright, Ziegler, 2017], но мы увеличили число деревьев с 500 до 1000. Оптимальное число деревьев рекомендуется разработчиком искать в пределах от 500 до 1000, и может оказаться, что 600 или 700 деревьев работают в итоге так же, как и 1000. Однако такая оптимизация займет неоправданно большое количество времени, а выигрыш в использовании RF с 600 деревьями вместо 1000 составит всего несколько секунд, поэтому мы решили использовать максимальное число деревьев из разумных пределов.

Результаты и их обсуждение

Более 60 % вклада в определение HSI, а также важности на перестановке — PI (permutation importance, которая оценивает относительное снижение AUC при случайном перемешивании значений предиктора) во всех испытаниях MaxEnt принадлежало переменным координат и порядковому номеру дня в году и их квадратам (для учета нелинейности). Следовательно, в рассматриваемом районе направление миграций достаточно часто повторяется, поэтому далее мы рассмотрим только факторы среды, которые и вызывали отклонения от общей по дням в году схемы миграции, которая обычно проходит против часовой стрелки.

В первом обучении MaxEnt (на 80 % случайно отобранных данных из первого набора) более 0,1 % вклада в определение HSI или PI получили 26 параметров среды (табл. 1) при AUC = 0,654. Последний в первом тесте (относительно оставшихся 20 % данных, не входивших в настройку MaxEnt) оказался немного выше, чем в обучении: AUC = 0,683.

Скорости течений попали в табл. 1, но только в северном направлении и с очень низким вкладом. Обучение на втором наборе данных, где вместо скоростей течений включен ЕКЕ, показало, что ЕКЕ также имеет низкий вклад в нахождение HSI (табл. 2). Во втором обучении AUC в тесте снова оказался выше (0,683), чем в обучающем наборе (0,652). В обоих случаях общая точность достигла 0,55, а индекс Каппа — 0,16.

В прежних SDM с помесечным разрешением вклад ЕКЕ в августе доходил до 28,1 %, снижаясь до 6,2 % в декабре [Syah et al., 2016]. Здесь SST получила максимальный вклад, но он тоже в разы ниже относительно прежних работ с помесечным разрешением. В таком сильном различии, вероятно, виновата высокая доля реальных «нулей» (*prevalence* = 0,2), поэтому в третьем обучении мы случайным образом снизили соотношение «нулей» и «единиц» с 4 : 1 до 2 : 1 и получили более низкий AUC

Таблица 1

Факторы среды, имевшие вклад в HSI или PI более 0,1 %, в первом обучении в MaxEnt

Table 1

Top environmental factors with contribution to the habitat suitability index (HSI) or permutation importance > 0.1 % in the first training set of MaxEnt

Фактор среды	Вклад в HSI, %	PI, %	Сдвиг дней назад в начале окна	Сдвиг дней назад в конце окна	Глубина, м	Метка
SST	8,9	3,7	0	2	0	Lag.02.MeanSST
SST	6,5	1,5	1	2	0	Lag.12.MeanSST
SST	3,2	4,7	0	0	0	Lag.00.MeanSST
SST	1,9	2,9	0	1	0	Lag.01.MeanSST
T	1,7	0,0	0	3	50	Lag.03.MeanT050
TG	1,2	0,0	2	3	50	Lag.23.MeanG050
TG	0,8	0,9	0	0	50	Lag.00.MeanG050
T	0,5	2,6	3	7	50	Lag.37.MeanT050
SSTG	0,2	0,0	3	4	0	Lag.34.MeanGRD
SST	0,2	1,1	0	7	0	Lag.07.MeanSST
T	0,2	0,6	3	5	100	Lag.35.MeanT100
SST	0,1	0,0	3	6	0	Lag.36.MeanSST
TG	0,1	0,0	2	4	50	Lag.24.MeanG050
T	0,1	0,8	0	7	50	Lag.07.MeanT050
v	0,1	0,2	0	1	0	Lag.01.MeanNcm
T	0,1	0,6	1	4	50	Lag.14.MeanT050
T	0,1	0,5	3	7	100	Lag.37.MeanT100
v	0,1	0,4	3	7	0	Lag.37.MeanNcm
TG	0,1	0,1	1	2	50	Lag.12.MeanG050
TG	0,1	0,5	3	7	50	Lag.37.MeanG050
SST	0,1	0,3	3	7	0	Lag.37.MeanSST
v	0,0	0,4	0	7	0	Lag.07.MeanNcm
TG	0,0	0,3	0	7	50	Lag.07.MeanG050
T	0,0	0,6	0	0	50	Lag.00.MeanT050
TG	0,0	0,3	0	0	0	Lag.00.MeanGRD
SST	0,0	0,2	2	3	0	Lag.23.MeanSST

как в обучающем наборе (0,625), так и в тесте (0,676) по сравнению с предыдущими испытаниями. Однако вклад в HSI и PI резко вырос у SST (табл. 3).

Во всех трех тестах (на данных, не входивших в настройку) AUC был выше, чем в обучении, следовательно, протестированные модели MaxEnt обладают хорошей обобщающей способностью и можно провести настройку на полном наборе. Наибольшее изменение вклада факторов среды в HSI и PI произошло при изменении соотношения «нулей» в обучающем наборе, поэтому мы исследовали влияние доли «нулей» на AUC и вклад в HSI и PI на полном наборе, а также в наборе с урезанием доли «нулей» до 66,6 и 50,0 %, но зафиксировали число факторов среды (18) из тех, что получали высокий вклад в предыдущих тестах (табл. 4).

В наших наборах данных оказалось, что при снижении доли «нулей» понижается не только AUC, но и вклад факторов среды в HSI и PI, если изначально выбрать самые показательные из них. Максимальный вклад EKE в HSI в полном наборе снова был низким (0,6 % при средней за неделю, предшествовавшей улову). При большом числе предикторов часть вклада в HSI самых важных на перестановке (PI) теряется, поэтому мы решили ограничить выбор потенциальных предикторов минимальным вкладом в HSI или PI в прежних испытаниях в 1 % (табл. 5).

Таким образом, получена финальная оценка выбранных предикторов в MaxEnt, который на полном наборе достиг AUC = 0,647. Графически найденные зависимости показаны на рис. 1.

Таблица 2

Факторы среды, имевшие вклад в HSI или PI более 0,1 %, во втором обучении в MaxEnt с EKE

Table 2

Top environmental factors (including the current velocity EKE) with contribution to the habitat suitability index (HSI) or permutation importance > 0.1 % in the second training of MaxEnt

Фактор среды	Вклад в HSI, %	PI, %	Сдвиг дней назад в начале окна	Сдвиг дней назад в конце окна	Глубина, м	Метка
SST	2,4	3,7	0	0	0	Lag.00.MeanSST
T	1,3	0,0	0	3	50	Lag.03.MeanT050
TG	1,3	0,0	0	4	50	Lag.04.MeanG050
TG	0,8	1,2	0	0	50	Lag.00.MeanG050
EKE	0,7	1,5	0	7	0	Lag.07.MeanEKE
T	0,6	4,2	3	7	50	Lag.37.MeanT050
T	0,6	0,6	0	6	50	Lag.06.MeanT050
EKE	0,3	0,7	0	0	0	Lag.00.MeanEKE
SST	0,3	1,4	0	7	0	Lag.07.MeanSST
SSTG	0,2	0,0	3	5	0	Lag.35.MeanGRD
SST	0,2	0,0	1	4	0	Lag.14.MeanSST
SSTG	0,2	0,0	1	5	50	Lag.15.MeanG050
EKE	0,1	0,8	2	3	0	Lag.23.MeanEKE
TG	0,1	0,4	0	7	50	Lag.07.MeanG050
T	0,1	1,0	3	5	100	Lag.35.MeanT100
TG	0,1	0,6	3	7	50	Lag.37.MeanG050
SST	0,1	0,4	3	7	0	Lag.37.MeanSST
T	0,0	0,5	3	7	100	Lag.37.MeanT100
EKE	0,0	0,7	1	2	0	Lag.12.MeanEKE
T	0,0	0,4	0	7	50	Lag.07.MeanT050
T	0,0	0,1	0	1	50	Lag.01.MeanT050
EKE	0,0	0,1	3	4	0	Lag.34.MeanEKE
TG	0,0	0,1	0	0	0	Lag.00.MeanGRD
T	0,0	0,7	0	0	50	Lag.00.MeanT050
SST	0,0	0,1	0	1	0	Lag.01.MeanSST
T	0,0	0,1	1	3	100	Lag.13.MeanT100

Далее мы используем эти предикторы, а также их более корректное отражение в 2D и 3D (вместо индексированных по дням независимых координат: jLon, jLon2 и jLat2) в GAM и RF.

Испытаны как аналогичная MaxEnt по структуре из одномерных предикторов биномиальная GAM (формула (12)) с логистической функцией связи, так и двумерное тензорное произведение — te (формула (13)) и трехмерное (формула (14)). Эти GAM различались способом учета общей схемы миграции сайры:

$$\begin{aligned}
 Catch \approx & f(\text{Lag.37.MeanT050}) + f(\text{Lag.37.MeanG050}) + f(\text{Lag.02.MeanSST}) + \\
 & + f(\text{Lag.00.MeanSST}) + f(\text{Lag.00.MeanT050}) + f(\text{Lag.00.MeanG050}) + \\
 & + f(j\text{Lon}) + f(j\text{Lon2}) + f(j\text{Lat2}) + f(j\text{Day2}),
 \end{aligned}
 \tag{12}$$

$$\begin{aligned}
 Catch \approx & f(\text{Lag.37.MeanT050}) + f(\text{Lag.37.MeanG050}) + f(\text{Lag.02.MeanSST}) + \\
 & + f(\text{Lag.00.MeanSST}) + f(\text{Lag.00.MeanT050}) + f(\text{Lag.00.MeanG050}) + \\
 & + te(\text{Lon}, j\text{Day}) + te(j\text{Day}, \text{Lat}),
 \end{aligned}
 \tag{13}$$

$$\begin{aligned}
 Catch \approx & f(\text{Lag.37.MeanT050}) + f(\text{Lag.37.MeanG050}) + f(\text{Lag.02.MeanSST}) + \\
 & + f(\text{Lag.00.MeanSST}) + f(\text{Lag.00.MeanT050}) + f(\text{Lag.00.MeanG050}) + \\
 & + te(\text{Lon}, \text{Lat}, j\text{Day}),
 \end{aligned}
 \tag{14}$$

где переменная ответа *Catch* принимала значения 1 в позициях на лову и 0 в позициях поиска сайры, а *f* — тонкоплечные сплайны с автоматическим перебором числа псевдоузлов до четырех методом обобщенной кроссвалидации в пакете *mgcv*.

Таблица 3

Факторы среды, имевшие вклад в HSI или PI более 0,1 %, в третьем обучении в MaxEnt

Table 3

Top environmental factors with contribution to the habitat suitability index (HSI) or permutation importance > 0.1 % in the third training set of MaxEnt

Фактор среды	Вклад в HSI, %	PI, %	Сдвиг дней назад в начале окна	Сдвиг дней назад в конце окна	Глубина, м	Метка
SST	17,0	12,6	0	2	0	Lag.02.MeanSST
SST	2,5	4,5	0	0	0	Lag.00.MeanSST
TG	1,2	0,0	0	4	50	Lag.04.MeanG050
T	0,8	0,3	0	3	50	Lag.03.MeanT050
T	0,6	4,2	3	7	50	Lag.37.MeanT050
TG	0,4	0,0	1	4	100	Lag.14.MeanG100
SSTG	0,2	0,5	0	0	0	Lag.00.MeanGRD
<i>u</i>	0,2	1,0	3	7	0	Lag.37.MeanEcm
SST	0,1	0,9	0	7	0	Lag.07.MeanSST
TG	0,1	1,5	0	0	50	Lag.00.MeanG050
SSTG	0,1	0,1	3	4	0	Lag.34.MeanGRD
<i>v</i>	0,1	0,2	0	1	0	Lag.01.MeanNcm
TG	0,1	0,0	1	5	50	Lag.15.MeanG050
TG	0,1	1,4	3	7	50	Lag.37.MeanG050
T	0,1	0,3	3	7	100	Lag.37.MeanT100
SST	0,1	0,0	3	5	0	Lag.35.MeanSST
<i>v</i>	0,1	0,4	3	7	0	Lag.37.MeanNcm
TG	0,1	0,0	1	2	50	Lag.12.MeanG050
T	0,0	1,9	0	0	50	Lag.00.MeanT050

Таблица 4

AUC, вклад в HSI и PI в MaxEnt в зависимости от доли «нулей» в обучении, %

Table 4

MaxEnt estimations of the area under the receiver operating characteristic curve (AUC), contribution to the habitat suitability index (HSI) and permutation importance (PI) in dependence on the portion of real sites with zero catches, %

AUC	0,650		0,624		0,591	
Число единиц	42224		42224		42224	
Всего наблюдений	216 489		126 672		84 448	
Доля «нулей», %	80,5		66,6		50,0	
Метка	Вклад в HSI	PI	Вклад в HSI	PI	Вклад в HSI	PI
Lag.02.MeanSST	17,0	11,2	16,8	12,0	14,6	13,7
Lag.00.MeanSST	3,0	4,4	2,3	4,9	1,7	6,2
Lag.04.MeanG050	1,4	0,0	0,0	0,0	0,9	0,0
Lag.03.MeanT050	1,3	0,0	1,0	0,5	1,0	1,3
Lag.00.MeanG050	1,1	2,1	1,4	2,5	0,5	2,2
Lag.37.MeanT050	0,7	5,3	0,7	4,4	0,5	6,1
Lag.00.MeanGRD	0,2	0,1	0,1	0,1	0,0	0,0
Lag.34.MeanGRD	0,2	0,1	0,2	0,1	0,1	0,2
Lag.37.MeanT100	0,2	0,6	0,0	0,3	0,0	0,1
Lag.00.MeanT050	0,1	2,2	0,0	0,9	0,0	0,2
Lag.07.MeanSST	0,1	0,7	0,1	0,7	0,0	0,2
Lag.12.MeanG050	0,1	0,0	0,0	0,0	0,0	0,0
Lag.15.MeanG050	0,1	0,0	0,0	0,0	0,1	0,0
Lag.37.MeanG050	0,1	1,4	0,3	1,1	0,1	1,9
Lag.37.MeanNcm	0,1	0,5	0,1	0,4	0,0	0,0
Lag.01.MeanNcm	0,0	0,1	0,0	0,1	0,0	0,0
Lag.14.MeanG100	0,0	0,0	0,5	0,0	0,0	0,0
Lag.37.MeanEcm	0,0	0,7	0,2	0,9	0,1	0,9

Таблица 5

Предикторы, оставшиеся в MaxEnt на полном наборе при последовательной их фильтрации по вкладу более 1 % в HSI или PI

Table 5

Predictors chosen by MaxEnt method from the full set of environmental factors after consecutive filtering by their contribution to the habitat suitability index or permutation importance > 1 %

Предиктор	Вклад в HSI, %	PI, %	Сдвиг дней назад в начале окна	Сдвиг дней назад в конце окна	Глубина, м	Метка
Долгота × 10 × День ^{2*}	64,8	31,2	0	0	—	jLon2
SST, К	16,5	14,7	0	2	0	Lag.02.MeanSST
День ²	6,0	21,9	0	0	—	jDay2
Долгота × 10 × День	4,1	11,4	0	0	—	jLon
SST, К	3,1	6,6	0	0	0	Lag.00.MeanSST
TG · 100, °С на 0,01 геоградуса ²	1,7	2,4	0	0	50	Lag.00.MeanG050
T · 100, °С	1,4	6,9	3	7	50	Lag.37.MeanT050
T · 100, °С	1,2	1,7	0	0	50	Lag.00.MeanT050
TG · 100, °С на 0,01 геоградуса ²	0,9	1,4	3	7	50	Lag.37.MeanG050
Широта × 10 × День ²	0,3	1,8	0	0	—	jLat2

* День (jDay) каждый год начинается с 1 и отображает только его порядковый номер в году, таким образом, здесь он лишь индексирует параболическую часть по долготе вне зависимости от года. Во избежание больших чисел в подписях осей на рисунках из jDay вычтено 100 дней, а jLon2 и jLat2 поделены на 100 000.

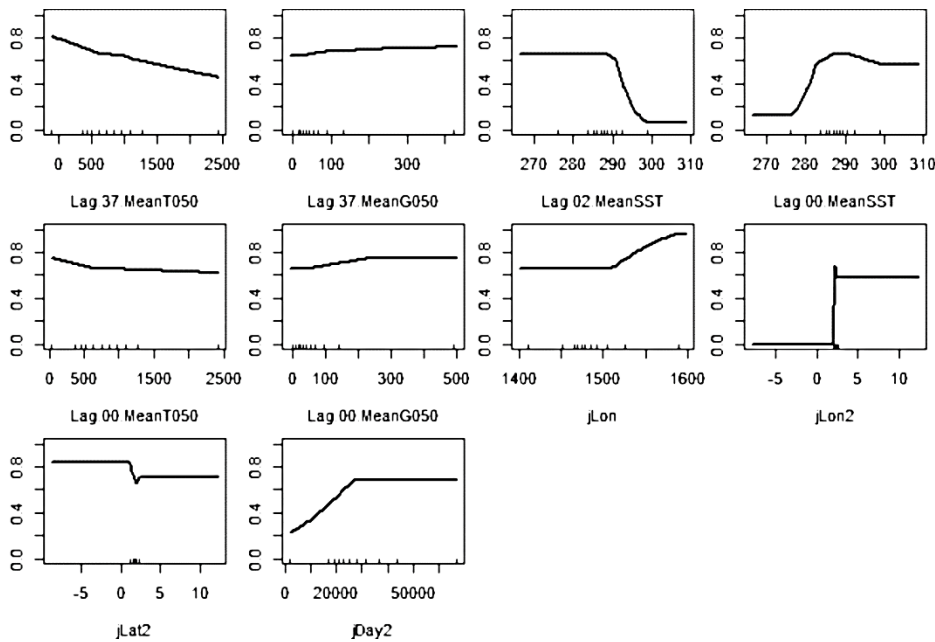


Рис. 1. Зависимости, определяющие вероятность улова сайры (по оси ординат) от различных предикторов (по оси абсцисс), в финальной модели MaxEnt

Fig. 1. Probability of pacific saury catch (y) in dependence on the predictors (x) finally chosen by MaxEnt method

Победила GAM по формуле (14) (3D GAM), обладавшая наименьшим BIC (193 775). GAM по формуле (13) (2D GAM) заняла второе место по BIC (195 172), а по формуле (12) — последнее (BIC = 197 286), несмотря на более чем двукратное увеличение числа параметров на каждом усложнении GAM по формулам (12–14) (соответственно 25, 53 и 116 параметров).

Тестирование показало, что при более низком AUC (0,602) общая точность 3D GAM оказалась выше (0,73) и индекс Каппа тоже выше (0,21), чем в финальной MaxEnt. При этом максимального значения индекс Каппа достигал при вероятности улова более 0,29, что в два раза ниже, чем в нашей финальной MaxEnt и ранних публикациях [Syah et al., 2014, 2016], где разделения улова и его отсутствия находилось на уровне вероятности 0,6.

В полном наборе данных 3D GAM получила более высокий AUC = 0,71 и общую точность 0,75, но максимальное значение индекса Каппа снова находилось при вероятности улова более 0,29, а при повышении вероятности разделения улова и его отсутствия до 0,50 понижается и AUC — до 0,34. Найденные в финальной 3D GAM зависимости, за исключением самой 3D части тензорного произведения координат и номера дня в году, показаны на рис. 2.

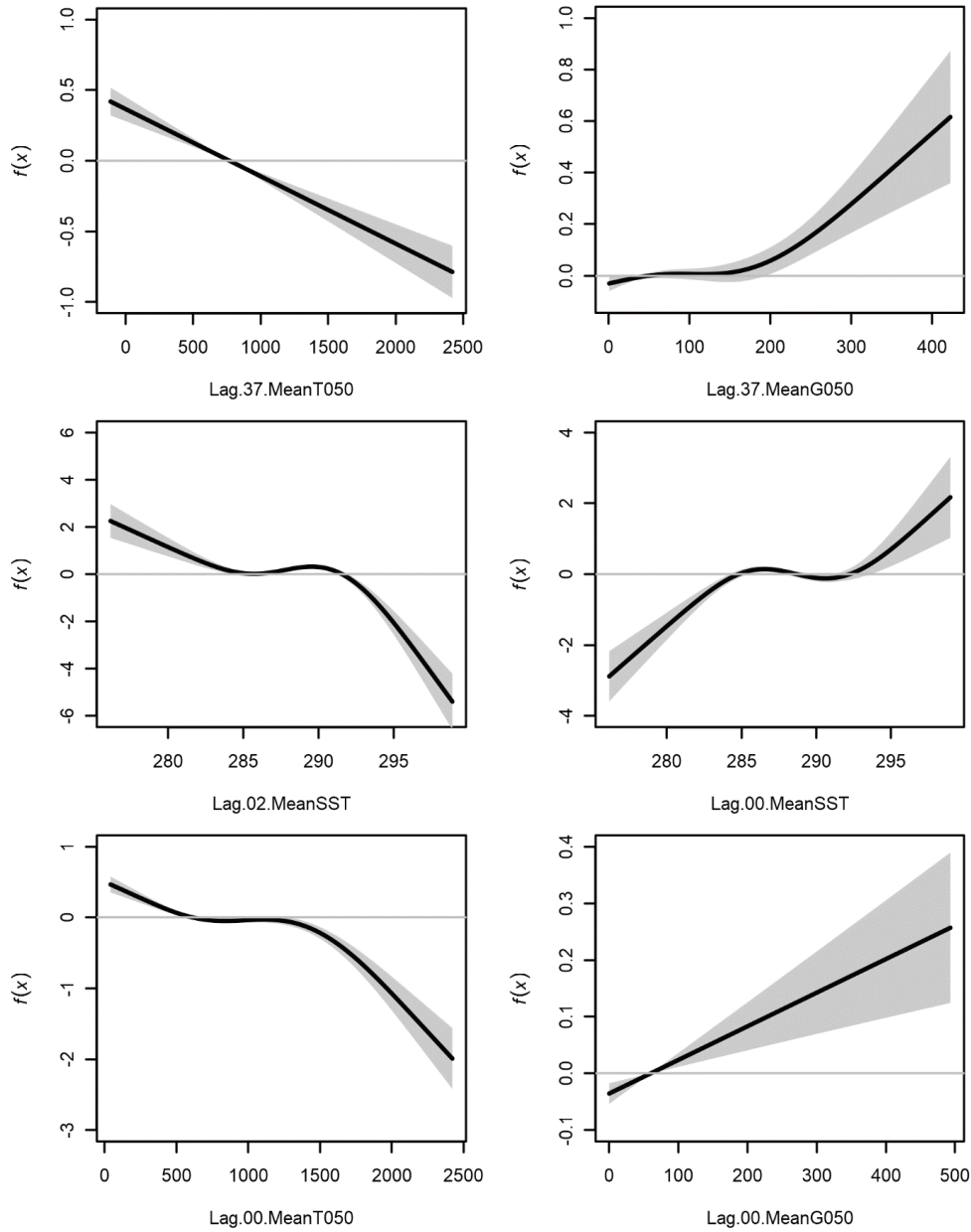


Рис. 2. Слагаемые (в лог-масштабе) члены финальной GAM. Серой заливкой показан интервал стандартной ошибки

Fig. 2. Additive terms in the final GAM (logarithmic scale; standard error interval is shaded)

Настройка RF проведена с использованием идентичного обучающего набора данных, как для финальной модели MaxEnt, тестирование показало более высокий AUC (0,696) и общую точность 0,861. При замене индексированных координат реальными их значениями отдельно от порядкового номера дня в году происходит снижение AUC до 0,688 и общей точности до 0,856. Индекс Каппа занял аналогичные места (0,47 против 0,45). В полном наборе данных расклад не изменился: индексированные координаты способствуют более высоким значениям AUC (0,713), индекса Каппа (0,50) и общей точности (0,866), чем реальные координаты (AUC = 0,703, κ = 0,478 и *accuracy* = 0,861).

Таким образом, RF превзошел по AUC, κ и общей точности 3D GAM и MaxEnt. Этот вывод подтверждается большинством остальных упомянутых в методике показателей (рис. 3), а также графиком Тэйлора (рис. 4).

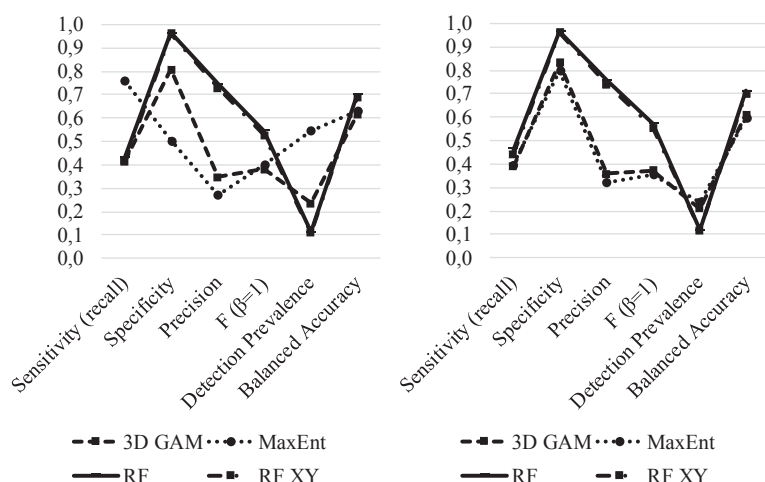


Рис. 3. Показатели матрицы неточностей в тестах (**слева**) и на полном наборе (**справа**), RF XY показывает RF, настроенный с истинными координатами, а не индексированными по дням в году

Fig. 3. Confusion matrix indices for the test set (**left**) and full set (**right**) of predictors. RF XY — results for the Random Forest technique with real coordinates instead of date-related ones

Как в тестах, так и в полном наборе RF предсказывает практически в два раза меньше мест улова (Detection Prevalence), чем MaxEnt и 3D GAM, но при этом они часто дают ложноположительные прогнозы, поэтому их положительная прогностическая ценность (precision) существенно ниже. Небольшое отставание RF по AUC с обычными координатами от индексированных по дням в году может объясняться, например, случайным совпадением зигзагообразного движения судов в поиске.

В целом превосходство ансамблевых методов типа RF широко известно аналитикам, работающим с большими данными, не только в биологии, где точность предсказания протеиновых функций в RF достигает уже 100 % [Lee et al., 2009], но и в других областях, например в распознавании образов [Shotton et al., 2013].

В данной работе мы столкнулись с тем, что в тестах оценки прогностического качества моделей оказывались выше, чем в обучающих выборках. В условиях большого количества наблюдений это не должно удивлять. Известны и более неожиданные выводы при использовании методов машинного обучения на больших данных, когда модели с глубоким обучением оказываются способны не только быстрее аппроксимировать физические решения в космологии, но и выдавать достаточно точные экстраполяции за пределы обучающих наборов [He et al., 2019].

Следует отметить, что существуют и другие подходы к моделированию распределения видов, например биофизические агентные (или супер-индивидуальные) модели (БАМ). Среди них наиболее известная разработка для сайры — NEMURO.FISH [Werner et al., 2007], но во избежание излишней параметризации такого рода моделей приходится огрублять пространственно-временной масштаб до величин, которые не

Рис. 4. Диаграмма Тэйлора по тестовым наборам (вверху) и полным (внизу)

Fig. 4. Taylor diagram for the test sets (top panel) and full sets (bottom panel) of predictors

позволили пока ее использовать, и предпочтение все еще отдается статистическим моделям и методам машинного обучения. Одним из препятствий на пути внедрения БАМ является то, что они должны разрабатываться сразу для нескольких видов на разных трофических уровнях, вовлеченных в трофические цепи целевого вида. В таких условиях наблюдений оказывается значительно меньше параметров, что ведет к отсутствию единого решения и прогноза.

В заключение мы решили проверить, насколько тесно связаны уловы и суммарные площади за год тех мест, где RF предсказывал нахождение сайры (вероятные места скоплений). Оказалось, что такая связь есть и она очень сильно ($r = 0,96$) выражена в годы экстремальных уловов — 2008, 2014, 2017 и 2018 (рис. 5).

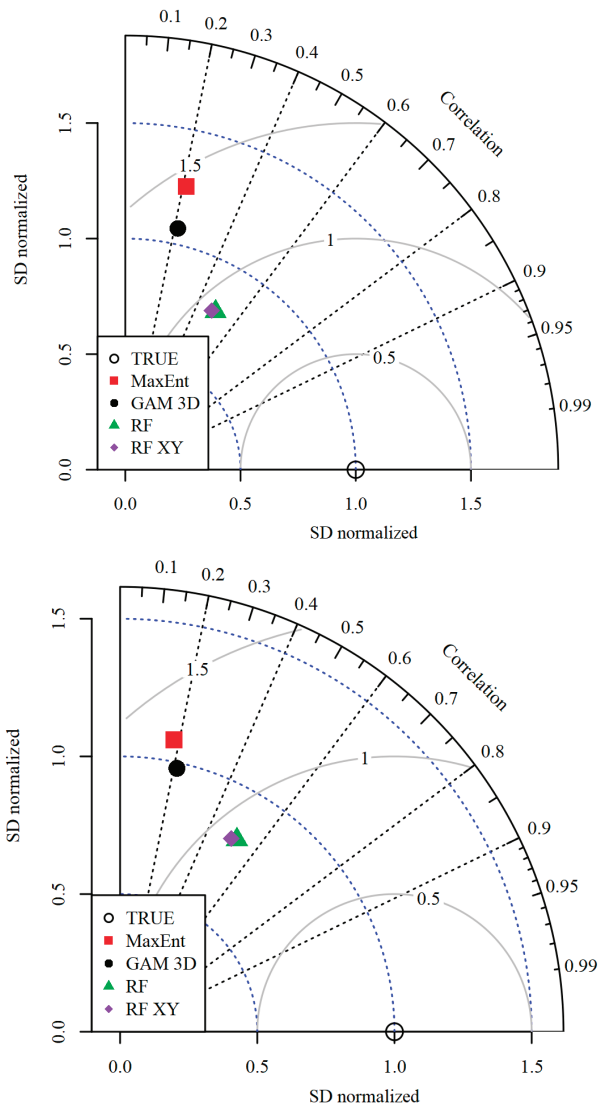
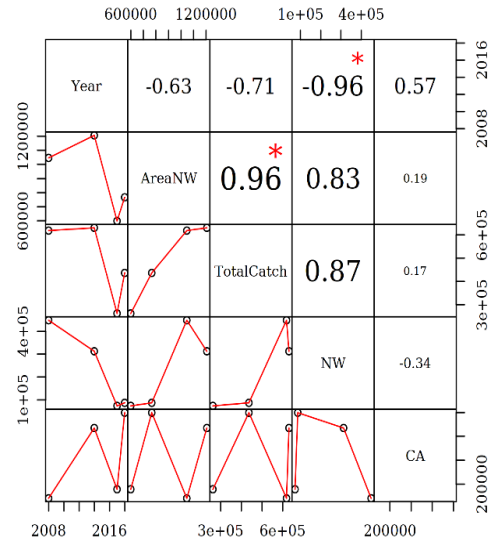


Рис. 5. Корреляция Пирсона между годами (Year), суммарной площадью за год тех мест, где RF предсказывал нахождение сайры в ИЭЗ (AreaNW), общим выловом всеми участниками NPFC (TotalCatch), выловом в ИЭЗ (NW) и конвенционном районе NPFC (CA) в 2008 г., когда был достигнут максимальный вылов в ИЭЗ; в 2014 г., когда был достигнут максимальный общий вылов; в 2017, когда вылов в ИЭЗ был минимальным, и в 2018 г., когда вылов в CA был максимальным. Звездочка показывает значимую корреляцию ($p < 0,05$)

Fig. 5. Pearson product-moment correlation between the years (Year), total area with predicted conditions favorable for the saury fishery within the EEZ of Russia and Japan (AreaNW), summary annual catch of saury by all NPFC countries (TotalCatch), annual catch in the EEZ of Russia and Japan (NW) and in the convention area of NPFC (CA) in 2008 (maximum NW), 2014 (maximum total catch), 2017 (minimum NW), and 2018 (maximum CA). The case with significant correlation ($p < 0.05$) is shown by asterisk



Однако по всем годам такой связи не обнаружено. Это может свидетельствовать о том, что на общий вылов гидрологическая ситуация влияет только в экстремальные годы, а в последнее время площадь за год, где RF предсказывал наличие сайры в ИЭЗ, была минимальной, что совпало с минимальными уловами.

Выводы

Из опробованных нами методов машинного обучения и GAM наилучшие результаты показал RF, общая точность которого достигла 0,86 при AUC, равном 0,7, с преимуществом во втором и третьем знаке после десятичного разделителя у RF с индексированными координатами. Стоит отметить, что это преимущество может являться случайным совпадением индекса по дням в году и долготе с зигзагообразным движением судов.

Во всех случаях в набор предикторов вошли значения параметров среды не только с задержкой в 3–7 дней, что позволяет использовать их для КППО, но и без задержки (SST, температура на горизонте 50 м и ее градиент), а также средняя SST за последние 3 дня, включая текущий. При этом вклад в определение HSI последних был самым высоким из факторов среды. Следовательно, чтобы появилась возможность в полной мере использовать настроенный RF на практике, например давать рекомендации флоту о поиске сайры в конкретных местах за 2–3 дня, нужны прогнозы SST, температуры воды и ее градиента на горизонте 50 м на 3–4 дня вперед. В лаборатории нелинейных динамических систем Тихоокеанского океанологического института им. В.И. Ильичева ДВО РАН ведется разработка моделей разномасштабных когерентных структур в дальневосточных морях России и в северо-западной части Тихого океана для оценки их значения для биопродуктивности и рыбного промысла. Эти модели смогут поставлять в высоком разрешении не только информацию о градиентах, но и о происхождении вод. Такая информация, конечно, будет в других единицах измерения (показатели Ляпунова) и типах данных (категориальные), но она сможет быть рассчитана вперед на несколько дней. В итоге потребуются повторить настройку RF уже с новыми данными.

Резюмируя, отметим, что в данной работе удалось показать возможность использования методов машинного обучения для КППО, но, как и в аналогичных работах месячного масштаба, прогнозы с использованием общедоступных существующих наборов данных будут отставать, только в отличие от аналогичных работ не на 1 мес, а на 2–3 дня.

Благодарности

Выражаем благодарность И.И. Шевченко и Е.А. Мирзязянову за выделение ресурсов в центре хранения и обработки данных ТИНРО, а также их настройку и поддержку в режиме онлайн и удалённого доступа, что позволило обработать терабайты исходных данных.

Финансирование работы

Работа В.В. Кулика по выполнению численных расчетов была поддержана Российским научным фондом (проект № 19-17-00006). Использование ресурсов в Центре хранения и обработки данных ТИНРО осуществлено в рамках государственного задания № 076-00005-19-00 ФГБНУ «ВНИРО» на 2019 г.

Соблюдение этических стандартов

Настоящая статья не содержит каких-либо исследований с использованием животных в качестве объектов. Все исходные данные получены в результате дистанционного зондирования Земли, мониторинга движения судов и последующего слияния информации судовых наблюдений, доступных по программе NEAR-GOOS и в отраслевой системе мониторинга информационной системы Росрыболовства.

Авторы заявляют, что у них нет конфликта интересов.

Информация о вкладе авторов

В.В. Кулик и А.А. Байталюк разработали постановку проблемы и вероятные пути решения, О.Н. Катугин и Е.И. Устинова участвовали в предварительном отборе исходных данных. В.В. Кулик проводил теоретические расчеты и обработку данных. Все авторы участвовали в написании текста статьи и в обсуждении результатов.

Список литературы

Байталюк А.А., Давыдова С.В. Распределение и пассивные миграции сайры в северной части Тихого океана // *Вопр. рыб-ва.* — 2002. — Т. 3, № 3(11). — С. 402–420.

Бочаров Л.Н. Информационная технология краткосрочного промыслового прогнозирования : дис. ... д-ра техн. наук в форме научного доклада. — Владивосток : ИАПУ ДВО РАН, 1993. — 68 с.

Истоки Ойясио : моногр. / под ред. В.Р.Фукса, А.Н. Мичурина. — СПб. : СПбГУ, 1997. — 248 с.

Кулик В.В. Использование температуры поверхности океана при оценке запаса тихоокеанской сайры (*Cololabis saira*) в Комиссии по рыболовству в северной части Тихого океана // *Процессы в геосредах.* — 2018. — Т. 3, № 17. — С. 76–77.

Кулик В.В., Катугин О.Н., Зуев М.А. Вероятное сокращение ареала северного кальмара (*Boreoteuthis borealis*) в северо-западной части Тихого океана при различных сценариях климатических изменений // *Комплексные исследования Мирового океана : мат-лы 2-й всерос. науч. конф. мол. ученых [Электронный ресурс]*. — М. : ИО РАН, 2017. — С. 362–363.

Парин Н.В. Ареал сайры (*Cololabis saira* Brev. — Scombresocidae, Pisces) и значение океанографических факторов для ее распространения // *ДАН СССР.* — 1960. — Т. 130, № 3. — С. 649–652.

Филатов В.Н. Миграции и формирование скоплений массовых пелагических гидробионтов (на примере тихоокеанской сайры) : моногр. — Ростов н/Д : ЮНЦ РАН, 2015. — 168 с.

Филатов В.Н. Определение перспективных для промысла сайры участков по данным полигонных съемок в Южно-Курильском районе // *Изв. ТИНРО.* — 1984. — Т. 109. — С. 35–40.

Филатов В.Н., Старцев А.В., Устинова Е.И., Еремин Ю.В. Тихоокеанская сайра. Научно-информационное обеспечение промысловой экспедиции : моногр. — Ростов н/Д : ЮНЦ РАН, 2011. — 120 с.

Baitaliuk A.A., Orlov A.M., Ermakov Yu.K. Characteristic features of ecology of the Pacific saury *Cololabis saira* (Scomberesocidae, Beloniformes) in open waters and in the northeast Pacific ocean // *J. Ichthyol.* — 2013. — Vol. 53, № 11. — P. 899–913. DOI: 10.1134/S0032945213110027.

Belkin I.M., O'Reilly J.E. An algorithm for oceanic front detection in chlorophyll and SST satellite imagery // *J. Mar. Syst.* — 2009. — Vol. 78, № 3. — P. 319–326. DOI: 10.1016/j.jmarsys.2008.11.018.

Breiman L. Random forests // *Mach. Learn.* — 2001. — Vol. 45, Iss. 1. — P. 5–32. DOI: 10.1023/A:1010933404324.

Chang Y.-J., Lan K.-W., Walsh W.A. et al. Modelling the impacts of environmental variation on habitat suitability for Pacific saury in the Northwestern Pacific Ocean // *Fish. Oceanogr.* — 2019. — Vol. 28, № 3. — P. 291–304. DOI: 10.1111/fog.12408.

Chin T.M., Vazquez-Cuervo J., Armstrong E.M. A multi-scale high-resolution analysis of global sea surface temperature // *Remote Sens. Environ.* — 2017. — Vol. 200, № 7. — P. 154–169. DOI: 10.1016/j.rse.2017.07.029.

Cohen J. A Coefficient of Agreement for Nominal Scales // *Educ. Psychol. Meas.* — 1960. — Vol. 20, № 1. — P. 37–46. DOI: 10.1177/001316446002000104.

Guillera-Arroita G., Lahoz-Monfort J.J., Elith J. Maxent is not a presence-absence method: a comment on Thibaud et al. // *Methods Ecol. Evol.* — 2014. — Vol. 5, № 11. — P. 1192–1197. DOI: 10.1111/2041-210X.12252.

He S., Li Y., Feng Y. et al. Learning to predict the cosmological structure formation // *Proc. Nat. Acad. Sci.* — 2019. — Vol. 116, № 28. — P. 13825–13832. DOI: 10.1073/pnas.1821458116.

Kawai H. Hydrography of the Kuroshio and the Oyashio (in Japanese) // *Physical Oceanography II, Fundamental Lectures of Oceanography.* — Japan : Tokai Univ. Press, 1972. — P. 129–320.

Kawai H. Statistical estimation of isotherms indicative of the Kuroshio axis // *Deep-Sea Res.* — 1969. — Vol. 16. — P. 109–115.

Kobayashi T., Wako M., Naito M. Studies on the life of the Pacific saury, *Cololabis saira* (Brevoort). I. Aggregative characteristics of adult of the autumn-spawning population // *Sci. Rep. Hokkaido Fish. Exp. Stn.* — 1968. — Vol. 9. — P. 1–45.

Kulik V.V., Katugin O.N., Zuev M.A. Climate change impacts on distribution patterns of boreopacific gonate squid (*Boreoteuthis borealis*) in the Northwest Pacific // Proceedings of the 25 Year of PICES: Celebrating the Past, Imagining the Future. — San Diego, CA, USA : North Pacific Marine Science Organization, 2016. — P. 148.

Lee B., Shin M., Oh Y. et al. Identification of protein functions using a machine-learning approach based on sequence-derived properties // Proteome Sci. — 2009. — Vol. 7, № 1. — P. 27. DOI: 10.1186/1477-5956-7-27.

Li W., Guo Q. How to assess the prediction accuracy of species presence-absence models without absence data? // Ecography. — 2013. — Vol. 36, № 7. — P. 788–799. DOI: 10.1111/j.1600-0587.2013.07585.x.

Michalsky J.J. The Astronomical Almanac's algorithm for approximate solar position (1950–2050) // Solar Energy. — 1988. — Vol. 40, № 3. — P. 227–235. DOI: 10.1016/0038-092X(88)90045-X.

Ogawa Y. Variation of the southern end latitude of the first Oyashio intrusion (in Japanese with English abstract and captions) // Bull. Tohoku Reg. Fish. Lab. — 1989. — Vol. 51. — P. 1–9.

Olofsson P., Foody G.M., Herold M. et al. Good practices for estimating area and assessing accuracy of land change // Remote Sensing of Environment. — 2014. — Vol. 148. — P. 42–57. DOI: 10.1016/j.rse.2014.02.015.

Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions // Ecol. Modell. — 2006. — Vol. 190, № 3–4. — P. 231–259. DOI: 10.1016/j.ecolmodel.2005.03.026.

Pontius R.G., Millones M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment // Int. J. Remote Sens. — 2011. — Vol. 32, № 15. — P. 4407–4429. DOI: 10.1080/01431161.2011.552923.

Sakamoto Y., Ishiguro M., Kitagawa G. Akaike Information Criterion Statistics : Mathematics and its Applications. — Springer Netherlands, 1986. — 290 p.

Shotton J., Sharp T., Kipman A. et al. Real-time human pose recognition in parts from single depth images // Commun. ACM. — 2013. — Vol. 56, № 1. — P. 116–124. DOI: 10.1145/2398356.2398381.

Syah A.F., Saitoh S.-I., Alabia I.D., Hirawake T. Detection of potential fishing zone for Pacific saury (*Cololabis saira*) using generalized additive model and remotely sensed data // IOP Conf. Ser.: Earth Environ. Sci. — 2017. — Vol. 54. — P. 012074. DOI: 10.1088/1755-1315/54/1/012074.

Syah A.F., Saitoh S.-I., Alabia I., Hirawake T. Habitat model development of Pacific saury (*Cololabis saira*) using satellite remotely sensed data in the Northwestern North Pacific // 12th Biennial Conference of Pan Ocean Remote Sensing Conference (PORSEC 2014). — Bali, Indonesia, 2014. — P. 1–10.

Syah A.F., Saitoh S.-I., Alabia I.D., Hirawake T. Predicting potential fishing zones for Pacific saury (*Cololabis saira*) with maximum entropy models and remotely sensed data // Fish. Bull. — 2016. — Vol. 114, № 3. — P. 330–342. DOI: 10.7755/FB.114.3.6.

Takasugi S., Yasuda I. Index temperature at 100 m depth of the Oyashio front in the Iwate coastal region (in Japanese with English abstract and captions) // Japan J. Fish. Oceanogr. — 1993. — Vol. 57. — P. 333–344.

Taylor K.E. Summarizing multiple aspects of model performance in a single diagram // J. Geophys. Res. Atmos. — 2001. — Vol. 106, № D7. — P. 7183–7192. DOI: 10.1029/2000JD900719.

Uda M. Fishing centre of “Sanrna” (*Cololabis saira*, Brevoort) correlated with the head of Oyashio cold current (In Japanese with English summary) // Bull. Jap. Soc. Sci. Fish. — 1936. — Vol. 5. — P. 236–238.

Usui N., Ishizaki S., Fujii Y. et al. Meteorological Research Institute multivariate ocean variational estimation (MOVE) system: Some early results // Adv. Sp. Res. — 2006. — Vol. 37, № 4. — P. 806–822. DOI: 10.1016/j.asr.2005.09.022.

Usui N., Wakamatsu T., Tanaka Yu. et al. Four-dimensional variational ocean reanalysis: a 30-year high-resolution dataset in the western North Pacific (FORA-WNP30) // J. Oceanogr. — 2017. — Vol. 73, Iss. 2. — P. 205–233. DOI: 10.1007/s10872-016-0398-5.

Werner F.E., Ito S.-I., Megrey B.A., Kishi M.J. Synthesis of the NEMURO model studies and future directions of marine ecosystem modeling // Ecol. Modell. — 2007. — Vol. 202, № 1–2. — P. 211–223. DOI: 10.1016/j.ecolmodel.2006.08.019.

Wood S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models // J. R. Statist. Soc. B (Statistical Methodology). — 2011. — Vol. 73, № 1. — P. 3–36. DOI: 10.1111/j.1467-9868.2010.00749.x.

Wood S.N. Generalized Additive Models: An Introduction with R. — N.Y. : Chapman and Hall/CRC, 2017. 2nd ed. — 476 p.

- Wood S.N.** Thin plate regression splines // J. R. Statist. Soc. B (Statistical Methodology). — 2003. — Vol. 65, № 1. — P. 95–114. DOI: 10.1111/1467-9868.00374.
- Wright M.N., Ziegler A.** ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R // J. Stat. Softw. — 2017. — Vol. 77, № 1. — P. 1–17. DOI: 10.18637/jss.v077.i01.
- Yasuda I., Watanabe Y.** On the relationship between the Oyashio front and saury fishing grounds in the north western Pacific: A forecasting method for fishing ground locations // Fish. Oceanogr. — 1994. — Vol. 3, № 3. — P. 172–181. DOI: 10.1111/j.1365-2419.1994.tb00094.x.
- Yoshida T.** Climatological seasonal variations of the distribution of Oyashio cold water // Umi to Sora. — 1992. — Vol. 68, № 2. — P. 79–88 (in Japanese with English abstract).

References

- Baitaliuk, A.A. and Davydova, S.V.**, Distribution and passive migration of Pacific saury *Cololabis saira* Brevoort in the northern part of the Pacific Ocean, *Vopr. Rybolov.*, 2002, vol. 3, no. 3(11), pp. 402–420.
- Bocharov, L.N.**, Information technology of short-term field forecasting, *Dr. Tech. Sci. Dissertation in the form of a scientific report*, Vladivostok: IAPU DVO RAN, 1993.
- Istoki Oyyasio** (The origins of Oyyashio), Fuchs, V.R., Michurina, A.N., Ed., St. Petersburg: S.-Peterb. Gos. Univ., 1997.
- Kulik, V.V.**, The sea-surface temperature usage during stock assessment of pacific saury (*Cololabis saira*) in the North Pacific fisheries commission, *Protsessy v geosredakh*, 2018, vol. 3, no. 17, pp. 76–77.
- Kulik, V.V., Katugin, O.N., and Zuev, M.A.**, Likely reduction of the range of northern squid (*Boreoteuthis borealis*) in the northwestern part of the Pacific Ocean under various scenarios of climatic changes, in *Mater. 2 Vseross. nauchn. konf. molodykh uch.* “Kompleksnyye issledovaniya Mirovogo okeana” (Integrated Studies of the World Ocean) [Elektronnyy resurs], Moscow: Inst. Okeanol. RAN, 2017, pp. 362–363.
- Parin, N.V.**, The range of saury (*Cololabis saira* Brev. — Scomberesocidae, Pisces) and the importance of oceanographic factors for its distribution, *Dokl. Akad. Nauk SSSR*, 1960, vol. 130, no. 3, pp. 649–652.
- Filatov, V.N.**, *Migratsii i formirovaniye skopleniy massovykh pelagicheskikh gidrobiontov (na primere tikhookeanskoy sayry)* (Migrations and the formation of clusters of massive pelagic aquatic organisms (on the example of Pacific saury)), Rostov-on-Don: Yuzhn. Nauchn. Tsentr, Ross. Akad. Nauk, 2015.
- Filatov, V.N.**, Determination of the promising saury fishery regions based on the polygon surveys in the South-Kuril area, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 1984, vol. 109, pp. 35–40.
- Filatov, V.N., Startsev, A.V., Ustinova, E.I. and Eremin, Yu.V.**, *Tikhookeanskaya sayra. Nauchno-informatsionnoye obespecheniye promyslovoy ekspeditsii* (Pacific saury. Scientific and informational support of a fishing expedition), Rostov-on-Don: Yuzhn. Nauchn. Tsentr, Ross. Akad. Nauk, 2011.
- Baitaliuk, A.A., Orlov, A.M., and Ermakov, Yu.K.**, Characteristic features of ecology of the Pacific saury *Cololabis saira* (Scomberesocidae, Beloniformes) in open waters and in the northeast Pacific ocean, *J. Ichthyol.*, 2013, vol. 53, no. 11, pp. 899–913. doi 10.1134/S0032945213110027
- Belkin, I.M. and O'Reilly, J.E.**, An algorithm for oceanic front detection in chlorophyll and SST satellite imagery, *J. Mar. Syst.*, 2009, vol. 78, no. 3, pp. 319–326. doi 10.1016/j.jmarsys.2008.11.018
- Breiman, L.** Random forests, *Mach. Learn.*, 2001, vol. 45, no. 1, pp. 5–32. doi 10.1023/A:1010933404324
- Chang Y.-J., Lan K.-W., Walsh W.A., Hsu J., Hsieh C.-H.** Modelling the impacts of environmental variation on habitat suitability for Pacific saury in the Northwestern Pacific Ocean, *Fish. Oceanogr.*, 2019, vol. 28, no. 3, pp. 291–304. doi 10.1111/fog.12408
- Chin, T.M., Vazquez-Cuervo, J., and Armstrong, E.M.**, A multi-scale high-resolution analysis of global sea surface temperature, *Remote Sens. Environ.*, 2017, vol. 200, no. 7, pp. 154–169. doi 10.1016/j.rse.2017.07.029
- Cohen, J.**, A Coefficient of Agreement for Nominal Scales, *Educ. Psychol. Meas.*, 1960, vol. 20, no. 1, pp. 37–46. doi 10.1177/001316446002000104
- Guillera-Arroita, G., Lahoz-Monfort, J.J., and Elith, J.**, Maxent is not a presence-absence method: a comment on Thibaud et al., *Methods Ecol. Evol.*, 2014, vol. 5, no. 11, pp. 1192–1197. doi 10.1111/2041-210X.12252

- He, S., Li, Y., Feng, Y., Ho, S., Ravanbakhsh, S., Chen, W., and Póczos, B.,** Learning to predict the cosmological structure formation, *Proceedings of the National Academy of Sciences*, 2019, vol. 116, no. 28, pp. 13825–13832. doi 10.1073/pnas.1821458116
- Kawai, H.,** Statistical estimation of isotherms indicative of the Kuroshio axis, *Deep-Sea Res.*, 1969, vol. 16, pp. 109–115.
- Kawai, H.,** Hydrography of the Kuroshio and the Oyashio (in Japanese), *Physical Oceanography II, Fundamental Lectures of Oceanography*, Japan: Tokai Univ. Press, 1972, pp. 129–320.
- Kobayashi, T., Wako, M., and Naito, M.,** Studies on the life of the Pacific saury, *Cololabis saira* (Brevoort). I. Aggregative characteristics of adult of the autumn-spawning population, *Sci. Rep. Hokkaido Fish. Exp. Stn.*, 1968, vol. 9, pp. 1–45.
- Kulik, V.V., Katugin, O.N., and Zuev, M.A.,** Climate change impacts on distribution patterns of boreopacific gonate squid (*Boreoteuthis borealis*) in the Northwest Pacific, *Proceedings of the 25 Year of PICES: Celebrating the Past, Imagining the Future*, San Diego, CA, USA: North Pacific Marine Science Organization, 2016, pp. 148.
- Lee, B., Shin, M., Oh, Y., Oh, H., and Ryu, K.,** Identification of protein functions using a machine-learning approach based on sequence-derived properties, *Proteome Sci.*, 2009, vol. 7, no. 1, pp. 27. doi 10.1186/1477-5956-7-27
- Li, W. and Guo, Q.,** How to assess the prediction accuracy of species presence-absence models without absence data?, *Ecography*, 2013, vol. 36, no. 7, pp. 788–799. doi 10.1111/j.1600-0587.2013.07585.x
- Michalsky, J.J.,** The Astronomical Almanac's algorithm for approximate solar position (1950–2050), *Solar Energy*, 1988, vol. 40, no. 3, pp. 227–235. doi 10.1016/0038-092X(88)90045-X
- Ogawa, Y.,** Variation of the southern end latitude of the first Oyashio intrusion, *Bull. Tohoku Reg. Fish. Lab.*, 1989, vol. 51, pp. 1–9.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., and Wulder, M.A.,** Good practices for estimating area and assessing accuracy of land change, *Remote Sensing of Environment*, 2014, vol. 148, pp. 42–57. doi 10.1016/j.rse.2014.02.015
- Phillips, S.J., Anderson R.P., and Schapire R.E.,** Maximum entropy modeling of species geographic distributions, *Ecol. Modell.*, 2006, vol. 190, no. 3–4, pp. 231–259. doi 10.1016/j.ecolmodel.2005.03.026
- Pontius, R.G. and Millones M.,** Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment, *Int. J. Remote Sens.*, 2011, vol. 32, no. 15, pp. 4407–4429. doi 10.1080/01431161.2011.552923
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G.,** Akaike Information Criterion Statistics, *Mathematics and its Applications*, Springer Netherlands, 1986.
- Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R.,** Real-time human pose recognition in parts from single depth images, *Commun. ACM*, 2013, vol. 56, no. 1, pp. 116–124. doi 10.1145/2398356.2398381
- Syah, A.F., Saitoh, S.-I., Alabia, I.D., and Hirawake, T.,** Detection of potential fishing zone for Pacific saury (*Cololabis saira*) using generalized additive model and remotely sensed data, *IOP Conf. Ser.: Earth Environ. Sci.*, 2017, vol. 54, pp. 012074. doi 10.1088/1755-1315/54/1/012074
- Syah, A.F., Saitoh, S.-I., Alabia, I., and Hirawake, T.,** Habitat model development of Pacific saury (*Cololabis saira*) using satellite remotely sensed data in the Northwestern North Pacific, *12th Biennial Conference of Pan Ocean Remote Sensing Conference (PORSEC 2014)*, Bali, Indonesia, 2014, pp. 1–10.
- Syah, A.F., Saitoh, S.-I., Alabia, I.D., and Hirawake, T.,** Predicting potential fishing zones for Pacific saury (*Cololabis saira*) with maximum entropy models and remotely sensed data, *Fish. Bull.*, 2016, vol. 114, no. 3, pp. 330–342. doi 10.7755/FB.114.3.6
- Takasugi, S. and Yasuda, I.,** Index temperature at 100 m depth of the Oyashio front in the Iwate coastal region (in Japanese with English abstract and captions), *Japan J. Fish. Oceanogr.*, 1993, vol. 57, pp. 333–344.
- Taylor, K.E.,** Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res. Atmos.*, 2001, vol. 106, no. D7, pp. 7183–7192. doi 10.1029/2000JD900719
- Uda, M.,** Fishing centre of “Sanrna” (*Cololabis saira*, Brevoort) correlated with the head of Oyashio cold current (In Japanese with English summary), *Bull. Jap. Soc. Sci. Fish.*, 1936, vol. 5, pp. 236–238.
- Usui, N. Ishizaki, S., Fujii, Y., Tsujino, H., Yasuda, T., and Kamachi, M.** Meteorological Research Institute multivariate ocean variational estimation (MOVE) system: Some early results, *Adv. Sp. Res.*, 2006, vol. 37, no. 4, pp. 806–822. doi 10.1016/j.asr.2005.09.022

Usui, N., Wakamatsu, T., Tanaka, Yu., Hirose, N., Toyoda, T., Nishikawa, S., Fujii, Y., Takatsuki, Y., Igarashi, H., Nishikawa, H., Ishikawa, Y., Kuragano, T., Kamachi, M., Four-dimensional variational ocean reanalysis: a 30-year high-resolution dataset in the western North Pacific (FORA-WNP30), *J. Oceanogr.*, 2017, vol. 73, no. 2, pp. 205–233. doi 10.1007/s10872-016-0398-5

Werner, F.E., Ito, S.-I., Megrey, B.A., and Kishi, M.J., Synthesis of the NEMURO model studies and future directions of marine ecosystem modeling, *Ecol. Modell.*, 2007, vol. 202, no. 1–2, pp. 211–223. doi 10.1016/j.ecolmodel.2006.08.019

Wood, S.N., Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *J. R. Statist. Soc. B (Statistical Methodology)*, 2011, vol. 73, no. 1, pp. 3–36. doi 10.1111/j.1467-9868.2010.00749.x

Wood, S.N., *Generalized Additive Models: An Introduction with R*, New York: Chapman and Hall/CRC, 2017, second edition.

Wood, S.N., Thin plate regression splines, *J. R. Statist. Soc. B (Statistical Methodology)*, 2003, vol. 65, no. 1, pp. 95–114. doi 10.1111/1467-9868.00374

Wright, M.N. and Ziegler, A., ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, *J. Stat. Softw.*, 2017, vol. 77, no. 1, pp. 1–17. doi 10.18637/jss.v077.i01

Yasuda, I. and Watanabe, Y., On the relationship between the Oyashio front and saury fishing grounds in the north western Pacific: A forecasting method for fishing ground locations, *Fish. Oceanogr.*, 1994, vol. 3, no. 3, pp. 172–181. doi 10.1111/j.1365-2419.1994.tb00094.x

Yoshida, T., Climatological seasonal variations of the distribution of Oyashio cold water, *Umi to Sora*, 1992, vol. 68, no. 2, pp. 79–88 (in Japanese with English abstract).

Поступила в редакцию 18.09.2019 г.

После доработки 17.10.2019 г.

Принята к публикации 29.10.2019 г.