

**МЕТОДИКА ИССЛЕДОВАНИЙ
METHODS OF INVESTIGATIONS**

Научная статья

УДК [639.2.053.7:519](265.5)**DOI: 10.26428/1606-9919-2022-202-1002-1014****EDN: IAVNBZ****ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ
ДЛЯ ОЦЕНКИ БИОМАССЫ ТРЕСКИ В СЕВЕРО-КУРИЛЬСКОЙ ЗОНЕ****В.В. Кулик, М.И. Горюнов***Тихоокеанский филиал ВНИРО (ТИНРО),
690091, г. Владивосток, пер. Шевченко, 4

Аннотация. Исследуются уловы тихоокеанской трески *Gadus macrocephalus* на промысле и в научной съемке с целью найти относительную биомассу трески и ее неопределенность в Северо-Курильской зоне в 2021 и 2022 гг. с использованием многофакторного подхода. Используется вся доступная в ТИНРО информация за эти годы для проведения сравнительного анализа посредством метода машинного обучения — ансамблевого метода случайного леса (Random forest) в процедуре множественного заполнения пропусков последовательными уравнениями — Multiple Imputation by Chained Equations (MICE). Коэффициент детерминации относительно тестовых данных превышал 0,8 при использовании данных научной съемки 2021 г., а при использовании промысловых данных он превышал 0,5. Тем не менее относительно исходной дисперсии: таковая в MICE была ниже более чем на 82 % по данным научной съемки. Выявлено, что биомасса трески осталась на прежнем уровне в 2022 г. относительно 2021 г. Предлагается расширить район исследований и список наблюдаемых орудий лова из-за их пространственной сегрегации для снижения вероятности искажения оценок в связи с большой площадью экстраполяции.

Ключевые слова: биомасса, тихоокеанская треска, Северо-Курильская зона, случайный лес, MICE

Для цитирования: Кулик В.В., Горюнов М.И. Применение метода машинного обучения для оценки биомассы трески в Северо-Курильской зоне // Изв. ТИНРО. — 2022. — Т. 202, вып. 4. — С. 1002–1014. DOI: 10.26428/1606-9919-2022-202-1002-1014. EDN: IAVNBZ.

Original article

**Application of the machine learning method to estimate the biomass of pacific cod
in the North Kuril zone****Vladimir V. Kulik*, Mikhail I. Goryunov****

*, ** Pacific branch of VNIRO (TINRO), 4, Shevchenko Alley, Vladivostok, 690091, Russia

* Ph.D., head of laboratory, vladimir.kulik@tinro-center.ru, ORCID 0000-0003-0920-5312

** leading specialist, mikhail.goryunov@tinro-center.ru

Abstract. The biomass of pacific cod (*Gadus macrocephalus*) in the North Kuril fishing zone is estimated using a multifactorial approach, with evaluation of uncertainty. For this

* Кулик Владимир Владимирович, кандидат биологических наук, заведующий лабораторией, vladimir.kulik@tinro-center.ru, ORCID 0000-0003-0920-5312; Горюнов Михаил Игоревич, ведущий специалист, mikhail.goryunov@tinro-center.ru.

© Кулик В.В., Горюнов М.И., 2022

purpose, the density of fish over entire zone is restored using the data on density obtained in 2022 compared with the data of previous surveys and fishery data obtained in 2021 and earlier, converted to the same scale, with application of the machine learning method, as the random forest in the multiple imputation by chained equations procedure (MICE). The coefficient of the restored data determination with out-of-bag (test set) data was > 0.8 with the data of scientific survey in 2021 and > 0.5 with the data of Danish seine observations. The cod density variance in MICE data was in 82 % lower than in the data of the scientific survey; therefore the biomass estimation with MICE data has lower uncertainty than that one calculated just from the mean density in survey. The study showed insignificant difference of the cod biomass in 2021 and 2022. Spatial segregation is revealed for fishing gears used for the pacific cod fishery. There is proposed to extend the list of fishing gears and to expand the study area for reducing possible bias in the biomass estimation due to large area of extrapolation.

Keywords: biomass, pacific cod, North Kuril fishing zone, random forest, MICE

For citation: Kulik V.V., Goryunov M.I. Application of the machine learning method to estimate the biomass of pacific cod in the North Kuril zone, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 2022, vol. 202, no. 4, pp. 1002–1014. DOI: 10.26428/1606-9919-2022-202-1002-1014. EDN: IAVNBZ.

Введение

Северо-Курильская промысловая зона характеризуется сложным рельефом дна, что препятствует выполнению научных съемок донными тралами (ДТ) с охватом всей ее акватории. Научные съемки ДТ по сетке станций проводятся там редко и только в доступной для тралений части. Последняя была проведена в 2021 г. на НИС «Дмитрий Песков». ТИНРО отправил в 2022 г. наблюдателя на снюрреводный промысел в Северо-Курильскую зону. Следовательно, необходимо разработать метод, позволяющий сравнить биомассы 2021 и 2022 гг., полученные из различных источников информации.

Доминирующим видом из придонных рыб в наблюденных уловах 2022 г. была тихоокеанская треска *Gadus macrocephalus*, поэтому мы выбрали ее в качестве объекта исследования.

Целью данной работы стало определение значимости изменения биомассы трески в Северо-Курильской зоне с использованием всей доступной нам информации за 2021 и 2022 гг.

Абсолютные биомассы в научных съемках рассчитываются с учетом коэффициента уловистости (КУ). Однако для снюрреводов КУ нам неизвестен, поэтому мы не будем использовать его в расчетах по данным 2021 и 2022 гг., а станем оперировать относительными биомассами.

Мы предлагаем использовать робастный метод, позволяющий моделировать наблюдения вне зависимости от их распределения и без смещения. Таковым является ансамблевый метод случайного леса (Random forest или RF) [Breiman, 2001]. Он уже позволил выявить верную тенденцию в динамике биомассы сайры в районе Курильских островов без использования данных научных съемок [Kulik et al., 2022]. Сайра совершает более протяженные миграции, чем треска, но для нее были доступны суточные лагранжевы характеристики на поверхности Тихого океана, отражающие динамику вод верхнего слоя воды, а треска обитает у дна. Это значит, что использованный ранее метод необходимо усовершенствовать с учетом известных нам важных факторов для трески по всей акватории Северо-Курильской зоны.

Материалы и методы

Сбор новых материалов проведен М.И. Горюновым на борту судна СТР-420 «Ветровой» в Северо-Курильской зоне за 37 судно-суток лова с конца мая по июнь 2022 г. В качестве вспомогательной информации мы привлекли всю доступную в ТИНРО промысловую статистику за это же время года в 2021–2022 гг. Информация о промысловых операциях получена из отраслевой системы мониторинга (ОСМ) Росрыболовства, которая ведется в Центре системы мониторинга рыболовства и связи [<http://cfmc.ru/>] и

сохраняется в ТИНРО при помощи сотрудников Института космических исследований РАН. Материалы научной съемки 2021 г. также служили в качестве входной информации по распределению относительной биомассы трески.

Важной константой, определяющей величину относительной биомассы (кг/км²), в научных съемках донными тралами является горизонтальное раскрытие трала. Оно может быть пересчитано от длины верхней подборки через коэффициент, равный 0,55 [Волвенко, 2013]. Такой коэффициент использовался в обобщениях по выбранному району исследований [Макрофауна..., 2014а, б], но не в рейсовых отчетах и не в материалах общего допустимого улова (ОДУ), где среднее горизонтальное раскрытие равнялось 16 м для стандартного ДТ с длиной верхней подборки, равной 27,1 м (ДТ 27,1/24,4), а не $27,1 \times 0,55 = 14,9$ м. Известно, что на шельфе западной части Берингова моря горизонтальное раскрытие у этого трала больше 16 м (16,4 м) [Захаров, Емелин, 2016]. Инструментальные наблюдения горизонтального раскрытия показали, что они не равны 16 м в 68 % случаев [Захаров и др., 2013]. Более того, горизонтальное раскрытие увеличивается с глубиной донного траления и длиной вытравленных ваеров нелинейно. Так, при длине ваеров 900 м оно составит уже 20,1 м [Захаров, Емелин, 2016]. Поэтому мы пересчитали горизонтальное раскрытие научных тралений ДТ 27,1 по опубликованным зависимостям в программе, разработанной авторами, которые их обнаружили [Захаров и др., 2019], а для остальных ДТ в ОСМ горизонтальное раскрытие получено через 0,55 длины верхней подборки [Волвенко, 2013].

Наблюдения ДТ сохранены в ТИНРО в базе данных (БД) «Морская биология»*. Из нее же мы получили описания грунтов в каждом ДТ. Грунты кодировались 2 цифрами. Разбив этот составной код, мы составили список из уникальных значений в каждой страте, добавив буквы «gr» в начале, чтобы было проще использовать их как названия переменных (прил. А табл. А.1)**. К сожалению, не все участники научных рейсов придерживались кодификатора грунтов, некоторые обозначали их буквами, но эти аббревиатуры мы перевели в цифровые коды сами.

Вспомогательными слоями для интерполяции служили также данные о придонной температуре, солености, скорости течений, нитратах, фосфатах, силикатах, растворенном кислороде, железе, хлорофилле, фитопланктоне, первичной продукции и интенсивности света у дна (они получены из системы Bio-ORACLE с разрешением в 5' [Tyberghein et al., 2012; Assis et al., 2018]), а также данные о физических особенностях дна: восточно-западная и северно-южная составляющие экспозиции, наклона, вогнутость морского дна, кривизна в плане и профиле и расстояние от берега получены из системы MARSPEC с разрешением в 30'' [Sbrocco, Barber, 2013]. С еще более высоким разрешением в сетке с ячейкой 100 × 100 м рассчитана морфометрия дна из цифровой модели рельефа GEBCO [<https://doi.org/10.5285/a29c5465-b138-234de053-6c86abc040b9>] с разрешением в 15'' после низкочастотного сглаживания радиусом до 2 км для исключения артефактов глубины из-за различий в ее источниках и билинейной интерполяции. Новая морфометрия рассчитана в геоинформационной системе (ГИС) SAGA версии 7.8.2 и показана на рис. Б.1–Б.7 в приложении Б. Зональная статистика со всех растров собрана взвешенно по площади снятия ячеек при помощи специально разработанного с этой целью пакета `exactextractr` [<https://CRAN.R-project.org/package=exactextractr>]. Площадь дна (прил. Б рис. Б.8), как и в наших прежних работах [Макрофауна..., 2014а, б], рассчитана в модуле ГИС SAGA*** на основе цифровой модели рельефа. Сетка агрегирования всех значений образована ромбами шириной 4 км, высотой 5 км и площадью 10 км² в конформной конической проекции

* Свидетельство № 6394 от 01.08.2000 г. о регистрации в Государственном регистре баз данных за № 0220006765.

** Все приложения размещены на странице статьи на сайте журнала [<http://izvestiya.tinro-center.ru>] как дополнительный файл.

*** Olaya V. Module Real Surface Area, SAGA-GIS, 2004.

Ламберта с опорными параллелями 48° и 50° с.ш. и центральным меридианом 155° в.д. Коррекция искажения площади поверхности из-за проецирования рассчитана в ГИС QGIS Desktop версии 3.26 (прил. Б рис. Б.9).

Основными слоями для интерполяции были наблюдения на промысле снюрреводом в 2022 г. Постановки снюрревода начинались со сброса буя за борт. Далее судно проходило по маршруту, совершало один разворот. Затем за борт опускался сам мешок снюрревода, после чего судно возвращалось за буюм, произведя еще один разворот. Таким образом, зона облова походила на четырехугольник со средней площадью 0,832 км². Мы использовали уникальную площадь каждого такого четырехугольника, рассчитанную в ГИС QGIS на сфероиде всемирной геодезической системы, а не среднюю. Процесс обмета осуществлялся за 20–25 мин. Затем снюрревод протягивали за судном для облова обнаруженной рыбы, что в зависимости от скорости хода судна и силы течения занимало от 30 до 60 мин. После этого начинался подъем снюрревода с уловом на борт, на что уходило около 30–40 мин. В зависимости от величины улова данный процесс мог занять от 10 до 25 мин. Таким образом, один замет снюрревода с момента сброса буя с грузом до подъема мешка с уловом занимал от 1 ч 30 мин до 2 ч 30 мин. Этот способ ведения лова может представлять научный интерес из-за его скоротечности и легко рассчитываемой площади облова.

Использованный метод многомерной аппроксимации базируется на RF, реализованном в пакете ranger [Wright, Ziegler, 2017] для языка программирования R [<https://www.R-project.org/>]. Восстановление вероятных плотностей гидробионтов (кг/км²) после логарифмической трансформации проведено многократным перезапуском (15 раз) по 100 наборам, которые каждый пропуск моделировали по всем имеющимся данным в случайном порядке в итерационной процедуре с удержанием общего математического ожидания (mean match). Такой подход называется Multiple Imputation by Chained Equations (MICE), что можно перевести как «множественное заполнение пропусков последовательными уравнениями» [<https://github.com/FarrellDay/miceRanger>]. Отличие от оригинальных установок ranger в MICE здесь заключалось в увеличении числа деревьев до 1000 (с 500) и глубины их роста (параметр mtry) до 100 (с 10). Оригинальные установки позволяют выполнять аналогичные расчеты быстрее, но менее точно.

Заметы снюрревода судном с наблюдателем в 2022 г. располагались кучно, что сильно затрудняет интерполяционную задачу. Поэтому мы привлекли всю доступную информацию из ОСМ по всем операциям в Северо-Курильской зоне в мае-июне 2020 и 2021 гг. В случае проблематичности расчета площади облова некоторых орудий лова в ОСМ мы использовали за единицу усилия часы застоя или 1000 крючков. Таким образом, получено еще 23 вспомогательные переменные (прил. А табл. А.2).

Привлечение промысловой статистики позволило заметно расширить акваторию, насыщенную хоть какой-то информацией (рис. 1).

Использование только наблюденных снюрреводов оставит для модели слишком много свободы в районах, где научных наблюдений на промысле не было. В связи со случайными цепочками, образующимися в каждом наборе (из 100), на каждой итерации (из 15), в каждом дереве (из 1000) можно установить условные и многомерные, интервальные связи между орудиями лова и содержать экстраполяцию плотности из наблюденных постановок снюрревода по плотностям уловов из других орудий лова. Использование среднеголетних долей и уловов на единицу усилия из БД «Ярусный промысел»* и БД «Морская биология» помогало решить эту задачу.

Все плотности из БД в ТИНРО были подвергнуты логарифмической трансформации для настройки и экспоненциальному инвертированию при обобщении результатов специальными функциями из языка R: $\log1p$ и $\expm1$, которые лучше работают со значениями, близкими к пределу машинной точности, чем простые преобразования $\ln(x+1)$ и $\exp(x)-1$ (прил. А табл. А.3). Такое преобразование необходимо потому, что

* Свидетельство № 2017620882 от 11.08.2017 г. о Государственной регистрации базы данных «Ярусный промысел».

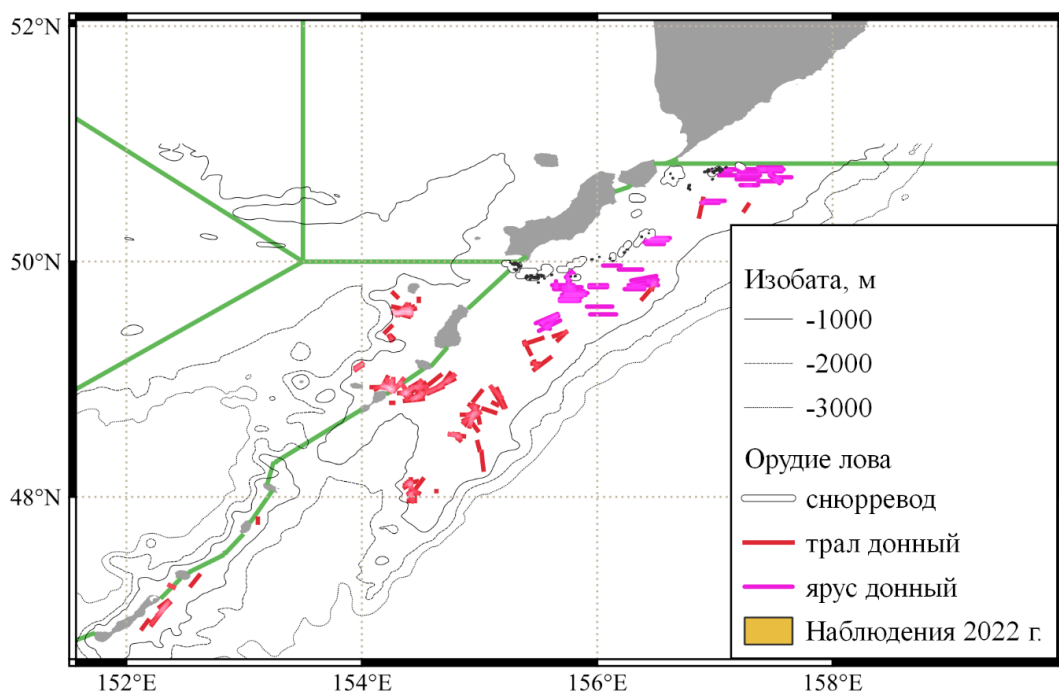


Рис. 1. Положения промысловых операций из ОСМ в Северо-Курильской зоне с мая по июнь 2020–2022 гг.

Fig. 1. Location of fishing operations in the North Kuril fishing zone in May-June of 2020–2022 obtained from the fishery monitoring system

распределение рыб не является нормальным, оно определяется условиями среды, наличием корма и т.д., поэтому в общем случае не может моделироваться нормальным законом, а средний арифметический улов, особенно при небольшом количестве станций облова, может содержать значительную случайную погрешность и оказаться неэффективной оценкой среднего улова. Следовательно, получение наиболее эффективной оценки величины популяции рыб и определения точности полученного результата зависит от закономерности распределения уловов [Аксютина, 1970а, б].

Сотрудники ТИНРО давно обращали внимание на эту проблему, отмечая, что склонность рыб к образованию скоплений с более или менее пустыми пространствами между ними ведет к тому, что большие уловы встречаются значительно реже, чем малые. Как следствие, данные траловой съемки явно не соответствуют нормальному распределению. Поэтому применение нормальных (параметрических) статистических процедур, даже таких элементарных, как расчет среднего и его стандартной ошибки, даст искаженные результаты [Волвенко, 1998].

В результате поиска путей решения проблемы ненормальности плотностей в траловых съемках многократно предлагалось логарифмическое преобразование $y = \lg(x+1)$, где x — исходные, а y — преобразованные данные [Аксютина, 1968]. Логарифмическое преобразование обеспечивает удовлетворительную нормализацию, следовательно, плотности гидробионтов имеют логнормальное распределение [Волвенко, 1998].

Промысловые уловы же в связи с отсутствием нулевых уловов при добыче целевого вида и примерно постоянными его плотностями, находимыми акустическими приборами и иными способами, наоборот, близки к нормальным при стратификации по орудиям лова. Нестабильность промысловых уловов может быть связана со множеством причин, одна из которых — истощение запаса. Наблюдаемое на промысле истощение запаса позволяет восстановить начальную численность рыб в районе промысла, например методом Лесли и де Лури с модификацией Рикера [Ricker, 1975]. Мы испытали эти

методы, прежде чем приступить к использованию метода машинного обучения — RF. Программная реализация методов Лесли и де Лури доступна в пакете FSA для языка R [<https://github.com/fishR-Core-Team/FSA>].

Результаты и их обсуждение

Видовой состав при снюрреводном промысле СТР-420 «Ветровой» в Северо-Курильской зоне в 2022 г. был представлен 34 видами рыб, 4 видами крабов, 1 видом осьминога и 1 видом кальмара (см. таблицу). Доминировал в уловах минтай *Gadus chalcogrammus*, а треска и двухлинейная камбала *Lepidopsetta polyxystra* были субдоминантами.

В связи с тем что минтай обитает у дна непостоянно и может встречаться по всей толще воды, мы не использовали его плотность как возможный фактор, определяющий вероятность поимки трески в MICE, а плотность другого придонного субдоминанта (двухлинейной камбалы) использовали.

Анализ динамики плотности рыб с применением методов истощения Лесли и де Лури с модификацией Рикера показал, что ни в целом, ни в отдельности по трем подрайонам (у южной, центральной и северной частях о. Парамушир) данные методы применять невозможно. Их выводы между собой не различались, поэтому приводятся результаты по методу де Лури (прил. В).

Наклон прямой истощения трески был либо положительным, что приводило к отрицательной начальной численности (прил. В рис. В.1 и рис. В.2), либо статистически незначимым ($p < 0,05$) (прил. В рис. В.3). В последнем случае оценка численности трески у северной части о. Парамушир, найденная методом де Лури, составила около $6,566 \pm 5,921$ тыс. экз. в пределах стандартной ошибки. У центральной части острова наблюдений было недостаточно, чтобы сделать хоть какое-то заключение.

Таким образом, применение методов истощения оказалось безуспешным. Применение метода машинного обучения — случайного леса — в MICE оказалось более надежным.

Качество (Q) восстановления в MICE плотности трески (kggad) в 2022 г. было выше, чем в 2021 г. (gad), как показано на рис. 2.

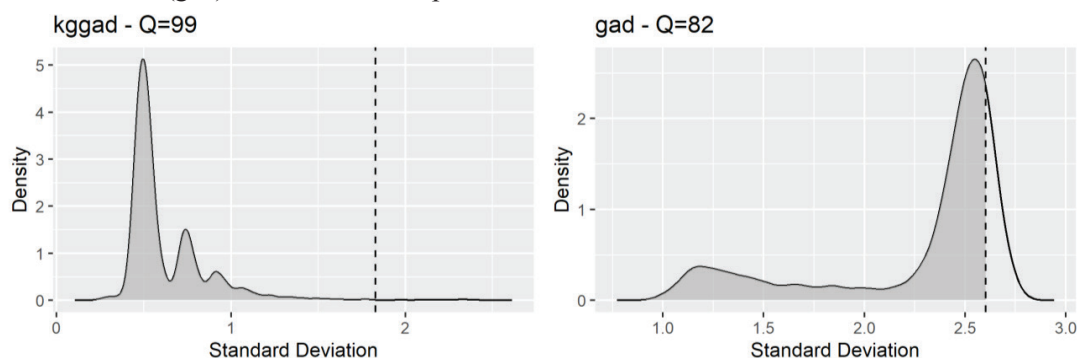


Рис. 2. Качество (Q) настройки: доля стандартного отклонения (Standard Deviation — SD) модельных данных ниже оригинального SD, отмеченного вертикальным пунктиром, %. Density — плотность распределения

Fig. 2. Quality (Q) of tuning: Percentage of the model samples with standard deviation (SD) below SD of observed data (vertical dashed line)

Коэффициент детерминации (R^2) модельных плотностей относительно наблюдаемых, но исключенных (OOB — out of bag) из тренировки RF, а затем включенных в тестирование, был выше для научных ДТ в 2021 г., чем для снюрреводных постановок в 2022 г. (рис. 3). Из этого можно сделать важный вывод — данные из научной съемки 2021 г., обладая более высокой дисперсией и, следовательно, меньшей определенностью, одновременно обладают большей обобщающей способностью и более пригодны для экстраполяции, чем данные наблюдений на снюрреводном промысле. При этом в последние годы оценки биомасс по всем 100 наборам были устойчивы, так как их

Наблюдаемый видовой состав уловов при снюрреводном промысле
в Северо-Курильской зоне в мае-июне 2022 г.
Species composition of the Danish seine catches in the North Kuril fishing zone in May-June 2022

Вид	Встре- чаемость, %	Длина, см		Сред- няя масса, г	Средний улов на судо-сутки		Доля от общего улова, % от массы
		От	До		Кол-во особей, экз.	Масса, кг	
<i>Gadus chalcogrammus</i>	99,17	21	67	431	49701	21404,78	73,19
<i>Lepidopsetta polyxystra</i>	94,21	17	47	328	5122	1680,71	6,05
<i>Gadus macrocephalus</i>	85,95	15	84,9	476	6968	3315,28	13,08
<i>Gymnocanthus detrisus</i>	58,68	15	38	203	673	136,77	0,79
<i>Erimacrus isenbeckii</i>	58,68	8	13	703	15	10,83	0,063
<i>Chionoecetes opilio</i>	57,85	4	18	313	73	22,91	0,134
<i>Hemilepidotus jordani</i>	55,37	18	51	859	413	354,45	2,171
<i>Pleurogrammus monopterygius</i>	42,15	21	49	599	523	313,25	2,520
<i>Hippoglossus stenolepis</i>	23,14	30	126	813	19	15,66	0,229
<i>Hexagrammos lagocephalus</i>	21,49	28	57	1761	34	59,61	0,941
<i>Myoxocephalus polyacanthocephalus</i>	19,83	21	75	694	18	12,65	0,216
<i>Limanda aspera</i>	14,05	20	42	465	15	6,84	0,165
<i>Hippoglossoides robustus</i>	12,40	17	45	436	12	5,14	0,141
<i>Paralithodes brevipes</i>	12,40	3	16	152	1	0,19	0,0053
<i>Paralithodes camtschaticus</i>	6,61	18	23	3210	0,025	0,08	0,0041
<i>Bathyraja parmifera</i>	4,96	59	83	1445	0,096	0,138	0,0095
<i>Hemilepidotus gilberti</i>	3,31	27	33	427	1,97	0,842	0,0864
<i>Eleginus gracilis</i>	3,31	37	41	391	1,22	0,477	0,0489
<i>Hemitripteris villosus</i>	3,31	39	51	2540	0,0056	0,0141	0,0014
<i>Berryteuthis magister</i>	3,31	22	26	448	0,0056	0,00249	0,00026
<i>Liparis ochotensis</i>	3,31	38	52	1283	0,0056	0,0071	0,00073
<i>Pleuronectes quadrituberculatus</i>	2,48	35	48	1176	0,39	0,458	0,0627
<i>Enophrys diceraus</i>	2,48	23	30	414	0,234	0,097	0,0133
<i>Eumicrotremus asperrimus</i>	2,48	8	9	101	0,023	0,0023	0,00032
<i>Podothecus sturioides</i>	2,48	19	31	150	0,0031	0,00047	0,000064
<i>Sebastes glaucus</i>	1,65	40	53	1211	0,096	0,116	0,0238
<i>Bathyraja aleutica</i>	1,65	105	122	10055	0,00139	0,014	0,00287
<i>Lycodes brunneofasciatus</i>	1,65	56	57	930	0,00139	0,00129	0,00027
<i>Percis japonica</i>	1,65	31	38	310	0,00139	0,00043	0,000088
<i>Triglops forficatus</i>	1,65	23	28	85	0,087	0,0074	0,00151
<i>Careproctus rastrinus</i>	0,83	26	33	527	0,184	0,097	0,04
<i>Icelus spiniger</i>	0,83	17	20	56	0,173	0,01	0,004
<i>Limanda sakhalinensis</i>	0,83	24	24	191	0,0118	0,0023	0,00093
<i>Bathyraja maculata</i>	0,83	63	63	1640	0,00035	0,00057	0,00023
<i>Aptocyclus ventricosus</i>	0,83	24	24	1540	0,00035	0,00053	0,00022
<i>Oncorhynchus</i> sp.	0,83	57	57	1900	0,00035	0,00066	0,00027
<i>Octopus conispadiceus</i>	0,83	–	–	950	0,00035	0,00033	0,00014
<i>Myzopsetta proboscidea</i>	0,83	36	36	530	0,00035	0,00018	0,000076
<i>Clupea pallasii</i>	0,83	30	30	200	0,00035	0,000069	0,000029
<i>Glyptocephalus stelleri</i>	0,83	33	33	200	0,00035	0,000069	0,000029

ошибки не проявили заметных сдвигов и тенденций между итерациями и случайными подвыборками данных, а доверительные интервалы были достаточно узкими для заключения о значимости или незначительности изменений относительной биомассы трески с 2021 по 2022 г.

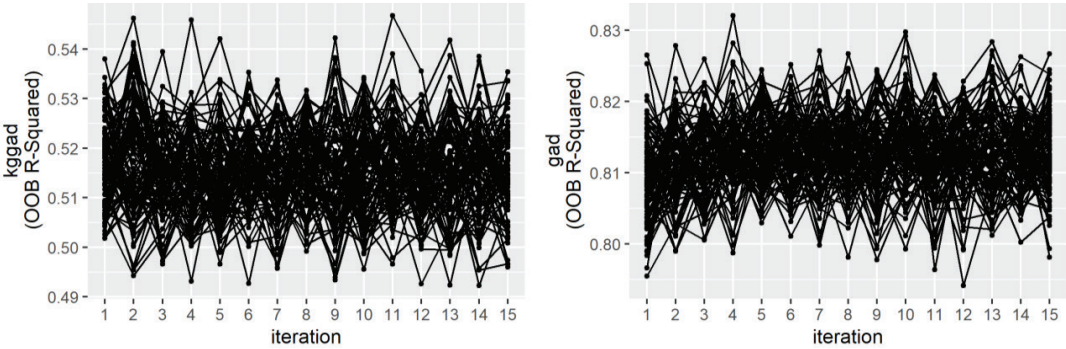


Рис. 3. Коэффициент детерминации (R-Squared) относительно тестовых данных, исключенных из настройки
Fig. 3. Coefficient of determination (R^2) of the model data relative to the observed data excluded multiple times from the tuning (at each iteration in each of 100 data sets)

Анализ динамики плотности трески относительно научных уловов ДТ 27,1 в 2021 г., которые использовались в MICE для заполнения пропусков в многомерном пространстве в 2022 г., показал, что биомасса трески значительно не изменилась. Медианы ее 100 оценок находились на уровне 27,8 тыс. т в 2021 г. и 27,1 тыс. т в 2022 г. Эта биомасса не является абсолютной в связи с тем, что в данном расчете КУ не применялись. Нам важнее межгодовые изменения и их 95 %-ные доверительные интервалы. Последние находились в пределах от 26,7 до 28,7 тыс. т в 2021 г. и от 26,6 до 27,7 тыс. т в 2022 г. Таким образом, мы можем утверждать, что биомасса трески в Северо-Курильской зоне не претерпела больших изменений в 2022 г. относительно 2021 г.

Картина распределения биомассы трески показана на рис. 4. Необходимо обратить внимание на важность научных наблюдений на промысле. Например, доля трески в уловах ярусами по историческим данным ТИНРО (рис. 5) сильно отличается по месту от таковой из аппроксимации по данным ОСМ (рис. 6).

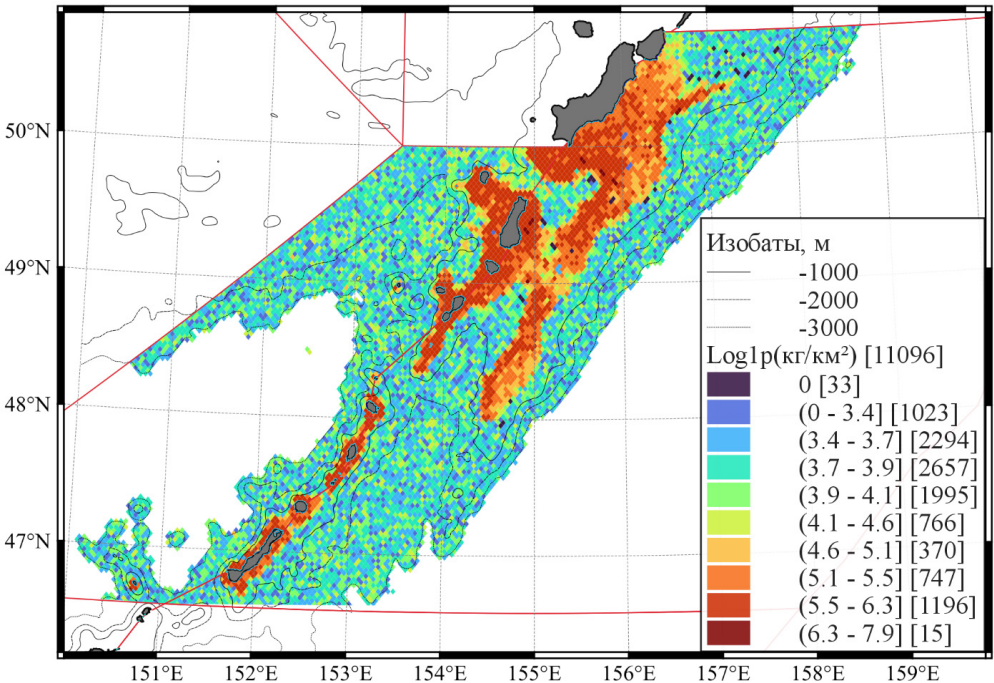


Рис. 4. Плотность трески в Северо-Курильской зоне по данным 2022 г.
Fig. 4. Observed data on the pacific cod density in the North Kuril fishing zone in 2022

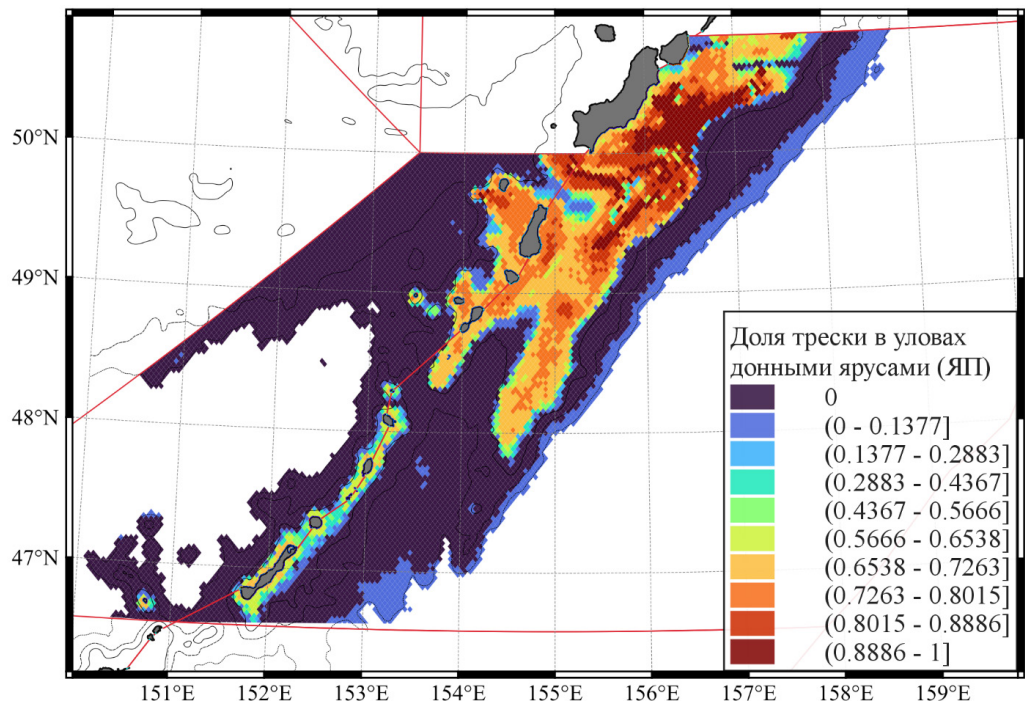


Рис. 5. Среднегодовная доля трески в ярусных уловах из БД ЯП
Fig. 5. The long-term average portion of pacific cod in the longline catches

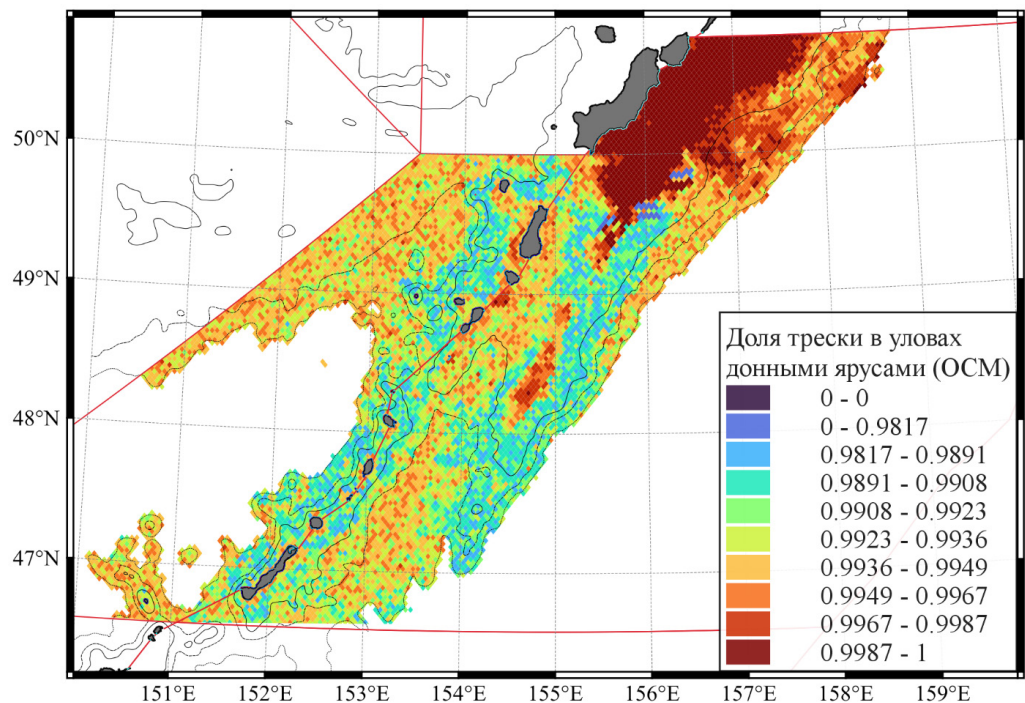


Рис. 6. Доля трески в ярусных уловах из ОСМ с мая по июнь в 2020–2022 гг.
Fig. 6. Portion of pacific cod in the longline catches in May-June of 2020–2022 (data of the fishery monitoring system)

Очевидно, что данные, интерполированные из научных наблюдений (в донных тралениях и на ярусном промысле), между собой хорошо сходятся, а с официальной статистикой (ОСМ) — не сходятся. Необходимо отметить, что промысловые данные

из ОСМ не позволяют должным образом учесть эффект глубины. В результате этого экстраполяция высокой плотности трески распространяется на глубины, где она не должна встречаться. Отсутствие там научных наблюдений вело к тому, что настройка RF такие сочетания диапазонов параметров не штрафовала, поэтому следует признать их артефактами экстраполяции. Следовательно, делать выводы о том, что на ярусном промысле трески приловы остальных видов незначительны по данным ОСМ, — преждевременно. В связи с этим мы можем утверждать следующее: нам необходимо расширить район исследований и список наблюдаемых орудий лова, если мы хотим снизить искажения оценок, которые сейчас возникают в связи с большой площадью экстраполяции.

Заключение

Простые методы истощения запасов Лесли и де Лури (с коррекцией по Рикеру) давали либо абсурдные (отрицательные) оценки начальной численности, либо неопределенные (с огромной ошибкой). Это может быть связано с коротким периодом наблюдений (менее 2 мес.), малым покрытием наблюдателями (всего 1) или положительным эффектом промыслового воздействия, которое лишь способствовало увеличению эффективности лова в мае-июне.

Метод машинного обучения — случайного леса, — запущенный в процедуре MICE, сработал значительно лучше простых средних и имел высокий коэффициент детерминации в тестовых прогонах. Огромное количество вспомогательных слоев информации (состав грунтов, морфометрия дна, придонные характеристики воды, вся информация из БД «Морская биология» и «Ярусный промысел», а также из ОСМ), доступной в каждой ячейке для интер- и экстраполяции, не помогло удержать оценки высоких плотностей в диапазонах глубин обитания трески при интерполировании промысловых данных из ОСМ. Это связано с тем, что промысловые операции недостаточно контрастны по признакам внешней среды в связи с их проведением на скоплениях рыб. В отличие от научных съемок, они не проводятся там, где рыбы не образуют достаточные для эффективного промысла концентрации. Так, глубина промысловых операций теряет роль определяющей распределение вида в пространстве при использовании в MICE.

В итоге мы все же можем с большой уверенностью допустить, что биомасса трески в 2022 г. осталась на уровне 2021 г., так как ее 95 %-ные доверительные интервалы в относительном масштабе (без КУ) почти полностью совпадали (от 26,7 до 28,7 тыс. т в 2021 г. и от 26,6 до 27,7 тыс. т в 2022 г.). Если отнести эти оценки к представленным Ким Сен Током в Сахалинском филиале ВНИРО в виде прогноза нерестового запаса в 2021 г. (85,93 тыс. т) и в 2022 г. (75,17 тыс. т) в Материалах ОДУ на 2022 г., то максимальный КУ для перевода относительной биомассы из MICE в абсолютную мог бы находиться в пределах от 0,33 до 0,36. Это очень близко к $KU = 0,4$ для взрослой трески [Макрофауна..., 2014а, б]. Наш вывод о незначительности изменения биомассы трески в Северо-Курильской зоне в 2022 г. соответствует прогнозу Ким Сен Тока, который ранее допускал, что формирование основы стада тремя доминирующими поколениями (2016–2018 годов рождения) позволит запасу сохраниться на стабильном уровне. Таким образом, наш вывод не противоречит прогнозу ответственного за подготовку материалов ОДУ трески в данном районе. Наш вывод получен иным методом: посредством MICE. Мы предлагаем этот метод не вместо традиционных, а в качестве дополнительного источника информации в условиях ее дефицита.

С целью снижения неопределенности в наших оценках потребуется расширить район исследований и список наблюдаемых орудий лова из-за их пространственной сегрегации. Расширение списка наблюдаемых орудий лова уже приводило к улучшению моделей распределения видов в восточной части Тихого океана [Thompson et al., 2022], поэтому мы надеемся на аналогичный эффект и в западной его части.

Необходимо помнить, что по наблюдениям в Северо-Курильской зоне на промысле снюрреводами в 2022 г. определенность значительно ниже (R^2 ООВ от 0,49 до 0,56), чем по данным научных ДТ в 2021 г. (R^2 ООВ от 0,80 до 0,85). При этом наращивание числа наблюдений на промысле не заставит рыбодобытчиков умышленно вести операции в зонах с низкой концентрацией рыб, а именно такие наблюдения крайне важны для более точной работы рассмотренного метода машинного обучения и любого другого метода моделирования распределения видов.

Благодарности (ACKNOWLEDGEMENTS)

Авторы выражают признательность всем участникам научных рейсов и наблюдателям на промысле, чьи материалы использованы для настройки моделей. Отдельная благодарность сотрудникам Центра системы мониторинга рыболовства и связи (ЦСМС) и Института космических исследований РАН (ИКИ РАН) за ведение базы данных ОСМ и предоставление доступа к ним.

The authors are thankful to all members of scientific cruises and observers whose materials were used to adjust the models. Special thanks to the staff of the Center for Fisheries Monitoring System and Communications and the Space Research Institute of the Russian Academy of Sciences (IKI RAS) for maintaining the OSM database and providing access to them.

Финансирование работы (FUNDING)

Материалы собраны и обработаны по договору на выполнение научно-исследовательских работ № 40–22 с АО «СК БСФ» от 13 апреля 2022 г.

The materials for the study were collected and processed under the contract for the performance of research work No. 40-22 with JSC «IC BSF» dated April 13, 2022.

Соблюдение этических стандартов (COMPLIANCE WITH ETHICAL STANDARDS)

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены. Авторы заявляют, что у них нет конфликта интересов.

All applicable international, national and/or institutional principles for the care and use of animals have been observed. The authors state that they have no conflict of interest.

Информация о вкладе авторов (AUTHOR CONTRIBUTIONS)

Концепция исследования, подготовка слоев для ГИС, настройка случайного леса, написание текста, редакторская правка — В.В. Кулик; сбор и обработка материалов исследований на снюрреводном промысле, перекодирование типов грунтов, написание введения и частично других разделов — М.И. Горюнов. Совместно обсуждены полученные результаты.

Research concept, data preparation for GIS, random forest tuning, text writing, editing — V.V. Kulik; collection and processing of the Danish seine data, recoding of the ground types, writing of introduction and partially other sections — M.I. Goryunov; all results were discussed jointly.

Список литературы

Аксютин З.М. Количественная оценка скопления рыб методом изолиний // Тр. ВНИРО. — 1970а. — Т. 71. — С. 302–308.

Аксютин З.М. Об использовании электронных вычислительных машин для анализа распределения и количественной оценки популяции рыб // Тр. ВНИРО. — 1970б. — Т. 71. — С. 309–318.

Аксютин З.М. Элементы математической оценки результатов наблюдений в биологических и рыбохозяйственных исследованиях : моногр. — М. : Пищ. пром-сть, 1968. — 288 с.

Волвенко И.В. Проблемы количественной оценки обилия рыб по данным траловых съемок // Изв. ТИНРО. — 1998. — Т. 124. — С. 473–500.

Волвенко И.В. Технические проблемы адекватной интерпретации результатов траловых съемок и пути их решения // Изв. ТИНРО. — 2013. — Т. 172. — С. 282–293.

Захаров Е.А., Емелин П.О. Актуальность использования аппаратуры контроля орудий лова для снижения погрешности в оценке численности гидробионтов // Изв. ТИНРО. — 2016. — Т. 186. — С. 198–206. DOI: 10.26428/1606-9919-2016-186-198-206.

Захаров Е.А., Кручинин О.Н., Мизюркин М.А., Сафронов В.А. Геометрические параметры донного трала 27,1/24,4 и возможные погрешности в оценке численности гидробионтов // Изв. ТИНРО. — 2013. — Т. 174. — С. 284–292.

Захаров Е.А., Кручинин О.Н., Шабельский Д.Л. Разработка и апробирование алгоритма расчета рабочих параметров траловых систем // Изв. ТИНРО. — 2019. — Т. 198. — С. 221–229. DOI: 10.26428/1606-9919-2019-198-221-229.

Макрофауна бентали Охотского моря: таблицы встречаемости, численности и биомассы. 1977–2010 / В.П. Шунтов, И.В. Волвенко, В.В. Кулик, Л.Н. Бочаров ; под ред. В.П. Шунтова и Л.Н. Бочарова. — Владивосток : ТИНРО-центр, 2014а. — 1052 с.

Макрофауна бентали северо-западной части Тихого океана: таблицы встречаемости, численности и биомассы. 1977–2008 / В.П. Шунтов, И.В. Волвенко, В.В. Кулик, Л.Н. Бочаров; под ред. В.П. Шунтова и Л.Н. Бочарова. — Владивосток : ТИНРО-центр, 2014б. — 554 с.

Assis J., Tyberghein L., Bosch S. et al. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling // Global Ecology and Biogeography. — 2018. — V. 27, Iss. 3. — P. 277–284. DOI: 10.1111/geb.12693.

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

Kulik V.V., Prants S.V., Uleysky M.Yu., Budyansky M.V. Lagrangian characteristics in the western North Pacific help to explain variability in Pacific saury fishery // Fish. Res. — 2022. — Vol. 252. — 106361. DOI: 10.1016/j.fishres.2022.106361.

Ricker W.E. Computation and Interpretation of Biological Statistics of Fish Populations. — Ottawa : Dep. Environ. Fish. Mar. Serv., 1975.

Sbrocco E.J., Barber P.H. MARSPEC: ocean climate layers for marine spatial ecology // Ecology. — 2013. — Vol. 94, Iss. 4. — P. 979. DOI: 10.1890/12-1358.1.

Thompson P.L., Anderson S.C., Nephin J. et al. Integrating trawl and longline surveys across British Columbia improves groundfish distribution predictions // Can. J. Fish. Aquat. Sci. — 2022. DOI: 10.1139/cjfas-2022-0108

Tyberghein L., Verbruggen H., Pauly K. et al. Bio-ORACLE: a global environmental dataset for marine species distribution modeling // Global Ecology and Biogeography. — 2012. — Vol. 21, Iss. 2. — P. 272–281. DOI: 10.1111/j.1466-8238.2011.00656.x.

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.

References

Aksyutina, Z.M., Quantitative assessment of fish aggregations by the method of contour lines, *Tr. Vses. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 1970, vol. 71, pp. 302–308.

Aksyutina, Z.M., On the use of electronic computers for the analysis of the distribution and quantification of fish populations // *Tr. Vseross. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 1970, vol. 71, pp. 309–318.

Aksyutina, Z.M., *Elementy matematicheskoi otsenki rezul'tatov nablyudenii v biologicheskikh i rybokhozyaistvennykh issledovaniyakh* (Elements of Mathematical Evaluation of the Results of Observations in Biological and Fishery Research), Moscow: Pishchevaya Promyshlennost', 1968.

Volvenko, I.V., Problems in quantitative estimation of fish abundance by trawl sampling, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 1998, vol. 124, pp. 473–500.

Volvenko, I.V., Technical problems of adequate interpretation the results of trawl surveys and ways to solve them, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 2013, vol. 172, pp. 282–293.

Zakharov, E.A. and Emelin, P.O., Relevance of using the equipment for control of fishing gear to reduce inaccuracy in evaluation of marine organisms abundance, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 2016, vol. 186, pp. 198–206. doi 10.26428/1606-9919-2016-186-198-206

Zakharov, E.A., Kruchinin, O.N., Mizurkin, M.A., and Safronov, V.A., Geometric parameters of the bottom trawl 27.1/24.4, and its possible errors in assessment of number of marine organisms, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 2013, vol. 174, pp. 284–292.

Zakharov, E.A., Kruchinin, O.N., and Shabelsky, D.L., Development and testing of algorithm for calculation of operating parameters of trawling systems, *Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.*, 2019, vol. 198, pp. 221–229. doi 10.26428/1606-9919-2019-198-221-229

Shuntov, V.P., Volvenko, I.V., Kulik, V.V., and Bocharov, L.N., *Makrofauna bentali Okhotskogo morya: tablitsy vstrechaemosti, chislennosti i biomassy. 1977–2010* (Benthic Macrofauna of the Okhotsk Sea: Occurrence, Abundance, and Biomass. 1977–2010), Shuntov, V.P. and Bocharov, L.N., eds, Vladivostok: TINRO-Tsentr, 2014.

Shuntov, V.P., Volvenko, I.V., Kulik, V.V., and Bocharov, L.N., *Makrofauna bentali severo-zapadnoi chasti Tikhogo okeana: tablitsy vstrechaemosti, chislennosti i biomassy. 1977–2008* (Benthic Macrofauna of the Northwestern Pacific: Occurrence, Abundance, and Biomass. 1977–2008), Shuntov, V.P. and Bocharov, L.N., eds, Vladivostok: TINRO-Tsentr, 2014.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A., and De Clerck, O., Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling, *Glob. Ecol. Biogeogr.*, 2018, vol. 27, no. 3, pp. 277–284. doi 10.1111/geb.12693

Breiman, L., Random forests, *Mach. Learn.*, 2001, vol. 45, no. 1, pp. 5–32. doi 10.1023/A:1010933404324

Kulik, V.V., Prants, S.V., Uleysky, M.Yu., and Budyansky, M.V., Lagrangian characteristics in the western North Pacific help to explain variability in Pacific saury fishery, *Fish. Res.*, 2022, vol. 252, 106361. doi 10.1016/j.fishres.2022.106361

Ricker, W.E., *Computation and Interpretation of Biological Statistics of Fish Populations*, Ottawa: Dep. Environ. Fish. Mar. Serv., 1975.

Sbrocco, E.J. and Barber, P.H., MARSPEC: ocean climate layers for marine spatial ecology, *Ecology*, 2013, vol. 94, no. 4, p. 979. doi 10.1890/12-1358.1

Thompson, P.L., Anderson, S.C., Nephin, J., Robb, C.K., Proudfoot, B., Park, A.E., Hagarty, D.R., and Rubidge, E., Integrating trawl and longline surveys across British Columbia improves groundfish distribution predictions, *Can. J. Fish. Aquat. Sci.*, 2022. doi 10.1139/cjfas-2022-0108

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O., Bio-ORACLE: a global environmental dataset for marine species distribution modelling, *Glob. Ecol. Biogeogr.*, 2012, vol. 21, no. 2, pp. 272–281. doi 10.1111/j.1466-8238.2011.00656.x

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

Svidetel'stvo № 6394 ot 01.08.2000 g. o registratsii v Gosudarstvennom registre bazy dannykh za № 0220006765 «Morskaja biologiya» (Certificate No. 6394 dated August 1, 2000 on registration in the State Register of the database No. 0220006765 “Marine biology”).

GEBCO Bathymetric Compilation Group 2020. The GEBCO_2020 Grid — a continuous terrain model of the global oceans and land, British Oceanographic Data Centre, National Oceanography Centre, NERC, UK. <https://doi.org/10.5285/a29c5465-b138-234d-e053-6c86abc040b9>. Cited July 20, 2022.

Baston, D., *exactextractr: Fast Extraction from Raster Datasets using Polygons*, R package version 0.7.2, 2021. <https://CRAN.R-project.org/package=exactextractr>. Cited January 27, 2022.

Olaya, V., *Module Real Surface Area*, SAGA-GIS Module Library Documentation (v2.1.3), 2004. *R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria, 2022, version 4.2.1. <https://www.R-project.org/>. Cited July 20, 2022.

Svidetel'stvo № 2017620882 ot 11.08.2017 g. o gosudarstvennoy registratsii bazy dannykh «Yarusnyy promysel» (Certificate No. 2017620882 dated August 11, 2017 on state registration of the “Longline Fishery” database).

Ogle, D.H., Doll J.C., Wheeler P., and Dinno A., *FSA: Fisheries Stock Analysis*, R package version 0.9.3. <https://github.com/fishR-Core-Team/FSA>. Cited July 26, 2022.

Поступила в редакцию 27.07.2022 г.

После доработки 11.10.2022 г.

Принята к публикации 21.11.2022 г.

The article was submitted 27.07.2022; approved after reviewing 11.10.2022; accepted for publication 21.11.2022