

Анализ сплочённости ледяного покрова в Карском море по данным радиометра MIRAS спутника SMOS с использованием методов машинного обучения

В. В. Тихонов^{1,3,4}, Д. Р. Катамадзе², Т. А. Алексеева^{3,1}, Е. В. Афанасьева^{3,1},
Ю. В. Соколова^{3,1}, И. В. Хвостов⁴, А. Н. Романов⁴

¹ *Институт космических исследований РАН, Москва, 117997, Россия*
E-mail: vtikhonov@asp.iki.rssi.ru

² *Московский государственный университет имени М. В. Ломоносова*
Москва, 119991, Россия
E-mail: DianaKatamadze@mail.ru

³ *Арктический и антарктический научно-исследовательский институт*
Санкт-Петербург, 199397, Россия
E-mail: taa@aari.ru

⁴ *Институт водных и экологических проблем СО РАН, Барнаул, 656038, Россия*
E-mail: romanov_alt@mail.ru

Представлены первые результаты исследования по использованию данных микроволнового радиометра MIRAS (*англ.* Microwave Imaging Radiometer using Aperture Synthesis) спутника SMOS (*англ.* Soil Moisture and Ocean Salinity) для определения сплочённости ледяного покрова в Карском море. По восьми областям в Карском море была исследована динамика яркостной температуры продукта SMOS L1C за 2022–2023 гг. Для определения сплочённости ледяного покрова в выбранных областях по данным продукта SMOS L1C были применены методы машинного обучения. В исследовании задействовано восемь моделей машинного обучения, использующих метод обучения с учителем. Обучение было выполнено по данным за 2022 г. Проверка качества прогнозов используемых моделей машинного обучения осуществлялась на основе данных за 2023 г. В результате анализа трёх метрик качества была выбрана оптимальная модель машинного обучения — XGBoost. Проведённые исследования показали перспективность использования машинного обучения для оценки состояния морского ледяного покрова Арктики по данным радиометра MIRAS спутника SMOS.

Ключевые слова: спутниковая микроволновая радиометрия, яркостная температура, сплочённость морского льда, видимый диапазон, инфракрасный диапазон, методы машинного обучения

Одобрена к печати: 29.11.2024

DOI: 10.21046/2070-7401-2024-21-6-344-355

Введение

Современный глобальный мониторинг морского льда в Северном Ледовитом океане осуществляется с помощью различных методов спутникового дистанционного зондирования, каждый из которых имеет свои преимущества и ограничения (Спутниковые..., 2011). Оптические методы позволяют получать высокодетализированные изображения морского ледяного покрова. Основным недостатком оптических методов заключается в их ограниченности условиями освещённости (полярная ночь) и метеорологической обстановкой (облачность, осадки, туман), что является существенным для арктического региона (Remote..., 2015). Радиолокаторы с синтезированной апертурой (РСА) обеспечивают возможность работы в любых погодных условиях и в любое время суток. Однако РСА требуют значительных энергозатрат, а дешифровка радиолокационных изображений не всегда бывает однозначной (Афанасьева и др., 2024; Заболотских и др., 2023).

Спутниковые микроволновые радиометры обеспечивают ежедневное полное покрытие полярных регионов Земли при любых метеорологических условиях в любое время суток. Основной характеристикой морского ледяного покрова, определяемой с помощью спутни-

ковой микроволновой радиометрии (СМР), является сплочённость морского льда (СМЛ). В настоящее время существует большое количество алгоритмов, используемых для определения СМЛ по данным СМР. Среди них нужно выделить шесть основных, это: NASA Team (NT), NASA Team 2 (NT2), ARTIST (*англ.* Arctic Radiation and Turbulence Interaction Study) Sea Ice (ASI), Bootstrap (BT), Bristol и TUD (*англ.* Technical University of Denmark). Кроме того, в последнее время всё чаще применяются гибридные алгоритмы, которые используют методики нескольких основных алгоритмов. Все эти алгоритмы основаны на данных о яркостной температуре $T_{\text{я}}$ на различных каналах или их комбинациях (поляризационные соотношения, градиентные соотношения, поляризационная разность и т. п.), полученных с космических радиометров SSMIS (*англ.* Special Sensor Microwave Imager/Sounder) или AMSR2 (*англ.* Advanced Microwave Scanning Radiometer 2). В настоящее время при определении сплочённости льда используются частоты от 18,7 ГГц и выше, а в усовершенствованном алгоритме BT также привлекается дополнительный канал на частоте 6,915 ГГц вертикальной поляризации. Однако все современные алгоритмы восстановления СМЛ по данным СМР имеют существенные погрешности, которые обусловлены сезонными, климатическими, географическими и другими факторами (Алексеева и др., 2022; Заболотских и др., 2023; Тихонов и др., 2016; Alekseeva et al., 2019; Wernecke et al., 2024).

В работе (Тихонов и др., 2023) авторами были представлены первые результаты исследований, показывающие, что $T_{\text{я}}$ на частоте 1,4 ГГц, принимаемая радиометром MIRAS (*англ.* Microwave Imaging Radiometer using Aperture Synthesis) спутника SMOS (*англ.* Soil Moisture and Ocean Salinity), очень чувствительна к изменениям состояния ледяного покрова Карского моря (сплочённость, таяние, замерзание). Обнаруженная зависимость позволила приступить к разработке методики анализа характеристик морского льда по данным радиометра MIRAS. В настоящей статье представлены результаты проведённых исследований по оценке сплочённости ледяного покрова Карского моря по данным радиометра MIRAS с использованием методов машинного обучения (МО).

Регион исследования и данные радиометра MIRAS

Регион исследования — акватория Карского моря (*рис. 1*, см. с. 346). В акватории было выбрано восемь ячеек (см. *рис. 1, таблицу*), по которым получены данные радиометра MIRAS спутника SMOS за 2022–2023 гг. Ячейки выбирались таким образом, чтобы охватить центральную часть акватории с юга на север, а также затронуть восточную область Карского моря, где в зимний период навигация практически отсутствует и ледяной покров не подвержен техногенному воздействию (Алексеева и др., 2024).

Выбранные ячейки исследования (см. *рис. 1*) и координаты их центра

Номер ячейки	Широта, ° с. ш.	Долгота, ° в. д.
p1	73,205	73,193
p2	74,007	76,092
p3	75,017	76,081
p4	76,021	76,144
p5	77,016	76,269
p6	77,948	75,902
p7	76,502	88,757
p8	77,583	99,751

В исследовании использовались данные продукта первого уровня SMOS L1C (*англ.* Level 1C). В этом продукте представлены значения яркостной температуры на горизонтальной и вертикальной поляризации для частоты 1,4 ГГц под углом зондирования 42,5° и с

пространственным разрешением 35×65 км. Данные продукта SMOS L1C привязаны к дискретной геодезической сетке DGG ISEA 4H9 (англ. Discrete Global Grid Icosahedral Snyder Equal Area). Линейный размер ячейки сетки составляет величину порядка 16 км, а площадь — около 195 км^2 . Подробнее о данных радиометра MIRAS спутника SMOS, а также об использовании в наших исследованиях продукта SMOS L1C изложено в работе (Тихонов и др., 2024).

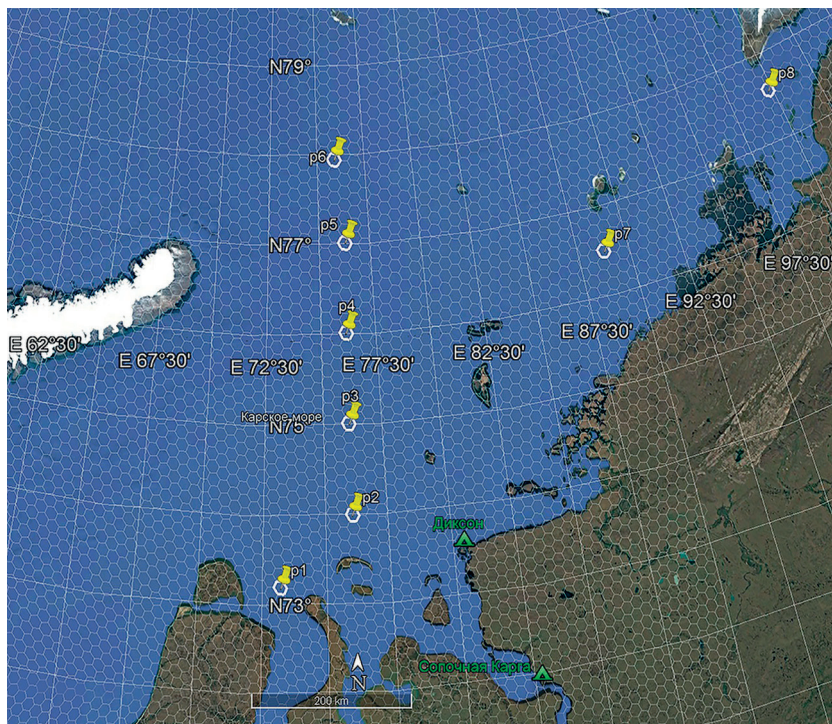


Рис. 1. Регион исследования. Белые шестиугольники — ячейки геодезической сетки DGG ISEA 4H9. Жирные белые шестиугольники, обозначенные жёлтыми кнопками с символами p1–p8, — исследуемые ячейки

Спутниковые данные о сплочённости морского льда, используемые для машинного обучения

Сплочённость ледяного покрова в выбранных ячейках была получена в результате комплексного дешифрирования спутниковых изображений видимого и инфракрасного (ИК) диапазонов, а также снимков, полученных с помощью РСА. Данные видимого диапазона в прибрежных районах арктических морей недоступны с ноября по февраль из-за полярной ночи. В тёмный период года их можно заменить ИК-данными. Данные РСА позволяют получать информацию о ледяном покрове вне зависимости от погодных условий и условий естественной освещённости (Спутниковые..., 2011). В работе для определения сплочённости морского льда был использован архив оптической и радиолокационной съёмки за 2022–2023 гг. Анализ ледовой обстановки выполнялся визуально с применением изображений РСА, а также снимков в видимом и ИК-диапазонах. Для этого использовались данные:

- спектрометрира MODIS (англ. Moderate Resolution Imaging Spectroradiometer) спутников Terra и Aqua (<https://worldview.earthdata.nasa.gov/>);
- радиометра VIIRS (англ. Visible Infrared Imaging Radiometer Suite) спутника Suomi NPP (англ. Suomi National Polar-orbiting Partnership). Данные получены с помощью выносного пункта приёма-передачи спутниковой информации Арктического и антарктического научно-исследовательского института в пос. Баренцбург на о. Западный Шпицберген;
- РСА Sentinel-1A (<http://seaice.dk/>).

В выбранных восьми ячейках, соответствующих геодезической сетке DGG ISEA 4H9 (см. рис. 1), были получены временные ряды общей сплочённости морского льда за 2022–2023 гг. Дискретность измерений определялась наличием безоблачных снимков в оптическом диапазоне, а также частотой радиолокационной съёмки анализируемых ячеек, но не чаще одного измерения в сутки. В зимние месяцы (с октября по май) помимо общей сплочённости были определены частные сплочённости однолетнего льда (>30 см), молодого льда (10–30 см) и ниласа (<10 см).

Модели машинного обучения

В ходе исследования использовались восемь моделей МО: Linear Regression (Linear), Ridge, LASSO (*англ.* Least Absolute Shrinkage and Selection Operator), ElasticNet (ElaNet), DecisionTreeRegressor (DTR), RandomForestRegressor (RFR), XGBoost (*англ.* EXtreme Gradient Boosting — XGB), KNeighborsRegressor (KNR). При этом устанавливалось соответствие информации о СМЛ и данных по яркостной температуре продукта SMOS L1C в каждой выбранной ячейке. В качестве атрибутов модели были выбраны четыре переменные: яркостная температура на горизонтальной (*англ.* Horizontal) T_H и вертикальной (*англ.* Vertical) T_V поляризации, их отношение T_H/T_V , а также поляризационное соотношение Pr , вычисляемое как $(T_V - T_H)/(T_V + T_H)$. Подобные комбинации $T_{\text{я}}$ на разных поляризациях часто используются при разработке алгоритмов определения СМЛ по данным СМР (Тихонов и др., 2016). Целевой переменной при МО являлась общая СМЛ. Все выбранные модели предсказывают значение целевой переменной на основе набора предикторов (T_H , T_V , T_H/T_V , Pr). Они используют методы обучения с учителем, что предполагает наличие размеченных данных, где целевая переменная данных для обучения известна. Основная задача таких моделей — найти зависимость между предикторами и целевой переменной, минимизируя ошибку прогнозирования. Каждая из моделей имеет свои методы регуляризации, настройки параметров и подходы к выбору наиболее релевантных признаков, что позволяет улучшить точность предсказаний и снизить риск переобучения. Это особенно важно при работе с большими объёмами данных (Флах, 2015).

Модели Linear Regression, Ridge, LASSO и ElasticNet относятся к классу линейных регрессионных методов. Общая черта всех этих моделей заключается в том, что они ищут оптимальные коэффициенты для атрибутов, при этом предполагая, что зависимость между атрибутами и целевой переменной можно описать линейной функцией (Рашка, Мирджалили, 2020).

Модель Linear Regression (линейная регрессия) — это один из самых простых и широко используемых методов в машинном обучении. Основная её задача заключается в подборе таких коэффициентов для атрибутов, чтобы минимизировать среднеквадратичную ошибку, т.е. разницу между фактическими и предсказанными значениями целевой переменной. Однако модель линейной регрессии может сталкиваться с проблемой переобучения, особенно если между атрибутами присутствует сильная корреляция. В таких случаях модель может плохо обобщать данные, что приводит к снижению качества прогнозов на новых данных (Mohri et al., 2018).

Модель Ridge (гребневая регрессия) — это модификация модели Linear Regression. Основное отличие заключается в применении L2-регуляризации. Регуляризация помогает сделать модель более устойчивой к шуму в данных и предотвращает чрезмерную подгонку под обучающую выборку, что позволяет избежать запоминания ненужных деталей, не имеющих отношения к общим закономерностям (Рашка, Мирджалили, 2020; Флах, 2015).

Модель LASSO — это разновидность линейной регрессии, использующая L1-регуляризацию. В отличие от модели Ridge, которая вводит штраф на сумму квадратов коэффициентов, LASSO добавляет штраф на сумму абсолютных значений коэффициентов атрибутов, что позволяет не только уменьшать величину коэффициентов, но и занулять некоторые из них. Это позволяет LASSO исключать неинформативные атрибуты, оставляя только самые важные. Однако её использование требует осторожности в ситуациях с высоко коррелированными

ными атрибутами или когда все признаки могут оказывать влияние на целевую переменную (Mohri et al., 2018).

Модель ElasticNet — это метод регрессии, который сочетает в себе элементы двух моделей: Ridge и LASSO. Модель ElasticNet строит линейную зависимость между атрибутами и целевой переменной, но использует сразу два типа регуляризации — L1 и L2; L1-регуляризация помогает занулять незначимые коэффициенты, тем самым отбирая только важные предикторы, а L2-регуляризация штрафует большие коэффициенты, что делает модель более стабильной и устойчивой. Это особенно полезно при большом количестве атрибутов, некоторые из которых могут быть сильно коррелированы друг с другом. Однако в ElasticNet появляется необходимость настройки двух параметров регуляризации — для L1 и L2, что может усложнять процесс подбора оптимальных значений. Также как и другие модели, использующие L1-, L2-регуляризацию, ElasticNet может не подходить для ситуаций, когда все признаки важны, поскольку она склонна снижать влияние менее значимых атрибутов (Рашка, Мирджалили, 2020; Mohri et al., 2018).

Модели DecisionTreeRegressor, RandomForestRegressor и XGBoost представляют класс моделей, основанных на методе дерева решений, которые используют иерархическую структуру для предсказания значений целевой переменной. Все три модели строят предсказания, основываясь на разбиении данных на подмножества, однако они отличаются по подходу к обучению и обработке данных (Рашка, Мирджалили, 2020; Ali et al., 2023; Mohri et al., 2018; Shalev-Shwartz, Ben-David, 2014).

Модель DecisionTreeRegressor (регрессия дерева решений), в отличие от линейных моделей, которые предполагают существование линейной зависимости между атрибутами и целевой переменной, разбивает данные на множество ветвей. На каждом шаге модель принимает решение на основе значений атрибутов, что позволяет ей постепенно разделять набор данных на более мелкие группы, улучшая точность предсказаний. Каждая «ветка» дерева соответствует определённому решению, основанному на значении атрибута, а «листья» дерева представляют собой итоговые предсказания. Такой подход позволяет модели эффективно улавливать сложные нелинейные зависимости между атрибутами и целевой переменной. Однако при этом DecisionTreeRegressor имеет риск создания излишне сложного дерева, которое точно подгоняет данные, но плохо работает на новых данных (Shalev-Shwartz, Ben-David, 2014).

Модель RandomForestRegressor (регрессия случайного леса) — это усовершенствованная версия метода дерева решений, которая использует ансамбль из нескольких моделей DecisionTreeRegressor для предсказания значений целевой переменной. Каждое дерево в случайном лесу строится на случайной подвыборке данных и случайном наборе атрибутов. Этот подход позволяет получать более точные и стабильные результаты прогнозирования и помогает избежать проблем с переобучением, когда одно дерево может слишком точно подстраиваться под обучающие данные. Основным недостатком RandomForestRegressor является вычислительная сложность, поскольку модель требует больше времени и ресурсов для обучения и выполнения из-за большого количества деревьев (Shalev-Shwartz, Ben-David, 2014).

Модель XGBoost, в отличие от RandomForestRegressor, строит множество независимых деревьев и объединяет их предсказания, используя последовательную схему. Модель строит деревья одно за другим, причём каждое новое дерево исправляет ошибки предыдущего; XGBoost делает это за счёт использования градиентного бустинга, при котором каждое новое дерево строится на основе градиента функции ошибки, минимизируя её шаг за шагом; XGBoost использует методы регуляризации (L1 и L2), что помогает снизить риск переобучения и улучшить обобщающую способность модели. Основным недостатком XGBoost — это сложность настройки. Модель имеет множество параметров, которые нужно тщательно подбирать, чтобы достичь оптимальной производительности (Ali et al., 2023).

Модель KNeighborsRegressor (регрессия на основе метода k -ближайших соседей) — это алгоритм машинного обучения, который предсказывает значения целевой переменной, основываясь на значениях ближайших «соседей» в пространстве признаков. Основным принципом работы KNeighborsRegressor заключается в том, что модель сначала определяет количество ближайших соседей (параметр k), а затем для каждого нового наблюдения находит k наиболее

похожих объектов в данных. После этого она предсказывает значение целевой переменной как среднее значение целевой переменной этих соседей. Производительность модели сильно зависит от выбора параметра k : если выбрать слишком маленькое значение, она может стать чувствительной к шуму, если слишком большое — может потерять способность к выявлению локальных особенностей данных (Shalev-Shwartz, Ben-David, 2014).

Эффективность моделей оценивалась с помощью различных метрик качества, каждая из которых предоставляет разнообразную информацию о точности прогнозов модели. Было использовано три метрики качества: R^2 , RMSE (*англ.* Root Mean Squared Error) и MAPE (*англ.* Mean Absolute Percentage Error).

Коэффициент детерминации (R^2) — это метрика качества, показывающая, насколько хорошо модель объясняет вариацию целевой переменной на основе атрибутов. Коэффициент детерминации вычисляется как единица минус отношение суммы квадратов остатков (разница между фактическими и предсказанными значениями) к сумме квадратов отклонений фактических значений от их среднего. Значение R^2 варьируется от 0 до 1. Если R^2 равен 1, это означает, что модель идеально предсказывает результаты. Однако высокая значимость R^2 не всегда гарантирует, что модель даёт хороший прогноз, так как метрика может быть чувствительна к числу атрибутов и не отражать наличие переобучения (Plevris et al., 2022).

Показатель RMSE — это метрика качества, применяемая в задачах регрессии для оценки точности модели; RMSE представляет собой квадратный корень из средней квадратичной ошибки. Значение RMSE показывает, насколько в среднем предсказания модели отклоняются от фактических величин. Высокое значение RMSE указывает на то, что модель имеет значительные ошибки в предсказаниях, в то время как низкое значение свидетельствует о высоком качестве модели. Однако RMSE имеет и свои ограничения. Во-первых, эта метрика измеряет только среднее отклонение предсказанных значений от фактических, не учитывая, как ошибки распределены по различным диапазонам данных. Во-вторых, RMSE может быть чувствительна к масштабу данных: если целевая переменная имеет широкий диапазон значений, RMSE может показывать большие значения, даже если относительные ошибки незначительны (Plevris et al., 2022).

Метрика MAPE измеряет среднюю абсолютную процентную ошибку между фактическими и предсказанными значениями целевой переменной и позволяет понять, насколько предсказания модели отличаются от реальных значений в относительном выражении. Однако эта метрика не может быть рассчитана для значений целевой переменной, равной нулю, так как это приведёт к делению на ноль, что делает MAPE неустойчивой в таких ситуациях (Kim, Kim, 2016).

Таким образом, каждая из метрик является необходимым инструментом для оценки качества модели регрессии. Однако для выбора оптимальной модели их следует использовать в сочетании друг с другом (Plevris et al., 2022).

Результаты и их обсуждение

В исходном наборе данных за 2022–2023 гг. по восьми ячейкам Карского моря информация о СМЛ была доступна не для всех записей яркостной температуры продукта SMOS L1C. Чтобы обеспечить корректное обучение моделей, отбирались только те данные, где СМЛ была определена. Для оценки эффективности моделей выполнено обучение на данных за 2022 г. Проверка качества прогнозов осуществлялась на основе информации за 2023 г.

На *рис. 2* (см. с. 350) представлены метрики качества для восьми используемых моделей. Из рисунка видно, что все обученные модели показывают высокие значения коэффициента детерминации (R^2), что свидетельствует об их способности точно реагировать на изменения в атрибутах и корректировать прогнозы (см. *рис. 2а*). Такие результаты вполне ожидаемы, поскольку яркостная температура хорошо коррелирует со сплочённостью морского льда (Тихонов и др., 2023). Метрика RMSE (см. *рис. 2б*) показывает, что наименее грубые ошибки допускает модель XGBoost. Метрика MAPE (см. *рис. 2в*) для XGBoost также демонстрирует

сравнительно хорошие результаты. Стоит заметить, что модель XGBoost обладает множеством гиперпараметров, при тонком подборе которых можно добиться значительных улучшений в качестве прогнозов. Она является наиболее современной и активно развивающейся из обученных моделей. Эта модель обладает регулярными обновлениями и расширениями возможностей, что делает её использование наиболее перспективным (Ali et al., 2023). Таким образом, при анализе метрик качества было выявлено, что XGBoost показывает результаты лучше, чем другие модели МО. Сравнение СМЛ за 2023 г. по выбранным восьми ячейкам (см. рис. 1), полученной по модели МО XGBoost с данными, определёнными методом визуального дешифрирования по спутниковым снимкам оптического диапазона и изображениям РСА (в дальнейшем будем называть эти данные реальной сплочённостью (PC)), представлены на рис. 3.

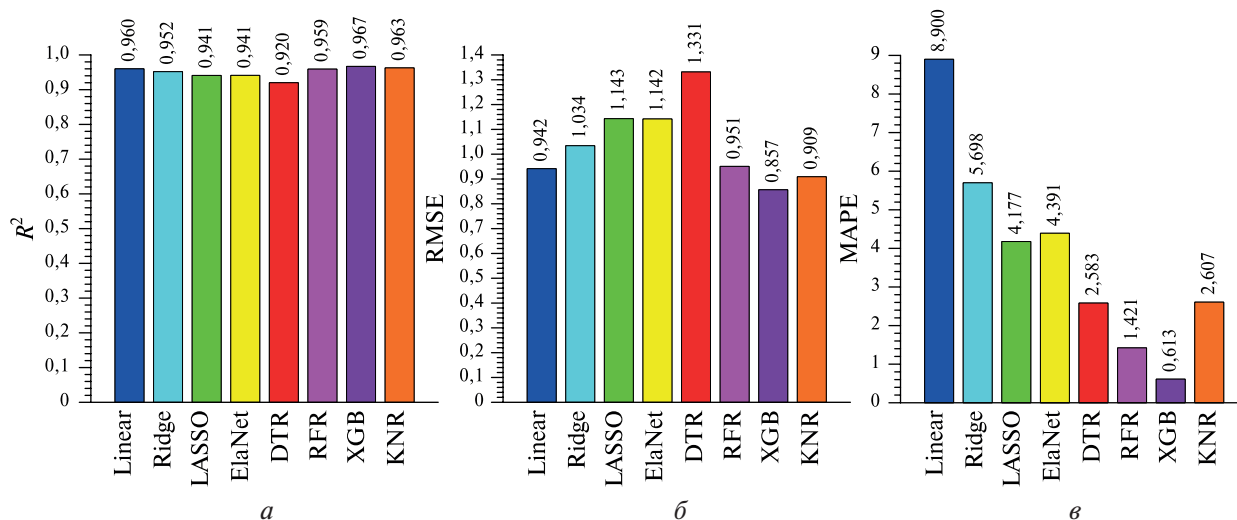


Рис. 2. Метрики качества для восьми моделей МО: а – R^2 ; б – RMSE; в – MAPE

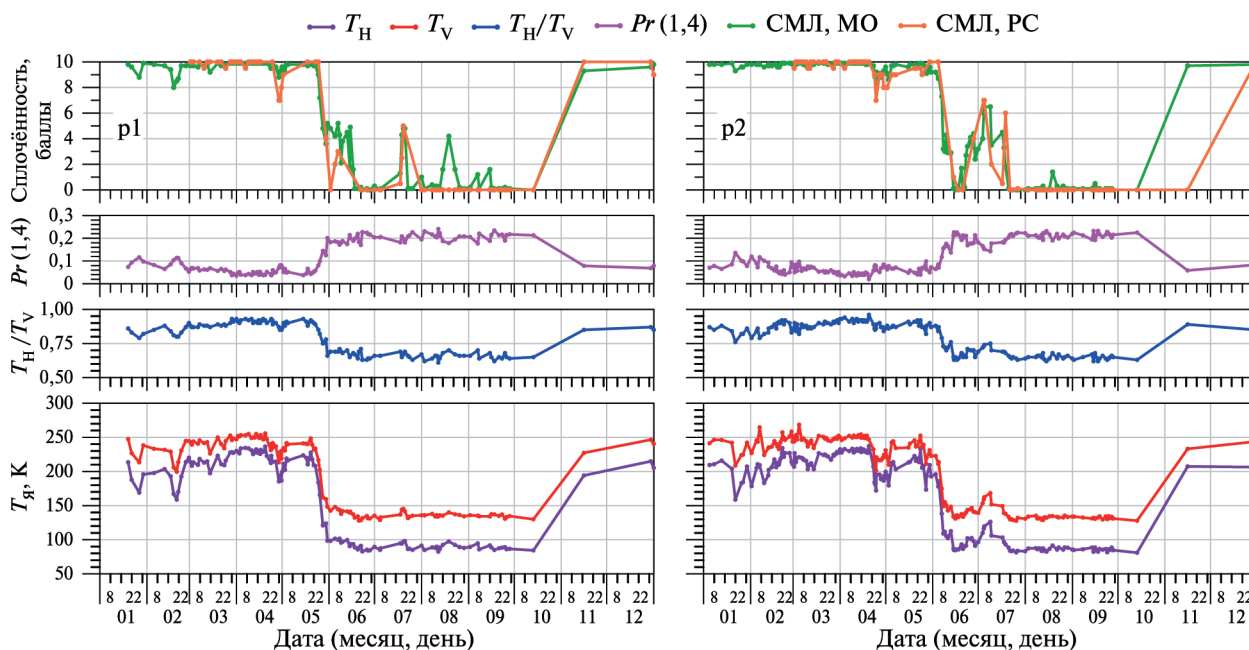
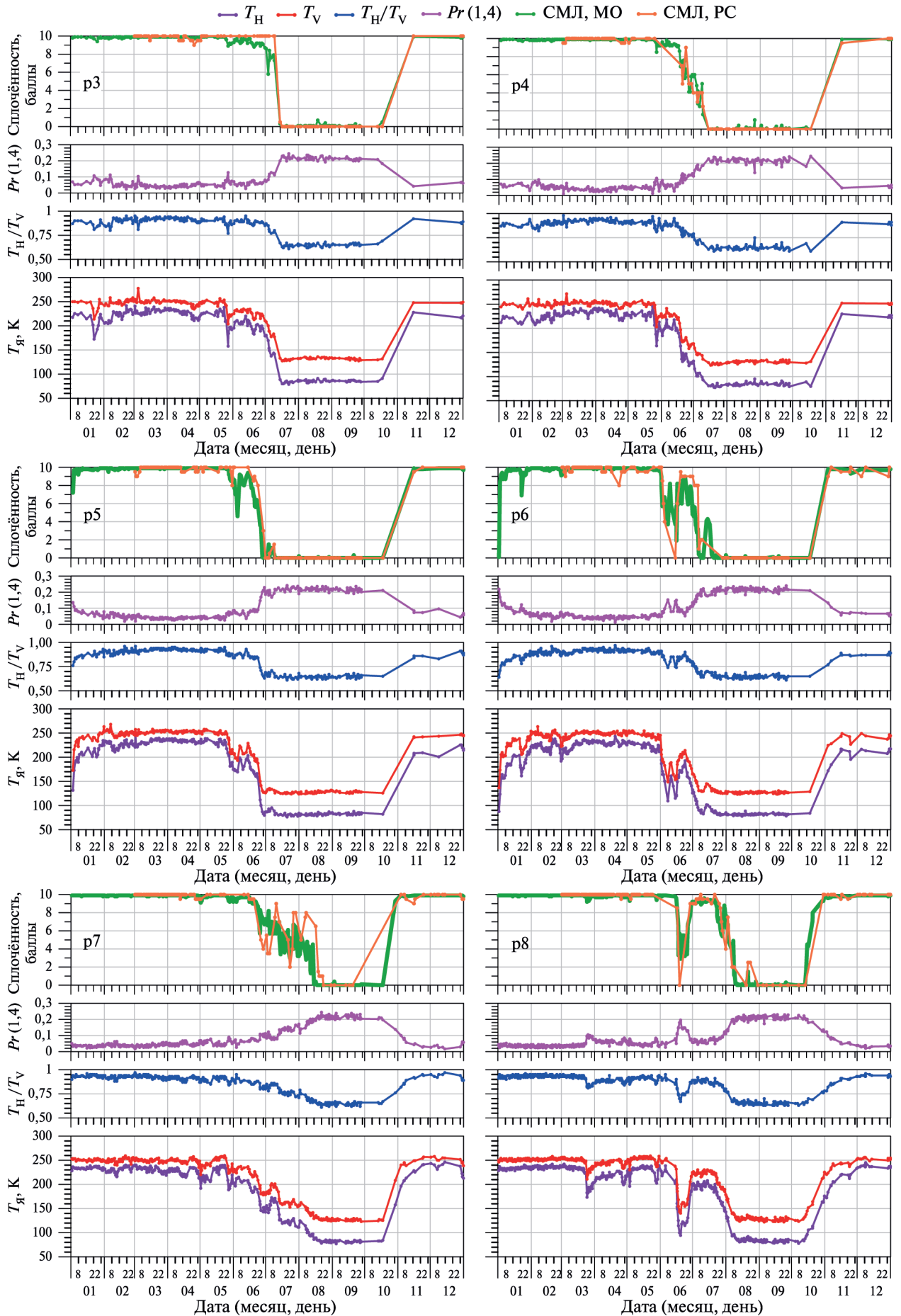


Рис. 3. Сезонные изменения T_H , T_V , T_H/T_V , Pr и СМЛ, полученной методом МО и методом визуального дешифрирования (РС), для ячеек p1, p2 за 2023 г.

Рис. 3. (Окончание, см. с.351) Сезонные изменения T_H , T_V , T_H/T_V , Pr и СМЛ, полученной методом МО и методом визуального дешифрирования (РС), для ячеек p3–p8 за 2023 г.



На рис. 3 для всех восьми ячеек Карского моря показана динамика за 2023 г.: T_H и T_V (нижние графики), T_H/T_V и Pr (второй и третий графики снизу соответственно), СМЛ (верхний график), полученной по модели МО XGBoost и РС. Рисунок 3 наглядно демонстрирует связь изменений T_H , T_V , T_H/T_V и Pr со СМЛ, полученной методом МО и по РС. Из рисунка видно, что СМЛ, определённая с помощью МО, хорошо соответствует данным по РС. Для оценки достоверности предсказанной сплочённости льда была построена диаграмма рассеяния СМЛ, полученная методом МО и РС (рис. 4). Анализ графика показывает, что рассеяние значений достаточно большое. Однако подобное рассеяние присутствует и на графиках для современных алгоритмов определения СМЛ по данным спутниковых радиометров AMSR2 и SSMIS (см., например, (Alekseeva et al., 2019; Kern et al., 2022)). Такие погрешности в получении СМЛ по данным СМР обусловлены многими факторами: сезонными, климатическими, географическими и т.д. (Тихонов и др., 2016; Alekseeva et al., 2019; Wernecke et al., 2024).

Высокий R^2 в данном случае (см. рис. 4) указывает, что СМЛ, определённая методом МО, может быть как занижена, так и завышена относительно значений СМЛ, полученных по спутниковым данным оптического диапазона и изображениям РСА. Однако из рис. 4 видно, что хуже всего метод МО работает при сплочённости льда 0 баллов. Завышение предсказанной сплочённости льда при её малых реальных значениях может быть связано с пенно-брызговой фазой на поверхности моря, которая возникает при ветре более 15 м/с. Присутствие таких образований на поверхности моря приводит к повышению яркостной температуры и, соответственно, к завышению предсказанных значений сплочённости ледяного покрова (Кутуза и др., 2016; Raizer, 2017).

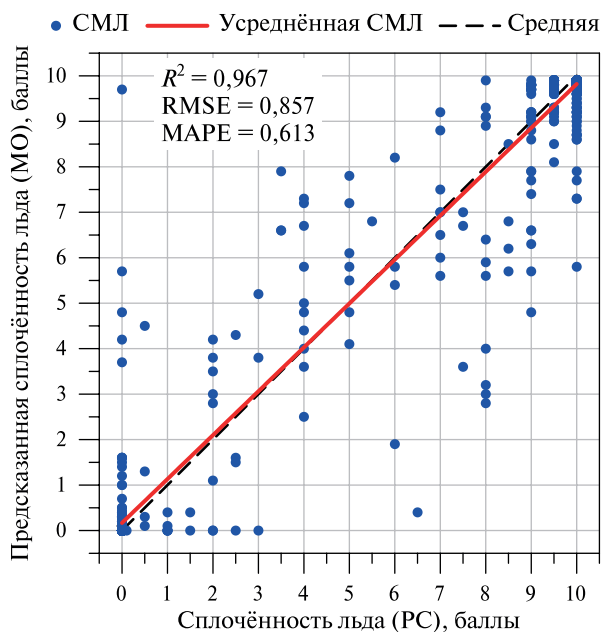


Рис. 4. Диаграмма рассеяния СМЛ, полученной по модели МО XGBoost и по РС

Заключение

Полученные результаты показывают перспективность использования данных радиометра MIRAS спутника SMOS для оценки состояния морского ледяного покрова Арктики. Для успешной реализации этих целей необходимо выполнить анализ причин завышения и занижения СМЛ, полученной методом МО, относительно реальных значений СМЛ. Поскольку радиометр MIRAS использует только частоту 1,4 ГГц, то дополнительное привлечение к исследованию высокочастотных каналов других спутниковых микроволновых радиометров должно привести к улучшению качества определения СМЛ. Связано это с различной

глубиной формирования излучения в снежно-ледовой толще для разной частоты излучения (Тихонов и др., 2020). Использование низкочастотного канала радиометра MIRAS позволит анализировать собственное микроволновое излучение более глубоких слоёв снежно-ледяного покрова, что в ряде случаев (оттепель, период таяния) может оказаться очень важным.

Совместное использование данных продукта SMOS L1C и данных радиометров AMSR2, SSMIS или МТВЗА-ГЯ (Модуль температурно-влажностного зондирования атмосферы, ГЯ — в память о Геннадии Яковлевиче Гуськове (1918–2002)) может способствовать значительному прогрессу в разработке дистанционных всепогодных методов контроля и диагностики морского ледяного покрова морей российской Арктики.

Работа выполнена при поддержке темы «Мониторинг» (гос. регистрация № 122042500031-8) (В. В. Тихонов, Т. А. Алексеева, Е. В. Афанасьева, Ю. В. Соколова), а также темы «Природные и природно-хозяйственные системы Сибири в условиях современных вызовов: диагностика состояний, адаптивные возможности, потенциал экосистемных услуг» (госзадание № FUFZ-2021-0007) (И. В. Хвостов, А. Н. Романов, В. В. Тихонов).

Литература

1. Алексеева Т. А., Соколова Ю. В., Афанасьева Е. В. и др. Влияние загрязнённости морского льда на ошибки в определении сплочённости в период таяния по данным спутниковой микроволновой радиометрии // Исслед. Земли из космоса. 2022. № 5. С. 30–46. <https://doi.org/10.31857/S0205961422050037>.
2. Алексеева Т. А., Сероветников С. С., Макаров Е. И. и др. Влияние интенсивного судоходства на изменение строения и динамики ледяного покрова в юго-западной части Карского моря // Проблемы Арктики и Антарктики. 2024. Т. 70. № 3. С. 323–337. <https://doi.org/10.30758/05552648-2024-70-3-323-337>.
3. Афанасьева Е. В., Соколова Ю. В., Тихонов В. В., Ермаков Д. М. Проблемы использования данных космической радиолокационной съёмки при решении задачи автоматизации ледового картирования // Современные проблемы дистанционного зондирования Земли из космоса. 2024. Т. 21. № 1. С. 9–27. DOI: 10.21046/2070-7401-2024-21-1-9-27.
4. Заболотских Е. В., Хворостовский К. С., Животовская М. А. и др. Спутниковое микроволновое зондирование морского льда Арктики: Обзор // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 1. С. 9–34. DOI: 10.21046/2070-7401-2023-20-1-9-34.
5. Кутуза Б. Г., Яковлев О. И., Данилычев М. И. Спутниковый мониторинг Земли: Микроволновая радиометрия атмосферы и поверхности. М.: ЛЕНАНД, 2016. 336 с.
6. Рашка С., Мирджалили В. Python и машинное обучение: Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2. 3-е изд. СПб.: ООО «Диалектика», 2020. 848 с.
7. Спутниковые методы определения характеристик ледяного покрова морей / под ред. В. Г. Смирнова. СПб.: ААНИИ, 2011. 240 с.
8. Тихонов В. В., Раев М. Д., Шарков Е. А. и др. Спутниковая микроволновая радиометрия морского льда полярных регионов: Обзор // Исслед. Земли из космоса. 2016. № 4. С. 65–84. DOI: 10.7868/S0205961416040072.
9. Тихонов В. В., Хвостов И. В., Романов А. Н. и др. Особенности собственного излучения Обской губы в L-диапазоне в период ледостава // Исслед. Земли из космоса. 2020. № 3. С. 59–76. DOI: 10.31857/S0205961420030070.
10. Тихонов В. В., Алексеева Т. А., Афанасьева Е. В. и др. О возможности определения сплочённости ледяного покрова арктических морей по данным спутника SMOS // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 6. С. 329–335. DOI: 10.21046/2070-7401-2023-20-6-329-335.
11. Тихонов В. В., Хвостов И. В., Романов А. Н., Шарков Е. А. Модель собственного микроволнового излучения устьевых областей арктических рек с учетом попадания в пиксель радиометра поверхности суши // Исслед. Земли из космоса. 2024. № 3. С. 16–29. DOI: 10.31857/S0205961424030026.
12. Флах П. Машинное обучение: Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2015. 400 с.
13. Alekseeva T., Tikhonov V., Frolov S. et al. Comparison of arctic sea ice concentrations from the NASA Team, ASI, and VASIA2 algorithms with summer and winter ship data // Remote Sensing. 2019. V. 11. No. 21. Article 2481. 31 p. <https://doi.org/10.3390/rs11212481>.

14. Ali Z. A., Abduljabbar Z. H., Tahir H. A. et al. Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review // Academic J. Nawroz University (AJNU). 2023. V. 12. No. 2. P. 320–334. DOI: 10.25007/ajnu.v12n2a1612.
15. Kern S., Lavergne T., Pedersen L. T. et al. Satellite passive microwave sea-ice concentration data set intercomparison using Landsat data // The Cryosphere. 2022. V. 16. No. 1. P. 349–378. <https://doi.org/10.5194/tc-16-349-2022>.
16. Kim S., Kim H. A new metric of absolute percentage error for intermittent demand forecasts // Intern. J. Forecasting. 2016. V. 32. No. 3. P. 669–679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>.
17. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of machine learning. 2nd ed. MIT Press, 2018. 498 p.
18. Plevris V., Solorzano G., Bakas N., Ben Seghier M. Investigation of performance metrics in regression analysis and machine learning-based prediction models // Proc. 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022). 5–9 June 2022, Oslo, Norway. 2022. 25 p. DOI: 10.23967/eccomas.2022.155.
19. Raizer V. Advances in passive microwave remote sensing of oceans. CRC Press, 2017. 262 p.
20. Remote sensing of the cryosphere / ed. Tedesco M. Oxford: John Wiley and Sons, 2015. 404 p.
21. Shalev-Shwartz S., Ben-David S. Understanding machine learning: From theory to algorithms. Cambridge: Cambridge University Press, 2014. 449 p.
22. Wernecke A., Notz D., Kern S., Lavergne T. Estimating the uncertainty of sea-ice area and sea-ice extent from satellite retrievals // The Cryosphere. 2024. V. 18. No. 5. P. 2473–2486. <https://doi.org/10.5194/tc-18-2473-2024>.

Analysis of ice concentration in the Kara Sea based on SMOS MIRAS data using machine learning methods

V. V. Tikhonov^{1,3,4}, D. R. Katamadze², T. A. Alekseeva^{3,1}, E. V. Afanasyeva^{3,1},
J. V. Sokolova^{3,1}, I. V. Khvostov⁴, A. N. Romanov⁴

¹ Space Research Institute RAS, Moscow 117997, Russia
E-mail: vtikhonov@asp.iki.rssi.ru

² Lomonosov Moscow State University, Moscow 119991, Russia
E-mail: DianaKatamadze@mail.ru

³ Arctic and Antarctic Research Institute, Saint Petersburg 199397, Russia
E-mail: taa@aari.ru

⁴ Institute for Water and Environmental SB RAS, Barnaul 656038, Russia
E-mail: romanov_alt@mail.ru

The paper presents the first results of a study on the use of data from Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) of Soil Moisture and Ocean Salinity (SMOS) satellite to determine ice concentration in the Kara Sea. The SMOS L1C brightness temperature dynamics in 2022–2023 has been studied for eight regions of the Kara Sea. Machine learning methods were used to determine ice concentration in the selected areas from the SMOS L1C data. The study involved eight machine learning models employing a supervised learning approach. The training was performed based on data for 2022. The quality of predictions by the machine learning models was verified using data for 2023. By analyzing the three quality metrics, the optimal machine learning model (XGBoost) was selected as capable of obtaining good results in determining ice concentration from SMOS MIRAS data. The conducted studies show the prospect of employing machine learning to assess the state of the Arctic sea ice cover using data from the MIRAS radiometer of the SMOS satellite.

Keywords: satellite microwave radiometry, brightness temperature, sea ice concentration, visible range, infrared range, machine learning methods

Accepted: 29.11.2024

DOI: 10.21046/2070-7401-2024-21-6-344-355

References

1. Alekseeva T. A., Sokolova J. V., Afanasyeva E. V. et al., The contribution of sea-ice contamination to inaccuracies in sea-ice concentration retrieval from satellite microwave radiometry data during the ice-melt period, *Izvestiya, Atmospheric and Oceanic Physics*, 2022, Vol. 58, No. 12, pp. 1470–1484, DOI: 10.1134/S0001433822120039.
2. Alekseeva T. A., Serovetnikov S. S., Makarov E. I. et al., The influence of heavy shipping traffic on the structure and dynamics of sea ice in the southwestern Kara Sea, *Arctic and Antarctic Research*, 2024; Vol. 70, No. 3, pp. 323–337 (in Russian), <https://doi.org/10.30758/05552648-2024-70-3-323-337>.
3. Afanasyeva E. V., Sokolova J. V., Tikhonov V. V., Ermakov D. M., Problems of using space-borne SAR data in solving the issue of ice charting automation, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2024, Vol. 21, No. 1, pp. 9–27 (in Russian), DOI: 10.21046/2070-7401-2024-21-1-9-27.
4. Zabolotskikh E. V., Khvorostovsky K. S., Zhivotovskaya M. A. et al., Satellite microwave remote sensing of the Arctic sea ice: Review, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2023, Vol. 20, No. 1, pp. 9–34 (in Russian), DOI: 10.21046/2070-7401-2023-20-1-9-34.
5. Kutuza B. G., Yakovlev O. I., Danilychev M. V. *Sputnikovyi monitoring Zemli: Microvolnovaya radiometriya atmosfery i poverkhnosti* (Satellite Earth monitoring: Microwave radiometry of the atmosphere and surface), Moscow: LENAND, 2016, 336 p. (in Russian).
6. Raschka S., Mirjalili V., Python machine learning: Machine learning and deep learning using Python, scikit-learn, and TensorFlow 2, 3rd ed., Birmingham: Packt Publishing, 2019, 770 p.
7. *Sputnikovye metody opredeleniya kharakteristik ledyanogo pokrova morei* (Satellite methods for determining sea ice cover characteristics), Smirnov V. G. (ed.), Saint Petersburg: Arctic and Antarctic Research Inst., 2011, 240 p. (in Russian).
8. Tikhonov V. V., Raev M. D., Sharkov E. A. et al., Satellite microwave radiometry of sea ice of polar regions: A review, *Izvestiya, Atmospheric and Oceanic Physics*, 2016, Vol. 52, No. 9, pp. 1012–1030, <https://doi.org/10.1134/S0001433816090267>.
9. Tikhonov V. V., Khvostov I. V., Romanov A. N. et al., Features of the intrinsic L-band radiation of the Gulf of Ob during the freeze-up period, *Izvestiya, Atmospheric and Oceanic Physics*, 2020, Vol. 56, No. 9, pp. 936–949, <https://doi.org/10.1134/S0001433820090236>.
10. Tikhonov V. V., Alekseeva T. A., Afanasyeva E. V. et al., On the possibility to determine the concentration of Arctic sea ice using SMOS satellite data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2023, Vol. 20, No. 6, pp. 329–335 (in Russian), DOI: 10.21046/2070-7401-2023-20-6-329-335.
11. Tikhonov V. V., Khvostov I. V., Romanov A. N., Sharkov E. A., A model of microwave emission from mouth regions of Arctic rivers providing for radiometer pixel land contamination, *Izvestiya, Atmospheric and Oceanic Physics*, 2024, Vol. 60, No. 9 (in press).
12. Flach P., *Machine Learning — The art and science of algorithms that make sense of data*, Cambridge: Cambridge University Press, 2012, 396 p.
13. Alekseeva T., Tikhonov V., Frolov S. et al., Comparison of arctic sea ice concentrations from the NASA Team, ASI, and VASIA2 algorithms with summer and winter ship data, *Remote Sensing*, 2019, Vol. 11, No. 21, Article 2481, 31 p., <https://doi.org/10.3390/rs11212481>.
14. Ali Z. A., Abduljabbar Z. H., Tahir H. A. et al., Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review, *Academic J. Nawroz University (AJNU)*, 2023, Vol. 12, No. 2, pp. 320–334, DOI: 10.25007/ajnu.v12n2a1612.
15. Kern S., Lavergne T., Pedersen L. T. et al., Satellite passive microwave sea-ice concentration data set intercomparison using Landsat data, *The Cryosphere*, 2022, Vol. 16, No. 1, pp. 349–378, <https://doi.org/10.5194/tc-16-349-2022>.
16. Kim S., Kim H., A new metric of absolute percentage error for intermittent demand forecasts, *Intern. J. Forecasting*, 2016, Vol. 32, No. 3, pp. 669–679, <https://doi.org/10.1016/j.ijforecast.2015.12.003>.
17. Mohri M., Rostamizadeh A., Talwalkar A., *Foundations of machine learning*, 2nd ed., MIT Press, 2018, 498 p.
18. Plevris V., Solorzano G., Bakas N., Ben Seghier M., Investigation of performance metrics in regression analysis and machine learning-based prediction models, *Proc. 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022)*, 5–9 June 2022, Oslo, Norway, 2022, 25 p., DOI: 10.23967/eccomas.2022.155.
19. Raizer V., *Advances in passive microwave remote sensing of oceans*, CRC Press, 2017, 262 p.
20. *Remote sensing of the cryosphere*, Tedesco M. (ed.), Oxford: John Wiley and Sons, 2015, 404 p.
21. Shalev-Shwartz S., Ben-David S., *Understanding machine learning: From theory to algorithms*, Cambridge: Cambridge University Press, 2014, 449 p.
22. Wernecke A., Notz D., Kern S., Lavergne T., Estimating the uncertainty of sea-ice area and sea-ice extent from satellite retrievals, *The Cryosphere*, 2024, Vol. 18, No. 5, pp. 2473–2486, <https://doi.org/10.5194/tc-18-2473-2024>.