

УЛУЧШЕНИЕ ТОЧНОСТИ ПРОГНОЗА СОСТОЯНИЯ МИРОВОГО ОКЕАНА ЗА СЧЕТ ОПТИМАЛЬНОГО РАСПОЛОЖЕНИЯ ИЗМЕРИТЕЛЕЙ

Н. А. Турко^{*1}, А. А. Лобашев², К. В. Ушаков^{1,3}, М. Н. Кауркин¹,
Л. Ю. Кальницкий⁴, С. В. Семин⁵, и Р. А. Ибраев^{1,3,6}

¹Институт океанологии им. П. П. Ширшова РАН, Москва, Россия

²Сколковский институт науки и технологий, Сколково, Россия

³Московский физико-технический институт, Долгопрудный, Россия

⁴Арктический и антарктический научно-исследовательский институт, Санкт-Петербург, Россия

⁵Институт проблем безопасного развития атомной энергетики РАН, Москва, Россия

⁶Институт вычислительной математики имени Г. И. Марчука РАН, Москва, Россия

* **Контакт:** Никита Андреевич Турко, nikitaturko@yandex.ru

В работе демонстрируется влияние расположения измерителей на точность оперативного прогноза состояния Мирового океана. Проводится сравнение различных методов расстановки измерителей, в том числе расстановка, полученная методом Concrete Autoencoder (CA). Для оценки влияния расположения датчиков на точность прогноза проводилось моделирование, имитирующее ситуацию, когда начальное состояние Мирового океана заметно отличается от реального. В эксперименте заменялись начальные условия для модели океана и льда, при этом атмосферный форсинг сохранялся из контрольного эксперимента. Затем производилось интегрирование модели с усвоением данных об «истинном» состоянии в точках расположения сенсоров. Результаты показали, что расстановка сенсоров, полученная при помощи методов глубокого обучения, превосходит в точности прогноза другие рассмотренные расстановки при сопоставимом числе сенсоров.

Ключевые слова: оперативный прогноз, Мировой океан, оптимальная расстановка измерителей, Concrete Autoencoder, усвоение данных

Цитирование: Турко, Н. А., А. А. Лобашев, К. В. Ушаков, М. Н. Кауркин, Л. Ю. Кальницкий, С. В. Семин, и Р. А. Ибраев Улучшение точности прогноза состояния Мирового океана за счет оптимального расположения измерителей // Russian Journal of Earth Sciences. — 2023. — Т. 23. — ES6005. — DOI: 10.2205/2023es000883 — EDN: UOMRDC

1. Введение

Увеличение точности прогнозирования состояния Мирового океана является критически важной задачей, которая во многом зависит от количества наблюдательных станций и их расположения. В рамках исследований по улучшению прогностических моделей понимание влияния числа измерителей и их расположения на точность прогноза имеет первостепенное значение. В данной работе мы исследуем влияние количества сенсоров и их расположения на точность прогноза состояния океана с использованием системы усвоения данных наблюдений. Задачей данного исследования является поиск оптимальной стратегии размещения датчиков, что будет способствовать эффективному планированию океанических станций сбора данных.

Актуальность данного исследования обусловлена двумя факторами. Во-первых, эффективное размещение датчиков позволяет ускорить обработку данных и вычисления при усвоении данных. Во-вторых, это сокращает необходимость в проведении обширных и затратных по времени наблюдений с большим числом измерительных станций. Как ранее было показано в работе [Кауркин и Ибраев, 2019], сокращение объема усваиваемых

<https://elibrary.ru/uomrdc>

Получено: 25 августа 2023 г.

Принято: 15 декабря 2023 г.

Опубликовано: 30 декабря 2023 г.



© 2023. Коллектив авторов.

данных наблюдений проекта АРГО [Wong et al., 2020] на 50–75 процентов не оказывает значительного влияния на точность прогноза при использовании метода оптимальной ансамблевой интерполяции. Аналогичные выводы были получены в работе [Turpin et al., 2016], где было показано, что усвоение только половины доступных наблюдений буев проекта АРГО приводит к незначительной деградации точности прогноза.

Классические методы размещения датчиков основаны на линейных приближениях и, обычно, используют информационную матрицу Фишера для количественной оценки ошибки между истинными и реконструированными полями [Nakai et al., 2021]. Данный класс методов использует сингулярное разложение матрицы ковариаций данных [Alonso et al., 2004; Krause et al., 2008; Kumar et al., 2014; Nagata et al., 2021; Nguyen et al., 2020; Saito et al., 2021], которое для вектора данных размера n масштабируется по количеству операций как $\mathcal{O}(n^3)$ для полного Singular Value Decomposition (SVD) разложения или как $\mathcal{O}(n^2)$ для поиска только первых нескольких сингулярных векторов [Li et al., 2019]. Однако эти методы сталкиваются с ограничениями при работе с многомерными данными, страдают от вычислительной неэффективности, и восстановленные физические поля часто не удовлетворяют физическим ограничениям.

В последние годы методы глубокого обучения стали чаще применяться в анализе климатических данных и системах оперативного прогнозирования, что позволяет проводить анализ с высоким разрешением и улучшать краткосрочные (с заблаговременностью до 3 суток) прогнозы состояния атмосферы. В то время как модели глубокого обучения с архитектурой трансформера демонстрируют многообещающие результаты для краткосрочного прогнозирования погоды [Pathak et al., 2022], классические численные модели по-прежнему превосходят их для среднесрочных (3–7 суток) и долгосрочных (от 8 суток до сезона) прогнозов.

Появление распределения Gumbel-softmax и континуально-дискретного (concrete) распределения [Jang et al., 2016; Maddison et al., 2016] облегчило оптимизацию параметров дискретных распределений вероятностей благодаря возможности использования более эффективных методов на основе градиентного спуска. Вскоре было предложено несколько методов, применяющих распределение Gumbel-softmax для оптимального размещения датчиков, таких как Concrete Autoencoder (CA) [Abid et al., 2019], глубокое вероятностное семплирование [Huijben et al., 2020] и динамический выбор признаков с максимизацией взаимной информации [Covert et al., 2023]. Эти методы продемонстрировали преимущества по сравнению с классическими методами на основе сингулярного разложения, а именно: повышенную эффективность использования памяти и отсутствие необходимости в дорогостоящем вычислении ковариационной матрицы данных.

В то же время исследования по оценке качества размещения датчиков с использованием статистических моделей ограничивались относительно простыми системами с сетками низкого разрешения [Clark et al., 2019; Manohar et al., 2018; Sun et al., 2019]. В этих исследованиях качество расположения датчиков оценивалось по норме ошибки восстановления мгновенного поля по данным наблюдений на тестовой выборке. Наше исследование выходит за эти ограничения, рассматривая более сложную систему океан–лед и используя более продвинутый метод оценки качества расположения сенсоров, в котором восстановленные поля удовлетворяют законам сохранения и термогидродинамическим уравнениям. Качество расстановки оценивается по результатам численного эксперимента с усвоением данных синтетических наблюдений методом оптимальной ансамблевой интерполяции в процессе интегрирования модели глобальной циркуляции океана с пространственным разрешением $0,25 \times 0,25$ градуса. Данный подход позволяет нам существенно углубить и уточнить анализ оптимальности размещения датчиков.

Преыдущие исследования [Abbasi et al., 2018; Turpin et al., 2016] показали, что усвоение данных о температуре, даже с одного горизонтального слоя, может улучшить точность прогноза. Таким образом, наш подход включает усвоение измерений только из одного горизонта океана, а именно в приповерхностном слое, чтобы минимизировать влияние атмосферного воздействия.

Использование исключительно модели Мирового океана высокого разрешения для решения задачи размещения датчиков методом прямого комбинаторного поиска нецелесообразно из-за высокой вычислительной мощности и большой продолжительности одного эксперимента. Методы же глубокого обучения сами по себе не позволяют во всех деталях учесть особенности физических процессов, законы сохранения и соответствие предсказаний уравнениям гидродинамики. В отличие от методов глубокого обучения, в процессе восстановления поля по разреженным измерениям с использованием модели океана и системы усвоения, мы получаем удовлетворяющие уравнениям гидродинамики полные трехмерные поля температуры и солености, трехмерные поля горизонтальных компонент скорости, а также двумерные поля сплоченности и толщины льда. Эта работа представляет новый, более глубокий взгляд на вопрос оптимального размещения датчиков для улучшения точности прогноза состояния океана.

Чтобы снизить вычислительную нагрузку, связанную с прямым комбинаторным поиском при моделировании Мирового океана с высоким разрешением, мы предлагаем гибридный подход. Наша методология сочетает статистическую обработку данных ретроспективного гидродинамического моделирования с использованием моделей нейронных сетей для захвата нелинейных зависимостей и последующую оценку эффективности расположения датчиков в ходе прогноза по модели общей циркуляции океана. Этот гибридный подход дает физически обоснованные оптимальные координаты датчика в пределах практических временных рамок. Кроме того, он демонстрирует линейную масштабируемость с размером вычислительной области и устраняет необходимость вычисления ковариационной матрицы численного решения. Таким образом, это оказывается выгодным для таких приложений, как модели глобальной циркуляции океана при разрешении 0,25 или 0,1 градуса, где ковариационная матрица может иметь векторные размеры порядка от 10^{12} до 10^{16} .

В контексте данного исследования мы оцениваем эффективность размещения датчиков, применяя процедуру усвоения данных в глобальной модели океана. Одним из основных вопросов, которым мы стремимся уделить внимание, является возможность экстраполяции расположения датчиков, определенного в процессе обучения нейронной сети на исторических данных [Lobashev et al., 2023], на глобальную модель прогноза океанской циркуляции [Ushakov and Ibrayev, 2018]. В частности, мы рассматриваем вопрос о том, будет ли конфигурация датчиков, обеспечивающая минимальную ошибку восстановления поля нейронной сетью, также оптимальной для прогнозирования состояния океана путем интегрирования модели с усвоением данных, полученных от этих датчиков.

В данном исследовании проводится сравнение метода Concrete Autoencoder с несколькими базовыми методами расстановки измерителей. Рассматриваемый вариант метода CA [Lobashev et al., 2023; Turko et al., 2022] включает в себя два этапа. На первом этапе проводится оценка пространственной вариабельности физического поля, аппроксимируя его информационную энтропию с использованием нейронной сети Conditional Pixel CNN. На втором этапе полученная энтропия применяется для инициализации начального расположения датчиков. Затем это расположение оптимизируется с помощью архитектуры CA, которая позволяет одновременно минимизировать количество датчиков и максимизировать точность восстановления. Мы исследуем расположение сенсоров, полученное в работе [Lobashev et al., 2023]. После исключения сенсоров, попадающих на сушу, мы усваиваем показания 1160 сенсоров. Сравнение производится с фиксированными расстановками, взятыми от свободно дрейфующих станций проекта АРГО за первые сутки и за первые четверо суток с момента начала эксперимента, а также с расстановкой в соответствии с расположением буев проекта АРГО, меняющимся каждый день. Число сенсоров равняется соответственно 302, 1299 и ~280–380. Также для сравнения производится регулярная расстановка по трехполярной локально-ортогональной решетке [Murray, 1996]. Всего рассматриваются четыре варианта регулярной расстановки с 1301, 3379, 6792 и 13 582 сенсорами.

В [разд. 1](#) содержится введение и обзор литературы. В [разд. 2](#) мы описываем постановку эксперимента. В [разд. 3](#) приводятся результаты. Наконец, [разд. 4](#) содержит обсуждение результатов и выводы.

2. Постановка численных экспериментов

Идея эксперимента. Для оценки влияния расположения датчиков на точность прогноза проводилось моделирование, имитирующее ситуацию, когда начальное состояние Мирового океана заметно отличается от реального. В эксперименте заменялись начальные условия для модели океана и льда, при этом атмосферный форсинг сохранялся из контрольного эксперимента. Затем производилось интегрирование модели с усвоением данных об истинном (взятом из контрольного эксперимента) состоянии поля температуры в приповерхностном слое в точках расположения сенсоров. Основная идея заключается в выборе такого расположения датчиков, которое позволит достичь максимального совпадения с состоянием контрольного эксперимента за минимальное число итераций усвоения. По скорости приближения текущего состояния модели океана к целевому и уменьшению ошибки можно будет сделать вывод о влиянии способа расстановки и количества измерителей на точность прогноза. Таким образом, в качестве основной метрики, характеризующей качество расстановки измерителей, мы используем среднюю точность прогноза состояния Мирового океана в приповерхностном слое. Модель океана совместно с системой усвоения выступают в таком случае как метод восстановления трехмерного поля температуры по пространственно разреженным данным измерений.

Способы расстановки измерителей. В большинстве современных систем прогнозирования, таких как участники проекта OceanPredict, ежедневно используются данные наблюдений, полученные с помощью измерительных буев проекта АРГО. Общее количество буев достигает приблизительно четырех тысяч. При этом ежедневно передается около 280–380 профилей температуры. В качестве базового варианта расстановки мы использовали координаты буев проекта АРГО за один день (эксперимент s23) и за четыре последовательных дня (эксперимент s24) для получения синтетических данных наблюдений. Также был проведен эксперимент с реальным положением буев проекта АРГО (эксперимент s25). Для оценки влияния количества датчиков на точность прогноза мы также исследовали расположение на равномерной (трехполярной локально-ортогональной) сетке из одной (эксперимент s36), трех (эксперимент s37), шести (эксперимент s38) и тринадцати тысяч (эксперимент s39) датчиков.

Основной метод размещения датчиков, который мы исследовали, состоит из двух этапов. На первом этапе проводится оценка пространственной вариабельности физического поля, аппроксимируя его информационную энтропию с помощью нейросети Conditional PixelCNN. На втором этапе энтропия используется для инициализации начального расположения датчиков, которое далее оптимизируется с помощью архитектуры Concrete Autoencoder, позволяющей одновременно минимизировать количество датчиков и максимизировать точность восстановления поля. Этот эксперимент обозначен как s44.

Для сравнения был проведен эксперимент s00 без усвоения данных, чтобы оценить влияние атмосферного форсинга на точность прогноза. Список проведенных экспериментов и их описания представлены в [табл. 1](#).

Постановка эксперимента. В данном исследовании мы использовали совместную модель Мирового океана с ледовым покровом. Модель океана разработана Институтом вычислительной математики РАН и Институтом океанологии РАН [[Кальницкий и др., 2020](#); [Ushakov and Ibrayev, 2018](#)], в качестве модели льда используется CICE 5.1 [[Hunke et al., 2015](#)]. Совмещение моделей (каплинг) выполнено с помощью платформы CMF [[Fadeev et al., 2018](#); [Kalmykov et al., 2018](#)]. Атмосферное воздействие осуществлялось с использованием данных реанализа ERA5 [[Hersbach et al., 2020](#)]. Усвоение данных

Таблица 1. Список экспериментов и соответствующее число сенсоров

Эксперимент	Метод расположения	Измерений в день
s00	без измерений	0
s23	АРГО, 1 день	302
s24	АРГО, 4 дня	1299
s25	АРГО, динамическое размещение	~ 280–380
s36	Регулярно – $5,25^\circ \times 6,25^\circ$	1301
s37	Регулярно – $5^\circ \times 2,5^\circ$	3379
s38	Регулярно – $2,5^\circ \times 2,5^\circ$	6792
s39	Регулярно – $2,5^\circ \times 1,5^\circ$	13 581
s44	Concrete Autoencoder	1160
a01	Реанализ с усвоением TS профилей, сплоченности льда, аномалии уровня поверхности океана	~11 200

проводилось методом оптимальной ансамблевой интерполяции EnOI [Kaurkin et al., 2016a,b] один раз в модельные сутки в 12:00. Параметры усвоения были подобраны так, чтобы корректировка модельного решения была максимально допустимой по соображениям устойчивости. Здесь требуется пояснение. Усвоение можно рассматривать как корректировку модельного решения в соответствии с данными наблюдений. В рассматриваемых экспериментах при усвоении корректировалось только поле температуры, а остальные физические поля не изменялись (соленость, компоненты скорости, уровень поверхности океана, сплоченность и толщина льда). Другими словами, при усвоении возмущалось исключительно поле температуры, что приводило к физическому несоответствию с остальными полями. Адаптация к возмущениям происходила в процессе дальнейшего интегрирования модели океана. Возмущение, внесенное при усвоении, можно воспринимать как шоковое изменение состояния, которое при слишком больших амплитудах может приводить к численной неустойчивости и развалу модели. Таким образом, максимизация возмущения при усвоении позволила сократить время выхода точности на квазистационарный режим в экспериментах. Но, с другой стороны, из условия численной стабильности работы модели океана после процедуры усвоения мы получили ограничение на количество измерителей снизу. Так, в экспериментах (не представленных в работе) с усвоением данных менее чем с ~ 250 измерителей модель разваливалась из-за шоков при усвоении. Более подробное описание использованной конфигурации приведено в работе [Lobashev et al., 2023].

Контрольный эксперимент a01 проводился с моделью Мирового океана, интегрированной в период с 2019-01-01 по 2020-12-01, начиная с климатических значений температуры и солености WOA2013 [Boyer et al., 2013]. С 2020-01-01 мы начали усвоение данных наблюдений: аномалии уровня поверхности океана [Desai, 2016], сплоченности льда [Lavergne et al., 2019] и профилей буев проекта АРГО [Wong et al., 2020], с использованием метода EnOI.

Для исследования влияния расположения датчиков на точность прогноза Мирового океана мы провели серию экспериментов (серия s**), в которых были одинаковы начальные и граничные условия. При этом начальное состояние отличалось от начального состояния в контрольном эксперименте, а именно, было возмущено за счет сдвига по времени на один год назад. В экспериментах варьировались количество датчиков и метод их расположения. В табл. 1 представлены названия экспериментов с указанием способов расположения датчиков и количества усваиваемых измерений.

В экспериментах серии s** в качестве начальных условий было выбрано состояние океана на 11 сентября 2019 года. Эксперименты выполнялись с атмосферным форсингом, начавшимся 11 сентября 2020 года и взятым из контрольного эксперимента a01.

Последующее усвоение данных производилось один раз в сутки, с дальнейшим интегрированием модели на 24 часа вперед, после этого вычислялась точность прогноза по сравнению с контрольным экспериментом. Для усвоения использовались синтетические данные наблюдений температуры в приповерхностном слое океана, полученные без добавления погрешностей из текущих полей температуры контрольного эксперимента a01, начиная также с 11 сентября 2020, года без усвоения альтиметрии и сплоченности льда. Точность суточного прогноза температурных полей Мирового океана оценивалась по сравнению с данными контрольного эксперимента a01 во всей трехмерной расчетной области, при этом использовались расчеты средней и среднеквадратичной ошибок [Ryan et al., 2015] по формулам, представленным в работах [Lobashev et al., 2023; Turko et al., 2022].

3. Результаты

Результаты проведенных экспериментов представлены в табл. 2. Оценка точности прогноза проводилась на основе двух основных метрик: смещения, или средней ошибки (Bias), и среднеквадратичной ошибки (Root Mean Square Error, RMSE). Bias является мерой систематической ошибки, присутствующей в прогнозе, в то время как RMSE является стандартной мерой различия между предсказанными и наблюдаемыми значениями, т.е. мерой точности, учитывающей как систематическую, так и случайную погрешности. Данные получены из десяти различных экспериментов, каждый из которых обозначен от s00 до s44, и количество используемых датчиков варьируется от 0 до 13 582. Значения Bias и RMSE представлены для каждого эксперимента за три последовательных дня (День 0, День 1 и День 2), начиная со дня старта, а также указаны средние значения за первые 20 дней для обеих метрик. Усреднение за 20 дней выбрано исходя из того, что после этого периода значения средней и среднеквадратичной ошибок выходят на квазистационарный уровень. Расчет ошибок проводился для слоя Мирового океана от поверхности до глубины 100 метров. В этом слое наиболее выражен эффект усвоения данных (рис. 10–14). С другой стороны, этот слой представляет интерес с точки зрения восстановления динамики Мирового океана. Верхний перемешанный слой, определяющий взаимодействие океана с атмосферой, летом характеризуется толщинами до 100 метров и возрастает в зимний период [Кошляков и Тараканов, 2014; Sallée et al., 2021]. Значения ошибок указаны до 3-го знака после запятой. При этом были проведены повторные эксперименты с одними и теми же условиями, которые показали, что численная ошибка модели океана и системы усвоения не влияет на значения точности прогноза вплоть до 4-го знака после запятой.

Таблица 2. Точность прогноза, усредненная по верхним 100 метрам, по сравнению с контрольным экспериментом a01 за первые 20 дней

Эксперимент	Сенсоры	Bias, °C				RMSE, °C			
		День 0	День 1	День 2	Среднее	День 0	День 1	День 2	Среднее
s00	0	-0,195	-0,224	-0,180	-0,222	1,139	1,173	1,197	1,128
s23	302	-0,195	-0,280	-0,192	-0,213	1,139	1,203	1,240	1,175
s24	1299	-0,195	-0,288	-0,217	-0,223	1,139	1,221	1,227	1,170
s25	~280–380	-0,195	-0,280	-0,262	-0,146	1,139	1,203	1,253	1,226
s36	1306	-0,195	-0,144	-0,053	-0,078	1,139	1,162	1,222	1,149
s37	3379	-0,195	-0,164	-0,065	-0,102	1,139	1,132	1,217	1,138
s38	6792	-0,195	-0,164	-0,071	-0,106	1,139	1,135	1,206	1,127
s39	13 581	-0,195	-0,160	-0,066	-0,100	1,139	1,136	1,210	1,127
s44	1160	-0,195	-0,162	-0,070	-0,100	1,139	1,160	1,194	1,139

Эксперимент s00, в котором не проводилось усвоение данных синтетических наблюдений, имеет большую среднюю ошибку (Bias) и одно из наименьших значений RMSE. Это объясняется отсутствием характерных для работы системы усвоения «шоков», которые присутствовали во всех остальных экспериментах серии s**, в силу настройки «агрессивного» усвоения.

Результаты численного моделирования в серии экспериментов s3* указывают на то, что при достижении порога в примерно три тысячи сенсоров дальнейшее увеличение числа измерительных устройств не влечет за собой существенного улучшения точности прогноза. Например, эксперименты, включающие усвоение данных с тремя, шестью и тринадцатью тысячами измерений в день, показывают минимальные различия в точности прогноза, с точки зрения как средней, так и среднеквадратичной ошибок. Эти данные подкрепляют гипотезу о возможности удаления примерно половины датчиков из набора s38 без ущерба для систематической и среднеквадратичной ошибок, что согласуется с результатами работ [Кауркин и Ибраев, 2019; Turpin et al., 2016], в которых было показано, что удаление примерно половины данных проекта АРГО при усвоении не приводит к существенной деградации точности прогноза.

Интересно, что при использовании метода Concrete Autoencoder для расположения 1160 сенсоров эксперименты демонстрируют аналогичное снижение средней ошибки прогноза, как и в случае регулярного расположения сенсоров в 3, 6 и 13 раз большим количеством. По значениям средней ошибки и RMSE расстановка s44 (Bias $-0,100$ и RMSE 1,139) наиболее близка к регулярной расстановке 3379 сенсоров в эксперименте s37 (Bias $-0,102$ и RMSE 1,138). При этом среднеквадратичная ошибка оказывается заметно ниже по сравнению с использованием регулярной или АРГО схемы расположения тысячи измерителей. Эти результаты говорят о том, что при оптимальном расположении сенсоров можно добиться увеличения точности прогноза без увеличения числа сенсоров.

При сравнении экспериментов s24, s36 и s44 с приблизительно одной тысячей сенсоров наибольшее значение как средней, так и среднеквадратичной ошибок имеет расположение буев проекта АРГО. В экспериментах s36 и s44 значение средней ошибки снижается уже после второго усвоения и остается низким в течение следующих 20 дней. Наименьшее значение RMSE имеет расстановка s44, полученная методом СА.

В серии экспериментов s2* наибольшее значение RMSE имеет динамическая расстановка, соответствующая реальному положению буев проекта АРГО. При этом динамическая расстановка имеет наименьшее значение Bias за счет того, что в среднем за 10 дней всплывает около четырех тысяч буев, приблизительно равномерно распределенных по Мировому океану, за исключением Арктики. Сравнивая между собой статические расстановки s23 и s24, можно заметить, что несмотря на более чем четырехкратное увеличение числа датчиков (с 302 до 1299) в эксперименте s24, средние значения Bias и RMSE демонстрируют незначительные изменения по сравнению с экспериментом s23.

Данные указывают на улучшение точности прогноза с увеличением количества датчиков, хотя эта связь не всегда строго линейна. Также они показывают, что расстановка Concrete Autoencoder может обеспечить низкое значение как Bias, так и RMSE по сравнению с другими вариантами расстановки с близким числом сенсоров.

Графики точности от времени. На рис. 1–8 представлены графики точности прогноза на 1 сутки по сравнению с контрольным экспериментом на разных горизонтах для экспериментов серии s** с усвоением. Можно заметить, что расположение сенсоров оказывает наибольшее влияние на точность прогноза в течение первых 20–30 итераций усвоения.

Анализ рис. 1–3 показывает, что все три эксперимента имеют схожую скорость сходимости RMSE в первые 20 дней. Однако в случае реального расположения АРГО средняя ошибка уменьшается быстрее, чем в сценариях s23 и s24 со статической расстановкой. Кроме того, наибольший эффект от увеличения числа датчиков с 302 до

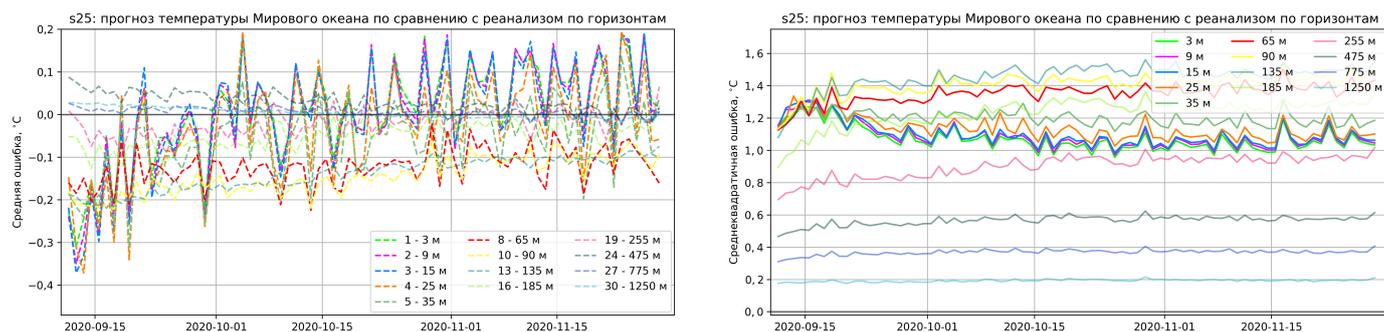


Рис. 1. Точность прогноза на 1 сутки для реального расположения ~280–380 сенсоров проекта АРГО на различных уровнях по сравнению с контрольным экспериментом. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

1299 наблюдается при рассмотрении графика RMSE через 20–25 дней после старта эксперимента. В то время как в эксперименте s23 наблюдаются признаки небольшого увеличения RMSE ближе к концу симуляции, график RMSE в эксперименте s24 достигает квазистационарных значений.

Анализ графиков, представленных на рис. 3 и 4, подтверждает важность использования алгоритмов оптимального размещения датчиков. В частности, сравнение сценариев s24 и s44 демонстрирует, что среднее значение ошибки (Bias) может быть улучшено с использованием меньшего количества датчиков при оптимальном размещении. Заметным является более быстрое уменьшение средней ошибки в эксперименте s44

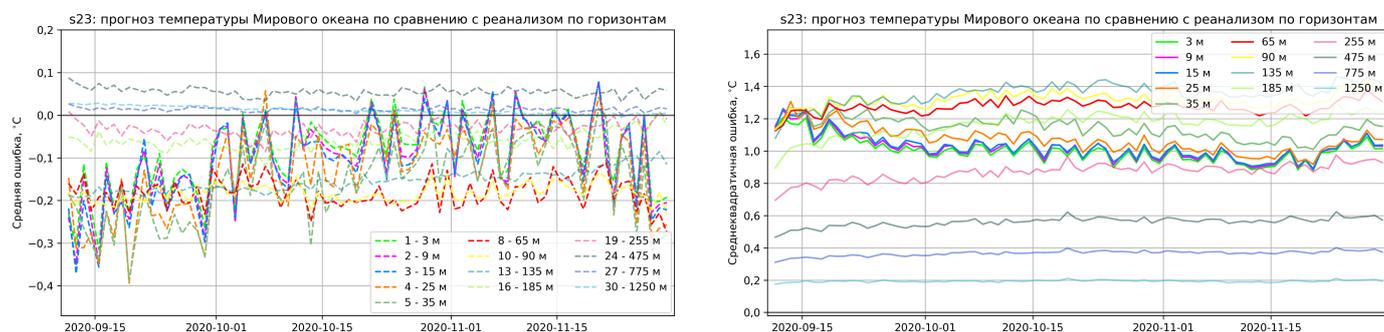


Рис. 2. Точность прогноза на 1 сутки для фиксированного расположения 302 сенсоров проекта АРГО на различных уровнях по сравнению с контрольным экспериментом. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

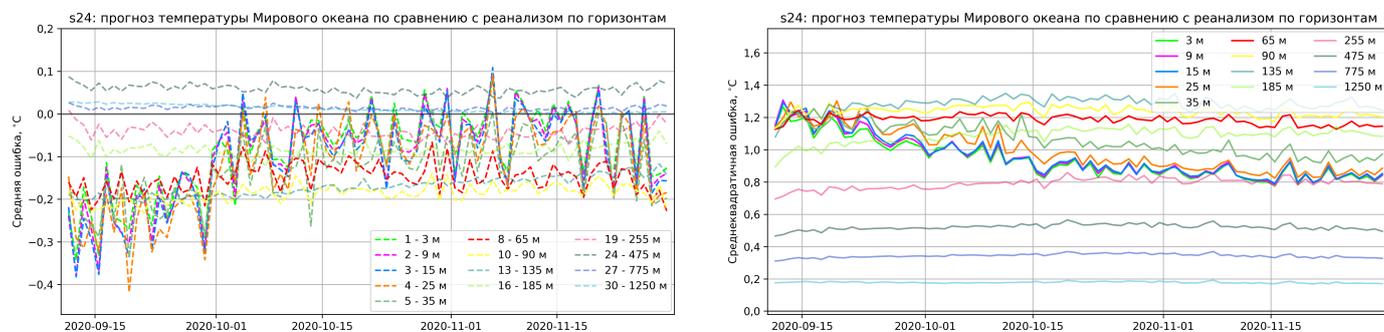


Рис. 3. Точность прогноза на 1 сутки для фиксированного расположения 1299 сенсоров проекта АРГО на различных уровнях по сравнению с контрольным экспериментом. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

по сравнению с s24. При сравнении рис. 4 и 5 можно видеть в целом схожую скорость сходимости как средней, так и среднеквадратичной ошибки в сценариях s44 и s36.

Анализ графиков на рис. 6–8 подчеркивает, что разница между использованием трех тысяч, семи тысяч и тринадцати тысяч сенсоров не является существенной. Графики ошибок в сценариях s36 и s37 указывают на то, что увеличение числа датчиков до 3 тысяч способствует значительному снижению среднеквадратичной ошибки по сравнению с 1 тысячей датчиков. С другой стороны, согласно сравнению s37 и s44, такого же снижения можно достичь без увеличения числа датчиков, за счет их грамотного расположения.

На рис. 9 приведено сравнение динамики ошибок в верхних 100 метрах в течение первых 20 суток для экспериментов s24 (фиксированное расположение измерителей проекта АРГО за 4 суток), s25 (сценарий с динамическим изменением расположения датчиков в соответствии с реальным положением буев проекта АРГО), s36 (регулярное распределение тысячи измерителей) и s44 (расстановка методом Concrete Autoencoder).

Bias, полученный при регулярной расстановке тысячи сенсоров, показывает несколько лучшие результаты в сравнении с s44. Это может быть связано с тем, что в эксперименте с регулярной расстановкой больше сенсоров попало в акваторию Северного Ледовитого океана. В отношении RMSE эксперимент s44 выделяется наилучшими показателями как сразу после старта моделирования, так и в течение всего 20-дневного периода. Также средняя ошибка (Bias) для реального расположения АРГО s25 через неделю после старта сходится к значениям, характерным для регулярной расстановки тысячи измерителей s36 и для расстановки Concrete Autoencoder s44. При этом реальная расстановка АРГО заметно хуже по показателю RMSE, чем другие рассматриваемые расстановки с большим числом сенсоров.

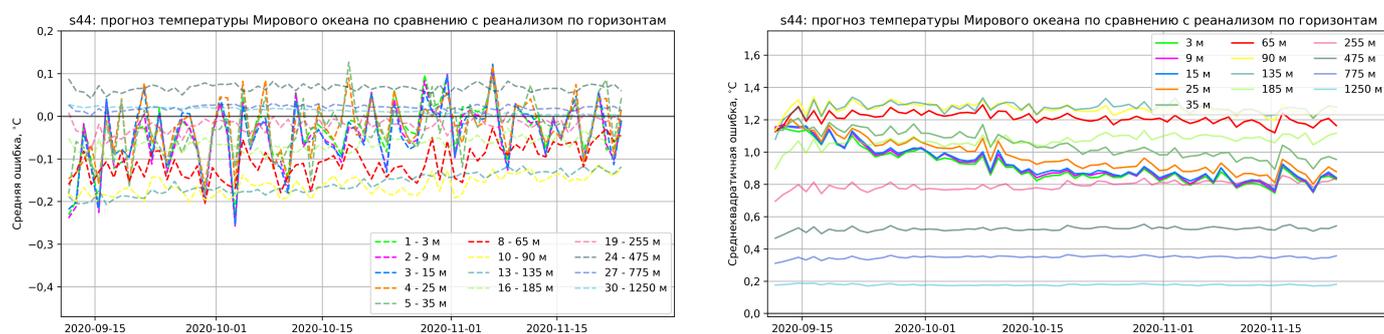


Рис. 4. Точность прогноза на 1 сутки для фиксированного расположения 1160 сенсоров, расставленных методом Concrete Autoencoder, на различных уровнях по сравнению с контрольным экспериментом. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

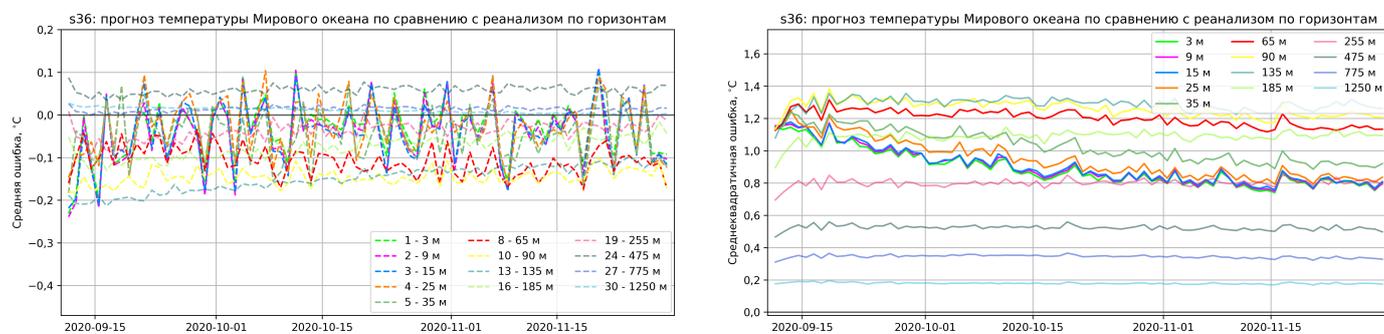


Рис. 5. Точность прогноза на 1 сутки по сравнению с контрольным экспериментом на различных уровнях для фиксированного расположения 1306 сенсоров, расставленных по равномерной решетке с шагом $5,25^\circ$ по долготе и $6,25^\circ$ по широте. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

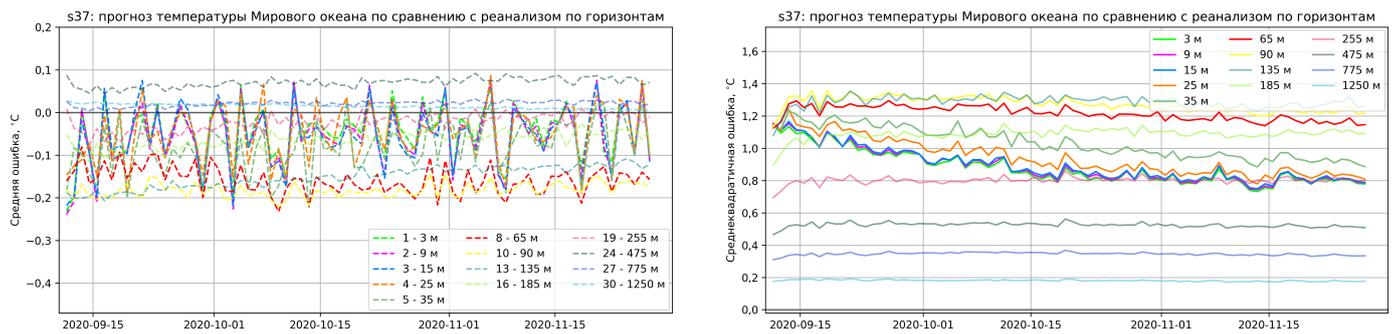


Рис. 6. Точность прогноза на 1 сутки по сравнению с контрольным экспериментом на различных уровнях для фиксированного расположения 3379 сенсоров, расставленных по равномерной решетке с шагом 5° по долготе и 2.5° по широте. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

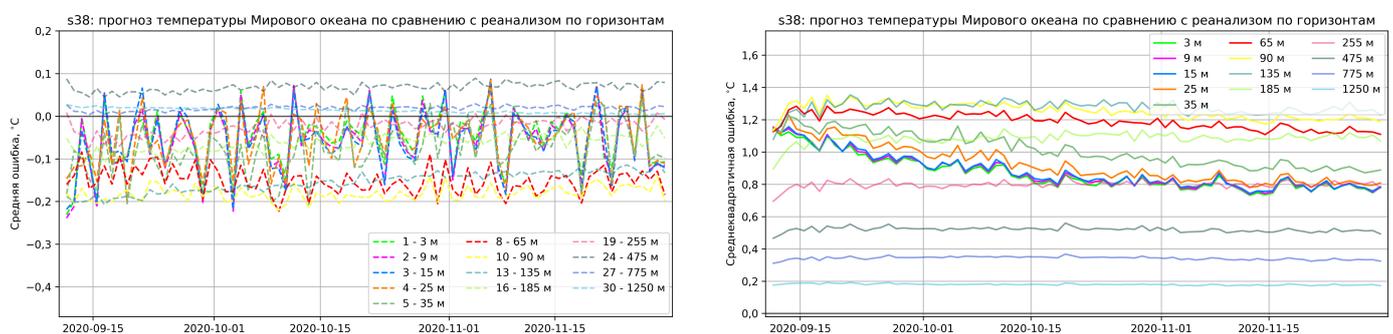


Рис. 7. Точность прогноза на 1 сутки по сравнению с контрольным экспериментом на различных уровнях для фиксированного расположения 6792 сенсоров, расставленных по равномерной решетке с шагом 2.5° по долготе и 2.5° по широте. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

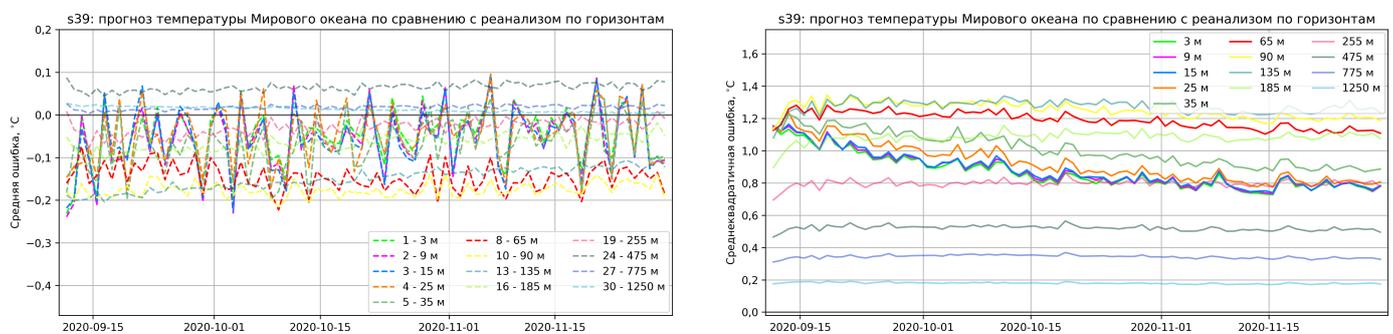


Рис. 8. Точность прогноза на 1 сутки по сравнению с контрольным экспериментом на различных уровнях для фиксированного расположения 13581 сенсоров, расставленных по равномерной решетке с шагом 2.5° по долготе и 1.5° по широте. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

Профили от времени. Отдельно следует обратить внимание на представленные ниже средние суточные профили по 13 уровням для проведенных экспериментов с усвоением, а также для контрольного эксперимента. Эти данные позволяют наблюдать динамику профилей различных экспериментов и детальнее оценить влияние различных методов расстановки.

Как можно видеть на рис. 10 и 11, в сценариях s44, s36, s37, s38 и s39 средний профиль в верхних 60 метрах уже в течение 2–3 дней становится близким к среднему профилю в контрольном эксперименте a01. С другой стороны, при сравнении динамики в первые дни для экспериментов с тысячей сенсоров (s24, s36 и s44) мы обнаруживаем, что в сценарии s24 профиль адаптируется не так быстро, как в s36 и s44.

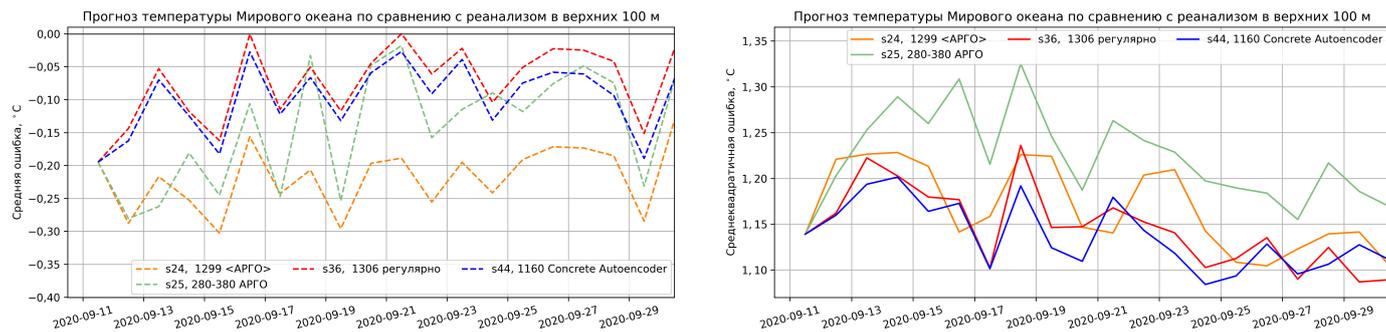


Рис. 9. Точность прогноза на 1 сутки по сравнению с контрольным экспериментом для экспериментов s24, s25, s36 и s44. Пунктирные линии показывают Bias (слева), сплошные линии соответствуют RMSE (справа).

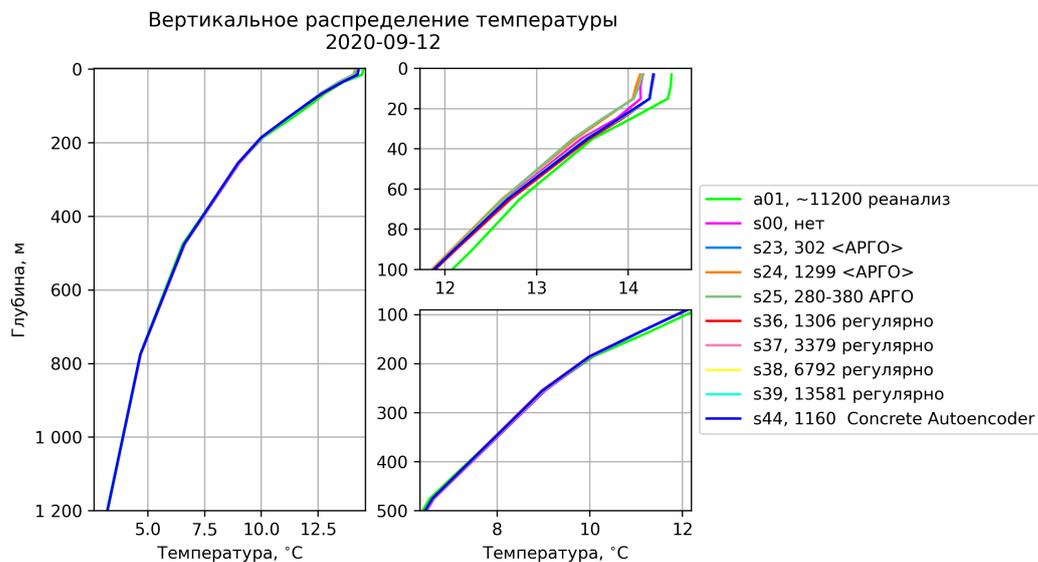


Рис. 10. Профили температуры 2020-09-12 после первой итерации усвоения.

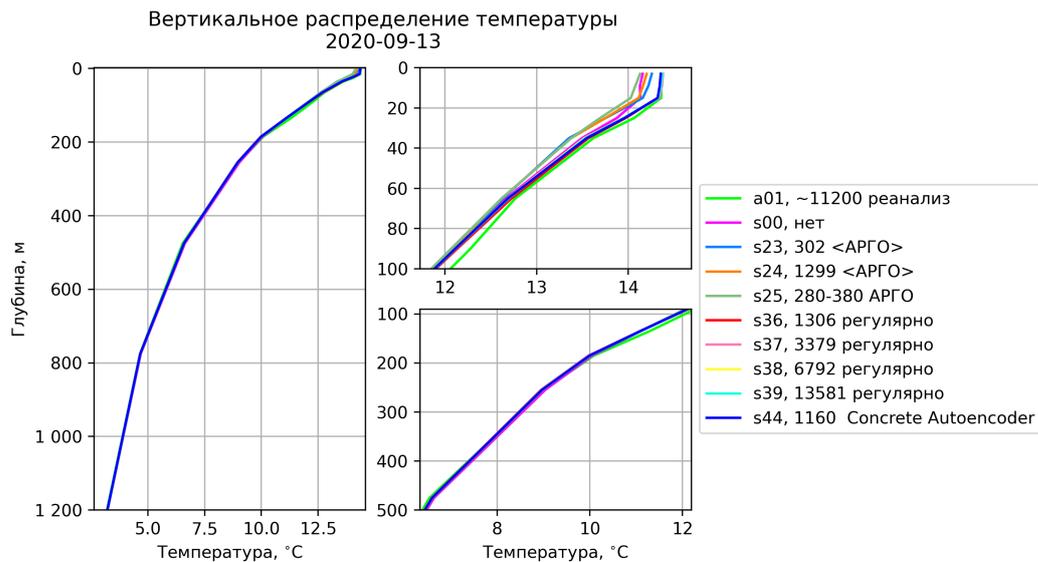


Рис. 11. Профили температуры 2020-09-13 после второй итерации усвоения.

Графики на рис. 12–14 иллюстрируют, что примерно после 20 дней всех экспериментов результаты всех применяемых методов сходятся, показывая схожую динамику.

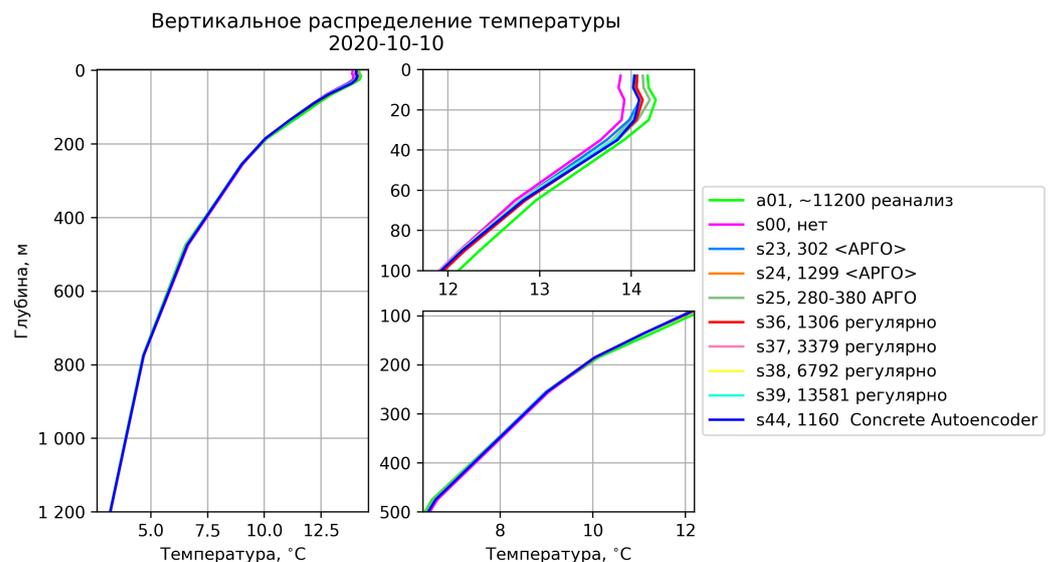


Рис. 12. Профили температуры 2020-10-10 после 29 дней работы усвоения.

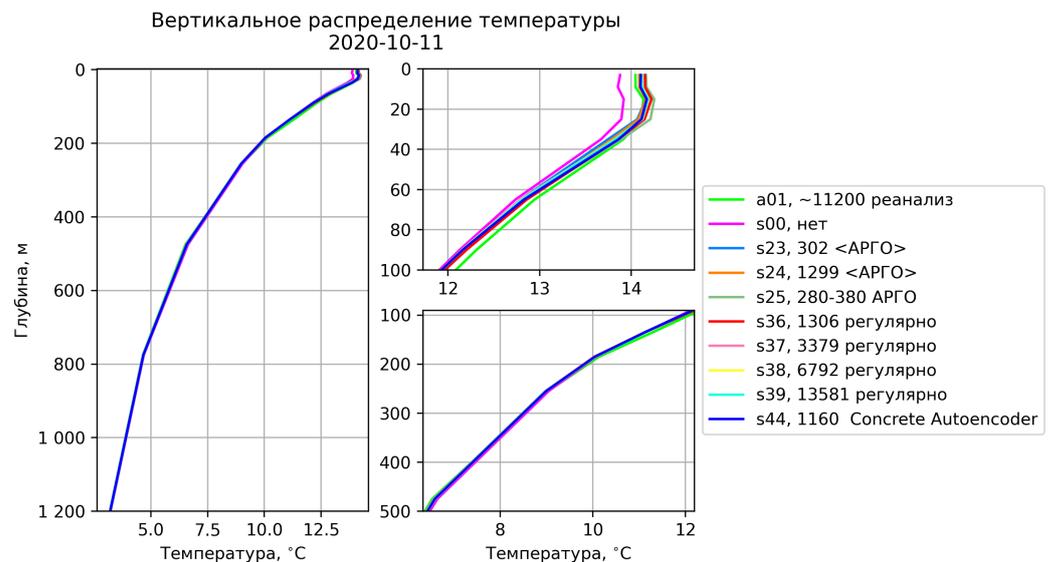


Рис. 13. Профили температуры 2020-10-11 после 30 дней работы усвоения.

Это подтверждает наше предположение о том, что продолжение эксперимента более чем на 30–40 дней не принесет дополнительной информации.

В установившемся через приблизительно 30 дней после старта моделирования квазистационарном режиме профили температур, полученные с использованием метода Concrete Autoencoder, регулярной расстановки на трех, шести и тринадцати тысячах сенсорах, совпадают с целевыми полями, с задержкой около одних суток.

Пространственное распределение ошибок. На рис. 15–18 показано пространственное распределение среднеквадратичной ошибки в верхних 100 метрах за первые 20 дней моделирования для сценариев s25, s24, s36 и s44 соответственно. В Баренцевом море и в Арктике было замечено, что ошибка прогноза при использовании метода СА и регулярной расстановки значительно меньше по сравнению с расстановкой АРГО s25 и s24. При этом регулярная расстановка (s36) имеет больше всего сенсоров по сравнению с другими способами в этом регионе. В других регионах ошибки прогноза для всех методов примерно одинаковы.

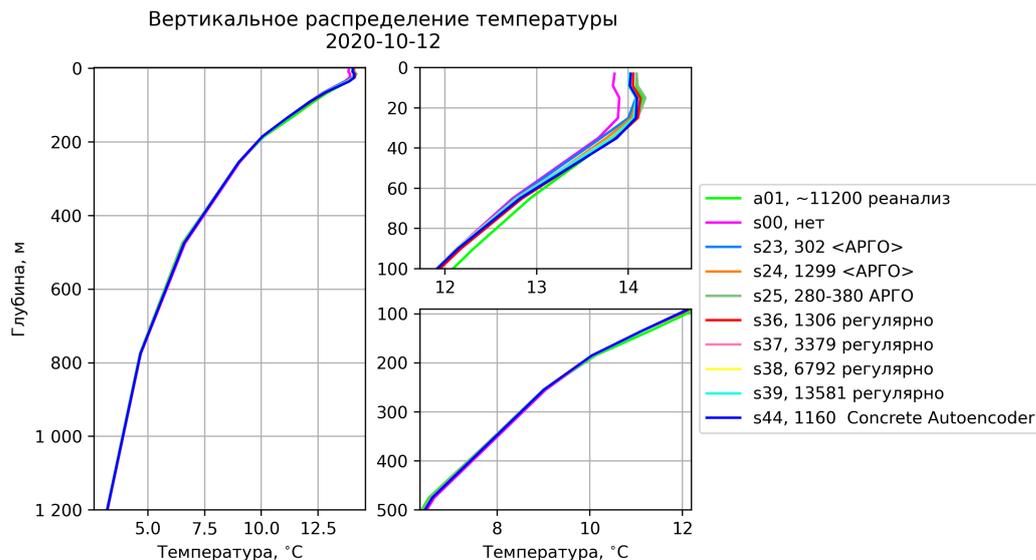


Рис. 14. Профили температуры 2020-10-12 после 31 дня работы усвоения.

s25: распределение ошибки за 20 дней, 280-380 АРГО

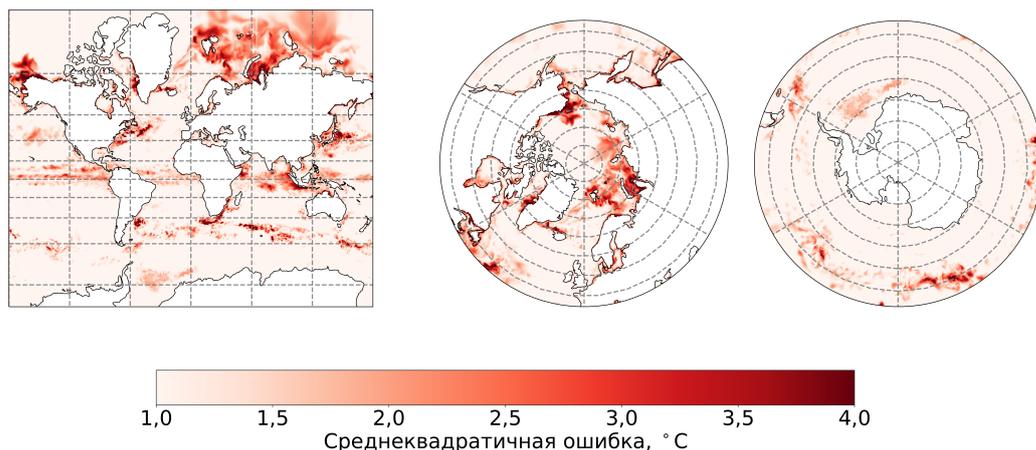


Рис. 15. Пространственное распределение ошибок температуры для изменяющейся расстановки в соответствии с реальными координатами буев проекта АРГО, 280–380 сенсоров.

Следует отметить, что паттерны ошибок в целом одинаковы для всех расстановок, но при этом варьируется амплитуда значений ошибки. Из этого можно сделать вывод, что мы заранее можем утверждать, в каких регионах прогноз получится более достоверным.

На рис. 19–21 представлены координаты сенсоров, которые использовались для наблюдений в проведенных экспериментах s44, s36 и s24 соответственно. Так как регулярное распределение и расстановка методом Concrete Autoencoder имеют большее число сенсоров в Арктике, мы можем также сделать вывод о том, что увеличение числа сенсоров в некотором регионе может значительно повысить точность прогноза.

4. Заключение

Экспериментальный анализ сценариев s00, s23, s24, s25, s36, s37, s38, s39 и s44 раскрыл ряд значимых наблюдений, улучшающих наше понимание влияния расположения и количества датчиков на значения средней (Bias) и среднеквадратичной (RMSE) ошибки.

s24: распределение ошибки за 20 дней, 1299 <АРГО>

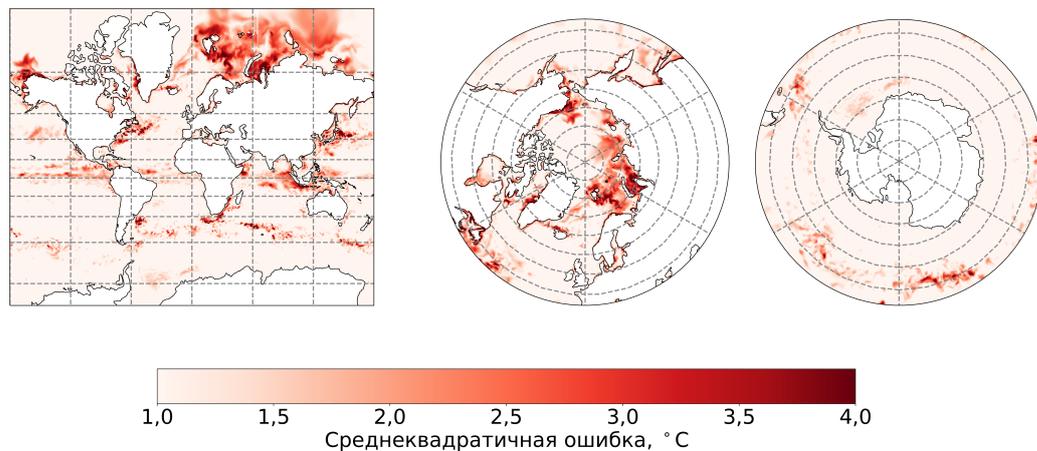


Рис. 16. Пространственное распределение ошибок температуры для фиксированной расстановки АРГО, 1299 сенсоров.

s36: распределение ошибки за 20 дней, 1306 регулярно

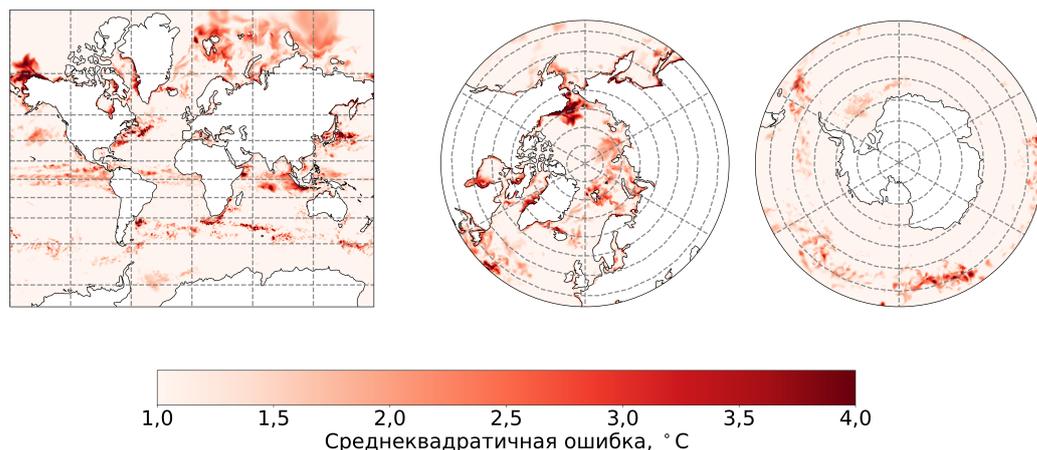


Рис. 17. Пространственное распределение ошибок температуры для фиксированной регулярной расстановки, 1306 сенсоров.

Первое наблюдение, основанное на сценариях s36 и s37 с регулярной расстановкой на 1306 и 3379 сенсоров соответственно, показывает, что увеличение числа датчиков может значительно улучшить показатели среднеквадратичной ошибки. Это подчеркивает, что при реализации сетей датчиков следует тщательно рассматривать возможность увеличения плотности расстановки датчиков, учитывая при этом другие факторы, такие как стоимость и энергопотребление. При этом, как показывает анализ экспериментов s37, s38 и s39, дальнейшее увеличение числа сенсоров более трех тысяч не приводит к заметному уменьшению средней и среднеквадратичной ошибок.

Второе наблюдение получается при сравнении сценариев s36 и s44: применение методов оптимизации при размещении датчиков может обеспечить улучшение показателя систематической ошибки при меньшем числе датчиков. Расстановка, полученная методом Concrete Autoencoder, имеет значительно меньшее значение средней ошибки, чем при регулярной расстановке. Результаты также подтверждают гипотезу, согласно которой примерно половина датчиков из сценария s38 может быть потенциально удалена без существенного ухудшения показателя систематической ошибки. Кроме того, расстановка, полученная методом Concrete Autoencoder с одной тысячей сенсоров,

s44: распределение ошибки за 20 дней, 1160 Concrete Autoencoder

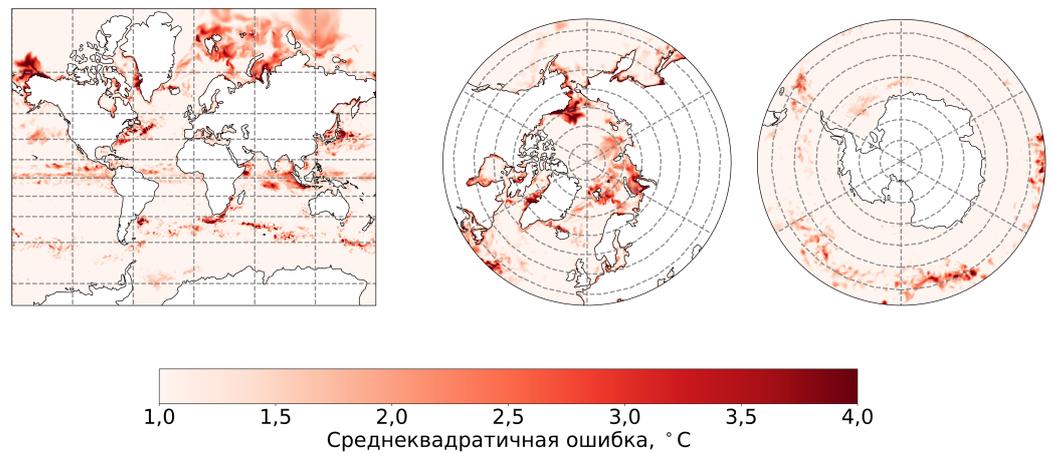


Рис. 18. Пространственное распределение ошибок температуры для фиксированной расстановки СА, 1160 сенсоров.

s44: расположение измерителей, 1160 Concrete Autoencoder

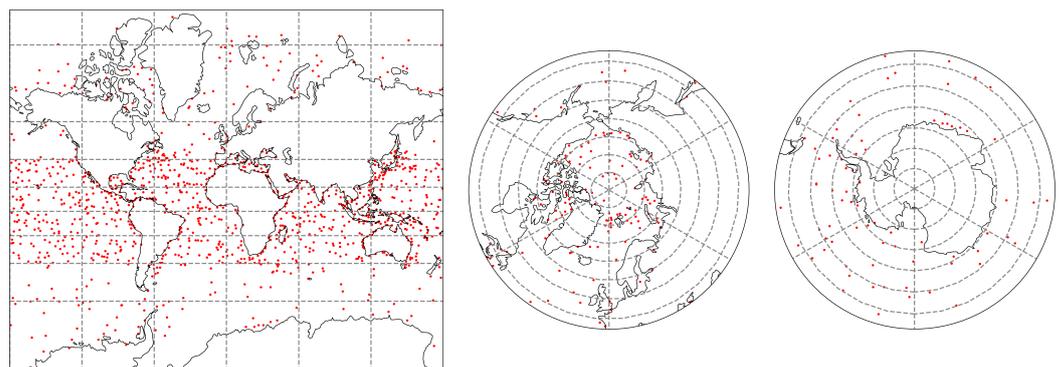


Рис. 19. Расположение 1160 сенсоров, полученное методом Concrete Autoencoder.

s36: расположение измерителей, 1306 регулярно

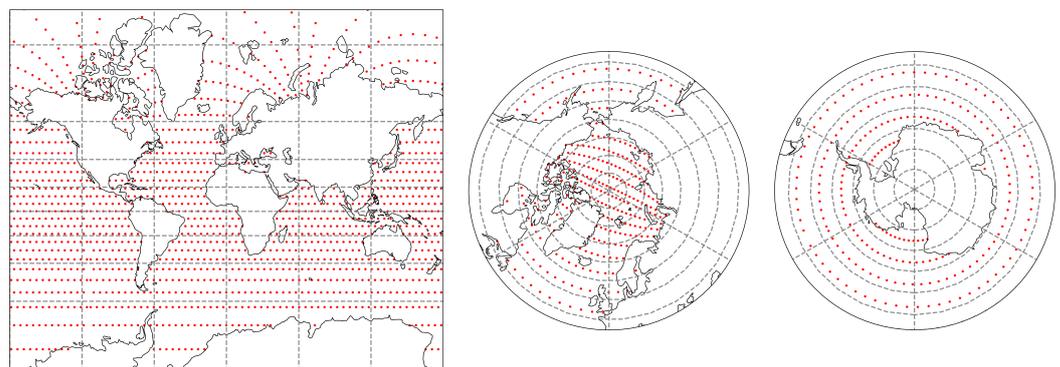


Рис. 20. Расположение 1306 сенсоров, расставленных регулярно с шагом $5,25^\circ$ по долготе и $6,25^\circ$ по широте.

незначительно уступает по RMSE регулярной расстановке с тремя тысячами сенсоров при сопоставимой средней ошибке.

s24: расположение измерителей, 1299 <АРГО>

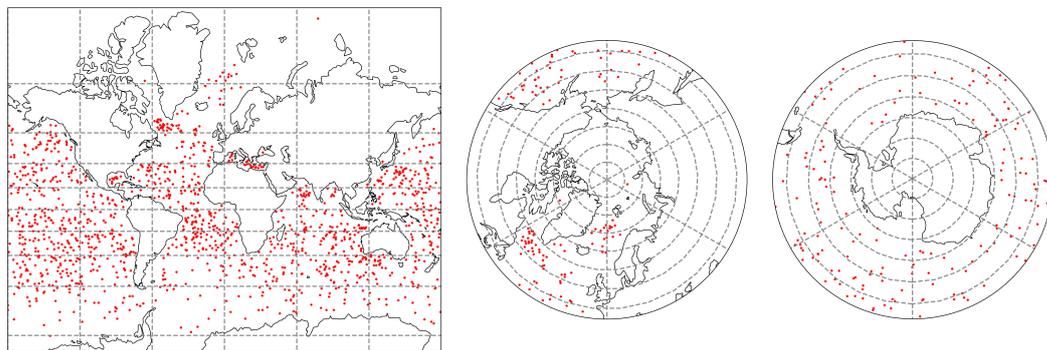


Рис. 21. Расположение 1299 сенсоров, расставленных в соответствии с координатами буев проекта АРГО, проводивших измерения в течение 4 суток, начиная с 2020-09-11.

Анализируя динамику средней ошибки в сценарии s38, мы видим, что примерно за 2–3 итерации усвоения средняя ошибка (Bias) достигает квазистационарных значений. За 20 дней все методы приближаются к одной и той же динамике профилей температуры, и значения среднеквадратичной ошибки выходят на квазистационарный уровень. Это подтверждает наше предположение о том, что продление эксперимента на более чем 30–40 дней не приводит к новым результатам.

В результате детального анализа пространственного распределения среднеквадратичных ошибок в проведенных экспериментах (выше приведены результаты для s24, s25, s36 и s44) было установлено, что, несмотря на различия в абсолютных значениях ошибок, их форма и пространственные паттерны остаются стабильными в различных конфигурациях. Это наблюдение позволяет сделать вывод о том, что в некоторых регионах достоверность прогноза будет систематически выше, в то время как в других она может быть снижена.

Следует отметить, что Concrete Autoencoder был обучен восстанавливать поле температуры в тот же момент времени, в котором проводились измерения, что несколько отличается от рассматриваемой постановки, где измерялась точность прогноза на одни сутки. Несмотря на это, из трех сценариев s24, s36 и s44 с приблизительно одинаковым числом сенсоров, расстановка s44, полученная методом Concrete Autoencoder, приводит к самой быстрой сходимости физических полей к состоянию контрольного эксперимента и наименьшему значению RMSE.

Полученные результаты подчеркивают важность оптимальной расстановки измерительных устройств при планировании наблюдательной сети, так как за счет правильного расположения можно максимизировать точность прогнозирования при одновременном снижении затрат в процессе сбора данных наблюдений.

Благодарности. Работа выполнена в рамках государственного задания ИВМ РАН (тема №075-01132-23-01 – настройка системы усвоения данных, эксперименты на устойчивость), государственного задания ИО РАН (тема №FMWE-2021-0003 – построение модельной конфигурации океана и морского льда, проведение основной серии экспериментов на точность прогноза, обработка и анализ их результатов) и при финансовой поддержке гранта Российского научного фонда (проект №20-19-00615 – построение динамического размещения измерителей, анализ восстановленных полей в Арктике, анализ профилей). Расчет численных экспериментов проводился с использованием ресурсов суперкомпьютера Межведомственного суперкомпьютерного центра РАН (МСЦ РАН, <https://www.jscc.ru/>).

Список литературы

- Кальницкий Л. Ю., Кауркин М. Н., Ушаков К. В. и др. Сезонная изменчивость циркуляции вод и морского льда в Северном Ледовитом океане в модели высокого разрешения // Известия Российской академии наук. Физика атмосферы и океана. — 2020. — Т. 56, № 5. — С. 598–610. — DOI: [10.31857/S0002351520050065](https://doi.org/10.31857/S0002351520050065).
- Кауркин М. Н., Ибраев Р. А. Исследование чувствительности алгоритма усвоения малочисленных данных наблюдений в модели динамики океана // Морской гидрофизический журнал. — 2019. — Т. 35, № 2. — DOI: [10.22449/0233-7584-2019-2-105-113](https://doi.org/10.22449/0233-7584-2019-2-105-113).
- Кошляков М. Н., Тараканов Р. Ю. Введение в физическую океанографию. — Москва : МФТИ, 2014. — 142 с.
- Abbasi M. R., Chegini V., Sadrinab M., *et al.* Correcting the Sea Surface Temperature by Data Assimilation Over the Persian Gulf // Iranian Journal of Science and Technology, Transactions A: Science. — 2018. — Vol. 43, no. 1. — P. 141–149. — DOI: [10.1007/s40995-017-0357-z](https://doi.org/10.1007/s40995-017-0357-z).
- Abid A., Balin M. F., Zou J. Concrete Autoencoders for Differentiable Feature Selection and Reconstruction // Cornell University. — 2019. — Vol. abs/1901.09346. — DOI: [10.48550/ARXIV.1901.09346](https://doi.org/10.48550/ARXIV.1901.09346).
- Alonso A. A., Frouzakis C. E., Kevrekidis I. G. Optimal sensor placement for state reconstruction of distributed process systems // AIChE Journal. — 2004. — Vol. 50, no. 7. — P. 1438–1452. — DOI: [10.1002/aic.10121](https://doi.org/10.1002/aic.10121).
- Boyer T. P., Antonov J. I., Baranova O. K., *et al.* World ocean database 2013. — U. S. Department of Commerce, National Oceanic, Atmospheric Administration, National Environmental Satellite, Data, Information Service, National Oceanographic Data Center, Ocean Climate Laboratory, 2013. — DOI: [10.7289/V5NZ85MT](https://doi.org/10.7289/V5NZ85MT).
- Clark E., Askham T., Brunton S. L., *et al.* Greedy Sensor Placement With Cost Constraints // IEEE Sensors Journal. — 2019. — Vol. 19, no. 7. — P. 2642–2656. — DOI: [10.1109/JSEN.2018.2887044](https://doi.org/10.1109/JSEN.2018.2887044).
- Covert I., Qiu W., Lu M., *et al.* Learning to Maximize Mutual Information for Dynamic Feature Selection // Proceedings of the 40th International Conference on Machine Learning. — Honolulu, Hawaii, USA : PMLR 202, 2023. — DOI: [10.48550/arXiv.2301.00557](https://doi.org/10.48550/arXiv.2301.00557).
- Desai S. Jason-3 GPS based orbit and SSHA OGDR. — 2016. — DOI: [10.5067/J3L2G-OGDRF](https://doi.org/10.5067/J3L2G-OGDRF).
- Fadeev R. Y., Ushakov K. V., Tolstykh M. A., *et al.* Design and development of the SLAV-INMIO-CICE coupled model for seasonal prediction and climate research // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2018. — Vol. 33, no. 6. — P. 333–340. — DOI: [10.1515/rnam-2018-0028](https://doi.org/10.1515/rnam-2018-0028).
- Hersbach H., Bell B., Berrisford P., *et al.* The ERA5 global reanalysis // Quarterly Journal of the Royal Meteorological Society. — 2020. — Vol. 146, no. 730. — P. 1999–2049. — DOI: [10.1002/qj.3803](https://doi.org/10.1002/qj.3803).
- Huijben I. A. M., Veeling B. S., Sloun R. J. G. van. Deep probabilistic subsampling for taskadaptivecompressed sensing // International Conference on Learning Representations 2020. — ICLR, 2020.
- Hunke E. C., Lipscomb W. H., Turner A. K., *et al.* CICE: The Los Alamos Sea ice Model Documentation and Software User's Manual Version 5 (Tech. Rep. LA-CC-06-012). — Los Alamos National Laboratory, 2015.
- Jang E., Gu S., Poole B. Categorical Reparameterization with Gumbel-Softmax // Cornell University. — 2016. — Vol. abs/1611.01144. — DOI: [10.48550/arXiv.1611.01144](https://doi.org/10.48550/arXiv.1611.01144).
- Kalmykov V. V., Ibrayev R. A., Kaurkin M. N., *et al.* Compact Modeling Framework v3.0 for high-resolution global ocean-ice-atmosphere models // Geoscientific Model Development. — 2018. — Vol. 11, no. 10. — P. 3983–3997. — DOI: [10.5194/gmd-11-3983-2018](https://doi.org/10.5194/gmd-11-3983-2018).
- Kaurkin M. N., Ibrayev R. A., Belyaev K. P. Data assimilation in the ocean circulation model of high spatial resolution using the methods of parallel programming // Russian Meteorology and Hydrology. — 2016a. — Vol. 41, no. 7. — P. 479–486. — DOI: [10.3103/S1068373916070050](https://doi.org/10.3103/S1068373916070050).
- Kaurkin M. N., Ibrayev R. A., Koromyslov A. EnOI-Based Data Assimilation Technology for Satellite Observations and ARGO Float Measurements in a High Resolution Global Ocean Model Using the CMF Platform // Supercomputing. — Springer International Publishing, 2016b. — P. 57–66. — DOI: [10.1007/978-3-319-55669-7_5](https://doi.org/10.1007/978-3-319-55669-7_5).
- Krause A., Singh A., Guestrin C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies // Journal of Machine Learning Research. — 2008. — Vol. 9, no. 2.
- Kumar P., El Sayed Y. M., Semaan R. Optimized sensor placement using stochastic estimation for a flow over a 2D airfoil with Coanda blowing // 7th AIAA Flow Control Conference. — American Institute of Aeronautics, Astronautics, 2014. — DOI: [10.2514/6.2014-2101](https://doi.org/10.2514/6.2014-2101).
- Lavergne T., Sørensen A. M., Kern S., *et al.* Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records // The Cryosphere. — 2019. — Vol. 13, no. 1. — P. 49–78. — DOI: [10.5194/tc-13-49-2019](https://doi.org/10.5194/tc-13-49-2019).
- Li X., Wang S., Cai Y. Tutorial: Complexity analysis of Singular Value Decomposition and its variants // Cornell University. — 2019. — Vol. abs/1906.12085. — DOI: [10.48550/arXiv.1906.12085](https://doi.org/10.48550/arXiv.1906.12085).

- Lobashev A. A., Turko N. A., Ushakov K. V., *et al.* Concrete Autoencoder for the Reconstruction of Sea Temperature Field from Sparse Measurements // Journal of Marine Science and Engineering. — 2023. — Vol. 11, no. 2. — P. 404. — DOI: [10.3390/jmse11020404](https://doi.org/10.3390/jmse11020404).
- Maddison C. J., Mnih A., Teh Y. W. The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables // Cornell University. — 2016. — Vol. abs/1611.00712. — DOI: [10.48550/arXiv.1611.00712](https://doi.org/10.48550/arXiv.1611.00712).
- Manohar K., Brunton B. W., Kutz J. N., *et al.* Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns // IEEE Control Systems. — 2018. — Vol. 38, no. 3. — P. 63–86. — DOI: [10.1109/MCS.2018.2810460](https://doi.org/10.1109/MCS.2018.2810460).
- Murray R. J. Explicit Generation of Orthogonal Grids for Ocean Models // Journal of Computational Physics. — 1996. — Vol. 126, no. 2. — P. 251–273. — DOI: [10.1006/jcph.1996.0136](https://doi.org/10.1006/jcph.1996.0136).
- Nagata T., Nonomura T., Nakai K., *et al.* Data-Driven Sparse Sensor Selection Based on A-Optimal Design of Experiment With ADMM // IEEE Sensors Journal. — 2021. — Vol. 21, no. 13. — P. 15248–15257. — DOI: [10.1109/JSEN.2021.3073978](https://doi.org/10.1109/JSEN.2021.3073978).
- Nakai K., Yamada K., Nagata T., *et al.* Effect of Objective Function on Data-Driven Greedy Sparse Sensor Optimization // IEEE Access. — 2021. — Vol. 9. — P. 46731–46743. — DOI: [10.1109/ACCESS.2021.3067712](https://doi.org/10.1109/ACCESS.2021.3067712).
- Nguyen L., Hu G., Spanos C. J. Efficient Sensor Deployments for Spatio-Temporal Environmental Monitoring // IEEE Transactions on Systems, Man, and Cybernetics: Systems. — 2020. — Vol. 50, no. 12. — P. 5306–5316. — DOI: [10.1109/TSMC.2018.2872041](https://doi.org/10.1109/TSMC.2018.2872041).
- Pathak J., Subramanian S., Harrington P., *et al.* FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators // Cornell University. — 2022. — Vol. abs/2202.11214. — DOI: [10.48550/arXiv.2202.11214](https://doi.org/10.48550/arXiv.2202.11214).
- Ryan A. G., Regnier C., Divakaran P., *et al.* GODAE OceanView Class 4 forecast verification framework: global ocean inter-comparison // Journal of Operational Oceanography. — 2015. — Vol. 8, sup1. — s98–s111. — DOI: [10.1080/1755876X.2015.1022330](https://doi.org/10.1080/1755876X.2015.1022330).
- Saito Y., Nonomura T., Yamada K., *et al.* Determinant-Based Fast Greedy Sensor Selection Algorithm // IEEE Access. — 2021. — Vol. 9. — P. 68535–68551. — DOI: [10.1109/ACCESS.2021.3076186](https://doi.org/10.1109/ACCESS.2021.3076186).
- Sallée J.-B., Pellichero V., Akhoudas C., *et al.* Summertime increases in upper-ocean stratification and mixed-layer depth // Nature. — 2021. — Vol. 591, no. 7851. — P. 592–598. — DOI: [10.1038/s41586-021-03303-x](https://doi.org/10.1038/s41586-021-03303-x).
- Sun S., Liu S., Liu J., *et al.* Wind Field Reconstruction Using Inverse Process With Optimal Sensor Placement // IEEE Transactions on Sustainable Energy. — 2019. — Vol. 10, no. 3. — P. 1290–1299. — DOI: [10.1109/TSTE.2018.2865512](https://doi.org/10.1109/TSTE.2018.2865512).
- Turko N., Lobashev A., Ushakov K., *et al.* Information Entropy Initialized Concrete Autoencoder for Optimal Sensor Placement and Reconstruction of Geophysical Fields // Supercomputing. — Springer International Publishing, 2022. — P. 167–184. — DOI: [10.1007/978-3-031-22941-1_12](https://doi.org/10.1007/978-3-031-22941-1_12).
- Turpin V., Remy E., Traon P. Y. L. How essential are Argo observations to constrain a global ocean data assimilation system? // Ocean Science. — 2016. — Vol. 12, no. 1. — P. 257–274. — DOI: [10.5194/os-12-257-2016](https://doi.org/10.5194/os-12-257-2016).
- Ushakov K. V., Ibrayev R. A. Assessment of mean world ocean meridional heat transport characteristics by a high-resolution model // Russian Journal of Earth Sciences. — 2018. — Vol. 18, no. 1. — DOI: [10.2205/2018ES000616](https://doi.org/10.2205/2018ES000616).
- Wong A. P. S., Wijffels S. E., Riser S. C., *et al.* Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats // Frontiers in Marine Science. — 2020. — Vol. 7. — DOI: [10.3389/fmars.2020.00700](https://doi.org/10.3389/fmars.2020.00700).

GLOBAL OCEAN FORECAST ACCURACY IMPROVEMENT DUE TO OPTIMAL SENSOR PLACEMENT

Nikita Turko^{**1} , Alexander Lobashev² , Konstantin Ushakov^{1,3} , Maxim Kaurkin¹ , Leonid Kalnitskii⁴ , Sergey Semin⁵ , Rashit Ibrayev^{1,3,6} 

¹Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

²Skolkovo Institute of Science and Technology, Moscow, Russia

³Moscow Institute of Physics and Technology, Dolgoprudny, Russia

⁴Arctic and Antarctic Research Institute, Saint Petersburg, Russia

⁵Nuclear Safety Institute, Russian Academy of Sciences, Moscow, Russia

⁶Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

****Correspondence to:** Nikita A. Turko, NikitaTurko@yandex.ru

The paper examines the impact of sensor placement on the accuracy of the Global ocean state forecasting. A comparison is made between various sensor placement methods, including the arrangement obtained by the Concrete Autoencoder method. To evaluate how sensor placement affects forecast accuracy, a simulation was conducted that emulates a scenario where the initial state of the global ocean significantly deviates from the ground truth. In the experiment, initial conditions for the ocean and ice model were altered, while atmospheric forcing was retained from the control experiment. Subsequently, the model was integrated with the assimilation of data about the ground truth state at the sensor locations. The results showed that the sensor placement obtained using deep learning methods is superior in forecast accuracy to other considered arrays with a comparable number of sensors.

Keywords: operational forecast, Global ocean, optimal sensor placement, Concrete Autoencoder, data assimilation

Received: 25 August 2023

Accepted: 15 December 2023

Published: 30 December 2023



© 2023. The Authors.

Citation: Turko, N., A. Lobashev, K. Ushakov, L. Kalnitskii, M. Kaurkin, S. Semin, and R. Ibrayev (2023), Global Ocean Forecast Accuracy Improvement Due to Optimal Sensor Placement, *Russian Journal of Earth Sciences*, 23, ES6005, <https://doi.org/10.2205/2023ES000883> EDN: UOMRDC

References

- Abbasi M. R., Chegini V., Sadriinasab M., *et al.* Correcting the Sea Surface Temperature by Data Assimilation Over the Persian Gulf // *Iranian Journal of Science and Technology, Transactions A: Science*. — 2018. — Vol. 43, no. 1. — P. 141–149. — DOI: [10.1007/s40995-017-0357-z](https://doi.org/10.1007/s40995-017-0357-z).
- Abid A., Balin M. F., Zou J. Concrete Autoencoders for Differentiable Feature Selection and Reconstruction // Cornell University. — 2019. — Vol. 1. — DOI: [10.48550/ARXIV.1901.09346](https://doi.org/10.48550/ARXIV.1901.09346).
- Alonso A. A., Frouzakis C. E., Kevrekidis I. G. Optimal sensor placement for state reconstruction of distributed process systems // *AICHE Journal*. — 2004. — Vol. 50, no. 7. — P. 1438–1452. — DOI: [10.1002/aic.10121](https://doi.org/10.1002/aic.10121).
- Boyer T. P., Antonov J. I., Baranova O. K., *et al.* World ocean database 2013. — U. S. Department of Commerce, National Oceanic, Atmospheric Administration, National Environmental Satellite, Data, Information Service, National Oceanographic Data Center, Ocean Climate Laboratory, 2013. — DOI: [10.7289/V5NZ85MT](https://doi.org/10.7289/V5NZ85MT).
- Clark E., Askham T., Brunton S. L., *et al.* Greedy Sensor Placement With Cost Constraints // *IEEE Sensors Journal*. — 2019. — Vol. 19, no. 7. — P. 2642–2656. — DOI: [10.1109/JSEN.2018.2887044](https://doi.org/10.1109/JSEN.2018.2887044).

- Covert I., Qiu W., Lu M., *et al.* Learning to Maximize Mutual Information for Dynamic Feature Selection // Proceedings of the 40 th International Conference on Machine Learning. — Honolulu, Hawaii, USA : PMLR 202, 2023. — DOI: [10.48550/arXiv.2301.00557](https://doi.org/10.48550/arXiv.2301.00557).
- Desai S. Jason-3 GPS based orbit and SSHA OGDR. — 2016. — DOI: [10.5067/J3L2G-OGDRF](https://doi.org/10.5067/J3L2G-OGDRF). — URL: https://podaac.jpl.nasa.gov/dataset/JASON_3_L2_OST_OGDR_GPS.
- Fadeev R. Y., Ushakov K. V., Tolstykh M. A., *et al.* Design and development of the SLAV-INMIO-CICE coupled model for seasonal prediction and climate research // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2018. — Vol. 33, no. 6. — P. 333–340. — DOI: [10.1515/rnam-2018-0028](https://doi.org/10.1515/rnam-2018-0028).
- Hersbach H., Bell B., Berrisford P., *et al.* The ERA5 global reanalysis // Quarterly Journal of the Royal Meteorological Society. — 2020. — Vol. 146, no. 730. — P. 1999–2049. — DOI: [10.1002/qj.3803](https://doi.org/10.1002/qj.3803).
- Huijben I. A. M., Veeling B. S., Sloun R. J. G. van. Deep probabilistic subsampling for taskadaptivecompressed sensing // International Conference on Learning Representations 2020. — ICLR, 2020.
- Hunke E. C., Lipscomb W. H., Turner A. K., *et al.* CICE: The Los Alamos Sea ice Model Documentation and Software User's Manual Version 5 (Tech. Rep. LA-CC-06-012). — Los Alamos National Laboratory, 2015.
- Jang E., Gu S., Poole B. Categorical Reparameterization with Gumbel-Softmax // Cornell University. — 2016. — Vol. abs/1611.01144. — DOI: [10.48550/arXiv.1611.01144](https://doi.org/10.48550/arXiv.1611.01144).
- Kalmykov V. V., Ibrayev R. A., Kaurkin M. N., *et al.* Compact Modeling Framework v3.0 for high-resolution global ocean-ice-atmosphere models // Geoscientific Model Development. — 2018. — Vol. 11, no. 10. — P. 3983–3997. — DOI: [10.5194/gmd-11-3983-2018](https://doi.org/10.5194/gmd-11-3983-2018).
- Kalnitiskii L. Y., Kaurkin M. N., Ushakov K. V., *et al.* Seasonal Variability of Water and Sea-Ice Circulation in the Arctic Ocean in a High-Resolution Model // Izvestiya, Atmospheric and Oceanic Physics. — 2020. — Vol. 56, no. 5. — P. 522–533. — DOI: [10.1134/S0001433820050060](https://doi.org/10.1134/S0001433820050060). — URL: <https://doi.org/10.1134/S0001433820050060>.
- Kaurkin M., Ibrayev R., Koromyslov A. EnOI-Based Data Assimilation Technology for Satellite Observations and ARGO Float Measurements in a High Resolution Global Ocean Model Using the CMF Platform // Supercomputing. — Springer International Publishing, 2016a. — P. 57–66. — DOI: [10.1007/978-3-319-55669-7_5](https://doi.org/10.1007/978-3-319-55669-7_5).
- Kaurkin M. N., Ibrayev R. A. Study of Sensitivity of the Algorithm for Assimilating Small Amount of Data in the Ocean Dynamics Model // Morskoy gidrofizicheskiy zhurnal. — 2019. — Vol. 35, no. 2. — DOI: [10.22449/0233-7584-2019-2-105-113](https://doi.org/10.22449/0233-7584-2019-2-105-113).
- Kaurkin M. N., Ibrayev R. A., Belyaev K. P. Data assimilation in the ocean circulation model of high spatial resolution using the methods of parallel programming // Russian Meteorology and Hydrology. — 2016b. — Vol. 41, no. 7. — P. 479–486. — DOI: [10.3103/S1068373916070050](https://doi.org/10.3103/S1068373916070050).
- Koshlyakov M. N., Tarakanov R. Y. Introduction to Physical Oceanography. — Moscow : MIPT, 2014. — 142 p.
- Krause A., Singh A., Guestrin C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies // Journal of Machine Learning Research. — 2008. — Vol. 9, no. 2.
- Kumar P., El Sayed Y. M., Semaan R. Optimized sensor placement using stochastic estimation for a flow over a 2D airfoil with Coanda blowing // 7th AIAA Flow Control Conference. — American Institute of Aeronautics, Astronautics, 2014. — DOI: [10.2514/6.2014-2101](https://doi.org/10.2514/6.2014-2101).
- Lavergne T., Sørensen A. M., Kern S., *et al.* Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records // The Cryosphere. — 2019. — Vol. 13, no. 1. — P. 49–78. — DOI: [10.5194/tc-13-49-2019](https://doi.org/10.5194/tc-13-49-2019).
- Li X., Wang S., Cai Y. Tutorial: Complexity analysis of Singular Value Decomposition and its variants // Cornell University. — 2019. — Vol. abs/1906.12085. — DOI: [10.48550/arXiv.1906.12085](https://doi.org/10.48550/arXiv.1906.12085).
- Lobashev A. A., Turko N. A., Ushakov K. V., *et al.* Concrete Autoencoder for the Reconstruction of Sea Temperature Field from Sparse Measurements // Journal of Marine Science and Engineering. — 2023. — Vol. 11, no. 2. — P. 404. — DOI: [10.3390/jmse11020404](https://doi.org/10.3390/jmse11020404).
- Maddison C. J., Mnih A., Teh Y. W. The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables // Cornell University. — 2016. — Vol. abs/1611.00712. — DOI: [10.48550/arXiv.1611.00712](https://doi.org/10.48550/arXiv.1611.00712).
- Manohar K., Brunton B. W., Kutz J. N., *et al.* Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns // IEEE Control Systems. — 2018. — Vol. 38, no. 3. — P. 63–86. — DOI: [10.1109/MCS.2018.2810460](https://doi.org/10.1109/MCS.2018.2810460).
- Murray R. J. Explicit Generation of Orthogonal Grids for Ocean Models // Journal of Computational Physics. — 1996. — Vol. 126, no. 2. — P. 251–273. — DOI: [10.1006/jcph.1996.0136](https://doi.org/10.1006/jcph.1996.0136).
- Nagata T., Nonomura T., Nakai K., *et al.* Data-Driven Sparse Sensor Selection Based on A-Optimal Design of Experiment With ADMM // IEEE Sensors Journal. — 2021. — Vol. 21, no. 13. — P. 15248–15257. — DOI: [10.1109/JSEN.2021.3073978](https://doi.org/10.1109/JSEN.2021.3073978).

- Nakai K., Yamada K., Nagata T., *et al.* Effect of Objective Function on Data-Driven Greedy Sparse Sensor Optimization // IEEE Access. — 2021. — Vol. 9. — P. 46731–46743. — DOI: [10.1109/ACCESS.2021.3067712](https://doi.org/10.1109/ACCESS.2021.3067712).
- Nguyen L., Hu G., Spanos C. J. Efficient Sensor Deployments for Spatio-Temporal Environmental Monitoring // IEEE Transactions on Systems, Man, and Cybernetics: Systems. — 2020. — Vol. 50, no. 12. — P. 5306–5316. — DOI: [10.1109/TSMC.2018.2872041](https://doi.org/10.1109/TSMC.2018.2872041).
- Pathak J., Subramanian S., Harrington P., *et al.* FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators // Cornell University. — 2022. — Vol. abs/2202.11214. — DOI: [10.48550/arXiv.2202.11214](https://doi.org/10.48550/arXiv.2202.11214).
- Ryan A. G., Regnier C., Divakaran P., *et al.* GODAE OceanView Class 4 forecast verification framework: global ocean inter-comparison // Journal of Operational Oceanography. — 2015. — Vol. 8, sup1. — s98–s111. — DOI: [10.1080/1755876X.2015.1022330](https://doi.org/10.1080/1755876X.2015.1022330).
- Saito Y., Nonomura T., Yamada K., *et al.* Determinant-Based Fast Greedy Sensor Selection Algorithm // IEEE Access. — 2021. — Vol. 9. — P. 68535–68551. — DOI: [10.1109/ACCESS.2021.3076186](https://doi.org/10.1109/ACCESS.2021.3076186).
- Sallée J.-B., Pellichero V., Akhoudas C., *et al.* Summertime increases in upper-ocean stratification and mixed-layer depth // Nature. — 2021. — Vol. 591, no. 7851. — P. 592–598. — DOI: [10.1038/s41586-021-03303-x](https://doi.org/10.1038/s41586-021-03303-x).
- Sun S., Liu S., Liu J., *et al.* Wind Field Reconstruction Using Inverse Process With Optimal Sensor Placement // IEEE Transactions on Sustainable Energy. — 2019. — Vol. 10, no. 3. — P. 1290–1299. — DOI: [10.1109/TSTE.2018.2865512](https://doi.org/10.1109/TSTE.2018.2865512).
- Turko N., Lobashev A., Ushakov K., *et al.* Information Entropy Initialized Concrete Autoencoder for Optimal Sensor Placement and Reconstruction of Geophysical Fields // Supercomputing. — Springer International Publishing, 2022. — P. 167–184. — DOI: [10.1007/978-3-031-22941-1_12](https://doi.org/10.1007/978-3-031-22941-1_12).
- Turpin V., Remy E., Traon P. Y. L. How essential are Argo observations to constrain a global ocean data assimilation system? // Ocean Science. — 2016. — Vol. 12, no. 1. — P. 257–274. — DOI: [10.5194/os-12-257-2016](https://doi.org/10.5194/os-12-257-2016).
- Ushakov K. V., Ibrayev R. A. Assessment of mean world ocean meridional heat transport characteristics by a high-resolution model // Russian Journal of Earth Sciences. — 2018. — Vol. 18, no. 1. — DOI: [10.2205/2018ES000616](https://doi.org/10.2205/2018ES000616).
- Wong A. P. S., Wijffels S. E., Riser S. C., *et al.* Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats // Frontiers in Marine Science. — 2020. — Vol. 7. — DOI: [10.3389/fmars.2020.00700](https://doi.org/10.3389/fmars.2020.00700).