

*На правах рукописи*

Кауркин Максим Николаевич

**Параллельный алгоритм ансамблевой оптимальной интерполяции  
усвоения данных наблюдений в модели динамики океана высокого  
пространственного разрешения**

05.13.18 «Математическое моделирование, численные методы и комплексы программ»

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата физико-математических наук

Москва — 2017

Работа выполнена в федеральном государственном бюджетном учреждении науки Институте вычислительной математики Российской академии наук (ИВМ РАН)

Научный руководитель: **Ибраев Рашит Ахметзиевич**,  
доктор физико-математических наук,  
член-корреспондент РАН.

Официальные оппоненты: **Кныш Василий Васильевич**,  
доктор физико-математических наук,  
профессор. Федеральное государственное  
бюджетное учреждение науки «Морской  
гидрофизический институт Российской академии  
наук» (МГИ РАН), ведущий научный  
сотрудник отдела динамики океанических  
процессов.

**Степаненко Виктор Михайлович**,  
кандидат физико-математических наук.  
Федеральное государственное бюджетное  
образовательное учреждение высшего  
образования «Московский государственный  
университет имени М.В.Ломоносова»,  
ведущий научный сотрудник Научно-  
исследовательского вычислительного  
центра (НИВЦ МГУ).

Ведущая организация: Федеральное государственное бюджетное  
учреждение науки Институт вычислительной  
математики и математической геофизики  
Сибирского отделения Российской  
академии наук (ИВМиМГ СО РАН)

Защита состоится 20 сентября 2017 года в 14<sup>00</sup> часов на заседании  
диссертационного совета Д 002.045.01 при федеральном государственном  
бюджетном учреждении науки Институте вычислительной математики  
Российской академии наук по адресу: 119333, г. Москва, ул. Губкина, д. 8.

С диссертацией можно ознакомиться в библиотеке и на сайте ИВМ РАН  
<http://www.inm.ras.ru>

Автореферат разослан \_\_\_\_\_ 2017 года.

Ученый секретарь  
диссертационного совета Д 002.045.01,  
доктор физико-математических наук

Бочаров Геннадий Алексеевич



## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы.** Модель динамики океана описывает временную эволюцию полей физических величин (температуры, солености, скорости и т.д.). В океанологии под усвоением данных в модели океана понимается процесс корректировки модельного решения, описывающего состояние океана, данными наблюдений, чтобы уменьшить расхождения между модельным полем и этими данными, т.е. приблизить решение к реальному состоянию океана. По определению, модель океана является несовершенным представлением реальной системы, использующим упрощенные математические параметризации для сложных процессов и содержащим ошибки, связанные с разрешением модельной сетки и используемых конечно-разностных методов. Даже в верно сформулированной и реализованной модели океана возникает расхождение с данными наблюдений из-за неопределенностей в начальных условиях и атмосферном форсинге (воздействии). Расхождение модели с реальным состоянием океана являются фундаментальным и неизбежным фактом. Разумеется, наблюдения также не обеспечивают полную картину состояния океана из-за ошибок измерений и малочисленности этих данных.

Задача эффективного использования растущих объемов данных наблюдений для понимания океанических процессов является актуальной. Например, уже сейчас доступны спутниковые данные наблюдений за поверхностной температурой и уровнем с разрешением менее 1 км. Это соответствует поступлению  $\sim 1$  Гб информации за одни сутки. Важно эту информацию правильно и быстро использовать для повышения качества модельного прогноза.

Для систем прогноза состояния океана, работающих в оперативном режиме, критичным становится вопрос времени решения уравнений модели и выполнения алгоритмов усвоения для построения среднесрочных и краткосрочных прогнозов. Интервал выдачи некоторых современных спутниковых данных наблюдений составляет  $O(1)$  часов. Высокое пространственное разрешение моделей океана позволяет моделировать поведение вихревых структур, а усвоение спутниковых данных наблюдений дает возможность своевременно их обнаружить, что, как следствие, позволяет предсказывать такие природные явления, как штормы, интенсивные вихри и течения.

Решение задачи прогноза состояния океана невозможно без использования параллельных вычислений на компьютерах с распределенной памятью в силу огромного объема получаемой и обрабатываемой информации. Поэтому при разработке и внедрении методов усвоения данных наблюдений необходимо решить задачу разработки масштабируемых алгоритмов и программных кодов.

**Целью диссертационной работы** является:

1. Разработка параллельного алгоритма усвоения данных наблюдений на основе динамико-стохастического подхода в модели динамики океана высокого пространственного разрешения.
2. Тестирование алгоритма усвоения данных спутниковых и дрейфтерных наблюдений с применением модели динамики океана ИВМИО для региона Северной Атлантики.

**Научная новизна.** Предложенная система усвоения данных наблюдений, основанная на ансамблевом подходе, имеет эффективную параллельную реализацию для компьютеров с распределенной памятью и практически не зависит от математической формулировки модели динамики океана. Система усвоения реализована в качестве программного сервиса усвоения данных вычислительной платформы CMF3.0 для использования в моделях высокого пространственного разрешения на массивно-параллельных компьютерах, пришедшего на смену использовавшемуся ранее интегрированному в модель океана методу многомерной оптимальной интерполяции.

Представленная в работе система усвоения данных океанических наблюдений, состоящая из математической модели динамики океана ИВМИО разрешения 0.1 градуса, программного комплекса совместного моделирования CMF3.0 и работающего на его базе сервиса усвоения данных наблюдений на основе метода ансамблевой оптимальной интерполяции, является первой в России системой, способной работать на пространственных глобальных сетках с разрешением 0.1° по горизонтали.

**Научная и практическая значимость.** Разработанная система усвоения данных позволяет эффективно усваивать различные данные спутниковых и дрейфтерных наблюдений и корректировать данные модельных расчетов, значительно повышая качество моделирования. Эффективность параллельной реализации программного сервиса усвоения данных (DAS) подтверждена тестами на современных суперкомпьютерах. С применением DAS в модели Северной Атлантики ИВМИО с пространственным разрешением 0.1° были усвоены данные спутниковой альтиметрии AVISO и данные о температуре и

солености с дрейфтеров ARGO. Показано, что ошибки прогноза после усвоения по сравнению с контрольным расчётом уменьшаются почти в два раза и в целом эффективность всей системы находится на мировом уровне в плане качества моделирования и параллельной масштабируемости. Предложенная система усвоения данных наблюдений на базе модели океана ИВМИО ориентирована на использование в среднесрочном и долгосрочном прогнозе.

**Личный вклад автора.** Представленная диссертация является самостоятельным законченным трудом автора. Научные результаты диссертации, выносимые на защиту, получены лично автором, являются новыми. Результаты других авторов, упомянутые в тексте диссертации, а также полученные в соавторстве, отмечены соответствующими ссылками и принадлежат соавторам в равной степени.

Основные положения, выносимые на защиту:

1. Разработан параллельный алгоритм многомерной оптимальной интерполяции (MVOI) для усвоения данных наблюдений ARGO в модели динамики океана ИВМИО.
2. Создана программная архитектура и разработан параллельный алгоритм ансамблевой оптимальной интерполяции (EnOI), масштабируемый для нерегулярных по пространству спутниковых и дрейфтерных данных наблюдений.
3. Написана программная реализация параллельного алгоритма усвоения данных EnOI в виде программного сервиса вычислительной платформы совместного моделирования CMF3.0. Вычислительная эффективность и параллельная масштабируемость разработанной системы подтверждена тестами на различных современных суперкомпьютерах массивно-параллельной архитектуры.
4. На основе серии численных экспериментов для модели динамики Северной Атлантики высокого разрешения ИВМИО с усвоением данных наблюдений о температуре и солености с дрейфтеров ARGO и данных спутниковой альтиметрии AVISO получены качественные и количественные оценки точности решения на основе независимых данных наблюдений, которые соответствуют результатам ведущих научных коллективов.

**Достоверность результатов** диссертационной работы обоснована использованием строгих математических выводов со ссылками на статьи других авторов, валидацией результатов прогноза состояния океана по данным

наблюдений, вычислительными экспериментами по масштабируемости разработанной программы и сравнением характеристик решения с результатами, полученными в других работах.

**Публикации.** Было опубликовано **10** работ, из них **6** публикаций, в которых излагаются основные научные результаты диссертации на соискание ученой степени кандидата наук: 3 статьи в журналах из перечня **ВАК** [1-3]; 1 статья в зарубежном сборнике, выпуски которого индексируются **Web of Science** [4]; 2 свидетельства о государственной регистрации программ для ЭВМ [5-6].

**Апробация работы.** Материалы, вошедшие в диссертационную работу, представлялись на следующих научных семинарах, международных и всероссийских конференциях: международная суперкомпьютерная конференция «Научный сервис в сети Интернет: многообразие суперкомпьютерных миров» (Абрау-Дюрсо, 2014); школа-семинар «Современные методы дистанционных исследований и прогноза параметров среды в Арктике» (Таруса, ИКИ РАН, 2015); всероссийская конференция «Комплексные исследования морей России: оперативная океанография и экспедиционные исследования» (Севастополь, ФГБУН МГИ, 2016); всероссийская Конференция «Научный сервис в сети интернет» (Абрау-Дюрсо, 2016); доклад на Ученом совете физического направления ИО РАН (03.06.2016); конференция «Суперкомпьютерные дни в России» (Москва, 2016); конференция «Мировой океан: модели, данные и оперативная океанология» (Севастополь, ФГБУН МГИ, 2016); доклад на семинаре в Гидрометцентре России (18.10.2016). Материалы диссертации содержатся в годовых отчетах о работе Суперкомпьютерного комплекса МГУ имени М.В. Ломоносова и Межведомственного суперкомпьютерного центра Российской академии наук.

**Структура диссертации.** Диссертация состоит из введения, 4 глав, заключения, списка литературы, включающего 135 наименований. Основной текст диссертации изложен на 126 страницах, содержит 25 рисунков.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

Во **введении** обоснована актуальность исследований, проводимых в рамках диссертационной работы, сформулированы ее цели и задачи, описана структура диссертации.

В **первой главе** приводится описание:

- математической модели динамики океана ИВМИО [Ибраев, Хабеев, Ушаков, 2012] высокого пространственного разрешения ( $0.1^\circ$  по горизонтали), позволяющей воспроизводить мезомасштабные вихри для решения задач прогноза состояния океана и исследования климата. Приводятся основные уравнения, граничные условия, кратко описываются используемые численные схемы и параллельный вычислительный алгоритм с перекрытиями для быстрого решения системы уравнений мелкой воды на массивно-параллельных компьютерах с распределенной памятью.
- модели термодинамики морского льда CICE5.1 [3]
- компактной вычислительной платформы для моделирования CMF3.0 (*eng. Compact Modeling Framework*), которая является программным комплексом для решения задач совместного моделирования Земной системы и ее компонентов с высоким разрешением на массивно-параллельных компьютерах [Калмыков дис., 2013]

Параллельный алгоритм усвоения рационально реализовать на базе CMF3.0, что дает возможность удовлетворить высоким требованиям по параллельной и вычислительной эффективности, условию абстрактности (минимальной зависимости метода усвоения от модели), а также позволит в дальнейшем использовать усвоение данных в различных моделях–компонентах Земной системы.

Во **второй главе** приводится обзор научной литературы и существующих систем усвоения данных наблюдений за океаном. Рассматриваются сильные и слабые стороны каждого из подходов к решению задачи усвоения, обосновывается выбор ансамблевого метода усвоения данных. Анализируются источники данных наблюдений за океаном, которые планируется использовать: спутниковые данные температуры поверхности океана (ТПО) и данные спутниковой альтиметрии, а также данные с дрейфтеров ARGO.

**Третья глава** посвящена описанию разработанных вычислительных параллельных алгоритмов усвоения данных на основе многомерной оптимальной интерполяции (MVOI) и ансамблевой оптимальной интерполяции (EnOI), представлены особенности их программной реализации для массивно-параллельных компьютеров с распределенной памятью. На описанные в этой главе программные продукты оформлены свидетельства о Государственной регистрации программ для ЭВМ [5], [6].

**Раздел 3.2** посвящен описанию основных уравнений, используемых при усвоении данных динамико-стохастическими методами [Evensen, 2009]:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}_{\text{obs}} - \mathbf{H}\mathbf{x}_b) \quad (3.1)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (3.2)$$

В формулах (3.1), (3.2):

$\mathbf{x}_a$ ,  $\mathbf{x}_b$  ( $n$ ) – векторы модельного решения после и до усвоения (анализа и прогноза, *eng. analysis and background*) размера  $n$ , где  $n$  – количество точек модельной сетки, умноженное на число корректируемых модельных величин (температура, соленость, уровень, скорости и т.д.), имеет порядок  $10^8$  для модели Мирового океана с разрешением 0.1 градуса;  $\mathbf{y}_{\text{obs}}$  ( $m$ ) – вектор наблюдений размера  $m$ , где  $m$  – число точек наблюдения, имеющее порядок  $10^3$ – $10^5$ . Может содержать различные данные наблюдений, такие как температура, соленость, уровень океана и т.д.;  $\mathbf{K}(n \times m)$  – весовая матрица (*eng. Kalman gain matrix*);  $\mathbf{R}(m \times m)$  – ковариационная матрица инструментальных ошибок наблюдений, которая имеет диагональный вид, так как предполагается, что инструментальные ошибки некоррелированы;  $\mathbf{H}(m \times n)$  – матрица проектирования значений модели в пространство данных наблюдений;  $\mathbf{B}$  – ковариационная матрица состояния модели.

В **разделе 3.3** описывается метод многомерной оптимальной интерполяции (*eng. multivariate optimum interpolation, MVOI*), впервые предложенный в работе [Гандина, 1961] и который до сих пор широко используется в системах прогнозирования погоды [Douville et al, 2000] и океана [Smith, Cummings, 2012]. В методе MVOI ковариационная матрица  $\mathbf{B}$  аппроксимируется функцией от координат. В работе [1] используется общая ковариационная матрица для температуры и солености, элементы которой задаются функцией расстояния между точками. Поскольку эта корреляционная зависимость строится одновременно для обеих величин, то это позволяет назвать этот общий метод – методом многомерной оптимальной интерполяции.

В **разделе 3.4** предлагается разработанный параллельный алгоритм для метода MVOI. Предположим для матрицы  $\mathbf{B}$ , что ковариация между достаточно удалёнными друг от друга точками сетки пренебрежимо мала. В модели динамики океана ИВМИО размеры двумерных процессорных подобластей метода декомпозиции [Калмыков, Ибраев, 2013], на которые разбивается вся область, варьируются в пределах 500 – 5 000 км в зависимости от числа используемых вычислительных ядер. Следовательно, можно считать, что  $B_{ij} \neq 0$ , только если точки  $i, j$  находятся в одной или соседних процессорных



подобластях (т.е. матрица  $B$  сильно разреженная). Это позволяет вычислять матрицу  $K$  локально для каждой подобласти. Численные эксперименты, выполненные в работе [1], показали приемлемую масштабируемость такого параллельного алгоритма усвоения данных, но выявили и ряд недостатков. В частности, неравномерность данных наблюдений приводит к падению ускорения при увеличении числа процессорных ядер, т.к. в одних процессорных подобластях данных наблюдений оказывается существенно меньше, чем в других. Другой же недостаток – это значительный объем памяти, занимаемый матрицами  $K$  и  $(HBN^T+R)$ . Также были выявлены некоторые архитектурно-программные проблемы, связанные с необходимостью чтения данных наблюдений всеми процессорными ядрами, на которых выполняется программный код модели океана ИВМИО, и, как следствие такой концепции, должно выполняться усвоение данных.

**Раздел 3.5** посвящен ансамблевым методам усвоения данных. Основная идея методов ансамблевого фильтра Калмана (*eng. Ensemble Kalman Filter, EnKF*) и ансамблевой оптимальной интерполяции (*eng. Ensemble Optimal Interpolation, EnOI*) заключается в том, что ковариационная матрица  $B$  не задается в явном виде или в виде функции, как в методе MVOI [1], а получается из ансамбля векторов состояния модели (выборки).

$$A_b = [x_b^1 \dots x_b^N] - [\bar{x}_b \dots \bar{x}_b] \quad (3.3)$$

$A_b$  – матрица размера  $n \times N$ , где  $N$  – количество элементов ансамбля ( $N \sim 10^2$ ), столбцы которой равны значениям состояния модели минус среднее по ансамблю. При этом ансамбль состояний модели в методе EnKF строится по модельным расчетам с разными начальными условиями. В методе EnOI для ансамбля состояний выбираются модельные расчеты за разные даты, сделанные при долговременном интегрировании модели [2].

Тогда матрица ковариации модели  $B$  аппроксимируется так:

$$B^{en} = \frac{1}{N-1} A_b (A_b)^T \quad (3.4)$$

В статье [Sakov et al, 2015] проводится сравнение EnKF и EnOI для модели океана MOM4 разрешения  $0.1^\circ$  и при значительно возрастающих (пропорционально  $N$ ) вычислительных затратах, точность прогноза с использованием EnKF выше лишь на 10 – 17 %.

В **разделе 3.6** дается математическая формулировка метода EnOI и приведен алгоритм его практического применения на основе работы [Evensen,

2003]. Соотношение между  $n$ ,  $m$  и  $N$  для рассматриваемой модели ИВМИО разрешения  $0.1^\circ$  следующее:

$$N \ll m \ll n, N \sim 10^2, m \sim 10^4, n \sim 10^8 \quad (3.5)$$

Введем матрицу  $S$ , в столбцах которой содержится проекция элементов ансамбля в точки данных наблюдений:

$$S = H A_b, (m \times N) \quad (3.6)$$

$\mathbf{d}$  – вектор инновации данных наблюдений:

$$\mathbf{d} = \mathbf{y}_{obs} - H \mathbf{x}_b, (m) \quad (3.7)$$

Также введем матрицу  $E$  ( $m \times N$ ) – ансамблевое представление матрицы  $R$ .  $E$  – псевдослучайная матрица, которая задается на основе требуемого уровня дисперсии и радиуса ковариации для конкретных данных наблюдений [Evensen, 2003].

$$R = \frac{EE^T}{N - 1} \quad (3.8)$$

Тогда из уравнений (3.1), (3.2) и (3.4) получаем:

$$\mathbf{x}_a = \mathbf{x}_b + A_b S^T (SS^T + EE^T)^{-1} \mathbf{d} \quad (3.9)$$

Ниже для формулы (3.9) приведены размеры матриц и векторов:

$$n = (n) + (n \times N) (N \times m) (m \times m) (m) \quad (3.10)$$

Для обращения симметричной матрицы  $(SS^T + EE^T)$  можно использовать ее спектральное разложение, но для его вычисления требуется порядка  $m^3$  операций с плавающей точкой.

$$SS^T + EE^T = Z \Lambda Z^T, (m \times m), O(m^3) \quad (3.11)$$

Такой подход неприемлем для большого количества данных наблюдений. Поэтому для больших  $m$  используется другой подход.

$$H A_b A_b^T H^T + EE^T = (H A_b + E)(H A_b + E)^T \quad (3.12)$$

В (3.12) учтено, что ансамбли модельных прогнозов и ошибок измерений не коррелируют, что эквивалентно предположению о том, что ошибки модельного прогноза не связаны с ошибками данных наблюдений:

$$H A_b E^T \equiv 0$$

Вычислим сингулярное разложение матрицы  $(S+E)$  ( $m \times N$ ) и левую матрицу собственных векторов  $U$ . Вычислительная стоимость такой процедуры составляет  $O(mN^2)$  операций для  $m > N$ .

$$S+E = U \Sigma V^T, m \times N, O(mN^2) - \text{операций} \quad (3.13)$$

Из (3.12) и (3.13) получаем:

$$SS^T + EE^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T \quad (3.14)$$

Здесь  $\Sigma\Sigma^T$  будет идентично левому верхнему квадрату  $(N \times N)$  матрицы  $\Lambda \sim \Sigma\Sigma^T$ ,  $(N \times N)$ , что соответствует  $N$  ненулевым собственным значениям. Кроме того,  $N$  сингулярных векторов, содержащихся в  $U$ , эквиваленты первым собственным векторам в  $Z$ .

Таким образом, вектор анализа можно получить из вектора прогноза по следующей формуле:

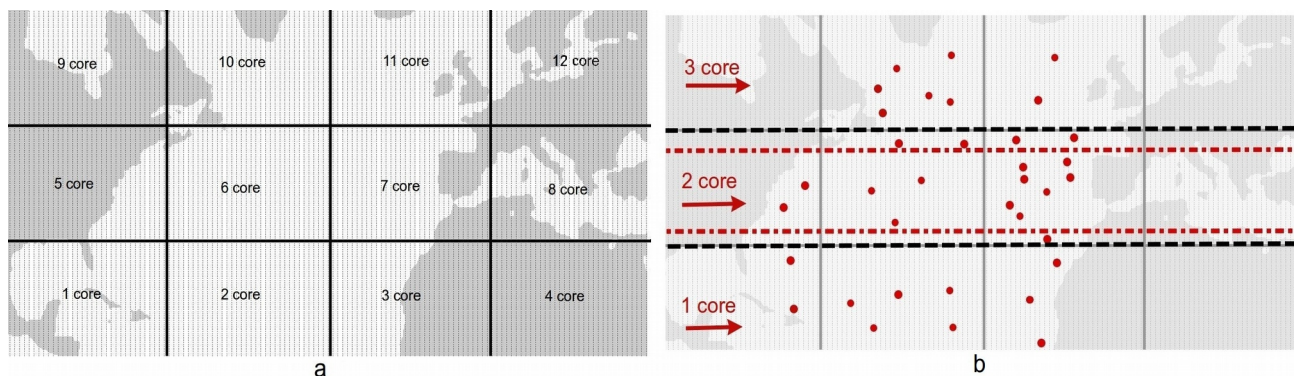
$$\mathbf{x}_a = \mathbf{x}_b + A_b S^T U \Lambda^{-1} U^T \mathbf{d} \quad (3.15)$$

Вычислительная стоимость метода растет линейно от размера задачи (параметров  $n$  и  $m$ ):

$$O((m+n)N^2) \text{ – операций} \quad (3.16)$$

В **разделе 3.7** предлагается оригинальный параллельный алгоритм метода EnOI.

Модель океана ИВМИО имеет параллельную реализацию для компьютеров с распределенной памятью на основе принципа декомпозиции области (**Рис. 3.1а**), где каждую подобласть обчисляет отдельное процессорное ядро. Но такой подход оказался не совсем подходящим для усвоения данных (см. **раздел 3.4**), и поэтому в дальнейшем при усвоении данных наблюдений используется другая процессорная декомпозиция области (ПДО) и отдельные от модели океана процессорные ядра.



**Рис. 3.1. (а)** Двумерная процессорная декомпозиция области, используемая в модели океана ИВМИО (показана серыми прямоугольниками);  
**(б)** одномерная процессорная декомпозиция, используемая при усвоении данных, показана красным и черным пунктиром. Красными точками показаны данные наблюдений.

Цветом будет обозначаться максимальный размер объекта (матрицы или вектора).

$$n \sim 10^8; \quad m \sim 10^4; \quad N \sim 10^2 \quad (3.17)$$

В параллельной реализации будут использоваться три ПДО:

- Двумерная ПДО для модели океана (**Рис. 3.1а**,  $^{2d}x$  использует ядра океана)
- Две одномерные ПДО ( $^{1d}x$  использует ядра, занимающиеся усвоением) :
  - равномерная по точкам модельной сетки (черный пунктир, **рис. 3.1b**)
  - равномерная по данным наблюдений (красный пунктир, **рис. 3.1b**)

Пусть изначально у нас есть вектор модельного решения и вектор данных наблюдений:

$$^{2d}x_b \quad y_o \quad (3.18)$$

Вектор модельного решение пересылается в блок усвоения данных.

$$^{2d}x_b \rightarrow ^{1d}x_b \quad (3.19)$$

Вычисляется инновация данных наблюдений.

$$^{1d}y_o \rightarrow ^{1d}d = ^{1d}y_o - Hx_b \quad (3.20)$$

Вычисляется матрица **S** по формуле (3.6).

$$^{1d}A_b \rightarrow ^{1d}S \quad (3.21)$$

Далее вычислим сингулярное разложение матрицы **(S+E)** и левую матрицу собственных векторов **U**. Благодаря тому, что данные наблюдений (и, как следствие, блоки матрицы **S** и **E**) распределены равномерно по процессорным ядрам, нагрузка на эти ядра распределяется также равномерно.

$$^{1d}(E+S) = ^{1d}U \Sigma V^T \quad (3.22)$$

$$\Sigma \Sigma^T \rightarrow \Lambda^{-1} \quad (3.23)$$

Далее параллельно выполняются операции, описанные формулой (3.15):

$$^{1d}G_1 = \Lambda^{-1} ^{1d}U^T \quad (3.24)$$

$$G_2 = ^{1d}G_1 ^{1d}d \quad (3.25)$$

$$^{1d}G_3 = ^{1d}U G_2 \quad (3.26)$$

$$G_4 = ^{1d}S^T ^{1d}G_3 \quad (3.27)$$

Отметим, что лишь в следующей формуле для оптимизации нагрузки на вычислительные ядра блока усвоения данных требуется использовать равномерную по точкам сетки ПДО.

$$^{1d}x_a = ^{1d}x_b + ^{1d}A G_4 \quad (3.28)$$

Вектор  $x_a$  отправляется в модель океана и раскладывается на ее двумерную процессорную декомпозицию:

$${}^1d\mathbf{x}_a \rightarrow {}^2d\mathbf{x}_a \quad (3.29)$$

В **разделе 3.8** описаны аспекты программной реализации параллельного алгоритма и приведена принципиальная схема передачи данных между моделью океана, сервисом усвоения данных, сервисом работы с файловой системой в рамках программного комплекса совместного моделирования CMF3.0. Блоку усвоения данных дано название DAS (*eng. Data Assimilation Service*) [4]. Программный сервис DAS исполняется на отдельных от модели океана вычислительных ядрах. Это позволяет лучше структурировать систему совместного моделирования, где каждая программная компонента решает свою задачу. При этом блок усвоения данных программно не зависит от модели океана: при усвоении используются лишь модельные результаты расчётов в виде векторов-элементов ансамбля, на основе которых аппроксимируется ковариационная матрица модели В. Данные от модели океана поступают в сервис (обычно один раз в модельные сутки) без обращения к файловой системе, что важно, поскольку размер трехмерных массивов состояния составляет несколько гигабайт.

Данные наблюдений (спутниковые или данные дрейфтеров) распределены крайне неравномерно по расчётной области океана, поэтому для их обсчета используется своя процессорная декомпозиция, равномерная по этим данным (**Рис. 3.1b**). Для реализации алгоритма, описанного выше, используются вызовы функций из пакетов BLAS и LAPACK (реализованные в Intel MKL) через API библиотеки Global Arrays, что существенно упростило написание межпроцессорных обменов.

В **четвертой главе** обсуждаются результаты работы параллельного алгоритма усвоения на данных спутниковой альтиметрии и данных о температуре и солености с дрейфтеров ARGO в модели динамики океана ИВМИО для Северной Атлантики. Экспериментально настраиваются параметры метода усвоения. Приводится качественная и количественная оценка эффективности работы системы. Оценивается параллельная эффективность предложенного подхода. Выполняется сравнение с независимыми данными наблюдений и результатами других научных коллективов. Анализируется, как усвоение аномалии уровня океана влияет на те параметры, которые не усваиваются непосредственно – температуру и соленость.

В **разделе 4.2** приводится описание численных экспериментов по тестированию параллельного алгоритма усвоения EnOI для Северной Атлантики в модели динамики океана ИВМИО.

Было проведено 4 основных численных эксперимента с атмосферным воздействием, соответствующим периоду с 01.05.2008 по 29.06.2008 (60 дней). Сводная информация по этим экспериментам приведена в **таблице 4.1**. *A01* — контрольный эксперимент (усвоение данных не выполняется); *A02* — усвоение данных дрейфтеров ARGO методом MVOI; *A03* — эксперимент с усвоением данных дрейфтеров ARGO методом EnOI; *A04* — усвоение данных альтиметрии AVISO методом EnOI. Во всех экспериментах модельное решение ежедневно сравнивается со спутниковыми данными альтиметрии AVISO и абсолютно независимыми от них данными дрейфтеров ARGO о температуре и солености.

В экспериментах *A02*, *A03* в качестве  $x_a$ ,  $x_b$  (см. **раздел 3.2**) используются векторы модельного решения, составленные из трехмерных полей температуры и солености;  $y_{obs}$  — вектор данных наблюдений состоит из данных дрейфтеров ARGO поступивших за модельные сутки (порядка  $10^2$  профилей в сутки). Также в экспериментах *A02*, *A03* в матрице  $B$  не будет учитываться вертикальная корреляция, так как данные наблюдений ARGO доступны для всех горизонтов на глубине до 1500 метров.

**Таблица 4.1.** Описание экспериментов по усвоению данных наблюдений ARGO и AVISO в северной Атлантике.

Эксп.	Усвоение T и S (ARGO)	Усвоение уровня	Сравнение с ARGO	Сравнение с AVISO
A01	-	-	+	+
A02	MVOI	-	+	-
A03	EnOI	-	+	-
A04	-	EnOI	+	+

В эксперименте *A04* в качестве  $x_a$ ,  $x_b$  используются векторы модельного решения, составленные из трехмерных полей температуры и солености и двумерного поля уровня океана (размер вектора  $1.2 \cdot 10^8$ ).  $y_{obs}$  — вектор данных наблюдений, состоящий из данных аномалии уровня (*eng. SLA, Sea Level Anomaly*) AVISO со спутника Jason-1 ( $\sim 10^3$  точек). В ковариационной матрице  $B$  будет учитываться корреляция (связь) между различными модельными величинами: аномалией уровня океана, температурой и соленостью на различных модельных горизонтах. Таким образом, усвоение данных аномалии уровня будет корректировать весь вектор модельного решения.

В **разделе 4.3** для проверки корректности работы реализованного метода выполнен эксперимент по усвоению синтетических данных наблюдений, полученных на основе известного состояния океана, которое будем считать «истинным». На основе контрольного эксперимента А01 были получены малочисленные ежесуточные профили температуры и солёности ( $Y_{sod}$ , *eng. synthetic observation data*). Затем проведены два эксперимента - Т01 и Т02 (**таблица 4.2**), в каждом из которых используется начальное условие, соответствующее модельному решению за 1 мая другого модельного года (2000-05-01), но атмосферный форсинг используется за май 2008 г., таким образом имитируется возмущение начального условия. В эксперименте Т02 будем ежедневно усваивать данные  $Y_{sod}$ .

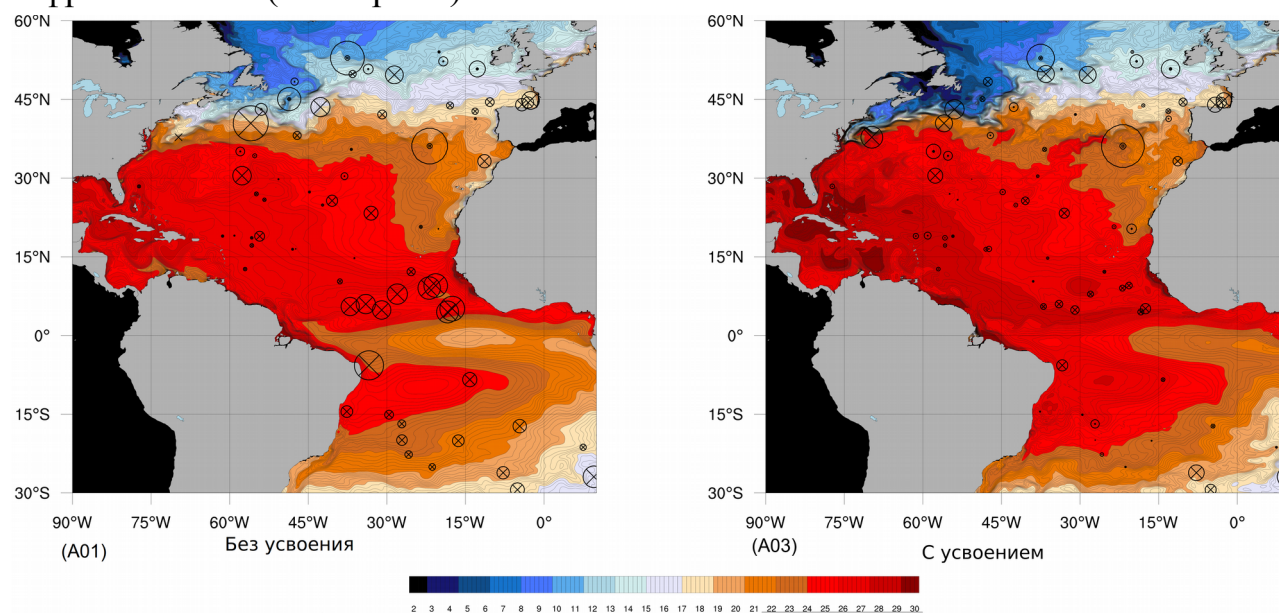
**Таблица 4.2** Описание экспериментов с использованием синтетических данных наблюдений  $Y_{sod}$ .

Эксп.	Начальное условие.	Использование данных наблюдений $Y_{sod}$	Атм. форсинг
А01	2008-05-01	Ежесуточное сохранение данных $Y_{sod}$	COREII 2008-05.01 — 2008-05-21
Т01	2000-05-01	Не используются	
Т02	2000-05-01	Ежесуточное усвоение $Y_{sod}$ методом EnOI	

Приведены графики среднеквадратичной и средней ошибок модельного поля ТПО в экспериментах Т01 и Т02 по отношению к полю ТПО А01, посчитанных по всем расчетным узлам модельной сетки, которые показывают, что усвоение  $Y_{sod}$  методом EnOI значительно приближает модельное поле в эксперименте Т02 к модельному полю А01. Это подтверждает корректную работу реализованного метода усвоения данных, использование которого на малочисленных данных наблюдений ( $Y_{sod}$ ) способно приблизить модельное решение к неизвестному «истинному» состоянию океана (А01).

В **разделе 4.4** выполняется качественный анализ модельного решения. Результаты экспериментов представлены на **рис. 4.2**. Показаны температура поверхности океана (ТПО) в базовом эксперименте (А01) и эксперименте с усвоением методом EnOI (А03). Кругами показано расположение дрейфтеров ARGO, профили температуры и солёности с которых поступили на 2008-06-29. Размер кругов пропорционален разнице между температурой с дрейфтеров и модельной температурой. Крестик в круге означает, что модельная температура ниже температуры с дрейфтера, точка в круге - модельная температура выше. В А03 хорошо воспроизводится струйный и фронтальный характер течения

Гольфстрим, его меандрирование, которое прослеживается севернее мыса Гаттерас. Анализируется решение с усвоением данных (A03), в котором не нарушается характер течения Гольфстрим, а именно струйный (в южной части) и фронтальный (в северной).



**Рис. 4.2** Температура поверхности океана в °C в модели Северной Атлантики за 2008-06-29 в базовом эксперименте (A01) и эксперименте с усвоением методом EnOI (A03).

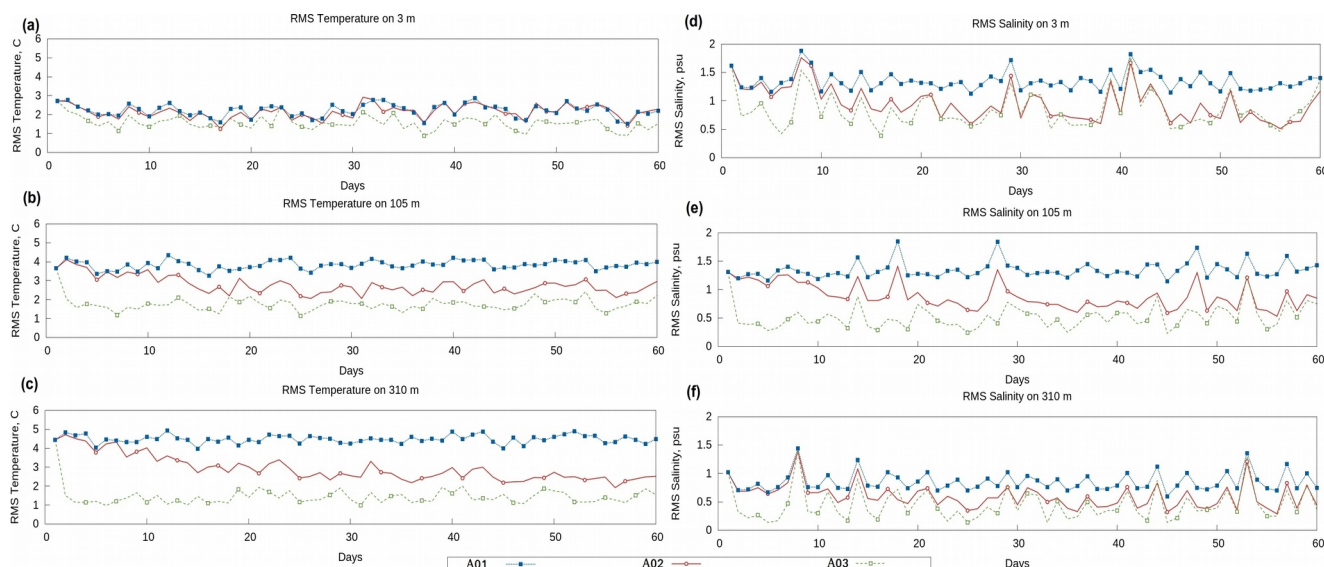
В **разделе 4.5** для количественной оценки эффекта от усвоения данных с дрейфтеров ARGO использована функция ошибки – среднеквадратичное отклонение модельного решения от данных наблюдения, посчитанное по всем дрейфтерам за конкретный день.

На **рис. 4.3** показано изменение этой функции с течением модельного времени для различной глубины (3, 105, 310 метров). Ошибка прогноза  $cost(\mathbf{H}\mathbf{x}_b - \mathbf{y}_o)$  считается каждый день в момент до усвоения данных методом EnOI (A03) или методом MVOI (A02). Ошибка контроля  $cost(\mathbf{H}\mathbf{x}_b - \mathbf{y}_o)$  без усвоения считается каждый день относительно данных наблюдения, но их усвоение не происходит (A01).

При сравнении графиков ошибки контроля для базового эксперимента (A01) и ошибки прогноза для экспериментов с усвоением (A02, A03) можно сделать вывод, что усвоение дает правильный знак коррекции на всех глубинах. Разница между ошибкой прогноза и контроля увеличивается с глубиной, при этом эффективность метода EnOI выше, чем у MVOI. Следует особо отметить, что близко к поверхности океана, на горизонте 5 метров, усвоение методом MVOI неэффективно из-за быстро меняющейся со временем температуры поверхности океана (ТПО) и сильного влияния атмосферного форсинга.



К концу эксперимента A03 (метод EnOI) среднеквадратичная ошибка прогноза температуры становится порядка  $1.5^\circ\text{C}$ , в то время как в эксперименте A02 (метод MVOI) она составляет порядка  $2.5^\circ\text{C}$ . Усвоение солёности также приносит пользу на всех горизонтах.



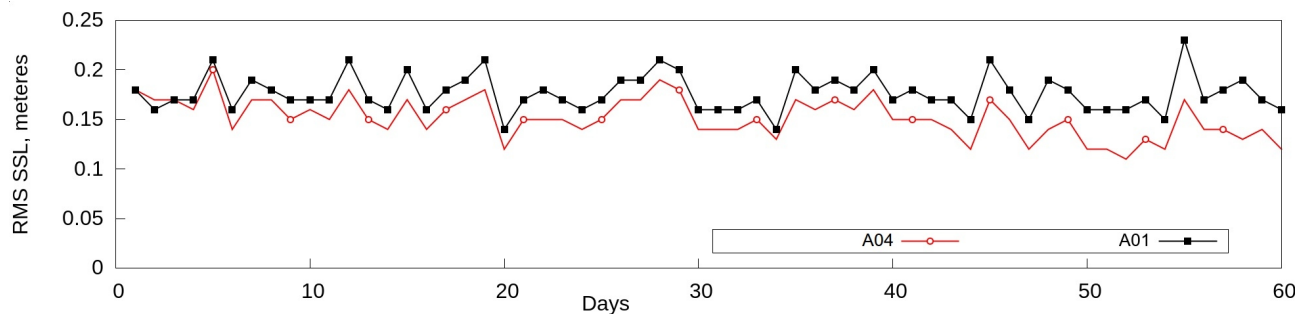
**Рис. 4.3.** Среднеквадратичная ошибка контроля для контрольного эксперимента (A01), ошибка прогноза для метода MVOI (A02) и ошибки прогноза для метода EnOI (A03) для поля температуры в  $^\circ\text{C}$  на глубине 3 (a), 105 (b) и 310 (c) метров и для поля солёности в psu на глубине 3 (d), 105 (e) и 310 (f) метров. По оси X — дни, по оси Y — ошибка температуры в  $^\circ\text{C}$  или ошибка солёности в psu.

В **разделе 4.6** выполнен анализ результатов усвоения данных аномалии уровня AVISO. На **рис. 4.4** показано изменение ошибки уровня океана относительно данных AVISO в зависимости от модельного времени. Приведены графики ошибок прогноза  $cost(\mathbf{H}\mathbf{x}_b - \mathbf{y}_o)$ , которая считается каждый день в момент до усвоения данных AVISO методом EnOI (A04), и контроля  $cost(\mathbf{H}\mathbf{x}_b - \mathbf{y}_o)$  (без усвоения данных, A01). Из графиков видно, что через 60 дней усвоения ошибка в уровне океана уменьшается примерно на 25 % (с 0.16 до 0.12 метров).

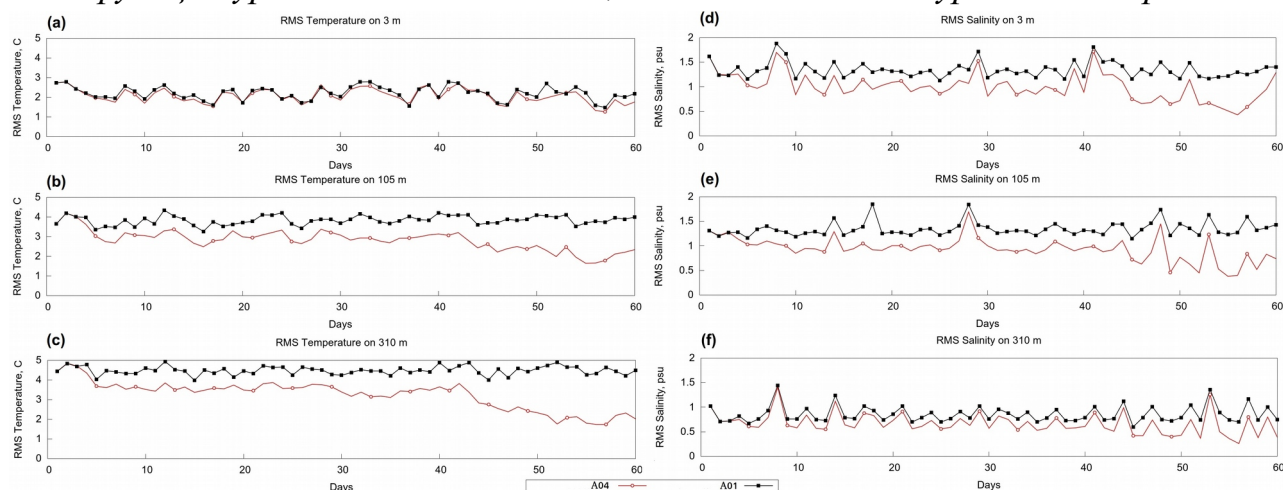
На **рис. 4.5** показано изменение функции ошибки прогноза (A04) в момент до усвоения данных альтиметрии и функции ошибки контроля (A01) для температуры и солёности по сравнению с независимыми (для этого эксперимента) данными ARGO по всем профилям, поступившим за текущие сутки для расчетной области. Ошибка считается каждый день для различной глубины (3, 105, 310 метров).

При сравнении графиков ошибки контроля (A01) и ошибки прогноза для экспериментов с усвоением уровня океана (A04) можно сделать вывод, что

усвоение дает правильный знак коррекции на всех глубинах для температуры, солености и для уровня океана, несмотря на то, что используются только данные альтиметрии.



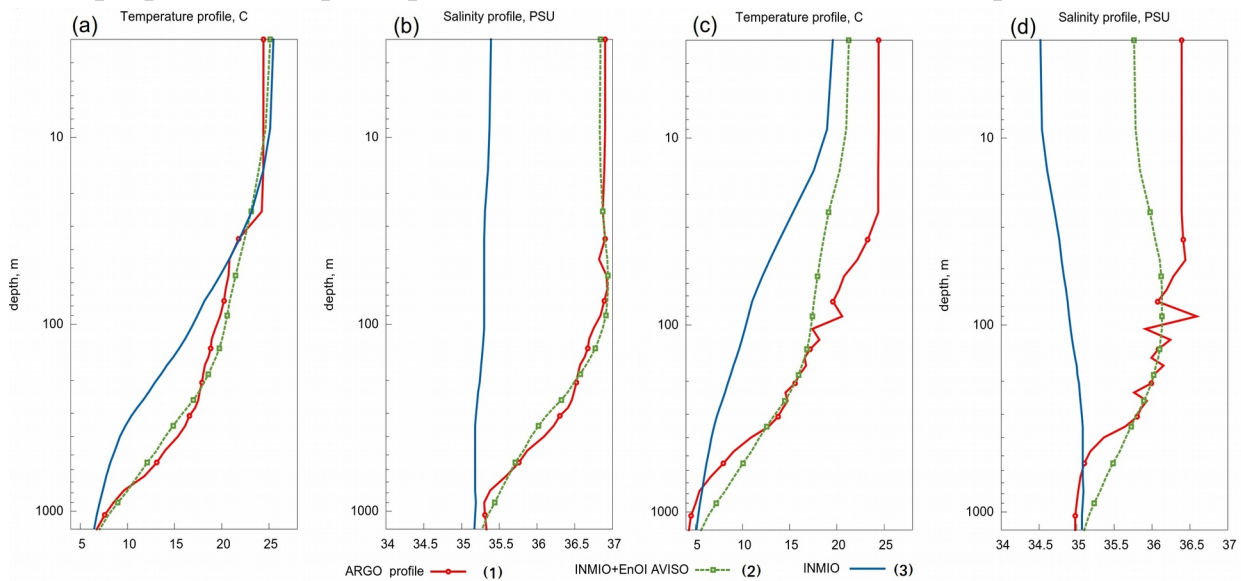
**Рис. 4.4.** Среднеквадратичная ошибка контроля для эксперимента без усвоения (A01) и ошибка прогноза при усвоении данных AVISO методом EnOI (A04) для функции уровня. По оси X — дни, по оси Y — ошибка уровня в метрах.



**Рис. 4.5.** Среднеквадратичная ошибка контроля для эксперимента без усвоения (A01) и ошибка прогноза при усвоении данных альтиметрии методом EnOI (A04) для поля температуры в °C на глубине 3 (a), 105 (b) и 310 (c) метров и для поля солености в psu на глубине 3 (d), 105 (e) и 310 (f) метров при сравнении с данными дрейфтеров ARGO. По оси X — дни, по оси Y — ошибка температуры в °C или ошибка солености в psu.

На рис. 4.6 приведены профили температуры и солености для двух дрейфтеров ARGO (D1 и D2) и модельные профили для экспериментов A01 (контрольный расчет) и A04 (усвоение данных AVISO), интерполированных в точки расположения дрейфтеров (данные за 2008-06-29). Дрейфтеры выбраны таким образом, чтобы D1 был в открытом океане, а D2 – в области течения Гольфстрим. Показано, что на глубине от 3 до 1000 метров в эксперименте с усвоением данных альтиметрии (A04) значительно лучше воспроизводится вертикальный профиль температуры и солености, представленный данными ARGO по сравнению с экспериментом без усвоения (A01), как для области

открытого океана вдали от интенсивных течений, так и для области течения Гольфстрим, в которой происходят активные динамические процессы.

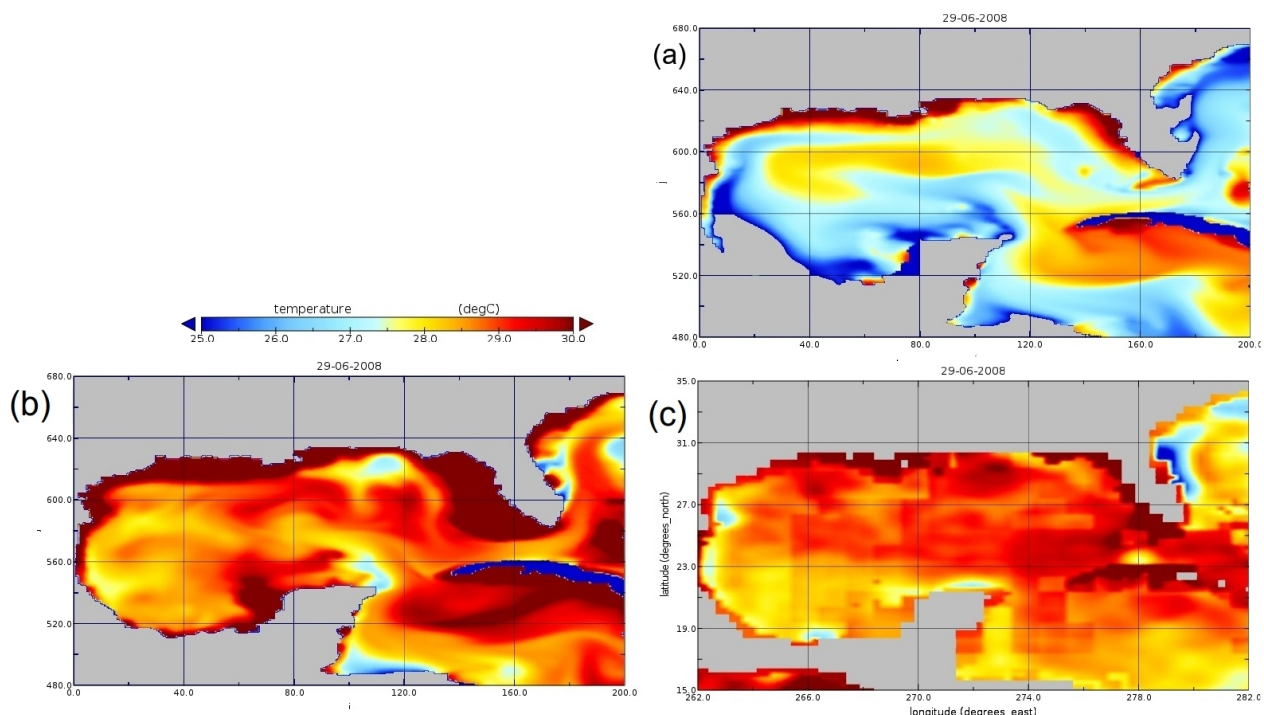


**Рис. 4.6.** Профили температуры и солёности (1) двух дрейфтеров ARGO D1 (a,b) и D2 (c,d); (2) модельные из эксперимента A04 с усвоением альтиметрии AVISO; (3) модельные из контрольного эксперимента A01

В **разделе 4.7** результаты сравниваются с опубликованными в других работах. В работе [Tanajura et al, 2014], в которой проводились эксперименты с тем же методом усвоения, но для модели NUCOM, показано, что усвоение аномалии уровня океана уменьшает ошибку прогноза для независимых наблюдений, в частности, для буев PIRATA по температуре на поверхности на 0.3 градуса в среднем, причем уменьшение ошибки имеет место и для глубоких слоев, до 500м включительно. Аналогичный эффект имеет место и для солёности, где величина коррекции ошибки прогноза достигает 0.5psu. В статье [Oke, 2010], где усвоение данных также выполняется методом EnOI в системе BODAS, ошибка температуры на глубине от 0 до 1500 метров по сравнению с данными дрейфтеров ARGO находится в пределах 1.5 градусов, что меньше на два градуса, чем для модели без усвоения, при этом в качестве данных для усвоения используются спутниковая ТПО и уровень океана, а также данные самих дрейфтеров ARGO. Таким образом, уменьшение ошибки модельного прогноза за счет усвоения данных спутниковой альтиметрии, полученное в данной работе методом EnOI, соответствует мировому уровню.

В **разделе 4.8** приведены результаты экспериментального исследования чувствительности функции ошибки к количеству элементов в ансамбле (параметр N) и коэффициенту на диагонали матрицы R (ошибок данных наблюдений) при усвоении данных ARGO для реализованного метода EnOI.

В **разделе 4.9** проведено сравнение с независимыми спутниковыми данными ТПО. На **рис. 4.7** показано, что ТПО в районе Мексиканского залива в эксперименте A03 (**Рис. 4.7b**) превышает значение ТПО контрольного эксперимента A01 (**Рис. 4.7a**), при этом разница в среднем составляет около  $2^{\circ}\text{C}$ . При сравнении с независимыми спутниковыми данными ТПО (*ARMOR 3d*, **рис. 4.7c**) видно, что расчетные поля A03 заметно ближе к наблюдениям, разница в среднем составляет меньше  $1^{\circ}\text{C}$ . Это свидетельствует о том, что схема усвоения совместно с моделью корректно приближает поля модели к измерениям, при этом сохраняя имеющийся в модельном решении баланс тепла и массы.



**Рис. 4.7.** ТПО в  $^{\circ}\text{C}$  в модели Северной Атлантики для Мексиканского залива за 2008-06-29; **(a)** в базовом эксперименте (A01); **(b)** в эксперименте с усвоением методом EnOI (A03) **(c)** спутниковые данные наблюдений *Armor3d*

**Раздел 4.10** посвящен анализу параллельной эффективности реализованного алгоритма. Благодаря эффективной реализации метода EnOI в качестве параллельного программного сервиса DAS, решение задачи усвоения данных масштабируется практически линейно.

В **заклучении** приводятся основные результаты диссертационной работы – разработан параллельный алгоритм и создана программная архитектура системы усвоения спутниковых и дрейфтерных данных наблюдений на основе динамико-стохастического подхода (MVOI, EnOI) в модели динамики океана высокого пространственного разрешения.

Получены следующие частные результаты:

1. Разработан параллельный алгоритм многомерной оптимальной интерполяции (MVOI) для усвоения данных наблюдений ARGO в модели динамики океана ИВМИО.
2. Создана программная архитектура и разработан параллельный алгоритм ансамблевой оптимальной интерполяции (EnOI), масштабируемый для нерегулярных по пространству спутниковых и дрейферных данных наблюдений.
3. Написана программная реализация параллельного алгоритма усвоения данных EnOI в виде программного сервиса вычислительной платформы совместного моделирования CMF3.0. Вычислительная эффективность и параллельная масштабируемость разработанной системы подтверждена тестами на различных современных суперкомпьютерах массивно-параллельной архитектуры («Ломоносов», «МВС-10П», «РСК Торнадо»).
4. На основе серии численных экспериментов для модели динамики Северной Атлантики высокого разрешения ИВМИО с усвоением данных наблюдений о температуре и солености с дрейферов ARGO и данных спутниковой альтиметрии AVISO получены качественные и количественные оценки точности решения на основе независимых данных наблюдений, которые соответствуют результатам ведущих научных коллективов.

### **Основные публикации автора по теме диссертации**

1. **Кауркин М.Н.**, Ибраев Р.А., Беляев К.П. Усвоение данных наблюдений в модели динамики океана высокого пространственного разрешения с применением методов параллельного программирования // Метеорология и гидрология. 2016. №7. С. 47 – 57 (ВАК, WoS, РИНЦ)
2. **Кауркин М.Н.**, Ибраев Р.А., Беляев К.П. Усвоение данных АРГО в модель динамики океана с высоким разрешением по методу ансамблевой оптимальной интерполяцией (EnOI) // Океанология. 2016. Т.56. №6. С. 852 – 860 (ВАК, WoS, РИНЦ)
3. Громов И.В., Коромыслов А.Ю., Ушаков К.В., **Кауркин М.Н.**, Ибраев Р.А. Совместная модель внутригодовой изменчивости циркуляции вод и льда Северного Ледовитого океана // Труды Гидрометцентра России. 2016. Выпуск 361. С. 29–46 (ВАК, РИНЦ)
4. **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. Eds. Voevodin V., Sobolev S. Series: Communications in Computer and Information Science. Springer – 2016. V. 687. P. 57 – 66 (*WoS*)

5. **Кауркин М.Н.**, Ибраев Р.А. Программа усвоения данных температуры, солености и спутниковой альтиметрии в модель мирового океана методом оптимальной интерполяции // ФГБУ "ГИДРОМЕТЦЕНТР РОССИИ". Свидетельство о государственной регистрации программы для ЭВМ №2015661002. Дата регистрации: 14.10.2015.

6. **Кауркин М.Н.**, Ибраев Р.А. Программа усвоения данных наблюдений методом EnOI для модели океана высокого пространственного разрешения на базе платформы CMF3.0 для моделирования на массивно-параллельных вычислительных системах // ФГБУ "ГИДРОМЕТЦЕНТР РОССИИ". Свидетельство о государственной регистрации программы для ЭВМ №2017610805. Дата регистрации: 18.01.2017.

7. **Кауркин М.Н.**, Тучкова Н.П., Беляев К.П., Михайлов Г.М., Ибраев Р.А., Сальников А.Н. Параллельное усвоение данных наблюдений в гидродинамических моделях высокого пространственного разрешения // «Научный сервис в сети Интернет: многообразие суперкомпьютерных миров». Труды Международной суперкомпьютерной конференции. 2014. С. 236–239 (*РИНЦ*)

8. **Кауркин М.Н.**, Ибраев Р.А. Разработка технологии усвоения данных наблюдений на основе ансамблевых фильтров Калмана в модель океана сверхвысокого пространственного разрешения. Материалы молодежной научной конференции «Комплексные исследования морей России: оперативная океанография и экспедиционные исследования». 2016. С. 76 – 81 (*РИНЦ*)

9. **Кауркин М.Н.**, Ибраев Р.А., Коромыслов А.Ю. Технология усвоения данных аэрокосмического мониторинга и измерений буев Арго методом EnOI в модели глобального океана высокого пространственного разрешения с использованием платформы CMF // Суперкомпьютерные дни в России: Труды международной конференции (26 – 27 сентября 2016 года, г. Москва) – М.: Изд-во МГУ, 2016. С.732 – 740 (*РИНЦ*)

10. **Кауркин М.Н.**, Ибраев Р.А., Беляев К.П., Тучкова Н.П. Параллельный сервис усвоения данных наблюдений в гидродинамическую модель высокого разрешения // «Научный сервис в сети Интернет». Труды XVIII Всероссийской научной конференции. ИПМ им. М.В. Келдыша РАН. 2016. С. 188–193 (*РИНЦ*)