

# ПАРАЛЛЕЛЬНОЕ УСВОЕНИЕ ДАННЫХ НАБЛЮДЕНИЙ В ГИДРОДИНАМИЧЕСКИХ МОДЕЛЯХ ВЫСОКОГО ПРОСТРАНСТВЕННОГО РАЗРЕШЕНИЯ

М.Н. Кауркин<sup>2,3</sup>, Г.М. Михайлов<sup>1</sup>, К.П. Беляев<sup>2,1</sup>, Р.А. Ибраев<sup>3,2</sup>, А.Н. Сальников<sup>4,1</sup>, Н.П. Тучкова<sup>1</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Вычислительный центр им. А. А. Дородницына РАН

<sup>2</sup> Федеральное государственное бюджетное учреждение науки Институт океанологии им. П.П. Ширшова РАН

<sup>3</sup> Федеральное государственное бюджетное учреждение науки Институт вычислительной математики РАН

<sup>4</sup> Федеральное государственное образовательное учреждение высшего профессионального образования "Московский государственный университет имени М.В.Ломоносова", факультет ВМиК

В работе рассматриваются две модели циркуляции высокого пространственного разрешения: модель Института Вычислительной математики (модель ИВМ-ИО, Ибраев, 2008) и МРІ-ОМ - модель института метеорологии общества Макса Планка, Германия (Naak, Jurgenclaus, 2004). Модель ИВМ-ИО работает с разрешением 0.1 градуса по горизонтали и 49 уровней по вертикали по всему глобусу. Модель МРІ-ОМ работает с разрешением 0.1 в северных широтах. Поэтому для их эффективной работы используется двумерная декомпозиция области (Калмыков, 2013) (Рис. 1) и параллельный расчет на отдельных процессорных ядрах с помощью МРІ -технологии (параллелизм до 10 тысяч МРІ-процессов). Для выполнения усвоения данных наблюдения и, как следствие, повышения точности даваемого прогноза для данной модели, работающей в таком высоком пространственном разрешении нужен эффективный параллельный алгоритм. В работе приводится пример реализации такого алгоритма и проверка его численной эффективности.

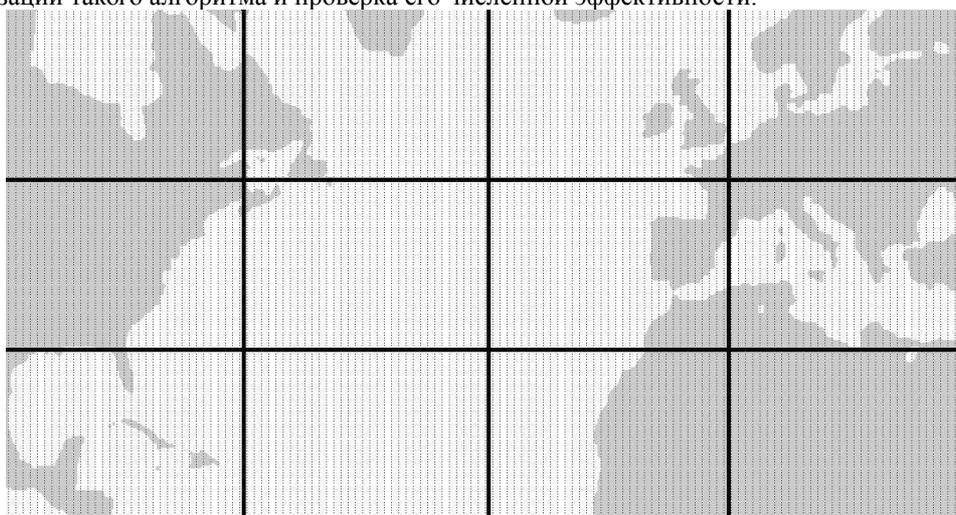


Рис. 1 Метод двумерной процессорной декомпозиции области используемый в модели Мирового океана ИВМ-ИО.

Основное уравнения метода следующее (Гандин, Каган, 1976):

$$X_a = X_b + K(Y - HX_b); K = BH^T(HBH^T + R)^{-1}, \text{ где } X_a, X_b - \text{ векторы модельной величины после и до}$$

усвоения (в английской литературе: analysis and background) размера  $N$  ( $\sim 10^8$ , для Океана с разрешением 0.1 градуса),  $Y$  - вектор наблюдений размера  $M$  ( $\sim 10^3$ , количество точек наблюдения),  $K(N \times M)$  - так называемая, весовая матрица (Kalman gain.),  $B(N \times N)$ ,  $R(M \times M)$ , диагональная),  $H(M \times N)$  - ковариационная матрица полей модели до усвоения (background covariance matrix), ковариационная матрица ошибок наблюдений ( в нашем случае диагональная) и матрица проектирования поля модели в пространство наблюдений, соответственно.

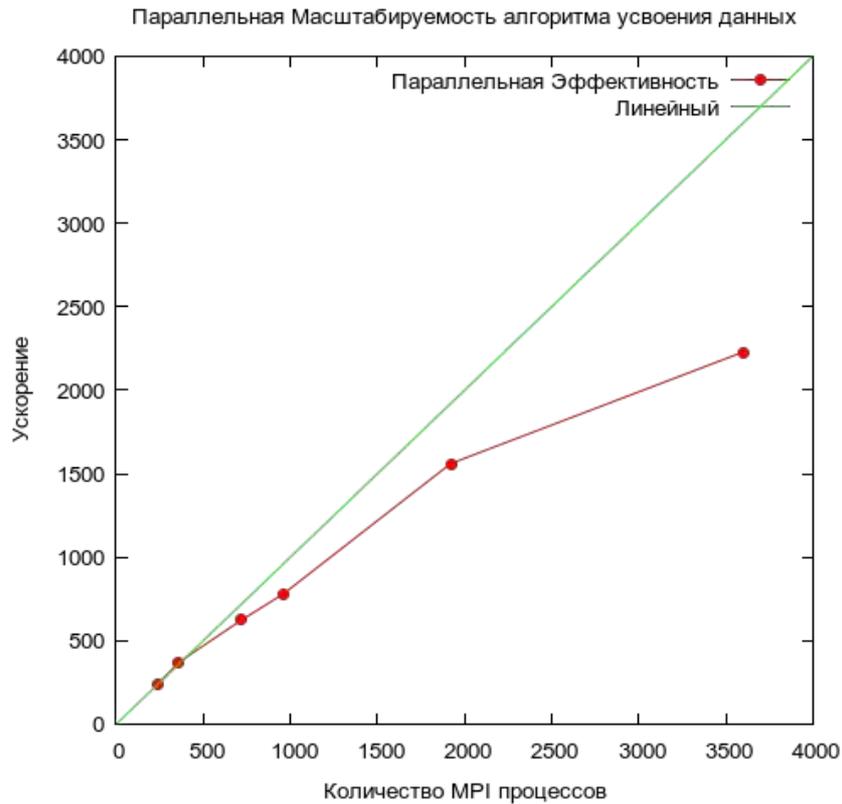


Рис 2. График показывающий масштабируемость параллельного алгоритма усвоения данных.

При усвоении основная проблема заключается в вычислении матриц **B**, **K**.  $B_{i,j}$  элемент матрицы показывает ковариацию между величинами в  $i$  и  $j$  узлах модельной сетки. Исходя из предположения, что ковариация между достаточно удаленными друг от друга точками сетки пренебрежимо мала, следовательно  $B_{i,j} \neq 0$  только если точки  $i$  и  $j$  находятся в одной или соседних процессорных подобластях (т. е. матрица **B** сильно разреженная) т.к. усвоение данных происходит ежедневно, а океаническое течение не способно преодолеть за сутки более 200 км, что меньше подобласти, которую обсмчитывает один MPI процесс ( характерный размер подобласти при 10000 MPI процессах  $\sim 300$  км, при уменьшении количества MPI процессов размер подобласти увеличивается). Это позволят вычислять матрицу **K** локально для каждой подобласти, при условии что мы получим от 8 соседних подобластей их локальные матрицы проекции  $H(M \times N)$  и векторы проекции модельных данных в точки наблюдения  $HX_b (M')$ , где  $M'$ -количество точек наблюдения в локальной подобласти и 8 соседних. Обращение матрицы  $(HbH^T + R)$  (размер  $M' \times M'$ ) - будем выполнять с помощью процедур **sytrf**, **sytri** (вычисление обратной матрицы через LU факторизацию для симметричных матриц) из пакета **Intel MKL**. Это дорогой, прямой метод, требующий порядка  $(2/3 M'^3)$  операций с плавающей точкой, но матрица  $(HbH^T + R)$  уже не является разреженной, при этом имеет не большой размер ( $M'$  не более 1000 ) и уменьшается по мере увеличения количества MPI процессов, поэтому выбор метода не является критичным. На Рис. 2 приведен график показывающий масштабируемость параллельного алгоритма усвоения данных наблюдения. График начинает загибаться из-за сравнительно небольшого количества данных наблюдения, которые распределены не равномерно (что характерно для океанологии в целом по сравнению с Атмосферными моделями).

Для количественной оценки пользы от усвоения данных будем использовать функцию ошибки (cost function, ср. кв. отклонение между модельным решением и данными наблюдения в точках наблюдения). На Рис. 3, 4 будут приведены графики изменения этой функции ( по оси X — дни , по оси Y — ошибка температуры в градусах или ошибка солёности в промилле ).

- **Ошибка контроля**  $cost(HX_b - Y_o)$  считается каждый день в момент до усвоения данных (красный цвет на графиках).
- **Ошибка контроля**  $cost(HX_b - Y_o)$  без усвоения считается каждый день в сравнении с данными наблюдения, но их усвоение не происходит (синий цвет на графиках).
- **Ошибка анализа**  $cost(HX_a - Y_o)$  считается сразу после усвоения данных в модель (зеленый цвет на графиках).

Как видно из графиков ( нужно сравнить ошибка контроля С и БЕЗ усвоения — красный и синий графики соответственно) усвоение дает пользу с точки зрения уменьшения ср. кв. отклонения температурного прогноза модели с 2.5 до 1.5 градусов, а для солёность с 0.7 до 0.3 промилле.

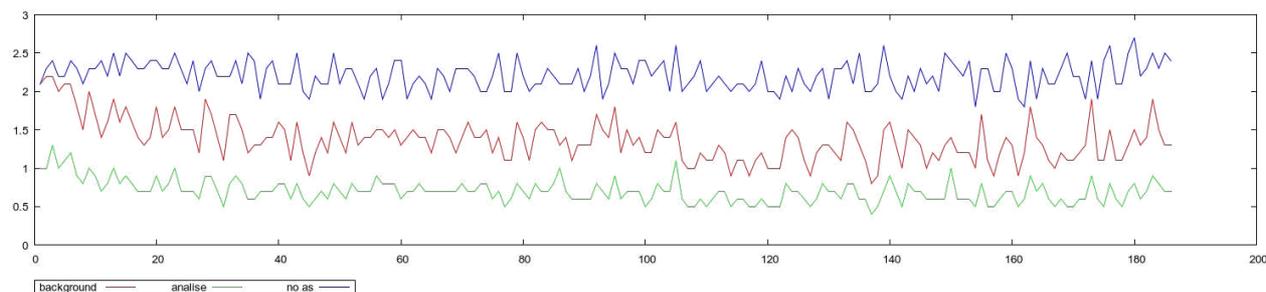


Рис. 3 Средне кв. ошибка анализа (зеленый график) и прогноза (С и Без усвоения данных наблюдения, красный и синий график соответственно) на глубине 100 м для поля температуры.

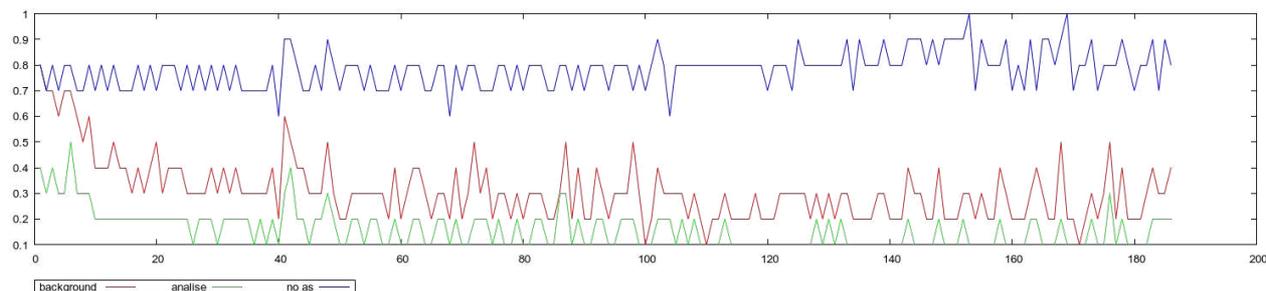


Рис. 4 Средне кв. ошибка анализа (зеленый график) и прогноза (С и Без усвоения данных наблюдения, красный и синий график соответственно) на глубине 100 м для поля солёности.

Для двух вышеперечисленных моделей по приведенным формулам осуществляется усвоение данных и, соответственно, коррекция модельного поля. Усваиваются данные альтиметрии со спутников Jason (см. Рис. 5, 6) и профилей температуры и солёности с дрейфтеров ARGO. Вновь построенные поля анализируются и сравниваются как с модельным расчетом без усвоения (контроль), так и с наблюдениями. Показывается, что в результате усвоения ошибка расчета относительно данных уменьшается. Приводятся соответствующие оценки и рисунки.

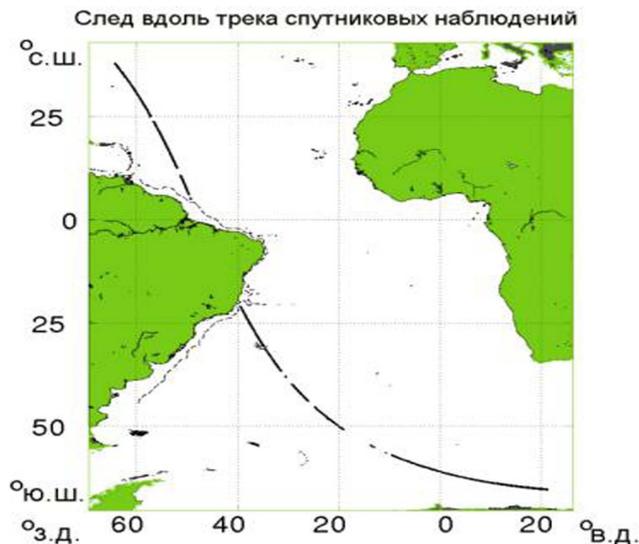


Рис. 5. Пример одного из треков спутника Jason 1 в Атлантике

NRT MSLA – Merged Product  
2002/01/01

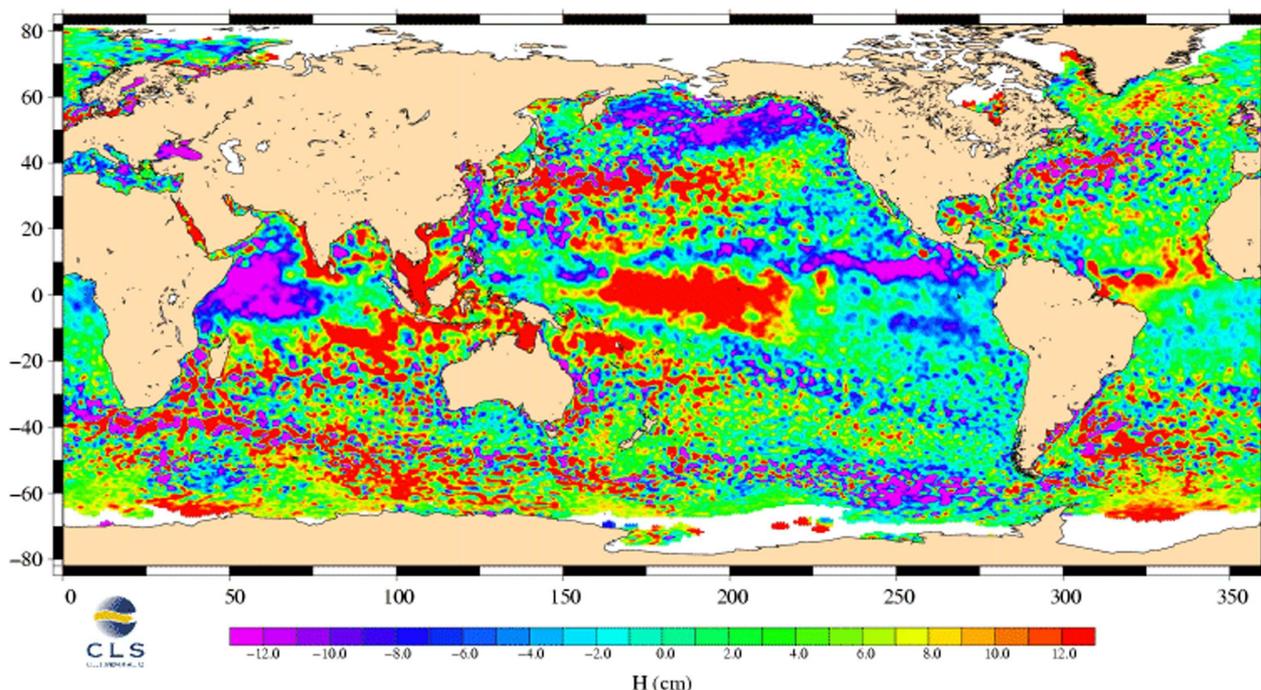


Рис 6. Карта реанализа данных альтиметрии в Мировом океане по данным спутников Jason 1, Jason 2

Кроме того, в работе строятся доверительные интервалы для усвоенных характеристик, в частности для полей уровня океана. Метод построения доверительных характеристик использует теорему для предельного поведения Марковской цепи, полученную в работе (Беляев и Тучкова, 2014). На основе исходных данных строятся величины вторых моментов, которые являются основными параметрами при передаче сигнала от исходных характеристик скорректированным.

Работа поддерживается РФФИ (проекты 14-07-00037, 14-05-00363).

ЛИТЕРАТУРА:

1. Р.А. Ибраев. Математическое моделирование термогидродинамических процессов в Каспийском море. М: ГЕОС, 2008.130с.
2. В.В. Калмыков. Программный комплекс численного моделирования совместной системы океан-атмосфера на массивно-параллельных компьютерах 05.13.18 — математическое моделирование, численные методы и комплексы программ Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. Москва, 2013.
3. Гандин Л.С., Каган Р.Л. Статистические методы интерпретации метеорологических данных. – Л.: Гидрометеиздат, 1976. – 360 с.
4. H. Naak. Simulation of Low-Frequency Climate Variability in the North Atlantic Ocean and the Arctic, Volume 1. Max Planck Institute for Meteorology, 2004.
5. P. Wetzel, H. Naak, J. Jungclaus, E. Maier-Reimer. The Max-Planck-Institute Global Ocean/Sea-Ice. Model [http://www.mpimet.mpg.de/fileadmin/models/MPIOM/DRAFT\\_MPIOM\\_TECHNICAL\\_REPORT.pdf](http://www.mpimet.mpg.de/fileadmin/models/MPIOM/DRAFT_MPIOM_TECHNICAL_REPORT.pdf) (Model MPI-OM. Technical report).
6. К.П. Беляев, Н.П. Тучкова. Предельные распределения для характеристик при усвоении данных наблюдений в стационарном режиме // Информатика и её применения, 2014 (в печати).