Разработка системы моделирования атмосферы с горизонтальным разрешением 10 км*

М.А. Толстых, А.В. Шляева, В.Г. Мизяк
Институт вычислительной математики РАН, Москва
Гидрометцентр России, Москва
МГТУ им. Баумана, Москва

Современная система моделирования атмосферы на временных масштабах от часа до нескольких месяцев состоит из прогностической модели атмосферы и системы усвоения данных наблюдений. Для реализации перспективной системы моделирования атмосферы на массивно-параллельных вычислительных системах решаются две задачи:

1. Разработка глобальной полулагранжевой модели атмосферы нового поколения. За основу принят динамический блок глобальной оперативной полулагранжевой модели атмосферы ПЛАВ [1].

2. Разработка схемы усвоения данных наблюдений для глобальной модели атмосферы на основе локального ансамблевого фильтра Калмана с преобразованием ансамбля (LETKF).

Глобальная полулагранжева модель атмосферы нового поколения

Основной моделью глобального среднесрочного прогноза погоды в России с конца 2009 г. является глобальная полулагранжева модель атмосферы ПЛАВ (полулагранжева, основанная на уравнении абсолютной завихренности) [1]. Модель разработана в Институте вычислительной математики РАН и Гидрометцентре России с помощью консорциума ALADIN/LACE, предоставившего параметризации процессов подсеточного масштаба. Внедрение этой модели позволило примерно в два раза сократить отставание России, по сравнению с лидирующей группой мировых прогностических центров в ошибках прогноза таких важных параметров, как давление на уровне моря, температура на уровне 850 гПа и высота поверхности 500 гПа. Краткое описание параллельной реализации модели приводится в [2].

По современным меркам разрешение оперативной версии модели является весьма грубым, поэтому коллективом разработчиков была реализована новая версия модели, имеющая горизонтальное разрешение порядка 20–25 км, а вертикальное – 51 уровень. Для повышения

* Работа выполнена при частичной поддержке гранта РФФИ 10-05-01066.

параллельной эффективности программного комплекса модели ПЛАВ было выполнено распараллеливание алгоритма постпроцессинга модели (расчета полей на стандартных изобарических поверхностях, которые и являются выходной прогностической продукцией модели). Были выполнены тестовые расчеты на вычислительной системе SGI Altix 4700, имеющей 1664 ядра процессоров Intel Itanium 9100, организованные в узлы по 128 ядер. Эта вычислительная система установлена в ГВЦ Росгидромета. Тестовые расчеты выполнялись для варианта модели с разрешением 0,225° по долготе, 0,17–0,24° по широте (примерно 20 км в средних широтах Северного полушария), 51 неравномерно расположенных вертикальных уровней в немонопольном режиме (запуск через систему очередей). Размерность сетки задачи – 1600x865x51. Значения параллельного ускорения в зависимости от числа используемых вычислительных ядер приведены на рис. 1.

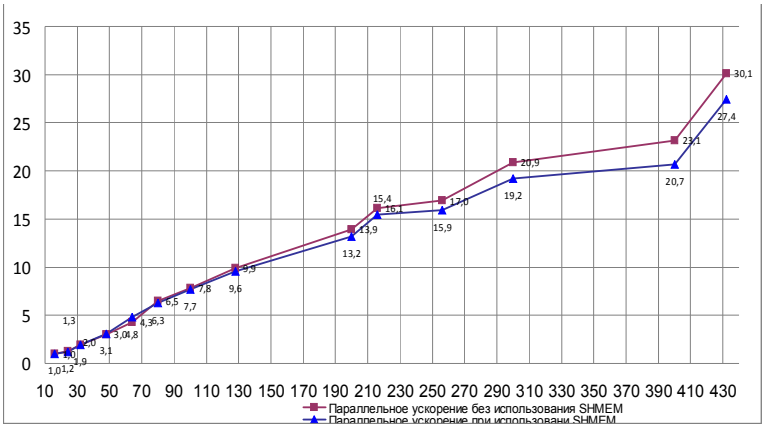

Рис. 1. Параллельное ускорение модели на вычислительной системе SGI Altix 4700 по сравнению со временем расчета на 16 процессорах как функция от количества процессоров

Мы видим, что модель позволяет эффективно использовать 432 процессора. Время расчета прогноза на сутки удовлетворяет требованиям оперативного счета. Однако дальнейшее увеличение числа процессоров не ведет к сокращению времени счета.

Часть вычислений блока решения уравнений динамики атмосферы выполняется в пространстве коэффициентов Фурье по долготе. В этом пространстве производится решение нескольких систем линейных алгебраических уравнений (СЛАУ) методом матричной прогонки. Этот метод требует наличия данных для всех широт на данном процессоре.

В то же время в основной части вычислений, осуществляемой в сеточном пространстве, данные распределены по широтам. Поэтому были выполнены модификации вычислительного алгоритма модели, которые позволили бы производить все вычисления при распределении данных по процессорам по широтам.

Для этого алгоритм векторной прогонки для решения блочно-трехдиагональных СЛАУ, получаемых в результате дискретизации уравнений диффузии вихря и дивергенции, полунеявной схемы интегрирования по времени и уравнения восстановления поля скорости по вихрю и дивергенции, был заменен на алгоритм дихотомии [3]. Двоичный логарифм параллельного ускорения этой части модели как функция двоичного логарифма числа процессоров приведен на рис. 2.

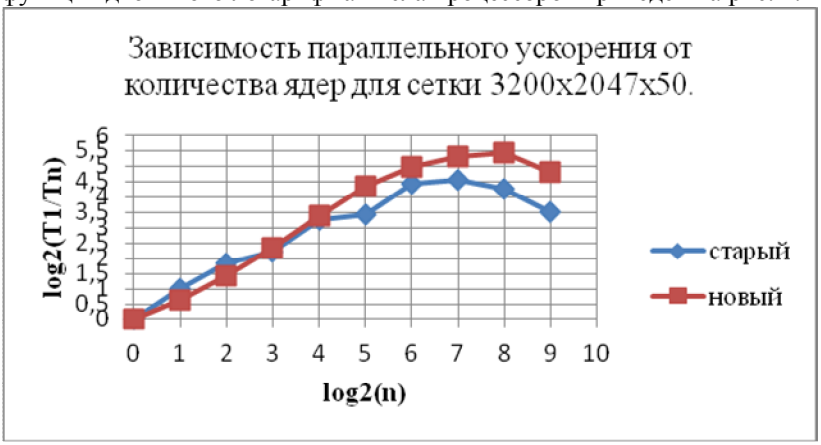

Рис. 2. Зависимость логарифма параллельного ускорения от логарифма количества ядер для сетки 3200x2047x50

Мы видим, что замена алгоритма привела к удвоению количества процессоров MPI, при котором наблюдается рост производительности.

Схема усвоения данных наблюдений для модели ПЛАВ

Система усвоения данных используется для вычисления начальных условий для старта прогноза (также называемых анализом) по краткосрочному прогнозу модели (первому приближению, размерности порядка $10^6 \times 10^8$) и данным наблюдений (размерности порядка 10^7), имеющим априори неизвестные ошибки.

В настоящее время широко распространены две группы методов усвоения данных: вариационное усвоение и усвоение с помощью ансамблевых фильтров Калмана.

Классический фильтр Калмана прогнозирует не только состояние системы, но и матрицу ковариаций. Однако его применение на практике для современных моделей атмосферы невозможно ввиду огромных размерностей вектора состояния. Идея ансамблевых фильтров Калмана [4] заключается в статистической оценке матрицы ковариаций по ансамблю первых приближений, размерность которого значительно меньше размерности модели (порядка $10^1 \times 10^2$):

Такой подход позволяет учесть реальные физические свойства потока, достаточно несложен в реализации, но предполагает дополнительные вычислительные затраты для расчёта большего числа прогнозов.

Нами разработана параллельная реализация схемы усвоения на базе локального ансамблевого фильтра Калмана с преобразованием ансамбля (Local Ensemble Transform Kalman Filter, LETKF) [5]. Одним из достоинств локализации является возможность эффективного распараллеливания алгоритма усвоения по данным, т.к. анализ можно вычислять параллельно в разных точках сетки. В нашей реализации данные (узлы регулярной широтно-долготной сетки) могут распределяться между вычислительными процессами равномерно по широте либо по косинусу широты.

Разработанная реализация была протестирована на задаче усвоения атмосферных данных для модели численного прогноза погоды ПЛАВ [1]. Для тестирования схемы усвоения данных использовалась версия модели ПЛАВ с горизонтальным разрешением 1,4°x1,125°, 28 уровней по вертикали. Радиус горизонтальной локализации – 1500 км, использовались наблюдения наземных станций, радиозондов и спутниковых наблюдений SATOB (общее число наблюдений – около 25000).

Были выполнены тестовые запуски прототипа LETKF с различным числом участников ансамбля (40 и 60) на вычислительной системе SGI Altix 4700.

Известно, что использование односторонних коммуникаций может ускорить работу параллельной программы. Были реализованы три версии коммуникаций: двухсторонние коммуникации MPI-I, односторонняя запись (MPI_PUT) и одностороннее чтение (MPI_GET) из стандарта MPI-II.

Графики параллельного ускорения расчетов при распределении обрабатываемых данных, равномерном по широте, представлены на рис. 3.

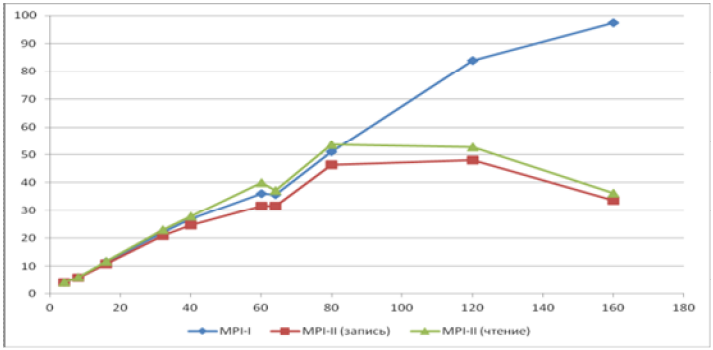

Рис. 3. Ускорение при использовании 60 участников ансамбля

Из графиков видно, что рост ускорения близок к линейному для 60 участников ансамбля. Для 40 участников ансамбля, начиная с числа процессов, равного 100, с ростом числа процессов ускорения не происходит. Это объясняется тем, что вычисления между процессами распределены неравномерно, так как вычислительная сложность задачи зависит от числа обрабатываемых локальных (расположенных на расстоянии не больше заданного) наблюдений, а наблюдения распределены неравномерно.

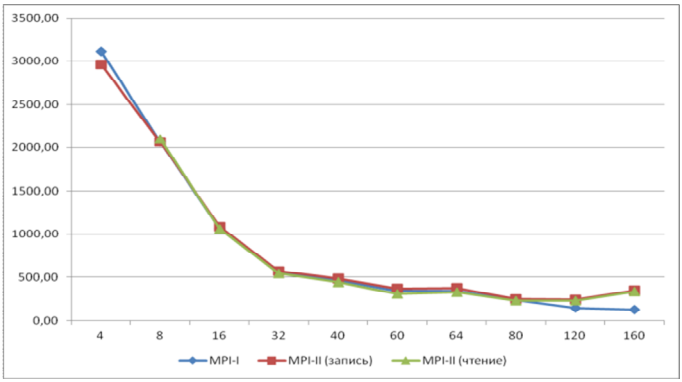

Рис. 4. Время выполнения на SGI Altix при использовании 40 участников ансамбля

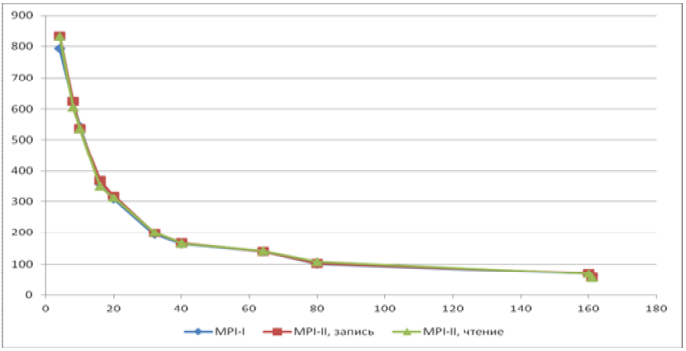

Рис. 5. Время выполнения на «РСК Торнадо» при использовании 40 участников ансамбля

Также были выполнены расчеты анализа для 40 участников ансамбля на пилотной кластерной системе РСК, установленной в октябре 2011 г. в Росгидромете. «РСК Торнадо» состоит из 96 вычислительных узлов, каждый из которых содержит по два шестиядерных процессора Intel Xeon 5680. Коммуникационная сеть построена на базе интерфейса Infiniband QDR. Время выполнения расчетов на SGI Altix и «РСК Торнадо» приведено на рис. 4 и 5 соответственно.

Как видно из графиков времени выполнения, использование односторонних коммуникаций не дало выигрыша во времени.

Выводы

Вычислительная сложность алгоритмов моделей атмосферы и усвоения данных наблюдений диктует необходимость их эффективной параллельной реализации.

Здесь представлены результаты работ по повышению параллельной эффективности программного комплекса модели атмосферы ПЛАВ. Показано, что применение метода дихотомии ведет к повышению в два раза числа процессоров, при котором наблюдается рост параллельного ускорения с ростом числа процессоров.

Нами разработан параллельный алгоритм локального ансамблевого фильтра Калмана для усвоения данных. Тестирование показало, что разработанная реализация позволяет достичь ускорения в 100 раз при распараллеливании на 160 процессах (размерность ансамбля равна 60). При меньшей размерности задачи (размерность ансамбля – 40 участников) эффективное распараллеливание ограничено 100 вычислительными процессами.

Литература

1. **Толстых М.А.** Глобальная полулагранжева модель численного прогноза погоды. – Москва; Обнинск: ОАО ФОП, 2011. – 111 с.
2. **Володин Е.М., Толстых М.А.** Параллельные вычисления в задачах моделирования климата и прогноза погоды // Вычислительные методы и программирование. – 2007. – Т. 8. – С. 113–122.
3. **Terekhov A.** Parallel Dichotomy Algorithm for solving tridiagonal system of linear equations with multiple right-hand sides // Parallel Computing. – 2010. – Vol. 36. – P. 423–438.
4. **Evensen G.** Sequential data assimilation with a nonlinear quasigeostrophic model using Monte-Carlo methods to forecast error statistic // J. Geophys. Res. – 1994. – Vol. 99. – P. 10143–10162.
5. **Hunt B.R., Kostelich E.J., Szunyogh I.** Efficient data assimilation for spatiotemporal chaos: a Local Ensemble Transform Kalman Filter // Physica D. – 2007. – Vol. 230. – P. 112–126.