

УДК 681.3.069; 681.324

МОДЕЛИРОВАНИЕ ЭКСТРЕМАЛЬНЫХ ЯВЛЕНИЙ В АТМОСФЕРЕ И ОКЕАНЕ КАК ЗАДАЧА ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

А. В. Бухановский¹, О. И. Зильберштейн², С. В. Иванов¹, С. В. Ковальчук¹,
Л. И. Лопатухин¹, С. К. Попов², М. М. Чумаков²

Рассматриваются особенности проектирования и реализации высокопроизводительного программного комплекса для моделирования экстремальных явлений в атмосфере и океане. Формулируется модель производительности задачи в целом и выполняется исследование зависимости эффективности распараллеливания от характеристик архитектуры вычислительной системы. Работа выполнена в рамках проекта ФЦП “Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2012 годы” по теме “Высокопроизводительный программный комплекс моделирования и прогноза экстремальных гидрометеорологических явлений и расчета воздействий на морские объекты и сооружения”, госконтракт № 02.514.11.4033. Статья подготовлена по материалам доклада авторов на международной научной конференции “Параллельные вычислительные технологии” (ПаВТ-2008; <http://agora.guru.ru/pavt>).

Ключевые слова: параллельные вычисления, комплексы программ, моделирование климата, параллельные вычислительные системы, моделирование экстремальных явлений.

1. Введение. Освоение пространства Мирового океана (прежде всего шельфовой зоны) требует информации о внешних нагрузках на морские объекты и сооружения. Внешние нагрузки обусловлены совместным действием ветра, волнения и морских течений (в некоторых случаях льда) при определенном уровне моря. Их характеристики подразделяются на оперативные (определяющие режимы эксплуатации объектов и сооружений) и экстремальные (характеризующие режим выживания) [1]. В отличие от оперативных характеристик, традиционно получаемых посредством статистического анализа исторических массивов гидрометеорологической информации, экстремальные характеристики обычно не обеспечены данными наблюдений и, как следствие, должны определяться с использованием инструментария математического моделирования [2, 3].

В практике проектирования морских сооружений в качестве экстремальных обычно рассматриваются расчетные характеристики гидрометеорологических явлений, возможных один раз в T лет, где T соответствует классу сооружения. Для судов и объектов океанотехники традиционно $T = 1 \div 100$ лет, а для некоторых гидротехнических сооружений (как, например, защитные сооружения Санкт-Петербурга от наводнений) могут рассматриваться периоды повторяемости до $T = 10$ тысяч лет. Тем самым, моделирование экстремальных гидрометеорологических явлений является существенно более ресурсоемкой задачей, чем традиционные гидрометеорологические задачи прогноза погоды и моделирования климата (см., например, [4–6]). Обычно при прогнозе погоды оперируют данными на объемной пространственной сетке, однако все расчеты выполняются в синоптическом диапазоне (в среднем от 2 до 5 суток). Задача моделирования климата охватывает временной интервал в десятки и сотни лет, однако она может оперировать декадными или даже среднемесячными данными. В то же время, современная концепция получения информации об экстремальных гидрометеорологических явлениях [7] основана на полностью синтетическом подходе [8]. Этот подход состоит в моделировании ансамбля реализаций пространственно-временных полей основных характеристик (ветрового волнения, скорости ветра и морских течений, уровня моря) за длительный временной интервал (50 и более лет) с дискретностью 0.5–3 часа. На основании этих данных строятся стохастические модели многомасштабной изменчивости, по которым с использованием метода Монте-Карло (порядка миллиона испытаний) предсказываются значения экстремальных сочетаний гидрометеорологических характеристик, возможных один раз в T лет.

¹ Научно-исследовательский институт наукоемких компьютерных технологий, Санкт-Петербургский государственный университет информационных технологий, механики и оптики, Биржевая линия, д. 4, 199034, г. Санкт-Петербург; e-mail: avb_mail@mail.ru, svivanov@mail.ifmo.ru, iterater@mail.ru, leonid-lop@yandex.ru

² Гидрометеорологический научно-исследовательский центр РФ, Большой Предтеченский переулок, д. 11–13, 123242, Москва; e-mail: info@infomarc.com, lmpi@yandex.ru

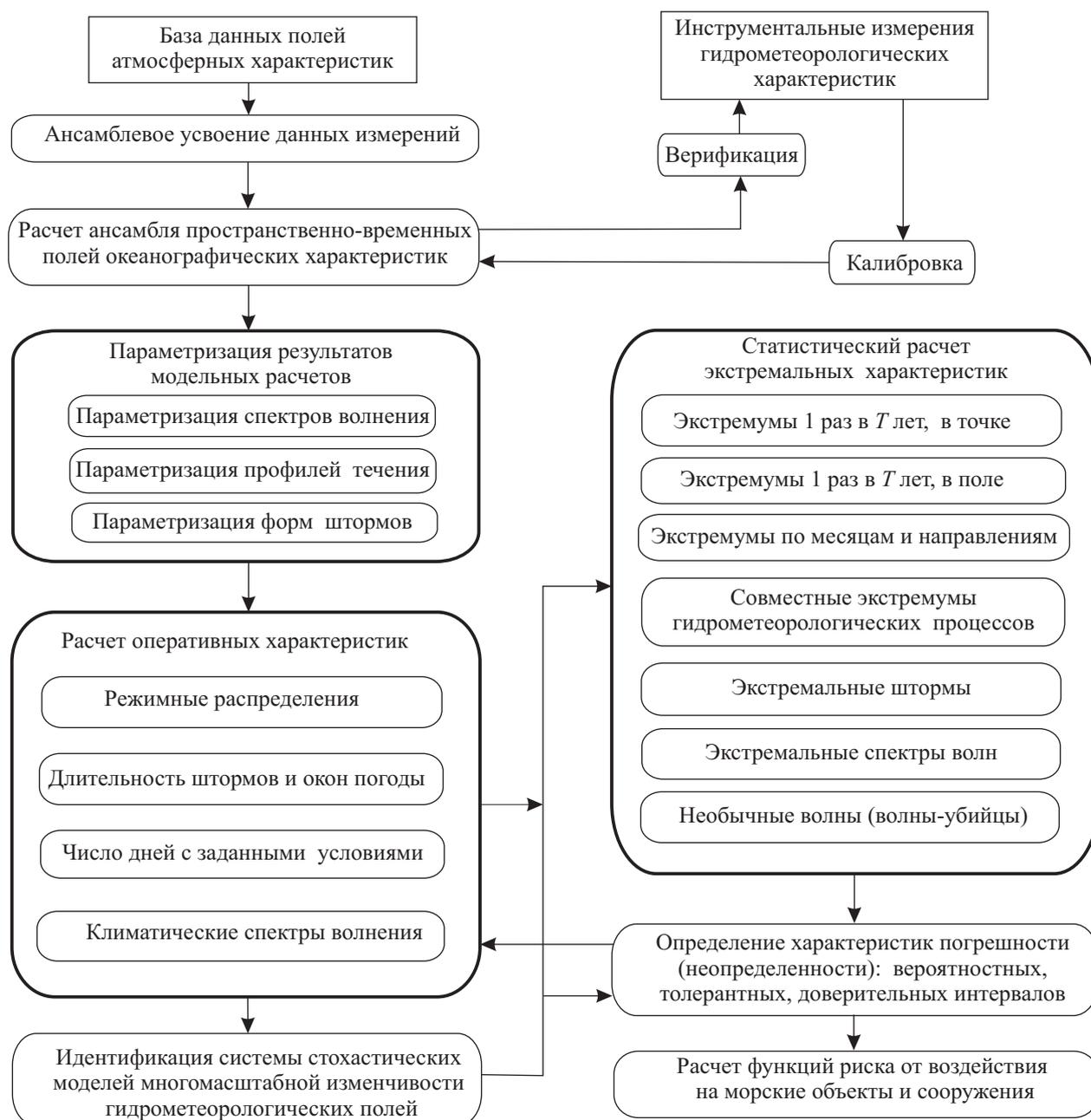


Рис. 1. Комплексная технология моделирования ветра, волнения, течений и уровня моря

Широкое распространение данного подхода сдерживается его ресурсоемкостью. Например, в [9] показано, что ресурсоемкость расчетов характеристик экстремальных явлений с периодом повторяемости не более 100 лет для акватории Каспийского моря составляет около 120 ПФлоп. Иными словами, для выполнения всех расчетов за одни сутки необходима реальная производительность вычислительных мощностей 1.4 ТФлоп. Как следствие, в настоящее время даже в ведущих зарубежных компаниях, выполняющих гидрометеорологические изыскания на шельфе Мирового океана (Ocean Weather (США), OCEANOR (Норвегия), IFREMER (Франция)), используются упрощенные подходы, тем не менее также требующие применения систем терафлопной производительности.

В настоящей статье рассматриваются аспекты разработки отечественного высокопроизводительного вычислительного комплекса, реализующего модельную концепцию получения информации об экстремальных гидрометеорологических явлениях.

2. Этапы технологии моделирования экстремальных гидрометеорологических явлений. Практическая реализация модельной концепции предусматривает последовательное решение четырех за-

дач, объединенных в вычислительную цепочку. Каждая задача дополнительно распадается на систему взаимосвязанных модулей (рис. 1).

2.1. Подготовка массива метеорологической информации (исходных данных). В качестве входных данных для расчета океанографических характеристик (ветрового волнения, течений и уровня моря) могут быть использованы поля атмосферного давления и скорости приводного ветра на основе данных реанализа. Однако для акваторий со сложной орографией (как, например, Каспийское море) данные реанализа существенно занижают экстремальные характеристики (в основном, скорости ветра) в сильных штормах. Это приводит к необходимости уточнения метеорологических полей над морем на основе данных измерений из различных источников (судовые наблюдения, наблюдения на гидрометеорологических станциях, спутниковые данные) с применением технологии ансамблевого усвоения данных.

Ансамблевое усвоение не связано напрямую с моделью динамики атмосферы; оно выполняет синтез обобщенного массива данных на основе нескольких измеренных реализаций одного и того же пространственно-временного поля, выполненных различными методами (и с различной погрешностью). Уравнение ансамблевого усвоения формулируется в виде фильтра Калмана; матричный коэффициент фильтрации Калмана определяется через пространственно-временные ковариационные функции полей атмосферных характеристик. Результатом применения процедуры усвоения данных наблюдений является синтетический массив атмосферных характеристик (атмосферного давления, температуры воздуха и скорости ветра) в узлах регулярной пространственно-временной сетки, шаг которой определяется спецификой расчетной акватории [10].

Процедура усвоения может проводиться как для каждой метеорологической величины независимо, так и путем пересчета одних величин через другие. Так, приводный ветер может быть определен из геострофического соотношения (по полям приземного атмосферного давления с усвоением) с учетом кривизны изобар, агеострофики и приводного трения. Эта методика используется в том случае, если данные реанализа собственно полей ветра отсутствуют или измерений ветра недостаточно для усвоения.

Реализация задачи ансамблевого усвоения является достаточно ресурсоемкой процедурой. В частности, для акватории сравнительного небольшого Каспийского моря в качестве исходных данных рассматривается массив реанализа длительностью 60 лет (1948–2007 гг.), каждые шесть часов на трехмильной пространственной сетке в 126×216 точек. В нем усваивается более 280 тысяч судовых измерений и срочные ряды наблюдений на 20 береговых и островных гидрометеорологических станциях.

2.2. Формирование массива полей океанографических характеристик на основе гидродинамических моделей. Используя массив метеорологической информации как входные данные, выполняется совместное гидродинамическое моделирование волнения, течений и уровня моря, что обусловлено необходимостью учета фактической глубины акватории по разгону волн. В том случае, когда штормовые колебания уровня моря обычно сопровождаются развитым ветровым волнением, этот эффект может привести к появлению существенно больших волн, чем предельно возможные при среднем многолетнем положении уровня моря.

Гидродинамическая модель волнения в спектральной форме представляется в виде уравнения баланса волновой энергии

$$\frac{\partial N}{\partial t} + \frac{\partial N}{\partial \varphi} \dot{\varphi} + \frac{\partial N}{\partial \theta} \dot{\theta} + \frac{\partial N}{\partial k} \dot{k} + \frac{\partial N}{\partial \beta} \dot{\beta} + \frac{\partial N}{\partial \omega} \dot{\omega} = G, \tag{1}$$

где N — спектральная плотность волнового действия; она является функцией от широты φ , долготы θ , волнового числа k и угла β между направлением волнового вектора и параллелью, а также от частоты ω и времени t . Это уравнение связывает между собой явления притока энергии от ветра, диссипации и ее перераспределения и нелинейного взаимодействия между частотными составляющими процесса волнения. Чаще всего функция источника G записывается в виде суммы трех компонент: $G = G_{in} + G_{nl} + G_{ds}$ (поступления энергии от ветра к волнам, слабонелинейного взаимодействия в спектре ветрового волнения и диссипации волновой энергии соответственно).

Для расчетов уровня моря и течений в общем случае используется трехмерная гидродинамическая бароклинная модель со свободной поверхностью. Исходная система уравнений в декартовой системе координат в приближении гидростатики и f -плоскости записывается в виде

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{2}$$

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} (uu) + \frac{\partial}{\partial y} (vu) + \frac{\partial}{\partial z} (wu) - fv = -\frac{1}{\rho} \frac{\partial p}{\partial x} + N_h \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + \frac{\partial}{\partial z} \left(N_z \frac{\partial u}{\partial z} \right), \tag{3}$$

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(uv) + \frac{\partial}{\partial y}(vv) + \frac{\partial}{\partial z}(wv) + fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} + N_h \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) + \frac{\partial}{\partial z} \left(N_z \frac{\partial v}{\partial z} \right), \quad (4)$$

$$\frac{\partial p}{\partial z} = \rho g, \quad (5)$$

$$\frac{\partial T}{\partial t} + \frac{\partial}{\partial x}(uT) + \frac{\partial}{\partial y}(vT) + \frac{\partial}{\partial z}(wT) = \frac{\partial}{\partial z} \left(K_z \frac{\partial T}{\partial z} \right), \quad (6)$$

$$\frac{\partial S}{\partial t} + \frac{\partial}{\partial x}(uS) + \frac{\partial}{\partial y}(vS) + \frac{\partial}{\partial z}(wS) = \frac{\partial}{\partial z} \left(K_z \frac{\partial S}{\partial z} \right), \quad (7)$$

$$\rho = \Psi(T, S, p), \quad (8)$$

где f — параметр Кориолиса; p — давление; ρ — плотность; T — температура; S — соленость; u, v, w — составляющие поля скорости течений по осям (x, y, z) ; N_z, N_h — коэффициенты вертикальной и горизонтальной вязкости; K_z — коэффициент вертикальной диффузии. Горизонтальная турбулентная диффузия не учитывается. На поверхности моря задаются касательные напряжения ветра, кинематическое условие и отсутствие потока плавучести. На дне принимается квадратичный закон трения, задаются кинематическое условие обтекания течением неровностей рельефа дна $H(x, y)$ и отсутствие потока плавучести. Для приливных морей в правой части (3), (4), (6), (7) дополнительно учитывается приливная составляющая.

Уравнения (1)–(8) интегрируются совместно. Реализация задачи расчета полей океанографических характеристик является более ресурсоемкой, чем задача ансамблевого усвоения. Расчет ведется с шагом интегрирования по времени от 5 до 15 минут, по пространству — от 0.5 до 30 миль (в зависимости от акватории).

2.3. Расчет оперативной статистики, идентификация системы стохастических моделей и оценивание экстремальных характеристик. Статистическая обработка информационной базы гидрометеорологических характеристик выполняется средствами многомерного статистического анализа (МСА) пространственно-временных полей. Аппарат МСА предназначен для решения трех задач: снижения мерности, выявления неоднородности и установления зависимости данных.

Для решения первой и второй задач использован общий подход на основе преобразования выборочных данных путем их проецирования на гиперповерхности, определяемые спецификой задачи. Он не только обобщает процедуры МСА, связанные с использованием канонических базисов (анализ главных компонент, канонический корреляционный анализ, канонический дискриминантный анализ) на модели не скалярного пространственно-временного поля, но и упрощает решение задачи установления зависимости. Для описания совокупной изменчивости системы гидрометеорологических полей под влиянием случайных управляющих факторов применен формализм линейных динамических систем, который позволяет с минимальной модификацией перенести технологии классического корреляционного, регрессионного и дисперсионного анализа на модель системы не скалярных пространственно-временных полей (скорости ветра, ветрового волнения, уровня и течений) [11].

Технологии МСА позволяют идентифицировать систему стохастических моделей гидрометеорологических полей. Она предназначена для того, чтобы методом Монте-Карло воспроизвести ансамбль реализаций заданного объема, аналогичный исходному в смысле вероятностного описания. Полученный ансамбль позволяет оценить достоверность точечных оценок (в форме вероятностных, толерантных, доверительных интервалов), а также выполнить измерения вероятностных характеристик в случаях, которые не обеспечены выборочными данными (например, расчет экстремальных значений за интервал T , превышающий непрерывную продолжительность информационной базы).

Учет многомасштабной (мелкомасштабной, синоптической, сезонной, межгодовой) изменчивости требует использования стохастических моделей разных классов. Для воспроизведения океанографических полей в мелкомасштабном (секунды–часы) диапазоне изменчивости применена модель в форме следующей динамической системы с учетом управляющих факторов η :

$$\zeta_v = \sum_{j=0}^N \Phi_j \zeta_{v-j} + \sum_{k=0}^P \Theta_k \varepsilon_{v-k} + \sum_{i=1}^M \Xi_i \eta_{v-i}. \quad (9)$$

Здесь Φ_j, Θ_k — матричные коэффициенты авторегрессии и скользящего среднего, ε — белый шум. Зависимость между η и ζ задается значениями весовой функции (функции отклика) Ξ_i .

Для воспроизведения синоптической изменчивости (перемежаемости штормов и окон погоды) исполь-

зовано импульсное представление пространственно-временного поля в форме

$$\zeta(\mathbf{r}, t) = \sum_j a_j W_j(t, \mathbf{r}) = \sum_j a_j W_j^{(k)}(\mathbf{r}, \Xi_j), \quad \Xi_j = \Xi_j(t), \tag{10}$$

где $W_j^{(k)}(\bullet)$ — пространственно-временные (временные) импульсы (выбросы выше и ниже уровня z), движущиеся вдоль определенной траектории $\mathbf{r} = \mathbf{r}(t)$, возникающие и исчезающие случайным образом. Коэффициенты a_k суть интенсивности импульсов. Форма импульса, управляемая параметром k , определяется задачей исследования; для описания экстремальных явлений допустимо аппроксимировать $W_j(\bullet)$ треугольным импульсом.

Описание сезонной и межгодовой изменчивости требует применения другого класса моделей на основе ортогональных разложений по каноническому базису (естественные ортогональные функции):

$$\zeta(\mathbf{r}, t) = m(\mathbf{r}, t) + \sum_{k=1}^s a_k(t) \phi_k(\mathbf{r}, t) + \varepsilon(\mathbf{r}, t). \tag{11}$$

Скалярные коэффициенты разложения $a_k = a_k(t)$ суть общие факторы, управляющие изменчивостью основных конфигураций поля, а ε — специфический фактор, характеризующий случайные вариации значений поля в каждой точке \mathbf{r} в момент времени t . Модель (11) обобщается для описания регрессионной взаимосвязи исходного поля $\zeta(\mathbf{r}, t)$ с несколькими управляющими факторами — в форме линейной динамической системы (9) относительно коэффициентов их разложения по каноническим базисам. Применение комплекса стохастических моделей (9)–(11) лежит в основе авторского метода оценивания экстремальных явлений BOLIVAR, одобренного для применения Всемирной метеорологической организацией в 2000 г. [12].

2.4. Оценка риска эксплуатации сооружений. Понятие экстремальности сочетаний основных характеристик гидрометеорологических явлений связано не только с вероятностью их появления (которая оценивается на этапе 2.3), но и с возможными последствиями для конкретных морских объектов и гидротехнических сооружений. Проблема выбора наиболее опасных условий в соответствующем районе зависит от конкретной прикладной задачи и допустимого риска, зависящего от набора многочисленных факторов. Построение функции риска требует знания особенностей конкретного проектируемого объекта или сооружения, выраженного в форме математической модели, описывающей изменение его характеристик (динамических, прочностных и пр.) под влиянием внешних возмущений. Структура модели определяется спецификой задачи и целевой переменной. Возможность оценки и ранжирования рисков, связанных с экстремальными гидрометеорологическими условиями, позволяет не только оценить степень возможного ущерба для инженерного объекта или сооружения, но сопоставлять сразу несколько альтернативных объектов с целью выбрать наиболее оптимальный для эксплуатации в заданных условиях (конкретном географическом регионе).

Характеристики ресурсоемкости вычислительных операций

№	Операция	Объем генерируемых данных	Относительное время выполнения
2.1	Усвоение метеорологической информации	1×10^{10}	6 %
2.2	Расчет полей характеристик морского волнения	5×10^{13}	28 %
	Расчет полей течений и уровня моря		40 %
2.3	Параметризация частотно-направленных спектров	3×10^{12}	13 %
	Стохастическое моделирование экстремумов	3×10^{15}	10 %
2.4	Расчет функций риска	5×10^7	3 %

3. Проектирование высокопроизводительного программного комплекса моделирования экстремальных явлений.

3.1. Специфика требований к организации программного комплекса. В таблице приведены оценки основных этапов, рассмотренных в разделе 2, по объемам генерируемых данных и относительному

времени вычислений. Для оценки использовались расчеты на акватории Каспийского моря, выполненные на кластере с 8 двудерными процессорами AMD Opteron 275. Как следствие, практическое использование описанной технологии требует привлечения вычислительных систем терафлопной производительности, имеющих несколько сотен или даже тысяч процессоров (как, например, вычислительная система Гидрометцентра России на основе SGI Altix 4700). При этом программная реализация данной технологии учитывает следующее.

1) В силу специфики задач гидродинамического и статистического моделирования океанографических процессов, применяемые алгоритмы допускают различные способы декомпозиции. Многие существующие на настоящий момент решения ориентированы на специфику современных архитектур с производительностью порядка 100 ГФлопс (кластеры с достаточно небольшим (16–32) количеством вычислителей). Перенос программного комплекса на систему терафлопной производительности с количеством вычислителей $\sim 10^3$ требует выбора альтернативных способов декомпозиции или их комбинаций.

2) Эффективность вычислительного комплекса может быть повышена за счет разумного использования особенностей его архитектуры. Специфика вычислительных алгоритмов и программной реализации должна быть ориентирована под гибридную, или NUMA-архитектуру (Non-Uniform Memory Access), соответствующую кластеру из однородных многопроцессорных узлов с общей памятью. Это учитывается в схеме распределения данных между узлами с целью минимизации обменов и переменной зернистости вычислительных задач, а также в программной реализации двухуровневой схемы распараллеливания на основе совместного использования технологий MPI и OpenMP.

3) Специфическая архитектура вычислительной системы, а также возможность использования ее части другими задачами (например, выполняющими оперативные функции), может привести к дисбалансу вычислительной нагрузки и снижению параллельной производительности. Потому в программную реализацию включен механизм динамической балансировки вычислительной нагрузки, который будет обеспечивать мониторинг параллельно выполняющихся задач, их распределение, а при необходимости — миграцию на другие узлы.

3.2. Сервисно-ориентированная архитектура программного комплекса. Реализация вышеупомянутых требований к программному комплексу упрощается при использовании сервисно-ориентированной архитектуры (SOA). Каждая из указанных на рис. 1 задач представляет собой сервис, обеспечивающий полное выполнение предполагаемых функциональных требований. Независимость сервисов в сочетании с фиксированным интерфейсом позволяет обеспечить возможность их легкой модификации и замены.

На рис. 2 представлена структура сервисов программной системы. Наиболее важным из сервисов с точки зрения рабочего процесса является модуль обработки задачи, использующий вычислительный модуль и модуль контроля вычислительного процесса, поскольку основная нагрузка возлагается именно на них. Указанные модули в свою очередь представимы в виде совокупности сервисов, исполняющихся на

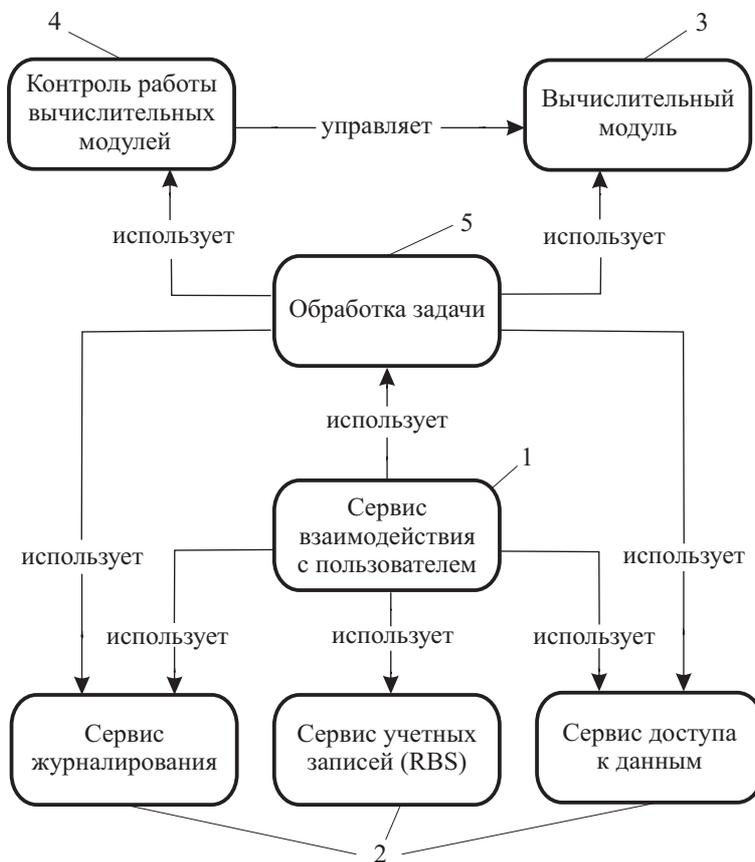


Рис. 2. Структура сервисов программной системы. Условные обозначения: 1 — коммуникационные сервисы (communication), 2 — сервисы сущностей (entity), 3 — сервисы активности (activity), 4 — управляющие сервисы (capability), 5 — сервисы-процессы (process)

отдельных узлах вычислительной системы и обеспечивающих наиболее полное использование вычислительных ресурсов. Модуль обработки задачи организует распределенную работу вычислительных моделей, самостоятельно выполняя декомпозицию задачи для распределенного вычисления и последующую обратную композицию. Следует отметить, что часть модулей присутствует в единственном экземпляре, обеспечивая централизованное управление комплексом моделей в целом, в то время как другая часть выполняется на каждом из узлов вычислительной системы.

Вычислительные модули представляют собой сервисы доступа к моделям, реализующим логику предметной области. С учетом того, что взаимодействие моделей организуется центральным модулем вычислительной системы, на него ложится ответственность за определение потоков данных между сервисами моделей. В то же время потоки данных в рамках сервисов также обладают довольно сложной структурой, обусловленной фиксированным интерфейсом модели. Таким образом, сервис, управляющий каждой из моделей, обязан предоставить интерфейс для унифицированного обращения с моделью, организуя потоки данных таким образом, чтобы эффективно использовать вычислительные ресурсы системы. Этот подход в общем виде представлен на рис. 3.

На диаграмме отражены основные компоненты системы организации распределенных вычислений: конвертер, система контроля удаленных вычислений и деконвертер.

Задача модуля “конвертер” состоит в анализе стандартной входной конфигурации модели и построении на ее основе заданий для удаленных вычислителей. Кроме того, к задачам этого модуля относится предварительное конфигурирование дальнейшего процесса распределенных вычислений.

Модуль “деконвертер” выполняет обратное действие над результатами работы модели на удаленных узлах: сшивку частичных выходных данных, полученных в результате работы каждой из моделей.

Распределенные вычисления осуществляются под управлением подсистемы, состоящей из модуля управления исполнителями и из исполнителей, контролирующей работу моделей на вычислительных узлах. Приведенная схема потоков данных может изменяться, адаптируясь под конкретную модель, однако предоставляемый ею интерфейс остается неизменным, обеспечивая тем самым универсальность, характерную для сервисно-ориентированной архитектуры.

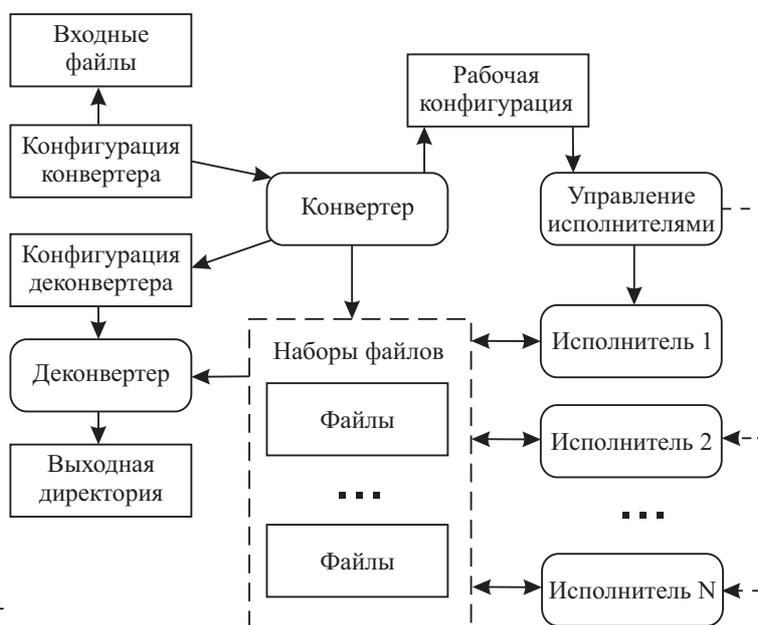


Рис. 3. Потоки данных (workflow) при организации параллельных вычислений

3.3. Иерархическая схема распараллеливания вычислений. Использование сервисно-ориентированного подхода позволяет гибко реализовывать различные схемы параллельных вычислений, допуская как одновременную обработку независимых задач, так и распараллеливание внутри каждой задачи в зависимости от специфики вычислительной архитектуры. Это дает возможность организовывать иерархические схемы параллельных вычислений на системах гибридной архитектуры. Например, на рис. 4 приведена принципиальная схема двухуровневого распараллеливания описанной выше технологии моделирования.

На узле 1 (ввода-вывода) выполняются предварительные операции с данными (например, извлечение данных из базы, подготовка батиметрии и пр.). Затем происходит распределение массива метеорологических характеристик (давление, температура воздуха и льда), объемом T лет, между Π вычислительными узлами с общей памятью. Батиметрия и прочие управляющие параметры моделирования рассылаются на все узлы. После этого выполняется процедура усвоения, а также по полям давления и температуры рассчитывается ветер. Эта процедура на каждом узле распараллеливается по пространству. Выполняется частичная репликация метеорологических данных (с перекрытием от нескольких недель до месяцев) между смежными узлами, что необходимо для последующей сшивки результатов расчетов по гидродина-

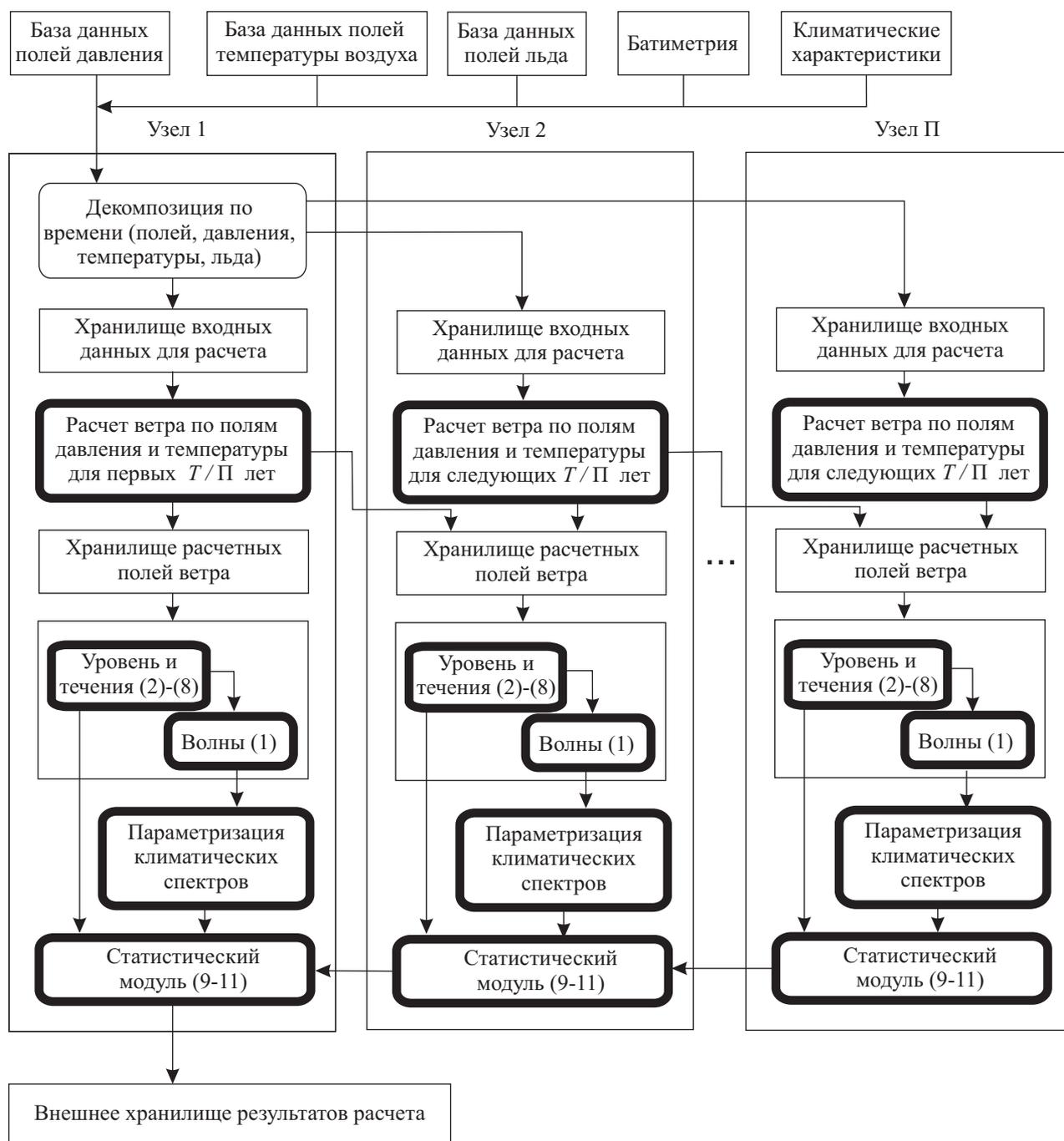


Рис. 4. Иерархическая схема распараллеливания вычислений (утолщенными линиями изображена работа с распараллеливанием по пространству на узлах с общей памятью)

мическим моделям. На каждом узле совместно (на каждом шаге) интегрируются уравнения (2)–(8) и (1), также распараллеленные по пространственной области в рамках модели общей памяти. Затем результаты расчетов передаются в модуль параметризации климатических спектров. Этот модуль распараллеливается внутри каждого узла по пространству с учетом балансировки нагрузки по причине географических (региональных) особенностей спектрального волнового климата. Полученные в итоге поля параметров передаются в статистический модуль (9)–(11), выполняется обработка результатов и расчет экстремумов. Статистические модули на каждом узле также допускают многоуровневое распараллеливание — по ансамблю, по индексующей переменной, по таксонам и на основе принципа перемешивания [13].

4. Анализ параллельной производительности программного комплекса на основе прототипирования. Описанный выше программный комплекс является весьма сложным многокомпонент-

ным объектом. Это требует изучения характеристик его функционирования (например параллельной производительности) еще на этапе проектирования на основе анализа результатов расчетов на системах-прототипах с небольшим количеством вычислителей и аналитических моделей параллельного ускорения.

В рамках схемы на рис. 4 можно приближенно записать время выполнения произвольной задачи из раздела 2 на p вычислителях внутри одного узла с общей памятью в форме

$$T_{\text{calc}}(p, \alpha, \gamma) = T_0 \left[\gamma \left(\alpha(p-1) + \frac{1}{p} \right) + (1-\gamma) \right], \quad (12)$$

где α — параметр, характеризующий рост накладных расходов с увеличением количества вычислителей, а γ — доля параллельных вычислений. Эти параметры могут быть оценены на основе прототипирования. Применительно к акватории Каспийского моря, например, процедуре подготовки метеорологической информации соответствует $\gamma = 0.96$, расчету волн — $\gamma = 0.95$, расчету течений и уровня моря — $\gamma = 0.59$, параметризации климатических спектров — $\gamma = 0.98$.

Общее время коммуникаций между узлами определим как сумму времени рассылки данных и их сбора, а также времени на частичную репликацию метеорологических данных:

$$T_{\text{comm}} = t_w \left[(3L_0\tau_0 + \Pi\mu L) + \Pi L\vartheta + (\eta\xi L + \nu(\xi L)^2)(\Pi - 1) \right]. \quad (13)$$

Здесь τ_0 , L_0 и τ , L — число точек расчетной сетки по времени и пространству для метеорологических (входных) и океанографических (выходных) данных соответственно, Π — количество узлов, η , ν , ξ , μ — коэффициенты, характеризующие кратность использования данных в моделях, а ϑ — доля репликации данных между узлами. Общее время вычислений T_0 каждой задачи (расчета ветра, волн, течений, параметризации спектров и расчета статистики) в (12) не зависит от топологии связей между узлами и определяется только спецификой алгоритма и производительностью узлов. Например, для одного ядра AMD Opteron 275 можно записать

$$\begin{aligned} T_{0(\text{wind})} &= \left[\psi_{\text{wind}}^{(1)} \frac{L_0\tau_0}{\Pi} + \psi_{\text{wind}}^{(2)} \frac{L\tau}{\Pi} \right], \\ T_{0(\text{waves})} &= \left[\psi_{\text{waves}} L \left(\frac{\tau}{\Pi} + \vartheta_{\text{waves}} \right) \right], \\ T_{0(\text{currents})} &= \left[\psi_{\text{currents}} L \left(\frac{\tau}{\Pi} + \vartheta_{\text{currents}} \right) \right], \\ T_{0(\text{spectra})} &= \left[\psi_{\text{spectra}} \xi L \left(\frac{\tau}{\Pi} \right) \right], \\ T_{0(\text{statistics})} &= \frac{\psi_{\text{statistics}} \xi (L\tau + 0.07 L^2 + 0.15 L\tau \ln(\tau))}{\Pi p}. \end{aligned} \quad (14)$$

В выражениях (14) коэффициенты ψ имеют смысл времени (в секундах) расчета элементарной операции (в пересчете на одну точку исходных или модельных данных). Приведем численные оценки этих коэффициентов для используемых моделей: $\psi_{\text{wind}}^{(1)} = 1.2 \times 10^{-5}$, $\psi_{\text{wind}}^{(2)} = 6. \times 10^{-4}$, $\psi_{\text{waves}} = 8.3 \times 10^{-4}$, $\psi_{\text{currents}} = 2.2 \times 10^{-4}$, $\psi_{\text{spectra}} = 1.6 \times 10^{-1}$, $\psi_{\text{statistics}} = 5. \times 10^{-5}$. Учитывая, что общее время вычислений является суммой времен из (13) и (14), этого достаточно, чтобы оценить параллельное ускорение как функцию от параметров.

На рис. 5а и 5б показана зависимость общего ускорения вычислений от особенностей гибридной архитектуры вычислительной системы.

При небольшом количестве узлов с общей памятью наиболее оптимальным является распараллеливание по времени с использованием отдельных процессоров (или ядер) узлов как независимых вычислителей. Однако с увеличением общего числа процессоров тенденция изменяется, и использование многопроцессорных узлов становится более предпочтительными. С другой стороны, чем больший объем данных (или число лет) обрабатывается, тем большее число процессоров необходимо для получения эффекта от использования многопроцессорных узлов. При этом очевидно, что чем большее число процессоров используется на узлах с общей памятью, тем дальше будет находиться точка заката кривой ускорения. С увеличением размера входного массива в результате увеличения гранулярности общее ускорение также возрастает.

На рис. 5в и 5г проиллюстрирован эффект влияния перекрытия независимых блоков данных (что является обязательным при распараллеливании по времени) на производительность. При уменьшении

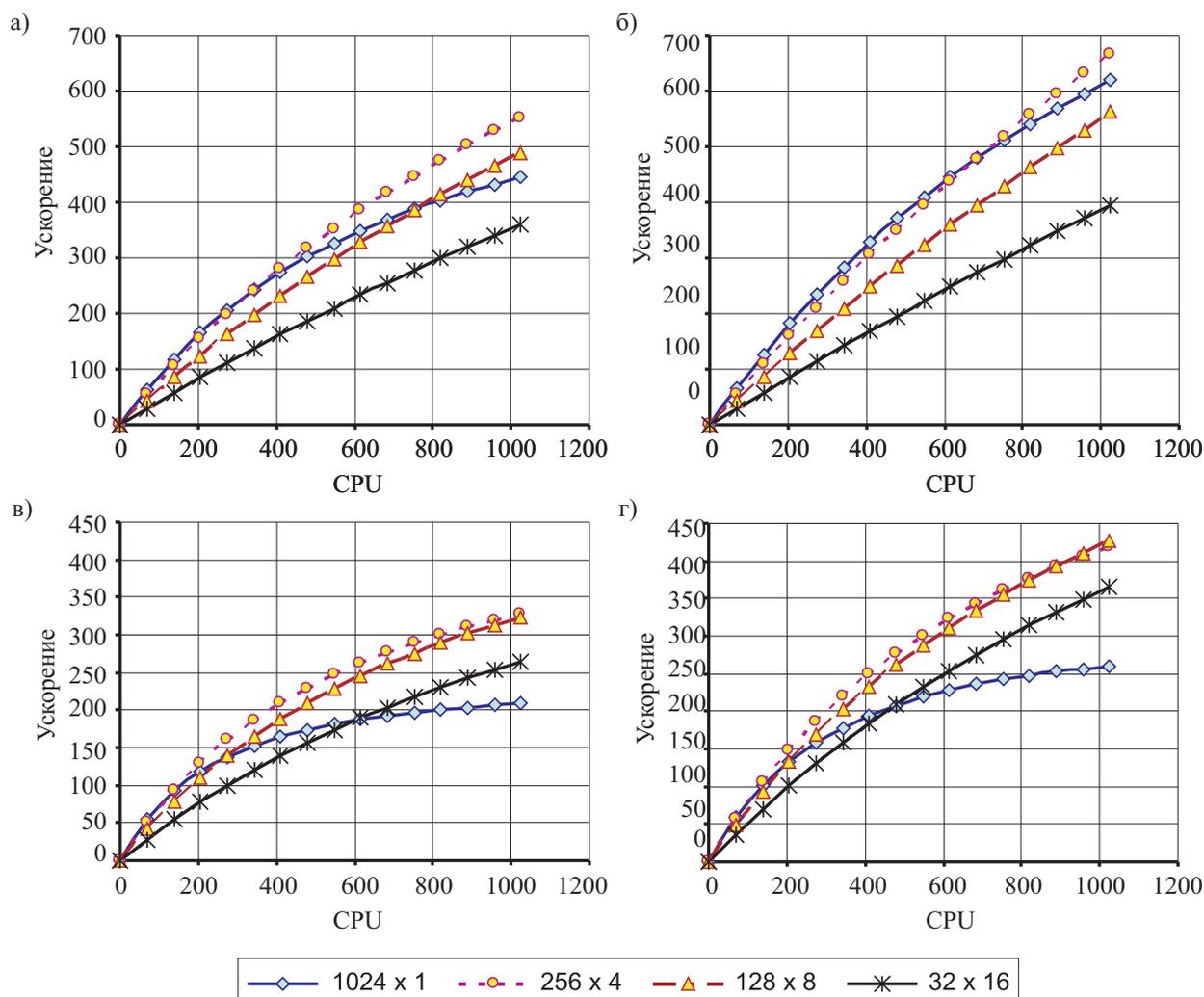


Рис. 5. Параллельное ускорение в зависимости от конфигурации вычислительных узлов, длины входного массива и размера перекрытия независимых блоков данных: а) 15 лет, перекрытие 1 месяц; б) 30 лет, перекрытие 1 месяц; в) 5 лет, перекрытие 3 месяца; г) 5 лет, перекрытие 1 месяц

величины перекрытия с трех месяцев до одного месяца ускорение в среднем вырастает примерно в полтора раза, а эффективность от использования многопроцессорных узлов сдвигается в зону большего числа узлов и меньшего числа процессоров на узле. Эта характеристика может меняться в зависимости от специфики акватории (океанские процессы обладают большим «временем жизни», чем процессы в закрытых морях), а также от особенностей моделируемого процесса (синоптический интервал атмосферных процессов и морского волнения составляет несколько суток, а эволюция вихрей водных масс может длиться несколько месяцев).

Результаты анализа, приведенного в данном разделе, могут рассматриваться как формализация знаний по управлению параллельными процессами с целью выбора оптимальной по производительности конфигурации для заданной задачи посредством интеллектуального сервиса управления вычислительными модулями (см. раздел 3.2).

5. Некоторые результаты расчетов. Прототип высокопроизводительного программного комплекса расчета экстремальных гидрометеорологических явлений был опробован на задаче оценивания экстремальных характеристик волнения в Баренцевом море. Исходными данными для расчета служили поля приводного ветра из массива реанализа NCEP/NCAR с 1967 по 2006 гг. Для усвоения использованы данные измерений ветра на пяти автоматических буйковых станциях в Баренцевом море (с 1989 по 1995 гг.). Учитывая, что волнение в Баренцевом море развивается в условиях, формируемых штормовой активностью в Северной Атлантике, для калибровки полей ветра дополнительно использовались данные наблюдений в Норвежском море, а также измерения судов погоды М, L в Атлантическом океане. Расчет полей

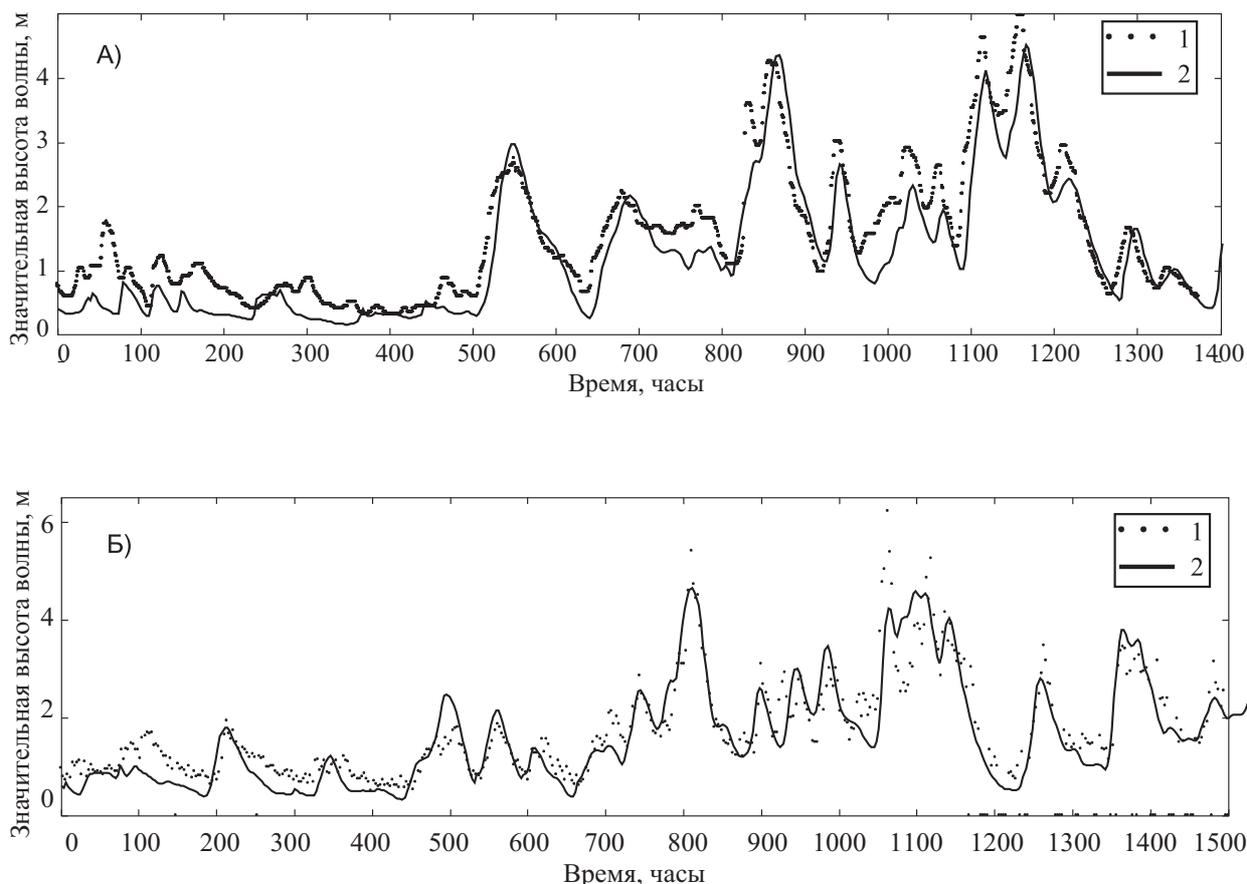


Рис. 6. Результаты верификации модельных расчетов высот волн по данным инструментальных измерений: 1 — данные измерений; 2 — расчеты по модели (1). Рисунки А) и Б) соответствуют точкам на рис. 7

волнения проводился на системе из трех вложенных пространственно-временных сеток посредством интегрирования (1) в форме спектральной модели IV поколения Wave Watch III [14]. Общее время расчета с учетом распараллеливания составило около 140 часов.

На рис. 6 приведены результаты верификации модельных расчетов высот волн по данным инструментальных измерений, выполненных в 2006 г. в двух точках восточной части Баренцева моря (рис. 7). В обоих случаях видно, что результаты моделирования достаточно хорошо согласуются с данными измерений, несмотря на то что за период расчетов усвоение данных по ветру не проводилось. Это дает основание для дальнейшего использования полученных расчетных данных с целью расчета экстремальных характеристик на основе системы моделей (9)–(11) и метода BOLIVAR.

На рис. 7 приведен пример расчетов волн наибольшей (0.1%) высоты в Баренцевом море, возможных один раз в 100 лет. Видно, что в западной части моря высота экстремальных волн может превышать 30 метров; по мере продвижения штормов к востоку (юго-востоку) эта величина несколько снижается, достигая в Печорском море только 8–16 метров.

6. Заключение. Таким образом, в данной работе представлен прототип комплекса совместного моделирования ветра, волнения, течений и уровня моря, позволяющего рассчитывать обширный набор оперативных и экстремальных статистик, характеризующих различные режимы эксплуатации морских объектов и сооружений. Отличительной особенностью комплекса является проектирование его структуры в рамках более общей концепции организации высокопроизводительных программных комплексов для моделирования сложных систем, к которым, несомненно, относится система “океан–атмосфера”. Среди основных аспектов проектирования стоит отметить следующее.

- 1) Использование сервисно-ориентированной архитектуры, отображающей взаимодействие управляющих и вычислительных компонентов, соответствующих моделям в различных диапазонах изменчивости.
- 2) Применение иерархической схемы отображения архитектуры программной системы на параллельную архитектуру вычислительной системы на аппаратном, системном, программном и логическом уров-

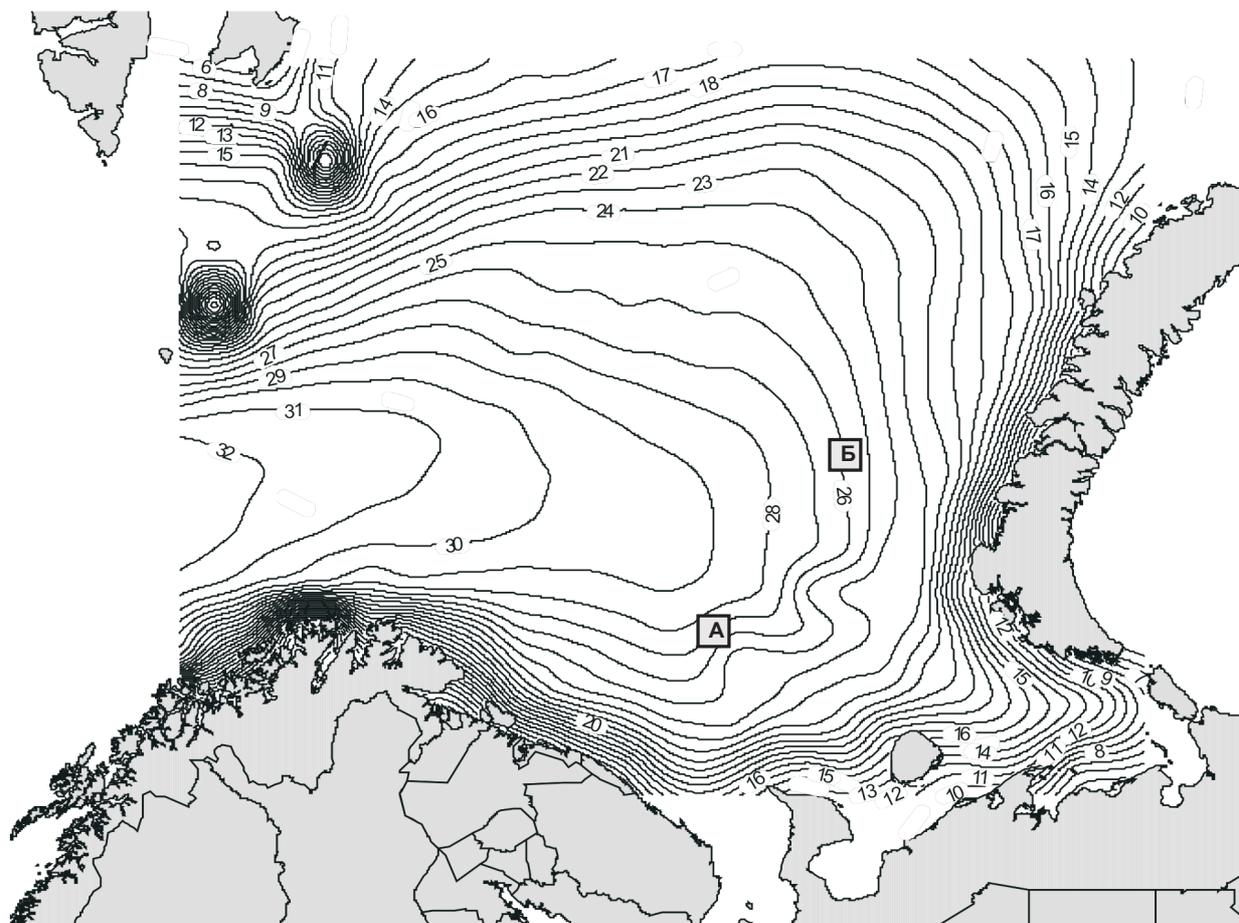


Рис. 7. Наибольшая (0,1 % обеспеченности) высота волны, возможная 1 раз в 100 лет, в Баренцевом море (значками нанесены точки измерений с рис. 6)

нях.

3) Использование аппарата параметрических моделей производительности, идентифицированных на основе прототипов, для оценки степени соответствия архитектуры программного комплекса проектным требованиям и оптимизации производительности программного комплекса за счет динамического выбора схемы отображения.

Несмотря на то что работы по созданию комплекса на данный момент не завершены, использование результатов прототипирования частей комплекса, а также априорных знаний о структуре моделей и методов позволяет обосновать архитектуру программного комплекса с учетом специфики суперкомпьютерных систем Гидрометцентра России на основе SGI Altix 4700 и SGI ICE 8200.

СПИСОК ЛИТЕРАТУРЫ

1. Справочные данные по режиму ветра и волнения Баренцева, Охотского и Каспийского морей / Под ред. Л. И. Лопатухина и др. Российский морской регистр судоходства. СПб: РМРС, 2003.
2. Бухановский А.В., Лопатухин Л.И., Чернышева Е.С. Подходы, опыт и некоторые результаты исследований волнового климата океанов и морей. II. Расчет волнения по гидродинамическим моделям, режимные распределения и климатические спектры волн // Вестник СПбГУ. 2005. Сер. 7, вып. 4. 56–69.
3. Справочные данные по режиму ветра и волнения Балтийского, Северного, Черного, Азовского и Средиземного морей // Под ред. Л. И. Лопатухина и др. Российский морской регистр судоходства. СПб: РМРС, 2006.
4. Володин Е.М., Толстых М.А. Параллельные вычисления в задачах моделирования климата и прогноза погоды // Вычислительные методы и программирование. 2007. 8, № 1. 117–126.
5. Joppich W., Quaas J. Coupling general circulation models on a meta-computer // Lecture Notes in Computer Science. Vol. 2657. Berlin: Springer, 2003. 161–170.
6. Stankova E., Zatevakhin M. Numerical simulation of cloud dynamics and microphysics // Lecture Notes in Computer Science. Vol. 2657. Berlin: Springer, 2003. 171–178.

7. *Мирзоев Д.А., Зильберштейн О.И., Лопатухин Л.И., Миронов Е.У., Михайлов Н.Н.* Концепция обеспечения специализированной гидрометеорологической информацией проектирования сооружений на шельфе арктических морей // Труды IV Международной конференции “Освоение шельфа арктических морей” (РАО 99). СПб, 1999. 311–318.
8. *Бухановский А.В., Лопатухин Л.И., Рожков В.А.* Подходы, опыт и некоторые результаты исследований волнового климата океанов и морей. III. Экстремальные и необычные волны // Вестник СПбГУ. 2006. Сер. 7, вып. 1. 58–69.
9. *Лопатухин Л.И., Бухановский А.В., Иванов С.В., Ковальчук С.В., Попов С.К., Лобов А.Л., Тихонова О.В., Чумаков М.М.* Современная технология комплексного моделирования ветра, волнения, течений и уровня моря // Труды VIII Международной конференции по освоению ресурсов нефти и газа Российской Арктики и континентального шельфа стран СНГ (РАО/CIS OFFSHORE 2007). СПб, 2007. 233.
10. *Бухановский А.В., Лопатухин Л.И., Иванов С.В.* Подходы, опыт и некоторые результаты исследований волнового климата океанов и морей. I. Постановка задачи и входные данные // Вестник СПбГУ. 2005. Сер. 7, вып. 3. 62–74.
11. *Boukhanovsky A.V.* Multivariate stochastic models of metocean fields: computational aspects and applications // Lecture Notes in Computer Science. Vol. 2329. Berlin: Springer, 2002. 216–225.
12. *Lopatoukhin L.I., Rozhkov V.A., Ryabinin V.E., Swail V.R., Boukhanovsky A.V., Degtyarev A.B.* Estimation of extreme wind wave heights. JCOMM Technical Report WMO/TD N 1041. 2000.
13. *Bogdanov A.V., Boukhanovsky A.V.* High performance parallel algorithms for data processing // Lecture Notes in Computer Science. Vol. 3036. Berlin: Springer, 2004. 239–246.
14. *Tolman H.L.* User manual and system documentation of WAVEWATCH-III version 2.22. NOAA / NWS / NCEP / MMAB Technical Note N. 222. 2002 (<http://polar.ncep.noaa.gov/waves>).

Поступила в редакцию
01.04.2008
