

УДК 528.854.2

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ В ЗАДАЧЕ ПРОСТРАНСТВЕННОГО МОДЕЛИРОВАНИЯ ИЗМЕНЕНИЙ ЛАНДШАФТНОГО ПОКРОВА

А.В. Замятин, А.А. Афанасьев

Томский политехнический университет
E-mail: zamyatin@tpu.ru, afanasyevaa@tpu.ru

Обсуждается задача моделирования изменений ландшафтного покрова, основанная на использовании марковских цепей и аппарата клеточных автоматов, увеличенная производительность решения которой достигается применением различных вариантов параллельной обработки данных при моделировании. Приведены результаты численных экспериментов по оценке различных вариантов параллельной обработки, полученные с использованием дорогостоящего суперкомпьютерного кластера и кластера из недорогих ПЭВМ в локальной сети, включая оценку параллельного ускорения, параллельной эффективности и общей производительности кластеров.

Ключевые слова:

Моделирование изменений ландшафтного покрова, параллельные вычисления, марковская цепь, клеточный автомат.

Key words:

Land cover changes modeling, parallel computing, Markov chain, cellular automata.

Введение

Один из наиболее перспективных и широко применяемых на практике подходов к моделированию, реализующих принципы построения систем с возможностями прогнозирования изменений ландшафтного покрова, основан на стохастических моделях, клеточных автоматах (КА) [1, 2] и аппарате марковских цепей и позволяет реализовать поведение сложных природно-территориальных комплексов и строить прогнозную тематическую карту T^t на некоторый момент времени t в будущем на базе разновременных тематических карт T^2 и T^1 на моменты времени t_2 и t_1 , матрицы вероятностей переходов $P=[p_{ij}]$ (стохастической матрицы) из типа (класса) ω_i в тип ω_j , матрицы $M_\Phi=[m_{ij}^\Phi]$ фактического числа переходов элементов на отрезке времени $\Delta t_2 \in [t_2; t_1]$ и матрицы $M_{\text{ож}}=[m_{ij}^{\text{ож}}]$ ожидаемого числа переходов элементов на отрезке $\Delta t_1 \in [t_1; t]$, а также с помощью сравнительно простых правил функционирования КА за десятки и сотни итераций алгоритма моделирования, требующих существенных вычислительных затрат [3–5]. Тематическую карту, на основе которой осуществляются последовательно необходимые преобразования (например, карта T^1 на момент времени t_1), будем называть базовой тематической картой. Период времени, соответствующий Δt_2 , называют *наблюдаемым*, а Δt_1 – *прогнозируемым*.

Правила функционирования КА могут быть сформированы различными способами [6]. В данном случае будем их задавать на вероятностной основе с использованием результирующей вероятности $p_{ij}^{\text{рез}}$, определяемой исходя из набора вероятностных компонентов $p_{ij}^{\text{вер}}$, $p_{ij}^{\text{пр}}$ и $p_{ij}^{\text{доп}}$. При допущении независимости событий, задаваемых $p_{ij}^{\text{вер}}$, $p_{ij}^{\text{пр}}$ и $p_{ij}^{\text{доп}}$, вероятность $p_{ij}^{\text{рез}}$ может быть найдена как

$$p_{ij}^{\text{рез}} = p_{ij}^{\text{вер}} \times p_{ij}^{\text{пр}} \times p_{ij}^{\text{доп}}. \quad (1)$$

Вероятностный компонент $p_{ij}^{\text{вер}}$ может быть определен как $p_{ij}^{\text{вер}} = p_{ij}$ или как $p_{ij}^{\text{вер}} = p_{ij} \cdot n_j$ на основе стохастической матрицы $P=[p_{ij}]$, где n_j – число эле-

ментов типа ω_j в окрестности заданного КА, а p_{ij} определяется как

$$p_{ij} = \frac{n_{ij}}{\sum_{j=1}^M n_{ij}}, \quad (2)$$

где n_{ij} – число элементов, перешедших из состояния ω_i в состояние ω_j , а M – число типов ландшафта.

Вероятностный компонент $p_{ij}^{\text{доп}}$ при наличии соответствующей априорной информации о буферных зонах, расстоянии до сетей инженерных коммуникаций и других аналогичных данных может быть сформирован с помощью вероятностных карт соответствия $Ptm_k = \{p_{ij}^k, k=1, 2, \dots, Sm, Sm \text{ – общее количество карт соответствия, } i=1, 2, \dots, Cn, j=1, 2, \dots, Rn, \text{ а } Ptm_k \text{ – матрица размером } [Cn \times Rn]\}$, построенных с использованием функций пространственного анализа. Если принять значимость (вес) каждого фактора влияния, с учетом которого построена каждая карта соответствия, идентичным для всех k , то вероятность $p_{ij}^{\text{доп}}$ будет определяться как

$$p_{ij}^{\text{доп}} = p_{ij}^1 \times p_{ij}^2 \times \dots \times p_{ij}^{Sm}.$$

Определение вероятностного компонента $p_{ij}^{\text{пр}}$ предлагается реализовать с использованием набора пространственных метрик. Один из возможных вариантов практического применения $p_{ij}^{\text{пр}}$ в моделировании изменений ландшафтного покрова изложен в [7].

Практические особенности формирования матриц, определяющих основные тенденции изменения ландшафтного покрова, заключаются в следующем. Матрицу фактических переходов определим как $M_\Phi=[m_{ij}^\Phi]$, где m_{ij}^Φ – число элементов, перешедших из поверхности типа ω_i в тип ω_j на изображениях T^2 и T^1 соответственно, $i, j=1, \dots, M$, M – число типов поверхности на исследуемой территории. С помощью выражения (2) и значений матрицы M_Φ ($n_{ij}=M_\Phi[i, j]$) получим стохастическую матрицу $P=[p_{ij}]$. Сумма элементов каждой строки матрицы M_Φ содержит информацию о количестве пиксе-

лей, принадлежащих соответствующим типам ландшафтного покрова на момент времени t_2 , а сумма элементов каждого столбца – ту же информацию на момент времени t_1 , причем $t_2 < t_1$.

Для того чтобы из \mathbf{M}_Φ получить матрицу ожидаемых переходов $\mathbf{M}_{\text{ож}}$, необходимо выполнить следующие действия:

1. Транспонировать \mathbf{M}_Φ и получить \mathbf{M}_Φ^T .
2. Получить вектор $\mathbf{S}=[s_1, s_2, \dots, s_M]^T$, причем

$$s_i = \sum_{j=1}^M (m_{ij})^T = \sum_{j=1}^M m_{ji}.$$

3. Получить $\mathbf{M}_{\text{ож}}=[m_{ij}^{\text{ож}}]=[s_i \times p_{ij}]$, $i, j=1, \dots, M$.

Как отмечено выше, изложенный подход к построению прогнозных карт с использованием моделирования изменений ландшафтного покрова является многоитерационной вычислительно затратной процедурой. Тем более вычислительные затраты возрастут при увеличении объемов используемых исходных данных (например при возрастании геометрических размеров разновременных растровых карт \mathbf{I}^2 и \mathbf{I}^1 или при увеличении числа M ландшафтных классов на исследуемой территории). Существенное снижение вычислительных затрат алгоритма моделирования может быть достигнуто путем применения *высокопроизводительных параллельных вычислений* [8–11].

Рассмотрим некоторые существующие возможности модификации алгоритма моделирования изменений ландшафтного покрова, позволяющие существенно увеличить производительность вычислений за счет параллельного варианта исполнения.

Организация параллельных вычислений при моделировании

При разработке параллельного варианта исполнения алгоритма моделирования целесообразно ориентироваться на следующие требования:

- сравнительная простота реализации параллельного варианта, не требующая значительных модификаций традиционного последовательного алгоритма;
- минимизация затрат на передачу данных между вычислительными узлами при распределенной обработке, увеличивающая общую производительность кластера [12];
- минимизация величины отклонений результатов параллельного алгоритма моделирования (в идеальном случае – их отсутствие) от результатов, которые были бы получены исходным алгоритмом при традиционном последовательном исполнении.

Реализация распределенного алгоритма предполагает наличие управляющего (головного) вычислительного узла (ВУ) (p_0) и N ВУ для параллельной обработки данных ($p_i, i=1, 2, \dots, N$). Поэтому логику возможных вариантов параллельного алгоритма моделирования будут определять варианты распределения данных между головным и прочими ВУ.

В самом простом варианте алгоритм моделирования оперирует следующими данными:

- *исходными изображениями* (тематическими картами): \mathbf{T}^{t-2} и \mathbf{T}^{t-1} , полученными на моменты времени $(t-2)$ и $(t-1)$ соответственно;
- *матрицами тенденций изменений*, рассчитываемыми на основе \mathbf{T}^{t-2} и \mathbf{T}^{t-1} : вероятностей переходов – \mathbf{P} , фактических переходов – \mathbf{M}_Φ и ожидаемых переходов – $\mathbf{M}_{\text{ож}}$, а также матрицей ранжирования – \mathbf{R} ;
- *прочими параметрами моделирования*, задаваемыми пользователем: порядком d анализируемой окрестности КА, критерием останова K_{stop} , а также величинами наблюдаемого Δt_2 и прогнозного Δt_1 интервалов времени;
- *промежуточными результатами моделирования (изображениями)*: $\mathbf{T}_{(k)}$, полученными в результате преобразований на $(k-1)$ -й итерации с использованием исходных изображений \mathbf{T}^{t-2} и \mathbf{T}^{t-1} , матриц \mathbf{P} , \mathbf{M}_Φ , $\mathbf{M}_{\text{ож}}$ и \mathbf{R} , а также с учетом прочих параметров моделирования; на основе $\mathbf{T}_{(k)}$ строится результирующее изображение \mathbf{T}^t на момент времени t .

Возможны два принципиально различных варианта организации параллельных вычислений при моделировании изменений ландшафтного покрова.

Первый вариант предполагает расчет вероятностных компонентов типа (1) на каждом из ВУ без фрагментации (разделения) и уменьшения геометрического размера исходных данных. Результат расчета на каждом ВУ p_i интерпретируется как значение вероятности $p_{ij}^{\text{рез}}$ и применяется в многоитерационном моделировании при определении правил функционирования КА. Результаты обработки каждым из ВУ p_i объединяются на головном ВУ p_0 тем или иным образом, в зависимости от того, в каком виде представлены результаты. Если результаты моделирования отдельными ВУ представлены в виде вероятностных карт для каждого из M типов ландшафта, результирующая карта может быть найдена путем произведения соответствующих вероятностей в каждой точке изображений. В этом случае тип ландшафта, для которого вероятность максимальна, является искомым результатом. Если результаты представлены в виде конечных тематических прогнозных карт, типы ландшафта в каждой точке результирующей прогноз-ной карты могут быть найдены, например, по мажоритарному правилу.

При такой организации параллельных вычислений моделирование изменений осуществляется независимо различными ВУ и достигается простота практической реализации. Однако при этом в процессе моделирования и обработки данных различными ВУ не учитывается взаимное влияние используемых вероятностных компонентов, что может негативно влиять на точность моделирования и адекватность представления информации о процессах и явлениях ландшафтного покрова на полученных прогнозных картах.

Другой вариант параллельной реализации алгоритма моделирования предполагает расчет вероятностей и моделирование таким образом, чтобы сохранить взаимное влияние вероятностных компонентов. В этом случае следует использовать весь доступный набор данных, которыми оперирует алгоритм моделирования, осуществляя на каждом ВУ моделирование для некоторого фрагмента исследуемой территории.

Поиск вариантов параллельной реализации алгоритма моделирования, удовлетворяющих сформулированному выше условию, показывает практическую сложность фрагментации указанных данных и разделения соответствующих параметров алгоритма по ВУ кластера таким образом, чтобы сохранить логику его традиционного последовательного исполнения. Это связано с тем, что использование геометрически уменьшенных фрагментов данных на каждом ВУ не всегда при моделировании (аналогичном тому, что исполняется традиционным последовательным вариантом алгоритма моделирования) позволяет отразить во фрагменте тенденции ландшафта, характерные для всей исследуемой области.

В связи с этим поиск способов адаптации рассматриваемого алгоритма моделирования для возможности параллельной обработки данных позволяет предложить варианты, характеризующиеся результатами моделирования, отличными в той или иной степени от результатов, которые были бы получены исходным алгоритмом при традиционном последовательном исполнении.

Рассмотрим два основных варианта возможной реализации параллельного алгоритма моделирования изменений ландшафтного покрова — с *собственными* и с *разделяемыми матрицами тенденций изменения* на ВУ кластера.

Вариант с собственными матрицами тенденций изменений

Вариант с *собственными матрицами тенденций изменения* предполагает следующий обобщенный порядок работы:

Шаг 1. На головном узле осуществляется разделение исходных изображений T^{t-2} и T^{t-1} на фрагменты T_i^{t-2} , T_i^{t-1} , $i=1, 2, \dots, N$, где N — число ВУ.

Шаг 2. На каждом p_i осуществляется расчет собственных матриц P_i , $M_{\Phi(i)}$, $M_{\text{ок}(i)}$ и $R_{(i)}$, которые используются при моделировании и построении фрагментов T_i^t результирующего изображения T^t на момент времени t .

Шаг 3. На головном узле осуществляется компоновка фрагментов T_i^t в единое результирующее изображение T^t .

Очевидно, что основным ограничением использования этого варианта параллельного алгоритма моделирования является потребность в обработке исходных изображений T^{t-2} и T^{t-1} , которая позволит при фрагментации изображений сохранить на отдельных p_i ландшафтно-классовую структуру, характерную для исходных изображе-

ний. При моделировании на отдельных ВУ это способствует сохранению тенденций изменения ландшафтного покрова, характерных для не фрагментированных исходных разновременных изображений. Это, в свою очередь, позволит снизить потенциальную величину отклонений результатов параллельного алгоритма от результатов, которые были бы получены исходным алгоритмом при традиционном последовательном исполнении. В противном случае, матрицы тенденций изменения, используемые каждым p_i , будут существенно отличаться от матриц, используемых традиционным последовательным алгоритмом, а, следовательно, обеспечивать соответствующие значительные отклонения в результатах моделирования.

Кроме того, независимая обработка данных различными ВУ может негативно влиять на общую производительность вычислений кластера. Действительно, объем вычислений каждым ВУ может существенно отличаться, а компоновка фрагментов T_i^t в единое результирующее изображение T^t будет выполнена только после завершения обработки данных каждым ВУ.

Вариант с разделяемыми матрицами тенденций изменения

В общем случае этот вариант алгоритма может иметь разновидности в зависимости от вида разделяемых матриц (например, разделяемыми всеми ВУ могут быть матрицы P и/или $M_{\text{ок}}$). Вариант с разделяемой матрицей P предполагает следующий обобщенный порядок работы:

Шаг 1. На головном узле p_0 на основе исходных изображений T^{t-2} и T^{t-1} осуществляется расчет матриц P , M_{Φ} , $M_{\text{ок}}$ и R ; выполняется разделение базового изображения T^{t-1} на фрагменты T_i^{t-1} , формируются матрицы ранжирования $R_i \in R$. Данные (P, T_i^{t-1}, R_i) передаются на соответствующие p_i .

Шаг 2. На каждом ВУ p_i на основе полученных данных осуществляется расчет собственных матриц $M_{\text{ок}(i)}$, выполняется моделирование и построение фрагментов T_i^t результирующего изображения T^t на момент времени t .

Шаг 3. На головном узле осуществляется компоновка фрагментов T_i^t в единое результирующее изображение T^t .

Изложенный вариант алгоритма позволяет ожидать менее значительную величину отклонений результатов параллельного алгоритма от результатов (по сравнению с вариантом алгоритма с собственными матрицами тенденций изменения), которые были бы получены исходным алгоритмом при традиционном последовательном исполнении. Главным образом это связано с тем, что в этом случае используемые каждым p_i матрицы тенденций изменения менее существенно отличаются от матриц, используемых традиционным последовательным алгоритмом.

В случае, если в качестве разделяемой матрицы выбрана $M_{\text{ок}}$, ее элементы могут быть найдены как $M_{\text{ок}} = \{m_{\text{ок}}[l, r] = m_{\text{ок}(i)}[l, r]/N, l, r = 1, 2, \dots, M, N$ — число

доступных ВУ, M – число типов ландшафта}, обеспечивая сниженный объем ожидаемых преобразований при моделировании на каждом ВУ.

Для обеспечения еще более значительного соответствия параллельного варианта исполнения его последовательному традиционному аналогу рассмотрим еще одну разновидность алгоритма с разделяемыми матрицами тенденций изменения – с *синхронизацией матриц тенденций изменения*. Такая синхронизация может выполняться через некоторый определенный временной интервал (Δt_{syn}), задаваемый либо временными параметрами моделирования, либо некоторой относительной долей выполненных преобразований (ΔS). Учитывая относительный характер величины ΔS , ее практическое использование при определении интервала синхронизации следует признать более предпочтительным.

Параллельный алгоритм моделирования с разделяемыми матрицами тенденций изменения и их синхронизацией предполагает следующий обобщенный порядок работы:

Шаг 1. На головном узле p_0 на основе исходных изображений T^{r-2} и T^{r-1} осуществляется расчет матриц P , M_{ϕ} , $M_{\text{ож}}$ и R ; выполняется разделение базового изображения T^{r-1} на фрагменты T_i^{r-1} , формируются матрицы ранжирования $R_i \in R$. Данные $(P, M_{\text{ож}}, R, T^r)$ передаются на соответствующие p_i .

Шаг 2. На каждом p_i на основе полученных данных осуществляется моделирование изменений ландшафтного покрова. Через заданные доли преобразования ΔS производим процедуру синхронизации:

- передаем матрицу $M_{\text{ож}(i)}$ с каждого p_i на p_j , $j=1, 2, \dots, N$ при $i \neq j$;
- приводим значения матрицы $M_{\text{ож}(i)}$ в соответствие с выполненными на всех p_i преобразованиях – $M_{\text{ож}(i)} = \{m_{\text{ож}(i)}[l, r] + m_{\text{ож}(j)}[l, r], l, r=1, 2, \dots, M, j=1, 2, \dots, N \text{ при } i \neq j\}$;
- определяем оставшийся объем преобразований – $M_{\text{ож}(i)}^0 = \{M_{\text{ож}}^0[l, r] - m_{\text{ож}(i)}[l, r], l, r=1, 2, \dots, M$, где $M_{\text{ож}}^0$ – матрица ожидаемых переходов $M_{\text{ож}}$, рассчитанная на шаге 1}.

Шаг 3. На головном узле осуществляется компоновка фрагментов T_i^r в единое результирующее изображение T^r .

В приведенном алгоритме матрица $M_{\text{ож}}$ без модификаций рассылается на все ВУ, а затем синхронизируется через заданные доли преобразования ΔS . При задании ΔS необходимо учитывать то, что более частая синхронизация (малое ΔS) потребует более значительных временных и вычислительных ресурсов, а менее частая не позволит поддерживать значения матрицы $M_{\text{ож}}$ адекватно отражающими объем выполненных преобразований всеми ВУ. Очевидно, этот вариант параллельного алгоритма позволит более равномерно обеспечивать загрузку различных ВУ кластера за счет обеспечения актуального состояния значений матрицы $M_{\text{ож}}$. Это позволит исключить ситуации, когда незначительное число из доступных ВУ кластера выполняет значи-

тельную долю вычислительной работы, а «незагруженные» ВУ простаивают в ожидании окончания этой обработки, что негативно влияет на общую производительность кластера. Пересылка целочисленных значений матрицы $M_{\text{ож}}$ размерности M по каналам связи вычислительного кластера не будет существенно влиять на задержки в обработке данных из-за незначительных объемов этих данных. Все эти особенности будут позитивно отражаться на общей производительности вычислительного процесса в независимости от особенностей исходных данных. Это позволяет выбрать данный вариант в качестве основного при проведении экспериментальных исследований.

Экспериментальные исследования

Для оценки предположений об уровне адекватности и точности распределенных версий алгоритма моделирования в сравнении с его традиционным последовательным аналогом распределенные версии алгоритма моделирования с разделяемыми и с собственными матрицами тенденций изменений реализованы на языке C++ с использованием компилятора g++ версии 3.4.4 и распространённого интерфейса параллельного программирования MPI (MPICH 2.0) [13–15].

Постановка задачи исследования. Определим ключевые параметры эффективности параллельных версий выбранного алгоритма моделирования:

- производительность кластеров различной конфигурации (недорогие пользовательские ПЭВМ, объединенные в локальной вычислительной сети невысокой пропускной способности; суперкомпьютер петафлопной производительности; различное число N доступных ВУ);
- ускорение S_N ($S_N = t_{\text{пол}}^{\text{пол}} / t_N^{\text{пар}}$, где $t_{\text{пол}}^{\text{пол}}$ – время работы последовательного алгоритма, $t_N^{\text{пар}}$ – время работы параллельного алгоритма на N ВУ при тех же входных данных) и эффективность E_N ($E_N = S_N / N$) параллельного моделирования на кластерах различной конфигурации [15];
- величины отклонений результатов параллельного алгоритма моделирования от результатов, которые получены исходным алгоритмом при традиционном последовательном исполнении. Используем для такой оценки каппа-индекс согласия (КИС) [16], рассчитываемый по матрице ошибок и широко применяемый для оценки точности классификации при сравнении эталонного и результирующего изображений.

Основные характеристики конфигураций вычислительных кластеров, используемых при проведении численных экспериментов, приведены в табл. 1.

При проведении экспериментов использованы два различных набора тестовых разновременных изображений, позволяющих осуществлять многоитерационное моделирование и построение по различным исходным изображениям серии прогнозных разновременных изображений. Так, в качестве одного набора использованы данные, поставляе-

мые в качестве примерных данных с растровой геоинформационной системой *Idrisi Kilimanjaro*, – файлы *Landuse71.rst* (изображение T^2), *Landuse85.rst* (изображение T^1) и *Landuse91.rst* (изображение T^0).

Таблица 1. Характеристики вычислительных кластеров

Номер конфигурации	Число ВУ	Сеть	Характеристики ВУ		
			Процессор	ОП	ОС
1	1	–	Pentium IV 3 ГГц	4 Гбайт	MS Windows7
2	2–11	FastEthernet, 100 Мбит/с	Pentium IV 3 ГГц	4 Гбайт	MS Windows7
3	1–24	Gigabit Ethernet, 1000 Мбит/с	2×Intel XEON 5150 (2×2 ядра)	8 Гбайт	Novell SLES 10

Эти данные (тематические растровые карты) имеют 9 ландшафтных классов и геометрические размеры – 565 столбцов и 452 строки, пространственное разрешение составляет 60 м. Другой набор разновременных данных представляет детальные (пространственное разрешение 25 м) тематические карты территории Португалии (изображение T^2) 2000 (изображение T^1) и 2006 гг. (изображение T^0), содержащие более 40 ландшафтных классов (Данные предоставлены Португальским географическим институтом в рамках реализации проекта анализа динамики территории Португалии при поддержке Португальского фонда научных исследований (PTDC/CS-GEO/101836/2008).) на 1990 г.). Для практического использования и проведения экспериментов по построению прогнозных карт при различных условиях данные подвергнуты предварительной обработке – с использованием экспертной генерализации количество ландшафтных типов сокращено до 16, а также путем уменьшения пространственного разрешения получен целый ряд разновременных наборов с различными геометрическими параметрами, но идентичной ландшафтно-классовой структурой. Основные характеристики используемого в экспериментальных исследованиях набора данных приведены в табл. 2.

Таблица 2. Характеристики тестовых разновременных данных

Номер набора (изображения T^2 , T^1 , T^0)	Размеры, пикс.	Число классов	Объем, Кбайт
1	271×581	16	154
2	531×1162	16	603
3	791×1774	16	1348
4	1168×2322	16	2422
5	1341×2906	16	3806
6	1646×3488	16	5607
7	1910×4069	16	7590
8	565×452	9	259

Для оценки общей производительности обработки данных, условного параллельного ускорения S_N и эффективности E_N на вычислительных кластерах различной конфигурации (табл. 1) использовались тестовые наборы данных 1–8 (табл. 2), иден-

тичные для последовательной версии алгоритма (табл. 1, конфиг. 1) и для параллельного алгоритма моделирования с синхронизацией матриц тенденций изменения, выбранного выше в качестве основного (табл. 1, конфиг. 2 и 3).

Процесс моделирования осуществлялся традиционно – на основе изображений T^2 и T^1 или T^2 и T^0 выполнялся расчет тенденций изменения ландшафтного покрова и производилось построение серии из 15 прогнозных изображений $T^{(i)} = \{T^i, T^{i+1}, \dots, T^{i+15}\}$.

Для получения статистически достоверных результатов каждый эксперимент повторялся тридцатикратно, а в качестве оценки использовалось среднее значение, полученное по результатам каждого из экспериментов, а доверительная вероятность при этом была принята равной 95 %. По результатам экспериментов ошибка определения среднего значения оцениваемой величины (математического ожидания) не превышала 3...4 %.

Результаты экспериментов

На рис. 1 приведены некоторые результаты оценки общей производительности (в данном случае определяемой через время t , затраченное на выполнение обработки данных и формирование результата), параллельного ускорения и параллельной эффективности для варианта параллельного алгоритма моделирования с синхронизацией матриц тенденций изменения, с использованием набора данных № 5 (табл. 2). Приведенные результаты демонстрируют увеличение производительности обработки данных с ростом числа доступных ВУ в 1,5–4,0 раза.

Следует отметить, что для подмножества тестовых данных с широким диапазоном характеристик достаточным является использование $N \approx 4...6$ ВУ. Это обусловлено тем, что дальнейший рост числа используемых ВУ не обеспечивает значительного увеличения производительности и параллельного ускорения из-за увеличивающегося влияния накладных расходов параллельной обработки данных. Поэтому, определив достаточное N , можно снизить влияние накладных расходов параллельной обработки и увеличить степень рационального использования оборудования кластера.

Важной характеристикой данных, также оказывающей влияние на производительность вычислений последовательного и параллельного вариантов исполнения алгоритма моделирования, является величина геометрического размера исходных разновременных изображений. Для оценки такого влияния проведены соответствующие эксперименты, фрагмент которых представлен на рис. 2 (наборы данных 1–8, табл. 2).

Приведенные на рис. 2 графики демонстрируют нелинейное (близкое к экспоненциальному) возрастающее увеличение затрат времени на производимые вычисления с увеличением геометрического размера исходных изображений. При этом наблюдается существенное возрастание преимущества параллельного варианта алгоритма (рис. 2, б).

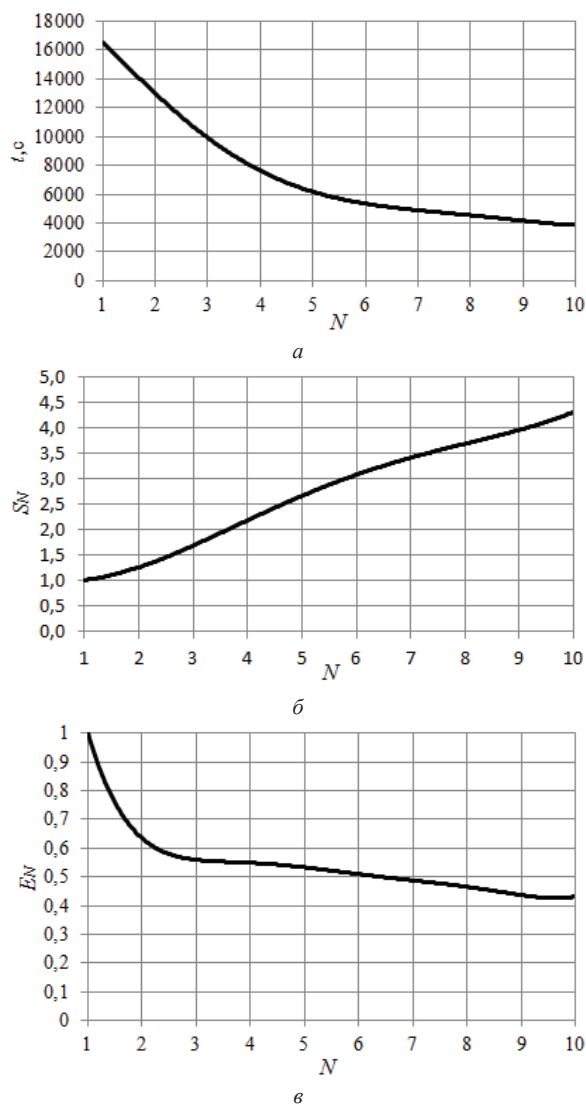


Рис. 1. Результаты оценки параллельных вариантов алгоритма моделирования: а) производительность; б) параллельное ускорение; в) параллельная эффективность

На рис. 3 приведен пример оценки адекватности параллельного варианта алгоритма моделирования его последовательному аналогу (набор данных № 5, табл. 2). Несмотря на то, что для данных с иными геометрическими и ландшафтно-классовыми характеристиками (рис. 3, а, использован набор данных № 3, табл. 2, при условии равенства наблюдаемого интервала прогнозируемому интервалу $|t_1 - t_2| = |t - t_1|$) и различных конфигураций вычислительного кластера (отличных от приведенных в табл. 1) оценки эффективности параллельного исполнения могут несколько отличаться, но в целом эти оценки являются характерными и позволяют сделать некоторые обобщенные выводы и сформулировать практические рекомендации.

В частности, результаты экспериментов (рис. 3, а) показывают, что с увеличением числа задействованных в моделировании ВУ адекватность параллельного варианта алгоритма последовательному

предсказуемо снижается из-за увеличивающегося несоответствия ландшафтно-классовой структуры обрабатываемых фрагментов и исходного изображения, не позволяющего в полной мере отразить тенденции изменения ландшафтного покрова, присутствующие на исследуемой территории в целом. Причем чем большее число ВУ задействовано в вычислениях, тем большее число фрагментов (уменьшенного геометрического размера) с более высокой степенью несоответствия исследуемой территории будет использовано при моделировании. При этом следует отметить, что алгоритм моделирования с синхронизацией матриц тенденций изменения позволяет существенно минимизировать такое отклонение. В приведенном на рис. 3, а примере максимальное отклонение при $N=8$ составило около 8 %. При использовании числа $N \approx 4-6$ ВУ, рекомендованного по результатам экспериментов как наиболее целесообразное с практической точки зрения, такое отклонение будет менее существенным.

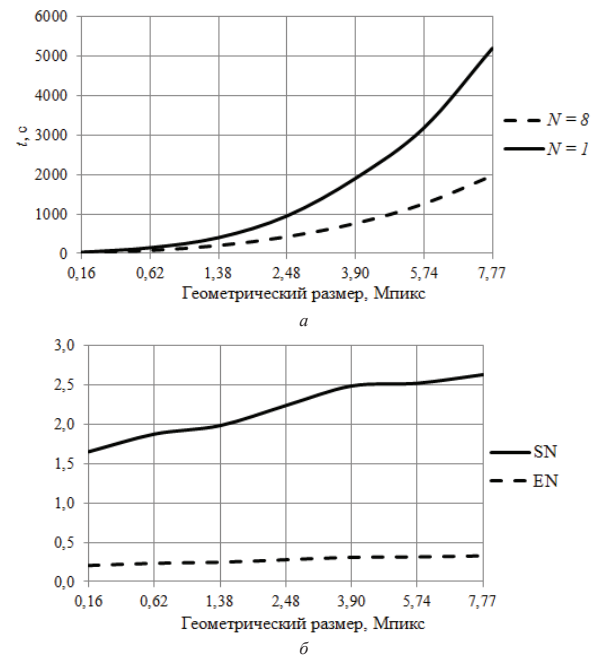


Рис. 2. Показатели эффективности параллельной обработки для данных различного геометрического размера: а) производительность; б) параллельное ускорение и эффективность

С увеличением горизонта прогнозирования адекватность получаемых результатов снижается. В примере на рис. 3, б при различных горизонтах прогнозирования ($Iter_{max} = 1, 2, \dots, 10$) степень несоответствия результатов, получаемых параллельной версией алгоритма и его последовательным аналогом, также не превысила 8 %.

Таким образом, потенциальную степень несоответствия результатов, получаемых параллельным алгоритмом и его последовательным аналогом для вычислительно сложной задачи пространственного ландшафтного прогнозирования, следует считать практически приемлемой.

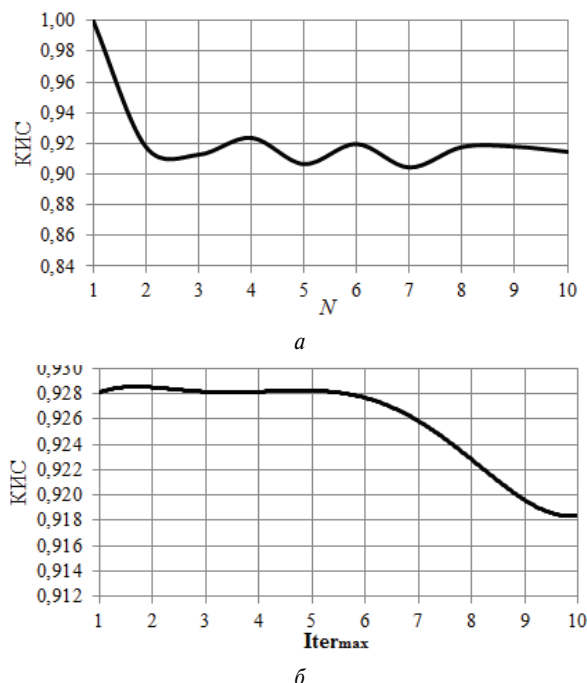


Рис. 3. Пример оценки адекватности параллельного алгоритма моделирования последовательному аналогу: а) при различном числе ВУ; б) при различном прогнозном интервале

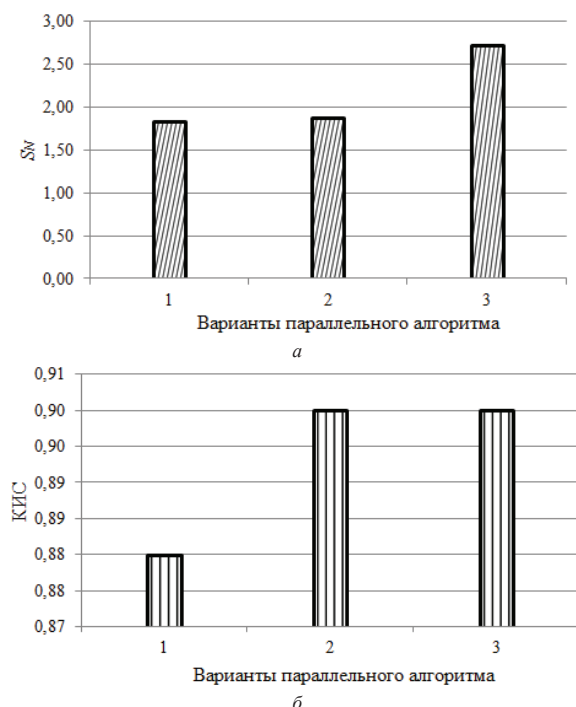


Рис. 4. Параллельное ускорение (а) и адекватность (б) различных вариантов параллельных алгоритмов последовательному исполнению алгоритма моделирования

На рис. 4 приведен фрагмент результатов сравнительного анализа, полученных при параллельном исполнении алгоритма моделирования (набор № 8, табл. 2). Эти результаты демонстрируют очевидное преимущество варианта с разделяемыми матрицами тенденций изменения и их синхронизации.

Скорость исполнения алгоритма моделирования при использовании дорогостоящего суперкомпьютерного кластера выше, чем при использовании недорогого кластера ПЭВМ. Однако параллельное ускорение, полученное на кластере ПЭВМ, будет заметно выше, чем на суперкомпьютерном кластере (рис. 5, набор № 5, табл. 2). Это связано с тем, что ВУ кластера ПЭВМ более «чувствителен» к уменьшению объема входных данных, чем ВУ суперкомпьютерного кластера.

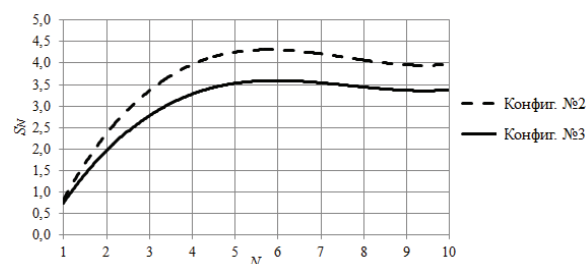


Рис. 5. Параллельное ускорение при использовании кластеров различной конфигурации

Заключение

Для существенного повышения производительности моделирования изменений ландшафтного покрова предложены варианты организации параллельного исполнения алгоритма моделирования, доступные к использованию на дорогостоящих суперкомпьютерах и недорогих кластерах ПЭВМ, отличающиеся тем, что в одном случае расчет вероятностных компонентов осуществляется на каждом ВУ независимо (без фрагментации и уменьшения геометрического размера исходных данных), а в другом предполагается моделирование с сохранением взаимного влияния вероятностных компонентов.

Для случая с сохранением взаимного влияния вероятностных компонентов предложены варианты параллельного исполнения алгоритма моделирования изменений ландшафтного покрова с собственными и с разделяемыми матрицами тенденций изменений на вычислительных узлах кластера. В качестве основного предложен и обоснован выбор варианта алгоритма с разделяемыми и синхронизируемыми матрицами тенденций изменения, характеризующегося более высокой адекватностью традиционному последовательному алгоритму моделирования.

Для оценки предположений об уровне адекватности распределенных версий основного алгоритма моделирования его традиционному последовательному аналогу проведены многочисленные эксперименты на тестовых разновременных данных с использованием дорогостоящей и недорогой конфигураций вычислительного кластера с варьируемым числом вычислительных узлов.

Результаты экспериментов позволили оценить производительность, параллельное ускорение и параллельную эффективность предложенного алгоритма, сформулировать следующие рекомендации по практическому применению параллельного ал-

горитма моделирования изменений ландшафтного покрова:

1. Несмотря на выбор в качестве основного алгоритма моделирования с синхронизацией матриц тенденций изменения, на практике возможны случаи, при которых может быть использован вариант алгоритма с собственными матрицами тенденций изменений (например, в случае низкой пропускной способности доступных каналов связи).
2. При возможности выбора доступного количества ВУ кластера следует учитывать геометрические размеры исходных тематических карт, обеспечивая в полученных после разделения фрагментах максимально возможное соответствие ландшафтно-классовой структуры фрагментов и исходного изображения.

СПИСОК ЛИТЕРАТУРЫ

1. Clarke K.C., Gaydos J. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore // *International Journal of Geographical information science*. – 1998. – V. 12. – № 7. – P. 699–714.
2. Syphard A.D., Clarke K.C., Franklin J. Using a cellular automaton model to forecast the effects of urban growth on habitat pattern in southern California // *Ecological Complexity*. – 2005. – № 2. – P. 185–203.
3. Замятин А.В. Анализ динамики ландшафтного покрова на основе данных дистанционного зондирования Земли // *Исследование Земли из космоса*. – 2006. – № 6. – С. 50–64.
4. Замятин А.В. Подход к моделированию изменений земной поверхности с использованием клеточных автоматов // *Известия Томского политехнического университета*. – 2005. – Т. 308. – № 6. – С. 164–169.
5. Наумов Л.А., Шалыто А.А. Клеточные автоматы. Реализация и эксперименты // *Мир ПЭВМ*. – 2003. – № 8. – С. 64–71.
6. Li X., Yeh A.G.O. Data mining of cellular automata's transition rules // *International Journal of Geographical Information Science*. – 2004. – № 18. – P. 723–744.
7. Cabral P., Zamyatin A. Advanced Spatial Metrics Analysis in Cellular Automata Land Cover Change Modeling // *DYNA, Sede Medellin*. – 2011. – V. 78. – № 170. – P. 42–50.
8. Plaza A.J., Chang C. *High Performance Computing in Remote Sensing*. – Chapman&Hall/CRC, 2008. – 496 p.
9. Kalluri S., JaJa J., Bader D.A., Zhang Z., Townshend J., Fallah-Adl H. High performance computing algorithms for land cover dynamics using remote sensing data // *International Journal of Remote Sensing*. – 2000. – V. 21. – № 6–7. – P. 1513–1536.

Также показано, что производительность обработки данных с помощью дорогостоящего вычислительного кластера в 3–3,5 раза выше, а с помощью кластера из недорогих типовых ПЭВМ в локальной сети в 4,0–4,2 раза выше, чем аналогичная обработка данных, выполненная последовательным аналогом, при использовании оптимального числа ВУ (4–6). Кроме того, показано, что адекватность алгоритма моделирования с синхронизацией матриц тенденций изменения последовательному аналогу будет различной для различных исходных данных и будет снижаться при увеличении используемых вычислительных узлов и горизонта прогнозирования.

Работа выполнена при поддержке РФФИ (грант № 11-07-00027а) и Госзадания (№ 8.8113.2013).

10. Ефимов С.С. Обзор методов распараллеливания алгоритмов решения некоторых задач вычислительной дискретной математики // *Математические структуры и моделирование*. – 2007. – № 17. – С. 72–93.
11. Eltgroth P.G., Bolstad J.H. Coupled Ocean / Atmosphere Modeling on High-Performance Computing Systems // *Eighth Society for Industrial & Applied Mathematics Conference on Parallel Processing for Scientific Computing* Minneapolis. – Minneapolis, 1997. – P. 1–10.
12. Ермаков С.Г. Метод устранения необходимости переключения вычислительных узлов при организации параллельной обработки информации // *Информационные технологии*. – 2007. – № 10. – С. 65–68.
13. Rokjin J.P., Daeok Y., et al. A general parallelization approach to improve computation efficiency in a global chemical transport model (GEOS-Chem) // *Geochemical Journal*. – 2010. – V. 44. – P. 323–329.
14. Sunderama V.S., Geist G.A. Heterogeneous parallel and distributed computing // *Parallel Computing*. – 1999. – № 25. – P. 1699–1721.
15. Воеводин В.В., Воеводин Вл.В. *Параллельные вычисления*. – СПб.: БХВ-Петербург, 2004. – 602 с.
16. Brennan R.L., Prediger D.J. Coefficient kappa: some uses, misuses, and alternatives // *Educational and Psychological Measurement*. – 1981. – № 41. – P. 687–699.

Поступила 21.11.2012 г.