

5. Gasser, L. MACE3J: Fast Flexible Distributed Simulation of Large, Large-Grain Multi-Agent Systems/L. Gasser, K. Kakugava [Электронный ресурс]. Режим доступа: <http://www.isrl.uiuc.edu/~gasser/papers/mace3j-aamas02-pap.pdf>
6. Macal, C. Tutorial On Agent-Based Modeling And Simulation/ C.Macal, M.North // Proceedings of the 2005 Winter Simulation Conference pp.
7. Bellifemine F.L. Developing Multi-Agent Systems with JADE/ F.L. Bellifemine, G.Caire, D.Greenwood // England: JOHN WILEY & SONS, LTD, 2007
8. Fujimoto, R.M. Parallel and Distributed Simulation Systems/R.M.Fujimoto//USA: Wiley, 2000. 300 p.
9. Wooldridge, M. Intelligent agents: Theory and Practice. / M.Wooldridge, N.Jennings // Knowledge Engineering Review. Vol 10, no. 2. ,pp. 115-152, Jun 1995.
10. Riley, A Distributed Agent Simulation EnvironmentWith Software-In-The-Loop Execu- tion/P.F.Riley, G.F.Spades // Proceedings of the 2003 Winter Simulation Conference
11. Nance, R.E. Distributed Simulation With Federated Models: Expectations, Realizations And Limitations/R.E.Nance // Proceedings of the 1999 Winter Simulation Conference. pp. 1026-1031.
12. Fujimoto, R.M. Distributed Simulation Systems/ R.M. Fujimoto // Proceedings of the 2003 Winter Simulation Conference. pp. 124-134
13. Шахов, В.В. Анализ средств противодействия одному виду атак типа "отказ в обслуживании"/ В.В.Шахов, А.С.Родионов // Вестник НГУ. Серия: Информационные технологии, том 6, вып. 2. – С. 80-88.

*Подкоарытов Д.И., аспирант,  
d.podkorytov@mail.ru - Институт вычислитель-  
ной математики и математической геофизики  
СО РАН, г.Новосибирск.*

УДК 504.064.37

## МАТЕМАТИЧЕСКИЙ МЕТОД И АЛГОРИТМ СЖАТИЯ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

С.Г. Катаев, М.Ю. Катаев

Рассматривается математический метод и алгоритм сжатия многомерных временных рядов, основанный на методе выделения структур. Метод относится к кластерным подходам основанных на графах. Математический подход апробирован на анализе многомерных временных рядов поведения профилей концентрации углекислого газа для некоторых районов земного шара. Представлены результаты работы алгоритма и оценки точностных характеристик работы алгоритма.

**Ключевые слова:** многомерные временные данные, сжатие, поле содержания газа

### Постановка задачи

Современные технические решения в области получения информации, в любых направлениях деятельности человека, позволяют добывать столько много информации, что сохранять ее становится сложно. Тем более, что в последствие эту информацию необходимо обработать и провести анализ полученных результатов. Поэтому задача сжатия информации является актуальной и необходимой для решения, внедрения в практику. Одним из направлений, где наблюдается избыточность добываемой и хранимой информации, является изучение атмосферы и поверхности Земли, а также климата. Для этих целей создаются банки данных о пространственно-временном распределении газовых составляющих, аэрозолей, термодинамических параметров атмосферы. К таким можно отнести центры хранения данных метеорологической информации NCEP (National Center for Environmental Prediction) и ECMWF (European Center Meteorological for Weather

Forecast). Помимо таких глобальных центров, в отдельных научных организациях создаются программы расчета, хранения получаемой информации о разнообразных физических параметрах, которые изменяются во времени. К таким расчетам можно отнести решение транспортной задачи переноса газов в атмосфере [1]. Данного вида расчеты возможно использовать при решении разнообразных задач оптики и физики атмосферы, экологии. Подобные варианты сочетания – расчет, избыточный по данным, и применение в практике – встречаются достаточно часто, что и побудило авторов к разработке математического подхода, представленного в данной статье.

Одним из наиболее развитых подходов изучения таких наборов данных является разработка многопараметрических моделей [2]. Однако, на наш взгляд, такие модели несут в себе методические погрешности, связанные с выбором типа модели, количеством параметров, нелинейностью и множества других факторов, определяющих точность.

### РАЗДЕЛ III. МОДЕЛИРОВАНИЕ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Нами предлагается использовать потенциал анализа многомерных полей данных непараметрическим подходом (метод выделения структур) с выделением составляющих, естественно представляющих изучаемое поле.

В данной статье приводится пример приложения математического подхода к анализу временных многомерных рядов углекислого газа с целью изучения однородности поведения в пространстве и времени и получении, в итоге, некоторой модели, которая позволяет сжать первичную информацию без существенной потери точности представления информации.

#### Описание метода

При решении модельных задач геофизики (расчет и обработка спутниковых сигналов, транспортные задачи переноса газов и аэрозолей, перенос излучения в атмосфере и различных средах и др.) на определенной территории, возникает ситуация, когда время вычислений становится значительным, что является следствием большого числа данных, необходимых для расчета. Результатом расчетов является функция  $F(t, x, y, z)$ , где  $t$  – время и  $\{x, y, z\}$  – пространственные переменные. Одним из решений по сокращению времени расчета, с таким набором данных, является применение мощной вычислительной техники (суперкомпьютер или вычислительный кластер). Однако такая техника дорога в использовании и еще не широко распространена. Нами предлагается подход, как решение задачи структуризации пространственно-временных наборов данных или поиска пространственно-временных структур, обладающих определенными свойствами, что позволяло бы существенно сократить объем данных и не потерять в точности при воспроизведении исходных массивов информации.

Для структуризации исходного набора данных предлагается использовать метод выделения структур (МВС), который хорошо зарекомендовал себя в задачах нахождения скрытых закономерностей в различных данных. Идеологически МВС близок к графовым методам кластерного анализа [3, 4], и поэтому алгоритм выделения структур, используемый нами для решения этой задачи, наиболее просто может быть описан на языке теории графов.

Соотнесем с каждым объектом  $a_i, i=1, \dots, N$  вершину  $x_i$  некоторого графа  $G=(X, E)$ , а каждому ребру  $e_{ij}=(x_i, x_j) \subset E, i, j=1, \dots, N$ , связывающему

вершины  $x_i$  и  $x_j$  в качестве стоимости  $c_{ij}$ , припишем величину, равную расстоянию  $d_{ij}$  между соответствующими объектами в пространстве признаков:  $c_{ij}=d_{ij}$ . Будем последовательно удалять из графа ребра  $e_{ij}$  со значениями, большими некоторого порога  $d_p^t$ , получая на каждом шаге  $G^t=(X^t, E^t)$  – граф на уровне  $t$ , множество ребер которого  $E^t=\{e_{ij} \in E: c_{ij} \leq d_p^t\}$ . Будем уменьшать величину  $d_p^t$ , и тогда, начиная с некоторого значения  $d_p^{t*}$ , исходный граф станет несвязным, появится  $k_t$  подграфов (компонент связности)  $G_1^t, G_2^t, \dots, G_{k_t}^t$  с различным числом вершин, среди которых могут быть подграфы, состоящие из одной вершины. Повторение процедуры дальнейшего уменьшения порогового значения  $d_p^t$  приводит к выделению все более связанных подграфов.

Будем понимать под структурой множество вершин графа, образующее на некотором шаге  $t$  компоненту связности и удовлетворяющее при этом некоторым наперед заданным аксиомам. В отличие от обычных методов кластерного анализа, МВС фиксирует компоненту связности графа как структуру, только, если определенные характеристики выделенного множества удовлетворяют сформулированным ограничениям – аксиомам структуры. Таким образом, МВС гарантирует, что каждая структура из результирующего набора будет обладать свойствами, заданными в аксиомах. Например, при исследовании полей какого-либо метеорологического параметра можно использовать следующую аксиому:

$$S = \{X_\alpha^t : G_\alpha^t = (X_\alpha^t, E_\alpha^t); \bar{R}(X_\alpha^t) \geq R_0, D(X_\alpha^t) \leq D_0\}.$$

Т.е., связность поля каждой структуры должна быть больше определенного значения  $R_0$ , а дисперсия – меньше  $D_0$ .

#### Описание задачи

В этой работе метод выделения структур был применен для исследования профилей концентрации CO<sub>2</sub> в различных регионах Земного шара, в частности, и с целью оценивания его возможностей нахождения областей с близким характером поведения CO<sub>2</sub>. В качестве исходных данных для моделирования были взяты результаты работы [1], представляющие собой 3-х часовые значения

ПОЛЗУНОВСКИЙ ВЕСТНИК № 2/1, 2012

профилей концентрации CO<sub>2</sub>, на пространственной сетке [0.5x0.5]. Профили даны на следующих 15 стандартных высотах (МПа): 996, 988, 968, 940, 900, 850, 800, 700, 600, 500, 400, 300, 200, 100, 30.

Были отобраны четыре, примерно однородных по свойствам подстилающей поверхности и равных по географическим размерам ареала, находящиеся на разных континентах: Западная Сибирь, австралийская пустыня (Австралия), штат Невада (Северная Америка) и дельта реки Амазонки (Южная Америка), а также две неоднородные области: районы Северной Европы и центральной Японии. Географические координаты рассматриваемых ареалов указаны в таблице 1.

Выбор районов с однородным типом ландшафта связан с возможностью спутникового мониторинга этих территорий и минимальным влиянием смеси типов поверхности на измеренный сигнал. Неоднородные территории по ландшафту позволяют оценить влияние значительных градиентов концентрации на результат работы метода.

В каждой области расположено по 377 точек сетки, покрывающих определенный регион согласно табл.1. Т.е., для всех изучаемых районов Земли (в терминах методов классификации) количество объектов  $N_{ob} = 377$ . Каждому объекту был поставлен в соответствие результирующий временной ряд, составленный из суточных временных рядов (8 значений) на каждой высоте. Таким образом, общее количество признаков  $N_f = 8 * 15 = 120$ .

Таблица 1

Но-мер	Область	Географические координаты
1	Западная Сибирь	(70-84)E; (56-62) N
2	Австралия (пустыня Гибсона и Большая пустыня Виктория)	(122 -136) E, (24-30) S
3	Амазонка	(50-64) W, (0-6) S
4	шт. Невада	(108-122) W, (42-36) N.
5	Северо-Западная Европа	(0-14) E; (50- 56) N
6	Япония	(130 -144) E, (40-34) N

Как уже отмечалось выше, МВС позволяет отыскивать в данных структуры с заранее определенными свойствами. Заявленным целям данного исследования – нахождение областей с похожим поведением профилей и проверка аппроксимационных возможностей метода, на взгляд авторов, отвечает требование к сильной связности поля в каждой вы-

деляемой структуре (классе), т.е. высокому значению среднего по структуре коэффициента детерминации  $R^2$ . Конкретное граничное значения  $R^2$ , при котором выявленное в результате работы алгоритма подмножество точек фиксировалось в качестве структуры, для каждого ареала подбиралось опытным путем таким образом, чтобы количество возникающих классов было не слишком большим и не слишком маленьким. Так, для Западной Сибири и австралийской пустыни граничное значение коэффициента детерминации  $R^2$  оказалось равным 0.99, для ареала Амазонки, штата Невада и Европы -  $R^2=0.93$ , для Японии  $R^2=0.90$ . Изолированные точки, т.е., точки, не попавшие ни в одну из структур, затем присоединялись к ближайшим сформированным классам. Количество таких точек колебалось в пределах от 5% (пустыни Австралии) до 15% (Западная Европа).

В итоге работы алгоритма по анализу многомерных временных рядов выделяются такие пространственно-временные структуры, средний профиль в которых с минимальной дисперсией отличаются от заданных в исходной базе данных. Таких областей (структур) получается примерно на порядок меньше, чем исходных точек.

#### Результаты работы алгоритма

Некоторые полученные результаты пространственно-временного анализа, приведены ниже. На рисунке помещены реальные профили (взяты из базы данных) для двух точек района Амазонки в сравнении с соответствующими типовыми профилями (полученными при восстановлении). Типовой профиль это усредненный по структуре временной ряд. Видно, что поведение профилей в различных в разных точках одного и того же региона существенно отличается друг от друга. Несмотря на это качество восстановленных профилей в каждой точке ареала по небольшому количеству типовых рядов оказывается неплохим (минимальное отличие от реального).

На рис.2 помещены значения средней квадратичной ошибки (СКО) во всех точках для района Амазонки. Видно, что эта величина не превышает 0.05% от среднего значения концентрации CO<sub>2</sub>.

В таблице 2 приведены некоторые характеристики выявленных структур для рассмотренных в статье ареалов

#### Обсуждение результатов

Ошибка восстановленных профилей концентрации CO<sub>2</sub> зависит от дисперсии и связности всего изучаемого поля конкретной области, а также от числа используемых клас-

### РАЗДЕЛ III. МОДЕЛИРОВАНИЕ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

сов. Так, из таблицы 2 следует, что наиболее однородные условия наблюдаются Западной Сибири и Австралийской пустыни, связность поля концентрации  $\text{CO}_2$  для которой равна 0.970. Неудивительно, что хорошую точность здесь удалось достичь с использованием всего 19 классов для территории Западной Сибири. Минимальное среднее значение СКО, практически на порядок меньше, чем для Западной Сибири получилось для Австралии, ареала с минимальной дисперсией.

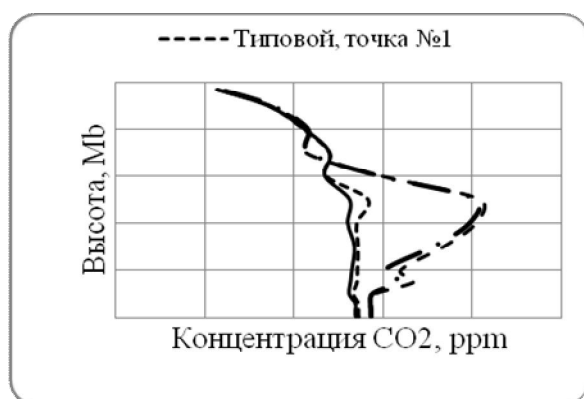


Рисунок 1 Сравнение типовых и реальных профилей для 1 и 51 точек. Район Амазонки.

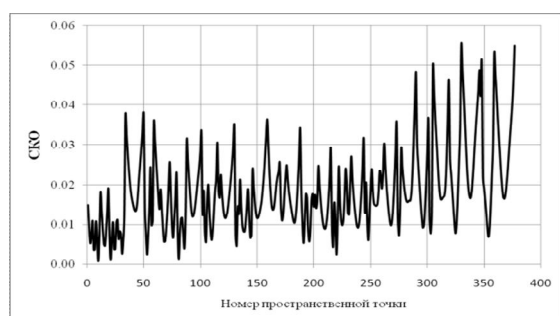


Рисунок 2. СКО для всех точек области. Район Амазонки

Выбранные для исследования регионы Северо-западной Европы и Японии включают в себя не только различные типы поверхности, но и крупные промышленные центры (Франкфурт, Берлин, Иокогама, Токио, Осака), т.е., характеризуются в этом плане существенной неоднородностью. Поэтому для этих районов качество аппроксимации профилей  $\text{CO}_2$  отличаются от однородных областей, что выражается повышенной СКО (примерно в четыре раза).

Для локализации областей с большим значением СКО была проделана следующая процедура. Каждой пространственной точке  $i$ , принадлежащей классу  $j$ , присваивалось среднеклассовое среднее значение  $A_j$  концентрации  $\text{CO}_2$ . Затем были построены карты

пространственного распределения  $A_j$ . В качестве примера на рис.3 приводится такая карта для территории Западной Европы.

Таблица 2. Некоторые характеристики выявленных структур для рассмотренных областей.

№	Область	Количество структур (классов)	Среднее значение концентрации исходного поля	Дисперсия исходного поля	Связность исходного поля (R2)	Среднее значение СКО
1	Западная Сибирь	19	390.639	12.533	0.970	0.0217
2	Австралийская пустыня	37	384.658	0.238	0.969	0.0027
3	Амазонка	42	387.643	2.035	0.814	0.0180
4	шт. Невада	37	389.611	7.294	0.881	0.0318
5	Западная Европа	44	391.000	29.337	0.704	0.0850
6	Япония	32	391.06	22.750	0.818	0.0839

Из рисунка 3 видно, что расположение классов хорошо коррелирует с типом подстилающей поверхности (земля-вода), и, как следовало ожидать, локализация областей с большой СКО, приходится на промышленно развитые районы вблизи Франкфурта и Берлина, а для территории Японии для Иокогамы-Токио и Осаки.

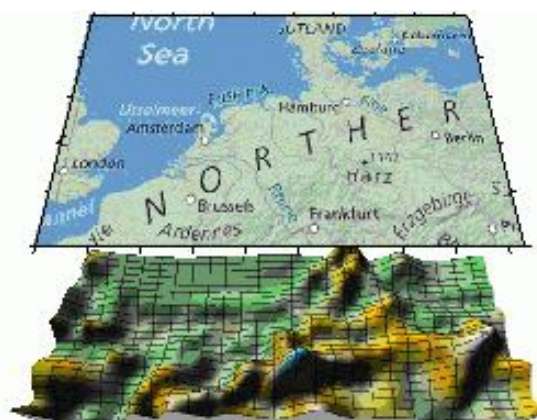


Рисунок 3 Пространственное распределение СКО района центральной Европы

Причина выбросов СКО в этих районах очевидна – большой градиент концентрации углекислого газа. Процедура полной класси-

## ПАРАЛЛЕЛЬНЫЙ ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ

фикации присоединила подобные точки к существующим классам, в результате чего им было поставлено в соответствие все типовые характеристики соответствующего класса, которые естественно сильно отличаются от индивидуальных показателей этих точек.

Отсюда следует, что в подобных ситуациях необходимо такого типа точки рассматривать либо в качестве отдельных классов, либо проводить в этих районах интерполяцию, увеличивая тем самым плотность точек.

### Заключение

Таким образом, предлагаемый в работе подход анализа многомерных временных рядов позволяет обеспечивать необходимую (наперед заданную) точность при аппроксимации временных рядов концентрации CO<sub>2</sub>. Разбив всю поверхность Земли на несколько сотен регионов, мы тем самым сокращаем необходимый объем вычислений в случае сетки 2.5°x2.5°, по крайней мере на 2 порядка.

Разбиение на районы выделяет области с близким поведением исследуемых функций, что дает возможность корректно проводить

процедуру интерполяции в любой точке поверхности Земли. Метод дает наглядное представление о поведении концентрации. Он универсален, для его применения не нужны априорные знания о природе исследуемого поля, характере его поведения.

### СПИСОК ЛИТЕРАТУРЫ

1. Maksyutov S. Atmospheric CO<sub>2</sub> simulations with a high resolution model and synoptic scale variability of CO<sub>2</sub> column / S. Maksyutov, R. Onishi, M. Naja, A. Yaremchuk, P.K. Patra, G. Inoue // CGER-I058-2007, v.14, 2007. – pp. 49-54
2. Состояние и комплексный мониторинг природной среды и климата. Пределы изменений – М.: Наука, 2001. – 242 С.
3. Айвазян С.А. Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин – М.: Финансы и статистика, 1989. – 607 с.
4. Мандель И. Д. Кластерный анализ. — М.: Финансы и Статистика, 1988. – 432 С.

*д.т.н., профессор Kamaev M.IU. – профессор каф АСУ ТУСУР, kataev.m@sibmail.com, 8(3822) 70-15-36*

УДК 517.9, 533.6.011

## ПАРАЛЛЕЛЬНЫЙ ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ

С.В.Тимченко

Рассматривается параллельный генетический алгоритм для решения задач многокритериальной оптимизации. При его помощи решена задача определения параметров траектории входа тела в атмосферу Земли, минимизирующих величину интегрального конвективного теплового потока в точке торможения затупленного тела при одновременной минимизации толщины теплозащитного покрытия. В качестве исходной математической модели для расчета теплового потока используются уравнения тонкого вязкого ударного слоя с учетом неравновесного характера протекания химических реакций, многокомпонентной диффузии и сопряженного характера теплообмена на поверхности тела.

**Ключевые слова:** генетический алгоритм, многокритериальная оптимизация, вязкий ударный слой, равновесная температура, параллельный алгоритм

### Введение

При решении большого числа практических задач приходится сталкиваться с необходимостью нахождения решений, удовлетворяющих нескольким, зачастую конфликтующим между собой, критериям. В связи с этим, решение задачи заключается не в нахождении какого-то одного решения, а в отыскании некоторого множества решений, каждое из которых будет превосходить другие хотя бы по одному критерию. Такие решения, как правило, называются оптимальными по

Парето. Необходимость отыскания целого множества решений чрезвычайно усложняет задачу оптимизации и делает практически непригодными большинство классических методов оптимизации. Задача усложняется еще и тем, что надо не только найти решения, максимально близкие к истинному множеству (или фронту) Парето, но и обеспечить максимально возможное различие между такими решениями (т.е. охватить возможно большую часть этого фронта).