

УДК 550.34, 551.4

ПРИМЕНЕНИЕ МНОГОМЕРНОГО ЕОФ-АНАЛИЗА В ГЕОИНФОРМАТИКЕ

Коротченко Р.А., к.т.н., старший научный сотрудник Тихоокеанского океанологического института им. В.И. Ильичева ДВО РАН, Владивосток, e-mail: rotankor@mail.ru

Самченко А.Н., научный сотрудник Тихоокеанского океанологического института им. В.И. Ильичева ДВО РАН, e-mail: samchenso@poi.dvo.ru

Ярошук И.О. д.ф.-м.н., заведующий лабораторией Тихоокеанского океанологического института им. В.И. Ильичева ДВО РАН, e-mail: yaroshchuk@poi.dvo.ru

Ключевые слова: геоинформатика, метод ортогональных функций, анализ полей данных, сингулярный, спектральный, модель рельефа, тектоника масштабов и структур.

Введение

Численная обработка экспериментальных данных является принципиальной основой геоинформатики, включающей разделы геофизики, океанологии, математической геологии, географии и многое другое. Геоинформатика, как наука и технология, опирается на цифровые модели, описывающие представление среды в виде суперпозиции большого количества взаимодействий между различными процессами или, иначе говоря, – разных степеней свободы.

Возможно использовать наиболее подробное описание происходящих процессов, полагаясь на ресурсы суперкомпьютеров, однако в этом случае объем информации становится колоссальным, а вопрос интерпретации данных остается открытым. Для анализа геоинформационных полей данных необходимы такие методы, которые позволяют уменьшить размерность системы и выявить структуры, наиболее полно объясняющие наблюдаемые вариации параметров.

Как правило, обрабатываемые данные содержат не только достоверные сведения, но и ошибки, связанные с условиями проведения наблюдений и особенностями применяемой аппаратуры. В таком случае требуется обработка данных с целью фильтрации «шумовых» составляющих наблюдений и селекции достоверной информации.

В процессе интерпретации наблюдений надлежит провести комплексный анализ морфологии выявленных структур и описать зависимость их характеристик от вариации внешних, «управляющих» параметров. Методы, основанные на оценках средних значений наблюдаемой величины и ее дисперсии, занимают первоочередное место в процессе обработки накопленного материала.

В рамках этого подхода метод естественных (эмпирических) ортогональных функций (ЕОФ) и его разновидности, используемые в анализе пространственно-временных данных, уже стали одним из основных рабочих инструментов в метеорологии, физике атмосферы и океана [1-3]. Однако многомерные варианты ЕОФ-

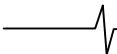
Рассмотрены принципы применения многомерного ЕОФ-анализа для обработки полей геоинформационных данных. Описаны процедурные этапы обработки методом двумерного сингулярного спектрального анализа (2D-SSA) для задач структурной пространственной декомпозиции скалярных полей. Для фрагмента северо-западной части шельфа Японского моря создана цифровая модель рельефа, выполнена двумерная ЕОФ-декомпозиция и реконструкция модели с интерпретацией выделенных в совокупности данных геоморфологии и тектоники пространственно-временных масштабов и структур.

анализа еще крайне редко применяются в практике обработки натурных наблюдений. Кроме работ с теоретическим акцентом [4], авторам не известны публикации о практических приложениях многомерных ЕОФ-методов в задачах геофизики и геоинформатики.

В отличие от геофизики и океанологии, геология и география являются в основном описательными науками и ограничиваются упрощенными способами работы с цифровыми данными. Применение современных численных методов в геоинформатике и, в частности – в геологии, качественно улучшает интерпретацию накопленных сведений на основе количественного анализа. Например, применение цифровых моделей в геоморфологии дает возможность перейти к пониманию процессов рельефообразования с позиций анализа пространственно-временных масштабов тектонических процессов, согласовывая шкалы времени и энергии, требуемых для развития структур рельефа различных пространственных размеров [5-7]. Рассмотренные в статье процедуры 2D-SSA и практика его применения для интерпретации данных геоморфологии самого крупного залива Японского моря демонстрируют практическую ценность подобной цифровой обработки.

Пространственные структуры и метод ЕОФ

Метод естественных ортогональных функций применяется для обработки и анализа скалярных полей данных, меняющихся во времени. Он позволяет выполнить статистическую обработку, которая вскрывает иерархию основных факторов и подавить влияние типичных погрешностей измерений, присутствующих в условиях натурных наблюдений. Метод выделяет пространственные структуры, описывающие изменчивость сигнала и оценивает относительный вклад каждой структуры в исходное поле. Теория и принципы ЕОФ-разложения и его разновидностей описаны в работах [1-4].



Основная идея, заложенная в методе, заключается в следующем. Предположим, что реализации полей данных представлены набором векторов $\{f_i, i = 1..k\}$ в N -мерном пространстве для последовательности моментов времени. Все вектора исходят из начала координат. Если исходные данные коррелированы, то вектора будут кластеризоваться вдоль каких-то выделенных направлений. Задача метода ЕОФ заключается в поиске такого ортогонального базиса $\{e_1, e_2, \dots, e_N\}$ в N -мерном пространстве, чтобы вектор e_1 был направлен к самому большому кластеру, вектор e_2 к следующему по величине и т.д. При этом сумма квадратов проекций всех векторов f на направления $\{e_1, e_2, \dots, e_N\}$ убывает строго последовательно. Из-за ортогональности векторов $\{e_1, e_2, \dots, e_N\}$ найденные структуры называются ортогональными функциями. А так как новый базис строится по самим данным, а не выбирается априорно, эти функции называются естественными или эмпирическими.

С прикладной точки зрения ключевым моментом является возможность подбора ортогонального базиса, построенного по самим данным и упорядоченного так, что «раскладка» исходных данных в базисе согласована по убыванию вклада в общий максимум вариации. Это означает, что ЕОФ расщепляет исходные поля данных на ортогональные составляющие – «моды данных», которые описывают статистически значимые, упорядоченные структуры. Эти моды связаны с собственными значениями, соответствующими масштабу вклада в исходное поле. В отличие от обычного спектрального анализа, где базис определен изначально набором синусов и косинусов, в ЕОФ функции базиса разложения определяются из самих данных и являются оптимальными в статистическом смысле.

Следует подчеркнуть, что метод расщепляет исходные поля данных на «моды данных», которые не всегда имеют прямое физическое толкование. Один и тот же физический процесс может давать вклад в разные моды, и одна и та же мода может быть результатом действия более чем одного физического процесса. Интерпретация выделенных мод зависит от физического смысла полной вариации. В некоторых случаях смысл достаточно прозрачен и связан с энергетическими параметрами исследуемых процессов, в иных случаях это следует выяснять более тщательно. Общее толкование состоит в декомпозиции поля данных на трендовую, низкочастотные и высокочастотные составляющие.

Обработка данных в методе 2D-SSA

В силу большого объема вычислений, в настоящее время наибольшую применимость нашли одномерные и двумерные разновидности метода. Двухмерный сингулярный спектральный анализ (2D-SSA) является вариантом ЕОФ. Он эффективно работает в задачах выделения сигнала из шумов, причем удаление шумов происходит без ухудшения сигнала, как обычно бывает при использовании сглаживающих фильтров. Детали алгоритма 2D-SSA рассмотрены в работе [4]. Ниже рассмотрена последовательность процедур обработки двумерных скалярных полей данных в рамках данного метода.

Пусть определено единовременное поле наблюдений скалярной величины f на пространственной сетке, значения которого могут быть представлены матрицей

$$F = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1N} \\ f_{21} & f_{22} & & f_{2N} \\ \dots & \dots & \ddots & \dots \\ f_{M1} & f_{M2} & \dots & f_{MN} \end{bmatrix}.$$

При высокой плотности сетки наблюдений предпочтительно выполнить статистическое усреднение по сегментам равной площади. В этом случае элементы матрицы будут соответствовать номеру сегмента.

Алгоритм включает два этапа – декомпозицию и реконструкцию. Параметрами алгоритма являются размеры окна сглаживания скользящим средним (n, m).

На этапе декомпозиции по данным поля формируется специальная траекторная матрица W . Для этого с помощью окна скользящего среднего выполняется отображение (векторизация) двумерной матрицы F в четырехмерное пространство: элементы поля f , попадающие в окно скользящего среднего, приводятся к одному столбцу траекторной матрицы W . Например, для окна скользящего среднего с $n = 2, m = 2$ отображение первого сегмента имеет вид

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \rightarrow (f_{11} \ f_{21} \ f_{12} \ f_{22})^T.$$

Построенная таким образом траекторная матрица далее приводится к блочной Ханкелевой структуре с симметрией блоков относительно главной диагонали [4].

На следующем шаге процедуры декомпозиции вычисляется сингулярное разложение блочной Ханкелевой траекторной матрицы вида

$$W = \sum_{i=1}^d W_i = \sum_{i=1}^d U_i \sqrt{\lambda_i} V_i^T,$$

где $\{\lambda_i\}$ – собственные значения, упорядоченные по модулю, $\{U_i\}, \{V_i\}$ – ортонормальные системы левых и правых сингулярных собственных векторов. В двухмерном случае матрицы собственных векторов называются собственными полями. Соответствие вклада проекции поля данных на подпространства собственных векторов в общую совокупность определяется амплитудами собственных значений. Относительный вклад в полную вариацию от сингулярного значения определяется как

$$\lambda_k \cdot \left(\sum_{i=1}^d \lambda_i \right)^{-1}.$$

Этап реконструкции предполагает селекцию набора собственных полей (мод) с целью формирования более простого представления анализируемых данных и фильтрации «шумовых» компонент. Реконструкция предполагает «сборку» поля данных в ограниченном наборе мод. Для этого выполняется суммирование выборки $\{W_i\}$ с последующим восстановлением блочной Ханкелевой структуры $\tilde{W}$ и далее выполняется трансформация к матрице обработанного поля данных наблюдений F :

$$\langle W \rangle = \sum_{i=1}^L W_i, \quad \langle W \rangle \rightarrow \tilde{W}, \quad \tilde{W} \rightarrow F.$$

Реконструкция и интерпретация результатов предполагает несколько вариантов «сборки», выделяя часть информации, заложенной в главной моде разложения с наибольшим собственным значением и в других наборах базисных векторов. Полученные пространственные структуры представляют декомпозицию исходного поля данных на сумму компонентов, включающую гладкую составляющую, осцилляции разного уровня и «шум».

При анализе двумерного поля данных метод позволяет выделить только фиксированные пространственные колебания. Для анализа пространственно-временных структур можно использовать модификацию метода по принципу расширенных ЕОФ (Extended EOF): в этом случае матрица F формируется более сложным образом – в начале формируется сама матрица F , затем к ней стыкуется ее копия, сдвинутая на лаг l , далее добавляется еще одна копия матрицы со сдвигом на $2l$ и т.д. К полученной большой матрице применяется рассмотренный выше алгоритм анализа. Сингулярные разложения позволяют восстановить последовательность сдвинутых карт с эволюционирующими структурами, связанных с выделенными модами.

Версия метода для анализа скалярных полей данных более высокой размерности сохраняет все принципиальные особенности алгоритма 2D-SSA с учетом степенного роста объема вычислений.

ЕОФ-анализ геоморфологии залива Петра Великого Японского моря

Геоморфология охватывает совокупность знаний о геологическом строении и происхождении рельефа [8]. Востребованность знаний о структуре и характеристиках строения рельефа связана с их важностью для исследований в близких областях: геологии и геофизике. Понимание пространственно-временных особенностей развития территорий позволяет исследовать историю тектонической активности – представлять ее прошлое и оценивать развитие возможных событий в ближайшем будущем. Возможность и обоснованность пространственно-временного анализа рельефообразующих процессов представлены в работах [5-7].

Создание сетки высот цифровой модели рельефа (ЦМР) является первым этапом исследования. Использовались цифровые данные спутниковой альтиметрии и совокупность экспериментальных наблюдений батиметрии [9-12]. Перепад высот на подготовленной сетке составил около 3000 м – от глубины 2250 м до высоты 920 м. Для уменьшения вычислительной нагрузки была сформирована расчетная сетка с ячейками 400×300 м. С целью удаления влияния на результаты расчетов значительных перепадов высот альтиметрия трансформировалась логарифмированием высот и глубин.

Согласно методу 2D-SSA, на первом этапе выполняется декомпозиция данных альтиметрии – разложение поверхности рельефа в базисе ортогональных полей (мод данных). Методом подбора были выбраны подходящие параметры окна скользящего среднего, позво-

ляющие реконструировать рельеф с надлежащим качеством. Из рассчитанных первых 40 мод оказалось достаточно использовать старшие 10 мод, чтобы выделить основные элементы рельефа (рис. 1). Суммарный относительный вклад от собственных значений 10 мод в полную вариацию составил 0.96 (рис. 2). Компоненты рельефа из более высоких мод разложения давали общий вклад 4%, рассматривались как «уровень шума» и были отфильтрованы.

Основной вопрос, связанный с применением метода ЕОФ, – физическая интерпретация результатов разложения – рассматривался в соответствии с понятием «энергии рельефа». Под энергией рельефа понимается показатель потенциальной интенсивности или возможного проявления тех или иных рельефообразующих процессов. Структуры, определяющие высоты местности, характеризуются различными уровнями энергии, требуемой для их натурального формирования [5,6]. Модовый состав предполагает комбинации таких структур. При этом полная вариация разложения пропорциональна потенциальной энергии рельефа, а собственные значения соответствуют вкладу модовой структуры в полную вариацию (рис. 1).

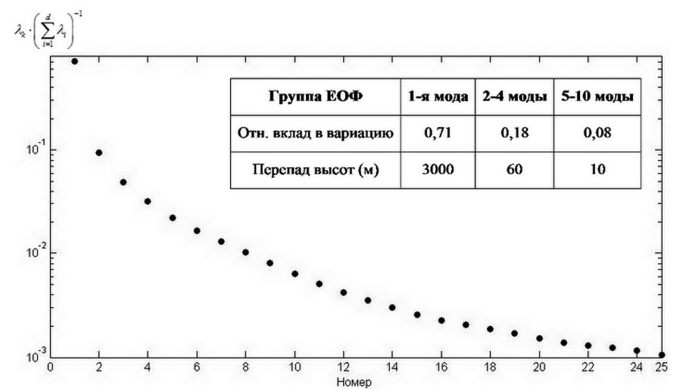


Рис. 1. Относительный вклад рассчитанных собственных значений в полную вариацию

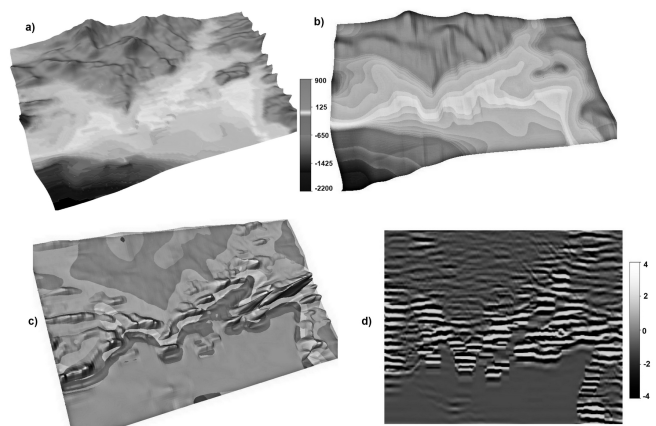


Рис. 2. ЕОФ-реконструкция ЦМР: а) первые 10 мод разложения, б) структура 1-й моды, в) изопроекция структур 2-4 мод на поверхность 1-й моды, д) структуры 5-10-й мод

Реконструкция в базисе 1-10 мод сохранила принципиальный рельеф батиметрии и побережья. Изолинии и тональный цвет на рис.2а отчетливо прорисовывают генеральную прибрежную зону и структуру глубин шельфа,

выделяют острова, внутренние заливы. Светлый фон соответствует прибрежному мелководью и зоне малых высот. Контрастно выделены свал глубин на юго-западе и отроги Сихотэ-Алиня на севере.

Поверхность, представленная 1-й модой разложения на рис. 2b представляет основную тектоническую форму. Относительно фундамента выделяется гористая деформация и глубоководный свал. Модель рельефа на основе 1-й моды разложения является репродукцией древней формы залива и прибрежной территории с обширной мелководной прибрежной областью.

Структуры 2-4 мод с вертикальными масштабами порядка десятков метров (рис. 2c) связаны с тектоническими процессами меньшего масштаба, происходивших в более позднее время и короткие сроки. Структуры протягиваются с юго-запада на северо-восток на дистанцию более сотни километров. ЕОФ-анализ явно выделяет формы, связанные с деформациями рельефа вдоль береговой зоны древнего залива – появлением островов, скалистых мысов и крупных бухт.

Выделенные на уровне 5-10 мод структуры рельефа (рис. 2d) имеют амплитуды высот до 5 м. Такие уровни несколько меньше высот природных образований, что связано со сглаживанием и вычислениями на грубой сетке. Интерпретация рельефа для масштаба 5-10 мод связана с завершающей стадией формирования прибрежной зоны в результате экзогенных процессов и элементов вулканизма.

Заключение

Многомерные ЕОФ-разложения являются достаточно новыми методами в арсенале методов цифровой обработки данных, в связи с вычислительной ресурсоемкостью и спецификой интерпретации. Однако, потенциальные возможности методов указывают на широкий спектр и актуальность применения при обработке данных не только для геоинформатики, но и в смежных областях знаний. Анализ пространственно-временной изменчивости и выделение масштабных шкал структурных составляющих полей данных с помощью ЕОФ-анализа расширяют возможности статистической обработки материалов.

Рассмотренные процедуры и алгоритмы нового метода 2D-SSA эффективно реализуются на практике. Комплексный подход, включающий создание цифровой геоинформационной модели, методы ЕОФ-анализа, совокупность описательных данных и характерные энергетические и временные особенности тектоники геосферы, существенно расширяет возможности количественного анализа и интерпретации природных данных в геоинформатике.

Литература

1. Глуховский А.Б., Фортус М.И. Оценка статистической надежности эмпирических ортогональных функций // Изв. АН СССР. Физ. Атмосф. и океана, 1982. Т.18 № 5. С. 451-459.
2. Elsner J.B., Tsonis A.A. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Plenum Press. New York, 1996. 164 p.
3. Navarra A., Simoncini V. A Guide to Empirical Orthogonal Functions for Climate Data Analysis. Springer, 2010. 152 p.
4. Golyandina N.E., Usevich K.D., Florinsky I.V. Filtering of Digital Terrain Models by Two Dimensional Singular Spectrum Analysis // International Journal of Ecology & Development, 2007. Vol. 8. No. F07. P. 81-94.
5. Голицын Г.С. Статистическое описание рельефа поверхности планеты и его эволюции // Физика Земли, 2003. № 7. С. 3-8.
6. Голицын Г.С. Природные процессы и явления: волны, планеты, конвекция, климат, статистика. – М.: ФИЗМАТЛИТ, 2004, 344 с.
7. Turcotte D.L. Fractals and chaos in geology and geophysics. Cambridge University Press, New York, 1997. 398 p.
8. Ветренников В.В. Геологическое строение Сихотэ-Алинского заповедника и центрального Сихотэ-Алиня. Владивосток: Дальневосточное книжное издательство, 1976. Вып. 6. 167 с.
9. Коротченко Р.А., Самченко А.Н., Ярошук И.О. Применение статистических методов в изучении рельефа шельфовой зоны на примере залива Посьета Японского моря // Вестник ДВО РАН, 2011. № 6. С. 54-59.
10. URL: <http://www.gdem.aster.ersdac.or.jp>.
11. URL: http://www.gebco.net/data_and_products/gridded_bathymetry_data.
12. URL: <http://www.ngdc.noaa.gov/mgg/global/global.html>.

APPLICATION OF MULTIDIMENSIONAL EOF-ANALYSIS IN GEOINFORMATICS

Korotchenko R.A., Samchenko A.N., Yaroshchuk I.O.

Principles of multidimension EOF-methods for processing and analyzing of geospatial data fields are considered. Procedural stages of two-dimensional singular spectrum analysis (2D-SSA) for spatial structural decomposition of the scalar fields are described. For a part of the Sea of Japan north-western shelf a digital terrain model was developed and two-dimensional EOF-decomposition and further reconstruction of the model were carried out to select and to interpret spatial and temporal scales for observed geomorphological and tectonics structures.