

Использование вариационного алгоритма фильтрации при восстановлении данных поверхностной температуры

В. С. Кочергин, С. В. Кочергин, С. В. Станичный

*Морской гидрофизический институт РАН, Севастополь, 299011, Россия
E-mail: vskocher@gmail.com*

Для восстановления пропущенных данных измерений поверхностной температуры рассматривается алгоритм фильтрации. В его основе лежит вариационный подход, поиск экстремума квадратичного функционала, представление функции, аппроксимирующей данные измерений в виде разложения по ортогональным функциям. Процедура основана на плоском вращении исходной системы при помощи максимизации квадратичного функционала, характеризующего соотношения коэффициентов системы и её правых частей. При этом используется априорная информация о возможных максимальных значениях коэффициентов разложения. В результате работы алгоритма получается эквивалентная система, в которой уравнения выстроены по рангу информативности. Осуществлена апробация метода на тестовых примерах и реальных данных. В качестве исходных данных был взят последовательный многолетний ряд поверхностной температуры для одного и того же района в акватории Чёрного моря на протяжении 35 лет. Данные измерений выбирались из имеющегося ряда случайным образом. Полученные при этом пропуски в данных имели продолжительность от одних до семи суток. Для двух произвольно выбранных лет были проведены эксперименты по восстановлению пропущенных данных. Процедура реализовывалась двумя способами: для самих значений температуры и для отклонений от средних значений. Были восстановлены данные о поверхностной температуре для моментов времени, когда наблюдались пропуски в измерениях. В результате численных экспериментов показана эффективность вариационного алгоритма фильтрации при реализации процедуры восстановления пропусков в данных измерений поверхностной температуры.

Ключевые слова: спутниковые данные, поверхностная температура, вариационный алгоритм

Одобрена к печати: 01.11.2018

DOI: 10.21046/2070-7401-2018-15-7-9-14

Введение

В настоящее время основным источником информации о температуре морской поверхности являются спутниковые данные, получаемые в ИК-диапазоне. Однако наличие облачности приводит к пропускам данных. Поэтому важной является задача восстановления данных для таких промежутков времени (gap filling). Существуют различные подходы к решению этой задачи, например, наиболее популярна процедура оптимальной интерполяции (Пухтыр и др., 2009; Reynolds et al., 2007).

В настоящей работе предлагается решать данную задачу на основе вариационного алгоритма фильтрации (Страхов, 1991). В качестве тестового массива использованы спутниковые данные о температуре поверхности (GHRSSST Level 4 AVHRR_OI Global Blended Sea Surface Temperature Analysis (GDS version 2)), полученные из архива PO DAAC с использованием сервиса LAS (<https://podaac-tools.jpl.nasa.gov/las/UI.vm>). Подобный вариационный алгоритм ранее применялся для восстановления амплитуд мод береговых захваченных волн (Кочергин, Янковский, 1995), а также при решении различных океанологических задач, связанных с представлением полей в виде разложения по некоторому ортогональному базису (Еремеев и др., 2002).

Описание алгоритма

Пусть в некоторые моменты времени t_i ($i = 1, \dots, N$) в течение интервала времени T мы имеем данные измерений температуры θ_δ^i в одной и той же точке. Разложив θ_δ в ряд по ортогональным функциям (синус и косинус), в итоге имеем переопределённую систему алгебраических уравнений:

$$Az = \theta_\delta, \quad \theta_\delta = \theta + \delta\theta, \tag{1}$$

где z — неизвестный вектор коэффициентов разложения; $\delta\theta$ — вектор, описывающий помеху.

Поиск этого неизвестного вектора осуществляется за счёт плоского вращения системы (1). Коэффициенты такого преобразования ищутся исходя из максимума квадратичного функционала специального вида. В итоге получается другая система алгебраических уравнений, эквивалентная (1):

$$\hat{A}z = \eta_\delta, \tag{2}$$

где $\hat{A} = V^T A$, $\eta_\delta = V^T \theta_\delta$, V^T — ортонормальная матрица, конструируемая как предел произведения элементарных матриц плоского вращения:

$$\left. \begin{aligned} V^T &= \lim_{n \rightarrow \infty} V_n^T, \\ V_n^T &= \prod_{k=1}^n T_{i(k),j(k)}(\varphi_k). \end{aligned} \right\} \tag{3}$$

Очевидно, что главная задача состоит в определении углов φ_k .

Положим $V_n^T \theta_\delta = w_\delta^{(n)}$:

$$\left. \begin{aligned} V_{n+1}^T \theta_\delta &= w_\delta^{(n+1)}, \\ w_\delta^{(n+1)} &= T_{i(n+1),j(n+1)}(\varphi_{n+1}) w_\delta^{(n)}. \end{aligned} \right\} \tag{4}$$

Для определения угла φ используем условие:

$$\left. \begin{aligned} \Phi_i(\varphi) &= \max_{\varphi}, \\ \Phi_i(\varphi) &= \left(\sum_{k=1}^M |a_{ik}^{(n+1)}|^2 \right) - \frac{|w_{i,\delta}^{(n+1)}|^2}{M^2}. \end{aligned} \right\} \tag{5}$$

В настоящей работе считается, что априорно известна следующая информация:

$$\|z\|_E^2 \leq M^2, \quad \|\delta\theta\|_E^2 = \delta^2, \tag{6}$$

где δ^2 и M^2 — заданные числа. Следует отметить, что число используемых гармоник меньше числа данных измерений.

После построения матрицы \hat{A} уравнения системы проверяются на выполнение следующего условия:

$$\|\hat{a}^{(i)}\|_E^2 M^2 \geq w_{i,\delta}^2, \tag{7}$$

где $\hat{a}^{(i)}$ — i -я вектор-строка матрицы \hat{A} . При решении поставленной задачи все уравнения нашей системы удовлетворяли этому условию. Уравнения системы сортируются по убыванию ранга информативности. То есть система выстраивается таким образом, что сверху располагаются наиболее информативные уравнения. Чем меньше величина $\rho_i = \|\hat{a}^{(i)}\|_E^2 M^2 - w_{i,\delta}^2$, тем меньшее значение имеет ранг уравнения ($\text{rang}(i) = 1, 2, \dots, N$). Далее из системы выбираются верхние n уравнений.

Таким образом, начальная система уравнений при помощи ортогонального преобразования плоских вращений трансформируется к эквивалентной системе. Углы поворота находятся из максимума функционала, характеризующего соотношение коэффициентов системы, априорной информации о решении и правых частях. Алгоритм осуществляется итерационно до тех пор, пока $\sin(\varphi) \neq 0$, а $\cos(\varphi) \neq 1$, где φ — угол поворота. В противном случае итерации прекращаются. Проведённые предварительные тестовые численные эксперименты показали, что при использовании точных данных и известных M и z любая комбинация n уравнений системы (1) даёт точное решение для коэффициентов разложения. В случае зашумления данных измерений без алгоритма фильтрации выбор любых n уравнений не даёт нам известного решения. После выполнения процедуры, когда в верхних n уравнениях учтена вся информация из данных измерений, полученное решение хорошо согласуется с первоначально заданным.

Численный эксперимент

В первом тестовом примере без ограничения общности в качестве данных измерений был взят временной ряд температуры для ограниченной области (33,8–34,1° в.д., 44,25–44,45° с.ш.) в северо-восточной части Чёрного моря за 1982 г. Анализировались как собственно значения температуры поверхности, так и отклонения от среднегодового хода. Осреднение проводилось для каждого дня с использованием информации за 35 лет. Из данных для выбранного года вычитался среднегодовой ход. Из полученного массива отклонений были сгенерированы случайным образом пропуски в измерениях от 1 до 7 сут. Для аппроксимации значений временной функции использовалось 15 мод разложения в ряд Фурье.

Особенностью реализованного алгоритма является необходимость задания априорной информации о решении. Для этой цели используется число M , которое характеризует возможные значения решения системы (1). Понятно, что оно может быть различным для разных периодов времени. Это число зависит и от интервала времени, на котором решается задача (в данном случае это год), и от того, насколько данный период является аномально тёплым или холодным относительно осреднённых величин. Численные эксперименты показали, что наилучшие результаты для выбранного года получаются при $M = 2,5$. На *рис. 1* представлены восстановленные значения отклонений температуры при $M = 2,5$ (жирная сплошная линия), $M = 3$ (пунктирная линия) и $M = 2$ (тонкая сплошная линия). Кружками изображены отклонения температуры для выбранного года от среднемноголетних значений.

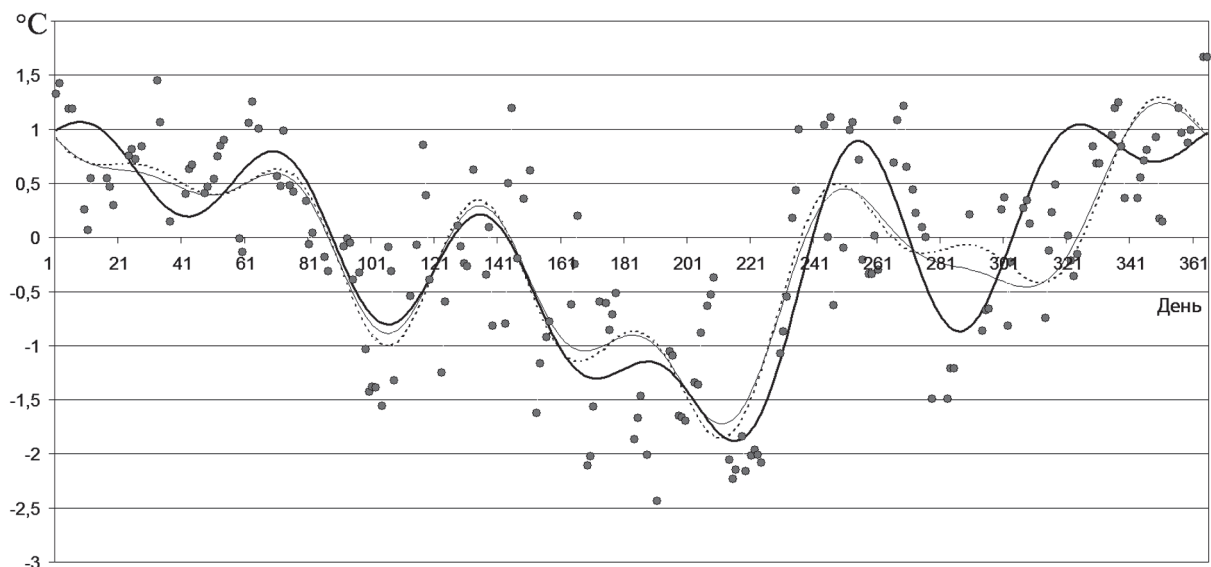


Рис. 1. Данные отклонений и восстановленные функции отклонений температуры: тонкая сплошная линия — $M = 2$, пунктирная — $M = 3$, жирная сплошная — $M = 2,5$

После восстановления функции отклонений на всём интервале времени мы можем, зная среднемесячные значения, вычислить значения температуры для исследуемого года (рис. 2). На этом рисунке жирной сплошной линией изображён результат восстановления, а кружками обозначены значения температуры в точках измерений. При этом среднеквадратичное отклонение от данных измерений $\bar{\sigma}^2 = 0,3631$. На рис. 2 также представлен результат при использовании в качестве данных измерений самой температуры (штриховая линия). В этом случае $\bar{\sigma}^2 = 1,5361$. При численной реализации алгоритма восстановления временной функции по данным температуры выбиралось $M = 20$. Численные эксперименты показали, что выбранное значение M в данном случае является оптимальным. Отметим, что M задаётся не строго, важно, чтобы выполнялось условие (7). На рис. 2 и по соответствующим значениям среднеквадратичных отклонений видно, что в случае использования первого подхода (аппроксимация отклонений) мы получаем результат, который лучше соответствует данным измерений. Это можно объяснить тем, что выбранное количество мод более оптимально используется для описания отклонений температуры без учёта внутригодовой изменчивости.

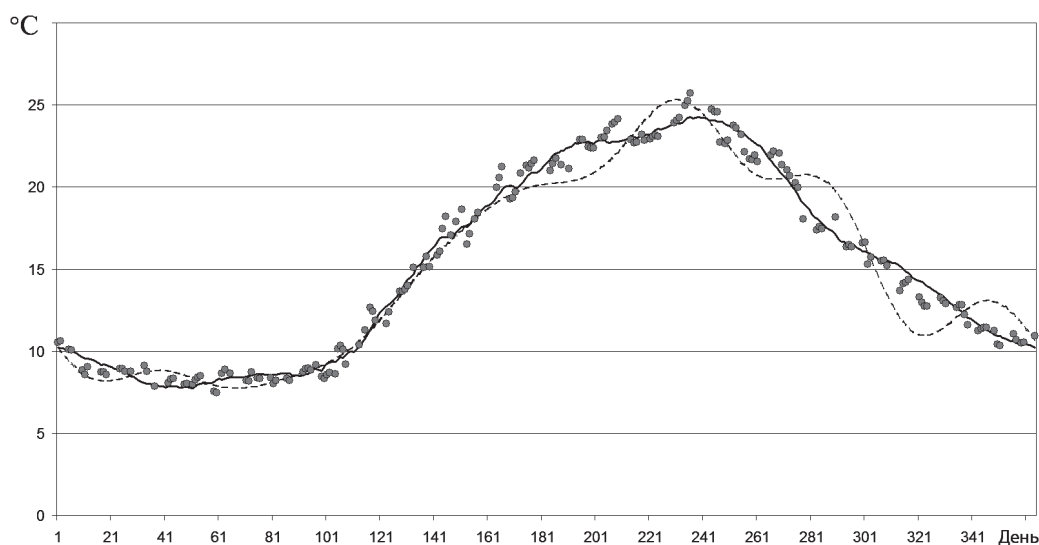


Рис. 2. Данные измерений и восстановленные функции: по отклонениям (жирная линия) и температуре (пунктирная)

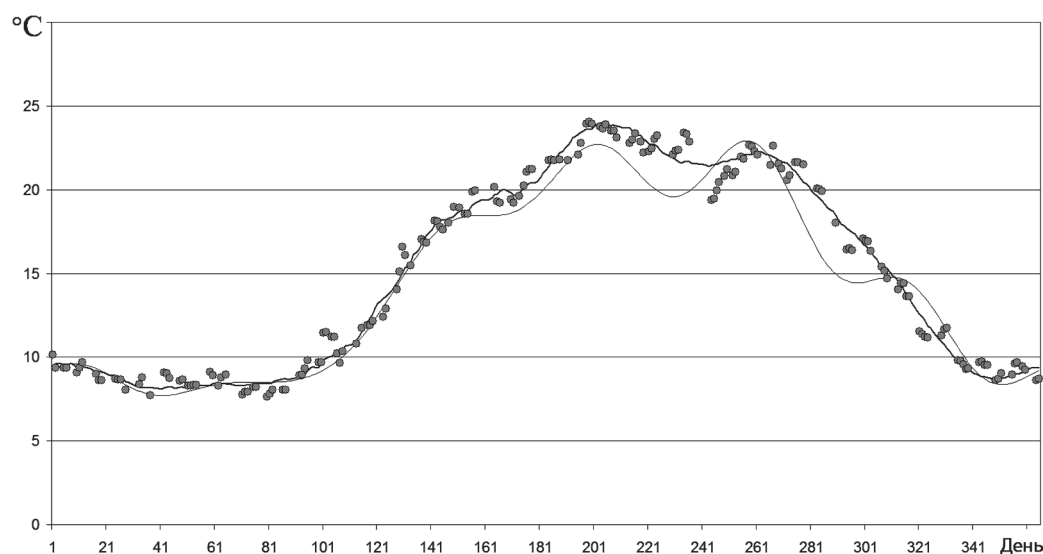


Рис. 3. Данные измерений и восстановленные функции: по отклонениям (жирная линия) и температуре (тонкая сплошная)

Конечно, и в другом случае результат может быть существенно улучшен за счёт увеличения общего количества используемых мод или уменьшения интервала времени, на котором решается задача. При этом общее количество данных измерений должно быть достаточным для осуществления процедуры фильтрации. Аналогичный результат был получен для другого (1984) года из 35-летнего ряда. *Рисунок 3* (см. с. 12) характеризует восстановленную функцию температуры в случае использования данных об отклонениях (жирная сплошная линия) и о самой температуре (тонкая сплошная линия; данные температуры отмечены кружками). Значения среднеквадратичных отклонений получились равными $\bar{\sigma}^2 = 0,4241$ при использовании отклонений и $\bar{\sigma}^2 = 1,8571$ во втором варианте расчёта, использующем сами значения температуры. Следует отметить, что в зависимости от особенностей температурного режима для каждого года величина M будет различной. Тем не менее проведённые численные эксперименты для 10 произвольно выбранных лет показали, что среднеквадратичное отклонение при реализации первого подхода (восстановление функции по отклонениям) от 1,5–5 раз меньше, чем при реализации алгоритма с использованием самих значений температуры.

Заключение

В результате проведённых численных экспериментов показана возможность использования вариационного алгоритма фильтрации для восстановления пропусков в данных измерений. При работе алгоритма фильтрации переопределённой системы уравнений результат учитывает всю информацию, имеющуюся в данных измерений, что приводит к улучшению получаемых оценок. В целом проведённые расчёты показали надёжную работу алгоритма. Применение предложенного метода для отклонений температуры и собственно её значений продемонстрировало более точное восстановление временных рядов по отклонениям температуры: среднеквадратичное отклонение уменьшалось в 4 раза. Выбор интервала времени, на котором решается задача восстановления, зависит от объёма имеющейся информации, а также необходимого количества мод разложения для корректного описания временной функции температуры. Предлагаемый алгоритм может быть использован для восстановления временных рядов с пропусками и более точной оценки средних характеристик.

Работа выполнена в рамках государственного задания по теме 0827-2018-0004 «Комплексные междисциплинарные исследования океанологических процессов, определяющих функционирование и эволюцию экосистем прибрежных зон Чёрного и Азовского морей» (шифр «Прибрежные исследования») и частично поддержана программой Президиума РАН № 1.2.50.

Литература

1. Еремеев В. Н., Кочергин В. П., Кочергин С. В., Скляр С. Н. Математическое моделирование гидродинамики глубоководных бассейнов. Севастополь: Экокси-Гидрофизика, 2002. 238 с.
2. Кочергин С. В., Янковский А. Е. Применение алгоритма фильтрации данных при восстановлении структуры захваченных волн на шельфе Крыма // Морской гидрофиз. журн. 1995. № 5. С. 62–65.
3. Пухтяр Л. Д., Станичный С. В., Тимченко И. Е. Оптимальная интерполяция данных дистанционного зондирования морской поверхности // Морской гидрофиз. журн. 2009. № 4. С. 34–50.
4. Страхов В. Н. Метод фильтрации систем линейных алгебраических уравнений — основа для решения линейных задач гравиметрии и магнитометрии // Доклады АН СССР. 1991. № 3. С. 595–599.
5. Reynolds R. W., Smith T. M., Liu C., Chelton D. B., Casey K. S., Schlax M. G. Daily High-Resolution-Blended Analyses for Sea Surface Temperature // J. Climate. 2007. No. 20. P. 5473–5496.

Using a variational filtering algorithm to fill in surface temperature data gaps

V. S. Kochergin, S. V. Kochergin, S. V. Stanichny

Marine Hydrophysical Institute RAS, Sevastopol 299011, Russia

E-mail: vskocher@gmail.com

To restore the missing surface temperature measurement data, a filtration algorithm is considered. It is based on the variational approach, the search for the extremum of a quadratic functional, the representation of a function that approximates the measurement data as an expansion in terms of orthogonal functions. The procedure is based on a flat rotation of the original system by maximizing the quadratic functional characterizing the ratio of the coefficients of the system and its right-hand sides. In this case, a priori information is used about the possible maximum values of the decomposition coefficients. As a result of the algorithm, an equivalent system is obtained, in which the equations are arranged according to the rank of informativity. Approbation of the method on test examples and real data was carried out. As a source of data, a consistent long-term range of surface temperatures was taken for the same area in the Black Sea for thirty-five years. The measurement data were selected from the existing series randomly. The resulting gaps in the data had a duration from one to seven days. For two randomly selected years, experiments were conducted to restore the missing data. The procedure was implemented in two ways for the temperature values themselves and for deviations from the mean values. The surface temperature data were restored for the time points when the gaps in the measurement data were observed. As a result of numerical experiments, the effectiveness of the variational filtering algorithm is shown in the implementation of the procedure for recovering gaps in surface temperature measurement data.

Keywords: satellite data, surface temperature, variation algorithm

Accepted: 01.11.2018

DOI: 10.21046/2070-7401-2018-15-7-9-14

References

1. Ereemeev V. N., Kochergin V. P., Kochergin S. V., Sklyar S. N., *Matematicheskoe modelirovanie gidrodinamiki glubokovodnykh basseinov* (Mathematical modeling of deep-water basin hydrodynamics), Sevastopol: Ekosi-Gidrofizika, 2002, 238 p.
2. Kochergin S. V., Yankovsky A. E., *Primenenie algoritma fil'tratsii dannykh pri vosstanovlenii struktury zakhvachennykh voln na shel'fe Kryma* (The application of the algorithm filtering data if you restore the structure of the trapped waves on the shelf of the Crimea), *Morskoi gidrofizicheskii zhurnal*, 1995, No. 5, pp. 62–65.
3. Pukhtyar L. D., Stanichny S. V., Timchenko I. E., *Optimal'naya interpolyatsiya dannykh distantsionnogo zondirovaniya morskoi poverkhnosti* (Optimal interpolation of sea surface remote sensing data), *Morskoi gidrofizicheskii zhurnal*, 2009, No. 4, pp. 34–50.
4. Strakhov V. N., *Metod fil'tratsii sistem lineinykh algebraicheskikh uravnenii — osnova dlya resheniya lineinykh zadach gravimetrii i magnitometrii* (The method of filtering systems of linear algebraic equations—the basis for solving linear problems of gravimetry and magnetometry), *Doklady AN SSSR*, 1991, No. 3, pp. 595–599.
5. Reynolds R. W., Smith T. M., Liu C., Chelton D. B., Casey K. S., Schlax M. G., *Daily High-Resolution-Blended Analyses for Sea Surface Temperature*, *J. Climate*, 2007, No. 20, pp. 5473–5496.