

Математическое моделирование морских систем

УДК 551.465.6

Е.Ф. Васечкина

Реконструкция полей температуры поверхности по фрагментарным данным дистанционного зондирования

Предлагается метод заполнения пропусков в данных дистанционного зондирования гидрофизических и биологических характеристик водной поверхности. Метод реконструкции опирается на представление полей поверхностных характеристик в виде суммы некоторого числа эмпирических ортогональных функций, имеющих наибольший вклад в общую дисперсию поля. По отрывочным данным, полученным в результате обработки спутниковых снимков, для летнего сезона года строятся оценки среднего поля и четырехмерной пространственной ковариационной функции температуры поверхности Черного моря. Коэффициенты разложения вычисляются методом наименьших квадратов либо находятся с помощью генетического алгоритма поиска. Результаты численных экспериментов свидетельствуют о перспективности применения метода в задачах заполнения пропусков в спутниковых данных.

Ключевые слова: спутниковые данные, пространственная ковариационная функция, эмпирические ортогональные функции.

Использование информации, получаемой со спутников, открывает широкие возможности для исследования пространственной динамики полей как физических, так и биологических характеристик поверхностного слоя моря. Однако ценность этой информации значительно снижается вследствие присутствия на снимках большого числа пропусков, «белых пятен», в основном обусловленных наличием облачности над морем. Известны различные подходы к решению данной проблемы, однако мы остановимся на методе, основанном на разложении полей по базису эмпирических ортогональных функций (ЭОФ). Этот метод хорошо изучен, он используется в гидрофизических исследованиях с 50-х годов прошлого века [1, 2]. Перспективность его применения в данной задаче связана с высокой скоростью сходимости рядов разложения по ЭОФ и следовательно – с относительно небольшим числом коэффициентов, которые необходимо будет рассчитывать для восстановления поля. Различным аспектам использования метода восстановления данных с применением системы ЭОФ посвящены работы, выполненные в отделе системного анализа МГИ НАН Украины [3 – 6]. Полученные результаты говорят о перспективности применения такого подхода к обработке спутниковой информации. Конечно, в том случае, когда облаками закрыто практически все море, восстановить утраченную информацию невозможно, однако существует довольно большая часть данных, которая после процедуры восстановления могла бы быть использована в задачах мониторинга или модельных расчетах для задания граничных условий.

© Е.Ф. Васечкина, 2011

Методика расчетов

Пусть произвольная реализация наблюдаемого поля $f(\mathbf{x}, t)$ представлена в виде

$$\begin{aligned} f(\mathbf{x}, t) &= [f(\mathbf{x}, t)] + f'(\mathbf{x}, t), \\ f'(\mathbf{x}, t) &= \sum_k a_k(t) \psi_k(\mathbf{x}), \end{aligned} \quad (1)$$

где $[f(\mathbf{x}, t)]$ – математическое ожидание $f(\mathbf{x}, t)$; ψ_k – собственные функции ковариационной матрицы $P(\mathbf{x}, \mathbf{y})$,

$$P(\mathbf{x}, \mathbf{y}) = \sum_t f'(\mathbf{x}, t) f'(\mathbf{y}, t). \quad (2)$$

Коэффициенты разложения $a_k(t)$ рассчитываются с помощью свертки

$$a_k(t) = \int_s f'(\mathbf{x}, t) \psi_k(\mathbf{x}) d\mathbf{x}. \quad (3)$$

Пусть нужно восполнить пропуски в функции $f(\mathbf{x}, t)$, являющейся реализацией некоего процесса или поля. На практике, когда речь идет о полях, построенных по спутниковым данным, реализации, в которых $f(\mathbf{x}, t)$ известна для каждого \mathbf{x} , являются редким исключением. На большинстве снимков присутствуют пропуски, «белые пятна», которые занимают большую площадь. Однако и в этом случае возможно с той или иной ошибкой оценить среднее поле и ковариационную матрицу, используя всю доступную информацию. При этом суммирование проводится во всех точках, где есть данные (здесь и далее речь идет о дискретных реализациях полей на сетке). Например, i -я реализация функции $f_i(\mathbf{x}_j, t)$ определена в точках \mathbf{x}_j ($j = 1, \dots, N$) и содержит пропуски, которые распределены таким образом, что значения функции f известны только в наборе точек K , $K < N$. Пусть в нашем распоряжении имеется набор из T реализаций, каждая из которых также имеет пропуски. Среднее значение функции можно вычислить, суммируя значения в точках, обеспеченных данными в каждой реализации:

$$S(\mathbf{x}_j) = \frac{1}{T_j} \sum_{i=1}^{T_j} f(\mathbf{x}_j, t_i), \quad (4)$$

где T_j – число реализаций, имеющих значение функции в точке \mathbf{x}_j . Оценка ковариационной матрицы будет тогда вычисляться по формуле

$$\begin{aligned} P(\mathbf{x}_j, \mathbf{x}_k) &= \frac{1}{T_{jk}} \sum_{i=1}^{T_{jk}} f'(\mathbf{x}_j, t_i) f'(\mathbf{x}_k, t_i), \\ f'(\mathbf{x}_j, t_i) &= f(\mathbf{x}_j, t_i) - S(\mathbf{x}_j), \end{aligned} \quad (5)$$

где суммирование выполняется только для тех реализаций (числом T_{jk}), в которых точки \mathbf{x}_j и \mathbf{x}_k содержат значения функции. Очевидно, что, поскольку суммирование производится по разному числу пар $f'(\mathbf{x}_j, t) f'(\mathbf{x}_k, t)$, ковариаци-

ционная матрица будет неодинаково обеспечена данными. Это необходимо учитывать при практических вычислениях.

Ансамбль реализаций, используемый для вычисления среднего и ковариационной матрицы, должен подбираться таким образом, чтобы выполнялись хотя бы приблизительно условия однородности и стационарности внешних воздействий. При ограниченном числе данных их выборка представляет собой отдельную проблему. Реализации не должны иметь большого смещения относительно среднего поля, этот фактор оказывается весьма значимым и существенно влияет на качество решения всей задачи. Обычно выборка делается в рамках одного сезона года, однако календарный сезон может не совпадать с климатическим. Например, в настоящей работе мы использовали выборку данных с 15 июня по 15 сентября, считая этот период летним сезоном года.

Построенная таким образом оценка ковариационной матрицы используется затем для расчета собственных функций Ψ и собственных чисел \mathbf{D} , удовлетворяющих уравнению $\mathbf{P}\Psi = \Psi\mathbf{D}$. Теперь для того чтобы получить полную реализацию без пропусков, нужно оценить коэффициенты разложения (3) для некоторого числа первых функций и произвести свертку по формуле (1). Если точек в реализации поля достаточно много, для нахождения коэффициентов можно воспользоваться методом наименьших квадратов, решая систему линейных уравнений

$$f'(\mathbf{x}_k, t_i) = \sum_j a_j(t_i)\psi_j(\mathbf{x}_k), \quad (6)$$

где $k = 1, \dots, K$ – номера точек с данными; $j = 1, \dots, n$ – номера первых n ЭОФ. Применение формулы (3) при наличии пропусков невозможно, поскольку пропущенные данные эквивалентны в этом выражении нулевым значениям $f'(\mathbf{x}, t)$. Число обеспеченных данными точек K в реализации должно быть не менее $10n$.

Если же данных мало, применение формулы (6) некорректно вследствие недостаточности информации для осреднения. В этом случае, как показала практика, хорошо работает генетический алгоритм [7], который отыскивает набор коэффициентов a_k , оптимальным образом удовлетворяющий данным и, кроме того, позволяющий учесть имеющуюся априорную информацию о функции. Подробная информация об этом варианте метода изложена в [4, 5]. Здесь же мы остановимся на нем лишь кратко.

В основе генетического алгоритма лежит идея поиска решения задачи путем имитации его «биологической» эволюции. Алгоритм генерирует и поддерживает популяцию потенциальных решений задачи – индивидуумов. Некоторые из этих решений используются для создания новых решений путем применения к ним специальных операторов, имитирующих процессы реальной биологической эволюции. Это операторы селекции, скрещивания (кроссовера) и мутации. Решения, обладающие лучшим качеством в смысле соответствия введенным критериям, имеют большие шансы «дать потомков» и пройти селекционный отбор в новое поколение. Новые поколения популяции потенциальных решений рассчитываются последовательно до выполне-

ния алгоритмом некоторого условия, обычно связанного с глубиной найденного экстремума функции качества.

Для запуска процедуры поиска нужно определить число функций, для которых требуется найти коэффициенты разложения, интервал изменчивости коэффициентов, необходимую точность их вычисления. Эти параметры задают длину строки так называемого генотипа потенциального решения задачи, за которое мы примем набор коэффициентов разложения. Генотип – это последовательность нулей и единиц, представляющая собой запись искомых коэффициентов в двоичном виде. Генотип разделен на участки, соответствующие отдельным коэффициентам. В каждом из участков определенные позиции отведены для кодировки знака коэффициента, его целой и дробной частей. Интервал возможной изменчивости коэффициентов определяет количество позиций, выделенных для кодировки целой части числа, а точность вычисления – число позиций, в которых будет записана его дробная часть.

Другими управляющими параметрами являются мощность популяции потенциальных решений задачи (индивидуумов), максимальное число генераций, вероятность мутации (замены «0» на «1» в некоторой случайным образом выбранной позиции генотипа). Исследователь также должен задать вид оператора кроссовера, процедуру отбора лучших решений при переходе популяции к следующей генерации, определить условия выхода из алгоритма (окончание поиска), а также условия, в которых популяция может быть элиминирована и заменена на новую, состоящую из случайных строк. Все эти предварительные установки определяют эффективность работы генетического алгоритма, т. е. скорость получения конечного решения задачи.

Особенно важным является задание вида функции качества получаемых решений. К очевидному требованию близости оценки к имеющимся данным в смысле минимума квадратической ошибки можно добавить некоторые априорные сведения о поле, которое необходимо получить. Например, в нашем случае использовалось требование минимума абсолютной величины лапласиана, вычисляемого для центральной части моря, т. е. требование относительной гладкости оценки $\hat{f}(\mathbf{x}, t)$:

$$Q = \frac{1}{L_i L_j} \sum_{i=1}^{L_i} \sum_{j=1}^{L_j} \left| 0,25(\hat{f}_{i+1,j} + \hat{f}_{i-1,j} + \hat{f}_{i,j+1} + \hat{f}_{i,j-1}) - \hat{f}_{i,j} \right| + \frac{1}{K} \sqrt{\sum_{k=1}^K (\hat{f}_k - f_k)^2}, \quad (7)$$

где Q – минимизируемая функция качества потенциальных решений задачи; L_i, L_j – размеры выделенной области расчетной сетки; $\hat{f}_{i,j}$ – значение оценки функции $f(\mathbf{x}, t)$ в узле сетки (i, j) ; \hat{f}_k – значение оценки в k -м узле из множества K , где имеются данные наблюдений.

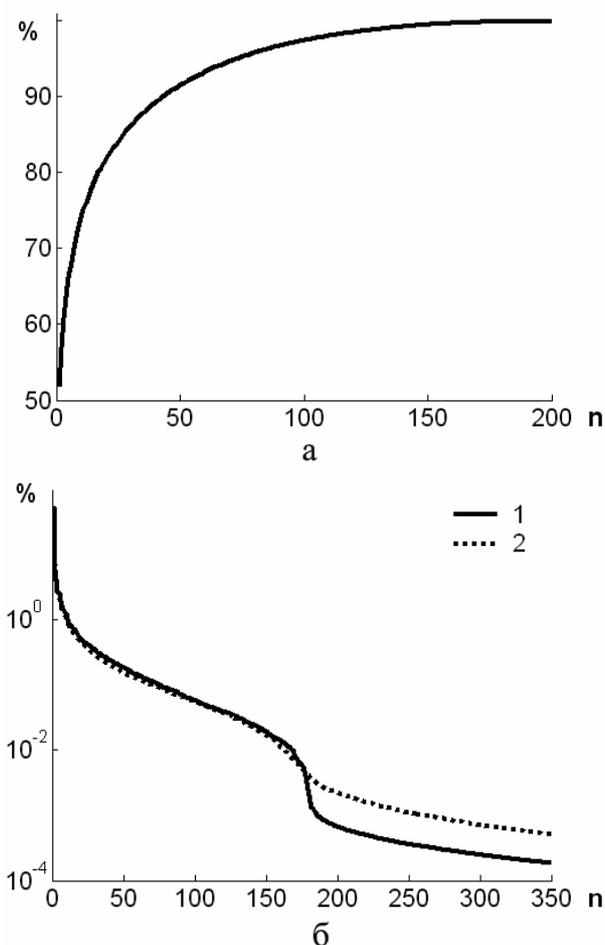
В работе использовалась модификация генетического алгоритма, обладающая следующими особенностями:

– скрещивание двух индивидуумов популяции производилось путем разбиения их генотипов на отдельные участки в двух случайно выбранных точках и последующего обмена этими участками (оператор двухточечного кроссовера);

- индивидум популяции «переходил» в следующую генерацию в том случае, если его функция качества была лучше, чем у трех случайно выбранных из популяции экземпляров;
- лучшие два решения проходили в следующую генерацию вне конкурса;
- популяция элиминировалась и заменялась на новую, если разница между средним и минимальным значениями функции качества в текущей популяции становилась меньше 0,1;
- в начале работы алгоритма искомые коэффициенты рассчитывались методом наименьших квадратов, после чего это решение кодировалось в строку генотипа и становилось «стартовой точкой» для поиска;
- алгоритм заканчивал работу, если лучшее полученное решение не менялось в 40 последовательных генерациях либо число генераций превышало заданный предел;
- параметры, использованные в расчетах: максимальное число генераций – 1000; мощность популяции – 200; длина участка генотипа, кодирующего один коэффициент – 15 бит (целая часть – 4 бита, дробная – 10, знак коэффициента – 1, точность представления коэффициента – 10^{-3}); вероятность мутации отдельного бита в генотипе – $5 \cdot 10^{-4}$.

После получения оценок коэффициентов разложения a_k остается применить формулы (1) для вычисления оценки поля $\hat{f}(\mathbf{x}, t)$. Полученная оценка, естественно, уже не имеет пропусков. Возникает вопрос о количестве функций, которые необходимо использовать для вычисления оценки поля. Этот вопрос решается на основе анализа зависимости величин собственных чисел от номера моды, характеризующего вклад соответствующей моды в общую дисперсию поля. В качестве примера рассмотрим график, полученный в данном исследовании (рис. 1). Как правило, на кривой собственных чисел (рис. 1, б) просматривается отчетливая «ступенька», после которой их значения резко падают. Эта «ступенька» приблизительно определяет необходимое число ЭОФ для оптимального восстановления поля. При меньшем количестве функций ошибка реконструкции поля заметно выше, при увеличении числа функций сверх необходимого минимума ошибка практически не меняется вплоть до момента использования полного набора ЭОФ, когда она обращается в ноль.

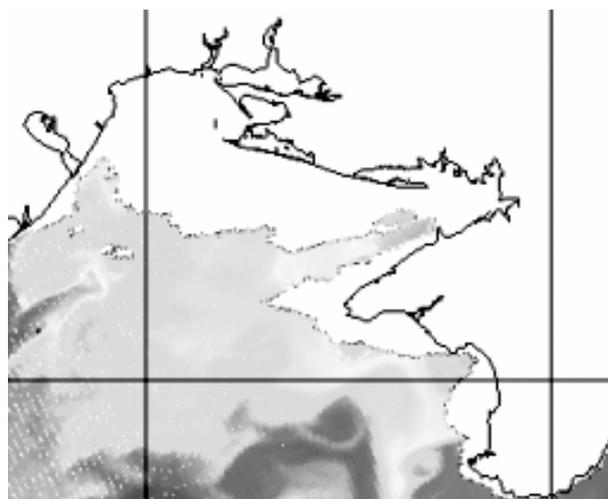
Практическое применение этого метода связано с двумя основными проблемами. Первая заключается в часто выявляемой недостаточности данных для построения ковариационной матрицы, вторая – в сложности ее расчета при большой размерности исследуемого поля. Накопление архивов спутниковых снимков снимет первую проблему в самом ближайшем будущем, вторая проблема связана с объемом памяти и быстродействием компьютеров, находящихся в пользовании исследователя, на сегодня эта проблема представляется более существенной. Рассмотрим вариант ее решения на примере задачи заполнения пропусков в данных дистанционного зондирования поверхности Черного моря [8].



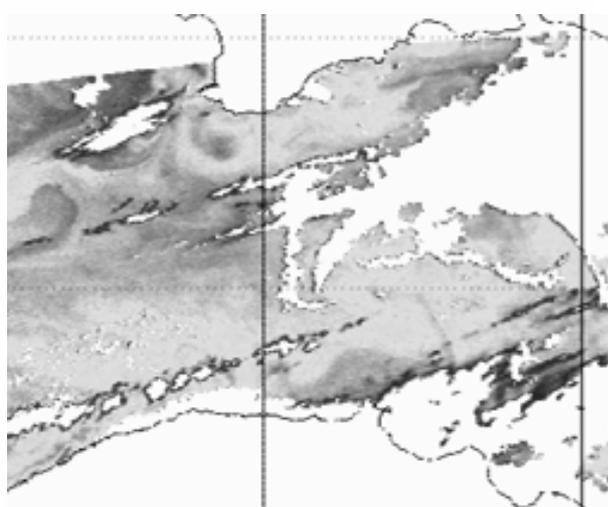
Р и с. 1. Распределение дисперсии изменчивости полей температуры поверхности (летний сезон 2007 – 2009 гг.) по собственным функциям ковариационной матрицы (в %): *a* – суммарная дисперсия; *б* – вклад в общую дисперсию последовательных мод (1 – для западной, 2 – для восточной части Черного моря)

Подготовка данных

При рассмотрении спутниковых снимков, имеющих пропуски вследствие наличия непрозрачных облаков над морем, выясняется, что на некоторых из них присутствуют искажения в данных на краях «белых пятен», вызываемые недостаточно корректным удалением отсчетов в области, скрытой облаками. Например, можно видеть такие снимки, как на рис. 2, где черные точки по краям «белых пятен» (пропусков) после преобразования в градусы Цельсия или единицы концентрации будут давать искаженные значения характеристик поверхности. Эти искаженные данные оказывают существенное влияние на результирующую ковариационную функцию, поэтому их необходимо тщательно удалять с тех снимков, где они имеются.



а

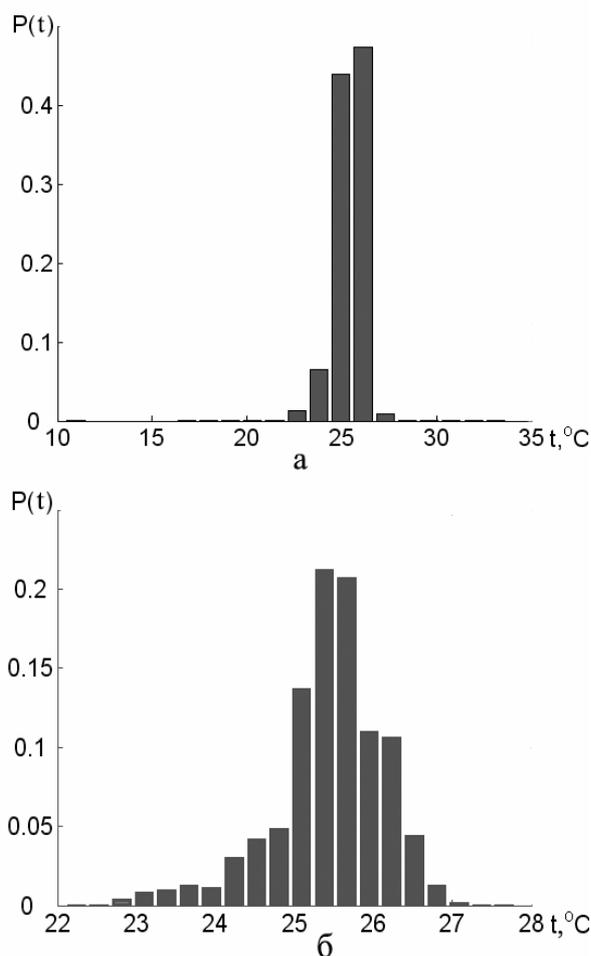


б

Р и с. 2. Примеры спутниковых изображений, содержащих искажения в данных на границе областей, закрытых облаками: *а* – фрагмент поля концентрации хлорофилла «а»; *б* – фрагмент поля температуры поверхности

Для фильтрации искажений была предложена следующая процедура. Для каждого из имеющихся полей строилась гистограмма распределения искомой характеристики. На снимках, содержащих ошибочные значения (как слишком высокие, так и слишком низкие), гистограмма имела длинные хвосты как вправо, так и влево, с малыми значениями частот встречаемости (рис. 3). Значения поля, имеющие низкую вероятность, фильтровались. Если амплитуда значений поля превышала заданную пороговую величину (например, максимальная и минимальная температура поверхности различалась более чем

на 5°C), производилась фильтрация и значения, имеющие частоту, меньшую одной сотой от максимальной частоты встречаемости, заменялись пропусками.



Р и с. 3. Выборочные распределения вероятности для одного из полей температуры поверхности, построенного по спутниковым данным, до (а) и после (б) применения процедуры фильтрации искажений

После такой обработки амплитуда изменчивости поля заметно сокращалась (в примере на рис. 3 с 23 до 6°C, что значительно более реалистично) и распределение приобретало вид, близкий к нормальному. Было установлено, что для полей содержания хлорофилла «а» такую обработку необходимо выполнять до пересчета данных в единицы концентрации, поскольку после пересчета гистограмма приобретает вид экспоненциального распределения и удалить ошибочно низкие и высокие значения уже не представляется возможным.

Таким образом были обработаны данные за летние сезоны 2007 – 2009 гг. по концентрации хлорофилла «а» и температуре поверхности. Полученные

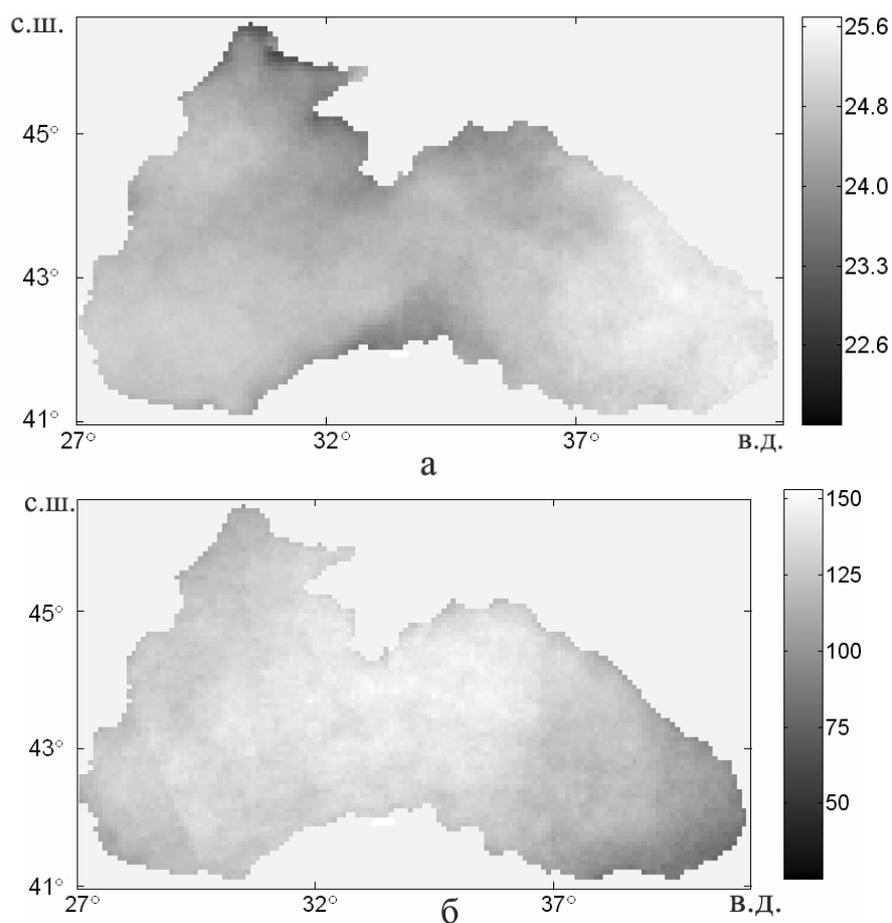
массивы использовались затем для расчета четырехмерной ковариационной матрицы и ее собственных функций. Поскольку принципиальных различий в этих расчетах нет, все дальнейшие выкладки настоящей работы будут касаться только полей температуры поверхности.

Вычисление четырехмерной ковариационной матрицы и ее собственных функций

Разрешение имеющихся в настоящее время данных *MODIS* о температуре поверхности составляет 1 км [8]. Каждый снимок за вычетом границ содержит более 300 000 точек, потенциально несущих полезную информацию. Понятно, что расчет четырехмерной пространственной ковариационной функции всего моря при сохранении разрешения 1 км невозможен, поэтому необходимо осреднять исходные поля, понижая размер ковариационной матрицы. В настоящем исследовании применялось осреднение исходных спутниковых полей с разрешением 1 км на сетке 7×7 км. Кроме того, вся акватория моря была поделена на два участка (западный и восточный), чтобы еще больше сократить размер получающейся ковариационной матрицы. Граница раздела приходилась на центральную часть моря (34° в. д.). Проблема соединения двух участков поля – западного и восточного – здесь специально не рассматривается, однако из опыта известно, что при корректном расчете ковариационных матриц и соответствующих собственных функций она не представляет особых затруднений. Западная часть моря после осреднения была представлена массивом данных 82×65 , в котором N_W узлов приходилось на область моря, восточная часть – массивом данных 62×76 с N_E значимыми узлами.

Рассмотрим теперь практические аспекты выполнения расчетов. Вначале определялся набор переменных, для которого и вычислялась в дальнейшем ковариационная матрица. Каждая переменная представляла собой последовательность значений наблюдаемого поля в некотором узле сетки в области моря $f(\mathbf{x}_k, t)$ ($k = 1 \dots N_W + N_E$). Для построения оценки ковариационной матрицы формировался ансамбль квазинезависимых реализаций поля, полученных обработкой спутниковых снимков в различные моменты времени с дискретностью не менее одних суток. В расчетах, приведенных ниже, ансамблем считался набор реализаций, полученных в летний период с 15 июня по 15 сентября 2007 – 2009 гг. Двумерное поле преобразовывалось в одномерную реализацию путем конкатенации строк массива данных, при этом исключались точки, лежащие за пределами моря. Полученные таким образом строки организовывались в массив $\mathbf{F}(i, k)$ ($i = 1, \dots, T$; $k = 1, \dots, M$), который представлял собой множество квазинезависимых значений переменных $f(\mathbf{x}_k, t)$ в различные моменты времени. Здесь T – мощность статистического ансамбля, или объем выборки двумерных реализаций поля; M – число переменных (для западной части $M = N_W$, для восточной $M = N_E$). Данный массив использовался для расчета среднего поля и ковариационной матрицы. Вычисленная ковариационная матрица имела размер $M \times M$. Каждая k -я строка такой матрицы представляла собой пространственную ковариационную функцию для k -го узла

сетки ($k = 1, \dots, M$). С помощью специальной процедуры такая запись могла преобразовываться в обычное представление в виде поля.

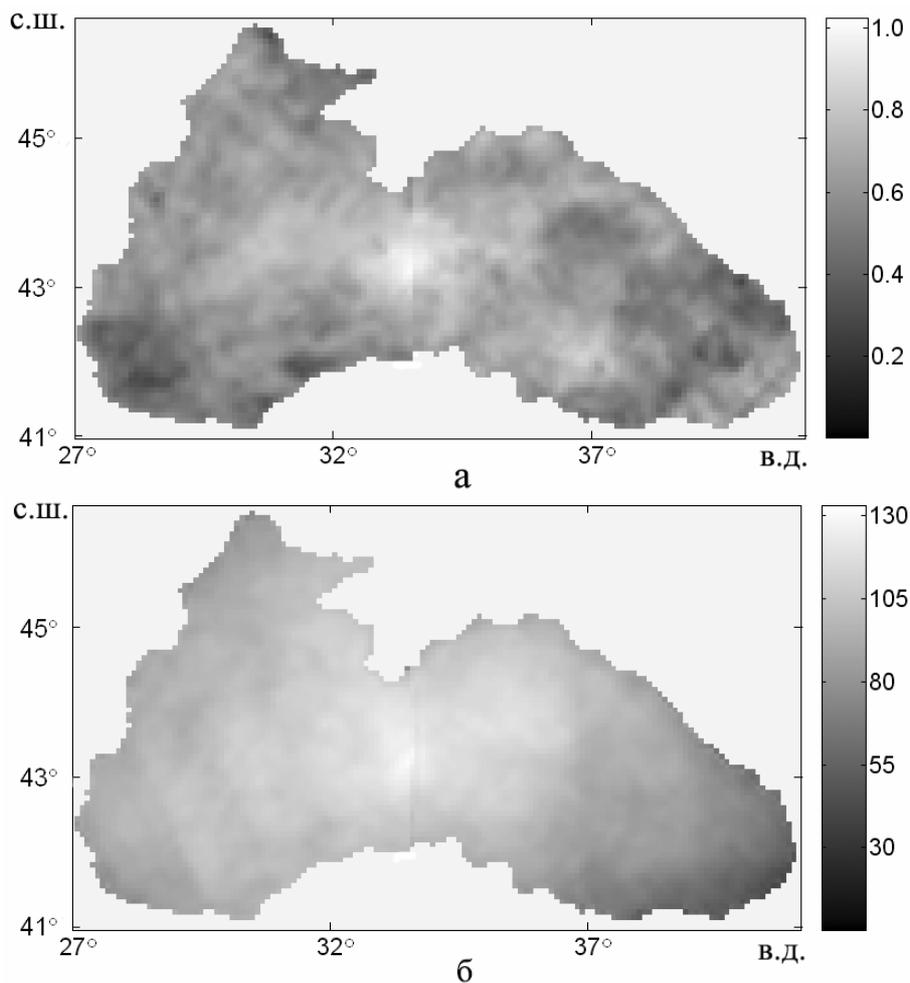


Р и с. 4. Среднее поле температуры поверхности Черного моря ($^{\circ}\text{C}$) на сетке с дискретностью 7 км, оцененное по выборке спутниковых снимков за 2007 – 2009 гг., – а и обеспеченность данными каждого узла этого поля – б

В процессе вычислений оценивалась обеспеченность данными той или иной статистической характеристики. На рис. 4 показано среднее поле температуры поверхности Черного моря в летний период и его обеспеченность данными. Хуже всего обстояло дело с данными на границе море – суша, там число отсчетов не превышало 15 – 20, поэтому приграничные точки были удалены из массивов. Плохо обеспечена данными юго-восточная часть моря, и это сказывается на репрезентативности ковариационной матрицы и в конечном счете – на качестве восстановления информации. Западная часть моря лучше обеспечена данными, и качество реконструкции полей температуры там выше. Очевидно, что количество данных, которым обеспечивается ковариационная функция для какого-либо узла сетки с координатами x_i, y_i , будет

меньше, чем при расчете среднего поля, ведь для того чтобы какая-либо реализация дала вклад в расчет коэффициента взаимной корреляции, нужно наличие в ней данных для пары узлов (x_i, y_i и x_j, y_j), тогда как при расчете среднего достаточно данных только в точке x_i, y_i .

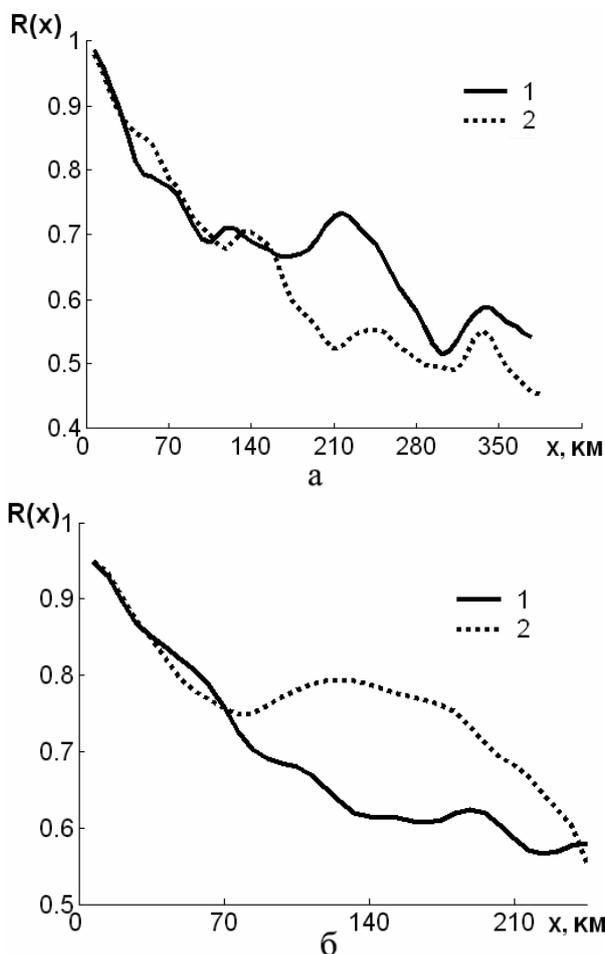
На рис. 5, *а* показана ковариационная функция для центральной области моря, где хорошо видна ее существенная анизотропия. Рис. 5, *б* иллюстрирует распределение количества данных, использованных для получения этой функции. Видно, что юго-восточная область обеспечена данными относительно хуже, чем западная, поэтому с уверенностью можно говорить об оценке ковариации только в центральной части моря. Подобные карты можно построить для любого узла сетки западной и восточной частей акватории.



Р и с. 5. Ковариационная функция температуры поверхности Черного моря, рассчитанная для центральной области по данным статистического ансамбля за июнь – сентябрь 2007 – 2009 гг., (осреднена по 12 центральным узлам сетки) – *а* и количество данных, использованных для расчета коэффициентов взаимной корреляции в каждом из узлов сетки, – *б*

Существенная анизотропия поля корреляции видна на рис. 6, где показаны профили ковариационной функции в широтном и меридиональном

направлениях. Широтные профили рассчитаны по данным, представленным на рис. 5, – от центральной области на запад и восток. Меридиональные разрезы построены по полю ковариационной функции для центральной области западной части моря в направлениях на север и юг.



Р и с. 6. Одномерные корреляционные функции температуры поверхности: *a* – широтный разрез по 43° с. ш. (от центральной области на запад – 1, на восток – 2); *б* – меридиональный разрез по 31° в. д. (от центральной области на юг – 1, на север – 2)

Численные эксперименты по восстановлению полей по фрагментарной информации

Следующий этап после вычисления четырехмерной ковариационной матрицы – расчет ее собственных функций, т. е. ЭОФ. Их относительный вклад в общую дисперсию показан на рис. 1. Первая мода описывает 52% общей дисперсии, вторая – 5%, третья – 4%. Более высокие моды вносят экспоненциально уменьшающийся вклад в общую дисперсию, однако для достижения оптимального результата в расчетах по восстановлению полей необходимо использовать в данном случае 150 – 160 функций («ступенька» на

графике зависимости собственных чисел от номера моды). Меньшая обеспеченность данными восточной части акватории отразилась на графике рис. 1, б – «ступенька» кривой собственных чисел для ковариационной матрицы восточного массива данных имеет меньший перепад, и остаточная дисперсия выше, чем для западного массива. На практике такие особенности приводят к более низкому качеству восстановления информации. Улучшить качество вычислений можно путем увеличения количества данных в исходном массиве.

Рассмотрим теперь конкретные результаты по заполнению пропусков. Эксперименты выполнялись следующим образом. Для обеспечения условий контролируемой точности реконструкции было отобрано несколько реализаций полей, имеющих небольшое количество пропусков, а также обладающих заметными особенностями пространственного распределения температуры. Они были обозначены как контрольные, удалены из массива данных и не учитывались при вычислении ковариационной матрицы. Параллельно отбирались реализации, имеющие различное число пропусков, – от 10 до 75%. Паттерн «белых пятен», обусловленных пропусками (маска облачности), с одной из этих реализаций накладывался на контрольную реализацию без пропусков, после чего поле подавалось на вход алгоритма реконструкции, описанного выше.

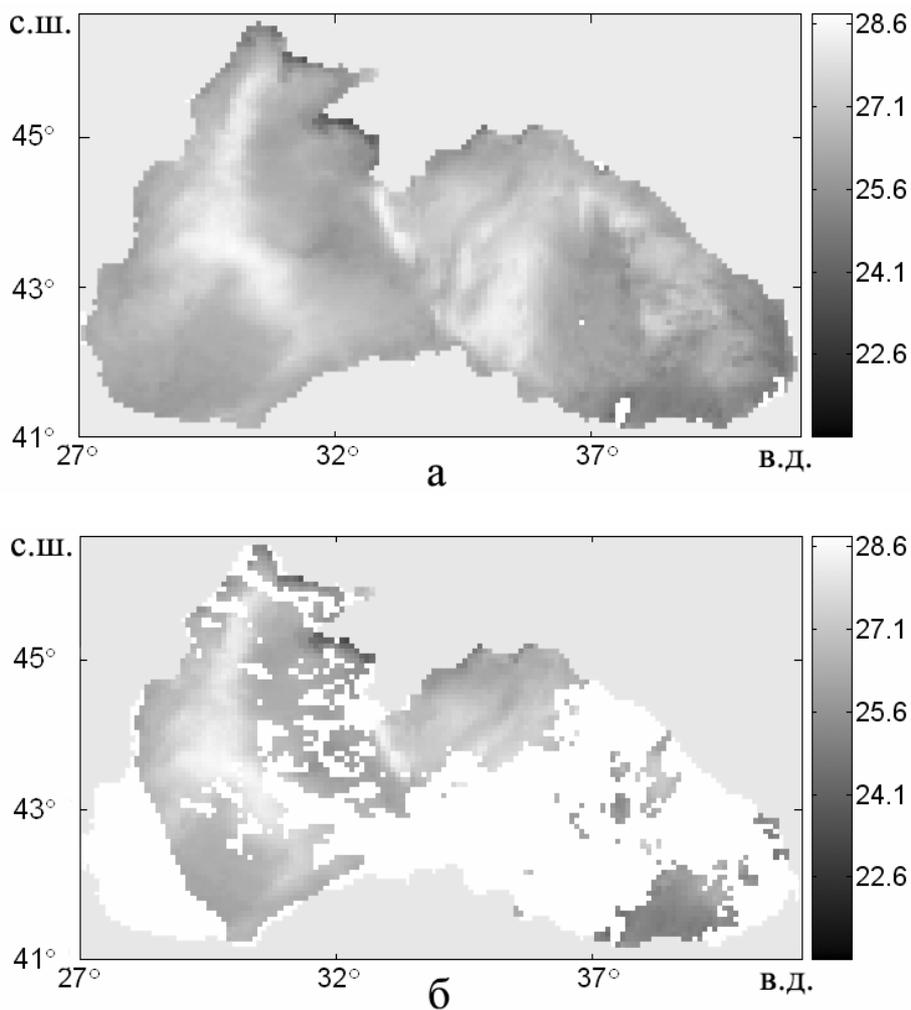
Последним этапом реконструкции являлась фильтрация полученных полей с помощью дискретного двумерного вейвлет-преобразования. Эта процедура удаления шумовой компоненты оказалась необходимой вследствие неизбежных ошибок вычислений, связанных с большим количеством пропусков в исходных полях температуры. Использовался двумерный вейвлет *db2* (*Daubechies*). Восстановленное поле температуры $\hat{f}(x, t)$ раскладывалось на аппроксимацию и три детализации – горизонтальную, вертикальную и диагональную (первый уровень разложения). Далее детализации подвергались фильтрации, в результате чего удалялись мелкомасштабные флуктуации. Отфильтрованный шум имел нормальное распределение с нулевым средним. Затем производилось обратное преобразование – реконструкция поля с использованием аппроксимации и модифицированных детализаций.

Поле, полученное на выходе процедуры, сравнивалось затем с исходным полем без пропусков, и вычислялись абсолютная и относительная ошибки. Относительная ошибка рассчитывалась как отношение среднеквадратической ошибки к среднеквадратическому отклонению случайной составляющей поля. На рис. 7 показан пример реконструкции поля температуры от 3 августа 2009 г. «Наблюдаемое» поле в данном случае имеет 55% пропусков и соответствует реальной ситуации наблюдения Черного моря со спутника от 14 сентября 2007 г.

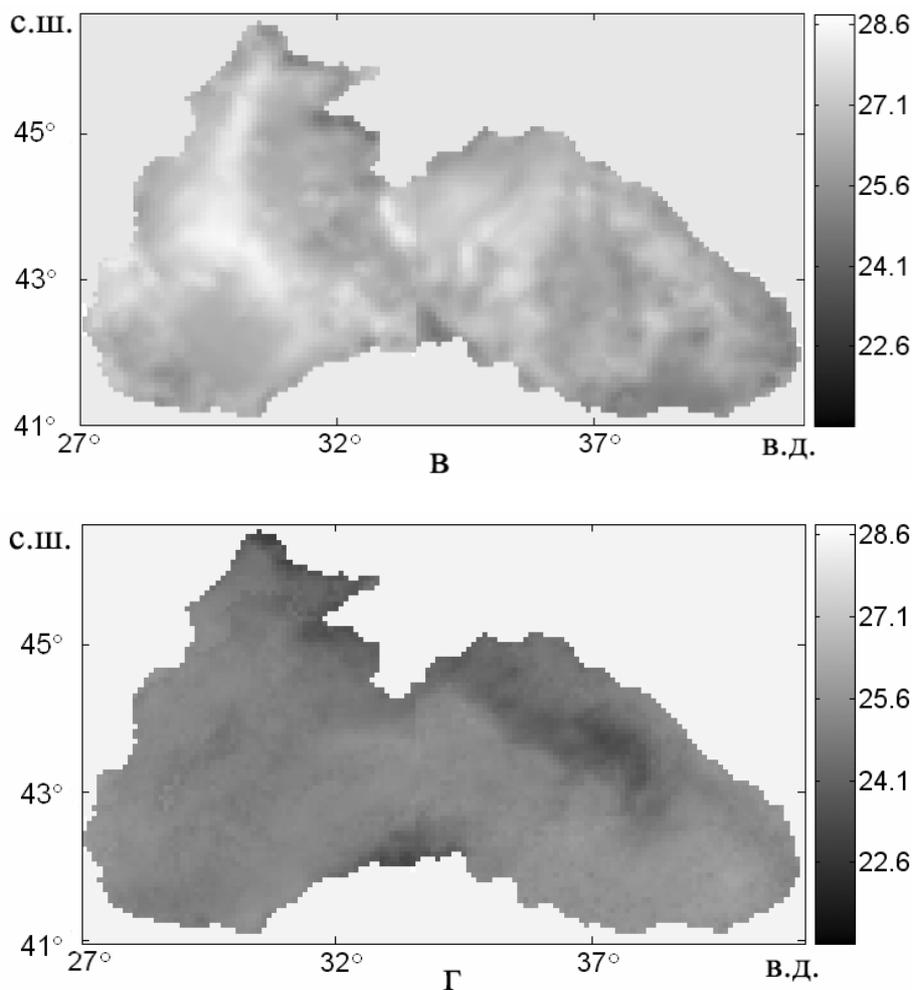
Относительная ошибка реконструкции составляет 0,22 среднеквадратического отклонения реального поля для его западной части и 0,28 – для восточной. Как видно из рис. 7, метод позволяет восстановить особенности пространственного распределения температуры, частично или полностью отсутствовавшие в «наблюдаемом» поле. Восточную часть моря удалось восстановить несколько хуже, чем западную. По всей видимости, это объясняется тем, что ковариационная матрица для восточного массива была оценена с большей погрешностью. Можно надеяться, что при увеличении мощности стати-

стического ансамбля, используемого для оценки ковариационной функции, ошибки реконструкции удастся уменьшить.

Восстановленное поле можно сравнить со средним полем поверхностной температуры для августа 2009 г. (рис. 7, *г*). Видно, что абсолютные значения среднего поля ниже, а характер пространственного распределения аномалий на рис. 7, *а, г* существенно различается. Холодное пятно (рис. 7, *а, в*) в юго-восточной части моря сместилось к северу, теплого пятна в западной части нет, и общий рисунок аномалий температуры заметно сглажен. Ошибка реконструкции относительно среднего поля за август 2009 г. составляет 0,24, абсолютная 0,48°C. Очевидно, что заполнение «белых пятен» в наблюдаемом поле среднемесячными значениями приведет к грубым искажениям рисунка пространственного распределения температуры.



Р и с. 7. Пример реконструкции поля температуры поверхности (°C) от 3 августа 2009 г.: *а* – реальное; *б* – то же поле после наложения маски облачности с 55% пропусков (имитация наблюдаемого поля); *в* – восстановленное; *г* – среднее за август 2009 г.



Р и с. 7. Окончание

Аналогичные эксперименты были проведены со всеми реализациями температуры поверхности, не имеющими пропусков. Маски облачности, которые использовались для имитации наблюдаемых полей, имели 10 – 75% пропусков. Абсолютные ошибки реконструкции поля находились в пределах 0,1 – 0,7°C, относительные – в пределах 0,06 – 0,60 стандартного отклонения аномалий. Погрешность зависела не только от количества пропусков, но и от характера их распределения в пространстве. На большинстве снимков пропуски были сплошными, т. е. часть акватории моря оказывалась полностью вне зоны наблюдения. В этом случае возникает задача экстраполяции, а не интерполяции. Если пропуски были рассредоточены по всему полю небольшими островками, т. е. поле в принципе можно было проинтерполировать, ошибка всегда была невелика. Однако таких снимков оказалось совсем немного, не более 15% от общего числа обработанных полей. В среднем зави-

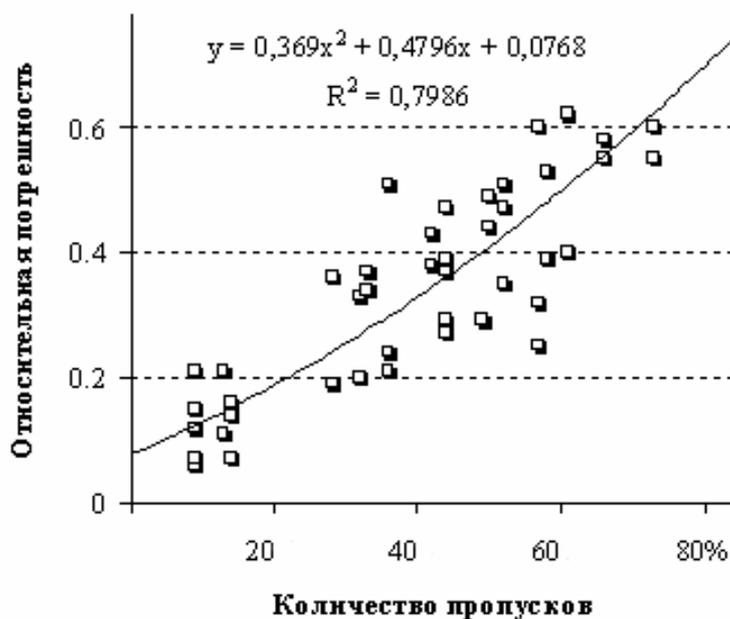
симось относительной ошибки реконструкции от количества пропусков в исходном поле можно аппроксимировать формулой

$$\varepsilon = 0,3690B^2 + 0,4796B + 0,0768,$$

где B – относительное число пропусков в наблюдаемом поле (рис. 8).

Если принять за число степеней свободы количество функций в формулах (1), использованных для восстановления конкретного поля, можно оценить доверительный интервал с заданной вероятностью. Например, если считать, что средняя ошибка реконструкции при использовании 160 ЭОФ составляет $0,3^\circ\text{C}$, то истинное поле с вероятностью 95% находится в интервале ($\hat{f}(\mathbf{x},t) - 0,5^\circ\text{C}$; $\hat{f}(\mathbf{x},t) + 0,5^\circ\text{C}$).

В работе [9] рассматривалась задача оптимальной интерполяции данных дистанционного зондирования температуры поверхности Черного моря. Приведена средняя ошибка интерполяции данных наблюдений при переходе от сетки с шагом 10 км на сетку с шагом 1 км, она равна $0,29^\circ\text{C}$. Средняя ошибка, полученная нами, составляет $0,38^\circ\text{C}$. Учитывая, что в нашем случае решается задача экстраполяции полей, в которых пропуски занимают до 70% площади, можно признать эту ошибку вполне удовлетворительной. Если же рассматривать только поля, где пропуски занимают не более 50% площади, то средняя ошибка реконструкции сократится до $0,32^\circ\text{C}$. Зависимость относительной погрешности от количества пропусков в исходных данных представлена на рис. 8.



Р и с. 8. Относительные ошибки реконструкции температуры поверхности (в долях среднеквадратического отклонения случайной составляющей) в зависимости от количества пропусков в исходных данных

Заключение

Предлагаемый метод восстановления информации на основе четырехмерной ковариационной матрицы наблюдаемых значений поля обеспечивает реконструкцию особенностей реального поля на качественном уровне. Что касается количественной оценки, то ошибка метода, как правило, не превышает одной трети стандартного отклонения характеристики от ее среднего распределения. При добавлении новых реализаций к ансамблю вид одно- и двумерных сечений ковариационной матрицы качественно не изменяется, т. е. построенная оценка достаточно устойчива. Предлагаемый метод может быть применен для реконструкции не только полей поверхностной температуры, но и других характеристик, получаемых по спутниковым данным.

СПИСОК ЛИТЕРАТУРЫ

1. *Everson R., Cornillon P., Sirovich L. et al.* An empirical eigenfunction analysis of sea surface temperatures in the Western North Atlantic // *J. Phys. Oceanogr.* – 1997. – 27, № 3. – P. 468 – 479.
2. *Репинская Р.П., Бабич Я.Б.* Аппроксимация рядами эмпирических ортогональных функций северополушарных полей облачности по спутниковым данным // *Исследования Земли из космоса.* – 1999. – № 6. – С. 8 – 15.
3. *Васечкина Е.Ф., Тимченко И.Е., Ярин В.Д.* Восстановление структуры гидрофизических полей по неполной информации // *Морской гидрофизический журнал.* – 1997. – № 2. – С. 37 – 44.
4. *Васечкина Е.Ф., Ярин В.Д.* Генетический алгоритм в задаче реконструкции гидрометеорологических полей // *Системы контроля окружающей среды.* – Севастополь: ЭКОСИ-Гидрофизика, 2002. – С. 141 – 145.
5. *Васечкина Е.Ф., Ярин В.Д.* Использование генетического алгоритма в задаче восстановления пропущенных данных // *Морской гидрофизический журнал.* – 2002. – № 4. – С. 30 – 39.
6. *Yarin V.D., Vasechkina Y.F.* Missing data recovery by using decomposition into empirical orthogonal functions // *Proc. of Intern. Conf. “Scientific and policy challengers towards an effective management of the marine environment”.* – Albena, Bulgaria, 2003. – P. 401 – 403.
7. *Whitley D.* A genetic algorithm tutorial // *Tech. Rep. CS-93-103.* – Colorado State University, 1993. – 38 p.
8. *Морской портал НКАУ – МГИ НАН Украины.* – <http://dvs.net.ua>.
9. *Пухтяр Л.Д., Станичный С.В., Тимченко И.Е.* Оптимальная интерполяция данных дистанционного зондирования морской поверхности // *Морской гидрофизический журнал.* – 2009. – № 4. – С. 34 – 50.

Морской гидрофизический институт НАН Украины,
Севастополь
E-mail: cpdi@rambler.ru

Материал поступил
в редакцию 19.01.10
После доработки 10.03.10

АНОТАЦІЯ Пропонується метод заповнення пропусків у даних дистанційного зондування гідрофізичних і біологічних характеристик водної поверхні. Метод реконструкції ґрунтується на представленні полів поверхневих характеристик у вигляді суми деякого числа емпіричних ортогональних функцій, які дають найбільший внесок в загальну дисперсію поля. За уривчастими даними, отриманими внаслідок обробки супутникових знімків, для літнього сезону року будуються оцінки середнього поля і чотиривимірної просторової коваріаційної функції температури поверхні Чорного моря. Коефіцієнти розкладання обчислюються методом найменших квадратів або знаходяться за допомогою генетичного алгоритму пошуку. Результати чисельних експериментів свідчать про перспективність застосування методу в задачах заповнення пропусків у супутникових даних.

Ключові слова: супутникові дані, просторова коваріаційна функція, емпіричні ортогональні функції.

ABSTRACT The method of gap-filling in remote sensing data on hydrophysical and biological characteristics of water surface is proposed. The method of reconstruction is based on presentation of fields of surface characteristics as a sum of a number of empirical orthogonal functions making the largest contribution to the field general dispersion. Based on fragmentary data resulted from processing of satellite images, estimated are the average field and four-dimensional spatial covariance matrix of the Black Sea surface temperature for a summer season. Decomposition coefficients are calculated by the least-squares method or obtained by the genetic algorithm. The results of numerical experiments testify to perspective application of the method in the tasks of gap-filling in satellite data.

Keywords: satellite data, spatial covariance matrix, empirical orthogonal functions.