

УДК 519.688

МЕТОДИКА КРАТКОСРОЧНОГО ПРОГНОЗА ИЗМЕНЕНИЯ ЗНАЧЕНИЙ ХАРАКТЕРИСТИК ДИНАМИЧЕСКОЙ СИСТЕМЫ НА ОСНОВЕ ЧИСЛЕННОГО АНАЛИЗА МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

*Запорожцев И.Ф., аспирант Мурманского государственного технического университета,
e-mail: zaporozhtsev.if@gmail.com;*

Серёда А.-В.И., д.т.н., e-mail: avis_14@mail.ru.

Ключевые слова: краткосрочный прогноз, пространственное распределение физической характеристики, многомерный временной ряд, сингулярный спектральный анализ, декомпозиция на эмпирические моды, модель TARX.

Введение

Задача разработки и совершенствования математических моделей и численных методов для анализа многомерных временных рядов, отражающих изменение во времени значений целевой характеристики для различных точек некоторого физического пространства, является на данный момент перспективным направлением исследований для специалистов из различных отраслей. Разработано большое количество теоретических и узко специализированных эмпирических моделей, допущений и эвристик, которые справедливы для конкретной прикладной задачи и предметной области. Тем не менее, ни одно из существующих решений не позволяет выстроить системный адаптивный подход к прогнозированию временной изменчивости пространственного распределения значений произвольной характеристики без учёта специфики анализируемых данных, обусловленной их природой. В данной статье делается определённый шаг в этом направлении: предлагается как общая концепция, так и детальная реализация авторской методики прогноза при работе с гридированными океанографическими данными. Для определения места этой методики среди имеющихся работ в первом разделе статьи приведена краткая характеристика некоторых групп моделей, положенных в основу современных концепций прогнозирования значений временных рядов. Затем (раздел 2) следует краткое описание многомерного ряда, с которым идёт работа. В разделе 3 представлены сведения о начальном внутрикластерном прогнозе, который является первым этапом предлагаемой методики. В разделе 4 вводятся так называемые кластерные характеристики, а в разделе 5 рассмотрена процедура исправления значений внутрикластерного прогноза. Общая схема методики прогноза приведена в разделе 6. Завершается статья анализом результатов вычислительных экспериментов, имевших целью апробацию методики на модельных данных (раздел 7).

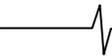
Представлена методика краткосрочного прогноза временной изменчивости пространственно распределённых характеристик природных динамических систем, основанная на анализе многомерных временных рядов. Предполагается, что значения характеристик заданы в узлах планарной равномерной сетки. Сформулирована как общая концепция этой методики, так и алгоритмическая детализация некоторых её этапов. На данный момент предлагаемая версия методики полностью сформирована и реализована в виде компьютерной технологии. Приводятся результаты численных экспериментов для модельных данных, а также их анализ.

1. Особенности проблемы классификации моделей прогноза

Стоит отметить, что классификация моделей и методов прогнозирования, как и классификация математических моделей, является в настоящее время не вполне устоявшейся. Некоторые критерии выделения и названия классификационных групп, предложенные отдельными авторами в последнее десятилетие и используемые нами в данной статье, не следует рассматривать как общепринятые.

В основу теоретических моделей анализа временной изменчивости характеристик природных динамических систем положены теории и методы решения задач математической физики. С развитием этого подхода стали появляться предметно-ориентированные модели, представляющие пространственную и временную изменчивость дифференциальными уравнениями, отражающими фундаментальные законы. Примеры подобных моделей, например, для динамики уровня поверхности моря можно найти в работах [1-3]. Эти модели могут быть достаточно легко адаптированы для применения в самых разных предметных областях. Одна из существенных объективных проблем их практической реализации вызвана отсутствием некоторых необходимых данных для расчётов или их недостаточной точностью. Также известной проблемой являются большие затраты временных ресурсов численной реализации этих моделей, которые не находят оправдание с точки зрения точности получаемых значений.

В основе другого подхода к прогнозированию лежат так называемые физико-статистические методы, по сути реализующие модели множественной линейной регрес-



сии по времени [2]. Для физико-статистических методов ставят целью получение статистических закономерностей изменения целевого параметра в зависимости от значений модельных параметров, имеющих чёткий физический смысл.

Аппроксимация результатов пассивных наблюдений и измерений, которая осуществляется подбором подходящего уравнения, включающего то или иное число независимых переменных, не несущих в себе никакой информации с точки зрения физики конкретных протекающих в реальной системе процессов, стала основой эмпирических моделей. Они являются итогом обработки результатов наблюдения изменения количественных значений наблюдаемых характеристик с целью обобщения результатов экспериментальных исследований для получения некоторой качественной информации и выработки управляющих решений. Наиболее распространённые эмпирические модели являются статистическими. Классическим примером для временных рядов является модель стационарных рядов ARIMA.

В настоящее время среди множества эмпирических регрессионных моделей выделяется классификационная единица, а именно так называемые структурные модели [4], не требующие выполнения условия стационарности. Они обычно используются в задачах декомпозиции временного ряда на аддитивные составляющие и прогноза его значений. Эти модели стали результатом усложнения процесса построения аналитических зависимостей, но по-прежнему отображают случайный (стохастический) характер некоторых параметров в представлении анализируемого временного ряда. Некоторые из этих моделей можно считать статистическими, например, модель эволюции выборочной плотности функции распределения [5]. Другие модели ориентированы на исключение случайных составляющих посредством задания количественных значений некоторых параметров без необходимости содержательной интерпретации этих значений, при условии, что модельные данные практически соответствуют реальным. Суть этого задания сводится к априорной предустановке значения отношения «сигнал-шум», как, например, в рамках сингулярного спектрального анализа [9-10]. Наконец, к структурным моделям можно отнести и нейронные сети, в частности, многослойный перцептрон, который вовсе не различает в обработке зашумлённые данные и данные, содержащие только полезный сигнал. Общей чертой всех структурных моделей является их алгоритмичность: необходимость выполнения ряда шагов, не все из которых хорошо формализованы. Неотъемлемой частью представления моделей подобного типа являются блок-схемы, структурные схемы, диаграммы состояний и т.д.

2. Многомерный временной ряд

Методика прогноза, предлагаемая нами, относится к классу структурных эмпирических моделей и основана на анализе многомерных временных рядов, каждый одномерный компонент которого поставлен в соответствие некоторому узлу регулярной сетки UG (uniform grid), заданной в прямоугольной области Ω . Предполагается, что одномерные временные ряды для каждого из узлов

сформированы единообразно: совпадают стартовый временной отсчёт, шаг по времени между измерениями и анализируемая длина временного ряда. Такой подход к формированию рядов позволяет говорить, что в области Ω задан многомерный временной ряд $F_N^{UG}(t) = (F_N^1(t), F_N^2(t), \dots, F_N^s(t))$ длины N , где s – число одномерных рядов. Каждый из одномерных рядов может быть также проиндексирован двумя координатами соответствующего ему узла сетки UG. На рис. 1 изображения узлов дублируются для всех моментов времени, что означает необходимость задания значений характеристики в узлах в эти моменты времени.

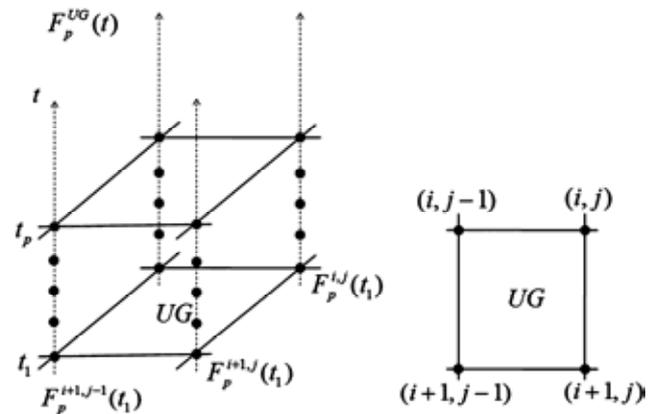


Рис. 1. Многомерный временной ряд

Основным ограничением в рамках формируемой методики является доступность для анализа только одного многомерного временного ряда (данных о значениях одной характеристики), что позволит использовать такую методику для прогноза различных характеристик, не учитывая природу данных.

3. Начальный внутрикластерный прогноз

Первым этапом анализа многомерного ряда в задаче прогноза является выделение кластеров узлов сетки UG; в дальнейшем временные ряды (кластерные ряды), которые им соответствуют, будут обрабатываться в отдельности. Разработанный авторами алгоритм кластеризации, представленный в статье [6], учитывает как взаимное расположение узлов, так и специфику временной изменчивости соответствующих им временных рядов.

Основные идеи, а также алгоритм начального внутрикластерного прогноза описаны детально в статьях [7, 8]. Здесь следует лишь упомянуть, что в основе алгоритма лежит модель и метод К-продолжения многомерного временного ряда, использующий результаты сингулярного спектрального анализа (МССА) [9, 10], а также (вариативно) многомерный вариант декомпозиции на эмпирические моды (МДЭМ) [11, 12], авторский опыт использования которого обобщён в статье [8]. Начальный внутрикластерный прогноз был протестирован, как на некоторых модельных данных, так и апробирован на экспериментальных данных геофизической природы: аномалиях уровня поверхности моря [7, 8]. В качестве новизны подхода следует отметить совместное использование математических моделей, которые лежат в основе современных технологий обработки временных рядов: МССА, МДЭМ и методики подбора оптимального

с точки зрения предустановленной границы ошибки прогноза длины фрагмента временного ряда, значения которого непосредственно участвуют в прогнозе (эта методика, основанная на выделении квазистационарных фрагментов ряда, изложена в книге [5]). Авторами разработан и реализован программно адаптированный алгоритм, который позволяет выстроить взаимодействие процедур, соответствующим трём упомянутым технологиям.

Предположим, что горизонт прогноза (число последовательных временных отсчётов, для которых необходимо построить прогноз значений характеристики) известен и равен τ . О точности прогноза значений характеристики по фрагменту кластерного многомерного ряда $F_N^C = (F_N^1, F_N^2, \dots, F_N^{s_c})$ длины T , для которого последним временным отсчётом является $(t-1)$, на горизонт τ будем судить по нормированной среднеквадратичной ошибке $\delta_{i,T,\tau}^C$ (в %):

$$\delta_{i,T,\tau}^C = \frac{100}{d(t)} \cdot \max_{i=1, \dots, s_c} \sqrt{\frac{\sum_{j=0}^{\tau-1} (F_N^i(t+j) - \tilde{F}_\tau^i(t+j))^2}{\tau}}, \quad (1)$$

где $\tilde{F}_\tau^C = (\tilde{F}_\tau^1, \tilde{F}_\tau^2, \dots, \tilde{F}_\tau^{s_c})$ – соответствующий многомерный ряд прогнозных значений длины τ (его значения соответствуют временным отсчётам $(t+j)$, $j = 0, \dots, \tau-1$), вычисленных по фрагменту ряда F_N^C длины T , а нормирующий множитель определяется разностью $d(t)$:

$$d(t) = \max_{i=1, \dots, s_c} \left\{ \max_{t' < t} F_N^i(t') - \min_{t' < t} F_N^i(t') \right\}. \quad (2)$$

В следующем разделе вводятся так называемые кластерные характеристики, которые будут участвовать в процедуре коррекции начального прогноза, формируя таким образом финальный внутрикластерный прогноз.

4. Кластерные характеристики

На множестве кластеров, а также на множестве упорядоченных пар кластеров зададим несколько функционалов, которые будем называть кластерными характеристиками. Их значения будут *использованы* на этапе коррекции результатов начального внутрикластерного прогноза.

Коэффициент кросс-корреляции. Коэффициент кросс-корреляции можно рассматривать как кластерную характеристику двумя способами: выбрать наименьшее его значение, вычисленное для всех *возможных* пар одномерных временных рядов одного кластера при нулевом лаге, или ввести для пары кластеров, определяя корреляционную связь для временных рядов различных кластеров.

Отношение сингулярных чисел. Множество сингулярных чисел может быть определено для каждого временного ряда, если использовать одномерный вариант сингулярного спектрального разложения траекторной матрицы в рамках метода МССА. Составим отношения первого сингулярного числа ко второму, второго к третьему и т.д. Затем выбираем для каждой пары номеров сингулярных чисел наибольшее и наименьшее значение

отношения среди всех рядов кластера. Представляет интерес рассмотрение таких отношений небольшого количества первых сингулярных чисел. Как критерий качества, эти отношения можно использовать следующим образом: чем меньше отличие наибольшего и наименьшего отношений для фиксированных номеров сингулярных чисел, тем в большей степени кластерный ряд является однородным. Также заметим, что значительная близость значений пары соседних сингулярных чисел показывает наличие гармонической составляющей в модели временного ряда [9]. Если одно значение в паре отношений сингулярных чисел (k_1, k_2) указывает на присутствие гармоники, а другое – нет, то кластер можно признать неоднородным. Подводя итог сказанному выше, будем считать кластер однородным, если имеет место один из следующих случаев:

- 1) $\min\{k_1, k_2\} \leq 1,1$ и $|k_1 - k_2| \leq 0,5$;
- 2) $\min\{k_1, k_2\} > 1,1$ и $\frac{\max\{k_1, k_2\} - 1}{\min\{k_1, k_2\} - 1} \leq 2$.

Очевидно, что представленные выше константы необходимо определять *эмпирически*. В приведённых условиях они зафиксированы определённым образом лишь для примера.

Дисперсия кластера (дисперсия пространственной изменчивости приращений характеристики). Проанализируем пространственную и временную изменчивость приращений исследуемой характеристики в пределах кластера C за небольшой временной промежуток τ . Для этого найдём разности

$$\Delta F^i(t) = F^i(t) - F^i(t-1), \quad (3)$$

где i – номер узла кластера C . Вычислим среднее значение для каждого временного ряда отдельно

$$\Delta \bar{F}^i(t|\tau) = \frac{\sum_{j=0}^{\tau-2} \Delta F^i(t-j)}{\tau}, \quad (4)$$

и нормируем найденные разности

$$\Delta f^i(t-j) = \Delta F^i(t-j) / \bar{F}^i(t|\tau). \quad (5)$$

При фиксированных значениях t_0 и τ эту процедуру можно распространить на все кластеры сетки UG, сформировав временные ряды и соответствующие карты $\Delta f(t)$ (множества единовременных значений в узлах сетки UG).

Вычислим для каждого момента времени $t \in [t_0 - \tau - 2; t_0]$ оценку дисперсии значений карты в области, ограниченной кластером. Наконец, найдём среднеквадратичную дисперсию:

$$D(C, t, \tau) = \sqrt{\frac{\sum_{j=0}^{\tau-2} D^2(\Delta f(t-j))}{\tau-1}}, \quad (6)$$

которую и будем считать дисперсией пространственной изменчивости приращений характеристики в рамках кластера C на момент t для временного масштаба τ (промежутка времени длины τ).

Градиент кластера. Для каждого узла кластера C в заданный момент времени находим градиент как век-

торную сумму конечных разностей значений характеристики для пары ближайших узлов вдоль горизонтальной и вертикальной осей. На рис. 2 приведён пример расчёта градиента при заданных значениях в соседних узлах: $(3-4; 2-(-1)) = (-1; 3)$. Векторную сумму градиентов для всех узлов одного кластера назовём градиентом этого кластера. Для простоты направление градиента кластера будем приводить к одному из четырёх вариантов, аналогичных расположению ближайших соседей отдельного узла сетки UG. Если предположить, что градиент в выбранном узле на рис. 2 совпадает с градиентом кластера, то результатом применения к нему такой аппроксимации будет направление на север.

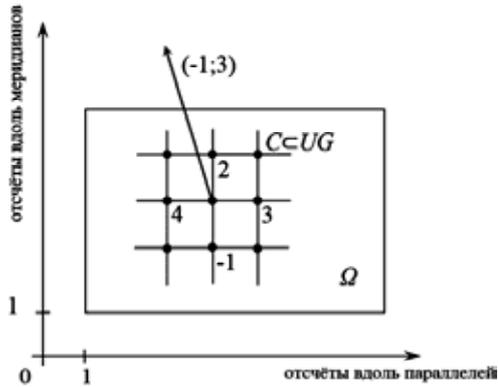


Рис. 2. Градиент характеристики в узле при заданных значениях в 4-х соседних узлах

Эту характеристику можно интерпретировать как критерий качества кластеризации по значению модуля градиента: чем больше его значение (при условии малости дисперсии $D(C, \tau, t)$), тем более однородным является кластер C в момент времени t для временного масштаба τ .

Отметим, что введённые характеристики можно разделить на две группы: меры однородности кластера как меры согласованности в динамике одномерных временных рядов представлены показателями локального (дисперсия и градиент) и глобального (корреляция и отношения сингулярных чисел) временного масштаба.

5. Коррекция результатов начального внутрикластерного прогноза

Метод прогноза значений ряда \tilde{F}_τ^C для кластера C , упомянутый в разделе 3, выполняется исключительно по значениям кластерного ряда F_τ^C , вследствие чего данный этап называется начальным внутрикластерным прогнозом. Методика коррекции его результатов должна использовать информацию о динамике других кластерных рядов, составленных для кластеров узлов сетки UG. Выполнение этого требования позволяет предположить повышение оправдываемости результирующего прогноза.

Принятая в данной работе коррекция результатов начального внутрикластерного прогноза, выполненного с помощью МССА, применяемого к исходному кластерному ряду или его модам МДЭМ, основана на модели одномерного временного ряда TARX (*Threshold AutoRegressive model with eXternal input*). Эта модель представляет собой авторегрессионную модель с дополнительной авторегрессионной частью относительно *внешних*

переменных, причём значения констант в линейной комбинации зависят в том числе от так называемого режима, определяемого моментом времени. Последний факт стал причиной другого названия модели TARX – модели с переключением режимов (*regime switching model*). На данный момент литература по этой модели были представлена исключительно зарубежными работами, например [13].

Общий вид модели TARX в приложении к решаемой задаче имеет следующий вид:

$$e_i(C) = \beta_0^r + \sum_{j=1}^{\tau} \beta_j^r e_{i-j}(C) + \sum_{i=1}^p \sum_{j=1}^{\tau} \beta_{ij}^r e_{i-j}(C_i) \quad (7)$$

где C – *корректируемый кластер*, то есть кластер, прогнозные значения \tilde{F}_τ^C которого необходимо корректировать; $e_i(C)$ – средняя ошибка начального внутрикластерного прогноза \tilde{F}_τ^C для момента t и $\tau=1$ значений ряда кластера C ; C_i – i -ый корректирующий кластер, то есть кластер, ошибка прогноза в котором выступает внешней переменной TARX; p – число корректирующих кластеров; r – номер режима модели TARX.

Отметим, что в модели TARX в данном случае используется один тип регрессора – ошибка начального внутрикластерного прогноза как среднее арифметическое разностей фактического и прогнозного значений, вычисленных для всех одномерных рядов кластера.

Для практического использования модели необходимо провести её структурную и параметрическую идентификацию, а именно выполнить следующие действия:

- 1) задать количество режимов и способы их переключения (смены);
- 2) определить число корректирующих кластеров и провести отбор этих кластеров из множества всех имеющихся;
- 3) найти значения параметров β .

Для пункта (3) из списка выше при условии выполнения двух предыдущих действий есть очевидное решение – метод МНК. Количество режимов зададим равным четырём – по количеству ближайших соседей неограниченных узлов сетки UG. Режим в момент времени t определим как направление градиента корректируемого кластера (см. раздел 4) в момент времени $(t-1)$. Число уравнений, реализующих модель TARX, и число корректирующих кластеров будем выбирать из условия получения квадратной матрицы коэффициентов системы линейных алгебраических уравнений, которую необходимо решать в рамках МНК. Для выбора корректирующих кластеров будем использовать правила, основанные на некоторых свойствах пары (корректируемый кластер C , потенциальный корректирующий кластер K) и значениях кластерных характеристик, введённых в разделе 4.

Случаи запрета использования кластера K :

- 1) непопадание кластера K в заданную окрестность кластера C в рамках сетки UG;
- 2) отставание временных рядов кластера K , то есть наличие пары одномерных рядов (F^i, F^j) , $F^i \in F^K$, $F^j \in F^C$, в которой первый отстаёт от второго в том

смысле, что наибольший коэффициент кросс-корреляции приходится не на нулевой лаг, причём фрагмент ряда F^j использует временные отсчёты, которые не представлены во фрагменте F^i .

Далее введём критерии для отбора p наилучших кластеров среди тех, которые не были опущены по запрещающим правилам. Правила отбора следует рассматривать в соответствии со следующим списком убывания их приоритета:

- 1) задача максимизации: число совпадений направления градиента кластеров K и C для промежутка времени длины τ , непосредственно предшествующего первой прогнозной точке t ;
- 2) задача минимизации: значение дисперсии кластера;
- 3) задача максимизации: значение коэффициента корреляции для пары кластеров (K, C), определяемое как минимум по всем возможным парам одномерных временных рядов из разных кластеров;

Таким образом, выделение корректирующих кластеров есть многокритериальная задача дискретной оптимизации. По каждому из значений критериев можно установить порог, превышение которого является необходимым условием для результата решения этой задачи. Ситуацию, при которой в TARX используется только авторегрессионная часть без внешних переменных, то есть множество корректирующих кластеров является пустым, будем именовать автокоррекцией.

Следует также упомянуть меру однородности кластера по динамике глобального временного масштаба – отношения сингулярных чисел. Наличие свойства однородности, что определяется по численному критерию, сформулированному в разделе 4, можно считать как необходимым условием в определении корректирующего кластера, так и дополнительной информацией при выборе из нескольких альтернатив.

6. Общая схема методики прогноза

Сформируем итоговый вариант внутрикластерного прогноза. Для удобства ссылок введём некоторые обозначения, используя рис. 3. Выполним разметку временной оси, предполагая, что t^* – первая прогнозная точка, T_{opt} – оптимальная длина фрагмента для прогноза, вычисленная по значениям, соответствующим $1, \dots, t^*$. Момент t_0 есть первая прогнозная точка первого начального внутрикластерного прогноза, который повторяется последовательно u раз для формирования данных, на которых настраивается модель TARX. Длина T_{opt} фрагмента временного ряда, который непосредственно участвует в каждом начальном внутрикластерном прогнозе, не изменяется и определяется из условия квазистационарности [5]. Моменту t_0 предшествует фрагмент длины τ , на котором определяется оптимальное количество значимых сингулярных чисел n_{opt} первого прогноза по фрагменту длины T_{opt} , далее для каждого из оставшихся $(u-1)$ прогнозов значение n_{opt} обновляется. В рамках начального внутрикластерного прогноза для обучающего временного ряда [7] используется заданное

значение горизонта τ , а для тестового будем указывать моменты времени, в которые определяются значения, например, «МССА: t_0 » означает вычисления для одного шага по времени (отсчёта t_0), то есть горизонт равен единице. Общая схема методики приведена на рис. 4.

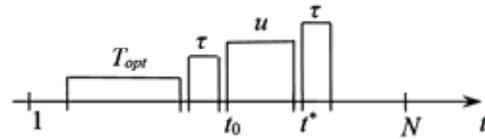


Рис. 3. Разметка временной оси

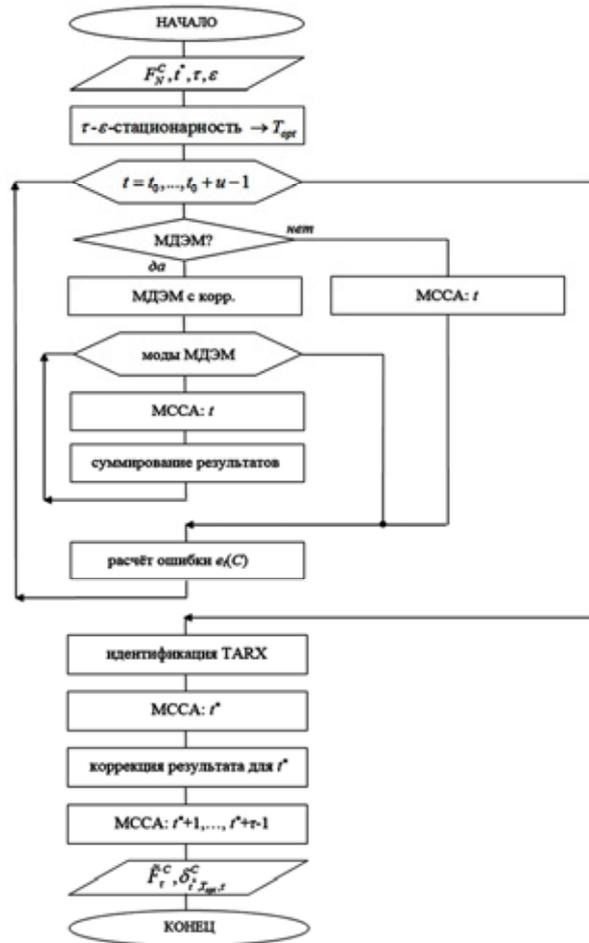


Рис. 4. Блок-схема методики прогноза

7. Вычислительный эксперимент

Апробация отдельных элементов авторской методики прогноза выполнялась для данных альтиметрии (масивов значений аномалий уровня поверхности моря) в работах [6-8]. Полный вариант алгоритма, представленного в предыдущем разделе, был полноценно апробирован на модельных данных. Результаты экспериментов и их анализ представлены в данном разделе.

Рассмотрим сетку UG размером (30x20) и построим для неё временной ряд $F_N^{UG}(t)$ длиной $N = 350$ отсчётов. Априори будем считать, что временные ряды в значительной степени согласованы и образуют несколько кластеров (в соответствии с авторским алгоритмом кластеризации). Каждый из рядов так или иначе строится по следующей модели полезного сигнала:

$$y(t) = P_4(\tilde{t}) + \sin \omega \tilde{t}, \quad (8)$$

для определённости примем, что

$$P_4(\tilde{t}) = 2,7356\tilde{t}^4 - 0,0714\tilde{t}^3 - 10,106\tilde{t}^2 - 4,0196\tilde{t} + 19,0154, \quad (9)$$

$$\tilde{t} = t/90 - 2, \quad t = 0, \dots, 349, \quad \omega = 2,5\pi.$$

К полезному сигналу будет добавляться шум $z(t)$ – случайная составляющая, распределённая по нормальному закону:

$$x(t) = y(t) + z(t), \quad z(t) \sim N(0, \sigma^2), \quad (10)$$

где σ^2 определяется из условия для оценки математического ожидания отношения шума к сигналу:

$$M\left(\frac{z(t)}{y(t)}\right) = 0,05. \quad \text{В данном случае } \sigma=0,7.$$

Априорные кластеры сформированы так:

- кластер 0 (15×10): модель $x(t)$;
- кластер 1 (15×5): линейная аппроксимация модели $x(t)$: $f(t) = 25 - 0,0667t$;
- кластер 2 (15×5): модель $x(t)$;
- кластер 3 (7×5): модель $x(t-10)$;
- кластер 4 (8×5): модель $y(t)$;
- кластер 5 (7×5): модель $x(t-5)$;
- кластер 6 (8×5): модель $x(t+3)$;
- кластер 7 (15×10): модель $x(t)$, в которой установлен более высокий уровень шума: $\sigma=0,9$.



Рис. 5. Профили временных рядов, являющихся представителями кластеров: 0 (зашумлённый), 1 (нелинейный), 4 (линейный)

Таким образом, одномерные временные ряды, составляющие кластерный ряд для кластера 1 (как и для кластера 4), в точности совпадают; для остальных кластеров они отличаются лишь значениями шумовой составляющей (рис. 5).

При выполнении кластеризации получено несколько отличающееся разбиение: кластер с наибольшими шу-

мовыми эффектами разделится на пять независимых кластеров (№№ 7-11, рис. 6). Эти результаты не противоречат закономерностям в модельной генерации рядов, поэтому именно это апостериорное множество кластеров далее используется для вычисления прогноза. Изображение сгенерировано созданной автором утилитой кластеризации (номера добавлены в результате постобработки вручную); прямоугольник наименьшей площади соответствует одному узлу.

Прогноз выполнялся независимо для целевого временного интервала I , содержащего 20 последовательных временных отсчётов (с 220 по 239, рис. 5) на горизонт, равный 5. Для данного примера оценкой математического ожидания оптимальной длины фрагмента временного ряда для прогноза с ошибкой не более 10% по условию квазистационарности из монографии [5] стало значение, близкое к 20 ($T = T_{opt} = 20$). Для экспериментального сравнения результатов прогноза взято ещё одно значение: $T = 30$.

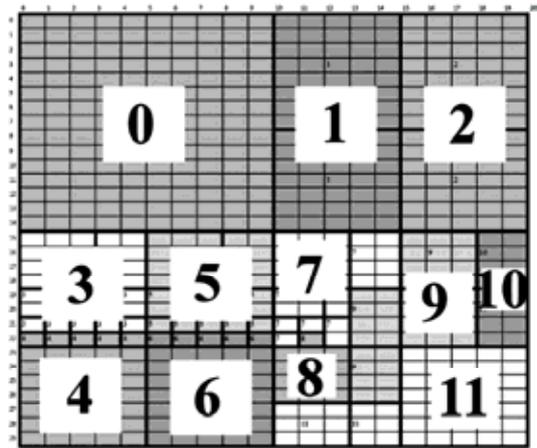


Рис. 6. Результаты кластеризации временных рядов и узлов сети UG

Была проведена серия опытов в предположении, что потенциально возможным корректирующим кластером является только кластер 4, как содержащий только идентичные ряды с полезным сигналом (без шума). Результаты показаны на рис. 7 и рис. 8. Изображения сгруппированы парами и строятся следующим образом: изображение слева отражает результаты коррекции, при которой используются сторонние (корректирующие) кластеры, а справа – коррекции по ошибкам прогноза в самом кластере (автокоррекция). На последующих рисунках тонкая линия показывает ошибку в процентах начального внут-

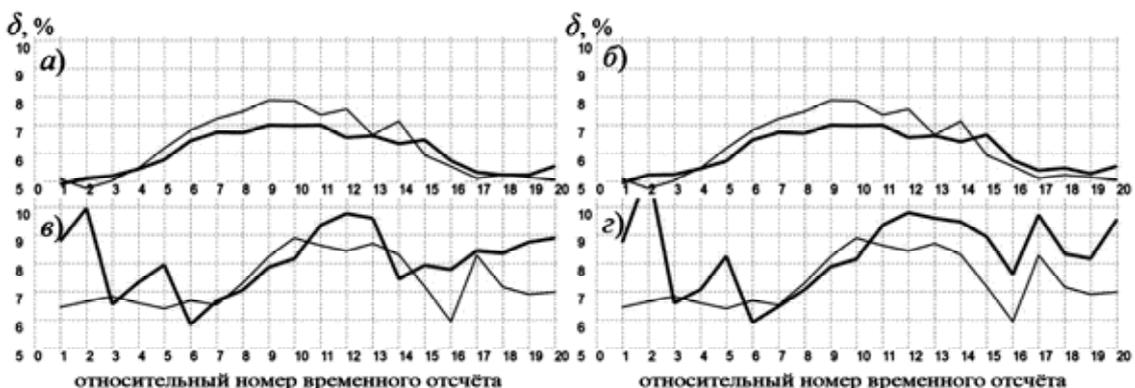


Рис. 7. Средняя ошибка прогноза при $T = 20$ (а), б) и $T = 30$ (в), г), автокоррекция – б) и г)

рикластерного прогноза, утолщённая – ошибку в процентах прогноза после коррекции. Для каждого кластера ошибки вычисляются по формуле (2), итоговое значение для графика выбирается из анализа всех кластеров: либо как средняя ошибка, либо как наибольшая ошибка.

Далее были введены такие ограничения на кластерные характеристики, которые сделали возможным использование большего количества кластеров для коррекции (но не более трёх, не считая сам корректируемый кластер). Результаты второй серии экспериментов приведены на рис. 9 и рис. 10.

Из анализа полученных результатов (рис. 7-10) можно сделать вывод, что начальный внутрикластерный прогноз даёт наибольшую ошибку в центральной части целевого интервала I – эта область соответствует существенному увеличению по модулю величины производной полезного сигнала (8), что видно на рис. 4.

Коррекция прогноза показала наилучшие результаты

при длине фрагмента в 20 отсчётов, что соответствует оценке, полученной из условия квази-стационарности. Стоит отметить, что при использовании более одного корректирующего кластера (рис. 9 а, 10 а), и средняя, и наибольшая ошибки на границах целевого интервала I уменьшаются по сравнению с результатами коррекции с помощью одного 4-го кластера (рис. 7 а, 8 а), а также меньше значений при автокоррекции (рис. 7 б, 8 б). Примечательно, что временные отсчёты, соответствующие границам целевого интервала I , характеризуются инерционностью – смены монотонности профиля временного ряда полезного сигнала (8) не происходит. На рис. 10 а наихудший результат имеет место для отсчёта 16 (отсчёт 235 на рис. 5), что соответствует точке перегиба для полезного сигнала (8). При $T = 30$ начальный внутрикластерный прогноз корректируется плохо во всех случаях, а значения ошибки прогноза без коррекции больше на 2-4% аналогичных значений при $T = 20$.

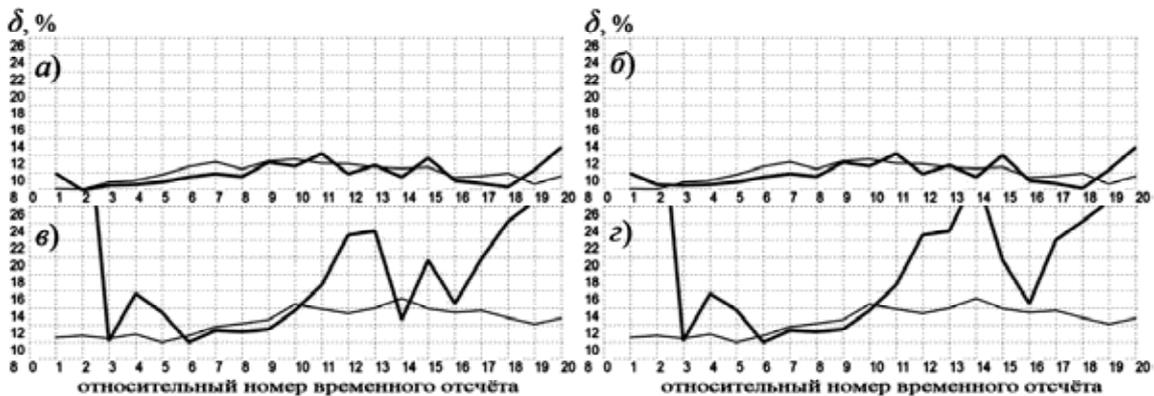


Рис. 8. Наибольшая ошибка прогноза при $T = 20$ (а, б)) и $T = 30$ (в, г)), автокоррекция – б) и г)

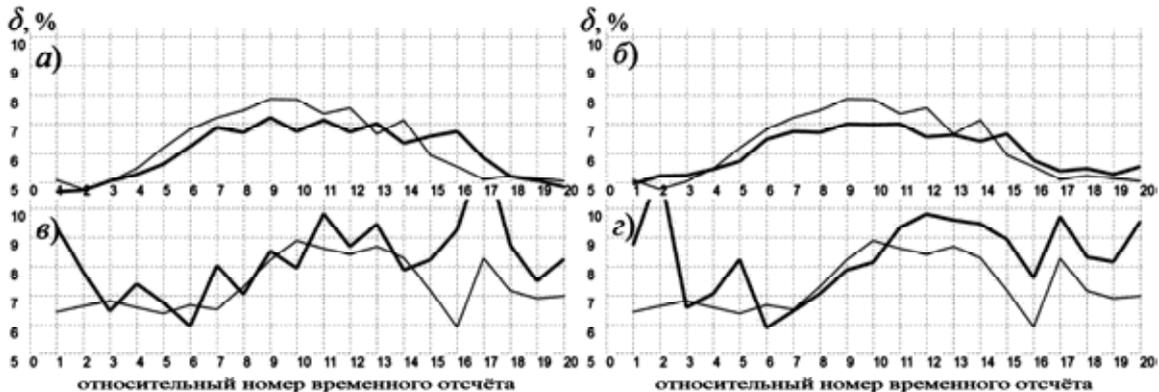


Рис. 9. Средняя ошибка прогноза при $T = 20$ (а, б)) и $T = 30$ (в, г)), автокоррекция – б) и г)

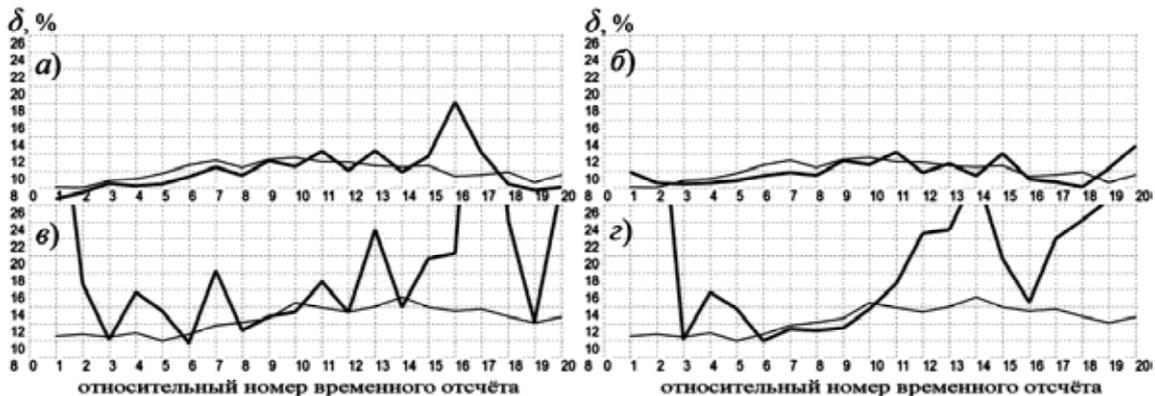


Рис. 10. Наибольшая ошибка прогноза при $T = 20$ (а, б)) и $T = 30$ (в, г)), автокоррекция – б) и г)



Таким образом, коррекция с участием нескольких корректирующих кластеров наиболее эффективна в случае продолжения монотонного участка профиля временного ряда и сохранения знака второй производной полезного сигнала (8). В окрестности точки перегиба начальный внутрикластерный прогноз оказался оптимальным. Наихудшие результаты в случае $T = 30$ как раз приходится на случай коррекции с участием более одного корректирующего кластера (рис 9 в, 10 в).

Заключение

Изложенная в статье методика прогноза является первым шагом авторского коллектива к построению оригинальной технологии краткосрочного прогноза временной изменчивости характеристик естественных процессов, которая удовлетворяла бы требованиям точности, надёжности, адаптивности к различным предметным областям исследований. Важно отметить, что в настоящей статье предложена как общая концепция методики, так и детализация её некоторых шагов. В настоящее время продолжается тестирование методики на модельных данных и начата её апробация в рамках решения задачи прогноза динамики уровня поверхности океана в акваториях Арктики с использованием реальных данных. Эти данные уже участвовали в расчётах по отдельным элементам методики. Стремление к высокой степени автоматизации обработки временных рядов с целью прогноза не исключает существенной роли экспертов в процессе использования данной методики. Предполагается, что её конечная версия будет содержать ряд параметров, для которых потребуется априорное задание рекомендуемых значений, а также предоставление возможности задания других конкретных значений в результате ручной настройки. Таким образом, методика будет реализована в виде интерактивной (на каждом шаге обработки данных) компьютерной технологии, применимой для краткосрочного прогнозирования пространственно-временной изменчивости различных физических характеристик при исследовании природных динамических процессов.

Литература

1. Ашик И.М. Численный гидродинамический метод прогноза колебаний уровня в юго-восточной части Баренцева и юго-западной части Карского морей [Электронный ресурс] / Режим доступа: <http://method.hydro-met.ru/publ/sb/sb31/sb31.html>, свободный. – 2005.
2. Вербицкая О.Н. Гидродинамический метод прогноза синоптических колебаний уровня и течений Каспийского моря: дис. канд. физ.-мат. наук [Электронный ресурс] // МГОУНБ. Диссертации. – 2009. – 175 с.
3. Попов С.К. Прилив в оперативной модели краткосрочного прогноза скоростей течений и уровня моря в Баренцевом и Белом морях / С.К. Попов [и др.] // Метеорология и гидрология. – 2013. – №6. – С. 68-82.
4. Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального подобия: дис. канд. техн. наук [Электронный ресурс] // МГОУНБ. Диссертации. – 2012. – 153 с.
5. Орлов Ю.Н., Осминин К.П. Нестационарные вре-

менные ряды: Методы прогнозирования с примерами анализа финансовых и сырьевых рынков. – М.: Либроком, 2011. – 384 с.

6. Запорожцев И.Ф. Кластеризация узлов планарной регулярной сетки в задаче прогноза временной изменчивости пространственного распределения физических характеристик [Электронный ресурс] / И.Ф. Запорожцев, А.-В.И. Середина // Наука и образование – 2014: материалы междунар. науч.-техн. конф., Мурманск, 24–28 марта 2014 г. / Мурман. гос. техн. ун-т. – Мурманск : МГТУ, 2014. – С. 382-385.

7. Zaporozhtsev I.F. Analysis and Forecast of Sea Level Anomalies Spatio-Temporal Variability in the Barents Sea / I.F. Zaporozhtsev, A.-V.I. Sereda // Computer Modeling and Simulation : труды междунар. науч.-техн. конф., Санкт-Петербург, 2-4 июля 2014 года. – СПб.: Изд-во Политехн. ун-та, 2014. – С. 81-85.

8. Запорожцев И.Ф. Декомпозиция на эмпирические моды в задаче краткосрочного прогнозирования многомерных временных рядов геофизической природы / И.Ф. Запорожцев, А.-В.И. Середина // Цифровая обработка сигналов. – 2014. – №2. – С. 34-40.

9. Голяндина Н.Э. Метод «Гусеница»-SSA: прогноз временных рядов: Учеб. пособие / Н.Э. Голяндина. – СПб.: С.-Петербургский государственный университет, 2004. – 52 с.

10. Степанов Д.А. Варианты метода «Гусеница»-SSA для прогноза многомерных временных рядов / Д.А. Степанов, Н.Э. Голяндина // Труды IV Международной конференции «Идентификация систем и задачи управления» SICPRO'05. – Москва, 2005.

11. Huang N.E. «The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis» [Электронный ресурс] / Режим доступа: http://keck.ucsf.edu/~schenk/Huang_etal98.pdf, свободный. – 1998.

12. Fleureau J. et al. Multivariate Empirical Mode Decomposition and Application to Multichannel Filtering // Signal Processing, vol. 91, pp. 2783-2792, 2011.

13. Tastu J. et al. Spatio-Temporal Analysis and Modeling of Short-Term Wind Power Forecast Errors // Wind Energy, vol. 14, pp. 43-60, 2011.

NUMERICAL METHOD FOR SHORT-TERM FORECASTING OF MULTIVARIATE TIME SERIES EVALUATED IN NODES OF THE PLANAR UNIFORM GRID

Zaporozhtsev I.F., Sereda A.-V.I.

Short-term temporal variability forecasting methodology for spatially distributed characteristics of natural processes based on multivariate time series analysis is presented in the article. Both its general concept and some stages algorithmic specification are designed. Methodology version introduced currently is formed completely and is implemented as a computer technology. The results of numerical experiments for model data and corresponding analysis are given.