

На правах рукописи



РАБАЙА ФУАД

**ИССЛЕДОВАНИЕ ОСОБЕННОСТЕЙ АНАЛИЗА И ПРОГНО-
ЗИРОВАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ
МЕТОДОМ SSA (НА ПРИМЕРЕ АСТРОФИЗИЧЕСКИХ И
ЭКОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ)**

Специальность 05.13.01 – Системный анализ, управление и обра-
ботка информации (в науке и технике)

Научный руководитель:
ПОРШНЕВ Сергей Владимирович,
доктор тех. наук, профессор

Автореферат
диссертации на соискание ученой
степени кандидата технических наук

ЕКАТЕРИНБУРГ – 2015

Работа выполнена на кафедре «Радиоэлектроника информационных систем» Федерального государственного автономного образовательного учреждения высшего профессионального образования «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина».

Научный руководитель: **Поршнеv Сергей Владимирович,**
доктор тех. наук, профессор

Официальные оппоненты: **Зобнин Борис Борисович,**
доктор технических наук, профессор, профессор
кафедры информатики ФГБОУ ВПО «Уральский
государственный горный университет»
Берг Дмитрий Борисович,
Доктор физико-математических наук, профессор,
главный научный сотрудник Института
промышленной экологии Уро РАН

Ведущая организация: **Федеральное казенное предприятие**
«Нижнетагильский институт испытания
металлов», г. Нижний Тагил

Защита состоится 24 декабря 2015 г. в 13 часов на заседании диссертационного совета Д 212.065.06 в ИжГТУ имени М.Т. Калашникова по адресу: 426069, г. Ижевск, ул. Студенческая, 7.

С диссертацией можно ознакомиться в библиотеке университета и на официальном сайте <http://istu.ru>. С авторефератом можно ознакомиться по адресу <http://vak.ed.gov.ru/ru/dissertation>.

Отзыв на автореферат в двух экземплярах, заверенный печатью организации, просим направлять по вышеуказанному адресу на имя ученого секретаря диссертационного совета. В отзыве указываются фамилия, имя, отчество лица, представившего отзыв на автореферат диссертации, почтовый адрес, телефон (при наличии), адрес электронной почты (при наличии), наименование организации, работником которой является указанное лицо, и должность в этой организации.

Автореферат разослан ___ноября 2015 г.

Ученый секретарь диссертационного совета,
кандидат технических наук

В.Н. Сяктерев

Актуальность темы исследования. Математические модели, описывающие статистические зависимости между величинами, изменение которых во времени обусловлено факторами (возможно случайными), скрытыми от наблюдателя, являются инструментом исследования сложных систем процессов, протекающих в окружающем нас мире. В большом количестве случаев исходная информация, используемая для построения данных моделей и оценки их адекватности, представляет собой временные ряды (ВР) – последовательности результатов измерений текущих значений одного или нескольких параметров, проводимых в упорядоченные моменты времени. Для объяснения причин того или иного поведения сложной системы или процесса, породивших данный ВР, выявления и объяснения закономерностей их динамики необходимо решить задачу анализа ВР, которые в подавляющем большинстве случаев оказываются нестационарными. Для прогнозирования динамики развития сложной системы или процесса, как правило, на основе результатов анализа ВР решается задача синтеза модели ВР, используя которую вычисляют прогнозируемые значения ВР. В связи с тем, что универсальных методов решения задач анализа и прогнозирования ВР, использование любого метода анализа и прогнозирования ВР требует проведения исследований с целью определения границ применимости, обоснованного выбора параметров данных методов и оценивания точности получаемых результатов. В этой связи тема исследования является актуальной.

Степень разработанности темы исследования. Анализ методов обработки данной информации позволяет разделить их на две группы: статистические методы, призванные либо по имеющимся данным находить оценки распределения, вид которого априори известен, или проверять статистическую гипотезу относительно вида распределения либо количественно описывать связи между величинами, как правило, находящимися в неочевидной функциональной связи; а также статистические методы анализа ВР, основанные, по сути, на использовании тех или иных математических моделей ВР и идентификации параметров данных модели на основе имеющихся экспериментальных данных.

В основу первой группы статистических методов положено априорное предположение о том, что генеральная совокупность изучаемой генеральной величины существует. Однако оно не выполняется для нестационарных процессов и процессов, физическая природа которых неизвестна, так как каждое значение анализируемого ВР оказывается принадлежащим другой генеральной совокупности, при этом оценка параметров распределения по одному значению случайной величины оказывается невозможной.

Среди основных методов анализа ВР, относящихся ко второй группе, следует отметить: методы выделения тренда (временного сглаживания), регрессионный, автокорреляционный, адаптивный (скользящих средних) будстреп (численного размножения выборок), нейросетовой метод. Необходимо отметить, что применимость данных методов корректно обоснована только для стационарных ВР, в то время как подавляющее большинство ВР, встречающихся на практике, оказывается нестационарными. Поэтому асимптотические критерии, гарантирующие увеличение точности оценок моделей ВР с увеличением

объема используемой выборки (за исключением случая, когда априори известно его функциональное описание), оказываются несостоятельными с точки зрения увеличения точности прогноза и, как следствие, требуют дополнительных проверок точности получаемых результатов.

Следует также отметить, что сегодня существует ряд методов анализа случайных ВР, которые не требуют предположения о стационарности анализируемого ВР, в том числе: вейвлет-преобразование (Meyer Y., Coifman R., Daubechies I., Coifman R., Новиков И.Я., Стечкин С.Б., Петухов А.Б. и др.); преобразование Хуанга-Гильберта (Huang N. E., R. C. Sharpley, V. Vatchev, Tao Q., Zhang L. и др.); метод сингулярного спектрального анализа (Singular Spectrum Analysis – SSA) (Elsner J., Tsonis A., Hassani H., Schoellhamer D.H., Данилов Д.Л., и Жиглявский А.А., Голяндина Н.Э. и др.). При этом метод SSA позволяет не только проводить анализ, но и решает весьма актуальную практически важную задачу прогнозирования нестационарных ВР, в связи с чем исследование данного метода, призванное научно обосновать рекомендации по выбору параметров метода SSA при анализе и прогнозировании нестационарных ВР, а также получить соответствующие оценки точности характеристик прогноза, является актуальным.

Объект исследования: метод сингулярного спектрального анализа (SSA).

Предмет исследования: одномерный ВР, представляющий собой последовательность чисел $F_N = (f_0, f_2, \dots, f_{N-1})$.

Цель диссертационной работы: исследование метода SSA с целью разработки научно обоснованных рекомендаций по выбору параметров метода при анализе и прогнозировании нестационарных ВР и их проверки на примере астрофизических и экономических ВР.

Для достижения поставленной цели решаются следующие **основные задачи исследования:**

1) Разработка научно обоснованных рекомендаций по выбору параметров метода SSA при анализе и прогнозировании реальных ВР.

2) Экспериментальная апробация разработанных рекомендаций по выбору параметров метода SSA при анализе и прогнозировании реальных астрофизических и экономических ВР.

3) Разработка алгоритма оценки точности прогнозирования нестационарных ВР методом SSA.

4) Получение количественных оценок точности прогнозирования нестационарных ВР, содержащих среднемесячные значения чисел Вольфа, а также часовых значений цен на электроэнергию на оптовом рынке электрической мощности (ОРЭМ).

Методы исследования. В работе были использованы статистическое моделирование, спектральный анализ, методы теории вероятности, математической статистики, метод SSA.

Научная новизна полученных результатов. К основным новым результатам, полученным в диссертации, можно отнести следующие:

1. Обоснован критерий выбора значений параметров метода SSA, обеспечивающих идентичность ГК временного ряда, содержащего дискретные значения периодической функции, и гармонических составляющих ряда Фурье.

2. В задаче анализа временного ряда $F_N = F_N^{(1)} + F_N^{(2)}$:

2.1. Предложено использовать зависимость разности между соответствующими собственными числами выборочной траекторной матрицы метода SSA, номера которых отличаются друг от друга на единицу, от значения параметра $L \cdot \Delta t$, L – размер окна сдвига, Δt – период дискретизации анализируемого временного ряда, позволяющей упростить разделение рядов $F_N^{(1)}$, $F_N^{(2)}$.

2.2. Уточнены понятия сильной и слабой разделимости рядов $F_N^{(1)}$, $F_N^{(2)}$.

2.3. На основе статистического моделирования проведено исследование проблемы разделимости ВР вида «шум + const», «шум + периодическая составляющая», «шум + тренд» при различных отношениях сигнал/шум.

3. Предложен алгоритм исследования точности прогнозирования ВР методом SSA.

4. Предложен алгоритм выбора собственных троек траекторной матрицы ВР, используемых для прогнозирования значений ВР.

5. Получено экспериментальное подтверждение целесообразности использования для краткосрочного прогнозирования нестационарных ВР полинома, аппроксимирующего изучаемый ВР, значения которого вычисляются восстановлением исходного ряда по набору собственных троек траекторной матрицы метода SSA, выбираемых в соответствии с предложенным в работе алгоритмом.

Практическая значимость работы

1. Проведен анализ ВР, содержащего среднемесячные значения чисел Вольфа, и ВР, содержащего часовые значения цен на электроэнергию на оптовом рынке электроэнергии и мощности (ОРЭМ), методом SSA, результаты которых подтверждают целесообразность использования зависимости разности между соответствующими собственными числами выборочной траекторной матрицы метода SSA, номера которых отличаются друг от друга на единицу, от значения параметра $L \cdot \Delta t$.

2. Обоснован выбор параметров метода SSA, обеспечивающих наилучшую точность прогнозирования ВР, содержащего среднемесячные значения чисел Вольфа, и ВР, содержащего часовые значения цен ОРЭМ, методом SSA.

3. Получены оценки точности прогнозирования методом SSA ВР, содержащего среднемесячные значения чисел Вольфа, и ВР, содержащего часовые значения цен на электроэнергию на ОРЭМ.

На защиту выносятся:

1. Результаты изучения свойств собственных чисел выборочной траекторной матрицы метода SSA временных рядов вида «периодическая составляющая + периодическая составляющая», «шум + const», «шум + периодическая составляющая», «шум + тренд» и разделимости составляющих данных ВР при различных отношениях сигнал/шум.

2. Результаты анализа методом SSA ВР, содержащего среднемесячные значения чисел Вольфа, и ВР, содержащего часовые значения цен на электроэнергию на ОРЭМ.

3. Алгоритмы исследования точности прогнозирования ВР методом SSA и выбора собственных троек траекторной матрицы ВР, используемых для прогнозирования значений ВР.

4. Результаты оценки точности прогнозирования методом SSA ВР, содержащего среднемесячные значения чисел Вольфа, и ВР, содержащего часовые значения цен на электроэнергию на ОРЭМ.

Достоверность полученных результатов подтверждается обоснованным применением методов статистического моделирования, математической статистики, метода SSA, их согласованностью с соответствующими результатами полученными другими известными методами анализа ВР.

Внедрение результатов диссертационного исследования

Результаты диссертационного исследования использованы в ФГОУ ВПО «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина» при подготовке бакалавров и магистров по направлению «Информационные системы и технологии», в ООО «Октоника» при разработке информационно-аналитических систем.

Апробация работы

Материалы работы докладывались на следующих научных конференциях: Международной научно-практической конференции «СВЯЗЬ-ПРОМЭКСПО 2011», Екатеринбург, 2011 г.; Международной научной конференции «Современные телекоммуникационные системы и компьютерные сети: перспективы развития», Санкт-Петербург, 2011 г.; Международной заочной научной конференции «Наука и образование в XXI веке», г. Тамбов, 2012 г.

Публикации по теме диссертации. По результатам исследований опубликовано 5 печатных работ, из которых в рекомендованных ВАК РФ периодических изданиях – 2.

Структура диссертационной работы. Диссертационная работа состоит из введения, трех разделов, заключения, списка использованных источников, содержащего 117 наименования. Общий объем работы составляет 167 страниц, в том числе 109 рисунков, 9 таблиц.

Содержание работы

Во **введении** изложена общая характеристика диссертационной работы: показана ее актуальность, сформулирована цель работы, отражена научная новизна и практическая ценность, а также кратко изложена суть метода SSA, состоящая в следующем.

Для выделения аддитивных составляющих ВР $F_N = F_N^{(1)} + F_N^{(2)}$ в соответствие анализируемому ряду ставится траекторная матрица X заданного размера $L \times K$, $1 < L < N$, $K = N - L + 1$ (L – длина окна сдвига), вычисляются собственные числа $\{\lambda_k\}_{i=k}^L$, собственные $\{U_k\}_{i=k}^L$ и факторные $\{V_i\}_{i=k}^L$ вектора выборочной траекторной матрицы $X X^T$, формируя сингулярное разложение

$X = \sum_k X_k$, $X_k = \sqrt{\lambda_k} U_k V_k^T$. Здесь набор $(\sqrt{\lambda_k}, U_k, V_k)$ называют k -ой собственной тройкой. Далее выделяется группа собственных троек из некоторого множества I и определяется матрица $X^{(1)} = \sum_{k \in I} X_k$, по которой с помощью диагонального усреднения восстанавливается ВР $\tilde{F}_N^{(1)}$ и, аналогично, ВР $\tilde{F}_N^{(2)}$. Данный алгоритм имеет два параметра длину окна сдвига L и набор собственных троек I .

Во второй главе диссертации проведено исследование особенностей собственных троек сингулярного разложения траекторной матрицы гармонического ВР, результаты которого позволили сделать следующие выводы:

1. При равенстве собственных чисел выборочной траекторной матрицы, номера которых отличаются друг от друга на единицу, выделяемая главная компонента (ГК) соответствует периодической функции, дискретные значения которой есть значения анализируемого ВР, однако, обратное не верно – неравенство собственных чисел не означает, их соответствия составляющим ВР с различными частотами, но требует дополнительно анализа (например, спектрального) собственных векторов.

2. При анализе ВР, содержащих значения реальных физических процессов, следует использовать не параметр сдвига L , но физически содержательную величину $L \cdot \Delta t \cdot f = n$, имеющую следующий смысл: при $L \cdot \Delta t \cdot f = n$, $n = 1, 2, \dots$ – количество периодов периодической функции с частотой f , укладываемогося на временном интервале $L \cdot \Delta t$; при $L \cdot \Delta t \cdot f = n$, $n > 1$, $n \notin \mathbb{N}$, где \mathbb{N} – множество целых чисел, – целое число периодов и доля периода периодической функции с частотой f , укладываемогося на временном интервале $L \cdot \Delta t$; при $L \cdot \Delta t \cdot f = n$, $n > 1$, $n \notin \mathbb{N}$, – множество целых чисел, – доля периода периодической функции с частотой f , укладываемогося на временном интервале $L \cdot \Delta t$.

3. Условие равенства значений собственных чисел выборочной траекторной матрицы метода SSA (параллельности отрезка, соединяющего данные собственные числа оси абсцисс на графике зависимости собственных чисел от их номера) имеет вид:

$$L \cdot \Delta t \cdot f = 0.5k, k = 1, 2, \dots \quad (1)$$

4. Для описания зависимости собственных векторов выборочной траекторной матрицы от времени $x(t), y(t)$ целесообразно использовать следующие математические модели:

$$\begin{aligned} x(t) &= a_x(t) \cos(\omega t + \psi), \\ y(t) &= a_y(t) \sin(\omega t + \psi) \end{aligned} \quad (2)$$

и дополнить алгоритм оценки частоты ГК на основе традиционно используемого в методе SSA анализа формы фигуры Лиссажу нахождением $a_1(t), a_2(t)$, например, с помощью метода аналитического сигнала (АС), и умножением выражений (2) на множители $1/a_x(t), 1/a_y(t)$, соответственно.

Далее проведен анализ разделимости аддитивных составляющих ВР, в том числе и ВР вида «const + изменяющийся во времени ВР», «изменяющийся во времени ВР + изменяющийся во времени ВР», результаты которого показывают, что для выявления собственных троек соответствующих данной ГК целесообразно использовать зависимости разностей между собственными числами выборочной траекторной матрицы метода SSA от параметра $L \cdot \Delta t$ и уточнены понятия сильной и слабой L -разделимости ВР.

а) Сильная разделимость имеет место, когда минимум зависимости разности между соответствующими собственными числами выборочной траекторной матрицы ряда $F_N = F_N^{(1)} + F_N^{(2)}$, номера которых отличаются друг от друга на единицу, оказывается близким к нулю, соответственно, данные собственные числа отличаются друг от друга на малую величину (менее 10% от максимального значения разности). (Пример сильно разделимого ВР представлен на рис. 1а; зависимость логарифма собственного числа выборочной траекторной матрицы анализируемого ВР от его номера – на рис. 1б; зависимость разностей между первым и вторым собственными числами выборочной траекторной анализируемого ВР параметра от параметра $L \cdot \Delta t$, $T_s = 1.0$ с – на рис. 1в; зависимость разностей между третьим и четвертым собственными числами выборочной траекторной анализируемого ВР от параметра $L \cdot \Delta t$, $T_s = 1.0$ с – на рис. 1г).

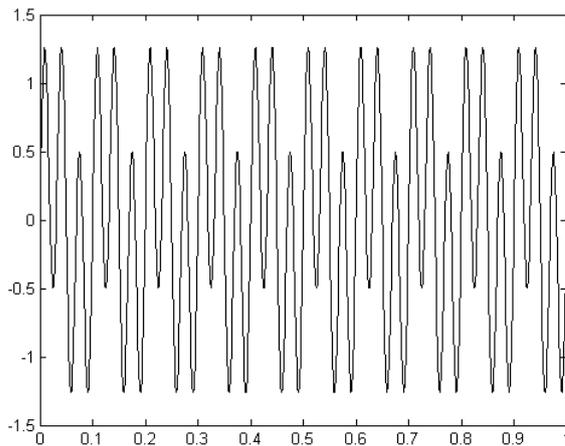


Рис. 1а: Временной ряд

$$F_{N_i} = A_1 \cos\left(2\pi f_1 \frac{T_s}{N-1}(i-1) + \varphi_1\right) + \dots$$

$$A_2 \cos\left(2\pi f_2 \frac{T_s}{N-1}(i-1) + \varphi_2\right),$$

$$A_1 = 1, A_2 = 0.5, f_1 = 10 \text{ Гц}, f_2 = 3f_1,$$

$$\varphi_1 = \varphi_2 = \pi/2 \text{ в узлах временной сетки}$$

$$t_i = \Delta t(i-1), \Delta t = T_s/(N-1), i = \overline{1, N},$$

$$N = 1000, T_s = 1 \text{ с}$$

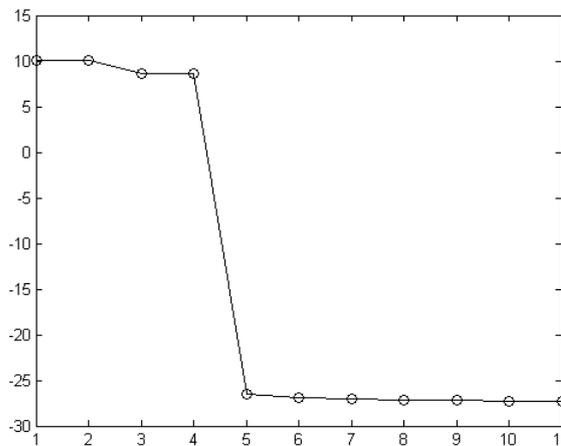


Рис. 1б: Зависимость логарифма собственного числа выборочной траекторной матрицы анализируемого ВР от его номера

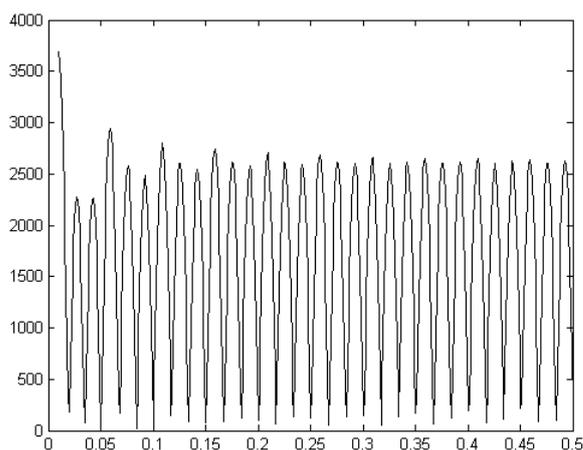


Рис. 1в: Зависимость разностей между первым и вторым собственными числами выборочной траекторной анализируемого ВР параметра от параметра $L \cdot \Delta t$, $T_s = 1.0$ с

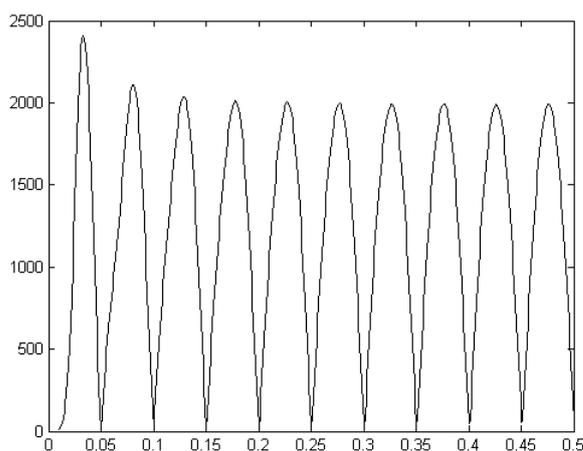


Рис. 1г.: Зависимость разностей между третьим и четвертым собственными числами выборочной траекторной анализируемого ВР от параметра $L \cdot \Delta t$, $T_s = 1.0$ с

б) Слабая разделяемость имеет место, когда минимум зависимости разности между соответствующими собственными числами выборочной траекторной матрицы ряда $F_N = F_N^{(1)} + F_N^{(2)}$, номера которых отличаются друг от друга на единицу, оказывается значимо отличным от нуля, соответственно, данные собственные числа выборочной траекторной матрицы ряда $F_N = F_N^{(1)} + F_N^{(2)}$, отличаются друг от друга на конечную величину (более 10% от максимального значения разности). (Пример слабо разделяемого ВР представлен на рис. 2).

Сводные результаты анализа разделяемости ВР представлены в табл. 1.

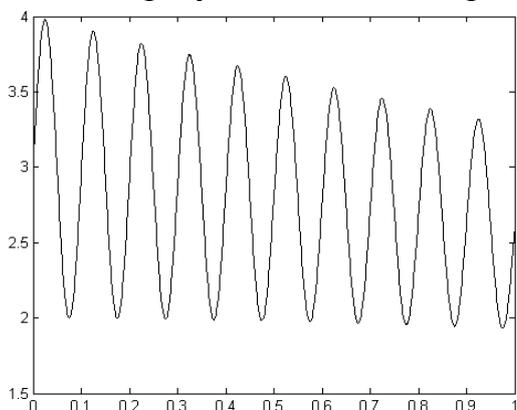


Рис. 2а: Зависимость значение временного ряда

$$F_{N_i} = \left(a \cdot \frac{T_s}{N-1} (i-1) + b \right) + \dots$$

$$A_2 \exp \left(\alpha_2 \frac{T_s}{N-1} (i-1) \right) \cdot$$

$$\cos \left(2\pi f \frac{T_s}{N-1} (i-1) + \varphi \right),$$

$a = -0.3$, $b = 3$, $A_2 = 1.0$, $\alpha_2 = -0.4$,
 $f = 10$ Гц, $\varphi = \pi/2$, $t_i = \Delta t (i-1)$,
 $\Delta t = T_s / (N-1)$, $i = \overline{1, N}$, $N = 1000$,
 $T_s = 1$ с от времени

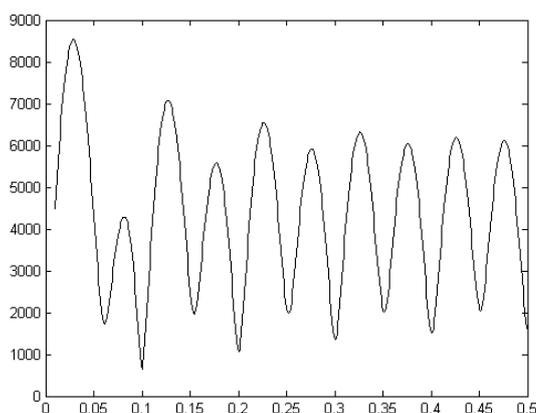


Рис. 2б: Зависимости разности первого и второго собственных чисел выборочной траекторной матрицы анализируемого ВР от величины $L \cdot \Delta t$

Таблица 1. Разделимость рядов $F_N^{(1)}$, $F_N^{(2)}$

	const	cos	exp	exp cos	$a \cdot k + b$
const	Неразделимы	Разделимы точно	Неразделимы	Разделимы точно	Неразделимы
cos	Разделимы точно ²	Разделимы точно ³	Разделимы точно ²	Разделимы точно ²	Разделимы точно ²
exp	Неразделимы	Разделимы точно ²	Неразделимы	Разделимы слабо	Разделимы слабо
exp cos	Разделимы точно	Разделимы точно ²	Разделимы слабо	Разделимы точно ²	Разделимы слабо
$a \cdot k + b$	Неразделимы	Разделимы точно ²	Разделимы слабо	Разделимы слабо	Неразделимы

Далее в работе исследованы особенности собственных чисел и собственных векторов выборочной траекторной матрицы метода SSA временного ряда, представляющего собой смесь шума и детерминированного сигнала, в том числе были изучены сигналы следующих типов: «шум + постоянная составляющая», «шум + периодическая составляющая», «шум + тренд», анализ которых позволил сделать следующие выводы:

1. Метод SSA оказывается не чувствительным в виду функции распределения случайной составляющей – ВР $F_N^{(1)}$. 2. Для получения более точных значений периодической составляющей ВР вида «шум + периодическая составляющая» следует использовать значения параметра сдвига обеспечивающих максимально возможные для данного ряда значения параметра сдвига L .

3. При анализе реальных ВР вида «шум + тренд» следует использовать такие параметры сдвига, чтобы выполнялось условие $L \cdot \Delta t > 0.3$.

В третьей главе диссертации изложены результаты применения метода SSA для анализа и прогнозирования нестационарного ВР, содержащего среднемесячные значения чисел Вольфа (рис. 3), (ВР № 1), и ВР, содержащего часовые значения цен на электроэнергию на ОРЭМ (рис. 4) (ВР № 2).

¹ В данной таблице термины «Разделимы точно», «Разделимы слабо» мы понимаем в смысле данных нами выше определений.

² За исключением случая равных амплитуд.

³ При выполнении условия (1).

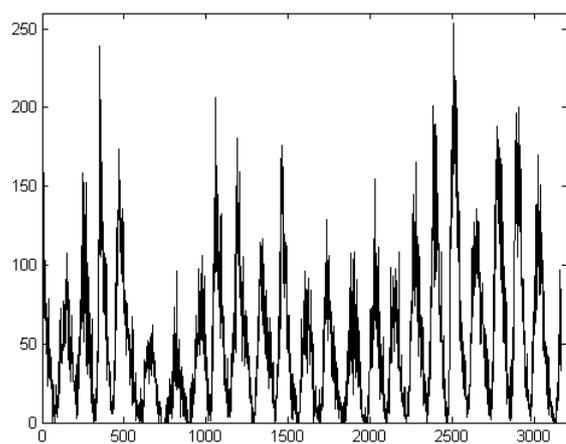


Рис. 3. Зависимость мгновенных среднемесячных значений чисел Вольфа от времени: период с января 1749 г. по май 2012 г., $N = 3551$

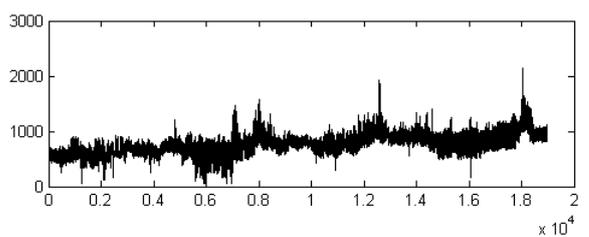


Рис. 4. Зависимость значений часовых цен на электроэнергию (руб./Мвт·час) в период с 1.03.2009 по 31.05.2011 гг. от времени (начало отсчета – 0 час. 1 марта 2009 г., период дискретизации ВР – 1 час) (сверху – исходный ВР, $N = 18831$, снизу – фрагмент ВР)

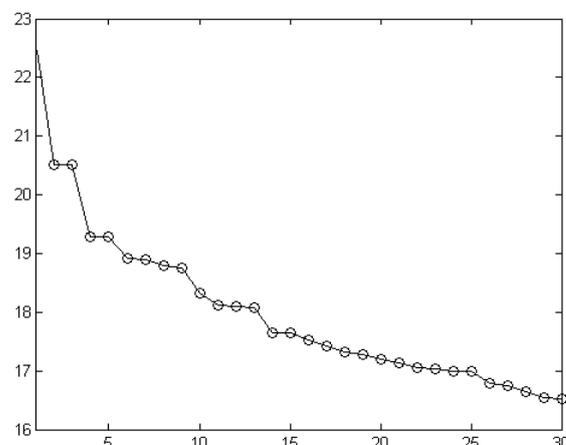
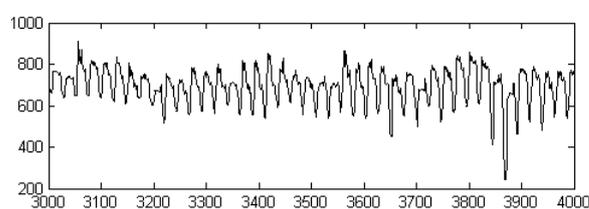


Рис. 5. Зависимость логарифма собственного числа выборочной траекторной матрицы ВР № 1 от его номера

Результаты анализа ВР № 1 представлены на рис. 5 и табл. 2, из которых видно, что для ВР № 1 количество пар собственных чисел, значения которых оказываются достаточно близкими весьма велико, например, 2–3, 4–5, 6–7, 8–9, 11–12 и т.д. Причем, с уменьшением значения собственного значения, появляются собственные векторы периодическая компонента которых имеют частоты, значения которых противоречат физическим представлениям о процессах, протекающих на Солнце (например, у собственных векторов № 20–23 период гармонической составляющей оказывается равным $\approx 5,8$ года). Однако привлечение дополнительного критерия, предложенного в диссертации, позволило существенно сократить количество ГК, выделенных при формальном анализе зависимости, представленной на рис. 5, и использовать далее только компоненты № 1, № 2–3, № 4–5, № 6–7; № 8–9, объясняющие $\approx 86,8\%$ дисперсии ВР № 1. ВР, полученный в результате группировки медленных изменяющихся во вре-

мени собственных векторов выборочной траекторной матрицы ВР № 1, № 1, № 6, 7 переставлен на рис. 6.

Таблица 2. Значения периодов максимальных гармоник периодограмм ГК, выделяемых при анализе зависимости, представленной на рис. 3б

№ ГК	Период, месяц	Период, год	Доля объясняемой дисперсии, %
2,3	130	10.83	15.60
4,5	141	11.75	4.70
8,9	119	9.92	2.60
11,12,14,15	95	7.92	2.40
18,19	183	15,25	0.70
20,21,22,23	68	5.77	1.10
24,25	155	12.92	0.45
30,31	66	5.46	0.28

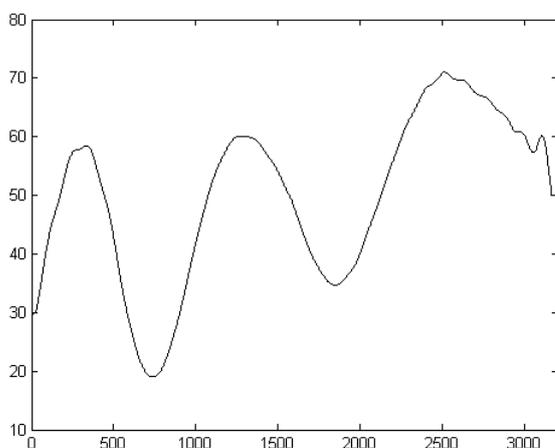


Рис. 6. Зависимость мгновенных значений составляющей анализируемого ВР, полученная группировкой собственных векторов № 1, 6, 7, от времени

Из рис. 6 видно, что в изучаемом ВР присутствует «низкочастотная» составляющая (результат группировки и восстановления компонент № 1, 6, 7), объясняющая 63,5% общей дисперсии временного ряда, которая может быть классифицирована как «вековая» (точнее 80–100 летняя) составляющая. Отметим, что ее период (его оценка может быть получена как длительность временного интервала, между последовательными максимумами рассматриваемой зависимости) обнаруживает тенденцию к увеличению с течением времени: с 1778 г. по 1850 г. равнялся ≈ 72 года, а с 1850 г. по 1980 г. – ≈ 125 лет. Кроме того, следует отметить, что происходит изменение максимального значения рассматриваемой зависимости, равного 71 год в 1778 г., 59 – в 1850 г., и 76 – в 1980 г., соответственно. Полученные результаты позволяют сделать предположение о присутствии в рассматриваемой зависимости периодической составляющей, которая ответственна за амплитудную модуляцию. Период модулирующей функции равняется ≈ 200 лет. Отметим, что полученные результаты вполне согласуются с ранее описанными циклами солнечной активности: 200-летним циклом Зюсса, долгопериодическим циклом Гляйсберга с периодом 80–90 лет, а также 11-летним циклом Швабе-Вольфа.

Зависимость мгновенных значений составляющей анализируемого ВР от времени, полученная группировкой собственных векторов, соответствующих собственным числам № 2–3, 4–5, 8–9, представлена на рис. 7. Зависимости пе-

риода ВР, полученного группировкой собственных векторов выборочной траекторной матрицы № 2–3, 4–5, 8–9, найденные с помощью описанных выше алгоритмов, представлены на рис. 3.8. (Зависимость «мгновенного» периода ВР от времени, вычисленная в соответствии с методом мгновенной частоты аналитического сигнала, для уменьшения шумов, возникающих в результате процедуры численного дифференцирования), была подвергнута процедуре скользящего сглаживания по 5 точкам).

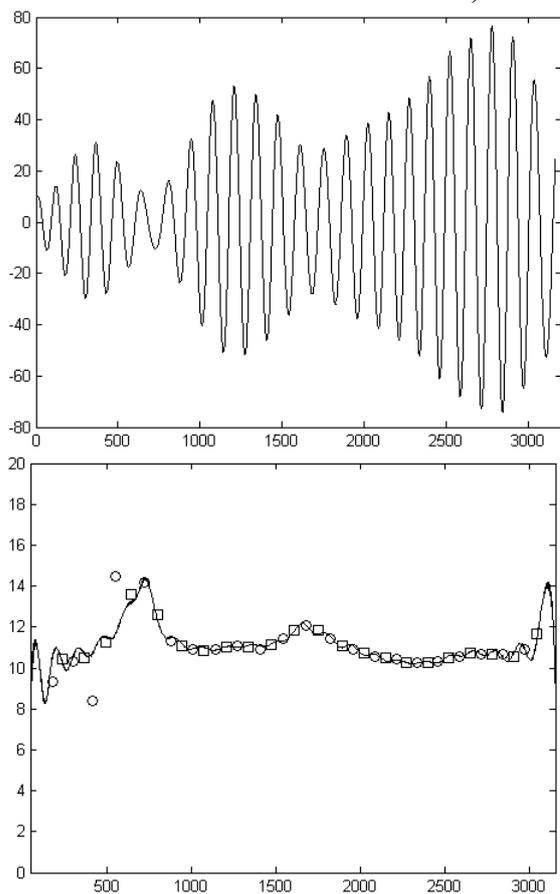


Рис. 7. Зависимость мгновенных значений составляющей анализируемого ВР, полученная группировкой собственных векторов № 2, 3, 4, 5, 7, 8, 9, от времени

Рис. 8. Закон изменения частоты ВР, представленного на рис. 7 (○ – зависимость «мгновенного» периода ВР, оцененное как значение временного интервала между последовательными максимумами ВР; □ – зависимость «мгновенного» периода ВР, оцененное как значение временного интервала между последовательными минимумами ВР; сплошная линия – метод мгновенной частоты аналитического сигнала), во времени

Из рис. 8 видно, что законы изменения периодов во времени, определенные с помощью описанных выше алгоритмов, достаточно хорошо согласуются друг с другом. Минимальное значение периода «одиннадцатилетнего» цикла, равное ≈ 8 годам, было достигнуто в 1763 г., максимальное, равное ≈ 13.6 года, – в 1800 г. Последующие минимумы периода «одиннадцатилетнего» цикла, равные 10.6, 10.2 и 10.5 года были достигнуты в 1842–1843 г., 1938–1939 г., 1995–1996 г., соответственно. Последующие максимумы периода «одиннадцатилетнего» цикла, равные 11.8 и 10.8 года были достигнуты в 1887–1888 г. и 1972–1973 г., соответственно. Отметим, что имеющие место увеличения периодов «одиннадцатилетнего» цикла на концах интервала наблюдения могут оказаться «краевыми» эффектами, присущим алгоритмам, использованных для анализа рассматриваемого ВР. Таким образом, привлечение дополнительного критерия для анализа собственных чисел сингулярного разложения траекторной матрицы ВР № 1 позволило построить адекватную математическую модель данного ВР. Отметим, что, используя зависимость «мгновенных» значений

«одиннадцатилетнего» периода от времени, можно делать прогноз об изменении периода «одиннадцатилетнего» цикла.

Для анализа ВР № 2 в работе использовались: зависимости разностей собственных значений выборочной траекторной матрицы анализируемого ВР от размера окна сдвига L ; зависимости мгновенных значений восстановленных ГК по собственным тройкам сингулярного разложения траекторной матрицы анализируемого ВР ($L = 300$); зависимость нормированной спектральной плотности мощности восстановленных ГК по собственным тройкам сингулярного разложения траекторной матрицы анализируемого ВР ($L = 300$) от частоты. Анализ данных зависимостей показал, что ГК № 1, объясняющая 97,31% дисперсии анализируемого ВР, является трендом, а периодическими оказываются ГК, полученные группировкой следующих собственных троек сингулярного разложения траекторной матрицы: № 2–3, № 4–5. Периоды данных ГК и объясняемые ими дисперсии анализируемого ВР, приведены в табл. 3.

Таблица 3. Количественные характеристики сгруппированных ГК, оказывающихся периодическими

Номера сгруппированных собственных троек	Период	Объясняемая дисперсия
2–3	24 час	1,46%
4–5	12 час	0,39%

Данный результат был также подтвержден проведенным в диссертации анализом ВР № 2 с помощью метода оконного преобразования Фурье, позволивший классифицировать ФВР на временных интервалах длительностью от полумесяца до трех месяцев как «квазистационарный» и выделить спектральные гармоники с аналогичными частотами.

Далее в работе выдвинута гипотеза о возможности использования для вычисления прогнозируемых значений нестационарных ВР дополнительно к известному рекурсивному алгоритму метода SSA, алгоритма, основанного на использовании аппроксимирующего полинома, значения которого вычисляются на основе использования выбранных собственных троек сингулярного разложения траекторной матрицы ВР F_N . Данный алгоритм реализуется следующей последовательностью действий:

1. Получить ВР $\tilde{F}_N^{(1)}$, аппроксимирующий анализируемый ВР $F_N^{(1)}$, используя набор собственных троек сингулярного разложения траекторной матрицы $(\sqrt{\lambda_i}, U_i, V_i)$, $i \in I = (j_1, \dots, j_r)$, из которого исключены собственные тройки, соответствующие шумовой составляющей исходного ВР.

2. Вычислить коэффициенты сплайна, аппроксимирующего ВР $\tilde{F}_N^{(1)}$.

3. Вычислить, используя соответствующие коэффициенты сплайна, прогнозируемые значения анализируемого ВР.

Для анализа точности прогнозирования ВР № 1 в соответствие с рекурсивным алгоритмом и с алгоритмом прогнозирования по полиному, аппроксими-

рующему анализируемому ВР (ряд $\tilde{F}_N^{(1)}$), было предложено использовать следующий алгоритм, описанный ниже на псевдокоде.

Алгоритм 1

Назначение: Исследование точности прогнозирования ВР

Вход:

F (вектор, содержащий значения анализируемого ВР)

Np (длина вектора *F*)

Nwind (размер отрезка анализируемого ВР, используемого для прогнозирования)

Start (точка начального положения окна прогнозирования)

x_Start = *Start* (номер начальной точки скользящего окна)

x_End = *x_Start* + *Start* - 1 (номер последней точки скользящего окна)

x = *F*(*x_Start*:*x_End*:) (вектор, содержащий значения, на основе которых вычисляются значения аппроксимирующего полинома)

In (вектор, содержащий номера группируемых троек собственных сингулярного разложения траекторной матрицы вектора)

Ni (длина вектора *In*)

L (размер сдвига)

Lp (число прогнозируемых точек)

Выход:

матрица *Zz* размерности $(Np - End - Lp) \cdot Ni \times 2 \cdot Lp$

Инициализация:

Km1 = *Np* - *End* - *Lp*;

Km = $(Np - End - Lp) \cdot Ni$;

k = 1;

for *j* = 1 : *Km1*

 for *i* = 1 : *Ni*

m = *In*(*i*);

 % вызов функции, возвращающую вектор-строку, содержащую

 % *LP* спрогнозированных значений в соответствие с

 % рекурсивным алгоритмом

Z1 = *SSA_rekurs*(*x*, *L*, *x_Start*, *x_End*, *Lp*, 1:*m*);

 % вызов функции, возвращающую вектор-строку, содержащую

 % *LP* спрогнозированных значений

 % по аппроксимирующему ряду $\tilde{F}_N^{(1)}$

Z2 = *SSA_appr*(*x*, *L*, *x_Start*, *x_End*, *Lp*, 1:*m*);

 % вызов функции, возвращающую вектор-строку, содержащую

 % *LP* спрогнозированных значений

 % по аппроксимирующему ряду $\tilde{F}_N^{(1)}$ в соответствие

 % с новым алгоритмом

Zz(*k*, 1:2·*Lp*) = *augment*(*Z1*, *Z2*); % объединение строк

k = *k* + 1;

 end;

x_Start = *x_Start* + 1;

x_End = *x_End* + 1;

end;

Средние и среднеквадратические значения разностей между точными и спрогнозированными значениями ВР № 1 представлены в табл. 4.

Таблица 4. Средние и среднеквадратические значения разностей между точными и спрогнозированными значениями ВР № 1

Группируемые компоненты	Прогнозирование с помощью рекурсивного алгоритма метода SSA					Прогнозирование на основе аппроксимирующего полинома				
	Номер прогнозируемой точки					Номер прогнозируемой точки				
	1	2	3	4	5	1	2	3	4	5
1–5	58±40	59±40	59±40	59±40	60±40	20±16	21±17	22±18	23±21	25±26
1–7	60±41	59±41	60±40	60±40	60±40	18±14	20±17	23±22	28±33	36±53
1–9	60±41	60±41	60±40	60±40	60±40	17±14	20±18	26±24	36±50	52±84
1–11	60±41	60±41	60±40	60±40	60±40	17±15	22±21	31±39	47±71	63±95
1–13	60±42	60±41	60±40	60±40	60±41	18±16	24±25	37±49	60±93	78±120
1–15	60±42	60±41	60±41	60±41	59±41	19±17	27±29	45±60	76±115	98±135

Из табл. 4 видно:

1) алгоритм прогнозирования на основе аппроксимирующего полинома при прогнозировании вперед на 1–3 точки оказывается в два-три раза точнее рекурсивного алгоритма;

2) при прогнозировании на 4–5 точек вперед увеличение числа собственных троек, используемых для вычисления значений аппроксимирующего полинома метод прогнозирования, приводит к уменьшению точности прогнозирования данным методом;

3) для практического использования метода прогнозирования на основе аппроксимирующего полинома требуются рекомендации по выбору числа собственных троек;

4) точность прогнозирования рекурсивным алгоритмом в рассмотренном диапазоне параметров оказывается независимой от количества группируемых собственных троек и числа прогнозируемых точек.

Таким образом, полученные результаты позволяют дать обоснованные рекомендации по выбору метода прогнозирования: при краткосрочном прогнозировании следует использовать алгоритм, основанный на использовании аппроксимирующего полинома, при долгосрочном – рекурсивный алгоритм метода SSA.

Анализ табл. 4 показывает, что точность прогнозирования оказывается зависящей от набора собственных троек $(\sqrt{\lambda_k}, U_k, V_k)$ сингулярного разложения траекторной матрицы X ВР F_N . В этой связи для обоснованного выбора числа собственных троек, используемых для группировки и, соответственно, вычисления значений аппроксимирующего полинома (ряд $\tilde{F}_N^{(1)}$) было предложено использовать следующий алгоритм, описанный ниже на псевдокоде.

Алгоритм 2

Назначение: Нахождение набора собственных троек, рекомендуемых для прогнозирования на основе аппроксимирующего полинома

Вход:

F (вектор, содержащий значения анализируемого ВР)

Nr (длина вектора F)

Nwind (размер отрезка анализируемого ВР, используемого для прогнозирования)
Start (точка начального положения окна прогнозирования)
x_Start = *Start* (номер начальной точки скользящего окна)
x_End = *x_Start* + *Start* - 1 (номер последней точки скользящего окна)
x = *F*(*x_Start*:*x_End*:) (вектор, содержащий значения, на основе которых вычисляются значения аппроксимирующего полинома)
In (вектор, содержащий номера группируемых троек собственных сингулярного разложения траекторной матрицы вектора)
Ni (длина вектора *In*)
L (размер сдвига)
Lp (число прогнозируемых точек)

Выход:

матрица *Zz* размерности $(Np - End - Lp) \cdot Ni \times 2 \cdot Lp$

Инициализация:

```

Km1=Np - End- Lp;
Km = (Np - End - Lp)·Ni;
k = 1;
for j = 1 : Km1
  for i = 1 : Ni
    m = In(i);
    % вызов функции, возвращающую вектор-строку, содержащую
    % разности между исходным и аппроксимирующим его ВР
    Z = SSA_res(x, L, x_Start, x_End, Lp, 1:m);
    % вычисление критерия Колмогорова в предположении, что
    % остатки распределены по нормальному закону для каждого
    % сочетания собственных троек
    [mu,sigma] = normfit(Z); % вычисление параметров нормального распределения
    z=sort(Z);
    Nn=length(z);
    S=std(z);
    S=sqrt((Nn-1)/Nn)*S;
    Dn_plus=[];
    Dn_minus=[];
    for is=1:Nn
      Dn_plus(is)=is/Nn-normcdf(z(is),mu,sigma);
      Dn_minus(is)=normcdf(z(is),mu,sigma) - (is-1)/Nn;
    end;
    D1=max(Dn_plus);
    D2=max(Dn_minus);
    Dn=max(D1,D2);
    Sk(i,1)=m;
    Sk(i,2)=(6*Nn*Dn+1)/(6*sqrt(Nn));
    k = k + 1;
  end;
  [Smin I]=min(Sk(:,2),[],1);
  Z2=SSA_appr(x,L,x_Start,x_End,Lp,1:In(I),false);
  Zz2(j,1:Lp)=Z2;
  Zz2(j,Lp+1)=In(I);
  Zz2(j,Lp+2)=Smin;
  x_Start=x_Start+1;
  x_End=x_End+1;

```

end;

Анализ результатов применения данного алгоритма к ВР № 1 показал, что средние значения отклонений между точными и спрогнозированными значениями в рассматриваемых случаях для первой, второй, третьей, четвертой, и пятой прогнозируемых точек составили: 19, 24, 36, 57, 86, соответственно, что согласуется с данными, представленными в табл. 4. Данный результат позволил сделать вывод о том, что для рассматриваемого ВР предложенный критерий выбора собственных троек, используемых для вычисления значений аппроксимирующего полинома, позволяет получать достоверные результаты.

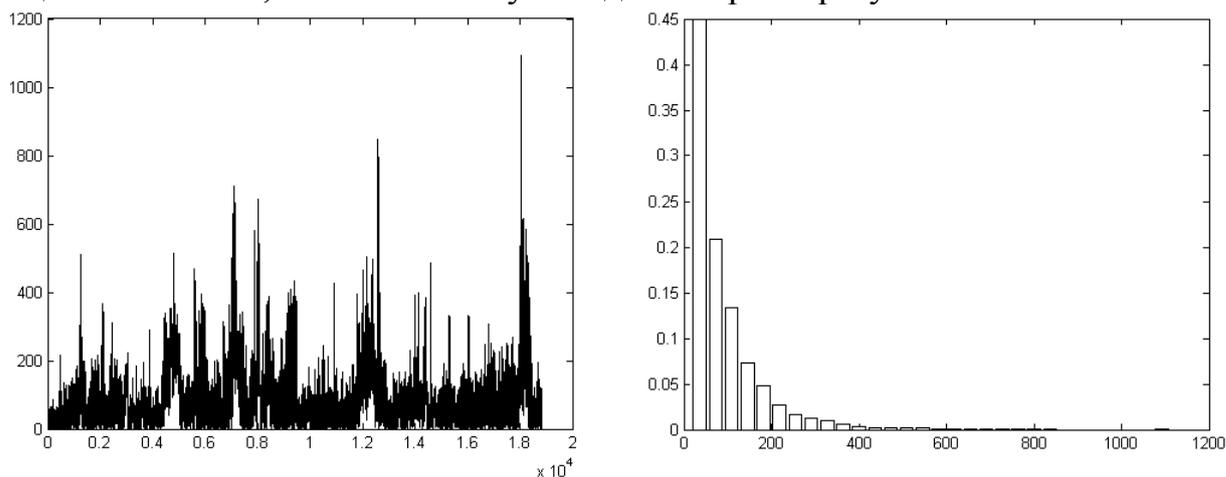


Рис. 9. Зависимость абсолютных значений разностей между фактическими значениями анализируемого ВР и значениями, спрогнозированными в соответствии с АлПр № 1, от номера прогнозируемой точки (слева) и ее гистограмма (справа)

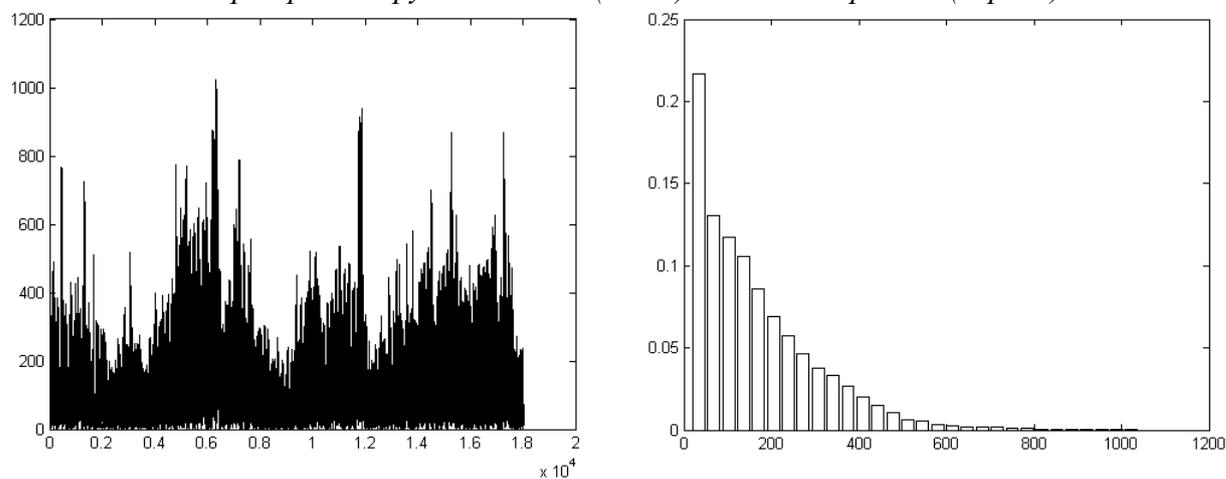


Рис. 10. Зависимость абсолютных значений разностей между фактическими значениями анализируемого ВР и значениями, спрогнозированными в соответствии с АлПр № 2, от номера прогнозируемой точки (слева), и ее гистограмма (справа)

Далее в работе в соответствии с алгоритмами № 1, 2 проведен сравнительный анализ точности прогнозирования ВР № 2 в соответствии с алгоритма прогнозирования данных, размещаемых на сайте <http://www.preprice.ru/> (АлПр № 1), рекурсивным алгоритмом (АлПр № 2), и алгоритмом прогнозирования по полиному, аппроксимирующему анализируемый ВР (ряд $\tilde{F}_N^{(1)}$) (АлПр № 3). ВР, содержащие абсолютные мгновенные значения разностей между фактическими и спрогнозированными значениями в соответствии с перечисленными выше ал-

горитмами прогнозирования, а также гистограммы их распределений представлены на рис. 9, 10, 11.

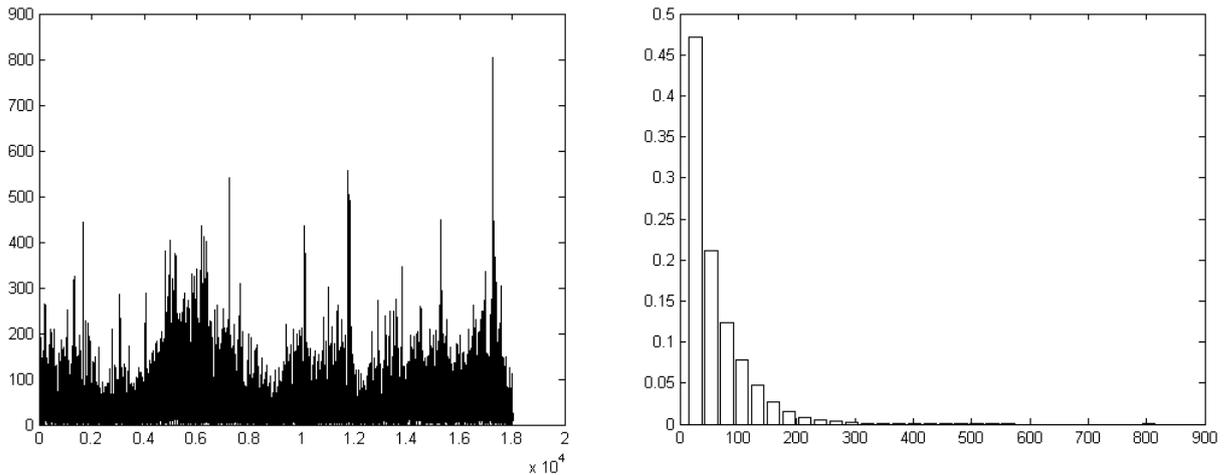


Рис. 11. Зависимость абсолютных значений разностей между фактическими значениями анализируемого ВР и значениями, спрогнозированными в соответствии с АлПр № 3, от времени (слева) и ее гистограмма (справа)

Из рис. 8–10 видно, что у обоих алгоритмов точность прогнозирования оказывается зависящей от локальных особенностей анализируемого ряда, которые априори неизвестны. При этом средние значения и среднеквадратические отклонения значения остатков, представленных на рис 8–10, составили 89 ± 90 , 158 ± 139 , 58 ± 55 , соответственно. Таким образом, в рассматриваемом случае качество прогнозирования на основе АлПр № 3 оказывается наилучшим из всех рассмотренных алгоритмов, а точность АлПр № 2, напротив оказывается ниже, чем у АлПр № 1. Таким образом, можно рекомендовать использовать для прогнозирования значений часовых цен на электроэнергию на ОРЭМ на один час вперед алгоритм, основанный на использовании аппроксимирующего полинома, значения которого вычисляются группировкой ГК № 1–7, вычисляемых для отрезка анализируемого ВР длина $N_{wind} = 800$ и значения сдвига $L = 40$.

Основные результаты диссертационного исследования

1) Исследованы особенности собственных чисел траекторной матрицы метода SSA временных рядов вида $F_N = F_N^{(1)} + F_N^{(2)}$ (как для детерминированных, так и случайных временных рядов), результаты которого свидетельствуют о целесообразности использования при выборе параметров метода SSA временных рядов, содержащих значения разностей между последовательными собственными числами, вычисляемыми для последовательных значений параметра $L \cdot \Delta t$, L – параметр сдвига, Δt – шаг временной сетки, соответствующей анализируемому временному ряду.

2) Проведена экспериментальная апробация предложенных рекомендаций по выбору параметров метода SSA при анализе и прогнозировании реальных временных рядов, содержащих среднемесячные значения чисел Вольфа, а также часовые значения цен на электроэнергию на оптовом рынке электрической мощности.

3) Предложено использовать для прогнозирования нестационарных временных рядов сплайн-интерполяцию аппроксимирующего низкочастотную со-

ставляющую анализируемого временного ряда полинома, значения которого вычисляются в соответствии с методом SSA.

4) Предложен алгоритм оценки точности прогнозирования нестационарных временных рядов методом SSA.

5) Получены оценки точности прогнозирования нестационарных ВР, содержащих среднемесячные значения чисел Вольфа, а также часовые значения цен на электроэнергию на оптовом рынке электрической мощности, подтверждающие преимущества предложенного в работе метода краткосрочного прогнозирования нестационарных временных рядов.

Публикации по теме диссертации

Статьи, опубликованные в журналах из списка ВАК

1. Фуад Рабайа Исследование особенностей временных рядов, содержащих часовые значения цен на оптовом рынке электроэнергии и мощности. [Текст] / С. В. Поршневу, Фуад Рабайа // Естественные и технические науки, 2012. – № 2. – С. 361–367.

2. Фуад Рабайа Об особенностях собственных чисел и собственных векторов выборочной траекторной матрицы в методе SSA. [Текст] / С. В. Поршневу, Фуад Рабайа // Научно-технический вестник Поволжья, 2012. – № 3. – С. 146–150.

Другие издания

1. Рабайа Ф. Исследование особенностей временных рядов, содержащих часовые значения цен на оптовом рынке электроэнергии и мощности. [Текст] / С. В. Поршневу, Фуад Рабайа // Современные телекоммуникационные системы и компьютерные сети: перспективы развития: Труды международной научн. конф. – СПб.: СПбГАСУ, 2011. – С. 109–123.

2. Рабайа Ф. Анализ особенностей временных рядов, содержащих среднесуточные значения цен на оптовом рынке электроэнергии и мощности, методом главных компонент Научные труды международной научно-практической конференции / С. В. Поршневу, Фуад Рабайа // «СВЯЗЬ-ПРОМ 2011» в рамках 8-го Евро-Азиатского форума «СВЯЗЬ-ПРОМЭКСПО 2011». – Екатеринбург: ООО «Компания Реал-Медиа», 2011. С. 84–86.

3. Рабайа Ф. О разделимости аддитивных составляющих в методе SSA. [Текст] / С. В. Поршневу, Фуад Рабайа // Наука и образование в XXI веке: сборник научных трудов по материалам Международной заочной научно-практической конференции 31 мая 2012 г.: в 5 частях. Мин. образования и науки Российской Федерации. Часть 2. – Тамбов: Изд-во ТРОО «Бизнес-Наука-Общество», 2012. – С. 113–115.

Подписано в печать _____.2015

Формат 60x84 1/16. Бумага офсетная.

Печать офсетная. Усл. печ. л. 1,0. Уч.-изд. л. 1,0.

Тираж 100 экз.

Типография «_____»

_____, г. Екатеринбург, ул. _____, _____