САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Александров Фёдор Игоревич

РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА АВТОМАТИЧЕСКОГО ВЫДЕЛЕНИЯ И ПРОГНОЗА АДДИТИВНЫХ КОМПОНЕНТ ВРЕМЕННЫХ РЯДОВ В РАМКАХ ПОДХОДА "ГУСЕНИЦА"-SSA

05.13.18 — Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург
2006

Работа выполнена на кафедре статистического моделирования
математико-механического факультета
Санкт-Петербургского государственного университета

Научный руководитель: доктор физико-математических наук,
профессор Ермаков Сергей Михайлович

Официальные оппоненты: доктор физико-математических наук,
профессор Егоров Владимир Алексеевич

доктор физико-математических наук,
доцент Тайбин Борис Залманович

Ведущая организация: Санкт-Петербургский государственный
политехнический университет

Защита состоится "___" _____________ 2006 г. в ___ часов на заседании диссертационного совета Д 212.232.51 по защите диссертаций на соискание ученой степени доктора наук при СПбГУ по адресу: 198504, Санкт-Петербург, Старый Петергоф, Университетский пр., 28.

С диссертацией можно ознакомиться в научной библиотеке им. М. Горького СПбГУ по адресу: 199034, Санкт-Петербург, Университетская наб., 7/9.

Автореферат разослан "___" _____________ 2006 г.

Ученый секретарь
диссертационного совета,
доктор физ.-мат. наук,
профессор Мартыненко Б. К.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Работа посвящена:

- разработке математических методов автоматической идентификации для задач выделения и прогноза аддитивных составляющих одномерных временных рядов в рамках общего подхода "Гусеница"-SSA;
- созданию на их базе алгоритмов, позволяющих решать задачи выделения аддитивных составляющих ряда как интерактивно, так и при пакетной обработке данных, и реализация их в виде программного комплекса;
- исследованию качества данных алгоритмов при решении поставленных задач;
- исследованию метода оценки коэффициентов линейной рекуррентной формулы, управляющей сигналом временного ряда для случая, когда ряд соответствует модели "сигнал плюс шум".

Актуальность темы следует из:

- практической значимости задач выделения и прогноза аддитивных составляющих одномерных временных и пространственных рядов в различных областях прикладной науки;
- необходимости автоматизации решения этих задач как при работе с одним рядом, так и для пакетной обработки;
- отсутствия на данный момент методов и алгоритмов, позволяющих решать данные задачи в рамках подхода "Гусеница"-SSA.

Целью работы является: 1) разработка новых теоретически обоснованных методов для выделения и прогноза таких аддитивных составляющих временных рядов как тренд и периодические составляющие; 2) разработка алгоритмов для автоматической обработки временных рядов на основе предложенных методов и реализация этих алгоритмов на ЭВМ; 3) статистическое исследование данных алгоритмов; 4) исследование методов оценки параметров сигнала ряда.

Методика исследования включает в себя применение подхода "Гусеница"-SSA, теорию сингулярного разложения матриц, использование свойств их собственных значений и собственных векторов, основы обработки цифровых сигналов, методы разложения Фурье, регрессионные методы. Численные оценки получались с помощью методов статистического моделирования. Для реализация алгоритмов использовались средства программирования Microsoft Visual C++ и Matlab.

Научная новизна. В работе получены следующие результаты:

- разработаны математические методы автоматического выделения и прогноза тренда и периодической составляющей ряда, управляемые заданием параметров и пороговых значений;
- описаны рекомендации по выбору параметров методов исходя из специфики поставленной задачи и предложены способы выбора пороговых значений для рассматриваемого ряда;
- на модельных примерах с помощью статистического моделирования исследовано качество предоставляемой автоматизации в рамках подхода "Гусеница"-SSA и качество получаемой аппроксимации;
- представлено решение задачи выделения тренда для множества рядов, исследованы вопросы проверки работоспособности методов в данном случае;
- проведено статистическое сравнение метода оценки коэффициентов линейной рекуррентной формулы, основанного на подходе "Гусеница"-SSA с регрессионным методом, основанным на оценке метода наименьших квадратов.

Практическая ценность. Предложенные алгоритмы выделения аддитивных составляющих могут быть использованы на практике для эффективного решения задач обработки как временных, так и пространственных рядов. Исследованные методы не только упрощают работу с методом "Гусеница"-SSA, но и позволяют расширить его применение на задачи автоматизированной (в частности, неинтерактивной) обработки данных.

Апробация работы. Основные результаты обсуждались на:

- семинарах кафедры статистического моделирования математико-механического факультета Санкт-Петербургского государственного университета;
- семинаре Time Series group в Astronomical Institute of Tübingen

(Тюбинген, Германия, март 2005);

- семинарах в Zentrum für Technomatematik, Bremen University (Бремен, Германия, март 2005, март 2006);

и докладывались на

- международной конференции System Identification and Control Problems'05, (Москва, январь 2005);
- международном семинаре 5th Workshop on Simulation (Санкт-Петербург, июнь 2005);
- международном семинаре Workshop on nonlinear and nonstationary time series in Kaiserslautern (Кайзерслаутерн, Германия, сентябрь 2005);
- международном симпозиуме 26th International Symposium on Forecasting (Сантандер, Испания, июнь 2006).

Работа над диссертацией была частично поддержана стипендией Правительства РФ для аспирантов за 2005-2006 гг. и грантами CRDF № RUB1-1556-ST-05, № RUB1-1643-ST-06.

Публикации. Основные результаты диссертации опубликованы в работах [1-5]. В совместных работах диссертанту принадлежит выбор методов решения, доказательства теорем и численное моделирование, а соавтору Голяндиной Н.Э. — постановки задач [1,4], а также идея метода низких частот идентификации трендовых составляющих [1] и описание схемы расчёта оптимальных пороговых значений для временных рядов фиксированной модели [2].

Структура и объем работы. Диссертация объёмом 152 страницы состоит из введения и четырёх глав, разбитых на разделы и параграфы. Содержит 145 рисунков, 36 таблиц и список цитируемой литературы из 69 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** приводится обоснование актуальности темы диссертации, формулируется цель и задачи работы, а также кратко описывается структура диссертации.

Основным объектом исследования является *одномерный временной ряд* – последовательность вещественных чисел $F = (f_0, f_1, \ldots, f_{N-1})$ длины N. Обычно это наблюдения в равноотстоящие моменты времени, но, несмотря на название, эта последовательность может иметь другую природу. Примеры пространственных рядов приводятся в диссертации. Предполагаем, что ряд состоит из суммы аддитивных составляющих, из которых наиболее интересны тренд ($F^{(\mathrm{T})}$) и периодическая составляющая с некоторым периодом ($F^{(\mathrm{P})}$):

$$F = F^{(\mathrm{T})} + F^{(\mathrm{P})} + F^{(\mathrm{R})}, \quad f_n = f_n^{(\mathrm{T})} + f_n^{(\mathrm{P})} + f_n^{(\mathrm{R})},$$

где в качестве $F^{(\mathrm{R})}$ обозначаем остаток, являющийся реализацией некоторой случайной последовательности. $F^{(\mathrm{R})}$ соответствует шуму, который имеет случайную природу и часто присутствует в наблюдениях. *Трендом* будем называть медленно меняющуюся составляющую ряда (для определённости полагаем, что она не является периодической). *Периодическую составляющую* $F^{(\mathrm{P})}$ с периодом T ($T \geqslant 2$) определим следующим образом:

$$f_n^{(\mathrm{P})} = \sum_{k=1}^{\lfloor T/2 \rfloor} A_k e^{\alpha_k n} \cos(2\pi nk/T + \phi_k), \quad \alpha_k, A_k \in \mathbb{R},\ \phi_k \in [0, 2\pi). \tag{1}$$

Основной задачей является выделение из известного ряда F неизвестных составляющих $F^{(\mathrm{T})}$ и $F^{(\mathrm{P})}$. Целью диссертации является разработка методов для решения этой задачи, которые можно использовать для автоматизированной обработки рядов. Для этого используется общий подход исследования временных рядов "Гусеница"-SSA.

Первая глава содержит информацию о подходе "Гусеница"-SSA, которая потребуется для решения поставленных задач. В рамках этого подхода описываются уже существующие интерактивные способы решения задач выделения тренда и периодических составляющих временных рядов, а также рассматриваются условия, при которых возможно решение этих задач.

Для того чтобы выделить аддитивную составляющую ряда F длины N, $F = F^{(1)} + F^{(2)}$, в подходе "Гусеница"-SSA по ряду F строится *траекторная матрица* $\mathbf{X}$ заданного размера $L \times K$, $1 < L < N$, $K = N - L + 1$ (L называется длиной окна), вычисляются собственные числа $\{\lambda_k\}_{k=1}^{L}$, собственные $\{U_k\}_{i=k}^L$ и факторные $\{V_i\}_{i=k}^L$ вектора матрицы $\mathbf{X}\mathbf{X}^T$, формируя *сингулярное разложение* $\mathbf{X} = \sum_k \mathbf{X}_k$, $\mathbf{X}_k = \sqrt{\lambda_k} U_k V_k^T$. Набор $(\sqrt{\lambda_k}, U_k, V_k)$ будем называть *k-ой собственной тройкой*.

Затем выделяется группа собственных троек с номерами из некоторого $\mathcal{I}$ и определяется матрица $\mathbf{X}^{(1)} = \sum_{k\in\mathcal{I}} \mathbf{X}_k$, по которой восстанавливается требуемая составляющая $\widetilde{F}^{(1)}$ с помощью диагонального усреднения (ганкелизации).

Эта схема имеет параметр – длину окна L и управляется выбором группы компонент $\mathcal{I}$, процесс этого выбора будем называть *идентификацией*. Нашей основной задачей будет автоматизация идентификации при решении поставленных задач.

Во **второй главе** предлагается и исследуется метод автоматического выделения тренда.

В разделе 2.1 приводятся данные, необходимые для определения тренда на языке разложения Фурье. Для элементов ряда F вводится *периодограмма* Π_F^N, определённая на дискретном множестве частот:

$$\Pi_F^N(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi\omega n} f_n \right|^2, \quad \omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}.$$

Так как по определению тренд – это последовательность, элементы которой медленно меняются, то его периодограмма в некоторой низкочастотной области будет иметь большие значения по сравнению с её значениями на остальной части частотного интервала $[0, 0.5]$, причём периодограммы собственных векторов, соответствующих тренду, обладают этим же свойством.

Введём ω_0 – границу интервала низких частот и в разделе 2.2 построим *метод низких частот* идентификации трендовых собственных троек, который для каждого собственного вектора будет проверять, достаточно ли велик вклад значений его периодограммы на интервале $[0, \omega_0]$. Критерий идентификации собственного вектора $U \in \mathbb{R}^L$ выглядит так:

$$\frac{\sum_{0\leqslant\omega\leqslant\omega_0} \Pi_U^L(\omega)}{\sum_{0\leqslant\omega\leqslant 0.5} \Pi_U^L(\omega)} \geqslant \mathcal{C}_0, \quad \omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor},$$

где $\mathcal{C}_0$ – пороговое значение метода. На основе этого метода построен алгоритм Trend для выделения тренда ряда, параметрами которого являются L, ω_0 и пороговое значение $\mathcal{C}_0$.

В разделе 2.3 изучается проблема выбора ω_0 и предлагаются варианты её решения в зависимости от специфики поставленной задачи и имеющейся информации о ряде. Основной способ состоит в исследовании периодограммы исходного ряда F, так как в утверждении 2.1 доказано, что в случае, если $F = G + H$:

$$\left|\Pi^N_{G+H}(k/N) - \Pi^N_G(k/N) - \Pi^N_H(k/N)\right| \leqslant 2\sqrt{\Pi^N_G(k/N)\Pi^N_H(k/N)}. \qquad (2)$$

Так как, в отличие от периодограмм других составляющих, периодограмма тренда имеет большие значения в низкочастотном интервале, то из (2) следует, что, при наличие в ряде тренда, значения Π^N_F в низкочастотном интервале также будут большими.

В разделе 2.4 проводится статистическое исследование алгоритма Trend для решения задачи выделения экспоненциального тренда в присутствии белого нормального шума с изменяющейся дисперсией. Для такого ряда можно имитировать процедуру визуального выделения тренда, так как известно, что тренду в этом случае соответствует первая собственная тройка ($\mathcal{I} = \{1\}$). Показано, что при наилучшем в среднем пороговом значении $\mathcal{C}_0$, во-первых, результаты алгоритма Trend близки к результатам имитированной визуальной (интерактивной) процедуры идентификации, и, во-вторых, качество предоставляемой им аппроксимации достаточно высоко. Качество оценивается с помощью как среднеквадратичной ошибки $||F^{(\mathrm{T})} - \widetilde{F}^{(\mathrm{T})}(\mathcal{C}_0)||^2_2$ (где $\widetilde{F}^{(\mathrm{T})}(\mathcal{C}_0)$ – тренд, выделенный с помощью Trend с $\mathcal{C}_0$), так и вероятности ошибки идентификации первого рода – когда $\mathcal{I}(\mathcal{C}_0) \not\supset \mathcal{I}$ (где $\mathcal{I}(\mathcal{C}_0)$ – группа идентифицированных с помощью Trend с $\mathcal{C}_0$ компонент).

Таким образом, установлено, что при некотором $\mathcal{C}_0$ процедура Trend действительно может быть использована для автоматического выделения тренда, и встаёт проблема выбора $\mathcal{C}_0$.

Далее исследуется вопрос оценки качества результата алгоритма Trend и в разделе 2.6 предлагается некоторая мера качества $\mathcal{R}(\mathcal{C}_0)$, которая рассчитывается только по результату, $\widetilde{F}^{(\mathrm{T})}(\mathcal{C}_0)$, и для которой показано, что в рассмотренных модельных случаях она согласована с

$||F^{(\mathrm{T})} - \widetilde{F}^{(\mathrm{T})}(\mathcal{C}_0)||_2$ таким образом, что используя меру $\mathcal{R}(\mathcal{C}_0)$, можно выбрать такое пороговое значение $\mathcal{C}_0$, что при нём идентифицируются все компоненты тренда. Предлагается алгоритм TrRmeas, использующий процедуру Trend с $\mathcal{C}_0$, выбранным с помощью $\mathcal{R}(\mathcal{C}_0)$.

В разделе 2.7 рассматривается случай, когда известна стохастическая модель остатка (шума). Показано, как при этом можно сузить область поиска порогового значения $\mathcal{C}_0$ и тем самым увеличить надежность процедуры поиска $\mathcal{C}_0$.

В разделе 2.8 приводятся примеры выделения тренда с помощью алгоритма TrRmeas для реальных рядов: ежемесячный уровень безработицы в различных штатах США и уровень экспрессии некоторых генов плодовой мушки Drosophila. Два примера изображены на рис. 1.

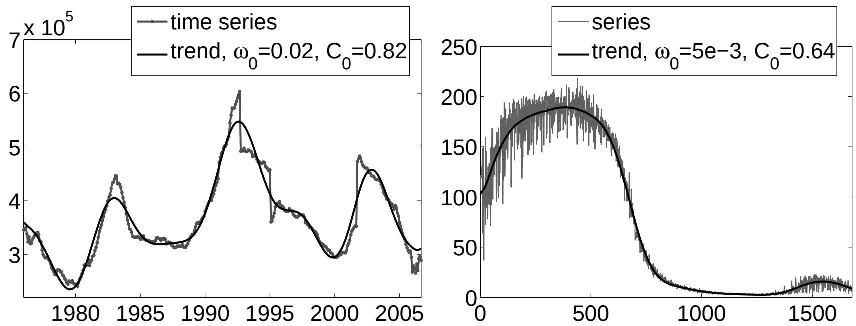

Рис. 1: Уровень безработицы во Флориде (слева) и пространственное распределение уровня экспрессии гена hunchb в эмбрионе плодовой мушки Drosophila (справа), тренды выделены с помощью TrRmeas

Для того чтобы применять алгоритм TrRmeas к ряду, необходимо, чтобы выполнялись некоторые требования, обусловленные свойствами как подхода “Гусеница”-SSA, так и метода низких частот и способа выбора $\mathcal{C}_0$. Не имея информации о модели ряда, проверить эти требования невозможно, и применение алгоритма TrRmeas, строго говоря, необоснованно. В разделе 2.9 исследуется эта проблема и рассматривается случай, для которого возможно проверить обоснованность применения TrRmeas – обработка множества рядов. Описывается математическая модель для этой задачи, и выводятся необходимые требования однородности для рядов из обрабатываемого множества.

В разделе 2.10 предлагается алгоритм TrFore автоматического прогноза тренда ряда, основанный на методе низких частот. Данный алгоритм использует линейно-рекуррентное представление тренда, которое может быть найдено по идентифицированным компонентам $\mathcal{I}(\mathcal{C}_0)$. С помощью статистического моделирования изучается проблема выбора для него порогового значения $\mathcal{C}_0$ и на примере экспоненциального тренда показано, что наилучшие в среднем пороговые значения $\mathcal{C}_0$ для аппроксимации и для прогноза – совпадают. Это позволяет пользоваться при прогнозе тренда теми же методами выбора $\mathcal{C}_0$, что и при аппроксимации. Как и для случая аппроксимации, показано, что при наилучшем в среднем $\mathcal{C}_0$ результаты алгоритма прогноза близки к результатам имитируемой визуальной процедуры идентификации, и что качество предоставляемого алгоритмам TrFore прогноза достаточно велико.

Третья глава посвящена автоматическому методу выделения периодической составляющей ряда.

Задача выделения периодической составляющей сильно отличается от задачи выделения тренда. Периодическая составляющая представляет собой известную параметрически-заданную функцию (1), в то время как тренд определялся по косвенному признаку – вкладу гармоник с низкими частотами.

Известно, что при выполнении определённых условий *экспоненциально-модулированная (э-м) гармоническая составляющая* с частотой ω (или просто *э-м гармоника*), n-ый элемент которой задаётся так:

$$Ae^{\alpha n}\cos(2\pi n\omega + \phi), \quad \alpha, A \in \mathbb{R},\ \phi \in [0, 2\pi),$$

порождает две собственных тройки (одну, если $\omega = 0.5$), элементы собственных векторов $U_j = (u_1^{(j)}, u_2^{(j)}, \ldots, u_L^{(j)})$ которых также задаются э-м гармонической функцией с теми же ω и α, причём одному соответствует э-м синус, а другому – э-м косинус. При более слабых условиях это будет выполняться приближённо.

Пользуясь этим фактом, построим в разделе 3.2 *метод Фурье* для идентификации собственных векторов э-м гармоники, основанный на исследовании периодограмм последовательностей их элементов $\Pi_{U_j}^L$. Метод Фурье состоит из двух частей, в первой для пар векторов вида U_i, U_{i+1} проверяется, что максимумы их периодограмм совпадают или достаточно близки:

$$L|\theta_i - \theta_{i+1}| \leqslant s_0, \quad \theta_j = \arg \max_{0 \leqslant k \leqslant L/2} \left\{ \Pi_{U_j}^L(k/L) \right\}, \tag{3}$$

где $s_0 \in \mathbb{Z}_+$ – пороговое значение. Во второй – для найденных пар векторов проверяется, что значение максимума превышает уровень, необходимый для того, чтобы они соответствовали э-м гармонике:

$$\max_{0 \leqslant k \leqslant L/2} \left\{ \Pi_{U_i}^L(k/L) + \Pi_{U_{i+1}}^L(k/L) \right\} \geqslant 2\rho_0, \tag{4}$$

где $\rho_0 \in [0, 1]$ – пороговое значение. На основе этого метода идентификации предлагается алгоритм Per выделения э-м гармонической составляющей ряда, параметрами которого являются L, s_0 и ρ_0. В разделе также численно и аналитически изучается поведение критериев (3), (4) для э-м гармоник с разными параметрами ω, α, и показано, что в большинстве случаев можно брать $s_0 = 1$.

В разделе 3.3 проводится проверка алгоритма Per с помощью статистического моделирования для модели э-м гармоники с белым нормальным шумом с изменяющейся дисперсией, причём схема исследования совпадает с использованной для Trend ранее. В нём показано, что при наилучшем в среднем ρ_0 алгоритм обеспечивает хорошее качество аппроксимации, а также исследуется изменение качества с тем, чтобы сформулировать принципы выбора ρ_0.

В следующих разделах описаны способы выбора ρ_0 при решении задачи выделения э-м гармонической компоненты с помощью алгоритма Per. Предлагается аналитический метод, позволяющий выбрать ρ_0 для выделения э-м гармоники с известным α при условии, что $L\omega \in \mathbb{N}$. Этот метод заключается в использовании для э-м гармонической последовательности с частотой ω оценки значения суммы периодограмм её собственных векторов в точке ω, которая была получена в разделе 3.2.2:

$$\Pi_{U_i}^L(\omega) + \Pi_{U_{i+1}}^L(\omega) \approxeq \frac{4}{\gamma} \frac{(e^\gamma - 1)}{(e^\gamma + 1)}, \quad \text{где} \quad \gamma = L\alpha.$$

Однако то, что α зачастую неизвестно, а также накладываемое ограничение $L\omega \in \mathbb{N}$ и другие причины делают этот метод полезным только в частных случаях. Поэтому далее предлагается эмпирический метод выбора ρ_0. На его основе предлагается алгоритм PerEmp автоматического выделения гармоники, параметрами которого являются: L и A – ограничение снизу на амплитуду искомой гармоники. При выделении э-м гармоники алгоритм можно сделать более точным, задав ограничение снизу на α и приблизительное значение частоты искомой э-м гармоники ω, если они известны.

Там же приведено статистическое исследование алгоритма PerEmp, которое подтверждает хорошее качество данного алгоритма как для модели $F^{(\mathrm{P})} + F^{(\varepsilon)}$, где $F^{(\varepsilon)}$ – шум, так и в присутствии некоторого тренда: $F^{(\mathrm{T})} + F^{(\mathrm{P})} + F^{(\varepsilon)}$.

Так как периодическая составляющая с периодом T состоит из суммы э-м гармонических компонент с частотами вида $k\omega$, $\omega = 1/T$ (см. (1)), то для её выделения необходимо уметь оценивать частоту найдённой э-м гармонической компоненты. В разделе 3.6 рассмотрен метод оценки частоты, который использует линейное рекуррентное представление э-м гармоники $f_n = Ae^{\alpha n}\cos(2\pi\omega n + \phi) = a_1 f_{n-1} + a_2 f_{n-2}$:

$$\omega(a_1, a_2) = (2\pi)^{-1} \arccos\big(a_1(2\sqrt{a_2})^{-1}\big). \tag{5}$$

В разделе 3.7 приводится пример применения алгоритма PerEmp к ряду, содержащему данные о ежемесячном уровне потребления электроэнергии в США. Приводится решение задачи разложения ряда на тренд, сезонную составляющую (периодическую составляющую с периодом $T \approx 12$) и остаток, см. рис. 2.

В разделе 3.8 представлен алгоритм автоматического прогноза периодической составляющей и, как и в главе 2, с помощью статистического моделирования показано, что данный алгоритм при наилучшем в среднем ρ_0 близок к имитированному визуальному (интерактивному) прогнозу и что это пороговое значение близко к наилучшему в среднем ρ_0 для аппроксимации. Это позволяет использовать при прогнозе уже сформулированные ранее для аппроксимации методы выбора ρ_0.

Четвертая глава посвящена задаче оценки коэффициентов линейной рекуррентной формулы (ЛРФ) сигнала ряда, где ряд $F = (f_0, \ldots, f_{N-1})$ соответствует модели “сигнал плюс шум”. Исследу-

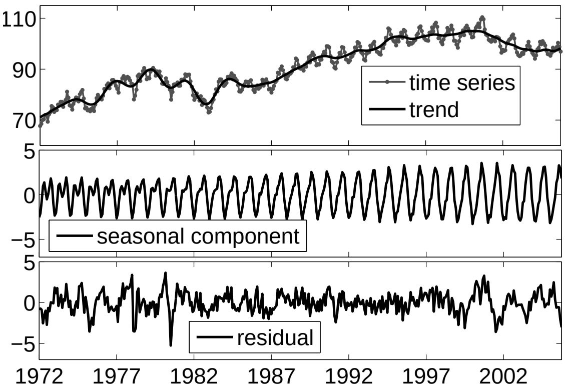

Рис. 2: Ежемесячный уровень потребления электроэнергии в США, тренд выделен с помощью TrRmeas, сезонная составляющая – с помощью PerEmp с параметрами $L = 198$, $A = 1$

ется случай, когда сигнал представляет собой гармонику:

$$f_n = s_n + \varepsilon_n, \quad s_n = A\sin(2\pi\omega n + \phi),$$

которой в рекуррентной записи соответствует ЛРФ порядка 2 (то есть $s_n = a_1 s_{n-1} + a_2 s_{n-2}$), а остаток $(\varepsilon_0, \ldots, \varepsilon_{N-1})$ – реализация белого нормального шума, каждый элемент которого распределён по $N(0, \sigma^2)$. Ставится задача оценки коэффициентов a_1, a_2 и через них частоты ω по (5).

В разделе 4.1 описаны два метода оценки a_1, a_2. Первый, SsaLrf, основан на подходе "Гусеница"-SSA, с помощью которого находится ЛРФ порядка 2 для сигнала. Второй метод, Regr, использует регрессионную запись для f_n и оценивает коэффициенты a_1, a_2 по методу наименьших квадратов.

Методы сравнивались с помощью статистического моделирования рядов F, для которых оценивались a_1 и ω и для каждого метода и параметра рассчитывались: абсолютное отклонение выборочного среднего от настоящего значения и выборочная дисперсия. Исследования проводились при различных значениях N, σ, причём при фиксированном значении одного из них другой изменялся с заданным шагом.

Метод SsaLrf в достаточно большой области параметров показал результаты лучшие, чем метод Regr (при не слишком больших σ или при достаточно большом N), что видно на рис. 3, на котором изображены результаты при $N = 60$, $\omega = 1/6$, $A = 1$, $\phi = 0$, оцененные на 10^4 реализациях.

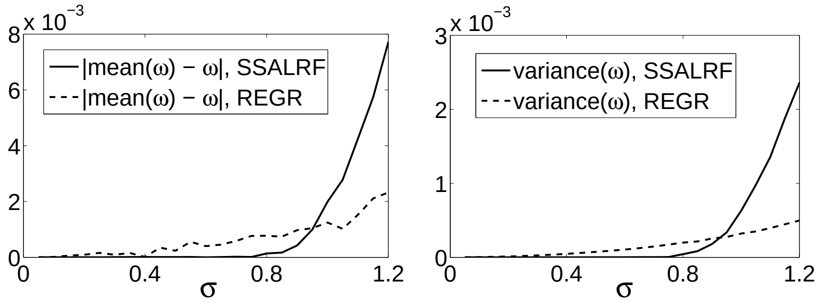

Рис. 3: Зависимость от стандарта шума σ характеристик оценок ω алгоритмами SsaLrf и Regr

Основные положения диссертации, выносимые на защиту.

- Разработаны математические методы автоматической идентификации собственных троек, соответствующих тренду или периодическим составляющим в рамках подхода "Гусеница"-SSA;
- Созданы и реализованы в виде программного комплекса алгоритмы выделения тренда и периодических составляющих, использующие предложенные методы;
- Рассмотрены способы выбора параметров и пороговых значений алгоритмов;
- Проведено статистическое исследование качества работы алгоритмов при наилучших в среднем пороговых значениях;
- Построена математическая модель применения алгоритма выделения тренда к множеству рядов;
- Представлены результаты обработки временных и пространственных рядов с помощью предложенных алгоритмов;

- Выполнено статистическое сравнение подхода "Гусеница"-SSA и регрессионного метода при оценки коэффициентов ЛРФ порядка 2 для периодического сигнала.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Александров, Ф. И.** Оценка коэффициентов линейной рекуррентной формулы порядка 2, управляющей сигналом / Ф. И. Александров // Мат. модели. Теория и приложения. — 2004. — вып. 5. — С. 50–61. — ISBN 5-9651-0082-5.
2. **Александров, Ф. И.** Выделение аддитивных компонент временного ряда при пакетной обработке методом "Гусеница"-SSA / Ф. И. Александров // Вестник СПбГУ, Сер. 1. Математика. Механика. Астрономия. — 2006. — № 2. — С. 71–74. — ISSN 1025-3106.
3. **Александров, Ф. И.** Автоматизация выделения трендовых и периодических составляющих временного ряда в рамках метода "Гусеница"-SSA / Ф. И. Александров, Н. Э. Голяндина // Exponenta Pro. Математика в приложениях. — 2004. — вып. 3-4. — С. 54–61.
4. **Александров, Ф. И.** Выбор параметров при автоматическом выделении трендовых и периодических составляющих временного ряда в рамках подхода "Гусеница"-SSA / Ф. И. Александров, Н. Э. Голяндина, IV Международная конференция "Идентификация систем и задачи управления" SICPRO'05, Москва, январь 2005 // Труды IV Международной конференции "Идентификация систем и задачи управления". – М.: 2005. — С. 1849–1864. — ISBN 5-201-14975-8.
5. **Alexandrov, F.** Automatic extraction and forecast of time series cyclic components within the framework of SSA / F. Alexandrov, N. Golyandina, 5th St.Petersburg Workshop on Simulation, St.-Petersburg, June 2005 // Proc. of the 5th St.Petersburg Workshop on Simulation. — SPb.: 2005. — p. 45–50. — ISBN 5-9651-0102-3.