

13. Engel, J. Conservatism, accuracy and efficiency: Comparing Value at Risk methods. Discussion paper 2. [Text] / J. Engel, M. Gizicki. – Australia: Australian Prudential Regulation Authority, Reserve Bank of Australia, 1999. – 64 p.
14. Longerstae, J. RiskMetrics technical document [Text] / J. Longerstae, M. Spenser. – 4th ed. – New York: J.P. Morgan/Reuters, 1996. – 281 p.
15. Лушавин, А. П. Обработка временных рядов параметров технологических процессов [Текст] / А. П. Лушавин. – LAP LAMBERT Academic Publishing, 2012. – 160 с.
16. Лоскутов, А. Ю. Применение метода локальной аппроксимации для прогноза экономических показателей [Текст] / А. Ю. Лоскутов, Д. И. Журавлев, О. Л. Котляров // Вопросы анализа и управления риском. – 2003. – Т. 1, № 1. – С. 21-31.
17. Чистякова, А. А. Информационная технология прогнозирования нестационарных временных рядов с использованием сингулярного спектрального анализа [Текст] / А. А. Чистякова, Б. В. Шамша // Восточно-Европейский журнал передовых технологий. – 2014. – № 2/4 (68). – С. 24-30. – Режим доступа: \www/URL: <http://journals.uran.ua/ejet/article/view/22158>.

В роботі запропоновано гібридні математичні моделі різної складності для прогнозування часових рядів і методи їх структурної ідентифікації. Методи прогнозування, що було розроблено, засновані на спільному використанні методів «Гусениця»-SSA та Бокса-Дженкінса, а гібридні математичні моделі, які синтезовано на їх основі, є пріоритетними на сьогоднішній день імовірно-детермінованими декомпозиційними моделями. Експериментальні результати показують ефективність запропонованих методів прогнозування

Ключові слова: прогнозування часових рядів, структурна ідентифікація моделі, декомпозиційна модель, метод Бокса-Дженкінса, метод «Гусениця»-SSA

В работе предложены гибридные математические модели различной сложности для прогнозирования временных рядов и методы их структурной идентификации. Разработанные методы прогнозирования основаны на совместном использовании методов «Гусеница»-SSA и Бокса-Дженкинса, а синтезированные на их основе гибридные математические модели являются приоритетными на сегодняшний день вероятностно-детерминированными декомпозиционными моделями. Экспериментальные результаты показывают эффективность предложенных методов прогнозирования

Ключевые слова: прогнозирование временных рядов, структурная идентификация модели, декомпозиционная модель, метод Бокса-Дженкинса, метод «Гусеница»-SSA

УДК 519.67

DOI: 10.15587/1729-4061.2014.28172

ГИБРИДНЫЕ МОДЕЛИ И МЕТОДЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МЕТОДОВ «ГУСЕНИЦА»-SSA И БОКСА-ДЖЕНКИНСА

В. Н. Щелкалин

Инженер 1-й категории

Кафедра прикладной математики

Харьковский национальный

университет радиозлектроники

пр. Ленина, 14, г. Харьков, Украина, 61166

E-mail: vitalii.shchelkalin@gmail.com

1. Введение

К настоящему моменту времени создано большое количество математических моделей и методов анализа и прогнозирования временных рядов (ВР) и наметилась тенденция их комбинирования с целью получения лучших характеристик комбинированной модели. А также прослеживается тенденция перевода теоретических наработок одних методов на другие, где это представляется возможным и целесообразным.

В задачах моделирования и прогнозирования ВР различной природы в 70–90-е годы прошлого века получили широкое применение статистические модели. Популярность моделей данного типа объясняется достаточно высокой степенью следующих показателей:

экономность по количеству оцениваемых параметров; высокая скорость процедуры идентификации модели; малая трудоёмкость и ресурсоемкость реализации при использовании на доступных тогда ЭВМ малой производительности; наилучшие результаты прогнозирования процессов в различных областях науки и техники в те времена; математическая обоснованность теории; адекватность для решения целого ряда задач теории и практики управления, контроля, моделирования и прогнозирования процессов в различных областях человеческой деятельности. Модель авторегрессии – проинтегрированного скользящего среднего (АРИСС, на англ. – ARIMA) является наиболее распространённой из классических статистических моделей прогнозирования.

Классические методы прогнозирования обычно предназначены для работы с линейными и стационарными ВР. Однако большинство естественных материальных процессов реальных физических систем являются нелинейными и нестационарными. Поэтому далее наметилась тенденция критического отношения к статистической постановке проблемы идентификации моделей ВР, особенно в случае, когда отсутствует возможность получения представительных выборок для построения математических моделей, статистических характеристик процессов и проверки их адекватности. Кроме того, статистическая теория использует операции осреднения по множеству реализаций, что в целом ряде случаев приводит к ухудшению математической модели, особенно в условиях малых и нестационарных выборок [1, 2]. В этих случаях эффективно использовать алгебраический, детерминированный, а не статистический подход к решению проблемы идентификации. И лишь в последние десятилетия достаточно активно начали развиваться методы прогнозирования линейных, но нестационарных ВР и нелинейных, но стационарных ВР. И относительно недавно начали появляться гибридные модели прогнозирования, синтезируемые достаточно сложными комбинациями методов, учитывающих как детерминированные, так и стохастические составляющие процессов.

Таким образом, существуют три основные группы моделей и методов моделирования ВР:

- 1) детерминированные (алгебраические);
- 2) вероятностные (статистические);
- 3) комбинированные вероятностно-детерминированные.

Актуальным представляется проведение синтеза классов математических моделей, основанных на совместном использовании идей двух достаточно хорошо теоретически обоснованных методов: детерминированного метода «Гусеница»-SSA и статистического метода Бокса-Дженкинса. Это связано с очевидными преимуществами метода «Гусеница»-SSA, среди которых можно выделить следующие: данным методом можно разложить ВР на интерпретируемые аддитивные составляющие, не требуя при этом стационарности ряда и знания модели тренда, а также сведений о наличии в ряде периодических составляющих и их периодах. Модели метода Бокса-Дженкинса включают в себе все приведенные выше преимущества статистических моделей прогнозирования. Однако данные методы имеют и ряд недостатков. Так, например, недостатками метода Бокса-Дженкинса можно считать заниженное количество оцениваемых параметров, используемых в модели, искажение долгосрочной зависимости ВР и недостаточное количество публикаций со сравнительным анализом прогнозирования данным методом. Среди недостатков же метода «Гусеница»-SSA можно выделить использование неоптимального с точки зрения точности воспроизведения некоторых ВР ортогонального базиса векторов траекторной матрицы ВР, существенно неавтоматическая группировка компонент сингулярного разложения траекторной матрицы ряда для получения составляющих исходного ряда, отсутствие модели не позволяет проверять гипотезы о наличии в ряде той или иной составляющей. Данный метод является непараметрическим. Для проверки подобных гипотез требуется построение модели, ко-

торое в свою очередь, может быть проведено на основе информации, получаемой с помощью метода «Гусеница»-SSA. Этот непараметрический метод позволяет получить результаты, часто лишь незначительно менее точные, чем многие параметрические методы при анализе ряда с известной моделью. Совместное же использование данных методов приводит к синергии, взаимно компенсируя противоположные по природе недостатки входящих в него моделей.

В статье приведены модели трендового подхода к прогнозированию ВР. Трендовый подход заключается в моделировании процесса как отклонения фактических значений ВР относительно трендовой составляющей, в роли которой в предлагаемой модели выступает линейная рекуррентная формула (ЛРФ) метода «Гусеница»-SSA.

Однако реальные ВР в различных отраслях науки и техники являются сложными и обычно порождаются двумя или более процессами, которые могут изменяться во времени. Каждый из этих процессов может быть описан различными моделями. Качественными признаками сложности являются нелинейность процесса, неопределённость его математической модели, наличие внешних возмущений. В результате перечисленных свойств построенная традиционными методами математическая модель сложного процесса, как единого и неделимого, является практически не реализуемой и малоприменимой для использования в задачах прогноза. Поэтому также в данной статье предлагаются модели и методы декомпозиционного подхода к прогнозированию, основанные на использовании методов «Гусеница»-SSA и Бокса-Дженкинса. Основной целью декомпозиционного подхода к прогнозированию является разделение исходного ВР на множество ВР, с более простой структурой, рассматриваемых независимо друг от друга. При этом декомпозиционная модель строится следующим образом:

- осуществление декомпозиции, т. е. непосредственно разложение исходного ВР;
- построение отдельных упрощённых, как правило, независимых моделей, соответствующих компонентам разложения;
- реализация общей декомпозиционной модели процесса как объединения построенных упрощённых моделей.

В статье производится сравнительный анализ предложенных математических моделей и методов прогнозирования на ВР потребления электроэнергии и природного газа.

2. Анализ литературных данных и постановка проблемы

В рассмотренной литературе [3–34] многочисленные исследования направлены на анализ математических моделей и методов среднесрочного и долгосрочного прогнозирования ВР в различных областях науки и техники. В [2] представлены аддитивные, мультипликативные и смешанные модели прогнозирования электропотребления. В [1] рассмотрены различные комбинированные математические модели для прогнозирования трендовой, сезонной, недельной, остаточной составляющих процессов потребления элект-

троэнергии и предложены декомпозиционный метод моделирования (ДММ) и метод окаймляющих функций. Среди предлагаемых моделей прогнозирования отдельных составляющих процесса используются экспоненциальное сглаживание, модель Хольта-Винтерса, двойного экспоненциального сглаживания, фильтр Калмана, ряды Фурье, модели спектрального анализа, экспоненциально взвешенного скользящего среднего и др. Для оценивания эффективности полученных прогнозов модели обычно сравнивались с моделями ARIMA [4–7].

BP потребления целевых продуктов в системах энергетики являются существенно нелинейными со средне- и долгосрочными зависимостями, поэтому аппроксимация таких сложных BP только моделью ARIMA не всегда является удовлетворительной. Было доказано, что для повышения точности прогнозирования нестационарных BP необходимо эффективно извлечь трендовую, гармонические и шумовую составляющие и далее подставлять их в модели прогнозирования. Таким образом, методы декомпозиции BP являются благоприятными для первичной обработки данных.

Сингулярный спектральный анализ (SSA) является одним из наиболее эффективных методов для анализа BP со сложными периодическими компонентами. Зарождение SSA обычно ассоциируется с публикацией работы [8], а также метод SSA независимо разработан в России и в нескольких группах в Великобритании и США. SSA стал широко используемым методом анализа периодических и трендовых составляющих BP [9–12]. В последнее время были разработаны новые модели для прогнозирования, основанные на SSA. В [13] разработана модель на основе сингулярного спектрального анализа и стандартной искусственной нейронной сети. В [14] создана гибридная модель прогнозирования BP на основе метода опорных векторов (SVM) в сочетании с SSA. В [15] предложили линейную рекуррентную формулу сингулярного спектрального анализа (SSA-LRF), и эта модель была использована для решения некоторых практических задач [16, 17].

В последнее время появилось достаточно большое количество работ, посвященных и другим гибридным декомпозиционным моделям и методам прогнозирования. В [18] SVM комбинировалась с вейвлет-анализом. В [19] предложен двухэтапный адаптивный подход для прогнозирования BP потребления электроэнергии. В этом подходе на первом этапе методом EMD (Empirical Mode Decomposition) производится декомпозиция прогнозируемого BP на собственные модальные функции и применяется к ним преобразование Гильберта. На втором этапе, полученные модальные функции и их мгновенные амплитуды подаются на вход прогнозной нейронной сети. В [20] разработан комбинированный метод прогнозирования тоже основанный на разложении Хуанга и искусственной нейронной сети. В [21] был предложен подобный метод, но в качестве метода разложения использовалось разложение на независимые компоненты ICA (independent component analysis). В [22] комбинируются модели SVR (support vector regression) и ARIMA. Достаточно много статей посвящено гибридным моделям, сочетающим нейронные сети с моделями ARIMA [23–26]. Гибридные модели

на основе SVM и ARIMA рассмотрены в [27–29]. В [30] EMD комбинируется с моделями ARIMA. В [31] ARIMA комбинируется с вейвлетами.

В данной статье разрабатываются гибридные математические модели и методы на основе методов «Гусеница»-SSA и Бокса-Дженкинса. В [32] предложена комбинированная аддитивная модель прогнозирования, в которой к ЛРФ добавлялась модель авторегрессии – проинтегрированного скользящего среднего. Исследования комбинирующие метод «Гусеница»-SSA с другими моделями можно встретить и в зарубежной литературе. Так, например, в [33] была разработана подобная модель, в которой ЛРФ комбинировалась с моделью авторегрессии и применялась такая гибридная модель для краткосрочного прогнозирования нагрузки иранской национальной энергосистемы. Другой примером совместного использования метода «Гусеница»-SSA с моделями Бокса-Дженкинса можно найти в [3]. Там методом «Гусеница»-SSA производилось разложение исходного BP, затем производилась группировка BP разложения на интерпретируемые составляющие, такие как трендовая, сезонные и шумовые. Далее для каждого сгруппированного BP идентифицировалась и оценивалась своя модель ARIMA. Остаточная ошибка моделирования корректировалась авторегрессионной моделью и находились прогнозы синтезированной моделью. Примерно в то же время в [34] была предложена подобная модель, однако, в отличие от [3], для идентификации и оценивания сгруппированных BP использовались сезонные модели авторегрессии – проинтегрированного скользящего среднего (САРПСС, на англ. – SARIMA), а также такая гибридная модель была обобщена уже для случая прогнозирования BP, зависящего от нескольких экзогенных BP.

Таким образом, анализ рассмотренной литературы позволяет сделать вывод, что в настоящее время происходит отход от статистической постановки задачи прогнозирования и эффективное прогнозирование нестационарных BP предполагает использование различных приёмов декомпозиции, синтеза прогнозных моделей и отбора их из множества альтернатив.

3. Цели и задачи исследования

Целью проведенных исследований является разработка перспективных по мнению автора декомпозиционных вероятностно-детерминированных математических моделей и методов их идентификации для прогнозирования нестационарных BP со сложной структурой и проверка эффективности прогнозирования разработанными моделями в сравнении с прогнозами, получаемыми вероятностными моделями SARIMA, обобщёнными для прогнозирования BP с несколькими сезонными составляющими [4].

Для этих целей в работе использованы методы «Гусеница»-SSA и Бокса-Дженкинса и на их основе разработаны гибридные математические модели и методы прогнозирования нестационарных BP. Метод «Гусеница»-SSA применялся для разложения исходного BP на BP с более простой структурой, а метод Бокса-Дженкинса – для построения математической

модели данных компонент разложения. Выбор различных параметров метода «Гусеница»-SSA приводит к созданию в сочетании с методом Бокса-Дженкинса различных вариантов методов прогнозирования и к синтезу на их основе гибридных математических моделей с различными структурами.

4. Математические модели и методы прогнозирования нестационарных временных рядов

В данном разделе сначала приведём короткое изложение двух классических методов прогнозирования нестационарных ВР: детерминированного метода «Гусеница»-SSA и статистического метода Бокса-Дженкинса. Приведенные в них сезонная модель авторегрессии – проинтегрированного скользящего среднего (САРПСС) метода Бокса-Дженкинса и линейная рекуррентная формула (ЛРФ) метода «Гусеница»-SSA необходимы далее для описания гибридных моделей на их основе.

4.1. Классические методы прогнозирования нестационарных временных рядов

4.1.1. Метод «Гусеница»-SSA

Рассмотрим ВР $y_t = (y_1, y_2, \dots, y_n)$ длины n . Основным параметром метода «Гусеница»-SSA является параметр L , $1 < L < n$ и называется длиной окна. Базовый алгоритм метода «Гусеница»-SSA состоит из следующих этапов [35].

1. Разложение ВР.

1.1. Вложение. На начальном этапе исходный ВР y_t , $t = \overline{1, n}$ переводится в последовательность $K = n - L + 1$ многомерных векторов вложения размерности L :

$$Y_i = (y_i, y_{i-1}, \dots, y_{i-L+1})^T, \quad 1 \leq i \leq K.$$

Затем из них строится так называемая траекторная матрица Y следующим образом:

$$Y = [Y_1 \ Y_2 \ \dots \ Y_K] \Rightarrow Y = (y_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} Y_0 & Y_1 & Y_2 & \dots & Y_{K-1} \\ Y_1 & Y_2 & Y_3 & \dots & Y_K \\ Y_2 & Y_3 & Y_4 & \dots & Y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{L-1} & Y_L & Y_{L+1} & \dots & Y_{n-1} \end{pmatrix}.$$

1.2. Сингулярное разложение. К траекторной матрице ВР применяется SVD-разложение:

$$Y = \sum_{i=1}^L \sqrt{\lambda_i} U_i V_i^T = \sum_{i=1}^L Y_i, \tag{1}$$

где $Y_i = \sqrt{\lambda_i} U_i V_i^T$; U_i , и V_i , $i = \overline{1, L}$ – левые и правые сингулярные вектора, соответствующие упорядоченным по убыванию собственным числам λ_i , $i = \overline{1, L}$ матрицы $S = Y Y^T$, $d = \text{rank } Y$.

2. Восстановление ВР.

2.1. Группировка. Всёмножество индексов $\{1, 2, \dots, L\}$ делится на r непересекающихся подмножеств I_1, I_2, \dots, I_r . Для каждого набора индексов $I_k = \{i_1^{(k)}, i_2^{(k)}, \dots, i_{p_k}^{(k)}\}$,

$k = \overline{1, r}$ определяется матрица $Y_{I_k} = \sum_{j=1}^{p_k} Y_{i_j^{(k)}}$. Тогда разложение (1) можно записать следующим образом:

$$Y = \sum_{k=1}^r Y_{I_k}.$$

Процедура выбора множеств I_1, I_2, \dots, I_r и называется группировкой собственных троек.

2.2. Диагональное усреднение. На этом этапе каждая матрица Y_{I_k} , $k = \overline{1, r}$ переводится во ВР $g_t^{(k)} = (g_1^{(k)}, g_2^{(k)}, \dots, g_n^{(k)})$, сопоставляя каждой k -ой «диагонали» матрицы среднее арифметическое значений данной «диагонали».

Применив диагональное усреднение к матрицам Y_{I_k} , $k = \overline{1, r}$ получаем ряды $\tilde{y}_t^{(k)} = (\tilde{y}_1^{(k)}, \tilde{y}_2^{(k)}, \dots, \tilde{y}_n^{(k)})$, $k = \overline{1, r}$. Тогда исходный ряд раскладывается в сумму r рядов:

$$y_t \approx \tilde{y}_t = \sum_{i=1}^r \tilde{y}_t^{(i)}. \tag{2}$$

3. Прогноз.

Обозначим $w_i = \tilde{y}_i$, $i = \overline{1, n}$. Рекуррентный прогноз на l шагов вперед определяется по следующей формуле:

$$w_i = \sum_{j=1}^{L-1} g_j w_{i-j}, \quad i = \overline{n+1, n+1+l}, \tag{3}$$

где коэффициенты g_j , $j = \overline{1, L-1}$ являются элементами вектора $\mathfrak{R} = (g_{L-1}, g_{L-2}, \dots, g_1)^T$, определяемого по формуле:

$$\mathfrak{R} = \frac{1}{1-v^2} \sum_{i=1}^r \pi_i U_i^v,$$

где π_i – последняя компонента вектора U_i , $i = \overline{1, r}$; U_i^v , $i = \overline{1, r}$ – вектор, состоящий из первых $L-1$ компонент вектора U_i ; $v^2 = \pi_1^2 + \pi_2^2 + \dots + \pi_r^2$. Формула (3) называется линейной рекуррентной формулой (ЛРФ).

Имеется и векторный вариант прогнозирования методом «Гусеница»-SSA, который в ряде случаев может давать более точные прогнозы в сравнении с вариантом метода рекуррентного прогнозирования [35].

4.1.2. Модель авторегрессии – проинтегрированного скользящего среднего

В методе Бокса-Дженкинса уделено особое внимание проблеме выбора модели и её оцениванию. В этом методе используются идеи, что нестационарные ВР можно преобразовать в стационарные путём перехода от исходного ВР к его разностям соответствующего порядка $d=1, 2, \dots$. Преобразованный таким образом ВР можно далее описать моделями авторегрессии (АР), скользящего среднего (СС) или смешанной АРСС моделью соответствующих порядков.

Сезонная модель АРПСС (САРПСС, на англ. – SARIMA) в операторной форме может быть представлена в следующем виде [4]:

$$y_t = \frac{\theta_q^*(B)}{\Phi_p^+(B)} a_t, \tag{4}$$

где y_t , $t = \overline{1, n}$ исходный или преобразованный (нормированный или прологарифмированный) центрированный ВР; n – объём выборки; B – опе-

ратор сдвига по времени на одну единицу назад, такой что $B^i x_k = x_{k-i}$; $()$ – обобщенный оператор

авторегрессии порядка $p^* = p^* + \sum_{i=1}^{n_s} D_i S_i$, $p^* = \sum_{i=1}^{n_s} p_i S_i$;

$\Phi_{p^*}^+(B) = \Phi_{p^*}^*(B) \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} \dots \nabla_{S_{n_s}}^{D_{n_s}}$; D_i , $i = \overline{1, n_s}$ – порядок

взятия разности S_i ; S_i , $i = \overline{1, n_s}$ – период i -й периодической компоненты, причем $S_1 = 1$; n_s – количество периодических компонент; ∇_{S_i} и B^{S_i} – упрощающие операторы такие, что $\nabla_{S_i} x_t = (1 - B^{S_i}) x_t = x_t - x_{t-S_i}$.

$\Phi_{p^*}^*(B)$ – обобщенный оператор авторегрессии порядка p^* вида $\Phi_{p^*}^*(B) = \prod_{i=1}^{n_s} \Phi_{p_i}^i(B^{S_i})$; $\Phi_{p_i}^i(B^{S_i})$, $i = \overline{1, n_s}$ –

полиномы от B^{S_i} степеней p_i соответственно, определяющие составляющие авторегрессии периодических компонент с периодами S_i соответственно;

$\theta_{q^*}^*(B)$ – обобщенный оператор скользящего среднего порядка $q^* = \sum_{i=1}^{n_s} q_i S_i$ вида $\theta_{q^*}^*(B) = \prod_{i=0}^{n_s} \theta_{q_i}^i(B^{S_i})$;

$\theta_{q_i}^i(B^{S_i})$, $i = \overline{1, n_s}$ – полиномы от B^{S_i} степеней q_i соответственно, определяющие составляющие скользящего среднего периодических компонент с периодами S_i соответственно;

a_t – случайный процесс типа белый шум.

Идентификация параметров p_i , $i = \overline{1, n_s}$ осуществляется путём анализа частной автокорреляционной функции (ЧАКФ), а параметров q_i , $i = \overline{1, n_s}$ – путём анализа автокорреляционной функции (АКФ). Параметры n_s , D_i , S_i также идентифицируются путём анализа АКФ и ЧАКФ процесса. Автоматизировать процесс структурной идентификации таких моделей можно перебором моделей с различной структурой и отбором наилучших по информационному критерию Акаике (AIC) и байесовскому информационному критерию Шварца (BIC).

4. 1. 2. 1. Алгоритм обобщённого метода Бокса-Дженкинса

Блок-схема алгоритма, обобщённого на случай нескольких сезонных составляющих ВР, метода Бокса-Дженкинса (рис. 1) состоит из следующих этапов [4]:

- 1) начало;
- 2) ввод обучающей выборки значений исходного ВР y_t в дискретные моменты времени $t = \overline{1, n}$;
- 3) идентификация порядков взятия разностей и преобразование ряда. На основании анализа вида автокорреляционной функции процесса y_t определяем порядок взятия разности D_{n_s} с периодом $S_{n_s} = 1$. Определяем периоды периодических компонент S_i с убывающими амплитудами. На основании анализа автокорреляционных функций процессов

$$y'_t = \nabla_{S_1}^{D_1} y_t; y''_t = \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} y_t; y'''_t = \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} \dots \nabla_{S_{n_s-1}}^{D_{n_s-1}} y_t,$$

определяем порядки взятия разностей D_i , $i = \overline{1, n_s - 1}$. По известным значениям D_i , S_i преобразуем исходную выборку y_t в

$$\omega_t = \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} \dots \nabla_{S_{n_s}}^{D_{n_s}} y_t,$$

- 4) идентификация параметров модели p_i , q_i , $i = \overline{1, n_s}$. Осуществляется по виду автокорреляционной и частной автокорреляционной или автоматически перебором моделей с различной структурой и отбором наилучших по информационному критерию Акаике (AIC) и байесовскому критерию Шварца (BIC);
- 5) грубая оценка параметров модели. Как показали экспериментальные исследования сходимости алгоритма параметрической оптимизации, для правильно идентифицированной модели сходимость этого алгоритма обеспечивается при задании в качестве грубых оценок параметров модели констант типа ± 0.1 ;
- 6) точная оценка параметров модели. Осуществляется в соответствии с алгоритмом параметрической оптимизации. В данной работе использовался метод Левенберга-Марквардта. В качестве начального приближения итерационной процедуры вначале используются грубые оценки параметров модели, в дальнейшем – точные оценки параметров модели, полученные для предыдущей обучающей выборки;
- 7) проверка адекватности полученной модели. Осуществляется в соответствии с критерием

$$Q = M \sum_{k=1}^n r_{aa}(k), \tag{5}$$

где $r_{aa}(k)$ – корреляционная функция остаточных ошибок a_t модели.

Сравнив значение данного критерия с табличными значениями χ^2 при заданном уровне значимости, приближенно проверяется гипотеза об адекватности полученной модели. Если полученная модель неадекватна исходным данным, то осуществляется переход к четвертому этапу. Если модель адекватна – к следующему этапу;

- 8) вычисление прогноза $\hat{y}_t(l)$, $l = \overline{1, L}$ с заданным упреждением L . Осуществляется по разностному уравнению

$$\hat{y}_t(l) = \theta_0 + \sum_{i=1}^{p^*} \Phi_i^+ [y_{t+l-i}] - \sum_{j=1}^{q^*} \theta_j^* [a_{t+l-j}],$$

где θ_0 – общая константа, позволяющая учесть детерминированный тренд. Условные математические ожидания величин y и a в моменты времени $t+l-i$ и $t+l-j$ соответственно равны

$$[y_{t+l-i}] = \begin{cases} \hat{y}_t(l-i), & l > i; \\ y_{t+l-i}, & l \leq i, \end{cases}$$

$$[a_{t+l-j}] = \begin{cases} 0, & l > j; \\ y_{t+l-j} - \hat{y}_{t+l-j}(l), & l \leq j, \end{cases}$$

где $\hat{y}_t(l-i)$ – прогноз величины y_t , вычисленной в момент времени t с упреждением $(l-i)$; Φ_i^+ , $i = \overline{1, p^*}$ и θ_j^* , $j = \overline{1, q^*}$ – коэффициенты полиномов $\Phi_{p^*}^+(B)$ и $\theta_{q^*}^*(B)$ соответственно.

Дисперсия прогноза $V(l)$, $l = \overline{1, L}$, а также его вероятностные пределы находим соответственно по уравнениям

$$V(l) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2,$$

$$y_{t+i}(\pm) = \hat{y}_t(l) \pm U_{\epsilon/2} \sqrt{V(l)},$$

где $U_{\epsilon/2}$ – квантиль уровня $1 - \epsilon/2$ стандартного нормального распределения. Веса ψ_i определяются по формуле

$$\psi_i = \begin{cases} 1, & i=1; \\ \sum_{j=1}^i \Phi_j^+ \psi_{j-1} - Q_j^*, & i > 1, \end{cases}$$

где $\Phi_i^+ = 0$, $i > p^+$; $Q_j^* = 0$, $j > q^*$.

- 9) печать прогноза $\hat{y}_t(l)$, $l = \overline{1, L}$;
- 10) ввод фактического значения y_{t+i} , $i = \overline{1, L}$;
- 11) вычисление остаточной ошибки. При вводе в систему каждого нового значения y_{t+i} , $i = \overline{1, L}$ определяется остаточная ошибка модели $a_{t+i} = y_{t+i} - \hat{y}_{t+i-1}(1)$;
- 12) проверка условия: если $i < L$, то осуществляется переход к следующему этапу, если нет – к 15-му этапу;
- 13) коррекция прогнозов. Используя веса ψ_t и остаточную ошибку модели a_{t+i} в момент времени $t+i$, производим коррекцию прогнозов $\hat{y}_t(k)$, $k = i+1, L$ в соответствии с уравнением $\hat{y}_{t+1}(1) = \hat{y}_t(1) + \psi_1 a_{t+1}$;
- 14) печать скорректированных прогнозов и возврат к 10-му этапу;
- 15) проверка условия продолжения вычислений. Если условие выполняется, то осуществляется переход к следующему этапу, если нет – к 17-му;
- 16) формирование новой обучающей выборки и нового ряда остаточных ошибок модели. Осуществляется путем отбрасывания L первых и прибавления L новых реализаций процесса и фактических ошибок прогнозов с

упреждением в единицу. Здесь же выполняем необходимые преобразования новой обучающей выборки y_{t+i} в

$$\omega_{t+i} = \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} \dots \nabla_{S_{ns}}^{D_{ns}} y_{t+i};$$

17) проверка адекватности модели.

Для ряда остаточных ошибок модели, соответствующих новой обучающей выборке, вычисляем значение критерия (5) и проверяем адекватность модели. Если модель адекватна, переходим к 8-му этапу. В противном случае, предполагая, что структура модели не изменилась, а изменились только значения параметров модели, переходим к 6-му этапу. Если новые данные привели к изменению стохастической структуры модели, это будет обнаружено на 7-м этапе (рис. 1).

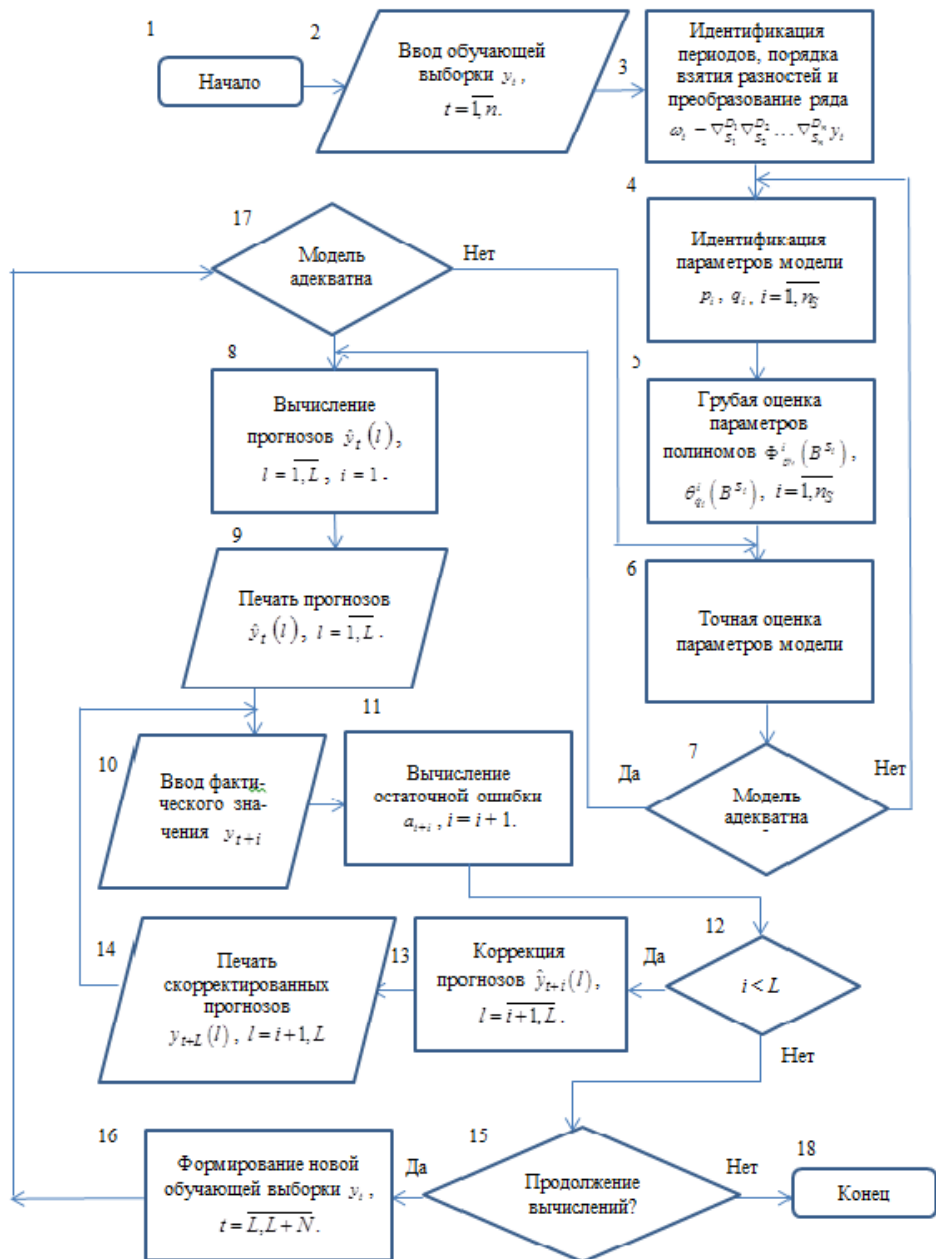


Рис. 1. Блок-схема алгоритма обобщённого метода Бокса-Дженкинса

Этот алгоритм обеспечил высокую точность прогнозов процессов потребления целевых продуктов в системах энергетики такими потребителями, для которых преобладающее влияние оказывают хронологические факторы [4].

4. 2. Гибридные математические модели и методы прогнозирования нестационарных временных рядов на основе методов «Гусеница»-SSA и Бокса-Дженкинса

В данном разделе рассмотрены трендовый и декомпозиционный подходы к прогнозированию ВР. Для каждого из подходов предложены соответствующие математические модели и методы их построения, основанные на совместном использовании двух методов: детерминированного метода «Гусеница»-SSA и статистического метода Бокса-Дженкинса.

4. 2. 1. Трендовый подход к прогнозированию нестационарных временных рядов

В общем случае ВР y_t ($t=1, 2, \dots, n$) может быть разложен на трендовую составляющую y_t^T (включая её разновидности – сезонную или гармоническую) и случайную составляющую ϵ_t :

$$y_t = y_t^T + \epsilon_t.$$

Очевидно, необходимо прогнозировать как трендовую, так и случайную составляющие.

Трендовый подход состоит в моделировании ВР как отклонения фактических значений ВР относительно его трендовой составляющей. Для прогнозирования трендовой составляющей наиболее широкое распространение получили следующие методы: метод наименьших квадратов (МНК), взвешенный МНК, аппроксимация полиномами, метод гармонических весов, метод экспоненциального сглаживания, адаптивного сглаживания, метод построения сплайн-функций, компенсация трендовой составляющей на основе вычислений разностей n -го порядка, методы скользящего среднего, методы оптимальной фильтрации, методы моделирования конечными рядами Фурье, спектральные методы, методы декомпозиции по ортогональным векторам или функциям, методы имитационного моделирования, методы нейро-сетевого и нечёткого моделирования, методы вариационного исчисления, метод группового учёта аргументов, метод «Гусеница»-SSA [1, 2] и др. Для прогнозирования же случайной составляющей используются методы вероятностного моделирования, регрессионные статистические модели, метод Цепей Маркова, модели экспоненциального сглаживания Брауна, взвешенного среднего, Хольта-Уинтерса, обобщённый фильтр Винера (метод прогнозирования Заде-Рагазини), спектральные статистические модели (канонические разложения, разложение Карунена-Лоэва, модели компонентного и факторного анализов), прогнозирующий фильтр Калмана-Бьюси, модели кластерного анализа и теории распознавания образов [1, 2] и др.

4. 2. 1. 1. Гибридные математические модели трендового подхода к прогнозированию нестационарных временных рядов на основе методов «Гусеница»-SSA и Бокса-Дженкинса

В настоящее время всё большее распространение начинают получать математические модели прогнози-

рования, являющиеся комбинацией статистических и детерминированных моделей.

Вернёмся к подпункту 1 этапа восстановления ВР метода «Гусеница»-SSA возьмём параметр $g=1$, т. е. из множества индексов $\{1, 2, \dots, L\}$ отберём некоторые или все индексы в подмножество $I_T \subseteq \{1, 2, \dots, L\}$. Тогда на этапе диагонального усреднения метода «Гусеница»-SSA исходному ВР будет поставлен в соответствие некоторый отфильтрованный (сглаженный) ряд \hat{y}_t^T . Т. е., согласно (2), имеем $y_t \approx \hat{y}_t^T$. Вычислим остаточный ВР \hat{y}_t^a как $\hat{y}_t^a = y_t - \hat{y}_t^T$. Таким образом, исходный ВР будет разложен на следующую сумму:

$$y_t = \hat{y}_t^T + \hat{y}_t^a.$$

Данное разложение нужно стараться производить так, чтобы ВР \hat{y}_t^T соответствовал тренд-сезонной составляющей исходного ВР, а остаточный ВР \hat{y}_t^a – шумовой. Это производится путём анализа собственных значений и собственных векторов траекторной матрицы [35]. При этом могут возникнуть трудности со слабой разделимостью некоторых ВР [35].

Вспомним про обозначение $w_t = \hat{y}_t^T$, $t=1, n$ и, для удобства записи, введём обозначение: $a_t = \hat{y}_t^a$.

В основе предлагаемой математической модели трендового подхода к прогнозированию ВР, основанного на совместном использовании методов «Гусеница»-SSA и Бокса-Дженкинса, лежит идея добавления к ЛРФ, вычисляемой методом «Гусеница»-SSA, модели АРПСС, определяемой методом Бокса-Дженкинса, построенной для остаточных ошибок a_t модели, определяемых путём вычитания из исходного ВР значе-

ний w_t . Введём обозначение $g(B) = \sum_{i=1}^{L-1} g_i B^i$, где $g(B)$ – полином от введенного выше оператора задержки B , коэффициенты g_i , $i=1, L-1$ которого равны коэффициентам ЛРФ $w_t = \sum_{i=1}^{L-1} g_i w_{t-i}$ метода «Гусеница»-SSA.

Если тренд-сезонную компоненту и остаточные ошибки прогнозировать отдельно, тогда описанная выше модель будет иметь вид (6), иначе – вид (7):

$$\epsilon_t = y_t - w_t, t = \overline{1, n};$$

$$\hat{\epsilon}_t = \frac{\theta_q(B)}{\nabla \phi_p(B)} a_t;$$

$$w_t = g(B) w_t;$$

$$\hat{y}_t(1) = \hat{w}_t(1) + \hat{\epsilon}_t(1), \tag{6}$$

где $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ – оператор авторегрессии порядка p ; $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ – оператор скользящего среднего порядка q ; $\hat{y}_t(1)$, $\hat{w}_t(1)$ и $\hat{\epsilon}_t(1)$ – прогнозы ВР y_t , w_t и ϵ_t соответственно в момент времени $t+1$. Прогнозы $\hat{w}_t(1)$ и $\hat{\epsilon}_t(1)$ вычисляются путём перехода от операторных уравнений

$$\epsilon_t = \frac{\theta_q(B)}{\nabla \phi_p(B)} a_t \text{ и } w_t = g(B) w_t \text{ к разностным:}$$

$$\hat{\varepsilon}_t(l) = \theta_0 + (\phi_1 + 1)[\varepsilon_{t+1-l}] + (\phi_2 - \phi_1)[\varepsilon_{t+2-l}] + (\phi_3 - \phi_2) \times [\varepsilon_{t+3-l}] + \dots + (\phi_p - \phi_{p-1})[\varepsilon_{t+l-p}] + \phi_p [\varepsilon_{t+l-(p+1)}] - \sum_{i=1}^q \theta_i [a_{t+l-i}]$$

где

$$\varepsilon_{t+l-i} = \begin{cases} \hat{y}_t(1-i), l > i, \\ y_{t+l-i}, l \leq i, \end{cases}$$

$$a_{t+l-j} = \begin{cases} 0, l > j, \\ y_{t+l-j} - \hat{y}_{t+l-j-1}(1), l \leq j \end{cases}$$

и

$$\hat{w}_t(l) = \sum_{i=1}^{l-1} g_i [w_{t+l-i}],$$

где

$$[w_{t+l-i}] = \begin{cases} \hat{w}_t(1-i), l > i, \\ w_{t+l-i}, l \leq i. \end{cases}$$

Запишем более общую модель:

$$y_t = g(B)w_t + \frac{\theta_q(B)}{\nabla\phi_p(B)}a_t, \tag{7}$$

где параметр L – длина окна метода «Гусеница»-SSA; $\phi_p(B)$ – оператор авторегрессии порядка p; $\theta_q(B)$ – оператор скользящего среднего порядка q; a_t – случайный процесс типа белый шум.

Оператор $g(B)$ может применяться непосредственно к процессу y_t , тогда коэффициенты полиномов $\theta_q(B)$ и $\phi_p(B)$ и ВР a_t изменятся, а модель запишется следующим образом:

$$y_t = g(B)y_t + \frac{\theta'_q(B)}{\nabla\phi'_p(B)}a'_t. \tag{8}$$

Если теперь ввести обозначение $g'(B) = 1 - g(B)$, то модель (8) примет вид:

$$g'(B)y_t = \frac{\theta'_q(B)}{\nabla\phi'_p(B)}a'_t. \tag{9}$$

Поясним наличие в формулах (7)–(9) оператора ∇ . После применения метода «Гусеница»-SSA остаточный ВР либо сразу становится стационарным, либо – после однократного применения оператора ∇ . Многократное же применение оператора ∇ может привести к появлению кратных корней характеристического полинома $g'(B) = 0$ модели, лежащих на границе устойчивости. Метод «Гусеница»-SSA в такой модели может применяться для выявления скрытых периодичностей во ВР.

Обобщим модели (7)–(9) добавляя к ЛРФ модель SARIMA, определяемую обобщённым методом Бокса-Дженкинса как в разделе 4. 1. 2. 1:

$$y_t = g(B)w_t + \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)}a_t, \tag{10}$$

$$y_t = g(B)y_t + \frac{\theta_{q^*}^{w*}(B)}{\Phi_{p^*}^{w+}(B)}a'_t, \tag{11}$$

$$g'(B)y_t = \frac{\theta_{q^*}^{w*}(B)}{\Phi_{p^*}^{w+}(B)}a'_t. \tag{12}$$

Назовём модель (10) моделью ЛРФ-SSA + SARIMA.

Метод «Гусеница»-SSA описывает достаточно широкий класс ВР, представляющих собой сумму произведений полиномов, экспонент и гармоник, подклассом которых являются ВР вида $w_n = \sum_k \beta_k e^{\alpha_k n} \cos(2\pi\omega_k n + \phi_k)$.

Поэтому выделенная из исследуемого процесса методом «Гусеница»-SSA детерминированная составляющая w_t исходного ВР достаточно хорошо описывается моделями SARIMA. Следовательно, целесообразно рассмотреть следующую модель:

$$w_t = \frac{\theta_q^w(B)}{\nabla\phi_p^w(B)}a_t^w;$$

$$\varepsilon_t = y_t - w_t;$$

$$\varepsilon_t = \frac{\theta_q(B)}{\nabla\phi_p(B)}a_t;$$

$$\hat{y}_t(l) = \hat{w}_t(l) + \hat{\varepsilon}_t(l).$$

или модель

$$y_t = \frac{\theta_{q^*}^w(B)}{\nabla\phi_{p^*}^{w+}(B)}a_t^w + \frac{\theta_q(B)}{\nabla\phi_p(B)}a_t,$$

где a_t^w , a_t – случайные процессы типа белый шум (остаточные ошибки моделей). Или в обобщённом виде эти модели запишутся следующим образом:

$$w_t = \frac{\theta_{q^{w*}}^*(B)}{\Phi_{p^{w+}}^{w+}(B)}a_t^w,$$

$$\varepsilon_t = y_t - w_t,$$

$$\varepsilon_t = \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)}a_t,$$

$$\hat{y}_t(l) = \hat{w}_t(l) + \hat{\varepsilon}_t(l) \tag{13}$$

и

$$y_t = \frac{\theta_{q^{w*}}^*(B)}{\Phi_{p^{w+}}^{w+}(B)}a_t^w + \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)}a_t. \tag{14}$$

Дадим модели (14) обозначение Тренд-SSA-SARIMA и поясним. В этой модели, в отличие от (10), тренд-сезонная составляющая описывается не ЛРФ, а моделью SARIMA.

4. 2. 1. 2 Метод на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса трендового подхода к прогнозированию нестационарных временных рядов

Ниже приведен алгоритм метода на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса трендового подхода к прогнозированию нестационарных ВР, а также его схема (рис. 2).

1. Применяя этапы 1 и 2 метода «Гусеница»-SSA к исходному ВР, определяется ВР $w_t = \hat{y}_t$, интерпретируемый как тренд-сезонная составляющая исходного ВР.

2. Для w_t определяется ЛРФ, управляющая данным рядом, если используем модели (10)–(12) или производится структурная и параметрическая идентификация модели САРПСС (используя алгоритм раздела 4. 1. 2. 1.) для ВР w_t : $w_t = \frac{\theta_{q^{**}}^*(B)}{\Phi_{p^{**}}^*(B)} a_t^w$, если же используется модель (13) или (14).

3. Вычисляется остаточный ВР ϵ_t следующим образом:

$\epsilon_t = y_t - w_t$, если используется модель (10);

$\epsilon_t = y_t - g(B)y_t$ для моделей (11), (12);

$\epsilon_t = y_t - w_t$, если же используется модель (13) или (14).

4. Строится САРПСС модель остаточного ВР ϵ_t :

$\epsilon_t = \frac{\theta_q^*(B)}{\Phi_p^*(B)} a_t^\epsilon$ (используя алгоритм раздела 4. 1. 2. 1).

5. Для w_t и ϵ_t вычисляются прогнозы $\hat{w}_t(l)$ и $\hat{\epsilon}_t(l)$ с заданным упреждением l для модели (13), используя соответствующие модели САРПСС.

6. Синтезируется общий прогноз $\hat{y}_t(l) = \hat{w}_t(l) + \hat{\epsilon}_t(l)$ как сумма частных прогнозов для модели (13). Или вычисляется прогноз по модели (14), предварительно перейдя от операторного к разностному уравнению.

4. 2. 2 Декомпозиционный подход к прогнозированию нестационарных временных рядов

Под термином «декомпозиция» при моделировании сложных ВР понимают разложение исходного ВР y_t на более простые y_t^i , $i=1, N$, но при этом совокупность этих более простых процессов эквивалентна исходному процессу:

$$y_t = \sum_{i=1}^N y_t^i.$$

Рассмотрим декомпозиционный подход в контексте концепции стандартизованного моделирования ВР, которая заключается в моделировании ВР y_t как совокупности детерминированного тренда y_t^T , различных колебательных составляющих $y_t^{S_i}$, $i=1, n_s$, модулированных по амплитуде и частоте и остаточной составляющей ϵ_t :

$$y_t = y_t^T + \sum_{i=1}^{n_s} y_t^{S_i} + \epsilon_t.$$

Декомпозиционный подход, как правило, используют при моделировании сложных ВР, когда использование традиционных подходов резко затруднено. В настоящее время эффективное моделирование процессов подразумевает использование различных приёмов декомпозиции модели, что позволяет повысить точность и адекватность моделирования в случае нелинейных нестационарных процессов, упростить и повысить устойчивость процесса идентификации [1].

На практике наблюдаемые ВР являются агрегированным представлением нескольких процессов, что ограничивает эффективность использования классических статистических моделей и методов для их анализа и прогнозирования. В большинстве случаев ВР, зависимые от этих порождающие процессов, могут быть разделены с применением гребенки фильтров, методов спектрального анализа, восстанавливающих определенную детерминированную часть ВР.

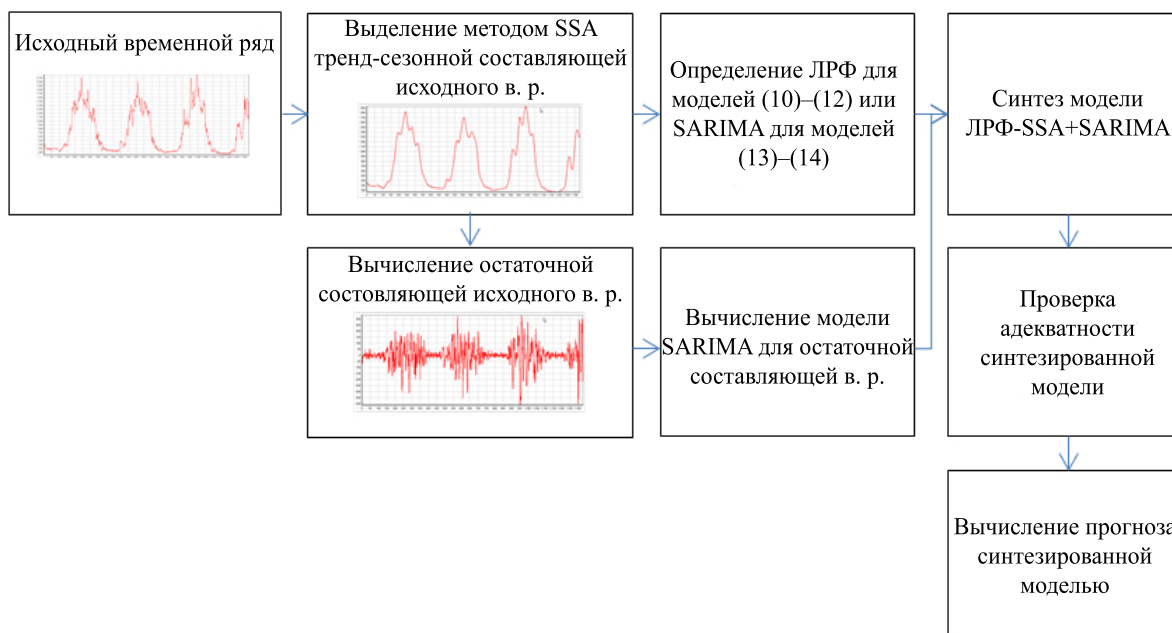


Рис. 2. Схема трендового подхода к прогнозированию нестационарных временных рядов на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса

Затем отфильтрованный ВР может быть спрогнозирован моделью, агрегированной при помощи специальных процедур отбора.

Таким образом, сложность процессов ассоциируется со следующим:

- значительное влияние внешних факторов;
- нелинейность, нестационарность или априорная неопределённость динамического поведения процесса.

4. 2. 2. 1 Гибридные математические модели декомпозиционного подхода к прогнозированию временных рядов на основе методов «Гусеница»-SSA и Бокса-Дженкинса

Для построения адекватных моделей нелинейных и нестационарных ВР необходимо иметь возможность формировать адаптивный базис, функционально зависящий от содержания ВР. Такой подход и реализуется в предлагаемом декомпозиционном подходе к прогнозированию на основе методов «Гусеница»-SSA и Бокса-Дженкинса.

Метод «Гусеница»-SSA используется для поиска скрытых взаимосвязей внутри ВР, являющегося смесью нескольких ВР, не имея никаких априорных знаний о механизме их смешивания. Для этого исходный ВР разлагается на несколько ВР с более простой структурой, применением к нему метода «Гусеница»-SSA. Далее каждый из ВР разложения прогнозируется методом Бокса-Дженкинса. Затем вычисляются прогнозы исходного ВР путем агрегирования прогнозов ВР разложения.

В подпункте 1 этапа восстановления ВР метода «Гусеница»-SSA возьмём параметр $r=L$, т. е. всё множество индексов $\{1, 2, \dots, L\}$ разделим на L подмножеств по одному элементу в каждом следующим образом $I_i = \{i\}$, $i = \overline{1, L}$. Тогда на этапе диагонального усреднения метода «Гусеница»-SSA исходный ВР будет разложен на следующую сумму:

$$y_t = \sum_{i=1}^L \tilde{y}_t^{(i)}.$$

Далее декомпозиционный подход к прогнозированию нестационарных ВР на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса предполагает идентифицировать и строить модели САРПСС каждого ряда разложения $\tilde{y}_t^{(i)}$, $i = \overline{1, L}$.

Таким образом, гибридная модель декомпозиционного подхода к прогнозированию ВР на основе методов «Гусеница»-SSA и Бокса-Дженкинса можно записать в следующем виде:

$$\tilde{y}_t^{(i)} = \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}}, i = \overline{1, L},$$

$$y_t = \sum_{i=1}^L \tilde{y}_t^{(i)},$$

$$\hat{y}_t(l) = \sum_{i=1}^L \hat{\tilde{y}}_t^{(i)}(l). \tag{15}$$

Или

$$y_t = \sum_{i=1}^L \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}}. \tag{16}$$

Теперь рассмотрим данную модель в контексте концепции стандартизованного моделирования нестационарных ВР. В подпункте 1 этап восстановления ВР метода «Гусеница»-SSA оставим без изменения как это описано в разделе 3. 1. 1. Т. е. Всё множество индексов $\{1, 2, \dots, L\}$ делится на r непересекающихся подмножеств I_1, I_2, \dots, I_r . Тогда на этапе диагонального усреднения метода «Гусеница»-SSA исходный ВР будет разложен на следующую сумму:

$$y_t = \sum_{i=1}^r \tilde{y}_t^{(i)}, \tag{17}$$

если в непересекающиеся подмножества I_1, I_2, \dots, I_r вошли все элементы из множества $\{1, 2, \dots, L\}$. Или

$$y_t \approx \sum_{i=1}^r \tilde{y}_t^{(i)} - \tag{18}$$

в противном случае. Причём ВР $\tilde{y}_t^{(i)}$, $i = \overline{1, r}$ могут интерпретироваться как трендовые, квазипериодические или шумовые составляющие исходного ВР.

Тогда гибридная модель прогнозирования нестационарных ВР, основанная на концепции стандартизованного моделирования и на методах «Гусеница»-SSA и Бокса-Дженкинса для случая (17) примет вид

$$\tilde{y}_t^{(i)} = \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}}, i = \overline{1, r},$$

$$y_t = \sum_{i=1}^r \tilde{y}_t^{(i)},$$

$$\hat{y}_t(l) = \sum_{i=1}^r \hat{\tilde{y}}_t^{(i)}(l). \tag{19}$$

Или

$$y_t = \sum_{i=1}^r \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}}. \tag{20}$$

А для случая (18) –

$$\tilde{y}_t^{(i)} = \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}}, i = \overline{1, r},$$

$$\tilde{y}_t = \sum_{i=1}^r \tilde{y}_t^{(i)};$$

$$\varepsilon_t = y_t - \tilde{y}_t,$$

$$\varepsilon_t = \frac{\theta_{q^*}^{\varepsilon}(B)}{\Phi_{p^*}^{\varepsilon}(B)} a_t^{\varepsilon},$$

$$\hat{y}_t(l) = \sum_{i=1}^r \hat{\tilde{y}}_t^{(i)}(l) + \varepsilon_t(l). \tag{21}$$

Или

$$y_t = \sum_{i=1}^r \frac{\theta^{y^{(i)*}}(B)}{\Phi^{y^{(i)+}}(B)} a_t^{y^{(i)}} + \frac{\theta_{q^*}^{\varepsilon}(B)}{\Phi_{p^*}^{\varepsilon}(B)} a_t^{\varepsilon}. \tag{22}$$

Дадим модели (16) обозначение SSA(L)-SARIMA, модели (20) – SSA-SARIMA, а модели (22) – SSA-SARIMA + SARIMA.

Рассмотрим модели (14), (16), (20), (22) и сделаем следующие выводы. Модели Бокса-Дженкинса основываются на гипотезе, что изучаемый процесс является выходом линейного фильтра, на вход которого подан процесс белого шума, т.е. что член ряда y_t является взвешенной суммой текущего и предыдущих значений входного потока:

$$y_t = \theta_0 + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots = \theta_0 + \theta(B)a_t.$$

Если же мы рассмотрим модели (14), (16), (20), (22), то изучаемый процесс в таком случае будет являться уже выходом гребёнки линейных фильтров, на входы которой поданы процессы белого шума.

4. 2. 2. 2 Метод на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса декомпозиционного подхода к прогнозированию нестационарных временных рядов

Приведём алгоритм и схему (рис. 3) метода на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса декомпозиционного подхода к прогнозированию нестационарных ВР:

1. Применяя этапы 1 и 2 метода «Гусеница»-SSA к исходному ВР определяются ВР

$$\tilde{y}_t^{(i)}, i = \overline{1, L} \text{ если используем модель (15) или (16);}$$

$$\tilde{y}_t^{(i)}, i = \overline{1, r} \text{, если используем модель (19) или (20);}$$

$$\tilde{y}_t^{(i)}, i = \overline{1, r}, \varepsilon_t = y_t - \sum_{i=1}^r \tilde{y}_t^{(i)}, \text{ если используется модель}$$

(21) или (22)

с более простой структурой и интерпретируемые как тренды, квазипериодические и шумовые составляющие.

2. Для каждого ВР

$$\tilde{y}_t^{(i)}, i = \overline{1, L} \text{, если используем модель (15) или (16);}$$

$$\tilde{y}_t^{(i)}, i = \overline{1, r} \text{, если используем модель (19) или (20);}$$

$$\tilde{y}_t^{(i)}, i = \overline{1, r}, \varepsilon_t \text{, если используется модель (21) или (22)}$$

производится структурная и параметрическая идентификация моделей САРПСС (используя алгоритм раздела 4. 1. 2. 1)

$$\hat{y}_t^{(i)} = \frac{\theta^{y^{(i)}}(B)}{\Phi^{y^{(i)}}(B)} a_t^{y^{(i)}}, i = \overline{1, L}, \text{ для модели (15) или (16);}$$

$$\hat{y}_t^{(i)} = \frac{\theta^{y^{(i)}}(B)}{\Phi^{y^{(i)}}(B)} a_t^{y^{(i)}}, i = \overline{1, r}, \text{ для модели (19) или (20);}$$

$$\hat{y}_t^{(i)} = \frac{\theta^{y^{(i)}}(B)}{\Phi^{y^{(i)}}(B)} a_t^{y^{(i)}}, i = \overline{1, r}, \varepsilon_t = \frac{\theta_{\varepsilon}^*(B)}{\Phi_{\varepsilon}^+(B)} a_t^{\varepsilon} \text{ для модели}$$

(21) или (22),

используя алгоритм, приведенный в пункте 4. 1. 2. 1.

3. Вычисляются прогнозы с заданным упреждением l

$$\hat{\hat{y}}_t^{(i)}(l), i = \overline{1, L}, \text{ для модели (15);}$$

$$\hat{\hat{y}}_t^{(i)}(l), i = \overline{1, r}, \text{ для модели (19) или прогнозы по модели (20);}$$

$\hat{\hat{y}}_t^{(i)}(l), i = \overline{1, r}, \hat{\hat{\varepsilon}}_t(l)$ для модели (21) или прогнозы по модели (22).

4. Синтезируется общий прогноз как сумма частных прогнозов

$$\hat{y}_t(l) = \sum_{i=1}^L \hat{\hat{y}}_t^{(i)}(l), \text{ для модели (15) или вычисляется}$$

прогноз по модели (16), предварительно приведя её к разностному выражению;

$$\hat{y}_t(l) = \sum_{i=1}^r \hat{\hat{y}}_t^{(i)}(l), \text{ для модели (19) или вычисляется}$$

прогнозы по модели (20), предварительно приведя её к разностному выражению;

$$\hat{y}_t(l) = \sum_{i=1}^r \hat{\hat{y}}_t^{(i)}(l) + \hat{\hat{\varepsilon}}_t(l), \text{ для модели (21) или вычисля-$$

ется прогнозы по модели (22), предварительно приведя её к разностному выражению.

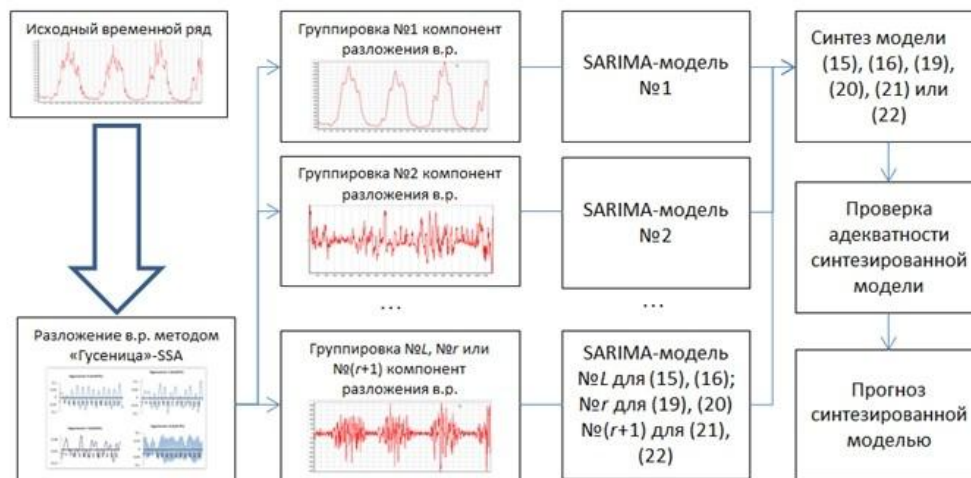


Рис. 3. Схема декомпозиционного подхода к прогнозированию нестационарных временных рядов на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса

5. Результаты исследований

Исследование предлагаемых гибридных моделей на основе методов «Гусеница»-SSA и Бокса-Дженкинса осуществим сопоставляя их результаты прогнозов с результатами прогнозов, получаемых классическими вероятностными моделями SARIMA, обобщёнными на случай нескольких сезонных компонент. Реализация рассмотренных моделей производилась в математическом пакете MATLAB R2014a.

Тестирование будем проводить на трёх ВР потребления целевых продуктов в системах энергетики. Сначала возьмём ВР часовых данных потребления электроэнергии за трое суток (рис. 4).

Модели будем идентифицировать на выборке объёма 50 значений и производить одношаговый прогноз на 1 час вперёд, а далее повторять процедуру скольжением окна вплоть до конца ВР.

На рис. 5 и рис. 6 приведены графики автокорреляционной и частной автокорреляционной функций

соответственно после следующего преобразования исходного ВР $y'_t = \nabla^1 y_t$. Анализ данных графиков совместно с отбором моделей SARIMA по критериям AIC и BIC определили следующую структуру модели:

$$SARIMA(1,1,1) \times (0,0,1)_{24} : y_t = \frac{(1+0.093B)(1+0.083B^{24})}{\nabla_1(1-0.476B)} a_t.$$

Определимся с выбором параметра длины окна метода «Гусеница»-SSA. Для базового алгоритма метода «Гусеница»-SSA чем больше длина окна, тем более детальным получается разложение исходного ВР. Однако, не имеет смысла брать длину окна, большую чем половина длины ряда, т.к. сингулярные разложения траекторной матрицы, соответствующие выбору длины окна L и N-L+1 эквивалентны. Таким образом, наиболее детальное разложение достигается при $L \approx n/2$ за исключением рядов конечного ранга. Маленькая длина окна может привести к смешиванию интерпретируемых компонент ряда [35]. Чтобы достичь лучшей разделительности аддитивных составляющих ВР, нужно выбирать большую длину окна.

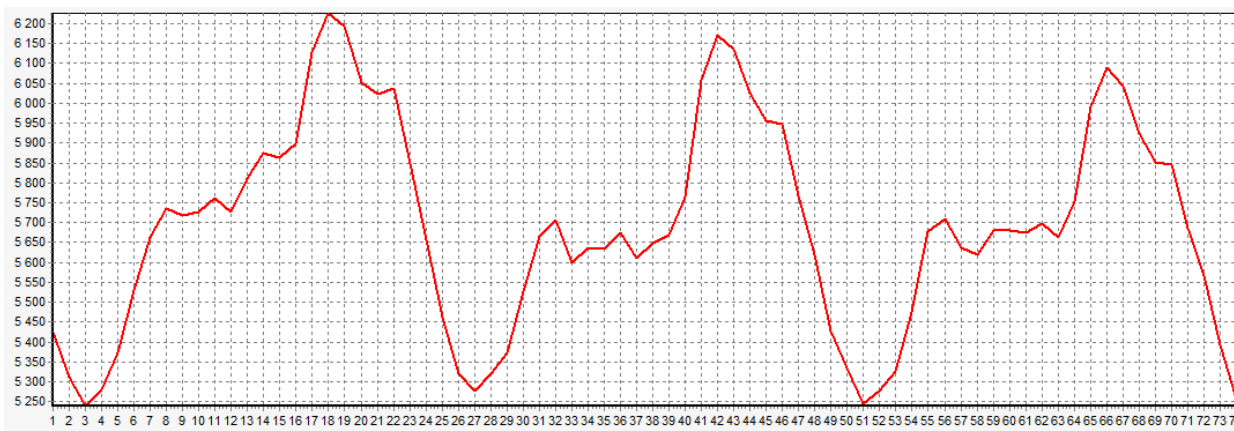


Рис. 4. Временной ряд часовых данных потребления электроэнергии за трое суток (в МВт)

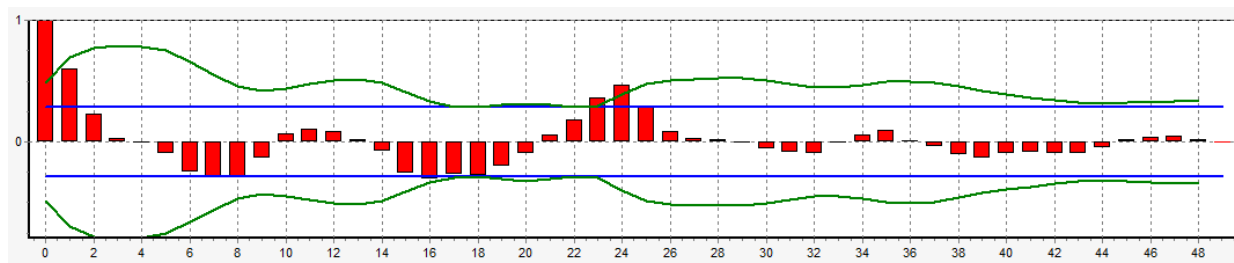


Рис. 5. Автокорреляционная функция временного ряда потребления электроэнергии (n=50; k=48)

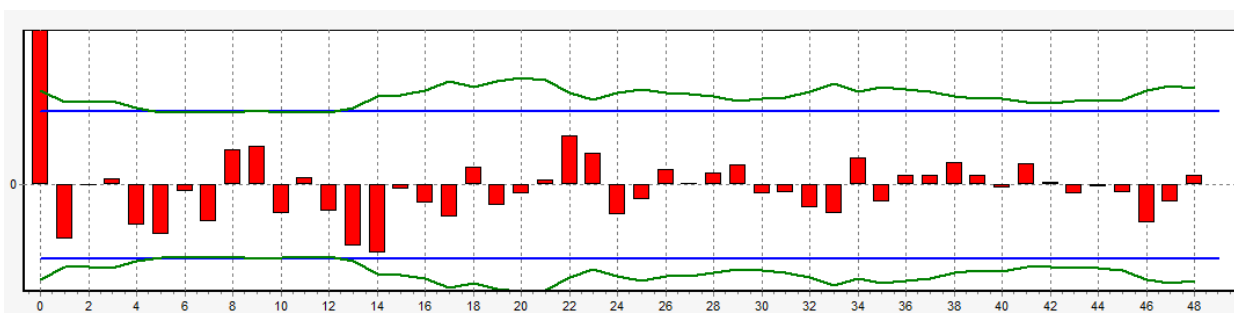


Рис. 6. Частная автокорреляционная функция временного ряда потребления электроэнергии (n=50; k=48)

Таким образом, выберем длину окна метода «Гусеница»-SSA $L=24$ и рассмотрим собственные значения траекторной матрицы (рис. 7).

Рассмотрим также графики рядов разложения (рис. 8). Анализ собственных значений траекторной матрицы, рядов разложения или факторных векторов помогает на этапе группировки так выбрать подмножества I_1, I_2, \dots, I_r , чтобы компоненты разложения были интерпретированы как трендовая, сезонная или шумовая составляющая исходного ВР.

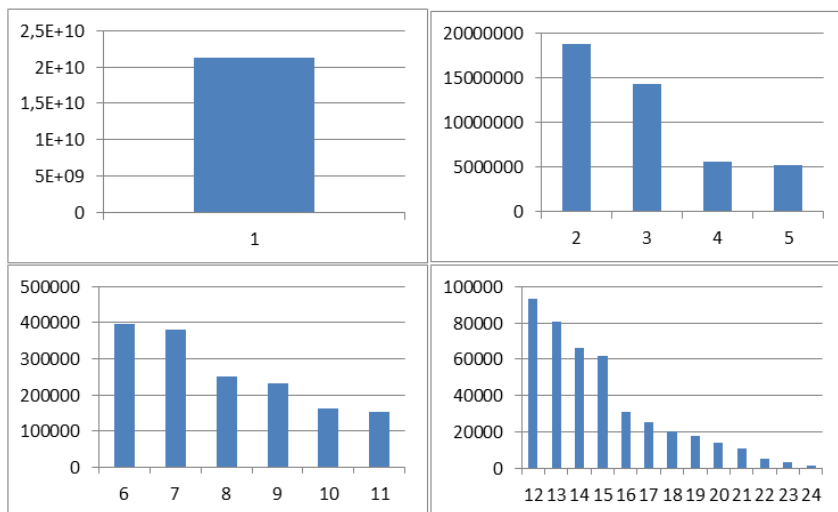


Рис. 7. Собственные значения траекторной матрицы временного ряда потребления электроэнергии

Сравнительный анализ эффективности прогнозирования рассмотренными моделями будем осуществлять при помощи наиболее широко используемой статистики RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{1}{n_1} \sum_{t=1}^{n_1} (y_t - \hat{y}_t)^2},$$

где n_1 – количество вычисленных прогнозов, y_t – фактические значения ВР, \hat{y}_t – прогнозные значения ВР.

Тестирование будет проводиться следующим образом. Длина окна метода «Гусеница»-SSA будет фиксированной, но модели одного класса будут отличаться количеством первых компонент разложения, используемых в них. Поэтому на приведенных ниже столбчатых диаграммах на оси абсцисс будем откладывать количество первых компонент разложения, используемых в конкретной модели, а на оси ординат – ошибку прогнозирования, выраженную в статистике RMSE.

Анализ диаграмм (рис. 9–11, 18, 19, 25–27) позволяет сделать вывод, что увеличение количества используемых в моделях компонент разложения постепенно снижает ошибку прогнозирования, а затем, начиная с некоторого количества используемых компонент разложения, эта ошибка начинает постепенно возрастать.

Анализ диаграмм (рис. 9, 18, 25) говорит о снижении ошибки прогнозирования при добавлении к ЛРФ метода «Гусеница»-SSA модели SARIMA, построенной на остаточных ошибках модели, определяемых путём вычитания из значений исходного ВР соответствующих значений, получаемых при помощи ЛРФ.

На следующих двух столбчатых диаграммах (рис. 10, 11) представлены зависимости ошибок прогнозирования моделей ЛРФ-SSA + SARIMA, Тренд-SSA-SARIMA и моделей ЛРФ-SSA + SARIMA и SSA-SARIMA соответственно.

Рассмотрение приведенных выше результатов позволяет сделать вывод, что анализ собственных значений траекторной матрицы и ВР разложения исходного ВР не позволяет точно определиться с количеством отбираемых в декомпозиционную модель ВР разложения, чтобы такая модель имела наименьшую ошибку прогнозирования, а позволяет только определиться с группировкой ВР разложения для модели SSA-SARIMA. Поэтому для следующих тестовых примеров не имеет смысла приводить собственные значения траекторной матрицы и серии разложения.

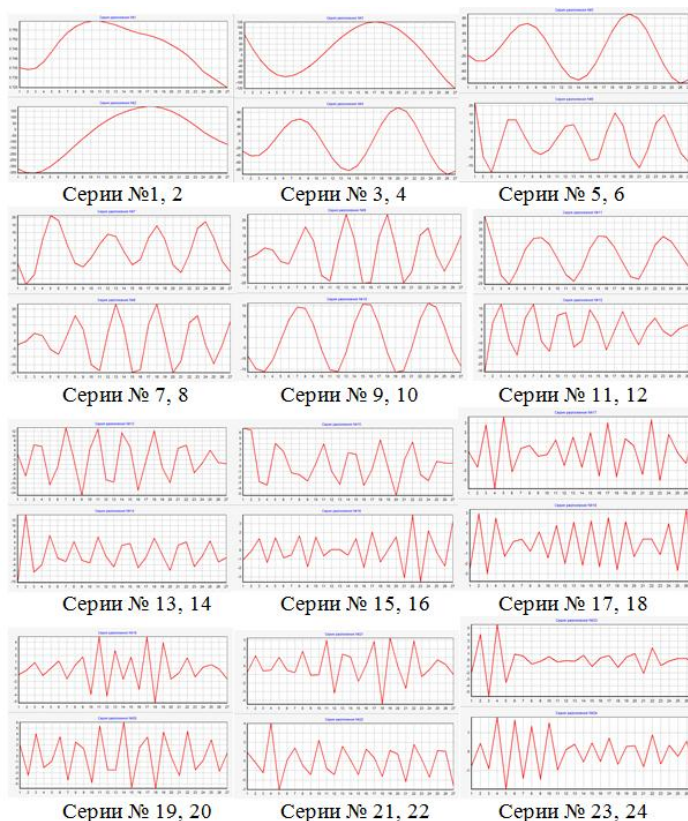


Рис. 8. Графики временных рядов разложения методом «Гусеница»-SSA временного ряда потребления электроэнергии

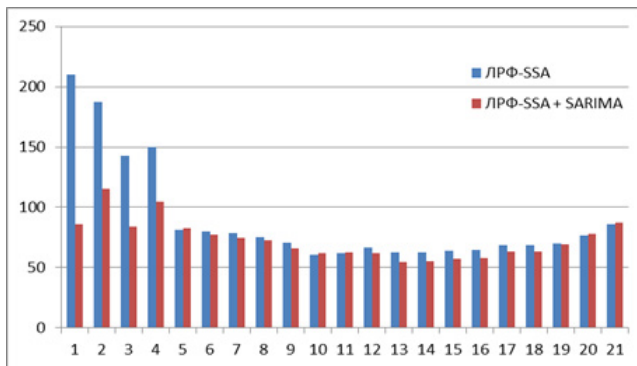


Рис. 9. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA и ЛРФ-SSA + SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

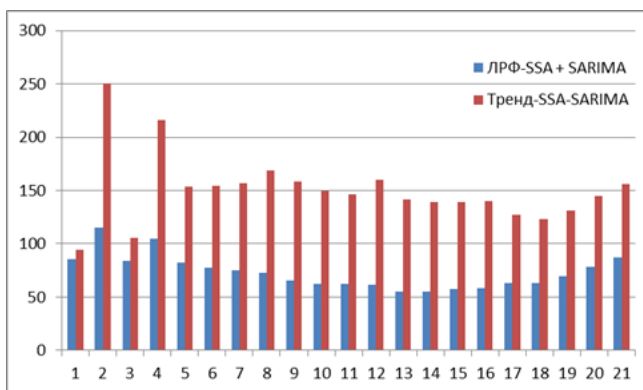


Рис. 10. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA + SARIMA и Тренд-SSA-SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

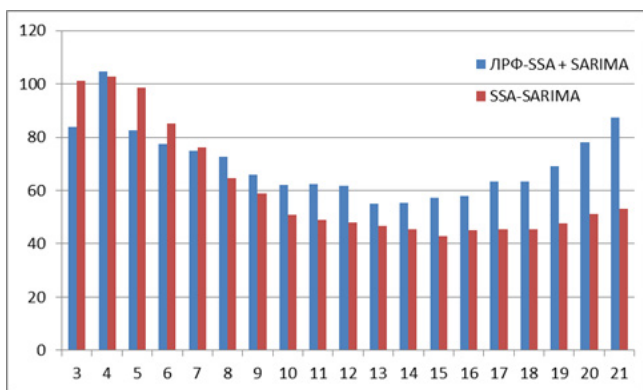


Рис. 11. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA + SARIMA и SSA-SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

Под структурной идентификацией приведенных декомпозиционных моделей будем подразумевать перебор различных вариантов разложения и выбор наилучшей по точности прогнозирования модели из множества альтернатив. Поэтому приоритетной задачей является разработка показателей, позволяющих автоматизировать и ускорить процесс определения

структуры наилучшей по точности прогнозирования декомпозиционной модели из определённого класса без перебора множества вариантов.

На следующей диаграмме (рис. 12) изображены ошибки прогнозирования RMSE для наилучших моделей из рассмотренных классов моделей. Наименьшую ошибку прогнозирования показала модель SSA-SARIMA.

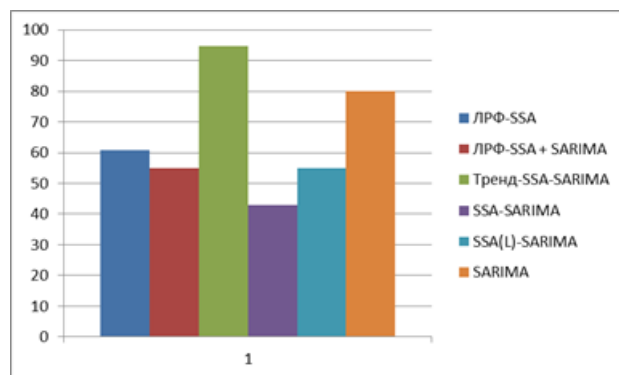


Рис. 12. Диаграмма минимальных ошибок прогнозирования RMSE для рассмотренных классов моделей

Как видно из диаграммы рис. 11 наилучшая по эффективности прогнозирования модель SSA-SARIMA включила в себя первые 15 компонент разложения методом «Гусеница»-SSA исходного ВР. При этом следует отметить, что группировка компонент разложения производилась как показано на следующих графиках (рис. 13).

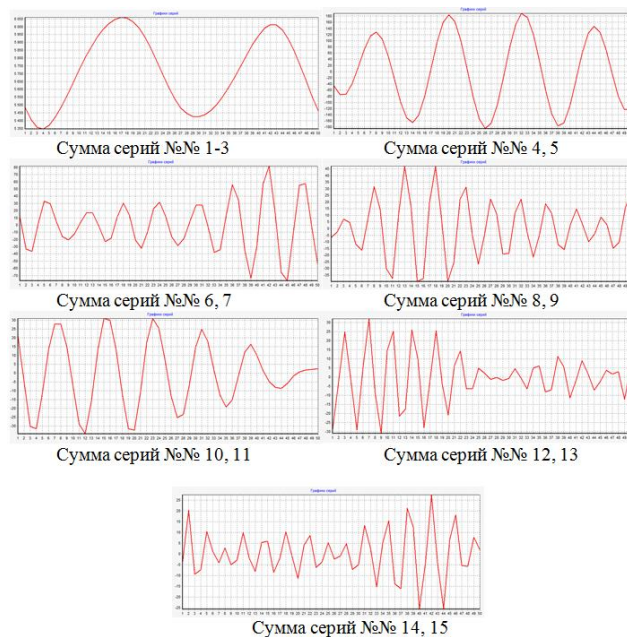


Рис. 13. Группировка компонент разложения для наилучшей по эффективности модели SSA-SARIMA

Приведём прогноз наилучшей из класса SSA-SARIMA модели (рис. 14).

Рассмотрим более длинный ВР часовых значений потребления электроэнергии объёма 1008 значений, что соответствует 6 неделям (рис. 15). Обучение мо-

делей будет производиться на выборке данных за 5 недель (840 значений), а тестирование – на данных последней недели. Прогноз тоже будет выполняться одношаговый и производится дальнейшее скольжение окна до последнего значения BP.

На рис. 16 и рис. 17 приведены графики автокорреляционной и частной автокорреляционной функций соответственно после следующего преобразования исходного BP $y'_t = \nabla_1^1 \nabla_{24}^1 y_t$. Анализ данных графиков совместно с отбором моделей SARIMA по критериям AIC и BIC определили следующую структуру модели: SARIMA(7,1,2) × (3,1,1)₂₄ × (1,0,1)₁₂₀ × (0,0,1)₁₆₈:

$$y_t = \frac{(1+0.495B+0.947B^2)(1-0.861B^{24})}{\nabla_1^1 \nabla_{24}^1 (1+0.396B+0.838B^2-0.011B^3+0.102B^4+0.076B^5+0.033B^6-0.082B^7)} \times \frac{(1-0.208B^{120})(1+0.336B^{168})}{(1-0.16B^{24}-0.001B^{48}+0.022B^{72})(1+0.040B^{120})} a_t.$$

Выберем длину окна метода «Гусеница»-SSA L=396. В модели будем отбирать от 1 до 24 первых компонент разложения.

Приведенная ниже столбчатая диаграмма (рис. 18), говорит о существенном снижении ошибки прогнозирования при добавлении к ЛРФ модели SARIMA, как это было описано для модели ЛРФ-SSA+SARIMA.

Из следующих диаграмм (рис. 19, 20) видно, что для данного тестового примера наиболее эффективный прогноз показала модель ЛРФ-SSA + SARIMA.

Приведём прогноз моделью с наименьшей ошибкой RMSE (рис. 21).

Рассмотрим теперь тестовый пример прогнозирования BP потребления природного газа. График данного BP среднесуточного объёма газопотребления представлен на рис. 22 (объём 758 значений). Обучение моделей будем проводить на данных за два года (объём 730 значений). Тестирование на данных за 4 недели. Прогноз будет производиться одношаговый, также как и для предыдущих двух примеров.

На рис. 23, 24 приведены графики автокорреляционной и частной автокорреляционной функций соответственно после следующего преобразования исходного BP $y'_t = \nabla_1^1 y_t$. Анализ данных графиков совместно с отбором моделей SARIMA по критериям AIC и

BIC определили следующую структуру модели:

SARIMA(7,1,1) × (1,0,1)₂₁ × (1,0,0)₃₅ × (1,0,0)₅₆ × (1,0,0)₇₀.

$$y_t = \frac{(1-0.851B)}{\nabla_1^1 (1-0.899B+0.058B^2-0.009B^3-0.016B^4+0.044B^5+0.013B^6+0.037B^7)} \times \frac{(1-0.609B^{21})}{(1-0.724B^{21})(1-0.08B^{35})(1-0.039B^{56})(1-0.066B^{70})} a_t.$$

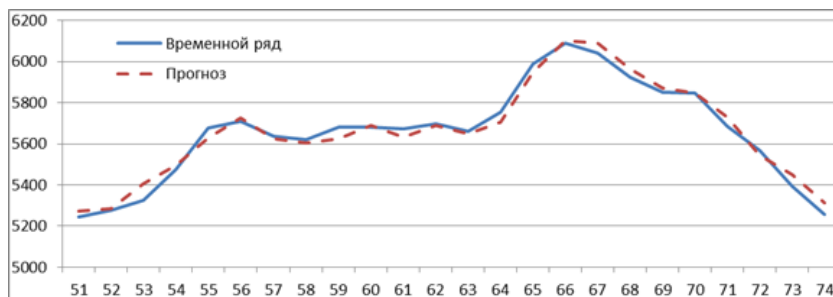


Рис. 14. График временного ряда потребления электроэнергии и прогнозы, полученные моделью с наименьшей ошибкой RMSE

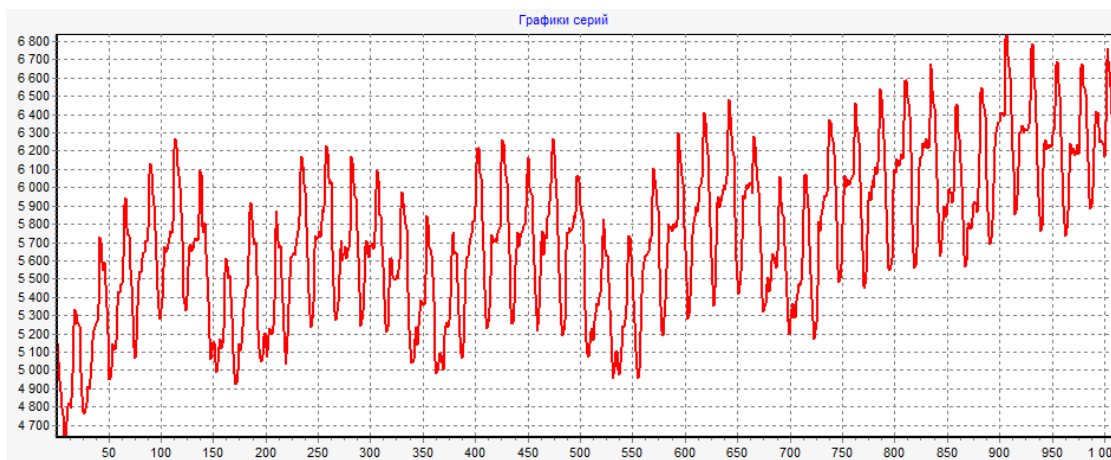


Рис. 15. График временного ряда часовых значений потребления электроэнергии за 6 недель (в МВт)



Рис. 16. Автокорреляционная функция временного ряда потребления электроэнергии (n=840; k=200)

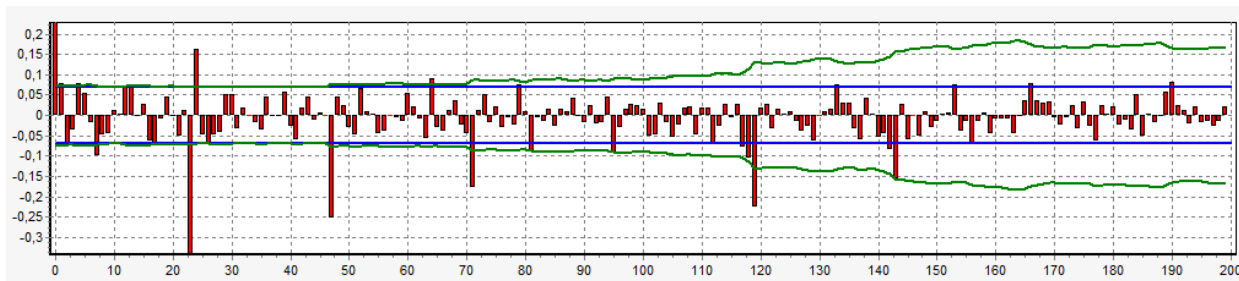


Рис. 17. Частная автокорреляционная функция временного ряда потребления электроэнергии (n=840; k=200)

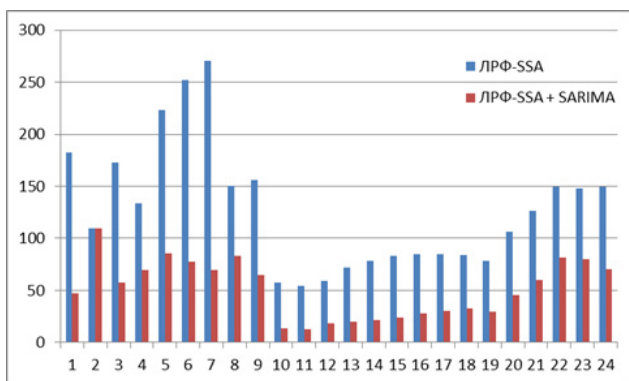


Рис. 18. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA и ЛРФ-SSA+ SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

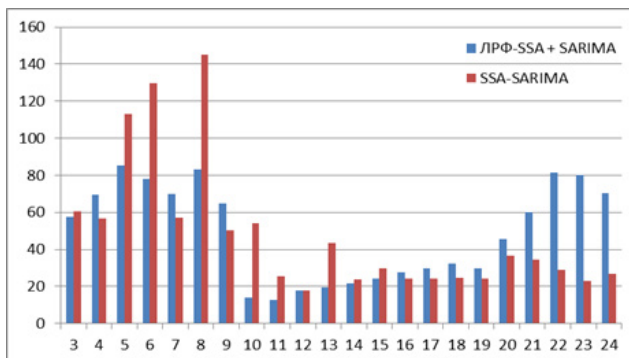


Рис. 19. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA+SARIMA и SSA-SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

Ниже приведена столбчатая диаграмма (рис. 25), которая говорит о снижении ошибки прогнозирования

при добавлении к ЛРФ модели SARIMA, как это было и для двух предыдущих тестов.

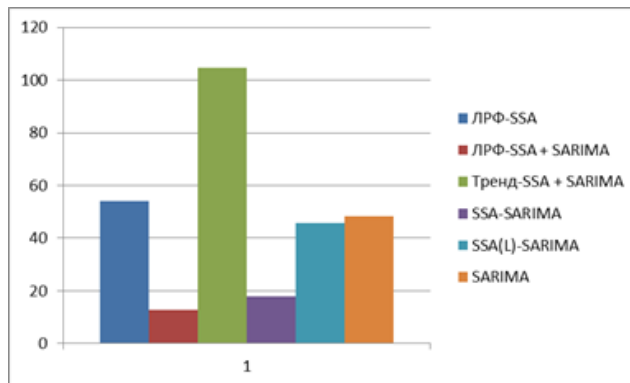


Рис. 20. Диаграмма минимальных ошибок прогнозирования RMSE для рассмотренных классов моделей

Для случаев прогнозирования ВР со сложными стохастическими трендами модели Тренд-SSA-SARIMA дают более эффективный прогноз в сравнении с моделями ЛРФ-SSA + SARIMA (рис. 26).

Кроме того, более простая модель Тренд-SSA-SARIMA может давать более эффективный прогноз в сравнении с моделью SSA-SARIMA (рис. 27, 28).

Для ВР потребления природного газа, который имеет полиномиальный, полигармонический и стохастический тренд со сложной корреляционной структурой наилучший результат выдала модель из класса Тренд-SSA-SARIMA (рис. 28).

Ниже приведём прогнозы наилучшей из класса Тренд-SSA-SARIMA модели (рис. 29).

Таким образом, для ВР различного объёма и структуры, наиболее эффективный прогноз дают модели из различных классов.

В дальнейшем целесообразно обобщит гибридные математические модели и методы, основанные на методах «Гусеница»-SSA и Бокса-Дженкина для прогнозирования взаимосвязанных нестационарных ВР.

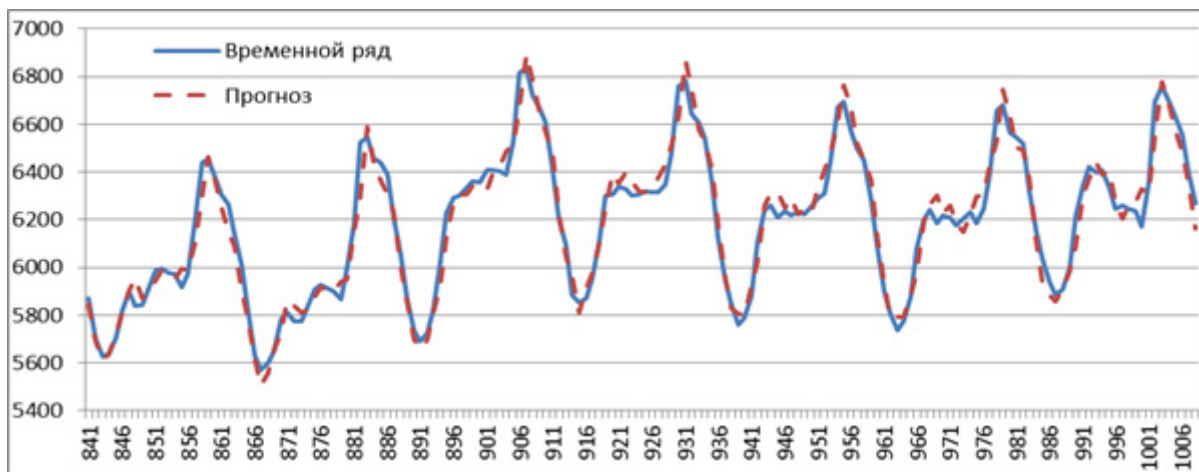


Рис. 21. График временного ряда потребления электроэнергии и прогнозы, полученные наилучшей из рассмотренных моделей

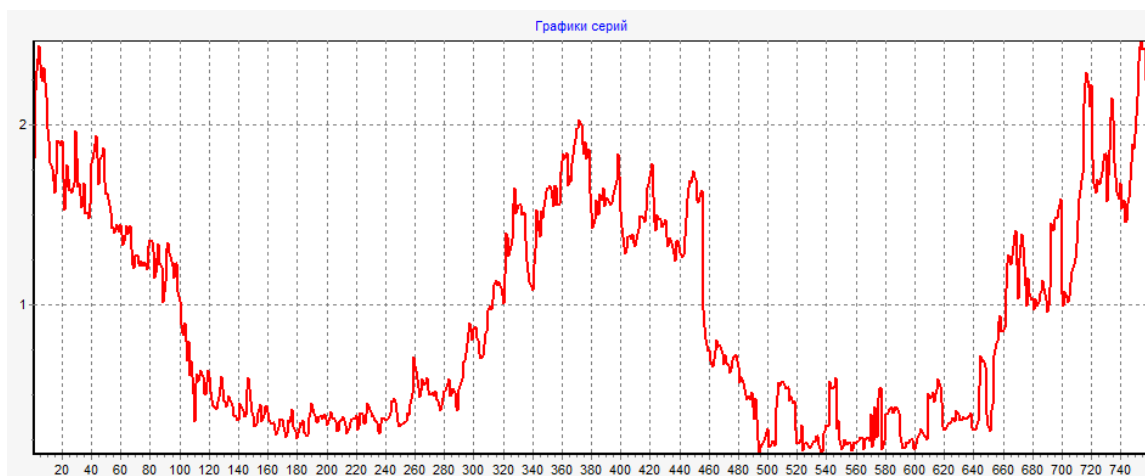


Рис. 22. Временной ряд суточных данных потребления природного газа объёмом 758 значений (в $\text{Мм}^3/\text{сут}$)

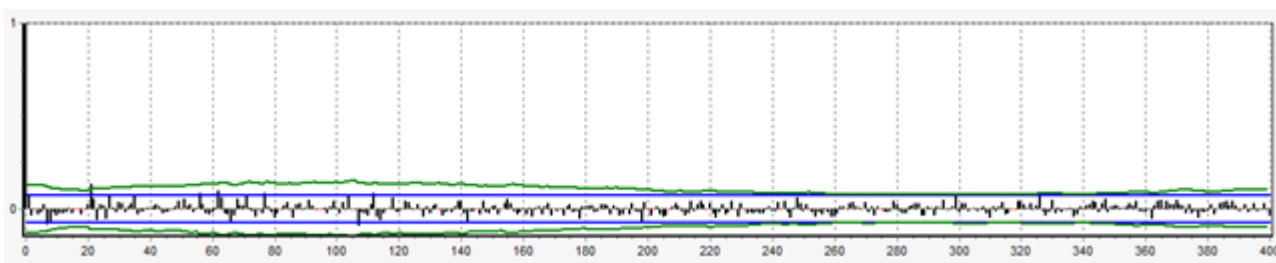


Рис. 23. Автокорреляционная функция временного ряда потребления природного газа ($n=730$; $k=400$)

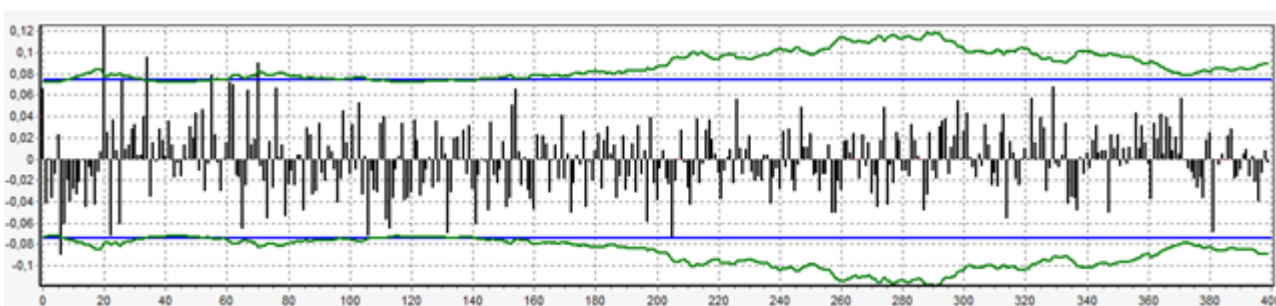


Рис. 24. Частная автокорреляционная функция временного ряда потребления электроэнергии ($n=840$; $k=200$)

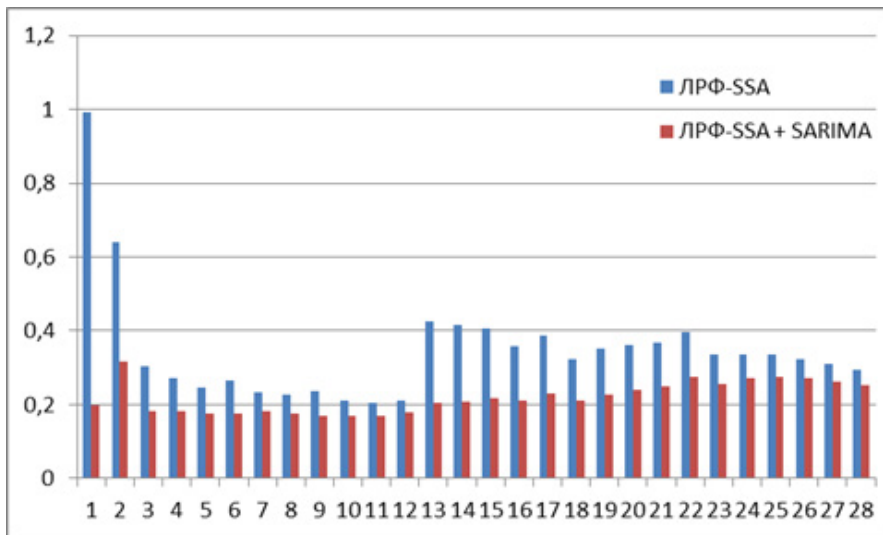


Рис. 25. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA и ЛРФ-SSA + SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

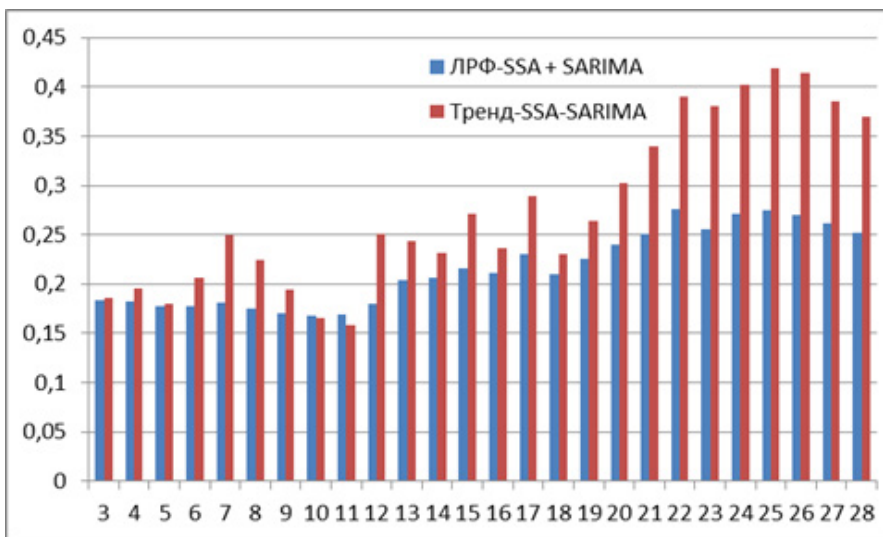


Рис. 26. Диаграммы зависимости ошибки прогнозирования RMSE моделей ЛРФ-SSA + SARIMA и Тренд-SSA-SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

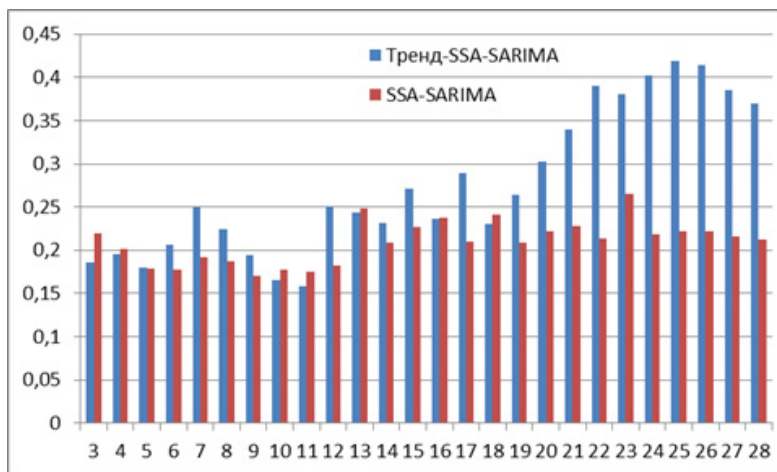


Рис. 27. Диаграммы зависимости ошибки прогнозирования RMSE моделей Тренд-SSA-SARIMA и SSA-SARIMA от количества используемых в них первых компонент разложения метода «Гусеница»-SSA

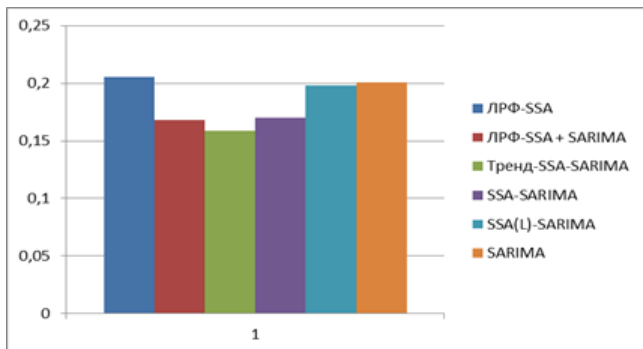


Рис. 28. Диаграмма минимальных ошибок прогнозирования RMSE для рассмотренных классов моделей

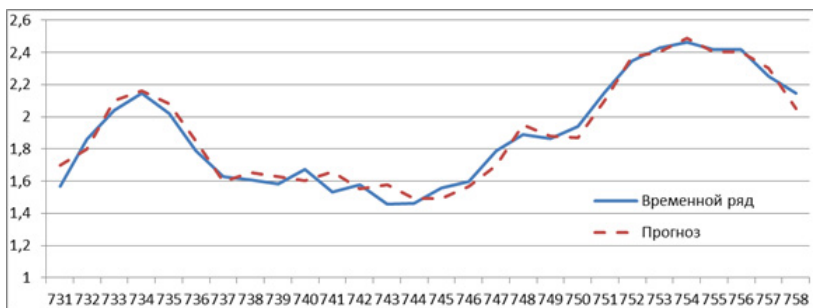


Рис. 29. График потребления природного газа и прогнозы, полученные наилучшей из рассмотренных моделей

6. Выводы

В данной статье рассмотрены трендовый и декомпозиционный подходы к прогнозированию нестационарных ВР. В соответствии с ними предложены декомпозиционные модели различной сложности для прогнозирования нестационарных ВР, а также методы идентификации этих моделей на основе совместного использования методов «Гусеница»-SSA и Бокса-Дженкинса, обобщённого для случая нескольких сезонных компонент. Гибридные математические модели трендового подхода к прогнозированию, основанные на методах «Гусеница»-SSA и Бокса-Дженкинса заключаются в моделировании процесса как отклонения фактических значений ВР относительно трендовой составляющей, в роли которой в предлагаемых моделях выступает линейная рекуррентная формула (ЛРФ) метода «Гусеница»-SSA или её аппроксимация моделью SARIMA. Основной же целью декомпозиционного подхода к прогнозированию на основе методов «Гусеница»-SSA и Бокса-Дженкинса является разложение методом «Гусеница»-SSA исходного ВР на множество ВР с более простой структурой, рассматриваемых независимо друг от друга; прогнозирование данных компонент разложения моделями SARIMA и вычисление общего прогноза, объединяя прогнозы построенных упрощённых моделей. Предложенные модели были протестированы на ВР потребления электроэнергии и природного газа, а их результаты прогнозирования были сравнены с результатами, полученными классическими вероятностными моделями SARIMA, обобщёнными на случай нескольких сезонных компонент.

Полученные результаты позволяют сделать заключение, что для получения эффективных прогнозов необходимо производить декомпозицию исследуемых ВР различными способами и отбирать наиболее эффективные из них для ВР определённой структуры. А также комбинировать различные модели, описывающие как статистические, так и детерминированные составляющие ВР, что позволяет достичь наилучшего качества прогнозирования. Предложенные модели учитывают следующие виды нестационарности: по тренду, по частоте и амплитуде колебаний, по дисперсии шумовой составляющей процесса.

Основным преимуществом предлагаемых методов построения моделей адекватных исследуемому процессу является их строгая формализация и, следовательно, возможность полной автоматизации всех этапов построения и использования.

Литература

1. Седов, А. В. Моделирование объектов с дискретно-распределёнными параметрами: декомпозиционный подход [Текст] / А. В. Седов. – Южный научный центр РАН. – М.: Наука, 2010. – 438 с.
2. Бэнн, Д. В. Сравнительные модели прогнозирования электрической нагрузки [Текст] / Д. В. Бэнн, Е. Д. Фармер; пер. с англ. – М.: Энергоатомиздат, 1987. – 200 с.
3. Qiang, Z. Singular Spectrum Analysis and ARIMA Hybrid Model for Annual Runoff Forecasting [Text] / Z. Qiang, D. W. Ben, H. Bin, P. Yong, L. R. Ming // Water Resour Manage. – 2011. – Vol. 25, Issue 11. – P. 2683–2703. doi: 10.1007/s11269-011-9833-y
4. Евдокимов, А. Г. Оперативное управление потоко-распределением в инженерных сетях [Текст] / А. Г. Евдокимов, А. Д. Тевяшев. – Х.: Вища школа, 1980. – 144 с.
5. Lawrance, A. J. Stochastic modelling of riverflow time series [Text] / A. J. Lawrance, N. T. Kottegoda // J. R. Stat. Soc. A. – 1977. – Vol. 140, issue 1. – P. 1–47. doi: 10.2307/2344516
6. Fernando, D. A. K. Generation and forecasting of monsoon rainfall data [Text] / D. A. K. Fernando, W. A. Jayawardena // In Proc. of the 20th WEDC conference. Colombo, Sri Lanka, 1994. – P. 310–313.
7. Yurekli, K. Application of linear stochastic models to monthly flow data of Kelkit Stream [Text] / K. Yurekli, A. Kurunca, F. Ozturkb // Ecol Model. – 2005. – Vol. 183, Issue 1. – P. 67–75. doi: 10.1016/j.ecolmodel.2004.08.001
8. Broomhead, D. S. Extracting qualitative dynamics from experimental data [Text] / D. S. Broomhead, G. P. King // Physica D. – 1986. – Vol. 20, Issue 2-3. – P. 217–236. doi: 10.1016/0167-2789(86)90031-x
9. Fraedrich, K. Estimating the dimension of weather and climate attractor [Text] / K. Fraedrich // J. Atmos Sci. – 1986. – Vol. 43. – P. 419–432.
10. Vautard, R. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series [Text] / R. Vautard, M. Ghil // Physica D. – 1989. – Vol. 35, Issue 3. – P. 395–424. doi: 10.1016/0167-2789(89)90077-8

11. Ghil, M. Interdecadal oscillations and the warming trend in global temperature time series [Text] / M. Ghil, R. Vautard // Nature. – 1991. – Vol. 350, Issue 6316. – P. 324–327. doi: 10.1038/350324a0
12. Yiou, P. Spectral analysis of climate data [Text] / P. Yiou, E. Baert, M.F. Loutre // Surv Geophys. – 1996. – Vol. 17, Issue 6. – P. 619–663. doi: 10.1007/bf01931784
13. Lisi, F. Combination of singular spectrum analysis and auto regressive model for short term load forecasting [Text] / F. Lisi, O. Nicolis, M. Sandri // Neural Process Lett. – 1995. – Vol. 2, Issue 4. – P. 6–10.
14. Sivapragasam, C. Rainfall and discharge forecasting with SSA-SVM approach [Text] / C. Sivapragasam, S.Y. Liong, M.F.K. Pasha // J. Hydroinform. – 2001. – Vol. 3, Issue 7. – P. 141–152.
15. Golyandina, N. Analysis of time series structure: SSA and related techniques [Text] / N. Golyandina, V. Nekrutkin, A. Zhigljavsky // Chapman and Hall/CRC, New York, 2001. doi: 10.1201/9781420035841
16. Marques, C. A. F. Singular spectrum analysis and forecasting of hydrological time series [Text] / C. A. F. Marques, J. A. Ferreira, A. Rocha, J. M. Castanheira // In Meeting of the European-Union-of-Geosciences. Vienna, Austria, 2005.
17. Hassani, H. Forecasting European industrial production with singular spectrum analysis [Text] / H. Hassani, S. Heravi, A. Zhigljavsky // Int. J. Forecast. – 2009. – Vol. 25, Issue 1. – P. 103–118. doi: 10.1016/j.ijforecast.2008.09.007
18. Wensheng, D. Financial Time Series Forecasting Using A Compound Model Based on Wavelet Frame and Support Vector Regression [Text] / D. Wensheng, Lu Chi-Jie // In the 4th International Conference on Neural Computation, 2008. – P. 328–332. doi: 10.1109/icnc.2008.455
19. Kurbatskii, V. G. On the Neural Network Approach for Forecasting of Nonstationary Time Series on The Basis of the Hilbert-Huang Transform [Text] / V. G. Kurbatskii, D. N. Sidorov, V. A. Spiryaev, N. V. Tomin // Automation and Remote Control. – 2011. – Vol. 72, Issue 7. – P. 1405–1414. doi: 10.1134/s0005117911070083
20. Zhang, W. Q. Time series forecasting method based on Huang transform and BP neural network [Text] / W. Q. Zhang, C. Xu // In Proc. Of the 7th International Conference on Computational Intelligence and Security, 2011. – P. 497–502. doi: 10.1109/cis.2011.116
21. Lu, C.-J. ICA-Based Signal Reconstruction Scheme with Neural Network in Time Series Forecasting [Text] / C.-J. Lu, J.-Yu Wu, T.-S. Lee // In First Conference on Intelligent Information and Database Systems, 2009. – P. 318–323. doi: 10.1109/aciids.2009.28
22. Xiang, L. A hybrid support vector regression for time series forecasting [Text] / L. Xiang, Y. Zhu, G.-J. Tang // In World Congress on Software Engineering, 2009. – P. 161–165. doi: 10.1109/wcse.2009.130
23. Sallehuddin, R. Hybridization Model of Linear and Nonlinear Time Series Data for Forecasting [Text] / R. Sallehuddin, S. M. Shamsuddin, S. Z. M. Hashim // In Second Asia International Conference on Modelling & Simulation, 2008. – P. 597–602. doi: 10.1109/ams.2008.142
24. Henrique, S. H. Combining Neural Networks and ARIMA Models for Hourly Temperature Forecast [Text] / H. S. Hippert, C. E. Pedreira, R. C. Souza // In Proceedings of the International Joint Conference on Neural Networks, 2000. – P. 1–6. doi: 10.1109/ijcnn.2000.860807
25. Xuemei, L. A Novel Air-conditioning Load Prediction Based on ARIMA and BPNN Model [Text] / L. Xuemei, D. Lixing, S. Ming, X. Gang, L. Jibin // In Asia-Pacific Conference on Information Processing, 2009. – P. 51–54. doi: 10.1109/apcip.2009.21
26. Tian, F. P. Forecast of Cerebral Infraction Incidence Rate Based on BP Neural Network and ARIMA Combined Model [Text] / F. P. Tian, L. L. Ma // In International Symposium on Intelligence Information Processing and Trusted Computing, 2010. – P. 216–219. doi: 10.1109/iptc.2010.7
27. Feng, K. Time Series Forecasting Model with Error Correction by Structure Adaptive Support Vector Machine [Text] / K. Feng, W. Xiaojuan // In International Conference on Computer Science and Software Engineering, 2008. – P. 1067–1070. doi: 10.1109/csse.2008.88
28. Lo, J.-H. A Data-Driven Model for Software Reliability Prediction [Text] / J.-H. Lo // In International Conference on Granular Computing, 2012. – P. 1–6. doi: 10.1109/grc.2012.6468581
29. He, Y. Research on Hybrid ARIMA and Support Vector Machine Model in Short Term Load Forecasting [Text] / Y. He, Y. Zhu, D. Duan // In Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications, 2006. – P. 1–5. doi: 10.1109/isda.2006.229
30. Linh, B. N. An Empirical Study on Forecasting using Decomposed Arrival Data of an Enterprise Computing System [Text] / B. N. Linh, A. Amy, H. Doug // In 9th International Conference on Information Technology- New Generations, 2012. – P. 756–763. doi: 10.1109/itng.2012.36
31. Hou, Z. Standardized Software for Wind Load Forecast Error Analyses and Predictions Based on Wavelet-ARIMA Models – Applications at Multiple Geographically Distributed Wind Farms [Text] / Z. Hou, Y. V. Makarov, N. A. Samaan, P. V. Etingov // In Hawaii International Conference on System Sciences, 2013. – P. 5005–5011. doi: 10.1109/hicss.2013.495
32. Щелкалин, В. Н. Трендовый подход прогнозирования временных рядов на основе метода «Гусеница»-SSA [Текст] / Материалы 14-й Международной научно-технической конференции САИТ 2012, Киев, 24 апреля 2012 г. / В. Н. Щелкалин // УНК «ИПСА» НТУУ «КПИ». – К.: УНК «ИПСА» НТУУ «КПИ», 2012. – С. 258 – 259.
33. Vahabie, A. H. Combination of Singular Spectrum Analysis and Autoregressive Model for short term load forecasting [Text] / A. H. Vahabie, M. M. R. Yousefi, B. N. Araabi, C. Lucas, S. Barghinia // IEEE LAUSANNE POWERTECH, 2007. – P. 1090–1093. doi: 10.1109/pct.2007.4538467
34. Щелкалин, В. Н. Декомпозиционный подход прогнозирования временных рядов на основе метода «Гусеница»-SSA [Текст]: матер. 14-й Междун. науч.-тех. конф. САИТ / В. Н. Щелкалин // УНК «ИПСА» НТУУ «КПИ». – К.: УНК «ИПСА» НТУУ «КПИ», 2012. – С. 260–261.
35. Голяндина, Н. Э. Метод «Гусеница»-SSA: прогноз временных рядов [Текст]: уч. пос. / Н. Э. Голяндина. – СПб.: С.-Петербургский государственный университет, 2004. – 52 с.