

В роботі запропоновано математичні моделі прогнозування нестационарних часових рядів з урахуванням зовнішніх чинників і методи їх структурної ідентифікації, засновані на спільному використанні багатовимірного варіанта методу «Гусениця»-SSA та моделей SARIMAX, розширених для прогнозування часових рядів з кількома періодичними компонентами і з урахуванням декількох екзогенних змінних. Експериментальні результати показують високу ефективність запропонованих моделей прогнозування при виборі відповідних структурних параметрів у порівнянні з моделями SARIMAX

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

В работе предложены математические модели прогнозирования нестационарных временных рядов с учётом внешних факторов и методы их структурной идентификации, основанные на совместном использовании многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX, расширенных для прогнозирования временных рядов с несколькими сезонными составляющими и с учётом нескольких экзогенных переменных. Экспериментальные результаты показывают высокую эффективность предложенных моделей прогнозирования при выборе подходящих структурных параметров в сравнении с моделями SARIMAX

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

УДК 519.254

DOI: 10.15587/1729-4061.2014.31729

ГИБРИДНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ ПРОГНОЗИ- РОВАНИЯ ВРЕМЕННЫХ РЯДОВ С УЧЁТОМ ВНЕШНИХ ФАКТОРОВ

В. Н. Щелкалин

Инженер

Кафедра прикладной математики

Харьковский национальный университет

радиоэлектроники

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

E-mail: vitalii.shchelkalin@gmail.com

1. Введение

Точность решения задач оперативного планирования режимов распределения целевых продуктов (воды, газа, электроэнергии) в инженерных сетях, а также возможность использования полученных результатов в практике диспетчерских служб управления инженерными сетями, определяются двумя факторами: степенью адекватности используемых математических моделей и точностью исходных данных. Важнейшие исходные данные для задач оперативного планирования – это прогнозы потребления целевого продукта в инженерных сетях. Недостаточное знание механизма процессов водо-, газо- и электропотребления, влияние на них огромного числа неконтролируемых и неуправляемых факторов приводят к тому, что эти процессы являются случайными процессами, которые в зависимости от типа потребителя могут носить как стационарный, так и нестационарный характер. Кроме того, они подвержены влиянию метеорологических, организационных и хронологических факторов, изменяющих их стохастическую структуру [1]. Это привело к появлению большого числа методов прогнозирования этих процессов, большинство из которых носит частный характер и позволяет получить удовлетворительные результаты только для ограниченного класса потребителей и типов прогнозов.

В настоящей работе в качестве моделей процессов потребления целевых продуктов (ППЦП) в инженерных сетях рассмотрим гибридные модели на основе многомерного варианта метода «Гусеница»-SSA (MSSA) и сезонных моделей авторегрессии – интегрированного скользящего среднего с экзогенными переменными, расширенных на случай прогнозирования ВР с несколькими сезонными компонентами и с учётом нескольких экзогенных переменных (далее такие расширенные модели будем обозначать САРПС-СЭ, на англ. – SARIMAX), позволяющих получить адекватные модели ППЦП, реальных потребителей с различной стохастической, детерминированной или вероятностно-детерминированной структурой.

Как уже отмечалось, ППЦП в инженерных сетях зависят от большого числа факторов трех основных типов: хронологических (время года, дни недели, сменность работы, режимы работы промышленных предприятий и т. п.), метеорологических (температура окружающей среды, скорость ветра, влажность наружного воздуха и т. п.), организационных (мероприятия, проводимые в сетевых системах, массовые мероприятия, проводимые с населением).

Соотношение между компонентами ППЦП определяется типом потребителя. Влияние хронологических факторов на ППЦП приводит, как правило, к появлению в последних периодических компонент. Их природа связана с характером технологических

процессов промышленных предприятий, временами года и т. п. Таким образом, появляется возможность для построения широкого класса моделей ППЦП, содержащих периодические компоненты либо в виде одной синусоиды, либо в виде отрезка ряда Фурье, либо в виде амплитудно-модулированного процесса. Влияние хронологических факторов для некоторых типов потребителей, а точнее, совместное влияние всех факторов может уменьшить или даже полностью исключить периодические компоненты из исследуемых процессов. Непрерывное увеличение потребления воды, газа или электроэнергии, а также влияние сезонных компонент с большими периодами приводит к появлению в этих процессах детерминированных и стохастических трендов и, следовательно, к нестационарности ППЦП. Установление этого факта позволило использовать модели и методы, позволяющие описывать ППЦП как нестационарные процессы.

Влияние на ППЦП температуры окружающей среды послужило основой для построения моделей, описывающих процессы газо- и электропотребления как некоторую детерминированную функцию от температуры окружающей среды.

В методе Бокса-Дженкинса для связи параметров ВР с экзогенными факторами используется мультипликативная модель SARIMAX, обладающая рациональной структурой и позволяющая адекватно описывать как стационарные, так и однородные нестационарные случайные процессы, содержащие n_s периодических компонент. Для связи параметров ВР используются модели линейных дискретных передаточных функций с запаздывающим аргументом. Рассмотрены методы проверки адекватности получаемых моделей, основанные на анализе автокорреляционных и взаимных корреляционных функций остаточных ошибок модели. Приведены алгоритмы вычисления и коррекции прогнозов и их дисперсий.

Отличительной чертой метода MSSA является то, что он не требует предварительного задания модели ряда. Тем не менее, он позволяет раскладывать временной ряд (ВР) на интерпретируемые составляющие, такие как тренд (в частности, линейный или экспоненциальный), периодические компоненты и шум. При этом не надо заранее знать параметрический вид тренда, а также о наличии колебательных компонент и их периодах [2]. Техническую основу метода составляет сингулярное разложение траекторной матрицы, столбцами которой являются вектора вложения – отрезки исходных ВР длины L , основного параметра метода, называемого длиной окна [3]. Анализ членов сингулярного разложения позволяет сначала классифицировать их как относящиеся к одной из компонент ряда, а затем выделить эту компоненту.

2. Анализ литературных данных и постановка проблемы

В последнее время появилось много работ, посвященных гибридным декомпозиционным моделям и методам прогнозирования, краткий обзор которых представлен в [4]. Для оценивания эффективности прогнозы обычно сравниваются с прогнозами, получаемыми моделями ARIMA [5–7].

ВР потребления целевых продуктов в инженерных сетях являются существенно нелинейными со средне- и долгосрочными зависимостями, поэтому аппроксимация таких сложных ВР только моделью ARIMA не всегда является удовлетворительной. Сингулярный спектральный анализ (SSA) является одним из наиболее эффективных методов для анализа ВР со сложными периодическими компонентами. SSA стал широко используемым методом анализа периодических и трендовых составляющих ВР [8–11]. В последнее время были разработаны новые модели для прогнозирования, основанные на SSA. В [12] предложили линейную рекуррентную формулу сингулярного спектрального анализа (SSA-LRF), и эта модель была использована для решения некоторых практических задач [13, 14].

В [15] предложена комбинированная аддитивная модель прогнозирования, в которой к ЛРФ добавляется модель авторегрессии – проинтегрированного скользящего среднего. Исследования, комбинирующие метод «Гусеница»-SSA с другими моделями можно встретить и в зарубежной литературе. Так, например, в [16] разработана подобная модель, в которой ЛРФ комбинируется с моделью авторегрессии и применяется для краткосрочного прогнозирования нагрузки иранской национальной энергосистемы. Другой пример совместного использования метода «Гусеница»-SSA с моделями Бокса-Дженкинса можно найти в [17]. Там методом «Гусеница»-SSA производится разложение исходного ВР, затем производится группировка ВР разложения на интерпретируемые составляющие, такие как трендовая, сезонные и шумовые. Далее для каждого сгруппированного ВР идентифицируется и оценивается модель ARIMA. Остаточная ошибка моделирования корректируется авторегрессионной моделью и определяются прогнозы синтезированной моделью. Примерно в то же время в [18] была предложена подобная модель, однако, в отличие от [17], для идентификации и оценивания сгруппированных ВР использовались сезонные модели авторегрессии – проинтегрированного скользящего среднего (САРПСС, *на англ.* – SARIMA), а также такая гибридная модель была обобщена уже для случая прогнозирования ВР, зависящего от нескольких экзогенных ВР. Гибридные модели на основе метода SSA, моделей ARIMA и искусственных нейронных сетей активно обсуждаются в [42–45].

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

3. Цели и задачи исследования

Целью проведенных исследований является разработка вероятностно-детерминированных и декомпозиционных математических моделей и методов их идентификации для прогнозирования нестационарных ВР со сложной структурой с учётом экзогенных факторов и проверка эффективности прогнозирования

разработанными моделями в сравнении с прогнозами, получаемыми вероятностными моделями SARIMAX, обобщёнными для прогнозирования ВР с несколькими сезонными составляющими и несколькими экзогенными переменными [1].

Для этих целей в работе использованы метод MSSA и модели SARIMAX. На их основе разработаны гибридные математические модели и методы прогнозирования нестационарных ВР с несколькими сезонными составляющими с учётом нескольких экзогенных факторов. Метод MSSA в работе применяется для разложения исходных прогнозируемого и экзогенных ВР на ВР с более простой структурой, а модели SARIMAX – для построения математических моделей данных компонент разложения и вычисления их прогнозов. Выбор различных параметров метода MSSA приводит к созданию, в сочетании с моделями SARIMAX, 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 с экзогенными переменными

В данном разделе сначала приведём краткое изложение методов и моделей прогнозирования нестационарных ВР, учитывающих внешние факторы: многомерного варианта метода «Гусеница»-SSA и модели SARIMAX, расширенной для прогнозирования ВР с учётом нескольких экзогенных переменных. Приведенные методы и модели необходимы далее для описания гибридных моделей на их основе.

4. 1. Модель авторегрессии – проинтегрированно скользящего среднего с экзогенными переменными

В [1] была предложена математическая модель многовязных нестационарных случайных процессов, позволяющая описывать широкий класс регулярных случайных процессов, содержащих полиномиальные, полигармонические и стохастические тренды и аддитивный цветной шум:

$$y_t = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)} x_{t-b_i}^{(i)} + \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^*(B)} a_t, \tag{1}$$

где y_t , $t = \overline{1, n}$ – исходный или преобразованный (нормированный или прологарифмированный) ВР; n – объём выборки; $x_t^{(i)}$, $t = \overline{1, n}$, $i = \overline{1, N}$ – исходные или преобразованные (нормированные или прологарифмированные) ВР изменения значений внешних (экзогенных) факторов; N – количество экзогенных факторов; B – оператор сдвига по времени на одну единицу назад, такой что $B^i x_t = x_{t-i}$; b_i – величина задержки i -го экзогенного ВР по времени относительно прогнозируемого ВР y_t ; $\delta_{r_i}^{(i)}(B)$ и $\omega_{c_i}^{(i)}(B)$ – полиномы передаточных функций от B степеней r_i и c_i соответственно; $\delta_{r_i}^{(i)}(B) = 1 - \delta_1^{(i)} B - \delta_2^{(i)} B^2 - \dots - \delta_{r_i}^{(i)} B^{r_i}$; $\omega_{c_i}^{(i)}(B) = \omega_0^{(i)} - \omega_1^{(i)} B - \dots - \omega_{c_i}^{(i)} B^{c_i}$; $\Phi_{p^*}^*(B)$ – обобщенный оператор авторегрессии (АР) 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 взят

тия 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 $\nabla_{S_i} y_t = (1 - B^{S_i}) y_t = y_t - y_{t-S_i}$; $\Phi_{p^*}^*(B)$ – обобщенный оператор АР 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 АР периодических компонент с периодами S_i соответственно;

$\theta_{q^*}^*(B)$ – обобщенный оператор скользящего среднего (СС) 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 СС периодических компонент с периодами S_i соответственно;

a_t – остаточные ошибки модели (1).

4. 1. 1. Проверка адекватности модели авторегрессии – проинтегрированного скользящего среднего с экзогенными переменными

Прежде чем воспользоваться моделью (1), необходимо проверить, является ли эта модель адекватной исследуемому процессу. Для этого необходимо установить, значимо ли отличаются остаточные ошибки модели от процесса "белого шума", значима ли величина взаимной корреляции между, приведенными к стационарному виду при помощи (26), процессами $x_t^{(i)}$ и a_t . Для принятия решения об адекватности модели используем метод, предложенный в работе [23].

Если структура модели (1) принята правильно и в соответствующие уравнения подставлены истинные значения параметров, остаточные ошибки будут "белым шумом", а выборочные корреляции будут распределены взаимонезависимо с нулевым средним значением и дисперсией, равной $\frac{1}{M'}$, где $M' = n' - u - p^*$,

где $n' = n - \sum_{i=1}^{n_s} D_i S_i$, $u = \max \left(\sum_{i=1}^N r_i, \sum_{i=1}^N (c_i + b_i) \right)$. Если вместо истинных значений параметров подставлены их выборочные оценки, распределение корреляций с малыми задержками искажается. В частности, дисперсия выборочных корреляций с малыми задержками может быть значительно меньше $\frac{1}{M'}$, а их значения

сильно коррелированы. Поэтому простое сравнение выборочных автокорреляций $r_{aa}(k)$ при малых k со "стандартной ошибкой" $\frac{1}{\sqrt{M'}}$ может создать ложное впечатление о малости первых. Кроме того, всплески выборочной автокорреляционной функции при малых задержках могут происходить просто из-за высокой корреляции между выборочными оценками. Если амплитуда таких всплесков мала по сравнению $\frac{1}{\sqrt{M'}}$, они могут быть чисто случайными и не свидетельствовать о каком-либо реальном поведении теоретических автокорреляций. Все это говорит о том, что предпочтительнее не рассматривать отдельные $r_{aa}(k)$, а оценить

неадекватность модели по совокупности первых автокорреляций. Можно показать [23], что если K достаточно велико, то для адекватной модели величина

$$Q = n \sum_{k=1}^K r_{aa}^2(k) \quad (2)$$

распределена примерно как χ^2 с $K - p^* - q^*$ степенями свободы, т. е. зависит только от параметров передаточной функции

$\frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)}$ и не зависит от $\frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}$. Сравнив значение Q с таблицей значений χ^2 при заданном уровне значимости, можно приближенно проверить гипотезу об адекватности полученной модели.

Для проверки адекватности модели передаточной функции $\frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}$ используем статистику, основанную на взаимной корреляционной функции $r_{\alpha^{(i)}a}(k)$ преобразованного процесса $\alpha_t^{(i)} = \Phi(B)x_t^{(i)}$ и остаточных ошибок модели, где $\Phi(B)$ – оператор метода предварительного выравнивания спектра, который позволяет преобразовать коррелированный процесс $x_t^{(i)}$ в некоррелированный – белый шум $\alpha_t^{(i)}$. Величина

$$S = M' \sum_{k=1}^K r_{\alpha^{(i)}a}^2(k) \quad (3)$$

приближенно распределена так же, как χ^2 с $K + 1 - \sum_{i=1}^N (r_i + c_i) + 1$ степенями свободы, где $\sum_{i=1}^N (r_i + c_i) + 1$ – число параметров в подгоняемой модели передаточной функции $\frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}$. Принятие решения об адекватности модели осуществляется путем сравнения значения S со значением χ^2 при заданном уровне значимости.

В автоматизированных системах вид выборочных функций $r_{\alpha^{(i)}a}(k)$ и $r_{aa}(k)$ в ряде случаев позволяет ЛПР определить причину неадекватности моделей $\frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}$, $\frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)}$ и обеспечить целенаправленный поиск адекватных моделей. В автоматических системах поиск адекватной модели может быть выполнен в ходе решения оптимизационной задачи поиска $min(S)$ и $min(Q)$ в ограниченном классе рациональных структур модели.

4. 1. 2. Вычисление прогнозов временных рядов моделей сезонной авторегрессии – проинтегрированно-скользящего среднего с экзогенными переменными

Запишем математическую модель (1) в более удобной форме:

$$\delta^*(B)y_t = \sum_{i=1}^N \omega^{(i)*}(B)x_{t-b_i}^{(i)} + \phi(B)a_t, \quad (4)$$

где

$$\delta^*(B) = \prod_{i=1}^N \delta_{r_i}^{(i)}(B)\Phi_{p^*}^+(B^{p^*}),$$

$$\omega^{(i)*}(B) = \omega_{c_i}^{(i)}(B)\Phi_{p^*}^+(B^{p^*}), \quad i = \overline{1, N},$$

$$\phi(B) = \prod_{i=1}^N \delta_{r_i}^{(i)}(B)\theta_{q^*}^*(B^{q^*}). \quad (5)$$

Предположим, что адекватная статистическая модель процесса $x_t^{(i)}$, $i = \overline{1, N}$ может быть экономично представлена в классе общей мультипликативной сезонной модели авторегрессии – проинтегрированного скользящего среднего, т. е.

$$x_t^{(i)} = \frac{\theta_{x_{q^*}^{(i)}}^*(B^{q^*})}{\Phi_{x_{p^*}^{(i)}}^+(B^{p^*})} \alpha_t^{(i)}, \quad i = \overline{1, N}, \quad (6)$$

где $\alpha_t^{(i)}$, $i = \overline{1, N}$ – процесс типа белого шума.

Подставляя (6) в (1) и обозначая

$$\lambda^{(i)}(B) = \frac{\omega_{c_i}^{(i)}(B)\theta_{x_{q^*}^{(i)}}^*(B^{q^*})B^{b_i}}{\delta_{r_i}^{(i)}(B)\Phi_{x_{p^*}^{(i)}}^+(B^{p^*})}, \quad (7)$$

$$\psi(B) = \frac{\theta_{q^*}^*(B^{q^*})}{\Phi_{p^*}^+(B^{p^*})}, \quad (8)$$

представим модель (1) в виде

$$y_t = \sum_{i=1}^N \lambda^{(i)}(B)\alpha_t^{(i)} + \psi(B)a_t = \sum_{i=1}^N \sum_{j=0}^{\infty} \lambda_j^{(i)} \alpha_{t-j}^{(i)} + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

где $\alpha_t^{(i)}$ и a_t статистически независимы $\forall i = \overline{1, N}$.

Прогноз $\hat{y}_t(l)$ величины y_{t+l} будем искать в виде суммы линейных функций вида

$$\hat{y}_t(l) = \sum_{i=1}^N \sum_{j=0}^{\infty} \lambda_{l+j}^{(i)*} \alpha_{t-j}^{(i)} + \sum_{j=0}^{\infty} \psi_{l+j}^* a_{t-j}.$$

Дисперсия ошибки прогноза с упреждением l равна

$$\begin{aligned} E[y_{t+l} - \hat{y}_t(l)]^2 &= \\ &= \sum_{i=1}^N \left([\lambda_0^{(i)}]^2 + [\lambda_1^{(i)}]^2 + \dots + [\lambda_{l-1}^{(i)}]^2 \right) \sigma_{\alpha^{(i)}}^2 + \\ &+ (1 + \psi_1^2 + \dots + \psi_{l-1}^2) \sigma_a^2 + \sum_{i=1}^N \sum_{j=0}^{\infty} (\lambda_{l+j}^{(i)} - \lambda_{l+j}^{(i)*})^2 \sigma_{\alpha^{(i)}}^2 + \\ &+ \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_a^2. \end{aligned} \quad (9)$$

Минимум выражения (9) достигается, если $\lambda_{l+j} = \lambda_{l+j}^*$, $\psi_{l+j} = \psi_{l+j}^*$. Таким образом, прогноз $\hat{y}_t(l)$ величины y_{t+l} в момент t с минимальной среднеквадратической ошибкой определяется как условное математическое ожидание y_{t+l} в момент t .

Обозначая, как и ранее, условные математические ожидания в момент времени t квадратными скобками, непосредственно из (4) получаем выражение для прогноза с упреждением l :

$$\begin{aligned} \hat{y}_t(l) &= \sum_{j=1}^{p^* + \sum_{i=1}^N r_i} \delta_j^* [y_{t+l-j}] - \\ &- \sum_{i=1}^N \sum_{j=0, i \neq j}^{p^* + c_i} \omega_j^{(i)*} [x_{t+l-b_i-j}^{(i)}] - \sum_{j=1}^{q^* + \sum_{i=1}^N r_i} \phi_j [a_{t+l-j}], \end{aligned} \quad (10)$$

где δ_j^* , $\omega_j^{(i)*}$, ϕ_j определяются из уравнения (5), а

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

$$[x_{t+j}^{(i)}] = \begin{cases} x_{t+j}^{(i)}, & j \leq 0; \\ \hat{x}_t^{(i)}(j), & j > 0, \end{cases} \quad i = \overline{1, N};$$

$$[a_{t+j}] = \begin{cases} a_{t+j}, & j \leq 0; \\ 0, & j > 0. \end{cases}$$

Прогнозы $\hat{x}_t^{(i)}(j)$ получают в соответствии с уравнением (6). Значения a_{t+j} , $j \leq 0$ могут быть взяты из (1) в три этапа.

1. Определяем значения y_t для $t \geq \sum_{i=1}^N (b_i + c_i) + 1$

$$\hat{y}_t = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)} x_{t-b_i}^{(i)} = \sum_{i=1}^N (\delta_1^{(i)} y_{t-1} + \delta_2^{(i)} y_{t-2} + \dots + \delta_{r_i}^{(i)} y_{t-r_i}) + \sum_{i=1}^N (\omega_0^{(i)} x_{t-b_i}^{(i)} - \omega_1^{(i)} x_{t-b_i-1}^{(i)} - \dots - \omega_{c_i}^{(i)} x_{t-b_i-c_i}^{(i)}).$$

Предшествующие значения y_t приравниваем к нулю.

2. Находим значения

$$\varepsilon_t = y_t - \hat{y}_t.$$

3. Непосредственно вычисление $a_t = \varepsilon_t^{(n)}$ осуществляется итерационно по формуле

$$\varepsilon_t^{(i)} = \varepsilon_t^{(i-1)} - \sum_{k=1}^{b_i} \Phi_k^{(i)} \varepsilon_{t-kS_i}^{(i-1)} + \sum_{j=1}^{q_i} \theta_j^{(i)} \varepsilon_{t-jS_i}^{(i-1)},$$

где $\varepsilon_t^{(0)} = \varepsilon_t - \mu$, $t \geq \sum_{i=1}^n (p_i + D_i) S_i$, μ – среднее значение ряда ε_t .

При вычислении дисперсии прогнозов с упреждением l необходимо определить веса $\lambda_j^{(i)}$ и ψ_j . Сначала находим параметры $\omega_0^{(i)*}$, $\omega_1^{(i)*}$, ..., $\omega_{q_i}^{(i)*}$ оператора

$$\begin{aligned} \omega_{q_i}^{(i)*}(B) &= \omega_{c_i}^{(i)}(B) \theta_{x_{q_i}^{(i)}}^* (B^{q_i^{x_i}^{(i)}}) B^{b_i} = \\ &= \omega_0^{(i)*} B^{b_i} - \omega_1^{(i)*} B^{b_i+1} - \omega_2^{(i)*} B^{b_i+2} - \dots - \omega_{q_i}^{(i)*} B^{b_i+q_i'}, \end{aligned} \quad (11)$$

где $q_i' = q_i^{x_i} + c_i$. Для этого используем формулу

$$\omega_l^{(i)*} = \begin{cases} \omega_0^{(i)}, & l = 0; \\ \sum_{j=0}^l \omega_j^{(i)} \theta_{x_{l-j}^{(i)}}^*, & 1 \leq l \leq c_i; \\ \sum_{j=0}^{c_i} \omega_j^{(i)} \theta_{x_{l-j}^{(i)}}^*, & c_i < l < q_i'. \end{cases} \quad (12)$$

Далее определяем параметры $\delta_1^{(i)*}$, $\delta_2^{(i)*}$, ..., $\delta_{p_i}^{(i)*}$ оператора

$$\begin{aligned} \delta_{p_i}^{(i)*}(B) &= \delta_{r_i}^{(i)}(B) \Phi_{x_{p_i}^{(i)}}^+ (B^{p_i^{x_i}^{(i)}}) = \\ &= 1 - \delta_1^{(i)*} B - \delta_2^{(i)*} B^2 - \dots - \delta_{p_i}^{(i)*} B^{p_i'}, \end{aligned} \quad (13)$$

где $p_i' = p_i^{x_i} - r_i$.

Для этого используем формулу

$$\delta_{p_i}^{(i)*} = - \sum_{j=0}^1 \delta_j^{(i)} \Phi_{x_{p_i-j}^{(i)}}^+, \quad (14)$$

где $\delta_0^{(i)} = -1$, $\Phi_{x_0}^+ = -1$, $\delta_j^{(i)} = 0 | j > r_i$, $\Phi_{x_l}^+ = 0 | l > p_i'$.

Вывод формул (12) и (14) достаточно прост и поэтому не приводится. Подставляя (11) и (13) в (7), имеем

$$\begin{aligned} (1 - \delta_1^{(i)*} B - \delta_2^{(i)*} B^2 - \dots - \delta_{p_i}^{(i)*} B^{p_i'}) (1 + \lambda_1^{(i)} B + \lambda_2^{(i)} B^2 + \dots) = \\ = (\omega_0^{(i)*} B^{b_i} - \omega_1^{(i)*} B^{b_i+1} - \omega_2^{(i)*} B^{b_i+2} - \dots - \omega_{q_i}^{(i)*} B^{b_i+q_i'}). \end{aligned} \quad (15)$$

Если $b_i = 0$, то из (15) получаем

$$\lambda_j^{(i)} = \begin{cases} \omega_0^{(i)*}, & j = 0; \\ \sum_{l=1}^j \delta_l^{(i)*} \lambda_{j-l}^{(i)} - \omega_j^{(i)*}, & j \geq 1, \end{cases} \quad (16)$$

где $\delta_j^{(i)*} = 0 | j > p_i'$, $\omega_j^{(i)*} = 0 | j > q_i'$ $\forall i = \overline{1, N}$.

Если $b_i > 0$, то

$$\lambda_j^{(i)} = \begin{cases} 0, & j < b_i; \\ \omega_0^{(i)*}, & j = b_i; \\ \sum_{l=1}^{j-b_i} \delta_l^{(i)*} \lambda_{j-l}^{(i)} - \omega_{j-b_i}^{(i)*}, & j > b_i, \end{cases} \quad (17)$$

где $\delta_j^{(i)*} = 0 | j > p_i'$, $\omega_j^{(i)*} = 0 | j > q_i'$.

Вес оператора (8) находим по формуле

$$\psi_j = \begin{cases} 1, & j = 1; \\ \sum_{i=1}^j \Phi_i^+ \psi_{j-i} - \theta_j^*, & j > 1, \end{cases} \quad (18)$$

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

4. 1. 3. Вычисление дисперсии прогнозов

Значение весов $\lambda_j^{(i)}$ и ψ_j позволяет найти дисперсию прогноза с упреждением l . При вычислении прогнозов по уравнению (10) дисперсия прогнозов (9) с упреждением l равна

$$V(l) = \sum_{i=1}^N \sum_{j=1}^{l-1} [\lambda_j^{(i)}]^2 \sigma_{\alpha^{(i)}}^2 + \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2, \quad (19)$$

где веса $\lambda_j^{(i)}$ и ψ_j могут быть получены в соответствии с уравнениями (16)–(18). Значения оценок дисперсии $\sigma_{\alpha^{(i)}}^2$ и σ_a^2 получают в результате решения задач параметрической оптимизации моделей (6), (1).

4. 1. 4. Коррекция прогнозов

Знание весов $\lambda_j^{(i)}$ и ψ_j позволяет не только вычислять дисперсию прогнозов, но и произвести их коррек-

цию внутри интервала упреждения L^y . При вводе в систему новых значений реализаций процессов y_{t+1} и $x_{t+1}^{(i)}$ можно вычислить фактические ошибки прогнозов и осуществить их коррекцию в момент времени $t+1$ для упреждений $l=1, 2, \dots, L^y - 1$.

Коррекция прогнозов, вычисляемых в соответствии с уравнением (10), несколько сложнее. Для получения формулы коррекции прогнозов представим уравнение (10) в виде

$$\hat{y}_t(l) = \sum_{i=1}^N \sum_{j=0}^{\infty} \lambda_{l+j}^{(i)} \alpha_{t-j}^{(i)} + \sum_{j=0}^{\infty} \psi_{l+j} a_{t-j}.$$

В этом случае

$$\hat{y}_{t+1}(l) = \sum_{i=1}^N (\lambda_{l+1}^{(i)} \alpha_t^{(i)} + \lambda_{l+2}^{(i)} \alpha_{t-1}^{(i)} + \dots) + \psi_{l+1} \alpha_{t+1} + \psi_{l+2} \alpha_t + \dots; \quad (20)$$

$$\hat{y}_{t+1}(l+1) = \sum_{i=1}^N (\lambda_{l+1}^{(i)} \alpha_t^{(i)} + \lambda_{l+2}^{(i)} \alpha_{t-1}^{(i)} + \lambda_{l+3}^{(i)} \alpha_{t-2}^{(i)} + \dots) + \psi_{l+1} \alpha_t + \psi_{l+2} \alpha_{t-1} + \dots; \quad (21)$$

Вычитая (21) из (20), имеем

$$\hat{y}_t(l) = \hat{y}_t(l+1) + \sum_{i=1}^N \lambda_1^{(i)} \alpha_t^{(i)} + \psi a_{t+1}. \quad (22)$$

Это уравнение означает, что прогноз величины y_{t+1} , вычисляемый в момент времени $t+1$, равен прогнозу этой величины в момент времени t плюс ошибки прогноза на шаг вперед и a_{t+1} с коэффициентами $\lambda_j^{(i)}, \psi_j$.

Уравнение (22) в дальнейшем используем для коррекции прогнозов. Величина $\alpha_{t+1}^{(i)}$ по определению равна

$$\alpha_{t+1}^{(i)} = x_{t+1}^{(i)} - \hat{x}_t^{(i)}(1). \quad (23)$$

a_{t+1} находим с помощью уравнения

$$a_{t+1} = \tilde{y}_{t+1} - \hat{y}_t(1), \quad (24)$$

где

$$\tilde{y}_{t+1} = y_{t+1} - \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)} x_{t+1}^{(i)}.$$

Из предположения, что ошибки $\alpha^{(i)}$ и a подчиняются нормальному закону распределения, следует, что при известных значениях процессов $x_t^{(i)}$ и y_t до момента времени t условное распределение вероятности $P(y_{t+1} | y_t, y_{t-1}, \dots, x_t^{(i)}, x_{t-1}^{(i)})$ будущего значения процесса y_{t+1} будет также нормальным с математическим ожиданием $\hat{y}_t(1)$ и дисперсией $V(1)$. Вероятностные пределы $y_{t+1}(t)$ для величины y_{t+1} имеют вид

$$y_{t+1}(t) = \hat{y}_t(1) \pm u_{\varepsilon/2} \sqrt{V(1)}. \quad (25)$$

где $u_{\varepsilon/2}$ квантиль уровня $1 - \varepsilon/2$ стандартного нормального распределения. При вычислении прогнозов по (10) значения $V(1)$ в (25) вычисляются соответственно по (19). Численные значения $u_{\varepsilon/2}$ выбираются

равными 0.68, 1.65, 1.96 или 2.58 в зависимости от того, находятся ли будущие значения y_{t+1} в интервале $\hat{y}_t(1) \pm u_{\varepsilon/2} \sqrt{V(1)}$ с вероятностью 0.5, 0.9, 0.95 или 0.99 соответственно.

4. 1. 5. Алгоритм построения модели Бокса-Дженкина, расширенной для прогнозирования временных рядов с несколькими сезонными составляющими и с учётом нескольких экзогенных переменных

Данный алгоритм основан на использовании в качестве математической модели ВР выражения (1). Блок-схема алгоритма представлена на рис. 1, 2.

Алгоритм состоит из 29 этапов:

- 1) начало;
- 2) ввод обучающих выборок $y_t, x_t^{(i)}, i = \overline{1, N}, t = \overline{1, n}$;
- 3) идентификация периодов S_1, S_2, \dots, S_{n_s} и порядка взятий разностей D_1, D_2, \dots, D_{n_s} . На основании анализа вида автокорреляционной функции процесса определяем порядок взятия разности с периодом $S_{n_s} = 1$. Используя алгоритм выявления скрытой периодичности [24], определяем периоды периодических компонент $S_1, S_2, \dots, S_{n_s-1}$. На основании анализа автокорреляционных функций процессов

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

определяем порядки взятия разностей $D_i, i = \overline{1, n_s - 1}$;

- 4) преобразование рядов y_t и $x_t^{(i)}, i = \overline{1, N}$. По известным значениям $D_i, S_i, i = \overline{1, n_s}$ осуществляем преобразование исходных выборок $y_t, x_t^{(i)}, i = \overline{1, N}$ в

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

$$x_t^{(i)'} = \nabla_{S_1}^{D_1} \nabla_{S_2}^{D_2} \dots \times \nabla_{S_{n_s}}^{D_{n_s}} x_t^{(i)}; \quad (26)$$

- 5) построена ли модель ряда x'_t ? Если «да», то переходим к 10-му этапу, если «нет» – к 6-му этапу.

Этапы 6–9-й соответствуют этапам 4–7-му алгоритма, приведенного в [4] и обеспечивают построение адекватной модели процесса $x_t^{(i)'}, i = \overline{1, N}$. Идентификацию модели процесса $x_t^{(i)'}, i = \overline{1, N}$ осуществляем в классе общей мультипликативной модели авторегрессии – проинтегрированного скользящего среднего, однако, можно использовать и другие модели, представленные в [4];

- 10) надо ли вычислять прогнозы $\hat{x}_t^{(i)}(l), i = \overline{1, N}$? Если «да», то переходим к 11-му этапу, если «нет» – к 12-му этапу;

- 11) модель $x_t^{(i)}, i = \overline{1, N}$ адекватна? Если «да», то переходим к 13-му этапу, если «нет» – к 8-му этапу;

- 12) вычисление прогнозов $\hat{x}_t^{(i)}(l), l = \overline{1, L^{x^{(i)}}}, i = \overline{1, N}$. Осуществляем в соответствии с разностным уравнением, приведенным в [1];

- 13) ввод прогнозов $\hat{x}_t^{(i)}(l), l = \overline{1, L^{x^{(i)}}}, i = \overline{1, N}$;

- 14) модель y_t адекватна? Если «да», то переходим к 19-му этапу, если «нет» – к 15-му этапу;

- 15) идентификация параметров моделей $\frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}, i = \overline{1, N}$ ($V_1(B) = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)}$). Производится методом выравнивания спектра компоненты процесса $y_t^{(i)}, i = \overline{1, N}$ в соответствии с видом и значениями параметров

модели процесса $x_t^{(i)}$, $i=1, \overline{N}$. Идентификация параметров передаточной функции $V_3(B) = \frac{\theta_q^*(B)}{\Phi_p^+(B)}$ выполняется в соответствии с алгоритмом, приведенным в [4];

16) грубая оценка параметров передаточной функции. Осуществляется путём решения системы линейных уравнений, представленных в [1]. В качестве грубых оценок параметров модели передаточной функции $\frac{\theta_q^*(B)}{\Phi_p^+(B)}$ используем константы типа $\pm 0,1$;

17) точная оценка параметров модели. Выполняется в соответствии с алгоритмом параметрической оптимизации Марквардта [1], Левенберга-Марквардта либо другими методами;

18) модель y_t адекватна? Если «да», то переходим к 10-му этапу, если «нет» – к 15-му этапу. Адекватность проверяем в соответствии со значениями статистик, определяемых уравнениями (2), (3);

19) вычисление прогнозов $\hat{y}_t(1)$, $l=1, L^y$. Осуществляем по уравнению (10). Дисперсию и вероятностные пределы прогнозов выполняем в соответствии с уравнением (19), (25);

20) печать прогнозов $\hat{y}_t(1)$, $l=1, L^y$;

21) ввод фактических значений реализаций процессов y_{t+i} и $x_{t+j}^{(i)}$, $i=1, \overline{N}$, в $t+j$ момент времени;

22) вычисление остаточных ошибок моделей $\alpha_{t+j}^{(i)}$, $j=1, \overline{L^{x^{(i)}}}$, $i=1, \overline{N}$ и a_{t+j} , $j=1, \overline{L^y}$ выполняется по уравнениям (23), (24);

23) проверка условия: если $i < L^y$, то переходим к 24-му этапу, в противном случае – к 26-му этапу;

24) коррекция прогнозов $y_{t+i}(1)$, $l=i+1, L^y$. Осуществляется в соответствии с (22);

25) печать скорректированных прогнозов $\hat{y}_{t+i}(1)$, $l=i+1, L^y$;

26) проверка условия «Продолжать ли вычисления?». Если «да», то переходим к 27-му этапу, если «нет» – к 29-му этапу;

27) формирование новых обучающих выборок y_t и $x_t^{(i)}$, $i=1, \overline{N}$ и соответствующих массивов остаточных

ошибок моделей a_t и $\alpha_t^{(i)}$, $i=1, \overline{N}$. Проводится путём отбрасывания их L^y и $L^{x^{(i)}}$, $i=1, \overline{N}$ соответственно старых значений и прибавлением L^y и $L^{x^{(i)}}$, $i=1, \overline{N}$ новых значений этих рядов;

28) модель y_t адекватна? Если «да», то переходим к 10-му этапу, если «нет», то, предполагая, что новые данные привели к изменению только параметров модели, переходим к 17-му этапу. Если новые данные изменили стохастическую структуру моделей

$\frac{\omega_{c_i}^{(i)}(B)}{\delta_{\tau_i}^{(i)}(B)} x_{t-b_i}^{(i)}$, $i=1, \overline{N}$ или $\frac{\theta_q^*(B)}{\Phi_p^+(B)} a_t$, то это будет обнаружено на 18-м этапе;

29) окончание алгоритма.

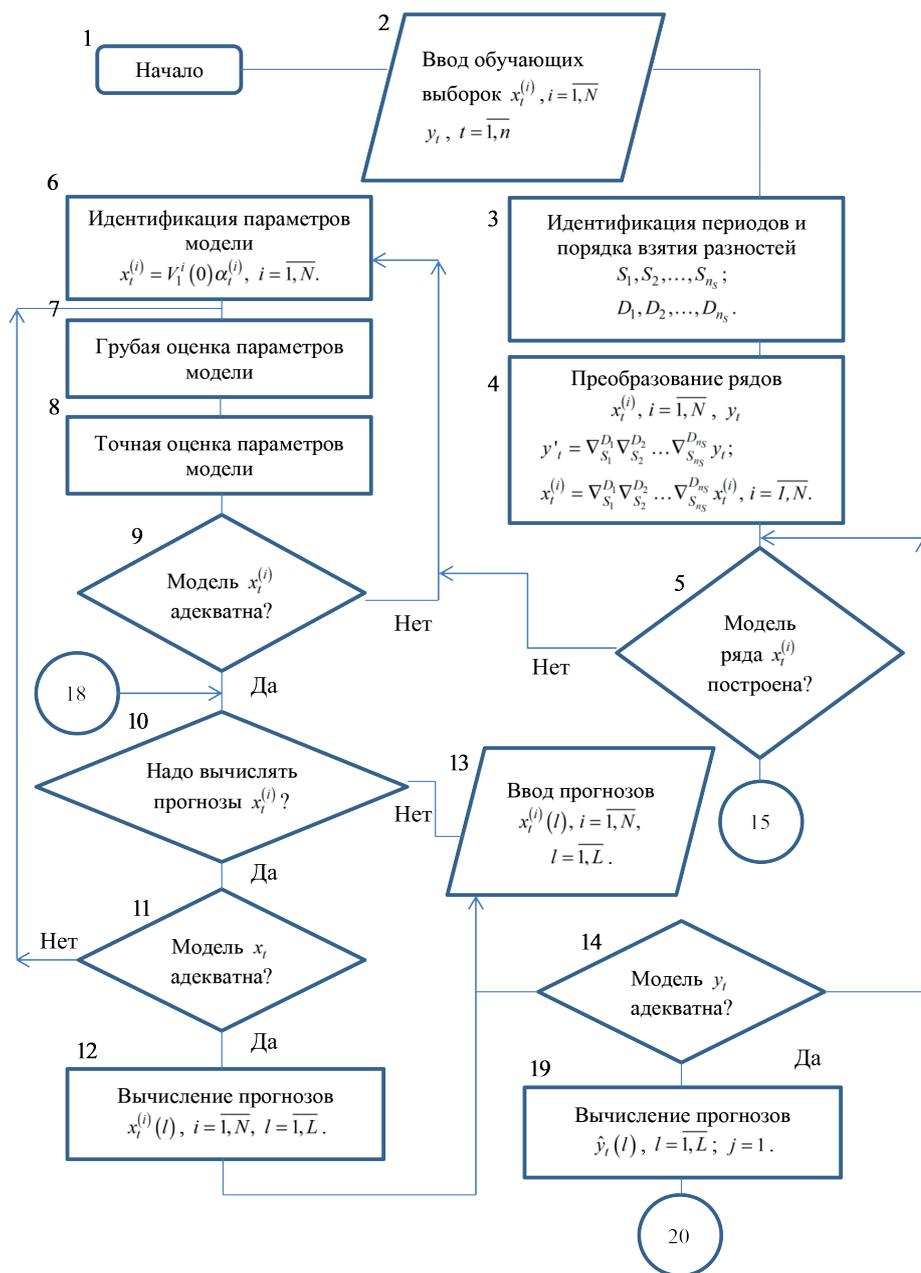


Рис. 1. Блок-схема адаптивного алгоритма прогнозирования BP моделями SARIMAX, расширенными для случая прогнозирования BP с несколькими сезонными компонентами и с учётом нескольких экзогенных переменных

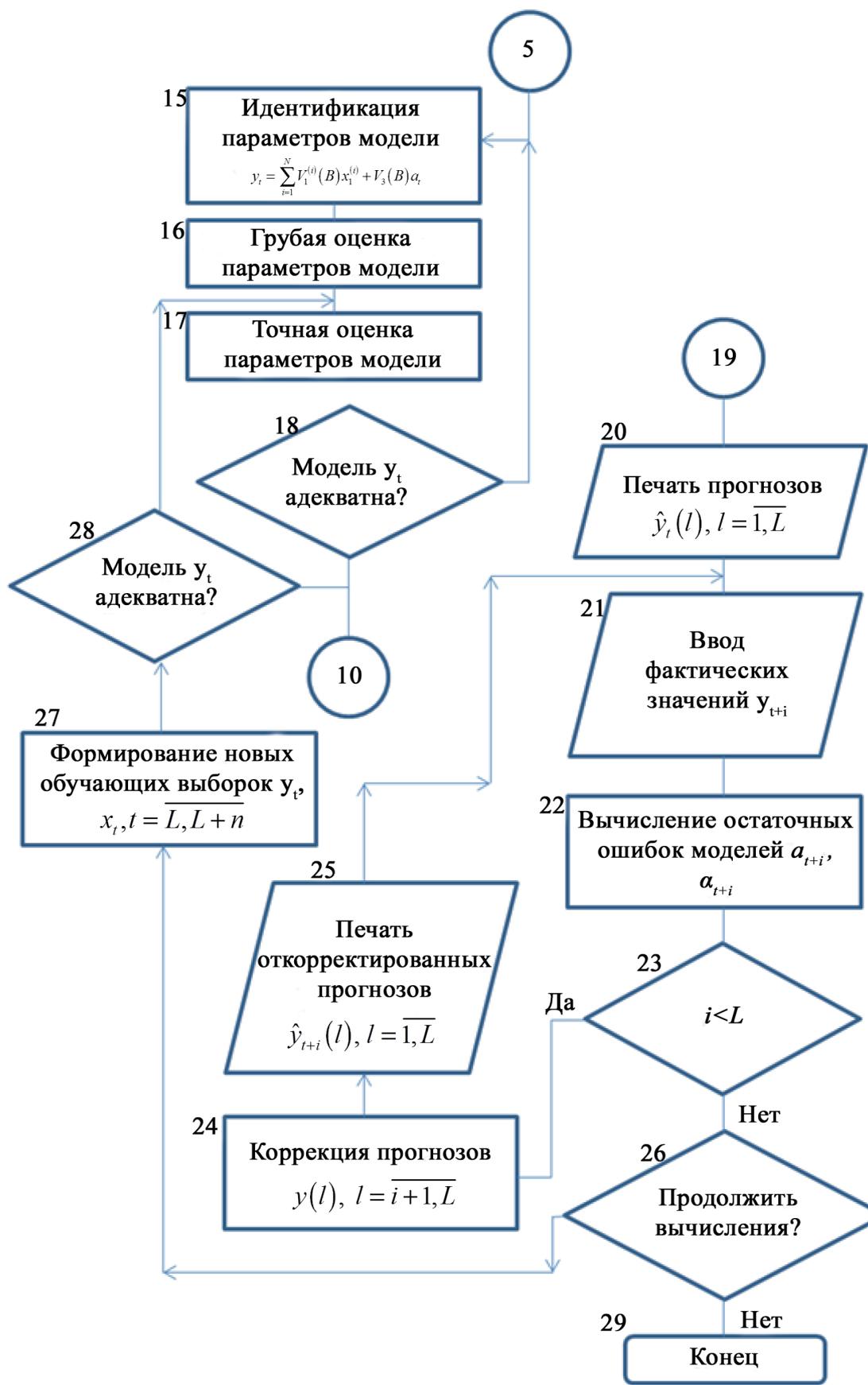


Рис. 2. Блок-схема адаптивного алгоритма прогнозирования ВР моделями SARIMAX, расширенными для случая прогнозирования ВР с несколькими сезонными компонентами и с учётом нескольких экзогенных переменных (продолжение)

Алгоритм повышает точность прогнозов ВР в сравнении с алгоритмом для построения моделей SARIMA [4], но требует значительно больших затрат машинного времени и дополнительной информации о внешних факторах.

4. 2. Математическая модель прогнозирования временных рядов с учётом внешних факторов и с локальной нерегулярностью

Необходимость выполнения предположений о стационарности модели АР и обратимости модели СС приводит к тому, что корни соответствующих характеристических полиномов должны быть вещественными и лежать вне или на единичной окружности, а соответствующие параметры моделей должны находиться в диапазоне от -1 до +1 [23]. Процессы, удовлетворяющие этим условиям, будем называть регулярными. В докладе рассматривается более широкий класс взаимосвязанных случайных процессов, как с локальной временной нерегулярностью воздействия внешних факторов, так и с локальной нерегулярностью по частоте, амплитуде и фазе для полигармонических трендов. Показано, что построение адекватных математических моделей таких процессов возможно в классе моделей вида (27) с дополнительным полиномом $g_{L-1}^{*a}(B) = 1 - g_{L-1}^a(B)$ от оператора задержки B , коэффициенты которого определяются как коэффициенты

линейной рекуррентной формулы $g_{L-1}^a(B)a_t = \sum_{i=1}^{L-1} g_i^a a_{t-i}$ метода «Гусеница»-SSA [25], применяемого к остаточным ошибкам a_t модели (1). Предложенная модель имеет следующий вид:

$$y_t^{SARIMAX-SSA} = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_i^{(i)}(B)} x_{t-b_i}^{(i)} + \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^*(B)} \cdot \frac{1}{g_{L-1}^{*a}(B)} e_t, \quad (27)$$

где e_t – остаточные ошибки модели (27), L – параметр, называемый длиной окна метода «Гусеница»-SSA [2]. При этом корни характеристического полинома $g_{L-1}^{*a}(B) = 0$ могут быть как вещественными, так и комплексными и лежать как снаружи, так и внутри единичной окружности [2, 25].

Вернёмся к рассмотрению модели ЛРФ-SSA+ SARIMA [4]. В более сложных случаях лучшие результаты прогнозирования можно получить, используя последовательный метод SSA, в котором после выделения одной из компонент с некоторым L_1 анализируется остаток (снова методом SSA) с возможно уже другой длиной окна L_2 и т. д. [2]. Тогда гибридная модель на основе последовательного метода SSA и моделей SARIMA будет иметь вид:

$$\epsilon_t^{(1)} = y_t - w_t^{(1)}, \quad t = \overline{1, n};$$

$$w_t^{(1)} = \sum_{i=1}^{L_1-1} g_i^{(1)} w_{t+1-i}^{(1)};$$

$$\epsilon_t^{(2)} = \epsilon_t^{(1)} - w_t^{(2)}, \quad t = \overline{1, n};$$

$$w_t^{(2)} = \sum_{i=1}^{L_2-1} g_i^{(2)} w_{t+1-i}^{(2)};$$

$$\epsilon_t^{(3)} = \epsilon_t^{(2)} - w_t^{(3)}, \quad t = \overline{1, n};$$

$$w_t^{(3)} = \sum_{i=1}^{L_3-1} g_i^{(3)} w_{t+1-i}^{(3)};$$

...

$$\epsilon_t^{(n_{seq})} = \epsilon_t^{(n_{seq}-1)} - w_t^{(n_{seq})}, \quad t = \overline{1, n};$$

$$w_t^{(n_{seq})} = \sum_{i=1}^{L_{n_{seq}}-1} g_i^{(n_{seq})} w_{t+1-i}^{(n_{seq})};$$

$$\epsilon_t^{(n_{seq})} = \frac{\theta_{q^{seq^*}}^{seq^*}(B)}{\Phi_{p^{seq^*}}^{seq^*}(B)} a_t^{seq^*};$$

$$\hat{y}_t^{seq-SSA}(1) = \sum_{i=1}^{n_{seq}} \hat{w}_t^{(n_{seq})}(1) + \hat{\epsilon}_t^{(n_{seq})}(1). \quad (28)$$

В мультипликативном виде данная модель запишется следующим образом:

$$y_t^{seq-SSA^*} = \sum_{i=1}^{n_{seq}} g_{L_i-1}^{(i)}(B) w_t^{(i)} + \frac{\theta_{q^{seq^*}}^{seq^*}(B)}{\Phi_{p^{seq^*}}^{seq^*}(B)} a_t^{seq^*}, \quad (29)$$

или

$$y_t^{seq-SSA^{**}} = \sum_{i=1}^{n_{seq}} g_{L_i-1}^{(i)}(B) \epsilon_t^{(i-1)} + \frac{\theta_{q^{seq^{**}}}^{seq^{**}}(B)}{\Phi_{p^{seq^{**}}}^{seq^{**}}(B)} a_t^{seq^{**}}, \quad (30)$$

или

$$\begin{aligned} &g_{L_1-1}^{(1)}(B) g_{L_2-1}^{(2)}(B) \dots g_{L_{n_{seq}}-1}^{(n_{seq})}(B) y_t^{seq-SSA^{***}} = \\ &= \frac{\theta_{q^{seq^{***}}}^{seq^{***}}(B)}{\Phi_{p^{seq^{***}}}^{seq^{***}}(B)} a_t^{seq^{***}}, \end{aligned} \quad (31)$$

где $g_{L_i-1}^{(i)}(B) = \sum_{j=1}^{L_i-1} g_j^{(i)} B^j$, $g_{L_i-1}^{(i)}(B)$, $i = \overline{1, n_{seq}}$ – полином от введенного выше оператора задержки B , коэффициенты $g_j^{(i)}$, $j = \overline{1, L_i-1}$, $i = \overline{1, n_{seq}}$ которого равны коэффициентам ЛРФ $w_t^{(i)} = \sum_{j=1}^{L_i-1} g_j^{(i)} w_{t-i}^{(i)}$, $i = \overline{1, n_{seq}}$ метода «Гусеница»-SSA; $w_t^{(i)}$, $i = \overline{1, n_{seq}}$ – компонента, определяемая на этапе диагонального усреднения метода «Гусеница»-SSA, применяемого к ВР $\epsilon_t^{(i-1)}$, $i = \overline{1, n_{seq}}$, $\epsilon_t^{(0)} = y_t$, $t = \overline{1, n}$.

Запишем (31) в следующем виде:

$$y_t^{seq-SSA^{***}} = \frac{\theta_{q^{seq^{***}}}^{seq^{***}}(B)}{g_{L_{n_{seq}}}^{*seq}(B) \Phi_{p^{seq^{***}}}^{seq^{***}}(B)} a_t^{seq^{***}}, \quad (32)$$

где

$$g_{L_{n_{seq}}}^{*seq}(B) = g_{L_1-1}^{(1)}(B) g_{L_2-1}^{(2)}(B) \dots g_{L_{n_{seq}}-1}^{(n_{seq})}(B), \quad L_{n_{seq}}^* = \sum_{i=1}^{n_{seq}} (L_i - 1).$$

Назовём такую модель seq-SSA.

Тогда гибридная модель (27) с использованием последовательного метода SSA примет следующий вид:

$$y_t^{\text{SARIMAX-Seq-SSA}} = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{f_i}^{(i)}(B)} X_{t-b_i}^{(i)} + \frac{\theta_{q^*}^*(B)}{\Phi_{p^*}^+(B)} \cdot \frac{1}{g_{L_{\text{seq}}}^{*\text{seq}}(B)} a_t^{\text{seq}}. \quad (33)$$

Назовём такую модель SARIMAX – Seq-SSA.

4. 3. Многомерный вариант метода «Гусеница»-SSA

В этом разделе рассматривается многомерное обобщение метода «Гусеница»-SSA и приводятся алгоритмы двух вариантов прогнозирования (многомерного продолжения) взаимосвязанных нестационарных ВР.

4. 3. 1. Алгоритм анализа временных рядов многомерным вариантом метода «Гусеница»-SSA

Пусть требуется получить прогноз N+1 взаимосвязанных нестационарных ВР $Y=y_1, y_2, \dots, y_{n_y}$ и $X^{(i)}=x_1^{(i)}, x_2^{(i)}, \dots, x_{n_x}^{(i)}, i=1, N$ произвольных длин n_y и $n_{x^{(i)}}, i=1, N$ соответственно.

1. Вложение

Выбираем длину окна L и строим траекторную матрицу

$$\mathbf{X} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_{K_y} & X_1^{(1)} & X_2^{(1)} & \dots \\ \dots & X_{K_{x^{(1)}}}^{(1)} & X_1^{(2)} & X_2^{(2)} & \dots & X_{K_{x^{(2)}}}^{(2)} & \dots \\ \dots & X_1^{(N)} & X_2^{(N)} & \dots & X_{K_{x^{(N)}}}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathbf{Y} & \mathbf{X}^{(1)} & \mathbf{X}^{(2)} & \dots & \mathbf{X}^{(N)} \end{bmatrix}$$

из векторов вложения $X_j^{(i)} = (x_j^{(i)}, x_{j-1}^{(i)}, \dots, x_{j+L-1}^{(i)})^T, 1 \leq j \leq K_{x^{(i)}}, K_{x^{(i)}} = n_{x^{(i)}} - L + 1, i = 1, N$. Здесь \mathbf{Y} – траекторная матрица ряда $Y, \mathbf{X}^{(i)}, i = 1, N$ – траекторная матрица ряда $X^{(i)}, i = 1, N$.

2. Сингулярное разложение

Сформируем матрицу $S = \mathbf{X}\mathbf{X}^T$ и произведём сингулярное разложение траекторной матрицы \mathbf{X} ВР.

Обозначим:

- $\lambda_1, \lambda_2, \dots, \lambda_L$ – собственные числа матрицы S, взятые в порядке убывания ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$);
- U_1, U_2, \dots, U_L – ортонормированная система собственных векторов матрицы S, соответствующих этим собственным числам.

Произведём разложение траекторной матрицы

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad (34)$$

где

$$\mathbf{X}_j = \sqrt{\lambda_j} U_j U_j^T; \quad V_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{X}^T U_j;$$

$$j=1, 2, \dots, d; \quad d = \max\{i | \lambda_i > 0\}.$$

3. Группировка

Разложение (34) в сгруппированном виде может быть записано следующим образом:

$$\mathbf{X} = \mathbf{X}_{i_1} + \mathbf{X}_{i_2} + \dots + \mathbf{X}_{i_\gamma}, \quad (35)$$

где $\mathbf{X}_{I_j} = \mathbf{X}_{i_{j1}} + \mathbf{X}_{i_{j2}} + \dots + \mathbf{X}_{i_{j\gamma_j}}; I_j = \{i_{j1}, i_{j2}, \dots, i_{j\gamma_j}\}, j = \overline{1, \gamma}; I_j$ – непересекающиеся подмножества множества индексов $\{1, 2, \dots, d\}$.

4. Диагональное усреднение

Матрицы $\mathbf{X}_{I_j}, j = \overline{1, \gamma}$ сгруппированного разложения переводятся в систему новых рядов длины n. Для этого они разбиваются следующим образом

$\mathbf{X}_{I_j} = [\mathbf{Y}_{I_j} \quad \mathbf{X}_{I_j}^{(1)} \quad \mathbf{X}_{I_j}^{(2)} \quad \dots \quad \mathbf{X}_{I_j}^{(N)}]$. Далее производится диагональное усреднение каждой из матриц \mathbf{Y}_{I_j} и $\mathbf{X}_{I_j}^{(k)}, k=1, 2, \dots, N, j = \overline{1, \gamma}$, преобразуя их в ВР \tilde{Y}_{I_j} и $\tilde{X}_{I_j}^{(k)}, k=1, 2, \dots, N, j = \overline{1, \gamma}$ соответственно. В результате каждая матрица \mathbf{X}_{I_j} порождает многомерный ВР $(\tilde{Y}_{I_j} \quad \tilde{X}_{I_j}^{(1)} \quad \tilde{X}_{I_j}^{(2)} \quad \dots \quad \tilde{X}_{I_j}^{(N)})$, $j = \overline{1, \gamma}$ – восстановленную аддитивную компоненту исходного ряда $(Y \quad X^{(1)} \quad X^{(2)} \quad \dots \quad X^{(N)})$.

4. 3. 2. Прогнозирование временных рядов многомерным вариантом метода «Гусеница»-SSA

В теории одномерного метода «Гусеница»-SSA вычисление прогноза ВР выполняется непосредственно по линейной рекуррентной формуле, которую порождает траекторное пространство ряда. Траекторная матрица многомерного ВР несимметрична, поэтому далее рассмотрим два варианта метода «Гусеница»-SSA для прогнозирования многомерных ВР.

Пусть имеется N+1 ВР $Y, X^{(1)}, X^{(2)}, \dots, X^{(N)}$ одинаковой длины n. Выберем длину окна $1 < L < n$.

L-продолжение многомерного ВР $(Y_{n+1} \quad X_{n+1}^{(1)} \quad X_{n+1}^{(2)} \quad \dots \quad X_{n+1}^{(N)})$ в пространстве столбцов траекторной матрицы имеет вид:

$$\mathbf{R}_{N+1} = \mathbf{Y}_\Delta \mathbf{S}_L, \quad (36)$$

где

$$\mathbf{S}_L = \frac{1}{1-v^2} \sum_{j=1}^d \pi_j U_j^v \in \mathbf{R}^{L-1}; \quad (37)$$

U_1, U_2, \dots, U_d – левые сингулярные вектора сингулярного разложения траекторной матрицы ряда;

X^v – вектор, состоящий из первых L – 1 компонент вектора X;

π_j – последняя координата вектора U_j .

$$v^2 = \pi_1^2 + \dots + \pi_d^2;$$

$$\mathbf{Y}_\Delta = \begin{bmatrix} y_{n-L+2} & y_{n-L+3} & \dots & y_n \\ x_{n-L+2}^{(1)} & x_{n-L+3}^{(1)} & \dots & x_n^{(1)} \\ x_{n-L+2}^{(2)} & x_{n-L+3}^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-L+2}^{(N)} & x_{n-L+3}^{(N)} & \dots & x_n^{(N)} \end{bmatrix}$$

– матрица из последних L–1 значений ВР.

K-продолжение многомерного ВР $(Y_{n+1} \quad X_{n+1}^{(1)} \quad X_{n+1}^{(2)} \quad \dots \quad X_{n+1}^{(N)})$ в пространстве строк траекторной матрицы имеет вид:

$$\mathbf{R}_{n+1} = (\mathbf{I}_{(N+1)(N+1)} - \mathbf{W}\mathbf{W}^T)^{-1} \mathbf{W}\mathbf{U}^T \mathbf{Z}, \quad (38)$$

где $X^{V(N+1)}$ – вектор, состоящий из всех компонент вектора X , кроме элементов с номерами, кратными K ; V_1, V_2, \dots, V_d – правые сингулярные вектора сингулярного разложения траекторной матрицы ряда;

$$W = \begin{pmatrix} \pi_1^y & \pi_2^y & \dots & \pi_d^y \\ \pi_1^{(1)} & \pi_2^{(1)} & \dots & \pi_d^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \dots & \pi_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1^{(N)} & \pi_2^{(N)} & \dots & \pi_d^{(N)} \end{pmatrix};$$

$$U = (U_1^{V(N+1)}, U_2^{V(N+1)}, \dots, U_d^{V(N+1)});$$

$$Z^y = (y_{n-K+2} \quad y_{n-K+3} \quad \dots \quad y_n)^T,$$

$$Z^{(i)} = (x_{n-K+2}^{(i)} \quad x_{n-K+3}^{(i)} \quad \dots \quad x_n^{(i)})^T,$$

$$i=1, 2, \dots, N;$$

$Z = (Z^y \quad Z^{(1)} \quad Z^{(2)} \quad \dots \quad Z^{(N)})^T$ – вектор последних значений всей совокупности рядов.

4. 4. Гибридные математические модели на основе многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX

В данном разделе предложено, используя формулы L - и K -продолжения формировать передаточные функции моделей (1) и (27).

При использовании формулы L -продолжения матрица Y_Δ формируется следующим образом:

$$Y_\Delta = \begin{pmatrix} y_{n-L+2} & y_{n-L+3} & \dots & y_n \\ x_{n+b_1-L+2}^{(1)} & x_{n+b_1-L+3}^{(1)} & \dots & \hat{x}_{n+b_1}^{(1)} \\ x_{n+b_2-L+2}^{(2)} & x_{n+b_2-L+3}^{(2)} & \dots & \hat{x}_{n+b_2}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+b_N-L+2}^{(N)} & x_{n+b_N-L+3}^{(N)} & \dots & \hat{x}_{n+b_N}^{(N)} \end{pmatrix},$$

где $\hat{x}_j^{(i)}$ – прогнозные значения BP $x_t^{(i)}$ в момент времени $j > n$. При использовании же формулы для K -продолжения векторы $Z^{(i)}$, $i=1, \overline{N}$ приобретают следующий вид $Z^{(i)} = (x_{n+b_{x^{(i)}}-K+2}^{(i)} \quad x_{n+b_{x^{(i)}}-K+3}^{(i)} \quad \dots \quad \hat{x}_{n+b_{x^{(i)}}}^{(i)})^T$, $i=1, \overline{N}$.

Прогнозные значения для каждого конкретного BP $x_t^{(i)}$, $i=1, \overline{N}$ определяются одномерным методом «Гусеница»-SSA [25], моделями SARIMA [1], их комбинациями моделями ЛРФ-SSA-SARIMA, Тренд-SSA-SARIMA [4] или декомпозиционными моделями SSA(L)-SARIMA, SSA-SARIMA, SSA-SARIMA+ SARIMA [4].

Гибридная математическая модель на основе выражения L -продолжения и модели SARIMA запишется следующим образом:

– в аддитивной форме:

$$\epsilon_t = y_t - y_t^{L-MSSA}, \quad t = \overline{1, n};$$

$$\epsilon_t = \frac{\theta_{q^{*'}}^{*'}(B)}{\Phi_{p^{*'}}^{*'}(B)} a_t^{*'};$$

$$\hat{y}_t^{L-MSSA-SARIMA}(1) = \hat{y}_t^{L-MSSA}(1) + \hat{\epsilon}_t(1); \tag{39}$$

– в мультипликативной форме:

$$y_t^{L-MSSA-SARIMA} = \frac{\theta_{q^{*''}}^{*''}(B)}{S_{L-1}(B)\Phi_{p^{*''}}^{*''}(B)} a_t^{*''}, \tag{40}$$

где $S_{L-1}(B) = 1 - [S_L]_{L-1} B - [S_L]_{L-2} B^2 - \dots - [S_L]_1 B^{L-1}$. Данное обозначение этой модели L -MSSA – SARIMA.

Теперь рассмотрим формулу K -продолжения (38). Сделаем следующие обозначения:

$$\Psi = (I_{(N+1)(N+1)} - WW^T)^{-1} WU^T; \tag{41}$$

$$\Psi_{K-1}^y(B) = 1 - \Psi_{K-1} B - \Psi_{K-2} B^2 - \dots - \Psi_1 B^{K-1};$$

$$\Psi_{K-1}^{x^{(i)}}(B) = \Psi_{(i+1)(K-1)} B - \Psi_{(i+1)(K-1)} B^2 - \dots - \Psi_{(i+1)(K-1)-(K-2)} B^{K-1},$$

$$i = \overline{1, N}.$$

Тогда выражение для гибридной модели на основе метода K -MSSA и модели SARIMA будет иметь вид:

– в аддитивной форме:

$$\epsilon_t = y_t - y_t^{K-MSSA}, \quad t = \overline{1, n};$$

$$\epsilon_t = \frac{\theta_{q^{*''''}}^{*''''}(B)}{\Phi_{p^{*''''}}^{*''''}(B)} a_t^{*''''};$$

$$\hat{y}_t^{K-MSSA-SARIMA}(1) = \hat{y}_t^{K-MSSA}(1) + \hat{\epsilon}_t^{K-MSSA}(1); \tag{42}$$

– в мультипликативной форме:

$$y_t^{K-MSSA-SARIMAX} = \frac{\sum_{i=1}^N \Psi_{K-1}^{x^{(i)}}(B)}{\Psi_{K-1}^y(B)} x_{t-b_i}^{(i)} + \frac{\theta_{q^{*IV}}^{*IV}(B)}{\Phi_{p^{*IV}}^{*IV}(B)} a_t^{*IV}. \tag{43}$$

Обозначим эту модель K -MSSA – SARIMAX.

Для модели (27) выражения, используемые для формирования передаточной функции формулы L -продолжения примут вид:

– в аддитивной форме:

$$\epsilon_t = y_t - y_t^{L-MSSA}, \quad t = \overline{1, n};$$

$$\epsilon_t = \frac{\theta_{q^{*a'}}^{*a'}(B)}{\Phi_{p^{*a'}}^{*a'}(B)} a_t^{*a'}; \quad a_t^{*a'} = \frac{1}{g_{L-1}^{*a'}(B)} e_t^{*a'};$$

$$\hat{y}_t^{L-MSSA-SARIMA-SSA}(1) = \hat{y}_t^{L-MSSA}(1) + \hat{a}_t^{*a'}(1); \tag{44}$$

– в мультипликативной форме:

$$y_t^{L-MSSA-SARIMA-SSA} = \frac{\theta_{q^{*a''}}^{*a''}(B)}{S_{L-1}(B)\Phi_{p^{*a''}}^{*a''}(B)} \times \frac{1}{g_{L-1}^{*a''}(B)} e_t^{*a''}. \tag{45}$$

Дадим обозначение этой модели L -MSSA-SARIMA-SSA.

Выражения, использующие для формирования передаточной функции модели (27) формулы K -продолжения примут вид:

– в аддитивной форме:

$$\varepsilon_t = y_t - y_t^{K-MSSA}, \quad t = \overline{1, n};$$

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

$$a_t^{*...} = \frac{1}{g_{L-1}^{*a...}(B)} e_t^{*...};$$

$$\hat{y}_t^{K-MSSA-SARIMA-SSA}(1) = \hat{y}_t^{K-MSSA}(1) + \hat{\varepsilon}_t(1); \quad (46)$$

– в мультипликативной форме:

$$y_t^{K-MSSA-SARIMAX-SSA} = \frac{\sum_{i=1}^N \Psi_{K-1}^{(i)}(B)}{\Psi_{K-1}^y(B)} x_{t-b_i}^{(i)} + \frac{\theta_{q^{*IV}}^{*IV}(B)}{\Phi_{p^{*IV}}^{*IV}(B)} \times \frac{1}{g_{L-1}^{*aIV}(B)} e_t^{*IV}. \quad (47)$$

Обозначим эту модель K -MSSA – SARIMAX – SSA.

4. 4. 1. Метод на основе совместного использования многомерного метода «Гусеница»-SSA и моделей SARIMAX

Ниже приведен алгоритм метода на основе совместного использования многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX, алгоритм идентификации которых приведен в пункте 4. 1. 5.

1. Если строим модель (39), (42), (44) или (46), то переходим к этапу 3.

Применяя многомерный метод «Гусеница»-SSA к исходным ВР, определяются:

– полиномы от оператора задержки $\omega_{c_i}^{(i)}(B)$, $\delta_{r_i}^{(i)}(B)$, $i = \overline{1, N}$, если строим модель (1) или (27);

– вектор S_L и полином $S_{L-1}(B)$, если строим модель (40) или (45);

– вектор Ψ и полиномы $\Psi_{K-1}^y(B)$ и $\Psi_{K-1}^{x(i)}(B)$, $i = \overline{1, N}$, если строим модель (43) или (47).

2. Определяется ВР \tilde{y}_t следующим образом:

$$- \tilde{y}_t = \sum_{i=1}^N \frac{\omega_{c_i}^{(i)}(B)}{\delta_{r_i}^{(i)}(B)} x_{t-b_i}^{(i)} = \sum_{i=1}^N (\delta_1^{(i)} y_{t-1} + \delta_2^{(i)} y_{t-2} + \dots + \delta_{r_i}^{(i)} y_{t-r_i}) + \sum_{i=1}^N (\omega_0^{(i)} x_{t-b_i}^{(i)} - \omega_1^{(i)} x_{t-b_i-1}^{(i)} - \dots - \omega_{c_i}^{(i)} x_{t-b_i-c_i}^{(i)}),$$

$$t \geq \sum_{i=1}^N (b_i + c_i) + 1, \text{ если строим модель (1) или (27);}$$

$$- \tilde{y}_t^{S_L} = [S_L]_{L-1} y_{t-1} + [S_L]_{L-2} y_{t-2} + \dots + [S_L]_1 y_{t-(L-1)},$$

$$t = \overline{L, n}, \text{ если строим модель (40) или (45);}$$

$$\tilde{y}_t^\Psi = \frac{\sum_{i=1}^N \Psi_{K-1}^{x(i)}(B)}{\Psi_{K-1}^y(B)} x_{t-b_i}^{(i)} = (\Psi_{K-1} y_{t-1} + \Psi_{K-2} y_{t-2} + \dots + \Psi_1 y_{t-(K-1)}) + \sum_{i=1}^N (\Psi_{(i+1)(K-1)} x_{t-b_i}^{(i)} - \Psi_{(i+1)(K-1)-1} x_{t-b_i-1}^{(i)} - \dots - \Psi_{(i+1)(K-1)-(K-2)} x_{t-b_i-(K-2)}^{(i)}),$$

$$t = \overline{K, n}, \text{ если строим модель (43) или (47).}$$

3. Вычисляется остаточный ВР ε_t следующим образом:

$$- \varepsilon_t^{L-MSSA} = y_t - \tilde{y}_t^{L-MSSA}, \quad t = \overline{1, n}, \text{ если строим модель (39) или (44);}$$

$$- \varepsilon_t^{K-MSSA} = y_t - \tilde{y}_t^{K-MSSA}, \quad t = \overline{1, n}, \text{ если строим модель (42) или (46);}$$

$$- \varepsilon_t^{S_L} = y_t - \tilde{y}_t^{S_L}, \quad t = \overline{1, n-L+1}, \text{ если строим модель (40) или (45);}$$

$$- \varepsilon_t^\Psi = y_t - \tilde{y}_t^\Psi, \quad t = \overline{1, n-K+1}, \text{ если строим модель (43) или (47).}$$

4. Идентифицируются модели SARIMA для остаточного ВР

$$- \varepsilon_t = \frac{\theta_q^*(B)}{\Phi_p^+(B)} a_t, \text{ если используется модель (1) или (27);}$$

$$- \varepsilon_t^{L-MSSA} = \frac{\theta_{q^{*...}}^{*...}(B)}{\Phi_{p^{*...}}^{*...}(B)} a_t^{*...}, \text{ если строим модель (39);}$$

$$- \varepsilon_t^{K-MSSA} = \frac{\theta_{q^{*...}}^{*...}(B)}{\Phi_{p^{*...}}^{*...}(B)} a_t^{*...}, \text{ если строим модель (42);}$$

$$- \varepsilon_t^{S_L} = \frac{\theta_{q^{*...}}^{*...}(B)}{\Phi_{p^{*...}}^{*...}(B)} a_t^{*...}, \text{ если строим модель (40);}$$

$$- \varepsilon_t^\Psi = \frac{\theta_{q^{*...}}^{*...}(B)}{\Phi_{p^{*...}}^{*...}(B)} a_t^{*...}, \text{ если строим модель (43);}$$

$$- \varepsilon_t^{L-MSSA} = \frac{\theta_{q^{*...}}^{*...}(B)}{\Phi_{p^{*...}}^{*...}(B)} a_t^{*...}.$$

5. Идентифицируем ЛРФ

$$- a_t = \frac{1}{g_{L-1}^{*a}(B)} e_t, \text{ если строим модель (27) или}$$

$$a_t = \frac{1}{g_{L_{*seq}}^{*a}(B)} a_t^{seq}, \text{ если строим модель (33);}$$

$$- a_t' = \frac{1}{g_{L-1}^{*a}(B)} e_t', \text{ если строим модель (44) или}$$

$$a_t' = \frac{1}{g_{L_{*seq}}^{*a}(B)} a_t^{seq'}, \text{ если для остаточных ошибок } a_t'$$

используем последовательный метод «Гусеница»-SSA;

$$- a_t'' = \frac{1}{g_{L-1}^{*a} (B)} e_t'' , \text{ если строим модель (45) или}$$

$$a_t'' = \frac{1}{g_{L-1}^{*seq} (B)} a_t^{seq''} , \text{ если для остаточных ошибок } a_t''$$

используем последовательный метод «Гусеница»-SSA;

$$- a_t''' = \frac{1}{g_{L-1}^{*a} (B)} e_t''' , \text{ если строим модель (46) или}$$

$$a_t''' = \frac{1}{g_{L-1}^{*seq} (B)} a_t^{seqIV} , \text{ если для остаточных ошибок}$$

a_t''' используем последовательный метод «Гусеница»-SSA;

$$- a_t^{IV} = \frac{1}{g_{L-1}^{*aIV} (B)} e_t^{IV} , \text{ если строим модель (47) или}$$

$$a_t^{IV} = \frac{1}{g_{L-1}^{*seqV} (B)} a_t^{seqV} , \text{ если для остаточных ошибок } a_t^{IV}$$

используем последовательный метод «Гусеница»-SSA.

6. Вычисляются прогнозы с заданным упреждением l :

$$- \hat{y}_t^{L-MSSA}(l), \hat{\epsilon}_t^{L-MSSA}(l), l = \overline{1, L}, \text{ если применяется модель (39);}$$

$$- \hat{y}_t^{K-MSSA}(l), \hat{\epsilon}_t^{K-MSSA}(l), l = \overline{1, L}, \text{ если применяется модель (42);}$$

- переходим к этапу 6, если используем модели (40) или (43).

7. Синтезируется общий прогноз как сумма частных прогнозов:

$$- \hat{y}_t^{L-MSSA-SARIMA}(l) = \hat{y}_t^{L-MSSA}(l) + \hat{\epsilon}_t^{L-MSSA}(l), \text{ если используется модель (39);}$$

$$- \hat{y}_t^{K-MSSA-SARIMA}(l) = \hat{y}_t^{K-MSSA}(l) + \hat{\epsilon}_t^{K-MSSA}(l), \text{ если используется модель (42).}$$

Или вычисляется прогноз по модели (40) или (43), предварительно перейдя от операторного к разностному уравнению.

4. 5. Гибридные математические модели и методы декомпозиционного подхода к прогнозированию временных рядов с учётом внешних факторов на основе многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX

По аналогии с алгоритмом декомпозиционного подхода к прогнозированию, приведенного в [4] в работе предложен алгоритм декомпозиционного подхода к прогнозированию ВР на основе совместного использования многомерного метода «Гусеница»-SSA и моделей SARIMAX. Повысить эффективность прогнозирования таких декомпозиционных моделей в большинстве случаев можно, используя вместо моделей SARIMAX искусственные нейронные сети (ИНС) для

прогнозирования каждой из компонент разложения. Схема декомпозиционного подхода к прогнозированию нестационарных ВР с учётом внешних факторов при этом примет следующий вид (рис. 3).

Как показано в [4] в зависимости от способа группировки матриц $X_{I_j}, j = \overline{1, g}$ можно рассмотреть следующие варианты гибридных моделей декомпозиционного подхода к прогнозированию ВР на основе многомерного метода «Гусеница»-SSA и моделей SARIMAX.

1. На этапе 3 «группировка» многомерного варианта метода «Гусеница»-SSA возьмём параметр $g=L$, т. е. всё множество индексов $\{1, 2, \dots, L\}$ разделим на L подмножеств по одному элементу в каждом следующим образом $I_j = \{j\}, j = \overline{1, L}$. Тогда на этапе 4 «диагональное усреднение» многомерного варианта метода «Гусеница»-SSA исходные прогнозируемый и экзогенные ВР будут разложены на следующие суммы:

$$y_t = \sum_{i=1}^L \tilde{y}_t^{(i)}, x_t^{(j)} = \sum_{i=1}^L \tilde{x}_t^{(ij)}, j = \overline{1, N}. \tag{48}$$

Далее декомпозиционный подход к прогнозированию нестационарных ВР на основе совместного использования многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX предполагает идентифицировать и строить модели SARIMAX для каждого ВР разложения $\tilde{y}_t^{(i)}, \tilde{x}_t^{(ij)}, i = \overline{1, L}, j = \overline{1, N}$.

Для удобства сделаем следующие обозначения:

$$w_t^{(i)} = \tilde{y}_t^{(i)}, i = \overline{1, L}, w_t^{(L+(j-1)L+i)} = \tilde{x}_t^{(ij)}, i = \overline{1, L}, j = \overline{1, N}.$$

Таким образом, гибридная модель декомпозиционного подхода к прогнозированию ВР на основе многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX можно записать в следующем виде:

$$w_t^{(j)} = \sum_{i \in M, i \neq j} \frac{\alpha_{w^{(j)}}^{w^{(i)}}(B)}{\delta_{w^{(j)}}^{w^{(i)}}(B)} w_{t-b_{w^{(j)}}}^{(i)} + \frac{\theta_{w^{(j)}}^*(B)}{\Phi_{p_{w^{(j)}}}^{w^{(j)}}(B)} a_t^{w^{(j)}},$$

$$M = \{1, 2, \dots, (N+1)L\};$$

$$\hat{y}_t(l) = \sum_{i=1}^{(N+1)L} \hat{w}_t^{(i)}(l). \tag{49}$$

2. Теперь рассмотрим данную модель в контексте концепции стандартизованного моделирования нестационарных ВР, приведенного в [4, 26]. На этапе 3 «группировка» многомерного метода «Гусеница»-SSA всё множество индексов $\{1, 2, \dots, L\}$ делится на g непересекающихся подмножеств I_1, I_2, \dots, I_g . Тогда на этапе диагонального усреднения метода «Гусеница»-SSA исходный ВР будет разложен на следующую сумму:

$$y_t = \sum_{i=1}^g \tilde{y}_t^{(i)}, x_t^{(j)} = \sum_{i=1}^g \tilde{x}_t^{(ij)}, j = \overline{1, N}. \tag{50}$$

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

$$y_t \approx \sum_{i=1}^g \tilde{y}_t^{(i)}, x_t^{(j)} \approx \sum_{i=1}^g \tilde{x}_t^{(ij)}, j = \overline{1, N}. \tag{51}$$

в противном случае. Причём ВР $\tilde{y}_t^{(i)}, \tilde{x}_t^{(j,i)}, i = \overline{1, \Gamma}, j = \overline{1, N}$ могут интерпретироваться как трендовые, квазипериодические или шумовые составляющие исходных ВР y_t и $x_t^{(j)}, j = \overline{1, N}$ соответственно.

Также вводим обозначения $w_t^{(i)} = \tilde{y}_t^{(i)}, i = \overline{1, L}, w_t^{(r+(j-1)r+i)} = \tilde{x}_t^{(j,i)}, i = \overline{1, \Gamma}, j = \overline{1, N}$.

Тогда гибридная модель прогнозирования нестационарных ВР, основанная на концепции стандартизованного моделирования и на многомерном методе «Гусеница»-SSA и моделях SARIMAX для случая (50) примет вид:

$$w_t^{(j)} = \sum_{i \in M, i \neq j} \frac{\omega_{w_t^{(j)}}^{w_t^{(i)}}(B)}{\delta_{w_t^{(j)}}^{w_t^{(i)}}(B)} w_t^{(i)} + \frac{\theta_{w_t^{(j)}}^{*}(B)}{\Phi_{p^{w_t^{(j)}}}(B)} a_t^{w_t^{(j)}},$$

$$M = \{1, 2, \dots, (N+1)r\};$$

$$\hat{y}_t(l) = \sum_{i=1}^{(N+1)r} \hat{w}_t^{(i)}(l). \quad (52)$$

А для случая (51) –

$$w_t^{(j)''} = \sum_{i \in M, i \neq j} \frac{\omega_{w_t^{(j)''}}^{w_t^{(i)''}}(B)}{\delta_{w_t^{(j)''}}^{w_t^{(i)''}}(B)} w_t^{(i)''} + \frac{\theta_{w_t^{(j)''}}^{*''}(B)}{\Phi_{p^{w_t^{(j)''}}}(B)} a_t^{w_t^{(j)''}},$$

$$M = \{1, 2, \dots, (N+1)r\};$$

$$y_t = \sum_{i=1}^{(N+1)r} \frac{\omega_{y_t}^{w_t^{(i)'}}(B)}{\delta_{y_t}^{w_t^{(i)'}}(B)} w_t^{(i)'} + \frac{\theta_{y_t}^{*'}(B)}{\Phi_{p^{y_t}}(B)} a_t^{y_t}. \quad (53)$$

Дадим модели (49) обозначение MSSA(L) – SARIMAX, модели (52) – MSSA – SARIMAX, а модели (53) – MSSA – SARIMAX+SARIMA.

Прогнозные значения компонент $\hat{w}_t^{(j)}$ в данной работе предлагается получать при помощи моделей (1). Однако, для тех же целей можно также использовать одномерный или многомерный вариант метода «Гусеница»-SSA [2, 25], модели SARIMA [1] или SARIMAX, комбинированные модели ЛРФ-SSA-SARIMA, Тренд-SSA-SARIMA [4] или декомпозиционные модели SSA(L)-SARIMA, SSA-SARIMA, SSA-SARIMA+SARIMA [4], а также, синтезированные в данной работе модели L-MSSA – SARIMA, K-MSSA – SARIMAX, L-MSSA – SARIMA – SSA, K-MSSA – SARIMAX – SSA или искусственные нейронные сети.

4. 5. 1. Метод декомпозиционного подхода к прогнозированию нестационарных временных рядов с учётом внешних факторов на основе совместного использования многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX

Приведём алгоритм и схему (рис. 3) метода на основе совместного использования многомерного метода «Гусеница»-SSA и моделей SARIMAX декомпозиционного подхода к прогнозированию нестационарных ВР с учётом экзогенных факторов.

1. Применяя многомерный вариант метода «Гусеница»-SSA к исходным прогнозируемому и экзогенным ВР определяются ВР

– $\tilde{y}_t^{(i)}, \tilde{x}_t^{(j,i)}, i = \overline{1, L}, j = \overline{1, N}$, если строится модель (49);

– $\tilde{y}_t^{(i)}, \tilde{x}_t^{(j,i)}, i = \overline{1, \Gamma}, j = \overline{1, N}$, если строится модель (52) или (53)

и вводятся обозначения

– $w_t^{(i)} = \tilde{y}_t^{(i)}, i = \overline{1, L}, w_t^{(L+(j-1)L+i)} = \tilde{x}_t^{(j,i)}, i = \overline{1, L}, j = \overline{1, N}$;

– $w_t^{(i)} = \tilde{y}_t^{(i)}, i = \overline{1, \Gamma}, w_t^{(r+(j-1)r+i)} = \tilde{x}_t^{(j,i)}, i = \overline{1, \Gamma}, j = \overline{1, N}$.

2. Для каждого ВР

– $w_t^{(i)}, i = \overline{1, (N+1)L}$, если используется модель (49);

– $w_t^{(i)}, i = \overline{1, (N+1)L}$, если используется модель (52) или (53)

производится структурная и параметрическая идентификация моделей SARIMAX (используя алгоритм пункта 4. 1. 5)

$$w_t^{(j)} = \sum_{i \in M, i \neq j} \frac{\omega_{w_t^{(j)}}^{w_t^{(i)}}(B)}{\delta_{w_t^{(j)}}^{w_t^{(i)}}(B)} w_t^{(i)} + \frac{\theta_{w_t^{(j)}}^{*}(B)}{\Phi_{p^{w_t^{(j)}}}(B)} a_t^{w_t^{(j)}},$$

$M = \{1, 2, \dots, (N+1)L\}$ для модели (49);

$$w_t^{(j)'} = \sum_{i \in M, i \neq j} \frac{\omega_{w_t^{(j)'}}^{w_t^{(i)'}}(B)}{\delta_{w_t^{(j)'}}^{w_t^{(i)'}}(B)} w_t^{(i)'} + \frac{\theta_{w_t^{(j)'}}^{*'}(B)}{\Phi_{p^{w_t^{(j)'}}}(B)} a_t^{w_t^{(j)'}},$$

$M = \{1, 2, \dots, (N+1)\Gamma\}$ для модели (52);

$$w_t^{(j)''} = \sum_{i \in M, i \neq j} \frac{\omega_{w_t^{(j)''}}^{w_t^{(i)''}}(B)}{\delta_{w_t^{(j)''}}^{w_t^{(i)''}}(B)} w_t^{(i)''} + \frac{\theta_{w_t^{(j)''}}^{*''}(B)}{\Phi_{p^{w_t^{(j)''}}}(B)} a_t^{w_t^{(j)''}},$$

$M = \{1, 2, \dots, (N+1)\Gamma\}$ для модели (53).

3. Вычисляются прогнозы с заданным упреждением l

– $\hat{w}_t^{(i)}(l), i = \overline{1, (N+1)L}$, для модели (49);

– $\hat{w}_t^{(i)}(l), i = \overline{1, (N+1)\Gamma}$, для модели (52) или (53).

4. Синтезируется общий прогноз как сумма частных прогнозов

– $\hat{y}_t(l) = \sum_{i=1}^{(N+1)L} \hat{w}_t^{(i)}(l)$ для модели (49);

– $\hat{y}_t(l) = \sum_{i=1}^{(N+1)\Gamma} \hat{w}_t^{(i)}(l)$ для модели (52)

или

– вычисляется прогноз по модели (53), предварительно приведя её к разностному выражению.

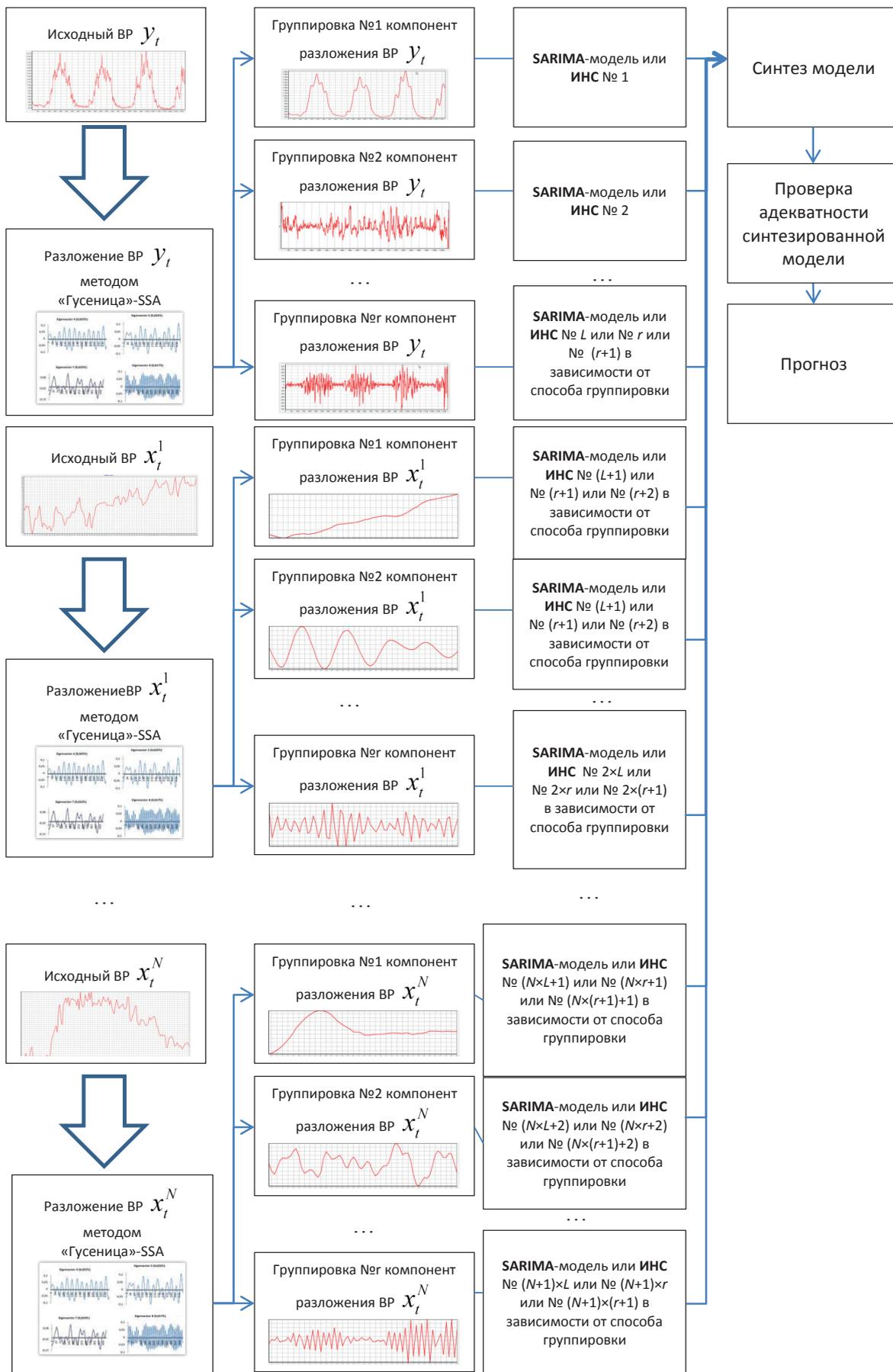


Рис. 3. Схема декомпозиционного подхода к прогнозированию нестационарных ВР с учётом внешних факторов

5. Результаты исследований эффективности предложенных моделей прогнозирования временных рядов с учётом внешних факторов

Исследование предлагаемых гибридных моделей на основе многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX осуществим сопоставляя их результаты прогнозов с результатами прогнозов, получаемых классическими вероятностными моделями SARIMAX, обобщёнными на случай прогнозирования ВР с несколькими сезонными составляющими и с учётом нескольких внешних факторов. Реализация рассмотренных моделей производилась в математическом пакете MATLAB R2014a.

Тестирование будем проводить на ВР часовых значений потребления электроэнергии объёма 1008 значений, что соответствует 6 неделям (рис. 4), с учётом изменения температуры воздуха (рис. 5). Обучения моделей будет производиться на выборках данных за 5 недель (840 значений), а тестирование – на данных последней недели. После проверки адекватности модели, прогноз будет выполняться одношаговый и производиться скольжение окна до последнего значения ВР.

Как отмечается в [4], структурная идентификация гибридных математических моделей на основе методов «Гусеница»-SSA и Бокса-Дженкинса является достаточно сложной задачей. Анализ собственных чисел траекторной матрицы ВР, автокорреляционной и частной автокорреляционной функций ВР не даёт точных значений структурных параметров наиболее эффективной модели. Поэтому под структурной идентификацией приведенных гибридных моделей будем подразумевать перебор различных вариантов разложения исходных ВР и группировки ВР разложения методом MSSA, идентификацию компонент разложения или остаточной компоненты моделями

SARIMAX, по приведенному в пункте 4.5.1 алгоритму и выбор наилучшей по точности прогнозирования синтезированной модели из множества альтернатив. Разложение исходного ВР получается тем более детальным, чем больше длина окна. Наиболее детальное разложение достигается при выборе длины окна, приблизительно равной половине длины ряда и кратной периоду периодической компоненты, за исключением рядов конечного ранга [2]. При одновременном выделении разных компонент ряда (трендовой, периодических, шумовых компонент) все аспекты выбора длины окна должны приниматься во внимание. Таким образом, имеет смысл начинать всегда с выбора длины окна близкой к половине ряда и кратной периоду периодической компоненты, если такая присутствует в ряде и её период известен.

Сравнительный анализ эффективности прогнозирования рассмотренными моделями будем осуществлять при помощи 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*.

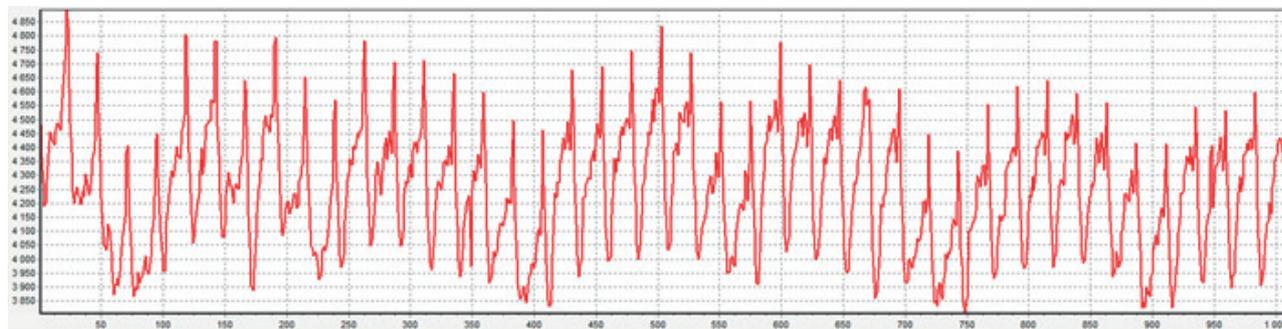


Рис. 4. График ВР часовых значений потребления электроэнергии за 6 недель (в МВт)

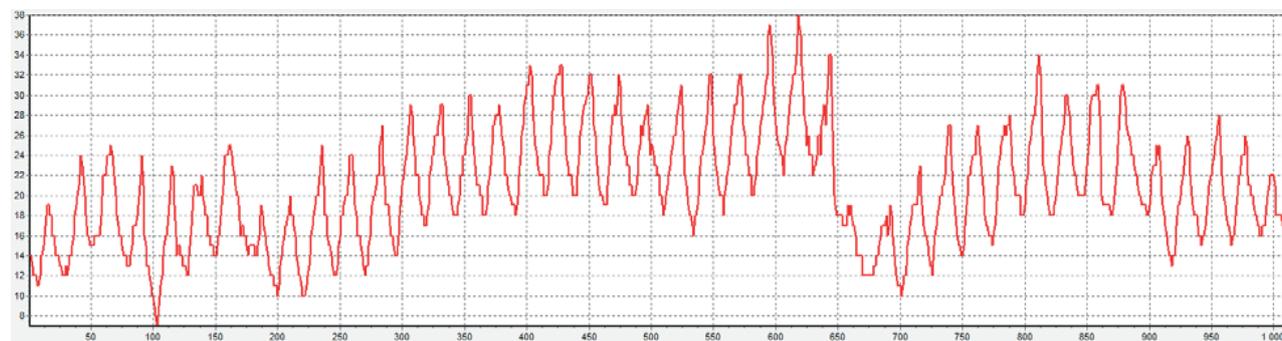


Рис. 5. График ВР часовых значений изменения температуры воздуха за 6 недель (в С°)

Выберем длину окна $L=396$. В модели будем отбирать количества первых компонент разложения кратные периоду 24. Приведенная ниже столбчатая диаграмма (рис. 6) говорит о снижении ошибки прогнозирования при добавлении к прогнозам, полученным при помощи L -продолжения многомерного варианта метода «Гусеница»-SSA, прогнозов, полученных моделями SARIMAX.

Анализ диаграммы (рис. 7) говорит о снижении ошибки прогнозирования при использовании модели $MSSA - SARIMAX + SARIMA$ в сравнении с прогнозами, полученными моделями $MSSA - SARIMAX$.

Наименьшую ошибку прогнозирования показала модель $K-MSSA - SARIMAX$ при использовании в модели 312 первых собственных троек SVD-разложения траекторной матрицы исходных ВР. Изменяя количество отбираемых в модель первых собственных троек вблизи значения $r=312$, удаётся найти более точную модель с $r=298$ ($RMSE=8.712$). На следующей диаграмме (рис. 9) изображены ошибки прогнозирования $RMSE$ для наилучших моделей из рассмотренных классов моделей. Наименьшую ошибку прогнозирования показала модель $K-MSSA - SARIMAX$.

В любом случае, всегда стоит попробовать несколько вариантов выбора длины окна. Для рядов со слишком сложной структурой слишком большая длина окна L может привести к нежелательному разложению компонент ряда, в частности, к смешиванию их с другими компонентами ряда. Даже маленькое изменение значения L может уменьшить смешивание компонент разложения и привести к лучшему разделению компонент ряда, т. е. обеспечить переход от слабой к сильной разделимости [2].

На рис. 10, 11 показано, что более точную модель $K-MSSA - SARIMAX$ можно получить, используя длину окна $L=72$ и количество первых собственных троек SVD-разложения траекторной матрицы исходных ВР метода $MSSA$ равное 46 ($RMSE=7.1304$). Незначительно по ошибке прогнозирования уступает более простая модель, использующая длину окна $L=48$ и количество первых собственных троек равное 34. Ошибка прогнозирования такой модели составила $RMSE=7.32$. Приведём прогнозы наилучшей по эффективности прогнозирования из класса $K-MSSA - SARIMAX$ модели (рис. 12).

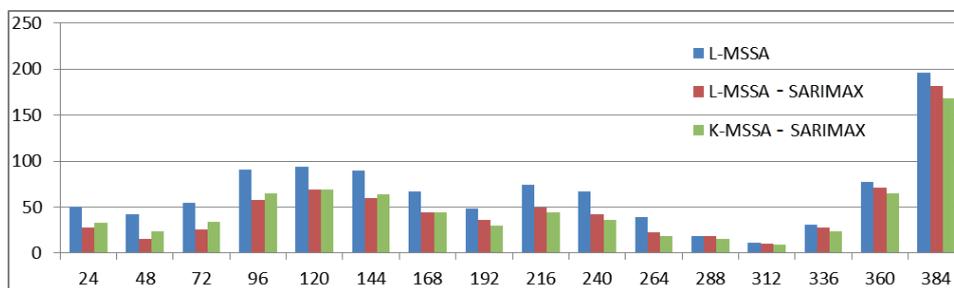


Рис. 6. Диаграммы зависимости ошибки прогнозирования $RMSE$ моделей $L-MSSA$, $L-MSSA - SARIMAX$ и $K-MSSA - SARIMAX$ от количества используемых в них первых компонент разложения метода $MSSA$

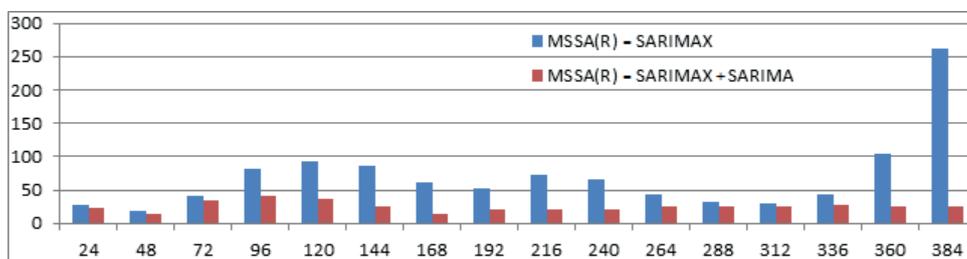


Рис. 7. Диаграммы зависимости ошибки прогнозирования $RMSE$ моделей $MSSA - SARIMAX$ и $MSSA - SARIMAX + SARIMA$ от количества используемых в них первых компонент разложения метода $MSSA$

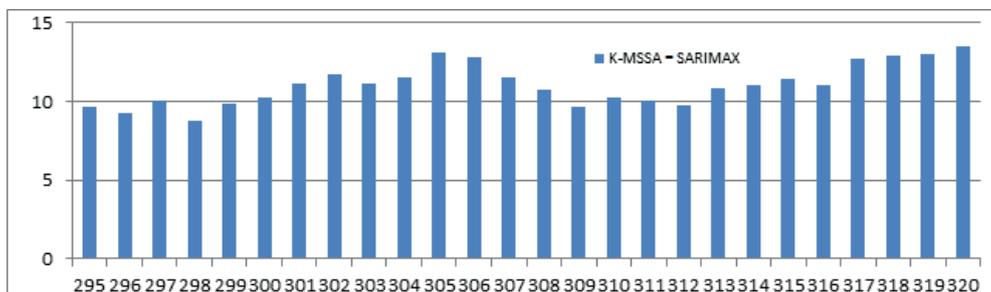


Рис. 8. Диаграммы зависимости ошибки прогнозирования $RMSE$ моделей $K-MSSA - SARIMAX$ от количества используемых в них первых компонент разложения метода $MSSA$

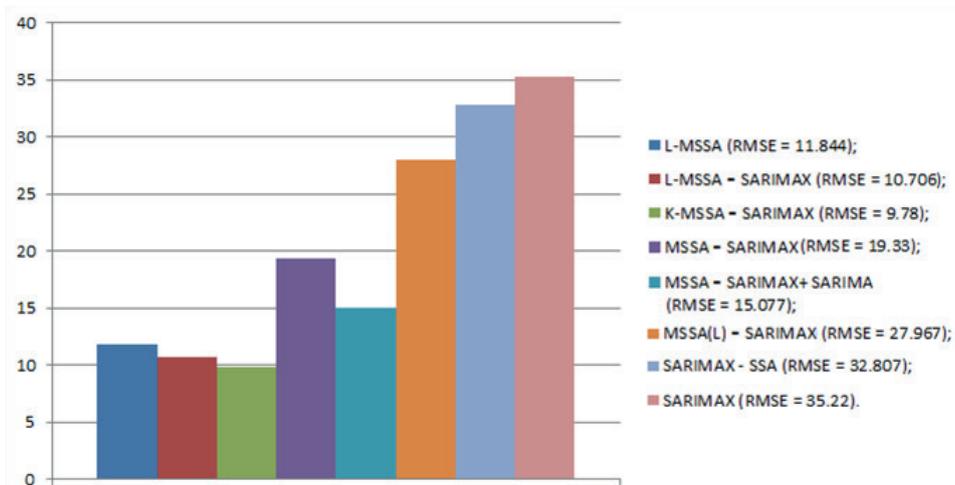


Рис. 9. Диаграмма минимальных ошибок прогнозирования RMSE для рассмотренных классов моделей (L=396)

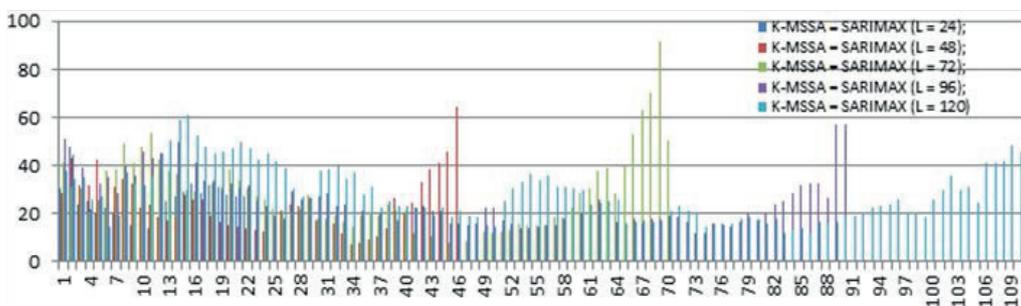


Рис. 10. Диаграммы зависимости ошибки прогнозирования RMSE моделей K-MSSA – SARIMAX от количества используемых в них первых компонент разложения методом MSSA для длин окна 24, 48, 72, 96 и 120 соответственно

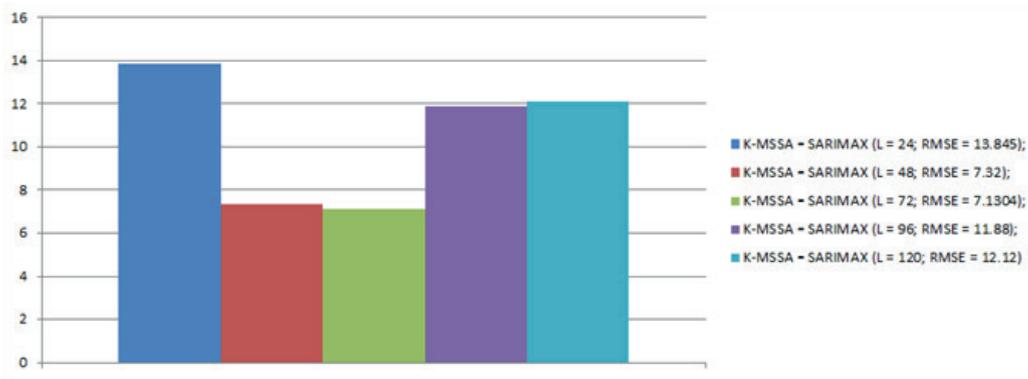


Рис. 11. Диаграмма минимальных ошибок прогнозирования RMSE для моделей K-MSSA – SARIMAX, использующих различные длины окон



Рис. 12. График временного ряда потребления электроэнергии и прогноза, полученные лучшей по эффективности прогнозирования из рассмотренных моделей

Рассмотрим пример прогнозирования ВР потребления природного газа. График данного ВР среднесуточного объёма газопотребления представлен на рис. 14. График ВР изменения температуры воздуха представлен на рис. 13. Обучение моделей будем проводить на данных за два года (объём 730 значений). После проверки адекватности модели, прогноз будет производиться одношаговый, также как и для предыдущего примера.

Для ВР потребления природного газа, который имеет полиномиальный, полигармонический и стоха-

стический тренд со сложной корреляционной структурой наилучший результат выдала модель из класса MSSA – SARIMAX+SARIMA (рис. 15). Ниже приведём прогнозы наилучшей по эффективности прогнозирования модели из класса MSSA – SARIMAX+SARIMA (рис. 16).

Таким образом, для ВР различной структуры, наиболее эффективный прогноз дают модели из различных классов.

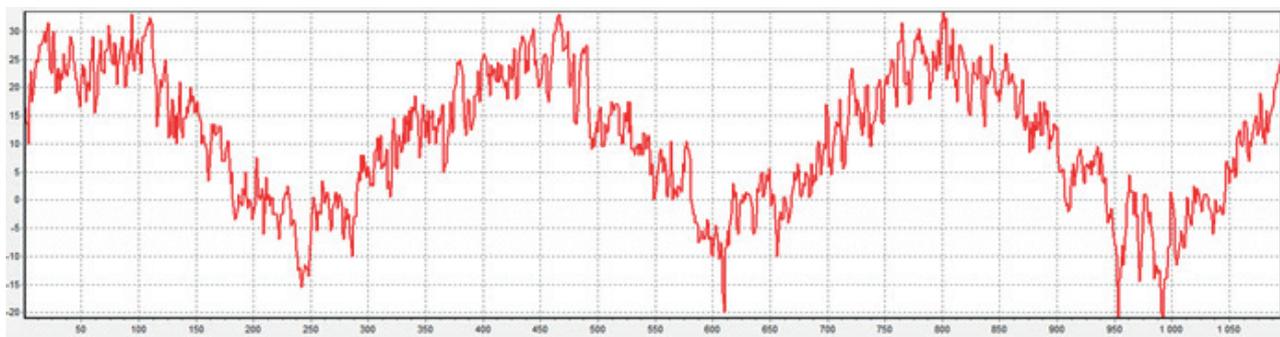


Рис. 13. График ВР среднесуточных значений изменения температуры воздуха (в С°)

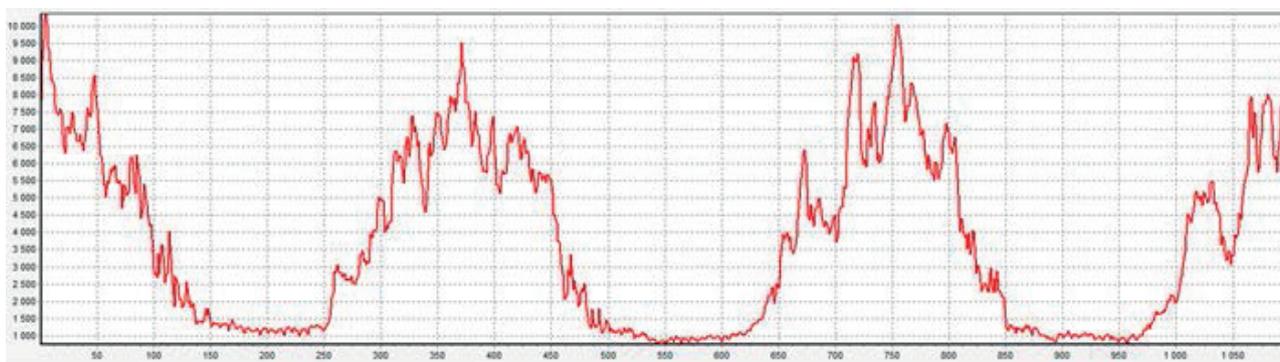


Рис. 14. График ВР среднесуточных данных потребления природного газа (в Мм³/сут)

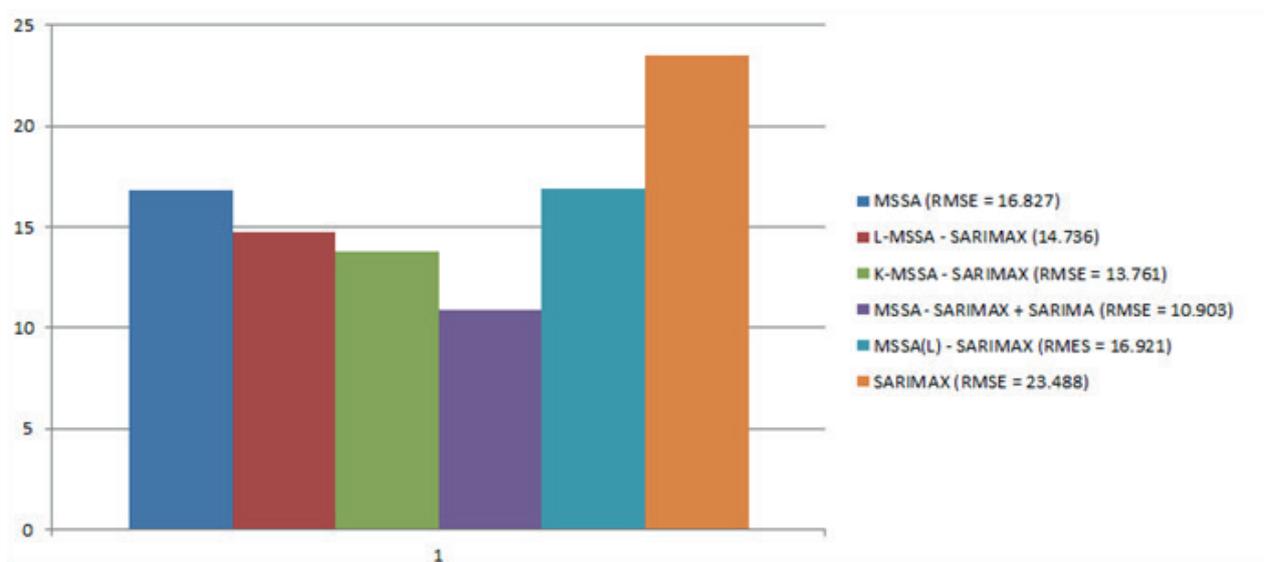


Рис. 15. Диаграмма минимальных ошибок прогнозирования RMSE для рассмотренных классов моделей (L=358)



Рис. 16. График потребления природного газа и прогнозы, полученные наилучшей из рассмотренных моделей

6. Выводы

В работе предложены модели различной сложности для прогнозирования нестационарных ВР с учётом экзогенных факторов, а также методы идентификации этих моделей на основе совместного использования многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX, обобщённых для случая прогнозирования ВР с несколькими сезонными компонентами и с учётом нескольких экзогенных переменных. Один класс гибридных математических методов, основанных на многомерном варианте метода «Гусеница»-SSA и моделях SARIMAX, заключается в структурной идентификации передаточной функции модели Бокса-Дженкинса, обобщённой на случай прогнозирования с учётом нескольких экзогенных факторов, используя формулы L- или K-продолжения многомерного варианта метода «Гусеница»-SSA. Декомпозиционный подход к прогнозированию ВР на основе многомерного варианта метода «Гусеница»-SSA и моделей SARIMAX заключается в разложении методом MSSA исходных прогнозируемого и экзогенных ВР на множество ВР с более простой структурой; прогнозировании данных компонент разложения моделями SARIMAX и вычислении общего прогноза, объединяя прогнозы построенных упрощённых моделей. Выбор

различных параметров многомерного варианта метода «Гусеница»-SSA приводит к созданию, в сочетании с моделями SARIMAX, различных вариантов методов прогнозирования и к синтезу на их основе гибридных математических моделей с различными структурами.

Предложенные модели были протестированы на ВР потребления электроэнергии и природного газа, а их результаты прогнозирования были сравнены с результатами, полученными вероятностными моделями SARIMAX. Экспериментальные результаты показывают высокую эффективность предложенных моделей прогнозирования при выборе подходящих структурных параметров.

Таким образом, в статье разработана методология моделирования и прогнозирования нестационарных ВР, использование которой позволило:

1. Учесть влияние на рассматриваемые ВР произвольного количества основных групп факторов (например, метеорологических, хронологических и организационных).
2. Получить экономичные по структуре и количеству оцениваемых параметров модели, описывающие широкий класс нестационарных ВР с детерминированными (полигармоническими и полиномиальными) или стохастическими трендами и локальной нерегулярностью.
3. Разработать эффективные методы и алгоритмы структурной идентификации моделей, проверки степени их адекватности реальным процессам, вычисления и коррекции прогнозов.

В дальнейшем целесообразно обобщит гибридные математические модели и методы, основанные на многомерном варианте метода «Гусеница»-SSA и моделях VARMAX для прогнозирования взаимосвязанных нестационарных ВР.

Литература

1. Евдокимов, А. Г. Оперативное управление потокораспределением в инженерных сетях [Текст] / А. Г. Евдокимов, А. Д. Тевяшев. – Х.: Вища школа, 1980. – 144 с.
2. Голяндина, Н. Э. Метод «Гусеница»-SSA: анализ временных рядов [Текст]: учеб. пос. / Н. Э. Голяндина. – СПб.: С.-Петербургский государственный университет, 2004. – 74 с.
3. Варианты метода «Гусеница»-SSA для прогноза многомерных временных рядов [Текст] : труды IV междунар. конф. / Москва «Идентификация систем и задачи управления». – Москва, 2005. – С. 1831–1848.
4. Щелкалин, В. Н. Гибридные модели и методы прогнозирования временных рядов на основе методов «Гусеница»-SSA и Бокса-Дженкинса [Текст] / В. Н. Щелкалин // Восточно-Европейский журнал передовых технологий. – 2014. – Т. 5, № 4 (71). – С. 43–62. – Режим доступа: <http://journals.uran.ua/eejet/article/view/28172/25654> doi: 10.15587/1729-4061.2014.28172
5. Lawrance, A. J. Stochastic modelling of riverflow time series [Text] / A. J. Lawrance, N. T. Kottegoda // Journal of the Royal Statistical Society. Series A (General). – 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, 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 // Ecological Modelling. – 2005. – Vol. 183, Issue 1. – P. 67–75. doi: 10.1016/j.ecolmodel.2004.08.001
8. Fraedrich, K. Estimating the dimension of weather and climate attractor [Text] / K. Fraedrich // Journal of the Atmospheric Sciences. – 1986. – Vol. 43. – P. 419–432.
9. Vautard, R. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series [Text] / R. Vautard, M. Ghil // Physica D: Nonlinear Phenomena. – 1989. – Vol. 35, Issue 3. – P. 395–424. doi: 10.1016/0167-2789(89)90077-8

10. 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
11. Yiou, P. Spectral analysis of climate data [Text] / P. Yiou, E. Baert, M. F. Loutre // Surveys in Geophysics. – 1996. – Vol. 17, Issue 6. – P. 619–663. doi: 10.1007/bf01931784
12. 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. – P. 320.
13. 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. – Vol. 31. – P. 1172–1179.
14. Hassani, H. Forecasting European industrial production with singular spectrum analysis [Text] / H. Hassani, S. Heravi, A. Zhigljavsky // International Journal of Forecasting. – 2009. – Vol. 25, Issue 1. – P.103–118. doi: 10.1016/j.ijforecast.2008.09.007
15. 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.
16. 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 Power Tech. 2007. – P. 1090–1093. doi: 10.1109/pct.2007.4538467
17. 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 Resources Management. – 2011. – Vol. 25, Issue 11. – P. 2683–2703. doi: 10.1007/s11269-011-9833-y
18. Щелкалин, В. Н. Декомпозиционный подход прогнозирования временных рядов на основе метода «Гусеница»-SSA [Текст] : Материалы 14-й Межд. науч.-тех. конф. САИТ / В.Н. Щелкалин // УНК «ИПСА» НТУУ «КПИ». – К.: УНК «ИПСА» НТУУ «КПИ», 2012. – С. 260–261.
19. Щелкалин, В. Н. От идей методов «Гусеница»-SSA и Бокса-Дженкинса до декомпозиционного метода прогнозирования и декомпозиционной ИНС [Текст] / В. Н. Щелкалин // Восточно-Европейский журнал передовых технологий. – 2011. – Т. 4, № 4 (52). – С. 59–69. – Режим доступа: http://www.nbu.gov.ua/old_jrn/natural/Vejpt/2011_4_4/2011_4_4/59-69.pdf
20. de Souza, R. M. SSA, Hierarchical Clustering and ARIMA Models in Forecasting Reservoir Inflows [Text] / R. M. de Souza, M. Barros, M. Menezes, J. F. M. Pessanha, L. A. T. Junior, K. M. Cassiano, R. C. Souza // The 32nd Annual International Symposium on Forecasting ISF – 2012, 2012. – P. 11–25.
21. de Souza, R. M. Streamflow Forecasting Using Singular Spectrum Analysis, ARIMA Model And Artificial Neural Network [Text] / R. M. de Souza, M. Menezes, R. C. Souza, J. F. M. Pessanha, L. A. T. Junior, K. M. Cassiano // The 32nd Annual International Symposium on Forecasting ISF – 2012, Boston, 2012. – P. 33–47.
22. Cassiano, K. ARIMA Model, Neural Networks and SSA in the Short Term Electric Load Forecast [Text] / K. Cassiano, M. Menezes, L. A. Junior, J. Pessanha, Raf. Souza, R. Souza // Economic Forecasting – Past, Present and Future: International Symposium of Forecasting ISF – 2014, Rotterdam, 2014. – P. 1 – 23.
23. Бокс, Дж. Анализ временных рядов. Прогноз и управление. Вып. I. Пер. с англ. [Текст] / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 406 с.
24. Евдокимов, А. Г. Об одном алгоритме выявления скрытой периодичности [Текст] / А. Г. Евдокимов, А. Д. Тевяшев // Автоматизированные системы управления и приборы автоматики. – 1977. – Вып. 43. – С. 60–64.
25. Голяндина, Н. Э. Метод «Гусеница»-SSA: прогноз временных рядов [Текст]: учеб. пособие / Н. Э. Голяндина – СПб. : С.-Петербургский государственный университет, 2004. – 52 с.
26. Седов, А. В. Моделирование объектов с дискретно-распределёнными параметрами: декомпозиционный подход [Текст] / А. В. Седов. – Южный научный центр РАН. – М. : Наука, 2010. – 438 с.
27. Чумак, О. В. Энтропии и фракталы в анализе данных [Текст] / О. В. Чумак. – НИЦ «Регулярная и хаотическая динамика». – М. Ижевск: Институт компьютерных исследований, 2011. – 164 с.