

У даній статті представляється огляд сучасних математичних моделей і методів, що стрімко розвиваються останнім часом автором у різних галузях науки і техніки, заснованих на спільному використанні ідей методу «Гусениця»-SSA і методу Бокса-Дженкінса. Пропоновані автором моделі є пріоритетними на сьогоднішній день ймовірно-детермінованими нелінійними декомпозиційними моделями.

Ключові слова: метод «Гусениця»-SSA, метод Бокса-Дженкінса, модель АРІКС, декомпозиційний метод моделювання та прогнозування, декомпозиційна штучна нейронна мережа.

В данной статье представляется обзор стремительно развивающихся в последнее время автором в различных отраслях науки и техники современных математических моделей и методов, основанных на совместном использовании идей метода «Гусеница»-SSA и метода Бокса-Дженкинса. Предлагаемые автором модели являются приоритетными на сегодняшний день вероятностно-детерминированными нелинейными декомпозиционными моделями.

Ключевые слова: метод «Гусеница»-SSA, метод Бокса-Дженкинса, модель АРІКС, декомпозиционный метод моделирования и прогнозирования, декомпозиционная искусственная нейронная сеть.

This paper presents an overview of the rapidly developing in recent years by the author in various fields of science and technology of modern mathematical models and methods based on joint usage of ideas of the «Caterpillar»-SSA and Box-Jenkins methods. The proposed by author models are a priority at present probabilistic and deterministic nonlinear decomposition models.

Keywords: «Caterpillar»-SSA method, Box-Jenkins method, ARIMA model, decomposition method of modeling and prediction, decomposition artificial neural network.

ОТ ИДЕЙ МЕТОДОВ «ГУСЕНИЦА»-SSA И БОКСА-ДЖЕНКИНСА ДО ДЕКОМПОЗИЦИОННОГО МЕТОДА ПРОГНОЗИРОВАНИЯ И ДЕКОМПОЗИЦИОННОЙ ИНС

В. Н. Щелкалин

Инженер 1-й категории

Кафедра прикладной математики

Харьковский национальный университет

радиоэлектроники

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

Контактный тел.: +38 (057) 7-19-91-09,

+38098-388-16-17

E-mail: vitalii.shchelkalin@gmail.com

1. Введение

В статье представлены наработки от отдельного использования классических методов прогнозирования, таких как: метод Бокса-Дженкинса, метод «Гусеница»-SSA, метод группового учета аргументов (МГУА) и их моделей и пр., комбинированных методов и моделей: метод «Гусеница»-SSA — АРСС, «Гусеница»-SSA — АРПСС, вероятностно-детерминированная модель совместного использования идей методов Бокса-Дженкинса и «Гусеница»-SSA, модель авторегрессии — спектрально проинтегрированного скользящего среднего и ее векторное обобщение — модель VARSIMAX, метод авторегрессии — спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщенной авторегрессионной условной гетероскедастичностью (АРСПСС — СПОАРУГ), метод «Гусеница»-SSA — АРПСС — СПОАРУГ, до предложенных в данной статье декомпозиционных методов прогнозирования и декомпозиционной нейронной сети [3, 4, 11—14, 17, 18].

Развитие теории математического моделирования определяется степенью математического описания процессов и явлений, имеющих место в различных отраслях

науки и техники. Научно-технический прогресс постоянно развивает экономические, технические, социальные, медицинские и прочие системы, усложняя их структуру и увеличивая количество их сложных внутренних взаимосвязей и внешних факторов, от которых они зависят и большинство из которых учесть невозможно. Поэтому актуальной является разработка и использование универсальных математических моделей и методик позволяющих моделировать, прогнозировать и управлять широким классом процессов, т. к. это позволит повысить эффективность управления и планирования режимов работы сложных систем, сэкономить значительные средства и ресурсы не разработкой новых энергосберегающих ресурсов, а путем всяких математических ухищрений, а также существенно разгружает работу соответствующего персонала различных организаций и позволяет предвидеть аварийные ситуации и привносит еще много прочей пользы.

Главными требованиями к построению математических моделей в 70—90-е годы прошлого века, являлись экономность по количеству параметров, скорость определения модели и ее ресурсоемкость для использования на доступных тогда ЭВМ малой производительности. Однако, современная вычислительная техника и методы

математического моделирования предоставляют большие возможности для анализа, моделирования, прогнозирования и управления временными рядами различной природы. Поэтому в настоящее время указанные требования не являются определяющими и современные вычислительные средства и системы позволяют выносить на первый план требования точности моделирования, качества анализа и прогнозирования. Одними из широко применяемых моделей, соответствующих вышеназванным требованиям являются модели ARIMAX.

В 80-е же годы прошлого века Гренджер и Джойо [1] предложили новый класс ARFIMA моделей, удобно описывающего финансово-экономические временные ряды с эффектами длинной и короткой памяти.

2000-е годы характеризуются применением большого спектра моделей для анализа и прогнозирования временных рядов, а также ансамблей моделей с различной структурой. С появлением высокоскоростных ЭВМ происходил и переход от ансамблей прогнозирующих моделей к ихней комбинации. Отличие комбинированных моделей от их ансамблей заключается в одновременном подстраивании параметров моделей.

Наиболее важными характеристиками моделей при анализе и выборе наиболее подходящих математических моделей являются следующие характеристики:

- способ моделирования трендовой составляющей временного ряда;
- способ нелинейного моделирования временного ряда;
- способ моделирования внутренних взаимосвязей между конструктивными компонентами, определяемыми при декомпозиции временных рядов;
- способ моделирования случайной составляющей временного ряда;
- способ учета влияния внешних факторов на процесс.

Поэтому приоритетным видом моделей являются комбинированные вероятностно-детерминированные нелинейные декомпозиционные модели, т. к. при этом в модели одновременно учитываются и используются как статистические, так и детерминированные составляющие и сложные внутренние взаимосвязи процессов, что позволяет достичь наилучшего качества прогнозирования [2]. Среди детерминированных моделей приоритетным вариантом является детерминированная модель спектрального разложения, реализующая моделирование на основе разложения по детерминированному ортонормированному базису, отличному от гармонических функций и используемая в методе «Гусеница»-SSA. В то время как наиболее приоритетной среди вероятностных моделей является модель авторегрессии — проинтегрированного скользящего среднего с экзогенными переменными.

В зарубежной литературе давно известны комбинированные вероятностно-детерминированные модели, представленные в [2]. В данной же статье предлагается обзор очередных модификаций моделей ARIMA и GARCH, метода «Гусеница»-SSA — АРПСС — СПОАРУГ, комбинированная вероятностно-детерминированная модель авторегрессии — спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщенной авторегрессионной условной гетероскедастичностью (АРПСС — СПОАРУГ, на англ. яз. ARSIMA — SIGARCH, на укр. яз. АРСІКС — СІУАРУГ) и метод ее построения, которые позволяют более гибко анализировать, моделировать и прогнозировать времен-

ные ряды в сравнении с моделями ARIMA — GARCH. А также представлены декомпозиционная модель прогнозирования и декомпозиционная нейронная сеть.

2. Изложение основного материала исследования

Суть зарождения всего теоретического материала автора сначала заключалась в многомерной декомпозиции экзогенных временных рядов и прогнозируемого временного ряда, на базисные латентные компоненты, включая их степени и комбинации, получаемые методами главных (PCA), гладких (SCA) и независимых компонентов (ICA) в отборе из этих базисных компонентов конструктивных методом быстрого ортогонального поиска (FOS), одного из эффективных и экономных по временным затратам методов в зарубежной литературе, и отсецинии деструктивных, тем самым сформировав передаточную функцию математической модели процесса; в дальнейшей идентификации шумовой части математической модели процесса благодаря сезонным моделям авторегрессии — скользящего среднего, и одновременной параметрической идентификации полученной структуры модели методом Левенберга-Марквардта [3]. Как стало известно позже, предложенный метод несколько схож с декомпозиционным методом моделирования (ДММ) [2]. Метод «Гусеница»-SSA тоже использует декомпозицию временных рядов по сингулярным значениям (SVD). Известны публикации использования метода «Гусеница»-SSA в различных отраслях науки и техники как метода достаточно хорошо описывающего нестационарные временные ряды с линейными, параболическими или экспоненциальными трендами с не всегда устойчивой колебательной составляющей, однако метод для моделирования использует неоптимальный с точки зрения точности воспроизведения некоторых временных рядов ортогональный базис векторов траекторной матрицы. Поэтому было предложено совместно использовать рекуррентную модель прогнозирования метода «Гусеница»-SSA и модель авторегрессии — скользящего среднего, обученных на конкурентной основе с учетом обобщенного критерия точности и адекватности. Использование такой комбинации было продиктовано тем, что отдельно эти подходы имеют ряд недостатков, а совместное их использование приводит к синергии, повышая их эффективность, робастность и адекватность. Однако при выделении тренда методом «Гусеница»-SSA, как и любым другим методом, остаточная составляющая ряда в большинстве случаев остается нестационарной, поэтому далее метод «Гусеница»-SSA использовался уже в комбинации с моделью авторегрессии — проинтегрированного скользящего среднего (АРПСС). При этом совместное использование приведенных выше методов подразумевает, что параметрическая идентификация осуществляет вычисление оценок параметров модели АРПСС и нелинейного обобщения авторегрессии на главных компонентах, минимизируя сумму квадратов отклонений ошибок моделирования с учетом временного ряда, получаемого методом «Гусеница»-SSA, который в течение моделирования не имеет в явном виде модели и соответственно параметров, а метод Левенберга-Марквардта, вычисляя параметры остальных аддитивных частей модели помогает определить внешний вид временного ряда метода «Гусеница»-SSA, выступающего в роли помощника определения детерминированной (трендовой) составляющей процесса.

Предлагаемый подход как раз и является одним из вариантов приоритетных на сегодняшний день комбинированных вероятностно-детерминированных подходов. Здесь также реализуется так называемый трендовый подход, когда процесс моделируется как отклонение фактических значений от тренда (который представлен здесь временным рядом, получаемым методом «Гусеница»-SSA) и который обеспечивает устойчивость получаемой модели и достаточную точность моделирования, тогда как ранее вероятностной моделью АРПСС пытались описать весь процесс, а в моделях АРПСС временные ряды приводятся к стационарному виду путем элиминирования большой мощности на низких частотах (нестационарной низкочастотной компоненты) фильтром первых разностей, максимум вторых. Поэтому метод сезонной АРПСС удовлетворительно моделировал и прогнозировал временные ряды только относительно простой структуры.

Таким образом, произведена успешная попытка спустя более тридцати лет после создания метода Бокса-Дженкинса и метода «Гусеница»-SSA объединить их. Однако для удовлетворения таких требований к моделям, как: скорость обучения, трудоемкость, ресурсоемкость, наглядность модели, простота использования и интерпретируемость, затратный по времени и ресурсам метод «Гусеница»-SSA в дальнейшем было предложено использовать лишь для предварительной структурной идентификации и грубой параметрической идентификации так называемого интегрирующего полинома от оператора задержки предлагаемой модели, а также для грубой структурной и параметрической идентификации полинома от оператора задержки, наличие которого отличает более общую полиномиальную модель от модели Бокса-Дженкинса, структура и коэффициенты которого сначала равны таковыми рекуррентной модели прогнозирования метода «Гусеница»-SSA.

Метод «Гусеница»-SSA также предлагается использовать для предварительного обобщенного коинтегрирования временных рядов при моделировании многосвязных процессов, а также для разделения отдельно на финитный и отдельно на апериодический регуляторы в случае использования предлагаемой модели в теории автоматического управления [4]. Также возможно нелинейное усложнение передаточной функции модели одним из способов: FOS, GMDH, RBF, LARS, построенных на главных компонентах, их степенях и сочетаниях. Поэтому, в первую очередь, предлагаемая выше автором модель направлена на теорию автоматического управления, моделирование и прогнозирование технических систем и технологических процессов, в связи с тем, что их передаточные функции более детерминированные и имеют сложную нелинейную структуру.

3. Описание предлагаемых математических моделей

Математическая модель процессов, зависящих от нескольких экзогенных факторов в операторной форме может быть представлена в виде модели сезонной авторегрессии — проинтегрированного скользящего среднего (САРПССЭ) [5]:

$$z_t^y = \sum_{i=1}^N \frac{b_{n_{bi}}^i(L)}{a_{n_{ai}}^i(L)} \cdot z_{t-m_i}^{x_i} + \frac{c_{n_c}^{\Pi}(L)}{d_{n_d}^{\nabla}(L)} \cdot e_t, \quad (1)$$

где L — оператор сдвига по времени на одну единицу назад, такой что $Lx_t = x_{t-1}$, N — количество экзогенных переменных; z_t^y — пронормированный от 0 до 1 по формуле $z_t^y = \frac{y_t - y_t^{\min}}{y_t^{\max} - y_t^{\min}}$ или каким-либо другим спосо-

бом преобразованный временной ряд y_t моделируемого и прогнозируемого процесса с вычтенным математическим ожиданием; $z_{t-m_i}^{x_i}$ — пронормированный таким же образом i -й экзогенный временной ряд x_t^i с вычтенным математическим ожиданием; m_i — задержка i -го экзогенного временного ряда x_t^i по времени относительно прогнозируемого временного ряда y_t ; $a_{n_{ai}}^i(L)$, $b_{n_{bi}}^i(L)$ — полиномы от L степеней n_{ai} и n_{bi} соответственно;

$c_{n_c}^{\Pi}(L) = c_{n_c^1}^1(L^{s_1}) \cdot c_{n_c^2}^2(L^{s_2}) \times \dots \times c_{n_c^{n_s}}^{n_s}(L^{s_{n_s}}) = \prod_{i=1}^{n_s} c_{n_c^i}^i(L^{s_i})$ — полином L^{s_i} степени n_c^i , определяющий составляющую скользящего среднего периодической компоненты с периодом s_i , $n_c^{\Sigma} = \sum_{i=1}^{n_s} n_c^i \cdot s_i$;

$$\begin{aligned} d_{n_d}^{\nabla}(L) &= d_{n_d^{\Sigma}}^{\Pi}(L) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \dots \nabla_{s_{n_s}}^{D_{n_s}} = \\ &= d_{n_d^1}^1(L^{s_1}) \cdot d_{n_d^2}^2(L^{s_2}) \cdot \dots \cdot d_{n_d^{n_s}}^{n_s}(L^{s_{n_s}}) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \dots \nabla_{s_{n_s}}^{D_{n_s}} = \\ &= \prod_{i=1}^{n_s} d_{n_d^i}^i(L^{s_i}) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \dots \nabla_{s_{n_s}}^{D_{n_s}}, \\ n_d^{\nabla} &= n_d^{\Sigma} + \sum_{i=1}^{n_s} D_i \cdot s_i, \end{aligned}$$

$$d_{n_d^{\Sigma}}^{\Pi}(L) = d_{n_d^1}^1(L^{s_1}) \cdot d_{n_d^2}^2(L^{s_2}) \times \dots \times d_{n_d^{n_s}}^{n_s}(L^{s_{n_s}}) = \prod_{i=1}^{n_s} d_{n_d^i}^i(L^{s_i})$$

— полином от L^{s_i} степени n_d^i , определяющий составляющую авторегрессии сезонной компоненты с периодом s_i ,

$n_d^{\Sigma} = \sum_{i=1}^{n_s} n_d^i \cdot s_i$; e_t — остаточные ошибки модели; D_i — порядок взятия разности s_i ; ∇_{s_i} и L^{s_i} — упрощающие операторы такие, что $\nabla_{s_i} y_t = (1 - L^{s_i}) \cdot y_t = y_t - y_{t-s_i}$.

Выражение (1) в более компактной форме можно представить в виде:

$$\tilde{a}(q) \cdot z_t^y = \sum_{i=1}^k \tilde{b}^i(q) \cdot z_{t-m_i}^{x_i} + \tilde{c}(q) \cdot e_t,$$

где

$$\tilde{a}(L) = d_{n_d}^{\nabla}(L) \cdot \prod_{i=1}^N a_{n_{ai}}^i(L),$$

$$\tilde{b}^i(L) = b_{n_{bi}}^i(L) \cdot d_{n_d}^{\nabla}(L) \cdot \prod_{j=1, j \neq i}^N a_{n_{aj}}^j(L),$$

$$\tilde{c}(q) = c_{n_c}^{\Pi}(q) \cdot \prod_{i=1}^N a_{n_{ai}}^i(q), \quad i = \overline{1, N}.$$

Выражение для прогноза с упреждением l при помощи предложенной модели совместного использования модели метода «Гусеница»-SSA и сезонной модели авторегрессии — проинтегрированного скользящего среднего с экзогенными переменными, после приведения ее от дробно-разностного (рационального) к разностному уравнению принимает вид:

$$\begin{aligned} \hat{y}_t(l) = & \hat{y}_t^{SSA}(l) + \sum_{j=1}^{n_d^{\Sigma} + \sum_{i=1}^N n_{ai}} \tilde{a}_j \cdot h_{t+1-j}^y + \\ & + \sum_{i=1}^N \left(\tilde{b}_0^i \cdot x_{t+1-m_i}^i - \sum_{j=1}^{n_d^{\Sigma} + n_{bi} + \sum_{p=1}^N n_{ap}} \tilde{b}_j^i \cdot x_{t+1-m_i-j}^i \right) - \\ & - \sum_{j=1}^{n_c^{\Sigma} + \sum_{i=1}^N n_{ai}} \tilde{c}_j \cdot e_{t+1-j}; \hat{x}_t(l) = \hat{x}_t^{kSSA}(l) + \\ & + \sum_{j=1}^{n_x^{k\Sigma}} \tilde{a}_{xkj} \cdot z_{t+1-j}^k + \sum_{j=1}^{n_c^{k\Sigma}} \tilde{c}_{xkj} \cdot e_{xk,t+1-j}, k = \overline{1, N}, \end{aligned} \quad (2)$$

где $a_{n_{ai}}^i(L)$, $b_{n_{bi}}^i(L)$ — полиномы от L степеней n_{ai} и n_{bi} соответственно;

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

$$e_{t+j} = \begin{cases} e_{t+j}, & j \leq 0, \\ 0, & j > 0; \end{cases} \quad h_t^y = y_t - \tilde{w}_t^{yN+1};$$

$$\hat{y}_t^{SSA}(i) = \sum_{j=1}^{L-1} f_j^y \cdot \tilde{w}_{t+i-j}^{yN+1}, \quad i = \overline{1, L}; \quad \tilde{w}_t^{yN+1} = \begin{cases} \hat{y}_t, & i > t, \\ \tilde{w}_t^{yN+1}, & i \leq t; \end{cases}$$

$$(f_{L-1}^y, f_{L-2}^y, \dots, f_1^y)^T = \frac{1}{1 - \sum_{i=1}^R (u_i^L)^2} \sum_{i=1}^R u_i^L \cdot (u_1^i \ u_2^i \ \dots \ u_{L-1}^i)^T,$$

где $(u_1^i \ u_2^i \ \dots \ u_{L-1}^i)^T$ — вектор, состоящий из первых $(L-1)$ элементов i -того собственного вектора U^i сингулярного разложения траекторной матрицы экзогенных и моделируемого процессов

$$\begin{aligned} X = (X_1 \ X_2 \ \dots \ X_N \ Y) &= (X_{1,1} \ X_{1,2} \ \dots \ X_{1,K} \ X_{2,1} \ X_{2,2} \ \dots \\ & \dots \ X_{2,K} \ \dots \ X_{N,1} \ X_{N,2} \ \dots \ X_{N,K} \ Y_1 \ Y_2 \ \dots \ Y_K); \\ Y_j &= (y_{j-1} \ y_{j-2} \ \dots \ y_{j+L-2})^T, \\ X_{i,j} &= (x_{j-1}^i \ x_{j-2}^i \ \dots \ x_{j+L-2}^i), \\ i &= \overline{1, N}, \quad j = \overline{1, K} \end{aligned}$$

— преобразования исходного $(N+1)$ -мерного временного ряда (N — количество экзогенных временных рядов и один прогнозируемый) в последовательность Ly — мерных векторов (Ly — ширина окна), число которых равно $(N+1) \cdot K$, $K = n - Ly + 1$, где n — длина временных рядов; u_L^i — последний элемент вектора U^i ; R — количество элементов сингулярного разложения; $\hat{y}_t^{SSA}(i) = \sum_{j=1}^{L-1} f_j^y \cdot \tilde{w}_{t+i-j}^{yN+1}$ — модель рекуррентного прогнозирования метода «Гусеница»-SSA временного ряда y_t , $t = \overline{0, n-1}$; которая в свою очередь нередко может быть экономно записана сезонной моделью АРСС (АРП) или при помощи модели распределенных лагов Алмон, где $\tilde{w}_0^{yN+1}, \tilde{w}_1^{yN+1}, \dots, \tilde{w}_{n-1}^{yN+1}$ — ряд соответствующий преобразованию прогнозируемого временного ряда y_t ; ряды $\tilde{w}_0^{xi}, \tilde{w}_1^{xi}, \dots, \tilde{w}_{n-1}^{xi}$, $i = \overline{1, N}$ соответствуют преоб-

разованиям i -х временных рядов, соответствующих экзогенным временным рядам, при помощи сингулярного спектрального анализа на этапе диагонального усреднения, переводящего матрицу, состоящую из K столбцов от $(i-1) \cdot K$ -го до $i \cdot K - 1$ -го матрицы Z в ряды $\tilde{w}_0^{xi}, \tilde{w}_1^{xi}, \dots, \tilde{w}_{n-1}^{xi}$, по формуле

$$\tilde{w}_t^{yk} = \begin{cases} \frac{1}{i+1} \sum_{j=1}^{k+1} \tilde{z}_{j,i-j+2}^i, & i = \overline{0, \min(L, K) - 2}; \\ \frac{1}{\min(L, K)} \sum_{j=1}^{\min(L, K)} \tilde{z}_{j,i-j+2}^i, & i = \overline{\min(L, K) - 1, \max(L, K) - 1}; \\ \frac{1}{n-i} \sum_{j=k-\max(L, K)+2}^{n-\max(L, K)+1} \tilde{z}_{j,i-j+2}^i, & i = \overline{\max(L, K), n-1}, \end{cases}$$

$k = \overline{1, N+1}$; где $Z = \tilde{Z}^1 + \dots + \tilde{Z}^i$ — сумма матриц разложения

$$\begin{aligned} \tilde{Z}^i &= (U^i \cdot (U^i)^T \cdot X_1 \quad U^i \cdot (U^i)^T \cdot X_2 \quad \dots \\ & \dots \quad U^i \cdot (U^i)^T \cdot X_N \quad U^i \cdot (U^i)^T \cdot Y), \end{aligned}$$

отобранных стандартным анализом собственных чисел траекторной матрицы в методе «Гусеница»-SSA; $\hat{x}_t^{kSSA}(i) = \sum_{j=1}^{Ly-1} f_j^{xk} \cdot \tilde{w}_{t+i-j}^{xkN+1}$, $k = \overline{1, N}$, тоже самое, только для k -того экзогенного временного ряда, а также все переменные и параметры с индексом x^i , $i = \overline{1, N}$ аналогично интерпретируются как для временного ряда y_t , $t = \overline{1, n}$. Имеется и модификация рекуррентного метода SSA-прогнозирования — векторное SSA-прогнозирование [6], которое в ряде случаев позволяет получать более точные прогнозы.

Предлагаемая модель позволяет моделировать и прогнозировать зашумленные циклические компоненты временного ряда, модулированных по амплитуде и частоте, благодаря своей возможности, в отличие от метода Фурье, выделять колебательные составляющие с модулированной амплитудой и частотой и различные шумовые составляющие с изменяющейся дисперсией.

С нелинейным усложнением передаточной функции выражение (2) принимает вид:

$$\begin{aligned} \hat{z}_t^y(l) = & \hat{y}_t^{SSA}(l) + \sum_{i=1}^r g_i \cdot p_t^i + \sum_{j=1}^{n_d^{\Sigma} + \sum_{i=1}^N n_{ai}} \tilde{a}_j \cdot h_{t+1-j}^y + \\ & + \sum_{i=1}^N \left(\tilde{b}_0^i \cdot z_{t+1-m_i}^i - \sum_{j=1}^{n_d^{\Sigma} + n_{bi} + \sum_{p=1}^N n_{ap}} \tilde{b}_j^i \cdot z_{t+1-m_i-j}^i \right) - \sum_{j=1}^{n_c^{\Sigma} + \sum_{i=1}^N n_{ai}} \tilde{c}_j \cdot e_{t+1-j}, \\ \hat{x}_t^i(l) = & \hat{x}_t^{kSSA}(l) + \sum_{j=1}^{n_x^{k\Sigma}} \tilde{a}_{xkj} \cdot z_{t+1-j}^k + \sum_{j=1}^{n_c^{k\Sigma}} \tilde{c}_{xkj} \cdot e_{xk,t+1-j}, k = \overline{1, N}, \end{aligned}$$

где $g_i \cdot p_t^i$ — члены полинома Колмогорова — Габора:

$$\begin{aligned} \sum_{i=1}^r g_i \cdot p_t^i = & \text{FOS} \left(\sum_{i=1}^M \tilde{g}_i \cdot \tilde{x}_t^i + \sum_{i=1}^M \sum_{j=1}^M \tilde{g}_{ij} \cdot \tilde{x}_t^i \cdot \tilde{x}_t^j \right. \\ & \left. + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \tilde{g}_{ijk} \cdot \tilde{x}_t^i \cdot \tilde{x}_t^j \cdot \tilde{x}_t^k + \dots \right), \quad g_n = \tilde{g}_{j \dots k}, \quad n = \overline{1, r} \end{aligned}$$

или

$$\sum_{i=1}^r g_i \cdot p_i^t = \text{FOS} \left(\sum_{i=1}^{M_2} \tilde{g}_i \cdot \varphi_i(\tilde{x}_t) \right),$$

где

$$\varphi_i(\tilde{x}_t) = \frac{1}{(2\pi)^{-\frac{M}{2}} \cdot |\Sigma_i|^{-\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\tilde{x}_t - \bar{c}_i) \Sigma_i^{-1} (\tilde{x}_t - \bar{c}_i)^T},$$

\bar{c}_i — вектор математических ожиданий временных рядов, представляющих главные компоненты, $\Sigma_i, i = \overline{1, M}$ — ковариационные матрицы; или

$$\sum_{i=1}^r g_i \cdot p_i^t = \text{GMDH}(\tilde{x}_t^1, \tilde{x}_t^2, \dots, \tilde{x}_t^M)$$

— структурная идентификация методом МГУА, $p_i^t = \tilde{x}_t^1 \cdot \tilde{x}_t^2 \cdot \dots \cdot \tilde{x}_t^M$, составленного из главных компонент временного ряда $z_i^y - \tilde{w}_i^{N+1}$ обозначенных $\tilde{x}_t^i, i = \overline{1, M}$, а также их степеней и сочетаний, отобранных при помощи FOS-алгоритма, определяющего нелинейную часть предложенной модели; FOS-функция структурного упрощения модели, записанной в ее аргументе, с учетом характера поведения временных рядов, при помощи алгоритма быстрого ортогонального поиска; $h_i^y = z_i^y - \tilde{w}_i^{N+1} - \sum_{i=1}^r g_i \cdot p_i^t$.

Процесс нахождения таких комбинированных моделей (2) совместного использования модели АРПСС и метода «Гусеница»-SSA может быть продолжительным в связи с ресурсоемкостью метода «Гусеница»-SSA. Поэтому метод «Гусеница»-SSA анализа предложено использовать лишь для предварительной структурной идентификации и грубой параметрической идентификации интегрирующего полинома $w(L)$ (отсюда и название модель авторегрессии — спектрально проинтегрированного скользящего среднего (АРСПСС)) от оператора задержки L , который также может быть интерпретирован как оператор перевода в пространство состояний, модели

$$f(L) \cdot w(L) \cdot y_t = \frac{c(L)}{d(L)} \cdot e_t \text{ — модель АРСПСС,} \quad (3)$$

а рекуррентный метод SSA-прогнозирования для грубой структурной и параметрической идентификации полинома $f(L)$, структура и коэффициенты которого сначала равны таковыми рекуррентной модели прогнозирования метода «Гусеница»-SSA и наличие которого отличает более общую полиномиальную модель от модели Бокса-Дженкинса, а также совместно с $w(L)$ ($w(L)$ вообще говоря может быть рационального вида) определяющих долгосрочную память модели, описывая более широкий класс процессов долгосрочной памяти, чем при фрактальном интегрировании в модели ARFIMA, которая в свою очередь была придумана для преодоления недостатка ARIMA моделей при моделировании и прогнозировании процессов в длинной памяти — потери (искажения) долгосрочной информации при взятии приращений. Полиномы $d(L)$ и $c(L)$, в свою очередь, определяют краткосрочные зависимости процесса. Модель АРСПСС может быть также построена опираясь не на метод «Гусеница»-SSA, а на метод главных компонент.

При анализе и прогнозировании временных рядов, зависящих от нескольких других необходима сбалансированность динамических свойств переменных, стоящих в левой и правой частях уравнения модели. В этом случае идеи метода «Гусеница»-SSA выступают для предвари-

тельной обобщенной коинтеграции временных рядов и модель разбивается следующим образом:

$$\begin{aligned} \hat{w}_t^y &= \frac{b_{n_{by}}^y(L)}{a_{n_{ay}}^y(L)} \cdot z_t^y + \sum_{i=1}^N \frac{b_{n_{bi}}^{wy}(L)}{a_{n_{ai}}^{wy}(L)} \cdot z_{t-m_i}^{x_i} + \frac{c_{n_{cy}}^{wy}(L)}{d_{n_{dy}}^{wy}(L)} \cdot e_t^{wy}; \\ z_t^y &= f^y(L) \cdot \hat{w}_t^y + \sum_{i=1}^N \frac{b_{n_{bi}}^i(L)}{a_{n_{ai}}^i(L)} \cdot z_{t-m_i}^{x_i} + \frac{c_{n_{ci}}^\pi(L)}{d_{n_{di}}^\pi(L)} \cdot e_t; \\ f^{x_i}(L) \cdot \omega(L) \cdot z_t^{x_i} &= \frac{c_{n_{xi}}^{x_i}(L)}{d_{n_{di}}^{x_i}(L)} \cdot e_t^{x_i}, \quad j = \overline{1, N}, \end{aligned} \quad (4)$$

\hat{w}_t^y — аппроксимация временным рядом \hat{y}_t и экзогенными временными рядами $x_t^j, j = \overline{1, N}$ временного ряда \tilde{w}_t^y при помощи модели сезонной АРПССЭ (или АРПЭ — проинтегрированной авторегрессии с экзогенными переменными), изначально полученного методом «Гусеница»-SSA, а впоследствии подстраиваемого оптимизационным методом при конкурентном обучении модели; $\omega(L)$ — интегрирующий полином, переводящий временной ряд x_t^j во временной ряд $\hat{w}_t^{x_j}$ — аппроксимация временного ряда $\tilde{w}_{x_t^j}^{N+1}$ моделью АРПСС; начальные грубые значения коэффициентов полиномов $f^y(L), f^{x_i}(L)$ и их количество берутся равными коэффициентам f_j^y и $f_j^{x_i}, j = \overline{1, N}$ моделей рекуррентного SSA-прогнозирования $\hat{y}_t^{SSA}(i) = \sum_{j=1}^{L_y-1} f_j^y \cdot \tilde{w}_{t+i-j}^{y, N+1}$ и $\hat{x}_t^{k, SSA}(i) = \sum_{j=1}^{L_{x_i}-1} f_j^{x_i} \cdot \tilde{w}_{t+i-j}^{x_i, N+1}$ соответственно; L_y и $L_{x_i}, i = \overline{1, N}$ — соответствующие длины окон; а затем итерационно подстраивать вместе с остальными коэффициентами модели (4) при помощи метода Левенберга-Марквардта. Модель (3) существенно выигрывает по времени обучения комбинированную модель совместного использования сезонной модели АРССЭ и метода «Гусеница»-SSA (2), но несколько уступает ей по статистическим свойствам и с учетом способа ее построения названа сезонной моделью авторегрессии — спектрально проинтегрированного скользящего среднего с экзогенными переменными (АРСПССЭ, ARSIMAX) и может быть записана следующим образом:

$$\begin{aligned} \hat{z}_t^y(l) &= \sum_{j=1}^{L_y + n_{by} + \sum_{i=1}^N n_{ai}^{wy} + n_{dy}^{wy} + \sum_{i=1}^N n_{ai} + n_{dy}} \tilde{a}_j^y \cdot z_{t+l-j}^y + \\ &+ \sum_{i=1}^N \left(\tilde{b}_0^i \cdot z_{t+l-m_i}^{x_i} - \sum_{j=1}^{n_{bi}^i} \tilde{b}_j^i \cdot z_{t+l-m_i-j}^{x_i} \right) - \\ &- \sum_{j=1}^{n_{cy}^{wy} + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + n_{dy}^{wy} + \sum_{i=1}^N n_{ai} + n_{dy}} \tilde{c}_j^y \cdot e_{t+l-j} - \\ &- \sum_{j=1}^{L_y + n_{wy}^c + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + \sum_{i=1}^N n_{ai} + n_{dy}} \tilde{d}_j^y \cdot e_{t+l-j}^y; \\ \hat{z}_t^{x_k}(l) &= \sum_{j=1}^{L_{x_k} + n_{x_k}^d} \tilde{a}_j^{x_k} \cdot z_{t+l-j}^{x_k} + \sum_{j=1}^{n_{x_k}^c} \tilde{c}_j^{x_k} \cdot e_{t+l-j}^{x_k}, \quad k = \overline{1, N}, \end{aligned}$$

где $\tilde{a}_j^y, j = \overline{1, L_y + n_{by} + \sum_{i=1}^N n_{ai}^{wy} + n_{dy}^{wy} + \sum_{i=1}^N n_{ai} + n_{dy}}$ — коэффициенты полинома

$$\tilde{a}(L) = f^y(L) \cdot b_{n_{by}}^y(L) \cdot \prod_{i=1}^N a_{n_{ai}^{wy}}^{wy,i}(L) \cdot d_{n_d^v}^v(L); \tilde{b}_j^i,$$

$$i = \overline{1, N}, j = 1, 2, \dots, n^{\tilde{b}};$$

$$n^{\tilde{b}} = Ly + \max \left(n_{bi}^{wy} + \sum_{j=1, j \neq i}^N n_{aj}^{wy} + n_{ay} + n^{wy} \nabla_d + \sum_{j=1, j \neq i}^N n_{aj} + n_d^v, \right.$$

$$\left. j = \overline{1, N}, n_{bi} + n_{ay} + \sum_{j=1, j \neq i}^N n_{aj}^{wy} + n^{wy} \nabla_d + \sum_{j=1, j \neq i}^N n_{aj} + n_d^v \right)$$

— коэффициенты полинома

$$\tilde{b}_i(L) = f^y(L) \cdot \sum_{i=1}^N \left(b_{n_{bi}^{wy}}^{wy,i}(L) \cdot \prod_{j=1, j \neq i}^N a_{n_{aj}^{wy}}^{wy,j}(L) \cdot a_{n_{ay}}^y(L) \cdot d_{n_d^v}^{wy \nabla}(L) \cdot \prod_{j=1, j \neq i}^N a_{n_{aj}}^j(L) \cdot d_{n_d^v}^v(L) \right) -$$

$$- \sum_{i=1}^N \left(b_{n_{bi}}^i(L) \cdot a_{n_{ay}}^y(L) \cdot \prod_{j=1, j \neq i}^N a_{n_{aj}^{wy}}^{wy,j}(L) \cdot d_{n_d^v}^{wy \nabla}(L) \cdot \prod_{j=1, j \neq i}^N a_{n_{aj}}^j(L) \cdot d_{n_d^v}^{wy \nabla}(L) \right);$$

$$\tilde{c}_j, j = 1, n^{\tilde{c}} + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + n^{wy} \nabla_d + \sum_{i=1}^N n_{ai} + n_d^v$$

— коэффициенты полинома

$$\tilde{c}(L) = c_{n_{\tilde{c}}}^{\Pi}(L) \cdot a_{n_{ay}}^y(L) \cdot \prod_{i=1}^N a_{n_{ai}^{wy}}^{wy,i}(L) \cdot d_{n_d^v}^{wy \nabla}(L) \times$$

$$\times \prod_{i=1}^N a_{n_{ai}}^i(L) \cdot d_{n_d^v}^v(L), i = \overline{1, N};$$

$$\tilde{d}_j, j = 1, Ly + n^{wy} \nabla_{\tilde{c}} + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + \sum_{i=1}^N n_{ai} + n_d^v$$

— коэффициенты полинома

$$\tilde{d}(L) = f^y(L) \cdot c_{n^{wy} \nabla_{\tilde{c}}}^{\Pi}(L) \cdot a_{n_{ay}}^y(L) \cdot \prod_{i=1}^N a_{n_{ai}^{wy}}^{wy,i}(L) \times$$

$$\times \prod_{i=1}^N a_{n_{ai}}^i(L) \cdot d_{n_d^v}^v(L), i = \overline{1, N};$$

$$\tilde{a}_j^{x^i}, j = 1, L^{x^k} + L^{\omega} + n^{x^k} \nabla_d$$

— коэффициенты полинома

$$\tilde{a}^{x^i}(L) = f^{x^i}(L) \cdot \omega(L) \cdot d_{n_{xi}^v}^{x^i \nabla}(L);$$

$$\tilde{c}_j^{x^i}, j = \overline{1, n^{x^k} \nabla_{\tilde{c}}} \text{ — коэффициенты полинома}$$

$$\tilde{c}^{x^i}(L) = c_{n^{x^i} \nabla_{\tilde{c}}}^{x^i \Pi}(L).$$

Для учета гетероскедастичности процесса (изменения дисперсии процесса во времени) применяется обобщенная модель с авторегрессионной условной гетероскедастичностью GARCH(m, r), имеющая вид [7]:

$$\sigma_t^2 = w + \theta(L) \varepsilon_t^2 + \varphi(L) \sigma_t^2,$$

где σ_t^2 — временной ряд изменения дисперсии процесса y_t ,

$$\theta(L) = \theta_1 L + \theta_2 L^2 + \dots + \theta_p L^m,$$

$$\varphi(L) = \varphi_1 L + \varphi_2 L^2 + \dots + \varphi_r L^r,$$

ε_t^2 — остаточные члены модели. Модель GARCH(m, r) может быть записана через модель APCC(s, m) следующим образом [8]:

$$\varepsilon_t^2 = \frac{w + (1 - \varphi(q))}{(1 - \theta(q) - \varphi(q))} v_t,$$

где $s = \max(r, m)$, $v_t = \varepsilon_t^2 - \sigma_t^2$.

Фрактально проинтегрированный процесс GARCH может быть записан следующим образом:

$$(1-L)^{2-H} \varepsilon_t^2 = \frac{w + (1 - \varphi(L))}{(1 - \theta(L) - \varphi(L)) \cdot (1-L)^{-1}} v_t,$$

где H — показатель Херста.

Предлагаемая спектрально проинтегрированная обобщенная модель с авторегрессионной условной гетероскедастичностью имеет вид:

$$f^{\varepsilon^2}(q) \cdot \omega^{\varepsilon^2}(q) \cdot \varepsilon_t^2 = \frac{w + (1 - \varphi(q))}{(1 - \theta(q) - \varphi(q))} \cdot v_t,$$

где $\omega^{\varepsilon^2}(L)$ — интегрирующий полином от оператора задержки, с помощью которого аппроксимируется временной ряд дисперсии шумов ε_t^2 в преобразованный сглаженный методом «Гусеница»-SSA временной ряд $w_t^{\varepsilon^2}$, а предварительная структурная идентификация и грубая параметрическая идентификация полинома от оператора задержки $f^{\varepsilon^2}(L)$, с помощью которого аппроксимируется сам ряд ε_t^2 производится при определении коэффициентов рекуррентной прогнозирующей формулы метода «Гусеница»-SSA; w — среднее значение или уровень временного ряда ε_t^2 .

Таким образом, модель авторегрессии — спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщенной авторегрессионной условной гетероскедастичностью и экзогенными переменными (АРСПСС — СПОАРУТ, ARSIMA — SIGARCH) принимает вид:

$$\hat{z}_t^y(l) = \sum_{j=1}^{Ly + n_{by} + \sum_{i=1}^N n_{ai}^{wy} + n^{wy} \nabla_d + \sum_{i=1}^N n_{ai} + n_d^v} \tilde{a}_j \cdot z_{t+1-j}^y +$$

$$+ \sum_{i=1}^N \left(\tilde{b}_0^i \cdot z_{t+1-m_i}^{x^i} - \sum_{j=1}^{n^{\tilde{b}}} \tilde{b}_j^i \cdot z_{t+1-m_i-j}^{x^i} \right) -$$

$$- \sum_{j=1}^{n_{\tilde{c}} + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + n^{wy} \nabla_d + \sum_{i=1}^N n_{ai} + n_d^v} \tilde{c}_j \cdot e_{t+1-j} -$$

$$- \sum_{j=1}^{Ly + n^{wy} \nabla_{\tilde{c}} + n_{ay} + \sum_{i=1}^N n_{ai}^{wy} + \sum_{i=1}^N n_{ai} + n_d^v} \tilde{d}_j \cdot e_{t+1-j};$$

$$\hat{z}_t^{x^k}(l) = \sum_{j=1}^{L^{x^k} + n_{x^k} \nabla_d} \tilde{a}_j^{x^k} \cdot z_{t+1-j}^{x^k} + \sum_{j=1}^{n^{x^k} \nabla_{\tilde{c}}} \tilde{c}_j^{x^k} \cdot e_{t+1-j}^{x^k}, k = \overline{1, N},$$

$$\varepsilon_t^2 = \sigma_t z_t,$$

$$z_t \sim N(0, 1),$$

$$f^{\varepsilon^2}(L) \cdot \omega^{\varepsilon^2}(L) \cdot \varepsilon_t^2 = \frac{w + (1 - \varphi(L))}{(1 - \theta(L) - \varphi(L))} \cdot v_t.$$

Модель может быть обобщена для многомерного случая.

Такие полиномиальные модели позволяют адекватно описывать чрезвычайно широкий класс стохастических систем и процессов и решать на их основе задачи моделирования, прогнозирования и управления [9].

4. Декомпозиционная модель прогнозирования (ДМП)

Возвращаясь к моделям (3), (4) предложено идентифицировать и строить АРПССЭ (ARIMAX) модели не ряда \widehat{w}_t^y и соответственно рядов \tilde{w}_0^{xi} , $\tilde{w}_1^{xi}, \dots, \tilde{w}_{n-1}^{xi}$, $i = \overline{1, N}$, полученных на четвертом этапе диагонального усреднения метода «Гусеница»-SSA матриц Z^i , $i = \overline{1, N+1}$, состоящих из K столбцов от $(i-1) \cdot K$ -го до $i \cdot K - 1$ -го матрицы Z , где $Z = \tilde{Z}^1 + \dots + \tilde{Z}^j$ — сумма матриц разложения

$$\tilde{Z}^i = \begin{pmatrix} U^i \cdot (U^i)^T \cdot X_1 & U^i \cdot (U^i)^T \cdot X_2 & \dots \\ \dots & U^i \cdot (U^i)^T \cdot X_N & U^i \cdot (U^i)^T \cdot Y \end{pmatrix},$$

отобранных стандартным анализом собственных чисел траекторной матрицы в методе «Гусеница»-SSA, а идентифицировать и строить АРПССЭ (ARIMAX) модели каждого из временных рядов $\tilde{w}_t^{(j)}$ диагонального усреднения подматриц \tilde{Z}^{ji} , $i = \overline{1, Ly}$, состоящих из K столбцов от $(j-1) \cdot K$ -го до $j \cdot K - 1$ -го матрицы \tilde{Z}^i , $i = \overline{1, Ly}$. Таким образом, потенциально можно сформировать $Ly \times (N+1)$ временных рядов разложения (компонент) исходного прогнозируемого временного ряда и экзогенных временных рядов (или $Ly \times (N+1)$ временных рядов факторных векторов, временные затраты на вычисление которых меньше), подлежащих, как было указано выше, моделированию, после отбора из этих компонент конструктивных и удаления деструктивных. Однако нет необходимости формировать все $Ly \times (N+1)$ моделей, а формируются только значимые, т. е. моделирующие конструктивные компоненты. Для отбора наиболее конструктивных компонент (наиболее коррелируемых с прогнозируемым временным рядом), как и в других работах автора, предлагается использовать алгоритм быстрого ортогонального поиска (FOS). При этом столбцами рабочей матрицы алгоритма являются значения полученных методом «Гусеница»-SSA компонент разложения прогнозируемого и экзогенных временных рядов, исходные значения прогнозируемого и экзогенных временных рядов, а также их задержки вплоть до m -той (где значение m предлагается выбирать не более 3). Эта процедура необходима для формирования передаточной функции ARIMAX модели второго слоя предлагаемой модели. Построение же моделей конструктивных компонент необходимо, для дальнейшего использования их как гребенки фильтров, широко применяемой в задачах моделирования радиосигналов.

$$\begin{aligned} \widehat{w}_t^{(j)} &= \frac{b_{n_b(j)}^{(j)}(L)}{a_{n_a(j)}^{(j)}(L)} \cdot z^y \cdot z^{t-m_i^{y w(j)}} + \sum_{i=1}^N \frac{b_x^{w(j)i}(L)}{a_x^{w(j)i}(L)} \cdot z^{x_i} \cdot z^{t-m_i^{x w(j)}} + \\ &\sum_{i=1, i \neq j}^M \frac{b_w^{w(j)i}(L)}{a_w^{w(j)i}(L)} \cdot \widehat{w}_{t-m_i^{w(i)}}^{(i)} + \frac{c_w^{(j)\pi}(L)}{d_w^{(j)\nu}(L)} \cdot e_t^{w(j)}; \\ &\vdots \\ \widehat{w}_t^{(k)} &= \frac{b_{n_b(k)}^{(k)}(L)}{a_{n_a(k)}^{(k)}(L)} \cdot z^y \cdot z^{t-m_i^{y w(k)}} + \sum_{i=1}^N \frac{b_x^{w(k)i}(L)}{a_x^{w(k)i}(L)} \cdot z^{x_i} \cdot z^{t-m_i^{x w(k)}} + \\ &\sum_{i=1, i \neq j}^M \frac{b_w^{w(k)i}(L)}{a_w^{w(k)i}(L)} \cdot \widehat{w}_{t-m_i^{w(i)}}^{(i)} + \frac{c_w^{(k)\pi}(L)}{d_w^{(k)\nu}(L)} \cdot e_t^{w(k)}; \\ z_t^y &= \sum_{i=1}^M \frac{b_{n_b(i)}^i(L)}{a_{n_a(i)}^i(L)} \cdot \widehat{w}_{t-m_i^y}^{(i)} + \sum_{i=1}^N \frac{b_{n_b(i)}^{y_i}(L)}{a_{n_a(i)}^{y_i}(L)} \cdot z^{x_i} + \frac{c_{n_b(i)}^\pi(L)}{d_{n_a(i)}^\nu(L)} \cdot e_t. \end{aligned} \quad (5)$$

Таким образом $\left(\widehat{w}_t^{(j)}, \dots, \widehat{w}_t^{(k)} \right)$ в каждый момент времени можно интерпретировать как значения модели в пространстве состояний, а совокупность их моделей как гребенку фильтров, подобно методам и моделям адаптивной цифровой фильтрации, моделям в пространстве состояний и прогнозирующим фильтрам Калмана-Бьюси, способных решать большой класс задач с минимальными вычислительными затратами. В книгах [5], [15] описано как идентифицировать передаточную функцию модели ARIMAX процедурой выбеливания спектра. Однако при идентификации передаточной функции ARIMAX модели с более чем одной экзогенной переменной рассмотрение процедуры выбеливания спектра перебирая все варианты функций отклика на единичный импульс является затруднительным. Поэтому для идентификации передаточной функции модели с большим количеством экзогенных переменных автор рекомендует использовать алгоритм быстрого ортогонального поиска, столбцы рабочей матрицы которого состоят из временных рядов разложения и их задержек не выше третьей. В остальном идентификация каждой из моделей входящей в состав выражения (5) на высоком уровне описано в книге [5].

На рис. 1 представлена структурная схема предложенной модели декомпозиционного метода прогнозирования (ДМП). В качестве математических моделей автор предлагает использовать модели ARIMAX (рис. 2) или их производные ARSIMAX, NARIMAX и пр. Для придания моделям ARIMAX нелинейности можно добавлять в них нелинейные составляющие, как это было указано выше, путем составления полинома Колмогорова — Габора $\sum_{i=1}^r g_i \cdot p_t^i$, где $g_i \cdot p_t^i$ — члены полинома Колмогорова — Габора:

$$\begin{aligned} \sum_{i=1}^r g_i \cdot p_t^i &= \text{FOS} \left(\sum_{i=1}^M \tilde{g}_i \cdot \tilde{x}_t^i + \sum_{i=1}^M \sum_{j=1}^M \tilde{g}_{ij} \cdot \tilde{x}_t^i \cdot \tilde{x}_t^j + \right. \\ &\left. + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \tilde{g}_{ijk} \cdot \tilde{x}_t^i \cdot \tilde{x}_t^j \cdot \tilde{x}_t^k + \dots \right), \quad g_n = \tilde{g}_{ij\dots k}, \quad n = \overline{1, r} \end{aligned}$$

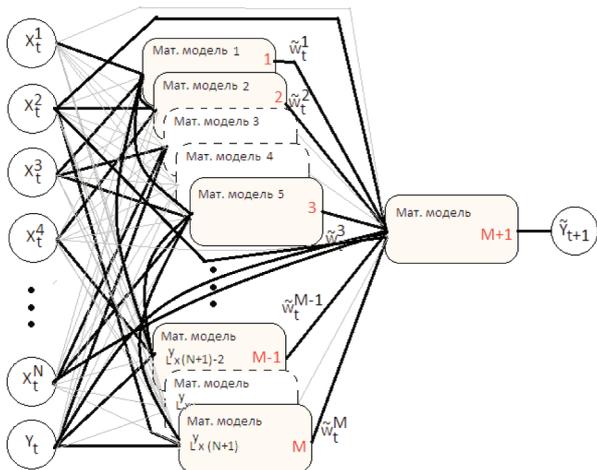


Рис. 1. Структурная схема модели декомпозиционного метода прогнозирования

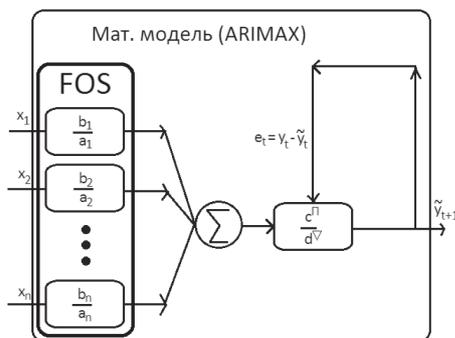


Рис. 2. Структурная схема ARIMAX модели

или

$$\sum_{i=1}^r g_i \cdot p_i^t = \text{FOS} \left(\sum_{i=1}^{M_2} \tilde{g}_i \cdot \varphi_i(\tilde{x}_t) \right),$$

где

$$\varphi_i(\tilde{x}_t) = \frac{1}{(2\pi)^{-\frac{M}{2}} \cdot |\Sigma_i|^{-\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\tilde{x}_t - \bar{c}_i) \cdot \Sigma_i^{-1} \cdot (\tilde{x}_t - \bar{c}_i)^T},$$

\bar{c}_i — вектор математических ожиданий временных рядов, представляющих главные компоненты, $\Sigma_i, i = 1, M$ — ковариационные матрицы; или

$$\sum_{i=1}^r g_i \cdot p_i^t = \text{GMDH}(\tilde{x}_t^1, \tilde{x}_t^2, \dots, \tilde{x}_t^M)$$

— структурная идентификация методом МГУА, $p_i^t = \tilde{x}_t^1 \cdot \tilde{x}_t^2 \cdot \dots \cdot \tilde{x}_t^k$, составленного из главных компонент временного ряда $z_i^y - \tilde{w}_i^{N+1}$ обозначенных $\tilde{x}_t^i, i = 1, M$, а также их степеней и сочетаний, отобранных при помощи FOS-алгоритма, определяющего нелинейную часть предложенной модели.

5. Декомпозиционная искусственная нейронная сеть

Используя такие же идеи декомпозиции временного ряда на конструктивные компоненты строится и декомпозиционная искусственная нейронная сеть. Методом «Гусеница»-SSA производится предварительная структурная идентификация значений выходов (аксонов) первого слоя искусственной нейронной сети, в свою очередь состоящего из искусственных нейронных сетей.

Предложенные декомпозиционные методы производят самоорганизацию моделей, определяя существенные связи и отбрасывая несущественные, производя тем самым структурную идентификацию предложенных моделей.

На рис. 3 представлена структурная схема предложенной декомпозиционной нейронной сети. В качестве искусственных нейронных сетей составляющих первый и второй слой общей нейронной сети автором предлагается использовать радиально-базисные искусственные нейронные сети, структурная и параметрическая идентификация которых, производится традиционно рекомендуемым автором алгоритмом быстрого ортогонального поиска (FOS), описанного в работе [6] или МГУА.

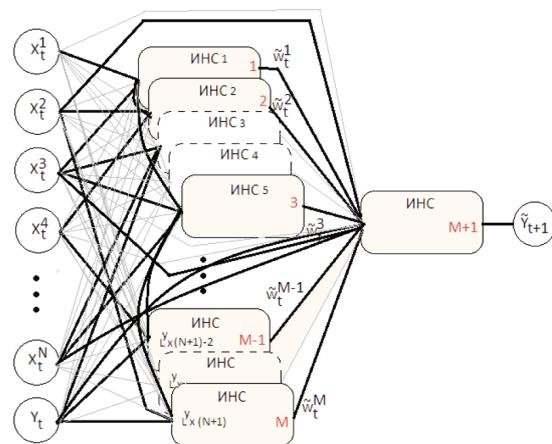


Рис. 3. Структурная схема декомпозиционной искусственной нейронной сети

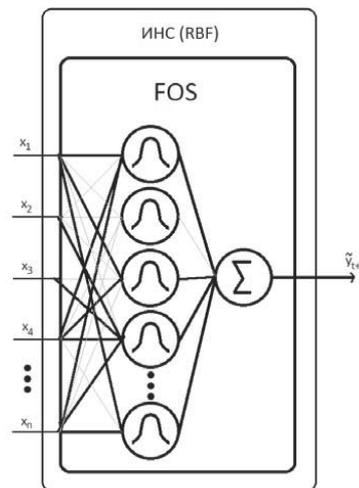


Рис. 4. Структурная схема радиально-базисной искусственной нейронной сети

6. Результаты исследований

Тестирование предложенной модели декомпозиционного метода прогнозирования проводилось на реальных среднесуточных данных потребления газа и изменения температуры воздуха за трехлетний интервал времени.

По факторным временным рядам, соответствующих большим и близким по величине собственным значениям информационной матрицы данных, можно судить о недельных и годовых сезонных составляющих данных временных рядов.

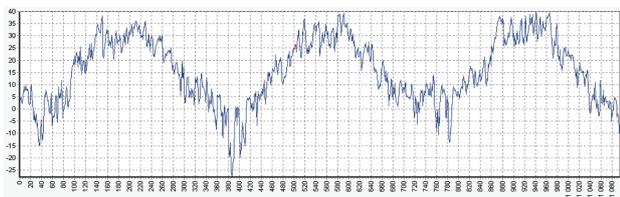


Рис. 5. График среднесуточных данных изменения температуры воздуха

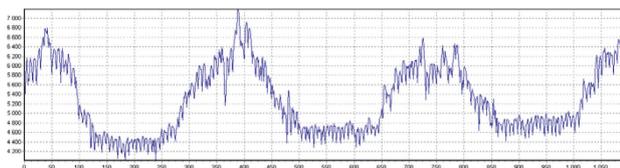


Рис. 6. График среднесуточных данных потребления газа

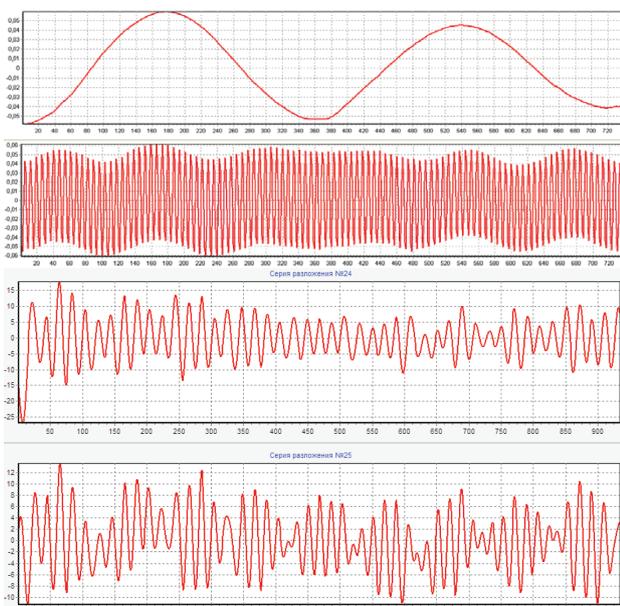


Рис. 7. Графики некоторых факторных временных рядов, приведенных процессов

После этапа разложения при структурной идентификации модели второго слоя предложенного декомпозиционного метода прогнозирования (при структурной идентификации последней модели, стоящей в выражении (5)) в ее состав предложенный метод не включил переменные разложения. Это говорит о том, что рассматриваемые процессы имеют достаточно простую структуру и необходимости в моделировании первого слоя общей модели декомпозиционного метода моделирования нет. В результате, предложенный декомпозиционный метод моделирования идентифицировал частный случай своей модели — сезонную модель АРССЭ.

Полученная модель САРССЭ для прогнозирования потребления природного газа с учетом изменения температуры воздуха приняла вид

$$\nabla \nabla_7 \nabla_{365} Y_t = \frac{-0,659 + 0,191L}{1 - 0,716L + 0,261L^2} \nabla \nabla_7 \nabla_{365} X_t + \frac{(1 - 0,915L) \cdot (1 - 0,748L + 0,254L^4)}{(1 - 0,773L - 0,155L^2) \cdot (1 - 0,401L^7) \cdot (1 + 0,188L^{365})} e_t.$$

Средний процент ошибки прогнозирования составил 1,87 %.

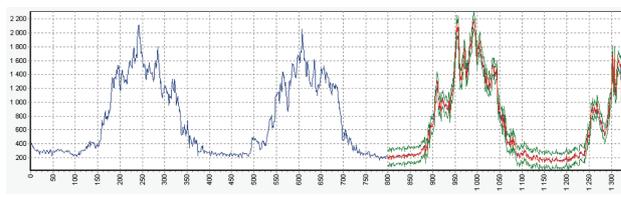


Рис. 8. Графики прогнозов потребления природного газа моделью САРССЭ и 95 % доверительные интервалы

Однако при моделировании временного ряда электроэнцефалограммы модель декомпозиционного метода прогнозирования получила существенно лучшие результаты по среднеквадратической ошибке, чем традиционная модель АРСС.

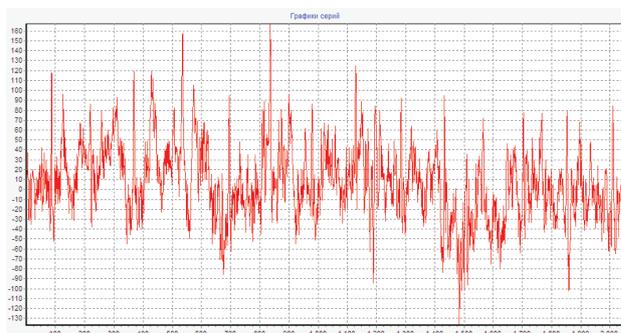


Рис. 9. График временного ряда электроэнцефалограммы

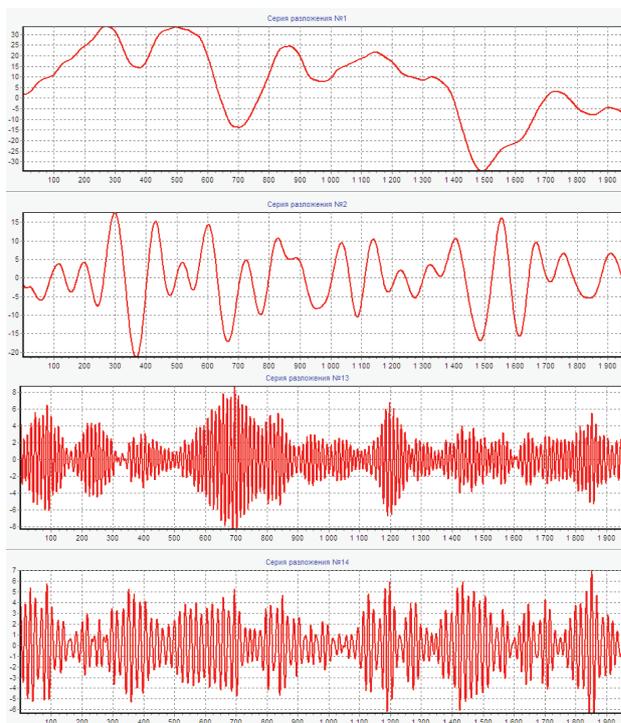


Рис. 10. Графики некоторых временных рядов разложения временного ряда электроэнцефалограммы

По результатам алгоритма быстрого ортогонального поиска можно судить о необходимости отбора в качестве конструктивных компонент временные ряды разложения с номерами 12, 4, 11, 1, 2, 3, 4, 5, 7 временные ряды разложения для дальнейшего их моделирования и включения

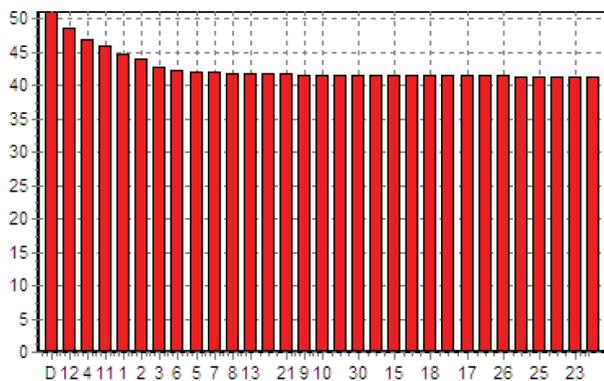


Рис. 11. Результаты применения алгоритма FOS для отбора конструктивных компонент разложения временного ряда электроэнцефалограммы

моделей в состав общей модели декомпозиционного метода прогнозирования.

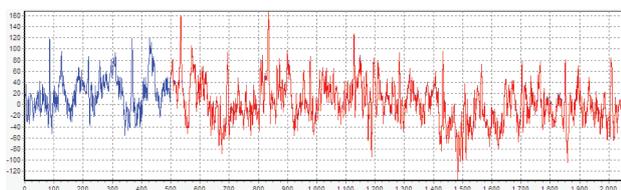


Рис. 12. Графики исходного временного ряда электроэнцефалограммы и восстановленного моделью декомпозиционного метода прогнозирования

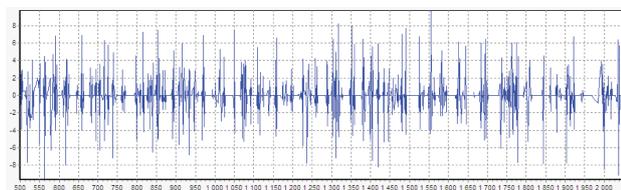


Рис. 13. График остаточных ошибок модели декомпозиционного метода прогнозирования

Среднеквадратическая ошибка моделирования составила 2,37.

Выводы

Таким образом, для получения адекватных моделей сложных процессов, высококачественных прогнозов и качественного управления необходимо комбинировать модели с разными структурами, включая нелинейные модели, которые являются взаимодополняющими при их конкурентном обучении.

Предлагаемый метод «Гусеница»-SSA — АРССС — СПОАРУГ является модификацией метода «Гусеница»-SSA с автоматическим отделением долгосрочной памяти от краткосрочной и периодических составляющих, а предложенная модель АРССС — СПОАРУГ, как и модель ARFIMA — FIGARCH, является очередной модификацией модели ARIMA — GARCH, а метод построения предлагаемой модели является развитием метода Бокса-Дженкинса, но для моделирования более широкого класса процессов. Модель декомпозиционного метода прогнозирования может быть интерпретирована как

развитие моделей в пространстве состояний, позволяет реализовать общую модель как совокупность взаимосвязанных более простых моделей. Такая структура модели позволяет повысить адекватность и точность моделирования и прогнозирования нелинейных нестационарных многомерных процессов, повысить устойчивость идентификации процесса и упростить ее.

Предложенная модель АРССС — СПОАРУГ, модель декомпозиционного метода прогнозирования и методики их построения являются некоторым промежуточным подходом между классическими регрессионными и современными нейросетевыми и более формализованные по выбору структуры, являясь при этом оптимальной, по мнению автора, по детализации с учетом существующих на сегодняшний день математических и машинных как достоинств и достижений так и недостатков и ограничений.

Подводя итоги описанным выше преимуществам предлагаемых математических моделей, еще раз следует отметить, что основная идея состоит в эффекте синергии, который возникает в результате комбинированного применения идей двух методов: метода «Гусеница»-SSA и метода Бокса-Дженкинса.

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

Литература

1. Granger C. W. J. An Introduction to Long-Memory Time Series Models and Fractional Differencing [Текст] / C. W. J. Granger, R. Joyeux // Journal of Time Series Analysis. — 1980. — N 1(1). — P. 15–29.
2. Седов А. В. Моделирование объектов с дискретно-распределенными параметрами: декомпозиционный подход [Текст] / А. В. Седов. — М.: Наука, 2010. — 438 с.
3. Автоматизированная система анализа и оперативного прогнозирования процессов потребления целевых продуктов в жилищно-коммунальном хозяйстве [Текст]: Международный конкурс инновационных проектов «Харьковские инициативы». — Харьков, 2010.
4. Модель авторегрессии — спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщенной авторегрессионной условной гетероскедастичностью для моделирования, фильтрации, прогнозирования и управления процессами в современных системах автоматизации [Текст]: труды Международной научно-практической конференции «Передовые информационные технологии, средства и системы автоматизации и их внедрение на российских предприятиях» АИТА-2011., 4–8 апреля 2011, г. Москва, 2011. — 1048 с.
5. Евдокимов А. Г. Оперативное управление потокораспределением в инженерных сетях [Текст] / А. Г. Евдокимов, А. Д. Тевяшев. — Х.: Вища школа, 1980. — 144 с.
6. Голяндина Н. Э. Метод «Гусеница»-SSA: прогноз временных рядов [Текст]: учеб. пособие / Н. Э. Голяндина. — СПб., 2004. — 52 с.

7. Перцовский О. Е. Моделирование валютных рынков на основе процессов с длинной памятью [Текст] / Н. Э. Перцовский : Препринт WP2/2004/03 — М. : ГУ ВШЭ, 2003. — 52 с.
8. Bollerslev T. Generalized autoregressive conditional heteroscedasticity / T. Bollerslev // *Journal of econometrics*. — 1986. — V. 31. — PP. 307–327.
9. Тевяшев А. Д. Системный анализ и управление большими системами энергетики [Текст] / А. Д. Тевяшев. — X. : 2009. — 507 с.
10. Juang J.-N. and Phan M. Q. Deadbeat Predictive Controllers [Текст] / J.-N. Juang, M. Q. Phan // *AIAA 35th Aerospace Sciences Meeting*. — AIAA paper number 97-0455. — Reno. — January 1997.
11. Цифровое моделирование радиосигналов комбинированными нелинейными моделями, основанными на моделях метода «Гусеница»-SSA и сезонной АРССЭ [Текст] : труды 13-й Международной конференции «Цифровая обработка сигналов и ее применение — DSPA-2011», 30 марта — 2 апреля 2011, г. Москва. — С. 165–168.
12. Модель VARSIMAX. Синтез финитного апериодического регулятора [Текст] : сборник статей одиннадцатой Международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности», 27–29 апреля 2011 г., Санкт-Петербург, Россия / под ред. А. П. Кудинова. — СПб. : Изд-во Политехн. ун-та, 2011. — С. 457–464.
13. Системный подход к автоматическому построению комбинированных математических моделей, основанных на совместном использовании моделей сезонной АРССЭ и метода «Гусеница»-SSA с нелинейным усложнением передаточной функции, для моделирования и прогнозирования широкого класса процессов, описанных временными рядами. Сезонная модель нелинейной авторегрессии — спектрально проинтегрированного скользящего среднего (сезонная нелинейная АРСПССЭ модель, seasonal NARSIMAX model) [Текст] : сборник статей десятой Международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности», 09–11 декабря 2010 г. Санкт-Петербург, Россия / под ред. А. П. Кудинова. — СПб. : Изд-во Политехн. ун-та, 2010. — С. 361–369.
14. Методика построения комбинированных математических моделей для описания и прогнозирования широкого класса физиологических и психофизиологических процессов [Текст] : сборник трудов первой Международной научно-практической конференции «Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине», 23–26 апреля 2010 г., Санкт-Петербург, Россия / под ред. А. П. Кудинова, Б. В. Крылова. — СПб. : Изд-во Политехн. ун-та, 2010. — С. 457–464.
15. Бокс, Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление [Текст] / Дж. Бокс, Г. Дженкинс ; пер. с англ. — М. : Мир, 1974. — Вып. II. — 197 с.
16. Ahid, Ahmed. Fast Orthogonal Search For Training Radial Basis Function Neural Networks / Ahmed Ahid // *Abstract of a Thesis Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science (in Electrican Engineering)*. August, 1994. — 69 p.
17. Метод «Гусеница»-SSA — АРПСС — СПОАРУГ и модель АРСПСС — СПОАРУГ для анализа и прогнозирования финансово-экономических временных рядов [Текст] : сборник трудов второй Международной научно-методической конференции «Математические методы, модели и информационные технологии в экономике», 4–6 мая 2011 г., Черновцы. — С. 306–308.
18. Тевяшев А. Д. Сравнительный анализ методов прогнозирования процессов потребления природного газа / А. Д. Тевяшев, В. Н. Щелкалин // *Восточно-европейский журнал передовых технологий* [Текст]. — 2009. — № 42. — С. 8–16.