

УДК 519.61+519.876.5:004

128

Е.А. Якимов, О.М. Демиденко, А.И. Якимов

## СИНГУЛЯРНЫЙ СПЕКТРАЛЬНЫЙ АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДАННЫХ НА ЭТАПЕ ЭКСПЛУАТАЦИИ ИМИТАЦИОННОЙ МОДЕЛИ

Представлены методика и результаты исследования погрешности восстановления аддитивных составляющих временного ряда методом сингулярного спектрального анализа. Даны рекомендации по практическому применению сингулярного спектрального анализа в имитационном моделировании на этапе эксплуатации имитационных моделей.

Ключевые слова: сингулярный спектральный анализ, имитационная модель, комплексная информационная система, временной ряд.

В настоящее время актуальным направлением совершенствования деятельности предприятия является использование имитационной модели (ИМ) в контуре управления. Построение имитационной модели реализуется последовательным выполнением следующих этапов [1]:

Этап 1. Разработка содержательного описания исследуемой системы и объектов в системе по результатам работы с заказчиком и экспертами в соответствии с целью моделирования.

Этап 2. На основе содержательного описания разрабатывается концептуальная модель с использованием, например, методологии функционального моделирования в нотации IDEF0.

Этап 3. Разработка формальной и построение имитационной модели системы. При этом IDEF0 используется не только для формализации системы, но и для изучения и анализа выполняемых ею функций, для описания имитационной модели, проработки вопросов синхронизации процессов и их взаимодействия с управляющей программой моделирования, задания начальных условий, организации сбора статистики и окончания имитации.

Этап 4. Программирование и отладка имитационной модели на основе концепции объектно-ориентированного программирования, что позволяет создавать библиотеку моделей объектов системы, расширять их функциональные возможности.

Этап 5. Испытание и исследование модели в составе комплексной информационной системы (КИС) предприятия, что требует разработки программного обеспечения для интеграции ИМ с КИС в реальных условиях производства для подготовки актуальных исходных данных, проведения экспериментов при решении поставленных задач.

Этап 6. Эксплуатация имитационной модели. Предполагает длительное использование ИМ в контуре управления промышленным предприятием. При этом на основе ИМ формируются управляющие решения, реализация которых приводит к изменению параметров и структуры системы, что, в свою очередь, требует доопределения ИМ и обеспечения ее актуальными исходными данными.

Эксплуатация имитационных моделей в производственных условиях затруднена отсутствием методов обработки накопленных данных и подготовки их для использования в моделях [2]. Эффективная подготовка входных данных для имитационного моделирования становится возможной при использовании данных, накопленных в информационных системах организаций. Для этого, например, разработан программный инструментарий, который позволяет подключаться к базе данных и извлекать их, представляя в CMSD – формате (CMSD – Core Manufacturing Simulation Data), пригодном для дискретно-событийных моделей [3].

В имитационном моделировании в основном используются статистические методы оценки последовательностей данных. В работе [4] представлен анализ программных средств для определения закона распределения случайных величин, используемых для ввода в имитационную модель (например, программа StatFit имеет 32 различных распределения, BestFit имеет 37 распределений вероятности, ExpertFit [5] имеет 40 распределений). Для наиболее точного определения типа распределения данных используются обобщенные законы распределения Пирсона, Джонсона и др. В последнее время все более часто применяются методы извлечения информации, основанные на искусственных нейронных сетях с использованием нечетких множеств и нечеткой логики, кластерном анализе и др. В то же время отсутствуют сведения о применении в имитационном моделировании сингулярного спектрального анализа (SSA – Singular Spectrum Analysis) последовательностей числовых данных, который позволяет получить дополнительную информацию о структуре данных разложением исходной последовательности на трендовую, периодическую и шумовую составляющие [6].

**Содержательное описание SSA.** Данные в комплексной информационной системе, представленные вещественными числами и определенные последовательно во времени, именуется временной последовательностью данных. В общем случае такую последовательность числовых данных, обычно упорядоченных во времени, называют временным рядом. При исследовании SSA трендом полагают медленно меняющуюся составляющую, описывающую глобальное поведение ряда. Под шумовой составляющей ряда понимают реализацию случайной последовательности чисел. Сингулярный спектральный анализ временных рядов включает следующие этапы: вложение, сингулярное разложение, группировку, диагональное усреднение. Чтобы выделить аддитивную составляющую ряда  $G$  длины  $n$  ( $G = G_T + G_H + G_N$ , причем  $G_T$ ,  $G_H$  и  $G_N$  – соответственно трендовая, гармоническая и шумовая составляющие), SSA-методом по ряду  $G$  строится траекторная матрица  $A$  заданного размера  $L \times K$ ,  $1 < L < n$ ,  $K = n - L + 1$  ( $L$  называется длиной окна), которая по правилам построения является ганкелевой. Для матрицы  $S = A \cdot A^T$  вычисляются решением характеристического уравнения  $S - \mu \cdot E = 0$  собственные числа  $\{\mu_k\}_{k=1}^L$ , а также левые  $\{U_k\}_{k=1}^L$  и правые ортонормированные собственные векторы  $\{V_k\}_{k=1}^L$ . Матрица  $A$  сингулярным разложением представляется в виде  $A = \sum_k A_k$ ,  $A_k = \sqrt{\mu_k} \cdot U_k \cdot V_k^T$ . Обозначим корень собственного числа матрицы  $S$  через  $\lambda_k = \sqrt{\mu_k}$ . Наборы  $\langle \lambda_k, U_k, V_k \rangle$  группируются для получения трендовой, гармонической и шумовой составляющих. Соответственно определяются матрицы  $A_T = \sum_{k \in T} A_k$ ,  $A_H = \sum_{k \in H} A_k$ ,  $A_N = \sum_{k \in N} A_k$ . После диагонального усреднения матриц  $A_T$ ,  $A_H$ ,  $A_N$  получают восстановленные аддитивные составляющие  $\hat{G}_T$ ,  $\hat{G}_H$ ,  $\hat{G}_N$ . Если  $G_T$ ,  $G_H$  и  $G_N$  заданы известными функциями соответственно  $F_T$ ,  $F_H$  и известен закон распределения  $G_N$ , то можно оценить показатели ошибки восстановления  $\Delta F_T = F_T - \hat{G}_T$ ,  $\Delta F_H = F_H - \hat{G}_H$  и тип распределения восстановленной шумовой составляющей  $\hat{G}_N$ . Показатели погрешности восстановления определяются характеристиками положения (математическое ожидание, оцениваемое средним (mean); медиана (med)) и характеристиками рассеяния (среднеквадратическое отклонение, оцениваемое стандартным отклонением (s); максимум (max); минимум (min), диапазон (range)) [7].

**Методика исследования.** Исследование сингулярного спектрального анализа временных рядов с комплексным применением информационных технологий основано на использовании табличного процессора MS Excel, математического пакета Mathcad, пакета статистической обработки данных Statistica и отличается применением сдвиговых процедур при формировании ганкелевой матрицы на этапе вложения и при

восстановлении составляющих исходного временного ряда на этапе диагонального усреднения [8]. На этапе группировки наборов  $\langle \lambda_k, U_k, V_k \rangle$  в MS Excel используется лепестковая диаграмма, которая является аналогом графика в полярной системе координат, отображая распределение значений относительно начала координат. По особенностям представления сингулярных векторов на лепестковой диаграмме принимается решение о принадлежности их одной группе (рис. 1 и 2).

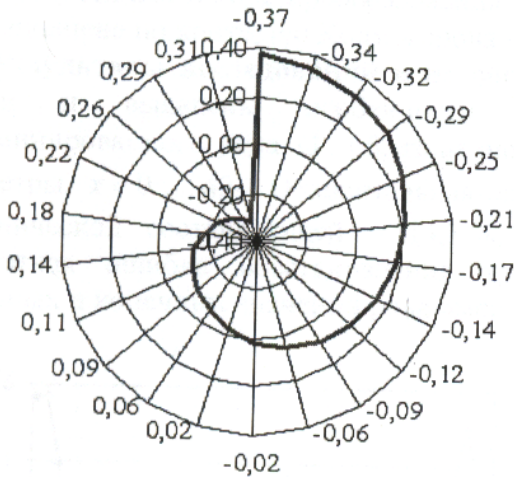


Рис. 1. Лепестковая диаграмма вектора тренда

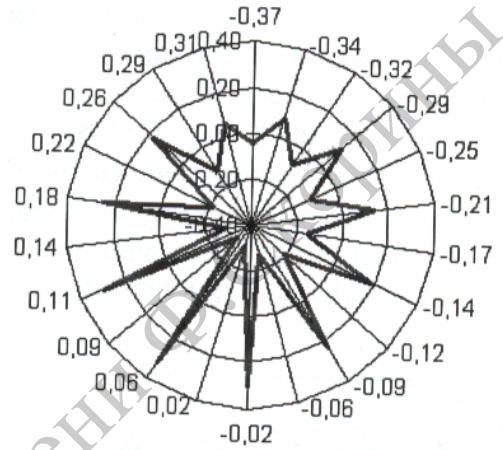


Рис. 2. Лепестковая диаграмма вектора шума

Исследования проведены для детерминированных числовых последовательностей и последовательностей, содержащих шумовую составляющую.

Для оценивания результатов исследования каждой из составляющих временного ряда используется сумма модулей значений элементов временного ряда, которая именуется  $\Phi$ -оценкой соответственно для трендовой, гармонической и шумовой составляющих:

$$\Phi_M = \sum_{i=1}^n |g_{Mi}|, \quad M = T, H, N,$$

где  $g_{Mi}$  – значение  $i$ -го элемента  $M$ -й составляющей временного ряда;  $n$  – длина временного ряда ( $n = 43$ );  $T, H, N$  – соответственно трендовая, гармоническая и шумовая составляющие.

При анализе временных рядов и сравнении результатов SSA-преобразований используются относительные  $\phi$ -оценки, определяемые соответственно для трендовой, гармонической и шумовой составляющих:

$$\phi_M = \Phi_M / \min(\Phi_T, \Phi_H, \Phi_N) \quad | \quad M = T, H, N.$$

Принятые  $\phi$ -оценки характеризуют соотношение между значениями составляющих временного ряда. Аналогично  $\Phi$ -оценкам и относительным  $\phi$ -оценкам составляющих временного ряда приняты оценки для сингулярной последовательности собственных чисел:

$$\Phi_{\lambda M} = \sum_{i=1}^k \lambda_i, \quad | \quad M = T, H, N;$$

$$\phi_{\lambda M} = \Phi_{\lambda M} / \min(\Phi_{\lambda T}, \Phi_{\lambda H}, \Phi_{\lambda N}) \quad | \quad M = T, H, N,$$

где  $k$  – количество собственных чисел.

В практике имитационного моделирования случайные последовательности определяются типами непрерывных распределений, относящимися к различным классам. При нормальном распределении случайная величина принимает значения на положительной и отрицательной полуосях, экспоненциальное распределение относится к классу распределений с возможными значениями на положительной полуоси, равномерное – к классу распределений с возможными значениями на ограниченном интервале. Проверка соответствия выполнена по критерию Колмогорова – Смирнова ( $p = 0,01$ ).

**Результаты исследований и их обсуждение.** Проведены исследования, получены оценки и выявлены особенности погрешности восстановления временного детерминированного ряда  $G$ , образованного функциями  $F(x) = ax + b \sin(x) + c$ ,  $a, b, c$  – параметры,  $x = 0, \dots, 42$ . Относительная ошибка восстановления тренда составляет 2,3%, гармонической составляющей – 14,3%. При анализе результатов обнаружено, что абсолютная ошибка восстановления тренда и гармонической составляющей для начальных и конечных значений ряда имеет наибольшие значения (рис. 3 и 4).

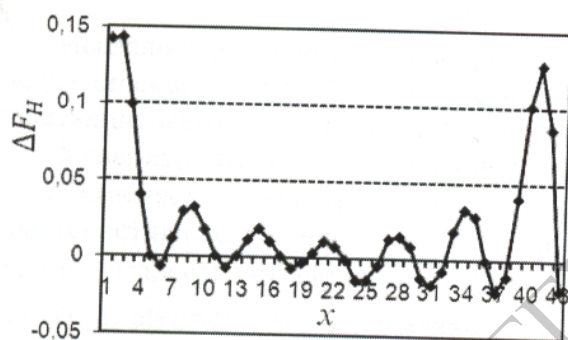


Рис. 3. Распределение абсолютной ошибки восстановления гармоники

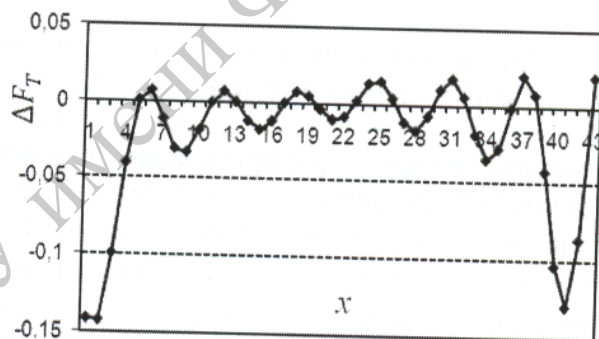


Рис. 4. Распределение абсолютной ошибки восстановления тренда

При исследовании методом сингулярного спектрального анализа временных рядов показано, что шумовая составляющая восстановленного ряда для равномерного, экспоненциального и нормального распределений смещается в сторону нуля на величину математического ожидания шумовой составляющей исходного ряда с погрешностью до 6% и возрастает до 30% при росте дисперсии шума исходного ряда до 25 раз. Стандартное отклонение шумовой составляющей восстановленного ряда отличается от исходного на 7–9% во всех случаях.

Показано, что восстановление гармонической составляющей затруднено при наличии значительной шумовой составляющей. Например, в исследованиях рассматривается временной ряд с соотношением  $\phi_{\lambda T}$ -оценок  $\phi_{\lambda T} : \phi_{\lambda H} : \phi_{\lambda N} = 28,6 : 1,0 : 10,5$ , т. е. шум на порядок превышает гармоническую составляющую. При этом восстанавливается лишь количество периодов, максимальная абсолютная ошибка превышает амплитуду гармонической составляющей более чем в два раза.

Для проверки соответствия восстановленных случайных величин  $\hat{G}_N$  теоретическому распределению используется критерий Колмогорова – Смирнова. Критическое значение  $\Delta_p$  для наибольшего отклонения эмпирического распределения от теоретического при  $p = 0,01$  и  $n = 43$  равно 0,24332. Поскольку наблюдаемое значение во всех случаях меньше критического, гипотеза  $H_0$  об известном распределении восстановленных случайных величин принимается.

Важным результатом исследования является вывод о независимости качества восстановления составляющих временного ряда от величины постоянной составляющей ис-

ходного временного ряда. Однако следует заметить, что при нулевой постоянной составляющей ( $G_{TC} = 0$ ) не удастся восстановить динамическую составляющую тренда исходного временного ряда (рис. 5 и 6).

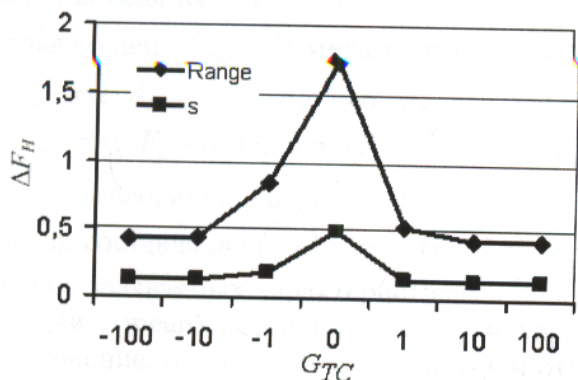


Рис. 5. Показатели ошибки  $\Delta F_H(x)$  восстановления  $F_H(x)$

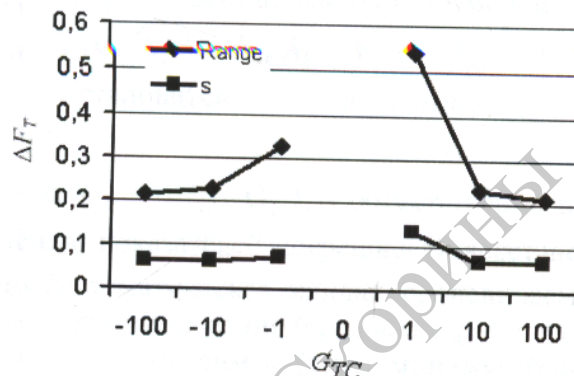


Рис. 6. Показатели ошибки  $\Delta F_T(x)$  восстановления  $F_T(x)$

Исследовано влияние процедуры сжатия/растяжения (пропорциональное уменьшение/увеличение) значений элементов временного ряда на качество восстановления его составляющих методом сингулярного спектрального анализа.

**Утверждение.** Изменение элементов временного ряда  $G$  пропорционально величине  $\alpha \in R$  изменяет показатели (характеристики положения: среднее (mean), медиана (med); характеристики рассеяния: стандартное отклонение (s), максимум (max), минимум (min), диапазон (range)) погрешности восстановления аддитивных трендовой  $G_T$ , гармонической  $G_H$  и шумовой  $G_N$  составляющих ( $G = G_T + G_H + G_N$ ) методом сингулярного спектрального анализа также пропорционально величине  $\alpha$ .

**Доказательство.** Пусть временной ряд  $G = G_T + G_H + G_N$ , причем  $G_T$ ,  $G_H$  и  $G_N$  заданы известными функциями соответственно  $F_T$ ,  $F_H$  и известен закон распределения  $G_N$ . Методом сингулярного спектрального анализа по ряду  $G$  построена ганкелева матрица  $A$  заданного размера  $L \times K$ . Для матрицы  $S = A \cdot A^T$  вычисляются решением характеристического уравнения  $S - \mu \cdot E = 0$  собственные числа  $\{\mu_k\}_{k=1}^L$ , а также левые  $\{U_k\}_{k=1}^L$  и правые векторы  $\{V_k\}_{k=1}^L$  [6], формируя разложение  $A = \sum_k A_k$ ,  $A_k = \sqrt{\mu_k} \cdot U_k \cdot V_k^T$ . Определены матрицы  $A_T = \sum_{k \in T} A_k$ ,  $A_H = \sum_{k \in H} A_k$ ,  $A_N = \sum_{k \in N} A_k$ . После диагонального усреднения матриц  $A_T$ ,  $A_H$ ,  $A_N$  получены восстановленные аддитивные составляющие  $\hat{G}_T$ ,  $\hat{G}_H$ ,  $\hat{G}_N$ , что позволяет оценить показатели ошибки восстановления  $\Delta F_T = F_T - \hat{G}_T$ ,  $\Delta F_H = F_H - \hat{G}_H$  и восстановленной шумовой составляющей  $\hat{G}_N$ .

Пусть далее  $G_\alpha = \alpha \cdot G = \alpha(G_T + G_H + G_N)$ , тогда  $A_\alpha = \alpha \cdot A$  и  $S_\alpha = A_\alpha \cdot A_\alpha^T = (\alpha \cdot A) \cdot (\alpha \cdot A)^T = \alpha \cdot A \cdot \alpha \cdot A^T = \alpha^2 \cdot A \cdot A^T$ , или  $S_\alpha = \alpha^2 \cdot S$ . Решением характеристического уравнения  $S_\alpha - \mu_\alpha \cdot E = 0$  находят собственные числа  $\{\mu_{\alpha k}\}_{k=1}^L$ , причем  $S_\alpha = \alpha^2 \cdot S$ , и тогда  $\alpha^2 \cdot S - \mu_\alpha \cdot E = 0 \rightarrow \mu_\alpha = \alpha^2 \cdot \mu$ . При этом определяют левые  $\{U_k\}_{k=1}^L$  и правые векторы  $\{V_k\}_{k=1}^L$ , формируя разложение  $A_\alpha = \sum_k A_{\alpha k}$ ,  $A_{\alpha k} = \sqrt{\alpha^2 \mu_k} \cdot U_k \cdot V_k^T = \alpha \cdot \sqrt{\mu_k} \cdot U_k \cdot V_k^T$ . Определены матрицы  $A_{\alpha T} = \sum_{k \in T} A_{\alpha k}$ ,  $A_{\alpha H} = \sum_{k \in H} A_{\alpha k}$ ,  $A_{\alpha N} = \sum_{k \in N} A_{\alpha k}$ .

После диагонального усреднения матриц  $A_{\alpha T}$ ,  $A_{\alpha H}$ ,  $A_{\alpha N}$  получены восстановленные аддитивные составляющие  $\hat{G}_{\alpha T}$ ,  $\hat{G}_{\alpha H}$ ,  $\hat{G}_{\alpha N}$ , что позволяет оценить показатели ошибки восстановления  $\Delta F_{\alpha T} = F_{\alpha T} - \hat{G}_{\alpha T}$ ,  $\Delta F_{\alpha H} = F_{\alpha H} - \hat{G}_{\alpha H}$  и восстановленной шумовой составляющей  $\hat{G}_{\alpha N}$ . Заметим, что  $A_{\alpha} = \sum_k A_{\alpha k} = \sum_k \alpha \cdot A_k = \alpha \cdot \sum_k A_k$ , т. е.  $A_{\alpha T} = \alpha \cdot A_T$ ,  $A_{\alpha H} = \alpha \cdot A_H$ ,  $A_{\alpha N} = \alpha \cdot A_N$ . Тогда становится очевидным, что  $\Delta F_{\alpha T} = \alpha \cdot F_T - \alpha \cdot \hat{G}_T = \alpha \cdot (F_T - \hat{G}_T) = \alpha \cdot \Delta F_T$ .

Аналогично  $\Delta F_{\alpha H} = \alpha \cdot F_H - \alpha \cdot \hat{G}_H = \alpha \cdot F_H - \alpha \cdot \hat{G}_H = \alpha (F_H - \hat{G}_H)$ . Соответственно шумовая составляющая  $\hat{G}_{\alpha N} = \alpha \cdot \hat{G}_N$ . По определению показателей погрешности восстановления, задаваемых характеристиками положения (математическое ожидание, оцениваемое средним (mean); медиана (med) [8]) и характеристиками рассеяния (среднеквадратическое отклонение, оцениваемое стандартным отклонением (s); максимум (max); минимум (min); диапазон (range) [8]), доказываемое исходное утверждение.

Доказательство закончено.

Таким образом, исследован нормированный временной ряд, определяемый по формуле  $G_n = \beta [G - \text{Mean}(G)] / \alpha + \gamma$ . Здесь  $\text{Mean}(G)$  – среднее элементов  $x_i$ ,  $i = 0, \dots, n-1$ , временного ряда  $G$ ;  $1/\alpha$  – коэффициент сжатия/растяжения (параметр  $\alpha$  определяется наибольшим по модулю значением элемента ряда  $G - \text{Mean}(G)$ );  $\beta$  – нормирующий параметр масштаба, определяемый верхней границей исследуемых значений элементов нормированного временного ряда (как правило,  $\beta = 1$  или  $\beta = 10$  (определяется исследователем));  $\gamma$  – параметр сдвига элементов временного ряда в область положительных вещественных чисел,  $\gamma \geq \beta$  [9].

Полученные результаты в области практического применения метода SSA ориентированы для применения в различных предметных областях, где естественно предположить существование тренда и/или периодическое поведение, и направлены на выявление особенностей в своей предметной области. Однако прежде всего результаты исследований направлены на повышение эффективности эксплуатации имитационных моделей на функционирующих предприятиях с развитой комплексной информационной системой, в которой накапливаются данные о разных сторонах функционирования предприятия в виде вещественных чисел. При этом в первую очередь полученные результаты ориентированы на имитационные модели случайных процессов. Эффективная подготовка входных данных для имитационного моделирования, построение моделей данных, подчиняющихся определенным зависимостям, повышает эффективность эксплуатации имитационных моделей. Поэтому в работе основное внимание уделено оценке погрешности восстановления трендовой, гармонической и шумовой составляющих SSA-методом.

Следует отметить, например, что проблема управления системами централизованного теплоснабжения охватывает широкий комплекс взаимосвязанных вопросов, касающихся технической политики в области развития систем, их оптимального проектирования и управления режимами работы в процессе эксплуатации. В имитационной модели системы теплоснабжения температура наружного воздуха (ТНВ) моделируется в среде имитационного моделирования PowerSim Studio. Модель ТНВ включает трендовую, гармоническую и шумовую составляющие. На примере работы индивидуального теплового пункта, оснащенного системой автоматического регулирования, показана эффективность капитальных вложений [10].

СПИСОК ЛИТЕРАТУРЫ

1. Якимов, А. И. Технология имитационного моделирования систем управления промышленных предприятий : монография / А. И. Якимов. – Могилев: Белорус.-Рос. ун-т, 2010. – 304 с.
2. Johansson, B. Information structure to support discrete event simulation in manufacturing systems / B. Johansson, J. Johansson, A. Kinnander // Proceedings of the 2003 Winter Simulation Conference, ed. S. Chick, P. J. Sanchez, D. Ferrin and D. J. Morrice. – New Orleans, Louisiana, 2003. – P. 1290–1295.
3. Bengtsson, N. Input Data Management Methodology for Discrete Event Simulation / N. Bengtsson, G. Shao, B. Johansson, T. Lee, S. Leong, A. Skoogh, C. McLean // Proceedings of 2009 Winter Simulation Conference. – 2009. – P. 1335–1344. – <http://www.informs-sim.org/wsc09papers/125.pdf>.
4. Biller, B. Introduction to Simulation Input Modeling / B. Biller, C. Gunes // Proceedings of 2010 Winter Simulation Conference. – 2010. – P. 49–58. – <http://www.informs-sim.org/wsc10papers/006.pdf>.
5. Law, A. M. ExpertFit Version 7 / A. M. Law // User's Guide.pdf. – 2006. – 144 p. – <http://www.averill-law.com/>.
6. Golyandina, N. Analysis of Time Series Structure: SSA and Related Techniques / N. Golyandina, V. Nekrutkin, A. Zhigljavsky. – Boca Raton: Chapman & Hall/CRC, 2001. – 310 p.
7. Якимов, Е. А. Исследование SSA-метода на основе комплексного применения информационных технологий / Е. А. Якимов // Доклады БГУИР. – 2010. – № 2(48). – С. 77–83.
8. Крамер, Г. Математические методы статистики : [пер. с англ.] / Г. Крамер. – 2-е изд. – М.: Мир, 1975. – 648 с.
9. Якимов, Е. А. О приведении последовательностей данных к нормированному виду для преобразования методом сингулярного спектрального анализа / Е. А. Якимов, О. М. Демиденко, Д. М. Албкеират, А. А. Ковалевич // Изв. Гомел. гос. ун-та им. Ф. Скорины. – 2011. – № 6(69). – С. 204–214.
10. Якимов, Е. А. Исследование системы теплоснабжения на основе имитационной модели / Е. А. Якимов // Изв. Гомел. гос. ун-та им. Ф. Скорины. – 2010. – № 5(62). – С. 94–99.

Материал поступил в редколлегию 04.12.12.