

УДК 004.8

Е.А. Якимов, В.Г. Замураев, А.И. Якимов

О ПРЕОБРАЗОВАНИИ МЕТОДОМ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДАННЫХ С РАВНОМЕРНЫМ ШУМОМ*

Рассматриваются методика и результаты исследования последовательностей данных с равномерным шумом путем преобразования методом сингулярного спектрального анализа (SSA-методом). Показываются особенности восстановления последовательностей данных, включающих линейный тренд, гармоническую составляющую и равномерный шум на основе сингулярного спектрального анализа с применением табличного процессора MS Excel, математического пакета Mathcad, статистического пакета Statistica.

Введение

Закон распределения случайной величины, представленный функцией распределения вероятностей, плотностью распределения и т. п., обязательно содержит хотя бы один параметр. Параметры закона распределения случайной величины принято делить на три вида: параметры положения μ , масштаба α , формы θ . Распределения вероятностей случайных величин X и Y называют однотипными, если существуют постоянные $a > 0$ и b , такие, что распределения случайных величин Y и $aX + b$ совпадают [1].

Пусть $X(0, 1)$ обозначает случайную величину с параметром положения μ , равным нулю, и параметром масштаба α , равным единице, а $X(\mu, \alpha)$ – случайную величину того же самого типа с параметром положения μ и параметром масштаба α . Тогда справедливо следующее соотношение:

$$X(\mu, \alpha) \sim \mu + \alpha X(0, 1). \quad (1)$$

Запись $X \sim Y$ обозначает, что случайные величины X и Y имеют одно и то же распределение, т. е. имеют одинаковые функции распределения вероятностей, плотности распределения и т. п.

Известные законы распределения случайных величин классифицируют на дискретные и непрерывные распределения, из которых отдельно рассматривают распределения с возможными значениями на положительной полуоси, распределения с возможными значениями на ограниченном интервале и распределения, используемые в математической статистике.

Среди дискретных распределений наибольшее применение получили дискретное равномерное распределение, распределение Пуассона, распределение Бернулли, биномиальное распределение, геометрическое распределение, гипергеометрическое распределение, отрицательное гипергеометрическое распределение, распределение Бореля – Таннера и др.

Для моделирования непрерывных распределений часто применяют нормальное распределение (распределение Гаусса – Лапласа), двустороннее показательное распределение, распределение Коши, распределение экстремального значения, распределение максимального значения, двойное показательное распределение и др.

К непрерывным распределениям с возможными значениями на положительной полуоси относятся показательное (экспоненциальное распределение), классическое (двухпараметрическое) γ -распределение, смещенное (трехпараметрическое) γ -распределение, распределение Эрланга m -го порядка, классическое (двухпараметрическое) распределение Вейбулла – Гнеденко (распределение минимального значения типа III), распределение Рэлея, распределение Максвелла, распределение Накагами, β -распределение второго рода, логарифмическое нормальное (логнормальное) распределение, распределение модуля нормальной случайной величины (от-

*Работа выполнена при финансовой поддержке гранта Ф09М-171 БРФФИ.

раженное нормальное распределение), усеченное нормальное распределение (одностороннее усечение).

К распределениям с возможными значениями на ограниченном интервале относятся равномерное (прямоугольное) распределение, классическое β -распределение, обобщенное β -распределение, параболическое распределение, распределение арксинуса, усеченное нормальное распределение (двустороннее усечение).

К распределениям, используемым в математической статистике, относятся χ^2 -распределение Пирсона, χ -распределение Пирсона, t -распределение Стьюдента, F -распределение Фишера – Снедекора (распределение дисперсионного отношения), z -распределение Фишера.

В практике имитационного моделирования исследованию входных последовательностей данных должно быть уделено значительное внимание [2]. Это подтверждается и анализом публикаций по материалам работы Winter Simulation Conference. Например, Л. Лимис выделяет отдельным этапом моделирование исходных данных [3], А. Гупта и Е. Парзен предлагают статистические методы моделирования входных данных. При этом выдвигается гипотеза о характере распределения, выбираются возможные модели, оцениваются параметры и проводится анализ наиболее подходящих из них [4]. Е.К. Лада проводит анализ методов моделирования входных данных, обращая особое внимание на β -распределение и распределение Джонсона, а также на непараметрические методы исследования [5]. В то же время практика требует дальнейших исследований входных последовательностей данных.

1. Описание SSA-метода преобразования последовательности данных

Данные в компьютерной информационной системе, представленные вещественными числами и определенные последовательно во времени, именуют временной последовательностью данных. В общем случае такую последовательность данных, обычно упорядоченных во времени, называют временным рядом. Поэтому в дальнейшем будем использовать понятие временного ряда, и для извлечения информации из последовательностей данных рассмотрим SSA-метод, который применяется для анализа временных рядов и может быть использован на каждом из этапов эксплуатации имитационной модели (ИМ). Он позволяет выделить составляющие временного ряда, которые описывают тренд и гармонические колебания, и те составляющие временного ряда, которые относят к «шуму». При этом метод не требует стационарности временного ряда, знания модели тренда, а также сведений о наличии периодических составляющих и их периодах [6]. Также с помощью данного метода можно определить модель тренда и использовать это знание для дальнейшей обработки временного ряда с уже известной моделью тренда, что важно, например, при автоматизации определения длительности переходного процесса в ИМ.

Математической основой SSA-метода является сингулярное разложение [6]. Для успешного применения SSA-метода следует последовательно пройти несколько шагов.

Вложение. На этом шаге выбирается ширина окна L , от выбора которой зависят результаты применения SSA-метода. Из-за того что нет общих рекомендаций по выбору ширины окна, параметр L зависит от решаемой задачи и предварительной информации, известной о временном ряде. Например, для выделения тренда рекомендуется выбирать ширину окна не слишком большой. С другой стороны, для выделения гармонических колебаний рекомендуется большая ширина окна. После выбора ширины окна в соответствии с L строится траекторная матрица A ряда, которая будет являться по условию ее построения ганкелевой [7].

Сингулярное разложение. Для матрицы $S = A \cdot A^T$ находятся собственные числа λ и ортонормированные собственные векторы U . Упорядоченные по убыванию собственные числа, которые больше нуля, часто называются сингулярными числами, а соответствующие им собственные векторы – левыми сингулярными векторами U . После этого вычисляются векторы V , которые называются правыми сингулярными векторами, и находятся элементарные матрицы, на сумму которых раскладывается первоначальная траекторная матрица.

Группировка. На данном этапе элементарные матрицы группируются по принципу принадлежности к тренду, гармоническим колебаниям или к шуму. Этот этап является наиболее сложным при применении SSA-метода. Для нахождения тренда на диаграммах собственных

векторов (по оси абсцисс откладывается порядковый номер координаты собственного вектора, а по оси ординат – значение координаты собственного вектора) выделяют медленно меняющиеся векторы. Сумма элементарных матриц, соответствующих этим векторам, будет являться траекторной матрицей тренда ряда. После этого восстанавливают гармонические колебания ряда. Для отделения шума можно воспользоваться несколькими замечаниями: нерегулярное поведение сингулярных векторов может говорить о принадлежности их к набору, порожденному шумовой компонентой; также об этом может свидетельствовать медленное, практически без скачков, убывание собственных чисел с некоторого номера.

Диагональное усреднение. Если полученные сгруппированные матрицы ганкелевы, то они являются траекторными матрицами некоторого временного ряда, который может быть легко по ним восстановлен. Однако обычно сгруппированные матрицы редко получаются ганкелевыми, поэтому для восстановления временного ряда прибегают к диагональному усреднению. В соответствии с этим этапом каждый член восстановленного временного ряда будет являться средним арифметическим соответствующей ему побочной диагонали траекторной матрицы.

В результате проделанных шагов получается несколько временных рядов: один описывает тренд первоначального временного ряда, другой – гармонические колебания, а третий – шумовые составляющие.

2. Методика исследования последовательностей данных с равномерным шумом

Перечисленные выше распределения случайных величин подвергнуты исследованию через преобразование SSA-методом. Основной целью работы являлось исследование SSA-метода при анализе временного ряда с линейным трендом, гармонической сезонной компонентой и шумовой составляющей, имеющей одно из указанных распределений. При исследованиях применялся табличный процессор MS Excel, а также пакеты Mathcad и Statistica [8]. Технология исследования включала формирование ганкелевой матрицы из исходного временного ряда, определение левых и правых собственных векторов матрицы, построение групп векторов с трендовой составляющей ряда, периодикой и шумом. Восстановление исходных составляющих временного ряда выполнялось путем диагонализации соответствующих матриц.

Равномерное распределение в научной литературе иногда называют прямоугольным. Случайная величина ξ называется равномерно распределенной на заданном интервале, если ее плотность вероятности $f_{\xi}(x)$ постоянна на этом интервале и равна нулю вне его.

Основные числовые характеристики закона распределения случайной величины:

- характеристики положения: математическое ожидание, оцениваемое средним; медиана;
- характеристики рассеяния: среднее квадратическое отклонение, оцениваемое стандартным отклонением; максимум; минимум; диапазон; коэффициенты асимметрии распределения и эксцесса.

Для предварительного исследования принят временной ряд G с равномерным шумом:

$$G = Rnd(0; 1)_n, n = 43, \quad (2)$$

где $Rnd(0; 1)_n$ – функция, возвращающая n случайных чисел, равномерно распределенных в интервале $[0; 1]$;

$G =$ 0,007; 0,059; 0,948; 0,536; 0,426; 0,756; 0,288; 0,939; 0,200; 0,588; 0,586;
0,397; 0,306; 0,139; 0,906; 0,756; 0,564; 0,333; 0,471; 0,000; 0,459; 0,449; 0,290;
0,997; 0,402; 0,181; 0,640; 0,256; 0,538; 0,747; 0,022; 0,603; 0,406; 0,684; 0,965;
0,518; 0,203; 0,077; 0,509; 1,000; 0,806; 0,198; 0,893.

После преобразования временного ряда G SSA-методом получен восстановленный временной ряд \hat{G} , определяемый двумя составляющими:

$$\widehat{G} = \widehat{G}_N + \widehat{G}_T, \quad (3)$$

где \widehat{G}_N , \widehat{G}_T – соответственно шумовая (рис. 1) и трендовая составляющие. При SSA-преобразовании получены собственные числа (рис. 2), где λ_1 соответствует трендовой составляющей \widehat{G}_T , а $\lambda_i, i = 2, \dots, 43$, определяют шумовую составляющую \widehat{G}_N .

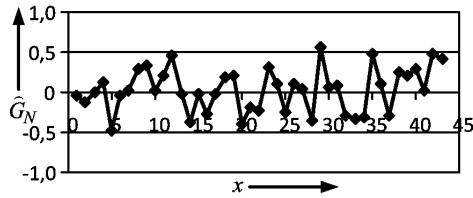


Рис. 1. Восстановленный шум временного ряда \widehat{G}

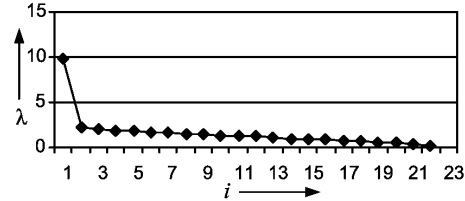


Рис. 2. Распределение собственных чисел

Проверка по критерию Колмогорова – Смирнова соответствия полученных случайных чисел теоретическому равномерному распределению показала, что шумовая составляющая \widehat{G}_N равномерно распределена на интервале $[-0,5; 0,5]$. При этом трендовая составляющая \widehat{G}_T восстановленного временного ряда \widehat{G} совпадает с математическим ожиданием исходного временного ряда G . В соответствии с (1) $\mu = 0,5$ и исходный временной ряд (2) представлен следующим образом:

$$Rnd(0; 1) \sim Rnd(-0,5; 0,5) + 0,5 \mid \alpha = 1. \quad (4)$$

Методика проведения экспериментальных исследований включает выбор исходного временного ряда G , который задан по известным функциям

$$G = F(x) = F_T(x) + F_H(x) + F_N(x); \quad (5)$$

$$G_T = F_T(x) = ax + b; \quad (6)$$

$$G_H = F_H(x) = csinx; \quad (7)$$

$$G_N = F_N(x) = Rnd(n_{inf}; n_{sup}), \quad (8)$$

где a, b, c – параметры; $Rnd(n_{inf}; n_{sup})$ – функция, возвращающая случайные равномерные числа в интервале $[n_{inf}; n_{sup}]$ (n_{inf} – нижняя граница значений случайных чисел, n_{sup} – верхняя граница значений случайных чисел); индексы T, H, N – соответственно трендовая, гармоническая и шумовая составляющие.

Для оценивания результатов исследования каждой из составляющих используется сумма модулей значений элементов временного ряда, которую будем именовать Φ -оценкой соответственно для трендовой, гармонической и шумовой составляющих:

$$\Phi_T = \sum_{i=1}^n |F_T(x_i)| = \sum_{i=1}^n |a \cdot x_i + b|; \quad (9)$$

$$\Phi_H = \sum_{i=1}^n |F_H(x_i)| = \sum_{i=1}^n |c \cdot \sin x_i|; \quad (10)$$

$$\Phi_N = \sum_{i=1}^n |F_N(x_i)| = \sum_{i=1}^n |Rnd(n_{inf}; n_{sup})|, \quad (11)$$

где n – число элементов временного ряда ($n = 43$).

Исходные данные для проведения исследований представлены в табл. 1.

Таблица 1
Исходные данные для исследования временных рядов с равномерным шумом

| $F_j(x)$ | $F_{Tj}(x)$ | $F_{Hj}(x)$ | $F_{Nj}(x)$ | Φ_{Tj} | Φ_{Hj} | Φ_{Nj} |
|----------|------------------|-------------------|--------------|-------------|-------------|-------------|
| $F_1(x)$ | $1 \cdot x + 2$ | $1 \cdot \sin x$ | $Rnd(0; 1)$ | 989 | 27,15 | 20,86 |
| $F_2(x)$ | $1 \cdot x + 2$ | $1 \cdot \sin x$ | $Rnd(0; 10)$ | 989 | 27,15 | 210,5 |
| $F_3(x)$ | $1 \cdot x + 2$ | $10 \cdot \sin x$ | $Rnd(0; 1)$ | 989 | 271,5 | 20,86 |
| $F_4(x)$ | $1 \cdot x + 2$ | $10 \cdot \sin x$ | $Rnd(0; 10)$ | 989 | 271,5 | 210,5 |
| $F_5(x)$ | $10 \cdot x + 2$ | $1 \cdot \sin x$ | $Rnd(0; 1)$ | 9030 | 27,15 | 20,86 |
| $F_6(x)$ | $10 \cdot x + 2$ | $1 \cdot \sin x$ | $Rnd(0; 10)$ | 9030 | 27,15 | 210,5 |
| $F_7(x)$ | $10 \cdot x + 2$ | $10 \cdot \sin x$ | $Rnd(0; 1)$ | 9030 | 271,5 | 20,86 |
| $F_8(x)$ | $10 \cdot x + 2$ | $10 \cdot \sin x$ | $Rnd(0; 10)$ | 9030 | 271,5 | 210,5 |

Для анализа временных рядов в практике имитационного моделирования и сравнения результатов SSA-преобразований удобно использовать относительные ф-оценки, определяемые соответственно для трендовой, гармонической и шумовой составляющих:

$$\varphi_T = \frac{\Phi_T}{\min(\Phi_T, \Phi_H, \Phi_N)}; \quad (12)$$

$$\varphi_H = \frac{\Phi_H}{\min(\Phi_T, \Phi_H, \Phi_N)}; \quad (13)$$

$$\varphi_N = \frac{\Phi_N}{\min(\Phi_T, \Phi_H, \Phi_N)}. \quad (14)$$

Принятые ф-оценки характеризуют соотношение между значениями составляющих временного ряда. Результаты расчета относительных ф-оценок представлены в табл. 2.

Таблица 2
Относительные ф-оценки исследуемых временных рядов $F(x)_j$

| Показатель | $F_1(x)$ | $F_2(x)$ | $F_3(x)$ | $F_4(x)$ | $F_5(x)$ | $F_6(x)$ | $F_7(x)$ | $F_8(x)$ |
|---------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|
| $\varphi_{Tj}, j = 1, \dots, 8$ | 47,4 | 36,4 | 47,4 | 4,7 | 432,9 | 332,6 | 432,9 | 42,9 |
| $\varphi_{Hj}, j = 1, \dots, 8$ | 1,3 | 1,0 | 13,0 | 1,3 | 1,3 | 1,0 | 13,0 | 1,3 |
| $\varphi_{Nj}, j = 1, \dots, 8$ | 1,0 | 7,8 | 1,0 | 1,0 | 1,0 | 7,8 | 1,0 | 1,0 |

3. Результаты и обсуждение

При сингулярном спектральном анализе получено распределение собственных чисел λ_i (табл. 3), являющееся одной из характеристик для оценки качества восстановления составляющих исходного временного ряда.

Значения собственных чисел в табл. 3, которым соответствуют векторы, восстанавливающие трендовую составляющую, выделены полужирным шрифтом. Собственные числа, соответствующие гармонической составляющей, выделены полужирным шрифтом с подчеркиванием. Остальные λ_i определяют шумовую составляющую.

Таблица 3

Собственные числа при сингулярном спектральном анализе временных рядов $F_j(x)$

| λ_i | $F_1(x)$ | $F_2(x)$ | $F_3(x)$ | $F_4(x)$ | $F_5(x)$ | $F_6(x)$ | $F_7(x)$ | $F_8(x)$ |
|----------------|---------------|---------------|---------------|---------------|--------------|--------------|---------------|---------------|
| λ_1 | 550,97 | 641,68 | 552,76 | 643,51 | 5059 | 5148 | 5060 | 5149 |
| λ_2 | 33,98 | 28,41 | 108,81 | 117,84 | 383,7 | 372,7 | 383,6 | 372,6 |
| λ_3 | 10,04 | 24,66 | 106,84 | 114,04 | 10,04 | 25,77 | 108,77 | 117,71 |
| λ_4 | 9,94 | 22,43 | 33,94 | 25,68 | 9,94 | 24,65 | 107,14 | 114,42 |
| λ_5 | 1,88 | 22,05 | 1,88 | 23,87 | 1,88 | 22,07 | 1,88 | 23,86 |
| λ_6 | 1,77 | 21,17 | 1,77 | 21,26 | 1,77 | 21,26 | 1,77 | 21,87 |
| λ_7 | 1,60 | 15,34 | 1,59 | 15,50 | 1,60 | 15,33 | 1,59 | 15,49 |
| λ_8 | 1,58 | 14,63 | 1,56 | 14,83 | 1,58 | 14,74 | 1,56 | 15,13 |
| λ_9 | 1,53 | 14,49 | 1,53 | 14,57 | 1,54 | 14,54 | 1,54 | 14,57 |
| λ_{10} | 1,51 | 14,41 | 1,526 | 14,47 | 1,51 | 14,42 | 1,53 | 14,47 |
| λ_{11} | 1,40 | 12,62 | 1,402 | 12,61 | 1,40 | 12,76 | 1,40 | 12,75 |
| λ_{12} | 1,25 | 11,67 | 1,253 | 11,93 | 1,25 | 11,67 | 1,25 | 11,93 |
| λ_{13} | 1,19 | 11,14 | 1,184 | 11,20 | 1,19 | 11,14 | 1,19 | 11,21 |
| λ_{14} | 1,06 | 9,70 | 1,051 | 9,87 | 1,06 | 9,69 | 1,05 | 9,86 |
| λ_{15} | 0,98 | 9,44 | 0,981 | 9,43 | 0,98 | 9,44 | 0,98 | 9,42 |
| λ_{16} | 0,89 | 8,59 | 0,898 | 9,19 | 0,89 | 8,62 | 0,90 | 9,27 |
| λ_{17} | 0,81 | 5,81 | 0,805 | 5,88 | 0,81 | 5,88 | 0,81 | 5,96 |
| λ_{18} | 0,77 | 4,70 | 0,764 | 4,78 | 0,77 | 4,65 | 0,76 | 4,72 |
| λ_{19} | 0,65 | 2,423 | 0,641 | 2,43 | 0,65 | 2,42 | 0,64 | 2,43 |
| λ_{20} | 0,53 | 2,417 | 0,521 | 2,42 | 0,53 | 2,418 | 0,52 | 2,42 |
| λ_{21} | 0,29 | 1,464 | 0,302 | 1,49 | 0,29 | 1,48 | 0,30 | 1,51 |
| λ_{22} | 0,25 | 1,456 | 0,238 | 1,43 | 0,25 | 1,42 | 0,24 | 1,39 |

Аналогично Φ -оценкам и относительным ϕ -оценкам составляющих временного ряда приняты оценки для сингулярной последовательности собственных чисел:

$$\Phi_{\lambda M} = \sum_{i=1}^k \lambda_i, | M = T, H, N; \tag{15}$$

$$\phi_{\lambda M} = \Phi_{\lambda M} / \min(\Phi_{\lambda T}, \Phi_{\lambda H}, \Phi_{\lambda N}) | M = T, H, N, \tag{16}$$

где k – количество собственных чисел.

Расчеты относительных ϕ_{λ} -оценок представлены в табл. 4.

Таблица 4

Относительные ϕ_{λ} -оценки сингулярной последовательности собственных чисел

| Показатель | $F_1(x)$ | $F_2(x)$ | $F_3(x)$ | $F_4(x)$ | $F_5(x)$ | $F_6(x)$ | $F_7(x)$ | $F_8(x)$ |
|-------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|
| $\phi_{\lambda T j}, j=1, \dots, 8$ | 29,3 | 28,6 | 29,5 | 3,0 | 272,8 | 640,5 | 273,4 | 29,3 |
| $\phi_{\lambda H j}, j=1, \dots, 8$ | 1,0 | 1,0 | 10,8 | 1,1 | 1,0 | 1,0 | 10,8 | 1,2 |
| $\phi_{\lambda N j}, j=1, \dots, 8$ | 1,0 | 10,5 | 1,0 | 1,0 | 1,0 | 26,2 | 1,0 | 1,0 |

Для определения качества восстановления составляющих временного ряда приняты следующие показатели (табл. 5):

$$\Delta F_{Tj}(x) = F_{Tj}(x) - \hat{G}_{Tj}, j = 1, \dots, 8; \quad (17)$$

$$\Delta F_{Hj}(x) = F_{Hj}(x) - \hat{G}_{Hj}, j = 1, \dots, 8, \quad (18)$$

где $F_{Tj}(x)$, $F_{Hj}(x)$ – трендовая и гармоническая составляющие исходного временного ряда; \hat{G}_{Tj} , \hat{G}_{Hj} – восстановленные трендовая и гармоническая составляющие временного ряда, $j = 1, \dots, 8$.

Восстановленная шумовая составляющая \hat{G}_{Nj} , а также $\Delta F_{Tj}(x)$ и $\Delta F_{Hj}(x)$, $j = 1, \dots, 8$, определяются как характеристиками положения: среднее (mean), медиана (med), так и характеристиками рассеяния: стандартное отклонение s , максимум (max), минимум (min), диапазон (range), коэффициенты асимметрии распределения (γ_3) и эксцесса (γ_4).

Таблица 5

Основные показатели качества восстановления составляющих исходного временного ряда SSA-методом

| Оценки | Mean | Med | Max | Min | Range | s | γ_3 | γ_4 |
|--------------------|-------|-------|-------|--------|-------|-------|------------|------------|
| $\Delta F_{T1}(x)$ | -0,50 | -0,48 | -0,36 | -0,76 | 0,40 | 0,123 | -0,49 | -1,03 |
| $\Delta F_{H1}(x)$ | 0,02 | 0,01 | 0,23 | -0,22 | 0,45 | 0,123 | -0,08 | -1,16 |
| \hat{G}_{N1} | 0 | -0,01 | 0,46 | -0,52 | 0,98 | 0,231 | -0,01 | -0,72 |
| $\Delta F_{T2}(x)$ | -5,73 | -5,44 | -3,46 | -9,29 | 5,83 | 1,747 | -0,49 | -0,97 |
| $\Delta F_{H2}(x)$ | 0,28 | 0,27 | 2,16 | -1,19 | 3,35 | 0,811 | 0,23 | -0,55 |
| \hat{G}_{N2} | -0,55 | -0,33 | 5,13 | -8,68 | 13,81 | 3,205 | -0,35 | -0,10 |
| $\Delta F_{T3}(x)$ | -0,70 | -0,54 | -0,13 | -2,12 | 1,98 | 0,497 | -1,51 | 1,54 |
| $\Delta F_{H3}(x)$ | 0,21 | 0,10 | 1,37 | -0,36 | 1,73 | 0,462 | 1,19 | 0,82 |
| \hat{G}_{N3} | 0 | 0 | 0,45 | -0,52 | 0,98 | 0,231 | -0,01 | -0,77 |
| $\Delta F_{T4}(x)$ | -5,92 | -5,52 | -3,47 | -10,32 | 6,85 | 2,023 | -0,66 | -0,69 |
| $\Delta F_{H4}(x)$ | 0,26 | 0,28 | 3,56 | -1,57 | 1,99 | 1,152 | 0,36 | -0,13 |
| \hat{G}_{N4} | -0,76 | -0,97 | 6,03 | -7,60 | 13,63 | 2,891 | -0,12 | 0,13 |
| $\Delta F_{T5}(x)$ | -0,50 | -0,48 | -0,36 | -0,75 | 0,39 | 0,122 | -0,48 | -1,11 |
| $\Delta F_{H5}(x)$ | 0,02 | 0,01 | 0,23 | -0,22 | 0,45 | 0,124 | -0,07 | -1,18 |
| \hat{G}_{N5} | 0 | -0,01 | 0,45 | -0,53 | 0,98 | 0,232 | -0,01 | -0,72 |
| $\Delta F_{T6}(x)$ | -4,97 | -4,87 | -3,86 | -6,23 | 2,37 | 0,522 | -0,69 | 0,43 |
| $\Delta F_{H6}(x)$ | 0,04 | 0,01 | 1,23 | -1,26 | 2,49 | 0,750 | -0,07 | -1,24 |
| \hat{G}_{N6} | -0,03 | -0,18 | 5,05 | -5,48 | 10,52 | 2,926 | 0,15 | -0,81 |
| $\Delta F_{T7}(x)$ | -0,67 | -0,51 | -0,10 | -1,94 | 1,84 | 0,434 | -1,54 | 1,67 |
| $\Delta F_{H7}(x)$ | 0,18 | 0,13 | 1,44 | -0,70 | 2,14 | 0,431 | 1,07 | 1,42 |
| \hat{G}_{N7} | 0 | 0 | 0,45 | -0,52 | 0,97 | 0,232 | -0,01 | -0,77 |
| $\Delta F_{T8}(x)$ | -5,13 | -4,89 | -4,36 | -7,07 | 2,71 | 0,692 | -1,25 | 1,19 |
| $\Delta F_{H8}(x)$ | 0,21 | 0,24 | 2,74 | -1,66 | 1,08 | 1,085 | 0,15 | -0,80 |
| \hat{G}_{N8} | -0,03 | -0,28 | 5,19 | -4,00 | 9,19 | 2,420 | 0,44 | -0,62 |

Для проверки соответствия восстановленных случайных величин \widehat{G}_{Nj} теоретическому равномерному распределению используется критерий Колмогорова – Смирнова. Критическое значение Δ_p для наибольшего отклонения эмпирического распределения от теоретического по табл. 6.2 [9, с. 347] при $p = 0,01$ и $n = 43$ равно $0,24332$. Если наблюдаемое значение Δ больше или равно критическому значению, нулевая гипотеза H_0 отклоняется. Поскольку наблюдаемое значение

$$\Delta = 0,071 - 0,187 | \forall \widehat{G}_{Nj}, j = 1, \dots, 8,$$

меньше критического, гипотеза H_0 о равномерном распределении восстановленных случайных величин принимается.

При исследовании выявлены следующие факты:

1. Шумовая составляющая восстанавливается со средним значением, близким к нулю. Например, для временного ряда $F_8(x)$ среднее значение восстановленной шумовой составляющей равно $-0,03$ (табл. 5), а для исходной шумовой составляющей, определяемой функцией $Rnd(0; 10)$, равно $4,89$.

2. Постоянная составляющая шума исключается при восстановлении и суммируется с трендовой составляющей исходного временного ряда.

3. Шумовая составляющая восстанавливается с погрешностью стандартного отклонения, не превышающей 32 %. Например, для временного ряда $F_8(x)$ с соотношением φ_λ -оценок

$$\varphi_{T8} : \varphi_{H8} : \varphi_{N8} = 29,3 : 1,2 : 1,0$$

стандартное отклонение для исходной составляющей равно $2,945$, а для восстановленной – $2,420$. При этом теоретическое значение стандартного отклонения для $Rnd(0, 10)$ равно $2,887$. Отклонение от теоретического значения для исходной шумовой составляющей равно $2,0 \%$, а для восстановленной – $16,2 \%$.

4. Наихудшим случаем восстановления гармонической составляющей является $F_2(x)$ с соотношением φ_λ -оценок

$$\varphi_{T2} : \varphi_{H2} : \varphi_{N2} = 28,6 : 1,0 : 10,5,$$

т. е. шум на порядок превышает гармоническую составляющую. Полностью восстанавливается лишь количество периодов (рис. 3). Максимальная ошибка превышает амплитуду гармонической составляющей в $2,16$ раза (см. табл. 5).

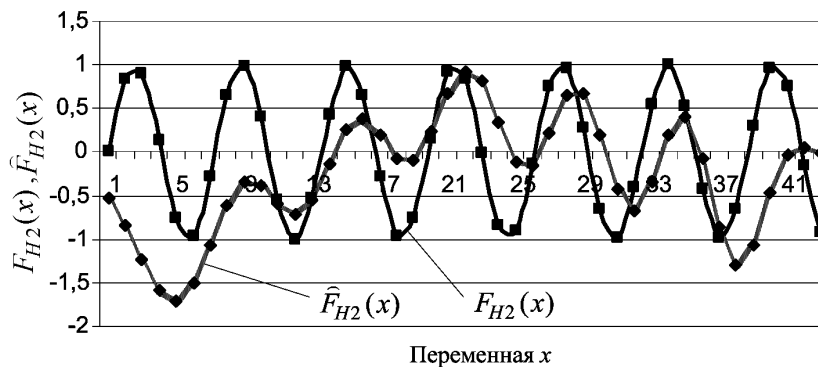


Рис. 3. Восстановление гармонической составляющей $F_{H2}(x)$

5. Ошибка восстановления трендовой составляющей распределяется в соотношении приблизительно $10 : 80 : 10$. Например, ошибка $\Delta F_{T3}(x)$ в середине восстановленного тренда меньше, чем для крайних значений тренда, в $1,8-1,94$ раза (рис. 4).

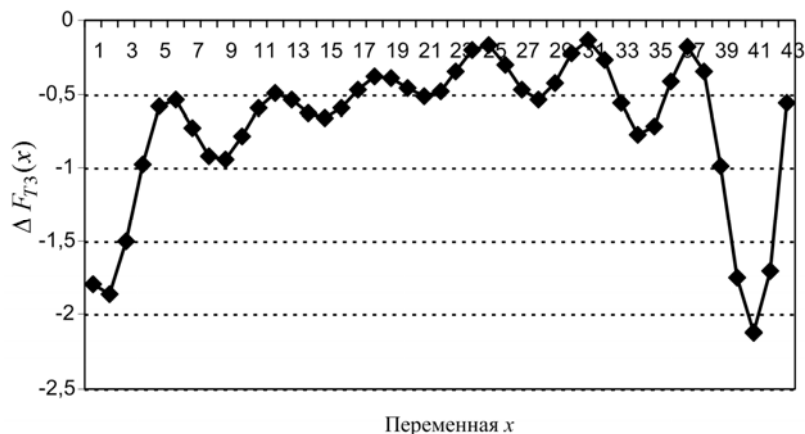


Рис. 4. Распределение абсолютной ошибки $\Delta F_{T3}(x)$ восстановления тренда $F_{T3}(x)$

6. Наихудшие результаты восстановления гармонических составляющих наблюдаются для рядов $F_2(x)$ и $F_6(x)$, характеризующихся значительной шумовой составляющей (см. табл. 2). Этот факт проявляется и для φ_λ -оценок сингулярной последовательности собственных чисел (см. табл. 4). При этом следует также обратить внимание на расположение собственных чисел, ответственных за восстановление гармонической составляющей (см. табл. 3). Эти числа располагаются среди ниспадающей части собственных чисел, определяющих шумовую составляющую. Данный факт является косвенным предупреждением о возможных значительных ошибках восстановления гармонической составляющей.

Заключение

Шумовая составляющая временного ряда восстанавливается со средним значением, близким к нулю. Постоянная составляющая равномерного шума исключается при восстановлении и суммируется с трендовой составляющей исходного временного ряда.

Наихудшим случаем восстановления гармонической составляющей является временной ряд с соотношением φ_λ -оценок $\varphi_{T2} : \varphi_{H2} : \varphi_{N2} = 28,6 : 1,0 : 10,5$, т. е. шум на порядок превышает гармоническую составляющую. При этом восстанавливается лишь количество периодов и максимальная абсолютная ошибка превышает амплитуду гармонической составляющей более чем в два раза.

Интересным фактом, ориентированным на практическое применение SSA-метода, является распределение ошибки восстановления трендовой составляющей в соотношении приблизительно 10 : 80 : 10. Например, в середине восстановленного тренда ошибка меньше, чем для крайних значений тренда, примерно в два раза.

Список литературы

1. Вадзинский, Р.Н. Справочник по вероятностным распределениям / Р.Н. Вадзинский. – СПб. : Наука, 2001. – 295 с.
2. Якимов, А.И. Технология имитационного моделирования систем управления промышленных предприятий / А.И. Якимов. – Могилев : Белорус.-Рос. ун-т, 2010. – 304 с.
3. Leemis, L. Input Modeling / L. Leemis // Proc. of 2003 Winter Simulation Conf. – New Orleans : IEEE Press, 2003. – P. 14–24.
4. Gupta, A. Input Modeling Using Quantile Statistical Methods / A. Gupta, E. Parzen // INFORMS Simulation Society[Electronic recourse]. – 2004. – Mode of access : <http://www.informs-sim.org/wsc04papers/089.pdf>. – Date of access : 23.03.2009.
5. Introduction to Modeling and Generating Probabilistic Input Processes for Simulation / E.K. Lada [et al.] // Proc. of 2005 Winter Simulation Conf. – Orlando : IEEE Press, 2005. – P. 41–55.
6. Голяндина, Н.Э. Метод «Гусеница»-SSA: анализ временных рядов : учеб. пособие / Н.Э. Голяндина. – СПб. : С.-Петербург. гос. ун-т, 2004. – 76 с.

7. Гантмахер, Ф.Р. Теория матриц / Ф.Р. Гантмахер. – 2-е изд., доп. – М. : Наука, 1966. – 576 с.
8. Якимов, Е.А. Исследование SSA-метода на основе комплексного применения информационных технологий / Е.А. Якимов // Доклады БГУИР. – 2010. – № 2 (48). – С. 77–83.
9. Большев, Л.Н. Таблицы математической статистики / Л.Н. Большев, Н.В. Смирнов. – М. : Наука, 1983. – 416 с.

Поступила 18.11.10

*Белорусско-Российский университет,
Могилев, пр. Мира, 43
e-mail: ykm@tut.by*

Y.A. Yakimau, V.G. Zamuraeu, A.I. Yakimau

**ON THE TRANSFORMATION OF SINGULAR SPECTRAL ANALYSYS
USING THE METHOD OF DATA SEQUENCES WITH UNIFORM NOISE**

A technique and results of investigation of sequences of data with uniform noise by transformation by a SSA-method are presented. Features of restoration of sequences of the data including a trend, a harmonious component and uniform noise on the basis of the Singular Spectrum Analysis with application of tabulated processor MS Excel, mathematical package Mathcad, statistical package Statistica are shown.