

ИССЛЕДОВАНИЕ МЕТОДА СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ BELSIM2#.SSA ¹

А. Н. Микулич, В. А. Долгий, В. О. Батыщиков, Е. А. Якимов

Аннотация. В статье представлены выполненные исследования, которые показывают неоднозначность восстановления трендовой, периодической и шумовой составляющей временного ряда. Полученные результаты восстановления необходимо учитывать при практическом применении метода анализа сингулярного спектра для принятия решений.

Ключевые слова: сингулярный спектральный анализ, программный модуль, нормированный временной ряд.

1. ВВЕДЕНИЕ

При исследовании временных рядов методом сингулярного спектрального анализа [1] применяется программный модуль BelSim2#.SSA [2], который предназначен для использования на действующих предприятиях с целью извлечения полезной информации из числовых данных, накопленных в комплексных информационных системах.

Программный модуль разработан на основе программной платформы *.NET Framework 4.0* в среде программирования *Microsoft Visual Studio 2008* с применением объектно-ориентированного языка программирования *C#*.

Файл с исходным временным рядом представляет собой текстовый файл, содержащий последовательность вещественных чисел. Выходной файл, создаваемый программой, представляет собой текстовый файл и графические файлы. Выходной текстовый файл хранит в себе следующую информацию: исходный временной ряд; численные значения трендовой, периодической и шумовой составляющей. Выходные графические файлы содержат графики исходного ряда; сингулярных векторов; сингулярных чисел; лепестковые диаграммы левых и правых сингулярных векторов; трендовой, периодической и шумовой составляющей.

Технические требования BelSim 2#.SSA: процессор Pentium 4x2 ГГц и выше; оперативная память 512 Мбайт и больше; пространство на жестком диске 1 Мбайт; операционная система Windows 2000/XP/Vista/7; дополнительное программное обеспечение: набор библиотек .NET Framework 4.0 Client Profile (41 Мбайт).

Особенность проведенных исследований в том, что в программном модуле BelSim2#.SSA реализовано аналитическое выражение для получения нормированного временного ряда [3].

2. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ КАЧЕСТВА ВОССТАНОВЛЕНИЯ АДДИТИВНЫХ СОСТАВЛЯЮЩИХ ВРЕМЕННОГО РЯДА

Для проведения исследований выбрана функция

$$G_i = G_H + G_T + G_N \quad (1)$$

где G_T – линейная функция, G_H – гармоническая функция, G_N – шумовая функ-

¹ Работа выполнена в порядке личной инициативы по НИРС

ция.

Ниже покажем формулы, по которым строится каждая функция. Представим гармоническую составляющую в общем случае с k периодами

$$G_H = K_H \cdot \sin(k \cdot \frac{2\pi(x-1)}{n-1}) | x = 1, \dots, n, k > 0 \quad (2)$$

где k – число периодов в ряду длины n , например, $k = 2$;

K_H – размах гармонической составляющей.

Трендовую составляющую G_T определим по формуле:

$$G_T = K_T \cdot (\frac{2x}{n-1} - 1) | x = 1, \dots, n \quad (3)$$

где K_T – размах линейной составляющей. Представим G_T в виде двух составляющих

$$G_T = G_{TC} + G_{TD},$$

где G_{TC} – постоянная составляющая,

G_{TD} – динамическая составляющая, причем полагаем $G_{TC} = 0$.

Чтобы не вносить дополнительных постоянных составляющих, G_N определим по формуле:

$$G_N = K_N \cdot \text{Rnd}(-1; 1), \quad (4)$$

где K_N – размах шумовой составляющей.

Особенность проведенных исследований в том, что в программном модуле BelSim2#.SSA реализовано аналитическое выражение для получения нормированного временного ряда, которое определяется по формуле

$$G_H = \beta \cdot [G - \text{Mean}(G)] / \alpha + \gamma, \quad (5)$$

где $\text{Mean}(G)$ – среднее элементов $x_i, i = 0, \dots, n-1$ временного ряда G ;

$1/\alpha$ – коэффициент сжатия/растяжения, при этом параметр α определяется наибольшим по модулю значением элемента ряда $G - \text{Mean}(G)$;

β – нормирующий параметр масштаба, определяемый верхней границей исследуемых значений элементов нормированного временного ряда, как правило $\beta = 1$;

γ – параметр сдвига элементов временного ряда в область положительных вещественных чисел, $\gamma \geq \beta$ [2].

Изучение свойств нормированного временного ряда позволяет независимым исследователям унифицировать результаты реальных исследований и создать базу знаний для выявления полезной информации с учетом особенностей применения метода сингулярного спектрального анализа.

Ставится задача количественной оценки качества восстановления аддитивных составляющих временного ряда. Проводится полный факторный эксперимент. Для сравнения результатов использовали оценки Φ_{TD} , Φ_H и Φ_N , определенные по формулам:

$$\Phi_{TD} = \frac{\sum_{TD} x_i^2}{n}, \quad \Phi_H = \frac{\sum_H x_i^2}{n}, \quad \Phi_N = \frac{\sum_N x_i^2}{n}, \quad (6)$$

где x_i^2 – квадрат i -ых элементов исходной и полученной составляющей ряда.

Приняты следующие порядковые номера экспериментов в зависимости от соотношения $\Phi_{TD} : \Phi_H : \Phi_N$:

- 1) $\Phi_{TD} : \Phi_H : \Phi_N = 1:1:1$;
- 2) $\Phi_{TD} : \Phi_H : \Phi_N = 1:1:10$;
- 3) $\Phi_{TD} : \Phi_H : \Phi_N = 1:10:1$;
- 4) $\Phi_{TD} : \Phi_H : \Phi_N = 1:10:10$;
- 5) $\Phi_{TD} : \Phi_H : \Phi_N = 10:1:1$;
- 6) $\Phi_{TD} : \Phi_H : \Phi_N = 10:1:10$;
- 7) $\Phi_{TD} : \Phi_H : \Phi_N = 10:10:1$;
- 8) $\Phi_{TD} : \Phi_H : \Phi_N = 10:10:10$.

Приняты следующие оценки: среднее (Mean), медиана (Med); характеристиками рассеяния: (стандартное отклонение s ; максимум (Max); минимум (Min), диапазон (Range), коэффициенты асимметрии распределения (γ_3) и эксцесса (γ_4) [4].

С целью выявления особенностей восстановления аддитивных составляющих ряда получены оценки погрешности восстановления аддитивных составляющих при длине ряда $n = 100, n = 200, n = 300$. В таблице 1 представлены оценки при $n = 100$.

Таблица 1. Оценки погрешности восстановления аддитивных составляющих при длине ряда $n = 100$

Оценки	Mean	Med	Max	Min	Range	s	γ_3	γ_4
$\Delta F_{T1}(x)$	-0,0423	-0,0508	1,12267	-1,1815	2,30415	0,75593	0,04936	-1,3535
$\Delta F_{H1}(x)$	-0,0027	0,04754	1,18676	-1,4252	2,61194	0,72501	0,01104	-1,2283
G_{N1}	0,01514	0,19167	1,00255	-1,0068	2,00937	0,57595	-0,2457	-1,1606
$\Delta F_{T2}(x)$	0,04209	0,0406	0,62154	-0,4874	1,1089	0,32016	0,03541	-1,2841
$\Delta F_{H2}(x)$	0,03287	-0,0374	1,07973	-0,7308	1,81049	0,5993	0,33539	-1,2649
G_{N2}	-0,0114	0,04763	3,15206	-3,2299	6,38199	1,82932	0,05431	-1,2929
$\Delta F_{T3}(x)$	-0,1248	-0,0572	3,11789	-3,261	6,37894	2,18799	-0,0011	-1,3985
$\Delta F_{H3}(x)$	-0,0132	0,09054	3,42581	-3,3491	6,77491	2,25527	-0,0013	-1,3498
G_{N3}	0,00711	0,02386	1,01403	-1,1777	2,19177	0,58178	-0,1815	-1,1085
$\Delta F_{T4}(x)$	-0,5059	-0,5594	1,66235	-2,3104	3,97273	1,26016	0,19317	-1,2669
$\Delta F_{H4}(x)$	0,23545	-0,2218	3,73842	-2,9363	6,67468	1,96021	0,24961	-1,1633
G_{N4}	0,14434	-0,209	10,724	-11,125	21,8486	5,54557	-0,1589	-0,93
$\Delta F_{T5}(x)$	0,02034	0,06779	1,27835	-1,4008	2,6792	0,86889	-0,1358	-1,34
$\Delta F_{H5}(x)$	0,02432	0,05703	1,57307	-1,0814	2,65443	0,75816	0,20358	-0,9638
G_{N5}	-0,0017	-0,0219	1,14921	-1,0053	2,1545	0,52547	0,10418	-0,879
$\Delta F_{T6}(x)$	0,38049	0,48226	2,01762	-1,0878	3,10543	1,06778	0,07782	-1,468
$\Delta F_{H6}(x)$	-0,0947	-0,1425	0,9612	-0,97	1,93123	0,62358	0,2356	-1,2691
G_{N6}	-0,1448	-0,3627	11,7902	-10,343	22,1329	5,87912	0,10212	-0,9957
$\Delta F_{T7}(x)$	-0,0849	-0,1037	3,44956	-3,3849	6,83451	2,31216	0,03995	-1,3866
$\Delta F_{H7}(x)$	-0,0206	-0,0739	3,52173	-3,7608	7,28257	2,31772	-0,0468	-1,2488
G_{N7}	-0,0038	0,13235	0,9586	-1,1316	2,09023	0,53007	-0,3507	-0,9764
$\Delta F_{T8}(x)$	-0,3656	-0,3463	3,37936	-4,2108	7,5902	2,53263	-0,0159	-1,3577
$\Delta F_{H8}(x)$	0,00539	0,03207	2,79986	-2,6299	5,42977	1,85481	0,00576	-1,4756
G_{N8}	0,02837	0,10099	6,21577	-6,5758	12,7915	2,19279	-0,0467	0,54157

Построенные графики, позволяющие визуально оценить качество восстановления при длине ряда $n = 100$, представлены на рисунках 1 – 8 при соотношении оценок $\Phi_T:\Phi_H:\Phi_N=1:1:1$. Например, на рисунке 1 представлена трендовая составляющая, определенная как линейная функция. На рисунке 2 представлен один период гармонической составляющей, определенной по формуле (2) при $k = 1$. На рисунке 3 – шумовая составляющая, определенная по формуле (3). На рисунке 4 временной ряд получен суммированием трендовой, гармонической и шумовой составляющей. На рисунке 5 представлен тренд, восстановленный программой BelSim2#.SSA, на рисунке 6 трендовая составляющая скорректирована к виду, близкому к исходному. На рисунках 5 и 6 видны особенности восстановления – наличие краевого эффекта [5, 6], при котором наблюдаются значительные искажения на крайних элементах временного ряда. Качество восстановления гармонической составляющей показано на рисунке 7, на котором видны значительные искажения исходной гармоникой. Вид распределения восстановленной шумовой составляющей не изменился (рисунок 8).

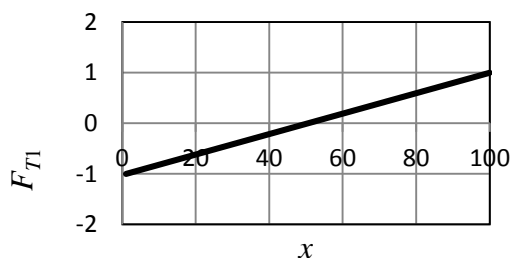


Рис. 1. Тренд исходный при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

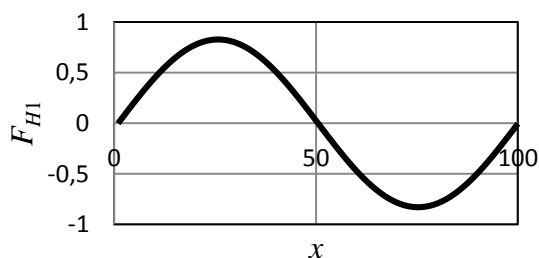


Рис. 2. Гармоника исходная при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

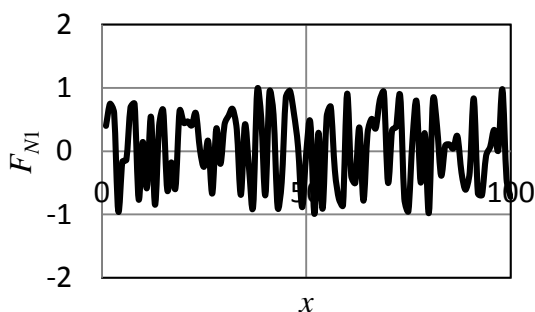


Рис. 3. Шум исходный при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

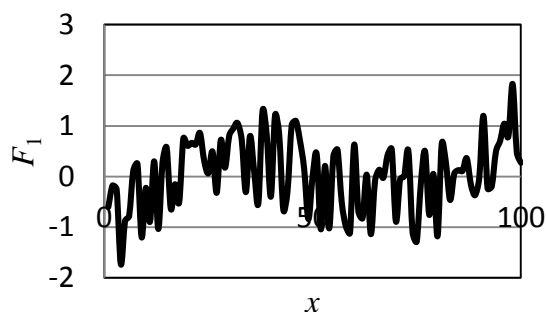


Рис. 4. Временной ряд исходный при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

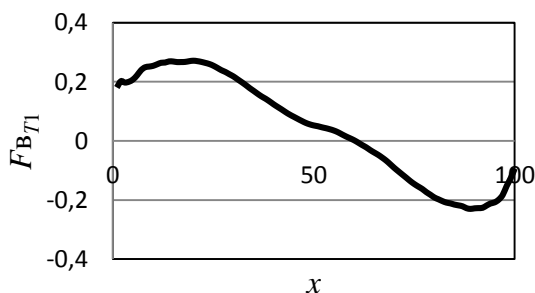


Рис. 5. Тренд восстановленный при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

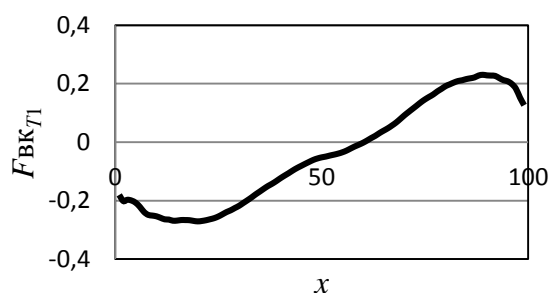


Рис. 6. Тренд восстановленный с коррекцией при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

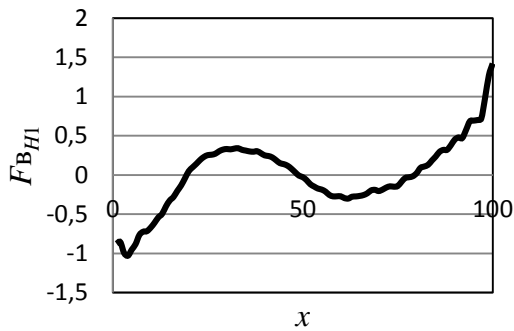


Рис. 7. Гармоника восстановленная при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

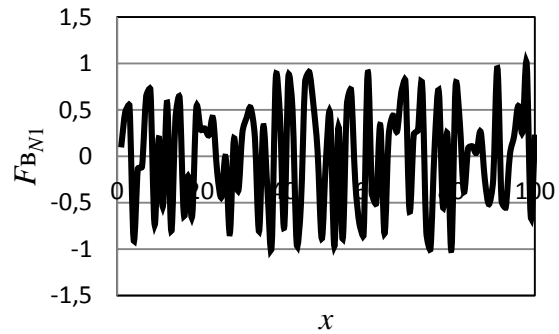


Рис. 8. Шум восстановленный при $\Phi_T:\Phi_H:\Phi_N=1:1:1$

3 ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Исследования проведены с нормированным временным рядом, в котором исключается постоянная составляющая тренда и используется динамическая трендовая составляющая G_{TD} . Для сравнения результатов использовали оценки Φ_{TD} , Φ_H и Φ_N , определенные по формулам (1).

В соответствии с результатами исследований отмечается, что трендовая составляющая восстанавливается в экспериментах с соотношением выбранных оценок $\Phi_T:\Phi_H:\Phi_N = 10:1:1$ при $n = 100, 200, 300$; $10:1:1$ при $n = 100, 200, 300$; $1:1:10$ при $n = 100$.

Гармоническая составляющая восстанавливается в экспериментах с соотношением оценок $\Phi_T:\Phi_H:\Phi_N = 1:1:1$ при $n = 100, 200, 300$; $1:10:1$ при $n = 100, 200, 300$; $10:1:1$ при $n = 100, n = 200$ (рисунок 1), $n = 300$; $10:10:1$ при $n = 100, 200, 300$.

Трендовая составляющая восстанавливается с изменением направления на противоположное в экспериментах с соотношением $\Phi_T:\Phi_H:\Phi_N = 1:1:1$ при $n = 100, 200, 300$; $1:1:10$ при $n = 200, 300$; $1:10:1$ при $n = 100, 200, 300$; $10:10:10$ при $n = 100, 200, 300$.

Гармоническая составляющая не восстанавливается (имеются значительные отклонения от амплитуды и числа периодов) в экспериментах с соотношением оценок $\Phi_T:\Phi_H:\Phi_N = 1:1:10$ при $n = 100, 200, 300$; $1:10:10$ при $n = 100, 200, 300$; $10:1:10$ при $n = 100, 200, 300$; $10:10:10$ при $n = 100, 200, 300$ [7].

4. ЗАКЛЮЧЕНИЕ

Выполненные исследования показывают неоднозначность восстановления при определенных соотношениях оценок, что необходимо учитывать при практическом применении метода сингулярного спектрального анализа для принятия решений.

Дальнейшие исследования предполагают определить, влияет ли размах линейной составляющей на качество восстановления. Для этого необходимо проверить погрешность восстановления временного ряда в зависимости от величины K_T , см. формулу (2). Другим направлением исследования является оценка влияния длины ряда на качество восстановления составляющих, что обусловлено наличием краевого эффекта при восстановлении составляющих временного ряда.

Выявленная неоднозначность восстановления трендовой составляющей требует проведения валидации программного продукта BelSim 2#.SSA с использованием математической системы MathCad по методике, предложенной в работе [8].

Литература

1. **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.
2. **Якимов, Е. А.** Программный модуль анализа сингулярного спектра числовых последовательностей данных «BelSim2#.SSA»: свидетельство о регистрации компьютерной программы № 518 / Е. А. Якимов, Д. М. Албкеират, Н. М. Чапаров, А. И. Якимов. – Минск: НЦИС, 2013. – Заявка № С20130042. – Дата подачи: 23.05.2013.
3. **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. Гомельс. гос. ун-та им. Ф. Скорины. – 2011. – № 6(69). – С. 204–214.
4. **Вадзинский, Р. Н.** Справочник по вероятностным распределениям / Р.Н. Вадзинский. – СПб.: Наука, 2001. – 295 с.
5. **Мартко, Е. О.** Сингулярный спектральный анализ как метод моделирования электрической нагрузки / Е. О. Мартко, И. В. Белицын // Ползуновский вестник. – 2009. – № 4. – С. 76–85.
6. **Якимов, Е. А.** О краевом эффекте при сингулярном спектральном анализе числовых последовательностей данных / Е. А. Якимов, Н. М. Чапаров, Д. М. Албкеират // Математическое и имитационное моделирование систем. МОДС 2012 : тез. докл. седьмой междунар. науч.-практич. конф., 25–28 июня 2012 г. – Чернигов-Жукин: ИПММС НАН Украины, 2012. – С. 384–386.
7. **Микулич, А. Н.** Особенности исследования метода сингулярного спектрального анализа временных рядов / А. Н. Микулич, В. А. Долгий, В. О. Батыщиков; науч. рук.: А. И. Якимов, Е. А. Якимов // 51-я студенческая научно-техническая конференция Белорусско-Российского университета: материалы конф., редкол.: И. С. Сазонов (гл. ред.) [и др.]; 21-22 мая 2015 г. – Могилев: Белорус.-Рос. ун-т, 2015. – С. 139.
8. **Якимов, Е. А.** Исследование SSA-метода на основе комплексного применения информационных технологий / Е. А. Якимов // Доклады БГУИР. – 2010. – № 2(48). – С. 77–83.

Микулич Алёна Николаевна

Студентка электротехнического факультета
Белорусско-Российский университет, г. Могилев
Тел.: +375(33) 691-51-73
E-mail: alionamikulich@gmail.com

Долгий Владимир Анатольевич

Студент электротехнического факультета
Белорусско-Российский университет, г. Могилев
Тел.: +375 (44) 592-64-34
E-mail: diastriker@list.ru

Батыщиков Владислав Олегович

Студент электротехнического факультета
Белорусско-Российский университет, г. Могилев
Тел.: +375 (44) 592-64-34
E-mail: maverick_css@mail.ru

Якимов Евгений Анатольевич

Старший преподаватель кафедры Автоматизированные системы управления, канд. техн. наук
Белорусско-Российский университет, г. Могилев
Тел.: +375(222) 25-24-47
E-mail: e-soft@bk.ru