

А.А. Кошулько, А.И. Кошулько

Тестирование полигармонического алгоритма МГУА

Рассматривается задача структурной идентификации колебательных процессов на основе полигармонического алгоритма МГУА. Приведены результаты тестирования программной реализации (язык C++) алгоритма полигармонического анализа временных рядов с применением тестовых и реальных данных. Представлен набор тестов, который можно использовать для оценки качества аналогичных программных средств

A problem of structural identification of oscillatory processes by the GMDH is considered. Results of testing the program realization (C++ language) of the algorithm of poly harmonic analysis of time series and real data are given. A set of tests which can be used for evaluation of the quality of the similar software is presented.

Розглядається задача структурної ідентифікації коливальних процесів за полігармонічним алгоритмом МГУА. Наведено результати тестування програмної реалізації (мова C++) алгоритму полігармонічного аналізу часових рядів із застосуванням тестових і реальних даних. Подано набір тестів, готовий для оцінки якості аналогічних програмних засобів.

Анализ литературы по МГУА говорит о том, что в последнее десятилетие гармонический алгоритм МГУА незаслуженно забыт, несмотря на то, что в свое время это было одно из лучших теоретических достижений в этой области [1], а сам алгоритм и программа показали себя эффективными во многих приложениях [2–4].

В данной статье приведена краткая характеристика полигармонического алгоритма (ПГА) и рассматривается его современная реализация на C++ для персональных компьютеров.

Построение алгоритма и работа программы основываются на принципах МГУА, при этом используются: разделение исходного ряда данных на обучающую и проверочную выборки, оценка качества модели на проверочной выборке по глубине минимума внешнего критерия.

Алгоритм [5, 6] ищет в исходных данных зависимость вида

$$f(a, b, \omega, I) = a_0 + \sum_{i=1}^m [a_i \cos(\omega_i I) + b_i \sin(\omega_i I)], \quad (1)$$

где $\cos \omega_1, \cos \omega_2, \dots, \cos \omega_m$ определяются как корни алгебраического уравнения

$$\begin{aligned} \cos(m\omega) - \alpha_1 \cos((m-1)\omega) - \dots \\ \dots - \alpha_{m-1} \cos \omega - 0,5\alpha_m = 0, \end{aligned} \quad (2)$$

коэффициенты α , которого должны удовлетворять системе линейных разностных уравнений

$$\sum_{j=1}^m (\alpha_j [f_{I+j} + f_{I-j}]) = f_I \quad (3)$$

где ω_i — частота колебаний ($0 < \omega_i < \pi$, $i = 1, 2, \dots, m$); I — время, связанное с началом отсчета наблюдений; a_i, b_i — коэффициенты, на основе которых вычисляется амплитуда; a_0 — свободный член (среднее выборки); m — количество частот.

Объясним значения левой части системы уравнений (3):

$$\sum_{j=1}^m (\alpha_j [f_{I+j} + f_{I-j}]) = \sum_{i=1}^m f_i 2 \sum_{j=1}^m (\alpha_j \cos(\omega_{ij})). \quad (4)$$

Для выполнения равенства (3) при любом I достаточно, чтобы

$$2 \sum_{j=1}^m (\alpha_j \cos(\omega_{ij})) = 1, \quad i = 1, 2, \dots, m. \quad (5)$$

Система уравнений (5) всегда имеет единственное решение, а константы $\alpha_1, \alpha_2, \dots, \alpha_m$ оцениваются с помощью ортогональной регрессии [6] для набора переменных t_{Ii} , $i = 0, 1, \dots, m$, где $t_{I0} = \sqrt{2} \cdot f_I$, $t_{Ij} = f_{I+j} + f_{I-j}$, $j = 1, 2, \dots, m$.

Оценки параметров $\omega_1, \omega_2, \dots, \omega_m$ вычисляются как корни на отрезке $[0, \pi]$ системы уравнений (5). Для этого с помощью формулы

$$\cos \omega j = 2 \cos \omega(j-1) \cdot \cos \omega - \cos \omega(j-2) \quad (6)$$

систему (5) приводят к виду

$$P_m(\cos \omega) = 0, \quad (7)$$

где P_m — некоторый многочлен степени m .

Уравнение (7) решают относительно $\cos \omega$. Оценки коэффициентов a_i, b_i находят из требования минимизации функционала

$$\sum_{k=1}^N [a_i \cos(\omega_i k) + b_i \sin(\omega_i k) - x_k]^2, \quad i = \overline{1, n}, \quad (8)$$

где x_k - k -й член выборки данных.

Одну итерацию алгоритма можно описать следующим образом: на основе обучающей выборки из таблицы исходных данных составляется система разностных уравнений (3); вычисляются их коэффициенты методом квадратного корня и корни многочлена (7) по методу Ньютона; оцениваются частоты и амплитуды гармоник. На проверочной выборке модель ряда наблюдений вида (1) оценивается по глубине минимума внешнего критерия, например:

$$KR = \left[\frac{\sum_{i=N_A+1}^{N_B} (x_i^P - x_i^M)^2}{\sum_{N_A}^{N_B} (x_i^P - x)^2} \right]^{1/2}, \quad (9)$$

где x_i^P, x_i^M — исходные значения ряда и полученные путем моделирования соответственно; x — среднее значение ряда на проверочной выборке; N_A, N_B — количество значений ряда в учебной и проверочной выборках соответственно. Количество итераций равно m и зависит от характера поведения внешнего критерия на проверочной выборке данных, т.е. вычислительный процесс завершается при достижении минимума внешнего критерия.

В связи с проблемами точности вычислений для разных реализаций алгоритма проводится тестирование программы полигармонического алгоритма МГУ А на специально подобранных по сложности тестах. Общая процедура тестирования заключается в том, что

по заданным уравнениям с известным числом гармоник насчитывается ряд значений, а потом по этому ряду данных восстанавливается исходное уравнение. При этом важна не только точность воспроизведения структуры исходного уравнения, но и точность, с которой это воспроизведение проведено. Семь тестов, приведенных в данной работе, отображают разные наборы частот и амплитуд, но необходимо отметить, что они, естественно, не исчерпывают всю сложность процессов, для анализа которых создавались алгоритм и программа.

Тестовые примеры построены по следующей схеме: для заданной тестовой функции F вычисляется ряд значений с единичным шагом аргумента; на основе полученного ряда значений программа ПГА восстанавливает функцию F . В примерах приводится: вид функции F , результаты работы программы и. в некоторых (наиболее сложных) примерах, график с наложением исходного и восстановленного уравнений. В отображении результатов работы программы приняты такие обозначения: N — количество точек в выборке входных данных; N_A — количество точек, использованных в алгоритме для обучения; m — количество частот колебаний; $A[*\cos]$ — коэффициент при косинусе; $B[*\sin]$ — коэффициент при синусе.

Пример 1 представлен одной гармоникой, разными коэффициентами при \cos и \sin и свободным членом, равным 10.

$$F = 10,00 + 4,00 \cdot \cos(0,35 \cdot i) + 2,00 \cdot \sin(0,35 \cdot i). \quad (10)$$

Воспроизведение этого примера программой ПГА на 200 точках обучающей выборки с нулевой ошибкой дало следующие результаты. Частота: 0,35, период: 17,952, амплитуда: 4,4721, $A[*\cos]$: 4,00, $B[*\sin]$: 2,00, фаза: 0,4636.

Пример 2 представлен одной гармоникой и свободным членом, равным 1,00.

$$F = 1,00 + 1,00 \cdot \cos(0,7954 \cdot i) + 0,00 \cdot \sin(0,7954 \cdot i). \quad (11)$$

Уравнение (11) воспроизведено программой ПГА на малой обучающей выборке из 10 точек с нулевой ошибкой. Частота: 0,7954, период: 7,8994, амплитуда: 1,00, $A[\cos]$: 1,00, $B[\sin]$: 0,00, фаза: 0,00.

Пример 3 составлен из двух гармоник с одинаковыми частотами и свободным членом, равным 1,33.

$$F = 1,33 + 1,00 \cdot \cos(0,544 \cdot i) + 0,00 \cdot \sin(0,544 \cdot i) + 0,00 \cdot \cos(0,744 \cdot i) + 1,00 \cdot \sin(0,744 \cdot i). \quad (12)$$

Уравнение (12) воспроизведено программой ПГА на 100 точках обучающей выборки с нулевой ошибкой (табл. 1, рис. 1).

Таблица 1

Свободный член = 1.3300 ($N = 100, NA = 100, m = 2$)					
: Частота : Период : Амплитуда : $A[\cos]$: $B[\sin]$: Фаза :					
0.544000	11.5500	1.0000	1.0000	0.0000	0.0000
0.744000	8.4451	1.0000	0.0000	1.0000	1.5708

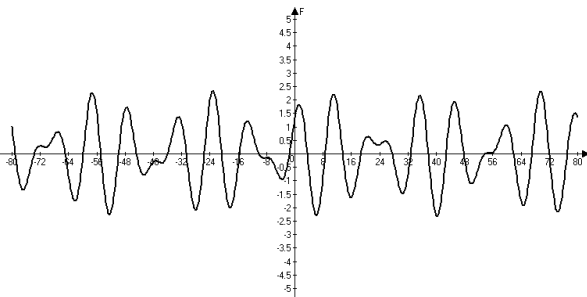


Рис. 1. Тестовый пример 3

Пример 4 (рис. 2) составлен из трех гармоник с разными частотами и свободным членом, равным -5.

$$F = -5,00 + 0,00 \cdot \cos(0,000001 \cdot i) + 1,00 \cdot \sin(0,000001 \cdot i) + 0,00 \cdot \cos(0,30 \cdot i) - 14,00 \cdot \sin(0,30 \cdot i) + 4,00 \cdot \cos(1,570796 \cdot i) + 0,00 \cdot \sin(1,570796 \cdot i) + 0,00 \cdot \cos(3,14159 \cdot i) - 7,00 \cdot \sin(3,14159 \cdot i). \quad (13)$$

Как показано в табл. 2, уравнение (13) воспроизведено программой ПГА на 200 точках обучающей выборки с удовлетворительной

точностью. В этом примере использованы частоты, близкие к 0, π , $\pi/2$.

Таблица 2

Свободный член = -4.9999 ($N = 200, NA = 200, m = 4$)					
: Частота : Период : Амплитуда : $A[\cos]$: $B[\sin]$: Фаза :					
0.300000	20.9440	14.0000	-0.0000	-14.0000	-1.5708
1.570796	4.0000	4.0000	4.0000	-0.0000	-0.0000
3.141589	2.0037	0.0037	-0.0002	-0.0037	-1.6165

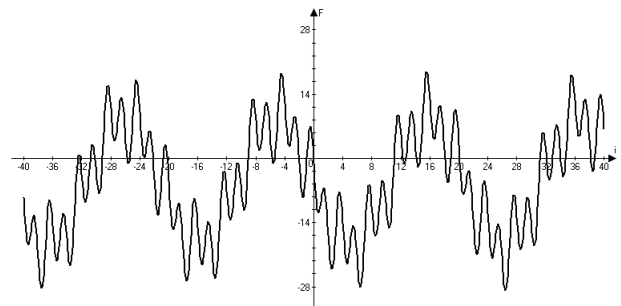


Рис. 2. Тестовый пример 4

Пример 5 (рис.3) составлен из пяти гармоник с разными частотами и нулевым свободным членом.

$$F = 0,00 + 0,00 \cdot \cos(0,30 \cdot i) - 1,00 \cdot \sin(0,30 \cdot i) + 30,00 \cdot \cos(0,60 \cdot i) + 0,00 \cdot \sin(0,60 \cdot i) + 0,00 \cdot \cos(1,20 \cdot i) - 30,00 \cdot \sin(1,20 \cdot i) + 80,00 \cdot \cos(2,70 \cdot i) + 0,00 \cdot \sin(2,70 \cdot i) + 0,00 \cdot \cos(3,14 \cdot i) - 80,00 \cdot \sin(3,14 \cdot i). \quad (14)$$

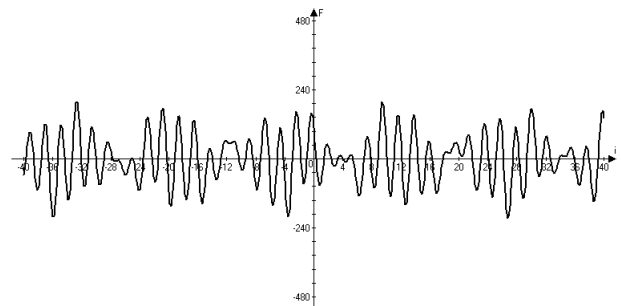


Рис. 3. Тестовый пример 5

Как показано в табл. 3, этот пример воспроизведен программой ПГА на 200 точках обучающей выборки с нулевой ошибкой.

Таблица 3

Свободный член = 0.0000 (N = 200, NA = 200, m = 5)					
: Частота : Период : Амплитуда : A[*cos] : B[*sin] : Фаза :					
0.300000	20.9439	1.0000	0.0000	-1.0000	-1.5708
0.600000	10.4720	30.0000	30.0000	0.0000	0.0000
1.200000	5.2360	30.0000	0.0000	-30.0000	-1.5708
2.700000	2.3271	80.0000	80.0000	0.0000	0.0000
3.110000	2.0203	79.9999	0.0009	-79.9999	-1.5708

Пример 6 представлен гармоникой со свободным членом, равным 1, и нулевым коэффициентом при sin.

$$F = 1,00 + 1,00 \cdot \cos(0,5236 \cdot i) + 0,00 \cdot \sin(0,5236 \cdot i). \quad (15)$$

Пример воспроизведен программой ПГА на 200 точках обучающей выборки с нулевой ошибкой. Частота: 0,5236, период: 12,00, амплитуда: 1,00, A[*cos]: 1,00, B[*sin]: -0,00, фаза: -0,00.

Пример 7 (рис. 4) составлен из десяти гармоник с разными частотами и свободным членом, равным 45,16.

$$F=45,16 - 7,73 \cdot \cos(0,11 \cdot i) - 5,84 \cdot \sin(0,11 \cdot i) + 5,21 \cdot \cos(0,15 \cdot i) + 0,848 \cdot \sin(0,15 \cdot i) - 9,44 \cdot \cos(0,49 \cdot i) + 5,302 \cdot \sin(0,49 \cdot i) + 3,64 \cdot \cos(0,51 \cdot i) + 4,21 \cdot \sin(0,51 \cdot i) - 4,66 \cdot \cos(0,52 \cdot i) + 4,649 \cdot \sin(0,52 \cdot i) + 1,06 \cdot \cos(0,57 \cdot i) + 19,700001 \cdot \sin(0,57 \cdot i) + 6,70 \cdot \cos(0,59 \cdot i) + 9,292 \cdot \sin(0,59 \cdot i) + 11,40 \cdot \cos(0,63 \cdot i) + 7,68 \cdot \sin(0,63 \cdot i) - 3,19 \cdot \cos(0,67 \cdot i) + 10,30 \cdot \sin(0,67 \cdot i) - 1,71 \cdot \cos(0,70 \cdot i) + 4,725 \cdot \sin(0,70 \cdot i). \quad (16)$$

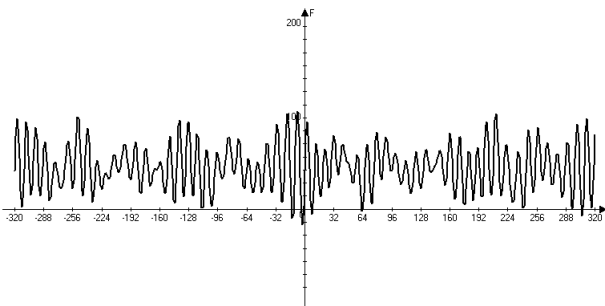


Рис. 4. Тестовый пример 7

Пример воспроизведен программой ПГА на 200 точках обучающей выборки с удовлетворительной точностью. Особенностью этого примера является то, что уравнение (16) было получено на фактических данных реального процесса и тем самым воссоздает возможную сложность процессов, которые будут моделироваться с помощью ПГА.

Таблица 4

Свободный член = 45.2841 (N = 200, NA = 200, m = 15)					
: Частота : Период : Амплитуда : A[*cos] : B[*sin] : Фаза :					
0.110059	57.0891	9.5489	-7.3856	-6.0526	-2.4551
0.150071	41.8682	5.5007	5.4178	0.9509	0.1737
0.490960	12.7978	9.1067	-9.1006	0.3317	3.1052
0.523455	12.0033	2.9702	-1.1128	-2.7539	-1.9548
0.574582	10.9352	6.0096	-3.8058	-4.6509	-2.2566
0.627346	10.0155	9.3345	6.9464	-6.2354	-0.7315
0.669377	9.3866	10.6606	-0.2386	10.6580	1.5932
0.699893	8.9774	4.9226	-2.0342	4.4826	1.9968
1.204492	5.2165	0.0592	-0.0115	0.0580	1.7668
2.772599	2.2662	0.0483	-0.0333	-0.0350	-2.3321
2.057358	3.0540	0.0279	-0.0109	-0.0257	-1.9733
3.041246	2.0660	0.0683	-0.0681	-0.0044	-3.0769
2.380885	2.6390	0.0713	-0.0629	0.0337	2.6497
1.761410	3.5671	0.0629	-0.0498	-0.0385	-2.4842
1.787922	3.5142	0.0040	0.0035	0.0021	0.5388

Покажем, насколько полезно использование полигармонического алгоритма МГУА для тестового моделирования некоторых гидрологических процессов, периодическая повторяемость которых проявляется в сезонных и многолетних колебаниях уровней воды. Выделение гармонических составляющих в начальном ряде наблюдений за некоторый период служит основой для поиска генетической связи между колебаниями уровня подземных вод и факторами, которые, возможно, эти колебания вызывают, а также для прогнозов [3]. Отметим, что ПГА, в отличие от классического Фурье-анализа, позволяет выделять гармоники с некрратными частотами, важные при анализе колебательных процессов, в которых отсутствует точная периодичность, характерная для большинства естественных процессов.

Полигармонический анализ многолетних колебаний пьезометрического уровня водоносного горизонта

Выделение по МГУА полигармонического тренда с некротными частотами в колебаниях пьезометрического уровня водоносного горизонта показано на примере анализа 240 среднемесячных наблюдений (рис. 5). В процессе работы алгоритма в исходном ряде наблюдений выделен тренд, который состоит из двух гармоник с периодами 119 лет и 1 год (табл. 5).

Гармоники выделяются по точкам обучающей последовательности, а выбор лучшего тренда осуществляется по минимуму критерия (9) на проверочной последовательности данных.

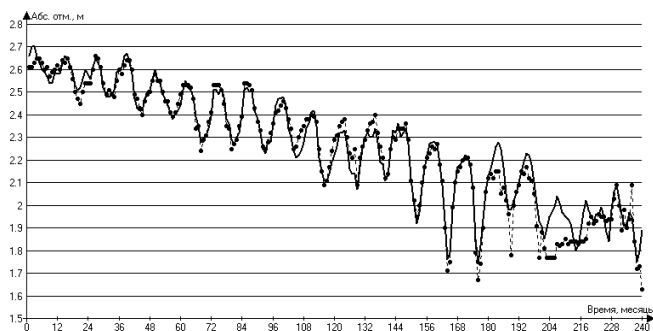


Рис. 5. Гармонический тренд пьезометрического уровня водоносного горизонта

Таблица 5

Частота	Коэффициенты		Период	Амплитуда (м)
	<i>a</i>	<i>b</i>		
0,0044	-0,6839	0,2890	119	0,74
0,5231	0,1144	0,0320	1	0,12

Полигармонический анализ многолетних рядов наблюдений за температурой воздуха и осадков

Колебательный характер таких рядов дает возможность использовать ПГА для выявления причин и оценки параметров этих процессов [4]. ПГА позволяет разложить исходный ряд наблюдений на составные (гармоники), оценить их параметры (частоту и амплитуду), вычислить период и фазу, а также связать эти параметры с известным или ожидаемым влиянием на атмосферу.

Исследуются результаты многолетних наблюдений за температурой воздуха и осадками, полученные на метеостанции г. Киева за 127 лет (1856-1982 гг., рис. 6 и 7).

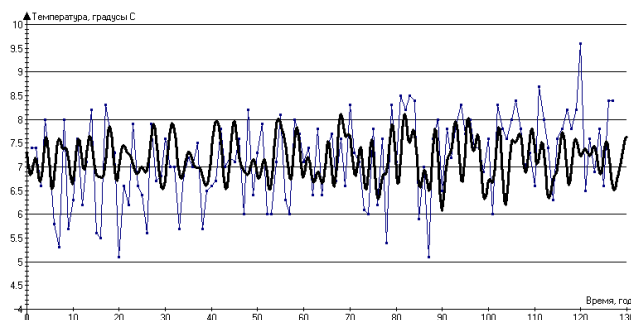


Рис. 6. График многолетних колебаний температуры воздуха: тонкая линия — измеренное значение, толстая — результат моделирования

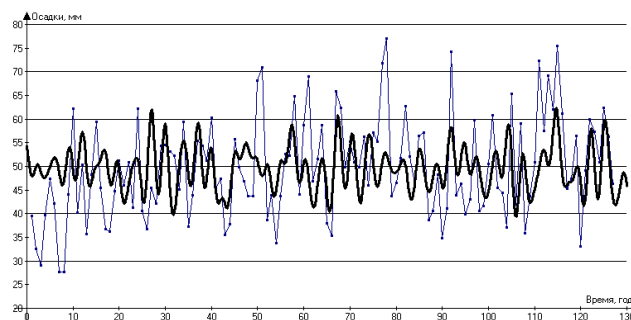


Рис. 7. График многолетних колебаний осадков: тонкая линия — измеренное значение, толстая — результат моделирования

Визуальный анализ этих данных не позволяет обнаружить какие-либо закономерности. Для поиска скрытых закономерностей используется ПГА на основе предложенной методики (параметры моделей приведены в табл. 6).

При помощи программы ПГА в измерениях температуры воздуха и осадков выделена гармоника с периодом 3 года и достаточно весомой амплитудой. Гармоника с наибольшей амплитудой для температуры воздуха имеет период 4,7 года, это в 1,63 раза превышает период гармоника с максимальной амплитудой для осадков. Характерно, что гармоники ряда температуры воздуха с частотами 2,2; 2,7; 2,8 (периоды от 2,3 до 2,9 года) могут обусловить явление биения. На это указывает и то, что

спектр температуры воздуха в границах частот от 2,2 до 2,8 «рассыпается» на три близкие частоты (2,2; 2,7; 2,8) с равными амплитудами (0,2). Менее выразительно это явление при анализе многолетнего ряда наблюдений за осадками, в спектре которого амплитуды меняются в незначительных границах — от 2,37 до 2,65.

Таблица 6

Температура воздуха (°С)		Осадки в мм	
Свободный член = 7,17; КР = 0,24		Свободный член = 49,62; КР = 0,50	
Амплитуда температуры	Период	Период	Амплитуда (мм)
0,18	13,67	21,20	2,08
0,22	12,24	11,04	2,38
0,13	6,93	9,57	2,67
0,32	4,70	5,66	2,40
0,13	4,34	3,68	2,15
0,27	3,38	3,40	3,13
0,20	2,86	3,07	3,37
0,20	2,33		
0,20	2,28		

Можно сделать вывод, что полигармоничность многолетнего ряда наблюдений за осадками в 2 раза меньше полигармоничности температурного, это подтверждает значение критерия (9), равное 0,5. В обоих рядах наблюдений за температурой воздуха и осадками выделено приблизительно по семь значимых гармоник. Длиннопериодические гармоники в многолетнем ряде температуры воздуха: 13, 7; 12, 2; 6, 9 лет, а осадков: 21; 11; 9, 6. В рядах наблюдений за осадками и температурой воздуха выделено только по одной гармонике, абсолютно совпадающих по частоте (частота 1,8, период 3,4). В обоих спектрах эта гармоника занимает второе место по амплитуде.

Заключение. Предложенная методика тестирования полигармонического алгоритма МГУА позволила выделить особенности и

подтвердить достаточную точность этого алгоритма при вычислениях. На нескольких примерах показана возможность применения ПГА для моделирования сложных природных процессов. Представлен набор тестов для оценки качества аналогичных программных средств.

1. Ивахненко А.Г., Высоцкий В.М., Чеберкус В.И. Долгосрочное прогнозирование колебательных процессов при помощи выделения гармонического тренда оптимальной сложности по критерию баланса переменных // Автоматика. — 1975. — № 1. — С. 23-31.
2. Ивахненко А.Г., Кротов Г.И., Чеберкус В.Л. Многорядный алгоритм самоорганизация долгосрочных прогнозов (на примере экологической системы оз.Байкал) // Автоматика. — 1980. — №4. — С. 28-47.
3. Ивахненко А.Г., Ивахненко М.А., Кошулько А.И. Индуктивные методы самоорганизации моделей на ЭВМ и сплайны в гидрогеологических исследованиях // Водные ресурсы. — 1985. — № 4. — С. 76-85.
4. Кошулько А.И., Попков П.В., Юрачковский Ю.П. Полигармонический анализ многолетних рядов температуры воздуха и осадков // Анализ и прогноз многолетних временных рядов.— Сб. науч. тр. ВАСХНИЛ. — Сиб. отделение СибНИИЗхим. Новосибирск, 1988. — № 6. — С. 45-53.
5. Юрачковский Ю.П., Чеберкус В.И. Самоорганизация моделей полигармонических полей // Автоматика. — 1981, — № 1. — С. 37-45.
6. Юрачковский Ю.П., Попков П.В. Один из подходов к оцениванию параметров в алгоритмах моделирования полигармонических процессов // Анализ и прогноз многолетних временных рядов. — Сб. науч. тр. ВАСХНИЛ. — Сиб. отделение СибНИИЗхим, Новосибирск, 1988. — №6. — С. 17-21.

Тел. для справок: 266-41-12 (Киев)
E-mail: kai@d461.icyb.kiev.ua
© А.А. Кошулько, А.И. Кошулько, 2003