

С.Н. Ефименко

**Построение систем прогнозных моделей многомерных взаимосвязанных процессов**

Разработаны теоретические основы рекуррентно-параллельных вычислений в комбинаторном алгоритме МГУА для моделирования и прогнозирования сложных многомерных взаимосвязанных процессов в классе моделей векторной авторегрессии. Продемонстрирована эффективность разработанного алгоритма.

Розроблено теоретичні основи рекуррентно-параллельних обчислень у комбінаторному алгоритмі МГУА для моделювання та прогнозування складних багатовимірних взаємозв'язаних процесів у класі моделей векторної авторегресії. Продемонстровано ефективність розробленого алгоритму.

**Введение.** В статье решается задача математического моделирования и прогнозирования многомерных взаимосвязанных временных рядов, которая находит свое применение, прежде всего, в экономической, экологической, социологической сферах [1, 2]. Если моделированию одномерных временных рядов в научной литературе уделяется внимание, то опыт моделирования многомерных временных рядов – недостаточен.

**Постановка задачи**

В случае прогнозирования векторного процесса, представленного в виде совокупности временных рядов (многомерного временного ряда), естественно ориентироваться на такой класс моделей, как векторная авторегрессия [3]. Рассмотрим один из возможных подходов к структурно-параметрической идентификации такого процесса, когда параметры для каждой модели оцениваются независимо. Недостаток такого подхода состоит в том, что параметры отдельных моделей взаимосвязанных процессов взаимозависимы. Для устранения этого недостатка в работе используется алгоритм, согласно которому для каждого из процессов выбирается не одна лучшая модель, а несколько. Это делается для того, чтобы из выбранных лучших моделей скомбинировать все возможные варианты систем и, согласно дополнительному критерию, выбрать лучшую.

**Модели векторной авторегрессии для прогнозирования многомерных взаимосвязанных процессов**

Модель векторной авторегрессии (*Vector Autoregression, VAR*) предложена Кристофером Симсом в 1980-м. Строится она по стационарным временным рядам. Это система уравнений,

в которой каждая переменная (компонента многомерного временного ряда) представлена в виде линейной комбинации всех переменных в предыдущие моменты времени. Порядок такой модели определяется порядком запаздываемых значений (лагов). В самом простом случае двух временных рядов с одним лагом модель *VAR* имеет такой вид:

$$\begin{aligned} x_1(t) &= \theta_{11}x_1(t-1) + \theta_{12}x_2(t-1); \\ x_2(t) &= \theta_{21}x_1(t-1) + \theta_{22}x_2(t-1), \end{aligned} \quad (1)$$

где  $\theta_{ij}$ ,  $i, j = 1, 2$  – параметры модели.

В общем случае для  $m$  временных рядов и  $k$  лагов такая модель имеет вид системы  $m$  уравнений

$$\begin{aligned} x_1(t) &= \theta_{11}x_1(t-1) + \dots + \theta_{1k}x_1(t-k) + \\ &+ \theta_{1,k+1}x_2(t-1) + \dots + \\ &+ \theta_{1,2k}x_2(t-k) + \dots + \theta_{1,mk}x_m(t-k); \\ &\dots \end{aligned} \quad (2)$$

$$\begin{aligned} x_m(t) &= \theta_{m1}x_1(t-1) + \dots + \theta_{mk}x_1(t-k) + \\ &+ \theta_{m,k+1}x_2(t-1) + \dots + \\ &+ \theta_{m,2k}x_2(t-k) + \dots + \theta_{m,mk}x_m(t-k) \end{aligned}$$

или в матричном виде

$$X(t) = \sum_{j=1}^k \Theta_j X(t-j), \quad (3)$$

где  $\Theta_j$ ,  $j = \overline{1, k}$  – матрицы коэффициентов модели размерности  $m \times m$ .

Структуру модели (3) и ее коэффициенты будем определять по комбинаторному алгоритму *COMBI* МГУА [4]. Модели формируются в виде системы линейных разностных уравнений. Число аргументов для каждого из  $m$  взаимосвязанных процессов составляет  $m \cdot k$ .

## Методика структурно-параметрической идентификации моделей векторной авторегрессии

Общая структура модели в виде системы  $m$  разностных уравнений определяется в результате выполнения таких операций:

Шаг 1. Исходя из количества  $m$  взаимосвязанных процессов и запаздываемых значений  $k$ , формируется массив данных из  $m \cdot k$  аргументов.

Шаг 2. Определяется максимальная сложность для перебора и выполняется моделирование по алгоритму *COMBI* МГУА с последовательным усложнением структур моделей на основе рекуррентно-параллельных вычислений. Для каждой исходной переменной выбирается  $F$  лучших (по значению критерия регулярности) моделей. В целом на следующий шаг передается  $F \cdot m$  моделей.

Шаг 3. Выполняется перебор  $G = F^m$  вариантов систем моделей и выбирается лучшая по значению системного интегрального критерия качества векторных моделей, вычисляемый на заданной части начальной выборки данных в режиме прогноза процесса на заданное количество шагов  $n_s$ :

$$B = \sum_{i=1}^{n_s} \sum_{j=1}^m (x_{ij} - x_{ij}^*)^2, \quad (4)$$

где  $x_{ij}^*$  – результат пошагового интегрирования системы из  $m$  уравнений.

При определении свободы выбора  $F$  (количества лучших моделей) следует учитывать, что время выполнения третьего шага (перебора всех вариантов систем моделей) быстро растет и может превысить приемлемое значение.

На рис. 1 показаны экспериментальные результаты для тестовой задачи с  $m = 11$  времен-

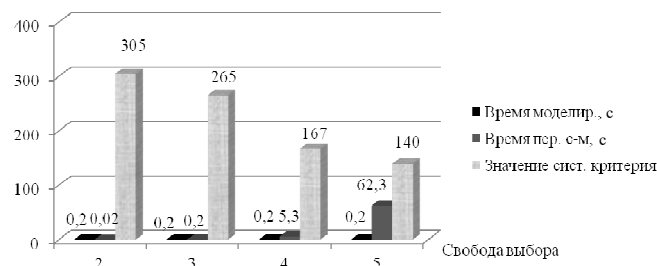


Рис. 1. Зависимость времени моделирования и значения системного интегрального критерия от свободы выбора  $F$

ными рядами и  $k = 2$  запаздываемыми значениями, согласно которым время перебора вариантов систем моделей (с вычислением значения системного интегрального критерия) растет экспоненциально.

## Построение моделей векторной авторегрессии на основе рекуррентно-параллельного алгоритма

Для построения моделей векторной авторегрессии используется комбинаторный алгоритм *COMBI* МГУА [5] с последовательным усложнением структур моделей, рекуррентным оцениванием параметров моделей с помощью модифицированного алгоритма Гаусса и распараллеливанием вычислений на кластерных многопроцессорных системах. Использование алгоритма *COMBI* предусматривает полный перебор всех возможных моделей и выбор лучшей по значению критерия. В процессе перебора для каждого из  $m$  взаимосвязанных процессов сравниваются линейные модели с  $g = m \cdot k$  (в нашем случае) аргументами (входами)

$$\hat{y}_v = X_v \hat{\theta}_v, \quad v = 1, \dots, 2^g - 1, \quad (5)$$

где десятичному числу  $v$  ставится в соответствие двоичный структурный вектор  $d_v = \{d_i\}$ ,  $i = \overline{1, g}$  с элементами 0 или 1 (включение в модель либо исключение из нее соответствующего аргумента). Последовательность генерации двоичных векторов организована таким образом, что сначала создаются все сочетания с одной единицей в составе структурного вектора (всего генерируется  $C_g^1 = g$  возможных вариантов), затем – с двумя единицами ( $C_g^2 = \frac{g(g-1)}{2}$  возможных вариантов) и т.д. к

одному варианту ( $C_g^g = 1$ ) включения в модель всех аргументов.

Способ равномерного разбиения общего количества моделей на все процессоры кластерной системы (при условии, что одинаковое общее количество аргументов будет приходиться на каждый процессор) разработан в [5]. Схема алгоритма *COMBI* с последовательным усложнением структур для построения моделей вектор-

ной авторегрессии выбрана в связи с тем, что она позволяет частично решать задачу полного перебора в случае, когда такой перебор (за приемлемое время моделирования) становится невозможным даже с распараллеливанием (ориентировочно при  $g > 50$ ). В таком случае полный перебор целесообразно выполнять не среди всех возможных моделей, а лишь моделей ограниченной сложности. Другой причиной применения ограничения на сложность может быть недостаточное количество точек учебной части выборки  $n_A$ . При  $n_A < g$  перебор будет выполняться среди моделей сложности не более чем  $n_A$ . Если же одновременно  $g < 50$  и  $g < n_A$ , то структуры моделей будут усложняться от 1 до  $g$ , т.е. будет реализован полный перебор среди всех возможных моделей.

Поскольку время построения моделей векторной авторегрессии фактически определяется временем перебора вариантов систем моделей (это четко видно по двум диаграммам для свободы выбора  $F = 5$  из рис. 1), то эффективность распараллеливания этого этапа алгоритма будет даже более приоритетной, нежели эффективность распараллеливания этапа построения  $F$  лучших моделей для каждого из  $m$  выходов (временных рядов). Поэтому рассмотрим способ распараллеливания именно этого этапа.

Поскольку выполняется перебор  $F^m$  вариантов систем моделей, то целесообразно использовать структурный вектор в виде числа с основой  $F$ , элементы  $\{f_i\}$ ,  $i = \overline{1, m}$  которого изменяются от 1 до  $F$  и указывают, которая из лучших (по значению критерия) моделей для  $i$ -го временного ряда будет включена в модель VAR. Покажем на простом примере ( $m = 2$ ,  $F = 3$ ) все варианты систем моделей (их будет  $G = 3^2 = 9$ ):

$\{1,1\}$  – для первого и второго выходов берутся первые из лучших моделей;

$\{1,2\}$  – для первого выхода берется первая модель, для второго – вторая;

$\{1,3\}$  – для первого выхода берется первая модель, для второго – третья;

$\{2,1\}$  – для первого выхода берется вторая модель, для второго – первая;

$\{2,2\}$  – для первого и второго выходов берутся вторые из лучших моделей;

$\{2,3\}$  – для первого выхода берется вторая модель, для второго – третья;

$\{3,1\}$  – для первого выхода берется первая модель, для второго – первая;

$\{3,2\}$  – для первого выхода берется третья модель, для второго – вторая;

$\{3,3\}$  – для первого и второго выходов берутся третьи по росту значения критерия модели.

Такая схема может быть достаточно легко применима к распараллеливанию на заданное количество процессоров. Идея равномерного разбиения общего количества системных моделей на  $p$  процессоров кластерной системы состоит в следующем:

- определяем общее количество системных моделей  $G = F^m$ ;

- количество моделей  $G$  равномерно распределяем между всеми  $p$  процессорами кластера;

- если  $G$  не делится нацело на  $p$ , то на первый процессор приходится большая нагрузка (но не более чем на  $p$  дополнительных моделей), нежели на другие процессоры.

### Тестирование работоспособности алгоритма

По системной модели

$$x_1(t) = 0,74x_3(t-2) + 42,93x_6(t-2) - 0,06x_9(t-2) + 7,96x_{11}(t-1) - 15,16x_{11}(t-2);$$

$$x_2(t) = 0,27x_2(t-1) - 0,93x_3(t-2) - 0,59x_9(t-1) - 0,44x_9(t-2) + 57,04x_{11}(t-2);$$

$$x_3(t) = -5x_1(t-1) + 4,28x_1(t-2) - 0,29x_7(t-2) - 0,5x_9(t-2) + 9,98x_{10}(t-2);$$

$$x_4(t) = 1,46x_4(t-2) + 0,08x_7(t-2) + 0,14x_8(t-2) - 0,04x_9(t-2) - 8,68x_{11}(t-2);$$

$$x_5(t) = 0,03x_2(t-2) - 0,003x_7(t-1) - 0,05x_8(t-1) + 0,01x_9(t-1) + 0,01x_9(t-2);$$

$$x_6(t) = 0,02x_1(t-1) - 0,03x_4(t-1) + 0,75x_6(t-2) + 0,004x_{10}(t-1) + 0,15x_{11}(t-1);$$

$$x_7(t) = -1,04x_2(t-1) + 0,78x_4(t-2) + 288,53x_6(t-1) - 249,19x_6(t-2) + 9,33x_{11}(t-2);$$

$$x_8(t) = 1,81x_2(t-1) - 4,71x_5(t-1) - 1,96x_9(t-1) - 1,86x_9(t-2) + 118,35x_{11}(t-2);$$

$$\begin{aligned}
x_9(t) &= 7,38x_1(t-2) - 13,94x_4(t-2) - \\
&- 1,46x_8(t-2) - 0,6x_9(t-2) + 41,94x_{10}(t-2); \\
x_{10}(t) &= 0,32x_4(t-2) - 11,34x_6(t-2) + \\
&+ 0,04x_8(t-2) - 0,04x_9(t-2) + 3,35x_{11}(t-2); \\
x_{11}(t) &= 0,08x_4(t-2) - 0,21x_5(t-1) - \\
&- 2,06x_6(t-2) + 0,01x_7(t-1) + 0,27x_{11}(t-1)
\end{aligned}$$

сгенерирована выборка и выполнено моделирование. Воспроизведена истинная системная модель со структурным вектором  $\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$ , для которой значение системного интегрального критерия качества векторных моделей  $B = 3,4e - 013$ .

### Тестирование эффективности распараллеливания алгоритма

Для экспериментального определения эффективности распараллеливания разработанного алгоритма выполнен тестовый эксперимент по решению задачи структурно-параметрической идентификации. Строилась системная модель для  $m = 11$  временных рядов с  $k = 2$  запаздываемыми значениями. Вычисление было распределено на пять потоков и последовательно выполнено на персональном компьютере с процессором *Intel Pentium M* с частотой 1,73 ГГц. Таким образом, получаем результат, приближенный к теоретическому, ввиду исключения потерь, связанных с межпроцессорным взаимодействием.

Для каждого временного ряда выбирались  $F = 5$  лучших моделей по критерию регулярности [4]. Измерялось время выполнения каждого из пяти потоков и время работы программы без распараллеливания. Результат эксперимента в виде диаграммы времени выполнения представлен на рис. 2.

Результаты эксперимента можно использовать для вычисления эффективности распараллеливания

$$E = \frac{T_1}{5 \times T_{5\max}} \times 100\% \quad (6)$$

и равномерности нагрузки

$$P = \left(1 - \frac{T_{5\max} - T_{5\min}}{T_{5\max}}\right) \times 100\%, \quad (7)$$

где  $T_1$  – время выполнения алгоритма с одним потоком (т.е. без распараллеливания),  $T_{5\max}$  –

время выполнения алгоритма с распараллеливанием на пять потоков (определяется как максимальное среди пяти потоков время выполнения программы),  $T_{5\min}$  – минимальное из пяти потоков время выполнения программы.

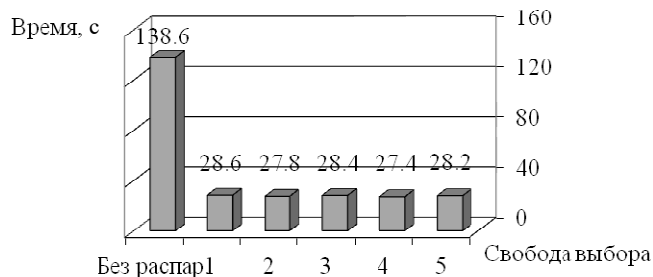


Рис. 2. Время выполнения алгоритма

Суть формулы (6) состоит в том, что если при использовании распараллеливания на  $n$  потоков (в данном случае пяти) время моделирования уменьшается в  $n$  раз, то эффективность распараллеливания составляет 100 процентов. В соответствии с (7), для обеспечения стопроцентной равномерности нагрузки все вычислители (ядра, процессоры, процессы) должны выполнять моделирование за одно и то же время.

В табл. 1 представлены экспериментальные значения показателей эффективности алгоритма

Т а б л и ц а 1. Экспериментальные результаты

Эффективность распараллеливания, %	97
Равномерность нагрузки, %	96

### Моделирование и прогнозирование показателей инвестиционной сферы Украины

Цель исследования – разработка экономико-математической модели для использования поддержки принятия эффективных управленческих решений.

**Описание исходных данных в поставленной задаче.** Были использованы данные Министерства экономики за 1996–2015 гг. (всего  $n = 20$  точек) для девяти показателей инвестиционной сферы Украины (за исключением пропущенного значения за 2015 г. для показателя  $x_2$ ) с целью построения модели векторной авторегрессии этих показателей в виде (3) и получения прогнозных значений на 2013–2015 гг. (которые не были использованы при моделировании):

$x_1$  – объем инвестиций в основной капитал в фактических ценах, млрд грн;

$x_2$  – степень износа основных фондов, процент к предыдущему году;

$x_3$  – чистый прирост прямых внешних инвестиций (ПВИ), млрд грн;

$x_4$  – доля иностранных инвестиций в общем объеме инвестиций в основной капитал, %;

$x_5$  – общий объем инвестиций, млрд грн;

$x_6$  – отношение объема инвестиций в основной капитал к валовому внутреннему продукту (ВВП), %;

$x_7$  – отношение чистого прироста ПВИ к ВВП, %;

$x_8$  – отношение общего объема инвестиций к ВВП, %;

$x_9$  – официальный обменный курс, грн./USD – в среднем за период.

### Построение прогнозных моделей векторной авторегрессии

Для получения модели векторной авторегрессии (3) для каждого из девяти показателей отбирались по пять лучших моделей (в виде системы линейных разностных уравнений), структура и коэффициенты которых определялись по алгоритму COMBI МГУА с последовательным усложнением. Значение свободы выбора  $F = 5$ , а также количество запаздываемых значений  $k = 2$  установлены с целью получения приемлемого времени моделирования. Поэтому в целом для каждого из девяти показателей выполнялся перебор среди 18 аргументов.

После отбора 45 лучших частных моделей выполнялся перебор  $G = 5^9$  вариантов систем моделей и отображена лучшая по значению системного интегрального критерия качества векторных моделей (4), который вычислялся в режиме прогноза процесса на пять шагов.

### Результаты прогнозирования и их анализ

Построена такая системная модель:

$$\begin{aligned} x_1(t) = & -0,317x_1(t-1) + 2,268x_2(t-2) + \\ & + 6,517x_3(t-1) - 5,719x_4(t-2) + 31,295x_6(t-1) + \\ & + 113,535x_6(t-2) + 149,356x_7(t-2) - \\ & - 21,236x_8(t-1) - 124,549x_8(t-2) - 0,302x_9(t-1); \end{aligned}$$

Т а б л и ц а 2. Данные показателей инвестиционной сферы Украины

	1996	1997	1998	1999	...	2010	2011	2012	2013	2014	2015
$x_1$	12,56	12,4	13,96	17,55	...	171,09	259,93	293,69	247,89	204,06	251,15
$x_2$	40	38	40,4	42,3	...	74,9	75,9	76,7	77,3	83,5	–
$x_3$	0,99	1,16	1,83	1,95	...	37,15	37,08	40,75	29,53	-64,89	-66,63
$x_4$	7,89	9,39	13,11	11,09	...	17,84	12,49	12,19	10,64	-46,62	-36,11
$x_5$	13,55	13,56	15,79	19,5	...	208,24	297,02	334,45	277,42	139,17	184,53
$x_6$	15,4	13,3	13,6	13,5	...	15,8	19,96	20,81	16,92	12,86	12,69
$x_7$	1,2	1,2	1,8	1,5	...	3,43	2,85	2,89	2,02	-4,09	-3,37
$x_8$	16,62	14,53	15,39	14,95	...	19,24	22,81	23,7	18,93	8,77	9,32
$x_9$	1,83	1,86	2,45	4,13	...	7,94	7,97	7,99	7,99	15,77	23,41

$$\begin{aligned} x_2(t) = & -0,08x_1(t-1) + 0,598x_2(t-2) + \\ & + 1,055x_3(t-1) + 1,564x_4(t-1) + 1,2x_6(t-1) - \\ & - 9,014x_7(t-1) - 0,321x_8(t-2) + 1,063x_9(t-1); \\ x_3(t) = & 0,044x_2(t-1) + 0,692x_3(t-1) + 1,223x_9(t-1); \\ x_4(t) = & -0,021x_5(t-1) + 0,766x_8(t-2) - \\ & - 0,539x_9(t-1) + 0,865x_9(t-2); \\ x_5(t) = & 5,339x_3(t-1) - 4,266x_4(t-1) + 19,413x_9(t-2); \\ x_6(t) = & 0,322x_2(t-1) + 0,782x_3(t-1) - \\ & - 0,675x_4(t-1) - 0,149x_5(t-1) + 1,4x_6(t-1) + \\ & + 12,377x_6(t-2) + 13,751x_7(t-2) - 13,077x_8(t-2); \\ x_7(t) = & 0,222x_4(t-1); \\ x_8(t) = & -0,029x_5(t-1) + 0,86x_7(t-2) + \\ & + 0,859x_8(t-1) + 0,964x_9(t-2); \\ x_9(t) = & 0,395x_4(t-1) + 0,003x_4(t-2) + \\ & + 0,013x_5(t-1) + 0,144x_6(t-2) - 1,277x_7(t-1). \end{aligned}$$

Значения средних относительных ошибок в процентах для системной модели в трех экзаменационных точках представлены в табл. 3.

Т а б л и ц а 3. Точность моделей в экзаменационных точках

Год	2013	2014	2015
Средняя относительная ошибка, %	16,7	87,1	146,1

Результаты моделирования и табличные значения показателей  $x_2$  и  $x_3$ , для которых получена соответственно наихудшая и наилучшая точность на экзаменационной выборке, представлены на рис. 3 и 4.

Построенная системная модель дает приемлемый прогноз на первую точку (2013 г.) для всех девяти показателей инвестиционной сферы Украины. Существенное расхождение реальных и прогнозных значений для второй и третьей прогнозных точек можно объяснить прежде всего экономическим кризисом 2014–2015 гг.,

а также возможным неучетом некоторых экономических и других влияющих факторов.

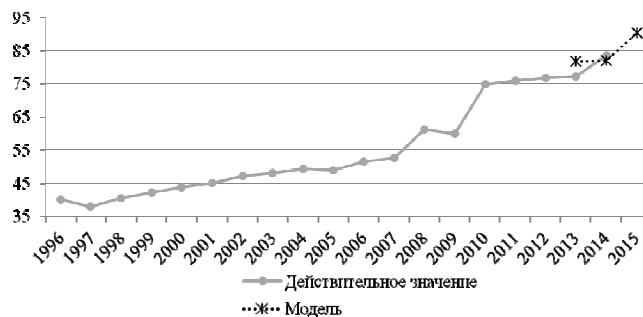


Рис. 3. Результаты моделирования для показателя  $x_2$

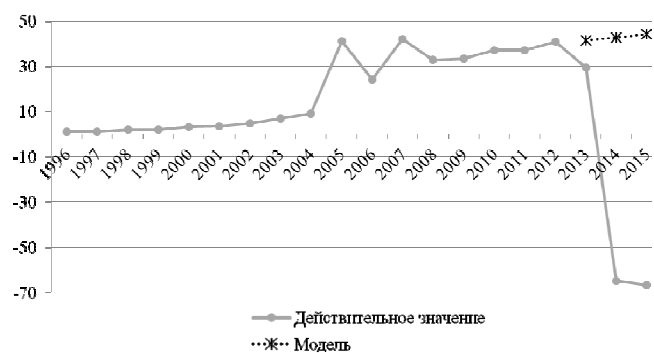


Рис. 4. Результаты моделирования для показателя  $x_3$

Таким образом, построенная системная модель может быть применена для краткосрочного прогноза.

**Заключение.** Разработанный принцип распараллеливания вычислений в комбинаторном алгоритме *COMBI* МГУА с рекуррентным оценением параметров моделей используется для построения дискретных прогнозных моделей динамики сложных многомерных взаимосвязанных процессов.

Особенность схемы распараллеливания алгоритма *COMBI* позволяет частично решать задачу полного перебора в случае, когда количество аргументов для перебора превышает возможности схемы алгоритма с полным перебором.

Разработанные программные средства для моделирования и прогнозирования сложных многомерных взаимосвязанных процессов на основе высокопродуктивного рекуррентно-параллельного алгоритма МГУА в классе дискретных динамических моделей векторной авторегрессии, применимых для моделирования и прогнозирования взаимосвязанных процессов в инвестиционной сфере Украины с целью информационной поддержки управленческих решений.

1. Єфіменко С.М., Кваша Т.К., Степашико В.С. Системне прогнозування динаміки взаємозалежних показників енергетичної сфери України // Індуктивне моделювання складних систем: Зб. наук. праць. – К.: МННЦ ІТ та С. – 2009. – 7. – С. 54–60.
2. Костенко Ю.В. Моделювання багатовимірних циклічних процесів за дворівневим алгоритмом МГУА // Там же. – 2011. – 3. – С. 99–109.
3. Гурский С.К. Адаптивное прогнозирование временных рядов в электроэнергетике. – Минск: Наука и техн., 1983. – 271 с.
4. Ивахненко А.Г., Степашико В.С. Помехоустойчивость моделирования. – Киев: Наук. думка, 1985. – 216 с.
5. Єфіменко С.М. Комбінаторний алгоритм МГУА з послідовним ускладненням структур моделей на основі рекуррентно-паралельних обчислень // Індуктивне моделювання складних систем: Зб. наук. праць. – К.: МННЦ ІТ та С, 2014. – 6. – С. 64–71.

E-mail: syefim@ukr.net  
© С.Н. Ефименко, 2016

UDC 519.163 + 681.5.015

S.M. Yefimenko

## System Modeling and Prediction of the Multidimensional Interrelated Processes

**Introduction.** The problem of the mathematical modelling and prediction of the multidimensional interrelated time series is considered. It is used in economy, ecology and sociology. While many scientific proceedings are dedicated to modelling of one-dimensional time series, the experience of multidimensional time series modelling is insufficient.

**Methods.** An approach to the structural and parameters identification of the multidimensional time series is considered when parameters for every model is estimated independently. An algorithm with selecting of more than one best model for every process is used. The purpose is to combine all possible variants of system models and to select the best one by additional criterion.

**Results.** Theoretical grounds of recurrent-and-parallel computing in combinatorial GMDH algorithm and software for modeling and prediction of complex multidimensional interrelated processes in the class of vector autoregression models are developed.

**Conclusion.** The scheme of paralleling for recurrent COMBI algorithm allows to solve the problem when arguments amount exceeds capability of scheme with the exhaustive search. The effectiveness of the constructed algorithm is demonstrated by prediction of the interrelated processes in the field of investment activity of Ukraine with the purpose of information support of administrative decisions.