УДК 517.9

Построение информационных множеств в моделях системной динамики

© С. И. Спивак[1], О. Г. Кантор[2], Г. Н. Юсупова[3]

Аннотация. Приведен алгоритм распараллеливания процесса перебора параметров, определенных в многомерной области, обеспечивающих приемлемую точность моделей системной динамики, позволяющий формировать информационные множества модели.

Ключевые слова: уравнения системной динамики, информационное множество, параллельное программирование, алгоритм распараллеливания

В моделях системной динамики для всех переменных пишутся уравнения одного и того же типа [5]:

$$\frac{dx}{dy} = y^{+} - y^{-}, \tag{1.1}$$

где y^{+} и y^{-} - положительный и отрицательный темпы скорости переменной y, называемой системным уровнем. Каждая из величин y^{+} и y^{-} включает в себя все факторы, вызывающие соответственно рост и убывание y. Предполагается, что y^{+} и y^{-} являются функциями только системных уровней.

Таким образом, уравнения системной динамики представляют собой дифференциальные уравнения вполне определенной структуры, общий вид которых в случае исследования модели с тремя переменными следующий

$$\begin{cases} \frac{dx}{dt} = a_1 x^{\alpha_1} y^{\beta_1} z^{\gamma_1} - a_2 x^{\alpha_2} y^{\beta_2} z^{\gamma_2}, \\ \frac{dy}{dt} = a_3 x^{\alpha_3} y^{\beta_3} z^{\gamma_3} - a_4 x^{\alpha_4} y^{\beta_4} z^{\gamma_4}, \\ \frac{dz}{dt} = a_5 x^{\alpha_5} y^{\beta_5} z^{\gamma_5} - a_6 x^{\alpha_6} y^{\beta_6} z^{\gamma_6}. \end{cases} \tag{1.2}$$

Традиционным подходом при решении задачи построения модели (1.1) является определение такого набора значений параметров $\{a_j, \alpha_j, \beta_j, \gamma_j\}$, $j = \overline{1,6}$, который обеспечивал бы приемлемую точность модели. При этом под точностью модели обычно понимается близость расчетных и экспериментальных данных согласно какому-либо числовому критерию. Именно такая задача решалась авторами при моделировании численности населения Российской Федерации [2], [3], [4]. На основе использования комплекса методов численного и эконометрического моделирования в сочетании со специально разработанными программными продуктами были получены значения параметров модели системной динамики

$$\begin{cases} \frac{dN}{dt} = 8.139 \cdot 10^{-22} \cdot \frac{N^{2.05} \cdot D^{2}}{I^{2}} - 64.1 \cdot \frac{N^{0.33} \cdot D^{0.3}}{I^{0.3}}, \\ \frac{dD}{dt} = 560 \cdot D^{0.35} - 9900 \cdot I, \\ \frac{dI}{dt} = 0.131 \cdot I^{-0.4} - 0.0072 \cdot \frac{N^{0.092} \cdot D^{0.092}}{I^{0.092}}. \end{cases} \tag{1.3}$$

где N - численность населения РФ, чел.; D - душевые доходы за год, руб./чел. в год; I - индекс потребительских цен, доля ед. Информационную базу исследования составили данные официальной статистической отчетности за период с 2000 по 2009 гг.

[1] Заведующий кафедрой математического моделирования, Башкирский государственный университет, г. Уфа; s. spivak@bashnet.ru.

[2] Старший научный сотрудник института социально-экономических исследований, Уфимский научный центр Российской академии наук, г. Уфа; o_kantor@mail.ru

[3] Аспирант, Башкирский государственный университет, г. Уфа; gulnur0104@rambler.ru

Модель (1.3) с достаточно высокой точностью описывает экспериментальные данные, что подтверждается значениями показателей средних ошибок аппроксимации по каждому уравнению $\{\overline{A_N}, \overline{A_D}, \overline{A_I}\}$ (рис 1.1)

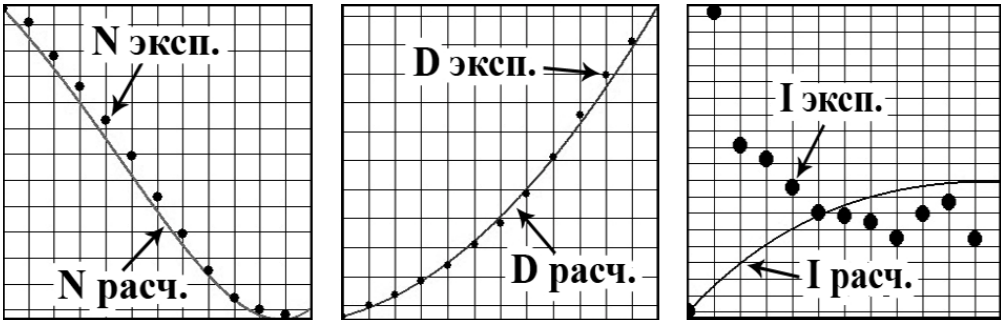


Рисунок **1.1**

Графическая иллюстрация результатов численного интегрирования системы (1.3) методом Рунге-Кутты

Отличительной особенностью примененного подхода являлось включение в модель ряда дополнительных условий, в том числе и на будущие значения переменных модели, вытекающих из смысла решаемой задачи, что не позволяло применять классические методы для определения параметров модели. Идеология реализованного подхода базировалась на использовании идеи Л.В. Канторовича, высказанной в работе [1]. Точность модели (1.3) в дальнейшем была подтверждена практикой (табл. 1).

Таблица 1: Сравнение прогнозных и фактических значений численности населения РФ, тыс. чел.

	На 1 янв. 2010 г.	На 1 янв. 2011 г.	На 1 янв. 2012 г.
По данным Федеральной службы государственной статистики	142833,0	142865,0	143056,0
Согласно модели системной динамики(3)	142025,2	142649,1	143793,4
Погрешность	807,8 (0,57%)	215,9(0,15%)	737,4(0,52%)

В целях обеспечения комплексного подхода к исследованию модели (1.3), а именно для реализации принципа множественности моделей, целесообразным является наряду с найденными точечными оценками параметров $\left\{a_j^0, \alpha_j^0, \beta_j^0, \gamma_j^0\right\}, j = \overline{1,6}$ определять множество альтернативных значений этих параметров, обеспечивающих соответствие модельных и экспериментальных данных согласно определенному критерию. Знание совокупности приемлемых значений параметров позволит исследователю не быть „привязанным“ к единственному виду модели и иметь дополнительную свободу при выборе значений ее параметров. Более того, на основе совокупности приемлемых значений параметров $\{a_j, \alpha_j, \beta_j, \gamma_j\}, j = \overline{1,6}$ могут проводиться расчеты, организуемые для проверки выдвигаемых гипотез относительно переменных модели (1.3).

Таким образом, актуальной является задача построения **информационного множества** для модели (1.3), под которым будем понимать совокупность значений параметров $\{a_j, \alpha_j, \beta_j, \gamma_j\}, j = \overline{1,6}$ принадлежащих некоторой окрестности их точечных оценок $\left\{a_j^0, \alpha_j^0, \beta_j^0, \gamma_j^0\right\}, j = \overline{1,6}$ и обеспечивающих приемлемое соответствие расчетных и экспериментальных данных. В качестве критерия, характеризующего близость расчетных и экспериментальных данных по каждому уравнению в отдельности, будем использовать средние ошибки аппроксимации $\{\overline{A_N}, \overline{A_D}, \overline{A_I}\}$.

Для определения информационного множества необходимо организовать вычислительную процедуру, цель которой заключается в формировании массива, содержащего значения параметров модели (1.2), удовлетворяющие заданным качественным характеристикам, на основе перебора значений параметров $\{a_j, \alpha_j, \beta_j, \gamma_j\}$, $j = \overline{1,6}$

Для достижения названной цели были поставлены и решены следующие задачи:

- построение алгоритма формирования наборов параметров, заданных в многомерной области, с требуемыми качественными характеристиками;

- построение алгоритма распараллеливания, заключающегося в разбиении перебора на несколько процессов, с возможностью идентификации набора параметров по номеру итерации;

- решение системы дифференциальных уравнений (1.2) для каждого из наборов параметров с обязательной проверкой на соответствие качественным критериям;

- решение системы дифференциальных уравнений (1.2) для каждого из наборов параметров с обязательной проверкой на соответствие качественным критериям;

Диапазоны вариации параметров модели (1.2) были взяты следующими:

$\alpha_1 \in$[0; 5],	$\beta_1 \in$[1,02; 5],	$\gamma_1 \in$[-5; 5]	$(a_1 - a_2) \in$[-14151439,74; 23,37],
$\alpha_2 \in$[0,11; 1,41],	$\beta_2 \in$[0; 1,03],	$\gamma_2 \in$[-2,06; 4,06]	$(a_3 - a_4) \in$ [7173,5; -3459,3],
$\alpha_3 \in$ [0; 0,13],	$\beta_3 \in$[0,32; 0,33],	$\gamma_3 \in$[-1,21; -1,14]	$(a_5 - a_6) \in$[1559,58; -686,59],
$\alpha_4 \in$[0; 2],	$\beta_4 \in$[0; 2],	$\gamma_4 \in$[-2; 2]	
$\alpha_5 \in$[0; 3],	$\beta_5 \in$[0; 2,99],	$\gamma_5 \in$[-1,99; 3]	
$\alpha_6 \in$[0; 3],	$\beta_6 \in$[0,01; 3],	$\gamma_6 \in$[-2; 3]	

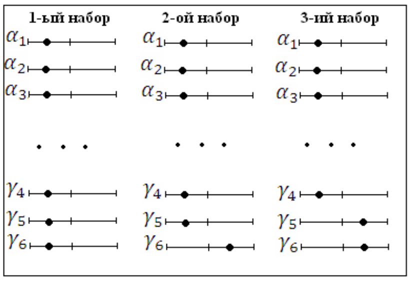
Как отмечалось ранее, в качестве критерия, характеризующего близость расчетных и экспериментальных данных по каждому уравнению в отдельности, будем использовать средние ошибки аппроксимации $\{\overline{A_N}, \overline{A_D}, \overline{A_I}\}$, приемлемые значения которых определим на уровне, не превышающем 10%.

Решение системы дифференциальных уравнений реализовано методом Рунге-Кутты. Эту часть программы назовем модулем решения системы.

Поскольку перебор параметров должен осуществляться в многомерной области (размерность в нашем случае равна 24), требуется очень большое количество итераций. Уже при рассмотрении только двух точек для каждого параметра из заданного для него интервала, возникает 2^{24} наборов параметров. Учитывая, что для каждого такого набора необходимо решить систему дифференциальных уравнений (1.2), получим, что модуль решения этой системы и соответствующий расчет погрешностей потребуется выполнить 2^{24} раз. Если осуществлять перебор напрямую, получится 23 вложенных цикла, в каждой итерации которого будет вызываться модуль решения системы. Очевидно, что такой способ решения поставленной задачи является очень трудоемким и затратным по времени.

Однако если этот перебор будут совершать несколько независимых процессов, время сократится в среднем пропорционально количеству независимых процессов. В связи с этим, актуальной является задача разбиения процесса перебора наборов параметров таким образом, чтобы каждый набор прогонялся (в том числе и через модуль решения системы) только в одном процессе. То есть необходимо упорядочить наборы таким образом, чтобы по их порядковому номеру однозначно определялся номер процесса, который будет проверять соответствие набора заданным условиям.

Поясним суть предлагаемого подхода для случая, когда интервалы задания параметров модели делятся на два отрезка и только одна точка из каждого отрезка участвует в построении наборов. Для удобства в качестве таких точек будем брать середины отрезков. На рисунке 1.2 дана иллюстрация этого случая (выделены точки, которые участвуют в построении наборов).



Р и с у н о к **1.2**

Графическая иллюстрация выбора точек из заданных интервалов при разбиении их на 2 равных отрезка

Очевидно, что в наборе по каждому параметру участвует либо точка из первого отрезка, либо из второго. Таким образом, в случае, когда точка выбирается из первого отрезка, ей может быть сопоставлена цифра 0, а если из второго - цифра 1. При таком подходе каждому набору параметров будет соответствовать набор чисел, состоящий из 0 и 1. Если интервалы, на которых определены параметры, будут делиться на 3 отрезка, в каждом из которых выбираются точки, то аналогично каждому набору параметров можно поставить в соответствие набор чисел, состоящий из 0, 1 и 2. Аналогичный подход может быть использован и при разбиении на большее число отрезков.

Таким образом, в числовом представлении набора участвуют только числа, используемые в системе счисления, соответствующей количеству отрезков, на которое производится разбиение интервалов.

Важным этапом является распараллеливание процесса перебора этих чисел так, чтобы одно и то же число (или, что то же самое, один и тот же набор) не проверялся несколькими процессами. Для непосредственной реализации распараллеливания была разработана специальная процедура. Суть ее заключается в следующем: внутри одной программы создаются несколько потоков, каждый из которых независимо от других выполняет свой перебор. Предположим, что программа разбивается на 5 процессов. Тогда задача сводится к тому, чтобы всю совокупность двоичных чисел (в случае разбиения интервалов на два отрезка), которые мы сопоставили набору точек, разбить на непересекающиеся множества, элементы каждого из которых будут проверяться внутри единственного процесса. При разбиении интервалов параметров модели на 2 отрезка возникает 2^{24} процессов. Переводя это число в двоичную систему счисления, получим 1 000 000 000 000 000 000 000 000. Как было сказано выше, 0 и 1 показывают из какого именно отрезка (первого или второго) необходимо взять точку. Поскольку в наборе участвуют 24 точки, значит, достаточно рассмотреть с конца 24 цифры. Все они равны 0, это означает, что данный набор параметров модели (2) формируется из точек, взятых только из первых отрезков соответствующих интервалов задания параметров. Набору параметров, все компоненты которого кроме последнего, взяты из первых отрезков, а последний компонент - из второго, будет соответствовать следующая последовательность нулей и единиц 000 000 000 000 000 000 000 001, что совпадает с десятичным представлением числа 1. Нетрудно заметить, что от 000 000 000 000 000 000 000 001 до 1 000 000 000 000 000 000 000 000 двоичные числа охватывают все возможные комбинации цифр 0 и 1.

Поэтому при делении интервалов задания параметров на 2 отрезка, если производить цикл от 1 до 2^{24} с переводом номера шага в двоичную систему, получается упорядоченный набор параметров всех возможных комбинаций. И по номеру итерации возможно однозначно определить, для какого набора осуществляется проверка на соответствие условию задачи.

В предлагаемой процедуре распараллеливания первый поток прогоняет цикл от 1 до целой части деления 2^{24} на 5 (поскольку номер итерации должен быть целым числом), то есть практически пятая часть общего количества наборов, аналогично второй поток будет считать следующие 20% вариантов и т.д. Внутри каждого потока номер итерации переводится в двоичную систему счисления и по каждому такому двоичному представлению однозначно устанавливается соответствие с набором параметров модели согласно описанному алгоритму.

Каждый поток вызывает модуль решения системы, осуществляющий решение системы методом Рунге-Кутты и подсчет средней ошибки аппроксимации. Если средняя ошибка аппроксимации по найденному набору менее 10%, то наборы параметров и соответствующее значение средней ошибка аппроксимации записываются в отдельный текстовый файл. Наименование файла представляет собой номер итерации (соответственно и порядковый номер набора) в двоичной системе, на которой был получен этот набор. Такая запись полученных данных обусловлена тем, что все потоки одновременно производят поиск и запись найденных файлов, а работа только с одним файлом может привести к ошибкам.

Поскольку при разбиении на два отрезка каждый параметр принимает лишь одно из двух значений, перебор получается очень грубым и вполне вероятно, что при таком подходе не удастся определить ни одного набора, удовлетворяющего критерию качества набора. Очевидно, что при разбиении интервалов значений параметров на большее число отрезков повышается точность «просмотра» многомерной области задания параметров. Однако при этом существенно возрастает необходимое количество итераций. Так, при разбиении указанных интервалов на 3 части количество итераций увеличивается примерно в 17 000 раз. Поэтому в такой ситуации целесообразным является использование принципа декомпозиции задачи, выражающегося в последовательном (возможно, более детальном) решении отдельных частей проблемы.

Описанный подход был апробирован при разбиении интервалов значений параметров только первого уравнения (кроме a_1 и a_2), с фиксацией значений остальных параметров на уровнях, соответствующих модели (1.3). Интервалы значений для шести варьируемых параметров разбивались на 10 равных частей, в качестве значений параметров $\{a_j, \alpha_j, \beta_j, \gamma_j\}$, $j = \overline{1,2}$ рассматривались центры интервалов. Общее количество всевозможных наборов составило 10^6 . Результаты применения описанного подхода позволили в качестве приемлемых вариантов значений варьируемых параметров модели (1.3) определить следующие (табл. 2).

Предложенный алгоритм распараллеливания процедуры перебора значений параметров модели (1.2), позволил определить 79 наборов значений параметров, удовлетворяющих введенным критериям качества модели, которые составляют часть информационного множества модели (1.3). Анализ этого подмножества показал, что изначально заданный интервал значений параметра α_2 при проведении дальнейших исследований может быть существенно сужен, а это, в свою очередь, позволит оптимизировать вычислительный процесс на следующих стадиях формирования информационного множества модели (1.3).

Таблица 2: Фрагмент информационного множества модели (1.3) по варьируемым параметрам $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2$

№	α_1	α_2	β_1	β_2	γ_1	γ_2
1	1,75	0,57	1,22	0,05	-4,5	-1,75
2	0,75	0,44	1,62	0,05	-3,5	-1,75
3	0,25	0,31	2,81	0,67	-3,5	-1,75
4	2,25	0,18	1,62	0,98	-3,5	-1,75
5	0,25	0,31	4,80	0,57	2,5	-1,75
6	0,75	0,18	3,21	0,67	4,5	3,14
7	2,75	0,18	1,22	0,05	-3,5	-1,14
8	1,75	0,18	1,62	0,36	3,5	3,75
9	0,25	0,31	4,40	0,36	3,5	3,75
10	4,25	0,18	2,81	0,36	4,5	3,75
Макс	**4,25**	**0,57**	**4,80**	**0,98**	**4,5**	**3,75**
Мин	**0,25**	**0,18**	**1,22**	**0,05**	**-4,5**	**-1,75**

Список литературы

1. Л. В. Канторович, "О некоторых новых подходах к вычислительным методам и обработке наблюдений", *Сибирский математический журнал*, **3**:5 (1962), 701–709.
2. С. И. Спивак,О. Г. Кантор, "Оценка параметров моделей системной динамики", *Журнал СВМО*, **13**:3 (2011), 107–113.
3. С. И. Спивак,О. Г. Кантор,И. Р. Салахов, "О программе, корректирующей систему уравнений", *Журнал СВМО*, **13**:4 (2011), 87–93.
4. С. И. Спивак,О. Г. Кантор, "Оценка качества спецификации моделей системной динамики", *Журнал СВМО*, **14**:2 (2012), 34–39.
5. Дж. Форрестер, *Мировая динамика*, Наука, М., 1978.

Construction of the information sets in models of system dynamics.

© **S. I. Spivak[4], O. G. Kantor[5], G. N. Yusupova[6]**

Abstract. An algorithm for parallelizing process of sorting parameters is defined in a multidimensional region, providing an acceptable accuracy of system dynamics models, allows you to shape the information sets model is given.

Key Words: equation of system dynamics, the information set, parallel programming, algorithm parallelizing

[4] Head of Mathematical modelling Chair, Bashkir State University, Ufa; s.spivak@bashnet.ru.
[5] Senior Research Scientist, Institute for Social and Economic Research, Ufa; o_kantor@mail.ru
[6] Postgraduate Student, Bashkir State University, Ufa; gulnur0104@rambler.ru