

# Идентификация моделей нестационарных распределенных процессов на основе многомерных временных рядов

М.Г. Матвеев<sup>а</sup>, А.В. Копытин<sup>а</sup>, Е.А. Сирота<sup>а</sup>, Е.А. Копытина<sup>а</sup>

<sup>а</sup> Воронежский государственный университет, 394018, Университетская площадь, 1, Воронеж, Россия

## Аннотация

Предложен метод идентификации уравнений математической физики, описывающих динамику пространственно-распределенных процессов, на основе экспериментальных многомерных временных рядов. Метод включает получение МНК-оценок параметров многомерной авторегрессии и построение вариантов систем алгебраических уравнений, связывающих оценки авторегрессии и параметры соответствующих дифференциальных уравнений. Система алгебраических уравнений, которой удовлетворяют полученные оценки, определяет структуру модели и соответствующие значения параметров дифференциального уравнения. Приводится численный пример идентификации процесса изменения температуры атмосферного воздуха.

*Ключевые слова:* уравнения в частных производных; структурная идентификация; параметрическая идентификация; многомерная авторегрессия; статистические гипотезы

## 1. Введение

Моделирование физических процессов обычно основывается на сочетании двух подходов: механистическом (аксиоматическом) и статистическом. Механистический подход предполагает, что механизмы функционирования процесса известны и могут быть отражены в виде соответствующих систем дифференциальных уравнений [1]. Статистический подход основан на построении эвристических моделей поведения процесса с использованием результатов наблюдений за переменными процесса. При этом широкое распространение получили линейные стохастические модели авторегрессии – скользящего среднего [2,3]. Если информации для принятия решений о точном механизме процесса недостаточно, т.е. допускается возможность построения вариантов (структурных и параметрических) систем дифференциальных уравнений, то дополнительное применение статистического подхода может в значительной степени уменьшить исходную неопределенность и повысить адекватность модели. Мы будем рассматривать широкий класс пространственно-распределенных динамических систем, для которых характерны диффузионные процессы, процессы адвекции или их сочетание. Источником информации о поведении системы являются данные измерений некоторой ее характеристики  $x_i^t$  в последовательные моменты времени  $t=0,1,\dots$  в узлах одномерной пространственной регулярной сетки  $i=0,1,\dots,n$ , т.е. многомерный временной ряд. Рассмотрение одномерной сетки ничем не ограничивает дальнейшие исследования, зато позволяет избежать громоздких построений, характерных для плоских и объемных пространств.

Задача заключается в разработке алгоритмов структурной и параметрической идентификации механистической модели с постоянными коэффициентами по наблюдаемым значениям  $x_i^t$ . Пусть в рассматриваемой системе могут протекать процессы диффузии и адвекции. Общий вид соответствующего уравнения:

$$\frac{\partial x}{\partial t} + v \frac{\partial x}{\partial l} = D \frac{\partial^2 x}{\partial l^2}, \quad (1)$$
$$x(0,l) = \varphi(l),$$

где  $v$  – скорость адвекции,  $D$  – коэффициент диффузии,  $l$  – пространственная координата.

Утверждать, что поведение системы определяется одним из указанных процессов или действуют оба процесса одновременно, нет достаточных оснований. Также неизвестны параметры (1), которые рассматриваются как константы, полученные в результате усреднения по пространству и времени. Необходимо получить информацию о структуре и параметрах рассматриваемой системы по результатам наблюдения  $x_i^t$ .

## 2. Разностные схемы для вариантов уравнений механистической модели

Для решения задачи составим явные трехточечные разностные схемы для уравнений каждого из вариантов структуры процессов:

диффузия и адвекция  $\frac{x_i^{t+1} - x_i^t}{\Delta t} + v \frac{x_{i+1}^t - x_{i-1}^t}{2\Delta l} = D \frac{x_{i+1}^t - 2x_i^t + x_{i-1}^t}{\Delta l^2}$ ,

$$x_i^{t+1} = (b_1 + b_2)x_{i-1}^t + (1 - 2b_2)x_i^t + (b_2 - b_1)x_{i+1}^t; \quad b_1 = \frac{v\Delta t}{2\Delta l}; \quad b_2 = \frac{D\Delta t}{\Delta l^2}; \quad (2)$$

диффузия в неподвижной среде  $\frac{x_i^{t+1} - x_i^t}{\Delta t} = D \frac{x_{i+1}^t - 2x_i^t + x_{i-1}^t}{\Delta l^2}$ ,

$$x_i^{t+1} = b_2 x_{i-1}^t + (1 - 2b_2)x_i^t + b_2 x_{i+1}^t; \quad (3)$$

$$\text{адвекция } \frac{x_i^{t+1} - x_i^t}{\Delta t} + \nu \frac{x_{i+1}^t - x_{i-1}^t}{2\Delta l} = 0,$$

$$x_i^{t+1} = b_1 x_{i-1}^t + x_i^t - b_1 x_{i+1}^t. \quad (4)$$

Для различия наблюдаемых и модельных значений переменной  $x$  введем следующие обозначения:  $x_i^t$  - уровень временного ряда, наблюдаемое значение;  $\hat{x}_i^t$  - значение, полученное в результате решения разностных уравнений (2) – (4);  $\tilde{x}_i^t$  - значение, полученное в результате моделирования временного ряда, которое понадобится нам в дальнейшем.

С учетом принятых обозначений представим разностные схемы (2) – (4) в виде системы рекуррентных зависимостей:

$$\hat{x}_{i-1}^{t+1} = \alpha_1 \hat{x}_{i-2}^t + \alpha_2 \hat{x}_{i-1}^t + \alpha_3 \hat{x}_i^t, \quad (5)$$

$$\hat{x}_i^{t+1} = \alpha_1 \hat{x}_{i-1}^t + \alpha_2 \hat{x}_i^t + \alpha_3 \hat{x}_{i+1}^t, \quad (6)$$

$$\hat{x}_{i+1}^{t+1} = \alpha_1 \hat{x}_i^t + \alpha_2 \hat{x}_{i+1}^t + \alpha_3 \hat{x}_{i+2}^t, \quad (7)$$

$$i = 2, \dots, n-2; t = 0, 1, \dots,$$

где  $\alpha_k$  - коэффициенты, вид которых определяется вариантом структуры процессов.

Наличие наблюдаемого многомерного временного ряда позволяет сформулировать начальные и граничные условия, соответственно

$$\hat{x}_i^0 = x_i^0, \quad i = 0, \dots, n,$$

$$\hat{x}_{i-2}^t = x_{i-2}^t, \quad \hat{x}_{i+2}^t = x_{i+2}^t, \quad t = 0, 1, \dots; i = 2, \dots, n-2.$$

Заметим, что в соответствии со свойством консервативности [4], сумма коэффициентов правой части выражений (2) – (4) равна единице.

С учетом граничных условий матричное представление системы (5) – (7) будет иметь вид

$$\hat{x}^{t+1} = A \hat{x}^t, \quad (8)$$

где  $\hat{x}^{t+1} = (x_{i-2}^t, \hat{x}_{i-1}^{t+1}, \hat{x}_i^{t+1}, \hat{x}_{i+1}^{t+1}, x_{i+2}^t)^T$ ,  $\hat{x}^t = (x_{i-2}^t, \hat{x}_{i-1}^t, \hat{x}_i^t, \hat{x}_{i+1}^t, x_{i+2}^t)^T$ . Матрица  $A$  имеет вид

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 \\ 0 & \alpha_1 & \alpha_2 & \alpha_3 & 0 \\ 0 & 0 & \alpha_1 & \alpha_2 & \alpha_3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \sum_{k=1}^3 \alpha_k = 1.$$

### 3. Построение модели многомерного временного ряда и ее сравнение с разностной схемой

Моделирование многомерного временного ряда будем осуществлять в классе линейных стохастических моделей авторегрессии [3]. Допустим, что в каждом узле регулярной сетки протекает марковский процесс без последствия и временные ряды в смежных узлах имеют высокие значения коэффициентов линейной корреляции. Такие допущения позволяют специфицировать стохастическую модель в виде

$$\tilde{x}^{t+1} = A \tilde{x}^t + e^{t+1}, \quad (9)$$

где  $\tilde{x}^{t+1} = (x_{i-2}^t, \tilde{x}_{i-1}^{t+1}, \tilde{x}_i^{t+1}, \tilde{x}_{i+1}^{t+1}, x_{i+2}^t)^T$ ,  $\tilde{x}^t = (x_{i-2}^t, \tilde{x}_{i-1}^t, \tilde{x}_i^t, \tilde{x}_{i+1}^t, x_{i+2}^t)^T$ ,  $e^{t+1}$  - векторный стационарный случайный процесс типа «белый шум». Матрица  $A$  имеет тот же вид, что и в уравнении (8).

Получение МНК-оценок элементов  $\alpha$  матрицы  $A$  хорошо изучено [3,5,6,7] и позволяет получать состоятельные и несмещенные оценки. Полученные оценки позволяют записать модель поведения переменной  $x$  в произвольном узле  $i$  в виде авторегрессии, т.е. условного математического ожидания

$$\tilde{x}_i^{t+1} = M(\tilde{x}_i^{t+1} | x_{i-1}^t, x_i^t, x_{i+1}^t) = a_1 x_{i-1}^t + a_2 x_i^t + a_3 x_{i+1}^t; \quad (10)$$

$$i = 2, \dots, n-1; t = 0, 1, \dots$$

Вычисления по выражениям (8) и (9) существенно различаются, несмотря на их кажущееся сходство. Выражение (8) представляет собой рекуррентную формулу для вычисления модельных значений переменной  $x$  в трех попарно смежных узлах сетки, с условной устойчивостью, а, следовательно, и сходимостью к решению соответствующего дифференциального уравнения. Выражение (9) позволяет вычислять модельные значения той же переменной в тех же узлах, но автономно для каждого узла в каждый момент времени как это определяется формулой (10). Различие в вычислительных алгоритмах (5) – (7) и (10) затрудняет сравнение результатов вычислений. Однако, для  $t=0$  такое сравнение произвести несложно. Запишем выражения (6) и (10) для нулевого момента времени

$$\hat{x}_i^1 = \alpha_1 x_{i-1}^0 + \alpha_2 x_i^0 + \alpha_3 x_{i+1}^0, \quad (11)$$

$$\tilde{x}_i^1 = a_1 x_{i-1}^0 + a_2 x_i^0 + a_3 x_{i+1}^0. \quad (12)$$

Очевидно, что результаты вычислений в приведенных формулах будут равны, если будут равны коэффициенты при одинаковых переменных в правой части. Поскольку мы допустили неизменность параметров модели по пространству и времени, то сравнение (11) и (12) с учетом (2) – (4) позволяет получить три варианта систем алгебраических уравнений относительно параметров разностных схем (2) – (4):

для диффузии и адвекции

$$\begin{cases} b_1 + b_2 = a_1, \\ 1 - 2b_2 = a_2, \\ b_2 - b_1 = a_3; \end{cases} \quad (13)$$

для диффузии

$$\begin{cases} b_2 = a_1, \\ 1 - 2b_2 = a_2, \\ b_2 = a_3; \end{cases} \quad (14)$$

для адвекции

$$\begin{cases} b_1 = a_1, \\ 1 = a_2, \\ -b_1 = a_3. \end{cases} \quad (15)$$

В полученных уравнениях в правой части везде стоят оценки, то есть значения случайных величин. Следовательно, решение каждого из уравнений следует проводить как проверку соответствующей статистической гипотезы в предположении нормального распределения оценок. Решение с проверкой статистической гипотезы будет показано в примере, приведенном в следующем разделе.

#### 4. Пример реализации предложенной методики

Диффузия и адвекция являются характерными процессами, определяющими изменение состояния земной атмосферы. В частности именно эти процессы определяют динамику температурных полей на заданных изобарических поверхностях [8,9]. Актуальной задачей метеорологии является определение структуры и параметров атмосферных процессов. Покажем, что решение этой задачи возможно по результатам регулярных наблюдений за температурой атмосферного воздуха над заданным участком земной поверхности.

Для экспериментальной апробации использовались статистические данные реанализа параметров атмосферы за 2012 год [10]. Данные представляют собой ежедневные значения температуры в узлах плоской регулярной сетки с шагом  $2,5^\circ$ . Рассматривались результаты наблюдений температуры при геопотенциале 300 ГПа в узле сетки с координатами  $35^\circ$  северной широты;  $7,5^\circ$  восточной долготы.

По указанным данным была построена одномерная модель множественной авторегрессии

$$\tilde{x}_{7,5}^{t+1} = 2,27x_5^t - 3,03x_{7,5}^t + 1,76x_{10}^t, \quad (16)$$

где  $t = 0, 1, \dots$ . Заметим, что сумма коэффициентов в выражении (16) равна единице, что можно рассматривать как признак того, что временной ряд отражает динамику одного из процессов, описываемых системой (8).

Полученные значения параметров лучше всего соответствуют процессу диффузии. Покажем это. В соответствии с системой (14) должно выполняться равенство  $a_1 = a_3$ . Учитывая, что  $a_1$  и  $a_3$  являются выборочными оценками, т.е. случайными величинами со стандартными ошибками  $\sigma_{a_1} = 0,28$  и  $\sigma_{a_3} = 0,27$ , примем нулевую гипотезу равенства нулю разности этих оценок. Расчетное значение  $t$ -статистики случайной величины  $a_1 - a_3$  составит 0,89. Критическое значение  $t$ -статистики приближенно равняется 2. Следовательно, нулевая гипотеза не отвергается. В этом случае найдем значение  $b_2$ , вычислив оценку математического ожидания:  $b_2 = (2,27 + 1,76) / 2 = 2,02$ . Полученное значение можно проверить, подставив его во второе уравнение системы (14). Результат подстановки дает  $-3,04 = -3,03$ , что со статистической точки зрения обеспечивает тождественность.

Для того чтобы убедиться в адекватности результатов структурной и параметрической идентификации было найдено численное решение уравнения диффузии по неявной разностной схеме, поскольку в данном случае условие устойчивости явной схемы не выполняется. Уровни наблюдаемого временного ряда температур, результаты моделирования с использованием авторегрессии (16) и решение уравнения диффузии с использованием неявной разностной схемы с найденным значением параметра  $b_2 = 2,02$  представлены на рис. 1.

Экспериментальным подтверждением правильности предложенной методики должно быть хорошее совпадение наблюдаемых значений температуры, значений, полученных по авторегрессионной модели (16), и численного решения уравнения диффузии. Качество авторегрессионной модели и модели диффузии характеризуется коэффициентом детерминации  $R^2$ . В обоих случаях этот коэффициент равен 0,78, что говорит об одинаковых качественных характеристиках моделей. Степень близости решений, полученных по этим моделям можно оценить коэффициентом парной корреляции, который составил в нашем случае 0,97, что еще раз подтверждает практическую эквивалентность найденных решений.

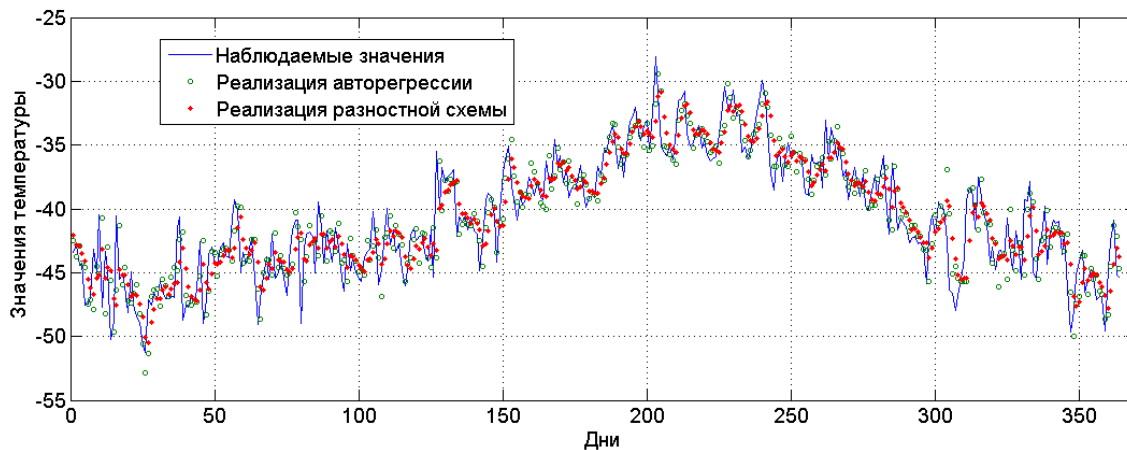


Рис. 1. Временной ряд и результаты моделирования.

## 5. Заключение

Для идентификации динамических моделей с распределенными переменными обычно используются либо непараметрические методы идентификации, например, на основе результатов активного эксперимента [11], который не всегда возможен, либо параметрические методы с аппроксимацией производных [12]. В последнем случае возникает опасность неадекватного приближения производных в условиях помех [13].

Предложенная методика идентификации основывается на сопоставительном анализе параметров авторегрессионной модели и разностного уравнения. Полученные в работе результаты показывают возможность использования такого подхода при моделировании зашумленных динамических объектов с распределенными переменными и обосновывают проведение дальнейших исследований в этом направлении.

## Литература

- [1] Захаров, Е.В. Уравнения математической физики: Университетский учебник /Е.В. Захаров, И.В. Дмитриева С.И. Орлик. – М.: Издательский центр «Академия», 2010. – 320 с.
- [2] Dickey, D.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root / D.A. Dickey and W.A. Fuller // Journal of the American Statistical Association. – 1979. – Vol. 74(366). – P. 427–431. DOI: 10.2307/2286348.
- [3] Носко, В.Н. Эконометрика. Кн. 2. Ч. 3, 4: Учебник / В.Н. Носко. – М.: Изд. дом «Дело» РАНХиГС, 2011. – 576 с.
- [4] Мареев, В.В. Основы методов конечных разностей / В.В. Мареев, Е.Н. Станкова. – СПб.: Изд-во С.-Петерб. ун-та, 2012. – 64 с.
- [5] Clements, M.P. Forecasting Non-stationary Economic Time series / M.P. Clements, D. F. Hendry. – Cambridge, Massachusetts, London: MIT Press, 1999. – 262 p.
- [6] Patterson, K. An Introduction to Applied Econometrics: A Time Series Approach /K. Patterson. – New York: Palgrave, 2000. – 832 p.
- [7] Конторович, Г.Г. Анализ временных рядов / Г.Г. Конторович // Экономический журнал ВШЭ. – 2002. – №3. – С. 379-401.
- [8] Хромов, С.П. Метеорология и Климатология. Учебник / С.П. Хромов, М.А. Петросянец. – М.: Изд. МГУ, 2001. – 527 с.
- [9] Матвеев, М.Г. Комбинированная прогностическая модель нестационарного многомерного временного ряда для построения пространственного профиля атмосферной температуры / М.Г. Матвеев, В.В. Михайлов, Е.А. Сирота // Информационные технологии. –2016. – Т. 22, № 2. – С. 89–94.
- [10] NCEP/DOE AMIP II Reanalysis [Электронный ресурс]. – Режим доступа: <http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis2.html> (03.01.2014).
- [11] Бойков, И.В. Методы идентификации динамических систем / И.В. Бойков, Н.П. Кривулин // Программные системы: теория и приложения. – 2014. – Т. 23, №5. – С. 79–96.
- [12] Bar, M. Fitting differential equations to space-time dynamics / M. Bar, R. Hegger, H. Kantz // Physical Review, E. – 1999. Vol. 59(1). – P. 337-342. DOI: 10.1103/PhysRevE.59.337.
- [13] Xun, X. Parameter estimation of partial differential equation models / X. Xun, J. Cao, B. Mallick, R. Carroll, J. Maity // Journal of the American Statistical Association. – 2013. – Vol. 108(503). – P. 1009–1020. DOI: 10.1080/01621459.2013.794730.