УДК 004.942

АНАЛИЗ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ АВТОРЕГРЕССИОННЫХ МОДЕЛЕЙ ДЛЯ ИДЕНТИФИКАЦИИ РАЗНОСТНЫХ УРАВНЕНИЙ НЕСТАЦИОНАРНЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМ

М. Г. Матвеев*, Е. А. Сирота*, И. В. Абрамов*, В. В. Синюков**

**Воронежский государственный университет*
***ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина» (г. Воронеж)*

Поступила в редакцию 13.11.2017 г.

Аннотация. В работе проводится анализ возможности применения авторегрессионных моделей для идентификации разностных уравнений нестационарных распределенных систем. Предлагаемый метод основан на сравнительном анализе конечно-разностных представлений исследуемой модели и авторегрессионного описания временных рядов наблюдений. В рамках данной работы показано, что даже при незначительном зашумлении экспериментальных данных достоверная идентификация параметров модели в рамках рассматриваемых методов описания динамических систем является неудовлетворительной, полученные МНК-оценки параметров являются смещенными. При этом может наблюдаться хорошее совпадение экспериментальных и модельных данных, но разница между истинными значениями и оценкой может быть достаточно большой.
Ключевые слова: авторегрессионная модель, конечно-разностные уравнения, идентификация, МНК-оценки, смещенные оценки параметров модели.
Annotation. The author analyzes the possibility of using autoregressive models to identify the difference equations of nonstationary distributed systems. The proposed method is based on a comparative analysis of finite-difference representations of the model under study and an autoregressive description of time series of observations. In the framework of this paper it is shown that even with a slight noisy experimental data, reliable identification of model parameters within the framework of the described methods for describing dynamical systems is unsatisfactory, the derived OLS estimates of the parameters are biased. In this case, good agreement between the experimental and model data can be observed, but the difference between the true values and the estimate can be quite large.
Keywords: autoregressive model, finite difference equations, identification, OLS estimates, biased estimates of model parameters.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Проблема достоверной идентификации параметров математической модели по экспериментальным данным является важнейшей проблемой в различных областях науки. В работах [1, 2] показано, что одним из эффективных методов параметрической идентификации динамических систем является метод, в основе которого лежит разработка и исследование линейно-параметрических дискретных моделей, описывающих в форме разностных схем многомерные динамические процессы. В работе представлено решение задачи для двумерного динамического процесса. Однако достаточное совпадение модельных и экспериментальных зависимостей еще не свидетельствует о действительно достоверной идентификации параметров в рассматриваемой модели.

В рамках данной работы показано, что даже при незначительном зашумлении экспериментальных данных достоверная идентификация параметров модели в рамках рассматриваемых методов описания динамических систем является неудовлетворительной, полученные МНК-оценки параметров являются смещенными. При этом может наблюдаться хорошее совпадение экспериментальных и модельных данных, но разница между истинными значениями и оценкой может быть достаточно большой.

Итак, рассмотрим подробно саму идею методики идентификации.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ. РАЗНОСТНЫЕ СХЕМЫ И ИХ СОПОСТАВЛЕНИЕ С МОДЕЛЬЮ МНОГОМЕРНОГО ВРЕМЕННОГО РЯДА

Будем рассматривать широкий класс пространственно-распределенных динамических систем, для которых характерны диффузионные процессы, процессы адвекции или их сочетание. Соответствующее дифференциальное уравнение в частных производных с начальными и граничными условиями имеет следующий общий вид:

$$\frac{\partial y}{\partial t} + \vartheta \frac{\partial y}{\partial l} = D \frac{\partial^2 y}{\partial l^2}, \qquad (1)$$

$$y(0,l) = \varphi(l), \;\; y(t,l^{\min}) = f_1(t), \;\; y(t,l^{\max}) = f_1(t),$$

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

Источником информации о поведении системы являются данные натурных измерений переменной y_i^t с погрешностью ε_i^t в виде «белого шума» – $x_i^t = y_i^t + \varepsilon_i^t$ в последовательные моменты времени $t = 0,1,\ldots$ в узлах одномерной пространственной регулярной сетки $i = 0,1,\ldots,n,$ т. е. многомерный временной ряд. Рассмотрение одномерной сетки ничем не ограничивает дальнейшие исследования, зато позволяет избежать громоздких построений, характерных для плоских и объемных пространств.

Пусть в рассматриваемой системе могут протекать процессы диффузии и адвекции. 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 времени.

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

диффузия и адвекция

$$\frac{y_i^{t+1} - y_i^t}{\Delta t} + \vartheta \frac{y_{i+1}^t - y_{i-1}^t}{2\Delta l} = D \frac{y_{i+1}^t - 2y_i^t + y_{i-1}^t}{\Delta l^2}, \qquad (2)$$

$$y_i^{t+1} = (b_1 + b_2) y_{i-1}^t + (1 - 2b_2) y_i^t + (b_2 - b_1) y_{i+1}^t;$$

$$b_1 = \frac{\vartheta \Delta t}{2\Delta l}; \;\; b_2 = \frac{D\Delta t}{\Delta l^2};$$

диффузия в неподвижной среде

$$\frac{y_i^{t+1} - y_i^t}{\Delta t} = D \frac{y_{i+1}^t - 2y_i^t + y_{i-1}^t}{\Delta l^2};$$

$$y_i^{t+1} = b_2 y_{i-1}^t + (1 - 2b_2) y_i^t + b_2 y_{i+1}^t; \qquad (3)$$

адвекция

$$\frac{y_i^{t+1} - y_i^t}{\Delta t} + \vartheta \frac{y_{i+1}^t - y_{i-1}^t}{2\Delta l} = 0, \qquad (4)$$

$$y_i^{t+1} = b_1 y_{i-1}^t + y_i^t - b_1 y_{i+1}^t.$$

Представим разностные схемы (2)–(4) в виде обобщенной рекуррентной зависимости для произвольного узла i с начальными и граничными условиями:

$$y_i^{t+1} = \alpha_1 y_{i-1}^t + \alpha_2 y_i^t + \alpha_3 y_{i+1}^t \qquad (5)$$

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

$$y^0 = \left(y_{i-1}^0; y_i^0; y_{i+1}^0\right),$$

$$y_{i-1} = \left(y_{i-1}^0; y_{i-1}^1; \ldots; y_{i-1}^k\right),$$

$$y_{i+1} = \left(y_{i+1}^0; y_{i+1}^1; \ldots y_{i+1}^k\right),$$

где α_i – коэффициенты, вид которых определяется вариантом структуры процессов.

Начальные и граничные условия задаются как результаты натурных измерений, поэтому

$$y^0 = x^0 = \left(x_{i-1}^0; x_i^0; x_{i+1}^0\right),$$

$$y_{i-1} = x_{i-1} = \left(x_{i-1}^0; x_{i-1}^1; \ldots; x_{i-1}^k\right),$$

$$y_{i+1} = x_{i+1} = \left(x_{i+1}^0; x_{i+1}^1; \ldots x_{i+1}^k\right).$$

Заметим, что в соответствии со свойством консервативности [3] сумма коэффициентов правой части выражений (2)–(4) равна единице. Соответственно $\sum_{j=1}^{3} \alpha_j = 1.$

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

$$\tilde{x}_i^{t+1} = M\left(x_i^{t+1} / x_{i-1}^t, x_i^t, x_{i+1}^t\right) = a_1 x_{i-1}^t + a_2 x_i^t + a_3 x_{i+1}^t; \quad i = 2,\dots,n-2;\ t = 0,1,\dots, \tag{6}$$

где a_1, a_2, a_3 – оценки параметров авторегрессии.

Вычисления по выражениям (5) и (6) существенно различаются, несмотря на их кажущееся сходство. Выражение (5) представляет собой рекуррентную формулу для вычисления модельных значений переменной y в трех попарно смежных узлах сетки, с условной устойчивостью, следовательно, и сходимостью к решению соответствующего дифференциального уравнения. Выражение (6) позволяет вычислять модельные значения случайной переменной x в тех же узлах, но автономно для каждого узла в каждый момент времени, как это определяется формулой (6). Различие в вычислительных алгоритмах (5) и (6) затрудняет сравнение результатов вычислений. В частности, для целей проводимого исследования важно показать, что выражение (6) отличается от классической авторегрессии на величину смещения, что в свою очередь порождает смещение МНК-оценки параметров модели.

Смещение МНК-оценок подробно описано в [4]. Применимо к нашему случаю проделаем ряд известных преобразований, чтобы показать смещенность оценок параметров исследуемой модели.

Перепишем модель (5) в виде

$$y_i^{k+1} = \left(y^k\right)^T \alpha \tag{7}$$

Запишем выражение (7) с учетом вида наблюдаемых значений

$$x_i^{k+1} = \left(x^k\right)^T \alpha - \left(\varepsilon^k\right)^T \alpha + \varepsilon_i^{k+1}, \tag{8}$$

где $\left(\varepsilon^k\right)^T = \left(\varepsilon_{i-1}^k, \varepsilon_i^k, \varepsilon_{i+1}^k\right)$.

Вычислим скалярное произведение $\left(\varepsilon^k\right)^T \alpha$:

$$\left(\varepsilon^k\right)^T \alpha = \tilde{\varepsilon}_i^{k+1}. \tag{9}$$

Для перехода к уравнению авторегрессии найдем условное математическое ожидание (8):

$$M\left(x_i^{k+1} / X\right) = \left(x^k\right)^T \alpha - M\left(\tilde{\varepsilon}_i^{k+1} / X\right) + M\left(\varepsilon_i^k / X\right) \tag{10}$$

или с учетом $M\left(\varepsilon_i^k / X\right) = M\left(\varepsilon_i^{k+1}\right) = 0$

$$\varepsilon_i^k \tilde{x}_i^{k+1} = \left(x^k\right)^T \alpha - M\left(\tilde{\varepsilon}_i^{k+1} / X\right), \tag{11}$$

где $X = X_{k\times 3}$ – матрица результатов наблюдений в узлах $i-1$, i, $i+1$, формируемая в моменты времени от 0 до k.

Выражение (11) отличается от классической авторегрессии на величину $M\left(\tilde{\varepsilon}_i^{k+1} / X\right)$, что является следствием зависимости помехи от объясняющих регрессоров. Это порождают смещение МНК-оценки параметров α.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Диффузия и адвекция являются характерными процессами, определяющими изменение состояния земной атмосферы. В частности, именно эти процессы определяют динамику температурных полей на заданных изобарических поверхностях [5, 6]. 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 [7]. Данные представляют собой ежедневные значения температуры в узлах плоской регулярной сетки с шагом 2,5°. Рассматривались результаты наблюдений температуры при геопотенциале 300 ГПа в узле сетки с координатами 35° северной широты; 7,5° восточной долготы.

По указанным данным была построена одномерная модель множественной авторегрессии

$$\tilde{x}_{7,5}^{t+1} = 2{,}27x_5^t - 3{,}03x_{7,5}^t + 1{,}76x_{10}^t, \qquad (12)$$

где $t = 0,1,\ldots$

В работах [8, 9] было показано, что полученные значения параметров лучше всего соответствуют процессу диффузии. Численное решение уравнения диффузии было получено по зашумленной неявной разностной схеме (условие устойчивости явной схемы не выполняется).

Уровни наблюдаемого временного ряда температур, результаты моделирования с использованием авторегрессии (12) и решение уравнения диффузии с использованием неявной разностной схемы представлены на рис. 1.

Визуальный анализ демонстрирует хорошее совпадение модельных и экспериментальных зависимостей. Тем не менее, сравнительный анализ параметров модели одномерной множественной авторегресии (12) и найденных мнк-оценок параметров зашумленной разностной схемы, представленный в табл. 1, показывает сильное расхождение полученных результатов.

ЗАКЛЮЧЕНИЕ

Таким образом, в рамках данной работы показано, что даже незначительное зашумление экспериментальных данных не позволяет проводить достоверную идентификацию параметров модели в рамках рассматриваемых методов описания динамических систем, при этом наблюдается хорошее совпадение модельных и экспериментальных зависимостей, однако полученные МНК-оценки являются смещенными, а значения параметров модели существенно различаются.

СПИСОК ЛИТЕРАТУРЫ

1. *Зотеев В. Е.* Параметрическая идентификация диссипативных механических систем на основе разностных уравнений / В. Е. Зотеев; под ред. В. П. Радченко. – М. : Машиностроение, 2009. – 344 с.

2. *Заусаева М. А.* Определение параметров двумерных динамических процессов на основе разностных схем / М. А. Заусаева, В. Е. Зотеев // Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки. – 2010. – № 1(20). – С. 154–161.

Таблица 1

Сравнительный анализ параметров модели одномерной множественной авторегресии и найденных мнк-оценок параметров зашумленной разностной схемы

	α_1	α_2	α_3
Параметры модели одномерной множественной авторегресии	2.27	–3.03	1.76
Мнк-оценки параметров зашумленной разностной схемы	0.82	0.12	0.05

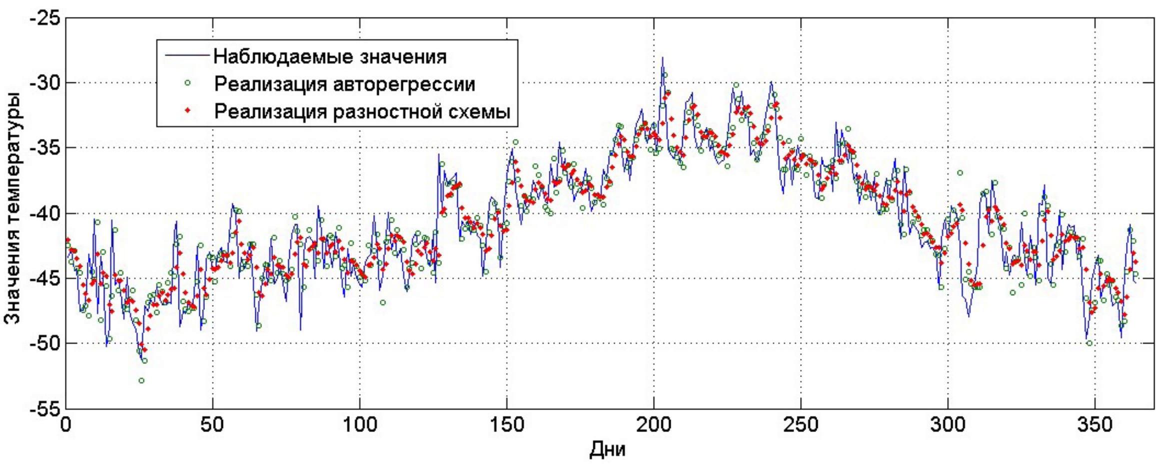

Рис. 1. Результаты работы иерархической нейронной сети

3. *Мареев В. В.* Основы методов конечных разностей / В. В. Мареев, Е. Н. Станкова. – СПб. : Изд-во С.-Петерб. ун-та, 2012. – 64 с.

4. *Носко В. Н.* Эконометрика. Кн. 2. Ч. 3, 4: Учебник / В. Н. Носко. – М. : Изд. дом «Дело» РАНХиГС, 2011. – 576 с.

5. *Хромов С. П.* Метеорология и Климатология. Учебник / С. П. Хромов, М. А. Петросянц. – М. : Изд. МГУ, 2001. – 527 с.

6. *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.

7. NCEP/DOE AMIP II Reanalysis [Электронный ресурс]. – Режим доступа: http:// www. cdc. noaa.gov/cdc/data.ncep.reanalysis2. html (03.01.2014).

8. *Матвеев М. Г.* Идентификация моделей нестационарных распределенных процессов на основе многомерных временных рядов / М. Г. Матвеев, А. В. Копытин, Е. А. Сирота, Е. А. Копытина // Информационные технологии и нанотехнологии (ИТНТ-2017). Сборник трудов III международной конференции и молодежной школы. Самарский национальный исследовательский университет имени академика С. П. Королева. – 2017. – С. 1059–1062.

9. *Матвеев М. Г.* Верификация процесса конвективной диффузии на основе анализа многомерных временных рядов / М. Г. Матвеев, Е. А. Сирота, М. Е. Семенов, А. В. Копытин // Аналитика и управления данными в областях с интенсивным использованием данных. Сборник научных трудов XIX Международной конференции DAMDID / RCDL'2017. – 10–13 октября 2017 г. г. Москва, МГУ. – С. 433–437.

Матвеев М. Г. – д-р техн. наук, профессор, зав. кафедрой информационных технологий управления, факультет компьютерных наук, Воронежский государственный университет.
E-mail: mgmatveev@yandex.ru

Сирота Е. А. – канд. физ.-мат. наук, доцент кафедры цифровых технологий, факультет компьютерных наук, Воронежский государственный университет.
E-mail: sirota_ea@sc.vsu.ru

Абрамов И. В. – канд. техн. наук, доцент кафедры информационных технологий управления, факультет компьютерных наук, Воронежский государственный университет.
E-mail: iva-dak.vrn@mail.ru

Синюков В. В. – канд. техн. наук, ВУНЦ ВВС «ВВА имени профессора Н. Е. Жуковского и Ю. А. Гагарина».
E-mail: sinukovhome@mail.ru

Matveev M. G. – doctor of technical sciences, professor department of programming and information technologies, Voronezh State University.
E-mail: mgmatveev@yandex.ru

Sirota E. A. – candidate of physics-math. sciences, assistant professor, the dept. of digital technologies Faculty of Computer Science, Voronezh State University.
E-mail: sirota_ea@sc.vsu.ru

Abramov I. V. – candidate of technical sciences, assistant professor, the dept. of programming and information technologies, Voronezh State University.
E-mail: iva-dak.vrn@mail.ru

Sinyukov V. V. – candidate of technical sciences, VUNTS VVS «VVA behalf of Professor N. E. Zhukovsky and Yu. A. Gagarin».
E-mail: sinukovhome@mail.ru