

Верификация процесса конвективной диффузии на основе анализа многомерных временных рядов

© М.Г. Матвеев © Е.А. Сирота © М.Е. Семенов © А.В. Копытин

Воронежский государственный университет,
Воронеж, Россия

mgmatveev@yandex.ru

atoris@list.ru

Аннотация. Предложен метод верификации процесса конвективной диффузии на основе анализа многомерных временных рядов. Он основан на сравнительном анализе конечно-разностных представлений исследуемой модели и авторегрессионного описания временных рядов наблюдений. 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. Приведен численный пример идентификации процесса изменения температуры атмосферного воздуха.

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

Verification of the Convective Diffusion Process Based on the Analysis of Multidimensional Time Series

© M.G. Matveev © E.A. Sirota © M.E. Semenov © A.V. Kopytin

Voronezh State University,
Voronezh, Russia

mgmatveev@yandex.ru

atoris@list.ru

Abstract. A method for verifying the process of convective diffusion based on the analysis of multidimensional time series is proposed. 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. The method includes the LSM (Least square method) estimates of the parameters of multidimensional autoregression and the construction of versions of systems of algebraic equations connecting the estimates of autoregression and the parameters of the corresponding differential equations. The system of algebraic equations satisfied by the obtained estimates determines the structure of the model and the corresponding values of the parameters of differential equation. A numerical example of identifying the process of changing the temperature of atmospheric air is given.

Keywords: partial differential equations, structural identification, parametric identification, multidimensional autoregression, statistical hypothesis.

1 Введение

Математические модели физических процессов основаны на априорном предположении о механизме функционирования этих процессов [1]. Однако такие предположения не всегда в достаточной степени обоснованы. Ошибки спецификации модели возникают как на структурном, так и на параметрическом уровнях. Эффективным инструментом проверки предполагаемых

механизмов процессов могут служить методы анализа временных рядов, образующихся при наблюдении за параметрами функционирования этих процессов [2, 3]. Предлагаемый метод основан на сравнительном анализе конечно-разностных представлений исследуемой модели процесса и авторегрессионного описания временных рядов наблюдений.

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:

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

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

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 $t = 0, 1, \dots$ 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.

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.

Задача заключается в верификации процессов конвективной диффузии на основе анализа многомерных временных рядов и разработке алгоритмов структурной и параметрической идентификации механистической модели с постоянными коэффициентами по наблюдаемым значениям 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:

$$\text{диффузия и адвекция}$$

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

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

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

диффузия в неподвижной среде

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

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

адвекция

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

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

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

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

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

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

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

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

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

Начальные и граничные условия задаются как результаты натурных измерений, поэтому

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

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

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

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

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 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

$$\tilde{x}_i^{t+1} = M \left(\frac{x_i^{t+1}}{x_{i-1}^t}, x_i^t, x_{i+1}^t \right) = \alpha_1 x_{i-1}^t + \alpha_2 x_i^t + \alpha_3 x_{i+1}^t;$$

$$i = 2, \dots, n - 2; \quad t = 0, 1, \dots, \quad (6)$$

где $\alpha_1, \alpha_2, \alpha_3$ – 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 математическое ожидание оценок α в авторегрессии (6) равно параметрам α в разностной схеме (5).

Найдем математические ожидания левой и правой части авторегрессионной зависимости, учитывая, что

математическое ожидание условного математического ожидания случайной величины есть математическое ожидание этой случайной величины:

$$M(x_i^{t+1}) = M(a_1)M(x_{i-1}^t) + M(a_2)M(x_i^t) + M(a_3)M(x_{i+1}^t).$$

Полученное выражение можно переписать с учетом граничных условий и очевидного условия

$$M(x_i^{t+1}) = M(y_i^{t+1}) + M(\varepsilon_i^{t+1}) = y_i^{t+1}$$

в виде

$$y_i^{t+1} = M(a_1)y_{i-1}^t + M(a_2)y_i^t + M(a_3)y_{i+1}^t,$$

что доказывает равенство математических ожиданий оценок параметров авторегрессии и параметров α разностной схемы (5). Полученный результат позволяет приравнять оценки параметров авторегрессии и параметров α и из полученных уравнений найти параметры b_1 и b_2 .

Трем вариантом сочетания процессов будут соответствовать три варианта систем алгебраических уравнений относительно параметров b_1 и b_2 разностных схем (2)–(4):

для диффузии и адвекции

$$\begin{cases} b_1 + b_2 = a_1, \\ 1 - 2b_2 = a_2, \\ b_2 - b_1 = a_3; \end{cases} \quad (7)$$

для диффузии

$$\begin{cases} b_2 = a_1, \\ 1 - 2b_2 = a_2, \\ b_2 = a_3; \end{cases} \quad (8)$$

для адвекции

$$\begin{cases} b_1 = a_1, \\ 1 = a_2, \\ -b_1 = a_3. \end{cases} \quad (9)$$

В полученных уравнениях в правой части всюду стоят оценки, т. е. значения случайных величин. Следовательно, решение каждого из уравнений следует проводить как проверку соответствующей статистической гипотезы в предположении нормального распределения оценок. Вариант системы уравнений, согласующийся с соответствующей гипотезой, определит адекватный процесс. Решение с проверкой статистической гипотезы 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 крайней мере, для среднегодовых значений, 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]. Эти 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° северной широты; $7,5^\circ$ восточной долготы. По указанным данным была построена одномерная модель множественной авторегрессии

$$\tilde{x}_{7,5}^{t+1} = 1,03x_5^t - 1,14x_{7,5}^t + 1,11x_{10}^t x_{10}^t, \quad (10)$$

где $t = 0, 1, \dots$. Заметим, что сумма коэффициентов в выражении (10) равна единице, что можно рассматривать как признак того, что временной ряд адекватно отражает динамику температуры, описанную разностной схемой (5). Для оценки процессов, оказывающих доминирующее влияние на динамику температуры, проведем статистический анализ систем (7)–(9).

Стандартные ошибки оценок равны $\sigma_{a_1} = 0,14$; $\sigma_{a_2} = 0,25$; $\sigma_{a_3} = 0,13$. Подставим полученные оценки a в систему (7), что соответствует принятию гипотезы о совместном влиянии на температурную динамику процессов диффузии и адвекции. Допустим, что ошибки оценок a равномерно распределены между коэффициентами b . Тогда последние два уравнения системы (7) позволяют вычислить оценки $b_2 = 1,07$ со стандартной ошибкой $\sigma_{b_2} = 0,07$ и $b_1 = -0,04$ со стандартной ошибкой $\sigma_{b_1} = 0,125$. Последнее равенство означает, что b_1 можно считать равным нулю. Поскольку коэффициент b_1 отвечает за адвекцию (см. раздел 2), нулевая гипотеза о совместном влиянии процессов диффузии и адвекции отклоняется.

Подставим полученные оценки в систему (8), что соответствует принятию нулевой гипотезы о влиянии диффузии. В этом случае нулевая гипотеза может быть принята, если подтвердятся гипотезы о равенстве a_1 и a_2 и тождественности второго равенства системы (8). Рассмотрим гипотезу о равенстве нулю разности a_1 и a_2 . Эта разность составляет 0,08 при стандартной ошибке a_1 и $\sigma_{a_2 - a_1} = 0,13 + 0,14 = 0,27$, что дает основание принять гипотезу о равенстве и улучшить оценку путем вычисления среднего значения $b_2 = (1,03 + 1,11)/2 = 1,07$. Для проверки гипотезы о тождественности второго равенства системы (8) подставим в него полученные оценки a_2 и b_2 и получим $-1,14 = -1,14$, что при стандартной ошибке $\sigma_{a_2} = 0,25$ позволяет принять гипотезу о тождественности, следовательно, и нулевую гипотезу о влиянии на динамику температуры процесса диффузии.

Подстановка оценок a в систему (9) для процесса адвекции позволяет сразу отклонить гипотезу о

влиянии на динамику температуры процесса адвекции из-за наличия противоречивых знаков оценок.

Окончательный вывод: на динамику температуры атмосферы в рассматриваемой точке в среднем в течение года доминирующее влияние оказывали процессы атмосферной диффузии.

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 (10) и решение уравнения диффузии с использованием неявной разностной схемы с найденным значением параметра $b_2 = 2.02$ представлены на рис. 1.

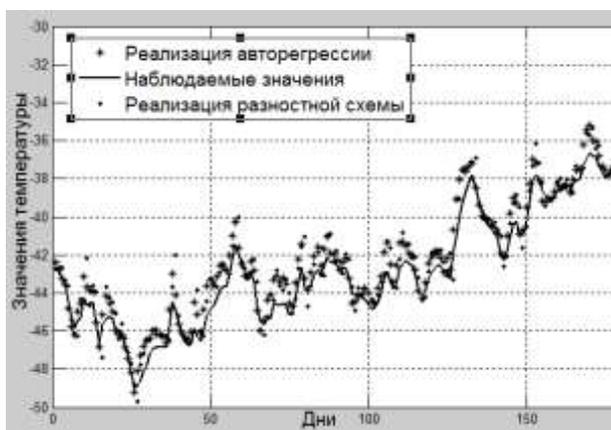


Рисунок 1 Временной ряд и результаты моделирования

Экспериментальным подтверждением правильности предложенной методики должно быть хорошее совпадение наблюдаемых значений температуры, значений, полученных по авторегрессионной модели (10), 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.

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]. Различные статистические методы были разработаны для оценки параметров моделей, описываемых обыкновенными дифференциальными уравнениями. Так, например, в [14–16] предложены иерархические байесовские подходы к этой проблеме. Эти методы требуют неоднократного численного решения соответствующих обыкновенных дифференциальных уравнений, что в ряде случаев требует значительного временного ресурса. Для оценки постоянных параметров модели в [17] предложен обобщенный сглаживающий подход, основанный на идеях метода максимального правдоподобия. В работе [18] рассматривается каскадный метод оценки изменяющихся по времени параметров. Этот метод оценивает параметры путем оптимизации определенного критерия, однако достижение глобального минимума проблематично с вычислительной точки зрения.

Другим подходом к оцениванию параметров обыкновенных дифференциальных уравнений (см. [19]) является двухэтапный метод, в котором на первом этапе, с использованием сглаживающих методов, по зашумленным данным оцениваются функция и ее производные, а затем на втором этапе строятся МНК-оценки параметров уравнения. Двухэтапный метод легко реализуется, однако он может быть статистически неэффективным, так как производные не могут быть точно оценены по зашумленным данным, особенно производные старших порядков.

Предложенная методика верификации основана на сопоставительном анализе параметров авторегрессионной модели и разностного уравнения. Полученные 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., Дмитриева И.В., Орлик С.И.: Уравнения математической физики. Университетский учебник. М.: Издательский центр «Академия», 320 с. (2010)
- [2] Dickey, D.A., Fuller, W.A.: Distribution of the Estimators for Autoregressive Time Series with a Unit Root. J. of the American Statistical Association, 74 (366), pp. 427-431 (1979). doi: 10.2307/2286348
- [3] Носко, В.Н.: Эконометрика. Кн. 2. Ч. 3, 4. Учебник. М.: Изд. дом «Дело» РАНХиГС, 576 с. (2011)
- [4] Мареев, В.В., Станкова, Е.Н.: Основы методов конечных разностей. СПб.: Изд-во С.-Петерб. ун-та, 64 с. (2012)
- [5] Clements, M.P., Hendry, D.F.: Forecasting Non-stationary Economic Time series. Cambridge, Massachusetts, London: MIT Press, 262 p. (1999)

- [6] Patterson, K.: An Introduction to Applied Econometrics: A Time Series Approach. New York: Palgrave, 832 p. (2000)
- [7] Канторович, Г.Г.: Анализ временных рядов. Экономический журнал ВШЭ. (3), сс. 379-401 (2003)
- [8] Хромов, С.П., Петросянец, М.А.: Метеорология и Климатология. Учебник. М.: Изд-во МГУ, 527 с. (2001)
- [9] Матвеев, М.Г., Михайлов, В.В., Сирота Е.А.: Комбинированная прогностическая модель нестационарного многомерного временного ряда для построения пространственного профиля атмосферной температуры. Информационные технологии, 22 (2), сс. 89-94 (2016)
- [10] NCEP/DOEAMIP2Reanalysis [Электронный ресурс]. <http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis2.html>
- [11] Бойков, И.В., Кривулин, Н.П.: Методы идентификации динамических систем. Программные системы: теория и приложения. 23 (5), С. 79-96 (2011)
- [12] Bar, M., Hegger, R, Kantz, H.: Fitting Differential Equations to Space-time Dynamics. Physical Review, E. 59 (1), pp.337-342 (1999). doi: 10.1103/PhysRevE.59.337
- [13] Xun, X., Cao, J., Maity, J.: Parameter Estimation of Partial Differential Equation Models. J. of the American Statistical Association. 108 (503), pp.1009-1020 (2013). doi: 10.1080/01621459.2013.794730
- [14] Putter, H., Lange S.: A Bayesian Approach to Parameter Estimation in HIV Dynamical Models. Statistics in Medicine Heisterkamp, pp. 2199-2214 (2000)
- [15] Huang, Y., Liu, D.: Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System, pp. 413-423 (2006)
- [16] Ramsay, J. O., Hooker, G., Campbell, D.: Parameter Estimation for Differential Equations: a Generalized Smoothing Approach (with discussion). J. of the Royal Statistical Society, Series B, pp. 741-796 (2007)
- [17] Cao, J., Huang, J., Wu, H.: Penalized Nonlinear Least Squares Estimation of Time-varying Parameters in Ordinary Differential Equations. J. of Computational and Graphical Statistics, pp. 42-56 (2012)
- [18] Liang, H., Wu, H.: Parameter Estimation for Differential Equation Models Using a Framework of Measurement Error in Regression Models. J. of the American Statistical Association, pp. 1570-1583 (2008)
- [19] Chen, J., Wu, H.: Efficient Local Estimation for Time-varying Coefficients in Deterministic Dynamic Models with Applications to HIV-1 Dynamics. J. of the American Statistical Association, pp. 369-384 (2008)