

МЕТОДИКА СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ СИСТЕМЫ ВРЕМЕННЫХ РЯДОВ

© 2013 Ю.Е. Кувайскова

Ульяновский государственный технический университет

Поступила в редакцию 26.09.2013

Описывается методика структурно-параметрической идентификации системы взаимосвязанных временных рядов. Методика основана на подходе адаптивного динамического регрессионного моделирования, предусматривающего при построении математических моделей временных рядов последовательную адаптацию модели к возможным нарушениям основных предположений регрессионного анализа. При этом построенная модель системы временных рядов адекватна реальной ситуации и позволяет прогнозировать с большей точностью будущее состояние процесса.

Ключевые слова: система временных рядов, адаптивное динамическое регрессионное моделирование, прогнозирование.

Одной из важных задач во многих отраслях промышленности является задача прогнозирования будущего состояния технического объекта. В большинстве случаев состояние объектов в различных технических приложениях характеризуется набором значений параметров, фиксируемых через определенные моменты времени и образующих систему временных рядов. При этом часто наблюдается высокая степень корреляционной зависимости между временными рядами значений параметров технического объекта. При моделировании динамики временных рядов нередко сталкиваются с такими нарушениями схем регрессионного анализа, как высокая степень автокорреляционной зависимости между последующими и предшествующими членами временного ряда, ненормальность распределения остатков и другие. В связи с этим возникает задача комплексной обработки системы взаимосвязанных временных рядов с обязательной проверкой соблюдения предположений нормальной схемы Гаусса-Маркова и последующей адаптацией к их нарушениям.

Для решения данной задачи разработана методика структурно-параметрической идентификации системы взаимосвязанных временных рядов на основе методологии адаптивного динамического регрессионного моделирования [1, 2]. Подход адаптивного динамического регрессионного моделирования представляет собой реализацию метода многоэтапной структурно-параметрической идентификации. На каждом этапе проводятся построение и анализ соответствующей компоненты модели временного ряда, оценка точности аппроксимации и прогнозирования,

Кувайскова Юлия Евгеньевна, кандидат технических наук, доцент кафедры «Прикладная математика и информатика». E-mail: u.kuvaiskova@mail.ru

диагностика свойств остатков и при необходимости – адаптация к нарушениям предположений о характеристиках этих свойств.

Рассмотрим технический объект, состояние которого характеризуется параметрами, значения которых регистрируются через определенные промежутки времени и образуют систему временных рядов, которую можно записать в виде $y(t) = \{y(t_1), y(t_2), \dots, y(t_n)\}$, где параметр t указывает на момент времени, в который зафиксировано значение, или номер наблюдения.

Пусть $t = 1, 2, \dots, n$ (т.е. количество наблюдений равно n), рассмотрим N временных рядов $y^1(t), y^2(t), \dots, y^N(t)$.

В соответствии с подходом адаптивного динамического регрессионного моделирования одномерный временной ряд $y^i(t)$, наблюдаемый в равноотстоящие моменты времени t_1, t_2, \dots, t_n представляется в виде:

$$y^i(t) = f^i(t) + g^i(t) + \psi^i(t) + \varepsilon^i(t), \quad (1)$$

при этом функция $f^i(t)$ – неслучайная (долговременная) функция тренда соответствующего ряда; $g^i(t)$ – неслучайная периодическая функция – совместная гармоническая составляющая ряда; $\psi^i(t)$ – случайная с элементами регулярности функция – векторная авторегрессия; $\varepsilon^i(t)$ – нерегулярная компонента (случайная величина, ошибка).

На первом этапе для системы временных рядов выполняется разведочный анализ (фрактальный, мультифрактальный) для выявления степени регулярности каждого ряда системы [3].

При выявлении заметной трендоустойчивости для соответствующих рядов выделяется функция тренда $f(t)$, которую обычно приближают полиномом достаточно низкой степени, также могут быть использованы и различные нелиней-

ные по оцениваемым параметрам выражения. Оптимальная функция трендовой составляющей ищется по критерию минимума внешнего среднеквадратического отклонения (СКО) σ_{Δ} .

Точность аппроксимации σ (внутреннее СКО) рассчитывается по формуле:

$$\sigma = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - p)}, \quad (2)$$

где n – количество наблюдений, p – число слагаемых в модели, y_i – результат i -го наблюдения, \hat{y}_i – прогнозируемое значение по построенной модели.

Для вычисления внешних мер (внешнего СКО σ_{Δ} и др.) [1] исходная выборка данных делится на две части – обучающую и контрольную. По обучающей выборке строится модель, по контрольной – внешние меры, характеризующие прогностические свойства модели.

Точность прогнозирования σ_{Δ} вычисляется по формуле:

$$\sigma_{\Delta} = \sqrt{\sum_{i=1}^k (\Delta_i - \bar{\Delta})^2 / (k - 1)}, \quad (3)$$

где $\Delta_i = y_i - \hat{y}_i$, $i = \overline{1, k}$ (k – количество элементов контрольной выборки),

$$\bar{\Delta} = \frac{\sum_{i=1}^k \Delta_i}{k}, \quad (4)$$

y_i – наблюдаемое значение отклика на контрольном интервале, \hat{y}_i – его прогноз – значения, вычисляемые по комплексной модели.

При этом для оценивания применяется линейный или нелинейный метод наименьших квадратов. Полученная модель для одного ряда может быть представлена в виде:

$$y^i(t) = f^i(t) + \varepsilon_trend^i(t), \quad (5)$$

где $y^i(t)$ – фактические значения i -го ряда; $f^i(t)$ – функция трендовой составляющей для i -го ряда, $\varepsilon_trend^i(t)$ – случайная величина, характеризующая отклонения реального значения ряда от теоретического, найденного по уравнению регрессии.

Далее из каждого ряда системы выделяются остатки $\varepsilon_trend^i(t)$:

$$\varepsilon_trend^i(t) = y^i(t) - f^i(t). \quad (6)$$

Дальнейшее сглаживание временных рядов производится методами гармонического анализа, основанного на использовании формул Фурье.

Для каждого временного ряда методом пошаговой регрессии находятся значимые гармоники. Затем последовательно ищутся совместные амплитуды и фазы только для тех временных рядов,

у которых на данном периоде T_k имеются значимые гармоники.

Совместную гармоническую составляющую можно представить в виде:

$$g^i(t) = \sum_{k=1}^{n/2} A_k \cdot \sin\left(\frac{2\pi t}{T_k} + \varphi_k\right), \quad (7)$$

где $t = 1, 2, \dots, n$, $A_k := \frac{\sum_{j=1}^N A_{j,k}}{T_k}$, $\varphi_k = \varphi_{2k} - \varphi_{1k}$.

Коэффициенты

$$A_{jk}, \varphi_{jk} \quad (j = 1, 2, \dots, N, k = 1, 2, \dots, n)$$

находим для каждого из N временных рядов по формулам:

$$\begin{cases} A_{jk} = \sqrt{w_{jk}^2 + v_{jk}^2} \\ \varphi_{jk} = \arctg\left(\frac{v_{jk}}{w_{jk}}\right), \end{cases} \quad (8)$$

$$w_{jk} = \sum_{d=1}^n \varepsilon_trend_{jd} \cdot \sin\left(\frac{2\pi d}{T_k}\right), \quad (9)$$

$$v_{jk} = \sum_{d=1}^n \varepsilon_trend_{jd} \cdot \cos\left(\frac{2\pi d}{T_k}\right), \quad (10)$$

$$T_k = \frac{n}{k}.$$

Далее находятся остатки для каждого временного ряда:

$$\varepsilon_g^i(t) = \varepsilon_trend^i(t) - g^i(t). \quad (11)$$

После выделения регулярных составляющих целью анализа системы временных рядов является моделирование остатков $\varepsilon_g^i(t)$ случайной с элементами регулярности функцией $\psi(t)$. Функция $\psi(t)$ для системы взаимосвязанных временных рядов может быть представлена в виде модели векторной авторегрессии.

Модель векторной авторегрессии – это система уравнений, каждое из которых представляет собой модель авторегрессии и распределенного лага. Пусть $x^i(t)$ $i = 1..N$ - i -й временной ряд. Модель векторной авторегрессии порядка p – модель для i -го временного ряда будет иметь вид

$$\begin{aligned} x^i(t) = & a_0^i + \sum_{j=1}^N a_{1j}^i x^j(t-1) + \\ & + \sum_{j=1}^N a_{2j}^i x^j(t-2) + \dots + \\ & + \sum_{j=1}^N a_{pj}^i x^j(t-p) + \varepsilon^i(t) \end{aligned} \quad (12)$$

Матричная запись модели векторной авторегрессии порядка p для системы временных рядов $X(t) = \{x_1(t), x_2(t), \dots, x_N(t)\}$ запишется в виде

$$X(t) = a_0 + A_1 X(t-1) + A_2 X(t-2) + \dots + A_p X(t-p) + \varepsilon(t) = a_0 + \sum_{m=1}^p A_m X(t-m) + \varepsilon(t) \quad (13)$$

где A_m – матрицы элементов a_{mj}^i .

Коэффициенты уравнений модели можно получить методом наименьших квадратов, пользуясь следующей формулой:

$$A = (X^T X)^{-1} X^T Y. \quad (14)$$

Здесь X – это матрица значений рядов, представленная в виде:

$$\begin{pmatrix} 1 & x^1(t-1) & \dots & x^1(t-p) & \dots & x^N(t-1) & \dots & x^N(t-p) \\ 1 & x^1(t-1) & \dots & x^1(t-p) & \dots & x^N(t-1) & \dots & x^N(t-p) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x^1(t-1) & \dots & x^1(t-p) & \dots & x^N(t-1) & \dots & x^N(t-p) \end{pmatrix} \quad (15)$$

единичный столбец соответствует нулевому коэффициенту.

Y – вектор последних значений всех рядов: $(x^1(t) \dots x^N(t))^T$. (16)

В результате операций над матрицами, получаем матрицу коэффициентов A , имеющую $(1+N \cdot p)$ строк и N столбцов.

$$\begin{pmatrix} a_0^1 & \dots & a_0^N \\ \vdots & \ddots & \vdots \\ a_{N \cdot p}^1 & \dots & a_{N \cdot p}^N \end{pmatrix}. \quad (17)$$

Каждый столбец здесь соответствует коэффициентам для i -ого уравнения модели.

Остатки $\varepsilon^i - g^i(t)$ после вычитания из исходного ряда регулярной составляющей представляются в виде модели векторной авторегрессии:

$$\begin{aligned} \varepsilon - g^i(t) &= a_0^i + \sum_{j=1}^N a_{1j}^i y^j(t-1) + \\ &+ \sum_{j=1}^N a_{2j}^i y^j(t-2) + \dots + \\ &+ \sum_{j=1}^N a_{pj}^i y^j(t-p) + \\ &+ \sum_{k=1}^p a_{ki}^i \varepsilon - g^i(t-k) + \varepsilon^i(t) \end{aligned} \quad (18)$$

Оптимальная модель векторной авторегрессии (18) ищется по критерию минимума внешнего СКО σ_Δ , рассчитанного по формуле (3). Далее из каждого ряда системы выделяются остатки:

$$\varepsilon^i(t) = \varepsilon - g^i(t) - \left(a_0^i + \sum_{j=1}^N a_{1j}^i y^j(t-1) + \sum_{j=1}^N a_{2j}^i y^j(t-2) + \dots + \sum_{j=1}^N a_{pj}^i y^j(t-p) + \sum_{k=1}^p a_{ki}^i \varepsilon - g^i(t-k) \right). \quad (19)$$

В итоге получаем для каждого из N временных рядов комплексную модель следующего вида:

$$\begin{aligned} y^i(t) &= f^i(t) + g^i(t) + a_0^i + \\ &+ \sum_{j=1}^N a_{1j}^i y^j(t-1) + \sum_{j=1}^N a_{2j}^i y^j(t-2) + \dots + \\ &+ \sum_{j=1}^N a_{pj}^i y^j(t-p) + \sum_{k=1}^p a_{ki}^i \varepsilon - g^i(t-k) + \varepsilon^i(t) \end{aligned} \quad (20)$$

Для построенных моделей по формулам (2) и (3) подсчитывается точность аппроксимации σ и прогнозирования σ_Δ .

После структурно-параметрической идентификации системы временных рядов проверяется соблюдение условий применения “регрессионного анализа-метода наименьших квадратов”. Если основные предположения соблюдаются, построенная комплексная модель системы временных рядов может быть использована для прогнозирования.

При несоблюдении предположений проводится соответствующая процедура адаптации к нарушению данного предположения [1].

Ниже представлены результаты анализа, моделирования и прогнозирования временных рядов вибраций гидроагрегата в режиме работы в сети. Наблюдения в режиме работы агрегата в сети длились 1200 секунд, показания снимались 2 раза в секунду, всего получено 2400 наблюдений по каждому из восьми каналов. После усреднения каждые 10 секунд получилось 120 наблюдений.

Применяя методику структурно-параметрической идентификации системы временных рядов для моделирования измерений вибраций по восьми каналам, получены следующие модели:

$$\begin{aligned} y^1(t) &= 641946 + 0,025 \cdot t + 1,022 \cdot \sin(-26,68 \cdot t + 65,093) + \\ &+ 0,001 \cdot y^2(t-1) - 0,001 \cdot y^3(t-1) - 0,002 \cdot y^4(t-1) + \\ &+ 0,001 \cdot y^5(t-1) + 0,0004 \cdot y^6(t-1) - 0,002 \cdot y^7(t-1) + \varepsilon^1(t), \end{aligned}$$

$$\begin{aligned} y^2(t) &= 698,017 - 0,003 \cdot t + 1,022 \cdot \sin(12,493 \cdot t - 22,256) - \\ &- 0,0002 \cdot y^1(t-1) - 0,006 \cdot y^3(t-1) + 0,007 \cdot y^4(t-1) - \\ &- 0,004 \cdot y^5(t-1) - 0,001 \cdot y^6(t-1) + \varepsilon^2(t), \end{aligned}$$

$$\begin{aligned} y^3(t) &= 225,93 + 0,997 \cdot \sin(0,0001 \cdot t + 11,456) - \\ &- 0,317 \cdot y^1(t-1) + 0,002 \cdot y^4(t-1) + 0,0001 \cdot y^5(t-1) - \\ &- 0,001 \cdot y^6(t-1) + 0,011 \cdot y^7(t-1) + 0,005 \cdot y^8(t-1) + \varepsilon^3(t), \end{aligned}$$

$$\begin{aligned} y^4(t) &= 44752 - 0,109 \cdot t + 1,021 \cdot \sin(236654 \cdot t + 170546) + \\ &+ 0,001 \cdot y^1(t-1) + 0,0002 \cdot y^2(t-1) + 0,008 \cdot y^3(t-1) - \\ &- 0,001 \cdot y^5(t-1) + 0,002 \cdot y^6(t-1) - 0,014 \cdot y^7(t-1) - \\ &- 0,007 \cdot y^8(t-1) + \varepsilon^4(t), \end{aligned}$$

$$y^5(t) = 29,174 + 0,008 \cdot t + 1,023 \cdot \sin(-1,299 \cdot t + 0,041) - 0,001 \cdot y^1(t-1) - 0,0001 \cdot y^2(t-1) + 0,0001 \cdot y^3(t-1) - 0,0003 \cdot y^6(t-1) + 0,0002 \cdot y^8(t-1) + \varepsilon^5(t),$$

$$y^6(t) = 3,943 \cdot t + 1,014 \cdot \sin(-0,191 \cdot t + 2,518) + 0,109 \cdot y^1(t-1) - 0,001 \cdot y^2(t-1) - 0,199 \cdot y^3(t-1) - 0,024 \cdot y^4(t-1) - 0,002 \cdot y^5(t-1) - 0,129 \cdot y^7(t-1) + \varepsilon^6(t),$$

$$y^7(t) = 26,049 + 1,001 \cdot \sin(0,040 \cdot t - 0,011) - 0,001 \cdot y^1(t-1) + 0,001 \cdot y^3(t-1) + 0,0002 \cdot y^4(t-1) - 0,001 \cdot y^5(t-1) - 0,001 \cdot y^6(t-1) + 0,004 \cdot y^8(t-1) + \varepsilon^7(t),$$

$$y^8(t) = 32,155 + 1,022 \cdot \sin(-3197,496 \cdot t + 734,684) - 0,0004 \cdot y^1(t-1) + 0,016 \cdot y^3(t-1) - 0,008 \cdot y^4(t-1) + 0,0003 \cdot y^5(t-1) + 0,0004 \cdot y^6(t-1) - 0,004 \cdot y^7(t-1) + \varepsilon^8(t).$$

На рис. 1 а показаны графики исходных данных (сплошная линия) и аппроксимации по соответствующей модели (штриховая линия), а также графики (Рис. 1 б) предсказанных (штриховая линия) и исходных (сплошная линия) значений временных рядов для данных вибраций по первому каналу.

Для сравнительного анализа эффективности разработанной методики моделирования системы временных рядов использовалась методика построения моделей авторегрессии проинтегрированного скользящего среднего (АРПСС), реализованная в пакете Statistica [4].

Построенные модели АРПСС в пакете Statistica, представлены ниже.

$$y^1(t) = 643,6562 + 0,2151y^1(t-1),$$

$$y^2(t) = 697,9956 - 0,3888 \cdot \varepsilon^2(t-1) - 0,1814 \cdot \varepsilon^2(t-2),$$

$$y^3(t) = 21,16052 + 0,51052 \cdot y^3(t-1),$$

$$y^4(t) = 39,69158 - 0,10275 \cdot \varepsilon^4(t-1) + 0,20859 \cdot \varepsilon^4(t-2),$$

$$y^5(t) = 28,76911 - 0,20671 \cdot \varepsilon^5(t-1) + 0,12513 \cdot \varepsilon^5(t-2),$$

$$y^6(t) = -42,3635 + 0,8185 \cdot y^6(t-1) + 0,1815 \cdot y^6(t-2) - 0,0973 \cdot \varepsilon^6(t-1) + 0,5629 \cdot \varepsilon^6(t-2),$$

$$y^7(t) = 25,28663 + 0,96942 \cdot y^7(t-1) + 0,61375 \cdot \varepsilon^7(t-1) + 0,22012 \cdot \varepsilon^7(t-2),$$

$$y^8(t) = 31,82976 - 0,06093 \cdot \varepsilon^8(t-1).$$

Результаты сравнения по точности прогнозирования (внешнее СКО) двух методик для системы временных рядов, полученных по восьми каналам измерения вибраций в режиме работы в сети гидроагрегата, приведены в табл. 1.

Таким образом, совместное описание характеристик технического объекта с применением методики многоэтапной структурно-параметрической идентификации с адаптацией к нарушениям основных предположений регрессионного анализа повышает точность прогнозирования по сравнению с точностью при использовании одномерных методов (методика АРПСС), что позволяет принимать с определенным запасом времени управленческие решения, направленные на предупреждение опасной ситуации.

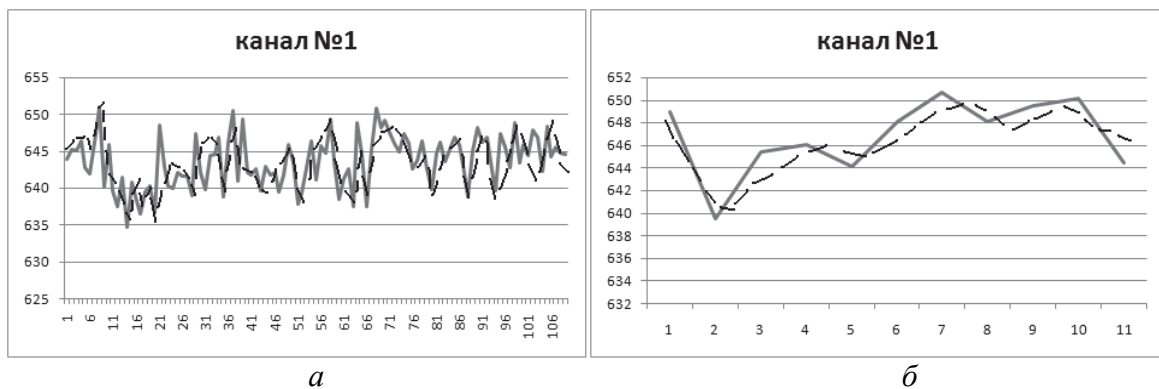


Рис. 1. Моделирование (а) и прогнозирование временного ряда (б)

Таблица 1. Оценка точности прогнозирования для системы временных рядов при использовании двух методик

Методика	Номер канала	Среднеквадратическое отклонение прогнозирования (σ_{Δ})							
		1	2	3	4	5	6	7	8
Методика АРПСС (Statistica)		4,99	3,32	2,68	15,91	8,21	878,55	6,01	2,72
Методика структурно-параметрической идентификации системы временных рядов		3,56	2,57	2,25	14,17	6,12	667,124	3,60	2,20

Исследование выполнено при поддержке Министерства образования и науки Российской Федерации, соглашение 14.В37.21.0672 (федеральная целевая программа "Научные и научно-педагогические кадры инновационной России").

СПИСОК ЛИТЕРАТУРЫ

1. Валеев С. Г. Регрессионное моделирование при обработке наблюдений. М.: Наука, 1991. 272 с.
2. Валеев С. Г., Кувайскова Ю. Е. Программное обеспечение обработки временных рядов техногенных характеристик // Обозрение прикладной и промышленной математики. 2009. Т. 16. Вып. 6. С. 1037-1038.
3. Валеев С. Г., Кувайскова Ю. Е., Губайдуллина С. А. Применение мультифрактального анализа при описании временных рядов в технике и экономике // Вестник Ульяновского государственного технического университета. 2008. №. 2. С. 23-27.
4. Боровиков В. П., Ивченко Г. И. Прогнозирование в системе STATISTICA в среде Windows. М.: Финансы и статистика, 1999. 384 с.

METHOD OF STRUCTURAL PARAMETRIC IDENTIFICATION OF TIME SERIES SYSTEM

© 2013 Ju.E. Kuvayskova

Ulyanovsk State Technical University

Method of structural parametric identification of relative time series system is described. The method based on procedure of adaptive dynamic regression modeling, providing at construction of mathematical models of time series consecutive adaptation of model to possible infringements of the basic assumptions of regression analysis. Thus constructed model of time series system is adequate to a real situation and allows predicting a process condition with greater exactness

Keywords: time series system, adaptive dynamic regression modeling, predicting.