

УДК 681.518.5

МАТЕМАТИЧЕСКИЕ И ИНФОРМАЦИОННО-СТРУКТУРНЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ ТЕХНИЧЕСКИ СЛОЖНЫХ ОБЪЕКТОВ

Статья поступила в редакцию 03.09.2015 г., в окончательном варианте 6.11.2015 г.

Михеев Михаил Юрьевич, доктор технических наук, профессор, Пензенский государственный технологический университет, 440039, Российская Федерация, г. Пенза, проезд Байдукова / ул. Гагарина, 1а / 11, e-mail: mix1959@gmail.com

Прокофьев Олег Владимирович, кандидат технических наук, доцент, Пензенский филиал Финансового университета при Правительстве Российской Федерации, 440052, Российская Федерация, г. Пенза, ул. Калинина, 33б, e-mail: prokof_ow@mail.ru

Савочкин Александр Евгеньевич, аспирант, Пензенский государственный технологический университет, 440039, Российская Федерация, г. Пенза, проезд Байдукова / ул. Гагарина, 1а / 11, e-mail: aebrat@mail.ru

Линкова Марина Алексеевна, магистр, Финансовый университет при Правительстве Российской Федерации, 125993, Российская Федерация, г. Москва, ГСП-3, Ленинградский проспект, 49, e-mail: m_linkova@mail.ru

В технически сложных объектах (ТСО) зачастую возникают ситуации, когда наблюдаемый процесс можно охарактеризовать только короткими временными рядами (ВР). Это быстропротекающие процессы, такие как удар, взрыв, краткосрочная вибрация, или ситуация, когда получение каждого наблюдения является либо дорогостоящей операцией, либо опасной для жизни и здоровья. По полученным данным необходимо составить прогноз дальнейшего развития состояния ТСО. Прогнозирование предаварийного или аварийного состояния ТСО осуществляется посредством выявления закономерностей в этих ВР. Отсутствие предварительных данных по процессу, необходимость быстрого получения результатов прогнозирования, малое число отсчетов во ВР являются существенными ограничениями для применения большинства известных инструментальных методов прогнозирования. В данной статье предлагаются к использованию математические и информационно-структурные модели прогнозирования технического состояния ТСО на основе проверки статистических гипотез для коротких ВР. Обоснована целесообразность разработки новых и синтез существующих математических моделей в рамках комплексного алгоритма анализа коротких ВР, для которых выявление определенных зависимостей может быть интерпретировано как проверка на возможное развитие ненормативных (нештатных) процессов в работе ТСО.

Ключевые слова: технически сложный объект, прогнозирование, короткий временной ряд, гетероскедастичность, статистические гипотезы, техническое состояние, тренд, излом тренда, предаварийная ситуация, аварийное состояние

MATHEMATICAL AND INFORMATION-STRUCTURAL PREDICTION MODELS STATE OF TECHNICALLY COMPLEX OBJECTS

Mikheev Mikhail Yu., D.Sc. (Engineering), Professor, Penza State Technological University, Baydukov Pr. / 1a / 11 Gagarin St., Penza, 440039, Russian Federation, e-mail: mix1959@gmail.com

Prokofev Oleg V., Ph.D. (Engineering), Associate Professor, Penza branch of Financial University under the Government of the Russian Federation, 33b Kalinina St., Penza, 440052, Russian Federation, e-mail: prokof_ow@mail.ru

Savochkin Aleksandr Ye., post-graduate student, Penza State Technological University, Baydukov Pr. / 1a / 11 Gagarin St., Penza, 440039, Russian Federation, e-mail: aebrat@mail.ru

Linkova Marina A., master, Financial University under the Government of the Russian Federation, GSP-3, 49 Leningradskiy Ave., Moscow, 125993, Russian Federation, e-mail: m_linkova@mail.ru

There are often situations in the technically complex objects (TCO) when observed process can be described only short time series (the fast-proceeding processes, such as shock, explosion, short-term vibration, or a situation when receiving each observation is either expensive business, or life-threatening and health). The obtained data is necessary to predict future status of TCO. Prediction of pre-emergency or emergency status of TCO happens by means of detection of regularities in these time series. Absence of preliminary data on process, need of fast receiving results of prediction, small number of levels for a time series are essential restrictions for application of the majority of known tools and forecasting methods. In this article mathematical and information and structural forecasting models of technical condition of TCO on the basis of check of statistical hypotheses of short time series are offered to use. Development new and synthesis of the existing mathematical models in a complex analysis algorithm of short time series existence of certain dependences in which can be interpreted as check on possible development of substandard processes in operation of TCO is offered.

Keywords: technically complicated object, forecasting, short time series, heteroscedasticity, statistical hypothesis, the technical condition of the trend, a break of the trend, pre-emergency situations, emergency condition

Введение. В связи с масштабной эксплуатацией на территории нашей страны и за рубежом значительного числа технически сложных объектов (ТСО), в том числе объектов, представляющих опасность для окружающих, предупреждение и прогнозирование предаварийных или аварийных ситуаций, прогнозирование тенденций к изменению состояния процесса или ТСО в целом на них является весьма актуальной задачей. Предаварийная ситуация характеризуется отклонением от номинального режима работы, что при неблагоприятных условиях ведет к аварии.

В ТСО зачастую возникают ситуации, когда наблюдаемый процесс, по которому необходимо составить прогноз, можно охарактеризовать только короткими временными рядами (ВР). Это могут быть быстропротекающие процессы (такие как удар или краткосрочная вибрация) или ситуация, когда получение каждого наблюдения является либо очень дорогостоящей операцией, либо опасно для жизни и здоровья. При этом количество отсчетов во ВР зачастую не превышает 30 значений. Такой малый объем данных и необходимость максимально быстрого получения прогнозного результата состояния ТСО накладывают ряд ограничений на применение известных методов анализа мониторинговых данных и прогнозирования.

Анализируя огромный перечень современных методов прогнозирования (например, [3]), можно сделать вывод, что большинство из них имеют ряд недостатков, затрудняющих или делающих невозможным применение классических алгоритмов, моделей и методов для прогнозирования состояния ТСО по коротким ВР. Основные причины: отсутствие предварительных данных по процессу; необходимость быстрого получения результатов прогнозирования; малое число отсчетов во временном ряду. Кроме того, некоторые методы (в т.ч. связанные с применением искусственных нейронных сетей) требуют достаточно сложного процесса адаптации/обучения для получения прогнозов и др.

Постановка задачи и ее практическая значимость. Прежде всего, дадим формальные определения терминов.

Предаварийная ситуация – ситуация, при которой отклонение от норм технологического режима, или состояния оборудования приводит к выходу за предаварийные граничные значения [21].

Аварийное состояние - категория технического состояния сооружения в целом, характеризующаяся повреждениями и деформациями, свидетельствующими об исчерпании

несущей способности и опасности обрушения; характеризующаяся кренами, которые могут вызвать потерю устойчивости объекта [7].

Своевременное выявление (а затем устранение) таких состояний, а также тенденций перехода в эти состояния позволяет обеспечить надежность эксплуатации ТСО, в том числе зданий и сооружений.

В данной работе основная задача состоит в модификации и синтезе существующих математических моделей в комплексный алгоритм анализа коротких ВР, характеризующих процессы, происходящие в ТСО. При этом выявление определённых зависимостей в таких ВР может быть интерпретировано как результат развития ненормативных (нештатных) процессов в работе ТСО.

Новизна и практическая значимость предлагаемого в данной работе решения заключаются в следующем:

- становится возможным составлять обоснованный прогноз состояния ТСО по коротким ВР;
- вычисления происходят в максимально короткие сроки, что в области обеспечения безопасности ТСО является одним из ключевых факторов;
- известные методы применяются в новой предметной области - в качестве отдельных операций в составе новых алгоритмов предупреждения предаварийного или аварийного состояния ТСО.

Базовая методология. В рамках разработанного алгоритма короткий ВР, характеризующий процессы на (в) ТСО, подвергается параллельному анализу на предмет наличия/отсутствия гетероскедастичности и обнаружения трендовой составляющей. Гетероскедастичность обычно связывают с «неоднородностью» наблюдений, выражающейся в неодинаковой дисперсии случайной ошибки для регрессионной модели.

В случае если тренд обнаружен, то проводится проверка ВР на определение излома (структурной нестабильности) тренда.

Проблема выявления гетероскедастичности не всегда может быть решена путём визуальных наблюдений за значениями выходного сигнала датчика по техническим (высокая скорость протекания процессов) и психологическим причинам (зрительные эффекты, разная интерпретация экспертами одинаковых графиков) [18]. Для решения этой проблемы в разработанном авторами алгоритме предлагается использовать методы проверки гипотезы о наличии функциональной связи между остатком и фактором (в данном случае, временем).

Если обнаруживается тенденция к увеличению значений ВР или выявляется гетероскедастичность с нарастанием дисперсии остатка, тогда нахождение выходного сигнала датчика в окрестностях предельных значений диапазона для объекта будет означать работу на границе предусмотренных технических характеристик функционирования. Иначе говоря, если наблюдается усиление рассеянности выходного сигнала датчика, то для наблюдаемого объекта это может означать функционирование на предельных возможностях работоспособности. Это и будет свидетельствовать о наступлении предаварийной или аварийной ситуации для всей системы в целом.

В качестве теста на определение гетероскедастичности (различные дисперсии [3]) в данной работе предлагается применить метод Парка. Для определения тренда мы используем метод средних, а для обнаружения излома в тренде используем модифицированный тест Чоу [16]. Все рассматриваемые в данной работе вычисления проводились с помощью разработанных авторами программ на встроенном языке программирования системы MATLAB.

У каждого метода (в том числе и у рассматриваемых в данной работе) есть определённые ограничения. Например, длина анализируемого ВР для теста Парка не должна быть меньше 7 значений. Поэтому прежде чем подвергать ВР анализу, он проходит предвари-

тельный этап предобработки. На этом этапе и определяется возможность применения к нему тех или иных методов анализа. В результате анализ проводится с использованием только тех инструментальных методов, на применение которых анализируемый ВР не накладывает никаких ограничений.

Разработка алгоритма прогнозирования технического состояния ТСО. На рисунке 1 представлен алгоритм прогнозирования технического состояния ТСО. Каждый этап может включать в себя несколько тестов. В результате выполнения всего алгоритма получается совокупность данных, на основе которой и можно будет судить о том, имеется ли у ТСО тенденция перехода к предаварийному или аварийному состоянию.

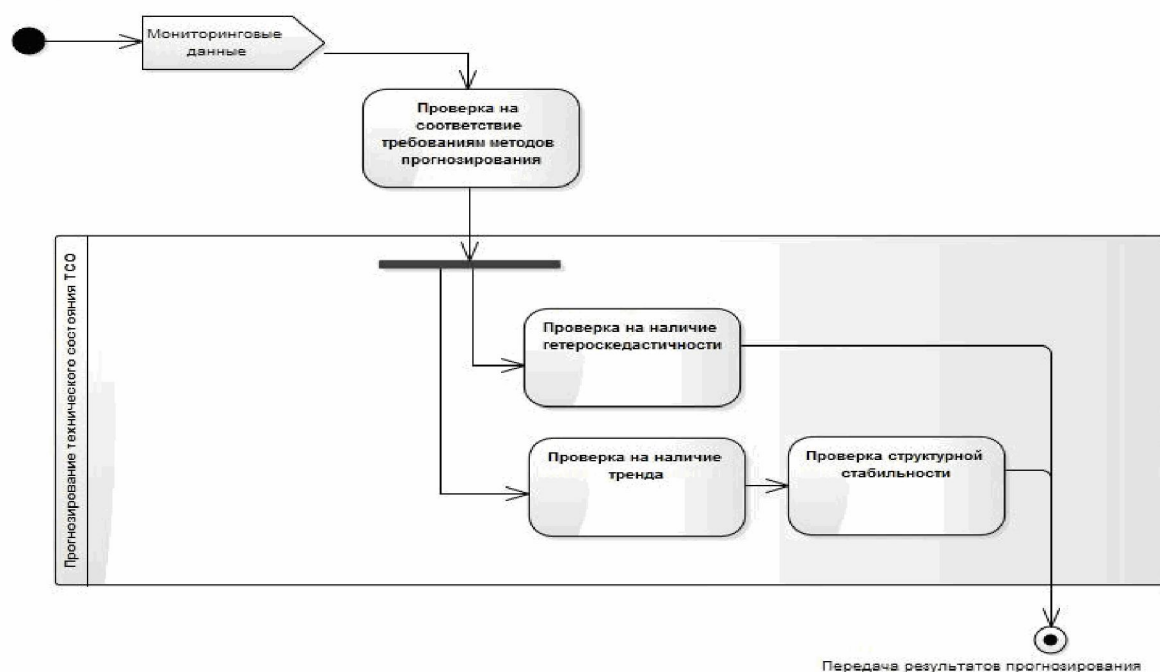


Рис. 1. Алгоритм прогнозирования технического состояния ТСО в нотации UML 2.0

Характеристика использованного экспериментального материала. Для анализа были взяты данные испытаний ударного акселерометра ВТ-51. Этот датчик предназначен для измерения линейных ускорений и составляющих ударных процессов. Указанный акселерометр является частью тракта преобразования в изделиях ракетно-космической техники [13]. Он служит для оценки возникающих в элементах конструкции перегрузок и применяется для решения прочностных задач. Источником данных послужили протоколы ускоренных испытаний на сохраняемость акселерометров ВТ-51, протокол испытаний на многоразовость и пожаровзрывозащищенность акселерометров ВТ-51, протокол лабораторных испытаний экспериментальных образцов акселерометров с металлическим подвесом. В данных протоколах отражены результаты испытаний 7-ми акселерометров. В рамках испытаний проводились проверка датчиков при нормальных условиях, при повышенных и пониженных температурах, при воздействии линейных ускорений, при вибрационном (ударно-импульсном) воздействии на датчики, при воздействии акустического шума.

Датчик на основе этого акселерометра выполнен в моноблочном исполнении – чувствительный элемент и преобразователь расположены в общем корпусе. Чувствительный элемент прибора дифференциально-трансформаторного типа с подвижным сердечником,

закреплённым на упругом подвесе, имеет жидкостное демпфирование нежелательных колебаний сердечника. Упругий подвес состоит из двух металлических пружин (рисунок 2), по наружному периметру жестко связанных с корпусом акселерометра, а в центральной части закрепленных на торцовых поверхностях инерционного элемента. Электронный преобразователь, собранный на бескорпусной базе, усиливает сигнал с чувствительного элемента до нормированного значения и имеет активный фильтр, настроенный на требуемый частотный диапазон [12].

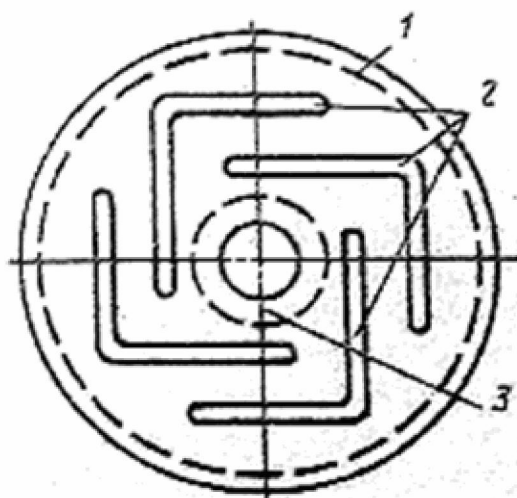


Рис. 2. Металлическая пружина акселерометра ВТ-51:
 1 – наружный периметр, 2 – сквозные пазы, 3 – центральная часть

Методология исследований и полученные результаты. Для прогнозирования состояния ТСО анализируется выходное напряжение (сигнал) датчика. С целью демонстрации эффективности предлагаемых методов (моделей) приведем пример анализа нескольких из имеющихся ВР испытаний акселерометра ВТ-51 (рис. 3).

На этапе проверки на гетероскедастичность ВР подвергается анализу несколькими независимыми друг от друга методами. При применении метода Парка предполагается, что дисперсия остатков связана со значением факторов функцией

$$\ln \varepsilon^2 = a + b \ln t, \quad (1)$$

где ε - остатки временного ряда; a, b – коэффициенты регрессии.

Проверяется значимость коэффициента регрессии по критериям Стьюдента (t) [25] и Фишера (F) [24]. Если коэффициент регрессии для $\ln \varepsilon^2$ окажется статистически значимым, то существует зависимость $\ln \varepsilon^2$ и $\ln t$, то есть имеет место гетероскедастичность остатков [18].

Для рассматриваемого ряда по рис.3 получаем следующее уравнение типа (1):

$$\ln \varepsilon^2 = -9,02 + 0,73 * \ln t$$

Кроме того определим расчетные значения критериев $F_{\text{расчетное}} = 4,34$ и $t_{\text{расчетное}} = 2,63$.

Табличные же значения при уровне значимости 0,05, степенях свободы 1 и 23 составляют $_{0,05}F_{1,23} = 4,279$, а $_{0,05}t_{23} = 2,068$. Напомним, что F-статистика - это отношение факторной дисперсии к остаточной. Число степеней свободы f_l факторной дисперсии равно m – т.е. количеству объясняющих переменных В нашем случае, это только фактор времени, т.е. $m = 1$, соответственно $f_l = 1$.

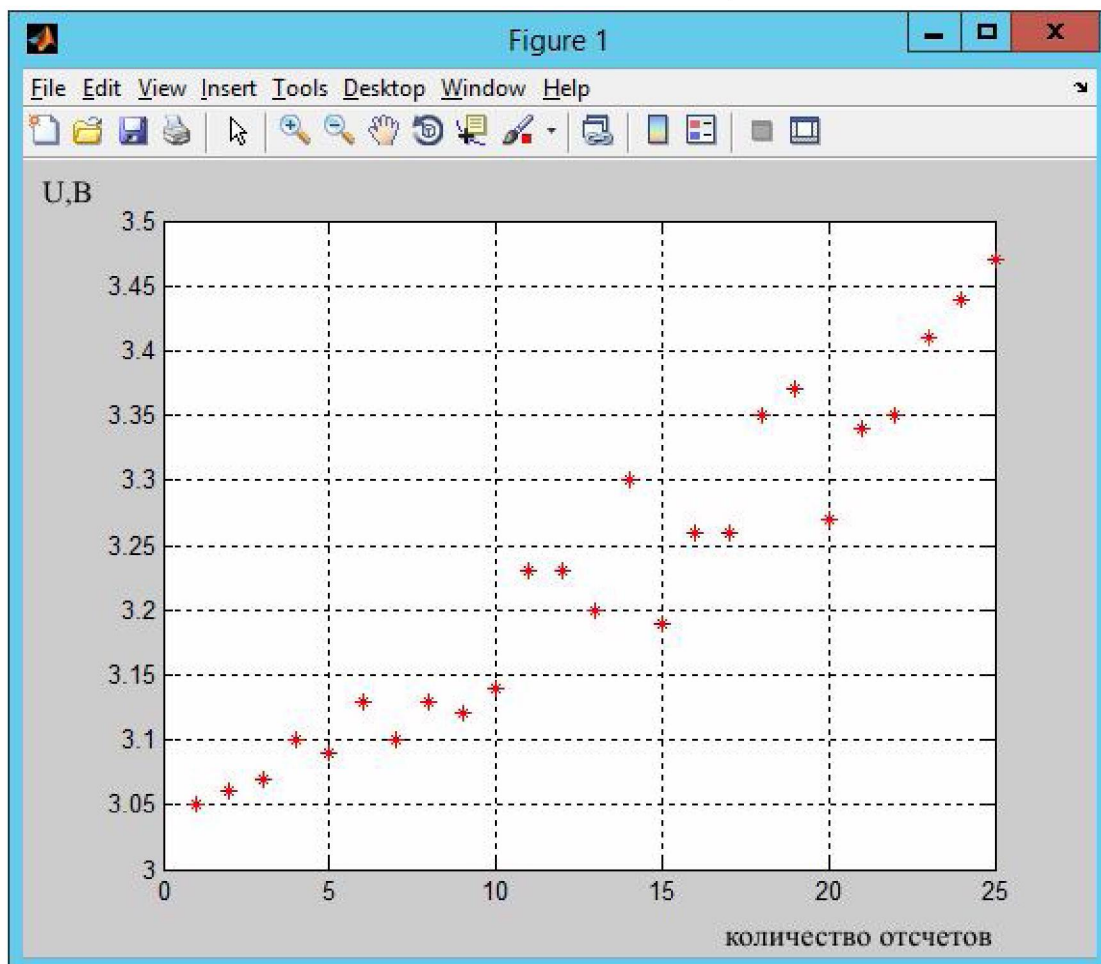


Рис. 3. График анализируемого временного ряда

Число степеней свободы остаточной дисперсии $f_2 = n - m - 1$, где n – размер исследуемого ВР [10]. Поскольку $F_{\text{расчетное}} > F_{\text{табл}}$ и $t_{\text{расчетное}} > t_{\text{табл}}$, то можно утверждать, что в данном ряду дисперсия остатков представляет собой статистически значимую степенную функцию от времени. Следовательно, гетероскедастичность в ВР присутствует.

Другой (параллельный) этап разработанного алгоритма – поиск трендовой составляющей методом проверки разности средних двух разных частей одного и того же ряда. С этой целью имеющийся ВР разбивается на две части, примерно равные по количеству экспериментальных точек. Обе части ВР рассматриваются как самостоятельные выборочные совокупности, которые должны подчиняться нормальному закону распределения.

Для проверки на нормальность воспользуемся коэффициентами эксцесса и асимметрии. Коэффициент эксцесса E (мера скученности) и коэффициент асимметрии A (мера скошенности) распределения случайной величины « x » определяются [22] по:

$$A = \mu_4 / \sigma^4 - 3, \quad (2)$$

$$E = \mu_3 / \sigma^3, \quad (3)$$

где:

$\mu_4 = M[(x - M[x])^4]$ – четвёртый центральный момент случайной величины X ;

$\mu_3 = M[(x - M[x])^3]$ - третий центральный момент случайной величины X ;
 $\sigma = \sqrt{M[(x - M[x])^2]}$ - второй центральный момент (стандартное отклонение) случайной величины X .

Если выполняются соотношения:

$$|A| \leq 1,5S_a \quad (4)$$

и

$$|E| \leq 1,5S_e, \quad (5)$$

где:

$$S_a = \sqrt{6(n-2)/((n+1)(n+3))}$$

$$S_e = \sqrt{24n(n-2)(n-3)/((n+1)^2(n+3)(n+5))}, \quad (n - \text{длина ряда}),$$

то считается, что распределение ряда не противоречит нормальному закону. Если же выполняется одно из соотношений:

$$|A| > 2S_a \quad (6)$$

или

$$|E| > 2S_e, \quad (7)$$

то распределение ряда не соответствует нормальному закону.

Если же значения данных величин попадают в зону неопределенности (интервал $(1,5S; 2S]$), то для оценки распределения применяется RS критерий:

$$RS = (\varepsilon_{\max} - \varepsilon_{\min}) / S, \quad (8)$$

где $\varepsilon_{\max}$ – максимальный уровень ряда остатков; $\varepsilon_{\min}$ – минимальный уровень ряда остатков; S – среднеквадратическое отклонение.

Полученные для анализируемого ВР значения рассмотренных показателей приведены в таблице 1.

Таблица 1

Результаты проверки частей ВР на нормальное распределение

Наименование показателя	1-ая часть ряда	2-ая часть ряда
коэф. асимметрии A	0,883	0,063
коэф. эксцесса E	-0,174	-0,959
показатель S_a	0,554	0,542
показатель S_e	0,775	0,779
RS-критерий	3,382 (при допустимом диапазоне [2,96;4,41])	не требуется
Нормальное распределение	да	да

Как видно из таблицы 1, обе части рассматриваемого ряда подчиняются нормальному закону распределения. Поэтому проверка на наличие тренда методом средних допустима. После анализа на нормальность для каждой из частей исходного ряда (рисунок 4) вычисляем средние значения:

$$\bar{Y}_1 = (\sum_{t=1}^{n_1} y_t) / n_1 \text{ и } \bar{Y}_2 = (\sum_{t=n_1+1}^n y_t) / n_2, \quad (8)$$

и дисперсии:

$$S^2_{y1} = (\sum_{t=1}^{n_1} (y_t - \bar{Y}_1)) / (n_1 - 1) \text{ и } S^2_{y2} = (\sum_{t=n_1+1}^n (y_t - \bar{Y}_2)) / (n_2 - 1). \quad (9)$$

Здесь n_1 и n_2 - размеры (количества значений) первой и второй частей ряда, y_1, y_2 - значения подвыборок, $\bar{Y}_1, \bar{Y}_2$ - средние значения подвыборок.

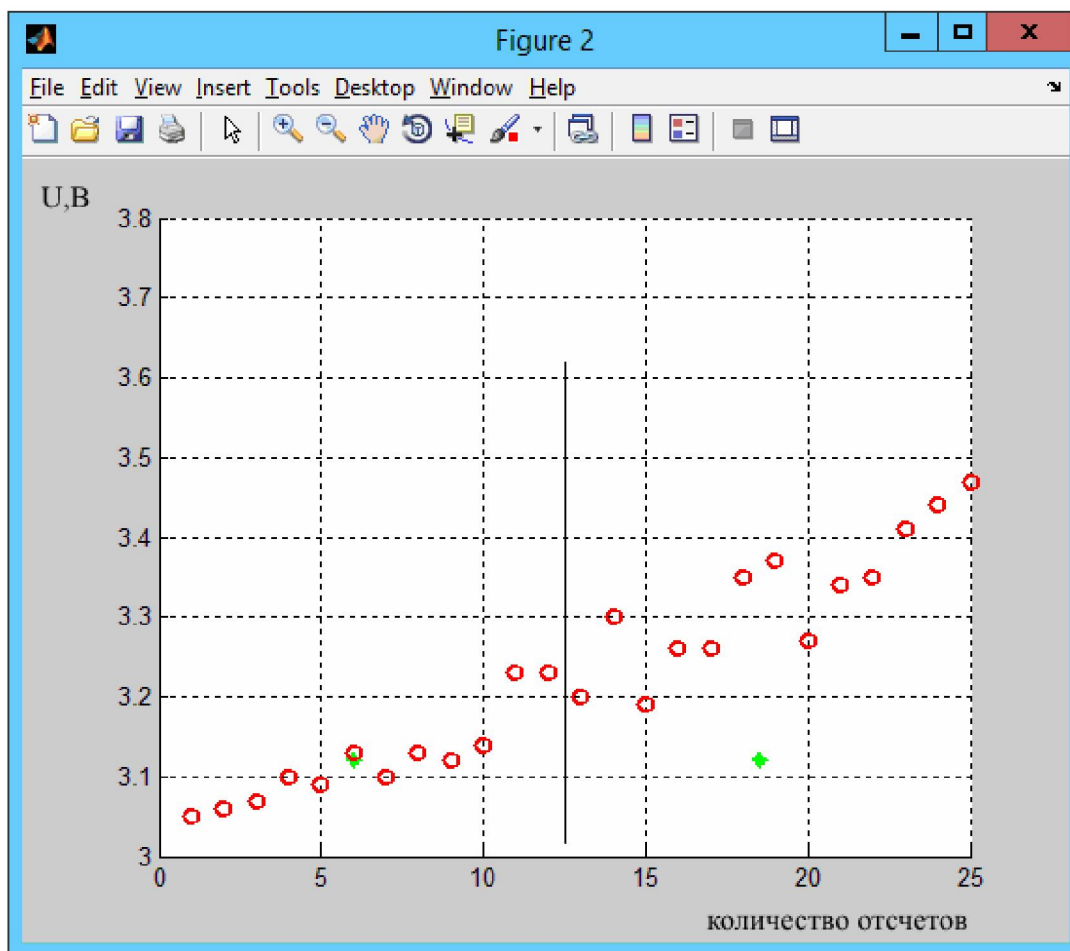


Рис. 4. Проверка наличия тренда во ВР методом средних

Проверяем гипотезу о равенстве (однородности) дисперсий обеих частей ряда, сравнивая рассчитанное $F_{расч}$ со значением F-критерия Фишера [14, 24].

$$F_{расч} = S^2_{y2} / S^2_{y1}. \quad (10)$$

Для рассматриваемого ВР получаем:

$$Y_1 = 3,12; Y_2 = 3,32; S_1 = 0,003; S_2 = 0,007; F_{расч} = 2,223; F_{кр} = 2,717; t_{расч} = -6,787; t_{кр} = 2,068.$$

Таким образом, $F_{расч} < F_{кр}$ и значит можно проверять основную гипотезу о равенстве средних значений с использованием t-критерия Стьюдента [25]. Получаем $|t_{расч}| > t_{кр}$, т.е. можно говорить о наличии в рассматриваемом ряду трендовой составляющей (рис. 5). Синим цветом дан общий тренд для всей области на основе уравнения линейной регрессии.

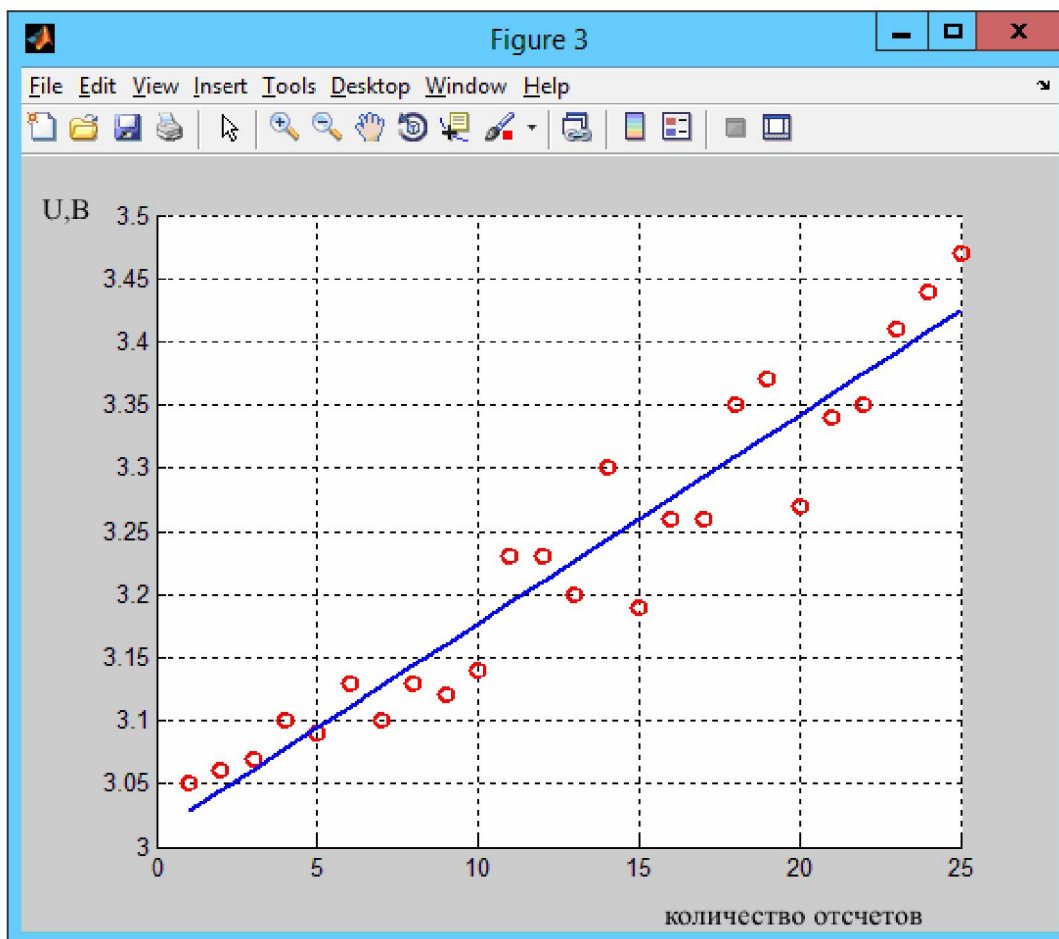


Рис. 5. Результат поиска трендовой составляющей в рассматриваемом временном ряду

Заключительная часть исследования ряда – определение структурной стабильности тренда, т.е. поиск (определение наличия) излома в трендовой составляющей. В качестве инструментария для поиска возможного излома тренда ВР без участия эксперта применялся модифицированный метод Чоу [16, 23].

Исходный ВР делится на две части. Имеющееся количество наблюдений в нашем случае мало, а «стенки» параболической, экспоненциальной, степенной и даже зачастую гиперболической зависимости на коротком участке стремятся к линейному виду (рис. 6) [2]. Поэтому будем рассматривать каждую часть ВР как ряд с линейной зависимостью (переход от нелинейных моделей к линейным – широко применяемый инструмент упрощения расчётов [22]).

По каждому из двух «субрядов» (подвыборок) строятся две линейные регрессионные модели:

$$Y_i = \beta_{1_0} + \sum \beta_{1_j} \cdot X_{ij} + \varepsilon_1 \quad i = \overline{1, n_1} \quad (11)$$

$$Y_i = \beta_{2_0} + \sum \beta_{2_j} \cdot X_{ij} + \varepsilon_2 \quad i = \overline{n_1 + 1, n_1 + n_2} \quad (12)$$

где X – значения ряда; β_1, β_2 – векторы параметров моделей двух частей ряда; $\varepsilon_1, \varepsilon_2$ – их случайные возмущения.

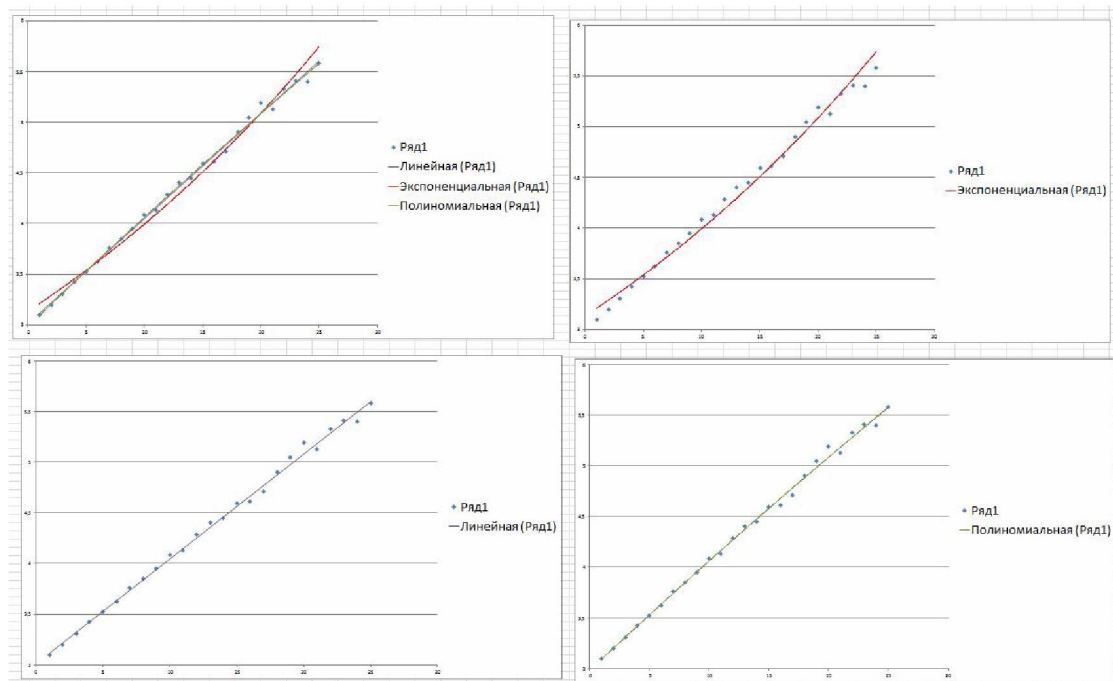


Рис. 6. Графики трендов разных форм для короткого ВР

1) Проверяется нулевая гипотеза:

$$H_0 : \beta_1 = \beta_2; D(\varepsilon_1) = D(\varepsilon_2) = \sigma^2.$$

2. Рассчитываются суммы квадратов остатков для линейных регрессий по этим подвыборкам (S_{e2}^2, S_{e1}^2).

3. Строится регрессия $Y_i = \beta_0 + \sum \beta_j \cdot X_{ij} + \varepsilon$ по всему ряду и рассчитывается сумма квадратов остатков S_e^2 .

Для автоматизированного поиска предполагаемого изменения линии тренда целесообразно решать оптимизационную задачу. Она состоит в нахождении максимального значения F-критерия Фишера:

$$F = (S_e^2 - S_{e1}^2 - S_{e2}^2)(n-4)/(2S_{e1}^2 + 2S_{e2}^2) \rightarrow \max \quad (13)$$

при соблюдении функциональных ограничений [10]:

$$\begin{cases} n_1 + n_2 = n \\ n_1 \geq 7 \\ n_2 \geq 7 \\ 0 < X_A < X_B \end{cases},$$

где n_1 и n_2 – объёмы подвыборок наблюдений, составляющих общий ряд объёмом n и не принимающие значений меньше рекомендуемых; X_A, X_B , соответственно, координаты точки излома тренда и крайней (последней) точки предсказанного значения ВР; $S_e^2, S_{e1}^2, S_{e2}^2$ – суммы квадратов остатков регрессионных моделей, соответственно, для модели, построенной по всем наблюдениям; для моделей по первой и второй подвыборкам наблюдений.

Найденное по формуле (13) значение $F_{\text{рассчитанное}}$ сравнивается с критическим (табличным) значением $F_{\text{кр}}$. Последнее определяется по таблице распределения Фишера – Снедекора в зависимости от уровня значимости и двух степеней свободы $f1 = m + 1$ и $f2 = n - 2m - 2$, где m – количество факторных переменных (в данном случае $m=1$, так как фактор только один – время); n – длина исходного временного ряда) [10].

Если расчётное значение больше или равно критическому, то результат значим, а граница между подвыборками соответствует моменту времени излома. Если расчётное значение меньше критического, то гипотеза о структурной стабильности тренда не отвергается, что говорит об отсутствии излома. В результате анализа рассматриваемого ряда, для каждой возможной точки излома было получено значение $F_{\text{рассчитанное}}$ (табл. 2).

Таблица 2

Вычисленные значения $F_{\text{рассчитанное}}$ для каждой возможной точки излома по анализируемому ВР

$F_{\text{расч}}$	1,1241	1,1340	1,3024	1,7251	0,7339	0,8108
Номер	7	8	9	10	11	12
$F_{\text{расч}}$	0,7621	2,1512	1,0822	1,0849	1,2359	1,1082
Номер	13	14	15	16	17	18

Как видно из таблицы 2, в данном ВР максимальное значение $F_{\text{рассчитанное}} = 2,1512$ (точка № 14). Критическое же значение $_{0,05}F_{\text{кр}2,21} = 19,4481$, то есть в рассматриваемом ВР излома нет ни в одной из точек. Следовательно, ряд должен рассматриваться как «структурно стабильный».

Проверка адекватности модели реальному явлению является важным этапом прогнозирования. Наиболее важными свойствами остаточной компоненты (случайных остатков) являются независимость последовательных значений ряда остатков, их случайность и соответствие нормальному закону распределения. В настоящее время разработан, проанализирован и апробирован широкий спектр критериев адекватности статистических моделей [6, 11, 14]. Проведем оценку адекватности рассматриваемой модели на основе исследования случайности отклонений ряда остатков по критерию пиков [4] (этот критерий имеет и другое название – «критерий поворотных точек»).

Количество поворотных точек для рассматриваемого ВР равно 17 (рис. 7).

Критерий случайности отклонений от тренда при уровне вероятности 0,95 можно представить, как:

$$p > [2(n-2)/3 - 1,96\sqrt{(16n-29)/90}] = 11, \quad (14)$$

где n – длина ряда.

Получаем $p = 17 > 11$. Таким образом, остатки случайны, неравенство выполняется ($17 > 11$). Модель по этому критерию адекватна.

Оценим адекватность модели на основе исследования независимости (отсутствия автокорреляции) значений ряда остатков по критерию Дарбина – Уотсона [14]. С этой целью построим dw-статистику, в основе которой лежит расчетная формула:

$$dw = \sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2 / \sum_{i=1}^n \varepsilon_i^2, \quad (15)$$

где ε_i – разница реального и прогнозного значений.

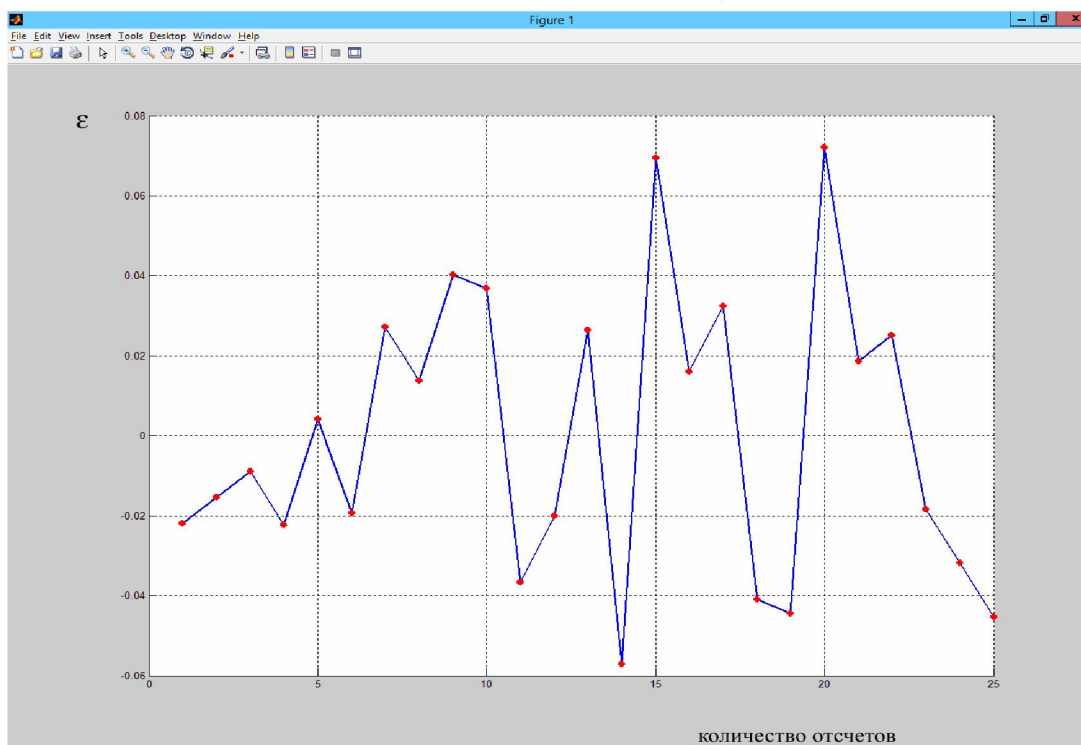


Рис. 7. График отклонений (случайных остатков) после вычитания тренда из ВР

Теоретическое основание применения этого критерия обусловлено тем, что в динамических рядах как сами наблюдения, так и отклонения от них распределяются в хронологическом порядке. При отсутствии автокорреляции значение dw примерно равно «2», а при полной автокорреляции – «0» или «4». Следовательно, оценки, получаемые по этому критерию, являются не точечными, а интервальными. Верхние (d_2) и нижние (d_1) критические значения, позволяющие принять или отвергнуть гипотезу об отсутствии автокорреляции, зависят от количества значений динамического ряда и числа независимых переменных модели. Значения этих границ для трех уровней значимости ($\alpha = 0,01$, $\alpha = 0,025$ и $\alpha = 0,05$) даны в специальных таблицах. При сравнении расчетного значения d-статистики с табличным могут возникнуть такие ситуации.

$d_2 < dw < 2$ – ряд остатков не коррелирован. В этом случае по данному критерию можно сделать вывод о выполнении свойства независимости. Следовательно, модель по этому критерию адекватна.

$0 < dw < d_1$ – остатки содержат автокорреляцию, то есть зависимы. Модель неадекватна.

$d_1 < dw < d_2$ – область неопределенности, когда нет оснований ни принять, ни отвергнуть гипотезу о существовании автокорреляции. При этом критерий Дарбина – Уотсона не дает ответа на вопрос о независимости значений ряда остатков. В таком случае необходимо воспользоваться другими критериями, например, проверить независимость уровней по первому коэффициенту автокорреляции.

$$r(1) = \frac{\sum_{i=2}^n (\varepsilon_i \times \varepsilon_{i-1})}{\sum_{i=1}^n \varepsilon_i^2}. \quad (16)$$

Если модуль рассчитанного значения первого коэффициента автокорреляции меньше критического значения $|r(1)| < r_{\text{таб}}$, то уровни ряда остатков независимы.

Если $dw > 2$, то это свидетельствует о наличии отрицательной автокорреляции. Перед входом в таблицу такие значения следует преобразовать по формуле $dw' = 4 - dw$.

В нашем случае получаем $dw = 2,044$, то есть наблюдается отрицательная автокорреляция. Значит, для проверки используем формулу $dw' = 4 - dw = 4 - 2,044 = 1,956$. При табличных $d_1 = 1,29$ и $d_2 = 1,45$, получаем: $d_2 < dw < 2$, т.е. модель по этому критерию адекватна.

Далее оценим адекватность модели на основе исследования нормальности закона распределения ряда остатков с помощью рассмотренных выше критериев эксцесса и асимметрии.

Для анализируемого ряда по формулам (2), (3) получаем:

$$A = 0,3121; S_a = 0,4354; E = -1,0860; S_e = 0,7312.$$

То есть выполняются оба неравенства:

$$|A| \leq 1,5S_a \text{ и } |E| \leq 1,5S_e \quad (0,3121 < 0,6531 \text{ и } 1,0860 < 1,0968).$$

Значит, ряд подчиняется нормальному закону распределения и модель по этому признаку адекватна.

В таблице 3 собраны (сгруппированы) данные анализа ряда остатков.

Таблица 3

Результаты проверки адекватности модели					
Проверяемое свойство	Используемые статистики		Граница		Вывод
	наименование	значение	нижняя	верхняя	
Независимость	dw-критерий Дарбина – Уотсона	dw = 2,044 dw* = 1,956	1,29	1,45	адекватна
Случайность	Критерий пиков (поворотных точек)	17	11		адекватна
Нормальность	Критерий асимметрии	0,3121	1,5 × 0,4354 = 0,6531		адекватна
	Критерий эксцесса	1,0860	1,5 × 0,7312 = 1,0968		
Вывод: Модель статистически адекватна					

Анализируя совокупность рассмотренных критериев можно сделать вывод, что модель является адекватной.

Итак, в результате проведенного анализа в рассматриваемом ВР было обнаружено наличие гетероскедастичности и трендовая составляющая. В отношении оценки технического состояния это может говорить о возможном ненормативном (нештатном) развитии процессов в ТСО. Иными словами, возможен переход к предаварийному или аварийному состоянию.

Анализ результатов экспериментов. В ходе экспериментов было проанализировано 273 выборки результатов испытаний ударного акселерометра ВТ-51. Фрагмент итоговой таблицы с результатами анализа представлен в таблице 4.

Таблица 4

Фрагмент итоговой таблицы результатов анализа						
Гетероскедастичность	Тренд	Излом тренда	Адекватность по критерию Дарбина – Уотсона	Адекватность по критерию пиков	Адекватность по критерию асимметрии и эксцесса	Прогнозируется угроза предаварийной ситуации
нет	да	да	да	да	да	да
да	нет	нет	да	нет	нет	нет
нет	да	нет	нет	да	да	нет
да	нет	нет	да	да	нет	нет
да	нет	да	да	да	да	да
да	да	да	да	да	да	да

нет	нет	нет	да	да	да	нет
нет	да	да	да	да	нет	да
да	да	да	да	да	нет	да
нет	да	нет	да	да	да	нет
нет	да	нет	да	да	да	нет
нет	нет	нет	нет	да	нет	нет
нет	да	да	да	да	да	да
нет	нет	нет	да	да	да	нет
нет	нет	нет	да	нет	нет	нет
да	да	да	да	да	да	да
нет	нет	нет	нет	нет	нет	нет
да	да	нет	да	да	да	да
нет	да	да	да	нет	да	да
нет	нет	нет	нет	нет	нет	нет
нет	нет	нет	нет	да	нет	нет
да	да	нет	да	да	да	да
нет	нет	нет	да	да	да	нет
нет	да	да	да	да	нет	да
да	нет	нет	да	да	да	нет

Для определения степени достоверности и точности предлагаемого алгоритма (рис. 1), воспользуемся U -статистикой Тейла, которая позволяет сравнивать прогностические способности моделей. Если значение U -статистики < 1 , то анализируемая модель лучше по прогностическим способностям по сравнению с первоначальной («эталонной») моделью; если > 1 – то хуже; если $= 1$, то анализируемая модель по прогностическим способностям эквивалентна первоначальной («эталонной») модели [20].

U -статистика Тейла, или резюмирующая статистика, основана на стандартной, симметрической функции потерь:

$$U = \sqrt{\sum_{t=t_0}^{t_0+T} (x_t - x_t^f)^2} / \sqrt{\sum_{t=t_0}^{t_0+T} (x_t - x_{t-1})^2}, \quad (17)$$

где x_t - реальное состояние объекта (взято из протокола испытаний), а x_t^f – результат прогнозирования состояния угрозы (ненормативной ситуации), полученный в результате применения алгоритма (столбец 8 табл. 4).

Получаем значение $U = 0,715$, что свидетельствует о высоких прогностических качествах применяемых моделей и алгоритмов.

Кроме того, для оценки качества прогнозирования, мы использовали показатель разброса экспериментальных (фактических) значений по отношению к расчетным - безразмерный коэффициент детерминации (coefficient of determination) R^2 [15]:

$$R^2 = \sum_{k=1}^n (Y_k^* - M[Y_k])^2 / \sum_{k=1}^n (Y_k - M[Y_k])^2 = 1 - \sum_{k=1}^n (Y_k^* - Y_k)^2 / \sum_{k=1}^n (Y_k - M[Y_k])^2, \quad (18)$$

где «М» – оператор математического ожидания.

В данном случае знаменатель дроби в (18) выступает индикатором для общего рассеивания Y_k относительно линии математического ожидания $M[Y_k]$. В числителе же (18)

представлена сумма квадратов отклонений, интерпретируемая как не объясненный данной моделью разброс. Таким образом, можно считать, что R^2 показывает, насколько хорошо выбранная модель объясняет имеющиеся данные. Из расчетных данных следует, что точность предсказания предаварийной ситуации составляет 74 %. Минимально допустимым значением (лимитом точности моделирования), как правило, считают $R^2 = 70 - 75 \%$ [15]. Таким образом, предложенная модель может считаться достаточно точной.

Выводы. Таким образом, были модифицированы и агрегированы в комплексный алгоритм анализа коротких ВР, характеризующих процессы ТСО, существующие математические модели.

На примере анализа данных с ударного акселерометра ВТ-51, являющегося частью тракта преобразования в изделиях ракетно-космической техники, установлено и подтверждено следующее.

- Модификация и агрегирование известных статистических моделей в комплексный алгоритм позволяют с достаточно высоким качеством спрогнозировать предаварийную ситуацию на ТСО по короткому набору данных.
- Для прогноза не требуются дополнительные данные с объекта – это позволяет анализировать ВР для уникальных неповторяющихся процессов.
- В связи с тем, что не требуется предварительная статистика по процессу и отсутствует этап обучения или адаптации, прогноз составляется в максимально короткие сроки.

В области предсказания предаварийной и аварийной ситуации на ТСО важен сам факт подтверждения рассмотренных гипотез, который даёт возможность сделать качественный вывод о возможном в будущем ненормативном состоянии контролируемого объекта.

Список литературы

1. Абзалов А.В., Жедунов Р.Р. Методика анализа предаварийных ситуаций на технологических объектах управления. // Прикаспийский журнал: управление и высокие технологии. – 2013, № 4 (24), с. 50-58.
2. Алтынов П.И., Баврин И.И. Звавич Л.И. Математика. Большой справочник для школьников и поступающих в вузы. М.: Дрофа - 2009, 768 с.
3. Артеменко М.В., Бабков А.С. Классификация методов прогнозирования поведения систем // Современные проблемы науки и образования - 2013, № 6, с. 13-19.
4. Архипова А.В. Применение критерия пиков для проверки адекватности трендовой модели при анализе прямых иностранных инвестиций в РФ // Организация работы с молодежью. – 2014, № 9, с. 13-18.
5. Брумштейн Ю.М., Тарков Д.А., Дюдилов И.А. Анализ моделей и методов выбора оптимальных совокупностей решений для задач планирования в условиях ресурсных ограничений и рисков // Прикаспийский журнал: управление и высокие технологии. – 2013, № 3 (23), с. 169-179.
6. Воронков Д.С. Проверка адекватности математической модели программного комплекса Locairheat для расчетов систем обеспечения теплового режима помещений воздушно-отопительными агрегатами // Современные проблемы науки и образования. – 2013, № 5. с 1-7.
7. ГОСТ Р 53778-2010. Здания и сооружения. правила обследования и мониторинга технического состояния.
8. ГОСТ Р 54317-2011. Комплексы стартовые и технические ракетно-космических комплексов. Требования безопасности.
9. Гуд А.Ю., Одейчук А.Н., Информационная технология построения моделей прогнозирования нелинейных временных рядов в условиях гетероскедастичности // Шамша, Восточно-Европейский журнал передовых технологий - 2010, №3, с. 58-61.
10. Елисеева И.И. Эконометрика. Учебник для бакалавриата и магистратуры. Издательство: М.:Издательство Юрайт - 2015, 449 с.

11. Королев М. Р., Лобейко В. И., Старусев А. В. Оценка адекватности комплексной модели сложной технической системы // Известия Волгоградского государственного технического университета - 2012, №13, с. 12-15.
12. Марова Н.В., Мешков Н.С., Тюрин М.В., Городнов В.И., Хохлова Г.Д. Низкочастотные линейные акселерометры прямого действия // Приборы и системы управления, 1990, №10, с.19-21.
13. Муха Ю.П., Королева И.Ю. Информационно-измерительные системы: Учебное пособие., Волгоградский государственный технический университет - 2015, 109 с.
14. Носков С.И., Торопов В.Д., Способ уточнения иерархической однородной математической модели статистического типа // Фундаментальные исследования - 2015, № 2, с. 2120 – 2123.
15. Паклин Н. Б. Бизнес-аналитика от данных к знаниям /Н.Б. Паклин, В.И. Орешков. – Санкт-Петербург : Питер - 2013, 704 с.
16. Прокофьев О.В., Савочкин, А.Е. Алгоритмическая модификация теста Чоу для автоматизированной проверки гипотезы о структурной стабильности тренда // XXI век: Итоги прошлого и проблемы настоящего плюс. Серия: Технические науки. Информационные технологии. – 2014, № 3 (19), с. 184–190.
17. Попов Г.А., Дедов А.В. Использование при принятии диагностических решений результатов статистического анализа влияния вирусно-бактериальных маркеров на клиническую картину заболевания // Прикаспийский журнал: управление и высокие технологии. – 2014, № 4, с. 119–133.
18. Прокофьев О.В., Савочкин, А.Е. Исследование изменения дисперсии остаточного компонента временного ряда // XXI век: Итоги прошлого и проблемы настоящего плюс. Серия: Технические науки. Информационные технологии. – 2013, № 10 (14), с. 135–144.
19. Прокофьев О.В., Сёмочкина И.Ю. Технологии и программные средства реализации интеллектуального анализа данных. Труды МНПК // Современные информационные технологии. - Пенза: Пензенский государственный технологический университет - 2013, № 17, с.31-35.
20. Савочкин, А.Е. Применение нейросетевого подхода при проектировании информационно-измерительных систем для определения степени повреждения технически сложных объектов // Прикаспийский журнал: управление и высокие технологии. – 2013, № 2 (22), с. 151–160.
21. Федоров Ю.Н. «Основы построения АСУТП взрывоопасных производств» В 2-х томах, [том 1] М.: Синтег - 2006, 255 с.
22. Федосеев В., Гармаш А., Орлова И., Экономико-математические методы и прикладные модели, Издательство: Юрайт - 2013, 336 с..
23. Chow G.C. Tests of Equality between Sets of Coefficients in Two Linear Regressions // Econometrica. – 1960. – Vol. 28, №3. – pp. 79–86.
24. Fisher R. A. On the mathematical foundations of theoretical statistics. Phil. Trans. Royal Soc. London A 222, 1922, pp.309–368.
25. Gosset W. S., Tables for estimating the probability that the mean of a unique sample lies etc. Biometrika, 1917, vol. 11, pp. 414 – 417.

References

1. Abzalov A.V., Zhedunov R.R. Metodika analiza predavarijnyh situacij na tehnologicheskikh ob"ektah upravlenija [The method of analysis preemergencies on engineering systems] // Prikaspijskij zhurnal: upravlenie i vysokie tehnologii [Caspian Journal: management and high technology]. – 2013, № 4 (24), pp. 50-58.
2. Altynov P.I., Bavrin I.I. Zvavich L.I. Matematika. Bol'shoj spravocnik dlja shkol'nikov i postupajushchih v vuzy [Mathematics. Great reference book for students and university entrants]. M.: Drofa 2009, 768 p.
3. Artemenko M.V., Babkov A.S. Klassifikacija metodov prognozirovaniya povedenija sistem [Classification of methods of predicting the behavior of systems] // Sovremennye problemy nauki i obrazovaniya [Modern problems of science and education]- 2013, № 6, pp. 13-19.
4. Arhipova A.V. Primenenie kriterija pikov dlja proverki adekvatnosti trendovoj modeli pri analize prjamyh inostrannyh investicij v RF [Application criteria peaks to check the adequacy of the model in the analysis of the trend in foreign direct investment in Russia] // Organizacija raboty s molodezh'ju [The organization works with youth]. – 2014, № 9, pp. 13-18.

5. Brumshtejn Ju.M., Tarkov D.A., Djudikov I.A. Analiz modelej i metodov vybora optimal'nyh sovokupnostej reshenij dlja zadach planirovanija v uslovijah resursnyh ogranichenij i riskov [Analysis of the models and methods of selection of optimum set of solutions for the problems of planning in terms of resource constraints and risks] // Prikaspijskij zhurnal: upravlenie i vysokie tehnologii [Caspian Journal: management and high technology]. – 2013, № 3 (23), pp. 169-179.
6. Voronkov D.S. Proverka adekvatnosti matematicheskoy modeli programmnogo kompleksa Locairheat dlja raschetov sistem obespechenija teplovogo rezhima pomeshhenij vozdušno-otopitel'nyimi agregatami [Check the adequacy of the mathematical model of software systems for calculations Locairheat thermal control rooms air-heating units] // Sovremennye problemy nauki i obrazovanija [Modern problems of science and education]. – 2013, № 5, pp 1-7.
7. GOST R 53778-2010. Zdanija i sooruzhenija. pravila obsledovanija i monitoringa tehničeskogo sostojanija [Buildings and constructions. Rules of inspection and condition monitoring].
8. GOST R 54317-2011. Kompleksy startovye i tehničeskie raketno-kosmicheskih kompleksov. Trebovanija bezopasnosti [Launching and technical complexes of rocket and space complexes. Safety requirements].
9. Gud A.Ju., Odejchuk A.N., Informacionnaja tehnologija postroenija modelej prognozirovaniya nelinejnyh vremennyh rjadov v uslovijah geteroskedastichnosti [Information technology build predictive models of nonlinear time series in Heteroskedasticity]// Shamsha, Vostočno-Evropejskij zhurnal peredovyh tehnologij [East European Journal of advanced technologies] - 2010, №3, pp. 58-61.
10. Eliseeva I.I. Jekonometrika. Uchebnik dlja bakalavriata i magistratury [Econometrics. The textbook for undergraduate and graduate]. Izdatel'stvo: M.:Izdatel'stvo Jurajt - 2015, 449 p.
11. Korolev M. R., Lobejko V. I., Starusev A. V. Ocenka adekvatnosti kompleksnoj modeli slozhnoj tehničeskoy sistemy [Assessing the adequacy of the integrated model of a complex technical system] // Izvestija volgogradskogo gosudarstvennogo tehničeskogo universiteta [News of Volgograd State Technical University] - 2012, №13, pp. 12-15.
12. Marova N.V., Meshkov N.S., Tjurin M.V., Gorodnov V.I., Hohlova G.D. Nizkochastotnye linejnye akselerometry prjamogo dejstvija [Low-frequency linear accelerometers direct] // Pribory i sistemy upravlenija [Instruments and Control Systems]. 1990, №10, pp.19-21.
13. Muha Ju.P., Koroleva I.Ju. Informacionno-izmeritel'nye sistemy: Uchebnoe posobie [Information-measuring system: Textbook] // Volgogradskij gosudarstvennyj tehničeskij universitet [Volgograd State Technical University] -2015, 109 p.
14. Noskov S.I., Toropov V.D., Sposob utocnenija ierarhicheskoj odnorodnoj matematicheskoy modeli statističeskogo tipa [A method of verifying a hierarchical uniform mathematical model of the type of statistical]// Fundamental'nye issledovanija [Basic Research] - 2015, № 2, pp. 2120 – 2123.
15. Paklin N. B. Biznes-analitika ot dannyh k znaniyam [Business analytics from data to knowledge]. Sankt-Peterburg : Piter - 2013, 704 p.
16. Prokof'ev O.V., Savochkin, A.E. Algoritmicheskaja modifikacija testa Chou dlja avtomatizirovannoj proverki gipotezy o strukturnoj stabil'nosti trenda [Algorithmic modification Chow test for automated test the hypothesis of structural stability trend] // XXI vek: Itogi proshlogo i problemy nastojashhego pljus. Serija: Tehničeskie nauki. Informacionnye tehnologii [XXI Century: The results of the past and the present problems of a plus. Series: Engineering. Information technology]. – 2014, № 3 (19), pp. 184–190.
17. Popov G.A., Dedov A.V. Ispol'zovanie pri prinjatii diagnostičeskikh reshenij rezul'tatov statističeskogo analiza vlijanija virusno-bakterial'nyh markerov na kliničeskiju kartinu zabojevanija [Use in making diagnostic decisions of the results of statistical analysis of the impact of viral and bacterial markers in the clinical picture of the disease] // Prikaspijskij zhurnal: upravlenie i vysokie tehnologii [Caspian Journal: management and high technology]. – 2014, № 4, pp. 119–133.
18. Prokof'ev O.V., Savochkin, A.E. Issledovanie izmenenija dispersii ostatochnogo komponenta vremennogo rjada [Study changes in the variance of the residual component of the time series] // XXI vek: Itogi proshlogo i problemy nastojashhego pljus. Serija: Tehničeskie nauki. Informacionnye tehnologii [XXI Century: The results of the past and the present problems of a plus. Series: Engineering. Information technology]. – 2013, № 10 (14), pp. 135–144.
19. Prokof'ev O.V., Sjomochkina I.Ju. Tehnologii i programmnye sredstva realizacii intellektual'nogo analiza dannyh. Trudy MNPK [Technology and software implementation of data mining. Proceedings of PBMCs] // Sovremennye informacionnye tehnologii. Penzenskij gosudarstvennyj tehnologičeskij universitet [Modern information technology. Penza State Technological University]. - 2013, № 17, pp.31-35.

20. Savochkin, A.E. *Primenenie nejrosetevogo podhoda pri proektirovanii informacionno-izmeritel'nyh sistem dlja opredelenija stepeni povrezhdenija tehnichecki slozhnyh ob"ektov* [Application of neural network approach in the design of information-measuring systems to determine the extent of damage is technically complex objects] // *Prikaspijskij zhurnal: upravlenie i vysokie tehnologii* [Caspian Journal: management and high technology]. – 2013, № 2 (22), pp. 151–160.

21. Fedorov Ju.N. *Osnovy postroenija APCS vzryvoopasnyh proizvodstv* [Fundamentals of process control hazardous areas]. V 2-h tomah, tom 1 M.: Sinteg. 2006, 255 p.

22. Fedoseev V., Garmash A., Orlova I., *Jekonomiko-matematicheskie metody i prikladnye modeli* [Economic-mathematical methods and applied models], Izdatel'stvo: Jurajt 2013, 336 p.

23. Chow G.C. Tests of Equality between Sets of Coefficients in Two Linear Regressions *Econometrica*. – 1960. – Vol. 28, №3. – pp. 79–86.

24. Fisher R. A. On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc. London A* 222, 1922, pp.309–368.

25. Gosset W. S., Tables for estimating the probability that the mean of a unique sample lies etc. *Biometrika*, 1917, vol. 11, pp. 414 – 417.

УДК 330.43, 339.743.44, 519.246.85

ИСПОЛЬЗОВАНИЕ МОДЕЛЕЙ ARIMA ДЛЯ ПРОГНОЗИРОВАНИЯ ВАЛЮТНОГО КУРСА

Статья поступила в редакцию 05.10.2015 г., в окончательном варианте 13.11.2015 г.

Пилюгина Анна Валерьевна, кандидат экономических наук, доцент, Московский государственный технический университет им. Н. Э. Баумана, 105005, Российская Федерация, г. Москва, ул. 2-ая Бауманская, 5, стр. 1, e-mail: pilyuginaanna@mail.ru

Бойко Андрей Алексеевич, аспирант, Московский государственный технический университет им. Н. Э. Баумана, 105005, Российская Федерация, г. Москва, ул. 2-ая Бауманская, 5, стр. 1, e-mail: boiko_andrew@mail.ru

Описаны модель ARIMA и методика ее применения для краткосрочного прогнозирования среднемесячного курса доллара США к российскому рублю. Приведена классификация временных рядов (ВР) с точки зрения их стационарности. Описана процедура проверки ВР на стационарность с помощью критерия Дики–Фуллера. Показано, что рассматриваемый ВР среднемесячного валютного курса USD / RUB с января 2009 г. по август 2015 г. является «интегрированным первого порядка». Операция дифференцирования обращает среднемесячный курс в стационарный в широком смысле ВР. Для определения порядка авторегрессии p и скользящего среднего q использован оптимизационный поиск на сетке $[0..5; 0..5]$. В качестве критерия оптимизации использованы байесовский информационный критерий (BIC) и информационный критерий Акаике (AIC). По данным критериям оптимальной признана модель $ARIMA(2,1,5)$. Кроме того, предложен алгоритм адаптивной идентификации модели с уточнением порядка p и величины q , а также соответствующих коэффициентов при поступлении каждого следующего отсчета. Однако оценка точности прогноза показала, что для минимизации средней относительной ошибки прогноза ($MAPE$) следует использовать самые простые модели – а именно $ARIMA(1,1,0)$ и $ARIMA(0,1,1)$.

Ключевые слова: валютный курс, прогнозирование, временной ряд, стационарный относительно детерминированного тренда ряд, стационарный относительно взятия разностей ряд, модель $ARIMA$, байесовский информационный критерий (BIC), информационный критерий Акаике (AIC), средняя относительная ошибка прогноза ($MAPE$)