УДК 004.942

В.Ф. Миргород, Е.В. Деренг, И.М. Гвоздева

ИНТЕРВАЛЬНАЯ ОЦЕНКА ТРЕНДОВОЙ КОМПОНЕНТЫ ВРЕМЕННЫХ РЯДОВ

Аннотация. В работе обосновывается подход к интервальной оценке трендовой компоненты временных рядов, образованных параметрами регистрации состояния силовых и энергетических установок в их длительной эксплуатации. Для многомерных массивов данных последовательно используются известные методы сингулярного спектрального и трендового анализа. Для временных рядов, матрица автокорреляции которых имеет превалирующее собственное значение, предложена интервальная оценка трендовой компоненты. Установлены свойства матрицы автокорреляции временного ряда. Решена задача интервальной оценки трендовой компоненты для временного ряда с линейным трендом.

Ключевые слова: временной ряд,статистическая модель, трендовый анализ, корреляционная матрица, интервальная оценка

Введение. Повышение надежности статистических выводов о техническом состоянии диагностируемых объектов является основной проблемой как для теории, так и для практики применения систем технической диагностики (СТД). Ее решение достигается на основе использования и развития методов трендового анализа временных рядов, образуемых регистрируемыми параметрами (измеряемыми переменными состояния и выходными переменными).

Задача выделения трендов в указанных временных рядах, установления их интервальных оценок на заданном уровне значимости имеет важное научно-прикладное значение, в частности, применительно к СТД силовых и энергетических установок на основе газотурбинных двигателей (ГТД), как общепромышленного, так и авиационного назначения.

Основная часть. Предметом исследования являются статистические модели (СМ) порождения данных, методы трендового анализа, позволяющие установить закономерности развития трендовой компоненты и ее доверительные интервалы [2,4,6].

Известные методы трендового контроля [4,6] позволяют установить лишь факт отсутствия тренда на заданном уровне значимости, поскольку именно так формулируется опорная гипотеза [9]. Опыт их применения свидетельствует о недопустимо высоком уровне ошибок как первого, так и второго рода. Методы трендового анализа, позволяющие непосредственно выделить трендовую компоненту временного ряда, дают возможность повысить надежность статистических выводов о техническом состоянии диагностируемых объектов. Однако зачастую объем выборки не является значительным, применяемые статистики являются выборочными, поэтому остается нерешенным вопрос об уровне значимости выделенной трендовой компоненты и доверительных интервалах ее оценки.

Целью настоящего исследования является обоснование подхода к интервальной оценке трендовой компоненты временного ряда, образованного совокупностью отклонений параметров регистрации состояния объекта диагностирования от его СМ, и установления уровня ее статистической значимости.

Центральной гипотезой исследования является предположение о том, что в процессе эксплуатации сложных энергетических объектов происходит постепенная и естественная деградация их характеристик, что неизбежно приводит к наличию в регестрируемых данных долговременного тренда параметров. Поэтому только отклонение от СМ позволяет получить информацию о техническом состоянии объекта. Подчиненная гипотеза состоит в том, что повышение надежности статистических выводов о техническом состоянии диагностируемых объектов достигается путем интервальной оценки трендовой компоненты на заданном уровне значимости.

Следуя методу “Гусеница” [3], полная информация об изменении временного ряда содержится в траекторной матрице X_t размером $n \times k$ [3]:

ISSN 1562-9945

$$\begin{aligned}
&\vec{x}_1 = [x_1 \;\; x_2 \; \ldots \qquad\quad x_n] \\
&\vec{x}_2 = [x_2 \;\; x_3 \; \ldots \qquad x_{n+1}] \\
&\ldots \\
&\vec{x}_k = [x_k \;\; x_{k+1} \ldots \; x_{n+k-1}]
\end{aligned} \tag{1}$$

В качестве СМ порождения данных предлагается следующая модель совокупности трендовой и шумовой компонент [6, 7]:

$$\vec{x}_k = [x_k \;\; x_{k+1} \;\; x_{k+2} \ldots x_{k+n-1}] = \vec{x}_{tr} + \vec{x}_{noise}. \tag{2}$$

Задача анализа выборки данных (1) состоит в разделении ее строк в виде (2) на заданном уровне значимости. Решение указанной задачи достигается последовательной реализацией следующих этапов:

1. Выполняется решение задачи на собственные значения

$$X_t X_t^T \vec{u}_i = \lambda_i \vec{u}_i, \tag{3}$$

где столбцы $\vec{u}_i$ образуют матрицу U из ортогональных векторов матрицы $X_t X_t^T$.

2. Определяется [1,5] матрица главных компонент временного ряда

$$F = U^T X_t, \tag{4}$$

где ее строки упорядочены по убыванию собственных чисел матрицы $X_t X_t^T$.

3. Выполняется разложение строк матрицы (1) по главным компонентам (4)

$$\vec{x}_s = \sum_{i=1}^{k} b_{si} \vec{f}_i, \tag{5}$$

где $s = \overline{1,k}$, b_{si} – коэффициенты влияния, определяемые решением переопределенной $(n > k)$ системы линейных алгебраических уравнений:

$$F^T \vec{b}_s = \vec{x}_s. \tag{6}$$

4. Отыскивается решение (6) с использованием псевдообратной матрицы [5]:

$$\vec{b}_s^T = \left(FF^T\right)^{-1} F \vec{x}_s^T. \tag{7}$$

Так как $\left(FF^T\right) = diag\{\lambda_i\}$, $i = \overline{1,k}$, то из (7) следует

$$b_{sj} = \vec{x}_s \vec{f}_j^{\,T} \Big/ \left(\vec{f}_j \vec{f}_j^{\,T} \right) = \lambda_j^{-1} \vec{x}_s \vec{f}_j^{\,T}, \qquad (8)$$

где $\vec{f}_j$ – строки матрицы главных компонент. Если собственное число $\lambda_1 = \lambda_{\max}$ соответствует трендовой компоненте, то согласно (8), получаем

$$\vec{x}_{tr,s} = b_{s1} \vec{f}_1, \qquad (9)$$

После выделения трендовой компоненты (9) выполняется установление статистических свойств остатка $\vec{x}_{noise}$. Если остаточный временной ряд удовлетворяет критериям принадлежности выборки из генеральной совокупности нормально распределенных случайных величин на заданном уровне значимости, то разделение ряда (2) выполнено корректно. Если при этом СКО остаточной выборки не превышает погрешностей измерений, то разделение ряда соответствует физическим особенностям процессов.

Как свидетельствует опыт практического применения [6,7,8] рассмотренного подхода выделения трендовой компоненты временного ряда, если в ряде присутствует статистически значимый тренд, то он связывается с максимальным собственным числом матрицы автокорреляций $X_t X_t^T$, которое существенно превосходит остальные собственные числа. В связи с этим может быть предложена гипотеза, что такое условие (существенное различие первого собственного значения от всех иных) может полагаться собственно условием наличия статистически значимого тренда в выборке.

Для подтверждения предлагаемой гипотезы и установления условий ее применения первоначально рассмотрим следующую Лемму.

Лемма. Если элементы r_{ij} матрицы R таковы, что

$$r_{ii} = 1, i = \overline{1,k},$$
$$r_{ij} = r, i \neq j,$$

то ее собственные значения имеют вид:

$$\lambda_1 = 1 + (k-1) r,$$
$$\lambda_i = 1 - r,$$

а элементы первого собственного вектора $\vec{u}_1$, соответствующего первому собственному значению, равны друг другу: $u_{11} = u_{12} = u_{13} ... = u_{1k}$.

 ISSN 1562-9945

Следствие. Если выполнено условие нормировки $\vec{u}_1\vec{u}_1^T = 1$, то элементы первого собственного вектора $\vec{u}_1$, соответствующего первому собственному значению, равны

$$u_{11} = u_{12} = u_{13} ... = u_{1k} = 1/\sqrt{k}.$$

Доказательство Леммы. Для доказательства рассмотрим уравнение $\det(\lambda E - R) = 0$ и установим, что в условиях Леммы справедливы следующие равенства

$$SpR = \sum_{i=1}^{k} \lambda_i,$$

$$\det R = \mathop{\Pi}_{i=1}^{k} \lambda_i.$$

Очевидно, что первое условие выполняется:

$$SpR = \sum_{i=1}^{k} \lambda_i = \lambda_1 + (k-1)\lambda_i = 1 + (k-1)r + (r-1)(1-r) = k.$$

Второе условие означает, что

$$\det R = \mathop{\Pi}_{i=1}^{k} \lambda_i = [1 + (k-1)r](1-r)^{k-1}. \qquad \textbf{(10)}$$

Вычтем из второй строки матрицы R первую, из третьей – вторую и т.д. В результате получим следующую матрицу

$$\begin{bmatrix} r-1 & 1-r & 0 & 0 ... & 0 \\ 0 & r-1 & 1-r & 0 ... & 0 \\ ... & & & & \\ r & r & r & r ... & 1 \end{bmatrix}$$

Определитель матрицы при этом не изменится. В полученной матрице прибавим все ее столбцы к первому:

$$\begin{bmatrix} 0 & 1-r & 0 & 0 ... & 0 \\ 0 & r-1 & 1-r & 0 ... & 0 \\ ... & & & & \\ 1+(k-1)r & r & r & r ... & 1 \end{bmatrix}$$

Раскрывая определитель полученной матрицы по первому столбцу, получим окончательно:

$$\det R = [1 + (k-1)r](1-r)^{k-1},$$

что полностью совпадает с выражением (10).

Так как сумма элементов любой из строк матрицы R равна первому собственному числу, то элементы первого собственного вектора $\vec{u}_1$, соответствующего первому собственному значению, равны друг другу. Лемма доказана.

Утверждение. Пусть задан временной ряд и его траекторная матрица (1). Если траекторная матрица удовлетворяет гипотезе о равнокоррелированности ее строк, то первая центрированная компонента (9) временного ряда по первой главной компоненте является скользящим средним этого ряда.

Такая гипотеза опровергается на заданном уровне значимости распределения χ^2 решающей статистикой коррелированности признаков (строк траекторной матрицы). Если, кроме того, шумовая компонента удовлетворяет условиям выборки из генеральной совокупности независимых нормально распределенных случайных величин, то доверительный интервал трендовой компоненты определяется интервальной оценкой среднего.

Если нормированная матрица автокорреляций $X_t X_t^T$ не принадлежит классу матриц R, но есть основания предполагать статистическую обоснованность такого предположения, то следует воспользоваться известной [1] статистикой коррелированности признаков.

Гипотеза коррелированности (нормированная матрица автокорреляций $X_t X_t^T$ принадлежит классу матриц R) опровергается при выполнении условия:

$$\zeta = (A - BC)(n-1)/(1-r)^2 > \chi^2(\alpha, N), \tag{11}$$

где α - уровень значимости,

$N = (k+1)(k-1)/2$ - число степеней свободы,

$$A = \sum_{\substack{i,j=1 \\ i \neq j}}^{k} \left(r_{ij} - r\right)^2,$$

$$B = \left(\sum_{i=1}^{k} \left(r_i - r\right)^2\right),$$

$$C = r(2-r)(k-1)^2 / \left[k - (k-2)(1-r)^2\right],$$

 ISSN 1562-9945

$$r = \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{k} r_{ij} \right) / \left(k \left(k-1 \right) \right),$$

$$r_i = \left(\sum_{\substack{j=1 \\ i \neq j}}^{k} \left(r_j \right) \right) / \left(k-1 \right),$$

r_{ij} - элементы нормированной матрицы автокорреляций $X_t X_t^T$.

Как показывает опыт практического применения, статистика (11) отвечает реальным временным рядам, если средний строчный коэффициент корреляции близок к 1 (как это имеет место в примере, рассмотренном в [1]), или к 0 (некоррелированность ряда, т.е. отсутствие тренда). В том случае, когда средний строчный коэффициент корреляции имеет промежуточное значение, может быть использован тест Стьюдента равенства средних применительно к строкам нормированной матрицы автокорреляций.

Для оценки условий применимости предлагаемого подхода выполнено тестовое компьютерное моделирование. Компьютерный эксперимент по численной оценке достоверности теоретических выводов заключался в том, что генерировался временной ряд по закону

w=(-0.5:0.005:0.585); wd=-0.5*w; w1=wd+0.1*randn(1,218)

и выполнялась оценка тренда для десяти независимых реализаций длиной 198 отчетов каждая, с окном в 20 отсчетов, двумя методами в сопоставлении с интервальной оценкой. Результаты компьютерного эксперимента приведены на рис. 1 и рис. 2. Интервалы скользящего среднего установлены на уровне доверительной вероятности 0,95. Как это следует из результатов обработки данных компьютерного эксперимента и видно на иллюстрациях, интервальная оценка тренда является статистически достоверной на заданном уровне значимости.

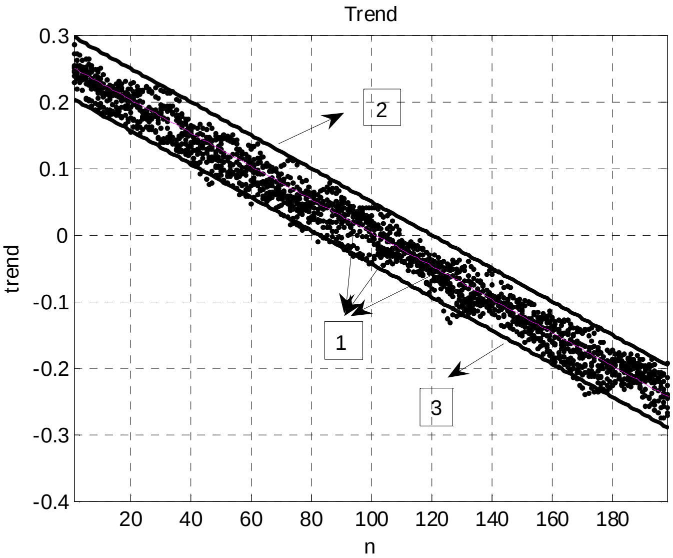

Рисунок 1 – Интервальная оценка тренда методом SSA:
1 – тренды, 2 – верхний доверительный интервал,
3 – нижний доверительный интервал

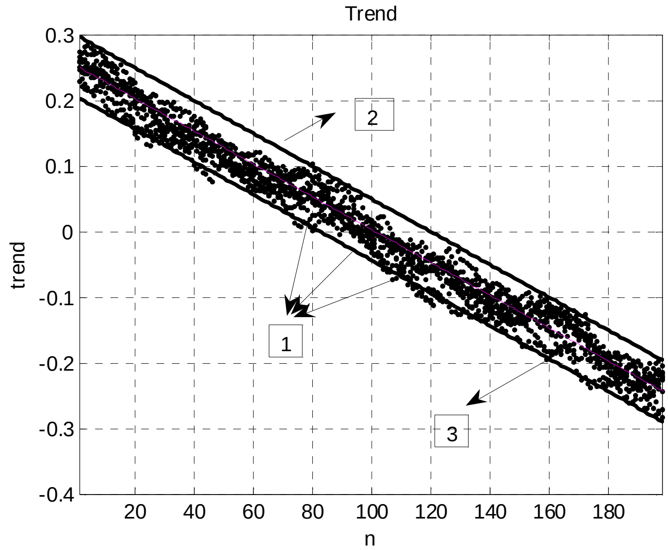

Рисунок 2 – Интервальная оценка тренда методом скользящего среднего: 1 – тренды, 2 – верхний доверительный интервал,
3 – нижний доверительный интервал

Заключение. Последовательное сочетание методов выделения трендовой компоненты и вероятностной оценки ее доверительных интервалов позволяет расширить признаковое пространство принятия решений и, тем самым, повысить надежность диагностических выводов о техническом состоянии сложных энергетических объектов. Такое расширение возможно на основе формирования многомерных массивов из данных регистрации технического состояния исследуемо-

 ISSN 1562-9945

го объекта и их аппроксимации методом главных компонент. Для временных рядов, матрица автокорреляции которых имеет превалирующее собственное значение, может быть найдена интервальная оценка трендовой компоненты.

Перспективы дальнейших исследований заключаются в разработке методов прогноза интервальной оценки трендов.

ЛИТЕРАТУРА

1. Айвазян С.А. Прикладная статистика. Классификация и снижение размерности [Текст] / В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.
2. Бендат Дж. Прикладной анализ случайных данных [Текст] / Дж. Бендат, А. Пирсон – М.: Мир, 1989. – 540 с.
3. Главные компоненты временных рядов: метод “Гусеница” [Текст] / Под ред. Д.Л. Данилова, А.А. Жиглявского. – С.-П. ун-т. – 1997.
4. Епифанов С.В. Синтез систем управления и диагностирования газотурбинных двигателей [Текст] / С.В. Епифанов, В.И. Кузнецов, И.И. Богаенко и др. // – К.: Техника, 1998. – 312 с.
5. Марпл мл. С.Л. Цифровой спектральный анализ и его приложения [Текст] / С.Л. Марпл мл. – М.: Мир, 1990. – 584 с.
6. Миргород В.Ф. Применение диагностических моделей и методов трендового анализа для оценки технического состояния газотурбинных двигателей [Текст] / В.Ф. Миргород, Г.С. Ранченко, В.М. Кравченко // Авіаційно-космічна техніка і технологія. – 2008. – 9(56) – С. 192-197.
7. Миргород В.Ф. Трендовый анализ на основе диагностических параллелепипедов [Текст] / В.Ф. Миргород, И.М. Гвоздева // Системні технології. Регіональний міжвузівський збірник наукових праць. – Дніпропетровськ.– Вип. 3(80).– 2012. – С. 97-104.
8. Деренг Е.В. Комбинированный метод ТАТ обработки многомерных временних рядов [Текст] / Е.В. Деренг, И.М. Гвоздева, В.Ф. Миргород // Системні технології. Регіональний міжвузівський збірник наукових праць. – Дніпропетровськ.– 2013. – Вип. 4(87). – С. 21-27.
9. Perron P. Trend and Random Walks in Macroeconomic Time Series: Furter Evidence from a New Approach [Text] / P. Perron. – Journal of Economic Dynamic and Control. – No. 12. – P. 297-332.