

Использование временных рядов для повышения надёжности статистических характеристик контролируемого параметра при прогнозировании

Ю. С. Сысоев, О. Ф. Цуверкалова

Волгодонский инженерно-технический институт – филиал Национального исследовательского ядерного университета «МИФИ», Волгодонск, Россия, e-mail: sysoev2004@mail.ru

На основе анализа временных рядов, составленных из значений контролируемого параметра, предложены варианты построения векторов (векторов-соседей), близких к стартовому вектору. Показано, что использование координат построенных векторов-соседей совместно с координатами стартового вектора позволяет увеличить объёмы выборок для определения применяющихся при прогнозе дрейфа параметра статистических характеристик и повысить надёжность оценок этих характеристик.

Ключевые слова: временные ряды, методы прогнозирования, дрейф параметров, векторы-соседи.

При изучении или контроле объектов различной природы на практике бывает сложно, а порой и невозможно построить аналитические модели этих объектов в виде оптимизационных задач, некоторых алгебраических, дифференциальных или других уравнений или неравенств, решаемых аналитическими или численными методами. Тогда для решения поставленных задач используют динамические системы, связанные с анализом временных рядов, в частности, скалярных временных рядов [1–4]. Такие ряды получают, фиксируя через некоторые временные промежутки числовые значения параметров, которые характеризуют работу контролируемого объекта. Изучение временных рядов способствует идентификации объектов и их динамики, а также диагностике данных объектов и прогнозированию состояний этих объектов. Прогнозирование состояний объектов осуществляют, контролируя и прогнозируя дрейф указанных параметров [1, 2, 5, 6]. Таким образом, состояние контролируемого объекта связано с текущими состояниями измеряемого параметра (например, дрейфа), по изменениям которого отслеживают и прогнозируют изменения состояний объекта в целом.

Для прогнозирования дрейфа параметра применяют, как правило, последние k значений (при некотором целом значении k) временного ряда. Эти значения представлены в виде вектора, называемого стартовым [2]. По последней координате стартового вектора или последнему значению временного ряда (что одно и то же) определяют текущее значение параметра в момент последнего измерения, а остальные координаты этого вектора используют для выявления динамики значений параметра на соответствующем промежутке.

В настоящей статье характеристикой надёжности выступает вероятность близости статистических и соответствующих фактических характеристик объекта. Из закона больших чисел следует, что данная вероятность растёт с увеличением объёма выборки, поэтому в представленной работе рассматриваются вопросы увеличения объёмов соответствующих выборок с указанием метода их получения, позволяющего избежать локализации этих выборок по времени.

Допустим, что измерения значений параметра (его мо-

нитинг) проводились через равные промежутки Δt в моменты времени $t_1 < t_2 < \dots < t_N$. В результате измерений получен временной ряд $x_1, x_2, \dots, x_N$, стартовым вектором которого является вектор

$$x_{N-k+1} = (x_{N-k+1}, x_{N-k+2}, \dots, x_N).$$

В дальнейшем предполагается, что на контролируемый объект действуют случайные возмущения, поэтому числовые значения, составляющие временной ряд, будем считать случайными. С учётом этого допущения для анализа динамики координат стартового вектора рассмотрим случайную величину ξ , которую определим ниже. Для упрощения изложения положим, что значения, составляющие временной ряд, положительны. Если это не так, то к каждому из указанных значений добавим одно и то же положительное число такое, чтобы для нового ряда указанное требование выполнялось. Работу продолжим с полученным временным рядом.

Обозначим через Δ любую из верхних границ значений анализируемого временного ряда. Разобьём промежутки $[0; \Delta)$ на N полуинтервалов длины $\Delta Z = \Delta/N$ каждый: $[0; \Delta Z)$, $[\Delta Z; 2\Delta Z)$, $[2\Delta Z; 3\Delta Z)$, ..., $[(N-1)\Delta Z; N\Delta Z)$. Рассмотрим бесконечную последовательность полуинтервалов $[0; \Delta Z)$, $[\Delta Z; 2\Delta Z)$, ..., $[(N-1)\Delta Z; N\Delta Z)$, $[N\Delta Z; (N+1)\Delta Z)$, Эти полуинтервалы определяют состояния контролируемого объекта.

Будем считать, что объект находится в состоянии E_j , $j=0, 1, 2, \dots$, если его параметр лежит в промежутке $[j\Delta Z; (j+1)\Delta Z)$. Поскольку изменения параметра носят случайный характер, то на состояния объекта можно смотреть как на случайные события. Определим смещение состояния объекта при переходе от состояния временного ряда, порождённого значением

$$x_s, s=N-k+1, \dots, N-1, \quad (1)$$

к состоянию, порождённому значением x_{s+1} .

Если значение x_s порождает состояние E_r , а x_{s+1} – состояние E_{j+r} , где r – некоторое целое число, то это число назовём смещением при переходе от значения x_s к значению x_{s+1} . Число r может быть отрицательным, положительным или нулём. Если значения координат стартового вектора обозначить согласно формуле (1) а l_s – смещение состояния параметра (объекта) при переходе во временном ряде от значения x_s к значению x_{s+1} , то с учётом приведённого определения и введённых обозначений для координат стартового вектора $\mathbf{x}_{N-k+1}$ составим табл. 1.

На основании табл. 1 построим ряд распределения или статистический ряд случайной величины ξ . Эта случайная величина принимает значения, равные смещениям координат стартового вектора [7]. Так как ряд распределения строится по координатам стартового вектора $\mathbf{x}_{N-k+1}$, то случайную величину ξ далее обозначим ξ_{N-k+1} .

Для построения ряда распределения случайной величины ξ_{N-k+1} введём следующие обозначения: $\omega_r=r$ – частота смещения состояний параметра (координат стартового вектора); p_r – вероятность того, что случайная величина примет значение числа r , равного смещению в стартовом векторе. Тогда ряд распределения случайной величины ξ_{N-k+1} можно представить в виде табл. 2.

Ряд распределения, представленный в табл. 2, будет конечным. Из табл. 2 следует, что случайная величина ξ_{N-k+1} определяет динамику изменения координат стартового вектора, т. е. динамику изменений состояний объекта в пределах временного промежутка $[t_{N-k+1}, t_N]$. В некоторых случаях при прогнозировании (фактически, на основе табл. 1) строят две функции распределения, характеризующие указанную динамику состояний объекта [6]. Однако для построения функции распределения необходимо иметь достаточное количество статистических данных. В описанном случае в качестве статистических данных выступают координаты стартового вектора, которых, как правило, недостаточно для получения надёжных статистических результатов. Например, в работе [8] для получения статистической дисперсии, которая с большой вероятностью будет близка к её истинному значению, необходимо иметь статистическую выборку объёма не менее 100 данных, хотя для некоторых статистических характеристик даже выборки, содержащие 300 данных, не всегда дают удовлетворительный результат.

Учитывая сказанное выше, в работе [2] для увеличения объёмов соответствующих выборок предложено использовать координаты не только стартового вектора, но и координаты векторов-соседей, т. е. векторов, в некотором смысле близких стартовому. Необходимо заранее определить численную оценку степени близости данных векторов к стартовому вектору, а именно – выяснить, меньше какой априори выбранной малой величины $\varepsilon > 0$ должно быть указанное расстояние. Необходимо, чтобы найденную численную оценку можно было применить с гарантией, что такие векторы существуют, однако сделать это достаточно сложно. Поэтому для определения векторов-соседей авторы настоящей статьи предлагают использовать уравнение, которое можно рассматривать и как оптимизационную за-

Т а б л и ц а 1
Таблица смещений для координат стартового вектора

Параметр	Значение						
Координата x_s	x_{N-k+1}	x_{N-k}	x_{N-k-1}	$\dots$	x_i	$\dots$	x_{N-1}
Смещение l_s	l_{N-k+1}	l_{N-k}	l_{N-k-1}	$\dots$	l_i	$\dots$	l_{N-1}

дачу [2]. Зададим заранее требуемое число m векторов-соседей, соседних стартовому вектору $\mathbf{x}_{N-k+1}$, и вычислим те m векторов $\mathbf{x}_{s1}, \mathbf{x}_{s2}, \dots, \mathbf{x}_{sm}$, которые удовлетворяют уравнению

$$\sum_{i=1}^m \rho(\mathbf{x}_{N-k+1}, \mathbf{x}_{s_i}) = \min_{Q \in \Omega} \left[\sum_{s \in Q} \rho(\mathbf{x}_{N-k+1}, \mathbf{x}_s) \right],$$

где Q – произвольный набор номеров длины m ; Ω – множество всех наборов номеров значений временного ряда длины m , кроме номеров координат стартового вектора; $\rho(\mathbf{x}_{N-k+1}, \mathbf{x}_s)$ – расстояние между $\mathbf{x}_{N-k+1}, \mathbf{x}_s$ при любом s .

Проведённый в работе [9] анализ определения понятия векторов-соседей показал, что если векторов, близких к стартовому вектору, окажется меньше m , то может случиться так, что какие-то векторы-соседи окажутся далекими от стартового вектора. В этом случае их применение при прогнозировании для определения статистических характеристик (включая соответствующие законы распределения) окажется нецелесообразным. Для устранения указанного недостатка авторы предлагают расширить понятие векторов-соседей описанным ниже способом.

При прогнозировании важно учитывать не только текущее состояние системы, определяющееся последней координатой стартового вектора (последним измеренным значением параметра или последним значением временного ряда), но и динамику изменений параметра. Следовательно, при известном текущем состоянии необходимо оценить динамику изменений состояний объекта, иными словами – построить табл. 1. При этом важны не столько значения временного ряда, сколько смещения состояния объекта при переходе в этом ряде от значения x_s к значению x_{s+1} . Для увеличения количества векторов, близких к стартовому, из которых далее надо выбрать векторы-соседи, в работе [9] авторы предлагают векторы-соседи не просто выбирать, но и строить. Для этого нужно использовать соответствующие векторы, полученные из последовательных значений временного ряда.

Т а б л и ц а 2
Ряд распределения случайной величины ξ_{N-k+1}

ξ_{N-k+1}	Значения ряда распределения							
	$\dots$	-2	-1	0	1	2	3	$\dots$
ω	$\dots$	ω_{-2}	ω_{-1}	ω_0	ω_1	ω_2	ω_3	$\dots$
p	$\dots$	p_2	p_1	p_0	p_1	p_2	p_3	$\dots$

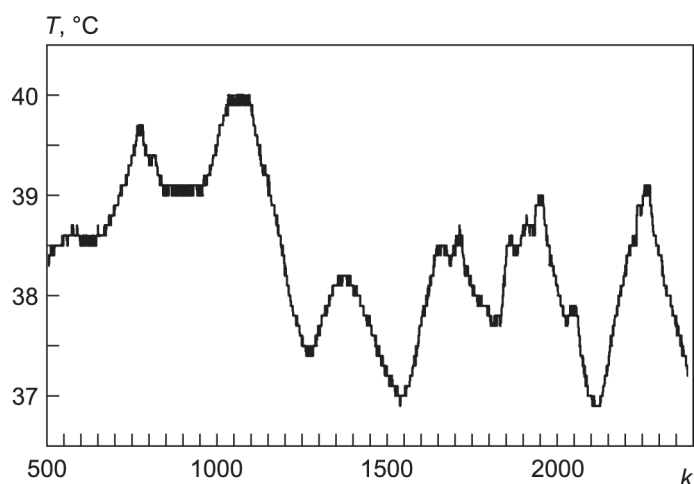


Рис. 1. График функции, определяющей температуру масла в масло-системе главного циркуляционного насоса: k – номер наблюдения; T – температура масла

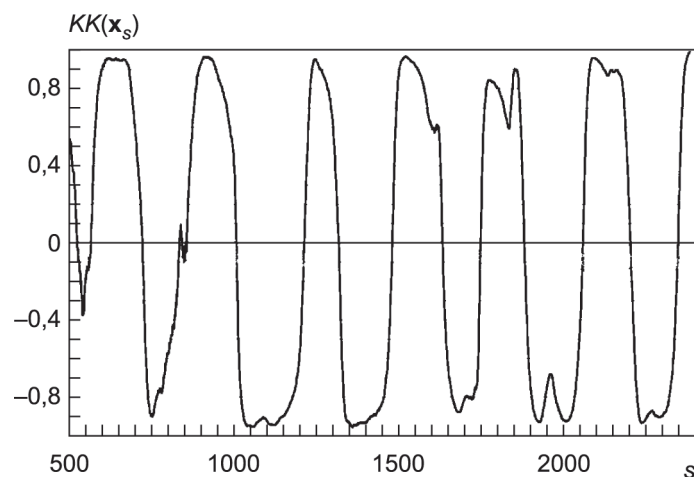


Рис. 2. Фрагмент графика функции $K=KK(x_s)$, определяющей коэффициент корреляции векторов $x_s = (x_s, x_{s+1}, \dots, x_{s+k-1})$, $s=1, 2, \dots, N-k+1$ и стартового вектора: s – номер вектора; $KK(x_s)$ – коэффициент корреляции

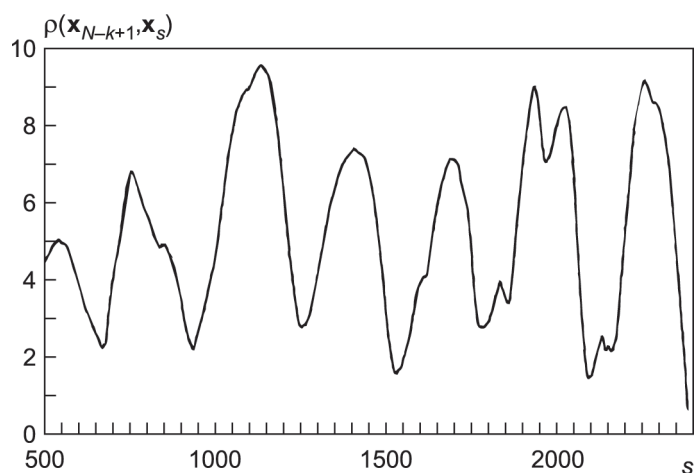


Рис. 3 – Фрагмент графика функции $\rho = \rho(x_{N-k+1}, x_s)$, задающей расстояния от векторов x_s , $s=1, 2, \dots, N-k+1$ до стартового вектора: s – номер вектора; $\rho(x_{N-k+1}, x_s)$ – расстояние до стартового вектора

Тогда для построения семейства векторов-соседей $y_{s1}, y_{s2}, \dots, y_{sm}$ стартового вектора нужно решить следующее уравнение:

$$\sum_{i=1}^m \rho(x_{N-k+1}, y_{si}) = \min_{Q \in \Omega} \left[\sum_{s \in Q} \rho(x_{N-k+1}, y_s) \right]. \quad (2)$$

В уравнении (2) вместо векторов x_s, x_{si} задействованы векторы y_s, y_{si} . Данные векторы найдены из векторов x_s, x_{si} с помощью их оптимальных сдвигов на постоянные векторы вида $c_s = (c_s, c_s, \dots, c_s)$ таким образом, чтобы расстояние от полученных после сдвига векторов до стартового вектора было минимальным.

Для определения векторов-соседей далее в работе применяется уравнение (2). Поэтому для удобства и краткости изложения везде, где говорится о расстоянии или близости произвольного вектора x_s к стартовому или другому вектору, подразумевается расстояние или близость соответствующего вектора y_s . Точно также, когда речь идёт о семействе $x_{s1}, x_{s2}, \dots, x_{sm}$ векторов-соседей, имеется в виду соответствующее семейство векторов $y_{s1}, y_{s2}, \dots, y_{sm}$, построенное по векторам $x_{s1}, x_{s2}, \dots, x_{sm}$.

Отметим, что для определения расстояния $\rho(x_{N-k+1}, x_s)$ в дальнейшем используем евклидову метрику, а в качестве единиц измерения – не обычные линейные единицы измерения, а единицы, в которых выражаются значения параметра, порождающего временной ряд.

Рассмотрим семейство всех векторов типа

$$x_s = (x_s, x_{s+1}, \dots, x_{s+k-1}), \quad s=1, 2, \dots, N-k+1.$$

Количество координат этих векторов такое же, как и стартового вектора. Найдём коэффициенты корреляции между этими векторами и стартовым вектором. Тем самым определим функцию $K=KK(x_s)$ дискретного аргумента. Естественно ожидать, что если вектор x_{s0} – близкий к стартовому, то коэффициент корреляции $K=KK(x_{s0})$ будет близким к единице, так как в этом случае выполняется необходимое условие близости к стартовому вектору. Поэтому близкие векторы, т. е. векторы-соседи целесообразно искать среди векторов x_s , в которых функция $K=KK(x_s)$ имеет локальные максимумы. Этот факт можно подтвердить соответствующим численным экспериментом, проведённым авторами.

В качестве примера рассмотрим прогнозирование изменения температуры масла в маслосистеме главного циркуляционного насоса. На рис. 1 приведен фрагмент графика температуры масла, построенный по результатам наблюдений. Мониторинг этой маслосистемы порождает временной ряд функции $K=KK(x_s)$, представленный на рис. 2, и функции, задающей расстояния от векторов x_s , $s=1, 2, \dots, N-k+1$ до стартового вектора, фрагмент графика которой приведен на рис. 3. В табл. 3 представлены расстояния от точек x_s максимума функции $K=KK(x_s)$ и близких к ним точек к стартовому вектору. Расстояние до стартового вектора в общем случае измеряется в тех же единицах, в которых измерена прогнозируемая величина. В табл. 3 выделены локальные максимумы коэф-

коэффициентов корреляции $KK(\mathbf{x}_s)$ и локальные минимумы расстояний $\rho(\mathbf{x}_{N-k+1}, \mathbf{x}_s)$ до стартового вектора.

Сравнение локальных экстремумов в табл. 3, а также визуальное сопоставление графиков на рис. 2, 3 показывает, что точки локального максимума функции $K=KK(\mathbf{x}_s)$ являются либо точками локального минимума функции $\rho=\rho(\mathbf{x}_{N-k+1}, \mathbf{x}_s)$, определяющей расстояния от векторов $\mathbf{x}_s$ до стартового вектора $\mathbf{x}_{N-k+1}$, либо близкими к указанным точкам минимума. В последнем случае расстояние от точек локального максимума функции $K=KK(\mathbf{x}_s)$ мало отличается от значения в соответствующей точке локального минимума функции $\rho=\rho(\mathbf{x}_{N-k+1}, \mathbf{x}_s)$. Таким образом, можно определить количество векторов-соседей, на которое можно рассчитывать при анализе рассматриваемого временного ряда.

Максимально возможное априори выбираемое число векторов-соседей должно зависеть от числа k координат стартового вектора, поскольку максимальное количество элементов множества всех векторов, близких к стартовому вектору (при определённой фиксированной степени близости), определяет монотонно невозрастающую функцию от параметра k . Заметим, однако, что оценки в виде ограничений снизу на число m векторов-соседей, приведённые в [2] и определяемые неравенством $m \geq 3(k+1)$, не вполне корректны, так как в реальных расчётах рекомендуется придерживаться соотношения $m \approx 3(k+1)$. Из указанных оценок следует, что чем больше число k , тем больше нужно выбирать число m . Эта рекомендация, по мнению авторов, является странной, поскольку с увеличением числа k координат стартового вектора число векторов, близких к стартовому вектору, может только убывать (но не возрастать). Поэтому, увеличивая число k , можно добиться такой ситуации, когда у стартового вектора вообще не будет близких векторов. Следовательно, увеличение числа m векторов-соседей для их поиска является рискованным действием и может создать такие условия, что в семействе найденных векторов-соседей не окажется ни одного вектора, близкого к стартовому вектору, а значит, использование координат таких векторов вместе с координатами стартового вектора может привести к абсурдным результатам.

В работе [9] представлен универсальный формальный алгоритм для определения априори выбранного количества векторов-соседей стартового вектора. Алгоритм подходит как для случая, когда применяется определение понятия вектора-соседа, приведённое в работе [2], так и для случая, когда данное понятие взято из работы [9]. Однако, как показывает практика, важен не только количественный состав семейства векторов-соседей, но и качественная характеристика этого семейства, но её указанный алгоритм учесть не может. Для решения проблемы качества семейства векторов-соседей необходимо избегать ситуации, когда всё семейство этих векторов-соседей состоит из векторов, расположенных в окрестности некоторого одного вектора $\mathbf{x}_{s0}$ – например вектора, в котором достигается максимум функции $K=KK(\mathbf{x}_s)$. Если избавиться от такой ситуации не получается, то влияние вектора $\mathbf{x}_{s0}$ на прогноз будет недопустимо доминировать над влиянием стартового вектора.

С учётом сказанного выше, авторы предлагают прежде чем назначать число m векторов-соседей, найти сначала функцию $K=KK(\mathbf{x}_s)$ и те точки её максимума, в которых функция превышает некоторое заданное число, как правило, 0,9, а затем из найденных точек выбирать векторы-соседи стартового вектора. По количеству точек максимума функции $K=KK(\mathbf{x}_s)$ со значениями $KK(\mathbf{x}_s) > 0,9$, можно приближённо оценить то количество векторов-соседей, на которое можно претендовать при их поиске. Для этого, например, можно количество указанных точек максимума разделить на два. При рассмотренном подходе возникает вопрос, будут ли полученные векторы удовлетворять уравнению (2). Из сказанного выше следует, что множество полученных векторов можно считать приближённым решением данного уравнения. Заметим, что при больших массивах точек максимума функции $K=KK(\mathbf{x}_s)$ выбор из этих массивов семейства векторов-соседей рекомендуется осуществлять с помощью того же алгоритма [9].

Заключение. На основе анализа временных рядов, составленных из значений контролируемого параметра, предложены варианты построения векторов, близких к стартовому вектору. Использование координат построенных векторов-соседей наряду с координатами стартового вектора позволяет увеличивать объёмы выборок для определения применяющихся при прогнозе дрейфа параметра статистических характеристик, а значит, и повышать надёжность оценок этих характеристик.

Т а б л и ц а 3
Расстояния от точек $\mathbf{x}_s$ максимума функции $K=KK(\mathbf{x}_s)$ и близких к ним точек до стартового вектора

Номер вектора	Коэффициент корреляции	Расстояние до стартового вектора
...	...	...
2085	0,95614	1,77028
2086	0,95865	1,70294
2087	0,96239	1,60932
2088	0,96058	1,59487
2089	0,96066	1,56240
2090	0,96350	1,48899
2091	0,96351	1,45794
2092	0,96194	1,45255
2093	0,96081	1,44568
2094	0,95811	1,46489
2095	0,95433	1,50253
...	...	...
2138	0,88376	2,32818
2139	0,88792	2,28438
2140	0,89293	2,22971
2141	0,89662	2,19041
2142	0,89671	2,18778
2143	0,89752	2,17649
2144	0,89646	2,18403
2145	0,89570	2,18749
2146	0,89203	2,22072
2147	0,88796	2,25679
2148	0,88727	2,25590
...	...	...

Л и т е р а т у р а

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974.
2. Лоскутов А. Ю., Михайлов А. С. Основы теории сложных систем. М.–Ижевск: Институт компьютерных исследований, 2007.
3. Абидова Е. А., Хегай Л. С., Чернов А. В., Драка О. Е., Пугачёва О. Ю. Диагностирование электроприводной арматуры с использованием энтропийных показателей // Глобальная ядерная безопасность. 2016. № 4. С. 69–77.
4. Муха Ю. П., Чернов А. В., Абидова Е. А., Хегай Л. С. Алгоритмизация процесса обработки диагностических сигналов электроприводной арматуры с учётом хаотических составляющих // Электронный научный журнал. Инженерный вестник Дона. 2017. № 2. С. 1–10. [Электрон. ресурс]: <http://ivdon.ru/ru/magazine/archive/n2y2017/4253> (дата обращения: 20.06.2018).
5. Сысоев Ю. С., Сальников А. А., Чернов А. В., Бекетов В. Г. Упреждающие методы диагностики технических объектов АЭС // Глобальная ядерная безопасность. 2017. № 3. С. 91–101.
6. Сысоев Ю. С., Симакова Н. А. Оценка длительности межповерочных интервалов измерительных устройств методами теории массового обслуживания // Измерительная техника. 2014. № 12. С. 10–15.
7. Пискунов Н. С. Дифференциальное и интегральное исчисления для втузов. Т. 2: Учебное пособие для втузов. М.: Наука, Главная редакция физико-математической литературы, 1985.
8. Крамер Г. Математические методы статистики. М.: Мир. 1975.
9. Сысоев Ю. С. Использование временных рядов для формирования промежутков однотипного поведения параметров объекта при различных способах прогнозирования // Измерительная техника. 2018. № 2. С. 8–12.

Дата принятия: 21.08.2018.

The time series using for the improving the controlled parameter statistic characteristics estimates reliability for forecasting

Yu. S. Sysoev, O. F. Tsuverkalova

*Volgodonsk Engineering-Technical Institute – Branch of National Research Nuclear University «MEPHI»,
Volgodonsk, Russia, e-mail: sysoev2004@mail.ru*

This paper gives the construction ways of vectors close to the initial vector (neighbour vectors) based on the analyses of time series composed of the controlled parameter values. Using constructed neighbour vectors coordinates together with the initial vector ones allows to increase sample volume to calculate statistic characteristics applied for parameter drift forecast and thus to improve these characteristics estimates reliability.

Key words: *time series, forecasting methods, parameter drift, neighbour vectors.*