УДК 681.518.3

АВТОМАТИЗИРОВАННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА КОРРЕЛЯЦИОННОГО АНАЛИЗА НЕСТАЦИОНАРНЫХ НЕЭКВИДИСТАНТНЫХ ВРЕМЕННЫХ РЯДОВ

С.А. Прохоров, А.А. Посконнова

Самарский государственный аэрокосмический университет им. ак. С.П. Королева
443086, Самара, Московское шоссе, 34

E-mail: sp@smr. ru, free 1. 0@mail. ru

Автоматизированная информационная система корреляционного анализа неэквидистантных временных рядов предоставляет возможность обработки экспериментальных данных либо сгенерированных временных рядов методом с использованием интервальной корреляционной функции. В системе предусмотрена возможность моделирования нестационарных временных рядов по математическому ожиданию и дисперсии, для анализа которых реализован аппроксимативный метод построения аналитической модели тренда и дисперсии нестационарного временного ряда с помощью ортогональных разложений.

Ключевые слова: *неэквидистантные временные ряды, нестационарные случайные процессы, аппроксимация, ортогональные полиномы, выделение тренда, центрирование, нормирование, корреляционная функция, интервальная корреляционная функция*

Введение

При решении многих научно-технических задач приходится сталкиваться с ситуациями, когда исследуемый процесс представлен неэквидистантным временным рядом, например:

- учет «дрожания» при датировании;
- неточность датирования в многоканальных системах;
- дискретизация с пропусками, сбоями наблюдений;
- передача информации по каналу объект-электронно-вычислительная машина;
- принципиальная невозможность получения равномерных отсчетов в ходе проведения эксперимента при теплофизических, океанологических и океанографических исследованиях и так далее [1].

Рассмотренные случаи являются примерами неумышленной дискретизации, которая присутствует независимо от желания исследователя, но иногда нерегулярность вносится в процесс дискретизации преднамеренно, с целью получения новых эффектов, полезных для практических применений. К ним относятся:

- упрощение технической реализации;
- уменьшение вероятностного округления чисел при выполнении повторных арифметических операций;
- нерегулярная дискретизация исследуемых случайных процессов, приводящая к сокращению объема измерительной информации, достаточной для его восстановления (в случае необходимости) с заданной погрешностью [1].

Прохоров Сергей Антонович - д.т.н., профессор.
Посконнова Анна Алексеевна - аспирант.

При преднамеренной дискретизации с известным алгоритмом восстановления информации на интервале дискретизации можно восстановить пропущенные отсчеты неэквидистантного временного ряда, а затем использовать классические алгоритмы оценивания вероятностных характеристик.

В случае непреднамеренной нерегулярной дискретизации случайных процессов оценка вероятностных характеристик, аргументом которых является время (интервал времени), применение классических алгоритмов невозможно. Одним из способов решения задачи оценивания является применение алгоритмов с использованием интервальной корреляционной функции, позволяющих оценить корреляционную функцию без восстановления пропущенных отсчетов неэквидистантного временного ряда [1].

Интервальная корреляционная функция (ИКФ) $C_x(t)$ определяет вероятность появления события в потоке X как функцию времени после данного события без учёта числа прошедших событий:

$$C_x(t)dt = \mathrm{P}\ [\text{событие в } X(t + dt) / \text{событие X в } 0].$$

Нестационарным неэквидистантным временным рядом (ННВР) будем называть временную последовательность случайного процесса, представляющую собой два массива данных: значений случайного процесса (существенных отсчетов) *Ypitp)* и соответствующих им временных отсчетов */y,-*, которую можно записать как

$$Mj\hat{}_N A\hat{}j\hat{}hMj_t)\backslash_{\mathrm{L}} \quad {}_{Jj=l...N}.$$

где *Atji* = $/_{:\,/+1}$ *-tj*, = *random* ; *j* - номер реализации ННВР. Фу/(/у/J - математиче-

ское ожидание **ННВР;** *a(t)* - среднее квадратическое отклонение **ННВР;** *X(t)* -

ское ожидание **ННВР;** *a(t)* - среднее квадратическое отклонение ННВР; X(/) -

НВР с нулевым математическим ожиданием *Ml Xji^jM^* O и единичной дисперси-

ей $D\,\mathrm{X}(/) = 1$.

Постановка задачи и разработка методов ее решения

В соответствии с данной моделью можно выделить несколько классов случайных процессов в зависимости от значений ф(/) и *o(t)* с различными видами нестационарности (см. таблицу) [2].

Приведенная в таблице классификация показывает наличие следующих задач анализа ННВР:

1) определение модели тренда ф(/) ННВР;

2) выделение центрированной составляющей ННВР- $o(t)\overset{\circ}{x}(t)$;

3) построение модели дисперсии а (/) ННВР;

4) нормирование центрированного случайного процесса с использованием полученной модели дисперсии;

5) оценка корреляционной функции (КФ) центрированного неэквидистантного временного ряда (НВР);

6) аппроксимация корреляционной функции НВР ортогональными функциями;

7) оценка спектральной плотности мощности НВР по параметрам ортогональных разложений корреляционной функции НВР.

Классификация нестационарных случайных процессов

<p(')	o(t)	Математическая модель	Класс процесса
0	1	Y(/)=X(/)	Стационарный центрированный случайный процесс
m_x	1	$Y(t) = m_x + X(t)$	Стационарный случайный процесс с ненулевым математическим ожиданием
ф(')	1	Y(/) = <p(f)+X(/)	Нестационарный по математическому ожиданию случайный процесс
0	o(t)	Y(/) = o(/)x(/)	Нестационарный по дисперсии случайный процесс
Ф(')	o(/)	**У(/) =ф(0 + o(/)x(0**	Случайный процесс, нестационарный по математическому ожиданию и дисперсии

Решение 1, 3, 5 и 6 задач возможно с применением аппроксимативных методов оценок вероятностных характеристик в различных ортогональных базисах.

Суть аппроксимативных методов состоит в нахождении аналитического выражения

*=0

л

где /(/) - оценка вероятностной характеристики f\t);

v|/^(r) - ортогональный полином k-того порядка, удовлетворяющий условию на отрезке [a;b] [2]

JV*M^(*M*)^=III Ё ..Л к^и>

где \i(x) - весовая функция.

Коэффициенты Фурье Р^, обеспечивающие минимум среднеквадратического отклонения

д=|/(0-£рт(') \i(t)dt mm ,
A=0

определяются в виде

h=^\f(^k(t)\i(t)dt.

С учетом неэквидистантности временного ряда выражение для оценки коэффициентов разложения представим в виде

$$\beta_k = \frac{(b-a)}{M\|\psi_k\|^2} \sum_{i=1}^{M-1} (t_{i+1} - t_i) \left[\frac{f(t_i)\psi_k(t_i)\mu(t_i) + f(t_{i+1})\psi_k(t_{i+1})\mu(t_{i+1})}{2} \right].$$

Таким образом,

$$\hat{\varphi}(t) = \sum_{k=0}^{m} \beta_k \psi_k(t),$$

$$\beta_k = \frac{(b-a)}{M\|\psi_k\|^2} \sum_{i=1}^{M-1} (t_{i+1} - t_i) \left[\frac{Y(t_i)\psi_k(t_i)\mu(t_i) + Y(t_{i+1})\psi_k(t_{i+1})\mu(t_{i+1})}{2} \right],$$

$$\overset{0}{y}(t_i) = y(t_i) - \hat{\varphi}(t_i),\ i = \overline{1, M},$$

$$\hat{\sigma}^2(t) = \mathrm{M}\left[\overset{o}{Y}{}^2(t) \right] = \sum_{k=0}^{n} C_k \psi_k(t),$$

$$C_k = \frac{(b-a)}{M\|\psi_k\|^2} \sum_{i=1}^{M-1} (t_{i+1} - t_i) \left[\frac{\overset{o}{Y}{}^2(t_i)\psi_k(t_i)\mu(t_i) + \overset{o}{Y}{}^2(t_{i+1})\psi_k(t_{i+1})\mu(t_{i+1})}{2} \right],$$

$$\overset{0}{x}_н(t_i) = \frac{\overset{0}{y}(t_i)}{\hat{\sigma}(t)},\ i = \overline{1, M}.$$

Получив центрированный, нормированный НВР, корреляционную функцию представим в виде [1]:

$$\hat{K}_{xj}(J) = \frac{\sum_{i=1}^{M_j} \overset{o}{x}_{ji_н} \sum_{s=0}^{L} \overset{o}{x}_{j,i+s_н} \delta_{j,i+s}}{\sum_{i=1}^{M_j} \sum_{s=0}^{L} \delta_{j,i+s}},$$

где

$$\delta_{j,i+s} = \begin{cases} 1, & если\ ent\left[\dfrac{t_{j,i+s} - t_{ji}}{\Delta\tau} + 0{,}5 \right] = J. \\ 0, & иначе. \end{cases}$$

Отметим, что выражение $\hat{C}_{xj}(J) = \frac{1}{M_j} \sum_{i=1}^{M_j} \sum_{s=0}^{L} \delta_{j,i+s}$ является текущей оценкой интервальной корреляционной функции и характеризует распределение отсчетов в потоке, находящихся на временном интервале $J\Delta t$.

Структурная схема системы

На рис. 1 представлена структурная схема разработанной автоматизированной системы, где цифровыми эквивалентами указаны соответствующие ей подсистемы:

1) подсистема генерации источника случайного процесса;
2) подсистема фильтрации;
3) подсистема центрирования нестационарного неэквидистантного временного ряда;
4) подсистема нормирования нестационарного неэквидистантного временного ряда;
5) подсистема формирования корреляционной функции и интервальной корреляционной функции;
6) подсистема расчета погрешностей;
7) подсистема ввода-вывода;
8) подсистема настройки оптимальных параметров.

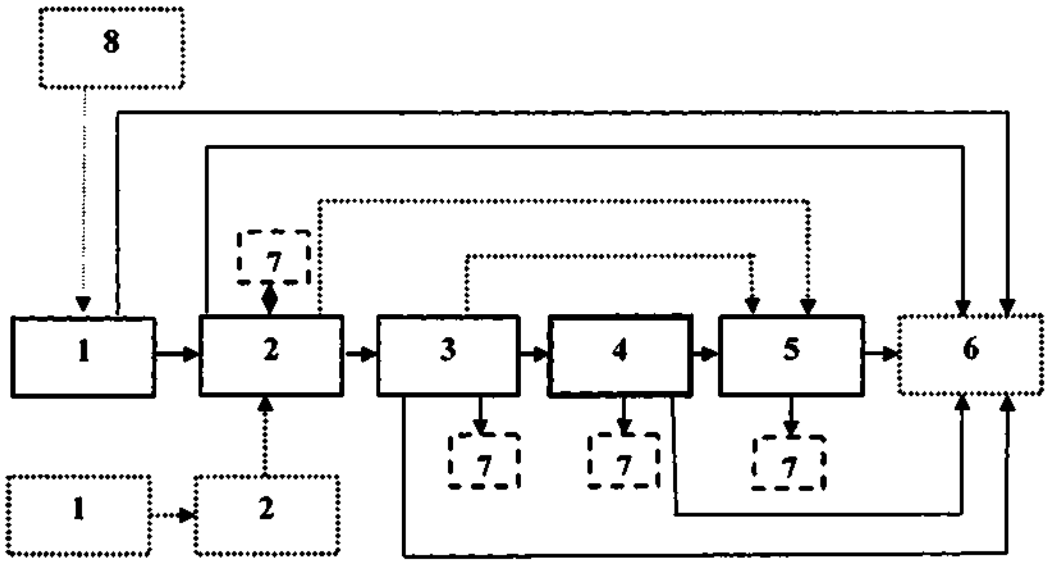


Рис. 1. Структурная схема автоматизированной системы

Примеры имитационного моделирования

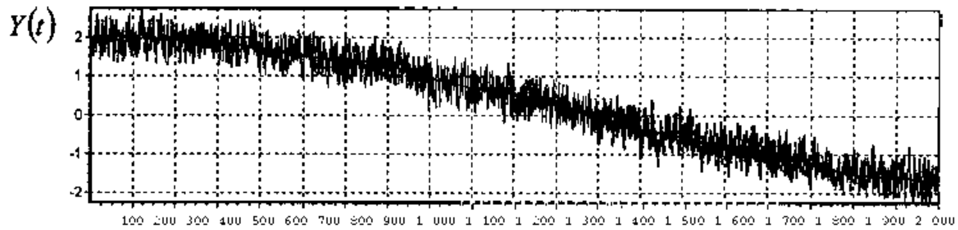
При проведении аппроксимативного анализа вероятностных характеристик неэквидистантных временных рядов в ортогональных базисах перед инженером-исследователем ставятся следующие задачи:

- выбор ортогонального базиса, наиболее подходящего для аппроксимации вероятностной характеристики данного неэквидистантного временного ряда [3];
- определение числа членов разложения, достаточного для описания данной характеристики [3].

Решить поставленные задачи можно с помощью имитационного моделирования нестационарных неэквидистантных временных рядов с заданным математическим ожиданием и дисперсией [3]. Результаты моделирования представлены на рис. 2, 3.

Обработка данных по температуре с «СП-36»

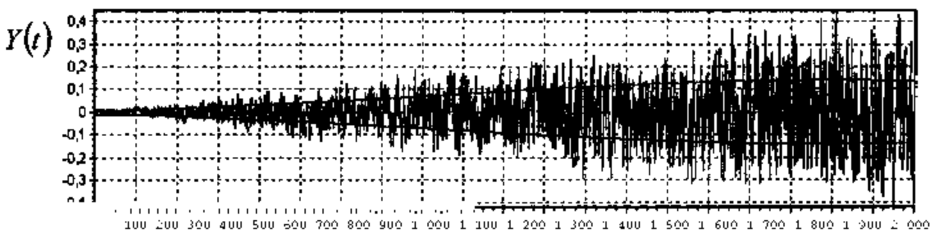
На рис. 4, 5 представлены результаты корреляционного анализа температуры, полученные с официального сайта ГУ «Арктический и антарктический научно-исследовательский институт» (ГУ ААНИИ) для дрейфующей станции «СП-36» [4].



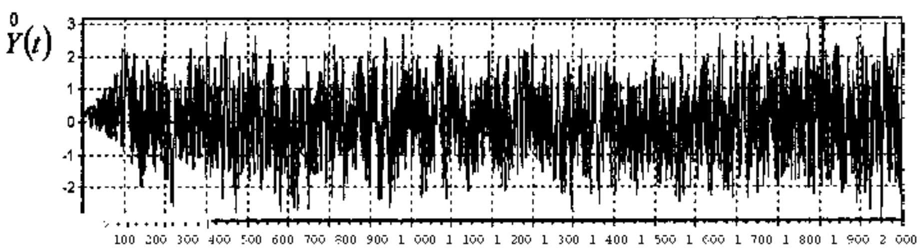
а - исходный процесс

б- центрированный процесс

Р и с. 2. Аппроксимация ННВР с аддитивным трендом вида *y(t)* = 2cos(1 3,2 • КГ /) полиномами Лежандра с числом членов ряда 18



а - центрированный процесс



б- центрированный и нормированный процесс

Р и с. 3. Аппроксимация ННВР с мультипликативным трендом вида д/) = 3 -10" *t* полиномами Лагерра с числом членов ряда 13

Анализируя полученные данные, видим, что ввиду каких-либо непредвиденных обстоятельств небольшая часть данных была утеряна, и некоторые измерения проведены не по расписанию, то есть обрабатываемые данные представляют собой нестационарные неэквидистантные временные ряды. Объем выборки в обоих случаях составляет 480 значений.

Видим, что данные представляют собой нестационарный по математическому ожиданию и дисперсии неэквидистантный временной ряд.

Для получения адекватной КФ загруженные данные передаем в подсистему центрирования и нормирования и подбираем параметры для получения максимально точного результата (см. рис. 4). Наилучшую аппроксимацию исходного (при центрировании) и центрированного (при нормировании) временных рядов получаем при использовнии полиномов Лежандра с числом членов ряда, равным 1.

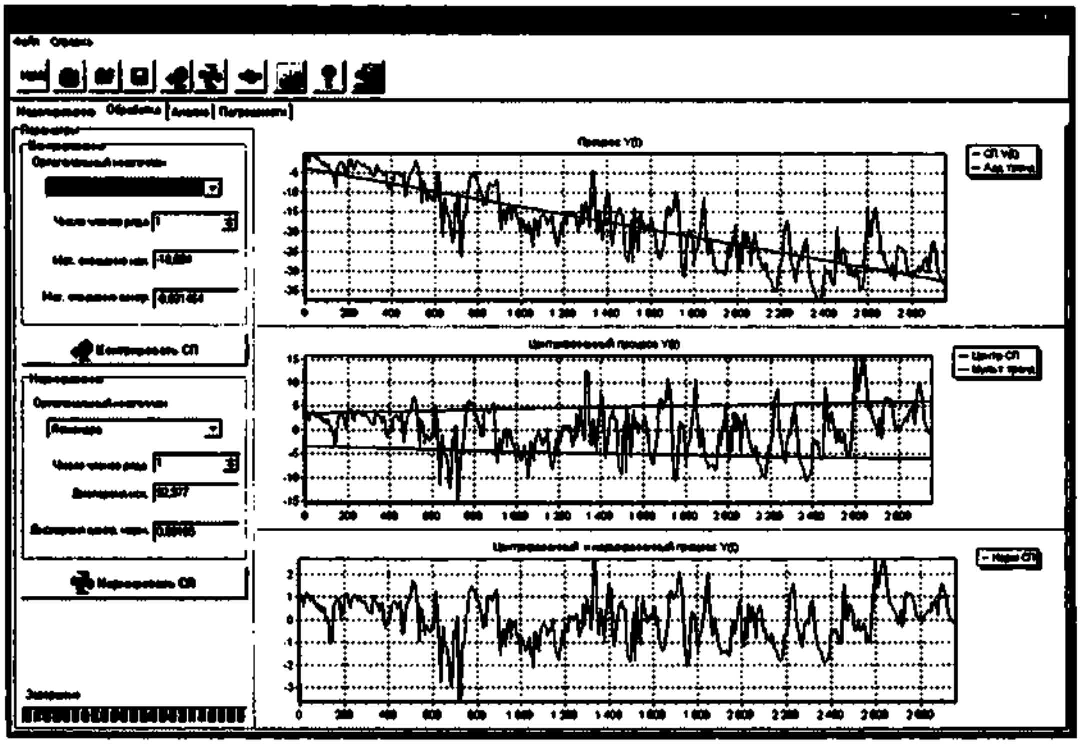


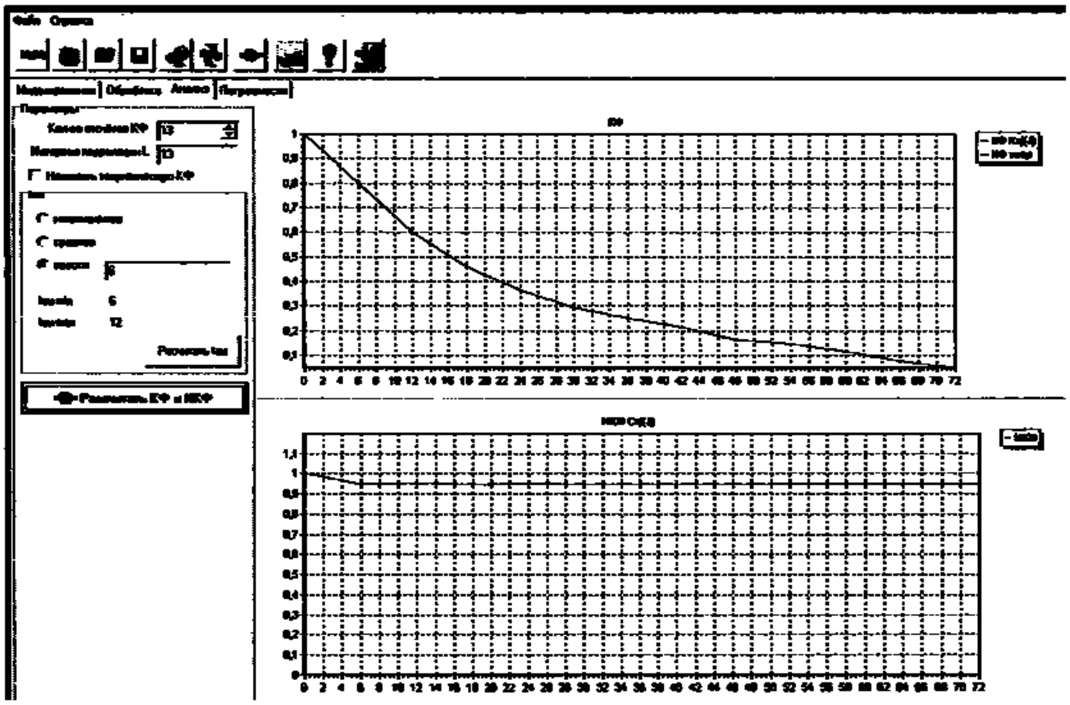
Рис. 4. Вид экранной формы нормирования и центрирования сведений о температуре с «СП-36»

Далее центрированный и нормированный временной ряд передаем в подсистему расчета корреляционной и интервальной корреляционной функции. Определив максимальный и минимальный интервал дискретизации входных данных, подбираем параметры корреляционной и интервальной корреляционной функции. Максимально адекватные результаты получаем, задавая количество отсчетов корреляционной функции и интервал корреляции равными 13 и интервал дискретизации равным 6 (см. рис. 5).

Анализ полученных функциональных зависимостей показывает, что полученная КФ близка к экспоненциальной, а ИКФ незначительно меньше единицы, то есть входные данные снимались с датчиков с небольшими отклонениями по времени и практически без пропусков наблюдений.

Заключение

Проведенные экспериментальные исследования свидетельствуют о применимости базисов ортогональных полиномов к описанию аддитивных и мультипликативных трендов нестационарных неэквидистантных временных рядов. Установлено, что погрешность получаемых в результате аппроксимации моделей трендов в значительной степени зависит от вида тренда, выбранного базиса и числа членов разложения.



Р и с. 5. Вид экранной формы построения корреляционной и интервальной корреляционной функции для данных по температуре с «СП-36»

В результате проведенных экспериментальных исследований установлено, что при использовании метода корреляционного анализа неэквидистантных временных рядов с использованием интервальной корреляционной функции погрешность получаемой оценки корреляционной функции в значительной степени зависит от процента существенных отсчетов ряда.

В дальнейшем планируется расширить систему, добавив в неё подсистему аппроксимации оценки КФ НВР.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Прохоров С.А.* Прикладной анализ неэквидистантных временных рядов. - Самара: Самар. гос. аэрокос. ун-т, 2001 -375 с.
2. Прикладной анализ случайных процессов / Под ред. С.А. Прохорова. - Самара: СНЦ РАН, 2007. - 582 с.
3. *Прохоров С.А.* Математическое описание и моделирование случайных процессов. - Самара: Самар. гос. аэрокос. ун-т, 2001 - 329 с.
4. Дрейфующяя станция «Северный полюс - 36» [Электронный ресурс] - www.aari.ru/resources/d0014/np36/default.asp?lang= I #meteo.

Статья поступила в редакцию 17 июня 2009 г.

UDC 681.518.3

AUTOMATED INFORMATION SYSTEM OF NONEQUIDISTANT CORRELATION ANALYSIS OF TIME SERIES

S.A. Prokhorov, A.A. Poskonnova

S.P. Korolyov Samara State Aerospace University
34, Moskovskoe sh., 443086, Samara

Automated information system of nonequidistant correlation analysis of time series provides an opportunity experimental processing data generated by a time-series method, using the interval correlation function. In a system capable of modeling nonstationary time series of mathematical expectation and dispersion, for which realized approximative method for constructing the analytical model of trend and the dispersion of nonstationary time series with the help of orthogonal decomposition.

Key words: *nonequidistant time series, nonstationary random processes, approximation, orthogonal polinomials, isolation trend, centering, normalization, correlation function, interval correlation function.*