

УДК 004.89

ПРИМЕНЕНИЕ МЕТОДА F -ПРЕОБРАЗОВАНИЯ ДЛЯ ПРОГНОЗА КОМПОНЕНТЫ ВЕКТОРНОГО ТРЕНДА И ЧИСЛОВОГО ПРЕДСТАВЛЕНИЯ ВРЕМЕННОГО РЯДА

© 2010 А.А. Романов

Ульяновский государственный технический университет

Поступила в редакцию 14.05.2010

В работе описывается метод прогнозирования компоненты векторного тренда, описывается алгоритм применения F -преобразования к временному ряду, нахождения остатков, обратного F -преобразования, восстановление временного ряда по спрогнозированному тренду, производится анализ работы метода.

Ключевые слова: F -преобразование, прогноз компоненты векторного тренда, временной ряд.

Нечеткое сглаживание временных рядов на основе нечеткого преобразования (F -преобразования) – методика, разработанная И. Перфильевой [1], которая может быть отнесена к методикам нечеткого приближения. Нечеткое преобразование (F -преобразование) представлено для непрерывных функций и функций на ограниченном наборе точек.

Предположим, что функция f известна в точках $p_1, \dots, p_N \in W$. Мы делим интервал W на множество равноудаленных узлов – $x_k = v_L + h(k-1), k = 1, \dots, n$,

где $N > n, h = \frac{v_R - v_L}{n-1}$ фиксированной длины. F -преобразование имеет две фазы.

Прямое F -преобразование. Определим n базисных функций A_1, \dots, A_n , которые покрывают W и разделим его на n неясных областей. Базисная функция должна удовлетворять следующим условиям ($k = 1, \dots, n$):

1. $A_k : W \rightarrow [0, 1], A_k(x_k) = 1$,
2. $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$, где мы формально полагаем $x_0 = x_1 = v_L, x_{n+1} = x_n = v_R$,
3. $A_k(x)$ непрерывна,
4. $A_k(x)$ монотонно возрастает на $[x_{k-1}, x_k]$ и монотонно убывает на $[x_k, x_{k+1}]$,
5. $\sum A_k(x) = 1$ для всех $x \in W$.

Используя базисные функции, мы преобразуем данную функцию f в кортеж из n действительных чисел $[F_1, \dots, F_n]$, определенных

$$F_k = \frac{\sum f(p_j)A_k(p_j)}{\sum A_k(p_j)}, \quad k = 1, \dots, n. \quad (1)$$

Построение модели ВР для генерации числовых оценок прогноза F, R и \hat{Y} по формуле линейной комбинации:

$$\begin{aligned} F_{k+1} &= \alpha F_k + \beta F_{k-p} \\ R_{k+1} &= \alpha R_k + \beta R_{k-p} \\ \hat{Y} &= F_{k+p} + R_{k+p} \end{aligned}$$

Для конкретной реализации F -преобразования был выбран следующий вид базисных функций:

$$A_k(x) = \frac{x - x_{k-1}}{x_k - x_{k-1}}, \text{ если } x_{k-1} \leq x \leq x_k$$

$$A_k(x) = \frac{x_{k+1} - x}{x_{k+1} - x_k}, \text{ если } x_k \leq x \leq x_{k+1}$$

$$A_k(x) = 0, \text{ в противном случае}$$

Далее строится матрица нечеткого разбиения и производится F -преобразование. F -компоненты представляют собой тренд временного ряда.

Вычисляются остатки – разность между трендом и исходным временным рядом:

$$r_t = f_t - F_t.$$

Прогноз $F_{(n+1)}$ -ой компоненты реализован несколькими методами:

- на основании системы двух линейных уравнений

$$F_{k-1} = \alpha F_{k-2} + \beta F_{k-3} + \gamma F_{k-4} \quad (2)$$

Найдя α, β, γ

$$F_{k+1} = \alpha F_k + \beta F_{k-1} + \gamma F_{k-2}.$$

- использованием нейронной сети

Для обучения сети и прогнозирования строилась регрессия: $F_{k+1} = \alpha F_k + \beta F_{k-1} + \gamma F_{k-2}$.

Для получения числового прогноза временного ряда необходимо сложить обратное F -преобразование с остатками. Остатки для прогноза получаются при помощи системы линейных уравнений, аналогичной (2).

Для оценки качества прогноза вычисляется критерий $MAPE$:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| * 100\%. \quad (3)$$

Также приводится оценка по критерию SMAPE:

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{(Y_t + \hat{Y}_t)/2} \right| * 100\% . \quad (4)$$

При его минимальном значении в автоматическом режиме делается вывод об оптимальных сочетаниях параметров. С данным сервисом был произведено порядка 40 экспериментов на ВР различного характера: стационарных, нестационарных, реальных, модельных. Полученные результаты (табл. 1) показали, что прогноз тренда имеет более выраженное направление, если шаг базисной функции имеет относительно большую величину (случай, когда весь временной ряд разбивается на 5-10 компонент) (рис.1, 2).

В первом случае *MAPE* тренда (внешняя) = 35%, во втором случае *MAPE* тренда (внешняя) = 12,9%. При использовании большого шага базисной функции мы хотя и добиваемся более

сглаженного тренда, но одновременно и удаляем негативные всплески ВР, акцентируя тренд на наиболее общую тенденцию.

Однако при увеличении шага базисной функции ухудшаются показатели качества прогноза остатков. Их оптимальные показатели при шаге базисной функции равной 5-9 единиц.

Производится прогнозирование одной компоненты тренда, начиная с наименьшего количества точек, которые можно подать на вход программы. В эксперименте принимали участие ряды, хранящиеся в БД. Искусственные ряды также присутствовали в эксперименте.

Целью данного эксперимента является выявление ограничений, накладываемых на использование метода *F*-преобразований.

Для начала рассмотрим искусственный ВР, длиной 150 точек (2-12) (рис. 3)

Покажем зависимость точности прогноза от количества точек ВР, участвующих в обучении (рис. 4)

Таблица 1. Сводные данные по оценкам прогноза тренда ВР

ВР (код)	Длина ВР	Шаг базисной функции	Метод	Тренд	
				Маре (внеш)	Smare(внеш)
2-6	40	11	НС	4,95	5
2-8	43	13	НС	9,68	9,84
2-9	42	7	СЛУ	2,3	2,3
2-10	90	11	СЛУ	5,99	5,82
2-11	57	9	НС	65,67	49,44
2-12	150	5	СЛУ	0,85	0,84
2-13	25	11	СЛУ	2,04	2,02
3-15	13	5	НС	94,99	180,94
3-16	13	5	НС	1,69	1,68
3-17	13	5	СЛУ	81,27	57,78
3-24	12	5	НС	19,72	21,88
3-25	12	5	НС	84,09	145,09

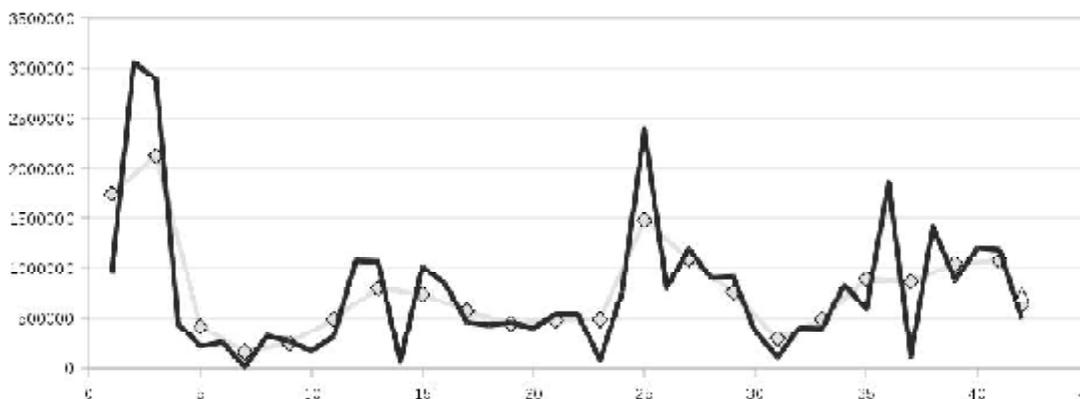


Рис. 1. Тренд построен при малом шаге базисной функции

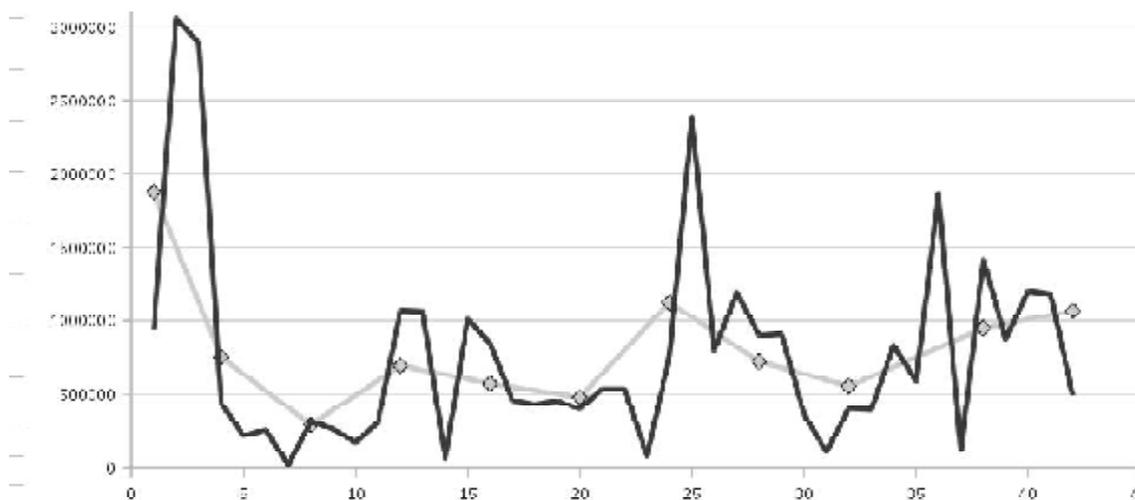


Рис. 2. Тренд построен при большом шаге базисной функции

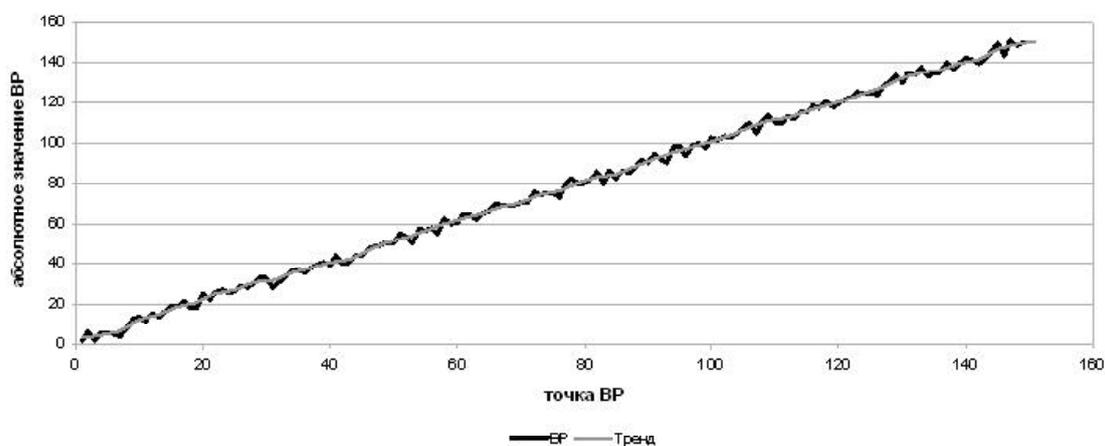


Рис. 3. BP и построенный на нем тренд

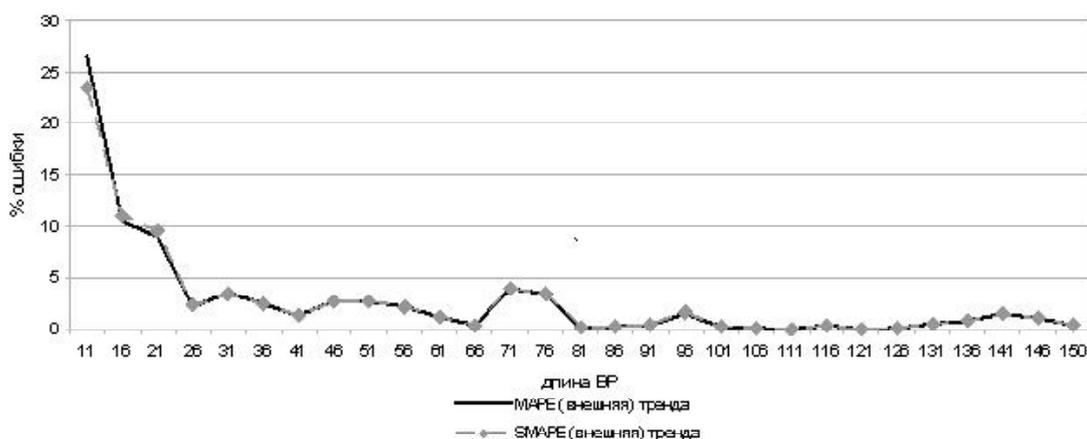


Рис. 4. Зависимость % ошибок от длины тренда

На рисунке показаны критерии оценки: *MAPE* (внешняя) тренда и *SMAPE* (внешняя) тренда. По рисунку можно выделить область, когда ошибка стабилизируется. Это наступает после 26 точек BP, используемых для обучения. Данный случай

- когда в прогнозе проявляется линейность тренда, участвующего в обучении. На графиках ошибок также (за редким исключением) проявляется линейная зависимость: чем больше длина BP для обучения — тем точнее прогноз.

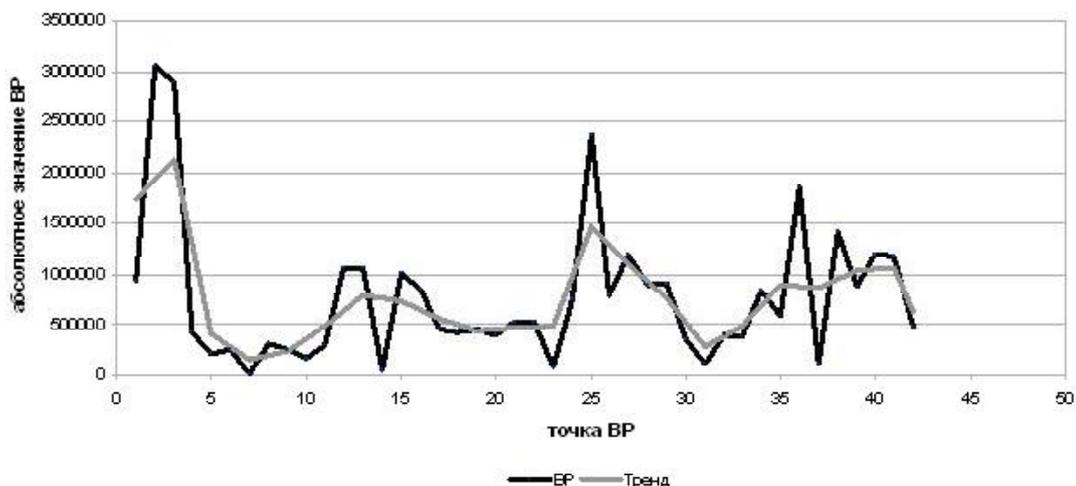


Рис. 5. BP и построенный на нем тренд

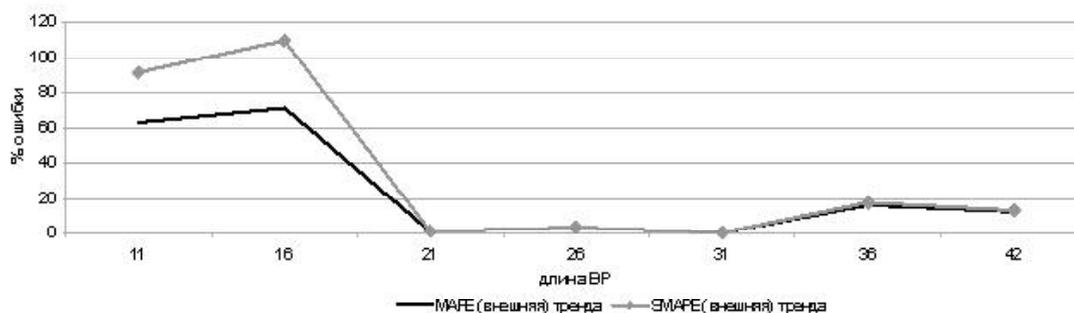


Рис. 6. Зависимость % ошибок от длины тренда

Теперь рассмотрим реальный BP, длиной 43 точки (2-8) (рис. 5).

И также построим зависимость показателей качества прогноза от длины BP (рис. 6)

На данном графике также видна тенденция, когда показатели качества входят в допустимую зону в районе 21-26 точек BP. Как показано в сводной таблице на данном BP достигнутая точ-

ность *MAPE* (внешняя) тренда ? 10% (на максимально возможном количестве точек данного ряда).

Рассмотрим ряд с выраженным трендом, но имеющим нестационарности (реальный BP) длиной 42 точки (2-9) (рис. 7, 8)

Видно, что ошибка практически не изменятся от длины ряда (за счет выраженного тренда)

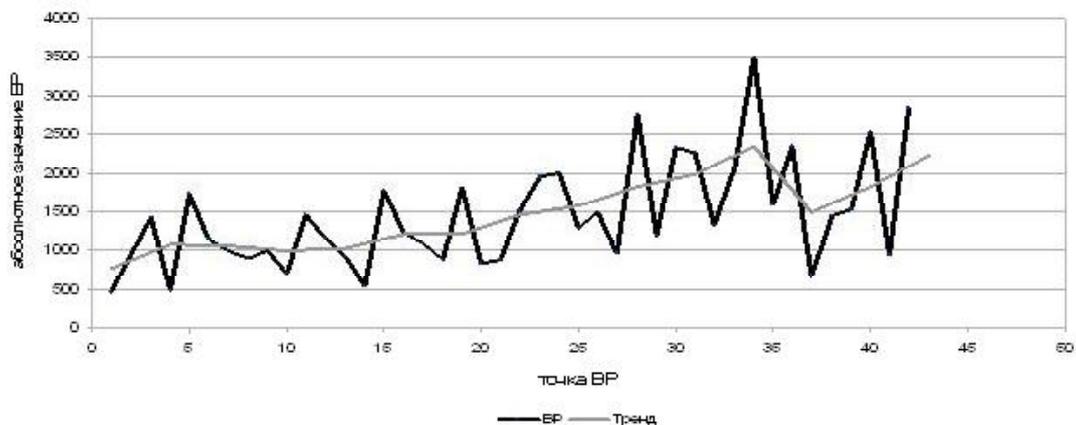


Рис. 7. BP и построенный на нем тренд

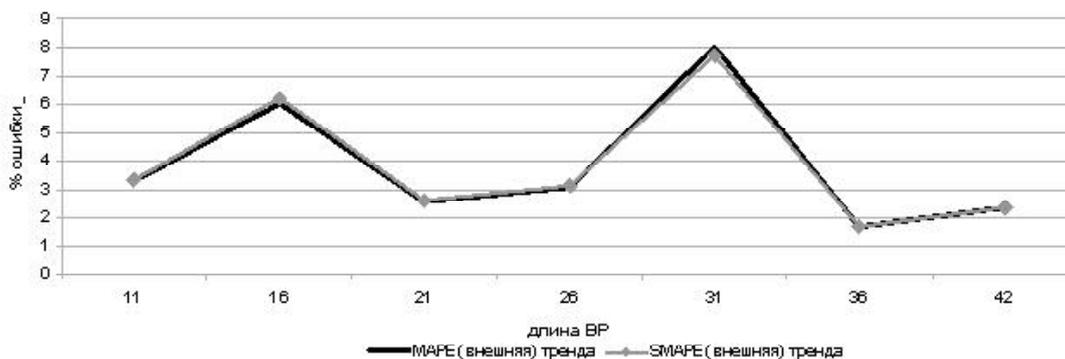


Рис. 8. Зависимость % ошибок от длины тренда

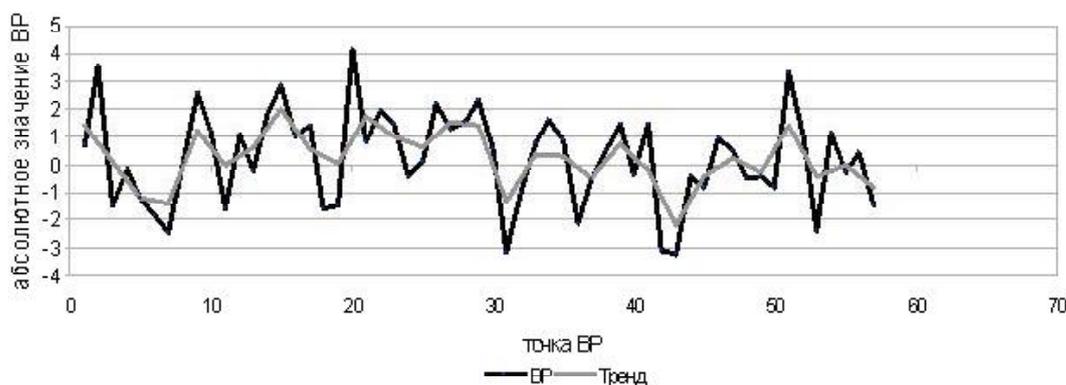


Рис. 9. ВР и построенный на нем тренд

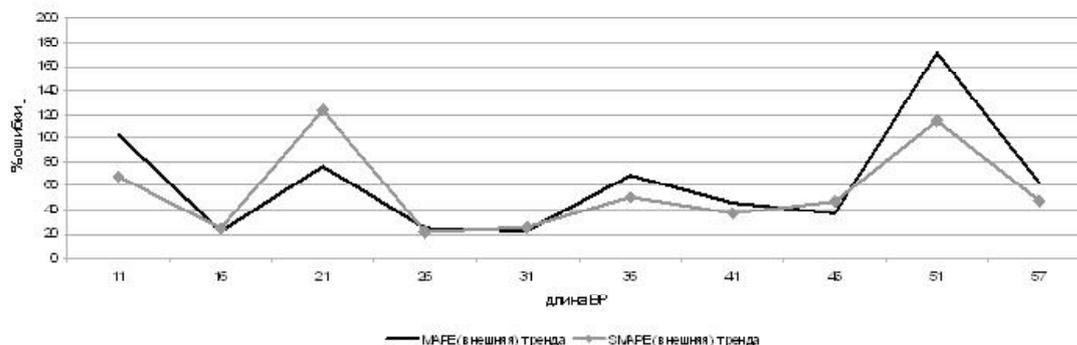


Рис. 10. Зависимость % ошибок от длины тренда

и имеет небольшую величину (до 10%) - линейность тренда. Только при переходе через излом тренда ошибка немного повышается — за счет того, что разность прогноза при обучении и реальной величины тренда имеет большую величину, чем на всем остальном ВР.

Ряд, без выраженного тренда длиной 57 точек (2-11) (рис. 9, 10):

Ошибка на таком ВР очень высока, и не зависит от длины ряда, участвующего в обучении. В такой ситуации снизить ошибку можно за счет сглаживания тренда — увеличения шага базисной функции. Но, полагаю, что такой шаг не име-

ет практической ценности, поскольку происходит вырождение тренда в прямую и теряется суть прогноза тренда — выявления тенденции.

Проведенные эксперименты показали, что данный метод достаточно хорошо работает (% ошибок укладывается допустимый интервал < 20%) на рядах с выраженным трендом и лишенных случайных перепадов значений. Заданная точность достигается уже при длине ВР более 25-30 точек. Вследствие этого необходимо дополнить использование данного метода способом определения характера ВР для предварительного анализа и принятия решения о целесообразности прогноза.

Исследование выполнено при поддержке ФЦП “Научные и научно-педагогические кадры инновационной России”, государственный контракт 02.740.11.5021. Автор выражает благодарность профессору г. Острава (Чехия) Перфильевой И.Г. за помощь в проведении исследования.

СПИСОК ЛИТЕРАТУРЫ

1. *Перфильева И.* Нечеткое преобразование: применительно к проблеме роста рифов. В кн: Демикко Р. Клир GJ и др. редакторы. Нечеткая логика в геологии. Амстердам: Академическая пресса: 2003, 275-300.
2. *Ярушкина Н.Г.* Основы теории нечетких и гибридных систем. Финансы и статистика, 2004. 320 с.

USING F-TRANSFORM METHOD FOR FORECAST COMPONENT OF THE VECTOR TREND AND NUMERIC REPRESENTATION OF THE TIME SERIES

© 2010 AA Romanov

Ulyanovsk State Technical University

The paper describes a method of predicting the trend components of the vector, is described by the algorithm of F-transform to the time series, find the remnants of the inverse F-transform reconstruction of the time series of those predicted trend, the analysis is made of the method.

Key words: F-transform, forecast component, the vector trend, time series