Short-term Traffic Flow Prediction Based on Embedding Phase-space and Blind Signal Separation

Hong Xie,Zhonghua Liu
College of Information Engineering
Shanghai Maritime University
Shanghai, China
hongxie@cie.shmtu.edu.cn

***Abstract*—Accurate traffic flow forecasting is one of the important issues for the research of Intelligent Transportation System (ITS).The capability to forecast traffic flow has been identified as a critical need for dynamic traffic control system. The embedding phase-space theory treats the dynamic evolution of traffic flow as a chaos time series, and this provides the possibility to forecast short-term traffic flow accurately. The theory of blind signal processing is widely used in the area of data mining. Practical historic traffic flow can be regarded as a blind signal mixed of real traffic flow and noise introduced by measurement tools. Blind signal separation is a good method to reduce noise and abstract principal components of historic traffic flow series. This paper proposes an approach based on embedding phase-space and blind signal separation, which enables us to realize the denosing and forecasting of the traffic flow synchronously, with another advantage of self-adaptive characteristic.**

***Index Terms*—Embedding phase-space;blind signal separating;prediction**

I. INTRODUCTION

Observation and research of traffic flow's variety regulation, as well as scientific and accurate traffic flow recognition in advance, namely traffic flow forecasting, have great importance to traffic navigation and management in ITS. Short-term traffic flow system forecasts the traffic flow condition in a few minutes in order to adopt proper traffic control and guidance measurements, using the forecasting algorithms to analyze traffic flow data. A variety of algorithms to forecast the short-term traffic flow have been presented based on statistical techniques, neural network theory, support vector machine theory, wavelet neural network theory and so on [1] [2] [3]. These algorithms play an important role in traffic flow forecasting, and each has its shortcoming. For example, the wavelet neural network theory develops a good idea to traffic flow prediction, but the edge effect of this algorithm could affect the accuracy of traffic flow forecasting.

Recent years, complicated dynamics character of nonlinear systems has been studied in deeply and used widely, the most interesting phenomenon among of them is chaos, it revealed that some variable seem to be random phenomenon but produced by nonlinear certain system essential in natural world, provide possibility for high precision in short-term traffic flow forecasting in theory. The practical traffic flows which exist in the form of time series data are obtained from the huge and complicated traffic system, so most of them have the characteristic of chaos and can be analyzed by chaos theory.

During the process of collecting the traffic flow, it can not avoid introducing measurement noise which can affect the prediction accuracy of the forecasting model to a certain extent. So before establishing the forecasting model in terms of the obtained historic traffic flow series, the noise should be removed from it. As so far, the most popular methods of smoothing noise consist of filters, wavelet analysis and so on, and all of them have their own advantages and disadvantages. Because the practically obtained traffic flow data is occurred by the huge and complicated unknown system, the traffic system could be seem as a blind source, and traffic flow series could be seen as a blind signal, a mixture of effective flow series and noise, generated by this blind source. Therefore, the theory of blind signal separation is suitable to analyze traffic flow series.

To overcome the problems mentioned above, a short-term traffic flow forecasting method based on phase-space construction and blind signal separation. Our main contribution of this paper is the de-noising and forecasting of the nonlinear time series. For the phase-space theory is the effective method to analyze nonlinear time series and the principal components analysis (PCA) method is widely used in blind signal processing and feature extracting as a data reduction technique, this paper put forwards a new forecasting model based on embedding phase-space and blind signal separation. First, reconstruct the phase-space of the historic time series. Second, extract the principal components of the series matrix after embedding phase-space reconstruction by an orthogonal transform, this process de-noise the data matrix synchronously. Third, establish suitable forecasting models to predict every principal component separately. At last, the next h ($h \epsilon Z$) steps' traffic flow forecasting data are obtained by an inverse transform.

II. EMBEDDING PHASE-SPACE PROBLEMS

A. Reconstruction of Embedding Phase-spce

Given the time series $\{x(n)\}_{n\epsilon Z}$, with zero mean (The practical time series such as traffic flow data are always non-zero-mean-value and should be removed mean value first. This paper assumes that the time series used in this paper is non-zero-mean-value), suppose the delay step and embedding dimension of the embedding phase-space be τ and d separately, and then the vector of embedding phase-space is defined as

$$X_i = \left(x\left(i\right), x\left(i-\tau\right), \cdots, x\left(i-\left(d-1\right)\tau\right)\right)^T \qquad (1)$$

where $i\epsilon((d-1)\tau+1,n)$.Time delay method is used to construct $(n-(d-1)\tau)\times d$ dimensional orbit matrix with series $\{x(1),x(2),\cdots,x(n)\}$. Let $m=n-(d-1)\tau$, matrix X is defined as

$$X=\begin{bmatrix} X^T_{(d-1)\tau+1} \\ X^T_{n-1} \\ \vdots \\ X^T_n \end{bmatrix}$$

$$=\begin{bmatrix} x((d-1)\tau+1) & x((d-2)\tau+1) & \dots & x(1) \\ \vdots & \vdots & \ddots & \dots \\ x(n-1) & x(n-1-\tau) & \dots & x(n-1-(d-1)\tau)) \\ x(n) & x(n-\tau) & \dots & x(n-(d-1)\tau)) \end{bmatrix} \tag{2}$$

Each row of matrix X represents a vector of embedding phase-space, namely a phase point in embedding phase-space. $m=n-(d-1)\tau$ is the number of points in the constructed phase-space attractor [4].

B. The Algorithm of Selecting The Optimal Delay Step And Embedding Dimension of Embedding Phase-space

The most important thing of embedding phase-space theory is how to select proper delay step and embedding dimension of the reconstructed embedding phase-space. Chaos theory always faces the problem of 'dimension disaster'. This paper adopts an algorithm of selecting the optimal delay step and embedding dimension synchronously to avoid this problem. Let delay step be τ, embedding phase-space dimension to be d, and then the vector of the embedding phase-space is defined as formula (1). Let dynamic system equation be:

$$X_{n+1}=F\left(X_n\right) \tag{3}$$

where $F:R^d\rightarrow R^d$ is a smooth mapping. For every reference point:

$$X_m=\left(x\left(m\right),x\left(m-\tau\right),\cdots,x\left(m-\left(d-1\right)\tau\right)\right)^T \tag{4}$$

take a small number r , define its r-neighborhood $U\left(X_m,r\right)$ as a set of points satisfying the following condition:

$$U\left(X_m,r\right)=\{X_n:\|X_n-X_m\|\leq r,n\neq m\} \tag{5}$$

where $\|\cdot\|$ is Euclid norm, that is:

$$\|X_n-X_m\|=\left(\sum_{i=0}^{d-1}\left(x\left(n-i\tau\right)-x\left(m-i\tau\right)\right)^2\right)^{\frac{1}{2}} \tag{6}$$

As F is a smooth mapping, if r is properly small, for an arbitrary point $X_s\in U\left(X_m,r\right)$, we have the following expression:

$$\begin{aligned} X_{s+1}-X_{m+1}&=F\left(X_s\right)-F\left(X_m\right) \\ &\approx J\left(X_m\right)\left(X_s-X_m\right) \end{aligned} \tag{7}$$

where $J\left(X_m\right)$ is the Jacobin matrix of F at point X_m. Define the first component of the above formula as:

$$x\left(s+1\right)-x\left(m+1\right)=\sum_{k=1}^{d}a_k\left(x\left(s-\left(k-1\right)\tau\right)-x\left(m\right)\right) \tag{8}$$

This is a linear equation. If there are enough points in neighborhood $U\left(X_m,r\right)$, we can estimate the coefficients $\{a_k\}(k=1,2,\cdots,d)$ of (8), using the least square method.

To estimate the coefficients of (8), the number of points included in $U\left(X_m,r\right)$ should not be less than d , at the same time we must ensure that the vector set $\{x\left(s\right)-x\left(m\right)\}$ is linear independent. In order to stably estimate the coefficients, the points included in $U\left(X_m,r\right)$ should be as more as possible, this usually needs to select a bigger neighborhood radius r and may lose the approximate linear character of the dynamic system described by formula (7). In order to balance this contradiction and make this algorithm to be effective all the time, we select r which satisfies that the number of points contained $U\left(X_m,r\right)$ is $2d$.

Estimate the coefficients of (8) by using the points in $U\left(X_m,r\right)$, and let $\delta\left(X_m\right)$ represent the mean square error of the estimation. Further, suppose all reference points selected are $\{X_{m_1},X_{m_2},\cdots,X_{m_D}\}$ $(D=2d)$, then we can define global approximating mean square error as:

$$\Gamma\left(\tau,d\right)=\frac{1}{D}\sum_{k=1}^{D}\delta\left(X_{m_k}\right) \tag{9}$$

The actual meaning of $\Gamma\left(\tau,d\right)$ is the average value of the mean square error of the local linear approximation. For reference points, if there are not too many phase points in the reconstructed phase-space, all phase points can be selected as reference points to calculate $\Gamma\left(\tau,d\right)$, otherwise some sample points can be randomly selected as reference points for the sake of saving computing time.

Then give a range of τ and a upper bound d_{max} of d, let $1\leq\tau\leq T_N$, $2\leq d\leq d_{max}$. For every group of τ and d, calculate $\Gamma\left(\tau,d\right)$, and then select the τ and d value which make $\Gamma\left(\tau,d\right)$ minimum as the optimal delay step and embedding dimension of the embedding phase-space.

For determining the range of delay step τ , the method presented in this paper is computing the autocorrelation coefficients of $\{x\left(n\right)\}$ first, and then let the corresponding delay time be T_N when the autocorrelation coefficients firstly arrive the negative local maximum, thus the searching range of τ is $1\leq\tau\leq T_N$. The maximal d value d_{max} is appointed to a proper number by experience.

III. PRINCIPAL COMPONENTS ANALYSIS

For matrix X defined in formula (2), every column of the orbit (data) matrix X is the observations of each delay coordinate at different moment, and the number of the coordinate delay variables is d. Coordinate system composed of these d variables is always non-orthogonal. PCA [5] is therefore the method to do an appropriate transform of matrix X, making the transform result satisfy the orthogonal characteristic in new coordinate system.

Principal Components Analysis (PCA) has been widely studied and used in pattern recognition and signal processing. In fact it is important in many engineering and scientific disciplines, e.g., in data compression and feature extraction, noise filtering, signal restoration and classification. Often the principal components (PCs) (i.e., directions on which the input data have the largest variances) are regarded as important, while those components with the small variances are minor components (MCs) are regarded as unimportant or associated with noise. PCs are also regarded as the components with lager energy of the input data, corresponding to the main physical quantities. Extracting and separating PCs equal to blindly extracting the main components from the blind signal (blindly eliminating other components). The method of extracting principal components from mixed signals called PCA is a basis method of blind signal processing.

Traditional PCA method analyzes the blind signal as a vector; this paper puts forward a new PCA method based on reconstructed phase-space of the blind signal.

For matrix X defined in formula (2), generally $d < m$, X can be decomposed as

$$X = VSU^T \tag{10}$$

where S is $d \times d$ dimensional diagonal matrix, its element is

$$S_{ij} = \delta_{ij}\sqrt{\lambda}_i, i, j = 1, 2, \dots, d \tag{11}$$

V is $m \times d$ dimensional matrix, satisfying column vector orthogonal

$$(V^T V)_{ij} = \delta_{ij} \tag{12}$$

U is $d \times d$ dimensional diagonal matrix, satisfying that

$$(U^T U)_{ij} = (UU^T)\delta_{ij} \tag{13}$$

The covariance matrix A of X is

$$A = (XX^T) \tag{14}$$

The covariance matrix A is $d \times d$ dimensional symmetric matrix, its element is

$$A_{ij} = \sum_{k=1}^{m}(x_{i+k-1}x_{j+k-1}), i, j = 1, 2, \dots, d \tag{15}$$

The value of (15) reflects the correlation degree of coordinate vector i and j. Using formula (10) instead of X in formula (14), (14) becomes

$$A = X^T X = USV^T VSU^T = US^2U^T \tag{16}$$

Thereby

$$U^T U = (XU)^T XU = S^2 = \Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_d\} \tag{17}$$

It's obvious that $U^T AU$ is covariance matrix of matrix Y, of which the elements equal to zero except those of diagonal. This means that the variable i and j of the new matrix Y, transformed from matrix by coordinate transformation $Y = XU$, are independent.Thus the coordinate system formed by variables of matrix Y has orthogonal characteristic.

In fact, $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_d\}$ (The d eigenvalues are organized with respect to decreasing values of themselves, i.e.$\lambda_1 > \lambda_2 > \cdots > \lambda_i > \cdots > \lambda_d$) is the d eigenvalues of covariance matrix A and $U = \begin{bmatrix} u_1 & u_2 & \dots & u_d \end{bmatrix}$ is the corresponding orthogonal eigenvector referred to as principal eigenvectors. Assume that the d eigenvalues Λ of A mutually different, the corresponding eigenvectors matrix U is exclusive.

Because the vector u has d values, the data matrix X has projections correspondingly, especially y_i is the projection of X to u_i

$$y_i = Xu_i, i = 1, 2, \dots, d \tag{18}$$

y_i is called principal component (PC).

The coordinate transformation to extract PCs is

$$Y_p = XU_p \tag{19}$$

where Y_p is the output matrix called the matrix of principal components (PCs), and $U_P = \begin{bmatrix} u_1 & u_2 & \dots & u_i \end{bmatrix}$ is the set of signal subspace eigenvectors. On the other hand, the $(d-i)$ minor components are given by

$$Y_M = XU_M \tag{20}$$

where $U_M = \begin{bmatrix} u_{i+1} & u_{i+2} & \dots & u_d \end{bmatrix}$ consists of the $(d-i)$ eigenvalues associated with the smallest eigenvalues.

For formula (19) and (20), an important thing to do is to determine the value of i, namely the number of the PCs. In terms of computing the "accumulative variance contribution rate"

$$Yp = \sum_{j=1}^{i}\lambda_j / \sum_{j=1}^{d}\lambda_j \tag{21}$$

of the first i components, we can judge the proper i value. If the first i components' 'accumulative variance contribution rate' is big enough, and has little change while i takes a bigger value, the other $(d-i)$ components can be regarded as noises, and i is the number of PCs.

IV. FORECASTING MODEL BASED ON EMBEDDING PHASE-SPACE AND BLIND SIGNAL SEPARATION

Having determined the optimal delay step τ and embedding dimension d of a time series' embedding phase-space, and having abstracted this phase-space's principal components matrix Y_p , the next thing to do is constructing the forecasting model of the time series $\{x(n)\}$.

For $U_P = \begin{bmatrix} u_1 & u_1 & \dots & u_i \end{bmatrix}$, formula (19) could be written

$$Y_P = X\begin{bmatrix} u_1 & u_1 & \dots & u_i \end{bmatrix} = \begin{bmatrix} Xu_1 & Xu_1 & \dots & Xu_i \end{bmatrix}$$

$$= \begin{bmatrix} y_1 & y_1 & \dots & y_i \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1i} \\ y_{21} & y_{22} & \cdots & y_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mi} \end{bmatrix} \quad (22)$$

For each column vector of Y_p , construct an appropriate forecasting model such as AR model and then proceed h steps prediction separately. The prediction matrix $\hat{Y}_p$ of Y_p is written

$$\hat{Y}_P = \begin{bmatrix} Y_P \\ Y_{Pf} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1i} \\ y_{21} & y_{22} & \cdots & y_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mi} \\ y_{(m+1),1} & y_{(m+1),2} & \cdots & y_{(m+1),i} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(m+2),1} & y_{(m+2),2} & \cdots & y_{(m+2),i} \end{bmatrix} \quad (23)$$

where $\hat{Y}_p$ is $h \times i$ dimensional matrix. Then the h ($h \epsilon Z$) steps' forecasting data of time series $\{x(n)\}$ are obtained by the inverse transform of $\hat{Y}_p$

$$\hat{X} = \hat{Y}_p U_P^T = \begin{bmatrix} Y_P U_P^T \\ Y_{Pf} U_P^T \end{bmatrix}$$

$$= \begin{bmatrix} \hat{x}((d-1)\tau+1) & \hat{x}((d-2)\tau+1) & \cdots & \hat{x}(1) \\ \hat{x}((d-1)\tau+2) & \hat{x}((d-2)\tau+2) & \cdots & \hat{x}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}(n) & \hat{x}((n-1) & \cdots & \hat{x}(n-d+1) \\ \hat{x}((n+1)) & \hat{x}(n) & \cdots & \hat{x}(n-d+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}((n+h)) & \hat{x}(n+h-1) & \cdots & \hat{x}(n-d+h) \end{bmatrix} \quad (24)$$

From formula (24), it is obvious that the last h values $(\hat{x}(n)), (\hat{x}(n+1)), \dots, (\hat{x}(n+h))$ of the first column of matrix $\hat{X}$ are just the $1 \sim h$ step forecasting value of $\{x(n)\}$. For accurate prediction, h should not be too large.

As mentioned above, we can see that this forecasting model has the advantage of simple computation and can realize h steps. As a conclusion, the algorithm approved by this paper can be summarized as Fig. 1.

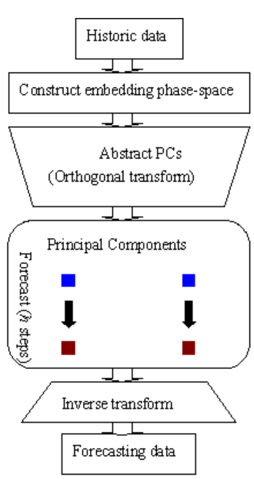

Fig. 1. Flow of chart of forecasting algorithm approved by this paper

TABLE I
PREDICTION ERRORS OF TWO METHODS(BASED ON DATA OF WEEKDAYS ONLY)

forecasting method	SSE	$MARE$
Forecasting algorithm approved by this paper	5.18185	2.2138
Wavelet neural network algorithm	7.9091	3.2778

V. EXPERIMENT RESULTS

For actual traffic flow data prediction, we use the analysis method and forecasting algorithm mentioned above. In order to obtain an accurate prediction, the traffic flow data for weekdays and weekends should be discussed separately because there are great differences between the characteristics of them. That is to say, the model based on the historical traffic flow data for weekdays (weekends) is used to forecast the next weekday's (weekend's) traffic flow only.

This paper discusses the real traffic flow data of a section of a highway that collected by the CATT Lab [10]. The historic data was collected from 0:00 of Dec.1 to 15:00 of Dec.28, 2006 (the sampling interval is five minutes). Use these data to establish the forecasting model approved by this paper and predict the traffic flow from 15:05 to 15:55 (the prediction interval is five minutes, too). During the process of constructing the forecasting model, we determine that the optimal delay step and embedding dimension of embedding phase-space are 12 and 5 separately, and the number of PCs is 3, by the algorithm approved in this paper. The forecasting results are plotted in Fig. 2. In order to verify the accuracy and validity of the method approved in this paper, we use a wavelet neural network model as comparison, the curve of 11 steps ahead forecasting data using wavelet neural network approach is also plotted in Fig.2, too.

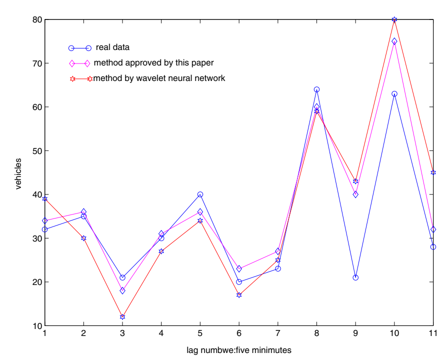

Fig. 2. Comparison 1 between real data and prediction data with two algorithms

This paper adopts two error measures: summary squares of prediction error (SSE), mean absolute relative error (MARE), the calculating result is shown in Table 1.

If the forecasting model is constructed with the historic flow data including the data of weekends, the forecasting traffic flow

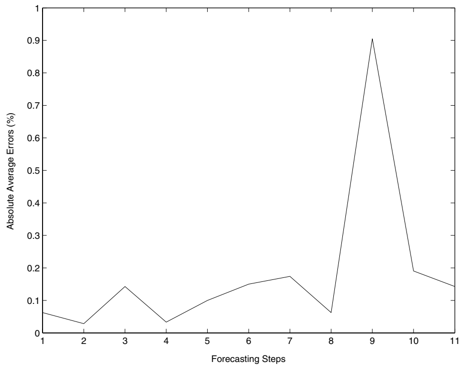

Fig. 3. Absolute average errors with method this paper approves

TABLE II
PREDICTION ERRORS OF TWO METHODS(BASED ON DATA OF ALL DAYS)

forecasting method	SSE	$MARE$
Forecasting algorithm approved by this paper	5.4545	2.2416
Wavelet neural network algorithm	8.7273	3.4985

from 15:05 to 15:55 of Dec.28, 2006 are plotted in Fig.4 and the prediction errors are shown in Table II.

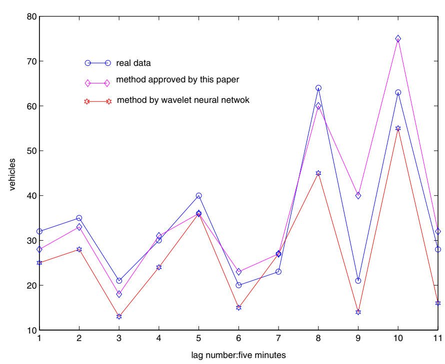

Fig. 4. Comparison 2 between real data and prediction data with two algorithms

From the tables and pictures listed above, we can verify that we should use forecast traffic flow of weekdays and weekends separately. Also we can see that both the error parameters of the algorithm approved in this paper are smaller than the method using wavelet neural network. So the algorithm approved in this paper is possibly superior to wavelet neural network method for short-term traffic flow forecasting, and has a certain application prospect. Fig.3 shows the relationship between the forecasting steps and the absolute average errors. It can be observed the forecasting steps of the model presented by this paper should not be too large, and this model can provide accurate short-term traffic flow forecasting.

VI. Conclusion

In recent years, people use a large number of forecasting methods to forecast traffic flow, and have obtained great achievements. This paper improves the forecasting model based on the embedding phase-space, firstly introduces the blind signal separation theory into the constructing of forecasting model, and then proposes a new algorithm which can de-noise and predict the historic traffic flow series. On the other hand, this forecasting model based on embedding phase-space and blind signal separation has the characteristic of self-adaptive by modifying the orthogonal transform matrix U_P real-time, and the analysis process of the historic data is simple. But the main weakness of this forecasting model is that we must use large number of historic data to construct embedding phase-space. Another weakness is that we consider no other factors that influence the traffic flow in this model. The status of roads and weather greatly influence the traffic flow, so how to include these factors in the traffic flow forecasting model and construct an accurate forecasting model with less data are problems need to be studied further.

References

[1] Chris Tampere , Bart van Arem,*Traffic flow theory and its applications in automated vehicle control: A review*, Intelligent Transportation Systems Conference Proceeding, Oakland (CA), USA, pp.25-29, August 2001.

[2] Weng XX , Han J , *Analysis of spectrum and prediction for short-term traffic flow*, Proceedings of the 26th Chinese control conference, Vol. 6 , pp.58-61, 2007.

[3] Su HW, Zhang L, Yu S, *Short-term traffic flow prediction based on incremental support vector regression*, Third international conference on Natural Computation, Vol. 1, pp.640-645, 2007.

[4] ShiXizhi, *Signal processing and soften-computing*, Higher Education Press, pp.318-320, 2003

[5] Cichocki, Amari, *Adaptive Blind Signal and Image Processing* , pp87-97.

[6] Daiheng Ni, *Determining Traffic-Flow Characteristics by Definition for Application in ITS*, IEEE transactions on international transportation systems, Vol.8, Issue 2, pp.181-187, June 2007.

[7] Han min, *Forecasting theory and methods of chaotic time series,* China Waterpower Press, pp.79-92 , 2007.

[8] Taehyung Kim, Hyoungsoo Kim and David J. Lovell, *Traffic Flow Forecasting: Overcoming Memoryless Property in Nearest Neighbor Non-Parametric Regression*, Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, pp.965-969, Vienna, Austria, September, 2005.

[9] S.Singh, *Pattern modeling in time-series forecasting*, Cybernetics and Systems, Vol. 31, Issue 1, pp. 49-66, 2000.

[10] http://www.cattlab.umd.edu