Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Tikhonov-type Regularization in Local Model for Noisy Chaotic Time Series Prediction

Zhiwei Shi and Min Han, *Senior Member, IEEE*

Abstract—**Tikhonov-type regularization method for noisy chaotic time series prediction is investigated. The current regularized local prediction method is interpreted as one kind of filter factors to decrease the variance of the predictor. One drawback in the interpretation is the ignorance of the random noise in coefficient matrix, another drawback is the relationship between the regularization parameter and the noise condition is not clearly explained, so the determination of regularization parameter has to resort to some techniques such as cross validation. In this study, local linear model is studied from the perceptive of the Errors-In-Variables (EIV) modeling, and the predictor is designed by considering the noise both in coefficient matrix and right-hand side. The optimal solution can be obtained by second order convex program (SOCP) if given a perturbation bound of the noise, and the solution can be reformulated as a form of Tikhonov regularization, and it will be shown how regularization parameter is related to the Frobenius norm of the noise containing in coefficient matrix and right-hand side. Two demonstrations are presented to show the validity of the results.**

I. Introduction

Time series analysis and prediction is an important problem encountered in many fields of practical applications, such as meteorology, hydrology, engineering and biology. Chaotic signal includes a broad class of time series, which is sensitive dependence on initial conditions and has positive Lyapunov exponent. Up to now, many time series proved to be chaotic and inherently deterministic. Chaotic time series is nonlinear, which exhibits irregular, unpredictable behavior. The unpredictability of the chaotic time series accumulates with time, so long term behavior is difficult to predict. Deterministic laws have been studied for scalar time series. According to these rules, many efforts have been focused on the chaotic time series modeling and forecasting[1, 2].

The modeling techniques can be generally cataloged into two basic classes, the local model[3, 4] and the global model [5,6]. The former uses local approximation for the nonlinear mapping and each point in the reconstructed phase space has its local linear model if there are enough neighbor points; the latter tries to fit the whole nonlinear attractor using nonlinear function. Neural networks play an important role in constructing a global model in the recent 20 years. The efforts include: Multi-Layer Perceptrons (MLPs)[5], Radial Basis Function (RBF)[6,7], Finite Impulse Response Neural Networks (FIRNN)[8]; kernel methods such as Support Vector Machines (SVMs)[9] and Gaussian Process (GP)[10]; Recurrent Neural Networks such as Nonlinear AutoRegressive (NAR) model[11], Recurrent Predictor Neural Network (RPNN)[12] and Echo State Networks (ESNs)[13].

One objective of these efforts is how to improve the prediction accuracy within the permission of chaotic rules. In practice, the noise containing in time series could be more harmful to the prediction model than the chaotic nature itself. The prediction performance could be bad because both the input and the target output of the predictor are subject to noise. To weaken the influence of noise, one can resort to the nonlinear noise reduction technique before modeling[12, 14-15]. The noise reduction has been proved to be an efficient preprocessor step in practical nonlinear time series prediction. However, the performance of noise reduction is not always ideal in many conditions, constructing a noise suppression model is still necessary.

In [7], the noise suppression is converted into the problem to determine the optimum hidden units number in RBF neural networks, and in [6], the regularization parameter is taken as an indicator of the sufficiency of the given training data, when the data contains considerable noise and uncertainness, the regularization level should be strong to apply the prior smoothness constraint on the model. However, quantitative analysis is lack in these methods to indicate how the error in the independent variables (input variables) influences the prediction.

In local linear method, the well-known method is regularization[4], which is interpreted as one kind of filter factors to decrease the variance of the predictor. One drawback in existing results is that the modeling is taken as a problem of ordinary least-square and the random noise in coefficient matrix is ignored, since the noisy chaotic time series modeling is inherently an Errors-In-Variables (EIV) problem[16]. Another drawback is the relationship between the regularization parameter and the noise condition is not clearly

This research is supported by the project (60674073) of the National Nature Science Foundation of China, the project (2006BAB14B05) of the National Key Technology R&D Program of China and the project (2006CB403405) of the National Basic Research Program of China (973 Program). All of these supports are appreciated.

The authors are with the School of Electronic and Information Engineering, Dalian University of Technology, Liaoning 116023, China (e-mail: minhan@dlut.edu.cn).

explained, so the determination of regularization parameter has to resort to some techniques such as cross validation.

In this study, local linear model is studied from the perceptive of the EIV modeling, and the predictor is designed by considering the noise both in coefficient matrix and right-hand side. The optimal solution can be obtained by second order convex programming (SOCP) if given a perturbation bound of the noise. It is easy to show that the solution can be reformulated as a form of Tikhonov-type regularization, and it will also be shown how the regularization parameter is related to the Frobenius norm of the noise containing in coefficient matrix and right-hand side.

This paper is organized as following: In section II, main results about the local linear model for noisy chaotic time series are presented, and simulation results are given in section III; Some conclusions are given in section IV.

II. Noisy Chaotic Time Series By Local Linear Model With Tikhonov Regularization

A. Noisy Chaotic Time Series Prediction Problem and Tikhonov Regularization

For a given chaotic time series $\{x(k), k = 1, 2, 3, \ldots\}$, the delay embedding state vector is defined as:

$$\mathbf{d}(k) = [\,x(k),\quad x(k-\tau),\quad x(k-2\tau),\quad \ldots,\quad x(k-(m-1)\tau)\,]^T$$

where m and τ are the embedding parameters of the time series. According to the Takens' theorem[17], if the embedding dimension m is large enough, the evolution of delay embedding vector $\mathbf{d}(k)\rightarrow\mathbf{d}(k+1)$ can recover the original dynamic system without ambiguity. It is assumed that the evolution of the delay embedding vector is described by: $\mathbf{d}(k+1) = \mathrm{F}(\mathbf{d}(k))$, and we assume $x(k)$ is available through a measurement function $g(\cdot)$:

$$x(k) = g(\mathbf{d}(k)).$$

In the method of global modeling, for example, to get an h-step ahead predictor, one can train a neural networks to approximate the mapping: $\mathbf{d}(k)\rightarrow x(k+h)$.

The noisy reference point at time k is denoted by:

$$\mathbf{d}^+(k) = [\,x(k)+\Delta x(k),\quad x(k-\tau)+\Delta x(k-\tau),\quad x(k-2\tau)+\Delta x(k-2\tau),\quad \ldots,\quad x(k-(m-1)\tau)+\Delta x(k-(m-1)\tau)\,]^T$$

where $\Delta x(k)$ is the noise part of the time series at time k.

Let $\{\mathbf{d}_1^+(k), \mathbf{d}_2^+(k), \mathbf{d}_3^+(k), \ldots, \mathbf{d}_M^+(k)\}$ be the M nearest neighbors of $\mathbf{d}^+(k)$ in the noisy reconstructed phase space, and let the coefficient matrix $X\in\mathrm{R}^{M,\,m}$:

$$X = \left[\mathbf{d}_1^+(k)\quad \mathbf{d}_2^+(k)\quad \ldots\quad \mathbf{d}_M^+(k)\right]^T = \begin{bmatrix} x_1(k)+\Delta x_1(k) & x_1(k-\tau)+\Delta x_1(k-\tau) & \cdots & x_1(k-(m-1)\tau)+\Delta x_1(k-(m-1)\tau) \\ x_2(k)+\Delta x_2(k) & x_2(k-\tau)+\Delta x_2(k-\tau) & \cdots & x_2(k-(m-1)\tau)+\Delta x_2(k-(m-1)\tau) \\ \vdots & \vdots & \cdots & \vdots \\ x_M(k)+\Delta x_M(k) & x_M(k-\tau)+\Delta x_M(k-\tau) & \cdots & x_M(k-(m-1)\tau)+\Delta x_M(k-(m-1)\tau) \end{bmatrix} \quad (1)$$

and vector $\mathbf{y}\in\mathrm{R}^M$:

$$\mathbf{y} = \left[x_1(k+h)+\Delta x_1(k+h)\quad x_2(k+h)+\Delta x_2(k+h)\quad \cdots\quad x_M(k+h)+\Delta x_M(k+h)\right]^T \quad (2)$$

Let $\{\mathbf{d}_1(k), \mathbf{d}_2(k), \mathbf{d}_3(k), \ldots, \mathbf{d}_M(k)\}$ be the above M phase points in the noiseless phase space (without the noise part), and let the coefficient matrix $X_0\in\mathrm{R}^{M,\,m}$:

$$X_0 = \left[\mathbf{d}_1(k)\quad \mathbf{d}_2(k)\quad \ldots\quad \mathbf{d}_M(k)\right]^T = \begin{bmatrix} x_1(k) & x_1(k-\tau) & \ldots & x_1(k-(m-1)\tau) \\ x_2(k) & x_2(k-\tau) & \ldots & x_2(k-(m-1)\tau) \\ & & \cdots & \\ x_M(k) & x_M(k-\tau) & \ldots & x_M(k-(m-1)\tau) \end{bmatrix}$$

and vector $\mathbf{y}_\mathrm{d}\in\mathrm{R}^M$:

$\mathbf{y}_0 = [x_1(k+h), x_2(k+h), x_3(k+h), \ldots, x_M(k+h)]^T$,

where $X=X_0+\Delta X$, $\mathbf{y}=\mathbf{y}_0+\Delta\mathbf{y}$.

The M phase points $\{\mathbf{d}_1(k), \mathbf{d}_2(k), \mathbf{d}_3(k), \ldots, \mathbf{d}_M(k)\}$ cannot be guaranteed to be the M nearest neighbors of noiseless phase point $\mathbf{d}(k)$ in the noiseless phase space. In the noisy reconstructed state space, both the reference point $\mathbf{d}^+(k)$ and potential neighbors are polluted by noise, and it will lead to systematic bias in the local linear model if the noise level is considerable, and it is the manner how the noise influence the local model. So it is necessary to perform a nonlinear noise reduction preprocessor step before the local model is built [15, 18]. If the noise level is small, the data set X_0 is approximately linear with $\mathbf{y}_0$, that is to say, one can find a local linear model with $\mathbf{w}\in\mathrm{R}^m$ and scalar $b\in\mathrm{R}$ to fit the linear model:

$$y_0 = X_0\cdot\mathbf{w}+b$$

and the h-step ahead prediction for the noiseless prediction origin $\mathbf{d}(k)$ is given by:

$$y_\mathrm{d} = \mathbf{w}^T\mathbf{d}(k)+b$$

and the h-step ahead prediction for the noisy prediction origin $\mathbf{d}^+(k)$ is given by:

$$y_{\mathrm{d}+} = \mathbf{w}^T\mathbf{d}(k)+b$$

In prediction with noisy data, the original least squares solution can have large variance. Regularization methods, designed to be robust against noise, may provide better results. The regularization is the technique of improving the original ill-posed estimation method with the goals of stabilizing the solution and obtaining a meaning solution [4]. There are mainly two categories of regularization techniques: truncation methods[19] and Tikhonov-type methods[20]. In truncation methods, "high frequency" components are cut out of the original problem so that the remaining part of the problem is well-conditioned. In Tikhonov-type regularization, a penalty term is added, for example, the norm of the variable, so that a trade-off obtained between the actual model fitting and the variations in the solution. The Tikhonov regularization problem in linear least square case is formulated as:

$$\min_{\mathbf{w}} \|X\mathbf{w}-\mathbf{y}\|_2^2 + \mu\|\mathbf{w}\|_2^2 \tag{3}$$

which has the closed-form solution:

$$\hat{\mathbf{w}} = [X^T X + \mu I]^{-1} X^T \mathbf{y} \tag{4}$$

where μ>0 is the regularization parameter. The method is also called ridge regression (RR). The ill-conditioning of the original least squares can be avoided in Tikhonov solution by a properly chosen value of μ.

Generalized cross-validation (GCV) [21] offers a way to estimate appropriate values for the regularization parameters μ. The GCV criterion is a constant multiple of:

$$V(\mu) = \frac{\|(I-H)\mathbf{y}\|^2}{(\mathbf{trace}(I-H))^2} \tag{5}$$

where $H=X(X^TX+\mu \mathrm{I})^{-1}X^T$. The GCV estimate is based on the minimization of the PRESS criterion, which is equivalent to a leave-one-out cross validation. There are also several other criterion[22] such as the Akaike information criterion and the more recently subspace information criterion[23]. In this paper, we will focus on the "filter factors" and "Errors-in-Variables" interpretations.

B. The "filter factors Interpretation" Revisited and the Limitations

The SVD of $X\in \mathrm{R}^{M,m}$ is defined as:

$$X = U\Sigma V^T, \quad U^T U = I, \quad V^T V = I,$$

$$\Sigma = diag(\sigma_1, \sigma_2, ..., \sigma_r),$$

Where $r \le \min(M, m)$ is the rank of X and $\sigma_1 \ge \sigma_2 \ge \ldots \sigma_r \ge 0$ are the non-zero singular values of X.

Consider the following estimator for $\mathbf{w}$:

$$\hat{\mathbf{w}} = V\Sigma^{-1}\Lambda U^T \mathbf{y} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}\left(\mathbf{u}_i^T \mathbf{y}\right)\mathbf{v}_i \tag{6}$$

Where $\Lambda = diag(\lambda_1,\ldots, \lambda_r)$ are the filter factors. The estimation differs in the choice of the diagonal elements of Λ in different techniques. The bias and variance for the prediction value y_{d+}(for noisy prediction origin $\mathbf{d}^+(k)$) are shown in [4]:

$$\mathrm{bias}(y_{d+}) = \sum_{i=1}^{r} (\lambda_i - 1)(\mathbf{v}_i^T \mathbf{d}(k))(\mathbf{v}_i^T \mathbf{w}) \tag{7}$$

$$\mathrm{var}(y_{d+}) = \sigma^2 \sum_{i=1}^{r} \frac{\lambda_i^2}{\sigma_i^2}\left\{\left((\mathbf{v}_i^T \mathbf{d}(k))\right)^2 + \sigma_i^2(\mathbf{v}_i^T \mathbf{w})^2 + \sigma^2\right\} \tag{8}$$

The filter factors for the Tikhonov regularization estimator(as shown in Eq.(4)) are given by:

$$\lambda_i = \frac{\sigma_i^2}{\sigma_i^2 + \mu} \tag{9}$$

Since $\mu > 0$, $0 < \lambda_i < 1$ for any i. It is clear from the Eqs.(7)-(8) that the reduction of the filter factors λ_i will decrease the variance of the predictor at the possible cost of introducing extra bias in Eq.(7). The optimal trade-off between bias and variance consists of finding the best filter factors, in other word, finding the regularization parameter μ for the Tikhonov regularization estimator (Eq.(4)).

In the results described above, there are some limitations and unsolved questions:

1) It has been shown that the existence of the optimal trade-off between bias and variance, but there is no quantitative explanation of what the trade-off actually is. Although the regularization parameter can be determined by cross-validation (for example, as shown by Eq.(5)), the relationships between the noise condition and the regularization parameter are remained unexplored.

2) The modeling of noisy chaotic time series is inherently an Errors-In-Variables problem, however, the explanation of filter factors ignores the random fluctuations in X, and takes it as a fixed value. It has been pointed in [4], the analysis of influence of noise on the prediction model is not available.

C. The "Errors-in-Variables Interpretation" and the Regularization Parameter

When the time series being processed contains noise, both the coefficient matrix X and the right-side vector $\mathbf{y}$ are corrupted by noise.

For a given chaotic time series $\{x(i), i=1, 2, 3,\ldots, N\}$, polluted by additive noise sequence $\{\Delta x(i), i=1, 2, 3,\ldots, N\}$. The coefficient matrix X and the right-hand side $\mathbf{y}$ of the local linear model are denoted by Eqs.(1)-(2). Since both the coefficient matrix X and the right-hand side $\mathbf{y}$ are corrupted by random noise, the total least square is a suitable method[24]. Because the uncertainness in X and y is the interest, the following formulation starts from the assumption of knowing the information of the matrix $[\Delta X\ \Delta \mathbf{y}]$. In the robust counterpart of total least square, the problem is solved by finding the local linear model to minimize the worst prediction error[25-27]:

$$\min_{\mathbf{w}} \max_{\|\Delta X\ \Delta \mathbf{y}\|_F \le \rho} \|(X+\Delta X)\mathbf{w} - (\mathbf{y}+\Delta\mathbf{y})\| \tag{10}$$

if the bound of Frobenius norm ρ of matrix $[\Delta X\, \Delta \mathbf{y}]$ is known.

By triangle inequality, the worst prediction error in Eq.(10) is:

$$\|X\cdot\mathbf{w}-\mathbf{y}\|\cdot\rho^{-1} + \sqrt{\|\mathbf{w}\|^2+1} \tag{11}$$

The minimization of Eq.(11) can be reformulated as the second-order cone program[28]:

$$\min \lambda$$
$$s.t.\ \|X\mathbf{w}-\mathbf{y}\| \le \rho(\lambda-\tau), \qquad (12)$$
$$\left\|\begin{bmatrix}\mathbf{w}\\ 1\end{bmatrix}\right\| \le \tau$$

The local linear model can be obtained if the solution to Eq.(12) is found. The second-order cone program belongs to a kind of convex optimization problems, and the solution can be effectively obtained.

The dual problem to Eq.(12) is:

$$\max\ (\mathbf{y}\rho^{-1})^T\mathbf{z}-\mathbf{v}$$
$$s.t.\ (X\rho^{-1})^T\mathbf{z}+\mathbf{u}=0$$
$$\|\mathbf{z}\|\le 1, \qquad (13)$$
$$\left\|\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}\right\| \le 1$$

Using the constraints, it is followed that

$$z=-\frac{X\mathbf{w}-\mathbf{y}}{\|X\mathbf{w}-\mathbf{y}\|} \text{ and } \begin{bmatrix}\mathbf{u}^T & \mathbf{v}\end{bmatrix}=-\frac{\begin{bmatrix}\mathbf{w}^T & 1\end{bmatrix}}{\sqrt{\|\mathbf{w}\|^2+1}} \qquad (14)$$

From the equation constraints, the optimal **w** is:

$$\mathbf{w}=((X\rho^{-1})^T X\rho^{-1}+\mu_0 I)^{-1}(X\rho^{-1})^T\mathbf{y}\rho^{-1},$$
$$\text{with } \mu_0=\frac{\lambda-\tau}{\tau}=\frac{\|X\rho^{-1}\mathbf{w}-\mathbf{y}\rho^{-1}\|}{\sqrt{\|\mathbf{w}\|^2+1}} \qquad (15)$$
$$\mathbf{w}=[X^TX+\mu I]^{-1}X^T\mathbf{y} \qquad (16)$$
$$\mu=\mu_0\rho^2=\frac{(\lambda-\tau)\rho^2}{\tau} \qquad (17)$$

From the Eq.(17), it is known that the regularization parameter is quantitatively related to the noise condition. In Eq.(17), λ and τ are the results of the second-order cone program of Eq.(12) and they can be determined if the norm information of ρ is provided.

As shown above, Tikhonov regularization in the local linear model for noisy chaotic time series can be well explained by the Errors-in-Variables method. The first advantage is that the relationship between the noise condition and the regularization parameter μ is clearly indicated, and another one is that the explanation is based on the Errors-In-Variables model.

Now, there are at least two kind of methods to obtain the local linear model: the one is to determine the regularization parameter μ by some techniques such as cross validation, and another one is to solve the second-order cone program posed by Eq.(12) using the norm information of the matrix [$\Delta X\ \Delta\mathbf{y}$].

III. SIMULATIONS

Since the Eq.(10) describes a local model, there exists a linear model for each point in the reconstructed phase space. To solve the problem of Eq.(12), it is the best choice to take the exact value of Frobenius norm of matrix [$\Delta X\ \Delta\mathbf{y}$] as the bound ρ for each local model. For the bound ρ, Smaller or larger values than $\|[\Delta X\ \Delta\mathbf{y}]\|_F$ are not good choices. The method is indicated by **SOCP by Exact** ρ in the following demonstration.

In practical application, the exact information of norm for matrix [$\Delta X\ \Delta\mathbf{y}$] is always unknown, since the noise is random. $\|[\Delta X\ \Delta\mathbf{y}]\|_F$ can be estimated from the distribution of noise. The method is indicated by **SOCP by Estimated** ρ.

To determine the local prediction model, the existing method is to use ridge regression, and the regularization parameter can be determined by GCV, and this method is indicated by **RR by GCV**. The results of the ordinary least squares (a non-EIV model) are also presented for comparisons. The method is indicated by **OLS**.

The simulation is conducted in the MATLAB environment running on HP Workstation xw8000, and the second-order cone program is implemented by CVX[29] and SeDuMi[30] toolbox.

A. Prediction of the Noisy Henon Time Series

The Henon system is described by the following coupled equations:

$$x(k+1)=1-a\cdot x^2(k)+y(k)$$
$$y(k+1)=b\cdot x(k)$$

where parameters a =1.4, b = 0.3, and initial observation (x_0, y_0) as (0, 0) are chosen to generate a time series. In constructing the local linear model, the embedding dimension is set to 4, and the noisy embedded data vector is:

$$\mathbf{d}^+(k)=[\,x(k)+\Delta x(k),\quad x(k-\tau)+\Delta x(k-\tau),$$
$$x(k-2\tau)+\Delta x(k-2\tau),\quad x(k-3\tau)+\Delta x(k-3\tau)\,]^T$$

at time k, where $\tau = 1$ in the simulation. The target output is the one-step-ahead value ($y(k) = x(k + 1)$) of the time series.

Length of the training sequence is 2000. Zero mean Gaussian noise is added to the original Mackey-Glass time series, and the noise level is 5%, the noise level is chosen the same as [4].

In this simulation, the prediction performance is measured by the root mean squared error on the new test sequence pairs normalized by the variance of the original time series (**NRMSE**).

$$\textbf{NRMSE}=\sqrt{\frac{\sum_{i=1}^{T_n}(y_{d+}(i)-y(i))^2}{T_n\cdot\sigma^2}}$$

where $y(i)$ denotes the target value, $y_{d+}(i)$ is the corresponding prediction output, T_n is the number of the test examples, and σ^2 is the variance of the original time series. T_n=500 in the simulation.

Table 1 contains the prediction results by the four kinds of methods. For each method, the number of neighbor M is chosen as 15. The NRMS errors are 7.8680e-2 and 8.0572e-2 for the **SOCP by Exact** ρ and **SOCP by Estimated** ρ, which is better than the result of the **RR by GCV**(8.3915e-2). The

NRMS error for the **OLS** is 9.0717e-2, which is the worst prediction.

Table 1 NRMS error of one-step prediction of the Henon time series pullulated by 5% Gaussian noise

Method	**NRMS** error
SOCP by Exact ρ	7.8680e-2
SOCP by Estimated ρ	8.0572e-2
RR by GCV	8.3915e-2
OLS	9.0717e-2

In the method of the **SOCP**, the key problem is the information about $[\Delta X\ \Delta\mathbf{y}]$. If the exact information of the uncertainness is known, the bound ρ for the matrix $[\Delta X\ \Delta\mathbf{y}]$ can be exactly determined for each local model, as a result, the model can be accurate. If the distribution of the noise is the only known information, one have to estimate ρ, as a result, many local models will share the same ρ if the number of neighbor is the same, and the prediction will be less accurate. The result is confirmed by the comparison in Table 1.

The existing method of the **RR by GCV** provides another method to determine the regularization parameter. In fact, GCV belongs to a kind of generalization error criteria for the model selection, and the optimal regularization parameter is chosen by the cross validation techniques. The method relies on the condition of the data, and does not use any prior information about the uncertainness. As shown in Table 1, it is not as good as the SOCP methods.

B. Prediction of the Noisy Mackey-Glass Time Series

The Mackey-Glass system is a time-delay differential system with the form:

$$\frac{dx}{dt} = \beta x(t) + \frac{\alpha x(t-\delta)}{1 + x(t-\delta)^{10}}$$

where $x(t)$ is the value of time-series at time t. The system is chaotic for $\delta > 16.8$. The parameter values are chosen as $\beta = 0.1$, $\alpha = 0.2$ and $\delta = 17$. The data set is constructed using second-order Runge-Kutta method with a step size of 0.1. The noisy embedded data vector consists of four values of the time series:

$$\mathbf{d}^+(k) = [\,x(k) + \Delta x(k), \quad x(k-\tau) + \Delta x(k-\tau), \quad x(k-2\tau) + \Delta x(k-2\tau), \quad x(k-3\tau) + \Delta x(k-3\tau)\,]^T$$

at time k, where $\tau = 6$ in the simulation. The target output is the 84-step-ahead value ($y(k) = x(k + 84)$) of the time series.

Length of the training sequence is 2000. Zero mean Gaussian noise is added to the original Mackey-Glass time series, and the noise level is 5%, and length of the test sequence is 500.

The number of neighbor M is chosen as 10 in each local model. Table 2 lists the NRMS error for the four methods. The NRMS errors of the two SOCP methods are respectively 0.1355 (**SOCP by Exact** ρ) and 0.1383 (**SOCP by Estimated** ρ), and the former is slightly better than the latter. The result of the **RR by GCV** is 0.1469, which is worse than the two SOCP methods. The worst result is obtained by the **OLS** method, and the NRMS error is 0.1702.

Table 2 NRMS error of 84-step prediction of Mackey-glass time series pullulated by 5% Gaussian noise

Method	**NRMS** error
SOCP by Exact ρ	0.1355
SOCP by Estimated ρ	0.1383
RR by GCV	0.1469
OLS	0.1702

C. The Analysis of the Results

As shown in the simulations, the SOCP methods obtain slightly better results than that of the **RR by GCV**. Both of them are better than the **OLS** method. In fact, as shown in Section II, the solutions for the SOCP methods and **RR by GCV** have the same form:

$$\mathbf{w} = [X^T X + \mu I]^{-1} X^T \mathbf{y}$$

as shown in Eq.(4) and Eq.(16).

The differences lie in the manner of the regularization parameter determination. In the methods of SOCP, the regularization parameter is automatically determined in the optimal process if provided the information of matrix $[\Delta X\ \Delta\mathbf{y}]$, as shown by Eq.(17). When the exact information of $[\Delta X\ \Delta\mathbf{y}]$ is known, the prior uncertainness is imposed on the local model exactly, and it is why it is better to known the exact ρ than the estimated ρ. In the **RR by GCV**, the regularization parameter is determined by the GCV method, without knowing any information of the uncertainness in the data, but the method is influenced by the condition of local data set.

In fact, the RR method can be derived from the perspective of Errors-In-Variables model, which has been shown from Eqs.(10)-(17). However, the "filter factors Interpretation" does not take the RR as one of the Errors-In-Variables method, and the "Errors-in-Variables Interpretation" gives the answer.

If provided the information of $[\Delta X\ \Delta\mathbf{y}]$, the Tikhonov regularization is implemented in the SOCP process. This point is similar to the fact that the Tikhonov regularization can also be implemented by quadratic program (QP) in support vector machines. The new results make use of the SOCP, the reason lies in the need to impose the information of uncertainness on the local model, and the explanation is from the perspective of the Errors-In-Variables method.

IV. Conclusions

This paper studies the chaotic time series prediction problem and Tikhonov-type regularization method for noisy chaotic time series. The existing "filter factors Interpretation" has two drawbacks in the explanation of Tikhonov regularization in local model. One drawback in the interpretation is the ignorance of the random noise in coefficient matrix, another drawback is the relationship between the regularization parameter and the noise condition is not clearly explained.

Based on the local data sets, the Errors-In-Variables model is constructed since both the coefficient matrix and the right-hand side are subject to noise. The optimal solution can be obtained by second order convex program if given a perturbation bound of the noise, and the solution can be reformulated as a form of ridge-regression, and it has been shown how the regularization parameter is related to the Frobenius norm of the noise containing in coefficient matrix and right-hand side. The results also indicate that the ridge-regression can also be derived from the perspective of the Errors-In-Variables model if the regularization is properly chosen.

ACKNOWLEDGMENT

This research is supported by the project (60674073) of the National Nature Science Foundation of China, the project (2006BAB14B05) of the National Key Technology R&D Program of China and the project (2006CB403405) of the National Basic Research Program of China (973 Program). All of these supports are appreciated.

REFERENCES

[1] M. Casdagli, “Nonlinear prediction of chaotic time series,” *Physica D*, vol.35, no.3, pp.335-356, 1989.

[2] J. C. Principe, A. Rathie, and J. M. Kuo, “Prediction of Chaotic Time Series with Neural Networks and the Issue of Dynamic Modeling,” *Int.J. Bifurcation and Chaos*, vol.2, pp.989-996, 1992.

[3] J. D. Farmer and J. J. Sidorowich. "Predicting Chaotic Time Series." *Phys. Rev. Lett.* vol.59, no.8, pp.845-848, 1987.

[4] D. Kugiumtzis, O. C. Lingjaerde, and N. Christophersen, "Regularized local linear prediction of chaotic time series," *Physica D*, vol. 112, no. 3-4, pp. 344-360, 1998.

[5] A. Lapades, and R. Farbar, “How neural nets work,” in *Advances in Neural Information Processing Systems*, pp.442-456, 1987.

[6] S. Haykin and J. Principe, "Making sense of a complex world," *IEEE Signal Processing Magazine*, vol. 15, no. 3, pp. 66-81, 1998.

[7] H. Leung, T. Lo, and S. Wang, "Prediction of noisy chaotic time series using an optimal radial basis function neural network," *IEEE Transactions on Neural Networks*, vol.12, no.5, pp. 1163-1172, 2001.

[8] E. A. Wan, “Time series prediction by using a connectionist network with internal delay lines (data set A),” in *Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis*, May 14-17 1992, pp. 195-217, 1993.

[9] K.-R. Müller, A. J. Smöla, G. Ratsch, B. Scholköpf, J. Kohlmorgen, and V. N. Vapnik, “Predicting time series with support vector machines,” in the *Proceedings of the 1997 7th International Conference on Artificial Neural Networks*, ICANN’97, Oct 8-10 1997,” Lecture Notes in Computer Science, vol. 1327 ed. Lausanne, Switz, 1997, pp. 999-1004.

[10] A. Girard, C. E. Rasmussen and J. Qui, “Gaussian Process Priors with Uncertain Inputs - Application to Multiple-Step Ahead Time Series Forecasting,“ in *Advances in Neural Information Processing Systems* 15, pp.529-536, Cambridge, MA: MIT Press, 2003.

[11] J. C. Principe, J. M. Kuo, “Dynamic modeling of chaotic time series with neural networks,” *Advances in Neural Information Processing Systems* 7, pp.311-318, Cambridge, MA: MIT Press, 1995.

[12] M. Han, J. Xi, S. Xu, and F.-L. Yin, "Prediction of chaotic time series based on the recurrent predictor neural network," *IEEE Transactions on Signal Processing*, vol. 52, no.12, pp. 3409-3416, 2004.

[13] H. Jaeger, and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication,” *Science*, Vol.304, No.5667, pp.78-80, 2004.

[14] K. Jinno, S. G. Xu, A. KAWAMURA and M. MATSUMOTO, “Prediction of Sunspots Using Reconstructed Chaotic System Equations,” *Journal of Geophysical Research-Space Physics*, vol.100, no.A8, pp.14773-14781, 1995.

[15] M. Han, Y. Liu, J. Xi, and W. Guo, "Noise Smoothing for Nonlinear Time Series Using Wavelet Soft Threshold," *IEEE Signal Processing Letters*, vol. 14, no. 1 pp. 62-65, 2007.

[16] H. Kantz and L. Jaeger, "Improved cost functions for modelling of noisy chaotic time series," *Physica D*, vol. 109, no. 1-2, pp. 59-69, 1997.

[17] F. Takens, Detecting Strange Attractors in Fluid Turbulence, in D.Rand and L.-S. Young(eds.) *Dynamical Systems and Turbulence*, Berlin, Springer, 1981.

[18] A. Leontitsis, T. Bountis, and J. Pagge, "An adaptive way for improving noise reduction using local geometric projection," *Chaos*, vol. 14, no. 1, pp. 106-110, 2004.

[19] R. D. Fierro, G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Regularization by truncated total least squares," *SIAM Journal on Scientific Computing*, vol. 18, no. 4, pp. 1223-1241, 1997.

[20] D. M. Sima, Regularization techniques in model fitting and parameter estimation, PhD thesis, Faculty of Engineering, K.U.Leuven, 2006.

[21] R. H. Myers, *Classical and Modern Regression with Applications*. Second Edition, Boston: PWS-KENT, 1990

[22] T. Nakamura, K. Judd, and A. Mees, "Refinements to model selection for nonlinear time series," *International Journal of Bifurcation and Chaos,* vol. 13, no.5, pp. 1263-1274, May 2003.

[23] M. Sugiyama and H. Ogawa, "Subspace information criterion for model selection," *Neural Computation,* vol. 13, no.8, pp. 1863-1889, Aug 2001.

[24] G. H. Golub, C. F. V. Loan, An Analysis of the Total Least Squares Problem, *SIAM Journal on Numerical Analysis*, vol. 17, no. 6, pp. 883-893, December 1980.

[25] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded modeling errors," *IEEE Signal Processing Letters*, vol. 4, no. 7, pp. 195-197, 1997.

[26] L. ElGhaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035-1064, 1997.

[27] G. A. Watson, Robust counterparts of errors-in-variables problems, in *Proceedings of 4th International Workshop on Total Least Squares and Errors-in-Variables Modeling*, Arenberg castle, Leuven, Belgium, August 21-23, 2006.

[28] S. Boyd, L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[29] M. Grant, S. Boyd, Y. Ye, CVX: Matlab Software for Disciplined Convex Programming, Version 1.0 beta 3 (June 2006), http://www.stanford.edu/~boyd/cvx, 2006.

[30] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," Optimization Methods & Software, vol. 11-2, no. 1-4 pp. 625-653, 1999.