Multi-step prediction of *Dst* index using singular spectrum analysis and locally linear neurofuzzy modeling

Javad Sharifi, Babak N. Araabi, and Caro Lucas

*Control and Intelligent Processing Center of Excellence, Electrical and Computer Engineering Department, University of Tehran,
Tehran, Iran and School of Intelligent Systems, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran*

Even one-step prediction of natural time series without delay especially in main phase of storm is difficult for many complicated time series such as *Dst* index. In this study, with a new method based on singular spectrum analysis, we extract the main components of the time series, model each component with a locally linear neurofuzzy network, and utilize the trained networks for multi-step ahead prediction of a validation set of data, and finally combine the predicted patterns for construction of general prediction. Our methods are compared with several previous studies for *Dst* index prediction. Several solar geomagnetic extreme events are predicted well with our state-of-the-art method; such as extreme events in 14 March 1989 that led to power black-out in Quebec, as well as other extreme storms.
Key words: Singular spectral analysis, locally linear neurofuzzy model, *Dst* prediction.

1. Introduction

Disturbance storm time (*Dst*) index is intended to be a direct measure of the symmetric ring current. It is a well known fact that rapid variations of the geomagnetic field can be harmful to technological systems. But the ring current variations are normally not very rapid compared to the sub storm effects at higher latitudes. Further, the magnitude of *Dst* variations is smaller than the magnitude of high-latitude magnetic disturbances. Why then should we make *Dst* forecasts? *Dst* approximately measures the strength of the ring current, and the ring current is one of the major current systems of the magnetosphere. An enhanced ring current, indicated by a decrease of the *Dst* index, indicates a major impact on the structure and location of the magnetosphere regions and the boundaries that separate them. On a global scale, the ring current generates a magnetic moment which augments the Earth's magnetic moment as presented to the solar wind. For these reasons, many *specification models*—i.e. models that specify the state of the near-Earth space environment—are parameterized in *Dst*. Operational *Dst* forecasts provide us with inputs to such models, and tell us much about current and upcoming *space weather* conditions. Another reason to make *Dst* forecasts is that some of the most adverse effects of space weather take place during magnetic storms, and the ring-current strength is the basic defining property of a magnetic storm. The *Dst* index can thus be used as a proxy for many types of disturbance that occur during a storm, even though the space-weather effects are not directly caused by the ring-current magnetic field variations. For prediction of geomagnetic storms, monitored by *Dst* index, a number of data driven tech-

niques are available (Kamide *et al.*, 1998; Detman and Vassilaidis, 1997; Joselyn, 1995). The more successful techniques include statistical time series analysis (Baker, 1986), linear prediction filters (Iyemori *et al.*, 1979), and linear and nonlinear autoregressive and moving average filters including local linear prediction (Vassiliadis *et al.*, 1995). Several of these techniques are being implemented and tested for reliable prediction of geomagnetic indices under real time conditions, e.g. a linear filter for predicting Kp. An AI technique is used internally in the Magnetospheric Specification and Forecast Model (MSFM) to predict *Dst* (Freeman *et al.*, 1994). Several neural models (Munsami, 2000; Gleisner *et al.*, 1996; Wintoft, 1997; Wintoft and Lundstedt, 1998; Wu *et al.*, 1998a, b) as well as analogue models (differential model) based upon physical knowledge (Nagatsuma, 2002; Burton *et al.*, 1975; Fenrich and Luhmann, 1998; O'Brien and McPherron, 2000) are applied for geomagnetic storms prediction. Most of these methods are using solar wind data (solar wind magnetic field component Bz, density n, and velocity V, etc) or a very large training dataset to make short-term prediction of some storms in the test set. Powerful black box neural and neurofuzzy methods fail when the prediction the time horizon increases and the number of training data decreases, especially when training set contains some few number of storms and contains minor storms but test set contains some extreme storms. In this article a novel combination of nonlinear and linear techniques that has been successfully utilized for long term prediction of space weather (Gholipour *et al.*, 2005) has been adopted for multi-step forecasting *Dst* values. The paper is organized as follows: In Section 2, singular spectral analysis (SSA) method is introduced and in Section 3, locally linear neurofuzzy (LLNF) model is introduced for Black-Box modeling. Section 4 covers the one-step prediction. In Section 5, LOLIMOT and LOLIMOT + SSA are applied for

multi-step prediction.

2. Singular Spectrum Analysis

SSA is a tool to extract information from short and noisy chaotic time series (Vautard *et al.*, 1992). It relies on the Karhunen-Loeve decomposition of an estimate of covariance matrix based on M lagged copies of the time series. Thus as the first step, the embedding procedure is applied to construct a sequence $\{\tilde{X}(t)\}$ of M-dimensional vectors from time series $\{X(t) : t = 1, \ldots, N\}$:

$$\tilde{X}(t) = (X(t), X(t+1), \ldots, X(t+M-1)),$$
$$t = 1, \ldots, N', \quad N' = N - M + 1. \quad (1)$$

The $N' \times M$ trajectory matrix (D) of the time series has the M dimensional vectors as its columns, and is obviously a Hankel matrix (the elements on the diagonals $j + j = constant$ are equal). In the second step, the $M \times M$ covariance matrix C_X is calculated as

$$C_X = \frac{1}{N'} D^T D \quad (2)$$

C_X Eigenelements can be determined by Singular Value Decomposition (SVD):

$$C_X = U \Sigma V^T; \quad U^T U = I, \quad V^T V = I. \quad (3)$$

The elements of diagonal matrix $\Sigma = [\text{diag}(\sigma_1, \ldots, \sigma_M)]$ are the singular values of D and are equal to square roots of the C_X eigenvalues. The C_X eigenelements $\{(\lambda_k, \rho_k) : k = 1, \ldots, M\}$ are obtained from

$$C_X \rho_k = \lambda_k \rho_k. \quad (4)$$

Each eigenvalue, λ_k, estimates the partial variance in the ρ_k direction, and the sum of all eigenvalues equals the total variance of the original time series. In the third step, the time series is projected onto each eigenvector, and yields the corresponding principal component (PC) for each $PC_k(t)$:

$$PC_k(t) = \sum_{j=1}^{M} X(t+j-1)\rho_k(j). \quad (5)$$

Each of the principal components, a nonlinear or linear trend, a periodic or quasi-periodic pattern, or a multi-periodic pattern, has a narrow band frequency spectrum and well defined characteristics to be estimated. As the fourth step, the time series is reconstructed by combining the associated principal components:

$$R_K(t) = \frac{1}{M_t} \sum_{k \in K} \sum_{j=L_t}^{U_t} PC_k(t-j+1)\rho_k(j). \quad (6)$$

The normalization factor (M_t), and the lower (L_t) and upper (U_t) bounds of reconstruction procedure differ for the center and edges of the time series, and are defined by following formula

$$(M_t, L_t, U_t) = \begin{cases} (t, 1, t), & 1 \leq t \leq M-1 \\ (M, 1, M), & M \leq t \leq N' \\ (\min(t, N-t+1), t-N+M, M), \\ \quad N'+1 \leq t \leq N. \end{cases}$$
$$(7)$$

To enhance signal to noise ratio, one can use the singular spectrum (the logarithmic scale plot of singular values of covariance matrix in decreasing order). The principal components related to lower singular values can be omitted in the reconstruction stage, to obtain adaptive noise cancellation. If all the components are used in reconstructing the time series, no information is lost.

3. Locally Linear Neurofuzzy with Model Tree Learning

The main idea for utilizing the locally linear neurofuzzy (LLNF) model for function approximation is dividing the input space into small linear subspaces with fuzzy validity functions, $\phi_i(\underline{u})$, which describe the validity of each linear model in its region. The validity function is used in this study is the normalized Gaussian function, defined as $\mu(x) = \exp(-\frac{(x-c)^2}{2\sigma^2})$, where c is the center and σ is the standard deviation of the Gaussian.

Each local linear subspace with its validity function is called a fuzzy neuron. Thus the total model is a neurofuzzy network with one hidden layer, and a linear neuron in the output layer which simply calculates the weighted sum of the outputs of locally linear models (LLMs) as follows:

$$\hat{y}_i = \omega_{i_0} + \omega_{i_1} u_1 + \omega_{i_2} u_2 + \cdots + \omega_{i_p} u_p,$$
$$\hat{y} = \sum_{i=1}^{M} \hat{y}_i \phi_i(\underline{u}). \quad (8)$$

Where $\underline{u} = [u_1 \, u_2 \, \cdots \, u_p]^T$ is the model input, M is the number of LLM neurons, and ω_{ij} denotes the LLM parameters of the ith neuron. The validity functions are chosen as normalized Gaussians; normalization is necessary for a proper interpretation of validity functions.

$$\phi_i(\underline{u}) = \frac{\mu_i(\underline{u})}{\sum_{j=1}^{M} \mu_j(\underline{u})} \quad (9)$$

$$\mu_i(\underline{u}) = \exp\left(-\frac{1}{2}\left(\frac{(u_1 - c_{i1})^2}{\sigma_{i1}^2} + \cdots + \frac{(u_p - c_{ip})^2}{\sigma_{ip}^2}\right)\right)$$
$$= \exp\left(-\frac{1}{2}\frac{(u_1 - c_{i1})^2}{\sigma_{i1}^2}\right)$$
$$\times \cdots \times \exp\left(-\frac{1}{2}\frac{(u_p - c_{ip})^2}{\sigma_{ip}^2}\right). \quad (10)$$

Each Gaussian validity function has two sets of parameters, centers (c_{ij}) and standard deviations (σ_{ij}) that are the parameters of the nonlinear hidden layer and the number of them is 2MP. Least square optimization method is used to adjust the parameters of local linear models (ω_{ij}) and learning algorithm would describe here could adjust the parameters of validity functions (c_{ij} and σ_{ij}). Global optimization of linear parameters is simply obtained by least squares technique. The complete parameter vector contains $M(p+1)$ elements:

$$\underline{\omega} = [\omega_{10} \, \omega_{11} \, \cdots \, \omega_{1p} \, \omega_{20} \, \omega_{21} \, \cdots \, \omega_{M0} \, \cdots \, \omega_{Mp}] \quad (11)$$

and the associated regression matrix $\underline{X}$ for N measured data

samples is

$$\underline{X} = [\, \underline{X}_1 \; \underline{X}_2 \; \cdots \; \underline{X}_M \,] \qquad (12)$$

$$\underline{X}_i = \begin{bmatrix} \phi_i(\underline{u}(1)) & u_1(1)\phi_i(\underline{u}(1)) & \cdots & u_p(1)\phi_i(\underline{u}(1)) \\ \phi_i(\underline{u}(2)) & u_1(2)\phi_i(\underline{u}(2)) & \cdots & u_p(2)\phi_i(\underline{u}(2)) \\ \vdots & \vdots & & \vdots \\ \phi_i(\underline{u}(N)) & u_1(N)\phi_i(\underline{u}(N)) & \cdots & u_p(N)\phi_i(\underline{u}(N)) \end{bmatrix}.$$

$$(13)$$

Thus

$$\hat{\underline{y}} = \underline{X}\hat{\underline{\omega}}; \quad \hat{\underline{\omega}} = (\underline{X}^T \underline{X} + \alpha I)^{-1} \underline{X}^T \underline{y}; \quad \alpha \ll 1 \qquad (14)$$

α is the regularization parameter for avoiding any near singularity of matrix $\underline{X}^T\underline{X}$ in Eq. (14). The remarkable properties of locally linear neurofuzzy model, its transparency and intuitive construction, lead to the use of least squares technique for rule antecedent parameters. In each iteration, the worst performing locally linear neuron is determined to be divided. All the possible divisions in the p dimensional input space are checked and the best is performed. The splitting ratio can be simply adjusted as 1/2, which means that the locally linear neuron is divided into two equal halves. The fuzzy validity functions for the new construction are updated; their centers are the centers of the new hypercubes, and the standard deviations are usually set as 0.7 times the width of the hypercube in that dimension. The learning algorithm is introduced as follows:

Learning Algorithm:

Locally Linear Model Tree (LOLIMOT) is an incremental tree-construction algorithm that partitions the input space by axis bisection in all directions of input space. It implements a heuristic search for the rule premise parameters and avoids a time consuming nonlinear optimization.

The LOLIMOT algorithm is described in five steps according to (Nelles, 1999, 2001):

1. **Start with an initial model**: start with a single LLM, which is a global linear model over the whole input space with $\Phi_1(\underline{u}) = 1$ and set $M = 1$. If there is a priori input space partitioning it can be used as the initial structure.
2. **Find the worst LLM**: Calculate a local loss function e.g. Mean Square Error (MSE) for each of the $i = 1, \ldots, M$ LLMs, and find the worst performing LLM.
3. **Check all divisions**: The worst LLM is considered for further refinement. The hyper rectangle of this LLM is split into two halves with an axis orthogonal split. Divisions in all dimensions are tried, and for each of the p divisions the following steps are carried out:

   a. Construction of the multi-dimensional membership functions for both generated hyper rectangles;
   b. Construction of all validity functions: in part a, only the membership function of LLM that is split would change and the membership function of other neurons not change, but all of the validity functions change that must be updated for all LLMs by Eq. (9).[1]
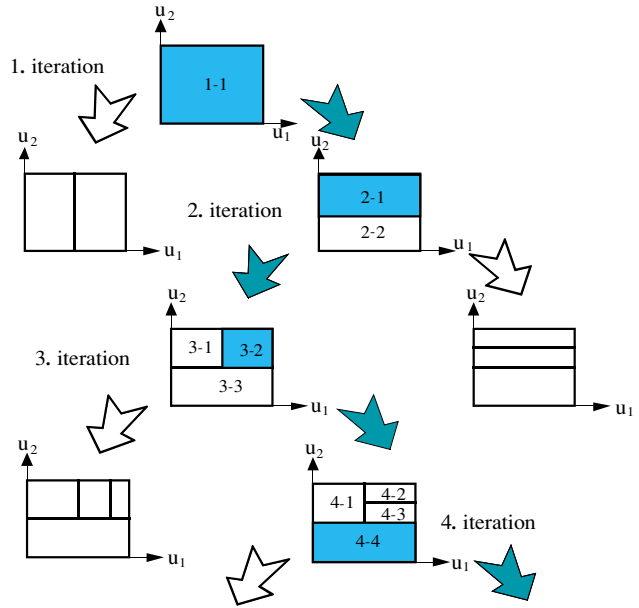

Fig. 1.   Illustration of LOLIMOT algorithm for two dimensional input space.

   c. Estimation of the rule consequent parameters for newly generated LLMs.
   d. Calculations of the loss function for the current overall model.

4. **Find the best division**: The best of the p alternatives checked in step 3 is selected, and the related validity functions and LLMs are constructed. The number of LLM neurons is incremented $M = M + 1$.
5. **Test the termination condition**: If the termination condition is met, then stop, else go to step 2.

The termination condition is reaching to a predefined error between output ($\underline{y}$) and LLNF output with M neuron ($\hat{\underline{y}}$), i.e. when the condition: $\|\underline{y} - \hat{\underline{y}}\| \le \varepsilon$ is satisfied. In practice we used a predefined number of neurons in LOLIMOT, plotted the error as a function of this number, and kept increasing the number of neurons until satisfactory performance was obtained.

In Fig. 1 the algorithm is represented for two dimensional input spaces.

4.   One-Step Ahead Prediction of *Dst* Index

Many neural and neurofuzzy methods like Multi Layer Perceptron (MLP) neural network (Kugblenu *et al.*, 1999) with back propagation learning algorithm (BP), Radial Basis Function (RBF) neural networks with orthogonal least square (OLS) for center learning of RBFs and recently Adaptive Network-Based Fuzzy Inference System (ANFIS) are widely used for nonlinear system identification and also for prediction of chaotic time series. In this section we utilize locally linear neurofuzzy models with Lo-

---

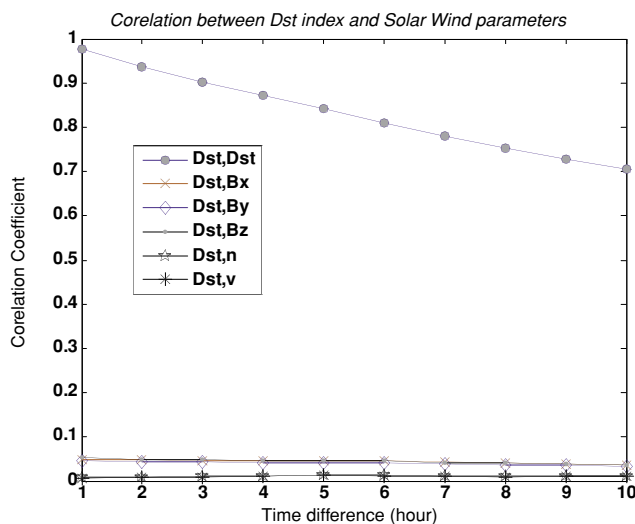[1]Although by increasing the number of neurons, only the membership

function of LLM that is split is change but the validity functions that are the normalized version of membership functions (Eqs. (9), (10)), then for each validity function all of other neurons membership functions have contribution, then all validity functions of neurons must update based on Eq. (9).

Table 1. Comparing LLNF with several neural methods for one-step prediction of Dst (Best Results).

Models	Learning Method	Neuron number	Epochs Number	Training set NMSE	Test set NMSE	Minimum Prediction Percentage
MLP	BP	63	120	0.0058	0.0336	83.67
RBF	OLS	20	24	0.0045	0.0286	97.76
ANFIS	LS + BP	16 Fuzzy rule	30	0.0051	0.031	92.41
LLNF	LOLIMOT	2	Epoch = Neuron	0.0044	0.0294	98.81

Table 2. Comparison of LLNF model with ENN for prediction of some storms during the year 1998 (error percentage for minimum Dst value prediction).

Method	Feb 10–13	Mar 9–12	May 2–5 May(2,4)	May 29 to June 1	Jul 28–31	Aug 5–8	Sep 24–27	Oct 16–19
ENN	1.85%	22%	22%, 21%	1.42%	7.6%	6%	18%	48%
LLNF	1.46%	8.08%	7.35%, 1.69%	0.74%	5.08%	4.35%	6.91%	13.88%

Fig. 2. Correlation coefficient between Dst and solar wind parameters.

Fig. 3. One-step prediction of some storms in year 1998 with LOLIMOT: observed Dst (solid lines) and one-step predicted values (dashed lines).

cally Linear Model Tree (LOLIMOT) learning algorithm for one-step ahead prediction of Dst index and compare the results with mentioned previous methods. We selected 11000 hours of Dst data for training and used the next 300 hours for test set; the test set interval contains the strongest geomagnetic storm between the years 1999 and 2000 with minimum value of -301 nT. Table 1 shows the result for best fit for one hour prediction of Dst time series for some neural and neurofuzzy methods and our method. For all cases, input to black-box model is $[\,Dst(t-1)\; Dst(t-2)\; Dst(t-3)\; Dst(t-4)\,]$ to predict $Dst(t)$. The NMSE is defined as follows:

$$\text{NMSE} = \sum_{i=1}^{N}(y(i) - \hat{y}(i))^2 / \sum_{i=1}^{N} y^2(i) \qquad (15)$$

where N is the number of test data and y is observed value and $\hat{y}$ is predicted value. In Table 1 Error percentage for prediction of Dst minimum on test set is defined as follows:

$$100 - 100 * (\min(Dst) - \text{Predicted } Dst \text{ minimum})/\min(Dst).$$

Also in (Watanabe *et al.*, 2002) used an Elman Neural Network (ENN) with Dst and solar wind parameters: speed (V), solar wind density (n), and the southward Interplanetary Magnetic Field (IMF) component (Bz) as inputs to ENN. The ENN model is trained with solar wind OMNI data from NASA's National Space Science Data Center (NSSDC) for the period from 1978–1982, which offers relatively continuous solar wind data for prediction of some storms over the year 1998 from January to September. In Fig. 2, correlation coefficient between Dst index and other solar wind parameters is plotted and shows poor correlation between Dst and solar wind data. Our method used only $Dst(t-1)$, $Dst(t-2)$, $Dst(t-3)$, $Dst(t-4)$ index as input for prediction of next Dst value. Table 2 shows the comparison of LLNF with ENN in the sense of error percentage for prediction of Dst minimum for some storms. In Fig. 3 observed Dst and predicted Dst with LLNF with LOLIMOT learning from February to October is depicted and in Fig. 4 the measured value versus predicted values for these intervals is depicted. The storm of October 16–19 is not more extreme than other storms of this year but prediction of this storm has much more error (13.88%) this
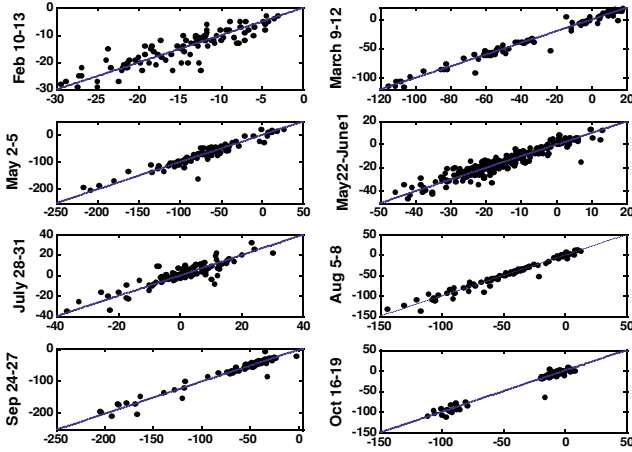
Fig. 4. Predicted values (horizontal axis) versus observed values (vertical axis) for storms in year 1998.

Table 3. Comparison of LLNF with some previous studies for prediction of storm during 2 May 1998.

Models	RMSE	Correlation
Sharifi	6.3	0.94
Lund *Dst* model	10.3	0.88
O'Brian and McPherron	12.3	0.83
Fenrich and Luhmann	15.3	0.78
Burton	16.4	0.76

is because of lack of data distribution between -20 nT and -70 nT as shown in Fig. 4, as a result each model would have bad interpolation. Also the correlation coefficient for these eight time interval of storms is as following:

{ 0.8879 0.9880 0.9517 0.9419 0.8817 0.9863 0.9758 0.9809 }.

This is much better than previous studies.

In Table 3 our method (LOLIMOT) is compared with several previous studies for predicting the storm in 2 May 1998 root mean square error in this table is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}. \quad (16)$$

Where N is the number of data points. Correlation coefficient value (0.94) in Table 3 differs from that reported on above (0.9517), because this value (0.94) is only for 2 May 1998 not for 2–4 May.

Many modeling methods (Vassiliadis *et al.*, 1999) relates *Dst* index to *Bz* negative part (*Bs*) and used as an important to system. *Bs* is defined as follows:

$$B_S = \begin{cases} 0; & B_z \geq 0 \\ |B_z|; & B_z \leq 0 \end{cases}. \quad (17)$$

In Fig. 5, correlation coefficient between *Dst* and its lagged values and *Bs* lagged values is plotted as a function of time for 2001 dataset. It is obvious from this figure that $Dst(t)$ is more correlated with $\{Dst(t-1), Dst(t-2), Dst(t-3), Dst(t-10), Dst(t-11), Dst(t-12), Dst(t-13)\}$ and also $\{Bs(t-1), Bs(t-2), Bs(t-3), Bs(t-4), Bs(t-$
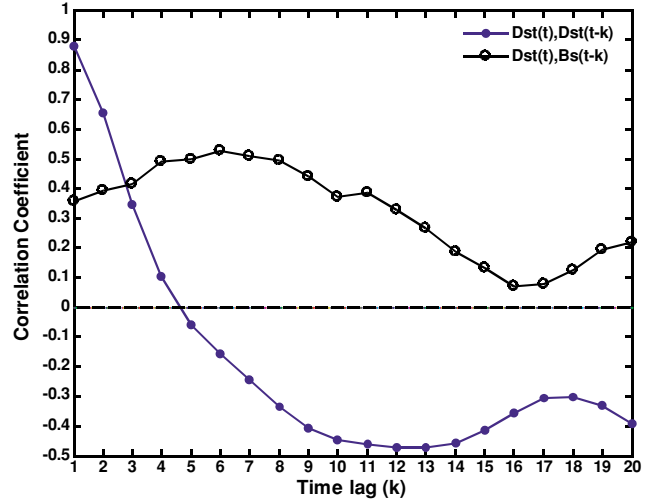

Fig. 5. Autocorrelation between $Dst(t)$, $Dst(t-k)$ and $Dst(t)$, $Bs(t-k)$; ($k = 1, \ldots, 20$) for 2001 dataset.
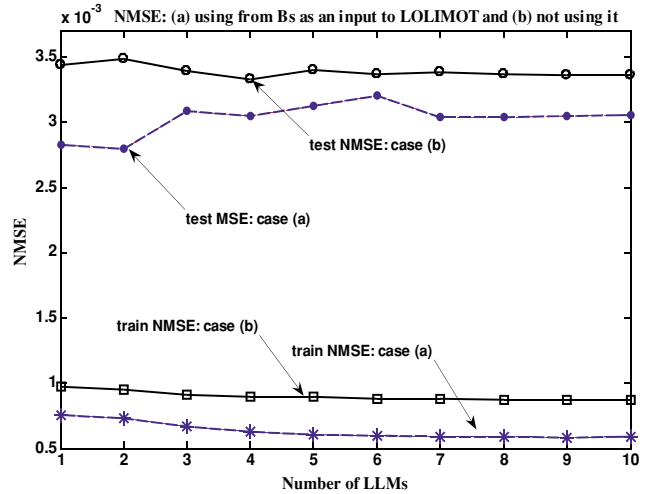

Fig. 6. NMSE for train data and test data for two cases: (a) using past values of Bs and *Dst* as inputs to LOLIMOT and (b) only using past values of *Dst*; training set contains data in year 2001 from first of year prior to 24 September 2001 and future 13 days for testing. Appropriate number of LLMs for case (a, b) is (2, 4); NMSE for test set is (0.0028, 0.0033) and correlation coefficient for test set is (0.978, 0.982) respectively. (Results is for one-step prediction).

5), $Bs(t-6)$, $Bs(t-7)\}$, we use these parameters as input to model (LOLIMOT) to predict $Dst(t)$. Data in year 2001 from first prior to 24 September is used as training to LOLIMOT and next 13 days for testing. In Fig. 6, NMSE for training data and test data for both cases is plotted as a function of LLMs that shows by using Bs as an external input to model, NMSE decreases for training and test data. Also many researchers have shown that *Dst* derivative $(d(Dst)/dt)$ is correlated with $V * Bs$ and also *Dst* is correlated with dynamic pressure. But in this study, we used only *Bs* as an external input to LOLIMOT model. In Fig. 7 the result of prediction of storm in 3–4 October 2001 is depicted for two cases; the first one employed only past *Dst* values and the second one employed past *Dst* values and also past *Bs* values as an external input to model. By using *Bs* as an external input, NMSE decreased from 0.0033 to
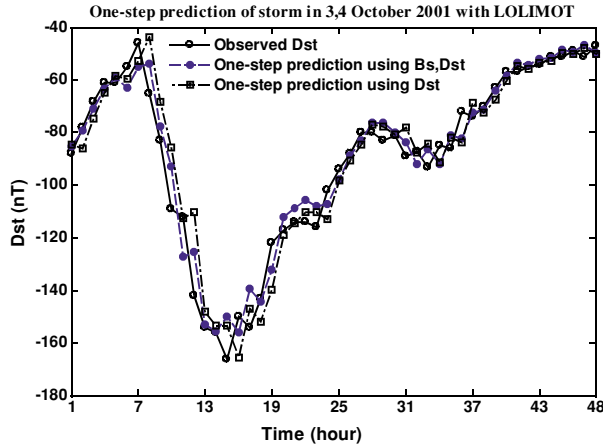
Fig. 7. One-step prediction of storm in 3, 4 October 2001 with LOLIMOT for both cases: using Bs, Dst past values and using only Dst past values. Training data are as the same with previous figure.
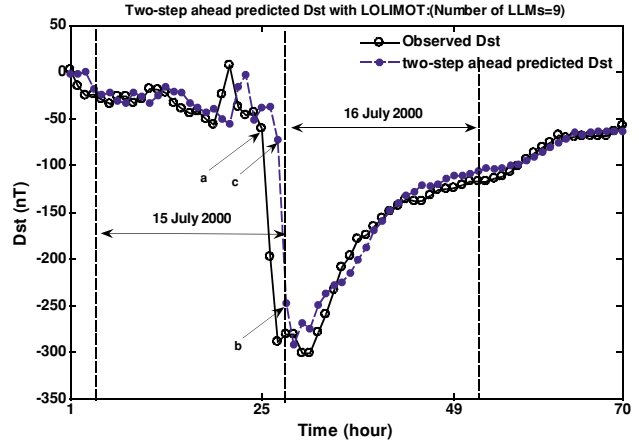

Fig. 9. Two-step ahead prediction of extreme storm of 15–16 July 2000 with LOLIMOT with 9 LLMs: the correlation coefficient of this prediction is 0.94 and the prediction of main phase of storm has two hours delay.
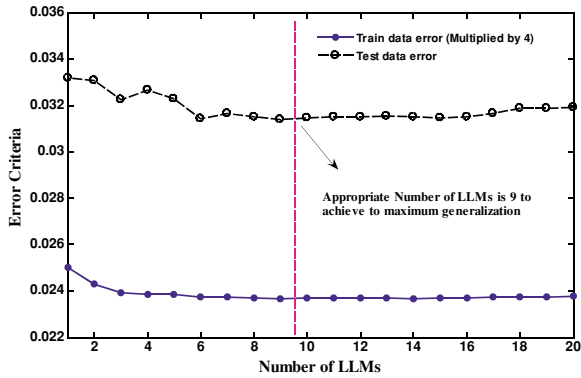

Fig. 8. NMSE versus the number of LLMs for train data (first 11000 hourly sample data from first of year 2000) and test data for two hour ahead prediction of storms of 14–16 July 2000, we select the number of LLMs where test error is approximately smooth and after it error increase with increasing the number of LLMs and in this case is 9 LLMs.
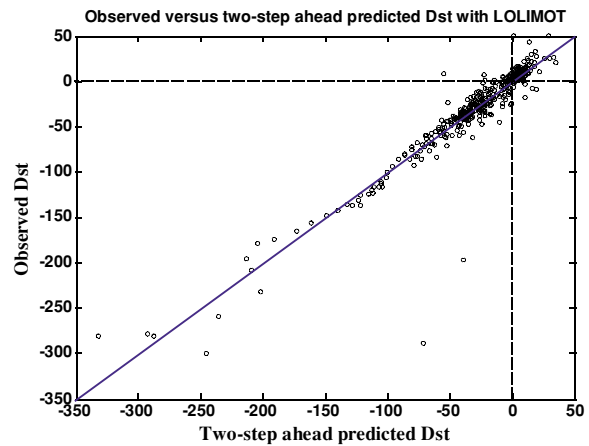
Fig. 10. Observed Dst versus two-step ahead predicted Dst of 14–16 July 2000 with LOLIMOT: correlation coefficient is 0.94.

0.0028 and correlation coefficient increased from 0.978 to 0.982; also the number of LLMs decreased from 4 to 2.

5.   Multi-Step Prediction of *Dst* Index

In this section, both the LOLIMOT and LOLIMOT+SSA method are applied for multi-step prediction of Dst index. By LOLIMOT method, input vector to Locally Linear Neurofuzzy Model is: $[Dst(t-2) \; Dst(t-3) \; Dst(t-4) \; Dst(t-5)]$ to predict. In Fig. 8 error criteria both for training data and test data is plotted as a function of LLMs. appropriate number of LLMs to achieve to maximum generalization, are 9. the result of two hour ahead prediction of geomagnetic storm with LOLIMOT method with 9 LLMs is depicted in Fig. 9, the Observed Dst values versus two-step ahead predicted Dst is depict in Fig. 10. Based on Fig. 9, neither seems by increasing the prediction horizon, even the LLNF model with a sophisticated training dataset (11000 hourly data of Dst index containing several big storms) nor can predict very good the storm in testing dataset and nor learn the dynamic of system in decay time but can learn the recovery phase of the storm.   In the remaining part of this section, in order to enhance the accuracy for multi-step

ahead prediction of Dst index, a combination of SSA and LLNF is used as depicted in Fig. 11. Eigenvalues of covariance matrix of Dst time series are depicted in Fig. 12 in logarithmic scale. It can be seen from this spectrum that the first fifty components have more significance and other components have small effect in prediction and furthermore can decrease the statistical reliability of the analysis. (Using all 150 components or 50 first ones have very similar results). Using Eq. (5), 50 PC's are constructed and then each PC constitutes a new series for the modeling of which a different LLNF model must be used. Thus for modeling $PC_k(t)$ the input vector of k'th LLNF ($LLNF_k$) is as follows:

$$[\, PC_k(t-2) \; PC_k(t-3) \; PC_k(t-4) \; PC_k(t-5) \,].$$

In order to achieve maximum generalization for each of the fifty components, as in Fig. 8, error criteria for each PCs must be used for selecting the appropriate number of LLMs for each PC. Number of LLMs for each PC for prediction of extreme storm of 14–16 July 2000 is as follows:

$$M = \{1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 2, 2, 1, 1, 3, 1, 1, 1, 1,$$
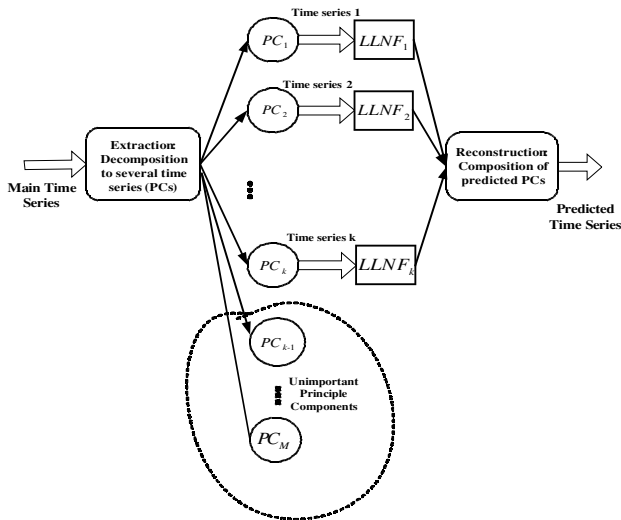$$2, 2, 2, 1, 1, 1, 3, 1, 2, 2, 1, 1, 3, 2, 1, 1, 1, 2, 4, 1,$$

Fig. 11. Block Diagram of SSA + LOLIMOT method for time series prediction: *M* principle component extracted and then for each PC a LLNF model should train then next value prediction of each PC obtained; finally predicted PCs combined for achievement to prediction of main series (Gholipour *et al.*, 2005).
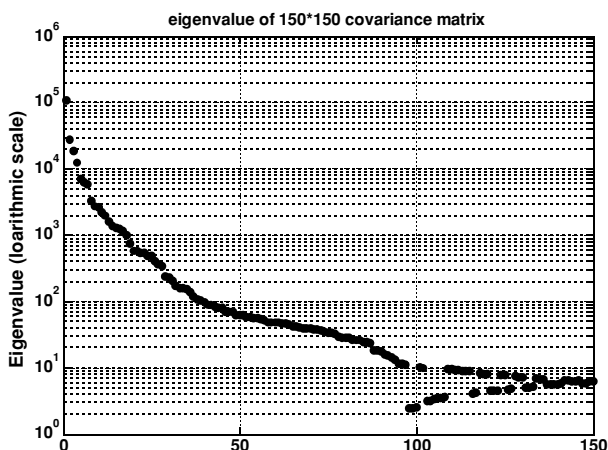

Fig. 12. Singular Spectrum of *Dst* training data in year 2000 with a 150×150 covariance matrix: (singular values is plotted in logarithmic scale).
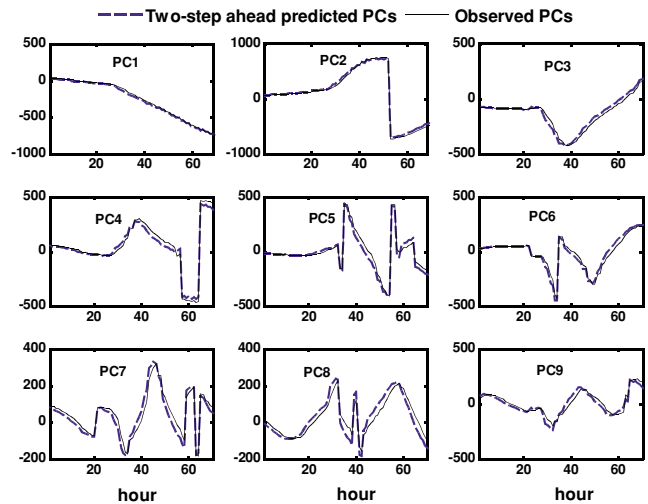

Fig. 13. Two-step ahead prediction of each principle component of *Dst* index in 14–16 July 2000 with LOLIMOT: (only nine PCs is shown).
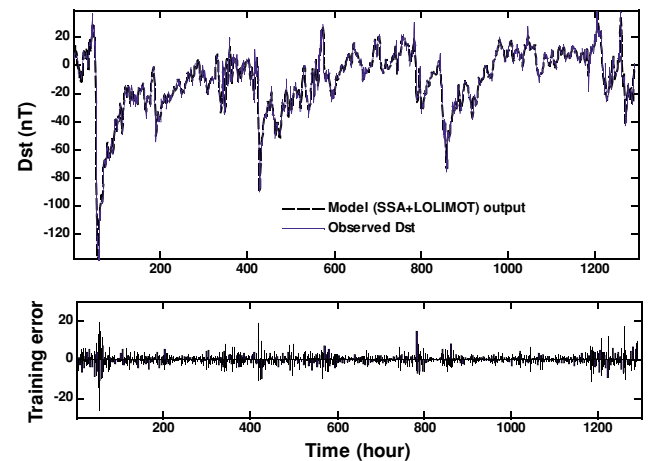

Fig. 14. Two-step ahead modeling of train data (about 1300 data points in year 2000 prior to event 14–16 July 2000) with LOLIMOT + SSA.

1, 1, 2, 1, 1, 2, 1, 1, 1, 1}.

Then for each PC this number of LLMs is trained for use in LLNF models for multi-step prediction of that PC. In Fig. 13 the result of two-step ahead prediction of PCs is depicted. Finally using from Eqs. (6), (7) to construct the Multi-step prediction of main *Dst* time series from predicted PCs that is let after training PCs. The result of combination of PCs both for training set and prediction is shown in Figs. 14 and 15, respectively. Comparing Fig. 15 with Fig. 9 shows that two-step ahead prediction of main phase of storm has one-hour delay with SSA + LOLIMOT against two-hour delay with LOLIMOT and many other methods. Some physical models (Nagatsuma, 2002) have achieved to correlation coefficient of 0.9 which is reported for one step prediction of *Dst* index, but our method correlation coefficient is 0.94 (LOLIMOT method) and 0.98 (LOLIMOT + SSA method) for one-step prediction. For

most of PCs the number of LLMs is only one, but why we use nonlinear process (LOLIMOT) for modeling? reply is that with the aim of using LOLIMOT for each PC we only could found that some PC for modeling only need one global linear model and also some PC need more that one line or in the other hand need locally linear model, then it's reasonable to doing nonlinear process for each PC to obtain best locally linear number for modeling to achieve maximum generalization.

Also most of the black-box models would predict the start and peak of storms with delay and with increasing the prediction horizon the delays increase, then its useful to achieve a method for prediction without delays for warning when the storm start. Our start-of-the art method is combination of SSA and LLNF is used as depicted in Fig. 11.

With SSA + LOLIMOT modeling method first SSA method with $M = 150$ in Eq. (1) is applied. Then a 150×150 covariance matrix is constructed based on Eq. (2) and eigenvalues of covariance matrix with logarithmic scale is plotted if Fig. 12 and approximately most important eigenvalues are first coming 50. Using from Eq. (5) 50 PCs
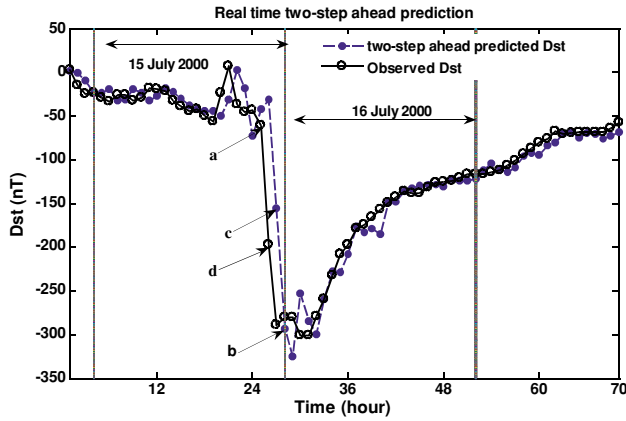
Fig. 15. Two-step ahead prediction of *Dst* index (15–16 July 2000) with LOLIMOT + SSA: the method predicted the start and decay phase of extreme storm with 1 hour delay from two hour ahead. When we are at time (a) in figure then the method must predict the point (b) that is near the event of two-next hour of point (a) but it predict the point (c) that is near the actual event of one-next hour (point d). Note that train data is shown in Fig. 14.
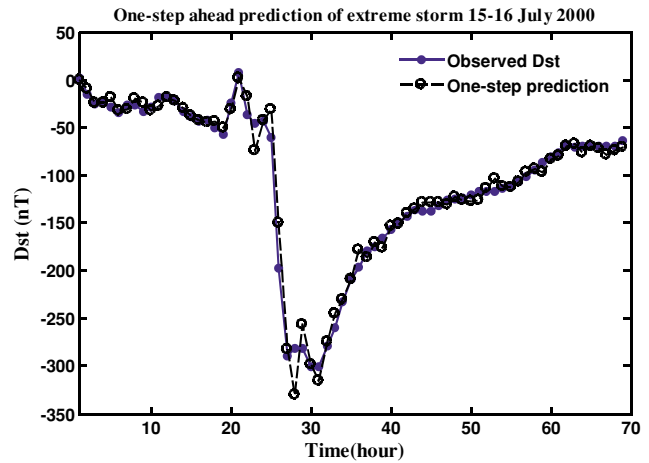

Fig. 17. One step ahead predicted *Dst* (shift back one-step the two-step ahead predicted *Dst* with LOLIMOT + SSA); this prediction is perfect and without delay for predicting the start and the main phase and also for prediction of recovery phase of the storm and the correlation of this prediction is 0.98.
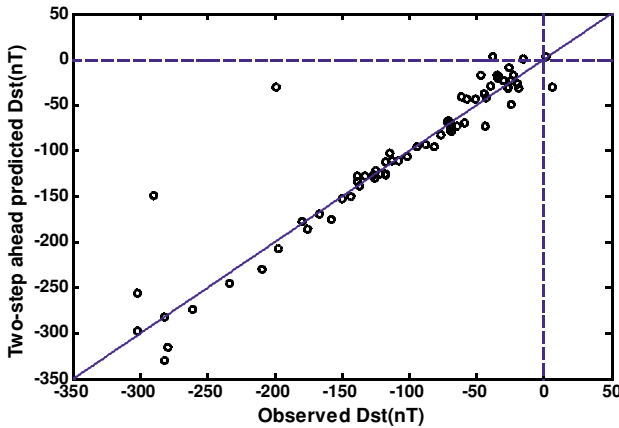

Fig. 16. Observed *Dst* versus two-step ahead predicted *Dst* (LOLIMOT + SSA) for prediction of storm 16 July 2000 (data from 15–18 July 2000): the correlation coefficient is 0.93.
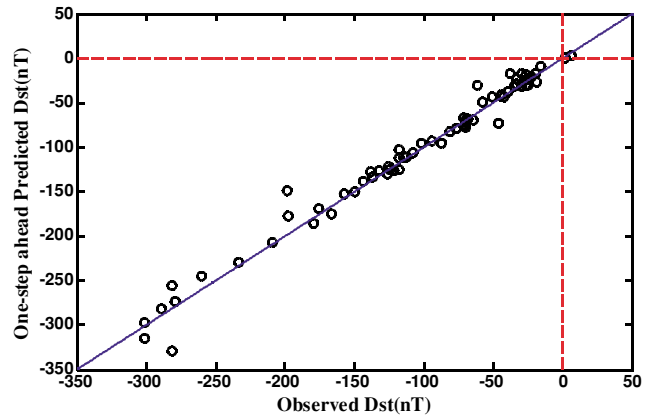

Fig. 18. Observed *Dst* versus one-step ahead predicted *Dst* (LOLIMOT + SSA) for prediction of storm 16 July 2000 (data from 14–16 July 2000): the correlation coefficient is 0.98.

is constructed and then each PC is a new series that must use for modeling of each, a LLNF model. For modeling of each PC the input vector to $LLNF_k$ for modeling of $PC_k(t)$ is the following inputs:

$$[\, PC_k(t-2)\ \ PC_k(t-3)\ \ PC_k(t-4)\ \ PC_k(t-5)\,].$$

An interesting result of using SSA + LOLIMOT is that for multi-step ahead prediction, it decrease one-step from the delay of prediction (especially the prediction of main phase of storm) that many methods could not, then if we predict the storm two step ahead many models predict the start and the main phase of storm with two hour delay, but our method for two-step ahead prediction have one-step delay (although for one-step prediction also our method have one-step delay for prediction of main phase of storm and from two-step ahead the recovery phase predict very good) and then the result of two next hour prediction of *Dst* index, is a point proximity to event that occur one next hour and not two next hour. Then we predict two-step ahead the storm and let it as the next hour event. The result of

such innovation is depicted in Fig. 17. Also in Fig. 18, Observed *Dst* versus one-step predicted of storm of 15–16 July 2000 is depicted that has much more compact error bar in comparison with Fig. 16.

The extreme geomagnetic storm in 14 March 1989 lead to Black-Out in Quebec and North America and as a result damage many expensive transformers and black-out for about 12 hours. Until now many methods could not predict such extreme storm to warn the human technical systems such as power grids, satellites, etc. our start of the art method (SSA + LOLIMOT) is applied here for two-hour ahead prediction of this extreme storm, the result of prediction is illustrated in Fig. 19, if analyze this figure, we found that the two-step ahead prediction (especially the start time and main phase of storm) only have one-step delay and some one heuristically point of view could assume the two-step ahead prediction of this storm as the event of next hour, the result of such insight is depicted in Fig. 20 that shows the benefices of such point of view for precise one-step prediction of this extreme storm for alerting the techni-
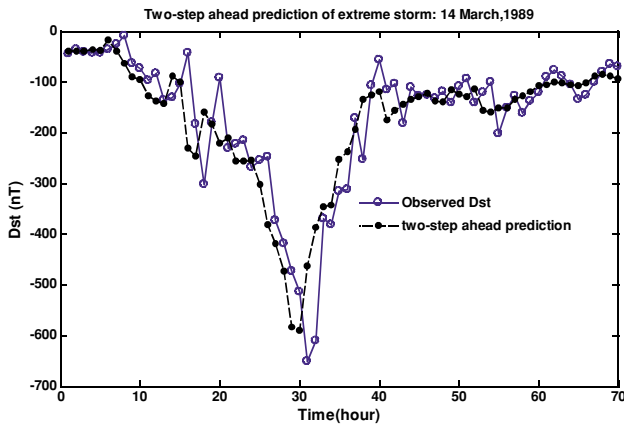
Fig. 19. Two step ahead prediction of extreme storm of 14 March 1989 that cause power grid black-out in Québec and US. (LOLIMOT + SSA method for prediction).
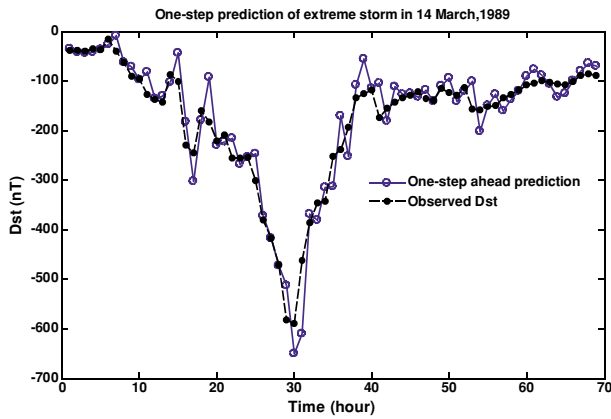

Fig. 20. One-step ahead prediction of extreme storm of 14 March 1989 by SSA + LOLIMOT: (this prediction is obtained by shifting one-step back the two-step ahead prediction).
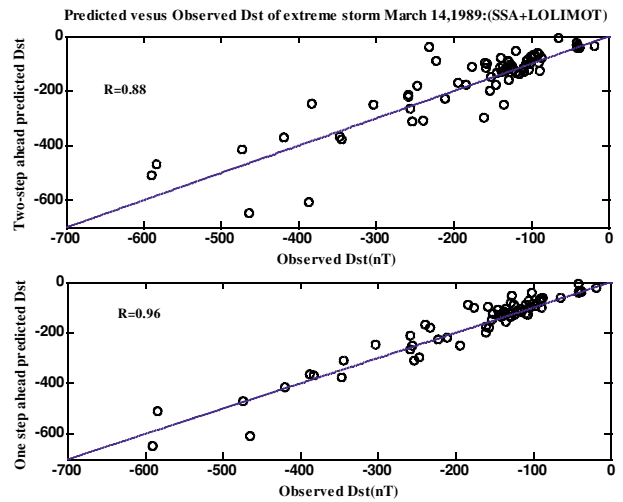

Fig. 21. Predicted versus observed *Dst* index of 14–16 March 1989 with correlation coefficient 0.88 and 0.96 for two step ahead and one step ahead, respectively.
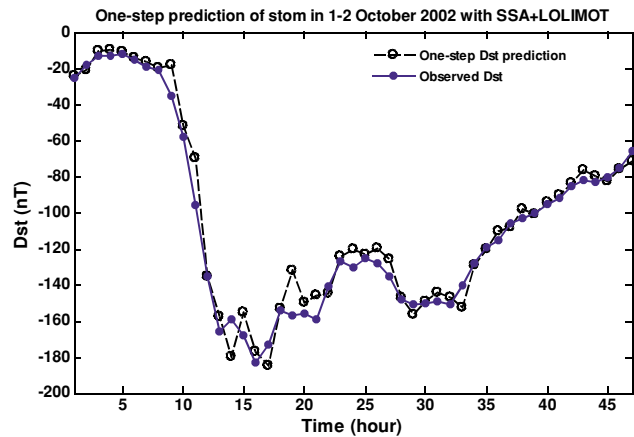

Fig. 22. Two-step ahead prediction of storm in 1, 2 October 2002 with SSA + LOLIMOT: data from 1 July 2002 prior to this event is used as training. (Note that this prediction obtained from one-step shifting back the two-step ahead prediction). Correlation coefficient is 0.98.

cal systems. Two-step and one-step ahead *Dst* prediction is depicted in Figs. 19 and 20 respectively. The correlation coefficient of two-step and one-step prediction with this novel insight is 0.88, 0.95 respectively. Two-step and one-step ahead predicted versus observed *Dst* values is plotted in Fig. 21. Also this method is used for prediction of storm in 1–2 October 2002 and the result of prediction is depicted in Fig. 22. The correlation coefficient of 0.98 is obtained for prediction of this storm with SSA + LOLIMOT method.

6. Conclusion

In this article we used a powerful combination of tools composed of singular spectrum analysis, and neuro-fuzzy locally linear network with model tree learning (LOLIMOT), for multi-step prediction of the *Dst* index. The method has also shown its effectiveness in the long term prediction of the sunspot number time series (Gholipour *et al.*, 2005). The use of singular spectral analysis results in the identification and separation of the main components with different periodicities. LOLIMOT, on the other hand, allows the capture of the dynamic behaviors of the different components and provides an excellent tool for nonlinear prediction of their future values. Taken together,

the methodology becomes a most powerful means for reliable multi-step and long term prediction of space weather indices without the need to assume a physical model. This is important for designing an alert system in order to protect the satellites that are not protected by the ionospheric and magnetospheric shields, as well as protecting high altitude flights, power and communication lines, etc. The tools can also be utilized for long term and black box prediction of other naturally occurring phenomena the physical properties of which are not well understood.

In assessing the improvement in predictability of the time series made possible by the SSA methodology, it should, in all fairness, be pointed out that SSA uses much more than a small number of data. So the prediction can no longer be claimed to have been achieved via lagged values including data for only a few hours. While more lagged values quickly lose usefulness in regression analysis, neural network models, LOLIMOT, and similar prediction methodologies, spectral methods like SSA are inherently using past

data for periods much longer than cyclicities in the time series. Although the training of the prediction tools like LOLIMOT use even more data than SSA. In fact, the utilization of SSA indirectly amounts to using external information about the behavior of the time series in the previous cycles as well. Therein lies the secret of the success in longer term prediction (Loskutov *et al.*, 2001a, b; Gholipour *et al.*, 2005). We seem to be doing the impossible by predicting multi-steps ahead by using only a few hours' past data; but we are, in fact, indirectly using other data on periodicities of the time series obtained through consideration of much longer term past values that could not have been directly used in prediction tools.

Another indirect advantage of this effort is the fact that the analysis of periodicities carried out by the SSA method paves the way for further physical investigations on the nature of *Dst*. These empirical data can possibly be a new source for improving our physical understanding of the currents represented by *Dst* (Campbell, 1996; Temerin and Li, 2002). Our concern, however, has been to show the possibility of predicting geomagnetic activity indices for periods longer than previously thought possible.

It is also very instructive, especially when SSA is not used and ones tries to predict the future values of the geomagnetic activity only by using past values of the same time series, to focus our attention to important events, rather than report general statistical properties of the prediction. It is during the storm conditions, that one can appreciate the pessimism of physicists with respect to the predictability of *Dst* given its past values. Indeed, our simulation results demonstrate the validity of that concern. Although LOLIMOT is very successful in the general prediction at least in the short term, it shows obvious lags when storms occur. Since no factor other than the past values of *Dst* is used as an input to the predictor, LOLIMOT has no way of knowing that a storm is about to begin until it sees its effect in past *Dst* values. But this means that it can only predict with considerable lag exactly when good and rapid prediction is important. This in no way belittles the achievement gained by using an excellent predictor. We can, of course, try even stronger predictors; for example, locally quadratic instead of locally linear. Although our group has reported the possibility of improving predictions by this method in other contexts, we did not achieve any significant improvement in the case under study. But the important thing is that even if future research finds much bettor predictors, neither of the tools can do the impossible. When one attempts to predict *Dst* by only using its past values, it cannot know that some other factor will cause a dramatic change until and unless that factor is somehow correlated with the past vales of *Dst* that are being used as predictors. We believe we have achieved an unbelievably successful predictive capability by the methodology discussed in this study. There are two possibilities for improving this result. One is to find a still more powerful predictor. But we have already mentioned that we already have achieved almost the best that is practically possible. The other possibility is to use more input signals. Our method has achieved predictive powers that are already better than those achieved by other researchers using physical data other than past *Dst* values. What if both past *Dst* values and other external factors are used as predictors? Our initial work indicates that other external factors, while having good predictive capability when used by themselves, cannot add to the predictive capability of the factors already used. Since highly nonlinear methodologies are being used, partial correlation analysis cannot be conclusive. The result obviously depends on what external signal is being used and in what way. So our negative answer cannot be the last word. Recent physical studies show much higher correlation between *Dst* and the negative part of Bz or the derivative of Dst ($d(Dst)/dt$) (Temerin and Li, 2002; Campbell, 1996). However, as already has been pointed out, it is not the predictive power of those factors in linear regression that is in question, but their added predictive capability to an already excellent prediction in a highly nonlinear fashion is what is really required. This calls for much research in the future. But we have actually found a third way of improved prediction. By using SSA we have, as explained in the previous paragraph, in effect been using previous cycles of the time series as external factors. The results have not only allowed much longer term prediction, but also better performance in crucial periods of solar storms, when LOLIMOT by itself did not have satisfactory performance.

References

Baker, D. N., Statistical analysis in the study of solar wind magnetosphere coupling, in *Solar Wind-Magnetosphere Coupling*, edited by Y. Kamide and J. A. Slavin, p. 17, Terra Sci., Tokyo, 1986.

Burton, R. K., R. L. McPherron, and C. T. Russell, An empirical relationship between interplanetary conditions and Dst, *J. Geophys. Res.*, **80**, 4204–4214, 1975.

Campbell, W., Geomagnetic storms, the Dst ring-current myth and lognormal distributions, *J. Aymospheric and Solar Terrestrial Physics*, **58**(10), 1171–1187, July 1996.

Detman, T. R. and D. Vassiliadis, Review of techniques for magnetic storm forecasting, in *Magnetic Storms, Geophys. Monogr. Ser.*, vol. 98, edited by B. T. Tsurutani, W. D. Gonzalez, Y. Kamide, and J. K. Arballo, p. 253, AGU, Washington D.C., 1997.

Fenrich, F. R. and J. G. Luhmann, Geomagnetic response to magnetic clouds of different polarity, *Geophys. Res. Lett.*, **25**, 2999–3002, 1998.

Freeman, J., A. Nagai, P. Reiff, W. Denig, S. Gussenhoven Shea, M. Heinermann, F. Rich, and M. Hairston, The use of neural networks to predict magnetospheric parameters for input to a magnetospheric forecast model, in *Artificial Intelligence Applications in Solar Terrestrial Physics*, edited by J. Joselyn, H. Lundstedt, and Trollinger, 167, Natl. Oceanic and Atmos. Admin., Boulder, Colorado, 1994.

Gleisner, H., H. Lundstedt, and P. Wintoft, Predicting geomagnetic storms from solar-wind data using time-delay neural networks, *Ann. Geophys.*, **14**, 679–686, 1996.

Gholipour, A., C. Lucas, M. Shafiee, and B. N. Araabi, Extracting the main patterns of natural time series for long term prediction, *J. Aymospheric and Solar Terrestrial Physics*, **67**(6), 595–603, 2005.

Iyemori, T., H. Maeda, and T. Kamei, Impulse response of geomagnetic indices to interplanetary magnetic fields, *J. Geomag. Geoelectr.*, **31**(1), 1979.

Joselyn, J. A., Geomagnetic activity forecasting: The state of the art, *Rev. Geophys.*, **33**, 383, 1995.

Kamide, Y., W. Baumjohann, I. A. Daglis, W. D. Gonzalez, M. Grande, J. A. Joselyn, R. L. McPherron, J. L. Phillips, E. G. D. Reeves, G. Ros-

toker, A. S. Sharma, H. J. Singer, B. T. Tsurutani, and V. M. Vasyliu-nas, Current understanding of magnetic storms: Storm-substorm relationships, *J. Geophys. Res.*, **103**, 17705–17728, 1998.

Kugblenu, S., S. Taguchi, and T. Okuzawa, Prediction of the geomagnetic storm associated Dst index using an artificial neural network algorithm, *Earth Planets Space*, **51**, 307–313, 1999.

Loskutov, A., I. A. Istomin, K. M. Kuzanyan, and O. L. Kotlyarov, Testing and forecasting the time series of the solar activity by singular spectrum analysis, *Nonlin. Phenomena in Complex Syst.*, **4**(1), 47–57, 2001a.

Loskutov, A., I. Istomin, O. Kotlyarov, and K. Kuzanyan, A study of the regularities in Solar magnetic activity by singular spectrum analysis, *Astronomy Letters*, **27**(11), 745–753, 2001b.

Loskutov, A., I. Istomin, and O. Kotlyarov, Data analysis: generalizations of the local approximation method by singular spectrum analysis, http://xxx.lanl.gov/abs/nlin.cd/0109022.

Munsami, V., Determination of the effects of substorms on the storm-time ring current using neural networks, *J. Geophys. Res.*, **105**, 27833, 2000.

Nagatsuma, T., Geomagnetic Storms, *Journal of the Communications Research Laboratory*, **49**(3), 2002.

Nelles, O., Nonlinear System Identification with Local Linear Neuro-Fuzzy Models, PhD Thesis, TU Darmstadt, Shaker Verlag, Aachen, Germany, 1999.

Nelles, O., Nonlinear system identification, Springer Verlag, Berlin, 2001.

O'Brien, T. P. and R. L. McPherron, An empirical phase space analysis of ring current dynamics: solar wind control of injection and decay, *J. Geophys. Res.*, **105**, 7707–7719, 2000a.

O'Brien, T. P. and R. L. McPherron, Forecasting the ring current index Dst in real time, *J. Atmospheric and Solar-Terrestrial Physics*, **62**, 1295–1299, 2000b.

Temerin, M. and X. Li, A New Model for the Prediction of Dst on the Basis of the Solar Wind, *J. Geophs. Res.*, **107**(A12), 1472, doi:10.1029/2001JA007532, 2002.

Vassiliadis, D., A. J. Klimas, D. N. Baker, and D. A. Roberts, A description of the solar wind magnetosphere coupling based on nonlinear prediction filters, *J. Geophys. Res.*, **100**, 3495, 1995.

Vassiliadis, D., A. J. Klimas, and D. N. Baker, Models of Dst Geomagnetic Activity and of its Coupling to Solar Wind Parameters, *Phys. Chem. Earth (C)*, **24**(1-3), 107-I 12, 1999.

Vautard, R., P. Yiou, and M. Ghil, Singular spectrum analysis: A toolkit for short noisy chaotic signals, *Physica D*, **58**, 95–126, 1992.

Watanabe, S., E. Sagawa, K. Ohtaka, and H. Shimazu, Prediction of the Dst index from solar wind parameters by a neural network method, *J. Communications Research Laboratory*, **49**(4), 2002.

Wintoft, P., Space weather physics—Prediction and classification of solar wind structures and geomagnetic activity using artificial neural networks, PhD Thesis, LUNFD6/(NFAS 1017)/1-86/(1997), 1997.

Wintoft, P. and H. Lundstedt, Identification of geoeffective solar wind structures with self-organized maps, AI Applications in Solar-Terrestrial Physics, Lund, Sweden, July 29–31, 1997, edited by I. Sandahl and E. Jonsson, ESA WPP-148, 151–157, 1998.

Wu, J.-G., H. Lundstedt, P. Wintoft, and T. R. Detman, Space weather forecasting on the 1997 January halo CME event using neural network models, AI Applications in Solar-Terrestrial Physics, Lund, Sweden, July 29–31, 1997, edited by I. Sandahl and E. Jonsson, ESA WPP-148, 145–150, 1998a.

Wu, J.-G., H. Lundstedt, P. Wintoft, and T. R. Detman, Neural network models predicting the magnetospheric response to the 1997 January halo-CME event, *Geophys. Res. Lett.*, **25**, 3,031–3,034, 1998b.

J. Sharifi (e-mail: j.sharifie@ece.ut.ac.ir), B. N. Araabi (e-mail: araabi@ut.ac.ir), and C. Lucas (e-mail: lucas@ipm.ir)