Research in Astron. Astrophys. **2009** Vol. **9** No. **6**, 694 – 702
http://www.raa-journal.org *http://www.iop.org/journals/raa*

***R**esearch in*
***A**stronomy and*
***A**strophysics*

Forecast daily indices of solar activity, F10.7, using support vector regression method

Cong Huang[1], Dan-Dan Liu[1] and Jing-Song Wang[1]

National Center For Space Weather, CMA, Beijing 100081, China; *huangc@cma.gov.cn; liudd@cma.gov.cn*

Received 2008 September 12; accepted 2009 March 9

Abstract The 10.7 cm solar radio flux (F10.7), the value of the solar radio emission flux density at a wavelength of 10.7 cm, is a useful index of solar activity as a proxy for solar extreme ultraviolet radiation. It is meaningful and important to predict F10.7 values accurately for both long-term (months-years) and short-term (days) forecasting, which are often used as inputs in space weather models. This study applies a novel neural network technique, support vector regression (SVR), to forecasting daily values of F10.7. The aim of this study is to examine the feasibility of SVR in short-term F10.7 forecasting. The approach, based on SVR, reduces the dimension of feature space in the training process by using a kernel-based learning algorithm. Thus, the complexity of the calculation becomes lower and a small amount of training data will be sufficient. The time series of F10.7 from 2002 to 2006 are employed as the data sets. The performance of the approach is estimated by calculating the norm mean square error and mean absolute percentage error. It is shown that our approach can perform well by using fewer training data points than the traditional neural network.

Key words: methods: data analysis — Sun: activity — Sun: radio radiation

1 INTRODUCTION

The 10.7 cm solar radio flux (F10.7) is a measurement of the intensity of solar radio emissions with a wavelength of 10.7 cm (a frequency of 2800 MHz). F10.7 is a general indicator of solar magnetic activity, solar ultraviolet and X-ray emissions, and even solar irradiance. F10.7 is used for a wide range of applications including astronomy, climate modeling, geophysics, meteorology, communications, satellite systems and so on. As a proxy for solar extreme ultraviolet (EUV) radiation, F10.7 is always an input parameter for calculating the atmospheric density for orbit determination in aerospace and predicting the influence of the ionosphere on communication. For example, F10.7 is used as a control parameter of an ionospheric model in calculating the variations of radio-signal characteristics (Ortikov et al. 2003).

A number of studies have investigated the application of data analysis techniques to space weather prediction problems. Mordvinov (1986) developed a multiplicative auto regression model for forecasting monthly values of F10.7. Dmitriev et al. (1999) applied artificial neural networks to solar activity forecasting. Zhong et al. (2005) applied the singular spectrum analysis method in combination with the similar cycle method for predicting F10.7. Zhao & Han (2008) reconstructed the development history of F10.7 and put forward a long-term prediction method of solar 10.7 cm radio flux.

In 1995, Vapnik developed a neural network algorithm called support vector machine (SVM), which is an efficient tool for high-dimensional data processing. SVM is a novel learning machine based on statistical learning theory, and adheres to the principle of structural risk minimization seeking to minimize an upper bound of the generalization error, rather than minimize the training error (the principle followed by traditional neural networks). The first application of SVM to space weather problems is the reliable prediction of large-amplitude substorm events from solar wind and interplanetary magnetic field data (Gavrishchaka & Ganguli 2001). Li et al. (2008) applied SVM combined with K-nearest neighbors to forecasting solar flare and solar proton events. SVM has been extended to solve nonlinear regression estimation problems, such as new techniques known as Support Vector Regression (SVR), which have been shown to exhibit excellent performance (Vapnik, Golowich & Smola 1997).

In this study, we apply an SVR modeling algorithm to F10.7 time series data, and examine the feasibility of SVR in short-term F10.7 forecasting. A brief overview of the SVR is given in Section 2. The results of the application of an SVR-based approach to short-term F10.7 forecasting are presented in Section 3. In Section 4, we give discussion and conclusions.

2 SUPPORT VECTOR REGRESSION

Recently, SVR has emerged as an alternative and powerful technique to solve regression problems by introducing an alternative loss function. The SVR follows the principle of structural risk minimization, seeking to minimize an upper bound of the generalization error rather than minimize the prediction error on the training set (the principle of empirical risk minimization). Thus, the SVR has greater potential to generalize the input-output relationship learnt during its training phase in order to make good predictions for new input data. Here, a brief description of SVR is given. Detailed descriptions of SVR can be found in the papers of Vapnik (1995, 1998), Vapnik et al. (1997), Schölkopf & Smola (2002) and Cristianini & Shawe-Taylor (2000).

In SVR, the basic idea is to map the data x into a high dimensional feature space F via a nonlinear mapping ϕ and to do linear regression in this space.

$$f(x) = (\omega \cdot \phi(x)) + b \qquad \text{with} \quad \phi : R^n \rightarrow F, \omega \in F, \tag{1}$$

where b is a threshold. Thus, the linear regression in high dimensional feature space F corresponds to nonlinear regression in the low dimensional input space R^n. Correspondingly, the original optimization problem involving nonlinear regression is transformed into finding the flattest function in the feature space F, and not in the input space, x (see Fig. 1).

The regression approximation is to estimate a function $f(x)$ according to a given data set $T = \{(x_i, y_i)\}_i^n$, where x_i denotes the input vector; y_i denotes the output (target) value and n denotes the total number of data patterns. The aim is to identify the regression function $f(x)$ that accurately predicts the outputs $\{y_i\}$ corresponding to a new set of input-output examples, $\{(x_i, y_i)\}$. Since ϕ is fixed, we determine ω from the data by minimizing the sum of empirical risk R_{emf} and a complexity term $\frac{1}{2}\|\omega\|^2$, which enforces flatness in feature space.

$$R_{\rm SVR}(C) = R_{\rm emf} + \frac{1}{2}\|\omega\|^2 = C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(f(x_i) - y_i) + \frac{1}{2}\|\omega\|^2, \tag{2}$$

where $L_\varepsilon(.)$ is a cost function. For a large set of cost functions, Equation (2) can be minimized by solving a quadratic programming problem, which is uniquely solvable (Schölkopf & Smola 1997, 1998). It can be shown that the vector ω can be written in terms of the data points with α_i and α_i^* being the solution of the aforementioned quadratic programming problem (Vapnik 1995).

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i), \tag{3}$$

where α_i and α_i^* have an intuitive interpretation (see Fig. 2) as forces pushing and pulling the estimate $f(x_i)$ towards the observations y_i. Taking Equations (3) and (1) into account, the whole problem in terms of dot products can be rewritten in the low dimensional input space.

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\phi(x_i)\phi(x)) + b = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b. \tag{4}$$

We use an RBF kernel $K(x, y) = \exp(- \parallel x - y \parallel^2 /(2\sigma^2))$ as the kernel function in this paper.

For this special cost function, the Lagrange multiplier α_i and α_i^* fulfill the Karush-Kuhn-Tucker conditions (Vapnik 1995). They result in non-zero values after the optimization only if they are on or outside the boundary (see Fig. 2). The ε-insensitive cost function is given by

$$L_\varepsilon(f(x_i) - y_i) = \begin{cases} \mid f(x_i) - y_i \mid -\varepsilon & \mid y_i - f(x_i) \mid \geq \varepsilon. \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The respective quadratic programming problem is defined as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j) + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i + (\alpha_i^2 + \alpha_i^{*2}), \\ \text{subject to} & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*). \end{array} \tag{6}$$

So basically all patterns become support vectors. For the proper choice of b, the key idea is to pick those values α_k and α_k^* for which the prediction error $\delta_k = f(x_k) - y_k$ can be determined uniquely. In the ε-insensitive case, the prediction error is

$$\delta_k = \varepsilon \cdot \text{sign}(\alpha_k - \alpha_k^*). \tag{7}$$

In principle, one x_k would be sufficient to compute b, but for stability purposes, it is recommended to take an average over all the points on the margin with

$$b = \text{average}_k\{\delta_k + y_k - \sum_i(\alpha_i - \alpha_i^*)K(x_i, x_k)\}. \tag{8}$$

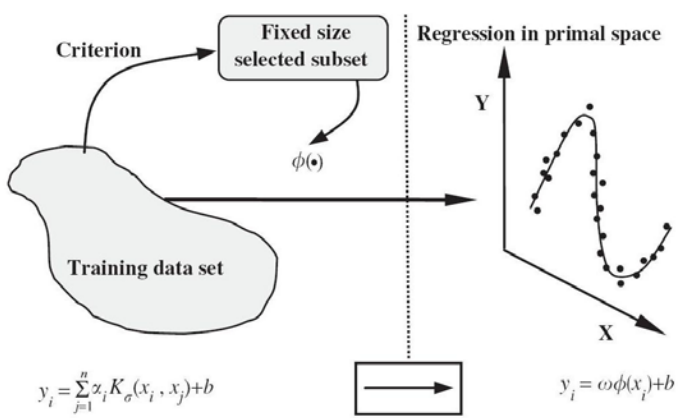

Fig. 1 Mapping input space x into a high-dimensional feature space (from Peng et al. 2004). Fixed size selected subset represents a given data set $T = \{(x_i, y_i)\}_i^n$. Primal space represents $f(x)$ which is the result of mapping the input data x into a high dimensional feature space F by nonlinear mapping ϕ.

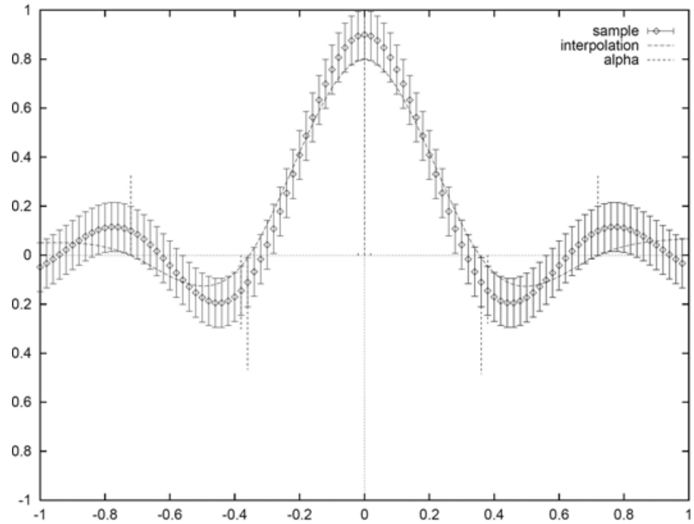

Fig. 2 Regression for the ε-insensitive case of the sinc function is the flattest with the ε tube around the data. α_i and α_i^* are drawn as positive and negative forces respectively. All points on the margin, where $f(x_i) - y_i = \varepsilon \cdot \mathrm{sign}(\alpha_i - \alpha_i^*)$, are used for the computation of b.

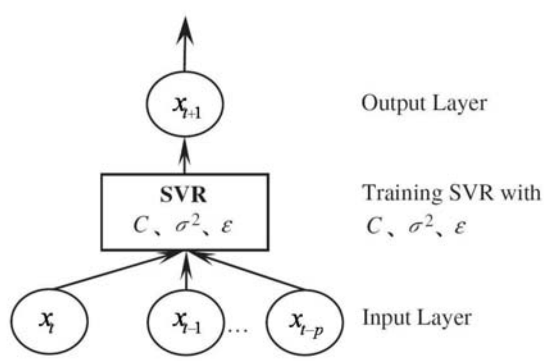

Fig. 3 Basic architecture of SVR model.

3 DATA ANALYSIS

For forecasting a univariate time series, the inputs of SVR are the past, lagged observations of the time series, and the outputs are the future values. Each set of input patterns is composed of moving a fixed-length window within the time series. Figure 3 shows the basic architecture of SVR. The mapping function can be described as below:

$$x_i = f(x_{i-1}, x_{i-2}, \ldots, x_{i-p}). \tag{9}$$

In the above equation, x_i is the observation of time t, and p is the number of past observations related to the future value. The SVR used for time series forecasting is actually featured with a general autoregressive model. The role of $f(x)$ is unspecified. This study aims to use SVR to fit $f(x)$ with the future values being well derived.

3.1 Data Set

In this study, we adopt daily values of F10.7[1] from 2002 to 2006 and predict the daily F10.7 within three days from 2003 to 2006. Since the set of input nodes precisely describes the correlation between lagged-time observations and future values, the number of input nodes is important in determining the structure of an autoregressive model with a time series. As the Sun rotation period is approximately 27 d, we choose 27 d as the window length of input patterns (i.e. the parameter p is 27, where $x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p})$). We find the training set length of 135 d (i.e. the parameter n is 135, which denotes the total number of data patterns) give a good result for the short-time prediction (within 3 d) model by assessing the relative performance of a prediction model. Chatterjee (2001) proposed a method based on traditional neural networks to predict the F10.7 values one day in advance. The data used for the test prediction were obtained during 1993, near the minimum of solar cycle 22, but the data used for training the network were obtained during the period from 1978 to 1988, over a complete solar cycle. Thus, for short-term prediction, the SVR-based model can work well by using fewer training data than the traditional neural network. Each data point is scaled by Equation (10) within the range of (0, 1). The scaling for the original data points helps to improve the forecasting accuracy.

$$X_s = 0.7\frac{X_t - X_{\min}}{X_{\max} - X_{\min}} + 0.15\,, \tag{10}$$

where X_t is the value of F10.7 at time t, $X_{\max}$ is the maximum of F10.7 during the period of data source and $X_{\min}$ is the minimum of F10.7 during the period of source data.

3.2 Performance Criteria

The performance of the prediction model based on SVR can be quantified in terms of some statistical metrics such as NMSE (Norm Mean Square Error), MAPE (Mean Absolute Percentage Error) and R. NMSE and MAPE are used to measure the deviation between the actual and predicted values. The smaller the values of NMSE and MAPE are, the closer the predicted values are to the actual values. The metric R is adopted to measure the correlation of the actual and the predicted values. The following shows their calculations

$$\mathrm{NMSE} = \frac{1}{\delta^2 n}\sum_{i=1}^{n}(y_i - x_i)^2, \qquad \delta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\,, \tag{11}$$

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n}|x_i - y_i|/x_i}{n} \times 100\%\,, \tag{12}$$

$$R = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}\,, \tag{13}$$

where x_i and y_i are the actual and predicted values, and $\bar{x}$ is the mean value of x_i, $i = 1 \dots n$.

3.3 Prediction Results

The results of SVR modeling of 10.7 cm solar radio flux are presented in Figures 4, 5 and 6. The initial data sets of daily F10.7 are presented by the solid line and the predicted values are presented by the dash-dot line. The relative errors are shown on the right. Tables 1, 2 and 3 show the performances of prediction model based on SVR. The predicted time series have good relationships with the measured F10.7 values, and the correlation coefficients reach as high as 0.99. In comparison to one step prediction, the NMSE is 0.15 and the MAPE is 2.71% during the best 1-year period of 2006, while they are 0.18 and 5.56% during the worst 1-year period of 2003; In comparison of the two step prediction, the NMSE

[1] *http://www.swpc.noaa.gov*

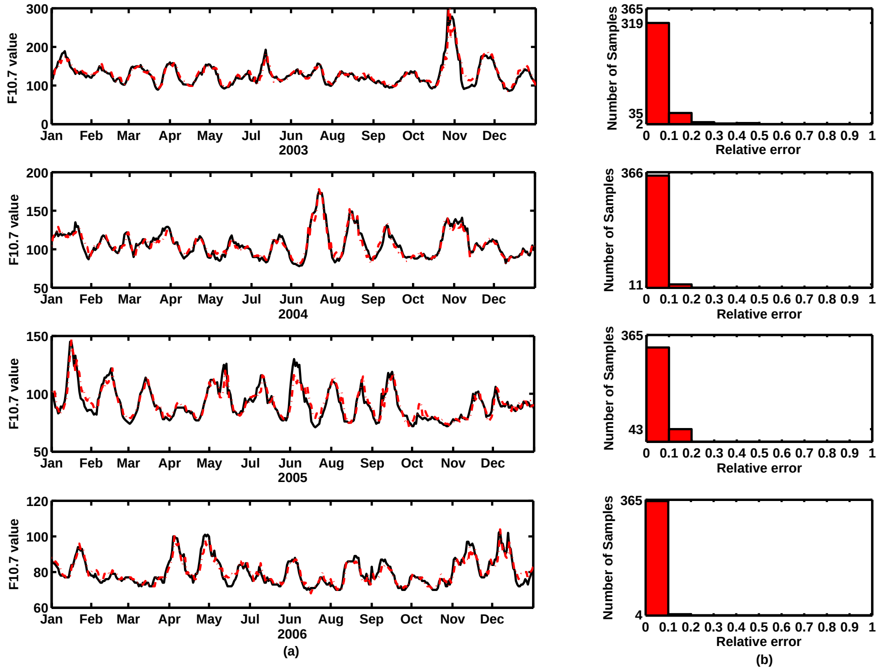

Fig. 4 One-step ahead (1 d) prediction with the SVR model. (a) The actual (solid line) and the predicted (dash-dot line) F10.7 value for 2003, 2004, 2005 and 2006. (b) The histogram of the relative differences of the predicted and the actual values.

is 0.28 and the MAPE is 3.60% during the best 1-year period of 2006, while they are 0.27 and 7.62% during the worst 1-year period of 2003; In comparison of the three step prediction, the NMSE is 0.42 and the MAPE is 4.48% during the best 1-year period of 2006, while they are 0.46 and 9.32% during the worst 1-year period of 2003. The performance of one-step prediction is superior to that of two-step prediction and three-step prediction.

4 DISCUSSION AND CONCLUSIONS

This study applied SVR to the forecasting fields of F10.7 values. The predictions of F10.7 agree well with the measured F10.7 values. From Tables 1, 2, and 3, it shows that the correlation between the observed and predicted data is about 99%, while they are 93% as referenced from the paper of Chatterjee (2001). Thus, the SVR model can perform well by using fewer training data points in short-term forecasting as compared with the traditional neural network based prediction techniques. However, the breaks of F10.7 values related to fierce solar activities are not predicted well.

Table 1 Norm mean square error (NMSE), mean absolute percentage error (MAPE) and correlation coefficients (R) on the test data set of one step ahead prediction from 2003 to 2006.

MODEL	MAPE(%)	NMSE	R
2003	5.56	0.18	0.9956
2004	3.69	0.16	0.9990
2005	4.58	0.17	0.9980
2006	2.71	0.15	0.9993

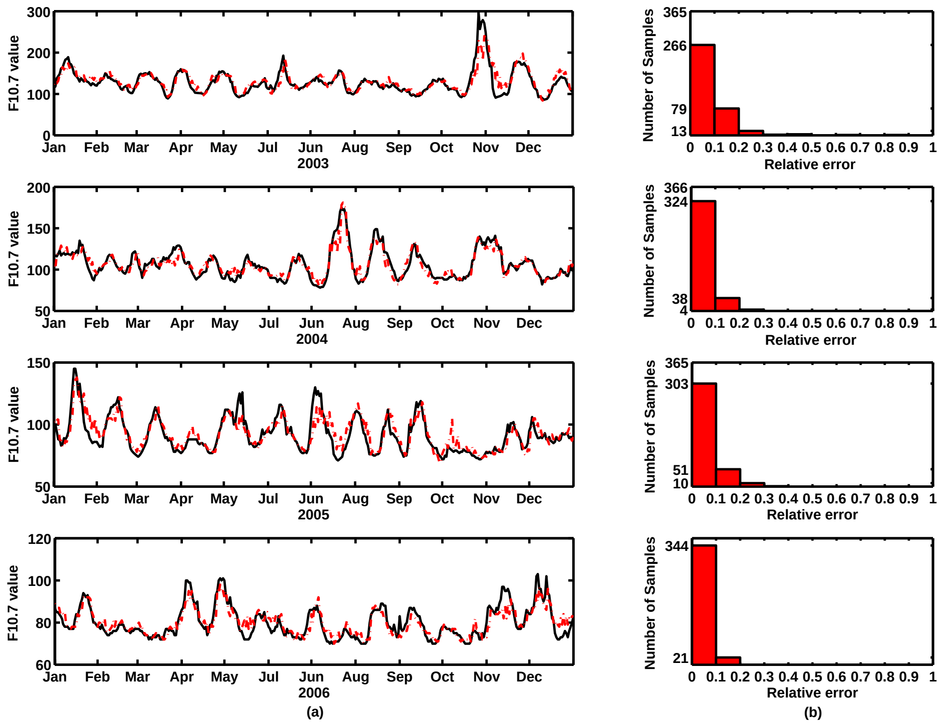

Fig. 5 Two-step ahead (2 d) prediction with the SVR model. (a) The actual (solid line) and the predicted (dash-dot line) F10.7 value for 2003, 2004, 2005 and 2006. (b) The histogram of the relative differences of the predicted and the actual values.

Table 2 Norm mean square error (NMSE), mean absolute percentage error (MAPE) and correlation coefficients (R) on the test data set of two step ahead prediction from 2003 to 2006.

MODEL	MAPE(%)	NMSE	R
2003	7.62	0.27	0.9933
2004	4.95	0.23	0.9979
2005	5.99	0.29	0.9966
2006	3.60	0.28	0.9988

Table 3 Norm mean square error (NMSE), mean absolute percentage error (MAPE) and correlation coefficients (R) on the test data set of three step ahead prediction from 2003 to 2006.

MODEL	MAPE(%)	NMSE	R
2003	9.32	0.46	0.9886
2004	7.28	0.38	0.9949
2005	6.83	0.38	0.9955
2006	4.48	0.42	0.9982

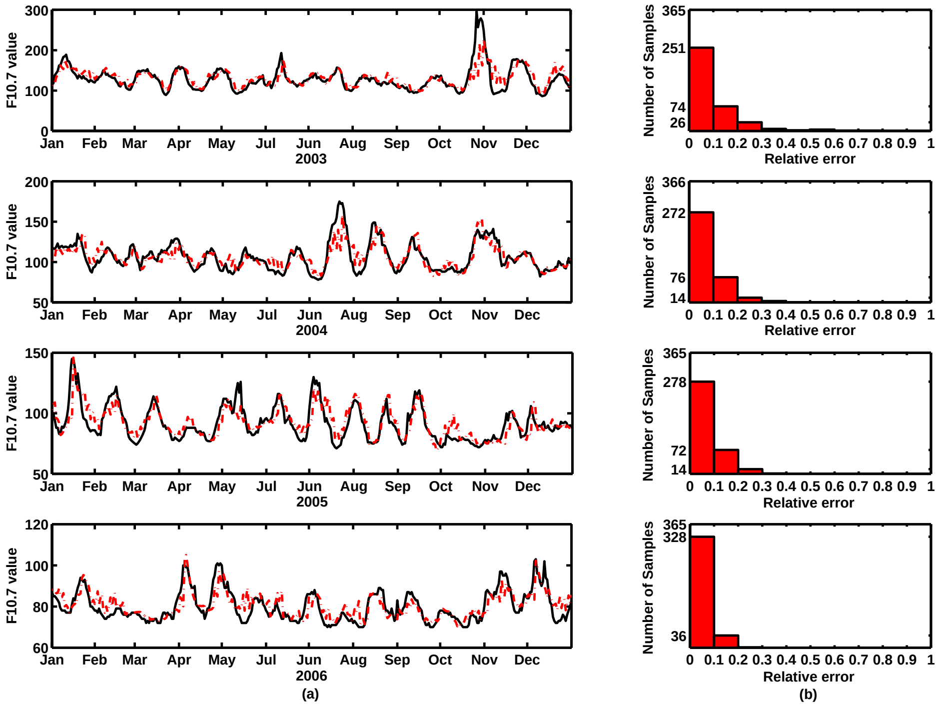

Fig. 6 Three-step ahead (3 d) prediction with the SVR model. (a) The actual (solid line) and the predicted (dash-dot line) F10.7 value for 2003, 2004, 2005 and 2006. (b) The histogram of the relative differences of the predicted and the actual values.

The prediction method based on SVR is general and we desire to implement future improvements. For example, the flux tube expansion factor, f_s, can be added to the input, which could improve the ability to predict the break of the F10.7 value caused by fierce solar activities. f_s is calculated by the equation $f_s = (\frac{R_{\rm ps}}{R_{\rm ss}})^2 \frac{B_{\rm ps}}{B_{\rm ss}}$, where $B_{\rm ps}$ and $B_{\rm ss}$ are the magnetic field strengths on the same field line at the photosphere and source surface respectively. $R_{\rm ps}$ is the radius of the photosphere, and $R_{\rm ss}$ is the radius of the source surface (Wintoft & Lundstedt 1999). It reflects the intensity of solar activity in some sense. The most important is that we can get the value of f_s per several minutes rather than per day. The parameter can reflect more delicate changes in the solar magnetic field. If it is added to the input of the SVR model, it will probably reduce the delay of prediction and improve the accurate prediction of large changes in F10.7. In the investigation of the prediction model, we found the method of scaling data within the range of (0,1) will influence the forecasting accuracy. Instead of Equation (10), we will try a more suitable equation to scale the data in further study.

References

Chatterjee, T. N. 2001, MNRAS, 323, 101
Cristianini, N., & Shawe-Taylor, J. 2000, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (Cambridge Univ. Press)
Dmitriev, A., Minaeva, Yu., Orlov, Yu., et al. 1999, the Hague, the Nethrelands: EGS XXIV General Assembly
Gavrishchaka, V. V., & Ganguli, S. B. 2001, J. Geophys. Res., 106, 29911
Li, R., Cui, Y. M., Han, H., & Wang, H. N. 2008, Advances in Space Research, 42(9), 1469
Mordvinov, A. V. 1985, Soln. Dannye, Byull., 12, 67

Ortikov, M. Yu., Shemelov, V. A., Shishigin, I. V., et al. 2003, Journal Atomospheric and Solar-Terrestrial Physics, 65, 1425
Peng, K. L., Wu, C. H., & Goo, Y. J. 2004, International Journal of Management, 21(4), 429
Schölkopf, B., & Smola, A. J. 2002, in Learning with Kernels (Cambridge: MIT Press)
Somla, A., & Schölkopf, B. 1997, Technical Report 1064, GMD
Smola, A., Schölkopf, B., & Müller, K. -R. 1998, in Proc. Of the Ninth Australian Conf. On Neural Networks, 79, eds. T. Downs, M. Frean, & M. Gallagher (Brisbane: University of Queensland)
Vapnik, V. 1995, The Nature of Statistical Learning Theory (New York: Springer Verlag)
Vapnik, V., Golowich, S., & Smola, A. 1997, in Advance in Neural Information Processing System, eds. M. Mozer, M. Jordan, & T. Petsche (Cambridge: MIT Press), 9, 281
Vapnik, V. 1998, Statistical Learning Theory (New York: Wiley)
Wintoft, P., & Lundstedt, H. 1999, J. Geophys. Res., 104, 6729
Zhao, J., & Han, Y. B. 2008, ChJAA (Chin. J. Astron. Astrophys.), 8, 472
Zhong, Q. Z., Liu, S. Q., He, J. X., et al. 2005, Chinese Journal of Space Science, 25, 199