Solar Physics
DOI: 10.1007/•••••-•••-•••-••••-•

Kalman Filter Technique for Improving Prediction of Smoothed Monthly Sunspot Numbers

T. Podladchikova[1] · **R. Van der Linden**[2]

Abstract In this work we develop a universal technique that improves medium-term prediction methods as they are monthly updated using the last available observations of smoothed sunspot numbers (see monthly output in real time at http://sidc.oma.be/products/kalfil/). The improvement of the predictions is provided by an adaptive Kalman filter based on using the last six monthly mean values of sunspot numbers, which cover the six months between the last available value of the 13-month running mean (the starting point for the predictions) and the current time. The proposed technique shifts the starting point for the predictions 6 months ahead and provides an increase of prediction accuracy for all the methods. Our technique has been tested on three medium-term methods of predictions from 6 to 18 months ahead which are now in operation: The McNish-Lincoln method (NGDC), the standard method (SIDC), the combined method (SIDC). The prediction accuracy for the McNish-Lincoln method (M&L) is increased by $17-30\%$, for the standard method (SM) by $5-21\%$ and for the combined method (CM) by $6-57\%$.

Keywords: Solar Cycle, Models; Solar Cycle, Observations; Sunspots, Statistics

1. Introduction

Medium-term prediction of the sunspot number has very high relevance in space weather applications. A number of different techniques have been proposed to predict future smoothed sunspot numbers of a solar cycle (usually the 13-month smoothed monthly mean sunspot number is used). The predictions that can be updated each month using the last available observations of the smoothed sunspot number and a method for their improvement are considered in this paper.

One of the better-known techniques for solar cycle prediction is the McNish-Lincoln method (McNish and Lincoln, 1949). It is based on averaging over past cycles to form a mean cycle. The difference between the smoothed sunspot numbers and the mean cycle is used to project future differences between

[1] First affiliation email: tatyana@oma.be
[2] Second affiliation email: ronald.vanderlinden@oma.be

predicted values and the mean cycle. The McNish-Lincoln regression technique originally used yearly values to get the predictions one year ahead. Steward and Ostrow (1970) have described a modification of the McNish-Lincoln technique that allowed prediction of monthly mean values. The Modified McNish-Lincoln auto-regression technique is used today to provide the prediction of the monthly smoothed sunspot number for the current solar cycle in the monthly Solar Geophysical Data reports published by the National Geophysical Data Center (NGDC).

The method to predict solar cycle profile from the steepness of the ascending phase of the cycle was proposed by Waldmeier (1968). The medium-term sunspot number prediction technique based on Waldmeier's interpolation curves is used now in monthly reports of the Solar Influence Data analysis Center (SIDC) as standard method. The currently observed smoothed curve is compared with the normalized smoothed monthly curves observed in the past, and choosing the normalized curve with the same slope yields predictions of the evolution of the cycle during future months (Koeckelenbergh, 1986).

A number of methods with different backgrounds and motivations have been compared to the McNish-Lincoln technique and the standard method. The 12 and 18 months predictions of the smoothed sunspot number based on Waldmeier's interpolation curves were compared with those given by a new method combining a regression model with predictions of the succeeding maximum based on precursors method (Denkmayr and Cugnon, 1997; Hanslmeier, Denkmayr, and Weiss, 1999). Predictions are obtained by computing the weighted average of the past cycles having a similar maximum, and are being gradually improved as new observations become available. The authors demonstrated the advantage of the new method in the ascending part of three cycles: 18, 21 and 22. This method is implemented today in SIDC as the combined method.

Relative comparisons between the McNish-Lincoln method and a neural network method were performed by Macpherson, Conway, and Brown (1995) and Fessant, Pierret, and Lantos (1996) who found that neural networks could provide a better prediction accuracy than the McNish-Lincoln method for cycle 22. Zhang (1996) showed that predictions given by nonlinear techniques are superior to that given by the standard method for part of cycle 22. Lantos (2006) analyzed the advantages of McNish-Lincoln method combined with the best of precursors.

However Conway (1998) and Lantos (2006), analyzing different medium-term prediction methods, indicated that it was difficult to decide definitely which prediction method was the best because different authors use different sets and formats of time series data and only relative comparisons between methods can be performed.

Lantos (2006) pointed that that the McNish-Lincoln method as well as all the precursors has rare but unacceptable errors of prediction. Thus any modified technique that improves the prediction of the current cycle can cause a risk of unacceptable prediction errors of other (past and future) cycles.

The objective of this paper is to develop a universal *auxiliary* technique that improves all the medium-term prediction methods as they are monthly updated using the last available observations of smoothed sunspot numbers without any risk of additional errors caused by this technique. The improvement of the predictions is provided by an adaptive Kalman filter based on using the last available six monthly mean observations of sunspot numbers, which cover the six months between the last available value of the 13-month running mean (the starting point for the predictions) and the current time. Usually, the last six monthly mean sunspot numbers are not directly used in prediction algorithms because of the large level of stochastic component included in them. Yet, they give significant information about future cycle evolution. A proper procedure of 'filtration' and their direct using for the prediction causes a shift of the starting point for the predictions 6 months ahead and therefore provides an increase of prediction accuracy for all the methods.

The prediction results obtained by any medium-term method are fed into the Kalman filter that causes an improvement of these predictions and are used for the construction of a state space stochastic model. Noise statistics of this model are determined on the basis of the developed identification method.

The proposed technique has been tested on the three medium-term methods of predictions that are now in operation: McNish-Lincoln method (reports published by National Geographic Data Center), standard method and combined method (reports published by Solar Influence Data analysis Center).

2. State space model

The last available observation of the smoothed sunspot number R_k is the starting point for the predictions at time k (in months). Let $\hat{R}_{k+j,k}$ denote the prediction of the smoothed sunspot number j months ahead on the basis of the original prediction method (which can be any method to get smoothed sunspot number prediction based on monthly updating) when the last observed smoothed sunspot number available is R_k. Let R^m_{k+j} $(j = 1, 2, \ldots, 6)$ denote the last six observations of the monthly mean sunspot numbers that are available.

To take into account in full measure the information contained in the last six available observations R^m_{k+j} $(j = 1, 2, \ldots, 6)$ we create the state space model of the sunspot cycle based on the predictions $\hat{R}_{k+j,k}$.

Let us describe the dynamics of smoothed sunspot numbers R_{k+j} in the form of the difference equation

$$R_{k+j} = \Phi_{k+j,k+j-1} R_{k+j-1} \quad (j = 1, 2, \ldots). \tag{1}$$

Here $\Phi_{k+j,k+j-1}$ is the transition function that reconstructs R_{k+j-1} at month $k + j - 1$ to R_{k+j} at month $k + j$.

The transition function depends on the model used and can be estimated by

$$\Phi_{k+1,k} = \frac{\hat{R}_{k+1,k}}{R_k},$$

$$\Phi_{k+j,k+j-1} = \frac{\hat{R}_{k+j,k}}{\hat{R}_{k+j-1,k}}, \quad (j = 2, 3, \ldots).$$

These transition functions $\Phi_{k+j,k+j-1}$ reconstruct estimates of the smoothed sunspot number $\hat{R}_{k+j-1,k}$ at month $k+j-1$ to estimates of the smoothed sunspot number $\hat{R}_{k+j,k}$ at month $k+j, (j=1,2,\ldots)$.

However, none of the models can be considered as fully reliable. As Conway (1998) pointed out, it is impossible to make precise predictions of the future because scientific theories are inseparable from random or stochastic effects, which can enter in one of two ways: either being intrinsic to the theory itself or as a consequence of measurement error.

The great variety of behavior of the 11-year sunspot cycle causes unpredictable deviation of the predicted value $\hat{R}_{k+j,k}$ from R_{k+j}. Therefore we introduce random noise w_{k+j} with unknown variance to the model (1), which allows us to take into account uncertainty and unpredictability of processes change in future.

Thus, the dynamics of process R_{k+j} is now governed by the stochastic difference equation

$$R_{k+j} = \Phi_{k+j,k+j-1} R_{k+j-1} + w_{k+j}, \quad (j=1,2,\ldots). \tag{2}$$

Here w_{k+j} represents the process noise intrinsic to the sunspot cycle. We consider it to be uncorrelated unbiased noise with variance $\sigma^2_{w,k+j}$ and assume it to be proportional to the smoothed sunspot number R_{k+j}

$$\sigma^2_{w,k+j} = a_w \cdot R_{k+j}, \tag{3}$$

where a_w is an unknown constant to be identified.

This assumption is related to results obtained by Hathaway et al (1994), who found that variance of monthly mean sunspot numbers is proportional to their values.

Equation (2) is the stochastic model which determinate base represents the origin model used to get the predictions $\hat{R}_{k+j,k}$. The process noise w_{k+j} allows us to take into account random deviations of predictions from the original model that can be estimated on the basis of the last six observations of the monthly mean sunspot number R^m_{k+j} $(j=1,2,\ldots,6)$.

The distortion of monthly mean sunspot numbers by measurement errors can be taken into account by construction of the measurement equation with additive measurement noise v_{k+j} that is given by

$$R^m_{k+j} = R_{k+j} + v_{k+j}, \;\; (j=1,2,\ldots,6). \tag{4}$$

Here the measurement noise v_{k+j} is supposed to be uncorrelated unbiased noise with variance

$$\sigma^2_{v,k+j} = a_v \cdot R_{j+k}, \tag{5}$$

where a_v is an unknown constant to be identified.

The improved prediction of R_{k+j} described by the state space model (2,4) can be obtained on the basis of a Kalman filter that provides an efficient computational means to estimate the state of a process in a way that minimizes the mean of the squared error by additionally using the last six available observations R^m_{k+j} $(j=1,2,\ldots,6)$.

3. Kalman filter algorithm

The Kalman filter (KF) is a very powerful technique including filtration and extrapolation to estimate current and future states quite precisely in conditions of incomplete information about the investigated processes (Kalman, 1960).

Let $\hat{R}^{KF}_{k+j,k+j}$ and $\hat{R}^{KF}_{k+j+1,k+j}$ $(j = 1, 2, \ldots, 6)$ represent respectively a filtered estimate of state R_{k+j} $(j = 1, 2, \ldots, 6)$ with variance of filtered error $\sigma^2_{k+j,k+j}$ and extrapolated estimate of state R_{k+j+1} $(j = 1, 2, \ldots, 6)$ with variance of extrapolated error $\sigma^2_{k+j+1,k+j}$ obtained on the basis of observations $R^m_{k+1}, \ldots, R^m_{k+6}$. Let $\hat{R}^{FK}_{k+j,k+6}$ $(j > 6)$ denote extrapolated estimates of the future states R_{k+j} $(j > 6)$ with variances of extrapolated error $\sigma^2_{k+j,k+6}$ $(j > 6)$ on the basis of observations $R^m_{k+1}, \ldots, R^m_{k+6}$.

In the actual implementation of the filter the variance of process noise $\sigma^2_{w,k+j}$ and variance of measurement noise $\sigma^2_{v,k+j}$ are usually determined prior to operation of the filter. To use the equations of the Kalman filter we begin with the goal of finding the unknown constants α_w and α_v used in the expressions (3,5) to determine these variances. Applying the developed identification technique of noise statistics we estimate the constants $a_w = 0.2$ and $\alpha_v = 2.6$ (see Appendix A).

To initiate the Kalman filter procedure we choose the last observed smoothed sunspot number R_k as the initial assessment of state $\hat{R}^{KF}_{k,k}$ with error variance $\sigma^2_{k,k} = 0$. Let $\hat{R}^{KF}_{k+1,k}$ denote an extrapolated estimate of state R_{k+1}.

The Kalman filter algorithm includes the following steps:

1) Extrapolation

The extrapolation equations for the next time step and estimates of the error variance are defined by

$$\hat{R}^{KF}_{k+j,k+j-1} = \Phi_{k+j,k+j-1} \cdot \hat{R}^{KF}_{k+j-1,k+j-1}, \quad j = (1, 2, \ldots, 6),$$

$$\sigma^2_{k+j,k+j-1} = \Phi^2_{k+j,k+j-1} \cdot \sigma^2_{k+j-1,k+j-1} + \sigma^2_{w,k+j}.$$

2) Filtration

Filtration equations incorporate a new measurement to obtain an improved estimate and are given by

$$\hat{R}^{KF}_{k+j,k+j} = \hat{R}^{KF}_{k+j,k+j-1} + K_{k+j} \cdot (R^m_{k+j} - \hat{R}^{KF}_{k+j,k+j-1}), \quad (j = 1, 2, \ldots, 6), \tag{6}$$

$$\sigma^2_{k+j,k+j} = (1 - K_{k+j}) \cdot \sigma^2_{k+j,k+j-1}, \tag{7}$$

where K is the filter gain and is represented by the following expression

$$K_{k+j} = \frac{\sigma^2_{k+j,k+j-1}}{\sigma^2_{k+j,k+j-1} + \sigma^2_{v,k+j}}.$$

Instead of the unknown R_{k+j} their estimates $\hat{R}^{KF}_{k+j,k+j-1}$ are substituted in the expressions (3,5) for variances $\sigma^2_{w,k+j}$ and $\sigma^2_{v,k+j}$.

As a result of the processing at time k of the last available six monthly mean observations R^m_{k+j} $(j = 1, 2, \ldots, 6)$ by the Kalman filter we obtain the improved estimate $\hat{R}^{KF}_{k+6,k+6}$ of state R_{k+6} in comparison with the value $\hat{R}_{k+6,k}$ obtained by the original prediction method. The variance of this estimate error is defined by $\sigma^2_{k+6,k+6}$.

The filtered estimate $\hat{R}^{KF}_{k+6,k+6}$ of state R_{k+6} is the new starting point for predictions that is shifted six month ahead in comparison with the last observed smoothed sunspot number R_k.

To determine improved predictions from 7 to 18 months ahead when observations of monthly mean sunspot numbers are not available yet we use the procedure of extrapolation given by

$$\hat{R}^{KF}_{k+j,k+6} = \Phi_{k+j,k+6} \cdot \hat{R}^{KF}_{k+6,k+6}, \ \ (j = 7, 8, \ldots, 18), \tag{8}$$

where $\Phi_{k+j,k+6} = \Phi_{k+j,k+j-1} \cdot \Phi_{k+j-1,k+6}$, $(j = 8, 9, \ldots, 18)$.
The variances of these estimates errors are defined in the recurrent way

$$\sigma^2_{k+j,k+6} = \Phi^2_{k+j,k+j-1} \cdot \sigma^2_{k+j-1,k+6} + \sigma^2_{w,k+j}, \ \ (j = 7, 8, \ldots, 18). \tag{9}$$

Thus, we obtain the improved predictions of smoothed sunspot numbers from 7 to 18 months ahead and variances of their estimates errors on the basis of expressions (8,9).

4. Exponential smoothing

The 13-month running mean obtained by averaging 13 monthly mean sunspot number isolates quite satisfactorily the component associated with the activity of a solar cycle. However the 6 month delay of estimation as compared to the last available monthly mean value is an essential drawback of this procedure. Applying the Kalman filter we can improve the prediction without delay taking into account the sunspot cycle dynamics characterized by the last six available monthly mean observations. However, because of the reduction of the sample size to 6 points, the filtration effectiveness of the stochastic component becomes lower and the prediction correction by the Kalman filter is characterized by an essentially fluctuating component. To obtain additional accuracy increase we apply exponential smoothing (ES) to Kalman filter estimates (Brown, 1963). The method of exponential smoothing is effective when there is an essentially random component in time series data and there is no great trend.

Applying the Kalman filter we obtained predictions, where

$$\left\{\hat{R}^{KF}_{k+6,k+6}, \hat{R}^{KF}_{k+7,k+6}, \ldots, \hat{R}^{KF}_{k+18,k+6}; (k = 1, 2, \ldots)\right\}$$

are the 13 time series of smoothed sunspot number predictions from 6 to 18 months ahead.

We apply exponential smoothing to these time series $\hat{R}^{KF}_{k+j,k+6}, \quad (k = 1, 2, \ldots)$ for every $j = 6, 7, \ldots, 18$ obtained for the estimates of R_{k+j} $(j = 6, 7, \ldots, 18)$.

Let $\hat{R}^{ES}_{k+j,k+6}\,(k = 1, 2, \ldots)$ for every $j = 6, 7, \ldots, 18$ represent the exponential moving average obtained as a result of exponential smoothing that in its simplest

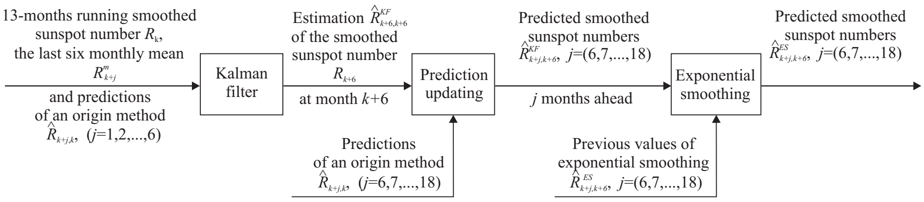

Figure 1. Schema of prediction algorithm.

form is given by

$$\hat{R}^{ES}_{k+j+1,k+7} = (1-\alpha) \cdot \hat{R}^{ES}_{k+j,k+6} + \alpha \cdot \hat{R}^{KF}_{k+j+1,k+7}, \quad (k = 1, 2, \ldots).$$

Here α is the exponential weight and $0 < \alpha < 1$.

5. Application to medium-term methods of predictions

The proposed prediction improvement technique of smoothed monthly sunspot numbers has been tested on three medium-term methods of predictions which are now in operation: The McNish-Lincoln method (predictions results published from August 1994 to May 2010), the standard method (prediction results published from January 1992 to May 2010) and the combined method (prediction results published from September 1997 to May 2010). To obtain improved prediction estimates we applied the adaptive Kalman filter (section 3) to the prediction results taking into account the last available six monthly mean observations of sunspot numbers. Additionally we used the procedure of exponential smoothing (section 4) to get the improved estimates of R_{k+j} ($j = 6, 7, ..., 18$). As the effectiveness of exponential smoothing decreases when there is a definite trend in the data, we also tested the proposed technique on prediction results obtained on with the combined method for cycle 22 (September 1986 – May 1996), which atypical ascending phase provides a good example of a clear trend. To obtain our results we started the procedure of exponential smoothing 23 months after the start of the cycle because the first half of an ascending phase of the 11-year sunspot cycle is usually characterized by a dominant trend. The prediction algorithm is schematically presented on Fig. 1.

The smoothed sunspot number R_k, the last available six observations of the monthly mean sunspot numbers R^m_{k+j} and the predictions $\hat{R}_{k+j,k}$, ($j = 1, 2, ..., 6$) from 1 to 6 months ahead of an original method are fed into to the Kalman filter. Applying the Kalman filter we obtain the improved estimate $\hat{R}^{KF}_{k+6,k+6}$ of state R_{k+6}. This estimate becomes the new starting point for the prediction updating that is shifted 6 months ahead in comparison with the last observed smoothed sunspot number R_k. On the basis of the improved estimate $\hat{R}^{KF}_{k+6,k+6}$ and predictions $\hat{R}_{k+j,k}(j = 6, 7, ...)$ of an original method from 6 to 18 months ahead we obtain improved predictions $\hat{R}^{KF}_{k+j,k+6}$, ($j = 7, 8, ..., 18$) from 7 to 18 months ahead. Then we apply the exponential smoothing to the Kalman filter estimates $\hat{R}^{KF}_{k+j,k+6}$, ($k = 1, 2, ...;\ j = 6, 7, ..., 18$).

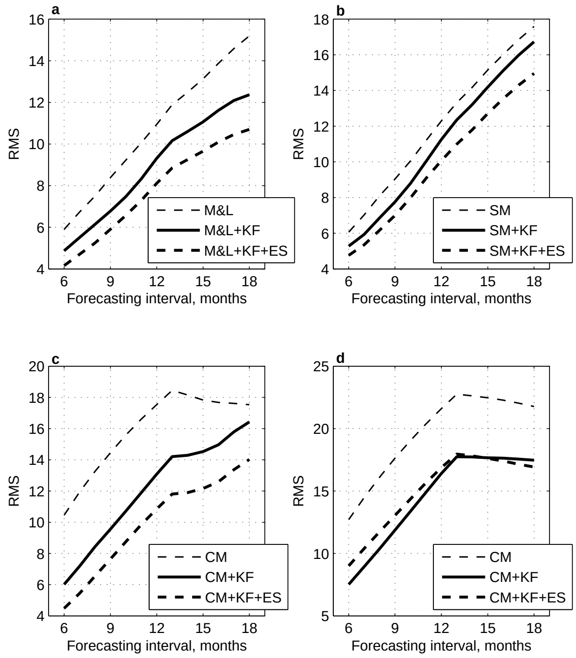

Figure 2. Root mean square errors of predictions from 6 to 18 months ahead. a) The McNish-Lincoln method (M&L); b) The standard method (SM); c) The combined method (CM) for the period from September 1997 to May 2010; d) The combined method (CM) for cycle 22 (September 1986 – May 1996). The thin dotted line shows RMS errors of predictions obtained on the basis of the original prediction method. RMS errors of the new prediction after applying the Kalman filter are marked by bold solid lines. Bold dotted lines depict the additional correction by the exponential smoothing (ES).

The root mean square errors (RMS) of the prediction from 6 to 18 months ahead are shown in Fig. 2.

As Fig. 2 shows, the Kalman filter statistically improves the predictions from 6 to 18 months ahead for all three methods. Exponential smoothing reduces additionally the errors of prediction (Fig. 2a,b,c). In the conditions of pronounced trend as in cycle 22 (Fig. 2d), the exponential smoothing insignificantly decreases the accuracy of prediction of the Kalman filter from 6 to 12 months ahead and gives smaller RMS errors from 13 to 18 months ahead. This gives grounds to use the exponential smoothing to improve the prediction without risk of additional errors caused by pronounced trends.

Fig. 3 shows predictions of the smoothed sunspot numbers 12 months ahead.

This figure shows that our technique improves the prediction obtained by the original method in the ascending phase for all the methods. In the descending phase the effectiveness of the proposed technique either increases or decreases in comparison with predictions of an original prediction method.

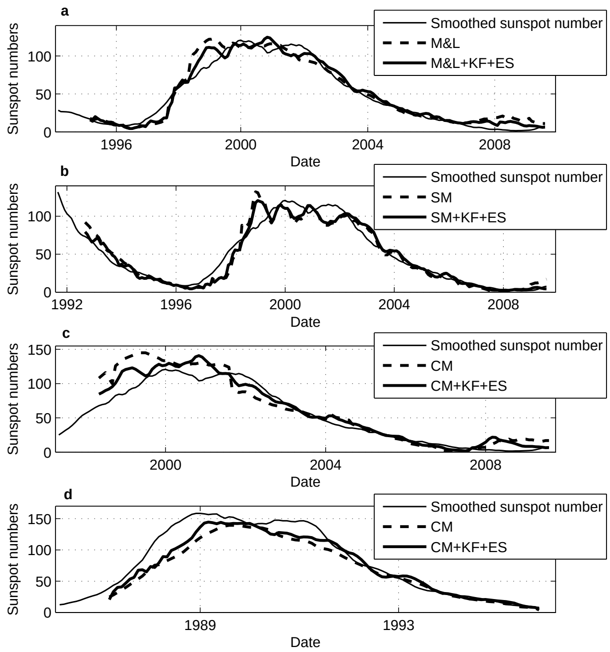

Figure 3. Prediction of smoothed sunspot numbers 12 months ahead. a) The McNish-Lincoln method (M&L); b) The standard method (SM); c) The combined method (CM) for the period from September 1997 to May 2010; d) The combined method (CM) for the cycle 22 (September 1986 - May 1996). Thin solid line shows 13-month running smoothed sunspot numbers. Dotted line depicts prediction obtained on the basis of origin prediction method. Prediction after applying Kalman filter (KF) and exponential smoothing (ES) are marked by bold solid line.

The Kalman filter technique provides not only estimates of present and future sunspot numbers, but also gives estimates of prediction accuracy. We obtain the variance of prediction error 6 months ahead using expression (7) when $j = 6$ and variances of prediction errors from 7 to 18 months ahead on the basis of expression (9). Supposing prediction errors are normally distributed we construct the 95% prediction interval as the range from $\hat{R}^{KF}_{k+j,k+6} - 2\sigma_{k+j,k+6}$ to $\hat{R}^{KF}_{k+j,k+6} + 2\sigma_{k+j,k+6}$, $(j = 6, 7, ..., 18)$, where $\hat{R}^{KF}_{k+j,k+6}$ is a prediction of smoothed sunspot number j months ahead obtained with the Kalman filter. Fig. 4 shows the estimate of the 95% prediction interval for the prediction 12 months ahead.

It is seen that almost all smoothed sunspot numbers fall in the 95% prediction interval. In the ascending phase of cycle 22 the smoothed sunspot number exceeds the limits of the 95% prediction interval during the year 1991 (Fig. 4d). This is evidently caused by the atypical ascending phase of the cycle.

Table 1 gives RMS errors of the predictions 6, 12 and 18 months ahead for the original prediction method and the correction on the basis of the Kalman filter with and without exponential smoothing and shows a reduction of RMS predic-

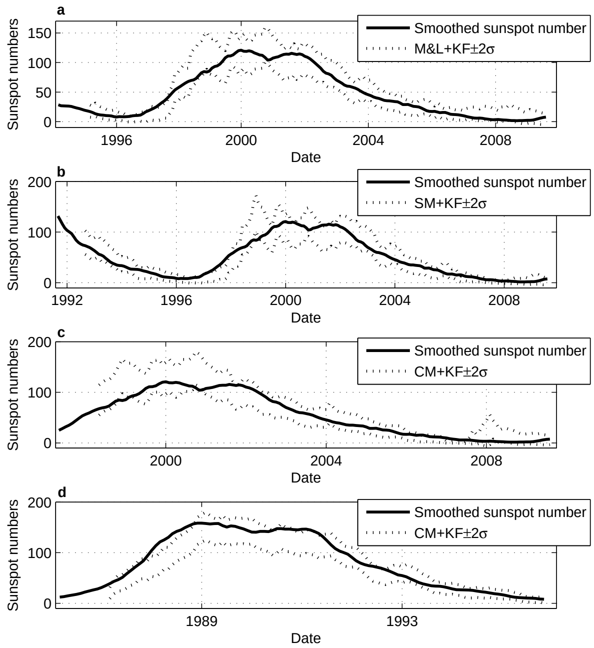

Figure 4. Estimate of the 95% prediction interval for the prediction 12 months ahead. The solid line shows the 13-month running smoothed sunspot numbers. The dotted lines depict the 95% prediction interval for a) The McNish-Lincoln method (M&L)+Kalman filter (KF); b) The standard method (SM)+Kalman filter (KF); c) The combined method (CM)+Kalman filter (KF) for the period from September 1997 to May 2010; d) The combined method (CM)+Kalman filter (KF) for cycle 22 (September 1986 - May 1996).

tion errors using the Kalman filter and exponential smoothing in comparison to the original prediction method.

As shown in Table 1, applying the Kalman filter and exponential smoothing performs well in all the methods. The proposed technique reduces RMS errors of predictions of an original method 6 months ahead from 5 to 57%. The exponential smoothing insignificantly decreases the effectiveness of Kalman filter when it is applied to obtain predictions 6 month ahead for cycle 22. However, even in this case we have essential improvements of prediction of 29% in comparison with the original prediction method. Therefore we propose to use additionally exponential smoothing after Kalman filter application without risk of important additional errors.

6. Conclusion and Discussion

Medium-term prediction methods as they are monthly updated using the last available observations of smoothed sunspot numbers (13-month running

Table 1. RMS errors of predictions 6, 12 and 18 months ahead and reduction of RMS prediction errors using a Kalman filter and exponential smoothing in comparison to an original prediction method.

Prediction method	6 months ahead prediction		12 months ahead prediction		18 months ahead prediction	
	RMS errors	ΔRMS	RMS errors	ΔRMS	RMS errors	ΔRMS
M&L	5.9		10.9		15.2	
M&l+KF	4.9	17%	9.3	15%	12.4	18%
M&l+KF+ES	4.2	29%	8.1	26%	10.7	30%
SM	6.1		12.3		17.6	
SM+KF	5.3	13%	11.3	8%	16.7	5%
SM+KF+ES	4.8	21%	10.1	18%	15.0	15%
CM (cycle 23)	10.4		17.5		17.5	
CM+KF (cycle 23)	6.0	42%	13.1	25%	16.4	6%
CM+KF+ES (cycle 23)	4.5	57%	10.9	38%	14.0	20%
CM (cycle 22)	12.7		21.6		21.8	
CM+KF (cycle 22)	7.5	41%	16.4	24%	17.5	20%
CM+KF+ES (cycle 22)	9.0	29%	16.9	22%	16.9	22%

smoothed sunspot number) do not directly use the last six observed monthly mean sunspot numbers because of the large level of stochastic component included in them, although they give significant information about cycle evolution in future.

Our study showed that the Kalman filter provides filtration of this stochastic component and predictions improvement of any original method by taking into account the dynamics of the sunspot cycle characterized by the last available six monthly mean observations of sunspot number, which cover the six months between the last available value of the 13-month running mean (the starting point for the predictions) and the current time. The proper procedure of their filtration and their direct use for the prediction shifts the starting point for prediction updating 6 months ahead and provides an increase of prediction accuracy for all the methods.

We created the state space model of sunspot activity, which determinate base was created on the basis of predictions of an original method. As any of the models can be considered as fully reliable and sunspot activity is inseparable from stochastic effects that are intrinsic to the process itself and are a consequence of measurement error, we introduced process and measurement noises to this model. To determine noise variances of the state space model we developed noise statistics identification methods, which allowed to take into account random deviations of predictions from an original prediction model using the Kalman filter and to correct the original predictions. The Kalman filter also allowed to obtain estimates of prediction accuracy.

To obtain additional increase of predictions accuracy we reduced the random component of the Kalman filter estimates applying exponential smoothing.

Our technique has been tested on three medium-term methods of predictions from 6 to 18 months ahead that are now in operation: the McNish-Lincoln method (NGDC), the standard method (SIDC), the combined method (SIDC). Test results showed that the Kalman filter improved predictions from 6 to 18 months ahead for all three methods. Exponential smoothing additionally reduced prediction errors. The prediction accuracy for the McNish-Lincoln method (M&L) is increased by 17-30%, for the standard method (SM) by 5-21% and for the combined method (CM) by 6-57%.

The real-time output of Kalman filter technique is in operation now at http://sidc.oma.be/products/kalfil/. Monthly reports are published by Solar Influence Data analysis Center.

Appendix

A. Identification of noise statistics

To determine noise statistics of the stochastic state space model (2,4) we use observations of monthly mean sunspot numbers R_j^m $(j = 3, 4, \ldots)$ from 1849 to 2009 (Podladchikova, 2006; Pankratova and Podladchikova, 2008). Let us rewrite the model (2,4) into

$$R_j = \Phi_{j,j-1} R_{j-1} + w_j, \tag{10}$$

$$R_j^m = R_j + v_j. \tag{11}$$

Here $\sigma_{w,j}^2 = a_w \cdot R_j$ and $\sigma_{v,j}^2 = a_v \cdot R_j$ are unknown variances.

To determine the constants α_w and a_v we create sequences $B_{1,j}$ and $B_{2,j}$

$$B_{1,j} = \left(R_j^m - \Phi_{j,j-1} R_{j-1}^m\right)^2,$$

$$B_{2,j} = \left(R_j^m - \Phi_{j,j-2} R_{j-2}^m\right)^2.$$

Using expressions (10) and (11) these sequences can be presented in the form of an explicit function of model noises w_j and η_j

$$B_{1,j} = \left(w_j + v_j - \Phi_{j,j-1} v_{j-1}\right)^2,$$

$$B_{2,j} = \left(\Phi_{j,j-1} w_{j-1} + w_j + v_j - \Phi_{j,j-2} v_{j-2}\right)^2.$$

Then the mathematical expectations of $B_{1,j}$ and $B_{2,j}$ are presented in the form of linear functions of constants α_w and α_v

$$E\left[B_{1,j}\right] = \alpha_w \cdot R_j + \alpha_v \cdot R_j + \Phi_{j,j-1}^2 \cdot \alpha_v \cdot R_{j-1}, \tag{12}$$

$$E\left[B_{2,j}\right] = \Phi_{j,j-1}^2 \cdot \alpha_w \cdot R_{j-1} + \alpha_w \cdot R_j + \alpha_v \cdot R_j + \Phi_{j,j-2}^2 \cdot \alpha_v \cdot R_{j-2}.$$

The difference of expectations is given by

$$E\left[B_{1,j}\right]-E\left[B_{2,j}\right]=\Phi_{j,j-1}^{2}\cdot\alpha_{v}\cdot R_{j-1}-\ \Phi_{j,j-1}^{2}\alpha_{w}\cdot R_{j-1}\ -\ \Phi_{j,j-2}^{2}\alpha_{v}\cdot R_{j-2}.$$

We divide the last expression by $\Phi_{j,j-1}^{2}$

$$\frac{E\left[B_{1,j}\right]-E\left[B_{2,j}\right]}{\Phi_{j,j-1}^{2}}=\alpha_{v}\cdot R_{j-1}-\ \alpha_{w}\cdot R_{j-1}\ -\ \Phi_{j-1,j-2}^{2}\alpha_{v}\cdot R_{j-2}.$$

Let us add to both parts of this last expression $E\left[B_{1,j-1}\right]$ defined by (12) to obtain the constant α_v in the following way

$$\alpha_{v}=\frac{1}{2\cdot R_{j-1}}E\left[\frac{\left(B_{1,j}-B_{2,j}\right)}{\Phi_{j,j-1}^{2}}+B_{1,j-1}\right]. \qquad (13)$$

Thus, the estimate $\hat{\alpha}_v$ of α_v is given by

$$\hat{\alpha}_{v}=\frac{1}{2}\sum_{j}\frac{1}{R_{j-1}}\left[\frac{\left(B_{1,j}-B_{2,j}\right)}{\Phi_{j,j-1}^{2}}+B_{1,j-1}\right].$$

We determine the constant α_w from the expression (12) taking into account expression (13).

$$\alpha_{w}=\frac{1}{2\cdot R_{j}}E\left[B_{2,j}-\frac{1}{\Phi_{j+1,j}^{2}}\left(B_{1,j+1}-B_{2,j+1}\right)-\Phi_{j,j-1}^{2}B_{1,j-1}\right].$$

Thus, the estimate $\hat{\alpha}_w$ of α_w is given by

$$\hat{\alpha}_{w}=\frac{1}{2}\sum_{j}\frac{1}{R_{j}}\left[B_{2,j}-\frac{1}{\Phi_{j+1,j}^{2}}\left(B_{1,j+1}-B_{2,j+1}\right)-\Phi_{j,j-1}^{2}B_{1,j-1}\right].$$

The proposed identification technique was applied to determine the unknown constants α_w and α_v.

The transition function in expression (10) was formed on the basis of analytic functions simulating cycle profiles that were introduced by Hathaway, Wilson, and Reichmann (1994). Then the estimate $\hat{R}_j$ of smoothed sunspot numbers R_j is given by

$$\hat{R}_{j}=\frac{a(j-j_{0})^{3}}{\exp\left({(j-j_{0})^{2}}/{b^{2}}\right)-0.71},$$

where time j is measured in months, j_0 denotes the initial starting time, a is a parameter connected directly with cycle amplitude, b depends on time from minimum to maximum and is connected with a in the following way

$$b\left(a\right)=27.12+\frac{25.15}{\left(a\cdot10^{3}\right)^{\frac{1}{4}}}.$$

Then the transition function that reconstructs the predicted state of the system at month j to the predicted state of the system at month $j+1$ is defined as

$$\Phi_{j+1,j} = \frac{\hat{R}_{j+1}}{\hat{R}_j} = \frac{\left(1+\frac{1}{j-j_0}\right)^3 \left(\exp\left(\frac{(j-j_0)^2}{b^2}\right) - 0.71\right)}{\exp\left(\frac{(j-j_0+1)^2}{b^2} - 0.71\right)}.$$

The value of parameter a is essential for the accuracy of curve fitting. However, the identification of constants α_w and α_v for different values of parameter a demonstrated an insensitivity of their estimates to the choice of parameter a. Irrespective of the value of a, the estimates of α_v vary in a quite narrow range from 1.8 to 3.4 for all the cycles from 10 to 23. Therefore in the prediction algorithm based on Kalman filter we used the average value for all the cycles $\alpha_v = 2.5$. The estimate of α_w does not exceed 0.2 for all the cycles. To provide robustness of Kalman filter we chose the maximum value $\alpha_w = 0.2$ for the prediction algorithm.

Acknowledgements The authors are grateful to S.Willems for online installation of operational programs, the observational collective of The Solar Influences Data analysis Center (SIDC) at the Royal Observatory of Belgium (ROB) and National Geophysical Data Center (NGDC) for data sets.

References

Brown, R.G.: *Smoothing forecasting and prediction in discrete time series*, Prentice Hall, New Jersey, 468 pp.

Conway, A.J.: 1998, Time series, neural networks and the future of the Sun. *New Astron. Rev.* **42**, 343 – 394.

Denkmayr, K., Cugnon, P.: 1997, About Sunspot Number Medium-Term Predictions. In: Heckman, G., Maruboshi, K., Shea, M.A., Smart, D.F., Thompson, R. (eds.) *Proceedings of the 5th Solar-Terrestrial Predictions Workshop*, Hiraiso Solar Terrestrial Research Center, Japan, 103.

Fessant, F., Pierret, C., Lantos, P.: 1996, Comparison of Neural Network and McNish and Lincoln Methods for the Prediction of the Smoothed Sunspot Index. *Solar Phys.* **168**, 423 – 433.

Hanslmeier, A., Denkmayr, K., Weiss: 1999, Longterm Prediction of Solar Activity Using the Combined Method. *Solar Phys.* **184**, 213 – 218.

Hathaway, D.H., Wilson, R.M., Reichmann, E.J.: 1994, The shape of the sunspot cycle. *Solar Phys.* **151**, 177 – 190.

Kalman, R.E.: 1960, A New Approach to Linear Filtering and Prediction Problems. *Trans. ASME-J. Basic Eng.* **82**.

Koeckelenbergh, A.: 1986, Comments on medium -term prediction of solar activity. In: Simon, P.A., Heckman, G., Shea, M.A. (eds.) *Solar-Terrestrial Predictions*, US Department of Commerce, NOAA, ERL, Boulder, 113 – 117.

Lantos, P.: 2006, Solar Cycle Prediction: Combining Precursor Methods with McNish and Lincoln Technique. *Solar Phys.* **236**, 399 – 413.

Macpherson, K.P., Conway, A.J., Brown, J.C.: 1995, Prediction of solar and geomagnetic activity data using neural networks. *J. Geophys. Res.* **100**, 21735 – 21744.

McNish, A.G., Lincoln, J.V.: 1949, Prediction of Sunspot Numbers. *Eos, Trans. Amer. Geophys. Union.* **30**, 673.

Pankratova, N.D., Podladchikova, T.V.: 2008, Estimation and prediction of difficult formalized processes of different physical nature. *J. Automn. Inform. Sci.* **40**, 23 – 35.

Podladchikova, T.: 2006, Identification of unknown noise statistics for non-stationary state space systems. In: *The Proceedings of Baltic Olympiad on Automatic Control*, 103 – 107.
Steward, F.G., Ostrow, S.M.: 1970, Improved version of the McNish-Lincon method for prediction of solar activity. *Telecomm.J.* **37**, 228.
Waldmeier, M.: 1968, Sonnenfleckenkurven und die Methode der Sonnenaktivittsprognose. *Astr. Mitt. Sternw.* **286**, 13.
Zhang, Q.: 1996, A nonlinear prediction of the smoothed monthly sunspot numbers. *Astron. Astrophys.* **310**, 646 – 650.