ENVIRONMENTAL SIGNAL PREDICTION

I.Šindelářová [1], *A.Procházka* [2], *and M.Mudrová* [2]

[1] University of Economics, Department of Econometrics
[2] Institute of Chemical Technology, Department of Computing and Control Engineering

Abstract

The paper presents the contribution to the analysis and prediction of dust particles concentration over the Czech Republic. The observations of a selected measuring station are in detail analyzed and the proposed method is then applied to further values observed by other measuring stations. The mathematical background of the paper includes wavelet transform, signal de-noising, and autoregressive modelling.

1 Introduction

The goal of the paper is to suggest an appropriate model for the real data forecasting. As a result an autoregressive model has been chosen for this purpose. To achieve better results of the given signal prediction it is useful to apply selected preprocessing methods including signal de-noising using selected thresholding methods. Mathematical tools used in the study include

- *wavelet transform* and *thresholding* for the data *de-noising*
- *autocorrelation* for the autoregressive model legitimacy
- *Akaike information criterion use* for the determination of the AR model order
- *autoregressive modelling* for the given signal forecasting

2 Signal De-Noising

The given real data are de-noised before their modelling and prediction. The following results contain analysis and prediction of dust particles concentration measured at the measuring station selected from the total number of 126 stations [4]. All observations are obtained with the sampling period of 2 hours. Fig. 1 illustrates dust particles concentration (black line) and its de-noised representation (blue line). Signal de-noising is based upon the wavelet transform [6]. The process of de-noising [5] is described by the following algorithm.

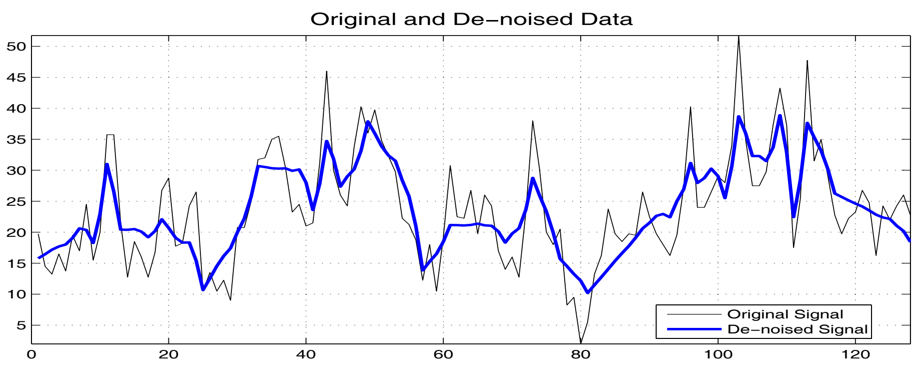

Figure 1: Real data of dust particles quantity measured each two hours at the selected station with its de-noised representation (soft thresholding, $\delta = 0.4$)

Algorithm:. Signal De-Noising based upon the Thresholding using DWT

- *Decomposition.* The discrete wavelet transform of the given signal resulting in the series of wavelet coefficients organized to the chosen number of levels.
- *Thresholding.* The selection of the thresholding level. The modification of wavelet coefficients by means of the soft or hard thresholding.
- *Reconstruction.* The application of the inverse discrete wavelet transform to the modified wavelet coefficients.

3 Signal Modelling and Prediction

Analysed signal part is relatively short (128 samples) and the segment can be considered as a stationary one. Values of the partial autocorrelation function (PACF) in Fig. 2(**b**) indicate that the signal can be studied as an autoregressive proces [1] of order $K = 1$. This result is verified by the Akaike function values illustrated in Fig. 3. An index with the minimal value of this function is index 1 for both cases of original and de-noised signal [2]. The future value of the series depends mainly on the previous one. A *random walk* is thus concerned in this case [1].

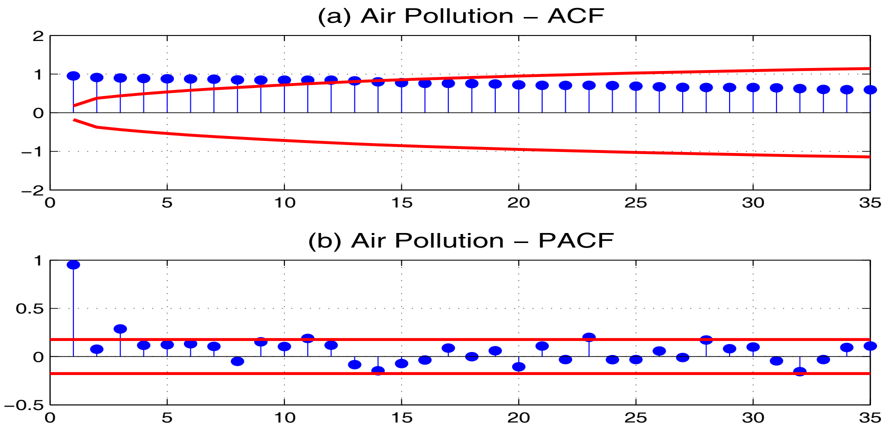

Figure 2: Results of processing of air pollution in chosen 10 days (sampling period is two hours) including **(a)** estimated values of autocorrelation function and **(b)** estimated values of the partial autocorrelation function - the significant level is meaningfully exceeded only by the first value

Further it is possible to remark in the case of the thresholded signal that values acceding 1 for model order are not so penalized. As a result the model order $K = 3$ can be considered for the de-noised data.

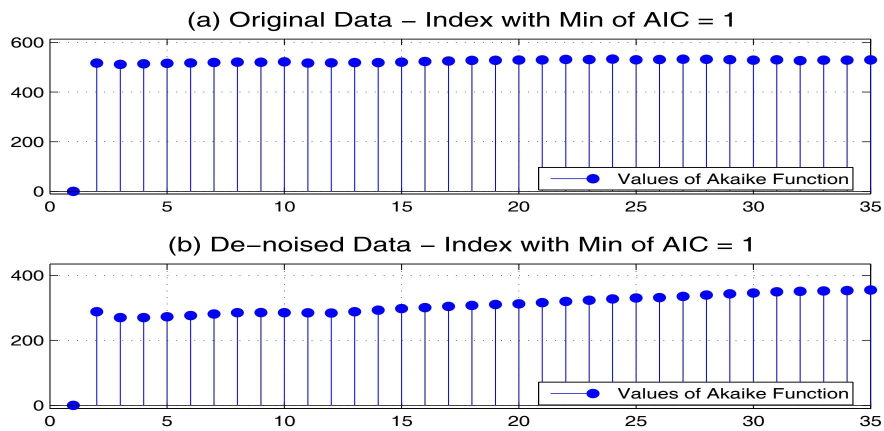

Figure 3: Values of Akaike function for AR model of dust particles concentration evaluated for **(a)** the original signal and **(b)** the de-noised signal (using soft thresholding, $\delta = 0.4$)

The Table 1 contains results of one step ahead prediction of dust particles concentration measured at one chosen station. Soft and hard thresholding has been used in this case. Threshold levels δ's changing from value 0.1 to 1.0 express the removed relative part of wavelet coefficients corresponding to noisy signal components.

Table 1: The model order K and the deviation ratio $MDevR$ of one step ahead prediction in dependence on threshold level and the manner of thresholding. In the case of the original signal (dust particles concentration) the AR model order is $K = 1$ and $MDevR = 27.99\%$

δ *(Soft)*	K	*MDevR*	δ *(Hard)*	K	*MDevR*
0.1	1	18.82	0.1	1	27.85
0.2	1	14.15	0.2	1	27.03
0.3	1	11.38	0.3	1	20.61
0.4	1	9.35	0.4	1	20.24
0.5	1	7.87	0.5	1	16.55
0.6	1	7.09	0.6	1	14.07
0.7	1	6.38	0.7	1	12.30
0.8	1	5.94	0.8	1	10.93
0.9	1	5.72	0.9	1	7.56
1.0	1	5.59	0.9	1	5.59

As it was possible to expect the mean deviation ratio descends with the growing value of δ. The use of the soft thresholding is more preferable in connection with the following prediction. Values of $MDevR$ descend faster than in the case of hard thresholding and additional hard thresholding changes the signal more radically for the same threshold levels than the soft version of thresholding. Fig. 4 illustrates the forecasting of soft thresholded signal for $\delta = 0.4$. The mean deviation ratio drops bellow 10%.

The AR model of order $K = 1$ stays for all listed thresholded series. The value of K is equal to 1 according to the Akaike criterion, which is certified by the dominant value of partial autocorrelation function for the data with index equal to 1. The only autoregressive coefficient in this signal modelling is $a_1 = 0.984$ and forecasted values can be evaluated by relation

$$x(n) = 0.984\, x\,(n-1) \tag{1}$$

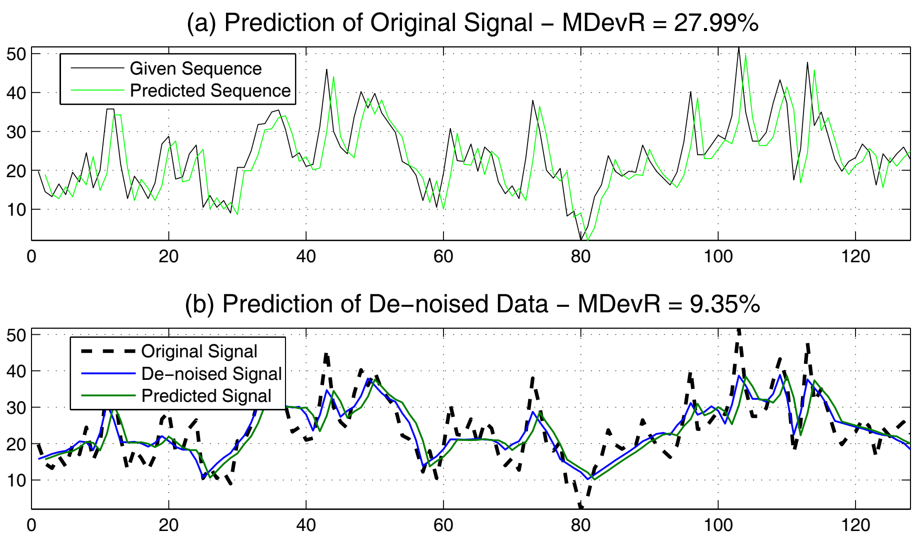

Figure 4: One step ahead real signal prediction (dust particles concentration) presenting the prediction of **(a)** the original data and **(b)** the de-noised data (soft thresholding, δ=0.4, K=1)

As it has been mentioned above the model order $K = 3$ is acceptable for de-noised data, too. Fig. 5 illustrates that the model of order $K = 3$ deteriorates mean deviation ratio in prediction of the original signal but for the thresholded signal the prediction is more successful in the sense of the mean deviation ratio - $MDevR = 8.89\,\%$ for $K = 3$ comparing to $MDevR = 9.35\,\%$ for $K = 1$. The predicted signal can be evaluated in this case by relation

$$x(n) = 1.325\,x(n-1) - 0.757\,x(n-2) + 0.428\,x(n-3) \tag{2}$$

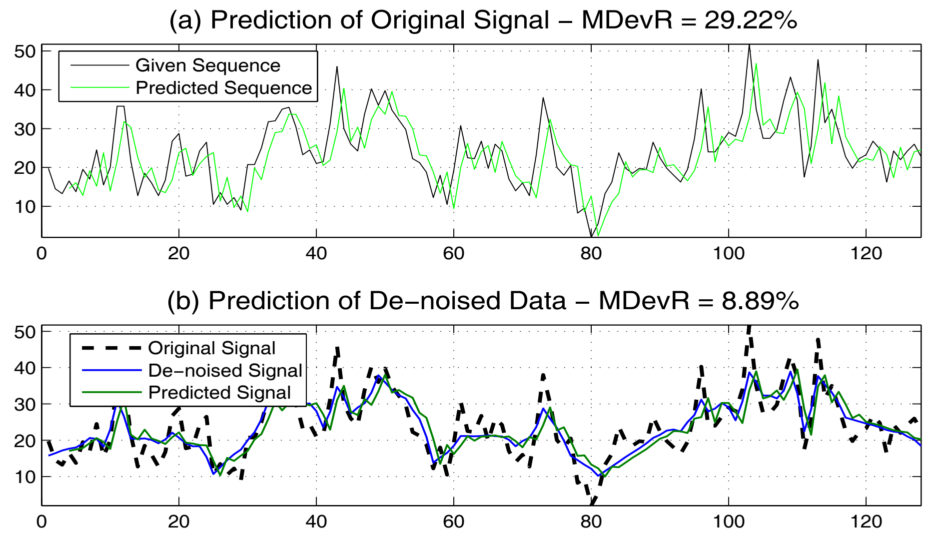

Figure 5: One step ahead real signal prediction (dust particles concentration) presenting the prediction of **(a)** the original data and **(b)** the de-noised data (soft thresholding, Th=0.4, K=3)

4 Conclusion

Methods mentioned above have been applied to selected measuring station in the Czech Republic. It is possible to assume that the data from all measuring stations have similar characteristics.

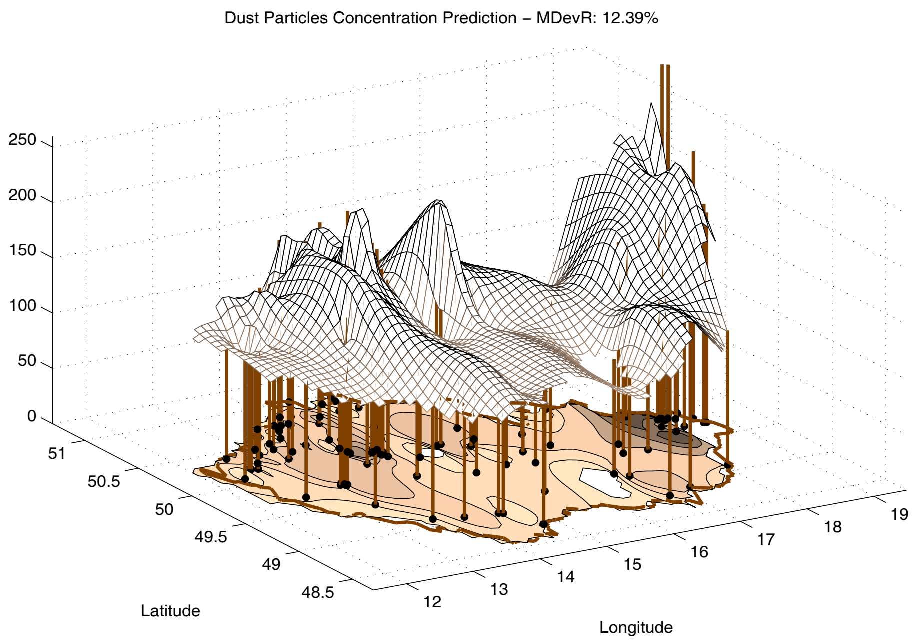

Figure 6: The two-dimensional interpolation of values of dust particles concentration using predicted values evaluated by the autoregressive model of the third order

Only the stations with more than 1000 missing values are omitted. Contingent missing values in considered series are replaced by interpolated values [3]. The numerical experiment proved the effect of signal de-noising to its prediction.

In summary it is suggested to use the following procedure for all selected data series of dust particle concentration: *For de-noising the soft thresholding method of the third decomposition level of wavelet coefficients is used with the threshold level* $\delta = 0.4$. *Time series containing the data of air pollution can be predicted with under 10% mean deviation ratio after the noise reduction.*

5 Acknowledgments

The work has been supported by the research grant of the Faculty of Chemical Engineering of the Institute of Chemical Technology, Prague No. MSM 6046137306. Data were kindly provided by Ing. A.Sieglerová from the Czech Hydrometeorological Institute. Authors express their thanks also to colleagues from the Department of Econometrics of the University of Economics in Prague.

References

[1] J. Arlt and M. Arltová. *Finanční časové řady.* Grada Publishing, Praha, first edition, 2003.

[2] P. P. Kanjilal. *Adaptive Prediction and Predictive Control.* Peter Peregrinus Ltd., IEE, U.K., 1995.

[3] M. Kolínová. *Environmental Image Analysis and Processing.* PhD thesis, Institute of Chemical Technology, Prague, 2001.

[4] M. Mudrová. *Lineární a nelineární metody analýzy a zpracování časových řad.* PhD thesis, Institute of Chemical Technology, Prague, 1999.

[5] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, USA, 1996.

[6] I. Šindelářová, J. Ptáček, and A. Procházka. Wavelet Use for Signal Denoising. In *The 13th International Conference on Process Control.* STU, Slovakia, 2001.

Mgr. Irena Šindelářová
University of Economics, Prague
Department of Econometrics
W. Churchill Sq. 4, 130 00 Prague 3
Phone: 00420-2240 95 443, Fax: 00420-2240 95 423
E-mail: isin@vse.cz

Prof. Aleš Procházka, Ing. Martina Mudrová, Ph.D.
Institute of Chemical Technology, Prague
Department of Computing and Control Engineering
Technická 1905, 166 28 Prague 6
Phone: 00420-22435 4198, Fax: 00420-22435 5053
E-mail: A.Prochazka@ieee.org, Martina.Mudrova@vscht.cz