Analysis of Air Pressure Distributions with Wavelets

Andreas Baierl

Institut für Statistik, OR und Computerverfahren
Universität Wien, Austria

Abstract: The article constitutes an alternative approach to transform air pressure distributions for downscaling climate parameters. The objectives include a low-parametric approximation of original pressure surfaces and, in the following, the analysis of the dependence structure of temporal successive distributions for simulation purpose. Wavelets are applied to carry out the transformation of the distributions. Based on these results, time series analysis is used for the further investigations.

The underlying air pressure data covers the relevant region for the central European weather. Daily historical data is available for the time period between 1946 and 1994.

Zusammenfassung: Der vorliegende Artikel präsentiert grundlegende Methoden der diskreten Waveletanalyse und eine spezielle Anwendung, Wavelets als Instrument der Datenanalyse und -reduktion auf dem Gebiet der Klimaforschung. Konkret werden Wavelets zur Approximation von Luftdruckverteilungen eingesetzt, vorrangig um die Anzahl der zur Beschreibung notwendigen Parameter effizient zu reduzieren. In Verbindung mit Temperatur- und Niederschlagsdaten können die approximierten Druckoberflächen zur Vorhersage und Simulation von Abflußmengen in einem ausgewählten Flußeinzugsgebiet verwendet werden. Zu Simulationszwecken ist weiters eine umfassende Untersuchung der Abhängigkeitsstruktur aufeinanderfolgender Luftdruckoberflächen mittels zeitreihenanalytischer Methoden erforderlich. Tägliche Aufzeichnungen der historischen Luftdruckverteilungen des für Mitteleuropa relevanten Gebietes stehen für den Zeitraum von 1946 bis 1994 zur Verfügung.

Keywords: Wavelets, Air Pressure Distributions, Downscaling, Time Series.

1 Introduction

The analysis of pressure distributions discussed in this article forms a part of a large hydrological and climatological model. The development is commissioned by the European Union under the title *CCHydro, Impact of climate change of river basin hydrology under different climate conditions* with participants from Austria, Germany, Greece and Italy.

The objectives of the project contain the prediction of *daily temperature* and *precipitation distributions* for the area of a river basin at *high resolution* under variable climate conditions (e.g. CO_2 level in the atmosphere). Based on these results, water off-flow models are estimated.

The major problem turns out to be the detailed prediction for a local region. Relevant data for the estimation comes from detailed historical observations and from appropriate climate simulation models, so called *global circulation models* (GCMs). The distributions

generated with GCMs cannot be directly applied to predict the corresponding temperature and precipitation, because of too low resolution.

The *downscaling* procedure connects simulated and historical temperature and precipitation data via corresponding pressure distributions that are available in both data sets at sufficient resolution. Given the pressure distribution, conditional temperature and precipitation distributions are estimated using historical data. In the following, the established model is applied to simulated pressure surfaces, and the procedure returns a detailed prediction of local climate variables.

The relevant pressure distributions in this specific case are observed at 336 points, and substantial spatio-temporal correlations of successive distributions are expected. Hence, in order to use pressure data as input for downscaling, the original surfaces have to be sufficiently approximated with a reasonable number of parameters. Researchers working at the project recommend a maximum of 30 parameters. Additionally, the dependence structure of successive distributions has to be modelled.

1.1 Methical Approach

CP-type classification constitutes a common method to solve the problem (see Matyasovszky et al., 1994, or Nachtnebel et al., 1995): Applying cluster analysis or similar explorative methods, a certain number of weather types, so called CP-types (usually a set of 20 to 40 different ones for the annual cycle), are derived from the high-dimensional air pressure data. They form the input of the downscaling procedure.

Shortcomings of this method include the loss of relevant information about the shape of the surface, especially if the nature of pressure data does not suggest classification (no typical patterns). Furthermore, the temporal course of pressure distributions, which means to estimate transmission probabilities and to take persistence behavior into account, proves difficult to model.

In this article, *wavelet analysis* is presented as an alternative approach to approximate pressure distributions giving a compressed functional description of the surface. This method is expected to handle the lacks of CP-type classifications in a better way.

2 Wavelets - Theory & Application

2.1 Introduction

The fundamental idea of wavelet analysis is to approximate an observed signal f by superposition of translated (varying positions) and dilated (varying scales) versions of a wavelet ψ.

Generally we define two procedures in wavelet analysis:

- *decomposition:* transformation of an observed signal f into a set of wavelet coefficients c_i, $i = 1, 2, ..., n$

- *reconstruction:* representation of f, using a set of wavelet coefficients and a specified wavelet ψ.

Applications of wavelet analysis include data reduction, for example image compression, as well as data analysis, e.g. denoising data, sound synthesis, time-series analysis and signal processing.

2.2   The Analysis of a Signal - Decomposition

The idea of wavelet decomposition is to use different filters to dissect an observed signal f into low- and high-frequency parts.

In practice, a *discrete* sequence of observations s_k of the signal is available.

$$s_k = s(k\,h), \; k \in \mathbf{Z}.$$

In the following, projections of the signal into subspaces $V_m, m = 0, 1, 2, ...$ (low frequencies) and $W_m, m = 0, 1, 2, ...$ (high frequencies) are derived using appropriate *orthogonal* sets of *scaling* functions $\varphi_{m,k} = 2^{-\frac{m}{2}} \varphi\left(2^{-m}x - k\right), k \in \mathbf{Z}$ and *wavelet* functions $\psi_{m,k} = 2^{-\frac{m}{2}} \psi\left(2^{-m}x - k\right), k \in \mathbf{Z}$. The decomposition algorithm separates the high-frequency parts *(details)* at each level m and proceeds with the remaining low-frequency representation *(approximation)*. Hence, the following equation holds:

$$V_0 = V_m \oplus W_m \oplus ... \oplus W_1. \tag{1}$$

Because of orthogonality of shifted scaling and wavelet functions, the scaling and wavelet *coefficients* can be derived independently in the following way:

$$c_k^m\left(\varphi, f\right) = \langle f, \varphi_{m,k} \rangle \tag{2}$$

$$c_k^m\left(\psi, f\right) = \langle f, \psi_{m,k} \rangle. \tag{3}$$

The decomposition level gives the degree of the approximation and has to be chosen according to the underlying problem.

2.3   Reconstruction

An increasing decomposition level m results in a decreasing number of non-zero wavelet and scaling coefficients that, multiplied with the corresponding wavelet and scaling functions, reconstruct the original signal f.

$$f_m^\varphi = P_m f = \sum_{k \in \mathbf{Z}} c_k^m(\varphi, f)\varphi_{m,k} \tag{4}$$

$$f_m^\psi = Q_m f = \sum_{k \in \mathbf{Z}} c_k^m(\psi, f)\psi_{m,k}. \tag{5}$$

The information (details or approximations) used to reconstruct the signal depends on the objective of the analysis. For the purpose of *data reduction,* the aim of this study, we intend a sufficient approximation of the original signal, while keeping m as large as possible.

2.4   Two-dimensional Wavelet Transformation

The two-dimensional wavelets are usually defined as tensor products of one-dimensional wavelets:

scaling function: $\varphi\left(x,y\right) = \varphi\left(x\right)\varphi\left(y\right)$

wavelet functions: $\psi_1 = \varphi\left(x\right)\psi\left(y\right)$, $\psi_2 = \psi\left(x\right)\varphi\left(y\right)$, $\psi_3 = \psi\left(x\right)\psi\left(y\right)$.

Hence equation (1) changes to

$$V_0 = V_m \oplus W_m^1 \oplus W_m^2 \oplus W_m^3 \oplus ... \oplus W_1^1 \oplus W_1^2 \oplus W_1^3. \tag{6}$$

2.5   Choosing a Wavelet

The capability of adaptation to various classes of problems forms one of the strengths of wavelets. This potential does exist due to the possibility of choosing from a wide variety of wavelet and corresponding scaling functions.

Discrete wavelet analysis requires a scaling function $\varphi \in L^2(\mathbf{R})$ that generates an orthonormal basis (ONB) of V_0. Generally, any function can be orthogonalized if its Fourier-transformed $\widehat{\varphi}$ is bounded:

$$A \leq \sum_{n \in \mathbf{Z}} \left|\widehat{\varphi}\left(\omega + 2\pi n\right)\right|^2 \leq B \text{ with A, B positve constants.} \tag{7}$$

Then every φ with non-vanishing first moment and

$$\widehat{\varphi}\left(2\pi k\right) = 0 \text{ for all } k \in \mathbf{Z}\backslash\left\{0\right\} \tag{8}$$

forms a candidate for discrete wavelet analysis (see Louis et al., 1994, pp. 110, 112).

If we allow a pair of scaling functions $(\varphi, \widetilde{\varphi})$, one function for decomposition and one for reconstruction, discrete wavelet analysis only requires *biorthogonality:*

$$\left\langle \varphi_{m,k}\widetilde{\varphi}_{m',k'} \right\rangle_{L^2} = \delta_{m,m'}\delta_{k,k'}. \tag{9}$$

From this large set of potential candidates, only certain wavelet families with specific convenient properties (zero moments, regularity, compact support) prove useful and are widely supported by statistical software.

In the following, *criteria for the selection* of an appropriate wavelet for the purpose of data reduction are discussed. Generally, the procedure requires the specification of two input arguments:

1. A certain pair of wavelet and scaling functions.

2. Decomposition level m - trade of between the number of parameters and degree of approximation.

Firstly the choice is restricted to wavelet families with convenient computational properties (existence of *fast algorithms*). A second criterion constitutes the *support length,* which largely affects the problem of *border distortions*. This means that all shifted versions of a chosen wavelet - still (partly) overlapping the support region of the surface

- must be included in order to correctly approximate the border regions of the surface. Naturally, a large support length causes a high number of adjusted wavelets not fully covering the surface and therefore decreases the efficiency of the analysis (in regard to data reduction).

Considering the group of potential wavelets satisfying the conditions above, the *shapes* of the wavelet is crucial for its capability to analyze or reconstruct a certain signal. In the case of reconstruction the shape of scaling function and signal (e.g. smooth, irregular) should match well. Apart from this quite intuitive selection criterion concerning the characteristic shape, the following distance value provides an objective measure for the correspondence of an original and reconstructed surface:

$$dist = \frac{1}{n} \sum_{i=1}^{n} \left((x_i - \overline{x}) - \left(\widehat{x}_i - \overline{\widehat{x}} \right) \right)^2 \tag{10}$$

$$\text{with } \overline{x} = \sum_{i=1}^{n} x_i \text{ and } \overline{\widehat{x}} = \sum_{i=1}^{n} \widehat{x}_i.$$

Finally, the goodness or sufficiency of the resulting approximation must be evaluated in relation to the objectives of a specific analysis.

2.6 Comparison of Fourier and Wavelet Transformation

Windowed Fourier analysis proves quite similar to wavelet analysis, where the latter does not restrict the set of functions used to analyze a signal is to the harmonic and exponential functions. It depends on the characteristics of the actual signal if it can be sufficiently described by Fourier analysis. Wavelets have advantages especially in analyzing physical situations, where the signal contains discontinuities and spikes. However, the wavelets stay competitive, even in contexts considered favorable for the Fourier technique.

2.7 Application

According to the subject of the study, wavelet analysis is applied to air pressure surfaces. The calculations are carried out with the software package MATLAB Toolbox Wavelets. The surface with rectangular support contains 21 times 16 observations. Wavelet analysis aims to approximate the original data catching the relevant and decisive characteristics for the prediction of temperature and precipitation distributions.

The selection of a specific wavelet is guided by criteria given in Section 2.5: All pairs of wavelet and scaling functions that allow discrete wavelet analysis and for which MATLAB provides fast algorithms are considered. Given these functions, we rule out those with support widths larger than 5, as a length of 5 means a maximum data reduction to 5 times 5 wavelet coefficients.

According to the distance measure, two candidates, the *biorthogonal wavelet of order 3.1* and the *symlet of order 2*, perform better compared to all other potential choices. Both have support length of 3, and therefore use 20 wavelets at decomposition level 2, and 42 at level 3 respectively to approximate the surface. Considering their shapes, the biorthogonal wavelet appears to reproduce the original data in a better way.
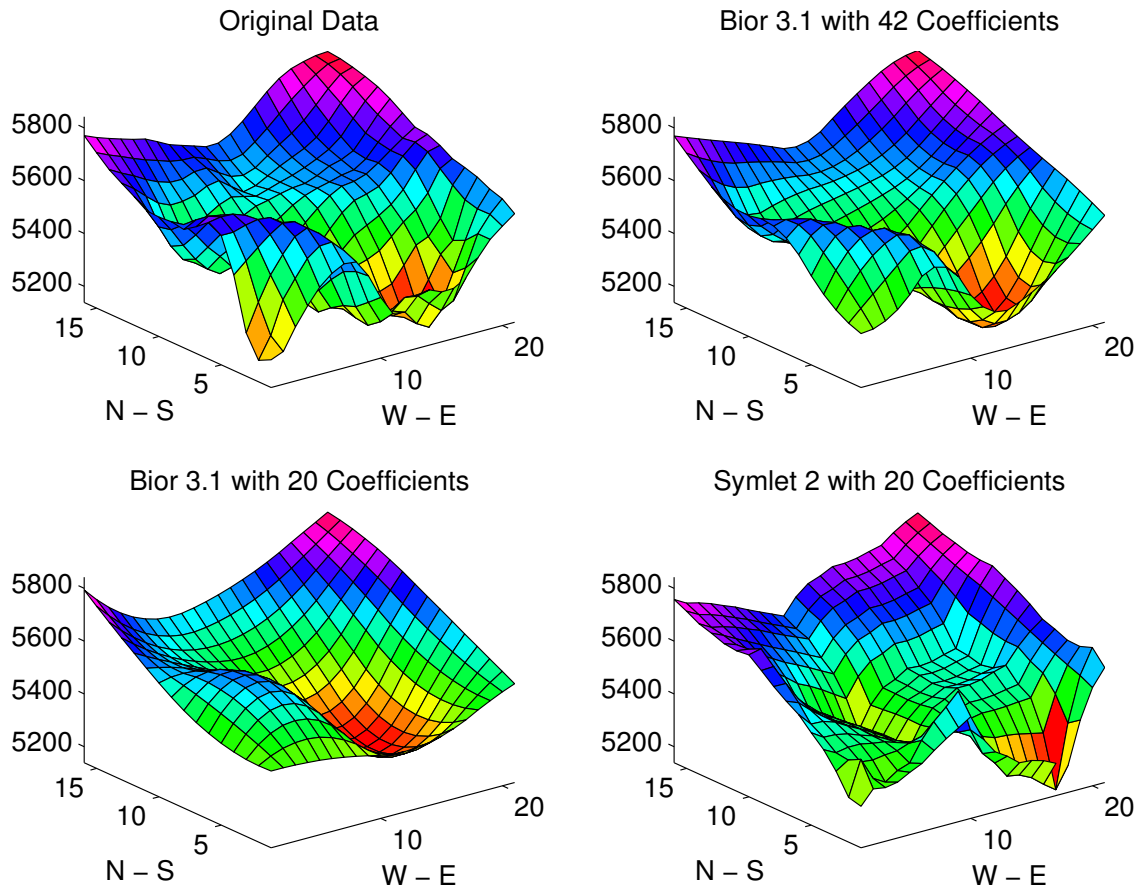
Figure 1: Original and reconstructed pressure surfaces. The z-axis denotes the height in meters of the 500 hPa pressure surface. The orientation of x- and y-axis is given in the plot.

The goodness of approximation always has to be evaluated in regard to the capability of predicting temperature and precipitation. This can also affect the restriction of using a maximum of 30 parameters. As this work has not been carried out yet, an optimal wavelet cannot be determined at this point.

3 Time Series Analysis of Wavelet Coefficients

The input of time series analysis consists of the output of the wavelet transformed pressure distributions. For the subsequent analysis all historical surfaces (2 records per day from 1946 to 1994) are transformed adjusting the biorthogonal wavelet of order 3.1 at decomposition level 3. Hence, we obtain 20 time series of length 35770 of equally spaced observations c_{it}: coefficient $i = 1, ..., 20$ at time point $t = 1, ..., 35770$.

The objective of the analysis consists of estimating a spatio-temporal distribution of pressure surfaces; i.e. to model the dependence structure of successive observations based on the wavelet transformed version. This provides a base to estimate transmission probabilities for simulation purpose and to investigate the dependence of local climate param-

eters on sequences of pressure distributions.

Examining the dependence structure of series of pressure surfaces, three different kinds of correlations are expected.

1. *temporal*: Given a fixed i equal to $\bar{i}$, the coefficients $c_{\bar{i}t}, t = 1, ..., 35770$, are temporally correlated.

2. *spatial*: For a fixed time point $\bar{t}$, the coefficients $c_{i\bar{t}}, i = 1, ..., 20$, are spatially correlated.

3. *spatio-temporal*: Correlations of coefficients c_{it} and $c_{i't'}$ for $i \neq i'$ and $t \neq t'$.

Concerning this article, only case 1, the analysis of the temporal dependence structure of the separated time series $\{c_{t1}\}, ..., \{c_{t20}\}$ is carried out.

3.1 Results

Observing graphs of all 20 time series, the plots show strong seasonality with a yearly cycle and weak trends. A non-parametric decomposition algorithm, namely STL (see Cleveland et al., 1990), is applied in order to filter a seasonal component identical for each yearly cycle, and to separate a trend component.

Analyzing autocorrelation functions of the remainders of the decomposed time series, substantial correlations are observed showing different characteristics for coefficients of the northern and southern area of the surface. Appropriate time series models are estimated for all 20 time series; namely ARMA(2,2) models for the series of coefficients located in the southern part of the support region and ARFIMA(2,d,2) models for time series lying in the north of the surface (see Figure 2). ARFIMA stands for fractionally-integrated ARMA models where the parameter d takes values between 0 and 0.5, and is determined by the algorithm (see Schlittgen and Streitberg, 1994, for further discussion). The time series analysis was carried out with the software package S-PLUS.

4 Conclusions and Outlook

The studies presented in the article at issue provide a low-parametric representation of pressure surfaces and describe parts of the spatio-temporal behavior (dependence structure). In further work this can be used to estimate conditional temperature and precipitation distributions (e.g. applying multiple regression models). In order to enable a simulation of successive pressure distributions and to detect changes in their temporal behavior, additional analysis is required. Namely, the complete spatio-temporal correlation structure has to be added to the model.
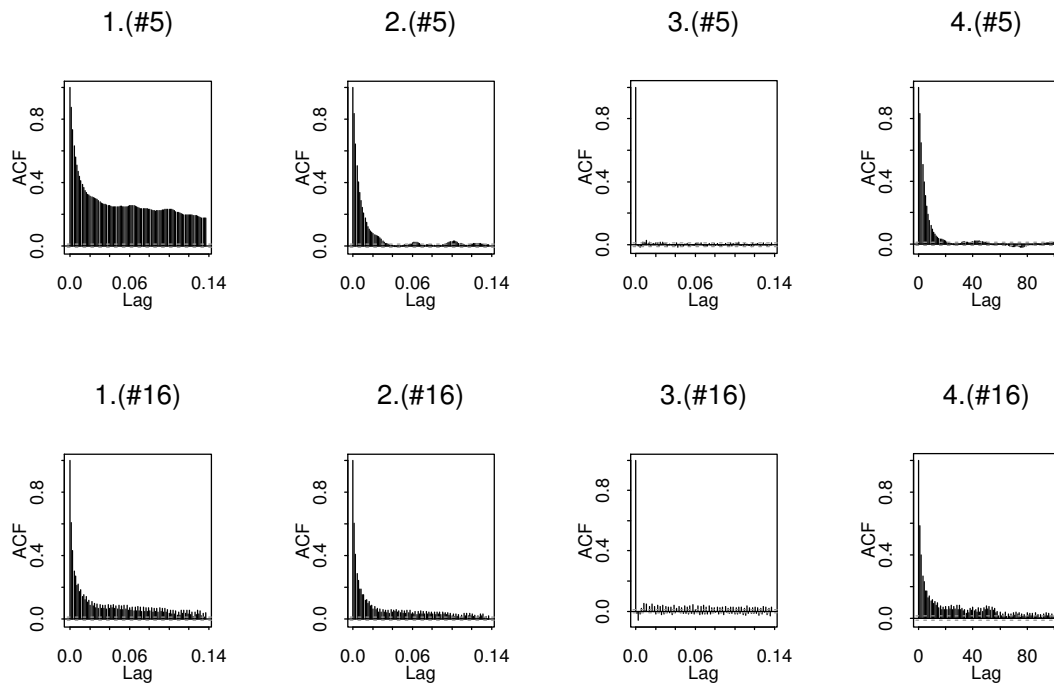
Figure 2: Autocorrelation functions (ACF) for time series #5 (located in the SW) and #16 (located in the NE). The first column shows the ACF of the original data; in the second step, the seasonal and trend component have been subtracted; the third column gives ACFs of the residuals of ARMA(2,2) models applied to both time series. The model fits the southern time series clearly better than the coefficient located in the north. The last column shows a simulation of the corresponding ARMA process for time series #5 and a simulated ARFIMA process for time series #16, which reproduces the ACF of the northern time series better than the ARMA process.

References

A. Baierl. *Analysis of air pressure distributions with wavelets*. Master's thesis, University of Vienna, 1997.

A. Bárdossy. *Modelle zur Abschätzung der regionalen hydrologischen Folgen einer Klimaänderung*. Institut für Hydrologie und Wasserwirtschaft, Karlsruhe, 1994.

A. Brian and B. Moshe. *Matlab für Ingenieure: systematische und praktische Einführung*. Addison-Wesley, Bonn, 1995.

P.J. Brockwell and R.A. Davis. *Time series: Theory and methods*. Springer Verlag, New York, 1991.

R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpening. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* **6**, 3-73, 1990.

P.J. Diggle. *Time series; a biostatistical introduction.* Oxford University Press, New York, 1990.

A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, **2**, Los Alamitos, CA, 1995.

A.K. Louis, P. Maaß, and A. Rieder. *Wavelets: Theorie und Anwendung.* Teubner, Stuttgart, 1994.

I. Matyasovszky, I. Bogardi, A. Bárdossy, and L. Duckstein. Space-time precipitation reflecting climate change. *Hydrological Science Journal* **38**, 539-558, 1993.

I. Matyasovszky, I. Bogardi, A. Bárdossy, and L. Duckstein. Local temperature estimation under climate change. *Theor. App. Climatology* **50**, 1-13, 1994.

M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi. *Wavelet Toolbox User's Guide.* The MathWorks, Inc., Natick, Mass, 1996.

H.P. Nachtnebel, K. Hebenstreit, I. Bogardi, and I. Matyasovszky. Auswirkungen großräumiger Klimaänderungen auf die Hydrologie eines alpinen Einzugsgebietes. *CCHYDRO Project*, 1995.

R. Schlittgen and B.H. Streitberg. *Zeitreihenanalyse.* Oldenburg, Wien, 1994.

Author's address:

Mag. Andreas Baierl
Hofzeile 11/3/11 A-1190 Vienna
Austria

E-mail: a9226277@unet.univie.ac.at