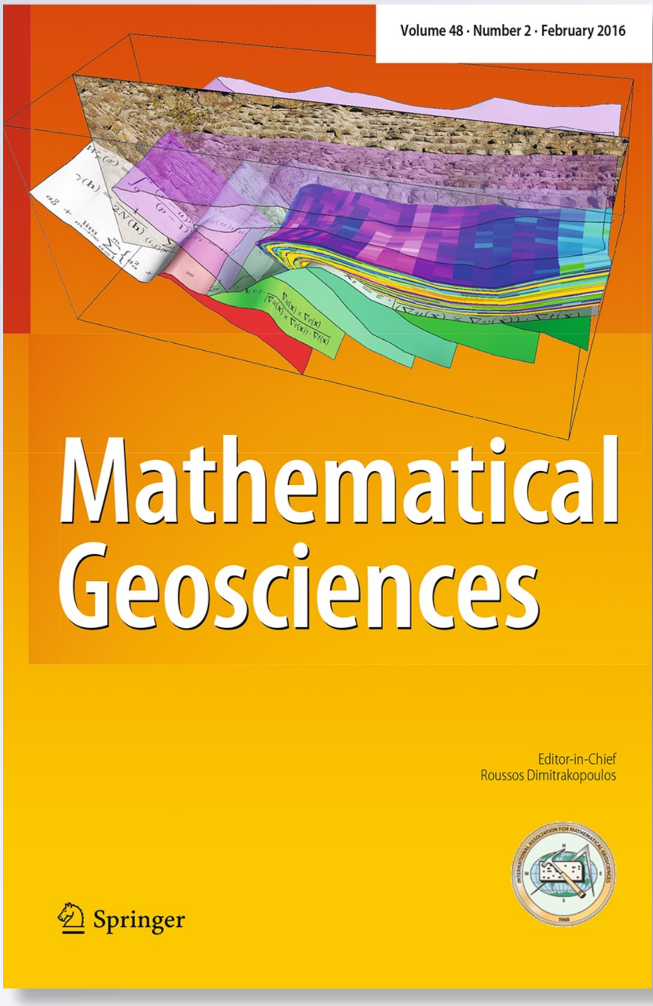
Wavelet-Based Clustering of Sea Level Records

S. M. Barbosa, S. Gouveia, M. G. Scotto & A. M. Alonso

Mathematical Geosciences

ISSN 1874-8961
Volume 48
Number 2

Math Geosci (2016) 48:149-162
DOI 10.1007/s11004-015-9623-9



Springer

Math Geosci (2016) 48:149–162
DOI 10.1007/s11004-015-9623-9

Wavelet-Based Clustering of Sea Level Records

S. M. Barbosa[1] · **S. Gouveia**[2] · **M. G. Scotto**[3] · **A. M. Alonso**[4]

Received: 6 April 2015 / Accepted: 26 October 2015 / Published online: 27 November 2015

Abstract The classification of multivariate time series in terms of their corresponding temporal dependence patterns is a common problem in geosciences, particularly for large datasets resulting from environmental monitoring networks. Here a wavelet-based clustering approach is applied to sea level and atmospheric pressure time series at tide gauge locations in the Baltic Sea. The resulting dendrogram discriminates three spatially-coherent groups of stations separating the southernmost tide gauges, reflecting mainly high-frequency variability driven by zonal wind, from the middle-basin stations and the northernmost stations dominated by lower-frequency variability and the response to atmospheric pressure.

Keywords Wavelets · Clustering · Sea level · Time series

1 Introduction

The monitoring of environmental parameters is often performed on a spatially irregular network of stations yielding large datasets of multivariate time series. The classification of such time series in terms of the characteristics of the corresponding temporal patterns

✉ S. M. Barbosa
susana.a.barbosa@inesctec.pt

1 INESC TEC-INESC Technology and Science, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

2 Institute of Electronics and Informatics Engineering of Aveiro (IEETA) and CIDMA, Universidade de Aveiro, Aveiro, Portugal

3 CEMAT, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

4 Department of Statistics and Instituto Flores de Lemus, Universidad Carlos III de Madrid, Madrid, Spain

is a common goal to summarise the information and identify relations among the different series.

Several clustering approaches have been developed for the classification of time series of a single variable (Alonso et al. 2006; Caiado et al. 2015; Díaz and Vilar 2010; Liao 2005; Liu et al. 2014; Scotto et al. 2009, 2010, 2011). However, monitoring stations often measure several different variables and clustering methods for the classification of multivariate time series have been recently developed (D'Urso and Maharaj 2012; D'Urso et al. 2014; Maharaj et al. 2010). The present study addresses the bivariate classification of sea level and sea level atmospheric pressure time series from the Baltic Sea using a wavelet-based clustering approach introduced by D'Urso et al. (2014). Their ideas will be extensively used throughout this work.

Sea level is an integrated environmental parameter reflecting both the state of the ocean and of the atmosphere above. Because of the multiple factors influencing the height of the sea surface, including atmospheric pressure, winds, water temperature and salinity, sea level displays variability on a wide range of spatial and temporal scales. Therefore, a scale-by-scale approach based on the discrete wavelet transform is particularly appealing for the analysis of sea level records (Barbosa et al. 2007; Bastos et al. 2013; Percival and Mojfeld 1997).

Atmospheric pressure also displays variability on multiple time scales and is amenable to wavelet-based analysis (Barbosa et al. 2009). The atmospheric pressure is responsible for a downward force acting on the sea surface which is compensated, to some extent, by corresponding sea-level variations. The static response of the sea surface can be modelled linearly by the inverse barometer model, a decrease (increase) in atmospheric pressure of 1 mb raising (depressing) sea level by 1 cm. However, ocean dynamics and wind effects introduce substantial deviations to the hydrostatic response of the sea surface to atmospheric pressure loading. Thus the association between sea level and atmospheric pressure is usually scale-dependent in time and variable from site to site. Furthermore, the inverse barometer law fails in the semi-enclosed Baltic Sea since the narrow straits connecting the Baltic to the North Sea prevent the fast response of the sea surface to local changes in atmospheric pressure (Kulikov and Medvedev 2013).

In this work the information from both sea level and atmospheric pressure time series is used to classify tide gauge records from the Baltic Sea. The classification of sea level time series allows to summarise the monitoring information in terms of common variability features of the records and to identify sub-regions with similar properties. The wavelet-based clustering approach allows to significantly reduce the dimensionality of the classification problem by reducing a large number of observations to a small number of wavelet coefficients while preserving the information on the temporal structure of the time series. Furthermore, the multivariate setting allows to include additional useful information into the clustering procedure by considering the interaction between pairs of components of the multivariate time series at each scale.

The rest of the paper is organized as follows: the wavelet-based clustering approach is described in Sect. 2. The sea level and pressure time series used in the application are presented in Sect. 3. The results are displayed in Sect. 4 and discussed in Sect. 5. Concluding remarks are provided in Sect. 6.

2 Methods

Section 2.1 provides an outline of discrete wavelet analysis and maximal overlap discrete wavelet transform (MODWT) useful in the present setting, referring the reader to Percival and Walden (2000), Percival (2008) and the references therein for a more detailed description. Furthermore, Sect. 2.2 describes the wavelet decomposition-based clustering approach used in this study to classify bivariate time series of sea level and atmospheric pressure. The purpose of this analysis was twofold: first, wavelet decomposition is applied to the time series to identify the most relevant scales in what concerns to variability and also joint variability for each tide gauge station. Second, based on such features the wavelet decomposition-based clustering method is applied for grouping stations with similar profiles.

2.1 Wavelet-Based Decomposition

The MODWT is a linear filtering operation which acts as the base of both a variance and an additive decomposition of a given time series $(X_t;\ t = 0, \ldots, T-1)$.

Let $(\delta_{j,l};\ l = 0, 1, \ldots, L_j - 1)$ be the MODWT wavelet filter of length L_j associated with the scale τ_j, where $L_j = (2^j - 1)(L-1) + 1$ and L is the width of the base filter (i.e., for $j = 1$). In addition assume that

$$W^X_{j,t} = \sum_{l=0}^{L_j-1} \delta_{j,l} X_{t-l},$$

represents the stochastic process by filtering the discrete parameter stochastic process X_t with the MODWT filter $\delta_{j,l}$. The time series X_t can be reexpressed as the sum of $J+1$ sub-series, that is $X_t = \sum_{j=1}^{J} D_j + S_J$, where the details $D_j;\ j = 1, \ldots, J$, derived from the wavelet coefficients $W^X_{j,t}$, correspond to the pass-band filtering scales $\tau_j = 2^{j-1}$ and the smooth S_J corresponds to the remaining parcel of the decomposition. The details D_j are associated with frequencies in the interval $\left[1/2^{j+1}, 1/2^j\right]$ and thus scale τ_j captures the dynamics over intervals with duration from 2^j to 2^{j+1} time units. Consequently, S_J includes information from all scales above 2^J time units.

The time independent MODWT wavelet variance at scale τ_j is defined as $\nu^2_X(\tau_j) := V(W^X_{j,t})$, provided that it exists and is finite. Thus

$$V(X_t) = \sum_{j=1}^{\infty} \nu^2_X(\tau_j),$$

which implies that the wavelet analysis decomposes the variance of (X_t) across wavelet scales. A similar decomposition can be obtained for the covariance between two stochastic processes X_t and Y_t with MODWT coefficients $W^X_{j,t}$ and $W^Y_{j,t}$, respectively, defined as

$$\mathrm{Cov}(X_t, Y_t) = \sum_{j=1}^{\infty} \mathrm{Cov}(W_{j,t}^X, W_{j,t}^Y) \equiv \sum_{j=1}^{\infty} \nu_{XY}(\tau_j).$$

In a bivariate framework, suppose that Z_t is a bivariate time series with components X_t and Y_t. The scale-by-scale wavelet variance/covariance quantification can be rearranged in the following symmetric matrix

$$\mathcal{C}_Z(\tau_j) := \begin{pmatrix} \nu_{X_Z}^2(\tau_j) & \nu_{X_Z Y_Z}(\tau_j) \\ \nu_{Y_Z X_Z}(\tau_j) & \nu_{Y_Z}^2(\tau_j) \end{pmatrix},$$

where X_Z and Y_Z represents the first and second component of the bivariate random vector Z, respectively.

After setting the base filter, $W_{j,t}^X$ can be straightforwardly computed by considering circular boundary conditions. Hence, $\mathcal{C}_Z(\tau_j)$ can be estimated through the unbiased empirical counterpart of its components, namely

$$\hat{\nu}_{X_Z}^2(\tau_j) := \frac{1}{M_j} \sum_{t=L_j-1}^{T-1} (\hat{W}_{j,t}^X)^2 \tag{1}$$

and

$$\hat{\nu}_{X_Z Y_Z}(\tau_j) \equiv \hat{\nu}_{Y_Z X_Z}(\tau_j) := \frac{1}{M_j} \sum_{t=L_j-1}^{T-1} \hat{W}_{j,t}^X \hat{W}_{j,t}^Y, \tag{2}$$

where $M_j = T - L_j + 1$, represents the number of wavelet coefficients excluding the boundary coefficients that are affected by the circular assumption of the wavelet filter.

The wavelet filter is selected to obtain an adequate variance decomposition and variance estimation across scales. In particular, the least asymmetric filter of width $L = 8$, i.e., LA(8), was adopted in this analysis since it yields coefficients that are approximately uncorrelated between scales while having a filter width short enough to keep the number of boundary coefficients small. Furthermore, LA filters exhibit approximately linear phase and thus allow to align the sinusoidal components in all scales with the original time series by time shift, for visualization purposes. Finally, the number of scales J is restricted by the length of the time series (T) and the filter width (L) through the inequality $J < \log_2(\frac{T}{L-1} + 1)$, which lead to $J \leq 10$ in the present work.

It is important to refer here that in this study the times series were analysed after the normalization X_t/s, where s represents the standard deviation of the original time series X_t. Since the MODWT allows partitioning the total variance of the original series by scale the variance of the normalized series associated with each scale corresponds to the percentage of X_t variance associated with such scale. The normalized series are also used to compute the wavelet covariance as measure of the association between the time series across scales.

2.2 Clustering of Bivariate Time Series

The clustering procedure builds a hierarchy from the individual elements by progressively merging more similar clusters, using an appropriate dissimilarity measure and a group linkage criterion (Everitt et al. 2011). Suppose that $Z_{1,t}, \ldots, Z_{n,t}$ are n bivariate time series. The $n \times n$ dissimilarity matrix, d_w, has entries $d_w(i, i'), i, i' = 1, \ldots, n$ corresponding to the pairwise comparison between $Z_{i,t}$ and $Z_{i',t}$. The comparison is based on their corresponding wavelet variance/covariance matrices, using the following distance measure proposed by D'Urso et al. (2014)

$$d_w(i, i') = \left\{\left(a_{WV} \cdot d_{wv}(i, i')\right)^2 + \left(a_{WC} \cdot d_{wc}(i, i')\right)^2\right\}^{\frac{1}{2}}, \tag{3}$$

where a_{WV} and a_{WC} are suitable non-negative and normalized weights for the wavelet variance-based distance and the wavelet covariance-based distance. Note that $d_{wv}(i, i')$ takes into account the differences in variance across scales for the objects i and i' as

$$d_{wv}(i, i') = \sum_{j=1}^{J} \|\mathrm{diag}(\mathcal{C}_{Z_i}(\tau_j)) - \mathrm{diag}(\mathcal{C}_{Z_i'}(\tau_j))\|, \tag{4}$$

where $\mathrm{diag}(\mathcal{A})$ denotes the principal diagonal of a matrix $\mathcal{A}$ and $\|\cdot\|$ represents the Euclidean norm. Moreover, the component

$$d_{wc}(i, i') = \sum_{j=1}^{J} \|\nu_{X_{Z_i} Y_{Z_i}}(\tau_j) - \nu_{X_{Z_{i'}} Y_{Z_{i'}}}(\tau_j)\| \tag{5}$$

quantifies the differences in wavelet covariances across scales. Distances $d_{wv}(i, i')$ and $d_{wc}(i, i')$ are estimated by replacing its components by their empirical counterparts from (1) and (2).

Finally, the clustering procedure involves obtaining a dendrogram based on the application of classical cluster techniques to the d_w matrix. In particular, unweighted average distance (average linkage), shortest distance (single) and furthest distance (complete) were considered for the group linkage criterion. The group linkage is chosen as to maximize the dendrogram's goodness-of-fit, evaluated through the cophenetic correlation coefficient between distances matrix d_w and distances represented in the cophenetic matrix (Everitt et al. 2011, p. 91). The closer the coefficient is to one, the more accurately the clustering procedure reflects the original data.

3 Data

Time series of daily-mean sea level (MSL) from tide gauges in the Baltic Sea are analysed for the period from January 1979 to December 2005. Daily data from the stations of Gedser and Hornbæk are provided by the Danish Meteorological Institute, DMI (Hansen 2007), the data for Furuögrund, Kungsholmsfort, Ölands Norra Udde

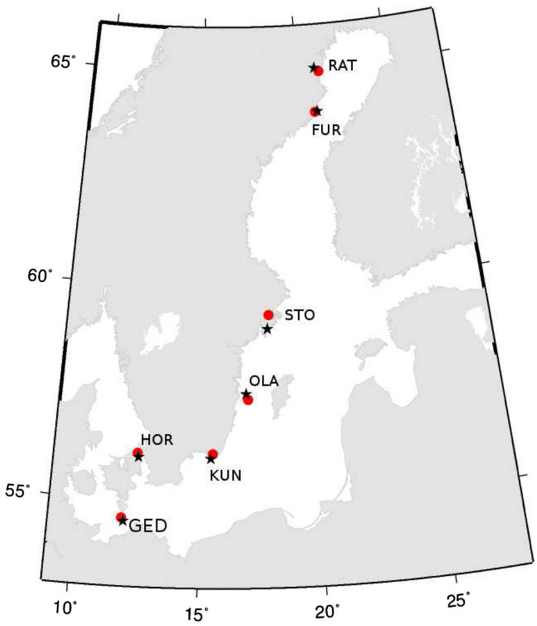


Fig. 1 Map of the Baltic Sea with location of tide gauges (*circle*) and reanalysis gridpoints (*star*)

Table 1 Analysed daily tide gauge records

Station name	Lon. (°E)	Lat. (°N)	Missing values (%)
Furuögrund (FUR)	21.23	64.92	0.04
Ratan (RAT)	20.90	63.99	0
Stockholm (STO)	18.08	59.32	3.7
Ölands Norra Udde (OLA)	17.10	57.37	0
Kungsholmsfort (KUN)	15.59	56.10	0
Hornbæk (HOR)	12.46	56.09	7.5
Gedser (GED)	11.93	54.57	4.8

and Ratan stations are provided by the Swedish Meteorological and Hydrological Institute, SMHI, and the data from Stockholm are provided by the University of Hawai Sea Level Center, UHSLC. Missing values in the tide gauge records are interpolated using the KNNinput method (Troyanskaya et al. 2001). Time series of daily-averaged sea-level pressure (SLP) are obtained from the ERA-interim reanalysis dataset (Dee et al. 2011). The analysed data are extracted from a 0.5° grid at the gridpoints closest to each tide gauge station (Fig. 1; Table 1).

As a pre-processing step all time series are linearly detrended and the annual seasonal cycle is removed. The analysed time series are shown in Fig. 2.

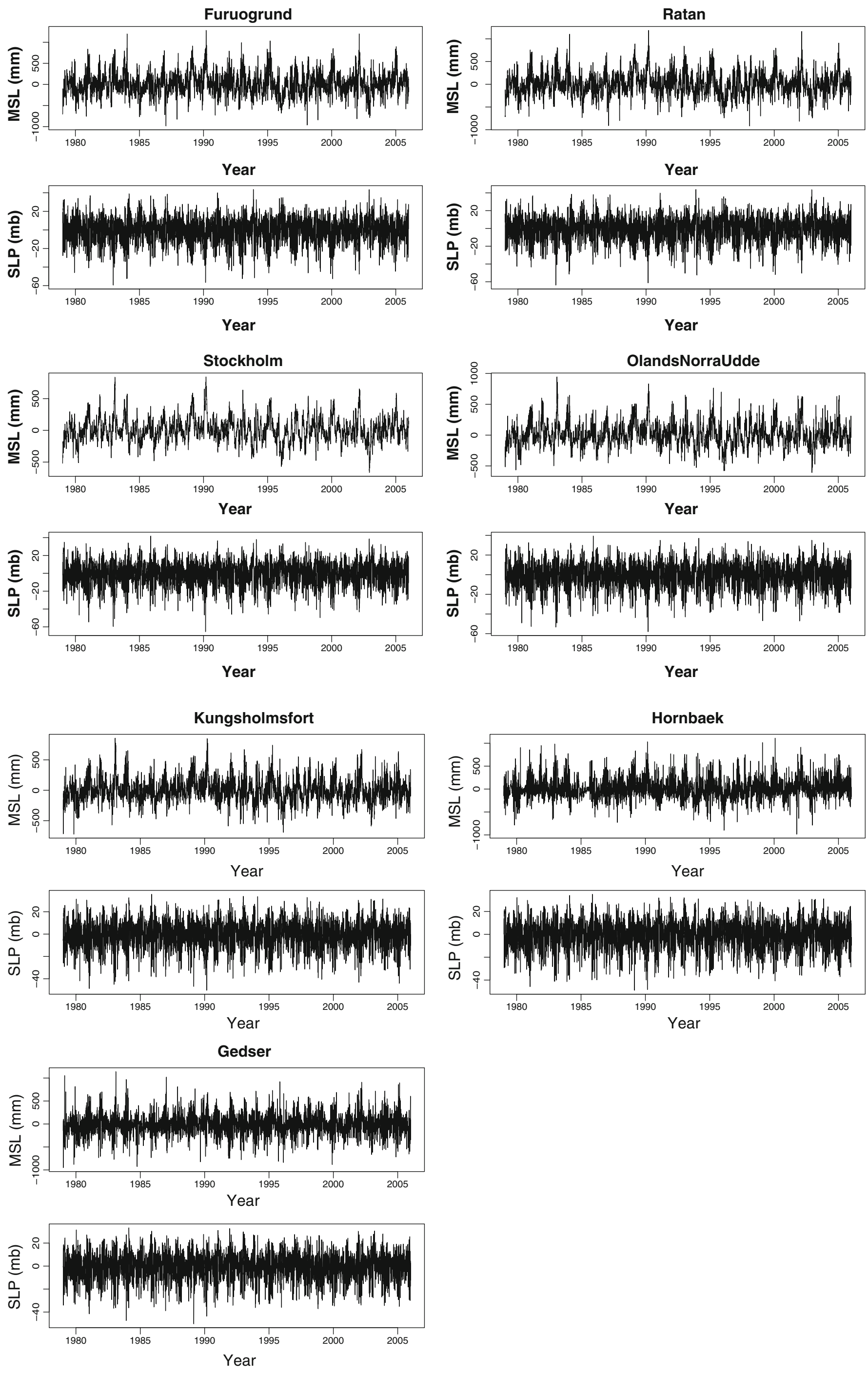


Fig. 2 Daily time series (1979–2005) of mean sea level (*MSL*) from Baltic tide gauges and sea-level pressure (*SLP*) from ERA-interim reanalysis

4 Results

4.1 Wavelet Decomposition

The scale-by-scale MODWT decomposition is the first step of the wavelet-based clustering procedure. The results are illustrated for the station Kungsholmsfort (KUN). Figures 3 and 4 show the MODWT decomposition of the MSL and SLP time series, respectively. The sub-series corresponding to a given wavelet scale j reflects variability on periods of 2^j to 2^{j+1} days, with $j = 1, \ldots, 10$. While for sea level the large scales significantly contribute to the series's variability, in the case of pressure the shorter scales (order of days) contribute significantly to the overall signal. The wavelet variance across scales is computed from each MODWT sub-series using Eq. (1) and the covariance is estimated from Eq. (2). The results are displayed in Fig. 5.

Sea level from the stations located at the Baltic entrance (GED and HOR) display a distinct temporal structure, with dominance of high-frequency short-scale

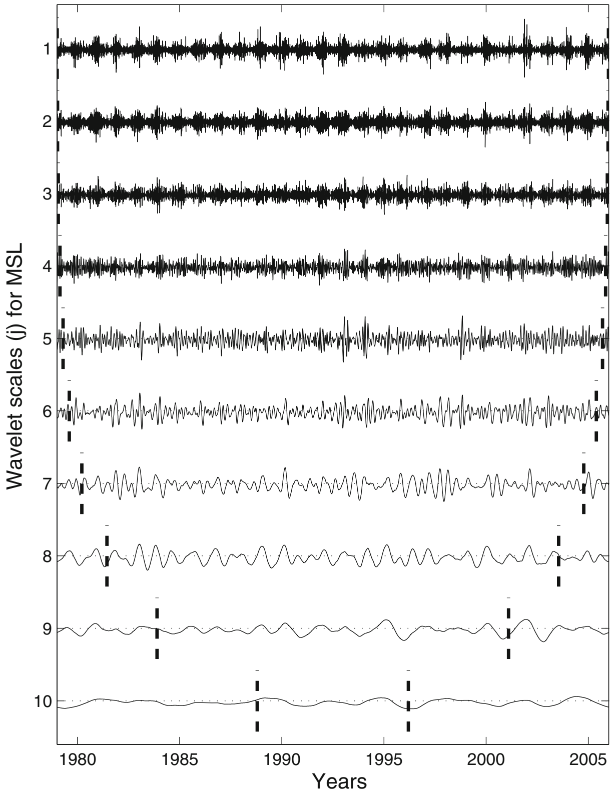


Fig. 3 MODWT decomposition for MSL time series in Kungsholmsfort (*KUN*) after phase shift for temporal alignment. *From top to bottom* components for wavelet scale $j = 1, \ldots, 10$. The *vertical dashed lines* delineate the boundary regions (wavelet coefficients outside of the lines are influenced to some degree by boundary conditions)

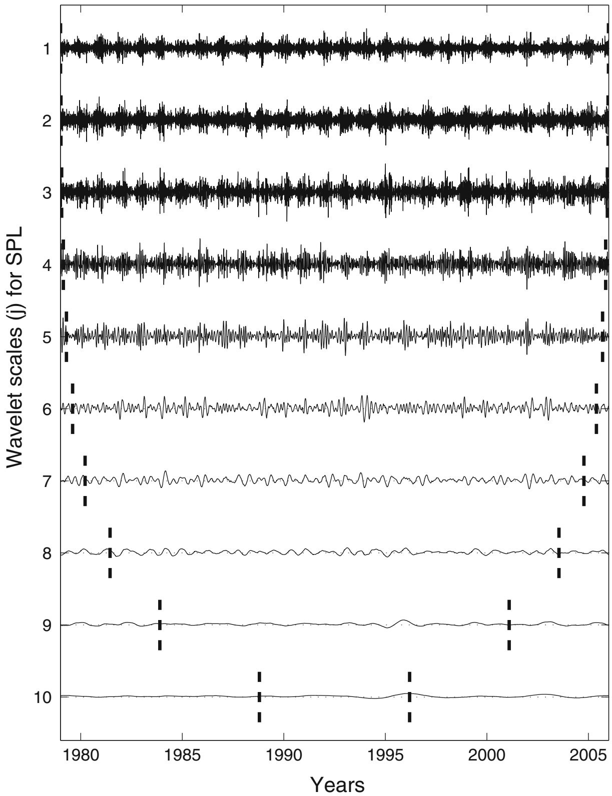


Fig. 4 As in Fig. 3 but for SLP time series in Kungsholmsfort (*KUN*)

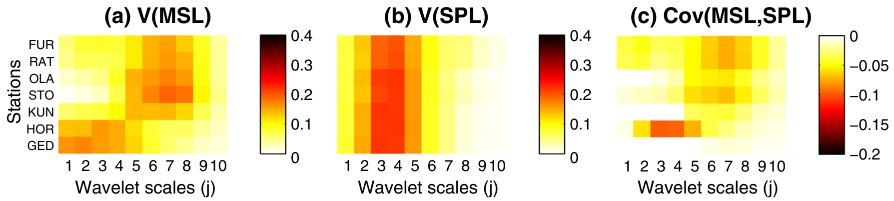


Fig. 5 Mosaic plot representing the contribution per wavelet scale to the total variance of **a** MSL, **b** SLP, and **c** covariance of MSL and SLP, for all stations (*darker colors* indicate higher absolute values)

components, in contrast with the remaining stations displaying higher contributions from lower-frequency signals (Fig. 5a). In the case of atmospheric pressure a more homogeneous picture emerges (Fig. 5b) with all stations exhibiting a similar temporal structure dominated by short range variability at wavelet scales 2–5. More than 40 % of the contribution to the overall variance of SLP series comes from scales 3 and 4 corresponding to periods of 8–32 days.

Figure 5c shows that for all stations the association of and SLP is not uniform, changing considerably with scale. Thus the scale-based picture is different from what

Table 2 Linear correlation between MSL and SLP time series and corresponding 95 % confidence interval assuming an effective sample size based on 1st order autocorrelations

Station	Correlation coefficient	95 % CI
FUR	−0.48	[−0.52, −0.45]
RAT	−0.47	[−0.50, −0.43]
STO	−0.38	[−0.42, −0.33]
OLA	−0.27	[−0.32, −0.22]
KUN	−0.13	[−0.17, −0.09]
HOR	−0.44	[−0.47, −0.41]
GED	+0.05	[0.021, 0.087]

emerges by just considering correlations between components of the bivariate time series (Table 2). The Gedser (GED) station displays virtually no correlation between SL and SLP, both for the whole series and across scales. For Hornbæk (HOR) the association of MSL and SLP is concentrated on a short range of scales about <30 days. For the remaining stations the co-variability of MSL and SLP is more important at large scales, typically of several months.

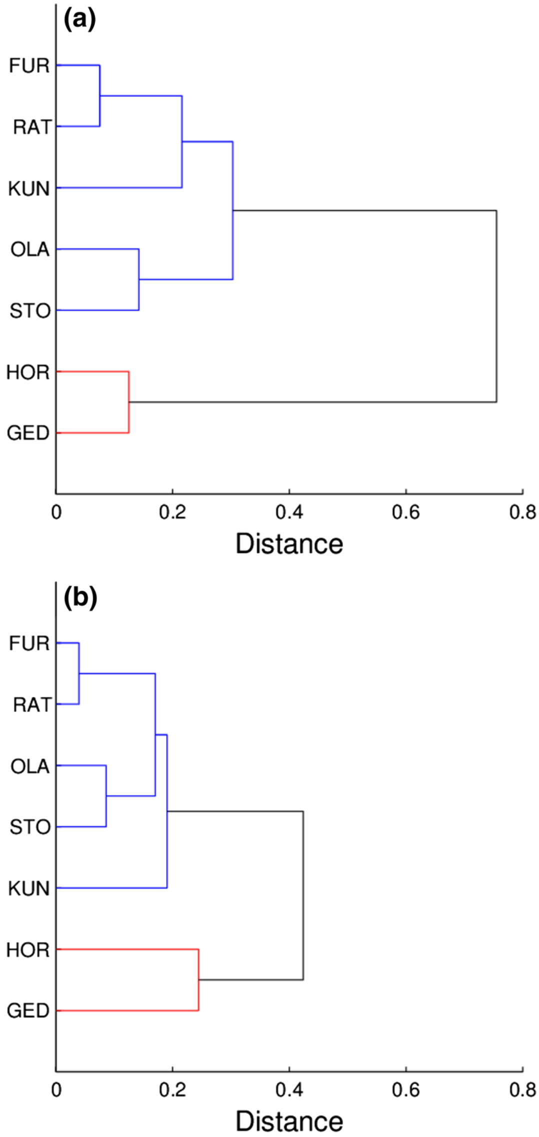
4.2 Clustering

The clustering procedure is based on the wavelet variance and covariance representations of the bivariate time series. For each individual time series the estimated wavelet variance at scale τ_j, $\hat{\nu}_X^2(\tau_j)$, is computed from the corresponding wavelet coefficients of the MODWT decomposition using Eq. (1). The wavelet variance-based distance $d_{wv}(i, i')$ can then be computed from Eq. (4) using the wavelet variances from MSL and SLP at stations i and i'. Likewise the wavelet covariance-based distance d_{wc} is computed using Eq. (5) from the wavelet covariance differences of MSL and SLP times across scales.

The dissimilarity matrix is then obtained from the weighted combination of d_{wv} and d_{wc} as in Eq. (3). Two distinct options are considered for the weights $a_{WV} = 1$ and $a_{WC} = 0$ (variance only) and $a_{WV} = 0.5$ and $a_{WC} = 0.5$ (equal weight of variance and covariance representations). Figure 6 shows the dendrograms obtained by applying the average linkage criterion to the dissimilarity matrix d_w with only variance-based distances (Fig. 6a) and with both variance and covariance distances (Fig. 6b).

Recall that the smaller the distance, the higher their similarity. The cophenetic correlation coefficient of the dendrograms is very similar, 0.92 and 0.90, respectively, indicating that the clustering is quite fit in both cases. Note that the dendrograms are very similar for the two sets of weights, the only difference being the relative position of station KUN (Kungsholmsfort). The value of the cophenetic coefficient for different covariance weights a_{WC} can be used to assess the choice of the coefficients of the wavelet-based distance measure. The cophenetic coefficient for weights $a_{WC} =$ 0, 0.25, 0.5, 0.75 and 1 is 0.92, 0.92, 0.90, 0.87 and 0.76, respectively, suggesting that equal weights ($a_{WC} = 0.5$) is a reasonable choice in the present application.

Fig. 6 Dendrogram of the hierarchical clustering of the stations (average linkage criterion): **a** variance-only; **b** variance + covariance



5 Discussion

The wavelet decomposition of MSL records (Figs. 3, 5) showed a distinct behaviour of the stations at the Baltic entrance, GED and HOR, dominated by high-frequency variability which is driven by zonal (east-west) winds. In contrast, the remaining stations are dominated by low-frequency contributions since fast variations are filtered out by the narrow and shallow Baltic entrance (Samuelsson and Stigebrandt 1996).

For pressure a similar behaviour across scales is obtained for all stations (Figs. 4, 5) reflecting the large-scale spatial structure of the pressure field and the dominance of short-term pressure fronts (Gustafsson and Andersson 2001).

Sea-level and atmospheric pressure tend to be anti-correlated reflecting the hydrostatic response of the sea surface to the weight of the atmosphere above it—higher/lower pressure corresponding to a larger/smaller downward force on the sea surface and thus a lower/higher sea-level. The atmospheric pressure also influences sea-level indirectly as a result of its impact on winds and ocean dynamics.

The wavelet-based dendrograms (Fig. 6) clearly discriminate the stations in the Baltic entrance from the remaining sea level records and groups together GED and HOR despite the absence of correlation of sea level with pressure at GED and the strong correlation at HOR.

Although the hierarchical clustering procedure yields apparently similar dendrograms, the dissimilarity measure taking into account both variance and covariance across scales produces a classification more consistent with the spatial setting of the stations. While the station KUN is placed closer to the group of the northernmost stations (FUR, RAT) in the variance-only classification, it occupies an intermediate position, more consistent with its geographical location, when both variance and covariance are taken into account. The southernmost stations (GED, HOR) belong to the same cluster in both cases. However, in the variance-only classification they are much closer to each other, and farther from the remaining group of stations, reflecting their similar variance structure across scales (Fig. 5a). In contrast, when covariance is also taken into account the distance between the two stations increases substantially reflecting the distinct covariance structure for the two stations (Fig. 5c).

6 Conclusions

The present work shows the application of a recently developed wavelet-based approach to the clustering of sea level records from the Baltic Sea. Taking into account both the variability of the multivariate time series across scales and the pairwise interaction at each scale allowed to obtain a classification more consistent with the regional setting. The clustering procedure evidenced two groups of stations with common variability features, discriminating the stations at the Baltic entrance (GED, HOR) from the remaining stations inside the Baltic. Furthermore, the intermediate station (KUN) was correctly assigned to the group of inward stations while being correctly placed farther from the stations in the central and northern Baltic area.

While the wavelet-based classification approach was applied here to sea-level records it is potentially useful to a wide range of geoscience applications dealing with multivariate time series such as the ones originated from environmental monitoring networks. In particular, simulation studies showed that the wavelet-based classification is able to adequately handle switching time series (Maharaj et al. 2010) making it particularly appealing, for example, for the classification of seismic data.

Acknowledgments Tide gauge data kindly provided by DMI (K. Madsen), SMHI (T. Hammarklint) and UHSLC. S.M. Barbosa acknowledges support of the FCT—Fundação para a Ciência e a Tecnologia (contract under programme IF2013 and project UID/EEA/50014/2013). This work was supported by the European Regional Development Fund (FEDER) through the COMPETE programme and by the Portuguese Government through the FCT, in the scope of the project UID/MAT/04106/2013 (Centro de I&D em Matemática e Aplicações, http://cidma.mat.ua.pt/) and projects PEst-OE/EEI/UI0127/2014 and UID/CEC/00127/2013 (Instituto de Engenharia Electrónica e Informática de Aveiro, IEETA/UA, http://www.ieeta.pt). S. Gouveia acknowledges the postdoctoral grant by FCT (ref. SFRH/BPD/87037/2012). A.M. Alonso acknowledges support of the Ministerio de Economía y Competitividad projects ECO2011-25706 and ECO2012-38442.

References

Alonso AM, Berrendero JR, Hernández A, Justel A (2006) Time series clustering based on forecast density. Comput Stat Data Anal 51:762–776

Barbosa SM, Fernandes MJ, Silva ME (2007) Scale-based comparison of sea level observations in the North Atlantic from satellite altimetry and tide gauges. In: Tregoning P, Rizos C (eds) Dynamic planet: monitoring and understanding a dynamic planet with geodetic and oceanographic tools. Springer, Berlin, Heidelberg, pp 63–66

Barbosa SM, Silva ME, Fernandes MJ (2009) Multi-scale variability patterns in NCEP/NCAR reanalysis sea-level pressure. Theor Appl Climatol 96:319–326

Bastos A, Trigo RM, Barbosa SM (2013) Discrete wavelet analysis of the influence of the North Atlantic Oscillation on Baltic sea level. Tellus A 65:20077

Caiado J, Maharaj EA, D'Urso P (2015) Time series clustering. In: Hennig CM, Meila M, Murtagh F, Rocci R (eds) Handbook of cluster analysis. Chapman and Hall/CRC, Boca Raton, Florida

Dee DP, Uppala SM, Simmons AJ et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597

Díaz SP, Vilar JA (2010) Comparing several parametric and nonparametric approaches to time series clustering: a simulation study. J Classif 27:333–362

D'Urso P, Maharaj EA (2012) Wavelets-based clustering of multivariate time series. Fuzzy Set Syst 193:33–61

D'Urso P, De Giovanni L, Maharaj EA, Massari R (2014) Wavelet-based self-organizing maps for classifying multivariate time series. Chemometrics 28:28–51

Everitt BS, Landau S, Leese M, Stahl D (2011) Cluster analysis. Wiley, West Sussex

Gustafsson BG, Andersson HC (2001) Modeling the exchange of the Baltic Sea from the meridional atmospheric pressure difference across the North Sea. J Geophys Res 106:19731–19744

Hansen L (2007) Hourly values of sea level observations from two stations in Denmark Hornbæk 1890–2005 and Gedser 1891–2005. DMI Technical Report No. 07-09, DMI

Kulikov EA, Medvedev IP (2013) Variability of the Baltic sea level and floods in the gulf of Finland. Oceanology 53:145–151

Liao TW (2005) Clustering of time series data—a survey. Pattern Recognit 38:1857–1874

Liu S, Maharaj EA, Inder B (2014) Polarization of forecast densities: a new approach to time series classification. Comput Stat Data Anal 70:345–361

Maharaj EA, D'Urso P, Galagedera DUA (2010) Wavelet-based fuzzy clustering of time series. J Classif 27:231–275

Percival D, Mojfeld H (1997) Analysis of subtidal coastal sea level fluctuations using wavelets. J Am Stat Assoc 92:868–880

Percival D, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge

Percival DB (2008) Analysis of geophysical time series using discrete wavelet transforms: an overview. In: Donner RV, Barbosa SM (eds) Nonlinear time series analysis in the geosciences—applications in climatology, geodynamics, and solar-terrestrial physics, pp 61–79. Springer, Berlin Heidelberg

Samuelsson M, Stigebrandt A (1996) Main characteristics of the long-term sea level variability in the Baltic sea. Tellus 48:672–683

Scotto MG, Barbosa SM, Alonso AM (2009) Model-based clustering of Baltic sea-level. Appl Ocean Res 31:4–11

Scotto MG, Alonso AM, Barbosa SM (2010) Clustering time series of sea levels: extreme value approach. J Waterw Port Coast Ocean Eng 136:215–225

Scotto MG, Barbosa SM, Alonso AM (2011) Model-based clustering of extreme sea level heights. In: Wright LL (ed) Sea level rise, coastal engineering, shorelines and tides, pp 277–293. Nova Science Publishers, New-York

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. Bioinformatics 17:520–525