Applied Ocean Research 31 (2009) 4–11

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

Model-based clustering of Baltic sea-level

M.G. Scotto [a,*], Susana M. Barbosa [b], Andrés M. Alonso [c]

[a] Departamento de Matemática, Universidade de Aveiro, Portugal

[b] Faculdade de Ciências, Universidade do Porto, Portugal

[c] Departamento de Estadística, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 19 June 2008
Received in revised form
10 March 2009
Accepted 11 March 2009
Available online 8 April 2009

Keywords:
Sea-level
Cluster analysis
Baltic Sea

ABSTRACT

Long (>30 years) monthly records of relative sea-level heights from tide gauges in the Baltic sea are analyzed. Time series clustering based on forecast densities is applied in order to describe regional sea-level variability in the Baltic Sea in terms of future relative heights. The tide gauge records are clustered on the basis of forecasts at 3-month and 6-month horizons. For the 3-month horizon, the results of the cluster analysis show a fairly spatial coherency in terms of grouping together locations from the same sub-basin, with the northern records in the Bothnian Sea and Gulf of Finland clustering together, followed by the tide gauges in the Baltic Proper and lastly the southernmost stations in the western Baltic. For the 6-month horizon, the results show a higher degree of homogeneity between different locations, but a clear separation between the stations at the Baltic entrance and the tide gauges inside the Baltic basin. Moreover, when considering detrended records, reflecting mainly the seasonal cycle, the clustering results are more homogeneous and suggest a distinct response of coastal sea-level in spring and in summer.

1. Introduction

The measurement of the sea-level has a long tradition in the Baltic Sea region, partly because of the ambition to study the post-glacial rebound of Fennoscandia [1,2]. Therefore, the Baltic Sea has one of the world's densest observational networks and a remarkable number of long and high-quality sea-level records from tide gauge measurements are available for geodetic and environmental studies. Hereafter, the term sea-level refers to relative sea-level heights.

The Baltic Sea is a shallow, semi-enclosed sea and one of the largest brackish water areas in the world. The Baltic is connected to the North Sea through shallow and narrow straits and includes three major sub-basins (the Baltic Proper, the Bothnian Sea and the Bothnian Bay), also separated by narrow and shallow passages that constrain the water exchange. The combined effect of limited exchange and high freshwater supply results in low salinity and a permanent salt-stratification, with surface salinity being largest in the Danish straits and decreasing northward. Due to this salinity gradient, the sea surface topography increases northward from the Kattegat to the Gulf of Finland and Bay of Bothnia [3,4]. Tides are negligible compared to other sea level variations and their influence decreases from the western Baltic areas to the Baltic Proper [5]. On time scales longer than one month, sea-level variations in the Baltic Sea are mainly driven by the exchange of water through the Danish Straits with maximum amplitudes in the North and minimum amplitudes at the Baltic entrance [6,7]. The water exchange between the Baltic Sea and the North Sea is mainly driven by the sea-level difference between the Kattegat and the Baltic Proper [7], which in turn is influenced by the zonal wind over the Skagerrak and the North Sea via a direct inverse barometer response and by set-up of sea-level against coasts [8]. Thus sea-level long term variability is mainly governed by persistent winds over the North Sea and the Baltic entrance driving sea water into or out of the Baltic, depending on the direction of the wind, the effect being larger at the innermost parts of the Baltic and smaller at the entrance [9]. Therefore the Baltic Sea is strongly influenced by North Atlantic atmospheric circulation patterns and long-term sea-level variability is correlated with the North Atlantic Oscillation (NAO) index (e.g. [10,11,8,12,13]), particularly at the central and northern tide gauge stations [14]—though a regional index is better for describing regional atmospheric conditions [8,15]. However, the association between the atmospheric pressure field and sea-level at Baltic tide gauges is both temporally and spatially heterogeneous [16] and other factors influence sea-level long-term variability, including precipitation and temperature [17,16]. Historical time series of relative sea-level measurements from the Baltic Sea constitute therefore a valuable record of regional climate and oceanographic variability.

* Corresponding address: Departamento de Matemática, Universidade de Aveiro, Campus de Santiago 3810-193, Aveiro, Portugal. Tel.: +351 234370359; fax: +351 234370683.

E-mail address: mscotto@ua.pt (M.G. Scotto).

0141-1187/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apor.2009.03.001

Since tide gauges measure sea-level relative to a vertical reference on land, long-term sea-level variability includes land movement effects, either due to uplift or subsidence of the Earth's crust. Most of the Baltic region experiences uplift due to postglacial rebound, which reaches a maximum in the northern Baltic area (∼9 mm/yr) and decreases southward to ∼3 mm/yr in the central Baltic. The southwestern Baltic region experiences subsidence on the order of ∼1 mm/yr [18].

In studies of regional sea-level variability, tide gauge records can be analyzed individually (e.g. [2,15,19,20]). An alternative approach is to consider simultaneously the whole dataset of sea-level records from a given region and characterize regional variability by means of techniques such as Empirical Orthogonal Functions (EOFs) (e.g. [21]) or Canonical Correlation Analysis (CCA), e.g. as applied by Heyen et al. [22] to identify atmospheric patterns responsible for variations in the modes of sea-level variability, and cross-correlation analysis(e.g. [23,24]). Cluster analysis is a useful approach for characterising regional variability in terms of locations exhibiting similar behavior. However, although clustering techniques have been popular for the analysis of (non-time series) environmental data, its extension to time series data is hindered by the serial dependence and high-dimensionality of the observations. Clustering time series, however, is a rapidly developing subject and it has been a topic of active research over the last few years, mainly due to its wide applicability to the analysis of environmental processes; see [25] for an analysis of historical data of CO_2 emissions in industrialized countries, [26] for a spatial clustering of daily ambient temperature, and [27] for an application to a rainfall problem. In contrast however, cluster analysis of sea-level measurements has not been applied yet. This paper aims to give a contribution towards this direction.

Other relevant references where cluster analysis is used to classify time series are Piccolo [28] and Maharaj [29,30] where time series are first modeled and then a distance is defined in terms of model parameters, and Kakizawa et al. [31] who characterize similarities among and differences between multivariate stationary time series in terms of the structure of the spectral matrices. More recently Vilar and Pértega [32] have considered the problem of clustering linear Gaussian processes and Fruhwirth-Schnatter and Kaufmann [33] proposed a Bayesian mixture model with a Markov Chain Monte Carlo (MCMC) algorithm for parameter estimation and data classification.

Clustering of time series can be based on cross-sectional information (ignoring the temporal evolution of the series), on models for the observations, or with respect to forecasts at a specific future time. The resulting groups will be different depending on the purpose, as illustrated in the diagram of Fig. 1, showing three simulated time series and the resulting three different solutions depending whether clustering is based on models, present information or forecasts at a future time. For instance, if we have the two following first autoregressive models, $X_{t+1} = \phi X_t + e_{t+1},\ t = 1, \ldots, N$ with e_t i.i.d. $\mathcal{N}(0, \sigma^2)$ and $Y_{t+1} = \phi Y_t + \upsilon_{t+1},\ t = 1, \ldots, N$ with υ_t i.i.d. $\mathcal{N}(0, \sigma^2)$, then X and Y will be in the same cluster when clustering by models is used. However, the closeness of their predictions will depend on the last observations, X_N and Y_N, since the first step ahead prediction densities are, in this case, $\mathcal{N}(\widehat{\phi}_X X_N, \widehat{\sigma}^2_X)$ and $\mathcal{N}(\widehat{\phi}_Y Y_N, \widehat{\sigma}^2_Y)$.

For the study of the regional variability of Baltic sea-level, the interest is on the temporal evolution of the records, therefore in this work the approach of Alonso et al. [25] is applied to group tide gauge locations in terms of sea-level forecasts at a future time. The novelty of this work is therefore the description of regional sea-level variability in the Baltic sea in terms of *future* relative sea-levels.

Future sea-level rise is an issue of considerable relevance in a climate change context. The Intergovernmental Panel on Climate Change (IPCC) AR4 estimates of sea level rise by 2100 are in the range of 18–59 cm [34], but these have been challenged on the basis that large ice sheets appear to be changing much more rapidly, and a recent study projects the sea level for the 2090–2099 period to be 0.9 to 1.3 m for the A1B scenario, with a low probability of the rise being within the IPCC confidence limits [35].

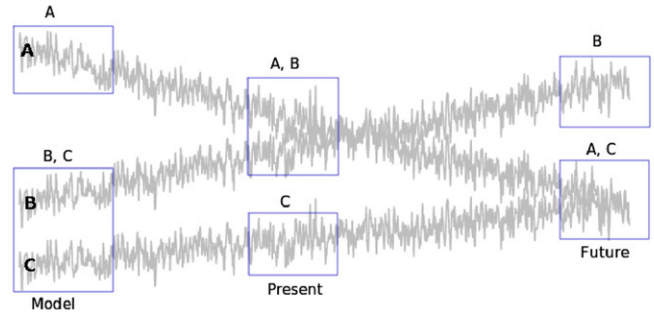

Fig. 1. Diagram showing three simulated time series (from top to bottom: A, B and C) and three different solutions depending whether clustering is based on models, present information or forecasts at a future time. Note: Future values are obtained using the models, but incorporate the information about the present values.

Projections of future sea-levels at the century scale can be obtained using physical models and considering future scenarios, rather than being estimated from present observations. On short time scales, sea-level prediction is very challenging. Most of the variability in Baltic tide gauge records on sub-monthly time-scales results from atmospheric-driven events, such as storm surges, that may last for several days. These events are dependent on the state of the North Atlantic Oscillation and on the atmospheric conditions over the North Atlantic, and are therefore difficult to predict. However, on monthly and longer time scales most of the variability in the level of the Baltic sea is determined by seasonal variations and long-term trends associated with climate variability, offering some possibilities of prediction. The estimation of future sea-levels is of highly practical relevance. Predictions at horizons of several months can be used as input in hydrodynamic, ecosystem and morphodynamic evolution models. Furthermore, the consequences of episodic events such as storm surges, which can have considerable practical impacts (including incidents of flooding, coastal erosion and damage to piers, buildings and coastal infrastructure) depend on the mean sea level.

The paper is organized as follows: the analyzed sea-level records from Baltic tide gauges are summarized in Section 2. The time series clustering approach is described in Section 3 and the results as groups of tide gauge locations extracted on the basis of forecasts at a specific future time are presented and discussed in Section 4. Finally, concluding remarks are given in Section 5.

2. Data

Long and continuous records of relative sea-level heights are obtained from the Permanent Service for Mean Sea Level (PSMSL) database of Revised Local Reference (RLR) tide gauge measurements [36]. A total of 14 stations (Fig. 2, Table 1) with records extending at least up to 2004 (clustering based on forecast densities requires time series data ending at the same time) are selected from the PSMSL dataset for the Baltic Sea.

The analyzed monthly time series are shown in Fig. 3. Except for the stations in the western Baltic (Wismar and Warnemünde) all records exhibit a decreasing trend resulting from postglacial rebound.

3. Model-based clustering time series

Let $\mathbf{X} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(T)})$ be a panel of T time series observed for, say $n_1, \ldots, n_T$, time units respectively. The approach pursued in this paper is based on the full forecast densities for each one

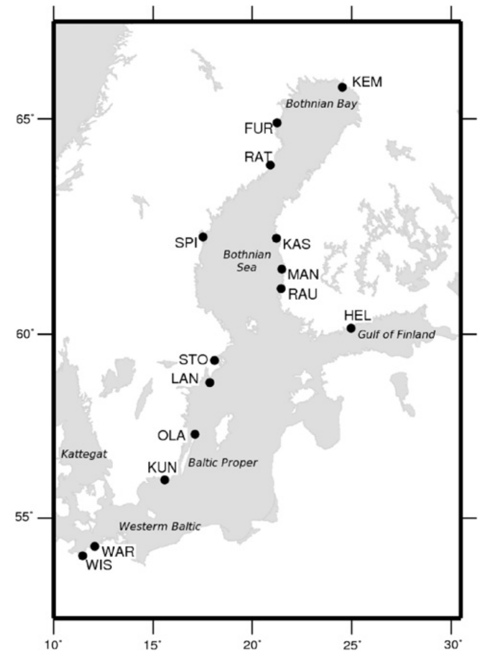

Fig. 2. Baltic Sea area and tide gauge locations. See Table 1 for further details.

Table 1
Analysed tide gauge records.

Tide gauge	Station code	Period	Longitude (°E)	Latitude (°N)
Furuögrund	FUR	Jan 1916–Dec 2004	21.23	64.92
Helsinki	HEL	Jan 1879–Dec 2004	24.97	60.15
Kaskinen	KAS	Jan 1927–Dec 2004	21.22	62.33
Kemi	KEM	Jan 1920–Dec 2004	24.52	65.67
Kungholmsfort	KUN	Jan 1887–Dec 2004	15.58	56.01
Landsort	LAN	Jan 1887–Dec 2004	17.87	58.75
Mäntyluoto	MAN	Jan 1911–Dec 2004	21.47	61.6
Ölands Norra Ude	OLA	Jan 1887–Dec 2004	17.01	57.37
Ratan	RAT	Jan 1892–Dec 2004	20.92	64.00
Rauma	RAU	Jan 1933–Dec 2004	21.43	61.13
Spikarna	SPI	Jan 1969–Dec 2004	17.53	62.37
Stockholm	STO	Jan 1889–Dec 2004	18.08	59.32
Warnemünde	WAR	Jan 1856–Dec 2004	12.08	54.18
Wismar	WIS	Jan 1849–Dec 2004	11.47	53.9

of the observed time series. Roughly speaking, a forecast density of a realization of a stochastic process at some future time is an estimate of the probability distribution of the possible future values of the process. It thus provides a complete description of the uncertainty associated with the prediction, and stands in contrast to a point forecast, which by itself contains no description of the associated uncertainty.

The implementation of the method for clustering time series is carried out in four stages: the algorithm starts defining a model for each one of the time series under analysis; next, based on

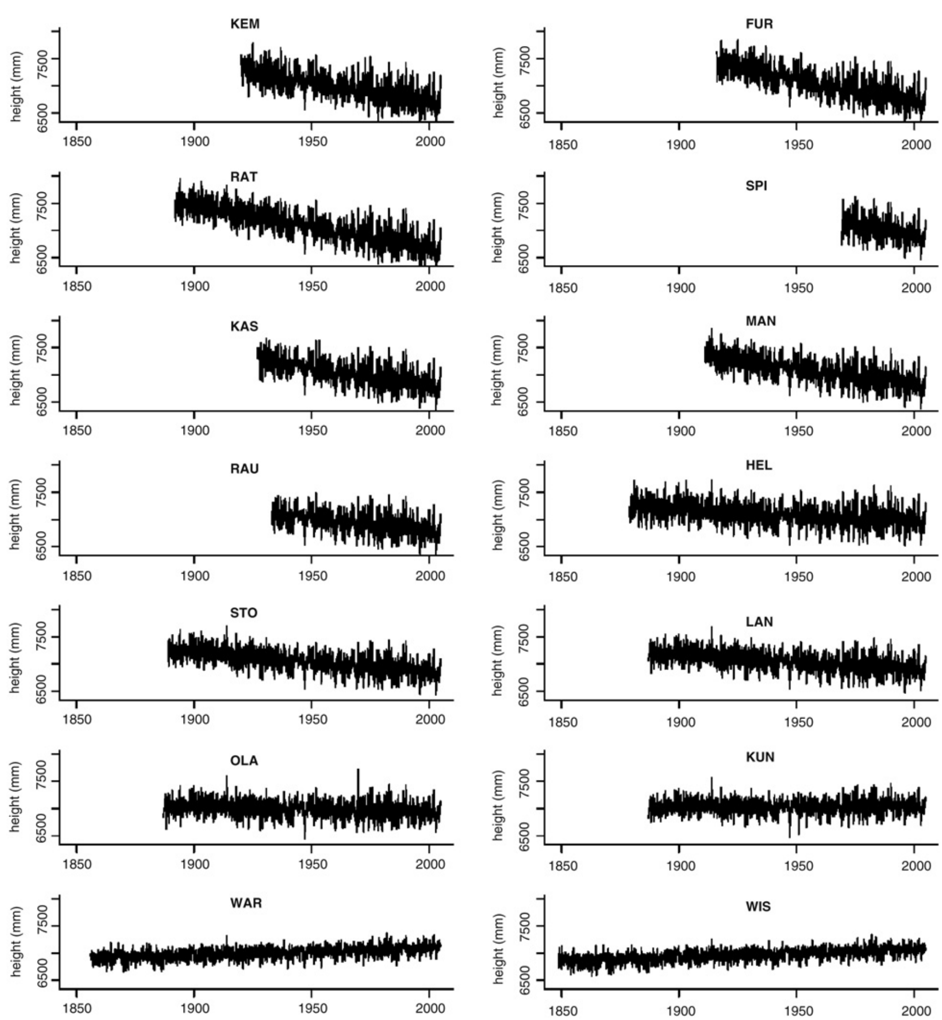

Fig. 3. Monthly time series of relative sea-level heights from Baltic tide gauges.

these models B copies of the h-step-ahead predicted values are calculated by *bootstrapping*, where the horizon h is selected by the user, in order to obtain the distances between all the time series required to built the dissimilarity matrix. Then a dendrogram based on the application of classical cluster techniques to the dissimilarity matrix is built and that gives us the different clusters formed by the models.

To fit a model we proceed as follows: first, a preliminary transformation of the data is used to remove the mean μ_t, the trend component m_t, and the seasonal component s_t. Then the estimated residuals associated to the j-time series $\mathbf{X}^{(j)}$, given by $\hat{\mathbf{Y}}_t^{(j)} = (\hat{Y}_1^{(j)}, \ldots, \hat{Y}_{n_j}^{(j)}) = (X_1^{(j)} - \hat{\mu}_t^{(j)} - \hat{m}_t^{(j)} - \hat{s}_j^{(j)}, \ldots, X_{n_j}^{(j)} - \hat{\mu}_t^{(j)} - \hat{m}_t^{(j)} - \hat{s}_j^{(j)})$, are modeled as a stationary time series which admits an autoregressive representation of the form $\epsilon_t^{(j)} = \hat{Y}_t^{(j)} - \sum_{i=1}^{p} \phi_i \hat{Y}_{t-i}^{(j)}$, for $t = p+1, \ldots, n_j$. The trend and the seasonal component are estimated by applying the iterative algorithm STL method (Seasonal-Trend decomposition procedure based on Loess) yielding a decomposition that is robust to extreme observations [37]. With these preliminaries out of the way, the bootstrap predictions can now be calculated for each time series, as follows: compute the estimated least square residuals

$$\hat{\epsilon}_t^{(j)} = \hat{Y}_t^{(j)} - \sum_{i=1}^{p} \hat{\phi}_i \hat{Y}_{t-i}^{(j)}, \quad t = p+1, \ldots, n_j, \tag{1}$$

with $(\hat{\phi}_1, \ldots, \hat{\phi}_p)$ being the least squares estimates of the autoregressive coefficients $(\phi_1, \ldots, \phi_p)$, and then the kernel density estimates, $\hat{f}_{\tilde{\epsilon}}(\cdot)$, of the centered residuals, where $\tilde{\epsilon}_t^{(j)} = \hat{\epsilon}_t^{(j)} - \bar{\hat{\epsilon}}^{(j)}$, being $\bar{\hat{\epsilon}}^{(j)}$ the mean of the $\hat{\epsilon}_t^{(j)}$'s, Compute B copies of the bootstrap future paths for $\hat{Y}_t^{(j)}$ and then, using the estimated $\hat{\mu}_t^{(j)}$, $\hat{m}_t^{(j)}$ and $\hat{s}_t^{(j)}$, obtain B copies of the bootstrap future paths for $X_t^{(j)}$. The resulting bootstrap sample $(X_{n_j+h}^{*(j,1)}, \ldots, X_{n_j+h}^{*(j,B)})$ is used to approximate the unknown distribution of $X_{n_j+h}^{(j)}$, given the observed sample. To compute the dissimilarity matrix, D, we consider the measure of distance between observation, D_{i_1,i_2} with $i_1, i_2 = 1, \ldots, T$, given in Alonso et al. [25], defined as

$$D_{i_1,i_2} = \int \left(f_{X_{n_j+h}^{(i_1)}}(x) - f_{X_{n_j+h}^{(i_2)}}(x) \right)^2 \mathrm{d}x. \tag{2}$$

The distances D_{i_1,i_2} are estimated through the expression

$$\widehat{D}_{i_1,i_2}^{*} = \int \left(\hat{f}_{X_{n_j+h}^{(i_1)*}}(x) - \hat{f}_{X_{n_j+h}^{(i_2)*}}(x) \right)^2 \mathrm{d}x, \tag{3}$$

with $\hat{f}_{X_{n_j+h}^{(i_1)*}}(x)$ being the h-step-ahead kernel density estimator at point x for the i_1-time series. Alonso et al. [25] proved that $\widehat{D}_{i_1,i_2}^{*}$ is a consistent estimator of D_{i_1,i_2}. Thus, the estimated dissimilarity matrix $\widehat{D}^{*} = (\widehat{D}_{i_1,i_2}^{*})$, $i_1, i_2 = 1, \ldots, T$ is used to perform the cluster analysis. In this case, however, classical clusters techniques such as the well-known standard k-means type algorithm can not be directly applied. Alternative procedures used as a grouping criteria in this work include agglomerative hierarchical methods with nearest distance (single linkage), average linkage and Ward's method [38,39]. The single linkage method utilizes a minimum-distance rule that starts out by first finding those two objects (in this case two time series) having the shortest distance. They constitute the first cluster. At the next stage one of two things can happen: either a third object will join the already formed cluster of two, or the two closest unclustered objects are joined to form a second cluster. The decision rests on whether the distance from one of the unclustered objects to the first cluster is shorter than the distance between the two closest unclustered objects. The process continues until all objects belong to a single cluster. For this method, the distance between clusters is defined as the distance between the nearest members. Moreover, motivation for considering the average linkage method comes from the fact that it is less dependent upon extreme values. Furthermore, it produces clusters with small within cluster variation and with approximately equal variances. Ward's method is distinct from the previous methods because it uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the sum of squares of any two (hypothetical) clusters that can be formed at each step. This method is regarded as very efficient, however, it tends to create clusters of small size; see [40,41] for details.

Table 2
Estimated autoregressive (AR(1)) models. AR: AutoRegressive, AIC: Akaike Information Criterion.

Station code	AR parameter	Std error	AIC
FUR	0.458	0.0272	13 714.9
HEL	0.442	0.0231	19 320.3
KAS	0.459	0.0290	11 937.8
KEM	0.314	0.0676	2 505.6
KUN	0.390	0.0245	17 299.1
LAN	0.426	0.0241	17 671.8
MAN	0.453	0.0266	14 326.5
OLA	0.411	0.0242	17 637.9
RAT	0.452	0.0242	17 299.4
RAU	0.476	0.0299	10 983.6
SPI	0.466	0.0456	4 830.3
STO	0.430	0.0242	17 410.4
WAR	0.306	0.0225	20 688.3
WIS	0.220	0.0235	20 028.4

4. Results

The approach described in Section 3 is applied to obtain clusters of the sea-level observations based on forecasts at a specific future time. First, in order to fit a linear model to the estimated residuals associated with each component of the panel of time series $\mathbf{X}$, different autoregressive moving average models were tested and their performance evaluated through the analysis of their corresponding residuals. It was found that an autoregressive model of order one (in short AR(1)) is adequate, in all cases, for modeling purposes. Table 2 shows the estimates of the parameters with their corresponding standard errors, and the values of the bias-corrected version of the Akaike Information Criterion (AIC) [42] suggested by Hurvich and Tsai [43].

To assess any significant deviation from Gaussian white noise sequence for the residuals, in Table 3 below the observed values of the Ljung–Box (LB) and the Kolmogorov–Smirnov–Lilliefors (KSL) statistic are presented.

Note that the assumption of independent and identically distributed sequences of normally distributed random variables is tenable for all stations.

The dendrograms based on the single and average linkage method and Ward's method, obtained for a 3-month and 6-month horizon (i.e., for March and June of 2005 respectively) are shown in Figs. 4 and 5.

The numbers of the vertical scale give the distance level where the clusters are joined. The lower values indicate that clusters are close. The dendrograms based on the single and average linkage method are very similar, as expected from the absence of outliers in the sea-level time series. Sea-level records from Helsinki (HEL) and Spikarna (SPI) are grouped as closest in terms of 3-months forecast. A second group includes Stockholm in the upper part of the Baltic Proper and the stations on the Finnish coast of the Bothnian Sea (KAS, MAN, RAU). To this group joins

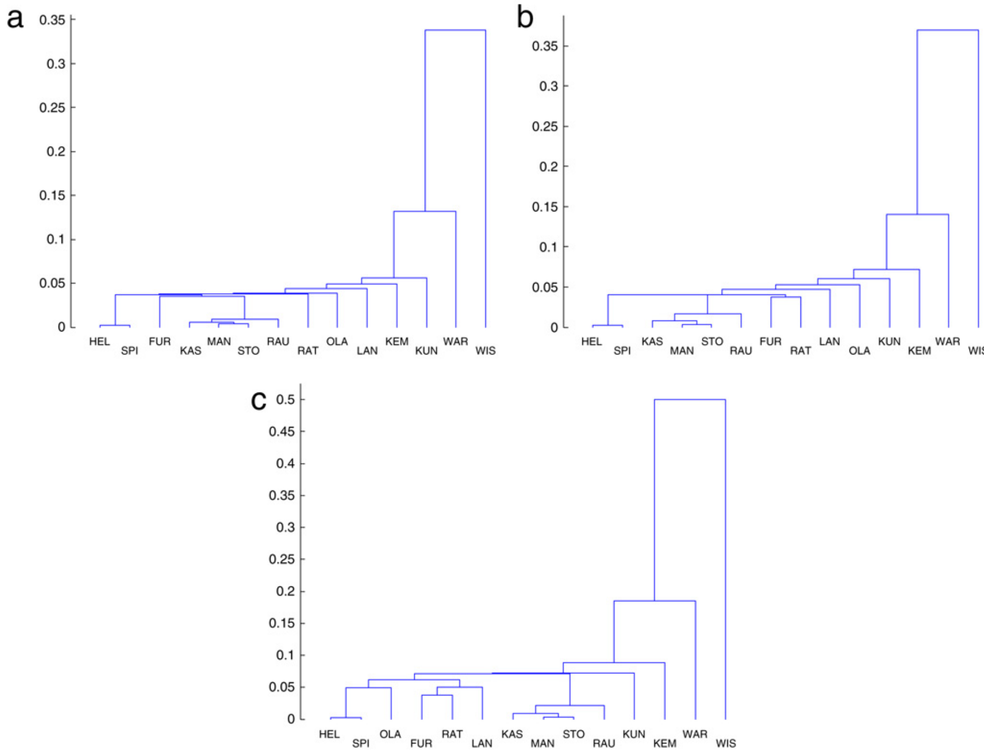

Fig. 4. Dendrogram for a $h = 3$ months horizon (March 2005). (a) Single linkage; (b) average linkage and (c) Ward's methods.

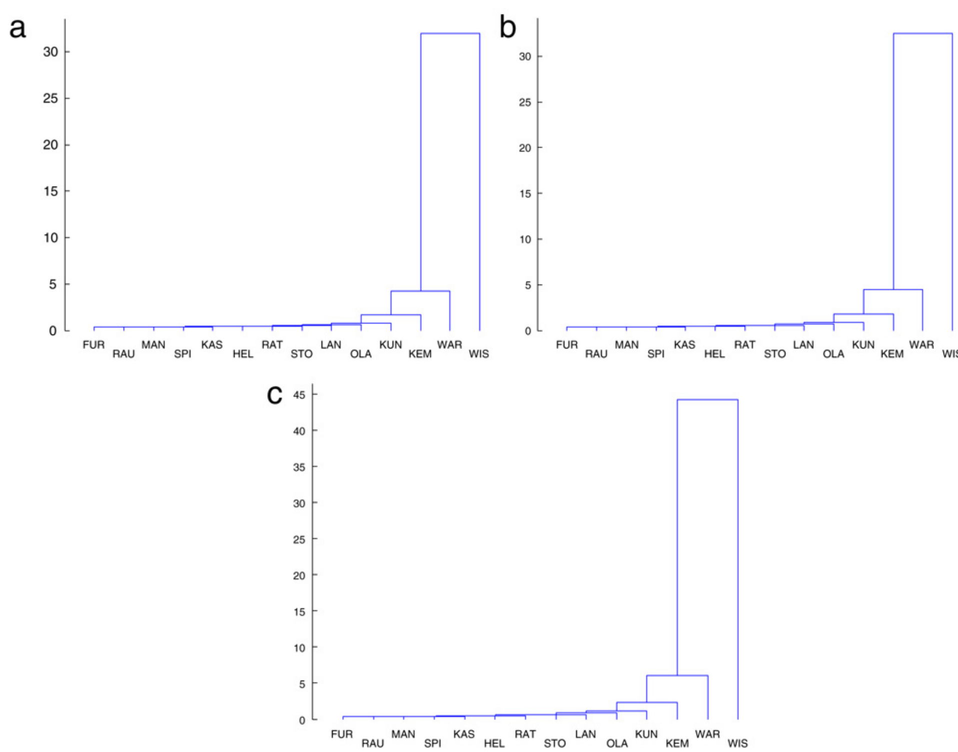

Fig. 5. Dendrogram for a $h = 6$ months horizon (June 2005). (a) Single linkage; (b) average linkage and (c) Ward's methods.

Furuögrund (FUR) and then Ratan (RAT) in the Bothnian Bay. Clustering follows spatial proximity for the stations in the Baltic Proper (OLA, LAN, KUN), with the last group including the German stations of Warnemünde (WAR) and Wismar (WIS) in the western Baltic coast. The northernmost location in the Bothnian Bay, Kemi (KEM), is grouped near the stations closest to the Danish straits. Thus, the results show a convergence to similar sea levels for these stations due to compensating influences of postglacial rebound in the north (KEM) and sea-level rise in the south (WAR, WIS). A closer look to the dendrogram based on the Ward's method shows a clearer separation between clusters. In this case OLA in the Baltic Proper joins the group of Helsinki and Spikarna and the neighboring station LAN joins the cluster of FUR and RAT from the Bothnian Bay. Still, the results are qualitatively similar to the ones

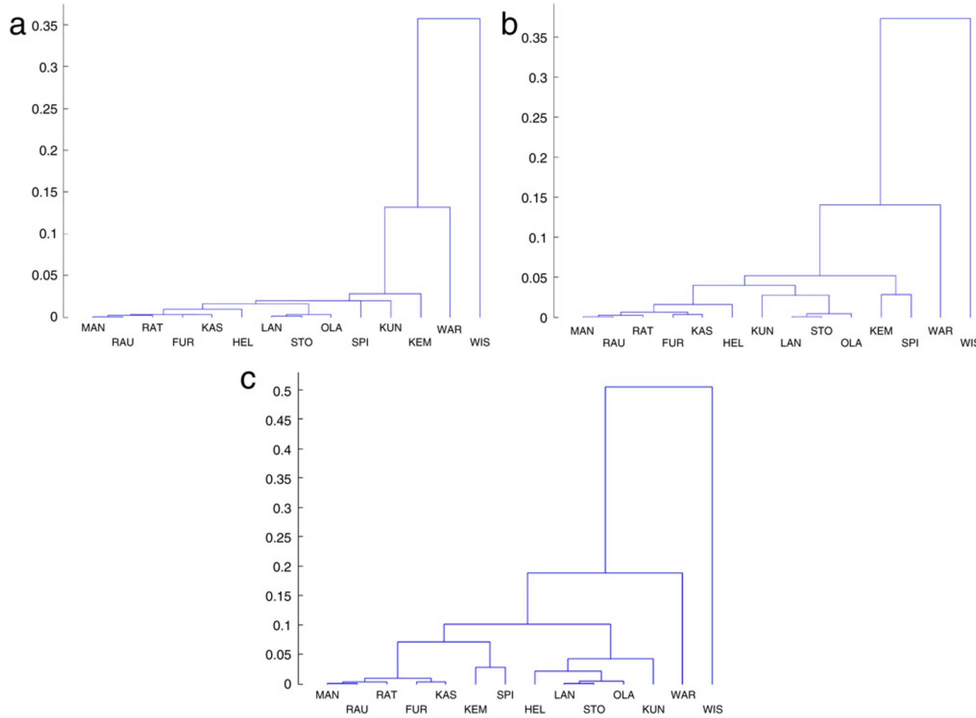

Fig. 6. Dendrogram from detrended signals for a $h = 3$ months horizon (March 2005). (a) Single linkage; (b) average linkage and (c) Ward's methods.

Table 3
Test of correlation and normality between residuals. LB: Ljung–Box, KSL: Kolmogorov–Smirnov–Lilliefors.

Station code	LB statistic	*p*-value	KSL statistics	*p*-value
FUR	19.21	0.50	0.89	0.40
HEL	18.84	0.53	1.15	0.14
KAS	19.75	0.47	1.11	0.16
KEM	31.80	0.45	1.20	0.11
KUN	19.06	0.51	0.93	0.34
LAN	16.62	0.67	0.98	0.29
MAN	19.22	0.50	0.91	0.37
OLA	24.36	0.21	0.99	0.28
RAT	23.52	0.26	0.90	0.38
RAU	19.24	0.50	0.91	0.37
SPI	24.95	0.20	0.73	0.64
STO	18.80	0.53	1.01	0.25
WAR	31.62	0.07	1.10	0.17
WIS	33.63	0.28	1.15	0.15

based on the single average and average linkage method, showing a clear separation between sea-level records near the Baltic entrance and inside the Baltic basin. For the 6-month horizon, the clustering pattern indicates a higher degree of homogeneity than for the 3-month horizon, with a clear distinction between the western Baltic stations on the German coast, Kemi in the north and the remaining tide gauge stations. Although there are some differences in terms of individual stations, results are similar to the 3-month forecasts in terms of general pattern, with the locations in the Bothnian Sea and Gulf of Finland clustering together, followed by the group of stations from the Baltic Proper and lastly the western Baltic group with the southernmost stations.

It is important to stress that the above results are based on forecast densities of future total sea-levels, including both trends and seasonal variability. Since the trend signal in the Baltic area is mostly due to postglacial rebound while seasonal variations are mainly associated with climate variability, it is instructive to consider clustering taking into account only the seasonal effects. The seasonal cycle of sea-level is affected by different climate factors, influencing both the amplitude and the phase of the seasonal pattern (e.g. [44–47]). Figs. 6 and 7 show the dendrograms from the three different clustering methods based on detrended sea-level time series. Thereby, the trends are estimated by ordinary least squares. The results are more homogeneous when considering detrended rather than total sea-level records. Furthermore, different clusters are obtained for the 3-month and 6-month horizons. While for the 3-month horizon (March 2005) the stations from the Bothnian Sea and Gulf of Finland are grouped first, for the 6-month horizon (June 2005) the stations from the Baltic Proper (LAN, STO, OLA) are clustered as closest. This suggests a distinct response of coastal sea-level in early spring and early summer, possibly due to spatial variability in hydrological discharge. Sea-levels in the Baltic area commonly exhibit an annual pattern with maximum in winter and minimum in spring.

5. Discussion and conclusions

The clustering of sea-level records with respect to forecasts at a specific future time shows as a general pattern three distinct groups: the northern stations in the Bothnian Sea and Gulf of Finland, the stations in the Baltic Proper, and the southern stations in the western Baltic. These results are consistent with the fact that for periods approximately longer than 1-month the Baltic acts as an open bay with increasing amplitudes of long-term variations from the entrance in the southern Baltic Sea inwards towards the Bothnian Bay [7,48].

Apparently unexpectedly, Kemi, the northernmost station in the Bothnian Bay, joins the cluster of the stations on the German coast closest to the Danish straits. The results from Kemi tide gauge must be interpreted with caution due to known instrumental problems, including problems in the registering of extremely low sea-levels in the early decades and change in the location of the tide gauge by nearly 7 km [10]. Still, the convergence to similar sea levels (relative to the RLR reference) of Kemi and the southern stations can be explained by the compensating influences of postglacial rebound in the north and sea-level rise in the south, in a similar fashion as depicted in Fig. 1 for time series A and C. Recent postglacial rebound in Fennoscandia based on sea-level

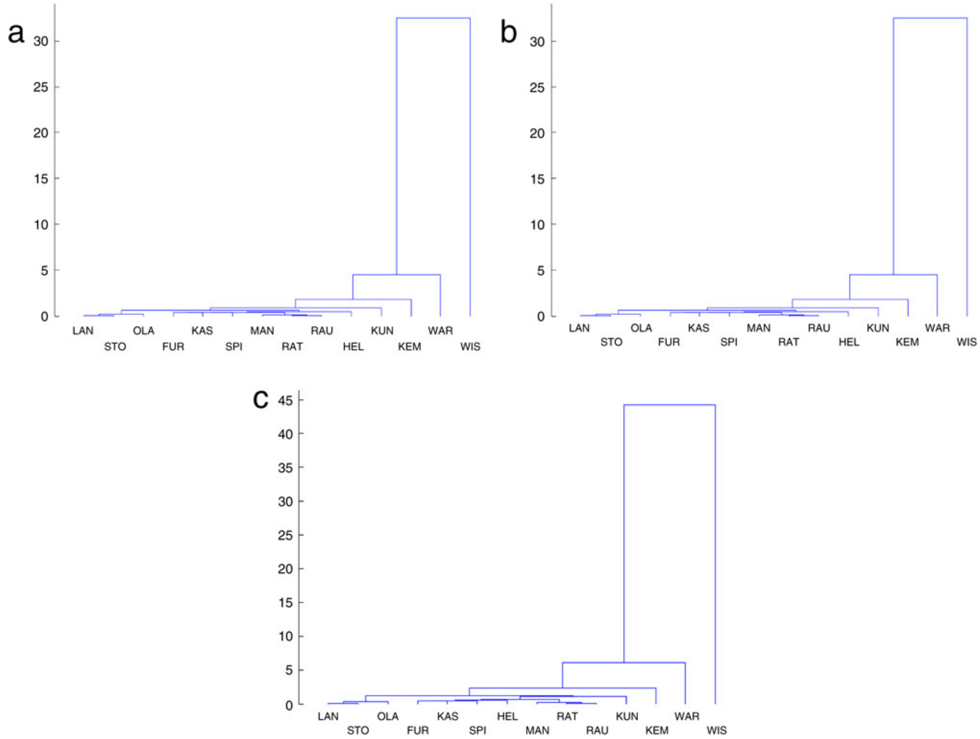

Fig. 7. Dendrogram from detrended signals for a $h = 6$ months horizon (June 2005). (a) Single linkage; (b) average linkage and (c) Ward's methods.

records shows a smooth pattern with a maximum apparent uplift in the Gulf of Bothnia of 9 mm/yr [6].

The interpretation of the differences between individual stations within the three major sub-basins is hindered by the diversity and complexity of factors influencing sea-level variations in the Baltic Sea. Although sea-level variability in the Baltic is mainly determined by the sea-level in the Kattegat and the winds controlling the exchange of waters through the Danish Straits [48,7], temperature and precipitation also contribute to sea-level variations [16], with an uneven distribution throughout the year [17]. Furthermore, the specific location of the tide gauge (at the coast and oftentimes in harbors) plays an important role in observed sea-level variations. This is particularly important in the Baltic Sea region since the morphology of the Baltic Sea coast is highly variable including open coasts, extensive archipelagos and deep embayments. Local effects associated with the coastal setting where the tide gauge measurements are taken is therefore an additional complicating factor [5].

Despite the multitude of factors influencing sea-level variability at each tide gauge location, the results of the cluster analysis show a fairly spatial coherency in terms of grouping together stations from the same sub-basin area for 3-month forecasts, while showing a high degree of similarity between most stations at a 6-month forecast and a clear separation from the stations closer to the entrance in the southwestern Baltic.

Clustering time series is a promising approach for the analysis of temporal variability within a spatial context, and specifically for linking future estimates and spatial distributions. In the case of sea-level this possibility is particularly relevant in terms of practical applications, since although mean sea-level is a relatively well known parameter on large to moderate spatial scales, its value at specific points on the coast is a complex function of a range of conditions, including coastal topography and morphology and non-linear shallow water tides and currents. The approach applied in the present study allows to identify the coastal locations displaying a similar behavior in terms of mean-sea level predictions. This information is relevant in practical applications (e.g. protection of coastal infrastructure) and complements and serves as input to models lacking the spatial resolution to identify such localized behavior.

Acknowledgments

We would like to express our gratitude to all three referees; they offered extremely valuable perspectives on our work and effective suggestions for improvements. The first author is supported by Unidade de Investigação Matemática e Aplicações of Universidade de Aveiro and by the project PTDC/MAT/64353/2006. The second author is supported by FCT (Fundação para a Ciência e Tecnologia, grant SFRH/BPD/23992/ 2005). The third author is supported by MEC (Ministerio de Educación y Ciencia, grants SEJ2005-06454 and SEJ2007-64500).

References

[1] Ekman M. The world's longest continued series of sea level observations. Pure Appl Geophys 1988;127(1):73–7.
[2] Ekman M. Climate changes detected through the world's longest sea level series. Global Planet Change 1999;21(4):215–24.
[3] Ekman M, Mäkinen J. Mean sea surface topography in the Baltic Sea and its transition area to the North Sea: A geodetic solution and comparisons with oceanographic models. J Geophys Res 1996;101(C5):11993–9.
[4] Carlsson M. Mean sea-level topography in the Baltic Sea determined by oceanographic methods. Mar Geod 1998;21(3):203–18.
[5] Novotny K, Liebsch G, Lehmann EL, Dietrich R. Variability of sea surface heights in the Baltic sea: An intercomparison of observations and model simulations. Mar Geod 2006;29(2):113–34.
[6] Ekman M. A consistent map of the postglacial uplift of Fennoscandia. Terra Nova 1996;8:158–65.
[7] Samuelsson M, Stigebrandt A. Main characteristics of the long-term sea level variability in the Baltic sea. Tellus 1996;48A:672–83.
[8] Andersson H. Influence of long-term regional and large-scale atmospheric circulation on the Baltic sea level. Tellus 2002;54A:76–88.
[9] Ekman M. A secular change in storm activity over the Baltic Sea detected through analysis of sea level data. Small Publ Historical Geophys 2007;16.
[10] Johansson M, Boman H, Kahma K, Launiainen J. Trends in sea level variability in the Baltic Sea. Boreal Environ Res 2001;6:159–79.
[11] Lehmann A, Krauss W, Hinrichsen H-H. Effects of remote and local atmospheric forcing on circulation and upwelling in the Baltic Sea. Tellus 2002;54A: 299–316.

[12] Omstedt A, Peterssen C, Rohde J, Winsor P. Baltic Sea climate: 200 years of data on air temperature sea level variation, ice cover, and atmospheric circulation. Cli Res 2004;25:205–16.
[13] Jevrejeva S, Moore JC, Woodworth PL, Grinsted A. Influence of large-scale atmospheric circulation on European sea level: Results based on the wavelet transform method. Tellus 2005;57A(2):183–93.
[14] Hünicke B, Luterbacher J, Pauling A, Zorita E. Regional differences in winter sea-level variations in the Baltic Sea for the past 200 yr. Tellus 2008;60A: 384–93.
[15] Ekman M. The world's longest sea level series and a winter oscillation index for Northern Europe 1744–2000. Small Publ Historical Geophys 2003;12:30.
[16] Hünicke B, Zorita E. Influence of temperature and precipitation on decadal Baltic Sea level variations in the 20th century. Tellus 2006;58A:141–53.
[17] Chen D, Omstedt A. Climate-induced variability of sea level in Stockholm: Influence of air temperature and atmospheric circulation. Adv Atmos Sci 2005; 22(5):655–64.
[18] Milne GA, Davis JL, Mitrovica JX, Scherneck HG, Johansson JM, Vermeer M, et al. Space-Geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia. Science 2001;291:2381.
[19] Woodworth PL, Pugh DT, Meredith MP, Blackman DL. Sea level changes at Port Stanley. Falkland Islands. J Geophys Res 2005;110:C06013.
[20] Barbosa SM. Quantile trends in Baltic sea-level. Geophys Res Lett 2008;35: L22704.
[21] Wroblewski A. Inter-annual oscillations of Baltic water volumes and sea levels. Oceanologia 1998;40:183–203.
[22] Heyen H, Zorita E, von storch H. Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus 1996; 48:312–23.
[23] Madsen KS, Høyer JL, Tscherning CC. Near-coastal satellite altimetry: Sea surface height variability in the North Sea–Baltic Sea area. Geophys Res Lett 2007;34:L14602007.
[24] Robinson JE. Crosscorrelation of sea levels. In: Bonham-Carter G, Cheng Q, editors. Progress in Geomathematics. Berlin, Heidelberg: Springer; 2008. p. 529–34.
[25] Alonso AM, Berrendero JR, Hernández A, Justel A. Time series clustering based on forecast densities. Comput Statist Data Anal 2006;51(2):762–76.
[26] Macchiato MF, La Rotonda L, Lapenna L, Ragosta M. Time modelling and spatial clustering of daily temperatures: An application in southern Italy. Environmetrics 1995;6(1):31–53.
[27] Cowpertwait PSP, Cox TF. Clustering population means under heterogeneity of variance with an application to a rainfall time series problem. The Statistician 1992;41(1):113–21.
[28] Piccolo DA. distance measure for classifying ARIMA models. J Time Ser Anal 1990;11:153–64.
[29] Maharaj EA. A significance test for classifying ARMA models. J Statist Comput Simul 1996;54(4):305–31.
[30] Maharaj EA. Clusters time series. J Classification 2000;17(2):297–314.
[31] Kakizawa Y, Shumway RH, Taniguchi M. Discriminant and clustering for multivariate time series. J Amer Statist Assoc 1998;93(441):328–40.
[32] Vilar JA, Pértega S. Discriminant and cluster analysis for Gaussian stationary processes: Local linear fitting approach. Nonparametr Stat 2004;16(3-4): 443–62.
[33] Fruhwirth-Schnatter S, Kaufmann S. Model-based clustering time series. J Bus Econom Statist 2008;26(1):78–89.
[34] Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, et al. Global climate projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, editors. Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge, (United Kingdom, NY, USA): Cambridge University Press; 2007.
[35] Grinsted A, Moore J C, Jevrejeva S. Reconstructing sea level from paleo and projected temperatures 200 to 2100AD. Clim Dyn 2009 [in press].
[36] Woodworth PL, Player R. The permanent service for mean sea level: An update to the 21st century. J Coastal Res 2003;19:287–95.
[37] Cleveland RB, Cleveland WS, Mcrae JE, Terpenning I. STL: A seasonal-trend decomposition procedure based on Loess. J Official Statist 1990;6(1):3–73.
[38] Ward JM. Hierarchical grouping to optimize an objective function. J Amer Statist Assoc 1963;58:236–44.
[39] Wishart D. An algorithm for hierarchical classifications. Biometrics 1969;25: 165–70.
[40] Everitt BS, Landau S, Leese M. Cluster analysis. fourth ed. London: Arnold; 2001.
[41] Wishart D. Efficient hierarchical cluster analysis for data mining and knowledge discovery. Comput Science Statist 1998;30:257–63.
[42] Akaike H. A new look at the statistical model identification. IEEE Trans Aut Contr 1974;AC-19:716–23.
[43] Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika 1989;76:297–307.
[44] Ekman M, Stigebrandt A. Secular change of the seasonal variation in sea level and of the pole tide in the Baltic Sea. J Geophys Res 1990;95(C4):5379–83.
[45] Plag HP, Tsimplis MN. Temporal variability of the seasonal sea level cycle in the north sea and Baltic sea in relation to climate variability. Global Planet Change 1999;20(2–3):173–203.
[46] Hünicke B, Zorita E. Trends in the amplitude of Baltic Sea level annual cycle. Tellus 2008;60A:154–64.
[47] Barbosa SM, Silva ME, Fernandes MJ. Changing seasonality in North Atlantic coastal sea level from the analysis of long tide gauge records. Tellus 2008;60A: 165–77.
[48] Ekman M. A common pattern for interannual and periodical sea-level variations in the Baltic Sea and adjacent waters. Geophysica 1996;32:261–72.