

<sup>1</sup> **Discrimination of Water Quality Monitoring Sites in  
2 River Vouga using a Mixed-Effect State Space Model**

<sup>3</sup> **Marco Costa · Magda Monteiro**

<sup>4</sup>  
<sup>5</sup> Received: date / Accepted: date

<sup>6</sup> **Abstract** The surface water quality monitoring is an important concern of pub-  
7 lic organizations due to its relevance to the public health. Statistical methods are  
8 taken as consistent and essential tools in the monitoring procedures in order to  
9 prevent and identify environmental problems. This work presents the study case of  
10 the hydrological basin of the river Vouga, in Portugal. The main goal is discrimi-  
11 nate the water monitoring sites using the monthly dissolved oxygen concentration  
12 dataset between January 2002 and May 2013. This is achieved through the extrac-  
13 tion of trend and seasonal components in a linear mixed-effect state space model.  
14 The parameters estimation is performed with both maximum likelihood method  
15 and distribution-free estimators in a two-step procedure. The application of the  
16 Kalman smoother algorithm allows to obtain predictions of the structural com-  
17 ponents as trend and seasonality. The water monitoring sites are discriminated  
18 through the structural components by a hierarchical agglomerative clustering pro-  
19 cedure. This procedure identified different homogenous groups relatively to the  
20 trend and seasonality components and some characteristics of the hydrological  
21 basin are presented in order to support the results.

<sup>22</sup> **Keywords** Water quality assessment · State space modeling · Kalman smoother ·  
23 Classification · Structural components · River Vouga

<sup>24</sup> **1 Introduction**

<sup>25</sup> The surface water quality assessment is an important part of the environment  
26 monitoring, whose evaluation can predict the water quality and avoid public health  
27 problems of various types and levels. The existence of an effective and efficient  
28 water quality monitoring system prevents the pollution of both water and soil.

---

M. Costa and M. Monteiro  
Escola Superior de Tecnologia e Gestão de Águeda  
Centro de Investigação e Desenvolvimento em Matemática e Aplicações  
Universidade de Aveiro  
Apartado 473, 3754 – 909 Águeda, Portugal  
E-mail: marco@ua.pt

29 There are several factors that contribute to water quality, some factors are known,  
30 others are unknown, which is a grey system ([30]).

31 Water quality monitoring procedures may be used in the decision-making process  
32 in order to support policy options. For this reason, several European Union  
33 (EU) countries have developed a national water quality system, considering characteris-  
34 tic structure of their own rivers and have used this type of indicators to evaluate the current  
35 situation of their water quality level. The management of water resources is regulated by EU  
36 directives and their transposition into national legislation. For instance, in Portugal, the Law n. 58/2005 (Law of Water)  
37 ensures the transposition into national law the Directive n. 2000/60/CE (the Water  
38 Framework Directive, WFD), which creates the institutional framework for sustainable  
39 management of surface, interior waters, transitional, coastal and even groundwater. The Decree-Law n. 77/2006 complements the WFD by characterizing  
40 the waters of a river basin. This regulatory instrument establishes the status  
41 of surface waters and groundwater and the ecological potential.

42 The knowledge of the dynamics of water quality surface can be achieved by  
43 studying the respective hydrological basin and its unique characteristics. The wa-  
44 ter quality assessment is, in general, based in a network of water quality monitor-  
45 ing sites which provides real-time water-quality measurements from surface-water  
46 monitoring locations. These sites can be fixed stations (usually to characterize a  
47 watershed); on a temporary basis (for instance, during the summer at bathing  
48 beaches) or on an emergency basis. This work focuses on the water quality as-  
49 sessment based on a set of fixed stations located in the hydrological basin of the  
50 river Vouga, Portugal. In this case, there is periodical data as frequent as possible  
51 in order to identify changes or trends in water quality over time or to devalue  
52 sporadic behavior in medium or long term analysis. Nevertheless, nowadays, the  
53 availability of knowledge about a watershed in a considerable period of time and  
54 with a reasonable spatial coverage enables a more efficient monitoring of water  
55 quality.

56 It is in the context of both legal framework and a significant investment effort  
57 in the water quality monitoring infrastructures in the river Vouga basin, in Portu-  
58 gal, that it is important to characterize the existing network. Thus, an adequate  
59 research in order to characterize the network can identify potential redundancies  
60 of monitoring sites. The minimization of these redundancies can bring a better  
61 use of resources maintaining the effectiveness of the monitoring process. So, this  
62 work aims to contribute to a better knowledge of the dynamics of the watershed  
63 to help in decision-making processes technical and policy that may be adopted in  
64 the near future.

65 An important role in the surface water quality monitoring is assigned to the  
66 dissolved oxygen (DO) concentration variable. Indeed, the amount of dissolved  
67 oxygen has been considered a relevant indicator of the water quality since it results  
68 from the impact of a set of environmental factors. These factors may be originate  
69 from a several conditions as the water temperature, air temperature and pressure,  
70 riverbed morphology, water cleanliness state, point and area sources of pollution  
71 of surface water, etc. Whence, several research is based on this variable.

72 This work presents a characterization of the river Vouga watershed, in Portugal,  
73 based on records of the DO concentration, in mg/l, identifying similarities or dis-  
74 similarities between monitoring sites. The statistical methodology classifies water  
75 monitoring sites according to both trend and seasonality time series components.

78 For each component, the obtained homogenous groups will be analyzed according  
79 to the watershed hydrology characteristics. The statistical approach combines  
80 time series analysis with the usual discrimination techniques as the cluster analy-  
81 sis. The time series analysis is performed through a state space modeling approach  
82 combined with the Kalman smoother in order to extract structural components  
83 which are used to investigate space-time patterns in the water quality monitoring  
84 sites network.

85 **2 Literature review**

86 Several studies have been developed on the river Vouga watershed or, particu-  
87 larly, on the Ria de Aveiro lagoon. The main focus of these works is related with  
88 ecological systems in the Ria de Aveiro as the diversity of flora and fauna or the  
89 contaminants into aquatic ecosystems (see, e.g., [1]). [6] presents a study in order  
90 to identify point sources of pollution and to assess the surface water quality in  
91 the Antuã basin by monitoring physicochemical variables. However, an analysis  
92 to characterize the main hydrological basin of the river Vouga according to the  
93 water quality in the monitoring sites in a discrimination view point has not been  
94 addressed yet. This work aims at giving a contribution towards this direction.

95 Several statistical techniques can be applied when the main goal is to charac-  
96 terize environmental variables through various temporal and spatial patterns. For  
97 instance, [12] presents a scheme for meteorological drought analysis at various tem-  
98 poral and spatial scales based on a spatial Bayesian interpolation of drought sever-  
99 ity derived from monthly precipitation data. [17] investigates both water quality  
100 evaluation in its time-space variations and the natural and anthropogenic origins of  
101 contaminants in surface or ground water. [4] presents the application of multivari-  
102 ate statistics for the interpretation of surface and groundwater data from Tarkwa.  
103 Both cluster analysis and principal component analysis were used to analyze the  
104 water quality in [28] and [29] in order to evaluate the temporal/spatial variations  
105 and to identify potential pollution sources. The factorial analysis was used in [9]  
106 in order to explain and evaluate the correlation structure between observed vari-  
107 ables in water quality sampling stations and to identify relevant factors. [15] uses  
108 cluster analysis and linear models to describe hydrological space–time series of  
109 quality variables and to detect changes in surface water quality before and after  
110 the installation of wastewater treatment plants. [8] applied clustering techniques  
111 based on Kullback Information, measures that are obtained in the state space  
112 modeling process and, for each homogeneous group, forecast models were com-  
113 pared with traditional linear models through the mean squared error of forecasts.  
114 Two approaches for clustering of time series oriented to large set of time series  
115 were proposed in [14]; the first is an approach based on a modification of classic  
116 state-space modeling while the second is based on functional clustering. In these  
117 works the discrimination procedure is performed directly on the environmental  
118 variables. The cluster analysis has been usefully applied also in [19] in order to  
119 differentiate between efficient and inefficient farms using a clustering model based  
120 on the imperialist competitive algorithm.

121 On the other hand, the DO concentration is a parameter frequently used to  
122 evaluate the water quality on different reservoirs and watersheds since it is strongly  
123 influenced by a combination of physical, chemical, and biological characteristics

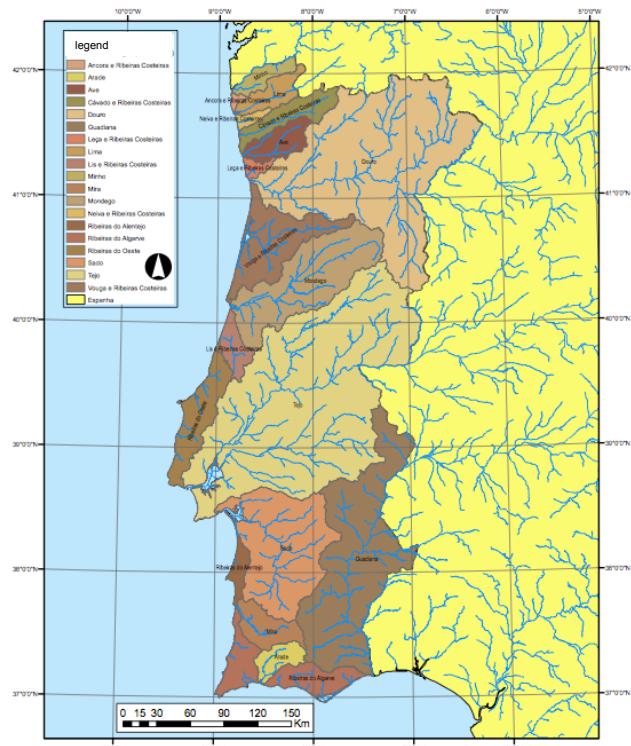
of streams. The DO is considered an index of water quality and was also used to estimate the effect of industrial and municipal effluents on the waters ([24], [25], [16]). With the same purpose, [22] validates a water quality model for the Ria de Aveiro, in order to better use it as a predictive tool in the study of the main water quality processes in the this lagoon, providing a sensitivity analysis of the model, which shows that the ocean remains the main source of oxygen as well as the main factor controlling the DO distribution throughout the main lagoon areas. Most recently, [27] uses dissolved oxygen (DO) indicators to calibrate the recharge potential analysis (RPA) parameters, which results indicated that defining the RPA parameters values based on DO indicators is necessary and important for accuracy. The ARIMA and ARFIMA models were applied in [3] to predict univariate DO time series for four water quality assessment stations at Stillaguamish River located in the state of Washington.

On the one hand, the approach proposed in this work has the potential of combining the temporal modeling of water quality variables evolution with a clustering analysis. Furthermore this approach allows, at the same time, a global characterization of water quality in the river basin and the identification of redundancies of water monitoring sites. On the other hand, the stochastic modeling is performed using a mixed linear state space model incorporating both fixed effects and random dynamics which has the advantage to model and forecast of non-stationary changes inherent in climate data ([20]). Other advantage of the State space approach is that it takes into account possible measurement errors measures which are minimized through the Kalman smoothers.

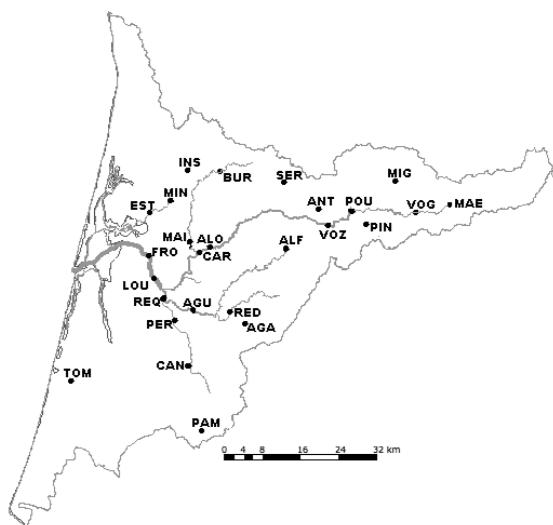
### 3 The river Vouga and data description

The hydrologic regime involves a summer low flow condition and the dynamic of the coastal lagoon is dominated by tidal oscillation. Ria de Aveiro is characterized by its rich biodiversity as well as by an increasing pressure of the anthropogenic activities near its margins, namely building and land occupation, agricultural and industrial activities. This has resulted in a significant change of the lagoon morphology, and in a constant input of a large volume of anthropogenic nutrients as well as of contaminant loads, with the consequent negative impact in the water circulation, as well as in the water quality of the lagoon ([21]). The construction, management and operation of Multi-municipality System Drainage of the Ria de Aveiro is of the responsibility of the SIMRIA - Integrated Sanitation of Municipalities of Ria, SA, which is a private company with majority public capital (established by Decree-Law n. 101/97 of 26 April). The Ria de Aveiro lagoon is inserted in the hydrological basin called by Vouga/Ribeiras Costeiras in the SNIRH (Portuguese national information system for water resources). In the annual report 2012 published by SNIRH, it is mentioned that the industrial activities with more units that contribute to the sources of urban pollution in the Vouga watershed come from manufacture of leather, manufacture of metal products and non-metallic, wood and cork industry, chemical manufacturing, food industries-oil, pulp and paper industry and metallurgical industries.

Vouga is a river situated in the center of Portugal and it rises at about 930m of altitude near the geodesic landmark Facho da Lapa, in Serra da Lapa, a mountain located in the district of Viseu; it flows 148 Km before emptying into Ria de Aveiro.



**Fig. 1** Hydrological basins of mainland Portugal (source SNIRH)



**Fig. 2** Water monitoring sites locations in the hydrological basin of river Vouga

**Table 1** Descriptive statistics of dissolved oxygen concentration between January 2002 and May 2013

Site	abbrev	obs	min	max	average	st dev
Agadão	AGA	111	5.8	11.0	8.74	1.26
Carvoeiro	CAR	112	6.2	11.0	8.79	1.18
Alombada	ALO	113	6.1	11.0	8.90	1.08
Captação Burgães	BUR	122	6.5	12.6	9.40	1.16
Captação Rio Ínsua	INS	122	6.4	12.4	9.31	1.05
Ponte Redonda	RED	112	4.6	11.5	8.88	1.22
Frossos	FRO	110	4.5	11.0	8.17	1.22
Pampilhosa	PAM	100	4.3	12.0	7.95	1.68
Ponte São João de Loure	LOU	112	5.4	11.0	8.24	1.25
Ponte Vale Maior	MAI	112	6.2	12.0	8.62	1.12
Ponte Águeda	AGU	111	5.1	11.0	8.39	1.20
São Tomé	TOM	118	5.0	11.0	7.88	1.16
Aç. Maeira	MAE	115	5.6	11.0	8.50	1.20
Aç. Rio Alfusqueiro	ALF	113	2.9	12.0	7.80	1.75
Pindelo Milagres	MIL	110	4.6	12.0	8.16	1.42
Ponte Antim	ANT	113	0.8	12.0	7.38	2.05
Ponte Pouves	POU	115	2.6	11.0	8.27	1.46
Ponte Vouzela	VOZ	109	1.8	13.0	8.10	1.91
São João Serra	SER	115	6.0	12.0	8.70	1.18
São Miguel Mato	MAT	111	4.3	12.0	8.44	1.53
Vouguinha	VOG	114	5.4	11.0	8.42	1.35
Estarreja	EST	114	3.4	11.0	7.62	1.32
Perrões	PER	111	4.6	9.8	7.19	1.17
Ponte Canha (Vouga)	CAN	114	2.6	10.1	6.89	1.92
Ponte Minhoteira	MIN	112	0.7	10.0	7.73	1.50
Ponte Requeixo	REQ	111	3.9	11.0	7.13	1.52

170 The watershed of the Vouga is the second largest basin of watercourses that run  
 171 exclusively in Portuguese territory comprising a total area of 3706 Km<sup>2</sup>. More  
 172 specifically, the Vouga basin is located in the transition zone between the North  
 173 and South of Portugal, i.e., between the watersheds of the Douro at north and  
 174 Mondego at south (see Fig.1).

175 The average flow of fresh water that flows into the Ria de Aveiro is about  
 176 40 m<sup>3</sup>/s. The Vouga and Antuã rivers are the main sources of fresh water, with  
 177 average annual flow of 24 m<sup>3</sup>/s and 2.4 m<sup>3</sup>/s, both rivers belonging to the Vouga  
 178 watershed ([23]). The main tributaries of the River Vouga are, from upstream to  
 179 downstream the River Mel, the Sul River, the Varoso, the river Teixeira, the river  
 180 Arões, the river Mau and the Caima river on the right bank. On its left bank  
 181 the river Ribamá, the Marnel, and the river Águeda with its major tributary, the  
 182 Alfusqueiro.

183 The dissolved oxygen concentration is available in a set of water monitoring  
 184 sites in the hydrological basin of river Vouga. However, some problems arise in the  
 185 statistical modeling, namely, some water monitoring sites have few data or missing  
 186 values. On the other hand, due to the lack of economic resources or some other  
 187 factor, the data collection was discontinued in some sites. In the SNIRH system  
 188 there are 78 water-monitoring sites registered on the hydrological basin of the river  
 189 Vouga. Unfortunately, the data collection is not continuous or some stations were  
 190 deactivated at some time. Relatively to the DO concentration 26 stations have a

191 significant data set until May 2013 (the last month available in the system). These  
192 water monitoring sites are represented in the Figure 2.

193 Data available in the SNIRH system is not temporal equidistant, that is, in  
194 some sites and for some months there are more than one measurement (for in-  
195 stance, two measurements for the same site in different days of the same month).  
196 The format of original dataset is improper to the statistical analysis, so it was  
197 changed to producing monthly data. The adopted methodology to produce the  
198 time series used to the purposes of this study is based on the average of mea-  
199 surements. When in a month/year there were more than one measurement it was  
200 considered their average to that month/year. Authors consider that an improve-  
201 ment in the data collection is desirable to increase statistical analyses accuracy.  
202 However, these improvements can only be applied to future collections of mea-  
203 surements. On the other hand, the way the data was collected does not jeopardize  
204 the results obtained in this work once, in general, data collection in the network  
205 has been followed a monthly scheme. That is, given the annual calendar and other  
206 constraints (holidays, weather conditions, etc.), the collection of samples remained  
207 monthly and, whenever possible, at the same time of the month at each water  
208 monitoring site.

209 Table 1 presents the descriptive statistics of the monthly DO concentration be-  
210 tween January 2002 and May 2013 according to the final dataset. An exploratory  
211 analysis shows that, in general, data are not normally distributed. Indeed, in some  
212 water monitoring sites, observations are leptokurtic. This fact must be taken in  
213 consideration in the modeling procedures since the Gaussian distribution is a usual  
214 assumption in several statistical analyses. Moreover, the box-plots of data identi-  
215 fied several moderate outliers in many sites, almost all in the left tail.

216 All graphical representations of the times series of the DO concentration show  
217 that there is a seasonal pattern. The monthly averages of each month (empirical  
218 seasonal coefficients) of the year indicate that DO concentration is greater in the  
219 winter months and lower in the summer months. This result is due to the hydro-  
220 meteorological conditions since the DO concentration is largely influenced by the  
221 precipitation amount and temperature. Furthermore, the variances of observations  
222 within each month of the year vary and they tend to be greater in winter months  
223 ([10]). This result indicates the existence of variance heterogeneity instead of the  
224 usual homocedasticity assumed in several models.

#### 225 **4 A linear mixed-effect state space model**

226 A preliminary work was performed based on the water monitoring site of Carvoeiro  
227 data ([11]). This work showed that when a linear regression model, which incor-  
228 porated a linear trend and seasonal coefficients, is applied, the residual series does  
229 not present a white noise behavior. In fact, the sample autocorrelation function  
230 (ACF) and the partial autocorrelation function (PACF) showed that residual se-  
231 ries follows an autoregressive process of order 1, AR(1), that is, there is a temporal  
232 correlation structure which were not explained by the linear model.

233 Thus, other models have to be considered in order to incorporate the structural  
234 components of the DO concentration as well as the time correlation structure. A  
235 proper choice is a linear mixed-effect state space (LMESS) modeling framework.  
236 The LMESS models have been applied in several modeling works ([20], [31]) with

good results. On the one hand, static statistical models with fixed effects are unlikely to have a good predictive accuracy, particularly in situations where the predictor and predictand relationship changes over time ([20]). On the other hand, the usual linear regression models are homocedastic which is a strong constraint regarding the results of the exploratory analysis. Thus, the LMESS allows to combine the simplicity of linear models with a temporal dynamic structure usually associated to the environmental variables.

Let  $Y_t$ , with  $t = 1, 2, \dots, n$ , be the DO concentration variable in a water monitoring site. The LMESS is specified by two equations: the observation equation and the state equation. The observation equation is given by

$$Y_t = \beta t + s_t X_t + e_t \quad (1)$$

where  $Y_t$  is the observed DO concentration at time  $t$  in a monitoring site,  $\beta$  is a slope parameter,  $s_t = s_{t \bmod 12} = s_i$ , with  $i = 0, \dots, 11$ , corresponding to the monthly seasonal coefficient (0- December, 1-January, ..., 11-November) and  $e_t$  is a white noise process ( $E(e_t) = 0$ ,  $\text{var}(e_t) = \sigma_e^2$  for all  $t$  and  $\text{cov}(e_t, e_r) = 0$  for all  $t \neq r$ ). In addition,  $X_t$  is an unobservable random variable, the state, which is assumed to follow an autoregressive process of order 1, AR(1), according to the state equation

$$X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t \quad (2)$$

where  $\mu$  is a parameter,  $\phi$  is the transition parameter and variables  $\varepsilon_t$  are a white noise process ( $E(\varepsilon_t) = 0$ ,  $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$  for all  $t$  and  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$  for all  $t \neq s$ ). It is assumed that the processes  $e_t$  and  $\varepsilon_t$  are uncorrelated,  $E(e_t \varepsilon_s) = 0$  for all  $t$  and  $s$ . When the state process  $\{X_t\}$  is stationary, that is  $|\phi| < 1$ , the parameter  $\mu$  represents the mean of the process.

The model defined by Eq. (1) and Eq. (2) can be interpreted as a linear regression model which incorporates a stochastic calibration factor in the seasonal component. In fact, the component  $s_t X_t$  includes the usual seasonal coefficients which are calibrated through a stochastic factor  $X_t$ . This formulation incorporates the heteroedasticity which was identified in the exploratory analysis. Indeed, it was checked, in an empirical analysis, that the monthly standard deviations of the detrended time series were greater in the months with a higher value of the DO concentration (winter months). Moreover, the LMESS model includes the usual linear trend.

The observation equation of the LMESS model (1)-(2) can be rearranged in order to emphasize the seasonal coefficients with the desirable property  $\sum_{i=0}^{11} s_i = 0$  as

$$Y_t = \alpha X_t + \beta t + s_t^* X_t + e_t \quad (3)$$

where  $\alpha = \frac{1}{12} \sum_{i=0}^{11} s_i$  and  $s_t^* = s_t - \alpha$ .

This formulation is equivalent to Eq. (1) but it is more useful for interpretation and modeling purposes. Indeed, this formulation shows a trend component,  $T_t = \alpha X_t + \beta t$ , with a constant slope but with a stochastic intercept and a stochastic seasonal component,  $S_t = s_t^* X_t$ , based on the overall seasonal coefficients but that allows its calibration dynamically. As the states  $X_t$  are unobservable random variables they must be predicted. This is done through the Kalman smoother ([26]). As usual,  $\hat{X}_{t|t-1}$ ,  $\hat{X}_{t|t}$  and  $\hat{X}_{t|n}$  represent the one-step-ahead forecast, the filtered prediction and the smoother prediction of  $X_t$  based on time up to  $t-1$ ,  $t$  and  $n$ , respectively.

**Table 2** Estimates of slopes and seasonal coefficients from the method of least squares

site	$\hat{\beta}$	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
AGA	-0.0081	10.83	10.78	10.33	10.65	9.51	8.86	8.33	8.17	8.26	8.64	9.61	10.55
AGU	-0.0067	10.56	10.09	9.97	9.29	9.03	7.98	7.85	7.77	7.89	7.88	9.09	10.15
ALF	-0.0129	10.87	10.52	9.94	9.82	9.13	8.02	7.55	7.42	7.11	7.29	8.91	9.91
ALO	-0.0036	10.33	9.99	9.72	9.38	9.01	8.37	8.60	8.39	8.48	8.47	9.30	10.14
ANT	-0.0064	10.27	9.72	9.46	9.24	8.85	7.60	6.13	4.70	5.36	7.12	8.34	9.06
BUR	-0.0080	11.20	10.84	10.75	10.2	9.57	9.20	9.07	9.04	8.90	9.79	10.21	10.97
CAN	-0.0139	9.23	9.97	9.74	8.79	8.68	7.96	6.97	6.13	6.05	6.10	7.39	8.93
CAR	-0.0020	10.34	10.11	9.68	9.11	8.70	7.90	8.29	8.97	8.07	7.98	9.10	10.08
EST	0.0008	9.03	8.88	8.35	7.85	7.12	6.92	7.42	7.22	6.86	5.90	7.65	8.71
FRO	-0.0046	9.92	9.87	9.57	9.01	8.29	7.35	7.61	7.97	7.52	7.44	8.59	9.83
INS	-0.0070	10.81	10.52	10.49	10.13	9.35	9.23	8.91	9.26	9.03	9.68	9.82	10.54
LOU	-0.0038	9.90	9.80	9.41	8.96	8.52	7.80	7.47	7.28	7.38	7.61	8.24	10.07
MAE	-0.0037	10.29	9.72	9.80	9.22	8.83	7.99	8.06	7.57	7.56	8.26	8.91	9.30
MAI	-0.0051	10.30	10.26	9.75	9.25	8.78	8.12	8.11	8.52	8.17	8.02	9.11	10.14
MIG	-0.0058	10.71	10.13	9.85	9.72	9.22	8.59	8.37	7.31	7.51	7.83	8.86	9.73
MIN	0.0028	8.92	9.11	8.73	8.01	7.31	6.90	7.04	7.11	6.63	5.64	7.46	8.98
PAM	-0.0169	11.10	10.75	10.48	9.91	9.22	7.68	8.07	8.01	8.79	8.04	8.87	9.54
PER	-0.0103	9.11	9.27	8.97	7.99	8.11	7.38	6.96	6.99	7.13	7.22	8.08	8.91
PIN	-0.0062	10.49	9.61	9.67	9.22	9.01	8.09	7.76	6.82	6.62	7.55	8.40	8.93
POU	-0.0039	10.21	9.88	9.56	9.28	8.94	8.16	8.00	6.81	7.67	8.21	8.17	9.55
RED	-0.0071	11.01	10.53	10.23	9.69	9.58	8.88	8.33	8.27	8.26	8.76	9.81	10.78
REQ	0.0062	8.39	8.53	8.06	7.17	6.63	5.96	5.74	5.36	5.49	5.57	6.66	7.93
SER	-0.0052	10.12	10.00	9.85	9.77	9.12	8.56	8.20	7.87	8.14	8.80	9.44	9.64
TOM	-0.0068	9.51	9.24	9.26	8.66	8.35	7.80	7.63	7.53	7.63	7.49	8.08	8.48
VOG	-0.0049	10.84	9.89	9.84	9.48	9.03	8.17	7.73	7.19	7.50	7.90	9.00	9.38
VOZ	-0.0011	10.91	9.72	9.19	9.41	8.61	7.32	7.32	6.61	5.85	7.72	8.67	9.47

## 281 5 Adjustment of the LMESS model to the DO concentration

282 The LMESS model formulated in (1)-(2) contains a set of unknown parameters  
 283 that must be estimated from data for each of the 26 times series. These parameters  
 284 are  $\Theta = \{\beta, s_i, \mu, \phi, \sigma_e^2, \sigma_\epsilon^2\}$ , with  $i = 0, 1, \dots, 11$  relatively to the twelve months  
 285 of the year. Parameters estimation of state space models is performed usually by  
 286 the maximum likelihood estimation. In the mixed-effect state space model fitting  
 287 context, [20] implemented the EM algorithm assuming the normality of errors, and  
 288 developing the updating equations for the M-step associated to the fixed effects  
 289 parameters.

290 We consider a classical decomposition approach ([5], p. 23) which combines the  
 291 least square estimation of the fixed-effects parameters with an estimation method  
 292 focused on state space models. So, in a first step, for each time series it was  
 293 applied the method of least squares in order to estimate the slope  $\beta$  and the  
 294 seasonal coefficients  $s_i$ , with  $i = 0, \dots, 11$  (corresponding to December, January, ...,  
 295 November) through the model

$$Y_t = \beta t + \sum_{i=0}^{11} d_{t,i} s_i + \omega_t \quad (4)$$

296 where  $\omega_t$  is the stochastic error,  $s_i$  the seasonal coefficients, with  $i = 0, \dots, 11$  and  
 297  $d_{t,i}$  is a dummy variable defined as,

$$d_{t,i} = \begin{cases} 1 & \text{if } i = t \bmod 12 \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

298 The estimates of  $\beta$  and  $s_i$ , with  $i = 0, \dots, 11$ , are obtained through the least  
 299 squares method and are presented in Table 2. The analysis of the trend estimates  
 300 will be performed after the global adjustment of the model and in the clustering  
 301 procedure.

302 The second step of the modeling procedure adjusts the state space framework  
 303 to the observations detrended by the regression modeling,  $Y_t^* = Y_t - \hat{\beta}t$ . However,  
 304 data set has missing values in all monitoring sites in the period of 137 monthly  
 305 measurements (see Table 1) which varies between an 11% up to 27% rate of ob-  
 306 servations. This is a problem to the implementation of the KF algorithm since it  
 307 is performed based on the one step-ahead predictions. Thus, the linear model ob-  
 308 tained in the first step was considered as a baseline model to complete the original  
 309 database. This methodology is simple and removes the problem of missing values  
 310 and does not change the data structure. Nevertheless, this procedure implies a  
 311 more careful reading of the inferential results that may be achieved, especially if  
 312 the aim is to get accurate forecasts, which is not the case in this work. However,  
 313 if a more accurate methodology is needed, the Kalman smoother and the EM  
 314 algorithm can be combined to estimate missing values ([2]).

315 After this procedure, the parameters  $\{\mu, \phi, \sigma_\varepsilon^2, \sigma_e^2\}$  of the state space models  
 316 must be estimated for each site. Usually, in the state space framework the pa-  
 317 rameters are estimated through the likelihood estimation (ML) performed by the  
 318 EM algorithm assuming that the disturbances  $e_t$  and  $\varepsilon_t$  are normally distributed.  
 319 Table 3 presents parameters estimates from ML estimation. However, the analysis  
 320 of the innovations series,  $\hat{\eta}_t = Y_t - (\hat{\beta}t + \hat{s}_t \hat{X}_{t|t-1})$ , resulted in the state space mod-  
 321 els fitting showed that the Gaussian distribution is rejected in several cases (see  
 322 p-values of both the Kolmogorov-Smirnov and the Shapiro-Wilk tests in Table 3).  
 323 Thus, other approach was considered in order to avoid distribution assumptions  
 324 in the errors distributions.

325 A non-parametric approach was applied taking distribution-free estimators  
 326 (DF) based on the generalized method of moments (GMM) proposed by [7] for uni-  
 327 variate state space models and later generalized to multivariate state space models  
 328 in [16]. While the ML method assumes the normality of errors, which is not a  
 329 reasonable assumption in certain environmental variables ([18]), the distribution-  
 330 free estimators does not have distributions assumptions and, in addition, only  
 331 depend on the lags between observations. Table 3 presents parameters estimates  
 332 distribution-free estimators. Note that, in general, the ML method overestimates  
 333 the autoregressive parameters and underestimates the state equation error vari-  
 334 ance relatively to the DF estimators ([7]).

335 Thus, since we are interested in the extraction of structural components (trend  
 336 and seasonality) we take the mixed-effect state space model with the DF estimates.  
 337 Indeed, the filtered prediction of the DO concentration can be interpreted as a pre-  
 338 diction where several variations besides the structural components are minimized,  
 339 as the instrumental errors from the devices or human errors (six water monitor-  
 340 ing sites are automatic, INS, MIN, LOU, MAI, AGU and TOM). Additionally,

**Table 3** Estimates of the state space parameters and p-values of both Kolmogorov-Smirnov (K-S) and Shapiro-Wilk tests to the assumption of Gaussian distribution of innovations in the ML estimation.

site	ML				DF				ML		
	$\hat{\mu}$	$\hat{\phi}$	$\hat{\sigma}_e^2 \cdot 10^{-3}$	$\hat{\sigma}_e^2$	$\hat{\mu}$	$\hat{\phi}$	$\hat{\sigma}_e^2 \cdot 10^{-3}$	$\hat{\sigma}_e^2$	K-S	S-W	
AGA	0.986	0.824	0.430	0.455	0.987	0.330	4.592	0.151	0.070	0.003	
AGU	1.002	0.677	0.863	0.273	1.002	0.339	4.427	0.048	0.000	0.000	
ALF	0.991	0.719	0.881	0.844	0.990	0.300	9.380	0.313	0.015	0.324	
ALO	1.004	0.746	0.922	0.412	1.004	0.559	2.131	0.356	0.000	0.000	
ANT	0.992	0.776	0.936	0.855	0.991	0.361	15.58	0.288	0.046	0.020	
BUR	0.994	0.332	4.026	0.100	0.994	0.340	4.092	0.094	0.059	0.564	
CAN	0.993	0.756	1.330	0.964	0.992	0.299	15.472	0.320	0.200	0.008	
CAR	1.001	0.591	2.892	0.455	1.001	0.585	3.423	0.151	0.015	0.002	
EST	1.002	0.735	1.376	0.668	1.003	0.446	7.755	0.401	0.000	0.000	
FRO	1.001	0.534	2.842	0.171	1.001	0.493	4.051	0.096	0.000	0.002	
INS	0.994	0.715	0.836	0.383	0.994	0.328	3.969	0.127	0.023	0.022	
LOU	1.007	0.770	0.892	0.348	1.008	0.420	4.766	0.120	0.001	0.019	
MAE	1.002	0.697	1.814	0.340	1.003	0.440	4.968	0.149	0.006	0.008	
MAI	1.002	0.675	1.694	0.215	1.002	0.609	2.553	0.157	0.026	0.035	
MIG	0.992	0.729	1.180	0.823	0.991	0.303	8.515	0.358	0.015	0.066	
MIN	1.000	0.805	1.034	0.829	1.002	0.470	9.065	0.601	0.000	0.000	
PAM	0.996	0.844	0.888	0.610	0.998	0.344	6.938	0.262	0.200	0.020	
PER	0.990	0.629	1.273	0.309	0.990	0.381	4.055	0.176	0.004	0.000	
PIN	1.008	0.711	1.997	0.400	1.007	0.371	6.974	0.111	0.070	0.046	
POU	0.994	0.800	0.652	0.819	0.992	0.222	6.921	0.468	0.000	0.000	
RED	0.997	0.706	0.975	0.316	0.998	0.407	3.861	0.114	0.000	0.000	
REQ	1.008	0.823	1.128	0.630	1.008	0.439	9.235	0.334	0.200	0.147	
SER	0.997	0.340	7.076	0.001	0.997	0.381	6.161	0.066	0.000	0.025	
TOM	1.002	0.797	0.592	0.557	1.002	0.496	1.264	0.590	0.006	0.019	
VOG	1.000	0.615	2.442	0.334	1.000	0.353	6.703	0.067	0.000	0.001	
VOZ	1.003	0.788	0.539	0.815	1.003	0.155	11.729	0.269	0.001	0.006	

341 series of innovations of the fitted models have a behavior of a white noise process  
 342 validating models adjustments.

## 343 6 Discrimination procedures

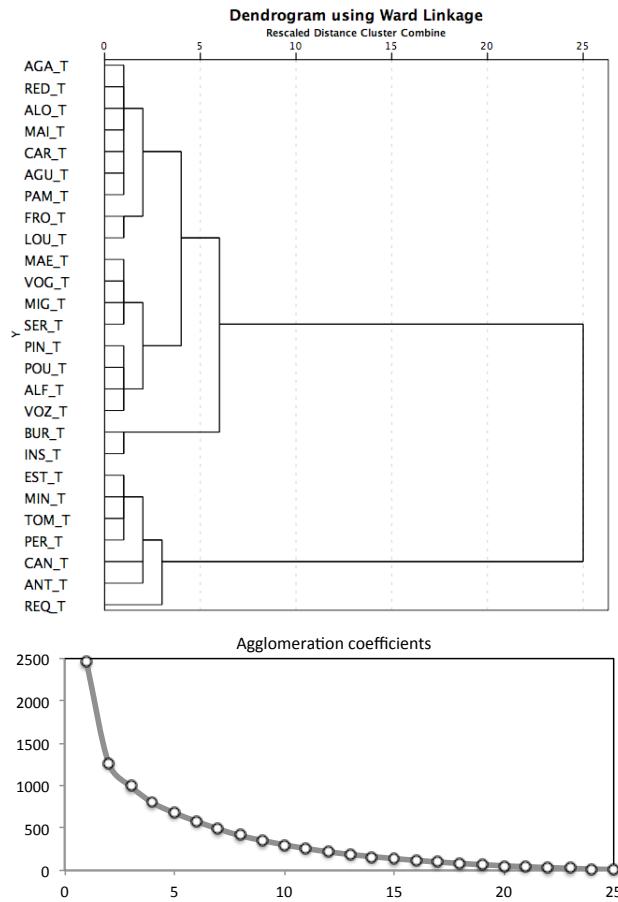
344 The Kalman smoother allows predicting the state  $X_t$  taking into account all available  
 345 data with the smallest mean square error within all linear estimators. These  
 346 predictions are used to compute smoothers predictions of the two main structural  
 347 components of the DO concentration: the trend and the seasonality, defined as  
 348 follows,

$$\hat{T}_{t|n} = \hat{\alpha} \hat{X}_{t|n} + \hat{\beta} t \quad (6)$$

349 and

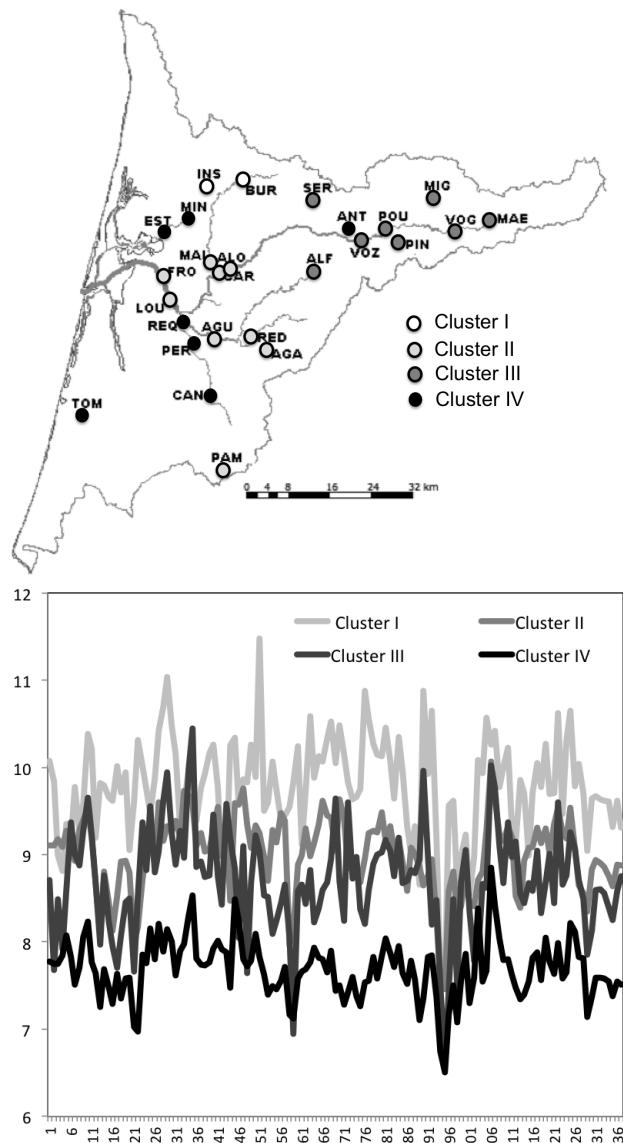
$$\hat{S}_{t|n} = \hat{s}_t^* \hat{X}_{t|n}. \quad (7)$$

350 Dynamic properties inherent in each site allow identifying patterns in order to  
 351 discriminate the water quality monitoring sites. This discrimination may not be  
 352 the same based on each component (trend and seasonality). Two procedures are  
 353 intended to identify patterns in each one of the structural components previously  
 354 predicted.



**Fig. 3** Dendrogram (top) and the agglomeration coefficients (bottom) of the extracted trend component based on the Ward's method

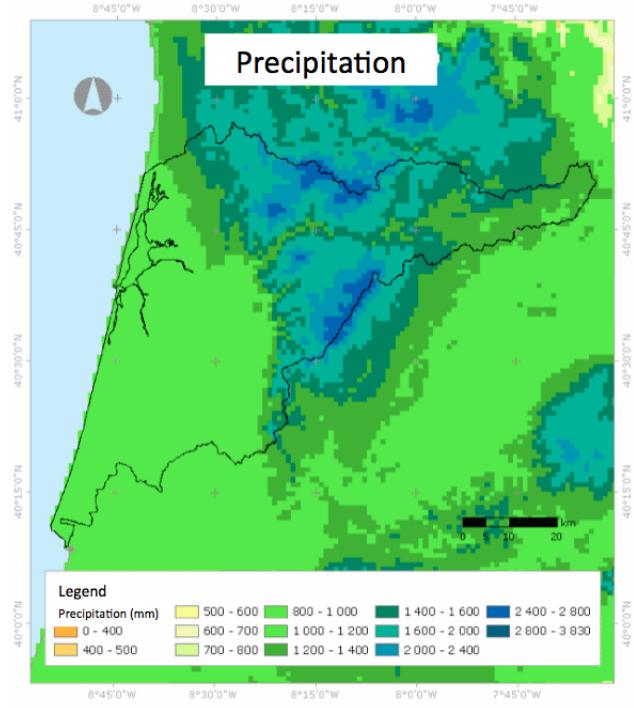
355 A hierarchical agglomerative clustering procedure is adopted since it is the  
 356 most common approach in discrimination and it is typically illustrated by a den-  
 357 drogram, which makes the analysis of results more easy. It is considered a hier-  
 358 archical agglomerative cluster analysis performed by means of Ward's method.  
 359 Ward's method uses a variance approach to evaluate the distances between clus-  
 360 ters, in an attempt to minimize the sum of squares of any two clusters that can be  
 361 formed at each step ([13]). Ward's minimum variance criterion minimizes the total  
 362 within-cluster variance. At each step the pair of clusters with minimum between-  
 363 cluster distance is merged. The initial cluster distances in Ward's minimum vari-  
 364 ance method are computed through the squared Euclidean distance.



**Fig. 4** Graphical representation of the solution with four clusters to the trend discrimination (top) and the monthly average within each cluster (bottom)

### 365 6.1 Discrimination using the trend component

366 Figure 3 represents the dendrogram and the agglomeration coefficients of the fil-  
 367 tered predictions of the trend component. Different levels were considered to cut  
 368 the dendrogram and the resulting hierarchical structures were analyzed in the  
 369 context of the basin. The solution that is considered acceptable and has an in-  
 370 terpretation in the basin context indicates four main clusters. This solution is



**Fig. 5** Total annual precipitation in mm (data is based on the SNIRH)

371 geographically represented in Fig 4 with the monthly average of the DO concen-  
 372 tration considering all the monitoring sites in each group.

373 On the one hand, this solution is reasonable since the number of clusters is  
 374 small and follows from the agglomeration schedule (see Fig 3). On the other hand,  
 375 the total annual precipitation in the region has an unequal distribution (see Fig 5).  
 376 As is well known, the hydrological conditions and the drainage areas are relevant  
 377 characteristics which influence the water quality. In this case, greater amount of  
 378 precipitation leads to a higher levels of DO concentration ([16]).

379 Considering that the cluster analysis produces homogenous groups of moni-  
 380 toring sites, a linear trend was adjusted to each cluster in order to estimate the  
 381 global linear trend of each group. Table 4 presents the least squares estimates  
 382 with the associated empirical 95% confidence intervals of both interceptions and  
 383 slopes of the global linear trends of each cluster. All clusters are discriminated by  
 384 the interceptions since all empirical confidence interval are disjuncted. Moreover,  
 385 this discrimination reflects the different average levels of each clusters. Clusters  
 386 II, III and IV have statistically significant negative slopes with similar empirical  
 387 confidence intervals while in the cluster I the slope estimate is not statistically sig-  
 388 nificant, i.e., in this cluster the average level of the DO concentration is constant.

389 Cluster I has only two monitoring sites: Captação Burgães (BUR) e Captação  
 390 Rio Ínsua (INS). This cluster corresponds to the sites with the highest DO concen-  
 391 tration levels, i.e., has the best water quality. In the other extreme, cluster IV with  
 392 the monitoring sites CAN, TOM, PER, REQ, EST, MIN and ANT has the overall

**Table 4** Least squares estimates with the empirical 95% confidence intervals of interceptions and slopes of global linear trends of clusters.

cluster	intercept		slope	
	estimate	C.I. 95%	estimate	C.I. 95%
I	9.850	[9.705, 9.994]	-0.00097	[-0.00279,0.00085]
II	9.088	[9.019, 9.158]	-0.00118	[-0.00206,-0.00030]
III	8.783	[8.698, 8.869]	-0.00113	[-0.00221,-0.00005]
IV	7.769	[7.680, 7.858]	-0.00114	[-0.00226,-0.00002]

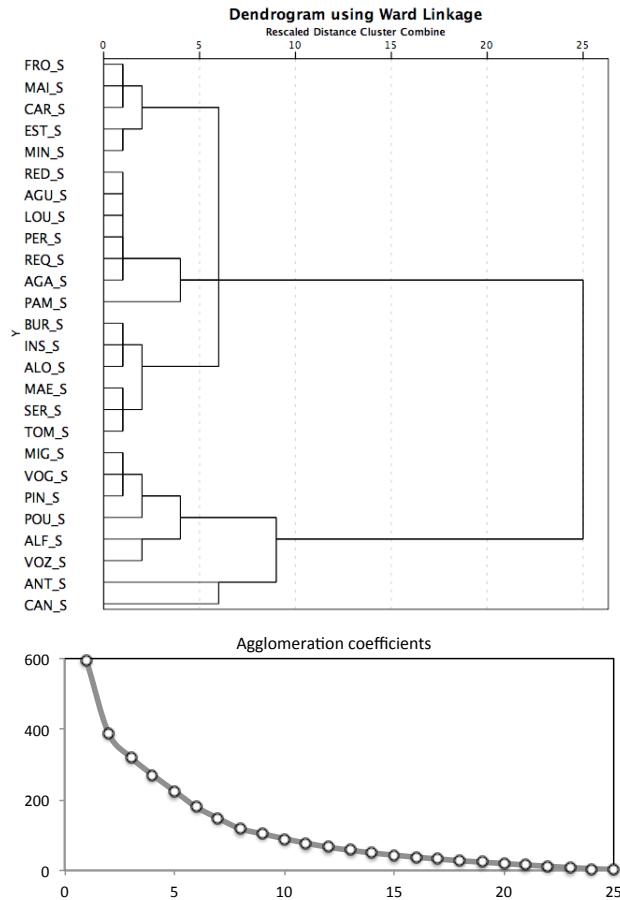
393 smallest values of the DO concentration. This cluster, which has the worst level of  
 394 the DO concentration, i.e. the worst water quality in view of the DO, contains a set  
 395 of monitoring sites located mainly in the industrial areas. In the site of Estarreja  
 396 (EST) there are several chemical industries, which can justified the poor quality of  
 397 surface water quality. For instance, the monitoring site of Ponte Minhoteira (MIN)  
 398 is located to a downstream from two industrial cities (São João da Madeira and  
 399 Oliveira de Azeméis) where there are a strong manufacture of shoes and associated  
 400 products. On the other hand, the majority of these sites correspond to locations  
 401 with a greater population density, thus, with a more intensive human activities.  
 402 In Ponte Requeixo (REQ) are located the main industrial activities of the city of  
 403 Aveiro, the capital district. The site that does not have these characteristics is the  
 404 Ponte Antim (ANT). This monitoring site is located in the municipality of São  
 405 Pedro do Sul, rural area and with a small population density. However, in this area  
 406 there is economic activities of poultry and lagomorphs slaughterhouses, which may  
 407 explain the lower DO concentration levels associated to pollutant discharges into  
 408 waterways.

409 Cluster II and cluster III are distinguished by the precipitation amount in the  
 410 respective drainage areas. Cluster II is located in the central area of the basin lo-  
 411 cated downstream from two relevant areas with high value of precipitation amount  
 412 while cluster III is located at upstream of the most rainier area, so is not influenced  
 413 by these high values of precipitation (see Fig. 5). These precipitation patterns are  
 414 associated to the topography of the region. Indeed, two locations with the highest  
 415 annual amount of precipitation in the region correspond to northeast of the Serra  
 416 da Freita mountain and to southeast of the Serra do Caramulo mountain.

## 417 6.2 Discrimination using the seasonal component

418 The discrimination of the water monitoring sites in order to the seasonal compo-  
 419 nent shows that there are less differentiation. Fig. 6 shows the dendrogram and the  
 420 agglomeration coefficients based on the Ward's method. It is very clear two main  
 421 groups: cluster I with the majority of the monitoring sites located in the west and  
 422 the remain sites in Cluster II concentrated to east (see Fig. 7). The discrimination  
 423 is evident in Fig. 7 where cluster I presents a seasonal component with a lower  
 424 amplitude instead of cluster II that has a higher annual range.

425 If we analyze the solutions with three or more clusters, the differences between  
 426 clusters are essentially in the summer months. Indeed, even in the solution with  
 427 two clusters the main differences are in the summer months. In cluster II, the  
 428 seasonal component has values near of -2 in the summer month instead of -1 in



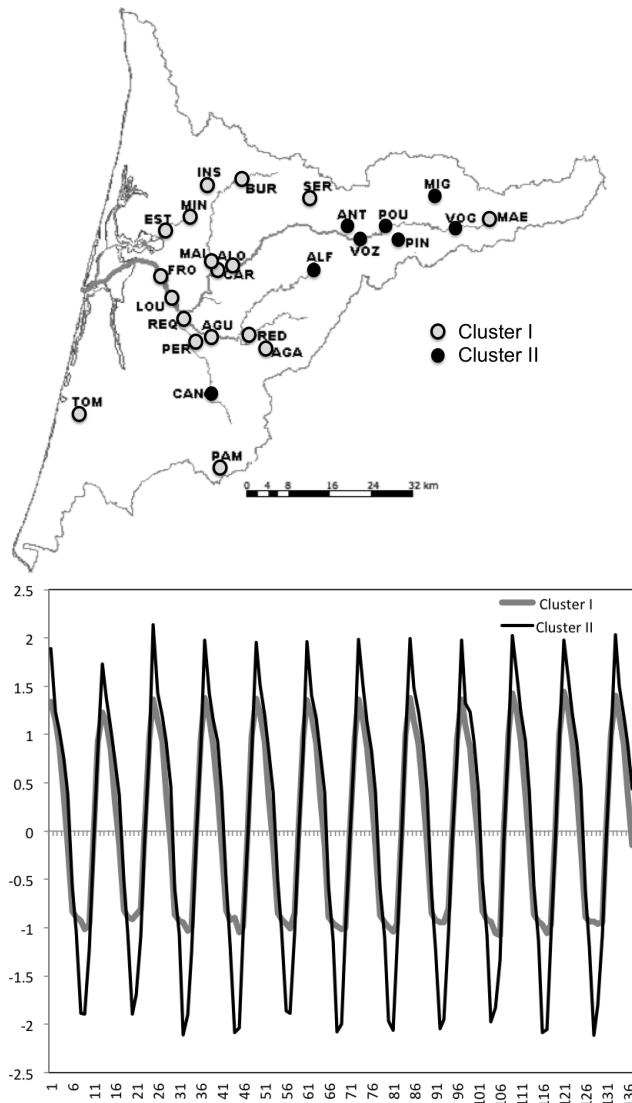
**Fig. 6** Dendrogram (top) and the agglomeration coefficients (bottom) of the extracted seasonality component based on the Ward's method

cluster I. However, this discrepancy is not so significant in the winter months since the seasonal component varies, approximately, between 1.5 and 2, respectively in clusters I and II.

On the one hand, if we want a parsimony solution, we consider that the solution with two groups is a reasonable discrimination solution mainly if we take into consideration that watershed of Vouga is a small hydrological basin. On the other hand, this solution is consistent with the annual average values of the real evapotranspiration in the region (see Fig. 8).

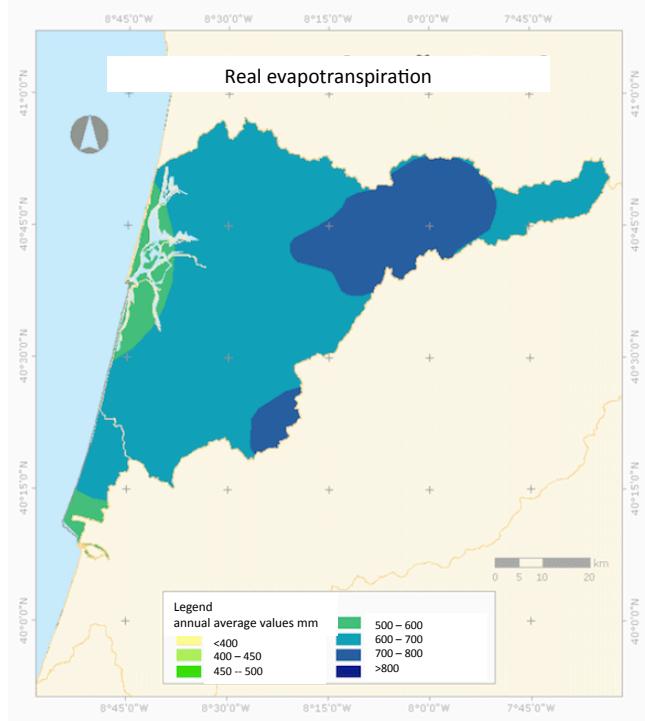
## 437 7 Conclusions

438 The linear mixed-effect state space approach shows to have versatility in order to  
 439 incorporate the usual trend and seasonality components of water quality variables.  
 440 This model combines the most useful properties of both multiple linear regression



**Fig. 7** Graphical representation of the solution with two clusters to the seasonality discrimination (top) and the monthly average within each cluster (bottom)

and state space models. This versatility accommodates a type of heterocedasticity which is present in the DO concentration at the same time that it takes into account the time correlation, of first order. The proposed models were fitted through a two-step parameter estimation procedure, which used the least square method combined with the state space parameters estimators. This approach is simple since combines parameters estimation procedures that are usually applied, having no additional complexity. On the other hand, the Kalman filter predictors provided predictions to the structural components as the trend and seasonality, which were used to classify the water monitoring sites. The filtered predictions



**Fig. 8** Annual average values of the real evapotranspiration in mm (data is based on the SNIRH)

of these components allowed to identify homogeneous groups of monitoring sites relatively to both trend/level and seasonal components. The level discrimination procedure provided four clusters with different levels. These clusters correspond to a four water quality levels in terms of the DO concentration. Mainly, the poor water quality is associated to industrial areas and with higher population densities while the major levels of the DO concentration are verified in the east of the hydrological basin, i.e., in the upstream locations or in areas with high levels of drained precipitation. **Besides, the cluster I which has the higher level of DO concentration shows a constant average level whereas the remaining clusters have negative trend.** The seasonal component is more related with environmental characteristics, as the real evapotranspiration, and less with human activities. An overall analysis of the models adjustments shows that the water quality has deteriorated in the sense of that the DO concentration has been decreasing slowly.

In addition to a global characterization of the evolution of water quality in the basin, the cluster analysis identified potential redundancies monitoring sites. Homogeneous groups of monitoring sites in terms of the evolution of DO were identified in both trend and seasonal components. The strategy that will be adopted to reduce the number of stations implies a combination between the statistical results and

470 both environmental and operational technical decisions, which must be  
471 framed in the political decision-making process.

472 **Acknowledgements** The authors would like to thank the Associated Editor and the anonymous  
473 reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. Authors were partially supported by Portuguese funds  
474 through the CIDMA - Center for Research and Development in Mathematics and Applications,  
475 and the Portuguese Foundation for Science and Technology ("FCT- Fundação para a Ciência  
476 e a Tecnologia"), within project UID/MAT/04106/2013.

478 **References**

- 479 1. Ahmad I, Mohmood I, Coelho JP, Pacheco M, Santos MA, Duarte AC, Pereira E (2012)  
480 Role of non-enzymatic antioxidants on the bivalves' adaptation to environmental mercury:  
481 Organ-specificities and age effect in *Scrobicularia plana* inhabiting a contaminated lagoon.  
482 *Environ Pollut* 163:218-225
- 483 2. Amisigo BA, Van De Giesen NC (2005) Using a spatio-temporal dynamic state-space model  
484 with the EM algorithm to patch gaps in daily riverflow series. *Hydrol Earth Syst Sc* 9:209-224
- 485 3. Arya FK, Zhang L (2015) Time series analysis of water quality parameters at Stillaguamish  
486 River using order series method. *Stoch Environ Res Risk Assess* 29(1):227-239
- 487 4. Ato AF, Samuel O, Oscar YD, Moi PA (2010) Mining and heavy metal pollution: assessment  
488 of aquatic environments in Tarkwa (Ghana) using multivariate statistical analysis. *J Environ  
489 Stat* 1:1-13
- 490 5. Brockwell, PJ, Davis RA (2002), *Introduction to Times Series and Forecasting*, 2th Edition,  
491 Springer-Verlag, New York
- 492 6. Cerqueira MA, Silva JF, Magalhães FP, Soares FM, Pato JJ (2008) Assessment of water  
493 pollution in the Antuã River basin (Northwestern Portugal). *Env Monit Assess* 142:325-335
- 494 7. Costa M, Alpuim T (2010) Parameter estimation of state space models for univariate ob-  
495 servations. *J Stat Plan Inference* 140:1889-1902
- 496 8. Costa M, Gonçalves AM (2011) Clustering and forecasting of dissolved oxygen concentration  
497 on a river basin. *Stoch Environ Res Risk Assess* 25:151-163
- 498 9. Costa M, Gonçalves AM (2012) Combining Statistical Methodologies in Water Quality  
499 Monitoring in a Hydrological Basin - Space and Time Approaches. In: Voudouris K and  
500 Voutsas D (ed) *Water Quality Monitoring and Assessment*, Intech, Croatia, pp 121-142
- 501 10. Costa M, Monteiro M (2015) Statistical Modeling of Water Quality Time Series - The  
502 River Vouga Basin Case Study. In: Lee TS (ed) *Water Quality*, Intech, Croatia, (in press)
- 503 11. Costa M, Monteiro M (2015) A mixed-effect state space model to environmental data.  
504 In: *Proceedings of Numerical Analysis and Applied Mathematics ICNAAM 2014*, In AIP  
505 Conference Proceedings, Rhodes, (in press)
- 506 12. Duan K, Xiao W, Mei Y, Liu D (2014) Multi-scale analysis of meteorological drought risks  
507 based on a Bayesian interpolation approach in Huai River basin, China. *Stoch Environ Res  
508 Risk Assess* 28(8):1985-1998
- 509 13. Everitt, B. S., Landau, S. and Leese, M. (2011), *Cluster Analysis*, 5th Edition, Wiley,  
510 Chichester
- 511 14. Finazzi F, Haggarty R, Miller C, Scott M, Fassò A (2014) A comparison of clustering  
512 approaches for the study of the temporal coherence of multiple time series. *Stoch Environ  
513 Res Risk Assess* DOI 10.1007/s00477-014-0931-2
- 514 15. Gonçalves AM, Alpuim T (2011) Water Quality Monitoring using Cluster Analysis and  
515 Linear Models. *Environmetrics* 22:933-945
- 516 16. Gonçalves AM, Costa M (2013) Predicting seasonal and hydro-meteorological impact in  
517 environmental variables modelling via Kalman filtering. *Stoch Environ Res Risk Assess*  
518 27(5):1021-1038
- 519 17. Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L (2000) Temporal evo-  
520 lution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal  
521 component analysis. *Wat Res* 34:807-816
- 522 18. Irincheeva I, Cantoni E, Genton MG (2012) A Non-Gaussian Spatial Generalized Linear  
523 Latent Variable Model. *J Agric Biol Envir S* 17:332-353

- 524 19. Khoshnevisan B, Bolandnazar E, Barak S, Shamshirband S, Maghsoudlou H, Altameem  
525 TA, Gani A (2014) A clustering model based on an evolutionary algorithm for better energy  
526 use in crop production. Stoch Environ Res Risk Assess 1-15. doi 10.1007/s00477-014-0972-6
- 527 20. Kokic P, Crimp S, Howden M (2011) Forecasting climate variables using a mixed-effect  
528 state-space model. Environmetrics 22:409-419
- 529 21. Lopes JF, Silva CI (2006) Temporal and spatial distribution of dissolved oxygen in the  
530 Ria de Aveiro lagoon. Ecol Model 197:67-88
- 531 22. Lopes JF, Silva CI, Cardoso AC (2008) Validation of a water quality model for the Ria de  
532 Aveiro lagoon, Portugal. Environ Modell Softw 23(4):479-494
- 533 23. MARETEC (2014) <http://maretec.mohid.com> (accessed 12 August 2014)
- 534 24. Rudolf A, Ahumada R, Pérez C (2002) Dissolved oxygen content as an index of water  
535 quality in San Vicente Bay, Chile (36 degrees, 450S). Env Monit Assess 78:89-100
- 536 25. Sánchez E, Colmenarejo MF, Vicente J, Rubio A, García MG, Travieso L, Borja R (2007)  
537 Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds  
538 pollution. Ecol Ind 7:315-328
- 539 26. Shumway RH, Stoffer D (2006) Time series analysis and its applications: with R examples.  
540 New York, Springer
- 541 27. Tsai JP, Chen YW, Chang LC, Chen WF, Chiang CJ, Chen YC (2015) The assessment  
542 of high recharge areas using DO indicators and recharge potential analysis: a case study of  
543 Taiwan's Pingtung plain. Stoch Environ Res Risk Assess 29(3):815-832
- 544 28. Varol M, Sen B (2009) Assessment of surface water quality using multivariate statistical  
545 techniques: a case study of Behrimaz Stream, Turkey. Environ Monit Assess 159:543-553
- 546 29. Zhang Y, Guo F, Meng W, Wang X-Q (2009). Water quality assessment and source iden-  
547 tification of Daliao river basin using multivariate statistical methods. Environ Monit Assess  
548 152:105-121
- 549 30. Zhang Y, Zhu C (2013) Water Quality Analysis in Jining City Using Clustering Methods.  
550 Nature Environment and Pollution Technology 12(4):685-690
- 551 31. Zhou J, Han L, Liu S (2013) Nonlinear mixed-effects state space models with applications  
552 to HIV dynamics. Stat Probabil Lett 83(5):1448-1456