Trend analysis of water quality series based on regression models with correlated errors

Abaurrea, Jesús [a], Asín, Jesús [a], Cebrián, Ana C. [a]

García-Vera, Miguel A. [b]

[a] *Dpto. Métodos Estadísticos. Universidad de Zaragoza (Spain)*

[b] *Confederación Hidrográfica del Ebro (Spain)*

Abstract

This work proposes a methodology for characterizing the time evolution of water quality time series taking into consideration the inherent problems that often appear in this type of data such as nonlinear trends, series having missing data, outliers, irregular measurement patterns, seasonal behavior, and serial correlation. The suggested approach, based on regression models with a Gaussian autoregressive moving average (ARMA) error, provides a framework where those problems can be dealt with simultaneously. Also the model takes into account the effect of influential factors, such as river flows, water temperature, and rainfall.

The proposed approach is general and can be applied to different types of water quality series. We applied the modeling framework to four monthly conductivity series recorded at the Ebro river basin (Spain). The results show that the model fits the data reasonably well, that time evolution of the conductivity series is non homogeneous over the year and, in some cases, non-monotonic. In addition, the results compared favorably over those obtained using simple linear regression, pre-whitening, and trend-free-pre-whitening techniques.

Preprint submitted to Elsevier *15 June 2011*

Key words: time series regression, ARMA error, water quality, trend analysis

1 Introduction

The assessment of long term water quality trends is a subject of growing interest. The 2000/60/EC directive of the European Parliament and of the Council sets the goal of achieving a "good status" for all Europe's surface waters and groundwater by 2015. According to the Water Framework Directive, Member States must establish surveillance monitoring programmes to provide information for the assessment of long-term changes in natural conditions and those resulting from widespread anthropogenic activity. In this framework, the detection and evaluation of the underlying trend due to anthropogenic activity is a primary issue. The identification of periods and locations where increasing pollution trends occur, would allow water management authorities to take adequate measures.

Our interest is in providing a statistical method able to detect the presence of a change in water quality, that can be attributable to anthropogenic behaviour. This is not an easy problem since those trends may be hidden by the effect of external factors, such as river flow, seasonality, water temperature or precipitation. We search for an approach that estimates and extracts the influence of these factors and, simultaneously, analyzes the presence of an underlying temporal trend. The statistical analysis can not assure that this trend is due to anthropogenic activities, but if the model properly eliminates all the possible environmental factors, it can be assumed that the remaining trend is related to human effect.

Alternative statistical methods such as non-parametric, time series and regression models have been used to assess trends in water quality (WQ in short) series. In this work it is proposed a methodology which tries to solve the disadvantages and limitations of those approaches. It enables us to take into account the influence of non human factors and also to deal with common problems in WQ series such as short records, missing data, outliers, irregularity in the measurement pattern and, particularly, serial correlation. It is based on the use of regression models with ARMA error and is successfully applied to the analysis of four monthly conductivity series. This model could also be used to simulate C behaviour under different environmental conditions. This analysis was performed at the request of the Confederación Hidrográfica del Ebro (CHE), the organization in charge of water management in the Ebro river basin (Spain).

The paper is organized as follows. In section 2 the most common approaches to the analysis of WQ series are reviewed, pointing out their advantages and limitations with regard to the problem of interest. In sections 3 and 4 the data and the model that are the basis of our approach are presented. Section 5 contains the core of the work, the description of how regression models with ARMA error are used to carry out the trend analysis of conductivity series and in section 6, the results for four conductivity series are discussed. The comparison of these results with those obtained from other approaches is shown in section 7 and the main conclusions are summarized in section 8.

2 Approaches to trend detection

In the WQ literature the most common approaches for detecting trends are non-parametric. Mann-Kendall or other tests of association between two variables are used to detect the existence of monotonic temporal trends. Different modifications have been made in order to deal with seasonality and serial correlation, see for example Lettenmaier (1976), Hirsch et al. (1982) and Hirsch and Slack (1984). Hirsch et al. (1991) examined some of the issues involved in estimating WQ temporal trends. Hipel and McLeod (1994) Chap. 23, offer a good review of the existing tests (seasonal Mann-Kendall, correlated seasonal Mann-Kendall, Spearman partial rank correlation, etc.). A more recent review on the identification of hydrological trends is provided by Khaliq et al. (2009). This revision includes methods for incorporating the effect of serial correlation, such as the pre-whitening approach proposed by Zhang et al. (2001), the trend-free-pre-whitening technique by Yue et al. (2002), the variance correction approach and the block boostrap. Other recent interesting references on this topic are Hamed (2008), who developed a new version of the Mann-Kendall test designed to account for the effect of scaling, and Bouza-Deaño et al. (2008), who analyzed the WQ trends in the Ebro river.

The main disadvantage of non-parametric analysis is the difficulty of detecting non monotonic trends. Moreover, most of the non-parametric tests do not allow the evolution of WQ parameters to be linked to influential factors; Libiseller and Grimvall (2002) developed a partial Mann-Kendall test for trend detection in the presence of covariates but it does not allow to model and quantify the effect of those covariates.

Since WQ data are sequentially observed and often serially correlated, time series techniques are used to model this kind of data. A wide review of time series models can be found in Hipel and McLeod (1994) and numerous references exist in WQ literature, such as Bhangu and Whitfield (1997), Worrall and Burt (1999), Ahmad et al. (2001) and Lehmann and Rode (2001), for example. However, time series models have some limitations, mainly that the data must be observed at equally spaced time intervals and that these models cannot easily deal with missing values, a common characteristic of WQ data. Standard time series models can not easily take into account the influence of external factors; Zetterqvist (1991) suggested the use of transfer functions to include covariate series.

Regression modeling is another tool to assess time trends. It enables modeling the effect of influential factors and it can easily deal with outliers, missing observations and irregular measure patterns. A thorough review of this technique applied to WQ series can be found in Hipel and McLeod (1994) Chap. 24, and some more recent articles are Antonopoulos et al. (2001), Scarsbrook et al. (2003) or Simeonova et al. (2003). More recently, Murdoch and Shanley (2006) used segmented regression analysis to assess WQ trends while Chang (2008) and Yang and Jin (2010) used spatial regression to incorporate the spatial correlation among the observations in the model estimation. However, regression models are not often used in the WQ literature since their assumptions (normality, constant variance, uncorrelation) are considered too restrictive for usual WQ data. Regarding this point, it must be considered that more general regression-based alternatives exist, such as the model suggested in this work, whose assumptions can be more easily met by WQ data.

3 Data

Electrical conductivity, also known as specific conductance and denoted by C herein, is a frequently used WQ indicator. Our data set consists of monthly conductivity series (microsiemens/cm) recorded at four rivers in the Ebro basin (North-East of the Iberian Peninsula). The interest of WQ studies in Mediterranean watersheds and in particular in the Ebro river basin is shown in the special issue edited by Barceló and Sabater (2010) that includes papers on the particular problems of WQ and resource scarcity in regions with a Mediterranean climate. The gauging stations were selected by CHE's managers in order to consider rivers of different flow, water use, basin size, etc., see Figure 1. Data for the Ebro river were gauged at Tortosa, close to its mouth, which represents the global behaviour of the basin; at Cinca river at Fraga; at Segre river at Vilanova; and at Guatizalema river at Peralta. The series were recorded from October 1980 to April 2003, except Peralta which started in October 1981. The instantaneous river flow F (m^3/s) and the water temperature WT (^{o} C) were also recorded.

Plots of the four C series are shown in Figure 2. The time points where the three variables C, F and WT are available are joined by segments to show the observations that could be used to fit a regression model with covariates F and WT. A smoothed signal showing the long term evolution (a lowess with a 35% window parameter) is also drawn. It can be seen that the lack of complete observations, mainly due to the missing flow values, is an important problem in Fraga and also in Vilanova and this problem makes the fit of a good model more difficult. The lowess signals reveal an increasing long-term trend in Tortosa while in Fraga and Vilanova after an increasing trend for the first third of the record the series appear to fluctuate around a constant value; in Peralta, a decreasing behaviour is observed from the late 90's. A thorough preliminary analysis revealed that the series in Tortosa shows an unlikely behaviour during the period 1988 to 1990, where all the observations are over the smoothed mean level. To confirm the different behaviour of that period, an indicator variable associated to the 3-year interval was introduced in the model; since it was significant at a 0.01 level, these anomalous observations were considered as missing values.

Some characteristics of the records are summarized in Table 1: the record length, the number of missing observations in C, F and WT, the number of complete observations in the three variables and the lag-1 correlation coefficients of the monthly C series $\hat{\rho}_C(1)$. The serial correlations appear to be significant and the highest values are observed for the Ebro (Tortosa) and Cinca (Fraga) rivers, which have the largest flows.

The mean conductivity varies as 918 and 904 $\mu s/cm$ for Tortosa and Fraga, and 550 and 503 $\mu s/cm$ for Vilanova and Peralta, respectively. The seasonal means and standard deviations are shown in Figure 3, which suggest that the series show some seasonality in the mean, although it is less evident in Peralta; the standard deviations do not seem to show a seasonal behaviour.

The annual and seasonal mean values of the flow are summarized in Table 2. This variable shows a clear seasonal behaviour with some common characteristics in the four rivers: lowest mean values (in italic in the table) are always observed in Summer, followed by the Autumn values; the highest mean values (in bold in the table) are observed in Winter except in Fraga, where it is observed in Spring.

4 Regression models for correlated data

As aforementioned, ordinary regression model is a tool capable of dealing with the usual problems in WQ data, except for the existence of serial correlation, a common characteristic in monthly data. In all the conductivity series of this study, residual correlation was detected after fitting an ordinary regression model.

It is well known that with serially correlated data, the least squares estimator is still unbiased but not BLUE (best linear unbiased estimator). Moreover, any selection procedure applied during the modeling process will be doubtful since, in the case of a positive correlation, the standard errors are underestimated and the conclusions obtained from the t or F tests may be incorrect. Some of the consequences of disregarding the error dependence in testing regression coefficients were shown by Vinod (1976).

A solution to the correlation problem, which is often applied in Econometrics and other fields, is some type of differentiation of the response variable, $y_t - y_{t-1}$ or $y_t - \rho(1) y_{t-1}$, where $\rho(1)$ is the lag-1 sample correlation coefficient. Other possible solutions are the use of the lagged response as a covariate or the more sophisticated ARMAX models, see Harvey (1990), that also include a linear combination of the last values of covariate time series.

A drawback of all these approaches is that, in order to get the serial correlation removed, the response variable is changed. The new response is a function of Y_t and its past values and this makes it difficult to obtain the time evolution of the original response. So, we opted for the use of a regression model with ARMA error as an adequate and useful tool to obtain the underlying trend in

WQ series.

4.1 Regression models with ARMA error

The equation of a regression model with ARMA error is:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{W} \tag{1}$$

where $\mathbf{Y}$ is the vector of the response variable, $\mathbf{X}$ the covariate matrix and $\mathbf{W}$ the error vector formed by values of a causal, zero-mean, ARMA(p,q) process, satisfying

$$\Phi(B)W_t = \Theta(B)Z_t \tag{2}$$

where $Z_t \sim N(0, \sigma^2)$ is an uncorrelated series of normal random variables with zero mean and constant variance, and $\Phi(B)$ and $\Theta(B)$ are polynomials of order p and q in the backshift operator B, $BZ_t = Z_{t-1}$.

Estimation and inference in these models are well known (e.g. Brockwell and Davis (2002) Chap. 6) and applications to various fields have been made. Reinsel et al. (1981) and Reinsel et al. (1988) analyze stratospheric ozone data for the detection of trends and Niu and Tiao (1995) generalize this approach to a space-time regression model also for ozone data. In Medicine, Greenhouse et al. (2006) develop a non-linear regression model with an ARMA error for fitting biological rhythm data which they apply to series of human core body temperature.

To our best knowledge this type of model has not been applied to WQ analysis. In the next section the modeling process of WQ data using this approach is presented, including a thorough description of its validation analysis; this is an important and necessary step of any modeling process which must not be underestimated.

5 Modeling process

5.1 The model

The regression model with ARMA error, succinctly described in (1), takes the following expression to represent the relationship between conductivity and the different covariates,

$$C_t = m_t(\beta^T)+f_1(WT_t;\beta^{WT})+f_2(F_t;\beta^F)+s_t(\beta^s)+f_3(D_t;\beta^D)+f_4(t\times D_t;\beta^{DT})+W_t$$

The vector of parameters β includes $\beta = (\beta^T, \beta^{WT}, \beta^F, \beta^s, \beta^D, \beta^{DT})$ where the superscripts indicate the covariates. The elements of the model are:

- C_t is the conductivity at time t.
- $m_t(\beta^T)$ represents the time evolution, modeled by a polynomial, $\beta_1^T t+\beta_2^T t^2+ \ldots + \beta_k^T t^k$, whose order is determined during the modeling process.
- $f_1(WT_t;\beta^{WT})$ and $f_2(F_t;\beta^F)$ are, respectively, polynomial functions of water temperature and river flow and, if necessary, in their lagged values, whose orders are also fixed during the modeling process.
- $s_t(\beta^s)$ denotes a seasonal term modeled by a sum of Fourier harmonics. Another way of describing seasonal effects is by functions of indicator variables, $f_3(D_t;\beta^D)$, linked to time periods such as seasons or months. Interaction of these variables and time terms, $f_4(t \times D_t;\beta^{DT})$, are also considered in order to enable the fit of different time evolutions in different periods of the year. Due to their seasonal character, the fitted WT and F functions help

to model the seasonal behaviour of C.

- W_t are the error terms having an ARMA(p,q) structure, see equation (2).

Invariance of the time definition. The zero of the time variable t is located in the middle of the record (January 1992) in order to avoid very high values, and the time unit is one year; for example the time variable for Tortosa, a series with 271 observations, is defined as t = -135/12, -134/12, ..., -2/12, -1/12, 0, 1/12, 2/12, ..., 135/12. Since any other time origin and unit would be equally adequate, a desirable property of the model would be its invariance to possible origin and scale time changes. To achieve this property in models including time interaction terms, a hierarchy principle must be applied: the covariate in the interaction term must be included in the model even if it is not statistically significant. For example, if D is not included, model (4) obtained by a linear transformation of t in model (3), is not equivalent to model (3),

$$Y_t = \beta_0 + \beta_1 \; t \times D_t + W_t \tag{3}$$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 \; (a + bt) \times D_t + W_t \\ &= \beta_0 + \beta_1 a D_t + \beta_1 b \; t \times D_t + W_t \end{aligned} \tag{4}$$

However, if D is included, models (5) and (6) are equivalent,

$$Y_t = \beta_0 + \beta_1 \; t \times D_t + \beta_2 D_t + W_t \tag{5}$$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 \; (a + bt) \times D_t + \beta_2 D_t + W_t \\ &= \beta_0 + \beta_1 b \; t \times D_t + (\beta_1 a + \beta_2) D_t + W_t \end{aligned} \tag{6}$$

Advantages of the use of a unified model with indicator variables. The usual approach in dealing with the seasonal character of WQ series is to fit separate models for each season by splitting the sample. However, the suggested model is able to simultaneously fit different structures for each season by using indicator variables (binary variables that identify the observations of a season or period of time) and their interactions with the covariates; see Weisberg (2005), section 6.2, for more details. It results in a more complex model but the set-up of a unified model has important advantages. First, it maintains the sample size that, otherwise, would be divided; moreover, it is more flexible since it allows us to consider not only the seasons but also other time periods formed by one, or more, months. Nevertheless, the most important benefit is that it provides a framework where the behaviour of the response in different seasons or periods can be compared. These comparisons are carried out using maximum likelihood ratio (MLR in short) tests that can objectively determine if the time evolution of the response in two periods is the same. The merging of the similar periods, using adequate indicator variables, will result in a reduction of the number of parameters and, consequently, in a simpler model.

5.2 Estimating the linear predictor

The parameter estimation is performed by restricted maximum likelihood, REML in short, see Harville (1977). Ordinary maximum likelihood could also be used but Cheang and Reinsel (2000) state that, for moderate sample lengths, the REML estimator is, generally, less biased than the MLE and leads to more accurate inference.

The covariate selection during the modeling process is mainly based on the values of the t-ratios, $\hat{\beta}$/standard error($\hat{\beta}$), associated to each covariate and, for more complicated hypothesis, on the MLR test. This test allows checking any hypothesis that involves the comparison of two models, one being a particular case of the other (nested models); for example, the inclusion of the two terms of a Fourier harmonic, the simultaneous effect of several terms, or the equality of two coefficients of the model, $H_0 : \beta_i = \beta_j$. The significance level in all the tests is 0.05.

The model selection is carried out in a systematic and iterative way, starting at the simplest model and then comparing models in a stepwise approach, until obtaining satisfactory fitting with the data. For each covariate, the hierarchy principle is applied in the following way: a polynomial function with the maximum order considered is first introduced; then, the greatest order term is checked and removed from the model if it is non significant. This process is repeated until a significant order term is found and then the simultaneous effect of the current fitted polynomial function is checked using a MLR test. Except when the sample size is too small, up to five order polynomials in time and four order polynomials in WT and F are considered.

A common problem in WQ series is the presence of influential observations (single cases or small group of cases that strongly influence the fit of the model) and outliers (observations that fall unusually out of the pattern in the relationship between the response and the covariates). In each step of the covariate selection the existence of influential observations and outliers is checked; the detected observations are removed from the sample to re-estimate the model since our aim is to obtain a model which characterizes the vast majority of the sample. The observations with a Cook distance higher than 1, or much higher than the rest, are considered influential and the observations with an absolute value of their standardized residuals much greater than 3 are identified as outliers; see Weisberg (2005) Chap. 9 and Jobson (1991) Chap. 3 for details on this topic.

5.3 Estimating the ARMA(p,q) error structure

In each step of the covariate selection process, given the values p and q, the coefficient vectors $\Phi = (\phi_1, \ldots, \phi_p)$, $\Theta = (\theta_1, \ldots, \theta_q)$ and β are estimated simultaneously by REML.

In the first step, the selection of p and q is based on the serial correlation structure of C, using the correlograms of the sample autocorrelation function (ACF) and the partial autocorrelation function (PACF), together with individual zero-correlation tests applied up to lag 24 (which corresponds to two complete seasonal cycles). In the later steps, the same tools are used to test if p and q values remain valid by analysing the residuals in that step. If there is some evidence that they are no longer adequate, a MLR test is used to compare the current model against the alternative suggested by the residual dependence structure.

5.4 Validation analysis

The validation analysis aims to check if the assumptions of the model (the relationship between the response and the covariates, the ARMA structure and the normality of the error terms) are correct since, otherwise, the conclusions obtained from the model could not be true.

Residuals. The validation analysis is based on the residuals e_t^* that result from filtering the raw residuals $e_t = y_t - \hat{y}_t$ with the estimated ARMA process. If the linear predictor and the ARMA structure are properly specified, the filtered residuals must be zero-mean normal uncorrelated variables and the usual regression diagnostics can be applied to them. When the ARMA error reduces to an AR(p), the filtering is simple,

$$e_t^* = e_t - \hat{\phi}_1 e_{t-1} - \ldots - \hat{\phi}_p e_{t-p}. \tag{7}$$

If the error structure is ARMA, the filtered residuals are

$$e_t^* = \hat{\Theta}(B)^{-1}\hat{\Phi}(B)e_t$$

and they should be calculated using for example a Kalman filter, see Brockwell and Davis (2002).

Checking serial correlation and normality. To check if the filtered residuals are uncorrelated, the correlograms of the ACF and PACF are plotted with an approximate confidence band. Individual zero-correlation tests are also performed up to lag 24 and, in addition, the Ljung-Box test (a portmanteau-type test) is calculated for lags 6, 12, 18 and 24. Normality is checked using the normal qqplot and the Shapiro-Wilk test.

Checking the linear predictor and homoscedasticity. The adequacy of fitted predictors is checked using scatter plots of residuals versus the fitted values, the covariates F, WT and time. In order to check if the seasonal behaviour is adequately modeled, the same plots are drawn for each season and, in the case of time, for each month. All these plots are also used to check homoscedasticity (constant variance); this graphical tool is complemented with the Breusch-Pagan test, see Cook and Weisberg (1993), which checks the hypothesis of constant error variance against the alternative that variance changes with the level of the fitted values or any covariate.

5.5 Missing observations

One of the advantages of regression models, including those with ARMA error, is that their estimation and inference can be easily performed when some observations are missing. As far as the validation analysis is concerned, the tools for checking the normality and the predictor equation can deal with missing observations but the analysis of serial correlation must be adapted.

Testing serial correlation with missing observations. The correlation function $\rho(h)$ has to be estimated using only the complete pairs of observations. This can significantly reduce the sample size, since each missing observation results in the elimination of up to two pairs, and make the samples available for each lag quite different. Because of this, the confidence intervals for $\rho(h)$ and the Ljung-Box tests are calculated using the real sample size n_h instead of the initial size n.

Temporal trend analysis in records with long missing periods. Another frequent problem in WQ series is the existence of long periods without data. These periods can make inadequate the use of a continuous polynomial function to represent the time evolution. To overcome this problem, we opted for fitting independent time polynomials in the isolated time periods. This is achieved by fitting interactions of the time terms with indicator variables marking the required periods. An example of this situation occurs in the Fraga series where there is a ten year missing period, from 1988 to 1997, due to the lack of river flow measurements. The use of two indicator variables linked to periods 1980-1988 and 1997-2003, respectively, allows fitting separate temporal trends without reducing the sample size.

6 Results

6.1 Fitted models

The most surprising result in the four fitted models was the complicated time polynomials needed to represent the temporal evolution in Tortosa in Winter (a 5-order polynomial function) and in Vilanova in Summer and Winter (a 4-order and 5-order polynomial functions, respectively), see Figure 4; in these plots sample observations are plotted over the fitted curve to distinguish the intervals where the estimation is based on fewer points. The evolutions suggest that some important information, whose effect is mixed with time, is missing in the linear predictors.

The non-monotonic behaviour of the temporal term fitted in Tortosa in Winter follows quite approximately the cycles observed in the evolution of Winter rainfall in the Eastern and Northeastern regions of the basin. The Winter and Summer behaviour in Vilanova could also be explained by the effect of the rainfall cycles in that region. So, two rainfall signals, $RainA$ and $RainB$, based on the anomalies of rainfall series observed in Tortosa-Tivissa and Lérida regions, respectively, were calculated.

To check if regional rainfall is an influential factor, the residuals from the models including only the significant covariates related to F and WT were plotted versus the corresponding Winter or Summer rainfall signals. The plots supported our hypothesis, see Figure 5 where Winter residuals for Vilanova are plotted versus $RainB$ together with a least square regression line. A new stepwise selection procedure, starting from the previous final models without the time terms and considering the rainfall covariates, was carried out for Tortosa and Vilanova.

Some summary measures of the four final fitted models are outlined in Table 3: the sample size, the number of outliers and influential observations eliminated during the modeling process, the estimated white noise standard deviation $\hat{\sigma}$ and, as a goodness-of-fit measure, the square correlation coefficient between the response and the fitted values, $Cor(C, \hat{C})^2$. The ARMA structure fitted to the error terms and the corresponding $\hat{\Phi}$ coefficients are shown in the last two rows. The covariates included in each linear predictor together with their coefficients and the p-values of their t-ratios are shown in Tables 4 (Fraga and Peralta) and 5 (Tortosa and Vilanova). It must be remembered that some non significant covariates according to t-test are kept in the model if they make part of a significant simultaneous effect or following the hierarchy principle, see section 5.2. The season indicator variables are denoted by Sp (Spring), Su (Summer), Au (Autumn) and Wi (Winter) and the variables associated with the months by the first three letters of their name, e.g. February is denoted by Feb. The fitted temporal trends for the four final models are shown in Figure 6.

The new model for Vilanova is quite simpler. The introduction of $RainB$ in Winter and Summer reduces the previous 5 and 4-order time polynomials to simple linear terms. Initially, two different slopes were fitted for Summer and Winter periods, 6.62 and 6.86 $(\mu s/cm)year^{-1}$ respectively, but finally both seasons were jointly modeled since a MLR test leads to conclude that there is not a significant difference between the slopes at a 0.05 level.

The model for Tortosa after adding the rainfall term is also simpler. The introduction of the $RainA$ signal in Winter reduces the previous five order time polynomial to a linear term with slope 11.24 $(\mu s/cm)year^{-1}$. The temporal slope for the Spring-Summer-November period in this model is 13.65 $(\mu s/cm)year^{-1}$ and since a MLR test does not reject the slope equality at the 0.05 level, the trends of both periods are jointly fitted. Hence, there is a linear trend in C except for the months of September and October, where no temporal evolution is detected. It is noteworthy that although the $Wi \times RainA$ and Wi terms in the Tortosa model, with p-values equal to 0.08 and 0.10 respectively, are not significant according to t-tests, the simultaneous effect of these two covariates is highly significant, with the p-value of the MLR test lower than 10^{-6}.

The period 1997-2003 in Fraga cannot be analyzed seasonally due to the scarce number of observations.

6.1.1 Comparing the fitted models

The river flow influences C in the four fitted models, although in different ways: linearly in Peralta, quadratically in Vilanova and as a third order polynomial in Fraga. The lagged flow is influential only in Tortosa where up to the second lag is significant. The greater complexity of this model is due to the fact that large rivers have much more persistence than do small streams and probably favored by the bigger sample size. The effect of WT is linear and positive in Fraga and Vilanova and also linear, but negative, in Tortosa. In Peralta, the temperature effect is quadratic.

As to seasonality, no harmonic term is included in any of the models since the seasonal effect is better described by F, WT and indicator variables. The fitted models show that a lower conductivity, with respect to the mean level of the series, is observed in Summer in Peralta, Vilanova and in Fraga during the period 1980-1987, with coefficients -52.7, -32.8 and -80.2 $\mu s/cm$, respectively. On the other hand, a higher conductivity level is observed in Vilanova in Winter and in Tortosa in the period AuDec (Autumn and December), with coefficients 106.0 and 216.1 $\mu s/cm$ respectively.

The fitted models reveal that the temporal evolution of C is not homogeneous throughout the year. In Vilanova there are linear increasing trends, with a common slope, in Summer and Winter, but C does not show any significant change in Spring nor in Autumn. In Tortosa there is also an increasing linear trend during all the year, except for September and October. In Peralta, an increasing linear trend is found in December and a similar evolution is observed between 1986 and 1997 in Spring and Autumn; however, the global evolution in these seasons is non monotonic and decreases from 1997. A decreasing linear trend from 1997 onwards is also detected in Fraga.

The observed increase of conductivity is not surprising since the use of water for agricultural purposes has grown in the Ebro basin during recent decades. The irrigated surface increased from 650 000 Ha in 1980 to 700 000 Ha in 2002. Concerning the simultaneous decreasing trends observed from 1997 onwards in the Guatizalema (Peralta) and Cinca (Fraga) rivers, a sound and common explanation could be the following. The Guatizalema basin, upstream of Peralta, is irrigated with water coming from the Cinca river through the Cinca channel, which starts downstream the system formed by El Grado and Mediano reservoirs. This system suffered an important and monotonic decrease of its water supply during the period 1995-2005. Apparently, if there is less water available, the farmers grow crops which demand less water and fertilizer and, given the lesser water return to the river, these crops are less pollutant. This explanation, provided by CHE experts, cannot be confirmed since agrarian surveys about crop history in these areas are not available.

With regard to the serial correlation, the fitted structure in the four models is an autoregressive process, which reduces to a simple AR(1) except for Peralta model.

6.2 Validation analysis

The four models have been fully validated with respect to serial correlation structure, homoscedasticity and the linear predictor specification (a possible misrepresentation of the covariates in the model formula or the absence of relevant terms). The normality of the residuals is also checked since it is a necessary assumption to make inference.

Serial correlation and normality. Table 6 shows the p-value of the lag-1 autocorrelation test for e_t^* and the number of lags, between 1 and 24, with a significant coefficient at the usual 0.05 level (the significant lags are shown in brackets). The p-values of the Ljung-Box test for lags 6, 12, 18 and 24, are also given and the ACF and PACF correlograms for Tortosa are shown in Figure 7 as an example. All the results provide a strong evidence on the uncorrelation of the filtered residuals. According to the normal qqplots and the Shapiro-Wilk test, see the p-values in Table 6, the normality of the filtered residuals can not be rejected in any model.

Linear predictor and homoscedasticity. The scatter plots of filtered residuals versus fitted values, water temperature, river flow and time are shown for the

Tortosa model in Figure 8; the linear predictor seems to be well specified since no pattern is detected in these plots. Concerning heteroscedasticity, although the variability seems to increase with the fitted values, see plot bottom right, the Breusch-Pagan test shows that this increase is not significant, see the p-value in Table 6. The seasonal structure also seems to be adequately modeled, see Figure 9. The same analyzes for the remaining models also turn out to be satisfactory.

6.3 Software

The modeling process is performed using the freeware R package, which can be downloaded from The Comprehensive R Archive network (CRAN) at *http://cran.r-project.org/*. The model estimation is carried out with the function *gls* from the library *nlme*. Some own functions and the package *car* are used in the validation analysis.

7 Comparison with other approaches

In order to compare the performance of the ARMA regression approach, the four C series were analyzed using other common approaches: a simple linear (SL) regression versus time, the pre-whitening (PW) approach proposed by Zhang et al. (2001) and the trend-free-pre-whitening (TFPW) technique by Yue et al. (2002). In the PW technique, if the lag-1 serial correlation $\hat{\rho}(1)$ is significant, the trend test is applied to the pre-whitened series $(y_t - \hat{\rho}(1)y_{t-1})$. Since this procedure affects the magnitude of the true slope in the original series, the TFPW tries to overcome this problem by calculating $\hat{\rho}(1)$ from the series detrended using the Sen slope estimator and, if the serial correlation coefficient is significant, the trend test is applied to the detrended pre-whitened series recombined with the Sen slope.

In Vilanova and Peralta, where the temporal evolution found by the ARMA regression approach is not the same in the different seasons, the three alternative methods are applied one season at a time. In Tortosa, since the ARMA regression finds a common time behaviour, except in September and October, the three methods are applied to the complete series. In Fraga two separate analysis for periods P1 and P2 are performed. The slopes estimated using the four approaches and their corresponding p-values are shown in Table 7. In all the cases, the null hypothesis is that there is no linear temporal trend, i.e. that the slope is equal to zero.

In Tortosa, the four approaches detect a significant linear temporal trend whose slope is underestimated if factors such as F an WT are not properly taken into account. The slope decreases from 13 with the ARMA regression to $8.1(\mu s/cm)year^{-1}$ with the TFPW method. In addition, the ARMA regression detects that this trend is not homogeneous over the whole year and identifies two months, September and October, with no trend.

In Vilanova, the four methods provide the same result for Spring and Autumn where no temporal trend is detected. In Summer and Winter, only the ARMA regression is able to detect an underlying positive temporal trend, after eliminating the wavy rainfall effect.

In Peralta, the ARMA regression detects a temporal trend in Autumn and Spring, described by a cubic polynomial, which decreases from about 1997 onwards. The other approaches, designed to identify monotonic trends, do not detect this behaviour. In Winter and Summer no trend is found by the methods that take into account the existing serial correlation. The possible mistakes resulting from not accounting for the serial correlation are evident in Winter, where the simple regression finds a statistically significant positive temporal trend.

In the P1 period of Fraga, the PW and the TFPW methods detect a positive trend, both with a p-value equal to 0.07, while the ARMA regression does not. These results are not comparable since the series used in the ARMA regression has 28 observations less, from a sample of 106, due to the lack of flow data. When the estimation and tests are applied to the same series, the PW and the TFPW methods do not find any significant trend, see values in brackets in Table 7. For the P2 period, 1998-2003, only the ARMA regression is able to detect a significant negative linear temporal trend in the C behaviour. The explanation of this fact is that the river flow shows a decreasing evolution during this period. The effect of a decreasing flow compensates the existing decreasing trend of conductivity and makes the PW and TFWP methods underestimate the real negative C trend with values -55.4 and -21.8 $(\mu s/cm)year^{-1}$ respectively, that are not significant. On the other hand, the temporal slope estimated by the ARMA regression, which takes into account the F effect, is -90.4 $(\mu s/cm)year^{-1}$ and significant.

The fit of temporal trends using the PW and the TFPW non-parametric procedures was carried out using the *zyp* library in R, developed by D. Bronaugh and A. Werner.

8 Conclusions

It can be concluded that the use of a regression model with a Gaussian ARMA error is a powerful tool to characterize, quantify and make inference on the temporal evolution of water quality parameters, under the difficult conditions often found in these series. Although the modeling process can be more time-consuming than other simpler approaches, it has important advantages as summarized below.

- It deals with the limitations of other common approaches to WQ trend analysis. More precisely,
  - It can be used in series with a serial correlation structure.
  - Because the proposed approach can characterize complex temporal evolutions such as non-monotonic trends, it offers an advantage over non-parametric trend analysis
  - It can be easily applied to series with an irregular recording pattern and missing observations. The Fraga series is an example of how it can be applied to data having long missing periods.
  - It can take into account any available influential factor.
- ARMA regression models are a flexible tool to analyze any temporal evolution. In particular,
  - The use of maximum likelihood ratio tests in combination with adequate indicator variables allows modeling different trends in different seasons, months or groups of months. It also allows describing these different evolutions in the simplest way by testing whether different periods have identical parameters.
  - The seasonal behaviour can be described by different and flexible tools:

  - harmonic terms, covariates with a seasonal character or indicator variables linked to the most convenient periods.
  - Using appropriate tests it is possible to determine which factors are influential and in which way they influence the response: linearly, through a polynomial function, through past values, etc.
- This approach uses the available information more efficiently, since the whole sample is used to estimate the model, whereas in other methods the sample must be split to undertake seasonal analysis or to distinguish different periods.
- The model validation process helps to improve the model in an iterative way since the residual analysis not only detects its failures but also suggests how to solve them. In our case, it helped detecting the influence of rainfall, a factor not considered initially.
- If influential natural factors, such as river flow, water temperature or rainfall, are not properly considered in the modeling process, as the ARMA regression does, the temporal trends due to anthropogenic activities can be hidden or distorted. The series of Vilanova in Winter and Summer and Tortosa are clear examples of this possibility.

As regards the results obtained in the four gauging stations in the Ebro basin, it is found that the temporal evolution of C series is not homogeneous within the year and sometimes even within seasons. The behaviour of C is not spatially homogeneous, although most of the series are stable or show an increasing trend most of the time; only Peralta and Fraga show a decreasing conductivity trend from about 1997. These differences are not surprising since the analyzed rivers present quite different characteristics regarding their flow, basin area and use of the water. From the six significant trends detected, five are linear and one is non-monotonic.

9 Acknowledgements

The authors want to acknowledge the 'Confederación Hidrográfica del Ebro' for its financial support and two anonymous reviewers and the Associate Editor for their helpful comments.

References

Ahmad, S., Khan, I. H., Parida B. P., 2001. Performance of stochastic approaches for forecasting river water quality. Water Research 35(18), 4261-4266.

Antonopoulos, V.Z., Papamichail, D.M., Mitsiou, K.A., 2001. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. Hydrology and Earth System Sciences 5(4), 679-691.

Barceló, D., Sabater, S., (Editors) 2010. Water quality and assessment under scarcity: Prospects and challenges in Mediterranean watersheds (Special Issue). Journal of Hydrology 383, 1-146.

Bhangu, I., Whitfield, P.H., 1997. Seasonal and long-term variations in water quality of the Skeena river at USK, British Columbia. Water Research 31(9), 2187-2194.

Bouza-Deaño, R., Ternero-Rodríguez, M., Fernández-Espinosa, A.J., 2008. Trend study and assessment of surface water quality in the Ebro river (Spain). Journal of Hydrology 361, 227-239.

Brockwell, P.J., Davis, R. A., 2002. Introduction to Time Series and Forecasting, second edition. Springer-Verlag.

Chang, H., 2008. Spatial analysis of water quality trends in the Han river basin, South Korea. Water Research 42, 3285-3304.

Cheang, W.K., Reinsel, G., 2000. Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. JASA 95, 1173-1184.

Cook, R. D., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. Biometrika 70, 1-10.

Greenhouse, J.B., Kass, R.E., Tsay, R.S., 2006. Fitting non linear models with ARMA errors to biological rhythm data. Statistics in Medicine 6(2), 167-183.

Hamed, K.H., 2008. Trend Detection in Hydrologic Data: The Mann-Kendall Trend Test under the Scaling Hypothesis. Journal of Hydrology 349, 350-363.

Harvey, A. C., 1990. The econometric analysis of time series. Cambridge, MA: MIT Press.

Harville, D. A., 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. JASA 72(358), 320-338.

Hipel, K.W., McLeod, A.I., 1994. Time series modelling of water resources and environmental systems. Elsevier.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18(1), 107-121.

Hirsch, R.M., Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20(6), 727-732.

Hirsch, R.M., Alexander, R.B., Smith, R.A., 1991. Selection of methods for the detection and estimation of trends in water quality. Water Resources Research 27(5), 803-813.

Jobson, J.D., 1991. Applied multivariate data analysis. Volume I: Regression and experimental design. Springer.

Khaliq, M. N., Ouarda, T. B., Gachon, P., Susham, L., St-Hilaire, A., 2009. Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. Journal of Hydrology 368, 117-30.

Lehmann, A., Rode, M., 2001. Long-term behaviour and cross-correlation water quality analysis of the river Elbe, Germany. Water Research 35(9), 2153-2160.

Lettenmaier, D.P., 1976. Detection of trends in water quality data from records with dependent observations. Water Resources Research 12(5), 1037-1046.

Libiseller, C., Grimvall, A., 2002. Performance of Partial Mann Kendall Tests for Trend Detection in the Presence of Covariates. Environmetrics 13, 71-84.

Murdoch, P.S., J.B. Shanley, 2006. Detection of water quality trends at high, median, and low flow in a Catskill Mountain stream, New York, through a new statistical method. Water Resources Research 42, 8407-8417.

Niu, X., Tiao, G., 1995. Modelling satellite ozone data. JASA 90(431), 969-983.

Reinsel, G., Tiao, G.C., Wang, M.N., Lewis, R., Nychka, D., 1981. Statistical Analysis of Stratospheric Ozone Data for the Detection of Trends. Atmospheric Environment 15 (9), 1569-1577.

Reinsel, G.C., Tiao, G.C., Ahn, S.K., Pugh, M., Basu, S., Deluisi, J.J., Mateer, C.L., Miller, A.J., Connell, P.S., Wuebbles, D.J., 1988. An analysis of the 7-years record of SBUV satellite ozone data: Global profile features and trends in total ozone. Journal of Geophysical Research 93, 1689-1703.

Scarsbrook, M.R., McBride, C.G., McBride, G.B., Bryers, G.G., 2003. Effects of Climate Variability on Rivers: Consequences for Long Term Water Quality Analysis. Journal of the American Water Resources Association 39(6), 1435-1447.

Simeonova, P., Simeonov, V., Andreev, G., 2003. Water Quality Study of the Struma River Basin, Bulgaria (1989 - 1998). Central European Journal of Chemistry 1(2), 121-136.

Vinod, H.D., 1976. Effects of ARMA Errors on the significance tests for Regression Coefficients. JASA 71(356), 929-933.

Weisberg, S., 2005. Applied linear regression, 3^{rd} Ed. Wiley.

Worrall, F., Burt, T.P., 1999. A univariate model of river water nitrate time series. Journal of Hydrology 214(1-4), 74-90.

Yang, X., Jin, W., 2010. GIS-based spatial regression and prediction of water quality in river networks: A case-study in Iowa. Journal of Environmental Management 91, 1943-1951.

Yue, S., Pilon, P., Phinney, B., Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes 16, 1807-1829.

Zetterqvist, L., 1991. Statistical estimation and interpretation of trends in water quality time series. Water Resources Research 27(7), 1637-1648.

Zhang, X., Harvey, K.D., Hogg, W.D., Yuzyk, T.R., 2001. Trends in Canadian streamflow. Water Resources Research 37(4), 987-998.

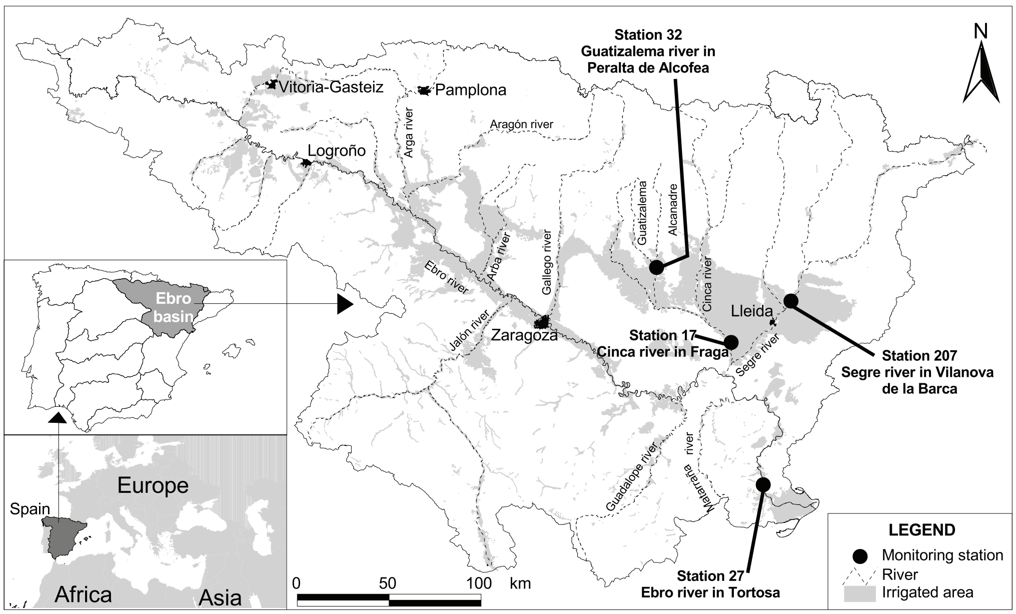

Fig. 1. Location of the Ebro river basin and the gauging stations.

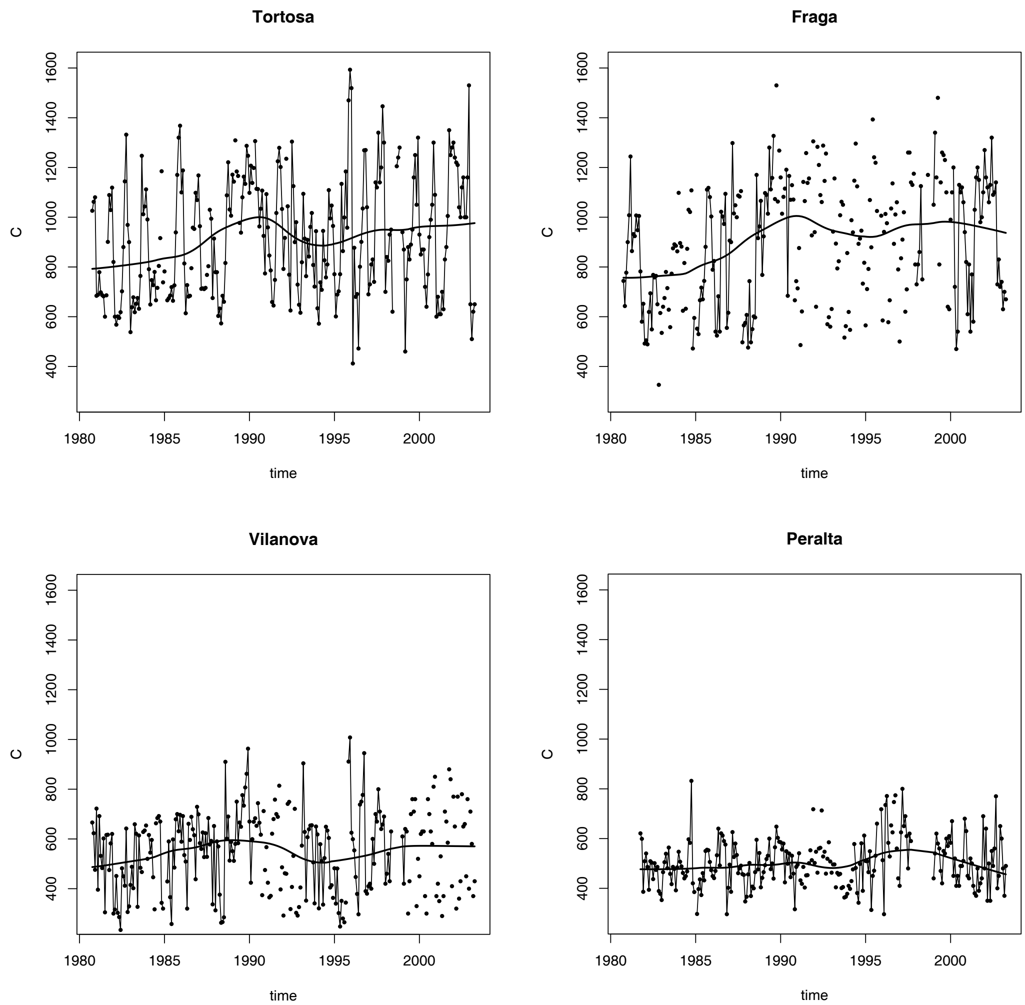

Fig. 2. Monthly series of conductivity (points) and long term evolution signals (lowess smoother). The segments joining the points indicate the periods where the three variables C, F and WT are available.

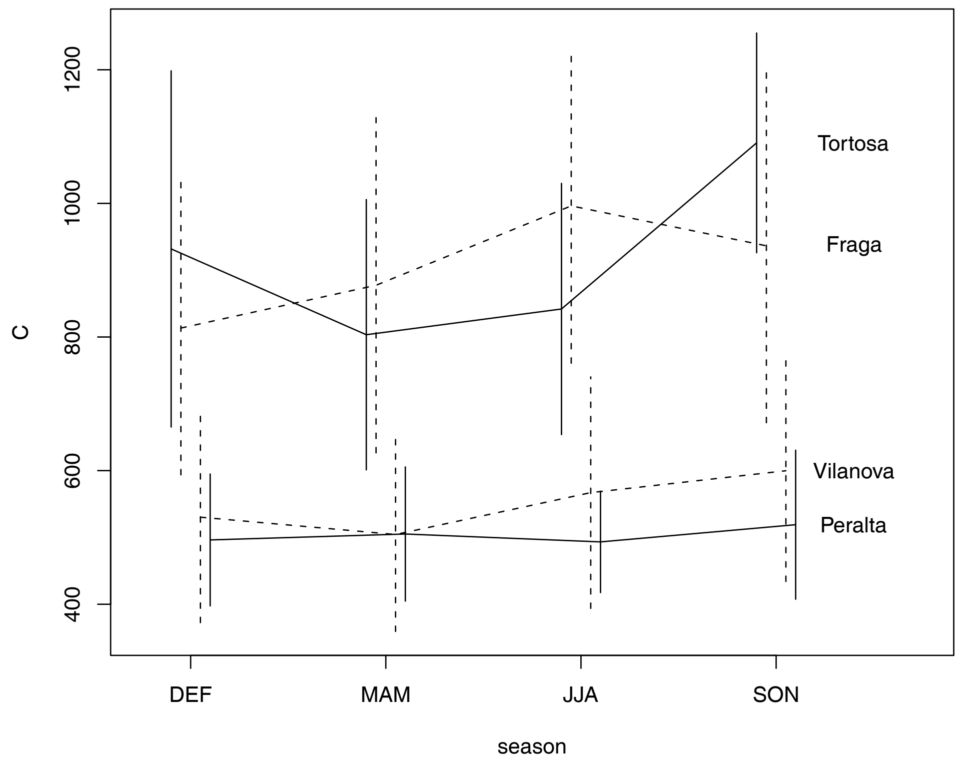

Fig. 3. Seasonal means ± standard deviation for the C series.

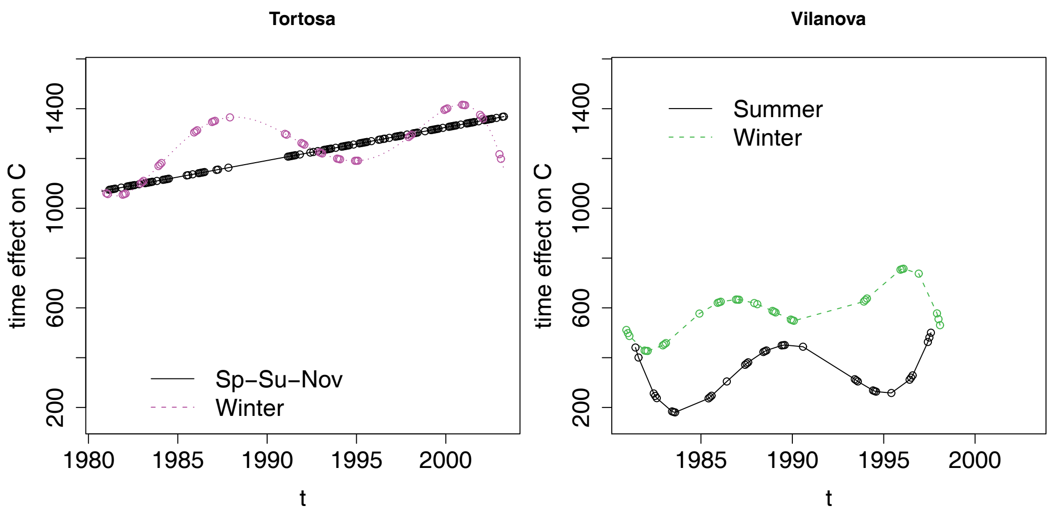

Fig. 4. Time evolutions fitted in the ARMA regression models (without rainfall signals), Tortosa and Vilanova.

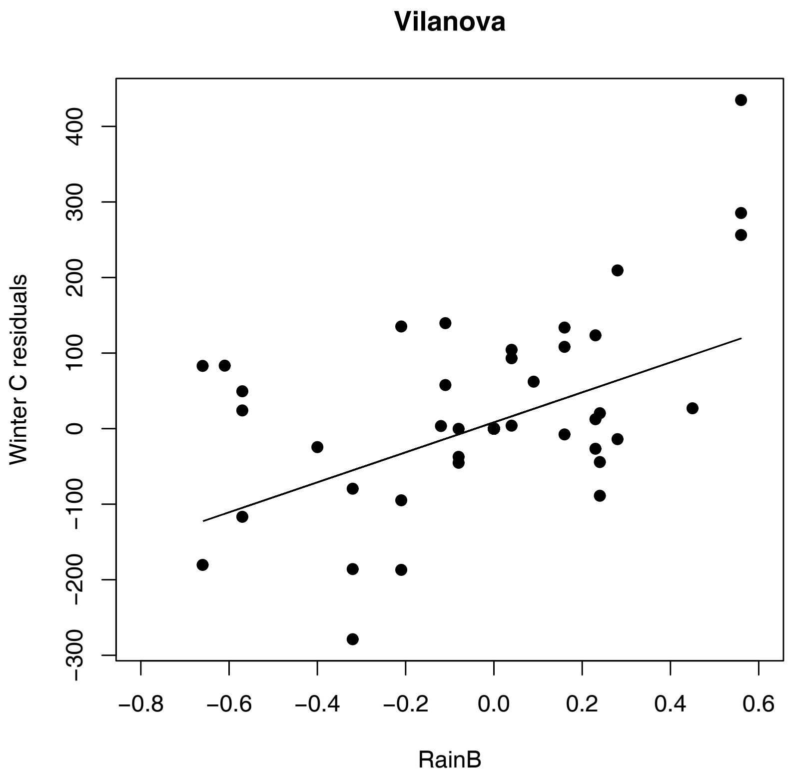

Fig. 5. Scatter plot of Winter conductivity residuals for Vilanova, from a model including only F and WT effects, versus the regional rainfall signal.

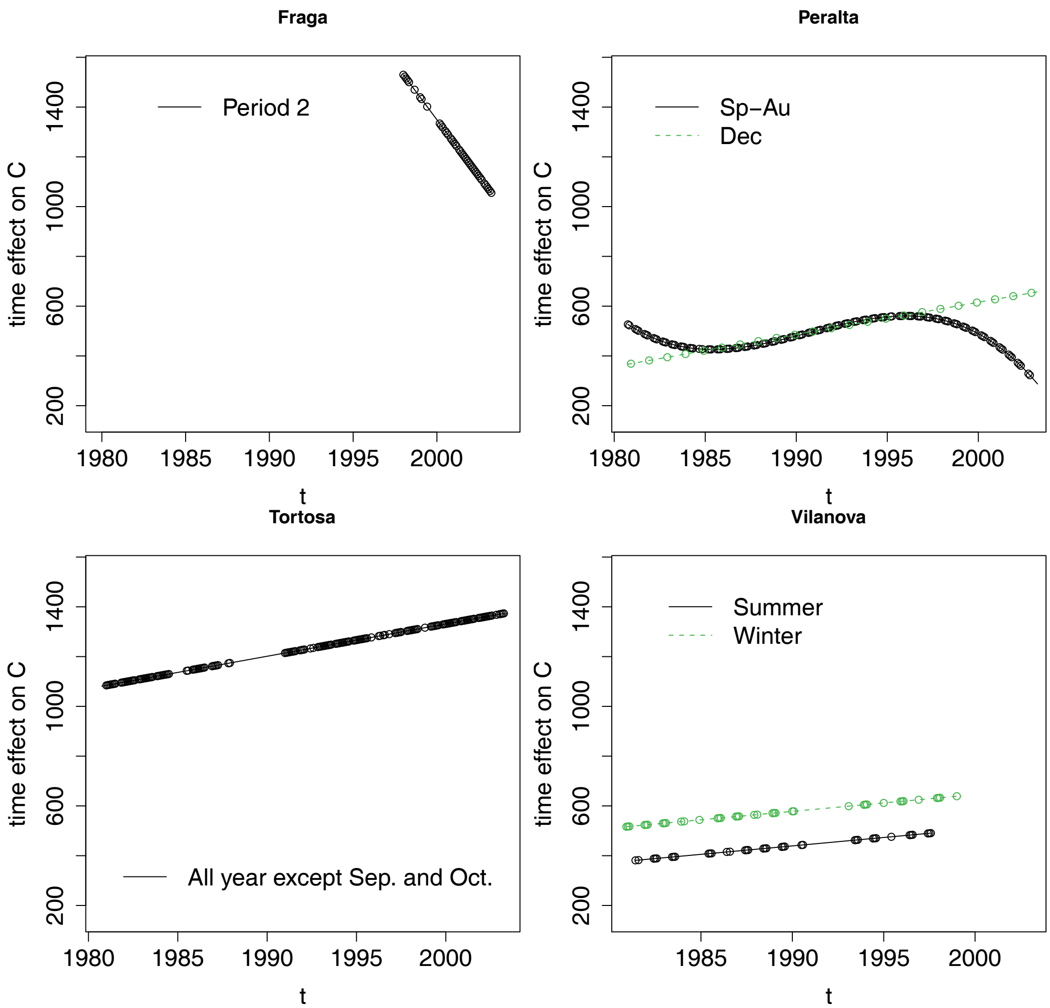

Fig. 6. Time evolutions fitted in the final ARMA regression models for the four series.

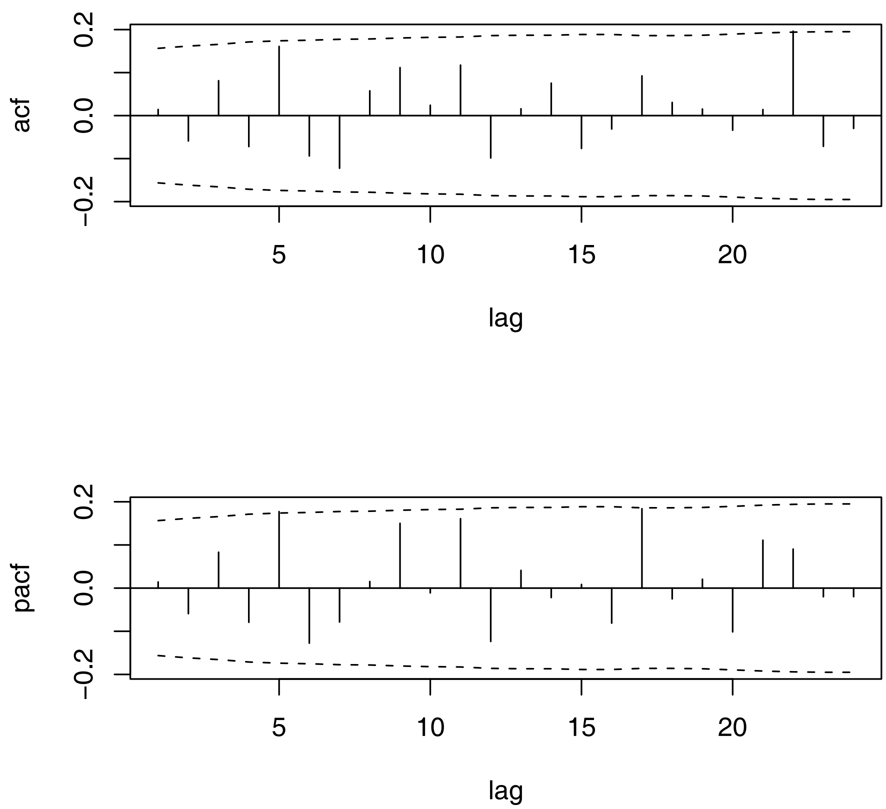

Fig. 7. Correlograms of the ACF and PACF of filtered residuals from the Tortosa model.

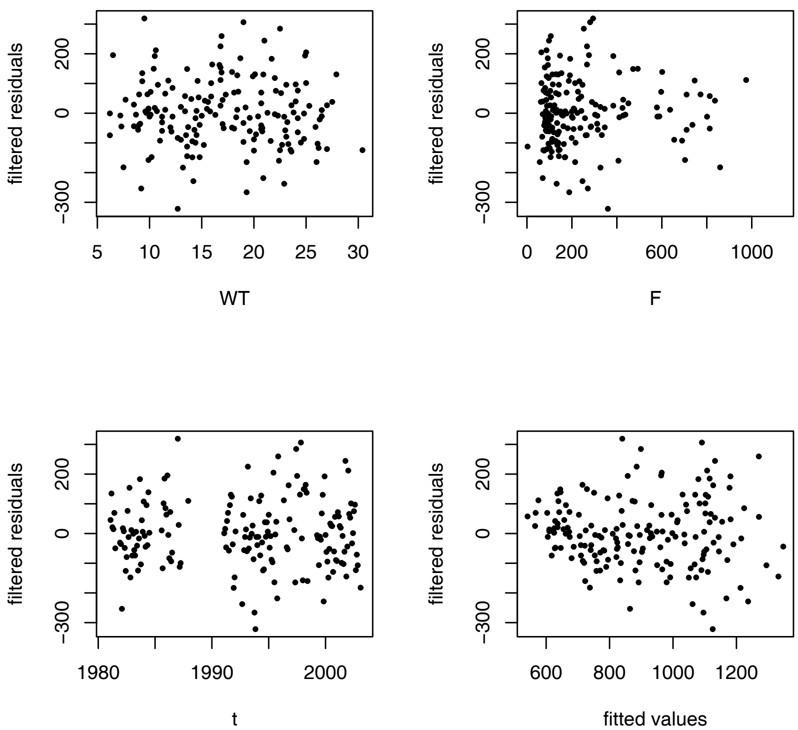

Fig. 8. Scatter plots of the filtered residuals versus fitted values and the covariates, water temperature, river flow and time for the Tortosa model.

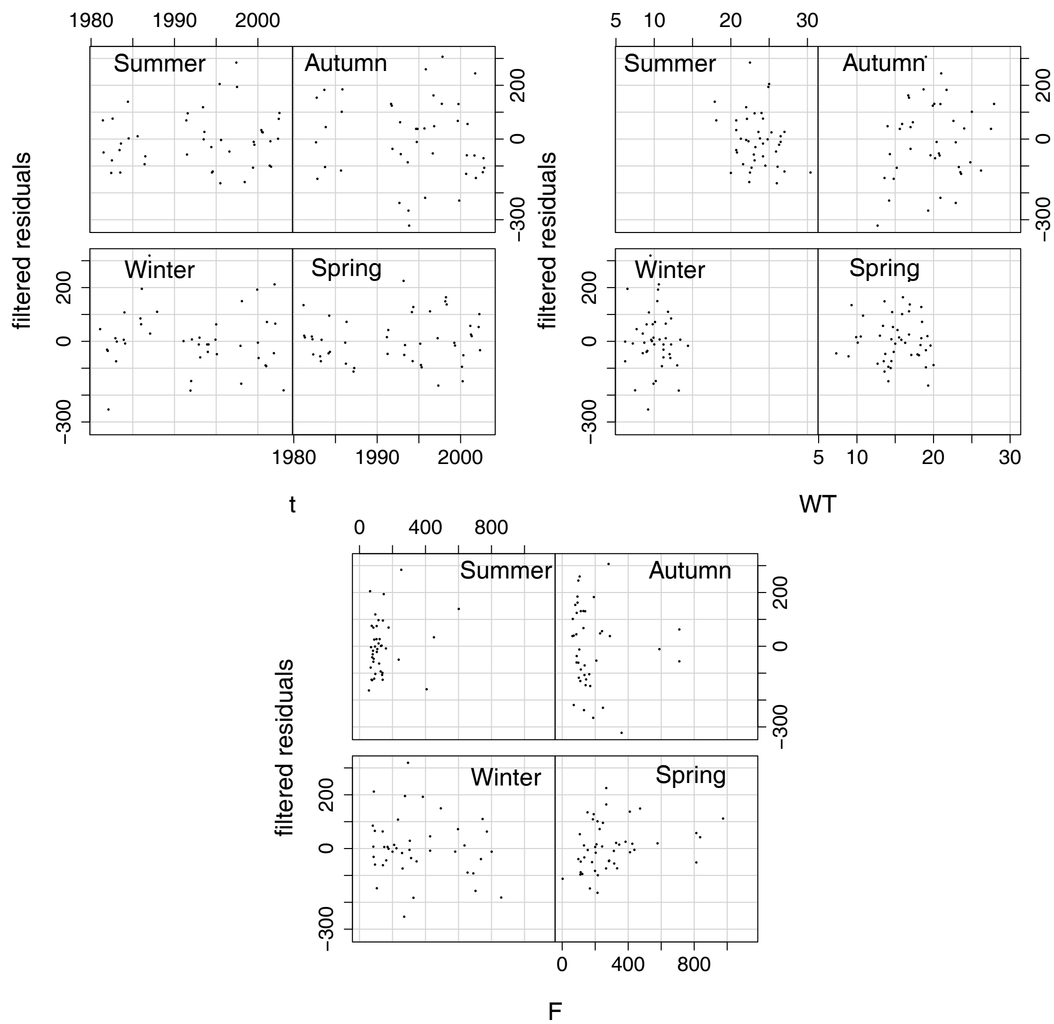

Fig. 9. Seasonal scatter plots of filtered residuals versus time, water temperature and river flow for Tortosa.

Table captions

Table 1. Some characteristics of the records of the four analyzed series.

Table 2. Annual and seasonal flow mean values; lowest seasonal values for each location are in italic and highest in bold.

Table 3. Summary measures of the fitted ARMA regression models.

Table 4. Coefficients and p-values of fitted linear predictors in Fraga and Peralta models.

Table 5. Coefficients and p-values of fitted linear predictors in Tortosa and Vilanova models.

Table 6. Summary of some validation measures based on e_t^* for the four final models.

Table 7. Estimated time slopes $(\mu s/cm) year^{-1}$ and p-values corresponding to null slope tests using four different approaches: the SL regression, the PW method, the TFPW method and the ARMA regression.

	Tortosa	Fraga	Vilanova	Peralta
Record length	271	271	271	259
# missing obs. in C	40	9	10	18
# missing obs. in F	18	143	100	59
# missing obs. in WT	4	10	10	18
# complete obs.	216	127	171	200
$\hat{\rho}_C(1)$	0.64	0.55	0.32	0.35

Table 1

	Tortosa	Fraga	Vilanova	Peralta
Annual mean F (m^3/s)	260.5	60.5	21.4	1.2
Winter mean F	**388.7**	61.1	**30.0**	**2.8**
Spring mean F	308.9	**68.9**	22.0	0.8
Summer mean F	*149.2*	*54.1*	*15.8*	*0.5*
Autumn mean F	184.2	56.6	16.6	0.6

Table 2

Series	Fraga	Peralta	Tortosa	Vilanova
Sample size	116	190	185	159
# influential obs.	5	8	10	7
# outliers	6	2	2	5
$\hat{\sigma}$	126.4	73.7	126.1	115.8
$Cor(C, \hat{C})^2$	0.76	0.42	0.71	0.48
error ARMA struct.	AR(1)	AR(2)	AR(1)	AR(1)
$\hat{\Phi}$	0.49	0.19, 0.27	0.29	0.20

Table 3

Fraga	$\hat{\beta}$	p-value
const.	1302.6	0.00
F	-18.1	0.00
F^2	0.1	0.00
F^3	0.0004	0.00
WT	10.3	0.00
P2$\times t$	-90.4	0.00
P2	769.8	0.00
Su$\times$ P1	-80.2	0.04

Peralta	$\hat{\beta}$	p-value
const.	481.4	0.00
F	-37.7	0.00
WT	9.2	0.02
WT^2	-0.2	0.07
AuSp$\times t$	17.9	0.00
AuSp$\times t^2$	-0.9	0.00
AuSp$\times t^3$	-0.2	0.00
AuSp	36.0	0.10
Dec$\times t$	12.9	0.00
Dec	30.7	0.10
Su	-52.7	0.04

Table 4

Tortosa	$\hat{\beta}$	p-value
const.	1191.8	0.00
F	-0.3	0.00
lag F	-0.4	0.00
lag2 F	-0.7	0.00
lag2 F^2	0.0005	0.00
WT	-6.9	0.01
SpSuWiNov$\times t$	13.0	0.00
SpSuWiNov	35.6	0.30
Wi$\times$RainA	140.7	0.08
Wi	53.4	0.10
AuDec	216.1	0.00

Vilanova	$\hat{\beta}$	p-value
const.	485.6	0.00
F	-5.6	0.00
F^2	0.03	0.00
WT	10.3	0.00
Su$\times$RainB	116.2	0.08
Su	-32.8	0.34
Wi$\times$RainB	173.7	0.00
WiSu$\times t$	6.8	0.01
Wi	106.0	0.00
May	-114.6	0.00

Table 5

Series	Tortosa	Fraga	Vilanova	Peralta
p-value of the correlation test	0.85	0.53	0.23	0.45
# signif. ACF lags (signif. lags)	1 (22)	1 (13)	0	1 (9)
Ljung-Box p-value: 6 lags; 12 lags	0.21; 0.19	0.13; 0.21	0.48; 0.25	0.74; 0.46
18 lags; 24 lags	0.39; 0.35	0.12; 0.08	0.25; 0.43	0.65; 0.66
Shapiro-Wilk p-value	0.50	0.10	0.19	0.96
Breusch-Pagan p-value	0.46	0.65	0.21	0.68

Table 6

Series	SL regression		PW method		TFPW method		ARMA regression	
	Slope	p-value	Slope	p-value	Slope	p-value	Slope	p-value
Tortosa	7.2	0.00	9.9	0.01	8.1	0.01	13.0	0.00
Vilanova, Spring	1.5	0.67	0.9	0.86	1.6	0.84	-	-
Vilanova, Summer	7.1	0.16	6.3	0.31	7.7	0.31	6.8	0.01
Vilanova, Automn	2.6	0.57	-1.4	0.77	2.6	0.79	-	-
Vilanova, Winter	-0.5	0.91	-6.1	0.35	-2.1	0.36	6.8	0.01
Peralta, Spring	1.1	0.57	1.7	0.68	0.7	0.69	cubic poly.	
Peralta, Summer	-0.4	0.79	-0.4	0.92	-0.1	0.92	-	-
Peralta, Automn	1.3	0.59	1.59	0.8	0.5	0.80	cubic poly.	
Peralta, Winter	6.8	0.00	0.67	0.8	6.8	0.53	-	-
Fraga, P1	22.7	0.01	26.2	0.07	24.9	0.07	-	-
(obs. with F value)	(17.0)	(0.07)	(24.4)	(0.24)	(18.7)	(0.22)	-	-
Fraga, P2	-19.4	0.37	-55.38	0.3	-21.8	0.26	-90.4	0.00

Table 7