Journal of Hydrology 400 (2011) 341–352

Contents lists available at ScienceDirect

Journal of Hydrology

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

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

Jesús Abaurrea [a], Jesús Asín [a], Ana C. Cebrián [a,*], Miguel A. García-Vera [b]

[a] Dpto. Métodos Estadísticos, Universidad de Zaragoza, Spain
[b] Confederación Hidrográfica del Ebro, Spain

ARTICLE INFO

Article history:
Received 13 April 2010
Received in revised form 3 December 2010
Accepted 25 January 2011
Available online 3 February 2011

This manuscript was handled by L. Charlet, Editor-in-Chief, with the assistance of Jose D. Salas, Associate Editor

Keywords:
Time series regression
ARMA error
Water quality
Trend analysis

SUMMARY

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 non-linear 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.

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 behavior. 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 cannot 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 behavior 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

* Corresponding author. Tel.: +34 976762885.
E-mail address: acebrian@unizar.es (A.C. Cebrián).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2011.01.049

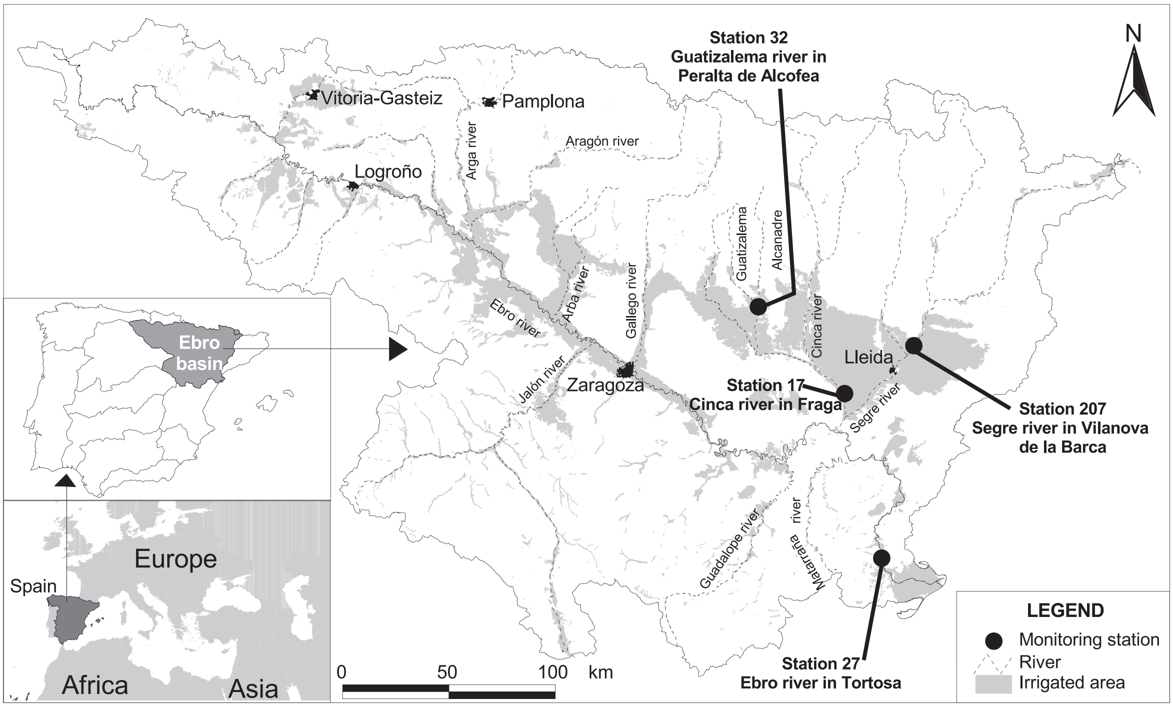

Fig. 1. Location of the Ebro river basin and the gauging stations.

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 cannot 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