A Methodological Framework for Characterizing the Spatiotemporal Variability of River Water-Quality Patterns Using Dynamic Factor Analysis

R. Aguilera[1,*], S. Sabater[1,2], and R. Marcé[1]

[1]*Catalan Institute for Water Research (ICRA), Scientific and Technological Park of the University of Girona, Girona 17003, Spain*
[2]*Institute of Aquatic Ecology, University of Girona, Faculty of Sciences, Girona 17071, Spain*

ABSTRACT. Water quality monitoring programs in many river basins have recorded data during several decades, but dealing with such environmental datasets is not an easy task. Uneven sampling frequency, missing observations and changes in monitoring strategies challenge the most basic data quality requirements and statistical assumptions of most time-series analysis methods. Lack of data consistency forces scientists to lean towards site-by-site approaches, avoiding simultaneous analysis of multiple time-series containing missing data. By using the appropriate tools and methods, however, common water quality patterns in a river basin can be identified and characterized in time and space. We introduce a collection of methodological steps for the detection and characterization of the spatiotemporal variability of river water-quality patterns in the context of global environmental change. Dynamic factor analysis (DFA) is used to extract underlying common patterns from sets of time-series with data gaps. The extracted patterns are further characterized using complementary methods such as frequency and trend analyses for the temporal dimension, together with regression and clustering analyses for the spatial dimension. We show how the application of these methods tackles the challenges of identifying patterns that act at different temporal and spatial scales, and we illustrate three case studies in Mediterranean basins where riverine nutrient patterns are unraveled and related to environmental drivers of change. Our methodological framework seeks to serve as a hypothesis-generation tool for further analyses of drivers of environmental change at the river basin scale.

Keywords: rivers, nutrients, time-series, spatiotemporal scales, dynamic factor analysis

1. Introduction

Global environmental change has intensified in the recent decades as a direct consequence of human activities (Meybeck, 2003). The detection of global change effects, defined as stating specific causal factor(s) and whether its origin relates to changes in environmental conditions and/or other external drivers (Hegerl et al. 2010), thus becomes crucial. The main objective in this context is to use available datasets and optimize tools and methods in order to identify change and its potential source or cause (Lyubchich et al., 2013). Furthermore, a key current challenge is to implement the use of complementary models and methods of different scales and complexity to pursue a full understanding of the interaction of global change phenomena (Huang and Chang, 2003) and their effect on the environment.

Freshwater systems are directly threatened by anthropogenic activities (Vörösmarty et al., 2010) in terms of both human water security and biodiversity. Consequently, the vulnerability of global water resources to ongoing environmental change should be a priority topic to assess (Vörösmarty et al., 2000), particularly at regional and global scales, taking advantage of knowledge obtained at the local scale and through case studies (Vörösmarty, 2002). Nutrient pollution remains one of the major threats to water quality worldwide, particularly in rivers and streams (Seitzinger et al., 2010), and it is therefore important to understand the spatial patterns of nutrient concentration and loading, their temporal variation, and the drivers behind their variability at the basin scale.

Time-series generally provide valuable information about physical, biological, or socio-economic systems (Ghil et al., 2002), and their interaction. Several methods can be used to extract key properties in time-series. Worldwide assessments of freshwater resources however rely to a great extent on fragmented data (Vörösmarty et al., 2010). The analysis of time-series is thus not straightforward since available datasets usually contain missing data and differing temporal resolutions, requiring the use of statistical techniques that produce equidistant observations and fill data gaps (e.g. interpolating missing values; Kandasamy et al., 2013). Long-term observations are also usually required to detect changes and diagnose causes. Furthermore, the assessment of trends and patterns in water quality is confounded by the interplay of many influencing

factors, acting at different temporal and spatial scales (Kundzewicz and Krysanova, 2010). These above mentioned constraints tend to narrow the choice of applicable methods and impose limitations.

Classical time-series analysis generally requires complete datasets, and often long time-scales (Zuur and Pierce, 2004; Bers et al., 2013). Methods like spectral analysis are limited to characterizing the spectral density to detect any periodicities in the data and to give information on cyclic patterns, and thus do not necessarily allow the user to identify complex patterns (e.g., non-monotonic trends, irregular cycles) embedded in a set of multiple time-series (Zuur et al., 2003a). Techniques like spectral and wavelet analyses, based on stationarity, need relatively long time-series and cannot cope well with missing values (Zuur and Pierce, 2004). Furthermore, most river water quality time-series can be nonstationary. Additionally, detecting water quality patterns and identifying their causes usually cannot be achieved through observations on a site-by-site basis (Ito, 2012), though some authors believe that this approach can be more successful than a regional one when studying water quality responses to climatic change (Benítez-Gilabert et al., 2010).

When observations of multiple time-series are considered, such as those recorded in different monitoring points in a river network, it is often reasonable to assume that there are common driving forces behind them (Márkus et al., 1998). Dynamic factor analysis (DFA) is a dimension-reduction method that extracts underlying common patterns in a set of multiple time-series (Zuur et al., 2003a; Figure 1), where the observations vary over time and cannot be dealt by conventional factor analysis (Geweke, 1977). DFA is able to treat time-series that have been recorded irregularly over time, or have short duration, by combining the state-space model framework and the Kalman Filter, and Expectation-Maximization methods (Harvey, 1989; Zuur et al., 2003a, Holmes, 2013). The extracted patterns (e.g., cycles and/or trends) are associated with factor loadings, which indicate the relevance that a given pattern has in a given time-series. These two end products, i.e., patterns and factor loadings, can then be analyzed in order to characterize the temporal and spatial variability of the extracted common patterns, thus facilitating the interpretation and the identification of potential drivers of change in the system. DFA has been widely used in econometric and psychological studies in the last decades, and has recently been applied to fisheries datasets (Zuur et al., 2003b; Zuur and Pierce, 2004; Vilizzi, 2012), and groundwater-quality data (Muñoz-Carpena et al., 2005, Kuo et al., 2013).

The present work describes the application of dynamic factor analysis and complementary methods in the identification of regional common patterns in river water-quality time-series, as well as the characterization of the spatiotemporal variability of such patterns in the context of global change, related mainly to climate and land use changes. The next section presents the different steps and procedures of our suggested methodological framework. Based on specific objectives and basin characteristics, a particular collection of such steps is

exemplified with particular case studies, using water quality monitoring data from three river basins in the Mediterranean region. We thus also provide examples of the versatility of the individual methods involved in this work, while discussing their main advantages and disadvantages and highlighting some future developments.
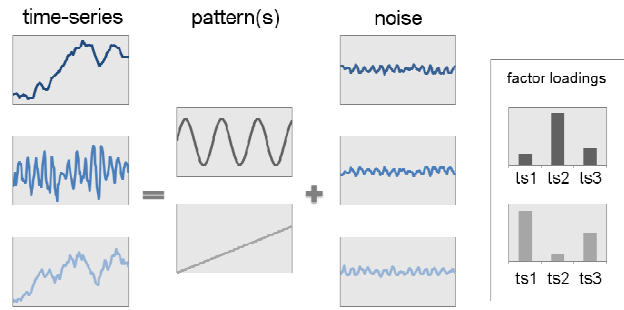

Figure 1. Dynamic factor analysis decomposes a set of *n* time-series into *m* common patterns and noise; factor loadings indicate the relevance of patterns for each original time-series (ts).

2. Methods

2.1. Rationale of the Methodological Framework and Workflow

The upcoming sections describe the full set of steps of our proposed methodological framework (illustrated in Figure 2) to detect common patterns and potentially attribute the effects of global change phenomena on river water-quality time-series. Key scripts in the R programming language (R Core Team, 2012), which is open to the scientific community, are included in the supplementary files in order to facilitate the implementation of our proposed methodology:

1) Data exploration is carried out for river water-quality datasets available from public water agencies or any other source. The main aim is to check the quality of the information and to remove any unreasonable values.

2) The Maximal Information Coefficient (MIC; Reshef et al., 2011 is used to find significant associations between variables and time in river monitoring large datasets. This optional step is intended to help in selecting relevant variables in large databases containing tens or hundreds of variables.

3) Subsequently, time-series with a significant relationship with time (based on MIC scores) are further explored by means of dynamic factor analysis (DFA; Zuur et al., 2003a), in order to extract patterns and their relevance (i.e., factor loadings) in a set of monitoring points.

4) Based on DFA results, the spatiotemporal variability of the identified river water-quality patterns is characterized using recently published time-series analysis tools for trend detection and for the identification of significant oscillations and relationships with streamflow.
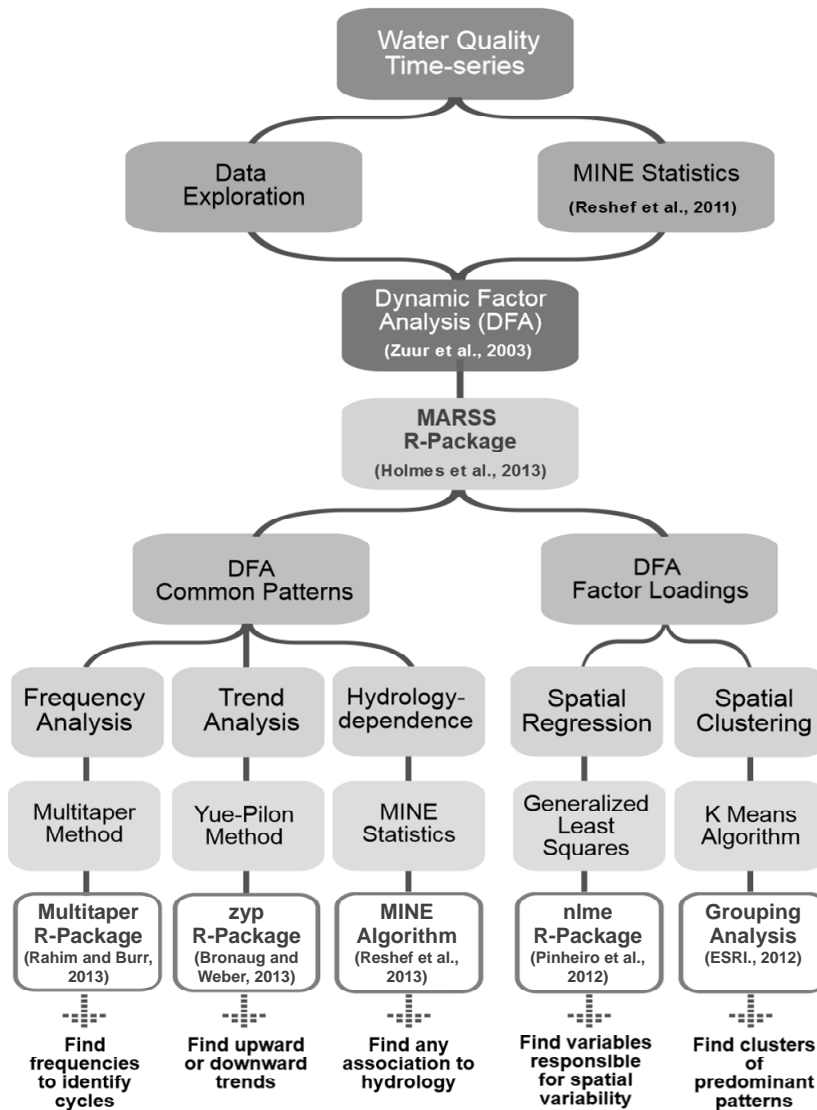
Figure 2. Relationships between individual components of our proposed methodological framework for unraveling spatiotemporal patterns of river water-quality at the basin scale.

5) Using the factor loadings as an expression of the importance of the extracted common patterns across river monitoring points, the drivers behind the observed behavior are identified using regression methods and cluster analysis.

2.2. Data Exploratory Analysis

The first step in our suggested methodological framework is an exploratory data analysis. We use the R v2.15 programming software (R Core Team, 2012) to standardize time-series, a step that allows comparison among time-series that might have different scales. Subsequently, a visual exploratory analysis of the series is performed using the plotting capabilities of R in order to determine the presence of outliers.

Potential outliers are detected and dealt with on an individual basis taking into consideration expected ranges of values for specific variables. Only outliers clearly indicating recording errors (i.e. misplacement of decimal point) were removed from the datasets in our case studies.

2.3. Exploring Significant Associations between Variable Pairs in a Large Dataset Using Mic

Being able to identify significant associations between pairs of variables in large datasets has increasingly gained importance within scientific studies. The Maximal Information Coefficient (MIC) method (Reshef et al., 2011) captures a broad range of associations, including but not limited to linear

relationships, and provides a score for functional relationships (i.e., hypothetical relationship that describes the effect of one or more independent variables on a dependent variable), and indicates the strength of the relationship (a value of 1 being the strongest). The basic strategy behind MIC is to calculate a measure of dependence for each pair of variables, ranking the pairs by scores and examining the top-scoring pairs (Reshef et al., 2011). The analysis is performed simultaneously for all possible variable-pair combinations, making it suitable for extracting information from large datasets. In our case, we used MIC to find significant relationships between river water-quality variables across time.

Besides MIC, the Maximal Information-based Nonparametric Exploration (MINE) indicators (Reshef et al., 2011) include several diagnostic statistics describing the shape of the relationship, such as a measure of monotonicity introduced as the Maximum Asymmetry Score (MAS), which can be used to characterize and select variables of interest to be further analyzed with more specialized or computationally intensive techniques (Reshef et al., 2011). We calculated the significant MIC scores at different levels of significance (0.05, 0.01, and 0.001) for any sample size using suitable power functions based on the p-values for various MIC scores at different sample sizes (available at http://www.exploredata.net/Downloads/P-Value-Tables).

Significant relationships between each river water-quality variable and time, tested simultaneously in all time-series included in the datasets, are identified based on resulting MIC scores and associated p-values. The focus is then set on the associations of variable-time pairs with the highest significant MIC scores that are present in the highest number of monitoring stations in the river basin. Apart from being a powerful additional exploratory analysis, the MIC analysis may be used to select the most significant variables containing temporal patterns, which can then be subjected to DFA for a more detailed study, if the analysis is not a priori focused on specific variables.

2.4. Extracting Common Patterns From a Set of Time-series Using DFA

DFA is used to detect common patterns in a set of time-series (Zuur et al., 2003a). DFA is based on the state-space model (Harvey, 1989), which treats each observed time-series as a linear combination of multiple state processes (Holmes et al., 2012), i.e., the time-series are modeled in an analogous way to linear regression as the sum of two terms: pattern(s) and error (Figure 1). DFA is applied to the database by means of the Multivariate Autoregressive State-Space (MARSS) R-package (version 3.4, Holmes et al., 2013) in order to explore common patterns in river water-quality time-series. The basis of the DFA formulation in MARSS includes a process model (x) and an observation model (y):

$$x_t = x_{t-1} + w_t \qquad w_t \sim MVN(0,Q) \qquad (1a)$$

$$y_t = Z_x + v_t \qquad v_t \sim MVN(0,R) \qquad (1b)$$

The observations (y_t) are modeled as a linear combination of hidden patterns (x_t) and factor loadings (Z) (Holmes et al., 2012). DFA is analogous to applying principal component analysis (PCA) to time-series: the DFA hidden patterns represent the principal components and the DFA factor loadings can be related to the loadings in PCA. The terms w_t and v_t represent the error associated with observations and hidden patterns, where Q and R are the covariance matrices for these error terms. The MARSS algorithm uses the Kalman filter and smoother, which operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state, to compute the best estimate for the hidden patterns given a set of predetermined assumed values for the unknown parameters Z, Q and R (Zuur et al., 2003a). The output of the Kalman filter is then used to compute the likelihood of the data given the better estimates for the set of parameters. MARSS estimates the maximum-likelihood via the Expectation-Maximization EM) method (Holmes, 2013).

Unlike common classical statistical approaches, DFA is able to deal with missing values, which are frequently encountered in time-series collected in river water-quality monitoring programs, and is also able to handle lagged correlation structures (Márkus et al., 1998). Depending on the complexity of the model and the number of time-series involved, however, the DFA method can be computationally intensive. Analyses included in this study were performed in a Linux-based High-Performance Computing Cluster, which allowed the simultaneous simulation of several DFA models. Nevertheless, computational times extended beyond weeks for analyses that included the longest records and the largest number of time-series.

Computational load is exacerbated by the fact that several algorithm settings can be adopted. Analyses may include several combinations of the number of common trends that can be potentially extracted (m) and different structures for the variance-covariance matrix for the observation errors. The number of common trends and the error structure affects the performance of the analysis and thus of the fitted model. Fortunately, the DFA implementation in the MARSS R-package already includes a best-model selection approach based on the Akaike Information Criterion (AIC).

The resulting best-fits, which are the linear combination of the extracted common patterns based on the corresponding factor loadings in each monitoring point (i.e., time-series), can be plotted against the original time-series in order to determine the extent of model performance. Best-fit plots, as well as variance values in the diagonal of the R matrix, can provide additional information about any monitoring point(s) that might not have been well represented by the best model, and thus identify any potential cause(s) behind such misrepresentation, whether it is based on data restrictions (e.g., low sampling frequency) or pattern behavior. The extracted common patterns and associated factor loadings can be studied in detail in order to characterize their spatiotemporal variability and

their relationship to global change phenomena as described in the following sections.

2.5. Characterization of DFA Resulting Common Patterns in Multiple Time-Series

2.5.1. Frequency Analysis

A time-series describes values of a response variable over time (Ghil et al., 2002). The trajectory of the response variable can be converted to a series of frequencies, based on how often the observed values fall within ordered ranges of magnitude during a specified period of time. Both deterministic and stochastic processes can, in principle, be characterrized by a function of frequency (instead of time), called the spectral density (Ghil et al., 2002). The interpretation of the spectral density function as the variance of the time-series over a given frequency band provides an intuitive explanation for its physical meaning (Shumway and Stoffer, 2010). Moreover, the concept of regularity of a series can be best expressed in terms of periodic variations of the underlying phenomenon that produced the series, expressed as Fourier frequencies (Shumway and Stoffer, 2010). The objective of time-frequency analysis in our methodological framework is to obtain significant frequencies present in the extracted common patterns that could be potentially linked to climatic or phenollogical signals (i.e., those related to climate influence on biological cycles) in our study basins.

The Multitaper Method (MTM) provides useful tools for the spectrum estimation and signal reconstruction of a time-series whose spectrum may contain both broadband (i.e., representing an irregular but continuous process) and line (i.e., periodic or quasiperiodic process) components (Ghil et al., 2002). MTM is non-parametric, since it does not use an apriori, parameter-dependent model of the process that generated the time-series. MTM reduces the variance of spectral estimates by using a small set of tapers (i.e., set of spectral windows), rather than the unique data taper or spectral window used in classical methods (Ghil et al., 2002).

MTM is applied to the resulting DFA patterns by means of the multitaper R-Package (Rahim and Burr, 2013). The time-bandwidth parameter (nw), which is the product of the temporal duration and spectral width (in frequency space) of a wave, and the number of tapers (k) are chosen by the method itself based on the characteristics of the input data. The constructed orthogonal tapers used in this method minimize the spectral leakage due to the finite length of the time-series (i.e., sharp discontinuities in the frequency domain caused by a limited observation interval), thus enabling the computation of independent estimates of the power spectrum (Ghil et al., 2002).

For each of these MTM analyses, the F test is also computed. The F distribution is a right-skewed distribution most commonly used in the analysis of variance. The F test is roughly based on the ratio of the variance captured by the filtered portion of the time series, and the resulting F statistic essentially represents a signal-to-noise ratio. Formally, the F test is a F variance-ratio test with 2 and $2k-2$ degrees of free-

dom (df; defined as *"dofs"* in the multitaper R package) for the significance of the estimated line component (Thomson, 1982). The higher the value of the F statistic, the higher is the significance of a specific frequency.

Based on the table of critical values for the F distribution at the $p = 0.05$ level of significance, critical values for frequencies in this atudy are obtained. This is done by looking at the F ratio and by identifying the critical value according to the rows and columns of the F table based on previously computed degrees of freedom for each frequency. If the MTM obtained value of F is equal or larger than this specific critical value, the result is significant at the level of $p = 0.05$.

2.5.2. Trend Analysis

The presence of positive serial dependence in the observations increases the probability of rejecting the no-trend hypothesis and can thereby cause trends to be detected that would not be found significant if the observations in the series were independent. Prewhitening will prevent the false detec0 tion of a non-existing trend, without a significant power loss in identifying a trend that exists (Bayazit and Önöz, 2007). The zyp R package (Bronaugh and Werner, 2013) contains an efficient implementation of Yue-Pilon's (Yue et al., 2002) prewhitening approach to determining trends in data.

In the Trend Free Pre-Whitening (TFPW) method, the slope is estimated with the Theil-Sen approach (TSA). If the slope is almost equal to zero, then it is not necessary to conduct the trend analysis. If it differs from zero, then it is assumed to be linear and the data is detrended by the slope and the auto-regressive AR(1) is computed for the detrended series (Yue et al., 2002). The residuals should be an independent series. The trend and residuals are then blended together through the sum of both terms. Finally, the Mann-Kendall test is applied to the blended series to assess the significance of the trend. Among a number of output fields, the Kendall's tau statistic and the Kendall's p-value computed for the final detrended time-series are reported after each Yue-Pilon analysis. The significant trends are thus identified and their increasing or decreasing behavior is set as a characteristic of individual patterns.

2.6. Distribution of DFA Factor Loadings across Monitoring Points

2.6.1. Drivers of the Spatiotemporal Variability of Water Quality

Factor loadings indicate the importance of a given pattern at a given river monitoring point (i.e. time-series). For each variable, factor loading values of individual extracted patterns can be plotted spatially, based on geographical coordinates, in order to visually identify clusters of patterns in space. Identifying potential causes of the distribution of patterns across a basin; however, requires a more rigorous approach than a merely visual examination of maps.

In our methodological framework, potential explanatory variables (e.g., characterizing climate or land-use) are related

to factor loadings by the generalized least squares (*gls*) regression model. The use of this type of model is advisable where errors are expected to be spatially correlated, which is usually the case in geographically-linked datasets (Pinheiro and Bates, 2000). The identification of an optimal set of explanatory variables for a regression model is however challenging and further complicated by the presence of collinearity in the explanatory variables.

To overcome these difficulties, we apply a combination of forward and backward selection in a stepwise multiple regression procedure. For each extracted pattern, we start by choosing the variables that gave the best predictions of the observed factor loadings at each monitoring point using gls models within the nlme R-Package (Pinheiro et al., 2012). Initially, the variable that resulted in the lowest significant p-value was chosen and kept in the regression; subsequent variables were selected based on the same criterion. If the inclusion of an additional variable rendered the initial variable non-significant, the latter was discarded and forward selection of the next significant variable took place until the predictions no longer improved with the addition of new variables.

The final gls models and their goodness-of-fit were assessed based on the generalized R^2 (Cox and Snell, 1989) by means of the r.squaredLR function included in the MuMIn R-Package (Barton, 2014). As a final product, we obtain the significant variables that act as potential drivers of the spatiotemporal variability of the DFA water-quality patterns extracted from our time-series.

2.6.2. Grouping Analysis of Predominant Patterns

In order to further assess any grouping or clustering of the extracted patterns in specific groups of monitoring points, we use the Grouping Analysis tool accessible in ArcGIS 10.1 (ESRI Inc.). The tool looks for a solution by which all the features (in our case, the factor loadings associated with the extracted patterns) within each group are as similar as possible, and all the groups themselves are as different as possible. Grouping analysis without any spatial constraints is performed by the K means algorithm and the decision of the most optimal number of clusters is based on the computed Calinski-Harabasz pseudo *F* statistic. The largest *F* statistic values indicate solutions that perform best at maximizing both within-group similarities and between-group differences.

3. Results

3.1. Case Studies

Data from river monitoring programs in three watersheds in Mediterranean Spain were used to illustrate our methodological framework in characterizing the spatiotemporal variability of water quality. These three basins were selected based on previous analysis carried out within the SCARCE Project Consortium (Navarro-Ortega et al., 2012): 1) The Llobregat River Basin (NE Spain) covers a drainage area of 4,948 km²

and hosts a population of ca. 3 million inhabitants (Marcé et al., 2012). The Llobregat River is a typical Mediterranean watercourse heavily impaired by dams, mines and pollution from diffuse and point sources, particularly in the lower part of its basin. 2) The Ebro River Basin covers a highly heterogeneous area of ca. 85,500 km². The hydrological regime of the Ebro River is dictated in part by its contrasting tributaries, from snowfed Pyrenean rivers to more typical Mediterranean tributaries in the southern part of the basin (Romaní et al., 2011). Large reservoirs and agricultural pollution are the two main pressures that influence biogeochemical characteristics of the river water in the basin. 3) Finally, the Júcar River Basin District (RBD) (43,000 km²) is located in the eastern part of the Iberian Peninsula and is formed by the aggregation of river basins that flow into the Mediterranean Sea. The hydrology of the Júcar RBD is typically Mediterranean, with rapid alternation between droughts and floods (Ferrer et al., 2012). In the middle section of the basin, agriculture leads to high nitrate concentrations in groundwater and surface water, whereas in the lower part there is a combination of agriculture, urban, and industrial pollution.

Data were collected from databases of public water agencies such as Confederación Hidrográfica del Júcar (CHJ), Agència Catalana de l'Aigua (ACA), and Confederación Hidrográfica del Ebro (CHE). We collected nitrate concentration data for the three basins, and for the Ebro River we additionally assembled time-series for oxygen concentration, pH, ammonium concentration, and biochemical oxygen demand. The time period was selected based on the longest possible availability of time series at as many monitoring stations as possible within each river basin. The frequency of sampling was monthly and the series length varied among basins: 17 years for the Júcar (starting in 1994), 11 (starting in 1995) for the Llobregat, and 31 for the Ebro (starting in 1980). Missing values were common in all three datasets.

The proposed methodology can be tailored to specific research objectives and characteristics of any given river basin, where the overall aim is to detect and characterize spatiotemporal patterns of water quality variability. Figure 3 introduces three applications of our methodological steps, based on specific objectives, and, to some extent, on the availability of water-quality data and the spatial density of monitoring locations. These case studies are presented in the following sections.

3.2. Examining Functional Relationships with MINE and DFA: Riverine Nitrate Concentration Patterns in the Júcar River Basin District

Water quality problems are common in the Júcar River Basin District (Júcar RBD). In the middle part, agriculture leads to high nitrate concentrations in ground and surface waters, whereas nutrient pollution in the lower part is a combination of agriculture, urban, and industrial sources (Ferrer et al., 2012). An initial exploratory analysis of nitrate time-series was carried out based on the Maximal Information Coefficient (MIC) to identify functional relationships between nitrate con-
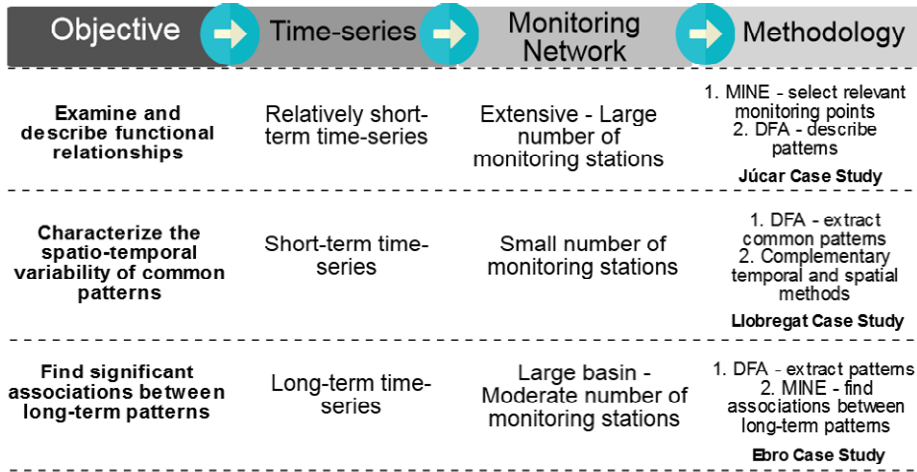
Figure 3. Methods tailored to specific objectives and basin characteristics of our case-study rivers.

centration and time in 90 monitoring points in the Júcar RBD. The analysis revealed a combination of agriculture, urban, and industrial sources (Ferrer et al., 2012). An initial exploratory analysis of nitrate time-series was carried out based on the Maximal Information Coefficient (MIC) to identify functional relationships between nitrate concentration and time in 90 mnitoring points available in the Júcar RBD. The analysis revealed that a strong relationship (according to MIC scores) of nitrate concentration with time was significant in 50 monitoring points (Figure 4). Consequently, we aimed to detect common nitrate patterns in the set of 50 time-series using dynamic factor analysis to further describe their shape and variability across time and space.

Among the four common extracted patterns by DFA, Pattern 1 had the largest importance in most monitoring points across the Júcar RBD (Figure 5), and it was mainly described by frequency analysis as a cycle with significant annual periodicity. MINE analysis and resulting statistics had initially identified a predominantly nonlinear, non-monotonic association between the nitrate and time variable-pair, which could be related to DFA Pattern 1. Conversely, the more linear and monotonic-like relationships previously found in the MINE analysis could be linked to the significant trend (Kendall's tau of -0.22, p < 0.001) found in DFA Pattern 4.

The geographical representation of the Maximum Asymmetry Scores (MAS; highest values indicate a clear departure from monotonicity) in Figure 4 shows that non-monotonic functions best described the time-series found along the Júcar River and in the upper Júcar RBD, both being mainly agricultural areas. The displayed non-monotonicity coincided, for the most part, with stations where cyclic patterns (e.g. DFA Pattern 1) predominated. These cycles are associated to the hydrological variability in the basin, which has been observed to modulate nitrate concentrations in rivers and streams. In fact, nitrate pollution related to fertilizer application is one of the major problems affecting water quality in the region (Ferrer et al., 2012), and is accentuated by the seasonal cycles of

hydrological variability (i.e., during dry periods, nitrate levels tend to increase, and in wet periods, nitrate levels tend to decrease).
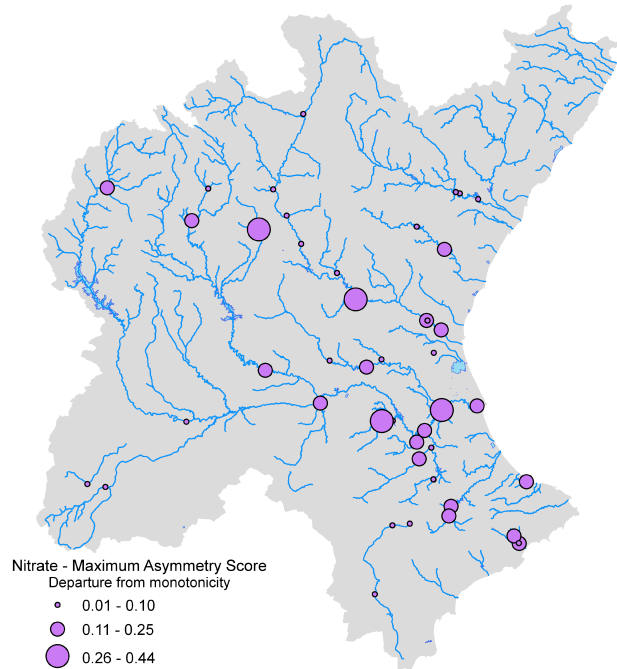

Figure 4. Maximum Asymmetry Scores (MAS) for the 50 nitrate time-series in the Júcar basin. The highest scores (larger circles) display the departure of the functional associations from monotonicity.

DFA results further explained the preliminary exploratory MINE results about the main characteristics of nitrate spatio-temporal variability (trend and cyclic behaviors and their geo-spatial importance). The initial MINE exploratory analysis was necessary to filter out those monitoring points with a sig-

nificant relationship between the nitrate and time variable-pair and thus avoid the cumbersome implementation of DFA involving a set of 90 time-series.

### 3.3. Characterizing the Spatiotemporal Variability of Nitrate Patterns: The Llobregat River Basin

The Llobregat River Basin (NE Spain) is a well-studied basin with a relatively small number of monitoring points of short-term data for water quantity and quality. During the last century, the Llobregat basin has suffered of progressive impairment by industrial and urban sewage as well as by runoff from agricultural areas that cannot be diluted by its natural flow. A comprehensive detection and characterization of nitrate concentration was carried out following all the steps shown in Figure 2. Since our objective was to specifically analyze patterns of nitrate dynamics and their drivers across the entire monitoring network, we skipped the MINE exploratory analysis step and worked with nitrate time-series from all available monitoring points.
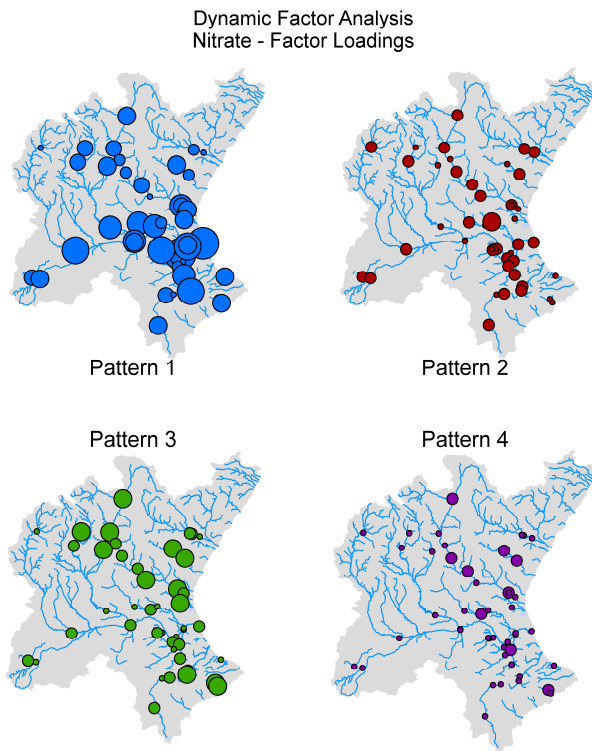
Figure 5. Factor loadings (absolute values; size of circles indicates the magnitude of relevance) associated with nitrate extracted patterns (1-4) in the Júcar basin.

The DFA model for nitrate time-series at 20 monitoring locations extracted four common patterns, quite dissimilar among one another (Figure 6). Frequency (multitaper) and trend (Yue-Pilon) analyses of these patterns characterized their temporal variability (Table 1). Detecting significant fre-

quencies in the extracted patterns facilitates the identification of seasonal cycles and oscillations potentially related to climate. In fact, the associations of El Niño Southern Oscillation (ENSO) with streamflow modifications (Marcé et al., 2010) and nitrate concentration dynamics (Vegas-Vilarrúbia et al., 2012) in the Iberian Peninsula have been clearly identified. Furthermore, detecting long-term significant trends can assist in identifying any phenomenon influencing nitrate concentration. Pattern 1, for instance, was described by a significant decreasing trend and significant periods of 2.1 and 1.7 yr. Such decreasing trend in nitrate concentration could be linked to improved agricultural practices in the region over time. No significant trend was observed for the remaining three patterns but Patterns 1 and 2 had significant cycles of 1 yr and of 3.4 yr, as well as Pattern 3. Pattern 4 presented a significant cycle of 10 months.
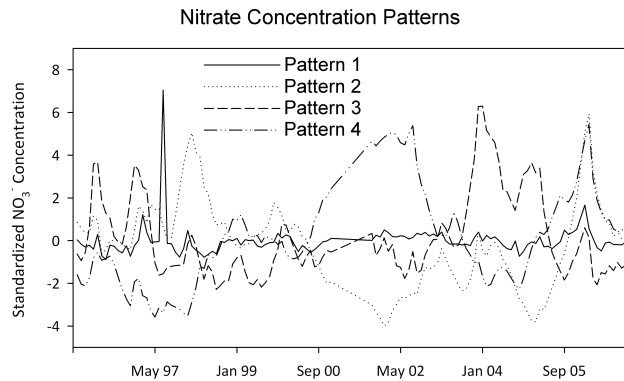
Figure 6. Common patterns extracted from 20 nitrate-concentration time series in the Llobregat basin (1995-2006).

The magnitude and sign of factor loadings, along with their spatial distribution, assist in the identification of predominant patterns in specific areas of the study basin. The factor loadings associated with Pattern 1 in the nitrate model had the largest values and were specifically located on the lower section of the Llobregat River (Figure 7). According to the gls spatial regression results, this pattern had positive significant relationships with total discharge upstream, as well as with local urban areas present in the basin (Table 1), among the different environmental variables included in the gls models. Here, a significant association with Pattern 1 also confirmed that it followed discharge variability, mainly in the lower section of the Llobregat River (highest factor loadings associated with Pattern 1). Since discharge relies on both precipitation and evapotranspiration, extreme events such as droughts and heat waves promoted by global atmospheric teleconnections can have significant effects on river water quality in the basin. These events and their link to climatic modes should be further investigated, especially in the Mediterranean region, where climate extreme events are predicted to increase (García-Ruiz et al., 2011).

The remaining patterns were more or less equally sca-

Table 1. Main Characteristics Found in the Resulting Nitrate Patterns and Associated Factor Loadings that Help in Describing the Spatiotemporal Variability in the Basin

| DFA MODEL	COMMON PATTERN			FACTOR LOADING	
Nitrate Model	Significant Cycle	Significant Trend	Hydrology-driven	Generalized least squares (gls) explanatory variable(s)	gls (p-value)
Pattern 1	2.1, 1.7 yr	Yes	Yes	Local Urban Area	0.0185
				Discharge	0.0013
Pattern 2	1 yr				
Pattern 3	3.9, 1 yr		Yes		
Pattern 4	10 mo			Discharge	0.0347

ttered across the entire basin, particularly Pattern 3, which presented high values for factor loadings in most monitoring points. Conversely, Pattern 4 predominated in the upper Anoia tributary (SW of the basin), a highly polluted and clearly differentiated area due to the presence of agricultural and industrial activities. The gls regression models for Patterns 2 and 3 did not identify any significant explanatory variables. Pattern 4, on the other hand, was negatively related to river discharge, suggesting higher nitrate concentration with lower values of discharge.

Grouping Analysis of the factor loadings associated with all four patterns extracted in the Llobregat basin, based on a distribution of two groups identified by the F-test, revealed no clear differentiation for the predominant Pattern 1 in specific monitoring points, but those points related to this cyclic and hydrology-driven pattern did remain part of the same group. Group 1 enclosed mainly those stations where Patterns 2 and 3 were predominant.

3.4. Exploring Associations between Long-Term Patterns of Water Quality: The Ebro River Basin

The biogeochemical characteristics of river water in the Ebro basin are highly influenced by anthropogenic activities. The main impacts are related to discharge regulation (i.e., the construction of the large reservoirs) and agricultural pollution (Romaní et al., 2011). Long-term patterns extracted from sets of water quality time-series provide information on significant trends in a river basin. The capabilities of the MINE algorithm were used in order to explore any significant functional relation-ship among patterns of different variables in the Ebro basin, where a sufficiently long dataset was available (31 years of monthly data). Such patterns were previously extracted in individual models (data not shown) for each variable: ammonium, nitrate, and dissolved oxygen concentrations, pH, and biochemical oxygen demand. The strongest of the significant associations between patterns, based on MIC scores, involved nitrate and dissolved oxygen in the Ebro basin (Table 2). Based on MINE statistics, the association was highly monotonic. Moreover, Pattern 3 of the DFA nitrate model was indeed a highly significant trend according to previous Yue-Pilon trend analysis (Kendall's tau of -0.53, $p < 0.001$). The trend component in Pattern 4 for dissolved oxygen was however only marginally significant (Kendall's tau of 0.08, $p < 0.05$). The de-

creasing or increasing nature of these long-term trends can, however, vary across space based on the sign of the corresponding factor loading, thus creating several possible relationships between the two variables (e.g., between nitrate and oxygen concentrations). Frequency analysis of both extracted common patterns showed significant 4-yr cycles.
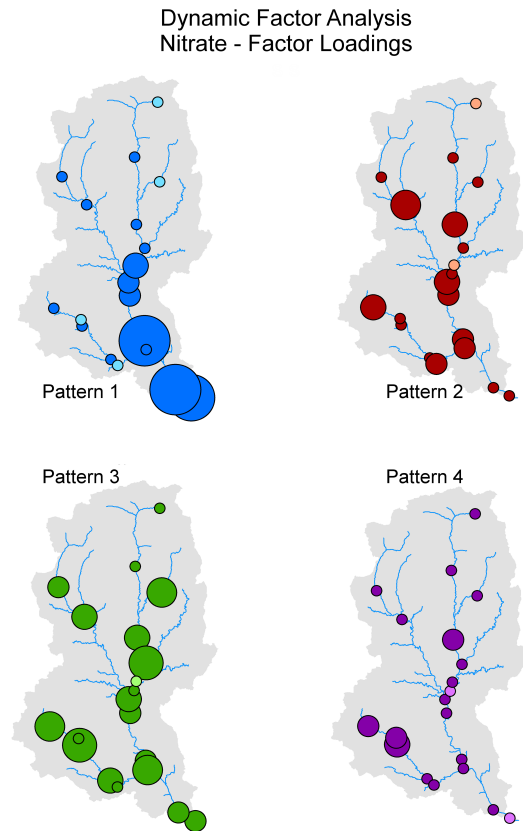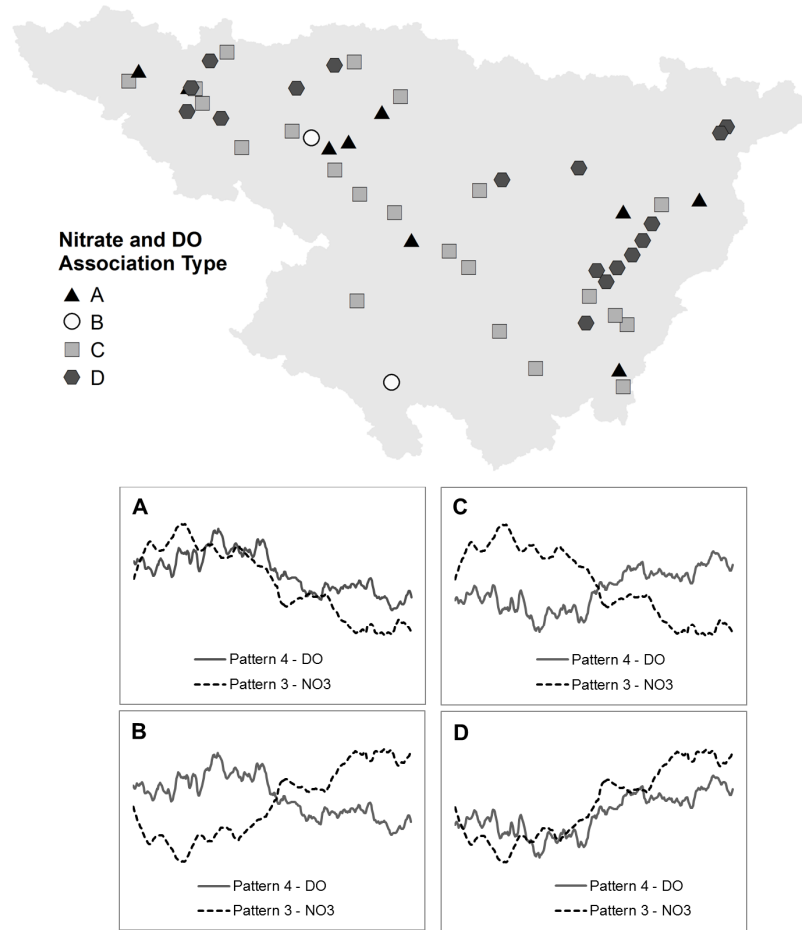

Figure 7. Factor loadings for nitrate patterns in the Llobregat basin (the size of circles indicates the magnitude; the light-colored circles represent negative factor loadings).

Figure 8 displays the different associations between the two patterns (Pattern 3 for nitrate and Pattern 4 for dissolved oxygen) and their geographical location in the basin. Asso-

Table 2. MIC Scores and Associated MINE Statistics for the Significant Relationships between Water Quality Patterns in the Ebro Basin

Variable 1	Variable 2	MIC (strength)	MIC-p^2 (nonlinearity)	MAS (non-monotonicity)	MEV (functionality)	MCN (complexity)
NO$_3$ Pattern 3	DO Pattern 4	0.99	0.33	0.03	0.99	5.08
NO$_3$ Pattern 3	PO$_4$ Pattern 3	0.90	0.40	0.11	0.90	5.08
NO$_3$ Pattern 3	NH$_4$ Pattern 1	0.83	0.45	0.05	0.82	5.08
NO$_3$ Pattern 3	BOD5 Pattern 3	0.75	0.25	0.09	0.75	4.81
NO$_3$ Pattern 3	pH Pattern 2	0.74	0.17	0.07	0.73	5.09

Figure 8. Association types (i.e., A, B, C, D) for NO3- Pattern-3 and DO Pattern-4 based on the sign of factor loadings and geographical location in the Ebro River Basin.

ciation type C, which shows the nitrate and oxygen patterns (standardized data) as they were extracted from the corresponding DFA models, is the predominant type of significant association (present in 22 out of 50 stations). Nitrate decreased over the 31 years included in the analysis, and the opposite happened with DO. Nitrate trends were possibly shaped by the application of agricultural practices that, in the last three decades, can be associated with a more rational fertilizer application (Lassaletta et al., 2012) in some areas of the basin. In addition, it is expected that nitrate and dissolved oxygen long-term trends behave in opposite directions at a particular point, since increasing nitrate concentration in rivers tends to promote eutrophication and the subsequent depletion of dissolved oxygen levels in freshwater systems. If nitrate concentration decreases across time, dissolved oxygen concentration levels should recover with time.

The most significant spatial agglomeration appeared to be the increasing nitrate and dissolved oxygen concentrations along the Segre tributary (a highly agriculturally active region located in the eastern part of the Ebro basin) within the 31-

year period included in the analysis. The different regional association types can be studied in detail in order to explain the varying relationships between nitrate and dissolved oxygen trends across the Ebro basin. Regression models involving environmental variables, for instance, could take part in identifying the main potential drivers of change.

4. Discussion

There is growing evidence that long-term climatic trends and changes in land cover have produced marked alterations in hydrological responses at the basin scale. These include the impact of climate change on rainfall and river flows, as well as the effect of changes in land use in recent decades (Milly et al., 2005; Gallart et al., 2011). Regardless of their size, river basins commonly encompass a variety of land use and climate characteristics. Even though anthropogenic activities in particular areas are usually determined by climatic and terrain conditions, the interaction of sources and drivers of change at different temporal and spatial scales challenges any analysis of variability and potential causality at the basin scale. The assessment of global change impacts on freshwater resources, which frequently relies on scantily available data, is further complicated by the complex interaction of processes within a system at multiple temporal and spatial scales (Qian et al., 2010). In order to address these issues, the proposed methodology tackles the different scales in time and space at which changes occur, making it possible to delineate areas where certain patterns predominate, or instances in time where specific trends were more relevant, which can all in turn be compared and contrasted with past and current global change phenomena in the study area. The case studies above illustrate the capabilities and the flexibility of our proposed methodological framework, which detects and characterizes river water-quality patterns at basin and regional scales. Even in absence of a clear a priori hypothesis, the MIC score is a very convenient tool to identify variables of interest and relevant monitoring points that could be subsequently studied in more detail.

One of the main limitations in the detection and assessment of changes in environmental time-series is the presence of observation gaps and the absence of sufficiently long records (Kundzewicz and Robson, 2004; Hegerl et al., 2010). Few available time-series techniques are able to readily overcome those two problems (Zuur et al., 2003b). In environmental datasets, shortcomings of these techniques include the irregular or uneven sampling frequency and the lack of a multivariate approach (Cazelles et al., 2008). Dynamic factor analysis can easily deal with short-term time-series that contain missing observations while providing a framework for detecting common patterns simultaneously in a set of multiple time-series within a basin or a region. Moreover, this method can incorporate time-series with differing sampling resolutions into the analysis. The DFA algorithm applied in this work is however currently computationally intensive and, subsequently, time-expensive. Working with the R-matrix structure of equal variance and covariance improved the speed of the algorithm. Reported inefficiencies of the MARSS algorithm are

being addressed by the R-package developers and some improvements have already been implemented in newer versions of the R-package (version 3.9 used in this study; Holmes et al., 2014).

The patterns detected by DFA describe mainly the temporal variability of the time-series and the magnitude and sign of factor loadings determine how the common trends are related to the original time series (Zuur et al., 2003b). By combining these two DFA outputs, we obtain a full description of the spatiotemporal variability of basin-wide water-quality patterns. Another advantage of the extracted patterns is that they are expressed as continuous time-series that can be easily analyzed with techniques that require complete observational records (e.g., spectral analysis). In this sense, dynamic factor analysis provides data to be further tested and studied in order to attribute changes to drivers acting at local and regional scales. On the other hand, the overall methodology is ultimately unable to give information about the physical foundation of the detected patterns of change.

Dynamic factor analysis allows the inclusion of covariates (i.e., potential explanatory variables) in the model. Previous applications of this particularity have been mainly focused on fisheries data in European seawaters and the influence of climate variables such as the North Atlantic Oscillation and Sea Surface Temperature (Zuur et al., 2003a; Zuur and Pierce, 2004), as well as groundwater nutrient concentration in south Florida (USA) affected by explanatory variables such as water table depth, enriched topsoil, and occurrence of a leaching rainfall event (Muñoz-Carpena et al., 2005). The covariate data are assumed to have no error and can not contain missing values, though there are ways to circumvent these problems (Holmes et al., 2013). Furthermore, other alternatives could be used instead of the specific tools and complementary methods for temporal and spatial analyses proposed in this paper. For instance, spectral and trend analyses include a myriad of possibilities (other than the multitaper and Yue-Pilon methods), and the user might choose a different method according to specific case requirements or preferences. This also applies to spatial clustering and regression methods. For our purposes, the K-means cluster seemed to suffice in order to pinpoint the main (few) groups of dominant patterns in our study basins, whereas the generalized least square approach for our regression models were adequate for our spatially correlated dataset. In terms of displaying the relevance and distribution of common patterns in the study area, we mapped the magnitude of the resulting factor loadings in the three basins (Figures 5, 7 and 8). Additionally, and depending on the area of study, isohyets can be included to indicate common areas where certain patterns are most relevant, such as in the analysis of lake water-quality by Magyar et al. (2013).

In general, the impact of global change on the dynamics of nutrient concentration in river basins usually combines regional and global factors, such as climatic events and agricultural practices, as well as local impacts, such as urban and industrial activities. The set of methods presented above deals mainly with the detection of changes in water quality and

the characterization of their variability across time and space within a region. In addition, a hint to subsequent steps of cause attribution has been provided in the comprehensive Llobregat River Basin case study.

Overall, the implemented methodology allows the identification and description of changes in water quality using datasets that would have been otherwise disregarded if conventional time-series techniques were to be used. We have illustrated the capabilities of the individual and of the collection of methods presented earlier in three Mediterranean basins, though our approach is not limited to this region since data availability restrictions and environmental change detection challenges are found worldwide. The same methodological framework is also applicable in other limnological and environmental contexts, as it allows the study of patterns in sets of time-series within a region or ecosystem. Finally, the uncertainties regarding field data collection and chemical analyses of solutes, which were undertaken by public/governmental environmental agencies, have been discussed elsewhere (e.g., Ibáñez et al., 2008). The uncertainty in our data analyses was not formally addressed but the extracted patterns and potential drivers of water quality in the three case studies were consistent with previous studies of riverine nutrient behavior in the three Mediterranean basins. The importance of uncertainty analysis and assessment in this context however warrants attention in future studies.

5. Conclusions

Some of the driving forces for global change, such as greenhouse gas emissions and the reach of global financial systems, operate at a global scale. Conversely, it also seems clear that several of the individual phenomena underlying environmental change processes, such as economic activities, resource use, and population dynamics, arise at a local scale (Wilbanks and Kates, 1999). With the improvement of computational resources and the knowledge gained through research on global change and river basins in the last few decades, present and near-future efforts should be focused on optimizing the tools and formalizing the methodology for change and impact assessment, from local to regional and global scales, in order to produce reliable forecasts that can assist policy makers and managers to ultimately secure freshwater resources. The proposed methodological framework addresses the spatiotemporal complexity of water-quality patterns in river basins. All in all, these methods offer the possibility of simultaneously extracting common patterns from a set of noisy and gap-containing time-series, providing not only a temporal interpretation of variability but also allowing researchers to characterize the spatial variability across basin and regional scales. Our methodological framework thus serves as a hypothesis-generation tool for further analyses of drivers of change in limnological and environmental patterns.

References

Barton, K. (2014). MuMIn: Multi-model inference. R Package version 1.10.0.

Bayazit, M., and Önöz, B. (2007). To prewhiten or not to prewhiten in trend analysis? *Hydrol. Sci. J.,* 52(4), 611-624. http://dx.doi.org/10.1623/hysj.52.4.611

Benítez-Gilabert, M., Alvarez-Cobelas, M., and Angeler, D.G. (2010). Effects of climatic change on stream water quality in Spain. *Clim. Change,* 103(3), 339-352. http://dx.doi.org/10.1007/s10584-009-9778-9

Bers, A.V., Momo, F., Scholss, I.R., and Abele, D. (2013). Analysis of trends and sudden changes in long-term environmental data from King George Island (Antartica): Relationships between global climatic oscillations and local system response. *Clim. Change,* 116(3), 789-803. http://dx.doi.org/10.1007/s10584-012-0523-4

Bronaugh, D., and Werner, A. (2013). zyp: Zhang + Yue-Pilon trends package. R package version 0.10-1 R package version 0.10-1, Pacific Climate Impacts Consortium.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S., and Stenseth, N.C. (2008). Wavelet analysis of ecological time series. *Oecologia,* 156(2), 287-304. http://dx.doi.org/10.1007/s00442-008-0993-2

Cox, D.R., and Snell, E.J. (1989). *The Analysis of Binary Data,* (2th Ed.), Chapman and Hall.

Ferrer, J., Pérez-Martín, M.A., Jiménez, S., Estrela, T., and Andreu, J. (2012). GIS-based models for water quantity and quality assessment in the Júcar River Basin, Spain, including climate change effects. *Sci. Total Environ.,* 440, 42-59. http://dx.doi.org/10.1016/j.scitotenv.2012.08.032

Gallart, F., Delgado, J., Beatson, S.J.V., Posner, H., Llorens, P., and Marcé, R. (2011). Analysing the effect of global change on the historical trends of water resources in the headwaters of the Llobregat and Ter river basins (Catalonia, Spain). *Phys. Chem. Earth,* 36(13), 655-661. http://dx.doi.org/10.1016/j.pce.2011.04.009

García-Ruiz, J.M., López-Moreno, J.I., Vicente-Serrano, S.M., Lasanta-Martínez, T., and Beguería, S. (2011). Mediterranean water resources in a global change scenario. *Earth-Sci. Rev.,* 105(3-4), 121-139. http://dx.doi.org/10.1016/j.earscirev.2011.01.006

Geweke, J.F. (1977). The dynamic factor analysis of economic time series, In D. Aigner and A. Goldberger (Eds.), *Latent Variables in Socio-economic North-Holland*, Amsterdam, pp. 365-383.

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A.W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P. (2002). Advanced Spectral Methods for Climatic Time Series. *Rev. Geophys.,* 40(1), 3-1-3-41. http://dx.doi.org/10.1029/2000rg000092

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Hegerl, G.C., Hoegh-Guldberg, O., Casassa, G., Hoerling, M., Kovats, S., Parmesan, C., Pierce, D., and Stott, P. (2010). Good practice guidance paper on detection and attribution related to anthropogenic climate change, in T. F. Stocker, C. Field, Q. Dahe, V. Barros, G. K. Plattner, M. Tignor, P. Midgley, and K. Ebi (Eds.), *IPCC Expert Meeting on Detection and Attribution Related to Anthropogenic Climate Change 2010*, Geneva, Switzerland, (1-8). 14-16 September, 2009.

Holmes, E.E. (2013). *Derivation of the EM Algorithm for Constrained and Unconstrained Multivariate Autoregressive State-space (MARSS) Models*, Technical Report.

Holmes, E.E., Ward, E.J. and Wills, K. (2012). MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *R Journal,* 4(1), 11-19.

Holmes, E.E., Ward, E.J. and Wills, K. (2013). MARSS: Multivariate Autoregressive State-Space Modeling, R package version 3.4.

Holmes, E.E., Ward, E.J. and Wills, K. (2014). MARSS: Multivariate

Autoregressive State-Space Modeling. R package version 3.9.

Huang, G.H., and Chang, N.B. (2003). The perspectives of environmental informatics and systems analysis. *J. Environ. Inf.,* 1(1), 1-7. http://dx.doi.org/10.3808/jei.200300001

Ibáñez, C., Fornell, N.P., Duran, C., Pardos, M., Torras, A.M., Andreu, R., Caiola, N., Cid, N., Hampel, H., Sáfael, R., and Pujadas, R.T. (2008). Changes in dissolved nutrients in the lower Ebro river: Causes and consequences. *Limnetica,* 27(1), 131-142.

Ito, A. (2012). Detection and attribution of global change and disturbance impacts on a tower-observed ecosystem carbon budget: a critical appraisal. *Environ. Res. Lett.,* 7(1), 1-6. http://dx.doi.org/10.1088/1748-9326/7/1/014013

Kandasamy, S., Baret, F., Verger, A., Neveux, P., and Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations-application to MODIS LAI products. *Biogeosciences,* 10(6), 4055-4071. http://dx.doi.org/10.5194/bg-10-4055-2013

Kundzewicz, Z.W., and Krysanova, V. (2010). Climate change and stream water quality in the multi-factor context: An editorial comment. *Clim. Change,* 103(3), 353-362. http://dx.doi.org/10.1007/s10584-010-9822-9

Kundzewicz, Z.W., and Robson, A.J. (2004). Change detection in hydrological records -- a review of the methodology / Revue méthodologique de la détection de changements dans les chroniques hydrologiques. *Hydrol. Sci. J.,* 49(1), 7-19. http://dx.doi.org/10.1623/hysj.49.1.7.53993

Kuo, Y.M., Jang, C.S., Yu, H.L., Chen, S.C., and Chu, H.J. (2013). Identifying nearshore groundwater and river hydrochemical variables influencing water quality of Kaoping River Estuary using dynamic factor analysis. *J. Hydrol.,* 486, 39-47. http://dx.doi.org/10.1016/j.jhydrol.2013.01.027

Lassaletta, L., Romero, E., Billen, G., Garnier, J., García-Gómez, H., and Rovira, J.V. (2012). Spatialized N budgets in a large agricultural Mediterranean watershed: High loading and low transfer, *Biogeosciences,* 9, 57-70. http://dx.doi.org/10.5194/bg-9-57-2012

Lyubchich, V., Gel, Y.R., and El-Shaarawi, A. (2013). On detecting non-monotonic trends in environmental time series: a fusion of local regression and bootstrap. *Environmetrics,* 24(4), 209-226. http://dx.doi.org/10.1002/env.2212

Magyar, N., Hatvani, I.G., Székely, I.K., Herzig, A., Dinka, M., and Kovács, J. (2013). Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See. *Ecol. Eng.,* 55, 82-92. http://dx.doi.org/10.1016/j.ecoleng.2013.02.005

Marcé, R., Rodríguez-Arias, M.A., García, J.C., and Armengol, J. (2010). El Niño Southern Oscillation and climate trends impact reservoir water quality. *Global Change Biol.,* 16(10), 2857-2865. http://dx.doi.org/10.1111/j.1365-2486.2010.02163.x

Marcé, R., Honey-Rosés, J., Manzano, A., Moragas, L., Catllar, B., and Sabater, S. (2012). The Llobregat River Basin: A paradigm of impaired rivers under climate change threats, in S. Sabater, A. Ginebreda and D. Barceló (Eds.), The Llobregat River: the story of a polluted river, vol. 21, *The Handbook of Environmental Chemistry*, Springer Verlag, Berlin Heidelberg, pp. 1-26. http://dx.doi.org/10.1007/698_2012_152

Márkus, L., Berke, O., Kovács, J., and Urfer, W. (1998). *Analysis of Spatial Structure of Latent Effects Governing Hydrogeological Phenomena (No. 1998, 37)*, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

Meybeck, M. (2003). Global analysis of river systems: From Earth system controls to Anthropocene syndromes. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.,* 358(1440), 1935-1955. http://dx.doi.org/10.1098/rstb.2003.1379

Milly, P.C.D., Dunne, K.A., and Vecchia, A.V. (2005). Global pattern

of trends in streamflow and water availability in a changing climate. *Nature,* 438, 347-350. http://dx.doi.org/10.1038/nature04312

Muñoz-Carpena, R., Ritter, A., and Li, Y.C. (2005). Dynamic factor analysis of groundwater quality trends in an agricultural area adjacent to Everglades National Park. *J. Contam. Hydrol.,* 80(1-2), 49-70. http://dx.doi.org/10.1016/j.jconhyd.2005.07.003

Navarro-Ortega, A., Acuña, V., Batalla, R.J., Blasco, J., Conde, C., Elorza, F.J., Elosegi, A., Francés, F., La-Roca, F., Muñoz, I., Petrovic, M., Picó, Y., Sabater, S., Sanchez-Vila, X., Schuhmacher, M., and Barceló, D. (2012). Assessing and forecasting the impacts of global change on Mediterranean rivers. The SCARCE Consolider project on Iberian basins. *Environ. Sci. Pollut. Res.,* 19(4), 918-933. http://dx.doi.org/10.1007/s11356-011-0566-5

Pinheiro, J.C., and Bates, D.M. (2000). *Mixed Effects Models in S and S-PLUS*, Springer, New York.

Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., and R Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models, R package version 3.1-105.

Qian, S.S., Cuffney, T.F., Alameddine, I., Mcmahon, G., and Reckhow, K.H. (2010). On the application of multilevel modeling in environmental and ecological studies. *Ecology,* 91, 355-361. http://dx.doi.org/10.1890/09-1043.1

Rahim, K., and Burr, W. (2013). Multitaper: Multitaper spectral analysis tools, R package version 1.0-8.

R Core Team (2012). R: A language and environment for statistical computing, R Foundation for Statistical Computing.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzernmacher, M., and Sabeti, P.C. (2011). Detecting novel associations in large datasets. *Science,* 334(6062), 1518-1524. http://dx.doi.org/10.1126/science.1205438

Romaní, A.M., Sabater, S., and Mu-oz, I. (2011). The physical framework and historic human influences in the Ebro River, in D. Barceló and M. Petrovic (Eds.), *The Ebro River Basin*, Springer-Verlag.

Seitzinger, S.P., Mayorga, E., Bouwman, A.F., Kroeze, C., Beusen, A. H.W., Billen, G., Drecht, G.V., Dumont, E., Fekete, B.M., Garnier, J., and Harrison, J.A. (2010). Global river nutrient export: A scenario analysis of past and future trends. *Global Biogeochem. Cycles,* 24(4). http://dx.doi.org/10.1029/2009GB003587

Shumway, R.H., and Stoffer, D.S. (2010). *Time Series Analysis and Its Applications: with R Examples*, Springer.

Thomson, D.J. (1982). Spectrum estimation and harmonic analysis. *Proc. of the IEEE,* 70(9), 1055-1096. http://dx.doi.org/10.1109/PROC.1982.12433

Vegas-Vilarrúbia, T., Sigró, J., and Giralt, S. (2012). Connection between El Niño-Southern Oscillation events and river nitrate concentrations in a Mediterranean river. *Sci. Total Environ.,* 426, 446-453. http://dx.doi.org/10.1016/j.scitotenv.2012.03.079

Vilizzi, L. (2012). Abundance trends in floodplain fish larvae: The role of annual flow characteristics in the absence of overbank flooding. *Fundam. Appl. Limnol./Arch. Hydrobiol.,* 181(3), 215-227. http://dx.doi.org/10.1127/1863-9135/2012/0394

Vörösmarty, C.J. (2002). Global water assessment and potential contributions from Earth Systems Science. *Aquat. Sci.,* 64, 328-351.

Vörösmarty, C.J., Green, P., Salisbury, J., and Lammers, R.B. (2000). Global water resources: Vulnerability from climate change and population growth. *Science,* 289(5477), 284-288. http://dx.doi.org/10.1126/science.289.5477.284

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., and Davies, P.M. (2010). Global threats to human water security and river biodiversity. *Nature,* 467, 555-561. http://dx.doi.org/10.1038/nature09440

Wilbanks, T.J., and Kates, R.W. (1999). Global change in local places:

How scale matters. *Clim. Change,* 43(3), 601-628. http://dx.doi. org/10.1023/A:1005418924748

Yue, S., Pilon, P., Phinney, B., and Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.,* 16(9), 1807-1829. http://dx.doi.org/10.10 02/hyp.1095

Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R., and Beukema, J.J. (2003a). Estimating common trends in multivariate time series us-

ing dynamic factor analysis. *Environmetrics,* 14(7), 665-685. http: //dx.doi.org/10.1002/env.611

Zuur, A.F., Tuck, I.D., and Bailey, N. (2003b). Dynamic factor analy-sis to estimate common trends in fisheries time series. *Can. J. Fish. Aquat. Sci.,* 60(5), 542-552. http://dx.doi.org/10.1139/f03-030

Zuur, A.F., and Pierce, G.J. (2004). Common trends in northeast Atlantic squid time series. *J. Sea Res.,* 52(1), 57-72. http://dx.doi. org/10.1016/j.seares.2003.08.008