Universitat de Girona

EFFECTS OF LAND USES AND CLIMATE VARIABILITY ON THE WATER QUALITY OF MEDITERRANEAN RIVERS: TOWARDS A REGIONAL VISION OF GLOBAL CHANGE

Rosana Aguilera Becker

Dipòsit legal: Gi. 522-2015
http://hdl.handle.net/10803/286633

Doctoral Thesis

Effects of land use and climate variability on the water quality of Mediterranean rivers: Towards a regional vision of global change

Rosana Aguilera Becker

2015

Doctoral Program in Water Science and Technology

Supervised by

Rafael Marcé Romero
Institut Català de Recerca de l'Aigua

Sergi Sabater Cortés
Institut d'Ecologia Aquàtica, UdG

Thesis submitted in fulfillment of the requirements for the Degree of Doctor (PhD) at the Universitat de Girona.

El Dr. Rafael Marcé Romero, investigador de l'Institut Català de Recerca de l'Aigua, i el Dr. Sergi Sabater Cortés, catedràtic d'Ecologia del Departament de Ciències Ambientals de la Universitat de Girona.

CERTIFIQUEN:

Que aquest treball titulat "Effects of land-use and climate variability on the water quality of Mediterranean rivers: towards a regional vision of global change", que presenta Rosana Aguilera Becker per a l'obtenció del títol de Doctor, ha estat realitzat sota la seva direcció.

Dr. Rafael Marcé Romero

Dr. Sergi Sabater Cortés

Rosana Aguilera Becker
Doctorand

Girona, Gener 2015

A mis padres, por darme el mejor ejemplo a seguir;
A mi hermana, por alentarme a creer en mí misma;
I a la meva parella, pel seu suport incondicional.

Acknowledgements

I would like to express my sincere gratitude to Sergi Sabater and Rafael Marcé for the opportunity of joining their research team, for enriching the work presented in this thesis with their ideas, knowledge, and experience, and for their sympathetic understanding in the most unforeseen situations.

I would also like to thank the Institut Català de Recerca de l'Aigua (ICRA) for the financial and institutional support provided throughout my experience as a predoctoral researcher.

This thesis was further supported by the Spanish Ministry of Economy and Competitiveness through the project SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and by a Doctoral Grant (FI-DGR 2012) awarded by the Catalan Government.

I would like to especially thank my colleagues at ICRA for offering their constructive criticism and for expressing their encouragement and friendship in numerous ways; they have helped me to endure and improve the many aspects of life as a PhD student, a journey that has proven to be challenging but by and large rewarding.

I thank the COST Action NETLAKE (Networking Lake Observatories in Europe), particularly to Rita Adrian and David M. Livingstone, for allowing me to have the experience every PhD student should have: working in a different environment and interacting with students and established scientists of other institutions.

Thanks to Santi Villalba for the wonderful cover design.

I am greatly indebted to my loved ones, Marta, Miguel, Melisa, and Xavier; I wouldn't have gotten to this point if it wasn't for them. *Rohayhu!*

"What you get by achieving your goals
is not as important as what
you become by achieving your goals."

- Henry David Thoreau

Scientific publications derived from this thesis

Aguilera, R., Sabater, S., Marcé R. A primer on the use of Dynamic Factor Analysis to study the spatiotemporal variability of river water quality series. *Submitted*.

Aguilera, R., Marcé, R., Sabater, S. Detection and attribution of global change effects on river nutrient dynamics in a large Mediterranean basin. *Submitted*.

Aguilera, R., Marcé, R., Sabater, S. (2013). Modeling nutrient retention at the watershed scale: does small stream research apply to the whole river network? *Journal of Geophysical Research-Biogeosciences*, 118 (2); 728-740

Aguilera, R., Marcé, R., Sabater, S. (2012). Linking in-stream nutrient flux to land use and inter-annual hydrological variability at the watershed scale. *Science of the Total Environment*, 440; 72-81.

Aguilera, R., Sabater, S., Marcé, R. (2012). In-Stream Nutrient Flux and Retention in Relation to Land Use in the Llobregat River Basin. In: The Llobregat: The Story of a Polluted Mediterranean River, S. Sabater et al. (eds), pp. 69-92. Springer Berlin Heidelberg.

LIST OF ACRONYMS

ACA	Catalan Water Agency
AIC	Akaike Information Criterion
ANCOVA	Analysis of covariance
ARIMA	Autoregressive integrated moving average model
BOD	Biochemical Oxygen Demand
CHE	Ebro Basin Authority
CHJ	Júcar Basin Authority
CI	Confidence Interval
DEM	Digital Elevation Model
DFA	Dynamic Factor Analysis
EL	Efficiency Loss concept
ENSO	El Niño Southern Oscillation
EM	Expectation-Maximization algorithm
GHG	Greenhouse gas(es)
GCM	Global Circulation Model
GIS	Geographical Information System
GLS	Generalized Least Squares
IPCC	Intergovernmental Panel on Climate Change
LOADEST	Load Estimator model
MARSS	Multivariate Autoregressive State-Space model
MIC	Maximal Information Coefficient
MIC-p^2	Natural measure of nonlinearity
MINE	Maximal Information-based Nonparametric Exploration
MAS	Maximum Asymmetry Score
MTM	Multitaper method
NAO	North Atlantic Oscillation
NWLS	Nonlinear Weighted Least Squares
RMSE	Root-Mean-Square Error
RBD	River Basin District
SRP	Soluble Reactive Phosphorus

SPARROW	Spatially-referenced regression on watershed attributes
SOI	Southern Oscillation Index
SAS	Statistical Analysis System software
TSA	Theil-Sen Approach
TFPW	Trend Free Pre-whitening procedure
WWZ	Weighted Wavelet Z-transform
WWTP	Wastewater Treatment Plant

CONTENTS

Summary 27

Resum 31

Resumen 35

General Introduction 41

General Objectives 53

Methods and Study Sites 57

Results

Chapter 1: *Modeling nutrient retention at the watershed scale: Does small stream research apply to the whole river network?* 77

Chapter 2: *Linking in-stream nutrient flux to land use and inter-annual hydrological variability at the watershed scale* 109

Chapter 3: *A primer on the use of Dynamic Factor Analysis to study the spatio-temporal variability of river water quality series* 135

Chapter 4: *Detection and attribution of global change effects on nutrient dynamics in a large Mediterranean basin* 167

General Discussion 193

General Conclusions 205

Appendices 209

References 245

LIST OF FIGURES

General Introduction

Figure 1: Main characteristics of Deterministic and Statistical Models. 48

Methods and Study Sites

Figure 1: The SPARROW model relates measured nutrient flux (i.e., monitoring points A and B) in rivers and streams to watershed characteristics associated to nutrient sources (point and diffuse) and to nutrient mobilization and decay processes (land-to-water delivery and in-stream decay). 59

Figure 2: The response of a particular stream subjected to periodic nutrient inputs may follow 1st order kinetics or Michaelis-Menten saturation model, but the response of streams receiving continuous loading could be best described by the Efficiency Loss model (O'Brien et al., 2007). 61

Figure 3: Dynamic Factor Analysis decomposes a set of time-series into a linear combination of common patterns plus associated noise terms. 64

Figure 4: Set of methods applied to the output of Dynamic Factor Analysis in order to attribute potential causes and drivers to the observed water-quality spatio-temporal variability. 65

Figure 5: Water quality data points in the Ebro (n=50), Júcar (n=90), and Llobregat (n=20) basins. Datasets and relevant geographical information were obtained from the corresponding public water agencies in charge of the management of these basins. 66

Figure 6: Main land-use types and wastewater treatment plants (WWTP) location in the Llobregat River Basin. 68

Figure 7: Main land-use types and WWTP location in the Ebro River Basin. 70

Figure 8: Main land cover and uses in the Júcar RBD. 71

Results

Chapter 1

Figure 1: Major land uses types found in the Llobregat River Basin within the delineated sub-watersheds and their corresponding reaches. Monitoring points and gauging stations are depicted for reference. 83

Figure 2: SPARROW-predicted relative to observed nutrient flux for 161 calibration measurements in 23 monitoring stations in the Llobregat River between 2000 and 2006 (Top, Nitrates; Bottom, Phosphates). 92

Figure 3: Relationship between uptake velocity (v_f) and nutrient concentration in literature studies (different symbols) and estimated in SPARROW for the Llobregat River Basin. (A, nitrates, B, phosphates). The correspondence between symbols for literature data and their reference can be found in the Appendix A. The power laws for literature data and the SPARROW results along with 95% confidence intervals are included. 93

Figure 4: Relationship between uptake rate (U) and nutrient concentration in literature studies (different symbols) and estimated in SPARROW for the Llobregat River Basin. (A, nitrates, B, phosphates). The correspondence between symbols for literature data and their reference can be found in the Appendix A. The power laws for literature data and the SPARROW results along with 95% confidence intervals are included. 98

Figure 5: Comparison of Functionally Equivalent and SPARROW Modeled v_f/HL ratio considering literature-based (0.5C-0.48) and model-driven (4.6C-1.2) power laws. 99

Figure 6: Resulting temporal-averaged uptake velocity (vf') from literature-based (low slope) and model-driven (high slope) power laws compared to their corresponding reference curves (bilogarithmic scale). 100

Figure 7: Nitrate concentration and streamflow data characteristics found in the literature compared to data used in the SPARROW models for the Llobregat River Basin. 102

Figure 8: Analysis of the effect of hydrology (represented by HL) and biological uptake variability (represented as different power relationships between v_f and C) on stream nutrient retention, calculated using Eq. 3. Three different relationships between v_f and C were tested: a constant v_f value (A), the power law fitted to bibliographical data gathered in this study (B), and the power law fitted by SPARROW for the Llobregat River Basin (C). 105

Chapter 2

Figure 1: Average Nitrate (a) and Phosphate (b) concentration in the selected monitoring sites (2000-2006). Dashed lines represent sub-watersheds division. Distribution of land use types (b) and Wastewater Treatment Plant (WWTP) location (a) are also shown. Urban land comprises 6% of the total watershed area, cultivated land, 24%, and natural land, the remaining 70%. 112

Figure 2: Mean in-stream observed annual loads (2000-2006) compared to model predicted loads at calibration sites (a: nitrate, b: phosphate). 119

Figure 3: Mean Observed Total Yield versus Predicted Total Yield in 23 monitoring points (2000-2006) for nitrate (a) and phosphate (b). Total Yield represents the total load divided by the total area of the watershed at a given monitoring station. 120

Figure 4: Share of incremental diffuse and point sources contribution to annual in-stream nitrate flux (a) and phosphate flux (b) in the Llobregat River Basin. 123

Figure 5: Mean Share Fraction for Diffuse and Point sources for nitrate (a) and phosphate (b) under wet and dry hydrological conditions in the Llobregat River Basin within the years 2000 – 2006. 128

Figure 6: Relationship between Mean Incremental Load for Nitrates and Phosphates and Mean Runoff generated in the watershed within 2000-2006. Data points represent the mean load generated in each of the seven years considered in this study for each nutrient and nutrient source, according to the associated mean runoff values. 128

Figure 7: Mean In-stream Nutrient Removed Fraction obtained from SPARROW reach decay specifications (a: nitrate model, b: phosphate model). Mean Total Predicted Load reaching the streams is obtained after land-to-water delivery and in-stream decay processes have taken place. The dashed line represents sub-watershed division. 130

Chapter 3

Figure 1: Methodological Flowchart. 139

Figure 2: Tailored methods to specific objectives and basin characteristics, and application to the three river basins used as study cases. 148

Figure 3: MAS scores for the 50 nitrate time-series. The highest scores display the departure of the functional associations from monotonicity. The predominant non-monotonic behavior is observed in the mid-western section of the basin. 150

Figure 4: Common patterns extracted from a set of 50 of 90 nitrate time-series where MINE analysis had indicated a significant functional association with time. 151

Figure 5: Factor loadings associated to the corresponding extracted patterns. Larger circles indicate larger weight or predominance of a specific pattern over the remaining three. Overall, Pattern 1 appeared to predominate across the Júcar RBD, specifically in the middle section along the Júcar River. 152

Figure 6: Nitrate common patterns extracted in the Llobregat basin (1995-2006). 155

Figure 7: The magnitude of the Factor Loadings is represented by the size of the circles for each monitoring point. Pattern 1 in the nitrate model was the most relevant in the basin, though the highest values concentrated along the mid and low Llobregat River sections. The light-colored circles represent negative Factor Loadings, which in turn describe the opposite behavior of the corresponding extracted pattern. 156

Figure 8: Association types for NO_3^- Pattern 3 and DO Pattern 4, both characterized mainly as trends, according to the sign of their corresponding factor loadings and their geographical location in the Ebro River basin. 159

Chapter 4

Figure 1: Top: DFA resulting patterns for nitrate (a) and phosphate (b) concentration. Center: Examples of observed time-series and fitted DFA models at two selected monitoring points for nitrate (c) and phosphate (d) concentration. The DFA models in panels (c) and (d) are the result of a linear combination of the patterns in panels (a) and (b), respectively. Bottom: Seasonal variation for nitrate Pattern 1 and streamflow (e), nitrate Pattern 2 and Temperature (f), and phosphate Pattern 1 and streamflow (g). Points depict monthly averages for the entire 31 year time-series. For temperature and streamflow, the average is for all time-series available. We only included standard deviations as error bars for the nutrient patterns to enhance readability. 177

Figure 2: Factor loadings associated to nitrate patterns (left column) and phosphate patterns (right column). Dark circles indicate positive factor loadings and light-colored circles represent negative factor loadings. The size of the circles represents the magnitude of the Factor Loading at each monitoring point. A map with the main land uses in the Ebro basin is enclosed in the lower left corner. 180

Figure 3: (a) Clustering analysis results for the spatial distribution of nitrate concentration patterns and associated explanatory variables. (b) Mean fraction of Factor Loadings (i.e., the overall weight of a specific pattern) found in each of the 4 clusters identified in the analysis. 182

Figure 4: (a) Clustering analysis results for the spatial distribution of phosphate concentration patterns and associated explanatory variables. (b) Mean fraction of Factor Loadings (i.e., the overall weight of a specific pattern) found in each of the 4 clusters identified in the analysis. 183

Figure 5: Global change factors, acting at different scales, that contribute to shape the spatio-temporal variability of nitrate and phosphate concentration in the Ebro basin. Numbered circles describe the relationship between nutrient concentration patterns and the identified factors and drivers of change. 189

General Discussion

Figure 1: Average nutrient source apportionment in the Llobregat basin within the 2000-2006 study period. 196

Figure 2: Summary of the main components identified in the extracted patterns of 50 time-series of nitrate and phosphate and their associated drivers. 198

LIST OF TABLES

General Introduction

Table 1: Types of global environmental change according to Turner et al. (1990). 41

Results

Chapter 1

Table 1: Nitrate Model Parameters for the Llobregat River Basin. Results for NLWS and bootstrap fitting are included. 94

Table 2: Phosphate Model Parameters for the Llobregat River Basin. Results for NLWS and bootstrap fitting are included. 95

Chapter 2

Table 1: Mean values for discharge and runoff found in the Llobregat River basin for each of the 7 calendar years included in this study. The highest values for these variables were observed in 2004; the lowest, in 2005. 117

Table 2: Nitrate Model Parameter Estimation (CIs correspond to Bootstrap Estimation). 121

Table 3: Phosphate Model Parameter Estimation (CIs correspond to Bootstrap Estimation). 121

Chapter 3

Table 1: Main characteristics found in the resulting nitrate patterns and associated factor loadings that help in describing the spatio-temporal variability in the basin. 155

Table 2: MIC scores and associated MINE statistics for the significant relationships between water quality patterns in the Ebro basin. 158

Chapter 4

Table 1: Characterization of the temporal variability and relationships with streamflow of nutrient patterns detected with DFA in the Ebro basin. 176

Table 2: GLS resulting potential drivers involved in the spatio-temporal variability of nitrate patterns in the Ebro basin. 179

Table 3: GLS resulting potential drivers explaining the spatio-temporal variability of phosphate patterns in the Ebro basin. 179

Summary
Resum
Resumen

SUMMARY

Scientists broadly agree that the predicted consequences of global change should be considered in any long-term plans prepared for the management of water resources. Although nutrient content in rivers has changed in the last decades due to increasing human activities, studies concerning the effects of environmental change on river water quality patterns still remain scarce. The significance of nutrient inputs is best expressed in terms of in-stream processes, compared to evaluating simple field measurements of nutrient inputs. Modeling tools are necessary to consider the complexity of river networks in the determination of the sources and processes by which nutrients are transported at the basin scale. Mediterranean rivers are especially vulnerable to climate change (potentially affected by the decrease in precipitation and increase of extreme events), and identifying and quantifying nutrient pollution sources and their spatial distribution can improve water resources management in the region.

One of the two main objectives in this thesis was to describe biogeochemical processes involving nutrients in impaired rivers by leaning towards a more integrative approach of biological and hydrological factors acting on in-stream retention at the river basin scale. In Chapters 1 and 2 of this thesis, a hybrid process-based and statistical model (SPARROW, spatially referenced regression on watershed attributes) was applied to a largely disturbed Mediterranean basin in NE Spain. The annual nitrate and phosphate loads reaching the drainage network of the Llobregat River Basin were estimated. The model emphasized the contribution of in-stream processes in nutrient transport and retention, even under disturbed circumstances, and the inter-annual (7 years) effects of hydrological variability on the export of nutrients from the landscape to water bodies. Although forest and grass land cover types predominate, agricultural activities and human agglomerations were significant sources of nutrient enrichment in the Llobregat basin.

The in-stream decay specifications of the SPARROW model were modified to include a partial saturation effect in uptake efficiency (expressed as a power law) and better

capture biological nutrient retention in river systems under high anthropogenic stress. This modification was conceptually framed under the efficiency loss (EL) dynamics, which establishes that the efficiency of process rates relative to available nutrient concentration in streams eventually declines. The stream decay coefficients were statistically significant in both nitrate and phosphate models, indicating the potential role of in-stream processing in limiting nutrient export. However, the EL concept did not reliably describe the patterns of nutrient uptake efficiency for the concentration gradient and streamflow values found in the Llobregat River Basin, putting in doubt its complete applicability to explain nutrient retention processes in stream networks comprising highly impaired rivers. The overall decrease in-stream nutrient removal in the lower parts of the Llobregat basin was probably worsened by the significant chemical and geomorphological impairment found in this area.

The second main objective was to detect common water quality signatures across a network of monitoring points and to attribute potential causes, while tackling the most commonly encountered challenges in classical time-series analysis. Water quality monitoring programs in many river basins have recorded data on dissolved nutrients during several decades. The analysis of such water quality time-series is a fundamental tool for the detection and attribution of global change effects at basin and regional scales. Dynamic Factor Analysis (DFA), a dimension-reduction technique, was implemented in Chapters 3 and 4 of this thesis in order to extract and assess underlying patterns in sets of water quality time-series. The resulting common patterns are then associated to factor loadings, which indicate the weight that each pattern partakes in every monitoring point. These were used to further characterize the temporal (cycles, trends and hydrological dependence) and the spatial (trough regression models and cluster analysis) variability of the extracted water quality signals. The resulting description of the extracted patterns thus facilitated the interpretation and the attribution of drivers behind the observed behavior of the system. DFA was thus used as the main tool to detect common patterns at the basin scale, and the application of complementary methods allowed the detection of

signatures of environmental change on water quality patterns in datasets from three Mediterranean basins.

Using the collection of methodological steps developed in Chapter 3, seasonal cycles and periodicities potentially related to climatic oscillations were identified in the nutrient patterns extracted from time-series in 50 monitoring points across the Ebro River Basin. The periodicities of 3-4 years identified in nitrate and phosphate patterns could potentially be related to climatic oscillations, such as the North Atlantic Oscillation (NAO) and El Niño Southern Oscillation (ENSO) teleconnections, which are known to be influential in the Iberian Peninsula. The overall decrease of phosphate concentration appeared to coincide with the implementation of sewage treatment and with the reduction of phosphate base fertilizers in agriculture. The spatio-temporal variability of nitrate concentration made a clear distinction among sites that had a relationship with hydrology and sites where other processes (e.g. phenology-related) prevailed. Conversely, the spatio-temporal variability of phosphate concentration was more complex and sites in the mid section of the Ebro River were underrepresented by the extracted common patterns.

The proposed collection of methodological steps was successful in providing a complete characterization of patterns and drivers of water quality signatures. This procedure might be applicable not only to Mediterranean basins, since it is able to deal with the most commonly encountered problems in the assessment of global change effects on freshwater ecosystems, such as the lack of continuous and long-term data and the ability to deal with different temporal and spatial scales at which changes and responses occur.

Overall, the work included in the current thesis elucidated the role of global change phenomena such as land-use and climate changes that occur both at regional and global scales in shaping water quality patterns at the river basin scale.

RESUM

El consens dins la comunitat científica estableix que les conseqüències previstes pel canvi global han de ser considerades en els plans a llarg termini elaborats per a la gestió dels recursos hídrics. Tot i que el contingut de nutrients en els rius ha canviat en les últimes dècades a causa de l'augment de les activitats humanes, els estudis sobre els efectes dels canvis ambientals sobre els patrons de qualitat de l'aigua als rius continuen sent escassos. La importància de les aportacions de nutrients s'expressa millor en termes de processos dins el riu, en comparació amb la simple avaluació de mesures de camp d'entrada de nutrients. Les eines de modelatge són necessàries per considerar la complexitat de les xarxes fluvials en la determinació de les fonts i els processos pels quals els nutrients són transportats a escala de conca. Els rius mediterranis són especialment vulnerables al canvi climàtic (caracteritzat per la disminució de les precipitacions i l'augment dels fenòmens extrems), i la identificació i quantificació de les fonts de contaminació de nutrients i la seva distribució espaial pot millorar la gestió dels recursos hídrics a la regió.

Un dels dos principals objectius d'aquesta tesi va ser descriure els processos biogeoquímics que involucren nutrients en rius antropitzats permetent un enfocament més integrador dels factors biològics i hidrològics que actuen sobre la retenció de nutrients en rius a escala de conca. En els dos primers capítols d'aquesta tesi, un model híbrid, entre determinista i estadístic (SPARROW), es va aplicar a una conca mediterrània humanitzada situada al nord-est d'Espanya. Les càrregues anuals de nitrats i fosfats que arriben a la xarxa de drenatge de la conca del riu Llobregat van ser estimades. El model va posar èmfasi en la contribució dels processos en el transport i la retenció de nutrients, fins i tot en circumstàncies alterades, i en l'efecte interanual (7 anys) de la variabilitat hidrològica en l'exportació de nutrients de la fase terrestre a l'aquàtica. Encara que els boscos i les pastures són els usos de sòl predominants, les activitats agrícoles i aglomeracions humanes són fonts importants d'enriquiment de nutrients a la conca del Llobregat.

Les especificacions de retenció en rius trobades al model SPARROW van ser modificades per incloure un efecte parcial de saturació en l'eficiència de la captació (expressat com una llei de potència) i la retenció de nutrients, i així poder representar el component biològic de retenció d'una millor manera en sistemes fluvials sota una alta pressió antropogènica. Aquesta modificació s'emmarca conceptualment en la pèrdua d'eficiència (Efficiency Loss, EL), dinàmica que estableix que l'eficàcia dels processos en relació amb la concentració de nutrients disponibles en rius disminueix amb el temps. Els coeficients de retenció als rius van ser estadísticament significatius en els models de nitrat i fosfat, el que indica el paper potencial de processament als rius en la limitació de l'exportació de nutrients riu avall. No obstant això, el concepte EL no va descriure de manera fiable els patrons d'eficiència en l'absorció de nutrients pel gradient de concentració i els valors de cabal característics de la conca del riu Llobregat, posant en dubte la seva aplicabilitat i la seva capacitat per explicar els processos de retenció de nutrients en les xarxes fluvials que comprenen els rius altament deteriorats. La disminució global en l'eliminació de nutrients al rius en les parts baixes de la conca del Llobregat es va veure agreujada probablement pel deteriorament químic i geomorfològic que es troba en aquesta zona.

El segon objectiu general va ser la detecció de signatures comuns de qualitat de l'aigua en punts de control de xarxes fluvials i la posterior atribució de causes potencials, tenint en compte els problemes més freqüents en l'anàlisi clàssic de sèries temporals. Els programes de monitoratge de qualitat de l'aigua en moltes conques hidrogràfiques van registrar dades sobre nutrients dissolts durant diverses dècades. L'anàlisi d'aquestes sèries temporals de qualitat de l'aigua és una eina fonamental per a la detecció i atribució dels efectes del canvi global a escales de conca i regionals. L'anàlisi dinàmica de factors (Dynamic Factor Analysis, DFA), una tècnica de reducció de dimensió, es va dur a terme en els dos últims capítols d'aquesta tesi per tal d'extreure i avaluar els patrons subjacents en els conjunts de sèries temporals de qualitat de l'aigua. Els patrons comuns resultants s'associen als factors de càrregues, que indiquen el pes amb el qual cada patró participa en cada punt de monitoratge . Aquests dos resultats de DFA es van utilitzar per caracteritzar millor les variacions

temporals (cicles, tendències i dependència hidrològica) i espaials (aplicant models de regressió i anàlisi de clústers) dels senyals de qualitat de l'aigua extrets. La descripció resultant dels patrons extrets d'aquesta manera facilita la interpretació i l'atribució dels factors darrere de la conducta observada al sistema. DFA s'utilitza així com la principal eina per detectar patrons comuns a escala de la conca, i l'aplicació de mètodes complementaris permet la detecció de signatures dels canvis ambientals sobre els patrons de qualitat de l'aigua en dades procedents de tres conques mediterrànies.
Gràcies a l'ús de la col·lecció de passos metodològics desenvolupats en aquest treball, es van identificar cicles estacionals i periodicitats potencialment relacionats amb oscil·lacions climàtiques en els patrons de nutrients extrets de sèries temporals en 50 punts de monitoratge en tota la conca del riu Ebre. Les periodicitats de 3-4 anys identificades en els patrons de nitrat i fosfat podrien potencialment estar relacionats, per exemple, amb l'Oscil·lació de l'Atlàntic Nord (NAO) i El Niño Oscil·lació del Sud (ENSO). Aquestes oscil·lacions són conegudes per ser influents a la Península Ibèrica. La disminució global de la concentració de fosfat va coincidir amb l'aplicació del tractament d'aigües residuals i amb la reducció de fertilitzants de base de fosfat en l'agricultura. La variabilitat espaial i temporal de la concentració de nitrats va fer una clara distinció entre els llocs que tenien una relació amb la hidrologia i els llocs on altres processos (relacionats, per exemple, a la fenologia) van prevaler. Per contra, la variabilitat espaial i temporal de la concentració de fosfat va resultat més complexa i els punts a la secció mitjana del riu Ebre estan insuficientment representats en els patrons comuns extrets.

La col·lecció de passos metodològics proposada en el capítol 3 va tenir èxit en la prestació d'una caracterització completa dels patrons i els controladors de senyals de qualitat de l'aigua. Aquest procediment podria ser aplicable no només a les conques mediterrànies, ja que és capaç de fer front a els reptes més comuns en l'avaluació dels efectes del canvi global sobre els ecosistemes d'aigua dolça, com ara la manca de dades contínues i de llarg termini i la capacitat d'incorporar les diferents escales temporals i espacials on es produeixen els canvis i les respostes del sistema.

El treball inclòs en la tesi actual va contribuir a dilucidar el paper dels fenòmens del canvi global, com el dels usos de sòl i els canvis climàtics, que es produeixen a escales regionals i globals en la configuració dels patrons o signatures de qualitat de l'aigua a escala de conca hidrogràfica.

RESUMEN

El consenso científico indica que las consecuencias del cambio global previstas deben ser consideradas en los planes de gestión a largo plazo de los recursos hídricos. A pesar de que la concentración de nutrientes en los ríos ha cambiado en las últimas décadas debido al aumento de las actividades humanas, los estudios sobre los efectos de los cambios medioambientales sobre patrones de calidad de agua de río siguen siendo escasos. La importancia de los aportes de nutrientes se expresa mejor en términos de procesos en el río, en comparación con la simple evaluación de las medidas de campo de entrada de nutrientes. Las herramientas de modelización son necesarias para considerar la complejidad de las redes fluviales en la determinación de las fuentes y los procesos por los cuales los nutrientes son transportados a escala de cuenca. Los ríos mediterráneos son especialmente vulnerables al cambio climático (caracterizado por la disminución de las precipitaciones y el aumento de los fenómenos extremos), y la identificación y cuantificación de las fuentes de contaminación de nutrientes y su distribución espacial pueden mejorar la gestión de los recursos hídricos en la región.

Uno de los dos principales objetivos de esta tesis fue el de describir los procesos biogeoquímicos que involucran nutrientes en ríos antropizados permitiendo un enfoque más integrador de los factores biológicos e hidrológicos que actúan sobre la retención de nutrientes en ríos a escala de cuenca. En los dos primeros capítulos de esta tesis, un modelo híbrido, entre determinista y estadístico (SPARROW), fue implementado en una cuenca mediterránea altamente humanizada, situada en el nordeste de España. De esta manera se estimaron las cargas anuales de nitratos y fosfatos que llegan a la red de drenaje de la cuenca del Río Llobregat. El modelo dio énfasis a la contribución de los procesos de transporte y retención de nutrientes, principalmente en circunstancias de cauces hídricos alterados, y al efecto interanual (7 años) de la variabilidad hidrológica en la exportación de nutrientes desde la fase terrestre a la acuática. A pesar de que bosques y pastizales predominan como usos de

suelo, las actividades agrícolas y las aglomeraciones humanas son también fuentes importantes de enriquecimiento de nutrientes en la cuenca del Río Llobregat.
Las especificaciones de retención en el río incorporadas en el modelo SPARROW fueron modificadas para incluir un efecto parcial de saturación en la eficiencia de la captación y la retención de nutrientes (expresada como una ley de potencia), de manera a capturar mejor el componente biológico en sistemas fluviales bajo alta presión antropogénica. Dicha modificación se enmarca conceptualmente en la pérdida de eficiencia (Efficiency Loss, EL), dinámica que establece que la eficacia de las tasas de procesos en relación con la concentración de nutrientes en ríos disminuye con el tiempo. Los coeficientes de retención en el río fueron estadísticamente significativos en los modelos de nitrato y fosfato, lo que indica el papel potencial de procesamiento en ríos en la limitación de la exportación de nutrientes río abajo. No obstante, el concepto EL no describió de forma fiable los patrones de eficiencia en la absorción de nutrientes dentro del gradiente de concentración y los valores de caudal característicos de la cuenca del Río Llobregat, poniendo en duda su capacidad para explicar los procesos de retención de nutrientes en las redes fluviales de ríos altamente deteriorados. La disminución global de la eliminación de nutrientes en la parte baja de la cuenca del Llobregat se vio probablemente agravada por el significativo deterioro químico y geomorfológico que predomina en dicha zona.

El segundo objetivo general fue la detección de patrones comunes de calidad del agua en puntos de control de redes fluviales y la posterior atribuición de sus causas potenciales, teniendo en cuenta los problemas más frecuentes en el análisis clásico de series temporales. Los programas de monitoreo de calidad de agua en cuencas hidrográficas han registrado datos sobre nutrientes disueltos durante varias décadas. El análisis de dichas series temporales es una herramienta fundamental para la detección y atribución de los efectos del cambio global a escalas de cuenca y regionales. El análisis dinámico de factores (Dynamic Factor Analysis, DFA), una técnica de reducción de dimensión, fue implementado en los dos últimos capítulos de esta tesis con el fin de extraer y evaluar los patrones subyacentes en los conjuntos de series temporales de calidad de agua en tres cuencas mediterráneas. Los patrones

comunes resultantes se asocian a factores de carga, que indican el peso con el que cada patrón participa en cada punto de monitoreo. Estos resultados se utilizaron para caracterizar mejor las variaciones temporales (ciclos, tendencias y dependencia hidrológica) y espaciales (modelos de regresión y análisis de clústeres) de los patrones comunes de calidad de agua. De este modo, la descripción resultante de los patrones extraídos facilita la interpretación y la atribución de los factores causantes de cambios en el sistema estudiado. El análisis de factores dinámicos (DFA) fue utilizado como herramienta principal para detectar patrones comunes a escala de cuenca, y la posterior implementación de métodos complementarios permitió la detección de señales de cambios medioambientales sobre los patrones de calidad de agua en datos procedentes de tres cuencas mediterráneas.

La colección de pasos metodológicos introducida en el capítulo 3 permitió la identificación de ciclos estacionales y de periodicidades potencialmente asociados a oscilaciones climáticas en los patrones de nutrientes extraídos de 50 series temporales en la cuenca del río Ebro. Concretamente, las periodicidades de 3-4 años identificadas en los patrones de nitrato y fosfato podrían estar relacionadas con modos característicos de la Oscilación del Atlántico Norte (NAO) y de El Niño-Oscilación del Sur (ENSO), conocidas por ser influyentes en la Península Ibérica. Asimismo, la disminución general de la concentración de fosfato coincidió aparentemente con la implementación del tratamiento de aguas residuales y con la reducción de fertilizantes de base de fosfato en la agricultura en la cuenca del Río Ebro. La variabilidad espacio-temporal de la concentración de nitratos hizo una clara distinción entre los sitios que tenían una relación con la hidrología y los sitios donde otros procesos (i.e., fenológicos) prevalecieron. Por el contrario, la variabilidad espacio-temporal de la concentración de fosfato resultó ser más compleja y los sitios en la sección media del Río Ebro no fueron correctamente representados en los patrones comunes extraídos.

La metodología propuesta logró una caracterización completa de los patrones y los factores envueltos en la variabilidad de la calidad del agua. Este procedimiento podría ser útil no sólo en cuencas mediterráneas, ya que es capaz de sortear los retos más

comunes en la evaluación de los efectos del cambio global sobre los ecosistemas de agua dulce, tales como la falta de datos continuos y de largo plazo y la capacidad de lidiar con las diferentes escalas temporales y espaciales en las que se producen los cambios y respuestas del sistema.

El trabajo incluido en esta tesis contribuyó a dilucidar el papel de los fenómenos del cambio global, como los usos de suelo y los cambios climáticos, que se producen tanto a escalas locales, regionales, y globales en la configuración de los patrones de calidad del agua a escala de cuenca.

General Introduction

General Introduction

Global environmental change

The interacting systems that make up the Earth constantly undergo changes and are thereby naturally dynamic. The rate and the intensity of recent changes however have no precedent since, at least, the beginning of the current geological epoch (Meybeck, 2003; García-Ruiz et al., 2011). Anthropogenic activities have been shown to add to the intrinsic natural variability of the Earth system and to counteract or enhance natural changes. Such transformations and alterations are components of what is currently known as global change (Rosenzweig et al., 2008; Rabalais et al., 2009). According to Turner et al. (1990), the term global change includes both globally systemic change and globally cumulative change (Table 1).

Global Environmental Change		
Type	**Characteristic**	**Examples**
Systemic	Direct impact on globally functioning system	Industrial and land use emissions of greenhouse gases. Industrial and consumer emissions of ozone-depleting gases Land cover changes in albedo
Cumulative	Impact through worldwide distribution of change	Groundwater pollution and depletion Species depletion/genetic alteration (biodiversity)
	Impact through magnitude of change (share of global resource)	Deforestation Industrial toxic pollutants Soil depletion on prime agricultural lands

Table 1: Types of global environmental change according to Turner et al. (1990)

In view of the current accelerated changes to the environment, a new geological epoch, the Anthropocene, has been proposed based on the idea that humans are creating a physical and biological environment that is distinct from anything before and that is likely to leave a substantial trace in the geological record of Earth's history (Ellis & Trachtenberg, 2014). Human population growth is, in fact, considered the main driver responsible for global change over time (Klein Goldewijk et al., 2011).

The aspect of global change regarding climate has gained increased attention from scientists and policy makers in the last decades. As a consequence, initiatives like the creation of the Intergovernmental Panel on Climate Change (IPCC) have taken place. In its most recent assessments, the IPCC clearly states, based on the existing literature on climate change and its consequences on our environment, that the consensus of scientific opinion is that Earth's climate is being affected by human activities (Parry, 2007 ; IPCC, 2014). Oreskes (2004), for instance, analyzed 928 abstracts published in refereed scientific journals between 1993 and 2003, and listed in the ISI database with the keywords 'global climate change', in order to assess the consensus on this topic. The author found that 75% of the papers either explicitly or implicitly accepted the consensus view, while the remaining 25% took no position on current anthropogenic climate change. Even the slight decrease in global warming in the last two decades (starting on 1994) has been shown to have a direct human origin (Estrada *et al.*, 2013a). Scientific evidence and available knowledge thereby render unquestionable the direct link between anthropogenic activities and current trends in climate change. Nevertheless, many details about climate interactions are still not well understood, and further research is vital to provide a better basis for understanding climate dynamics (Oreskes, 2004).

Conversely, land use has initially been considered a local environmental matter, though it is presently becoming an issue of concern at regional and global scales. Land-use activities, whether converting natural landscapes for human use or changing land management practices, have indeed transformed a large proportion of the Earth's surface (Foley et al., 2005). Land-use changes can also interact with climate on the

regional scale through changes in surface energy and water balance (Kalnay & Cai, 2003). For instance, agricultural lands and practices contribute to the emission of greenhouse gases (GHG), which, in turn favors global warming and triggers changes in climatic patterns at local and regional scales. Even more subtle changes in land management and land-use practices, such as forest management and carbon emission or sequestration, can have important consequences on climate (Verburg et al., 2011).

Local changes in land surface properties associated with changes in vegetation can have impacts on continental and atmospheric circulation at the global scale, with potential influence on regional and continental climate. The effects of landscape alterations upon climate comprise distant impacts upon local circulation regimes known as land-use driven 'teleconnections' because effects are connected to remote regions from the actual changes in surface characteristics (Eastman et al., 2001). The overall consequences of global change related to land use may indeed rival climate change in environmental and societal impacts (Tilman et al., 2001). In general, however, they act and contribute synergistically to environmental change across all spatial scales.

Even though water is considered the most essential of natural resources, freshwater systems across the globe remain to be directly threatened by human activities (Vörösmarty *et al.*, 2010). Anthropogenic pressures on river systems have indeed accelerated in the past 50 years, and are now balancing the Earth system natural forcing (Meybeck, 2003). Since freshwater availability is the basis to sustain life, water resources and their vulnerability to ongoing global environmental change should be a priority topic to assess (Vörösmarty et al., 2000).

The vast majority of studies has focused on water supply and availability under current and projected patterns of global environmental change. It is also vital, however, to study the consequences of such patterns on the quality of water resources. Land use and climate are generally the main determinants of hydrology and water quality, and are therefore considered to be key drivers of functional and structural

changes in freshwater ecosystems (Stevenson & Sabater, 2010). Therefore, investigating the consequences of changes in land uses and climate on freshwater systems is an important first step to describe current and potential global change impacts.

Environmental change in the Mediterranean region

Mediterranean watersheds are historically amongst the most heavily impacted by anthropogenic activities (Liquete et al., 2009; Ludwig et al., 2009). The variability of water resources availability is one of the critical factors that determine the socioeconomic development of the Mediterranean region. In spite of this variability being inherent to climate, it has been recently exacerbated by the increase of human uses and alterations on the hydrologic cycle. Alterations to climate include the decrease of rainy days, as well as the increase of extreme events (Hirabayashi et al., 2008; Sillmann & Roeckner, 2008). Furthermore, temperature in the Mediterranean region is predicted to increase within the current century, as well as the occurrence of droughts (Giorgi & Lionello, 2008). The resulting imbalance between the available water resources during long drought periods and the increasing water demand has actually become a recurrent predicament that governments in this region are seeking to solve.

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). For instance, a generalized decrease in annual precipitation has been reported in Mediterranean basins and several studies have pointed to alterations in precipitation intensity and drought occurrence (García-Ruiz et al., 2011).

Landscape changes in the region alter the already highly variable hydrological cycle by increasing peak flood flows, decreasing lag times between rainfall and runoff, and

decreasing dry season flows (Cooper et al., 2013). Additionally, a decline in overland flow has been observed in experimental basins throughout the world as a consequence of changing the land cover towards plantations (Farley et al., 2005). A large part of rivers and streams in the Mediterranean region have been dammed or diverted to guarantee water supply for human activities (Gasith & Resh, 1999). In Spain, for instance, the construction rate of large dams peaked during the 1960s and 1970s (Cooper et al., 2013). All these alterations add up to the characteristic hydrological variability and water scarcity in the region, and create further pressures on freshwater systems.

Population growth during the last century and the beginning of the current one resulted in the expansion of agriculture and the decrease in forested areas (Tilman et al., 2001). Particularly, human population density and urban area in the Mediterranean Basin increased during the decade of the 1990s (+17% change in urban areas contrasted with +1% in expansion of agricultural areas; Cooper et al., 2013). In the recent decades, however, social and economic changes (particularly at high land elevations) have resulted in the abandonment of most unimportant agricultural fields and the subsequent return to land covers such as forests and shrublands.

In addition to the past and present observed pressures in the region, the Mediterranean has been identified as one of the most vulnerable regions to projected global change events. In a study by Schröter et al. (2005), multiple potential impacts such as water shortages, increased risk of forest fires, shifts in the distribution of typical tree species, and losses of agricultural potential were projected based on predictions from Global Circulation Models (GCM) linked to future scenarios for ecosystem services in Europe. All these forecasted pressures are indeed mainly related to an increase in temperatures and a reduction in precipitation.

Scientific literature extensively reports the consequences of global change (understood as the synergy between climate change and direct action of human activities on the territory; U.S. Global Change Research Act, 1990) on river hydrology. However,

studies dealing with the consequences of global change on the water quality of rivers are scarce. Thus, water agencies have little evidence to develop long-term strategies to avoid the likely negative effects of global change on rivers. This is especially true in Mediterranean regions, where there is a dearth of knowledge compared to temperate climate regime studies (e.g. Qian et al., 2010). Mediterranean streams and rivers can be particularly vulnerable to water pollution due to the presence of additional pressures such as damming, water extraction and urbanization (Sabater & Tockner, 2010). The study of the consequences derived from land-use alterations and climate change on water quality is therefore essential to achieve the sustainable use of water resources; that is, one that benefits both the environment and the human population in the region.

Freshwater quality and nutrient pollution

The economic development has favored the enrichment of continental and coastal waters with nitrogen and phosphorus derived from anthropogenic sources. Globally, rivers export about 59 Tg N yr^{-1}, of which 11 Tg N yr^{-1} are transported to dry lands and inland waters, and the remaining 48 Tg N yr^{-1} are further transported to the coastal zone (Boyer *et al.*, 2006b). The identification and quantification of nutrient inputs and their sources is therefore vital to achieve good water quality. Further understanding of the coupling between terrestrial nutrient sources and in-stream transport within a river network is fundamental for the management of receiving water bodies (Ensign & Doyle, 2006; Wollheim et al., 2006).

In Mediterranean rivers, nutrient export contributes to eutrophication because of the co-existence of naturally-occurring low flows and high water demand (Caille, 2009). Undesirable levels of in-stream nutrient concentrations are generally the major cause behind proliferation of algal masses in rivers and streams. Wastewaters are also major source of nitrogen and phosphorus, particularly in urbanized catchments. During seasonal or longer-term droughts, wastewater discharges can be the major or solely provider of flow and nutrient flux, often accounting for more than 50% of nitrogen

and phosphorus in-stream loads (Martí et al., 2004; Cooper et al., 2013). Water pollution acts as an additional stressor on the frequently scarce water supply and aggravates the damage to stream health in the Mediterranean region (Gasith & Resh, 1999). Moreover, the existence of high nutrient concentration does not only affect the structure and function of the aquatic ecosystem but may also be harmful to human health.

Analyzing the influence of nutrient sources on water bodies, such as land use and wastewater effluents, often requires the evaluation of nutrient cycling processes at a larger spatial scale (e.g., at the basin scale) than the reach level (Dodds & Oakes, 2006). Considering that a reach is part of a stream or river which, in turn, belongs to a network of streams within a basin, it becomes evident that conditions found at the reach scale are affected by those found at larger spatial scales (Allan et al., 1997). Nutrient transport and retention processes have been mainly studied at the reach scale and characterized by metrics provided by the Nutrient Spiraling concept (Stream Solute Workshop, 1990). According to this conceptual framework, the complete cycle of a nutrient atom is examined as it travels downstream, including its transformation by biological uptake and its subsequent release to the water column. Overall, nutrient spiraling metrics combine both the hydrological and biological factors that are involved in nutrient processing in streams. Current modeling tools provide a valuable means for unraveling such processes in systems larger than headwater streams. Models assist in the assessment of nutrient cycling from a stream network perspective by up-scaling the most relevant mechanisms to river basin and regional scales.

Models: tools for visualizing and analyzing water quality patterns at the basin scale

Much of the investigation into freshwater quality has been carried out at the individual monitoring location (Miller et al., 2014). The development and implementation of spatio-temporal statistical water quality models at the basin scale can provide more powerful tools to investigate patterns of long-term change in nutrient concentration. Such models make possible the study of trends and seasonality in nutrients using

combined information from the hydrologically-defined basin area, rather than on an individual site basis. This approach increases the sample size considerably and hence the statistical power to detect patterns in water quality records.

Appropriate tools and techniques are needed in order to identify the stressors that alter water quality and, most importantly, to link the stream's response to the responsible stressor (Allan, 2004). Modeling has been increasingly used to portray current water quality and predict potential impacts of future scenarios. Several hydrologic and water-quality models have been used to particularly describe contaminant sources and transport in surface waters. The level of complexity or detail represented by model descriptions of hydrologic and biogeochemical processes commonly varies with the extent to which methods are used, whether deterministic or statistical, to describe and estimate these processes (Schwarz et al., 2006).

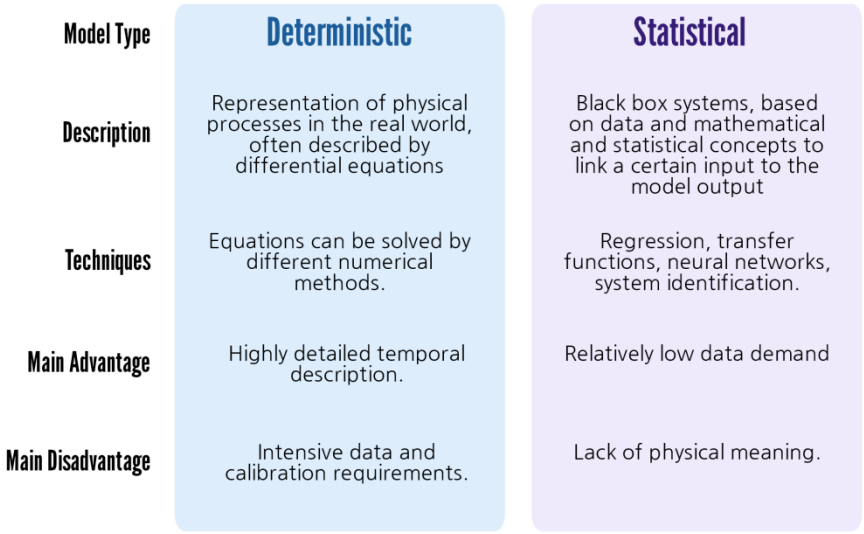


Figure 1: Main characteristics of Deterministic and Statistical Models

Statistical models usually reflect more simplistic model structures and assume limited a priori knowledge of a variety of processes (Schwarz et al., 2006). Their advantage is that they can be readily applied in large basins (often relying on available monitoring records from water agencies) and are able to quantify the errors in model parameters and predictions. These models commonly provide an excellent fit to the observations, but have the disadvantage of providing little understanding of the processes that affect contaminant transport. Contrastingly, mechanistic water-quality models have a very complex mass-balance structure describing hydrologic and contaminant transport processes, often according to fine temporal scales (Schwarz et al., 2006).

The complexity involved in mechanistic models tend to generate intensive data and calibration requirements and to limit their application to relatively small watersheds. Because these models are frequently calibrated manually, robust measures of uncertainty in model parameters and predictions cannot be estimated (Schwarz et al., 2006). Despite the common use of highly mechanistic models, especially among water-resource managers, there are growing concerns about whether sufficient water-resources data and knowledge of biogeochemical processes exist to reliably support the general use of such highly complex descriptions of processes (Beven, 2002). A balance between data-hungry mechanistic models and over-simplistic statistical models should be sought in order to adjust the modeling objectives to data availability.

Water quality monitoring programs in many watersheds have recorded data on dissolved nutrients during several decades, thus making it possible to carry out analyses of periods long enough to show significant changes in concentration and loading (Ibáñez et al., 2008). Studies have shown, for instance, that nutrient content and loading of rivers has changed in the last decades due to increasing human population and economic activities (Grizzetti et al., 2011). However, working with such environmental datasets is seldom an easy task. Uneven frequency sampling and changes in monitoring strategy challenge the most basic data quality requirements and statistical assumptions of most of the available modeling methods.

Commonly used model in time-series analysis are limited to characterizing the spectral density to detect any periodicities in the data, and thus do not necessarily allow the user to identify underlying common patterns in environmental records (Zuur *et al.*, 2003; Zuur *et al.*, 2007). Moreover, most of the methods above do not easily accommodate missing observations, which are extremely abundant in environmental databases, and are not suitable for extracting patterns of spatio-temporal variability or for unraveling any interactions between the observations involved in the analysis. These issues should be addressed since the impact of global environmental change on a given ecosystem is frequently comprised of overlapping stressors acting at both regional and local scales.

Methodologies that explicitly take into account the inextricable relationship between temporal and spatial patterns of environmental change are essential to identify the drivers and the processes linked to the observed changes. Furthermore, methods and tools that are able to deal with low-quality databases are needed to explore global change patterns in basins where data scarcity prevails. The use of appropriate tools for the identification of water quality patterns at the basin and regional scales enables the visualization and interpretation of the governing processes across space and time, and allows the investigation of further hypotheses under changing scenarios.

General Objectives

General Objectives

The work in this thesis aimed to support the existing knowledge of the impact of global change on the water quality of Mediterranean rivers and to generate original contributions within the field of detection of climate and land-use effects on river ecosystems in the region.

The main objectives were the following:

- To identify and quantify the main nitrate and phosphate sources and link their variability across time and space to land-use and climatic conditions in a Mediterranean basin.

- To describe biogeochemical processes involving nutrients in impaired rivers by leaning towards a more integrative approach of biological and hydrological factors acting on in-stream retention at the river basin scale.

- To detect and characterize common water quality signatures across a network of monitoring points while tackling the most commonly encountered challenges in classical time-series analysis.

- To characterize the spatio-temporal variability of nitrate and phosphate concentration in a Mediterranean basin and attribute the potential drivers behind the underlying patterns in the context of global change.

Methods and Study Sites

Modeling methods

Two main modeling methods were applied in this thesis in order to describe the link between nutrient variability and changes in land-uses and climate observed in Mediterranean basins. These two models are briefly described in the current section.

Firstly, the SPARROW model was applied to study in-stream nutrient processing and the subsequent up-scaling to the basin scale as well as to identify the variability and apportionment of nutrients and their sources. The detailed methodology is included in Chapters 1 and 2 of this thesis. Secondly, Dynamic Factor Analysis was used as multivariate analysis approach to investigate common patterns in a set of water quality time-series in the context of global environmental change. An exhaustive description of this method is included in Chapter 3 and further exemplified in Chapter 4.

Nutrient load estimation and source apportionment at the basin scale: *The SPARROW Model*

The relative importance of nutrient sources associated to human activities is better expressed in terms of the processes that occur within the rivers compared to merely analyzing measurements at individual locations. Modeling tools become useful in determining the sources and the processes by which pollutants are transported at the basin scale (Smith & Alexander, 2000; Grizzetti *et al.*, 2005).

When selecting an appropriate modeling tool, a balance between model complexity and data availability must be reached. Most modeling efforts at the basin level have employed mechanistic models, which require enormous amount of data at highly detailed temporal and spatial scales, but that in turn are able to supply dynamic responses at fine temporal and spatial resolutions (Schoumans et al., 2009). Detailed input data are seldom readily available and thus a greater number of assumptions must be made to implement the mechanistic approach. On the other hand, conventional empirical models frequently come across the challenge of up-scaling the

biogeochemical processes that affect nutrient fluxes along stream networks and reservoirs within a basin, with no guarantee that measurements performed at one particular scale will be valid to parameterize processes at different scales.

The hybrid statistical and process-based model SPARROW (Spatially Referenced Regression On Watershed attributes, developed by the United States Geological Survey) attempts to address the aforementioned issues by spatially linking watershed characteristics to the stream network and by using mechanistic functions based on empirical observations to describe nutrient dynamics. The mechanistic mass transport components of SPARROW (Schwarz et al., 2006) include surface-water flow paths (channel time of travel, reservoirs), non-conservative transport processes (i.e., first-order in-stream and reservoir decay), and mass-balance constraints on model inputs (sources), losses (terrestrial and aquatic losses/storage), and outputs (riverine nutrient export).

Parameters are estimated by spatially correlating water quality observations with geographic data on pollutant sources and climatic and hydrogeologic properties. Parameter estimation ensures that the calibrated model will not be more complex than can be supported by the data and SPARROW provides robust measures of uncertainty (Schwarz et al., 2006). Furthermore, the robustness of the parameter estimates can be evaluated by applying a bootstrap algorithm incorporated in SPARROW.

The fundamental unit of study of the SPARROW model is the stream reach and its associated incremental watershed. Conceptually, the nutrient load leaving a reach is the sum of two components, the load generated within the upstream reaches and transported to the reach via the stream network and the load originated within the reach's incremental watershed and delivered to the reach segment (Schwarz et al., 2006). Nutrient flux generated within the incremental watershed is subject to attenuation in the terrestrial phase (represented by a land-to-water delivery factor) and to in-stream and reservoir decay within the reach. The load generated upstream undergoes additional in-stream attenuation in the reach being evaluated.

Figure 1: The SPARROW model relates measured nutrient flux (i.e., monitoring points A and B) in rivers and streams to watershed characteristics associated to nutrient sources (point and diffuse) and to nutrient mobilization and decay processes (land-to-water delivery and in-stream decay).

The SPARROW model works at the annual time-step. Limitations therefore include the consideration of lower-than-annual variations in nutrient flux. The annual resolution is however generally sufficient to locate high risk areas within the basin that are vulnerable to particular sources of pollution (Schoumans et al., 2009). The SPARROW model has been previously applied to the regional assessment of water quality in the United States of America (Preston *et al.*, 2011b) and in the evaluation of nutrient loads in New Zealand (Alexander *et al.*, 2002a). As an example, the delivery of nitrogen to the Gulf of Mexico (where eutrophication and hypoxic areas are a recognized problem) and the relationship with nutrient aquatic processing and stream size was characterized by applying the SPARROW model in the Mississippi River basin (Alexander et al., 2000).

In-stream nutrient decay in SPARROW models

SPARROW models have implemented a first-order decay equation to quantify the proportion of constituent being removed in a particular reach (Alexander et al., 2000). This implies that nutrient removal rates would increase linearly with nutrient concentration, with no signs of potential saturation. However, it has recently been argued that the concentration of nutrients in streams determines the rate of uptake, which could eventually reach partial saturation. Data of nutrient spiraling metrics have shown that the efficiency of process rates relative to increasing concentration eventually declines.

According to a recent study that evaluated the saturation of nitrogen cycling in a broad range of streams in Kansas (USA; O'Brien et al., 2007), N uptake increases with NO_3^- concentration, but at a slower rate than the increase in concentration thus resulting in a loss of nitrogen processing efficiency. Within this Efficiency Loss (EL) concept, log-transformed areal uptake rates increase with log-transformed nutrient concentration, where the slope of the relationship is less than one. Several subsequent studies have also observed that the empirical relationship between areal uptake rate and increasing NO_3^- concentration in streams shows a partial saturation effect (Hall *et al.*, 2009; Mulholland *et al.*, 2008; Mulholland *et al.*, 2009).

The formulation for reach nutrient attenuation in SPARROW was thereby modified (Aguilera et al., 2013; Chapter 1) in this thesis to include a wider range of stream uptake dynamics (including first-order kinetics and EL; Figure 2). The new model specification was based on an empirically derived power law related to available nutrient concentration. Here, biological factors affecting nutrient removal were also taken into account and were represented by the uptake velocity, which is the downward velocity at which nutrients are removed from the water column by migrating into the benthic compartment. The hydrological conditions of the reach being evaluated were reflected in the hydraulic load, estimated as the ratio between reach time-of-travel (water residence time) and water depth in a given reach.

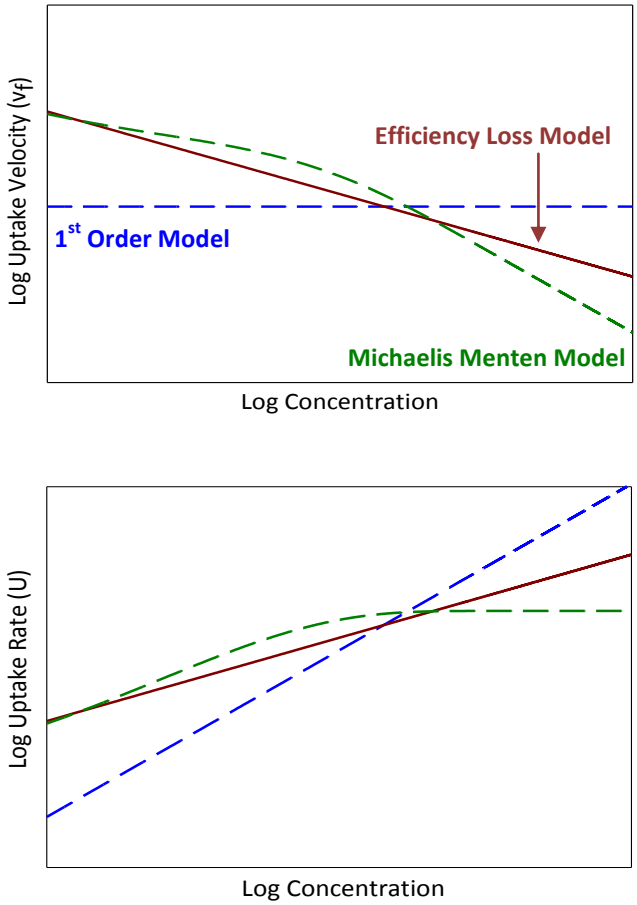


Figure 2: The response of a particular stream subjected to periodic nutrient inputs may follow 1st order kinetics or Michaelis-Menten saturation model, but the response of streams receiving continuous loading could be best described by the Efficiency Loss model (O'Brien et al., 2007).

Detection and characterization of nutrient patterns in a river basin: *Dynamic Factor Analysis*

Ecosystem processes involve interactions at multiple scales, where processes at one temporal or spatial scale (e.g., a river basin) interact with processes at another scale (e.g., regional patterns in temperature and precipitation). Understanding the interplay

between fine- and broad-scale patterns and processes is key to understanding ecosystem dynamics (Qian et al., 2010). This can be examined by using long-term series of data of the respective variables. The idea that a time-series exhibits repetitive or regular behavior over time is of fundamental importance because it distinguishes time-series analysis from classical statistics, which assumes complete independence over time (Shumway & Stoffer, 1982).

In spite of the recent inclination towards generating long-term high-frequency environmental records, the reality is that currently available water quality data in most river basins are usually characterized by being relatively short and discontinuous. This is the reality from many monitoring long-term schemes carried out by water agencies. Therefore, a compromise has to be reached, where scientists and policy makers often need to rely on the best available data and discard other potentially useful datasets due to the presence of data gaps and the length and frequency of data sampling.

The varying sampling frequency and the presence of gaps in datasets are in fact among the first encountered challenges when selecting the appropriate method for analysis. Commonly used techniques to characterize environmental time-series require complete datasets, and often even long time-scales (Bers et al., 2013). Few of the currently available methods are able to cope with the challenges mentioned above.

Schoellhamer, (2001), for instance, has suggested a modified singular-spectrum analysis algorithm to obtain spectral estimates from records with a large fraction of missing data. Also, the weighted wavelet Z-transform (WWZ) method can handle unevenly sampled data but large gaps in datasets would not be well compensated for identifying temporal pattern (Foster, 1996). Another example is the seasonal Kendall trend analysis, which allows missing values (Hirsch et al., 1982).

Nevertheless, commonly used methods such as spectral analysis, wavelet analysis, autoregressive integrated moving average (ARIMA) and Box-Jenkins models usually require stationary and continuous time-series (Zuur et al., 2003a). In addition,

methods like spectral analysis are limited to characterizing the spectral density to detect periodicities in the data, and thus do not necessarily allow the user to identify common patterns embedded in multivariate time-series (Zuur *et al.*, 2003b)

Dynamic Factor Analysis (DFA) is a dimension-reduction method that estimates underlying common patterns in a set of time-series (Zuur et al., 2003a). DFA is not a new technique and it has been widely and initially used in econometric and psychological fields (Harvey, 1989). Most recently, DFA has been applied to fisheries (Zuur *et al.*, 2003b; Zuur & Pierce, 2004; Villizzi, 2012) and groundwater datasets (Muñoz-Carpena et al., 2005; Kuo et al., 2013).

DFA is based on structural time-series models handled within the state-space model framework. A considerable advantage of the state-space approach is the ease with which missing observations can be dealt with. Holmes (2013) performed a series of modifications in the Expectation Maximization (EM) update equations for missing values in Multivariate Autoregressive State-Space (MARSS) models, based on the work published by Shumway & Stoffer (1982) and Zuur et al. (2003a). Furthermore, DFA allows for multivariate analysis whereas frequently used techniques are based on a univariate approach. The main disadvantage of DFA is that it can be computationally intensive, depending mainly on the number of variables involved and on the length and quality of the records.

DFA is able to decompose n number of time-series into common patterns and associated error terms (Fig. 3). In DFA, the estimation of common patterns is done simultaneously using the maximum likelihood method. These patterns are modeled as random walks and thus are allowed to be stochastic. This means that their shape can change over time and that it is not necessarily restricted to be a straight-line trend or a cosine-function cyclic or seasonal component (Zuur et al., 2007).

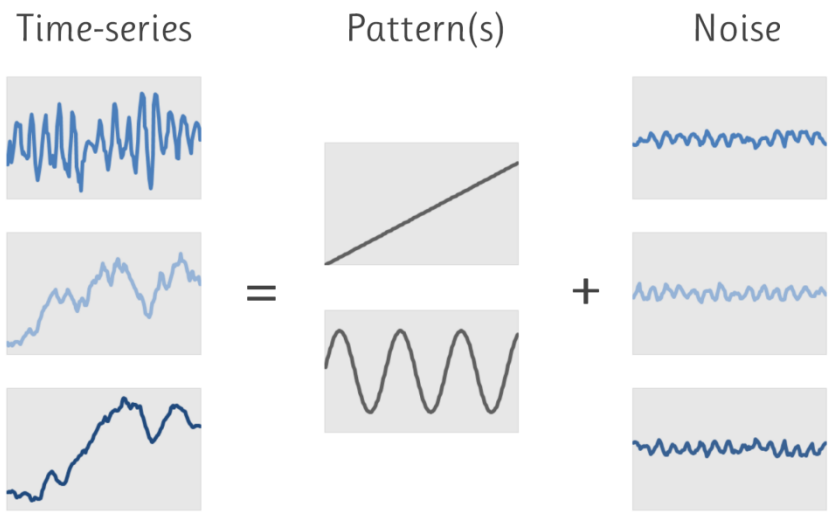


Figure 3: Dynamic Factor Analysis decomposes a set of time-series into a linear combination of common patterns plus associated noise terms.

The resulting patterns are in turn related to factor loadings, which indicate the weight or importance that each pattern has in every monitoring point included in the analysis. These two end products are then further analyzed in order to characterize the temporal variability (cycles, trends and hydrological dependence) and the spatial distribution of the extracted water quality signals. The overall result is then a full description of the water quality signals present in different points across a river basin. From this point, potential causes can be attributed to the observed spatio-temporal variability of the water quality signals in the context of environmental change at regional and global scales.

An example of process mentioned above can be observed in the following scheme:

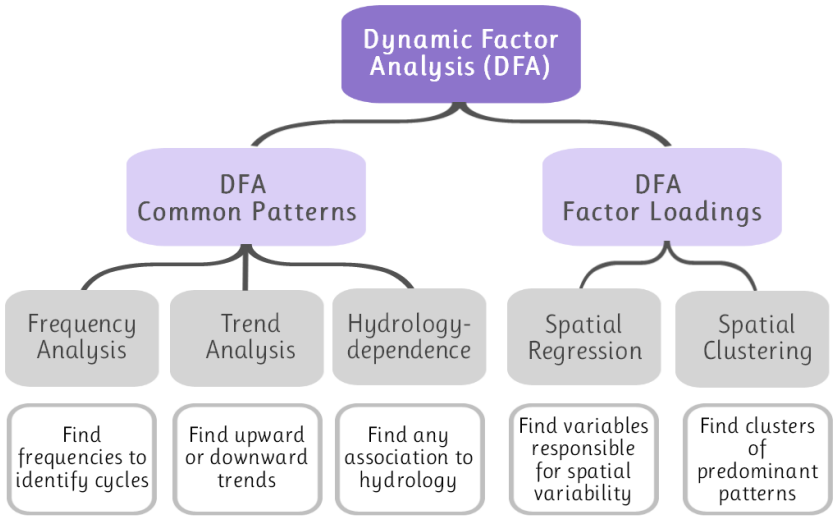


Figure 4: Set of methods applied to the output of Dynamic Factor Analysis in order to attribute potential causes and drivers to the observed water-quality spatio-temporal variability.

Study Sites

This work encompasses three river basins in Spain discharging into the Mediterranean Sea (Fig. 5). Chapters 1 and 2 of this thesis deal with the Llobregat River Basin (NE Spain), whereas Chapters 3 and 4 also include the Ebro River Basin and the Júcar River Basin District. These three basins were selected based on previous analysis carried out within the SCARCE Project Consortium.

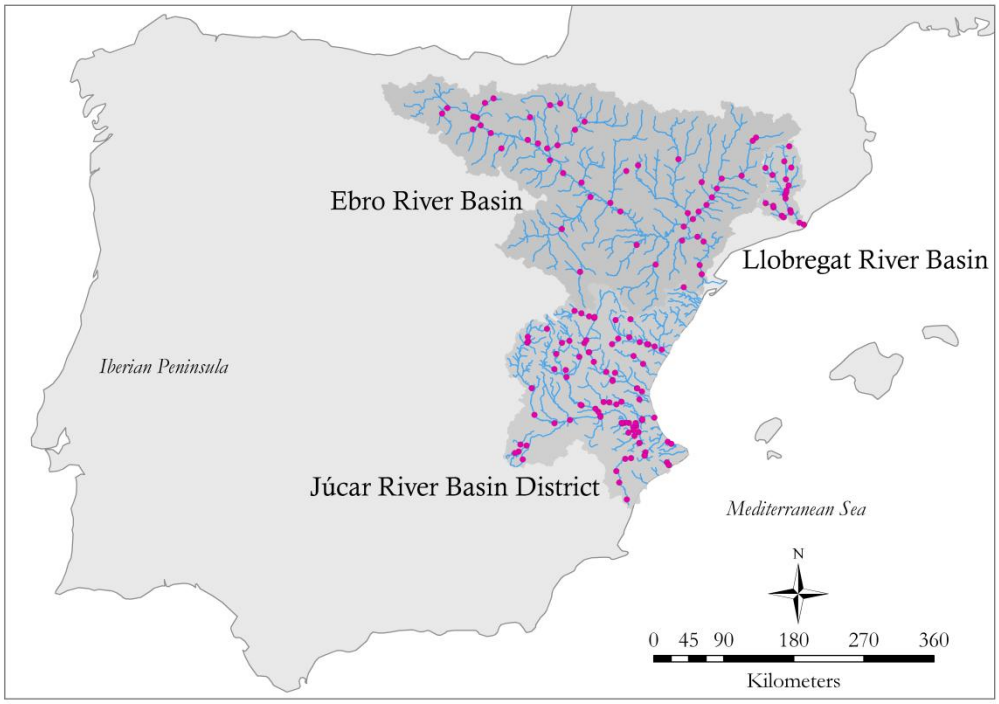


Figure 5: Water quality data points in the Ebro (n=50), Júcar (n=90), and Llobregat (n=20) basins (NE Spain). Datasets and relevant geographical information were obtained from the corresponding public water agencies in charge of the management of these basins.

Llobregat River Basin

The Llobregat River runs across NE Spain and has a total length of 156 km. The Llobregat hosts more than 3 million people in its watershed, and its waters are intensively managed for drinking, agricultural, and industrial purposes (Marcé et al., 2012). In fact, this river is the main drinking water resource of the city of Barcelona. The hydrological regime of the river is typically Mediterranean and it is defined by low basal flows (5–14 $m^3 s^{-1}$), extremely high peak events and a wide range (<2 to 130 $m^3 s^{-1}$) of monthly discharge values (Sabater & Tockner, 2010).

The drainage area spreads over and extension of 4,948 km^2 and it is mainly characterized by a calcareous substratum that determines the high water alkalinity (Sabater *et al.*, 1987). Average annual rainfall in the watershed is 610 mm and average annual temperature is around 15 °C. The Llobregat is regulated through three large reservoirs located in its upper part, as well as by several weirs scattered along the main river channel. The main land cover types in the watershed include forested and shrub land areas, situated mainly in the middle and upper parts of the basin (Fig.6). Dryland crops and vineyards tend to predominate in the mid section and along the Anoia River region (SW of the Llobregat basin; Fig. 6).

During the last century, the Llobregat River Basin has suffered of progressive affection by industrial and urban sewage as well as by runoff from agricultural areas that cannot be diluted by its natural flow. Thus, high concentration of nutrients, pesticides and pharmaceuticals compounds are found in rivers and streams of the Llobregat basin (Muñoz *et al.*, 2009b). The increasing presence of pollutants in the environment and the subsequent effect on the functioning or the river ecosystem are exacerbated during drought periods. Furthermore, the river is heavily managed in its lower course and water that once run to the sea is now pumped upstream to increase the natural flow, recharge the delta wetlands and control seawater intrusion (Marcé et al., 2012).

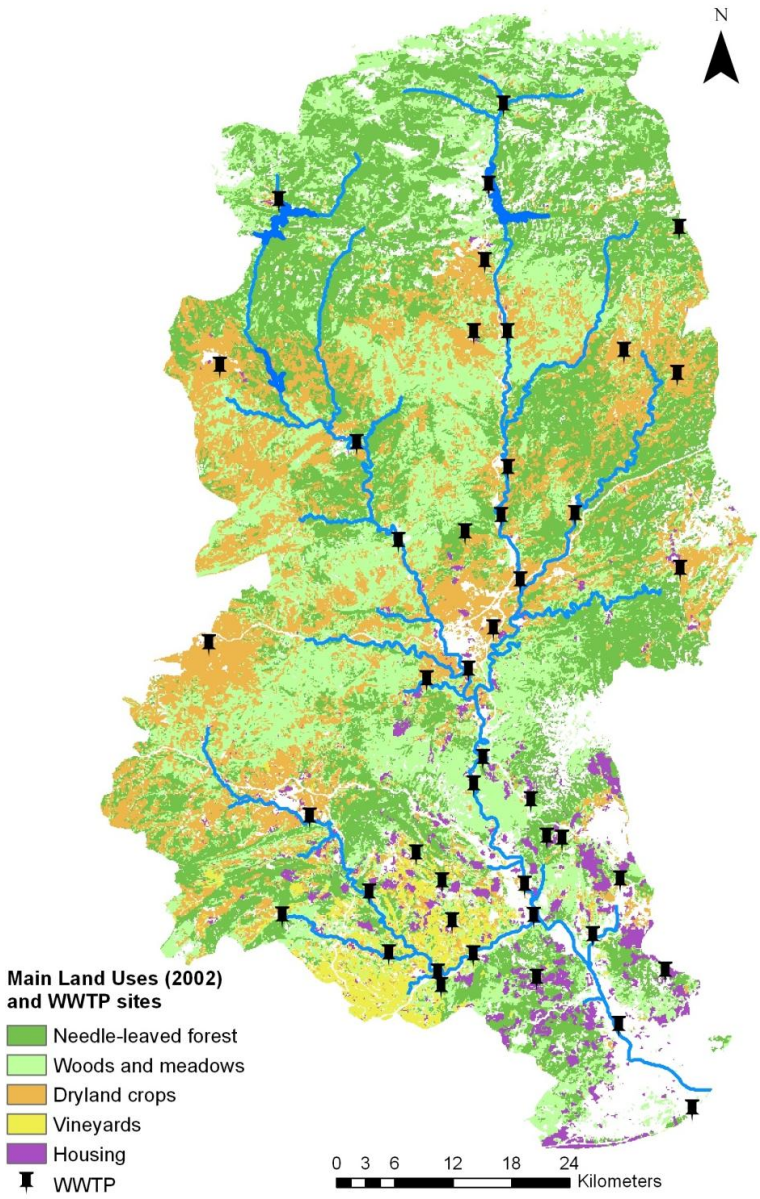


Figure 6: Main land-use types and wastewater treatment plants (WWTP) location in the Llobregat River Basin

Ebro River Basin

The Ebro River Basin covers a highly heterogeneous area of ca. 85,500 km^2, which extends from the southern-facing slopes of the Cantabrian range and Pyrenees and the northern-facing slopes of the Iberian Massif until the river reaches the Mediterranean Sea (Batalla et al., 2004). The hydrological regime of the Ebro River, which is 910 km long, is dictated in part by its contrasting tributaries, from snow-fed Pyrenean Rivers to more typical Mediterranean tributaries in the southern part of the basin. The mean annual runoff at the outlet is 13,408 hm^3.

The geography of the Ebro River determines a large range of climatic conditions (Romaní et al., 2010). For instance, mean annual precipitation varies from over 2000 mm in the Pyrenees to less than 400 mm in the arid interior. Overall, silicic materials are located in the uppermost altitudes while calcareous materials are found at lower elevations (Lassaletta et al., 2009). Despite being much larger than the Llobregat basin, the Ebro basin has a smaller population, some 2.7 million inhabitants, of which 45% live in the five biggest cities in the region.

The biogeochemical characteristics of the river water are highly influenced by anthropogenic activities. The main impacts are those related to discharge regulation (i.e., the construction of the large reservoirs) and agriculture (determining increases in nitrate concentration; Romaní et al., 2010). The intense use of water throughout the basin (Boithias et al., 2014) puts the Ebro River under strong pressure and affects its ecological functionality. The Ebro delta hosts both intensive agricultural areas where rice, fruit (in particular citric), and vegetables are harvested, as well as the Ebro Delta Natural Park, one of the most important coastal wetlands in the Western Mediterranean region.

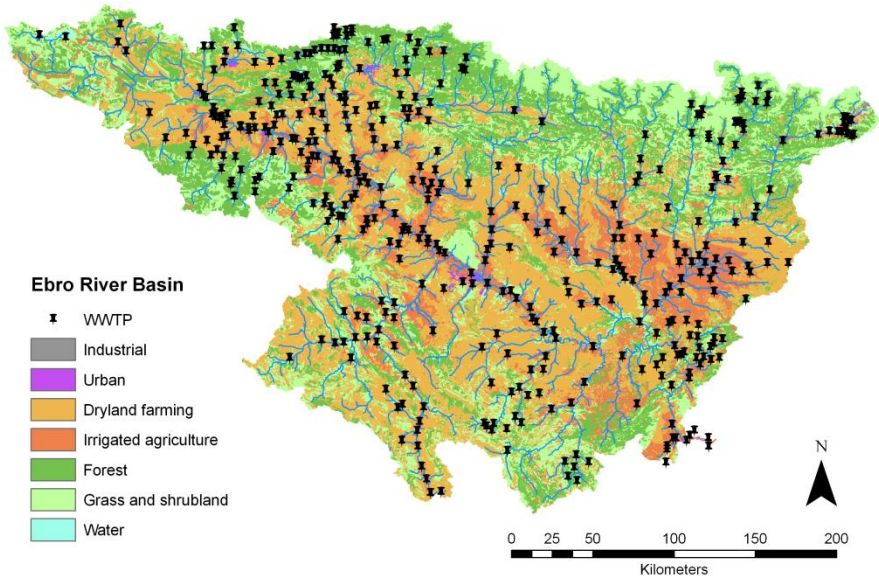


Figure 7: Main land-use types and WWTP location in the Ebro River Basin

Júcar River Basin District

The Júcar River Basin District (RBD) is located in the eastern part of the Iberian Peninsula and is formed by the aggregation of watersheds that flow into the Mediterranean Sea (total area of 43,000 km^2). The Júcar River is 498 km long and drains to the Albufera de Valencia, a large coastal lagoon and a protected wetland (Navarro-Ortega et al., 2012). The other two main rivers within the river district are the Turia and Mijares. The hydrology of the basin is typically Mediterranean, with rapid alternation between droughts and floods. Groundwater is particularly relevant in the Júcar RBD, with 90 defined groundwater bodies and with over 70% of the streamflow coming from groundwater sources. Climatic conditions are highly heterogeneous, both in space and time. The average annual rainfall is of 500 mm, varying between 320 mm in the driest years and nearly 800 mm in the wettest years. In southern regions, however, the average annual rainfall is lower than 300 mm (Ferrer et al., 2012).

The Júcar basin is vastly regulated, with a total reservoir capacity of 2,643 hm^3 (Navarro-Ortega et al., 2012). The management of the system is very complex, with many users and considerable hydrologic variability. The human population living in the basin (over one million people) makes an intensive use of the available water, the demand being often in excess of water supply. Nowadays, agriculture accounts for nearly 80% of water demand (1,394 hm^3 y^{-1} for 200,000 ha of irrigated crops), but agricultural demand appears to be stabilized or decreasing, whereas urban/industrial demand is forecasted to rise.

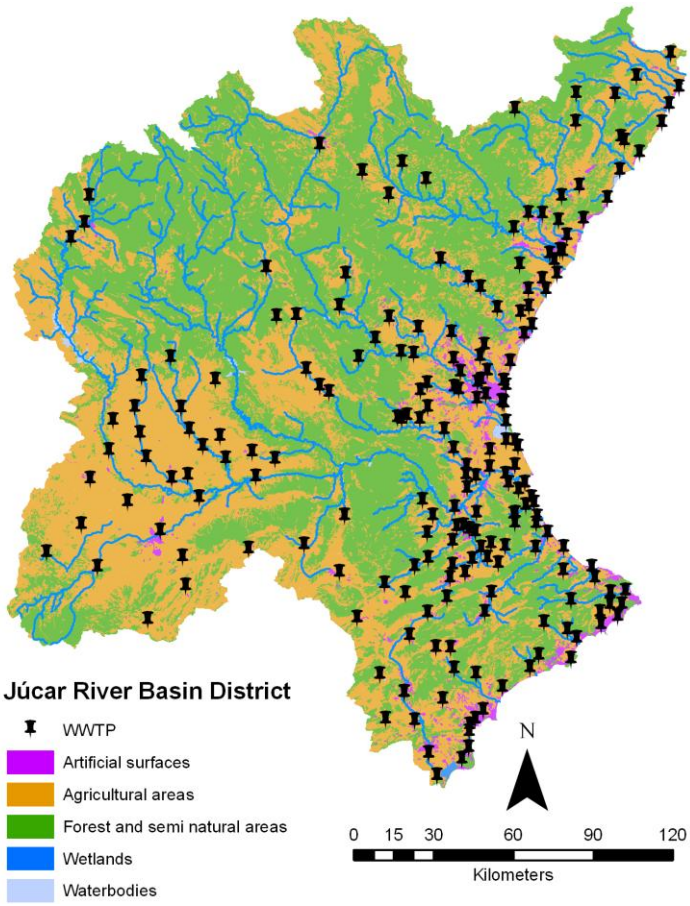


Figure 8: Main land cover and uses in the Júcar RBD

Water quality problems are common in the medium and lower parts of the basin. In the medium 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. Furthermore, groundwater overexploitation has disturbed the hydrological system, severely affecting the river flow, and causing water scarcity and salinisation in the surroundings of the river outlet (Ferrer et al., 2012).

Results

Chapter 1

Modeling nutrient retention at the watershed scale: Does small stream research apply to the whole river network?

State of the Science

Streams and rivers transport nutrients from their watersheds and thereby serve as sites for nutrient transformation, storage and removal, thus turning river networks into regulators of exported material to downstream aquatic ecosystems (Ensign et al., 2006). Understanding the coupling between terrestrial nutrient sources and the aquatic transport within a watershed is vital for the management of receiving water bodies downstream, including marine environments (Ensign & Doyle, 2006; Wollheim *et al.*, 2006). However, nutrient transport and retention processes have been mainly studied in low order streams with relatively low discharge (< 0.2 m^3s^{-1}; Doyle, 2005; Tank *et al.*, 2008). This bias towards small streams has received strong support regarding retention efficiency at the reach scale (Peterson et al., 2001), but large rivers also play a fundamental role on river export downstream in terms of total load at greater scales (Seitzinger *et al.*, 2002; Wollheim *et al.*, 2006). Nevertheless, such role has been largely unnoticed and even more so from the perspective of river networks (Tank et al., 2008).

Modeling tools are a valuable means for unraveling nutrient processing in large river systems. Models assist in the evaluation of nutrient cycling from the stream network perspective by upscaling the most relevant mechanisms occurring at the reach scale to the watershed scale. Although the differentiation a priori between hydrologic and non-hydrologic processes could offer a better understanding of the complex nature of nutrient dynamics at the watershed scale (Stream Solute Workshop, 1990), most modeling exercises have focused on the effects of hydrological variability on nutrient retention processes (Doyle, 2005; Wollheim *et al.*, 2006; Basu *et al.*, 2011). Hydrological conditions can be particularly relevant when considering nutrient retention at a single reach over time. However, the hypothesis that hydrological variability dominates nutrient loss processes across sites is supported neither by theory (Doyle et al., 2003) nor by empirical data. For instance, taking the most complete and homogeneous database about nitrogen retention across sites (Mulholland *et al.*, 2008), a simple variance analysis shows that the variability of the overall uptake rate coefficient (k) is mainly explained by nutrient uptake velocity (v_f) variability (77%),

which is a biological measure of nutrient uptake, while water depth (*h*, a proxy of hydrology) accounted for the remaining 23%. Even within biomes, the variability of k is most frequently explained by v_f (ranging between 35% and 90%), a finding that is also supported by Doyle et al. (2003). Therefore, when modeling in-stream processes, it is important to consider those factors that are germane to biological uptake instead of relying only on hydrology-related variables such as channel depth or discharge.

Data gathered within the framework of the nutrient spiraling metrics (Stream Solute Workshop, 1990) have shown that the efficiency of process rates relative to increasing nutrient concentration eventually declines. O'Brien et al. (2007) reported that the loss in nitrogen processing efficiency could be described by an Efficiency Loss (EL) model, which emphasizes the link between decreasing biological nitrogen uptake with increasing NO_3^- concentration. Subsequent studies also supported the EL concept by identifying a partial saturation effect in nutrient uptake (Mulholland *et al.*, 2008; Mulholland *et al.*, 2009; Hall *et al.*, 2009). Using modeling techniques at the watershed scale, Alexander et al. (2009) and Marcé & Armengol (2009) found that EL dynamics applied for nitrate and phosphate, concluding that streams of all sizes can be negatively affected in terms of their nutrient retention capacity. Such observations of EL dynamics contrast with other studies that indicate that nutrient uptake completely saturates at a threshold concentration or follows first order kinetics instead (O'Brien & Dodds, 2010).

On the other hand, Bernot et al. (2006) showed that biological uptake of nitrate saturated with higher concentrations, while the opposite occurred with soluble reactive phosphorus (SRP), indicating a potential saturation of process rates. A similar response was observed by Newbold et al. (2006) in streams that were, in most cases, under the influence of urbanization. Overall, a particular river subjected to periodic nutrient inputs may behave as the theoretical curves for first order kinetics and Michaelis-Menten saturation models, but the response of streams receiving continuous loading is best described by the EL concept (O'Brien et al., 2007). Several uptake kinetics could therefore occur, and it remains difficult to anticipate which will be the

dominant dynamics in a particular system, especially when considering large spatial and temporal scales.

Nowadays the status of many watersheds is largely altered, mainly in the downstream areas where human-driven pressures produce changes in chemistry, geomorphology, hydrology, and biota (Bukaveckas, 2007). This has implications for nutrient processing and its modeling at large scales since the overall retention capacity of the system and its distribution across the network may vary dramatically (Wollheim et al., 2006). However, we still lack a reliable theoretical or empirical framework to upscale the processes related to biological activity (represented by v_f) across different scales in stream networks including both non-altered and impaired reaches, mainly because data from impaired, high order reaches is scarce (e.g. Martí *et al.*, 2004; Haggard *et al.*, 2005). Moreover, the EL dynamics is conceptualized as a non-truncated power law relating nutrient uptake and nutrient concentration. Therefore, if in-stream nutrient uptake models consider a partial saturation effect under chronic nutrient inputs, a highly unlikely infinite retention capacity would occur under strong chronic nutrient inputs (Bernot & Dodds, 2005).

This study presents a heuristic approach to model nutrient retention in a watershed with major anthropogenic stress exerted along the stream network. Using a statistical-mechanistic modeling tool, we aimed to characterize nutrient retention across a watershed that includes impaired reaches ranging from first to fourth stream order. The main objective of this chapter was thus to describe biogeochemical processes involving nutrients in impaired rivers by leaning towards a more integrative approach of biological and hydrological factors acting on in-stream retention at the river basin scale.

Methods

Model general description

The selected modeling approach, SPARROW (SPAtially Referenced Regression On Watershed attributes), is a hybrid mechanistic-statistical technique that estimates pollutant sources and describes contaminant transport within a detailed network of stream reaches and their corresponding sub-watersheds. Available monitoring points and watershed characteristics are spatially referenced to corresponding reaches in the drainage network. Flux can be expressed as (Schwarz et al., 2006):

$$L_i = \underbrace{[\Sigma_{j\in J(i)} L'_j]\, F(Z_i^A;\theta_A)}_{\textbf{A: Upstream Load}} + \overbrace{[\sum_{n=1}^{Ns} S_{n,i}\, \alpha_n\, D_n\, (Z_i^D;\theta_D)]\; F'(Z_i^A;\theta_A)}^{\textbf{B: Sub-watershed Load}} \qquad (1)$$

The nutrient load L_i at the downstream end of a given reach is defined as the sum of all nutrient loads from upstream reaches (L'j, representing either a measure available at the calibration points or the model-estimated flux in kg yr-^{1}; term A in Eq. 1) plus the load originated within the sub-watershed of each river reach (term B in Eq. 1). In both terms, the load entering the reach is subjected to attenuation processes and the fraction remained in transport (F and F'; unitless) is expressed as a function of variables Z_i^A and parameters θ_A describing aquatic loss in streams and reservoirs. The summation term in B represents the diffuse and point nutrient source variables (S_n) and the source-specific coefficients (α_n). Diffuse sources are regulated by a land-to-water delivery factor D_n (unitless), estimated as a function of delivery variables (Z_i^D) and parameters θ_D. The sections below include a detailed description of these functions in our case. A comprehensive description of SPARROW can be found elsewhere (Schwarz et al., 2006).

Study Area

We applied the SPARROW model in the Llobregat River Basin (NE Spain), which has a drainage area of 4.948 km^2 and is characterized by a calcareous substratum and Mediterranean climate (Sabater *et al.*, 1987). The Llobregat River has its origin at the Pre-Pyrenean Mountains and ends in the Mediterranean Sea, near the city of Barcelona. The hydrological regime of the river, characterized by low basal flows (5-14 m^3s^{-1}), extremely high peak events, and a wide range of monthly discharge values (<2 to 130 m^3s^{-1}), is typically Mediterranean (Muñoz *et al.*, 2009; Sabater & Tockner, 2010). The most important tributaries to the Llobregat River are the Cardener and Anoia rivers, the two being highly polluted. Urbanized and industrialized clusters are predominantly located in the middle and lower parts of the watershed. A lack of dilution of point sources occurs after the reception of urban and industrial wastewaters (Prat & Rieradevall, 2006; Sabater & Tockner, 2010). Forested lands are mostly associated to upper headwater reaches (Marcé et al., 2012) and agricultural lands spread mostly throughout the middle and lower sections.

The drainage network and associated sub-watersheds were delineated in a GIS platform (Miramon v6.4p) based on a 100 m Digital Elevation Model (DEM). A total of 79 reaches and sub-watersheds were established (Fig. 1). Three reservoirs located on the upper part of the basin were included in SPARROW as reservoir reaches. Mean annual river discharge was estimated by means of the Drainage-Area Ratio Method (Emerson et al., 2005), based on daily measurements from 17 gauging stations in the Llobregat River Basin (Fig. 1) supplied by the Catalan Water Agency (ACA) and the area for each of the 79 sub-watersheds. A value for each of the 7 years within the study period (2000-2006) was assigned to each reach.

Nutrient loads

Nitrate and phosphate concentration data were obtained from locations monitored by the ACA (Fig. 1). The average nutrient concentration in the 23 sampling stations within the period 2000-2006 were 11.2 (± 9.1) mg L^{-1} NO_3^- and 0.90 (± 1.4) mg L^{-1} PO_4^{3-}. The water quality data (monthly resolution in most stations) in combination

with corresponding daily discharge data were used to calculate nitrate and phosphate daily loads by means of the software Load Estimator (LOADEST; Runkel et al., 2004). Regression-based rating curves such as those included in LOADEST have been applied in previous SPARROW modeling exercises to estimate nutrient load at monitoring stations (e.g., Smith et al., 1997; Alexander et al., 2002a; Wellen et al., 2012). Within LOADEST, the model to estimate loads was set to be automatically selected from models 1-9 (Table A1, Appendix A). To select the best models, LOADEST calculates model coefficients for several predefined regression models using each calibration dataset, and models with the lowest Akaike Information Criterion values are selected for load estimations. The models used to estimate nitrate and phosphate daily loads for each station, along with some performance measures, are shown in Table A2 of the Appendix A.

Most of the selected models for nitrate and phosphate load estimation considered nonlinear patterns within the options presented in LOADEST, and different performance measures and inspection of model residuals normality and lack of significant serial correlation guaranteed the use of these results to calibrate our SPARROW models (Table A2 in Appendix A). Moreover, our principal focus was to use SPARROW as a heuristic platform to challenge different hypotheses about river functioning, not to have a model performing perfectly at every sampling site. However, it is worth mentioning that there were some stations with large errors associated to the load estimation. Removing these stations from the SPARROW models did not improve the performance of the models, and the SPARROW parameters did not vary appreciably. Therefore, we decided to keep these stations in our models.

Daily loading information for every station was averaged for each of the 7 years considered in this study and then multiplied by 365 to obtain total annual loads. The calculated 161 annual loads (23 sampling locations times 7 years) were used as the dependent variable in Eq. 1 (in kg yr^{-1}), considering separate models for nitrate and phosphate.

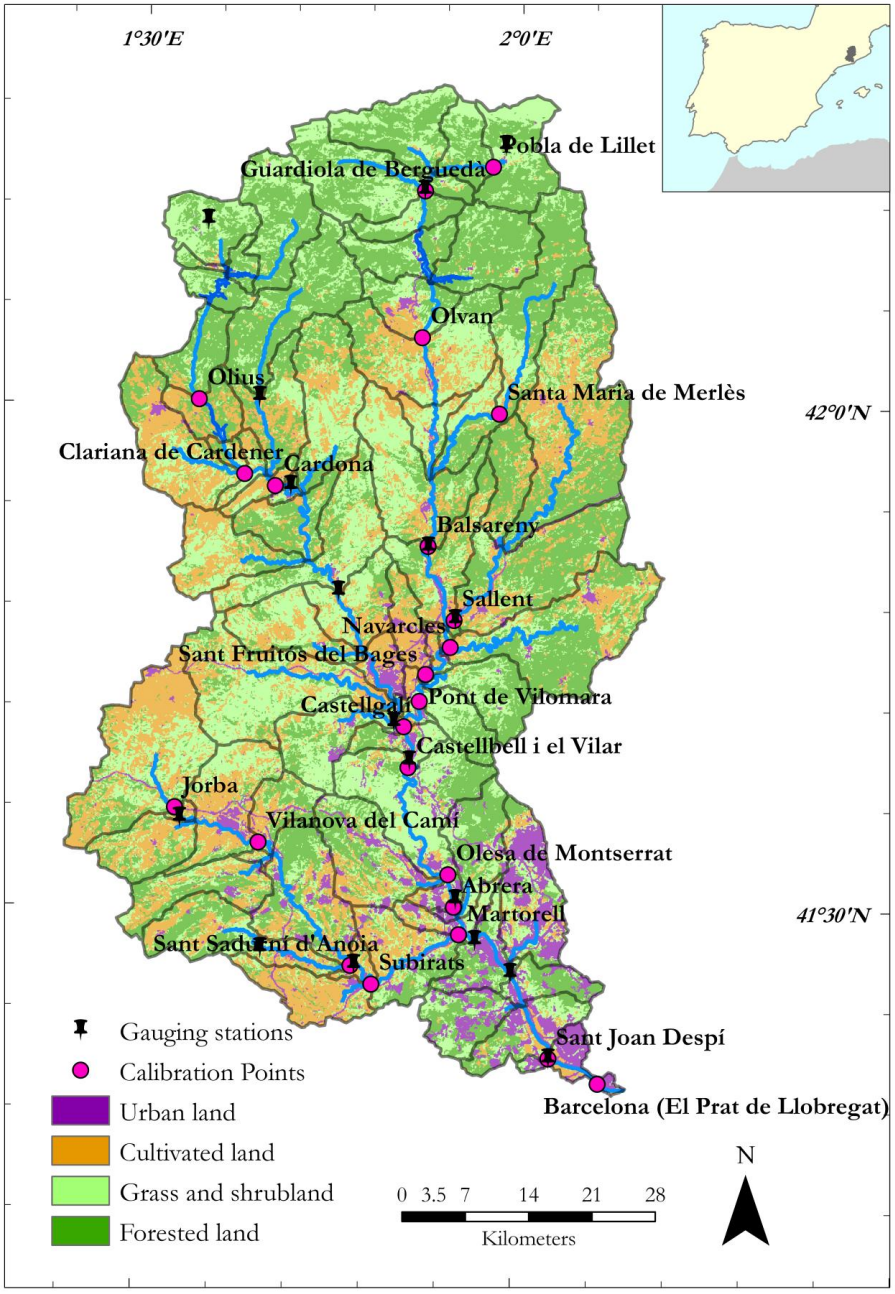


Figure 1: Major land uses types (obtained by grouping land uses and covers into 4 main categories) in the Llobregat River Basin within the delineated sub-watersheds and their corresponding reaches. Monitoring points and gauging stations are depicted for reference.

Nutrient sources

Nutrient sources (S in Eq. 1) were represented by four different land use types and point sources. Land use definitions were based on data captured by the Thematic Mapper sensor of the Landsat satellite during year 2002 (30m grid resolution, http://dmah.nexusgeografics.com). The four types were urban land (6% of the total basin area), cultivated land (24%), forested land (38%), and grass and shrub land (32%; Fig. 1). These diffuse sources were represented in the SPARROW models as S_{URBAN}, $S_{CULTIVATED}$, S_{FOREST} and S_{GRASS} (in km^2) with their associated α coefficients (Eq. 1, kg km^{-2} yr^{-1}).

The land-to-water delivery factor considered in the model (Z^D in Eq. 1) was average runoff (m yr^{-1}), along with the corresponding coefficient θ_D. Z^D was estimated as the area-weighted runoff produced in each sub-watershed. Since runoff may reflect the effect of climate, soil properties, lateral flow paths, and water storage on nutrient export from the land phase (Alexander *et al.*, 2002b), it was used as a proxy to describe the main characteristics involved in the terrestrial transport of nutrients. Indeed, mean runoff values were correlated to annual precipitation and temperature values, as well as to terrain slope values for each of the sub-watersheds. In addition, the use of runoff allowed us to include in our analysis the climatic inter-annual variability among the 7 years included in the study period.

Point source loads (S_{POINT}, in kg yr^{-1}) were obtained from waste water and industrial effluents data (ACA) and incorporated into our nitrate and phosphate models. These nutrients were not subjected to land-to-water delivery transformations. However, a coefficient to correct potential monotonic biases in the database is included in SPARROW (α for S_{POINT} following terminology in Eq. 1). In our case, two different coefficients were considered: one associated to point source loads upstream the Abrera monitoring station ($S_{POINT\_UPSTREAM}$; Fig. 1) and a second parameter associated to the point sources downstream ($S_{POINT\_DOWNSTREAM}$). The rationale for this distinction is that the final fate of several industrial effluents downstream Abrera is not accurately described in the available information (i.e., we were not sure the effluents actually

arrived to the watercourses). Therefore, the point-source coefficient for the downstream section could take values significantly below one.

In-stream retention processes

Models that simulate nutrient transport and retention at the watershed level range from highly detailed deterministic approaches to less complex source-transport models that are based on statistical methods and empirically-derived functions (Boyer *et al.*, 2006a). Regardless, both types of models usually rely on reaction rate expressions to describe nutrient loss in terrestrial and aquatic ecosystems, though mass flux rate formulations are also used. Most of the rate expressions implemented in modeling assume first-order kinetics in the behavior of nutrient uptake, which means that the rate of nutrient loss is proportional to the load (or concentration) of the constituent being modeled (Alexander *et al.*, 2000; Wollheim *et al.*, 2006). The proportion of nutrient removed in a given time step (F in Eq. 1) has been expressed on a volumetric basis (depth-dependent) as an exponential function of the rate k (θ_A in Eq. 1) and the time-of-travel τ (Z^A in Eq. 1) of the solute in a given reach:

$$F = 1 - e^{(-k\tau)} \tag{2}$$

Alternatively, the mass-transfer coefficient, v_f, could be used to describe the nutrient migration to the streambed sediment. This depth-independent measure quantifies the vertical velocity of the solute expressed as an exponential function related to the ratio of the water residence time and the mean depth (d):

$$F = 1 - e^{\left(-v_f \times \frac{\tau}{d}\right)} = 1 - e^{\left(-\frac{v_f}{H_L}\right)} \tag{3}$$

where the inverse of the hydraulic load (H_L) is the ratio between water depth and time-of-travel (water residence time) in a given reach. Average time-of-travel was calculated by dividing channel length by mean velocity values for each reach. The velocity values were previously obtained based on mean annual discharge and river channel sections provided by ACA.

Previous SPARROW applications have implemented a first-order decay equation to quantify the proportion of constituent being removed in a particular reach. This implies that nutrient areal uptake rates would increase linearly with nutrient concentration, with no signs of potential saturation. Since the nutrient concentration in streams determines the uptake rate, which could eventually reach partial saturation, we modified the SPARROW in-stream nutrient decay formulation to include a wide range of stream uptake dynamics, including first-order kinetics and EL.

The new model specification was based on an empirically derived power law related to available nutrient concentration. The uptake velocity (v_f) was calculated as follows:

$$v_f = a \times C^b \tag{4}$$

where *a* and *b* are parameters of the power law and C is the nutrient concentration in the reach. In this way, in Eq. 3, the variability in biological factors affecting nutrient retention are represented by v_f, as defined in Eq 4, while H_L reflects the hydrological conditions of the reach being evaluated. In this formulation, the value of *b* is of vital importance, since it defines the nature of the nutrient uptake dynamics. A first-order model will correspond to $b=0$, while an EL dynamics arise when $0>b>-1$ (O'Brien *et al.*, 2007; Hall *et al.*, 2009).

Given that the power law parameters were set to be boundless during calibration, any nutrient retention dynamics could have potentially arisen. However, some highly non-linear dynamics, specifically the Michaelis-Menten function, were not included in this framework. We purposely omitted such function to work with a power law that offered us an elegant and parsimonious procedure to include a wide range of dynamics. Nonetheless, it has been repeatedly found that it is very difficult to differentiate between Michaelis–Menten and EL dynamics based on the examination of empirical data at the reach scale (Wollheim *et al.*, 2008a; O'Brien & Dodds, 2010). Finally, the areal uptake rate (U, mg day^{-1} m^{-2}) was calculated using the model-derived

uptake velocity (v_f, m day^{-1}) and the measured annual mean nutrient concentration (C, mg l^{-1}) at each reach:

$$U = v_f \times C \quad (5)$$

For the three reaches considered as reservoirs, the fraction of the nutrient mass transported through the reservoir segment downstream was estimated as a function of the reciprocal of the annual areal hydraulic load and a fitted apparent settling velocity coefficient (v_R, in m yr^{-1}). This formulation, adapted from Reckhow & Chapra (1983), is the default option in SPARROW. However, in this Chapter we focus analyses on nutrient processing in stream reaches.

SPARROW model calibration

Parameters associated with nutrient sources and decay terms in Eq. 1 were estimated through the SPARROW automated capabilities, which include a gradient algorithm for nonlinear weighted least squares (NWLS). Additionally, and also as implemented in SPARROW, NWLS was applied in a bootstrapping framework to infer significance levels and uncertainty of fitted parameters, represented as non-linear confidence intervals (see Schwarz et al., 2006 for a comprehensive description of both calculations). Nitrate and phosphate models were fitted separately using the 161 annual estimated log-transformed loads available in each nutrient model.

Bibliographical data

Values for uptake velocity and uptake rates (U) with their corresponding nitrate and phosphate concentrations across a wide range of systems, were obtained from publications on compiled nutrient spiraling data (Ensign & Doyle, 2006, Mulholland *et al.*, 2008). Phosphorus spiraling metrics were also obtained from a set of data collected by Marcé & Armengol (2009; Section A1 in Appendix 1 includes a complete list of references). Our aim was to compare the relationship between uptake metrics (v_f and U) and nutrient concentration showed by the bibliographical data with the relationship fitted in our SPARROW model. Since the nutrient spiraling metrics fitted

in SPARROW correspond to net annual retention, we tried to confine our literature search to values derived from isotopic tracer techniques and mass balance approaches that could account for a reasonable representation of net retention. We purposely omitted all nutrient spiraling data derived from short-term addition studies, since they most likely reflect gross estimates of nutrient retention.

Assessment of model specifications

Since the construction of a SPARROW model asks for a number of decisions and inclusion of data that may prove inappropriate, we explored the specifications used in our models to ensure that they would guarantee that the results were unique to the sample data and representative of in-stream nutrient removal processes in the Llobregat basin.

As mentioned by Qian et al. (2005), a mathematical model is merely an approximation and therefore cannot include all relevant processes that affect the system under evaluation. These excluded parameters tend to be spatially correlated and so are the residuals of the model, compromising the meaning of the fitted parameters. Previous work by Qian et al. (2005) and Wellen et al. (2012) have included a regional random effect in each sub-watershed of the SPARROW model network to account for the error introduced by the absence of possibly relevant explanatory variables, spatially variable processes, or caused by systematic error in load calculations. For this purpose, we performed supplementary runs of our SPARROW models with a random error term, consisting of additional point sources, to detect whether fitted nutrient retention metrics were sensitive to spatially clustered errors. A point source for every sub-watershed with field data was defined in such a way that it affected the given sub-watershed and all neighboring ones, while the value for every point source was adjusted during calibration (a single value for the seven years simulated at every sub-basin). In this way we defined a procedure similar to the autoregressive modeling approach (CAR) by Qian et al. (2005), but without leaving the SAS environment where the public SPARROW release is programmed.

Furthermore, the nitrate and phosphate models were run with a runoff coefficient of 0 (i.e. delivered fraction of 1 for all sub-watersheds, thus withdrawing the land-to-water delivery load modulation) to test the effect of the runoff variable on the in-stream decay parameters, namely *b*, the main focus of this Chapter.

Finally, we performed a series of trial simulations where parameters for both upstream and downstream point sources were set to arbitrary values between 0 and 1 to assess the robustness of the values for stream decay parameters in Eq. 4.

Evaluation of temporal scale effects on model parameters

Results obtained with SPARROW (which apply for annual periods) and those from the literature (spanning relatively shorter periods) cannot be compared without thorough consideration of the potential confounding effects arising from the differing temporal scale that applies in each case. In order to discard that such differences were a reflection of different temporal scales, we used the functionally equivalent discharge (Q_{fed}), a new metric based on the concept of effective discharge used in sediment transport studies (Doyle, 2005). Considering a given relationship between streamflow and nutrient retention, Q_{fed} is the single discharge that reflects the magnitude of the nutrient retention generated by the full spectrum of discharge during a period of time (e.g., yearly). Doyle (2005) showed that as discharge variability increases, Q_{fed} becomes increasingly different from typical measures of the modal discharge (e.g., mean or median). In our case, if the mean annual flow is far from the annual Q_{fed}, calibration of the SPARROW model may accommodate such a difference by adjusting the retention metric parameters to biased values, which may confound the comparison with retention metrics obtained in short-term field experiments performed under stable flow conditions.

However, in our case, hydrological variability was not the sole factor considered to affect in-stream nutrient loss. Thus, we need not only to consider Q_{fed} (using H_L as an appropriate proxy), but also to account for a functionally equivalent v_f, defining a

more complex two-dimensional problem (Eq. 3). For the sake of mathematical simplicity, we reduced the problem to one dimension searching for potential effects of temporal averaging using the functionally equivalent v_f/H_L ratio ($[v_f/H_L]^{fe}$ hereafter). That is, we compared $[v_f/H_L]^{fe}$ with the annual mean v_f/H_L ratio as used in SPARROW in our study basin.

To determine to which extent temporal averaging could affect the values of the retention metrics fitted in SPARROW, it must be understood that the calibration process seeks to fit the modeled retention (Eq. 3) with the observed retention by modifying the power law in Eq. 4 (since H_L is an observational variable not subjected to calibration). In other words, the model attempts the following:

$$a\,(C^m)^b / H_L^m = [v_f / H_L]^{fe} \tag{6}$$

being H_L^m the mean annual hydraulic load and C^m the mean annual concentration as calculated in our SPARROW models. This relationship can be rearranged as $v_f'/H_L^m = [v_f/H_L]^{fe}$; arithmetically, we have that $v_f' = [v_f/H_L]^{fe} \times H_L^m$, where v_f' is the "temporal-averaged uptake velocity" calibrated by the model. Note that if H_L^m equals the functionally equivalent HL, then v_f' equals the functionally equivalent v_f, and no deleterious effect of temporal averaging on the retention metrics is expected during calibration.

Since we did not have a daily series of measured v_f/H_L ratio to calculate $[v_f/H_L]^{fe}$ and a mean v_f/H_L ratio, we built a synthetic 7-year daily trace of the v_f/H_L ratio based on the measured daily streamflow, the geomorphological data, and the daily nutrient concentration data generated by LOADEST at every sampling station, assuming that v_f varies with C (Eq. 4). We duplicated this experiment using two different relationships between v_f and C, one with the slope showed by the literature data points, and another with the slope calibrated by SPARROW for the nitrate model. Once a daily v_f/H_L ratio was obtained, a reference annual retention value (F_{ref}) was calculated for every station and year, solving the integral problem given by Doyle

(2005). After obtaining F_{ref}, the corresponding $[v_f/H_L]^{fe}$ was estimated with Eq. 3. On the other hand, the annual mean vf/HL ratio was calculated following the approach considered in our SPARROW exercises; that is, calculating the mean annual hydraulic load (H_L^m) and the mean annual concentration (C^m) to estimate the mean annual uptake velocity (v_f^m) with Eq. 4. Any departure between $[v_f/H_L]^{fe}$ and v_f^m/H_L^m would indicate a potential bias in the fitted nutrient retention metrics promoted by temporal averaging.

Results

Model fit

Calibration results provided a reasonable and statistically significant fit between measured and predicted loads for both nitrate and phosphate models (Fig. 2). The nitrate model was able to reproduce observed loads, as shown by a coefficient of determination (R^2) of 0.86 (RMSE of ≈ 9×105 kg NO_3^- yr^{-1}). In the case of the phosphate model, R^2 was 0.80 (RMSE of ≈ 4×104 kg PO_4^{3-} yr^{-1}). The relationship between observed loads and those predicted by the model calibration for each nutrient was acceptable according to results found in previous SPARROW model applications (e.g. Preston *et al.*, 2011a).

Sources apportionment

The nutrient sources associated with agricultural uses played a major role in explaining nitrate loads (Table 1); the export coefficient value was 3550 kg km^{-2} yr^{-1} (annual average of 35.5 kg NO_3^- per hectare). In general, diffuse inputs largely contributed to the nitrate leaching into the stream network, although two sources (Urban and Grass land) were consistently related to a zero export coefficient. Additional model runs excluding point sources loads (not shown) suggested that at least in the case of SURBAN the zero values were promoted by strong correlation between point sources location and SURBAN spatial distribution. Diffuse sources such as agricultural and forested land remained significant in the phosphate model as well (p-values <0.001; Table 2). In the phosphate model Point_Upstream (1.60) and

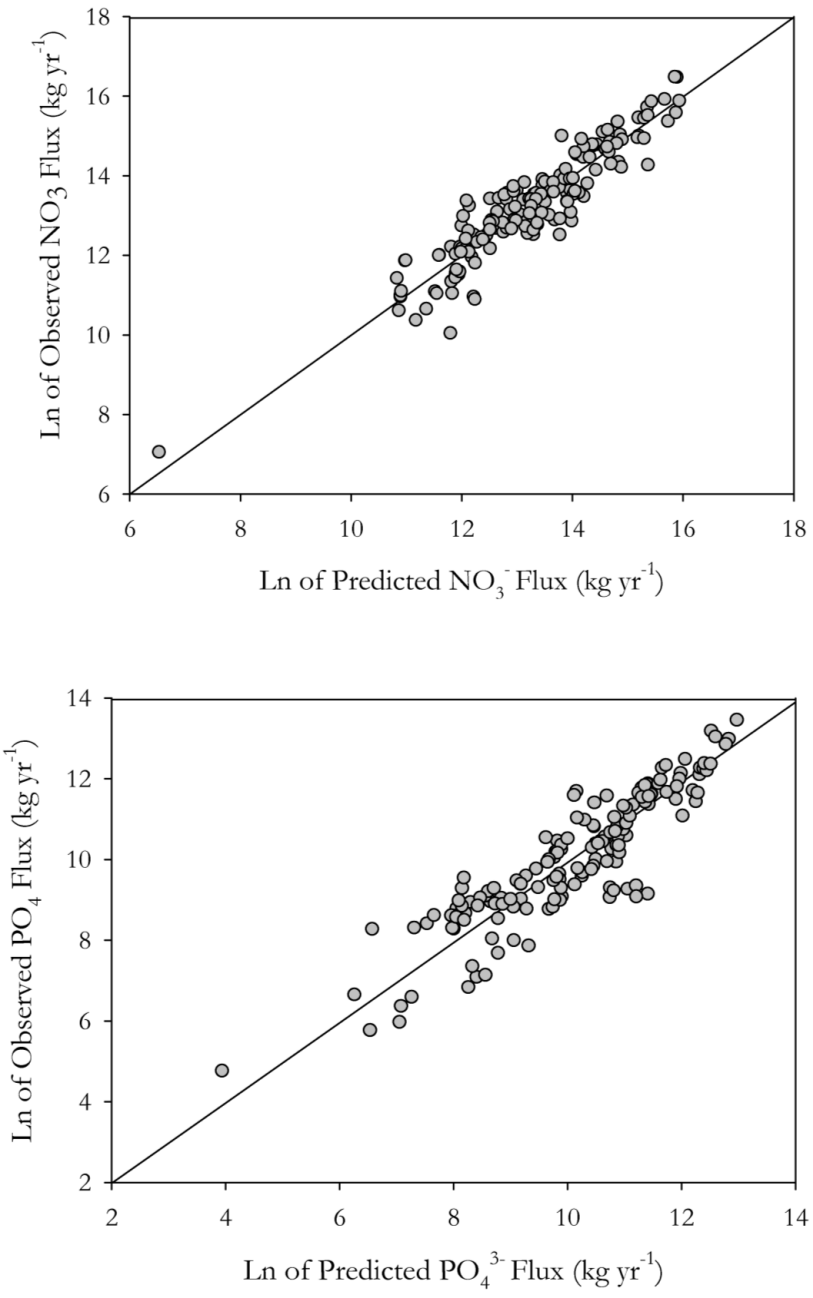


Figure 2: SPARROW-predicted relative to observed nutrient flux for 161 calibration measurements in 23 monitoring stations in the Llobregat River between 2000 and 2006 (Top, Nitrates; Bottom, Phosphates).

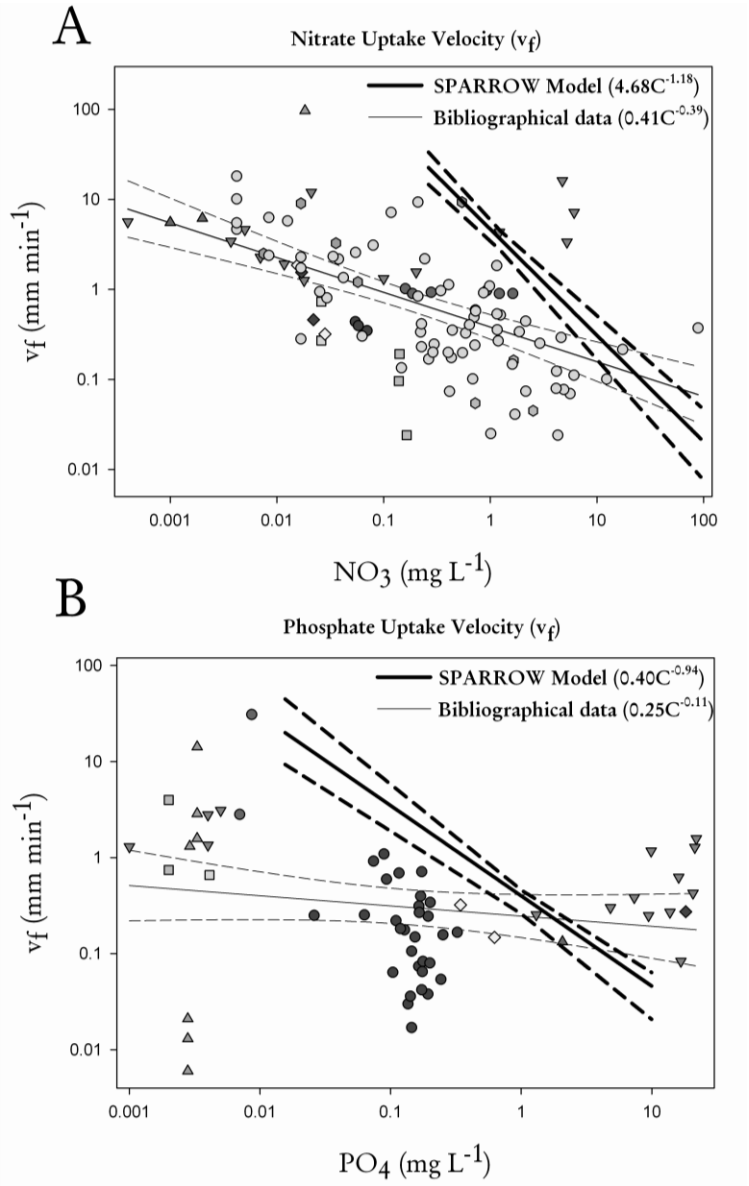


Figure 3: Relationship between uptake velocity (vf) and nutrient concentration in literature studies (different symbols) and estimated in SPARROW for the Llobregat River Basin. (A, nitrates, B, phosphates). The correspondence between symbols for literature data and their reference can be found in the Appendix A. The power laws for literature data and the SPARROW results along with 95% confidence intervals are included.

Point_Downstream (0.80) coefficients were respectively ca. two-fold and five-fold higher than the values obtained in the nitrate model, indicating potential non-homogeneous biases in the point sources inventories.

The point source coefficients assigned to the upstream part of the basin were higher than that of the downstream area. Point_Downstream coefficient was particularly low in the nitrate model (in fact, non-significantly different from zero, Table 1). In general, the value of these coefficients suggested a potential overestimation of point loads, except for phosphate in the upstream section, which seemed to be underestimated.

Table 1: Nitrate Model Parameters for the Llobregat River Basin. Results for NLWS and bootstrap fitting are included.

Model Parameter	Parametric Coefficient	Parametric *p*-value	Bootstrap Coefficient	Bootstrap *p*-value	Lower 90% CI	Higher 90% CI	Units
Sources							
α for S_{URBAN}	**0**	-	**0**	-	0	0	kg km^{-2} yr^{-1}
α for $S_{CULTIVATED}$	**3550.59**	< 0.001	**3451.01**	<0.001	1689.25	5203.36	kg km^{-2} yr^{-1}
α for S_{GRASS}	**0**	-	**0**	-	0	0	kg km^{-2} yr^{-1}
α for S_{FOREST}	**2342.96**	< 0.001	**2212.83**	<0.001	1632.97	2683.90	kg km^{-2} yr^{-1}
α for $S_{POINT\_UPSTREAM}$	**0.69**	0.01	**0.68**	<0.001	0.37	1.26	dimensionless
α for $S_{POINT\_DOWNSTREAM}$	**0.15**	0.75	**0.15**	0.02	0.15	0.15	dimensionless
In-stream attenuation							
Reach decay intercept (*a* in Eq. 4)	**4.68**	< 0.001	**4.64**	<0.001	3.66	5.61	dimensionless
Reach decay slope (*b* in Eq. 4)	**-1.18**	< 0.001	**-1.21**	<0.001	-1.34	-1.04	dimensionless
Reservoir settling velocity	**-2.29**	0.87	**-1.35**	0.54	-20.06	10.49	m yr^{1}
Land-to-water delivery							
Delivery coefficient	**2.19**	< 0.001	**2.13**	<0.001	1.77	2.43	yr m^{-1}

Table 2: Phosphate Model Parameters for the Llobregat River Basin. Results for NLWS and bootstrap fitting are included

Model Parameter	Parametric Coefficient	Parametric *p*-value	Bootstrap Coefficient	Bootstrap *p*-value	Lower 90% CI	Higher 90% CI	Units
Sources							
α for S_{URBAN}	**0**	-	**-84.65**	0.16	-400.81	0	kg km^{-2} yr^{-1}
α for $S_{CULTIVATED}$	**409.28**	< 0.001	**395.99**	<0.001	173.56	665.82	kg km^{-2} yr^{-1}
α for S_{GRASS}	**0**	-	**-13.55**	0.18	-100.23	0	kg km^{-2} yr^{-1}
α for S_{FOREST}	**150.35**	< 0.001	**147.59**	<0.001	100.57	205.38	kg km^{-2} yr^{-1}
α for $S_{POINT\_UPSTREAM}$	**1.60**	0.0043	**1.70**	<0.001	1.41	2.40	dimensionless
α for $S_{POINT\_DOWNSTREAM}$	**0.80**	0.0140	**0.79**	<0.001	0.53	1.09	dimensionless
In-stream attenuation							
Reach decay intercept (*a* in Eq. 4)	**0.40**	< 0.001	**0.36**	0.02	0.26	0.46	dimensionless
Reach decay slope (*b* in Eq. 4)	**-0.94**	< 0.001	**-0.98**	<0.001	-1.10	-0.86	dimensionless
Reservoir settling velocity	**-30.78**	< 0.001	**-41.30**	<0.001	-55.25	-28.56	m yr^{-1}
Land-to-water delivery							
Delivery coefficient	**2.71**	< 0.001	**2.81**	<0.001	2.13	3.48	yr m^{-1}

In-stream retention

In-stream decay coefficient *b* was statistically significant in both models (Tables 1 and 2). The empirically driven power law relationship that estimated in-stream decay in SPARROW showed a decline in uptake velocity with increasing concentration (Fig. 3). The decreasing values of the uptake velocity along the concentration gradient confirmed a loss in uptake efficiency in streams of all sizes in the Llobregat River Basin. However, the slopes were much steeper (Fig. 3) than those corresponding to the relationship described by the bibliographical data. This result was consistent in both nitrate and phosphate models, and was confirmed by ANCOVA analysis test for differences in slope values ($p<0.001$). The curves using the bibliographical data followed the expected power law function with negative exponent (Fig. 3) and a value between zero and -1 (*b*=-0.39 for nitrate and -0.11 for phosphate) attributable to an EL

dynamics. However, while the slopes were different, the v_f values derived by our results in SPARROW fitted within the range of the literature data.

The areal uptake rate curves reported a deviation from the expected relationship (increasing nutrient uptake with increasing nutrient concentration, Fig. 4). U slightly decreased with increasing concentration in the nitrate model while the curves for phosphate U plotted against increasing concentration showed a slope close to zero (Fig. 4).

Effect of model specifications on in-stream decay

The inclusion of a spatial random error term in both nitrate and phosphate model successfully accounted for the spatial errors included in the basic SPARROW models. Although no apparent spatial clustering (i.e. downstream or upstream accumulation of errors) of model residuals was observed, some stations consistently over or under-predicted nutrient loads (Fig. A1 in Appendix A). The inclusion of the spatial error term in the models centered errors across stations on zero, and explained variance of the models (R^2 = 0.93 for nitrate and 0.95 for phosphate) and statistical significance greatly improved (overall regression $p<0.001$), despite the inclusion of many more parameters in the calibration process. The models including the error term did not yield different in-stream coefficient values for the slope b (-1.09 ± 0.08 for nitrate and -0.90 ± 0.05 for phosphate, $p < 0.0001$ in both cases, compare with values in Table 1 and 2). Therefore, we can consider our fitted nutrient retention metrics free of biases arising from model errors associated to the absence of relevant explanatory variables, spatially variable processes, or systematic error in load calculations. Interestingly, while nutrient retention metrics did not show significant differences between models with and without spatial error terms, values for parameters related to nutrient sources from land uses showed major changes both for nitrate and phosphate models. It seems that the source terms were more sensitive to model errors than the in-stream processes.

Concerning the original error structure (the models fitted without spatial error term), the points that deviate the most from the 0 line in the phosphate model belong to the Cardona monitoring station load predictions, where the SE % error of LOADEST was low (0.14), whereas Sant Fruitós de Bages, Clariana de Cardener and Navarcles stations had LOADEST SE % of 30, 54 and 30, respectively. In terms of the nitrate model residuals, the points that deviate the most from the 0 line are related to the Jorba monitoring station (where the SE % error of LOADEST was high, i.e.,76), Cardona station (SE % 14), Sant Sadurní d'Anoia (SE % 56), and Martorell stations (SE % 25). Correlation between mean SPARROW model residuals and LOADEST mean standard errors for the nitrate model was 0.17 if compared to the initial model residuals and 0.12 if compared to model residuals after random error was taken into account in model estimation. For the phosphates model residuals, the correlation was equal to 0.47 before the inclusion of spatial model errors and 0.34 before that, suggesting that some degree of dependence on errors associated to LOADEST estimation still remained.

The suppression of the land-to-water delivery factor did not vary the slope values for in-stream decay considerably (detailed results are included in Table A3 in Appendix A). Finally, setting arbitrary values to point source coefficients did not change the relationships between nutrient concentration and uptake capacity (i.e., the reach decay parameters did not significantly diverge and varied only at the decimal level), even after substantial alteration of the values for point source coefficients in both upper and lower parts of the basin.

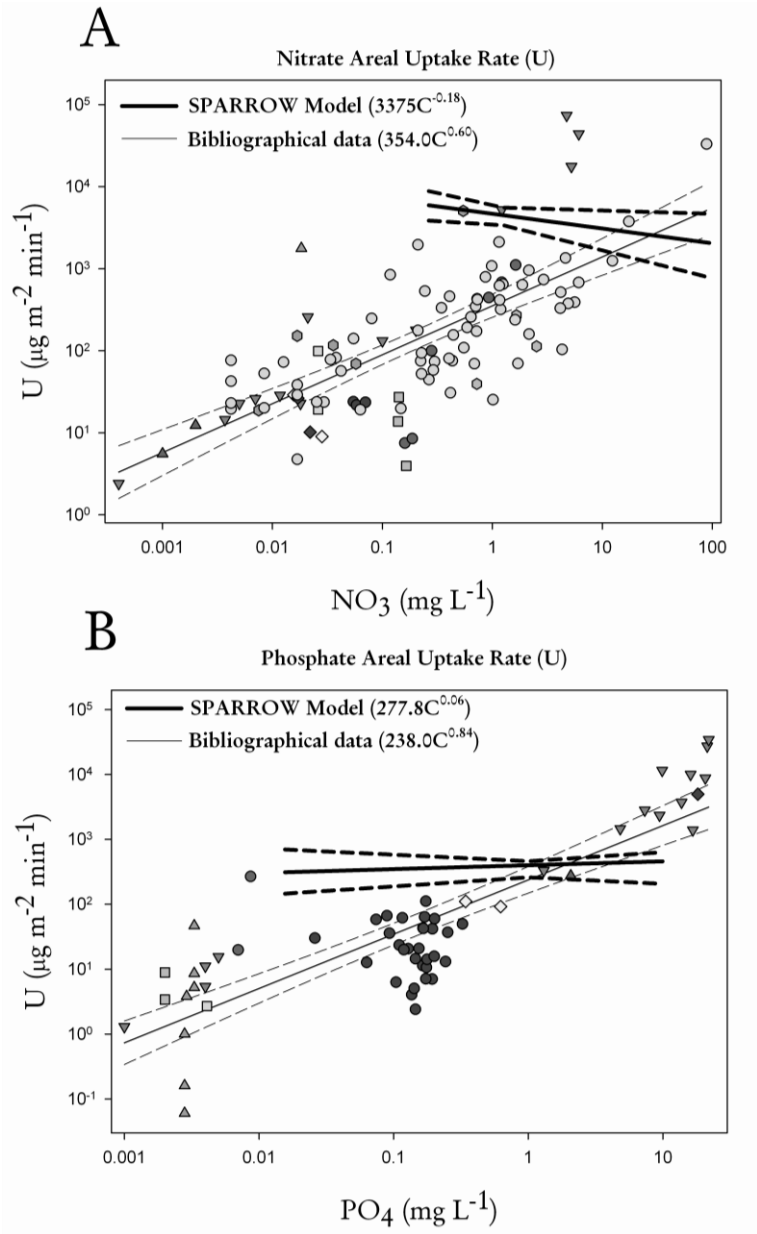


Figure 4: Relationship between uptake rate (U) and nutrient concentration in literature studies (different symbols) and estimated in SPARROW for the Llobregat River Basin. (A, nitrates, B, phosphates). The correspondence between symbols for literature data and their reference can be found in the Appendix A. The power laws for literature data and the SPARROW results along with 95% confidence intervals are included.

Effect of temporal averaging on fitted v_f

Regardless of which v_f vs. concentration relationship was used to built the synthetic v_f/H_L series, $[v_f/H_L]^{fe}$ was always larger than v_f^m/H_L^m, which suggests that temporal averaging could have indeed affected our results (Fig. 5). Confirming this, the relationship between v_f' and C resulting from the application of temporal averaging as applied in SPARROW differed from the curves used to generate the two synthetic series (i.e., the reference v_f vs. C relationships; Fig. 6). Deviations between v_f' and the reference v_f were observed for both power laws (high and low slopes). However, although the deviation is quite severe in a few cases, in general terms the resulting power laws do not differ from reference ones as substantially to explain the differences found between our SPARROW results in the Llobregat River Basin and bibliographical values (Fig. 3). Furthermore, as seen in Fig. 6, the net result of temporal averaging in our study was most likely a slight decrease in the slope of the v_f vs. C relationship, while the slope of our SPARROW relationships was larger than that of the bibliographical data (Fig. 3).

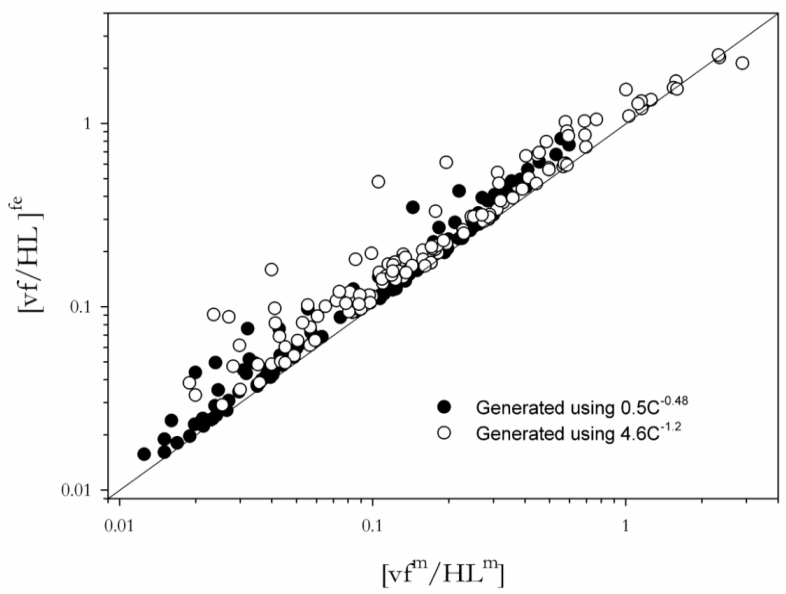


Figure 5: Comparison of Functionally Equivalent and SPARROW Modeled vf/HL ratio considering literature-based ($0.5C^{-0.48}$) and model-driven ($4.6C^{-1.2}$) power laws.

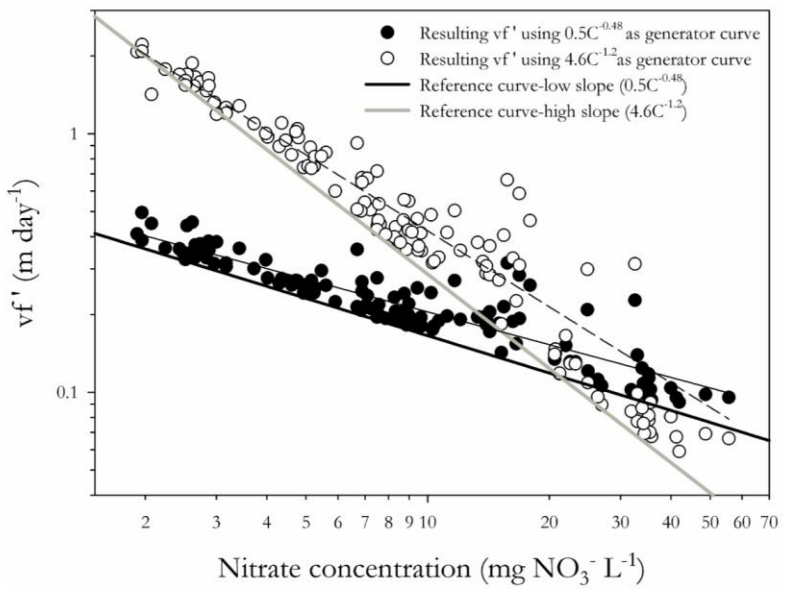


Figure 6: Resulting temporal-averaged uptake velocity (*vf'*) from literature-based (low slope) and model-driven (high slope) power laws compared to their corresponding reference curves (bilogarithmic scale).

Discussion

Our study in an impaired river network showed that the calibrated in-stream retention parameters in the SPARROW models did not fully reflect an EL dynamics of nutrient retention. It has been suggested that the EL model best describes the response across streams receiving chronic nutrient inputs (O'Brien et al., 2007), but this concept appears to be inadequate to explain in-stream retention responses for the nutrient concentration gradient and streamflow values found in the Llobregat River Basin. Instead, uptake velocity values in the Llobregat River Basin steeply decreased with increasing nutrient concentration, contrasting with other river systems. In the nitrate model, the uptake rate slightly decreased with increasing concentration (negative slope, Fig. 4), contrary to what would have been expected based on bibliographical data and the EL concept. In the case of phosphate, the behavior is at the limit of an EL dynamics, with an almost imperceptible increase of U with nutrient concentration. It is worth mentioning that some differences in behavior between bibliographical and modeled data could be attributable to the possibility of unintentionally having

included gross retention values under the reference case. Also, the environmental setting of sampling sites in our modeling exercise (large streams and rivers) differs, in most cases, among sites included in the literature values (with headwater reaches being predominant). Overall, the analysis of the ratio $[v_f / H_L]^{fe}$ allows to safely discard temporal averaging as the generator of the differences between our model results and the bibliographical data, but does not guarantee the possibility of acute effects of temporal averaging in other settings.

Streams in the network of the Llobregat River Basin are characterized by nutrient concentration and discharge values outside of the range covered by bibliographical data gathered in this study (Fig. 7). This seems to indicate that the difference in slopes might represent a genuine expression of underlying biogeochemical processes involved in nutrient retention. It is also true that the lack or weakening of EL dynamics at high nutrient concentrations can be interpreted as a possible saturation at highly polluted conditions. This perspective would be more reasonable under an ecological point of view than an ever growing U with increasing C under chronic input regime (Bernot & Dodds, 2005; Bernot *et al.*, 2006).

It is not clear whether nutrient concentration is the ultimate driver of the modeled responses in our basin. The presence of additional factors that usually covariate with nutrient concentration in impaired basins (e.g., flow alteration, river bed modifications, riparian forest degradation) can negatively affect nutrient retention capacity (Doyle *et al.*, 2003; Martí *et al.*, 2006). However, there is a dearth of knowledge about nutrient retention in large streams subjected to chronic nutrient loading and additional impairment contributors. It can be argued that there are, in fact, several responses (i.e., v_f vs. C slopes) for different types of streams, which may be generated according to climate, geology, their current level of pollution, and the geomorphologic (or possibly other) changes they have undergone. Recently, Helton et al. (2011), using data from Mulholland *et al.* (2008), fitted several power laws relating v_f for denitrification and nitrate concentration in each catchment involved in the LINX II experiment. Helton et al. (2011) confirmed the presence of EL in denitrification

with increasing nitrate concentration, but also suggested that the strength of this relationship varied significantly when the response of each of the eight catchments in the study was considered individually (i.e., nutrient concentration was correlated with v_f in some catchments, but not in others, and showed different slopes). However, we have to take into account that models were developed for only one particular form of the nutrients, nitrate and phosphate, rather than for the complete suite of nutrient forms or the total mass of nutrients. Thus, the processing and fate of the total mass of nutrients are not fully accounted for by the model in terms of the sources and sinks in the watershed. This is more probably affecting the phosphate model than the nitrate model (Ludwig et al., 2009), and should be considered to avoid over-interpreting our results.

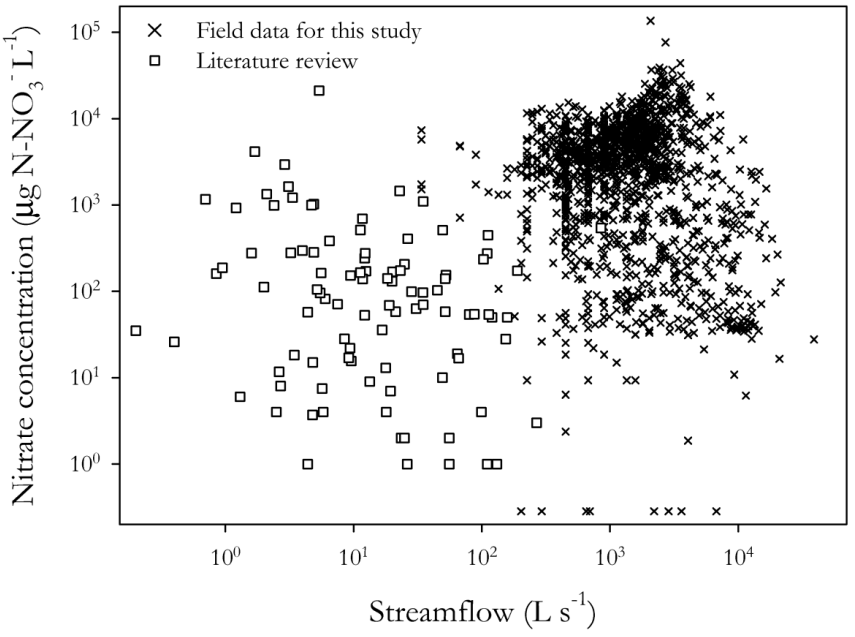


Figure 7: Nitrate concentration and streamflow data characteristics found in the literature compared to data used in the SPARROW models for the Llobregat River Basin.

These findings and our results imply that a diverse set of responses could emerge from the differences in chemical, physical, and biological conditions among streams and rivers, including the collapse of EL dynamics at some still undefined impairment threshold. It is already known that biotic communities subjected to nutrient enrichment conditions could eventually adjust their response to higher nutrient concentrations (O'Brien *et al.*, 2007; O'Brien & Dodds, 2010), which in fact is the keystone of the EL concept. The potential loss of EL dynamics at even larger nutrient concentrations suggested here might be explained by a subsidy-stress pattern (Odum *et al.*, 1979; Niyogi *et al.*, 2007), where anthropogenic activities might not harm stream diversity or function until the effects of stressors prevails over any subsidy effects (Niyogi et al., 2007).

In river networks, nutrient concentration is not the only potential driver of subsidy-stress responses. The role of biological communities and niche partitioning in net nutrient retention in streams was recently evaluated by Cardinale (2011), who found that algal diversity related to heterogeneous habitat increased nutrient uptake. Therefore, it deems reasonable to hypothesize that in rivers homogenized by human activities (for example, by means of altered flow regimes, channeling, etc.), such as in the Llobregat River Basin, niche differentiation could be lower than in less impaired systems and therefore the nutrient retention capabilities of biological communities might be diminished. Niche partitioning related to habitat heterogeneity may potentially affect other means of in-stream retention (e.g. denitrification). However, to our knowledge, this has not yet been systematized at the watershed scale, and we should seek better understanding of the factors influencing lumped concepts such as the nutrient spiraling metric v_f at larger spatial scales.

In general, assessments of the nutrient retention capacity within streams at the watershed scale have been related to hydrological and geomorphologic properties that determine the time a solute remains in contact with reactive surfaces (Doyle, 2005; Wollheim *et al.*, 2006; Basu *et al.*, 2011). But it is also obvious that biological activity will contribute to shape nutrient retention and transport to downstream water bodies.

This has profound implications in the way nutrient retention is extrapolated across scales. The assumption that nutrient retention is at least a two dimensional problem (i.e., hydrology and biological activity should be taken into account), implies that different approaches to estimate v_f would yield distinct dynamics for nutrient retention, especially when considering entire watersheds. By applying a constant v_f, the nutrient loss fraction would be exclusively dependent on the hydrological conditions (represented by H_L in Fig. 8A) of the streams under evaluation, and only situations of low hydraulic load (most probably in small headwater streams) will support high retention capacity. However, if v_f is dependent on nutrient concentration, higher hydraulic loads maintain higher nutrient retention as we move from an EL model (Fig. 8B) to the dynamics found in the Llobregat Basin (Fig. 8C). One consequence of this is that the use of hydrological-driven formulations of nutrient retention might have contributed to a biased view of the relative role of headwater streams on nutrient retention at large scales. Tables A4 and A5 in the Appendix A include examples of the consequences of changing from a first-order rate reaction to the one fitted in the Llobregat basin.

Finally, we acknowledge that some of our interpretations were based exclusively on modeling results, and that empirical confirmation through appropriate nutrient retention measurement in the field would be desirable. Nevertheless, the use of models as heuristic tools to stimulate critical thinking is a powerful mean of generating new hypotheses (Oreskes et al., 1994), and in any case models are already extensively used to upscale nutrient retention processes at large scales (Alexander *et al.*, 2000; Seitzinger *et al.*, 2002; Alexander *et al.*, 2009). One of the main problems in upscaling nutrient retention to the watershed scale is the scarcity of published empirical work related to large impaired systems, which may dominate the river landscape in many regions of the world. Considering that a small fraction of these rivers remains unaffected by humans (Vörösmarty *et al.*, 2010), detailed field work is desirable in impaired rivers with substantial flow, since most studies have been undertaken in reaches under pristine (or near-pristine) conditions, usually characterized by low flows.

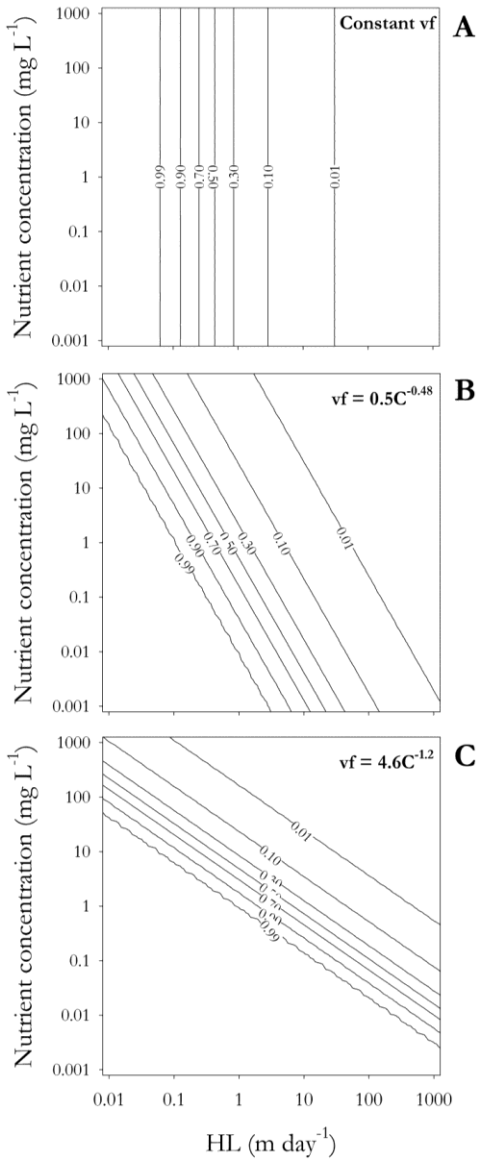


Figure 8: Analysis of the effect of hydrology (represented by HL) and biological uptake variability (represented as different power relationships between vf and C) on stream nutrient retention, calculated using Eq. 3. Three different relationships between vf and C were tested: a constant *vf* value (A), the power law fitted to bibliographical data gathered in this study (B), and the power law fitted by SPARROW for the Llobregat River Basin (C).

We further acknowledge that several challenges exist when dealing with large rivers, due mainly to elevated costs and difficulties in conducting particular analyses such as isotopic N measurements under high flow conditions. However, measurement of nutrient transport and removal in large rivers could be approached as recently shown by Tank et al. (2008), where a nutrient pulse addition method was used to estimate inorganic N uptake in a seventh order stream in Wyoming, USA. This and similar alternatives should be explored, coupled with empirical modeling efforts, in order to improve our understanding of nutrient in-stream processing at the river network scale. Results from this work suggest that researchers should carefully consider the formulations included in their empirical models, especially when working at large scales.

Chapter 2

Linking in-stream nutrient flux to land use and inter-annual hydrological variability at the watershed scale

State of the Science

Global change can be understood as the combination of land use and climate change that can alter the Earth's capacity for sustaining life (U.S. Global Change Research Act, 1990). One of the most commonly encountered problems is excess nutrient export from natural and anthropogenic sources, which are transported through the fluvial network and cause the eutrophication of downstream water bodies (Dodds & Welch, 2000). High nutrient concentrations along the stream network do affect the structure and function of the aquatic ecosystem, threatening the provision of ecosystem services to humans. Nutrient enrichment is coupled to other dramatic changes driven by human activities such as water abstraction (Oki & Kanae, 2006), dam construction (Nilsson *et al.*, 2005; Sabater & Tockner, 2010), and water pollution (Schwarzenbach et al., 2006), which define extremely complex scenarios for reliable management. Describing and understanding the present status of the world's watercourses is a necessary step in delineating management measures under possible global change scenarios.

According to a recent analysis of water quality trends in Europe (Bouraoui & Grizzetti, 2011), freshwater ecosystems in Spain have experienced a constant increase in nitrogen and phosphorus surplus within the period 1960-2000. This increase of nutrient surplus is due to the application of fertilizers in agriculture, and adds to the increasing relevance of point source emissions (Bouraoui & Grizzetti, 2011). Mediterranean rivers are typically subjected to geomorphologic alterations such as channelization and reservoir creation (Sabater & Tockner, 2010). In this context, water pollution acts as an additional stressor on the actual water supply and aggravates the damage to stream health in these regions (Gasith & Resh, 1999; Ricart et al., 2010). In the case of densely populated areas such as Catalonia (NE Spain), both agricultural and industrial activities contribute to widespread pollution of river ecosystems. Agriculture typically generates high nutrient exports, estimated to pollute up to 57% of the rivers in Catalonia (Catalan Water Agency, 2008). Point sources are

also detrimental to the water quality of streams, particularly where water flow is insufficient to achieve effluent dilution (Martí et al., 2004).

Determining the influence that human activities have on river ecosystems requires the evaluation of the nutrient cycling processes at spatial scales larger (e.g., at the watershed scale) than the reach scale (Dodds & Oakes, 2006). Suitable tools and techniques are therefore needed to identify stressors and to link the stream's response to the responsible stressor (Allan, 2004). Modelling tools can be useful in identifying nutrient sources as well as relevant nutrient cycling processes at the watershed scale (Smith & Alexander, 2000; Grizzetti *et al.*, 2005). However, a compromise between model complexity and data availability is needed to select the most appropriate modelling tool for the intended application (Paudel & Jawitz, 2012). Mechanistic models are often employed at the watershed level because they are able to supply dynamic responses at a fine temporal and spatial resolution. However, these models require enormous amount of water-resources data at a highly detailed temporal and spatial scale (Schoumans et al., 2009), which may not be readily available in medium or large-sized watersheds. In these cases, a greater number of assumptions must be made during model development, which results in higher uncertainty. On the other hand, conventional empirical models face the challenge of upscaling the relevant biogeochemical processes that affect nutrient fluxes within a basin, without guaranteeing that measurements performed at a particular scale will be valid to parameterize processes at other scales (Quinn, 2004).

The hybrid statistical and process-based model SPARROW (SPAtially Referenced Regression On Watershed attributes) spatially links watershed characteristics to the river network and uses mechanistic functions based on empirical observations to describe nutrient dynamics (Alexander *et al.*, 2002a). Models such as SPARROW can provide valuable information for the implementation of mitigation actions at the basin scale (e.g., for the Water Framework Directive Basin Management Plans) and the basis to perform heuristic exercises to study the role of pristine and impaired watercourses in the fate of nutrients. In this Chapter, we used the SPARROW

approach to estimate the annual nitrate and phosphate loads reaching the drainage network and its spatial variations in a Mediterranean watershed. Since inter-annual variability is projected to increase in the Mediterranean region (Giorgi & Lionello, 2008) we also studied the relative influence of wetter and dryer years on annual nutrient flux, in order to provide tools for the better management of water resources as well as for contributing to the delineation of sound policies in the Mediterranean region.

Study Area

The Llobregat River is a Mediterranean river that runs across NE Spain. The hydrological regime of the river is defined by low basal flows (5-14 m^3s^{-1}), extremely high peak events and a wide range of monthly discharge values (<2 to 130 m^3s^{-1}; Muñoz *et al.*, 2009; Sabater & Tockner, 2010). The river's drainage area of 4,948 km^2 is mainly characterized by a calcareous substratum that determines the high water alkalinity (Sabater *et al.*, 1987). The average rainfall in the watershed is 610 mm yr^{-1} and the mean annual temperature is around 15° C. The occurrence of natural salt formations submitted to mining activity at the upper part of the watershed increases the water salinity in the downstream sections. The Llobregat is regulated through three large reservoirs located in its upper part (La Baells in the Llobregat River, and La Llosa del Cavall and Sant Ponç in the Cardener River), as well as by several weirs scattered along the main river channel.

The Llobregat hosts more than 3 million people in its watershed, and its waters are intensively managed for drinking, agricultural, and industrial purposes (Marcé et al., 2012). The main land cover types in the watershed include forested and shrub land areas, crop fields, and impervious surfaces found in urban and industrial agglomerations. Urban and industrial nuclei are mainly located in the downstream part of the watershed, whereas forested lands are situated, for the most part, near the upper headwater reaches. Agricultural lands spread mostly throughout the middle and lower sections, particularly by the Anoia River and its tributaries (Fig. 1b). During the

last few decades, the Llobregat River Basin has been heavily polluted by industrial and urban sewage as well as by runoff from agricultural areas. Together with its two major tributaries, the Cardener and the Anoia, the river now receives input from multiple wastewater treatment effluents as well as diffuse inputs from agricultural sources.

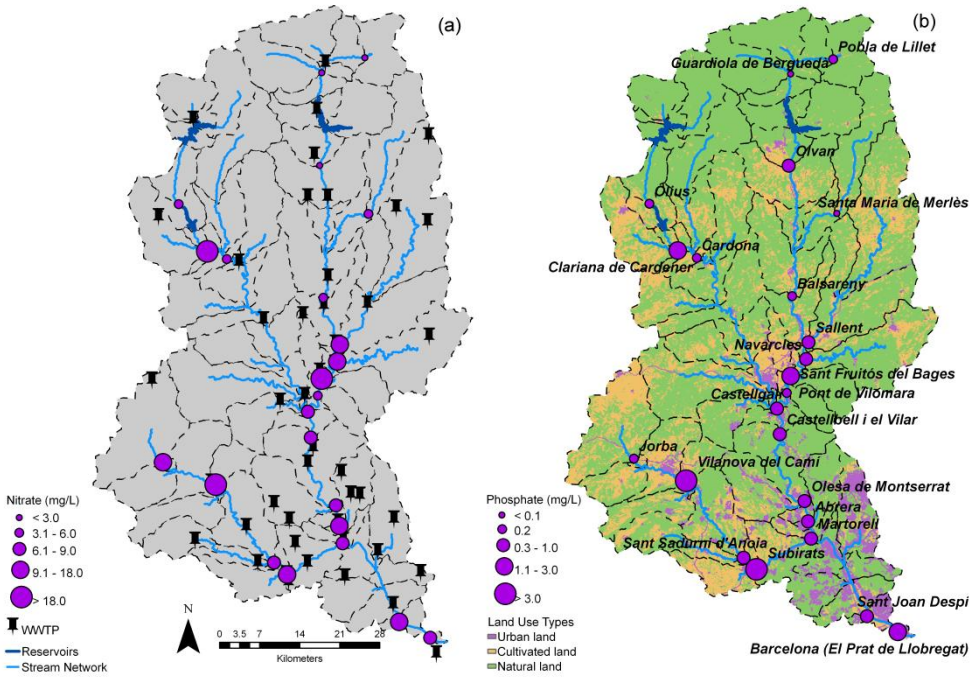


Figure 1: Average Nitrate (a) and Phosphate (b) concentration in the selected monitoring sites (2000-2006). Dashed lines represent sub-watersheds division. Distribution of land use types (b) and Wastewater Treatment Plant (WWTP) location (a) are also shown. Urban land comprises 6% of the total watershed area, cultivated land, 24%, and natural land, the remaining 70%.

Methods

Model description

The SPARROW model (SPAtially Referenced Regression On Watershed Attributes), is a hybrid mechanistic-statistical technique that estimates pollutant sources and

describes contaminant transport within a detailed network of stream reaches and corresponding sub-watersheds. A version of the expression for nutrient load, the dependent variable of the SPARROW regression, adapted from Schwarz et al. (2006) is:

$$L_i = [\Sigma_{j \in J(i)} L_j^{'}]\, F(Z_i^A; \theta_A) + [\sum_{n=1}^{Ns} S_{n,i}\, \alpha_n\, D_n\, (Z_i^D; \theta_D)]\; F^{'}(Z_i^A; \theta_A) \quad (1)$$

The nutrient load L_i at the downstream end of a given reach is defined as the sum of all attenuated nutrient sources in the watersheds draining to the upstream reaches (first term in Equation 1) plus the load originated within the watershed of the evaluated reach (second term in Equation 1). In both terms, the load generated is subjected to attenuation processes and the fraction transported to reach i (F or F') is expressed as a function of variables (Z_i^A) and parameters (θ_A) describing nutrient removal in streams and reservoirs. The summation term in the second part of the equation represents the nutrient source variables S_n and the source-specific coefficients α_n. Those sources associated to landscape processes (e.g. transport and retention) are regulated by the land-to-water delivery factor D_n, estimated as a function of delivery variables (Z_i^D) and parameters (θ_D).

The steady-state mass-balance structure of SPARROW quantifies the long-term net effects of nutrient flux (Schwarz et al., 2006). This establishes a practical limit for the shortest time span that can be used in simulations because transient dynamics are not considered in the model. We worked at the annual time step as a trade-off between avoiding transient dynamics and maintaining the integrity of the information collected during dryer and wetter years. The annual time scale is sufficient to locate high risk areas that are vulnerable to particular sources of pollution (Schoumans et al., 2009).

Channel transport characteristics

SPARROW decay estimation reflects the net annual removal of nutrients in river channels and in reservoirs. The in-stream attenuation of N considers denitrification and long-term storage as the most significant processes influencing retention at the

mean annual scale (Wollheim *et al.*, 2008b). Phosphorus attenuation considers the removal from the water column through biological uptake and biogeochemical processes in the sediment. Previous SPARROW in-stream decay specifications considered 1st order kinetics as an exponential function of water time-of-travel and an attenuation rate (Schwarz et al., 2006). We modified this in order to include parameters defining the uptake Efficiency Loss concept (Aguilera et al., 2013). In summary, our modification considers that nutrient uptake in a given reach depends on the available nutrient concentration, and is estimated based on an empirically derived power law (O'Brien et al., 2007). Thus, the fraction (F in Eq. 1) of nutrient flux delivered to the downstream reach is estimated as a function of the hydraulic load and the uptake velocity values found in a given stream reach.

The nutrient fraction retained in reservoir reaches was calculated as a function of the reciprocal hydraulic load and settling velocity in a given reservoir reach (Schwarz et al., 2006). However, removal percentages in these reaches were not expected to be significant in our results due to the small number of reaches identified as reservoirs and to their position nearby headwater reaches. Seitzinger et al. (2002) found that the location that reservoirs occupy in the river network can also be important when evaluating nutrient removal, since downstream reservoirs trap comparatively more nutrients that reservoirs upstream.

Data sources

Drainage network and sub-watersheds

A total of 79 reaches and sub-watersheds were delineated in a GIS platform (Fig. 1), using a Digital Elevation Model (DEM) of the Llobregat River Basin. The average size of each sub-watershed was about 60.2 km^2, ranging between 1.2 and 240.5 km^2.

Streamflow and water quality monitoring data

Mean annual river discharge at the outlet of the 79 sub-watersheds for each of the seven hydrological periods within 2000-2006 was estimated based on the Drainage-

Area Ratio Method (Emerson et al., 2005), using river streamflow data from 17 gauge stations managed by the Catalan Water Agency (ACA).

Monthly nitrate and phosphate concentration (measured by standard procedures; APHA, 1995) data were obtained from locations monitored by ACA along the Llobregat River and its tributaries (Cardener, Anoia, Gavarresa and Rubí). The average nutrient concentrations of the selected 23 monitoring sites (Fig. 1) during the period of 2000-2006 were 11.2 (± 9.1) mg NO_3^- L^{-1} and 0.9 (± 1.4) mg PO_4^{3-} L^{-1}. The spatial distribution of the highest and lowest concentration values slightly differs among nutrients (Fig. 1). Higher average concentration of nitrate (>15 mg L^{-1}) were observed in the middle section of the Llobregat River (vicinity of Manresa) and in reaches of the Anoia and Cardener Rivers (particularly at Clariana de Cardener). The concentration of phosphate was also high in these reaches, mainly at the lower part of the Anoia River.

Water quality data and daily discharge measurements, were used to calculate nitrate and phosphate annual loads (dependent variable for SPARROW model regression, in kg yr^{-1}) at each of the 23 monitoring sites using the program Load Estimator (LOADEST; Runkel et al., 2004). In total, the model included 161 annual load values for calibration (23 sites × 7 years).

Nutrient source terms

Diffuse nutrient sources were represented by three different land use types based on the land use map captured by the Thematic Mapper sensor of the Landsat satellite during 2002 (30x30m grid resolution, http://dmah.nexusgeografics.com/). The three types were distributed as urban areas (6% of the total catchment area), cultivated land (24%), and natural land (70%). The latter grouped forested and grass land areas into a unique explanatory variable in order to avoid collinearity between these two highly correlated explanatory variables. Diffuse sources were represented in the SPARROW models as S_{Urban}, $S_{Cultivated}$ and $S_{Natural}$ (km^2) with the associated α coefficients as in Equation (1).

Point source loads (S_{POINT} explanatory variable in Eq. 1, in kg yr^{-1}) were obtained from wastewater and industrial effluent data. These nutrient-rich effluents are considered to be discharged directly into the river and are not subjected to attenuation processes within the landscape. SPARROW includes a point source coefficient to solve for any systematic bias in the point source database. Because the point sources are discharged directly to streams and their loads are expressed in units of mass (same as the SPARROW response variable) a fitted point-source coefficient of 1 (θ_D in Equation 1) is expected (Alexander *et al.*, 2002a). Point source coefficients in the Llobregat River Basin were separated into two parameters, one associated to point source loads in the watershed upstream ($\alpha_{Point\_Upstream}$) the Abrera monitoring station (including the Anoia), and a second parameter ($\alpha_{Point\_Downstream}$) attributed to the downstream point sources. This distinction aimed to properly account for the industrial effluents discharged downstream Abrera, where it was assumed that their final fate was not accurately described in the database. In that section the point-source coefficient may take values considerably lower than one.

Land-to-water delivery

Land-to-water delivery variables (Z_i^D in Equation 1) and parameters (θ_D) describe landscape properties related to climatic, natural or anthropogenic factors affecting terrestrial nutrient transport (Schwarz et al., 2006). SPARROW assumes steady-state conditions when estimating land-to-water transport, reflecting the net effect of the processes involved. The land-to-water variable considered in our model was runoff (in m yr^{-1}), estimated as an area-weighted amount of overflow produced in each incremental watershed. The estimated annual runoff in the Llobregat River Basin, along with mean discharge values for each sub-watershed, accounted for the climate inter-annual variability in this study (Table 1).

Table 1: Mean values for discharge and runoff found in the Llobregat River basin for each of the 7 calendar years included in this study. The highest values for these variables were observed in 2004; the lowest, in 2005.

Year	Discharge [$m^3 s^{-1}$]	Runoff [$m\ yr^{-1}$]
2000	4.44	0.24
2001	2.92	0.20
2002	3.10	0.18
2003	4.41	0.27
2004	5.17	0.29
2005	2.36	0.17
2006	3.27	0.19

Modelling and calibration strategy

Nutrient source and aquatic decay parameters were estimated through the SPARROW automated capabilities based on the nonlinear weighted least squares (NWLS) method. Observed load in all 161 monitoring locations were included in the calibration.

Parametric estimation in SPARROW provided measures of model accuracy and predictive power as well as robust measures of uncertainty in statistical estimation of the regression parameters (Schwarz et al., 2006). The robustness of the parameter estimates can be further evaluated by applying the bootstrap algorithm incorporated in SPARROW. Bootstrap coefficients are usually more robust since they are estimated based on resampling with replacement from the set of mean nutrient fluxes at the calibration points, fitting separate NWLS regression models to the resampled data for the total amount of iterations specified in the model (Preston & Brakebill, 1999).

Results and Discussion

Model fit

The nitrate model calibration results provided a reasonable fit between measured and predicted loads and yields (Fig. 2 and 3). The nitrate model was considerably efficient as shown by its coefficient of determination (R^2 = 0.86; RMSE of 0.55). Fig. 2 illustrates the relationship between observed and predicted annual load values for each of the years included in this study.

The annual yield-R^2 value indicated that 66% of the variance of observed nitrates yield is explained by the estimated model. In the case of the phosphate model, the values of load-R^2 and yield-R^2 were 0.80 (RMSE of 0.73) and 0.60, respectively. When averaged over the seven-year period of the study, the greatest discrepancy between observed and SPARROW-predicted nitrate total yields (Fig. 3a) corresponded to monitoring points located in reaches in the Cardener River (Clariana de Cardener [CC] and Cardona stations [CD]) which are nearby wastewater treatment plant (WWTP) sites and downstream reservoirs. Also, reaches in the upper and middle part of the watershed (Pobla de Lillet [PL], Navarcles [NA]), and reaches in the Anoia River (Martorell [MA] and Sant Sadurní d'Anoia [SS]) showed considerable deviations between measured and predicted loads. In the phosphate model, CC and CD continued to have large differences between observed and predicted yield. The average predicted yield at Jorba [JO] in the Anoia River is also much larger than that observed.

Model coefficients

Significant explanatory variables for the nitrate and phosphate models and their associated parameters are given in Tables 2 and 3. Bootstrap estimates obtained after 100 iterations are also shown, along with the related 90% confidence intervals (CI). The obtained mean bootstrap coefficients were similar to the parametric coefficients, which indicated robustness in parameter estimation.

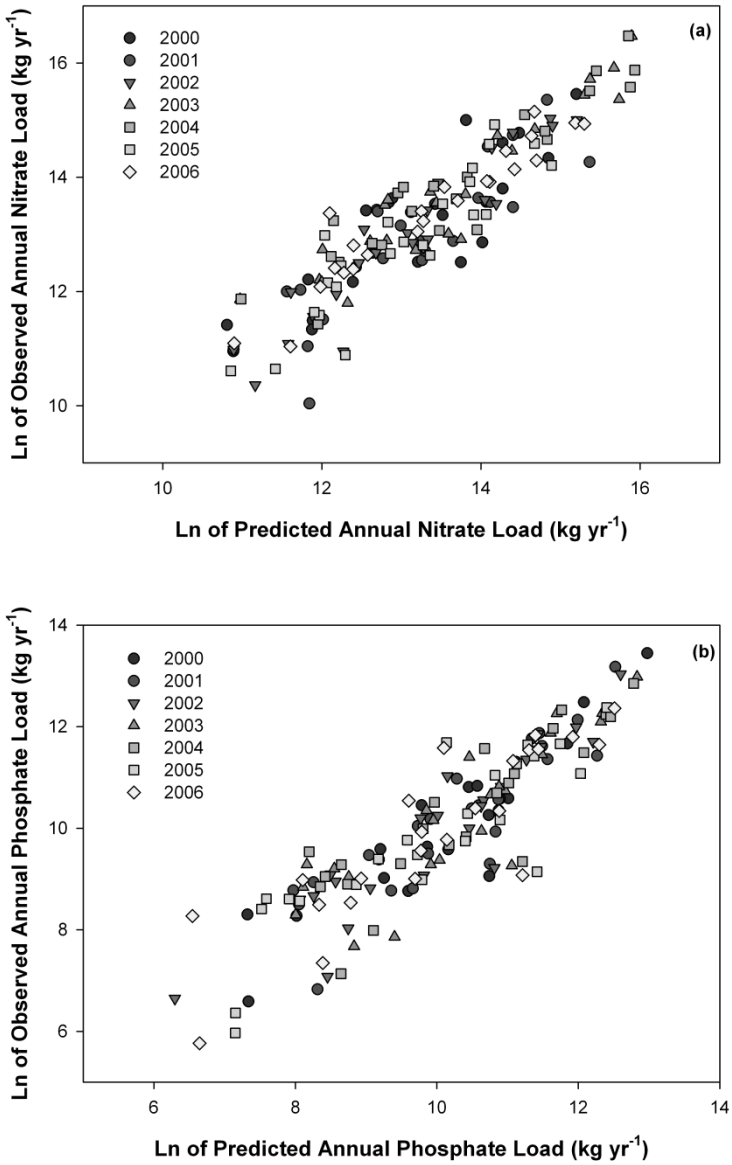


Figure 2: Mean in-stream observed annual loads (2000-2006) compared to model predicted loads at calibration sites (a: nitrate, b: phosphate).

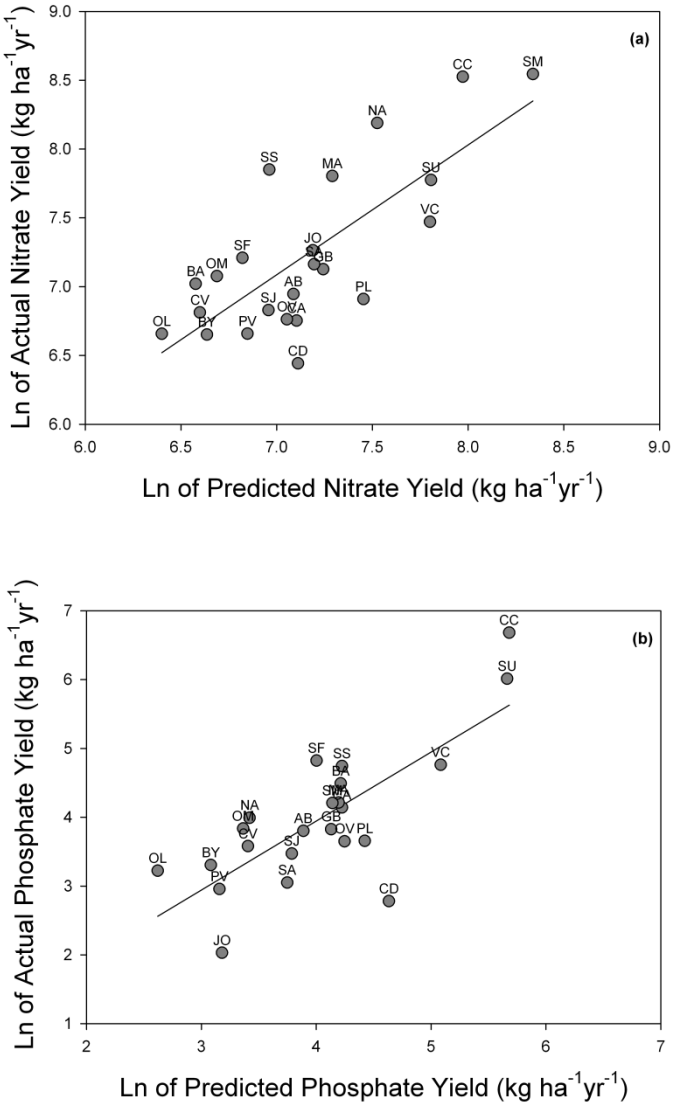


Figure 3: Mean Observed Total Yield versus Predicted Total Yield in 23 monitoring points (2000-2006) for nitrate (a) and phosphate (b). Total Yield represents the total load divided by the total area of the watershed at a given monitoring station.

The mean predicted yield for total flux leaving the reach generated in SPARROW between the years 2000 and 2006 was 4.03 kg N ha^{-1} yr^{-1} for the nitrate model and 0.26 kg P ha^{-1} yr^{-1} for phosphates. The averaged flow-weighted concentration values predicted by the model were 4.5 mg N L^{-1} and 0.3 mg P L^{-1}.

Table 2: Nitrate Model Parameter Estimation (CIs correspond to Bootstrap Estimation)

Explanatory Variable	Parametric Coefficient	*p*-value	Bootstrap Coefficient	Lower 90% CI	Higher 90% CI	Units
Nutrient Sources						
Urban land	**0**	-				kg N ha^{-1} yr^{-1}
Cultivated land	**7.43**	<0.001	7.26	3.34	11.28	kg N ha^{-1} yr^{-1}
Natural land	**3.34**	<0.001	3.14	2.28	3.89	kg N ha^{-1} yr^{-1}
Point_Upstream	**0.76**	0.005	0.78	0.52	1.36	dimensionless
Point_Downstream	**0.15**	0.76	0.15	0.15	0.15	dimensionless
Land-to-water delivery						
Runoff	**2.25**	<0.001	2.22	1.79	2.56	yr m^{-1}

Table 3: Phosphate Model Parameter Estimation (CIs correspond to Bootstrap Estimation)

Explanatory Variable	Parametric Coefficient	*p*-value	Bootstrap Coefficient	Lower 90% CI	Higher 90% CI	Units
Nutrient Sources						
Urban land	**0**	-				kg P ha^{-1} yr^{-1}
Cultivated land	**0.83**	<0.001	0.82	0.31	1.22	kg P ha^{-1} yr^{-1}
Natural land	**0.21**	<0.001	0.19	0.10	0.27	kg P ha^{-1} yr^{-1}
Point_Upstream	**1.54**	0.006	1.60	1.29	2.36	dimensionless
Point_Downstream	**0.79**	0.01	0.80	0.56	1.10	dimensionless
Land-to-water delivery						
Runoff	**2.73**	<0.001	2.82	2.19	3.48	yr m^{-1}

Nutrient sources apportionment

The SPARROW model developed for the Llobregat River identified that natural sources of the two nutrients predominated in the upper part of the watershed, whereas agricultural activities were more relevant in its middle section and the upper Anoia River (Fig. 4). Point sources were large contributors of nutrient loads, especially phosphates, in the lower part of the watershed. Such apportionment pattern broadly agrees with the land use distribution in the Llobregat River Basin (Marcé et al., 2012). Diffuse sources in the watershed accounted for the major part of nitrate load arriving to rivers (natural land, 42% and cultivated land, 33%). In the phosphate model, point sources accounted for the greatest portion of the total average load generated in the watershed (48%). These different patterns follow those described in large European watersheds where agricultural sources contribute the greatest load of nitrogen and urban and point sources are major sources of phosphate loads (Grizzetti & Bouraoui, 2006; Schoumans *et al.*, 2009). On the other hand, percentages of nitrogen diffuse sources attributed to the Llobregat River Basin (resulting from extrapolation from the European scale) range between 0 and 25 % in a study by Grizzetti & Bouraoui (2006). This is probably an underestimation of diffuse emissions of nitrates, and could be related to the differences in the land cover characteristics considered in the latter study. These watersheds had significantly larger percentages of their areas occupied by intense farming and crop cultivation than those found in the Llobregat River Basin. Thus, when extrapolated to the European scale, the values for diffuse sources in Grizzetti and Bouraoui (2006) could have been undermined in our study basin.

Nutrient export

Urban land and point sources contribution

Urban runoff generated after rainfall events washes off the contaminants accumulated on urban components (vegetation, automobiles, houses, shopping centers, streets, and industries), transporting contaminants into gutters and storm sewers to the nearest natural or man-made water course (Beaulac & Reckhow, 1982). Point sources mainly refer to the wastewater inputs associated with industrial and urban sewage inputs.

Point source parameter estimates should approximate the value of 1 since the field estimates of point sources in our models were expressed in units of loads. Point source coefficients related to nitrate pollution were greater in the upper part of the watershed (coefficient value of 0.76), than in the lower section (0.15). The uncertainty in the final disposition of the WWTP and industrial effluents in the lower section (downstream Abrera station), which in some areas are deviated and do not reach the Llobregat River, could justify an overestimation prior to model calibration. In the phosphate model, point sources parameters were higher than those in the nitrate model, yielding coefficient values of 1.54 for the upper part and 0.79 for the lower part.

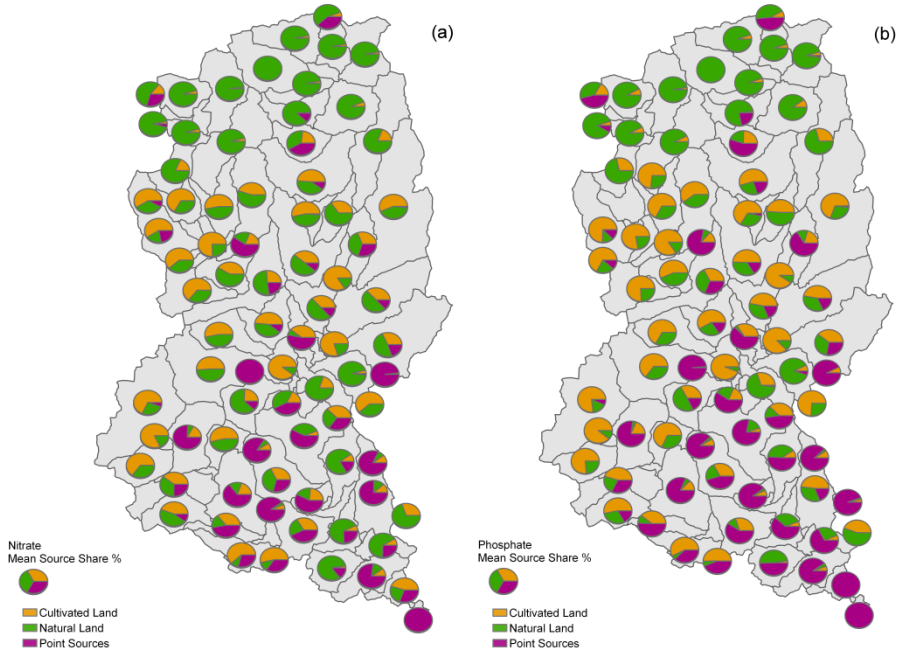


Figure 4: Share of incremental diffuse and point sources contribution to annual in-stream nitrate flux (a) and phosphate flux (b) in the Llobregat River Basin.

The explanatory SPARROW variable *Urban land*, which aimed to reflect the amount of nutrients transported by urban runoff, yielded a value of 0 in both nitrate and phosphate models (Tables 2 and 3). The proximity of point source locations to urban and industrial areas in the watershed could have masked the effect of nutrient export from *Urban land*. Point sources (expressed in loads) usually take preference in explaining nutrient flux deriving from urbanized lands, most likely due to the large masses of nutrients discharged directly into streams (Preston et al., 2011a). The connection between the networks collecting urban runoff and wastewater systems does also contribute to the uncertainty: sewage and rainwater are not separated in some areas of the Llobregat basin (Prat & Rieradevall, 2006). Furthermore, the annual scale considered in SPARROW does not accurately capture the peaks of nutrient transport by urban runoff, since these usually occur during and immediately after rainfall events. A long-term mechanism of urban diffuse pollution could occur due to infiltration but as impervious surfaces usually dominate urban areas, the cumulative infiltration is actually sharply reduced (Beaulac & Reckhow, 1982).

In order to determine the relevance of urban areas on nutrient flux, we tested the effect of point sources on the estimation of the urban land parameter in both nitrate and phosphate models. We removed the explanatory variables associated to point sources and set urban land as a diffuse source of nutrient pollution (not affected by land-to-water delivery). Doing so, the coefficient for urban land yielded higher values than other sources (26.2 kg N ha^{-1} yr^{-1} and 5.0 kg P ha^{-1} yr^{-1}), which approximated the values that point sources would have obtained in the previous models (29.9 kg N ha^{-1} yr^{-1} and 6.4 kg P ha^{-1} yr^{-1}) if they were associated to the extent of urban lands in the watershed (roughly 300 km^2). All other parameter values related to cultivated and natural land sources, in-stream nutrient retention, and land-to-water delivery remained comparable to the previous model exercises.

Cultivated land

One of the most representative dryland crops in the lower Llobregat River Basin is the vineyard. Nutrient export from vineyards in the region occurs during flood events and nutrient losses at the annual scale are indeed considerable, ranging from 8.0 kg ha^{-1} yr^{-1} of N and 6.5 kg ha^{-1} yr^{-1} of P (Ramos & Martínez-Casasnovas, 2005). Although Cultivated land represents only 24% of the total watershed area, the associated coefficient played an important role in explaining nitrate yield. The annual export coefficient value was 7.4 kg N ha^{-1} yr^{-1}, a value lower than the median export coefficient for mixed agriculture (including crops and pastures, and accounting for 14.0 kg N ha^{-1} yr^{-1}) reported by Beaulac & Reckhow (1982). Higher nitrogen export from diffuse sources such as agriculture occurs in watersheds with higher annual precipitation than the Llobregat River Basin. This is certainly associated to the biogeochemical behavior of NO_3^- , which is highly soluble and prone to be transported with surface runoff generated during rainfall events (Oeurng et al., 2010). The SPARROW N export coefficient is also much lower than the range of 21 and 35 kg N ha^{-1} yr^{-1} found in European watersheds characterized by larger areas of intensive agricultural activities (Grizzetti *et al.*, 2005).

The phosphate export coefficient associated to cultivated land was 0.8 kg P ha^{-1} yr^{-1}. This figure agrees with the median export coefficient for mixed agriculture of 1.0 kg P ha^{-1} yr^{-1} (range 0.5-1.5) reported by Beaulac & Reckhow (1982). Unlike N, P has low solubility and its transport is mostly in particulate form and not as much in the dissolved form (House & Denison, 1998).

Kronvang et al. (2009) reported that the range for average annual nitrogen and phosphorus loss from agricultural areas in 17 European catchments amounted 19.1 - 34.6 kg N ha^{-1} and 0.12 - 1.67 kg P ha^{-1}, respectively. These values are slightly higher than our export coefficients, particularly for nitrates. Differences might be due to the larger agricultural occupation in the watersheds considered by Kronvang et al. (2009), where a mean 51.1% of the total area of each of the 17 watersheds was covered by agricultural land.

Natural land

Nutrients derived from Natural land correspond both to diffuse anthropogenic sources and/or natural background concentrations of the nutrients under evaluation. Atmospheric deposition can be a negligible contributor of nitrogen in impaired systems, but where other nutrient sources (agriculture, urban agglomerations) are not present, deposition is usually the largest source of nitrogen that reaches streams (Preston et al., 2011b). Therefore, high N export could be expected from atmospheric or background sources in our model, and this is also the case for other impaired river watersheds such as those located in the south-central, mid-Atlantic and northeastern United States (Preston et al., 2011b).

The nitrate export coefficient for natural land estimated in SPARROW held a slightly higher value (3.3 kg N ha^{-1} yr^{-1}) than the median observed in forests of North America (2.5 kg N ha^{-1}). Nevertheless, the expected ranges for natural sources reported by Schwarz et al. (2006) are quite broad, 0.3-12 kg N ha^{-1} yr^{-1} for forest land and 0.5-25 kg N ha^{-1} yr^{-1} for grass and shrub land. In a forested watershed in the Montseny Mountains (NE Spain), Avila et al. (2002) estimated that total atmospheric deposition was 15 kg N ha^{-1} yr^{-1}. Atmospheric N deposition calculated for the year 2001 (Leip et al., 2011) ranged between 10 – 14 kg N ha^{-1} yr^{-1} in the upper Llobregat River Basin and between 5 and 10 kg N ha^{-1} yr^{-1} in the lower part of the watershed. These values are among the highest reported in Spain, and appear to be concentrated in the industrialized Northeast Spain. In fact, the natural rate of atmospheric nitrogen fixation is 10–35 kg N ha^{-1} yr^{-1} in Mediterranean shrub land, much higher than in other European forests (Billen et al., 2011).

The export coefficient attributed to natural land in the phosphate model (0.2 kg P ha^{-1} yr^{-1}) fell within the range reported in the literature (Beaulac & Reckhow, 1982). Forest land has been identified in previous SPARROW applications as a statistically significant source of the phosphorus exported to water bodies (Preston et al., 2011a).

Nutrient sources related to inter-annual hydrological variability

Climate-associated factors are considered the main determinants of nutrient export from terrestrial sources to water bodies, particularly in areas where forests predominate (Beaulac & Reckhow, 1982). An evaluation of regional SPARROW models in the United States recognized that climatic variables played a major role in the delivery of nutrients to streams (Preston et al., 2011a). In our models, the export of nutrient loads in sub-watersheds with lower runoff (a surrogate for precipitation values) appeared to be minimal or smaller than in areas characterized by a higher mean runoff value.

Inter-annual hydrological variability was associated to differences of mean nutrient load apportionment among nutrient sources (Fig. 5).Point source pollution was greater under dry conditions both for N and P. A high percentage of the measured water flow in streams and rivers under dry climatic conditions (Fig. 5b) is composed of point source effluents, a condition that frequently occurs in Mediterranean rivers (Martí et al., 2010), and has implications for pollutants other than nutrients (Petrovic et al., 2010).

Greater loads were predicted during the wetter periods (i.e., years with highest discharge and runoff values). Although the relationship between loads and runoff was the same for both diffuse sources (increasing load with increasing runoff, Fig. 6), the annual nutrient load delivered to reaches from their sub-watersheds was subjected to availability of nominal loads and runoff magnitude. Thus, the proportion of land use types and mean hydrological conditions, which differ among sub-watersheds, shaped the nutrient fluxes exported from the land phase. Phosphate loads from agricultural sources were greater than those derived from natural sources (Fig. 6). This could be the reason why the coefficient associated to runoff (land-to-water delivery) in the phosphate model (2.73 yr m^{-1}) was slightly higher than the one in the nitrate model (2.25 yr m^{-1}), since the natural land area is substantially larger than that occupied by cultivated land.

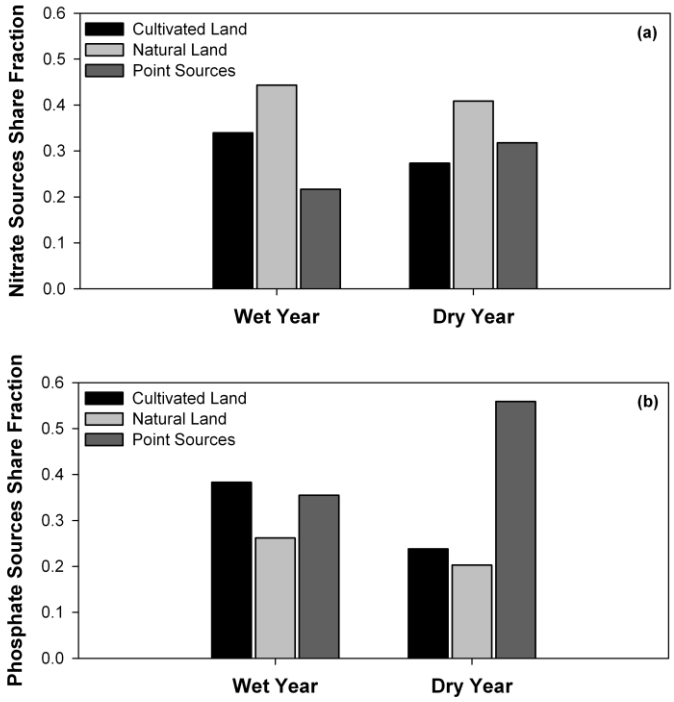


Figure 5: Mean Share Fraction for Diffuse and Point sources for nitrate (a) and phosphate (b) under wet and dry hydrological conditions in the Llobregat River Basin within the years 2000 – 2006.

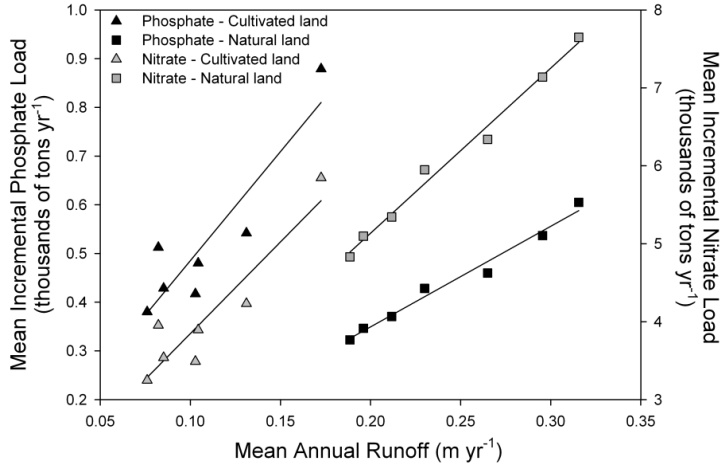


Figure 6: Relationship between Mean Incremental Load for Nitrates and Phosphates and Mean Runoff generated in the watershed within 2000-2006. Data points represent the mean load generated in each of the seven years considered in this study for each nutrient and nutrient source, according to the associated mean runoff values.

Nutrient in-stream retention

The fraction of flux removed (Fig. 7) is the calculated nutrient in-stream attenuation in each of the 73 river reaches during the seven years considered in this study. The mean annual percentage of in-stream nutrient removal between 2000 and 2006 was 23% for nitrates and 40% for phosphates. A modelling study performed in a Mediterranean small watershed in NE Spain also reported high mean retention factors (46% for P and 52% for N (Caille et al., 2011). The results obtained in Mediterranean rivers are therefore higher than others of river nitrogen removal in European watersheds, estimated around 10-20% (Grizzetti *et al.*, 2005). The low streamflow characteristic of the Mediterranean watersheds could favor higher nutrient uptake, contrasting with higher discharge values in larger European watersheds (Grizzetti *et al.*, 2005).

SPARROW estimates defined the removed nutrient fraction as inversely related to increasing nutrient concentration (Aguilera et al., 2013), mostly due to decreasing uptake velocity values. In-stream nitrate removal was greater in the upper Llobregat tributaries (lower order streams), where removal accounted for more than 60%. Instead, the higher order river sections in the Llobregat showed a practically negligible removal fraction (Fig. 7a). Low nitrate in-stream retention in the middle and lower sections of the Llobregat could be also related to interactions among several stressors (e.g. hydrogeomorphological alterations, low cover of riparian vegetation, saline effluents). In the phosphate model, high removal fractions were also observed in upper and middle reaches of the Anoia River (Fig. 7b), while low phosphate in-stream retention coincided with the location of urban and industrial nuclei (i.e. the presence of point source discharges, Fig. 1). Results provided by Alexander et al. (2009) also indicated that nitrate removal efficiency could be substantially reduced during periods and locations with high nitrate loadings, concentrations, and discharges.

Point source effects in systems such as the Llobregat River Basin, where the natural flow of the river is low, are exacerbated by the fact that wastewater inputs usually account for a large part of the streamflow. Though streams that receive wastewater

inputs are able to maintain a capacity to retain and transform nutrients, their self-purification ability is usually lower than that of pristine streams (Martí et al., 2010). The high loadings coming from point sources can mask and even minimize the high stream nutrient retention naturally occurring in Mediterranean streams.

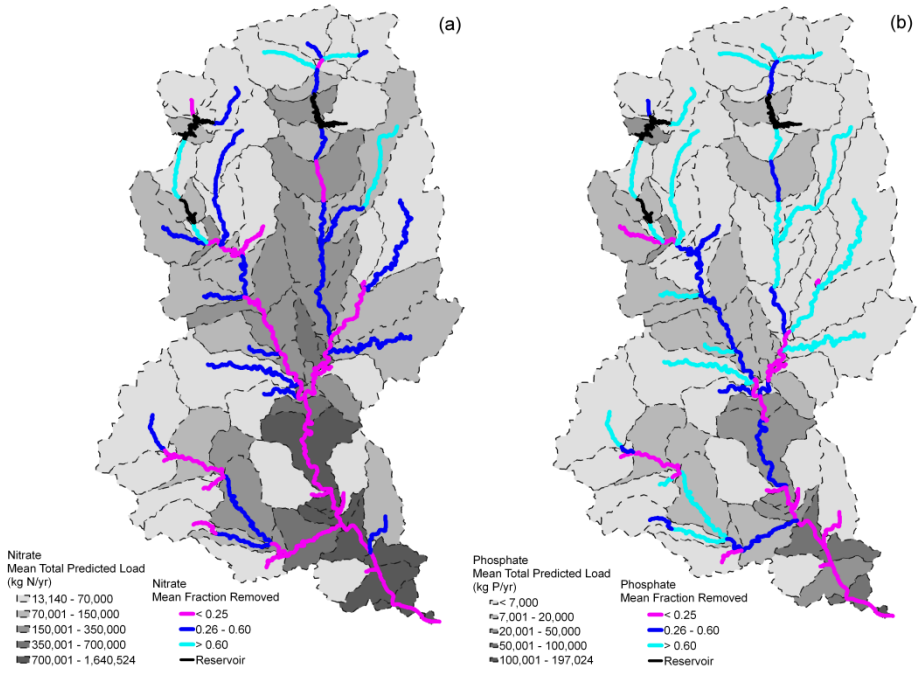


Figure 7: Mean In-stream Nutrient Removed Fraction obtained from SPARROW reach decay specifications (a: nitrate model, b: phosphate model). Mean Total Predicted Load reaching the streams for each model is obtained after land-to-water delivery and in-stream decay processes have taken place. The dashed line represents sub-watershed division.

Conclusions

The application of SPARROW in the Llobregat River Basin showed that models working at a level of intermediate complexity may offer a good compromise between process description and treatment of data available from water agencies. Exported nutrient load in the basin increased in the downstream direction, coinciding with the decrease of nutrient removal capacity occurring in the lower part of the watershed. Such decline in the natural depurative ability of streams is likely caused by the substantial chemical and geomorphological impairment that characterizes the lower Llobregat and Anoia reaches. Our models identified diffuse pollution as the major source of nitrate export, and point sources as the largest contributors of phosphate. Overall, nutrient export and apportionment reflected the inter-annual hydrological variability: larger nutrient loads from diffuse sources were observed in wetter years and point sources accounted for the greatest portion of nutrient loading under dry conditions. The Llobregat River is an important source of drinking water supply to the Barcelona Metropolitan Area (Mujeriego et al., 2008). By identifying the major sources of nutrients and their spatial distribution, water authorities could be able to plan and implement specific policies and measures to improve water quality and availability. Furthermore, having in mind the particular nature of the Mediterranean region and the associated predicted decrease in precipitation (Giorgi & Lionello, 2008), our model results could be used in management and decision-making processes related to the improvement of water quality under dry climatic conditions. Based on these results, such strategies should focus mainly on minimizing nutrient export to downstream river reaches, in order to allow these to restore their natural nutrient assimilation. Since point sources contribute substantially to nutrient excess in downstream reaches, their regulation and management adaptation to the Mediterranean regime could lead to an improvement in water quality conditions.

Chapter 3

A primer on the use of Dynamic Factor Analysis to study the spatio-temporal variability of river water quality series

State of the Science

Global environmental change has intensified in the recent decades as a direct consequence of human activities (Meybeck 2003). Regardless of the nature, whether natural or anthropogenic, of recent changes and transformations in all systems comprising the Earth, there is an increasing need for using collected data and optimizing the tools and methods for the detection and attribution of environmental change effects. Hegerl et al. (2010) define this process as stating specific causal factor(s) to which a particular change is being attributed and whether its origin relates to changes in environmental conditions and/or other external drivers. The main objective in environmental data analysis in this context is thus to identify change and attribute it to a potential source or cause (Lyubchich et al. 2013).

Freshwater systems are directly threatened by anthropogenic activities (Vörösmarty et al. 2010) in terms of both human water security and biodiversity. Consequently, vulnerability of global water resources to ongoing environmental change should be a priority topic to assess (Vörösmarty et al. 2000). Furthermore, there is a growing need to evaluate the nature of water resources vulnerability at regional and global scales while taking advantage of knowledge obtained at the local scale and through case studies (Vörösmarty 2002). Even though water quality and quantity are intimately linked, they are rarely measured simultaneously (Carr and Neary 2008). The assessment of quality, and not only quantity, of current and future freshwater resources is of upmost importance for the health and well-being of humans and ecosystems.

Time-series provide valuable information about the physical, biological, or socio-economical system that shaped them (Ghil et al. 2002). The idea that a time-series exhibits repetitive or regular behavior over time is of fundamental importance because it distinguishes time-series analysis from classical statistics, which assumes complete independence over time (Shumway and Stoffer 1982). The purpose of time-series analysis is therefore to extract key properties of the system under study by quantifying

certain features of the time-series. These properties can help understand and forecast the future behavior of the system (Ghil et al. 2002). Time-series analysis is thereby of particular interest in the study of environmental change and its effects on water quality.

Worldwide assessments of water resources rely to a great extent on fragmented data (Vörösmarty et al. 2010). The analysis of environmental time-series is thus not straightforward since available datasets are usually populated with missing observations and characterized by heterogeneous temporal resolutions. These commonly encountered issues usually require the preliminary use of statistical techniques to obtain even-spaced time-series and to fill data gaps, frequently done by interpolating missing values. Long-term observations are also usually required to detect any change and further attribute its driver(s). Furthermore, the assessment of trends and patterns in water quality is confounded by the interplay of many influencing factors, acting on different temporal and spatial scales (Kundzewicz and Krysanova 2010). These constraints tend to narrow the choice of available methods and impose limitations on the datasets that can be ultimately included in the analysis. Despite all the uncertainties involved in water quality modeling with limited input data, models are very important tools to support water managers and policy makers (Kundzewicz and Krysanova 2010).

Several time-series analysis techniques can be used to study responses and patterns embedded in datasets. However, in environmental datasets containing data gaps (those from monitoring programs; for instance), most common types of times-series analyses are not applicable. Classical time-series analysis requires complete datasets, and often even long time-scales (Bers et al. 2013). Furthermore, methods like spectral analysis are limited to characterizing the spectral density to detect any periodicities in the data, and thus do not necessarily allow the user to identify complex patterns (e.g., non-monotonic trends, irregular cycles) embedded in a set of time-series (Zuur et al. 2003a). In fact, the detection and attribution of global change effects on water quality 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).

Dynamic Factor Analysis (DFA) is a dimension-reduction method that estimates underlying common patterns in a set of time-series (Zuur et al. 2003a). An attractive feature of DFA is its ability to treat time-series that have been recorded irregularly over time. Moreover, DFA allows time-series to be short and thus the lack of sufficiently long records does not represent a problem (Zuur and Pierce 2004). The extracted patterns (e.g., cycles and/or trends) are associated to factor loadings, which indicate the weight that each pattern partakes in every monitoring point. These two end products can be then analyzed in order to characterize the temporal and spatial variability of the extracted water quality signals. The resulting description of the extracted patterns thus facilitates the interpretation and the attribution of drivers behind the observed behavior of the system.

The present work describes the application of Dynamic Factor Analysis and complementary methods in the identification of regional patterns enclosed in water quality time-series, with the aim of detecting and attributing the effects of global change phenomena on river water quality. The next section provides detailed explanations of the different steps and procedures of our suggested methodology. 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.

Methods

Rationale of the method and workflow

The upcoming sections describe the full set of steps of our proposed methodology (illustrated in Fig. 1; key R scripts included in Appendix B) to detect and potentially attribute the effects of global change phenomena on river water quality time-series:

1) Firstly, data exploration is carried out for 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 unreasonable values based on expected ranges.
2) Maximal Information-based Nonparametric Exploration (MINE) is used to find significant associations between variables and time in large river monitoring datasets. This optional step is intended to help in selecting relevant variables in large databases containing tens or hundreds of water quality variables.
3) Subsequently, time-series with a significant relationship with time are further explored by means of DFA, in order to extract patterns and their relevance (i.e., factor loadings) in a set of monitoring points.
4) Based on DFA results, the spatio-temporal variability of the identified water quality patterns is characterized using state-of-the-art time-series analysis for trend detection and for the identification of significant oscillations and relationships with streamflow.
5) Using the Factor Loadings as an expression of the relevance of the different identified patterns across sampling points, drivers behind the observed behavior are identified using regression methods and cluster analysis.

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. Subsequently, a visual exploratory analysis of the series is performed using the plotting capabilities of R. Possible outliers are detected and dealt with on an individual basis taking into consideration expected ranges of values for specific variables.

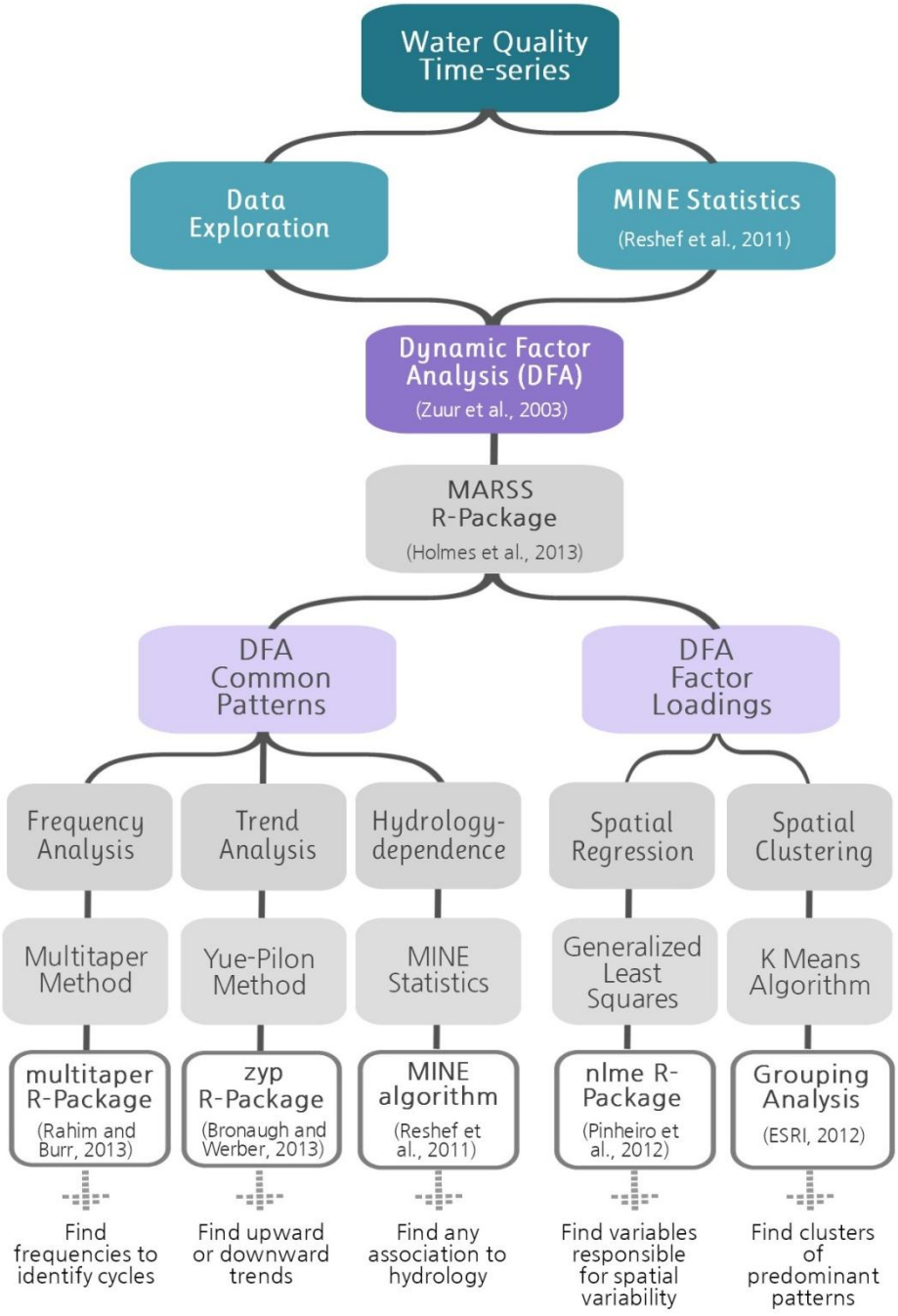


Figure 1: Methodological Flowchart

Exploring significant associations between variable pairs in a large dataset using MINE

In order to obtain significant associations for two-variable relationships in our datasets, time-series from all monitoring points can be subjected to the Maximal Information Coefficient (MIC) method (Reshef et al. 2011). MIC captures a broad range of associations, including but not limited to linear relationships, and provides a score for functional relationships equivalent to the coefficient of determination. The basic strategy behind MIC is to calculate a measure of dependence for each pair of variables, then ranking the pairs by scores and examining the top-scoring pairs (Reshef et al. 2011).

MIC belongs to a larger class of Maximal Information-based Nonparametric Exploration (MINE) statistics (Reshef et al. 2011) for identifying and classifying relationships in large data sets. The MIC score indicates the strength of the relationship (a value of 1 being the strongest). The MINE statistics include several diagnostic statistics describing the shape of the relationship, such as the natural measure of nonlinearity (MIC-p^2) and the deviation from monotonicity measured by the Maximum Asymmetry Score (MAS). These can be used to characterize relationships that, in turn, may be further studied with more specialized or computationally intensive techniques (Reshef et al. 2011).

MIC is a rank-order statistic and thereby the uncorrected p-value under a null hypothesis of statistical independence depends only on the MIC score and on the sample size of the relationship. The p-values for various MIC scores at different sample sizes are available at http://www.exploredata.net/Downloads/P-Value-Tables. Based on this table, we calculated the MIC significant scores at different significance levels (0.05, 0.01, and 0.001) for any sample size fitting power functions.

Relationships between all possible pairs of water quality variables and the relationship between each variable and time are tested in all monitoring stations with MINE to identify significant MIC scores. The focus is then set on the associations with the

highest significant MIC scores and that are present in the highest number of monitoring stations. The collection of significant MIC scores informs about the most relevant associations between variables in the dataset, and particularly about the existence of temporal patterns in the time-series. Apart from being a powerful additional exploratory analysis, this information may be used to select relevant variables to include in DFA if the analysis is not a priori focused on specific variables.

Extracting common patterns in 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 in the so-called state-space model (Harvey 1989), which treats each observed time-series as a linear combination of multiple state processes (Holmes et al. 2012). 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):

$$\begin{aligned} x_t &= x_{t-1} + w_t \qquad & w_t &\sim MVN(0,Q) \\ y_t &= Z_x + v_t \qquad & v_t &\sim MVN(0,R) \end{aligned} \qquad (1)$$

The observations (y_t) are modeled as a linear combination of hidden patterns (x_t) and factor loadings (Z; Holmes et al. 2012). The terms w_t and v_t contain the error associated to the observations and the hidden patterns, where Q and R are the covariance matrices associated to these error terms. The MARSS algorithm uses the Kalman filter and smoother to compute the best estimate for the hidden patterns given a set of assumed values for the unknown parameters Z, Q and R. 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).

The main advantage of DFA is its ability to deal with missing values, which are frequently encountered in time-series collected in river water-quality monitoring

programs. Depending on the complexity of the model and the number of time-series involved, however, the DFA method can be computationally expensive. Analyses included in this chapter were performed in a Linux-based High-Performance Computing Cluster, which allowed the simultaneous analyses of several DFA models. Still, computational times extended beyond several weeks for analyses that included the longest records and the largest number of time-series.

This limitation is exacerbated by the fact that several algorithm settings can be adopted. Analyses may include several combinations of the maximum number of common trends that can be potentially extracted (m) and different structures for the variance-covariance R 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. This is translates into small or large diagonal elements in R (the latter indicating that the corresponding time-series are not fitted well). Conversely, the off-diagonal elements represent the interactions between response variables that are not captured by the common patterns. Fortunately, the DFA implementation in the MARSS R-package already includes a best-model selection approach based on the Akaike Information Criterion (AIC). After DFA for all tested algorithm settings are completed and the best model is identified, the extracted common patterns and the associated factor loadings can be studied in detail as described and illustrated in the following sections.

Characterization of Common Patterns identified by DFA

Frequency Analysis

Both deterministic and stochastic processes can, in principle, be characterized 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 at this stage of the exercise is to obtain significant frequencies present in the extracted common patterns that could be potentially linked to climatic or phenological signals 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 and line components (Ghil et al. 2002). MTM is non-parametric, since it does not use an a priori, 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 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-bandwith parameter (nw) and the number of tapers (k) are chosen by the method itself based on the input data. The constructed orthogonal tapers used in this method minimize the spectral leakage due to the finite length of the time-series, 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 compares the values of the background spectra with the power in a line component resulting in the test. Formally, the F test is a F variance-ratio test with 2 and 2k-2 degrees of freedom (dof.) for the significance of the estimated line component (Thomson 1982). When referencing the F distribution, the numerator degrees of freedom are always given first, as switching the order of degrees of freedom changes the distribution. The higher the value of the F statistic, the higher is the significance of a specific frequency.

Normally, the critical value for the F test is computed using the baseline, non-adaptively-weighted weights: 2 and 2k – 2 dof. Since the F test statistic is the ratio of two chi-squared, the numerator is a 2 dof chi-squared (e.g., a squared normal random variable) and the denominator is the set of all of the tapered spectra. The dofs output in the multitaper R-Package indicates approximately how many degrees of freedom are left after the process has finished. Therefore, to find critical values for certain frequencies (freq), the estimation must be based on F(2, dof[which(freq)]) instead of just doing a F(2, 2k-2; W. Burr, personal communication, October 8, 2013).

Based on the table of critical values for the F distribution at the p=0.05 level of probability, critical values for our particular frequencies 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.

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 series were independent. 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 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. 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.

Hydrology-driven dynamics

Significant associations between resulting DFA common patterns and DFA reconstructed river discharge series (Q) are assessed by means of the MINE algorithm (Reshef et al. 2011). The DFA reconstructed Q series are the extracted DFA patterns from the original river discharge time-series multiplied by the corresponding factor loadings. The result is a reconstructed, filtered time-series with an enhanced signal-to-noise ratio and free of data gaps. Relationships between common patterns and reconstructed Q series showing high and significant MIC scores can be further explored to assess the characteristics of the hydrological dependence.

Distribution of DFA Factor Loadings across monitoring points

Drivers of the spatio-temporal variability of water quality

Factor loadings are the weights or levels of importance of a particular pattern at each monitoring point. For each variable, factor loading values of individual extracted patterns can be plotted spatially, based on geographical coordinates, in order to visually identify potential clusters of patterns in space. The identification of the 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., climate forcing and land-use distribution) 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. The identification of an optimal set of explanatory

variables for a regression model is not an easy task, and the presence of collinearity in the explanatory variables can further complicate the process.

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-Squared (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 spatio-temporal variability of individual DFA patterns extracted from our time-series.

Grouping analysis of predominant patterns

In order to further assess any grouping or clustering of 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 where all the features (in our case, the factor loadings associated to 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 more efficient number of clusters is solved using the Calinski-Harabasz pseudo F-statistic (ratio reflecting within-group similarity and between-group differences).

Case Study Sites

Data from water quality monitoring programs in three watersheds in Mediterranean Spain was used to illustrate our methodological framework in describing the spatio-temporal variability of water quality. 1) The Llobregat River Basin (NE Spain) covers a drainage area of 4,948 km^2 and hosts a population of ca. 3 million inhabitants (Marcé et al. 2012). The Llobregat River is a typical Mediterranean water course heavily impaired by damming, mining activities and pollution coming 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 km^2. The hydrological regime of the Ebro River is dictated in part by its contrasting tributaries, from snow-fed 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 km^2) 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 medium part 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 main criterion for selecting the time period was the longest possible time-series available at as many as possible monitoring stations within each river basin. The frequency of sampling was monthly, and the series length varied among

basins: 17 years for the Júcar, 11 for the Llobregat, and 31 for the Ebro. The majority of time-series were nevertheless characterized by having missing observations.

Results

The proposed methodology can be tailored to concrete objectives and characteristics identified in a concrete river basin, where the overall aim is to detect and characterize spatio-temporal patterns of water quality variability. Figure 2 introduces three situations of applicability of such tailored methodological steps, based on specific objectives, and, to some extent, on the availability of water quality data and the coverage of the monitoring network regarding basin size. These situations are then exemplified in three different river basins in the following sections.

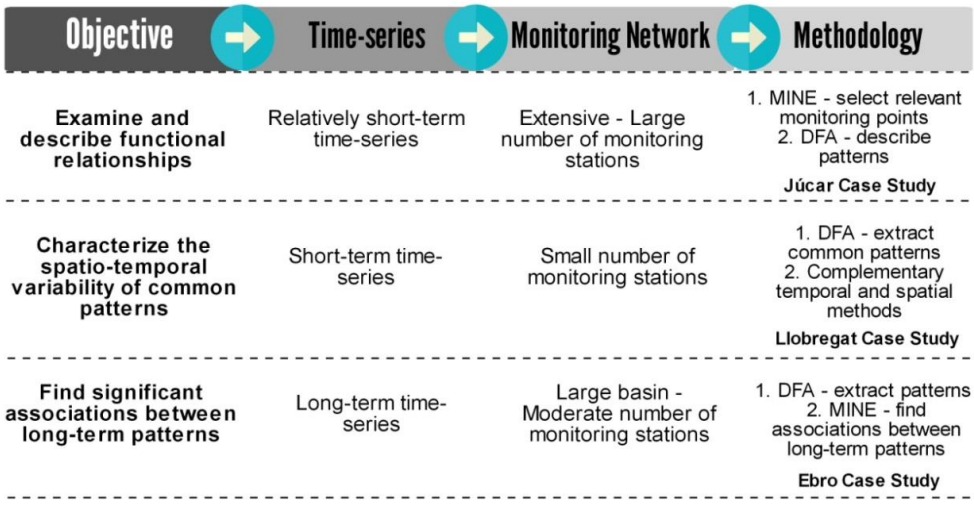


Figure 2: Tailored methods to specific objectives and basin characteristics, and application to the three river basins used as study cases.

Examining functional relationships with MINE and DFA: Nitrate common patterns in the Júcar River Basin District

River basin districts (RBD) are defined as the area made up of one or more neighboring river basins. Water quality monitoring networks are usually organized at the river basin level, though management plans occasionally tend to group basins of a particular region for a conjunct administration of water resources. Such grouping can elevate the number of monitoring points within datasets, impeding the joint analysis of patterns of a high number of environmental time-series. Ideally, a method capable of finding functional relationships between variable pairs as well as variables and time should be implemented to locate and identify relevant points and variables where significant associations are found. The Júcar River Basin District (Spain) is formed by the aggregation of watersheds that flow into the Mediterranean Sea. An initial exploratory analysis of nitrate time-series was carried out using the MINE algorithm in order to identify relevant monitoring points out of the 90 available in the Júcar RBD. The analysis revealed that a strong relationship of nitrate concentration with Time was significant in 50 monitoring points (Fig. 3). Consequently, we aimed to detect common nitrate patterns among time-series using Dynamic Factor Analysis and further characterize their shape and variability across time and space in those particular 50 points with significant functional relationships across time.

Among the four common extracted patterns by DFA (Fig. 4), Pattern 1 had the largest relevance in most monitoring points across the Júcar RBD (Fig. 5), and it was mainly described by frequency analysis as a cycle with significant annual periodicity. MINE statistics had initially identified a predominantly nonlinear, non-monotonic association between the variables Nitrate and Time, which could be related principally to Pattern 1 detected by DFA. 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 departure from monotonicity) in Figure 3 shows that non-monotonic

functions best described the time-series found along the Júcar River and in the upper Júcar district. The displayed non-monotonicity coincided, for the most part, with stations where cyclic patterns, such as DFA Pattern 1, predominated. In this way, DFA results further explained the preliminary exploratory MINE results about the main characteristics of nitrate spatio-temporal variability, and described the shape and key properties of the time-series represented in the extracted patterns. The initial MINE exploratory analysis was necessary to filter only those monitoring points with a significant functional relationship across time in order to avoid the cumbersome implementation of DFA involving a set of 90 time-series.

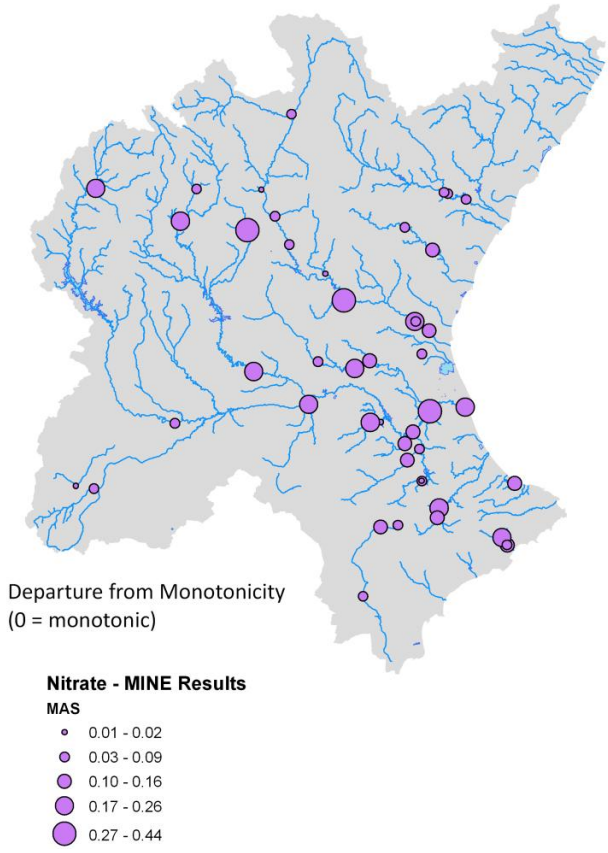


Figure 3: MAS scores for the 50 nitrate time-series. The highest scores display the departure of the functional associations from monotonicity. The predominant non-monotonic behavior is observed in the mid-western section of the basin.

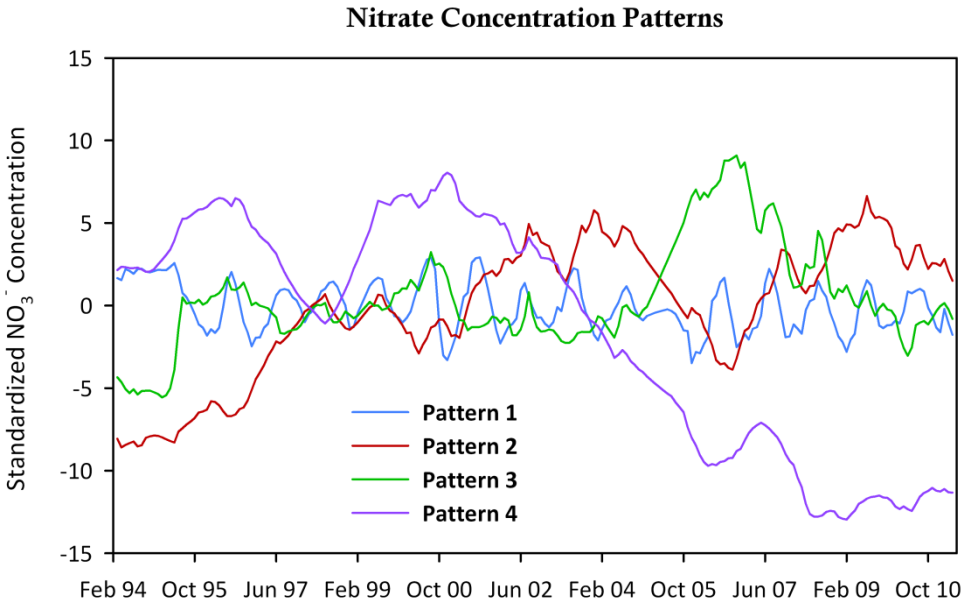


Figure 4: Common patterns extracted from a set of 50 of 90 nitrate time-series where previous MINE analysis had indicated a significant functional association with time.

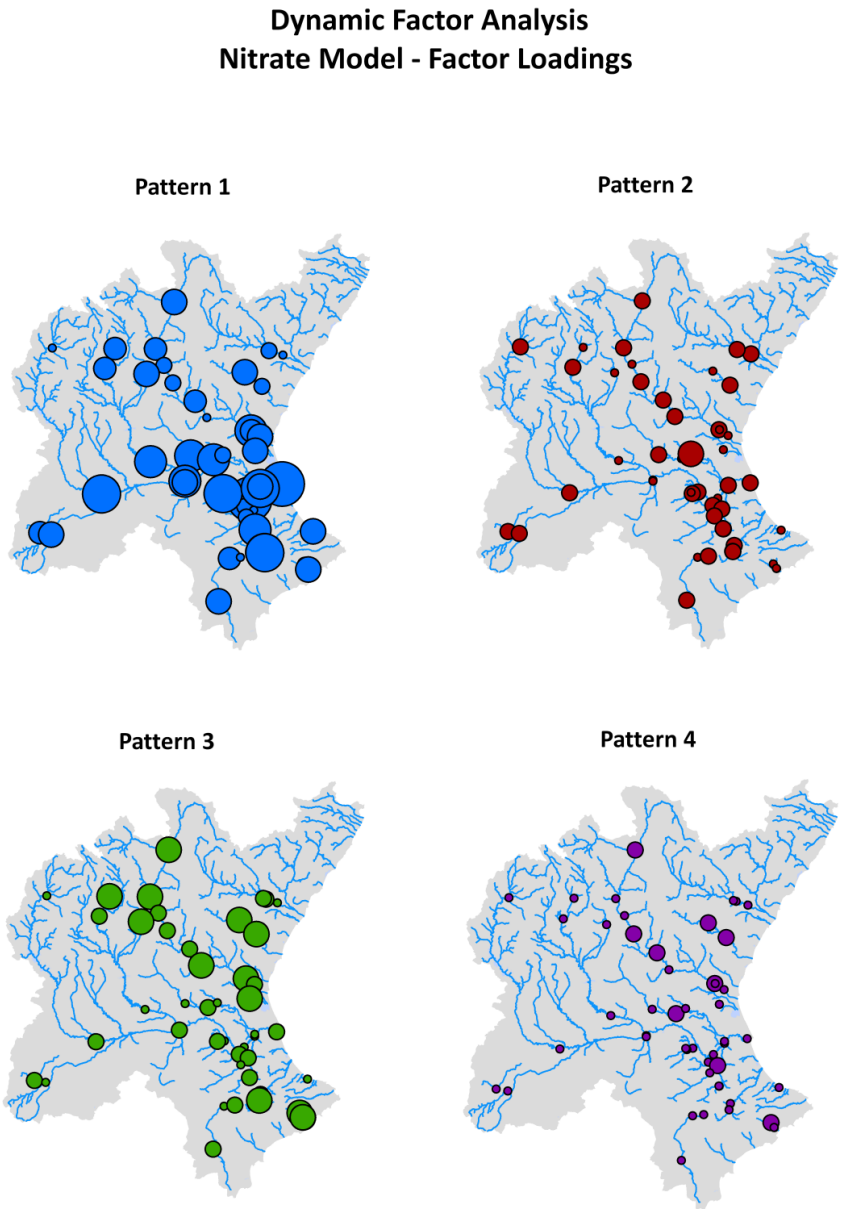


Figure 5: Factor loadings associated to the corresponding extracted patterns. Larger circles indicate larger weight or predominance of a specific pattern over the remaining three. Overall, Pattern 1 appeared to predominate across the Júcar RBD, specifically in the middle section along the Júcar River.

Characterizing the spatio-temporal variability of nitrate patterns: The Llobregat River Basin

River basins, regardless of their size, commonly encompass a variety of land uses and climate characteristics. Even though anthropogenic activities in particular areas within a basin 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 Llobregat River Basin (NE Spain) is a well-studied basin with a manageable number of monitoring points of water quantity and quality short-term data. A comprehensive detection and characterization of nitrate concentration was carried out following all the steps underlined in Figure 1. Since our objective was to specifically analyze patterns of nitrate dynamics and their drivers across the entire monitoring network, we skipped the MINE analysis step (which aids in choosing relevant time-series from a large database), and worked with time-series from all available monitoring points.

The DFA model for nitrate time-series in 20 points extracted four common patterns, quite dissimilar among one another (Fig. 6). Frequency (Multitaper) and trend (Yue-Pilon) analyses of these patterns resulted in the full characterization of their temporal variability (Table 1). On one hand, detecting significant frequencies in the extracted patterns can facilitate the identification of seasonal cycles and oscillations potentially related to regional and global climatic modes. On the other hand, long-term significant trends can assist in identifying any relevant event that, in this, case, could have facilitated either the increase or decrease of nitrate concentration across the study period. Pattern 1, for instance, was described by a significant decreasing trend and significant periods of 2.1 and 1.7 yr. 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.

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 to Pattern 1 in the nitrate model had the largest values, and

they were specifically located on the lower section of the Llobregat River (Fig. 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 varying environmental variables included in the gls models. The MINE analysis for associations can further elucidate the relationship (if any) between the extracted patterns and hydrological variability. 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 to Pattern 1). The remaining patterns were more or less equally scattered across the entire basin, particularly Pattern 3, which presented high values for Factor Loadings in most monitoring points. Conversely, Pattern 4 had a higher relevance in the upper Anoia tributary (SW of the basin). The gls regression models for Patterns 2 and 3 did not identify any significant explanatory variables. Pattern 4, on the other hand, was related to negatively to river discharge.

Methods like clustering analysis provide information on spatial grouping of predominant patterns within the study basin. Grouping Analysis of the factor loadings associated to all four patterns extracted in the Llobregat basin, based on a distribution of two groups revealed no clear differentiation for the predominant Pattern 1 in specific stations, but the points related to this cyclic and hydrology-driven pattern did remain part of the same group (not shown). Group 1 enclosed mainly those stations where Patterns 2 and 3 were the most relevant.

The best-fits, which are the linear combination of the extracted patterns based on the corresponding factor loadings in each monitoring point, can be plotted against the original time-series in order to determine the extent of model performance. Best-fit plots, as well as 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 eventually identifying any potential cause(s) behind such misrepresentation, whether it is based on data restrictions or pattern behavior.

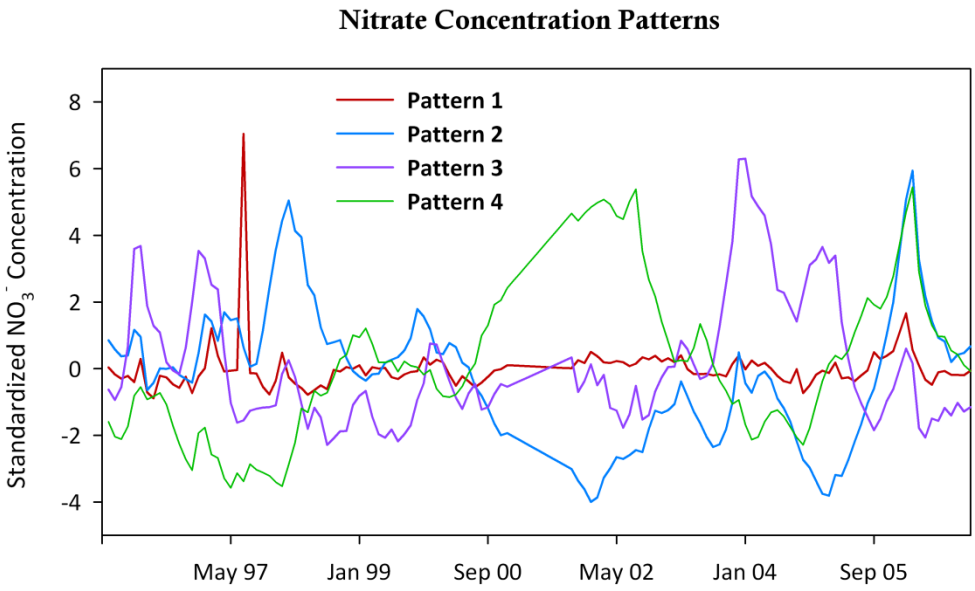


Figure 6: Nitrate common patterns extracted in the Llobregat basin (1995-2006)

Table 1: Main characteristics found in the resulting nitrate patterns and associated factor loadings that help in describing the spatio-temporal variability in the basin

DFA MODEL	COMMON PATTERNS			FACTOR LOADINGS	
Nitrate	Significant Cycle	Significant Trend	Hydrology-driven	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

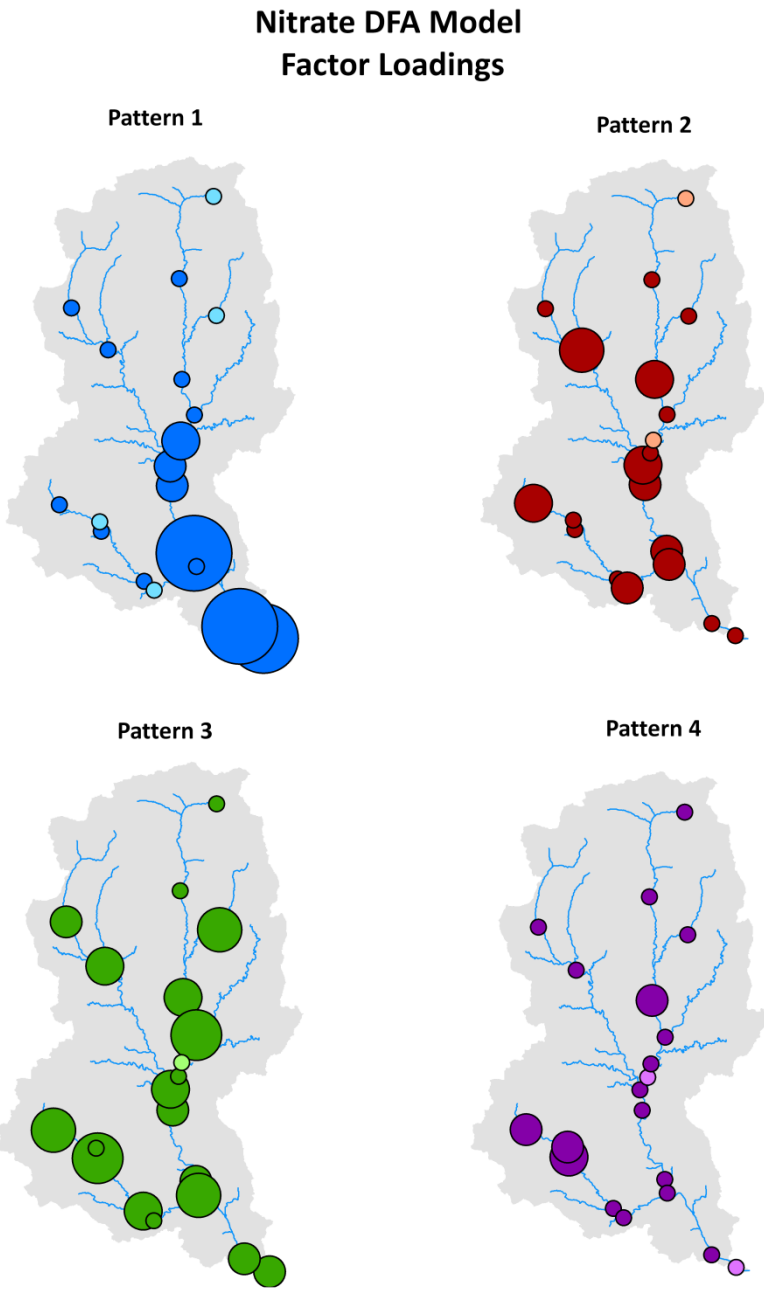


Figure 7: The magnitude of the Factor Loadings is represented by the size of the circles for each monitoring point. Pattern 1 in the nitrate model was the most relevant in the basin, though the highest values concentrated along the mid and low Llobregat River sections. The light-colored circles represent negative Factor Loadings, which in turn describe the opposite behavior of the corresponding extracted pattern.

Exploring associations between patterns of nitrate and dissolved oxygen: The Ebro River Basin

Water quality monitoring programs in many basins worldwide have recorded data on dissolved nutrients during several decades, thus making it possible to carry out analyses of periods long enough to show significant changes in concentration and loading (Ibáñez et al., 2008). Long-term patterns extracted from sets of such time-series can thus provide information on significant trends in water quality of individual variables, as well as between variable pairs. The capabilities of the MINE algorithm were used in order to explore any significant functional association among patterns of different variables in the Ebro River Basin, where a sufficiently long dataset was available (30 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 most remarkable of the significant associations between patterns involved nitrate and dissolved oxygen in rivers and streams (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, $p<0.001$), though the trend component in Pattern 4 for dissolved oxygen was only marginally significant (Kendall's tau of 0.08, $p<0.05$). The decreasing 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). Furthermore, the frequency analysis for both patterns presented significant periodicities associated to 4-yr cycles.

Figure 8 displays the different associations between the two patterns and their location on the basin monitoring network. Association 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). Here, the nitrate pattern describes a decreasing trend for nitrate

concentration along the 30 years included in the analysis, and the opposite happens with DO. It is actually 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 (eastern part of the basin) within the 30-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 drivers of change.

Table 2: MIC scores and associated MINE statistics for the significant relationships between water quality patterns in the Ebro basin.

X var	Y var	MIC (strength)	MIC-p^2 (nonlinearity)	MAS (non-monotonicity)
NO_3 Pattern 3	DO Pattern 4	0.99	0.33	0.03
NO_3 Pattern 3	PO_4 Pattern 3	0.90	0.40	0.11
NO_3 Pattern 3	NH_4 Pattern 1	0.83	0.45	0.05
NO_3 Pattern 3	BOD5 Pattern 3	0.75	0.25	0.09
NO_3 Pattern 3	pH Pattern 2	0.74	0.17	0.07

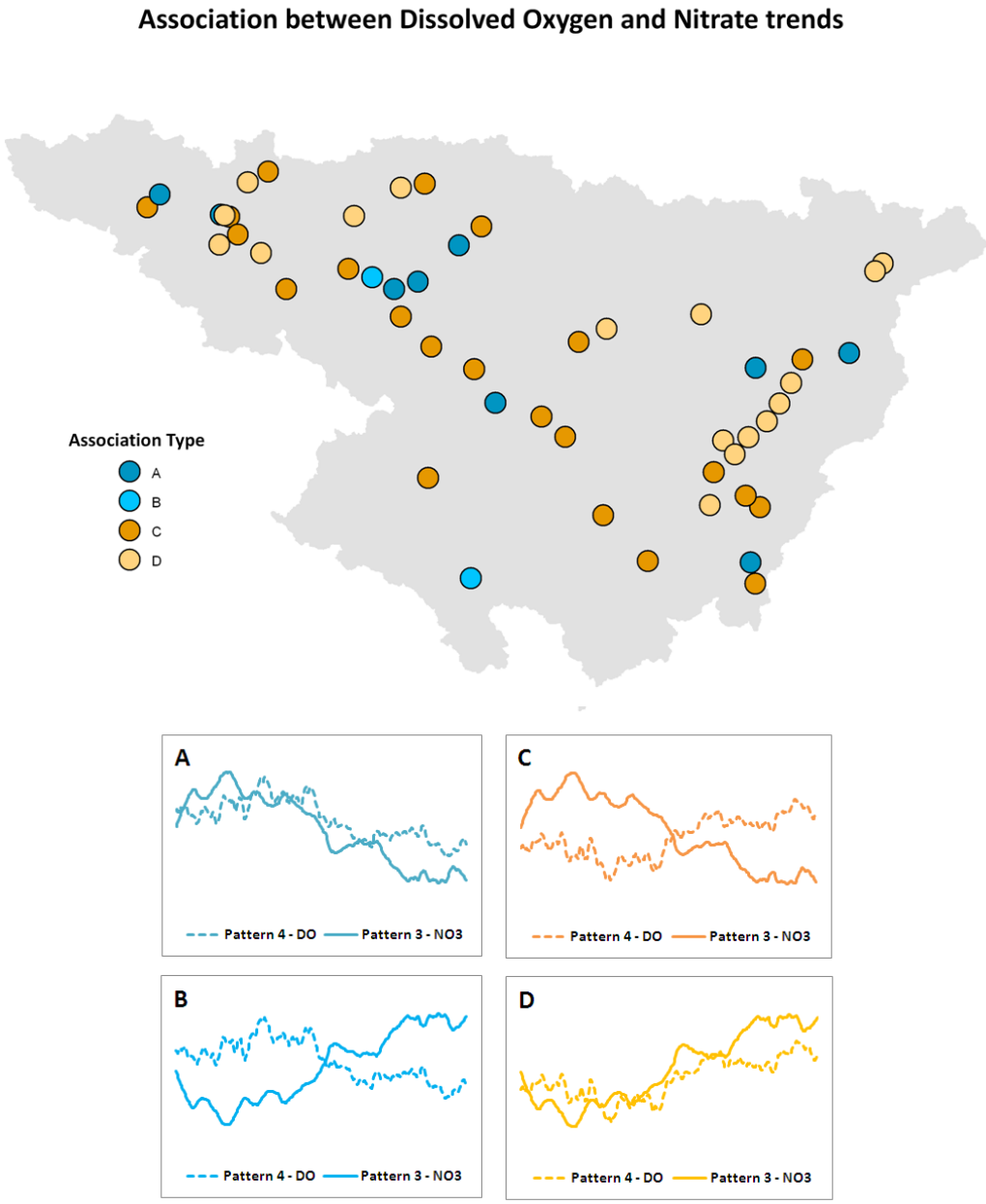


Figure 8: Association types for NO_3^- Pattern 3 and DO Pattern 4, both characterized mainly as trends, according to the sign of their corresponding factor loadings and their geographical location in the Ebro River basin.

Discussion

The case studies above illustrate the capabilities and the flexibility of the individual methods as well as of the collection of methodological steps intended to detect and characterize water quality patterns at basin and regional scales. Contrary to the complete set of methodological steps introduced in the Llobregat basin case study, in the Júcar RBD, MINE was used to find significant nitrate associations with time within a set of 90 time-series and to provide a preliminary description of the existing temporal patterns. DFA was then implemented only on the 50 monitoring points with highest significant MIC scores. Conversely, DFA was applied to all time-series available at the Ebro River Basin in order to extract common patterns for different sets of water quality variables. Subsequently, MINE pinpointed the most significant associations between pairs of extracted patterns, with the association between nitrate and oxygen long-term trends being the most important (highest MIC score).

The description obtained by these methods can be further analyzed in order to identify the main causes and drivers behind the spatio-temporal variability of water quality patterns, which could be related, for instance, to climate forcing or land management, or a combination of the two. Common patterns can be checked for cycles by time-frequency analysis to find any association to climate forcing factors such as global and regional climatic oscillations. In terms of land practices and their influence on river water quality signals, land-use type areal percentages could be used as an indicator.

Being able to identify interesting relationships between pairs of variables in large datasets has increasingly gained importance within scientific studies (Reshef et al. 2011). The measure of dependence for two-variable relationships, i.e., the Maximal Information Coefficient (MIC), provides a means to elucidate significant associations among variable pairs of river basin monitoring datasets, without applying any gap-filling or data pre-treatment procedures. The remaining MINE statistics aided in characterizing the functional associations based on scores expressing linearity and monotonicity. Thus, as a preliminary step for the exploration and detection of patterns and in absence of a clear a priori hypothesis, the MINE algorithm is very convenient

tool to pinpoint 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, such as those involving spectral analysis, are able to readily overcome those two problems (Zuur et al. 2003b). In an ecological context, shortcomings of the above and other similar techniques include the irregular or uneven sampling resolution and the lack of a multivariate approach (Cazelles et al. 2008). Dynamic Factor Analysis can easily deal with short-term, gappy time-series while providing a framework for detecting common patterns in a set of 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 (currently 3.9; 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 spatio-temporal 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 presented in this work 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). The covariate data are assumed to have no error and cannot contain missing values, though there is a way to circumvent these problems (Holmes et al. 2013). Furthermore, other alternatives could be used instead of the specific tools and complementary methods herein proposed for temporal and spatial analyses. 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. The same applies to spatial clustering and regression methods.

Conclusions

The set of methods presented above deal mainly with the detection of changes in water quality and the characterization of their variability across time and space within a region or river basin. 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 has allowed 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.

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, economic activities, resource use, and population dynamics arise at a local scale (Wilbanks and Kates 1999). The proposed methodological framework addresses the complexity of interacting processes found at different scales in global change phenomena. All in all, these methods offer the possibility of extracting common water quality signatures from a set of noisy and gappy time-series, providing not only a temporal interpretation of variability but also allowing researchers to include the spatial variability across basin and regional scales.

Chapter 4

Detection and attribution of global change effects on river nutrient dynamics in a large Mediterranean basin

State of the Science

The Earth system is intrinsically dynamic but the intensity and rate of recent environmental changes are overall unprecedented (Meybeck, 2003; García-Ruiz et al., 2011). Land-use change and management practices, pollution, human demography shifts, and climate change are components of global environmental change (Rosenzweig et al., 2008). Freshwaters are at the forefront of the phenomena associated to global change (Vörösmarty et al., 2010), and impacts on water resources availability as well as on their quality are extensive (Parmesan and Yohe, 2003; Milly et al., 2005; Grimm et al., 2008; Rabalais et al., 2009; Gallart et al., 2011).

A fundamental concern is to understand the spatial patterns of nutrient concentration and loading in rivers, their variation during the last decades, and whether these are promoted by the increasing human activities (Grizzetti et al., 2011), or may be associated to climate change (Marcé et al., 2010). This is particularly relevant in Mediterranean regions where the imbalance between available water resources and increased demands has become a growing problem (Milly et al., 2005; Bovolo et al., 2011), and where streams and rivers bear concurrent additional pressures such as damming, water extraction, and urbanization (Sabater and Tockner, 2010).

However, it is challenging to attribute changes in nutrient concentration dynamics to specific factors, because factors of change exist and act at different temporal and spatial scales (Kundzewicz and Krysanova, 2010). Identifying factors and causes often requires examining long-term series of data, which should be consistent and of good quality. Such detailed analysis of time-series is essential to obtain insight of the physical, biological, or socioeconomical systems that produce them (Ghil et al., 2002). It is realistic to consider that temporal trends and spatial patterns reveal emerging environmental problems (Lane et al., 1994; Lovett et al., 2007; Marcé et al., 2010; Estrada et al., 2013). Data consistency can be however affected by the lack of observations of relevant water quality variables and the low or uneven sampling frequency, which are common characteristics of many water quality monitoring

schemes worldwide. These impede the appropriate analysis of the time-series available from long-term monitoring, eventually affecting management decisions on the minimization of effects of global change, and particularly in Mediterranean regions, where there is a dearth of knowledge compared to other temperate regions (Benítez-Gilabert et al., 2010).

The vast majority of studies of global change impact based on the analysis of long-term data use time-series methods like the Mann-Kendall and the Seasonal Kendall analyses for trend detection (Chang, 2008; Bouza-Deaño et al., 2008; Argerich et al., 2013); wavelet analysis for temporal patterns (Kang and Lin, 2007); and combinations of statistical models such as univariate and multivariate regressions (Tilman et al., 2001); and analysis of variance and variography (Bernal et al., 2013). However, methods like spectral analysis are limited to characterizing the spectral density to detect any periodicities in the data and do not necessarily allow the identification of common patterns embedded in a collection of time-series (Zuur et al., 2003). Furthermore, most of these methods do not easily accommodate missing observations, which are extremely abundant in most public environmental databases. These limitations make the analysis of water quality datasets in large regions difficult and cumbersome, compromising the spatial and temporal coverage of the analyses, while in most occasions the impact of global change on a given ecosystem consists in the overlap of multiple stressors acting at both regional and local scales. Therefore, it is necessary using methodologies that explicitly consider the inextricable link between temporal and spatial patterns of change and that are able to accommodate missing values.

We use a combination of Dynamic Factor Analysis (DFA), classical time-series methods, and spatial regression models to extract underlying common patterns in a set of time-series and to depict their relationships with local and global scale phenomena. We apply the above to a set of river nutrient concentration time-series within a Mediterranean basin in order to identify temporal and spatial patterns at the basin-wide scale, and to understand how global change shapes these patterns. Both nitrate

and dissolved phosphate dynamics were analyzed in order to disentangle the role of hydrology, land-use practices, and climate phenomena on the observed patterns, with the final aim of understanding how the different aspects of global change may affect nutrient variability (and hence water quality) in the basin.

Study Area

The Ebro River is one of the main tributaries of the Mediterranean Sea. The mean annual runoff at the outlet is 13,408 hm3. The basin covers a highly heterogeneous area of ca. 85,500 km2, which extends from the southern-facing side of the Cantabrian range and Pyrenees and the northern-facing side of the Iberian Massif until the river reaches the Mediterranean Sea (Sabater et al., 2009). The geographical setting of the Ebro River determines a large range of climatic conditions (Sabater et al., 2009). Mean annual precipitation varies from over 2000 mm in the Pyrenees to less than 400 mm in the arid interior. Overall, silicic materials are located in the uppermost altitudes while calcareous materials occur at lower elevations (Lassaletta et al., 2009). The water biogeochemical characteristics are highly influenced by anthropogenic activities. The main effects are those due to water discharge regulation (i.e., the construction of large reservoirs) and agriculture (determining increases in nitrate concentration; Romaní et al., 2010). The intense use of water throughout the basin (Boithias et al., 2014) puts the Ebro River under strong pressure particularly in the most downstream sections during dry annual periods, when irrigation is widespread. The basin started a sanitation plan during the 90s that progressively covered most of the local inputs.

Methods

Time-series data

Existing nitrate and phosphate concentration (analyzed based on standard procedures; APHA, 1995) as well as water discharge data of the Ebro River Basin were collected from public databases (Ebro Basin Authority - CHE). The frequency of sampling was monthly. We selected 50 monitoring points distributed all across the basin that

showed the longest time-series, consisting in 31-year-long (1980-2011) monthly data. Thus, these time-series had a maximum length of 372 data points, although most of the stations contained observation gaps. Some unreasonable values were manually removed considering expected ranges of values for each nutrient. Discharge time-series were available in 37 of the sampling sites.

Detection and attribution of global change effects

The first step in defining global change effects on nutrient time-series was to detect common patterns in the 50 nutrient time series (nitrate or phosphate) using DFA (Zuur et al., 2003). Once the common patterns for nitrate and phosphate were identified, we described the significant cycles and trends present in those patterns with classical frequency and trend analyses. Subsequently, the potential dependence on hydrological variability was sought by exploring any significant association between patterns and water discharge time-series. We finally assessed the spatial variability of these patterns and their relationship to environmental change drivers in the region by means of spatial regression models and clustering.

Extraction of common nutrient concentration patterns

Dynamic Factor Analysis (DFA; Zuur et al., 2003) is a dimension-reduction method that estimates underlying common patterns in a set of time-series. It is based in the so-called state-space model, which treats each observed time-series as a linear combination of multiple state processes (Holmes et al., 2012). A considerable advantage of the state-space approach is the ease with which missing observations can be dealt with. The main disadvantage of DFA is that it can be computationally expensive.

DFA decomposes the observed time-series from all sampling points included in the analysis into common patterns and their associated error terms (Holmes et al., 2012). The resulting patterns are in turn related to factor loadings, which indicate the weight that each pattern has at every monitoring point included in the analysis. In other

words, DFA models the different time-series as a linear combination of common temporal patterns, in a similar way a Principal Component Analysis reduces an n-dimensional problem into a few manageable axes. Both the identified common patterns and their relevance at each sampling point (i.e., the factor loadings) were subsequently analyzed using additional time-series and regression techniques. DFA was applied to our database by means of the MARSS v3.4 R-package (Holmes et al., 2013).

We also used DFA to enhance the signal to noise ratio of the measured streamflow time-series which in turn facilitated the identification of characteristic oscillations and potential relationships between streamflow and other variables. After DFA, we reconstructed the streamflow time-series at each sampling point using the best linear combination of the common patterns identified during DFA. This procedure is equivalent to other signal to noise ratio enhancement methods, like reconstruction using Singular Spectrum Analysis (Ghil et al., 2002), with the difference that our approach enhances the features shared by the different time-series.

Identification of significant oscillations in the common nutrient concentration patterns

We analyzed all significant frequencies present in the common patterns identified by DFA using frequency analysis. We specifically aimed to identify frequencies that could be linked to intra-annual and seasonal cycles (6 and 12 months period) and climatic interannual oscillations. We chose the Multitaper Method (MTM) due to its reduced variance of spectral estimates compared to classical methods (Ghil et al., 2002). Frequencies significantly different from noise at the $p<0.05$ level were identified using the F test for spectral frequencies. MTM was applied using the Multitaper R-Package (Rahim and Burr, 2013).

Identification of significant temporal trends in the common nutrient concentration patterns

Since the common patterns are allowed to be stochastic in DFA, they can also contain significant trends that are not necessarily a straight line (Zuur et al., 2007). We therefore sought to identify the significant trends present in individual patterns and to characterize such trends as increasing or decreasing over time. We used the implementation of the Yue-Pilon's (Yue et al., 2002) prewhitening approach included in the zyp R-package (Rahim and Burr, 2013) to determine the trends in data that are serially correlated. The method computes both the Kendall's tau statistic and the Kendall's p-value.

Relationships between streamflow and the common nutrient concentration patterns

The relationships between streamflow and nitrate and phosphate concentration patterns from the DFA analysis were assessed with the Maximal Information Coefficient (MIC) method (Reshef et al., 2011). MIC captures a wide range of associations which are not restricted to be linear, and without the need to define a model a priori. MIC provides a score that roughly corresponds to the coefficient of determination of the data relative to the regression function, and a significance level. In our case, we calculated MIC scores and significance levels for each paired combination of common nutrient concentration patterns and the DFA reconstructed streamflow series measured at each sampling station. We used these filtered streamflow time-series instead of the original ones due to the continuity of the resulting filtered series and in order to enhance the signal to noise ratio.

Attribution of drivers for spatio-temporal variabilility of the common nutrient concentration patterns

Factor loadings are the multiplication factors that determine the linear combination of the common patterns to produce a best-fit nutrient concentration time-series (Zuur et al., 2003). Factor loadings can take positive or negative values when specific time-

series behave in an opposite way to that described by the extracted pattern. Therefore, the geographical distribution of factor loading values across monitoring points inform about the spatial development of the processes responsible for the extracted patterns.

To evaluate the relationship between the relevance (i.e., factor loading) of the extracted patterns at each sampling point and the environmental change drivers, we selected a set of potential explanatory variables that were spatially distributed. These included meteorological variables (mean annual air temperature and precipitation), reservoir capacity and location, wastewater treatment plants (WWTP) discharge and location, specific streamflow (runoff index), mean river nutrient concentration in the sampling point, land use distribution, and five variables related to nitrogen loads and their sources obtained by (Lassaletta et al., 2012): application of synthetic fertilizers, application of manure, inputs by biological fixation, total exported N, and point sources. For each sampling point we calculated mean values or percent areas considering two different regions: the total basin upstream from the sampling point, and a buffer area of 10 km surrounding the point, aimed at capturing the more local conditions.

The potential explanatory variables were related to factor loadings measured at each sampling site by the Generalized Least Squares (gls) regression model (Pinheiro and Bates, 2000). The use of generalized least squares for regression modeling is advisable when neighboring values of the response variable tend to be spatially correlated. In our case, we assumed a spatial error structure, the Gaussian structure being the most appropriate among the (six) available options based on Akaike Information Criterion (AIC) values. A combination of forward and backward selection was used to identify the significant explanatory variables, using the AIC criterion to identify the best model. We fitted different gls models for sampling stations showing opposite signs of the factor loading for a given Pattern (e.g., stations showing positive and negative factor loadings for Pattern 1 of nitrate concentration were treated separately). The rationale of this procedure is that many fundamental features of the patterns (phase of the time-series, relationships with streamflow and other variables, direction of the

trends) change when the pattern is flipped due to a change of the factor loading sign, potentially implying different generating mechanisms. The gls models were fitted using the nlme R-Package (Pinheiro et al., 2012). In order to assess the model fit and the variance explained, we calculated a Generalized R-Squared based on (Cox and Snell, 1989) using the r.squaredLR function included in the MuMIn R-Package (Barton, 2014).

Spatial aggregation of common nutrient concentration patterns and explanatory variables

We assessed the clustering of the spatial distribution of nutrient concentration patterns and the significant explanatory variables found in gls regression models. We used the clustering analysis to portray homogeneous regions in terms of the presence of concrete nutrient concentration patterns and their likely drivers. Our final aim was to highlight the most relevant cause-effect mechanisms that define vulnerable regions to the effects of global change. We used the implementation of the unsupervised k-Means algorithm in Orange 2.7, which uses the between-cluster-distances score to assess the most effective grouping. The method looks for a solution where all the features (in our case, the value of all factor loadings and significant explanatory variables found during gls modeling) within each group are as similar as possible, and all the groups themselves are as different as possible. Thus, it is not necessary to define the number of desired cluster beforehand. We applied the k-means algorithm without any spatial constraints. Although explicit spatial relationships actually exist between sampling points along a river network, our aim was to identify clusters exclusively based on the information contained in the factor loadings and explanatory variables.

Results

Common nutrient concentration patterns in the basin

The DFA analysis for nitrate concentration extracted 3 common patterns (Fig. 1a), where the order of the extracted patterns has no implication on the importance or

weight of a particular pattern. Patterns 1 and 2 identified in nitrate time-series had a marked seasonal component appreciated visually (Fig. 1a) and further confirmed by the significant 12-month cycles found in the frequency analysis (Table 1). The seasonal evolution of Pattern 1 was clearly associated with the seasonal streamflow pattern (Fig. 1e), suggesting that it was hydrology-driven. The MINE analysis also detected significant associations between Pattern 1 of nitrate concentration and the DFA reconstructed streamflow series in almost all sites across the basin (Table 1). Nitrate concentration increased with streamflow (sites showing positive factor loadings), and was affected by a dilution dynamics (negative factor loadings). In contrast, Pattern 2 was strongly associated with the seasonal evolution of the mean air temperature in the basin (Fig. 1f), suggesting its connection to phenological processes (lower values during the growing season).

Pattern 1 of nitrate concentration was also associated to a ca. 2.6 year periodicity according to the MTM analysis, and Pattern 3 showed a significant 3.5 yr oscillation period (Table 1). Pattern 3 also included a significant decreasing trend (Table 1). The signs associated to DFA factor loadings of Pattern 3 indicated that 20 of the 50 stations were in fact following the opposite trend.

DFA extracted four common patterns from the 50 dissolved phosphate concentration time-series included in the analysis (Fig. 1b). The 1990s represented a shift-time point for phosphate patterns. In all four patterns, a sharp decrease in the phosphate concentration occurred in the early 1990s, and shifted to a steady behavior till the end of the study period, but the four patterns differed in peak timing before the 1990s. Despite the overall decrease, only Pattern 2 had a highly significant trend while trend in Pattern 4 was only marginally significant (Table 1). Pattern 1 had a marked seasonal cycle, potentially driven by streamflow (suggested by the significant relationship between the seasonal evolution of the pattern and streamflow; Fig. 1b). However, the MINE algorithm detected just 2 significant associations between this pattern and the DFA reconstructed streamflow time-series from the sampling sites (Table 1). Pattern 3 showed cycles of ca. 4.3 and 1.6 yr (Table 1).

The frequency analysis of the 37 DFA reconstructed streamflow series revealed several characteristic oscillations. Apart from the strong seasonal signal, there were significant oscillations at 1.5, 2.2, 3.2, 4.2, and 9 years in several sampling stations. Periods from 1.5 to 4.2 years were highly coherent with the oscillations found in the common patterns of nitrate and phosphate concentration (Table 1), suggesting that multi-year oscillations in nutrients concentration were related to streamflow variability. Interestingly, nitrate and phosphate patterns showing at least one significant oscillation with period longer than one year also showed many significant MINE associations with streamflow across sites (Table 1). No significant trend was detected in the streamflow series.

Table 1: Characterization of the temporal variability and relationships with streamflow of nutrient patterns detected with DFA in the Ebro basin.

Nutrient	Trend (Kendall tau)	Significant oscillations (years)	Significant MINE relationships with streamflow (number of sites out of 37)	Other relationships with streamflow
Nitrate				
Pattern 1	ns	1 and 2.6	34	• Strong seasonal coherence (Fig. 1e) • Coincident and significant streamflow oscillation at 2.2
Pattern 2	ns	1	12	• Nothing to remark
Pattern 3	-0.53***	3.5	22	• Trend NOT related to streamflow • Coincident and significant streamflow oscillation at 3.2 years
Phosphate				
Pattern 1	ns	1	2	• Moderate seasonal coherence (Fig. 1g)
Pattern 2	-0.09**	ns	4	• Trend NOT related to streamflow
Pattern 3	ns	1.6 and 4.3	25	• Coincident and significant streamflow oscillations at 1.5 and 4.2 years
Pattern 4	-0.08*	ns	10	• Trend NOT related to streamflow

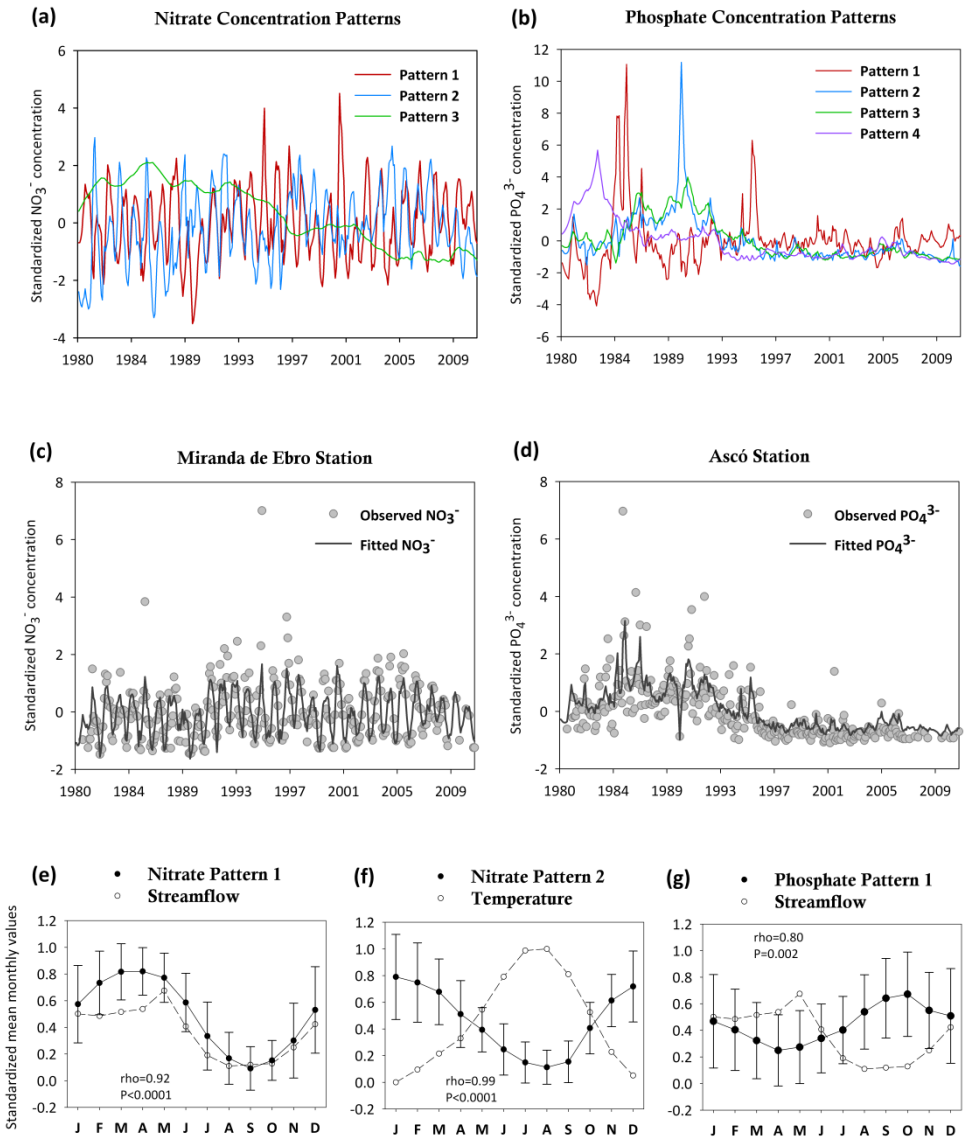


Figure 1: *Top:* DFA resulting patterns for nitrate (a) and phosphate (b) concentration. *Center:* Examples of observed time-series and fitted DFA models at two selected monitoring points for nitrate (c) and phosphate (d) concentration. The DFA models in panels (c) and (d) are the result of a linear combination of the patterns in panels (a) and (b), respectively. *Bottom:* Seasonal variation for nitrate Pattern 1 and streamflow (e), nitrate Pattern 2 and Temperature (f), and phosphate Pattern 1 and streamflow (g). Points depict monthly averages for the entire 31 year time-series. For temperature and streamflow, the average is for all time-series available. We only included standard deviations as error bars for the nutrient patterns to enhance readability.

Factors explaining the distribution of nutrient concentration patterns

The gls regression models for the distribution of factor loadings for each Pattern identified several significant explanatory variables (Tables 2 and 3). Since nitrate concentration Patterns 1 and 2 showed contrasted positive and negative factor loadings across sites, we considered different models for sites showing positive and negative factor loadings. The distribution of positive factor loadings for Pattern 1 strongly related to the total area of water (mainly reservoirs). The higher the total area occupied by water upstream, the higher the weight of Pattern 1.

Other associations were also significant, although their prediction weights on the model were less important: a negative relationship with mean annual air temperature upstream from the sampling point; a positive relationship with dryland farming area around the sampling point; and a negative association with the industrial areas upstream from the sampling point (Table 2). Negative factor loadings of Pattern 1 were related to the presence of irrigated agricultural lands and to the mean annual precipitation received upstream. The reservoir water capacity upstream the sampling point had a small and marginally significant effect.

Factor loadings for Pattern 2 of nitrate were strongly associated to sites with irrigated agricultural areas upstream from the sampling point. The distribution of Pattern 2 was also weakly related to the annual mean precipitation and the presence of irrigated lands. Finally, the distribution of factor loading values for Pattern 3 was spatially associated to industrial areas. The main difference between models for negative and positive factor loadings for this pattern was dictated by the relevance of distinct sources of nitrogen being used in the area, namely, synthetic fertilizers and manure (Table 2).

Globally, the explanatory power of the gls models for the distribution of phosphate patterns was much lower than for nitrate concentration models (Table 3). Pseudo-R2 values were one third of those found in nitrate models, except for Pattern 1 that

reached similar explanatory power. The distribution of the factor loadings of Pattern 1 was explained by a complex combination of synthetic fertilizer load and industrial area upstream from the sampling point, the runoff index associated to it, and the mean river phosphate concentration in the site. Overall, the distribution of the phosphate patterns was hardly explained by the set of explanatory variables considered in this study, and was mainly explained by the presence of industrial areas upstream of the sampling points (Table 3).

Table 2: GLS resulting potential drivers involved in the spatio-temporal variability of nitrate patterns in the Ebro basin.

Nitrate Patterns	Pseudo R^2	Explanatory Variable	Coefficient	Std. Error	t-value	p-value
Pattern 1 - Positive Factor Loadings	0.65	Mean Air Temperature (°C) UPSTREAM	-1.42	0.30	-4.66	0.0001
		Water area (km^2) UPSTREAM	0.06	0.00	13.75	0.0000
		Dryland Farming (%) LOCAL	0.00	0.00	3.23	0.0035
		Industrial area (%) UPSTREAM	-0.12	0.04	-2.91	0.0074
Pattern 1 - Negative Factor Loadings	0.61	Reservoir Capacity (hm^3) LOCAL	-0.05	0.02	-2.64	0.0166
		Irrigated agriculture area (%)UPSTREAM	0.30	0.05	6.37	0.0000
		Mean Annual Precipitation (m) UPSTREAM	0.48	0.11	4.42	0.0003
Pattern 2	0.59	Irrigated agriculture area (km^2) UPSTREAM	0.11	0.01	19.06	0.0000
		Irrigated agriculture area (%) LOCAL	0.00	0.00	-2.59	0.0127
		Mean Daily Precipitation (m) LOCAL	0.16	0.07	2.29	0.0269
Pattern 3 - Positive Factor Loadings	0.57	Industrial area (%) UPSTREAM	0.04	0.01	6.53	0.0000
		Synthetic Fertilizer Load UPSTREAM	-0.01	0.00	-3.45	0.0018
Pattern 3 - Negative Factor Loadings	0.56	Industrial area (%) UPSTREAM	0.04	0.01	4.81	0.0001
		Areal Manure Load UPSTREAM	0.04	0.01	3.01	0.0063
		Water area (%) UPSTREAM	0.01	0.01	2.14	0.0428

Table 3: GLS resulting potential drivers explaining the spatio-temporal variability of phosphate patterns in the Ebro basin.

Phosphate Patterns	Pseudo R^2	Explanatory Variable	Coefficient	Std. Error	t-value	p-value
Pattern 1	0.62	Synthetic Fertilizer Load UPSTREAM	0.46	0.08	6.18	0.0000
		Mean river phosphate concentration	-0.07	0.02	-3.97	0.0003
		Runoff Index UPSTREAM	-0.03	0.01	-3.69	0.0006
		Industrial area (%) UPSTREAM	0.19	0.05	4.02	0.0002
Pattern 2 - Positive Factor Loadings	0.20	Industrial area (km^2) UPSTREAM	0.03	0.02	2.22	0.0384
Pattern 2 - Negative Factor Loadings	0.17	Grass and shrubland area (%) LOCAL	0.01	0.00	2.24	0.0339
Pattern 3	0.21	Industrial area (km^2) UPSTREAM	0.05	0.01	3.60	0.0008
Pattern 4	0.14	Industrial area (%) UPSTREAM	0.05	0.02	2.75	0.0083

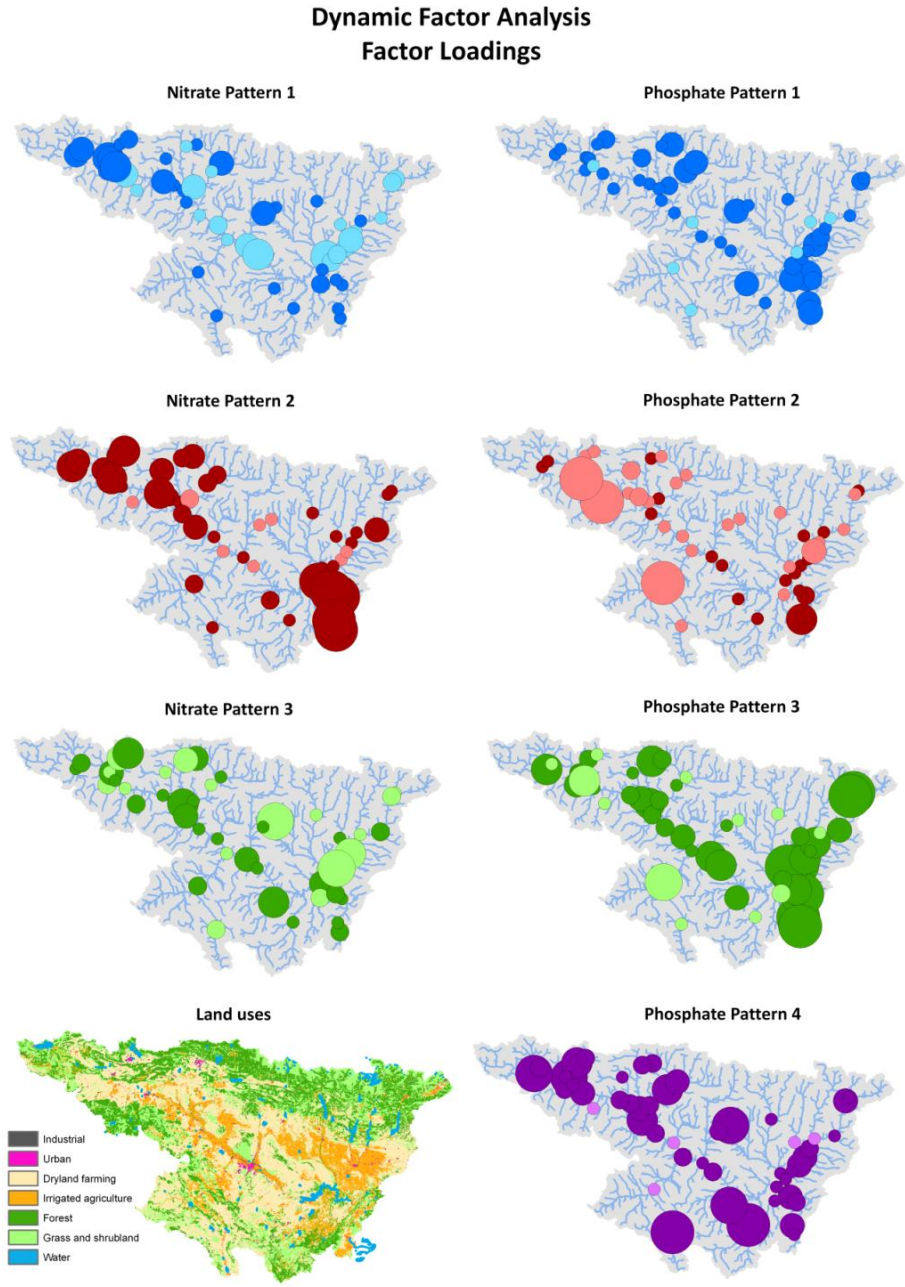


Figure 2: Factor loadings associated to nitrate patterns (left column) and phosphate patterns (right column). Dark circles indicate positive factor loadings and light-colored circles represent negative factor loadings. The size of the circles represents the magnitude of the Factor Loading at each monitoring point. A map with the main land uses in the Ebro basin, based on the CORINE Land Cover dataset, is enclosed in the lower left corner.

The joint spatial distribution of the nutrient concentration patterns and explanatory factors

The clustering analysis for the spatial distribution of the nitrate patterns and the significant explanatory variables found 4 aggregations among the 50 sampling sites (Fig. 3a). Cluster 1 contained sampling points located mainly in downstream sections of major tributaries of the Ebro River (particularly along the Segre River); Cluster 2 included points in upstream locations of tributaries and in the main Ebro; Cluster 3 comprised points located even more upstream; and Cluster 4 collected the downstream sites of the main stem of the Ebro River. These clusters were characterized by significant differences in the absolute values of the factor loadings for Pattern 1 (Fig. 3b, non parametric Wilcoxon test for mean comparison, $p=0.011$), and Pattern 2 ($p=0.017$). Cluster 1 showed the largest relevance of Pattern 1, Cluster 4 for Pattern 2, and Cluster 3 for Pattern 3. Therefore the most fundamental regional difference in the dynamics of nitrate concentration in the basin was a switch from a streamflow-dominated dynamics in Cluster 1 to a phenology-dominated of Cluster 4. The preeminence of Pattern 3 in Cluster 3 was also a significant spatial pattern extracted from the clustering analysis. These differences between clustering groups were coincident with significant differences for many explanatory variables, particularly the extension of irrigated agriculture ($p<0.0001$), the presence of reservoirs upstream the sampling point ($p<0.0001$), and the application of synthetic fertilizers ($p<0.0001$). Cluster 3 showed the minimum values for these variables, followed by Cluster 2 and Cluster 1, whereas Cluster 4 showed the largest values.

Contrastingly, the clustering analysis for the phosphate concentration resulted in a poor regionalization with only 2 different aggregations (Fig. 4a), one including just 5 sampling points. There were no obvious spatial clusters beyond Cluster 2, which included points with higher values for Pattern 4 ($p=0.006$). This coincided with very high phosphate concentrations ($p=0.002$) and extensive industrial areas ($p=0.001$) related to the sampling points. The poor regionalization in the phosphate case stressed again the apparently idiosyncratic behavior of phosphate concentration across sampling sites.

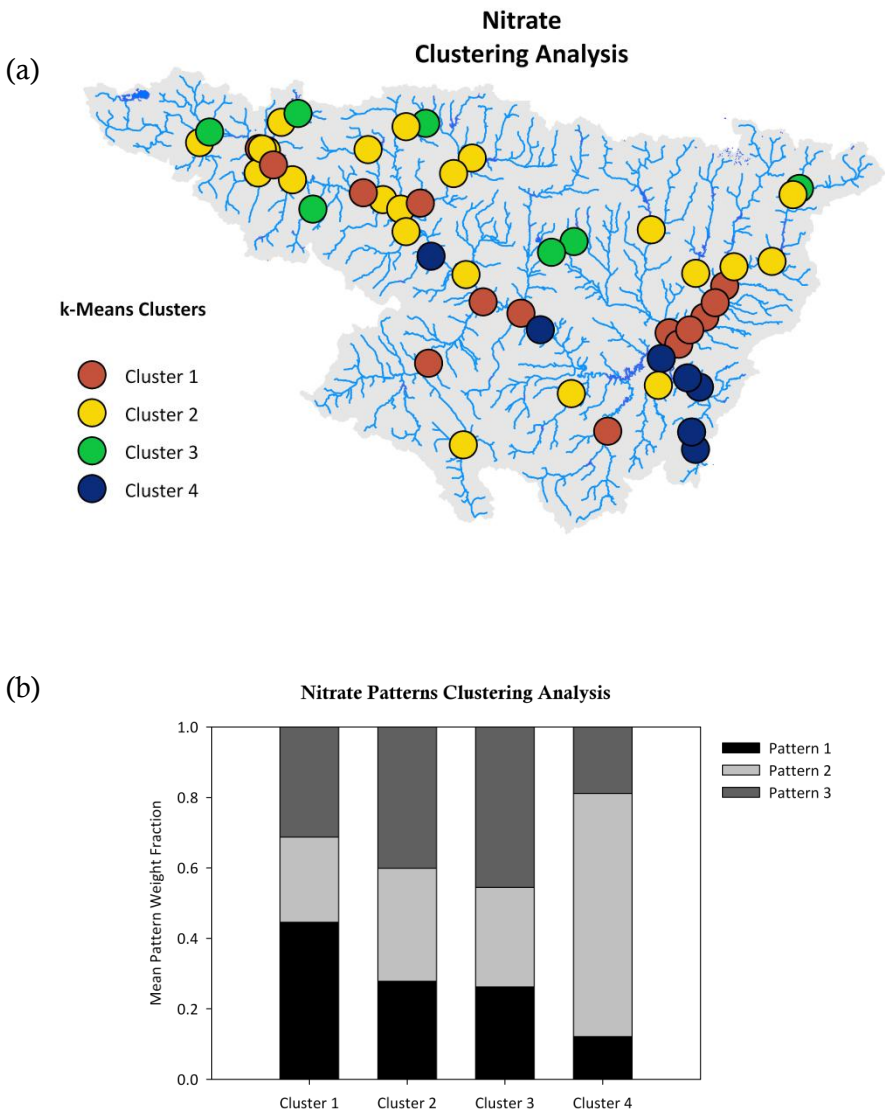


Figure 3: (a) Clustering analysis results for the spatial distribution of nitrate concentration patterns and associated explanatory variables. (b) Mean fraction of Factor Loadings (i.e., the overall weight of a specific pattern) found in each of the 4 clusters identified in the analysis.

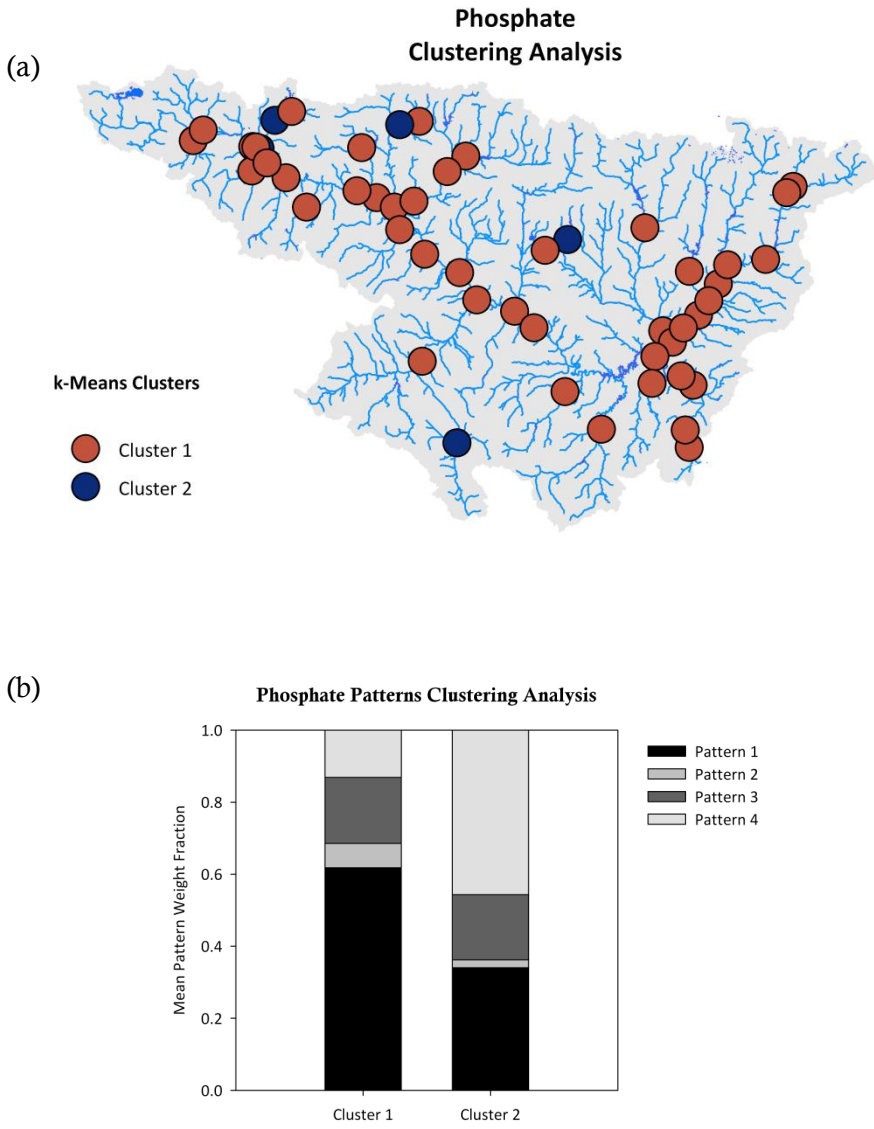


Figure 4: (a) Clustering analysis results for the spatial distribution of phosphate concentration patterns and associated explanatory variables. (b) Mean fraction of Factor Loadings (i.e., the overall weight of a specific pattern) found in each of the 4 clusters identified in the analysis.

Discussion

The nature of nutrient concentration patterns in the Ebro basin

The analysis of the impacts of global change on freshwater ecosystems requires the use of appropriate tools to identify the main regional trends and modes present in hydrological and water quality variables. Results of this study show that the combination of DFA, traditional time-series analysis, and regression methods is a convenient approach and several features of the time series shared by many sampling points across the Ebro basin can be detected.

The analysis of the nutrient concentration time-series detected the existence of seasonal patterns related to hydrology. Although the common hydrological relation with nutrient dynamics (Donner et al., 2002) may hide the detection of other seasonal cycles not related to streamflow, our analysis also detected seasonality that was unrelated to hydrology. While Pattern 1 of nitrate concentration was related to streamflow, the nitrate dynamics in the basin was also related to the phenological cycles of the adjacent terrestrial ecosystems or other water bodies upstream (Pattern 2). The actual mechanism behind the association between nitrate concentration and air temperature may be complex, and in fact it may differ at different sampling points, since air temperature can co-vary with many other factors. In the case of nitrate concentration, assimilation by freshwater primary producers during summer (Carpenter and Dunham, 1985) and the seasonal evolution of leaf fall and decomposition (González et al., 2012) could have taken a major role. However, other factors may contribute to lower concentrations, like the seasonal cycle of denitrification in the adjacent terrestrial ecosystems and upstream water bodies during summer months (Tatariw et al., 2013).

Nutrient concentrations showed multiple associations with streamflow spanning from the seasonal to the interannual scale. One of the most prominent features of nitrate concentration time-series was the existence of a switching relationship with streamflow (expressed by the changing sign of factor loadings for Pattern 1). This

implies a fundamental change of the dynamics of nitrate concentration and suggests a major change in the sources of nitrogen to freshwaters. The positive relationship between nutrient concentration and streamflow suggest the preponderance of diffusive inputs from the terrestrial ecosystems and non-irrigated agricultural fields, whereas the negative relationship pointed to a dilution mechanism typical of locations having point sources. The gls models further identified the land fraction occupied by irrigated agriculture as the main factor associated to the presence of negative factors loadings for Pattern 1 of nitrate concentration. Summer irrigation is a common agricultural practice in Mediterranean areas that can disrupt the relationship with the natural flow regime as well as the nitrate dynamics. This has been already observed in the Ebro basin where the intra-annual N export differed among rainfed and irrigated crops, the former following the flow regime, the latter modifying it (Lassaletta et al., 2012). In addition, irrigation has the capability of altering local and regional precipitation behavior through changes in soil moisture and heat budgets (Boucher et al., 2004), particularly in downstream areas (Huber et al., 2014). However, none of these regional climate effects has been confirmed in the Ebro basin.

The absence of seasonal relationships between nitrate concentration and streamflow (i.e., very low absolute values for Pattern 1) can also be related to the proximity to large reservoirs in the lower section of the basin, where the seasonal nitrate concentration cycles seem to be highly influenced by the water released from the reservoirs.

The supra-annual frequencies detected in the nitrate and phosphate concentration patterns in the Ebro point out to associations with climatic oscillations identified in the Mediterranean region. The North Atlantic Oscillation (NAO) has multiple modes starting at 1.4 years, while the El Niño-Southern Oscillation (ENSO) has modes between 2.4 and 5.2 years (Rodó et al., 1997). Both the nutrient patterns and the streamflow series showed oscillations coherent with those from the ENSO and NAO, which are known to modify, through teleconnections, the magnitude and frequency of precipitation in a heterogeneous manner (Rodó et al., 1997). The impact of ENSO on

nitrate river concentrations is, in fact, not uncommon in areas under indirect ENSO effects, such as the SE United States (Keener et al., 2010). Moreover, the associations of ENSO with streamflow modifications (Marcé et al. 2010) and nitrate concentration dynamics (Vegas-Vilarrúbia et al., 2012) in the Iberian Peninsula have been unambiguously stated. Indeed, all nutrient concentration patterns showing significant supra-annual frequencies also showed significant relationships with streamflow in many sites across the basin. In our view this indicates that the effect of atmospheric teleconnections on nitrate and phosphate concentrations was driven by modifications in the streamflow. Since streamflow 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. Indeed, the relationship between the partially predictable global climate modes and the occurrence and frequency of extreme events is a very active topic in the literature (Coumou and Rahmstorf, 2012), and their links with water quality crisis episodes should be further investigated, especially in the Mediterranean region, where climate extreme events are predicted to increase (García-Ruiz et al., 2011).

Nutrient trends and local management practices

Nitrate concentration showed decreasing and increasing trends in areas across the basin. The association of the trends with sampling points affected by large loads of synthetic fertilizer and manure indicates that trends were possibly promoted 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). Also, the implementation of sewage treatment schemes in the basin can be partly invoked to justify this decrease (Romaní et al., 2010). The dominant role of nitrate concentration trends in the more upstream locations of the basin (mostly included in Cluster 3) suggest that the impact of human activities upstream sampling points were higher in streams, and that these water courses were the most vulnerable to increasing trends. On the other hand, decreasing trends also dominated the time-series in some of the sampling points included in Cluster 3, suggesting that upstream locations are also

prone to improvement due to remediation measures and best management practices. Particularly, our analysis suggests that the application of synthetic fertilizers precluded the existence of a decreasing trend in some areas of the basin, but the application of manure as a fertilizer actively promoted increasing nitrate concentration trends. While during the last decades manure application has dramatically grown in some specific areas (Terrado et al., 2010), there has been a more rational application of synthetic fertilizers in the basin (Lassaletta et al., 2012).

The overall decrease of phosphate concentration in the Ebro basin since the early 1990s was highlighted by all four extracted patterns. This decreasing trend coincides with the improvement of urban sewage treatment in the most important cities of the Ebro basin (Ibáñez et al., 2008), since most patterns of phosphate dynamics derive from point sources. The reduction of phosphate fertilizers in the agriculture could have also resulted in the reduction of phosphate loads exported to rivers and streams (Bouza-Deaño et al., 2008). A similar pattern has been observed in the Loire River (France), where the wastewater treatment plants and the concurrent ban on phosphorus content in washing powders (Floury et al., 2012) were highly effective. Severe reductions of riverine phosphorus loads were common in Europe during the 1990s, while nitrate concentrations decrease has been limited to recent years (Ludwig et al., 2009).

Overall, the significant trends identified in nitrate and phosphate concentration across the Ebro basin appear to be modulated by local management practices associated to the different anthropogenic activities that have co-existed in the basin during the study period, but no climatic factor seemed to play any relevant role in shaping decreasing or increasing trends of nutrient concentration.

Conclusion

The results presented in this chapter imply that the impact of global change on the dynamics of nitrate concentration across the Ebro basin is a multifaceted process that includes regional and global factors while impacts on phosphate concentration depend more on local impacts and less on large-scale factors (Figure 5).

In the case of nitrate, our analyses have identified the presence of irrigated agriculture and its corresponding fertilization management practices (synthetic fertilizers or manure), the presence of industrial activities in the basin, and damming as the main global change factors. Other climatic processes linked to streamflow variability were also identified, but the impact of climate changes on these processes is uncertain and could not be disentangled in this study. These factors shape a complex dynamics including temporal trends, and interannual and seasonal cycles, with either strong or vanishing relationships with streamflow, and links with phenological processes in terrestrial ecosystems and reservoirs. Interestingly, the impact of identified factors on this rich dynamics was not homogenous across the basin, but clustered in 4 regions not entirely coherent from a geographic perspective (Figure 3).

In contrast, phosphate concentration showed a more idiosyncratic behavior. The only relevant global change mechanism acting at large scales is the presence of industrial activities and the application of synthetic fertilizers, which defines higher phosphate concentrations in Cluster 2. The explanatory power of our models was low in the case of phosphate concentration dynamics, meaning that most variability was accounted by factors not considered in our models. Although these factors may include some relevant regional drivers, the contrasting results from the nitrate analysis imply that the ultimate reason of the lower performance of the phosphate models is the absence of the more local factors, such as the different timing of implementation of wastewater treatment technologies.

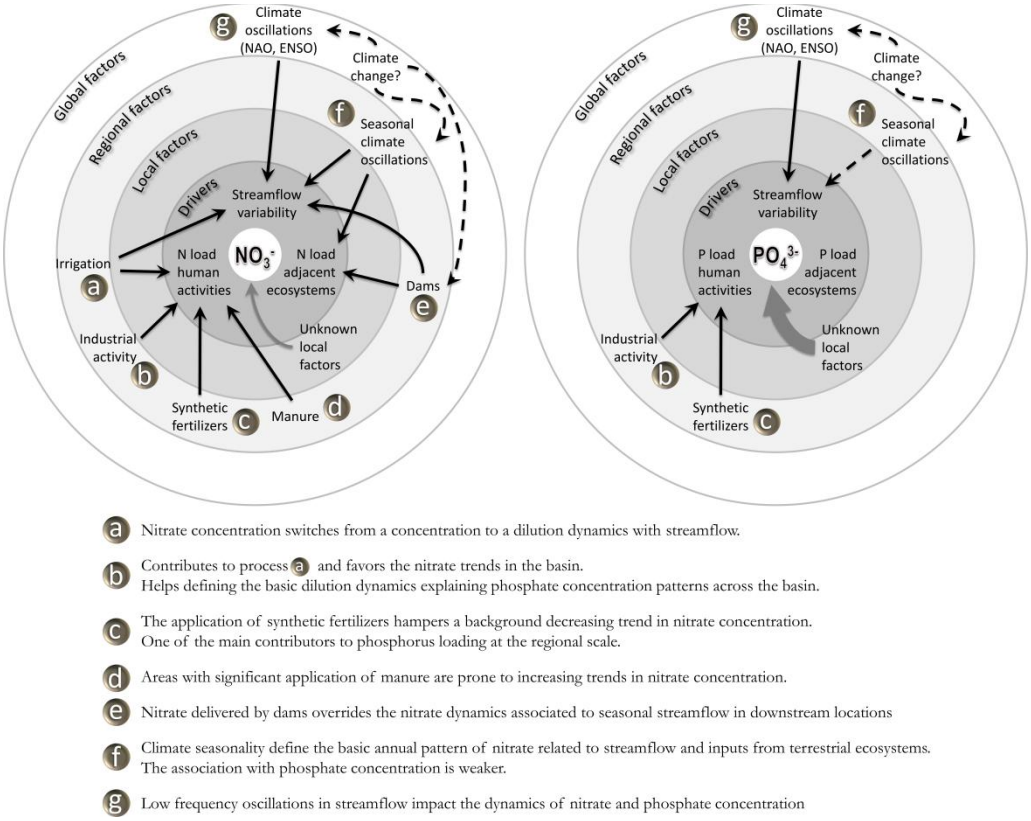


Figure 5: Global change factors, acting at different scales, that contribute to shape the spatio-temporal variability of nitrate and phosphate concentration in the Ebro basin. Numbered circles describe the relationship between nutrient concentration patterns and the identified factors and drivers of change.

Overall, our analysis shows that nitrate concentration dynamics is more responsive to regional and global factors, while global change impacts on phosphate concentration dynamics operate at the small scales of point sources. Anthropogenic land uses seem to play the most relevant role, and appropriate fertilization management may aid in stabilizing temporal trends, thus avoiding future nitrate concentration increases.

The relevance of the inter-annual signals in our nutrient concentration series suggest that any impact of climate change on the intensity and timing of global climate phenomena driving inter-annual streamflow oscillations can also exert a significant impact on river nutrient dynamics. This would be expressed more likely in variations of the prevalence of extreme events in streamflow that would impact nutrient dynamics. This may add to a multi-stressor situation typical from freshwaters in Mediterranean countries, guaranteeing future research on this topic.

General Discussion

General Discussion

In-stream nutrient processes at the basin scale

The intensity of human-induced global change, which has accelerated in the mid twentieth century, has marked the beginning of a new geological epoch, the Anthropocene (Crutzen & Stoermer, 2000). The environment has since suffered major alterations primarily due to a boost in industrial and agricultural activities. Although most of these changes occur mainly at specific local and regional scales, the accumulation of localized changes can reach a global scale through widespread human activity (Turner et al., 1990). Global environmental change is thereby not restricted to alterations that operate globally through the major components of the Earth system.

Freshwater ecosystems have continually shown significant responses to environmental changes across scales. Consequently, the sustainable management of inland resources continues to gain relevance, especially since human demand for high-quality water keeps increasing (Straile et al., 2003). The Mediterranean region has particularly undergone a myriad of alterations that have compromised the sustainability of freshwater resources and have further impaired the quality of water resources in the region.

Notwithstanding the conjunct efforts at local and regional management levels, nutrient pollution continues to detrimentally impact the quality of water in rivers and streams. The study of nutrient loads and their associated sources requires the use of holistic approaches that go beyond the reach scale and the evaluation of isolated observations within a river basin or a region. Conversely, since empirical work at large scales is logistically and economically challenging, reach-scale data and knowledge are vital as they set the foundation for upscaling key hydrological and biogeochemical processes to the basin scale. Basin modeling tools can provide useful information by linking observations to processes and changes occurring at a broad range of scales.

The models developed to simulate nutrient transport and retention at the basin level usually rely on reaction rate expressions in order to describe nutrient loss in terrestrial and aquatic ecosystems. Most of the rate expressions implemented in modeling assume first-order kinetics in the behavior of nutrient uptake, which means that the rate of nutrient loss is proportional to the load (or concentration) of the constituent(s) being modeled (Alexander et al., 2000; Wollheim et al., 2006).

It has recently been argued that the concentration of nutrients in streams determines the rate of uptake, which could eventually reach partial saturation (O'Brien et al., 2007). Data of nutrient spiraling metrics have shown that the efficiency of process rates relative to increasing concentration eventually declines. Within this Efficiency Loss (EL) concept, log-transformed areal uptake rates (U) increase with log-transformed nutrient concentration, where the slope of the relationship is less than one.

The formulation for reach nutrient attenuation in SPARROW has been modified (Chapter 1, Aguilera et al., 2013) to include a wider range of stream uptake dynamics (including first-order kinetics and EL). The decreasing values of the uptake velocity along the concentration gradient confirmed a loss in uptake efficiency in streams of all sizes in the Llobregat River basin. However, the slopes were much steeper than those corresponding to the relationship described by the bibliographical data. Thus, nutrient retention capacity decreased with increasing nutrient concentration, particularly in the nitrate model. Increased exported nutrient load and decreased in-stream nutrient removal coincide in a downstream direction, probably due to the significant chemical and geomorphological changes particularly suffered in the lower Llobregat River basin.

In general, variations in the nutrient retention capacity within the river networks are related to hydrological and geomorphological properties that determine the time a solute remains in contact with reactive surfaces (Bukaveckas, 2007). Biological activity is also likely to shape nutrient retention and transport to downstream waterbodies; for

instance, a recent study found that biological activity in streams under the influence of agricultural activities was higher relative to pristine streams (Bernot et al., 2006), although nutrient retention efficiency tends to be lower at nutrient enriched streams.

In order to describe the effective nutrient retention capabilities of river networks, we need to better understand the factors influencing lumped concepts and variables, such as nutrient spiraling metrics at larger scales (other than exclusively studying nutrient retention at the reach scale). Reaches in the lower section of the Llobregat River basin do not only receive chronic nutrient inputs but have been also modified by natural and anthropogenic processes and activities, which further influence the self-purification capacity of streams and rivers.

Nutrient source apportionment

The results in Chapter 2 have shown that natural sources of the two nutrients predominated in the upper part of the Llobregat basin, whereas agricultural activities were more relevant in its middle section and the upper Anoia River. Point sources were large contributors of nutrient loads, especially phosphates, in the lower part of the basin. Such apportionment pattern broadly agrees with the land use distribution in the Llobregat basin (Marcé et al., 2012).

Diffuse sources accounted for the major part of nitrate load arriving to rivers (natural land, 42% and cultivated land, 33%). In the phosphate model, point sources accounted for the greatest portion (48%) of the total average load generated in the watershed (Fig. 1). These different patterns follow those described in large European watersheds where agricultural sources contribute the greatest load of nitrogen and urban and point sources are major sources of phosphate loads (Grizzetti & Bouraoui, 2006; Schoumans *et al.*, 2009).

Nutrient apportionment varies according to nutrients and periods of wetter and dryer hydrological conditions. Diffuse nutrient loads derived from agricultural and natural

land in the Llobregat River basin increased during years with higher runoff values and decreased when runoff was lower. Overall, climate-associated factors are generally considered the main determinants of nutrient export, particularly in forested land (Beaulac & Reckhow, 1982). An evaluation of regional SPARROW models in the United States recognized that climatic variables played a major role in the delivery of nutrients to streams (Preston *et al.*, 2011a).

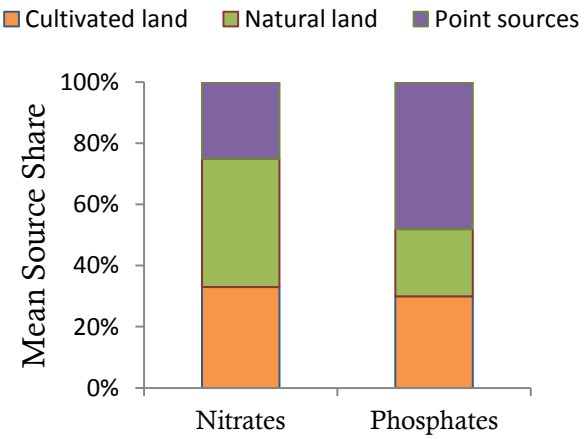


Figure 1: Average nutrient source apportionment in the Llobregat basin within the 2000-2006 study period.

Inter-annual hydrological variability was associated to differences of mean nutrient load apportionment among nutrient sources. Point source pollution was greater under dry conditions both for nitrate and phosphate. A high percentage of the measured water flow in streams and rivers under dry climatic conditions is composed of point source effluents, a condition that frequently occurs in Mediterranean rivers (Martí et al., 2010), and has implications for pollutants other than nutrients (Petrovic et al., 2010).

The results in Chapters 1 and 2 of this thesis support the idea of using modeling tools within a heuristic approach to better understand nutrient transport and retention through a river network. This implies that model development is a dynamic process,

where new questions and issues arise as a consequence of changes in management practices and continuous improvement of knowledge about processes at fine and broad scales. In this sense, the iterative refinement of current modeling tools generate new solutions and further questions to address in order to guide and support management decisions that seek to improve the quality of water resources at local, regional and global scales.

Regional river water quality signatures

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, Chapter 3 of the current thesis has introduced a collection of methodological steps for detecting and potentially attributing global environmental change effects on water quality patterns, taking into consideration both temporal and spatial variability of responses and drivers.

The detection and attribution of global change effects on water quality cannot be achieved through observation alone (Ito, 2012). Dynamic Factor Analysis (DFA) provided the methodological framework to extract underlying common patterns in nutrient time-series with missing observations, a commonly encountered problem in environmental databases. Most importantly, DFA guaranteed the explicit consideration of the interaction between temporal and spatial patterns of change necessary to investigate the drivers and processes that shape them. The extracted patterns were further characterized using complementary methods such as frequency and trend analyses for the temporal dimension, together with regression and grouping analysis of spatially distributed data.

Chapter 4 has introduced a comprehensive implementation of the methodology in the Ebro River Basin, a large basin that encloses a range of climatic characteristics as well

as different land uses and covers. Nitrate and phosphate common patterns were extracted from a set of time-series scattered across the basin by means of DFA. The extracted nitrate concentration patterns described a large proportion of the observed variability at the basin scale. Conversely, phosphate concentration patterns did not fully describe the behavior of all monitoring points included in the analysis. The differing components (cycles, trends) identified within the extracted patterns of the time-series respond to different drivers (Fig. 2), as has been found in other studies dealing with univariate time-series decomposition (Halliday et al., 2012; Worrall et al., 2014).

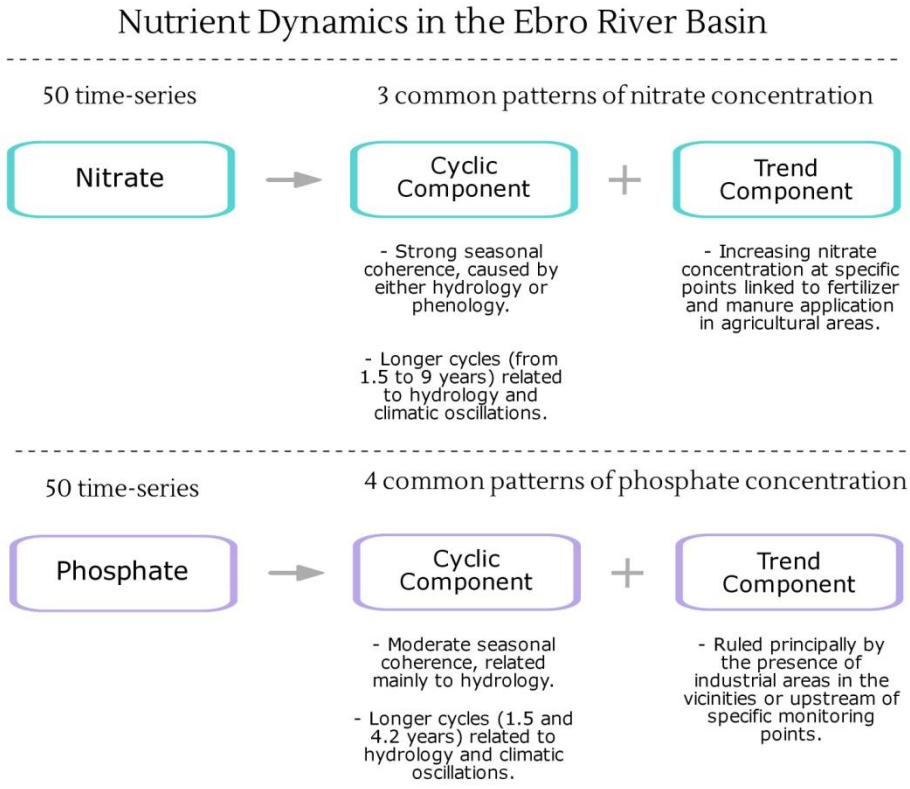


Figure 2: Summary of the main components identified in the extracted patterns of 50 time-series of nitrate and phosphate and their associated drivers.

The seasonality of nitrate patterns was either driven by hydrological or phenological processes, depending on the geographical location of the monitoring point. The

frequencies of the cycles identified in nitrate concentration patterns were further potentially linked to climatic oscillations, such as the El Niño Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). Because the hydrology of inland waters is intimately linked to precipitation, the influence of climatic oscillations on river discharge would be expected. Physical responses to climatic oscillations might result in secondary chemical responses, due for instance to changes in river inflow or in mixing or upwelling processes. Alterations in winter temperature and precipitation as a response of changes in climatic modes might, in turn, result in changes in nutrient release from soil in the basin area (Straile et al., 2003).

The significant trends identified in the nutrient common patterns were mainly associated to local land and water management practices. For instance, land uses linked to fertilizer application further modulated the increasing and decreasing nitrate trends scattered across the basin. Decreasing phosphate trends observed across space and time coincided with changes in land-use management practices in the study basin such as the implementation and the improvement of sewage water treatment facilities (a shared characteristic with river basins across Europe) and the reduction of phosphate content in the composition of fertilizers.

The added value of the work presented in Chapters 3 and 4 is that the implemented methodology allows the simultaneous analysis of sets of time-series of a particular water quality variable, and that it further characterizes the spatio-temporal variability and seeks to provide an explanation for the main patterns observed across a wide range of environmental conditions and scales. The results therein emphasize the synergistic interaction between land use and climate changes that shape both temporal and spatial patterns in water quality as a consequence of local and global anthropogenic activities (Pielke, 2005).

Previous studies dealing with detection of changes and the subsequent attribution of their potential drivers and/or forcings have been mainly performed within the context of climate change. Efforts towards the formalization of detection and attribution of

impacts of climate change on human and natural environments have been made (Hegerl et al., 2010; Stone et al., 2013). Among the reported key problems for current assessments, the limited availability of continuous and long-term observations has been addressed in this thesis. In addition, the methodological framework applied herein has dealt with the challenge of studying patterns and impacts that occur heterogeneously in space and time (Ito, 2012). Although this thesis has focused on the river water quality in the Mediterranean region due to the inherent and recently aggravated water stress in arid and semi-arid zones, the series of steps outlined and exemplified in Chapters 3 and 4 could be applied to any basin or region in the world.

Future Perspectives

Global environmental changes are complex and so are the responses of the Earth system across multiple temporal and spatial scales. Models, on the other hand, have been defined as a simplified representation of the real system under study. Counterintuitively, models are needed to investigate the processes involved in ecosystem response to changes in the sense that they are able to interpret our understanding of the system into a behavioral outcome (Stone et al., 2013).

Particularly in the case of nutrient spiraling and water quality at the basin scale, given basin-wide measurement and sampling difficulties, river network models of nutrient transport and transformation are a suitable alternative tool to unravel and upscale the most essential mechanisms found at the reach scale to the basin and regional scales. Here, the application of the SPARROW model avoided the use of excessively complex model constructs while providing a combination of the main advantages of process-based (mechanistic understanding) and statistical techniques (parameter estimation based on data availability). This exercise is however not exempted of a certain level of uncertainty in parameter estimation and model consistency. For these and other purposes, Bayesian calibration frameworks have been proposed within SPARROW (Qian et al., 2005; Wellen et al., 2014) and for water-quality models in general (Rode et al., 2010). The initial model uncertainty could thus be iteratively

overcome based on new knowledge and new questions that emerge from previous modeling exercises (Wellen et al., 2014).

An example of recently acquired insight by scientists that has been implemented in this thesis would be the inclusion of Efficiency Loss kinetics, which permitted the study of nutrient in-stream decay within a wide range of nutrient concentration and streamflow characteristics in a highly humanized river basin. As a complementary and supportive approach to the findings reported on this matter in the current thesis, field work in large and impaired reaches and river networks would be ideal.

In terms of detection and attribution of global change impacts on freshwater quality, similar efforts to those applied in climate change studies should be sought in order to have a formal framework for the assessment process. The current thesis has dealt with some of the reported issues that usually hinder the assessment of impacts of global change phenomena on the Earth System, one of them being continuous data availability and sampling frequency. An additional analysis summarized in Appendix C has shown, however, that a coarse temporal resolution, i.e. monthly, can compromise the analysis of patterns in time-series of biological variables, which tend to exhibit larger high-frequency variability than physico-chemical variables.

One of the disadvantages of using Dynamic Factor Analysis as the core of water quality patterns detection was the computationally intensive algorithm involved in the Expectation-Maximization iterative process of parameter estimation. While the R Project for Statistical Computing offers many solutions and advantages to the community of R users without any financial cost, it is worth keeping in mind that several available R-Packages are current research projects themselves. Achieving the most efficient algorithm is often part of an ongoing process that involves developers and users reporting on bugs and inefficiencies observed in highly complex algorithms such as those involved in MARSS (Multivariate Autoregressive Space-State models; (Holmes *et al.*, 2012). The work of this thesis was mainly based on the MARSS R-Package version 3.4, released in February 2013. Since then, many improvements have

been reported and the EM algorithm is currently being re-written in C++ to gain more speed.

Despite the few drawbacks of this method, Dynamic Factor Analysis offers the possibility to extract water quality signatures from multivariate noisy and gappy time-series, providing not only a temporal interpretation of variability but also allowing researchers to include the spatial variability of patterns across basin and regional scales.

With the improvement of computational resources and the knowledge gained through research on global change and freshwater ecosystems 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 global scales, in order to produce reliable forecasts for the availability of water resources that can in turn assist policy-makers and managers to ultimately secure the sustainability of the most important natural resource on Earth.

Harmonizing science, management practices, and policy however remains a challenging task and a somewhat cumbersome obstacle to overcome in order to protect water resources worldwide. Perhaps one way to deal with this issue is to advocate for projects that involve all parties from the very beginning, so that the objectives can be established having in mind a specific problem, with specific stakeholders, and specific benefits for the wellbeing of freshwater ecosystems and those who directly benefit from them: humans. In this picture, models could help to visualize the existent knowledge on changes on the environment and its consequences, and to generate new questions to be investigated based on the already available information. The existent trend in reconciling water science with the socioeconomic component of society could be a path for drawing the attention of governmental agencies, and of the public in general, towards the promotion of sustainable water resources.

General Conclusions

General Conclusions

1. The calibrated in-stream retention parameters in the SPARROW nitrate and phosphate models did not fully reflect an Efficiency Loss (EL) dynamics of nutrient retention in the Llobregat River basin.

2. Uptake velocity values steeply decreased with increasing nutrient concentration, contrasting with other river systems. In the nitrate model, the uptake rate slightly decreased with increasing concentration, contrary to what would have been expected based on bibliographical data and the EL concept. In the case of phosphate, the behavior was at the limit of an EL dynamics, with an almost imperceptible increase of uptake rate with increasing nutrient concentration.

3. Increased exported nutrient load and decreased in-stream nutrient removal coincide in a downstream direction, probably due to the significant chemical and geomorphological changes particularly suffered in the lower Llobregat River basin.

4. The SPARROW models identified diffuse pollution as the major source of nitrate export in the Llobregat River Basin, and point sources as the largest contributors of phosphate.

5. Overall, nutrient export and apportionment reflected the inter-annual hydrological variability: larger nutrient loads from diffuse sources were observed in wetter years and point sources accounted for the greatest portion of nutrient loading under dry conditions.

6. The application of SPARROW in the Llobregat River basin showed that models working at a level of intermediate complexity may offer a good compromise between process description and treatment of data available from water agencies.

7. The implemented methodology for change detection allowed 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.

8. As a preliminary step for the exploration and detection of patterns, the Maximal Information-based Nonparametric Exploration (MINE) algorithm was able to pinpoint water quality variables of interest that could be subsequently studied in more detail.

9. By combining the information provided by the patterns detected by Dynamic Factor Analysis (DFA) and the associated factor loadings, a full description of the spatio-temporal variability of basin-wide water quality patterns was obtained in three Mediterranean basins.

10. The proposed methodological framework for change detection and attribution allowed dealing with different temporal and spatial scales at which interacting processes associated to global change phenomena usually take place.

11. In the Ebro River Basin, the extracted nitrate concentration patterns described a large proportion of the observed variability at the basin scale, whereas phosphate concentration patterns did not fully depict the behavior of all monitoring points included in the analysis.

12. The impact of global change on the dynamics of nitrate concentration across the Ebro basin was found to be a multifaceted process including regional and global factors, whereas impacts on phosphate concentration depend more on local impacts and less on large-scale factors.

Appendices

Appendix A

MODELING NUTRIENT RETENTION AT THE WATERSHED SCALE: SUPPLEMENTARY MATERIAL

Table A1: Model formulations used in LOADEST to estimate nutrient loads (Runkel et al., 2004). The specified value is set under the MODNO option in LOADEST; we chose the value of 0 for each monitoring station, which in turn chooses the best of the model formulations between 1 and 9.

Specified	Regression model included in LOADEST
0	automatically select best model from models 1-9.
1	$a_0 + a_1 \ln Q$
2	$a_0 + a_1 \ln Q + a_2 \ln Q^2$
3	$a_0 + a_1 \ln Q + a_2 dtime$
4	$a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime)$
5	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 dtime$
6	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime)$
7	$a_0 + a_1 \ln Q + a^2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4 dtime$
8	$a_0 + a_1 \ln Q + a_2 \ln Q_2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime$
9	$a_0 + a_1 \ln Q + a_2 \ln Q_2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime +$
10	$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q\ per$
11	$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q\ per + a_4 \ln Q2 + a_5 \ln Q^2\ per$
99	user defined

Table A2: Selected models for load estimation in monitoring stations in the Llobregat River basin along with some performance measures. AIC = Akaike Information Criterion. SE = Standard Error. AMLE = Adjusted Maximum Likelihood Estimation. Turnbull-Weiss is a normality test of the residual.

Phosphate		Model		AMLE Load Estimation			AMLE Regression Statistics			AMLE	Turnbull-Weiss
Station	N of Observations	Number	AIC	Mean Load	SE	% SE	R-Squared [%]	Prob. Plot Corr. Coeff. (PPCC)	Serial Correlation of Residuals	Residual Variance	p-value
Abrera	78	2	1.44	430	29	7	38	0.93	0.29	0.23	0.1760
Balsareny	80	7	2.05	75	7	9	49	0.98	0.03	0.42	0.0143
Barcelona	68	3	2.10	1155	124	11	21	0.96	0.23	0.44	0.3677
Cardona	46	1	2.43	30	5	16	4	0.99	0.13	0.80	0.0001
Castellbell	82	1	1.93	320	28	9	14	0.89	0.05	0.38	< 0.0001
Castellgalí	81	1	2.75	237	31	13	2	0.94	0.27	0.88	0.0089
Clariana de Cardener	19	3	3.38	252	137	54	7	0.98	0.31	1.50	0.3983
Guardiola de Berguedà	43	3	2.18	32	5	16	61	0.97	-0.15	0.51	< 0.0001
Jorba	79	5	2.60	4	1	21	76	1.00	0.01	0.94	0.0002
La Pobla de Lillet	64	3	1.39	7	1	12	85	0.96	-0.15	0.38	< 0.0001
Martorell	79	9	2.45	152	18	12	43	0.95	0.14	0.60	< 0.0001
Navarcles	19	3	2.67	25	7	29	41	0.92	-0.09	0.70	< 0.0001
Olesa de Montserrat	78	4	1.26	431	29	7	39	0.96	0.21	0.20	0.4847
Olius	65	1	2.20	17	4	25	15	0.87	-0.02	1.55	< 0.0001
Olvan	82	5	2.55	57	8	14	29	0.95	-0.01	0.95	< 0.0001
Pont de Vilomara	81	4	2.33	97	10	11	12	0.96	0.08	0.57	0.0068
Sallent	37	3	2.34	20	4	20	68	0.98	0.20	0.58	0.1830
Sant Fruitós de Bages	49	3	2.34	20	6	30	28	0.98	0.14	0.57	0.2554
Sant Joan Despí	78	4	1.76	416	49	12	68	0.92	0.32	0.31	0.0001
Sant Sadurní d'Anoia	48	9	2.96	26	15	58	93	0.92	-0.06	0.96	0.0001
Santa Maria de Merlès	46	7	1.94	23	3	13	77	0.97	-0.08	0.38	0.0076
Subirats	50	7	2.29	70	15	21	31	0.98	-0.03	0.52	0.5770
Vilanova del Camí	79	8	2.65	112	15	13	35	0.97	0.18	0.76	0.2400

Table A3: SPARROW results for the in-stream decay slope b after the model estimation that tested the "no effect" variant of the runoff variable.

Parametric	Std Err	Pr > |t|	Bootstrap	Std Errr	Pr > |t|	90% CI LB	90% CI UB
			Nitrate Model				
-1.08	0.09	< 0.001	-1.24	0.27	< 0.001	-1.73	-0.94
			Phosphate Model				
-0.72	0.09	< 0.001	-0.78	0.13	< 0.001	-1.00	-0.58

Table A4: In-stream retained mass (kg) and fraction averaged for the 7 years included in the nitrate model, resulted from the first-order in-stream decay as well as from our model (power law).

Reach classified as stream	First-order kinetics		Our Model	
	Mass	Fraction	Mass	Fraction
1	20943.01	0.11	28263.92	0.14
2	115579.77	0.37	233128.94	0.58
3	47994.32	0.20	140232.01	0.42
4	223625.04	0.46	320221.04	0.53
5	9138.25	0.11	5559.50	0.08
6	46376.81	0.20	69722.64	0.27
9	304519.86	0.62	325597.87	0.62
10	-125649.80	-0.26	-180631.09	-0.42
11	598066.86	0.47	463784.51	0.41
12	104306.42	0.38	48897.57	0.22
13	1353183.02	0.64	1707526.50	0.78
14	27757.65	0.06	324976.71	0.39
15	1457203.76	0.77	1287327.22	0.77
16	56465.54	0.21	31580.05	0.13
17	23638.68	0.13	1155.01	0.01
18	518661.41	0.82	150178.65	0.35
19	1714132.18	0.74	1485182.40	0.73
20	64458.97	0.16	170971.00	0.22
21	2410059.69	0.75	2919730.39	0.81
22	861779.79	0.72	271132.67	0.33
23	3300738.56	0.70	3370537.75	0.73
24	2231066.44	0.72	2017106.06	0.68
25	-91727.84	-0.14	-151754.82	-0.26
26	35756.26	0.12	153843.79	0.24
27	3343164.14	0.62	3339154.98	0.65
28	-964.32	-0.01	-21698.82	-0.35
29	2502935.31	0.73	2468411.06	0.69
30	148108.65	0.47	89808.07	0.36
31	32561.54	0.27	22634.27	0.23
32	2771622.07	0.64	2632584.00	0.62
33	2543373.64	0.69	2601000.07	0.70
34	4250071.11	0.74	4000743.70	0.74
35	129805.61	0.33	17223.77	0.07
36	31098.97	0.18	4643.36	0.03
37	644421.17	0.52	296918.46	0.34
38	41807.63	0.18	20663.65	0.12
39	4991266.83	0.56	4896157.95	0.57
40	40723.84	0.15	5645.14	0.03
41	5799406.76	0.62	5587737.01	0.61
42	25690.42	0.21	4891.24	0.05
43	7286765.27	0.64	6538197.17	0.62
44	1796559.25	0.71	840324.87	0.49

Table A4 (Continued)

Reach classified as stream	First-order kinetics		Our Model	
	Mass	Fraction	Mass	Fraction
45	-51580.59	-0.17	-115962.35	-0.52
46	9453531.07	0.63	7418514.73	0.57
47	95942.02	0.20	135613.97	0.29
48	1892033.05	0.66	864902.92	0.44
49	895502.61	0.32	121568.53	0.08
50	77900.67	0.35	18719.37	0.14
51	10240123.25	0.66	8002646.88	0.59
52	9682.58	0.14	1740.15	0.03
53	8083136.84	0.62	7015931.48	0.60
54	66883.87	0.46	71973.43	0.48
55	378771.56	0.41	630905.36	0.57
56	689523.58	0.55	330343.97	0.37
57	900834.68	0.65	1081721.22	0.68
58	1188487.31	0.58	1396610.31	0.64
59	252583.42	0.34	708008.73	0.63
60	1866366.60	0.71	2321209.94	0.74
61	3408668.05	0.61	3393945.63	0.64
62	3413435.52	0.71	3382860.94	0.71
63	2000364.77	0.68	1794444.68	0.62
64	1166255.00	0.75	1057332.47	0.73
65	215568.25	0.44	148523.07	0.35
68	672193.46	0.86	607748.00	0.80
71	5692122.98	0.67	5697819.97	0.67
72	7040652.71	0.63	6431473.98	0.62
73	132981.90	0.27	213683.89	0.46
74	9012537.91	0.68	7948825.06	0.66
75	9933876.15	0.59	8026956.07	0.56
76	244769.12	0.50	34936.25	0.12
77	897506.05	0.57	362885.28	0.34
78	82330.55	0.27	28953.81	0.13
79	70384.47	0.45	9488.07	0.09

Table A5: In-stream retained mass (kg) and fraction averaged for the 7 years included in the phosphate model, resulted from the first-order in-stream decay as well as from our model (power law).

Reach classified as stream	First-order kinetics		Our Model	
	Mass	Fraction	Mass	Fraction
1	1205.36	0.16	5960.29	0.37
2	4921.86	0.39	28360.87	0.81
3	2474.58	0.27	22283.85	0.82
4	7127.57	0.43	38089.94	0.78
5	581.43	0.15	1158.76	0.23
6	2502.69	0.27	13158.89	0.68
9	10190.12	0.63	32157.86	0.84
10	-69214.50	-4.59	-49183.30	-1.55
11	-60227.13	-1.01	-2776.29	-0.05
12	5873.84	0.50	9644.87	0.65
13	81539.91	0.65	189253.41	0.90
14	4881.35	0.38	70814.48	0.89
15	71548.60	0.80	142450.10	0.93
16	3223.15	0.29	8294.81	0.34
17	4366.14	0.18	444.86	0.02
18	28333.55	0.91	23305.17	0.62
19	85749.39	0.80	168191.19	0.90
20	3046.23	0.22	52886.85	0.44
21	135958.69	0.83	337536.65	0.94
22	61972.18	0.78	58504.93	0.72
23	212261.00	0.84	418548.11	0.94
24	113560.23	0.77	282334.60	0.87
25	18814.43	0.67	36809.35	0.80
26	1746.86	0.17	43868.37	0.40
27	247315.82	0.82	472358.69	0.94
28	-4365.86	-1.41	-3771.49	-1.00
29	128316.25	0.77	364677.38	0.85
30	9971.73	0.59	14527.95	0.54
31	3458.84	0.37	2807.46	0.28
32	146491.21	0.64	391056.88	0.80
33	65021.46	0.43	328080.23	0.77
34	290147.49	0.85	508648.95	0.94
35	15451.82	0.92	26934.49	0.95
36	1877.56	0.24	1653.17	0.11
37	36940.01	0.48	54296.33	0.57
38	4420.06	0.23	11345.13	0.56
39	264966.99	0.63	752049.12	0.83
40	6606.98	0.20	864.34	0.04
41	300365.59	0.66	783993.00	0.83
42	4927.17	0.29	723.77	0.06
43	505083.27	0.64	963553.51	0.81
44	193257.25	0.74	172749.02	0.78
45	5051.71	0.33	9885.67	0.50

Table A5 (Continued)

Reach classified as stream	First-order kinetics		Our Model	
	Mass	Fraction	Mass	Fraction
46	842715.23	0.73	1252237.94	0.84
47	-238.65	0.27	322.31	0.08
48	218459.03	0.75	205550.95	0.82
49	170989.33	0.76	199853.79	0.78
50	-10256.83	-0.66	-5417.11	-0.26
51	873831.78	0.74	1252112.52	0.84
52	3099.38	0.19	132.86	0.02
53	587325.51	0.58	1059003.64	0.71
54	1787.86	0.37	7930.51	0.72
55	17435.62	0.44	78046.14	0.84
56	46888.44	0.83	63144.84	0.87
57	36893.43	0.64	110006.13	0.83
58	67599.78	0.56	153239.15	0.75
59	14853.91	0.56	112770.01	0.94
60	80103.99	0.75	264878.60	0.91
61	250164.15	0.76	475937.27	0.90
62	179540.04	0.84	410972.51	0.92
63	100647.55	0.73	246757.34	0.79
64	43430.30	0.81	113424.37	0.92
65	11536.88	0.54	22649.59	0.65
68	28525.52	0.84	48717.49	0.76
71	257978.23	0.69	751193.13	0.86
72	480355.47	0.66	951343.42	0.83
73	328.72	0.36	-235.64	0.14
74	660936.77	0.81	1137203.34	0.88
75	977190.95	0.46	1285439.22	0.58
76	25241.52	0.93	33618.41	0.96
77	74466.06	0.58	69746.60	0.58
78	8963.82	0.31	13612.25	0.48
79	5260.60	0.54	1806.74	0.17

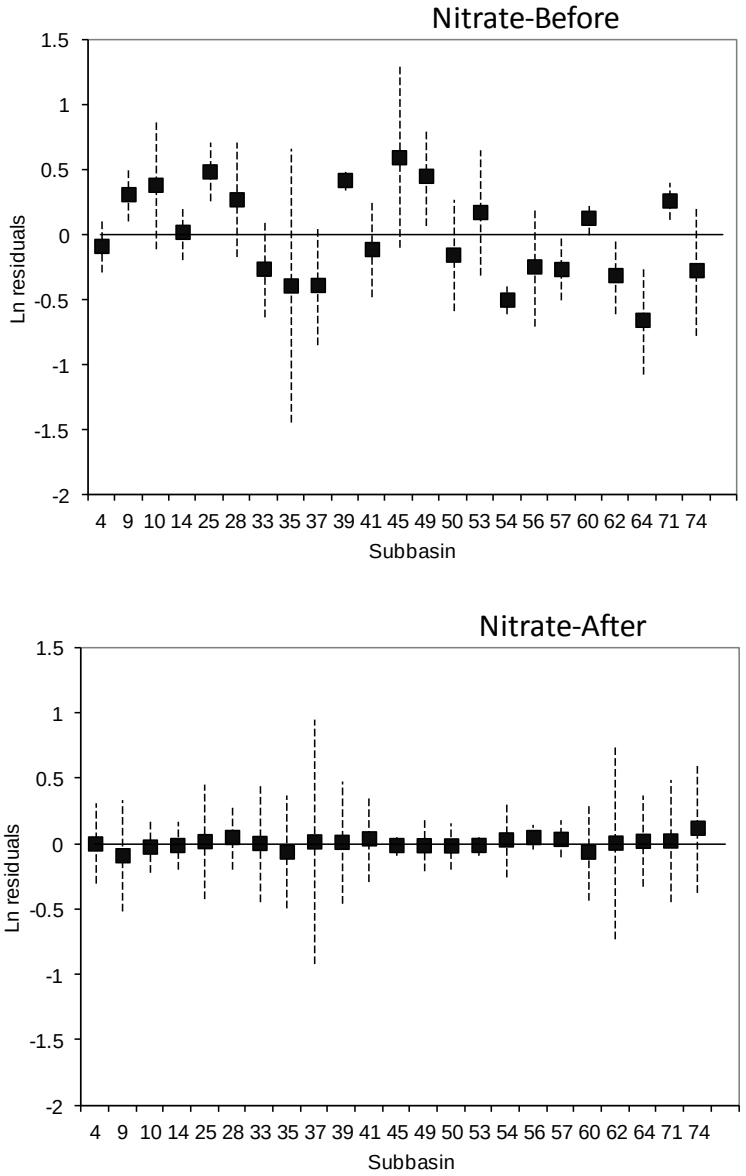


Figure A1: Nitrate mean model residuals before and after the inclusion of the random error term in SPARROW.

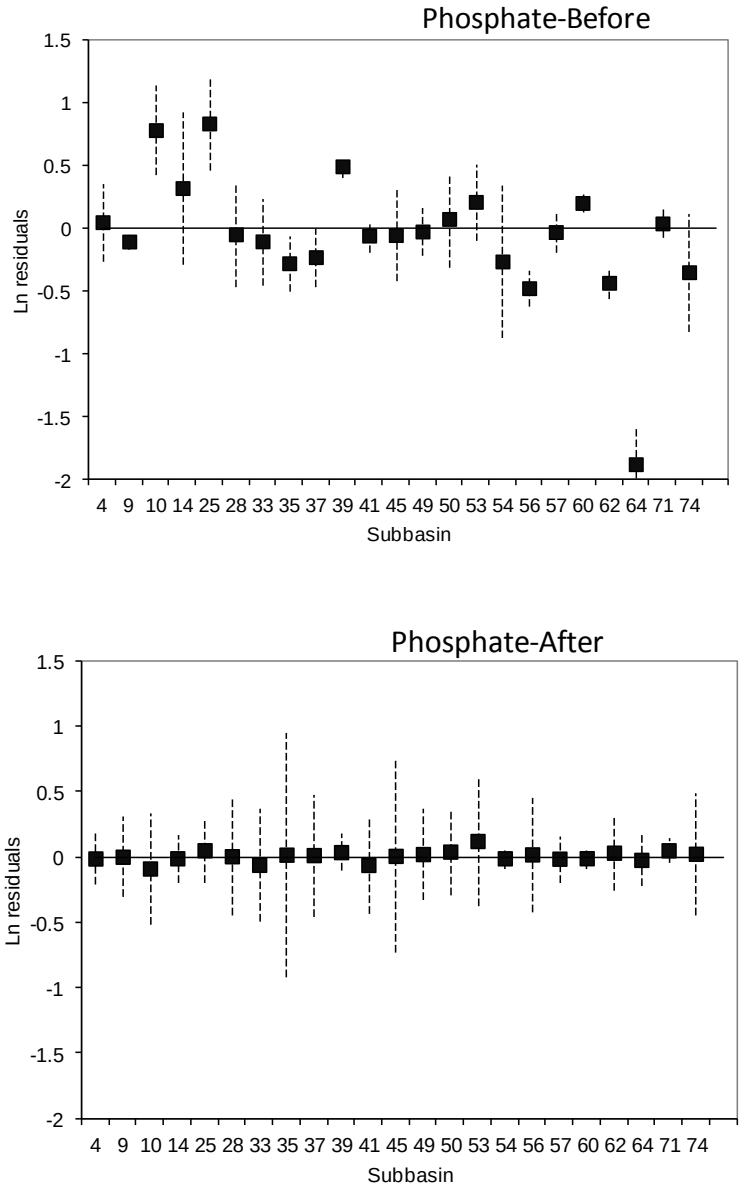


Figure A2: Phosphate mean model residuals before and after the inclusion of the random error term in SPARROW.

Section A1: References used to compile the Bibliographical Data and symbol correspondence in Figures 3 and 4 of the manuscript.

Symbol Correspondence

Nitrate model

●	A	Ashkenas et al. (2004)
●	B	Cooper and Cooke (1984)
▼	C	Grimm et al. (2005)
▲	D	Maltchik et al. (1994)
■	E	Merriam et al. (2002)
□	F	Mulholland et al. (2004)
◇	G	Mulholland et al. (2000 a)
◆	H	Mulholland et al. (2000 b)
▲	I	Tank et al. (2000)
▼	J	Thomas et al. (2003)
⬢	K	Webster et al. (2003)
⬡	L	von Schiller et al. (2008)
○	M	Mulholland et al. (2008)

Phosphate model

●	A	Doyle et al. (2003)
●	B	Maltchik et al. (1994)
▼	C	Mulholland et al. (1985)
▲	D	Mulholland et al. (1990)
■	E	Mulholland et al. (1997)
□	F	Newbold et al. (1983)
◇	G	Gücker and Pusch (2006)
◆	H	Haggard et al. (2005)
▲	I	Ruggiero et al. (2006)
▼	J	Martí et al. (2004)

References (listed alphabetically)

Ashkenas, L.R., S.L. Johnson, S.V. Gregory, J.L. Tank, W.M. Wollheim. 2004. A stable isotope tracer study of nitrogen uptake and transformation in an old-growth forest stream. Ecology 85(6):1725-1739.

Cooper, A.B., J.G. Cooke. 1984. Nitrate loss and transformation in 2 vegetated headwater stream. New Zeal. J. Mar. Fresh. 18:441-450.

Doyle, M.W., E.H. Stanley, J.M. Harbor. 2003. Hydrogeomorphic controls on phosphorus retention. Water Resour. Res. 39(6):1147.

Grimm, N.B., R.W. Sheibley, C.L. Crenshaw, C.N. Dahm, W.J. Roach, L.H. Zeglin. 2005. N retention and transformation in urban streams. J. N. Am. Benthol. Soc. 24(3):626-642.

Gücker, B.,Pusch, M.T. 2006. Regulation of nutrient uptake in eutrophic lowland streams. Limnol. Oceanogr. 51:1443-1453

Haggard, B.E., E.H. Stanley, D.E. Storm. 2005. Nutrient retention in a point-source-enriched stream. J.N. Am. Benthlo. Soc. 24:29-47.

Maltchik, L., S. Molla, C. Casado, C. Montes. 1994. Measurement of nutrient spiralling in a Mediterranean stream: comparison of two extreme hydrological periods. Archiv. Hyrobiol. 130(2):215-227.

Martí, E., J. Aumatell, L. Godé, M. Poch, F. Sabater. 2004. Nutrient retention efficiency in streams reciving inputs from wastewater treatment plants. J. Environ. Qual. 33:285-293.

Merriam, J.L., W.H. McDowell, J.L. Tank, W.M. Wollheim, C.L. Crenshaw, S.L. Johnson. 2002. Characterizing nitrogen dynamics, retention and transport in a tropical rainforest stream using an in situ 15N addition. Freshwater Biol. 47:143-160.

Mulholland, P.J., J.D. Newbold, J.W. Elwood, L.A. Ferren. 1985. Phosphorus spiralling in a woodland stream: seasonal variations. Ecology 66(3):1012-1023.

Mulholland, P.J., A.D. Steinman, J.W. Elwood. 1990. Measurement of phosphorus uptake length in streams: comparison of radiotracer and stable PO4 releases. Can. J. Fish. Aquat. Sci. 47:2351-2357.

Mulholland, P.J., E.R. Marzolf, J.R. Webster, D.R. Hart, S.P. Hendricks. 1997. Evidence that hyporheic zones increase heterotrophic metabolism and phosphorus uptake in forest streams. Limnol. Oceanogr. 42(3):443-451.

Mulholland, P.J., J.L. Tank, D.M. Sanzone, J.R. Webster, W.M. Wollheim, B.J. Peterson, J.L. Meyer. 2000. Ammonium and nitrate uptake lengths in a small forested stream determined by 15N and short-term nutrient enrichment experiments. Verh. Internat. Verein. Limnol. 27: 1320-1325.

Mulholland, P.J., J.L. Tank, D.M. Sanzone, W.M. Wollheim, B.J. Peterson, J.R. Webster, J.L. Meyer. 2000. Nitrogen cycling in a forest stream determined by a 15N tracer addition. Ecol. Monogr. 70(3):471-493.

Mulholland, P.J., H.M. Valett, J.R. Webster, S.A. Thomas, L.W. Cooper, S.K. Hamilton, B.J. Peterson. 2004. Stream denitrification and total nitrate uptake rates measured using a field 15N tracer addition approach. Limnol. Oceanogr. 49(3):809-820.

Mulholland, P.J., A.M. Helton, G.C. Poole, et al. 2008. Stream denitrification across biomes and its response to anthropogenic nitrate loading. Nature. 452:202-206.

Newbold, J.D., J.W. Elwood, R.V. O'Neill, A.L. Sheldon. 1983. Phosphorus dynamics in a woodland stream ecosystem: a study of nutrient spiralling. Ecology 64(5):1249-1265.

Ruggiero, A., A.G. Solimini, G. Carchini. 2006. Effects of a waste water treatment plant on organic matter dynamics and ecosystem functioning in a Mediterranean stream. Ann. Limnol.-Int.J. Lim. 42:97-107.

Tank, J.L., J.L. Meyer, D.M. Sanzone, P.J. Mulholland, J.R. Webster, B.J. Peterson, W.M. Wollheim, N.E. Leonard. 2000. Analysis of nitrogen cycling in a forest stream during autumn using a 15N-tracer addition. Limnol. Oceanogr. 45(5):1013-1029.

Thomas, S.A., H.M. Valett, J.R. Webster, P.J. Mulholland. 2003. A regression approach to estimating reactive solute uptake in advective and transient storage zones of stream ecosystems. Adv. Water Resour. 26:965-976.

von Schiller, D., E. Martí, J.L. Riera. 2008. Nitrate retention and removal in Mediterranean streams with contrasting land uses: a 15N tracer study. Biogeosciences. 6: 181-196.

Webster, J.R., P.J. Mulholland, J.L. Tank, H.M. Valett, W.K. Dodds, B.J. Peterson, W.B. Bowden, C.N. Dahm, S.Findlay, S.V. Gregory, N.B. Grimm, S.K. Hamilton, S.L. Johnson, E. Marti, W.H. McDowell, J.L. Meyer, D.D. Morrall, S.A. Thomas, W.M. Wollheim. 2003. Factors affecting ammonium uptake in streams - an interbiome perspective. Freshwater Biol. 48:1329-1352.

Appendix B

A PRIMER ON THE STUDY OF THE SPATIO-TEMPORAL VARIABILITY OF RIVER WATER QUALITY SERIES USING DYNAMIC FACTOR ANALYSIS: R SCRIPTS

Section B1. Standardize variables within datasets

The script TS_STAND.R standardizes the observed values in individual monitoring stations, based on identification codes, included in water quality datasets with all stations and prepares an output table to be used for plotting purposes.

Input Data
R_LLOB.csv file

Code	*Name*	*Date*	*NH4*	*Chlor*	*Cond*	*P2O5*	*NO3*	*COD.MnO4*	*DO*	*pH*	*Temp.water*
LL0001	Abrera	37193	0.6	223.2	1300	0.5	8.2	5.4	3.5	7.9	15.8
LL0001	Abrera	37200	0.6	256.5	1392	0.5	9.6	4.5	9.5	8.1	14
LL0001	Abrera	37235	1	229.3	1535	0.4	9.2	3	13.1	7.9	7.5
...											

R Script

```
################
## Read input data ##
################
input<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\WQ Time Series\\Copy of
Data\\R_LLOB.csv",header=T)

summary(input)
nrow(input);ncol(input)

#######################
## WQ stations - unique codes ##
#######################
codes<-as.matrix(unique(input$Code))
codes

nrow(codes)
dim(codes)

#####################################
##  Function to calculate column mean and cocients  ##
#####################################
mean.sweep<-function(x){
x.1<-as.matrix(apply(x,2,mean,na.rm=T))
x.2<-as.matrix(sweep(x,2,x.1,"-"))
x.3<-as.matrix(apply(x,2,sd,na.rm=T))
x.4<-as.matrix(sweep(x.2,2,x.3,"/"))

} # end function
```

```
#######################################
##  Number of columns to perform standardization on ##
#######################################
col<-ncol(input)- 2
col

############################
## Create empty matrix to store results ##
############################
output<-matrix(nrow=0,ncol=col,byrow=T)
ncol(output)

out.n.row<-matrix(nrow=0,ncol=1,byrow=F)

##############################################
## Loop to calculate mean.sweep function for each WQ station ##
##############################################
for (i in 1:nrow(codes)){
w<-which(input[,c(1)]==codes[i])
stand<-mean.sweep(input[w,c(3:(ncol(input)))])

output<-rbind(output,stand)

################################
##  How many rows per monitoring station ##
################################
n.row<-nrow(stand)
out.n.row<-rbind(out.n.row,n.row)

} # end loop

###############
##  Export output ##
###############
output.stand<-cbind(input[1:2],output)
names(output.stand)
nrow(output.stand); ncol(output.stand)

setwd("C:\\Documents and Settings\\raguilera\\My Documents\\WQ Time Series\\Output.stand")
write.table(output.stand, file="OUTPUT.standL.csv", sep=",", qmethod="double", row.names=F)
write.table(out.n.row, file="OUT.nrowL.csv", sep=",", qmethod="double", row.names=F, col.names="n.row")
```

Output Data

OUTPUT.standL.csv

Code	*Name*	*Date*	*NH4*	*Chlor*	*Cond*	*P2O5*	*NO3*	*COD.MnO4*	*DO*	*pH*	*Temp.water*
LL0001	Abrera	37193	-0.33053	-0.82107	-0.53217	-0.0116	-0.1774	0.209727	-3.22313	-1.04664	-0.02714
LL0001	Abrera	37200	-0.33053	-0.35599	-0.30708	-0.0116	0.051129	-0.30192	0.114471	-0.07977	-0.32778
LL0001	Abrera	37235	0.133372	-0.73587	0.042786	-0.35974	-0.01417	-1.15465	2.117033	-1.04664	-1.41343
...											

Section B2. Graph time-series

The script TS_PLOT.R uses the previous standardized time-series of individual monitoring stations within the entire river basin dataset to plot water quality variables of interest.

Input Data

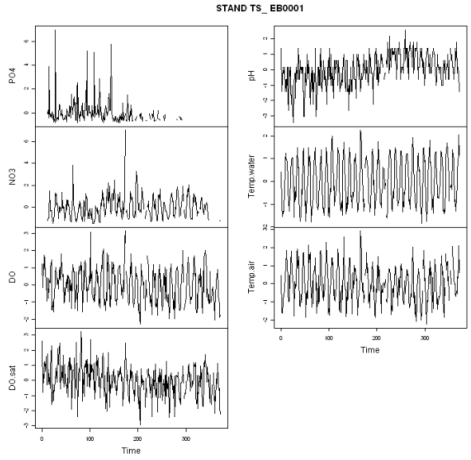
OUTPUT.standEmonth.csv

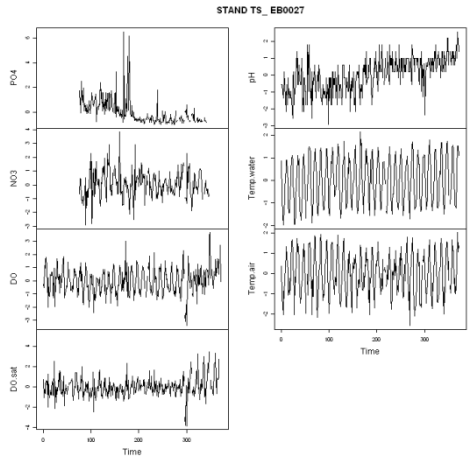
Code	*Date*	*NH4*	*Chlor*	*Cond*	*BOD5*	*COD.MnO4*	*PO4*	*NO3*	*DO*	*DO.sat*	*pH*	*Temp.water*	*Temp.air*
EB0001	198201	-0.57149	-0.68249	-1.05703	-0.97701	-0.71659	-0.00398	1.494959	0.172041	-0.01017	-2.64115	-0.85059	-0.98759
EB0001	198202	-0.78651	-0.79102	-1.67104	-0.97701	0.126159	-0.31699	-0.15853	0.116074	-0.53991	0.182277	-1.29948	-0.95934
EB0001	198203	0.134992	-1.33367	-2.05795	-0.41932	0.126159	-0.04311	-0.37506	1.06751	0.615894	-0.62442	-1.24561	0.453062
...													

R Script

```
################
## Read input data ##
################
input<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\WQ Time
Series\\Output.stand\\OUTPUT.standEmonth.csv",header=T)

summary(input)
nrow(input);ncol(input)
class(input)
names(input)

########################
##  WQ stations - unique codes ##
########################
codes<-as.matrix(unique(input$Code))
codes

############################
##  Loop  - Plot variables per station ##
############################
for (i in 1:nrow(codes)){
w<-which(input[,c(1)]==codes[i])

series<-ts(input[w,c(8:14)])  ##ncol(input)##

mypath <- file.path("C:\\Documents and Settings\\raguilera\\My Documents\\WQ Time Series\\Graphical
output\\EBRO STAND TS",paste("StandEmonth__", codes[i], ".jpg", sep = ""))

tiff(file=mypath, width = 700, height = 700)

mytitle = paste("STAND TS_", codes[i])

plot(series, main = mytitle)

dev.off()

} # end loop
```

Output Plot
Examples: *StandEmonth__EB0001.jpg* and *StandEmonth__EB0027.jpg*





Section B3. MINE

B3.1. Apply MINE to time-series in each water quality monitoring station

The script MINE_STATION.R applies the MINE algorithm to all available values of time-series in individual monitoring stations within a dataset.

Input data
R_JUCAR_2_monthly.csv

Code	*Name*	*Date*	*Actual.Date*	*Q*	*Chlor*	*Cond*	*BOD5*	*COD*	*PO4*	*NO3*	*NO2*	*Temp.air*	*Temp.water*
JUO502	Monovar (antigua COCA)	199403	34411	0.1	1142	3320	62	153	NA	2.7	1.87	26.4	8.2
JUO502	Monovar (antigua COCA)	199404	34444	0.4	886	3380	5	186	NA	2.37	NA	15.9	9.7
JUO502	Monovar (antigua COCA)	199405	34471	0.5	1106	4680	70	78	NA	0.6	NA	21.9	15.4
…													

R Script

```
#####################
##  Install rJava and MINE ##
#####################
setwd("C:/Documents and Settings/raguilera/My Documents/R Files/MINE")
install.packages("rJava") # 1-time initialization step
source("MINE.R")
```

```
########################################
##  Save station data as individual files in new directory ##
########################################
input<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\WQ Time Series\\Copy of
Data\\R_JUCAR_2_monthly.csv",header=T)
nrow(input);ncol(input)

#######################
## WQ stations - unique codes ##
#######################
codes<-as.matrix(unique(input$Code))
dim(codes)

############################
## Create empty matrix to store results ##
############################
station<-matrix(nrow=0,ncol=0,byrow=F)

############################
## Loop to read and save separate files ##
############################
for (i in 1:nrow(codes)){
w<-which(input[,c(1)]==codes[i])
station<-(input[w,])

setwd("C:\\Documents and Settings\\raguilera\\My Documents\\R Files\\MINE per
station\\JUCAR\\Stations")

write.csv(station, file=paste("Station_",codes[i],".csv", sep=""),row.names=F)

#######################
## Another way of saving files:  ##
#######################
## mypath <- file.path("C:\\Documents and Settings\\raguilera\\My Documents\\R Files\\MINE per
station\\JUCAR\\Stations_JUCAR",paste("Station_",codes[i],".csv", sep=""),sep = ",")
## write.csv(station, file=mypath, row.names=F)

} # end loop

############################
## Create a list of all individual files  ##
############################
mylist<-list.files("C:\\Documents and Settings\\raguilera\\My Documents\\R Files\\MINE per
station\\JUCAR\\Stations")
length(mylist)

##############################
##  Run MINE for individual station files ##
##############################
for (i in 1:length(mylist)) {
MINE(mylist[i],"all.pairs",0)
} # end loop

# 'all.pairs' will cause MINE to compare all pairs of variables against each other
```

B3.2. Calculate minimum MIC score for time-series in each water quality monitoring station

The script MINE_VN.R finds the minimum MIC scores required for the results in individual monitoring stations within a dataset to be significant.

R Script

```
####################
##  Set wd and create list  ##
####################
setwd("C:\\Documents and Settings\\raguilera\\My Documents\\R Files\\MINE per
station\\EBRO\\Stations")
mylist<-list.files(getwd())

########################################
## Create list of data frame names without the ".csv" part ##
########################################
names <-substr(mylist,1,16)
names

n<-matrix(nrow=0,ncol=0,byrow=F)
MIC_0.05<-matrix(nrow=0,ncol=0,byrow=F)
MIC_0.01<-matrix(nrow=0,ncol=0,byrow=F)
MIC_0.001<-matrix(nrow=0,ncol=0,byrow=F)
vn<-matrix(nrow=0,ncol=0,byrow=F)

##############
## Load all files ##
##############
for(i in names){

filepath <- file.path("C:\\Documents and Settings\\raguilera\\My Documents\\R Files\\MINE per
station\\EBRO\\Stations",paste(i,".csv",sep=""))

y<-as.matrix(assign(i, read.csv(filepath)))

##############################
## Remove "Code" and "Name" columns ##
##############################
y<-(y[,-(1:2)])

########################
## Coerce from "character" type ##
########################
mode(y)<-"double"

##################################
## Calculate n and MIC for different p-values ##
##################################
n<-as.matrix(combn(ncol(y),2,function(x) sum(y[,x[1]] & y[,x[2]],na.rm=T)))
print(n)

# save(n, file = "n.RData")
# write.csv(n,file="nloop.csv")

} # end loop
```

```
MIC_0.05<-as.matrix(lapply(n,function(x) 1.2594*(x)^-0.306))
#print(MIC_0.05)
#write.csv(MIC_0.05,file="MIC_0.05.csv")

MIC_0.01<-as.matrix(lapply(n,function(x) 1.7217*(x)^-0.348))
#print(MIC_0.01)
#write.csv(MIC_0.01,file="MIC_0.01.csv")

MIC_0.001<-as.matrix(lapply(n,function(x) 2.4104*(x)^-0.393))
#print(MIC_0.001)
#write.csv(MIC_0.001,file="MIC_0.001.csv")

######################
## Bind n and MIC p-values ##
######################
vn<-as.matrix(cbind.data.frame(n,MIC_0.05,MIC_0.01,MIC_0.001))
#print(vn)
#write.csv(vn,file="vnloop.csv")
```

Section B4. Dynamic Factor Analysis

The script DFA_MARSS.R is based on the Case Study 4 included in the MARSS User Guide (Holmes et al., 2013). This script applies Dynamic Factor Analysis to all time-series of a specific water quality variable (in this case, Nitrate). The input data include individual monitoring station tables (.csv format).

R Script

```
############################
## Code chunk number 1: RUNFIRST##
############################
require(MARSS)
options(prompt=" ", continue=" ")
library(xtable)
tabledir="tables/"

############################
## Code chunk number 2: read.in.data ##
############################
## Load the data ##
mylist<-list.files("/home/raguilera/MARSS34/Data/EBRO")

## Create list of data frame names with the code part ##
names <-substr(mylist,1,14)
names <-names[c(1:50)]

## Create variable/station names ##
variables <-substr(names,9,14)

## Create empty matrices to store the data ##
my_data<-matrix(nrow=372,ncol=0,byrow=F)

## Load all files ##
for(i in names){

 filepath <- file.path("/home/raguilera/MARSS34/Data/EBRO",paste(i,".csv",sep=""))
 y<-as.matrix(assign(i, read.csv(filepath)))
```

```
  ## Select variable(s) of interest columns ##
  y<-(y[,("NO3")])

  ## Coerce from "character" type ##
  mode(y)<-"double"

  ## Bind data ##
  my_data<-(cbind(my_data,y))

} # end loop

##############################
## Code chunk number 3: transpose.data  ##
##############################
## Transpose data so time goes across columns ##
my_data = t(my_data)
rownames(my_data) = variables
## Get number of time series ##
N.ts = dim(my_data)[1]
N.ts
## Get length of time series ##
TT = dim(my_data)[2]
TT

#########################
## code chunk number 4: z.score  ##
#########################
Sigma = sqrt(apply(my_data, 1, var, na.rm=TRUE))
y.bar = apply(my_data, 1, mean, na.rm=TRUE)
dat.z = (my_data - y.bar) * (1/Sigma)
rownames(dat.z) = variables
dim(dat.z)

#########################
## Code chunk number 5: plotdata ##
#########################
par(mfcol=c(10,5), mar=c(3,4,1.5,0.5), oma=c(0.4,1,1,1))
for(i in variables){
  plot(dat.z[i,],xlab="",ylab="", bty="L", xaxt="n", pch=16, col="darkgreen", type="b")
  axis(1,12*(0:dim(my_data)[2])+1,1995+0:dim(my_data)[2])
  title(i)
  }

###################################
## Code chunk number 6: set.up.many.trends.echo ##
###################################
cntl.list = list(minit=200, maxit=10000000, conv.test.slope.tol=0.1, MCInit=FALSE, allow.degen=FALSE,
safe=FALSE)

## Set up forms of R matrices ##
levels.R = c("equalvarcov")

model.data = data.frame()
## Fit lots of models & store results ##
for(R in levels.R) {
  for(m in 3:3) {
    dfa.model = list(A="zero", R=R, m=m)

    kemz = MARSS(my_data, model=dfa.model, control=cntl.list,
              form="dfa", z.score=TRUE)
```

```
  model.data = rbind(model.data,
                data.frame(R=R,
                        m=m,
                        logLik=kemz$logLik,
                        K=kemz$num.params,
                        AICc=kemz$AICc,
                        stringsAsFactors=FALSE))
  assign(paste("kemz", m, R, sep="."), kemz)

 } # end m loop
} # end R loop


################################
##  Code chunk number 7: makemodeltable  ##
################################
## Calculate delta-AICc ##
model.data$delta.AICc = model.data$AICc - min(model.data$AICc)
## Calculate Akaike weights ##
wt = exp(-0.5*model.data$delta.AICc)
model.data$Ak.wt = wt/sum(wt)
## Sort results ##
model.tbl = model.data[order(model.data$AICc),-4]
## Calculate cumulative wts ##
model.tbl$Ak.wt.cum = cumsum(model.tbl$Ak.wt)
tmpaln = "c" #figure out the number of cols automatically
for(i in 1:ncol(model.tbl)) {tmpaln = paste(tmpaln,"c",sep="")}
thetable = xtable(model.tbl,caption='Model selection results.', label='tab:tablefits', align=tmpaln,
digits=c(1,1,1,1,0,1,2,2))
align(thetable) = "cp{3.5cm}p{0.5cm}p{1.5cm}p{0.9cm}ccc"
setwd("/home/raguilera/MARSS37/EBRO")
print(thetable, type = "latex", file = paste(tabledir,"tablefitNO3EBRO_BMALL_37d.tex",sep=""),
include.rownames=FALSE,include.colnames=TRUE, caption.placement="top",table.placement="htp",
sanitize.text.function = function(x){x},hline.after = c(-1,0,nrow(model.data)))

############################
## Code chunk number 8: save.results ##
############################
setwd("/home/raguilera/MARSS37/EBRO/tables")
save(file="NO3EBRO_BMALL_37d--model.fits.Rdata",list=c("model.data", ls(pattern="^kemz.")))

############################
## Code chunk number 9: getbestmodel ##
############################
## Get the "best" model ##
best.model = model.tbl[1,]
fitname = paste("kemz",best.model$m,best.model$R,sep=".")
best.fit = get(fitname)

############################
## Code chunk number 10: varimax  ##
############################
## Get the inverse of the rotation matrix ##
H.inv = varimax(coef(best.fit, type="matrix")$Z)$rotmat


############################
## Code chunk number 11: rotations ##
############################
## Rotate factor loadings ##
```

```
Z.rot = coef(best.fit, type="matrix")$Z %*% H.inv
## Rotate trends ##
trends.rot = solve(H.inv) %*% best.fit$states

################################
## Code chunk number 12: plotfacloadings ##
################################
minZ = 0.05
ylims = c(-1.1*max(abs(Z.rot)), 1.1*max(abs(Z.rot)))
par(mfrow=c(2,2), mar=c(2,4,1.5,0.5), oma=c(0.4,1,1,1))
for(i in 1:best.model$m) {
      plot(c(1:N.ts)[abs(Z.rot[,i])>minZ], as.vector(Z.rot[abs(Z.rot[,i])>minZ,i]),
            type="h", lwd=2, xlab="", ylab="", xaxt="n", ylim=ylims, xlim=c(0,N.ts+1))
      for(j in 1:N.ts) {
            if(Z.rot[j,i] > minZ) {text(j, -0.05, variables[j], srt=90, adj=1, cex=0.9)}
            if(Z.rot[j,i] < -minZ) {text(j, 0.05, variables[j], srt=90, adj=0, cex=0.9)}
      abline(h=0, lwd=1, col="gray")
   } # end j loop
   mtext(paste("EBRO_NO3", "Factor loadings on trend",i,sep=" "),side=3,line=.5)
} # end i loop


############################
## Code chunk number 13: plottrends  ##
############################
## Get ts of trends ##
ts.trends = t(trends.rot)
#par(mfrow=c(ceiling(dim(ts.trends)[2]/2),2), mar=c(3,4,1.5,0.5), oma=c(0.4,1,1,1))
par(mfrow=c(2,2), mar=c(2,4,1.5,0.5), oma=c(0.4,1,1,1))
## Loop over each trend ##
for(i in 1:dim(ts.trends)[2]) {
      ## Set up plot area ##
      plot(ts.trends[,i],
            ylim=c(-1.1,1.1)*max(abs(ts.trends)),
            type="n", lwd=2, bty="L",
            xlab="", ylab="", xaxt="n", yaxt="n")
      ## Draw zero-line ##
      abline(h=0, col="gray")
      ## Plot trend line ##
      par(new=TRUE)
      plot(ts.trends[,i],
            ylim=c(-1.1,1.1)*max(abs(ts.trends)),
            type="l", lwd=2, bty="L",
            xlab="", ylab="", xaxt="n")
      ## Add panel labels ##
      mtext(paste("Trend",i,sep=" "), side=3, line=0.5)
      axis(1,12*(0:dim(my_data)[2])+1,1995+0:dim(my_data)[2])
      } # end i loop (trends)


#########################################################
### code chunk number 14: plotbestfits
#########################################################
par.mat=coef(best.fit, type="matrix")
fit.b = par.mat$Z %*% best.fit$states + matrix(par.mat$A, nrow=N.ts, ncol=TT)
variables = rownames(dat.z)


par(mfcol=c(10,5), mar=c(3,4,1.5,0.5), oma=c(0.4,1,1,1))
for(i in 1:length(variables)){
  plot(dat.z[i,],xlab="",ylab="",bty="L", xaxt="n", ylim=c(-4,3), pch=16, col="blue")
```

```
 axis(1,12*(0:dim(dat.z)[2])+1,1995+0:dim(dat.z)[2])
 lines(fit.b[i,], lwd=2)
 title(variables[i])
}
```

Section B5. Multitaper Spectral Analysis

B5.1. Estimate the power spectra of each DFA extracted pattern

The script SPEC_MTM.R calculates the power spectra of the DFA extracted patterns. Output data also includes F test values and degrees of freedom.

```
#############
## Read data  ##
#############
JUCAR<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\R Files
(cluster)\\TRENDS\\trends.JUCAR.csv",header=T)

Names<-names(JUCAR)
n.J<-ncol(JUCAR)

############################
##  Multitaper spectral analysis tools ##
############################
require(multitaper)

output_J<-matrix(nrow=257,ncol=0,byrow=T)   # Empty matrix to store results

## Loop to calculate spectra with multitaper method ##
for (i in 1:n.J){
 par(mar=c(2.5, 2.5, 2.5, 2.5))
 w<-which(colnames(JUCAR)==Names[i])
 spectra_J<-spec.mtm(ts(JUCAR[,w]),deltat=1/12, dtUnits="month", Ftest=TRUE)

 Ftest_<-spectra_J$mtm$Ftest
 Freq_<-spectra_J$freq
 dofs_<-spectra_J$mtm$dofs
 FF<-cbind(Freq_,Ftest_)

   output_J<-cbind(output_J,FF)

 # png(file=paste("mygraphic", sep = "_",Names[i], "png"),width=400,height=350)

} # end loop
 col.names_<-rep(Names, each = 2)

##############
## Save output   ##
##############

 FF_J<-rbind(col.names_,output_J)
 write.table(FF_J,file="FF_JUCAR5.csv", sep=",", row.names=FALSE, col.names=FALSE)
```

B5.2. Find critical values based on the F distribution table and degrees of freedom (dofs) of MTM results

The script FTEST.R finds critical values for the F distribution (p=0.05)

Input data

ftable.csv

0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	161.448	199.5	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19	19.164	19.247	19.296	19.33	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
…										

R Script

```
#############
## Read data  ##
#############
ftable<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\R Files
(cluster)\\TRENDS\\multitaper\\ftable.csv",header=T)

dofs<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\R Files
(cluster)\\TRENDS\\multitaper\\dofs_LLOBtemp.csv",header=F)

NN<-nrow(dofs)

my.dofs<-round(dofs, digits=0)
my.dofs

out<-vector(length=0)

##################
##  Find critical value  ##
##################
for (i in 1:NN){
  w<-which(ftable[,1]== my.dofs[i,1])
  critical.value<-ftable[w,3]

  out<-cbind(out, critical.value)

} # end loop

##############
## Save output  ##
##############
dim(out)

out

write.table(t(out),file="critical_value_LLOBtemp.csv", sep=",", row.names=FALSE, col.names=FALSE)
```

Section B6. Yue-Pilon Trend Analysis

The script TRENDS_YUE.R finds significant trends embedded in DFA extracted patterns.

R Script

```
#############
## Read data  ##
#############
EBRO<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\R Files
(cluster)\\TRENDS\\trends.EBRO.csv",header=T)

names(EBRO)
dim(EBRO)

########################################
##  Yue-Pilon - Prewhitened Nonlinear Trend Analysis ##
########################################

require(zyp)

EBROdf<-as.data.frame(EBRO)
EBRO.YUE<-apply(EBRO,2,zyp.yuepilon)

##############
## Save output  ##
##############
write.table(EBRO.YUE,file="EBRO.YUE.csv", sep=",")
```

Section B7. Generalized Least Squares

The script GLS_FL.R fits gls models while performing forward stepwise selection of significant explanatory variables.

R Script

```
################
## Load R-package ##
################
require(nlme)

###############
## Load my data  ##
###############
EBRO_GLS_PO4<-read.csv("C:\\Documents and Settings\\raguilera\\My Documents\\R
Files\\GLS_DFA_May2014\\EBRO\\PO4pointsGLSMay2014_log.csv",header=T)

names(EBRO_GLS_PO4)
dim(EBRO_GLS_PO4)
```

```
##############################
##  Run gls with all explanatory variables ##
##############################
for(i in 09:length(EBRO_GLS_PO4)){

  x <- as.matrix(EBRO_GLS_PO4[i])
  gls_<-gls(EBFL_1_PO4~ x, correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)

  print(summary(gls_)$tTable)
} # end loop

##############################
## Run gls with most significant variable  ##
##############################
## check x ##
gls_1check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US , correlation=corGaus(form = ~
PMSPX30+PMSPY30), EBRO_GLS_PO4)
summary(gls_1check)

#######################
## Find 2nd significant variable ##
#######################
## 2 expvar ##

for(i in 09:length(EBRO_GLS_PO4)){
  if (i == 45) next # skip log_Areal_Sum_MINE_KG_US as x2
  else{

  x2 <- as.matrix(EBRO_GLS_PO4[i])
  gls_2<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + x2, correlation=corGaus(form = ~
PMSPX30+PMSPY30), EBRO_GLS_PO4)

  print(summary(gls_2)$tTable)
} # end if
} # end loop

## check x2 ##
gls_2check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC,
correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)
summary(gls_2check)

#######################
## Find 3rd significant variable ##
#######################
## 3 expvar ##

for(i in 9:length(EBRO_GLS_PO4)){
  if (i == 45  | | i == 9) next # skip log_Areal_Sum_MINE_KG_US + PO4_meanCONC
  else{

    x3 <- as.matrix(EBRO_GLS_PO4[i])
    gls_3<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + x3,
correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)

    print(summary(gls_3)$tTable)
}
}
```

```
## check x3 ##
gls_3check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US,
correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)
summary(gls_3check)

#########################
## Find 4th significant variable ##
#########################
## 4 expvar ##

for(i in 9:length(EBRO_GLS_PO4)){
  if (i == 45  || i == 9 || i == 54) next # skip log_Areal_Sum_MINE_KG_US + PO4_meanCONC +
runoff_index_US
  else{

    x4 <- as.matrix(EBRO_GLS_PO4[i])
    gls_4<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US + x4,
correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)

    print(summary(gls_4)$tTable)
  } end if
} end loop

## check x4 ##
gls_4check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc, correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)
summary(gls_4check)

#########################
## Find 5th significant variable ##
#########################

for(i in 9:length(EBRO_GLS_PO4)){
  if (i == 45  || i == 9 || i == 54 || i == 24) next # skip log_Areal_Sum_MINE_KG_US + PO4_meanCONC
+ runoff_index_US + log_Industrial_pc
  else{

    x5 <- as.matrix(EBRO_GLS_PO4[i])
    gls_5<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc + x5, correlation=corGaus(form = ~ PMSPX30+PMSPY30), EBRO_GLS_PO4)

    print(summary(gls_5)$tTable)
  } end if
} end loop

## check x5 ##
gls_5check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc + log_Areal.Sum_Point_US, correlation=corGaus(form = ~ PMSPX30+PMSPY30),
EBRO_GLS_PO4)
summary(gls_5check)

#########################
## Find 6th significant variable ##
#########################
## 6 expvar ##
# drop PO4_meanCONC #

for(i in 9:length(EBRO_GLS_PO4)){
  if (i == 45  || i == 9 || i == 54 || i == 24 || i == 49) next # skip log_Areal_Sum_MINE_KG_US +
PO4_meanCONC + runoff_index_US + log_Industrial_pc + log_Areal.Sum_Point_US
  else{
```

```
   x6 <- as.matrix(EBRO_GLS_PO4[i])
   gls_6<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc + log_Areal.Sum_Point_US + x6, correlation=corGaus(form = ~ PMSPX30+PMSPY30),
EBRO_GLS_PO4)

   print(summary(gls_6)$tTable)
  } end if
} end loop

## check x6  ##
gls_6check<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc + log_Areal.Sum_Point_US + log_buf_PMEDIAD, correlation=corGaus(form = ~
PMSPX30+PMSPY30), EBRO_GLS_PO4)
summary(gls_6check)

########################
## Find 7th significant variable ##
########################
## 7 expvar ##
# drop PO4_meanCONC #

for(i in 9:length(EBRO_GLS_PO4)){
  if (i == 45  | | i == 9 | | i == 54 | | i == 24 | | i == 49 | | i == 11) next # skip log_Areal_Sum_MINE_KG_US
+ PO4_meanCONC + runoff_index_US + log_Industrial_pc + log_Areal.Sum_Point_US + log_buf_PMEDIAD
  else{

   x7 <- as.matrix(EBRO_GLS_PO4[i])
   gls_7<-gls(EBFL_1_PO4~ log_Areal_Sum_MINE_KG_US + PO4_meanCONC + runoff_index_US +
log_Industrial_pc + log_Areal.Sum_Point_US + log_buf_PMEDIAD + x7, correlation=corGaus(form = ~
PMSPX30+PMSPY30), EBRO_GLS_PO4)

   print(summary(gls_7)$tTable)
  } end if
} end loop

#####################################
##  Calculate pseudo-R-squared for final gls model  ##
#####################################

# r.squaredLR {MuMIn} #
# Likelihood-ratio based pseudo-R-squared #

require (MuMIn)

r.squaredLR(gls_6check)

plot(gls_6check)
```

Appendix C

TIME-SERIES DECOMPOSITION AND THE EFFECTS OF SAMPLING DATA RESOLUTION

In order to address the issue of sampling intervals and gaps in limnological time-series, we took advantage of an ongoing long-term research program at Lake Müggelsee (Berlin, Germany), in which data on plankton and physico-chemical variables are collected at high time resolution. We tested whether Dynamic Factor Analysis was able to reproduce the same underlying patterns if the resolution of the observations was reduced from weekly to monthly. Furthermore, analyses of biological variables were contrasted with those of physico-chemical variables as they typically differ in frequency variability. Finally, we aimed to determine whether the complex dynamics involving a set of variables with well-known interactions could be described using DFA and differing sampling frequencies.

Effect of sampling resolution

Univariate analysis

Univariate analyses were performed so that patterns would be extracted from time-series of different types (biological, physical, and chemical variables) as well as different sampling resolutions (weekly, biweekly, and monthly data). Thus, not only the effect of changing sampling resolution was assessed but also the behavior of different types of variables.

Multivariate analysis

A well established and studied relationship between measured variables at Lake Müggelsee was selected to perform multivariate DFA for weekly and monthly data. The study deals with the abrupt increase of a copepods species, *Cyclops kolensis*, when abundances of competing larger *Cyclops vicinus* fell below a critical threshold, which in turn was triggered by a change in the abundance of cryptophytes (Scharfenberger *et al.*, 2013).

Results

Univariate analysis

The reduction in resolution (from weekly to biweekly and monthly) did not significantly change the DFA results for a physical variable such as water temperature (data not shown), which exhibits relatively little high-frequency variability. The analysis of the total algal biomass in Lake Müggelsee, however, revealed a drastic change in the overall pattern of the time-series after the resolution was reduced from weekly (Fig.C1a), to biweekly (Fig.C1b) and monthly (Fig.C1c).

Multivariate analysis

Two coexisting copepods abundances (ind L^{-1}), *Cyclops vicinus* and *Cyclops kolensis*, and cryptophytes availability (Scharfenberger *et al.*, 2013) with weekly and monthly data resolution were subjected to DFA. Model selection based on AIC determined that weekly data were best described by two common patterns (Fig. C2a). According to the factor loadings, *Cyclops vicinus* (Vicinus) and cryptophytes (Crypto) seemed to be best described by Pattern 1, which decreased across time (Fig. C2b). On the other hand, *Cyclops kolensis* (Kolensis) was best characterized by the Pattern 2, which can be said to be fairly opposite to the first one.

The monthly data yielded one common pattern, where Vicinus and Crypto followed a decreasing trend with time and Kolensis followed the opposite trend (negative factor loading, depicted as dotted line in Fig. C3a). The factor loadings of the monthly data model (Fig. C3b) are about one order of magnitude lower than those in the previous weekly model. Therefore, monthly data was not able to represent the underlying common pattern(s) among the three time-series as well as the weekly data resolution model. Nevertheless, the overall pattern seemed to coincide with what Scharfenberger et al. (2013) have previously stated, which is that *C. vicinus* decline followed the driver threshold scenario of regime shift theory, caused by an abrupt change in a driver (in this case, decline in cryptophytes availability).

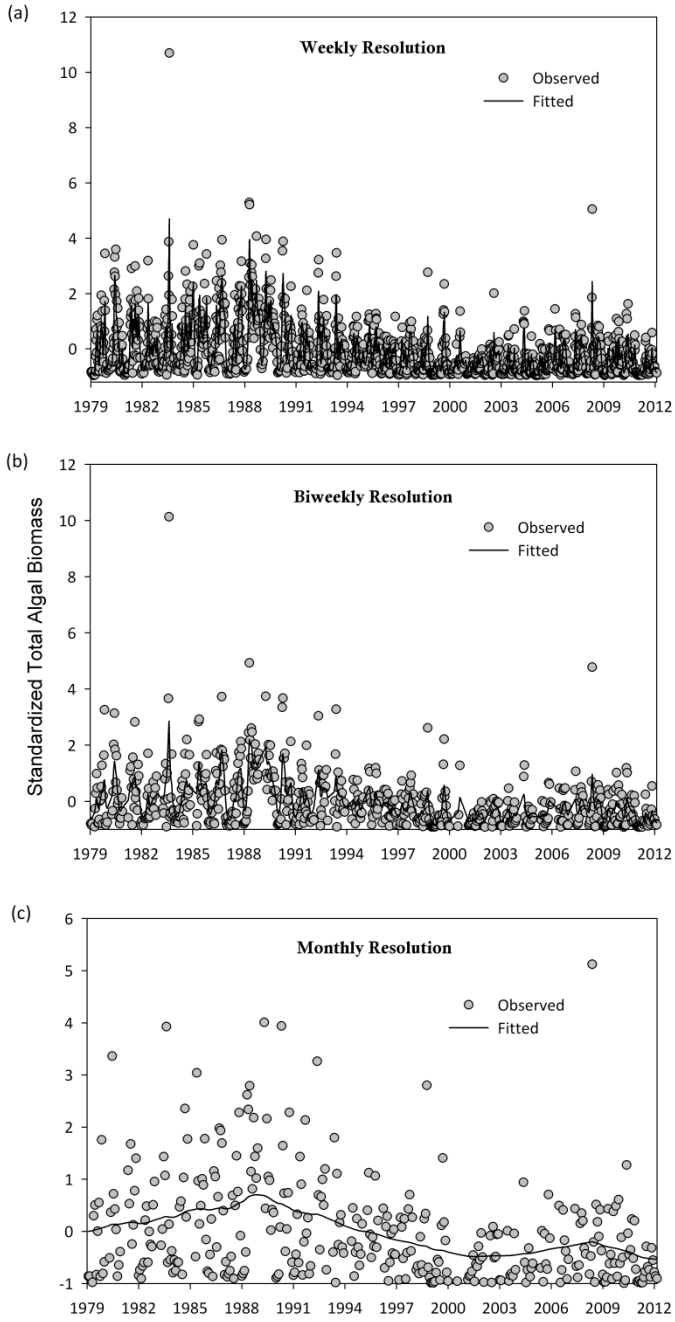


Figure C1: Extracted patterns (lines) fitted to the weekly (*a),* biweekly *(b)* and monthly (*c*) standardized data points (dots) for total algal biomass in Müggelsee.

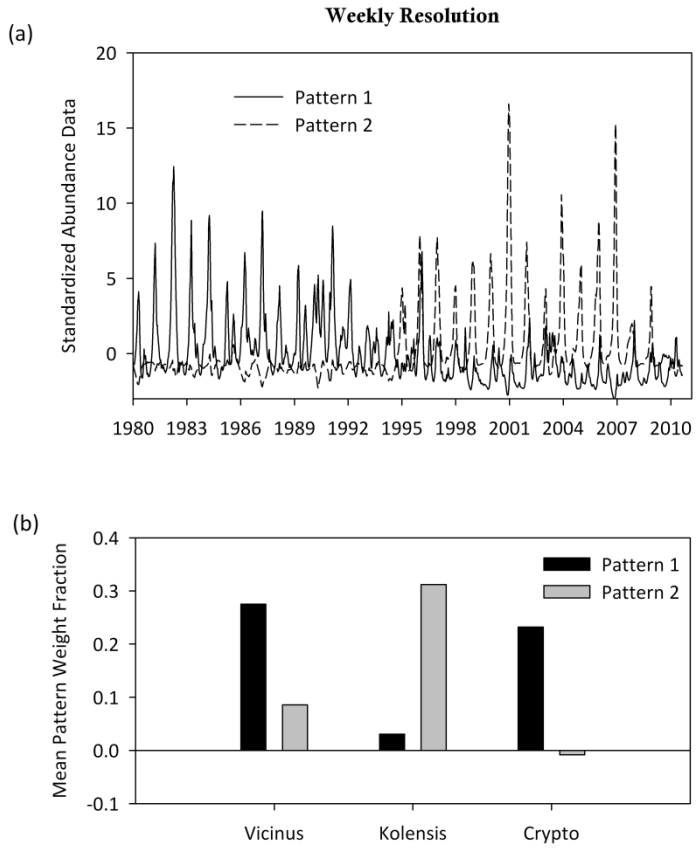


Figure C2: *(a)* The weekly data resolution model extracted two common patterns. *(b)* Pattern 1 mainly describes the *C. vicinus* and cryptophytes time-series whereas Pattern 2 describes the temporal variability of *C. kolensis* abundance, according to the highest corresponding Factor Loadings (weights) for each variable.

Based on the current results, it can be said that sampling frequency could be a limiting factor when studying interactions and processes in aquatic systems, and that monthly resolution can be insufficient. This is particularly true for variables that involve species counts or absence/presence data. As previously stated by (Adrian *et al.*, 2012), not only are sufficiently long datasets required to study complex systems, but the data also have to be measured at an appropriate temporal resolution.

Nonetheless, if the question at hand is to detect long-term patterns, the monthly resolution has proved to be adequate in representing the overall decline in some variables in total algal biomass (Fig.C1c) starting in the 1990s, which has been observed in previous studies (Köhler *et al.*, 2005).

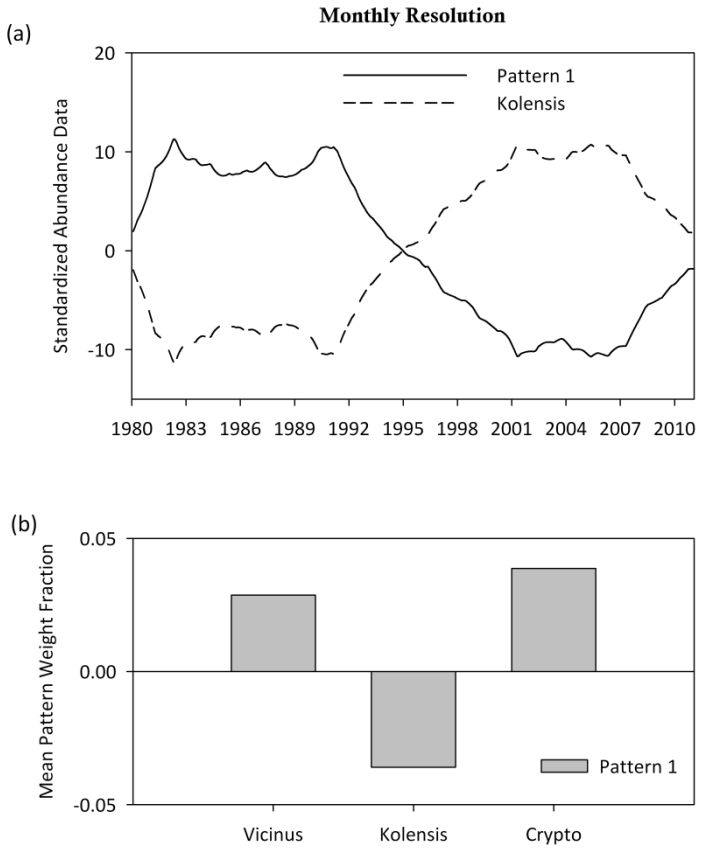


Figure C3: Best model fit for the monthly data. *(a)* One common pattern was extracted; the dotted line represents the opposite pattern related to the negative Factor Loading of *C. kolensis (b)*.

Acknowledgements

This work was part of a Short-Term Scientific Mission at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (Berlin, Germany) funded by the ESSEM COST Action ES1201 NETLAKE (Networking Lake Observatories in Europe) and supervised by Rita Adrian and David M. Livingstone.

References

References

Adrian R, Gerten D, Huber V, Wagner C, Schmidt S (2012) Windows of change: temporal scale of analysis is decisive to detect ecosystem responses to climate change. *Marine Biology,* **159**, 2533-2542.

Aguilera R, Marcé R, Sabater S (2013) Modeling nutrient retention at the watershed scale: Does small stream research apply to the whole river network? *Journal of Geophysical Research: Biogeosciences,* **118(2)**, 728-740.

Alexander RB, Böhlke JK, Boyer EW *et al.* (2009) Dynamic modeling of nitrogen losses in river networks unravels the coupled effects of hydrological and biogeochemical processes *Biogeochemistry,* **93,1/2**, 91-116.

Alexander RB, Elliott AH, Shankar U, Mcbride GB (2002a) Estimating the sources and transport of nutrients in the Waikato River Basin, New Zealand. *Water Resources Research,* **38 (12)**, 4-1.

Alexander RB, Johnes PJ, Boyer EW, Smith RA (2002b) A Comparison of Models for Estimating the Riverine Export of Nitrogen From Large Watersheds. *Biogeochemistry,* **57 / 58**, 295- 339.

Alexander RB, Smith RA, Schwarz GE (2000) Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. *Nature,* **403**, 758-761.

Allan JD (2004) Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annu. Rev. Ecol. Evol. Syst.,* **35**, 257-284.

Allan JD, Erickson DL, Fay J (1997) The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biology,* **37**, 149-161.

American Public Health Association (APHA) (1995) Standard methods for the examination of water and wastewater. *American Public Health Association, American Water Works Association, and Water Pollution Control Federation*. 19th edition, Washington, D.C.

Argerich A, Johnson SL, Sebestyen SD *et al.* (2013) Trends in stream nitrogen concentrations for forested reference catchments across the USA. *Environmental Research Letters,* **8**, 8.

Avila A, Rodrigo A, Rodà F (2002) Nitrogen circulation in a Mediterranean holm oak forest, La Castanya, Montseny, northeastern Spain. *Hydrol. Earth Syst. Sci.,* **6 (3)**, 551-557.

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

Basu NB, Suresh P, Rao C *et al.* (2011) Spatiotemporal averaging of instream solute removal dynamics. *Water Resources Research,* **47**.

Batalla RJ, Gómez CM, Kondolf GM (2004) Reservoir-induced hydrological changes in the Ebro River basin (NE Spain). *Journal of Hydrology,* **290**, 117-136.

Beaulac MN, Reckhow KH (1982) An examination of land use-nutrient export relationships. *Water Resources Bulletin,* **18**, 1013-1024.

Benítez-Gilabert M, Alvarez-Cobelas M, Angeler Dg (2010) Effects of climatic change on stream water quality in Spain. *Climatic Change,* **103**, 339-352.

Bernal S, Belillas C, Ibáñez JJ, Àvila A (2013) Exploring the long-term response of undisturbed Mediterranean catchments to changes in atmospheric inputs through time series analysis. *Science of the Total Environment,* **458**, 535-545.

Bernot MJ, Dodds WK (2005) Nitrogen Retention, Removal, and Saturation in Lotic Ecosystems. *Ecosystems,* **8**, 442-453.

Bernot MJ, Tank JL, Royer TV, David MB (2006) Nutrient uptake in streams draining agricultural catchments of the midwestern United States. *Freshwater Biology,* **51**, 499-509.

Bers AV, Momo F, Scholss IR, 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. *Climatic Change,* **116**, 789-803.

Beven KJ (2002) Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes,* **16**, 189-206.

Billen G, Silvestre M, Grizzetti B *et al.* (2011) Nitrogen flows from European regional watersheds to coastal marine waters. In: *The European Nitrogen Assessment.* (eds Sutton Ma, Howard Cm, Erisman Jw, Billen G, Bleeker A, Grennfelt P, Grinsven Hv, Grizzetti. B). pp. 271-297.Cambridge University Press.

Boithias L, Acuña V, Vergoñós L, Ziv G, Marcé R, Sabater S (2014) Assessment of the water supply:demand ratios in a Mediterranean basin under different global change scenarios and mitigation alternatives. *Science of the Total Environment,* **470-471**, 567-577.

Boucher O, Myhre G, Myhre A (2004) Direct human influence of irrigation on atmospheric water vapour and climate. *Climate Dynamics,* **22**, 597-603.

Bouraoui F, Grizzetti B (2011) Long term change of nutrient concentrations of rivers discharging in European seas. *Science of the Total Environment,* **409** 4899-4916.

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

Bovolo CI, Blenkinsop S, Majone B *et al.* (2011) Climate change, water resources and pollution in the Ebro Basin: Towards an integrated approach. In: *The Ebro River Basin.* (eds Barceló D, Petrovic M). pp. 295-329. Berlin Heidelberg, Springer-Verlag.

Boyer EW, Alexander RB, Parton WJ *et al.* (2006a) Modeling Denitrification in Terrestrial and Aquatic Ecosystems at Regional Scales. *Ecological Applications,* **16**, 2123-2142.

Boyer EW, Howarth RW, Galloway JN, Dentener FJ, Green PA, Vörösmarty CJ (2006b) Riverine nitrogen export from the continents to the coasts. *Global Biogeochemical Cycles,* **20**.

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

Bukaveckas PA (2007) Effects of Channel Restoration on Water Velocity, Transient Storage, and Nutrient Uptake in a Channelized Stream *Environ. Sci. Technol.,* **41** 1570-1576.

Caille F (2009) Integrated environmental assessment of nutrient emissions in a Mediterranean catchment: A case study in La Tordera, Catalonia. PhD Thesis, Autonomous University of Barcelona, Bellaterra, Spain.

Caille F, Riera JL, Rosell-Melé A (2011) Modelling nitrogen and phosphorus loads in a Mediterranean river catchment (La Tordera, NE Spain). *Hydrol. Earth Syst. Sci. Discuss.,* **8**, 7555-7594.

Cardinale BJ (2011) Biodiversity improves water quality through niche partitioning. *Nature,* **472 (7341)**, 86-89.

Carpenter EJ, Dunham S (1985) Nitrogenous nutrient uptake, primary production, and species composition of phytoplankton in the Carmans River Estuary, Long Island, New York. *Limnology and Oceanography,* **30**, 513-526.

Carr GM, Neary JP (2008) *Water quality for ecosystem and human health*, UNEP/Earthprint.

Catalan Water Agency (2008) Water in Catalonia: Diagnosis and proposed actions. In: *Significant water management issues raised within the compilation of the River Basin District Management Plan for Catalonia.* (ed Catalan Regional Government) pp Page, Catalan Water Agency (ACA).

Cazelles, B., Chavez, M., Berteaux, D., *et al.* (2008). Wavelet analysis of ecological time series. Oecologia **156**: 287–304.

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

Cooper SD, Lake PS, Sabater S, Melack JM, Sabo JL (2013) The effects of land use changes on streams and rivers in mediterranean climates. *Hydrobiologia,* **719**, 383-425.

Coumou D, Rahmstorf S (2012) A decade of weather extremes. *Nature Climate Change,* **2**, 491-496.

Cox DR, Snell EJ (1989) *The analysis of binary data,* London, Chapman and Hall.

Crutzen PJ, Stoermer EF (2000) The "Anthropocene". *Global Change Newsl.,* **41**, 17-18.

Dodds WK, Oakes RM (2006) Controls on Nutrients Across a Prairie Stream Watershed: Land Use and Riparian Cover Effects. *Environmental Management,* **37**, 634-646.

Dodds WK, Welch EB (2000) Establishing nutrient criteria in streams. *J. N. Am. Benthol. Soc.,* **19**, 186-196.

Donner SD, Coe MT, Lenters JD, Twine TE, Foley JA (2002) Modeling the impact of hydrological changes on nitrate transport in the Mississippi River Basin from 1955 to 1994. *Global Biogeochemical Cycles,* **16**, 1-18.

Doyle MW (2005) Incorporating hydrologic variability into nutrient spiraling. *Journal of Geophysical Research,* **110** (G1).

Doyle MW, Stanley EH, Harbor JM (2003) Hydrogeomorphic controls on phosphorus retention in streams. *Water Resources Research,* **39** (6).

Eastman JL, Coughenour MB, Pielke RAS (2001) The regional effects of CO2 and landscape change using a coupled plant and meteorological model. *Global Change Biology,* **7**, 797-815.

Ellis MA, Trachtenberg Z (2014) Which Anthropocene is it to be? Beyond geology to a moral and public discourse. *Earth's Future,* **2**, 122-125.

Emerson DG, Vecchia AV, Dahl AL (2005) Evaluation of Drainage-Area Ratio Method Used to Estimate Streamflow for the Red River of the North Basin, North Dakota and Minnesota. *Scientific Investigations Report,* **2005-5017**.

Ensign SH, Doyle MW (2006) Nutrient spiraling in streams and river networks. *J. Geophys. Res,* **111** (G4).

Ensign SH, Mcmillan SK, Thompson SP, Piehler MF (2006) Nitrogen and Phosphorus Attenuation within the Stream Network of a Coastal, Agricultural Watershed. *J. Environ. Qual.,* **35**.

Estrada F, Perron P, Martínez-López B (2013a) Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience,* **6**, 1050-1055.

Estrada F, Perron P, Gay-García C, Martínez-López B (2013b) A Time-Series Analysis of the 20th Century Climate Simulations Produced for the IPCC's Fourth Assessment Report. *Plos One,* **8**, 1-10.

Farley KA, Jobbágy EG, Jackson RB (2005) Effects of afforestation on water yield: a global synthesis with implications for policy. *Global Change Biology,* **11**, 1565-1576.

Ferrer J, Pérez-Martín MA, Jiménez S, Estrela T, Andreu J (2012) GIS-based models for water quantity and quality assessment in the Júcar River Basin, Spain, including climate change effects. *Science of the Total Environment,* **440**, 42-59.

Floury M, Delattre C, Ormerod SJ, Souchon Y (2012) Global versus local change effects on a large European river. *Science of the Total Environment,* **441**, 220-229.

Foley JA, Defries R, Asner GP *et al.* (2005) Global consequences of land use. science. *Science,* **309**, 570-574.

Foster G (1996) Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal,* **112**, 1709.

Gallart F, Delgado J, Beatson SJV, Posner H, Llorens P, 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). *Physics and Chemistry of the Earth,* **36**, 655-661.

García-Ruiz JM, López-Moreno JI, Vicente-Serrano SM, Lasanta–Martínez T, Beguería S (2011) Mediterranean water resources in a global change scenario. *Earth-Science Reviews,* **105**, 121-139.

Gasith A, Resh VH (1999) Streams in Mediterranean Climate Regions: Abiotic influences and biotic responses to predictable seasonal events. *Annu. Rev. Ecol. Syst.*, 51-81.

Ghil M, Allen MR, Dettinger MD *et al.* (2002) Advanced Spectral Methdos for Climatic Time Series. *Reviews of Geophysics,* **40**, 1-41.

Giorgi F, Lionello P (2008) Climate change projections for the Mediterranean region. *Global and Planetary Change,* **63**, 90-104.

Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu J, Bai X, Briggs JM (2008) Global Change and the Ecology of Cities. *Science,* **319**, 756-760.

Grizzetti B, Bouraoui F, Marsily GD, Bidoglio G (2005) A statistical method for source apportionment of riverine nitrogen loads. *Journal of Hydrology,* **304**, 302-315.

Grizzetti B, Bouraoui F (2006) Assessment of Nitrogen and Phosphorus Environmental Pressure at European Scale. (eds Commission E, Centre D-Gjr). Institute for Environment and Sustainability - Rural, Water and Ecosystem Resources Unit. Report EUR 22526 EN.

Grizzetti B, Bouraoui F, Aloe A (2011) Changes of nitrogen and phosphorus loads to European seas. *Global Change Biology,* **18**, 769-782.

Haggard BE, Stanley EH, Storm DE (2005) Nutrient retention in a point-source-enriched stream. *J. N. Am. Benthol. Soc.,* **24 (1)**, 29-47.

Hall RO, Tank JL, Sobota DJ *et al.* (2009) Nitrate removal in stream ecosystems measured by 15N addition experiments: Total uptake. *Limnol. Oceanogr,* **54**, 653-665.

Halliday SJ, Wade AJ, Skeffington RA *et al.* (2012) An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. *Science of the Total Environment,* **434**, 186-200.

Harvey AC (1989) *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Hegerl GC, Hoegh-Guldberg O, Casassa G *et al.* (2010) Good Practice Guidance Paper on Detection and Attribution Related to Anthropogenic Climate Change. In: *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change.* (eds Stocker Tf, Field Cb, Qin D, Barros V, Plattner Gk, Tignor M, Midgley Pm, Ebi Kl), IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.

Helton AM, Poole GC, Meyer JL *et al.* (2011) Thinking outside the channel: modeling nitrogen cycling in networked river ecosystems. *Front Ecol Environ,* **9(4)**, 229-238.

Hirabayashi Y, Kanae S, Emori S, Oki T, Kimoto M (2008) Global projections of changing risks of floods and droughts in a changing climate. *Hydrological Sciences Journal,* **53**, 754-772.

Hirsch RM, Slack JR, Smith RA (1982) Techniques of trend analysis for monthly water-quality data. *Water Resources Research,* **18**, 107-121.

Holmes EE (2013) Derivation of the EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models. *arXiv preprint arXiv:1302.3919.*

Holmes EE, Ward E, Wills K (2012) MARSS: Multivariate Autoregressive State-space Models for analyzing Time-series Data. *The R Journal,* **4**, 11-19.

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

Holmes EE, Ward E, Wills K (2014) MARSS: Multivariate Autoregressive State-Space Modeling. R package version 3.9.

House WA, Denison FH (1998) Phosphorus dynamics in a lowland river. *Water Research,* **Jun 32**, 1819-1830.

Huber DB, Mechem DB, Brunsell NA (2014) The Effects of Great Plains Irrigation on the Surface Energy Balance, Regional Circulation, and Precipitation. *Climate,* **2**, 103-128.

Ibáñez C, Prat N, Duran C *et al.* (2008) Changes in dissolved nutrients in the lower Ebro river: causes and consequences. *Limnetica,* **27**, 131-142.

IPCC (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (eds Field Cb, Barros Vr, Dokken Dj, Mach Kj, Mastrandrea Md, Bilir Te, Chatterjee M, Ebi Kl, Estrada Yo, Genova Rc, Girma B, Kissel Es, Levy An, Maccracken S, Mastrandrea Pr, White Ll) pp Page, Cambridge, United Kingdom and New York, NY, USA.

Ito A (2012) Detection and attribution of global change and disturbance impacts on a tower-observed ecosystem carbon budget: a critical appraisal. *Environmental Research Letters,* **7**, 1-6.

Kalnay E, Cai M (2003) Impact of urbanization and land-use change on climate. *Nature,* **423**, 528-531.

Kang S, Lin H (2007) Wavelet analysis of hydrological and water quality signals in an agricultural watershed. *Journal of Hydrology,* **338**, 1-14.

Keener VW, Feyereisen GW, Lall U, Jones JW, Bosch DD, Lowrance R (2010) El-Niño/Southern Oscillation (ENSO) influences on monthly NO3 load and concentration, stream flow and precipitation in the Little River Watershed, Tifton, Georgia (GA). *Journal of Hydrology,* **381**, 352-363.

Klein Goldewijk K, Beusen A, Van Drecht G, De Vos M (2011) The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography,* **20**, 73-86.

Köhler J, Hilt S, Adrian R, Nicklisch A, Kozerski HP, Walz N (2005) Long-term response of a shallow, moderately flushed lake to reduced external phosphorus and nitrogen loading. *Freshwater Biology,* **50**, 1639-1650.

Kronvang B, Behrendt H, Andersen HE *et al.* (2009) Ensemble modelling of nutrient loads and nutrient load partitioning in 17 European catchments. *Journal of Environmental Monitoring,* **11**, 572–583.

Kundzewicz ZW, Krysanova V (2010) Climate change and stream water quality in the multi-factor context: An editorial comment. *Climatic Change,* **103**, 353-362.

Kundzewicz ZW, Robson AJ (2004) Change detection in hydrological records—a review of the methodology / Revue méthodologique de la détection de changements dans les chroniques hydrologiques. *Hydrological Sciences Journal,* **49**, 7-19.

Kuo YM, Jang CS, Yu HL, Chen SC, Chu HJ (2013) Identifying nearshore groundwater and river hydrochemical variables influencing water quality of Kaoping River Estuary using dynamic factor analysis. *Journal of Hydrology,* **486**, 39-47.

Lane LJ, Nichols MH, Osborn HB (1994) Time series analyses of global change data. *Environmental Pollution,* **83**, 63-68.

Lassaletta L, García-Gómez H, Gimeno BS, Rovira JV (2009) Agriculture-induced increase in nitrate concentrations in stream waters of a large Mediterranean catchment over 25 years (1981–2005). *Science of the Total Environment,* **407**, 6034-6043.

Lassaletta L, Romero E, Billen G, Garnier J, García-Gómez H, Rovira JV (2012) Spatialized N budgets in a large agricultural Mediterranean watershed: high loading and low transfer. *Biogeosciences,* **9**, 57-70.

Leip A, Achermann B, Billen G *et al.* (2011) Integrating nitrogen fluxes at the European scale. In: *The European Nitrogen Assessment.* (eds Sutton Ma, Howard Cm, Erisman Jw, Billen G, Bleeker A, Grennfelt P, Grinsven Hv, Grizzetti. B) pp 345-376. Cambrigde University Press.

Liquete C, Canals M, Ludwig W, Arnau P (2009) Sediment discharge of the rivers of Catalonia, NE Spain, and the influence of human impacts. *Journal of Hydrology,* **366**, 76-88.

Lovett GM, Burns DA, Driscoll CT *et al.* (2007) Who needs environmental monitoring? *Frontiers in Ecology and the Environment,* **5**, 253-260.

Ludwig W, Dumont E, Meybeck M, Heussner S (2009) River discharges of water and nutrients to the Mediterranean and Black Sea: Major drivers for ecosystem changes during past and future decades? *Progress in Oceanography,* **80**, 199-217.

Lyubchich V, Gel YR, El-Shaarawi A (2013) On detecting non-monotonic trends in environmental time series: a fusion of local regression and bootstrap. . *Environmetrics,* **24**, 209-226.

Marcé R, Armengol J (2009) Modeling nutrient in-stream processes at the watershed scale using nutrient spiralling metrics. *Hydrol. Earth Syst. Sci.,* **13**, 953-967.

Marcé R, Honey-Rosés J, Manzano A, Moragas L, Catllar B, Sabater S (2012) The Llobregat River Basin: A Paradigm of Impaired Rivers Under Climate Change Threats. In: *The Llobregat* (eds. Sabater S, Ginebreda A, Barceló D.) pp. 1-26. Springer Berlin Heidelberg.

Marcé R, Rodríguez-Arias MA, García JC, Armengol J (2010) El Niño Southern Oscillation and climate trends impact reservoir water quality. *Global Change Biology,* **16**, 2857–2865.

Martí E, Aumatell J, Godé L, Poch M, Sabater F (2004) Nutrient retention efficiency in streams receiving inputs from wastewater treatment plants. *J. Environ. Qual.,* **33**, 285-293.

Martí E, Riera JL, Sabater F (2010) Water Scarcity in the Mediterranean: Effects of Wastewater Treatment Plants on Stream Nutrient Dynamics Under Water Scarcity Conditions. *The Handbook of Environmental Chemistry,* **8**, 173-195.

Martí E, Sabater F, Riera JL *et al.* (2006) Fluvial nutrient dynamics in a humanized landscape. Insights from a hierarchical perspective. *Limnetica,* **25**, 513-526.

Meybeck M (2003) Global analysis of river systems: from Earth system controls to Anthropocene syndromes. *Phil. Trans. R. Soc. Lond. B,* **358**, 1935-1955.

Miller C, Magdalina A, Willows RI *et al.* (2014) Spatiotemporal statistical modelling of long-term change in river nutrient concentrations in England & Wales. *Science of the Total Environment,* **466-467**, 914-923.

Milly PCD, Dunne KA, Vecchia AV (2005) Global pattern of trends in streamflow and water availability in a changing climate. *Nature,* **438**, 347-350.

Mujeriego R, Compte J, Cazurra T, Gullón M (2008) The water reclamation and reuse project of El Prat de Llobregat, Barcelona, Spain. *Water Science and Technology,* **57**, 567-574.

Mulholland PJ, Helton AM, Poole GC *et al.* (2008) Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature,* **452**, 202-206.

Mulholland PJ, Robert O. Hall J, Sobota DJ *et al.* (2009) Nitrate removal in stream ecosystems measured by 15N addition experiments: Denitrification. *Limnol. Oceanogr,* **54**, 666-680.

Muñoz-Carpena R, Ritter A, Li YC (2005) Dynamic factor analysis of groundwater quality trends in an agricultural area adjacent to Everglades National Park. *Journal of Contaminant Hydrology,* **80**, 49-70.

Muñoz I, López-Doval JC, Ricart M *et al.* (2009) Bridging levels of pharmaceuticals in river water with biological community structure in the Llobregat river basin (Northeast Spain). *Environmental Toxicology and Chemistry,* **28**, 2706-2714.

Navarro-Ortega A, Acuña V, Batalla RJ *et al.* (2012) Assessing and forecasting the impacts of global change on Mediterranean rivers. The SCARCE Consolider project on Iberian basins. *Environmental science and pollution research international,* **19**, 918-933.

Newbold JD, Bott TL, Kaplan LA *et al.* (2006) Uptake of nutrients and organic C in streams in New York City drinking-water-supply watersheds. *J. N. Am. Benthol. Soc.,* **25**, 998-1017.

Nilsson C, Reidy CA, Dynesius M, Revenga C (2005) Fragmentation and flow regulation of the world's large river systems. *Science,* **308**, 405-408.

Niyogi DK, Koren M, Arbuckle CJ, Townsend CR (2007) Stream Communities Along a Catchment Land-Use Gradient: Subsidy-Stress Response to Pastoral Development. *Environ. Manage.,* **39**, 213-225.

O'Brien JM, Dodds WK (2010) Saturation of NO_3 uptake in prairie streams as a function of acute and chronic N exposure. *J. N. Am. Benthol. Soc.,* **29**, 627-635.

O'Brien JM, Dodds WK, C.Wilson K, Murdock JN, Eichmiller J (2007) The saturation of N cycling in Central Plains streams: 15N experiments across a broad gradient of nitrate concentrations. *Biogeochemistry,* **84**, 31-49.

Odum EP, Finn JT, Franz EH (1979) Perturbation Theory and the Subsidy-Stress Gradient. *BioScience,* **29**, 349-352.

Oeurng C, Sauvage S, Sánchez-Pérez JM (2010) Temporal variability of nitrate transport through hydrological response during flood events within a large agricultural catchment in south-west France. *Science of the Total Environment,* **409**, 140-149.

Oki T, Kanae S (2006) Global Hydrological Cycles and World Water Resources. *Science,* **313**, 1068-1072.

Oreskes N (2004) The scientific consensus on climate change. *Science,* **306**, 1686-1686.

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences *Science,* **263**, 641-646.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature,* **421**, 37-42.

Parry ML (ed) (2007) *Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Paudel R, Jawitz JW (2012) Does increased model complexity improve description of phosphorus dyanmics in a large treatment wetland? . *Ecological Engineering,* **42**, 283-294.

Peterson BJ, Wollheim WM, Mulholland PJ *et al.* (2001) Control of Nitrogen Export from Watersheds by Headwater Streams. *Science,* **292**, 86-90.

Petrovic M, Postigo C, Alda MLD, Ginebreda A, Gros M, Radjenovic J, Barceló D (2010) Occurrence and Fate of Pharmaceuticals and Illicit Drugs Under Water Scarcity. In: *Water Scarcity in the Mediterranean* (eds. Sabater S, Barceló D.) pp. 197-228. Springer Berlin Heidelberg.

Pielke RA (2005) Land Use and Climate Change. *Science,* **310**, 1625.

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

Pinheiro JC, Bates DM (2000) *Mixed effects models in S and S-PLUS*, Springer.

Prat N, Rieradevall M (2006) 25-years of biomonitoring in two mediterranean streams (Llobregat and Besòs basins, NE Spain). *Limnetica,* **25 (1-2)**, 541-550.

Preston SD, Alexander RB, Schwarz GE, Crawford CG (2011a) Factors affecting stream nutrient loads: A synthesis of regional Sparrow model results for the continental United States. *Journal of the American Water Resources Association (JAWRA),* **47**, 891-975.

Preston SD, Alexander RB, Woodside MD (2011b) Regional assessments of the Nation's water quality-Improved understanding of stream nutrient sources through enhanced modeling capabilities. US Department of the Interior, US Geological Survey.

Preston SD, Brakebill JW (1999) Application of Spatially Referenced Regression Modeling for the Evaluation of Total Nitrogen Loading in the Chesapeake Bay Watershed In: *Water Resources Investigations.* Baltimore, Maryland, U.S. Geological Survey.

Qian SS, Cuffney TF, Alameddine I, Mcmahon G, Reckhow KH (2010) On the application of multilevel modeling in environmental and ecological studies. *Ecology,* **91**, 355-361.

Qian SS, Reckhow KH, Zhai J, Mcmahon G (2005) Nonlinear regression modeling of nutrient loads in streams: A Bayesian approach. *Water Resources Research,* **41 (7)**.

Quinn P (2004) Scale appropriate modelling: representing cause-and-effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning. *Journal of Hydrology,* **291**, 197-217.

Rabalais NN, Turner RE, Díaz RJ, Justic D (2009) Global change and eutrophication of coastal waters. *ICES Journal of Marine Science,* **66**, 1528-1537.

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

Ramos MC, Martínez-Casasnovas JA (2005) Nutrient losses by runoff in vineyards of the Mediterranean Alt Penedès region (NE Spain). *Agriculture, Ecosystems and Environment,* **113**, 356-363.

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

Reckhow KH, Chapra SC (1983) *Engineering approaches for lake management. Volume 1 -- Data analysis and empirical modeling. ,* Boston, MA, Butterworth publishers.

Reshef D, Reshef Y, Finucan H *et al.* (2011) Detecting novel associations in large datasets. *Science,* **334** (6062), 1518-1524.

Ricart M, Guasch H, Barceló D *et al.* (2010) Primary and complex stressors in polluted mediterranean rivers: Pesticide effects on biological communities. *Journal of Hydrology,* **383**, 52-61.

Rode M, Arhonditsis G, Balin D, Kebede T, Krysanova V, Griensven AV, Zee Seatmvd (2010) New challenges in integrated water quality modelling. *Hydrological Processes,* **24**, 3447–3461.

Rodó X, Baert E, Comín FA (1997) Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation. *Climate Dynamics,* **13**, 275-284.

Romaní AM, Sabater S, Muñoz I (2011) The Physical Framework and Historic Human Influences in the Ebro River. In: *The Ebro River Basin.* (eds Barceló D, Petrovic M) pp. 1-20. Berlin Heidelberg, Springer-Verlag.

Rosenzweig C, Karoly D, Vicarelli M *et al.* (2008) Attributing physical and biological impacts to anthropogenic climate change. *Nature,* **453**, 353-357.

Runkel RL, Crawford CG, Cohn TA (2004) Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers. *U.S. Geological Survey Techniques and Methods,* **Book 4** 69 p.

Sabater S, Feio MJ, Graça MAS, Muñoz I, Romaní A (2009) The Iberian Rivers. In: *Rivers of Europe.* (eds Tockner K, Robinson C, Uhlinger U) pp Page., Academic Press.

Sabater S, Sabater F, Tomàs X (1987) Water quality and diatom communities in two Catalan rivers (N.E. Spain). *Water. Res.,* **21**, 901-911.

Sabater S, Tockner K (2010) Effects of Hydrologic Alterations on the Ecological Quality of River Ecosystems. In: *Water Scarcity in the Mediterranean: Perspectives Under Global Change.* (eds Sabater S, Barceló D) pp. 15-39. Berlin Heidelberg, Springer – Verlag.

Scharfenberger U, Mahdy A, Adrian R (2013) Threshold-driven shifts in two copepod species: Testing ecological theory with observational data. *Limnol Oceanogr,* **52**, 741-752.

Schoellhamer DH (2001) Singular spectrum analysis for time series with missing data. *Geophysical Research Letters,* **28**, 3187-3190.

Schoumans OF, Silgram M, Groenendijk P *et al.* (2009) Description of nine nutrient loss models: capabilities and suitability based on their characteristics. *Journal of Environmental Monitoring,* **11**, 506-514.

Schröter D, Cramer W, Leemans R *et al.* (2005) Ecosystem Service Supply and Vulnerability to Global Change in Europe. *Science,* **310**, 1333-1337.

Schwarz GE, Hoos AB, Alexander RB, Smith RA (2006) The SPARROW Surface Water-Quality Model-Theory, Applications and User Documentation. *U.S. Geological Survey, Techniques and Methods,* **6-B3**, 248 p. and CD -ROM.

Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, Gunten UV, Wehrli B (2006) Global Hydrological Cycles and World Water Resources *Science,* **313**, 1072-1077.

Seitzinger SP, Styles RV, Boyer EW *et al.* (2002) Nitrogen retention in rivers: model development and application to watersheds in the northeastern U.S.A. *Biogeochemistry,* **57/58**, 199-237.

Shumway RH, Stoffer DS (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis,* **3**, 253-264.

Shumway RH, Stoffer DS (2010) *Time series analysis and its applications: with R examples,* New York, Springer.

Sillmann J, Roeckner E (2008) Indices for extreme events in projections of anthropogenic climate change. *Climatic Change,* **86**, 83-104.

Smith RA, Alexander RB (2000) Sources of Nutrients in the Nation's Watersheds In: Managing Nutrients and Pathogens from Animal Agriculture In: *Proceedings from the Natural Resource, Agriculture, and Engineering Service Conference for Nutrient Management Consultants, Extension Educators, and Producer Advisors.* Camp Hill, Pennsylvania

Stevenson RJ, Sabater S (2010) Understanding effects of global change on river ecosystems: science to support policy in a changing world. *Hydrobiologia,* **657**, 3-18.

Stone D, Auffhammer M, Carey M *et al.* (2013) The challenge to detect and attribute effects of climate change on human and natural systems. *Climatic Change,* **121**, 381-395.

Straile D, Livingstone DM, Weyhenmeyer GA, George DG (eds) (2003) The response of freshwater ecosystems to climate variability associated with the North Atlantic Oscillation. In: *The North Atlantic Oscillation: Climate significance and environmental impact*. (eds. J. W. Hurrell, Y. Kushnir, G. Ottersen and M. Visbeck). pp. 263-279. *Amer. Geophys. Union, Geophys. Monogr. Ser.* **134**.

Stream Solute Workshop T (1990) Concepts and methods for assessing solute dynamics in stream ecosystems. *J. N. Am. Benthol. Soc.,* **9**, 95-119.

Tank JL, Rosi-Marshall EJ, Baker MA, Hall RO (2008) Are rivers just big streams? A pulse method to quantify nitrogen demand in a large river. *Ecology,* **89** 2935-2945.

Tatariw C, Chapman EL, Sponseller RA, Mortazavi B, Edmonds JW (2013) Denitrification in a large river: consideration of geomorphic controls on microbial activity and community structure. *Ecology,* **94**, 2249-2262.

Terrado M, Barceló D, Tauler R (2010) Multivariate curve resolution of organic pollution patterns in the Ebro River surface water-groundwater-sediment-soil system. *Analytica chimica acta,* **657(1)**, 19-27.

Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proceedings of the IEEE,* **70**, 1055-1096.

Tilman D, Fargione J, Wolff B *et al.* (2001) Forecasting Agriculturally Driven Global Environmental Change. *Science,* **292**.

Turner BL, Kasperson RE, Meyer WB *et al.* (1990) Two types of global environmental change. *Global Environmental Change,* **1**, 14-22.

U.S. Global Change Research Act (1990) In: *Public Law 101-606(11/16/90) 104 Stat. 3096-3104.* pp Page.

Vegas-Vilarrúbia T, Sigró J, Giralt S (2012) Connection between El Niño-Southern Oscillation events and river nitrate concentrations in a Mediterranean river. *Science of the Total Environment,* **426**, 446-453.

Verburg PH, Neumann K, Nol L (2011) Challenges in using land use and land cover data for global change studies. *Global Change Biology,* **17**, 974-989.

Villizzi L (2012) Abundance trends in floodplain fish larvae: the role of annual flow characteristics in the absence of overbank flooding. *Fundamental and Applied Limnology/Archiv für Hydrobiologie,* **181**, 215-227.

Vörösmarty CJ (2002) Global water assessment and potential contributions from Earth Systems Science. *Aquatic Sciences,* **64**, 328-351.

Vörösmarty CJ, Green P, Salisbury J, Lammers RB (2000) Global water resources: vulnerability from climate change and population growth. *Science,* **289**, 284-288

Vörösmarty CJ, Mcintyre PB, Gessner MO *et al.* (2010) Global threats to human water security and river biodiversity. *Nature,* **467**, 555-561.

Wellen C, Arhonditsis GB, Labencki T, Boyd D (2014) Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring. *Hydrological Processes,* **28**, 1260–1283.

Wilbanks TJ, Kates RW (1999) Global Change in Local Places: How Scale Matters. *Climatic Change,* **43**, 601-628.

Wollheim WM, Peterson BJ, Thomas SM, Hopkinson CH, Vörösmarty CJ (2008a) Dynamics of N removal over annual time periods in a suburban river network *J. Geophys. Res.,* **113** (G3).

Wollheim WM, Vörösmarty CJ, Bouwman AF *et al.* (2008b) Global N removal by freshwater aquatic systems using a spatially distributed, within-basin approach. *Global Biogeochemical Cycles,* **22**, 14.

Wollheim WM, Vörösmarty CJ, Peterson BJ, Seitzinger SP, Hopkinson CS (2006) Relationship between river size and nutrient removal. *Geophys. Res. Lett,* **33**.

Worrall F, Howden NJK, Burt TP (2014) Time series analysis of the world's longest fluvial nitrate record: evidence for changing states of catchment saturation. *Hydrological Processes*. doi: 10.1002/hyp.10164.

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.

Zuur AF, Fryer RJ, Jolliffe IT, Dekker R, Beukema JJ (2003) Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics,* **14**, 665-685.

Zuur AF, Ieno EN, Smith GM (2007) *Analysing ecological data,* New York, Springer.

Zuur AF, Pierce GJ (2004) Common trends in Northeast Atlantic Squid time series. *Journal of Sea Research,* **52**, 57-72.

Zuur AF, Tuck ID, Bailey N (2003b) Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences,* **60**, 542-552.