Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Contaminant Hydrology 80 (2005) 49–70
JOURNAL OF Contaminant Hydrology
www.elsevier.com/locate/jconhyd

# Dynamic factor analysis of groundwater quality trends in an agricultural area adjacent to Everglades National Park

R. Muñoz-Carpena [a,*], A. Ritter [a,b], Y.C. Li [c]

[a] *Agricultural and Biological Engineering Department, University of Florida, 101 Frazier Rogers Hall, PO Box 110570 Gainesville, FL 32611-0570, USA*
[b] *Dep. Ingeniería, Producción y Economía Agraria, Universidad de La Laguna, Ctra. Geneto, 2., 38200 La Laguna, Spain*
[c] *Soil and Water Science Department, Tropical Research and Education Center, University of Florida, 18905 SW 280 Street, Homestead, FL 33031, USA*

Received 9 September 2004; received in revised form 6 July 2005; accepted 13 July 2005

## Abstract

The extensive eastern boundary of Everglades National Park (ENP) in south Florida (USA) is subject to one of the most expensive and ambitious environmental restoration projects in history. Understanding and predicting the water quality interactions between the shallow aquifer and surface water is a key component in meeting current environmental regulations and fine-tuning ENP wetland restoration while still maintaining flood protection for the adjacent developed areas. Dynamic factor analysis (DFA), a recent technique for the study of multivariate non-stationary time-series, was applied to study fluctuations in groundwater quality in the area. More than two years of hydrological and water quality time series (rainfall; water table depth; and soil, ground and surface water concentrations of $N–NO_3^-$, $N–NH_4^+$, $P–PO_4^{3-}$, Total P, $F^-$and $Cl^-$) from a small agricultural watershed adjacent to the ENP were selected for the study. The unexplained variability required for determining the concentration of each chemical in the 16 wells was greatly reduced by including in the analysis some of the observed time series as explanatory variables (rainfall, water table depth, and soil and canal water chemical concentration). DFA results showed that groundwater concentration of three of the agrochemical species studied ($N–NO_3^-$, $P–PO_4^{3-}$and Total P) were affected by the same explanatory variables (water table depth, enriched topsoil, and

---

* Corresponding author.
*E-mail address:* carpena@ufl.edu (R. Muñoz-Carpena).

doi:10.1016/j.jconhyd.2005.07.003

occurrence of a leaching rainfall event, in order of decreasing relative importance). This indicates that leaching by rainfall is the main mechanism explaining concentration peaks in groundwater. In the case of N–$NH_4^+$, in addition to leaching, groundwater concentration is governed by lateral exchange with canals. $F^-$and $Cl^-$ are mainly affected by periods of dilution by rainfall recharge, and by exchange with the canals. The unstructured nature of the common trends found suggests that these are related to the complex spatially and temporally varying land use patterns in the watershed. The results indicate that peak concentrations of agrochemicals in groundwater could be reduced by improving fertilization practices (by splitting and modifying timing of applications) and by operating the regional canal system to maintain the water table low, especially during the rainy periods.

*Keywords:* Hydrology; Groundwater; Surface water; Water quality; Non-point source pollution; Dynamic factor analysis; Multivariate time series; Monitoring; Field methods; Everglades

## 1. Introduction

In the first half of the 20th century a complex drainage canal system was built in south Florida to protect urban and agricultural areas against flooding. However, this regional water management also led to the draining of protected natural wetland areas in the adjacent Everglades National Park (ENP) creating a negative impact on the environment. In an attempt to restore the wetland ecosystem of the ENP, the Combined Structural and Operational Project (CSOP) and the Comprehensive Everglades Restoration Plan (CERP) are being implemented along the extensive eastern boundary with the developed area (agricultural and urban) (SFWMD, 2004). The goal of these plans is to enhance water deliveries into the ENP while maintaining flood protection for developed areas. In addition, water quality is at the core of the restoration effort. Surface waters entering the ENP must not exceed a maximum regulatory level of total phosphorous of 0.010 mg $l^{-1}$ and other chemicals must be monitored as well (Florida Senate Bill 0626ER, 2003). Implementation of these projects is complex and requires detailed understanding of the hydrological processes involved. Predicting the water quality interactions between surface water flow in the canals and the shallow and extremely permeable Biscayne Aquifer (Fish and Stewart, 1991) is a special priority for ecosystem restoration of the Everglades and flood protection of urban and agricultural areas. Previous studies in the area (Genereux and Guardiario, 1998, 2001; Genereux and Slater, 1999) have shown the complexity of the groundwater system with extremely permeable materials and evidence of a very dynamic interaction between canals and the aquifer. Muñoz-Carpena et al. (2003), based on preliminary hydrological data (1-year) obtained in an agricultural area located at the boundary of the ENP, reported the almost instantaneous response of the groundwater to canal and rainfall inputs in the area as well as evidence of water quality interaction between canals, the shallow aquifer and land use. Detailed data sets containing temporal variation of hydrological and water quality variables have the potential to be used to understand the surface–groundwater–land use interactions in the area. However, interpretation of results from data analysis based on visual inspection and descriptive statistics is difficult and may not be sufficient, especially when dealing with multivariate time series.

Chemical fluctuations in shallow groundwater typically result from different cumulative effects, such as land use and associated chemical concentration in the topsoil, net vertical recharge (affected by leaching rainfall), local depth to groundwater, lateral recharge from ground or surface water sources, etc. Although some of these effects can be measured accurately, it is impractical to measure others, i.e., those with unstructured spatial and temporal distribution. An example of this is land use in an intensive commercial horticulture setting managed by different farmers. Typically land parcels can be combined or used independently for different crops and management practices (chemical application times and rates, irrigation, etc.), which vary from farmer to farmer. These combinations change from year to year depending on marketing, farmer specialty or preferences, etc. This generates the need for estimations by indirect methods applied to observed water quality data at fixed observation sites (Márkus et al., 1999).

Although standard multivariate analysis techniques are useful tools and can be adapted to analyze time series to obtain information about the interactions between variables, the time component of the data is ignored. A preferred method for studying multivariate time series is dynamic factor analysis (DFA), because it allows estimating common patterns and interactions in several time series and studying the effect of explanatory time-dependent variables as well (Zuur et al., 2003b). Multivariate time series may be analyzed as response variables assuming that there are common driving forces behind them, i.e., latent effects that determine the variation of the individual observations with time. These latent effects can be described by trends and/or explanatory variables. Dynamic factor analysis is a specialized time series technique originally developed for the study of economic models (Geweke, 1977) that has been recently used with variations in disciplines like psychology (e.g., Molenaar, 1985, 1989; Molenaar et al., 1992) and economics (Harvey, 1989; Lütkepohl, 1991). Lately, Zuur et al. (2003a), Zuur and Pierce (2004) and Erzini (2005) used dynamic factor analysis for fisheries applications, while Mendelssohn and Schwing (2002) applied it to large oceanographical time series. Although dynamic factor analysis has been recently applied in hydrology to identify common patterns of groundwater level (Márkus et al., 1999; Ritter and Muñoz-Carpena, 2005), there is no previous application to water quality studies. Analysis of large water quality datasets is complex because of the many effects affecting the chemical concentration variation in the system. DFA can be an effective methodology to handle such datasets and identifying the dominant effects controlling such variation.

The objective of this study is to apply DFA to study the interactions between monthly water quality time series and other hydrological variables obtained at an intensively monitored small agricultural watershed along the boundary of the ENP. Four agrochemical species of nitrogen and phosphorus, plus two natural tracers, Fluoride ($F^-$) and Chloride ($Cl^-$), were included in the analysis. The analysis was conducted in three steps: i) identification of common trends of groundwater quality; ii) inclusion of explanatory variables in the dynamic factor model; and iii) study of interactions between ground and surface water quality and canal management, hydrology and land use components.

## 2. Materials and methods

### 2.1. Experimental set-up

The study was conducted in the Frog Pond area, a small watershed of 2023 ha located along the boundary of Everglades National Park (ENP) in Homestead, Florida (Fig. 1). This public land was leased for the last 11 years to a group of growers that farmed under restricted conditions (low inputs and limited flood protection). The area is delimited by two canals that belong to the South Florida Water Management District (SFWMD) regional network: C-111 (West) and L-31W (East) (Fig. 1). Water level in both canals is regulated by remotely operated structures S-177 (spillway) and S-175 (culvert), respectively. Under CSOP operations, water level in canal L-31W is maintained high in order to increase water delivery into the ENP, while pushing agricultural return flows away to the east. This system influences surface and groundwater flow patterns and elevation in the area. Although farming practices vary with crop (sweet potato, sweet corn, green beans, malanga, okra, squash) and by individual farmer, the cropping season for the entire Frog Pond extends from the end of September through April, coinciding with the dry season.

An extensive monitoring network distributed across the southern portion of the Frog Pond watershed (780 ha south of the Torcise ditch, Fig. 1) was developed for this study. The first experimental phase of the University of Florida (UF) monitoring network was initiated in February 2002 with the installation of 10 instrumented wells for measuring water elevation, two rain gauges, soil moisture sensors and an automatic weather station along a 1.6 km transect. Groundwater levels were registered every 15 min by auto-logging pressure transducers compensated for temperature effects and atmospheric pressure (Solinst Inc., Canada). Fifteen minute rainfall readings were made with two auto-logging tipping-bucket rain gauges (Onset Computer Corp., USA) located at points 1/3 and 2/3 of the way along the main transect. A sampling location for each canal (C-111 and L-31W) was selected at each end of the transect.

In a second experimental phase started in February 2003, six additional instrumented wells were added north and south of the original transect and included in the water quality protocol described below (Fig. 1). These new wells were added to study the possible perturbations introduced by the newly constructed detention pond when operation started in summer 2003. To date, the detention pond has only been filled in June 2003. Surface water elevations in the canals were continuously recorded by a simple self-contained automatic recorder developed for this purpose (Schumann and Muñoz-Carpena, 2003). The loggers in the two canals were attached to custom-made steel and wood platforms (6 × 1 m) supported by pillars anchored in the banks and the bottom of the canal. Further details about the experimental area and set-up can be found in Ritter and Muñoz-Carpena (2005).

Surface and groundwater quality grab samples were collected in acid cleaned and pre-labeled 500 ml bottles every two weeks at each monitoring location (2 canals and 16 wells for the second phase; 2 canals and 10 wells for the first phase). QA/QC field and laboratory procedures were followed at all times (FL-DEP, 2002). The samples were stored in the field in a cooler with ice and transported to the laboratory within 2 h. The water

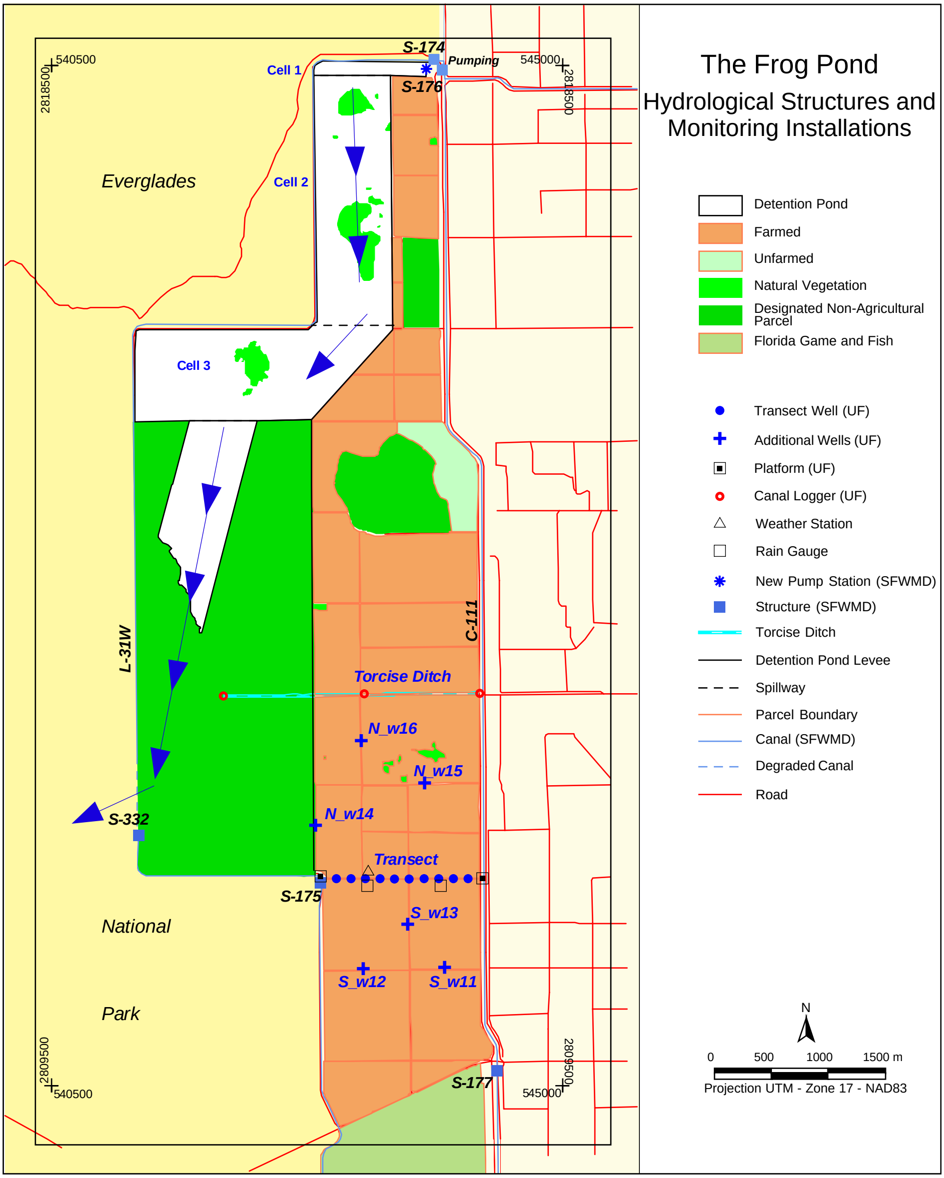

Fig. 1. The Frog Pond water quality monitoring network. Water quality samples were obtained from the transect and additional wells (UF), platforms and canal loggers locations (UF).

samples were prepared immediately on receipt and transferred in refrigeration before analysis. The samples were analyzed for concentrations of orthophosphate [$P{-}PO_4^{3-}$], total phosphorus [TP], ammonia–nitrogen [$N{-}NH_4^+$] and nitrate–nitrogen [$N{-}NO_3^-$] using an Autoanalyzer (AA3, Bran+Luebbe, Buffalo Grove, IL). In addition, fluoride [$F^-$] and chloride [$Cl^-$] were analyzed by ion chromatography (Dionex 500, Dionex Corporation, Sunnyvale, CA). Analytical precision for these elements was better than 3% RSD (Relative Standard Deviation).

Soil samples were collected every 6 months, at the beginning and end of the cropping season (i.e., at the end and beginning of the rainy season), from the land adjacent to each well. The soil samples were air-dried, grinded, sieved (<2 mm) and stored in plastic-lined paper bags before chemical analysis. Soil samples were digested according to US-EPA method 3050A and analyzed for TP. Nitrogen species (N–$NO_3^-$ and N–$NH_4^+$) in soils were extracted with 2 M KCl and analyzed using an Autoanalyzer. Fluoride, chloride and water soluble P in soil were extracted with water (1 : 5 soil to water ratio) and analyzed using an ion chromatograph.

### *2.2. Dynamic factor analysis, DFA*

Time series are time dependent data showing a systematic and a non-systematic variation. These are usually analyzed by decomposing the information, so that both types of variations (systematic and non-systematic) can be characterized. DFA is a statistical technique for the analysis of multivariate time series that first received this name from the pioneering work of Geweke (1977). It has been designed to identify underlying common trends or latent effects in several time series and interactions among them. Moreover, the DFA scheme used here (Zuur et al., 2003b) allows for evaluation of the effect of explanatory variables. DFA is similar to other dimension reduction techniques like factor analysis or redundancy analysis, but it takes into account the time component and thus it is designed to be used with non-stationary time series. Notice that these conventional multivariate statistical methods usually require independent observations, which is not the case for time series (Márkus et al., 1999). In addition, the order in time series is an important characteristic that must be taken into account, however conventional methods handle unordered data. The difference between DFA and the latter techniques is that in DFA the axes are restricted to be latent smoothing functions over time. The analysis is based on the so-called structural time series models (Harvey, 1989) that allow describing the time series of measured data of $N$ response variables with a Dynamic Factor Model (DFM) given by:

$$\begin{aligned} &N \text{ time series} = \text{linear combination of } M \text{ common trends} + \text{level parameter} \\ &\qquad + \text{explanatory variables} + \text{error component}. \end{aligned} \tag{1}$$

The aim of DFA is to choose $M$ as small as possible but still obtaining a reasonable fit. $M$ should be much smaller than $N$, because although increasing the numbers of common trends leads to a better model fit, it results in more information that needs to be interpreted. The mathematical formulation of this DFM is given by (Lütkepohl, 1991; Zuur et al., 2003b):

$$S_n(t) = \sum_{m=1}^{M} \gamma_{m,n}\alpha_m(t) + \mu_n(t) + \sum_{k=1}^{K} \beta_{k,n}v_k(t) + \varepsilon_n(t) \tag{2}$$

$$\alpha_m(t) = \alpha_m(t-1) + \eta_m(t) \tag{3}$$

where $s_n(t)$ is the value of the $n$th response variable at time $t$ (with $1 \leq n \leq N$); $\alpha_m(t)$ is the $m$th unknown trend (with $1 \leq m \leq M$) at time $t$; $\gamma_{m,n}$ represents the unknown factor loadings; $\mu_n$ is the $n$th constant level parameter for displacing up and down each linear combination of common trends (i.e., it is the intercept term in the regression DFM); $\beta_{k,n}$ stands for the unknown regression parameters (with $1 \leq k \leq K$) for the $K$ explanatory variables $v_k(t)$; $\varepsilon_n(t)$ and $\eta_m(t)$ are error components that are assumed to be independent of each other and normally distributed with zero mean and unknown covariance matrix. The error covariance matrix was selected here as a diagonal matrix. Notice that with this DFM (Eqs. (2) and (3)) if seasonal or cyclic components are present in the time series, they will be masked and included in the trend component (Eq. (3)). The unknown parameters were estimated using the Expectation Maximization (EM) algorithm (Dempster et al., 1977; Shumway and Stoffer, 1982; Wu et al., 1996). Technically, within the DFA framework, the trends are modeled as a random walk (Harvey, 1989) and estimations are performed using the Kalman filter/smoothing algorithm and the EM method, while the regression parameters associated with the explanatory variables are modeled as in linear regression (Zuur and Pierce, 2004). It is worth noting that the incorporation of explanatory variables results in a complete, unified description of the DFM within the EM framework (Zuur et al., 2003b). These techniques are implemented in the statistical software package Brodgar v2.3.3 (www.brodgar.com), which was used in this study. A complete and detailed description of this technique is given in Zuur et al. (2003b).

Results from the DFA were interpreted in terms of the estimated parameters $\gamma_{m,n}$ and $\beta_{k,n}$, the canonical correlations, and match between model estimations and observed values. The goodness-of-fit of the model can be assessed by visual inspection, the coefficient of efficiency (Nash and Sutcliffe, 1970) and Akaike's Information Criterion (AIC) (Akaike, 1974). The coefficient of efficiency ($C_{\text{eff}}$) has been widely used to evaluate the performance of hydrologic models. It compares the variance about the 1 : 1 line (perfect agreement) to the variance of the observed data (see Appendix A). The AIC is a statistical criterion for model selection. It combines the measure of fit with a penalty term based on the number of parameters used in the model. If more parameters (i.e., number of trends or explanatory variables) are used, the model fit is better, but the penalty for the extra parameters is higher as well. The smallest AIC indicates the most appropriate model.

The common trends, $\alpha_m(t)$, are functions that represent the patterns in the data that cannot be described with the explanatory variables included in the model. Factor loadings $\gamma_{m,n}$ indicate the weight of a particular common trend in the response time series, $s_n$. In addition, the comparison of factor loadings of different time series allows for detection of interactions between the different $s_n$. Canonical correlations coefficients ($\rho_{m,n}$) are used to quantify the cross-correlation between the response variables ($s_n$) and the common trends ($\alpha_m$). The terms "high", "moderate", and "weak" correlation are usually applied to $|\rho_{m,n}| > 0.75$, 0.50–0.75, and 0.30–0.50, respectively. The influence or weight of each explanatory variable $v_k$ on each $s_n$ is given by the regression parameters, $\beta_{k,n}$. Standard errors for the regression parameters are also included.

## 2.3. Water quality and hydrological time series and analysis procedure

### 2.3.1. Response variables

Sixteen groundwater chemical concentration (mg $l^{-1}$) time series for each chemical were obtained from the wells located along the main transect (T_w1–T_w10), south of the transect (S_w11, S_w12, S_w13) and north of it (N_w14, N_w15, N_w16) (Fig. 1). These were considered as response variables. Each of these biweekly time series was averaged monthly. This smoothing procedure favors the underlying common trends against local peaks and thus facilitates the analysis. Therefore, monthly averaged data (non-stationary) from a period of over 2-years (26 months, April 2002–May 2004) were used.

### 2.3.2. Explanatory variables

From a practical standpoint groundwater chemical variation is a function of chemical inputs, outputs and transformation. In the case of drained agricultural lands like those in the study, we can differentiate between two groups of chemicals based on their source. Products not used in agricultural production (here $F^-$and $Cl^-$) constitute the first group. Typically, the concentration changes for this group will be driven by lateral inflow and outflow to and from the canals, atmospheric deposition in coastal areas followed by rainfall leaching, chemical transformation, etc. For the second group, the agrochemicals (here N–$NH_4^+$, N–$NO_3^-$, P–$PO_4^3$, and TP), the relatively large concentration at which they are applied will frequently mask most of their natural variability. Shallow groundwater concentration for this group will be dominated by leaching from the topsoil which in turn depends on crop applications, mobility of the product, topsoil enrichment (saturation), rainfall, and the length of the transport flow path (water table depth), among other effects.

Based on this, five observed time series were used as potential explanatory variables in the DFA: a) rainfall (*aR*) (mm $day^{-1}$); b) water table depth (*WTD*) (m NGVD 29); c) soil chemical concentration (*Soil*) (mg $kg^{-1}$); and d) chemical concentrations in the canals bordering the area (*C-111* and *L-31W*) (mg $l^{-1}$).

To approximate the rainfall that can potentially produce leaching of a chemical to the aquifer, the adjusted rainfall (*aR*) was calculated as the ratio between monthly rainfall and number of rainy days in the month. Typically a four-month rainy season occurs in the area from June–September, where over 60% of the total annual precipitation is collected. Since during the wet season it rains almost daily in the area, the adjustment only affects the dry season when sometimes intense and isolated events with a large leaching potential occur. Due to the high cross-correlation between the two field rain gauges (0.95, $p < 0.001$), the average of the two time series from both devices was used (Fig. 1). The *Soil*, *WTD*, *C-111* and *L-31W* explanatory variables were obtained directly from field observations and sample analyses.

### 2.3.3. Analysis procedure

DFA was applied on standardized time series, because this facilitates the interpretation of factor loadings and the comparison of regression parameters. Although normality of data is beneficial for DFA, it is not strictly necessary (Zuur et al., 2003a). The analysis was conducted in three incremental steps. First, an exploratory analysis was conducted by visual inspection of the observed data and calculation of cross-correlation among all variables (response and explanatory) for each chemical, with the aim of identifying relevant explanatory variables for the agricultural and non-agricultural chemicals being studied. Second, different DFMs were compared based on AIC and $C_{\mathrm{eff}}$. These models were derived by incrementally adding the number of common trends and by testing different combinations of explanatory variables. To choose the 'best' model, a compromise was sought between AIC, goodness-of-fit ($C_{\mathrm{eff}}$) and minimum number of common trends and explanatory variables needed. Third, results from the DFA performed for each chemical with the selected models were discussed.

## 3. Results and discussion

### 3.1. Experimental time series

A total of 772 water quality samples (excluding field and instrument blanks) were collected during the experimental period, resulting in a total of 5404 concentration values used in this study. Figs. 2 and 3 depict the standardized values for the chemicals studied and Table 1 summarizes the results for the ground (wells) and surface (canal C-111 and L-31W) samples. These figures allow for a quick visual comparison among the elements and the potential explanatory variables identified in the first step of the analysis.

$P–PO_4^{3-}$ and TP average concentrations and ranges were markedly different in surface and groundwater (Table 1). Mean concentrations of TP in surface waters exceeded the 0.010 mg $l^{-1}$ regulatory level, in 70–74% of the canal samples (40/57 and 42/57 samples for *C-111* and *L-31W*, respectively). Average concentrations and ranges of both P analyses from canal L-31W closely matched those obtained from C-111 canal. These concentrations in water samples from the monitoring wells were the highest during June–September (summer rainy season), although some isolated peaks occurred in both winter crop seasons, typically associated with large rainfall events (Figs. 2 and 3). The June–September high concentrations indicate a rapid mobilization (leaching) from the topsoil enriched by fertilizers after the crop season. On the other hand, the peaks at the beginning of the crop season can be attributed to the fertilizer just applied to the soil (in pre-planting) and leached by the intense rainfall event.

Average $[N–NO_3^-]$ in all surface and groundwater samples were below 10 mg $l^{-1}$ (U.S. drinking water standard) except for one sample collected in well 2 (June 5, 2002) and another in well 3 (June 19, 2002). On a monthly basis, the higher groundwater nitrate concentrations were again observed during the rainy seasons, with a second (smaller) increase at the beginning of the winter crop seasons (Fig. 3). Nitrate concentrations in the canals were lower than in the groundwater by around one order of magnitude.

$[N–NH_4^+]$ in groundwater suggests an inverse pattern to that of nitrate, i.e. the peak ammonia concentrations were generally higher when the nitrate was low (Fig. 3). Average ammonia concentrations in both ground and surface waters were similar (Table 1).

Average concentrations of other natural tracer elements analyzed ($F^-$and $Cl^-$) were low and within natural and regulatory levels (McCutcheon et al., 1992). Surface and groundwater concentrations were similar for both elements. The similar concentration

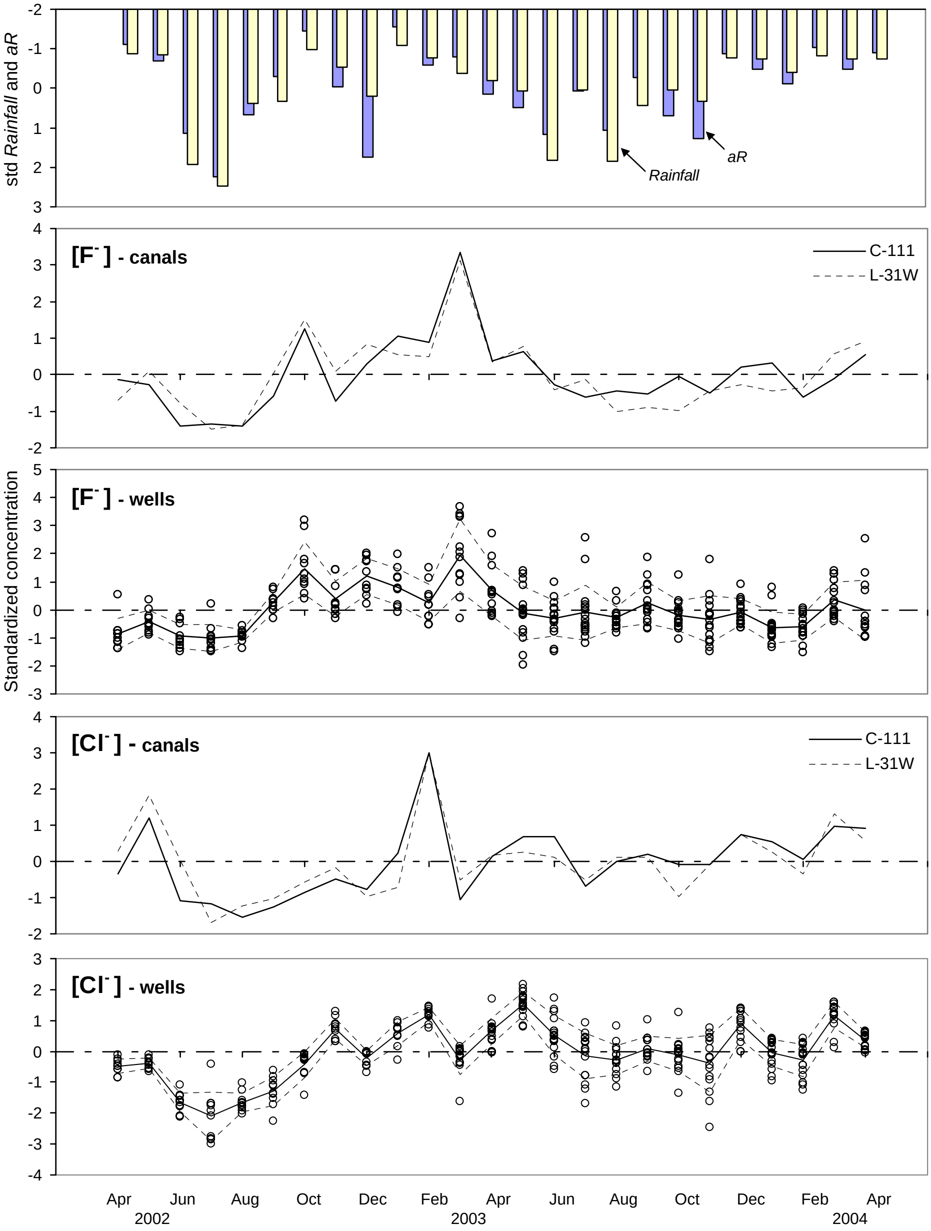

Fig. 2. Standarized time series for the explanatory hydrological variables (rainfall, adjusted rainfall (*aR*)) and chemical concentrations for the $F^-$ and $Cl^-$ obtained in the 2 canals (C-111 and L-31W) and the 16 experimental wells (symbols). Average time series (solid line) and ± standard deviation (dashed lines) are included for each chemical from the wells.

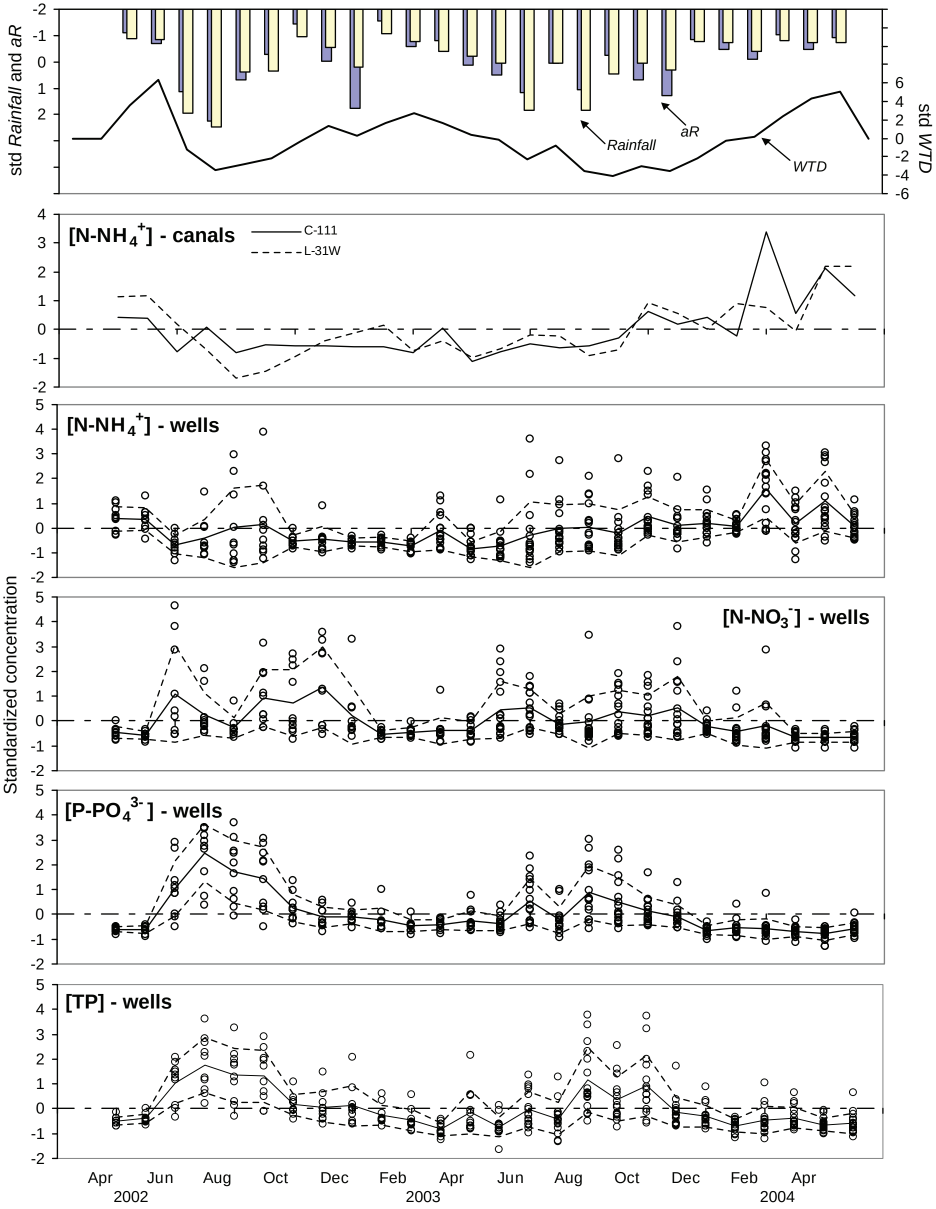

Fig. 3. Standarized time series for the explanatory hydrological variables (rainfall, adjusted rainfall (*aR*), water table depth (*WTD*)) and agrochemical concentrations obtained in the 2 canals (C-111 and L-31W) and 16 experimental wells (symbols). Average time series (solid line) and ± standard deviation (dashed lines) are included for each chemical from the wells.

Table 1
Descriptive statistics for chemicals studied in ground and surface waters[a]

| | Wells | *C-111* | *L-31W* |
|---|---|---|---|
| [$F^-$] | 0.18 ± 0.08 (0.06–0.74) | 0.21 ± 0.08 (0.11–0.44) | 0.19 ± 0.08 (0.09–0.41) |
| [$Cl^-$] | 39.70 ± 8.51 (8.00–60.73) | 46.76 ± 10.84 (30.00–79.38) | 44.25 ± 11.65 (25.00–79.23) |
| [$N–NH_4^+$] | 0.20 ± 0.15 (0.01–1.03) | 0.12 ± 0.08 (0.03 –0.40) | 0.13 ± 0.06 (0.03–0.26) |
| [$N–NO_3^-$] | 0.42 ± 0.89 (0.002–10.46) | 0.05 ± 0.03 (0.02–0.10) | 0.05 ± 0.04 (0.01–0.14) |
| $P–PO_4^{3-}$ | 0.04 ± 0.06 (0.001–0.42) | 0.003 ± 0.002 (0.001–0.01) | 0.003 ± 0.002 (0.001–0.01) |
| [TP] | 0.08 ± 0.09 (0.01–0.60) | 0.02 ± 0.01 (0.003–0.04) | 0.02 ± 0.01 (0.002–0.04) |

[a] Average ± standard deviation; Range in parenthesis; Values expressed in mg $l^{-1}$.

ranges in surface and groundwater for [$N–NH_4^+$], [$Cl^-$], and [$F^-$], suggest a possible interaction between canals and wells. Although $N–NH_4^+$ is considerably less mobile than $Cl^-$and $F^-$, transport of this element can be facilitated by the large hydraulic conductivity and preferential flow paths of the gravelly soil and porous limestone rock (Genereux and Guardiario, 1998, 2001).

### 3.2. Dynamic factor analysis

#### 3.2.1. Analysis of cross-correlation

Cross-correlation results (not shown) were obtained for both [$Cl^-$] and [$F^-$]. In general, concentrations of these anions in all of the wells had a moderate to high correlation among each other. [$F^-$] and [$Cl^-$] in the canals also showed moderate cross-correlations with the concentrations in 70% of the wells. The cross-correlations between the two canals were high for both chemicals (0.88).

As with [$F^-$] and [$Cl^-$], the canals were cross-correlated for ammonia [$N–NH_4^+$] (0.68), and 75% of wells generally presented a moderate cross-correlation coefficient with the concentrations in the canals. In addition, 25% of the wells were cross-correlated with *WTD*. Nitrate concentrations [$N–NO_3^-$] were correlated to *aR*, *Soil*, and *WTD* for 25–44% of the wells, regardless of their location. Both [$P–PO_4^{3-}$] and [TP] were correlated with *aR*, *Soil*, or *WTD* for all the wells but one.

Because concentrations in canals C-111 and L-31W for some chemicals ($N–NH_4^+$, $F^-$and $Cl^-$) were correlated, only the corresponding concentration in canal C-111 was used in the DFAs herein. This explanatory variable will be denoted with the label *Canal*.

#### 3.2.2. DFM selection

Various models can be analyzed according to the number of common trends used and the different combinations of explanatory variables added to the DFM. Table 2 summarizes the models tested for describing the concentrations of each chemical at the observation wells. [$F^-$] were best described with three common trends (minimum AIC = 768), whereas a DFM with a single trend and *aR,* and [$F^-$]$_{Canal}$ as explanatory variables resulted in a similar AIC = 763. Using a model with a single common trend would be satisfactory to predict [$Cl^-$] (minimum AIC = 513, $C_{eff}$=0.81), but the addition of the explanatory variables *aR*, and [$Cl^-$]$_{Canal}$ in the model further decreased the AIC to 485. When no explanatory variables were considered, the AIC values suggested that [$N–NH_4^+$] are best described with two common trends. The model could be improved by

Table 2
Selection of dynamic factor models based on performance coefficients

| Chemical | Trends | $v_k$ | AIC | $C_{eff}$[a] |
|---|---|---|---|---|
| $F^-$ | 1 | | 795 | 0.52 |
| | 2 | | 778 | 0.64 |
| | 3 | | 768 | 0.76 |
| | 4 | | 780 | 0.83 |
| | **1** | ***aR*, *Canal*** | **763** | **0.64** |
| $Cl^-$ | 1 | | 513 | 0.81 |
| | 2 | | 518 | 0.86 |
| | **1** | ***aR*, *Canal*** | **485** | **0.84** |
| $N–NO_3^-$ | 1 | | 924 | 0.34 |
| | 2 | | 886 | 0.53 |
| | 3 | | 856 | 0.70 |
| | 4 | | 837 | 0.74 |
| | 5 | | 828 | 0.81 |
| | 6 | | 824 | 0.83 |
| | 7 | | 847 | 0.83 |
| | 1 | *aR*, *Soil*, *WTD* | 902 | 0.52 |
| | 2 | *aR*, *Soil*, *WTD* | 842 | 0.66 |
| | **3** | ***aR*, *Soil*, *WTD*** | **826** | **0.77** |
| $N–NH_4^+$ | 1 | | 935 | 0.35 |
| | 2 | | 800 | 0.62 |
| | 3 | | 807 | 0.68 |
| | 1 | *aR*, *Soil*, *WTD* | 971 | 0.44 |
| | 1 | *Soil*, *WTD*, *Canal* | 782 | 0.59 |
| | 2 | *aR*, *Soil*, *WTD* | 836 | 0.66 |
| | **2** | ***Soil*, *WTD*, *Canal*** | **779** | **0.69** |
| $P–PO_4^{3-}$ | 1 | | 716 | 0.67 |
| | 2 | | 666 | 0.79 |
| | 3 | | 658 | 0.84 |
| | 4 | | 633 | 0.88 |
| | 5 | | 649 | 0.90 |
| | 1 | *aR*, *Soil*, *WTD* | 689 | 0.75 |
| | **2** | ***aR*, *Soil*, *WTD*** | **632** | **0.83** |
| TP | 1 | | 844 | 0.51 |
| | 2 | | 838 | 0.60 |
| | 3 | | 825 | 0.68 |
| | 4 | | 828 | 0.78 |
| | **1** | ***aR*, *Soil*, *WTD*** | **822** | **0.63** |
| | 2 | *aR*, *Soil*, *WTD* | 794 | 0.72 |

Best model indicated in bold characters.

[a] $C_{eff}$ was calculated with the combined set of predictive vs. observed values for all the wells.

including the following explanatory variables: *Soil*, *WTD* and $[N–NH_4^+]_{Canal}$. This is not the case for $[N–NO_3^-]$ where, if only common trends were considered, six were required to obtain the best model (minimum AIC = 824). However, by adding the best combination of explanatory variables (*aR*, *Soil*, and *WTD*) only three common trends were needed to obtain a similar AIC. For [$P–PO_4^{3-}$], the model containing no explanatory variables that resulted in the lowest AIC used four common trends, while only two trends were necessary when including *aR*, *Soil*, and *WTD* as explanatory variables. The DFM without explanatory variables that best described [TP] in the wells used three common trends. When adding the same explanatory variables as for [$P–PO_4^{3-}$] (i.e. *aR*, *Soil*, and *WTD*) one common trend would be sufficient to reach the same low AIC.

For all chemicals it was shown that by including the explanatory variables in the DFMs, these contributed to explaining the variation in concentration and thereby the number of common trends had been reduced. The introduction of the explanatory variables changes the $C_{\mathrm{eff}}$ by increasing its value for ammonia and chloride and decreasing it for the other chemicals. Note that in all cases, the resulting $C_{\mathrm{eff}}$ were acceptable (0.63–0.84). Based on this, the best DFMs using the corresponding explanatory variables were selected (in bold characters, Table 2). Among all the chemicals studied, results derived from the DFA are only included for three of them: the natural tracer [$F^-$], [$N–NO_3^-$] and [$P–PO_4^{3-}$]. Thereby, Tables 3–5 summarize the results from the best models obtained to predict chemical variations in the sixteen wells.

[$F^-$] was predicted satisfactorily ($0.50 < C_{\mathrm{eff}} < 0.89$) in 81% of the wells (Table 3). The factor loadings ($\gamma_{1,n}$) and the regression parameters ($\beta_{\mathrm{aR},n}$ and $\beta_{\mathrm{Canal},n}$) represent the relative weight of the common trend and each explanatory variable in the model, respectively. For most of these wells, the $\gamma_{1,n}$ and the $\rho_{1,n}$ indicate that the inclusion of the explanatory variables does not reduce the importance of the common trends, so that these contained information that is necessary to determine [$F^-$] variations in the area. However, both *aR* and *Canal* have influence in the concentration changes observed in the wells. The results from DFA of [$Cl^-$] (not shown here) indicate a satisfactory model fit for all the sixteen wells ($0.64 < C_{\mathrm{eff}} < 1.00$). The $\gamma_{1,n}$ and the $\rho_{1,n}$ suggest that the common trend is also important for

Table 3
DFA results for [$F^-$] in the sixteen wells

| $s_n$ | $\gamma_{1,n}$ | $\mu_n$ | $\beta_{\mathrm{aR},n}$ | $\beta_{\mathrm{Canal},n}$ | $\rho_{1,n}$ | $C_{\mathrm{eff},n}$ |
|---|---|---|---|---|---|---|
| T_w1 | 0.54 | 0.00 ± 0.54 | **0.25 ± 0.15** | **0.54 ± 0.16** | 0.80 | 0.81 |
| T_w2 | 0.33 | 0.00 ± 0.34 | 0.00 ± 0.13 | **0.70 ± 0.13** | 0.58 | 0.78 |
| T_w3 | 0.49 | 0.00 ± 0.50 | −0.11 ± 0.18 | 0.30 ± 0.19 | 0.66 | 0.56 |
| T_w4 | 0.55 | 0.00 ± 0.55 | 0.11 ± 0.15 | **0.47 ± 0.17** | 0.79 | 0.77 |
| T_w5 | 0.49 | 0.00 ± 0.50 | −0.20 ± 0.19 | 0.19 ± 0.20 | 0.63 | 0.50 |
| T_w6 | 0.58 | 0.00 ± 0.58 | **−0.35 ± 0.17** | 0.16 ± 0.18 | 0.72 | 0.72 |
| T_w7 | 0.28 | 0.00 ± 0.32 | −0.15 ± 0.21 | 0.18 ± 0.22 | 0.37 | 0.22 |
| T_w8 | 0.31 | 0.00 ± 0.33 | **−0.49 ± 0.16** | 0.23 ± 0.17 | 0.42 | 0.58 |
| T_w9 | 0.41 | 0.00 ± 0.41 | **0.29 ± 0.13** | **0.70 ± 0.14** | 0.69 | 0.78 |
| T_w10 | 0.30 | −0.02 ± 0.31 | 0.08 ± 0.10 | **0.77 ± 0.11** | 0.63 | 0.87 |
| S_w11 | 0.25 | −0.11 ± 0.28 | 0.10 ± 0.18 | **0.81 ± 0.14** | 0.50 | 0.81 |
| S_w12 | 0.41 | 0.04 ± 0.47 | −0.37 ± 0.35 | −0.07 ± 0.27 | 0.35 | 0.19 |
| S_w13 | 0.60 | 0.33 ± 0.63 | **−0.71 ± 0.33** | 0.32 ± 0.55 | 0.22 | 0.38 |
| N_w14 | 1.29 | 0.65 ± 1.28 | 0.24 ± 0.32 | **1.61 ± 0.48** | 0.70 | 0.89 |
| N_w15 | −0.09 | −0.12 ± 0.21 | **−0.99 ± 0.25** | **−0.90 ± 0.44** | −0.06 | 0.64 |
| N_w16 | −0.35 | 0.02 ± 0.38 | **−0.48 ± 0.22** | **0.91 ± 0.38** | −0.67 | 0.70 |

Table 4
DFA results for $[\mathrm{N{-}NO_3^-}]$ in the sixteen wells

| $s_n$ | $\gamma_{1,n}$ | $\gamma_{2,n}$ | $\gamma_{3,n}$ | $\mu_n$ | $\beta_{\mathrm{aR},n}$ | $\beta_{\mathrm{Soil},n}$ | $\beta_{\mathrm{WTD},n}$ | $\rho_{1,n}$ | $\rho_{2,n}$ | $\rho_{3,n}$ | $C_{\mathrm{eff},n}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T_w1 | 0.14 | 0.09 | −0.05 | 0.01 ± 0.22 | 0.14 ± 0.17 | −0.02 ± 0.25 | **−0.64 ± 0.27** | 0.14 | 0.03 | 0.03 | 0.53 |
| T_w2 | 0.52 | 0.52 | −0.15 | 0.04 ± 0.75 | 0.11 ± 0.13 | **0.61 ± 0.28** | **−1.11 ± 0.24** | 0.70 | 0.00 | 0.03 | 0.99 |
| T_w3 | 0.52 | 0.56 | −0.15 | 0.04 ± 0.78 | −0.13 ± 0.18 | **0.83 ± 0.34** | **−1.09 ± 0.31** | 0.65 | 0.04 | −0.08 | 0.76 |
| T_w4 | 0.11 | 0.71 | 0.09 | −0.01 ± 0.71 | −0.08 ± 0.19 | 0.44 ± 0.34 | −0.49 ± 0.32 | 0.19 | 0.67 | −0.42 | 0.70 |
| T_w5 | −0.29 | 0.35 | 0.24 | −0.04 ± 0.54 | −0.07 ± 0.18 | 0.01 ± 0.30 | −0.09 ± 0.30 | −0.39 | 0.74 | −0.38 | 0.59 |
| T_w6 | −0.52 | 0.17 | 0.11 | −0.04 ± 0.62 | **0.45 ± 0.15** | 0.14 ± 0.27 | 0.28 ± 0.26 | −0.48 | 0.76 | −0.56 | 0.80 |
| T_w7 | 0.01 | 0.59 | 0.41 | −0.04 ± 0.75 | −0.19 ± 0.19 | 0.08 ± 0.35 | −0.28 ± 0.33 | −0.10 | 0.75 | −0.21 | 0.66 |
| T_w8 | 0.26 | −0.08 | 0.94 | −0.06 ± 1.24 | 0.16 ± 0.19 | 0.26 ± 0.39 | −0.37 ± 0.34 | −0.18 | 0.05 | 0.75 | 0.96 |
| T_w9 | 0.01 | 0.25 | 0.61 | −0.05 ± 0.78 | 0.14 ± 0.20 | 0.21 ± 0.34 | −0.36 ± 0.33 | −0.23 | 0.50 | 0.14 | 0.60 |
| T_w10 | −0.36 | 0.38 | 0.06 | −0.04 ± 0.58 | **−0.35 ± 0.15** | −0.25 ± 0.27 | 0.00 ± 0.26 | −0.44 | 0.79 | −0.57 | 0.81 |
| S_w11 | 0.65 | −0.22 | −0.12 | −0.40 ± 0.77 | **0.51 ± 0.18** | 0.09 ± 0.33 | **−0.81 ± 0.28** | −0.14 | −0.11 | 0.60 | 0.96 |
| S_w12 | 0.82 | 0.00 | 0.09 | −0.35 ± 0.96 | −0.03 ± 0.20 | 0.47 ± 0.38 | **−1.34 ± 0.33** | −0.19 | −0.01 | 0.72 | 0.94 |
| S_w13 | 0.84 | 0.30 | −0.29 | −0.46 ± 0.98 | −0.10 ± 0.18 | −0.21 ± 0.38 | **−1.26 ± 0.32** | −0.08 | −0.07 | 0.32 | 1.00 |
| N_w14 | −0.29 | 0.38 | −0.34 | 0.72 ± 0.78 | 0.13 ± 0.40 | 0.96 ± 0.70 | −0.86 ± 0.53 | −0.20 | 0.11 | −0.32 | 0.26 |
| N_w15 | −0.53 | 0.05 | 0.99 | 0.16 ± 1.24 | **0.48 ± 0.23** | 0.38 ± 0.48 | 0.44 ± 0.40 | −0.43 | 0.68 | 0.73 | 1.00 |
| N_w16 | 0.02 | 1.54 | 0.01 | 1.37 ± 1.52 | 0.42 ± 0.32 | **1.82 ± 0.65** | **−1.44 ± 0.55** | 0.15 | 0.83 | 0.02 | 1.00 |

Table 5
DFA results for [P–$PO_4^{3-}$] in the sixteen wells

| $s_n$ | $\gamma_{1,n}$ | $\gamma_{2,n}$ | $\mu_n$ | $\beta_{\mathrm{aR},n}$ | $\beta_{\mathrm{Soil},n}$ | $\beta_{\mathrm{WTD},n}$ | $\rho_{1,n}$ | $\rho_{2,n}$ | $C_{\mathrm{eff},n}$ |
|---|---|---|---|---|---|---|---|---|---|
| T_w1 | 0.31 | 0.05 | −0.12 ± 0.36 | −0.02 ± 0.12 | **1.17 ± 0.20** | −0.18 ± 0.15 | −0.27 | 0.04 | 0.83 |
| T_w2 | −0.15 | 0.35 | −0.04 ± 0.37 | 0.07 ± 0.18 | 0.03 ± 0.24 | **−0.67 ± 0.20** | −0.22 | 0.15 | 0.52 |
| T_w3 | 0.03 | 0.30 | −0.09 ± 0.32 | **0.18 ± 0.09** | **0.71 ± 0.17** | **−0.33 ± 0.12** | −0.44 | 0.29 | 0.93 |
| T_w4 | 0.16 | 0.21 | −0.11 ± 0.32 | 0.03 ± 0.11 | **0.94 ± 0.18** | **−0.25 ± 0.14** | −0.35 | 0.24 | 0.84 |
| T_w5 | −0.09 | 0.33 | −0.05 ± 0.33 | **0.30 ± 0.12** | **0.32 ± 0.19** | **−0.48 ± 0.14** | −0.39 | 0.20 | 0.85 |
| T_w6 | −0.47 | 0.51 | 0.03 ± 0.63 | **0.31 ± 0.13** | −0.33 ± 0.34 | **−0.65 ± 0.21** | −0.60 | 0.33 | 0.99 |
| T_w7 | 0.46 | −0.11 | −0.13 ± 0.49 | −0.06 ± 0.16 | **1.20 ± 0.27** | −0.09 ± 0.20 | −0.01 | −0.09 | 0.70 |
| T_w8 | 0.13 | −0.01 | −0.04 ± 0.16 | **0.26 ± 0.12** | **0.74 ± 0.12** | **−0.28 ± 0.12** | −0.31 | −0.11 | 0.79 |
| T_w9 | −0.35 | 0.43 | 0.01 ± 0.50 | 0.08 ± 0.16 | −0.05 ± 0.29 | **−0.63 ± 0.21** | −0.54 | 0.29 | 0.75 |
| T_w10 | −0.19 | 0.43 | −0.05 ± 0.44 | −0.04 ± 0.14 | 0.15 ± 0.25 | **−0.79 ± 0.19** | −0.37 | 0.26 | 0.80 |
| S_w11 | 0.07 | 1.11 | −0.68 ± 1.15 | 0.13 ± 0.25 | −0.44 ± 0.94 | **−2.09 ± 0.40** | 0.31 | −0.18 | 0.99 |
| S_w12 | 0.60 | 0.80 | −0.81 ± 1.16 | 0.03 ± 0.26 | 0.52 ± 1.05 | **−1.66 ± 0.40** | 0.49 | −0.20 | 0.97 |
| S_w13 | 1.40 | −0.18 | −0.62 ± 1.44 | −0.49 ± 0.30 | **2.62 ± 1.11** | −0.50 ± 0.49 | 0.47 | −0.45 | 1.00 |
| N_w14 | 0.87 | −0.21 | −2.38 ± 1.13 | −0.55 ± 0.37 | −1.54 ± 1.37 | **−0.88 ± 0.43** | 0.15 | −0.38 | 0.48 |
| N_w15 | 0.00 | 0.49 | 2.00 ± 0.52 | **0.40 ± 0.12** | **3.50 ± 0.47** | **−0.37 ± 0.18** | 0.15 | −0.26 | 0.99 |
| N_w16 | −0.65 | 1.22 | 0.52 ± 1.27 | **0.55 ± 0.28** | −0.29 ± 0.94 | **−1.55 ± 0.44** | 0.12 | 0.07 | 0.96 |

this model, especially for the transect wells. Regression parameters for the concentration in the canals ($\beta_{\mathrm{Canal},n}$) presented the largest values, but [$Cl^-$] are also affected by the *aR*.

The model fit for [N–$NH_4^+$] was successful in 81% of the wells with $C_{\mathrm{eff}}$ values ranging from 0.55 to 1.00 (results not shown). Canonical correlation coefficients indicate that the two common trends are only important for describing [N–$NH_4^+$] in the wells south and north of the transect ($0.36<|\rho_{1,n}|<0.83$). While these wells have a positive correlation with the first trend ($0.36<\rho_{1,n}<0.68$), they are negatively correlated to the second trend ($-0.49>\rho_{1,n}>-0.83$). For the transect wells the explanatory variables generally determine the observed variation in [N–$NH_4^+$], while the influence of the common trends is minor. Regression parameters show that *Canal* is the most important variable in the transect wells, while *Soil* has more impact in the southern wells and *WTD* in the northern wells. Results obtained from the DFA on [N–$NO_3^-$] (Table 4) show a satisfactory model fit ($C_{\mathrm{eff}}>0.50$) in all the wells but N_w14. The low $C_{\mathrm{eff}}$ value obtained for this well reduces the global $C_{\mathrm{eff}}$ compared to the one for the DFM without explanatory variables (0.83 vs 0.77, Table 2). Although the inclusion of the explanatory variables partially describes [N–$NO_3^-$] changes, each of the three common trends remained important for explaining the [N–$NO_3^-$] variability. No relationship between these common trends and the spatial layout of the different wells is observed (cf. $\rho_{m,n}$ in Table 4). In general, *WTD* and *aR* are the explanatory variables with the largest and smallest influence, respectively.

The [P–$PO_4^{3-}$] model performance (Table 5) was satisfactory, with $C_{\mathrm{eff}}$ above 0.70 in all but two wells (T_w2 and N_w14). The explanatory variables included in the model contain information necessary to describe [P–$PO_4^{3-}$] changes in the well, so that the influence of the common trends is reduced to 25% of the wells. *WTD* and *Soil* are the variables that most affect [P–$PO_4^{3-}$] in groundwater, while *aR* has less influence. This is also the case for TP, but the trend still affects 44% of the wells (results not shown). Model performance for [TP] was acceptable for 81% of the wells with $C_{\mathrm{eff}}$ ranging from 0.53 to 0.99.

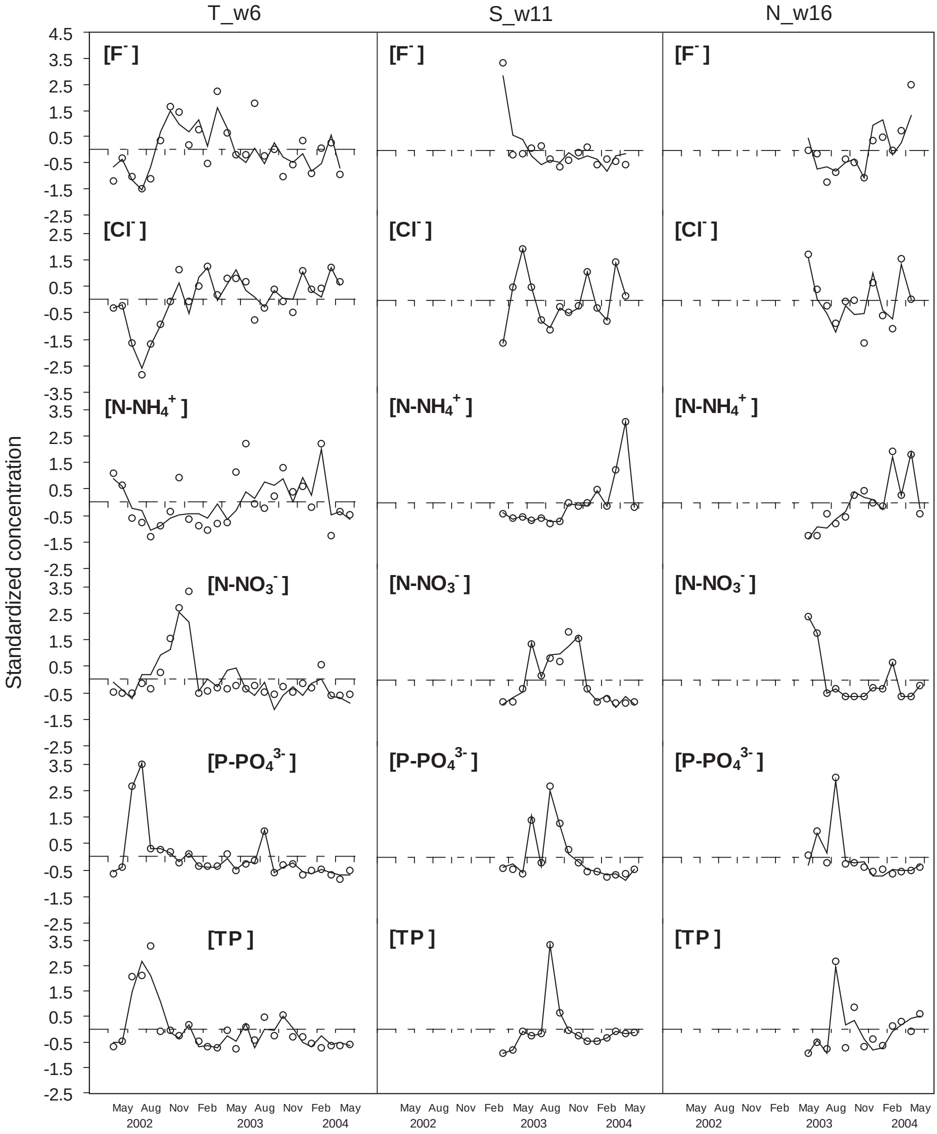

Fig. 4. Model fit for the chemicals studied in three representative wells, located in the transect (T_w6), south of the transect (S_w11) and north of it (N_w16). North and south wells installed in second experimental phase (February 2003).

Fig. 4 illustrates the satisfactory model fit derived from the DFA with explanatory variables for each chemical. The three representative wells presented correspond to a location in the transect itself (T_w6), a location south of the transect (S_w11) and one north (N_w16) of the experimental area.

Table 6
Summary of relative effect of explanatory variables and trends on groundwater chemical variation

| Chemical | *aR*[a] | *Canal*[a] | *WTD*[a] | *Soil*[a] | Trends[b] | $C_{\mathrm{eff}}$ |
|---|---|---|---|---|---|---|
| $F^-$ | 2 | 1 | – | – | (1)** | 0.64 |
| $Cl^-$ | 2 | 1 | – | – | (1)** | 0.84 |
| $N–NH_4^+$ | – | 1 [transect] | 2 | 1 [south] | (2)** [north/south] | 0.69 |
| $N–NO_3^-$ | 3 | – | 1 | 2 | (3)** | 0.77 |
| $P–PO_4^{3-}$ | 3 | – | 1 | 2 | (2)** | 0.83 |
| TP | 3 | – | 1 | 2 | (1)* | 0.63 |

[a] The relative importance for each explanatory variable is quantified in increasing order from 1 to 3.

[b] The number of common trends is given in parenthesis. Asterisks denote the relative importance of the common trends based on $\rho_n$, where *, **, *** correspond to an average $\rho_n$=0.3–0.5, 0.5–0.75, >0.75, respectively.

### 3.2.3. *Influence of the explanatory variables on groundwater quality*

Table 6 presents a matrix summarizing the interactions of the model components for each chemical studied. Three of the agrochemical species analyzed ($N–NO_3^-$, $P–PO_4^{3-}$ and TP) were affected by the same explanatory variables. In order of decreasing relative importance a small water table depth, followed by enriched topsoil and occurrence of a leaching rainfall event affected the increase in groundwater concentration for these chemicals. These variables indicate that leaching by rainfall is the main mechanism explaining concentration peaks. Topsoil and canal concentrations, followed by water table depth resulted in variation of the ammonia groundwater concentration. This suggests that the dominant processes affecting variation in this case are different than for the other agrochemicals. In addition to leaching, variation is induced by lateral exchange between canals and groundwater for this element. Ammonia transport is facilitated in this environment by the large hydraulic conductivity and preferential flow paths of the gravelly soil and porous limestone rock (Genereux and Guardiario, 1998, 2001). Water table depth (related to leaching opportunity when couple with rainfall) affected wells north and south of the transect differently (Table 6). Previous work (Ritter and Muñoz-Carpena, 2005) has shown that the water table in the northern wells is consistently higher (smaller WTD) than in the southern wells. [$F^-$] and [$Cl^-$] are mainly affected by periods of dilution by rainfall and by exchange with the canals. Since model fit ($C_{\mathrm{eff}}$ in Table 6) and the correlation of the common trends was generally only moderate (weak for TP), it can be concluded that much of the groundwater chemical variation observed is successfully accounted for by the explanatory variables included in each model. The remaining effect of the common trends did not show spatial structure across the area (except for ammonia). Since the combination of different land parcels, crops and farmers results in unstructured agrochemical use patterns (crop, irrigation and fertilization scheduling and rates, etc.) across space and time, this suggests that the land use effect is encompassed in the unexplained variability represented by the common trends.

## 4. Summary and conclusions

Multivariate time series of hydrological and water quality variables were obtained from a small agricultural watershed located at the boundary of Everglades National Park

(ENP). Two drainage canals from the regional water management network delimit this area and are operated under environmental restrictions. Several projects seek the restoration of the wetland ecosystems of the ENP by enhancing water deliveries into the Park while maintaining flood protection in the adjacent agricultural fields. In addition, surface water entering the ENP must satisfy the current regulatory standards ($<0.010$ mg $l^{-1}$ of Total P). In this context, monitoring and analysis of the variation of chemical concentrations in surface and groundwaters can provide a better understanding of land use and natural variables affecting the water quality in the area. Dynamic Factor Analysis was performed on monthly averaged time series of soil and ground and surface water concentrations of different chemicals (fluoride, chloride, ammonia–nitrogen, nitrate–nitrogen, orthophosphate, and total phosphorus) from 18 locations in the area. A technique for the analysis of multivariate non-stationary time-series, DFA was conducted as follows. Firstly, cross-correlations among all time series for each chemical were determined. This preliminary procedure allowed identifying relevant explanatory variables for each chemical. However, although cross-correlation coefficients serve as an exploratory tool and provide a measure of the relationship between paired data sets, it does not properly capture simultaneous interactions of multivariate time series. This can be achieved with Dynamic Factor Analysis (DFA). Secondly, a series of DFAs was performed for each chemical to identify the combination of common trends that best describes changes in concentration over time and in sixteen wells across the field. Both, orthophosphate and nitrate concentrations required the largest numbers of trends (four and six, respectively). This suggests that various latent effects influence in a different way the groundwater concentration of these agrochemicals across the area. The number of common trends required to determine the concentration of each chemical in the wells was greatly reduced by including time series of explanatory variables in the DFA. These were rainfall (*aR*), water table depth (*WTD*), agrochemicals concentration in the soil (*Soil*) and concentrations in the canal bordering the watershed (*Canal*). Fluoride and chloride concentrations were influenced by rainfall and especially by the canal concentration, and by one trend representing an unidentified (latent) effect. Ammonia–nitrogen concentrations were affected by *WTD*, *Soil* and *Canal*, but *Canal* was especially important in describing concentrations in the transect wells, while *Soil* had more effect in the southern wells and *WTD* in the northern wells. This corresponds with the more frequent flooding conditions present in the northern part of the area, promoting direct transport from the topsoil to the groundwater. Two trends were needed, which represent unidentified effects governing ammonia concentrations in the wells south and north of the transect. For the rest of the agrochemicals (nitrate, orthophosphate and total phosphorus), concentrations were affected by the same explanatory variables that were, in order of decreasing importance, *WTD*, *Soil* and *aR*. However, these variables only partially described concentration changes for some of the wells, so that common trends were also required. The common trends affected the groundwater concentrations of these agrochemicals without any spatial structure. This is likely a consequence of varying land use patterns in this watershed. The combination of different land parcels, crops and farmers results in different management patterns (crop, irrigation and fertilizer schedules and rates, etc.) across space and time. This land use effect is encompassed in the unexplained variability represented by the common trends.

Since *WTD* and *Soil* are conditioned by land management practices, peak concentrations of agrochemicals in groundwater could be reduced by improving fertilization practices (reduction and splitting) and by maintaining a low water table, especially during the rainy periods. In this context it is worth mentioning that observed concentrations of total phosphorus in canal surface waters exceeded the regulatory standards in 70–74% of the samples.

## Acknowledgments

The authors wish to recognize the rest of the research team that collaborated in the field and laboratory efforts: Tina T. Dispenza, Martin Morawietz, Harry Trafford, Michael Gutiérrez, Guiqing Yu and Li Ma. This project was partially funded by the South Dade Soil and Water Conservation District (SDSWCD) and a University of Florida's Center for Natural Resources 2003 Mini-Grant. Dr. A. Ritter wants to thank the DGUI de la Consejería de Educación Cultura y Deportes del Gobierno de Canarias for the funds provided. The team also wishes to acknowledge the collaboration of SDSWCD in setting up the experimental canal platforms constructed for this study. Bruce Schaffer (UF TREC) generously shared his staff to help in field sampling tasks. Karen Minkowski provided GIS and mapping support to this project, and Mr. James Beadman, Registered Surveyor with the State of Florida, donated his time to survey the hydrological instruments. Special thanks go to Julia Lacy, Senior Engineer with the South Florida Water Management District, for her continuous support and for acting as an effective link with the agency. This research was supported by the Florida Agricultural Experiment Station, and approved for publication as Journal Series No. R-10388.

## Appendix A. Coefficient of efficiency

The coefficient of efficiency, $C_{\mathrm{eff}}$ (Nash and Sutcliffe, 1970), also known as the *Nash–Sutcliffe coefficient*, is defined as the ratio of the mean square error to the variance of the observed data, subtracted from unity (Legates and McCabe, 1999). Here, it was expressed as follows:

$$C_{\mathrm{eff}} = 1 - \frac{\mathrm{MSE}}{(\sigma^*)^2} = 1 - \frac{\sum_{i=1}^{l_s} [s(t_i)^* - s(t_i)]^2}{\sum_{i=1}^{l_s} [s(t_i)^* - \bar{s}^*]^2} \qquad (A.1)$$

where $s(t_i)^*$ and $s(t_i)$ are the observed and the predicted chemical concentrations in water samples obtained from the monitoring wells at time $t_i$, respectively; $l_s$ is the length of the observed data set; and $(\sigma^*)^2$ is the variance of the observed chemical concentration. The coefficient of efficiency compares the variance about the 1 : 1 line (perfect agreement) to the variance of the observed data. Notice that for non-regression models the $C_{\mathrm{eff}}$ does not represent the proportion of sum squares (i.e., deviation of the observed values to their mean) explained by the model, and it ranges from $-\infty$ to 1. Thereby $C_{\mathrm{eff}}=1$ implies that the plot of predicted vs. observed values matches the 1 : 1 line. $C_{\mathrm{eff}}$ values $<0.5$ indicate that the model is not adequate.

## References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723.

Erzini, K., 2005. Trends in NE atlantic landings environmental landings. Fisheries Oceanogr. 14 (3), 195–209.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B 39, 1–38.

Fish, J.E., Stewart, M., 1991. Hydrogeology of the Surficial Aquifer System, Dade County, Florida. WRI 90-4108. U.S. Geological Survey, Tallahassee, FL.

FL-DEP, 2002. DEP Standard Operating Procedures for Field Activities. Florida Department of Environmental Protection, Tallahassee, FL. DEP-SOP-001/01, January 2002.

Florida Senate Bill 0626ER, 2003. Amended Everglades Forever Act. May 20, 2003.

Genereux, D., Guardiario, J., 1998. A canal drawdown experiment for determination of aquifer parameters. J. Hydrol. Eng. 3 (4), 294–302.

Genereux, D., Guardiario, J., 2001. A borehole flowmeter investigation of small-scale hydraulic conductivity variation in the Biscayne Aquifer, Florida. Water Resour. Res. 37 (5), 1511–1517.

Genereux, D., Slater, E., 1999. Water exchange between canals and surrounding aquifer and wetlands in the Southern Everglades, USA. J. Hydrol. 219 (3–4), 153–168.

Geweke, J.F., 1977. The dynamic factor analysis of economic time series models. In: Aigner, D.J., Goldberger, A.S. (Eds.), Latent Variables in Socio-economic Models. Amsterdam, North-Holland, pp. 365–382.

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

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.

Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin.

Márkus, L., Berke, O., Kovács, J., Urfer, W., 1999. Spatial prediction of the intensity of latent effects governing hydrogeological phenomena. Environmetrics 10, 633–654.

McCutcheon, S.C., Martin, J.L., Barnwell, T.O., 1992. Water quality. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 11.1–11.73.

Mendelssohn, R., Schwing, F.B., 2002. Common and uncommon trends in SST and wind stress in the California and Peru–Chile current systems. Prog. Oceanogr. 53 (2–4), 141–162.

Molenaar, P.C.M., 1985. A dynamic factor model for the analysis of multivariate time series. Psychometrika 50, 181–202.

Molenaar, P.C.M., 1989. Aspects of dynamic factor analysis. Analysis of Statistical Information. The Institute of Statistical Mathematics, Tokyo, pp. 183–199.

Molenaar, P.C.M., de Gooijer, J.G., Schmitz, B., 1992. Dynamic factor analysis of non-stationary multivariate time series. Psychometrika 57, 333–349.

Muñoz-Carpena, R., Li, Y.C., Dispensa, T.T., Morawietz, M., 2003. Hydrological and water quality monitoring network in a watershed adjacent to Everglades National Park. ASAE Paper No. 032048ASAE, St. Joseph, Mich.

Nash, J.E., Sutcliffe, J.V., 1996. River flow forecasting through conceptual models. Part 1-A discussion of Principles. J. Hydrol. 10, 282–290.

Ritter, A., Muñoz-Carpena, R., 2005. Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park. J. Hydrol. 80, 49–70. doi:10.1016/j.jhydrol.2005.05.025.

SFWMD, 2004. Everglades Consolidated Report, 2004 Executive Summary. South Florida Water Management District, West Palm Beach, FL.

Schumann, A., Muñoz-Carpena, R., 2003. A simple, self-contained canal stage recorder. Appl. Eng. Agric. 18 (6), 20–25.

Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3, 253–264.

Wu, L.S.-Y., Pai, J.S., Hosking, J.R.M., 1996. An algorithm for estimating parameters of state–space models. Stat. Probab. Lett. 28, 99–106.

Zuur, A.F., Pierce, G.J., 2004. Common trends in northeast Atlantic squid time series. J. Sea Res. 52, 57–72.

Zuur, A.F., Tuck, I.D., Bailey, N., 2003a. Dynamic factor analysis to estimate common trends in fisheries time series. Can. J. Fish. Aquat. Sci. 60, 542–552.

Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R., Beukema, J.J., 2003b. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics 14 (7), 665–685.