HYDROLOGICAL PROCESSES
Hydrol. Process. **21**, 2464–2473 (2007)
Published online 30 December 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.6483

Agricultural land use and hydrology affect variability of shallow groundwater nitrate concentration in South Florida

A. Ritter,[1,2] R. Muñoz-Carpena,[1]* D. D. Bosch,[3] B. Schaffer[4] and T. L. Potter[3]

[1] *Agricultural and Biological Engineering Dept. University of Florida, 101 Frazier Rogers Hall, Gainesville, FL 32611-0570, USA*
[2] *Instituto Canario de Investigaciones Agrarias (ICIA), Apdo. 60 La Laguna, 38200 Tenerife, Spain*
[3] *USDA-ARS, SEWRL, 2375 Rainwater Road, Tifton, GA 31794, USA*
[4] *TREC, University of Florida, 18905 SW 280 St, Homestead, FL 33031, USA*

Abstract:

South Florida's Miami-Dade agricultural area is located between two protected natural areas, the Biscayne and Everglades National Parks, subject to the costliest environmental restoration project in history. Agriculture, an important economic activity in the region, competes for land and water resources with the restoration efforts and Miami's urban sprawl. The objective of this study, understanding water quality interactions between agricultural land use and the shallow regional aquifer, is critical to the reduction of agriculture's potentially negative impacts. A study was conducted in a 4-ha square field containing 0·9 ha of corn surrounded by fallow land. The crop rows were oriented NW–SE along the dominant groundwater flow in the area. A network of 18 monitoring wells was distributed across the field. Shallow groundwater nitrate–nitrogen concentration [$N\text{-}NO_3{}^-$] was analyzed on samples collected from the wells biweekly for 3 years. Detailed hydrological (water table elevation [*WTE*] at each well, groundwater flow direction [*GwFD*], rainfall) and crop (irrigation, fertilization, calendar) data were also recorded *in situ*. Flow direction is locally affected by seasonal regional drainage through canal management exercised by the local water authority. The data set was analyzed by dynamic factor analysis (DFA), a specialized time series statistical technique only recently applied in hydrology. In a first step, the observed nitrate variation was successfully described by five common trends representing the unexplained variability. By including the measured hydrological series as explanatory variables the trends were reduced to only three. The analysis yields a quantification of the effects of hydrological factors over local groundwater nitrate concentration. Furthermore, a spatial structure across the field, matching land use, was found in the five remaining common trends whereby the groundwater [$N\text{-}NO_3{}^-$] in wells within the corn rows could be generally separated from those in fallow land NW and SE of the crop strip. Fertilization, masked by soil/water/plant-delayed processes, had no discernible effect on groundwater nitrate levels. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS groundwater; hydrology; land use; nitrate; nonpoint source pollution; dynamic factor analysis; multivariate time series; water quality

Received 20 December 2005; Accepted 16 June 2006

INTRODUCTION

The South Florida (USA) agricultural region is located in an environmentally sensitive area, adjacent to the Everglades and Biscayne National Parks. This region is subject to the Comprehensive Everglades Restoration Project (CERP), the most costly natural ecosystem restoration project in history with an estimated budget of US $10 000 million from federal and state funds. While some perceive agriculture in the region as a threat to the restoration process, many prefer it as an alternative to establishment of exotic invasive plants on land left fallow, or to an extension of Miami's urban sprawl. Within Florida's South Miami-Dade Basin, intensive vegetable, tropical fruit, and nursery crops are produced on an area of approximately 32·4 km^2. These crops are grown on extremely permeable and thin soils on top of the shallow unconfined Biscayne aquifer that underlies the entire region providing potable water. High rates of fertilizer and pesticide applications make the aquifer vulnerable to nonpoint source agrochemical contamination (Harman-Fetcho *et al.*, 2005). It has also been hypothesized that leaching to groundwater may be a main pathway for contaminant transport to surface water, since this process may be coupled with subsurface transport, and seepage into the large network of drainage canals (Genereux and Slater, 1999). This has implications for on-going Everglades restoration efforts, which focus on providing increased freshwater deliveries to the National Park. To better understand the environmental costs and benefits of agriculture in Florida's South Miami-Dade Basin, the contribution of agrochemicals to nonpoint source pollution of ground and surface water must be clearly defined (Potter *et al.*, 2004).

Nitrate fluctuations in shallow groundwater typically result from the cumulative effects of different factors, such as land use and associated nitrate concentration in the topsoil, net vertical recharge (affected by leaching rainfall and excess irrigation), local depth to groundwater, lateral recharge from ground or surface water sources, etc. (Muñoz-Carpena *et al.*, 2005). Although

*Correspondence to: R. Muñoz-Carpena, Agricultural and Biological Engineering Dept. University of Florida, 101 Frazier Rogers Hall, Gainesville, FL 32611-0570, USA. E-mail: carpena@ufl.edu

some of these effects can be measured accurately, it is impractical to measure others, i.e. those with unstructured spatial and temporal distribution. In this context, the application of indirect methods to observed water quality data at fixed observation sites becomes an interesting alternative (Márkus *et al.*, 1999). Detailed data sets containing temporal variation of hydrological and water quality variables have the potential to facilitate the understanding of surface–groundwater–land use interactions in the area. However, interpretation of results from data analysis based on visual inspection and comparative statistics is difficult and may not be sufficient, especially when dealing with multivariate time series. An innovative technique for studying hydrological and water quality multivariate time series responses is dynamic factor analysis (DFA). DFA is a specialized time series technique originally developed for the study of economic models (Geweke, 1977) and recently applied to identify common patterns in groundwater levels (Márkus *et al.*, 1999; Ritter and Muñoz-Carpena, 2006), and interactions between hydrological variables and groundwater quality trends (Muñoz-Carpena *et al.*, 2005). It is a dimension reduction technique that takes into account the time component. In DFA, multivariate time series responses may be analyzed as response variables assuming that there are common driving forces behind them, i.e. underlying latent effects that determine the variation of the individual observations with time. These latent effects can be described by common patterns or trends (representing unexplained variability) and/or explanatory variables consisting of other observed time series. The analysis provides information about whether (i) there are any underlying common patterns in the response time series, (ii) there are interactions between the response time series, and (iii) these are affected by the explanatory variables considered. Since the analysis of large water quality data sets is complex because of the many characteristics affecting the variation of chemical concentration in the system, DFA can be an effective tool to handle such data sets and to identify the dominant effects controlling the observed variation.

The purpose of this study was to assess the influence of agricultural land use and hydrological variables on groundwater nitrate concentrations monitored in 18 wells distributed across an agricultural field within Florida's South Miami-Dade Basin. Particular focus was on determining to what extend the groundwater nitrate fluctuations could be explained by hydrological variables (including water table elevation [*WTE*], groundwater flow direction [*GwFD*], rainfall and irrigation), and location related to land use (agricultural or fallow). A three-step procedure was used to analyze the data: (i) exploratory analysis by visual inspection and cross-correlation analysis of the time series; (ii) identification of common trends of groundwater quality with DFA; and (iii) inclusion of explanatory variables in a dynamic factor model (DFM).

MATERIALS AND METHODS

Area of study and experimental setup

The study was conducted on a 4-ha field at the University of Florida's Tropical Research and Education Center (UF-TREC) located on approximately 32·4 km^2 of prime farmland in Homestead (Florida) (80·50 °W, 25·51 °N). This area is adjacent to the South Florida Water Management District (SFWMD) C-103 drainage canal, which is located N of the field, curving E about 400 m from the field (Figure 1). The field is essentially flat with Krome soil type (loamy-skeletal, carbonatic, hyperthermic Lithic Udorthents) of 30-cm depth overlaying porous limestone bedrock (Noble *et al.*, 1996). This is an artificial gravelly loam soil created by 'rock-plowing' the underlying porous limestone bedrock (Genereux and Guardiario, 1998, 2001). Krome soils are used for fruit and vegetable crops and urban and residential development (USDA-NRCS, 2004). The soil characteristics in this study were very low organic carbon content (<1%), high fraction of particles >2 mm (>50%), low water holding capacity, and high permeability (Al-Yahyai *et al.*, 2006). The bedrock was characterized by a very high transmissivity with a high potential for leaching both nutrients and pesticides (Harman-Fetcho *et al.*, 2005).

A rectangular strip (192 by 47 m) of sweet corn (*Zea mays* L.) diagonally oriented with rows running NW–SE (Figure 1) was planted within the 4-ha field so that the widthwise dimension (47 m) paralleled the predominant direction of the natural groundwater flow (NW–SE) identified during a previous hydrogeological investigation conducted at UF-TREC (ES&E, 1996). The cornfield was divided into six subplots of 27 by 47 m each oriented along the diagonal. Corn was planted on all subplots

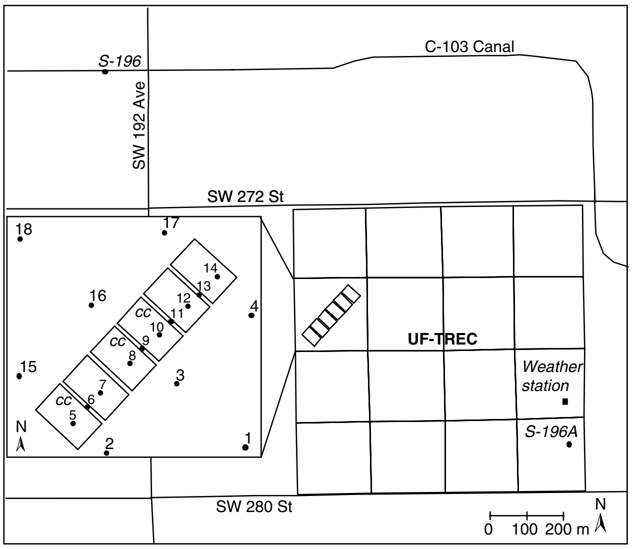

Figure 1. Location of experimental site at UF-TREC showing canal C-103 with the SFWMD's S-196 canal recording gated culvert (upper left corner), and the USGS' S-196A recording well next to the automatic weather station (lower right corner). Each square of the grid represents a 4-ha research field (200 by 200 m). The inset depicts the location of the 18 sampling wells and the corn crop strip along the diagonal (six rectangles). Plots marked with '*CC*' used Sunn Hemp as summer cover crop

DOI: 10.1002/hyp

in October–November of each year and harvested the following January–March at maturity (about 100 days). After harvest, all plots were mowed and cultivated with a disk harrow. In half of the plots, Sunn Hemp (*Crotalaria juncea* L.) was then planted as a summer cover crop. In October of each year, the cover crop plots were mowed and the residue was then worked into the soil by repeated disking. Potter *et al*. (2004) found no significant differences between nitrate concentrations in shallow wells placed on plots with cover and on plots without cover, so they concluded that the cover crop had no measurable effect on shallow ground water nitrate leaching. Crop irrigation and agrochemical applications records were obtained for three crop seasons of the study (1999–2000, 2000–2001, 2001–2002). Types and amounts of pesticides and fertilizers used reflected commercial grower recommendations (Li *et al*., 2002).

A network of 18 wells (4·6–6·1-m depth) was installed for water quality monitoring (Figure 1). These were distributed across the field so that water quality as a function of farming practices and groundwater dynamics could be systematically characterized. During February 2000–2003, shallow groundwater quality grab samples at each monitoring well were collected in acid cleaned and prelabeled 500-ml bottles approximately every 2 weeks and after major rainfall events. Field and laboratory QA/QC standards (FL-DEP, 2002) were followed at all times. The sample bottles (including field and instrument blanks) were placed in a cooler with ice immediately after collection in the field, and transported to the laboratory within 2 hours. The water samples were prepared immediately upon receipt and transferred on ice for refrigeration before analysis. The samples were analyzed for concentrations of nitrate–nitrogen ($N\text{-}NO_3^-$) using an Autoanalyzer (AA3, Bran + Luebbe, Buffalo Grove, IL). Analytical precision for these elements was better than 3% Relative Standard Deviation (RSD).

Figure 1 depicts the location of the instrumentation used to collect the hydrological variables. Detailed (15 min) meteorological rainfall data were obtained at the University of Florida's FAWN (Florida Automated Weather Network, http://fawn.ifas.ufl.edu) station located in UF-TREC's property less than 700 m away from the plots. Continuous records of surface levels on the distal NW corner of the area were obtained from the SFWMD's culvert structure S-196. Groundwater table elevation in the distal SE corner of the property was obtained from well S-196A maintained by the US Geological Service. Water table elevation inside the experimental field was continuously recorded at three of the wells (*W2*, *W4*, and *W18*) equipped with auto-logging pressure transducers compensating for temperature effects and atmospheric pressure (Solinst Inc., Canada). In addition, water table depth was recorded manually at all field wells before each sampling event.

Dynamic factor analysis

Time series consist of time dependent data containing deterministic and random variation. The analysis of this type of data is generally based on decomposing the information to characterize the deterministic and random variation. The analysis of multivariate time series with classical time series methods is difficult because they are techniques based on stationarity of data, cannot handle missing values and need relatively long time series. To overcome the nonstationarity problem, the time series can be detrended or analyzed as integrated series; however, trends may contain important information about the underlying effects in the time series. On the other hand, to address the multivariate nature of the response variables, conventional multivariate techniques (such as principal component analysis, factor analysis or redundancy analysis) are sometimes used. However, since these techniques are not designed for the analysis of time series, interpretation of the results is likely to be difficult (Zuur *et al*., 2003a), or even unreliable or misleading results are obtained (Márkus *et al*., 1999). In addition, these methods cannot handle missing values properly and do not take the time order of the data into account (Solow, 1994; Zuur *et al*., 2003a). DFA is a more appropriate methodology, because it is a statistical dimension reduction technique especially designed for the analysis of multivariate time series data. Applying DFA to a set of time series allows determining common patterns (underlying effects) in the series, whether there are interactions between the time series, and the relationships between the series and selected explanatory variables. 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 DFM in which the elements are allowed to be stochastic. Following Lütkepohl (1991) and Zuur *et al*. (2003b) the DFM is given by

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

The aim of DFA is to choose M as small as possible while still obtaining a reasonable fit. Although increasing the number of common trends (M) leads to a better model fit, it results in more information that needs to be interpreted. Therefore, M should be much smaller than N. According to Zuur *et al*. (2003b), the scheme described in Equation (1) allows for evaluating the effect of explanatory variables in the N time series. The mathematical form of this DFM is given by

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

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

where $s_n(t)$ is the value of the nth response variable at time t (with $1 \le n \le N$); $\alpha_m(t)$ is the mth unknown trend (with $1 \le m \le M$) at time t; $\gamma_{m,n}$ represents the

DOI: 10.1002/hyp

unknown factor loadings; μ_n is the nth constant level parameter for shifting up or down each linear combination of common trends (i.e. it is the intercept term in the regression DFM); $\beta_{k,n}$ represents the unknown regression parameters (with $1 \leq k \leq K$) for the K explanatory variables $\upsilon_k(t)$; $\varepsilon_n(t)$ 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 simplest approach to model the error covariance matrix is to use a diagonal matrix. Regarding the estimation of the unknown parameters in the model, these are obtained with the Expectation Maximization (EM) algorithm (Dempster *et al*., 1977; Shumway and Stoffer, 1982; Wu *et al*., 1996) instead of using direct optimization of a maximum likelihood criterion (Harvey, 1989). This allows performing the DFA with a larger number of response variables. 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 v2·4·5 (Highland Statistics Ltd., Newburgh, UK, www.brodgar.com), which was used in this study. A complete and detailed description of DFA can be found in Zuur *et al*. (2003b).

The results from the DFA are interpreted in terms of the regression parameters $\beta_{k,n}$, the canonical correlation coefficients, and the match between model estimations and observed values. The goodness-of-fit of the model was assessed by visual inspection of the observed versus predicted N time series, the coefficient of efficiency (*Ceff*) (Nash and Sutcliffe, 1970) and Akaike's Information Criterion (AIC) (Akaike, 1974). The *Ceff* 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. The *Ceff* was calculated here by using the formulation given in Muñoz-Carpena *et al*. (2005). The AIC is a statistical criterion for model selection that combines the measure of fit with a penalty term based on the number of parameters used in the DFM. The more parameters (i.e. number of trends or explanatory variables) used, the better the fit, but the penalty increases for each parameter added into the model. Thereby, the DFM with the smallest AIC represents an appropriate model. The degree to which each of the response time series (s_n) is related to each of the common trends (α_m) that result from the DFA was assessed by canonical correlation coefficients ($\rho_{m,n}$). These allow for quantifying the cross-correlation between a response variable and a common trend, so that if $\rho_{m,n}$ is close to 1, it indicates that the corresponding response variable follows the pattern of the common trend. The terms 'high', 'moderate', and 'weak' correlation were used to denote $|\rho_{m,n}| > 0{\cdot}75$, 0·50–0·75, and 0·30–0·50, respectively. The influence or weight of each explanatory variable υ_k on each s_n is given by the regression parameters, $\beta_{k,n}$. Thus, standard errors for each $\beta_{k,n}$ obtained from the Kalman filter were used to assess whether the response variables were related to the explanatory variables.

DFA variables and analysis procedure

Response variables, s_n. The analysis was conducted using, as response variables, the 18 time series of groundwater [N-NO_3^-] (mg l^{-1}) that were obtained individually from each of the wells distributed across the field (Figure 1). At every well, the [N-NO_3^-] values corresponded to the 56 sampling dates during a 3-year period (15/02/2000–18/02/2003).

Explanatory variables, υ_k. From a practical standpoint, groundwater chemical variation is a function of chemical inputs, outputs, and transformation. In the case of extremely permeable gravelly soils above porous limestone rocks and drained agricultural lands like those in the study, groundwater [N-NO_3^-] changes are likely to be driven by different processes: lateral inflow and outflow to and from the canals; chemical transformations; leaching from the topsoil, which in turn depends on crop fertilizer applications; topsoil enrichment (saturation); rainfall and irrigation intensity; and the length of the transport flow path (water table depth), among other effects. On the basis of this, five observed time series were used as potential explanatory variables in the DFA: (i) *WTE* (m NGVD 1929 datum); (ii) *GwFD*; (iii) rainfall totals (cm) between sampling periods, *Prec*; (iv) irrigation totals (cm), *Irr*; and (v) nitrogen fertilizer applications, *Fert* (mg N l^{-1}). The *WTE* was computed as the average *WTE* recorded at wells *W2*, *W4*, and *W18*. *GwFD* is locally affected by seasonal regional drainage through canal management exercised by the regional water authority (SFWMD). This time series was obtained as the difference between the distal surface water levels in canal C-103 (measured northwest from the field at the S-196 structure) and the well S-196A water elevation located southeast from the field (Figure 1). Thus, positive *GwFD* values (Figure 2(a)) indicate that groundwater is flowing to the southeast (recharge from the canal). On the other hand, negative values indicate that groundwater is flowing to the northwest because the low stage at canal C-103 is draining the area. The *Prec, Irr*, and *Fert* were obtained directly from field records.

Analysis procedure. The DFA was applied to data in their original units (i.e. nonstandardized data). Although this makes the factor loadings interpretation difficult, we found that by using nonstandardized data, the DFM performance was improved. In addition, this has the benefit that DFM's predictions are in the original units of the data. Normality of data is not strictly necessary, but it is beneficial for DFA (Zuur *et al*., 2003a). On the basis of probability plots, data were found to be

DOI: 10.1002/hyp

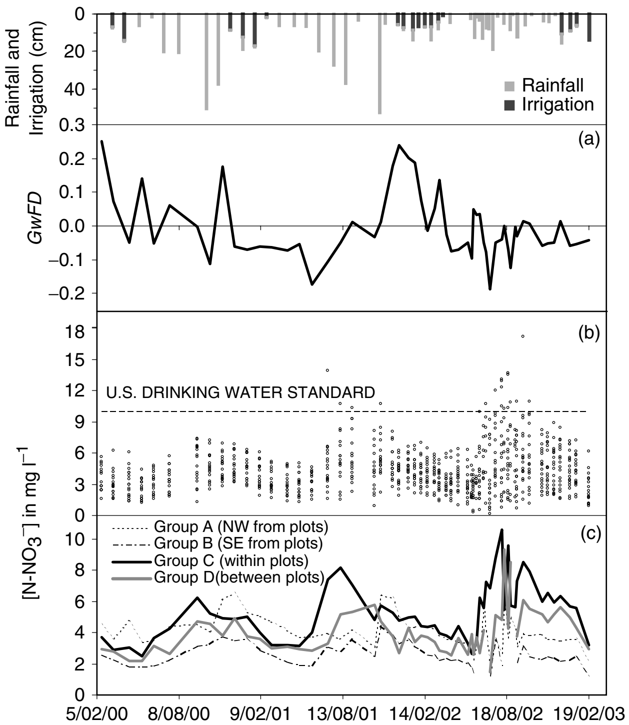

Figure 2. Summary of groundwater [N-NO_3^-] time series at the monitoring wells. (a) GwFD time series; (b) Values for all samples collected; (c) average trends for the selected groups of wells

normally distributed, so no data transformation was necessary. The analysis was conducted in three incremental steps. Firstly, with the aim of identifying relevant and appropriate explanatory variables, an exploratory analysis was conducted by visual inspection of the observed data and calculation of cross-correlation among all variables (response and explanatory). Secondly, different DFMs were compared on the basis of AIC and *Ceff.* 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 (*Ceff*), and minimum number of common trends and explanatory variables needed. Thereby the model with the smallest AIC was taken to be the appropriate model if the *Ceff* was close enough to unity. We followed this approach because AIC proposes a balance between the fit of the model and the number of parameters that the model is using to achieve this fit, while *Ceff* is a dimensionless measure that provides a relative assessment of the model performance. Thirdly, the resulting DFM was examined with respect to local land use and regional water management.

RESULTS AND DISCUSSION

Experimental time series

Figure 2(b) shows the time series of [N-NO_3^-] resulting from a total of 1008 water quality samples (excluding field and instrument blanks) along with the total rainfall and irrigation between sampling periods. Generally, [N-NO_3^-] in all these samples were below 10 mg l^{-1} (US drinking water standard) except for some samples collected during the summer and fall 2001 and 2002.

Four groups of wells were established with respect to the diagonal cropping strip: Group A, wells located NW from the plots (*W15, W16, W17, W18*); Group B, wells located SE from the plots (*W1, W2, W3, W4*); Group C, wells within the cultivated plots (*W5, W7, W8, W10, W12, W14*); and Group D, wells between the plots (*W6, W9, W11, W13*). Figure 2(c) depicts the average fluctuations of [N-NO_3^-] for the four groups of wells. Average, minimal, and maximal concentrations during the 3-year period and for each of the four groups are given in Table I. In general, groundwater [N-NO_3^-] was low during the sweet corn growing period (November–March), which coincides with the lower rainfall periods. The [N-NO_3^-] peaks can be related to summer rainfall events and, to a lesser degree, to irrigation events. Samples collected from the wells located within the cultivated plots (Group C) had higher [N-NO_3^-] than those obtained from the wells located between plots (Group D). In addition, during the rainy periods, the [N-NO_3^-] corresponding to well Groups C and D were greater than those observed in groups A and B. This suggests that not all the nitrate present in the soil was absorbed by the corn plants, so it accumulated in the soil and was later leached as a consequence of heavy rains. Figure 2(c) also depicts a similar pattern in the average [N-NO_3^-] of well groups A and B, where average [N-NO_3^-] values for Group A were about 2 mg l^{-1} higher than those for Group B. All these data give an indication of the complexity of the aquifer system, where three groundwater flow classes have been reported (Cunningham *et al.*, 2004) on the basis of unique categories of lithology and kinds of pore systems that characterize the heterogeneous and anisotropic porosity of the underlying coral limestone that make up the top of the Biscayne Aquifer.

Dynamic factor analysis

Analysis of cross-correlation. Relevant cross-correlations (results not shown) were obtained between the response time series. On the one hand, computed coefficients indicated cross-correlation for [N-NO_3^-] among the Group A wells (0·48–0·93); among the Group B wells

Table I. Well grouping according to location within the field

Group	Description	Number of wells	[N-NO_3^-][a] (mg l^{-1})	Range (mg l^{-1})
A	NW from plots	4	$4·07 \pm 0·14$	1·95–6·55
B	SE from plots	4	$2·55 \pm 0·10$	1·20–4·40
C	Within cultivated plots	6	$5·23 \pm 0·25$	2·5–10·62
D	Between cultivated plots	4	$3·99 \pm 0·20$	1·48–9·34

[a] Average concentration ± standard error calculated for the whole period.

DOI: 10.1002/hyp

(0·44–0·51); and among the Group C wells (0·53–0·85). In Group D, this was also the case for all the wells (0·48–0·69) except for *W13* that showed correlation only with *W4* with a cross-correlation coefficient of 0·47. On the other hand, cross-correlations in [N-NO_3^-] were found between wells in groups A and B (0·51–0·76), and between wells in groups C and D (0·48–0·66) as well. The explanatory variables (*Fert, WTE, GwFD, Prec,* and *Irr*) showed very low cross-correlations (0·02–0·31), indicating that they were independent and thus appropriate for the DFA.

DFM selection. Bearing in mind that the DFA aims to model the variation observed in a set of time series in terms of several components, according to the number of common trends and combinations of explanatory variables considered, the analysis could proceed with different alternative models (Table II). The DFM with a minimal number of common trends and no explanatory variables that best described [N-NO_3^-] in the 18 wells (minimum AIC = 2731 and *Ceff* = 0·819) included five common trends ($M = 5$). In contrast to visual inspection of the [N-NO_3^-] time series that showed no discernible effects (Figure 2(c)), this result indicates that five latent effects influence the chemical concentration over time. This means that the variation observed in the response variables can be represented by five underlying pattern or effects, which are common to all or part of the response time series. These common trends represent latent effects influencing the time series, but their physical explanation remains unknown. The inclusion of explanatory variables in the DFM may be useful to reduce this nonidentified variability, because the resulting DFM thereby accounts for the variation that is a consequence of known physical processes (i.e. the explanatory time series). Thus,

Table II. Selection of dynamic factor models based on performance coefficients

Explanatory variables, v_k	Trends	AIC	*Ceff*[a]
—	**5**	**2731**	**0.819**
WTE, GwFD, Prec	**3**	**2720**	**0.780**
WTE, GwFD, Prec, Irr	3	2730	0·790
Fert, WTE, GwFD, Prec, Irr	4	2736	0·813
WTE, GwFD	3	2740	0·789
WTE, Prec	3	2745	0·793
WTE, Irr, Prec	3	2749	0·798
Prec	3	2761	0·736
WTE	3	2770	0·764
GwFD, Prec	2	2773	0·736
WTE, GwFD, Irr	3	2776	0·769
Irr	3	2779	0·742
GwFD	3	2781	0·746
GwFD, Irr, Prec	2	2783	0·743
WTE, Irr	3	2785	0·767
GwFD, Irr	3	2796	0·763
Fert	3	2798	0·740

[a] *Ceff* was calculated with the combined set of predictive versus observed values for all the wells. Bold characters indicate the best DFMs for both approaches, with and without explanatory variables (v_k).

in an attempt to reduce M, explanatory variables were added. The results obtained for the models with different combinations of explanatory variables and common trends are given in Table II. Notice that by including the *WTE, GwFD,* and *Prec*, M was reduced to three common trends, resulting in the lowest AIC. Although the inclusion of these explanatory variables decreased *Ceff*, the obtained *Ceff* = 0·780 was still considered satisfactory. Interestingly, these results indicate that, compared to the other explanatory variables included in the study, the information available about irrigation and fertilizer applications is not found relevant for modeling the groundwater concentration variation during the experimental period. This likely indicates that soil/water/plant-delayed processes masked these factors. Although irrigation application did not result in sufficient leaching during the growing season, the soil enrichment likely prepared the soil for high leaching during the intense summer rainy season.

DFA of the best DFM. According to the previous section, the best DFM contained three common trends and the explanatory (hydrological) variables. The estimated regression parameters of the best DFM for these variables (*WTE, GwFD,* and *Prec*) were calculated, as well as the factor loadings and level parameters (results not shown). Table III presents a matrix summarizing the interactions of the model components for the time series in the 18 wells.

Outside the cultivated area (except for *W16*), the [N-NO_3^-] in all of the wells was affected by one or more of the selected hydrological variables. *GwFD* influenced all the SE wells (Group B) and the well *W17*. Because of the high porosity and shallow water table, *Prec* and *WTE* were expected to have a large influence on the observed [N-NO_3^-]. However, outside the corn plots, *Prec* was found to strongly influence the concentration at just three wells, *W3, W15*, and *W18*. The variable *WTE* had some effect on the concentrations at *W17* and *W18*. Only the [N-NO_3^-] in *W3* was under the influence of all three explanatory variables. For the well groups in the crop strip (Groups C and D), the effect of the explanatory variables was observed in half of the wells (Table III). The effect of these hydrological variables on the [N-NO_3^-] in the individual wells appeared to be a function of the well location within the whole field. *WTE* affected wells located at the northern boundary plus some wells at the center of the field. The important variable in the western and southern section of the field (i.e. the areas that would be considered southeast from the plots) was *GwFD*. *Prec* predominated in the eastern part of the field. Within and between the cultivated plots the influence of these explanatory variables was not consistent. This is likely due to nitrate leaching and lateral groundwater transport from the cultivated plots that may mask the influence of the hydrological variables on the [N-NO_3^-] in the corresponding wells.

So far, we have discussed the groundwater [N-NO_3^-] time dependent variation that can be explained by the

DOI: 10.1002/hyp

Table III. Summary of relative effect of explanatory variables and trends on groundwater [N-NO_3^-]

Location/land use	s_n	*WTE*[a]	*GwFD*[a]	*Prec*[a]	Trend 1[b]	Trend 2[b]	Trend 3[b]	*Ceff*
SE/fallow (Group B)	*W1*	—	+	—	***	** (−)	* (−)	0·749
	W2	—	+	—	**	—	—	0·749
	W3	+	+	+	*	—	—	0·761
	W4	—	+	—	**	—	—	0·478
Crop strip (Groups C and D)	*W5*	—	+	—	—	**	—	0·522
	W6	—	—	—	—	**	**	0·823
	W7	—	—	—	—	***	*	0·856
	W8	+	+	—	—	***	**	0·728
	W9	—	—	+	—	*	*	0·462
	W10	+	—	—	—	***	**	0·621
	W11	—	—	—	—	*	**	0·597
	W12	—	—	—	—	**	*	0·820
	W13	—	+	+	**	—	* (−)	0·704
	W14	—	—	—	—	**	—	0·532
NW/fallow (Group A)	*W15*	—	—	+	*	—	—	0·632
	W16	—	—	—	***	—	—	0·778
	W17	+	+	—	***	** (−)	* (−)	0·985
	W18	+	—	+	**	** (−)	* (−)	0·574

[a] + indicates a significant strong influence (t-value >2) of the explanatory time series (v_k) on the corresponding response variable (s_n).
[b] * denote the relative importance of the common trends based on ρ_n, where *,**,** correspond to a $\rho_n = 0{\cdot}3–0{\cdot}5$, $0{\cdot}5–0{\cdot}75$, $>0{\cdot}75$, respectively.
(−) Indicates a negative correlation.

hydrological time series. However, the information contained in the selected explanatory time series alone is not sufficient to model the groundwater [N-NO_3^-] in the wells. The DFA detected an unexplained variation in the response series too, which is reduced to and represented by the resulting common trends. This means that, in addition to the influence of the hydrological time series, groundwater [N-NO_3^-] in this area was also affected by the three underlying common patterns shown in Figure 3. Although interpretation of these trends is difficult, the corresponding canonical correlation coefficients, $\rho_{m,n}$, included in Figure 3 give information about how the [N-NO_3^-] in each of the wells is related to each of the trends. Thus, while the first trend was only correlated with wells outside the cultivated plots (and *W13* inside) (Figure 3a), the second and the third trends were positively correlated to all the wells in the cultivated area (except for *W13*). According to these results, two main groups can be clearly distinguished according to their position within the cultivated area or the surrounding fallow area. *W13* was not influenced by the same latent effects as the remainder of the wells in the cultivated area. The [N-NO_3^-] in wells from Group C (*W5, W7, W8, W10, W12* and *W14*) were better correlated with the second trend than those of Group D (*W6, W9* and *W11*) (Figure 3b). This was not observed with the third trend (Figure 3c). The [N-NO_3^-] in wells *W1, W17*, and *W18* showed also a negative weak correlation to the second and third trends (Figure 3b and 3c). Since the three common trends encompass sources of unexplained variability, these represent other aspects not included in the model that have some influence on the groundwater [N-NO_3^-]. Examples of this might be [N-NO_3^-] in the adjacent canal, and [N-NO_3^-] that accumulates in the soil and is leached to the groundwater (Muñoz-Carpena *et al*., 2005).

The DFA was useful for studying the observed time series of groundwater [N-NO_3^-] in the area of study. Although the visual inspection of these time series showed no discernible effects, the DFA indicated that [N-NO_3^-] patterns in the 18 wells depend on the location of the monitoring well within the field. Thereby, subareas in the experimental site were identified where groundwater nitrate concentration was affected by common effects (hydrological variables and unknown effects). Further interpretation of results for explaining how the system works is not possible, especially because of the complex nature of the coral limestone aquifer system, which is characterized by heterogeneous and anisotropic porosity that gives rise to a mosaic of semiconfining units and preferential flow zones (Cunningham *et al*., 2004).

The performance of the selected best DFM to describe the [N-NO_3^-] in the 18 wells is given in Figure 4. Although at some of the wells the DFM failed to match isolated concentration peaks, in general the selected DFM described satisfactorily the groundwater [N-NO_3^-] fluctuations, as expected from the high *Ceff* obtained.

CONCLUSIONS

In a 3-year study in an agricultural field in South Florida, observation of groundwater [N-NO_3^-] in 18 wells showed no discernible effects. The application of DFA successfully identified the extent to which groundwater nitrate fluctuations were explained by the hydrological variables studied. The [N-NO_3^-] patterns observed from samples collected during the 3-year period at the 18 wells were influenced by three hydrological variables:

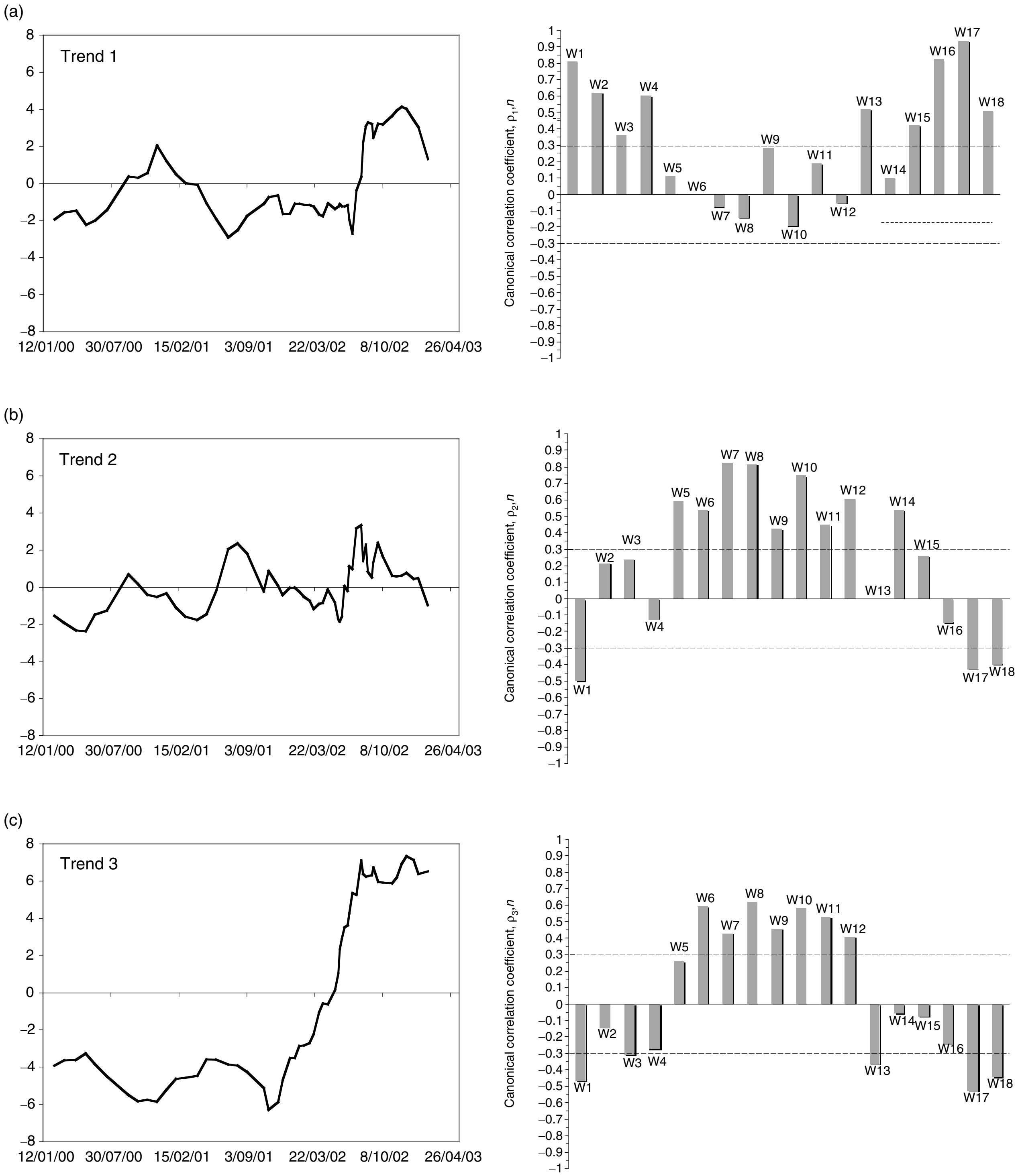

Figure 3. Common trends and corresponding canonical coefficients for the best DFM. Dashed lines indicate the threshold for weak correlations. Trend units multiplied by factor loadings units yield mg l^{-1}

water table elevation, groundwater flow direction, and precipitation. The groundwater gradient had the greatest impact upon well concentrations southeast from the plots. While precipitation and water table elevation affected the concentrations observed at five wells on the cultivated plots, their influence was not as large as what may have been expected because of the highly porous nature of the soil. Although not all the wells were similarly affected by these variables, they could be grouped according to their spatial location and the hydrological variables having influence on them. It can be concluded that much of the groundwater [$N\text{-}NO_3^-$] variation observed was successfully accounted for by these hydrological time series. The remaining variation was related to the unexplained variability represented by three common trends. The effect of these common trends on the observed [$N\text{-}NO_3^-$] showed a spatial structure across the field area allowing the effective grouping of the wells by land use,

DOI: 10.1002/hyp

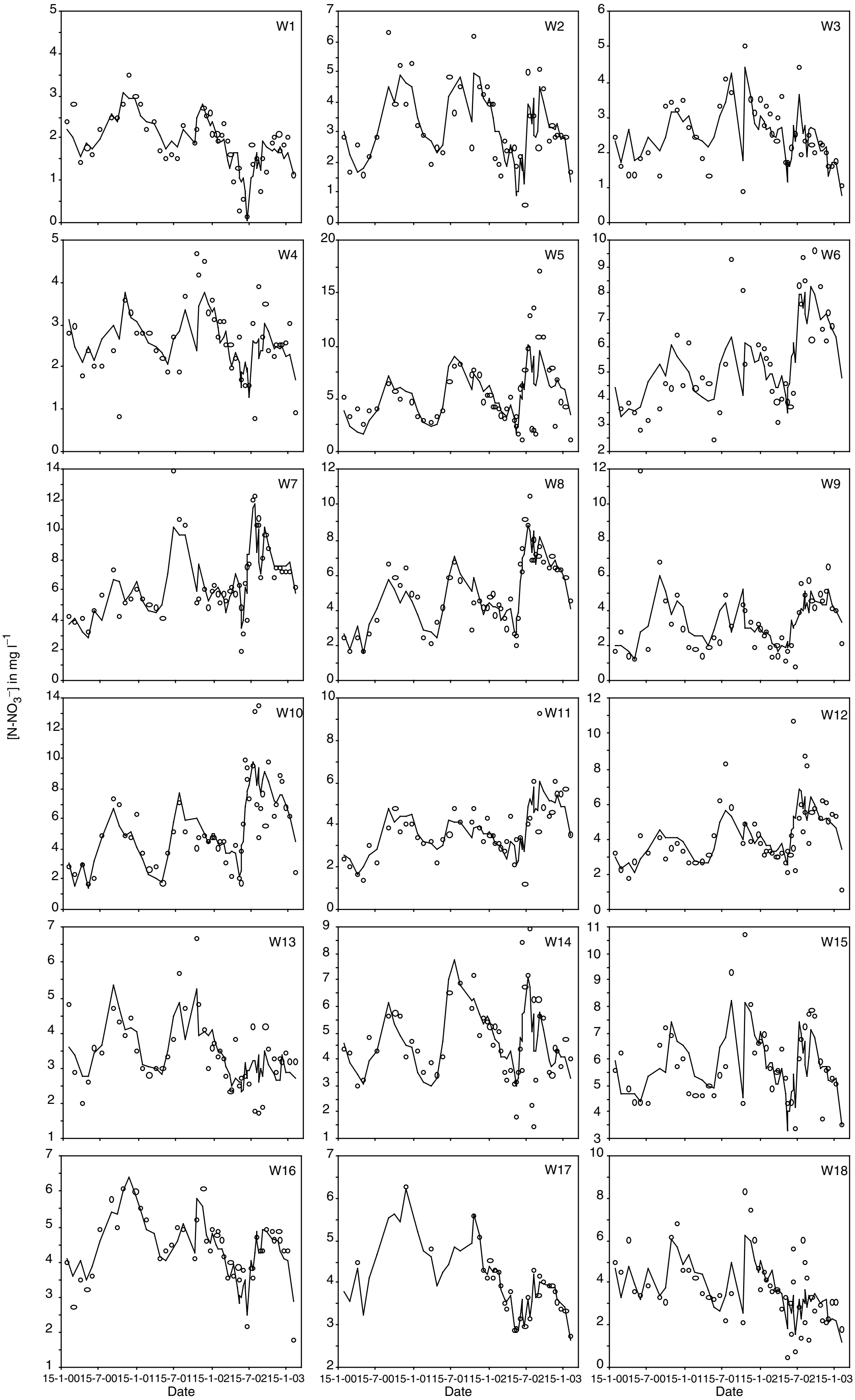

Figure 4. Dynamic Factor Model fit for [N-NO_3^-] in mg l^{-1} for the 18 experimental wells

DOI: 10.1002/hyp

i.e. cultivated and fallow. Chemical aspects such as the [N-NO_3^-] in the soil and the adjacent canal may also have affected the groundwater [N-NO_3^-]. This effect is likely encompassed by the common trends.

The results of this study indicate the complex nature of the coral limestone aquifer system in this region. Shallow water table conditions, rapid infiltration, and extremely fast groundwater dynamics make it difficult to evaluate the factors that most directly influence groundwater loading using visual inspection or comparative statistical techniques. These confounding effects are in fact present to different degrees in most groundwater quality studies. Despite this complexity, DFA was able to identify the wells expected to be influenced by the crop. In addition, groundwater concentrations southeast from the plots were influenced by flow direction resulting from the regional water management exercised by the regional water authority (SFWMD) through the operation of the drainage canal network. This means that local land use/management effects (agriculture/fallow) can be effectively separated from those of a more regional nature. This information should prove useful in devising water quality strategies and studies in the area.

In addition to representing a useful technique for studying the interactions among the variables affecting the complex aquifer system, DFA provides a model that can also prove useful for agricultural and groundwater management.

ACKNOWLEDGMENTS

The authors would like to thank the field technicians Mr Michael Gutierrez and Mrs Tina T. Dispenza for their contribution. Partial support was received from the University of Florida Center for Natural Resources (UF-CNR) and the South Florida Water Management District (SFWMD). This research was also supported by the Florida Agricultural Experiment Station, and approved for publication as Journal Series No. R-11051.

REFERENCES

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–723.

Al-Yahyai R, Schaffer B, Davies FS, Muñoz-Carpena R. 2006. Characterization of soil-water retention of a very gravelly-loam soil varied with determination method. *Soil Science* **171**(2): 85–93. DOI:10.1097/01.ss.0000187372.53896.9d.

Cunningham KJ, Carlson JL, Wingard GL, Robinson E, Wacker MA. 2004. Characterization of aquifer heterogeneity using cyclostratigraphy and geophysical methods in the upper part of the karstic Biscayne aquifer, southeastern Florida. US Geological Survey Water Resources Investigations Report 03–4208, US Geological Survey: Tallahassee, Florida; 66.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society Series B* **39**: 1–38.

ES&E. 1996. *Contamination Assessment Report Addendum for the Tropical Research and Education Center, Homestead, FL.* Report prepared for the Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL by Environmental Science & Engineering, Inc.: Gainesville, FL.

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

Genereux D, Guardiario J. 1998. A canal drawdown experiment for determination of aquifer parameters. *Journal of Hydrologic Engineering* **3**: 294–302.

Genereux D, Guardiario J. 2001. A borehole flowmeter investigation of small-scale hydraulic conductivity variation in the Biscayne Aquifer, Florida. *Water Resources Research* **37**: 1511–1517.

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

Geweke JF. 1977. The dynamic factor analysis of economic time series models. In *Latent Variables in Socio-economic Models*, Aigner DJ, Goldberger AS (eds). North-Holland: Amsterdam; 365–382.

Harman-Fetcho JA, Hapeman CJ, McConnell LL, Potter TL, Rice CP, Sadeghi AM, Smith RD, Bialek K, Sefton KA, Schaffer BA, Curry R. 2005. Pesticide occurrence in selected South Florida canals and Biscayne Bay during high agricultural activity. *Journal of Agricultural Food & Chemistry* **53**: 6040–6048.

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: New York.

Li YC, Bryan HH, Klassen W, Lamberts M, Olczyk T. 2002. *Sweet Corn Production in Miami-dade County, Florida. Fact Sheet HS-862, Florida Cooperative Extension Service*. Institute of Food and Agricultural Sciences, University of Florida: Gainesville, Florida; http://edis.ifas.ufl.edu/TR013, (last accessed July 25, 2005).

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.

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**(1–2): 49–70, DOI:10·1016/j.jconhyd.2005·07·003.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models. Part 1-A discussion of Principles. *Journal of Hydrology* **10**: 282–290.

Noble CV, Drew RW, Slabaugh V. 1996. *Soil Survey of Dade County Area*. Florida. U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, DC.

Potter TL, Bosch DD, Joo H, Schaffer B, Muñoz-Carpena R. 2004. *Demonstration and Evaluation of South Dade Basin Vegetable Crop Best Management Practices: Summer Cover Crops to Control Herbicide and FertilizerLeaching*, Contract Number: C-12331. USDA Agreement Number: 58-6602-1-211. USDA Miscellaneous Publication 1343, USDA-ARS Southeast Watershed Research: Tifton, GA.

Ritter A, Muñoz-Carpena R. 2006. Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park. *Journal of Hydrology* **317**(3–4): 340–354, DOI:10·1016/j.jhydrol.2005·05·025.

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.

Solow AR. 1994. Detecting change in the composition of a multispecies community. *Biometrics* **50**: 556–565.

USDA-NRCS. 2004. *Classification of the Soils (Soil Survey of Dade, Florida)*. USDA-Natural Resources Conservation Service, Washington, DC, available on-line at http://soils.usda.gov/survey/online_surveys/florida, (verified on March 14, 2004).

Wu LS-Y, Pai JS, Hosking JRM. 1996. An algorithm for estimating parameters of state-space models. *Statistics & Probability Letters* **28**: 99–106.

Zuur AF, Fryer RJ, Jolliffe IT, Dekker R, Beukema JJ. 2003b. Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics* **14**: 665–685.

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. 2003a. Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences* **60**: 542–552.