



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Hydrology 317 (2006) 340–354

Journal
of
Hydrology

www.elsevier.com/locate/jhydrol

Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park

A. Ritter^{a,b}, R. Muñoz-Carpena^{a,*}

^a*Agricultural and Biological Engineering Department, University of Florida, 101 Frazier Rogers Hall,
P.O. Box 110570 Gainesville, FL 32611-0570, USA*

^b*Departamento de Ingeniería, Producción y Economía Agraria, Universidad de La Laguna, Ctra. Geneto, 2, 38200 La Laguna, Spain*

Received 9 September 2004; revised 15 April 2005; accepted 25 May 2005

Abstract

The extensive eastern boundary of Everglades National Park (ENP) in south Florida (USA) is subject to one the most expensive and ambitious environmental restoration projects in history. Understanding and predicting the interaction between the shallow aquifer and surface water is a key component for fine-tuning the process. The Frog Pond is an intensively instrumented agricultural 2023 ha area adjacent to ENP. The interactions among 21 multivariate daily time series (ground and surface water elevations, rainfall and evapotranspiration) available from this area were studied by means of dynamic factor analysis, a novel technique in the field of hydrology. This method is designed to determine latent or background effects governing variability or fluctuations in non-stationary time series. Water levels in 16 wells and two drainage ditch locations inside the area were selected as response variables, and canal levels and net recharge as explanatory variables. Elevations in the two canals delimiting the Frog Pond area were found to be the main factors explaining the response variables. This influence of canal elevations on water levels inside the area was complementary and inversely related to the distance between the observation point and each canal. Rainfall events do not affect daily water levels significantly but are responsible for instantaneous or localized groundwater responses that in some cases can be directly associated with the risk of flooding. This close coupling between surface and groundwater levels, that corroborates that found by other authors using different methods, could hinder on-going environmental restoration efforts in the area by bypassing the function of wetlands and other surface features. An empirical model with a reduced set of parameters was successfully developed and validated in the area by interpolating the results from the dynamic factor analysis across the spatial domain (coefficient of efficiency across the domain: 0.66–0.99). Although specific to the area, the resulting model is deemed useful for water management within the wide range of conditions similar to those present during the experimental period.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Hydrology; Groundwater; Surface water; Dynamic factor analysis; Multivariate time series; Dynamic factor modeling; Computer simulation; Hydrological monitoring field methods; Everglades

* Corresponding author. Tel.: +1 352 392 1864; fax: +1 352 392 4052.

E-mail addresses: aritter@ull.es (A. Ritter), carpena@ufl.edu (R. Muñoz-Carpena).

1. Introduction

In the first half of the 20th century a complex drainage canal system was constructed in south

Florida to protect urban and agricultural areas against flooding. However, this regional water management also led to draining of protected natural wetland areas in the adjacent Everglades National Park, creating a negative impact on the ecosystem. In an attempt to restore the wetland environment of the Everglades, the Combined Structural and Operational Project and the Comprehensive Everglades Restoration Plan are being developed along the Park's extensive eastern boundary with the developed area (agricultural and urban). The goal of these projects is to enhance water deliveries into the Everglades while maintaining flood protection of developed areas. Implementation of these projects is a complex problem that requires detailed understanding of the hydrological processes involved. Especially, predicting the interactions between surface water flow in the canals and the shallow and extremely permeable Biscayne aquifer (Fish and Stewart, 1991) is a priority for fine-tuning the objectives at each side of the park boundary, i.e. environmental restoration and flood protection. In this context, 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 Everglades National Park, reported the almost instantaneous response of the groundwater to canal and rainfall inputs in the area. Long-term data sets containing temporal variation of hydrological variables have the potential to be used to study the surface-groundwater interactions in the area. However, data analysis based on visual inspection and descriptive statistics can be complex and may not be sufficient, especially when dealing with multiple variables. 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 as well as studying the effect of explanatory time-dependent variables (Zuur et al., 2003b). Multivariate time series

may be analyzed as response functions assuming that there are common driving forces behind them, i.e. factors or latent effects that determine the variation of the individual observations. These factors 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 time series models (Geweke, 1977) that has been recently used with variations in disciplines like psychology (Molenaar, 1985, 1989, 1993; Molenaar and de Gooijer, 1988; Molenaar et al., 1992, 1999) and economics (Harvey, 1989; Lütkepohl, 1991). Lately, Zuur and Pierce (2004) used dynamic factor analysis for fisheries applications, while Mendelsohn and Schwing (2002) applied it to large oceanographical time series. Márkus et al. (1999) applied dynamic factor analysis in hydrology to identify common patterns of groundwater level in a karstic area of Hungary. Although they identified two common trends as recharge (infiltration) and extraction (pumping), no explanatory variables were included in the model. In addition, since the timing of water level measurements available in the study was not systematic, this study considered only annual average water elevation, and information related to seasonality was lost.

The objective of this study was to apply dynamic factor analysis and modeling to study the interactions between daily time series of hydrological variables obtained from a heavily instrumented small agricultural area in the boundary of the Everglades National Park. The analysis was conducted in three steps: (i) identification of common trends of ground and surface water levels; (ii) inclusion of explanatory variables in the multilinear DFA model; (iii) extension of results to the spatial field domain to simulate observed values.

2. Materials and methods

2.1. Experimental set-up

The study was conducted at the Frog Pond (Fig. 1), a small area of 2023 ha located at the boundary of Everglades National Park in Homestead, Florida. This public land was leased for the last 11 years to a group of growers that farm under restricted conditions (low inputs and limited flood protection). The area adjoins

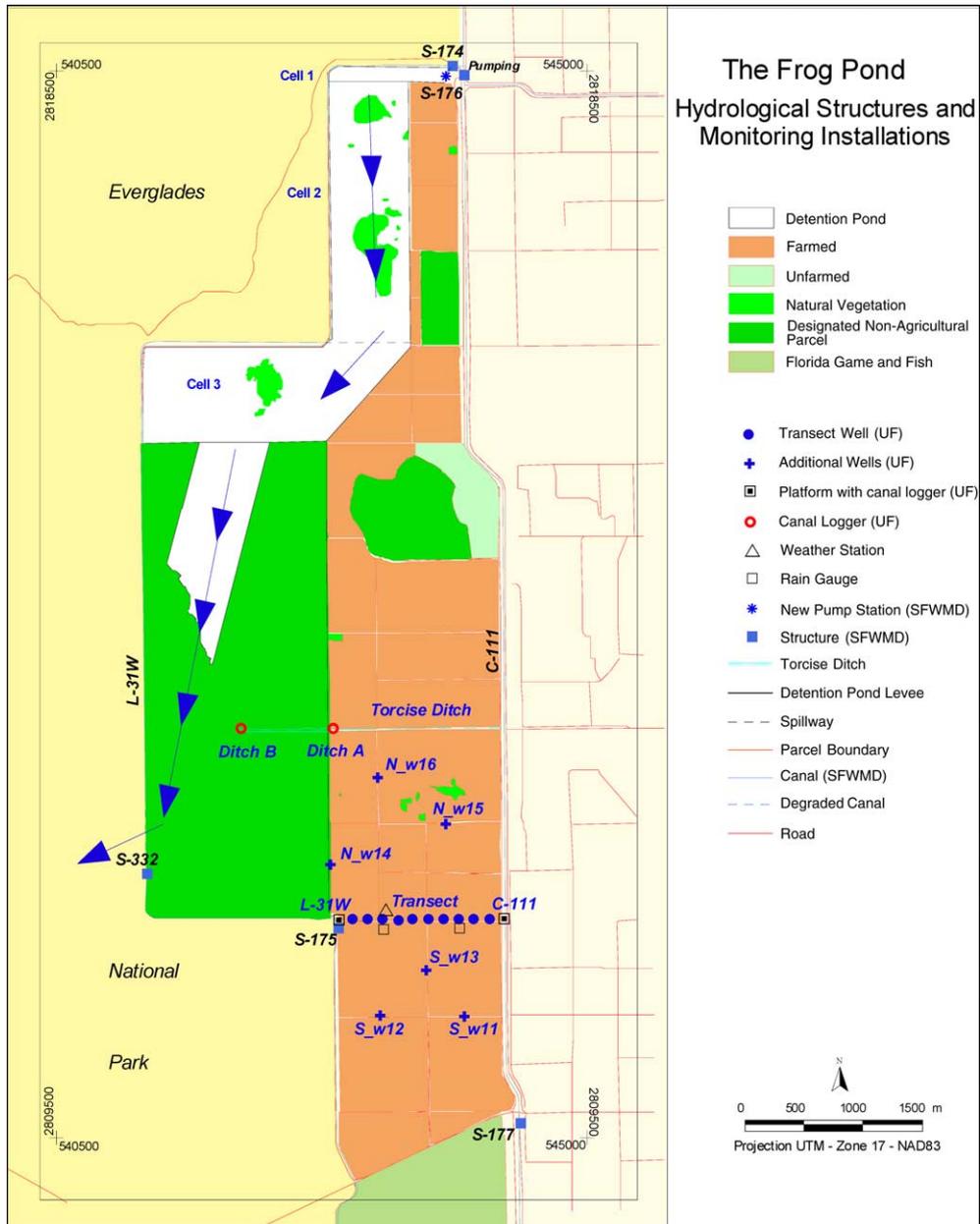


Fig. 1. Frog Pond area monitoring network.

two canals managed by the South Florida Water Management District (SFWMD) regional network: C-111 (west) and L-31W (east). Water level in both canals is regulated by remotely operated structures S-177 (spillway) and S-175 (culvert), respectively. Under the Combined Structural and Operational

Project operations, the water level in canal L-31W is maintained high in order to increase water delivery into the Everglades, while pushing agricultural return flows away to the east. This is achieved by keeping the gate at structure S-175 permanently closed while pumping water from canal C-111 into the L-31W in

the northern part of the Frog Pond. In addition a system of detention ponds and a $14.2 \text{ m}^3/\text{s}$ pumping station (Fig. 1) was constructed in June 2002 in accordance with new environmental regulations in the area (USACE, 2002). This system influences surface and groundwater flow patterns and elevation in the area.

An extensive monitoring network distributed across the southern portion of the Frog Pond area (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 a 1.6 km transect with 10 fully instrumented wells for water elevation, two rain gauges, soil moisture sensors and an automatic weather station (Fig. 2). Wells were 5 m deep cased with 5 cm diameter PVC pipe screened at the lower 4 m section. The pipe was surrounded by a clean silica sand envelope to fill the 20 cm wide borehole and sealed at the top with 20 cm bentonite and 10 cm concrete layers where a 30 cm wide cast iron manhole was set. Wells and manholes were kept closed and locked at all but maintenance times. Groundwater levels were registered every 15 min by auto-logging pressure transducers compensated for temperature effects and atmospheric pressure (Solinst, Inc., Canada). Fifteen minutes rainfall readings were made with two auto-logging tipping-bucket rain gauges (Onset Computer Corp., USA), one each located at 1/3 and 2/3 of the distance along the main transect. Penman-Monteith potential evapotranspiration was estimated based on 15-min weather data (wind speed, solar radiation, relative humidity, air temperature, atmospheric pressure) measured with an automatic weather station (Onset Computer Corp., USA) placed on the well transect at a point 2/3 of the total length.

Canal levels at both ends of the transect (C-111 and L-31W) were obtained in this first phase from the SFWMD's online records (SFWMD, 2004) at structures S-175, S-176 and S-177.

In a second experimental phase started in February 2003, six additional fully instrumented wells and two canal (C-111 and L-31W) and two stage recorders at Torcise drainage ditch (*DitchA* and *DitchB*) were added north and south of the original transect and included in the 15-min monitoring protocol (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 canals and in the Torcise ditch were 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 ($6 \times 1 \text{ m}$) supported by pillars anchored to the banks and the bottom of the canal.

All monitoring stations (wells, ditch and canal loggers) were surveyed and georeferenced by a registered surveyor (horizontal coordinates, UTM-Zone 17 WGS-84/NAD83; elevation, m NGVD29).

2.2. Dynamic factor analysis, DFA

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, it allows for evaluation of the effect of explanatory variables. DFA is similar to

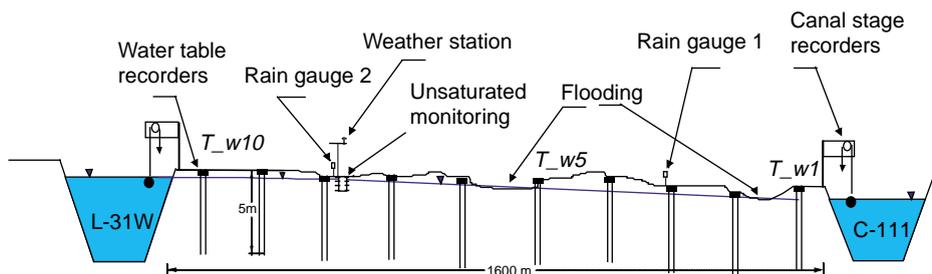


Fig. 2. Details of transect monitoring network well transect showing on-site instrumentation.

other dimension reduction techniques like factor analysis or redundancy analysis, but since it preserves temporal structure it can be used with non-stationary time series. Notice that these conventional multivariate methods usually require independent observations, which is not the case for hydrological time series (Márkus et al., 1999). Another difference between DFA and the latter techniques is that in DFA the axes or factors 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:

N time series

$$\begin{aligned} &= \text{linear combination of } M \text{ common trends} \\ &+ \text{level parameter} + \text{explanatory variables} \\ &+ \text{error component} \end{aligned} \quad (1)$$

The aim of DFA is to choose M as small as possible while 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) as

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

$$\alpha_m(t) = \alpha_m(t-1) + \eta_m(t) \quad (3)$$

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; μ_n is the n th constant level parameter for displacing up and down each linear combination of common trends; $\beta_{k,n}$ 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, which are assumed to be independent of each other and normally distributed with zero mean and unknown covariance matrix \mathbf{H} and \mathbf{Q} , for $\varepsilon_n(t)$ and $\eta_m(t)$, respectively. 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 can be 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 noticing 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) used in the 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. To assess the significance of the regression parameters, standard errors for them are included. Low values for the standard error indicate the statistical significance of the corresponding parameter. The goodness-of-fit of the model can be assessed by visual inspection, the coefficient of efficiency (Nash and Sutcliffe, 1970) and the Akaike's Information Criterion, AIC (Akaike, 1974). The coefficient of efficiency C_e compares the variance about the 1:1 line (perfect agreement) to the variance of the observed data (see Appendix A). Notice that for non-regression models the C_e 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 (Wilson, 2001). Thereby $C_e = 1$ implies that the plot of predicted vs. observed matches the 1:1 line. Statistical significance (p -value) for C_e was estimated with the bootstrap percentile-t method (Zoubir and Boashash, 1998). 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. When comparing two or more models, 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 (α_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, $\beta_{k,n}$.

2.3. Time series of hydrological variables and analysis procedure

The 21 daily time series used in the analysis were: (a) groundwater table elevations (*WTE*) given in m NGVD29 at sixteen wells; (b) surface water level (*SWL*) given in m NGVD29 in the two canals (C-111 and L-31W) and in the two Torcise ditch locations; and (c) average net recharge (*nRech*) in mm/day. *WTE* in the wells located along the transect (T_{w1} – T_{w10}), south (S_{w11} , S_{w12} , S_{w13}) and north of it (N_{w14} , N_{w15} ; N_{w16}), and *SWL* data from the two stage recorders at the ditch (*DitchA* and *DitchB*) were considered as response variables, while canals’ *SWL*, and *nRech* were selected as explanatory variables. Although the analysis of time series can be conducted using daily increments of the hydrological variables, this would partially remove information about the underlying patterns that might be important (Márkus et al., 1999). Therefore, daily-averaged data (non-stationary) were used from a period of over 2 years (796 days, 28/03/2002–31/05/2004). *nRech* contains the rainfall and evapotranspiration (ET_o) information and was calculated as the difference between cumulative daily rainfall and evapotranspiration data. Due to the high cross-correlation between the two rain gauges (0.95 , $p < 0.001$), the average of the two time series was used.

The DFA was conducted in three incremental steps. First, no explanatory variables and up to four trends were simultaneously considered in order to detect which wells were influenced by the same

underlying effects (common trends). To reduce the number of model parameters needed, the smaller number of trends to adequately represent the response variables was investigated. DFA was applied on standardized time series, because this facilitates the interpretation of factor loadings and the comparison of regression parameters. It is worth noting that although normality of data is beneficial for DFA, it is not strictly necessary (Zuur et al., 2003a). Second, the analysis was repeated, taking into account the three explanatory variables described above to look at a possible reduction in the influence of the common trends obtained in the previous step. Finally, in order to predict *WTE*, DFA was also conducted with non-standardized data and using the explanatory variables with the most impact. To assess the robustness of this model, the DFM in this last step was developed from a reduced data set used for calibration, and then validated using independent data (unused portion of the data set).

The error component ε_t in Eq. (2) is determined by the covariance matrix \mathbf{H} , whose elements represent information that cannot be explained by the common trends or the explanatory variables. Using a symmetric, non-diagonal \mathbf{H} , translates into a smaller number of common trends needed for an adequate model fit (Zuur et al., 2003a), because it also contains off-diagonal elements, that account for the joint information in two response variables that is not explained by the other terms included in the DFM. Large off-diagonal elements are an indication that the corresponding time series are not fitted well. Thereby, for all DFA the option to use a symmetric, non-diagonal covariance matrix of the error term ε_t was selected.

2.4. Spatial modeling of water levels

The DFM obtained from the last step in the analysis procedure described above is space-dependent, so that the resulting factor loadings and regression parameters are limited to each observation site and thus cannot be used at intermediate locations of the domain. Assuming that the observed response variables represent a finite sample from an infinite collection of time series continuously distributed all over the study area, empirical spatial functions of these DFM parameters can be obtained by

interpolation. In this context we performed least-square surface fitting on the relevant DFM parameters across the domain. The resulting model is site specific and should be applied only within the wide range of conditions similar to those found during the experimental period.

3. Results and discussion

3.1. Experimental time series

Daily time series of the hydrological variables (rainfall potential evapotranspiration, surface- and groundwater levels) measured at the experimental site are presented in Fig. 3.

Fig. 3a shows the seasonal variation of rainfall and ET_o , both being higher in spring and summer and lower in fall and winter. Typically, a 4-month rainy season occurs in the area from June to September, where over 60% of the total annual precipitation is collected. WTE in the wells along the transect (Fig. 3b) seem to follow the same pattern around the value of 1 m NGVD29. The higher levels were observed during the rainy season, while in May 2002 and May 2004 the levels dropped gradually to around 0.5 m NGVD29. Fig. 3b includes also WTE in the wells located north and south of the transect (in the second experimental phase). Observations in these wells differed from those in the transect. Wells located south of the transect (S_{w11} , S_{w12} , S_{w13}) showed lower WTE than the transect wells, while those to the north of the transect (N_{w14} , N_{w15} , N_{w16}) presented the highest WTE values. Fig. 3c shows the temporal variations of surface water level in the canals (C-111 and L-31W) and in the ditch. The visual inspection of SWL in C-111 and L-31W indicates the effect of the Combined Structural and Operational Project interim management strategy, so that SWL_{L-31W} is generally 0.1–0.7 m higher than SWL_{C-111} except during February 2003 when the gate at structure S-175 in canal L-31W was opened to accommodate a canal-drawdown study carried out by the SFWMD. In addition, SWL observed in the ditch at the stage recorder *DitchB* closely matches the SWL_{L-31W} . Generally, Fig. 3 suggests that rainfall and canal management can potentially affect WTE ; rainfall is especially important for explaining sharp rises

observed in WTE . As an example, complete surface flooding was experienced in at least four wells after the extreme rainfall event (84 mm) in December 2002 (T_{w5} and T_{w6}), and during the third and fourth quarters of 2003 (N_{w14} and N_{w16}). At least one time during the experimental period, 50% of the wells presented water table depth within the top 15 cm, which usually corresponds to the agricultural soil layer.

3.2. Dynamic factor analysis

DFA using water elevations at all the wells and at the ditch stations (*DitchA* and *DitchB*) as response variables, Model I (Table 1) showed that just one common trend, $\alpha_1(t)$, was sufficient for an adequate model fit ($AIC = -24755$; visual inspection; and $Ce_n = 0.94 \pm 0.02$ across the domain). The estimated factor loadings ($\gamma_{m,n}$), constant level parameters (μ_n), canonical correlation coefficients ($\rho_{m,n}$) and coefficients of efficiency (Ce_n) for this model are summarized in Table 1. The statistical significance of Ce_n was $p < 0.001$ for all response variables.

Factor loadings in Table 2 indicate that WTE in all the wells have similar positive relation to the common trend (average $\gamma_{1,n}$ value: 0.287 ± 0.013), and this correlation is very high in all cases (average $\rho_{1,n} = 0.97 \pm 0.01$). In addition, the high Ce_n showed that WTE in the area can be described satisfactorily by the common trend presented in Fig. 4a plus the constant level parameter. In this context, notice the similarity between Figs. 3b and 4a. Diagonal and off-diagonal elements of \mathbf{H} were small, so that all the information contained in the response variables is fitted well and can be explained with a single common trend.

Table 3 summarizes the results obtained from the DFA using $nRech$, SWL_{C-111} , and SWL_{L-31W} as explanatory variables (Model II). These results include the regression parameters ($\beta_{k,n}$) and their standard error for each explanatory variable. Again, the DFM with just one trend, $\alpha_1(t)$ (Fig. 4b) provided an adequate fit ($AIC = -29250$; visual inspection; and average $Cd_n = 0.96 \pm 0.02$, $p < 0.001$).

WTE in all the wells have a similar positive relation to the common trend (average $\gamma_{1,n}$ value: 0.037 ± 0.018), but the inclusion of explanatory variables reduced these factor loadings by an order of

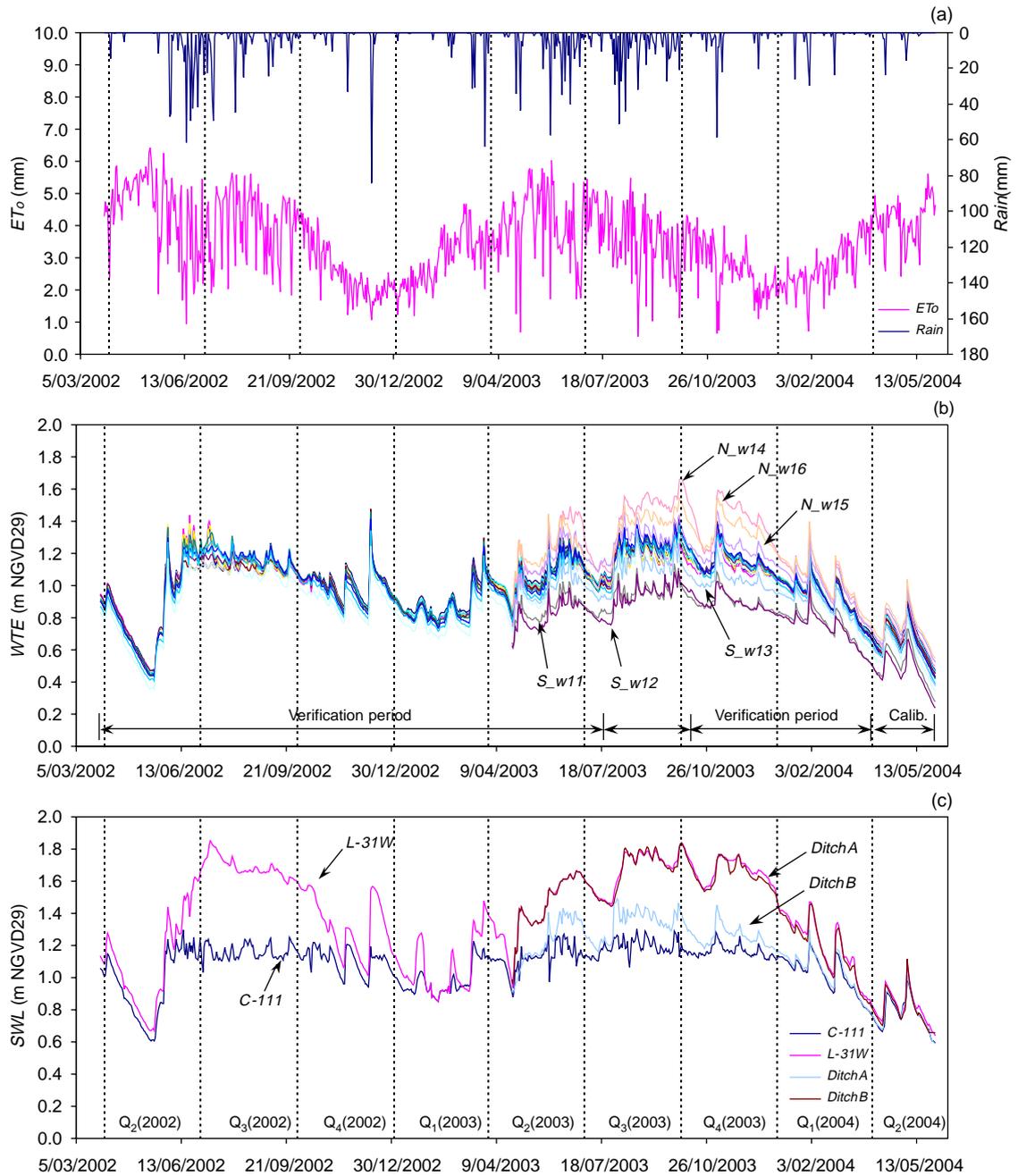


Fig. 3. Summary of hydrological time series obtained at the experimental site for the whole monitoring period.

magnitude compared to those of the previous DFM (Model I). This effect of the trend is weak ($0.41 < \rho_{1,n} < 0.64$), suggesting that patterns observed in the wells may be sufficiently described by the

explanatory variables. A visual comparison of the common trend (Fig. 4b) and the WTE time series (Fig. 3b) showed less similarity between them compared with that of Model I (Fig. 4a).

Table 1
DFM models tested in the study (see explanation in the text)

Model	No. of trends	Explanatory variables	Regression parameters	Total number of parameters	C_e
I	1	None	–	36	0.94
II	1	$nRech$, SWL_{C-111} , SWL_{L-31W}	From DFA	90	0.96
III	0	SWL_{C-111} , SWL_{L-31W}	From DFA	54	0.91
IV	0	SWL_{C-111} , SWL_{L-31W}	Interpolation	22	0.87

Regression parameters represent the intensity of the corresponding explanatory variable. Comparison of $\beta_{k,n}$ provides evidence that SWL in the canals influence the temporal changes of WTE more than the rainfall. In addition, the β_{L-31W} for $DitchB$ is 0.97 ± 0.01 , which agrees with the good match between SWL_{L-31W} and $DitchB$ observed in Fig. 3c. Keeping in mind that T_{w1} is near the eastern canal C-111 and T_{w10} is near the western canal L-31W, notice how the $\beta_{C-111,n}$ decreased from T_{w1} to T_{w10} , while for $\beta_{L-31W,n}$ an increase in the same direction was observed. This effect is also consistent north and south of the main well transect when moving from S_{w11} to S_{w13} or N_{w15} to N_{w14} . This showed that the influence of canal elevations on WTE at each well was inversely related to the distance between the well and each canal. In an attempt to visualize the effect of both canals on surface- and groundwater in the area, isolines were generated by kriging based on the β_{C-111} and β_{L-31W} in Table 3. Notice the similarity of the β isolines between Fig. 5a and b, except for their complementary values. Again, this suggests that both canals are interacting one against the other through groundwater flow, thus affecting WTE in the area. North of the transect, β isolines are almost parallel, which can be related to the general west to east groundwater flow that results from the Combined Structural and Operational Project modifications. However, in the southern part of the Frog Pond, β isolines are perturbed. A possible explanation for this effect was reported by Muñoz-Carpena et al. (2003). Based on computer simulations with the numerical model MODFLOW (McDonald and Harbaugh, 1988), they concluded that the permanently closed gate at structure S-175 (associated with the Combined Structural and Operational Project operations) has a general impact on groundwater flow south of the transect, so that the sharp gradient of around 1 m between head and tail waters of the canal structure shifts the general west to east flow to turn around the

structure with increasing speed and eventually west towards the Everglades. Genereux and Slater (1999), studying discharge data collected with acoustic velocity meters installed in the canals C-111 and L-31W, also reported a very quick transfer of groundwater from north to south of this area through the canals. These authors concluded that this transfer would bypass the overlaying wetlands water quality function and thus hinder on-going ecosystem restoration efforts.

Regression parameters for $nRech$ (β_{nRech}) in Table 3 were low compared to the canals', so that this variable, which combines the effect of rainfall and ET_o , has less impact on daily water levels. Based on this, an additional DFA was conducted with non-standardized time series (in m NGVD29) and using only SWL_{C-111} and SWL_{L-31W} as explanatory variables. This analysis was performed with a reduced data set (14% of total time series) used for

Table 2
Output results from DFA without explanatory variables (Model I)

s_n	$\gamma_{1,n}$	μ_n^a	$\rho_{1,n}$	C_{e_n}
T_{w1}	0.290	0.00 ± 0.43	0.960	0.922
T_{w2}	0.290	0.00 ± 0.43	0.957	0.917
T_{w3}	0.291	0.00 ± 0.43	0.962	0.925
T_{w4}	0.294	0.00 ± 0.43	0.972	0.945
T_{w5}	0.294	0.00 ± 0.43	0.971	0.943
T_{w6}	0.293	0.00 ± 0.43	0.968	0.937
T_{w7}	0.295	0.00 ± 0.43	0.978	0.956
T_{w8}	0.298	0.00 ± 0.43	0.984	0.968
T_{w9}	0.299	0.00 ± 0.43	0.988	0.976
T_{w10}	0.300	0.00 ± 0.43	0.991	0.982
S_{w11}	0.287	-0.07 ± 0.42	0.965	0.921
S_{w12}	0.292	-0.06 ± 0.43	0.977	0.942
S_{w13}	0.289	-0.08 ± 0.42	0.968	0.929
N_{w14}	0.272	-0.21 ± 0.40	0.978	0.950
N_{w15}	0.257	-0.21 ± 0.38	0.975	0.934
N_{w16}	0.269	-0.23 ± 0.39	0.965	0.922
$DitchA$	0.260	-0.17 ± 0.38	0.965	0.922
$DitchB$	0.299	-0.09 ± 0.44	0.982	0.956

^a Mean \pm standard error.

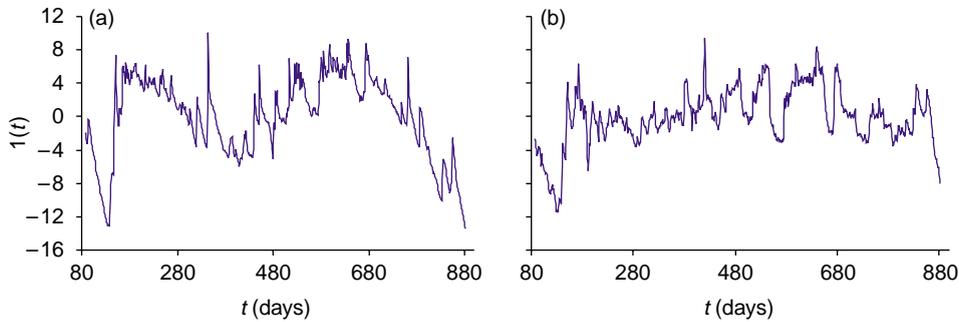


Fig. 4. Common trend $\alpha_1(t)$ obtained by DFA: (a) without explanatory variables (Model I); (b) with explanatory variables (Model II).

calibration that consisted of information from the rainy and dry seasons (periods [20/07/2003-10/11/2003] and [1/04/2004-31/05/2004]) (see Fig. 3b). The low factor loadings and weak canonical correlations obtained (columns 2–4 in Table 4) support a further simplification of the model by excluding the common trend (Model III). This DFM (using columns 3, 5 and 6 in Table 4) predicted WTE in all the wells and SWL in the ditch satisfactorily for both the calibration ($0.95 < Ce_n < 0.99$, column 7) and the verification period ($0.69 < Ce_n < 0.99$, column 8) with $p < 0.001$ in all cases. The model was validated using the remainder of the time series not used for calibration (Fig. 3b). In

addition, removing $nRech$ and the common trend from the model drastically reduced the total number of parameters required (Table 1).

Fig. 6 shows the WTE observed and predicted by Model III at the side and center main transect wells and north and south of it. The DFM successfully predicts the variation across time and space. Although the model does match the sharp rises in water elevation created by large and isolated rainfall events, it is responsive to them.

3.3. Spatial modeling

In order to extend the practical application of this DFM, regression- and constant level parameters for

Table 3
Output results from DFA with explanatory variables (Model II)

s_n	$\gamma_{1,n}$	μ_n	$\rho_{1,n}$	$\beta_{nRech,n}$	$\beta_{C-111,n}$	$\beta_{L-31W,n}$	Ce_n
T_w1	0.032	0 ± 0.08	0.450	0.10 ± 0.01	0.59 ± 0.02	0.34 ± 0.02	0.961
T_w2	0.017	0 ± 0.04	0.406	0.11 ± 0.01	0.57 ± 0.02	0.38 ± 0.02	0.945
T_w3	0.024	0 ± 0.06	0.427	0.10 ± 0.01	0.53 ± 0.02	0.41 ± 0.02	0.954
T_w4	0.037	0 ± 0.09	0.467	0.08 ± 0.01	0.47 ± 0.02	0.46 ± 0.02	0.970
T_w5	0.033	0 ± 0.08	0.456	0.08 ± 0.01	0.45 ± 0.02	0.49 ± 0.02	0.969
T_w6	0.055	0 ± 0.13	0.516	0.07 ± 0.01	0.38 ± 0.02	0.52 ± 0.02	0.967
T_w7	0.033	0 ± 0.08	0.453	0.09 ± 0.01	0.31 ± 0.02	0.62 ± 0.02	0.968
T_w8	0.039	0 ± 0.09	0.469	0.08 ± 0.01	0.25 ± 0.02	0.68 ± 0.02	0.976
T_w9	0.042	0 ± 0.10	0.475	0.07 ± 0.01	0.20 ± 0.02	0.72 ± 0.02	0.975
T_w10	0.052	0 ± 0.13	0.506	0.05 ± 0.01	0.20 ± 0.01	0.72 ± 0.02	0.976
S_w11	0.029	-0.06 ± 0.07	0.555	0.12 ± 0.01	0.64 ± 0.02	0.27 ± 0.03	0.963
S_w12	0.025	-0.05 ± 0.06	0.603	0.15 ± 0.01	0.34 ± 0.03	0.58 ± 0.03	0.945
S_w13	0.032	-0.07 ± 0.08	0.583	0.15 ± 0.01	0.48 ± 0.03	0.42 ± 0.03	0.945
N_w14	0.071	-0.24 ± 0.17	0.645	-0.05 ± 0.01	-0.01 ± 0.02	0.80 ± 0.03	0.920
N_w15	0.042	-0.20 ± 0.10	0.534	0.01 ± 0.01	0.42 ± 0.02	0.41 ± 0.03	0.954
N_w16	0.07	-0.24 ± 0.17	0.636	-0.03 ± 0.01	0.24 ± 0.03	0.57 ± 0.04	0.903
$DitchA$	0.029	-0.16 ± 0.07	0.502	0.06 ± 0.01	0.47 ± 0.02	0.38 ± 0.02	0.952
$DitchB$	-0.003	-0.12 ± 0.01	0.513	0.01 ± 0.00	0.01 ± 0.01	0.97 ± 0.01	0.995

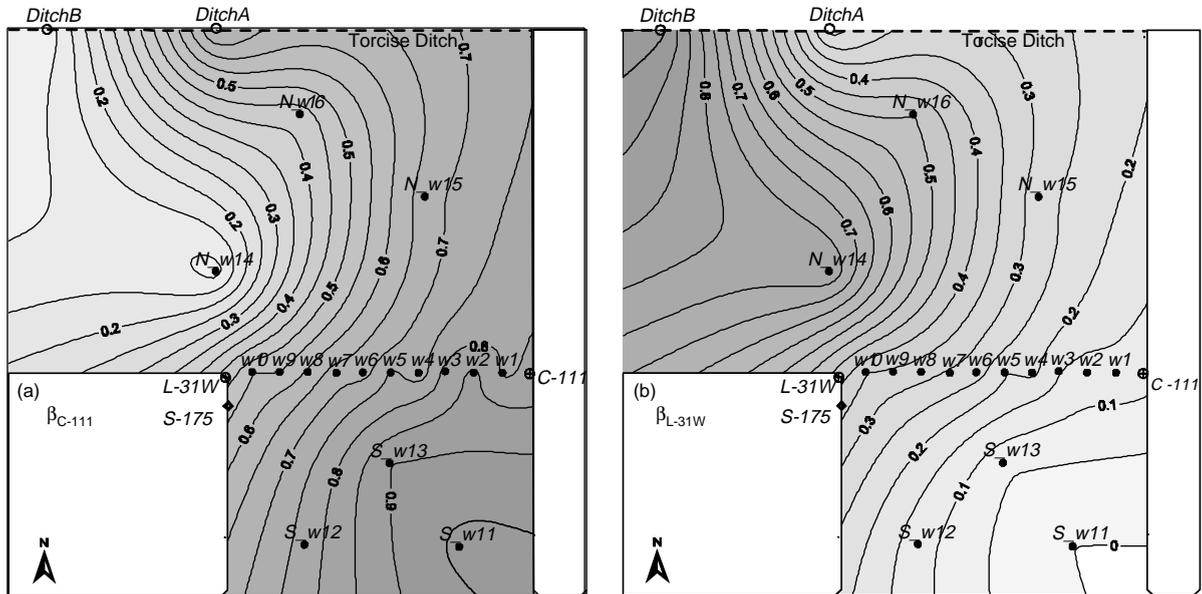


Fig. 5. Canal effects on the Frog Pond's surface- and groundwater based on β : (a) β_{C-111} ; (b) β_{L-31W} .

Table 4

Output results from DFA with non-standardized time series (Models III and IV, see explanation on text)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
s_n	$\gamma_{1,n}$	μ_n	$\rho_{1,n}$	$\beta_{C-111,n}$	$\beta_{L-31W,n}$	Ce_n^a	Ce_n^b	Ce_n^c
T_w1	0.006	-0.14 ± 0.03	0.245	0.85 ± 0.04	0.19 ± 0.02	0.978	0.908	0.787
T_w2	0.005	-0.16 ± 0.02	0.220	0.83 ± 0.04	0.20 ± 0.02	0.980	0.910	0.883
T_w3	0.005	-0.16 ± 0.03	0.226	0.80 ± 0.04	0.23 ± 0.02	0.974	0.914	0.904
T_w4	0.006	-0.14 ± 0.03	0.225	0.75 ± 0.04	0.27 ± 0.02	0.976	0.912	0.898
T_w5	0.006	-0.15 ± 0.03	0.226	0.07 ± 0.04	0.31 ± 0.02	0.975	0.911	0.921
T_w6	0.006	-0.15 ± 0.04	0.217	0.56 ± 0.05	0.42 ± 0.03	0.968	0.858	0.918
T_w7	0.006	-0.16 ± 0.03	0.217	0.51 ± 0.04	0.46 ± 0.02	0.976	0.905	0.900
T_w8	0.006	-0.14 ± 0.03	0.218	0.43 ± 0.04	0.51 ± 0.02	0.975	0.907	0.912
T_w9	0.007	-0.13 ± 0.04	0.233	0.40 ± 0.04	0.50 ± 0.02	0.973	0.917	0.932
T_w10	0.006	-0.17 ± 0.04	0.217	0.36 ± 0.04	0.52 ± 0.02	0.972	0.911	0.932
S_w11	0.006	-0.29 ± 0.03	0.259	0.87 ± 0.04	0.12 ± 0.02	0.969	0.797	0.826
S_w12	0.007	-0.23 ± 0.04	0.264	0.52 ± 0.05	0.32 ± 0.03	0.958	0.689	0.891
S_w13	0.006	-0.17 ± 0.04	0.248	0.70 ± 0.06	0.26 ± 0.03	0.952	0.724	0.845
N_w14	0.006	-0.04 ± 0.05	0.186	-0.08 ± 0.05	0.93 ± 0.03	0.972	0.909	0.715
N_w15	0.004	-0.08 ± 0.02	0.195	0.65 ± 0.04	0.35 ± 0.02	0.981	0.897	0.885
N_w16	0.005	0.01 ± 0.05	0.192	0.24 ± 0.05	0.64 ± 0.03	0.959	0.795	0.832
DitchA	0.003	-0.06 ± 0.02	0.180	0.77 ± 0.04	0.29 ± 0.02	0.979	0.825	0.660
DitchB	0.001	-0.01 ± 0.01	0.118	-0.06 ± 0.02	1.05 ± 0.01	0.998	0.991	0.988

^a Coefficient of efficiency for Model III applied to the calibration period (see Fig. 3).

^b Coefficient of efficiency for Model III applied to the verification period (see Fig. 3).

^c Coefficient of efficiency for Model IV applied to the whole period.

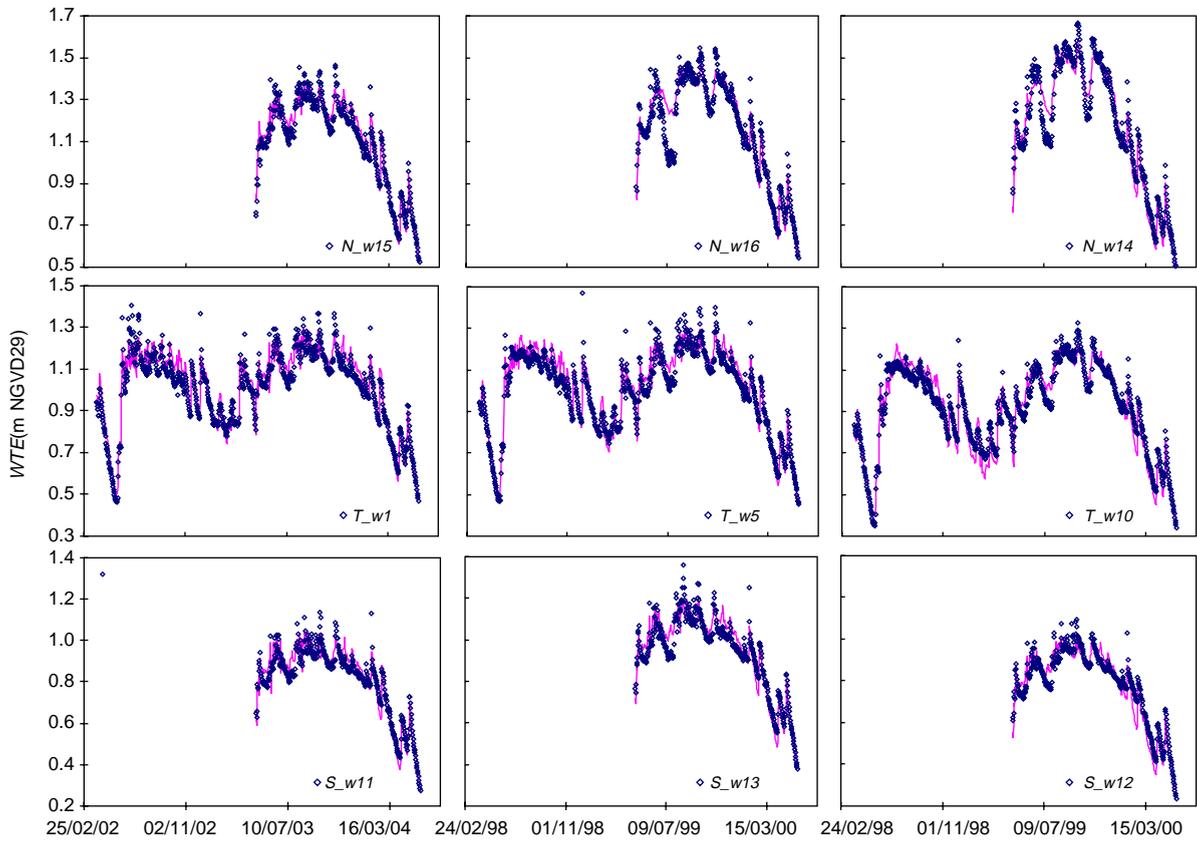


Fig. 6. Observed (symbols) and predicted (lines) WTE at the transect and at the wells north and south of it obtained with the DFA model with no trend and canal levels as explanatory variables (Model III).

both canals can be obtained at every location by using the following equations and the fitted parameters given in Table 5 (Model IV)

$$WL(X, Y, t) = SWL_{C-111}(t)\beta_{C-111}(X, Y) + SWL_{L-31W}(t)\beta_{L-31W}(X, Y) + \mu(X, Y) \quad (4)$$

$$\beta_k(X, Y) = \frac{a + b \ln X + cY + dY^2 + hY^3}{1 + e \ln X + fY + gY^2 + iY^3} \quad (5)$$

$$\mu(X, Y) = a + bX \ln X + c\sqrt{X} \ln X + \frac{d \ln X}{X^2} + \frac{e}{\sqrt{Y}} + \frac{f}{Y} \quad (6)$$

where WL (m NGVD29) stands for surface and groundwater levels across the domain. Coordinates (X, Y) are expressed in UTM (meters) and correspond

to northing and easting from WGS-84 (NAD-83), respectively. Eq. (4) derives from the general DFM Eq. (2) after applying the simplifying assumptions from Model III, while Eqs. (5) and (6) were obtained from the least-squares surface interpolation of the parameters in columns 3, 5 and 6 of Table 4. These were selected as the equations, which yielded the best fit from a set of different equations tried. The performance of this model was verified by estimating water elevations in the eighteen monitoring locations (wells and ditch). The corresponding Ce_n (column 9, Table 4) indicated that, in general, the model was acceptable ($p < 0.001$) and at the same time it required less numbers of parameters than any of the previous models tested (Table 1). The expected error (root mean square error) in predictions across the domain was found to be 0.07 ± 0.03 m.

Table 5
Empirical parameters for Eqs. (5) and (6) (Model IV)

DFM parameters	Eq.	a	b	c	d	e	f	g	h	i	Ce
β_{C-111} , $k=C-111$	(5)	-1.374	2.706×10^{-7}	1.469×10^{-6}	-5.233×10^{-13}	1.113×10^{-7}	-1.065×10^{-6}	3.778×10^{-13}	6.214×10^{-20}	-4.470×10^{-20}	0.899
β_{L-31W} , $k=L-31W$	(5)	-0.1678	-2.286×10^{-5}	1.196×10^{-7}	-2.128×10^{-14}	9.164×10^{-6}	-7.112×10^{-7}	1.264×10^{-13}	0	0	0.961
μ	(6)	-2.997×10^7	-2.922	4838.810	5.307×10^{16}	4.878 × 10 ⁹	-4.091×10^{12}	-	-	-	0.843

4. Conclusions

Detailed hydrological multivariate time series, obtained at an agricultural area located at the boundary of Everglades National Park in south Florida, were studied and modeled using dynamic factor analysis (DFA). The analysis was successfully applied to understand the hydrological trends in this area, which is affected by an ongoing large scale environmental restoration project. The technique proved to be a powerful tool for the study of interactions among 21 long-term, non-stationary hydrological time series. Elevations in canals surrounding the area were found to be the main factors responsible for groundwater profiles, while rainfall events were only responsible for instantaneous or localized groundwater responses that in some cases can be directly associated with the flooding risk. This substantiates the impact of the regional water management system on the local hydrological conditions of the area and corroborates previous results by other authors using different methods (Genereux and Slater, 1999). This close coupling between surface and groundwaters could hinder on-going environmental restoration efforts in the area by bypassing the function of wetlands and other surface features. The Dynamic Factor Model (DFM) resulting from the DFA was validated with acceptable results (coefficient of efficiency 0.69–0.99). The regression parameters of the DFM obtained for each observation point were interpolated by fitting to empirical functions in UTM (X,Y) coordinates in an effort to extend the model across the spatial domain. This second empirical model has an added benefit that the total number of parameters required is greatly reduced. The comparison of model predictions with observed data yielded also satisfactory results (coefficient of efficiency 0.66–0.99) with an expected prediction error of 0.07 ± 0.03 m across the domain. This empirical model is deemed useful for area management in conditions similar to those present in the area during the experimental period. Using this tool on different canal management alternatives could be explored and optimized in terms of flooding risk and the on-going environmental restoration goals for the Everglades National Park.

Acknowledgements

The authors wish to recognize the rest of the team that collaborated in the field data collection and instrumentation effort: Tina T. Dispenza, Martin Morawietz, Harry Trafford and Michael Gutiérrez. 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-10389.

Appendix A. Coefficient of efficiency

The coefficient of efficiency, C_e (Nash and Sutcliffe, 1970), also known as the *Nash-Sutcliffe coefficient*, was calculated from the normalized mean squared error ($nMSE$) (Gershenfeld and Weigend, 1993; Berthouex and Brown, 2002) as follows

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

where $s(t_i)^*$ and $s(t_i)$ are the observed and the predicted values, respectively, of the surface or groundwater levels at time t_i ; l_s is the length of the observed data set; and $(\sigma^*)^2$ is the variance of the observed data.

References

- Akaike, H., 1974. A new look at the Statistical Model Identification. IEEE Trans. Automat. Control 19, 716–723.
- Berthouex, P.M., Brown, L.C., 2002. Statistics for Environmental Engineers. Lewis Publishers, Boca Raton (pp. 345–353).
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from incomplete data via the EM Algorithm. J. R. Stat. Soc., Ser. 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 1991.
- 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. 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.
- Gershenfeld, N.A., Weigend, A.S., 1993. The future of time series: learning and understanding. In: Gershenfeld, N.A., Weigend, A.S. (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past SFI Studies in the Sciences of Complexity, Proc Vol XV. Addison-Wesley, Reading, MA.
- 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. North-Holland, Amsterdam, pp. 365–382.
- Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
- Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer, Berlin.
- McDonald, M.G., Harbaugh, A.W., 1988. A modular three-dimensional finite-difference ground-water flow model U.S. Geological Survey Techniques of Water-Resources Investigations, book 6 1988 chap. A1, 586 pp..
- Márkus, L., Berke, O., Kovács, J., Urfer, W., 1999. Spatial prediction of the intensity of latent effects governing hydrogeological phenomena. Environmetrics 10, 633–654.
- 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 (p. 183-199).
- Molenaar, P.C.M., 1993. Dynamic factor analysis of psychophysiological signals. In: Jennings, J.R., Ackels, P., Coles, M.G.H. (Eds.), Advances in Psychophysiology. Jessica Kingsley Publishers, London.
- Molenaar, P.C.M., de Gooijer, J.G., 1988. On the identification of the latent covariance structure in dynamic nonstationary factor models. In: Jansen, M.G.H., van Schuur, W.H. (Eds.), The

- Many Faces of Multivariate Analysis, vol. 1. Society for Multivariate Analysis in the Behavioral Sciences, Groningen, pp. 196–209.
- Molenaar, P.C.M., de Gooijer, J.G., Schmitz, B., 1992. Dynamic factor analysis of non-stationary multivariate time series. *Psychometrika* 57, 333–349.
- Molenaar, P.C.M., Rovine, M.J., Corneal, S.E., 1999. Dynamic factor analysis of emotional dispositions of adolescent stepsons towards their stepfathers. In: Silbereisen, R.K. (Ed.), *Growing Up in Times of Social Change*. De Gruyter, Berlin, pp. 287–318.
- 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, St Joseph, MI (ASAE Paper No. 032048).
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part 1-A discussion of principles. *J. Hydrol.* 10, 282–290.
- Schumann, A., Muñoz-Carpena, R., 2003. A simple, self-contained canal stage recorder. *Appl. Eng. Agric.* 18 (6), 20–25.
- SFWMD, 2004. Environmental Database (DBHYDRO). Available at www.sfwmd.gov/org/ema/dbhydro (accessed 15 Jun. 2004). South Florida Water Management District, West Palm Beach, FL.
- Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Anal.* 3, 253–264.
- USACE, 2002. Anon., 2002. Interim Operational Plan (IOP) for Protection of the Cape Sable Seaside Sparrow-Final Environmental Impact Statement. Anon., 2002. US Army Corps of Engineers. Jacksonville, Jacksonville District.
- Wilson, B.N., 2001. Evaluation of hydrologic models using statistical methods. ASAE Paper 01-012207. Presented at 2001 ASAE Annual International Meeting, St. Joseph, MI.
- Wu, L.S.-Y., Pai, J.S., Hosking, J.R.M., 1996. An algorithm for estimating parameters of state-space models. *Statist. Probab. Lett.* 28, 99–106.
- Zoubir, A.M., Boashash, B., 1998. The bootstrap and its application in signal processing. *IEEE Signal Process. Mag.* 15 (1), 55–76.
- 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.