1-11-2012

Operational Prediction of Groundwater-phosphorous Interaction Over Surficial Aquifers of South Florida

Yirgalem A. Chebud
Florida International University, ycheb002@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

OPERATIONAL PREDICTION OF GROUNDWATER-PHOSPHOROUS

INTERACTION OVER SURFICIAL AQUIFERS OF SOUTH FLORIDA

A dissertation submitted in partial fulfillment of

the requirement for the degree of

DOCTOR OF PHILOSOPHY

in

GEOSCIENCES

by

Yirgalem Assegid Chebud

2012

To: Dean Kenneth G. Furton
    College of Arts and Sciences

This dissertation, written by Yirgalem Assegid Chebud, and entitled Operational Prediction of Groundwater-phosphorus Interaction over Surficial Aquifers of South Florida, having been approved in respect to style and intellectual content is referred to you for judgment

We have read this dissertation and recommend that it be approved.

_____
Dean Whitman

_____
Ping Zhu

_____
Fernando Miralles

_____
Georgio Tachiev

_____
Assefa M. Melesse, Major Professor

Date of defense: January 11, 2012

The dissertation of Yirgalem Assegid Chebud is approved.

_____
Dean Kenneth G. Furton
College of Arts and Sciences

_____
Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2012

DEDICATION

This piece of work is dedicated to my late mother, Fentaye Birhanu, whose life was a blessing that will resonate in me for ever. She has protected and realized the vision that was shared with my late father, Assegid Chebud.

ACKNOWLEDGMENT

ABSTRACT OF THE DISSERTATION

OPERATIONAL PREDICTION OF GROUNDWATER-PHOSPHOROUS

INTERACTION OVER SURFICIAL AQUIFERS OF SOUTH FLORIDA

by

Yirgalem Assegid Chebud

Florida International University, 2012

Miami, Florida

Professor Assefa M. Melesse, Major Professor

South Florida has transformed from a natural to a managed ecosystem upon channelization of Kissimmee River and the wetlands in the 1960's. The drainage has resulted in fast transport of water and nutrient, and subsequently eutrophication of the downstream water bodies. The intervention required: intensive management of the shallow groundwater to balance ecological water requirement; and nutrient removal, namely phosphorus, to minimize eutrophication.

The study was set to examine and develop an operational prediction method for groundwater-phosphorus interactions to support the wetlands management. Accordingly, a point scale and a spatio-temporal groundwater level was simulated using sequence based Markovian stochastic analysis and dynamic factor analysis methods respectively. A root mean square error of 0.12m and 0.15m was observed for a point and spatio-temporal groundwater prediction.

Soluble and sequestered phosphorus were also simulated at 13% error using a watershed based model called ArcWAM. A spatial analysis on simulated soluble phosphorus and groundwater level indicated similarity of patterns (spatial correlation)

99% of the time. A geographically weighted multivariate analysis of soluble phosphorus using predictors of groundwater level, total phosphorus of surficial water, and distance from Kissimmee River showed a goodness of fit (R^2) of 0.2 – 0.7. Amongst the factors, the groundwater explained 70% of the soluble phosphorus variability.

In summary, an increase in soluble phosphorus was observed with groundwater rise and a decrease during groundwater recession. A reversed relationship was identified for the total phosphorus. Presumably, organic matter in the root zone has contributed to increased soluble phosphorus with the rise in groundwater. On the other hand, solubility of calcium carbonate from the karst aquifers seems to fix and precipitate phosphorus during recession of groundwater. The least sequestration of phosphorus, observed in oversaturated wetlands also suggested that nutrient removal on karst hydrogeology could be risky unless a check is made using vegetation strip to enhance phosphorus uptake.

The study concluded that phosphorus could be operationally predicted associated with forecasting of groundwater fluctuation. Further research is recommended to explore factors that could be derived either empirically or from satellite data for prediction of soluble phosphorus at minimum cost.

Key words: groundwater, phosphorus, spatiotemporal, stochastic, prediction

TABLE OF CONTENTS

CHAPTER   PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AIC	Akaike Information Criterion
ATCOR	Atmospheric and Topographic Correction
BIC	Baye's Information Criterion
BLASROUTE	Basin Land Area to Stream Routing
BUC	Basin Unique Cell
DFA	Dynamic Factor Analysis
EAAMOD	Everglades Agricultural Area Model
ENR	Everglades Nutrient Removal
EPA	Environmental Protection Agency
FDEP	Florida Department of Environmental Protection
FHANTM	Field Hydrology and Nutrient Movement
GLEAMS	Groundwater Loading Effects of Agricultural Management Systems
GLDAS	Global Land Data Assimilation System
IDW	Inverse Distance Weighted
LO	Lake Okeechobee
LOAA	Lake Okeechobee Agricultural Area
MCL	Maximum Contaminant Level
MODIS	Moderate Resolution Imaging Spectroradiometer
MLR	Maximum Likelihood Ratio
N	Nitrogen
NLDAS	National Land Data Assimilation System
ANN	Artificial Neural Network

P	Phosphorus
PCA	Principal Component Analysis
RMSE	Root Mean Square Error
SAR	Satellite Aperture Radar
STA	Storm-water Treatment Areas
SFWMD	South Florida Water Management District
SVD	Singular Value Decomposition
SWET	Soil Water Engineering Technology
TMDL	Total Maximum Daily Load
USGS-EROS	United States Geological Survey - Earth Resources Observation and Science
WAM	Watershed Assessment Model
WCA	Water Conservation Area

CHAPTER I

1.1. Introduction

The functioning of wetlands to attenuate nutrient and reduce influx to surface water bodies is a widely established fact. Such functioning of wetlands has been exploited by utilizing artificial wetland systems as storm water treatment areas. However, wetlands can be effective, if and only if the underlying hydrogeology has the capability to attenuate nutrient leaching that could otherwise percolate to groundwater. Also, the efficiency of a wetland is dependent on the permeability of the underlying aquifer. The lesser the permeability, the higher will be residence time of water and nutrient and hence improved attenuation of nutrients. Also, aquifers with abundant Fe and Al ions could attenuate higher amount of nutrients from overlying wetlands.

In karst aquifers such as in south Florida (SF), the effectiveness of wetlands for nutrient removal is questionable for the fact that the aquifer is highly transmissive draining nutrients to water bodies. Even though the dissolution of calcium carbonate can fix nutrients such as phosphorus, the precipitated nutrients are subject to transportation because of the porous nature of the aquifer. Currently, nutrient removal is being implemented over the karst aquifers of SF regardless of the contextual effectiveness of the wetlands. The practice will change wetlands from oligotrophic to hypertrophic condition unless research supports the decision making.

Since the 1960's, SF wetlands have undergone a transformation from a natural to a managed ecosystem because of the channelization of the Kissimmee River for flood control (SFWMD, 2010). The change has had consequences on the vegetation pattern, surface-groundwater interactions and nutrient dynamics (Koebel et al., 1999). First,

preventing the natural flooding had drained excessive load of nutrient to the downstream causing eutrophication of water bodies namely to Lake Okeechobee (Chimney et al., 1999). Second, a relocation of the drained storm water into water conservation areas (WCAs) for nutrient removal and flood protection has turned oligotrophic wetland systems to hypertrophic ecosystems and has caused unintended vegetation succession (Kadlec, 2006). The vegetation succession namely by cattail was in part the consequence of nutrient imbalances. Third, canal dredging has caused loss of natural flow directionality (Variano et al., 2008) and hence has influenced the surface-groundwater interaction. Finally, expansion of agriculture specifically dairy cattle farming, with over 587 farms between 1995 to 2000 (Zhang et al., 2002), and sod farming in the upstream of Lake Okeechobee (Chebud et al., 2011) has caused excessive nutrient loads exacerbating hypertrophic condition of the lake (O'Connar and Sarker, 2001).

To curb ecological degradation, the South Florida Water Management District (SFWMD) promoted Best Management Practices (BMP's) in the upstream agricultural areas, and also developed storm water treatment areas (STAs) at the micro scale, and WCAs (at macro scale) (Hiscock et al., 2003). Chimney et al. (1999) reported effectiveness of STAs reducing nutrient load by up to 80%, though ecological damages on water conservation areas were impacted by vegetation succession (Kadlec, 2006). In contrast, Moreno-Mateos et al. (2009) reported the reduction in another wetland was 99% on Nitrate (NO_3-N) with no reduction on phosphorus.

The intensive groundwater table management attenuated with recurring drought is reported to increase groundwater table fluctuation (Chebud and Melesse, 2011) which in turn triggers redox process of phosphorus mineralization and ultimately their transport

(Moreno-Mateos et al., 2009). Soluble phosphorus was observed reaching highest level of load with the occurrence of rainfall preceded by extended dry season (Young et al., 2008).

Phosphorus leaching and its fast drainage was known to be facilitated by the shallow water table depth, the coarse nature of SF soils and the high transmissivity of the Karst aquifer that reduces time for adsorption of phosphate (PO_3^-) ion by Fe and Al (O'connor et al., 2002; Salzar et al., 2011). The assumption that the calcareous surficial aquifer would contribute to the sorption of the soluble and labile phosphate ions (PO_3^-) and precipitate it to a stable calcium phosphate was probably outweighed by the groundwater fluctuation enhanced redox processes of mineralization and fast drainage processes. In summary, the managed SF ecosystem had influence on the groundwater fluctuation which enhanced redox processes of mineralization that could easily be laden by intense rainfall events and contributes much more than its natural state of basin wide flooding.

Evidently, a key contribution of research related to effectiveness of SF wetlands functioning on nutrient removal is to capture spatio-temporal processes of nutrient attenuation, and loading by subsurface flow. The study proposed a broader research framework of developing cheaper methods of groundwater-phosphorus interaction monitoring at operational level. A specific goal was set to synthesize groundwater table fluctuation and use it as a predictor of soluble phosphorus. The study supports existing groundwater and nutrient monitoring activities by SFWMD and Everglades Foundation which is reported to be expensive and logistically difficult. The study has particularly benefited the Everglades Foundation that sought quantification of phosphorus load,

before and after the Kissimmee River restoration program (before 1999 and after 2009), and also identification of hotspots of phosphorus sources.

1.2. Hypothesis

It was hypothesized that estimate of soluble phosphorus load using groundwater fluctuation as a predictor and the measured soluble phosphorus load should show similarity at 90% of the time. In other words, there should not be significant difference between measured and predicted phosphorus concentration more than 10 percent of the time (p=0.1).

Retrieval of phosphorus using groundwater level as a predictor is relevant in a shallow groundwater system as in SF where subsurface flow dictates nutrient transport. Also, understanding the relationship between groundwater fluctuation and phosphorus load both at point and regional scales supports decision making on groundwater table management, one of the main duties of SFWMD.

1.3. Objectives of the study

In this study, specific objectives were set to:

- Develop a model that simulates a point scale groundwater level in the major surficial hydrogeological regions of south Florida which would be an input for operational prediction of soluble phosphorus.

- Develop a spatio-temporal model that simulates groundwater fluctuation which would be an input for spatial analysis of groundwater-phosphorus interaction.

- Conduct a case study of watershed based nutrient transport assessment that estimates soluble and sequestered P which would be an input for spatial analysis of groundwater-phosphorus interaction..

- Develop a remote sensing approach for phosphorus level estimation on surficial water bodies that support the prediction of soluble phosphorus load from sources, and

- Conduct a multivariate analysis on the predictability of soluble phosphorus

1.4. Methodological framework and structure of the documentation

A Point scale groundwater fluctuation simulation (objective 1) was met using sequence based Markovian stochastic method, a probabilistic approach discussed in Chapter 3. The simulation is an input for a regional scale spatiotemporal groundwater forecasting. The spatiotemporal groundwater fluctuation simulation (objective 2) discussed in Chapter 4, involved identification of explanatory and latent factors for the synthesis of the groundwater level. The approach has a comparative advantage where watershed based models are not applicable. The spatiotemporal groundwater simulation output was also used for spatial pattern analysis and as a predictor to soluble phosphorus load (Chapter 7).

A gridded data of soluble phosphorus was vital to validate against the newly predictive model introduced in this study. The data was derived from a watershed based physical model called ArcWAM, calibrated for SF context and endorsed by SFWMD. Kissimmee basin was identified for a case study and objective 3 (Chapter 5) was met simulating soluble and sequestrated phosphorus in the wetlands.

Another input presumed to serve groundwater-phosphorus interaction was the phosphorus level of surficial water bodies. The notion is that the Kissimmee River phosphorus level is a key feedback from the groundwater-phosphorus interaction, and

hence an input for the spatiotemporal soluble phosphorus simulation. Given the spread of surficial water bodies at regional scale, it necessitated application of remote sensing approach for cost effectiveness. So, phosphorus level of surface water bodies (objective 4) discussed in Chapter 6 was met by analyzing Landsat Thematic Mapper satellite imagery using Artificial Neural Network (ANN).

Finally, a spatial pattern analysis of groundwater level and soluble phosphorus load was conducted using chi-square analysis. Also, geographically weighted multivariate analysis was tested to identify key inputs for groundwater-phosphorus relationship development. The significance of the method is also tested and deduced in Chapter 7.

CHAPTER II

2. Literature Review

2.1. Phosphorus cycle and the effect of groundwater fluctuation in wetland ecosystems

Raisin et al. (1999) reported that groundwater-fed wetlands are nutrient sources to the surficial water bodies during base flow as well as storm flow conditions. The result was supported by Corbett et al. (2002), reporting groundwater loading of phosphorus (P), "2 to 7 times the natural level, although 50% reduction was observed from the background concentration". The increased release of phosphorus from the water column and soil matrices upon wetting and drying (fluctuation of groundwater) was reported by Moreno-Mateous et al. (2009). A report shows that the median P concentration in most aquifers was above the thresholds of the 'ecological requirement' (Kadlec, 2006).

The phosphate-groundwater dynamics in the wetland system of karst hydrogeology site is dominated by the Ca-P dissolution and precipitation processes. The process could be conceptualized by a framework of adsorption vs. desorption, precipitation vs. dissolution, mineralization vs. immobilization, plant uptake and leaching (Kadlec, 2006; Nahra, 2006). Part of the framework for the wetland system includes influx of nutrient from catchments vs. outflow (discharge) to water bodies both in surficial and groundwater systems (Kadlec, 2006).

Adsorption vs. desorption is the fastest process that involves the interaction of the phosphate ion with the soil particles governed by an equilibrium balance of the soils charged surfaces with the solution ionic content (Nahra, 2006). The proportion of the ionic content in the soil solution against the adsorbed amount is dependent on the type of soil with the general characteristic of finer soils having higher adsorption capacity

(Nahra, 2006). The precipitation vs. dissolution process is known to be the slowest process that involves a reaction of phosphate ion with calcium in calcareous soils and with Fe/Al in acidic soils forming a stable calcium phosphate or Fe/Al phosphate, respectively (Nahra, 2006). The precipitation process in karst areas is dominated by formation of calcium phosphate. The process takes place until saturation is reached and after which excessive supply of phosphate or calcium has little effect on mobility of the phosphate ion. The mineralization process is a conversion of organic phosphorus to ionic phosphate making it available to mobility and plant uptake. The process is favored with the subsidence of shallow water table that increases microbial activity. Immobilization is a reciprocal process of mineralization involving consumption of phosphate ions by microorganism and making immobile organic phosphorus. The framework indicated on Figure 2.1 is a simplified form of the complex processes of phosphorus adapted after Kadlec (2006), Charles et al., (2005) and Nahara (2006).
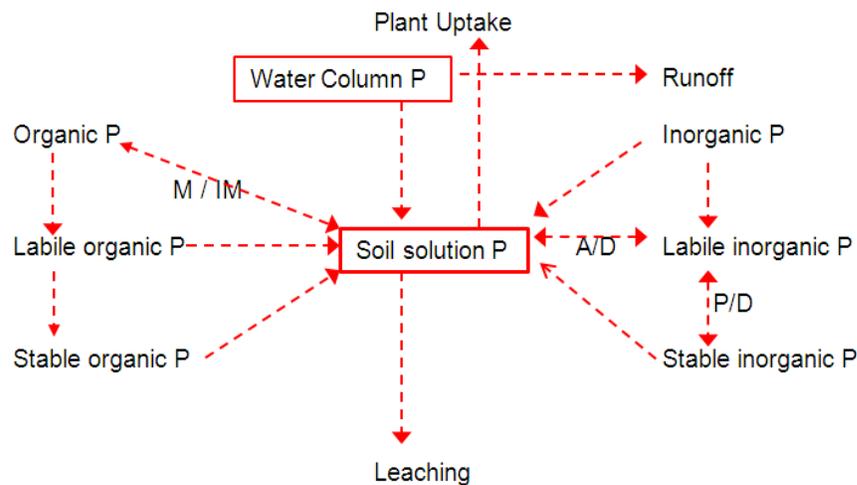
Figure 1.1 Phosphorus dynamics framework

A/D – Adsorption/Desorption; P/D – Precipitation/Dissolution; M/IM – Mineralization / Immobilization

In general, the process of conversion and mobility of phosphorus is dependent on the soil type, microbial activity, sources of phosphorus, vegetation, temperature and hydrological processes (drying and wetting) (Raisin et al., 1999). The higher the vegetation diversity, the higher would be the uptake of phosphorus. Similarly, the drier the subsurface, the greater will be the aerobic microbial activity that facilitates mineralization. If intense rainfall is preceded by drier period that favor mineralization, the mineralized phosphate certainly impacts water bodies because of its high mobility (Moreno_Mateos et al., 2009). The same report indicated that phosphorus load was enhanced by the drying and wetting cycles during a summer season. In Fe/Al rich soils, the mineralized phosphate would be attenuated by an adsorption process. However, in the case of calcareous soil, drier condition rather allow availability of freely mobile orthophosphate ions; the maximum being in times of extended summer that allows excessive microbial activity (Young et al., 2006). The same report argued that increased microbial activity and maximum phosphorus loads are associated with extended summer (climate change).

The pragmatic way forward is to monitor phosphorus load over shallow water bodies and understand complex processes between groundwater and phosphorus dynamics in a wetland ecosystem (Young et al., 2006). A remote sensing approach is known to be the cheapest approach with no compromise on the quality of the monitoring.

2.2. Phosphorus monitoring over water bodies: A remote sensing approach

Applicability of remote sensing to monitor vegetation, precipitation and water quality has been reported by Griffth (2002) and Dekker et al. (1996). The relevance in a large wetland system is evident for its cost and efficiency to capture basin-wide data. A

remote sensing is a necessity to capture water quality parameters in the surficial water bodies and relate the phosphorus level against the background groundwater concentration. Its importance in a large wetland system in terms of reducing cost of monitoring was reported by Chebud et al. (2011) given a recent transition from airborne to optical sensors capability to capture water quality parameters. Landsat Thematic Mapper (TM) capability to retrieve water quality parameters of chlorophyll-a, turbidity and phosphorus was reported by Zhang (2002).

The different methods used to predict phosphorus include multiple regression analyses, neural network methods (Zhang, 2002), and empirical methods (Ritchie et al., 2003) are commonly used to retrieve water quality parameters. The neural network approaches is considered inclusive for its capability to use all the bands as information containers (regardless of redundancy) (Zhang, 2002). Importance of neural network methods is evident especially on shallow water bodies where the back ground effect from sediment needs to be trained.

2.3. Groundwater and solute transport: A deterministic Approach

Groundwater and solute transport are coupled by mass balances of the flow and hydrogeologic property of the aquifer namely the hydraulic conductivity and diffusivity parameters (Equations 2.1 and 2.2). Two of the approaches for quantification of groundwater and solute transport, commonly called deterministic methods, use Darcy's Law and Fick's Law respectively.

$$\nabla(K\nabla h) = I + S_y \frac{\partial h}{\partial t} \tag{2.1}$$

Where K is hydraulic conductivity of the medium; h is head of water; I, recharge,

S_y is the specific yield, t is time and ∇ the gradient operator;

$$-v\frac{\partial C}{\partial x}+\nabla(D\nabla C)=R\frac{\partial C}{\partial t}$$

(2.2)

Where v is Kinematic viscosity; D is dispersion coefficient; R is the retardation factor given by $1+\rho_b K_d/\theta$; ρ_b is bulk density; K_d is adsorption isotherm and; θ is porosity.

The retardation factor (R) is the most important factor that captures chemical reaction processes of phosphors explained in the percentage of attenuation (degradation) in the transport. Applying a finite difference method and boundary conditions to the partial differential equation and populating parameters, a spatiotemporal dynamics of head of groundwater and phosphorous transport could be retrieved. The relationship is a basis for watershed based hydrological models that are used to model surface groundwater interaction and nutrient transport deterministically.

In the SF context, first, the approach would suffer at least from the absence of clearly delineated watershed boundary because of the flat topography with no flow dividing lines, except the Kissimmee River basin, which constitutes less than 25% of the land mass of SF. Second, the watershed-based models are less practical for operational prediction in karst hydrogeology because of the dynamic nature of the hydrogeological factors K in equation 2.1; and also D and R in equation 2.2, which raise parameterization uncertainty whenever changes occur (Denic`-Jukic` et al., 2003). Third, in karst hydrogeologic areas, quantifying surface–groundwater interaction using distributed hydrological models is a difficult exercise since the watershed based hydrological models are computationally intensive requiring frequent recalibration whenever new data inputs are entered..

2.4. Data driven stochastic simulation of groundwater fluctuation

A storage change on groundwater transport models is conceptualized as the net balance of inflow and outflow from aquifers (Anderson and Woessner, 2002). Inflow sources are rainfall, soil moisture in the profile, surface water bodies, and regional groundwater flow. The outflow is captured in the form of evapotranspiration, abstraction by different users and leak from aquifers (Anderson and Woessner, 2002). Rainfall being one of the drivers of the shallow groundwater budget, its random nature dictates the nature of water table fluctuation stochastically (Chebud and Melesse, 2011). Laio et al. (2007) has shown that infiltration and evapotranspiration have a random behavior because of the rainfall stochasticity.

The random nature of the water table fluctuation is actually not only from the rainfall and stream flow but also because of the spatial variability of soils (subsurface aquifers) and their respective hydraulic variables. Texture, porosity, and conductivity of medium dictate the spatial stochastic nature (Iturbe et al., 2008). The implication is that nutrient transport would also be dictated by the stochastic behavior because of the dependence of the diffusivity factor on hydrogeological properties.

In contexts of large wetland systems, karst hydrogeology and complex interaction with surface and oceanic water, effectiveness of data driven and dynamic stochastic models were reported by Zhang et al. (1999; 2002). Markovian probabilistic simulation observations have proven predictions up to a month time span (Chebud and Melesse, 2011). Unlike deterministic-based stochastic method that suffers from the error propagation, data driven stochastic simulation were observed to minimize error propagation. The data driven approach is effective for forecasting at monitoring sites, if

the relationship is developed from long-term observed data and the model is updated whenever new observations are made. In the context of SF, significant hydrological monitoring data had helped to develop a simplistic data-based model on groundwater fluctuation and stream flow (Zhang et al., 1999).

2.5. Perspectives of spatiotemporal groundwater-phosphorus interaction simulation

Given the shallow nature of groundwater and the flat topography of SF, the water table can be retrieved on the basis of signals from rainfall, soil moisture, and probabilistically from past data. Beldring et al. (1999) and Jacobs et al. (2004) have shown the stability of the soil moisture variance with time and predictable relationship with water table fluctuation. A study by Itrube et al. (2006) has shown that soil moisture has a power probabilistic distribution regardless of the land use and soil distribution. Integration of a spatial structure on point scale processes as well as identification of explanatory and latent factors is a key aspect of spatiotemporal modeling.

However, since capturing all the input variables would not be attainable, the latent factor will explain the aggregate of such unidentified factors whereas the exploratory factors would be selected using the minimum Akaike Information Criterion (AIC). Both the latent and explanatory variables could be determined using dynamic factor analysis that helps to identify the input variables as well as their loadings at different spatial and temporal scale. The loadings derived at different points could contribute to the regional scale forecasting as the weight of variables change over space and time. Finally, the spatiotemporal synthesis of groundwater level would support identification of a spatial correlation between groundwater and phosphorus concentration at grid points in space and ultimately simulation of phosphorus load.

REFERENCES

Ackerman J R, (2011). Quantifying Nutrient removal from groundwater seepage out of a constructed wetlands receiving wastewater effluent, MSc Thesis, Department of Geography-Geology, Illionis State University, 2011

Balla D, Gabriel O and Kalettka T (2006). Estimating phosphous fluxes in wetland groundwater-surface water environments with respect to eutrophication. HydroEco-Karlovy Vary, Czech Republic; Kovar-Hrkal-Bruthans eds.

Chebud Y A, Naja M, Rivero R, (2010), Phosphorus run-off assessment in a watershed, *Journal of Environmental Monitoring*, 13: 66-73.

Chebud Y A., Melesse A. (2011). Operational Prediction of Groundwater Fluctuation in South Florida using Sequence based Markovian Stochastic Model, *Water Resource Management*, Volume 25, Number 9, 2279-2294, DOI: 10.1007/s11269-011-9808-z

Corbet D. R., Dillion K., Burnett W., Schaefer G. (2002). The spatial variability of nitrogen and phosphorus concentration in a sand aquifer influenced by onsite sewage treatment and disposal systems: a case study on St. George Island, Florida, Environmental Pollution 117(2002)337-345.

Chimney M J. Nungesser M, Newman J, Pietro K, Germain G, Lynch T, Goforth G and Moustafa M Z, (1999). CHAPTER 6: storm water treatment areas – status of research and monitoring to optimize effectiveness of nutrient removal and annual report on operational compliance: Everglades Consolidated Report, Status of STA's
http://www.garygoforth.net/ch6%202000%20ECR.pdf

Denic-Juki` V, Jukic` D. (2003). Composite transfer functions for Karst aquifers, Journal of Hydrology 274: 80-94

El Kahloun M, Gerard M, Meire P. (2005). Phosphorus and nitrogen cycling in fen vegetation along different trophic conditions in the Biebrza valley, Poland, Ecohydrology and Hydrobiology, 5(1): 67-78

Florida Department of Environmental Protection. 2008 Reuse Inventory. Tallahassee, Florida: Florida Department of Environmental Protection. 2010.

Iturbe I, A Porporato, Laio F, Ridofi R, (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress, I scope and general outline, Advances in Water Resources Research 24: 695-705

Jukic` D, Denic-Juki` V, (2004). A frequency domain approach to groundwater recharge estimation in Karst, Journal of Hydrology 289: 95-110

Jiang Y, Woodbury A, Painter S, (2004). Full-Bayesian Inversion of the Edwaress Aquifer, Ground water, 42(5):724-733

Kadlec R. H. (2006). Free surface wetlands for phosphorus removal: The position of the Everglades Nutrient Removal Project, Ecological Engineering 27:361-379

Laio F, Porporato A, Illescas F, Iturbe I R, Ridofi R. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress, IV, Discussion of real cases, Advances in water resources research 24: 745-762

Moreno-Mateos D., Perocchi C., Comi'n, (2009). Effect of wetland construction on water quality in a semi-arid catchment degraded by intensive agricultural use, *Ecological Engineering*, 1479: 1-9

O'Connor, G.A., H.A. Elliott, and P. Lu. 2002. Characterizing water treatment residuals phosphorus retention. Soil Crop Sci. Soc. FL Proc. 61:67-73.

Porporato A., Laio F, Ridofi R, Iturbe I R. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress, III Vegetation stress, Advances in Water Resources Research 24: 725-744.

Price, R., Swart, P. and Fourqurean, J. (2006). Coastal groundwater discharge – an additional source of phosphorus for the oligotrophic wetlands of the Everglades. Hydrobiologia,569:23–36.

Raisin G., Bartley J., Croome R., (1999). Groundwater influence on the water balance and nutrient budget of a small natural wetland in Northeastern Victoria, Australia Ecological Engineering 12: 133–147

Ritchie, J. C., Zimba, P. V., and Everitt, J. H. (2003). Remote sensing techniques to assess water quality. *Photogrammetric Engineering and Remote Sensing, 69*(6), 695-704.

Salazar Osvaldo, Wesstrom Ingrid, Joel A. (2011). Identification of hydrological factors Controlling phosphorus concentration in drainage water in sandy soils, J. Soil Sci. Plant Nutr 11(2): 31-46

Seeling B, DeKeyser S. (2006) Water Quality and Wetland Function in the Northern Prairie Pothole Region, North Dakota State University Extension Service manual, North Dakota, Fargo, August 2006

SFWMD-South Florida Water Management District,(2008). Lake Okeechobee Watershed Construction Project – Phase II Technical Plan, South Florida Water Management District.
http://www.sfwmd.gov/portal/page/portal/xrepository/sfwmd_repository_pdf/ne_lakeo_watershed_phase2_summary.pdf

Shapiro C, Kovar J, (2005). Agricultural Phosphorus Management and Water Quality protection in the Midwest, The Heartland Regional Water Coordination Initiative, United States Environmental Protection Agency (USEPA)

Spruill, T.B., Bratton, J.F., 2007, Estimation of groundwater and nutrient fluxes to the Neuse River Estuary, North Carolina: Estuaries and Coasts (2008) 31:501–520.

Surridge B W J, Heathwaite A L, Baird A J, (2006). Groundwater-surface water interaction and Phosphorus biogeochemistry in river floodplains: Geophysical Research Abstracts Vol 8, 05972, 2006

Tiemeyer B, Frings J, Kahle P, Lennatz B, (2007). Spatial variability of soil properties and shallow groundwater solute concentrations in degraded peatland. Geophysical Research Abstracts, Vol 9, 03236, 2007

Variano E.A, D.T. Ho, V.C. Engel, P.J. Shmieder, M.C.Reid, (2009). Flow and Mixing Dynamics in a Patterned Wetland; Kilomenter-Scale Tracer Releases in the Everglades WATER RESOURCES RESEARCH, VOL. 45, W08422, doi:10.1029/2008WR007216, 2009

Wortman, Helmers M, Mallarino, Barden C, Devlin D, Pierzynski G, Ray m, Lory J, Holz J, Young E O., Briggs D R., (2008). Phosphorus Concentrations in Soil and Subsuface Water: A field study among cropland and riparian buffers. , Journal of Environmental Quality, 37:69-78

Zhang, Y., Pulliainen, J., Koponen, S., and Hallikainen, M. (2002). Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment, 81*(2-3), 327-336.

Zhang J., Gornak S.I (1999). Evaluation of field-scale water quality models for the lake Okeechobee regulatory program, Applied Engineering in Agriculture, 15:441-447.

CHAPTER III

3. Operational Prediction of Groundwater Using Markovian Stochastic model

3.1. Abstract

The ecosystem of south Florida is characterized by a large wetland system, karst surficial hydrogeology, and extended coastal boundary. The ecosystem is poised under risks of: ecological failure as a result of increased fragmentation by urbanization; groundwater flow disruption because of sinkhole formation; and intrusion of oceanic water with decreasing water table head because of drought and over pumping. Synthesizing a spatiotemporal state of the groundwater hydrology and developing a forecasting model was found important to support the intensive management and monitoring in place. An objective was set to develop a stochastic sequence model capable of forecasting groundwater levels on a monthly span at a daily time scale. The groundwater level simulation was conceptualized as a sequence of daily fluctuating states of magnitudes and patterns that has a defined probability of occurrence. The model setup involved representation of daily fluctuation magnitudes in 10 states and pattern changes in 3 states. The sequential occurrence of states of magnitudes and patterns at each time step was used for estimation of the transitional probabilities and employed in a hidden Markov model frame work for ensemble generation and estimation of posterior probabilities. A realization was chosen using the highest maximum likelihood ratio of 90% and smallest root mean square error of $0.05 - 0.12$ m against the historical data. A monthly forecasting

at daily time step was done dynamically incorporating observed data at each time step and revising prior and posterior probability estimation in the hidden Markov model formulation. A case study was conducted at three well sites, which are situated at three different hydrogeologic settings. The model not only reproduced annual groundwater fluctuation patterns but also forecasted monthly fluctuations at maximum likelihood ratio above 90% and root mean square error below 0.15m. A further study was recommended first to analyze break point parametric estimation for seasonal analysis, and secondly to integrate the approach in other hydrological models for the purpose of synthetic groundwater fluctuation generation.

Key words: Groundwater, Stochastic Sequence , Hidden Markov Model, South Florida

3.2. Introduction

The ecosystem of south Florida is dictated by a large wetland system, karst surficial hydrogeology, and extended coastal boundary with the Atlantic Ocean (Gunderson, 2001). The wetland serves to temporarily retain excess surface water during the wet season and releases it to streams during the dry season (Paynter et al., 2011). However, because of the anthropogenic interference through canal dredging, channelization, and drainage it has resulted in fast movement of water to the downstream side entailing a risk of disconnectivity of groundwater dependent wetlands during extended dry seasons (Ali, 2009). On the other hand, the shallow nature of the groundwater and the flat topography makes the urban areas susceptible to the groundwater table rise and flooding problems. Regardless of the water level management operations, the risk of groundwater rise in urban areas and flooding is prevalent at times of high rainfall events, mainly in hurricane seasons.

The karst hydrogeology dictates the direction of surface-groundwater interaction of the wetlands and canals with occasional disruption of groundwater flow because of frequent sinkhole formation (Genereux et al., 1999). The same report indicated re-connectivity problems of the wetlands which could be exacerbated by sinkhole formation posing both spatial and temporal uncertainty of the groundwater supply to the ecosystem. Practices such as surface water abstraction for agricultural practices (Ahmad et al., 2010) and domestic use from the surficial aquifer potentially exacerbate the already fragmented wetlands through anisotropy and groundwater flow disruption.

The oceanic coastal boundary poses saltwater intrusion risk whenever the groundwater head drops in times of drought or because of over pumping. The thin vadoze zone especially along the urbanized coastal areas makes the groundwater susceptible to salt water intrusion. The coastal areas naturally maintain their equilibrium through flooding in the wet season that counter balances saltwater intrusion. Nevertheless, with the expansion of urbanization, the water is partly drained, and in part siphoned for industrial or agricultural use, decreasing the water table head further and consequently allowing intrusion of ocean water.

The dryness and wetness of the subsurface because of groundwater fluctuation also dictates microbial activity and hence enhances redox processes of nutrient mineralization and subsequently its mobility. Apparently, a prediction model on groundwater fluctuation would help to predict phosphorus load.

Currently, the natural equilibrium of SF ecosystem is poised under intensified artificial management. South Florida Water Management (SFWMD) employs extensive monitoring with over 1834 stations as part of the management effort to cope against the

spatial and temporal uncertainty of the groundwater level dynamics (Florida Council of 100, 2003). Evidently, availability of data is an opportunity to support the management style through synthesis of the spatiotemporal surface-groundwater data and predicting the risks related to ecological failure because of groundwater flow disruption; flooding in urban areas from groundwater surge in extreme rainfall events; intrusion of saltwater with groundwater head drop at times of drought; and nutrient load to surficial water bodies.

In this research, it was intended to develop a point scale stochastic groundwater prediction model that could in the longer term evolve to serve as a component of a regional scale spatiotemporal surface-groundwater forecasting model. The rationale for choosing the stochastic approach was (1) its capability to predict any risk probabilistically, (2) the stochastic approach is preferable to watershed based hydrological models because of the flat topography and extended boundary with the Atlantic Ocean which makes delineation of flow dividing lines difficult for the latter approach. Only the Kissimmee River basin which covers less than 25% of the district could be practically modeled using watershed based approach, and (3) despite the fact that south Florida ecosystem is highly managed, rainfall is one of the driving forces of groundwater level fluctuation of the surficial aquifer system and hence the stochastic approach is acceptable for the stochasticity of the rainfall embedded in the system. The assumption holds true, especially on monitoring stations that are reasonably at far distance from surface water bodies, whose daily variation would not be influenced at short time scale.

A specific objective was set to develop a sequence based Markovian stochastic model that predicts groundwater level at a daily time scale over a span of a month. The

study hypothesized that no significant difference would be observed at P value of 0.01 between a state sequence representation of groundwater fluctuation the measured groundwater level.

Several authors introduced application of Markovian stochastic models for synthetic generation of stream flow, power and wind speed (Aksoy, 2004; Aksoy, 2003; Szilagyi et. al., 2006; Zahabiyoun, 2005; Shamshad et. al., 2005). The analogy of groundwater fluctuation patterns with stream flow pattern would be intuitive in view of the similarities of flow patterns that show increasing, decreasing or constant flow (Aksoy, 2004) versus the rising, falling and constant head of groundwater levels. In other words, groundwater fluctuation could be represented by the states of the daily fluctuation magnitudes and daily change of patterns similar to the stream flow states representation by Aksoy (2004).

The approach is conservative since the size of states and patterns are from the long term daily fluctuation and hence less affected by error propagation unlike other stochastic methods such as the perturbation method. Moreover, the fact that the approach is sequence based and is reliant on historical data makes it unique over other stochastic approaches such as the Monte Carlo method that relates probabilistically only the preceding and the proceeding events.

The model development was conceptualized by designating states of fluctuation magnitudes and patterns with a prior and posterior parametric formulation on the basis of experiences from the stream flow generation studies with a purpose to adapt it to the groundwater simulation. The assumption was that stream flow and the groundwater fluctuation have analogous patterns of daily increase or decrease. Upon model

development, choice for a realization was made using maximum likelihood ratio (MLR) and the root mean square error (RMSE) requirements which were tested to maintain statistical and pattern consistency of the observed and generated data. The analysis draws achievements on monthly forecasting at daily time step and the adaptive behavior of the model. Further analysis on its efficiency as part of a regional scale spatiotemporal model is also recommended. Last, a break point analysis for seasonal simulation and integration of the approach to other hydrological models is suggested.

3.3. Background

One of the factors influencing the water table fluctuation in surficial aquifers is precipitation (Park et al., 2008). Forced by the random nature of the rainfall, the magnitude and pattern of the water table fluctuation assumes a stochastic behavior. As a matter of fact the magnitude of the fluctuation is the size of the change whereas the pattern is characterized by the state of rise (recharging), drop (discharging) or remaining constant. A stochastic simulation captures both daily magnitudes and patterns of groundwater fluctuation by their respective states which represent a range of variation that has a defined probability of occurrence in time series. A sequence (time series) of states of fluctuation magnitudes and patterns are chosen when the joint probability is maximum and also the pattern agrees with observations (Janssen et al., 2006).

3.3.1. Magnitude of states of groundwater fluctuation

The daily groundwater fluctuation varies randomly as indicated by schematic drops and rises (l1, l2, l3 in Fig 3.1). If large datasets are available, Najib et al. (2007) suggested a coarse resolution of fluctuation states defining peak-over-threshold (POT) approach which uses ascending and descending patterns occurring at consecutive time

steps. The POT needs to satisfy the condition that consecutive peaks be independent (Najib et al., 2007).

The fluctuation magnitudes could be represented by a category of states (s_i) and a prior probability of occurrence. For a length of a study period, the groundwater fluctuation could be viewed as a sequence of the states which occur at each time step shifting from state s_i to state s_j in accordance with their transitional probability M_{sisj} formulated in a Markov notation (equation 3.1).
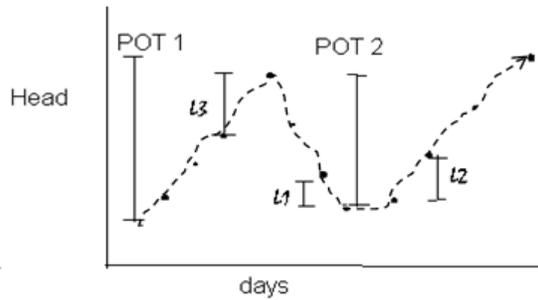
Figure 3.1 Daily fluctuation magnitudes, and peak-to-threshold representation (POT) of groundwater fluctuation. POT1, POT2 are (peak-to-threshold at times 1 and 2)

$$M_{S_iS_j} = \begin{vmatrix} M_{S_1S_1} & \circ & \circ & M_{S_1S_n} \\ \circ & \circ & & \circ \\ \circ & & \circ & \circ \\ M_{S_nS_1} & \circ & \circ & M_{S_nS_n} \end{vmatrix} \qquad (3.1)$$

3.3.2. Directional Patterns of groundwater fluctuation

Adopting states of intermittent flow pattern suggested by Aksoy (2004), the pattern of the groundwater fluctuation could be categorized as the recharge mode (increasing water level), discharge mode (decreasing water level), or constant (for no

changes in water level). Schematically, it could be represented as shown in Fig. 3.2, with notations of a, d and c to represent ascending, descending or constant. Each of the patterns are observed ('emitted') associated with the states of the magnitude s_i. For each occurrence of state of magnitude s_i an associated emission probability e_p as is tabulated described in Table 3.
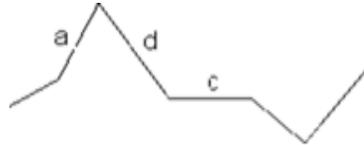
Figure 3.2. Groundwater fluctuation pattern

Table 3.1 Emission probability representation

$e_{a,si}$ is the probability of state s_i on pattern a

e_p =

	A	D	C
S1	$e_{a,s1}$	$e_{d,s1}$	$e_{c,s1}$
S2	$e_{a,s2}$	$e_{d,s2}$	$e_{c,s2}$
.	.	.	.
Sn	$e_{a,sn}$	$e_{d,sn}$	$e_{c,sn}$

3.4. Methodology

3.4.1. Building a state sequence model from the groundwater time series data

The daily time series of groundwater fluctuation was conceptualized as sequence of states of magnitude and direction that have a joint probability of occurrence. So, any realization of a sequence (time series of groundwater fluctuation) constitutes a combination of n states of magnitude, s_1, s_2, ..., s_n , that have a probability of occurrence derived from the joint probabilities of the transition of states. Using an analogous relationship of genome sequencing from Durbin et al. (2001), the probability of a sequence could be defined by (equation 3.2)

$$\text{Pr (Seq | States)} \quad = \quad \text{Pr } (s_j{}^t, s_r{}^{t\text{-}1}, \ldots s_i{}^k, s_1{}^1) \tag{3.2}$$

$$= \text{Pr } (s_j{}^t \mid s_r{}^{t\text{-}1}, \ldots s_i{}^k, s^1) \; \text{Pr}(s^{t\text{-}1} \mid s^{t\text{-}2} \ldots s^1) \ldots \text{Pr}(s^1)$$

where $s_i{}^k$ is the ith state of magnitude of fluctuation that occurred at the kth order

of a t length sequence (time in this context). For the first order Markov chain, the

probability of each state of magnitude at the kth order (s^k) is dependent only on the (k-1)[th]

order state $(s^{k\text{-}1})$ in the sequence ; leading to a simplistic formulation (equation.3. 3)

whose final representation uses the notation of equation.3.1.

$$\text{Pr (Seq | States)} = \text{Pr}(S^1)\prod_{k=2}^{t} \text{Pr}(S^k \mid S^{k-1}) \tag{3.3}$$

$$= (M_{s^1})M_{S^1S^2} M_{S^2S^3} \cdots M_{S^{t-1}S^t}$$

In the kth order of the sequence, a given state of magnitude has three possible states of

patterns with emission probability, e_p (Table 3.2). Assuming independence of occurrence

of the patterns, the probability of an observable pattern of the fluctuation over a given

sequence of states (s) of magnitudes is a conditional probability derivable from (equation

3.4) and implemented as shown in (equation 3.4).

$$\text{Pr}(\; pattern \; \mid states \;) = \prod_{k=1}^{t} P(e_k \mid s_k) \tag{3.4}$$

$$= e_{s^1} \cdot e_{s^2} \cdots e_{s^k}$$

Where e_{s^k} is the emission probability for each states at the kth time order. The

probability of a realizable sequence can be determined using the joint probabilities of the

states of magnitude and patterns represented by the transitional and emission probabilities

respectively as shown in equation 3. 5.

$$\Pr([Pattern, States] | Seq) = \Pr(S^1) \cdot e_{s^1} \cdot M_{S^1 S^2} \cdot e_{S^2} \cdot M_{S^2 S^3} \cdots e_{s^k} \cdot M_{S^{t-1} S^t} \quad (3.5)$$

3.4.2. Choice of a realization of a sequence

A realization of magnitude and pattern at each order of the sequence depends on the maximum probabilistic value determined from the joint probability (equation 3.5). In other words, the sequence will be selected using a critical path that connects states and patterns whose joint probability is maximum at each time step (in the order of the sequence) (Durbin et al., 2001). Evaluation of probabilities at each step in the sequence and selection of states and emission is called the Viterbi critical path approach, demonstrated in Figure 3.3. The Viterbi critical path choice for the maximum joint probability 'V' at the kth order for the t long sequence is commonly done using a requirement (equation 3.7) (Durbin et al., (2001).

$$V_j(k) = \underbrace{Max}_{1, 2 \cdots n} [V_i(k-1) \cdot M_{S_i S_j}] \cdot e_{S_j}(p) \quad (3.6)$$

For the first order of the sequence, a state that gives the maximum joint probability with all combinations of emission probabilities is chosen. However, in most practical cases, the iteration begins by assigning a given state with a probability 1 to initiate. The different possible states also take initial probability of 1 whenever new sequence is started. For conservative steps, the product of each state probability $P(s_i)$ and emission probabilities (e_{si}) will be compared and the state and emission with the maximum product will be chosen. In order to choose a state for the second order in sequence, the maximum product should constitute: (i) the transitional probabilities M_{SiSj} from the Markov model that assigns its value on the basis of the chosen state in step 1 to all possible new states of magnitude 's_i'; (ii) possible emission probabilities from Table

3.2; and, (iii) the previous probability (determined at the 1st step) that serves as a weighting factor. In general, the choice of the state and pattern at the kth order (other than first step) has to use a product of preceding maximum probability and the joint probability of the transition from previous state i-1 to the next state i defined by the transitional and emission probabilities (equation 3.5). The output of the iteration will be an ensemble of a sequence of states and patterns (time series of fluctuation).

The steps followed in selection Viterbi critical path are as follows

1. Start by choosing an initial state in the sequence that has maximum joint probability in terms of magnitude and pattern:

$$V_j(1) = \underbrace{Max}_{1,2\cdots n}[P_{S_i}] \cdot e_{S_j} \tag{3.7}$$

2. The states and pattern in the 2nd order of the sequence is chosen as

$$V_j(2) = \underbrace{Max}_{1,2\cdots n}[V_i(1) \cdot M_{S_i S_j}] \cdot e_{S_j} \tag{3.8}$$

3. The states and pattern in the kth order of the sequence is chosen as:

$$V_j(k) = \underbrace{Max}_{1,2\cdots n}[V_i(k-1) \cdot M_{S_i S_j}] \cdot e_{S_j}(p) \tag{3.9}$$

4. Continue the selection till the end of the sequence length.
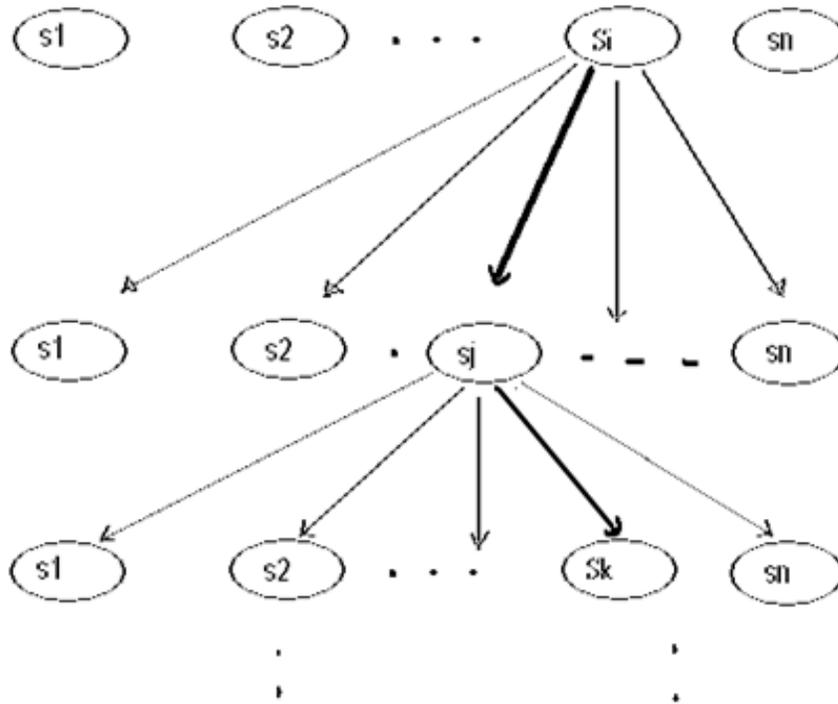
Figure 3.3. Trellis diagram of the Viterbi critical path selection

3.4.3. Wells Selection

The four major surficial hydrogeological units categorized by SFWMD are geographically classified as the lower east coast area, the lower west coast area, Kissimmee basin, and the upper east coast (Figure 3.4). The hydrogeological unit at the upper east coast called the Anstasia formation was dropped for its representation of very small area in the district. The lower eastern coast area was identified by the highly transmissive Miami limestone; the Kissimmee basin has unidentified formations overlaid by clastic sandy medium soil that has high drainage characteristic; and the western coastal area was identified by Forthompson formations that has poor transmissivity.

Wells farther than 2 miles from riverside and canals were selected (to capture the natural fluctuation) from the three hydrogeologic units using the Geographic Information

Systems- ArcGIS analyst. Three well sites, G3356, POS-12, and ST2-G stations, which are situated at lower east coast, Kissimmee Basin, and the west coast area in order were selected. The selection of specific well was randomized from a number of wells within the hydrogeologic unit. Groundwater level measured by the SFWMD was collected from the SFWMD DBHYDRO database.
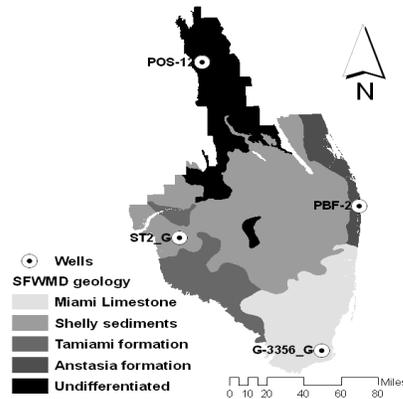
Figure 3.4 Wells selected from SFWMD

3.5. Results and Discussion

The raw data of groundwater fluctuation head showed extended patterns of discharging with some episodes of recharge for months from October to May (a dry season in SF). Whereas the period from June to September could be characterized by frequent recharging and rise of the water level. Multiple years of data from POS-12 station (Figure. 3.5) showed almost the same pattern of seasonal variation, which was similar across all other sites. The pattern from June to September showed rapid fluctuations while months from October to May are dominated by decreasing head pattern. Despite the seasonal pattern, the annual data set from year to year showed very small differences of variances $(0.1 - 0.4)$, and the assumption of stationarity was held for the Markov stochastic model application.
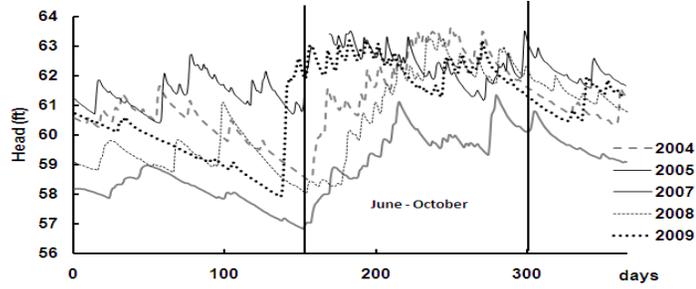
Figure 3.5 Groundwater head fluctuation at POS-12 station, 2004-2009

3.5.1. States representation and prior probability estimation

First, daily fluctuations were calculated from the total head data taking difference of heads from consecutive time steps using equation (3.11). Second, the maximum daily fluctuation was chosen from the daily averaged data for the study period of each of the sites (2000-2009 for G3356 and ST2-G stations, and 2004-2009 for POS-12 station). Thirdly, states of water level changes were represented by segmenting the maximum observed daily water level change into 10 states at 10% interval. So the first state represented water level changes that are below 10% the maximum daily fluctuation; the second state represented water level changes between 10-20 % of the maximum daily fluctuation etc., and continued up to the 10th state representing 90 - 100% of the maximum daily water level change (equation 3.12).

$$h_n = H_n - H_{n-1}, \qquad (3.10)$$

$$S_i = [10 \cdot i \cdot \text{Maximum} (h_n)] \ \%. \qquad (3.11)$$

n = 1, ….,365 , h_n is the nth date fluctuation, H_n, head of groundwater level on n^{th} day and S_i is the ith state

The prior transitional probability was estimated for the 3 sites from the data (Janssen et al., 2006) and denoted by Ms_i,s_j to represent the probability of transitions

from each of the 10 states $s_1, s_2, \ldots, s_{10}$ (Tables 3.2a, 3.3a and 3.3a). Also emission probabilities that represent states of the pattern were derived from daily differences of water levels (Janssen et al., 2006) assigning positive changes (ascending) as "a", negative changes (descending) as "d", and no changes in water levels (constant) as "c" (Tables 3.2b, 3.3b and 3.4b).

Table 3.2 Transitional (a) and emission (b) probabilities, station G3356

$M_{SI,SJ}$	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	$e_{si, a, d}$	a	d	c
s_1	0.31	0.30	0.21	0.11	0.03	0.03	0.01	0.00	0.00	0.00	S_1	0.56	0.44	0.00
s_2	0.29	0.40	0.20	0.05	0.02	0.02	0.02	0.01	0.00	0.01	S_2	0.64	0.36	0.00
s_3	0.30	0.35	0.16	0.16	0.00	0.01	0.00	0.01	0.00	0.00	S_3	0.59	0.41	0.00
s_4	0.23	0.28	0.18	0.23	0.03	0.03	0.00	0.00	0.00	0.05	S_4	0.55	0.45	0.00
s_5	0.22	0.44	0.22	0.00	0.11	0.00	0.00	0.00	0.00	0.00	S_5	0.44	0.56	0.00
s_6	0.00	0.43	0.29	0.14	0.14	0.00	0.00	0.00	0.00	0.00	S_6	0.00	1.00	0.00
s_7	0.25	0.50	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	S_7	0.00	1.00	0.00
s_8	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.33	0.00	0.00	S_8	0.33	0.67	0.00
s_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_9	0.00	0.00	0.00
s_{10}	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	S_{10}	0.00	1.00	0.00

Table 3.3 Transitional (a) and emission (b) probabilities, station POS12

$M_{SI,SJ}$	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	$e_{si, a, d}$	a	d	c
s_1	0.57	0.29	0.08	0.03	0.01	0.01	0.01	0.01	0.00	0.01	S_1	0.32	0.68	0.01
s_2	0.36	0.50	0.05	0.06	0.01	0.01	0.01	0.00	0.00	0.01	S_2	0.29	0.71	0.00
s_3	0.17	0.37	0.26	0.03	0.09	0.00	0.00	0.06	0.00	0.03	S_3	0.46	0.54	0.00
s_4	0.20	0.40	0.20	0.13	0.07	0.00	0.00	0.00	0.00	0.00	S_4	0.67	0.33	0.00
s_5	0.57	0.29	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_5	1.00	0.00	0.00
s_6	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	S_6	1.00	0.00	0.00
s_7	0.00	0.00	0.67	0.00	0.00	0.33	0.00	0.00	0.00	0.00	S_7	1.00	0.00	0.00
s_8	0.25	0.25	0.00	0.00	0.00	0.00	0.25	0.00	0.25	0.00	S_8	1.00	0.00	0.00
s_9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_9	1.00	0.00	0.00
s_{10}	0.33	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_{10}	1.00	0.00	0.00

Table 3.4 Transitional (a) and emission (b) probabilities, station ST2-G

$M_{SI,SJ}$	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	$e_{si,a,d}$	a	d	c
s_1	0.80	0.11	0.04	0.02	0.03	0.00	0.00	0.00	0.00	0.00	S_1	0.21	0.77	0.02
s_2	0.52	0.36	0.03	0.03	0.03	0.02	0.00	0.00	0.00	0.00	S_2	0.26	0.74	0.00
s_3	0.29	0.43	0.10	0.10	0.05	0.00	0.05	0.00	0.00	0.00	S_3	0.71	0.29	0.00
s_4	0.75	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_4	0.75	0.25	0.00
s_5	0.42	0.17	0.33	0.00	0.08	0.00	0.00	0.00	0.00	0.00	S_5	1.00	0.00	0.00
s_6	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_6	1.00	0.00	0.00
s_7	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.50	S_7	1.00	0.00	0.00
s_8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_8	0.00	0.00	0.00
s_9	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	S_9	1.00	0.00	0.00
s_{10}	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	S_{10}	1.00	0.00	0.00

3.5.2. State sequence simulation

In order to observe relationship of states sequence representation with the observed data, the annual head of the state sequence (derived from the data) was fitted to the observed annual pattern of the groundwater level. The state sequence derivation from the data required substitution of the actual daily groundwater fluctuation by the states of magnitude and pattern, and adding them at daily time step to obtain the annual fluctuation. The simulation was done using Matlab script and fitted to the observed data. The close agreement confirmed that the assumption of state sequence conceptualization was realistic as shown for the case of G3356 station in Figure. 3.6b, and for ST2-G and POS-12 stations in Figures 3,7a and 3.7b, respectively. Three of the sites showed a good agreement with the observed data. The specific fit confirms the possibility to represent the groundwater fluctuation using state sequences.

Using the prior transitional and emission probabilities, random sequences were generated that assigns each new sequence a posterior joint probability derived from the transitional and emission probabilities (Figure 3.6c). The selection of a realization was

done on the basis of the maximum likelihood ratio criteria. Since the joint probability of the sequence (implemented using (equation 3.5)) had very small values, known to cause numeric underflow problems (Macdonald, 1997), the logarithmic likelihood ratio was used (equation 3.12). The MLR selection criterion was set at 90% to limit the increased computational time and also avoid failure of convergence as the requirement was increased.

$$MLR = \frac{\sum\limits_{sample} \log \Pr(S_1) + \log e_{S_1} + \cdots \log M_{S_i S_j} + \log e_{S_j})}{\sum\limits_{observed} \log \Pr(S_1) + \log e_{S_1} + \cdots \log M_{S_i S_j} + \log e_{S_j})} \geq 90\% \qquad (3.12)$$

MLR is the probabilistic ratio of the sample sequence against the observed data. However, selecting a sequence that had the same probability as the observed posed a pattern irregularity problem on the realizations at detailed points in the sequence order. In other words, the MLR ratio did not guarantee the resemblance of the pattern. So, a root mean square error (RMSE ≤ 0.15m) was employed as second criteria at least to narrow the gaps at peak and threshold patterns with the observed data. The result was a closer fit of the simulated fluctuation with the observed pattern for all sites (Figure 3.6d, 3.7c and 3.7d), whose RMSE and MLR statistic showing the agreement (Table 3.6). The graphical fit was acceptable given the variability of fluctuations over a year that has moment of discharging and recharging.

Table 3.5 RMSE and MLR of the simulated against the observed

Site	RMSE	MLR
G3356	0.08	90%
POS-12	0.13	90%
ST2-G	0.14	90%
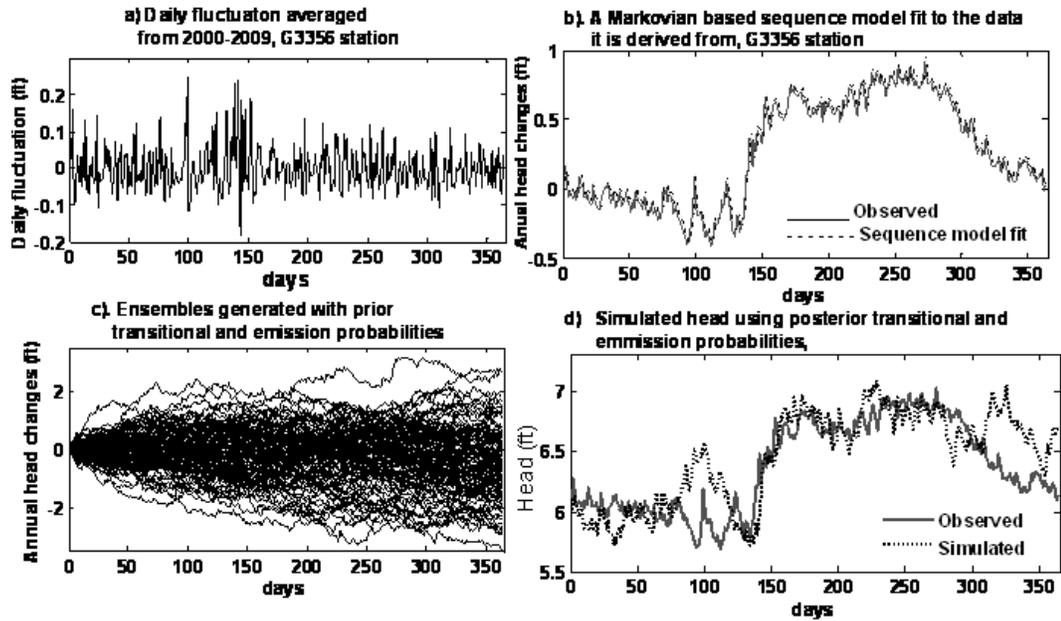
Figure 3.6 Station G3356: a) Daily fluctuation b) state sequence representation c) ensemble generation and d) Simulated fit
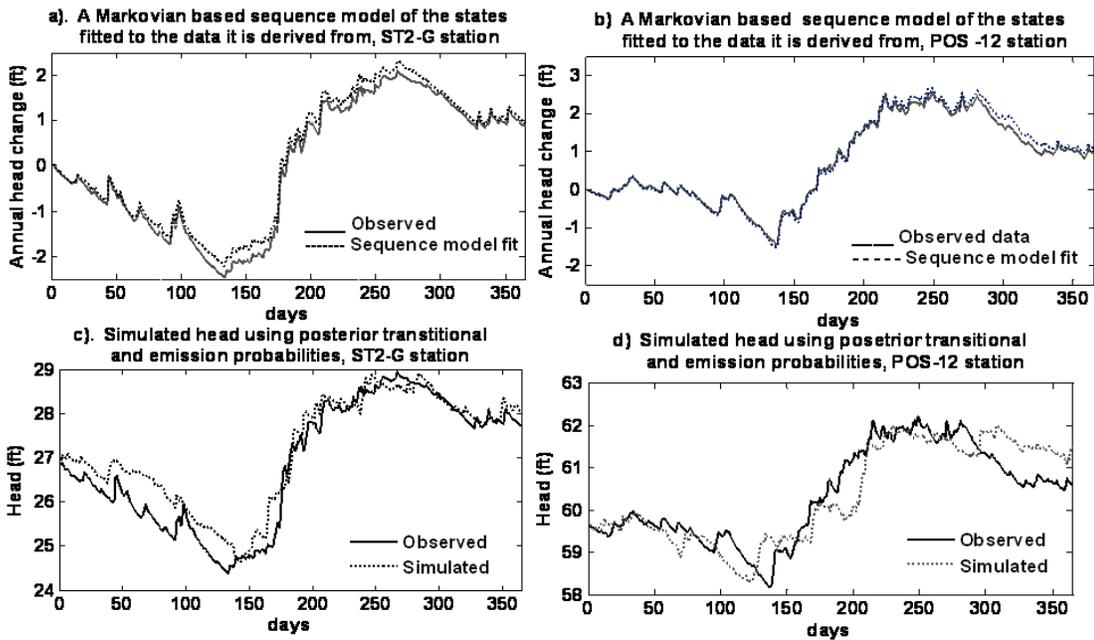
Figure 3.7 Station POS-12: a) state sequence representation, ST2-G station: b) state sequence representation, POS-12 station: c) simulated fit, ST2-G station: d) Simulated fit

3.5.3. Integration of the Stochastic Sequence Model for operational prediction

One key application of the groundwater fluctuation modeling was for operational prediction. In this analysis, the observed daily fluctuation data was first segmented for each month. The first month data (January for this case) was considered as an observed data used for selection criteria test of a realization that had to satisfy the RMSE and MLR to accept the forecasting for February. In order to forecast for March, the prior transitional and emission probabilities were re-estimated from the observed historical data upon incorporating the new observations from February. The process continued adapting the transitional and emission probabilities as more observations are made through the period of study (12 months). In all the 12 month forecasts, the MLR was kept above 90% while the RMSE was below 0.15. The result showed a consistent forecasting with the observed data except at the timings of seasonal shift (in this case June and October), Figure. 3.8. The first order Markov model depends only on a one step prior state. So, the seasonal shift might have caused a bias on prior probability estimation as a drainage dominated month of May was expected to predict a recharging (wet season) month of June, and vice versa for the case of October. Even then, the observations in both June and October automatically corrected the forecasting as the observation was incorporated for the forecast of the preceding months. Separate analysis of forecasting on half-yearly basis starting once in October (beginning of drier season) and the other starting in June (beginning of wet season) might give an improved RMSE and a pattern of fit at all cases compared to the annual cycle forecast. However for simplicity (a requirement of operational prediction) annual cycle prediction could be acceptable as long as the RMSE is less than 0.15m.
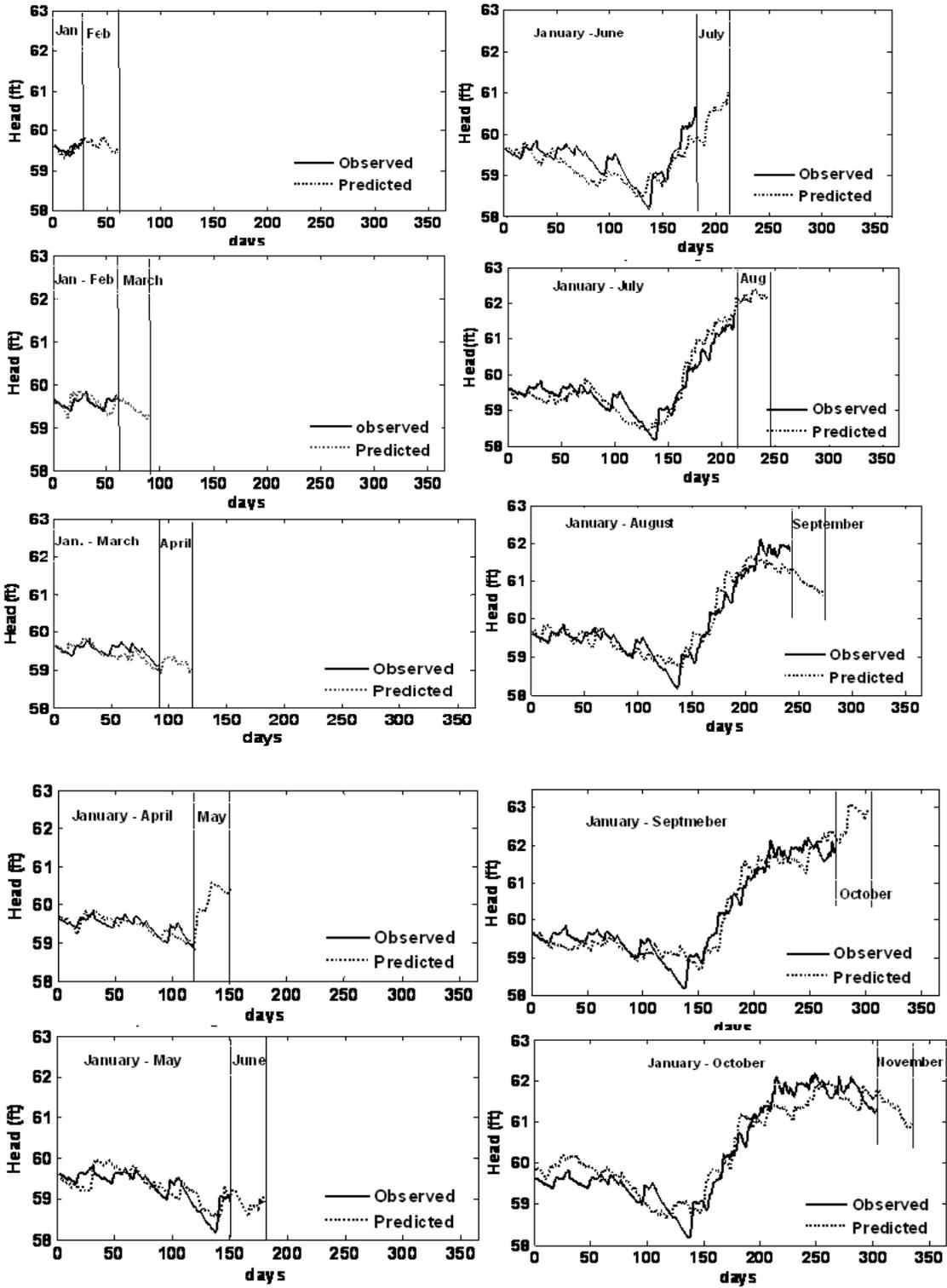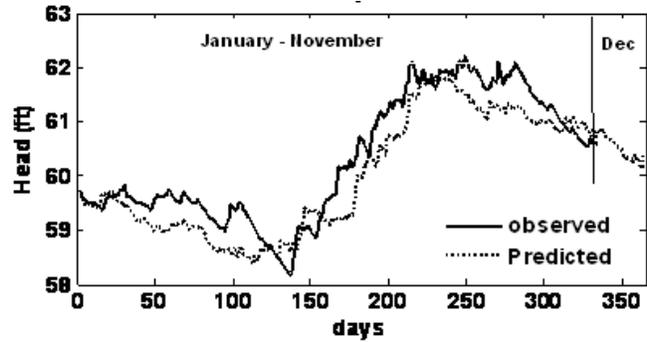
Figure 3.8 Monthly spanned forecasted groundwater level at POS-12 station

3.6. Conclusion

The agreement between state representation of the daily fluctuations and the observation data confirmed that the sequence based stochastic model offers an alternative approach for simulating and forecasting the water table. The states representation simulates both the pattern and magnitude acceptably, and its adaptability was proven via the Markov model specification. The method could be employed at all sites as demonstrated in the three different hydrogeological units which almost cover the district.

For a more accurate simulation and forecasting, both MLR and RMSE criteria had to be employed to retrieve the pattern with good agreements on peak and thresholds as well. Acceptable limits of MLR and RMSE need to be set to retrieve a realization from a possible random sequence generation with a defined sequence probability. Raising the requirement of RMSE has led to convergence compromising MLR and vice versa. In other words, as MLR criteria was raised and the computational time grew exponentially to satisfy RMSE, in some cases with failure of convergence. So requirements of MLR and RMSE at 90% and 0.15 m, respectively required optimal computation time without compromising the results.

37

The model testing for forecasting through segmented daily fluctuation data on monthly basis showed that except for months when seasonal changes occur, the results were in good agreement. However, experts knowledge of seasonal changes and conscious analysis would help till more data from the new season could be obtained and incorporated.

The model could be used for point scale forecasting at different sites over SFWMD. Apparently, interpolating forecasted data obtained from point scale simulation would help to develop regional contours that would need to be developed for comprehensive synthesis of the hydrological state. Such an approach is vital for SF which requires close monitoring and understanding of the surface-groundwater interaction over coastal areas, urban zones, and wetland ecosystems. In other words, it could serve as a synthetic groundwater data generating component for a regional scale spatiotemporal forecasting models inclusive of watershed based models. Finally, observation on seasonal versus annual cycle forecasting would be relevant for a further critical analysis and parameterization of breakpoints.

Acknowledgment

REFERENCES

Ahmad S, Prashar D (2010). Evaluating Municipal Water Conservation Policies Using a Dynamic Simulation Model, Water Resour Manage 24(2010)3371–3395

Aksoy, H. (2003) Markov chain-based modeling techniques for stochastic generation of daily intermittent stream flows. *Adv. Water Resources.* 26(6), 663–671.

Aksoy H. (2004). Pulse generation models for daily intermittent Stream flows, *Hydrological Sciences–Journal–des Sciences Hydrologiques*, 49**(**3) 399-411

Ali A. (2009). Nonlinear multivariate rainfall-stage model for large wetland systems, *Journal of Hydrology*, 374(2009)338-350

Durbin R., Eddy S., Krogh A., Mitchison G. (1998). Biological Sequence Analysis, Probabilistic Models of proteins and nucleic acids, Cambridge University Press, pp 356

Florida Council of 100 (2003). Improving Florida's Water Supply Management Structure**;** Ensuring and Sustaining Environmentally Sound Water Supplies and Resources to Meet Current and Future Needs (http://www.fc100.org/documents/waterreportfinal.pdf).

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

Gunderson L H, (2001). South Florida: The reality of change and prospects for Sustainability; Managing surprising ecosystems in southern Florida, Ecological Economics 37:371-378

Janssen J., Manca R., (2006). Applied Semi-Markov Processes, Springer, New York; pp 302

Maliva R. G., Missimer T. M., Walker C. W., Owosina E. S., Dickson J. A. D. and Fallick A.E., (2001), Carbonate diagenesis in a high transmissivity coastal aquifer, Biscayne aquifer, southeastern Florida, USA. *Sedimentary Geology*, Volume 143, Issues 3-4, 287-301

MacDonald L I, Zucchini, 1997. Hidden Markov and Other Models for Discrete-valued Time series; Monographs on Statistics and Applied Probability 70, CHAPMAN & HALL, London,

Najib K, Jourde H., Pistre S. (2007). A methodlogy for Extreme groundwater surge predetermination in carbonate aquifers: groundwater flood frequency analysis, *Journal of Hydrology* Volume 352, Issues 1-2, (1-15)

Park E, Parker J.C. (2008). A simple model for water table fluctuations in response to precipitation, Journal of Hydrology 356 (2008)344-349

Paynter p Nachabe M , Yanev M. (2011). Statistical Changes of Lake Stages in Two Rapidly Urbanizing Watersheds, Water Resour Manage 25(2011)21–39

Shamshad A, Bawadi M.A., Hussin W.M.A., Majid T.A., Sanusi S.A.M. (2005). First and second order Markov chain models for synthetic generation of wind speed time series, Energy 30 (2005) 693–708.

Szilagyi J., Balint G., Csik A. (2006). Hybrid, Markov Chain-Based Model for Daily Stream   flow Generation at Multiple Catchment Sites, *Journal of Hydrologic Engineering*, 11(3): 245-256.

Zahabiyoun B. (2005). Stochastic generation of stream flow data; International Symposium on Stochastic Hydraulics, 23 and 24 May 2005, Nijmegen, The Netherlands.

Zhang, Y., Pulliainen, J., Koponen, S., & Hallikainen, M. (2002). Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment, 81*(2-3), 327-336.

CHAPTER IV

4. Spatiotemporal surface-groundwater interaction simulation in South Florida

[1]Chebud, Y., [1] Melesse, A.,

[1]Department of Earth and Environment, Florida International University, FL, USA

Chebud Y., e-mail: ycheb002@fiu.edu; Tel: 305-348-7664, Fax; 305-348-6137

Melesse, A., e-mail: melessea@fiu.edu; Tel: 305-348-6518: Fax; 305-348-6137

4.1. Abstract

South Florida ecosystem is dictated by a large wetland, karst hydrogeology and extended coastal boundary with the Atlantic Ocean. The risks related to the ecosystem include: disruption of groundwater flow as a result of frequent sinkhole formation; flooding in urban areas as a result of the shallow water table; saltwater intrusion from the ocean; and excessive nutrient load to surficial water bodies and subsequently eutrophication because of the intensive utilization of wetlands for nutrient removal. Attempts to understand ecohydrological processes primarily focus on extensive monitoring and use of distributed hydrological models. However, the relatively flat nature of the region and also the extended coastal boundary with the ocean, makes watershed based approaches less realistic. A regional spatiotemporal groundwater level modeling approach was attempted using a Dynamic Factor Analysis (DFA) method. The daily water levels of 13 monitoring well sites from major hydrogeologic regions and different land uses were used to conduct the DFA analysis, and six dynamic factors were identified using minimum Akaike Information Criterion (AIC). Further exploratory analysis to relate the dynamic factors

with physically attributable explanatory variables has helped to identify five of the major factors that govern the groundwater dynamics in south Florida. Three of the factors were attributable to the Lake Kissimmee water level in the north, Caloosahatchee River water level in the west, and Hillsboro canal in the east. The other two factors identified were the regional averaged rainfall and soil moisture. The spatiotemporal simulation involved interpolation of the loadings of the dynamic factors using an inverse distance weighted method and convoluting with the dynamic factors. The result has shown a good fit with the maximum RMSE of 0.12m. Given the fact that the factors, namely rainfall, soil moisture, and surface water level are derivable from satellite imagery, spatiotemporal operational modeling of the groundwater level is achievable. Making use of point scale groundwater level predictive models as a sub-component, the modeling approach could be used for regional scale operational forecasting.

Key words: Surface-groundwater interaction, Groundwater, South Florida, Akaike Information Criterion, Spatiotemporal modeling, Dynamic Factor Analysis

4.2. Introduction

Surface-groundwater interaction in southern Florida is dictated by a wetland ecosystem, karst surficial aquifer, and the coastal boundary with the Atlantic Ocean (Gunderson, 2001). The wetland ecosystem interacts with the stream flow retaining flood during the wet season and discharging subsurface water to the streams during the dry season. The karst hydrogeology dictates direction of groundwater flow as well as the efficiency of nutrient removal vis a vis the dissolution of limestone and hence formation of sinkholes (Genereux and Slater, 1999). The coastal boundary with the Atlantic Ocean poses saltwater intrusion, especially when groundwater head drops.

Apparently any subsurface water management in the SF should take into account: the risks associated with ecological failure because of the connectivity of groundwater to the wetlands at times of drought; the risk of contamination by saltwater in the coastal areas where there is very thin vadose zone and over pumping of groundwater; flooding of urban areas from groundwater surge because of the flat terrain and shallow water table; and the risks associated with the use of STAs in the karst areas where nutrient attenuation could not be achieved.

Recognizing such risks, South Florida Water Management District (SFWMD) uses over 1854 networked stations as a tool to monitor the ecohydrology of the region. Harvey et al. (2004) reported the logistical challenge to monitor connectivity of groundwater fed micro-ecosystems attributed to the large expanse of the wetland. The karst hydrogeologic setting of the district increases the complexity of monitoring because of the dynamic property of the hydraulic conductivity potentially causing flow directional changes (anisotropy), (Genereux and Slater, 1999). Evidently, real-time analysis and synthesis of the groundwater hydrology specifically spatiotemporal modeling of surface-groundwater interaction at regional scale is vital for decision making. Such analysis would not be met using watershed based distributed hydrological models because of the absence of clearly delineated watershed boundary (indiscernible from the flat topography) with no flow dividing lines, except for the Kissimmee River basin that constitutes less than 25% of the land mass. In addition, the watershed based models are computationally expensive for operational prediction in karst hydrogeologic region, as in SF, because of the dynamic nature of the hydrogeological factors which raise

parameterization uncertainty whenever changes occur. For the purpose of operational modeling and prediction, low inputs and minimum parameterization is required.

In this study, a regional scale spatiotemporal model capable of simulating the daily groundwater level was developed making use of existing data. The regional scale spatiotemporal model is intended to be linked with point scale groundwater operational prediction for regional groundwater level forecasting. The model will be ultimately used to identify hot spots of groundwater drop/rise and also as a predictor for soluble phosphorus load. The point scale prediction was already developed for major hydrogeologic regions from which the model presupposes to use (Chebud and Melesse, 2011). Therefore, the specific objectives of the study were to:

- analyze the factors that influence regional groundwater level dynamics,

- develop a regional scale spatiotemporal modeling approach for future use on spatial analysis of nutrient load, salt water intrusion, flooding etc.

- develop regional scale operational prediction of groundwater dynamics.

4.3. Factor Analysis

Water table fluctuation in karst regions is affected by precipitation (Park and Parker, 2008, Knotters and Bierkens, 2001), air pressure changes and Moon's tide (Ma`rkus et al., 1999), the soil moisture (Hoogland et al., 2010), surface water level (Ritter and Mun'oz-Carpena, 2006, Genereux and Slater, 1999). The effects of some of the variables are measureable and categorized as explanatory while the background effects of other hidden variables are assumed as latent (Markus et al., 1999). The latent factors could also represent unidentified explanatory factors (Zurr et al., 2003), or lack of

proper parameterization. The concept helps to reduce a complex interplay of factors to a minimum number through factor reduction method (Aggarwal, 1998).

For explanatory variables reduction, Ramsey and Shaffer (2002) suggested sequential forward/backward stepwise regression methods imposing Akaike's Information Criteria (AIC) or the Baye's Information Criterion (BIC). The AIC and BIC inform the log-likelihood of a model penalizing it for every addition of a parameter as indicated on Equation (4.1) and Equation (4.2) (Claeskens and Hjort, 2008). The difference in the two criterion is that in the case of BIC, penalization increases with the increased number of data and hence it is stricter than AIC. The penalization is used to avoid complication from increased number of parameters.

$$AIC = 2\ln(\theta) - 2\text{length}(\theta) \tag{4.1}$$

$$BIC = 2\ln(\theta)\text{-log(n) length}(\theta) \tag{4.2}$$

where θ is the parameter and n is the number of the data, length is the number of components in the vector θ.

Another factor reduction method is a multivariate principal component analysis (PCA) that trade off with 'factor analysis' method (Davis, 2002). The PCA identifies the eigen structure of factors using Singular Value Decomposition approach (SVD) and retain the factors that give the maximum efficiency (highest content of the explainable variance) and discards less informative ones. The method could theoretically identify latent factors with prior knowledge of explanatory variables and factor loadings. However, variables such as groundwater fluctuation would be affected by the lag factors failing to meet independence assumption of factor analysis using PCA (Markus et al., 1999).

Lopes et al. (2008), Zurr and Pierce (2004) and Zurr et al. (2003) reported effectiveness of a dynamic factor analysis method to identify the latent factors and explanatory variables from a time series of the response variables. The method is reported to have flexibility and capture any patterns such as periodic, non-periodic and multiple jumps of time series response variables (Zurr et al., 2003). The method masks the seasonal elements in the model and hence the representation ultimately captures all season simulation (http://www.brodgar.com/index.htm). The DFA was used as a tool for spatiotemporal dynamic forecasting of groundwater fluctuation (Hoogland, 2010). Ritter and Mun'oz-Carpena (2006) reported employment of dynamic factor analysis for prediction of surface groundwater interaction in the Agricultural areas, north of Everglades National Park.

The dynamic factor analysis formulates the response variable as a combination of the effects of latent variables and explanatory variables as indicated on Equation (4.3) after Ritter and Mun`oz-Carpena (2006).

$$Y_j(t) = \sum_{i=1}^{I} \lambda_{i,j} \rho_i(t) + \mu_j + \sum_{k=1}^{K} \omega_{k,j} \psi_k(t) + \varepsilon_j(t) \tag{4.3}$$

$$\rho_i(t) = \rho_i(t-1) + \varphi_i(t) \tag{4.4}$$

where $Y(t)$ is the jth response variable at site j and time t; $\rho_i(t)$ is the ith unknown trend at time t; $\lambda_{i,j}$ represents the unknown factor loadings ; μ_j is the trend parameter, $\omega_{k,j}$ is a regression parameter; ψ_k is the kth explanatory variable; ε_j and φ_i are the observation error and systemic error terms that are independent, normally distributed, and each of them with zero mean but unknown covariance matrix.

In order to use DFA on surface-groundwater interaction, prior knowledge of the explanatory variables would simplify the model uncertainty. The most common explanatory variables for the groundwater fluctuation includes rainfall (Park and Parker 2008), surface-water level (Genereux and Slater, 1999) and pedosphere soil moisture (Hoogland et al., 2010, Visser et al., 2010). The unidentified hydrologic and climatic variables as well as the dynamic karst aquifer properties could be captured by the latent factors.

4.4. Dynamic factor analysis and spatial structure

A peculiar application of dynamic factor analysis is to understand the common trends of a large area that influence the response variable and their loadings. An important aspect of common regional factors versus their loadings is the possibility of incorporating the spatial structure into the later for a spatiotemporal dynamic modeling. Gamerman (2010), and Lopes et al. (2008) suggested implicit framing of the spatial structure into the loading 'λ_t' as indicated in Equation (4.5), while the common dynamic factors 'ρ_t' in Equation (4.4) capture the temporal changes.

$$\lambda_j \sim N(\beta_j, \sigma_j^2 \, \Omega_j)$$

$$\lambda_t(x) = \lambda_{t-1}(x) + w_t(x) \tag{4.5}$$

where λ_j is the loading from Equation (4.4) is independent and distance-based Gaussian Random field. β_j an N-dimensional mean vector. Ω_j is given by a correlation function $f(Is_l\text{-}s_kI)$ which could be exponential, spherical etc. t is time and x is a coordinate point

Other methods that exploit the spatial relationship of the loading factors for spatiotemporal dynamic formulation include geostatistical approach (Hoogland et al.,

2010; Visser et al., 2005; Knotters and Bierkens, 2001), Bayesian methods (Diggle and Edith, 2010; Lopes et al., 2008), Neural and Networks (Nayak et al., 2005). Amongst the methods, using Bayesian method is comparatively less advantageous for the intensity of computation whereas the geostatistical approach is the most commonly used (Hoogland et al., 2010).

4.5. Study Area

The study was conducted over the geographical boundary of SFWMD that distinctively constitutes the largest wetland, and extended coastal boundary with the Atlantic Ocean. The surficial aquifers of the SFWMD could be broadly categorized as the Biscayne aquifer and the undifferentiated aquifer system (Randazzo and Jones, 1997). The same report indicates the Biscayne aquifer covers the south eastern part of the district namely the Miami Dade, Broward, southern Palm Beach and eastern Marone counties. The lithographic layering constitutes up to 6m thick organic soils, 14 m Pamlico sand layer, 14 m thick Miami limestone, 150 m thick Forthompson formation, 40 m thick layer Ansatsia formation and 20 m thick Key Largo formation in consecutive order from top to bottom. Other part of the district is dominated by the undifferentiated formation whose Lithographic layering is similar to the Biscayne aquifer except that the top layer is alluvium unlike the organic soils of Biscayne and also absence of the Key Largo formation. The geographical exposition of the lithographic units in SFWMD is dominated by Miami limestone, Tamiami formation, Undifferentiated and the Anstasia formations as shown in (Figure 4.1b). The Miami limestone extends in the Eastern and south eastern part while the Tamimai formation covers the western and south western. The northern

part is a clastic sandy unidentified formation. The Anstasia formation, the smallest in proportion, covers the north eastern part of the district.

The land use of the district is categorically proportioned as 60% wetland, 20% pastureland, 10 % cropland, 6% forest land, 1% urban, and the rest 3% are other environmental service areas (Chebud et al., 2011).
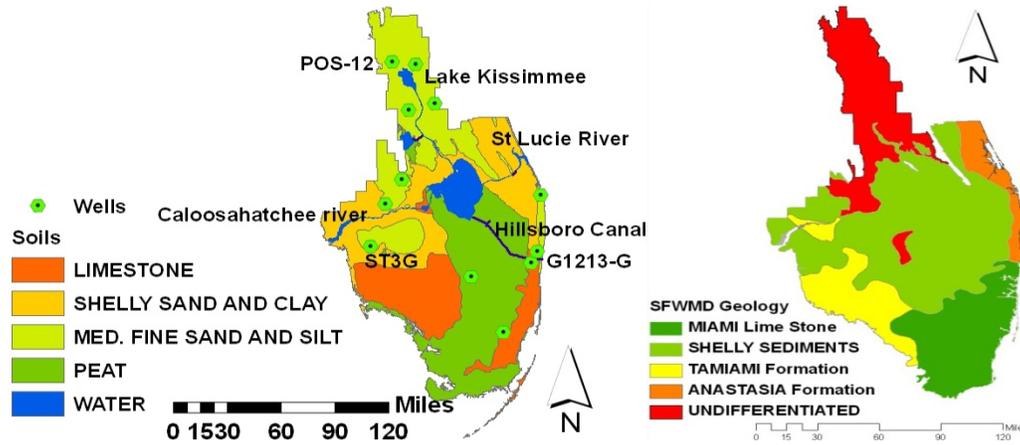
Figure 4.1 a Rivers, canals, lakes and well sites, Figure 4.1 b Hydrogeological units

4.6. Methodology

A stratified sampling was proposed at the outset to select groundwater wells from the three major hydrogeologic units splitting by soil type and four major land uses. The selection was made from sites which were at least 5 miles from canals to avoid daily influences of the canal operation. About 13 wells (Table 4.1a and Figure 4.1b) were selected from the intersection of a given hydrogeological unit, soil type, and land use using the Geographic Information Systems tool- ArcGIS analyst.

Table 4.1 Sampling distribution of wells

	Miami Limestone			Tamiami formation			Undifferentiated		
	Peat	Sandy	Bare	peat	Sandy	bare	peat	Sandy	bare
Urban		1	1		1	1	1	1	
Forest				1		1			
Crops				1				1	
Scrub/Range		1						1	1

The DFA was conducted for the normalized daily groundwater level from the 13 wells using Equation (4.3). A coupled statistical package of R was used accordingly as suggested by Zurr et al. (2003). Upon the DFA analysis, a search for the relationship of the identified factors against explanatory factors was made to give the identified factors a physical meaning and help for the development of a unified spatiotemporal model.

As a preliminary analysis of the explanatory factor search, the relationship of groundwater fluctuation with rainfall, canal water level, and soil moisture were observed for three stations (ST3G, POS-12, G1213-G, Figure 4.1a), situated in three different hydrogeologic units. Daily rainfall as well as daily canal water level readings were obtained from the nearest station of each well site (Figure 4.1b). The data sets of groundwater level, rainfall and canal water level were taken from the SFWMD DBHYDRO data base. The daily soil moisture data for multiple depths below surface (10cm, 40cm and 100cm) were obtained from the NOAA-Global Land Data Assimilation System (GLDAS) website. The soil moisture is available at 0.125^o for the USA from the National Land Data Assimilation System (NLDAS).

Backward and forward stepwise regression method was applied for each groundwater monitoring station to identify potential exploratory variables using the AIC

Criterion. Upon parameterizing rainfall and soil moisture by their respective regional average and normalizing the nearest canal / river water levels, a curve fitting method was used to analyze patterns of the explanatory variables and dynamic factor trends. Finally, invoking a spatial structure into the loading through interpolation using Inverse Distance Weighted (IDW) method, a regional groundwater level was developed at daily scale.

A descriptive framework of the methodological approach for the spatiotemporal modeling is indicated in Figure 4.2.
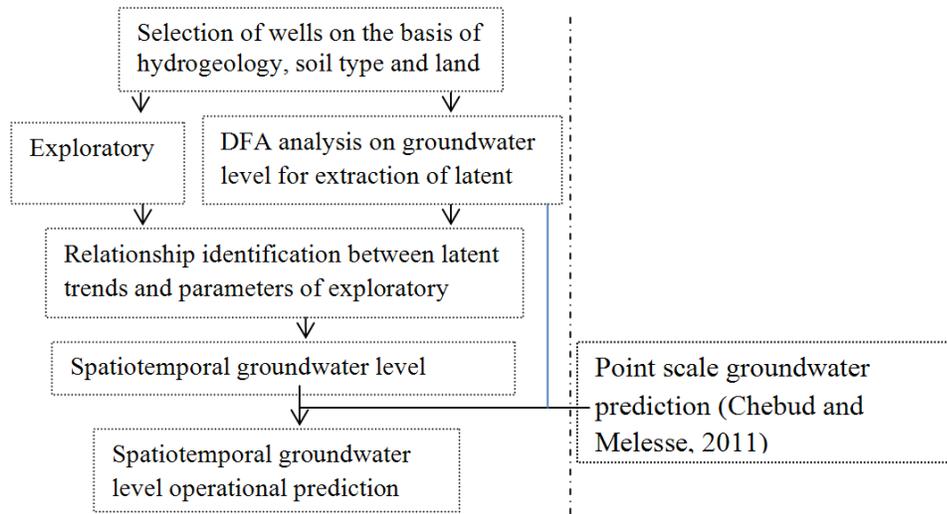
Figure 4.2 Methodological framework for the spatio-temporal groundwater modeling

4.7. Analysis

Using Equation (4.3), which offers several model types depending on the experimental design, a framework was chosen assuming all factors are latent. So the component of explanatory variable in Equation (4.3) was substituted by latent factor and formulated the model as:

Response variable (groundwater level) =   m common trends + noise   (4.6)

Applying the assumption, m > 1, the model served for factor analysis on smaller data sets (~20 events) as suggested in (http://www.brodgar.com/index.htm). The DFA analysis was employed on each month separately in 2004 for which complete data was available for the majority of the wells. The minimum AIC as well as the RMSE of the model was employed as a criterion to determine the number of common factors 'm' for the regional groundwater fluctuation.

The DFA analysis was done using normalized groundwater level data series from the 13 wells selected. The normalization is required expecting different units of the dynamic factors, for instance rainfall (mm), soil moisture (kg/m^2) and water level (m). Also the normalization removed the local trend μ_j from Equation (4.3), and reduce the complication which could rather be added on the output. The analysis was conducted using the averaged daily data from 2004-2009 in iterative fashion increasing the number of common factors mentioned in Equation (4.3) starting from 1. The analysis of the DFA for each month showed that minimum AIC was achieved when dynamic common factors (trends) are six (Table 4.2).

Table 4.2 AIC values of the DFA

m	Jan.	Feb.	March	April	May	June	July	August	Sep.	Oct	Nov	Dec
1	187	612	827	634	851	692	650	886	882	590	331	928
2	178	711	685	506	546	567	563	730	698	439	221	794
3	171	584	574	366	454	487	509	657	619	327	167	730
4	230	542	470	322	359	450	471	615	544	229	37	651
5	243	520	464	373	359	407	452	634	516	205	26	599
6	142	492	437	311	347	361	402	538	507	189	5	563
7	195	493	447	366	370	497	400	681	520	271	61	578

The analysis was further conducted for 2004, a year with complete dataset. The common dynamic factors were fixed this time to 6 and analysis was made each month using normalized daily data. The DFA analysis extracts a fit of the observed and simulated groundwater levels (Figures 4.3a and 4.4a) and the six dynamic trends (Figures 4.3b and 4.4b). In this paper, only outputs for the month of January and July are presented with the intention to represent dry and wet seasons. The maximum RMSE was only 0.12m for the month of July showing a good fit of the observed and simulated.

Table 4.3 RMSE values at each well site for January and July 2004

Stations	January	July
PB_1642	0.02	0.04
PB_1491_G	0.06	0.04
OSF_66	0.06	0.02
MUSEWELLS	0.04	0.01
MAXCEY_N_G	0.02	0.08
H_15A_G	0.03	0.12
G3554	0	0.04
AVON_P_G	0.07	0.01
var3AN1W4_GW2	0.01	0.03
SKYLAKE_G	0.03	0
POS_12	0.01	0
ST3G_G	0.04	0
G_12_13_G	0	0.05
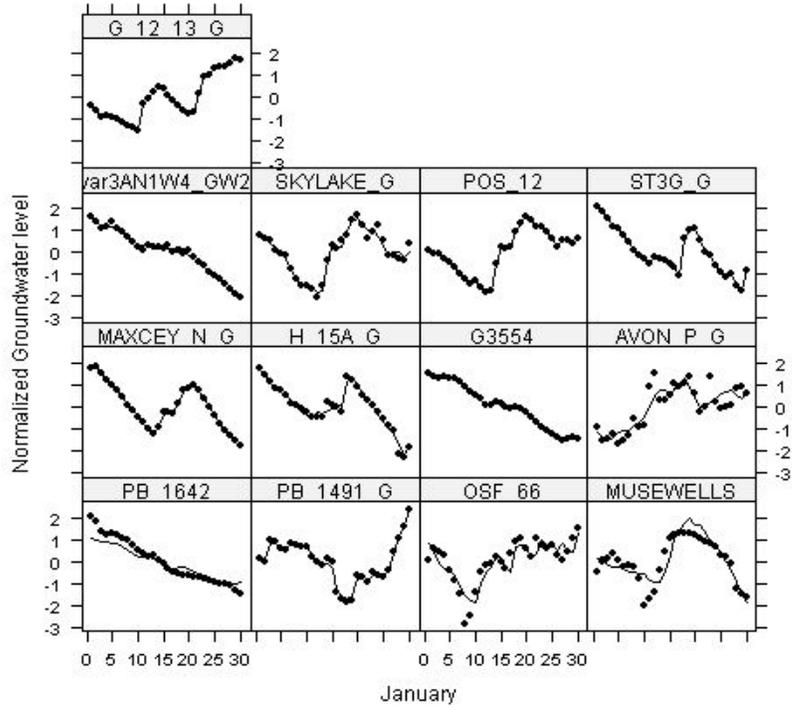
Figure 4.3 a Simulated and observed groundwater level, January 2004
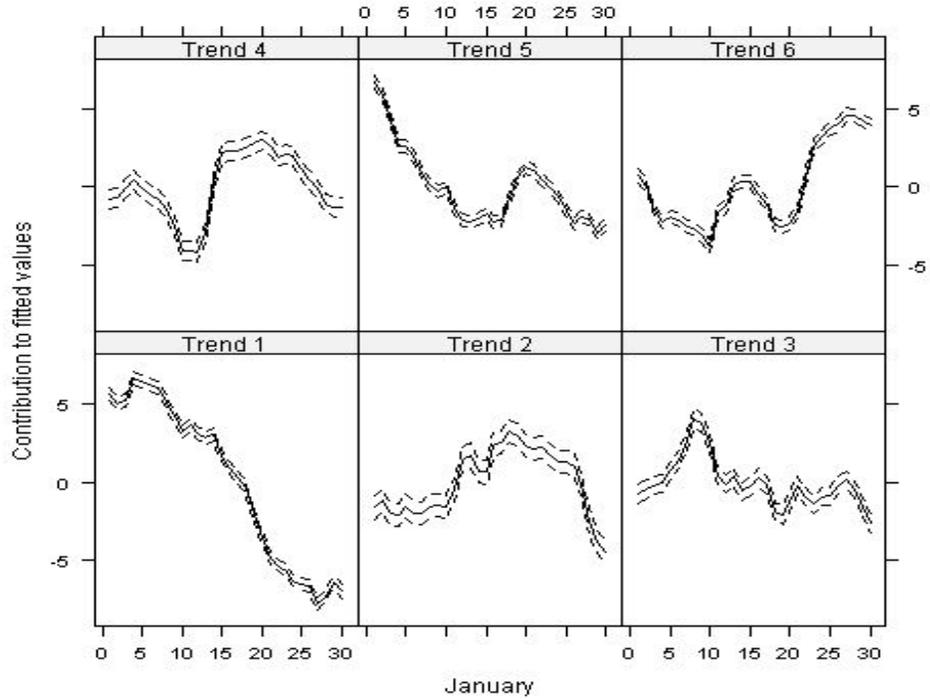
Figure 4.3b Dynamic trends (factors), January 2004
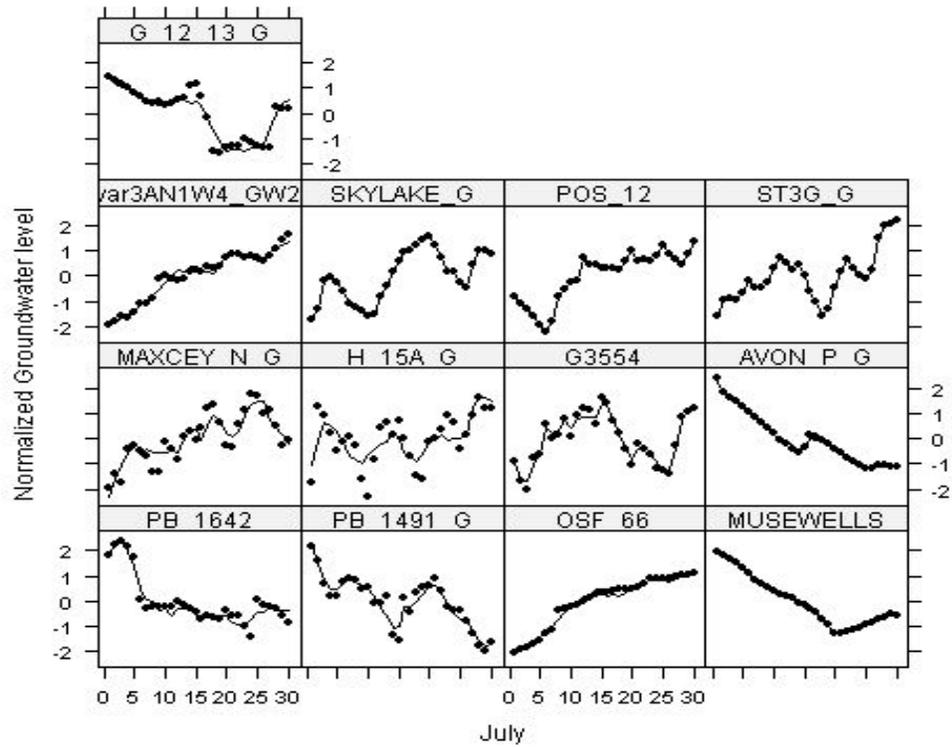
Figure 4.4a Simulated and observed groundwater level, July 2004
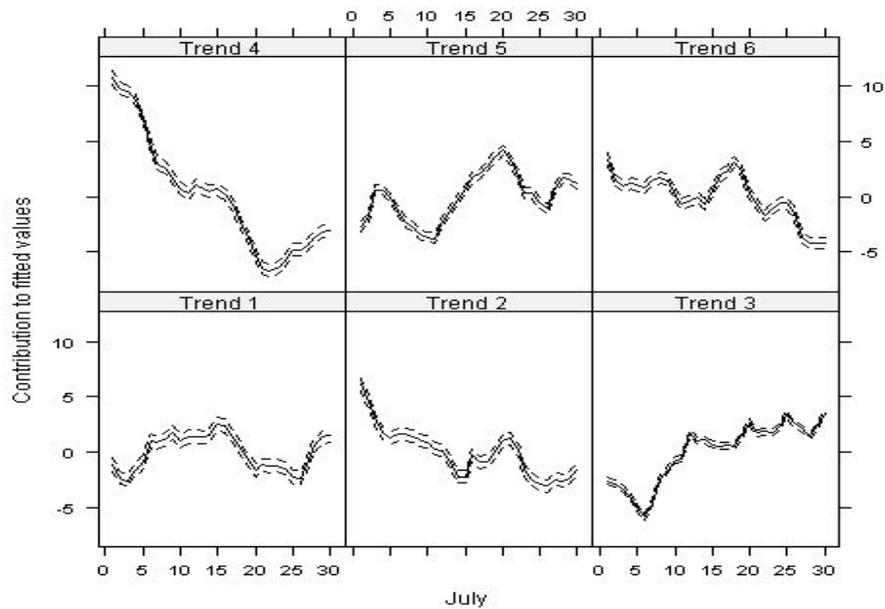
Figure 4.4b Dynamic trends (factors), July 2004

Associated with the six identified factors are their loadings that have similar meaning to a weight factor. The retrievability of the loadings at each observation point in space (Tables 4.4a and 4.4b) offers an opportunity to introduce the spatial structure through interpolation in the regional spatiotemporal simulation process. The loadings on Tables 4.4a and 4.4b, related to the January and July dynamic trends indicated in Figures 4.3b and 4.4b, are derived on condition of minimum AIC.

Table 4.4a Loadings (weights) of each trend for the month of January

Stations	Trend 1	Trend 2	Trend 3	Trend 4	Trend 5	Trend 6
PB_1642	0.144	0.035	0.104	-0.047	0.187	0.09
PB_1491_G	-0.01	-0.404	-0.002	-0.07	-0.036	0.001
OSF_66	-0.009	-0.017	-0.539	0.032	0.017	0.03
MUSEWELLS	0.056	0.163	0.046	0.392	0.013	0.038
MAXCEY_N_G	-0.001	0.056	0.063	0.14	0.345	-0.052
H_15A_G	0.063	0.17	-0.052	0.095	0.241	-0.071
G3554	0.147	-0.015	0.016	0.04	0.105	-0.038
AVON_P_G	-0.025	0.128	-0.299	-0.077	-0.238	-0.069
var3AN1W4_GW2	0.154	0.109	0.051	0.035	0.143	0.012
SKYLAKE_G	-0.062	-0.053	-0.166	0.314	0.162	-0.043
POS_12	-0.138	-0.08	-0.132	0.264	0.107	-0.09
ST3G_G	0.025	0.034	-0.12	0.019	0.307	-0.107
G_12_13_G	0.013	-0.008	-0.141	0.002	-0.119	0.292

Table 4.4b Loadings (weights) of each trend for the month of July

Stations	Trend 1	Trend 2	Trend 3	Trend 4	Trend 5	Trend 6
PB_1642	-0.246	0.006	0.022	0.204	0.036	-0.083
PB_1491_G	-0.119	0.358	-0.046	-0.098	-0.099	0.196
OSF_66	0.098	-0.076	0.129	-0.083	0.016	-0.032
MUSEWELLS	-0.002	0.029	-0.066	0.149	-0.067	-0.056
MAXCEY_N_G	-0.165	-0.38	-0.053	-0.08	0.02	0.163
H_15A_G	-0.045	-0.1	-0.012	0.108	0.151	-0.369
G3554	0.587	0.014	0.018	0	0.043	-0.077
AVON_P_G	-0.027	0.141	-0.078	0.086	0.021	0.078
var3AN1W4_GW2	0.124	-0.004	0.112	-0.095	0.057	-0.101
SKYLAKE_G	0.088	-0.058	0.001	-0.001	0.429	-0.019
POS_12	0.033	0.013	0.403	0.009	0	0.019
ST3G_G	0.164	0.046	-0.005	-0.017	-0.036	-0.48
G_12_13_G	0.234	-0.001	0.052	0.211	-0.007	-0.136

Explaining the dynamic factors in terms of the physically known variables is the key aspect of the development of a generic spatiotemporal model. Most physically known variables of the water budget that dictate groundwater level include soil moisture, rainfall and surface water level (Ritter and Mun`oz-Carpena, 2006). Assumption was made that evapotranspiration is embedded in the soil moisture storage estimation and hence it is an already reduced factor. Accordingly, exploratory factor analysis was done in a stepwise regression of monthly averaged groundwater level against monthly averaged values of rainfall, soil moisture, and canal water level. Factor selection was on the basis of the condition of the AIC (Table 4.5). The analysis was conducted on three hydrogeologic areas of the Miami Limestone Surficial aquifer system represented by G1213-G station, the Tamiami surficial aquifer system represented by ST3G station and the unidentified surficial aquifer system represented by POS-12 station. Explanatory variables for three of the hydrogeologic were found different suggesting a stronger filtering effect of the subsurface hydrogeology (Figure 4.5). The explanatory variables influencing groundwater level fluctuation over Miami limestone area (station G1213_G) are only the rainfall and canal water level suggesting that the subsurface layer is serving as a conduit which has no delaying effect by soil moisture (Figure 4.5 and Table 4.5).

An extreme opposite subsurface profile in the peat overlying Tamiami formation at station ST3G showed that only the deepest moisture profile and canal water level are explanatory variables. In other words, neither rainfall nor top soil moisture profiles could explain to the groundwater level change (Figure 4.5 and Table 4.5). The rainfall seems to infiltrate less and rather accumulate as runoff causing little direct effect on groundwater level rise, as observed on ST3G station of Figure 4.5 and Table 4.5 which showed no

direct relationship of rainfall and groundwater level fluctuation. Similarly, the low effect of moisture at 10 cm and 40 cm on groundwater suggests that moisture is either retained and hence its effect is lagged for longer period (more than a month), or during such elongated time it is mostly siphoned as evapotranspiration.

For the unconsolidated sandy region (Station POS-12), rainfall is identified as one of the factors attributable to the higher infiltration rate for the medium sandy top layer (Figure 4.5 and Table 4.5). Also the moisture at the top and deepest layer are amongst the identified variables suggesting that infiltration at the surface and percolation at the deepest layers are key factors while the middle vadose zone is serving only a as a fast draining medium.

In all the hydrogeological regions, the surface water level is identified as the major variable dictating the groundwater level at all sites. Rainfall amount and soil moisture are important explanatory factors, although the weight of the explanatory factors vary spatially.
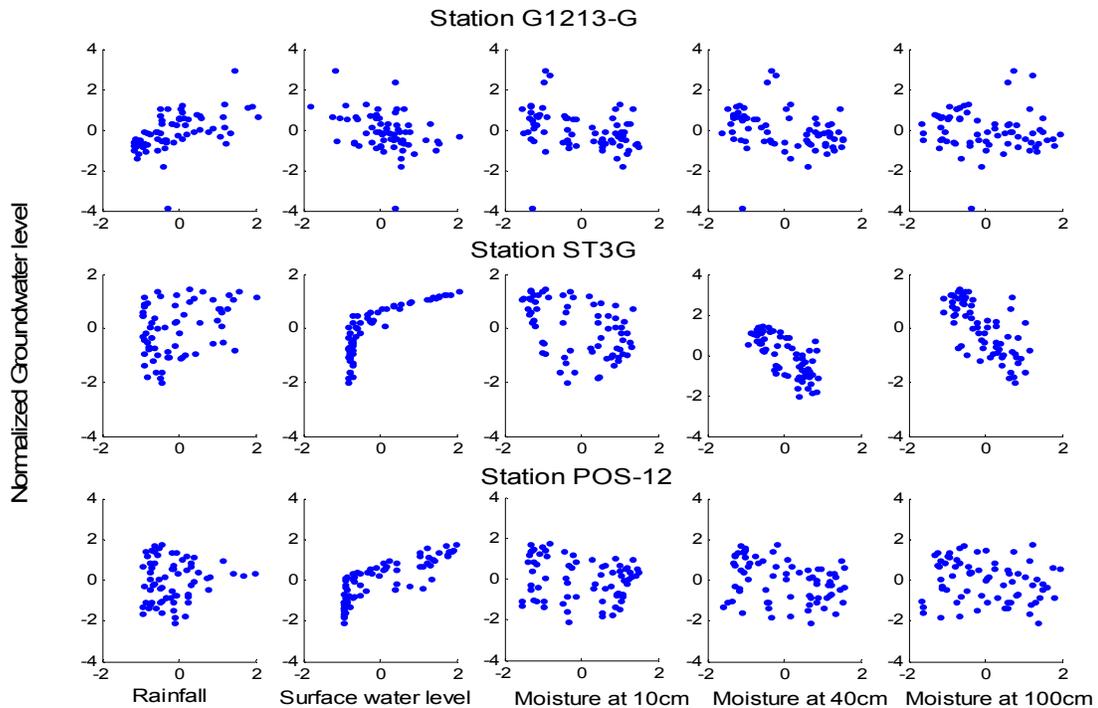
Figure 4.5 Scatter plots of groundwater level against rainfall, surface water level, and soil moisture at different stations

Table 4.5 AIC values of the stepwise regression

Hydrogeology	Station	Rain	θ_{10cm}	θ_{40cm}	θ_{100cm}	Stage	AIC (a)
Tamiami formation,	ST3G				X	X	-101.2
Miami limestone	G1213_G	X				X	-28.94
Undifferentiated,	POS12	X	X		X	X	-84.27

The relationship of the dynamic factors obtained from DFA and possible explanatory variables was established fitting normalized and parameterized time series of regional average rainfall, regional average soil moisture and surface water level. Results showed that the three dynamic factors were found related to three surface water bodies namely: the Lake Kissimmee water level in the north (Figures 4.6a and 4.6b); the Caloosahatchee River water level in the west (Figures 4.6e and 4.6f); and the Hillsboro canal water level in the east (Figures 4.6i and 4.6j) that drains from Lake Okeechobee to

the east coastal area. The two other factors identified from the DFA were explainable by the regional averaged rainfall (Figures 4.6m and 4.6n), and regional averaged soil moisture (Figures 4.6q and 4.6r). The curve fitting of the dynamic factors against the explanatory variables show similar trends at all seasons. Only one factor was found unexplainable (Figures 4.6u and 4.6v) and assumed as latent factor (Zurr et al., 2003). From the point of view of the water balance approach, water abstraction would partially explain the latent factor, though not tested for lack of data.

The spatial association of the loadings to the explanatory factors was also found in agreement with the exploratory factor analysis (Figure 4.5) earlier. The analysis showed a localized higher weight in the north as shown in Figures 4.6c and 4.6d, attributable to factor 1 (Lake Kissimmee water level); higher weight in the south, Figures 4.6g and 4.6h, attributable to factor 2 (Caloosahatchee River water level); and discernible higher local weight in the east, Figures 6k and 6l, attributable to factor 3 (the Hillsboro canal water level). The effect of factor 4 (rainfall average) is spatially variable and event dependent, and hence assumes the loading spatial variability accordingly as shown in Figures 4.6o and 4.6p. The effect of factor 5 (soil moisture) in most part of the region is observed shifting with seasonal changes as shown in Figures 4.6s and 4.6t.

The dynamicity and spatiotemporal implicit nature of the loadings was observed from the changes over the seasons confirming the applicability of Equation (4.5) after Gamerman (2010). The increased loading of the soil moisture from east (Miami limestone area) to west as observed on Figures 4.6s and 4.6t suggests the lagging effect of the overlying clayey soil on recharge. The higher loading of most factors in the

extreme southern part supports the fact that the area is under high water table over most

seasons of the year.
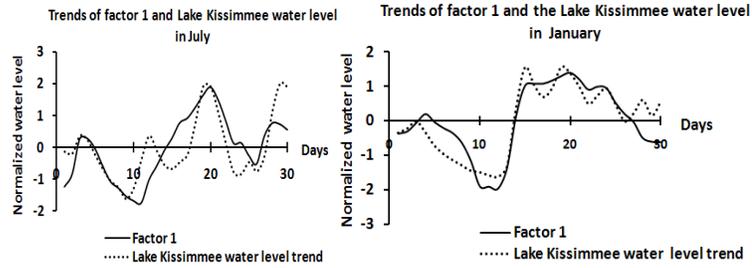
Figure 4.6 a Factor 1 Fit – July, 4.6b Factor 1 fit - January
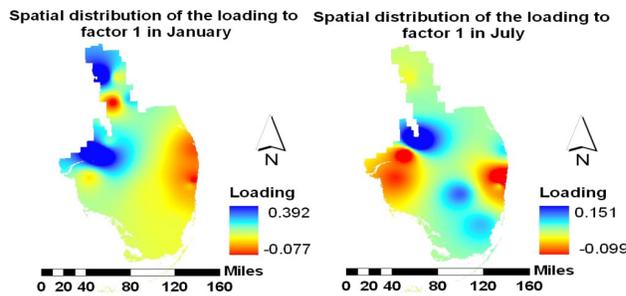
Figure 4.6 c Factor 1 Loading - January, 4.6d Factor 1 loading –July
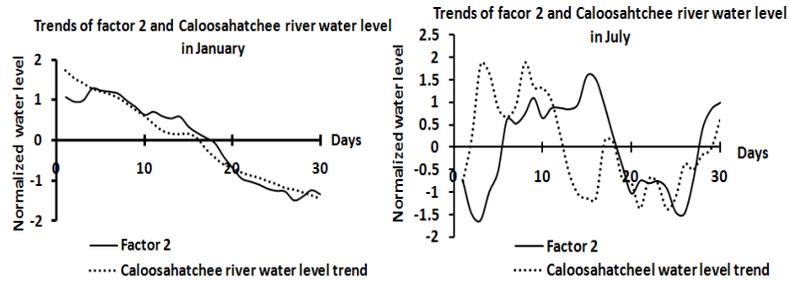
Figure 4.6 e Factor 2 fit- January, 4.6f Factor 2 fit - July
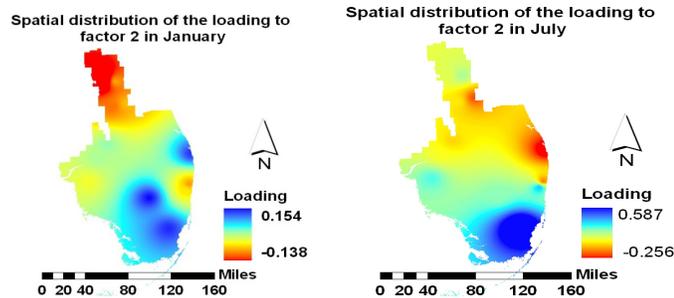
Figure 4.6 g Factor 2 Loading – January, 4.6h Factor 2 Loading – July
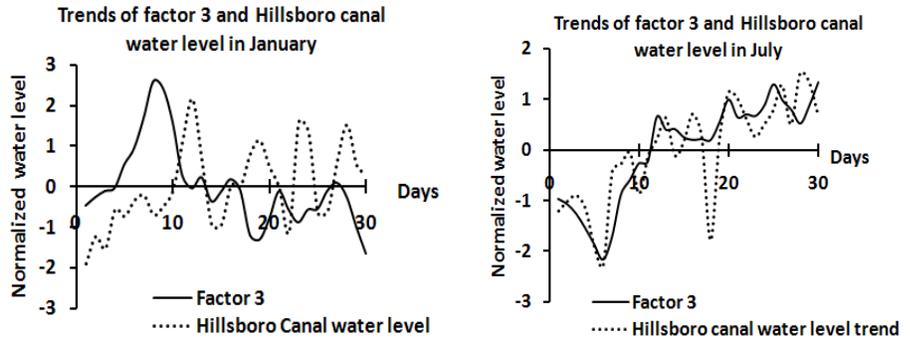
Figure 4.6 i Factor 3 fit – January, 4.6j Factor 3 - July
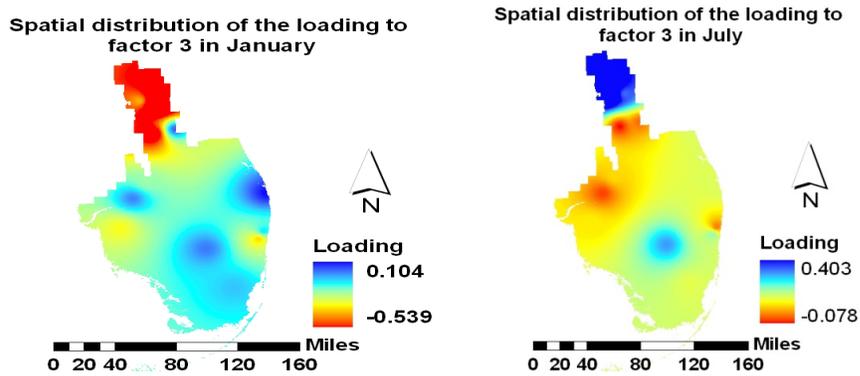
Figure 4.6 k Factor 3 Loading – January, 4.6l Factor 3 loading - July
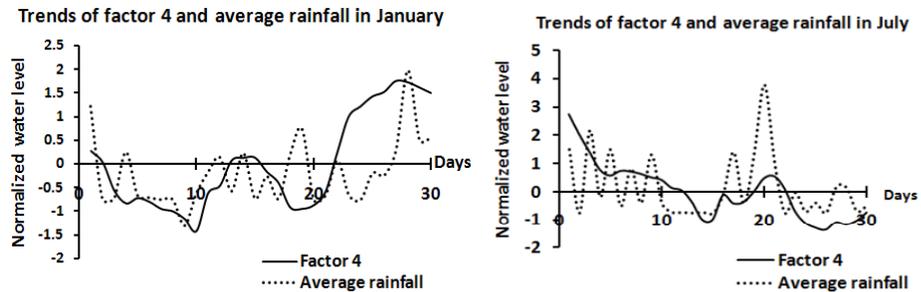
Figure 4.6 m Factor 4 fit – January,   4.6n Factor 4 fit - July
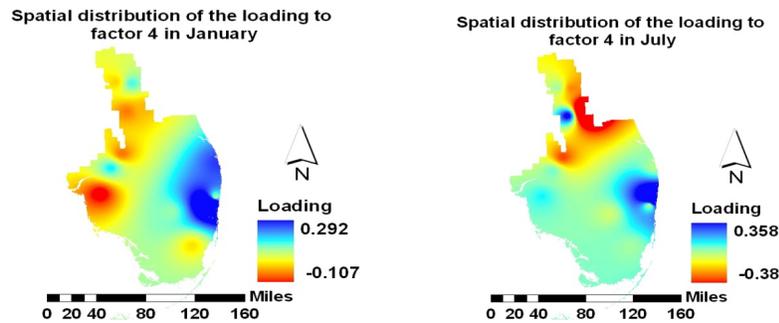
Figure 4.6 o Factor 4 Loading – January,   4.6p Factor 4 Loading - July
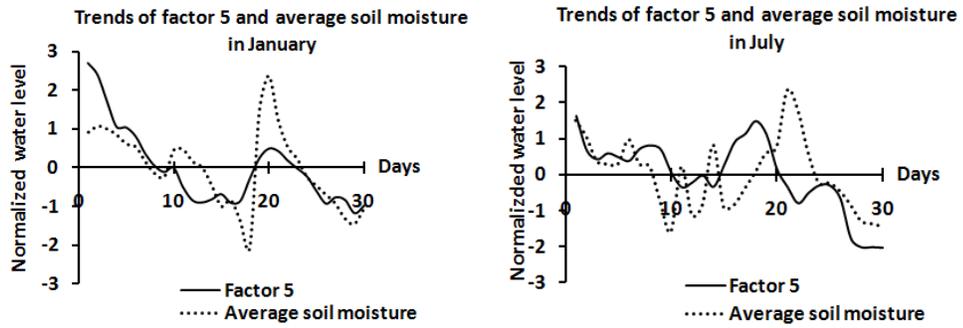
Figure 4.6 q Factor 5 fit – January, 4.6r Factor 5 fit - July
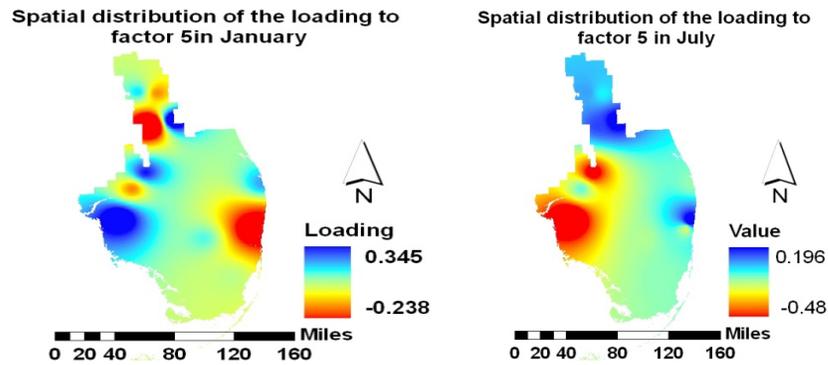
Figure 4.6 s Factor 5 loading – January, 4.6t Factor 5 Loading
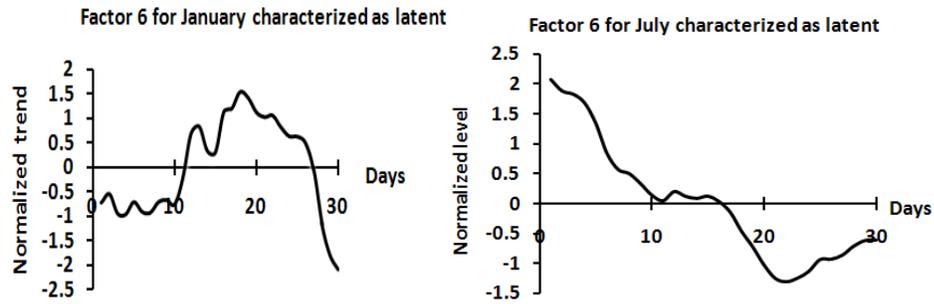
Figure 4.6 u Factor 6 fit – January, 4.6v Factor 6 fit - July
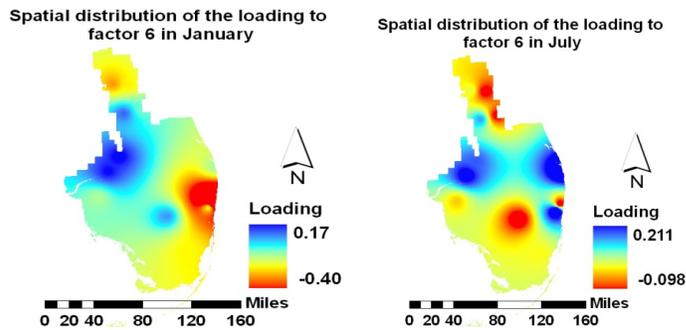
Figure 4.6 w Factor 6 loading – January, 4.6x Factor 6 Loading -July

4.8. Simulated Groundwater level

Convolving the dynamic factors with the spatially interpolated loadings, a regional scale spatiotemporal groundwater level was simulated. Cases were observed for four arbitrary dates of January 15, January 30, July 15 and July 30 as shown in Figures 4.7a, 4.7b, 4.7c and 4.7d, respectively, all from year 2004. The results show those highest water levels in the east, west and northern part, seemingly a result of the management effect. The lowest drop of groundwater level below the monthly average, on January 30 (drier month), was -60.42 cm in the east coast as shown in Figure 4.7b. The pattern could be traced for each day of the drier months and observe risks of oceanic water intrusion during the dry season. Similarly, the maximum groundwater rise on July 30 2004 (Figure 4.7d) was 67 cm above the monthly average and such trends could be traced for the wettest months to oversee flooding risks. The validation conducted on three wells, not included in the simulation, show that the maximum RMSE is 17cm.

Figure 4.7 Predicted groundwater level
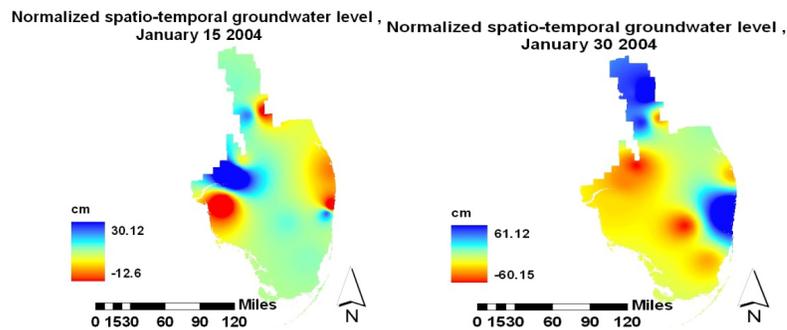
(4.7 a) 15 Jan 2004, (4.7b) 30 Jan 2004
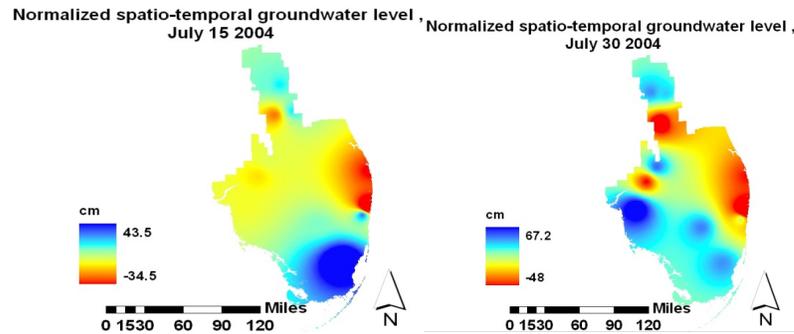
(4.7c) 15 July 2004, (4.7d) July 30 2004

4.9. Conclusion and Recommendation

The spatiotemporal model was developed extracting dynamic factors and their respective loadings from time series observations of groundwater level using a dynamic factor analysis (DFA) method, as described in Zhurr et al. (2003). The DFA served to extract common latent factors at regional scale together with their loadings in which the latter allow incorporation of the spatial structure for a spatiotemporal modeling of groundwater levels. The extracted factors were related to explanatory variables to explain the simulation.

The results of the analysis showed that the surface ground-water interaction in the SFWMD is governed by regional averaged rainfall, regional averaged soil moisture, Lake Kissimmee water level in the north, Caloosahatchee River in the west and Hillsboro canal water level in the East. Only one factor was found latent (hidden) and presumed to be related to the water extraction in which case daily data was not available for analysis. Further research on data acquisition of subsurface water utilization would be vital to observe its relationship with the latent factor.

A regional water level was simulated using a combination of the factors and interpolated loadings at regional level. The validation conducted at arbitrary regional

points of well sites result shows that a sound agreement exists between the observed groundwater level and the simulated water level with maximum RMSE of 0.12m. The model in combination with the point scale groundwater prediction model (Chebud and Melesse, 2011) would serve for operational prediction and decision making.

Spatiotemporal forecasting requires factor reduction to increase its efficiency. On the other hand, forecasting requires identification of explanatory as well as latent factors that are often embedded in observations. A DFA is a powerful tool to retrieve such dynamic factors which are not necessarily independent and hence could have an embedded lagging effect.

The availability of data in south Florida offers an opportunity to test applicability of regional spatiotemporal surface-groundwater interaction modeling and forecasting. The DFA analysis indicated that the three well known surface water levels namely, the Lake Kissimmee water level, the Caloosahatchee River water level, and the Hillsboro canal water level as well as the regional averaged soil moisture, and rainfall are key to the regional spatiotemporal groundwater level modeling. Apparently, the ease of any operational prediction could be achieved as long as data is obtainable 'economically'.

For regional spatiotemporal surface-groundwater operational modeling in SFWMD, nearly all explanatory variables are obtainable from satellite imagery (cheap and real-time). Rainfall is measured both on the ground as well as using radar; soil moisture is currently available from programs such as GLDAS. Also surface water levels are captured by satellite imagery (TOPIX) in the case of lakes. In extreme cases, such as canals and river water levels monitoring networks as in SFWMD could support the effort.

Employing DFA over SFWMD and testing its workability is viable given its low computational intensive nature and possible readjustment of dynamic factors through rigorous analysis. The study concludes that DFA is an efficient tool for its ease of synthesis of regional spatiotemporal groundwater level and also capability of forecasting using minimum parameters at reasonable accuracy. Further study into the inclusion of water extraction as part of the explanatory variables should help for decision making.

In summary, spatiotemporal modeling was achieved at regional scale in a reasonably flat topography, where watershed based modeling was less applicable and direction of flow is subjected to be disrupted by karst hydrogeology. Apart from synthesizing the regional scale water level dynamics, the modeling approach would help for predictions of spatiotemporal groundwater level fluctuation coupled with the point scale dynamic forecasting as reported by Chebud and Melesse, (2011).

Acknowledgment

REFERENCES

Aggarwal Y.P. (1998). Statistical Methods Concepts, Application and Computation, 3rd Edition, Sterling Publishers Pvt.Ltd, New Delhi, PP387

Claeskens G and Hjort N L. (2008). Model Selection and Model Averaging: Cambridge Series in statistical and Probabilistic Mathematics, Cambridge University Press, UK pp 305

Chebud Y A and Melesse A M.(2011) Operational Prediction of Groundwater Fluctuation in south Florida Using Sequence Based Markovian Stochastic Model, Water Resour Manag. 25(9): 2279-2294, DOI: 10.1007/s11269-011-9808-z

Gamerman D. (2010). Dynamic Spatial Models Including Spatial Time Series: Handbook of Spatial Statistics, Edited by Gelfand A E., Diggle P J, Fuentes M, Guttorp P. Chapman & Hall / CRC Boston, USA.

Davis J C, (2002). Statistics and Data Analysis in Geology, 3rd Edition, John Wiley & Sons, New York, pp 638

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

Gunderson L H. (2001). South Florida: The reality of change and prospects for sustainability; Managing surprising ecosystems in southern Florida, Ecological Economics 37(2001) 371-378

Harvey J W., Krupa S L. and Krest J M. (2004). **G**round water Recharge and Discharge in the central Everglades, *Vol. 42, No. 7 –GROUND WATER-Oceans Issue* p (1090-1102)**.**

Hoogland T, Heuvelink B.M.G, Knotters M. (2010). Mapping Water-table depths over time to assess desiccation of groundwater-dependent ecosystems in the Netherlands, Wetlands 30(2010)137-147

Knotters M and Bierkens M F P. (2001). Predicting water table depths in space and time using a regionalized time series model, *Geoderma* 103(2001) 51-77

Lopes F H, Salzar E, Gamerman D. (2008). Spatial dynamic factor analysis, Bayesian Analysis 3(4)759-792

Ma'rkus L, Berke O, Kovacs J, Urfer W. (1999). Spatial prediction of the intensity of latent effects governing hydrological phenomena, Environmetrics 10(1999)633-654

Nayak P C, Rao S.Y.R, Sudheer K.P. (2006). Groundwater level forecasting in shallow aquifer using artificial neural network approach, water resources management 20 (2006)77-90

Park E and Parker J.C. (2008). A simple model for water table fluctuations in response to precipitation, Journal of Hydrology 356(2008)344-349

Ramsey F. L .and Shafer D.W. (2002). A Statistical Sleuth: A Course in Methods of Data Analysis, 2nd edition, DUXBURY; Thomson Learning, California USA pp 698

Ritter A , and Mun'oz-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 (2006) 340-354

Visser A, Stuurman R, Biierkens F.P M. (2005). Real-time forecasting of water table depth and soil moisture profiles, *Advances in Water Resources* 29(2006)692-706

Zurr A F, Pierce G.J. (2004). Common trends in northeast Atlantic squid time series, Journal of Sea Research 52(2004)57-72

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

CHAPTER V

5. Phosphorus run-off assessment in a watershed

5.1. Abstract

The Watershed Assessment Model was used to simulate the runoff volume, peak flows, and non-point source phosphorus loadings from the 5870 km^2 Lake Okeechobee watershed as a case study. The results were compared to on-site monitoring to verify the accuracy of the method and to estimate the observed/ simulated error. In 2008, the total simulated phosphorus contribution was 9634, 6524 and 3908 kg (P) y^{-1} from sod farms, citrus farms and row crop farmlands, respectively. Although the dairies represent less than 1% of the total area of Kissimmee basin, the simulated P load from the dairies (9283 kg (P) y^{-1} in 2008) made up 5.4% of the total P load during 2008. On average, the modeled P yield rates from dairies, sod farms and row crop farmlands are 3.85, 2.01 and 0.86 kg (P) ha^{-1} y-1, respectively. The maximum sediment simulated phosphorus yield rate is about 2 kg (P) ha^{-1} and the particulate simulated phosphorus contribution from urban, improved pastures and dairies to the total phosphorus load was estimated at 9%, 3.5%, and 1%, respectively. Land parcels with P oversaturated soil as well as the land parcels with high phosphorus assimilation and high total phosphorus contribution were located. The most critical sub-basin was identified for eventual targeting by enforced agricultural best management practices. Phosphorus load, including stream assimilation, incoming to Lake Okeechobee from two selected dairies was also determined.

Key words: Phosphorus, WAM, Lake Okeechobee, Kissimmee basin, Sub-basin

5.2. Introduction

Water quality degradation in many watersheds caused by phosphorus and nitrogen-laden runoffs from farmed fields is known to be responsible for significant ecological changes throughout entire ecosystems. Freshwater as well as marine ecosystems are

carelessly damaged by the introduction of increasing levels of nutrients leading to eutrophication. Phosphorus (P) levels in Lake Erie (Canada and USA) have been steadily increasing since the 1990s causing severe blue-green algae blooms, Schwab e al. (2009). The 730 square-mile Lake Okeechobee (LO) ecosystem (south Central Florida, USA) has become more eutrophic and less efficient at retaining nutrients because of its exposure to higher nutrient levels from agriculture and urban activities within its watershed (Breaonik et al., 1998; Redfield et al., 1998; Richardson et al., 2007). The total maximum phosphorus load (TMDL) for LO has been established at 140 mtons y^{-1} (equivalent to 383.5 kg day^{-1}) by the Florida Department of Environmental Protection (FEDP) (FEDP, 2001). During 2004 and 2005, the monitored total phosphorus load brought into LO exceeded 930 and 830 mtons y^{-1}, respectively (SFWMD, 2008). Although best management practices (BMPs) in agriculture have been implemented in LO watershed since 1970s, the phosphorus concentration in the lake increased from ~60 mg L^{-1} in 1975 to ~210 mg L^{-1} in 2008 (Zhang et al., 2009). Research on agricultural management practices that aim to reduce phosphorus in runoff from agricultural land, dairy farms and residential areas has been hampered by the need to study large watersheds over relatively long time periods to account for both the temporal and spatial effects of scale. The diffuse

non-point phosphorus inputs derived from many sources makes it difficult to manage its levels when compared to point source pollution.

5.3. Environmental impact

A novel environmental modeling and analytical method was presented to assess the non-point source phosphorus loadings from a large watershed draining into a lake and the nutrient transport through the stream network. Specific locations where elevated levels of non-point source phosphorus may be expected were pointed out. The method has broad applicability for the assessment of nutrient and non-point source loadings into large shallow lakes where monitored data may be lacking. It could also be used in the control of non-point source discharges so as to facilitate the meeting of regulatory rules.[i]

Lake Okeechobee watersheds are characterized by surface water and groundwater flows, most of the time occurring in an integrated fashion because of the relatively low land surface elevations coupled with shallow water tables. Significant degradation of the surface water quality originates from non-point source polluting discharges, in addition to direct surface water discharges (Donigian et al., 1995). For many non-point source pollutants, measurement of the discharges is not technically or economically feasible and monitoring non-point source loadings is particularly difficult and expensive. Several spatial computer models using the geographic information system (GIS) have been developed to assess the water quality (for point and non-point sources) and nutrient transport on the basis of land-use conditions, soil type, topography, hydrology, and other factors (Morse et al., 1994).

The objective of the present work is to assess the phosphorus loading from the Kissimmee basin, north of LO, using the Watershed Assessment Model (WAM), a water

quality/ hydrology model with a GIS interface. The predictive capabilities of the WAM model with respect to runoff volume, peak flows, non-point source phosphorus loading (soluble and particulate) to LO from the 5870 km^2 watershed was estimated from the period between 1998 and 2008. The results have been compared to those from on-site monitoring to verify the accuracy of the method. The nutrient transport over the landscape and through a stream network was simulated for the purposes of the watershed nutrient pollution investigation. Specific locations where elevated levels of phosphorus may be expected were highlighted and specific phosphorus loadings to the lake from two selected dairies were verified. Although this study was directed specifically at quantification of P loadings from agricultural fields in a specific watershed for a single lake, results have important implications for many other lakes in similar landscapes that are facing the same problems. In addition, the cost-effective method described here has broad applicability for the assessment of nonpoint source nutrient loadings into large shallow lakes.

5.4. Experimental section: Watershed Assessment Model (WAM) description

The WAM model, schematically represented in the support document SI[1] Figure S1[1], developed and calibrated by the Soil Water Engineering Technology, Inc. (SWET: www.swet.com, Gainesville, FL), was GIS based continuous time model that operates on a daily time step (Bottcher, Hiscook and Jacobson (2002); Jacobson et al., (1998). The

[1] Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011.
http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art%20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrada.pdf

WAM model was used to assess the phosphorus loadings from the Kissimmee basin to LO. Details regarding the model are reported in the SI[2] . Briefly, the model uses a grid-based system to assess the spatial impact of existing and modified land uses on water quality and quantity. The grid cell representation allows for the identification of surface and groundwater flows and phosphorus concentrations for each cell. The BUCSHELL (Basin Unique Cell) cell model integrates three well-known subroutine modules that are identified as (1) Everglades Agricultural Area Model (EAAMOD) to model transport in high water table soils (Bottcher et al., 1998) (2) Groundwater Loading Effects of Agricultural Management Systems (GLEAMS) to simulate the nutrient transport in deeper water table areas (Leonard et al., 1987) and (3) Field Hydrology and Nutrient Movement (FHANTM for dairy runoffs and a category of special functions to capture the functionality of wetlands, mining, aquaculture and urban areas (Campbell et al., 1995). The cell models are used to simulate the hydrologic contaminant transport by first modeling the grid cell combinations of land use, soil types and rain zone. The model then ''routes'' the surface water and groundwater flows from the cells to assess the flow and phosphorus levels throughout the watershed. The dynamic routing is performed using the BLASROUTE module for the determination of the attenuation coefficients. The attenuation factors for each water quality parameter are related to a complex combination of transport, dispersion, and assimilation factors dependent on land uses and wetland

---

[2] Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011.
http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art%20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrada.pdf

types and are chosen on the basis of published assimilation rates (Brown et al., 1995; Cooper et al., 1993). The boundary conditions are related to the studied area and its outlets.

5.5. Study area and model input/output parameters

A phosphorus runoff simulation was performed to predict the soluble phosphorus and sediment phosphorus-enriched loads to the adjacent streams at a grid size of $0.01km^2$. The selected area for the present study was the Kissimmee basin within the LO watershed (Figure S2[3]) that encompasses the lower Kissimmee (S65A-E basins, 1737 km2) and the upper Kissimmee (4130 km^2) basins, including different types of land uses (Figure 5.1a) as well as 16 water quality, 21 flow and 30 rainfall monitoring stations (Figure 5.1b). The study period covered 11 years (from January 1, 1998 to December 31, 2008). Three monitoring stations are located at the area of discharge into LO (Stations S72, S71 and S65E).

The data input into the model included the land use, soil and BMP types, topography, hydrography and rainfall zones, basin boundaries, climate data, point sources and service area coverage (as detailed in Figure S1[1]). The inputs for monitored parameters such as the rainfall (from 30 stations), land use practices, phosphorus and flow monitoring data were downloaded for each of the stations from the south Florida Water Management District database that is available from the web site

---

[3] Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011.
http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art%20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrada.pdf
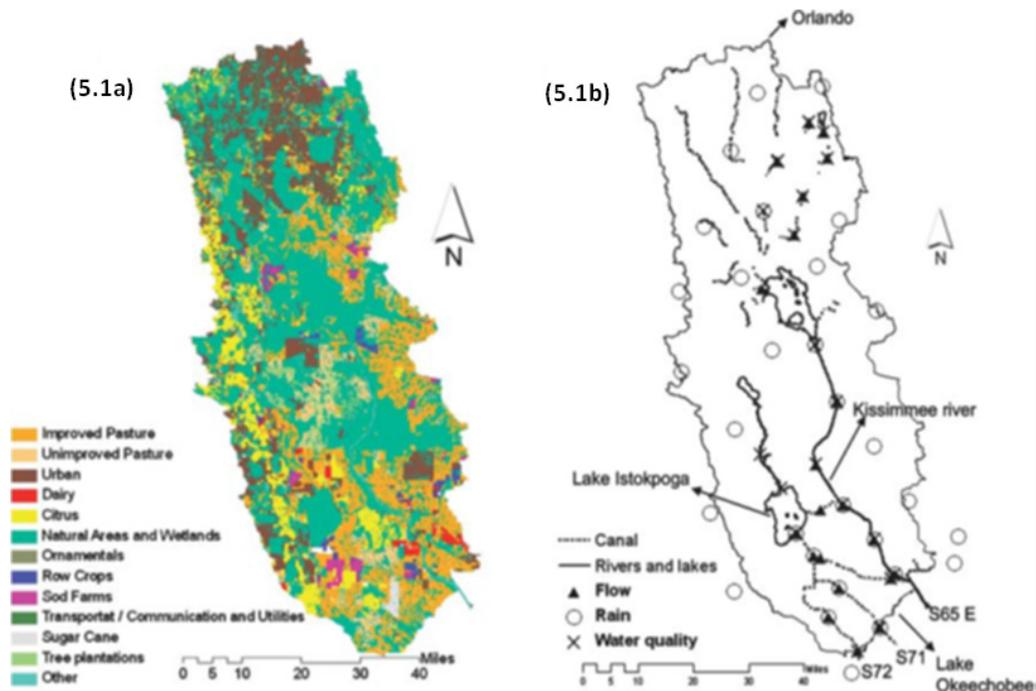
Figure 5.1 Kissimmee basin map (a) Land use distribution and location (b) Monitoring stations (flow, rain and water quality) and outlets (S65 E, S-71 and S-72)

The monitoring station data were fed into the model-the flow was measured on a daily basis and the nutrient concentration biweekly. The observed mean annual rainfall in the Kissimmee basin averaged around 1255 mm and the temperature ranged between min -1 $^{\circ}$C and max 30.5 $^{\circ}$C, with an average around 22.3 $^{\circ}$C. It is worth noting that the P loading to Lake Okeechobee mainly occurs during the wet season (June–October, characterized by heavy rainfall), thus making the non-point runoff a challenging problem from a P management perspective.

The simulated variables were a daily time series of the surface and groundwater flow, and the water quality at source cells, sub-basins and individual stream reaches. The model-simulated water quality parameters derived from the model run were the

suspended solids, sediment nitrogen, sediment phosphorus, soluble nitrogen, soluble phosphorus and the biochemical oxygen demand. The model also provided several other output data such as the ranking of land uses by load source and the comparative displays of different BMP scenarios.

The simulated P loadings were derived from daily simulated flows and concentrations. The estimation of P loads at the basin outlet is determined using a routing process combining transport, dispersion, and assimilation that, in most cases, will result in an attenuation, or decrease, in P concentration. Most of the attenuation will occur in the surface water. Once the runoff has left the source cell, it is attenuated to the streams depending on the flow rate, characteristics of flow path, flow distance and land use. Separate coefficients for different land uses and wetland types and background concentrations are stored within the WAM model (Bottcher, Hiscook and Jacobson, 2002).

5.6. Model testing and accuracy

The WAM model accuracy was tested taking into account the changes in land uses throughout the years and their effect on water quality, nutrient runoff and soil properties. The simulated time series and cumulative flow (total and deviation points) derived from water flow (peaks, base and transitions) and stage (minimum and maximum time series) were compared to the data points from monitoring stations that were analyzed for spatial and temporal patterns. A ground-truth visit was also conducted to the land parcels showing the highest total phosphorus load for a land-use adjustment. The ground-truthing was done on crop lands, dairies, citrus farms, tree nurseries, sod farms and improved pastures.

5.7. Results and discussion

Basin land use as phosphorus surface runoffs are directly linked to land uses in the region, it is important to accurately determine the land use distribution. The land use distribution considered in this work corresponded to the most recent one acquired by the South Florida Water Management District in 2006. Figure 5.2 shows the land use distribution in the Kissimmee basin watershed (Figure 5.2a) compared to the land use of the entire LO watershed (Figure 5.2b). Similar land use categories are found in the Kissimmee basin and in the entire LO watershed. The dominant land use is clearly the ''natural areas'' that respectively represent 46% and 36% of the total area in the Kissimmee basin and the LO watershed.

Improved (18%) and unimproved pastures (4%), woodland and rangeland (7%), urban (16%) and citrus (5%) represent 50% of the total land use in the Kissimmee basin (Figure 5.2a). Improved (20%) and unimproved pastures (4%), woodland and rangeland (5%), urban (11%) and citrus (7%) represent 47% of the total land use in the entire LO watershed (Figure 5.2b). Some land uses are worth noting such as dairies, sod farms or row crops because of their high phosphorus load to LO despite their relatively small surface (Bottcher, 2008).
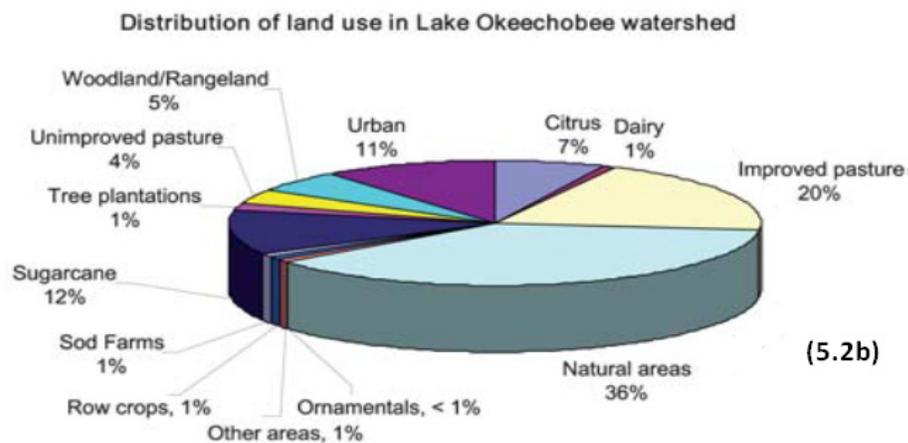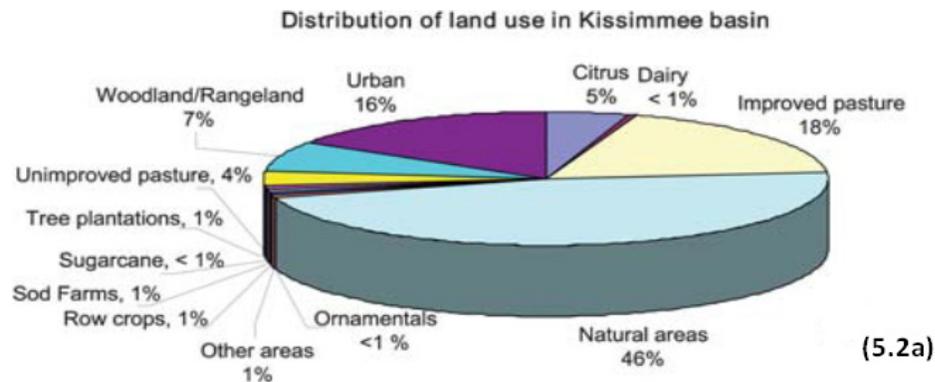
Figure 5.2 Land use surface area distribution (%) in the Kissimmee basin (5.2a) compared to the entire Lake Okeechobee watershed (5.2b).

5.8. Phosphorus load and yield rate

All the land uses described above are generating phosphorus and contributing to the total phosphorus load to LO (Cisar et al., 1992). Table 5.1 summarizes the average simulated P load from land-use classes in 2008, calculated from the daily simulated flows and concentrations. The total simulated P load (170 mtons y^{-1}) from the Kissimmee basin exceeded the TMDL (total maximum daily load) for LO by 22%. In 2008, the highest simulated contribution was 46.118 kg (P) y^{-1} and 42,468 kg (P) y^{-1} from improved pastures and urban areas, respectively. Jointly, they accounted for 52% of the total P load

coming from the studied basin. In 2008, the total simulated incoming P load was 9,634 kg (P) y^{-1} from sod farms and 3,908 kg (P) y^{-1} from row crops. Although some BMPs target agricultural areas, applications of high-phosphorus chemical and organic fertilizers to agricultural soils most often exceed the crop requirements and they are responsible for heavily phosphorus loaded runoff (Sharpely et al., 1999). Dairies also generate significant phosphorus quantities through cow manure production which is rich in P. Although dairies represent less than 1% of the total surface of the Kissimmee basin, the simulated P load from the dairies (9,283 kg (P) y^{-1} in 2008) represented 5.4% of the total LO phosphorus load during 2008.

Simulated phosphorus load contributions to the adjacent streams are presented in Figure 5.3a. Phosphorus loads to adjacent streams, driven by flows and concentrations, was not homogeneous in the basin. Natural areas were among the lowest P contributors to the streams with only <0.5 kg (P) ha^{-1} y^{-1} yield rate and <0.05 mg L^{-1} concentrations. Some lakes (Figure 5.3a) actually represented a negative load because of their assimilation of phosphorus (biota) and retention in sediments. The land uses contributing the highest total P load to the nearby streams were the dairies, row crops, sod farms and improved pastures. Table 5.1 lists the average simulated total phosphorus yield rates by region and by land use as an output of the WAM model for the Kissimmee basin in 2008. The averages were calculated from the total simulated P load by land use and by property. Generally, the simulated P load is proportional to the individual farm area (having the same stocking density). The larger the farmed or urban area, the higher the P loads originating from it as run-off. The highest phosphorus load per hectare originated from the dairies, sod farms and row crops, confirming the values represented in Figure

5.3a. As an average, the simulated P yield rates of dairies, sod farms and row crop farmlands were 3.85 kg (P) ha^{-1} y^{-1} , 2.01 kg (P) ha^{-1} y^{-1} and 0.86 kg (P) ha^{-1} y^{-1} , respectively. The urban areas in the basin include high, medium and low densities and their respective contribution of P to the streams depended on the area density (Table 5.1).

The values for the entire LO watershed (except upper Kissimmee) published by Bottcher (2008) (Table 5.1) and those in the present paper fall in the same range for many land uses. It should not be forgotten that when studying P load from farms and dairies, the loads differ depending on the basin and the region. Indeed, the surface runoff and sediment P load (or export) coefficients are highly dependent on many variables such as rain, topography, basin, tributary location, groundwater elevation and on whether a high percentage of the basin farmers are applying some BMPs to control the runoff. The quantities could be averaged on a basin or on a sub-basin level but averaging them at a watershed level would lead to a higher uncertainty because of the spatial and temporal variability leading to larger uncertainties in the load estimation. Published data by Shrestha et al. (2008) and Penn et al. (2007) demonstrated the spatial variability of P in soil across agricultural landscapes and within agricultural fields, leading to a spatial heterogeneity affecting the phosphorus loss to runoff. The heterogeneities are highlighted in Table 5.1 which also presents the simulated phosphorus load determined for S65 and for the upper Kissimmee basins. The values could be compared to the overall one for the entire Kissimmee basin. Dairies and sod farms are perfect examples. Dairies, more concentrated in the Table 5.1 The total simulated phosphorus (P) load from land use classes in the Kissimmee basin during 2008. The average simulated P yield rate kg ha^{-1} y^{-1} values by the land use and by the region (upper Kissimmee and S65) are compared to

other studies. Land uses Area/ha Total P load/kg y^{-1} S65 basin, showed a higher modeled

P yield rate (3.89 kg (P) ha^{-1} y^{-1}) when compared to the dairies in the upper Kissimmee

basin(2.43 kg (P) ha^{-1} y^{-1}). The same observation applies to the sod farms that had an

average modeled yield rate of 2.44 kg (P) ha^{-1} y^{-1} in S65 and only 1.90 kg (P) ha^{-1} y^{-1} in

the upper Kissimmee basin. The spatial heterogeneity generally tends to increase the P

loading estimation uncertainty when averaging the values on a watershed level.

Table 5.1 Total simulated phosphorus (P) load from land use classes in the Kissimmee basin during 2008. The simulated P yield rate kg ha^{-1}y^{-1} values by the land use and by the region (upper Kissimmee and S65) are compared to other studies

| Land uses | Area/ha | Total P load/kg y^{-1} | Average P yield rate/kg ha^{-1} y^{-1} | | | |
			Overall	Upper Kissimmee	S65	Bottcher, 2008
Improved pasture	108, 589	46 118 (27%)	0.42			0.43
Unimproved pasture	20 759	5556 (3.3%)	0.27			0.29
Urban	87 883	42 468 (25%)	0.48	0.46	0.43	0.39
Dairies	2408	9283 (5.4%)	3.85	2.43	3.89	2.03
Citrus	27 982	6524 (3.8%)	0.23	0.26	0.09	0.97
Natural areas[a]	273 376	40 455(23.7%)	0.14			0.12
Ornamentals	80	365 (0.2%)	4.56			2.46
Row crops	4535	3908 (2.3%)	0.86	1.78	0.45	3.79
Sod farms	4791	9634 (5.6%)	2.01	1.90	2.44	1.51
Woodland/Rangeland	47 730	1101 (0.65%)	0.02			0.16
Transportation[b]	4694	1311 (0.77%)	0.28			—
Other areas	4352	3680 (2.2%)	0.84			0.42
Total	587 179	170 405 (100%)				

[a] Natural areas include the wetlands. [b] Transportation areas include the communication and utilities. [c] P load per 2006 land use for Lake Okeechobee watershed area excluding upper Kissimmee basin—2006 values transformed to fit the last five year average load to Lake Okeechobee.

Figure 5.3 Maps of the total simulated phosphorus yield rate (a) simulated sediment phosphorus yield rate (b), and total simulated phosphorus amount sequestered in soil (c) from the different lands of the nearby streams

5.9. Phosphorus sequestration in soil and sediment transport

The phosphorus load into LO is composed of soluble phosphorus and phosphorus in particulate forms transported with the runoff that ends up in the receiving lake (Metha and Hwang, 1989) as demonstrated in Figure 5.3b. Haygarth and Jarvis (1999) elaborated on the three processes for P transfer from agricultural fields namely dissolution, incidental and physical transfers. The physical transfer is the opposite of dissolution, related to the mechanisms of detachment and soil erosion involving the physical displacement and entrainment of colloids and submicron-sized material. The particles contain phosphorus molecules, bound with soil minerals and other organic compounds, which could be remobilized when biological and/or physico-chemical conditions change. The proportion of particulate P in the total phosphorus in surface runoff varies with soil types and hydrological conditions (Uusitalo et al., 2001). Figure 5.3b shows the modeling results concerning phosphorus yield rate to the streams originating from sediment transport. The highest modeled sediment phosphorus yield rate is about 1 kg (P) ha^{-1} y^{-1} transported via rivers, streams and water bodies as seen from Figure 5.3b.

The model estimated the particulate phosphorus contribution to the total phosphorus load from urban, improved pasture areas and dairies, respectively at 9%, 3.5%, and 1%. The urban areas (south of Orlando) in the north LO watershed showed the highest phosphorus contribution through particulate transport (Figure 5.3b). Some dairies, sod farms and improved pastures that generate a high phosphorus load in a dissolved form also contribute phosphorus through particulate transport (Figure 5.3b), indicating that the soil within the land uses is phosphorus saturated (heavy leaching) with some erosion occurring. He et al. (2006) studied the transport of phosphorus through surface

runoff from agriculture lands in south Florida. Sandy soils, the dominant soil types north of LO, usually contain low levels of clay and organic matter for binding to phosphorus. There is, therefore, a good reason for concern about P losses from sandy soils in field-drains as the potential of P leaching from them is usually high (Zhang et al., 2002). The total particulate phosphorus in the runoff water reported by He et al. (2006) varied widely, ranging from 0.01 to 19.94 mg L^{-1}, with a median of 0.23 mg L^{-1}. The numbers agree with the modeled particulate phosphorus contribution observed in the present work (Figure 5.3b), whereby the highest concentration of phosphorus in the sediment was around 0.5 mg L^{-1}.

According to Graetz and Nair (1995) the maximum P storage in surface soil of the LO watershed was 2 and 0.42 kg (P) ha^{-1} y^{-1} for the areas with pastures and non-impacted areas, respectively. The present investigation determined (by modeling) the maximum of P sequestered in soil around 2 kg ha^{-1} y^{-1} (Figure 5.3c) agreeing well with the values stated above. Comparison of Figure5.3a and 5.3c indicates land parcels where the soil was oversaturated and leaching phosphorus (dark in Figure 5.3a) with low P assimilation (light in Figure 5.3c). Some of the land parcels are identified by a square in Figure 4.3c. Only a few locations (improved pastures and sod farms) contributed high phosphorus to the water bodies (dark in Figure 5.3a) and featured high phosphorus soil assimilation (dark in Figure 5.3c). Some of the land parce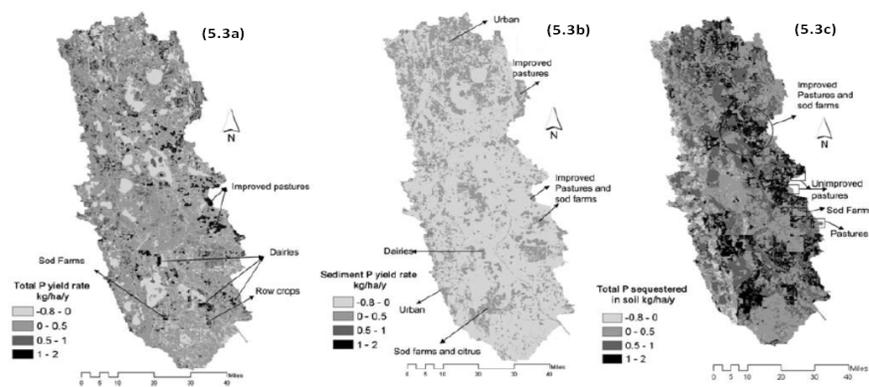ls are circled in Figure 5.3c. Figure 5.3 Maps of the total simulated phosphorus yield rate (3a); simulated sediment phosphorus yield rate (3b); and total simulated phosphorus amount sequestered in soil (3c) from the different lands to the nearby streams.

5.10. Water balance and annual phosphorus load, modeling vs. monitoring

To verify the accuracy of the WAM model (Sohrabi et al., 2003), the predicted P concentration levels from the three outlets were compared to the annual measured values (Figure 5.4). Moreover, the total daily simulated flow and P load were compared to the daily monitored values as shown in Figure S3 and S4, respectively. † Visual comparisons of measured and simulated values showed very good matches, especially during rainfall events (Figure S3 and S4 †). The yearly monitored and simulated phosphorus concentration standard deviations differed depending on the outlet. The estimated observed/simulated error was around 13.7% for S65E, 10% for S72 and increased to 31% for S71 for the 1998–2008 period. It is worth noticing that while the simulation gives the phosphorus concentration daily, the actual measured values were obtained from bi-weekly sampling by monitoring stations. As an average, there were only 115, 33 and 42 phosphorus monitoring points per year for S65E, S72 and S71 outlets, respectively (Figure 5.1b). The values for monitored daily concentrations were calculated by interpolating between the grab samples and the loads were then calculated from daily measured flows and so calculated interpolated concentrations. The technique would give reasonable estimates of phosphorus loads but with errors ranging between 25 and 50% of the actual loads (Jacobson, 2002). The average measured (monitored) load to LO from S72 was 13.8 mtons (P) y^{-1} compared to 11.5 mtons (P) y^{-1} (simulated) for the 11-year period. While S65 outlet was loading 139.5 mtons (P) y^{-1} (monitored) to the lake compared to the simulated 92 mtons (P) y^{-1}, S71 outlet was loading 47.75 mtons (P) y^{-1} (monitored) to the lake compared to the simulated 30.5 mtons (P) y^{-1}. Figure 5.4 indicates that P concentration in the runoff from S65slightly decreased during the last two

years, from S71 it remained constant, and from S72 it slightly increased. The trends are to be followed during the next few years to assess the BMP application effectiveness in decreasing the P load to LO.
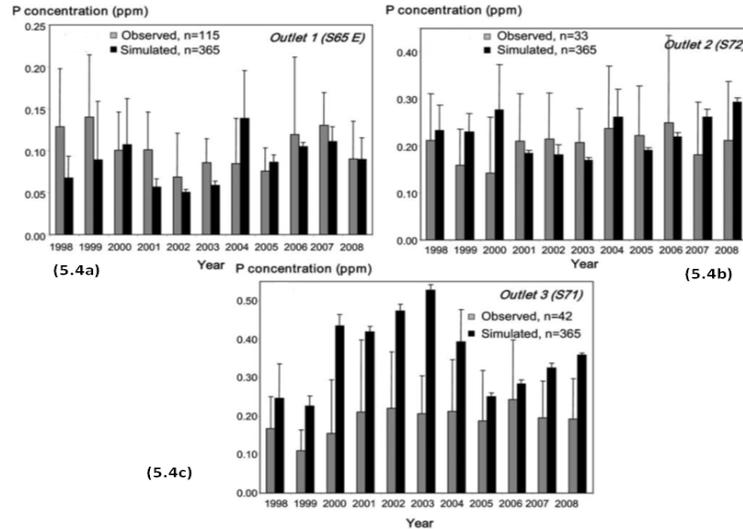
Figure 5.4 Yearly simulated/observed averaged phosphorus concentration at the three outlets S65 E (4.4a), S-72 (5.4b) and S-71 (5.4c). The number (n) of the observed or simulated sampling points is indicated on each figure.

Figure 5.5 Maps of the priority sub-basin. Land use (5.5a) and simulated phosphorus yield rate (5.5b) from the different lands to the nearby streams.

5.11. Identification and prioritization of critical sub-basins

Looking at the P-loading and soil P-sequestration maps, the most critical sub-basin identified is presented in Figure 5.5 and it encompasses a variety of land uses near Kissimmee River and Lake Istokpoga. The sub-basin was visited for ground-truthing and indeed some sod farms, row crops and dairies were identified that should be prioritized for future BMP applications. The model indicated that some of the lands in the basin, mainly dairy and row crop farms (Figure 5.5b), contributed as much as 2 kg (P) ha^{-1} day^{-1} to the nearby streams. Since their locations near the Kissimmee River, the run-off P from them will not be assimilated and will reach directly the stream and the lake. The ground-truth visit corroborated the results of the modeling whereby the non-commitment to the regulations was visually noticed. The properties should be targeted with enforced BMPs to decrease their P contribution to the streams and the lake.

5.12. Phosphorus assimilation and load to LO from two selected dairies

The phosphorus assimilation capacity of a specific land use depends on complex physical, chemical and biological parameters. The modeled loads to LO from two selected properties (Figure S5[4] dairy 1, upper watershed, latitude: 27°33.90' N; longitude: 81° 21.10' W and dairy 2, lower watershed, latitude: 27°18.80' N; longitude: 81°4.090' W) were determined knowing the routing process and estimating the attenuation of phosphorus concentration. The method was already used by Zhang et al., (2002) to

---

[4] Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011. http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art%20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrada.pdf

calculate the assimilation capacity of several sites using a simple function relating the phosphorus concentration at the outlet to the total length of transport and one single assimilation factor. A first-order decay formula (equation 5.1) was used in the present study to estimate the attenuation whereby once the runoff had left the source cell, it was attenuated to the streams related to flow rate, characteristics of flow path, and flow distance:

$$dC = (C-Cb)\ e^{(-aq-bd)} \hspace{4cm} 5.1$$

where dC is the change in concentration (mg L^{-1}), C and Cb are the current and background concentrations, respectively (mg L^{-1}), a is the attenuation multiplier, b is the attenuation exponent, q is the flow rate (m3 s^{-1}) and d is the flow distance (m). The attenuation coefficients a and b as well as Cb were used as determined in the WAM model. Values of a and b (Table S2[5]), depend on the stream conveyance feature such as wetlands, lakes, or others. The dairies 1 and 2 have a surface area of 3–3.5 km^2

each and contributed 487 kg (P) y^{-1} (0.33 mg L^{-1} as soluble P) and 1612 kg (P) y^{-1} (1.4 mg L^{-1} as soluble P), respectively, to the nearby stream-modeling values. The phosphorus in the runoffs from dairy 1 and dairy 2 reaches LO after getting assimilated

along 61 km and 16.3 km of different types of streams (freshwater marshes, shrub and brush land,.), respectively. The present calculation showed the modeled soluble P concentration originating from dairy 1 and reaching LO was 0.06 mg L^{-1} . The modeled

[5] Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011. http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art %20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrad a.pdf

soluble P concentration originating from dairy 2 and reaching LO was 0.21 mg L^{-1}. The modeling results indicated that dairy 1, discharging 487 kg (P) y^{-1} (1.6 kg (P) ha^{-1} y^{-1}) into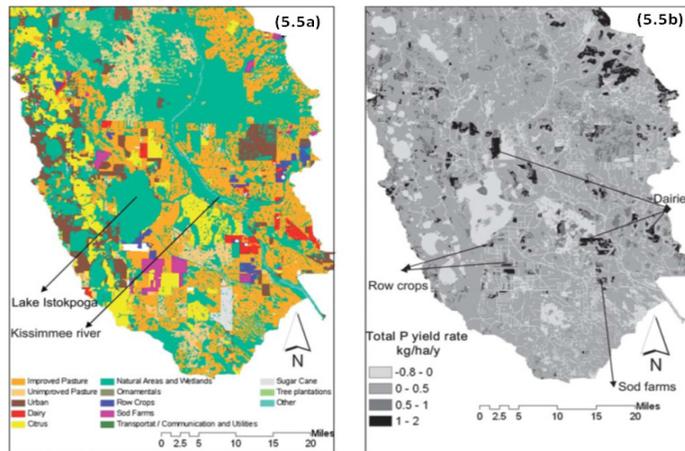 Figure 5.5 Maps of the priority sub-basin. Land use (5a) and simulated phosphorus yield rate (5b) from the different lands to the nearby streams.

The nearby stream, contributed 91 kg (P) y^{-1} (0.3 kg ha^{-1} y^{-1}) into LO whereby dairy 2, discharging 1612 kg (P) y^{-1} (4.5 kg (P) ha^{-1} y^{-1}) into the nearby stream, contributed 250 kg (P) y^{-1} (0.7 kg (P) ha^{-1} y^{-1}) into LO. The remaining phosphorus was assimilated in the farmland system and the farther away the dairy was, the more phosphorus was assimilated.

5.12. Conclusion

For many non-point source pollutants, measurements of nutrient discharges are not technically or economically feasible and monitoring non-point source loadings is particularly difficult and expensive. The method described in this paper was successful in quantifying the phosphorus loads from agricultural run-off. Specific locations where elevated levels of non-point-source phosphorus may be expected were highlighted and the phosphorus loadings from two selected dairies to the lake were estimated.

The estimation of how much each land type contributes to the total phosphorus load received by LO could be used as a basis for regulating the non-point source discharges in such a way as to ensure that the total maximum daily load (TMDL) of the lake would be met. Although the study was directed specifically toward quantifying P loadings from agricultural fields in a specific watershed for a single lake, it obviously has a broad applicability for the assessment of nutrient and non-point source loadings into large shallow lakes where monitored data may be lacking.

Acknowledgment

REFERENCES

Schwab D. J., Beletsky D., DePinto J. and Dolan D. M. (2009). A hydrodynamic approach to modeling phosphorus distribution in Lake Erie, Journal of Great Lakes Research, 35 (2009) 50–60.

Brezonik P. L. and Engstrom D. R. (1998). Modern and historic accumulation rates of phosphorus in Lake Okeechobee Journal of Paleolimnology., 20, 31–46.

Redfield G. W. (2000). Ecological research for aquatic science and environmental restoration in south Florida, Ecological Applications, 10 (4) 990–1005.

Richardson C. J.,. King R. S, Qian S. S., Vaithiyanathan P., Quall R. G. and Stow C. A. (2007). Estimating ecological thresholds for phosphorus in the Everglades Environ. Sci. Technol., 2007, 41, 8084–8091.

FEDP – Florida Department of Environmental Protection (2001).Total Maximum Daily Load for Total Phosphorus Lake Okeechobee, Florida Department of Environmental Protection, Tallahassee, FL.,

SFWMD – South Florida Water Management District (2008). Lake Okeechobee Watershed Construction Project – Phase II Technical Plan, South Florida Water Management District, Florida Department of Environmental Protection and Florida, Department of Agriculture and Consumer Services, West Palm Beach, FL.

Zhang J., James R. T. and McCormick P. (2009). Volume I: The South Florida Environment, South Florida Water Management District, West Palm Beach, FL. 2009, p. 48/84.

Donigian S. J., Bicknell B. R.and Linker L. C. (1995). Proceedings of the International Symposium on Water Quality Modeling, American Society of Agricultural Engineers, Orlando, FL., 1995.

Morse G., Eatherall A.and Jenkins A. (1994). Managing Agricultural Pollution Using a Linked Geographical Information System and Non-Point Source Pollution Model Water and Environment Journal, Vol (8) Issue 3, 277–286.

Bottcher A. B., Hiscook J. G. and Jacobson B. M. (2002). Proceedings of the Watershed 2002, Pre-Conference Modeling Workshop, Fort Lauderdale, FL., 2002.

Jacobson B. M., Bottcher A. B., Pickering N. B. and Hiscook J. G., (1998). Modeling Phosphorus Load Reductions of Agricultural Water Management Practices on a Beef Cattle Ranch, 1998. ASAE, American Society of Agricultural Engineers, (06) 982-237

Bottcher B., Pickering N. B. and Cooper A. B., (1998). Drainage in the 21st Century: Food Production and the Environment, 599–606.

Leonard R. A., Knisel W. G. and Still D. A. (1987). GLEAMS: Groundwater Loading Effects of Agricultural Management System. Trans. ASAE, (30) 1403–1418

Campbell K. L., Capece J. C. and Tremwel T. K. (1994). Surface/subsurface hydrology and phosphorus transport. Ecol. Eng., 1995, 5, 301–330

Brown M. T., Water quality and wildlife values for wetlands, Soil and Water Engineering Technology, Inc., Gainesville, FL, 1995.

Cooper A. B. and Bottcher A. B. (1993). Basin Scale Modeling as a Tool for Water Resource. Planning. Journal of Water Resources Planning and Management, J. Water Resour. Plann. Manage., 1993, 119, 306-323.

Bottcher D. (2008). Task 3 Report: Legacy P Abatement Plan, Technica Assistance in Review and Analysis of Existing Data for Evaluation of Legacy Phosphorus in the Lake Okeechobee Watershed, Soil and Water Engineering Technology, Inc., South Florida Water Management District and JGH Engineering, West Palm Beach, FL., 2008.

Cisar, J.L., G.H. Snyder, and G.S. Swanson. (1992). Nitrogen, phosphorus, and potassium fertilization for histosol-grown St. Augustine grass sod. agron. J. 84:475–479

Sharpley A., Daniel T., Sims T., Lemunyon J. (1999). Stevens R. and Parry R., Agricultural phosphorus and eutrophication, US Department of Agriculture, Agricultural Research Service-149, USA, 1999

Shrestha S., Kazama F. and Newham L. T. H., Environ. Model Software, (2008). 23, 182–194.

Penn C. J., Bryant R. B., Needelman B. and Kleinman P. (2007). Soil Sci., 172, 797-810.

Mehta A. J. and Hwang K., Fine sediment erodibility in Lake Okeechobee, Florida, University of Florida Coastal and Oceanographic Engineering Department, Gainesville, FL, 1989, p.161.

Haygarth P. M. and Jarvis S. C. (1999). Temporal variability in phosphorus transfers: classifying concentration-discharge even dynamics Adv. Agron., (66) 195–249.

Uusitalo R., Turtola E., Kauppila T. and Lilja T. (2001). Particulate phosphorus and sediment in surface runoff and drainflow from clayey soils. J. Environ. Qual., 30, 589-595.

He Z. L., Zhang M. K., Stoffella P. J., Yang X. E. and Banks D. J. (2006). Transport of heavy metals in surface runoff from vegetable and citrus fields in Florida, Soil Sci. Soc. Am. J.

He, Z. L., M. K. Zhang, D. V. Calvert, P. J. Stoffella, X. E. Yang, and S. Yu. (2004). USA Soil Sci. Soc. Am. J. 68: 1662-1669.

Zhang, M. K., He Z. L., Stoffella P. J., Calvert D. VYang., X. E., Xia Y. P.and Wilson S. B. (2004). Solubility of Phosphorus and Heavy Metals in Potting Media Amended with Yard Waste- Biosolids Compost. J. Environ. Qual. 33:373-379

Toxic/Hazard. Subst. Environmental Engineering, 2002, 37, 793–809.

D. Graetz A. and Nair V. D. (1995). Fate of phosphorus in Florida Spodosols contaminated with cattle manure , Ecological Engineering, 5, 163–181.

Sohrabi, T.M., Shirmohammadi A., Chu T.W., Montas H., and NejadhashemiM A.P., (2003). Uncertainty Analysis of Hydrologic and Water Quality Predictions for a Small Watershed Using SWAT2000. Environmental Forensics, 4:229-238.

Jacobson B. M., (2002) Hydrologic/water quality characterization of the watershed, Soil and Water Engineering Technology, Inc., Gainesville, FL., , p. 41.

Zhang J., Ray S. A. F. and Steinman A., (2002) J. Am. Water Resour. Assoc. , 38, 1613–1624.

Supporting Information (SI) for the paper Phosphorus run-off assessment in a watershed by Chebud Y, Naja M., Rivero R., 2011.
http://www.dcs.ufla.br/site/_adm/upload/file/pdf/Prof%20Marx/Aula%205/Outro%20art%20interesse/Chebud%20et%20al%202010_v%C3%A1rios%20modelos_P%20enxurrada.pdf

6. Water quality monitoring using remote sensing and artificial neural network

Y. Chebud[, a], G.M. Naja[b] *, R. Rivero[b] and A. M. Melesse[a]

[b] *Everglades Foundation, Science Department, Miami, Florida 33157, USA*

[a] *Florida International University, Department of Earth and Environment, Miami, Florida 33199, USA*

In review

6.1. Abstract

In remotely located watersheds or large water bodies, monitoring water quality parameters is often not feasible because of high costs and site inaccessibility. A cost-effective remote sensing based methodology was developed to predict water quality parameters over a large and logistically difficult area. Landsat spectral data were used as a proxy and a neural network model was developed to quantify water quality parameters, namely chlorophyll-*a*, turbidity and phosphorus before and after ecosystem restoration and during the wet and dry seasons. The results demonstrate that the developed neural network model provided an excellent relationship between the observed and simulated water quality parameters correlated at $r^2 > 0.95$ in 1998–1999 (dry and wet seasons, before restoration) and in 2009–2010 (dry and wet seasons, after restoration). Moreover, the root mean square error values for phosphorus, turbidity and chlorophyll-*a* were below 0.03 mg·L^{-1}, 0.5 NTU and 0.17 mg·m^{-3}, respectively, for all seasons at the neural network training and validation phases. Using the developed methodology, the trends for temporal and spatial changes of the selected water quality parameters were investigated. Moreover, the amounts of phosphorus and chlorophyll-*a* stored in the water column were calculated.

Keywords: neural network, Landsat TM, Kissimmee River, remote sensing.

6.2. Introduction

High phosphorus and nitrogen concentrations mainly originating from fertilizer rich agricultural runoffs and effluents from wastewater treatment plants are threatening many worldwide ecosystems (Reed-Andersen et al., 2000). Increases of the primary productivity have already been observed in several water bodies (Reddy et al., 1996). Kissimmee River in Florida is one of such ecosystems characterized by high phosphorus and nitrogen loadings originating mainly from agricultural runoffs. The situation has been aggravated by the channelization of the river in 1960, transforming the 167 km long meandering river into a wide canal only 90 km long (Koebel et al., 1999). Water quality parameters, generally used as indicators of ecosystem degradation or restoration, have nowadays been intensively monitored over the vast wetlands of the Kissimmee basin and the Kissimmee River to assess the river state. The South Florida Water Management District (SFWMD) database (South Florida Water Management District, 2010a) shows that about 1,834 sampling stations are currently monitoring the water quality status in the basin. The frequent monitoring program is necessary to capture pulses of water quality changes. However, frequent and extensive monitoring of waterbodies and wetland areas presents logistical difficulties and challenges in terms of cost and manageability (Dekker et al., 1996).

Remotely sensed data coupled with the application of the Geographic Information System (GIS) was proposed as a complementary technique to conduct dynamic monitoring and prediction of water quality parameters at operational levels (Glasgow et al., 2004). Several authors suggested that water quality data could be retrieved from satellite imagery upon training existing monitored data. Hu et al. (2004) applied a space

borne optical remote sensing technique on the shallow waterbodies of south Floridian estuaries. The same method was also used by Ritchie et al. (2003) to examine the Mississippi River basin. The studies have reported a successful determination of water quality parameters namely chlorophyll-*a* (Chl-*a*), suspended sediments (TS) and total phosphorus (TP) at multiple spatial and temporal scales. In these studies, Chl-*a*, turbidity and TP served as indicators of water quality status for the fact that the parameters can be used as a proxy for phytoplankton biomass growth rate and hence eutrophication (Wool et al., 2007).

The application of remote sensing techniques to landscape analysis and habitat monitoring has advanced in the past three decades (Schowengerdt 2007). The use of the technique in ecology and water quality monitoring is considered as providing an "out of the box" solution to resource managers (Sawaya et al., 2003). The main objective of the present work was to apply a remote sensing technique to assess and monitor the spatio-temporal dynamics of Chl-*a*, turbidity and TP in the Kissimmee River, south Florida. Although this study focused on the Kissimmee River in Florida, the technique could be generalized and applied to other water bodies. The methodology would optimize the cost of monitoring while significantly reducing the number of grab samples. The changes in water quality were assessed pre (1998-1999) and post (2009-2010) Kissimmee River restoration during both wet and dry seasons. However, the main purpose of this paper is not to evaluate the restoration impacts on water quality in the Kissimmee River but to present an alternative cost effective methodology for monitoring water quality parameters of large water bodies.

6.3. Background

Several studies reported that Chl-*a* and turbidity observations could be retrieved from the visible, mid-infrared and infrared bands (Akbar et al., 2009). Because of the characteristic color of Chl-*a*, its relationship with water radiance at 500 and 600 μm wavelengths was documented using optical sensors at laboratory scale by Ritchie et al. (2003) and at field scale by Sass and Creed (2008), Oyama et al. (2009), and Zhang et al. (2003). Ritchie et al. (2003) reported the signatures of Chl-*a* and suspended sediments and showed their higher reflectance at the green and thermal bands. The same paper indicated that the reflectance value increased when the concentration of Chl-*a* increased which was also true for the suspended sediments within the same band (Ritchie et al., 2003).

Different water quality parameters were retrieved using specific space borne sensors with radiometric resolutions such as Landsat Thematic mapper (TM) (Zhang et al., 2003), Satellite Aperture Radar (SAR-1) (Ritchie et al., 2003), (Sass and Creed 2008) and (Oyama et al., 2009), Moderate Resolution Imaging Spectroradiometer (MODIS) (Hu 2009) and Sea-viewing Wide Field-of-view Sensor (SeaWiFS) (Soto et al., 2009). The possibility of retrieving TP water quality data was documented by Shafique et al. (2006) over the Ohio River using hyperspectral bands of air-borne optical sensors at 554 nm, 564 nm, 710 nm and 740 nm wavelengths. Han and Jordan (2005) suggested the use of MODIS and Landsat TM as preferable because they are economical and available for the public use. Furthermore, MODIS has higher radiometric and temporal resolution (capturing data on a daily basis), whereas, Landsat TM has an advantage in its higher

spatial resolution of 30 m (to be compared to 250 m for MODIS) enabling to obtain observations for individual canals and rivers.

Landsat TM has seven bands of which three are in the visible range, two in the infrared and mid-infrared, and two in the thermal region. According to Akbar et al. (2009), Chl-*a* and turbidity water quality parameters are potentially separable at a defined ratio using the green (0.50-0.60 μm) and red bands (0.60-0.70 μm) of Landsat TM. The greenish characteristic of Chl-*a* driven algal system would have higher reflectance at the green band and higher absorption at the red band. In the same paper (Akbar et al., 2009), the TP levels were determined through a combination of the green (0.545-0.565 μm) and red bands (0.62-0.67 μm) of Landsat TM.

Despite knowing the water quality parameter signatures, the practical retrieval of their concentration from shallow water bodies using optical sensors is often challenging because of the background effect from the reservoir floor, embankments, rocks and sediment deposits. Hu (2009) and Zhang et al. (2003) indicated that water quality parameter retrieval methods were found somehow dependent on the waterbody class and water depth requiring multiple bands with different information content. Zhang et al. (2003) indicated the need of comparing multiple approaches namely deterministic, semi-empirical, and data driven regression analysis methods when retrieving water quality parameters. In Zhang et al. (2003), a framework for Chl-*a* data retrieval was suggested from the blue and green band combinations of the Landsat TM imagery upon calibrating an empirical relationship as shown in Equations (6.1) and (6.2):

$$\text{Chl-}a = C_o + C_k(DN_i/DN_j)^k; \tag{6.1}$$

$$\text{Chl-}a = C_o + \sum C_i\, DN_i + \sum C_j\, comb DN_j \tag{6.2}$$

where C_o, C_j and C_k are constants; DN_i and DN_j are the digital number (DN) values of the optical bands, and $combDN_j$ is the band combination (ratios or differences).

Ritchie et al. (2003) suggested the applicability of a non-linear relationship model (Equation 6.3) from the results obtained after examining several water bodies in Arkansas and Mississippi, USA.

$$R = C_o \, C^Q \tag{6.3}$$

where R is the sensor reflectance value, C_o and C are constants and Q is the water quality parameter (Chl-*a*, turbidity or TP) that could be determined upon linearizing and inverting Equation (6.3).

Brezonik et al. (2005) used a regression analysis of multiple band combinations and showed that no single band combination has uniqueness, implying the need to explore all possible band combinations. The idea of using neural network approaches was suggested by Jensen (2005), demonstrating the superiority of the method over the traditional statistical methods. To capture both linear and non-linear relationships, Sudheer et al. (2006) also suggested a neural network approach as a more flexible method to retrieve water quality parameters. The advantage of neural networks over non-linear regression approaches is attributable to the fact that the latter demands a prior knowledge of the parameter relationships while the neural network does not. The approach is vital in view of the complexity of retrieving water quality parameters affected by atmospheric and other background factors under non-ideal contexts causing uncertainties to be accounted for in Equations (6.1), (6.2) and (6.3).

The most widely used neural network model is the McCulloch and Pitt model (Hu and Hwang 2002) constituted of neurons accumulating a weighted sum of multiple inputs

called 'input function' that will later be subjected to a transfer function called 'activation function' to synthesize the outputs as represented in Figure 6.1. The net function of the McCulloch and Pitt's neuron model (Hu and Hwang 2002) is represented as a weighted linear function as indicated in Equation (6.4). Other higher order net functions could be used for complex models replacing the single input term by the product of two or more inputs.

$$u = \sum_{j=1}^{N} w_j x_j + \theta \qquad (6.4)$$

where w_j are weight parameters, x_i are network inputs, θ is the bias (threshold).
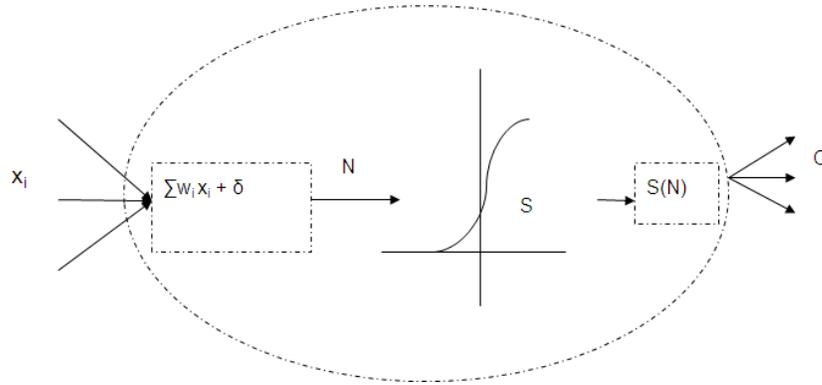
Figure 6.1 Sketch of the neural network modeling approach.

N is the input function at each neuron that gives the network to model the linear behavior; w the weight for each input x; δ is the noise which controls the activation level of the sigmoid; S is a sigmoid function as described in Hu and Hwang (2002) that serves to model non-linear relationships (Zhang et al., 2002); and O is the output.

The net output of the McCulloch and Pitt's neuron model is a nonlinear transformation of the weighted linear function u (Equation (6.4)) by a sigmoid distribution called activation function as shown in Equation (6.5).

$$f(u) = \frac{1}{1 + e^{-u/T}}$$

(6.5)

where T is a target parameter used by the sigmoid distribution to fit the output.

Other models that serve the same purpose of transformation could be linear, Gaussian radial basis, inverse tangent and hyperbolic tangent. Among the different types of activation functions, Hu and Hwang (2002) rated the sigmoid type as the most commonly used.

The objectives of this paper are 1) to develop a new cost effective technique coupling remote sensing with an artificial neural network to monitor water quality parameters of shallow water bodies and 2) to assess the effectiveness of the new methodology while examining changes in water quality in Kissimmee River during wet and dry seasons pre (1998-1999) and post (2009-2010) restoration.

6.4. Methodology

6.4.1. Study area

The selected area for this study was the Kissimmee River basin in south Florida (North of Lake Okeechobee and south of Lake Kissimmee), within the Northern Lake Okeechobee watershed (Figure 6.2). This study was conducted focusing on the lower Kissimmee River basin where the meandering river channel was straightened by a canal (C-38) in 1960 with a resulting flood plain of 16,200 hectares dried out. An important restoration strategy was developed in 1990, implemented by the SFWMD over multiple phases starting from 1999. Three construction phases are now complete, and continuous water flow has been reestablished to 37 km of the meandering Kissimmee River. According to the SFWMD report (South Florida Water Management District 2010b), the first phase (I)

backfilled 12.9 km of the C-38 canal, while the second (II) and third (III) phases backfilled 14.5 km of the canal. The fourth (IV) phase reestablished flow to 9.7 km of reconnected river channel. So far this resulted in 4,960 hectares of wetlands restored. Phases IV, I and II/III correspond to the upper section, the middle section and the downstream sections of the river, respectively (Figure 6.2). Technically, the restoration involves water release in the river original meandering route, backfilling of canals, excavating the river channel and removing some water control structures and locks. The four phased restoration cross-sections, the pre and post restoration channels and the floodplain are indicated in Figure6.2.
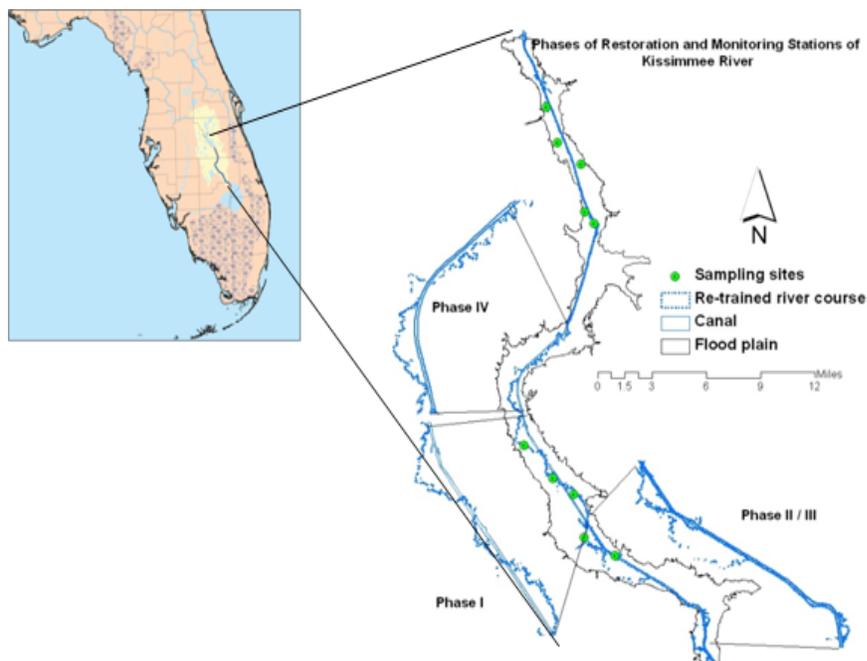
Figure 6.2 Restoration phases of the Kissimmee River

The present work has focused on four selected days during the dry (November 1 to April 30) and wet (May 1 to October 31) seasons before (dry 1998 – wet 1999) and after restoration (dry 2009 – wet 2010). The selected years are considered as average years as could be noticed when comparing the average temperature and rainfall. The average

annual temperature and rainfall during the wet season of the selected years are around 23.2°C and 228 mm, respectively. The average temperature and rainfall during the dry season of the selected years are around 16.2°C and 53.5 mm, respectively.

6.5. Data compilation and processing

The stretch from Lake Kissimmee outlet down to S65E monitoring station, constituting both the old and retrained courses of the river, was delineated from the 2006 land-use GIS database of the SFWMD using ArcGIS 9.3 suite (Figure 6.2). Upon delineation, water quality data from the monitoring stations were downloaded from the SFWMD web site (DBHYDRO database (South Florida Water Management District 2010a)), overlaid onto the river basin and selected for further data screening and collection. Accordingly, a total of 25 stations were identified along the Kissimmee River. From these stations, 6 stations have only TP (lacking Chl-*a* and turbidity) and other 7 stations were inactive since 1997 or earlier. Therefore, only 11 stations were used to capture water quality parameters namely Chl-*a*, turbidity and TP as indicated in Figure6.2. Water quality data were downloaded from the DBHYDRO database for the dates matching the satellite pass time (dry season of 1998-1999: December 15, 1998; wet season of 1998-1999: July 24, 1999; dry season of 2009-2010: November 27, 2009; wet season of 2009-2010: July 9, 2010) (Table 6.1). The water quality monitoring for Chl-*a*, turbidity and TP in south Florida are conducted biweekly at most stations, thus meeting the temporal resolutions of the Landsat TM which passes overhead every 16 days. Most of the selected water sampling dates for the dry and wet seasons of 1998 -1999 and 2009 - 2010 were within 96 hours of the satellite pass time and hence considered acceptable (Zhang et al., 2002) (Table 6.1).

Table 6.1 Water quality data obtained from stations identified in Figure 6a

	Dry season			Wet season			Dry season			Wet season		
Satellite passing date	12/15/1998			7/24/1999			11/27/2009			7/9/2010		
Water sampling	12/17/1998			7/22/1999			12/1/2009			7/8/2010		
Station Name	Chl-*a*	Turb	TP	Chl-*a*	Turb	TP	Chl-*a*	Turb	TP	Chl-*a*	Turb	TP
KREA91	2.5	1.31	0.019	30.1	2.53	0.098	1.1	32	0.048	NA	3.1	0.085
KREA92	3.5	4	0.171	14.4	2.1	0.06	0.9	35	0.034	2	2.4	0.148
KREA93	4.6	1.31	0.118	11.3	2.9	0.038	7	8.5	0.074	4	5	0.151
KREA94	2.4	0.605	0.075	72.7	3.84	0.067	8	0.073	5	5	8.1	0.175
KREA95	7.1	1.04	0.043	17.3	3.8	0.047	5	4.1	0.044	6	1.6	0.048
KREA97				3.8								
KREA98	5.4	0.714	0.034	5.7	0.761	0.025	8	10.1	0.083	4	5.6	0.138
KREA79	7.2	1.99	0.036	26.9	1.81	0.048	19	0.9	0.031	3	5.2	0.079
S65	12.16	23.7	0.151	9.6	5.25	0.096	8	1.9	0.031	14	3	0.111
S65A	NA	2.08	0.085	2.6	1.38	0.089	5	1	0.03	6	3.1	0.126
S65D	1.4	1.3	0.112	2.9	1.46	0.27	2	4	0.064	4	2.8	0.078

Chl-*a* stands for chlorophyll–*a* (mg m^{-3}), TP for total phosphorus (mg L^{-1}), Turb for Turbidity (NTU) and NA for not applicable.

The Landsat TM images provide adequate spatial resolution for regional coverage with rather negligible image distortion. The Landsat images collected from the US Geological Survey - Earth Resources Observation and Science (USGS-EROS) were already preprocessed for geometrical correction and hence did not need any further action. Kissimmee River (Florida) lies in a large wetland system where high humidity (water vapor) affects the visible spectrum. Indeed, the high humidity makes it difficult to obtain cloud-free images (Yang 2009) particularly during the wet season in Florida (Rivero et al., 2009). For water quality assessment, it is critical to use imagery without

cloud cover or haze because it can affect the spectral-radiometric responses and cause erroneous results (Olmanson et al., 2008). The images had cloud coverage and atmospheric correction was conducted using the Atmospheric and Topographic Correction software (ATCOR) implemented into the ERDAS Imagine 9.3 interface. The atmospherically corrected images were then used to capture a subset of the Kissimmee River before and after restoration.

The river has a width larger than 107 m enabling thus the capture of 3 or more pixels (each with 30 m dimension) of the imagery within the river. After overlaying the water quality monitoring stations onto the reflectance images, each pixel was read from the seven Landsat TM bands. The reflectance values from the different bands were used as inputs to train the neural network while the corresponding readings of the water quality parameters were used as targets as shown schematically in Figure 6.3. The training and simulation were conducted using the MATLAB software (www.mathworks.com).
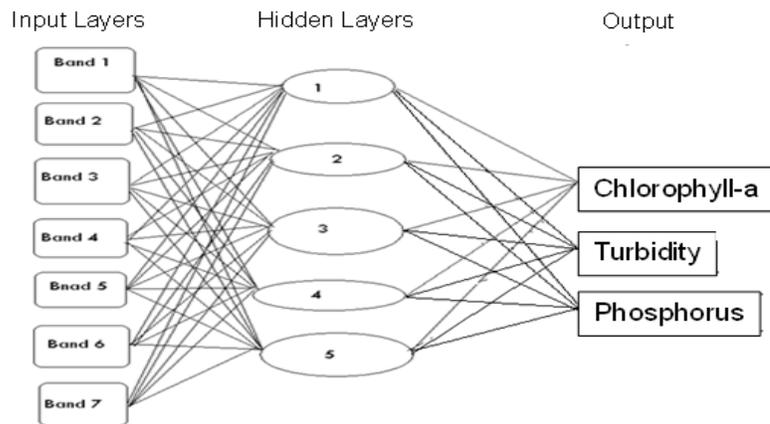

Figure 6.3 Schematic diagram of the input and output layers of ANN model

The neural network technique was applied after splitting the data into 70 % for the training and 30 % for the validation purposes. The iteration was made changing the hidden number of neurons from 3-5 while minimizing the root mean square error

(RMSE) at the training stage and testing the validation separately with a requirement setting the maximum RMSE criterion for TP less than 0.05 mg L^{-1} corresponding to the 10 % of the Maximum Contaminant Level (MCL) value published by the US-Environmental Protection Agency (EPA 2000). The Chl-*a* and turbidity maximum RMSE values were set at 10 % of the average observed data.

6.6. Results and Discussion

6.6.1. The remote sensing image quality and the neural network results

Landsat TM images have been widely used for environmental monitoring, resource management, and land cover/land use classification (Haack et al., 1987; Park and Stenstrom, 2006). The images obtained from the 2009 - 2010 wet and dry seasons showed only little residual haziness with much more improvement after haze removal and atmospheric corrections. The 1998 dry season image had cloud coverage at the northern tip portion of the river where no river restoration was conducted and that portion could thus be removed from the analysis. The image taken in the 1998 wet season contained scattered clouds (not over the river course) and the haze removal and atmospheric correction had decreased the cloud coverage in this image to 1 % considered as an acceptable value (Tucker et al., 1985).

The results from the neural network technique showed that the RMSE values for TP were below 0.03 mg L^{-1} for all seasons at the neural network training and validation phases (Table 6.2). Similarly, the average RMSE values for Chl-*a* and turbidity were 0.17 mg m^{-3} (ranging from 0.03 to 0.54 mg m^{-3}) and 0.5 NTU (ranging from 0.03 to 1.38 NTU), respectively, at the training stage and at the validation stage (Table 6.2) below the 10 % level of the observed data (0.91 mg m^{-3} and 0.51 NTU, respectively). The

correlation coefficients (r^2) obtained after graphically fitting the observed and simulated results were above 0.95. Figure 6.4 presents some of the graphical fit results for the three selected water quality parameters, indicating the relatively good agreement between the monitored and the simulated values and revealing that the present methodology can be used as a suitable technique to predict water quality data in the Kissimmee River.
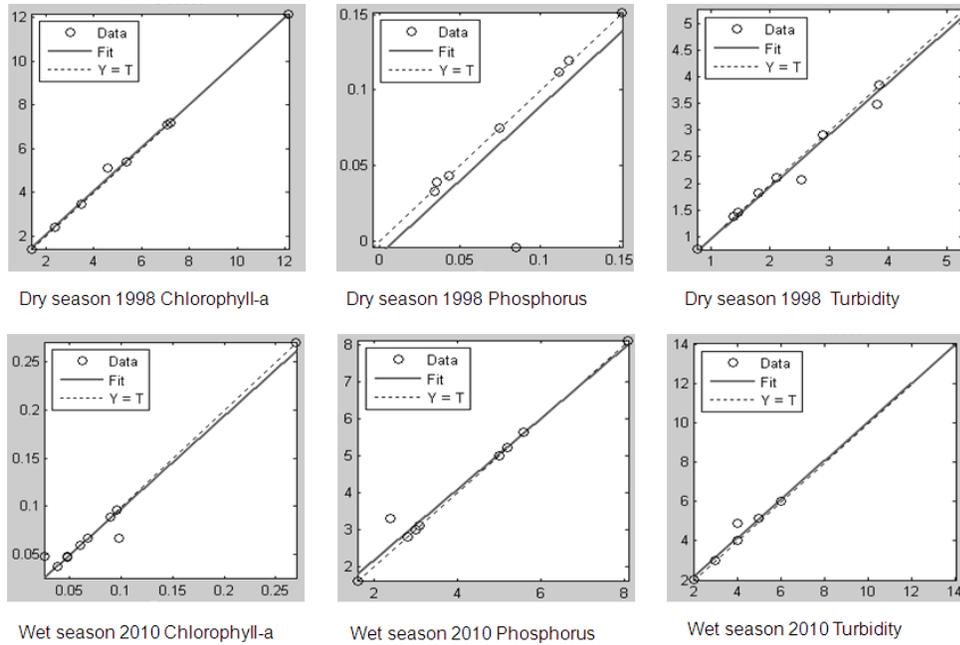

Figure 6.4 Graphical fit of the water quality parameters against 1-1 line of predicted (Y) and targeted (T) for the different seasons of 1998-1999 and 2009-2010.

Table 6.2 RMSE (m) and averaged weights of the ANN inputs

		Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	RMSE Calibration	RMSE Validation
1998 Dry	Chl-*a*	0.57	0.85	0.15	0.44	0.49	0.11	0.41	0.2	0.03
	Turbidity	0.55	0.98	0.48	0.73	0.49	0.48	0.063	0.44	1.38
	TP	0.43	0.88	0.68	0.79	0.88	0.56	0.42	0.03	0.03
1998 Wet	Chl-*a*	0.65	0.72	0.75	0.93	0.49	0.7	0.32	0.54	0.14
	Turbidity	0.03	0.2	0.0019	0.3	0.97	0.64	0.69	0.03	0.44
	TP	0.073	0.64	0.25	0.35	0.59	0.9	0.056	0.03	0.03
2009 Dry	Chl-*a*	0.073	0.064	0.51	0.65	0.96	0.39	0.13	0.22	0.14
	Turbidity	0.35	0.76	0.13	0.11	0.73	0.33	0.39	0.83	0.83
	TP	0.91	0.96	0.14	0.22	0.98	0.95	0.17	0.03	0.03
2010 Wet	Chl-*a*	0.99	0.5	0.36	0.4	0.68	0.23	0.4	0.03	0.04
	Turbidity	0.99	0.076	0.5	0.94	0.99	0.17	0.89	0.03	0.03
	TP	0.26	0.42	0.42	0.62	0.46	0.49	0.68	0.03	0.03

Chl-*a* stands for chlorophyll–*a* (mg m^{-3}), TP for total phosphorus (mg L^{-1}), Turb for Turbidity (NTU) and NA for not applicable. The unit of the RMSE is the unit of the quantity being estimated.

Several other authors attempted to employ remote sensing image data in water quality mapping. Nas et al. (2010) used the Landsat imagery to investigate spatial water quality patterns in Lake Beysehir. Their best model led to regression correlation coefficient values ranging between 0.61 and 0.71 when comparing the measured and estimated values of water quality parameters. Wu et al. (2010) attempted to develop empirical models to estimate TP concentration in the mainstream of the Qiantang River in China using Landsat data. The optimal regression model produced a correlation coefficient of 0.77. Olmanson et al. (2008) conducted a 20-year assessment to evaluate the accuracy of a Landsat method to obtain a comprehensive spatial and temporal coverage of key water

quality characteristics that can be used to detect trends at different geographic scales. In the same study, the reliability of the data was evaluated by examining the precision of repeated measurements on individual lakes within short time periods and by comparing water clarity computed from Landsat data to field-collected Secchi depth data. The agreement between satellite data and field measurements of Secchi depth within Landsat paths was strong (average r^2 of 0.83 and range 0.71–0.96).

The nonlinear neural network model enables to produce more accurate estimations than can be obtained with the more conventional methods. Especially so when only sparse, gapped data are available for model training (Abrahart and See 2000; Li et al., 2010; Srivastava et al., 2006). The main advantage of using the neural network is that it allows considering nonlinear multi-parameter relationship between reflectance from the different spectral bands and the water quality parameters (Ceyhun and Yalcin, 2010). The weight of each of the bands used in the linear part of the neural model averaged over the multiple neurons in each model is presented in Table 6.2. The results indicated that band 1 weighted the highest during the wet season in 2009 – 2010, whereas the weight of bands 3 and 7 were mostly lower when compared to other band weight. Similarly, band 2 and band 5 showed dominant weights and closer ties averaging for each parameter type over the years. Band 2 and band 5 is attributable to the fact that similar information content is captured by the different bands, and unlike other multivariate methods, the neural network would not remove redundancy (Benediktsson et al., 1990). Indeed, although each individual spectral band reveals different characteristics of the examined parameters, the dataset as a whole contains a degree of redundancy because of the correlation between the spectral bands (Kavzoglu and Mather 2000). Taking into account

this redundancy and after comparing the neural network model to a traditional regression model, Wang et al. (2008) estimated the error to below 25 % using the neural network while estimating water quality parameters such as suspended solids, dissolved oxygen, total nitrogen, TP and Chl-*a*.

6.6.2. Water quality parameters monitoring

The main purpose of the Kissimmee restoration was the reestablishment of hydrologic conditions and fluctuations while restoring the Kissimmee River and its floodplain. However, the impacts of the floodplain restoration on the water quality remain uncertain (South Florida Water Management District 2010b). It is expected that the flooding of the natural wetland zones adjacent to the Kissimmee River would impact the selected water quality parameters (TP, Chl-*a* and turbidity). Settlement of sediments, attenuation of nutrient levels and change of vegetation pattern could be expected in the long run. Colangelo (2007) examined the response of the river metabolism to restoration of flows and reported that the post-restoration dissolved oxygen of the Kissimmee River increased from < 2 mg/L to 4.70 mg/L thus meeting the target values derived from free flowing and minimally impacted reference streams. The present work focused on Chl-*a*, TP and turbidity as water quality parameters.

Chl-*a* is an important parameter considered as an indicator of a waterbody trophic status. Spatial patterns of chl-*a* tend to be complex, being a result of biological as well as physical and chemical factors. Figure 6.5a represents the levels of Chl-*a* during the four days of the dry and wet seasons of 1998 -1999 and 2009 - 2010. The modeling results showed that during 2009-2010, Chl-*a* decreased from the 5-40 mg m^{-3} level to the 0-5 mg m^{-3} level during the dry season and from the 15-40 mg m^{-3} level to the 0-15 mg m^{-3} level

during the wet season (Figure 6.5a). During the two selected days representative of the 1998 -1999 dry and wet seasons, the Chl-*a* across the river showed higher values (25-40 mg m^{-3}) at the river bank. The values decreased to 5-15 mg m^{-3} in the center of the river, agreeing with the physical process of sedimentation along the banks and algae encroachment (Sabater et al., 2008). During the two selected days representative of the 2009 - 2010 season, much of Chl-*a* was retained in the meandering section of the river course thus decreasing the Chl-*a* levels in the river downstream (Sabater et al., 2008). While Colangelo (2007) reported a general increase in the concentration of Chl-*a* pre and post-restoration, the same authors did not compare the amount of Chl-*a* stored in the water column during the pre and post-restoration periods. The results obtained from the maps generated in the present study could be further processed to obtain the amount of Chl-*a* stored in the water column as summarized in Table 6.3 (taking into account the averaged monitored water depth and the pixel area 30x30 m).

Data on discharge was not available for all the stations consistently and thus could not be considered. When examining the upper section of the river, the amount of Chl-*a* decreased by 67.5 % and 88.9 % when comparing the four selected days of the 1998 - 1999 and 2009 - 2010 dry and wet seasons, respectively. Except for the middle section of the river and when comparing the 1998 - 1999 to the 2009 - 2010 wet and dry seasons, the same decreasing tendency could be observed for the middle and lower sections of the river (Table 6.3). However, it is worth noticing that the results are on the basis of comparison between two days of the wet and dry seasons before and after restoration. A longer time period should be investigated for better establishment of the general increase or decrease tendency for Chl-*a* in the water column.

The second modeled parameter was the turbidity generally caused by the presence of suspended and dissolved matter such as clay, finely divided organic matter, plankton, other microscopic organisms and organic acids. Unlike Chl-*a*, turbidity did not exhibit a cross sectional concentration variation and showed a decreasing trend from upstream to downstream as observed in Figure 6.5b. This is attributable to the fact that Kissimmee River meanders in the flat plain which gives enough time for the sediments to settle down with a resulting decrease in turbidity toward downstream locations.

Similar to Chl-*a*, the TP concentration decreased from the bank towards the center, confirming their theoretical relationship in trophic waterbodies (Smith and Shapiro 1981) (Figure 6.5c). During the day selected as representative of the 2009 - 2010 wet season, TP relocated from the river course to the wetlands (higher concentrations in the floodplains generally considered as nutrient attenuation zones). This elevated TP levels could originate from the legacy phosphorus of the dried-up wetlands. Indeed, after channelization of the Kissimmee River, the primary land use of the drained floodplain was as cattle pasture and included also some sod farms (van der Valk et al., 2009) that would certainly leave behind an important legacy phosphorus deposits in the soil. The overall trend in TP stored in the water column presented in Table 6.3 is not conclusive since an increase and a decrease in the TP levels were observed in different sections of the river.

Figure 6.5 Modeling results of chlorophyll-*a* (6.5a), turbidity (6.5b) and phosphorus (6.5c) levels during the wet and dry seasons and before (1998-1999) and after restoration (2009-2010)
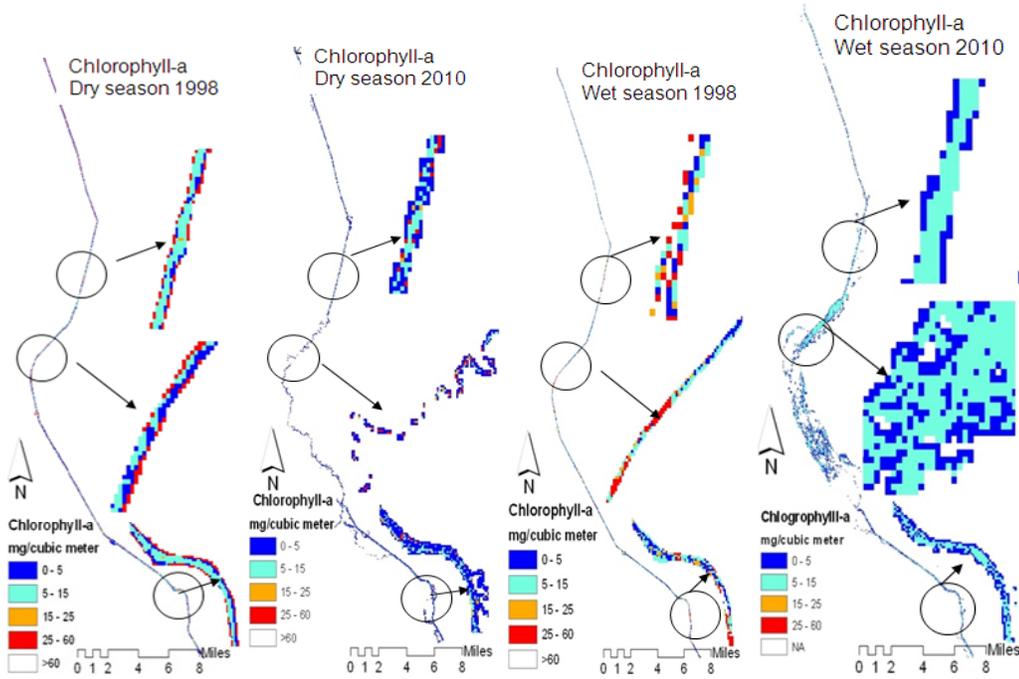
(6.5a)
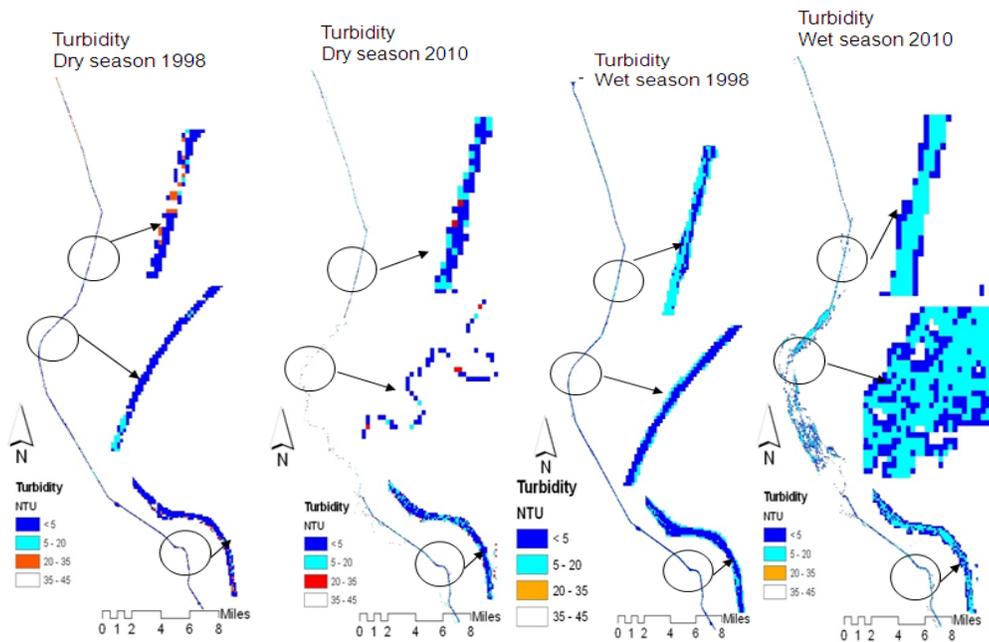
(6.5b)
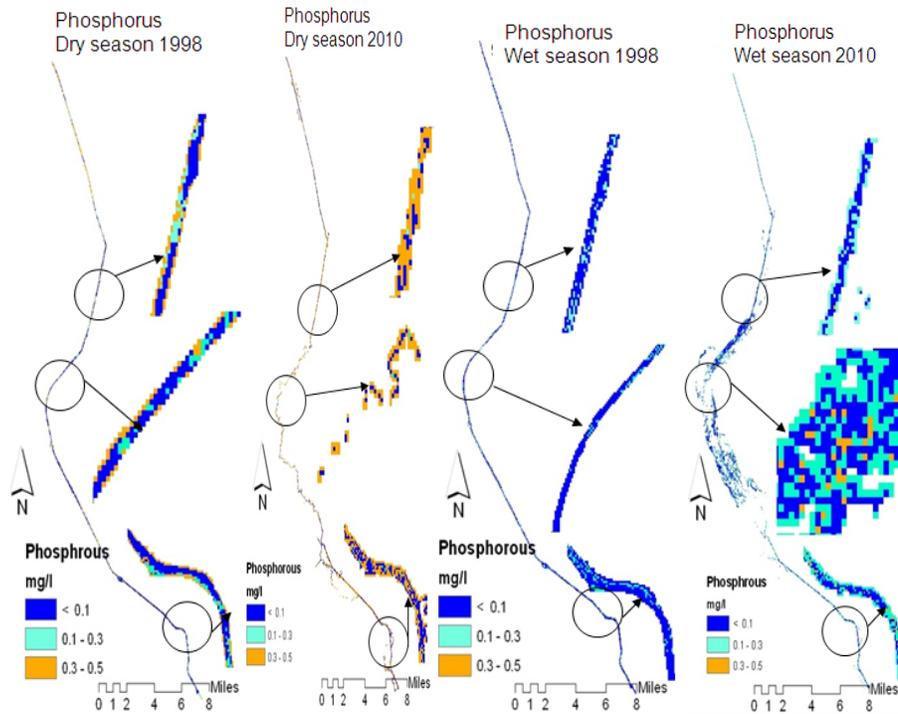
(6.5c)

Table 6.3 Percentage of increase(+) or decrease (-) in the phosphorus and chlorophyll-a amounts stored in the water column in the upper, middle and lower section of the river.

	Phosphorus (%)	Chlorophyll-a (%)
	Upper section of the River	
1998 Dry - 2010 Dry	- 78.8	- 67.5
1998 Wet - 2010 Wet	+ 7.4	- 88.9
	Middle section of the River	
1998 Dry - 2010 Dry	-99.8	- 49.8
1998 Wet - 2010 Wet	+2190	+ 59.1
	Lower section of the River	
1998 Dry - 2010 Dry	- 85.2	- 76.0
1998 Wet - 2010 Wet	- 33.5	- 94.8

116

6.8. Conclusions

Optical sensors such Landsat TM could be used as a cost effective technique to monitor water quality parameters at a basin scale. The developed neural network model reduced the uncertainty from the exclusion of any of the bands and also captured both the linear and non-linear complex relationships. The neural network model provided an excellent relationship between the observed and simulated water quality parameters correlated at r^2 > 0.95. The results showed that the root mean square error values for phosphorus, turbidity and chlorophyll-*a* were below 0.03 mg L^{-1}, 0.5 NTU and 0.17 mg m^{-3}, respectively, for all seasons at the neural network training and validation phases. The trends for changes in the selected water quality parameters were established while noticing that the results were on the basis of the comparison between two days during the wet and dry seasons before and after restoration. A longer time period should be investigated to follow the general increase or decrease tendency of selected water quality parameters in the water column, taking into account the difference in water depth and surface area before and after restoration.

Acknowledgment

REFERENCES

Abrahart, R. J., & See, L. (2000). Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes, 14*(11-12), 2157-2172.

Akbar, T. A., Hassan, Q. A. and Achari, G. (2009). A Remote sensing based framework for predicting water quality of different source waters. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 34, Part XXX (30).*

Benediktsson, J., Swain, P. and Esroy, O. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing, 28*, 540-552.

Brezonik, P., Menken, K. and Bauer, M. (2005). Landsat based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). *Lake and Reservoir Management, 21*(4), 373-382.

Ceyhun, O., & Yalcin, A. (2010). Remote sensing of water depths in shallow waters via artificial neural networks. *Estuarine Coastal and Shelf Science, 89*(1), 89-96.

Colangelo, D. J. (2007). Response of river metabolism to restoration of flow in the Kissimmee River, Florida, USA. *Freshwater Biology, 52*(3), 459-470.

Dekker, A. G., Zamurovic-Nenad, Z., Hoogenboom, H. J., & Peters, S. W. M. (1996). Remote sensing, ecological water quality modeling and in situ measurements: a case study in shallow lakes. *Hydrological Sciences, 41*(4), 531-547.

Environmental Protection Agency, US-EPA. (2000). Ambient Water Quality Criteria Recommendations, Information Supporting the Development of State and Tribal Nutrient Criteria. *EPA 822-B-00-014*. Washington, DC.

Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., & Kleinman, J. E. (2004). Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology, 300*(1-2), 409-448.

Haack, B., Bryant, N., Adams, S., & Yang, X. (1987). An assessment of Landsat MSS and TM data for urban and near-urban land-cover digital classification. *Remote Sensing of Environment, 21*, 201–213.

Han, L. H., & Jordan, K. J. (2005). Estimating and mapping chlorophyll-a concentration in Pensacola Bay, Florida using Landsat ETM plus data. *International Journal of Remote Sensing, 26*(23), 5245-5254.

Hu, C. M. (2009). A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment, 113*(10), 2118-2129.

Hu, C. M., Chen, Z. Q., Clayton, T. D., Swarzenski, P., Brock, J. C., & Muller-Karger, F. E. (2004). Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sensing of Environment, 93*(3), 423-441.

Hu, Y. H., & Hwang, J. N. (2002). *Introduction to Neural Networks for Signal Processing, Handbook of Neural Networks Signal Processing*. The Electrical Engineering and Applied Signal Processing Series: CRC Press.

Jensen, J. R. (2005). *Introductory Digital Image Processing, A Remote Sensing Perspective*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kavzoglu, T., & Mather, P. M. The use of feature selection techniques in the context of artificial neural networks. In *Proceedings of the 26th Annual Conference of the Remote Sensing Society, Leicester, UK, 2000*

Koebel, J. W., Jones, B. L., & Arrington, D. A. (1999). Restoration of the Kissimmee River, Florida: Water quality impacts from canal backfilling. *Environmental Monitoring and Assessment, 57*(1), 85-107.

Li, X., Nour, M. H., Smith, D. W., & Prepas, E. E. (2010). Neural networks modeling of nitrogen export: model development and application to unmonitored boreal forest watersheds. *Environmental Technology, 31*(5), 495-510.

Nas, B., Ekercin, S., Karabork, H., Berktay, A., & Mulla, D. J. (2010). An Application of Landsat-5TM Image Data for Water Quality Mapping in Lake Beysehir, Turkey. *Water Air and Soil Pollution, 212*(1-4), 183-197.

Olmanson, L. G., Bauer, M. E., & Brezonik, P. L. (2008). A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sensing of Environment, 112*(11), 4086-4097.

Oyama, Y., Matsushita, B., Fukushima, T., Matsushige, K., & Imai, A. (2009). Application of spectral decomposition algorithm for mapping water quality in a turbid lake (Lake Kasumigaura, Japan) from Landsat TM data. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(1), 73-85.

Park, M. H., & Stenstrom, M. K. (2006). Using satellite imagery for stormwater pollution management with Bayesian networks. *Water Research, 40*(18), 3429-3438.

Reddy, K. R., Flaig, E. G., & Graetz, D. A. (1996). Phosphorus storage capacity of uplands, wetlands and streams of the Lake Okeechobee Watershed, Florida. *Agriculture, Ecosystems and Environment, 59*(3), 203-216.

Reed-Andersen, T., Carpenter, S. R., & Lathrop, R. C. (2000). Phosphorus flow in a watershed-lake ecosystem. *Ecosystems, 3*(6), 561-573.

Ritchie, J. C., Zimba, P. V., & Everitt, J. H. (2003). Remote sensing techniques to assess water quality. *Photogrammetric Engineering and Remote Sensing, 69*(6), 695-704.

Rivero, R. G., Grunwald, S., Binford, M. W., & Osborne, T. Z. (2009). Integrating spectral indices into prediction models of soil phosphorus in a subtropical wetland. *Remote Sensing of Environment, 113*(11), 2389-2402.

Sabater, S., Artigas, J., Duran, C., Pardos, M., Romani, A. M., Tornes E.(2008). Longitudinal development of chlorophyll and phytoplankton assemblages in a regulated large river (the Ebro River). *Science of the Total Environment, 404*(1), 196-206.

Sass, G. Z., & Creed, I. F. (2008). Characterizing hydrodynamics on boreal landscapes using archived synthetic aperture radar imagery. *Hydrological Processes, 22*(11), 1687-1699.

Sawaya, K. E., Olmanson, L. G., Heinert, N. J., Brezonik, P. L., & Bauer, M. E. (2003). Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery. *Remote Sensing of Environment, 88*(1-2), 144-156.

Schowengerdt, R. A. (2007). *Remote Sensing: Models and Methods for Image Processing*. New York: Academic Press.

Shafique, N. A., Fulk, F., Autrey, B. C., & Flotemersch, J. Hyperspectral remote sensing of water quality parameters for large rivers in the Ohio River basin. In *The Second Interagency Conference on Research in the Watersheds, North Carolina, 2006* (pp. 216=221)

Smith, V. H., & Shapiro, J. (1981). Chlorophyll-phosphorus relations in individual lakes. Their importance to lake restoration strategies. *Environmental Science and Technology, 15*(4), 444-451.

Soto, I., Andrefouet, S., Hu, C., Muller-Karger F. E., Wall C. C., Sheng J., (2009). Physical connectivity in the Mesoamerican Barrier Reef System inferred from 9 years of ocean color observations. *Coral Reefs, 28*(2), 415-425.

South Florida Water Management District (2010a). DBHYDRO: http://www.sfwmd.gov/dbhydroplsql/show_dbkey_info_main_menu.

South Florida Water Management District (2010b). Environmental Report, Chapter 11: Kissimmee Basin. Available at: http://www.sfwmd.gov/portal/page/portal/pg_grp_sfwmd_sfer/portlet_sfer/tab2236037/2232010%2236020report/v2236031/chapters/v2236031_ch2236011.pdf). West Palm Beach, FL.

Srivastava, P., McVair, J. N., & Johnson, T. E. (2006). Comparison of process-based and artificial neural network approaches for stream flow modeling in an agricultural watershed. *Journal of the American Water Resources Association, 42*(3), 545-563.

Sudheer, K. P., Chaubey, I., & Garg, V. (2006). Lake water quality assessment from Landsat thematic mapper data using neural network: An approach to optimal band combination selection. *Journal of the American Water Resources Association, 42*(6), 1683-1695.

Tucker, C. J., Townshend, J. R. G., & Goff, T. E. (1985). African land-cover classification using satellite data. *Science, 227*(4685), 369-375.

Van der Valk, A. G., Toth, L. A., Gibney, E. B., Mason, D. H., & Wetzel, P. R. (2009). Potential propagule sources for reestablishing vegetation on the floodplain of the Kissimmee River, Florida, USA. *Wetlands, 29*(3), 976-987.

Wang, J. P., Cheng, S. T., & Jia, H. F. Application of artificial neural network technology in water color remote sensing inversion of inland waterbody using TM data. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXV, Beijing, China, 2008* (pp. 426)

Wool, T., Abrose, R., & Martin, J. (2007). WASP7 Multiple Algae Model Theory and Users' Guide Supplement to Water Quality Analysis Simulation Program (WAS) User Documentation. Washington, DC: U.S. Environmental Protection Agency Office of Research and Development.

Wu, C. F., Wu, J. P., Qi, J. G., Zhang, L. S., Huang, H. Q., Lou, L. P., et al. (2010). Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data. *International Journal of Remote Sensing, 31*(9), 2309-2324.

Yang, X. (2009). *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin: Springer.

Zhang, Y., Pulliainen, J., Koponen, S., & Hallikainen, M. (2002). Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment, 81*(2-3), 327-336.

Zhang, Y. Z., Pulliainen, J. T., Koponen, S. S., & Hallikainen, M. T. (2003). Water quality retrievals from combined Landsat TM data and ERS-2 SAR data, in the Gulf of Finland. *IEEE Transactions on Geosciences and Remote Sensing, 41*(3), 622-629.

CHAPTER VII

7. Spatial Relationship of Groundwater-phosphorus Interaction

Chebud, Y. A., Melesse, A.M.

Department of Earth and Environment,

Florida International University

7.1. Abstract

After the channelization of the Kissimmee River in the 1960's, the water that naturally inundates the flood plains was drained to the downstream Storm Water Treatment Areas and Water Conservation Areas for nutrient removal. The shift from a natural equilibrium ecosystem to a managed ecosystem has required operational prediction both on water and nutrient dynamics. In this research, operational prediction of soluble phosphorous that used groundwater fluctuation as a predicto was analyzed. A graphical fit of soluble phosphorus with groundwater level showed goodness of fit (R^2, 0.8 - 0.99). Also, a chi-square analysis showed similarity of pattern between groundwater fluctuation and soluble phosphorus level 99% of the time. Finally, a geographically weighted multivariate analysis of soluble phosphorus was conducted using inputs of groundwater fluctuation, water quality of adjacent surface water bodies (discharge sites) and distance from source to discharging sites. The result indicated no significant difference between the simulated and observed soluble phosphorus level at p-value of 0.01. The groundwater fluctuation was capable to explain up to 70% of the phosphorus variation. Amongst the inputs, distance to surface water bodies and their respective water quality (feedback input) showed significant weight only within a 5-mile range from the

water body. The study showed predictability of soluble phosphorus using groundwater fluctuation in the wetlands of south Florida.

Key: Multivariate analysis, Soluble Phosphorus, Total Phosphorus, Groundwater fluctuation.

7.2. Introduction and Background

Chebud et al. (2011) reported a 13% accuracy of watershed based models to estimate soluble and sequestered P in SF. However the high resolution of inputs required such as land use, management practices, soils and climatic data pose the challenge to use them for operational prediction at routine level. Moreover, in a large wetland expanse on flat topography, watershed-based analysis is neither effective nor affordable. Apparently, a methodology is necessary for spatiotemporal prediction of phosphorus associated with the groundwater level prediction and inputs from surficial water quality.

In a large wetland system, where flow directionality is dynamic, accounting all inputs is unaffordable in terms of data acquisition and computational costs (Chebud et al., 2011). The technical complexity and logistical problems of monitoring such a large wetland system is reported by Harvey et al., (2004). One option would be developing simplistic monitoring approach namely, using data driven simulation (Chebud and Melesse, 2011) supported with cheaper data acquisition techniques of remote sensing.

Therefore, the study hypothesized that soluble phosphorus could be predicted using groundwater fluctuation that embeds the intensive processes of wetting and drying; and also a feedback from the phosphorus level in surficial water bodies with no-significance difference between the observed and predicted values at 90% of the time (p-value of 0.1).

125

The specific objective of the study was set to conduct a point and spatiotemporal analysis of the groundwater-phosphorus interaction that could ultimately contribute to spatiotemporal prediction of soluble phosphorus coupled with groundwater fluctuation.

In a groundwater driven hydrological and nutrient transport systems, coupling nutrient dynamics with groundwater level prediction would be realizable given availability of a point and spatiotemporal groundwater fluctuation models (Chebud and Melesse, 2011).

7.4. Study Area

The study area for the point scale groundwater-phosphorus interaction was Everglades Nutrient Removal (ENR) project sites (Figure 7.1). The project sites are part of storm water treatment areas (STA). The sites are strategically situated in the downstream side of Lake Okeechobee Agricultural Area; on the upstream side of water conservation area 1 (WCA-1), and north of Everglades National Park (Figure 7.1). The stations chosen and availability of data are shown in Table 7.1.

Table 7.1. Phosphorus monitoring wells used for analysis

| Stations	Site/Area	Data	Type of data
ENR40215,ENR40230, ENR40245, ENR40260 ENR40315, ENR40315, ENR40315 ENR40360 ENR40415,ENR40430, ENR40445, ENR40460	Upstream of WCA-1	1993 – 1994	Soluble P
3AN1-GW1, 3AN1-GW2, 3AN1-GW3 3AN1-PZ1, 3AN1-PZ2, 3AN1-PZ3	Downstream side of LOAA	2002-2005	Total P
3AS3-GW1 , 3AS3-GW2, 3AS3-GW3 3AS3-PZ1, 3AS3-PZ2, 3AS3-PZ3	Upstream of Big Cypress	2000-2005	Total P
3BS1-GW1, 3BS1-GW2, 3BS1-GW3, 3BS1-PZ1, 3BS1-PZ2, 3BS1-PZ3	Upstream of Everglades National Park	2000-2005	Total P
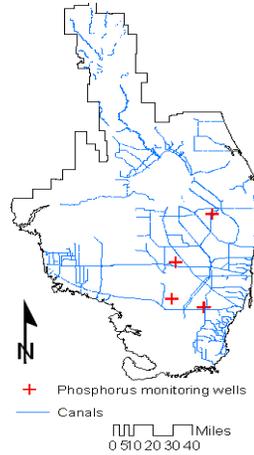
Figure 7.1 Location of phosphorus monitoring well sites over SFWMD

The study area for the spatial pattern analysis of the phosphorus dynamics against the groundwater fluctuation was tested for the S-65 sub-basin commonly called Lower Kissimmee Basin (Figure 6.2, Chapter 6). Soluble phosphorus for the S-65 sub-basin (Figure 7.2a) was obtained from soluble P of the watershed based simulated output (the simulated result of Chebud et al. (2011)). Also the groundwater level change was extracted from the simulated spatiotemporal groundwater level reported in Chapter 4 of this dissertation (Figure 7.2b).
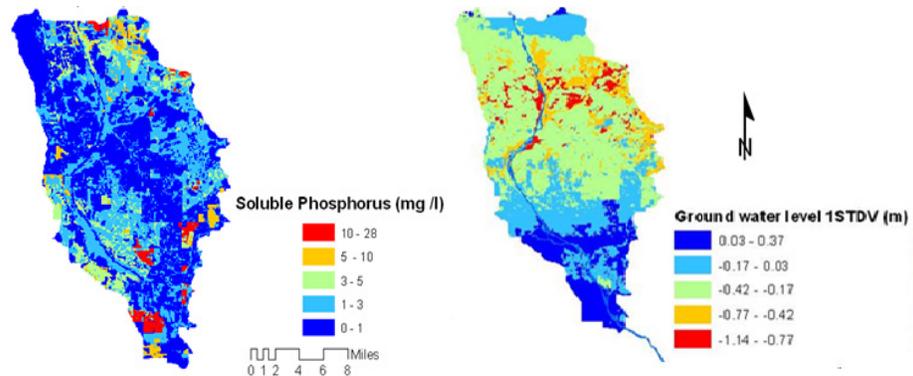

Figure 7.2 S-65 sub-basin (a) Soluble phosphorus, (b) groundwater levels

7.5. Methodology

Analysis of phosphorus dynamics against groundwater fluctuation was tested using data obtained from ENR. Data on dissolved phosphate, total phosphorus and groundwater level are collected on monthly basis and only for certain time of the year (with no time pattern). Despite the non-regular pattern of data collection, the collected data constitutes both groundwater level (measured from surface), total phosphorus and soluble phosphorus. In some of the wells, measurements are taken at groundwater depth of 0.15, 0.3, 0.45 and 0.6 m for shallow water tables. The depth of monitoring reaches up to 2.5 m over deeper water table sites.

A spatial pattern relationship between phosphorus levels and groundwater table depth spatial patterns was first observed using a cross-tabulation method which was a measure of joint frequencies obtained for a matrix of classes of the two patterns. Chi-square was used as a measure of test for the cross-classified patterns, evaluated using equation (7.1). Groundwater level and soluble phosphorus were classified in 10 classes for a chi-square analysis at regional scale.

$$\chi^2 = \sum \frac{(O - E)^2}{E}; \quad \chi^2 = \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{(x_{i,j} - (\sum_{j}^{N} x_i \sum_{i}^{M} x_j) / \sum_{1}^{N} \sum_{1}^{M} x_{i,j})^2}{\sum_{j}^{N} x_i \sum_{i}^{M} x_j / \sum_{1}^{N} \sum_{1}^{M} x_{i,j}} \qquad (7.1)$$

where O and E are observed and expected frequencies, respectively, in a matrix of N and M classes of the phosphorus and groundwater level spatial patterns to be compared; x_{ij} is frequency of observation on classes i (phosphorus) and j (groundwater level).

A geographically distributed multivariate analysis that fit to the local system was conducted using the method of least squares varying the number of neighboring cell sizes until a reasonable fit was obtained. The analysis was done on ArcGIS suite 10 and S+ statistical software package.

7.6. Inputs for operational prediction

Groundwater level was chosen as one of the factors after a report by Young et al. (1998) which states that groundwater fluctuation has a direct effect on redox process and hence on soluble phosphorus. The Kissimmee River phosphorus level (Figure 7.3b) was also considered as an input because of the dominance of groundwater transport dictated by the flat topography not only during base flow but also in wet season flow. Distance from the river (Figure 7.3c) was chosen because of the direct relationship with attenuation processes of phosphorus as it drains to the river.
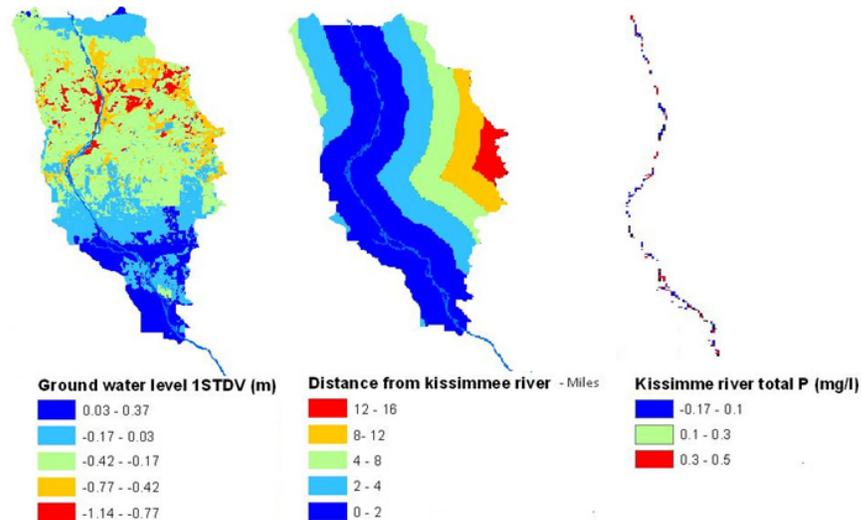

Figure 7.3 Inputs for the soluble phosphorus prediction a) Groundwater level deviation from mean b) distance from Kissimmee River c) Phosphorus level of Kissimmee River.

7.6. Analysis

### 7.6.1. Point scale groundwater level vs. water quality parametric relationship

Graphical plot of soluble phosphorus against groundwater table (measured from surface) showed increased solubility with rise of water table (Figures 7.4a, 724, 7.4e). The relationship could be explained by increased saturation and solubility from top soil (with higher organic matter) as well as proximity of water table to the surface that enhances quick percolation to the groundwater. Groundwater recession will be accompanied by soluble phosphorus fixation because of the increase in solubility of calcium carbonate from the limestone formations. The relationship was the same across all sites showing a power decrease of soluble phosphorus with receding groundwater level from surface. The fit of the curves showed an R-squared value of 0.96 - 0.99 (Table 7.1).

In contrast, the graphical fit of total phosphorus versus groundwater table depth showed an increase in total phosphorus with recession of the groundwater table (Figures 7.4b, 7.4d, 7.4f). As mentioned earlier, the dissolved calcium carbonate from the karst aquifer might have decreased the soluble phosphorus forming a particulate calcium phosphate. Total phosphorus has shown a logarithmic increase with groundwater recession at R-squared value of 0.63 - 0.99 (Table 7.1).
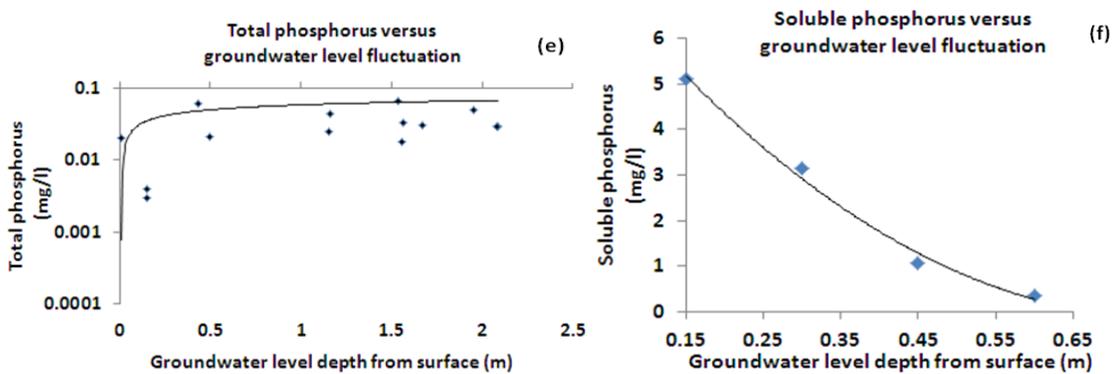
Figure 7.4 Total and soluble phosphorus relationships with groundwater table
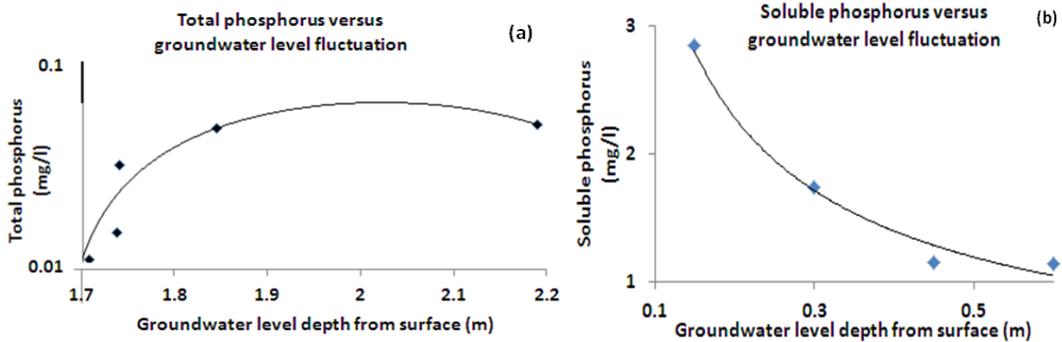
7.5 Total phosphorus (7.4b) Soluble phosphorus at 3AN1-GW / PZ * and ENR40*
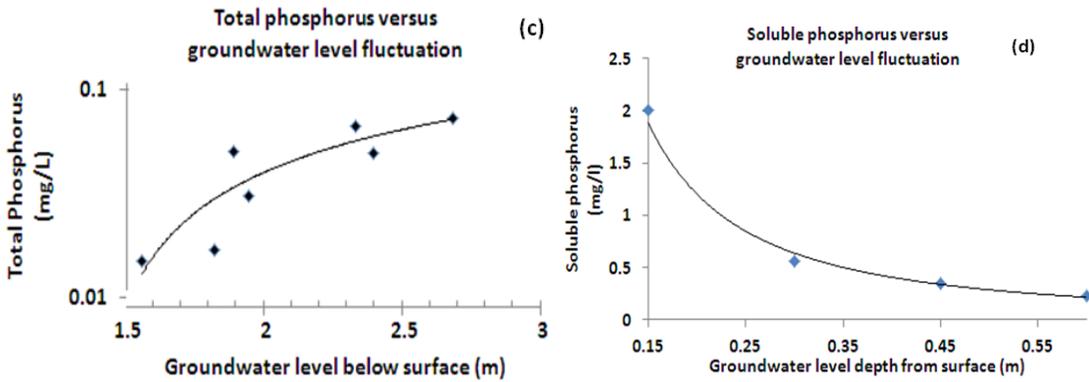
7.6c Total phosphorus (7.4d) soluble phosphorus at 3AS3-GW / PZ and ENR403*

7.7 Total phosphorus (7.4f) soluble phosphorus at SBS1-GW/ PZ* and ENR404*

Table 7.2 R-squared for the testing sites

Total phosphorus against groundwater table depth (R^2)		Soluble phosphorus against groundwater table depth (R^2)	
Stations	R^2	Stations	R^2
3AN1 - GW* / PZ *	0.80	EN40415, EN40430, EN40445, EN40460	0.96
3AS3 – GW * / PZ *	0.63	EN40315, EN40330, EN30445, EN40360	0.99
3BS1 – GW * / PZ*	0.92	EN40415, EN40430, EN40445, EN40460	0.99

Because of the complexity of the phosphorus dynamics being affected by vegetation, soil type, drainage and hydrogeology, the analysis could not extend and interpret that groundwater has a causative relationship. However, the given exploratory analysis of the patterns of relationship is sufficient to test whether phosphorus dynamics holds the same relationship with groundwater fluctuation at watershed / regional scale. Finally, a spatial analysis of the soluble phosphorus against groundwater level using the chi-square analysis showed similarity of patterns at 99% of the time.

7.6.2. Soluble phosphorus prediction: A multivariate analysis

The point scale as well as chi-square analyses mentioned earlier strengthen the fact that groundwater fluctuation (Figure 7.4a) could be explanatory variable for local level change of phosphorus. Accordingly a multivariate analysis of soluble phosphorus using groundwater level, distance from water bodies and P concentration on surface water bodies showed a visual similarity of predicted soluble phosphorous with a pattern of the ArcWAM simulated soluble phosphorus (Figures 7.5a and 7.5b). The analysis showed a spatial dependence of variability with goodness of fit (R^2) ranging 0.2 – 0.7 (Figure 7.6). Hotspot occurrences (maximum phosphorus sources) as well as phosphorus sinks (vast

wetlands) were captured (Figures 7.5a and 7.5b). The highest correlation was observed on hot-spots of phosphorus sources and sinks alike which seem to be related to local homogeneity of observations. The predicted versus observed fit (Figure 7.6, plotted graphs) at different sections of the basin indicated slight over prediction (confirming the visual observation, Figure 7.5b).
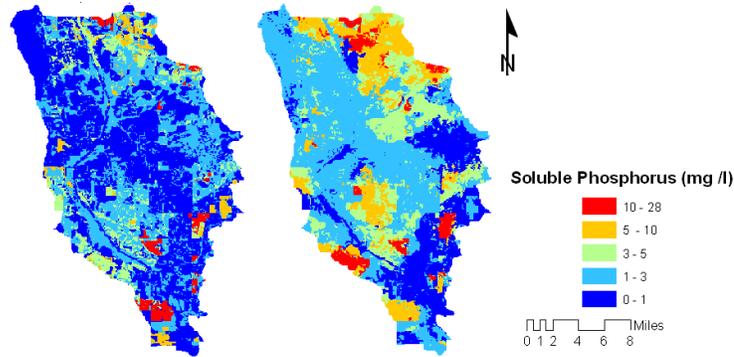

Figure 7.5 a) Phosphorus extracted from watershed based estimation, 'observed' b) Predicted soluble phosphorus
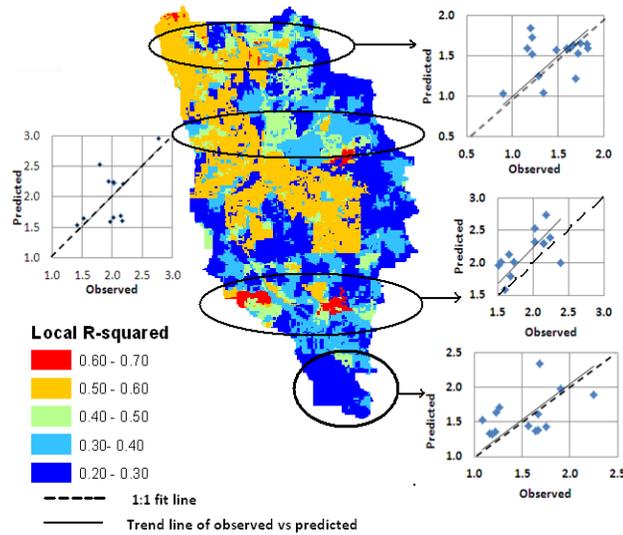

Figure 7.6 The goodness of fit (R^2) and graphical fit of observed versus predicted

A residual analysis was also done using the difference between observed soluble phosphorus and the predicted. The residual spatial patterns showed highest and lowest at transitional sites between hotspots of sources or sinks, respectively as it might be expected (Figure 7.7a). However, the residuals were also found fitting a normal distribution (Figure 7.7b) with mean value of 0 revealing a non-significant difference between the observed and simulated phosphorus levels at p-value of 0.01.
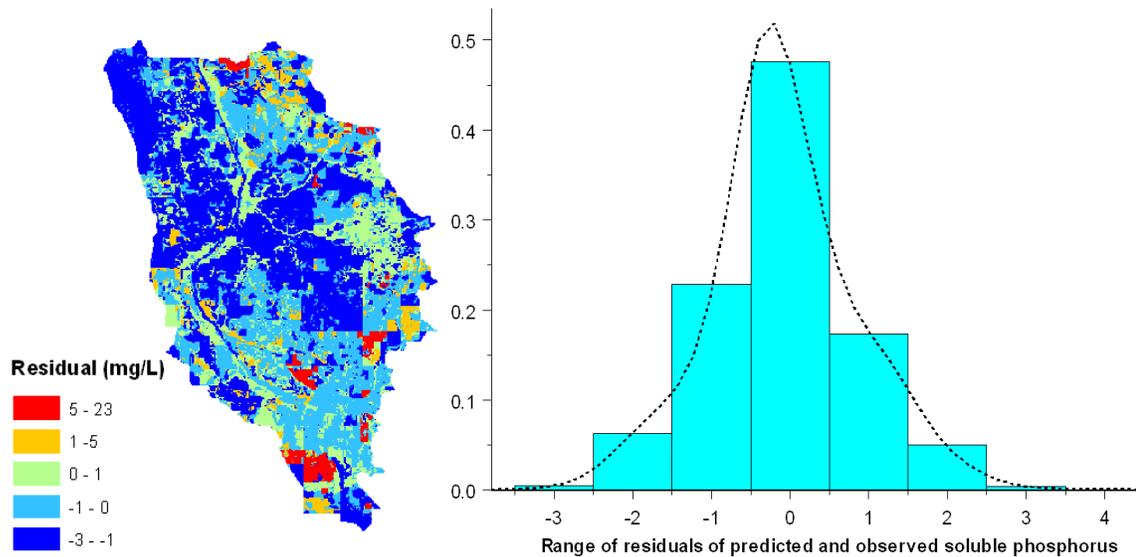
Figure 7. 7a Residual of observed and predicted soluble phosphorus

Figure 7.7b Histogram of residuals fitted with normal distribution

The average weight given to groundwater level for the phosphorus prediction was 0.4 with 99% confidence of occurrence. Whereas the weights of Kissimmee River phosphorus level and the distance from the river course were small showing no significant difference with null value at p-value of 0.2. However, the weights were 95% of the time significantly different near the Kissimmee River (within 5 miles range).

7.8. Conclusion

This study revealed the possibility of predicting soluble P with groundwater spatiotemporal simulation. Point scale relationship of phosphorus dynamics with groundwater level changes were observed over the wetlands of south Florida that showed increased soluble phosphorus with the rise of groundwater. In contrast, total phosphorus was observed to decrease with rise in groundwater and vice versa. A geographically weighted multivariate analysis using inputs of regional groundwater level change, distance from discharging sites and water quality of discharging sites (surficial water bodies) showed that groundwater could explain phosphorus 70% of the time at p-value of 0.01. A chi-square analysis revealed similarity of pattern between groundwater fluctuation pattern and soluble phosphorus over 99% of the time. Analysis of the residual on the predicted and observed soluble phosphorus showed no significant difference at p-value of 0.01 exceeding the expected p-value of 0.1. Therefore, the ultimate goal of predicting phosphorus for operational purposes coupled with groundwater fluctuation was found possible. The research recommended on further data acquisition and fine tuning input factors.

REFERENCES

Chebud, Y., Melesse, A.M (2011). Operational Prediction of Groundwater Fluctuation in South Florida using Sequence based Markovian Stochastic Model, Water Resources Management, 25(9): 2279-2294, DOI: 10.1007/s11269-011-9808-z

Chebud, Y., Melesse, A. (2011b) Spatiotemporal Surface-Groundwater Interaction Simulation over the Karst Hydrogeological Region of South Florida, in review. Water Resources Management.

Chebud Y., Naja G.M., Rivero R. (2011) Phosphorus run-off assessment in a watershed†, J. Environ. Monit., 2011 (13) 66, DOI: 10.1039/c0em00321b

Chebud Y., Naja G.M., Rivero R. Melesse A. M. (2011a) Water Quality Monitoring Using Remote Sensing and Artificial Neural Network, (Accepted)

Moreno-Mateos D, Perocchi C, Comi'n (2009). Effect of wetland construction on water quality in a semi-arid catchment degraded by intensive agricultural use, *Ecological Engineering*, 2009 (1479) 1-9
research grant proposal, USGS

Kadlec R. H. (2006). Free surface wetlands for phosphorus removal: The position of the Everglades Nutrient Removal Project, Ecological Engineering 27(2006)361-379

O'Connor, G.A., H.A. Elliott, and P. Lu. 2002. Characterizing water treatment residuals phosphorus retention. Soil Crop Sci. Soc. FL Proc. 61:67-73.

South Florida Water Management District (2010a). DBHYDRO: http://www.sfwmd.gov/dbhydroplsql/show_dbkey_info_main_menu.

South Florida Water Management District (2010b). Environmental Report, Chapter 11: Kissimmee Basin. Available at: http://www.sfwmd.gov/portal/page/portal/pg_grp_sfwmd_sfer/portlet_sfer/tab2236037/2 232010%2236020report/v2236031/chapters/v2236031_ch2236011.pdf). West Palm Beach, FL.

CHAPTER VIII

Conclusion and Recommendation

The research has introduced point and regional scale groundwater fluctuation model that uses the available data in SFWMD. However, the approach could equally help data scarce areas through synthetic generation of groundwater fluctuation vis a vis scenario analysis. The synthetic generation of state sequences at point scale is the core of prediction in addition to the dynamic nature of the model that updates whenever new observations are made.

Further research on point scale process of phosphorus state retrieval with groundwater level over a range of scenarios would contribute to the development of minimum input operational prediction of phosphorus at point scale as it was observed for groundwater. The way forward would be simulating point scale phosphorus dynamics with analogical approach to groundwater fluctuation simulation since groundwater fluctuation has shown similarly of patterns with soluble phosphorus at 99% the time.

The spatiotemporal groundwater fluctuation simulation has contributed identifying explanatory factors which could later serve for future spatiotemporal simulation of P. Since all the factors identified from spatiotemporal groundwater simulation could easily be captured using a remote sensing approach, the developed model is applicable to use for operational prediction purposes. The rainfall, soil moisture, and water levels from lakes and canals could be captured from satellite imagery as discussed elsewhere.

Because of the extended management strategy of groundwater level by SFWMD, identification of the relationship between the groundwater fluctuation and phosphorus dynamics and the methodological approach developed for soluble phosphorus estimation using groundwater as predictor serves as part of the decision support system

The methodological approach for phosphorus prediction was realized using groundwater fluctuation that explained 60-70% of the variation. The hypothesis was met showing no significant differences between groundwater based prediction of soluble phosphorus and the results from the physical based model at p-value of 0.01, exceeding the hypothesized p-value of 0.1. Further identification of explanatory variables that could be captured with minimum cost (using satellite imagery) as in the case of the spatiotemporal groundwater fluctuation simulation would enhance operational prediction for a greater precision.

The relationship of depth of groundwater and soluble phosphorus draw the attention because of the fact that the site is dominated by karst hydrogeology (excessively transmissive - 300 m/day) to drain phosphorus to surficial water bodies. Because of the observed low sequestration of P in oversaturated areas and subsequently an inevitable transport of soluble and particulate phosphorus from karst hydrogeology to water bodies, a management decision targeted improving uptake of soluble phosphorus would be realistic. The urgency to use vegetation is pronounced, especially in storm water treatment areas where excessive load of phosphorus needs to be siphoned off before it gets at depth and ultimately transported, though in particulate form. A physical management system being practiced currently to keep the groundwater table at a

specified depth from surface to reduce the soluble phosphorous would serve less on nutrient removal.

The limitation of the study was the absence of primary generated data. The study was fully dependent on the published data of SFWMD. As a result, data was not captured at varying hydrogeological and land use settings. Similarly, the threshold depth where soluble phosphorus shows changes (increase /decrease) at different hydrogeological contexts could not be comprehended because of the scarcity of data. So, future work recommended include:

1. Capturing soluble and total phosphorous at varying hydrogeology, soil type, land use outside of the wet lands to understand the causative relationship for the groundwater-phosphorus dynamics.

2. Capturing temporal variation of the relationship between groundwater and phosphorus dynamics was averaged out because of the scarcity of the data. Therefore, generating the groundwater and phosphorus dynamics over meaningful temporal ranges at least for 3 to 5 years would help for conclusiveness of the data

3. Since phosphorus levels of surface water bodies was identified as a feedback input to the groundwater phosphorus prediction, emphasis need to be given to fine-tuning the neural network methods over a reasonable number of shallow surface water bodies.

VITA

Yirgalem A Chebud

Born, Debre-Markos, Ethiopia

1990 – 1993	B.Sc., Applied Physics Addis Ababa University, Ethiopia
1994-1995	Physics and Mathematics Instructor Mekelle University, Ethiopia
1996-1997	M.Sc., Land Reclamation Engineering, University of Southampton
1998-1999	Soil physics and hydrology instructor Mekelle University, Ethiopia
2000	Assistant materials Engineer, WANBO Construction
2001	Research consultant, FARMAFRICA
2002-2004	Netherlands Development Organization (SNV) Research advisor
2005-2006	Research and Development Officer, Consultative group for International Research (CGIAR-ILRI), Ethiopia

PUBLICAIONS AND PRESENTATIONS

PUBLISHED/IN PRESS

Chebud Y A. Melesse A, 2011. *Operational Prediction of Groundwater Fluctuation in South Florida using Sequence based Markovian Stochastic Model*, Water Resources Management, 25(9): 2279-2294, DOI: 10.1007/s11269-011-9808-z

Chebud Y A, Naja M, Rivero R, 2010, *Phosphorus run-off assessment in a watershed*, Journal of Environal Monitoring, 2011, **13**, 66-73

Chebud Y A and Melesse M.A, 2009 *Modelling lake stage and water balance of Lake Tana,* Hydrol. Process. 23, 3534–3544

Chebud Y A and Melesse M.A, (2009), *Numerical modeling of the groundwater flow system of the Gumera sub-basin in Lake Tana basin*, Hydrol. Process. 23, 3694–3704

Chebud Y A, Melesse A, 2012. *Stage level, Volume and Information Content of Lake Tana Using Stochastic Simulation and Wavelet Analysis Methods*, in press, Hydrological Processes.

PUBLICATIONS IN REVIEW

Chebud Y., Naja G.M., Rivero R. Melesse A. M. (2012) Water Quality Monitoring Using Remote Sensing and Artificial Neural Network, (in review)

Chebud, Y., Melesse, A. (2012) Spatiotemporal Surface-Groundwater Interaction Simulation over the Karst Hydrogeological Region of South Florida. Water Resources Management (In review)

PRESENTATIONS

Chebud Y and Melesse A, Application of Remote Sensing and Neural Network Approaches For Prediction of Chlorophyll-a over The Shallow Water Bodies of South Florida, Presented For the Water SCAPE, NASA project, FIU, USA

Chebud Y and Melesse A, (2008) Hydrological Balance of Lake Tana, Presented at AGU meeting, USA

Chebud Y and Melesse A, Water Management Strategic Changes from State Owned to Household Level and its Economic Efficiency, Presented for the Agroecological Research, FIU, USA

Chebud Y and Melesse A, 2008, Hydrological balance of Lake Tana, presented to a workshop on Ecology of the Blue Nile Under Extreme Conditions. Addis Ababa, Ethiopia

Chebud Y and Melesse A, 2008, The groundwater flow system of Gumara sub-basin, presented to a Workshop on Ecology of the Blue Nile Under Extreme Conditions. Addis Ababa, Ethiopia.