

**DRE #323
TECHNICAL MEMORANDUM
May 1995**

**FORECASTING & MANAGEMENT
OF
GROUNDWATER DROUGHT
CONDITIONS IN COLLIER COUNTY**

(DRE #323)

TECHNICAL MEMORANDUM

**FORECASTING AND MANAGEMENT OF
GROUNDWATER DROUGHT CONDITIONS IN
COLLIER COUNTY**

by

Hosung Ahn

May 1995

**Hydrogeology Division
Water Resources Evaluation Department
South Florida Water Management District
West Palm Beach, Florida**

THIS PAGE INTENTIONALLY BLANK

EXECUTIVE SUMMARY

An important issue to water managers is the management of groundwater droughts which is accompanied by the need to forecast and manage the groundwater resources during drought conditions. Drought management at the South Florida Water Management District has generally been a reactive approach. Groundwater levels are routinely monitored, and when critical levels are reached, action is taken in the form of water use restrictions. When water levels recover, the restrictions are lifted. This project is an attempt to make drought management more proactive. By forecasting future groundwater levels based on historical trends, the District may be able to recommend less stringent water use restrictions before critical levels are reached. The result may be that critical water levels, and the more drastic water use restrictions that accompany them, may be able to be avoided using this proactive approach.

There exists a variety of forecast and control methods, among which the Kalman filter algorithm, associated with a stochastic time series model, is one of the most promising options. Since monthly groundwater flow is highly dependent on space and time domains, the underlying stochastic time series model should incorporate the correlation structure of the system. However, there currently exists no tools or examples of such multivariate space-time modeling techniques. Thus, the objective of this study is to develop a state-of-the-art methodology for forecasting and managing regional groundwater droughts, with the goal of providing consistent and subjective rules for a drought management plan based on sound and advanced technology of the stochastic time series theory.

The general methodology adopted for the drought management plan presented here is as follows: First, a state-space form of stochastic time series model was developed using the long-term historical groundwater head data. Using the most current groundwater heads, the spatial groundwater heads for the next month are forecasted by the Kalman filter algorithm with the fitted time series model. Then, the deviations from the target heads are computed, from which the spatial pumpage reduction rates are obtained. The pumpage reduction rates are considered as a management option to meet the target groundwater heads. As a pilot study, a multi-layered

freshwater groundwater system located in Collier County, Florida were used, in which 115 groundwater monitoring wells were selected and historical month-end groundwater heads whose periods of record range from 7 to 35 years were used.

This study consists with the following four parts:

- Part I: Theory of the maximum likelihood fitting of the stochastic space-time autoregression with exogenous variables (STARX) model,
- Part II: Application of the STARX model to the multi-layered aquifer system in Collier County, Florida,
- Part III: Development of a feedforward control scheme for the regional groundwater drought problem, and
- Part IV: Drought management decision support system (DMDSS).

The scope of Part I included development of a new form of the STARX model with a theory of maximum likelihood (ML) estimation of the proposed STARX model. The estimation method, with an assumption of complete-data (no-missing), is based on the expectation-maximization algorithm with a simplified smoother estimator (EMSSE). Several experiments with the EMSSE algorithm were performed in order to investigate its convergence behaviors, sensitivity of the initial parameter assumption, and computational loads. Part II addressed the fundamental problems in identification, calibration, and verification of the STARX model applied for the multi-layered aquifer system in the model area. Also addressed is the correlation structure of the groundwater system by the space-time correlation function, and the statistics of the first spatial order neighbors defined by the Thiessen polygons created by the gaging station network. In Part III, a feedforward control scheme was developed to control the groundwater head during the anticipated drought condition. This control scheme consists of a forecasting equation developed in Part II and a control equation developed by the empirical relationship between the conceptual Pumpage/Recharge terms and the corresponding head changes. Using the deviations from the target defined by the 2-in-10-year historical heads, the control equation is used to compute the spatial pumpage reduction rates by the composite of deviations in each layer. As an ultimate utility software, a drought management decision support system (DMDSS) was

developed in Part IV. DMDSS is an ARC/INFO based, user-friendly, and menu-driven interface that integrates functionalities of data input, forecasting, control schemes, interpolation, and graphical display of both input and output data as well as all relevant GIS coverages.

The major scientific contributions of this study is development of a new form of the STARX model with its parameter calibration procedure, which may have other potential applications in regional groundwater forecasting as well as in general scientific and engineering problems. In addition, the formulation of the space-time correlation function, a conceptual PR function, and development of a control equation applied for the feedforward control scheme may be credited as major accomplishments of this study.

One of the inherent limitations of the proposed STARX modeling approach is that even though it was designed for a large scale problem, this approach has dimensional limitations in the state variable. That is, a STARX model having more than approximately 100 state variables causes an extreme inefficiency in calibration. The other minor deficiencies in the proposed control scheme are that of using a lumped-layer control scheme rather than layer-by-layer, that of the spatial independence assumption in the pumpage/recharge equation, and that of short period of records (minimum 7 years) in calibration.

It is highly recommended that this type of groundwater drought management model be expanded to the other areas. Results of those models will be useful not only as drought management tools during anticipated drought periods, but also as an invaluable database of long-term historical groundwater heads that may be valuable data sources for the other water resources management purposes. Additional recommendations for implementation of this model are:

- (1) The District's Water Shortage Plan would need to be modified to allow this approach, and actual experience with the proposed management plan would need to be obtained in order to test the effectiveness of this approach.
- (2) A real time data acquisition system for monthly groundwater head monitoring is required in the model area since the real time forecasting is achieved by the most current groundwater head information. The current monthly groundwater head data should be

directly ported to the DMDSS system to forecast the next month heads and the corresponding recommended spatial pumpage reduction rates.

- (3) Collection of historical pumpage data should be extended to non-public water uses, including agricultural water uses. This data will be a good resource for the verification of the proposed feedforward control scheme as well as the general groundwater models.
- (4) The forecasting model should be continuously recalibrated since the most current information result in the most accurate forecasting model. It is highly recommended to recalibrate the forecasting model every two years or after experiencing a severe drought event. Since the feedforward control scheme is constructed on the theory of water budget and assumptions of 2-in-10-year rainfall and permitted groundwater uses, it is desirable to evaluate the current control scheme using real drought events, from which improvements and updates of the proposed scheme may be possible.

LIST OF CONTENTS

Part I. Theory of the maximum likelihood fitting of the stochastic space-time autoregression with exogenous variables (STARX) model

1. Introduction	I-3
2. The Proposed STARX Model	I-6
3. Spatial Index Matrix (SIM)	I-9
4. Maximum Likelihood (ML) Fitting of the STARM Model	I-13
4.1. State-Space Formulation	I-15
4.2. Filtering, Smoothing, and Forecasting	I-16
4.3. The Modified Kalman Smoothed Estimator	I-17
4.4. Expectation-Maximization (EM) Algorithm	I-19
5. Simplified Kalman Smoother for Complete-Data Set	I-24
6. Properties of the EMSSE Algorithm	I-26
6.1. Convergence	I-27
6.2. ML Function Versus Number of Neighbors, N_n	I-29
6.3. Sensitivity of the Initial Parameter Assumptions	I-31
6.4. Computational Load	I-33
7. Summary and Conclusions	I-35
REFERENCES	I-37

Part II. Application of the STARX model to the multi-layered aquifer system in Collier County, Florida

1. Introduction	II-3
2. Description of the Model Area with Historical Data	II-4
3. Preliminary Considerations to Build Stochastic Time Series Model	II-8
4. Space-Time Correlation Functions	II-9
Results and Discussion	II-18
5. Treatment of the Raw Data	II-19
5.1. Seasonality	II-19

5.2. Normality	II-22
5.3. Temporal Trend	II-23
6. Designing the Forecasting Model Structure	II-25
6.1. Determination of Neighbor Sites by the Thiessen Polygons	II-26
6.2. Statistics on the First Order Neighbors	II-29
7. Calibration of the STARX Model	II-31
8. Verification	II-34
8.1. Data Generation	II-34
8.2. Forecasting with Its Error Covariance	II-36
9. Summary and Conclusion	II-45
REFERENCES	II-46
Appendix II-A. Gaging Station Information	II-47

**Part III. Development of a feedforward control scheme for the regional
 groundwater drought problem**

1. Introduction	III-3
2. Existing Rules for Groundwater Use Permits and Drought Contingency Plan	III-4
3. System Control Theory	III-7
4. Feedforward Control Scheme	III-8
4.1. Operation of the System Equation	III-11
4.2. Supplemental Water Requirement (SWR) for Groundwater Use Permits	III-12
4.3. Rainfall-Groundwater Head Relationship	III-16
5. A Conceptual Pumpage/Recharge (PR) Term	III-17
6. Optimization of a PR Parameter Set	III-24
7. Develop a Control Equation	III-25
8. Target Water Level	III-26
9. Simulation and Discussion	III-28
10. Conclusion	III-33
REFERENCES	III-34

Part IV. Drought management decision support system (DMDSS) User's Manual
Version 1.0

1. Introduction	IV-3
2. How to Start the Application	IV-4
3. Overview of the Main Menu	IV-6
4. Description of Menus	IV-8
4.1. Model Data	IV-8
1.a. Historical data	IV-8
A. Contour Map	IV-9
B. Time Series	IV-12
Monthly Statistics	IV-12
Yearly Statistics	IV-12
Correlations	IV-12
1.b. GIS Coverage	IV-15
4.2. Parameter Calibration	IV-18
2.a. Spatial Index Matrix (SIM)	IV-18
2.b. Update Yearly Data	IV-21
2.c. Update Monthly Data	IV-21
2.d. Parameter Calibration	IV-21
4.3. Forecasting	IV-22
3.a. Parameters	IV-22
3.b. Present Condition	IV-22
3.c. Execute Forecasting Model	IV-22
4.4. Output	IV-24
4.a. Spatial Mapping	IV-24
4.b. Time Series	IV-27

LIST OF FIGURES

Part I.	
Figure 1.	Spatial order with influence boundary for a regularly spaced square system. I-10
Figure 2.	Gaging station network with the spatial index matrices specified for the example problem. I-12
Figure 3.	Convergence of ML function by the EMSSE method. I-28
Figure 4.	Convergence of ML function with different number of neighbors N_n and temporal order N_q for the example problem. I-30
Figure 5.	The estimated ML values with different initial parameters $\Xi(0)$ for the example problem. I-32
Figure 6.	Average CPU times, in seconds per iteration, by the EMSSE algorithm. . I-34
 Part II.	
Figure 1.	Location of model area II-5
Figure 2.	Selected groundwater gaging stations with their ID numbers. II-6
Figure 3.	An inter-space matrix with block notations of both space-time correlation functions and semi-variograms. II-12
Figure 4.	The fitted space-time correlation functions of groundwater head and rainfall II-13
Figure 5.	Space-time correlation functions for the upper-triangular blocks II-15
Figure 6.	Time lagged semi-variogram functions of groundwater head and rainfall II-16
Figure 7.	Spatially averaged monthly mean and standard deviation of groundwater heads. II-20
Figure 8.	Thiessen polygons with their gaging station ID numbers. II-28
Figure 9.	Statistics of the generated data versus time step II-35
Figure 10.	Space-time correlation functions of the generated groundwater heads for the Surficial aquifer II-35
Figure 11.	Comparison of forecasted and historical heads, Layer 1 & Lead time 1 . . II-39

Figure 12.	Comparison of forecasted and historical heads, Layer 1 & Lead time 2 . .	II-40
Figure 13.	Comparison of forecasted and historical heads, Layer 1 & Lead time 4 . .	II-41
Figure 14.	Comparison of forecasted and historical heads, Layer 2 & Lead time 1 . .	II-42
Figure 15.	Comparison of forecasted and historical heads, Layer 3 & Lead time 1 . .	II-43
Figure 16.	Comparison of forecasted and historical heads, Layer 4 & Lead time 1 . .	II-44

Part III.

Figure 1.	A feedforward control scheme for groundwater drought management. . . .	III-9
Figure 2.	Concept of real time forecasting using the Kalman filtering algorithm. .	III-11
Figure 3.	Location of study area with rainfall and ET stations	III-14
Figure 4.	Time series plots of rainfall-groundwater heads for the selective stations.	III-15
Figure 5.	Historical pan ET with a 27-year average ET at Naples	III-18
Figure 6.	Public water supply wells with their one-mile buffer zones	III-22
Figure 7.	Agricultural landuses with Thiessen polygons	III-23
Figure 8.	The computed 2-in-10-year Recharge/Pumpage (inches/month)	III-29
Figure 9.	Statistics of the pumpage reduction versus return period of drought . . .	III-30
Figure 10.	An example of recommended pumpage reduction (percent) in case of an 1-in-20-year drought event.	III-31

Part IV.

Figure 1.	Sample output for historical contour map.	IV-11
Figure 2.	Sample output for historical time series.	IV-14
Figure 3.	Model area with basemap	IV-17
Figure 4.	Display of the selected neighbors by the different options..	IV-20
Figure 5.	Output contour maps showing 12-month lead forecasting	IV-26
Figure 6.	An example of forecasted results for station ID #21 over next twelve months.	IV-29

LIST OF TABLES

Part II

Table 1.	Schematics of aquifer system in the model area with their hydraulic properties.	II-7
Table 2.	Statistics of the first spatial order neighbors.	II-30
Table 3.	Test statistics of the alternative models with their estimated ML values. .	II-33
Table 4.	Spatial means and standard deviations of the forecasting errors p_t^t	II-38

Part III.

Table 1.	Monthly crop factors with the estimated parameters for the conceptual PR model	III-21
----------	--	--------

Part I

Theory of Maximum Likelihood Fitting of the Stochastic Space-time Autoregression with Exogenous Variables (STARX) Model

THIS PAGE INTENTIONALLY BLANK

1. Introduction

Forecasting future groundwater heads is an important issue to water managers who are concerned with management of the limited water resources during drought periods. There exist a variety of statistical forecasting techniques. One promising option is the Kalman filter forecasting with an underlying system model, which might be either deterministic or stochastic.

The deterministic system model uses either a full scale or simplified scale physical law of the system. This approach is physically-based and accurate, but complications arise when applied with the Kalman filter forecasting. Another problem in this approach is that it should accompany the forecasting of inputs to the physical system. However, some input variables, like rainfall recharged to the groundwater system, are so uncertain that the forecasting them results in significant uncertainties which makes forecasting itself virtually meaningless. That is why the Kalman filtering with a stochastic time series model is commonly used in practical forecasting problems, and by the same reason, the deterministic system model was eliminated from the consideration in this study.

When a time series model is fitted to the system, identification of model structure and calibration of its parameters may become the main concerns. If spatial correlation is dominant, as in the case of regional groundwater heads, either multivariate autoregressive moving average (MARMA) models or space-time ARMA (STARMA) model can be applied. Along with a family of univariate ARMA models, numerous discussions on the theory of MARMA models are available [see *Box and Jenkins, 1970; Salas et al., 1980; Bras and Rodriguez-Iturbe, 1985; Blockwell and Davis, 1987; Shumway, 1988; and others*]. However, the main drawback of MARMA models is that they do not properly account for the spatial structure of the system.

The spatial structure can be expressed by the statistical correlation with respect to the spatial components such as distance, direction, or elevation. Without properly discriminating neighbor stations, the pure MARMA model applied to large systems may spread its weights to the entire system due to a large degree of freedom, enough to diminish the significance of the

nearby neighbors. Also, the mathematics involved with the sophisticated estimation procedures of the large scale MARMA model become burdensome. That is why a family of univariate ARMA models have been extensively used in practice, even though the system is a multivariate nature in space. An example pertaining to groundwater hydrology, *Shih et al.* [1992] forecasted groundwater heads in Collier County, Florida, by applying a set of single site transfer function models. Also, *Graham and Tankersley* [1993] applied the Kalman filtering with a set of univariate ARMA models to forecast groundwater level in the upper Floridan aquifer in Florida. The deficiency of their approaches is that they never account for the spatial dependency of the system.

A family of STARMA models offers a way of generalizing both the ARMA time series models and the simultaneously specified spatial model. It utilizes a hierarchical spatial ordering of the neighbors. *Cressie* [1991, page 449] summarized several existing STARMA models, among them the following two models are of interest since the structure of the proposed model here resembles them. The STARMA model considered in *Pfeifer and Deutch* [1980] is of the form

$$x_t = \sum_{k=1}^{Np} \sum_{i=0}^{\lambda_k} \phi_{ki} W^{(i)} x_{t-k} + w_t - \sum_{k=1}^{Nq} \sum_{i=0}^{m_k} \theta_{ki} W^{(i)} w_{t-k} \quad (1)$$

where, for the system of nx fixed locations in space, x_t is the $(nx \times 1)$ state vector at time t , w_t is the white noise vector, Np is the temporal autoregressive order, Nq is the temporal moving average order, λ_k and m_k are the k -th spatial orders of autoregressive and moving average terms, ϕ_{ki} and θ_{ki} are the autoregressive and moving average parameters (scalar) at the k -th temporal order and the i -th spatial order, respectively, and $W^{(i)}$ is the $(nx \times nx)$ matrix of weight for the i -th spatial order. The other form of STARMA model with exogenous term given by *Stoffer* [1985] is defined as

$$x_t = \sum_{i=1}^{Nq} \Lambda_i D_i x_{t-i} + \sum_{i=0}^{Nk} \Psi_i z_{t-i} + w_t \quad (2)$$

where, for the n_x fixed stations, A_i is the $(n_x \times n_x)$ diagonal space-time transition intensity matrix at the i -th temporal lag, D_i is a known $(n_x \times n_x)$ distance matrix which expresses the spatial relationship between the random field x_i at lag i , z_i is the $(n_z \times 1)$ covariate vector at time t with (n_z) fixed locations in a covariate space, and Ψ_i is the $(n_x \times n_z)$ regression matrix.

In the above two models, the spatial parameters are pre-determined by either the inverse distance weighing or spatial statistics such as covariance or semi-variogram. Then, parameter calibration is used to find the optimal time dependent multipliers to the predefined spatial structure. A difficulty in applying the above approaches is that it is hard to imagine how spatial dependence can arise other sources than integration of causation over the system [Cressie, 1991, page 450]. That is, defining the spatial weights $W^{(i)}$ or D_i explicitly, rather than from space-time correlation structure itself, might circumvent the real system, and the lumped spatial parameters may lose the integrity of the individual correlation structure of the true system. Conclusively, spatial parameters can be used to define a structure of spatial system in the model, but should not be used to determine the parameters themselves.

To overcome these problems, this study proposes a new form of the Space-Time Auto-Regression with an Exogenous variables (STARX) model suitable for a large scale problem. As a matter of fact, there exists no subjective definition of the large scale system in the time series model, but the large scale system referred to here is a system which has more than 20 state variables, enough to cause divergence problems occasionally during parameter calibration.

The subsequent sections introduce structure and properties of the proposed STARX model, with its parameter estimation method based on the Expectation-Maximization algorithm with a Kalman smoother estimator. Also developed is an EM algorithm with a simplified smoother estimator (EMSSE) for the complete-data. Properties of the estimation algorithm, including convergence, sensitivity to the initial conditions, and computing time were investigated sequentially.

2. The Proposed STARX Model

The proposed STARX model is based on the theory of multivariate time series model with incorporation of the spatial structure via the concept of spatial neighborhood. It is obvious that adding the exogenous term or the so-called covariate, improves estimation of the state variable (variate) and becomes more a physically-based model. The other advantage of using covariate, compared to the simultaneous modeling of both variate and covariate, is that it is possible to reduce the dimension of the system equation which is very critical in the calibration of a large scale model. The reason for eliminating the moving average term in the proposed model is that the moving average term causes severe non-linearity, which adds difficulties in parameter calibration. The STARX model possesses enough degree of freedom, when applied for a large system, so that the autoregressive term with its covariate may be enough to represent the entire system.

In order to formulate the STARX model, suppose that a spatial random vector denoted by x_t at time t , with $t=1, \dots, T$, is of interest to an investigator. With the n_x fixed locations, x_t may be decomposed into components $x_{t,j}$ denoting the state at time t and at the j -th spatial variate, so that $x_t' = (x_{t,1}, \dots, x_{t,n_x})$, where the notation ($'$) indicating the transpose of either vector or matrix. Further suppose that a spatial random vector of covariates denoted by $z_t' = (z_{t,1}, \dots, z_{t,n_z})$ at time t may be measured concurrently from the n_z fixed locations. For a regional groundwater head forecasting problem, $x_{t,j}$ may be spatial potentiometric head at time t and the j -th station in a given aquifer, while $z_{t,j}$ may be either rainfall, evapotranspiration, temperature, adjacent layer's heads, or a composite of them, at time t and the j -th covariate station. With these definitions, a proposed STARX model which describes the current state x_t in terms of the previous states $\{x_{t-1}, \dots, x_{t-N_q}\}$ and the covariate $\{z_t, z_{t-1}, \dots, z_{t-N_k}\}$ is of the form

$$x_t = \sum_{i=1}^{N_q} D_i \circ \Lambda_i x_{t-i} + \sum_{j=0}^{N_k} E_j \circ \Omega_j z_{t-j} + w_t \quad (3.a)$$

where Nq and Nk are the temporal order of regressions for x_t and z_t , respectively, Λ_i is a $(nx \times nx)$ matrix of parameters for the vector x_{t-i} , Ω_j is a $(nx \times nz)$ matrix of parameters for the covariate z_{t-j} , D_i is a known $(nx \times nx)$ spatial index matrix (SIM) for the x_{t-i} , E_j is a known $(nx \times nz)$ SIM for the z_{t-j} vector, and w_t is a $(nx \times 1)$ white noise vector having a covariance of Q . Both D_i and E_j need not be a symmetric matrix. Notation (\circ) is the Hadamard product which is an element-wise product of two matrices of the same size [Horn and Johnson, 1985]. Let us assume that the Hadamard product has higher precedence than that of matrix multiplication. Then, the relationships of both $A \circ BC = (A \circ B)C$ and $A \circ BC \neq A \circ (BC)$ are also satisfied, where A , B , and C are the subset of $M_{m,n}$ with $M_{m,n}$ is a $(m \times n)$ vector space.

In order to encode the spatial structure into the SIM matrix, let us define that the m -th row and n -th column element $d_{i,mn}$ in D_i matrix is of the form

$$d_{i,mn} = \begin{cases} 1 & \text{if the } m\text{-th and } n\text{-th stations are an } i\text{-th time lag neighbor} \\ 0 & \text{otherwise} \end{cases}$$

The same rule is applied for the element $e_{j,mn}$ in E_j matrix, indicating that $z_{t-j,n}$ and $x_{t,m}$ are an j -th time lag neighbor. Determination of the neighboring sites will be discussed in the next section.

The STARX model in equation (3.a) can be rewritten in the block matrix form of

$$x_t = [D_1 \ \dots \ D_{Nq}] \circ [\Lambda_1 \ \dots \ \Lambda_{Nq}] \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-Nq} \end{bmatrix} + [E_0 \ \dots \ E_{Nk}] \circ [\Omega_0 \ \dots \ \Omega_{Nk}] \begin{bmatrix} z_t \\ \vdots \\ z_{t-Nk} \end{bmatrix} + w_t \quad (3.b)$$

or simply,

$$x_t = D \circ \Lambda \ x(t-1) + E \circ \Omega \ z(t) + w_t \quad (3.c)$$

where Λ and Ω are the block matrices, whose components corresponds to those in equation (3.b).

The dimensions of the matrices in equation (3.c) are as follows: D $(nx \times ns)$, E $(nx \times nc)$, Λ

$(ns \times ns)$, and Ω ($nx \times nc$), $x(t-1)$ ($ns \times 1$), and $z(t)$ ($nc \times 1$), where $ns = nx \times Nq$, $nc = nz \times (Nk+1)$, and $ng = ns + nc$. Letting $\Phi_i = D_i \circ \Lambda_i$ and $\Psi_j = E_j \circ \Omega_j$ will further simplify the above equation to

$$x_t = \Phi x(t-1) + \Psi z(t) + w_t \quad (3.d)$$

where $\Phi = [\Phi_1, \dots, \Phi_{Nq}]$ and $\Psi = [\Psi_0, \dots, \Psi_{Nk}]$, which have dimensions of $(nx \times ns)$ and $(nx \times nc)$, respectively. The equation (3.d) is exactly the same format as that of a multivariate autoregressive model with covariate term [see *Blockwell and Davis, 1987*; and *Shumway, 1988*]. However, it should be noted that both Φ and Ψ are sparse matrices due to the Hadamard product by the SIM's, and its behavior is different from that of a full matrix. With properly determined temporal orders Nq and Nk by model identification procedure, let us define that the above model is a STARX model of order (Nq, Nk) , or simply a STARX(Nq, Nk) process.

The proposed STARX model is more flexible in its format than the pure multivariate time series model. That is, if the D_i 's in equation (3.b) are diagonal matrices with no covariate, the STARX model is a set of univariate autoregressive models. However, forecasting by the state-space formulation of the diagonal STARX model is different from that of a set of pure univariate ARMA models, due to the fact that the former uses non-diagonal elements of model noise covariance Q . Also, if the D_i 's are unit matrices, the STARX model is nothing but a multivariate autoregressive model with an exogenous term.

3. Spatial Index Matrix (SIM)

One of advantageous features of the STARX model, compared to MARMA models, is that it can include the spatial structure of the system via the SIM's. However, it is difficult to set up a generalized rule on SIM, mainly because it depends on the spatial structure as well as the scaling of model. That is why the determination of spatial structure into the model must be left to the investigator of the space-time system [Bennett, 1979, p. 477; Pfeifer and Deutsch, 1980(b); Stoffer, 1985]. This section will briefly discuss the possible SIM schemes to be used for the STARX model. In most practical cases, a spatial monitoring network, or sampling grids, plays a key role in determining the spatial structure. Sampling grids can be either regular or irregular spacings, and the regular grid might be either an equilateral triangular grid, a square grid, or a hexagonal grid as discussed in Cressie [1991, p. 318].

For regularly spaced grids, spatial order is assigned by equi-distance from the given location and equally scaled weighing of nearby stations is typically employed [Besag, 1974; Bennett, 1979, p. 477; and Pfeifer and Deutsch, 1980]. In this case, the definition of spatial order represents an increasing order of all possible Euclidean distances from the location of interest. In the case of a one-dimensional regular space system, the number of neighbor sites of the given spatial order is just two times the spatial order. Figure 1 shows the spatial ordering scheme for the regularly spaced square grid system. This figure displays only the first six spatial orders, but it can be extended infinitely. In this ordering scheme, high spatial orders do not have distinct distance differences while the number of neighbor sites for higher spatial orders increases dramatically. Thus, in practical spatial modeling, the first spatial order is commonly used [Bennett, 1979, p.477]. If the spatial weighing scheme is needed, as in the cases of equation (1) and (2), weights are assigned by the inverse of the spatial order if the site is within the boundary, or set to zero if the sites are outside of boundary.

For an irregularly spaced system, either the spatial statistics or the spatial ordering scheme is commonly used. As in the former approach, the semi-variogram can be used after assuming

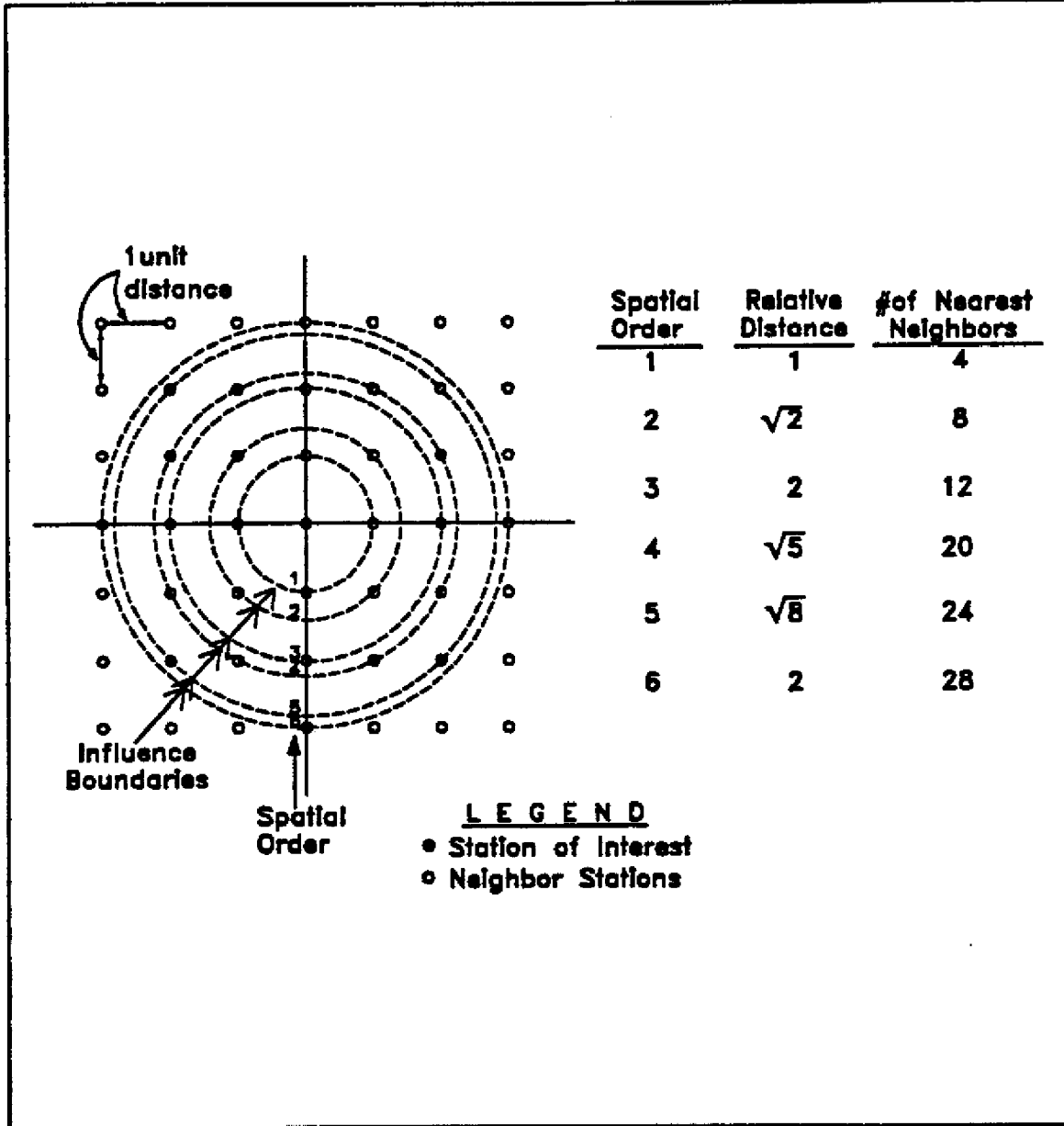


Figure 1. Spatial order with influence boundary for a regularly spaced square system.

that spatial mean and variance are stationary. If a fitted semi-variogram has a significant sill value after a certain distance, the range of the semi-variogram can be used as a boundary for the neighbors. However, it is usually difficult to find such a distinct sill or range in practical cases. Moreover, an assumption of spatial stationarity makes the procedure simple, but loses the true individual correlation system due to a lumped spatial function. Thus, the spatial ordering scheme is more preferred in practical spatial modeling.

The spatial ordering scheme is determined by the lagging of adjacency of boundary cells [Bennett, 1979 p. 481]. That is, the first order sites are the ones which have right adjoined common boundaries to the interest site, and this rule can be extended to the second and higher orders. If two cells are adjoined by a point, they are considered second order neighbors. The site boundary is determined by the natural geographic boundary, or by Thiessen polygons if such a geographic boundary does not exist as in the case of groundwater head monitoring network. Advantages of using the Thiessen polygon method are that this method is easy to apply and that it accounts for both distance and directional components simultaneously in selecting the neighbors. Instead, if the nearest distance criterion is used to select neighbors, the selected neighbor might be skewed to a certain direction which is not desirable. A disadvantage of the Thiessen polygon method is that the numbers of neighbors are not uniform when the monitoring network is irregular.

Bennett [1979 p.484] pointed out that the problem of spatial order schemes in irregular systems is that "extension to high-order spatial orders becomes very dubious since there is no simple or objective rule by which the contiguity counts can be made". Thus, instead of using the higher spatial orders, the first spatial order is strongly recommended for SIMs in the STARX model.

Let us define N_n as the number of neighbors for the given site. N_n within a space may vary from one site to another or one time lag to another time lag, but for simplicity it is assumed that a uniform N_n is used throughout a STARX model. Figure 2 shows an example of constructing SIM's from the first spatial order defined by the Thiessen polygon, where $N_n=3$ was

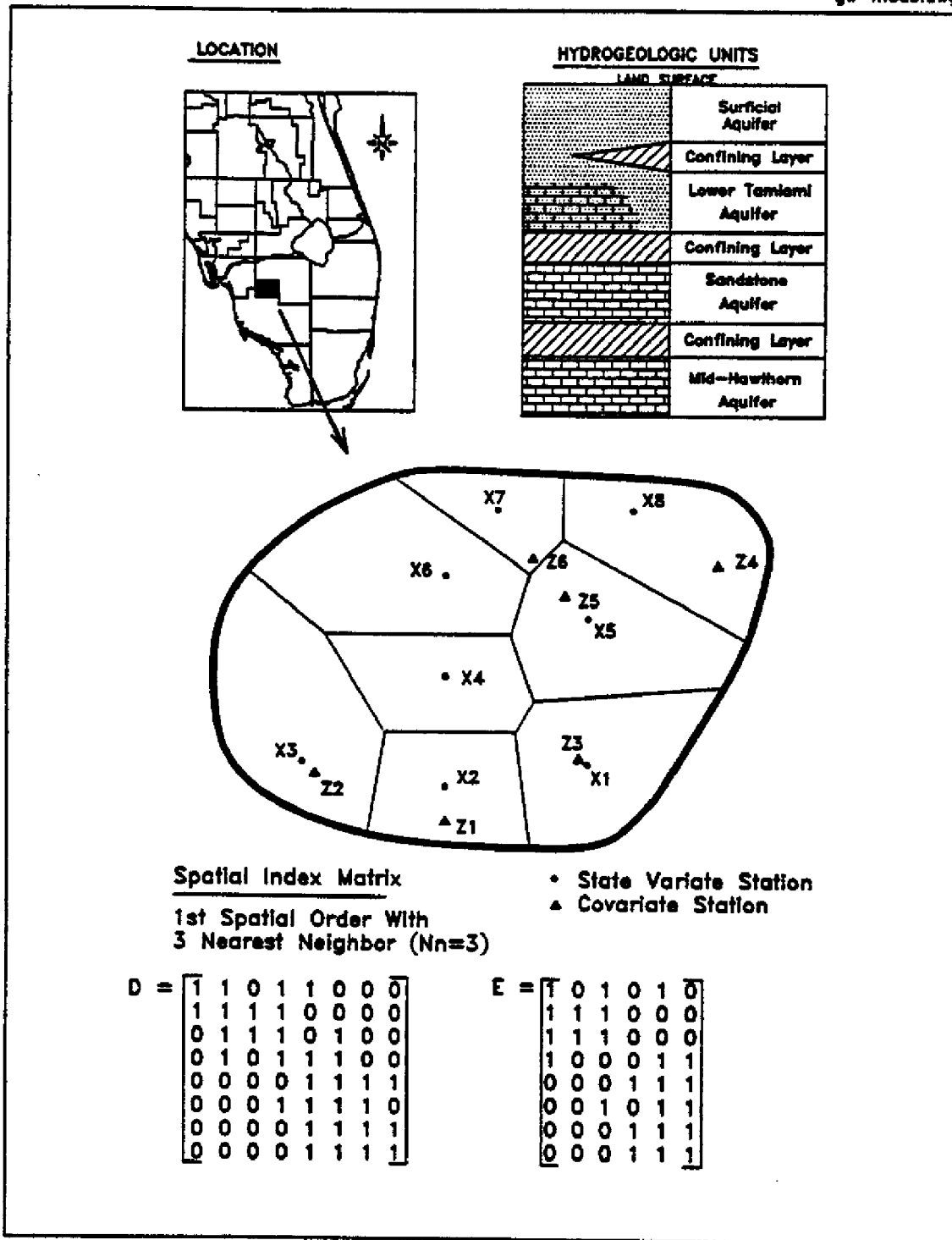


Figure 2. Gaging station network with the spatial index matrices specified for the example problem.

used uniformly for both variate and covariate cases.

The SIM's in the STARX model use a concept of binary indexing of stations to discriminate whether a site is neighbored to the site of interest (1=abuttal) or not (0=no abuttal). The neighbors are determined by the boundary of the Thiessen polygon. In the case of the uniform Nn scheme, the Thiessen polygon can still be used to add or subtract the neighbors to make uniform Nn. In this case, the combination of the Euclidean distance and direction can be used as the neighborhood selection criteria. For instance, the X4 site in Figure 2 has 5 first-order neighbors, but the three nearest neighbors were selected based on the distances. Also, X8 has only two neighbors, but X6 was added to make Nn=3. In the case of a covariate SIM E_i , the Thiessen polygon should be constructed one-by-one for each state variable.

4. Maximum Likelihood (ML) Fitting of the STARX Model

There are several different techniques available to calibrate parameters of stochastic time series models, such as the method of moment, least square method, ML method, etc. The method of moment uses the correlation structure or the partial auto-correlation structure of the data. This method is relatively simple and widely used for model identification process, but needs to refine the parameters further by the other sophisticated methods. It is possible to apply one of the least square methods as in the linear regression model fitting. The least square method uses only the first two moments of the measured time series. *Harvey* [1990, p.84] pointed out that, if the form of the distribution is specified, restricting attention to the first two moments may be statistically inefficient.

The ML method attempts to incorporate all the information into a model by working with the complete distribution of the measurement. This is why the ML method is preferred over the other methods for the time series modeling. To illustrate the ML function, suppose that a set of the measured random variables, y_1, \dots, y_T are available. The statistical time series model specifies a distribution for y_1, \dots, y_T , known as the joint density function, which depends on n unknown

parameters in a vector $\psi = (\psi_1, \dots, \psi_n)'$. Once the sample has been drawn, y_1, \dots, y_T becomes a set of fixed numbers. The expression for the joint density can then be re-interpreted as a function of ψ , where ψ is any admissible value of the parameter vector, rather than the true value. It therefore indicates the plausibility of different values of ψ , given the sample. Viewed in this way, the expression for the joint density function is called the likelihood function, and is denoted by $L(\psi)$. In general, $L(\psi)$ is the continuous function of ψ and the ML estimator can be found by the various optimization algorithms based on the differential calculus of the ML function.

If the objective is just fitting of the STARX model, one may use the Gauss-Newton algorithm or the scoring algorithm as introduced in *Harvey* [1990, p.134]. However, the ultimate objective of fitting the proposed STARX model here is forecasting the future state by the Kalman Filter algorithm. Kalman filter seeks to provide the minimum error variance estimator for the state vector with balancing both the model and output errors, given the measurements. The Kalman filter forecasting requires a set of state model parameters as well as noise covariances and initial state with its error covariance, all of which should be calibrated simultaneously. Thus, a special algorithm, rather than conventional optimization algorithms is requested for forecasting the future state.

It is known that the expectation-maximization (EM) algorithm proposed by *Dempster et al.* [1977] can be applied in conjunction with modified Kalman smoothed estimators to derive a simple recursive procedure for the ML fitting of multivariate ARMA models [*Shumway and Stoffer, 1982; Stoffer, 1985; Shumway, 1988* (p.173); etc.]. Particularly, the EM algorithm, even though it is somewhat complicated, is ideally suited to the multivariate time series applications involving unobserved components or irregularly observed data. Furthermore, the recursive estimation by Kalman filtering can effectively eliminate the divergence of optimization that is commonly occurred by the false initial parameters. The next subsections will discuss the basic assumptions, formulation of general EM algorithm with the Kalman smoother estimator, and the simplified Kalman smoothed estimator for the complete-data set.

4.1. State-Space Formulation

To apply the Kalman filter recursion, the STARX model needs to be transformed into the state-space form which consists with the state and measurement equations. The STARX model in equation (3.d) can be written in the form of the state equation

$$x(t) = \Phi x(t-1) + \Psi z(t) + w(t) \quad (3.e)$$

where $x(t)' = [x_t', \dots, x_{t-N_q+1}']$, $z(t)' = [z_t', \dots, z_{t-N_k}']$, and $w(t)' = [w_t', 0, \dots, 0]$ are augmented vectors having $(ns \times 1)$ dimensions, and the parameter matrices Φ and Ψ are defined by

$$\Phi = \begin{bmatrix} \Phi_1 & \dots & \dots & \Phi_{N_q} \\ I & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}, \quad \text{and} \quad \Psi = \begin{bmatrix} \Psi_0 & \dots & \Psi_{N_k} \\ 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

with I is the identity matrix and 0 is the zero matrix. The model noise vector $w(t)$ is a multi-Gaussian white noise with $w(t) \approx N(0, Q^*)$, where $Q^* = E[w(t) w(t)']$. For more simplicity, the equation (3.e) can be rewritten as

$$x(t) = [\Phi \ \Psi] \begin{bmatrix} x(t-1) \\ z(t) \end{bmatrix} + w(t) = \Theta X(t-1) + w(t) \quad (3.f)$$

which will be used for the maximization step in the EM algorithm.

Now, to allow for the possibility of missing data and the existence of measurement noise, the measurement equation is expressed by

$$y_t = M(t) x(t) + v_t \quad (4)$$

where y_t is an $(n_x \times 1)$ incompletely measured vector at time t , the measurement noise v_t is a $(n_x \times 1)$ multi-Gaussian white noise having $v_t \approx N(0, R)$. The measurement matrix $M(t) = [M_t, 0, \dots, 0]$

has the dimension of $(n \times ns)$, and M_t is a non-random bounded matrix whose element $m_{t,ij}$ at time t and i -th variate is defined by

$$m_{t,ij} = \begin{cases} 1 & \text{if the corresponding state is measured} \\ 0 & \text{otherwise} \end{cases}$$

In the state-space form, it is not possible to measure the true state vector x_t . Only y_t is measurable and x_t is expressed implicitly by the surrogate measurement vector y_t .

4.2. Filtering, Smoothing, and Forecasting

The problem of estimating the discrete state vector $x(t)$ in equation (3.e) can be approached by the expectation of $x(t)$ conditioned on the measured data y_1, \dots, y_T and the concurrently measured covariate z_1, \dots, z_T , where T is the number of total discrete measurements. In order to specify this procedure, consider the general conditional mean of the $(ns \times 1)$ state vector

$$\hat{x}_t^s = E [x(t) \mid y_1, \dots, y_s, z_1, \dots, z_s] \quad (5)$$

where $E[\cdot]$ denotes the expectation operator, s is the span of the measured data and t is the time at which state vector $x(t)$ is estimated. Defining that the estimation error \tilde{x}_t is given by the true value x_t^* minus the estimated value \hat{x}_t^s , that is, $\tilde{x}_t = (x_t^* - \hat{x}_t^s)$ which has a dimension of $(ns \times 1)$. Also let us define that the $(ns \times ns)$ error covariance of the state estimates is

$$P_t^s = E [(x_t^* - \hat{x}_t^s)(x_t^* - \hat{x}_t^s)^T \mid y_1, \dots, y_s, z_1, \dots, z_s] \quad (6)$$

and the error covariance for one-time lagging state vectors is

$$P_{t,t-1}^s = E [(x_t^* - \hat{x}_t^s)(x_{t-1}^* - \hat{x}_{t-1}^s)^T \mid y_1, \dots, y_s, z_1, \dots, z_s]. \quad (7)$$

Then, the following three problems occur when estimating the above three estimates \hat{x}_t^s , P_t^s , and $P_{t,t-1}^s$: if $t=s$, it is called the **filtering** problem; if $t < s$, it is the **smoothing** problem; and

if $t > 0$, it is the forecasting problem. With definitions of the above conditional expectations and some assumptions, the modified Kalman smoothed estimator is derived. The step-by-step derivations can be found in *Jazwinski* [1970 p. 200] and *Shumway and Stoffer* [1981], and this subsection introduces only the fundamental assumptions with their final results.

4.3. The modified Kalman smoothed estimator

To develop the Kalman filter recursion, it is assumed that the initial state $x(0)$ is a Gaussian vector, with $x(0) \approx N(\mu, \Sigma)$, and that w_t and v_t are mutually independent, so that $E[w(t) v_t'] = 0$. Furthermore, it is assumed that both $w(t)$ and v_t are independent to both $x(t)$ and $z(t)$. The calculation of the Kalman filter estimators proceeds by starting the forward recursion. For $t=1, \dots, T$, the predicted state of the STARX model in equation (3.e) becomes

$$\hat{x}_t^{t-1} = \Phi \hat{x}_{t-1}^{t-1} + \Psi z(t) \quad (8)$$

where Φ is $(ns \times ns)$ nonsingular state transition matrix. The predicted error covariance is

$$p_t^{t-1} = \Phi p_{t-1}^{t-1} \Phi' + Q^* \quad (9)$$

where $Q^* = E[w(t) w(t)'] = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$, and $Q = E[w_t w_t']$.

The filter gain, or so called the Kalman gain, is

$$K_t = p_t^{t-1} M(t)' [M(t) p_t^{t-1} M(t)' + R]^{-1} \quad (10)$$

where $R (= E[v_t v_t'])$ is any positive semi-definite matrix which satisfies the $x' R x \geq 0$ condition for all nonzero x , and superscript (-1) denotes the matrix inversion. Then, the state update, or the state estimate, is

$$\hat{x}_t^t = \hat{x}_t^{t-1} + K_t [y_t - M(t) \hat{x}_t^{t-1}] \quad (11)$$

and the state error covariance is

$$P_t^t = P_t^{t-1} + K_t M(t) P_t^{t-1} \quad (12)$$

Equations (8) through (12) are the forward recursion. The initial conditions for the forward recursion are taken to be $x_0^0 = \mu$ and $p_0^0 = \Sigma$, where Σ should be the positive semi-definite matrices to assure the matrix inversion operation at the initial time step.

The general purpose of smoothing is to refine the forward estimators \hat{x}_t^t and P_t^t based on all measured data y_1, \dots, y_T and z_1, \dots, z_T . Assuming that the filtering solution \hat{x}_t^t , P_t^{t+1} , and P_t^t are available from the forward recursion, the backward recursion can be performed for $t=T, T-1, \dots, 1$ using the equations

$$J_{t-1} = P_{t-1}^{t-1} \Phi' [P_t^{t-1}]^{-1} \quad (13)$$

$$\hat{x}_{t-1}^T = \hat{x}_{t-1}^{t-1} + J_{t-1} [\hat{x}_t^T - \Phi \hat{x}_{t-1}^{t-1}] \quad (14)$$

$$P_{t-1}^T = P_{t-1}^{t-1} + J_{t-1} [P_t^T - P_t^{t-1}] J_{t-1}' \quad (15)$$

Also, it is requested to compute the smoothed error covariance for the one-time lagging state $p_{t-1,t-1}^T$ in equation (7) for the expectation step in the EM algorithm. The resulting backward recursion step from *Shumway and Stoffer* [1981] for $t=T, T-1, \dots, 2$, is

$$P_{t-1,t-2}^T = P_{t-1}^{t-1} J_{t-2}' + J_{t-1} [P_{t,t-1}^T - \Phi P_{t-1}^{t-1}] J_{t-2}' \quad (16)$$

and the initial error covariance is computed by

$$P_{T,T-1}^T = [I - K_T M(T)] \Phi P_{T-1}^{T-1} \quad (17)$$

where K_T and p_{T-1}^{T-1} are those from the last time step in the forward recursion. The above forward and backward recursions give a convenient means for calculating the conditional expectations of both the state vector and associated error covariance, which are of interest when estimating parameters in the STARX model. The data are not required to be regularly spaced or complete series in the above formulations.

Unlike equation (8), equation (9) does not account for the covariate effect in estimating the state error covariance. The underlying assumptions used in equation (9) are that covariate is a deterministic term and both $E[\tilde{x}_t' z(t)']$ and $E[z(t) z(t)']$ are set to zero. If the covariate effect is considered, the right hand side of equation (9) should add the $(\Phi E[\tilde{x}_t' z(t)'] \Psi' + \Psi E[\tilde{x}_t' z(t)'] \Phi' + \Psi E[z(t) z(t)'] \Psi')$ term. With this term, the equation (9) becomes similar to that of the Schmidt-Kalman filter [Jazwinski, 1970, p. 285]. The Schmidt-Kalman filter was basically derived for the state-space equation with uncertain parameters that is regarded as a random variable with known a priori statistics. In order to estimate the above term, the predicted state error term \tilde{x}_t should be estimated recursively along with the state vector.

4.4. Expectation-Maximization (EM) Algorithm

The EM algorithm, introduced by *Dempster et al.* [1977], is an alternative non-linear optimization algorithm which enables estimation of model parameter with an incomplete data set using the ML method. The term "incomplete data" implies the existence of two sample spaces: a measured part, and an unmeasured part.

In the time series context, it is common to measure some incomplete data Y as a known function of a signal process X and a noise process V . Both X and V are not directly measured, but only indirectly expressed through Y . Let us define $\ln L(X, V | \theta)$ as a log likelihood based on the complete data, and $\ln L(Y | \theta)$ as a log likelihood based on the incomplete data, where θ is the parameter set of the model to be estimated. Then, the EM algorithm is designed to find θ iteratively, which maximizes the expectation of the complete-data log likelihood conditioned on the measured data Y . Each iteration of the EM algorithm involves two steps; the expectation step

(E-step) and maximization step (M-step). Let us write such a conditional expectation process as

$$Q(\theta|\theta_i) = E_i [\ln L(X, V; \theta) | Y; \theta_i], \quad (18)$$

where i denotes i -th iteration. The E-step computes $Q(\theta | \theta_i)$ from the above expression and the M-step chooses θ_{i+1} to maximize $Q(\theta | \theta_i)$ using one of the optimization techniques. Since the X process cannot be measured directly, the above equation can be written in terms of the Kalman smoothed estimators defined in the previous subsection. The E-step in the case when vectors y_1, \dots, y_T , and z_1, \dots, z_T are fully observed is given by [see *Shumway and Stoffer, 1982*]

$$\begin{aligned} Q(\theta|\theta_i) = & -\frac{1}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \{ \Sigma^{-1} [p^T(0) + (x^T(0) - \mu)(x^T(0) - \mu)'] \} \\ & - \frac{T}{2} \ln |Q| - \frac{1}{2} \text{tr} \{ Q^{-1} (C - \theta S_{i-1}(1)' - S_{i-1}(1)\theta' + \theta S_{i-1}(0)\theta') \} \\ & - \frac{T}{2} \ln |R| - \frac{1}{2} \text{tr} \{ R^{-1} \sum_{t=1}^T [(y_t - M(t)\hat{x}_t^T)(y_t - M(t)\hat{x}_t^T)' + M(t) P_t^T M(t)'] \} \end{aligned} \quad (19)$$

where $x^T(0) \approx N(\mu, \Sigma)$, $|A|$ is the determinant of the matrix A , $\text{tr}A$ is the trace of a matrix A given by $\sum_{i=1}^n a_{ii}$ for $A = \{a_{ij}\}$ with $A \in R^n$, and

$$S_{i-1}(0) = \begin{bmatrix} A & F \\ F' & H \end{bmatrix} \quad (20)$$

$$S_{i-1}(1) = [B \ G] \quad (21)$$

where

$$A = \sum_{t=1}^T [P_{t-1}^T + \hat{x}_{t-1}^T \hat{x}_{t-1}^T] \quad (22)$$

$$B = \sum_{t=1}^T [P_{t,t-1}^T + \hat{x}_{t,1}^T \hat{x}_{t-1,1}^T] \quad (23)$$

$$C = \sum_{t=1}^T [p_t^T + \hat{x}_{t,1}^T \hat{x}_{t,1}^T] \quad (24)$$

$$F = \sum_{t=1}^T [\hat{x}_{t-1}^T z(t)'] \quad (25)$$

$$G = \sum_{t=1}^T [\hat{x}_{t,1}^T z(t)'] \quad (26)$$

$$H = \sum_{t=1}^T [z(t) z(t)'] \quad (27)$$

Again $\hat{x}_{t,1}^T$ is from the Kalman smoothed estimator, and $\hat{x}_{t,1}^T$ is the first sub-vector in the $\hat{x}_t^T = [\hat{x}_{t,1}^T, \dots, \hat{x}_{t,N_q+1}^T]$. The dimensions of each matrix in equations (20) through (28) are as follows: $S_{t-1}(0)$ ($na \times na$), $S_t(1)$ ($nx \times na$), A ($ns \times ns$), B ($nx \times ns$), C ($nx \times nx$), F ($ns \times nc$), G ($nx \times nc$), and H ($nc \times nc$), with $na = ns + nc$.

The M-step is now easily applied by maximizing equation (19) with respect to the parameters θ , Q , and R . The M-step yields the following regression estimators:

$$\theta_{t+1} = D^* \circ [S_t(1) S_{t-1}(0)'] \quad (28)$$

$$Q_{t+1} = \frac{1}{T} (C - \theta_{t+1} S_t(1)' - S_t(1) \theta_{t+1}' + \theta_{t+1} S_{t-1}(0) \theta_{t+1}') \quad (29)$$

$$R_{t+1} = \frac{1}{T} \sum_{t=1}^T [(y_t - M(t) \hat{x}_t^T) (y_t - M(t) \hat{x}_t^T)' + M(t) P_t^T M(t)'] \quad (30)$$

where, from equation (3.f), D^* is defined by

$$D^* = \begin{bmatrix} D_1 & \dots & D_{Nq-1} & D_{Nq} & E_0 & \dots & E_{Nk} \\ I & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & 0 & \vdots & \ddots & \vdots \\ 0 & \dots & I & 0 & 0 & \dots & 0 \end{bmatrix}$$

Also, it is possible to update either $\mu_{i+1} \{=\hat{x}^T(0)\}$ or $\Sigma_{i+1} \{=p^T(0)+(x^T(0)-\mu)(x^T(0)-\mu)'\}$, or both of them at each iteration. However, it is known that both μ and Σ are relatively insensitive to the variations in the initial conditions [Stoffer, 1985]. In case of missing data (incomplete-data) at a given time step, the measurement vector should be partitioned and rearranged into the measured and missing parts by $y_t' = [y_t^{(1)}, y_t^{(2)'}]$ and the measurement noise covariance should be estimated in the partitioned form. The R_{i+1} term in this case given by Stoffer [1985] is

$$R_{i+1} = \frac{1}{T} \sum_{t=1}^T \Pi_t C_t \Pi_t' \quad (31)$$

with

$$C_t = \begin{bmatrix} S_t^{(1)} & S_t^{(1)} R_{2,1}' \\ R_{2,1} S_t^{(1)} & R_{2,1} S_t^{(1)} R_{2,1}' + R_{22,1} \end{bmatrix} \quad (32)$$

$$R_{2,1} = R_{21} R_{11}^{-1} \quad (33)$$

$$R_{22,1} = R_{22} - R_{21} R_{11}^{-1} R_{12} \quad (34)$$

and

$$S_t^{(1)} = (y_t^{(1)} - M(t)^{(1)} x_t^T) (y_t^{(1)} - M(t)^{(1)} x_t^T)' + M(t)^{(1)} P_t^T M(t)^{(1)'} \quad (35)$$

where $R_{ij} = \text{cov}\{v_i^{(i)} v_j^{(j)}\}$, with $v_i^{(i)}$ is the partitioned measurement noise corresponding to $y_t^{(i)}$, R_{ij} is computed first by the complete-data estimator (equation 30) and reshaped into the measured and missing parts like $y_t^{(i)}$. The matrix Π_t in equation (31) is a permutation matrix which is

obtained after reordering the $M(t)$ matrix. A simplification introduced in *Shumway and Stoffer* [1982] is for the case where the missing part and measured part are not correlated, so that $R_{2,1}$ is a zero matrix and the update for the missing part is given by just $R_{2,2}$. If the measurement vector y_t is completely missing, then the term C_t is simply R .

The incomplete-data log-likelihood given by *Gupta and Mehra* [1974] is of the form

$$\begin{aligned}
 -2\ln L(Y; \Theta) \approx & \sum_{t=1}^T \ln |M(t) P_t^{t-1} M(t)'| + \\
 & \sum_{t=1}^T (y_t - M(t)\hat{x}_t^{t-1})' [M(t) P_t^{t-1} M(t)' + R]^{-1} (y_t - M(t)\hat{x}_t^{t-1})
 \end{aligned} \tag{36}$$

where \hat{x}_t^{t-1} and P_t^{t-1} are from the forward recursion. The reason why the log likelihood function is multiplied by (-2) is that this likelihood function is the same as that of the least square objective function. Then, the objective function of the optimization is the minimizing equation (36).

To summarize the above procedure, each M -step recalibrates the following forecasting parameter set

$$\Xi = \{\Theta, Q^*, R, \mu, \text{ and } \Sigma\}$$

based on the E-step by the Kalman smoothed estimator. The entire EM procedure can be computed as follows:

- (I) Start with initial assumption of Ξ_0 .
- (II) On the i -th iteration, run the forward and backward recursions using equations (8) through (17) with Ξ_i from the previous iteration.
- (III) Update Ξ_{i+1} by the M -step in equations (28) through (35).
- (IV) Compute a likelihood function given by equation (36) and check convergence. Stop iteration when the incomplete-data log-likelihood and the parameter set stabilize, otherwise repeat step (II) and (III) for the $i+1$ iteration.

5. Simplified Kalman Smoother for Complete-Data Set

Although the EM algorithm has its greatest advantage when it is used for incomplete-data sets, it is possible to simplify the Kalman smoothing procedure with an assumption of complete-data. The EM algorithm with a simplified smoothed estimator (EMSSE) enables one not only to increase the convergence rate, but to significantly reduce computational time and memory requirements. The simplification of the Kalman smoothed estimator is accomplished by using time-invariant state error covariance terms. The concept of time-invariance in Kalman filtering was briefly discussed by *Lainiotis* [1978], where the partitioned filter is solved in the form of a time-invariance and steady-state (Weiner) filter.

Without missing data, the measurement equation can be rewritten by

$$y_t = M x(t) + v_t \quad (37)$$

where the time-invariant measurement matrix ($n \times ns$) becomes $M=[I,0,\dots,0]$, with I and 0 are identity and zero matrices having ($n \times n$) dimensions, respectively. Since the measurement matrix M is time-invariant, p_t^{-1} , p_t^t , and K_t in the forward recursion also become time-invariant variables. Let us redefine these variables as the time-invariant notations; $p_1=p_t^{-1}$, $p_0=p_t^t$, and $K=K_t$. Then, equations (9), (10), and (12) can be rewritten by

$$\left\{ \begin{array}{ll} p_1 = \Phi p_0 \Phi' + Q & \dots(a) \\ K = p_1 M' [M p_1 M' + R]^{-1} & \dots(b) \\ p_0 = (I - KM) p_1 & \dots(c) \end{array} \right. \quad (38)$$

which can be estimated independently from the state estimation. This is a set of simultaneous equations having three unknown terms and three known terms (Φ , Q , and R). The explicit solution of the above equation is unknown, but an iterative method can be used to solve it since the Kalman forward recursion is designed to obtain the minimum mean-squared error estimation. That is, with an initial assumption of p_0 , which satisfies the positive semi-definition matrix, the above equations are solved for p_1 , K , and P_0 sequentially and repeatedly until the estimators are stabilized. The experiment in this study reveals that less than 10 iteration leads to a good

convergent solution.

The same iteration method can be applied for the backward recursion. That is, the time-invariant estimator J ($=J_t$) is computed by

$$J = p_0 \Phi' [p_1]^{-1} \quad (38)$$

Also, with time-invariant notations of p_0^T ($=p^T$) and p_1^T ($=p_{t-1,t-2}^T$), the smoothed error covariance matrices are rewritten by

$$p_0^T - J p_0^T J' = p_0 - J p_1 J' \quad (39)$$

$$p_1^T - J p_1^T J' = p_0 J' - J \Phi p_0 J' \quad (40)$$

Again, both equations (39) and (40) were derived under the condition of minimum mean-squared error estimation, so that an iteration of each equation with an initial assumption which satisfies the positive semi-definition condition will result in the optimal solutions.

Alternatively, both equations (39) and (40) can be solved explicitly by using the Kronecker sum properties [Bellman, 1970, p. 125]. That is, since each equation has the form

$$X - UXV = W \quad (41)$$

where X is an unknown matrix, and U, V , and W are known coefficient matrices, the solution of the above equation to X is given by

$$X = \sum_{k=0}^{\infty} U^k W V^k \quad (42)$$

In practice, summation of the above equation continues until the term $(U^k W V^k)$ at the k -th summation is insignificant. The condition of equation (42) is that the sum of any two characteristic roots (eigenvalues) of U and V is nonzero. This condition is automatically satisfied when both U and V are the stability matrices, which means that all of the eigenvalues of U and

V have negative real parts. Since $V=U'$ for both equations (39) and (40), the characteristic polynomials of both U and V are the same and the above stability condition is automatically satisfied as long as J , p_0 , and p_1 are all positive semi-definite matrices. After solving the error covariances in equations (38) through (40), the forward recursion to estimate the state vector is given by equation (8) and (11) using the time-invariant matrices K and M , and the backward recursion is computed by equation (14) with the J matrix.

The great advantage of EMSSE is that whenever complete-data is available, the computational load is significantly reduced due to the reduced number of matrix inversions; the larger the system is, the greater is the reduction of computational loads. The other advantage is that use of time-invariant parameters eliminates the effect of the initial assumption of the error covariances, and consequently escalates the convergence rate of the EM algorithm. These advantages have significant meaning for a large scale model where divergence of the forward recursion may occurred in early iteration step.

6. Properties of the EMSSE Algorithm

A number of mathematical properties of the general EM algorithm are available in *Dempster et al.* [1977] and *Wu* [1983]. The key property is that, under the continuity and differentiability of the ML function $L(Y;\theta)$, the EM sequence of the form $Q(\theta|\theta_i)$ in equation (18) converges to the likelihood value L^* which maximizes $L(Y;\theta)$ giving an optimal parameter θ^* . If $Q(\theta|\theta_i)$ is continuous, then L^* is the stationary value of $L(Y;\theta)$. If θ_i converges to θ^* , then θ^* is a stationary point under continuity of the partial derivative of Q with respect to θ .

The intention of this section is to investigate the behaviors of the proposed EMSSE algorithm, particularly when it is used to calibrate the forecasting parameter set Ξ . Four experiments were conducted: convergence, ML function versus a number of neighbor N_n , sensitivity of initial parameter set $\Xi(0)$, and computational loads. For these experiments, an example problem was set up as in Figure 2, which is an actual groundwater monitoring network

taken from the northern part of Collier County, Florida. It consists of 8 variate stations ($n_x=8$) from the sandstone aquifer (3rd layer from the top) and 6 covariate stations ($n_z=6$) from the lower Tamiami aquifer (2nd layer). The standardized historical month-end head data from January 1987 to August 1993 were used.

6.1. Convergence

One of the interesting aspects of the EMSSE algorithm is that it eventually diverges after a certain convergence. The larger the system is, the faster the divergence becomes. It should be noted that divergence of the EMSSE algorithm is somewhat different from filter divergence. The latter one occurs when the Kalman filter is constructed on the basis of an erroneous model; some treatment techniques of filter divergence are available in *Jazwinski* [1970 p.301].

The convergence pattern of ML function by the EMSSE algorithm has three distinct phases as shown in Figure 3; let us define them as the **rapid convergence**, **gradual convergence**, and **divergence phases**. With any assumed initial parameter set $\Xi(0)$, the ML function $-2\ln L$ dramatically decreases at the first iteration and continues to decrease significantly during a few more iterations (until 4-th iteration in Figure 3) mainly due to the rapid convergence of the first term of ML function in equation (36). This is the rapid convergence phase. After this phase, the algorithm still converges at a relatively small rate. This is the gradual convergence phase. The optimal parameter set is obtained at the end of the gradual convergence phase. With continuing iterations, the ML function begins to oscillate with a very small amplitude at the beginning but increases gradually. This is the beginning of the divergence phase.

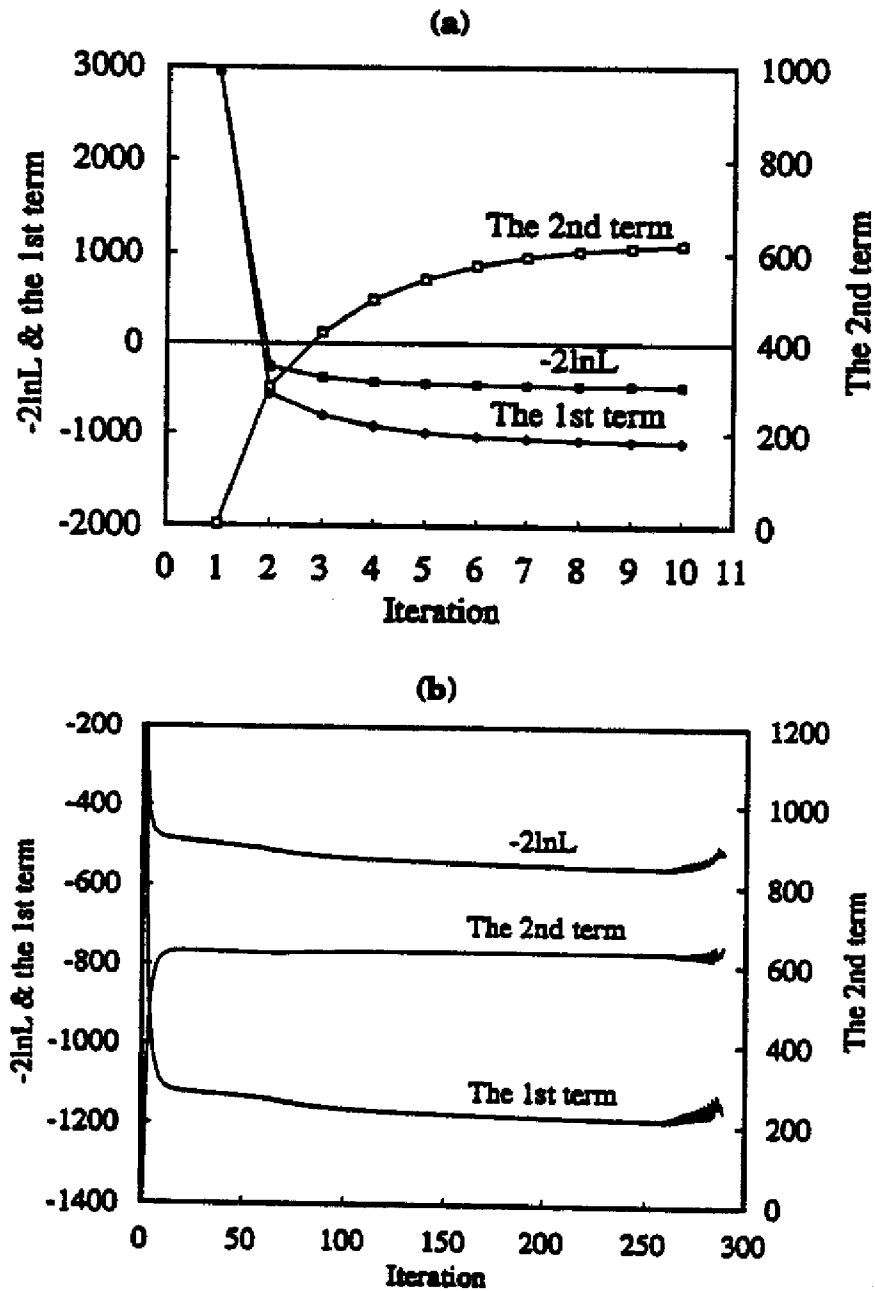


Figure 3. Convergence of ML function by the EMSSE method. Calibration conditions are $n_x=8$, $n_z=6$, $N_q=1$, $T=80$, with full SIM system: (a) Initial stage (up to the 10-th iteration), (b) Overall convergence (up to the 289-th iteration).

In Figure 3(b), the actual oscillation started at the 216-th iteration and it increased dramatically after about 260-th iteration. The EMSSE algorithm was terminated automatically at 289-th iteration due to an unrealistically large ML value, specifically the determinant in the first term of ML function.

The second term in the ML function (36) behaves opposite to its first term. That is, during both the rapid and gradual convergence phases, the second term increases while $-2\ln L$ and its first term decrease. In the case of a large system, the magnitude of the second term is relatively small compared to the first term. However, the starting of divergence of the $-2\ln L$ is slightly different from that of the first term due to the effect of second term.

6.2 ML Function Versus Number of Neighbors, N_n

In the second experiment, calibrations were conducted with different SIMs created by different N_n 's ranging from 0 to 7(full SIM). For three different temporal orders ($N_q=1, 2,$ and 3), the optimal ML values, as well as the first term of the ML functions were computed and plotted in Figure 4. In general, $(-2\ln L)$ is inversely proportional to N_q when the SIM is a full or near full matrix, since the higher N_q means more degrees of freedom in model. However, this rule does not hold for the sparse SIMs which behaves more or less non-linearly. For example, the model with $N_q=3$ outperformed than that of $N_q=2$, even though the first terms of the ML function were almost identical regardless of N_q .

The reasons for such non-linearity are as follows: The EMSSE algorithm identifies its model parameter θ from the full SIM system at first hand, then it performs the Hadamard product of the SIM and θ at the end of each iteration to get the parameter matrix that has a sparse format. That is, even though θ is calibrated for the full system, the uncorrelated portion of θ are ignored by the Hadamard product and the next iteration uses only the correlated portion of θ by the predefined neighborhood concept. Thus, the EMSSE algorithm with the sparse SIM model still converges at a certain degree, but an optimal ML value for the sparse SIM model is higher than that of the full matrix model.

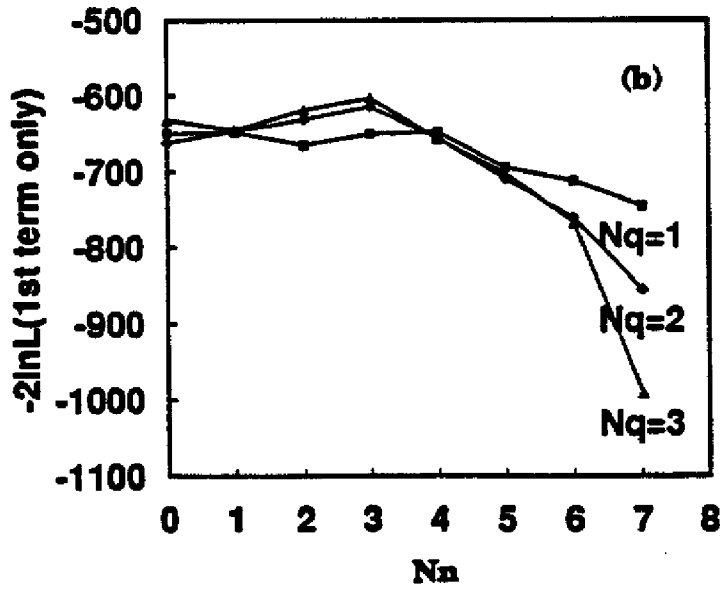
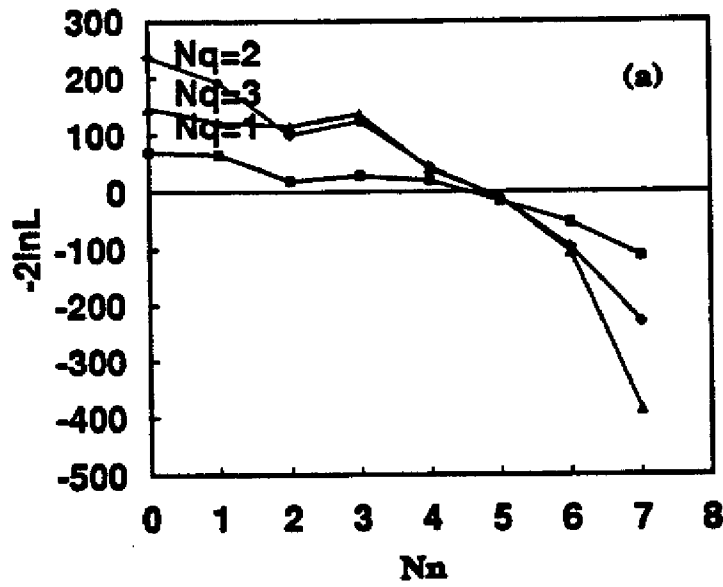


Figure 4. Convergence of ML function with different number of neighbors N_n and temporal order N_q for the example problem.

Alternatively, it is possible to use a sparse system approach in the Kalman smoother. That is, at each EM step, the error covariance terms, p_1 , p_0 , p_0^T , and p_1^T , are obtained by utilizing the Hadamard products. That is, for any sets of two sites that are not neighbors, their error covariance terms are not set to zeros in the EM algorithm. This scheme seems to be more logical, but in most cases, it causes extreme instability in filtering, and such instability hinders the convergence of the EMSSE algorithm.

6.3. Sensitivity of the Initial Parameter Assumptions

The EMSSE algorithm starts with the initial parameter set $\Xi(0)=\{\mu(0), \Sigma(0), \theta(0), Q(0), \text{ and } R(0)\}$, where (0) indicates the initial step before iteration. It was known that the EM algorithm with Kalman smoothed estimator is relatively insensitive to the variations of the initial conditions $\mu(0)=\hat{\lambda}_0^0$ and $\Sigma(0)=p_0^0$ [Stoffer, 1985]. This fact is still valid for the EMSSE algorithm. Furthermore, an assumption of $\Sigma(0)=p_0$ is needed, but this initial value never affect the final convergence. The third experiment of the EMSSE algorithm was conducted to analyze the sensitivities of the ML function with respect to the three initial parameters $\theta(0)$, $Q(0)$, and $R(0)$. The resulting ML values with different initial parameters were plotted with respect to the number of iterations in the Figure 5.

The results of this experiment indicated that the ML function converged to the same values after certain iterations regardless of the initial $\theta(0)$'s and $Q(0)$'s. Among the initial parameters, $R(0)$ was the most sensitive. The exact values of $-2\ln L$'s at the 50-th iteration (which are not shown in Figure 5(c)) are as follows.

Diagonal of $R(0)$	0.2	0.5	1.0	2.0	4.0	10.0
$-2\ln L$	-510.2	-514.2	-514.5	-514.1	-513.5	-512.6

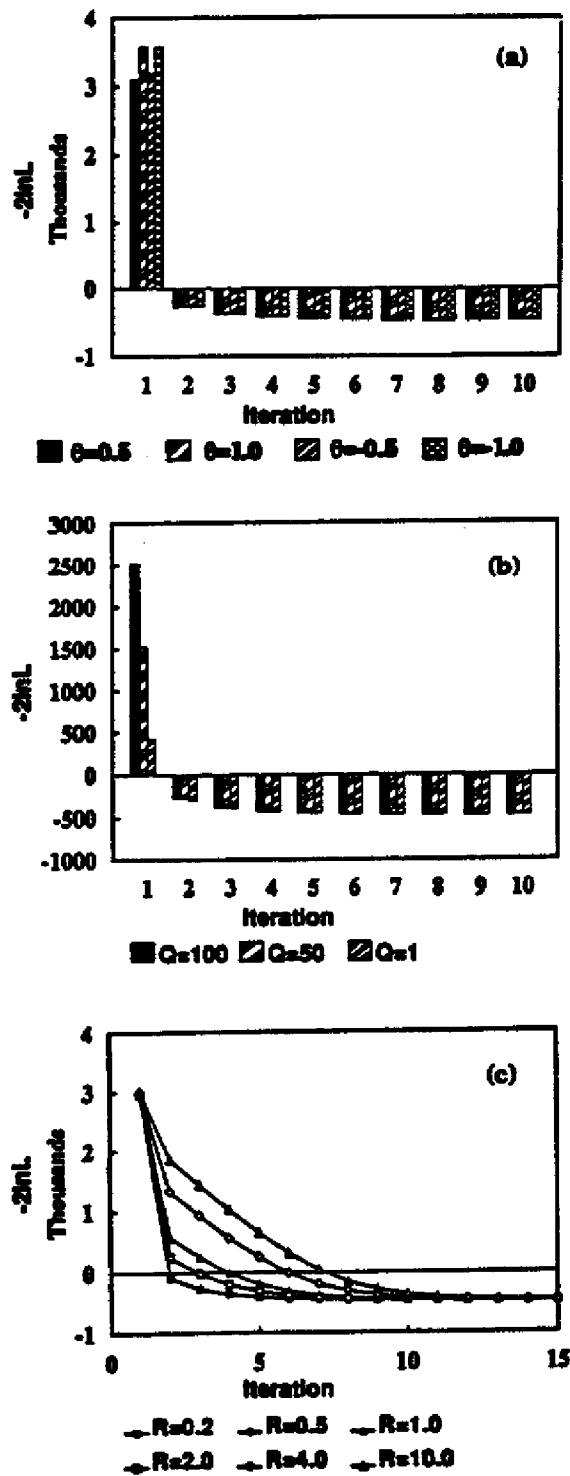


Figure 5. The estimated ML values with different initial parameters $\Xi(0)$ for the example problem, where $\mu_i=0$, $\Sigma_{ii}=1$, $\theta_{ij}=0.1$, $Q_{ii}=100$, $R_{ii}=0.1$, and all off-diagonal in Σ , Q and R are zero: (a) Changing $\theta(0)$, (b) Changing $Q(0)$, and (c) Changing $R(0)$.

Particularly, the influence of $R(0)$ was very significant during the rapid convergence phase. The small $R(0)$ induces fast convergence of the EMSSE algorithm, but may cause quick divergence, particularly for the large scale model. It was verified that, for small or medium scale problems (approximately $n_x < 20$), initial $R(0)$ of less than 0.1 may give good convergence, and for the large scale problem ($n_x > 20$), $R(0)$ of 0.5 or higher, may produce stabilized ML values with optimal parameter estimation without divergence during the rapid convergence phase. From this experiment, the following conclusions are drawn:

(1) The initial parameters $\theta(0)$, $Q(0)$, and $\mu(0)$ are insensitive to the final ML function, even though the ML values at the first few iterations are quite different. $\Sigma(0)$ can be any positive semi-definite value if the time-invariant scheme is used in the EMSSE algorithm.

(2) The optimal ML function by the EMSSE algorithm is not significantly affected by $\theta(0)$. That is, regardless of any $\theta(0)$'s, the algorithm converges to the L^* giving an optimal model parameter θ^* .

(3) Among the parameter set Ξ , $R(0)$ is the most sensitive parameter to the ML function in the EMSSE algorithm. In most cases, diagonals of $R(0)$ of less than 0.2 leads to an optimal solution, but for the large scale model, diagonals of $R(0)$ of 0.5 or higher are recommended to avoid divergence during the rapid convergence phase.

6.4. Computational Load

The EMSSE algorithm requires considerable computational time due to numerous matrix manipulations. Nevertheless, discussions of the exact computation time (CPU time) is inadequate because the CPU time is subject to change depending on the hardware system, programming habitat (especially matrix manipulations), and the input-output options in the program. In this paper, the experiment was focused on how much computational loads are increased, in a relative sense, by increasing the dimension n_x and temporal order N_q .

The EMSSE algorithm was coded in FORTRAN with some IMSL subroutines and executed on a SUN Sparc-10 workstation. During the calibration of the model with different

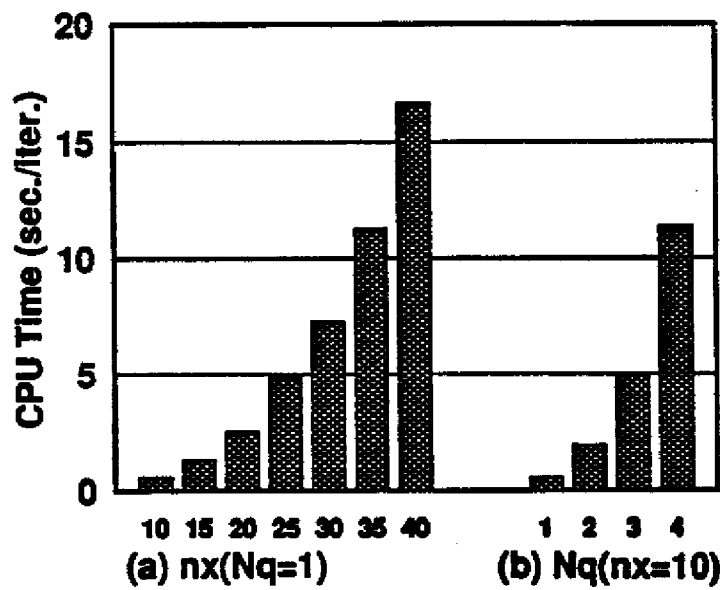


Figure 6. Average CPU times, in second per iteration, by the EMSSE algorithm: (a) Changing n_x with a fixed $N_q=1$, (b) Changing N_q with a fixed $n_x=10$.

n_x and N_q , the CPU times were monitored. Figure 6 shows CPU times in seconds per iteration, which is an average of 20 iterations of the EMSSE algorithm. This experiment revealed that CPU time increases exponentially with proportional to either n_x or N_q , mainly due to the increased dimensions of the matrices. The fitted quadratic equations for CPU time ω in seconds per iteration in both cases were given by

$$\omega = 1.614 - 0.290n_x + 0.016n_x^2, \quad \text{with } R^2 = 0.998, \quad (44)$$

$$\omega = 2.038 - 2.643N_q + 1.238N_q^2, \quad \text{with } R^2 = 0.994, \quad (45)$$

respectively. When increasing both n_x and N_q , or using covariate in the model, the computation time increases dramatically. For instance, if $n_x=48$, $N_q=4$, and $T=72$, the total CPU time for 25 iterations took about 9.6 hours.

7. Summary and Conclusion

Presenting a methodology to fit a stochastic time series model in the space-time domain has been the main goal of this paper, which will ultimately be used to forecast the regional groundwater head. This study developed a new form of the space-time autoregression with exogenous variables (STARX) model, which uses the Hadamard product of parameter matrices and spatial index matrices to incorporate the spatial correlation of the system. After discussing the structures of the STARX model, efforts were concentrated on the ML fitting, which is based on the expectation-maximization algorithm with a simplified smoother estimator (EMSSE). The EMSSE algorithm, if the complete-data set is available, not only enables the reduction of the computational load, but also achieves accurate parameter estimation by using time-invariant error covariance matrices.

To verify the EMSSE algorithm, several experiments were conducted, which included the convergence fashion, ML versus number of neighbors N_n , sensitivity of the initial parameter assumption $\Xi(0) = \{\mu(0), \Sigma(0), \theta(0), Q(0), \text{ and } R(0)\}$, and computational load. After defining three convergence phases based on the ML function obtained by the EMSSE algorithm,

comparisons of convergence rates were made for the full system versus different sparse systems. The sensitivity analysis to the initial parameter assumption revealed that $\theta(0)$, $Q(0)$, $\mu(0)$, and $\Sigma(0)$ are insensitive to the ML function, but $R(0)$, particularly its diagonal elements, is the most sensitive to the ML function in the EMSSE algorithm. Computational load increases approximately proportional to either nx^2 or Nq^2 .

REFERENCES

- Bellman, R., *Methods of Nonlinear Analysis*, Volume I, Academic Press, New York, 1970.
- Bennett, R. J. *Spatial Time Series: Analysis-Forecasting-Control*, Pion Ltd., 207 Brondesbury Park, London, 1979.
- Besag, J., Spatial Interaction and the Statistical Analysis of Lattice System, *Journal of Royal Statistical Society B*, 36, 192-235, 1974.
- Box, G.E.P., and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.
- Bras, R. L., and I. Rodriguez-Iturbe, *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass., 1985.
- Brockwell, P.J., and R.A. Davis, *Time Series Theory and Methods*, Springer-Verlag, N.Y., 1987.
- Cressie, N. A. C. *Statistics for Spatial Data*. John Wiley & Sons Inc., New York, 1991.
- Dempster, A. P., N.M. Laird, and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of Royal Statistical Society B*, 39, 1-38, 1977.
- Graham, W.D., and C. Tankersley, Forecasting Piezometric Head Levels in the Floridan Aquifer: A Kalman Filtering Approach, *Water Resour. Res.*, 29(11), 3791-3800, 1993.
- Gupta, N.K., and Mehra, R.K., Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculation, *IEEE Transactions on Automatic Control*, 19, 774-783, 1970.
- Harvey, A. C., *The Econometric Analysis of Time Series*, Second Edition, The MIT Press, Cambridge, Massachusetts, 1990.
- Horn, R.A., and C.R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, N.Y., 1985.
- Jazwinski, A.H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- Jones, R. H., and Vecchia, Fitting Continuous ARMA Models to Unequally Spaced Spatial Data, *Journal of American Statistical Association*, Vol. 88, No. 423, Theory and Methods, September, 1993.
- Lainiotis, D. G., Partitioned Filters, *Applications of Kalman Filter to Hydrology, Hydraulics, and*

- Water Resources*, page 71-81, edited by C. Chiu, Univ. of Pittsburgh, Pittsburgh, PA, 1978.
- Pfeifer, P. E. and Deutsch, S. J. Identification and Interpretation of First Order Space-Time ARMA Models. *Technometrics*, 22, 397-408, 1980a.
- Pfeifer, P. E. and Deutsch, S. J. A Three-Stage Iterative Procedure for Space-Time Modelling. *Technometrics*, 22, 35-47, 1980b.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and L.W. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Fort Collins, Colo., 1980.
- Shih, G., W.S. Burns, and R.F. Bower, A ground Water Drought Management Model for Collier County, Florida, Tech. Pub. #92-01, South Florida Water Mgt. District, West Palm Beach Florida, 1992.
- Shumway, R.H., *Applied Statistical Time Series Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1988.
- Shumway, R. H., and D. S. Stoffer, Time Series Smoothing and Forecasting using the EM Algorithm, Tech. Report No. 27, Div of Statistics, University of California, Davis, California, 1981.
- Shumway, R. H., and D. S. Stoffer, An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm. *J. of Time Series Anal.*, 3, 253-264, 1982.
- Stoffer, D. S. Maximum Likelihood Fitting of STARMAX Models to Incomplete Space-Time Series Data. In *Time Series Analysis: Theory and Practice 6*, page 283-296, edited by O.. Anderson, J. K. Ord, and E. A. Robinson, North-Holland, Amsterdam, 1985.
- Tankersley, C. D., W. D. Graham, Comparison of Univariate and Transfer Function Models of Groundwater Fluctuations, *Water Resour. Res.*, 29(10), 3517-3533, 1993.
- Wu, C.F., On the Convergence Properties of the EM Algorithm, *Ann. Statistics*, 11, 95-103, 1983.

Part II

Application of the STARX Model to the Multi-layered Aquifer System in Collier County, Florida

THIS PAGE INTENTIONALLY BLANK

1. Introduction

Like most types of modeling, procedures involved in stochastic time series modeling include identification, calibration, and verification. The algorithm of the calibration process for the proposed STARX model was presented in Part I. The objective of Part II is to address the state-of-the-art methodologies of the modeling procedure as well as the practical issues that arose during the STARX model building. As a pilot study for the groundwater drought management problem, a model area was selected which is located in the western Collier County, Florida. The month-end groundwater head data from 115 monitoring wells in the four freshwater aquifers were collected, whose periods of record vary from 7 to 35 years.

To identify the structure of the STARX model, the spatial statistics of the space-time correlation function and the time lagged semi-variogram function between the different space domains were analyzed, along with the statistics of the first spatial neighbors defined by the Thiessen polygon created by the gaging station network. Discussions also includes transformation processes of the raw data to meet the underlying assumptions in the STARX model. Those transformations includes standardization to remove seasonality, the Box-Cox transformation for normality, and use of linear models to remove temporal trends.

Then, parameters of several candidate STARX models having different temporal orders were calibrated along with error covariance matrices, from which the best models in each layer were selected based on the Akaike information criterion. After investigating the abilities of data generation to verify the fitted STARX model, 12-month lead time forecasting was performed, whose result as well as error covariances were displayed along with the historical data. Results show that the fitted STARX models preserve first four moments and the low-order correlation coefficient of the historical data. Some of forecasting results were displayed with the corresponding historic data which shows good matches between them for the smaller lead times.

2. Description of the Model Area with Historical Data

The model area consists of the western half of Collier County and a small portion of southwestern Lee County, Florida (Figure 1). Its boundary is defined by;

	X-coordinate	Y-coordinate	Longitude	Latitude
Minimum	221734	574947	W 81°51'	N 25°55'
Maximum	421499	788160	W 81°14'	N 26°30'

This area is characterized by moderately drained sandy soils with extensive agricultural and urban development. This region receives an average of 54.3 inches of rainfall annually, with about 60 percent occurring during the wet season (mid May-October). Without having a significant surface water inflow source or surface water storage, the area relies mainly on the groundwater sources to meet water use demands. Thus, this area typically experiences water shortage problems during the dry season. Also, the groundwater uses in this area are very sensitive to the conflicting issues of water supply, salt water intrusion, and environmental impacts on the extensive wetlands which are located along the western edge of the model area.

The area's hydrogeology consists of four underlying fresh water aquifers as described in Table 1, which provide most of the region's water demands. The top three aquifers are separated from each other by semi-confining beds having low permeable material, however a significant amount of inter-aquifer flow takes place due to the leaky structure. These three aquifers are extensively developed due to their low salinity content. The mid-Hawthorn aquifer is separated from the sandstone aquifer by a low permeable confining unit, and not highly utilized due to increased well depth, high salinity, and low yield.

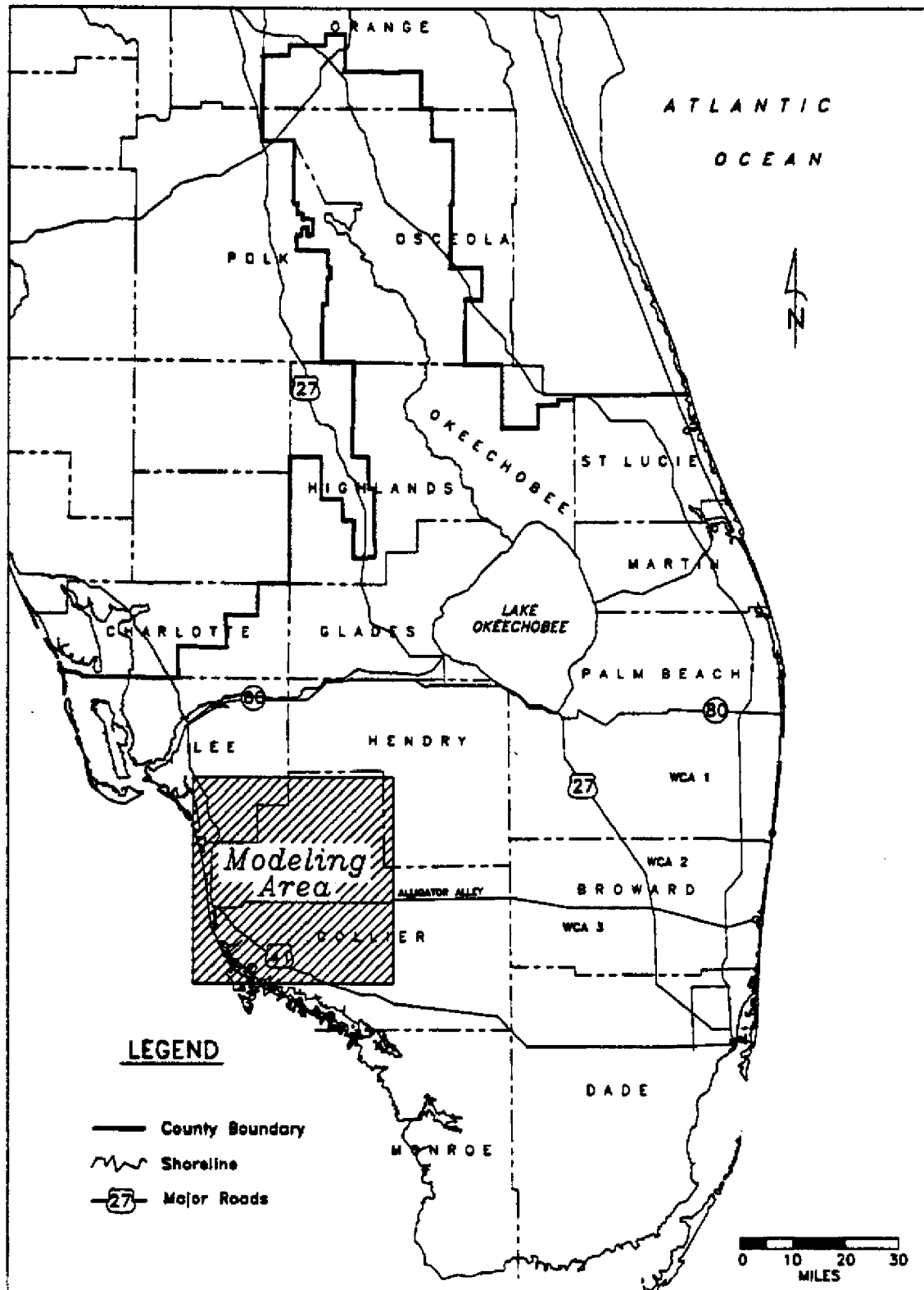


Figure 1. Location of model area.

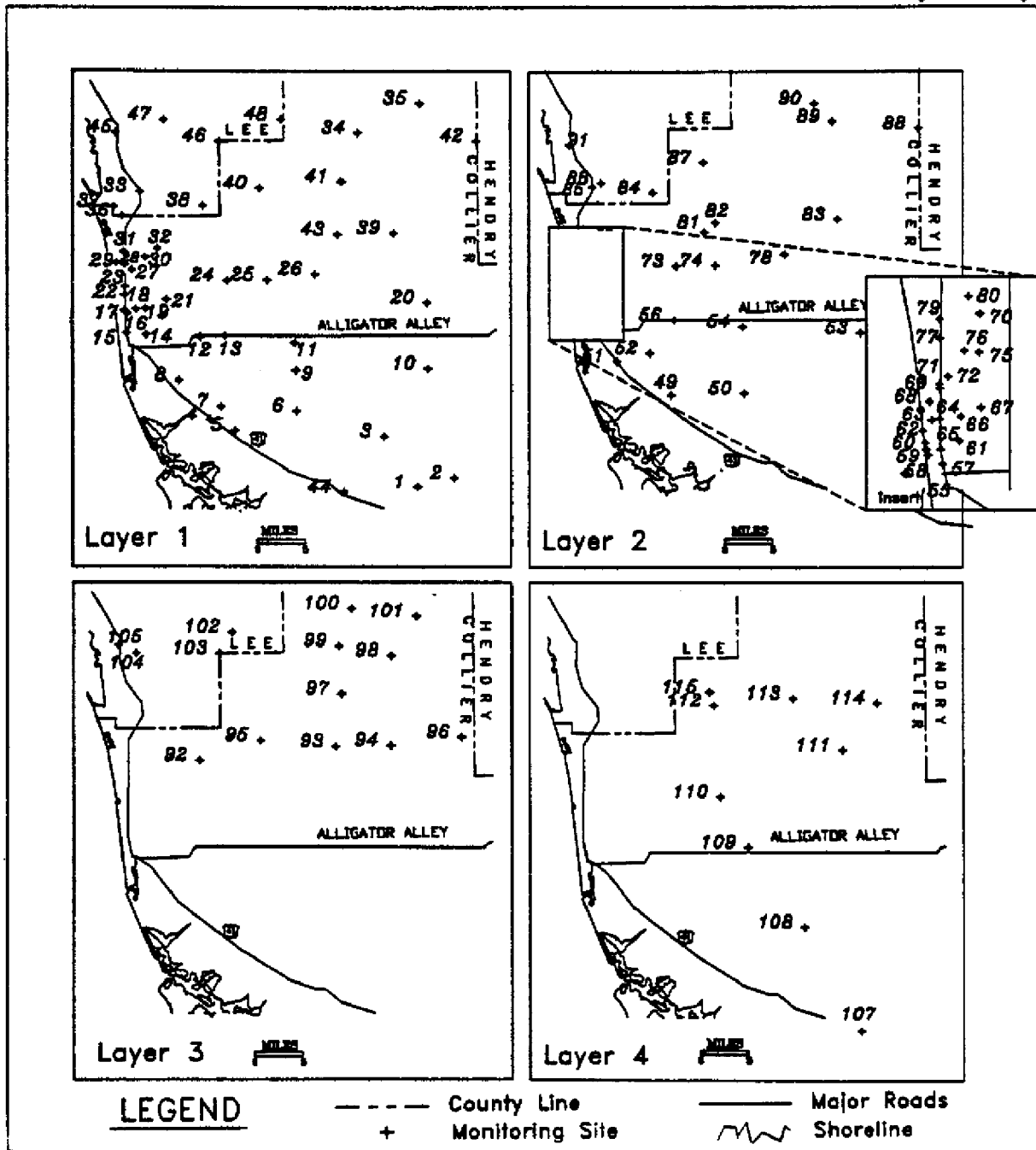


Figure 2. Selected groundwater gaging stations with their ID numbers: (a) Surficial aquifer, (b) Lower Tamiami aquifer, (c) Sandstone aquifer, and (d) Mid-Hawthorn aquifer.

Table 1. Schematics of aquifer system in the model area with their hydraulic properties

Layer ID	Name of aquifer	Aquifer thickness b(ft)	K_H or T (ft ² /day)	# of gaging stations selected
Layer 1	Surficial aquifer	40-60	100-3500	48
Layer 2	Lower Tamiami aquifer	70-120	10000-320000	43
Layer 3	Sandstone aquifer	150-250	160-25000	14
Layer 4	Mid-Hawthorn aquifer	160-410	500-1200	10

Ref.: K_H is the horizontal hydraulic conductivity, and T is the transmissivity.

Currently, the U.S. Geological Survey (USGS) maintains approximately 150 groundwater monitoring wells in this area, from which piezometric heads are collected at either regular or random time intervals. Also, some local agencies and private companies (mainly agricultural operations) have collected groundwater head data, but those are not included in this study due to their unknown reliability and inconsistency in data collection and reporting. The periods of record range from the early 1900's through August 1993. However, only a few wells have records extending back to the mid 1950s. The active USGS monitoring program in this area started during the mid-1970s, and a full-fledged monitoring program began in 1987; from that time, a complete-data set is available. After eliminating stations whose records are inconsistent or having sampling intervals greater than one month, only 115 gaging stations were selected. Table 1 and Figure 2 show the distribution of these stations for each aquifer. The groundwater head in the area, along with the land surface elevation, are usually highest in the northeast region and gradually decreases to the southwest coastal region, where the fresh groundwater interfaces with ambient saltwater. The groundwater flow direction follows this natural head gradient, except for the mid-Hawthorn aquifer where the groundwater flows in a more east-west direction.

3. Preliminary Considerations to Build Stochastic Time Series Model

The governing equation describing the three dimensional transient flow of a slightly compressible fluid in a non-homogeneous anisotropic aquifer can be written [McDonald and Harbaugh, 1988, p98] as;

$$\frac{\partial}{\partial x}(K_{xx} \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}(K_{yy} \frac{\partial h}{\partial y}) + \frac{\partial}{\partial z}(K_{zz} \frac{\partial h}{\partial z}) = S_0 \frac{\partial h}{\partial t} + W \quad (1)$$

where K_{xx} , K_{yy} , and K_{zz} are values of hydraulic conductivity (LT^{-1}) along the x, y, and z coordinate axes, S_0 is the specific storage (L^{-1}), h is the potentiometric head (L), t is the time (T), and W is a volumetric flux per unit volume (take a positive sign for outflow from the system, T^{-1}) which is characterized by pumpage, recharge, leakages from and to surface-water bodies, flux from adjacent formations, evapotranspiration, etc. Numerous analytical and numerical solutions to the above governing equation are known for various system scales as well as different boundary conditions, and selecting a solution method is largely dependant on the purpose of the analysis and data availability.

Bennett [1992] simulated the regional groundwater flow in the model area by applying the MODFLOW model developed by McDonald and Harbaugh [1988]. This kind of physical model, even though it does not have any forecasting capability, provides invaluable information on the groundwater system and usually is used to estimate the impact of alternative groundwater uses. The sensitivity analysis in his report showed that head in the Surficial aquifer is the most sensitive to the changes of rainfall, and that heads at the Lower Tamiami, Sandstone, and Mid-Hawthorn aquifers are mostly affected by the vertical hydraulic conductance of the semi-confining zones.

Instead of using a mathematical model describing the above full dynamics of groundwater flow, this study used a type of stochastic time series model. The stochastic time series model can not incorporate all the above physical components due to its dimensional limitations. Furthermore, the forecasting error increases proportional to the increasing forecasting

lead time in general. These two facts justify the use of time series model with only limited independent variables. The important thing in the time series model is not a detailed physical groundwater flow mechanism, but a statistical correlation structure of the system, both auto-correlation and cross-correlation.

Pumpage is a sensitive component to the groundwater system. Historical records showed that irrigation pumping for agricultural and landscape is approximately 78 percent of the groundwater uses in this area based on the 1988 estimates [Bennett, 1992]. However, only a limited number of actual pumping records are available in the study area. Thus, the pumpage term was unable to include in the time series model. To investigate the cause and effect of the simplified groundwater system, the correlation analyses were focused on the heads between the layers, as well as rainfall.

4. Space-Time Correlation Functions

The two most commonly used spatial statistics are the correlation coefficient and semi-variogram. As long as the maximum likelihood method is used in parameter calibration, those spatial statistics are not directly used in the time series model. However, the spatial statistics are very critical to identify the structure of the STARX model. Also, investigating the spatial statistics of historical heads will help to understand the whole groundwater flow system correctly in the model area.

In order to effectively illustrate a multi-dimensional groundwater system in the model area, the following definitions of space-time correlation are made: First, the correlation coefficient, ρ_{XY} of two random variables X and Y is defined by

$$\rho_{XY} = \frac{cov[X,Y]}{\sigma_X\sigma_Y} \quad (2)$$

with $cov[X,Y]=E[(X-\mu_X)(Y-\mu_Y)]$, and μ_X , σ_X , μ_Y , and σ_Y are the means and standard deviations

of X and Y, respectively. ρ_{XY} is a unitless measure of the linear relationship of X and Y, and satisfies $-1 \leq \rho_{XY} \leq 1$ condition. Using the standardized series of X and Y (which is dividing its standard deviation after subtracting its mean from the original series), the correlation coefficient is the same as the covariance, that is, $\rho_{XY} = \text{cov}[X, Y] = E[XY]$.

Introducing the time-lag l , let us define the space-time correlation coefficient $\gamma_{ij}(l)$ of two standardized random variables x_i and x_j as

$$\gamma_{ij}(l) = E[x_{t,i} x_{t-l,j}] = \frac{1}{T^*} \sum_{t=1}^{T^*} (x_{t,i} x_{t-l,j}) \quad (3)$$

where T^* is the number of effective pairs of measurements. For a complete-data, T^* is $(T-l)$ with T as the period of record. If missing data are presented in either $x_{t,i}$ or $x_{t-l,j}$, or in both, $\gamma_{ij}(l)$ is computed by only the non-missing pairs. The basic properties of the space-time correlation coefficient are: $\gamma_{ij}(l) = \gamma_{ji}(-l)$, but $\gamma_{ij}(l) \neq \gamma_{ji}(l)$, and $\gamma_{ij}(l) \neq \gamma_{ij}(-l)$; If $i=j$, $\gamma_{ij}(l)$ is the auto-correlation coefficient, with $\gamma_{ij}(l) = \gamma_{ij}(-l)$; and if $l=0$, $\gamma_{ij}(0)$ is the cross correlation coefficient.

Expanding (3) to a state vector $x'_t = [x_{t,1}, \dots, x_{t,nx}]$, where nx is the number of sites, results in the space-time correlation matrix $\Gamma_{XX}(l)$ of

$$\Gamma_{XX}(l) = \Gamma_X(l) = E[x_t x'_{t-l}] = \begin{bmatrix} E[x_{t,1} x_{t-l,1}] & \dots & E[x_{t,1} x_{t-l,nx}] \\ \vdots & \ddots & \vdots \\ E[x_{t,nx} x_{t-l,1}] & \dots & E[x_{t,nx} x_{t-l,nx}] \end{bmatrix} = \begin{bmatrix} \gamma_{11}(l) & \dots & \gamma_{1,nx}(l) \\ \vdots & \ddots & \vdots \\ \gamma_{nx,1}(l) & \dots & \gamma_{nx,nx}(l) \end{bmatrix} \quad (4)$$

The $\Gamma_X(l)$ is a $(nx \times nx)$ bounded (± 1) matrix. If the covariate $z'_t = [z_{t,1}, \dots, z_{t,nz}]$ is presented where nz is the number of covariate sites, $\Gamma_Z(l)$ ($nz \times nz$) as well as the corresponding inter-space correlation matrices $\Gamma_{XZ}(l)$ ($nx \times nz$) and $\Gamma_{ZX}(l)$ ($nz \times nx$) can also be defined similar to (3), where the equality $\Gamma_{XZ}(l) = \Gamma_{ZX}(-l)'$ holds. The space-time correlation matrix is not only useful in displaying the correlations systematically, but also can be applied for the parameter estimates by the method of moment. For instance of the latter case, let us assume that we have a multivariate AR(1) model $x_t = \phi x_{t-1} + w_t$, where ϕ is the parameter matrix, and w_t is the model noise.

Then, the Yule-Walker equation is written by $\Gamma_x(-1)=\phi\Gamma_x(0)$, from which the moment estimate ϕ can be obtained by solving the system of equation.

Finally, defining $\gamma_k(l,d)$, $k=1,\dots,N(l,d)$, as the k -th $\gamma_{ij}(l)$ whose Euclidean distance between site i in m -space domain and j in n -space domain is $d\pm\delta$ with δ is a bound of d , a four-dimensional space-time correlation function (STCF) $\xi(m,n,l,d)$ can be given by

$$\xi(m,n,l,d) = \frac{1}{N(l,d)} \sum_{k=1}^{N(l,d)} \gamma_k(l,d). \quad (5)$$

Computationally, the function $\xi(m,n,l,d)$ with respect to d can be obtained easily by the curve fitting of $\gamma_k(l,d)$'s versus d 's. Like the correlation coefficient, the STCF values range from +1 to -1.

Another commonly used spatial function is the semi-variogram. Under the assumptions of the second-order spatial stationary and $E[x_{t,i}-x_{t-l,j}]=0$, the variogram for the time-lag l can be defined by

$$\begin{aligned} \eta(m,n,l,d) &= \frac{1}{2} \text{var}[x_{t,i} - x_{t-l,j}] \\ &= \frac{1}{2N(l,d)} \sum_{k=1}^{N(l,d)} (x_{t,i} - x_{t-l,j})^2 \end{aligned} \quad (6)$$

where $N(l,d)$ is the number of paired-measurements which has a time lag l and distance d apart, and $\text{var}[\cdot]$ is the variance operator. The estimated $\eta(m,n,l,d)$'s with respect to d 's from the sample are commonly fitted to a theoretical functional model. The widely used functional models are linear, polynomial, exponential, Gaussian, and spherical models. Subsequent sections will show both the fitted SFDCF and variogram functions to the multi-aquifer groundwater heads as well as rainfall spaces.

m\n	Rainfall	Layer 1	Layer 2	Layer 3	Layer 4
Rainfall	(a) $\xi(R,R,l,d)$ or $\eta(R,R,l,d)$				
Layer 1	(b) $\xi(1,R,l,d)$ or $\eta(1,R,l,d)$	(c) $\xi(1,1,l,d)$ or $\eta(1,1,l,d)$	Figure 5(a) $\xi(1,2,l,d)$		
Layer 2	(d) $\xi(2,R,l,d)$ or $\eta(2,R,l,d)$	(e) $\xi(2,1,l,d)$ or $\eta(2,1,l,d)$	(f) $\xi(2,2,l,d)$ or $\eta(2,2,l,d)$	Figure 5(b) $\xi(2,3,l,d)$	
Layer 3	(g) $\xi(3,R,l,d)$ or $\eta(3,R,l,d)$		(h) $\xi(3,2,l,d)$ or $\eta(3,2,l,d)$	(i) $\xi(3,3,l,d)$ or $\eta(3,3,l,d)$	Figure 5(c) $\xi(3,4,l,d)$
Layer 4	(j) $\xi(4,R,l,d)$ or $\eta(4,R,l,d)$			(k) $\xi(4,3,l,d)$ or $\eta(4,3,l,d)$	(l) $\xi(4,4,l,d)$ or $\eta(4,4,l,d)$

Figure 3. An inter-space matrix with the block notations of both space-time correlation functions and semi-variograms.

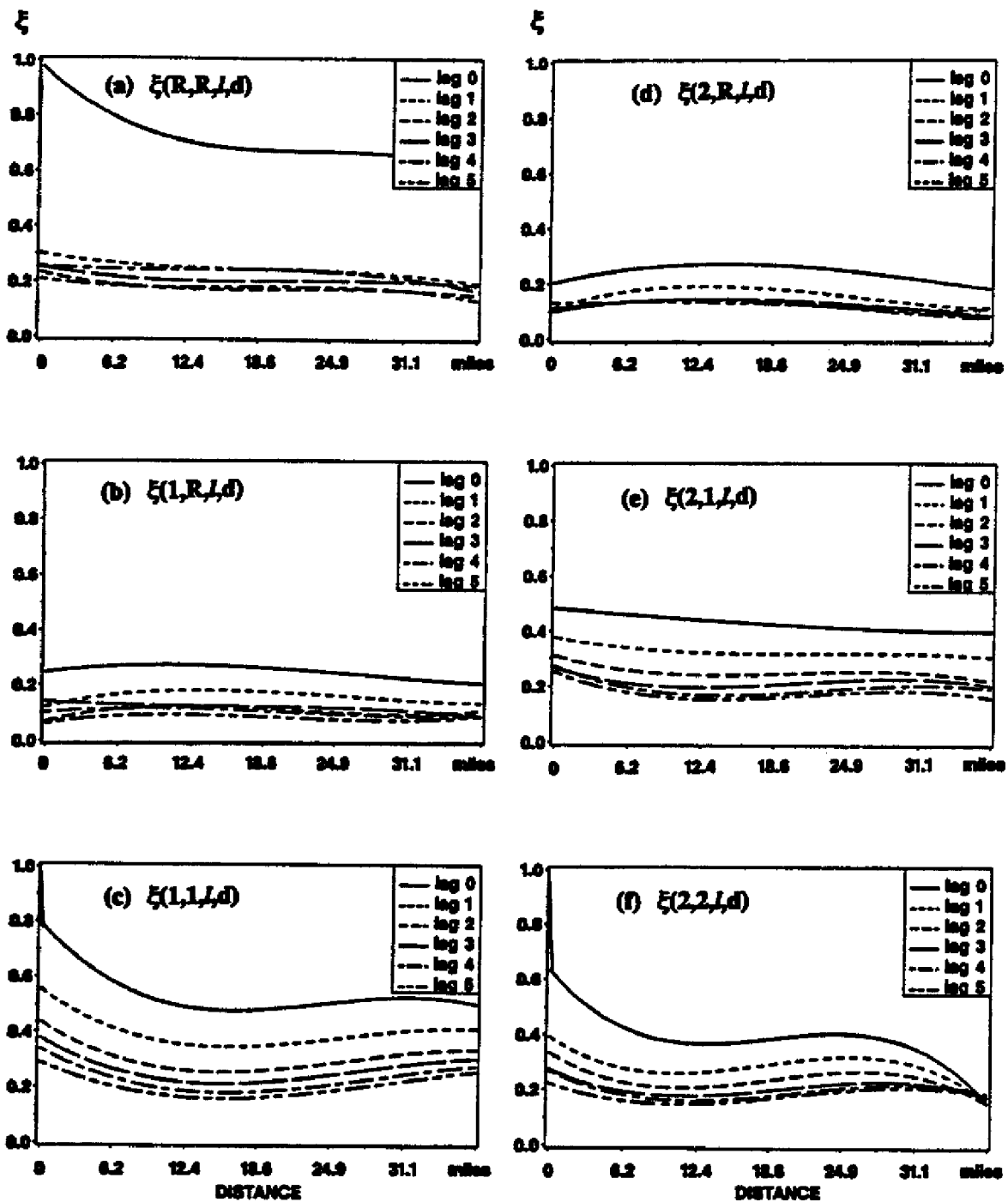
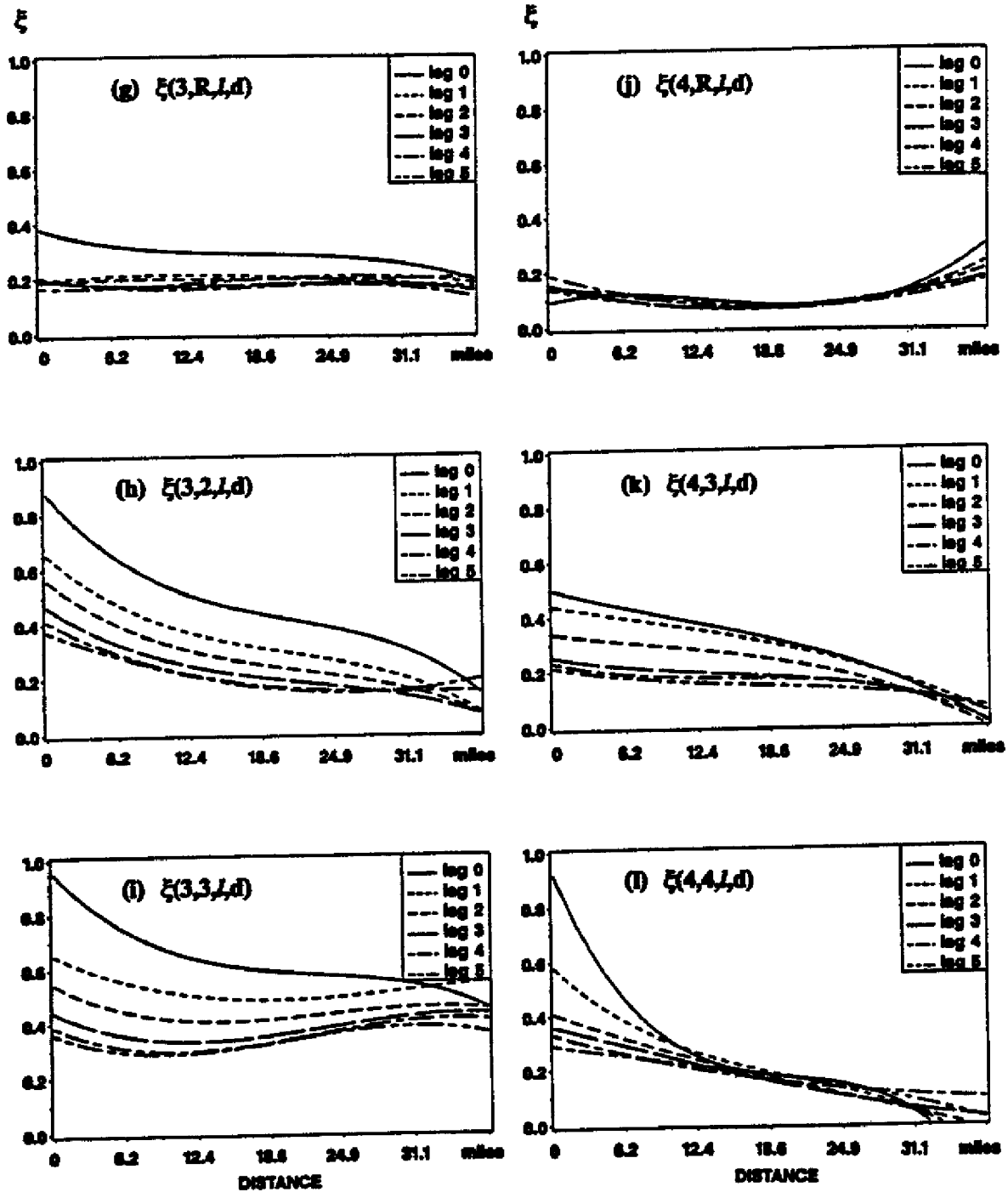


Figure 4. The fitted space-time correlation functions of groundwater head and rainfall.

(Figure 4 continue)



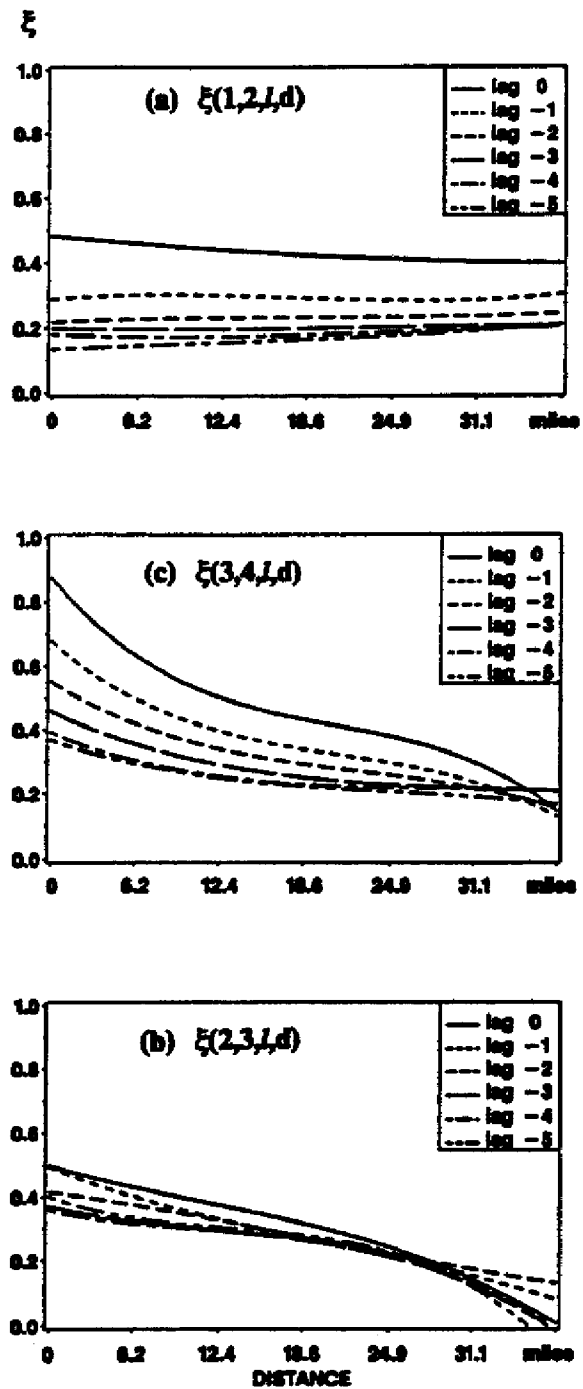


Figure 5. Space-time correlation functions for the upper-triangular blocks.

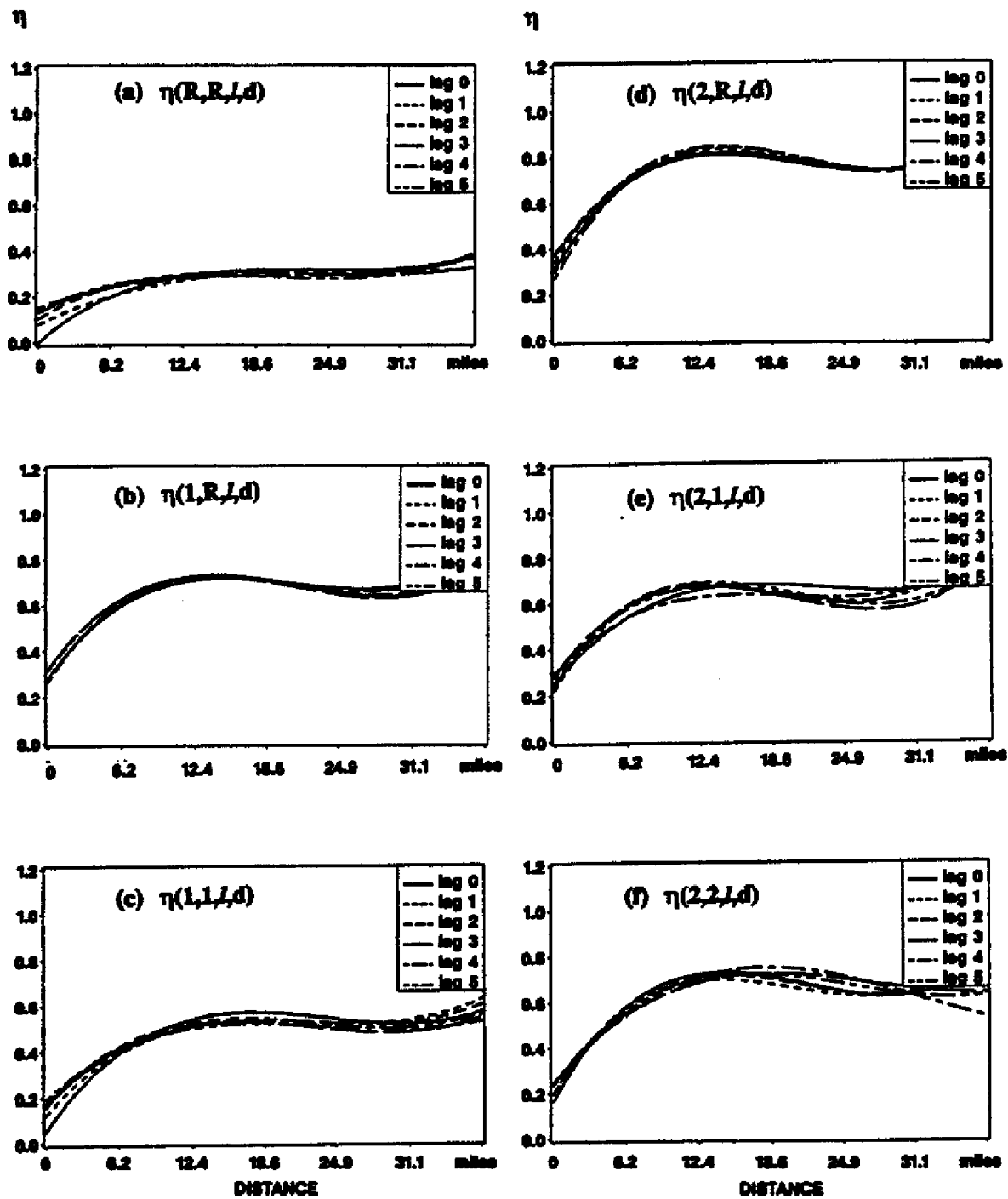
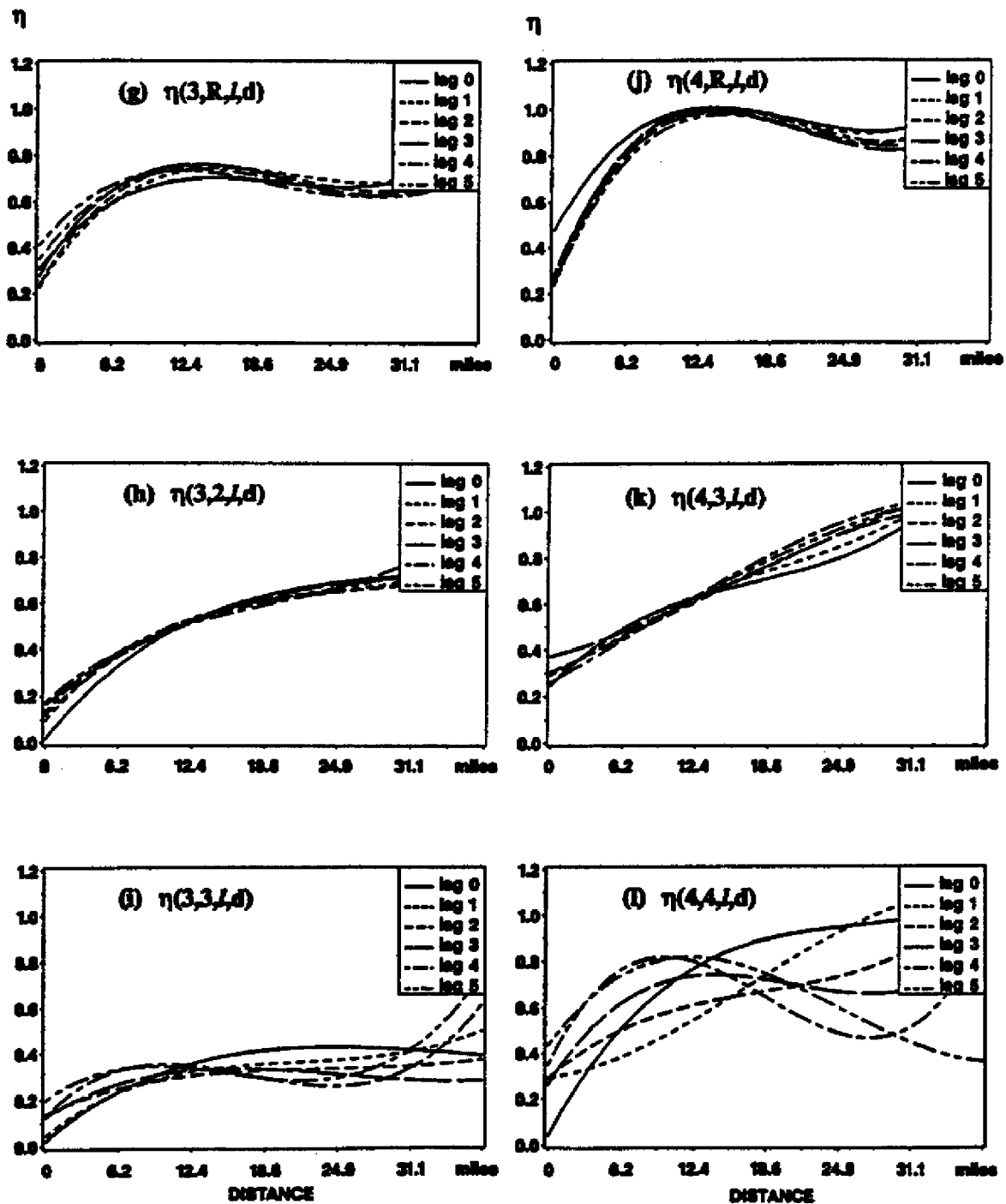


Figure 6. Time lagged semi-variogram functions of groundwater head and rainfall.

(Figure 6 continue)



Results and Discussions.

Even though rainfall is not used to forecast the future groundwater head here, it is worthwhile to investigate the correlations between groundwater heads and rainfall since rainfall is a main driving factor in the groundwater flow system. In order to compute the STCF $\xi(m,n,l,d)$'s, the rainfall gaging stations as well as groundwater monitoring stations from Layer 1, 2, 3, and 4 in the model area were stacked sequentially to form a vector X (133×1) so that $E[XX']$ produces a covariance matrix having 25 block matrices. The notations of the block matrices are shown in Figure 3. From the covariance matrix associated with the Euclidean distance d , the fitted function $\xi(m,n,l,d)$ versus d for each time lag $l(=0, \dots, 5)$ was computed by the cubic regression analysis. For example, when $m=1$, $n=1$, and $l=0$, a fitted STCF curve is given by

$$\xi(1,1,0,d) = \begin{cases} 0.789 - 0.0441d + 0.0019d^2 - 0.000026d^3 & \text{if } d > 0.0 \\ 1.0 & \text{if } d = 0.0 \end{cases} \quad (7)$$

The fitted curves of $\xi(m,n,l,d)$ versus d for the lower triangular blocks in Figure 3 are shown in Figure 4. The block elements $\xi(3,1,l,d)$, $\xi(4,1,l,d)$, and $\xi(4,2,l,d)$ were excluded in Figure 4 because those are not adjacent layers. Also, Figure 5 shows the fitted $\xi(m,n,l,d)$'s which are from the upper triangular blocks in Figure 3 for comparison with Figure 4. It should be noted that the STCFs for Layer 4 was somewhat inconsistent due to the small number of gaging stations. Figure 6 shows the fitted semi-variograms with respect to distance for different time lags.

From both STCF and variogram curves, the following conclusions on the spatial correlation structures were drawn. The STCFs of the monthly rainfall in the region were very high (>0.65). However the temporal correlations on them were insignificant. In other words, the spatial variation of the monthly rainfall within the scale of a county was not dominant. Furthermore, the correlation coefficient between rainfall and groundwater head were generally lower than that of neighboring layer's heads. The STCFs of groundwater head decreased proportional to both temporal lag l and the Euclidean distance d , which were distinct for low temporal orders, but

when $l > 2$ those function were not significantly varied. In terms of spatially averaged STCF's, the equality of $\xi(m,n,l,d) = \xi(n,m,l,d)$ holds, even though each element was somewhat different. That is, a matrix in Figure 3 was symmetric in the sense of spatial scale. Unlike the STCFs, the semi-variogram functions did not show any differences between the time lags, implying that this function can hardly be used for time series analysis. In most cases, the range of the semi-variogram was about 12 miles, with sill values ranging from 0.5 to 1.0.

5. Treatments of the Raw Data

The underlying assumptions of the STARX model is that both the variate $x_{t,i}$ and covariate $z_{t,i}$ are independent and identically distributed (i.i.d.) normal random variables, and that time series should be stationary in time. To meet the stationary condition, the time series should not include any temporal trend, inconsistency, or seasonality. There exists a variety of statistical tests to detect the above conditions. When the above assumptions are not meet, the stochastic time series model should be developed either with those abnormality conditions, or after appropriate transformations of raw data. Examples of the former cases include the ARIMA or the seasonal ARMA models, in which the modeling procedure is more complicated and a dimensional increase is inevitable. Thus, the later transformation method is preferred for the STARX model since the dimension of the STARX model is critical for the parameter calibration.

5.1. Seasonality

Seasonality or periodicity in time series is a regular change in the data values that occurs at the same time in a given period, so it can be characterized by a strong serial correlation at the seasonal lag. Typically, the monthly groundwater head in the model area had a strong seasonality as shown in the Figure 7. The monthly means showed higher head during the wet season due to intensive summer rainfall.

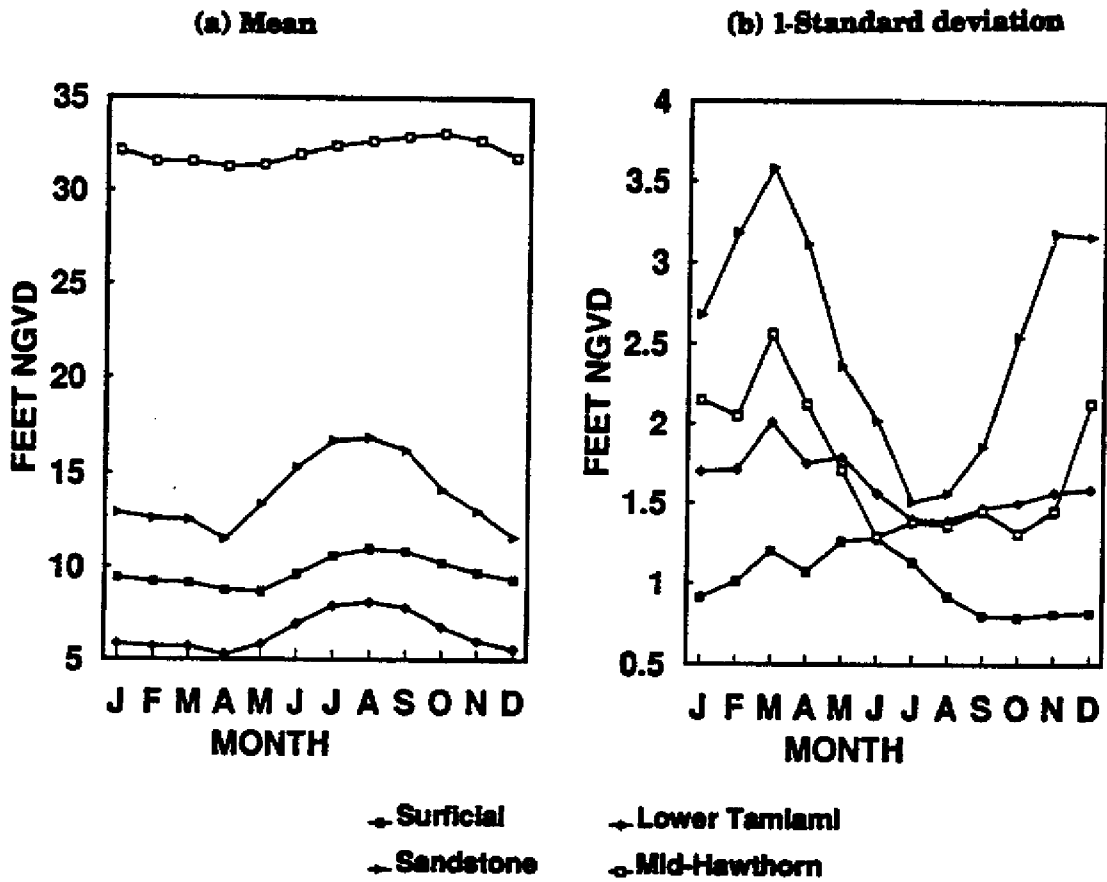


Figure 7. Spatially averaged monthly mean and standard deviation of groundwater heads.

However, the monthly standard deviations during the wet season were smaller than those of the dry season except for the Surficial aquifer. The reason is that the regional summer rainfall induces near maximum groundwater heads in each aquifer.

Several methods are available to handle the seasonality. One approach is to take differences, using a 12 month time lag for monthly time steps, and to model by the ARIMA procedure [Box and Jenkins, 1976, chapter 4; Salas et al., 1985, chapter 6]. A deficiency of this approach is that it loses a year of data by the seasonal differencing, which is critical when the period of record is short. Another approach is to use the seasonal parameters in the time series model as introduced by Salas et al. [1985, section 5.3], but this approach increases the number of parameters dramatically and is not applicable to the large scale STARX model. A third approach to seasonality is to include indicator variables for the season as explanatory variables. This approach is very simple in nature and does not increase the number of model parameters, thus this study adopts the multiplicative explanatory variable approach by the standardization procedure introduced by Salas et al. [1980, page 241] as follows.

Consider a time series $u_{t,i}$ (untransformed) at a site i is measured with an equal time step t with $t=1, \dots, T$. This series can be rewritten in the form of $u_{\tau,v,i}$ using the seasonal notation $\tau(1, \dots, \omega)$ and year $v(=1, \dots, N)$, where ω is the number of season in a year, N is the period of record in years, and the number of measurement T is ωN . Then, the moment estimate of the seasonal mean $\mu_{\tau,i}$ is

$$\mu_{\tau,i} = \frac{1}{N} \sum_{v=1}^N u_{\tau,v,i} \quad (8)$$

and the unbiased seasonal standard deviation $\sigma_{\tau,i}$ is

$$\sigma_{\tau,i} = \sqrt{\frac{1}{N-1} \sum_{v=1}^N (u_{\tau,v,i} - \mu_{\tau,i})^2} \quad (9)$$

with $\tau=1, \dots, 12$ when a monthly time step is used. The transformed time series $x_{\tau,v,i}$ after removing the seasonality, or so-called the standardized series, at site i is given by

$$x_{\tau,v,j} = \frac{u_{\tau,v,j} - \mu_{\tau,j}}{\sigma_{\tau,j}} \quad (10)$$

After standardization, the series $x_{\tau,v,i}$ is considered $x_{t,i}$, $t=1, \dots, T$, to fit for the STARX model. This standardized series $x_{t,i}$ has a zero mean and variance of one if the series meet the normality condition. For standardized series, the correlation coefficient is the same as covariance.

5.2. Normality

Several statistical tests are available for testing the hypothesis that a given series is normally distributed. The skewness test of normality given by *Salas et al.* [1985, page 92] is based on the fact that the skewness coefficient for a normal variable is zero. An estimation of the skewness of a time series $u_{t,i}$, $t=1, \dots, T$ is

$$\gamma_i = \frac{\frac{1}{T} \sum_{t=1}^T (u_{t,i} - \mu_i)^3}{\left[\frac{1}{T} \sum_{t=1}^T (u_{t,i} - \mu_i)^2 \right]^{3/2}} \quad (11)$$

where μ_i is the sample mean for site i . The $(1-\alpha)$ probability limits may be defined by $\pm u_{1-\alpha/2}(6/T)^{1/2}$, where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. If the estimated γ_i is within the probability limit, the hypothesis of normality is accepted, or otherwise, rejected. In case of non-normal series, several transformation techniques are available.

Since some groundwater head time series do not satisfy the hypothesis of normality, the Box-Cox transformation [*Bras and Rodriguez-Iturbe*, 1985, page 73] was performed, which is given by

$$x_{i,j} = \begin{cases} \frac{(u_{i,j})^{\lambda_i} - 1}{\lambda_i} & \lambda_i \neq 0 \\ \ln(u_{i,j}) & \lambda_i = 0 \end{cases} \quad (12)$$

where λ_i is the transformation parameter at site i to be estimated. Using a simple trial and error method, optimal λ_i 's, $i=1, \dots, 115$, were calibrated, whose values range from -0.6 to 2.2, with an average of 0.43. After the Box-Cox transformation, the γ_i 's (mean \pm 1 standard deviation) are improved from -0.065 ± 0.35 to -0.042 ± 0.26 , while the number of sites which were accepted by the normality assumption with $\alpha=0.1$ was increased from 92 sites to 106 sites (92.2% accepted). In addition, the other transformations, including logarithm, exponential, and power transformations, were conducted, but the results demonstrated that the Box-Cox transformation was superior to the other methods.

5.3 Temporal Trend

Stochastic time series analysis is generally based on the assumption of stationarity in time. The first order stationarity means that the expected values do not vary with time, while the second order stationarity is defined as stationary in both mean and covariance [Salas *et al.* 1985, page 3]. It should be noted that stationarity in data is different from the stationarity condition of the time series model itself. As an example of the latter case, to ensure the second order stationarity condition of the autoregressive model $x_t = \psi(B)w_t$, a characteristic polynomial $\psi(B)$ should be on or within the unit circle.

One of the common non-stationarity in the groundwater head data is a temporal trend which is a long, consistent change in the time series values from beginning to end. Temporal trends in data can be detected by visual inspection of time series plots, differenced data, autocorrelations by parts, or by the test of the unit-root hypothesis developed by *Dickey and Fuller* [1979]. Temporal trend can be modeled by either linear, quadratic, or cubic patterns. A simple linear trend model introduced by *Shumway* [1988 page 124] is

$$u_{t,i} = \beta_{0,i} + \beta_{1,i} t + e_{t,i} \quad (13)$$

where $t=1,\dots,T$, $e_{t,i}$ is the model noise, $\beta_{0,i}$ and $\beta_{1,i}$ are regression parameters, which can be estimated by

$$\beta_{1,i} = \frac{\sum_{t=1}^T (t-t^*)(u_{t,i}-u_i^*)}{\sum_{t=1}^T (t-t^*)^2} \quad (14)$$

$$\beta_{0,i} = u_i^* - \beta_{1,i} t^* \quad (15)$$

where t^* and u_i^* are the sample means over the times axis and observed series, respectively. Then, the detrended series can be obtained by

$$\begin{aligned} x_{t,i} &= u_{t,i} - \beta_{0,i} - \beta_{1,i} t \\ &= (u_{t,i} - u_i^*) - \beta_{1,i} (t-t^*) \end{aligned} \quad (16)$$

After standardization and the Box-Cox transformation, the temporal trends were removed before the calibration of STARX model. For $i=1,\dots,115$, the estimated $\beta_{0,i}$'s have -0.04649 and the standard deviation of 0.1437, and a maximum of 0.318, while the estimated $\beta_{1,i}$'s have a mean of 0.0036, a standard deviation of 0.00899, and a maximum of 0.028.

Since time series were transformed to remove seasonality, non-normality, and temporal trend, the forecasted time series should be back transformed for the comparison and control of the groundwater drought problem. Those back transformations can be done sequentially by inverting the transformation equations in terms of the original series $u_{t,i}$.

6. Designing the Forecasting Model Structure

To illustrate groundwater interactions in the multi-layered aquifer system in the model area, let us define h_1 , h_2 , h_3 , and h_4 as the representative groundwater heads at Layers 1, 2, 3, and 4, respectively. The historical data showed the relationships of; $h_1 > h_2$; $h_2 > h_3$; and $(h_1, h_2, \text{ or } h_3) \ll h_4$. Similar to the land surface elevation, the groundwater heads in Layer 1, 2, and 3 ranged from about 30 feet NGVD in the northeast region to near zero feet NGVD in the southwest coastal region, and the groundwater flow direction was generally from the northeast to the southwest. Heads in Layer 4 ranged approximately from 40 feet NGVD at the eastern edge of model area to 25 feet NGVD at the west coastal region.

A volumetric water budget analysis by *Bennett* [1992] showed that Layer 1 receives 95% of its inflow from local rainfall and releases about 16% of its outflow to Layer 2. Layer 2 receives about 82% of its inflow from the Layer 1 and releases only 9% of its outflow to Layer 3. Layer 3 receives 54% of its inflow from Layer 2 and 43% of it from Layer 4 while releasing 40% of its outflow to Layer 2. Layer 4 receives most of its inflow from through upward leakage from deeper aquifers and releases most of its outflow to Layer 3. These results indicate that Layer 1 and 4 are not significantly affected by the adjacent layers, while Layer 2 is influenced by Layer 1, and Layer 3 is affected by both Layers 2 and 4. The correlation analysis in Section 4 also support these conclusions.

Experimentation with the EMSSE algorithm in Part I showed that a system having more than 50 state variables is extremely inefficient in calibration. Thus, instead of building a STARX model comprising all aquifer layers at once, the following three separate STARX models were developed:

- Model I : for Layer 1, with no-covariate,
- Model II : for Layer 2, with Layer 1 as covariate, and
- Model III : for Layers 3 and 4, with Layer 2 as covariate.

Although this approach contradicts simultaneous simulation and forecasting of the multi-aquifer system, it accounts for the main cause and effect of groundwater flows with minimizing the structural error of the model.

6.1. Determination of Neighbor Sites by the Thiessen Polygons.

A matrix form of the STARX model describing the current groundwater head vector x_t in terms of the previous heads x_{t-1}, \dots, x_{t-Nq} and the covariate of the above layer heads $z_t, z_{t-1}, \dots, z_{t-Nk}$ may be expressed in the form

$$x_t = \sum_{i=1}^{Nq} D_i \circ \Lambda_i x_{t-i} + \sum_{j=0}^{Nk} E_j \circ \Omega_j z_{t-j} + w_t \quad (17)$$

where Nq and Nk are the temporal order of regressions for x_t and z_t , respectively, Λ_i ($nx \times nx$) and Ω_j ($nx \times nz$) are the parameter matrices, D_i ($nx \times nx$) and E_j ($nx \times nz$) are the spatial index matrix (SIM) for state vector and covariate, respectively, w_t is the ($nx \times 1$) white noise vector having covariance Q , and notation (\circ) is the Hadamard product that is the element-wise product of the two same size matrices. Both D_i and E_j in the above equation represent the spatial correlation structure of the system, in the forms of the 0-1 binary elements. For instance, $\{d_{i,mn}\}$, which is an element in matrix D_i , is 1 if site n is the i -th temporal order neighbor of the site of interest m (abuttal), and 0 otherwise (non-abuttal).

The factors in determining neighbor sites are the Euclidean distance of two sites, direction from the site of interest, spatial statistics such as variogram or spatial correlation, etc. For a regularly spaced system, neighbors having the same spatial order can be determined objectively by the equi-distance. But for an irregularly spaced system, choosing either spatial order or neighbors is difficult. Among several possible options, the Thiessen polygon method was used here because this method incorporates both distance and directional components simultaneously. Thiessen polygon is a convex polygon whose boundary defines the area that is closest to the point of interest relative to all other points [*Environmental System Research Institute, 1992*]. It

is geometrically defined by the perpendicular bisectors of the lines which connect all neighbor sites as shown in Figure 8(a), which is directly used to construct D_i for Model I. When the covariate z_i is used in a STARX model, the covariate Thiessen polygons should be constructed site by site. That is, if the n_x state vector and n_z covariate vector are used, a total n_x maps should be prepared each of which is constructed on the covariate space with adding a station from the state vector. Figure 8(b) shows some examples of the Thiessen polygons of the covariate space in the Model II. For instance, by introducing site 52 (also 78, or 89) into the covariate space of layer 1, a new polygon (hatched) for site 52 was created as in Figure 8(b).

From these polygons, all neighbors whose boundaries are joined by lines are first order neighbors. If a pair of neighbors is joined by a point (for example, site ID 7 and 12, or 26 and 40, in Figure 9), they are considered second order neighbors, as in the case of a regularly spaced system. Even though first order neighbors can easily be found, neighbors having a spatial orders greater than one are very difficult to determine. An advantage of using the Thiessen polygon method in determining the neighbors is that this method is objective since the Thiessen polygon is uniquely determined, which cannot be achieved for the nearest distance criteria. For instance, the nearest distance criteria cannot link the dense network system located in the western area with the sparse network in the eastern portion of the model area.

LEGEND

- Thiessen Polygon
- Monitoring Site
- ▨ Polygon for Covariate

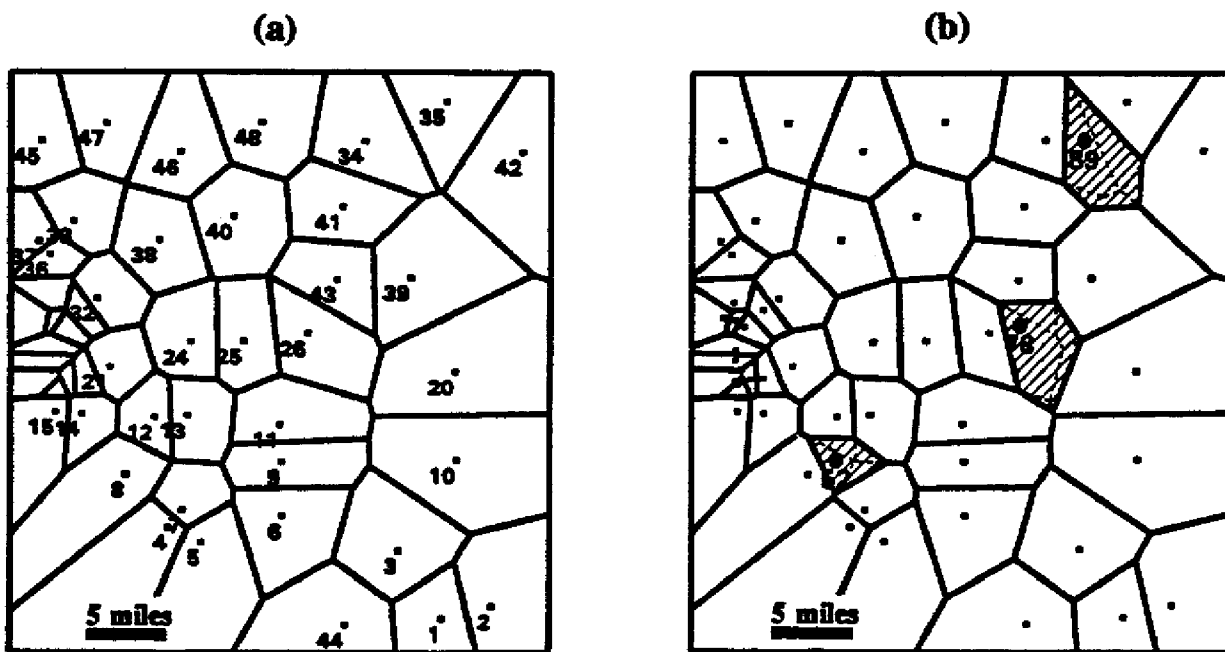


Figure 8. Thiessen polygons with their gaging station ID numbers: (a) For D_i matrix in Model I (Layer 1), (b) For E_i matrix in Model II, where the hatched areas are the covariate Thiessen polygons created by variate sites (site ID 52, 78, and 89) of Layer 2, while the dashed lines show the original polygons of Layer 1.

6.2. Statistics on the First Order Neighbors

The Thiessen polygons can be classified into two classes: The exterior polygon which faces directly to the model boundary, and the interior polygon which is surrounded completely by other polygons. From the Thiessen polygon map of each aquifer layer, the first order neighbors of each site were counted which are listed in Table 2. This table revealed that the most probable number of neighbors of the first spatial order for the interior polygons is either 5 or 6 with a range from 4 to 8, while that for the exterior polygons is 3 or 4 with a range from 1 to 6.

To investigate the spatial neighborhood statistics, let us define that n_i^M (The superscript M is neither power nor combinatorial notations) is the number of polygons, each of which has i neighbors defined by the M spatial order. The sum of all possible n_i^M , $i=1, \dots, \infty$, is equal to the number of gaging stations nx . Let us also assume that the selected neighbors of each station does not exceed the limiting number of stations n . If a polygon has more than n neighbors, only n neighbors are selected for the given station based on either direction or distance criteria. This scheme is useful to restrict the number of neighbors when they are a large number which is not desirable for the STARX model. With this neighbor definition, the number of total neighbors within a space is $(nx \times n)$ and the total number of neighbors excluded by the limiting maximum neighbor n is $\sum_{k=1}^{n-1} n_k^M (n-k)$. Then, N_n^M , which is the total number of neighbors selected by the limiting neighbor numbers n , is given by

$$N_n^M = (nx \times n) - \sum_{k=1}^{n-1} n_k^M (n-k) \quad (18)$$

and p_n^M , which is the probability of being selected by the restriction scheme versus that of the non-restriction scheme, is given by

$$p_n^M = \frac{N_n^M}{N_{n \rightarrow \infty}^M} \quad (19)$$

The p_n^M is the probability of neighbors included when the n uniform neighbor scheme with M spatial order is defined. The statistics of p_n^M are also listed in Table 2. This table shows that if the gaging network is large enough (meaning sufficient gaging stations) the most probable number of the n was 5. p_n^M statistics showed that the first spatial order with a 5-uniform-neighbors scheme covered approximately over 90% of the all first order stations. Likewise, that of 6-neighbor scheme covered over 95% of the first order stations.

Table 2. Statistics of the first spatial order neighbors.

n	1	2	3	4	5	6	7	8	sum
Layer 1									
n_n^M for interior polygons,A	0	0	0	4	10	6	5	1	25
n_n^M for exterior polygons,B	0	1	8	7	5	1	0		23
$n_n^M, (A+B)$	0	1	8	11	15	7	5	1	48
N_n^M	48	96	143	182	210	223	229	230	23
$p_n^M(\%)$	21	42	66	79	91	97	100	100	0
Layer 2									
n_n^M for interior polygons,A	0	0	0	4	7	6	5	1	23
n_n^M for exterior polygons,B	0	4	5	7	3	1	0		20
$n_n^M, (A+B)$	0	4	5	11	10	7	5	1	43
N_n^M	43	86	125	159	182	195	201	202	20
$p_n^M(\%)$	21	43	62	79	90	97	100	100	2
Layer 3									
n_n^M for interior polygons,A	0	0	0	0		3			3
n_n^M for exterior polygons,B	1	2	4	4		0			11
$n_n^M, (A+B)$	1	2	4	4	-	3			14
N_n^M	14	27	38	45	48	51	51	51	51
$p_n^M(\%)$	28	53	75	88	94	100	100	100	
Layer 4									
n_n^M for exterior polygons	2	2	3	2	1	-	-	-	10
N_n^M	10	18	24	27	28	28	28	28	28
$p_n^M(\%)$	36	64	86	96	100	100	100	100	

7. Calibration of the STARX Model

As identification processes of the stochastic time series model, the previous sections discussed characteristics of the groundwater system, groundwater head data, pre-treatment of the data, and spatial structure of the system. *Box and Jenkins* [1976, page 173] also included determination of model orders and preliminary estimation of model parameters as the model identification process. In their framework, identification is performed by a simple estimation method such as the method of moment, then the model parameters are calibrated by the maximum likelihood (ML) method. Since the ML estimates by the EMSSE algorithm was adopted in here, both determination of model order and calibration of model parameters were performed simultaneously. That is, the parameters of several alternative models having different model orders were calibrated with some information criterion statistics, which are used to select the best model order.

For information statistics to determine the model order for the multivariate autoregressive model, either the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) [Shumway, 1988, page 167] can be used. Using the residual sum of products defined by $RSP = \sum_{t=1}^T [w_t' w_t]$, where w_t is the model noise estimated by the ML method, the form of AIC is given by

$$AIC(Nq) = \ln \left| \frac{RSP}{T} \right| + \frac{2 \, nx^2 \, Nq}{T} \quad (21)$$

where nx is the dimension of the state vector, and T is the period of record of data. Especially, $\ln RSP$ is the so called reduced likelihood [Blockwell and Davis, 1987, page 280], which has similarity to the $-2 \ln L$ estimated by the EMSSE algorithm. Note that if the conditional fitting procedure leading to the RSP is applied, T in the above two equations may be replaced by the effective observation ($T - Nq$). For a multivariate autoregressive model, the number of effective parameters is $(nx^2 Nq)$, or for the STARX model with covariate term, the number of effective parameters will be $(nx^2 Nq + nx \times nz \times (Nk + 1))$. The BIC, or the so called Schwarz criterion, chooses

the model order Nq that minimizes

$$BIC(Nq) = \ln \left| \frac{RSP}{T} \right| + \frac{nx^2 Nq \ln T}{T} \quad (20)$$

During the calibration of several alternative models, the ML functions as well as AIC and BIC values were also computed as summarized in Table 3. The period of record of data used for calibration is from January 1987 to December 1992 (72 months). Based on the AIC and the first term of the ML function, STARX(2,no), STARX(1,0), and STARX(3,0) were selected (marked) for Model-I, II, and III, respectively. It should be noted that the determination of model order by the estimated BIC is somewhat different from that of the AIC, indicating that the further study should be focused on those criterion statistics for the family of the multivariate autoregressive models.

Table 3. Test statistics of the alternative models with their estimated ML values.

Nq	Nk	The 1st term of -2lnL,A	The 2nd term of -2lnL,B	-2lnL, A+B	The 2nd term of AIC	AIC	The 2nd term of BIC	BIC
Model I nx=48								
1	-	-13678	2977	-10701	58	-91.0	126	-22.4
2	-	-15956	613	-15343	115	-97.9	252	39.3
3	-	-15638	236	-15402	173	-41.1	379	164.7
4	-	-16288	122	-16166	230	5.9	505	280.3
Model II nx=43, nz=48								
1	-	-11667	3111	-8555	46	-72.6	101	-17.5
2	-	-16558	853	-15705	93	-125.6	203	-15.1
3	-	-17628	344	-17284	139	-101.4	304	63.8
1	0	-17814	1420	-16394	81	-147.1	177	-51.0
2	0	-15076	334	-14742	127	-77.9	278	73.2
3	0	-15792	140	-15652	173	-44.3	379	161.8
1	1	-16418	193	-16225	115	-110.3	252	26.7
2	1	-12188	62	-12125	161	-7.2	353	184.9
3	1	-14546	34	-14512	207	5.9	455	253.0
Model III nx=24, nz=43								
1	-	-3673	1871	-1802	14	-10.6	32	6.5
2	-	-5644	1802	-3843	29	-24.6	63	9.7
3	-	-8765	1274	-7491	43	-60.8	95	-9.4
1	0	-5350	1833	-3517	25	-23.6	55	6.4
2	0	-8406	1555	-6852	40	-55.6	87	-8.4
3	0	-9716	600	-9116	54	-72.6	118	-8.3
1	1	-7530	1753	-5777	36	-44.2	79	-1.4
2	1	-9336	662	-8674	50	-70.1	110	-10.0
3	1	-8941	307	-8634	65	-55.1	142	22.1
2	2	-8452	218	-8234	61	-53.2	134	19.7
3	2	-8779	99	8680	76	-45.0	166	45.1
3	3	-6753	15	-6738	86	-7.2	189	95.7

8. Verification

Two of the most common applications of the stochastic time series model are stochastic data generation and forecasting. If certain sample statistics are preserved in either generated or forecasted data, the null hypothesis of building a particular type of stochastic time series model is accepted, or otherwise rejected. Thus, verification of the calibrated STARX model is focused on those two aspects.

8.1. Data generation

In water supply plans or hydraulic structure designs, the frequency of the design hydrologic event is occasionally defined by a yearly or larger time step. However, within that time step, there exist infinitely many realizations when data are discretized by monthly or weekly time steps. In reality, the historical data cannot incorporate all possible realizations due to the limited period of records. Instead, the stochastic data generation method provides an ideal tool to replicate such infinite realizations, each of which possess similar sample statistics. The main mechanism for generating the different realization in the stochastic time series model is the random noise term.

The calibration of the STARX model using the EMSSE algorithm assumes that there exists measurement noise v_t ; then the measurement equation in the state-space form is given by $y_t = M_t x_t + v_t$. In case of data generation, M_t is nothing but an identity matrix. Substituting all x_t 's in equation (17) by $(y_t - v_t)$'s and rearranging with respect to y_t results in the following recursion equation:

$$y_t = \sum_{i=1}^{Nq} D_i \circ \Lambda_i y_{t-i} + \sum_{j=0}^{Nk} E_j \circ \Omega_j z_{t-j} + w_t + v_t - \sum_{i=1}^{Nq} D_i \circ \Lambda_i y_{t-i} \quad (22)$$

where w_t and v_t are the multi-Gaussian random vectors having zero means and variance Q

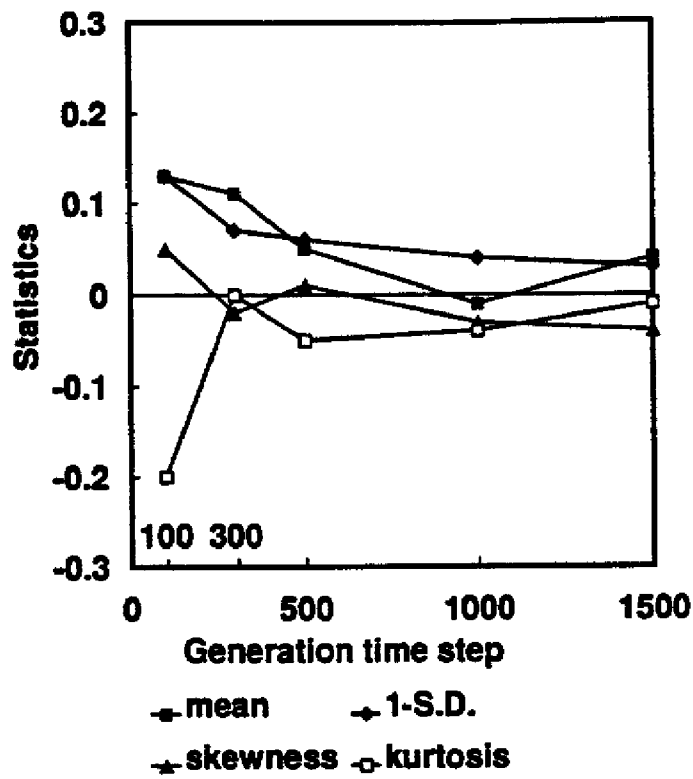


Figure 9. Statistics of the generated data versus time step.

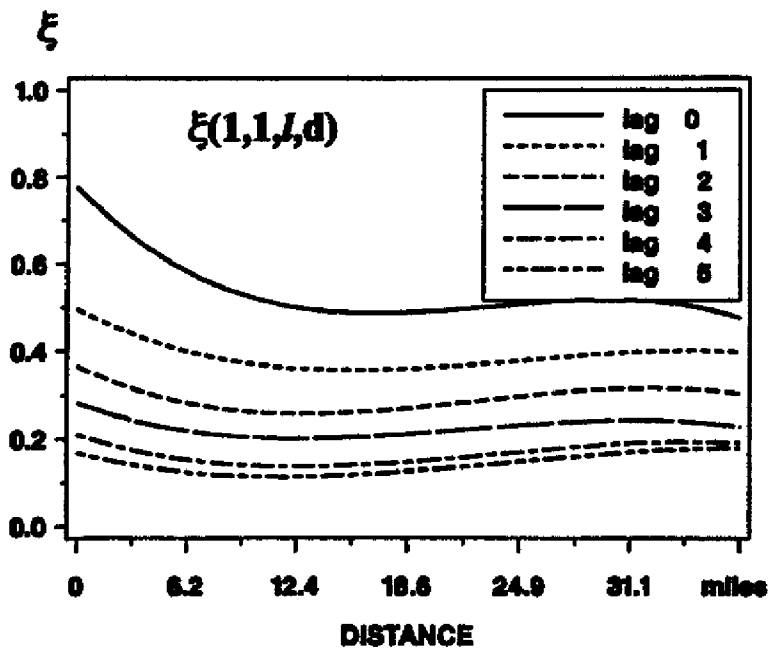


Figure 10. Space-time correlation functions of the generated groundwater heads for the Surficial aquifer (1500 generation time step was used)

and R , respectively. Since $x(0)$ is set to $N(\mu, \Sigma)$, the initial condition of generated series y_0 can be $N(\mu, \Sigma + R)$. With covariate series z_t (which should be generated beforehand) as well as generated noise series w_t and v_t and assumption of initial state y_0 , the state vector y_t are generated by (22), recursively. In order to eliminate the impact of the initial data assumption, the data generated by the first 100 time steps were discarded.

Using the calibrated Model I, five different data sets having different numbers of time steps were generated. Figure 9 shows the spatial averages of the following four moments with respect to the generated time steps: mean, standard deviation (1-SD), skewness, and Kurtosis (coefficient of excess). Theoretical values of the above four statistics should be zero. Results showed asymptotic convergence to zero implying that the first four moments of the time series were preserved by the fitted STARX model. Also, an investigation of preserving the space-time correlation in the generated time series was performed. Figure 10 is the fitted STCF for Layer 1 (Surficial aquifer) computed from the generation data having 1500 time steps. This result can be compared with the original data in Figure 4(c). The comparison revealed that the STCF's with time lags up to 3 were well preserved for distances less than 12 miles, but the STCFs with distances greater than 12 miles were usually lower than that of the actual data. In reality, only the first spatial order neighbor is commonly used in the fitted STARX model. As a summary, the STARX model preserves the first four moments as well as space-time correlations for the designated space-time model orders.

8.2 Forecasting with Its Error Covariance

By definition, an unconditional prediction (ex ante prediction) is called forecasting [Harvey, 1991]. Even though the terms forecasting and prediction have been used interchangeably throughout the previous literatures, this study uses the term forecasting as the prediction of the future by the stochastic time series model without having state measurement after the calibration period. The adoptive mode in modeling uses previous model inputs as well as the previous measured outputs in calculating current model output. The current model output is expressed as a function of previous measured outputs as well as model input, or using the

discrepancy between the last measured and model output as feedback to calculate the future state.

Since the STARX model was calibrated in a state-space format, the Kalman forward recursion can easily be applied to forecasting the future state using the predicted state equation and its error covariance. That is, if $x(t)$ is the stacked state vector $x(t)'=[x_1', \dots, x_{r,Nq}']$, then the forecasted state is given by

$$x(t) = \Phi x(t-1) + \Psi z(t) \quad (23)$$

which is the minimum variance forecasting. The forecasted error variance is also given by

$$p_t^i = \Phi p_{t-1}^{i-1} \Phi' + Q^* \quad (24)$$

where $Q^* = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$ and $Q = E[w_t w_t']$.

In order to verify the forecasting ability of the fitted three STARX models, forecasting was performed at the end of December 1992 using the forecasting lead time $l=1, \dots, 12$. Based on the forecast error covariance p_t^i , or p_0 in the EMSSE algorithm, Table 4 displays the spatial mean and standard deviation of the forecasted error, which is the square root of the diagonal terms of error covariance. The first row, $l < 1$, is for the calibration case, which is constant during the period of calibration since complete data were used. An interesting thing to note is that the forecasting errors increased during the first three or four steps then remain constant. The data in Table 4 is standardized data having a mean of zero and variance of one. Figure 11 shows the contour maps of both historical and forecasted groundwater heads for each layer for the selected months ($l=1, 2$, and 4 month). Due to the irregular distribution of gaging stations, layers below the Surficial Aquifer display only a portion of the heads. This contour maps shows a good match between the forecasted with historical heads for $l=1$, but the forecasting errors increase with respect to the increasing lead times.

Table 4. Spatial means and standard deviations of the forecasting error, p_t^1 .

Lead Time l	Model I		Model II		Model III	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
<1	0.260	0.044	0.251	0.062	0.242	0.066
1	0.421	0.067	0.640	0.232	0.641	0.339
2	0.484	0.071	0.651	0.232	0.684	0.319
3	0.526	0.075	0.652	0.232	0.720	0.292
4	0.544	0.078	0.652	0.232	0.734	0.281
5	0.559	0.081	0.652	0.232	0.738	0.279
6	0.570	0.084	0.652	0.232	0.742	0.276
7	0.579	0.087	0.652	0.232	0.743	0.275
8	0.586	0.092	0.652	0.232	0.744	0.275
9	0.592	0.096	0.652	0.232	0.744	0.274
10	0.596	0.099	0.652	0.232	0.744	0.274
11	0.601	0.103	0.652	0.232	0.744	0.274
12	0.604	0.106	0.652	0.232	0.744	0.274

Ref.: S.D. = 1 standard deviation

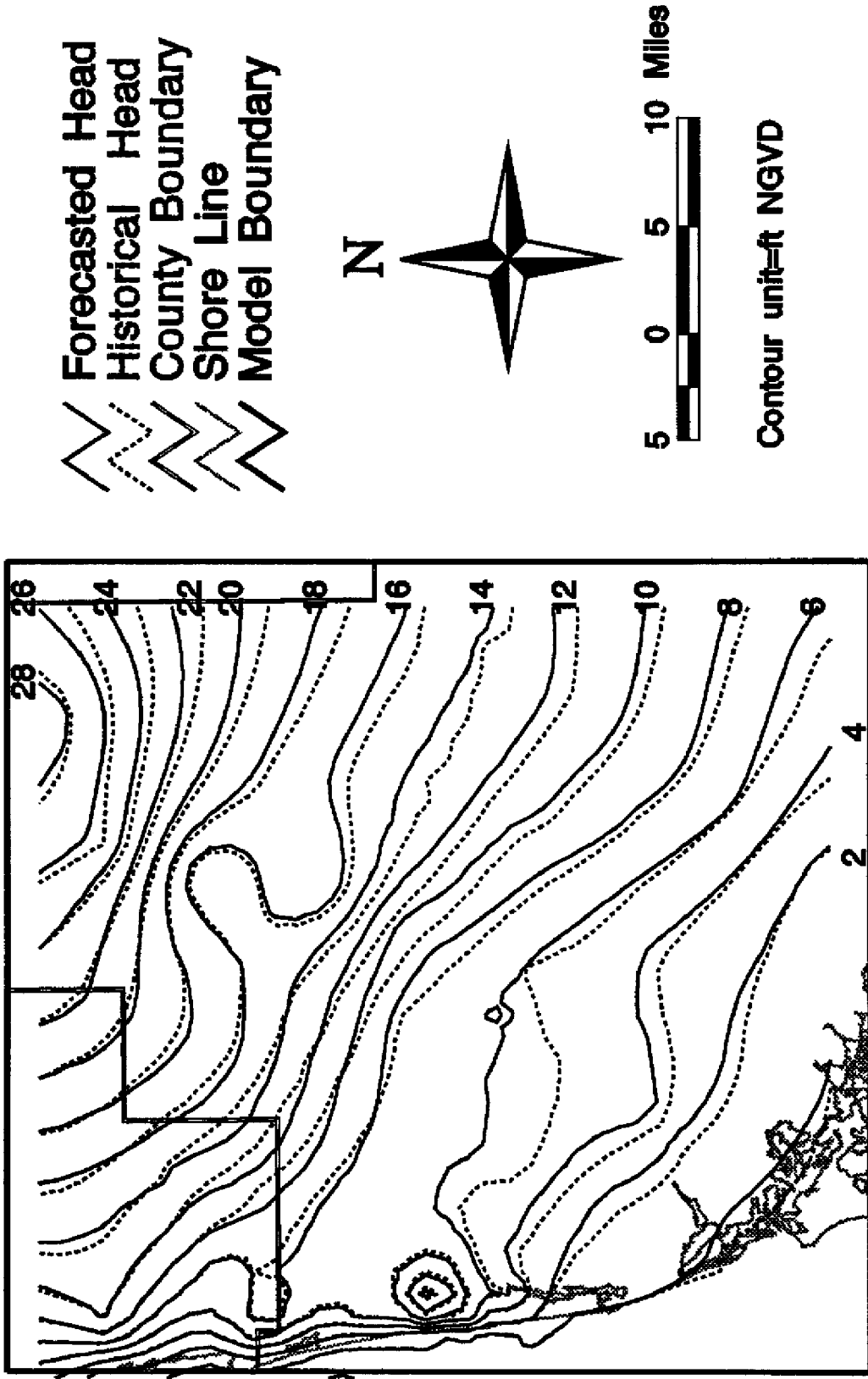


Figure 11. Comparison of forecasted and historical heads for Layer 1 & Lead time 1.

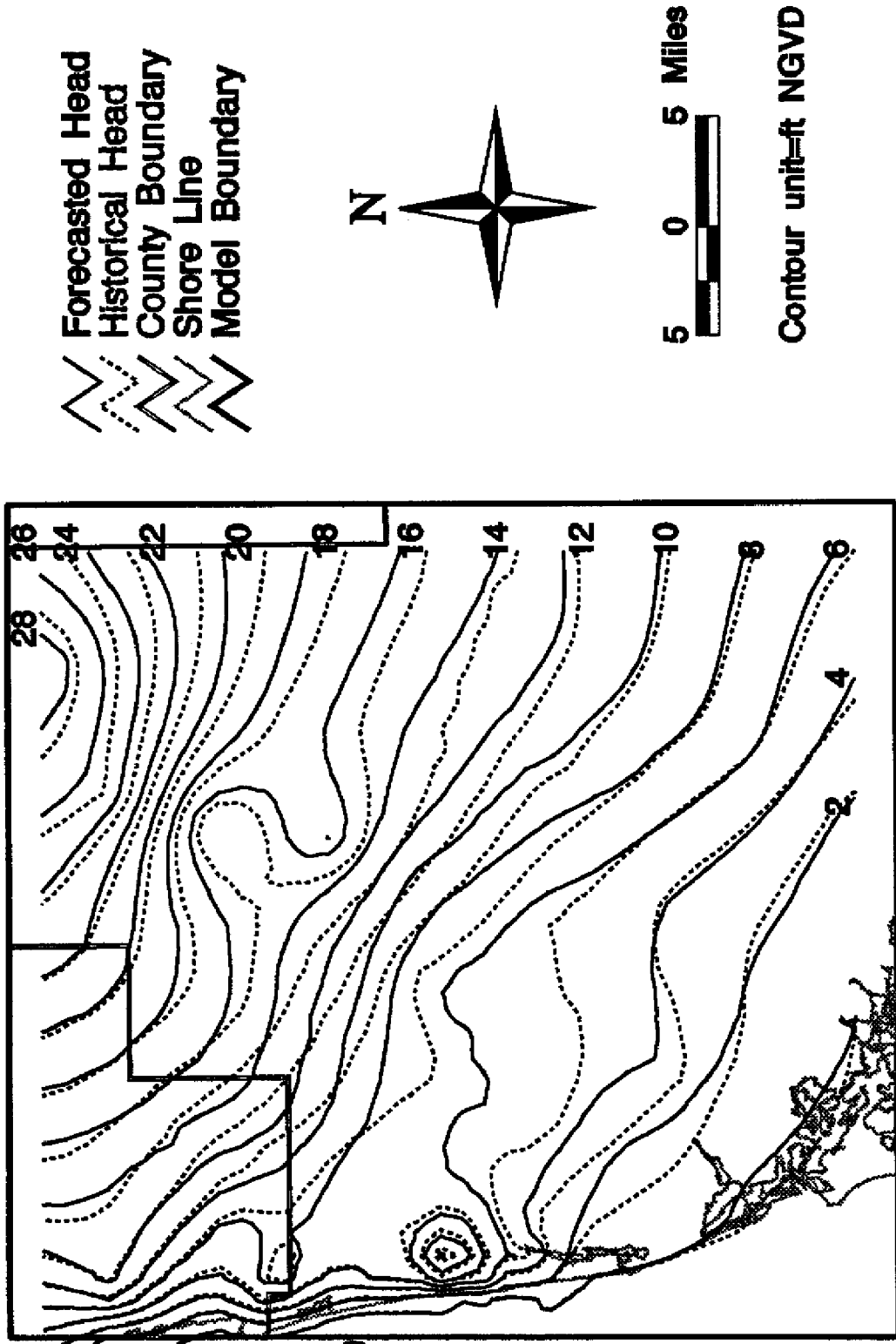


Figure 12. Comparison of forecasted and historical heads for Layer 1 & Lead time 2.

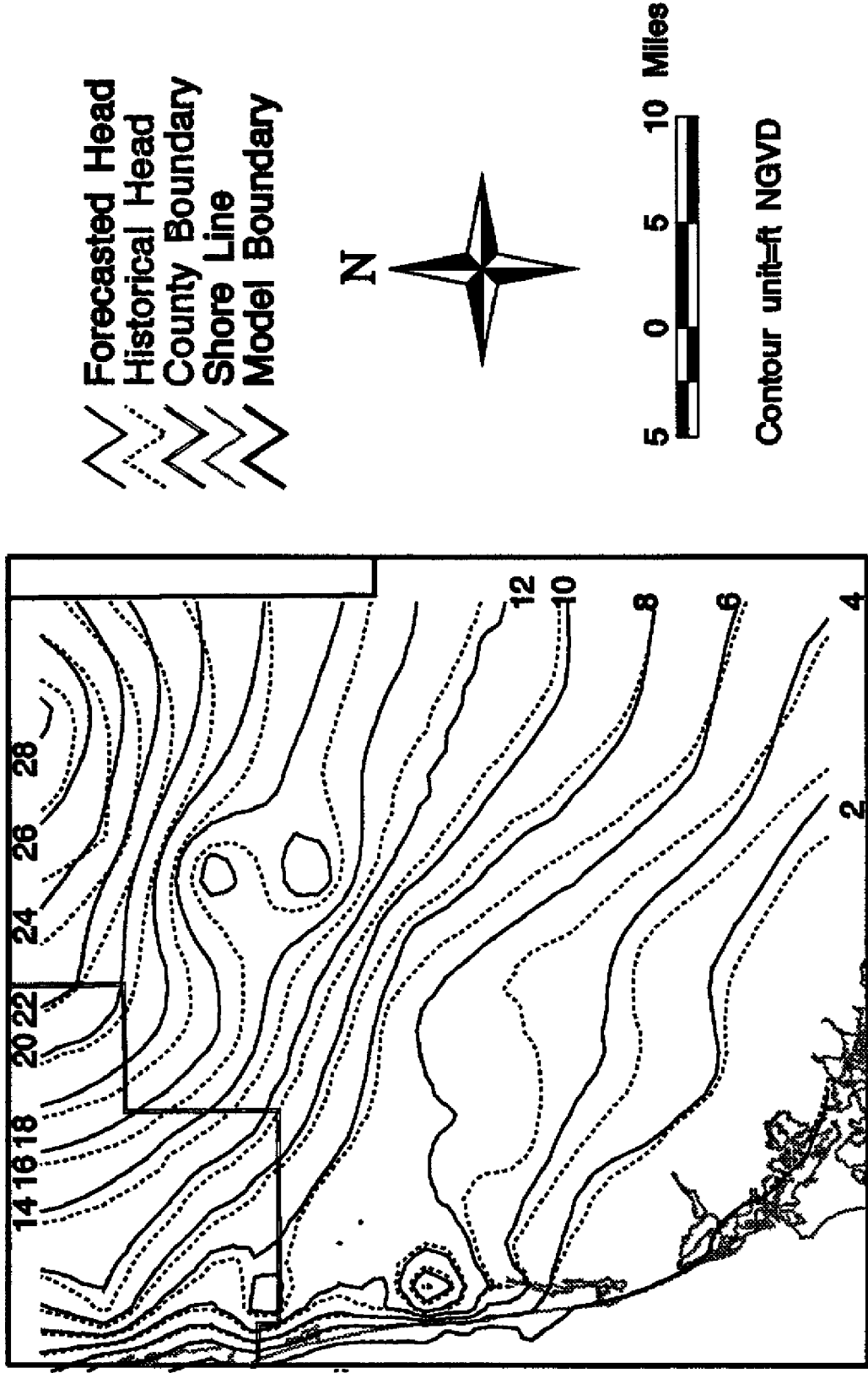


Figure 13. Comparison of forecasted and historical heads for Layer 1 & Lead time 4.

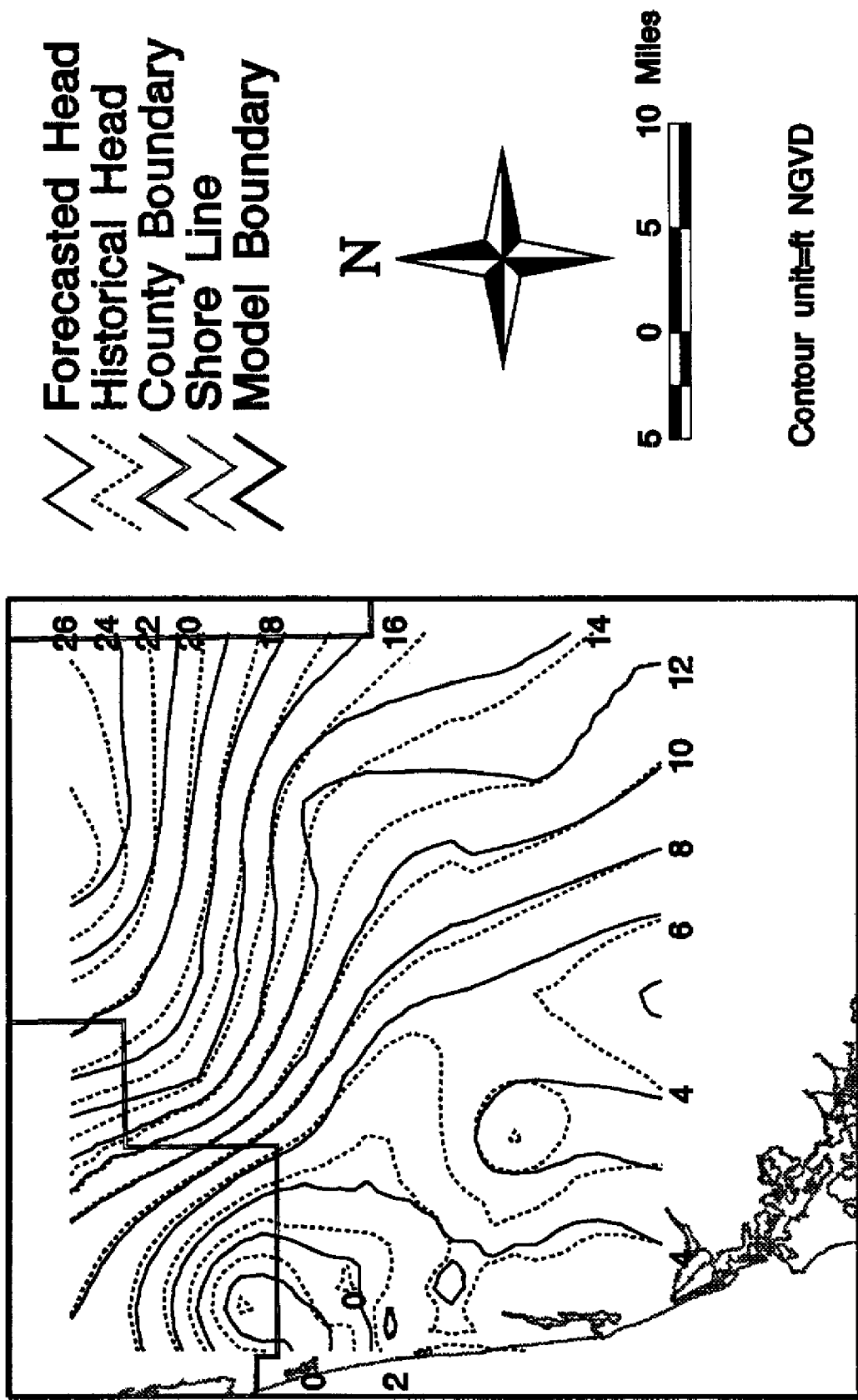


Figure 14. Comparison of forecasted and historical heads for Layer 2 & Lead time 1.

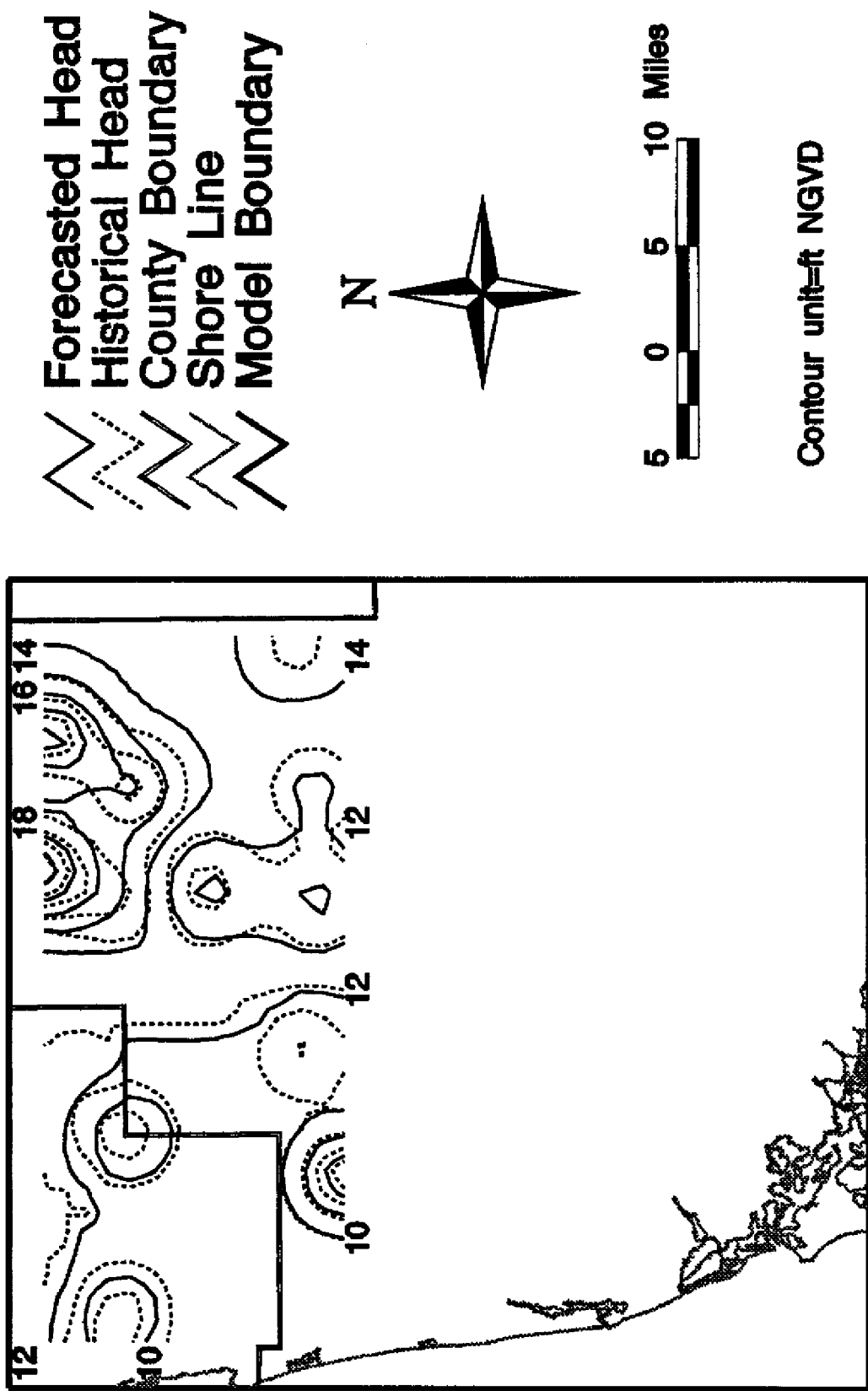


Figure 15. Comparison of forecasted and historical heads for Layer 3 & Lead time 1.

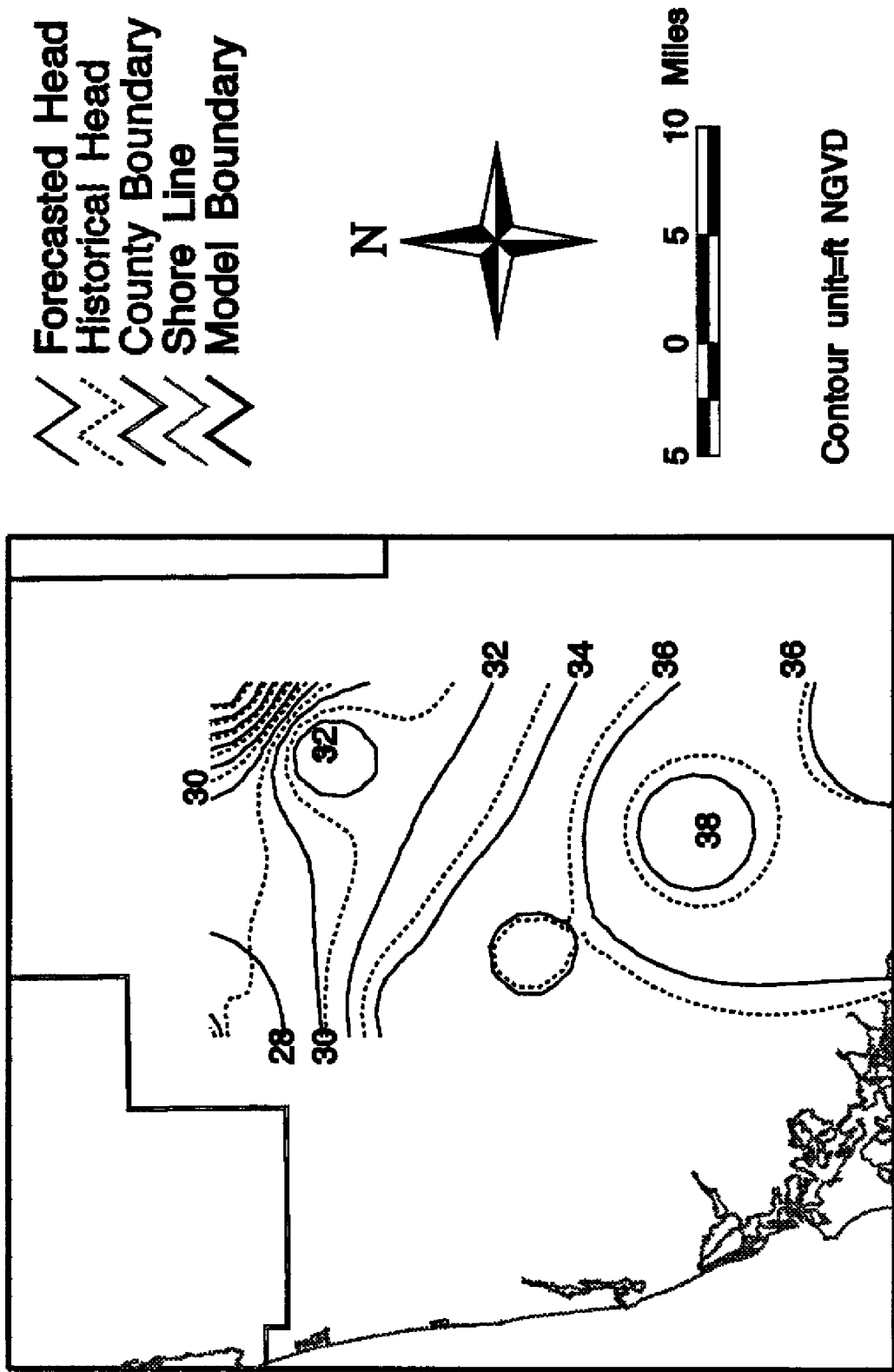


Figure 16. Comparison of forecasted and historical heads for Layer 4 & Lead time 1.

9. Summary and Conclusion

To forecast groundwater heads in Collier County, Florida, three STARX models were built. As an identification process, several properties of groundwater flow in the region were discussed, which includes the groundwater flow characters in the model area, selection of spatial neighbors, statistics of the first spatial neighbors, and the spatial statistics of space-time correlation function (STCF) and time lagged semi-variogram function. The results of the STCF analysis indicates that the spatial correlation of monthly heads were stronger than the temporal correlation, while the one-to-one correlation between rainfall and head were relatively low.

In order to satisfy the underlying stationary assumption of raw data which is required in the STARX model, the following three transformations of raw data were performed: standardization to remove seasonality, Box-Cox transformation for normality, and use of a linear model to remove temporal trends. Using the maximum likelihood fitting procedure based on the EMSSE algorithm, three STARX models were calibrated whose temporal orders selected by the Akaike information criterion are: STARX(2,no) for Layer 1, STARX(1,0) for Layer 2, and STARX(3,0) for Layer 3 and 4.

The verification of the fitted STARX model focused on the data generation and forecasting ability. Results shows that the STARX model preserved the first four moments as well as space-time correlations for the designated space-time model orders. The 12-month ahead forecasting from January 1993 were performed and compared with historically measured data; resulting in a good match.

REFERENCES

- Bear, J., *Hydraulics of groundwater*, McGraw-Hill, New York, 1979.
- Bennet, M. W., A three-dimensional finite difference ground water flow model of western Collier county, Florida, *Technical Pub. 92-04*, South Florida Water Management District, West Palm Beach, Florida, 1992.
- Box, G.E.P., and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.
- Bras, R. L., and Rodriguez-Iturbe, *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass., 1985.
- Brockwell, P.J., and R.A. Davis, *Time Series Theory and Methods*, Springer-Verlag, N.Y., 1987.
- Dickey, D.A. and W.A., Fuller, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.
- Environmental Systems Research Institute, *Understanding GIS: The Arc/Info Method*, Environmental Systems Research Institute Inc., CA, 1992.
- McDonald, M.G., and Harbaugh, A.W., A Modular three-dimensional finite-difference ground water flow model: U.S. Geological Survey Techniques of Water Resources Investigations, Book 6, chapter A1. 1988.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and L.W. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Fort Collins, Colo., 1980.
- Shumway, R.H., *Applied Statistical Time Series Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1988.
- Stoffer, D. S. Estimation and identification of space-time ARMAX models in the presence of missing data. *Journal of the American Statistical Association*, 81, 762-772, 1986.

Appendix II-A

Gaging Station Information

THIS PAGE INTENTIONALLY BLANK

Heading Index

ID = well identification number
 x-coor = x-coordinate(ft)
 y-coor = y-coordinate(ft)
 name = USGS well name
 elev. = surface elevation (ft. NGVD)
 wd = well depth (ft)
 cd = casing depth (ft)
 aq = aquifer type
 SU = Surficial aquifer
 LT = Lower Tamiami aquifer
 SA = sandstone aquifer
 MH = Mid-Hawthorn aquifer
 type = instrument type
 gstape = USGS taping
 usgs = USGS recoder
 b-yr = begining year of record
 e-yr = ending year of record

ID	x-coor	y-coor	name	elev	wd	cd	aq	type	b-yr	e-yr
Layer-1										
1	381523	587958	C-995	8.00	37.0	28.0	SU	gstape	1985	1993
2	399789	592454	C-495	6.58	70.0	8.0	SU	usgs	1971	1993
3	365081	613044	C-496	13.59	57.0	8.0	SU	usgs	1974	1993
4	279826	621373	C-969	5.09	72.0	25.0	SU	usgs	1985	1993
5	292113	615960	C-1063	6.08	55.0	30.0	SU	gstape	1987	1993
6	322074	625619	C-1067	5.40	65.0	30.0	SU	gstape	1987	1993
7	285149	627807	C-968	6.50	23.0	8.0	SU	usgs	1985	1993
8	264244	640639	C-967	5.04	140.0	19.0	SU	usgs	1985	1993
9	321791	645307	C-690	8.64	48.0	43.0	SU	usgs	1981	1993
10	386697	646196	C-296	14.10	45.0	8.0	SU	usgs	1974	1993
11	321210	658839	C-972	11.18	40.0	25.0	SU	gstape	1985	1993
12	274748	661583	C-996	10.53	23.0	13.5	SU	gstape	1985	1993
13	286871	662227	C-976	10.95	40.0	10.0	SU	gstape	1985	1993
14	248055	662843	C-1052	7.26	25.0	10.0	SU	gstape	1987	1993
15	238400	663507	C-1062	10.77	24.0	10.0	SU	gstape	1987	1993
16	238643	673400	C-1000	11.50	24.0	14.0	SU	gstape	1985	1993

Id	x-coor	y-coor	name	elev	wd	cd	aq	type	b-yr	e-yr
17	237375	674720	C-1001	12.65	24.0	14.0	SU	gstape	1985	1993
18	243028	675494	C-392	10.00	30.0	28.0	SU	usgs	1974	1993
19	247676	675769	C-1054	8.83	25.0	10.0	SU	gstape	1987	1993
20	386237	678505	C-986	16.39	40.0	28.0	SU	gstape	1985	1993
21	258359	680149	C-1055	9.87	25.0	10.0	SU	gstape	1987	1993
22	237514	682494	C-1026	16.67	38.0	28.0	SU	gstape	1986	1993
23	237629	686330	C-1061	14.88	25.0	10.0	SU	gstape	1987	1993
24	288008	689280	C-980	13.37	30.0	15.0	SU	gstape	1985	1993
25	307679	689590	C-953	12.35	40.0	12.0	SU	usgs	1985	1993
26	331186	692519	C-598	13.36	36.5	32.5	SU	usgs	1981	1993
27	240959	694791	C-321	11.21	20.3	20.3	SU	gstape	1972	1993
28	237423	697236	C-1060	11.71	25.0	10.0	SU	gstape	1987	1993
29	233332	698170	C-999	8.74	23.0	13.0	SU	gstape	1985	1993
30	247644	701011	C-1057	10.69	10.5	8.0	SU	gstape	1987	1993
31	237009	703802	C-1059	9.42	25.0	10.0	SU	usgs	1987	1993
32	253586	705318	C-384	12.70	58.0	9.7	SU	usgs	1974	1993
33	244928	733842	L-5722	11.36	21.0	11.0	SU	usgs	1986	1993
34	352285	763216	C-1078	31.91	38.0	13.0	SU	gstape	1987	1993
35	382596	777660	C-1075	30.64	28.0	8.0	SU	gstape	1987	1993
36	236208	721476	L-5726	11.00	32.0	22.0	SU	usgs	1986	1993
37	232234	726247	L-5724	11.96	35.0	25.0	SU	usgs	1986	1993
38	276094	726800	L-1997	14.90	20.0	10.0	SU	usgs	1975	1993
39	369492	712981	C-503	17.47	20.4	8.0	SU	usgs	1974	1993
40	303976	735445	C-978	19.06	40.0	15.0	SU	gstape	1985	1993
41	344105	739014	C-981	15.34	60.0	40.0	SU	gstape	1985	1993
42	410999	759316	C-131	26.60	54.0	22.0	SU	usgs	1956	1993
43	342095	712266	C-984	20.30	40.0	30.0	SU	gstape	1985	1993
44	345368	585447	C-1065	3.47	50.0	27.0	SU	gstape	1987	1993
45	233014	763197	L-2308	15.49	13.5	12.0	SU	gstape	1976	1993

ID	x-coor	y-coor	name	elev	wd	cd	aq	type	b-yr	e-yr
46	284175	759067	L-5665	20.00	37.0	32.0	SU	gstape	1983	1993
47	256870	769616	L-739	18.65	20.0	18.0	SU	gstape	1973	1993
48	314407	769929	L-1138	25.19	20.0	15.0	SU	gstape	1976	1993
Layer-2										
49	285316	624777	C-975	6.70	150.0	60.0	LT	gstape	1985	1993
50	322074	625619	C-1068	5.40	200.0	120.0	LT	gstape	1987	1993
51	257593	641585	C-600	5.39	52.0	48.0	LT	usgs	1981	1993
52	274481	645632	C-599	8.92	50.0	46.0	LT	usgs	1981	1993
53	381254	655600	C-1070	13.37	205.0	100.0	LT	gstape	1987	1993
54	321210	658839	C-973	11.18	150.0	90.0	LT	gstape	1985	1993
55	237201	661192	C-130	5.49	71.5	69.0	LT	gstape	1976	1993
56	286871	662227	C-977	10.95	140.0	75.0	LT	gstape	1985	1993
57	238308	663507	C-472A	14.69	70.2	70.2	LT	gstape	1983	1993
58	234587	665852	C-524	4.29	80.0	63.0	LT	gstape	1976	1993
59	234231	667268	C-525	6.90	83.0	63.0	LT	gstape	1975	1993
60	233877	668886	C-526	5.71	68.0	63.0	LT	gstape	1983	1993
61	242444	669338	C-161	3.48	165.0	140.0	LT	gstape	1975	1993
62	233258	671919	C-527	5.90	71.5	63.0	LT	gstape	1977	1993
63	235552	674530	C-474A	6.67	72.0	63.0	LT	gstape	1976	1993
64	237558	674820	C-491	12.60	71.0	70.0	LT	gstape	1979	1993
65	237785	667347	C-123	11.44	157.0	96.6	LT	usgs	1975	1993
66	242846	675495	C-391	9.38	75.0	70.0	LT	usgs	1974	1993
67	247960	677685	C-430	8.50	65.0	63.0	LT	usgs	1976	1993
68	234943	679178	C-528	4.39	80.0	63.0	LT	gstape	1976	1993
69	237514	682494	C-506A	16.67	70.7	62.5	LT	usgs	1976	1993
70	247644	701011	C-1058	10.69	80.0	62.0	LT	usgs	1987	1993
71	237429	683504	C-490	16.55	71.0	70.0	LT	gstape	1976	1993
72	239627	685409	C-489	15.20	83.0	63.0	LT	usgs	1974	1993
73	288008	689280	C-956	13.37	260.0	60.0	LT	gstape	1987	1993

ID	x-coor	y-coor	name	elev	wd	cd	aq	type	b-yr	e-yr
74	307679	689590	C-951	12.35	170.0	120.0	LT	usgs	1985	1993
75	247404	691319	C-458	9.17	63.0	63.0	LT	gstape	1976	1993
76	243673	691746	C-460	10.39	66.0	64.0	LT	usgs	1975	1993
77	237500	694913	C-1003	18.03	61.0	51.0	LT	gstape	1985	1993
78	342578	695202	C-988	15.66	160.0	95.0	LT	usgs	1985	1993
79	237439	699760	C-424	11.00	132.0	126.0	LT	gstape	1977	1993
80	244756	705370	C-1004	9.92	60.0	52.0	LT	usgs	1985	1993
81	302204	706578	C-304	15.59	130.0	125.0	LT	gstape	1983	1993
82	307686	711197	C-971	15.54	150.0	100.0	LT	gstape	1985	1993
83	369492	712880	C-1073	18.80	160.0	100.0	LT	gstape	1987	1993
84	276094	726800	L-1996	15.03	259.0	65.0	LT	usgs	1976	1993
85	245086	729802	L-738	9.16	75.0	61.0	LT	usgs	1980	1993
86	249828	731793	L-1691	12.49	69.0	58.0	LT	usgs	1974	1993
87	301825	742220	C-492	17.50	64.0	60.0	LT	usgs	1974	1993
88	410817	759114	C-1074	26.71	130.0	100.0	LT	usgs	1987	1993
89	366828	762865	C-363	34.10	119.0	84.0	LT	gstape	1983	1993
90	357586	771881	C-462	34.10	110.0	50.0	LT	usgs	1974	1993
91	233210	750979	L-5731	15.67	120.0	90.0	LT	usgs	1987	1993
Layer-3										
92	273978	705305	C-303	13.45	300.0	232.0	SA	gstape	1983	1993
93	342095	712266	C-989	20.30	270.0	240.0	SA	usgs	1985	1993
94	369492	712880	C-689	18.80	265.0	230.0	SA	gstape	1983	1993
95	304520	715351	C-688	16.73	242.0	220.0	SA	gstape	1983	1993
96	405450	717125	C-1072	19.29	260.0	140.0	SA	usgs	1987	1993
97	345014	739011	C-1079	15.34	390.0	298.0	SA	usgs	1987	1993
98	369813	758009	C-298	30.67	303.0	254.0	SA	gstape	1983	1993
99	343829	762843	C-687	22.98	310.0	290.0	SA	gstape	1983	1993
100	350077	781499	C-531	41.84	240.0	210.0	SA	usgs	1977	1993
101	382596	777660	C-1077	30.64	246.0	170.0	SA	usgs	1987	1993

ID	x-coor	y-coor	name	elev	wd	cd	aq	type	b-yr	e-yr
102	290501	769637	L-2192	27.26	184.0	155.0	SA	usgs	1976	1993
103	284175	759067	L-5664	20.00	300.0	180.0	SA	usgs	1983	1993
104	241990	759204	L-5668	15.62	155.0	106.0	SA	usgs	1984	1993
105	233014	763197	L-741	15.26	119.0	102.0	SA	usgs	1976	1993
Layer-4										
106	372433	539020	C-39	10.00	484.0	436.0	MH	gstape	1985	1993
107	378559	572519	C-311	4.94	450.0	430.0	MH	gstape	1982	1993
108	349888	625007	C-987	9.30	370.0	280.0	MH	gstape	1985	1993
109	321235	664694	C-974	10.10	460.0	400.0	MH	gstape	1985	1993
110	307224	689592	C-948	12.35	420.0	370.0	MH	gstape	1985	1993
111	369492	712880	C-684	17.46	490.0	440.0	MH	gstape	1982	1993
112	303976	735445	C-963	19.06	340.0	340.0	MH	gstape	1985	1993
113	344105	739014	C-983	15.34	520.0	480.0	MH	gstape	1985	1993
114	387029	736656	C-965	21.96	458.0	438.0	MH	gstape	1985	1993
115	301825	742220	C-1080	17.50	309.0	238.0	MH	gstape	1987	1993

THIS PAGE IS INTENTIONALLY BLANK

Part III

Development of a Feedforward Control Scheme for the Regional Groundwater Drought Problem

THIS PAGE INTENTIONALLY BLANK

1. Introduction

The water manager's goal of implementing a groundwater drought plan is to protect the groundwater resources and to assure equitable distribution of water to the different users during the anticipated drought condition so that adverse economic, social, environmental, and health impacts from the water shortage will be minimized. Since drought management plans should be based on the past, present, and future water conditions, forecasting of future droughts is required. Forecasting in general is accompanied by considerable uncertainty and the following control problem should also take into account the uncertainty. That is why many control problems have been handled by the stochastic time series topics. Along with stochastic forecasting theory, there exist a variety of stochastic control theories [Box and Jenkins, 1976, page 423, Bennett, 1979, page 533, and others]. However, to apply these control theories to a practical problem, specifically for large scale problems, several difficulties, such as appropriate forecasting model structure and scaling of the real system arise; none of which are addressed in the literature. Thus, the objective of this study is to develop a state-of-art methodology to handle the groundwater drought problem for practical management purposes. The methodology is based on the stochastic forecasting and control scheme in the system theory applied to a large scale groundwater system.

To manage the groundwater resources during anticipated drought conditions, a feedforward control scheme was adopted. This scheme consists of a forecasting equation and a control equation. The forecasting equation was built by the Kalman filter algorithm associated with the state-space form of space-time autoregressive model with exogenous variable (STARX), which was discussed in Parts I and II of this report. The control equation, which will be used to estimate recommendations for reducing permitted groundwater use based on the anticipated deviation from the target water level, was developed by the empirical relationship between the head change and a conceptual pumpage/recharge (PR) function. The PR function was defined by surface water budget components including effective rainfall, potential evapotranspiration, and spatial landuse patterns.

An alternative to this control equation can be physically-based groundwater flow models such as MODFLOW [McDonald and Harbaugh, 1988]. However, difficulties in this alternative approach are that model input, including rainfall and boundary conditions should be forecasted (which creates another difficulty when those variables are multivariate in nature), that aquifer characteristics should be pre-defined (which usually contain considerable uncertainties), and that the forecasting itself contains a lot of uncertainty so that the use of sophisticated physically-based models does not increase accuracy proportional to the increased work load associated with creating and running a physically-based model. For these reasons, a simple regression equation was chosen to be used as a control equation.

The structure of the proposed feedforward control scheme is that, after computing deviations from the target heads at the next predicted month for each aquifer layer, those deviations are converted to the recommended reduction of groundwater pumping by the control equation, and the spatially-distributed percentage reduction rates are computed based on the permitted groundwater use estimated by the 2-in-10-year drought scenario. As a final result, the control scheme provides a spatially varied water use restriction map in each month, which can be used as a management tool during anticipated drought periods. This reduction plan is a regional and monthly guideline. The proposed control scheme was tested in western Collier County, Florida, under several generated drought conditions. The test showed promising results for managing the groundwater resources during the drought periods.

2. Existing Rules for Groundwater Use Permits and Drought Contingency Plan

A definition of hydrologic drought given by *Dracup et al.* [1980] is "a water shortage with reference to a specified need for water in a conceptual supply and demand relationship". This definition implies that drought is defined by a relative sense usually in terms of frequency of historical hydrologic records. Statistically, the probability of a drought event p , which is opposite to that of flood, is defined by the following non-exceedance cumulative density function $F(\cdot)$:

$$p = F_X(x) = \text{prob.}\{X \leq x\} \quad (1)$$

where X is the random variable denoting any hydrologic event such as rainfall depth or groundwater head, and x is any given number. Then, the return period T_r of the drought event is given by $1/p$, and the complementary probability q is given by $(1-p)$ which is the exceedance probability. For the definition of reference to a specified need for water, it is necessary to know the existing rules for groundwater use and drought management in the model area, from which the recommended groundwater pumping reductions during the anticipated drought condition can be estimated by the ratio of the pumping reduction required to meet the target head under existing permit conditions. The model area referred to here consists of the northwest portion of Collier County and the southwest portion of Lee County as described in Figure 3 and Part-II of this report.

The Florida Statutes (Part II of Chapter 373) state that the South Florida Water Management District (The District) is responsible for the permitting of the use of both surface and ground water within its jurisdictional boundaries [SFWMD, 1993]. The major groundwater uses in the model area are public, industrial, mining, and irrigation. However, Bennett [1992] estimated that agricultural and landscape irrigation withdrawals account for approximately 78 percent of the total groundwater use based on 1988 estimates. Thus, the discussion hereafter will be focused on the method of estimating the agricultural water use. The permit information manual Volume III [SFWMD, 1993] specifies that the reasonable need for irrigation water use is defined by the supplemental water requirement (SWR) with the irrigation efficiency, which varies from 50% to 85% depending on the irrigation methods. This manual uses the modified Blaney-Criddle equation for the evapotranspiration and the SCS method for the effective rainfall to estimate the SWR for crops. The monthly SWR is computed by the 2-in-10-year frequency rainfall, and the maximum monthly allocation is determined by choosing the irrigated month with the largest supplemental water requirement.

The water shortage plan, which is one of the water management rules listed in Volume

III, provides consistent rules, principles and restrictions that apply to groundwater users, facilitating the management and enforcement of droughts in the area. This plan also provides for variances from these restrictions. If there is a possibility that there will not be sufficient water available within a source class to meet the anticipated demands, water managers issue a water shortage order to the users. This order is activated when a drought is foreseen and remains active as long as water restrictions are in effect.

Specifically, Chapter 40E-21.221 in the water shortage plan outlines the drought contingency plan by stating that "the current data shall be compared to historical data to determine whether estimated present and anticipated available water supply will be insufficient to meet the estimated present and anticipated demand...". This plan establishes the severity of the groundwater drought condition with the reduction of water uses as follows:

Water shortage (W/S) Phase	Color code	% reduction in overall demand
Moderate W/S	Yellow	Less than 15%
Severe W/S	Orange	Less than 30%
Extreme W/S	Red	Less than 45%
Critical W/S	Purple	Less than 60%

However, there exists no objective definition or tool to determine the above water shortage phases. In the past, a uniform water use reduction scheme at the regional scale based on the simple frequency analysis of rainfall event was used as a management tool. Thus, this study was intended to develop a state-of-the-art methodology to determine the spatial water shortage condition from current and anticipated droughts, as well as a spatial groundwater reduction scheme to meet the goal of groundwater resource protection. The system control theory was adopted to resolve this problem.

3. System Control Theory

As mentioned before, the ultimate goal of groundwater drought management is to protect the groundwater resources during anticipated drought conditions, while maximizing permitted water uses. In terms of system theory, this management activity can be formulated and solved by the control theory. System control serves to specify what system inputs are required to achieve given output levels. One of the common ways of modeling and controlling the system is by the stochastic time series model as a represented system. Numerous discussions on this topic are available from *Box and Jenkins* [1976, page 423], *Bennett* [1979, page 533], and others.

In system theory, there exist a variety of control schemes, but three basic forms prevail; open-loop control, closed-loop control, and feedforward control. In the open-loop control scheme, a control rule is preset on the basis of available experience. This scheme can be adjusted only infrequently and tends to allow unexpected fluctuations. The closed-loop control scheme, or so called feedback device, compares the system output and the specified target and makes adjustments based on the deviation. The feedforward control scheme offers the advantages of detecting disturbances before they affect system operations and a control action is then initiated to compensate for potential deviations in the output.

The District's past drought management practice can be classified as an open-loop scheme since the drought frequency was computed from past and present drought conditions and the empirical water use reduction was applied uniformly over the local area. The next section describes how to build a feedforward control scheme to manage groundwater drought problems. Of particular interest is how to build a control equation based on the relationship between the conceptual PR function and head changes in multi-layered aquifers.

4. Feedforward Control Scheme

In terms of a stochastic time series framework, groundwater flow is governed by the endogenous variable as well as several exogenous variables. Exogenous variables to the groundwater system include rainfall, evapotranspiration, pumpage, seepage from the surface water bodies, and regional groundwater flow. Among them, pumpage is the only controllable variable in the sense of the system theory. However, the historical pumping records in the model area, specifically those of agricultural uses, are not available. Thus, the historical pumpages for crop irrigation were estimated from the surface water budget.

Using the forecasting model and the empirical PR-groundwater head relationship, a feedforward control scheme was developed to apply to groundwater drought management, as shown in Figure 1. Output from the control system is simulated groundwater heads in space at the next time step. System inputs include the most recently measured heads, specified target heads (seasonal) and water use permit information, as well as the unmeasured disturbances. The unmeasured disturbance, which is not an artificial input, is the source of disturbances other than the measured disturbance [*Box and Jenkins, 1976, page 424*]. The main components of the proposed feedforward control scheme are a system equation and a control equation. The system equation forecasts the future groundwater heads in each layer using the Kalman filter algorithm associated with the STARX models developed in Part II of this report. The main function of the control equation is to manipulate the system so that the state variable reaches the target level. If head change is explicitly modeled by the PR function, it is possible to build a control equation by combining the STARX model and the head-PR equation as described in *Box and Jenkins [1976, page 424]*. However, this approach is not practical since a multi-layered aquifer system creates a huge state dimension which impedes an appropriate parameter estimation and the handling of those equations. Instead, a control equation was developed based on the empirical relationship between the PR function and head changes in multi-layered aquifers.

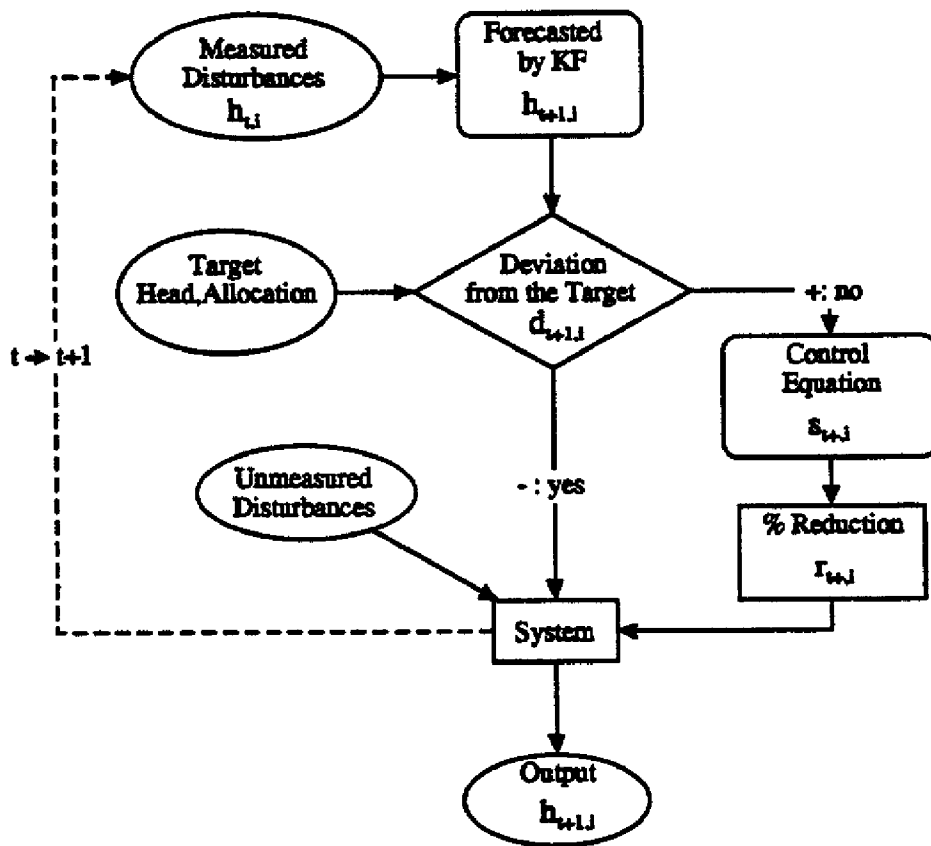


Figure 1. A feedforward control scheme for groundwater drought management.

Defining that $h_{t+1,i}^m$ is the forecasted head at time t , site i , and layer m , and that $H_{k+1,i}^m$ is the corresponding seasonal target head at month $k+1$ ($k=1,\dots,12$) of time $t+1$, the deviation from the target $d_{t+1,i}^m$ is given by

$$d_{t+1,i}^m = H_{k+1,i}^m - h_{t+1,i}^m \quad (2)$$

Positive $d_{t+1,i}^m$ means that deficit of water is predicted and a control action is needed. Using the control equation with the computed deviations in every layer, the PR term $s_{t+,i}$, where $t+$ indicates a time period during t to $t+1$ since it is a cumulative term during that period, is computed by the control equation. Then, the percent reduction of pumpage $r_{t+,i}$ is computed by

$$r_{t+,i} = \frac{S_{t+,i}}{S_i^T} \times 100 \quad (\%) \quad (3)$$

where s_i^T is the permitted groundwater use at site i with the superscript T denoting the target level.

If no control scheme is used, the forecasting can be done for any lead time l even though the forecast error increases with increasing l . If the feedforward control scheme is continuously used, either one of the following two cases occurs depending on the availability of input data: The first case is when the complete- or partial-data are available at time t , it should be used to forecast (update) the state at time $t+1$ and to compute a reduction rate during t to $t+1$. The second case is if state measurements at time t are not available at all but the system was controlled at $t-1$, the reduction rates at $t+1$ can be computed after updating the state at time t using the target heads as the second best alternative.

The following sections will illustrate how to operate a system equation operation, how to set up a control equation, and how to specify the target groundwater heads along with the permitted groundwater conditions based on the existing rules.

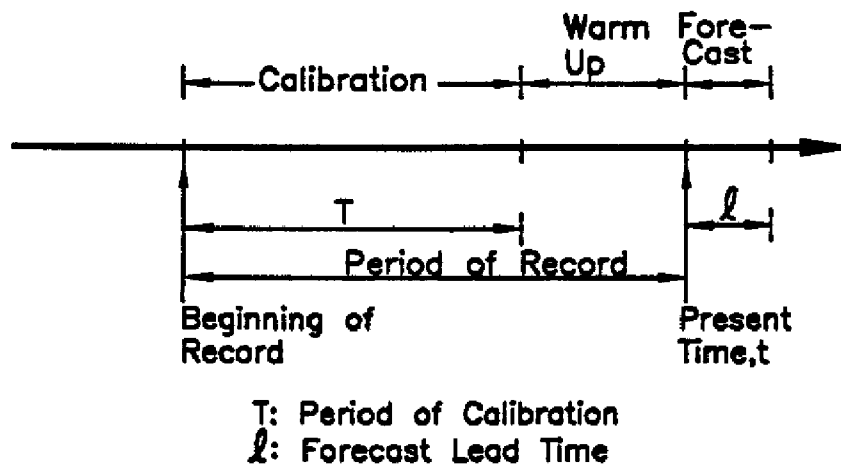


Figure 2. Concept of real time forecasting using the Kalman filtering algorithm.

4.1. Operation of the System Equation

Forecasting by the Kalman filter algorithm requires a set of predefined optimal parameters for a conceived forecasting model. The forecasting parameters include parameters for the STARX model, state and measurement noise covariances, and updated state and associated error covariance. Ideally, these forecasting parameters can be calibrated using the up-to-date measured data, but this option is impractical since the calibration process of a large scale model needs a great deal of time and effort. Moreover, if the system is more or less stationary in time, a new calibration with a small amount of updated data should not differ from the previous calibration, as indicated in the sensitivity analyses of the expectation-maximization algorithm given by *Stoffer* [1986] and discussion in Part II of this report.

Fortunately, the Kalman filter algorithm allows updating the state and its associated error covariance continuously using the most currently measured data before forecasting. Let us define such an update of the state vector and its error covariance as the warm-up process. If a complete-data set is available during the warm-up period, the state error covariance remains constant (time-invariant) and only the state vector will be updated. Figure 2 illustrates the concept of real time forecasting in the time horizon. The real time forecasting by the Kalman

filter algorithm is an adaptive mode: the forecasting is based on the previous model inputs as well as the previous measured model output. The forecast lead time l is mainly determined by the error covariances of the forecasted states. If l is too long, the forecasting errors will be the same as variance of sample data.

For the practical purpose of the groundwater drought management problem, it is recommended to calibrate forecasting parameters every other year, mainly during the off-drought period or non-cultivating season. Since the model is not verified using the actual drought records, a process of model evaluation including the forecasting model is necessary after certain years (for example every 5 years) or after experiencing a severe drought event, whichever may come first. The model structures including the spatial index matrixes, covariance structure, and temporal order of model can be refined based on this evaluation.

4.2. Supplemental Water Requirement (SWR) for Groundwater Use Permits

Since agricultural water use is the largest of the groundwater uses in the area, discussion here is focused on the permitting process of agricultural water use, from which a conceptual PR function was derived. When little or no measurements of water requirements for crops are available, the SWR for a crop is usually used to estimate the pumping requirement. The SWR for a crop is computed by subtracting effective rainfall from the potential ET, both of which are not directly measurable. In order to determine the SWR for the groundwater use permit, the permit information manual Volume-III [SFWMD,1993] recommends the modified Blaney-Criddle equation for estimation of evapotranspiration (ET) E_t at time t , which is given by [Jensen *et al.*, 1990, page 103]

$$E_t = k_t k_c T_t \frac{P_t}{100} \quad (4)$$

where k_t is a coefficient related to the mean monthly air temperature, ($=0.0173T_t-0.314$), k_c is a monthly factor reflecting the growth stage of the crop type, T_t is the mean monthly temperature at the month t (°F), and p_t is the percent of daytime hours of the year that occurred during

the month t .

The Soil Conservation Service method [Jensen *et al.*, 1990, page 67; SFWMD, 1993] to compute effective rainfall RE_t^* at the month t is

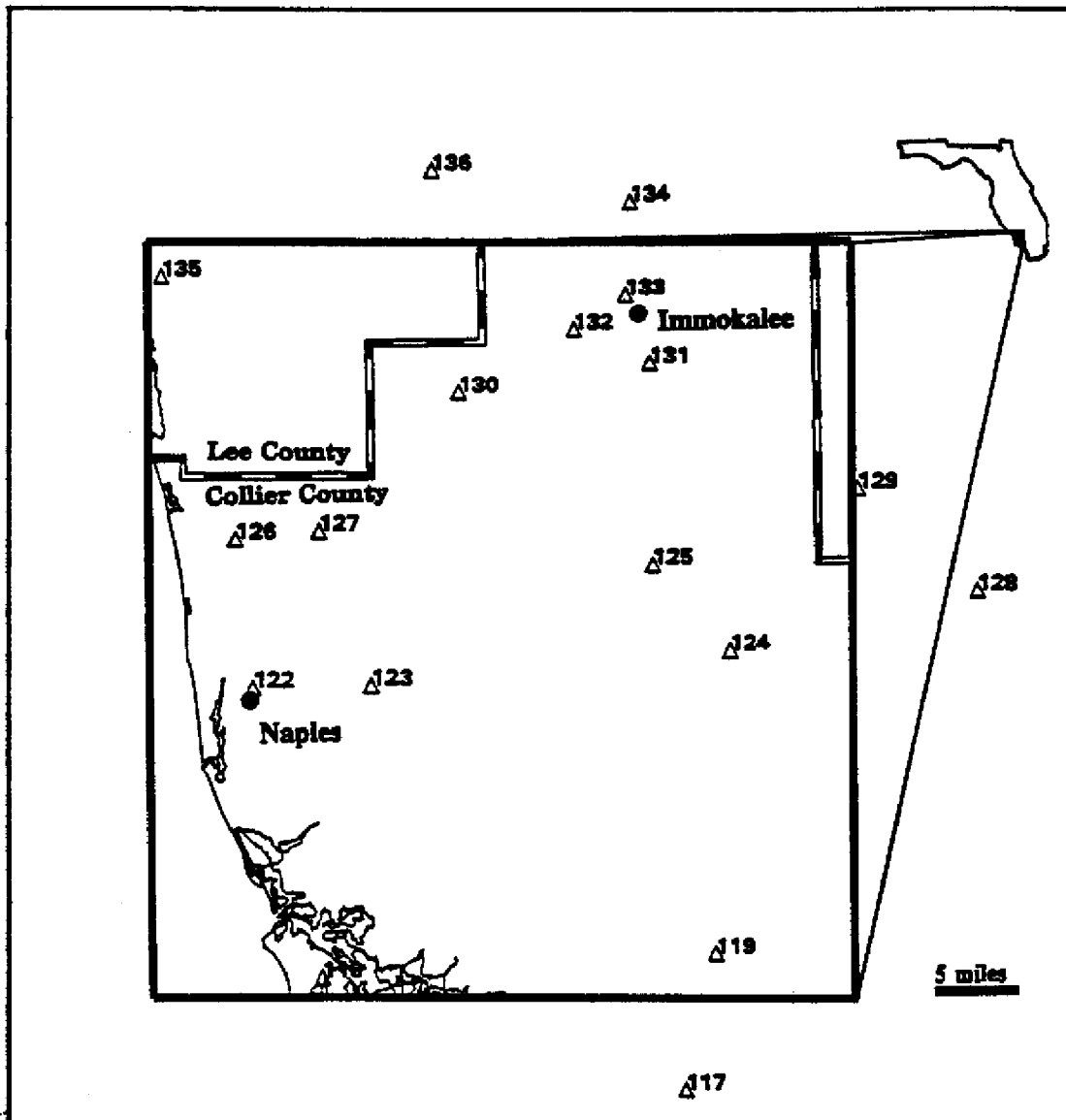
$$RE_t^* = f_s f_e E_t^* (0.7092 R_t^{0.8242} - 0.1156) \quad (5)$$

where f_s is a soil factor given by $f_s = 0.5317 + 0.2952d - 0.0577d^2 + 0.0038d^3$ with d as the net depth of application, f_e is a conversion factor from an average rainfall to a 2-in-10-year frequency rainfall, E_t^* is the monthly effective ET expressed by $E_t^* = 10^{0.02426E_t}$, and R_t is the 2-in-10-year monthly rainfall depth at month t . Then, the monthly SWR for the crop at time t , s_t , is given by

$$s_t = E_t - RE_t^* \quad (6)$$

where s_t is a depth unit (in inches). The total volume of SWR at month t can be obtained by multiplying total irrigation area and dividing by the irrigation efficiency which is dependant on the irrigation method. Finally, the permit allocation for groundwater use is determined by selecting the month which has the largest SWR.

As a summary, the 2-in-10-year SWR is computed by the spatial parameter set, $\{i_{em}, A, d, f_e, k_c, R_p, T_p, \text{ and } p_t\}$, among which the last three terms are space-time dependent variables. An attempt was made to compute the historical SWR series by the above method with historical rainfall ($f_e=1.0$) and temperature data, and related the SWR with the corresponding historical groundwater head change. However, the result was unsatisfactory, implying that the SWR method, particularly the RE_t^* term by the SCS method, does not adequately estimate the SWR in the model area. Thus, after thoroughly investigating the relationship between groundwater head and the corresponding rainfall (which is the most important driving force in both surface and subsurface hydrologic systems), a conceptual PR function was developed, along with an estimation of the associated parameters by optimization methods.



LEGEND

- County Boundary
- ~ Shoreline
- △ Rainfall Station
- ET Station

Figure 3. Location of study area with rainfall and ET stations.

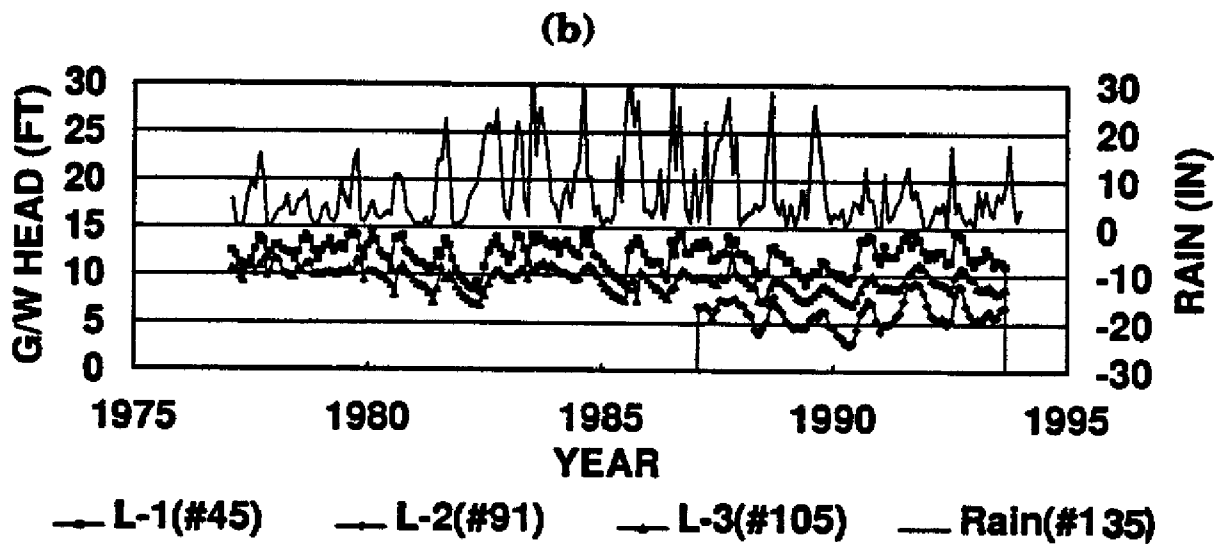
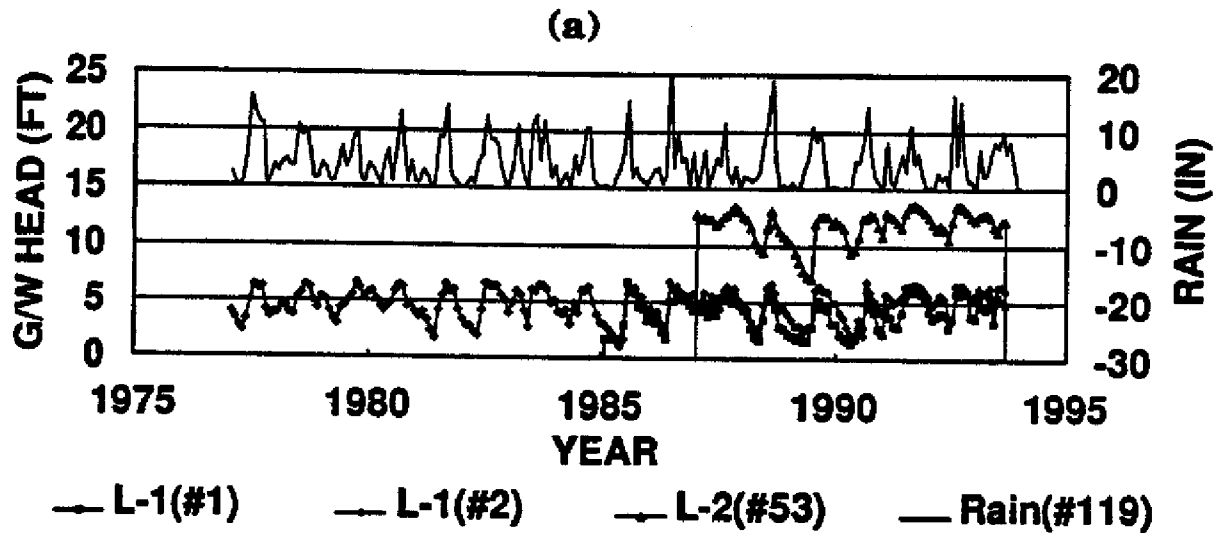


Figure 4. Time series plots of rainfall-groundwater heads for the selective stations, where L-i(#j) stands for the selected gaging station j at aquifer i.

4.3. Rainfall-Groundwater Head Relationship

Although rainfall is the most important driving variable to the groundwater system, one-to-one correlation between rainfall and heads are not very significant as shown in Part II of this report. This is a main cause of failure when a simple linear regression model is fitted to the rainfall-groundwater head system. This subsection will investigate the cause and effect of rainfall and groundwater heads from the multi-aquifer systems in the model area, from which it is possible to conceptualize their relationship for building a control equation.

More often than not, time series plots provide invaluable information on the statistical characteristics of raw data, those of which can not be detected from the lumped correlation statistics, such as correlation coefficient or variogram. Figure 4 shows time series, raw data without any transformations, for arbitrarily chosen rainfall (point rainfall) and nearby groundwater head stations (instantaneous head at the end of each month) from the underlying aquifers. Particularly, the location of the gaging station in graph (a) is at the southeast corner of model area, while that of (b) is at the northwest corner. From the examination of these time series, the following observations were made;

(1) The general patterns of monthly rainfall and the corresponding groundwater head hydrograph in each layer matched very well. A large rainfall during the summer season induced high groundwater heads as expected, but there exists a threshold in rainfall depth so that the head does not exceed the threshold. This phenomenon can be explained by the infiltration capacity theory, where the excess rainfall over the infiltration capacity will be drained in the form of overland or channel flows.

(2) The groundwater heads decreased consistently when the amount of rainfall is small. These drawdown is mainly due to pumpage and regional groundwater outflow exceeding the inflow to the groundwater system. This phenomenon is another cause of failure when one seeks a direct relationship between rainfall and groundwater head.

(3) The rising and falling limbs of the groundwater hydrographs behave differently. The rising limb was steeper and increased linearly, while the falling limb was milder and decayed exponentially. There exists a time lagging effect between the high intensity rainfall and the associated peak of the head hydrograph. Also, the groundwater head maintained constantly high levels for one or two more months after a high intensity rainfall, then it gradually decreases.

(4) The deeper the aquifer layer, the less sensitive the head was to the rainfall. The temporal dependence of heads were higher than that of rainfall, which is more significant in the deeper aquifers. This implies that the applicability of the stochastic time series model to groundwater head is much higher than to that of rainfall.

5. A Conceptual Pumpage/Recharge (PR) term

Based on the above observations, a conceptual PR model was formulated and its seasonal parameter set was estimated by the optimization algorithm. A problem arose on what kind of spatial scale should be used to compute PR function. As mentioned before, the forecasting was done for the fixed gaging stations in each aquifer layer, whose two-dimensional results are displayed after a spatial interpolation. Thus, a Thiessen polygon created by the Surficial Aquifer's monitoring network was used as the base spatial scale, since this aquifer has the most dense monitoring network among the four freshwater aquifers. That is, whenever site i is cited in the space domain, i ranges from 1 to n_x (=48 stations).

Let us assume that, like the SWR in equation (6), the PR term, $s_{t,i}$, at time t and site i is defined by

$$s_{t,i} = E_{t,i}^* - R_{t,i}^* \quad (7)$$

where $E_{t,i}^*$ is the average ET depth at polygon i , and $R_{t,i}^*$ is the effective rainfall contributed to the groundwater system. Incorporating the infiltration capacity of soil in the discussion (1) of the previous subsection, the effective rainfall can be expressed by

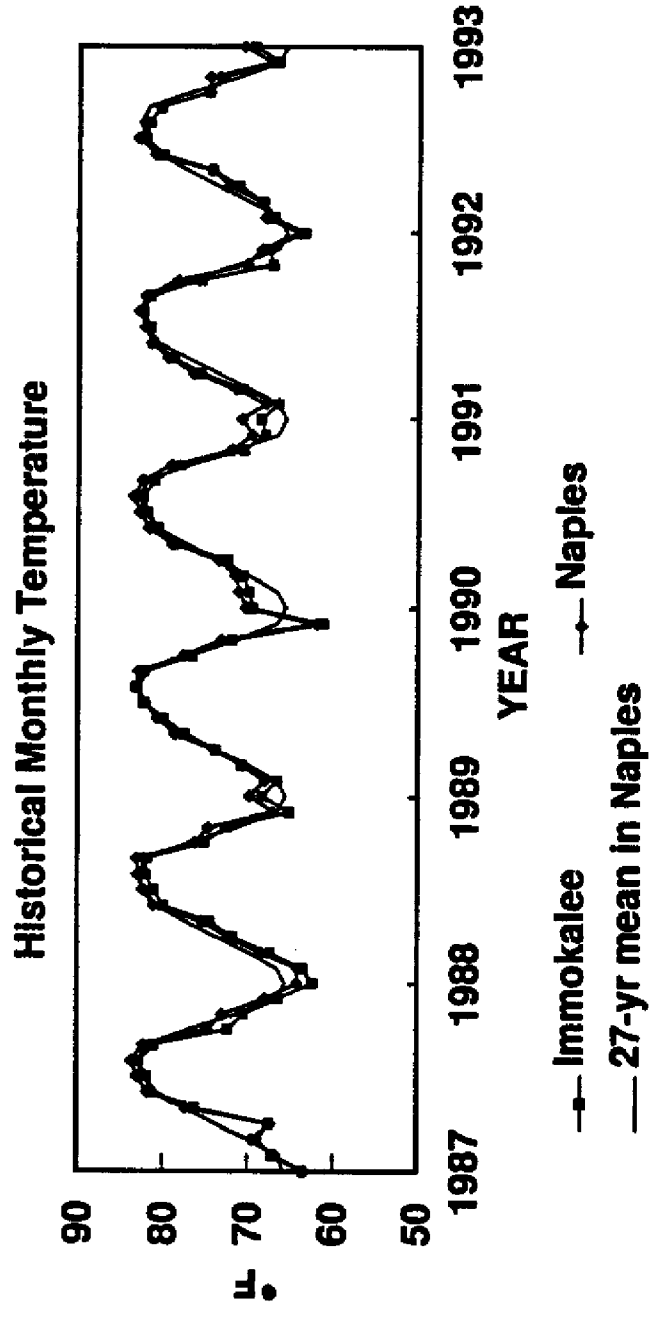


Figure 5. Historical pan ET with a 27-year average ET at Naples.

$$R_{t,i}^* = \begin{cases} R_{t,i} & \text{if } R_{t,i} \leq R_{max,k} \\ R_{max,k} & \text{otherwise} \end{cases} \quad (8)$$

where $R_{max,k}$ is the maximum rainfall depth contributing to the groundwater system at month k ($k=1, \dots, 12$) for time i , and $R_{t,i}$ is the measured monthly rainfall depth at time t and site i . Within the study area, 18 rainfall stations were available (Figure 3) and $R_{t,i}$ for a given site i was obtained by selecting the rainfall station which is closest to the center of polygon i .

The Blaney-Criddle method in equation (6) indicates that SWR in the agricultural area is computed by the ET that is the function of the crop growth factor k_c and the spatial soil type d . The second largest groundwater use was the public water supply. Individual installed capacity of public water supply wells was much greater than those of agricultural wells. Combining both agricultural use and public water supply, the average transformed ET term in depth units is formulated as:

$$E_{t,i}^* = d_i k_c(k,i) E_k + E_{pub}(i) \quad (9)$$

where $k=1, \dots, 12$, d_i is the net soil depth of application at site i , ranging from 0.4 to 3.6 depending on the soil type [refer to *SFWMD*, 1993], $k_c(k,i)$ is the factor reflecting the growth stage of the crop at site i and month k , and E_k is the monthly pan ET rate at month k . Although day-by-day temperatures fluctuate significantly, the monthly mean temperatures in the region were more or less stationary in space and time as shown in Figure 5. Figure 5 displays the monthly temperature time series measured from two stations separated by approximately 30 miles, along with the 27-year average temperature for the Naples station. This temperature stationarity justifies the use of the monthly average ET in estimating the effective ET.

$E_{pub}(i)$ is the permitted public water supply converted to the equivalent ET depth at polygon i . In order to add the pumping effects by public water supply wells in the PR function, the following two assumptions were made: First, the monthly pumping rates for public water supply wells are constant over the year (non-seasonal). Second, the influence boundary of the drawdown by public water supply well is approximately defined by an one-mile buffer zone.

Then, the $E_{pub}(i)$ is expressed by

$$E_{pub}(i) = k_p \frac{Q_i}{A_i} \quad (10)$$

where k_p is the global factor for converting the total public water supply pumpage at polygon i to the equivalent ET depth, A_i is the total area of a one-mile buffer zones created by public water supply wells at polygon i , and Q_i is the total public water supply allocation for polygon i . Figure 6 shows major public water supply wells with their one-mile buffer areas. Even though the assumption of uniform one-mile buffer zones is not objective, optimizing k_p from the conceptual PR term will compensate for the weakness of this assumption. Although $E_{pub}(i)$ is not very sensitive to the pumpage reduction, inclusion of $E_{pub}(i)$ in the PR term enables one to apply a pumpage reduction scheme to the public water supply wells during drought periods.

There are a variety of types of agricultural practices within the model area, as well as each Thiessen polygon. However those agricultural practice types were classified into one of five classes as shown in Figure 7. If n_c is the total number of different crops including any non-agricultural zones in a polygon i , a composite crop factor within polygon i at month k , denoted by $k_c(k,i)$, is computed by the areal weighing scheme as

$$k_c(k,i) = \sum_{j=1}^{n_c} w_j k_c(k,i,j) \quad (11)$$

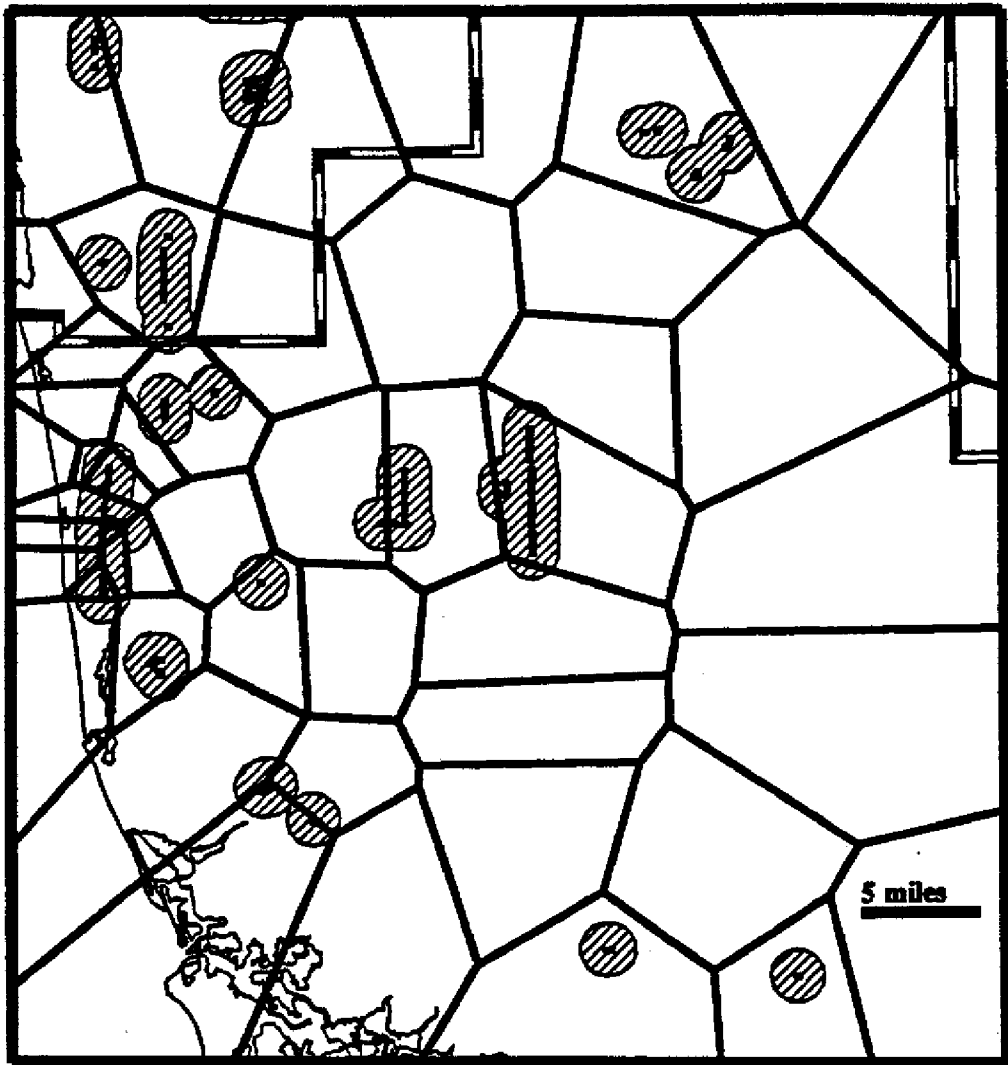
where w_j is the areal weight for the particular crop type j , that is, $w_j = A_{ij}/A_i$ with A_{ij} is the area covered by the j -th crop type within the i -th polygon, and A_i is the total area. Table 1 lists k_c 's for the distinct crop types. The remaining land uses other than agricultural are mainly urban and forest areas, whose $k_c(k,i,j)$ values are assumed the same as that of AM landuse type (groves, nurseries, and tropical fruit). To summarize, the historical PR series during the period of record are estimated from equations (7) through (11), where major inputs include the historical rainfall, spatial landuse and soil maps, public water supply well information, and an optimal parameter set, $\{R_{max,k}, E_x, \text{ and } k_p, k=1, \dots, 12\}$.

Table 1. Monthly crop factors with the estimated parameters for the conceptual PR model

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
k_c	AC	0.63	0.73	0.86	0.99	1.08	1.13	1.22	1.06	0.99	0.91	0.78	0.64
	AP	0.46	0.60	0.63	0.68	0.70	0.53	0.56	0.58	0.52	0.53	0.49	0.44
	AM	0.63	0.66	0.68	0.70	0.71	0.71	0.71	0.71	0.70	0.68	0.67	0.64
	AG	0.49	0.57	0.73	0.85	0.90	0.92	0.92	0.91	0.87	0.79	0.67	0.55
$R_{mean}(in)$	1.56	3.91	4.05	2.50	4.26	7.62	7.43	7.50	8.74	4.09	1.19	1.45	
$R_{max,k}(in)$	3.00	6.00	5.50	6.00	6.00	7.50	9.50	5.50	4.50	8.50	3.00	11.0	
$E_k(in)$	4.50	5.30	5.30	5.20	2.90	0.20	5.20	7.60	8.00	9.50	7.10	5.20	
k_p	0.832												

Ref.: - Landuse codes: AC=cropland; AP=pasture; AM=groves, ornamentals, nurseries, tropical fruits; AG(AF and UO)=grass [from *SFWMD*, 1993].

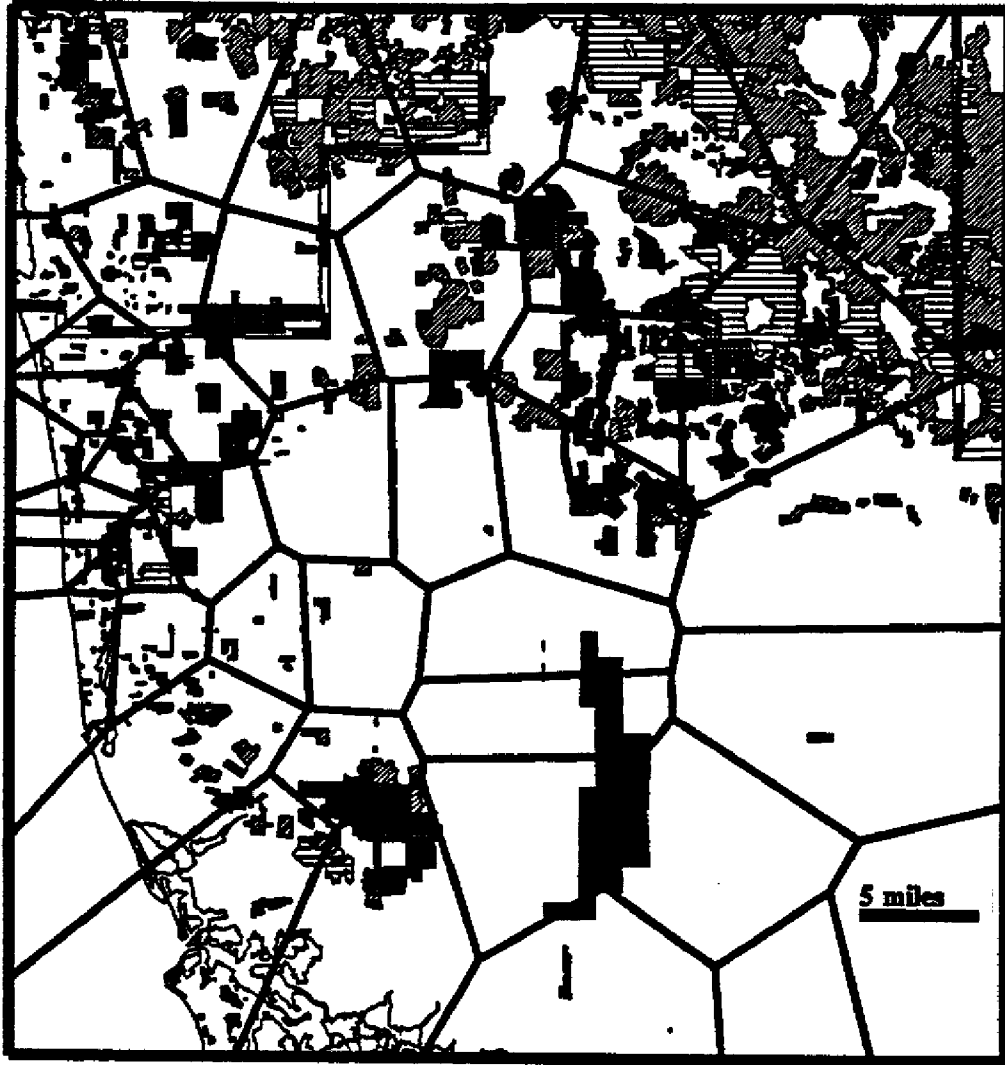
- R_{mean} is the historical monthly average rainfall, $R_{max,k}$ is the maximum rainfall contributed to the groundwater system, and E_k is the monthly pan ET depth, all three quantities are in inches per month.



LEGEND

- Public Water Supply Well
- ⊙ 1-mile Buffer Zone
- County Boundary
- ◇ Thiessen Polygon
- ⚡ Shoreline

Figure 6. Public water supply wells with their one-mile buffer zones.



LEGEND









-  Thiessen Polygon
-  County Boundary
-  Shoreline
-  AC (crop land)
-  AP (pasture)
-  AM (grove, nursery)
-  AF (dairy farm)
-  UO (golf, parks, rec. area)

Figure 7. Agricultural landuse with Thiessen polygons.

6. Optimization of a PR Parameter Set

The parameter set $\{R_{\max,k}, E_k, \text{ and } k_p, k=1, \dots, 12\}$ in the previous conceptual PR model was calibrated by an optimization method. The purpose of this optimization was to obtain an optimal parameter set which maximizes the correlation between the historical PR term and the corresponding head changes, so that the functional relationship of those two variables can be used to estimate the pumpage reduction by the deviation from the anticipated target head. Particularly, the head change $\Delta h_{t,i}^m$ is defined by

$$\Delta h_{t,i}^m = h_{t-1,i}^m - h_{t,i}^m \quad (12)$$

where, $h_{t,i}^m$ is the groundwater head at time t , site i , and aquifer layer m . Equations (2), (7), and (12) may have either positive or negative quantities whose interpretations are different as follows:

	Pumpage/ Recharge	Head Change	Deviation from the target
Equation	$s_{t,i} = E_{t,i} - R_{t,i}$	$\Delta h_{t,i}^m = h_{t-1,i}^m - h_{t,i}^m$	$d_{t,i}^m = H_{k+1,i}^m - h_{t,i}^m$
Positive(+)	SWR	Drawdown	Reduction of pumping
Negative(-)	Recharge	Head increase	No action required

Let us define $\rho_{\Delta h,s}$ as the average correlation coefficient between the historical PR and head change at Layer 1 (because it is the most sensitive to the PR term). $\rho_{\Delta h,s}$ is computed by

$$\rho_{\Delta h,s} = \frac{1}{nx} \sum_{i=1}^{nx} \rho_{\Delta h,s}(i) = \frac{1}{nx} \sum_{i=1}^{nx} \frac{\text{cov}[\Delta h_i, s_i]}{\sigma_{\Delta h_i} \sigma_{s_i}} \quad (13)$$

where nx is the number of Thiessen polygons for the Layer 1, $\text{cov}[\cdot]$ and σ are the sample covariance and standard deviation, respectively. Then, the objective function of the optimization

is to maximize $\rho_{\Delta h,s}$. The unconstrained nonlinear least square with the finite-difference Jacobian method was used to get the optimal parameter set [IMSL, 1991]. For the optimization, the historical $\Delta h_{t,j}^m$ series from January 1987 to September 1993 were prepared, some of which have missing values. The final results obtained are listed in the last three rows in Table 1, where the final objective function for an optimal parameter set was $\rho_{\Delta h,s}=0.619$.

Also practiced was optimization of the conceptual model itself. Alternative conceptual models considered include the direct rainfall, groundwater head, ET by the Blaney-Criddle method, effective rainfall by the SCS method, and the logarithm transformations of one or more of them. The details of these optimizations are not described here, but the results shows that the conceptual PR function was superior to the alternative models.

7. Develop a Control Equation

In order to make the system controllable, let us set up the following time-lagged linear regression equation of $s_{t+,j}$ using the head change in each layer $\Delta h_{t+,j}^m$ as independent variables:

$$s_{t+,j} = \beta_{0,j} + \sum_{m=1}^L \sum_{j=0}^{Nk} \beta_{m,j,j} \Delta h_{t-j+,j}^m + e_{t+,j} \quad (14)$$

where $t+$ means a period between t and $t+1$, $\beta_{0,i}$ is a intercept of regression for polygon i , $\beta_{m,j,i}$ is a regression coefficient for layer m and temporal order j at polygon i , and $e_{t,i}$ is the Gaussian white noise having a mean of zero and a variance of $\sigma_{e,i}$. If the current time is t , $s_{t+,j}$ is given by $\Delta h_{t+,j}^m = (h_{t+,j}^m - h_{t,j}^m)$, where $h_{t+,j}^m$ is the mean square error forecasting conditioned on the previous measurements. This control equation does not use any of the spatial correlation structure of the system. The reason is that introducing a spatial correlation structure in the multi-layered aquifer system results in too many parameters which hinder optimal parameter calibration as well as the manipulation of the equation. However, the term $h_{t+,j}^m$ already includes such spatial correlations via the STARX model, so that estimated $s_{t+,j}$ will more or less include the

spatial characters.

The spatial order Nk , which is considered as a global parameter in the space domain, can be determined by the identification procedure of the stochastic time series model. As a criterion statistics for the candidate temporal orders $Nk=0,\dots,3$, the spatial average $AIC(Nk)$'s were computed by [also refer to the original form given by *Salas et al.*, 1985, page 97];

$$AIC(Nk) = \frac{1}{nx} AIC_t(Nk) = \frac{T}{nx} \ln(\sigma_{e,t}^2) + 2Nk \quad (15)$$

where T is the sample size and $\sigma_{e,t}^2$ is the residual error variance computed by the least square linear regression estimator. The period of record used in this identification procedure was from January 1987 to September 1993 ($T=80$). With sets of optimal parameters of four candidate models, the $AIC(Nk)$'s were computed as: $AIC(0)=145.96$, $AIC(1)=122.17$, $AIC(2)=122.83$, and $AIC(3)=122.56$. Finally, based on the AIC 's, $Nk=1$ was selected for which the spatial average of R^2 statistics was 0.657 and the spatial average of the sum of square error was 360.6 ft². For example, the 28-th polygon which is located at the western coastal area has a control equation of

$$S_{t,28} = 1.614 + 0.942\Delta h_{t+1,28}^1 + 0.728\Delta h_{t+1,28}^2 + 0.258\Delta h_{t+1,28}^3 + 0.374\Delta h_{t,28}^1 + 0.489\Delta h_{t,28}^2 + 0.259\Delta h_{t,28}^3 \quad (16)$$

with a R^2 of 0.765. It should be noted that the superscripts in equation (16) are not the power notations but the indicators for the specified aquifer layer.

8. Target Water level

In addition to the most currently measured heads, the other required inputs to the feedforward control scheme are target groundwater levels, $H_{k,i}^m$ and the permitted groundwater use in terms of PR function, s_i^T . Since the groundwater uses in the model area have been allocated based on the estimated SWR which is computed by the 2-in-10-year drought rainfall,

it is logical to set the target water level as the 2-in-10-year groundwater head. Also, it was assumed that the target water levels are a monthly distributed spatial function. That is, the desired target is subject to change by the crop irrigation requirement. To explain the procedure, let us define a 4-dimensional random variable h_{ijk}^m as the groundwater head at site i , year j , month k , and layer m . After estimating the seasonal parameters by fitting the normal distribution functions at site i month k and layer m , the target water levels h_{ij}^m 's are computed from the quantile corresponding to the 2-in-10-year drought frequency. The period of record in this frequency analysis was from January 1977 to August 1993. The two parameter normal distribution was fitted by the method of moment, and Chi-square statistics was used to justify the goodness-of-fit of the distribution function, resulting in 92% of fittings accepted by the null hypothesis of the normal distribution with a 5% significant level.

In addition, the following constraints were used to adjust the target levels: in order to prevent salt water intrusion, the target water level (for Layer 1, 2, and 3) is adjusted to zero feet NGVD if a site is located within five miles of the coast and the target level is below zero feet NGVD. Also, in the confined aquifer (Layer 4), the target water level is adjusted to the top elevation of the confined aquifer if the target level is below the top of confined aquifer. This latter constraint is important to maintain the structural integrity of the limestone aquifer, in which the hydrodynamic pressure of the groundwater provides a significant amount of support against collapse and possible sinkhole formation.

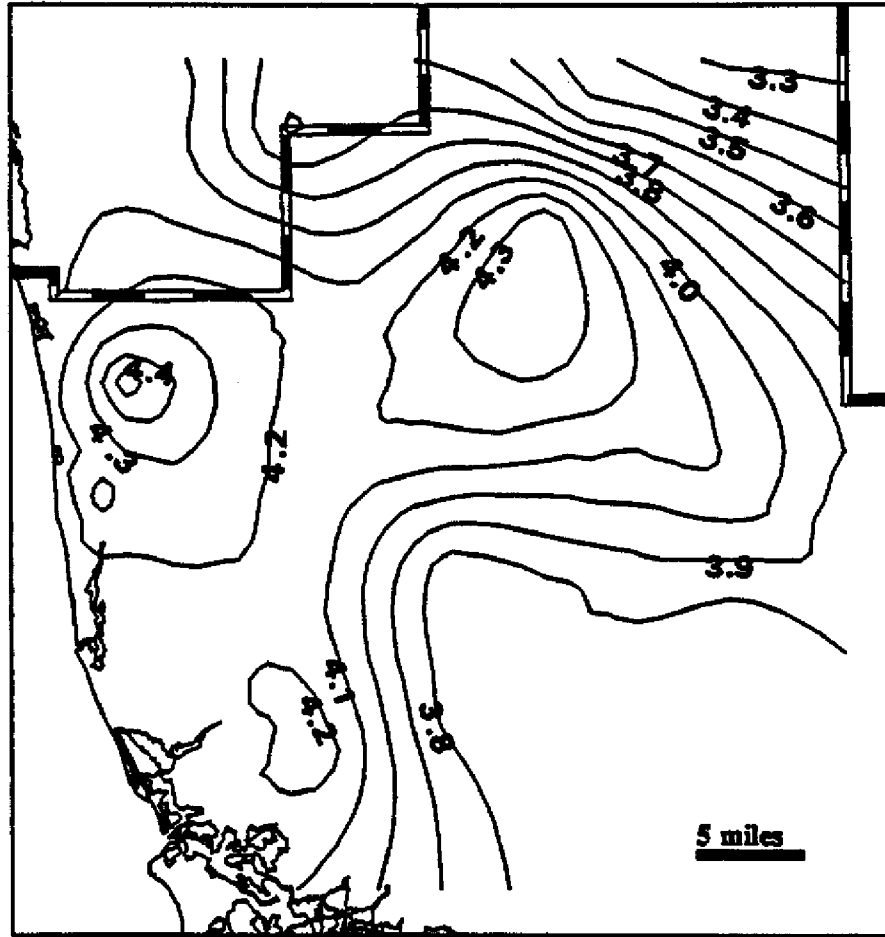
The monthly 2-in-10-year PR functions were computed by the PR procedure with the average 2-in-10-year rainfall depth in Table 1, which was computed from the 27-year historical rainfall in the region. Then, the s_i^T for polygon i was determined by selecting the largest such PR function at polygon i . Figure 8 is a contour map showing the spatial distribution of s_i^T , which indicates that the larger groundwater uses are located in the northern portion of the model area.

9. Simulation and Discussion

The proposed feedforward control scheme was tested for various drought conditions to investigate variations of pumpage reduction with respect to the different drought frequencies. As mentioned before, the calibration period for both the forecasting and control equations was from January 1987 to August 1993. A one-year warm-up period was used for this test simulation. Without having an actual drought event during that warm-up period, expected drought events having five different return frequencies ($Tr=2, 5, 10, 20,$ and 50 years) were generated. Such a data can be generated by the following two methods. First, the monthly groundwater head series $h_{t,i}^m(p)$, $t=1, \dots, 12$ having an average probability is p can be generated by means of the STARX model. But due to seasonality, that series is nothing but one realization, and there exists an infinite number of such series. An average of those series by month becomes an expected drought event, $E[h_{t,i}^m(p)]$, $t=1, \dots, 12$. Thus, as a second approach, the expected drought series having probability p can be obtained, not by the above data generation method, but from the frequency information of the historical data. The latter method was used to create the drought events during the warm-up period. In either case, the space-time correlation of the system can not be preserved.

Since the seasonal target heads were used in the feedforward control scheme, the simulation was performed in both May (end of dry season) and September (middle of wet season). For instance, when the present time t is in May, the warm-up period was from the previous June to this May and heads at $t+1$ (=June) were forecasted spatially, then the recommended pumpage reduction $r_{t,i}$ during June was computed.

As summary statistics of the above test, the mean $\mu_r=E[r_{t,i}]$ and standard deviation $\mu_r=\{\text{var}[r_{t,i}]\}^{1/2}$ of pumpage reduction rates were computed for different frequencies p during the warm-up period. This simulation was performed for both wet or dry conditions and the results are shown in Figure 9. Also, Figure 10 shows an example of pumpage reduction contour map during the June case with the return period of 20 years.



LEGEND




-  Contour
-  County Boundary
-  Shoreline

Figure 8. The computed 2-in-10-year recharge/pumpage (inches/month).

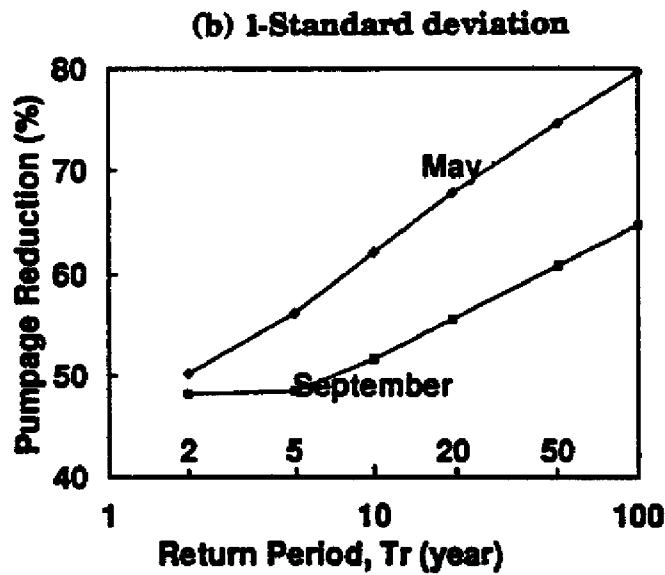
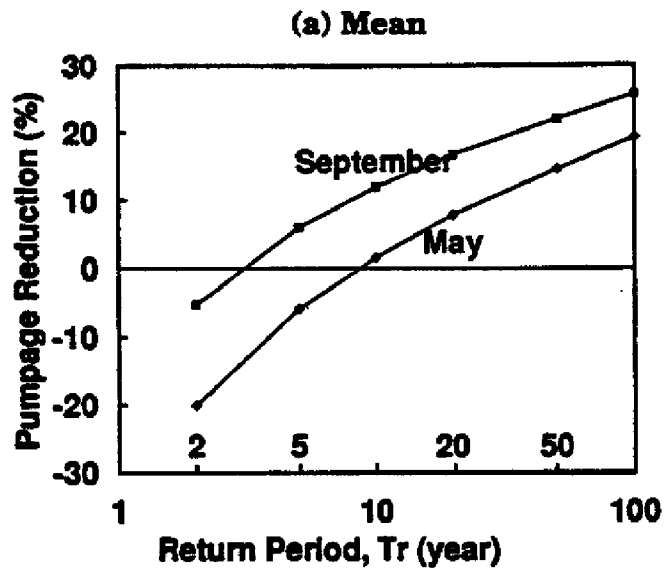
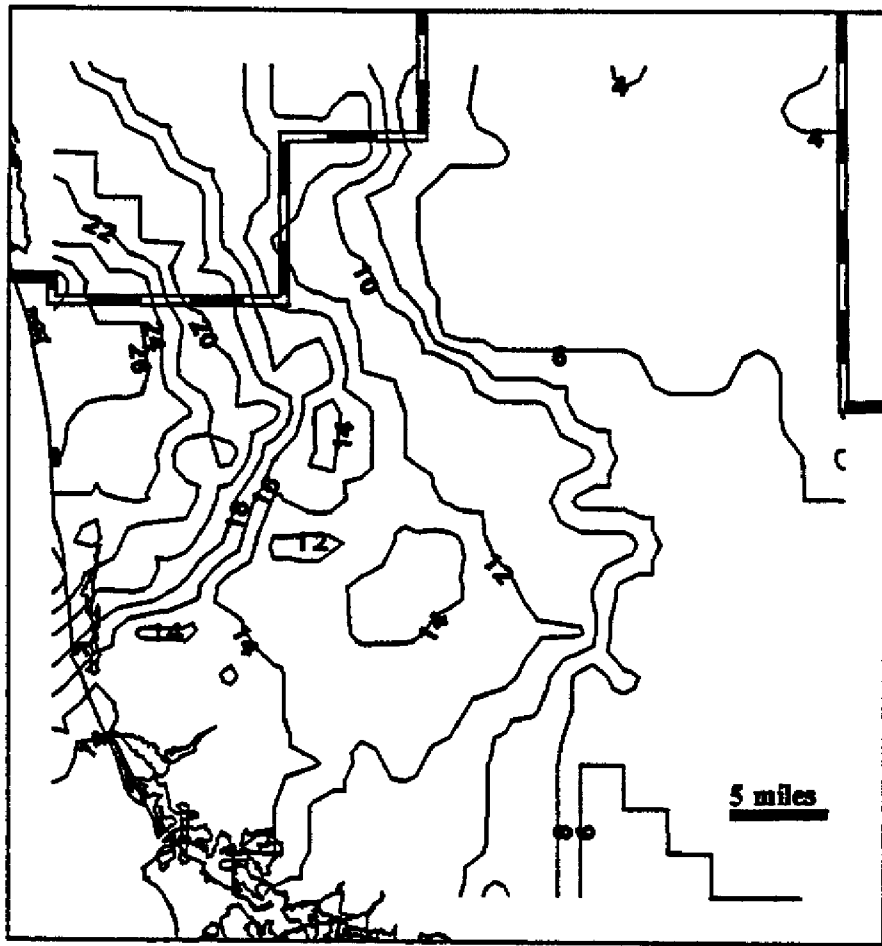


Figure 9. Statistics of the pumpage reduction versus return period of drought.



LEGEND




-  Contour
-  County Boundary
-  Shoreline

Figure 10. An example of recommended pumpage reduction (percent) in case of a 1-in-20-year drought event.

From these simulations, the following results were noted:

(1) The relationship between pumpage reduction and return period T_r of antecedent head condition during the warm-up period behaved nonlinearly. However, the pumping reduction rates, in terms of both μ and σ , were approximately linearly proportional to $\ln(T_r)$ for $t_r > 5$ years, and this linearity was more acceptable when the groundwater head is less than the 2-in-10-year target level.

(2) The spatial variation of the pumpage reduction was much greater than that by the frequency variation. This result strongly supports the concept of spatial forecasting and management of groundwater in the model area, instead of the uniform pumpage reduction previously practiced by the District.

(3) The average pumping reduction rate μ during the wet season was higher than that during the dry season, despite the higher expected rainfall during the wet season. This is mainly due to the increasing supplemental water requirement from the agricultural fields during the summer season. However, the spatial variance of reduction rate σ during the wet season was smaller than that of the dry season.

(4) The average pumping reduction rate of both the wet and dry seasons was near zero when T_r is 5 years. That is, the estimated pumpage reduction rate based on the groundwater head difference matches well with the historical target level. This result validates the proposed feedforward control scheme, specifically the assumptions of using 2-in-10-year seasonal target water levels along with the 2-in-10-year groundwater use permit based on the proposed PR function.

10. Conclusion

A feedforward control scheme was developed to handle the regional groundwater drought problem. This scheme forecasts the future groundwater heads in the multi-layered aquifer system by the STARX model and computes the deviations from the seasonal target level. Using the empirical control equation based on the relationship of head change and the conceptual pumpage/recharge function, spatial pumpage reduction rates are computed from the present and forecasted heads. The recommended pumpage reduction scheme can be applied to the all different kinds of groundwater users since the pumpage/recharge function accounts for a variety of groundwater uses components. As model verification, simulations were performed for different seasons and different drought frequencies. Results validate the feedforward control scheme along with the empirical control equation based on the proposed pumpage/recharge function.

Advantages of the proposed control scheme are that it is simple in nature, thus more intuitive, and that it is possible to control the drought spatially, instead of uniformly as in the previous method. Another advantage comes from the use of the conceptual pumpage/recharge function in estimating the spatial recharge in the groundwater system. Limitations of this control scheme are that the recommended pumpage reduction is that of the lumped layers, not layer-by-layer, that it does not provide reduction rates well-by-well or different water uses classifications, and that adaptability of the forecasting model, such as changing model structure, is difficult. To use the proposed feedforward control scheme, collection of timely measured groundwater head data in the model area is critical.

REFERENCES

- Bear, J., *Hydraulics of groundwater*, McGraw-Hill, New York, 1979.
- Bennett, R. J., *Spatial Time Series: Analysis-Forecasting-Control*, Pion Ltd. 207 Brondesbury Park, London, 1979.
- Bennett, M. W., "A Three-dimensional finite difference ground water flow model of western Collier county, Florida", Technical Pub. 92-04, South Florida Water Management District, West Palm Beach, Florida, 1992.
- Box, G.E.P., and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.
- Dracup, J.A., K. S. Lee, and E. G. Paulson, Jr., On the definition of droughts, *Water Resources Researches*, 16(2), 297-302, 1980.
- IMSL, *Math/Library User's Manual*, FORTRAN subroutines for mathematical applications. Version 2.0, Houston, Texas, 1991.
- Jensen, M.E., R.D. Burman, and R.G. Allen, *Evapotranspiration and Irrigation Water Requirements*, ASCE Manuals and Reports on Engineering Practice No. 70, ASCE, New York, 1990.
- Kite, G. W. *Frequency and Risk Analysis in Hydrology*, Water Resources Publications, Fort Collins, Colorado, 1977.
- McDonald, M.G., and Harbaugh, A.W., *A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model: U.S. Geological Survey Techniques of Water Resources Investigations*, Book 6, Chapter A1. 1988.
- South Florida Water Management District, Permit Information Manual Volume, III (Revised version), South Florida Water Management District, West Palm Beach, FL, January 1993.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and L.W. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Fort Collins, Colo., 1980.
- Stoffer, D. S. Estimation and identification of space-time ARMAX models in the presence of missing data. *Journal of the American Statistical Association*, 81, 762-772, 1986.

Part IV

Drought Management Decision Support System (DMDSS)

User's Manual

Version 1.0 (October 1994)

THIS PAGE INTENTIONALLY BLANK

1. Introduction

The Drought Management Decision Support System (DMDSS) is a user-friendly application designed to help in the prediction of future ground water levels for Collier county, Florida. The forecasting method is based on Kalman filtering associated with Space-Time Autoregression with exogenous variables (STARX) model. Historical data for calibration of the STARX model consists of 17 years (1977-1993) of monthly groundwater heads from 115 monitoring stations. However, the DMDSS database contains 44 years (1950-1993) of data from the above 115 stations (record lengths vary station-by-station), as well as 44-year of monthly rainfall data from 21 stations, to let the user understand the long term historical trend of groundwater heads in the model area.

The DMDSS allows users to look at the temporal and spatial distribution of rainfall and groundwater head data across the model area, set parameters for existing conditions, and execute the forecasting model for the next twelve-month period. Results of the model can be displayed graphically with different options including spatial interpolation, contouring of the future groundwater head, displaying time series by two-dimensional graphs, etc. It also possesses the capability of saving and printing the analysis and the results.

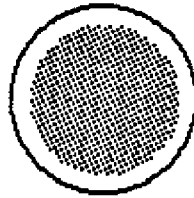
The DMDSS, requiring spatial analysis for its analyses, makes extensive use of GIS technology. Arc/Info (version 7.0), the standard GIS package used at the South Florida Water Management District, was selected as the basis for the development of graphical interface. Considering the complexity of Arc/Info package itself and wide range of users type, the application focuses on simple point and click operation. It consists of a series of menus written in Arc Macro Language (AML) which allows users to perform the operations without requiring any in-depth knowledge of the ARC/INFO process involved. It also includes supporting programs for data manipulation and statistical computation written in FORTRAN.

DMDSS

DMDSS

*Draught Management
Decision Support System*

Forecasts and Management of Groundwater droughts
In Collier County, Florida



South Florida Water Management District
Water Resources Evaluation Department
Hydrogeology Division

Version 1.0 (Oct. 1994)

Continue)

Disclaimer)

Help)

Quit)

2. How to Start the Application

The DMDSS program is operated from within the Arc/Info software. If unfamiliar with the command for invoking Arc/Info on your system, users should ask their UNIX system manager as well as the GIS manager to set up the appropriate Arc/Info environments. To run this application, user should start the Arc/Info application and type "gwmanarc" (without quotes) at the UNIX prompt. It will by default create a directory "dmdss" in the user's home directory and change the workspace to that directory. If the "dmdss" directory already exists, it will simply change the workspace to it.

```
Example :   muir% runarc6  
           arc: gwmanarc
```

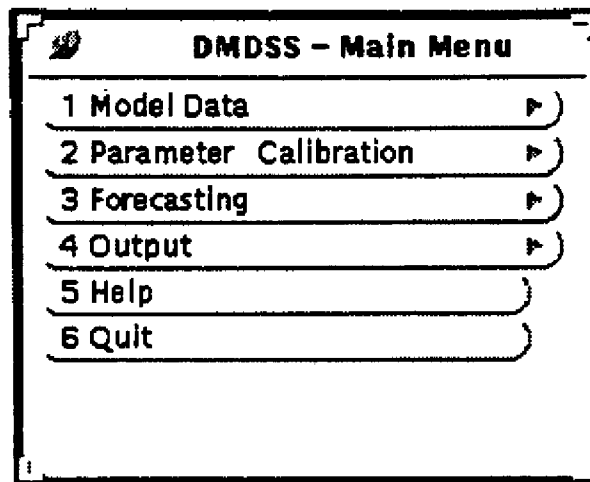
The above process will result in popping up of the following title menu on the screen. The four available options at this title menu are:

- continue* will result in displaying the main menu,
- disclaimer* of the software,
- popping up a *help* menu,
- quit* the application, respectively.

The main functionality of the title menu is initiation or re-initiation of the DMDSS application. In case of re-initiation, all pre-set parameters and options from the previous run will be changed to the system defaults.

3. Overview of the Main Menu

The main menu comprises four major components as shown in the following figure:



Each submenu allows the following options:

1. **Model data:** Display historical data stored in the info database and all pertinent GIS coverages. The purpose of this option is to let the user understand the characteristics of the historical groundwater head and rainfall, as well as to familiarize the user with the geography of the model area.
2. **Parameter Calibrations:** Perform calibration of the STARX model parameters using updated information. This option be used only when the input data are significantly updated. It is recommended to recalibrate the STARX model every two years or after experiencing a severe drought event, whichever comes first.
3. **Forecasting:** Run a STARX model in the forecasting mode to predict the next twelve month groundwater heads after updating the present condition.

4. Output: Display the forecasting results, as well as the errors and confidence intervals in temporally and spatially.

Each of these functions involves one or more sub-menus, that can be invoked from the main menu by simply clicking the desired choice. Subsequent choices under the four main functions represent the following various features offered to the user. It should be noted that the *historical data* menu can be viewed separately for spatial mapping and time series.

1. Model Data

- 1.a. Historical data

- A. Contour Map

- B. Time Series

- Monthly Statistics

- Yearly Statistics

- Correlation

- 1.b. GIS Coverage

2. Parameter Calibration

- 2.a. Spatial Index Matrix (SIM)

- 2.b. Update Yearly Data

- 2.c. Update Monthly Data

- 2.d. Parameter Calibration

3. Forecasting

- 3.a. Parameters

- 3.b. Present Conditions

- 3.c. Execute Forecasting Model

4. Output

- 4.a. Spatial Mapping

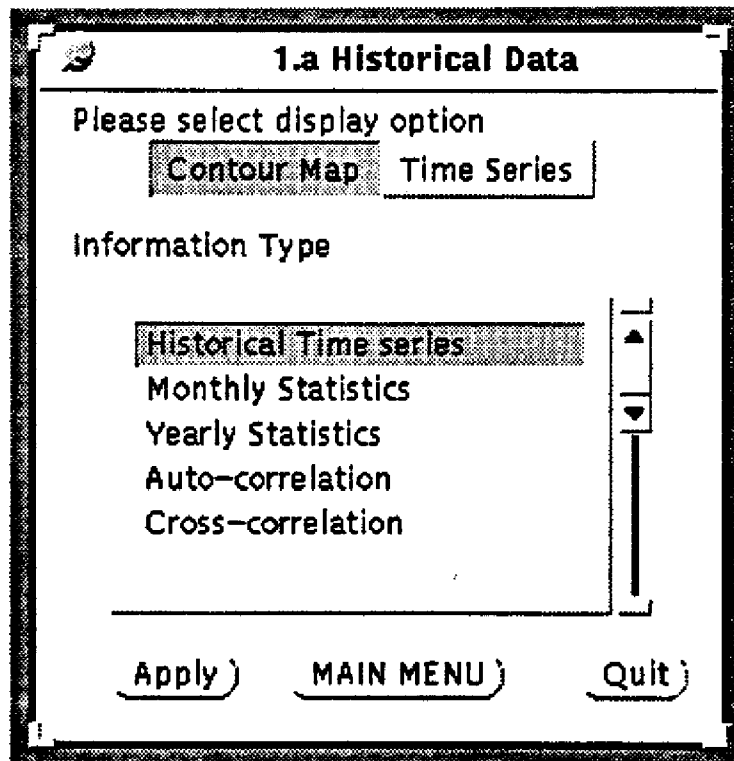
- 4.b. Time Series

4. Description of Menus

4.1. Model Data

1.a. Historical data

The historical data menu as shown in the following figure allows the user to display contour maps and time series for five different options, resulting in ten different menus. The purpose of this submenu is to display historical data stored in the info database which has the groundwater heads from 115 monitoring stations and rainfall from 21 stations in the model area. To initiate the desired menu, the user should select the desired display option first (contour map or time series) and then select the information type. Next the user must click on the "Apply" button, which will pop up another appropriate submenu. The *Main menu* button returns to the main menu while the *Quit* button returns to the main menu.



A. Contour Map

This menu helps in creating a contour map of desired data type for a specific time. The process of creating contours involves interpolation of point data to produce a continuous distribution for the model area before generating contours. The method used for interpolation is the universal kriging with the linear semi-variogram function to model the drift. Final contour output includes the interpolated contour map as well as the map showing standard error for the model area. It also has the capability to *save* and *print* the generated contour map. This submenu permits the user to change the workspace and view any coverage(s). Following are the steps to create a contour map.

- Step 1: Select the appropriate data type (Rainfall, Pumpage, or Potent_head).
- Step 2: Set appropriate month and year by clicking month and year scroll bar.
- Step 3: Choose desired contour interval.
- Step 4: Click the *Apply* button.

A brief description of other menu buttons is given below.

Save:	Saves the generated map under different name.
Print:	Prints a hard copy of the created map.
Up:	Moves up a workspace
Down:	Moves down a workspace
Draw:	Draw the selected coverages in the scrollbar above.
Main menu:	Returns to the main menu
Previous menu:	Returns to last menu.
Quit:	Returns to the title menu.

The user should note that the contour map display option for the other options, such as monthly statistics, yearly statistics, auto-correlation, and cross-correlation, are similar in nature and can be used by the above process.

Historical Data - Contour

DATA TYPE

Rainfall

Pumpage

Potent_head

YEAR

Jan

Feb

Mar

April

May

MONTH

1950

1951

1952

1953

1954

COVERAGES

└ Basemaps

draw12_93

draw2_93

draw4_93

CONTOUR INTERVAL

5 10

Please Select/Type A Workspace

/usr/users/hahn/dmdss/

APPLY)

MAIN MENU)

SAVE)

PREVIOUS MENU)

PRINT)

HELP)

UP) DOWN) DRAW)

QUIT)

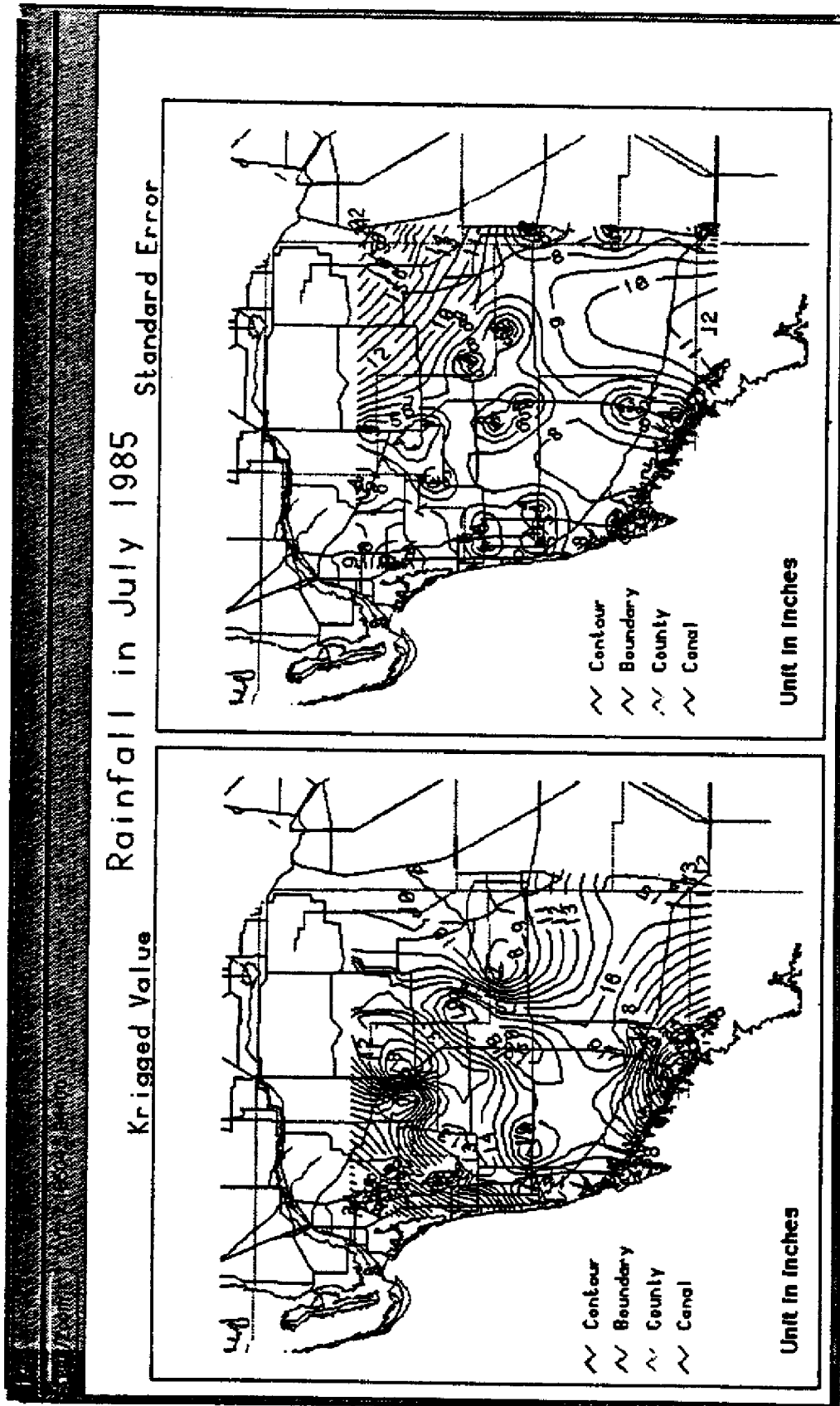


Figure 1. Sample output for historical contour map.

B. Time Series

The purpose of this menu is to let the user see seasonal variations of one or more station for a specified time period. It plots a graph for the selected station point between time and corresponding values. Described below are the steps to use this menu.

- Step 1: Select desired data type by clicking on the appropriate button (rainfall, pumpage, or Potent_head). A small menu will appear requiring input for the number of stations to be viewed. User then can select the stations by clicking on the desired gaging stations from the active Arc/Plot window.
- Step 2: Repeat the process for other data type if required.
- Step 3: Click on *Display* button.

A brief description of other menu buttons is given below.

- Save: Saves the generated time series plot under different name.
- Print: Prints a hard copy of the generated time series plot.
- Help: Pops up a help menu.
- Main menu: Returns to the main menu
- Previous menu: Returns to last menu.
- Quit: Returns to the title menu.

Monthly Statistics

Same as the subsection 1.a

Yearly Statistics

Same as the subsection 1.a

Correlations

Not available in the current version

Historical Data - Time Series			
Year(from)	Month(from)	Year(to)	Month(to)
1950	Jan	1986	Aug
1951	Feb	1987	Sept
1952	Mar	1988	Oct
1953	April	1989	Nov
1954	May	1990	Dec

Select Data Type

DISPLAY)
 RAINFALL)
 PUMPAGE)
 POTENT_HEAD)
 SAVE)

PRINT)
 PREVIOUS MENU)
 MAIN MENU)
 HELP)
 QUIT)

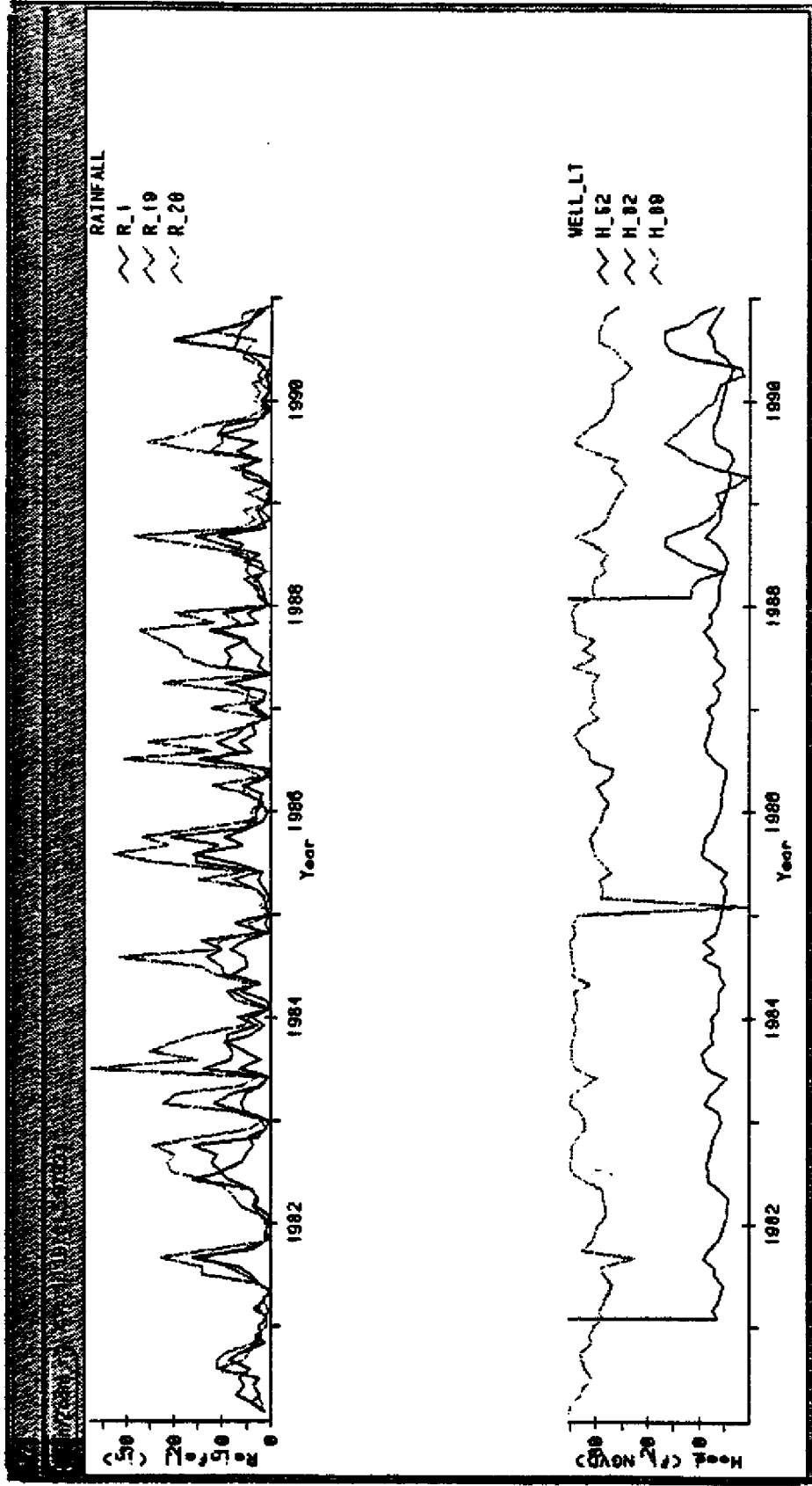


Figure 2. Sample output for historical time series.

1.b. GIS Coverage

GIS data is a display menu permitting the user to view base maps, gaging stations, aquifer elevations, and other pertinent GIS coverages in the model area. The base maps include district boundary, canal, lake, county boundary, and major roads. The *GIS data* option helps to set Tics which are registration or geographic control points for GIS coverages that represents known locations on the appropriate coordinate system. The Tics allow the coverage to be standardized in a particular coordinate system, allows map sheets to be registered for digitizing, and also serves as the basis for transforming features into new coordinate systems. It is generally advantageous to have each coverage registered to the same set of Tics.

To use this menu the user must click on the desired checkboxes and hit the *Draw* button. The identify button allows the user to get information in detail about any particular feature of the displayed coverage. Please refer to the Arc/Info documentation (Map projections and coordinate management volume) for more details on tics, projections, and transformations.

A brief description of other menu buttons is given below.

- Up: Moves up a workspace from the present working directory to search for appropriate GIS coverages.
- Down: Moves down a workspace.
- Draw: Draw the selected coverages in the scrollbar.
- Help: Pops up a help menu.
- Identify: Identifies any particular feature of displayed coverage in detail. It will invoke a pop up window telling the type of GIS coverage (point, arc, tin, lattice, etc.).
- Main menu: Returns to the main menu
- Quit: Returns to the title menu.

1b. GIS Data

Historical Data Set for Model

Base Map

- Boundary
- Canal
- Lake
- County
- Roads

Acquifer Elevation

- Surface_Elev.
- Top of Layer2
- Top of Layer3
- Top of Layer4

Gaging Station

- Rain
- Pumpage
- Well 1: Surficial
- Well 2: LTamiami
- Well 3: Sandstone
- Well 4: M.Hawthorn

Modflow Coverages

- draw12_93
- draw2_93
- draw4_93
- draw5_93

DRAW) CLEAR)

IDENTIFY)

MAIN MENU)

Please Select a valid path
◀ |rs/hahn/dmdss//draw1 | ▶

UP) DOWN)

HELP) QUIT)

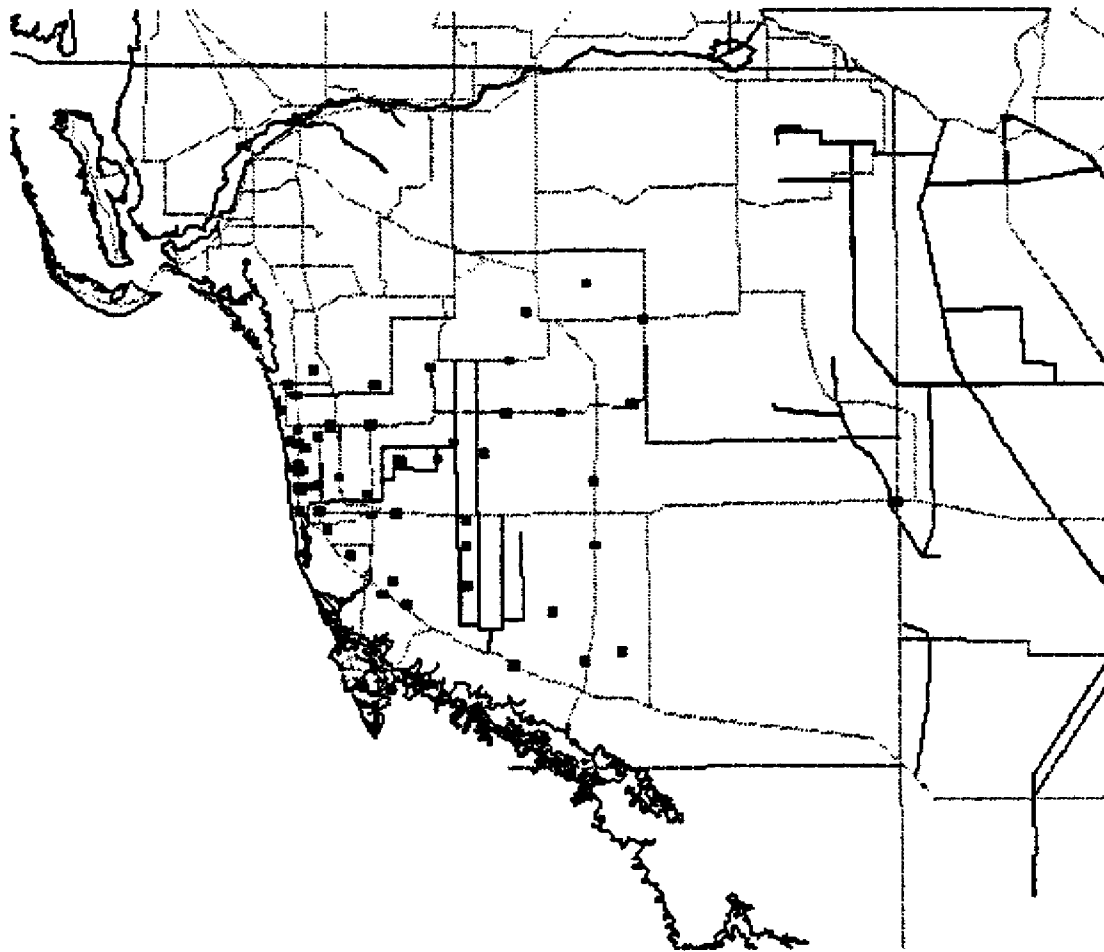


Figure 3. Model area with basemap.

4.2. Parameter Calibration

2.a. Spatial Index Matrices (SIM)

SIM is a method to establish the correlation of station points to five neighbor points in the current layer, as well as with the layer above, the layer below, and rainfall. This menu allows the user to understand the relationship by visually displaying the related station points on the screen. It makes a convex polygon of the five selected points from which user can easily identify the boundary of the site interested. The convex polygon is a polygon in which no internal angle is more than 180 degree. To use this menu the user should select an appropriate layer and display option, then click on the *Apply* button.

The three display options are available:

- Overall: Displays the relationship for all the station points at the same time.
- Sequential: Displays convex polygons one-by-one in fixed sequence of station identification numbers which were assigned from west-south to east-north direction in general.
- Interactive: Displays a convex polygon for a selected gaging station, providing more flexible views to the user.

Other menu buttons are:

- Main menu: Returns to the main menu.
- Apply: Initiate display on the Arc/Plot window.
- Quit: Returns to the title menu.
- Help: Pops up a help menu.

2a. DW Matrices

SELECT LAYER

Layer1
Layer2
Layer3
Layer4

DISPLAY OPTIONS

Overall
Sequential
Interactive

APPLY HELP
MAIN MENU QUIT

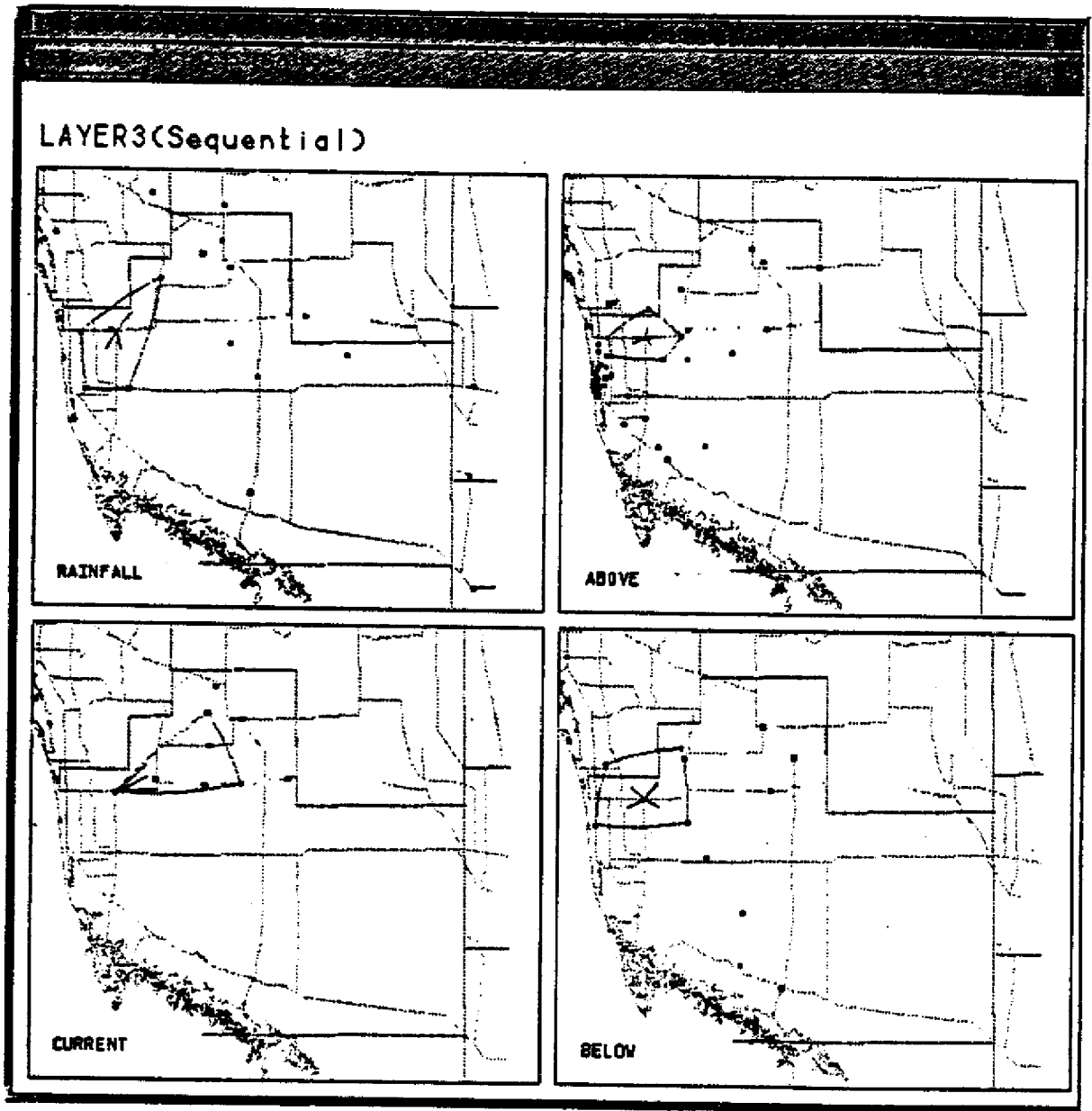
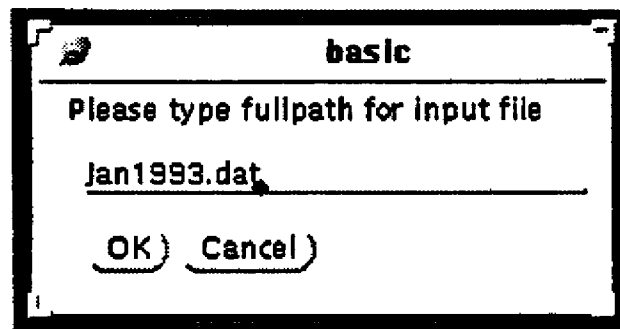


Figure 4. Display the selected neighbors by the different options.

2.b. Update Yearly Data

The update yearly data menu simply inputs the specified name of the ASCII file containing yearly data in a column format (ID, Station ID, values) and updates the existing info database. The user must have a write permission for the data files to be able to update them. Since this option adds the data to the central database, only an authorized user (DMDSS manager) can do it. If unauthorized users try it, it will kick back to the previous menu with a warning message.



2.c. Update Monthly Data

Same as the *update yearly data* option.

2.d. Parameter Calibration

This menu allows the user to recalibrate the parameters for the STARX model using the expectation-maximization algorithm associated with a simplified smoother estimator (EMSSE) method which is coded by FORTRAN and attached to the DMDSS. (Not available in the current version).

4.3. Forecasting

3.a. Parameters

This option allows the user to look at the calibrated parameters for the STARX model in the form of ASCII file from the pop up window, from which the user can understand the way the space-time forecasting is done in the STARX model (Not available in the current version).

3.b. Present Condition

The main functionality of this option is to display currently updated data. Both extended time series and contour maps can be generated by this option, with the stations which have missing values at the present time (Not available in the current version).

3.c. Execute Forecasting Model

This menu runs the forecasting model for a period of the next twelve months and import the forecasted values and associated errors into the info database automatically. The user simply defines the current time by clicking the appropriate month and year and hitting the *Execute* button which results in the forecasting of groundwater heads for next 12 months. Other menu buttons descriptions are:

Main menu:	Returns to the main menu
Quit:	Returns to the title menu.
Help:	Pops up a help menu.

3c. Forecasting Model

Please Select Current Date

MONTH	YEAR
Jan	1993
Feb	1994
Mar	1995
April	1996
May	1997

EXECUTE) HELP)

MAIN MENU) QUIT)

4.4. Output

4.a. Spatial Mapping

The spatial mapping menu let the user see the forecasted output for two different times and compare them with current and historical conditions. It also calculates and displays drawdown and water table depth from the land surface elevation for forecasted times. It is necessary to execute the forecasting model before output is displayed. The following steps describes the usage of this menu.

- Step 1: Select the data type. (Drawdown and WT_depth cannot be calculated unless forecast map previously have been created for that period).
- Step 2: Type/Select the current time for which the forecasting model has been executed previously.
- Step 3: Set appropriate contour intervals.
- Step 4: Choose the lead time for forecasting.
- Step 5: This step has two options: The first option allows the user to display all the maps at the same time by clicking the *All* button. This option is available only for forecasting and not for the drawdown and water table depth display. The second option allows the user to display each map individually by clicking appropriate *Do* button.

Other menu button description are

- Save: Saves the generated map under different name.
- Print: Print a hard copy of a map.
- Help: Pops up a help menu.
- Main menu: Returns to the main menu
- Quit: Returns to the title menu.

4a. Spatial Mapping

Please Select Data Type

Rainfall	▲
Potent_head	▼
Draw_down	
WT_Depth	

Please Type Current Date

Month Year

CURRENT CONDITIONS

Contour interval
1 1 10

FORECAST_1

Contour Interval
1 1 10

Lead Time

1	2	3	4	5	6	7	8	9	10	11	12
---	---	---	---	---	---	---	---	---	----	----	----

FORECAST_2

Contour Interval
1 1 10

Lead Time

1	2	3	4	5	6	7	8	9	10	11	12
---	---	---	---	---	---	---	---	---	----	----	----

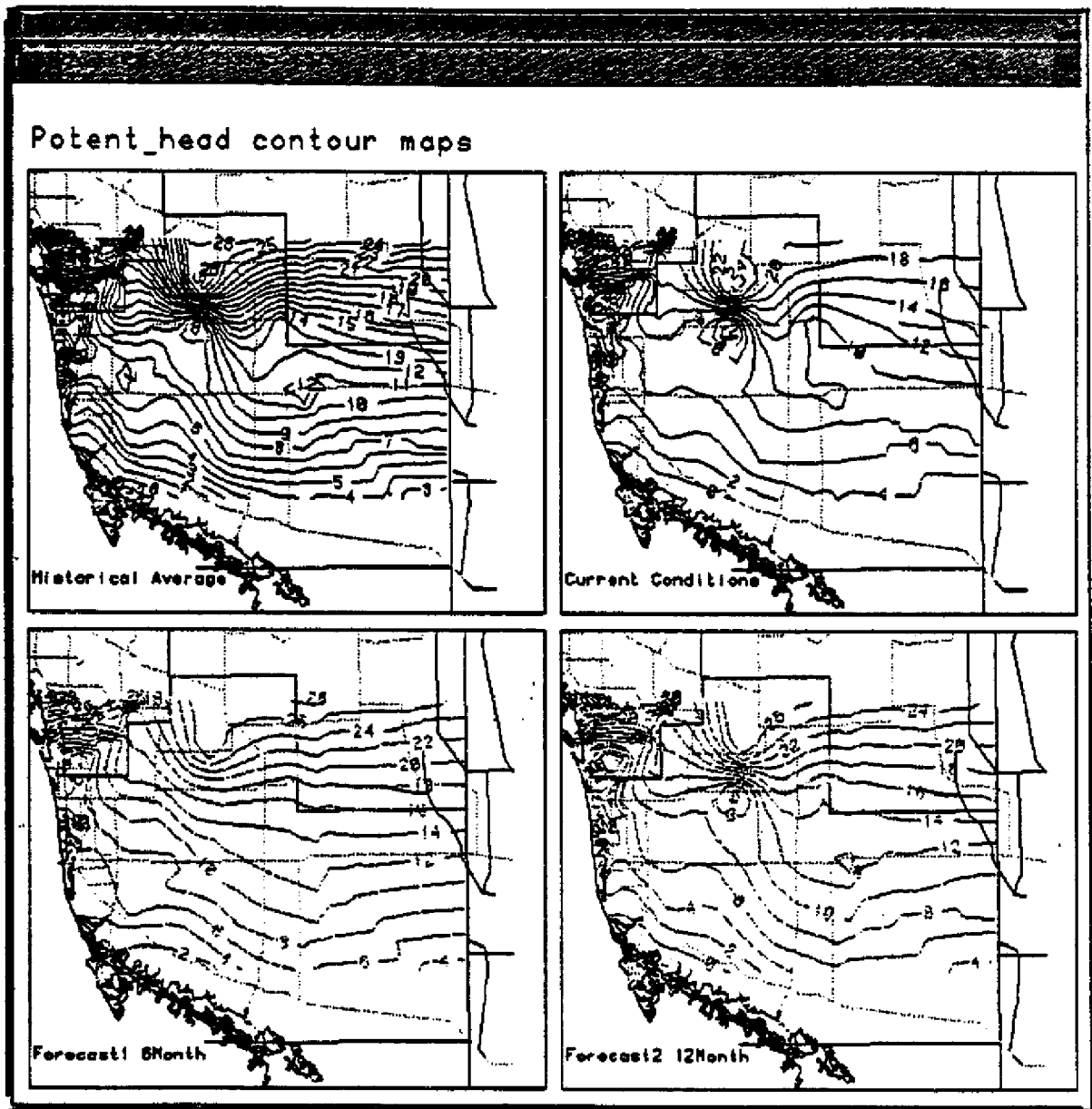


Figure 5. Output contour maps showing 12-month lead forecasting.

4.b Time Series

This output option displays the time series projection for next twelve month period with confidence intervals of $\pm 50\%$ and $\pm 95\%$, along with the previous twelve month data. The procedure to use this menu is described below.

- Step 1: Type/Select the current date (month/year) for which the forecasting model has been previously executed. The current date should be the same as that specified in the forecasting option.
- Step 2: Select desired data type by clicking on the appropriate button (Rainfall, Potent_head). If the user selects the *Rainfall* button, it will display only the previous twelve months data. The user then can select the stations by clicking on the desired point. The process is repeated if other data types also need to be selected.
- Step 3: Click the "Display" button to plot the time series on screen.

A brief description of other menu buttons is given below.

- Save: Saves the generated plot file under different name.
- Print: Print a hard copy of a map.
- Help: Pops up a help menu.
- Main menu: Returns to the main menu.
- Quit: Returns to the title menu to finish the application.

4b. Time Series

Please Select/Type Current Date
Mon: _____ Year: _____

Select Data Type
RAINFALL) POTENT_HEAD)

DISPLAY) SAVE) PRINT)

MAIN MENU) HELP) QUIT)

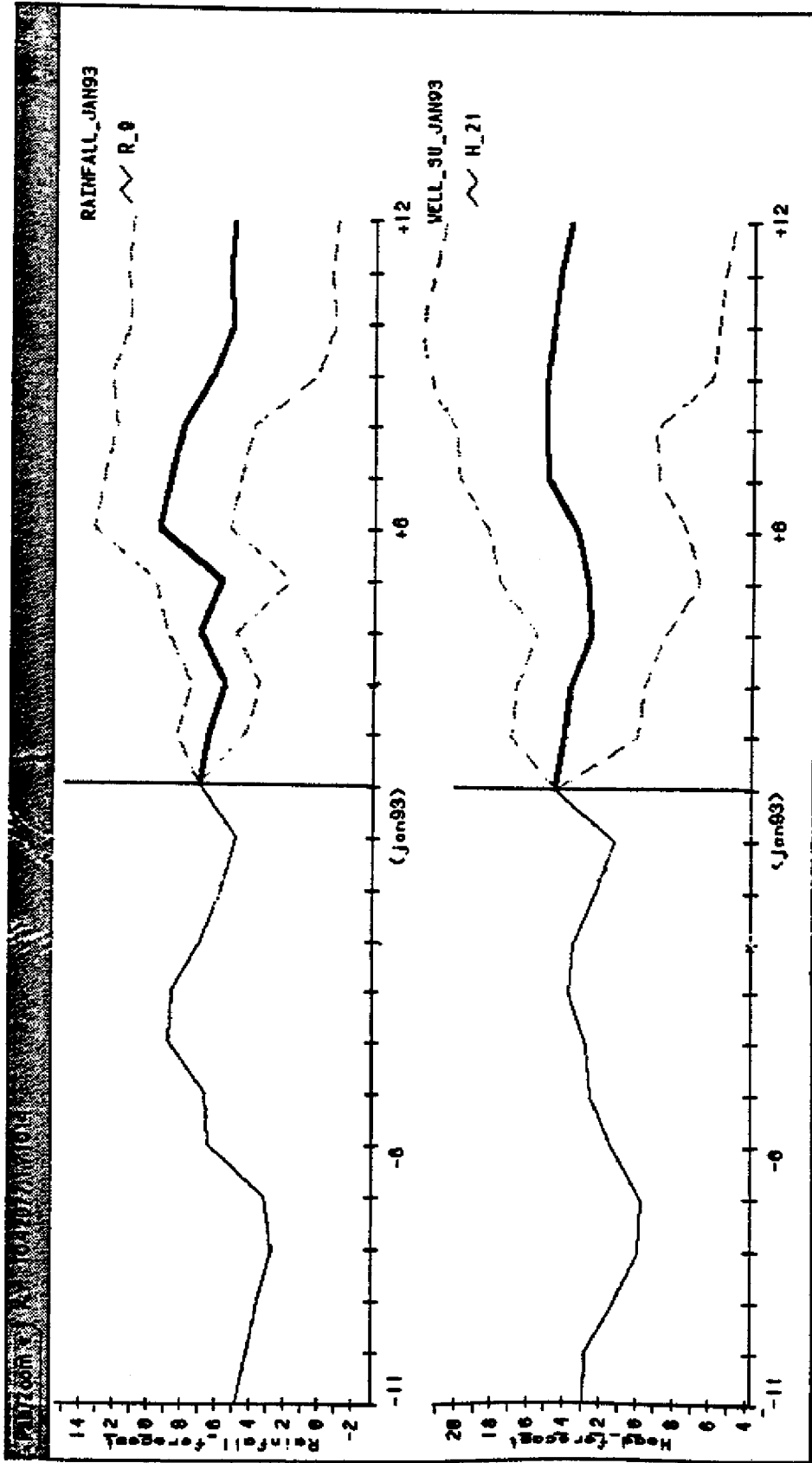


Figure 6. An example of forecasted results for station ID #21 over next twelve months.