

Special Publication SJ92-SP8

STATISTICAL CHARACTERIZATION OF GROUND WATER FLUCTUATIONS  
AND HYDROGEOLOGICAL PROPERTIES IN THE FLORIDAN AQUIFER

for

St. Johns River Water Management District  
Palatka, Florida

by

Claude D. Tankersley<sup>1</sup>  
Wendy D. Graham<sup>2</sup>  
Kirk H. Hatfield<sup>1</sup>

Departments of

<sup>1</sup>Civil Engineering  
<sup>2</sup>Agricultural Engineering

University of Florida  
Gainesville, Florida

September, 1991

Revisions to Special Publication SJ92-SP8  
 Statistical Characterization of Ground Water Fluctuations  
 and Hydrogeological Properties in the Floridan Aquifer  
 June 1992

Page	Description
IV-21	Paragraph 1, change third sentence to read: "Thus for small distances on the order of $0.01L/2$ , where $L$ is the dimension of ...."
IV-40	Paragraph 2, change first sentence to read: "It will be recalled that grouping data ... particularly significant for small distances (i.e., on the order of $0.01L/2$ )."
IV-56	Table 4.5, change variance for original data from 0.00000 to $1.12 \times 10^{-6}$
V-51	Paragraph 1, line 8, change (5.89) to (5.87)
V-84	Equations (5.97a) and (5.97b), change 0.638 to 0.683
V-91	Table 5.5, Well 21, second to last column, change -0.12641 to 0.12641 (drop the negative)
V-93	Replace data in Table 5.6 as noted below

Lead-1	Lead-3	Lead-6
Actual (ft)	Actual (ft)	Actual (ft)
4.27911	5.03234	5.22820
0.43608	1.07905	1.70084
0.88798	1.87703	2.34453
0.61741	1.35791	2.05172
0.39660	1.11668	1.81831
1.44866	1.75001	1.86715
0.57333	0.99247	1.20514
0.35740	0.71990	0.87110
0.52358	1.30178	1.97632
0.62423	1.49833	2.26891
0.92918	1.64148	1.87394
0.56869	0.91645	1.01957
0.93286	1.57532	1.93579
0.54761	0.85909	0.95534
1.13117	2.03125	2.64032
0.68346	1.12839	1.29738
1.02299	1.48112	1.57691
0.46985	0.81831	1.01720
0.62896	1.08533	1.26723
0.74253	1.23014	1.43072
0.89946	1.21393	1.37955
0.72110	1.28370	1.62490

V-95 Table 5.7, change values for Well 11 as noted below in order from left to right

0.829, 42.29, 5.37, 40.68, 0.056, 17.71, 2.81, 39.07, 0.013, 79.36, and 1.81

VI-8 Paragraph 1, last line, change "... observation at location j for location x." to "... forecast at location j for location x."

AB-1 Replace data in Appendix B as noted below

Well Number	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
2	40.35		44.77	35.96
11	42.29	1.61	46.12	38.55
12	29.67	1.12	31.77	25.27

AC-2 Replace data in Table C.2 as noted below

Month	Mean (ft)	Maximum (ft)	Minimum (ft)
January	41.04	42.74	38.78
February	41.33	43.34	38.46
March	41.67	44.03	38.22
April	41.52	44.77	37.85
May	40.65	44.11	36.97
June	39.70	43.39	36.03
July	39.14	42.46	35.96
August	39.18	42.05	36.09
September	39.55	41.79	36.83
October	39.84	42.12	37.56
November	40.18	42.32	37.96
December	40.44	42.36	37.74

AC-6 Second Table on page AC-6 should read "Table C.12 Monthly sample statistics for Well 12 (SJRWMD Well V-0101)." and not "Table C.11 Monthly sample statistics for Well 11 (SJRWMD Well SU-0013)."

AC-7 First Table on page AC-7 should read "Table C.11 Monthly sample statistics for Well 11 (SJRWMD Well SU-0013)." and not "Table C.12 Monthly sample statistics for Well 12 (SJRWMD Well V-0101)."

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	iv
LIST OF TABLES . . . . .	xi
CHAPTERS	
I. EXECUTIVE SUMMARY . . . . .	I-1
II. INTRODUCTION . . . . .	II-1
III. HYDROGEOLOGY OF THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT . . . . .	III-1
Introduction . . . . .	III-1
Geologic Setting . . . . .	III-1
The Floridan Aquifer . . . . .	III-4
Transmissivity . . . . .	III-5
Storativity . . . . .	III-6
Leakance . . . . .	III-7
Regional Flow System . . . . .	III-7
The Surficial Aquifer . . . . .	III-9
The Intermediate Aquifer . . . . .	III-9
IV. GEOSTATISTICAL ANALYSIS OF AQUIFER PARAMETERS . . . . .	IV-1
Introduction . . . . .	IV-1
Theory . . . . .	IV-2
Random Fields . . . . .	IV-2
Properties of the Variogram . . . . .	IV-8
Inter-Correlated Random Fields . . . . .	IV-11
Properties of the Cross-Variogram . . . . .	IV-12
Theoretical Variogram Models . . . . .	IV-12
Models of Anisotropy . . . . .	IV-17
Constructing a Variogram or Cross-Variogram . . . . .	IV-20
Kriging Theory . . . . .	IV-21
Cokriging . . . . .	IV-26
Log-Kriging . . . . .	IV-27
Description of Software Packages Used . . . . .	IV-29
Hydrogeologic Data Description and Analysis . . . . .	IV-31
Results and Discussion . . . . .	IV-38
Estimation of Transmissivity in the Upper Floridan Aquifer . . . . .	IV-38

Estimation of Storativity in the Upper Floridan Aquifer . . . . .	IV-55
Conclusions . . . . .	IV-64
V. TIME SERIES ANALYSIS OF PIEZOMETRIC HEAD . . . . .	V-1
Introduction . . . . .	V-1
Theory . . . . .	V-1
Stochastic Processes . . . . .	V-1
Moments of Stationary Stochastic Process . . . . .	V-4
The Spectral Density Function and Spectrum . . . . .	V-11
General Modeling Basics . . . . .	V-15
The Autoregressive Process Model . . . . .	V-16
The Moving Average Process Model . . . . .	V-23
The Mixed Autoregressive-Moving Average Process Model . . . . .	V-28
The Autoregressive Integrated Moving Average Process Model . . . . .	V-32
Seasonal Models . . . . .	V-36
Process Identification . . . . .	V-38
Model Parameter Estimation . . . . .	V-42
Model Verification . . . . .	V-44
Forecasting . . . . .	V-44
Excursion Analysis . . . . .	V-49
Recurrence Analysis . . . . .	V-51
Description of Software Packages Used . . . . .	V-52
Hydrogeologic Time Series Data Description and Analysis	V-54
Results and Discussion . . . . .	V-58
Case Study 1: Development of a Stochastic Model for a Steady-State Well . . . . .	V-58
Case Study 2: Development of a Stochastic Model for a Well with a Long-Term Trend . . . . .	V-76
Stochastic Modeling Results for all 21 Wells . . . . .	V-87
Conclusions . . . . .	V-94
VI. STOCHASTIC MODELING OF REGIONAL PIEZOMETRIC HEAD . . . . .	VI-1
Introduction . . . . .	VI-1
Theory . . . . .	VI-2
Hydrogeologic Data Description . . . . .	VI-8
Results and Discussion . . . . .	VI-9
Conclusions . . . . .	VI-52
VII. LITERATURE REVIEW . . . . .	VII-1
Introduction . . . . .	VII-1
Optimal Estimation of Spatially Variable Hydrogeologic Parameters . . . . .	VII-1
Optimal Estimation of Spatially and Temporally Variable Hydrogeologic Parameters . . . . .	VII-5
Optimal Estimation with Augmented Data . . . . .	VII-9
Kriging Coupled with Linear Regression . . . . .	VII-10
Kriging with an External Drift . . . . .	VII-11
Kriging with a Guess Field . . . . .	VII-12

Cokriging with Statistically Estimated or Physically-based Correlation Structure . . . . .	VII-13
Derivation of Physically-based Correlation Structure . . . . .	VII-15
Monte Carlo Methods . . . . .	VII-15
Spectral Techniques . . . . .	VII-17
Numerical Small-Perturbation Approximations . . . . .	VII-21
VIII. CONCLUSIONS AND RECOMMENDATIONS . . . . .	VIII-1
Conclusions . . . . .	VIII-1
Recommendations . . . . .	VIII-3
REFERENCES . . . . .	R-1
APPENDIX A: LATITUDE, LONGITUDE, TRANSMISSIVITY, STORATIVITY, AND LEAKANCE VALUES FOR WELLS PROVIDED BY THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT . . . . .	AA-1
APPENDIX B: ANNUAL MEAN, STANDARD DEVIATION, MAXIMUM AND MINIMUM PIEZOMETRIC HEAD LEVELS FOR 21 WELLS WITHIN THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT . . . . .	AB-1
APPENDIX C: MONTHLY MEAN, STANDARD DEVIATION, MAXIMUM AND MINIMUM PIEZOMETRIC HEAD LEVELS FOR 21 WELLS WITHIN THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT . . . . .	AC-1

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
3.1 The St. Johns River Water Management District and ground water basins. . . . .	III-2
4.1 Typical transition model with sill $C_0$ and range $a$ . . . . .	IV-9
4.2 Behavior near the origin of the semivariogram. (a) Parabolic behavior; (b) linear behavior; (c) nugget effect; (d) pure nugget effect (after Journel and Huijbregts 1978). . . . .	IV-10
4.3 Three general transition models (after Journel and Huijbregts 1978). . . . .	IV-15
4.4 Semivariogram and directional graph for geometric anisotropy (after Journel and Huijbregts 1978). . . . .	IV-18
4.5 Spatial distribution of transmissivity data for Upper Floridan aquifer. . . . .	IV-34
4.6 Spatial distribution of storativity data for Upper Floridan aquifer. . . . .	IV-35
4.7 Spatial distribution of leakance data for Upper Floridan aquifer. . . . .	IV-36
4.8 Histogram of transmissivity data for the Upper Floridan aquifer. . . . .	IV-39
4.9 Histogram of log-transformed Upper Floridan transmissivity data. . . . .	IV-39
4.10 Spherical model fitted to a semivariogram of districtwide log transformed transmissivity values in the Upper Floridan aquifer. . . . .	IV-41
4.11 Transmissivity estimation map for the Upper Floridan aquifer based on a districtwide semivariogram. . . . .	IV-43
4.12 Standard deviations of transmissivity estimates based on a districtwide semivariogram. . . . .	IV-44
4.13 Northern and southern subdistricts. . . . .	IV-45

<u>Figure</u>	<u>Page</u>
4.14 Spherical model fitted to a semivariogram of log transformed transmissivity values for the northern subdistrict. . . . .	IV-47
4.15 Spherical model fitted to a semivariogram of log transformed transmissivity values for the southern subdistrict. . . . .	IV-49
4.16 Transmissivity estimation composite map based on the northern and southern subdistrict semivariograms. . . .	IV-51
4.17 Standard deviations of transmissivity estimates based on the northern and southern subdistrict semivariograms. .	IV-52
4.18 Model derived transmissivities for the Upper Floridan aquifer (after Bush and Johnston 1988). . . . .	IV-54
4.19 Histogram of storativity data for the Upper Floridan aquifer. . . . .	IV-55
4.20 Histogram of log-transformed Upper Floridan storativity data. . . . .	IV-56
4.21 Spherical model fitted to a semivariogram of log transformed storativity values for the Upper Floridan aquifer. . . . .	IV-57
4.22 Storativity estimation for the Upper Floridan aquifer. . . . .	IV-59
4.23 Standard deviations of storativity estimates. . . . .	IV-60
4.24 Spherical model fitted to a cross-semivariogram of log transformed storativity and transmissivity values for the Upper Floridan aquifer. . . . .	IV-61
4.25 Storativity estimation for the Upper Floridan aquifer based on storativity-transmissivity co-kriging. . . . .	IV-62
4.26 Standard deviations of storativity estimates based on storativity-transmissivity co-kriging. . . . .	IV-63
5.1 An example of a discrete time series. . . . .	V-2
5.2 An autocorrelation matrix and the resulting autocorrelation function (after Box and Jenkins 1976). . .	V-8
5.3 Realizations from AR(1) processes and their corresponding theoretical autocorrelation functions and spectrums (after Box and Jenkins 1976). . . . .	V-20

<u>Figure</u>	<u>Page</u>
5.4 Typical autocorrelation functions $\rho_k$ and partial autocorrelation functions $\phi_{kk}$ for various AR(2) processes (after Box and Jenkins 1976). . . . .	V-22
5.5 Estimated partial autocorrelation function together with two standard error limits calculated for an AR(1) model (after Box and Jenkins 1976). . . . .	V-24
5.6 Typical autocorrelation functions $\rho_k$ and partial autocorrelation functions $\phi_{kk}$ for various MA(2) processes (after Box and Jenkins 1976). . . . .	V-29
5.7 Time series plots. (a) Original series with nonstationary mean; (b) First-order differenced series with nonzero mean; (c) Second-order differenced series with zero mean.	V-35
5.8 Iterative approach to model building (after Bras and Rodriguez-Iturbe 1985). . . . .	V-39
5.9 Definition sketch for continuous time series (after Nordin and Rosbjerg 1970). . . . .	V-50
5.10 Spatial distribution of 21 wells. . . . .	V-56
5.11 Piezometric head levels for Well 17. . . . .	V-60
5.12 Average long-term and monthly head levels for Well 17. . .	V-61
5.13 Autocorrelation function and partial autocorrelation function for Well 17 original time series. . . . .	V-62
5.14 Spectrum of Well 17 original time series. . . . .	V-63
5.15 Seasonally differenced piezometric head levels for Well 17. . . . .	V-64
5.16 Autocorrelation function and partial autocorrelation function for Well 17 seasonally differenced time series. .	V-66
5.17 Spectrum of Well 17 seasonally differenced time series. .	V-67
5.18 Residuals of an ARIMA(2,0,0)x(0,1,1) <sub>12</sub> adjusted to the original series of Well 17. . . . .	V-69
5.19 Autocorrelation function and spectrum of the residuals of an ARIMA(2,0,0)x(0,1,1) <sub>12</sub> adjusted to the original series of Well 17. . . . .	V-70
5.20 One month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 17. . . . .	V-72

<u>Figure</u>	<u>Page</u>
5.21 Three month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 17. . . . .	V-73
5.22 Six month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 17. . . . .	V-74
5.23 Piezometric head levels for Well 13. . . . .	V-77
5.24 Average yearly and monthly head levels for Well 13. . . . .	V-78
5.25 Autocorrelation function and partial autocorrelation function for Well 13 original time series. . . . .	V-79
5.26 Spectrum of Well 13 original time series. . . . .	V-80
5.27 Seasonally differenced piezometric head levels for Well 13. . . . .	V-81
5.28 Autocorrelation function and partial autocorrelation function for Well 13 seasonally differenced time series. . . . .	V-82
5.29 Spectrum of Well 13 seasonally differenced time series. . . . .	V-83
5.30 Residuals of an ARIMA(2,0,0)x(0,1,1) <sub>12</sub> adjusted to the original series of Well 13. . . . .	V-85
5.31 Autocorrelation function and spectrum of the residuals of an ARIMA(2,0,0)x(0,1,1) <sub>12</sub> adjusted to the original series of Well 13. . . . .	V-86
5.32 One month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 13. . . . .	V-88
5.33 Three month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 13. . . . .	V-89
5.34 Six month lead forecasts of the ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model for Well 13. . . . .	V-90
6.1 Spatial distribution of 636 wells. . . . .	VI-10
6.2 Observed piezometric head map for September 1988. . . . .	VI-11
6.3 Observed piezometric head map for May 1989. . . . .	VI-12
6.4 Semivariogram of observed piezometric head for September 1988. . . . .	VI-13
6.5 Semivariogram of observed piezometric head for May 1989. . . . .	VI-14
6.6 Differenced observed piezometric head map for September 1988. . . . .	VI-15

<u>Figure</u>	<u>Page</u>
6.7	Differenced observed piezometric head map for May 1989. VI-16
6.8	Experimental cdf for differenced observed piezometric head for September 1988. . . . . VI-18
6.9	Experimental cdf for differenced observed piezometric head for May 1989. . . . . VI-18
6.10	Linear model fitted to a semivariogram of differenced observed piezometric head levels for September 1989. . . VI-21
6.11	Linear model fitted to a semivariogram of differenced observed piezometric head levels for May 1989. . . . . VI-21
6.12	Linear model fitted to a combined semivariogram of differenced observed piezometric head levels. . . . . VI-22
6.13	Piezometric head estimates for September 1988 created by kriging over 636 well network from actual observations over 21 well network. . . . . VI-24
6.14	Piezometric head estimates for May 1989 created by kriging over 636 well network from actual observations over 21 well network. . . . . VI-25
6.15	Standard deviations of piezometric head estimates for September 1988 created by kriging over 636 well network from actual observations over 21 well network. . . . . VI-26
6.16	Standard deviations of piezometric head estimates for May 1989 created by kriging over 636 well network from actual observations over 21 well network. . . . . VI-27
6.17	Actual prediction errors for September 1988 calculated by subtracting the kriged head estimate from the observed head measured at each of the 636 wells. . . . . VI-28
6.18	Actual prediction errors for May 1989 calculated by subtracting the kriged head estimate from the observed head measured at each of the 636 wells. . . . . VI-29
6.19	Experimental cdf for normalized actual piezometric head kriging errors for September 1988. . . . . VI-31
6.20	Experimental cdf for normalized actual piezometric head kriging errors for May 1989. . . . . VI-31
6.21	Six month lead regional forecast of piezometric head for September 1988. . . . . VI-33

<u>Figure</u>	<u>Page</u>
6.22 Actual forecast error for September 1988 calculated by subtracting the 6-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-34
6.23 Experimental cdf for the actual 6-month lead forecast errors for September 1988. . . . .	VI-35
6.24 Three month lead regional forecast of piezometric head for September 1988. . . . .	VI-37
6.25 Actual forecast error for September 1988 calculated by subtracting the 3-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-38
6.26 Experimental cdf for the actual 3-month lead forecast errors for September 1988. . . . .	VI-39
6.27 One month lead regional forecast of piezometric head for September 1988. . . . .	VI-40
6.28 Actual forecast error for September 1988 calculated by subtracting the 1-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-41
6.29 Experimental cdf for the actual 1-month lead forecast errors for September 1988. . . . .	VI-42
6.30 Six month lead regional forecast of piezometric head for May 1989. . . . .	VI-43
6.31 Actual forecast error for May 1989 calculated by subtracting the 6-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-44
6.32 Experimental cdf for the actual 6-month lead forecast errors for May 1989. . . . .	VI-45
6.33 Three month lead regional forecast of piezometric head for May 1989. . . . .	VI-46
6.34 Actual forecast error for May 1989 calculated by subtracting the 3-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-47
6.35 Experimental cdf for the actual 3-month lead forecast errors for May 1989. . . . .	VI-48
6.36 One month lead regional forecast of piezometric head for May 1989. . . . .	VI-49

<u>Figure</u>		<u>Page</u>
6.37	Actual forecast error for May 1989 calculated by subtracting the 1-month lead forecast from the actual observed head over each of the 636 wells. . . . .	VI-50
6.38	Experimental cdf for the actual 1-month lead forecast errors for May 1989. . . . .	VI-51

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
4.1 Useful hydrogeologic data provided per aquifer by the St. Johns River Water Management District. . . . .	IV-32
4.2 Sample statistics for Upper Floridan transmissivity data for St. Johns River Water Management District. . . . .	IV-40
4.3 Sample statistics for Upper Floridan transmissivity data for northern subdistrict. . . . .	IV-46
4.4 Sample statistics for Upper Floridan transmissivity data for southern subdistrict. . . . .	IV-46
4.5 Sample statistics for Upper Floridan storativity data. . . . .	IV-56
5.1 Summary of properties of autoregressive, moving average, and mixed ARMA processes (after Box and Jenkins 1976). . . . .	V-33
5.2 Identification, location, years of data, and number of monthly observations $n$ for 21 wells provided by the St. Johns River Water Management District. . . . .	V-57
5.3 Minimum, maximum, mean, and standard deviation values for Well 17. . . . .	V-61
5.4 Minimum, maximum, mean, and standard deviation values for Well 13. . . . .	V-78
5.5 ARIMA(2,0,0)x(0,1,1) <sub>12</sub> model parameters fitted to 21 wells provided by the St. Johns River Water Management District. . . . .	V-91
5.6 Actual versus predicted model error $a_t$ statistics. . . . .	V-93
5.7 Excursion analysis results on the normalized series for the 8 nontrending wells . . . . .	V-95
5.8 Recursion analysis results for the 8 nontrending wells. . . . .	V-96
6.1 Differenced head field sample statistics . . . . .	VI-19
6.2 Experimental variogram model parameters. . . . .	VI-20
6.3 Normalized kriging error statistics. . . . .	VI-32

<u>Table</u>	<u>Page</u>
A.1 Values for shallow aquifer. . . . .	AA-1
A.2 Values for intermediate aquifer. . . . .	AA-2
A.3 Values for Upper Floridan aquifer. . . . .	AA-2
A.3--continued. . . . .	AA-3
A.3--continued. . . . .	AA-4
A.3--continued. . . . .	AA-5
C.1 Monthly sample statistics for Well 1 (SJRWMD Well N-0003). . . . .	AC-1
C.2 Monthly sample statistics for Well 2 (SJRWMD Well D-0348). . . . .	AC-2
C.3 Monthly sample statistics for Well 3 (SJRWMD Well D-0160). . . . .	AC-2
C.4 Monthly sample statistics for Well 4 (SJRWMD Well U-0001). . . . .	AC-3
C.5 Monthly sample statistics for Well 5 (SJRWMD Well C-0120). . . . .	AC-3
C.6 Monthly sample statistics for Well 6 (SJRWMD Well P-0172). . . . .	AC-4
C.7 Monthly sample statistics for Well 7 (SJRWMD Well SJ-0104). . . . .	AC-4
C.8 Monthly sample statistics for Well 8 (SJRWMD Well F-0087). . . . .	AC-5
C.9 Monthly sample statistics for Well 9 (SJRWMD Well M-0048). . . . .	AC-5
C.10 Monthly sample statistics for Well 10 (SJRWMD Well M-0013). . . . .	AC-6
C.11 Monthly sample statistics for Well 11 (SJRWMD Well SU-0013). . . . .	AC-6
C.12 Monthly sample statistics for Well 12 (SJRWMD Well V-0101). . . . .	AC-7
C.13 Monthly sample statistics for Well 13 (SJRWMD Well S-0125). . . . .	AC-7

<u>Table</u>	<u>Page</u>
C.14 Monthly sample statistics for Well 14 (SJRWMD Well L-0062) . . . . .	AC-8
C.15 Monthly sample statistics for Well 15 (SJRWMD Well OR-0047) . . . . .	AC-8
C.16 Monthly sample statistics for Well 16 (SJRWMD Well OR-0007) . . . . .	AC-9
C.17 Monthly sample statistics for Well 17 (SJRWMD Well SU-0002) . . . . .	AC-9
C.18 Monthly sample statistics for Well 18 (SJRWMD Well OR-0064) . . . . .	AC-10
C.19 Monthly sample statistics for Well 19 (SJRWMD Well BR-0202) . . . . .	AC-10
C.20 Monthly sample statistics for Well 20 (SJRWMD Well OS-0001) . . . . .	AC-11
C.21 Monthly sample statistics for Well 21 (SJRWMD Well PO-0006) . . . . .	AC-11

## CHAPTER I EXECUTIVE SUMMARY

The management of water resources to meet water supply demands while maintaining minimum water levels and quality, is of paramount importance. Florida is almost totally dependent on groundwater for commercial, industrial, agricultural, and residential needs. Over the years, Florida's water resource management concerns have been amplified due to periods of drought and to ever increasing demands for more water.

In the past two decades groundwater models have become common tools to aid decision makers with problems of resource management. Traditional deterministic groundwater models require information throughout the hydrologic system that is never known but must be estimated. The error inherent in this information may undermine the quality of the model predictions needed by decision makers. Furthermore, deterministic models cannot provide information on model prediction uncertainty.

Recent developments in hydrogeologic modeling have emphasized stochastic modeling techniques. These models, founded on the principles of probability theory, provide a systematic method for estimating and predicting water levels and recharge rates which vary erratically in space and in time. More importantly, stochastic models calculate optimal predictions based on finite and limited measurements, and quantify the accuracy and quality of these predictions.

As part of The St. Johns River Water Management District's responsibility to manage water resources within the District, this study was initiated to develop the capability of forecasting critical low groundwater levels in space and time. Research efforts were focused to:

- 1) develop empirical stochastic models capable of forecasting groundwater levels at point locations in time and throughout the District in space and time;
- 2) geostatistically characterize the spatial variability of groundwater levels, transmissivities, and storativities in the Floridan aquifer; and
- 3) identify and evaluate alternative physically-based stochastic regional groundwater modeling techniques which are consistent with the quality and quantity of existing hydrogeologic information available throughout the District, and which may further aid the District in fulfilling its goals.

In Chapters V and VI, methods for predicting groundwater levels in space and time, and methods for obtaining the recurrence interval and the duration of critical events are identified and presented. Results from critical event investigations are given for selected wells distributed throughout the District. Geostatistical information on regional Floridan aquifer parameters is presented in Chapter IV. Methods of using this information in physically-based stochastic models are examined in Chapter VII. Beyond the methods and results already mentioned, this project produced the following deliverables:

- 1) 21 stochastic time-series models which can forecast groundwater levels (with forecast confidence intervals) at

21 sites distributed throughout the District. For these models, the prediction uncertainty decreases significantly when forecast lead times are limited to less than seven months.

- 2) A geostatistical model which extrapolates the 21 site specific groundwater level forecasts over space in order to provide optimal forecasts (with confidence intervals) of groundwater levels throughout the District. This model should be useful for predicting regional water shortages and it should provide a basis for drought management decisions.
- 3) District-wide maps of optimal estimates of the spatial distribution of transmissivity and storativity in the Floridan aquifer (with confidence intervals). These maps were developed from geostatistical analysis of available transmissivity and storativity measurements, and should provide valuable information for both defining physically-based deterministic and/or stochastic groundwater model input parameters, and for designing field studies to gather additional data.

## CHAPTER II INTRODUCTION

In the field of groundwater management, the design and operation of cost-effective well fields to meet water supply demands while maintaining minimum groundwater levels and quality is of paramount importance. In the state of Florida, such concerns are magnified by the almost total dependence on groundwater for commercial, industrial, agricultural, and residential needs. Further, Florida's tremendous population growth combined with recent drought periods have strained the existing water supply systems.

The majority of groundwater management models are physically-based deterministic models which depend on spatially variable hydrogeologic parameters such as recharge, transmissivity, and storativity. However, these variables, as well as system demands, initial conditions, and boundary conditions, are never precisely known and must be estimated, usually in an ad-hoc manner, from incomplete and inexact data. The error inherent in these estimations may undermine management goals.

Stochastic modeling techniques, founded on the principles of probability theory, provide a systematic method for estimating and predicting the behavior of unknown spatially variable hydrogeologic parameters. Recognizing the variability of hydrogeologic variables in space and time, stochastic techniques, based on well-defined assumptions, calculate optimal predictions based on finite and limited measurements, and quantify the accuracy and quality of these

predictions.

As part of the St. Johns River Water Management District's (SJRWMD) responsibility to manage the water resources within the District, this research project was designed to (1) develop empirical stochastic models capable of forecasting groundwater levels at point locations in time and throughout the District in space and time, (2) characterize the spatial variability and correlation structure of available piezometric head, transmissivity, and storativity measurements in the Floridan aquifer using geostatistical methods, and (3) identify and evaluate alternative physically-based stochastic regional groundwater modeling techniques which are consistent with the quality and quantity of existing hydrogeologic information available throughout the District, and which may further aid the District in fulfilling its goals.

Characterization of temporal potentiometric head fluctuations at 21 wells located throughout the District was conducted using the techniques of time series analysis (Box and Jenkins 1976). Wells were selected, from those provided by the District, with lengthy and continuous records of measured piezometric head at frequent time intervals. The data was visually summarized using time series plots of the raw data as well as plots of long-term monthly and annual means and standard deviations. After analyzing these plots for long-term and seasonal trends, spectral and autocorrelations analyses were performed on the detrended series.

These efforts revealed an appropriate form of the time series model for each well based on the temporal correlation structure of the potentiometric head values. Parameters for these models were estimated

using estimation techniques outlined by Box and Jenkins (1976). The models were verified by examining the statistics of the model residual series and the statistics of the model forecast errors.

The stochastic time series models described above provide short-term, temporal piezometric head forecasts for the particular wells being modeled. To extrapolate these site-specific forecasts over space, and thus obtain a regional forecast, a geostatistical analysis of the spatial variability and correlation structure of piezometric head was conducted. Geostatistical analysis of the raw piezometric head data detected a distinct nonstationary spatial trend due to the recharge and discharge patterns within the District. The piezometric head values were therefore detrended using a temporal differencing procedure and a steady-state, spatially stationary, geostatistical structure emerged. The 21 site-specific time series models were then coupled with the geostatistical model to produce optimal forecasts of May and September piezometric head levels throughout the District.

An excursion analysis was performed on each of the 21 modeled wells which did not exhibit a long-term declining trend in piezometric head level, to estimate probabilities of recurrence intervals of pre-selected "critical" groundwater levels. Crossing-theory (Nordin and Rosbjerg 1970; Bras and Rodriguez-Iturbe 1985) was used to estimate the expected number of excursions below these critical levels in a given time period and the expected duration of these excursions. Recursion analysis of these selected wells was performed to determine expected low piezometric head values over given time intervals.

The spatial variability and correlation structure of available transmissivity and storativity measurements were analyzed using the

method of geostatistics. The spatial distribution of point measurements of these properties cannot be predicted exactly as they vary erratically in space due to the natural geologic variability and measurement error. Geostatistical techniques were thus used to obtain optimal (i.e., minimum variance/unbiased) estimates of the spatial distribution of these properties using knowledge of point measurements of the process, and its mean, variance and correlation scale. Maps of the estimated spatial distribution of these hydrologic properties were developed, along with maps of the standard deviation of these estimates. These maps should provide valuable information for both defining physically-based deterministic and/or stochastic groundwater model input parameters and for designing field studies to gather additional data.

A review was conducted of recently developed methods for using historic piezometric head data, point measurements of hydrogeologic parameters, and geostatistical information to develop physically-based stochastic regional groundwater models. Physically-based stochastic modeling techniques are often preferable to empirical stochastic modeling techniques (such as time-series methods or traditional geostatistical analysis) since their parameters are based on measurable aquifer characteristics (rather than fit from historical data) and their predictions incorporate the underlying physics of groundwater flow. However, physically-based stochastic models are typically more computationally intensive and require significantly more prior information than empirical models. Results of this review should assist the District in selecting modeling techniques for future studies which are most suited to its needs.

CHAPTER III  
HYDROGEOLOGY OF THE ST JOHNS RIVER  
WATER MANAGEMENT DISTRICT

Introduction

The St. Johns River Water Management District (SJRWMD), created by the Florida Legislature in 1972 as one of five water management districts in the state, includes all or part of nineteen counties in northeast Florida (McKenzie-Arenberg and Toth 1990). The District is further divided into five ground water basins. A map of the District and its five ground water basins is shown in Figure 3.1. Hydrogeologic information on each of the five ground water basins is available from SJRWMD technical publications SJ 89-4, SJ 90-8, SJ 90-10, and SJ 90-11 (McKenzie-Arenberg 1989; Huff and McKenzie-Arenberg 1990; McKenzie-Arenberg and Toth 1990; McKenzie-Arenberg and Szell 1990).

Ground water resources in the District are provided by three general aquifer systems defined by the geologic setting. These are the surficial (unconfined), the intermediate, and the Floridan aquifer systems. The Floridan aquifer provides the primary source of ground water for most of the District.

Geologic Setting

The District is located on the Coastal Plain province of the Southeastern United States. The Coastal Plain province is underlain by a sequence of unconsolidated to semiconsolidated sedimentary rocks (Bush

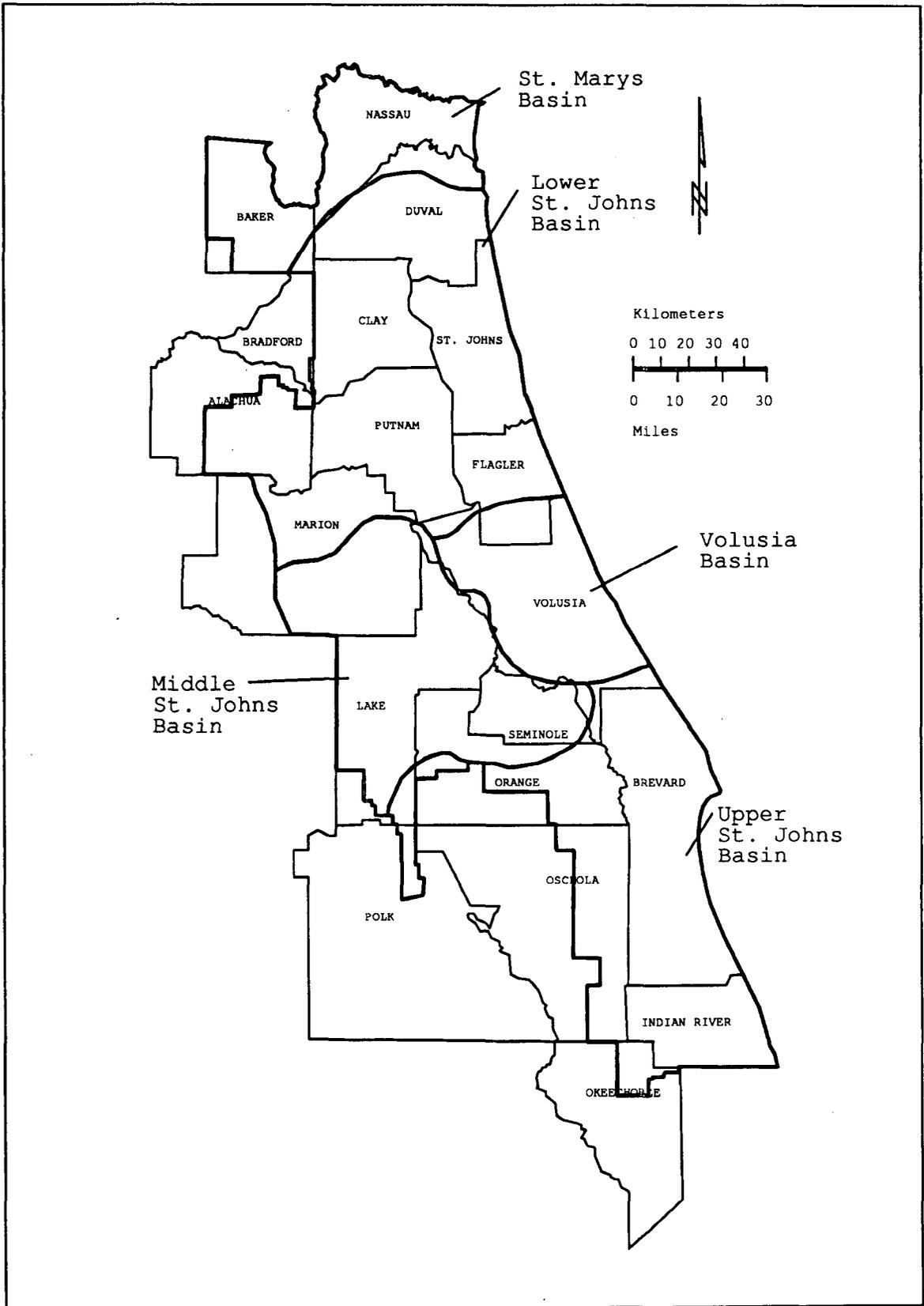


Figure 3.1 The St. Johns River Water Management District and ground water basins.

and Johnston 1988). These sediments were laid down upon an eroded surface of igneous intrusive rocks, metamorphic rocks, and consolidated sedimentary rocks from the Paleozoic to Early Jurassic age. These sediments begin as a razoredge where they crop out against the older metamorphic and igneous rocks of the Appalachian and Piedmont provinces. From there they thicken seaward to a thickness of approximately 20,000 to 25,000 feet in southern Alabama and south Florida. The underlying rocks generally dip gently toward the Atlantic Ocean and the Gulf of Mexico.

The poorly consolidated sedimentary deposits of the Coastal Plain are easily eroded. When the sediments consist of carbonate rocks (principally limestone with some dolomite) at or near the surface, karst topography is developed by the downward-percolating water. A series of Pleistocene age sandy marine terraces has been developed in much of the area. The sediments consist of two main facies: (1) generally clastic rocks containing minor amounts of limestone, and (2) thick continuous units of shallow-water platform carbonate rocks. In general, limestone facies of successively younger units (through the end of the Oligocene age) extend progressively farther and farther updip, overlapping the older clastic rocks. These carbonate rocks are then overlain, except where removed by erosion, by clastic facies of Miocene and younger rocks.

Three categories of sedimentary rocks comprise most of the rocks underlying the District: Quaternary sedimentary rocks, Pliocene and Miocene sedimentary rocks, and Oligocene and Eocene sedimentary rocks. The Floridan aquifer is found in the Oligocene and Eocene layer and in

the lower Miocene rocks. The intermediate and surficial aquifers are found in the upper Pliocene rocks and in the Quaternary layer.

### The Floridan Aquifer

The Floridan aquifer system is a sequence of hydraulically connected carbonate rocks of generally high permeability that are of Tertiary age. The aquifer system generally consists of an upper and lower aquifer separated by a less permeable confining unit. Where that confining unit is very thin or nonexistent, as in north Florida and southwest Georgia, the Floridan is effectively one continuous aquifer. The two aquifers, the Upper Floridan and the Lower Floridan, are defined on the basis of permeability and not on time-stratigraphic or rock-stratigraphic units.

The Floridan aquifer system is overlain by a clastic rock layer of low-permeability. The degree of confinement of the aquifer system is determined by the thickness and integrity of this layer. Areas of recharge, discharge, and groundwater flow in the Floridan are influenced by the presence or absence of this confining layer.

Within SJRWMD, the base of the Floridan aquifer consists of a bed of hard brown dolostone approximately 100 ft. thick (Johnson, Frazee, and Fenzel 1982). This is the Lake City Limestone formation. This formation is then overlain by the Avon Park Limestone, Ocala Limestone, and occasionally the Suwannee Limestone formations. In some areas, the Lower portion of the Hawthorn Formation exists.

## Transmissivity

Transmissivity  $T$  defines the rate at which a unit volume of water will flow across a unit width of aquifer under unit hydraulic gradient and is expressed in dimensions of volume per length-time. The transmissivity of the Floridan aquifer is derived from openings ranging from small fossil hashes to the large cavernous openings common to karst topography. Thus, transmissivity values within the Floridan aquifer range from low to very high. Transmissivities are highest (greater than 1,000,000 square feet per day) in major spring areas and unconfined karst areas of central and northern Florida, where conduit flow is most likely to occur. Transmissivities of less than 50,000 square feet per day occur in panhandle Florida and southernmost Florida where the Upper Floridan is confined by thick clay layers. The aquifer in this area contains large amounts of micritic limestone that has very low permeability (Miller 1986). The hydraulic properties of the Lower Floridan are not as well known, but areas of high transmissivity have been attributed to paleokarst development.

Transmissivity values can be estimated from aquifer tests, specific-capacity data, or model simulation. Most transmissivity values obtained from specific-capacity and aquifer test data are much lower than model derived values (Bush and Johnston 1988). This is likely due to a combination of factors including (1) the effects of well losses during pump tests, (2) the site-specific nature of transmissivity values derived from pump tests, and (3) the difficulties associated with using models to calibrate unique transmissivity values by matching observed hydraulic head distributions. The most common symptom of a non-unique solution to this calibration problem is for the transmissivities to

exhibit spatial oscillations which are higher than those anticipated from a geological standpoint (Neuman and Yakowitz 1979).

Calculated transmissivities, based on pump tests, within the District range between 1,600 ft<sup>2</sup>/day in northeast Volusia County to 758,000 ft<sup>2</sup>/day in Orange County. Model simulated transmissivity values within the District range from less than 10,000 ft<sup>2</sup>/day to more than 1,000,000 ft<sup>2</sup>/day (Bush and Johnston 1988).

### Storativity

Storativity  $S$  is defined as the volume of water that an aquifer releases from storage per unit surface area of aquifer per unit decline in the component of hydraulic head normal to that surface (Freeze and Cherry 1979). Storativity in confined aquifers is directly proportional to aquifer thickness, and is defined as the integral of the aquifer specific storage  $S_s$  over depth

$$S = \int_0^b S_s(z) dz \quad . \quad (3.1)$$

In the Floridan aquifer system, however, reported storativity coefficients appear to have no discernable relation to the thickness of the aquifer on a regional basis. The storativity values calculated from aquifer test data range from a low of 0.00001 to a high of 0.02. Most values fall in the 0.001-0.0001 range. The higher values, 0.01-0.001, reflect the unconfined and semiconfined nature of parts of the system, particularly where the aquifer lies very close to land surface (Bush and Johnston 1988).

## Leakance

Leakance  $L$  is defined as the hydraulic conductivity per unit depth of the overlying or underlying aquifer confining bed, and is usually given in units of inverse time. The leakance coefficient of the upper confining unit of the Floridan aquifer ranges from 0.02 per day to 0.0001 per day, based on field-determined leaky aquifer tests. Digitally simulated estimates of leakance calibrated from measured flow between the surficial aquifer and the Upper Floridan aquifer range from  $2.28 \times 10^{-4}$  per day to  $2.28 \times 10^{-6}$  per day. Within the District, the leakance has been calculated to range from 0.04 per day to  $1.3 \times 10^{-7}$  per day. Generally, leakance values calculated from aquifer test data are two orders of magnitude larger than those obtained from simulation. For most test sites, leakance values calculated from aquifer tests are too large to realistically characterize the true rate of water migration from the surficial aquifer into the Upper Floridan aquifer. These values, it has been found, reflect not only the downward leakance from the surficial aquifer into the Upper Floridan aquifer, but also leakance into the aquifer from other sources below and adjacent to it. Furthermore, wells in the Floridan aquifer are usually partially penetrating, and often intersect local units of low-permeability. Thus, those leakance values calculated from aquifer tests characterize leakance from all sources and do not accurately represent downward leakance.

## Regional Flow System

The dominant feature of the Floridan flow system is the springs which occur in unconfined or semiconfined parts of the Upper Floridan in

Florida. Spring flow to surface water bodies comprised 88 percent of the total aquifer discharge prior to groundwater development. The remaining 12 percent occurred as diffuse upward leakage. Total redevelopment discharge stood at approximately 21,500 cubic feet per second (Bush and Johnston 1988).

Recharge to the Floridan aquifer system occurs mostly where the Upper Floridan is unconfined or semiconfined, and averages 10-20 inches per year. The high discharge areas are along the coasts, rivers, and streams. High recharge areas may also appear in the same proximity as the high discharge areas, indicating a well-developed shallow flow system within the unconfined to semiconfined areas of the aquifer system. Within the District, recharge areas are located toward the center of the state along piezometric highs relative to higher topographic elevations and coincident with prominent karst or paleobeach ridges (Johnson, Frazee, and Fenzel 1982). Discharge areas are located principally in low-lying areas, along the Atlantic coast, or along streams and rivers.

In those areas of the aquifer system that are tightly confined and deeply buried, groundwater flow is very sluggish, and discharge to springs and surfacewater bodies is practically nonexistent. The primary means of discharge in these areas is by diffuse upward leakage through the thick confining layers.

Although the regional flow system has not been appreciably altered by groundwater development, increased pumping has resulted in long-term regional water-level decline in some areas. This decline has surpassed 10 feet in coastal South Carolina and Georgia, northeast Florida, west-

central Florida, and panhandle Florida. Local saltwater intrusion as a result of pumping has occurred in some coastal areas.

The increase in pumping in the Upper Floridan is offset primarily by reduction of natural discharge and increased recharge rather than a decrease in aquifer storage. Discharge to springs, streams, and lakes has been reduced to 95 percent of predevelopment flow and upward leakage has been reduced to 70 percent of predevelopment conditions. Natural recharge has been increased by about 12 percent (Bush and Johnston 1988).

#### The Surficial Aquifer

The surficial aquifer consists of late Miocene, Pleistocene, and Recent Age deposits (Johnson, Frazee, and Fenzel 1982). It is an unconfined system ranging from 20 to 200 ft in depth. It is directly recharged by rainfall, with the top of the aquifer defined by the water table under atmospheric pressure. The water table marks the line below which all pore spaces are filled with water (Mckenzie-Arenberg and Szell 1990).

Transmissivity values in the shallow aquifer within the District range from 240 ft<sup>2</sup>/day to 40,200 ft<sup>2</sup>/day. Specific storage values range from 0.28 (unconfined) to 0.0001 (semi-confined). Ground water flow in the surficial aquifer usually follows the land surface topography.

#### The Intermediate Aquifer

The intermediate aquifer lies between the surficial and the Floridan aquifer, and is usually found within the confining unit of the Floridan. Composed of clays, and thin water bearing-zones of sand,

shell, and limestone, the intermediate aquifer is a low-yielding aquifer throughout most of the district. Where the Floridan aquifer contains water of marginal quality, however, this aquifer can be an important source of ground water.

## CHAPTER IV GEOSTATISTICAL ANALYSIS OF AQUIFER PARAMETERS

### Introduction

Geostatistical techniques rely upon the premise that information about a phenomenon in space can be deduced by sampling and examining a small representative set of data from an infinitely larger set of potential observations on the phenomenon (Journel and Huijbregts 1978). In hydrogeology, available data on hydrogeologic properties is extremely limited in comparison to the vastness and complex variability of these properties in space. Exact determination of the spatial distribution of hydrogeologic properties could only be accomplished with an infinite sampling set over the space in question. This being obviously impossible, the hydrogeologist must utilize geostatistical techniques to obtain reasonable estimates of the spatial distribution of the phenomenon from a set of relatively few point measurements.

Since geostatistics involves making inferences with limited available data, it relies heavily on probability theory. For most phenomena, there exist many possible outcomes or realizations. Each of these has an associated probability of occurrence. Knowledge of the range of possible realizations and their probabilities can lead to confident estimations of future behavior or past states of the phenomenon.

## Theory

### Random Fields

Many hydrogeologic parameters such as piezometric head, transmissivity, storativity, and leakance exhibit significant spatial variability over relatively small distances. Examination of these data indicates, however, that these parameters are not completely disordered in space. There is some spatial structure which must be taken into account. For example, transmissivity values measured over a regional aquifer will vary erratically around the regional mean. However, it is likely that clusters of high transmissivity measurements and clusters of low transmissivity measurements will occur in different regions of the aquifer.

The duality of the random and structured aspects of hydrogeologic parameters can be conveniently described using random field theory. A random field  $Z_x$  is defined as a collection or ensemble of random variables that vary over space, any one realization of which is observed at a given time and location. The random field  $Z_x$  can be expressed as

$$Z_x = \bar{Z}_x + \xi_x, \quad (4.1)$$

where  $\bar{Z}_x$  is the population mean, or expected value, of the random field at location  $x$ , and  $\xi_x$  is the zero-mean random fluctuation or residual at  $x$ . The mean  $\bar{Z}_x$  characterizes the large-scale variations of the natural phenomenon while the residual  $\xi_x$  characterizes the usually small-scale random variations around the mean value. In general, random variables  $Z_1$  and  $Z_2$  occurring at locations  $x_1$  and  $x_2$  will be statistically

correlated with one another. This correlation can be dependent on both the magnitude and direction of the vector  $h$  separating the two points, as well as the absolute locations of the points within the random field.

A complete probabilistic description of the random field  $Z_x$  requires knowledge of the joint probability density function between the random variables at all points in space. For practical applications, however, where only one realization of the random field is generally available, this infinite order probability density function will never be known. Therefore we must settle for more easily obtainable descriptions of the random field such as its mean, variance, and covariance or variogram.

The population mean, or expected value, of a random field is defined as the sum of all the values its random variables may take, each weighted by the probability with which that value is taken. For a continuously distributed random field with known probability density function  $f_z(Z_x)$ , the population mean is defined by the following equation:

$$E\{Z_x\} = \bar{Z}_x = \int_{-\infty}^{\infty} Z_x f_z(Z_x) dZ_x, \quad (4.2)$$

where  $E\{ \}$  is the expectation operator. A random field is said to exhibit a stationary mean if the mean does not vary with location, i.e.,  $E\{Z_x\} = \bar{Z}_x = \bar{Z}$ . An estimate of the population mean of a stationary random field may be calculated from a finite sampling of  $n$  random variables  $Z_i$  ( $i = 1, 2, \dots, n$ ) taken from one realization of the random field using the following equation:

$$\hat{\bar{Z}} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad (4.3)$$

where  $\hat{\bar{Z}}$  is the estimated population mean of the stationary random field  $Z_x$ .

The population variance  $\sigma_z^2(x)$  of a random field provides a measure of the expected deviation of each of its random variables from its respective population mean, and is defined as

$$E\{(Z_x - \bar{Z}_x)^2\} = \sigma_z^2(x) = \int_{-\infty}^{\infty} (Z_x - \bar{Z}_x)^2 f_z(Z_x) dZ_x. \quad (4.4)$$

The population variance is stationary if its value does not vary with location, i.e.,  $E\{(Z_x - \bar{Z}_x)^2\} = \sigma_z^2(x) = \sigma_z^2$ . An estimate of the population variance of a stationary random field may be calculated from a finite sampling of  $n$  random variables taken from one realization of the random field using the following equation:

$$\hat{\sigma}_z^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \hat{\bar{Z}})^2. \quad (4.5)$$

The standard deviation  $\sigma_z$  of a random field is simply the square root of its variance  $\sigma_z^2$ .

The population covariance function of a random field describes the expected similarity, or spatial correlation, of two values of the random field  $Z_1$  and  $Z_2$  observed at locations  $x_1$  and  $x_2$ . A high covariance between the pair of random variables  $Z_1$  and  $Z_2$  indicates that measurement of one of these random variables provides a significant

amount of information about the other. Theoretical determination of the covariance function requires knowledge of the joint probability density function  $f_{1,2}(Z_1, Z_2)$ , and is defined as

$$C_{1,2}(x_1, x_2) = E\{(Z_1 - \bar{Z}_1)(Z_2 - \bar{Z}_2)\} \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z_1 - \bar{Z}_1)(Z_2 - \bar{Z}_2) f_{1,2}(Z_1, Z_2) dZ_1 dZ_2, \quad (4.6)$$

where  $C_{1,2}(x_1, x_2)$  is the covariance of the two values of the random field  $Z_1$  and  $Z_2$  at locations  $x_1$  and  $x_2$ .

In general, the covariance function depends simultaneously on the two spatial locations  $x_1$  and  $x_2$ . However, if the population covariance is to be inferred from one realization of the random field, it must be assumed to be stationary. The covariance function is defined to be stationary if it depends only on the vector modulus  $|h| = |x_1 - x_2|$  separating two points in the random field, and not the actual locations of these points, i.e.,

$$C_h = C_{1,2}(|x_1 - x_2|) = C_{1,2}(x_1, x_2), \quad (4.7)$$

where  $C_h$  is the stationary covariance function expressed as a function of the separation distance  $|h|$ . The population covariance function for the stationary case may be estimated using the following equation:

$$\hat{C}_h = \frac{1}{N(h)-1} \sum_{i=1}^{N(h)} [(Z_{xi+h} - \hat{\bar{Z}})(Z_{xi} - \hat{\bar{Z}})], \quad (4.8)$$

where  $\hat{C}_h$  is the estimate of the population covariance function and  $N(h)$  is the total number of pairs of points separated by distance  $|h|$  that have been sampled from a particular realization.

An alternative summary of the spatial correlation of a random field is the population variogram. The population variogram  $2\gamma_{1,2}(x_1, x_2)$  describes the expected variability of two values of the random field  $Z_1$  and  $Z_2$  observed at locations  $x_1$  and  $x_2$ , and is defined as

$$2\gamma_{1,2}(x_1, x_2) = E\{(Z_1 - Z_2)^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z_1 - Z_2)^2 f_{1,2}(Z_1, Z_2) dZ_1 dZ_2 . \quad (4.9)$$

In general, the variogram, like the covariance function, simultaneously depends on the two spatial locations  $x_1$  and  $x_2$ . For the stationary case, however, the variogram depends only on the vector modulus  $|h|$  separating two points in the random field, and not the actual point locations:

$$2\gamma_h = 2\gamma_{1,2}(|x_1 - x_2|) = 2\gamma_{1,2}(x_1, x_2) , \quad (4.10)$$

where  $2\gamma_h$  is the variogram expressed as a function of the separation distance  $|h|$ . An estimate of the variogram may be obtained for the stationary case using the following equation:

$$2\hat{\gamma}_h = \frac{1}{N(h)-1} \sum_{i=1}^{N(h)} (Z_{x_i+h} - Z_{x_i})^2 . \quad (4.11)$$

where  $2\hat{\gamma}_h$  is the estimate of the population variogram and  $N(h)$  is the

total number of pairs of points separated by distance  $|h|$  that have been sampled from a particular realization.

A random field is said to be second-order stationary when its mean, variance, and covariance are all stationary. In this case, the covariance and variogram functions are equivalent tools for describing the spatial correlation of the random field  $Z_x$ , and have the following relation:

$$\gamma_h = \sigma_z^2 - C_h \quad (4.12)$$

where  $\gamma_h$  is termed the semivariogram. Thus, while the covariance function increases with increased correlation, the variogram decreases with increased correlation.

It should be noted that second-order stationarity implies stationarity of the variogram, but the converse is not true. Stationarity of the variogram requires only the statistics of the increment  $(Z_1 - Z_2)$  to be stationary. This is a less restrictive assumption than second-order stationarity since it allows for a linear trending mean. Notice also that in (4.8) knowledge of the population mean is required to estimate the covariance function, while in (4.11) the differencing procedure employed in estimating the variogram eliminates the need to estimate the population mean. Thus, it is often preferable to use the variogram to describe the spatial correlation structure of a statistically random field, particularly when there is limited field data available.

## Properties of the Variogram

The vector  $h$  of the variogram  $2\gamma_h$  has a modulus of  $|h|$  and a direction of  $\alpha$ . In general, for any direction  $\alpha$  the variogram increases with  $|h|$  from its origin  $\gamma_0=0$ . This behavior of the variogram is a representation of the fact that, in general, the expected difference between any two point values increases as the distance  $|h|$  separating them increases. For small values of  $|h|$ , the manner in which the variogram increases characterizes the spatial continuity of the phenomenon. In a sedimentary deposit, for example, spatial changes in hydrogeologic properties usually occur very slowly, resulting in variograms with a gentle, regular growth from zero. In other cases, such as in fractured media, hydrogeologic properties will change rapidly over short distances. This results in what will later be defined as the Nugget Effect. The Pure Nugget Effect occurs when there exists no spatial correlation regardless of the distance between two samples, i.e., they are completely independent of each other.

In any given direction  $\alpha$ , the variogram may stabilize at a plateau beyond some distance  $|h| = a$ , called the range or zone of influence. Beyond this range  $a$ , the influence of a sample disappears and the two quantities  $Z_x$  and  $Z_{x+h}$  are no longer correlated. In an isotropic situation, the range is the same in all directions  $\alpha$  of space. However, many geologic deposits are anisotropic, so the range becomes dependent on  $\alpha$ . For instance, if the vertical variogram presents a stronger variability (i.e., shorter range and higher plateau) than the horizontal variogram, the horizontal sedimentary character of the phenomenon is expressed.

The plateau of constant value beyond the range  $a$  of the semivariogram function is called the "sill,"  $C_0$ , and is equal to the population variance of the random field

$$C_0 = \gamma_\infty = \text{Var}\{Z_x\} = \sigma_z^2, \quad (4.13)$$

where  $\gamma_\infty$  is the semivariogram value for very large distances (greater than the zone of influence  $a$ ). Semivariograms which exhibit both a sill and a range are called "transition" models, cf. Figure 4.1. Note that the existence of a sill implies a finite variance which reveals that second-order stationarity is valid.

Behavior of the variogram near the origin characterizes the continuity in space of the random field  $Z_x$ . Four main types of behavior can be categorized, cf. Figure 4.2.

- (a) Parabolic:  $\gamma_h \sim A|h|^2$  when  $h \rightarrow 0$ . Parabolic behavior indicates highly regular spatial variability.

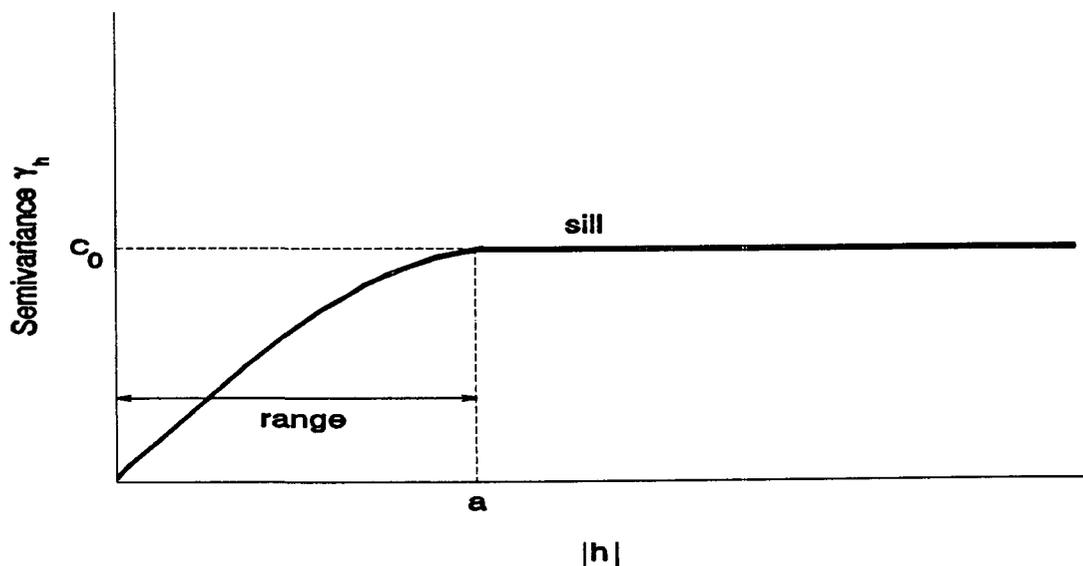


Figure 4.1 Typical transition model with sill  $C_0$  and range  $a$ .

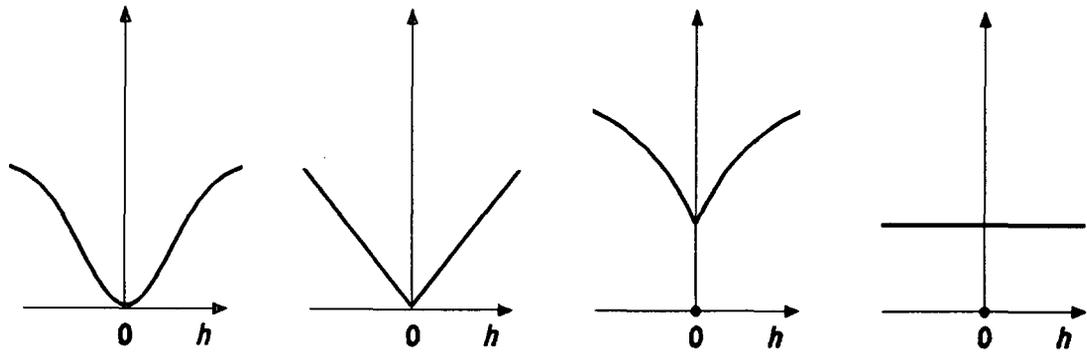


Figure 4.2 Behavior near the origin of the semivariogram.  
 (a) Parabolic behavior; (b) linear behavior;  
 (c) nugget effect; (d) pure nugget effect (after  
 Journel and Huijbregts 1978).

(b) Linear:  $\gamma_h \sim A|h|$  when  $h \rightarrow 0$ .

(c) Nugget Effect:  $\gamma_h$  is discontinuous at  $h=0$ , where  $\gamma_0 = 0$ .

At a distance  $0+\xi$ , where  $\xi$  is very small in relation to data observation distances, the semivariogram function jumps to  $\gamma_{0+\xi} = \gamma_{00}$ , a constant non-zero value called the nugget. As  $|h|$  increases to distances larger than  $\xi$ , the variability becomes more continuous (see Figure 4.2c). This discontinuity of the variogram at the origin is due both to micro-variabilities (i.e., variability over distances smaller than the scale of measurement) of the geologic formation and to human measurement errors.

(d) Pure Nugget Effect:  $\gamma_0 = 0$  and  $\gamma_h = \gamma_{00} = C_0 = \sigma_z^2$  when  $h > \epsilon$ , where  $\epsilon$  is very small in relation to data observation distances. The pure nugget effect corresponds to the absence of auto-correlation. For all experimental distances, no matter how small, the random variables  $Z_x$  and  $Z_{x+h}$  are uncorrelated. This behavior is very rare.

## Inter-Correlated Random Fields

Often, a phenomenon can be described by several independent and/or inter-correlated random fields. It may provide valuable insight to study the relationship between these random fields, especially if they are strongly correlated.

An aquifer, for example, can be characterized by the properties of piezometric head, transmissivity, and storativity. Each property independently can describe an important aspect of the aquifer, while all three together can provide a more complete picture of the aquifer's spatial structure and resource potential. Often, however, information on all three properties is not available for each data point. Therefore, it would be useful to use the information provided by one variable to compensate for missing information on another.

The theoretical development of the probabilistic description of correlated random fields parallels that of a single random field. Assuming a set of  $K$  inter-correlated stationary random fields, the following conditions apply:

- (i) for each random field  $Z_x^k$ , the mathematical expectation exists and is constant,

$$E\{Z_x^k\} = \bar{Z}^k = \text{constant}, \quad \text{for all } x;$$

- (ii) for each pair of random fields  $[Z_x^k, Z_x^{k'}]$ , the cross-covariance exists and depends only on the vector  $h$ ,

$$C_h^{kk'} = E\{Z_x^k \cdot Z_{x+h}^{k'}\} - \bar{Z}^k \bar{Z}^{k'}, \quad \text{for all } x; \quad (4.14)$$

(iii) a cross-variogram exists,

$$2\gamma_h^{kk'} = E\{[Z_x^k - Z_{x+h}^k][Z_x^{k'} - Z_{x+h}^{k'}]\}, \text{ for all } x. \quad (4.15)$$

### Properties of the Cross-Variogram

The properties of the cross-variogram are similar to those of the direct variogram with a few exceptions. The direct variogram is actually a special case of the cross-variogram. When  $k'=k$ , equations (4.14) and (4.15) revert to the definitions of the covariance and variogram,  $C_h$  and  $2\gamma_h$ , equations (4.6) and (4.9).

Whereas a direct semivariogram is always positive, a cross-semivariogram  $\gamma_{kk'}(h)$  can take negative values. A positive value of the cross-semivariogram indicates that a positive increase in one of the variables ( $k$ ) indicates a positive increase in the other ( $k'$ ). A negative value of the cross-semivariogram indicates a negative correlation between variables, i.e., as  $k$  increases,  $k'$  decreases.

As with the direct variogram, the cross-variogram is related to the cross-covariance of second-order stationary random fields by the equation

$$2\gamma_h^{kk'} = 2(\sigma^{kk'})^2 - C_h^{kk'} - C_h^{k'k}. \quad (4.16)$$

### Theoretical Variogram Models

To statistically characterize a particular realization of a random field, it is necessary to perform a structural analysis. It has already been shown that a variogram provides a description of the variability structure of a random field. An experimental variogram consists of a

discontinuous series of formulated data points to which a theoretical continuous function model must be fit. An experimental variogram is incomplete without a fitted model. Therefore, every geostatistical study must begin with the construction of a variogram model. The variogram models developed to describe a random field can also be applied to cross-variograms of two intercorrelated random fields.

For a random field  $Z_x$  with a stationary semivariogram  $\gamma_h$ , there exists two main semivariogram characteristics: the behavior at the origin and the presence or absence of a sill.

Semivariogram models which exhibit a sill are called transition models. The sill value  $C_0 = \sigma_z^2$  of a transition model is defined as the variance of the stationary random field  $Z_x$ , with the covariance being  $C_h = \gamma_\infty - \gamma_h = C_0 - \gamma_h$ . Transition models can be categorized by their behavior at the origin: linear or parabolic, with or without a nugget.

Parabolic behavior at the origin indicates highly regular structural variability and is rarely found in hydrogeology. The most common model used to describe parabolic behavior at the origin is the gaussian model:

$$\gamma_h = C_0 \cdot (1 - \exp(-h^2/a^2)). \quad (4.17)$$

The sill of the semivariogram is reached asymptotically and the parameter  $a$  governs the rate at which the sill is approached. For a gaussian model, the magnitude of the practical range  $a_g = a\sqrt{3}$  gives an indication of the distance over which the random field is correlated, i.e.,  $a_g$  can be considered the distance at which  $\gamma_{a_g} = 0.95C_0 = C_0$  (Journel and Huijbregts 1978).

Linear behavior at the origin is a frequently encountered type of behavior in hydrogeology. Two models describing such behavior are the spherical and exponential models. The spherical model of the semivariogram  $\gamma_h$  is defined as:

$$\gamma_h = C_0 \cdot \begin{cases} \frac{3}{2} \frac{h}{a} - \frac{1}{2} \frac{h^3}{a^3} & 0 \leq h \leq a, \\ 1 = \text{sill} & h \geq a. \end{cases} \quad (4.18)$$

The spherical model effectively reaches its sill at a distance  $h = a$ , beyond which no spatial correlation exists.

The exponential model approaches its sill asymptotically, and is defined by:

$$\gamma_h = C_0 \cdot (1 - \exp(-h/a)), \quad (4.19)$$

where again the parameter  $a$  governs the rate at which the sill is approached. For the exponential model, the practical range  $a_e = 3a$  gives an indication of the distance over which the random field is correlated, i.e.,  $a_e$  can be considered the distance at which  $\gamma_{ae} = 0.95C_0 \sim C_0$  (Journel and Huijbregts 1978). Because of the asymptotic behavior of the exponential model, it reaches its sill more slowly than the spherical model.

When there exists a nugget effect  $\gamma_{00}$ , then the semivariogram model  $\gamma_h$  becomes a nested model composed of the fitted model (for instance, an exponential model) plus the nugget  $\gamma_{00}$ , i.e.:

$$\gamma_h = \gamma_{00} + C_0 \cdot (1 - \exp(-h/a)) \quad (4.20)$$

David (1977) notes that, in general, geologists tend to use the spherical model. In geologic problems, and by inference hydrogeologic problems, the exponential model displays too gentle a growth to adequately characterize the variability structure of a regionalized variable.

The three transitional models can be compared by assigning each model a sill value of  $C_0 = \sigma_z^2 = 1$ . It can then be seen (as was discussed above) that the correlation parameters of the gaussian and exponential models,  $a_g$  and  $a_e$ , can be compared to the correlation parameter  $a$  of the spherical model by the following relationships:

$$a_g = a\sqrt{3}, \quad (4.21)$$

$$a_e = 3a. \quad (4.22)$$

Figure 4.3 shows the relationship of these three sill models.

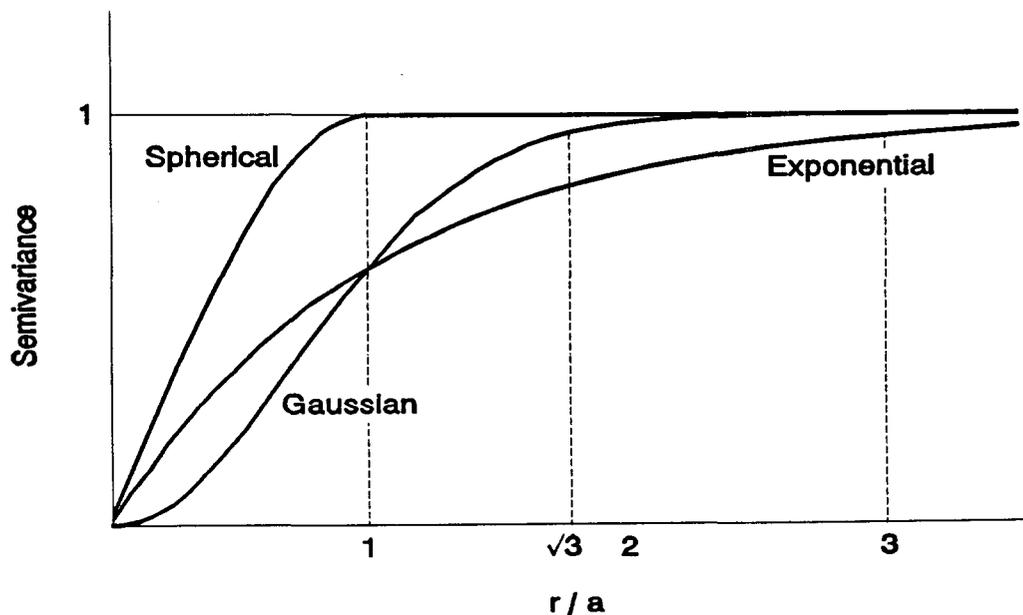


Figure 4.3 Three general transition models (after Journel and Huijbregts 1978).

Two models which do not exhibit a sill are the power and linear models. These nonstationary models correspond to random fields with unlimited capacity for spatial variability (Journel and Huijbregts 1978). Neither the variance nor the covariance of the random field  $Z_x$  can be defined.

The power model is defined as

$$\gamma_h = h^r \quad 0 < r < 2, \quad (4.23)$$

with the limits 0 and 2 excluded from the set of possible values of  $r$ . As  $r$  increases between one and two, the behavior of  $\gamma_h$  at the origin becomes parabolic, indicating a random field which is very continuous in its spatial variability. In practice, power models for  $1 < r < 2$  are seldom used since they indicate a random field with a linearly trending mean. Instead, the trend is usually accounted for with a trend function, and the resulting detrended random field is then modeled as stationary.

A special case of the power model which is frequently used to describe only incrementally stationary, or intrinsic, random fields is the linear model

$$\gamma_h = \omega h \quad (4.24)$$

where  $\omega$  is the slope at the origin and  $r = 1$ . This model, when used for small distances ( $h \rightarrow 0$ ), can also be fitted to spherical and exponential models that have linear behavior at the origin.

## Models of Anisotropy

A phenomenon is said to be isotropic when its semivariogram function  $\gamma_h$  depends only on the modulus  $|h|$  of the vector  $h$ , i.e., the variability of the phenomenon is the same in every direction. If the phenomenon's variability is dependent on direction, then it is anisotropic, and the semivariogram function  $\gamma_h(\alpha, \varphi)$  is direction and vector dependent. The method of modeling an anisotropic semivariogram depends on the type of anisotropy of the phenomenon: either geometric or zonal.

For  $n$  directional semivariograms, geometric anisotropy is present if:

- (a) each of the  $n$  semivariograms can be modeled by  $n$  transition models of the same type;
- (b) each of the  $n$  semivariograms have the same sill;
- (c) the ranges of the  $n$  semivariograms form an elliptical-shaped directional graph in two dimensions or ellipsoidal in three dimensions, cf. Figure 4.4.

For example, consider two semivariograms in two particular directions in space,  $\alpha_1$  and  $\alpha_2$ , which follow the above guidelines. The two semivariograms are unique only in their direction and range. To transform the two anisotropic models  $\gamma_{h_1}(\alpha_1)$  and  $\gamma_{h_2}(\alpha_2)$  into one isotropic representative model  $\gamma_h$ , the two models are equated by the ratio of their ranges  $a_{\alpha_1}/a_{\alpha_2}$ . For the two models

$$\gamma_{h_1}(\alpha_1) = \frac{3}{2a_{\alpha_1}}h_1 - \frac{1}{2a_{\alpha_1}^3}h_1^3, \quad h_1 \leq a_{\alpha_1}, \quad (4.25)$$

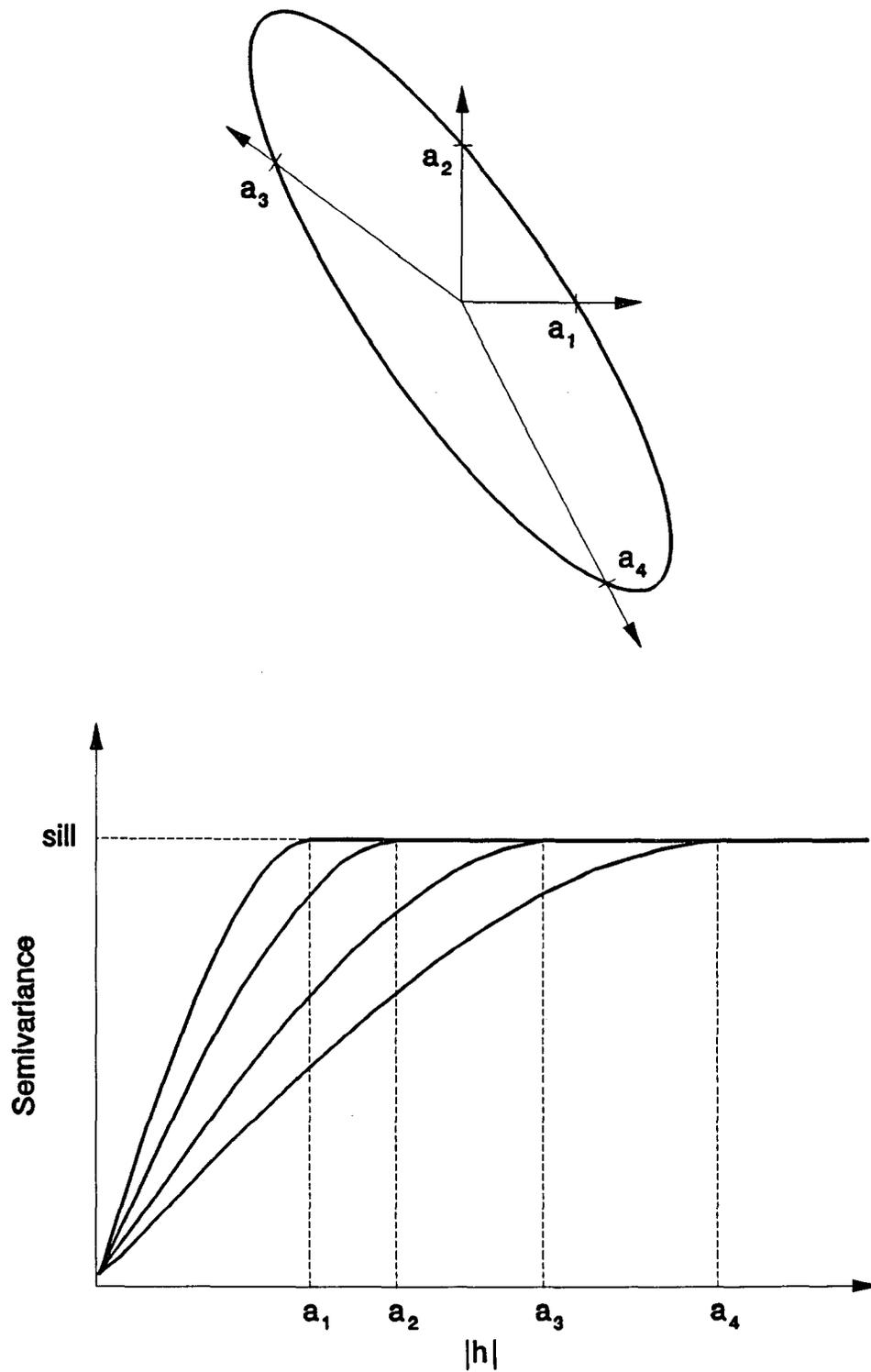


Figure 4.4 Semivariogram and directional graph for geometric anisotropy (after Journel and Huijbregts 1978).

and

$$\gamma_{h_2}(\alpha_2) = \frac{3}{2a_{\alpha_2}} h_2 - \frac{1}{2a_{\alpha_2}^3} h_2^3, \quad h_2 \leq a_{\alpha_2}, \quad (4.26)$$

a single spherical model with a range  $a_{\alpha_2}$  is derived:

$$\gamma_{h'}(\alpha_2) = \frac{3}{2a_{\alpha_2}} h' - \frac{1}{2a_{\alpha_2}^3} h'^3, \quad (4.27)$$

where  $h' = h_1 \frac{a_{\alpha_2}}{a_{\alpha_1}}$ , i.e.,

$$\gamma_{h_1}(\alpha_1) = \gamma_{h'}(\alpha_2).$$

When the anisotropy cannot be reduced by linear transformation, a model of zonal anisotropy must be used. This model is the one most currently used in practice (Journel and Huijbregts 1978). A zonal model of anisotropy is simply a nested structure of anisotropic component structures. Given a coordinate system of  $(u, v, w)$ , consider two semivariogram models  $\gamma_{h_u}$ , constructed in direction  $h_u$ , and  $\gamma_{h_{vw}}$  ( $h_{vw} = \sqrt{h_v^2 + h_w^2}$ ), isotropic for directions  $h_v$  and  $h_w$  at right angles to  $h_u$ . A nested model in three dimensions can be adopted which is just the sum of the first two:

$$\gamma_{h_u, h_v, h_w} = \gamma_{h_u} + \gamma_{h_{vw}}. \quad (4.28)$$

## Constructing a Variogram or Cross-Variogram

When performing a geostatistical analysis on field data, many factors are involved in constructing a meaningful variogram and/or cross-variogram. First, there must be enough sampled data to assure proper statistical analysis. The ideal data sampling configuration for constructing a variogram is a systematic grid with each data point separated from its neighbors by a vector  $h$ . However, in reality, data distribution is often somewhat random and unstructured due to the expense of sampling and availability of existing data. Thus, for nonaligned data, distinct direction paths for variogram computation will not contain enough data points for proper analysis. Instead, the data must be grouped into angle classes and distance classes.

An angle class is defined by the arc  $[\alpha \pm d\alpha]$ , where  $\alpha$  is the direction angle. The increment  $d\alpha$  is appropriately chosen to facilitate proper analysis according to the amount of data points available. Each data value lying within the angle class is associated with every other value located within the arc. Within the angle class, distance classes can be created to compensate for lack of uniform spacing of data. Every data pair that is separated by a distance  $[h \pm \epsilon(h)]$  is used to estimate the value  $\gamma_h$ , where  $\epsilon(h)$  is the distance tolerance. Careful analysis of experimental variograms in four or more directions with appropriate angle classes can reveal anisotropic behavior.

When angle and distance classes are used, a smoothing of the experimental variogram relative to the theoretical variogram occurs. The tolerance  $\epsilon(h)$  should be large enough to ensure that each distance class has enough pairs for variogram reliability, yet small enough to allow at least three or four classes before the sill is reached. For

each angle and distance class combination, there should be on the order of 30 to 50 pairs for the computation of a reliable semivariogram data point  $\gamma_h$ . The grouping into distance classes creates a smoothing effect which is most significant at very small distances  $h$ . Thus for  $bh < 0.1L$ , where  $L$  is the dimension of the field of study and  $b=[1,2,\dots]$ , the experimental variogram data points  $[\gamma_h, \gamma_{2h}, \dots, \gamma_{bh}]$  should be discarded, since they will not be reliable estimates of the actual variogram.

Journel and Huijbregts (1978) state that an experimental variogram should only be considered for distances less than half the dimension of the field of study, i.e.,  $|h| < L/2$ . This distance  $L/2$  is known as the limit of reliability. This limit is a result of the fact that at distances  $|h| > L/2$ , the uncertainty of the variogram estimate is large and thus virtually any theoretical model can be used to fit the semivariogram and statistically pass a reliability-of-fit test.

The procedures and rules of thumb discussed above for the construction of an experimental variogram are also valid for the construction of the experimental cross-variogram.

### Kriging Theory

Kriging, after D. G. Krige, a South African mining engineer who pioneered the use of geostatistics, was first developed as a linear interpolation tool (Rouhani et al. 1990a). Although many estimation and interpolation techniques exist, kriging provides the best (i.e., minimum variance) linear unbiased estimate of the phenomenon being studied as well as a measure of the accuracy of this estimate. It is a local estimation technique which requires only the knowledge of the measured data set and the covariance or variogram of the underlying random field.

For a second-order stationary random field  $Z_x$  with mean  $m$  and covariance function  $C_h$ , the kriging linear estimator  $\hat{Z}_0$  is a weighted sum of the  $n$  measurements of the random field:

$$\hat{Z}_0 = \sum_{i=1}^n \lambda_i Z_{xi} . \quad (4.29)$$

The kriging weights,  $\lambda_i$ , are selected to achieve the desired unbiased minimum variance estimation properties. To ensure that the estimator is unbiased, the expected value of the estimate  $\hat{Z}_0$  must be equal to the true expected value  $m$  of the random field. Applying the expectation operator to (4.29) gives:

$$E\{\hat{Z}_0\} = \sum_i \lambda_i E\{Z_i\} = \sum_i \lambda_i m. \quad (4.30)$$

Therefore, to provide an unbiased estimator, it is required that

$$\sum_{i=1}^n \lambda_i = 1 , \quad (4.31)$$

To achieve the "best" linear estimate possible requires that the  $\lambda_i$ 's also be selected to minimize the estimation error, or estimation variance. The estimation variance  $\sigma_{k0}^2$  is the expected value of the squared difference between  $Z_0$  and  $\hat{Z}_0$ . The expectation of the estimation variance can be written:

$$E\{[Z_0 - \hat{Z}_0]^2\} = E\{Z_0^2\} - 2E\{Z_0\hat{Z}_0\} + E\{\hat{Z}_0^2\}. \quad (4.32)$$

Substituting (4.29) into (4.32) and taking the expected value gives:

$$\sigma_{k0}^2 = \sigma_0^2 - 2 \sum_{i=1}^n \lambda_i C_{0i} + \sum_i \sum_j \lambda_i \lambda_j C_{ij} , \quad (4.33)$$

where:

$\sigma_0^2$  is the variance of  $Z_0$ ,

$C_{0i}$  is the covariance of  $Z_0$  and  $Z_{xi}$ ,

$C_{ij}$  is the covariance of  $Z_{xi}$  and  $Z_{xj}$ .

Thus, to obtain the best linear unbiased estimate, the estimation variance (4.33) is minimized by optimal selection of the  $\lambda_i$ 's subject to the unbiasedness constraint expressed in (4.31). This minimization is easy to achieve using standard lagrangian techniques which adjoin the constraint to the function being minimized, to form a new objective function  $F$  (David 1977):

$$F = \sigma_0^2 - 2 \sum_{i=1}^n \lambda_i C_{0i} + \sum_i \sum_j \lambda_i \lambda_j C_{ij} + 2\mu \left( \sum_i \lambda_i - 1 \right) . \quad (4.34)$$

By taking the derivatives of the adjoined objective function (4.34) with respect to the unknowns ( $\lambda_i$ 's and  $\mu$ ) and setting them equal to 0, a linear system of equations referred to as the kriging system is produced:

$$\begin{cases} \sum_j \lambda_j C_{ij} + \mu = C_{0i} & (i = 1, \dots, n), \\ \sum_i \lambda_i = 1 , \end{cases} \quad (4.35)$$

which in matrix form can be written as  $[K][\lambda] = [M]$ , where  $K$ ,  $\lambda$ , and  $M$  are defined as:

$$[K] = \begin{bmatrix} C_{11} & C_{12} & \cdot & \cdot & \cdot & C_{1n} & 1 \\ C_{21} & C_{22} & \cdot & \cdot & \cdot & C_{2n} & 1 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ C_{n1} & C_{n2} & \cdot & \cdot & \cdot & C_{nn} & 1 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix} \quad [\lambda] = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix} \quad [M] = \begin{bmatrix} C_{01} \\ C_{02} \\ \cdot \\ \cdot \\ \cdot \\ C_{0n} \\ 1 \end{bmatrix} .$$

The kriging matrix  $[K]$  is symmetric, i.e.,  $C_{ij} = C_{ji}$ , and depends only on the spatial configuration of the sampled data points.

Both the  $[K]$  and  $[M]$  matrices can be computed from the covariance, and thus the solution to the kriging equations is  $[\lambda] = [K]^{-1}[M]$ . Once  $[\lambda]$  is known, of course, then the kriging linear estimator  $\hat{Z}_0$  of equation (4.29) can be computed.

The minimum estimation variance as defined in equation (4.33) can be simplified mathematically by recognizing that

$$\sum_i \sum_j \lambda_i \lambda_j C_{ij} + \mu = \sum_i \lambda_i C_{0i} , \quad (4.36)$$

and thus:

$$\sigma_{k0}^2 = \sigma_0^2 - \sum_{i=1}^n \lambda_i C_{0i} - \mu . \quad (4.37)$$

The kriging system of  $[K][\lambda] = [M]$  which includes the Lagrange multiplier  $\mu$  is termed ordinary kriging and does not require prior

knowledge of the mean value  $m$  of the random field  $Z_x$  (Rouhani et al. 1990a).

It has previously been shown that  $C_h = \sigma_z^2 - \gamma_h$  for a second-order stationary random field; thus it can be shown that  $C_{ij} = \sigma_z^2 - \gamma_{ij}$ , where  $\gamma_{ij}$  is defined as the semivariogram of  $Z_{x_i}$  and  $Z_{x_j}$ . With this equality, the ordinary kriging system can be rewritten as:

$$\begin{cases} \sum_j \lambda_j \gamma_{ij} + \mu = \gamma_{0i} , \\ \sum_i \lambda_i = 1 , \end{cases} \quad (4.38)$$

where  $(i = 1, \dots, n)$ .

When the kriging systems are defined with semivariograms, the estimation variance of ordinary kriging becomes:

$$\sigma_{k0}^2 = \sum_{i=1}^n \lambda_i \gamma_{0i} - \mu - \gamma_{00}, \quad (4.39)$$

where  $\gamma_{00}$  is the nugget effect, if any.

The kriging system has a unique solution and the kriging variance is non-negative if and only if the covariance matrix  $[M]$  is positive definite. In addition to being an unbiased estimator, kriging is also an exact interpolator if the measurements are assumed to be error free. Thus, the estimated value at a sampled point is exactly equal to its measured value, and the kriging variance is zero.

The kriging system is valid for any support structure, be it isotropic, anisotropic, nested or not. The kriging system and the kriging variance do not depend on the particular values of the data, but

only on the model structure  $C_h$  or  $\gamma_h$  and the spatial configuration of the data network.

The influence of neighboring data upon the estimated value is not linear in relation to their distance from  $Z_0$ . Although in general it can be said that the influence of  $Z_{x_i}$  on  $Z_0$  decreases as the distance between  $x_i$  and  $x_0$  increases, the influence is also dependent on the relation of the other sample points to the  $(Z_0, Z_{x_i})$  pair. If no other sampled data lie between  $Z_0$  and  $Z_{x_i}$ , then  $Z_{x_i}$  will have a large kriging weight  $\lambda_i$  associated with it, and thus will have a significant effect on  $Z_0$ . However, if other data lie between  $Z_0$  and  $Z_{x_i}$ , then  $\lambda_i$  will take on a smaller value relative to the  $\lambda_j$ 's associated with other nearby measurements  $Z_{x_j}$ . Thus, undue weight will not be given to clusters of data.

### Cokriging

As previously discussed, it is often useful to correlate two or more variables, when there exists a relationship among them, to improve the quality of the structural variability model. This procedure is especially useful when there exists more sampled data on one random field than on another. Once an experimental cross-variogram has been computed and modeled, then estimated values for one of the random fields can be interpolated using information from other fields. This procedure is called cokriging (Journel and Huijbregts 1978; David 1977).

For two random fields, the cokriging estimator is defined as a linear combination of the available sampled data:

$$\hat{Z}_0^k = \sum_{i=1}^n \lambda_i Z_{x_i}^k + \sum_{j=1}^{n'} \theta_j Z_{x_j}^{k'}, \quad (4.40)$$

where  $n$  and  $n'$  are the number of samples of  $Z_x^k$  and  $Z_x^{k'}$ , respectively;  $\lambda_i$  and  $\theta_j$  are the associated weight factors for  $Z_x^k$  and  $Z_x^{k'}$ , respectively. The unbiased condition applied to single variable kriging is similarly required for cokriging for minimum estimation error. Thus, applying the expectation operator to (4.40) gives:

$$E\{\hat{Z}_0^k = \sum_{i=1}^n \lambda_i Z_{x_i}^k + \sum_{j=1}^{n'} \theta_j Z_{x_j}^{k'}\} = \sum_{i=1}^n \lambda_i m^k + \sum_{j=1}^{n'} \theta_j m^{k'}, \quad (4.41)$$

where  $m^k$  is the expected value of the random field  $k$ , and  $m^{k'}$  is the expected value of the random field  $k'$ . To provide an unbiased cokriging estimator, the following conditions must be met:

$$\sum_{i=1}^n \lambda_i = 1 \quad \text{and} \quad \sum_{j=1}^{n'} \theta_j = 0. \quad (4.42)$$

Cokriging allows an improvement in the estimation of a poorly sampled variable when there exists a correlation with a more heavily sampled variable. Rouhani et al. (1990a) state that although it is theoretically possible to use an unlimited number of variables in cokriging, in practice, only three or four variables can be handled efficiently. It is important to note that the cokriging system can be written in terms of the cross-semivariogram only if the cross-covariances are symmetric, i.e.,  $C_h^{kk'} = C_h^{k'k}$ .

### Log-Kriging

The distributions of hydrogeologic variables such as transmissivity are often skewed because they are inherently non-negative. Such variables may not be accurately described using linear

estimators which often yield both positive and negative results. In such cases, a simple logarithmic transformation produces symmetrical, approximately normally distributed values which are often better suited to linear kriging applications. De Marsily (1986) noted that the use of logarithms to transform data often produces better-defined semivariograms with stronger correlations. Log-kriging also preserves the non-negative and skewed nature of the variable.

Consider the stationary random field  $Z_x$  with a multivariate lognormal distribution. It has an expectation  $m$ , a covariance  $C_h$ , and a variance  $\sigma^2$ . To transform the lognormal function to a normal function, the natural logarithm of  $Z_x$  is taken, and the random field  $Y_x$  is defined as  $Y_x = \ln(Z_x)$ . This new function  $Y_x$  has a multivariate normal distribution with an expectation of  $m'$ , covariance function  $C_h'$ , and variance  $\sigma'^2$ . The moments of these random fields are related by the following formulae:

$$m = \exp(m' + \sigma'^2/2), \quad (4.43a)$$

$$C_h = m^2[\exp(C_h') - 1] \Rightarrow \sigma^2 = m^2[\exp(\sigma'^2) - 1] \quad (4.43b)$$

Once the random field  $Z_x$  has been lognormalized to  $Y_x$ , the semivariogram of the random field  $Y_x$  can be determined and the linear kriging estimator  $\hat{Y}_0$  can be computed. The kriging estimate of the non-transformed variable is then defined as:

$$\hat{Z}_0 = \exp[\hat{Y}_0 + \sigma'^2/2]. \quad (4.44)$$

The corresponding estimation (kriging) variance of the non-transformed variable can be expressed as:

$$\sigma_{k0}^2 = m^2 \exp(\sigma'^2) [1 - \exp(-\sigma'_{k0}^2)]. \quad (4.45)$$

Journel and Huijbregts (1978) note that the estimator  $\hat{Z}_0$  as defined in equation (4.55) does not always fulfill the nonbiased condition. To account for this, they suggest multiplying  $\hat{Z}_0$  by a corrective factor  $K_0$  equal to the ratio of the arithmetic mean of  $\hat{Z}_0$  and the expectation  $m$  of the sampled data. Thus the linear estimator becomes:

$$\hat{Z}_0 = K_0 \exp[\hat{Y}_0 + \sigma'_{k0}^2/2]. \quad (4.46)$$

This divergence of the estimator  $\hat{Z}_0 = \exp[\hat{Y}_0 + \sigma'_{k0}^2/2]$  from the expectation  $m$  is due to the lack of robustness of  $\hat{Z}_0$  with regard to the multivariate hypothesis. In other words, while the univariate distribution of  $Z_x$  can be fitted to a lognormal distribution, its multivariate distribution is not necessarily lognormal. Furthermore, it is usually not possible to test the character of the multivariate distribution of  $Z_x$  because, in practice, only one realization  $Z_x$  from a few data points is available.

#### Description of Software Packages Used

The geostatistical analysis of the provided data was facilitated by the use of three computer software packages: SPECTRUM<sup>®</sup>, GEOPACK

(Version 1.0) (Yates and Yates 1990), ARC/INFO<sup>®</sup> and ARC/INFO TIN<sup>®</sup>.<sup>1</sup>

SPECTRUM is a coordinate conversion package, GEOPACK is a geostatistical software package, and ARC/INFO is a geographic information system software package.

SPECTRUM is an interactive geodetic-to-plane coordinate conversion software package which allows geodetic coordinate conversion to one of twelve predefined coordinate systems or one user-defined coordinate system. It also allows for the choice of three coordinate unit systems: meters, U.S. survey feet, and international feet.

GEOPACK is an interactive geostatistical software system developed by S. R. Yates and M. V. Yates, under contract to the U. S. Environmental Protection Agency's R. S. Kerr Environmental Research Laboratory, Ada, Oklahoma. GEOPACK is a package of programs for conducting spatial variability analyses of one or more random fields. The system is menu driven and provides graphic output in a variety of forms. GEOPACK includes programs for the computation of basic statistics, variography, and linear and nonlinear kriging estimation.

The method used by the GEOPACK system in determining the experimental semivariogram is similar to that outlined in this report and by Journel and Huijbregts (1978). Semivariogram model fitting can be performed by using the nonlinear least-squares fitting procedure of Marquardt (1963). The program also allows for the traditional iterative method of manually selecting the model coefficients and then visually inspecting its fit to the experimental semivariogram.

---

<sup>1</sup>SPECTRUM<sup>®</sup> copyright 1990,1991 by RAM-SOFT, Inc., Gainesville, FL, USA. ARC/INFO and ARC/INFO TIN are registered trademarks of Environmental Systems Research Institute, Inc., Redlands, CA, USA.

The GEOPACK system includes programs for the determination of ordinary kriging and cokriging estimates along with their associated estimation variance. The system can also model geometric and zonal anisotropies. In addition, an option is provided to cross-validate and optimize the spatial correlation structure by eliminating measured values, one by one, and comparing kriged estimates to actual measured values.

ARC/INFO is a geographic information system used to automate, manipulate, analyze, and display geographic data in digital form. ARC/INFO TIN is a contouring package which stores, manages, and performs analyses of three-dimensional surfaces. The acronym TIN stands for Triangulated Irregular Network - a set of adjacent, nonoverlapping triangles developed from irregularly spaced points having x, y, and z values. Each triangle is similar to a facet of a surface. Once a TIN surface has been created, contour lines can be generated representing the z values. The two methods available for interpolating contour lines from a TIN surface are linear and bivariate quintic interpolation. TIN generated contour maps can then be plotted using the ARCPLOT package of the ARC/INFO system.

#### Hydrogeologic Data Description and Analysis

The St. Johns River Water Management District provided location and hydrogeologic data on 194 wells within the district and surrounding counties. For each well, the District provided data on latitude, longitude, and transmissivity. In addition, storativity and leakance values were provided for some wells.

Some of the wells were located in clusters of two or more with each well having the same latitude and longitude designations. In such instances, it was feasible to combine the information from those well clusters to describe a single composite well at that location. Thus, the original list of 194 wells was reduced to 188 wells. Of those 188 wells, 21 wells penetrated the shallow aquifer, 15 wells penetrated the intermediate aquifer, and 152 wells penetrated the upper Floridan Aquifer.

Occasionally more than one value of transmissivity, storativity, or leakance was provided for a particular well. These values corresponded to multiple tests performed on the same well on separate occasions. When multiple values were provided, the arithmetic mean of the sampled values was used to characterize the aquifer at that well location. Table 4.1 lists the quantity of useful data provided per aquifer by the District. For the Upper Floridan, 69 values of storativity were originally provided. For analytical simplicity, it was assumed that the Upper Floridan is a confined aquifer throughout the district. Bouwer (1978) states that storativity values for a confined aquifer lie in the range of 0.01 to 0.00005. Consequently, for reliability, 7 of the 69 storativity values provided by the District for

Table 4.1 Useful hydrogeologic data provided per aquifer by the St. Johns River Water Management District.

Aquifer	Number of Wells	Number of Wells with Values for		
		Transmissivity	Storativity	Leakance
Shallow	21	21	9	2
Intermediate	15	15	9	4
Upper Floridan	152	152	62	32

the Upper Floridan were discarded for lying outside of this range. Thus, only 62 storativity values for the Upper Floridan were deemed useful. Latitude, longitude, transmissivity, storativity, and leakance data values for each of the 188 wells are listed in Appendix A according to aquifer penetration.

Upon receiving the well data from the St. Johns River Water Management District, computer files containing the data were prepared with the data separated according to aquifer penetration. Using the SPECTRUM software, the latitude and longitude values for each well were converted to East Florida state plane coordinates in units of U.S. Survey feet.

It became immediately obvious that there was not sufficient data to conduct a variogram analysis for the shallow and intermediate aquifers. It will be recalled that for each distance and angle class, it is necessary to have on the order of 30 to 50 data pairs for the reliable computation of an experimental semivariogram. Additionally, it is important to have sufficient experimental semivariogram points to adequately model the variability structure of the phenomenon. For both the shallow and intermediate aquifers, even when  $h$  was set large enough to allow for a minimum of 10 experimental semivariogram points, each point only represented an average  $N'$  of 14 data pairs. Consequently, structural analysis was performed only on those wells which penetrated the Upper Floridan aquifer.

Figures 4.5, 4.6, and 4.7 show the spatial distribution of transmissivity, storativity, and leakance data, respectively, for the Upper Floridan aquifer. The sparsity of the leakance data should be noted. It was decided that there was not sufficient Upper Floridan

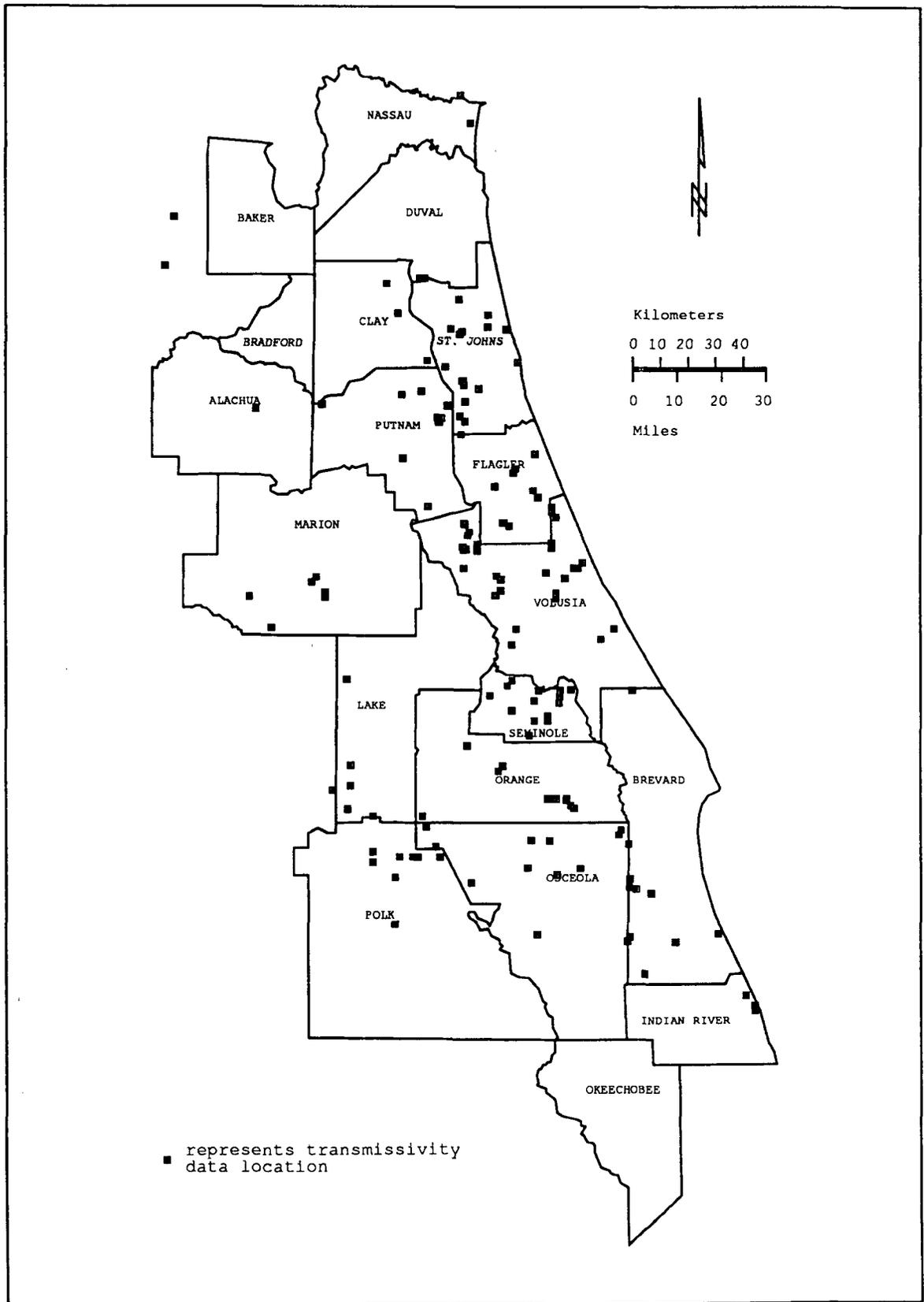


Figure 4.5 Spatial distribution of transmissivity data for Upper Floridan aquifer.

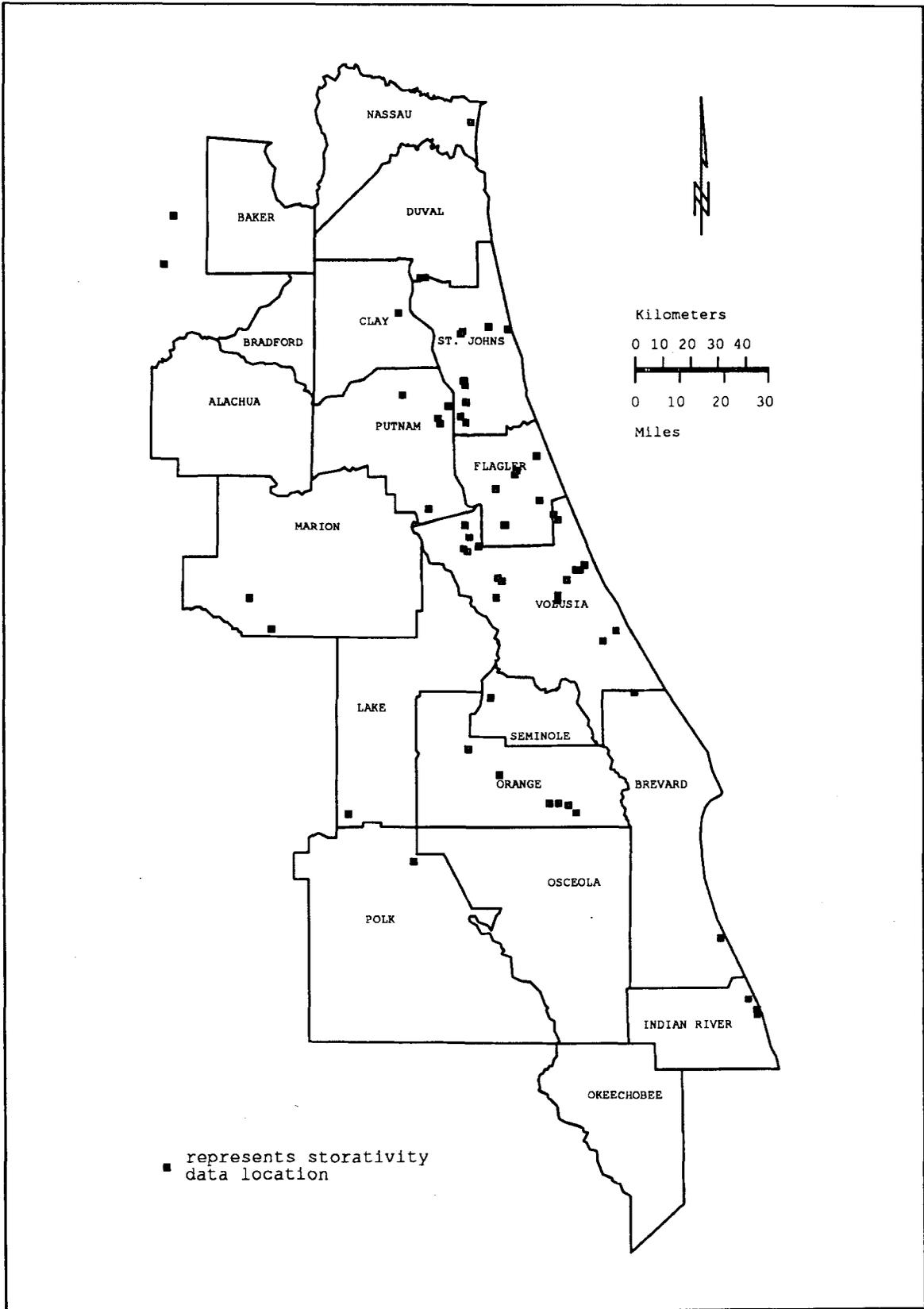


Figure 4.6 Spatial distribution of storativity data for Upper Floridan aquifer.

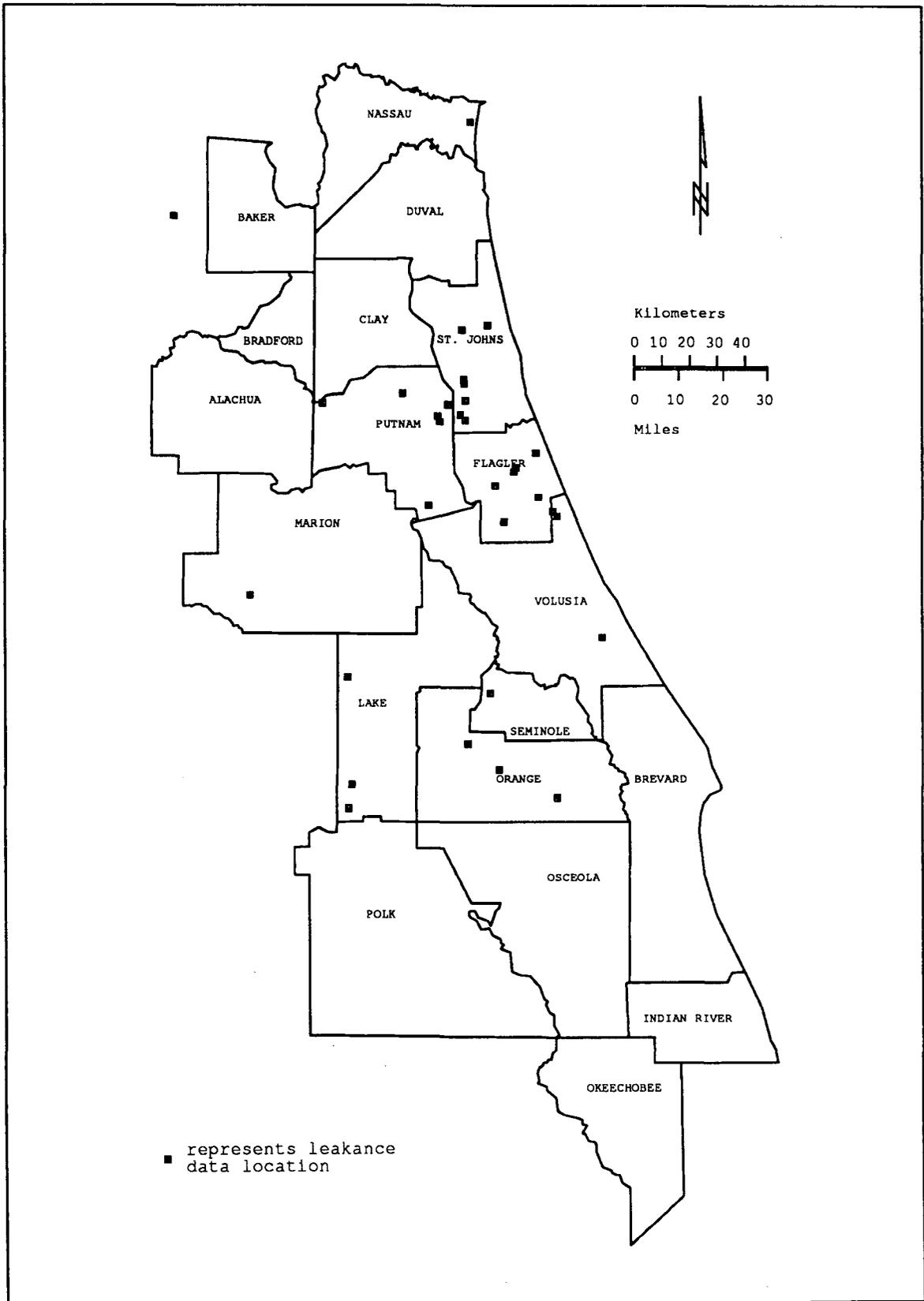


Figure 4.7 Spatial distribution of leakance data for Upper Floridan aquifer.

aquifer leakance data to perform a reliable structural analysis. With only 35 leakance values available for analysis,  $N'$  was limited to just 13 to 57 pairs. While some experimental semivariogram points were represented by an adequate  $N'$  pairs, there did not exist enough semivariogram points for the reliable calculation of a structural model.

The GEOPACK software was used to statistically analyze the hydrogeologic data for the Upper Floridan aquifer. The original transmissivity and storativity data was found not to have a normal distribution. Although a normal distribution is not necessary for variogram analysis, it does allow for certain simplifying assumptions to be made in the computation of the experimental variogram. Thus, a logarithmic transformation of the data was performed.

Experimental variograms of the transformed data were computed using the GEOPACK variogram package. Semivariogram models were then fit using GEOPACK's nonlinear least-squares fitting procedure. Estimates for the model coefficients were thus generated and were then used in a cross-validation program to optimize the model fit. Plots of the experimental semivariograms with fitted models were then created for visual inspection.

Once the spatial variability of the system was evaluated, the selected semivariogram model was used to perform kriging over a very large grid covering the District. The GEOPACK ordinary kriging package was used to prepare the kriging estimates and the kriging variances. If cokriging was desired, the GEOPACK cokriging package was employed.

The kriging estimate and kriging variance data generated by GEOPACK was then converted into the correct format for processing in the ARC/INFO TIN contouring package. If kriging was performed on

logarithmically transformed data, the inverse transform was performed on the kriging estimates and variances before mapping. TIN was then used to create maps of the estimates and their standard deviations which were computed from the estimate variances.

## Results and Discussion

### Estimation of Transmissivity in the Upper Floridan Aquifer

A complete structural analysis was first performed on the transmissivity data from the Upper Floridan aquifer. A statistical analysis of the original data showed it to have a mean value of  $\hat{Z} = 45,206 \text{ ft}^2/\text{day}$ , a standard deviation of  $98,351 \text{ ft}^2/\text{day}$ , and a skewness of 5.15. A Kolmogorov-Smirnov test for normal distribution indicated that the original data was not normally distributed (Figure 4.8) with the data failing the test at the 99, 95, and 90 percent confidence levels. It was decided to transform the transmissivity data by taking the natural logarithm of the data,  $Y_{xi} = \ln(Z_{xi})$ . With a mean value of  $\hat{Y} = 9.8611$ , a standard deviation of 1.24, and a skewness of 0.22, the transformed data (Figure 4.9) was not rejected by the Kolmogorov-Smirnov test for normality at the 99, 95, or 90 percent confidence levels. Table 4.2 lists the districtwide sample statistics for both the original and the transformed transmissivity data.

A review of the transmissivity data distribution revealed a field of study length of approximately  $L = 1,200,000 \text{ ft}$  for the district. Thus, the limit of reliability for variogram calculation was set at  $L/2 = 600,000 \text{ ft}$ . A distance class  $[h \pm \epsilon(h)]$  was chosen with  $h = 6000 \text{ ft}$  and  $\epsilon(h) = 3000 \text{ ft}$ , thus allowing for the computation of 100 experimental semivariogram points. It can be seen in Figure 4.5 that the shape of

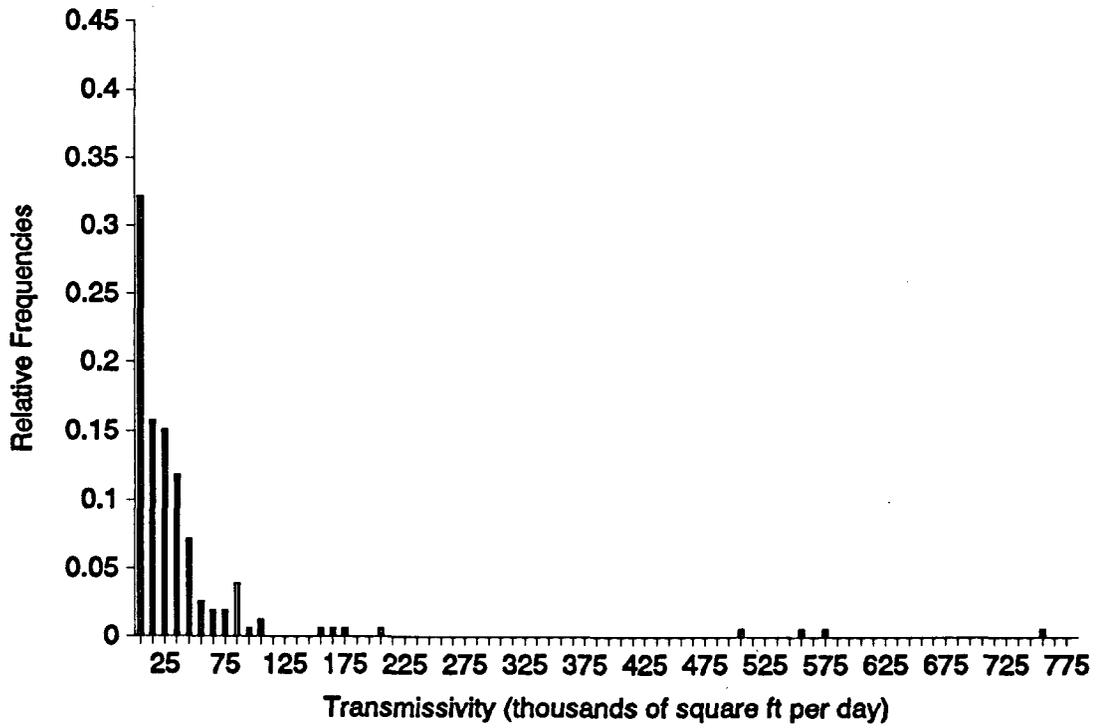


Figure 4.8 Histogram of transmissivity data for the Upper Floridan aquifer.

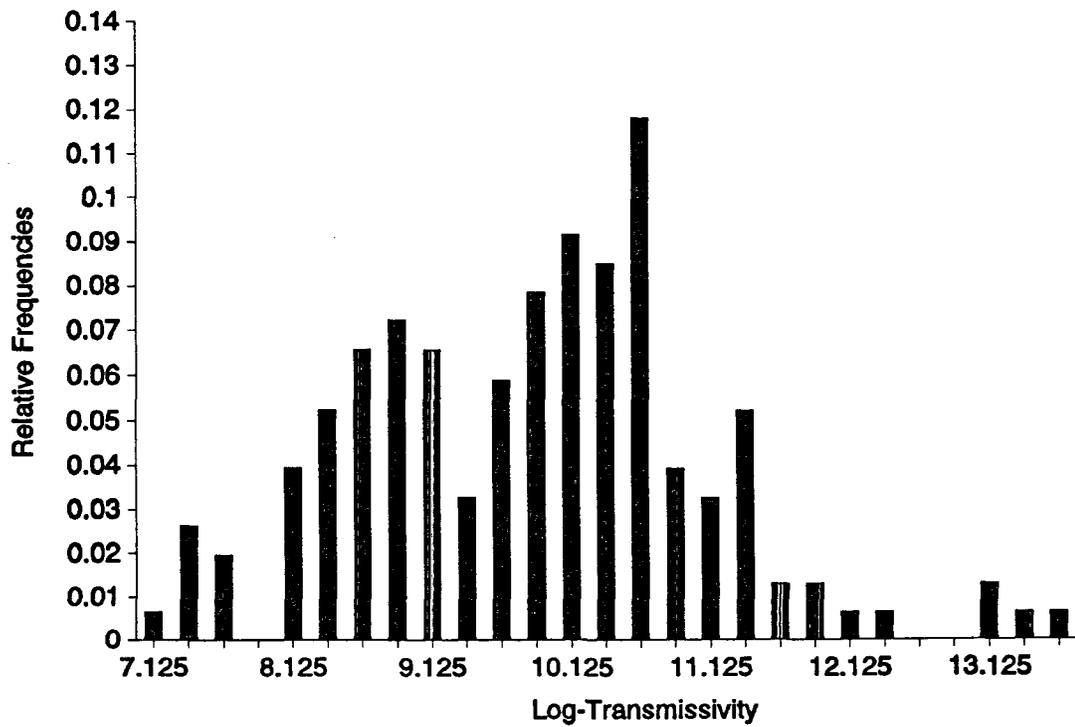


Figure 4.9 Histogram of log-transformed Upper Floridan transmissivity data.

Table 4.2 Sample statistics for Upper Floridan transmissivity data for St. Johns River Water Management District.

	Original Data	Transformed Data
Number of Data Values =	152	152
Mean Value =	45,205.63000	9.86107
Median Value =	22,002.50000	9.99891
Standard Deviation =	98,350.77000	1.24002
Variance =	9,609,238,000.00000	1.52752
Skewness =	5.15055	0.22073
Kurtosis =	31.58301	3.19784
Minimum Data Value =	1,216.58000	7.10380
Maximum Data Value =	758,016.00000	13.53850

the district created a roughly rectangular data distribution, with more data points lying on a north-south axis than on an east-west axis. Due to this uneven spatial distribution, insufficient data existed in the east-west direction to allow for the computation of experimental directional semivariograms. To determine structural anisotropy, experimental directional semivariograms must be computed along several axes using angle class increments of  $d\alpha \leq 45^\circ$ . Given the limitation of  $N' > 30$  to 50 pairs and the skewed distribution of the transmissivity data, angle class increments of this size could not be used to produce reliable experimental directional semivariograms. Thus, it was decided that only an isotropic experimental variogram should be produced, and the angle class increment was set to  $d\alpha = 90^\circ$ .

It will be recalled that grouping data into distance and angle classes creates a smoothing effect particularly significant for distances less than  $0.1L$ . For this reason, the first point of the experimental semivariogram, i.e.  $\hat{\gamma}_{3000}$ , was rejected. Otherwise, all the other experimental semivariogram points met the required limitations. Figure 4.10 shows the districtwide isotropic experimental semivariogram

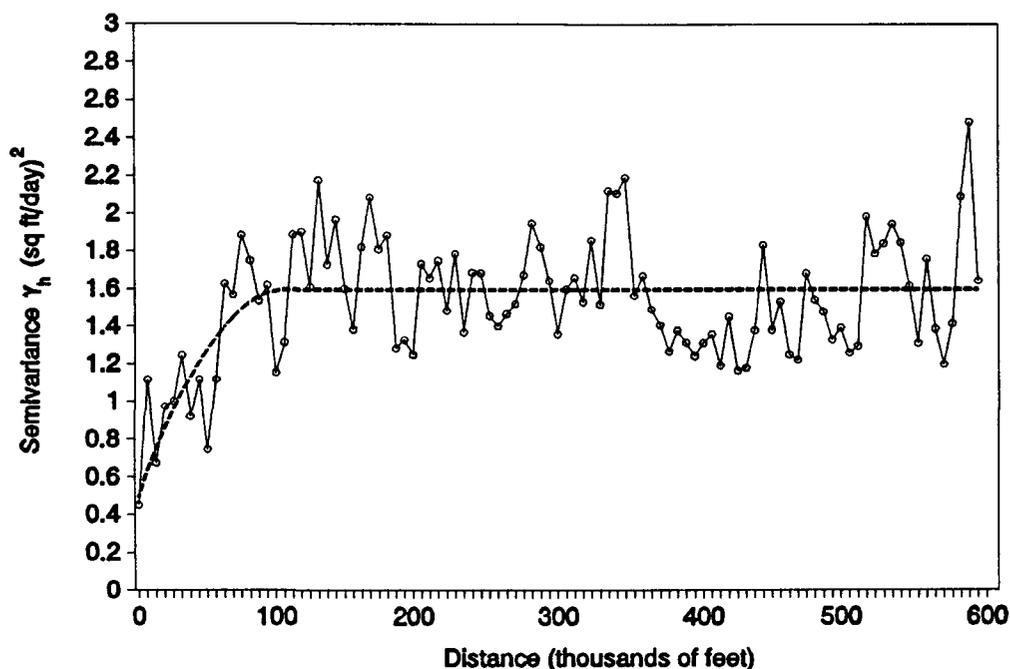


Figure 4.10 Spherical model fitted to a semivariogram of districtwide log transformed transmissivity values in the Upper Floridan aquifer.

computed for all 152 Upper Floridan log-transformed transmissivity data values.

A cross-validated spherical model with a nugget effect was fitted to the experimental semivariogram with  $\gamma_{00} = 0.5$  for the nugget constant, a sill of  $C_0 = 1.6$ , and a correlation parameter of  $a = 101,000$  ft. Note that the model sill is close to the sample variance  $\sigma_z^2 = 1.53$  of the districtwide log-transformed data, as expected.

The fitted semivariogram model was then utilized in a kriging routine to estimate log-transformed transmissivity values over a grid of 12,672 points. The estimation grid consisted of 88 estimation points in the east-west direction and 144 points in the north-south direction, with a separation distance in both directions of 10,000 ft. Before the

kriged values could be used in producing maps, an inverse transformation of the kriged estimates and variances had to be performed. Employing (4.45) and (4.46), transmissivity kriged estimates and variances were produced from the log-transformed data. The mean of the estimated values  $\hat{Z}_0$  was determined to be 50,742 ft<sup>2</sup>/d, while the expectation  $m$  estimated from the original data was 45,206 ft<sup>2</sup>/d. Thus a corrective factor of  $K_0 = 45,206/50,724 = 0.89$  was applied. Figures 4.11 and 4.12 show the resulting transmissivity estimation and standard deviation maps for the Upper Floridan aquifer based on a districtwide semivariogram. It should be noted in Figure 4.11 that in regions far from measurement locations the kriging estimator approaches the sample mean of the data measurements.

The available data revealed a concentration of very high transmissivity values in the Orange county region. This cluster of high transmissivity values noticeably influenced the sample mean  $m$  of the transmissivity measurements taken over the District as a whole. Consequently, kriging with a single semivariogram over the District could tend to overestimate transmissivity values in some regions of the District while underestimating transmissivity values in others if the mean transmissivity exhibits a noticeable spatial trend. As this cluster of high values was located in the southern half of the District, it was decided to divide the District into a northern subdistrict and a southern subdistrict and to statistically analyze the transmissivity data in each. Figure 4.13 shows the areas covered by the two subdistricts.

Statistical analysis of the data in the two subdistricts revealed skewed distributions in both. Therefore, log transformations of both

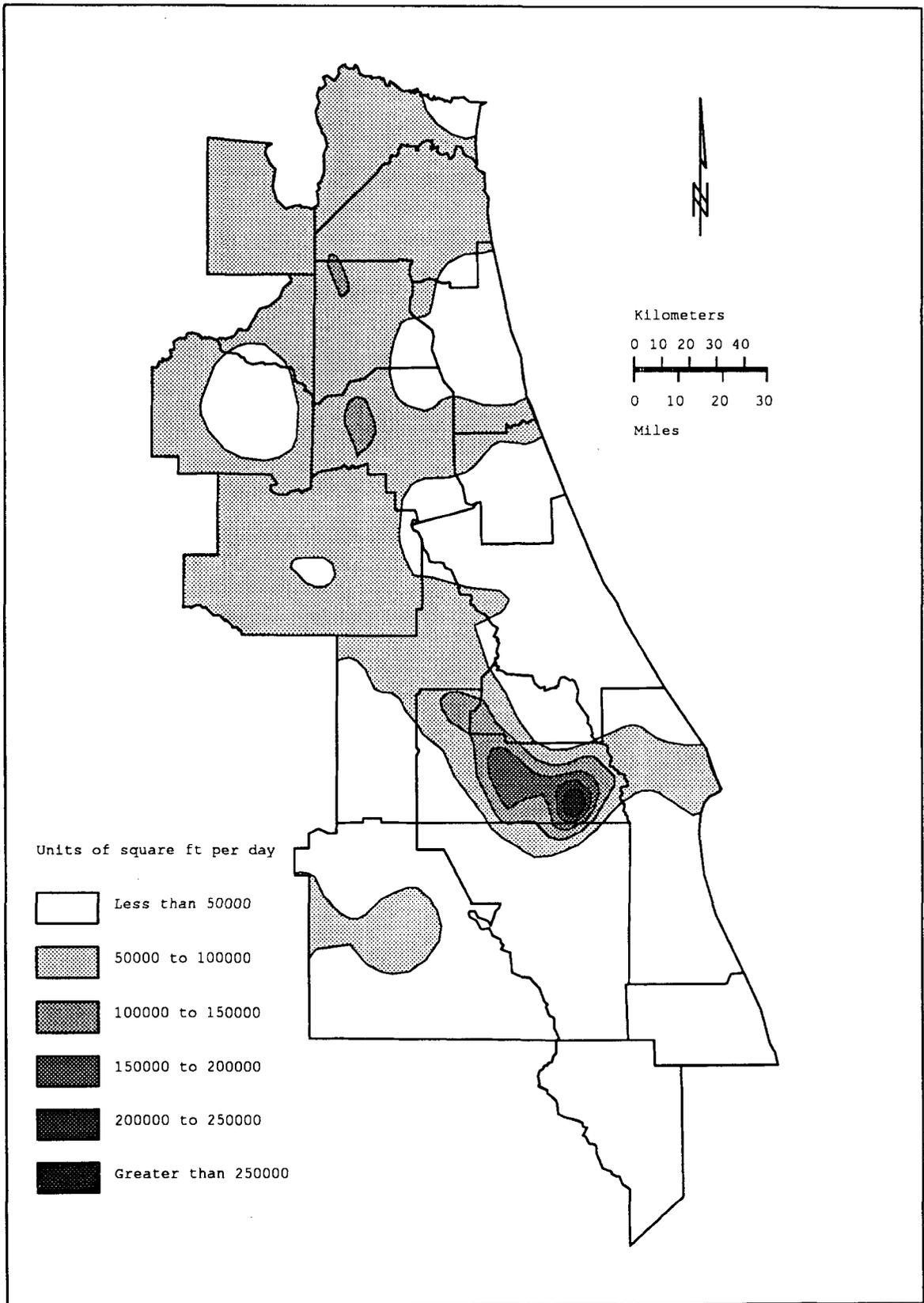


Figure 4.11 Transmissivity estimation map for the Upper Floridan aquifer based on a districtwide semivariogram.

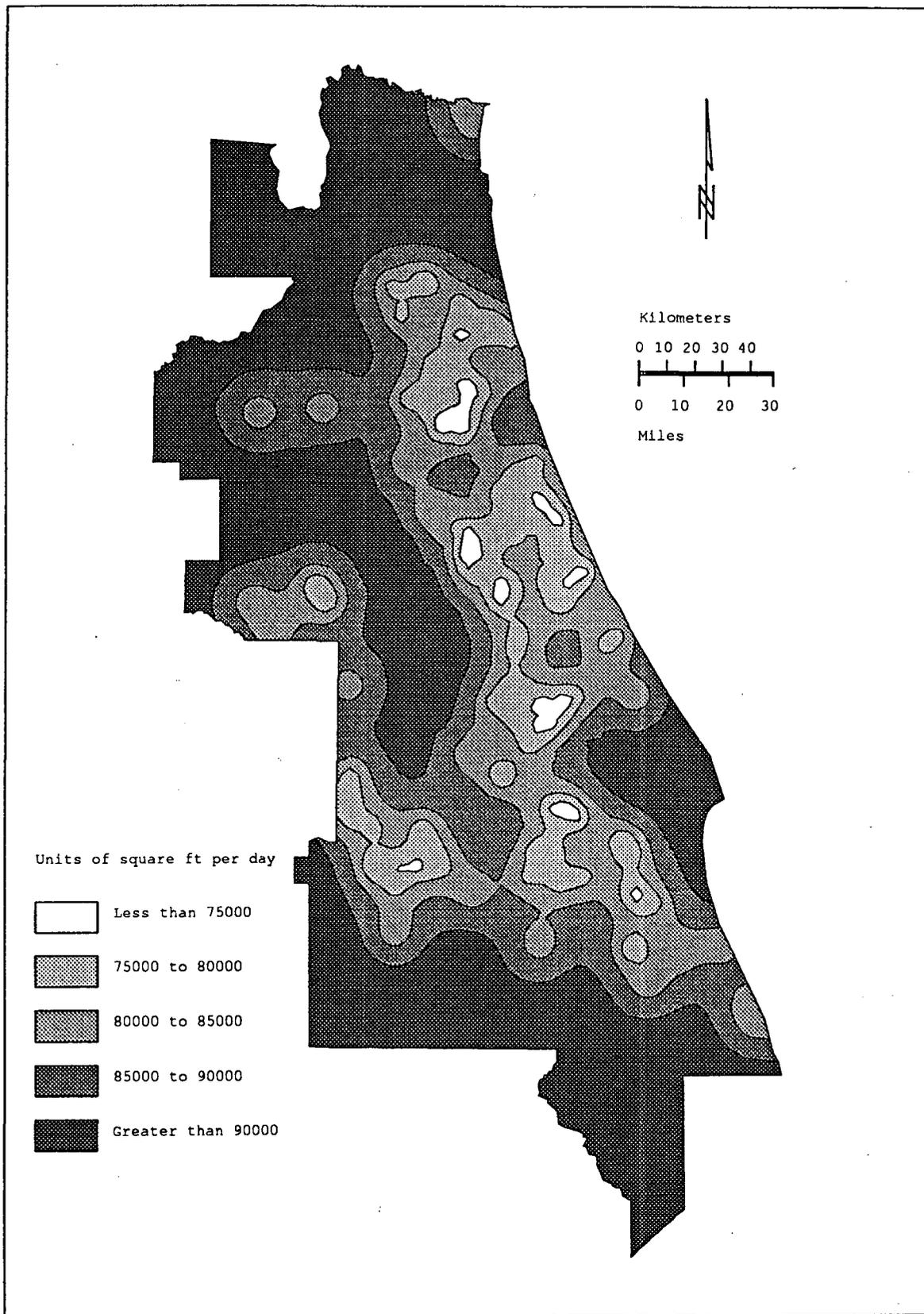


Figure 4.12 Standard deviations of transmissivity estimates based on a districtwide semivariogram.

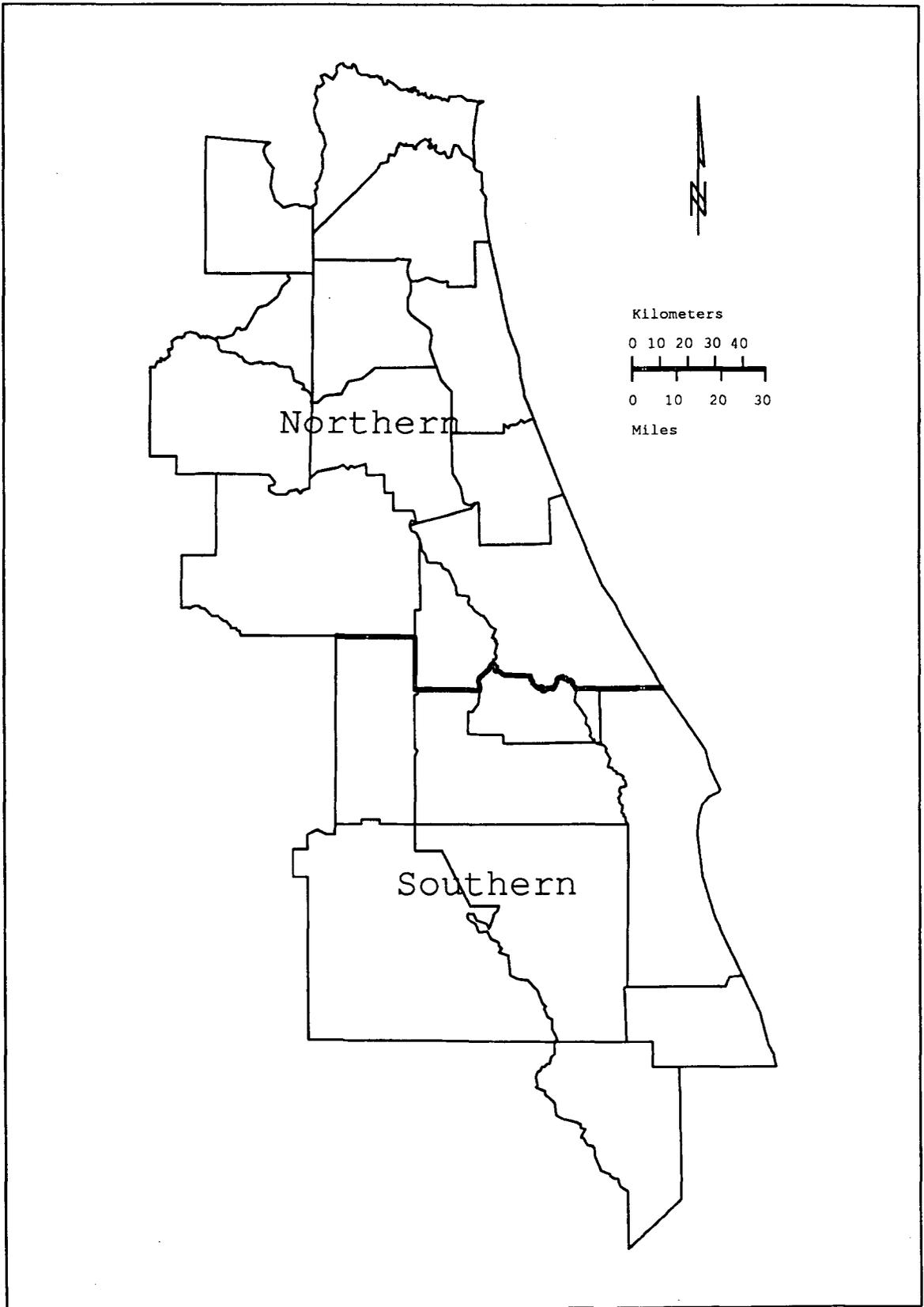


Figure 4.13 Northern and southern subdistricts.

Table 4.3 Sample statistics for Upper Floridan transmissivity data for northern subdistrict.

	Original Data	Transformed Data
Number of Data Values =	88	88
Mean Value =	31,193.29000	9.95146
Median Value =	24,749.30000	10.11650
Standard Deviation =	27,583.39000	0.98589
Variance =	752,197,700.00000	0.96016
Skewness =	1.89673	-0.60075
Kurtosis =	7.80988	3.03772
Minimum Data Value =	1,604.00000	7.38043
Maximum Data Value =	160,026.00000	11.98310

sets of data were performed for the purposes of kriging. Table 4.3 lists the sample statistics for both the original and the transformed transmissivity data for the northern subdistrict, while Table 4.4 lists the sample statistics for the southern subdistrict. It should be noted that the sample mean of the northern subdistrict was about half the sample mean of the southern subdistrict, reflecting the large influence of the cluster of high transmissivity values. Also, the sample variance of the northern subdistrict was significantly lower than the sample variance of either the southern subdistrict or District as a whole.

Table 4.4 Sample statistics for Upper Floridan transmissivity data for southern subdistrict.

	Original Data	Transformed Data
Number of Data Values =	64	64
Mean Value =	64,472.58000	9.73679
Median Value =	13,296.65000	9.49516
Standard Deviation =	146,567.90000	1.52223
Variance =	21,146,480,000.00000	2.28097
Skewness =	3.30317	0.64738
Kurtosis =	13.35923	2.82378
Minimum Data Value =	1,217.00000	7.10380
Maximum Data Value =	758,016.00000	13.53850

A review of the transmissivity data distribution in the northern subdistrict revealed a field of study length of approximately  $L = 720,000$  ft. Thus, the limit of reliability was set at  $L/2 = 360,000$  ft. A distance class  $[h \pm \varepsilon(h)]$  was chosen with  $h = 6000$  ft and  $\varepsilon(h) = 3000$  ft, thus allowing for the computation of 60 experimental semivariogram points. An isotropic experimental semivariogram was produced with an angle class increment of  $d\alpha = 90^\circ$ .

Due to the smoothing effect of distance and angle classes, the first point of the experimental semivariogram, i.e.  $\hat{\gamma}_{3000}$ , was rejected. Further, points  $\hat{\gamma}_{18000}$  and  $\hat{\gamma}_{348000}$  were rejected for having  $N'$  less than 30 pairs. Otherwise, all the other experimental semivariogram points met the required limitations. Figure 4.14 shows the isotropic experimental

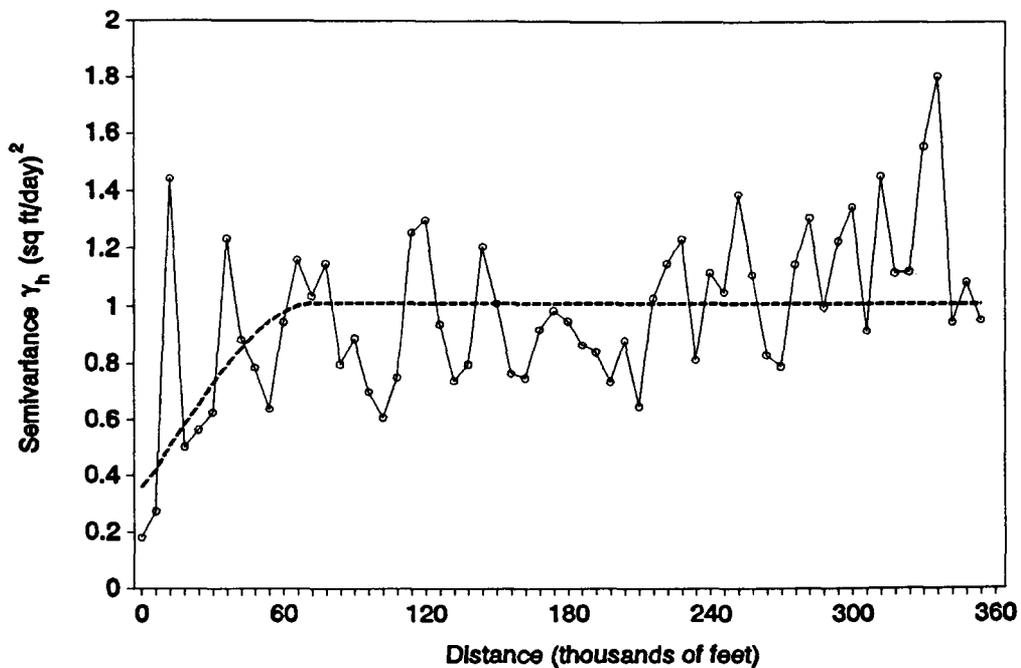


Figure 4.14 Spherical model fitted to a semivariogram of log transformed transmissivity values for the northern subdistrict.

semivariogram computed for the log-transformed transmissivity data values of the northern subdistrict.

A cross-validated spherical model with a nugget effect was fitted to the northern subdistrict experimental semivariogram with a nugget constant of  $\gamma_{00} = 0.34$ , a sill of  $C_0 = 1.0$ , and a correlation parameter of  $a = 73,000$  ft. Once again, note that the model sill is very close to the sample variance  $\sigma_z^2 = 0.96$  of the northern subdistrict log-transformed data. It should also be noted that the sill of the northern subdistrict is smaller than the sill of the District as a whole, indicating less data variability in the northern subdistrict. The range of influence in the northern subdistrict is smaller than for the District as a whole.

The field of study length in the southern subdistrict was determined to be  $L = 620,000$ , giving a limit of reliability of 310,000 ft. A distance class  $[h \pm \epsilon(h)]$  was chosen with  $h = 10,000$  ft and  $\epsilon(h) = 5000$  ft and the angle class increment was set to  $d\alpha = 90^\circ$ . An isotropic experimental semivariogram was thus produced with 31 points. None of the experimental semivariogram points were rejected. Figure 4.15 shows the isotropic experimental semivariogram computed for the log-transformed transmissivity data values of the southern subdistrict.

A cross-validated spherical model with a nugget effect was fitted to the southern subdistrict experimental semivariogram. For this model, the nugget constant was determined to be  $\gamma_{00} = 0.33$ , with the sill set to  $C_0 = 2.67$ , and the correlation parameter computed as  $a = 93,000$  ft. Note that the sill is somewhat larger than the southern subdistrict sample variance of  $\sigma_z^2 = 2.28$ . This discrepancy indicates that the log-transformed transmissivity data in the southern subdistrict may still

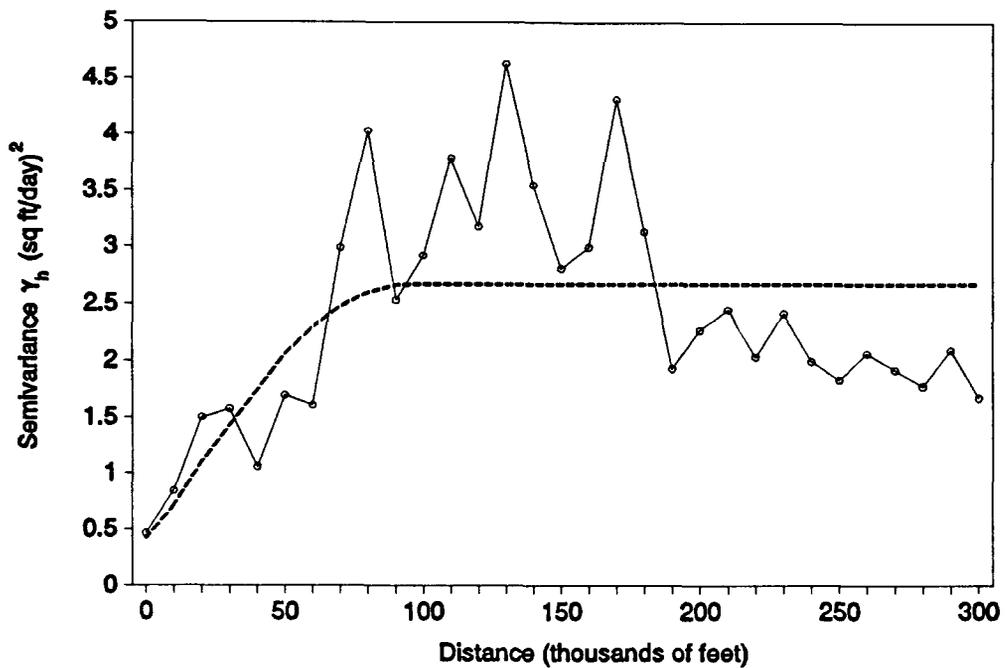


Figure 4.15 Spherical model fitted to a semivariogram of log transformed transmissivity values for the southern subdistrict.

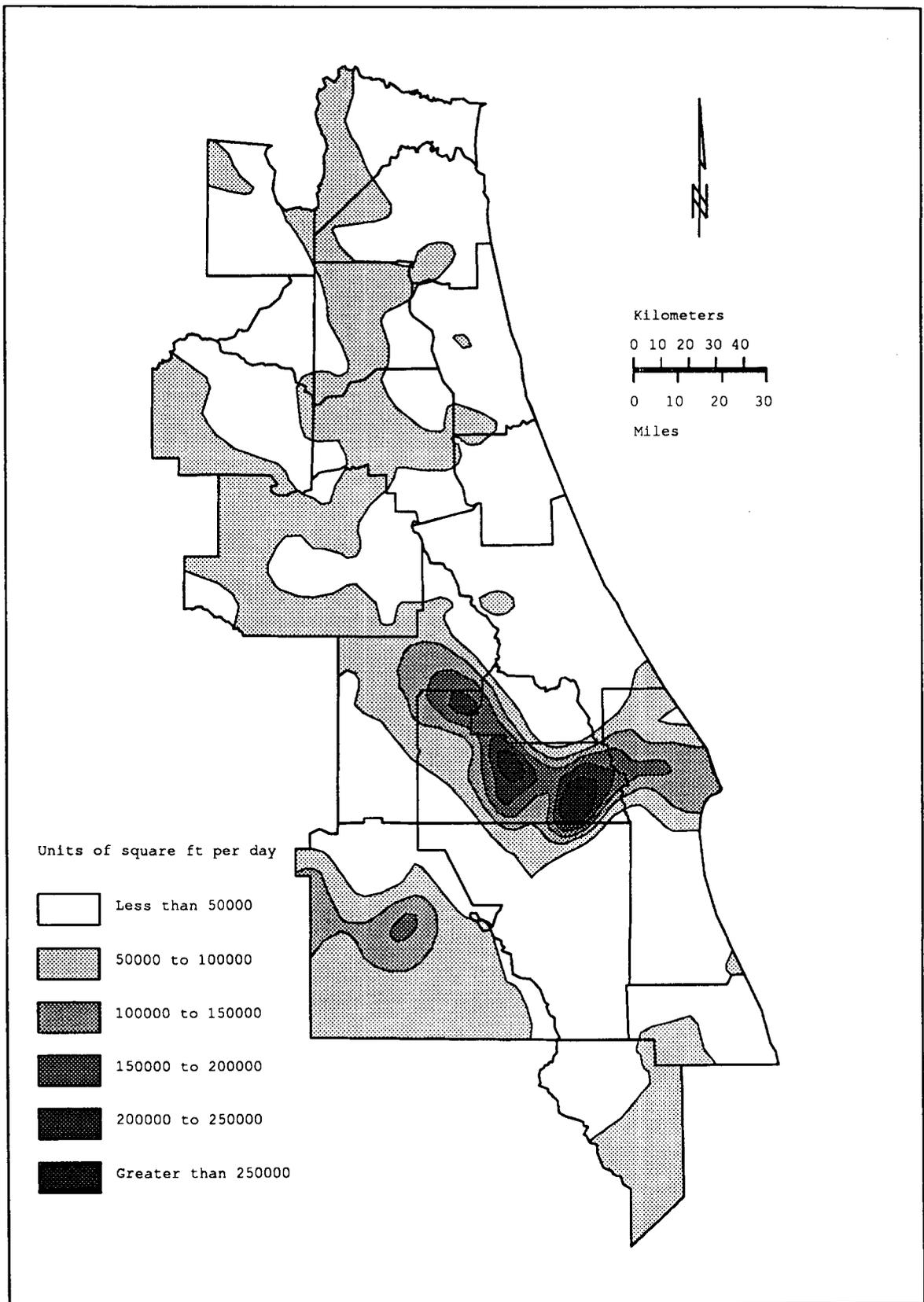
exhibit some nonstationarity due to the cluster of high data values around Orange county. This nonstationarity is further revealed by the "hump" present between lag distances of 60,000 ft and 190,000 ft in the semivariogram of Figure 4.15. The southern subdistrict sill is much larger than the sills of the northern subdistrict and the District as a whole, indicating greater data variability in the southern region of the District. The southern subdistrict range of influence is larger than the northern subdistrict range of influence, yet is still somewhat smaller than that of the District as a whole.

The two fitted semivariogram models for the northern and southern subdistricts were utilized in a kriging routine to estimate log-transformed transmissivity values. The estimated values were then

inversely transformed using (4.45) and (4.46). A corrective factor of  $K_0 = 0.75$  was applied for the northern subdistrict, while a corrective factor of  $K_0 = 1.00$  was applied for the southern subdistrict. Figures 4.16 and 4.17 show the resulting transmissivity estimation and standard deviation maps for the Upper Floridan aquifer based on the northern and southern subdistrict semivariograms.

The kriged transmissivity values created from the northern and southern semivariograms are more variable than those values kriged from the districtwide semivariogram. Comparison of Figures 4.16 and 4.11 reveals that, compared with kriging over the whole District with a single variogram, kriging with the northern and southern semivariograms produced lower estimated transmissivity values in the north and higher estimated transmissivity values in the south. Additionally, kriging with the northern and southern semivariograms produced smaller estimation uncertainty in the north and greater estimation uncertainty in the south, cf. Figures 4.17 and 4.12.

The radial nature of the estimates and their variance is a result of modelling the random field as isotropic. The estimation variance is smallest in areas surrounding clusters of sampled data points and increases radially outward, as can be seen in Figures 4.12 and 4.17. The estimation variance is greatest where no sampled data points fall within the semivariogram correlation distance  $a$  of the point of estimation. It should be noted that the coefficient of variation  $\sigma_0/m_0$  (where  $m_0$  is the conditional mean, or estimated value, and  $\sigma_0$  is the estimation standard deviation at point  $x_0$ ) is much greater than one in several areas of the estimation maps. These large uncertainties are due to the high variability of the available data.



**Figure 4.16** Transmissivity estimation composite map based on the northern and southern subdistrict semivariograms.

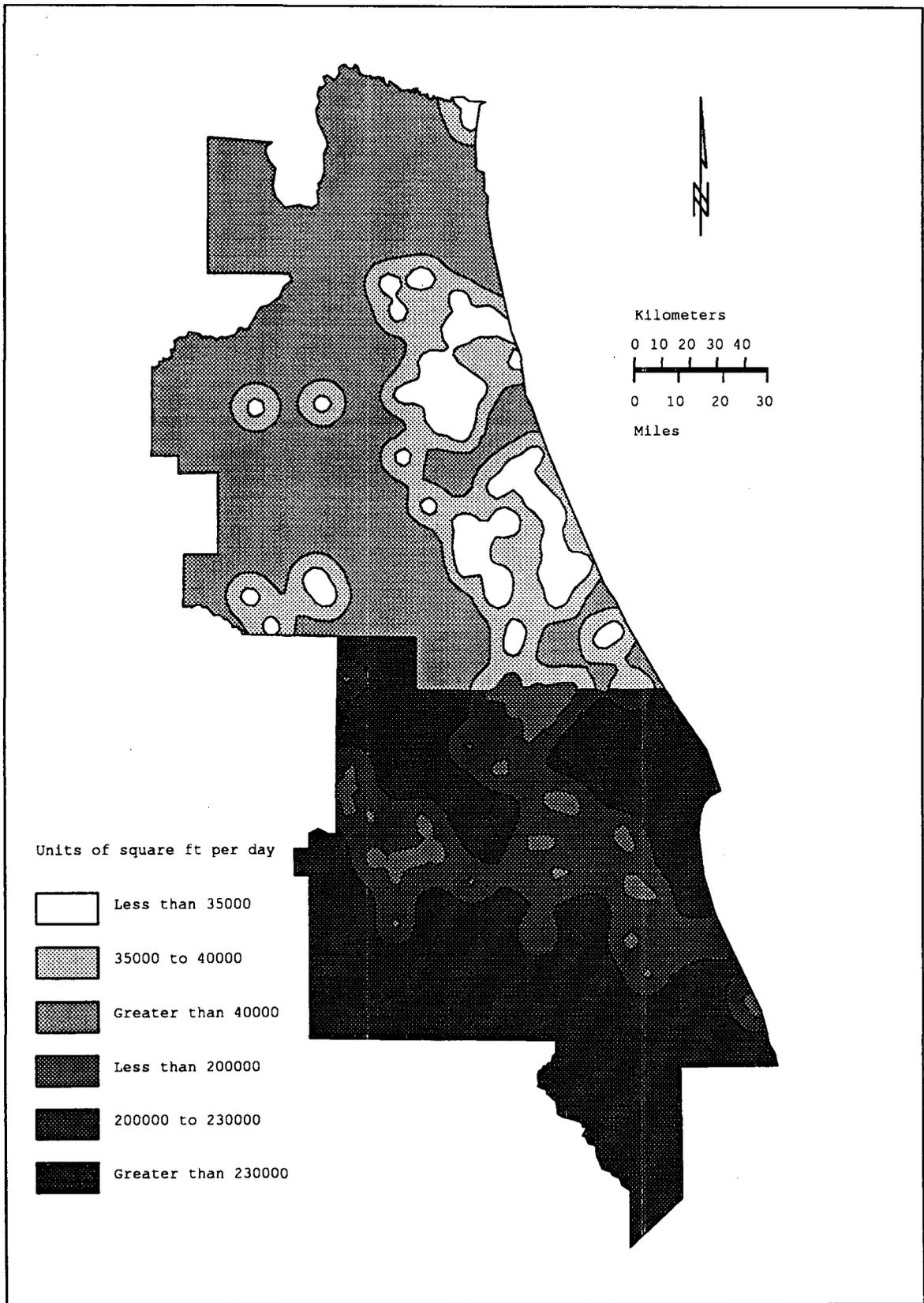


Figure 4.17 Standard deviations of transmissivity estimates based on the northern and southern subdistrict semivariograms.

At first glance, the estimated values created from the northern and southern semivariograms seem more representative of the observed transmissivity distribution within the District. However, this observation is purely subjective in light of the fact that the estimates were produced from only one realization of the random field. In addition, the northern and southern semivariograms exhibit larger fluctuations and are less reliable (based on number of data pairs) than the districtwide semivariogram. The southern semivariogram exhibits a "hump" which may indicate that the data should be detrended in that region to produce a more reliable semivariogram. Thus, it is felt that the districtwide semivariogram produces a more reliable estimation of the transmissivity distribution within the District.

Figure 4.18 shows model derived estimated transmissivities for the St. Johns River Water Management District (Bush and Johnston 1988). These transmissivity estimates were derived using a three-dimensional finite-difference computer model. The estimation of the transmissivities was a two-phase process described in detail by Bush and Johnston (1988). Essentially, transmissivity estimates were calibrated against the estimated steady-state predevelopment potentiometric surface of the Floridan aquifer and the estimated discharge to surface water bodies. The estimates were then adjusted by superimposing average 1980 pumping on the calibrated predevelopment model and then comparing the observed May 1980 potentiometric surface of the Upper Floridan aquifer with the modeled potentiometric surface.

The reliability of the model derived transmissivity estimates is often low due to the insensitivity of model-computed heads to transmissivity. Comparison of Figure 4.18 with Figure 4.16 shows that,

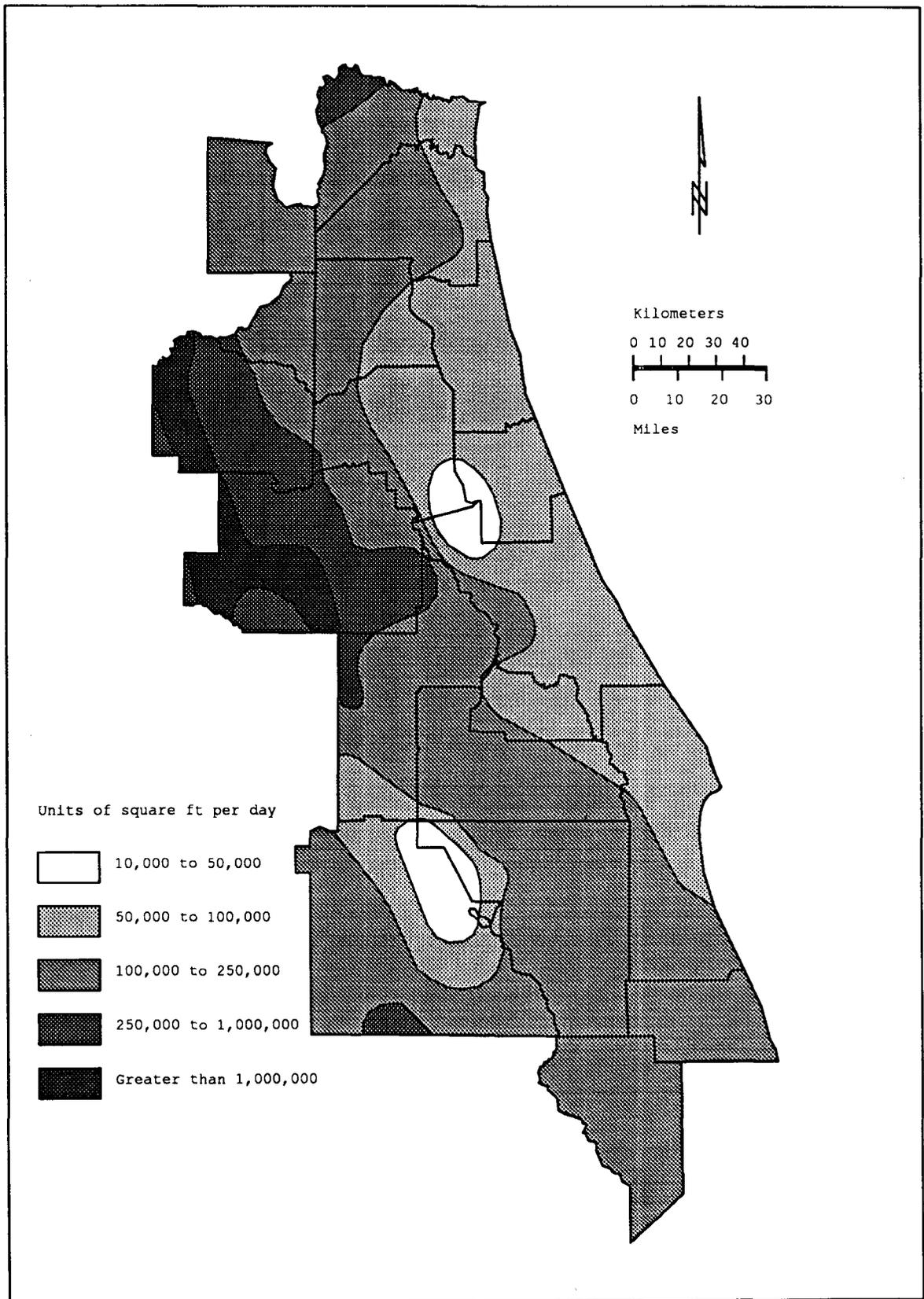


Figure 4.18 Model derived transmissivities for the Upper Floridan aquifer (after Bush and Johnston 1988).

in general, the model derived transmissivities are an order of magnitude greater than those estimated using field aquifer tests and geostatistical analysis. It is felt that geostatistical analysis provides a more objective and reliable method to estimate the transmissivity distribution within the District than calibration using groundwater models.

#### Estimation of Storativity in the Upper Floridan Aquifer

A statistical analysis of the storativity data for the Upper Floridan was performed. Like the transmissivity data, the storativity data was found to be skewed, cf. Figure 4.19. A log transformation was performed on the original data to create a data set with an

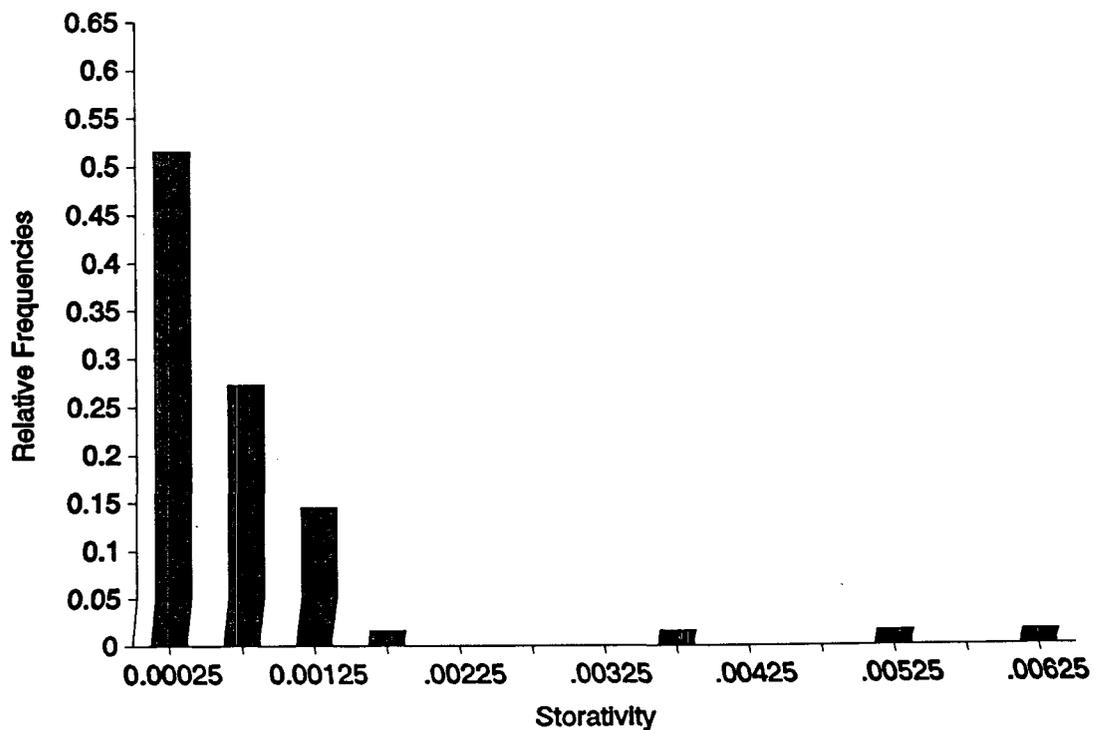


Figure 4.19 Histogram of storativity data for the Upper Floridan aquifer.

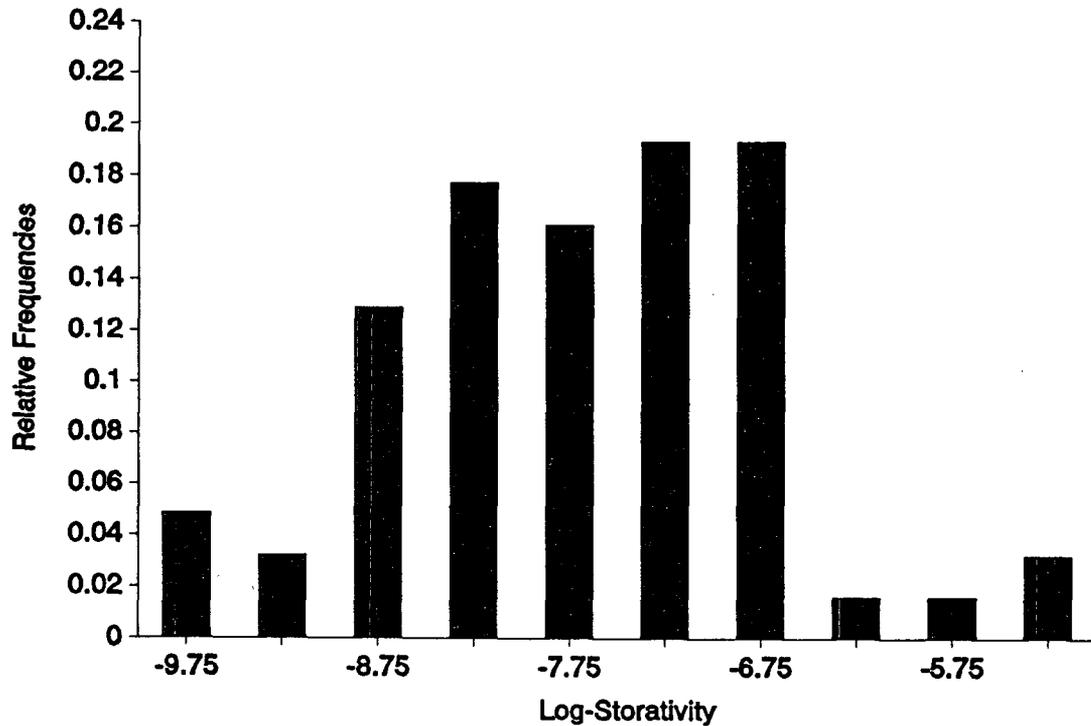


Figure 4.20 Histogram of log-transformed Upper Floridan storativity data.

approximately normal distribution, cf. Figure 4.20. Table 4.5 lists the sample statistics for both original and transformed storativity data.

After reviewing the available storativity data and its spatial distribution, it was determined that experimental semivariogram field of

Table 4.5 Sample statistics for Upper Floridan storativity data.

	Original Data	Transformed Data
Number of Data Values =	62	62
Mean Value =	0.00077	-7.67607
Median Value =	0.00048	-7.63484
Standard Deviation =	0.00106	0.97032
Variance =	0.00000	0.92634
Skewness =	3.57884	0.25273
Kurtosis =	16.78627	3.01702
Minimum Data Value =	0.00007	-9.56701
Maximum Data Value =	0.00620	-5.08320

study length should be set to  $L = 1,120,000$  ft, thus giving a limit of reliability of  $L/2 = 560,000$  ft. A distance class was chosen with  $h = 20,000$  ft and  $\varepsilon(h) = 10,000$  ft and the angle class increment was set to  $d\alpha = 90^\circ$ , thus producing an isotropic experimental semivariogram with 28 points. One of these experimental semivariogram points,  $\hat{\gamma}_{480000}$ , was rejected with an  $N'$  of less than 30 pairs. Figure 4.21 shows the isotropic experimental semivariogram computed for the log-transformed storativity data values for the Upper Floridan aquifer.

A cross-validated spherical model with a nugget effect was fitted to the storativity experimental semivariogram. The nugget constant for this model was  $\gamma_{00} = 0.55$ , with the sill being  $C_0 = 1.13$ , and the correlation parameter determined to be  $a = 511,000$  ft.

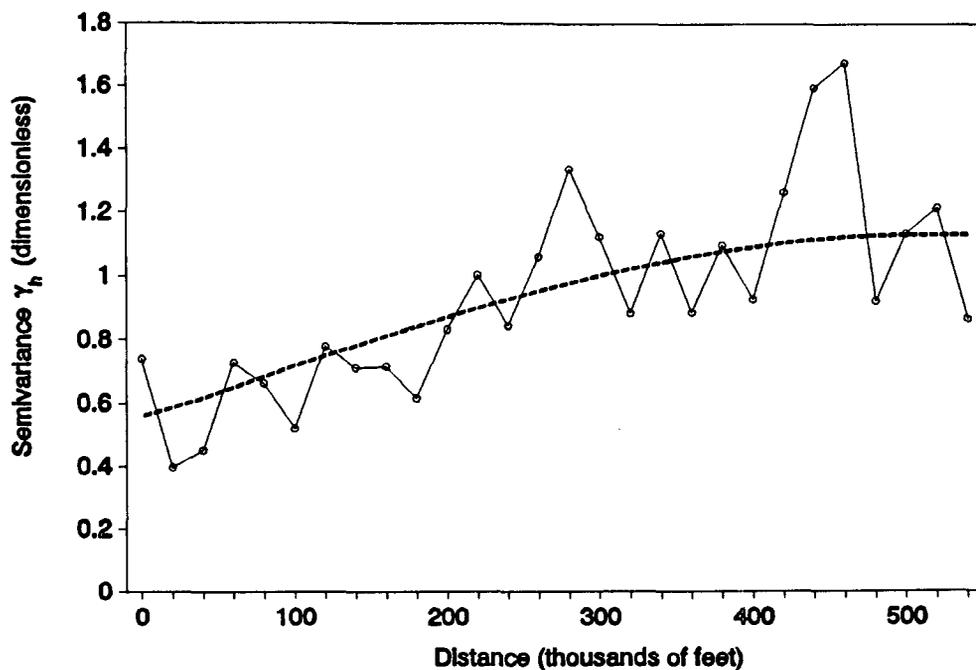


Figure 4.21 Spherical model fitted to a semivariogram of log transformed storativity values for the Upper Floridan aquifer.

The storativity kriging estimates, determined with this semivariogram model, were kriged over the same grid as the transmissivity estimates. Setting  $K_0 = 0.77$ , the kriged estimates were inversely transformed, along with the standard deviation values. Figures 4.22 and 4.23 show the resulting storativity estimation and standard deviation maps for the Upper Floridan aquifer.

It was decided to investigate the correlation of the storativity and transmissivity variables in the Upper Floridan aquifer. It was noticed that the storativity values tended to increase and decrease directly with the transmissivity values, although this relationship was in general only approximate. Therefore, it was decided that co-kriging the storativity values with the transmissivity values may improve storativity estimations, given that there was twice the available transmissivity data than available storativity data. A storativity-transmissivity experimental cross-semivariogram was developed using the same distance and angle classes used in the development of the storativity experimental semivariogram. None of the experimental semivariogram points were rejected, resulting in the experimental cross-semivariogram shown in Figure 4.24. A cross-validated spherical model with a nugget effect was fit, with  $\gamma_{00} = 0.12$ ,  $C_0 = 0.40$ , and  $a = 126,000$  ft.

Co-kriging was performed on the storativity and transmissivity data using the direct storativity and transmissivity semivariograms and their cross-semivariogram. Setting  $K_0 = 0.84$ , the co-kriged storativity estimates were inversely transformed, along with the standard deviation values. Figures 4.25 and 4.26 show the resulting storativity estimation

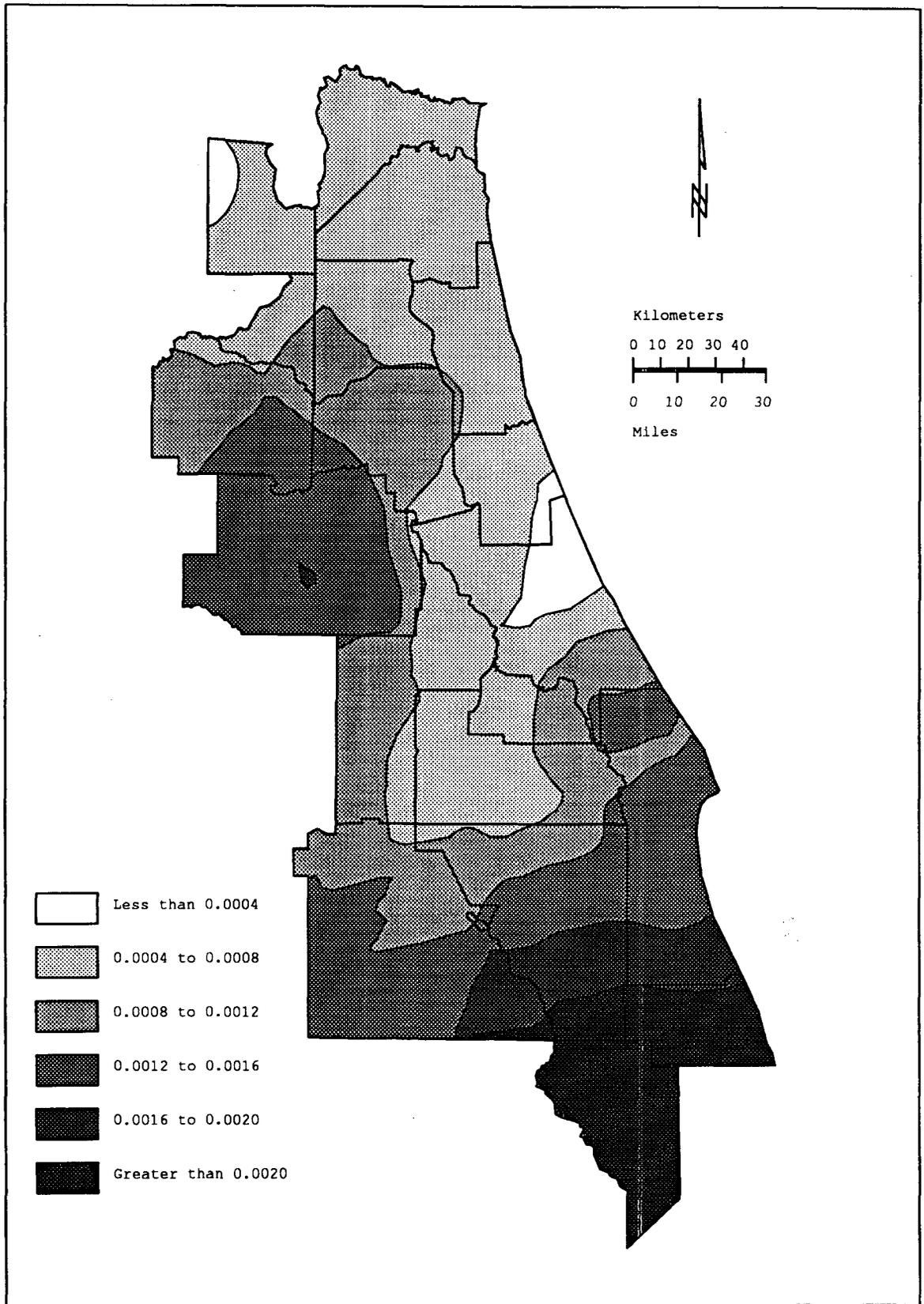


Figure 4.22 Storativity estimation for the Upper Floridan aquifer.

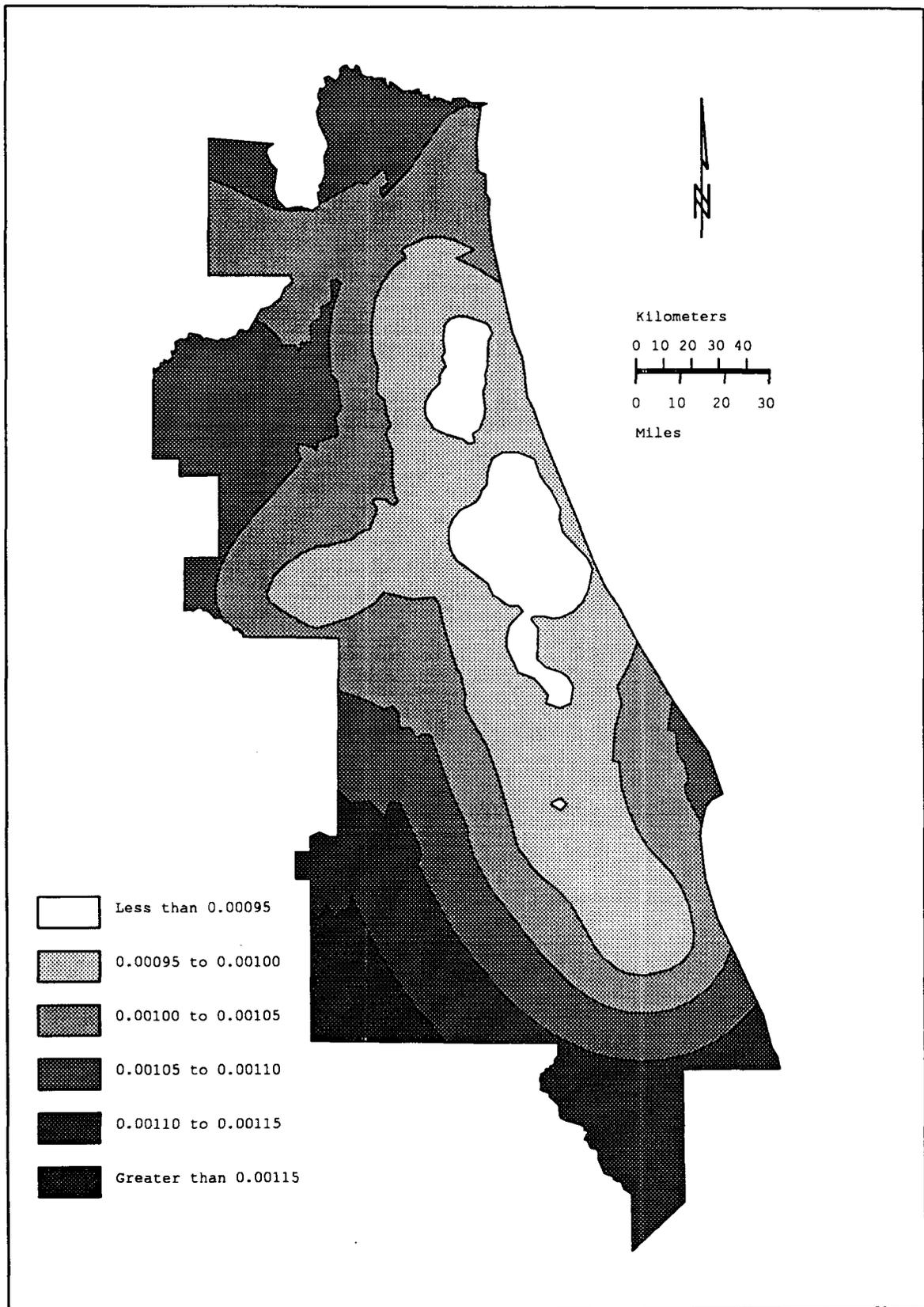


Figure 4.23 Standard deviations of storativity estimates.

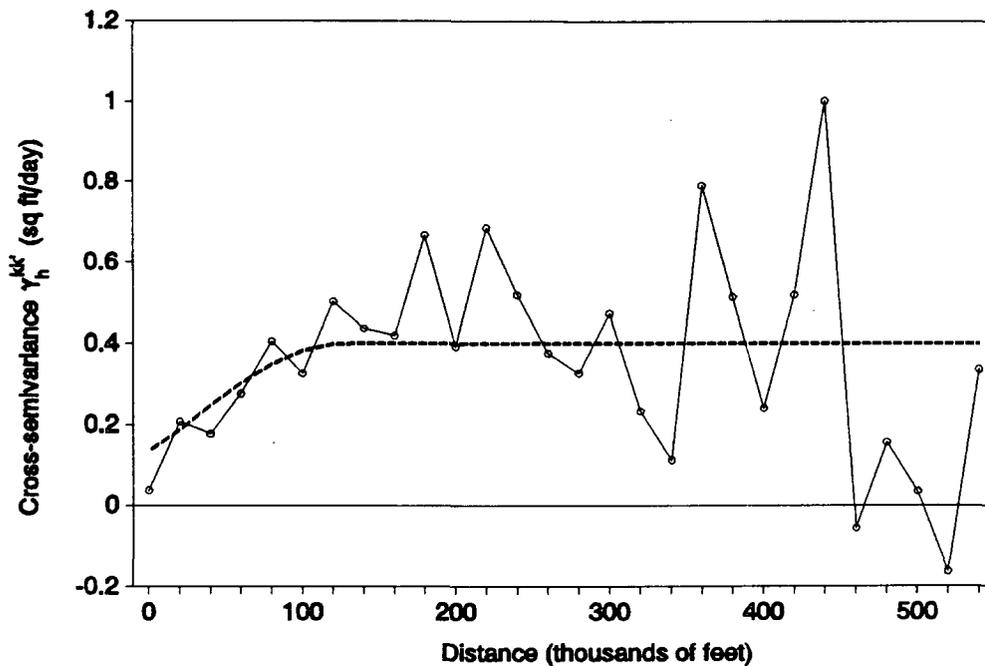


Figure 4.24 Spherical model fitted to a cross-semivariogram of log transformed storativity and transmissivity values for the Upper Floridan aquifer.

and standard deviation maps for the Upper Floridan aquifer based on storativity-transmissivity coregionalization.

Comparison of Figures 4.22 and 4.25 shows that the co-kriged estimates exhibit more variability than the estimates kriged directly from the storativity data. This is a result of the variability of the transmissivity data field. Inspection of Figures 4.23 and 4.26 shows that, in general, co-kriging did not improve the estimation standard deviations over direct kriging. This indicates a lack of significant correlation between storativity and transmissivity values. Thus, little information about storativity in the Floridan aquifer can be derived from transmissivity data.

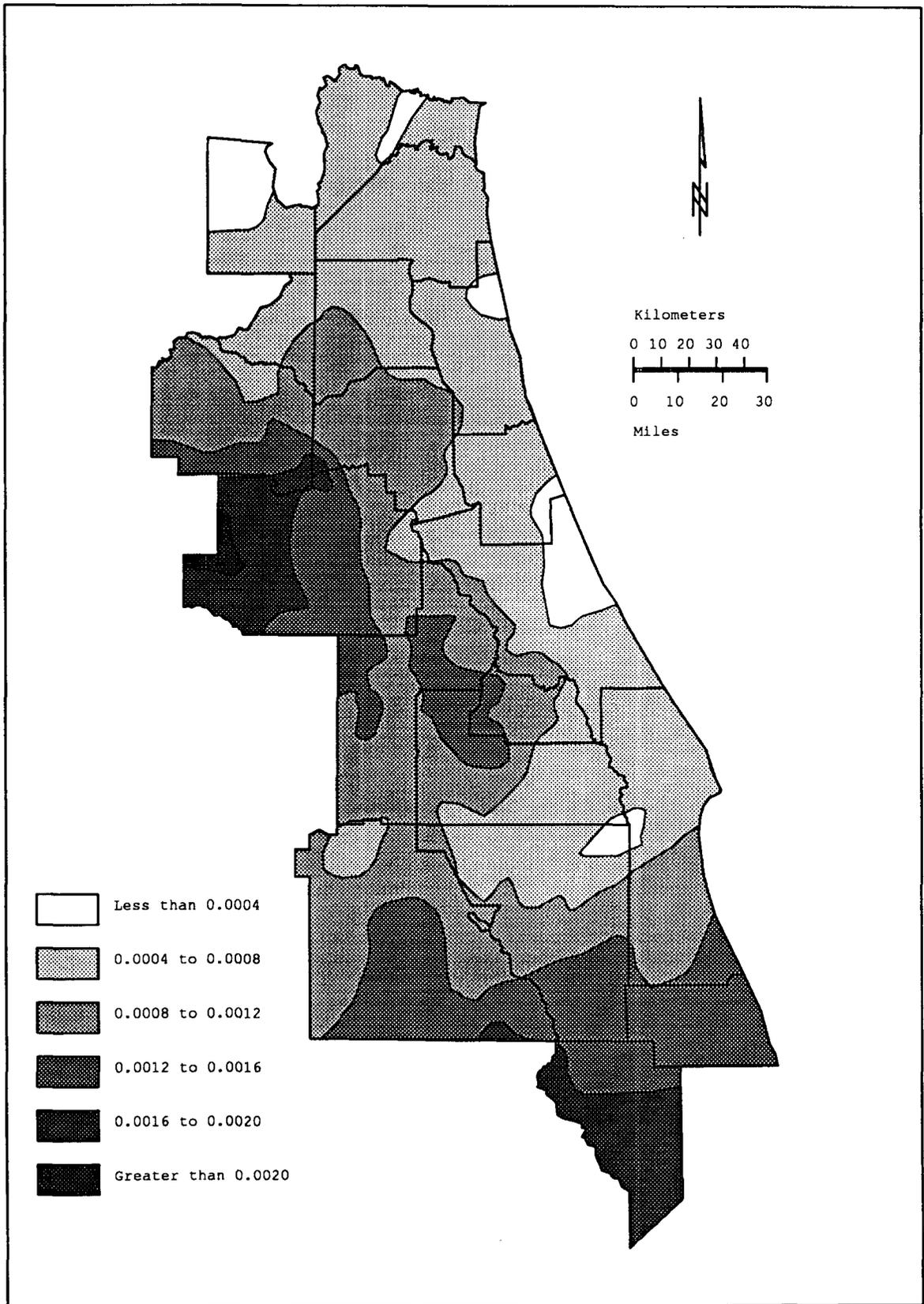


Figure 4.25 Storativity estimation for the Upper Floridan aquifer based on storativity-transmissivity co-kriging.  
IV-62

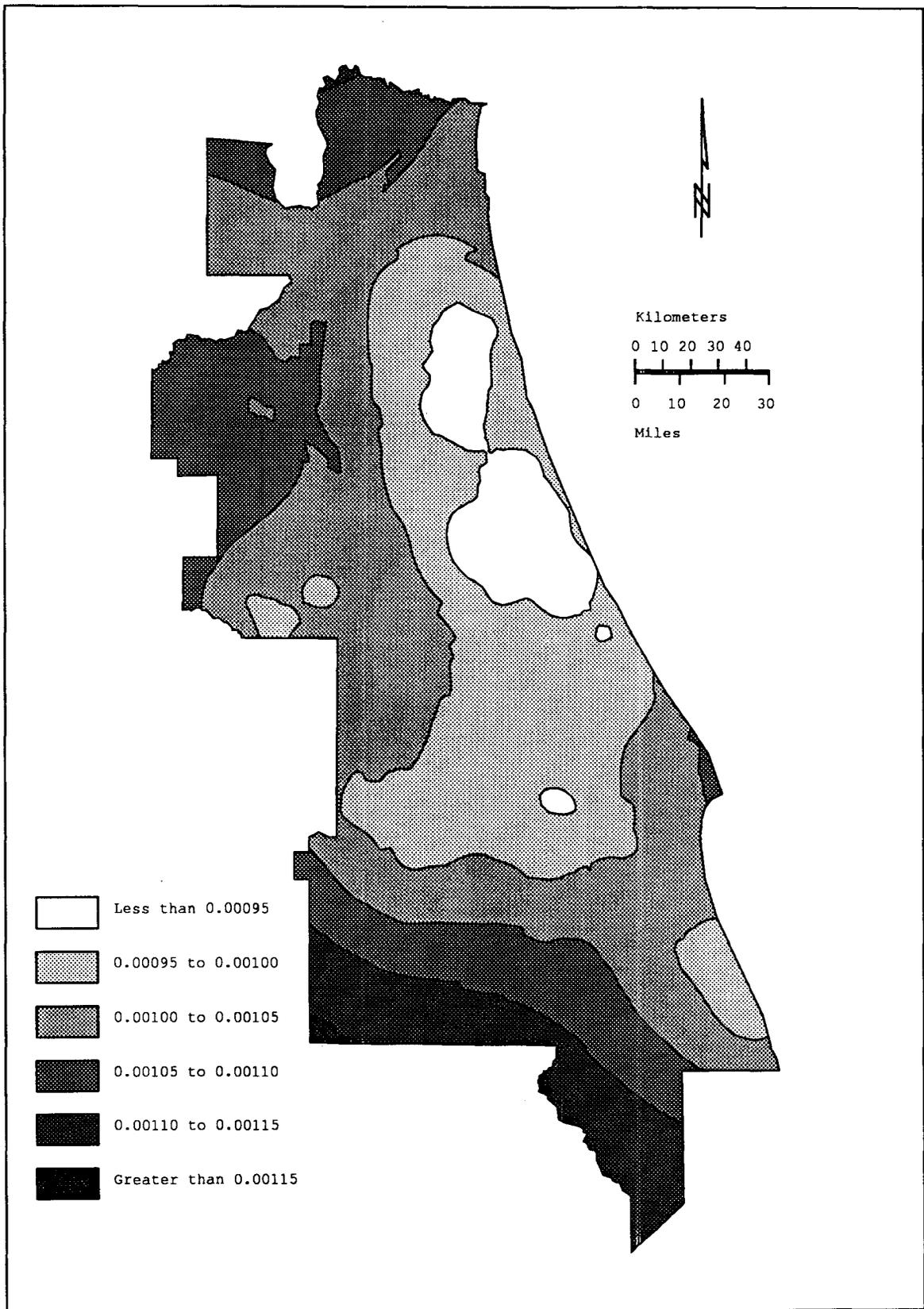


Figure 4.26 Standard deviations of storativity estimates based on storativity-transmissivity co-kriging.

## Conclusions

In this Chapter, geostatistical analysis of transmissivities and storativities within the District produced spatial correlation structure models for each variable. Knowledge of these models allowed the best linear unbiased estimates of transmissivities and storativities, and the associated error in those estimates, to be produced throughout the District. Maps were generated showing the estimated spatial distribution of transmissivity and storativity, along with maps of the standard deviations of these estimates. These maps should provide valuable information for both defining physically-based deterministic and/or stochastic groundwater model input parameters and for designing field studies to gather additional data.

Transmissivities and storativities were found to be highly variable, and in areas with few measurements higher estimation variances indicate greater uncertainty. Thus, more measurements should be taken in regions where estimates have the greatest uncertainty. A comparison of transmissivity estimates obtained through kriging versus estimates obtained through regional flow modeling shows little resemblance between the two. However, the accuracy of the flow model estimates are uncertain since they were obtained by trial and error calibration of a regional flow model to estimated steady-state predevelopment conditions.

## CHAPTER V TIME SERIES ANALYSIS OF PIEZOMETRIC HEAD

### Introduction

Data that are obtained from sequential observations of a phenomenon over time, such as monthly piezometric head fluctuations or daily precipitation, are called time series. Time series analysis involves building, identifying, fitting and checking stochastic models for such data series (Box and Jenkins 1976). Analysis of a time series is helpful in understanding the mechanisms that give rise to an observed process. Time series analysis is also useful for forecasting future events based on past behavior. In hydrogeology, time series analysis can provide a basis for (1) forecasting hydrological events such as precipitation, streamflow, and groundwater level in time, (2) optimization and control of the use of water resources.

### Theory

#### Stochastic Processes

It was shown in Chapter IV that spatially variable hydrogeologic parameters can be characterized as random fields, and that the behavior of these random fields can be described by their first- and second-order spatial moments, i.e., the spatial mean, variance, covariance, and variogram. Likewise, temporally variable hydrogeologic parameters can be characterized as random processes that may vary over both space and

time. The temporal variability of the random process at a particular location can be described by its first- and second-order temporal moments, i.e., the temporal mean and variance, as well as the temporal covariance function. The temporal moments are completely analogous to their spatial counterparts defined in Chapter IV.

Consider a series of observations  $[z_1, z_2, \dots, z_t, \dots, z_n]$  of the random process  $Z_t$  taken at a certain point  $x_0$  at times  $[t_1, t_2, \dots, t_k, \dots, t_n]$ , cf. Figure 5.1. These observations constitute a discrete time series. For a discrete series,  $\tau_0$  is defined as the time origin,  $\Delta t$  as the incremental time unit, and  $z_k$  as the observation at the  $k$ th time increment  $t_k = \tau_0 + k\Delta t$ .

If a time series can be exactly described by a mathematical function, the time series is said to be deterministic. Future values of a deterministic time series can be exactly determined without error. A time series which cannot be exactly described by a mathematical

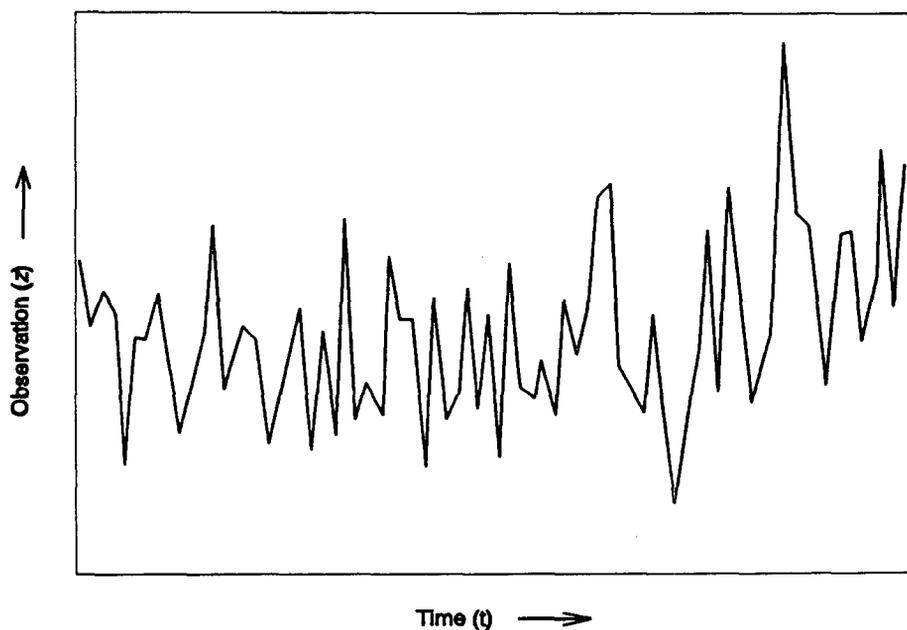


Figure 5.1 An example of a discrete time series.

function may be considered to be a stochastic time series. Its future values can only be described in terms of a probability distribution and/or moments of this distribution. Due to the vast scale and complexity of the physical mechanisms which govern hydrogeologic phenomena, stochastic time series models are often useful for estimating future values and behavior of such phenomena.

A time series of random observations  $[z_1, z_2, \dots, z_n]$  is only one particular realization of the series of random variables  $[Z_1, Z_2, \dots, Z_n]$  generated by the random process  $Z_t$  at a particular location  $x_0$ . As discussed in Chapter IV for random fields, the complete probabilistic structure of a random process can only be determined from the infinite ensemble of distributions of all possible series generated by the random process. In practice, however, only one set of observations is generally available. Therefore, we must assume the random process to be stationary in time and that its probability structure can be estimated from the statistical properties of the one realization that is available.

To describe the past behavior of a stochastic process and to predict future behavior, a probability or stochastic model is developed. A stochastic model can be used to calculate the probability of a future value lying between two specified limits, or confidence intervals. Several established classes of stochastic models have been developed to describe the behavior of stochastic time series. Various analytical tools are utilized to determine which class of stochastic model should be employed to describe a process. Some of these tools are the mean and variance of the process, the autocovariance function, the autocorrelation function, the partial autocorrelation function, and the

spectrum. This chapter will discuss these tools and the general classes of stochastic models.

### Moments of Stationary Stochastic Processes

As defined in Chapter IV, a process which is in a particular state of statistical equilibrium is called a stationary process. Strict stationarity for a temporally variable random process implies that all joint probability density functions and associated first- and second-order moments of the process are unaffected by a change of time origin  $\tau_0$ . Thus, a stationary stochastic process is said to have a temporally constant expected value, or mean,  $\bar{Z}$  which defines the level about which it fluctuates in time:

$$E\{Z_t\} = \int_{-\infty}^{\infty} Z_t f_z(Z_t) dZ_t = \bar{Z}, \quad (5.1)$$

where  $f_z(Z_t)$  is the process probability function. The temporally constant variance of the process is defined by

$$\sigma_z^2 = E\{(Z_t - \bar{Z}_t)^2\} = \int_{-\infty}^{\infty} (Z_t - \bar{Z}_t)^2 f_z(Z_t) dZ_t, \quad (5.2)$$

which measures the spread of the time series about its mean.

A random process is said to exhibit a stationary mean if the mean does not vary with location, i.e.,  $E\{Z_t\} = \bar{Z}_t = \bar{Z}$ . Since stationarity implies that the probability distribution  $f_z(Z_t)$  is the same for all times  $t$ , its shape can be inferred from a histogram of the observed time series  $[z_1, z_2, \dots, z_n]$ , which is just one realization of the random

process. Additionally, the mean of the stochastic process can be estimated from the mean of the observed time series

$$\hat{\bar{Z}} = \frac{1}{n} \sum_{t=1}^n z_t, \quad (5.3)$$

and the variance of the process can be estimated from the variance of the time series

$$\hat{\sigma}_z^2 = \frac{1}{n-1} \sum_{t=1}^n (z_t - \hat{\bar{Z}})^2. \quad (5.4)$$

Under the stationarity assumption, the joint probability distribution  $f_{z_1 z_2}(Z_{t_1}, Z_{t_2})$  is implied to be the same for all times  $t_1, t_2$ , separated by a constant time interval  $k$ , or lag  $k$ . The nature of this joint distribution can be inferred from the covariance between  $z_t$  and  $z_{t+k}$ , called the autocovariance or covariance  $C_k$  at lag  $k$ , defined as

$$\begin{aligned} C_k &= E\{(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z_{t_1} - \bar{Z}_{t_1})(Z_{t_2} - \bar{Z}_{t_2}) f_{z_1 z_2}(Z_{t_1}, Z_{t_2}) dZ_{t_1} dZ_{t_2}. \end{aligned} \quad (5.5)$$

Similarly, the autocorrelation  $\rho_k$  at lag  $k$  measures the correlation between  $Z_t$  and  $Z_{t+k}$ , and is defined by

$$\rho_k = \frac{E\{(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})\}}{\sqrt{E\{(Z_t - \bar{Z})^2\}E\{(Z_{t+k} - \bar{Z})^2\}}} = \frac{C_k}{C_0}, \quad (5.6)$$

where

$$C_0 = E\{(Z_t - \bar{Z})(Z_t - \bar{Z})\} = E\{(Z_t - \bar{Z})^2\} = \sigma_z^2 . \quad (5.7)$$

It becomes obvious that the autocovariance  $C_k$  is related to the autocorrelation  $\rho_k$  by the following relationship:

$$C_k = \sigma_z^2 \rho_k = C_0 \rho_k . \quad (5.8)$$

With  $n$  successive observations  $[z_1, z_2, \dots, z_n]$  of a stationary process, the relationship expressed in equation (5.8) can be rewritten with the autocovariance and autocorrelation matrices:

$$\begin{bmatrix} C_0 & C_1 & C_2 & \cdot & \cdot & \cdot & C_{n-1} \\ C_1 & C_0 & C_1 & \cdot & \cdot & \cdot & C_{n-2} \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ C_{n-1} & C_{n-2} & C_{n-3} & \cdot & \cdot & \cdot & C_0 \end{bmatrix} = \sigma_z^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{n-2} \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdot & \cdot & \cdot & 1 \end{bmatrix} .$$

Both the autocovariance and autocorrelation matrices are symmetric with constant elements on any diagonal and are positive-definite.

The following properties result from the definitions of the autocovariance and autocorrelation coefficients:

$$\begin{aligned} C_0 = \sigma_z^2 &\Rightarrow \rho_0 = 1 \\ C_k = C_{-k} &\Rightarrow \rho_k = \rho_{-k} \\ |C_k| \leq \sigma_z^2 &\Rightarrow |\rho_k| \leq 1 \end{aligned} \quad (5.9)$$

A strong correlation is indicated by values of  $\rho_k$  near  $\pm 1$ , while a weak correlation is indicated by values of  $\rho_k$  near 0. If  $Z_t$  and  $Z_{t+k}$  are independent and uncorrelated, then  $\rho_k = 0$ .

The correlation, or linear dependence, between two values  $Z_t$  and  $Z_{t+k}$  a distance  $k$  apart is measured by the autocorrelation coefficient  $\rho_k$  at lag  $k$ . This correlation can also be inferred from the autocovariance coefficient  $C_k$  at lag  $k$ . The plot of  $\rho_k$  versus the lag  $k$  is called the autocorrelation function, or correlation function, of the stochastic process. The autocovariance function, or covariance function, is similarly defined as the plot of  $C_k$  versus the lag  $k$ . It should be noted that given the relationship  $C_k = \rho_k \sigma_z^2$ , the autocorrelation coefficient  $\rho_k$  is dimensionless by definition. Thus the autocorrelation function is independent of the unit scale of the time series, making it a very useful analytical tool.

Figure 5.2 shows the autocorrelation function as it relates to the diagonals of the autocorrelation matrix. The autocorrelation function is symmetric about zero with  $\rho_k = \rho_{-k}$ , thus in practice only the plot of the positive half of the autocorrelation function is necessary.

If the probability distribution of a process is a multivariate normal distribution, the process is called a Gaussian process. A Gaussian stationary process, then, is completely characterized by its mean  $\bar{Z}$ , variance  $\sigma_z^2$ , and autocorrelation function  $\rho_k$ , or equally by its mean  $\bar{Z}$  and autocovariance function  $C_k$ . The autocorrelation function (or the autocovariance function) is essential in identifying the model which generates the stochastic process.

As the theoretical autocorrelation function describes a conceptual process, it can only be estimated from the experimental autocorrelation

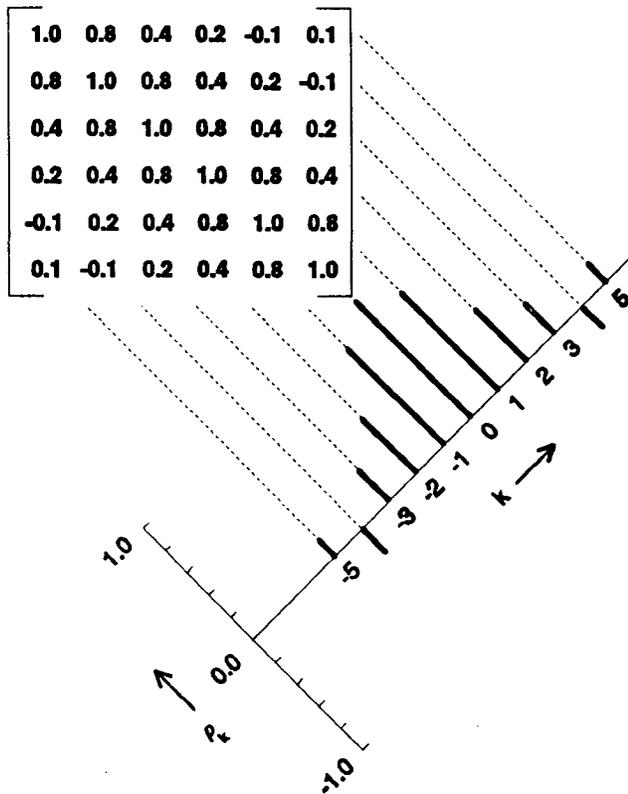


Figure 5.2 An autocorrelation matrix and the resulting autocorrelation function (after Box and Jenkins 1976).

function. Consider the finite time series  $[z_1, z_2, \dots, z_n]$  of  $n$  observations. The estimate of the  $k$ th lag autocorrelation  $\rho_k$  is

$$\hat{\rho}_k = \frac{\hat{C}_k}{\hat{C}_0}, \quad (5.10)$$

where

$$\hat{C}_k = \frac{1}{n-1} \sum_{t=1}^{n-k} (z_t - \hat{\bar{z}})(z_{t+k} - \hat{\bar{z}}), \quad k = 0, 1, 2, \dots, K \quad (5.11)$$

is the estimate of the autocovariance  $C_k$  and  $\hat{C}_0 = \hat{\sigma}_z^2$  as defined in

equation (5.4). To obtain a useful estimate of the autocorrelation function, it is necessary to have at least thirty to fifty pairs of observations separated by lag  $k$ .

The experimental correlation values estimated with equation (5.10) will differ somewhat from the theoretical correlation values. A measure of this difference is the variance of the estimated autocorrelation coefficient given by Box and Jenkins (1976):

$$\text{VAR}\{\hat{\rho}_k\} = \frac{1}{n} \sum_{v=-\infty}^{+\infty} \left\{ \rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2 \right\}. \quad (5.12)$$

Given a process for which the autocorrelation coefficients  $\rho_k$  are effectively zero when  $k > q$ , the variance approximation simplifies to

$$\text{VAR}\{\hat{\rho}_k\} = \frac{1}{n} \left\{ 1 + 2 \sum_{v=1}^q \rho_v^2 \right\}, \quad k > q. \quad (5.13)$$

To compute  $\text{VAR}\{\hat{\rho}_k\}$  for lags greater than  $q$ , the estimated autocorrelations  $\hat{\rho}_k$  for  $k = 1, 2, \dots, q$  are substituted for the theoretical autocorrelations  $\rho_k$  in equation (5.13). The square root of (5.13) approximates the standard deviation of  $\hat{\rho}_k$  and is termed the large-lag standard error of the autocorrelation estimate.

The correlation between two values  $Z_t$  and  $Z_{t+k}$  is sometimes influenced by the correlation between the values  $Z_t$  and  $Z_{t+j}$ , where  $j = 1, 2, \dots, k-1$ . The autocorrelation coefficient  $\rho_k$ , consequently, can then be expressed as a weighted linear combination of the  $j$  previous autocorrelation coefficients:

$$\rho_k = \sum_{i=1}^k \phi_{ki} \rho_{k-i} \quad k > 0, \quad (5.14)$$

where  $\phi_{kk}$  is called the partial autocorrelation coefficient. The other coefficients  $\phi_{kj}$ , where  $j = 1, 2, \dots, k-1$ , are used only to calculate the partial autocorrelation coefficients and are disregarded. Equation (5.14) leads to a set of linear equations called the Yule-Walker equations (Box and Jenkins 1976). Since both  $\rho_k$  and  $\rho_{k-1}$  are known, then  $\phi_{ki}$  can be determined from this set of linear equations.

Once the partial autocorrelation coefficients  $\phi_{kk}$  have been determined for  $k$  lags, the plot of the resulting matrix

$$[\phi_{kk}] = \begin{bmatrix} \phi_{11} & \phi_{22} & \phi_{33} & \cdot & \cdot & \cdot & \phi_{nn} \\ \phi_{22} & \phi_{11} & \phi_{22} & \cdot & \cdot & \cdot & \phi_{n-1n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{nn} & \phi_{n-1n-1} & \phi_{n-2n-2} & \cdot & \cdot & \cdot & \phi_{11} \end{bmatrix}$$

results in the partial autocorrelation function. Analogous to the autocorrelation function, the partial autocorrelation function is symmetric about zero with  $\phi_{kk} = \phi_{-k-k}$ .

To estimate the theoretical partial autocorrelation function, the experimental partial autocorrelation coefficients  $\hat{\phi}_{kk}$  are computed, with equation (5.14) becoming

$$\hat{\rho}_k = \sum_{i=1}^k \hat{\phi}_{ki} \hat{\rho}_{k-i} \quad k > 0, \quad (5.15)$$

which can be solved for  $\hat{\phi}_{ki}$  given estimates  $\hat{\rho}_k$ . For a useful estimate of the partial autocorrelation function, it is necessary to have at least 50 observations and the partial autocorrelations  $\hat{\phi}_{kk}$  computed for  $k = 1, 2, \dots, K$  where  $K$  is not larger than  $n/4$ . Like the autocorrelation function, the partial autocorrelation function is useful in determining the model which describes the stochastic generating mechanism of the process.

Box and Jenkins (1976) state that it has been shown that the estimated partial autocorrelations are approximately independently distributed. Given a time series of  $n$  observations for which the partial autocorrelation coefficients  $\phi_{kk}$  are effectively zero when  $k > p$ , the variance of  $\phi_{kk}$  can then be estimated as

$$\text{VAR}\{\hat{\phi}_{kk}\} = \frac{1}{n}, \quad k > p. \quad (5.16)$$

Thus, the standard error of the partial autocorrelation estimate  $\hat{\phi}_{kk}$  is just the square root of (5.16).

### The Spectral Density Function and Spectrum

Another way of analyzing a time series is based on the assumption that its behavior can be described with different frequency properties. The spectral density function, then, is a tool used in the analysis of the frequency domain of the stochastic process. In order to understand the spectral density function, the frequency of the periodic variation of a time series must first be considered.

Consider a zero mean time series which contains a periodic component at a known frequency. To describe this periodicity, the

following model is employed:

$$Z_t = R \cos(\omega t + \theta) + e_t, \quad (5.17)$$

where  $\omega$  is called the frequency of the periodic variation,  $R$  is called the amplitude,  $\theta$  is called the phase, and  $e_t$  is denotes a stationary random residual (Chatfield 1989). This model is a simple model describing the behavior of a time series containing only one periodic frequency. To describe a stochastic process displaying variations caused by  $k$  different frequencies, (5.17) can be generalized to

$$Z_t = \sum_{j=1}^k R_j \cos(\omega_j t + \theta_j) + e_t, \quad (5.18)$$

where  $R_j$  is the amplitude at frequency  $\omega_j$ .

Recalling the mathematical relationship of  $\cos(\omega t + \theta) = \cos \omega t \cos \theta - \sin \omega t \sin \theta$ , (5.18) can be rewritten as

$$Z_t = \sum_{j=1}^k (a_j \cos \omega_j t + b_j \sin \omega_j t) + e_t, \quad (5.19)$$

where  $a_j = R_j \cos \theta_j$  and  $b_j = -R_j \sin \theta_j$ . When the process does not have a zero mean, the mean  $m$  is added to the right side of (5.19).

When  $k \rightarrow \infty$  for any discrete stationary process with zero mean measured at unit intervals, (5.19) becomes

$$Z_t = \int_0^{\pi} \cos \omega t d u(\omega) + \int_0^{\pi} \sin \omega t d v(\omega), \quad (5.20)$$

where  $u(\omega)$  and  $v(\omega)$  are uncorrelated continuous processes which are defined for all  $\omega$  in the range  $(0, \pi)$  (Chatfield 1989). Although  $k$  approaches infinity, the integrals are taken from 0 to  $\pi$  due to the periodic repetition of frequencies about  $\pi$ .

Chatfield (1989) states that to since  $u(\omega)$  and  $v(\omega)$  in (5.20) have no direct practical interest, the spectral density function  $F(\omega)$  is introduced. The spectral density function is related to  $u(\omega)$  and  $v(\omega)$  and arises from the Wiener-Khinchine theorem which states that for any stationary stochastic process with autocovariance function  $C_k$ , there exists a monotonically increasing function  $F(\omega)$  such that

$$C_k = \int_0^{\pi} \cos \omega k dF(\omega) . \quad (5.21)$$

For a unit time interval, the highest possible frequency is a multiple of  $\pi$  and so all variation is accounted for by frequencies between 0 and  $\pi$ . Thus the variance of the stochastic process is equal to

$$\sigma_z^2 = F(\pi) = C_0 = \int_0^{\pi} dF(\omega) . \quad (5.22)$$

Differentiating  $F(\pi)$  leads to

$$C_k = \int_0^{\pi} \cos \omega k f(\omega) d(\omega) , \quad (5.23)$$

where

$$f(\omega) = \frac{dF(\omega)}{d\omega} . \quad (5.24)$$

The function  $f(\omega)$  is called the power spectral density function, or spectrum for short. It becomes obvious that the variance of the process is related to  $f(\omega)$  by

$$\sigma_z^2 = \int_0^{\pi} f(\omega) d\omega . \quad (5.25)$$

The spectrum  $f(\omega)$  represents the contribution to the process variance of components with frequencies in the  $(\omega, \omega+d\omega)$  range. A peak in the plot of the spectrum, then, will indicate an important contribution to the variance at that frequency. The total area under the plot of  $f(\omega)$  is equal to the variance of the process.

To determine the spectrum from the autocovariance function  $C_k$ , the inverse of (5.23) is taken:

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} C_k e^{-i\omega k} , \quad (5.26)$$

so that  $f(\omega)$  is the Fourier transform of the autocovariance function. Equivalently, since  $C_k = C_{-k}$ , (5.26) can be written as

$$f(\omega) = \frac{1}{\pi} \left[ C_0 + 2 \sum_{k=1}^{\infty} C_k \cos \omega k \right] . \quad (5.27)$$

The spectrum can be estimated from the experimental autocovariance function by replacing the theoretical autocovariances  $C_k$  with the estimated autocovariances  $\hat{C}_k$ . Further, the spectrum can be determined from the experimental autocorrelation function by making the substitution  $\hat{C}_k = \hat{C}_0 \hat{\rho}_k$ , resulting in the experimental spectrum  $\hat{f}(\omega)$ :

$$\hat{f}(\omega) = \frac{1}{\pi} \left[ 1 + 2 \sum_{k=1}^{\infty} \hat{p}_k \cos \omega k \right]. \quad (5.28)$$

The autocorrelation function (or autocovariance function) and the spectrum are equivalent ways of describing a stationary stochastic process and are complimentary to each other. While the autocorrelation function expresses the time-domain behavior of the process, the spectrum expresses the frequency-domain behavior. Like the autocorrelation function, the spectrum is useful in identifying the model which drives the generating mechanism of the process.

### General Modeling Basics

Before discussing particular stochastic process models, it is necessary to define a few modeling basics. Stochastic models can be generally defined as either stationary or nonstationary. Stationary models assume that the process remains in statistical equilibrium about a constant mean level. The covariance of any two values  $Z_t$  and  $Z_{t+k}$  of a process described by a stationary model depends only on the time lag  $k$  and not on the actual times  $t$  and  $t+k$  (Cryer 1986). Nonstationary models assume that the statistical properties of the process are time dependent.

In the process of defining stochastic models, a couple of simple operators shall be employed. The backward shift operator  $B$  is defined by  $BZ_t = Z_{t-1}$ ; thus  $B^m Z_t = Z_{t-m}$ . The backward shift operator is linear and it is easy to see that for constants  $b$  and  $c$

$$B(bZ_t + ca_t) = b[BZ_t] + c[Ba_t] = bZ_{t-1} + ca_{t-1}. \quad (5.29)$$

Another important operator is the backward difference operator  $\nabla$  defined as

$$\nabla Z_t = Z_t - Z_{t-1} = (1 - B)Z_t. \quad (5.30)$$

The backward difference operator is useful in detrending a process, i.e., changing a nonstationary process into a stationary process of zero mean.

An important concept in stochastic modeling is the white noise process. A series of random variables  $a_t, a_{t-1}, a_{t-2}, \dots$  from a white noise process is normally distributed and has zero mean and a variance  $\sigma_a^2$ . White noise is purely random and exhibits no correlation between values. A frequency analysis of white noise shows that all frequencies contribute equally to the variance  $\sigma_z^2$  of the process (Cryer 1986).

### The Autoregressive Process Model

A stochastic model very useful in describing certain common series is the autoregressive model. In the autoregressive model, the current value of the process is described by a finite linear combination of previous values of the process and a white noise residual  $a_t$ . Consider a stationary gaussian process represented by  $Z_t, Z_{t-1}, Z_{t-2}, \dots$  at equally spaced times  $t, t-1, t-2, \dots$ . Defining  $\bar{Z}_t = Z_t - m$  as the deviations of the process about its mean, then:

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2} + \dots + \phi_p \bar{Z}_{t-p} + a_t \quad (5.31)$$

is called an autoregressive process of order  $p$ , or an  $AR(p)$  process;

termed autoregressive because the process is regressed on previous values of itself. The weight coefficients  $\phi_1, \phi_2, \dots, \phi_p$ , are finite in number and can be defined by the autoregressive operator  $\phi(B)$  of order  $p$  by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p . \quad (5.32)$$

Thus, the autoregressive model can be written as

$$\phi(B)\bar{Z}_t = a_t , \quad (5.33a)$$

or

$$\bar{Z}_t = \phi^{-1}(B)a_t . \quad (5.33b)$$

The AR( $p$ ) model has  $p+2$  parameters  $\bar{Z}$ ,  $\phi_1, \phi_2, \dots, \phi_p$ , and  $\sigma_a^2$  which must be estimated from the time series data.

To determine the values for  $\phi_1, \phi_2, \dots, \phi_p$ , it is enough to multiply (5.31) by  $\bar{Z}_{t-k}$  to obtain

$$\bar{Z}_{t-k}\bar{Z}_t = \phi_1\bar{Z}_{t-k}\bar{Z}_{t-1} + \phi_2\bar{Z}_{t-k}\bar{Z}_{t-2} + \dots + \phi_p\bar{Z}_{t-k}\bar{Z}_{t-p} + \bar{Z}_{t-k}a_t . \quad (5.34)$$

By taking the expected values of (5.34), and noting that  $a_t$  is independent of  $\bar{Z}_{t-k}$  (Cryer 1986), the autoregressive operator coefficients  $\phi_k$  can be expressed in terms of the autocovariance coefficients  $C_k$ :

$$C_k = \phi_1 C_{k-1} + \phi_2 C_{k-2} + \dots + \phi_p C_{k-p} \quad k > 0. \quad (5.35)$$

On then dividing (5.35) by  $C_0$ , the autocorrelation function  $\rho_k$  satisfies the difference equation:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k > 0. \quad (5.36)$$

which is analogous to the original (5.31) for the process  $\bar{Z}_t$ .

By substituting  $k = 1, 2, \dots, p$  in (5.36), a set of linear Yule-Walker equations for  $\phi_1, \phi_2, \dots, \phi_p$ , in terms of the autocorrelation function are obtained which can be written in matrix form as

$$\begin{bmatrix} \rho_1 & \rho_2 & \dots & \rho_p \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & & & & \\ \vdots & & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} .$$

from which the  $\phi_k$  are easily determined. Estimates for the autocorrelation parameters can be obtained by substituting estimates  $\hat{\rho}_k$  for  $\rho_k$  in (5.36) and solving the resulting equations.

Of particular importance are the first- and second- order autoregressive processes. The first-order autoregressive process, AR(1), is

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + a_t \quad (5.37)$$

For the process to be stationary,  $\phi_1$  must satisfy the condition  $-1 < \phi_1 < 1$  (Box and Jenkins 1976). Given (5.36), the first-order

autocorrelation function is

$$\rho_k = \phi_1 \rho_{k-1} \quad k > 0 . \quad (5.38)$$

With  $\rho_0 = 1$ , the solution to first-order difference equation is

$$\rho_k = \phi_1^k \quad k \geq 0 , \quad (5.39)$$

and in particular

$$\rho_1 = \phi_1 . \quad (5.40)$$

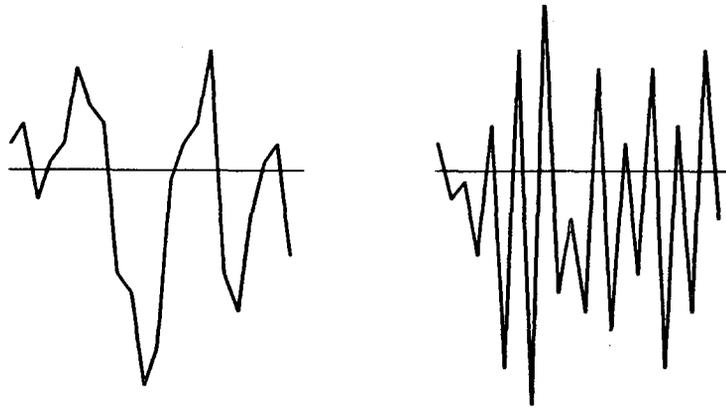
When  $0 \leq \phi_1 \leq 1$  (positive), the autocorrelation function decays exponentially to zero while the spectrum is dominated by low frequencies, cf. Figure 5.3. When  $-1 \leq \phi_1 < 0$  (negative), the autocorrelation function decays exponentially to zero while oscillating in sign and the spectrum is dominated by higher frequencies. Note that  $\omega$  is in units of  $\pi$  so that  $0 \leq \omega/\pi \leq 1$ .

The second-order autoregressive process AR(2) is written as

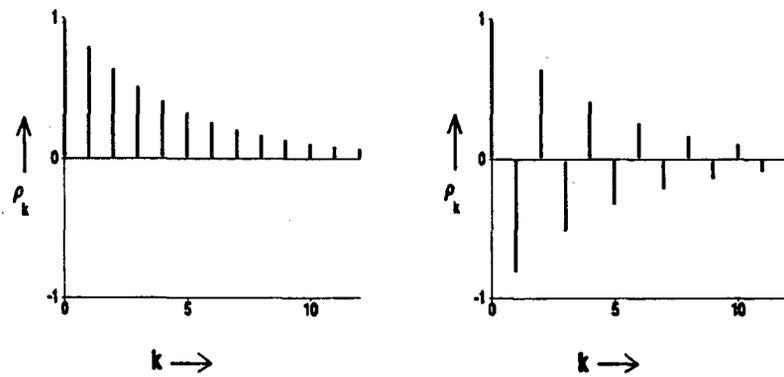
$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + a_t . \quad (5.41)$$

For stationarity, Box and Jenkins (1976) show that the roots of  $\phi(B) = 0$  must lie outside of the unit circle, which results in the parameters  $\phi_1$  and  $\phi_2$  having the following restrictions:

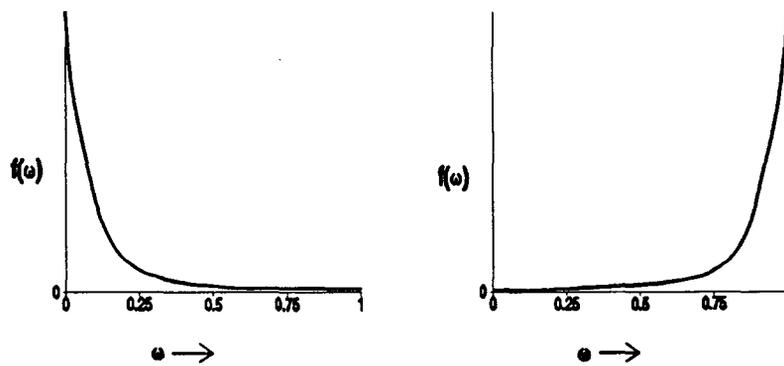
$$\begin{aligned} \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1 \end{aligned} \quad (5.42)$$



Theoretical AR(1) Processes



Theoretical Autocorrelation Functions



Theoretical Spectrums

Figure 5.3 Realizations from AR(1) processes and their corresponding theoretical autocorrelation functions and spectrums (after Box and Jenkins 1976).

The second-order autocorrelation function can be written as

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k > 0 \quad (5.43)$$

which, when solved for  $\phi_1$  and  $\phi_2$ , give

$$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad (5.44a)$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (5.44b)$$

Whereas the autocorrelation function indicates the general form of the stochastic model, the partial autocorrelation function is useful in determining the order. For an autoregressive process of order  $p$ , the partial autocorrelation function displays a cutoff after lag  $p$ , i.e., the parameter  $\phi_{kk}$  will be nonzero for  $k \leq p$  and zero for  $k > p$ .

Figure 5.4 shows the four general shapes of the autocorrelation function of the AR(2) process with their corresponding partial autocorrelation functions. Note that the partial autocorrelation function coefficient  $\phi_{kk}$  is zero for  $k > p$ . When both autocorrelation coefficients  $\phi_1$  and  $\phi_2$  are positive, the autocorrelation function decays exponentially to zero while the partial autocorrelation function exhibits positive coefficients  $\phi_{11}$  and  $\phi_{22}$ . When  $\phi_1$  is positive and  $\phi_2$  is negative, the autocorrelation function decays exponentially to zero while oscillating in sign and the partial autocorrelation function has a negative  $\phi_{11}$  and a positive  $\phi_{22}$ . When  $\phi_1$  is negative and  $\phi_2$  is positive,

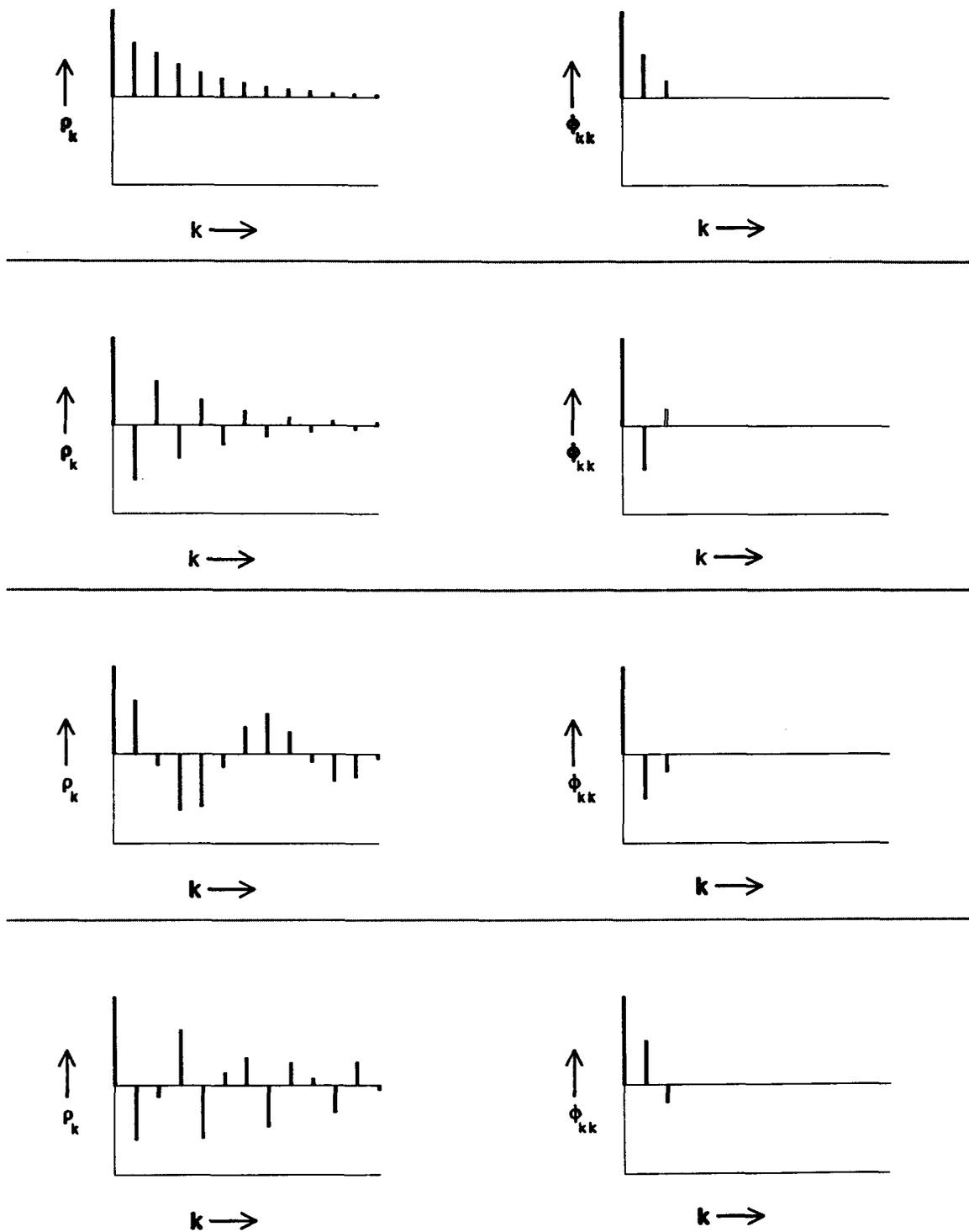


Figure 5.4 Typical autocorrelation functions  $\rho_k$  and partial autocorrelation functions  $\phi_{kk}$  for various AR(2) processes (after Box and Jenkins 1976).

the autocorrelation function decays in a damped sine wave and the partial autocorrelation function exhibits negative coefficients  $\phi_{11}$  and  $\phi_{22}$ . When both autocorrelation coefficients  $\phi_1$  and  $\phi_2$  are negative, the autocorrelation function decays in a damped sine wave while oscillating in sign and the partial autocorrelation function has a positive  $\phi_{11}$  and a negative  $\phi_{22}$ .

It should be noted that the experimental autocorrelation function and partial autocorrelation function may never truly approach zero as they are only estimates of the theoretical functions. Thus, the standard error can be used to determine when  $\rho_k$  or  $\phi_{kk}$  effectively becomes zero. When  $\rho_k$  or  $\phi_{kk}$  is greater than two standard errors, it can be assumed to be nonzero. Those estimated values which are much smaller than two standard errors can be assumed to be effectively zero. Values close to two standard errors require consideration as to their relevancy to the process. Figure 5.5 shows a partial autocorrelation function for a autoregressive process which is definitely of order 1 and possibly of order 2. Final determination of order is sometimes subjective and depends on the accuracy of forecasted values generated by a fitted model.

### The Moving Average Process Model

Another stochastic model very useful in describing certain common series is the moving average model. In the moving average model, the current value of the process is described by a finite linear combination of previous white noise values  $a_t$ . The finite moving average model is equivalent to an infinite autoregressive model (Bras and Rodríguez-Iturbe 1985). A moving average process of order  $q$  is represented by

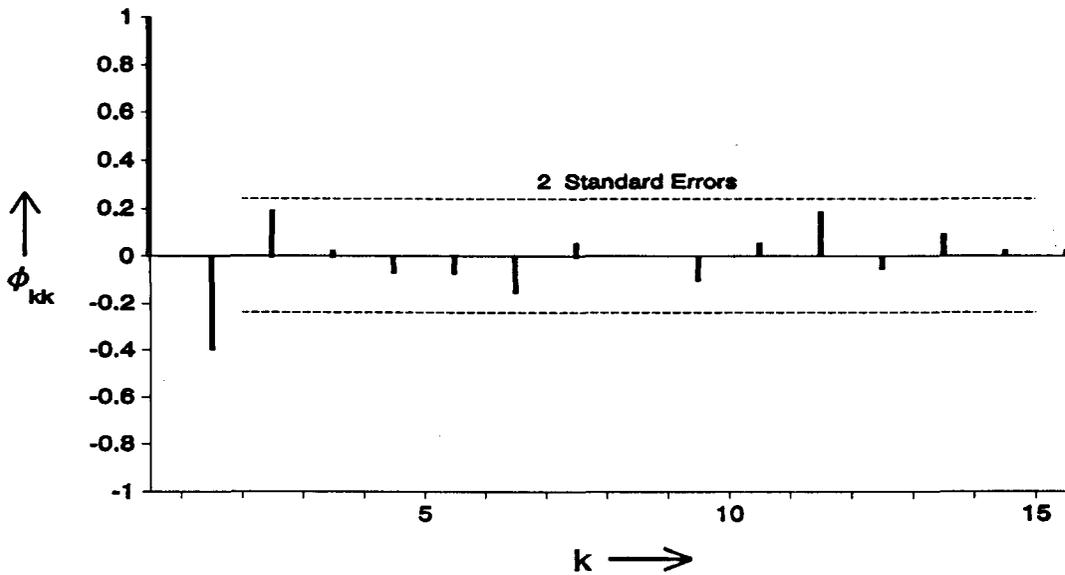


Figure 5.5 Estimated partial autocorrelation function together with two standard error limits calculated for an AR(1) model (after Box and Jenkins 1976).

$$\begin{aligned}
 \bar{z}_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} & (5.45) \\
 &= (1 - \theta_1 B - \dots - \theta_q B^q) a_t \\
 &= \theta(B) a_t .
 \end{aligned}$$

Box and Jenkins (1976) show that the roots of the equation  $\theta(B) = 0$  must lie outside of the unit circle for stationarity of the moving average model.

The autocovariance function  $C_k$  of a MA( $q$ ) process is derived by multiplying both sides of (5.45) by  $\bar{z}_{t-k}$  and then taking the expectation. Recalling that white noise has a zero mean and a variance of  $\sigma_a^2$ , the variance of the process becomes

$$C_0 = \sigma_z^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 . \quad (5.46)$$

When  $k$  is nonzero, the autocovariance function becomes

$$C_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q)\sigma_a^2 & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (5.47)$$

To compute the autocorrelation function  $\rho_k$  it is enough to divide the autocovariance function by the process variance  $C_0$ :

$$A_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (5.48)$$

Unlike the autoregressive process parameters, the moving average parameters cannot be solved using the Yule-Walker equations, but require iterative least-squares procedures (Box and Jenkins 1976). Initial estimates for the moving average parameters can be obtained by substituting estimates  $\hat{\rho}_k$  for  $\rho_k$  in (5.48) and solving the resulting equations.

Of considerable practical importance are the first- and second-order moving average processes. The first-order moving average model MA(1) has the form

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} \quad (5.49a)$$

or alternatively

$$\tilde{Z}_t = (1 - \theta_1 B)a_t \quad (5.49b)$$

The MA(1) process is stationary for all  $\theta_1$ , but it is necessary that  $|\theta_1| < 1$  in order to associate present events with past happenings in a sensible manner (Box and Jenkins 1976). When  $|\theta_1| > 1$ , then the present events are only associated with the present and future events of the process, an undesirable condition. This aspect of the modeling of time series is called invertibility and is discussed in further detail in Box and Jenkins (1976).

From (5.46), the variance of the process is defined as

$$C_0 = (1 + \theta_1^2) \sigma_a^2, \quad (5.50)$$

and the autocorrelation function as

$$\rho_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k \geq 2 \end{cases} \quad (5.51)$$

From (5.51) it can be seen that the solution for  $\theta_1$  takes the form of the quadratic equation

$$\theta_1^2 + \frac{\theta_1}{\rho_1} + 1 = 0. \quad (5.52)$$

Thus, for stationarity,  $\theta_1$  can assume one of two values, the roots of the quadratic. However, for invertibility, Box and Jenkins (1976) show that only one of the roots is valid.

The autocorrelation function  $\rho_k$  of a MA(1) process has a cutoff after the first lag while the partial autocorrelation function tails off

and is dominated by a damped exponential. In general,  $\rho_1$  is positive for a negative  $\theta_1$  and the spectrum is dominated by low frequencies. When  $\theta_1$  is positive,  $\rho_1$  is negative with the spectrum dominated by high frequencies.

The second-order moving average process is defined as

$$\bar{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (5.53)$$

and is stationary for all values of  $\theta_1$  and  $\theta_2$ . However, the MA(2) is invertible for only those values which conform to the following conditions:

$$\begin{aligned} \theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 &< \theta_2 < 1 \end{aligned} \quad (5.54)$$

which are parallel to the stationarity conditions of an AR(2) process.

From (5.46), the variance of the MA(2) process is defined as

$$C_0 = (1 + \theta_1^2 + \theta_2^2) \sigma_a^2, \quad (5.55)$$

and the corresponding autocorrelation function as

$$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \quad (5.56a)$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad (5.56b)$$

$$A_k = 0 \quad k \geq 3 . \quad (5.56c)$$

Values of  $\theta_1$  and  $\theta_2$  must be determined by solving the nonlinear equations (5.56).

It should be noted from this discussion that there exists a duality between moving average and autocorrelation processes. Figure 5.6 shows the four general shapes of the autocorrelation function of the MA(2) process with their corresponding partial autocorrelation functions. When compared to Figure 5.4, it can be seen that the autocorrelation function of a moving average process behaves similarly to the partial autocorrelation function of the corresponding autoregressive process, and vice versa. Also, the spectrum of a moving average process has an inverse relationship to the spectrum of the corresponding autoregressive process (Box and Jenkins 1976).

#### The Mixed Autoregressive-Moving Average Process Model

Processes that cannot be modeled exclusively by either the autoregressive or the moving average models may sometimes be represented by a combination of the two. The autoregressive-moving average model ARMA( $p, q$ ) takes the form

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (5.57)$$

$$\phi(B) \tilde{Z}_t = \theta(B) a_t . \quad (5.58)$$

This relationship defines a stationary model when the roots of  $\phi(B) = 0$  lie outside the unit circle. It also defines an invertible model when the roots of  $\theta(B) = 0$  lie outside the unit circle.

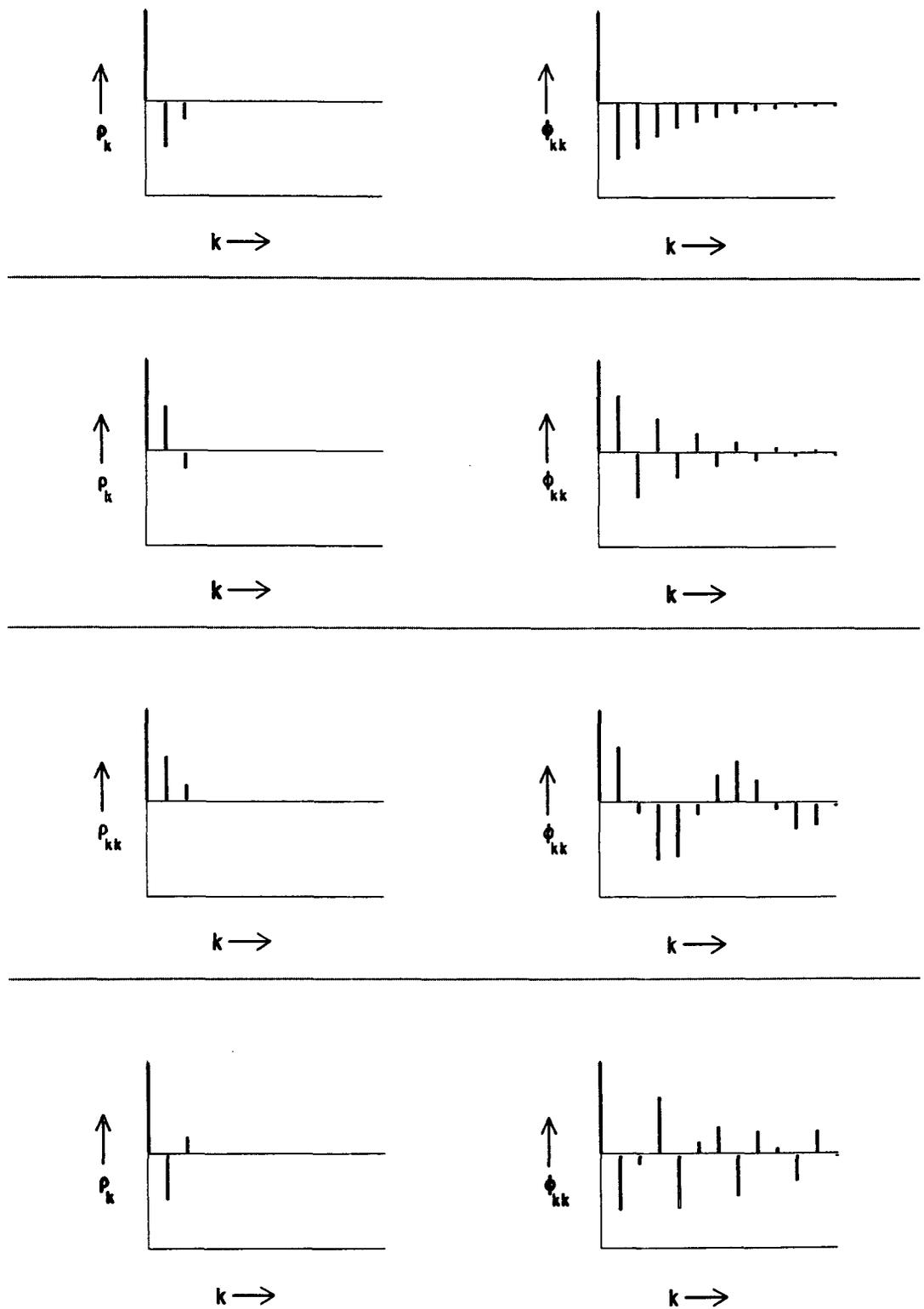


Figure 5.6 Typical autocorrelation functions  $\rho_k$  and partial autocorrelation functions  $\phi_{kk}$  for various MA(2) processes (after Box and Jenkins 1976).

Following the same procedure as for the autoregressive and moving average models, the autocorrelation function for the ARMA( $p, q$ ) model is developed to be

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k \geq q + 1 \quad (5.59a)$$

or

$$\phi(B)\rho_k = 0 \quad k \geq q + 1. \quad (5.59b)$$

In identifying mixed series, it should be noted that the autocorrelation function will behave differently according to whether  $q - p$  is positive or negative. If  $q - p < 0$  the entire autocorrelation function will consist of a mixture of damped exponentials and/or damped sine waves. If, however,  $q - p \geq 0$  then the autocorrelation function will consist of  $q - p + 1$  initial values  $\rho_0, \rho_1, \dots, \rho_{q-p}$ , which will not follow this form. The ARMA( $p, q$ ) model, then, has the convenient property that its first  $q$  autocorrelations depend on both moving average terms as well as autoregressive terms (Bras and Rodríguez-Iturbe 1985). Autoregressive behavior then dominates after  $q$  lags. The partial autocorrelation function of an ARMA( $p, q$ ) model is infinite in extent and behaves with a mixture of autoregressive and moving average behavior, depending on the parameters  $\phi_k$  and  $\theta_k$ .

The ARMA(1,1) process is of particular importance and is defined by

$$\bar{z}_t - \phi_1 \bar{z}_{t-1} = a_t - \theta_1 a_{t-1} \quad (5.60)$$

which can also be written as

$$(1 - \phi_1 B)\bar{Z}_t = (1 - \theta_1 B)a_t .$$

For stationarity and invertibility, the process parameters must fall in the ranges of  $-1 < \phi_1 < 1$  and  $-1 < \theta_1 < 1$ .

The first two autocorrelation coefficients of the process can be expressed in terms of the parameters  $\phi_1$  and  $\theta_1$  as follows

$$\rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1} \quad (5.61a)$$

$$\rho_2 = \phi_1 \rho_1 . \quad (5.61b)$$

For stationarity and invertibility, it can be shown that  $\rho_1$  and  $\rho_2$  must lie in the region

$$\begin{aligned} |\rho_2| &< |\rho_1| \\ \rho_2 &> \rho_1(2\rho_1 + 1) \quad \rho_1 < 0 \\ \rho_2 &> \rho_1(2\rho_1 - 1) \quad \rho_1 > 0. \end{aligned} \quad (5.62)$$

The partial autocorrelation function of the mixed ARMA(1,1) process consists of a single initial value  $\phi_{11} = \rho_1$ , after which it behaves like a pure MA(1) process. Thus when  $\theta_1$  is positive, the partial autocorrelation function is dominated by a smoothly damped exponential which decays from  $\phi_{11}$ . When  $\theta_1$  is negative, the partial autocorrelation function is dominated after  $\phi_{11}$  by an exponential which oscillates as it decays.

Table 5.1 summarizes the properties of an autoregressive-moving average process. It also provides a comparison of the three models which are restricted to being both stationary and invertible. Extensions of these three basic models can describe nonstationary type of behavior and will be examined next.

### The Autoregressive Integrated Moving Average Process Model

Many time series display homogenous behavior yet are nonstationary in the mean of the process. Thus, although one part of the series may behave very much like any other, its mean is not stationary over time. A nonstationary series can be homogenous in local behavior but not in local level, such that one part of it looks much the same as another except in vertical translation. Homogenous nonstationarity such as this can be modeled by differencing the process  $d$  times, i.e.  $\nabla^d Z_t = (1-B)^d Z_t$ . The differenced process can then be modeled as a stationary mixed autoregressive-moving average process. This particular formulation is called the autoregressive integrated moving average process of orders  $p$ ,  $d$ , and  $q$ , ARIMA( $p,d,q$ ). It should be noted that

$$\text{ARMA}(p,q) = \text{ARIMA}(p,0,q)$$

$$\text{AR}(p) = \text{ARIMA}(p,0,0)$$

$$\text{MA}(q) = \text{ARIMA}(0,0,q) .$$

It is possible to eliminate an unknown but constant mean from a time series by taking a first-order difference. If the process was otherwise homogenous, the differenced data should then be stationary with zero mean. By taking second-order differences, i.e. differences of

Table 5.1 Summary of properties of autoregressive, moving average, and mixed ARMA processes (after Box and Jenkins 1976).

	Autoregressive Processes	Moving Average Processes	Mixed Processes
Model in terms of previous $\bar{Z}$ 's	$\phi(B)\bar{Z}_t = a_t$	$\theta^{-1}(B)\bar{Z}_t = a_t$	$\theta^{-1}(B)\phi(B)\bar{Z}_t = a_t$
Model in terms of previous $a$ 's	$\bar{Z}_t = \phi^{-1}(B)a_t$	$\bar{Z}_t = \theta(B)a_t$	$\bar{Z}_t = \phi^{-1}(B)\theta(B)a_t$
Stationarity condition	roots of $\phi(B) = 0$ lie outside unit circle	always stationary	roots of $\phi(B) = 0$ lie outside unit circle
Invertibility condition	always invertible	roots of $\theta(B) = 0$ lie outside unit circle	roots of $\theta(B) = 0$ lie outside unit circle
Autocorrelation function	infinite (damped exponentials and/or damped sine waves)	finite	infinite (damped exponentials and/or damped sine waves after first $q - p$ lags)
	tails off	cuts off	tails off
Partial autocorrelation function	finite	infinite (dominated by damped exponentials and/or sine waves)	infinite (dominated by damped exponentials and/or sine waves after first $p - q$ lags)
	cuts off	tails off	tails off

a differenced series, it is possible to eliminate unknown linear trends if the process otherwise exhibits homogeneity. The first difference would eliminate the unknown slope and produce a stationary process with nonzero mean. The second difference would then eliminate the unknown but constant mean. These concepts are illustrated in Figure 5.7.

The general form of the ARIMA( $p,d,q$ ) model is

$$\phi(B)\nabla^d\bar{Z}_t = \theta(B)a_t, \quad (5.63)$$

where

$$\nabla^d\bar{Z}_t = (1 - B)^d\bar{Z}_t$$

Recalling that  $\bar{Z}_t = Z_t - m$ , it becomes obvious that  $\nabla^d\bar{Z}_t = \nabla^dZ_t$  for  $d \geq 1$ , and the ARIMA( $p,d,q$ ) model can then be written as

$$\phi(B)\nabla^dZ_t = \theta(B)a_t. \quad (5.64)$$

The general form of the autoregressive integrated moving average model used to describe time series is a slightly modified version of (5.64). The model described by (5.64) is capable of representing processes which have stochastic trends, i.e., trends typified by random changes in slope or level. However, sometimes it becomes desirable to include a deterministic constant  $\theta_0$  which allows for a deterministic polynomial trend of degree  $d$ . For example,  $\theta_0$  can be used to estimate a possible deterministic linear trend when  $d = 1$ . Thus, (5.64) then becomes

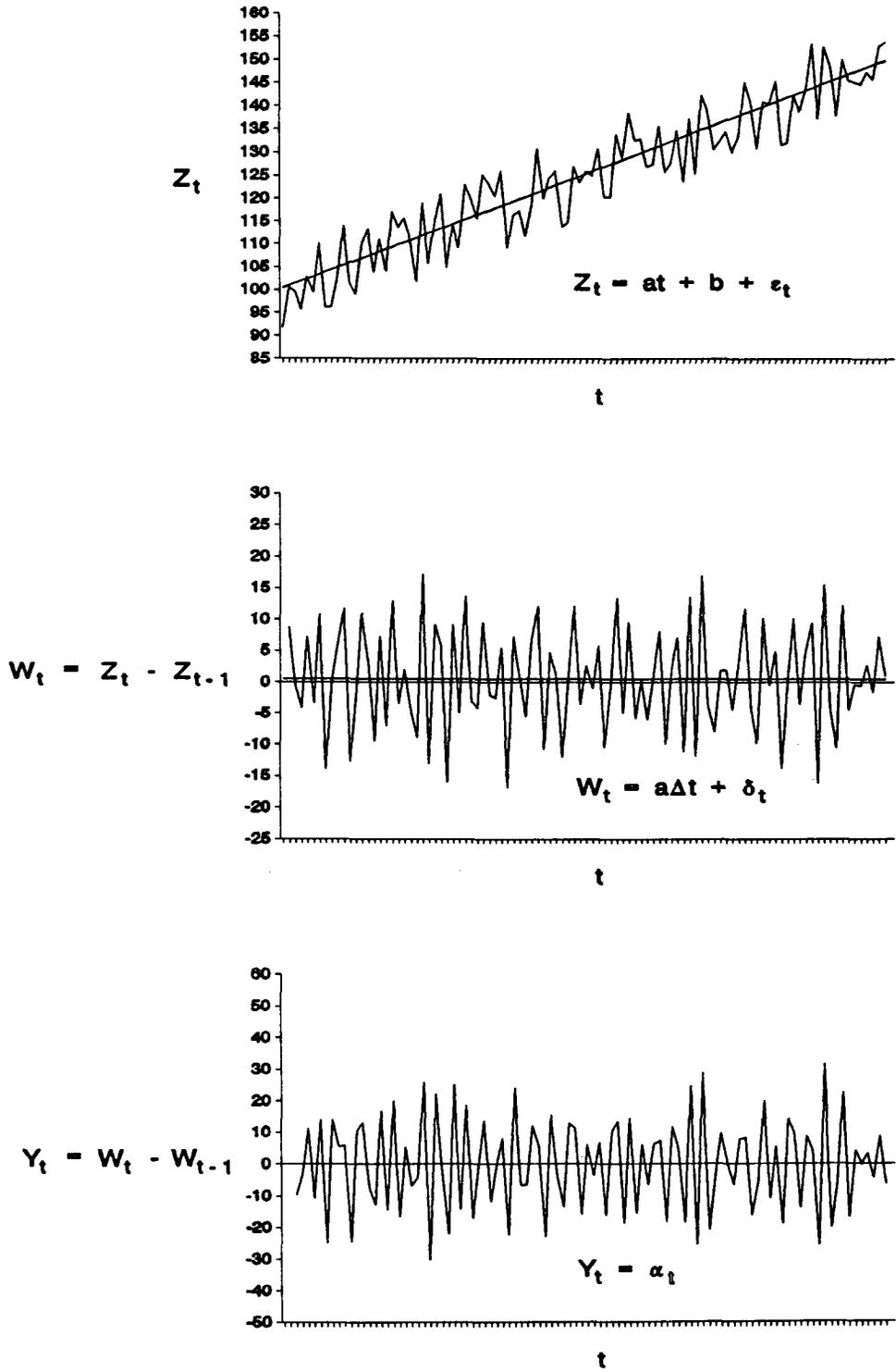


Figure 5.7 Time series plots. (a) Original series with nonstationary mean; (b) First-order differenced series with nonzero mean; (c) Second-order differenced series with zero mean.

$$\phi(B)\nabla^d Z_t = \theta_0 + \theta(B)a_t . \quad (5.65)$$

An alternate way of expressing this general model is to allow  $W_t = \nabla^d Z_t$ . The constant  $\theta_0$  can then be defined as

$$E\{W_t\} = E\{\nabla^d Z_t\} = \mu = \theta_0 / (1 - \phi_1 - \dots - \phi_p) .$$

Defining  $\tilde{W}_t = W_t - \mu$  as the deviations of  $W_t$  about its mean  $\mu$ , then the ARIMA( $p, d, q$ ) model can be written in its most general form as

$$\phi(B)\tilde{W}_t = \theta(B)a_t . \quad (5.66)$$

The similarity between the general form of the ARIMA( $p, d, q$ ) model (5.66) and the general form of the ARMA( $p, q$ ) model (5.58) is now clearly evident. Thus, it follows that the ARIMA( $p, d, q$ ) model behaves the same as the ARMA( $p, q$ ) and is characterized by similar autocorrelation and partial autocorrelation functions (see Table 5.1).

### Seasonal Models

Many time series, particularly hydrologic time series, show marked seasonal patterns or nonstationarity. For example, a time series of piezometric head levels will display a similarity in monthly variations from year to year. This annual pattern of groundwater fluctuation is evident in the almost perfect correlation between piezometric head levels at 12 months' lag. In general a series is said to exhibit

periodic behavior with period  $s$  when similarities in the series occur after  $s$  basic time intervals (Box and Jenkins 1976).

Processes which exhibit seasonal behavior represent a particular type of nonstationarity which may be modeled with the simplifying operation

$$\nabla_s Z_t = (1 - B^s)Z_t = Z_t - Z_{t-s} . \quad (5.67)$$

Thus, a purely seasonal model may be thought of as a variation on the general ARIMA( $p,d,q$ ) and may be written as

$$\phi(B^s)\nabla_s^D Z_t = \theta(B^s)\alpha_t , \quad (5.68)$$

where  $\nabla_s^D$  represents seasonal differences  $s$  of order  $D$ , and  $\phi(B^s)$ ,  $\theta(B^s)$  are polynomials in  $B^s$  of degrees  $P$  and  $Q$ , respectively. This seasonal model, then, is called an ARIMA( $P,D,Q$ ) model, and satisfies stationarity and invertibility conditions.

The error component  $\alpha_t$  of (5.68) in general will be correlated with previous errors  $\alpha_{t-1}$ ,  $\alpha_{t-2}$ ,.... Thus it would be expected that modelling these errors with an ARIMA( $p,d,q$ ) would take care of the correlations. Therefore, a second model is introduced

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t , \quad (5.69)$$

where  $a_t$  represents a white noise process,  $\nabla^d$  represents nonseasonal differences of order  $d$ , and  $\phi(B)$ ,  $\theta(B)$  are polynomials in  $B$  of degrees  $p$

and  $q$ , respectively. Stationarity and invertibility conditions are satisfied by this model.

By substituting (5.69) into (5.68), a general multiplicative model is obtained

$$\phi(B)\phi(B^s)\nabla^d\nabla_s^D Z_t = \theta(B)\theta(B^s)a_t . \quad (5.70)$$

The resulting multiplicative model is termed an  $ARIMA(p,d,q)\times(P,D,Q)_s$  model. The multiplicative model emphasizes the fact that in periodic processes, there exist multiple time intervals of importance. For example, in the piezometric head level timeseries, correlations are seen to exist between observations for successive months in a particular year and between observations for the same month in successive years. Models with three or more periodic components can be developed to describe processes with multiple seasonalities.

### Process Identification

Figure 5.8 illustrates an iterative algorithm of the three main steps in stochastic model building: identification, estimation, and verification. The identification phase involves the rough estimate of the general form of the representational model to be studied. By estimating initial values of  $p$ ,  $d$ ,  $q$ , and/or  $P$ ,  $D$ ,  $Q$  of the general ARIMA model, a tentative model can be obtained. This tentative model can then serve as a starting point for model parameter estimation.

The first step in model identification is the visual inspection of a plot of the original time series. Nonstationarities, seasonal trends, and extreme values should be evident from this plot. The general form

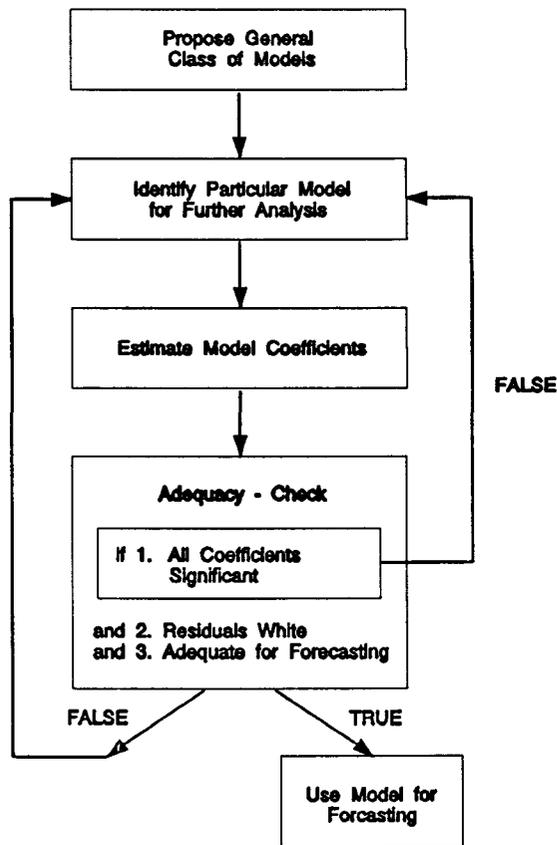


Figure 5.8 Iterative approach to model building (after Bras and Rodriguez-Iturbe 1985).

of the tentative model can also be inferred from the behavior of the original data. If strong trends or seasonalities are evident, then local or seasonal differencing may be required.

Plots of the estimated autocorrelation function  $\hat{\rho}_k$  and the estimated spectrum  $\hat{f}(\omega)$  of the original data can yield further information. It should be recalled from (5.10) and (5.11) that estimates of the autocorrelation function  $\rho_k$  can be obtained from

$$\hat{\rho}_k = \frac{\frac{1}{n-1} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\hat{\sigma}_z^2}, \quad (5.71)$$

where  $\bar{z}$  and  $\hat{\sigma}_z^2$  are given in equations (5.3) and (5.4), and  $n$  is the number of available data values. As  $k$  increases, the above estimator becomes increasingly variable, thus Box and Jenkins (1976) recommend that the autocorrelation should only be estimated up to lag  $k = n/4$ . A plot of (5.71) may reveal nonstationarities or seasonalities that may or may not be evident in the plot of the original time series. A failure of the estimated autocorrelation function to dampen with increasing lag suggests that the underlying stochastic process is nonstationary. Thus differencing of first-order or higher may be required. Further, if a wave pattern with period  $s$  is detected, seasonality would be indicated and seasonal differencing would be necessary.

A plot of the estimated spectrum  $\hat{f}(\omega)$  as computed from (5.28) can also reveal whether seasonality exists in the original data. As the spectrum represents the contribution to the process variance of components with frequencies in the  $(\omega, \omega+d\omega)$  range, a peak in the plot of the spectrum will indicate an important contribution to the variance at that frequency. Thus, for a seasonal process, peaks should be expected at frequencies corresponding to important seasonal time lags. For example, a time series which cycles on a twelve month basis, with highs alternating with lows every six months, may have a spectrum with peaks at frequencies corresponding to periods of three, six, and twelve months.

Once the original data has been detrended by differencing (if detrending was necessary), then a review of the estimated autocorrelation function and estimated spectrum of the stationary data can reveal information on the general form of the model to be used. A plot of the partial autocorrelation function becomes useful at this point. Tentative values for the partial autocorrelations can be estimated from the recursive formula given by Box and Jenkins (1976)

$$\hat{\phi}_{11} = \hat{\rho}_1, \quad (5.72a)$$

$$\hat{\phi}_{p+1,p+1} = \frac{\rho_{p+1} - \sum_{j=1}^p \hat{\phi}_{p,j} \rho_{p-j+1}}{1 - \sum_{j=1}^p \hat{\phi}_{p,j} \hat{\rho}_j}, \quad (5.72b)$$

where

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{p,j} - \hat{\phi}_{p+1,p+1} \hat{\phi}_{p,p-j+1} \quad j = 1, 2, \dots, p.$$

An autoregressive process of order  $p$  is indicated if the autocorrelation function tails off while the partial autocorrelation function cuts off after lag  $p$ . Conversely, a moving average process of order  $q$  is indicated if the autocorrelation function cuts off after lag  $q$  while the partial autocorrelation function tails off. If both the autocorrelation function and partial autocorrelation function tail off, a mixed process is indicated. A review of Table 5.1 will help in the model identification process.

## Model Parameter Estimation

Once a tentative model has been chosen, the next step in the model building process is parameter estimation. Parameters are first roughly estimated and then refined using iterative procedures. For a general ARIMA( $p, d, q$ ) model, the calculation of initial estimates of the nonseasonal parameters  $\phi$  and  $\theta$  is a three stage process.

- (1) The autoregressive parameters  $\phi_i (i = 0, 1, \dots, p)$  are estimated from the autocorrelations  $\hat{\rho}_j (j = q-p+1, \dots, q+p)$  using the relationship of (5.36).
- (2) Using the  $\hat{\phi}$  estimates calculated in (1), the first  $q+1$  autocorrelations  $\hat{\rho}'_j (j = 0, 1, \dots, q)$  of the derived series

$$W'_t = W_t - \hat{\phi}_1 W_{t-1} - \dots - \hat{\phi}_p W_{t-p}$$

are calculated.

- (3) The autocorrelations  $\hat{\rho}'_j (j = 0, 1, \dots, q)$  are used in an iterative Newton-Raphson algorithm to compute initial estimates of the moving average parameters  $\theta_i (i = 0, 1, \dots, q)$  and of the residual variance  $\sigma_a^2$ .

The calculation of initial estimates of the seasonal parameters  $\phi$  and  $\theta$  follows precisely the same three step process described for nonseasonal parameters. The reader is referred to Box and Jenkins (1976), Bras and Rodríguez-Iturbe (1985), and SAS Institute, Inc. (1988) for more detailed reviews of the calculation of initial parameter estimates and the Newton-Raphson iteration algorithm.

Once initial estimates for the model parameters have been made, these rough estimates must be refined. The refinement procedure used by

Box and Jenkins (1976) involves least squares estimates of the parameters  $\mu$ ,  $\phi$ ,  $\theta$ ,  $\phi$ ,  $\theta$ , and  $\sigma_a^2$  in the general multiplicative model

$$\phi(B)\phi(B^s)(W_t - \mu) = \theta(B)\theta(B^s)a_t$$

where  $W_t = \nabla^d \nabla_s^D Z_t$ ,

$\mu$  is the mean value of the  $W$  series.

To calculate the residual sum of squares, it is necessary to back forecast initial values of  $W$ 's so that residuals may be calculated in a two step process:

$$a_t = (W_t - \mu) - \sum_{i=1}^p \phi_i (W_{t-i} - \mu) + \sum_{j=1}^q \theta_j a_{t-j}$$

$$a_t = \alpha_t - \sum_{i=1}^p \phi_i \alpha_{t-is} + \sum_{j=1}^q \theta_j a_{t-js}$$

where  $t = t', t' + 1, \dots, n$  and  $t'$  is a negative origin for  $t$  for the back forecasts. For given parameters  $(\mu, \phi, \theta, \phi, \theta)$ , the residual sum of squares is defined as

$$S(\mu, \phi, \theta, \phi, \theta) = - \sum_{t=t'}^n a_t^2 . \quad (5.73)$$

The values of these parameters which minimize the residual sum of squares are obtained using the method proposed by Marquardt (1963).

## Model Verification

Upon estimating the model parameters, verification of the adequacy of the model is necessary. Theoretically, the model residuals  $a_t$  should form a white noise random sequence (Bras and Rodríguez-Iturbe 1985). Model verification determines whether the actual residuals have this expected white noise behavior.

White noise behavior would require that the residuals  $a_t$  be uncorrelated. Correlation can be checked by plotting the autocorrelation function of the residual series. A white noise series would be indicated if the autocorrelation function has a cutoff to zero after lag 0. Additionally, the spectrum of a white noise residual series would be uniform with no predominate frequency peaks.

A white noise series is normally distributed, and thus a test of residual normality can verify the validity of the proposed model. The residuals should possess an approximately zero mean and an approximately zero skewness to ensure an accurate model.

## Forecasting

Forecasting future values of a process based on present and past values is the primary goal of stochastic model building. Once a model has been identified, fit, and verified, the process of forecasting future values may begin. By definition, stochastic models forecast the expected value of a random future event. Thus any forecast should be accompanied by information about the expected deviation around this forecast, which is summarized by the forecast variance or standard deviation.

Although several equivalent forecasting algorithms exist, forecasting in terms of the difference equation is the simplest and most elegant. Consider the simple multiplicative model  $ARIMA(p,0,0) \times (0,1,Q)_s$ ,

$$\phi(B)W_t = \theta_0 + \theta(B^s)a_t . \quad (5.74)$$

This model can be used to predict future values from origin  $t$  for lead times  $l$ . Thus, for  $l \geq 1$ , and recalling that  $W_t = \nabla_s Z_t$ , the equation for  $Z_{t+l}$  may be written as

$$\begin{aligned} Z_{t+l} = & \theta_0 + Z_{t+l-s} + \phi_1(Z_{t+l-1} - Z_{t+l-1-s}) \\ & + \dots + \phi_p(Z_{t+l-p} - Z_{t+l-p-s}) + a_{t+l} \\ & - \theta_1 a_{t+l-s} - \dots - \theta_q a_{t+l-s-q} \end{aligned} \quad (5.75)$$

Since  $a_{t+l}$  is an unknown random variable prior to time  $t+l$ , only the expected value of  $Z_{t+l}$  can be forecast. Thus, the forecast of  $Z_{t+l}$  is written as

$$\begin{aligned} E\{Z_{t+l}\} = \hat{Z}_t(l) = & \theta_0 + Z_{t+l-s} + \phi_1(Z_{t+l-1} - Z_{t+l-1-s}) \\ & + \dots + \phi_p(Z_{t+l-p} - Z_{t+l-p-s}) - \theta_1 a_{t+l-s} \\ & - \dots - \theta_q a_{t+l-s-q} \end{aligned} \quad (5.76)$$

where  $\hat{Z}_t(l)$  is the expected value or forecast of  $Z_{t+l}$ . Note that all known (and thus deterministic) quantities have expected values equal to their observed value, however  $E\{a_{t+l}\}$  is equal to zero and thus does not appear in (5.76).

Equation (5.76) is valid for  $l = 1$ , but for lead values of  $l$  greater than one, it must be further modified by replacing unknown  $Z$ 's with their respective expected values (forecasts). To illustrate this further, consider the case of the ARIMA(1,0,0)x(0,1,1)<sub>12</sub> model

$$(1 - \phi_1)(Z_t - Z_{t-12}) = \theta_0 + (1 - \theta_{12})a_t . \quad (5.77)$$

The lead one forecast can be written following (5.76)

$$\hat{Z}_t(1) = \theta_0 + Z_{t+11} + \phi_1(Z_t - Z_{t-12}) - \theta_1 a_{t-11} . \quad (5.78)$$

However, the lead two forecast would have to be

$$E\{Z_{t+2}\} = \hat{Z}_t(2) = E\{\theta_0 + Z_{t+10} + \phi_1(Z_{t+1} - Z_{t-11}) - \theta_1 a_{t-10} + a_{t+2}\} . \quad (5.79)$$

In equation (5.79)  $Z_{t+1}$  and  $a_{t+2}$  are the only random variables. Given that  $E\{Z_{t+1}\} = \hat{Z}_t(1)$  and  $E\{a_{t+2}\} = 0$ , the lead two forecast becomes:

$$\hat{Z}_t(2) = \theta_0 + Z_{t+10} + \phi_1(\hat{Z}_t(1) - Z_{t-11}) - \theta_1 a_{t-10} . \quad (5.80)$$

Equation (5.80) expresses the forecast in terms of previous  $Z$ 's and previous forecasts of  $Z$ 's. It should be clear that as  $l$  increases, the forecast estimation error will also increase as each forecast becomes increasingly dependent on previous forecasts.

Although separate expressions for each lead time  $l$  may be written, forecast computation can be best carried out by using a general expression. Thus, a general forecast equation for the ARIMA(1,0,0)x

$(0,1,1)_{12}$  model may be written as

$$\begin{aligned} \hat{Z}_t(l) &= \theta_0 + Z_{t+l-12} + \phi_1(Z_{t+l-1} - Z_{t+l-13}) \\ &\quad - a_{t+l-12} + \theta_1 a_{t+l-13} \end{aligned} \quad (5.81a)$$

where

$$Z_{t+j} = \begin{cases} Z_{t+j} & j \leq 0 \\ \hat{Z}_t(j) & j > 0 \end{cases} \quad (5.81b)$$

$$a_{t+j} = \begin{cases} a_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} \quad (5.81c)$$

Consider that forecasts at lead times  $1, 2, \dots, L$  are required. These estimates have little meaning unless they are accompanied by a forecast variance.

Box and Jenkins (1976) show that the forecast variance for a general  $ARIMA(p, d, q) \times (P, D, Q)_s$  is given by

$$V(l) = \left\{ 1 + \sum_{j=1}^{l-1} \Psi_j^2 \right\} \sigma_a^2 . \quad (5.82)$$

The weights  $\Psi_1, \Psi_2, \dots, \Psi_L$  are calculated utilizing the following relationship:

$$\phi(B) \Psi(B) = \theta(B) , \quad (5.83)$$

where  $\phi(B) = \phi(B)\phi(B^s)\nabla^d\nabla_s^D$ ,

$$\theta(B) = \theta(B)\theta(B^s).$$

The operator  $\phi(B)$  is a polynomial in the  $p^*$ th degree

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_{p^*} B^{p^*}, \quad (5.84)$$

where  $p^* = p + sP + d + sD$ . Similarly,  $\theta(B)$  is a polynomial in the  $q^*$ th degree

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_{q^*} B^{q^*}, \quad (5.85)$$

where  $q^* = q + sQ$ . Knowing the values of the  $\phi$ 's and  $\theta$ 's, the  $F$ 's may be calculated using

$$F_j = \sum_{i=1}^j \phi_i F_{j-i} - \theta_j \quad j = 1, 2, \dots, L \quad (5.86)$$

where  $F_0 = 1$ ,  $F_j = 0$  for  $j < 0$  and  $\theta_j = 0$  for  $j > q^*$ .

Note that the variance of the residuals  $\sigma_a^2$  in equation (5.82) can be substituted with estimate  $\hat{\sigma}_a^2$  when such an estimate is calculated from at least 50 observations.

Assuming that the  $a$ 's are normally distributed, then the probability distribution of the future value  $Z_{t+l}$  of the process will be normally distributed with mean  $\hat{Z}_t(l)$  and standard deviation  $\{V(l)\}^{1/2}$  (Box and Jenkins 1976). It should be noted that the probabilities calculated apply only to the forecasts produced given the information available at origin  $t$ . The probabilities calculated for a lead time  $l_0$  from origin  $t$  will be unique and different from the probabilities calculated for a

lead time 5 from origin  $t+5$ . Also, while it is true that the actual value of  $Z_{t+10}$  may fall within the 95% probability limits of the forecast  $\hat{Z}_t(10)$ , the value  $Z_{t+5}$  may not fall within the 95% probability limits of the forecast  $\hat{Z}_{t+5}(5)$ .

### Excursion Analysis

Excursion analysis was developed by S. O. Rice in 1945 in connection with statistical analysis of white noise time series (Nordin and Rosbjerg 1970). For those time series which are approximately Gaussian, excursion analysis can provide an estimate of the length of time that the process runs above or below a given level. In hydrogeology, this is particularly helpful for predicting the duration of an excursion of piezometric head levels above or below a fixed critical level  $h$ , which is measured in units of standard deviations. Obviously,  $h$  can be chosen so that excursions below the level are representative of drought conditions.

Figure 5.9 shows a sketch of a continuous time series  $y(t)$  which has been normalized to have a zero mean and standard deviation  $h$ . The number of excursions below the level  $-h$  by the process in the interval  $(0,T)$  represents the number of droughts in the time interval  $T$ , while the expected duration of the excursion below  $-h$ ,  $E\{T_h^-\}$ , is the mean duration of each drought. Similarly, the expected interval between zero crossings,  $E\{T_0\}$ , is simply the average time interval that the piezometric head levels are above or below average.

Nordin and Rosbjerg (1970) show that the expected interval between zero crossings as given by Rice (1945) can be expressed as

$$E\{T_0\} = \pi \left[ \frac{\rho_0}{-\rho_0''} \right]^{1/2}, \quad (5.87)$$

where  $\rho_0$  is the correlation of the normalized random process at the origin and  $\rho_0''$  is the second derivative of  $\rho_0$ . The average number of excursions below  $-h$  is given by

$$E\{N\} = \frac{1}{2} e^{-\frac{1}{2} h^2} [E\{T_0\}]^{-1}, \quad (5.88)$$

and the expected duration of an excursion below the fixed level  $-h$  is

$$E\{T_h^-\} = 2E\{T_0\} \text{pr}\{y(0) > h\} e^{\frac{1}{2} h^2} \quad (5.89)$$

where

$$\text{pr}\{y(0) > h\} = \frac{1}{\sqrt{2\pi}} \int_h^\infty e^{-\frac{1}{2} x^2} dx \quad (5.90)$$

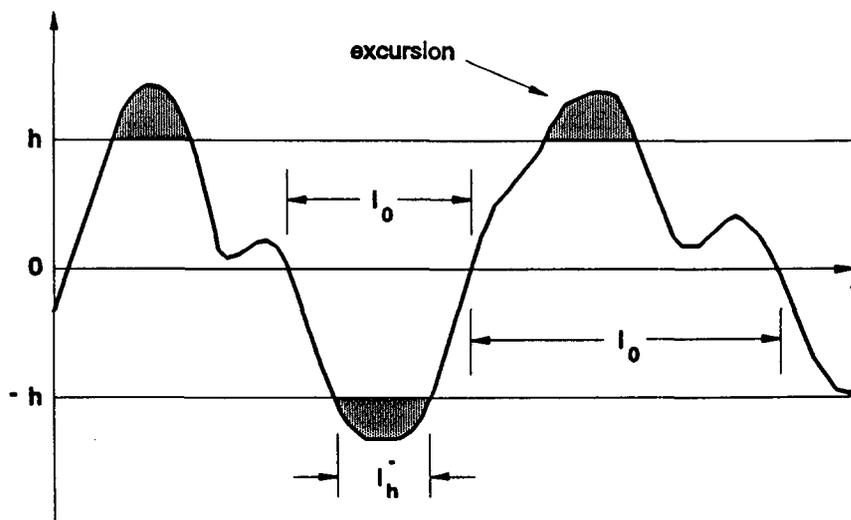


Figure 5.9 Definition sketch for continuous time series (after Nordin and Rosbjerg 1970).

In practice, hydrogeologic time series are not continuous but discrete, with time intervals of  $\Delta t$  and process values which are averaged over  $\Delta t$ . In this case, Nordin and Rosbjerg (1970) give the value of  $E\{I_0\}$  as

$$E\{I_0\} = \frac{\pi}{\sqrt{2}(1 - \rho_1)^{\frac{1}{2}}} \quad (5.91)$$

where  $\rho_1 = \rho_{\Delta t}$ , the value of the autocorrelation function for the first lag interval. Since the time series is given as a series of mean values assumed to apply over a given time interval  $\Delta t$ , Nordin and Rosbjerg (1970) show that (5.89) can be approximated by

$$E\{I_0\} = \left( \frac{1}{2} - \frac{1}{\pi} \arcsin \rho_1 \right)^{-1} . \quad (5.92)$$

The average number of excursions below a fixed level  $-h$  and the expected duration of each excursion can then be calculated using (5.88) and (5.89).

Although excursion analysis was developed for Gaussian processes, moderate departures from normality do not seem to greatly influence the applicability of the theory (Nordin and Rosbjerg 1970).

### Recurrence Analysis

For a stationary stochastic process, the probability of extreme high or low values occurring can be determined from the experimental cumulative distribution function calculated from a time series observed at interval  $\Delta t$ .

The experimental cumulative distribution function (cdf) can be calculated by

$$\hat{F}(Z_0) = \text{prob}(Z \leq Z_0) = \frac{\text{Rank}(Z_0)}{n + 1} \quad (5.93)$$

where  $\text{prob}(Z \leq Z_0)$  is the probability of a data value being less than or equal to a given value  $Z_0$  in a given observation interval  $\Delta t$ ,  $\text{Rank}(Z_0)$  is the rank of the actual data value corresponding to  $Z_0$ , and  $n$  is the number of data available. The smallest data value is given a rank of 1, and the largest data value is given a rank of  $n$ .

Recurrence intervals can be inferred once the experimental cdf has been calculated according to the following equation:

$$R(Z_0) = \frac{1}{\hat{F}(Z_0)} \Delta t \quad (5.94)$$

where  $R(Z_0)$  is the recurrence interval associated with the value  $Z_0$  (measured in units of  $\Delta t$ ),  $\hat{F}(Z_0)$  is calculated according to (5.93), and  $\Delta t$  is the observation interval of the process  $Z_t$ .

#### Description of Software Packages Used

The computer software packages used for the analysis of the provided time series data were the base SAS<sup>®</sup> System, and the SAS/GRAPH<sup>®</sup> and SAS/ETS<sup>®</sup> packages.<sup>1</sup> The SAS system provides data retrieval and

---

<sup>1</sup>SAS, SAS/GRAPH, and SAS/ETS are registered trademarks of SAS Institute Inc., Cary, NC, USA.

management, programming, statistical, and reporting capabilities.

SAS/GRAPH is a graphics package for producing two and three dimensional graphs, charts, and figures. SAS/ETS is an econometrics and time series analysis package. The primary SAS/ETS procedures employed in building the stochastic models developed for this report were the ARIMA and SPECTRA procedures.

The SAS/ETS ARIMA procedure analyses and forecasts univariate time series data, transfer function data, and intervention data using the autoregressive integrated moving average model of Box and Jenkins (1976). The ARIMA procedure provides tools for univariate time series model identification, parameter estimation, and forecasting, and allows for seasonal, subset, and factored models (SAS Institute, Inc. 1988).

Corresponding to the model building stages described by Box and Jenkins (1976), the SAS/ETS ARIMA procedure is divided into three stages: IDENTIFY, ESTIMATE, and FORECAST. The IDENTIFY statement identifies the model. It specifies the time series data to be analyzed, differences them if applicable, and computes autocorrelations, inverse autocorrelations, partial autocorrelations, and crosscorrelations. The IDENTIFY statement computes statistics of the response variable as well as test statistics for the white noise hypothesis of the series.

The ESTIMATE statement estimates the model parameters and diagnoses the fit. Preliminary estimates of the parameters are either specified by the user, computed from the autocorrelations developed in the IDENTIFY stage, or arbitrarily set to values that produce stable polynomials. These preliminary estimates are the starting values in an iterative algorithm to compute the model parameter estimates. Three methods of estimation are provided: the maximum-likelihood estimation

method, the unconditional least-squares estimation method, and the conditional least-squares estimation method, which is the default method. The ESTIMATE statement provides sample statistics on the estimated parameters as well as the estimation of the standard deviation of the residuals  $\sigma_a$ , the correlation matrix of the parameters, test statistics for the white noise hypothesis of residuals, and a summary of the estimated model.

The FORECAST statement computes future values of the time series based on the estimated model and generates confidence intervals for these forecasts. It provides a table of forecasts, forecast standard errors, approximate 95 percent confidence limits for the forecasts, values of the response variable (actual process values), and the forecast residuals.

The SPECTRA procedure produces estimates of the spectral densities of the univariate or multivariate time series. These estimates are produced using a finite Fourier transform. SPECTRA can also test the data for the white noise hypothesis.

### Hydrogeologic Time Series Data Description and Analysis

The St. Johns River Water Management District provided latitude, longitude, and piezometric head level information for 25 wells within the district and surrounding counties. The piezometric head level data was relatively consistent, although a few wells had data gaps of a few months to a few years. All wells had occasional data gaps of a few days to a few weeks. In general, daily piezometric head levels were provided for each well, although some wells had periods of weekly, biweekly, or monthly readings. Of the 25 wells provided, 4 were discarded due to

lack of sufficient data or extremely large data gaps. Table 5.2 lists the 21 wells which were modeled, giving identification numbers, latitude and longitude information, years of relatively uninterrupted head level readings, and number of monthly observations  $n$ . Figure 5.10 shows the spatial distribution of the wells in the district and surrounding counties.

Upon receiving the well data from the District, computer files containing the piezometric head levels were prepared. Where applicable, monthly head values were computed from the average of the daily, weekly, or biweekly readings provided for each well, creating time series of  $n$  monthly observations ( $\Delta t = 1$  month). These time series were then plotted for visual inspection. It was noted that 13 of the 21 wells exhibited steady downward trends in head levels, while the remaining 8 had no detectable downward trend. Seasonal trends were readily apparent from all the time series plots, with seasonal lows occurring around May of each year and seasonal highs occurring around September. For each well, the annual mean, standard deviation, high head level and low head level were computed for the each year of record, cf. Appendix B. These plots were superimposed on the long-term mean head and long-term standard deviation to aid in the detection of long-term trends. Additionally, mean, standard deviation, high and low values were computed for each month of the year over the entire data record, cf. Appendix C. These plots were also superimposed on the long-term mean head and standard deviation to aid in the identification of seasonal trends.

Autocorrelation functions and partial autocorrelation functions of the original time series were computed for each well using the IDENTIFY

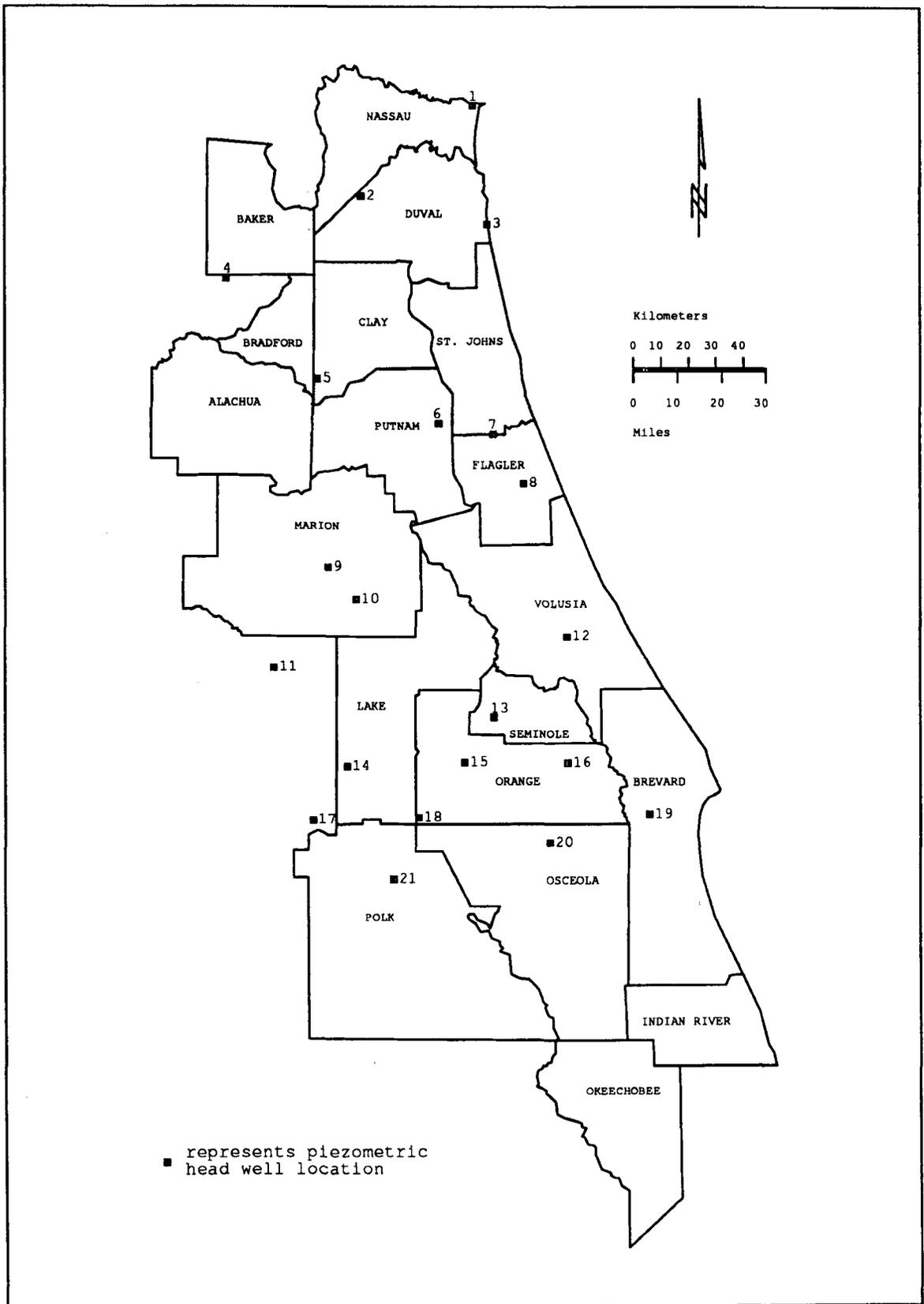


Figure 5.10 Spatial distribution of 21 wells.

Table 5.2 Identification, location, years of data, and number of monthly observations  $n$  for 21 wells provided by the St. Johns River Water Management District.

Well Number	SJRWMD Well Number	Latitude	Longitude	Years of Data	Number of Years	$n$	S/T*
1	N-0003	304210	812708	1977-1990	13	157	T
2	D-0348	302416	815226	1976-1990	13	158	T
3	D-0160	301852	812342	1945-1990	44	529	T
4	U-0001	300747	822258	1959-1982	23	284	T
5	C-0120	294807	820209	1974-1989	14	172	S
6	P-0172	293933	813428	1976-1990	13	163	S
7	SJ-0104	293729	812212	1959-1989	30	368	T
8	F-0087	292750	811520	1937-1990	53	643	T
9	M-0048	291115	815925	1933-1989	56	681	T
10	M-0013	290555	815304	1975-1990	14	175	S
11	SU-0013	285121	821122	1973-1990	16	198	S
12	V-0101	285745	810540	1951-1990	39	469	T
13	S-0125	284147	812202	1953-1990	37	444	T
14	L-0062	283204	815449	1959-1990	31	371	S
15	OR-0047	283252	812835	1943-1990	46	557	T
16	OR-0007	283249	810540	1961-1990	29	347	T
17	SU-0002	282127	820225	1959-1990	31	370	S
18	OR-0064	282202	813846	1959-1990	31	370	T
19	BR-0202	282245	804716	1955-1989	34	406	T
20	OS-0001	281714	810930	1976-1990	14	169	S
21	PO-0006	281008	814418	1960-1990	30	359	S

\*T indicates a long-term trend, S indicates no apparent long-term trend.

statement of the SAS/ETS ARIMA procedure. Spectrums of the original data were produced using the SAS SPECTRA procedure. Inspection of the autocorrelation functions, partial autocorrelation functions, and spectrums was used to confirm the apparent seasonality and long-term trends evident in the original data.

The original data was seasonally detrended by differencing the time series with a period of  $s = 12$  months. Autocorrelation functions, partial autocorrelation functions, and spectra of the differenced time series were then computed for each well. A review of each well's

autocorrelation function and partial autocorrelation function for the differenced data suggested the general stochastic model form to be fit. A cursory review of the data revealed that it was feasible to use the ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model for all 21 wells, providing continuity and ease of modeling for the District. Using the ESTIMATE statement of the SAS/ETS ARIMA procedure, model parameter estimates were computed for each well. Using these parameters, the FORECAST statement produced back forecast estimates of the original data. The estimated model was then verified by checking the white noise properties of the model residuals.

Back forecasts for each well were produced and compared to the actual data record to check the accuracy and usefulness of the estimated model. Plots were developed of one-month, three-month, and six-month lead predictions. Each plot detailed predictions over the ten year period 1980-1990 and showed the predicted head values, the actual head values, and the 95 percent confidence limits for the forecasts.

For those time series which were approximately Gaussian, an excursion analysis was performed for  $h = 1$  and  $h = 2$  standard deviations. Expected drought piezometric head levels for these wells were determined for 30, 20, 10, and 5 year recurrences from the cumulative distribution function of the original data.

## Results and Discussion

### Case Study 1: Development of a Stochastic Model for a "Steady-State" Well

Well number 17 (SJRWMD number SU-0002) is a steady-state well (i.e. a well which shows no evidence of a long-term decline in piezometric head level) located in Sumter County, Florida. Continuous

piezometric head level data were available for this well from April 1959 through January 1990, a period of approximately 31 years. The monthly piezometric head level time series data for this well consisted of  $n = 370$  observations. Figure 5.11 shows a plot of the original time series. Note the absence of an apparent long term trend in the data.

The original data exhibited a seasonal cycle of 12 months, with a subcycle of 6 months, cf. Figure 5.12. Beginning with annual low head levels in May, the levels would rise to an annual peak in September. They would then dip to secondary lows in November, rise to secondary highs in March, before returning once again to yearly lows in May. Table 5.3 lists the minimum, maximum, mean, and standard deviation values for the May piezometric head series, September piezometric head series, and the long-term time series.

Using the SAS/ETS ARIMA and SPECTRA procedures, the autocorrelation function, partial autocorrelation function, and spectrum for the original data were computed. Inspection of the autocorrelation function, Figure 5.13, revealed a sinusoidal wave pattern with particularly strong correlations at 6 month and 12 month lag intervals. This confirmed the apparent seasonality of the original data series. The failure of the autocorrelation function to die out even after an approximately 48 month lag suggests that the process exhibits nonstationarity. Thus, the process must be differenced to produce a stationary time series.

Figure 5.13 also shows the partial autocorrelation function for the original data. The sinusoidal behavior of this function with periodic significant partial autocorrelations at lags up to 48 months suggests the presence of a moving average component in the process.

09-A

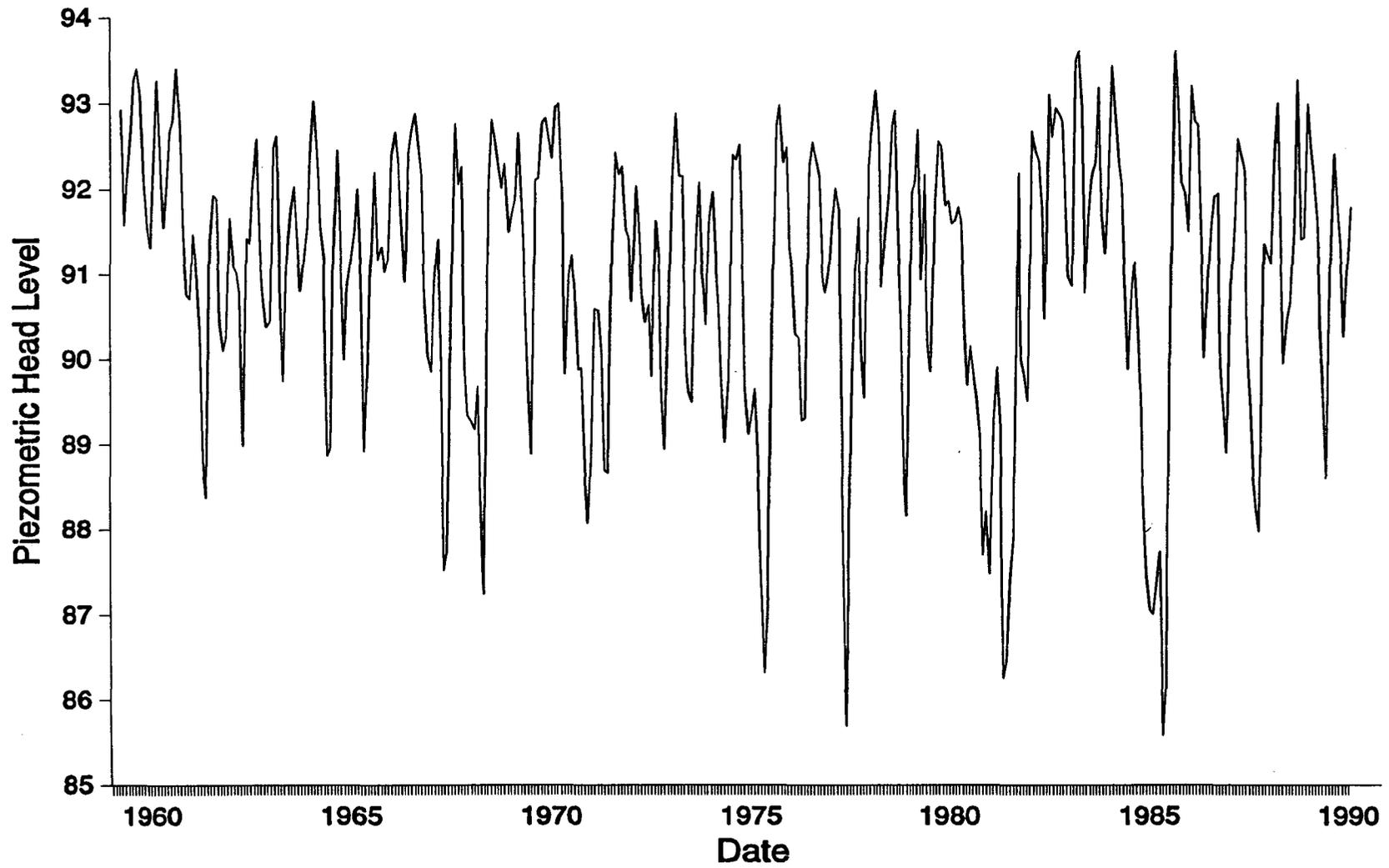


Figure 5.11 Piezometric head levels for Well 17.

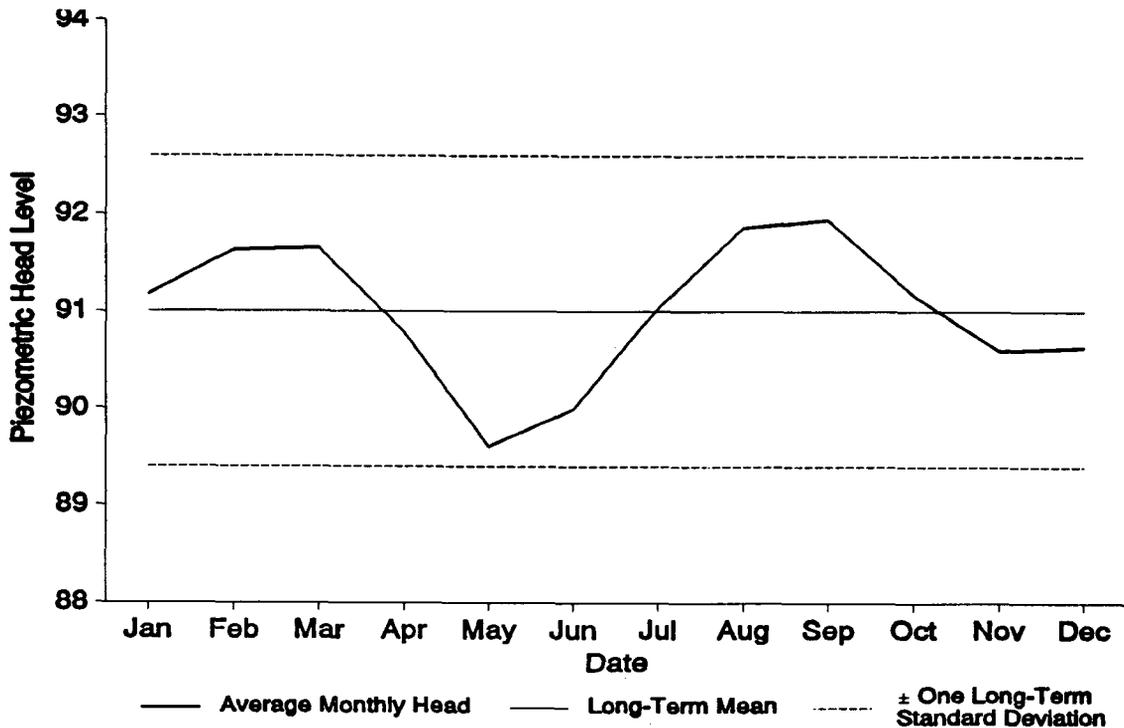


Figure 5.12 Average long-term and monthly head levels for Well 17.

More pronounced, however, is the generalized AR(2) component indicated by the abrupt drop off of the partial autocorrelations after lag 2. Thus a mixed autoregressive moving average model is indicated. Inspection of the spectrum of the original series revealed that it was dominated by low frequencies, confirming the positive correlation between successive lags and indicating a slowly oscillating process series, cf. Figure 5.14. Positive lag correlation indicates a direct

Table 5.3 Minimum, maximum, mean, and standard deviation values for Well 17.

	Minimum	Maximum	Mean	Std Dev
Long-Term Head Level	85.58	93.61	91.00	1.60
May Head Level	85.58	92.23	89.61	1.75
September Head Level	88.23	93.41	91.95	1.17

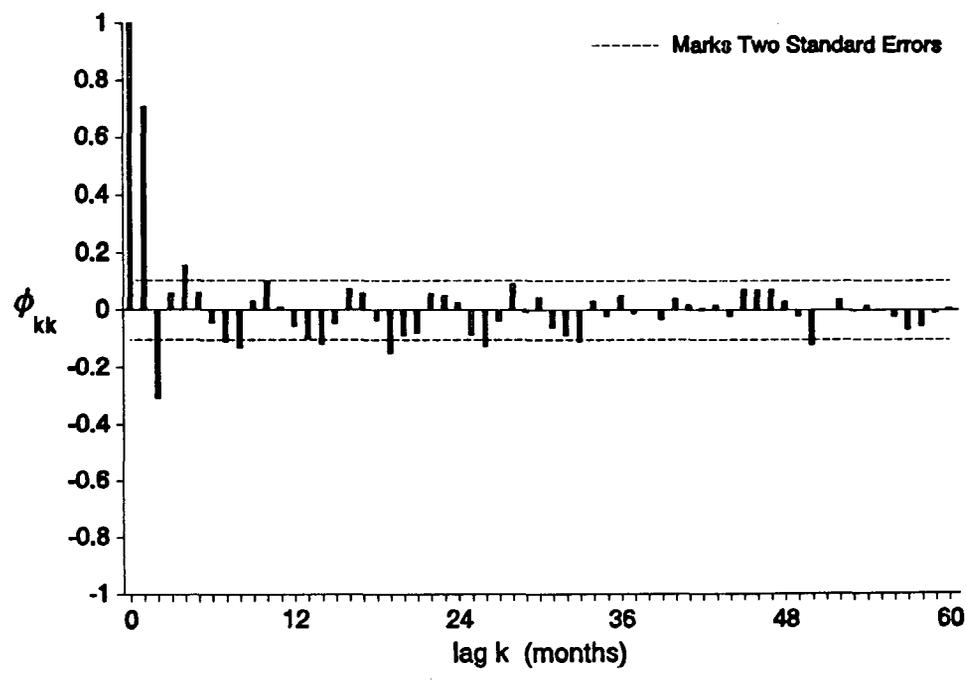
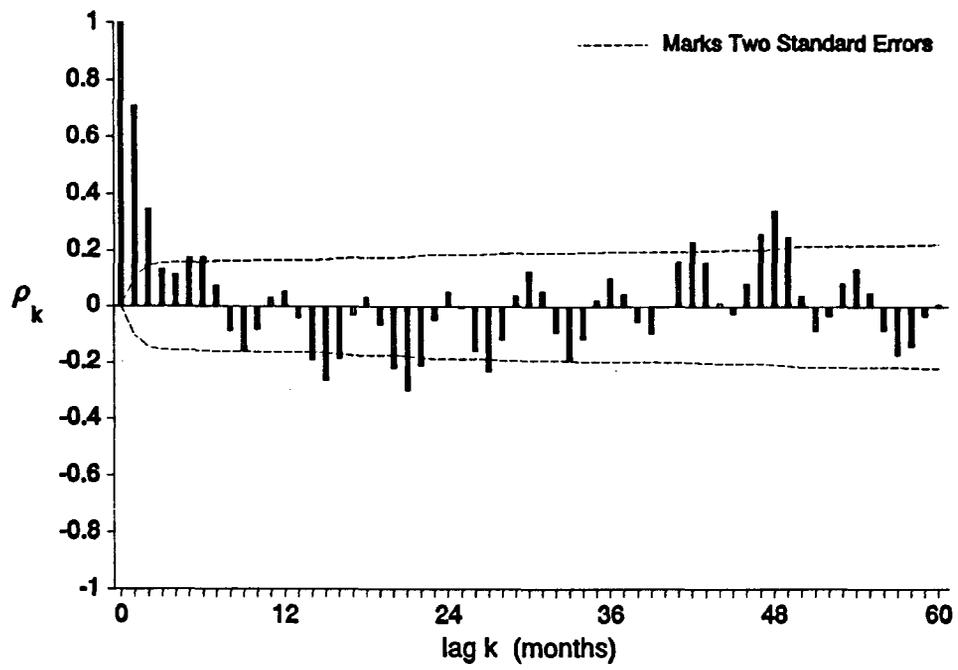


Figure 5.13 Autocorrelation function and partial autocorrelation function for Well 17 original time series.

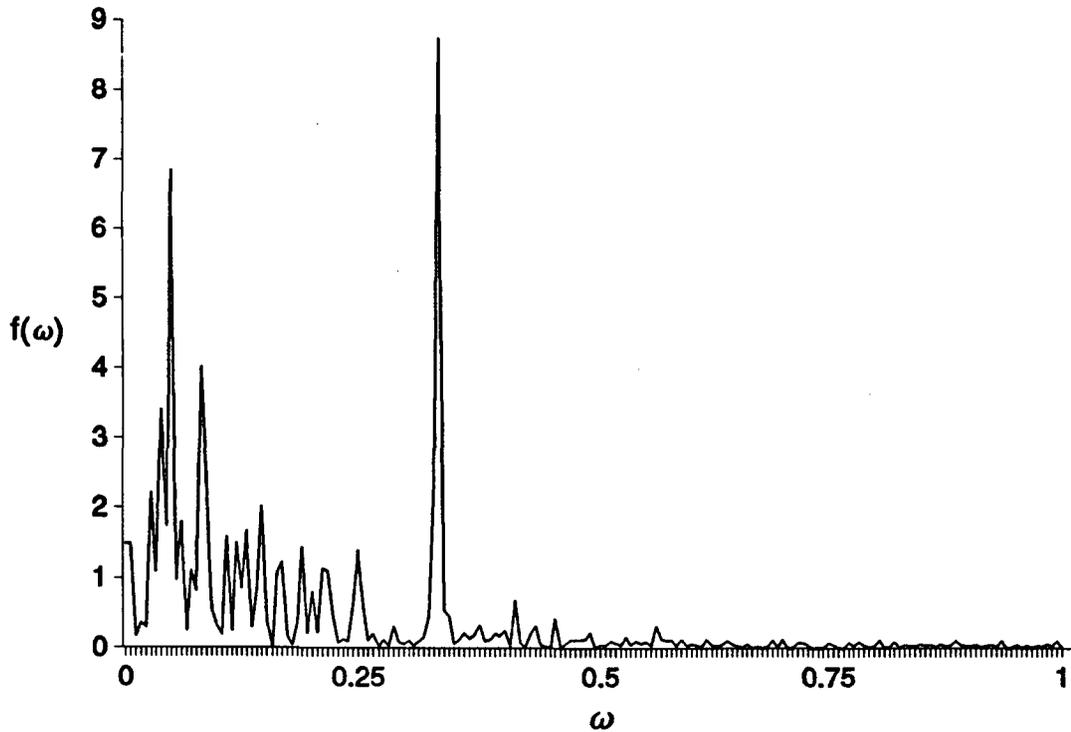


Figure 5.14 Spectrum of Well 17 original time series.

correlation between values  $Z_t$  and  $Z_{t+k}$ , i.e., if  $Z_t$  is less than the mean  $\bar{z}$  then there is a high probability that  $Z_{t+k}$  will be also. Further inspection of the spectrum showed a major peak occurring at a frequency of  $\omega = \pi/\Delta t = 0.333$  corresponding to an interval of 3 months. A peak in the spectrum indicates an important contribution to the process variance at that frequency. Thus, a peak in the spectrum corresponding to a 3 month frequency indicates a cycle of high values followed by low values 3 months later, or a complete cycle every 6 months. This corresponds well with the observed 6 month subcycle of the original data.

The original data were seasonally differenced with a period of  $s = 12$ . This reduced the number of monthly observations for the differenced time series to  $n = 358$ . A plot of the differenced time series is shown in Figure 5.15.

V-64

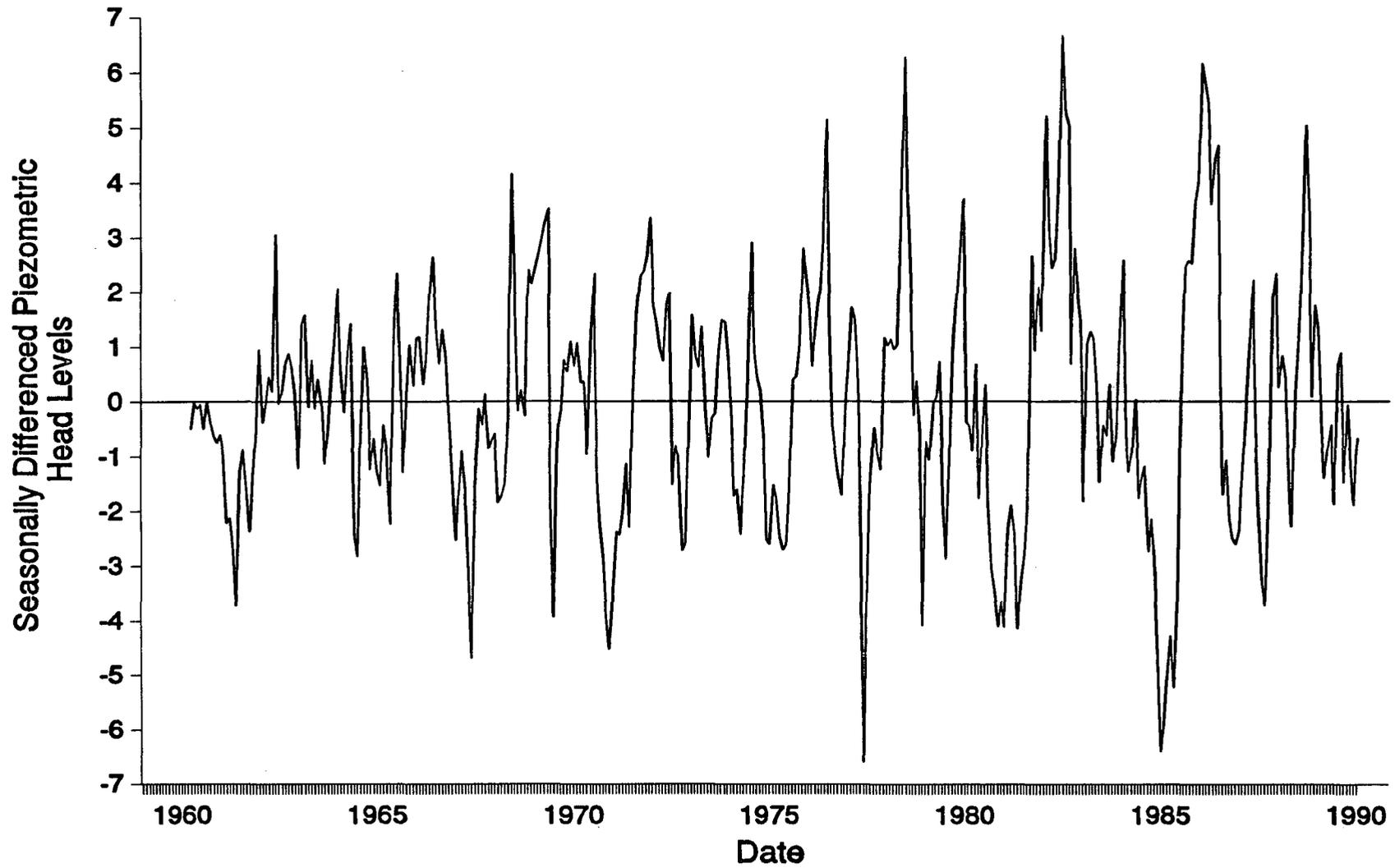


Figure 5.15 Seasonally differenced piezometric head levels for Well 17.

Figure 5.16 shows the autocorrelation function and partial autocorrelation function for the differenced time series. The autocorrelation plot for the differenced series is much smoother than the original series autocorrelation plot, and exhibits a more easily identifiable structure. Note that the autocorrelation function falls off exponentially at first and then tails off in a damped sine wave after lag 12 indicating a mixed autoregressive-moving average process (see Table 5.1). An inspection of the partial autocorrelation function also reveals characteristics of both autoregressive and moving average behavior. First, it appears that an AR(1) or AR(2) component is indicated since correlations drop significantly after lag 1, or possibly lag 2. Additionally, the partial autocorrelation function exhibits a damped exponential decay at lag multiples of approximately 12. This behavior is indicative of a lag-one seasonal moving average component. Thus an  $ARIMA(1,0,0) \times (0,1,1)_{12}$  or an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model is indicated.

Inspection of the spectrum for the differenced time series, cf. Figure 5.17, reveals that the peak at frequency  $\omega = 0.333$ , corresponding to a period of  $\Delta t = 3$  months, has disappeared. This suggests that differencing of the original data has removed this periodic component of the process.

Having established that the general form of the process model was  $ARIMA(p,0,0) \times (0,1,1)_{12}$ , it was decided to model the process using  $p = 2$ . Although the improvements achieved by an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model over an  $ARIMA(1,0,0) \times (0,1,1)_{12}$  would be relatively minor for this well, the more complex model was chosen to be consistent with the fitted models of the other 20 wells.

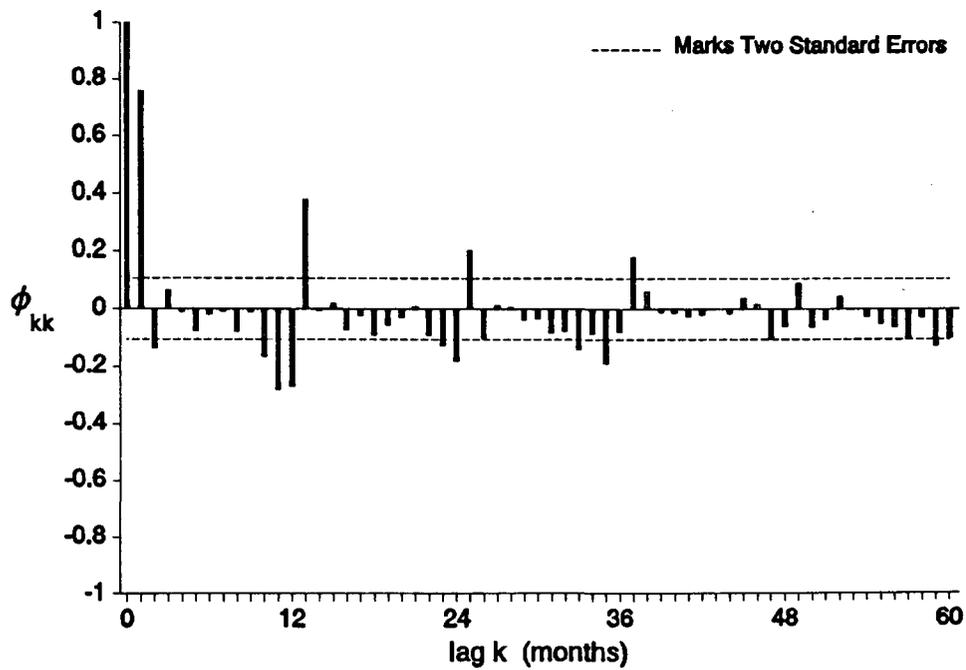
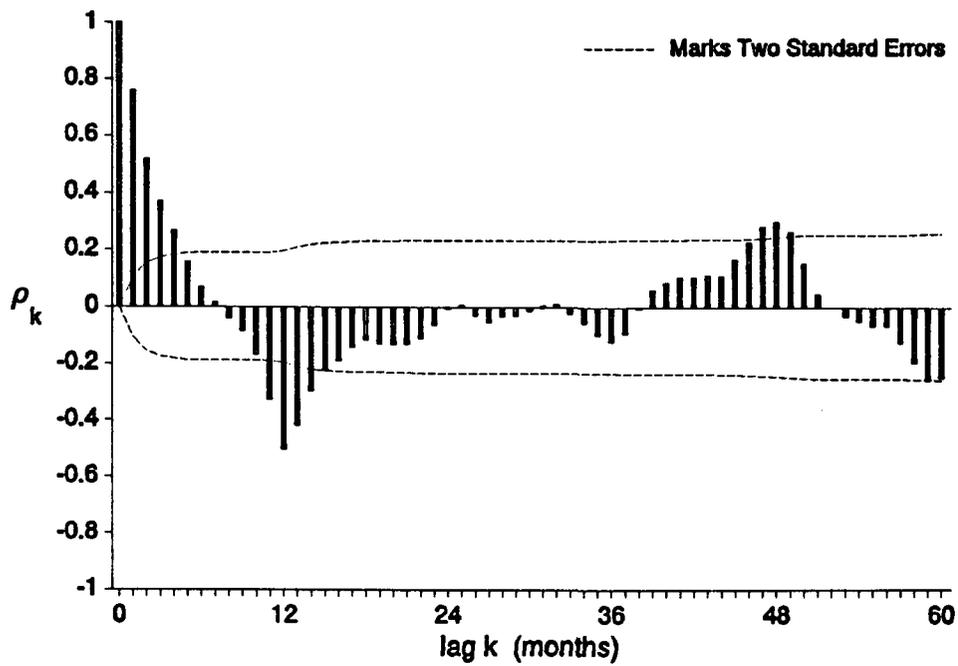


Figure 5.16 Autocorrelation function and partial autocorrelation function for Well 17 seasonally differenced time series.

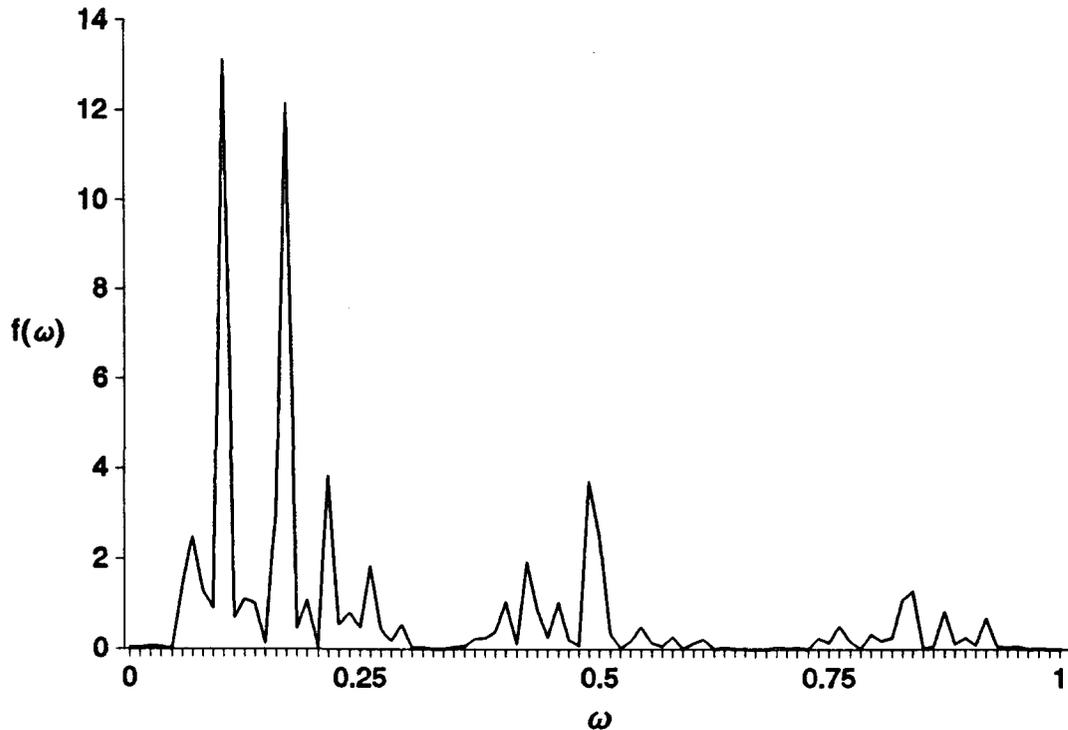


Figure 5.17 Spectrum of Well 17 seasonally differenced time series.

Using the ESTIMATE statement of the SAS/ETS ARIMA procedure, model parameters were estimated for the differenced time series using the conditional least squares technique. The estimated parameters of the ARIMA(2,0,0) $\times$ (0,1,1) $_{12}$  model were  $\hat{\phi}_1 = 0.856$ ,  $\hat{\phi}_2 = -0.113$ ,  $\hat{\theta}_1 = 0.920$ ,  $\hat{\theta}_0 = -0.014$ ,  $\hat{\sigma}_a^2 = 1.008 \text{ ft}^2$ .

To verify that an appropriate model had been identified, one month lead forecasts were generated for the months of May 1960 through January 1990. Model error residuals were then calculated by subtracting the forecast piezometric head level from the observed piezometric head level for each month. If the modeling procedure had extracted all available information from the data series, these model residuals should represent a zero mean white noise (i.e. uncorrelated) process.

Figure 5.18 shows a plot of the model error residuals. Visual inspection of this plot confirmed that the residuals resembled white noise with a sample mean of 0.07 ft and a sample standard deviation of 0.99 ft. The residual autocorrelation function and spectrum were computed to test for correlation, cf. Figure 5.19. The autocorrelation function showed no significant violation of the  $2(1/\sqrt{n})$  limit that would imply correlated residuals. In addition, the spectrum of the residuals showed relatively equal weighting over the frequency distribution. Thus the portmanteau lack-of-fit test did not reject the noncorrelation hypothesis of residuals at the 0.05 level of significance. The residual normality hypothesis was also not rejected using a  $\chi$ -squared distribution test, also at the 0.05 level of significance. Thus, the final form of the fitted ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model was  $\phi(B)\nabla^{12}Z_t = \theta(B^{12})a_t$  or

$$(1 - 0.856B + 0.113B^2)\nabla_{12}Z_t = -0.014 + (1 - 0.920B^{12})a_t. \quad (5.95)$$

Equation (5.95) can be written in a less compact, but more understandable form as:

$$Z_t = Z_{t-12} + 0.856(Z_{t-1} - Z_{t-13}) - 0.113(Z_{t-2} - Z_{t-14}) - 0.014 - 0.920a_{t-12} + a_t. \quad (5.96)$$

Equation (5.96) indicates that the piezometric head level at time  $t$  ( $Z_t$ ) can be predicted based on knowledge of the piezometric head level that occurred twelve months ago ( $Z_{t-12}$ ), as well as knowledge of how different the previous two months piezometric head levels were from their

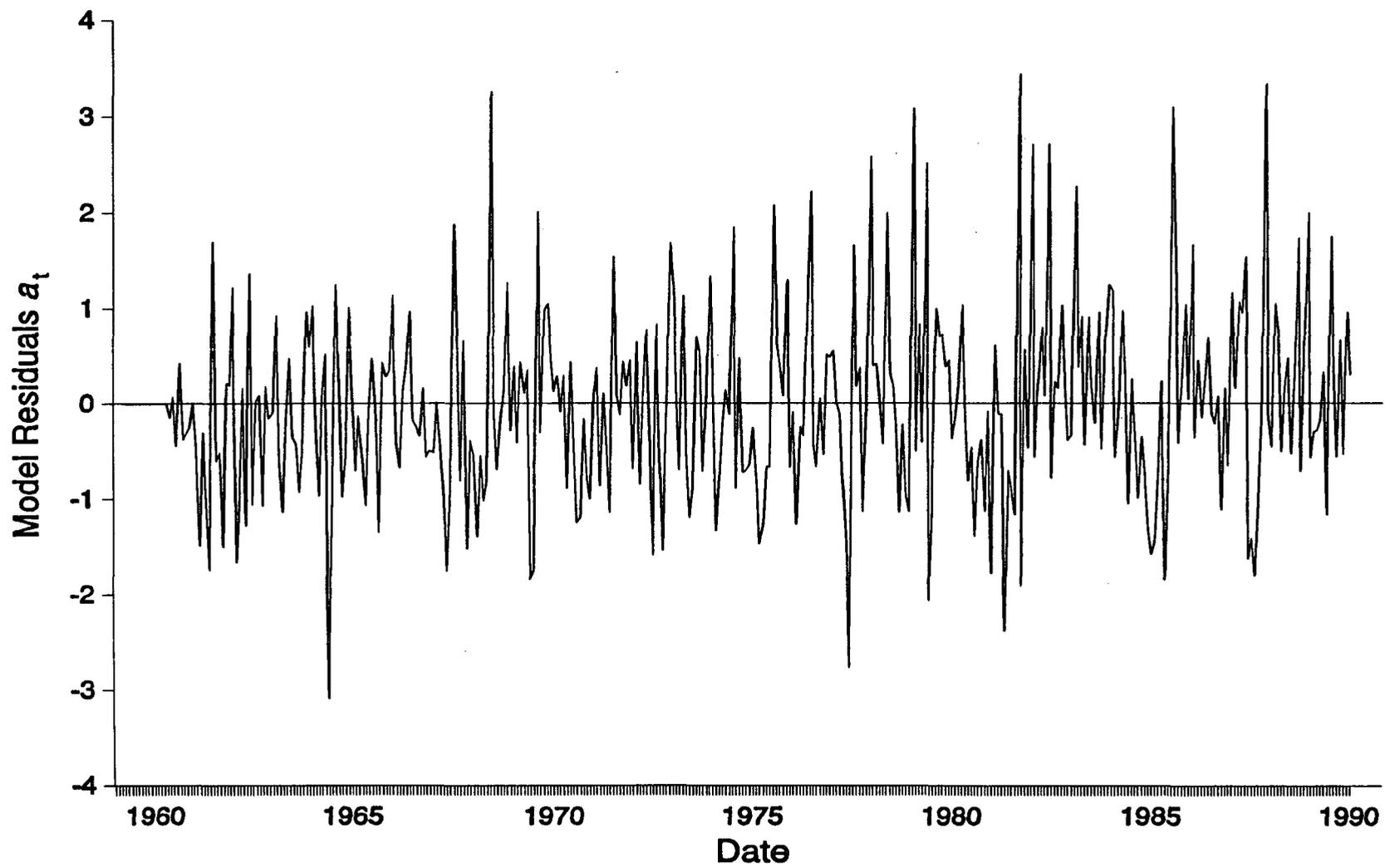


Figure 5.18 Residuals of an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  adjusted to the original series of Well 17.

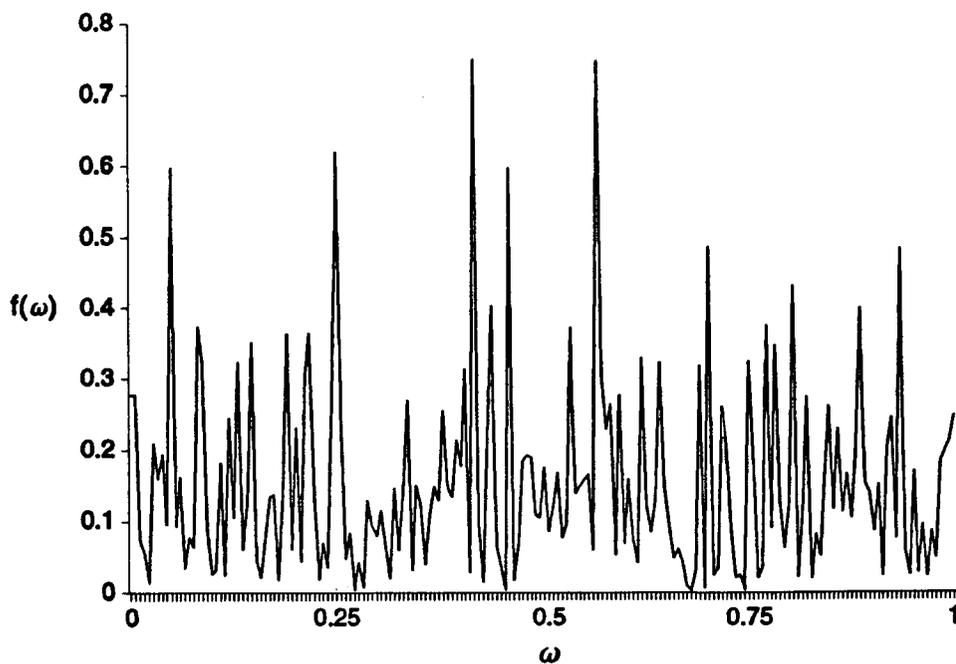
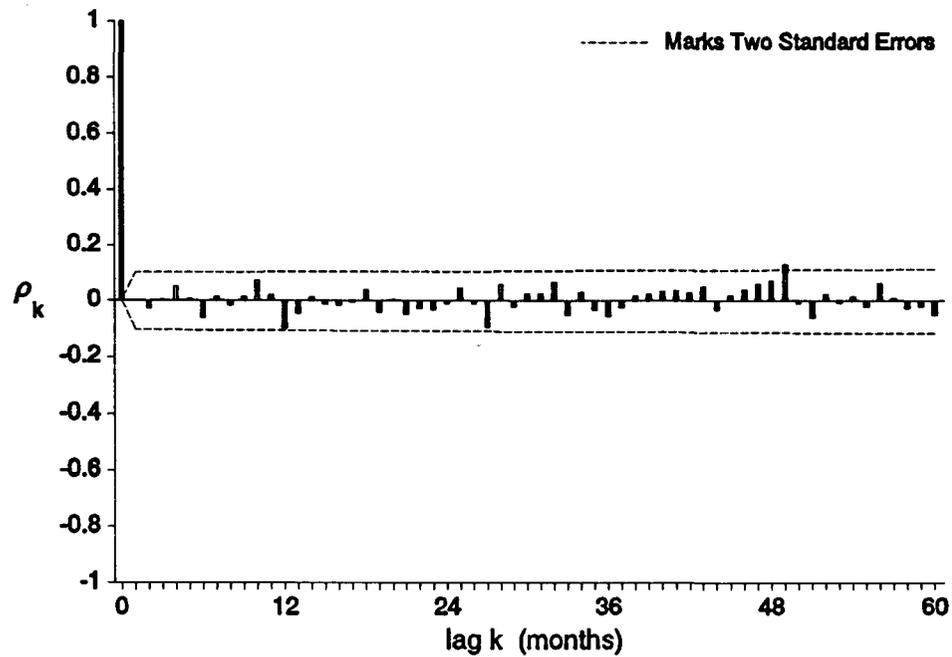


Figure 5.19 Autocorrelation function and spectrum of the residuals of an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  adjusted to the original series of Well 17.

respective values that occurred twelve months prior (i.e.  $(Z_{t-1} - Z_{t-13})$  and  $(Z_{t-2} - Z_{t-14})$ ). The moving average term  $a_{t-12}$  adjusts this prediction based on the actual 1 month forecast error that occurred 12 months ago. The final term in (5.96),  $a_t$ , accounts for the unknown random inputs/outputs to the system at time  $t$ . Possible deviations in inputs/outputs to the groundwater system as a result from lower than normal (or higher than normal) rainfall and/or pumping are represented in this term.

The identified model was used to make lead-1, lead-3, and lead-6 forecasts for Well 17 over a period of 1980-1990. Figures 5.20, 5.21, and 5.22 show these predictions along with the estimated forecast standard deviation. If the model is performing as expected, the model forecasts should fall within one forecast standard deviation of the actual head level 68 percent of the time, and two forecast standard deviations of the actual head level 95 percent of the time. Figures 5.20, 5.21, and 5.22 indicate that this is indeed the case and thus the model's estimate of the distribution of its errors is accurate.

The actual versus predicted model error statistics for this and the other 20 wells are shown in Table 5.6. This table indicates that, for Well 17, one month lead forecasts are expected to be within 1.00 ft of the actual values, three month lead forecasts are expected to be within 1.47 ft of the actual values, and six month lead forecasts are expected to be within 1.60 ft of the actual values.

Due to the fact that it did not exhibit a long-term mean, Well 17 was a candidate for excursion and recursion analyses. To determine if an excursion analysis would be feasible, the original time series data was normalized by subtracting the monthly means and dividing by the monthly

V-72

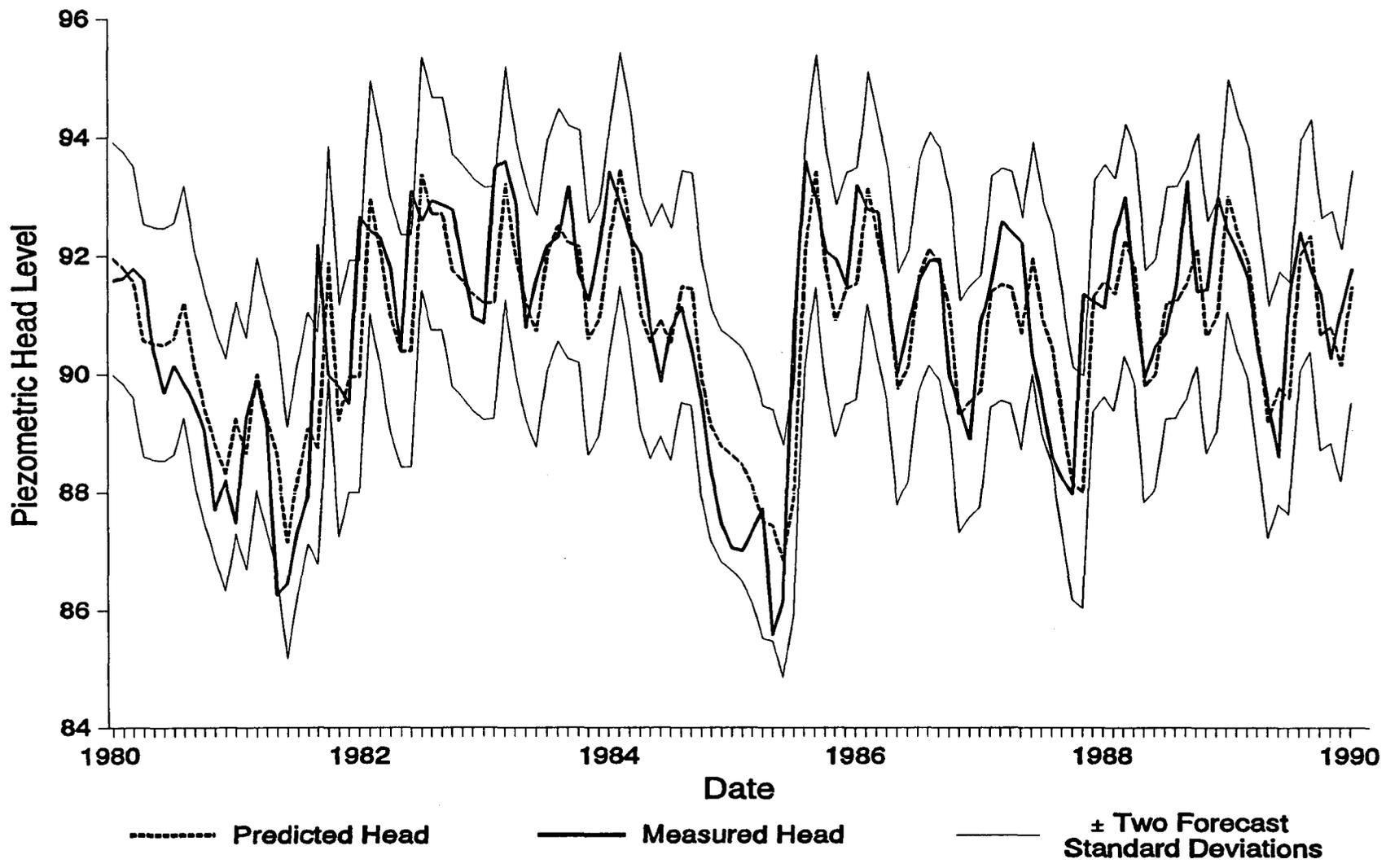


Figure 5.20 One month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 17.

V-73

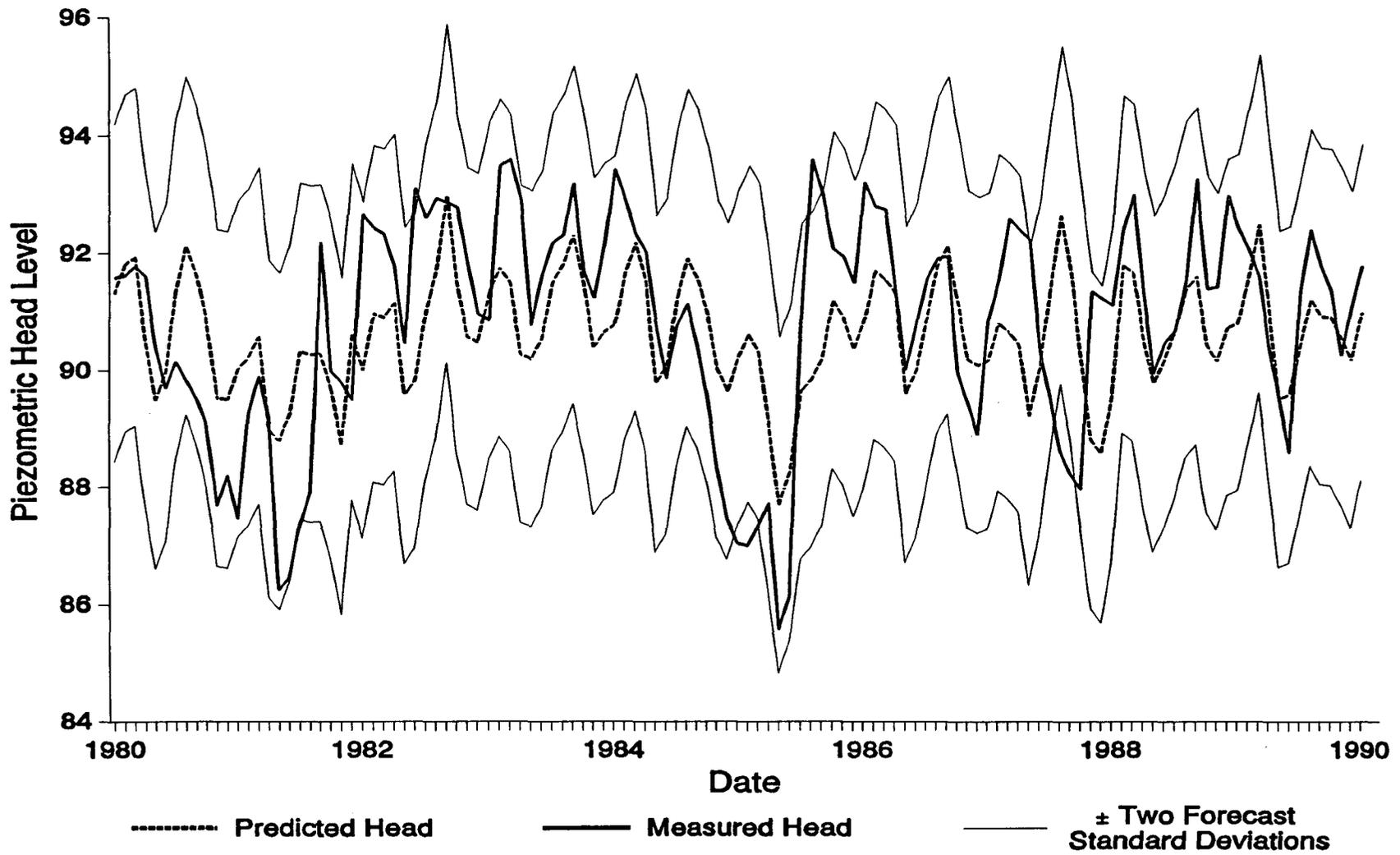


Figure 5.21 Three month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 17.

V-74

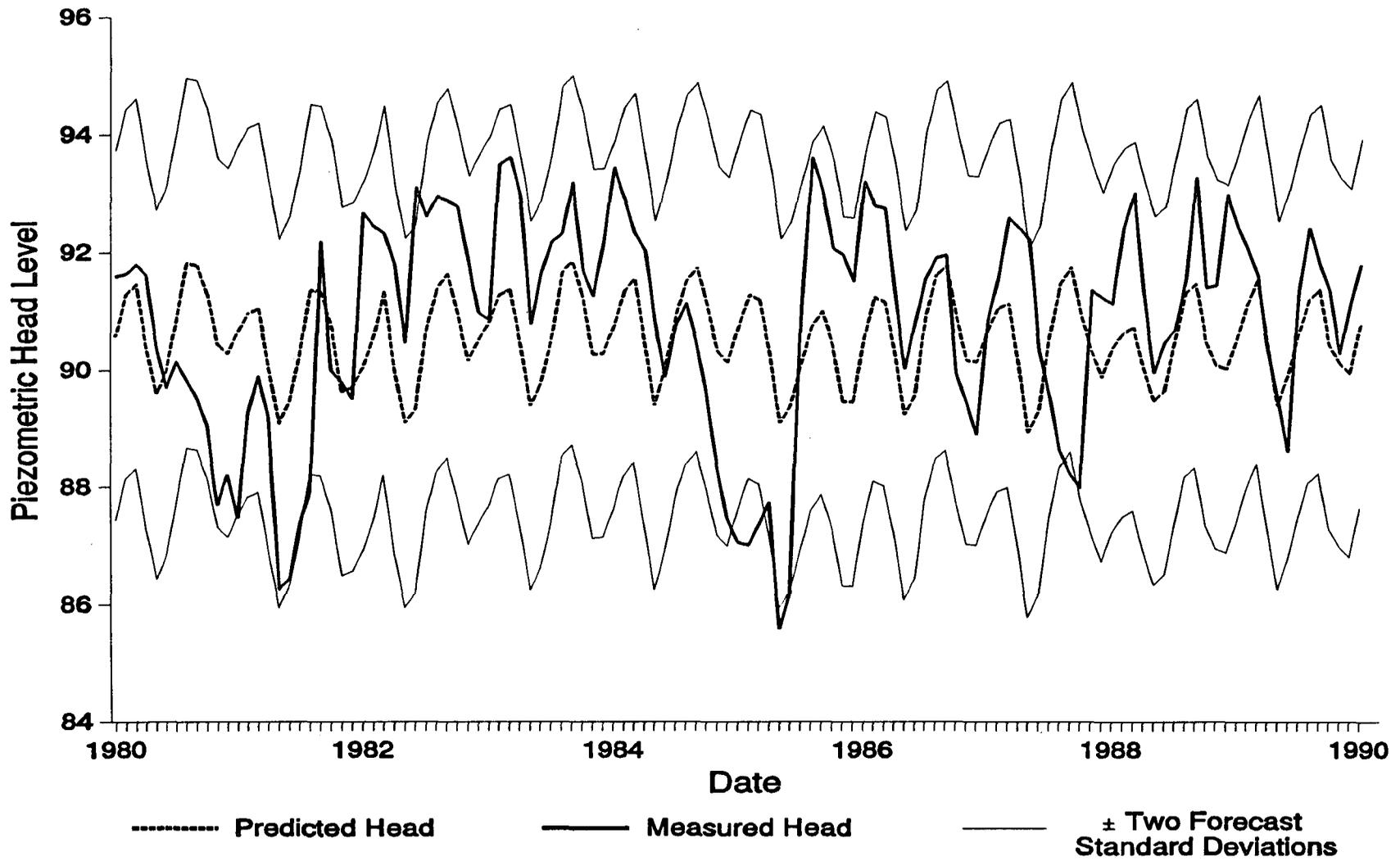


Figure 5.22 Six month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 17.

standard deviations. The Kolmogorov-Smirnoff test for normality of the series was rejected at the 0.05 level of significance, but accepted at the 0.01 significance level. It will be recalled that moderate departures from normality do not seem to greatly influence the applicability of the excursion analysis theory. Thus an excursion analysis was performed on the normalized time series for Well 17.

The normalized time series was found to have a zero mean, a variance of 0.97, and a lag-1 autocorrelation  $\hat{\rho}_1 = 0.775$ . An excursion analysis revealed that it had an expected interval between zero crossings of  $E\{T_0\} = 4.59$  months. Information from an excursion analysis of a normalized series can be transferred to the original series. Thus, in real terms,  $E\{T_0\}$  is simply the average time interval (4.59 months) that the head level in Well 17 can be expected to remain below (or above) a mean level of 91.00 ft.

The expected number of excursions below  $h = 1$  standard deviation was determined to be  $E\{N\} = 0.066$  per month, or one every 15.15 months, with an expected duration of  $E\{T_h\} = 2.40$  months. In real terms, the head level in Well 17 can be expected to fall below a level of 89.40 ft (long-term mean minus one standard deviation) once every 15.15 months, remaining below this level for 2.40 months. For  $h = 2$  standard deviations, the head level in Well 17 can be expected to fall below a level of 87.80 ft (long-term mean minus two standard deviations) once every 67.88 months ( $E\{N\} = 0.015$  per month), remaining below this level for  $E\{T_h\} = 1.55$  months.

A recursion analysis performed on Well 17 revealed that the piezometric head level can be expected to drop below 85.58 ft once every 30 years, to drop below 85.69 ft once every 20 years, to drop below

86.16 ft once every 10 years, and to drop below 86.44 ft once every 5 years. Tables 5.7 and 5.8 tabulate the excursion and recursion analyses results of this well and the other 7 steady-state wells.

Case Study 2: Development of a Stochastic Model for a Well with a Long-Term Trend

Well number 13 (SJRWMD number S-0125), located in Seminole County, Florida, displays a long-term downward trend. Continuous piezometric head level data were available for this well from January 1953 through December 1989, a period of 37 years. The monthly piezometric time series data for this well consisted of  $n = 444$  observations. Figure 5.23 shows a plot of the original monthly time series. The series displays a marked downward trend of approximately 0.25 ft per year.

A plot of the average monthly versus average long-term head levels reveals a seasonal cycle of 12 months, cf. Figure 5.24. The cycle begins with yearly low head levels in May, rises to a yearly peak in September, and then drops again to lows in May. Table 5.4 lists the minimum, maximum, mean, and standard deviation values for the May series, September series, and the long-term time series. The standard deviation values are quite large due to the long-term downward trend.

Using the SAS/ETS ARIMA and SPECTRA procedures, the autocorrelation function, partial autocorrelation function, and spectrum for the original data were computed. The autocorrelations of the series, shown in Figure 5.25, suggest that at least one seasonal differencing is necessary due to the failure of the function to die out rapidly. The pronounced wave pattern has an interval of 12 lags, confirms a seasonal cycle of 12 months.

V-77

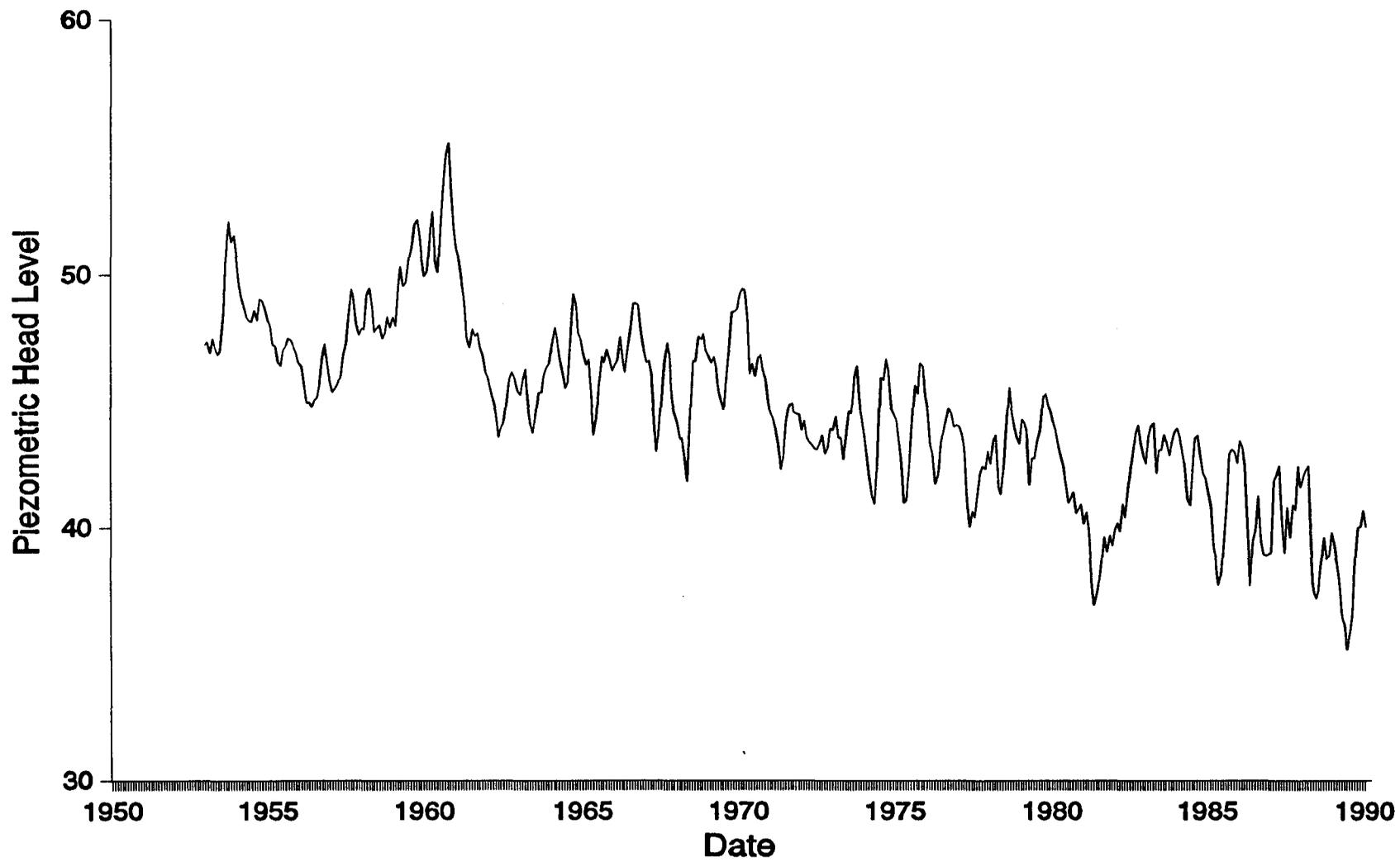


Figure 5.23 Piezometric head levels for Well 13.

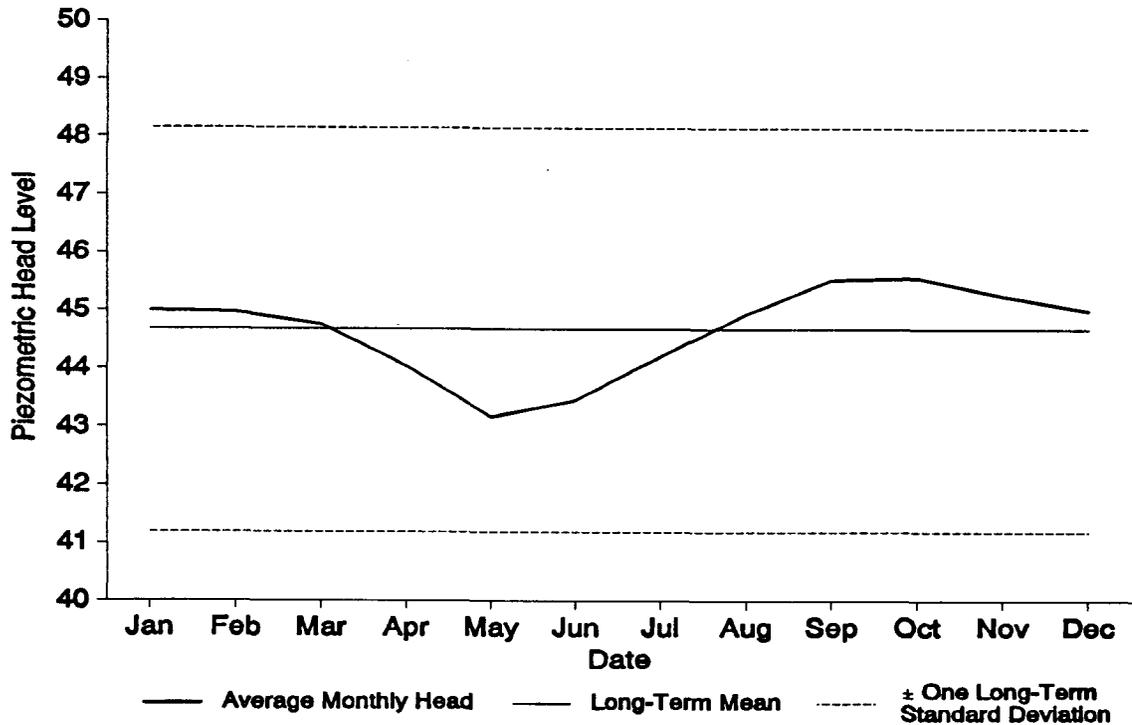


Figure 5.24 Average yearly and monthly head levels for Well 13.

Figure 5.25 also shows the partial autocorrelation function for the original data. The abrupt cut off of the partial autocorrelations after lag 2 would suggest a generalized AR(2) model. The possible presence of a moving average component in the process is also suggested by the periodic significant partial autocorrelations at large lags.

Figure 5.26 shows the spectrum of the original series. Once again, the spectrum is dominated by low frequencies. This low frequency

Table 5.4 Minimum, maximum, mean, and standard deviation values for Well 13.

	Minimum	Maximum	Mean	Std Dev
Long-Term Head Level	35.15	55.19	44.67	3.48
May Head Level	36.14	50.59	43.17	3.62
September Head Level	38.72	54.79	45.53	3.58

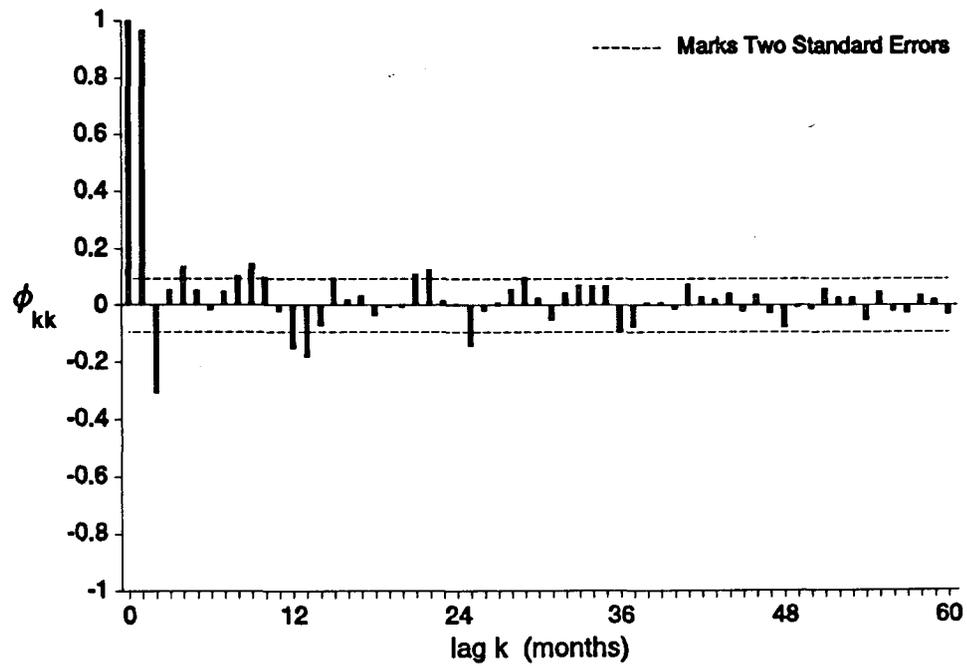
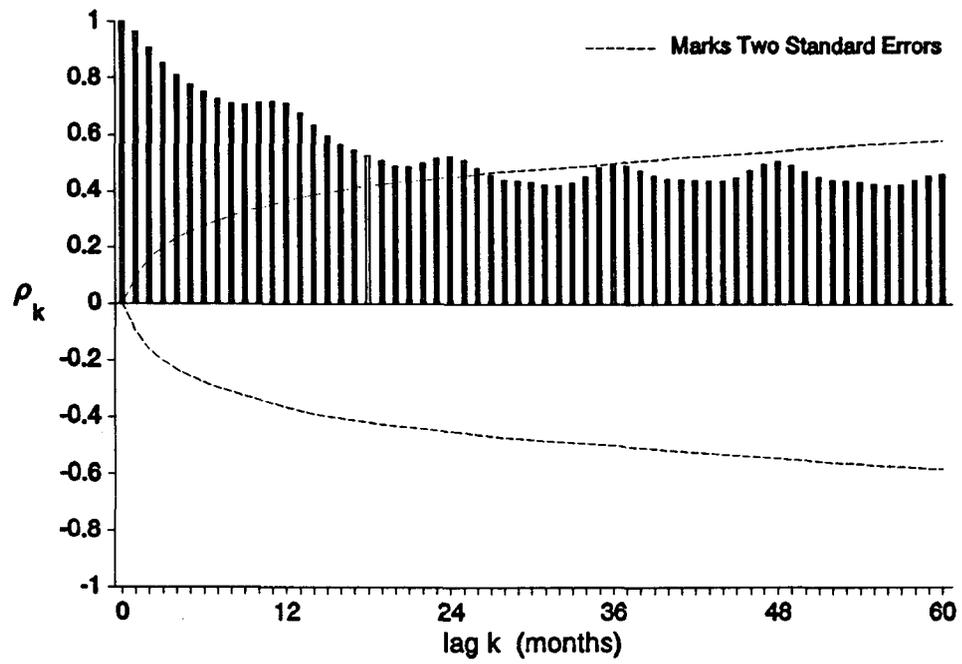


Figure 5.25 Autocorrelation function and partial autocorrelation function for Well 13 original time series.

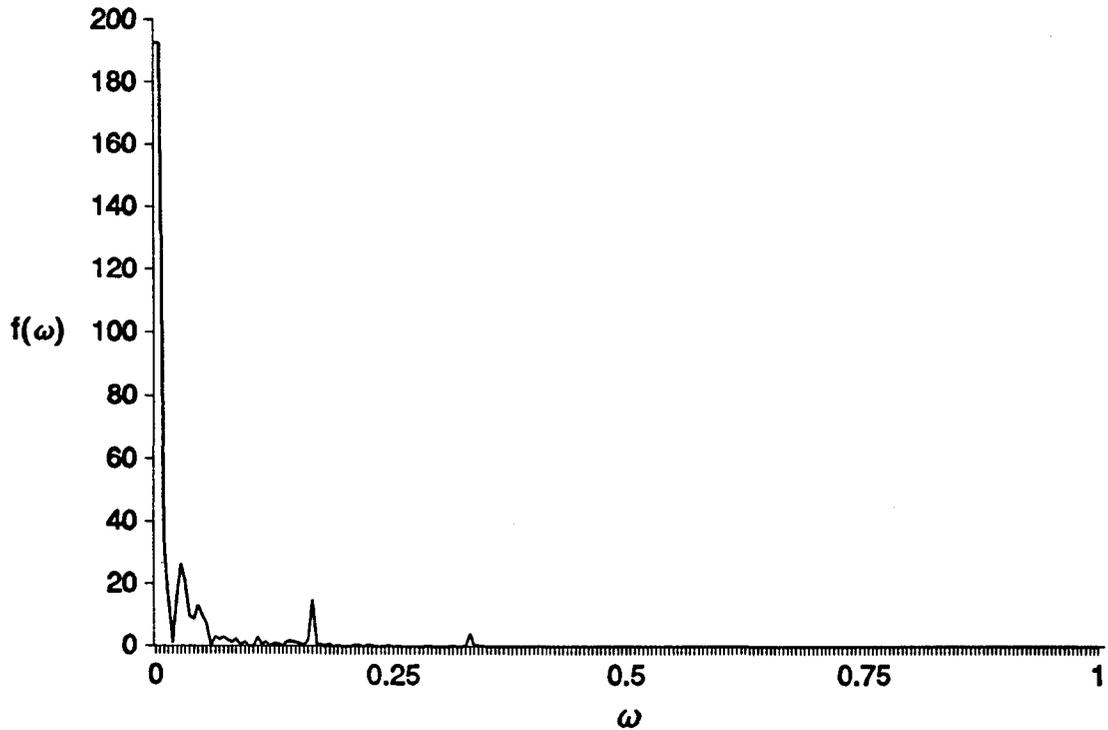


Figure 5.26 Spectrum of Well 13 original time series.

domination is particularly pronounced due to the downward trend of the series, i.e., the covariance of data values separated by  $\Delta t$  decays very slowly as  $\Delta t$  increases. The spectrum dies off toward higher frequencies but is not negligible at frequencies  $\omega$  of 0.167 and 0.333, corresponding to periods of 6 and 3 months, respectively.

The original data was seasonally differenced with a period of  $s = 12$ . This reduced the number of monthly observations for the differenced time series to  $n = 432$ . A plot of the differenced time series is shown in Figure 5.27. Note that seasonally differencing the time series over a 12 month period has effectively removed the long-term as well as seasonal trends.

Figure 5.28 shows the autocorrelation function and partial autocorrelation function for the differenced time series. Note that the

18-1

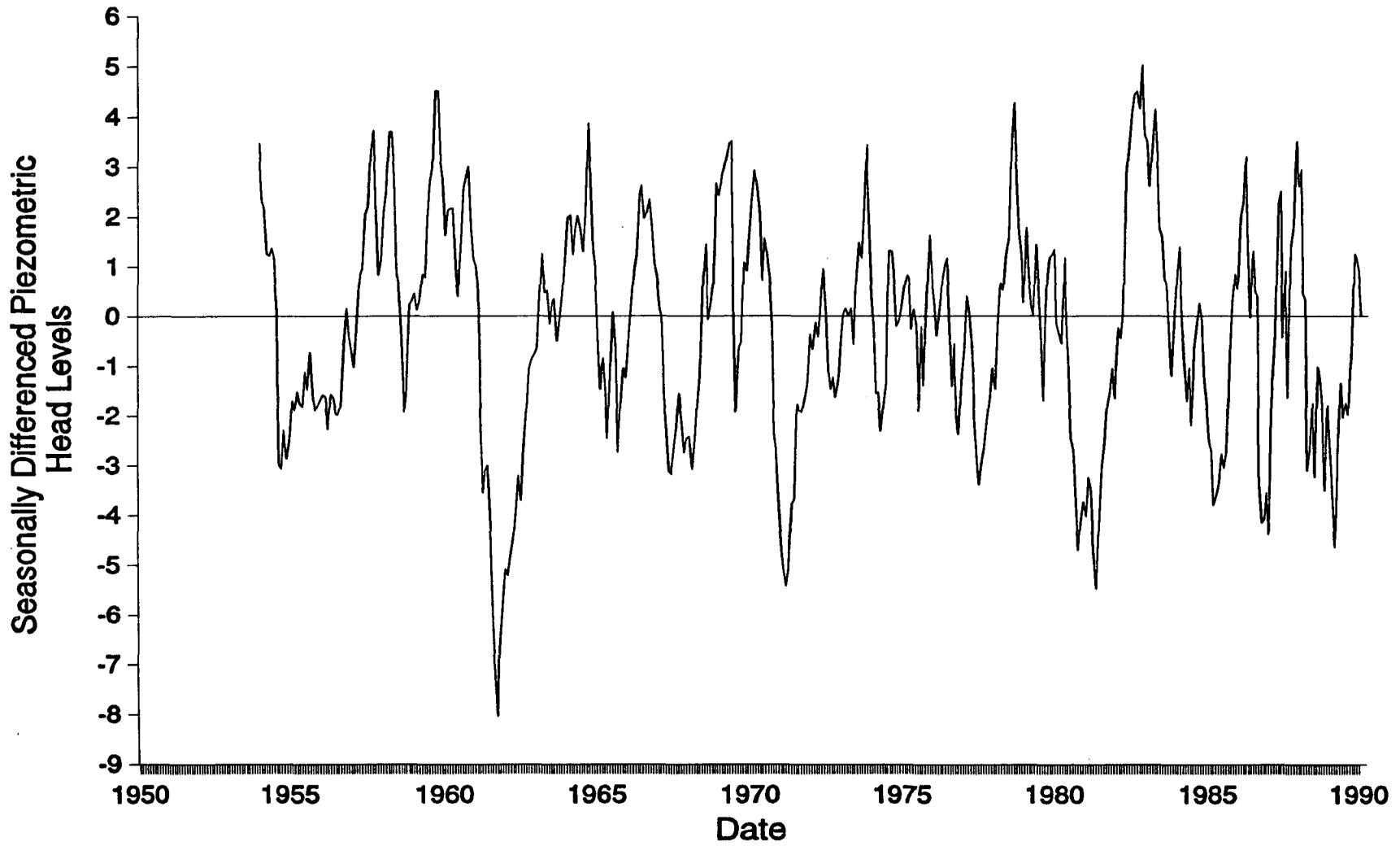


Figure 5.27 Seasonally differenced piezometric head levels for Well 13.

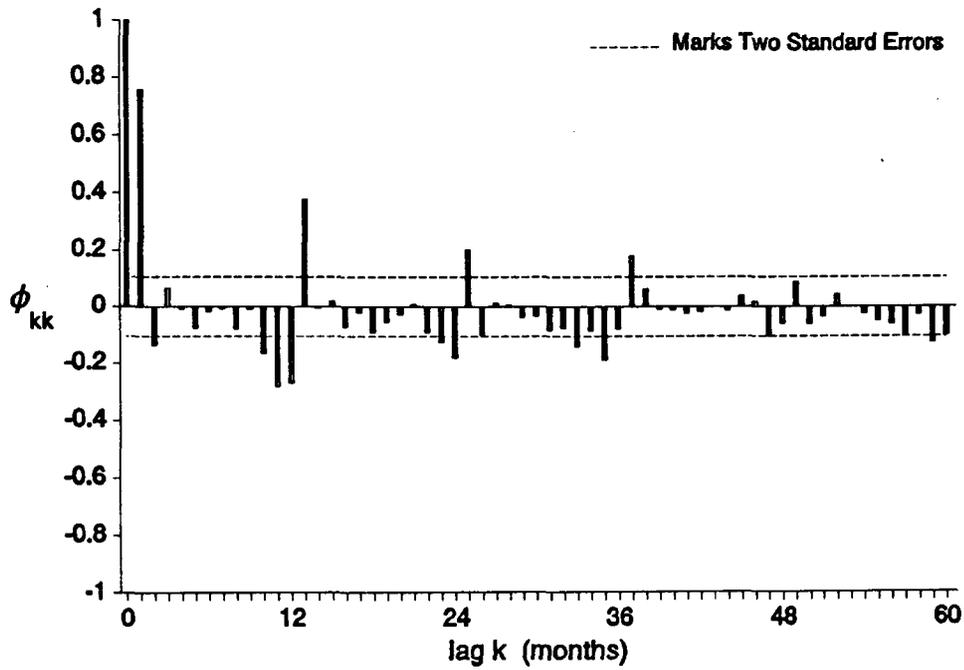
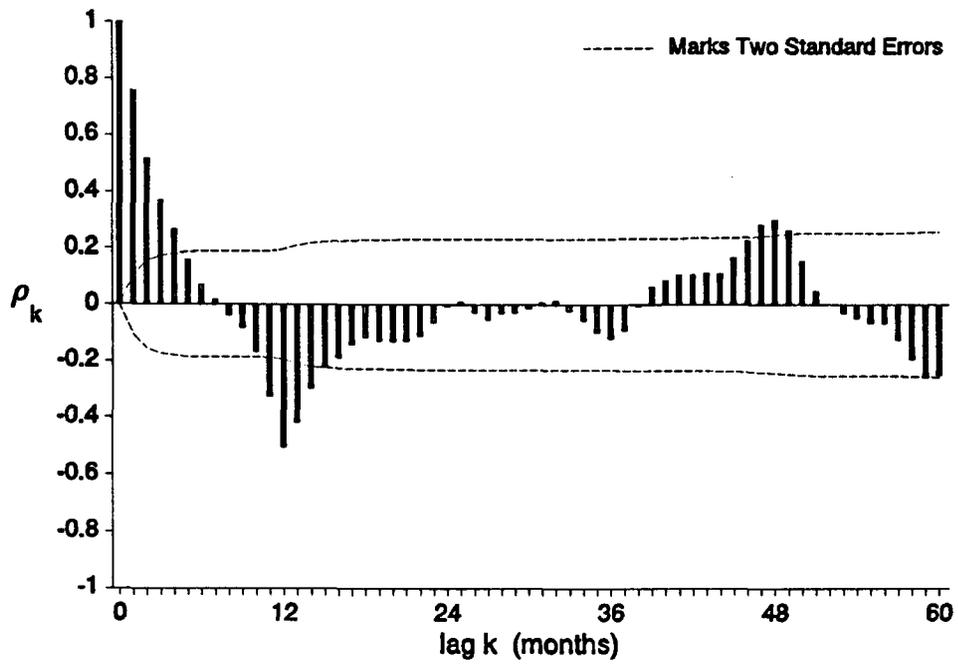


Figure 5.28 Autocorrelation function and partial autocorrelation function for Well 13 seasonally differenced time series.

autocorrelation function falls off exponentially at first and then tails off in a damped sine wave after lag 12. An inspection of the partial autocorrelation function reveals both autoregressive and moving average behavior. First, it appears that an AR(2) component is indicated since correlations drop off quickly after lag 2. Additionally, the partial autocorrelation function exhibits a damped exponential decay at lag multiples of 12, indicative of a lag-one seasonal moving average component. Thus an ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model is indicated. Inspection of the spectrum for the differenced time series, cf. Figure 5.29, reveals that the peaks at frequencies  $\omega$  of 0.167 and 0.333 have disappeared. This suggests that differencing of the original data has removed this periodic component of the process.

Using the ESTIMATE statement of the SAS/ETS ARIMA procedure, model parameters were estimated for the differenced time series using the

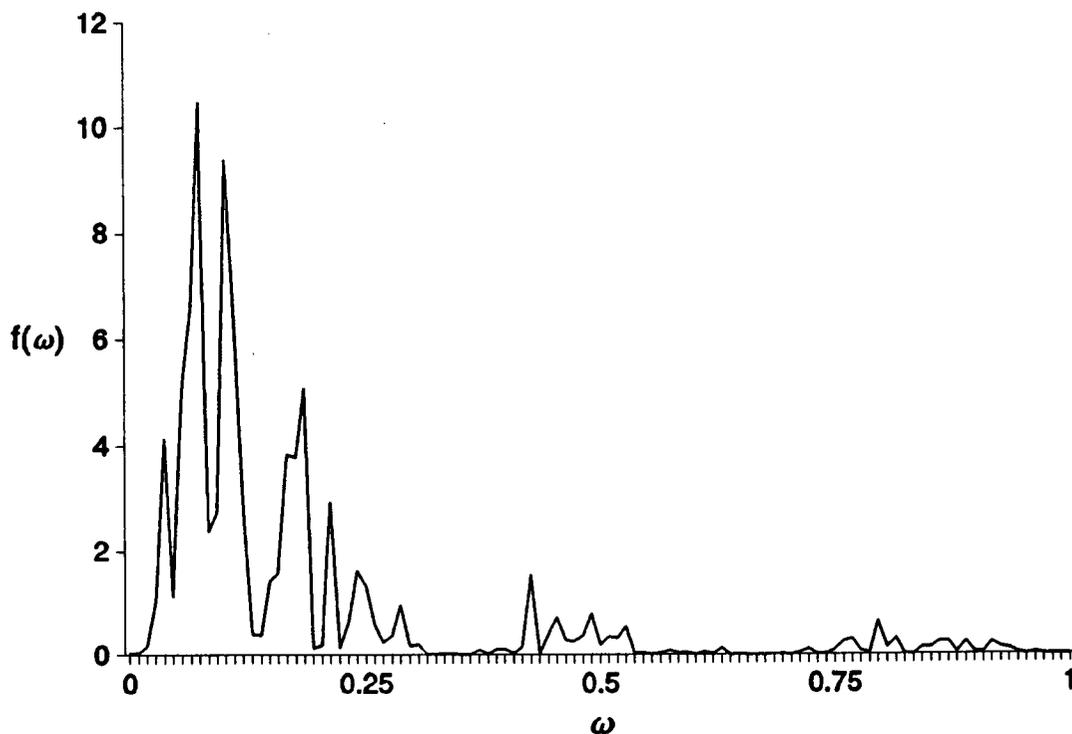


Figure 5.29 Spectrum of Well 13 seasonally differenced time series.

conditional least squares technique. The estimated parameters of the ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model were  $\hat{\phi}_1 = 1.050$ ,  $\hat{\phi}_2 = -0.136$ ,  $\hat{\theta}_1 = 0.683$ ,  $\hat{\theta}_0 = -0.002$ ,  $\hat{\sigma}_a^2 = 0.679 \text{ ft}^2$ .

To verify the model structure, one month lead forecasts were generated for the months of March 1954 through December 1989. Model error residuals were then calculated by subtracting the forecast piezometric head level from the observed piezometric head level for each month. Figure 5.30 shows a plot of the model error residuals. Visual inspection of this plot confirmed the white noise behavior of the residuals with a sample mean of -0.07 ft and a sample standard deviation of 0.76 ft.

The residual autocorrelation function and spectrum of residuals were computed to test for correlation, cf. Figure 5.31. The autocorrelation function showed no significant violation of the  $2(1/\sqrt{n})$  limit that would imply correlated residuals. Additionally, the spectrum of the residuals showed relatively equal weighting over the entire frequency distribution. The portmanteau lack-of-fit test did not reject the noncorrelation hypothesis of residuals at the 0.05 level of significance. The normality hypothesis on residuals was also not rejected using a  $\chi$ -squared test, also at the 0.05 level of significance. Thus the fitted ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model was not rejected and had a final form of  $\phi(B)\nabla^{12}Z_t = \theta(B^{12})a_t$  or

$$(1 - 1.050B + 0.136B^2)\nabla_{12}Z_t = -0.002 + (1 - 0.638B^{12})a_t. \quad (5.97a)$$

$$Z_t = Z_{t-12} + 1.050(Z_{t-1} - Z_{t-13}) - 0.136(Z_{t-2} - Z_{t-14}) - 0.002 - 0.638a_{t-12} + a_t. \quad (5.97b)$$

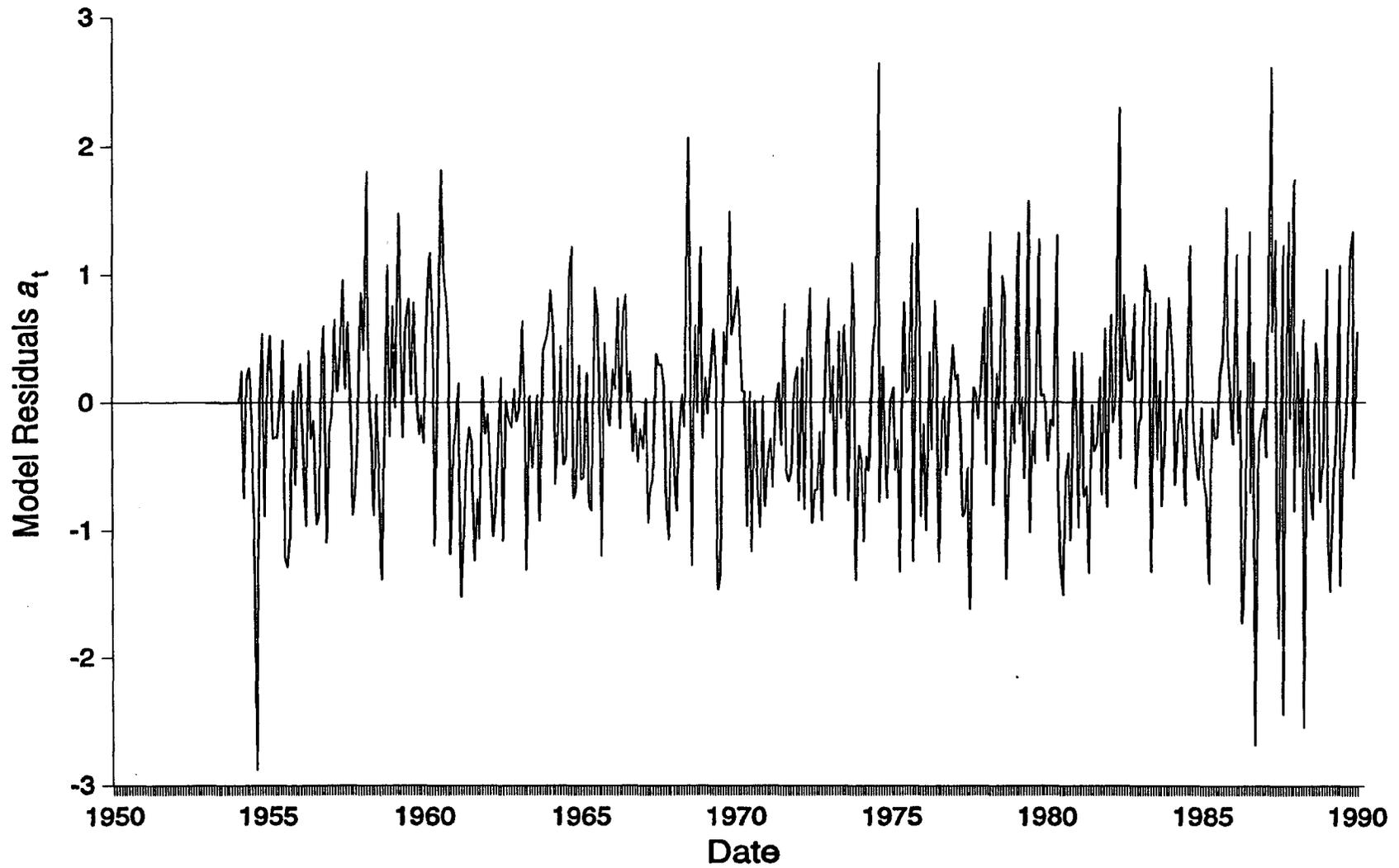


Figure 5.30 Residuals of an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  adjusted to the original series of Well 13.

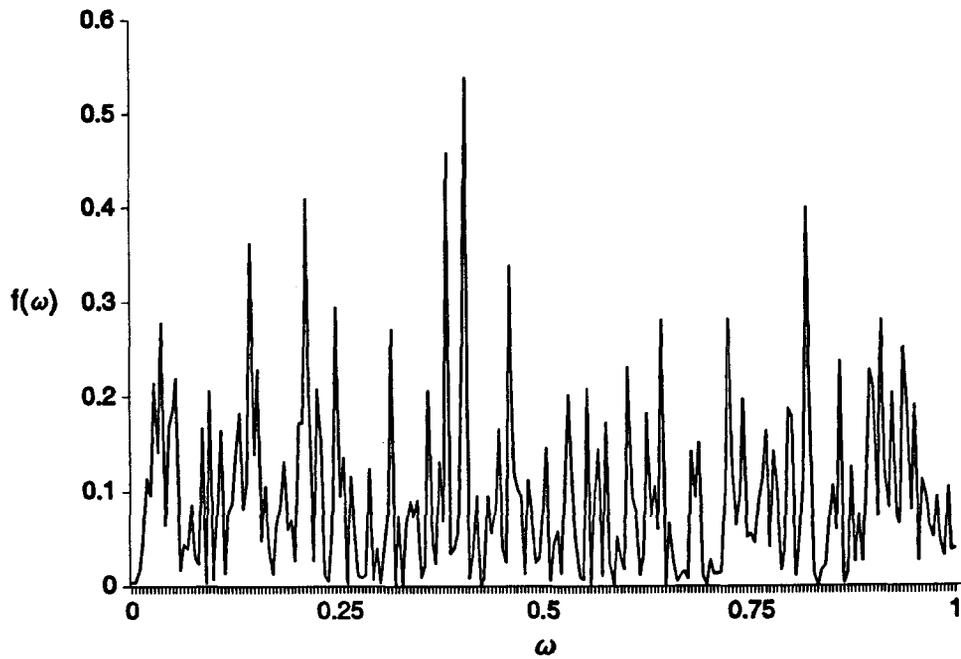
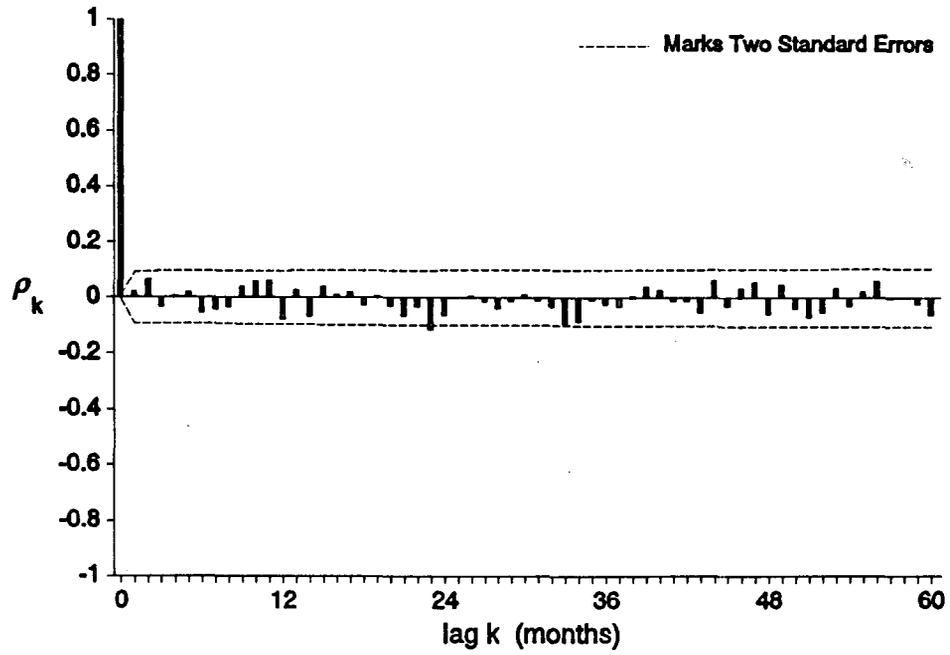


Figure 5.31 Autocorrelation function and spectrum of the residuals of an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  adjusted to the original series of Well 13.

As discussed in Case Study 1, The expanded form of (5.97) indicates that the piezometric head level at time  $t$  can be predicted based on knowledge of the piezometric head level that occurred twelve months ago, as well as knowledge of how different the previous two months piezometric head levels were from their respective values that occurred twelve months prior. The moving average term  $a_{t-12}$  adjusts this prediction based on the actual 1 month forecast error that occurred 12 months ago. The final term in (5.97b),  $a_t$ , accounts for the unknown random inputs/outputs (i.e. rainfall, pumping, etc.) to the system at time  $t$ .

The identified model was used to make lead-1, lead-3, and lead-6 forecasts for Well 13 over a period of 1980-1990. Figures 5.32, 5.33, and 5.34 show these forecasts along with the estimated forecast standard deviation. Table 5.6 tabulates the actual versus model error statistics for this and the other 20 wells. This table indicates that, for Well 13, one month lead forecasts are expected to be within 0.82 ft of the actual values, three month lead forecasts are expected to be within 1.43 ft of the actual values, and six month lead forecasts are expected to be within 1.83 ft of the actual values.

Due to its pronounced long-term trend, Well 13 was not a candidate for excursion and recursion analyses.

#### Stochastic Modeling Results for all 21 Wells

Following the general procedures outlined in the previous two case studies, an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model was fitted to each of the 21 wells. It will be recalled that an  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model has the form

88-A

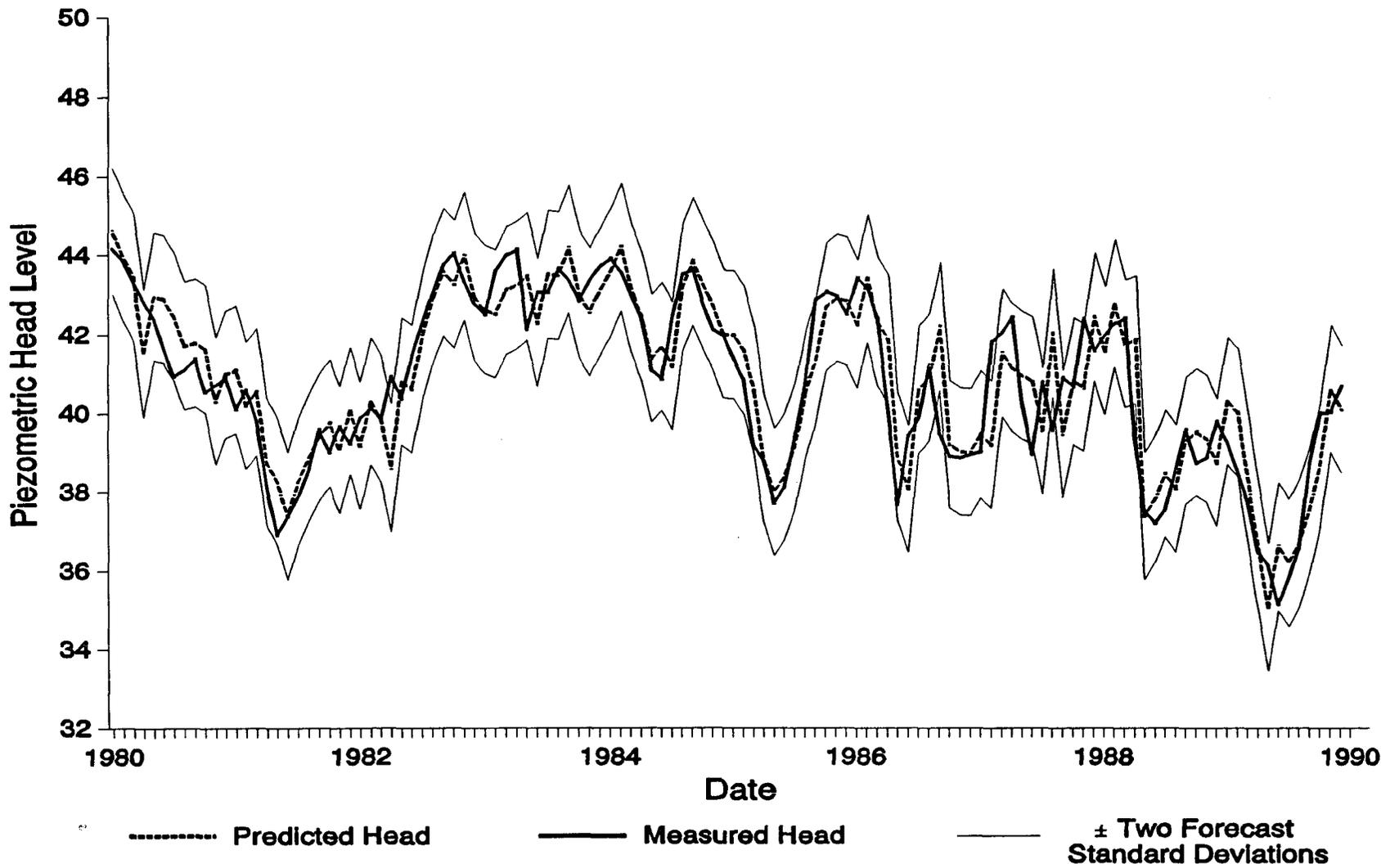


Figure 5.32 One month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 13.

68-1

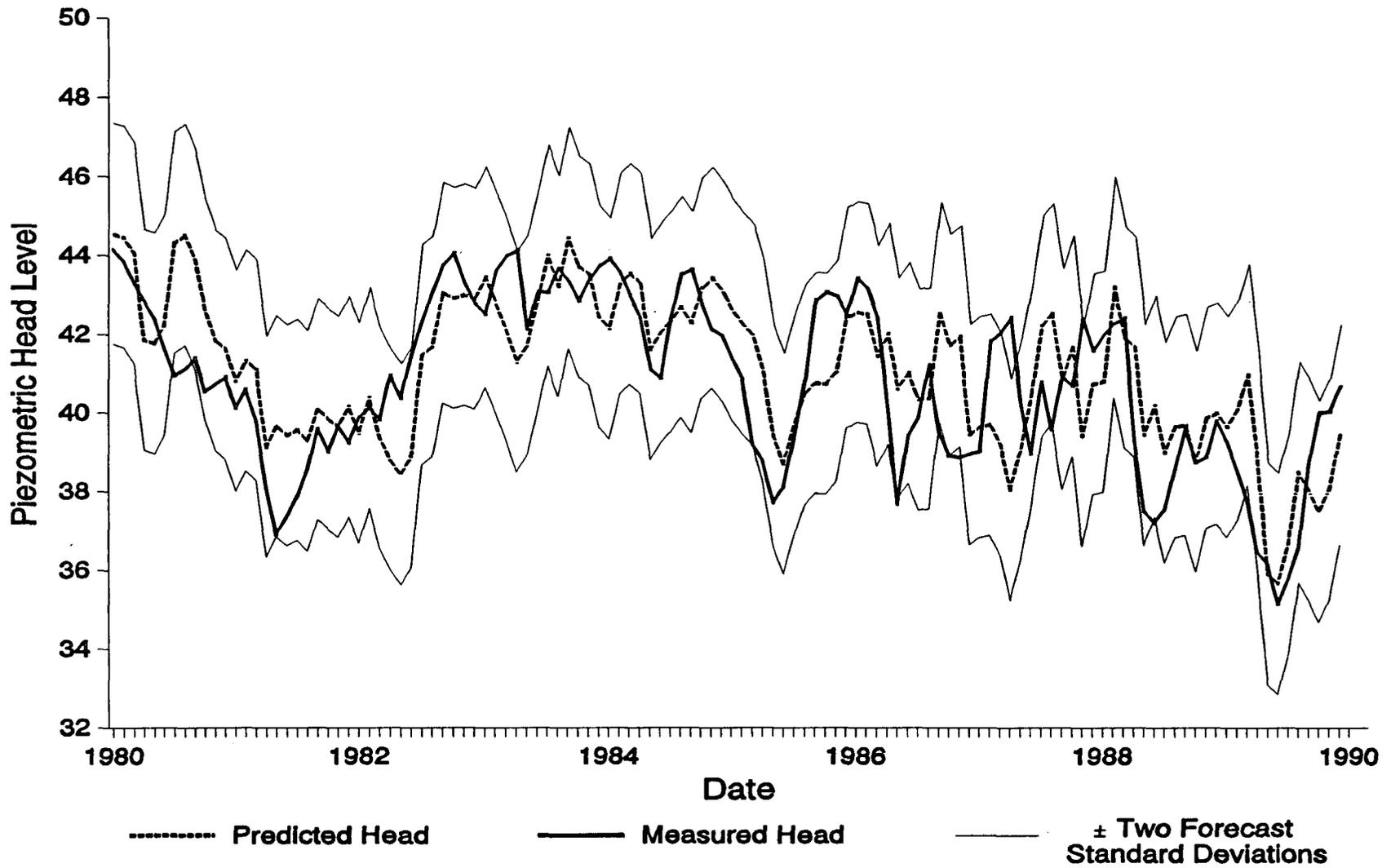


Figure 5.33 Three month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 13.

06-A

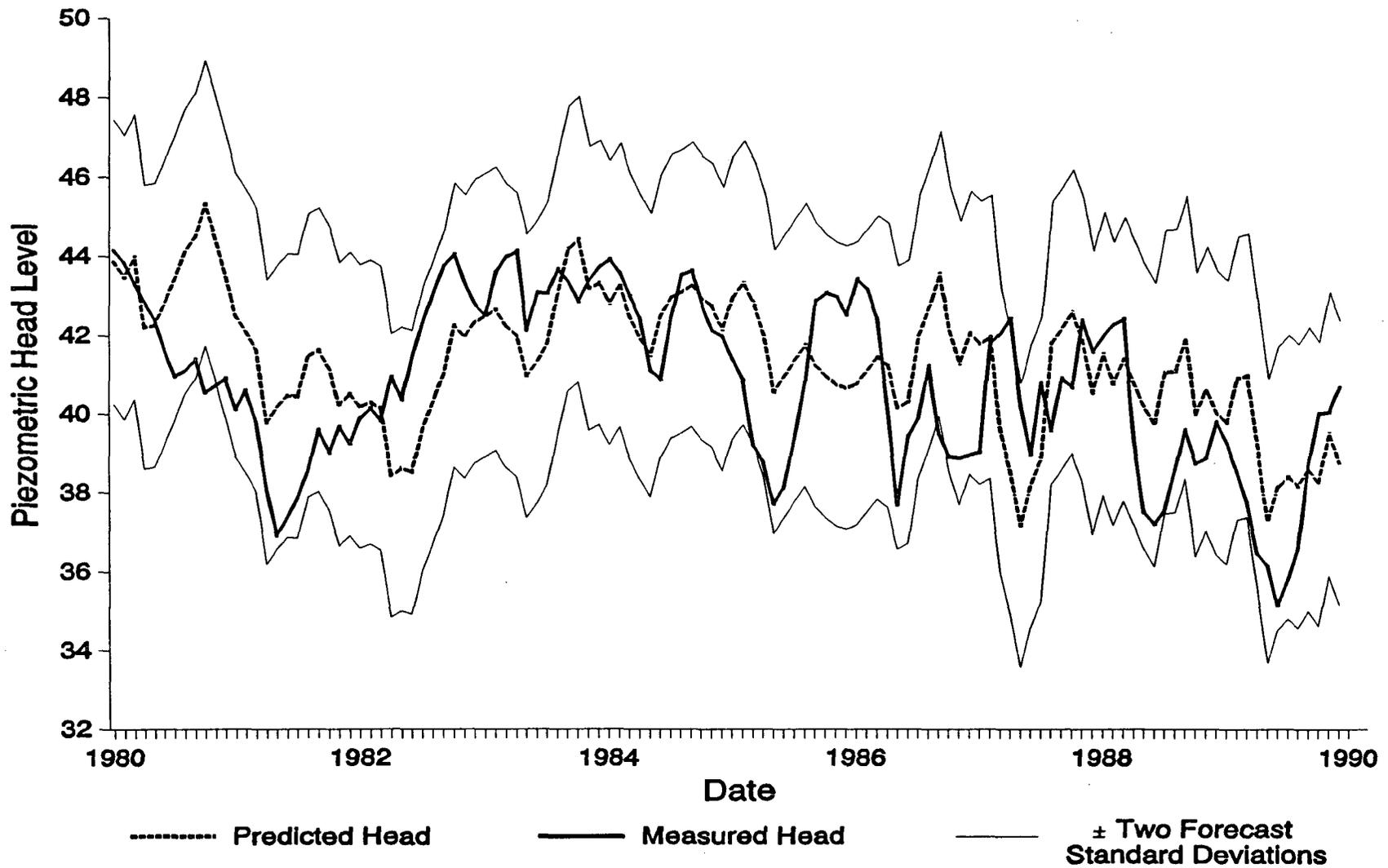


Figure 5.34 Six month lead forecasts of the  $ARIMA(2,0,0) \times (0,1,1)_{12}$  model for Well 13.

Table 5.5 ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model parameters fitted to 21 wells provided by the St. Johns River Water Management District.

Well No.	$\sigma_a^2$	$\theta_0$	$\phi_1$	$\phi_2$	$\theta_{12}$
1	17.361214	0.282726	0.57742	0.09093	0.63745
2	0.184538	0.024125	1.50572	-0.55845	0.65203
3	0.745759	-0.049966	1.29360	-0.40283	0.84598
4	0.356887	0.008398	1.33961	-0.34435	0.70951
5	0.154069	-0.005633	1.60244	-0.60507	0.78577
6	1.790165	-0.002409	0.61322	0.11001	0.79495
7	0.333215	-0.019422	1.12883	-0.24837	0.74941
8	0.094658	-0.010656	1.38132	-0.51990	0.78955
9	0.278584	0.002065	1.46743	-0.46798	0.76215
10	0.323197	-0.000000	1.47815	-0.47815	0.83049
11	0.894661	-0.003613	1.10231	-0.29921	0.71556
12	0.289917	-0.007582	0.97836	-0.20473	0.84134
13	0.678634	-0.002312	1.04987	-0.13643	0.68303
14	0.275027	-0.004345	0.98589	-0.15210	0.90336
15	1.164497	-0.009827	1.12982	-0.19465	0.84116
16	0.499200	-0.035206	0.99690	-0.18319	0.68089
17	1.007870	-0.013605	0.85623	-0.11260	0.92045
18	0.196385	-0.006658	1.04562	-0.11558	0.87371
19	0.397356	-0.020838	1.07507	-0.22719	0.76960
20	0.575690	0.000763	1.06876	-0.24323	0.63362
21	0.777525	-0.014430	0.71970	-0.12641	0.82432

$$(1 - \phi_1 B - \phi_2 B^2) \nabla_{12} Z_t = \theta_0 + (1 - \theta_{12} B^{12}) a_t \quad (5.98)$$

A listing of each well with its fitted model parameters is shown in Table 5.5.

Several of the wells could have been adequately modeled with a more parsimonious ARIMA(1,0,0)x(0,1,1)<sub>12</sub> model, such as Wells 1, 6, 13, 17, 18, and 21. All of these wells have small second-order autoregressive parameters of  $\hat{\phi}_2 = -0.13$  or less. The forecasting accuracy improvements afforded by the use of the ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model for these wells were minor. For Wells 2, 3, 5, 8, 9, and 10, however, the second-order autoregressive terms are significant at values of  $\hat{\phi}_2 \geq -0.40$ . Forecast improvements for these wells using the

ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model over the ARIMA(1,0,0)x(0,1,1)<sub>12</sub> model were not negligible. Thus it was decided to model all of the wells with an appropriate ARIMA(2,0,0)x(0,1,1)<sub>12</sub> model to allow for overall modeling consistency.

The actual versus predicted model error statistics for each of the 21 wells is shown in Table 5.6. This table indicates that, on the average, one month lead predictions can be made within 0.70 ft of the actual values, three month lead predictions can be made within 1.22 ft of the actual values, and six month lead predictions can be made within 1.53 ft of the actual values for these 21 wells.

The univariate models developed for these wells were based on the assumption that the present piezometric head level can be described by some weighted linear combination of previous piezometric head levels and an independent random shock, or residual,  $a_t$ . The influence of the residual on the process is described by its variance. Thus a process with a large residual variance is dominated by white noise input and is difficult to predict. The residual variance for Well 1, for instance, is very large at  $\hat{\sigma}_a^2 = 17.361214 \text{ ft}^2$ . A review of the forecast error standard deviations, Table 5.6, shows that the forecast errors for Well 1 have correspondingly large standard deviations. Thus, it can be concluded that the process which describes Well 1 is dominated by unpredictable random input. Consequently, model predictions for Well 1 may produce large forecast errors.

Table 5.6 shows that the forecast error increases as the lead time  $l$  increases. This results from the additive nature of the forecast errors for  $l > 1$ . For instance, the forecast  $\hat{Z}_t(6)$  for  $Z_{t+6}$  is based on past observations and forecasts. As borne out by equation (5.92), the

Table 5.6 Actual versus predicted model error  $a_t$  statistics.

Well Number	$\sigma_a$ (ft)	Mean Forecast Error (ft)	Forecast Error Standard Deviations					
			Lead-1		Lead-3		Lead-6	
			Actual (ft)	Estimated (ft)	Actual (ft)	Estimated (ft)	Actual (ft)	Estimated (ft)
1	4.16668	0.76843	3.97135	4.16668	4.77414	5.07300	4.99286	5.25706
2	0.42958	-0.08937	0.40598	0.42598	0.97000	0.98911	1.35992	1.37432
3	0.86357	0.86716	0.85716	0.86375	1.72528	1.72741	2.21015	2.21083
4	0.59740	-0.00431	0.57249	0.59740	1.22661	1.27739	1.85361	1.87060
5	0.39252	0.04294	0.37750	0.39252	0.96864	0.97120	1.43256	1.45310
6	1.33797	-0.05740	1.27603	1.33797	1.53947	1.67793	1.65906	1.77379
7	0.57725	0.04992	0.51975	0.57725	0.96096	1.04258	1.18435	1.30126
8	0.30767	0.04547	0.30323	0.30767	0.63124	0.63976	0.79630	0.79831
9	0.52781	-0.01643	0.52236	0.52781	1.21212	1.21561	1.79188	1.80871
10	0.56850	-0.01326	0.56739	0.56850	1.31121	1.31739	1.89167	1.96316
11	0.94587	0.04342	0.90377	0.94587	1.56982	1.62795	1.79948	1.90129
12	0.53844	-0.05088	0.52669	0.53844	0.82476	0.85645	0.89798	0.96334
13	0.82379	-0.06624	0.76161	0.82379	1.37682	1.43205	1.77596	1.83536
14	0.52443	-0.02796	0.51915	0.52443	0.82543	0.85334	0.96619	1.00481
15	1.07912	-0.02426	1.03877	1.07912	1.94309	1.98749	2.56210	2.63537
16	0.70654	0.05669	0.66938	0.70654	1.08701	1.15056	1.24939	1.33441
17	1.00393	0.06597	0.62917	1.00393	1.44026	1.46812	1.57302	1.60479
18	0.44315	-0.01975	0.44396	0.44315	0.77471	0.77254	0.98350	1.00828
19	0.63084	0.09112	0.61863	0.63084	1.06551	1.09021	1.24243	1.31149
20	0.75874	-0.06236	0.69491	0.75874	1.14703	1.29669	1.33215	1.52866
21	0.88177	-0.02621	0.86715	0.88177	1.12924	1.14979	1.16965	1.18138
Average*	0.69677	0.00729	0.65375	0.69677	1.18646	1.22718	1.48657	1.54316

\*Well 1 not included in average.

forecast error ( $\hat{Z}_t(6) - Z_{t+6}$ ) will then incorporate a weighted sum of previous errors.

Note that the actual model error standard deviations are smaller than the predicted model error standard deviations, thus confirming that the models are performing as expected. The forecast errors resemble a white noise process with, on the average, an approximately zero mean of 0.007 ft over all 21 wells. This indicates that the models have extracted all available information from the data series, and will produce optimal short-term site-specific predictions based on previous piezometric head level measurements.

Excursion and recursion analyses were performed on Wells 5, 6, 10, 11, 14, 17, 20, and 21. With the exception of Well 17 (see Case Study 1 above) the normality hypothesis of the original series for each well was not rejected using the Kolmogorov-Smirnoff test at a 0.05 level of significance. Table 5.7 lists the excursion analysis results. Table 5.8 lists the recursion analysis results. It should be recalled that for normally distributed data  $Z_t$  with mean  $\bar{Z}_t$  and standard deviation  $\sigma_t$ ,  $E\{I_0\}$  is the expected interval between mean crossings in months;  $E\{N\}$  is the average number of downcrossings of  $\bar{Z}_t - h\sigma_t$  ( $h = 1$  or  $2$ ) per month; conversely,  $1/E\{N\}$  is the number of months between downcrossings of  $\bar{Z}_t - h\sigma_t$ ;  $E\{I_h^- \}$  is the expected length of series excursion below  $\bar{Z}_t - h\sigma_t$ .

### Conclusions

In this Chapter, site-specific empirical stochastic models were developed for 21 wells within the District that are capable of producing accurate, optimal monthly forecasts of piezometric head levels for up to 6 month lead times. The developed models forecast future piezometric

Table 5.7 Excursion analysis results on the normalized series for the 8 nontrending wells.

Well Number	$\rho_1$	$\bar{Z}_t$ (ft)	$E\{T_0\}$ (months)	$h = 1$				$h = 2$			
				$\bar{Z}_t - \sigma_t$ (ft)	$E\{N\}$ (1/month)	$1/E\{N\}$ (months)	$E\{T_h\}$ (months)	$\bar{Z}_t - 2\sigma_t$ (ft)	$E\{N\}$ (1/month)	$1/E\{N\}$ (months)	$E\{T_h\}$ (months)
5	0.952	83.07	10.04	81.44	0.030	33.31	5.25	79.81	0.007	148.37	3.38
6	0.793	17.28	4.79	14.05	0.063	15.80	2.51	10.82	0.014	70.83	1.61
10	0.938	53.04	8.84	51.21	0.034	29.16	4.63	49.38	0.008	130.70	2.98
11	0.869	29.67	6.07	28.55	0.050	20.02	3.18	27.43	0.011	89.71	2.05
14	0.810	100.05	5.01	99.10	0.061	16.52	2.62	98.15	0.014	74.05	1.69
17	0.775	91.00	4.59	89.40	0.066	15.15	2.40	87.80	0.015	67.88	1.55
20	0.853	43.75	5.73	42.22	0.053	18.88	3.00	40.69	0.012	84.62	1.93
21	0.776	127.04	4.60	125.44	0.066	15.18	2.41	123.84	0.015	68.05	1.55

Table 5.8 Recursion analysis results for the 8 nontrending wells.

Well Number	Recurrence Piezometric Head Level (ft)			
	5 year	10 year	20 year	30 year
5	79.64	79.53	—*	—
6	3.01	2.84	—	—
10	49.20	49.12	—	—
11	38.66	38.55	—	—
14	97.64	97.31	97.12	97.03
17	86.44	86.16	85.69	85.58
20	40.17	38.83	—	—
21	123.02	122.38	121.84	121.59

\* Insufficient data to compute value.

head levels based on behavior of the groundwater flow system observed over the previous 12 to 14 months. Comparison of 1-month, 3-month, and 6-month model forecasts with observed piezometric head levels for all 21 wells showed that the forecast errors are unbiased (i.e., symmetrically distributed about a zero mean). An analysis of the model forecasts for 20 of the 21 wells showed that average standard deviations for the 1-month, 3-month, and 6-month forecasts are 0.70, 1.23, and 1.54 feet respectively. The model developed for Well 1 (SJRWMD Well N-0003) exhibits much larger forecast standard deviations, thus producing less accurate predictions. For all 21 models, actual model forecast errors fell consistently within the model's expected 95 percent confidence intervals. These time series models should provide short-term, temporal piezometric head forecasts for the particular wells modeled. To use these models, the District must continue to sample the 21 wells on a monthly basis. Since the model coefficients are empirical parameters

fit to historical data, estimates of these parameters should be updated on an annual or biannual basis.

For wells showing no long-term trends of decreasing or increasing water levels, recursion analysis produced expected low water level estimates at several of the 21 wells for several recurrence intervals. For those same wells, excursion analysis produced estimates of the recurrence intervals and durations of groundwater excursions below levels of one and two standard deviations below the mean water level.

## CHAPTER VI STOCHASTIC MODELING OF REGIONAL PIEZOMETRIC HEAD

### Introduction

As a part of their responsibility to evaluate and manage water resources, regional water management agencies must develop models capable of forecasting groundwater levels in space and time. As illustrated in Chapter IV, regional estimates of transmissivity, storativity and recharge (or leakance) are quite uncertain due to the sparseness, high spatial variability, and unknown reliability of available measurements of these parameters. As a result conventional deterministic numerical models of regional groundwater movement which require complete knowledge of these input parameters can produce highly uncertain results. Stochastic groundwater modeling techniques, however, provide a mechanism to account for both input parameter and model uncertainty in model predictions. The reliability of stochastic model predictions may be assessed directly by examining model prediction variances. Model predictions, together with model confidence intervals, can then be used to make more informed water resource management decisions.

The empirical time series models described in Chapter V provide short-term, temporal piezometric head forecasts for the particular wells being modeled. Univariate models were developed for each well based on the observed temporal correlation structure of  $Z(x_i, t)$  at each well

location  $x_i^*$ . Multivariate time series models, which incorporate temporal cross-correlations of piezometric head levels between wells, have the potential to improve piezometric head forecasts at wells that are sampled less frequently than others nearby. However, since each of the 21 wells modeled in this study is sampled at least monthly, site-specific multivariate time series models were not necessary.

To extrapolate the univariate site-specific model predictions over space, and thus forecast groundwater levels at unmonitored locations, the spatial correlation structure of  $Z(x,t)$  must be determined. This can be accomplished within the framework of geostatistical modeling techniques outlined in Chapter IV.

### Theory

Piezometric head varies over both time and space and therefore must, in general, be considered as a spatiotemporal random field  $Z(x,t)$ :

$$Z(x,t) = \bar{Z}(x,t) + \xi(x,t) , \quad (6.1)$$

where  $\bar{Z}(x,t)$  is the population mean, or expected value, of the random head field at location vector  $x$  and time  $t$ , and  $\xi(x,t)$  is the zero mean random fluctuation (residual) at location vector  $x$  and time  $t$ . In general, the covariance function describing the spatiotemporal correlation of the random head field will depend simultaneously on the two spatial locations  $x_1$  and  $x_2$  and the two times  $t_1$  and  $t_2$ , i.e.:

---

\* In Chapter IV, we spoke of the spatial random field  $Z_x$ , while in Chapter V we spoke of the temporal random process  $Z_t$ . In this chapter, to emphasize its spatial and temporal dependence, the spatiotemporal random field will be represented by  $Z(x,t)$ .

$$E\{[Z(x_1, t_1) - \bar{Z}(x_1, t_1)][Z(x_2, t_2) - \bar{Z}(x_2, t_2)]\} = C_2(x_1, x_2; t_1, t_2) . \quad (6.2)$$

In practice, it is often necessary to simplify this spatiotemporal covariance function by assuming the spatial and temporal correlations can be separated (Rodríguez-Iturbe and Mejia 1974; Egbert and Lettenmaier 1986; Rouhani and Wackeragel 1990) for example:

$$C_2(x_1, x_2; t_1, t_2) = C_2^t(t_1, t_2)C_2^x(x_1, x_2) , \quad (6.3)$$

where  $C_2^t(t_1, t_2)$  is the temporal correlation structure which is assumed to be independent of space and  $C_2^x(x_1, x_2)$  is the spatial correlation structure which is assumed to be independent of time.

We established in Chapter V that the piezometric head field  $Z(x, t)$  is non-stationary in time. Therefore a 12 month differencing procedure was proposed which can remove both seasonal and linear trends over time, i.e.:

$$\begin{aligned} \delta Z(x, t) &= Z(x, t) - Z(x, t-12) \\ &= (\bar{Z}(x, t) + \xi(x, t)) - (\bar{Z}(x, t-12) + \xi(x, t-12)) , \end{aligned} \quad (6.4a)$$

$$\delta Z(x, t) = \overline{\delta Z}(x, t) + \eta(x, t) . \quad (6.4b)$$

where  $t$  is measured in months and  $\delta Z(x, t)$  is the differenced random head at location vector  $x$  and time  $t$ , with mean  $\overline{\delta Z}(x, t) = \bar{Z}(x, t) - \bar{Z}(x, t-12)$  and zero mean residual  $\eta(x, t) = \xi(x, t) - \xi(x, t-12)$ . Examination of the piezometric time series for the 21 well locations,  $x_i$ , showed that the mean and correlation structure of the differenced head series was temporally stationary at each location i.e.:

$$E\{\delta Z(x_i, t)\} = \overline{\delta Z}(x_i, t) = \overline{\delta Z}(x_i) , \quad (6.5)$$

$$C_{\delta Z}(x_i, x_i; t_1, t_2) = C_{\delta Z}^t(t_1 - t_2) C_{\delta Z}^x(x_i, x_i) . \quad (6.6)$$

Evaluation of the temporal correlation structure required time series of at least 200 to 300 monthly head levels at a series of locations,  $x_i$ . Similarly, evaluation of the spatial correlation structure (ie the spatial mean and spatial covariance ) requires a spatial network of at least 200 to 300 head levels distributed throughout the region of interest at a series of times,  $t_n$ . In the St. Johns River Water Management District, piezometric head levels are measured over a dense network of 500 to 600 wells in May and September of every year. Thus the spatial correlation structure of piezometric head levels throughout the District can be evaluated using data from these sampling episodes and the geostatistical techniques summarized in Chapter IV.

As discussed in Chapter IV, inference of statistical spatial structure from one realization of a spatial random field requires that the random field exhibit either second-order stationarity or intrinsic (incremental) stationarity. However for the case of piezometric head, different spatial snapshots of the random field are available for May and September of every year. Evaluation of the temporal correlation structure of the differenced piezometric head series showed that temporal correlation is negligible after approximately 6 months in all of the 21 wells examined. Thus each May and September series of differenced piezometric head measurements taken over the District can be considered to be an approximately independent realization of the spatially random differenced head field. The existence of multiple

independent realizations of the random field provides a means to account for possible non-stationarities in the spatial correlation structure of the piezometric head field.

It can easily be shown that if  $Z(x,t)$  is a random field which is non-stationary in space and only seasonally non-stationary in time, then seasonally differencing two independent realizations of this random field removes both the unknown spatial mean and the unknown seasonal mean, leaving a stationary zero mean residual. For example if  $Z(x,t)$  and  $Z(x,t-12)$  are two independent spatial realizations of the random field taken during the same month of two successive years then :

$$\begin{aligned} Z(x,t) &= \bar{Z}(x,t) + \xi(x,t) , \\ Z(x,t-12) &= \bar{Z}(x,t-12) + \xi(x,t-12) . \end{aligned} \quad (6.7)$$

However since  $t$  and  $t-12$  represent the same month:

$$\bar{Z}(x,t) = \bar{Z}(x,t-12) = \bar{Z}(x) , \quad (6.8)$$

where  $\bar{Z}(x)$  is the spatially variable, but temporally constant mean for the month under consideration. Thus:

$$\begin{aligned} \delta Z(x,t) &= Z(x,t) - Z(x,t-12) \\ &= (\bar{Z}(x) + \xi(x,t)) - (\bar{Z}(x) + \xi(x,t-12)) , \end{aligned} \quad (6.9a)$$

$$\delta Z(x,t) = \xi(x,t) - \xi(x,t-12) . \quad (6.9b)$$

where  $\delta Z(x,t)$  is now a detrended random field with a stationary zero mean in both time and space for all months.

If there are  $n$  realizations of the original piezometric head field initially available for geostatistical analysis, the differencing procedure described above produces  $n-1$  realizations of the detrended field. The spatial structure of each of these  $n-1$  realizations can be analyzed using the methods outlined in Chapter IV. If the spatial structure of the differenced random field is truly stationary over time, the same general geostatistical structure should emerge from each realization. However, since they are inferred from a limited data set, the experimental variograms determined from each realization of the random field are only estimates of the underlying population variogram. Therefore the geostatistical parameters may not be identical from realization to realization. If the  $n-1$  variograms are not found to be significantly different (ie the spatial structure is in fact invariant over time but estimated parameters vary slightly between realizations) an improved estimate of the population variogram can be obtained by grouping the data from the  $n-1$  realizations. Then a single variogram, which is a more accurate estimate of the population variogram, can be determined and applied for each realization (Journel and Huijbregts, 1978).

Once the spatial correlation structure of piezometric head throughout the District has been determined, measurements taken at the 21 modeled wells can be extrapolated over space using the kriging technique. The optimal estimation equations for kriging piezometric head throughout the region, based on the variogram of the differenced head field and head observations at only the 21 modeled well locations, can be written:

$$\hat{Z}(x, t|t) = Z(x, t-12) + \sum_{j=1}^{21} \lambda_j(x) [Z(x_j, t) - Z(x_j, t-12)] , \quad (6.10)$$

where  $\hat{Z}(x, t|t)$  is the kriged estimate of piezometric head at an unmeasured location  $x$  and time  $t$  given observations of piezometric head at the 21 wells at time  $t$ ,  $Z(x, t-12)$  is the actual observed piezometric head that occurred at location  $x$  twelve months prior at time  $t-12$ ,  $Z(x_j, t)$  is the observed piezometric head at modeled well  $j$  at time  $t$ ,  $Z(x_j, t-12)$  is the observed piezometric head that occurred at well  $j$  twelve months prior at time  $t-12$ , and  $\lambda_j(x)$  is the kriging weight assigned to the observation at location  $j$  for location  $x$  determined from the kriging system of equations presented in Chapter IV. Equation (6.10) indicates that predictions of piezometric head at an unmeasured location  $x$ , for the current time  $t$ , may be estimated from observations made at the same location 12 months ago plus a weighted sum of the differences between heads observed at measured locations  $x_j$  this month and heads observed at locations  $x_j$  12 months ago.

Piezometric head forecasts for the 21 modeled wells can also be extrapolated, or kriged, over space in a similar manner. The optimal forecast equation for predicting piezometric head throughout the region, based on the variogram of the differenced head field and head forecasts at the 21 modeled well locations, can be written:

$$\hat{Z}(x, t+1|t) = Z(x, t+1-12) + \sum_{j=1}^{21} \lambda_j(x) [\hat{Z}(x_j, t+1|t) - Z(x_j, t+1-12)] , \quad (6.11)$$

where  $\hat{Z}(x, t+1|t)$  is the kriged forecast of piezometric head at location vector  $x$  and time  $t+1$  given forecasts of piezometric head at the modeled

21 wells for time  $t+1$ ,  $Z(x,t+1-12)$  is the actual observed piezometric head that occurred at location vector  $x$  twelve months prior to the forecast time,  $\hat{Z}(x_j,t+1|t)$  is the forecast piezometric head at well  $j$  for time  $t+1$  given observations at time  $t$ ,  $Z(x_j,t+1-12)$  is the observed piezometric head that occurred at well  $j$  twelve months prior to the forecast time, and  $\lambda_j(x)$  is the kriging weight assigned to the observation at location  $j$  for location  $x$ .

Equations 6.10 and 6.11 indicate that to kriging piezometric head over the district using this method requires that observations of head must be available at the location where a kriged estimate is desired at a time 12 months prior to when the kriged estimate is needed. This essentially restricts the use of the regional model to forecasting regional head levels for the months of May and September (when intensive piezometric head sampling is conducted). Thus the stochastic models developed in this study can be used to generate up to 6 month lead forecasts of piezometric head for any month of the year at the 21 wells that were modeled, or up to 6 month lead forecasts of piezometric head throughout the District for the months of May and September. Since May is typically the month that head levels reach their annual low and September is typically the month that head levels reach their annual high, regional forecasts for these months should provide sufficient information for water resource management purposes.

#### Hydrogeologic Data Description

The St. Johns River Water Management District provided ASCII data files of piezometric head levels for 636 wells within the Floridan aquifer for a series of May and September sampling episodes. Figure 6.1

shows the distribution of these wells throughout the District. Geostatistical analysis was conducted on 19 spatial data series beginning with the September 1981 sampling episode and ending with the September 1990 sampling episode. Multiple observations of head levels were given for some wells on various sampling dates. When multiple observations occurred for a single location on a particular date, the mean and the standard deviation of the observations were taken. If the standard deviation of the observations was less than one foot, the mean value for that location was used in the geostatistical analysis. If the standard deviation of the observations was greater than one foot, the data for that location was thrown out of the analysis for that sampling period. The data set was reduced in this manner to prevent the use of data that represented observations from different wells located at approximately the same latitude and longitude. It should be noted that not all 636 wells were sampled on each date, thus the number of observations used in the geostatistical analysis ranged from a low of 161 for the September 1986 sampling episode to a high of 537 for the September 1988 sampling episode.

### Results and Discussion

Plots of the spatial configuration of actual piezometric head levels for the entire series of May and September sampling episodes revealed an obvious spatial trend in the original data which is consistent over time. Figures 6.2 and 6.3 illustrate this trend for September 1988 and May 1989 respectively. The trend is characterized by generally higher piezometric heads near the western boundary of the District and generally lower piezometric heads near the Atlantic coast,

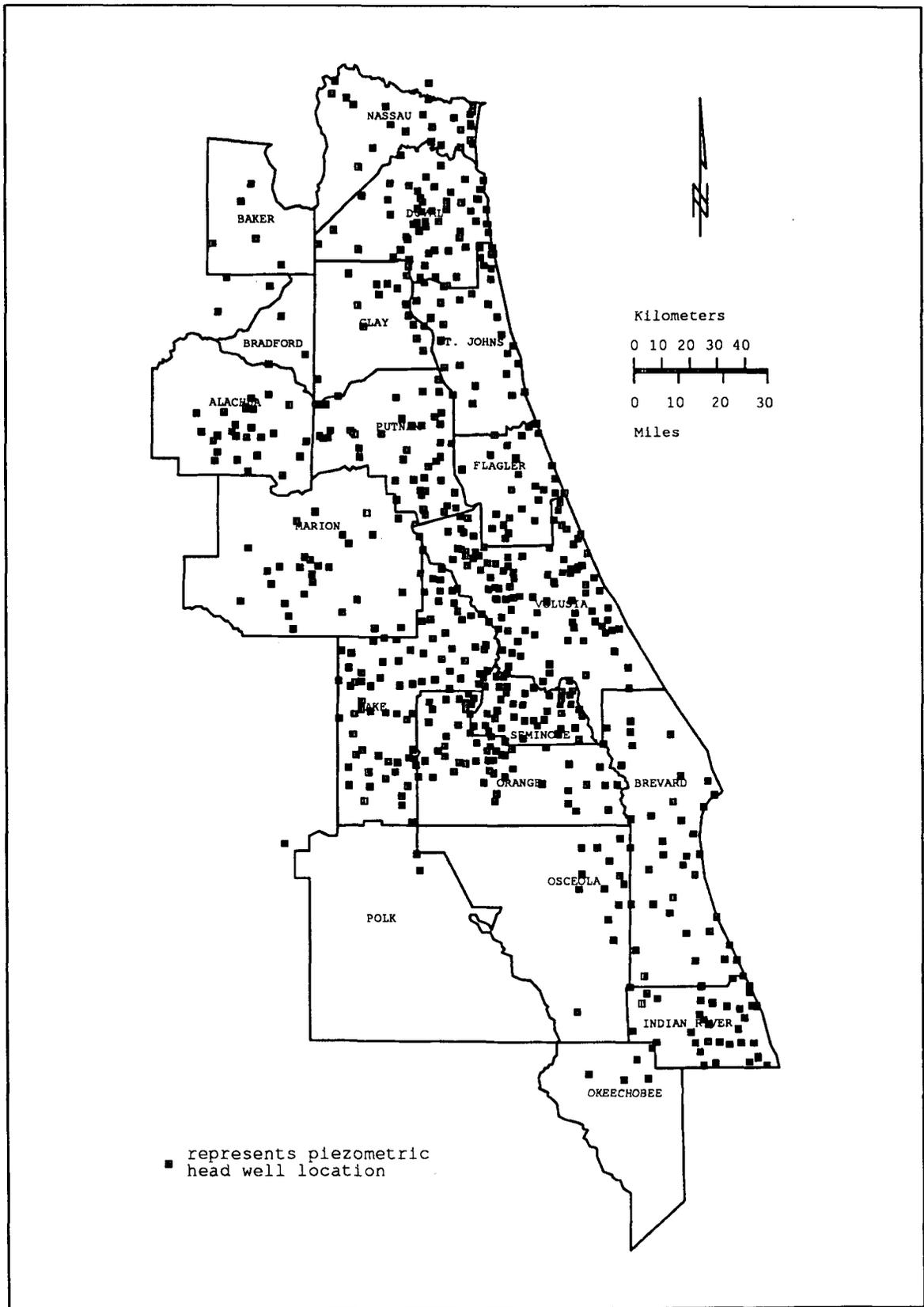


Figure 6.1 Spatial distribution of 636 wells.

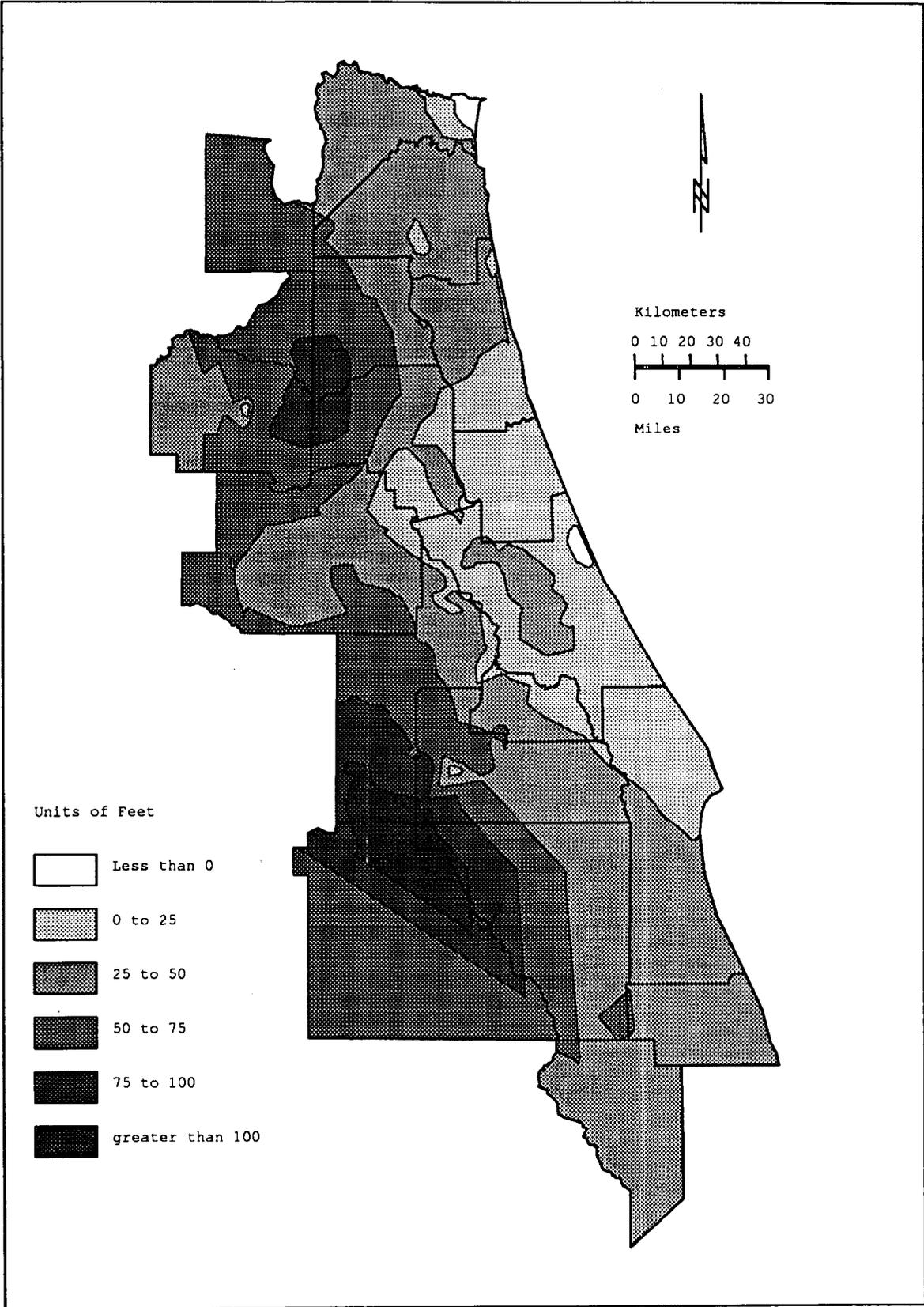


Figure 6.2 Observed piezometric head map for September 1988.

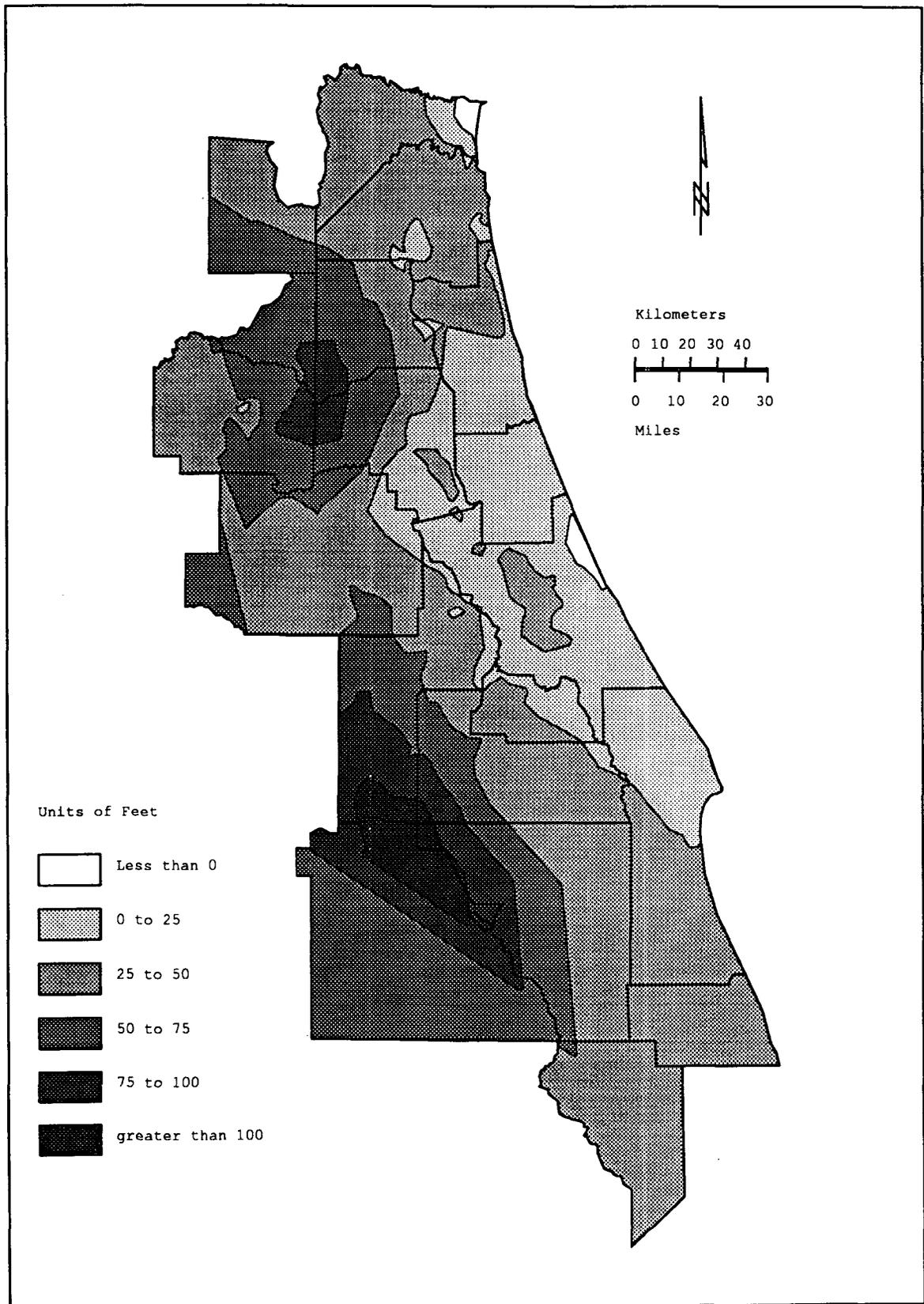


Figure 6.3 Observed piezometric head map for May 1989.

as well as a local piezometric high in central Volusia county. Random deviations around the approximately temporally constant spatial trend will occur from month to month due to random variations in the spatial rainfall, recharge and pumping patterns.

The non-stationary nature of the original (undifferenced) piezometric head field in space is also evident in the experimental variograms for the September 1988 and May 1989 samplings shown in Figures 6.4 and 6.5. These figures illustrate the distinct undulating non-stationary variogram shape that was found for each May and September sampling episode. This shape can likely be attributed to the fact that two wells separated by a given distance in the east-west direction would be expected to have significantly different measured head values due to the deterministic drop of piezometric head from the western boundary of the District toward the coast. However, two wells separated by a

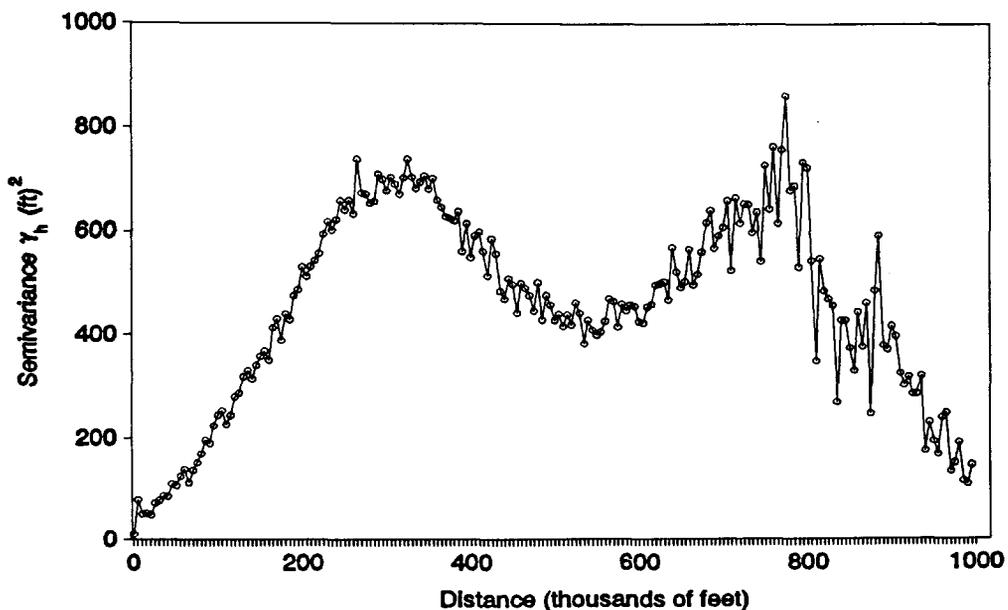


Figure 6.4 Semivariogram of observed piezometric head for September 1988.

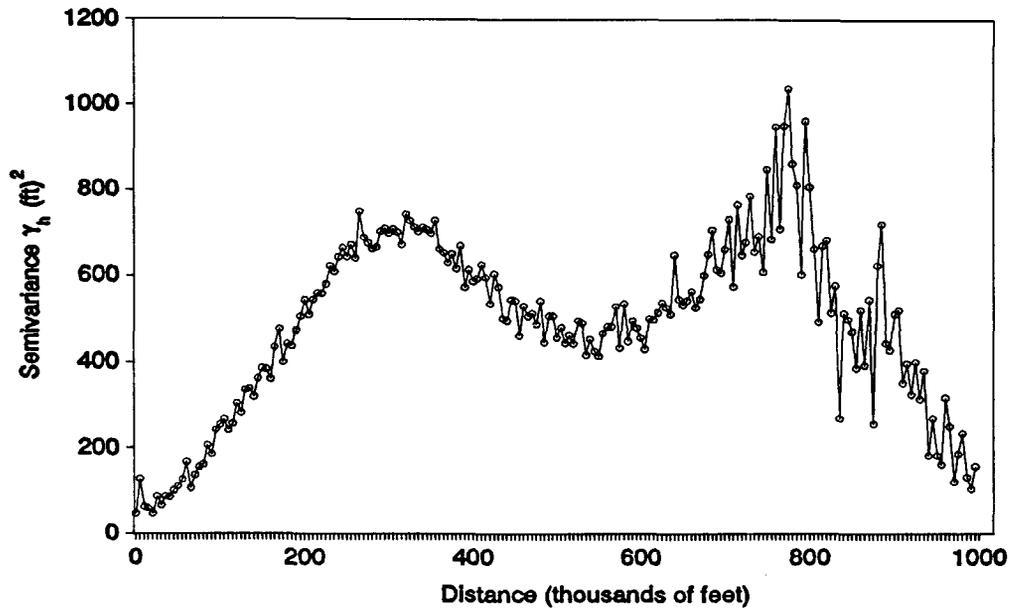


Figure 6.5 Semivariogram of observed piezometric head for May 1989.

greater distance in the north-south direction would be expected to have more similar piezometric head values, since they would both be located on the central ridge of the piezometric head field. Thus the variogram tends to dip up and down with increasing direction and does not reach the level plateau indicative of a second-order stationary random field.

Fortunately, as discussed in the theory section above, the seasonal differencing procedure used to detrend the random field over time also has the potential to detrend the random field over space. Figures 6.6 and 6.7 show the spatial distribution of the differenced head fields for September 1988 and May 1989 respectively. Recall that:

$$\delta Z(x, \text{September } 1988) = Z(x, \text{September } 1988) - Z(x, \text{September } 1987) \quad (6.12)$$

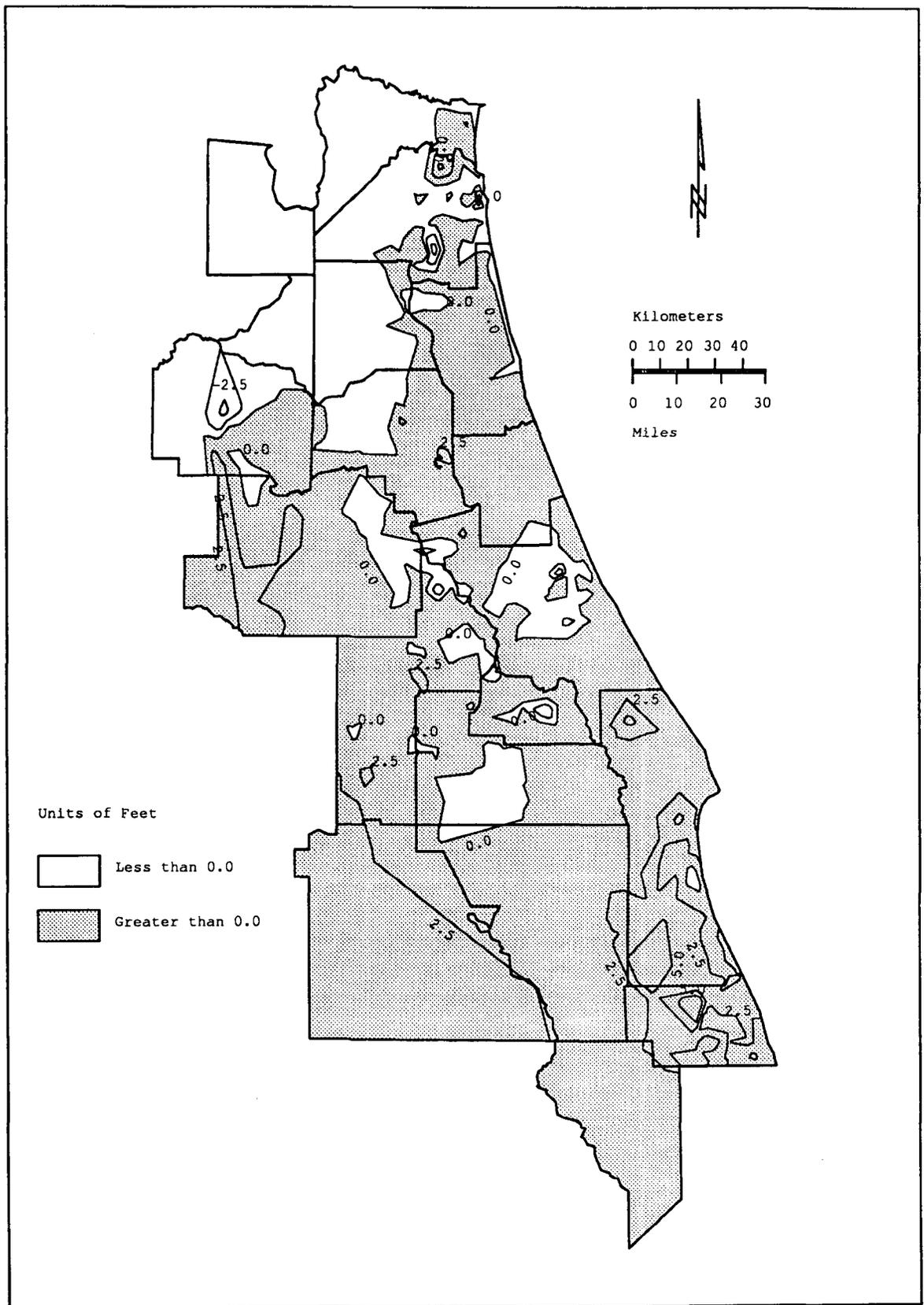


Figure 6.6 Differenced observed piezometric head map for September 1988.

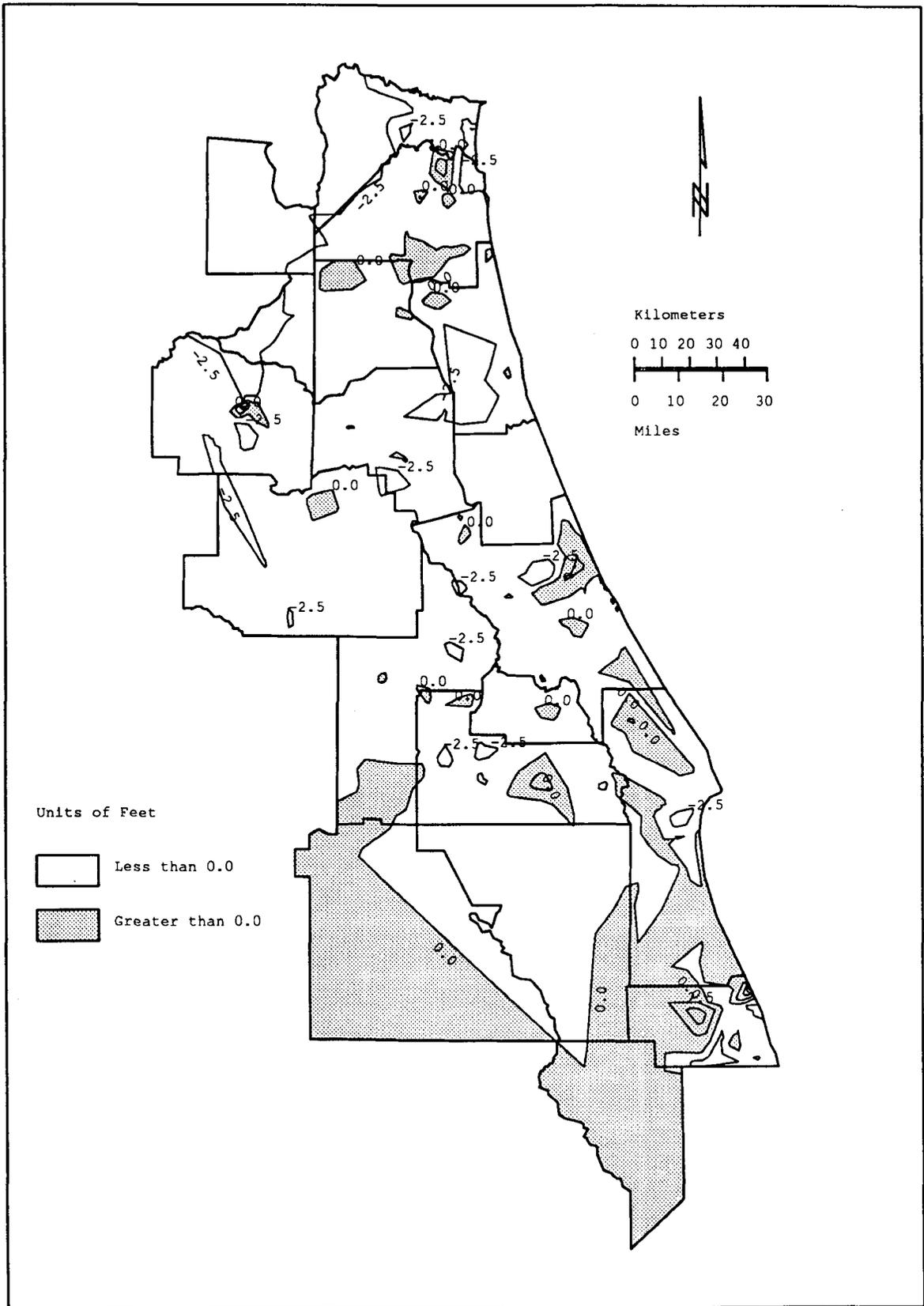


Figure 6.7 Differenced observed piezometric head map for May 1989.

and

$$\delta Z(x, \text{May } 1989) = Z(x, \text{May } 1989) - Z(x, \text{May } 1988) \quad (6.13)$$

A comparison of Figures 6.2 and 6.3 to Figures 6.6 and 6.7 shows that while the original random field varied from less than zero to greater than one hundred feet, the differenced head field varies from -10 to 10 feet. Thus the differencing procedure produces a random field that has a lower range of variability and is distributed around an approximately zero mean. The experimental cumulative distribution function (cdf) for the differenced head field can be calculated for each sampling episode according to the following equation:

$$\hat{F}(\delta Z) = \frac{\text{Rank}(\delta Z)}{n+1} \quad (6.14)$$

where  $\hat{F}(\delta Z)$  is the experimental cumulative distribution function value for the differenced head value  $\delta Z$ ,  $\text{Rank}(\delta Z)$  is the numerical rank of the  $\delta Z$  value assigned after the data set has been sorted into increasing order, and  $n$  is the total number of  $\delta Z$  observations used to estimate the experimental cdf. Figures 6.8 and 6.9 show the experimental cdf calculated for the differenced head fields for September 1988 and May 1989 respectively. These figures indicate that the differenced head field is approximately symmetrically distributed around an approximately zero mean. Seventeen differenced head fields, corresponding to sampling episodes September 1982 through September 1990, were calculated from the 19 original head fields that were initially available. Table 6.1 summarizes the sample mean and sample standard deviations calculated from each of these differenced head fields. The sample statistics

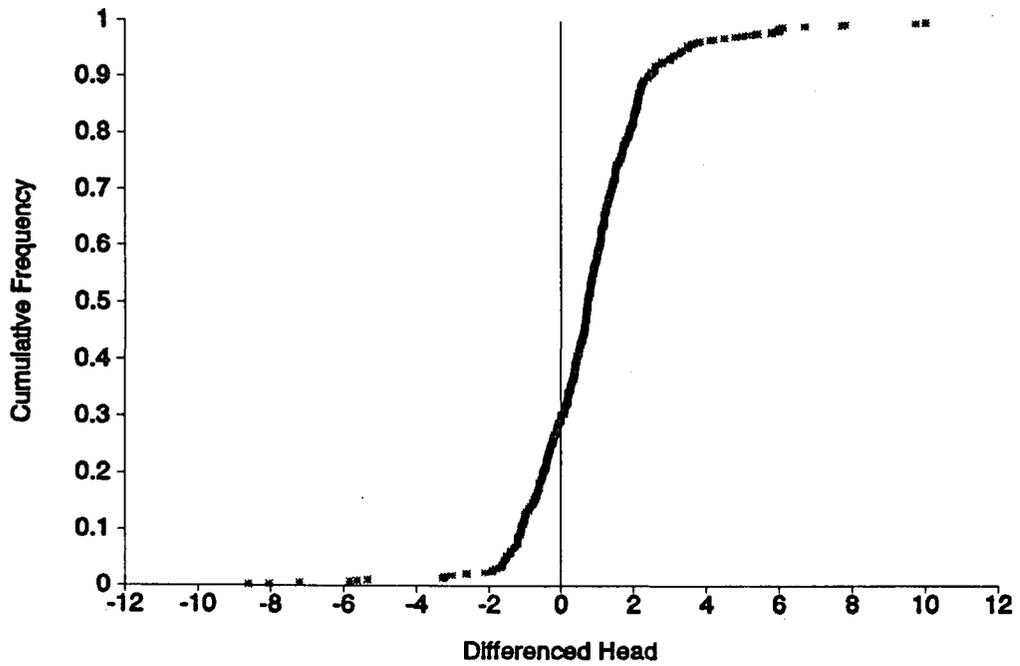


Figure 6.8 Experimental cdf for differenced observed piezometric head for September 1988.

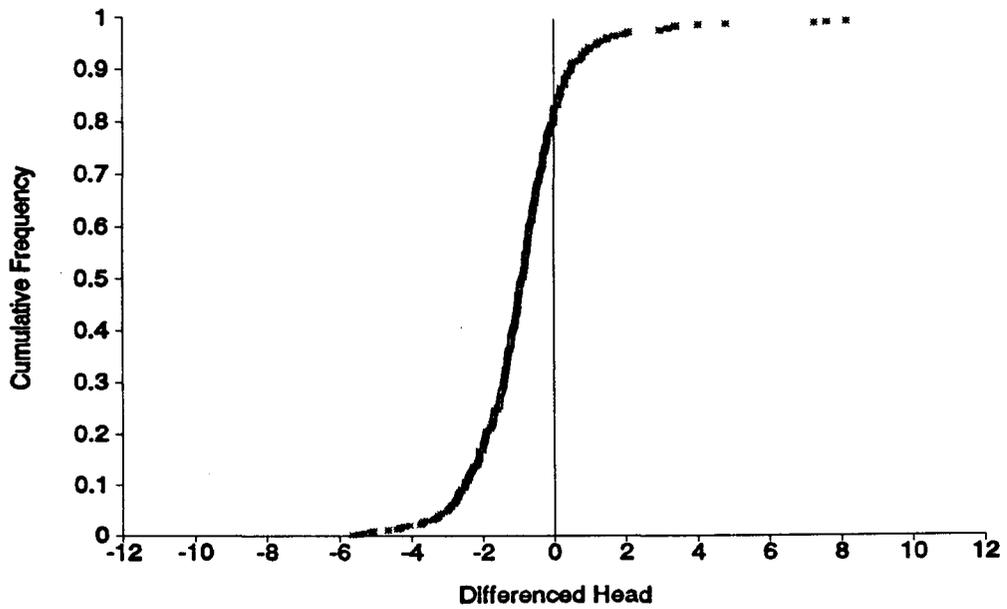


Figure 6.9 Experimental cdf for differenced observed piezometric head for May 1989.

Table 6.1 Differenced head field sample statistics.

Realization	Number of Observations	Mean	Standard Deviation
Sept 1982	517	3.06	2.03
May 1983	500	1.97	2.30
Sept 1983	551	-2.36	1.97
May 1984	543	-0.89	1.85
Sept 1984	537	0.22	1.39
May 1985	550	-2.53	2.17
Sept 1985	549	-1.33	1.56
May 1986	567	0.68	1.79
Sept 1986	186	-0.22	1.53
May 1987	552	1.59	1.76
Sept 1987	561	-0.57	1.90
May 1988	555	-1.22	1.97
Sept 1988	564	0.74	1.80
May 1989	555	-0.90	1.46
Sept 1989	538	-1.50	1.79
May 1990	543	-0.96	1.59
Sept 1990	556	-1.20	1.59
Average		-0.19	1.79

fluctuate around an average mean of -0.19 ft, and have an average standard deviation of 1.79 ft. Thus although  $Z(x,t)$  is a non-stationary random field in both space and time, the seasonally differenced field  $\delta Z(x,t)$  has a stationary approximately zero mean.

Experimental variograms were calculated for each of the 17 differenced fields to determine whether the geostatistical spatial structure was consistent (i.e. stationary) over time. Each of the 17 variograms was found to be well fit by a simple linear variogram with a nugget effect. Table 6.2 summarizes the linear variogram model parameters fit to each sampling episode using GEOPACK. Figures 6.10 and 6.11 show plots of the experimental variograms for September 1989 and May 1989 respectively.

Table 6.2 Experimental variogram model parameters.

Realization	Number of Observations	Model	Nugget (ft <sup>2</sup> )	Slope (ft <sup>2</sup> /ft)	Range of Validity (ft)
Sept 1982	517	linear	1.53	0.0016	200,000
May 1983	500	linear	2.40	0.0086	500,000
Sept 1983	551	linear	1.81	0.0053	500,000
May 1984	543	linear	1.96	0.0045	500,000
Sept 1984	537	linear	1.23	0.0018	500,000
May 1985	550	linear	2.11	0.0063	500,000
Sept 1985	549	linear	1.73	0.0019	500,000
May 1986	567	linear	1.64	0.0044	500,000
Sept 1986	186	linear	2.15	0.0003	500,000
May 1987	552	linear	1.88	0.0026	500,000
Sept 1987	561	linear	1.91	0.0087	300,000
May 1988	555	linear	1.98	0.0049	500,000
Sept 1988	564	linear	1.78	0.0031	500,000
May 1989	555	linear	1.33	0.0018	500,000
Sept 1989	538	linear	2.16	0.0032	500,000
May 1990	543	linear	1.11	0.0035	500,000
Sept 1990	556	linear	1.47	0.0028	500,000
Combined		linear	1.90	0.0036	500,000

Since the 17 experimental variograms were not found to be significantly different (i.e. the spatial structure is apparently invariant over time but estimated parameters vary slightly between realizations) an improved estimate of the population variogram was obtained by grouping the data from the 17 realizations. Figure 6.12 shows a plot of the combined variogram, which is a more accurate estimate of the population variogram, and which can be used for the geostatistical analysis of the differenced head field for any May or September. The variogram model parameters fit by GEOPACK to this combined experimental variogram are included in Table 6.2.

Using the spatial correlation structure of the differenced piezometric head described by the combined variogram, measurements or

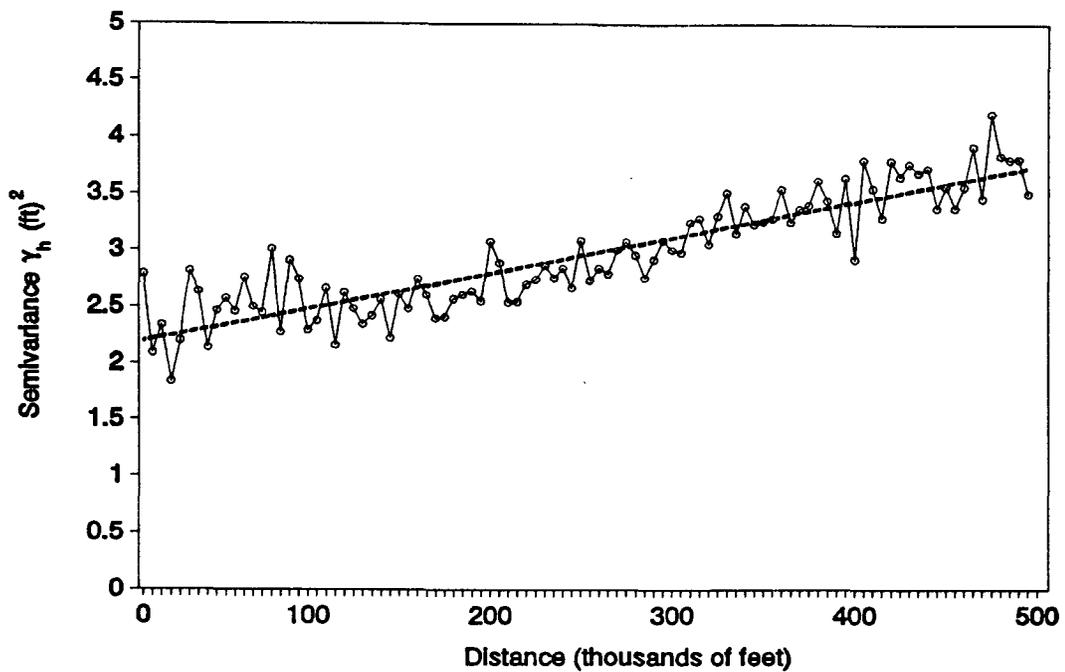


Figure 6.10 Linear model fitted to a semivariogram of differenced observed piezometric head levels for September 1989.

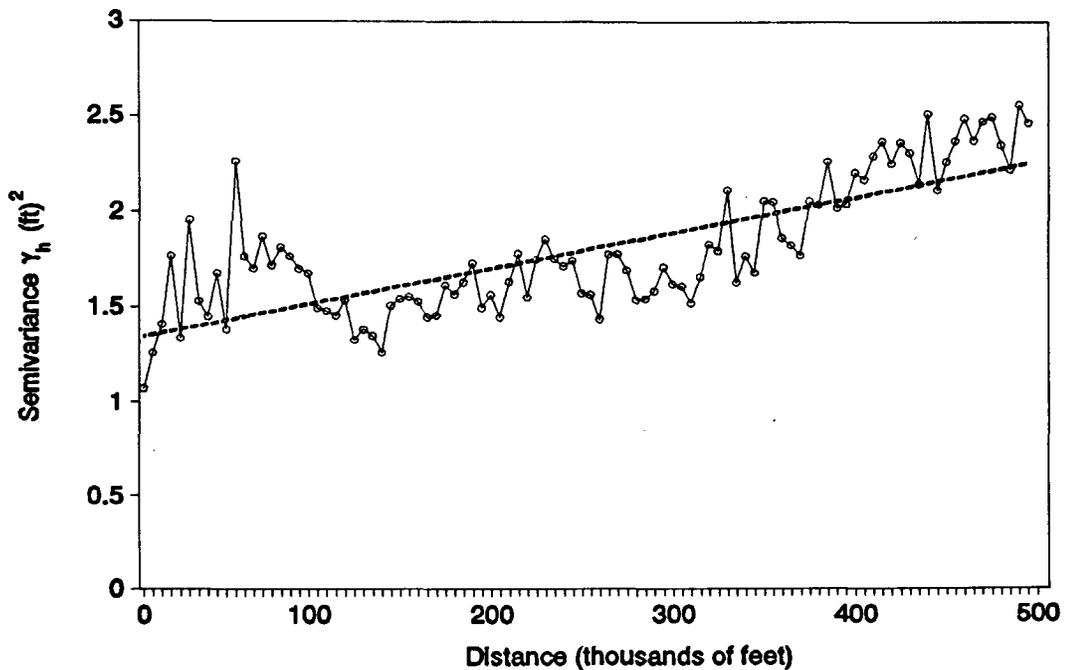


Figure 6.11 Linear model fitted to a semivariogram of differenced observed piezometric head levels for May 1989.

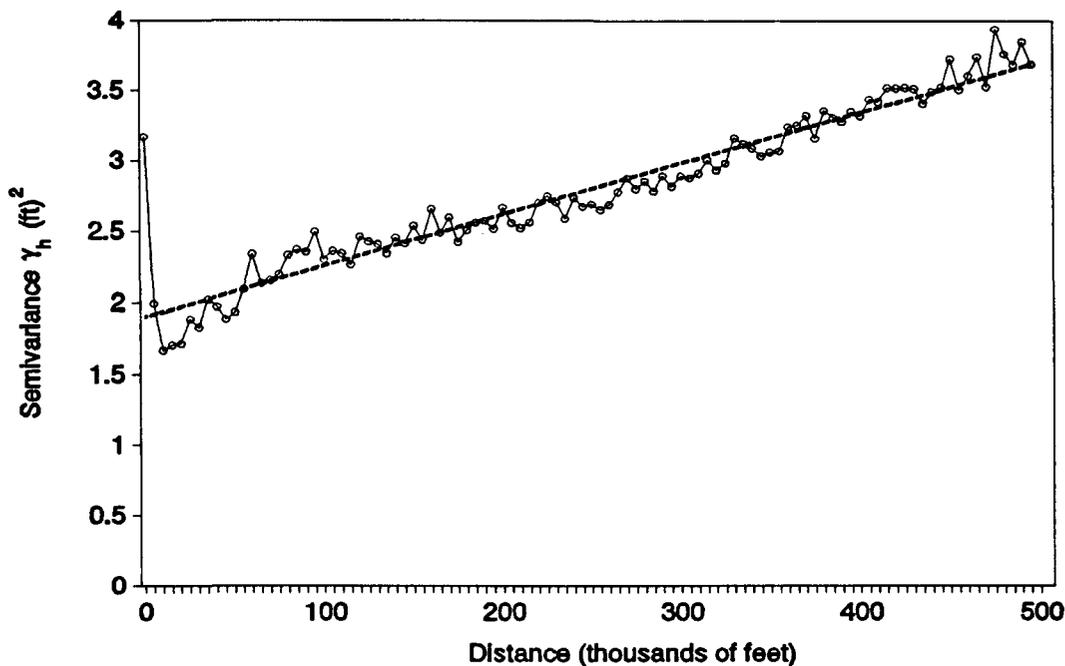


Figure 6.12 Linear model fitted to a combined semivariogram of differenced observed piezometric head levels.

forecasts at the 21 modeled wells can be kriged over the District to produce a regional map of estimated piezometric head. In addition the kriging technique produces a regional map of the expected error (i.e. the standard deviation of the predicted error) of the piezometric head estimate. It should be recalled that the differencing technique required to detrend the piezometric head field requires that observed head levels at time  $t-12$  be available for any point in the District where a kriged estimate is desired at time  $t$ . This essentially restricts the model to kriging May and September measurements or forecasts from the 21 modeled wells over the grid defined by the 636 well network.

To verify the performance of the geostatistical model, a regional map for each May and September sampling episode was created by kriging

an estimate of piezometric head level over the 636 well network from actual observations over only the 21 well network. Figures 6.13 and 6.14 show two examples of these kriged head maps for September 1988 and May 1989 respectively, while Figures 6.15 and 6.16 show the kriging standard deviations for these estimates. It should be noted that the kriging standard deviation maps are identical for all sampling episodes. This is due to the fact that the kriging standard deviation depends only on the spatial configuration of the 21 well measurement well network, not on the actual values measured.

Figures 6.15 and 6.16 indicate that, except around clusters of wells very near the 21 measured wells, the predicted kriging error standard deviation over most of the District is between 1.5 and 2.0 feet. This value is just slightly larger than the square root of the nugget of the combined variogram. As discussed in Chapter IV, the nugget effect represents unresolved variability occurring over distances smaller than the separation of the closest measurements used to create the variogram. Thus geostatistical models which contain a nugget factor predict kriging estimation errors which increase from zero at the measurement point to the square root of the nugget effect over very small distances.

Figures 6.17 and 6.18 show maps of the actual prediction errors calculated by subtracting the kriged head estimate from the observed head measured at each of the 636 wells. These plots, which are typical of those observed for each of the 17 sampling episodes, indicate that the actual prediction errors oscillate up and down around zero in a random fashion. Recall from the theory discussion in Chapter IV that the kriging estimator is designed to produce unbiased estimates whose

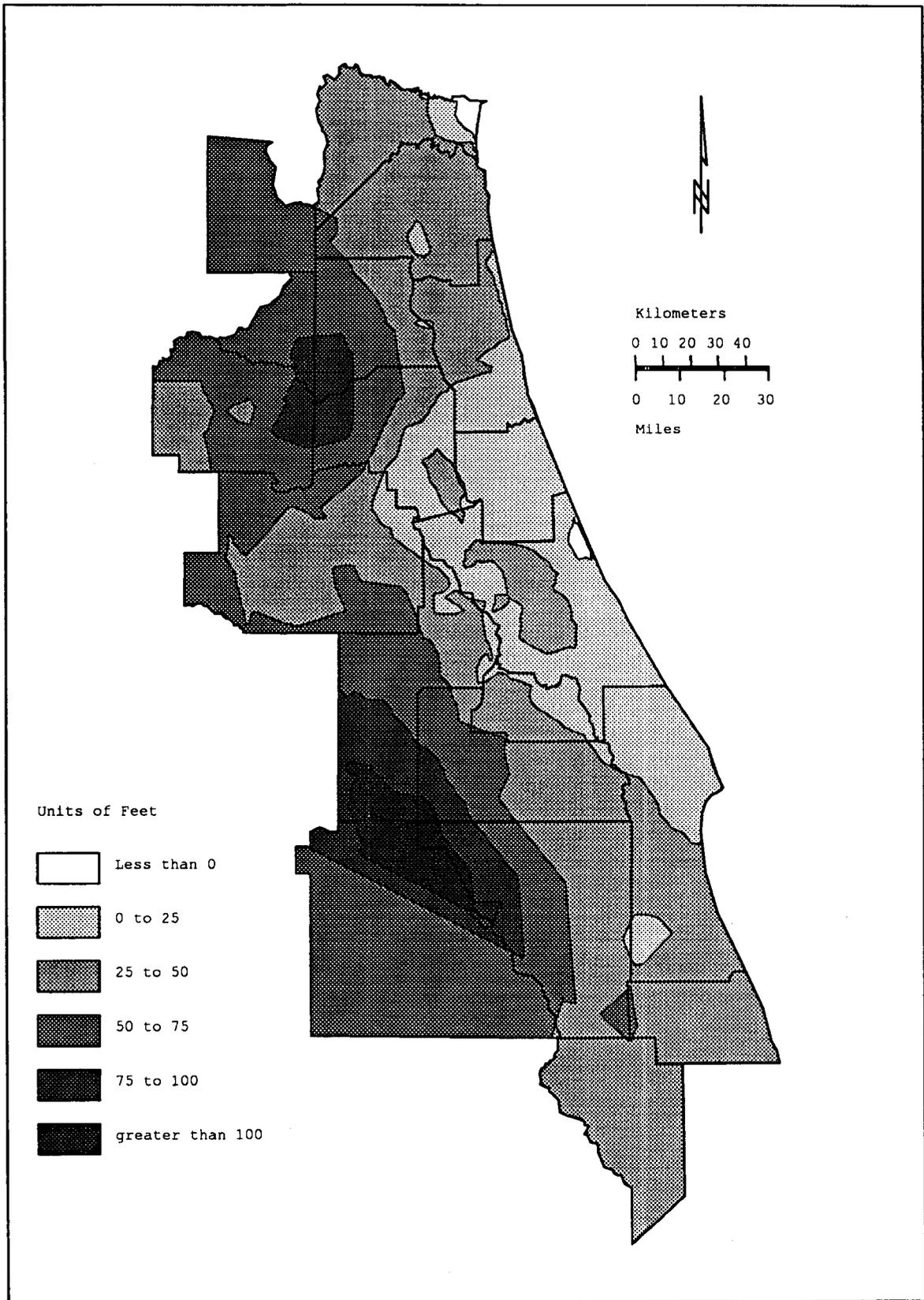


Figure 6.13 Piezometric head estimates for September 1988 created by kriging over 636 well network from actual observations over 21 well network.

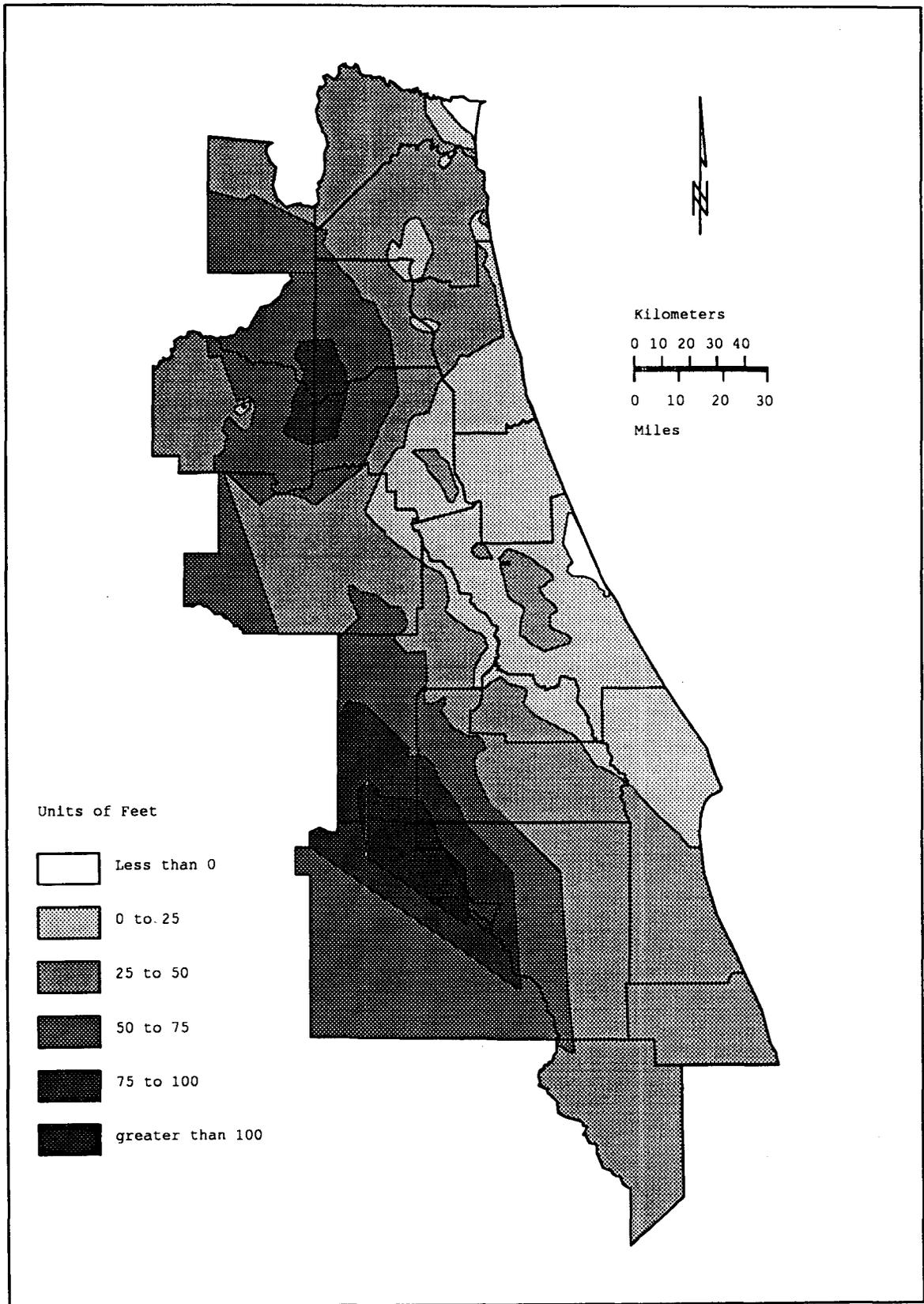


Figure 6.14 Piezometric head estimates for May 1989 created by kriging over 636 well network from actual observations over 21 well network.

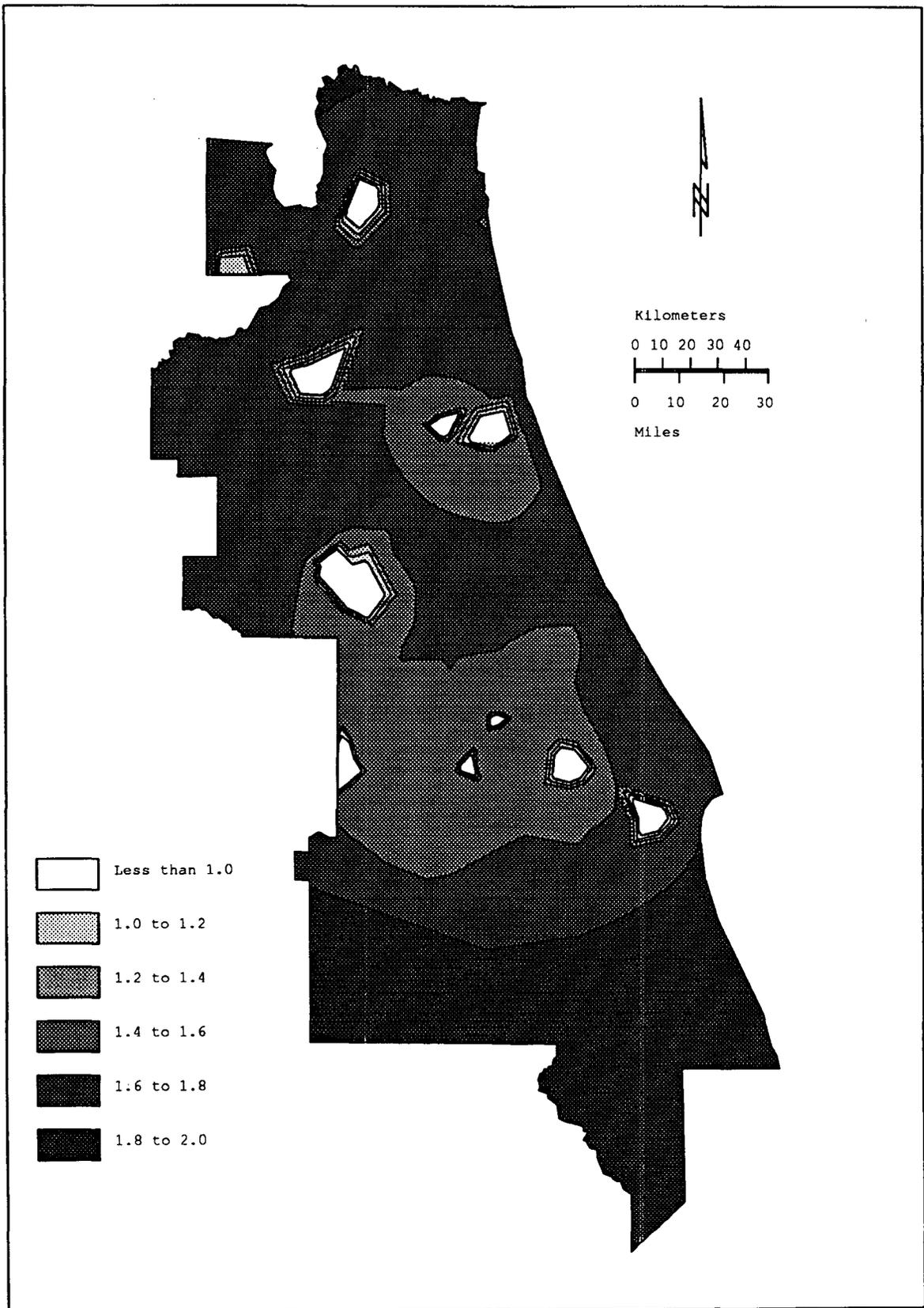


Figure 6.15 Standard deviations of piezometric head estimates for September 1988 created by kriging over 636 well network from actual observations over 21 well network.

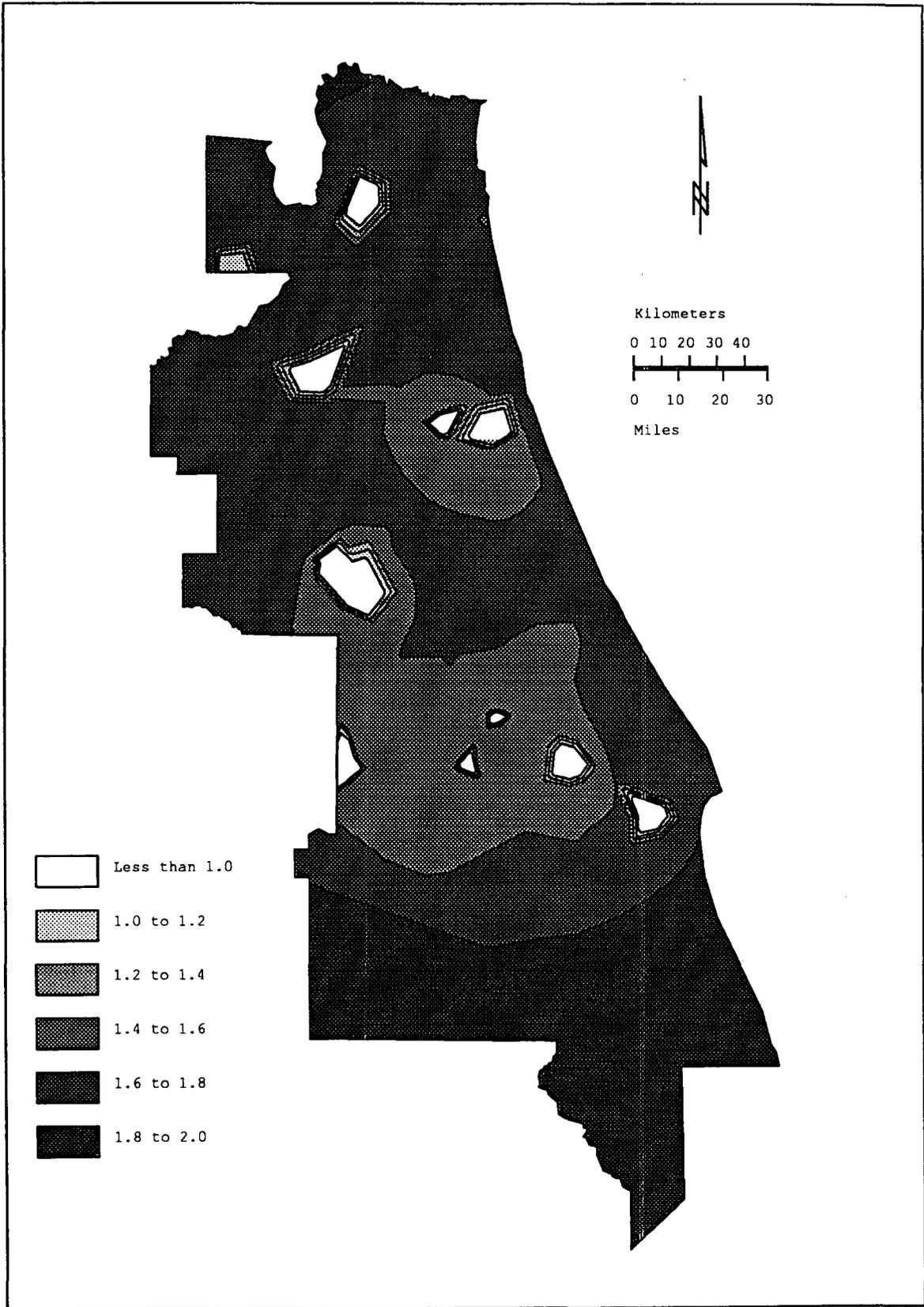


Figure 6.16 Standard deviations of piezometric head estimates for May 1989 created by kriging over 636 well network from actual observations over 21 well network.

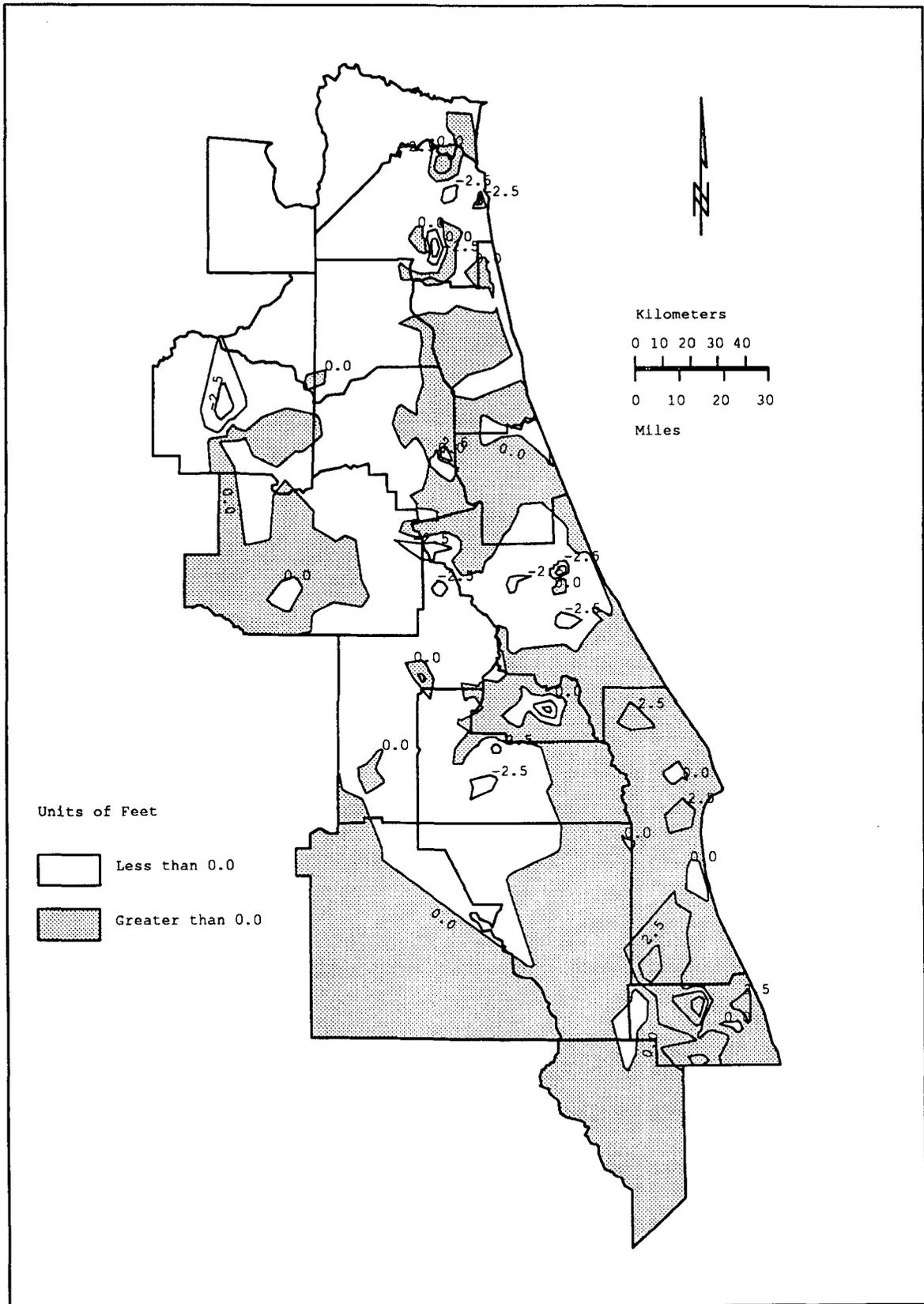


Figure 6.17 Actual prediction errors for September 1988 calculated by subtracting the kriged head estimate from the observed head measured at each of the 636 wells.

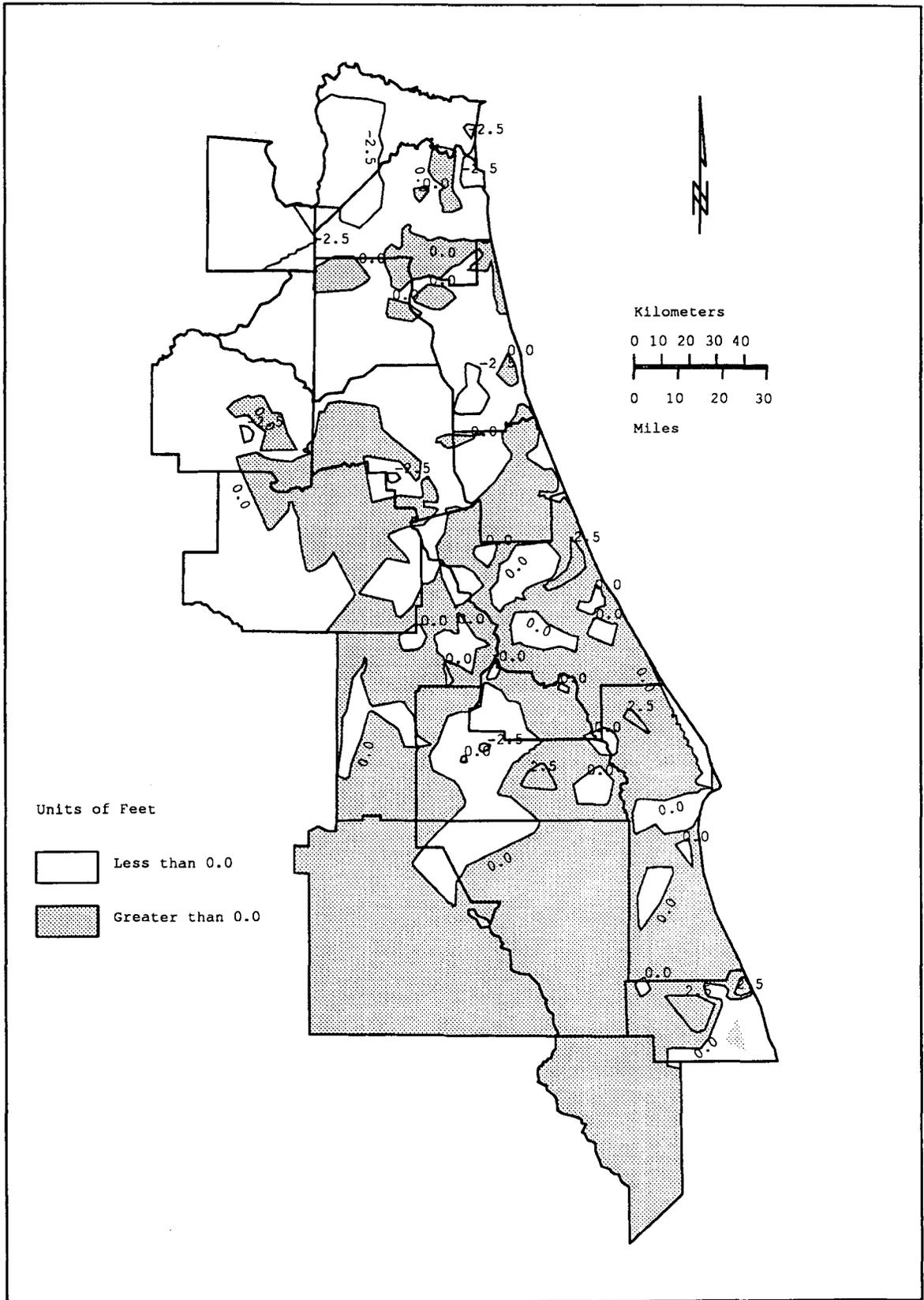


Figure 6.18 Actual prediction errors for May 1989 calculated by subtracting the kriged head estimate from the observed head measured at each of the 636 wells.

actual errors have a mean of zero and a standard deviation equal to the kriging standard deviation. Thus the accuracy of the kriging estimator can be evaluated by examining the statistics of its actual prediction errors.

Figures 6.19 and 6.20 show the experimental cdf's of the actual kriging errors which have been normalized by subtracting the kriging estimate from the observed head and dividing by the kriging standard deviation at each of the 636 well locations,  $x_i$ , i.e.:

$$\varepsilon(x_i, t) = \frac{Z(x_i, t) - \hat{Z}(x_i, t|t)}{\sigma_k(x_i, t)}, \quad (6.15)$$

where  $\varepsilon(x_i, t)$  is the normalized kriging error,  $Z(x_i, t)$  is the actual observed head,  $\hat{Z}(x_i, t|t)$  is the kriged estimate of head, and  $\sigma_k(x_i, t)$  is the predicted kriging standard deviation, all at locations  $x_i$  and times  $t$ . If the kriging estimator is performing as designed these normalized errors should be symmetrically distributed around a mean of zero and have a standard deviation of 1.0. Figures 6.19 and 6.20 indicate that the actual kriging errors are indeed symmetrically distributed with an approximately zero mean and unit standard deviation. Table 6.3 summarizes the statistics of the normalized kriging errors for each of the 17 realizations.

The above analysis indicates that the combined variogram provides an accurate description of the spatial structure of the differenced random head field, and the geostatistical model developed using this variogram provides accurate estimates of the regional piezometric head given only 21 head measurements distributed throughout the District.

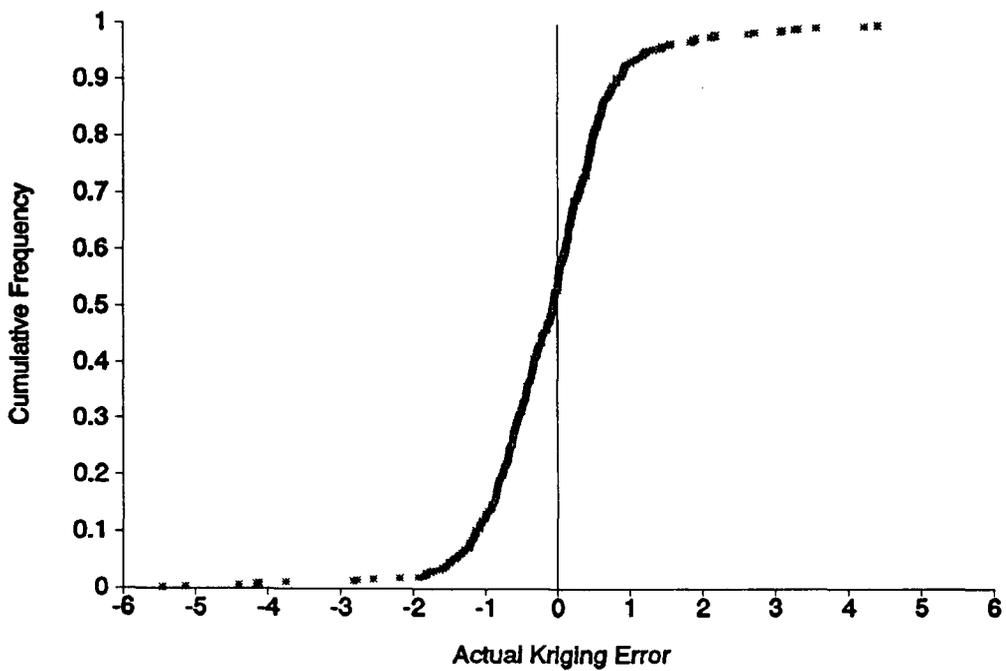


Figure 6.19 Experimental cdf for normalized actual piezometric head kriging errors for September 1988.

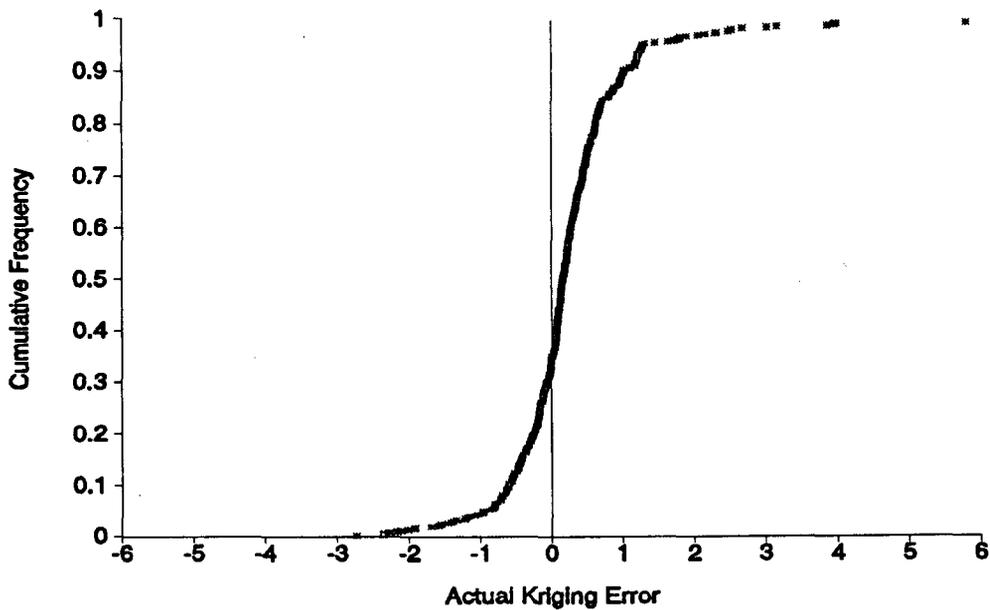


Figure 6.20 Experimental cdf for normalized actual piezometric head kriging errors for May 1989.

Table 6.3 Normalized kriging error statistics.

Realization	Number of Observations	Mean (ft)	Standard Deviation (ft)
Sept 1982	456	-0.18	1.10
May 1983	442	0.47	1.41
Sept 1983	485	0.32	1.23
May 1984	457	0.18	1.06
Sept 1984	493	0.31	0.77
May 1985	490	-0.01	1.09
Sept 1985	489	-0.37	0.93
May 1986	505	-0.01	0.93
Sept 1986	161	0.22	0.93
May 1987	517	-0.03	1.07
Sept 1987	170	0.02	1.05
May 1988	497	0.10	1.01
Sept 1988	526	-0.11	0.97
May 1989	514	0.20	0.80
Sept 1989	506	0.04	0.96
May 1990	514	0.27	0.73
Sept 1990	500	-0.08	0.83
Average	454	0.08	0.99

Next, the accuracy of kriging forecast piezometric heads from the 21 wells over the 636 well network will be examined.

Figure 6.21 shows the 6-month lead regional forecast of piezometric head for September 1988. This map was produced using 6-month lead forecasts (i.e. forecasts of September head levels given March measurements) for each of the 21 modeled wells to kriging estimates over the 636 network according to equation 6.11. Figure 6.22 shows the actual forecast error calculated by subtracting the 6-month lead forecast for September 1988 from the actual observed head for September 1988 over each of the 636 wells. Figure 6.23 shows the experimental cdf for the actual forecast errors. A comparison of Figure 6.19 (the cdf for the kriged estimate, or 0-lead forecast) to Figure 6.23 shows that

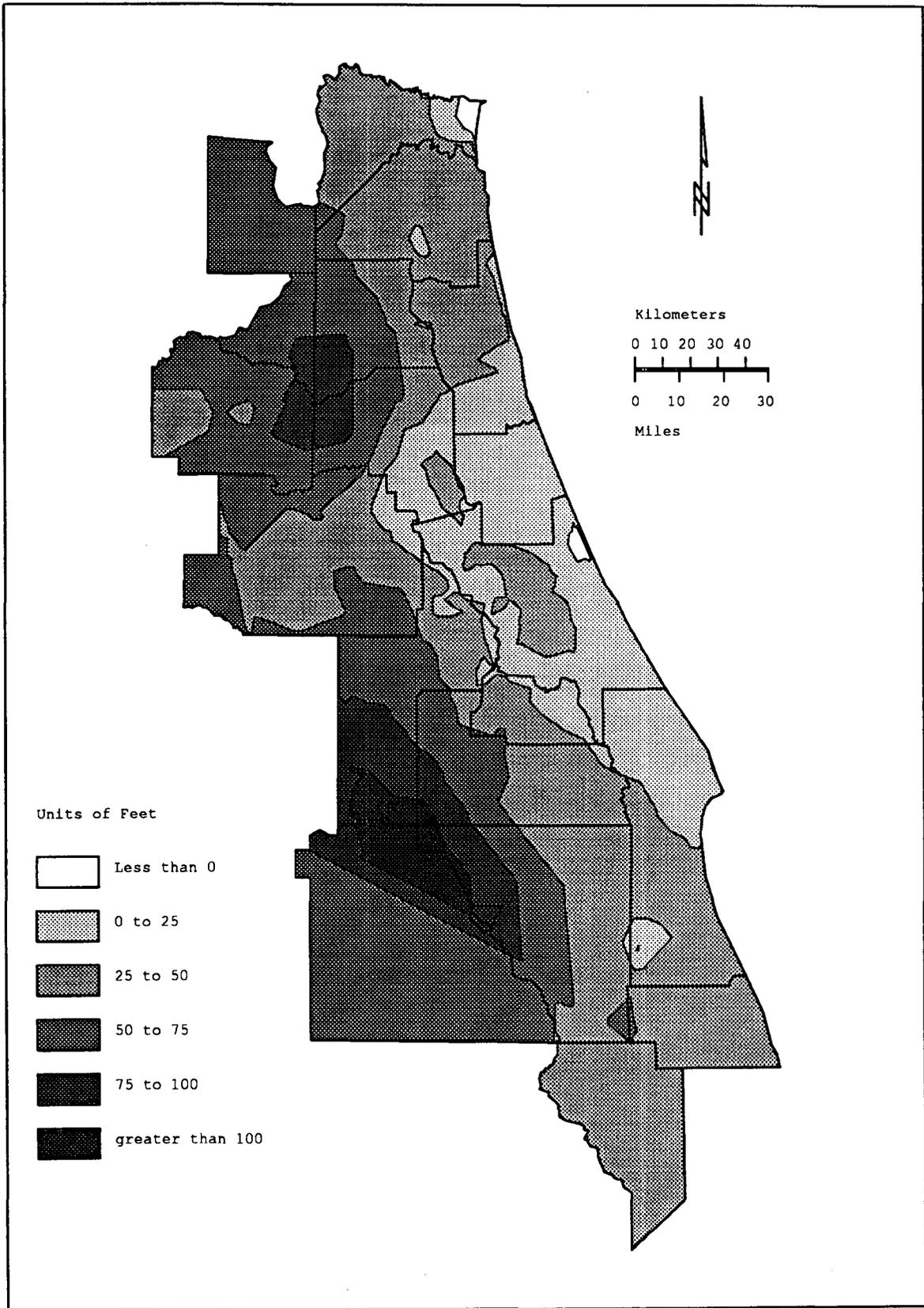


Figure 6.21 Six month lead regional forecast of piezometric head for September 1988.

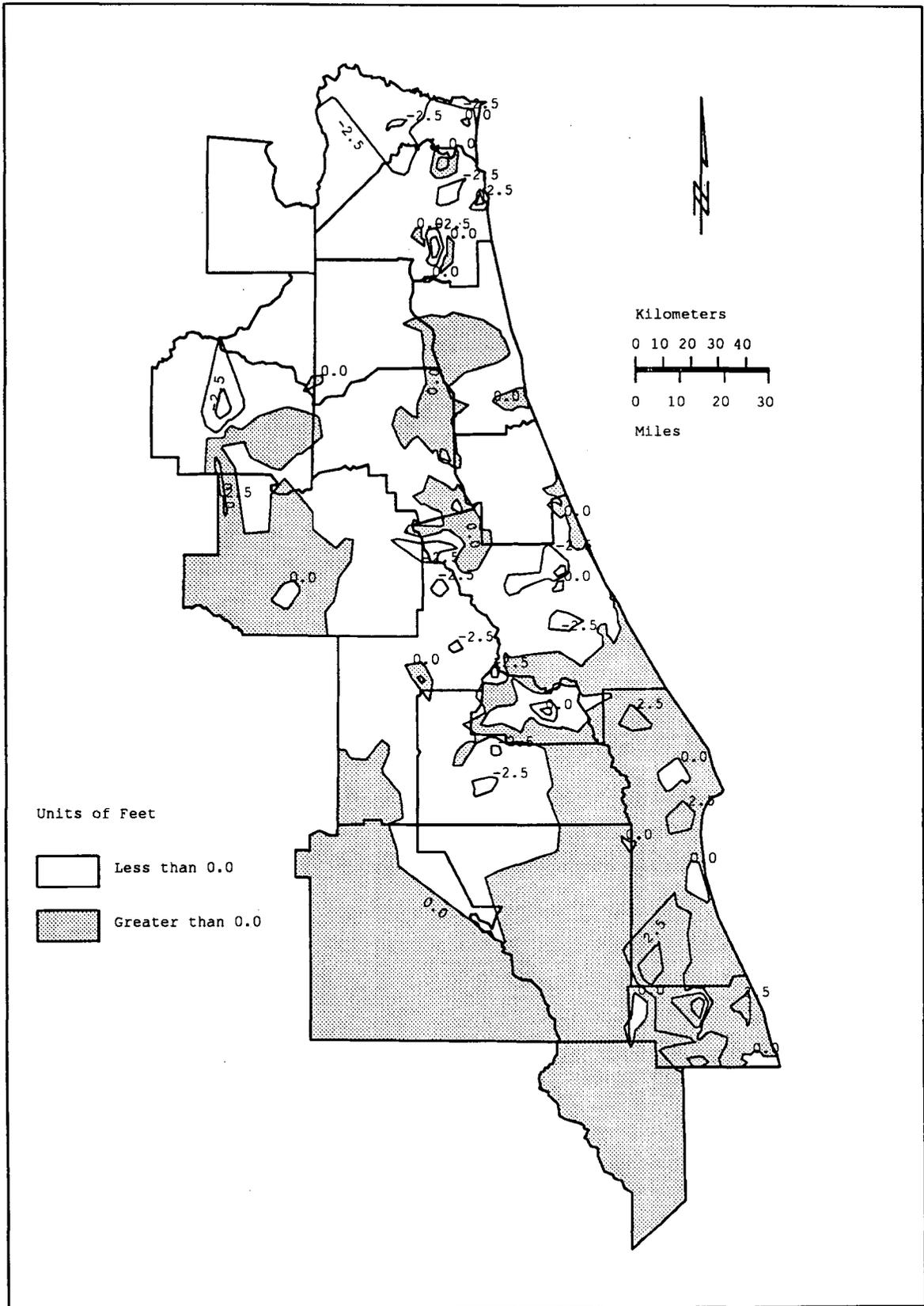


Figure 6.22 Actual forecast error for September 1988 calculated by subtracting the 6-month lead forecast from the actual observed head over each of the 636 wells.

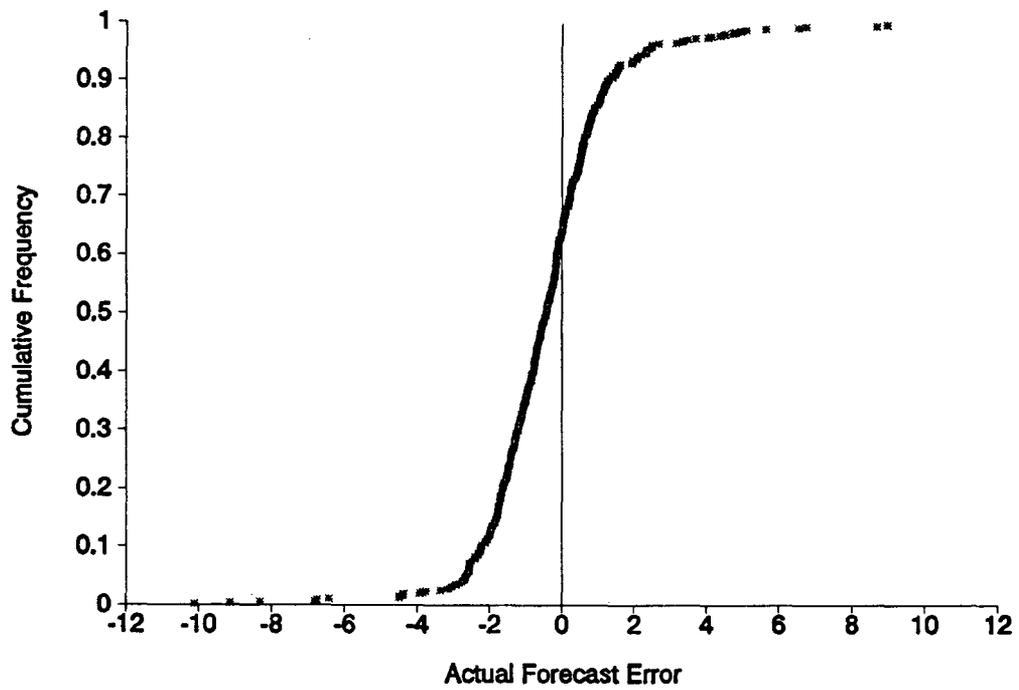


Figure 6.23 Experimental cdf for the actual 6-month lead forecast errors for September 1988.

both sets of prediction errors are approximately symmetrically distributed around zero, but that standard deviation of the 6-month forecast error is larger since it contains error contributions from both the time series forecasting and the spatial kriging processes.

Figures 6.24 through 6.26 and Figures 6.27 through 6.29 show the same series of plots for the 3-month lead forecast and the 1-month lead forecast for September 1988 respectively. Note that the error distribution remains symmetric and the mean error gets closer to zero as the lead time of the forecast decreases. This indicates that the forecast becomes more unbiased as the lead time decreases. The standard deviation of the actual forecast error, however, remains relatively constant as the lead time decreases. This may indicate that the kriging errors, which remain constant over time, dominate the time series forecasting errors which should decrease as lead time decreases.

Figures 6.30 through 6.38 show the regional forecasts, the actual forecast errors, and the experimental cdf's of the actual forecast errors for the 6-month, 3-month and 1-month lead forecasts for May 1989 respectively. Again the forecast errors are approximately symmetrically distributed around a zero mean for all times, and become more unbiased as the lead time decreases. The standard deviation of the actual forecast error ranges from 1.58 to 1.71 feet and again remains relatively constant with time.

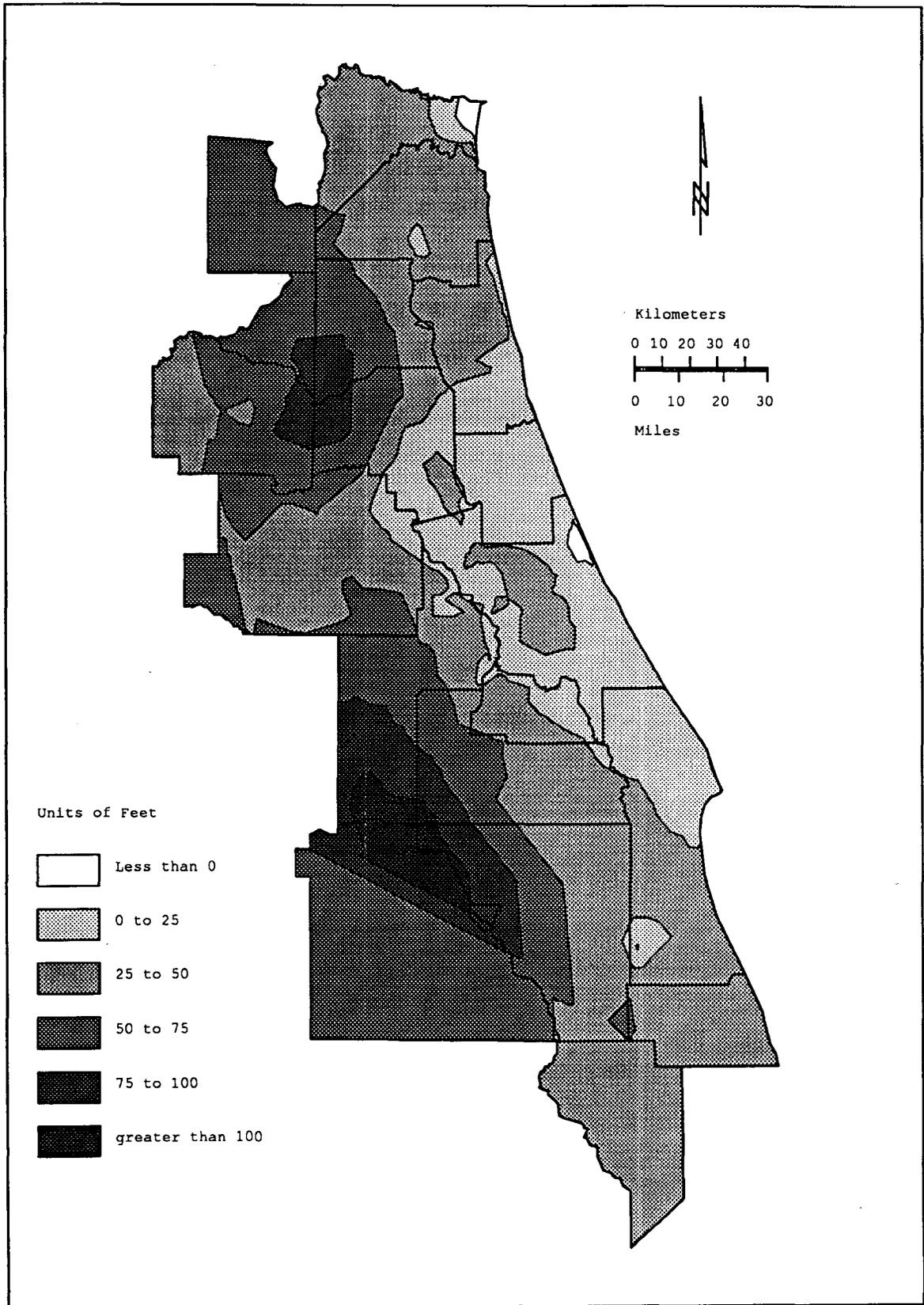


Figure 6.24 Three month lead regional forecast of piezometric head for September 1988.

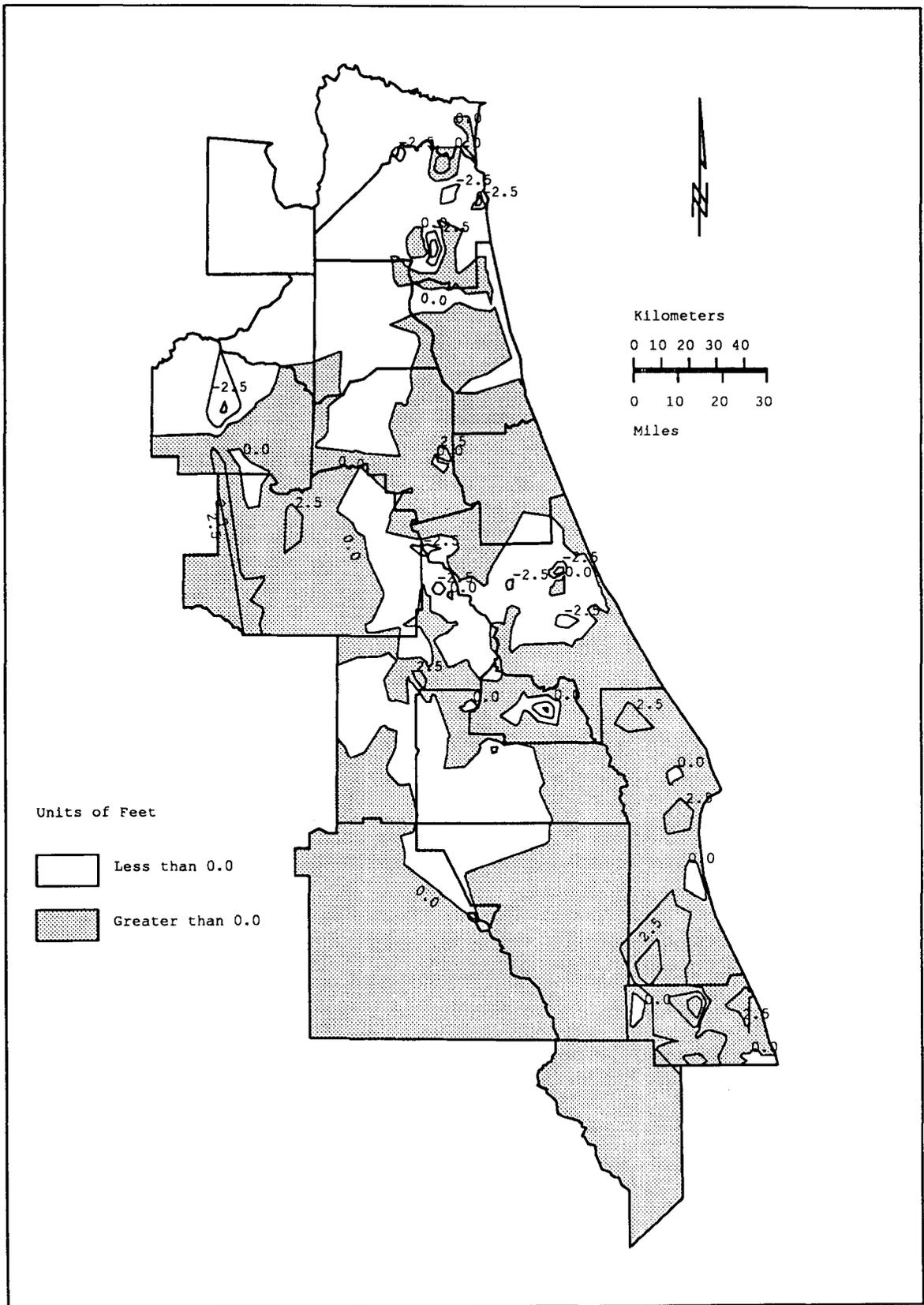


Figure 6.25 Actual forecast error for September 1988 calculated by subtracting the 3-month lead forecast from the actual observed head over each of the 636 wells.

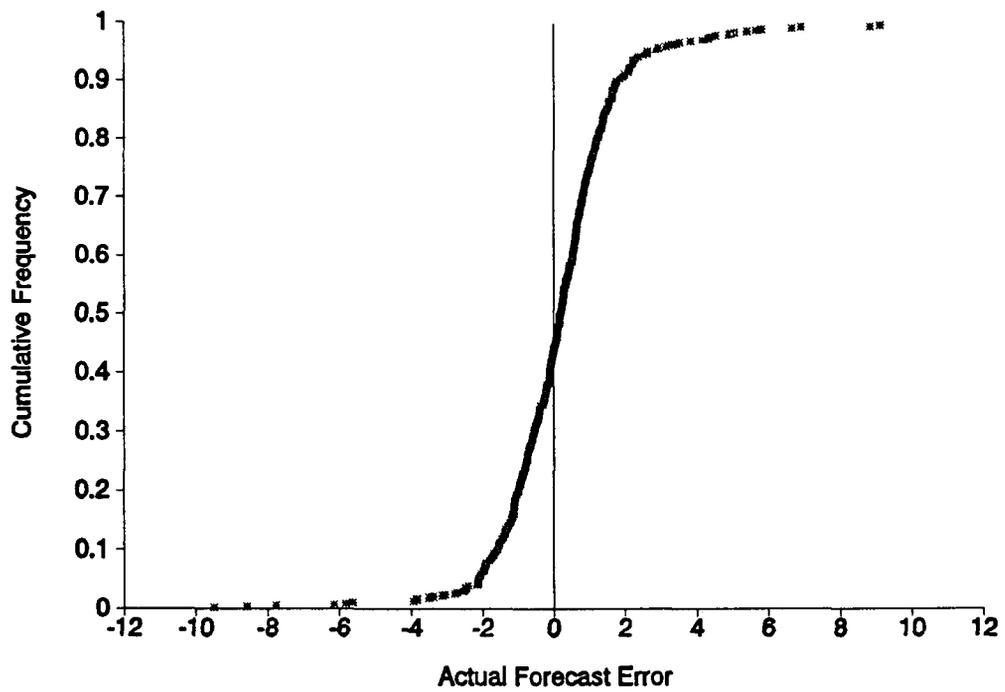


Figure 6.26 Experimental cdf for the actual 3-month lead forecast errors for September 1988.

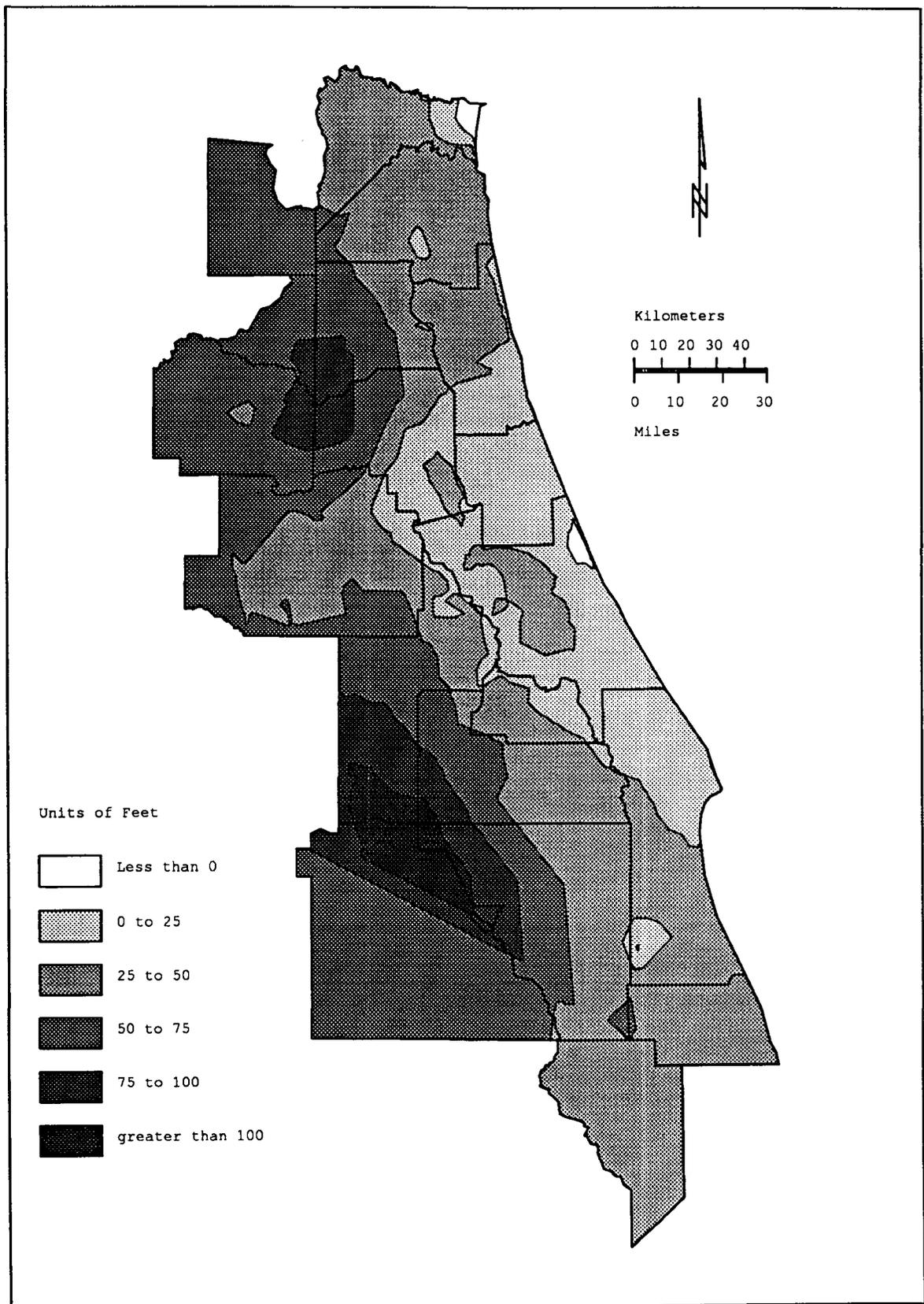


Figure 6.27 One month lead regional forecast of piezometric head for September 1988.

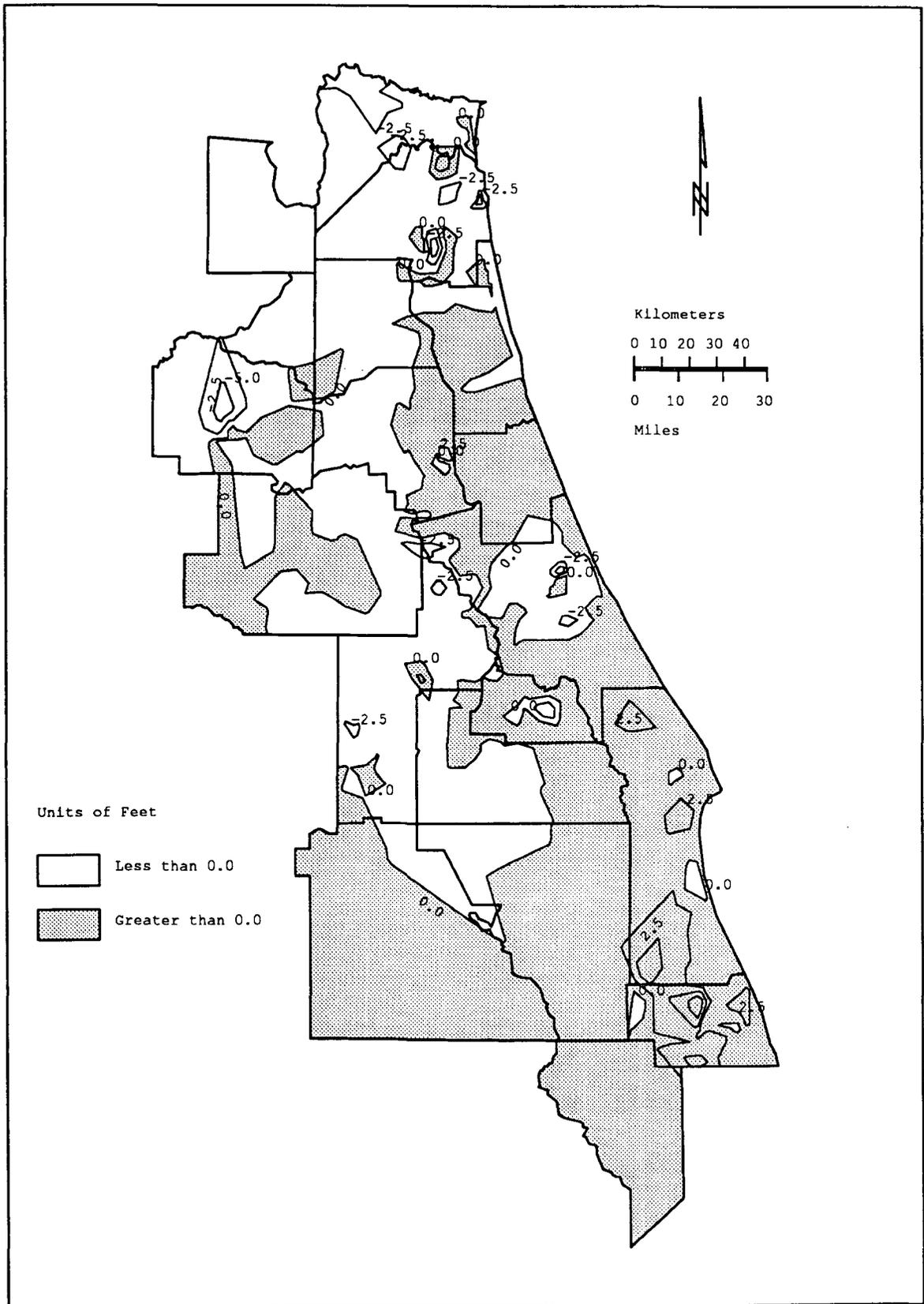


Figure 6.28 Actual forecast error for September 1988 calculated by subtracting the 1-month lead forecast from the actual observed head over each of the 636 wells.

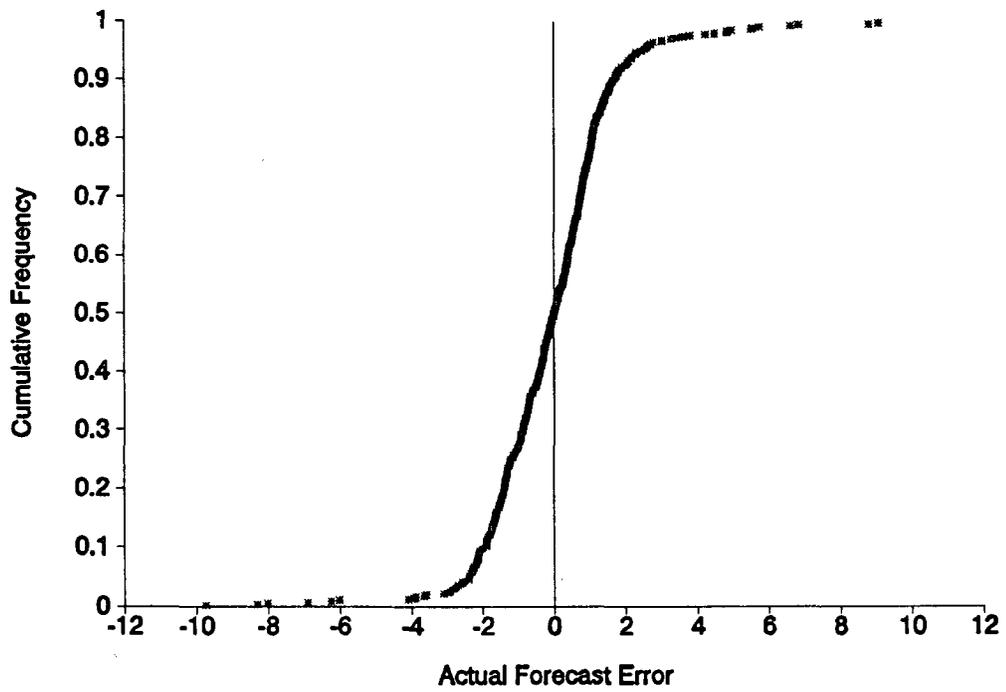


Figure 6.29 Experimental cdf for the actual 1-month lead forecast errors for September 1988.

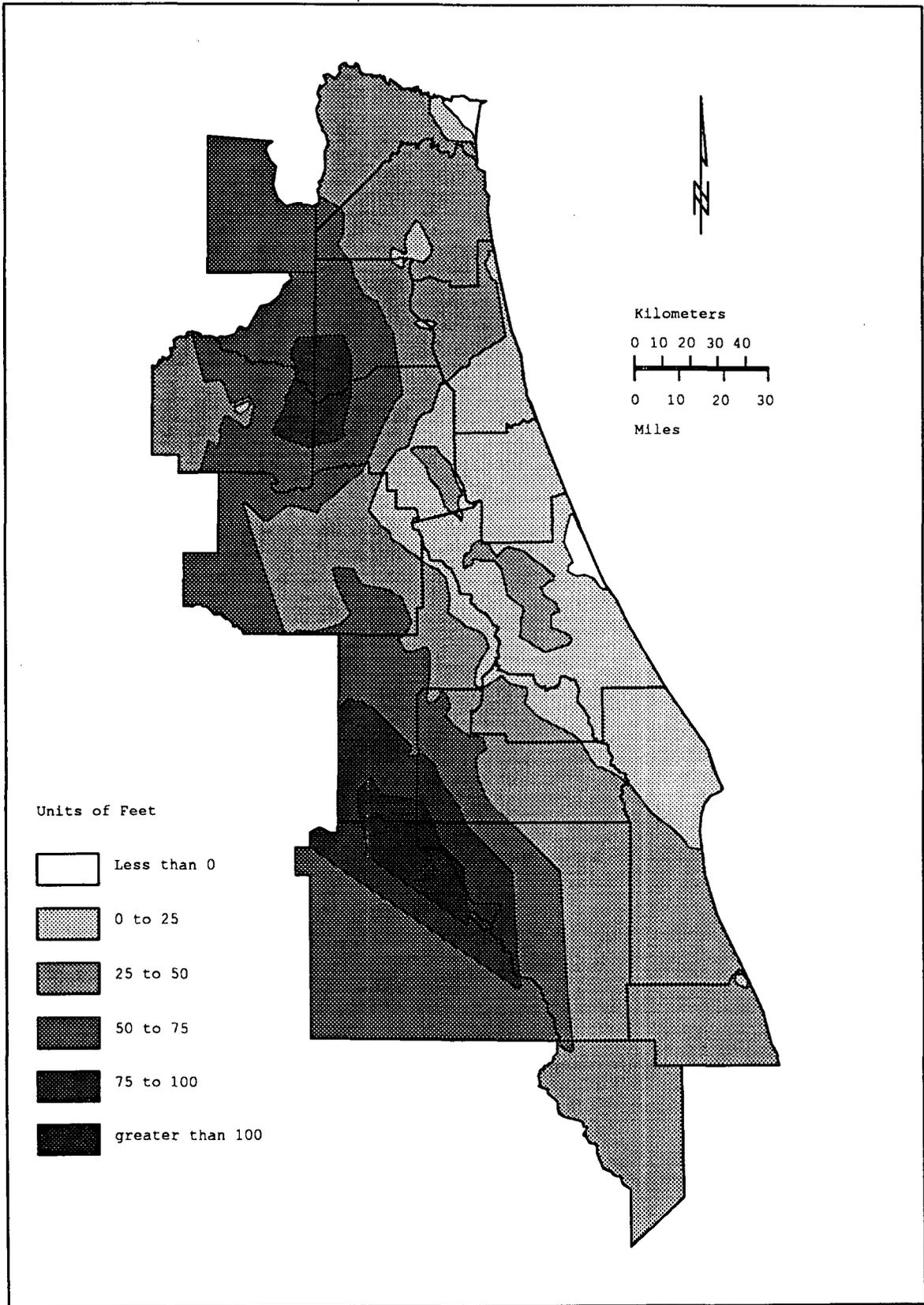


Figure 6.30 Six month lead regional forecast of piezometric head for May 1989.

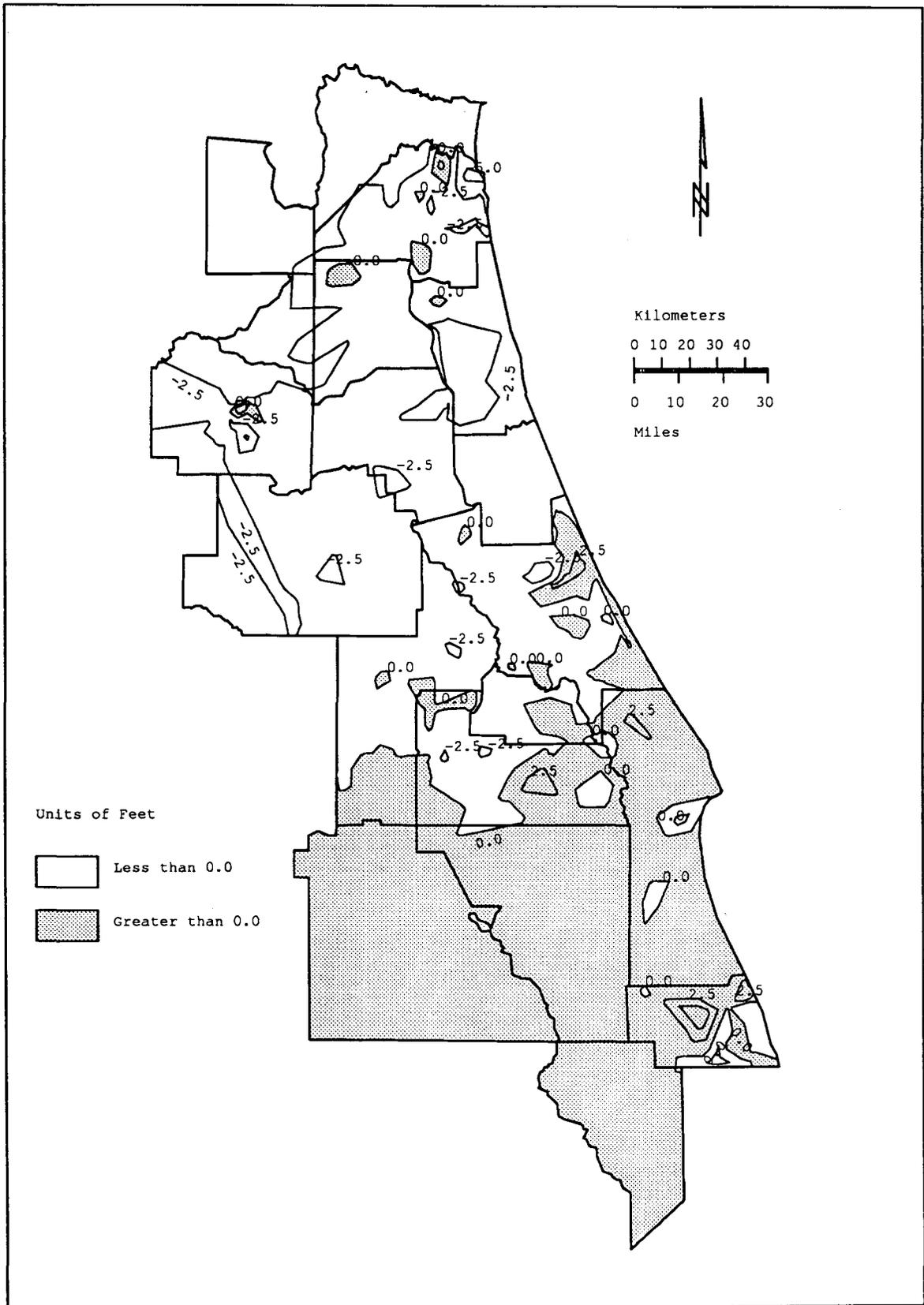


Figure 6.31 Actual forecast error for May 1989 calculated by subtracting the 6-month lead forecast from the actual observed head over each of the 636 wells.

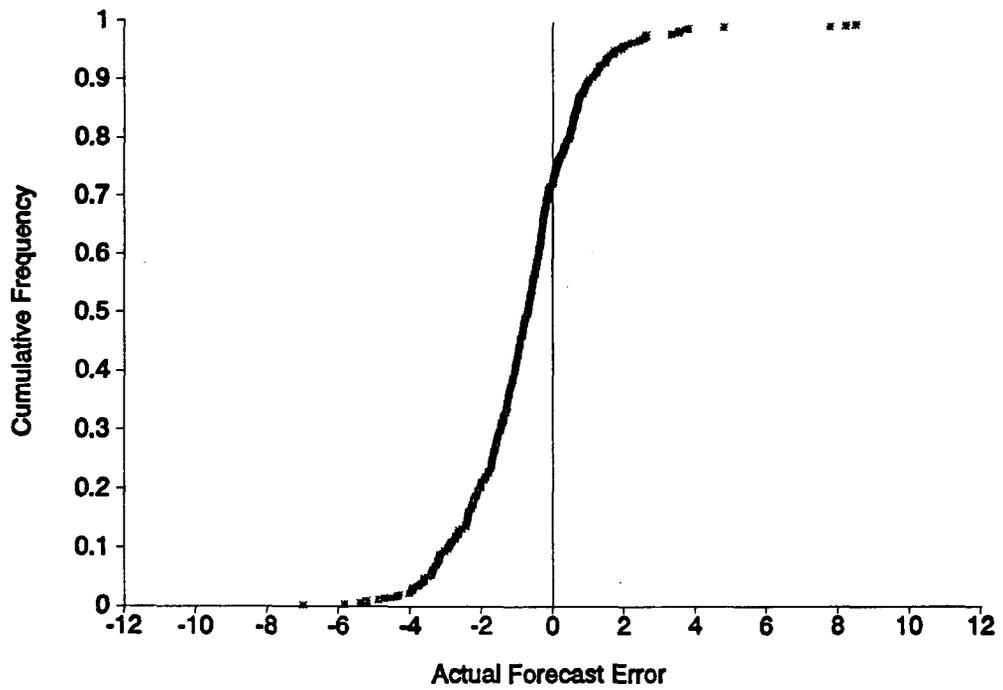


Figure 6.32 Experimental cdf for the actual 6-month lead forecast errors for May 1989.

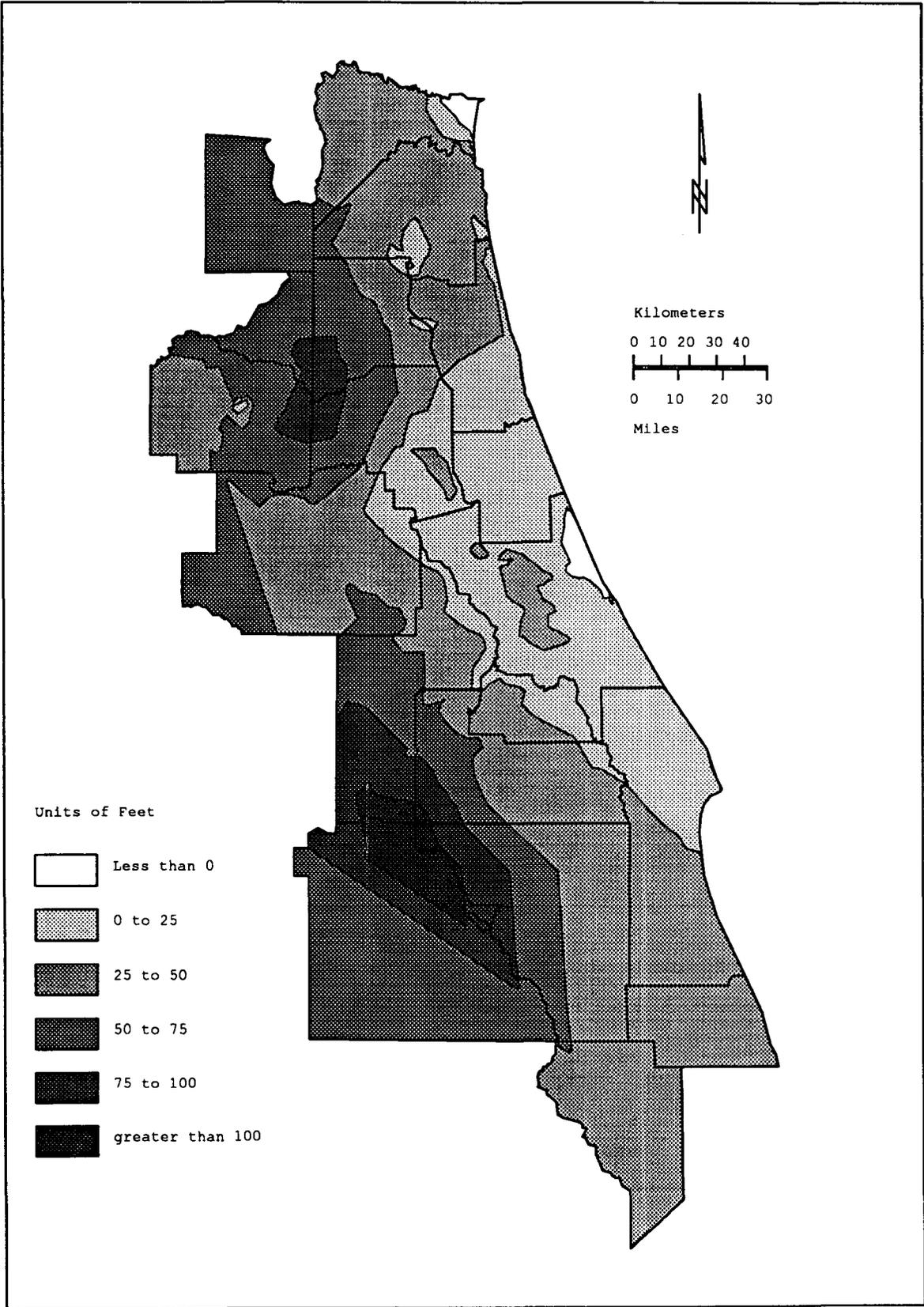


Figure 6.33 Three month lead regional forecast of piezometric head for May 1989.

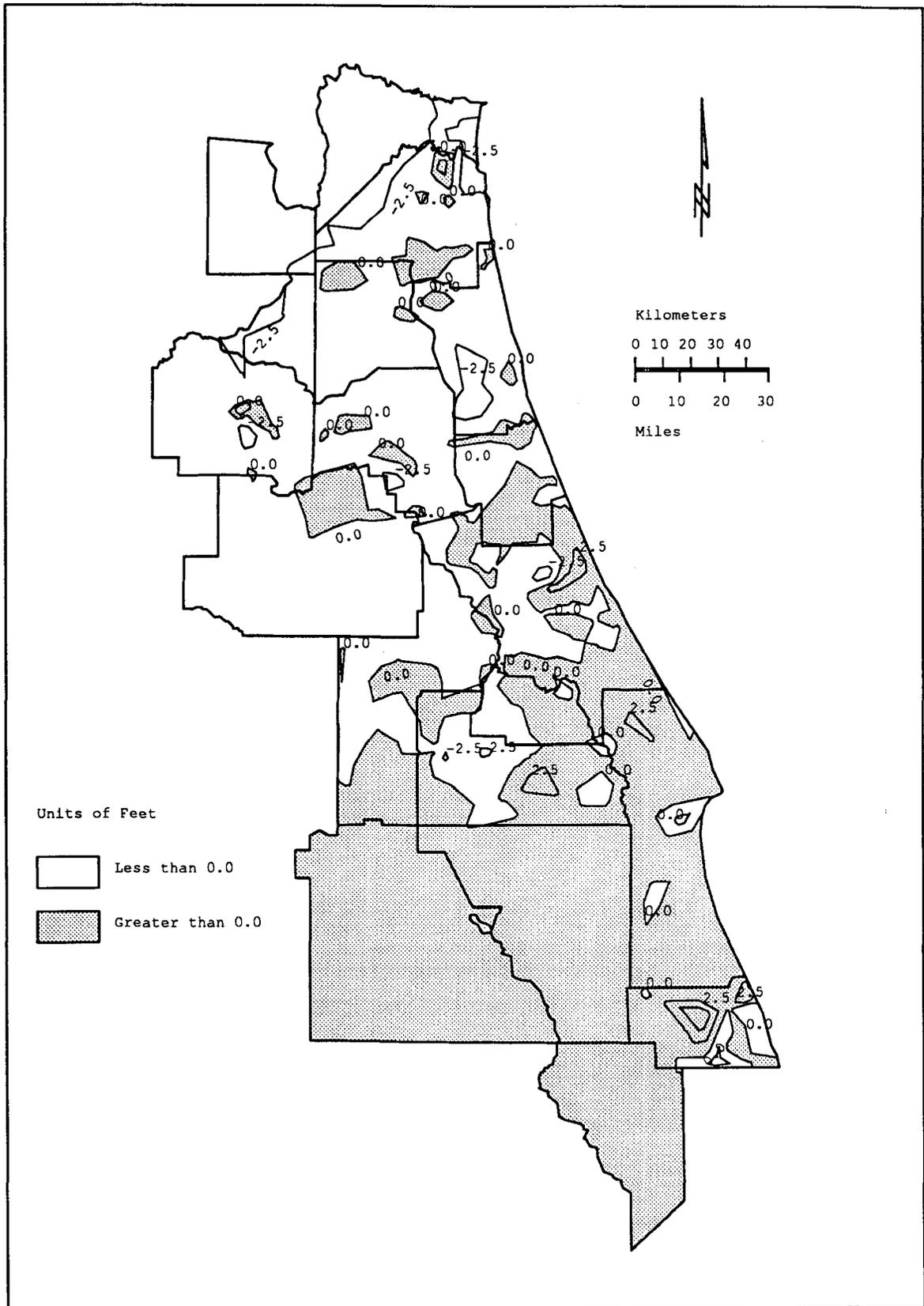


Figure 6.34 Actual forecast error for May 1989 calculated by subtracting the 3-month lead forecast from the actual observed head over each of the 636 wells.

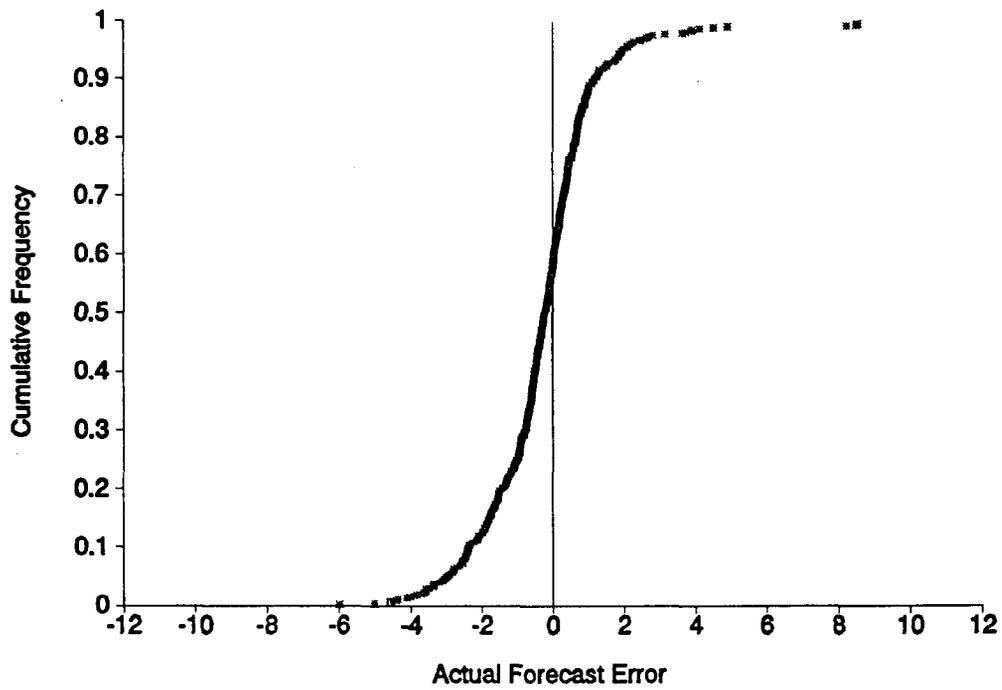


Figure 6.35 Experimental cdf for the actual 3-month lead forecast errors for May 1989.

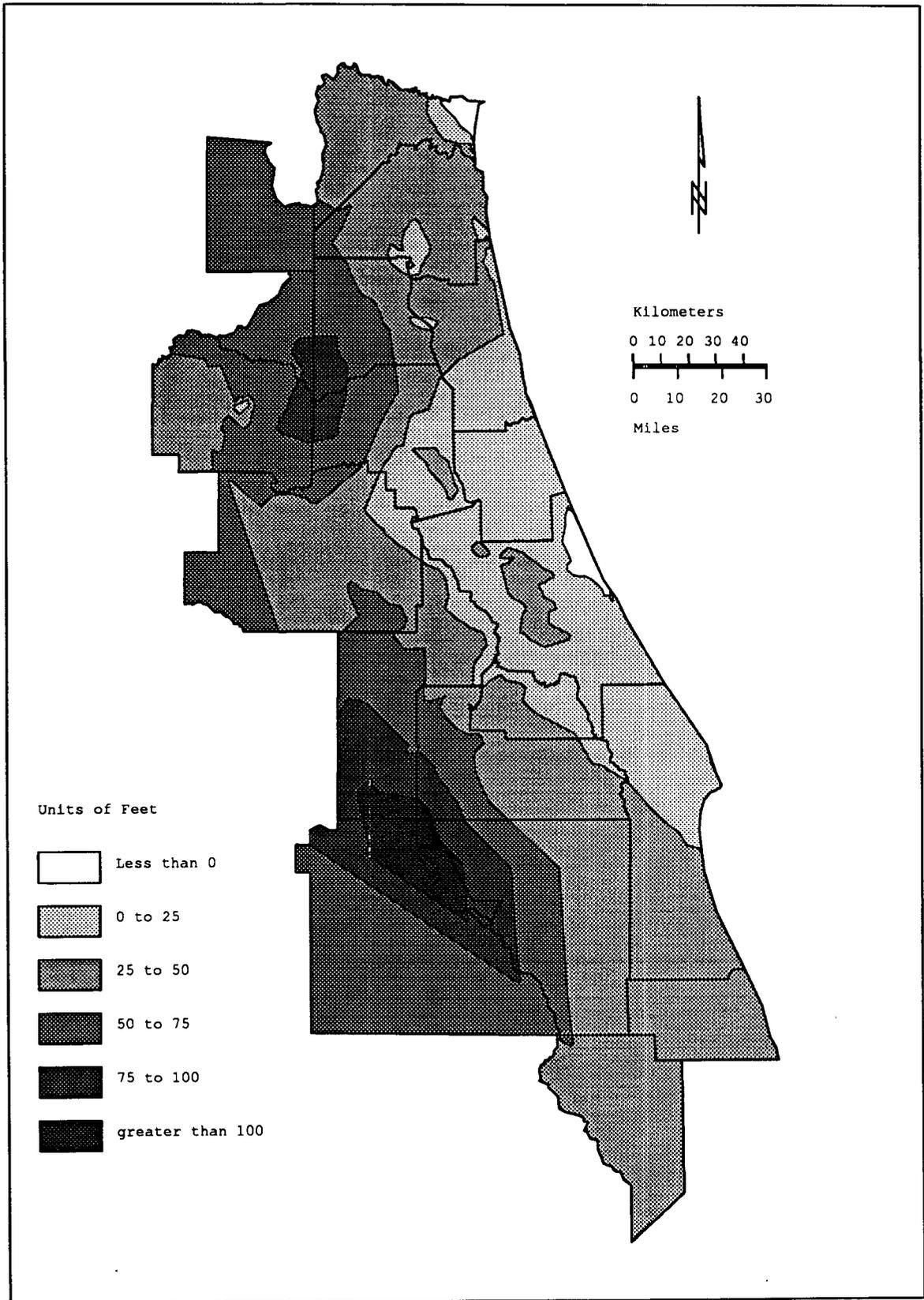


Figure 6.36 One month lead regional forecast of piezometric head for May 1989.

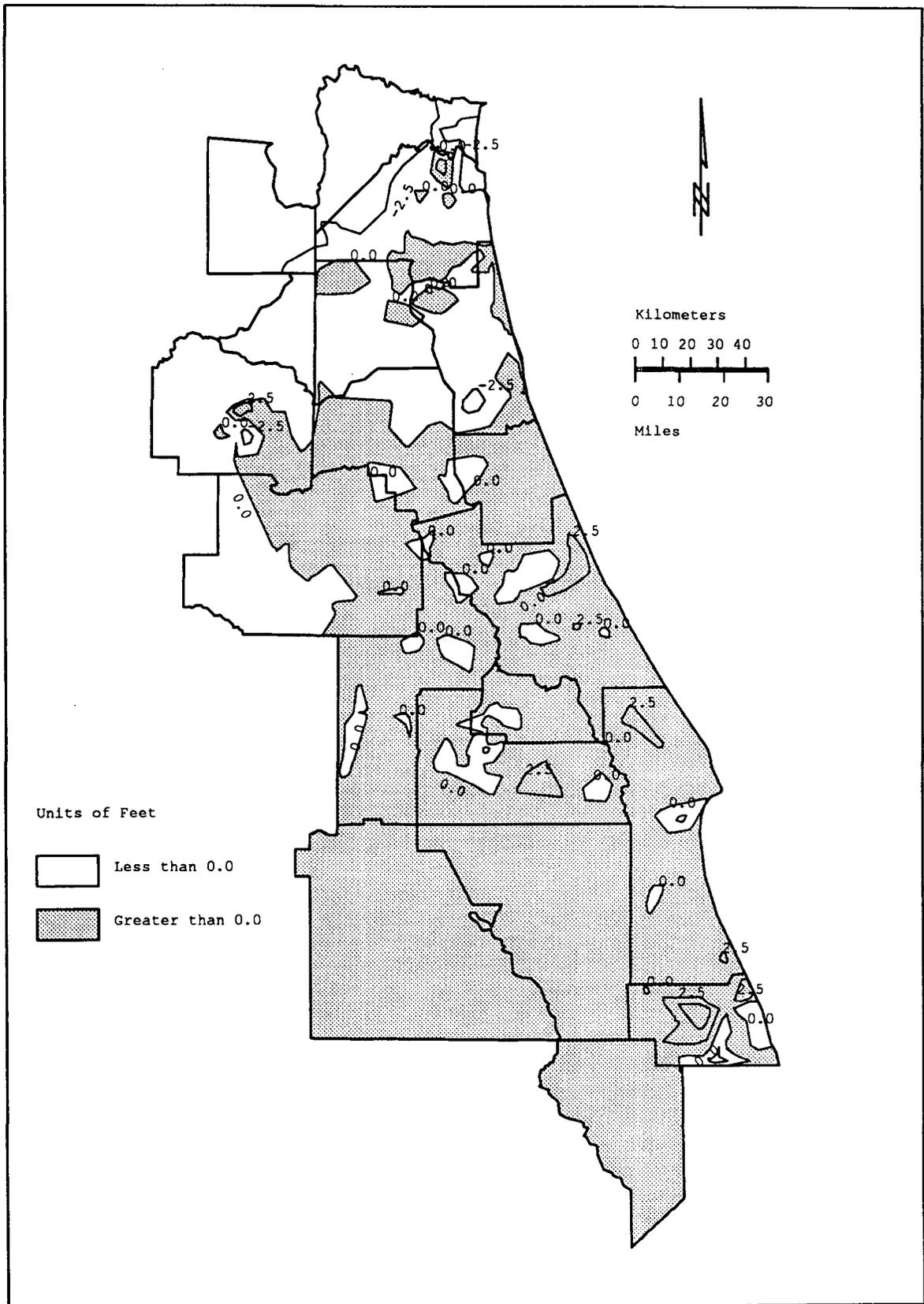


Figure 6.37 Actual forecast error for May 1989 calculated by subtracting the 1-month lead forecast from the actual observed head over each of the 636 wells.

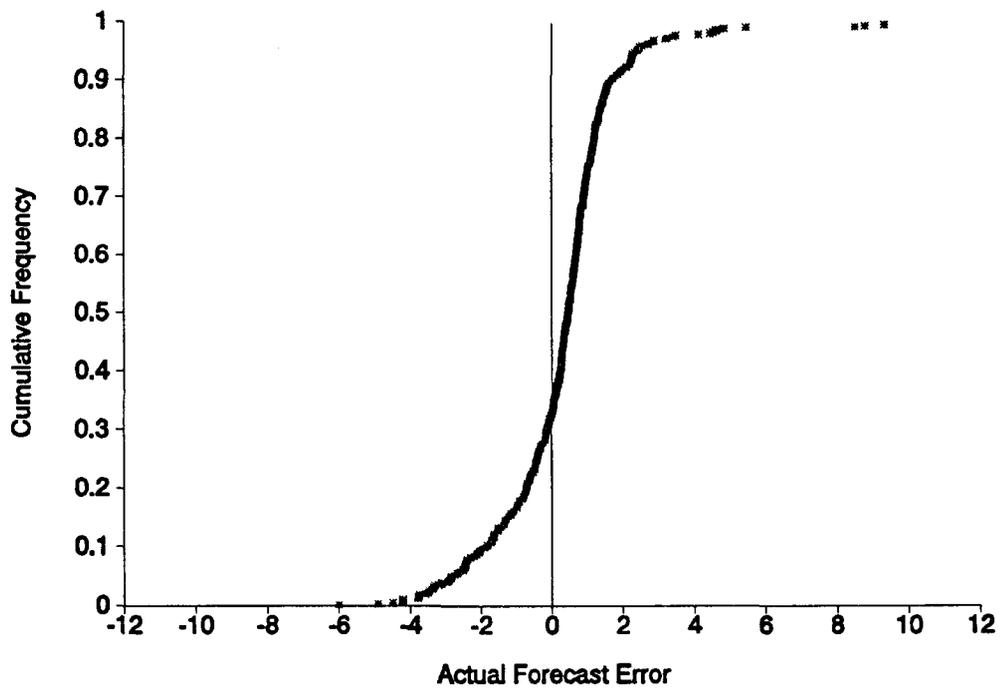


Figure 6.38 Experimental cdf for the actual 1-month lead forecast errors for May 1989.

## Conclusions

In this Chapter, a regional stochastic model was developed that is capable of producing accurate, optimal forecasts of May and September piezometric head levels throughout the District for up to 6 month lead times. The model couples site-specific time series models for 21 wells distributed throughout the District with a regional geostatistical model of the spatial correlation structure of piezometric head. Comparison of regional 1-month, 3-month and 6-month forecasts to observed regional piezometric head levels showed that the forecast errors are unbiased (i.e. symmetrically distributed around a zero mean ) with a standard deviation which ranges from 1.5 to 1.8 feet. The model forecasts future May and September head levels based on behavior of the groundwater flow system observed over the previous 12 to 14 months, and stochastic model parameters fit from long term time series and dense spatial sampling records. To use this model the District must continue to measure piezometric head levels in the 636 well network in May and September of each year, however monthly levels of piezometric head need only be measured for the 21 well network.

## CHAPTER VII LITERATURE REVIEW

### Introduction

The development of physically-based stochastic groundwater models is often preferable to empirical time-series models or traditional geostatistical analysis; first, because their parameters are based on measurable aquifer characteristics rather than a fit to data series, and second, because their predictions incorporate the underlying physics of ground water flow. This section reviews methods of using potentiometric head measurements, point estimates of hydrogeologic variables, and information on the spatial and/or temporal correlation structure of these variables to develop deterministic and/or stochastic regional ground water models.

### Optimal Estimation of Spatially Variable Hydrogeologic Parameters

Development of regional models of groundwater systems often requires point estimates of hydrogeologic variables; these variables include state variables which vary in both space and time (i.e., piezometric head, unconfined aquifer transmissivities, and solute concentrations) and physically-based modeling parameters which vary in space alone (i.e., hydraulic conductivity or confined aquifer transmissivity). Frequently, measurements of state variables and parameters are available from a limited number of locations within a study area; and based on these measurements, estimates are made at all

other required locations. In the past, these estimates were made in a rather ad-hoc manner. Over the last 15 years however, there has been increasing applications of geostatistical methods to groundwater hydrologic problems (i.e., in the estimation of one or more spatially distributed state variables or system parameters).

Kriging is a popular geostatistical technique for making spatial interpolations and mapping variables. The theoretical foundations of the technique are based in random field theory presented in Chapter IV. Measurements represent one realization of the random field of the variable of interest. Traditional kriging uses these measured values and the spatial correlation structure inferred from these measurements, to provide a minimum variance, unbiased linear estimate of the random field at any specified location. In addition, because kriging is a linear estimator, the algorithm can be used to calculate the variance of an estimated random field (conditional variance) for a particular observation scheme before any data are actually collected. Thus, the algorithm could be used in monitoring network design.

Kriging has several advantages over the other applied methods of estimating spatial parameters. First, this method is an "exact interpolator"; meaning it preserves measured variable values at sampling locations within the mapped region. A second important advantage is that kriging produces estimation variances at all locations within the study site. These variances can be used to assess the accuracy of generated maps (Rouhani et al. 1990b) or examine the uncertainty in transmissivity and piezometric heads as a function of sample points (Delhomme 1976 and 1979). Traditional kriging can also be modified to incorporate both measurement errors and large-scale spatial trends into

variable estimates (Gambolati and Volpi 1979). Finally, kriged estimates are more robust than other mapping techniques such as least square estimators and distance weighting methods, i.e., they are less sensitive to small changes in the data (Rouhani 1986).

Early applications of kriging were simply to estimate regional transmissivities from point measurements (Delhomme 1974, 1976, and 1978 and De Marsily 1978). Since these early applications, kriging has also been used to interpolate steady-state piezometric heads (Aboufirassi and Marino 1983; Chirlin and Dagan 1980; Dunlap and Spinazola 1981; and Sophocleous et al. 1982), steady-state hydraulic head gradients (Philip and Kitanidis 1989), soil water content (Yates and Warrick 1987), subsurface electrical conductivities (Yates et al. 1986), and steady-state groundwater contaminant concentrations (Cooper and Istok 1988b and 1988c). Furthermore, the kriging technique has become more general since the early applications; for example estimates can now reflect measurement errors and observed large-scale trends.

Gambolati and Volpi (1979) kriged hydraulic head distributions in three aquifers underlying the Venetian Lagoon. This study was among the first to utilize a kriging method which accounts for a regional trend in the data. The hydraulic head field was modeled as the sum of stochastic and deterministic components, where the stochastic component describes natural variability of head around the deterministic trend. Later, Neuman and Jacobson (1984) developed an iterative method to simultaneously estimate a regionally trending mean and covariance from point observations that improves on the method used by Gambolati and Volpi (1979).

Geostatistics have been used in aquifer parameter estimation modeling or "inverse modeling". The inverse problem of estimating transmissivities from head measurements is generally ill posed and thus transmissivity estimates are extremely sensitive to small changes in the head measurements. Neuman and Yakowitz (1979) used kriged transmissivities as initial input to an inverse model estimating regional transmissivity from steady-state water levels. Similar applications of kriged transmissivities can also be found in Neuman et al. (1980) and Neuman (1980). Inverse modeling predicated on geostatistical foundations tends to produce the most robust estimates (Kitanidis and Vomvoris 1983 and Dagan 1985a). A more thorough review of inverse modeling literature is beyond the scope of this project; although, such reviews already exist (Townley 1983 and Yeh 1986).

Recently, kriging has been used to characterize the spatial distribution of groundwater contaminants. Pollutant concentrations for two dimensional horizontal interpolations must be depth averaged and expressed in units of mass per unit aquifer volume (or mass); otherwise concentrations are not additive as required with linear kriging. Moore and McLaughlin (1980) were among the first to kriged groundwater contaminant concentrations. They investigated the spatial distribution of ruthenium-106 in a plume. Cooper and Istok (1988a) examined the work of Moore and McLaughlin, and they concluded that the spatial analysis was incomplete because pollutant concentrations were not additive.

Cooper and Istok (1988a) described a four-step procedure to spatially characterize steady-state, spatially stationary contaminant plumes with kriging. Steps of the procedure include: 1) an initial conversion of contaminant concentration data into additive contaminant

densities; 2) calculation of the experimental variogram from converted data; 3) fitting a semivariogram model to the experimental variogram; and 4) interpolating unknown contaminant densities with kriging.

In a second paper, Cooper and Istok (1988b) use the above four step procedure, to evaluate groundwater concentrations of boron, barium, iron, manganese, zinc, and total volatile organic carbon (TVOC) at the Chem-Dyne toxic waste site located in Hamilton, Ohio. And finally, in a third paper Cooper and Istok (1988c) present a method of obtaining global estimates of mean contaminant concentrations. Global estimates are obtained from integrating depth average contaminant concentrations over a specified region of the plume. These estimates can be used to locate recovery systems or to simply calculate the mass of contaminant represented by a plume. The method of Cooper and Istok obtains not only global estimates but also global estimation errors for steady-state, stationary contaminant distributions.

#### Optimal Estimation of Spatially and Temporally Variable Hydrogeologic Parameters

Typically, in practice, state variables and hydrogeologic variables have been measured to understand state variable or parameter variations in either space or time. For example it is often the case that water levels are measured at a few locations at frequent intervals, while measurements at neighboring wells are taken less frequently and possibly at irregular intervals. At wells with frequent and regular measurements stochastic time series analyses of temporal water level fluctuations are often conducted (Houston 1983). Whereas, when water levels are measured simultaneously from a network of wells for any given moment, then a geostatistical analysis of the spatial distribution of

water levels may be conducted (i.e., Chirlin and Dagan 1980). In general, however, it is preferable to make optimal estimates (forecasts) of spatially and temporally variable processes using techniques that can account for both temporal and spatial correlation properties.

To date, optimal estimates of hydrogeologic variables have rarely been produced in both space and time. Kriging has generally been used to only interpolate temporally constant parameter and state variable values over space (Rouhani and Hall 1988b). This is because kriging is a static linear estimation technique which is best suited to processes where temporal changes occur relatively slowly (i.e., transmissivities). Rouhani and Hall (1988a) and Rouhani and Wackernagel (1990) however, recently attempted to expand kriging into the space-time domain by using time as the third dimension in the geostatistical analysis of a two dimensional sampling regime for piezometric head.

Consistent time-space mapping can be achieved with Kalman filtering, which is an alternate state variable estimation technique capable of predicting optimal estimates and variances of these estimates for time and space varying stochastic processes. This technique shows promise for use in optimal forecasting of groundwater flow systems or subsurface contaminant plumes which are changing over time. The Kalman filter, like kriging, is a linear estimation algorithm which can be used to combine measurements, taken at discrete points in space and/or time, to obtain minimum variance estimates of the states of a random process whenever and wherever measurement are unavailable. The heart of the Kalman filter is a linear mathematical model, which consists of a system of stochastic equations, formulated in state-space, which describe the propagation of the random process through space and/or time. The form

of the mathematical model can either be estimated empirically from historical data, or derived theoretically based on physical processes.

At measurement times, the Kalman filter computes optimal gains which blend measurements and prior state estimates predicted from the governing stochastic model, to obtain an updated minimum variance state estimate, and an updated state covariance matrix. The updated state estimate and covariance are the conditional mean and variance of the random process based on the measurement information. Since the Kalman filter is a linear estimator, the updated state covariance matrix can be calculated before the measurements are actually taken. Thus, it could be used for the design of efficient data collection networks in both space and time. The mathematical concepts of Kalman filtering and its derivatives (such as extended and iterative Kalman filtering) are developed in detail in Gelb (1974). It can easily be shown that conventional kriging is a simplification of Kalman filtering for a steady-state stochastic model (Chirlin and Wood 1982).

The Kalman filter has been applied to problems in subsurface hydrology by Wilson et al. (1978), Townley (1983), and Graham and McLaughlin (1989a and 1989b). Wilson and his colleagues used an extended Kalman filter and a 2-dimensional groundwater flow model to combine prior information about the head and transmissivity fields with transient head measurements to provide optimal estimates of the transmissivity distribution. Model uncertainty, measurement error, and uncertainty in the transmissivity and head fields were all accounted for in their analysis. An extended Kalman filter was required since the transmissivity and head fields are related nonlinearly in the groundwater flow equations. The main problem Wilson et al. encountered

was the computational burden associated with the large augmented state vector which results from a discretized flow model. Nevertheless, they showed that the extended Kalman filter can provide an effective means of estimating both the mean head and transmissivity fields, based only on head measurements, assumed information about measurement and model error, and the initial state vector and covariance matrix.

Townley (1983) continued the work of Wilson et al. and compared the use of an extended (linearized) Kalman filter to a non-linear weighted least squares approach in simultaneously estimating transmissivities, heads, boundary values and inputs for 1- and 2-dimensional numerical models of aquifer flow. The main difference between the two approaches is that in the weighted least squares approach the transmissivity is not considered to be a random field. Both unconditioned prior estimates of the mean and covariance of the head and transmissivity fields, and prior estimates conditioned on transmissivity measurements (using a kriging algorithm) were tested. As expected, the conditional priors gave superior results in all cases. Townley found that the extended Kalman filter was extremely sensitive to the choice of model noise, and cost twice as much to run as the weighted least squares algorithm -- mainly due to the predictive portion of the filter. Thus he concluded that there was no clear advantage in using the extended Kalman filter when the number of unknown parameters was large.

Graham and McLaughlin (1989a and 1989b) used an extended Kalman filter, based on a physical 2-dimensional solute transport model, to provide optimal estimates and estimation variances of the trajectory of a solute plume in an unknown spatially variable hydraulic conductivity

field. The parameter and state estimates predicted by the physically based stochastic model were updated when measurements of hydraulic conductivity, hydraulic head and/or solute concentrations became available. The performance of the filter was demonstrated using both a synthetically generated solute plume, and field data from an experimental tracer test site. The Kalman filter was found to accurately predict the trajectory of these plumes, and accurately estimate the reliability of its predictions. Again the major problem encountered in the implementation of the Kalman filter for the solute transport problem were the long computation times and large computer memory requirements.

It should be noted that the regional stochastic model for piezometric head level presented in Chapter VI of this report can be interpreted as a simplified Kalman filter. In this case the stochastic model is the set of empirical time series models at 21 discrete locations in space. Updates to each model are made whenever new forecasts or measurements are available. The updated model forecasts are then optimally interpolated over space using the temporally-constant spatial covariance structure of the piezometric head field.

#### Optimal Estimation with Augmented Data

Frequently it is the case that estimates of a random field cannot be predicted accurately from measurements of a single hydrogeologic variable. This occurs because of too few measurements, or because measurements are not distributed throughout the region of study. To overcome these data limitations, measurements can be augmented with data or information on other hydrogeologic variables before spatial or

temporal interpolations are performed with kriging or Kalman filtering. Several methods exist by which data from one measured variable can be supplemented with measurement or information on an ancillary variable. The most obvious methods include: 1) kriging combined with linear regression modeling, 2) kriging that incorporates external drift of a secondary variable, 3) kriging with "guess field", and 4) simple cokriging. Ahmed and De Marsily (1987) compared the above four methods in a case study of interpolating transmissivities from data on both transmissivity and specific capacity.

#### Kriging Coupled with Linear Regression

Kriging combined with linear regression modeling is the simplest approach to augmenting a data set. This method can be used if point values of a primary variable can be estimated from a simple linear regression model involving the secondary variable. First, a linear regression model is derived from point locations where both the specific capacity and transmissivity are measured. Then additional transmissivity estimates are calculated from the regression relationship at locations where only the specific capacity is known. The errors associated with regression model predictions are also considered. Delhomme (1974, 1976, and 1978) developed the initial kriging equations that incorporate measurement or variable error. Ahmed and De Marsily (1987) give a complete system of equations required when kriging a set of variable values developed from point measurements and regression model estimates.

Kriging coupled with linear regression works for no more than two variables, and it requires that sufficient data exist to: 1) create a

regression model, and 2) estimate the primary variable from the secondary variable measured at locations where the primary variable has not been measured. Ahmed and De Marsily (1987) tested this method to supplement measured transmissivities in a random field with additional transmissivities estimated from available specific capacity measurements. They found the method easy to apply; although, a poorly fitted regression model will increase the uncertainty of kriged variables.

#### Kriging with an External Drift

Kriging with external drift is another method of augmenting available data. The name of this method does not imply that the primary variable is nonstationary (Ahmed and De Marsily 1987), only that the mean of the primary variable is strongly correlated with a well sampled secondary variable. The conditional expectation of the primary variable is written as a linear function of the secondary variable; consequently, this alters the classical system of kriging equations to a system directly dependent on the secondary variable. A linear model is used to estimate the primary variable from the secondary variable. Prediction residuals are calculated by subtracting linear model estimates of the primary variable from measured values. Generally variograms for the residuals and the primary and secondary variables are used in this method. But for the stationary case, variograms of the main variables are sufficient. Ahmed and De Marsily examined this method to augment transmissivity data with specific capacity data. They found the method to be complicated but effective when the primary variable is both under and equally sampled with respect to the secondary variable. Variograms

of both variables are needed, but the method does not require measurements of both variables at common locations.

#### Kriging with Guess Field

Another simple technique of supplementing data with additional hydrologic information, is kriging with a "guess field." This method was first proposed by Delhomme (1978) to improving mapping accuracy. In this procedure, an estimate of a spatial variable is first obtained throughout the region of interest with an independent model (i.e., piezometric surface could be estimated with a numerical model using constant parameters). Next deviations between the estimates and point measurements are calculated over the domain, and a variogram of the deviations is calculated. A map of deviations is then kriged and added to the map of initial estimates to obtain a corrected variable map.

Ahmed and De Marsily (1987) used a linear regression model to estimate transmissivities from specific capacity estimates. Transmissivities estimated from the regression model were initially kriged to obtain a "guess field". Next, at point locations where both transmissivity and specific capacity were known, regression model prediction errors were calculated. These model errors were kriged. Finally, the interpolated errors were added to the guess field to give a corrected transmissivity field. Ahmed and De Marsily concluded that the guess field approach is best applied the if secondary variable has been measured (or predicted independently) at all locations where the primary variable has been measured.

## Cokriging with Statistically Estimated or Physically-Based Correlation Structures

Cokriging is an extension of kriging which can improve the accuracy of random field estimation by sampling from, and considering the correlation between two or more related random fields. Ahmed and De Marsily (1987) found cokriging effective for undersampled as well as equally sampled primary and secondary variables. However, to statistically estimate cross-variograms from data requires a large number of points where both variables have been simultaneously measured. For stationary fields where mean parameter values are known, the cross-covariances can be used in lieu of the cross-variograms. The statistical estimation of the cross-covariance does not require that both variables be simultaneously measured at each location.

Before cokriging, it is necessary to characterize the correlation structure between primary and secondary variables through appropriate joint first and second moments (i.e., the mean and covariance of the secondary variable and the cross-covariances function between the variables). With sufficient data, these moment functions can be estimated by developing direct and cross variograms from the data (Journel and Huijbregts 1978).

In real-life applications however, there is often insufficient data available to obtain reliable statistical estimates of the mean, covariances, and cross-covariances. In these cases, the use of physically-based analytically derived covariance and cross-covariance functions is desirable. Moments are obtained for a dependent variable (e.g. piezometric head) given a model solution to an applicable boundary value problem (e.g. a flow model) and the first and second moments

(i.e., mean and covariance) of the independent variable (e.g. transmissivity). After the mean and the covariance functions for the dependent variable have been characterized from field data, and the cross-covariance function for the dependent-independent variables has been derived from a physical model, the estimation problem is solved under the same framework of cokriging (Rouhani et al. 1990b). In a typical application, the mean and covariance functions for transmissivities would be specified and then a flow model would be used to characterize the mean and covariance of the piezometric surface and the cross-covariance function.

Several methods have been used to derive required first and second moments of the dependent variable including: Monte Carlo simulations; analytical spectral methods; numerical small-perturbation approximations; and lagrangian methods. Kitanidis and Vomvoris (1983) and Hoeksema and Kitanidis (1984) used physically-based covariances and cross-covariances, derived from numerical small-perturbation approximations, to cokrige a transmissivity field based on head and transmissivity observations. Gutjahr and Wilson (1985) demonstrated the cokriging of hydraulic head field from head and transmissivity measurements based on spectrally derived covariance and cross-covariance functions. McLaughlin and Graham (1986) used cokriging techniques to estimate a steady-state concentration field based on observations of hydraulic conductivity, head, and concentration, and spectrally derived covariance and cross-covariance functions. Graham and McLaughlin (1989a and 1989b) extended the method to estimate a transient concentration field based on observations of hydraulic conductivity, head, and concentration, using a Kalman Filtering technique. In this study

physically-based covariances and cross-covariances were derived using numerical small-perturbation approximations. Details on alternative methods of deriving physically-based covariances and cross-covariances are discussed in the next section.

### Derivation of Physically-Based Correlation Structures

#### Monte Carlo Methods

Some of the earliest work in stochastic subsurface hydrology concentrated on Monte Carlo simulations of flow in heterogeneous porous media. Simulations of this type require a random field generator to create multiple realizations of the spatially correlated independent variable (i.e., transmissivity), a numerical flow and/or transport model for the boundary-value problem of interest, and software for calculating the mean, covariance, and cross-covariance of the resulting realizations of the dependent variable (Rouhani et al. 1990b).

Freeze (1975) and Smith and Freeze (1979a and 1979b) used Monte Carlo simulations to derive the first and second moments of the hydraulic head distribution, based on 1- and 2-dimensional flow models and a random hydraulic conductivity field with known statistics. They found that the variance of the hydraulic head increased with increasing hydraulic conductivity variance, increasing hydraulic conductivity correlation scale, and increasing mean hydraulic gradient. In addition, they found that the variance of hydraulic head decreased by approximately one half when going from 1- to a similar 2-dimensional problem. These results, which have also been found by others (Dagan 1979; and Gelhar 1982), point out that correctly predicting the derived moments for a particular problem depends on correctly determining the

appropriate dimensionality, and recognizing the inter-dependence of the first and second moments of the derived random fields.

Smith and Schwartz (1980, 1981a, and 1981b) used Monte Carlo methods to study contaminant transport in randomly generated, spatially autocorrelated hydraulic conductivity fields. First they investigated the relationship between random variations in hydraulic conductivity and macroscopic dispersion using a lagrangian particle tracking algorithm. They found that, in general, a constant macrodispersivity could not be defined either for each individual realization of the Monte Carlo simulations or for the ensemble behavior as a whole. Increasing the variance and correlation scale of the hydraulic conductivity was found to lead to greater variability in the concentration distribution. The concentration distribution was found to be very sensitive to the specific arrangement of hydraulic conductivity in each realization.

Smith and Schwartz also analyzed the uncertainty in mass transport predictions as characterized by the frequency distributions for initial breakthrough, time of maximum mass and last arrival of mass at a particular boundary. Again they found that transport uncertainties were highly sensitive to particular features of the spatial structure of the hydraulic conductivity field. Finally they conducted a conditional Monte Carlo simulation which investigated the effect of uniformly spaced hydraulic conductivity measurements in reducing uncertainty in mass transport predictions. They concluded that although gathering hydraulic conductivity measurements could be useful in defining ensemble statistics for the hydraulic conductivity field, mass transport prediction uncertainties did not decrease significantly even for their most dense measurement grid.

Smith and Schwartz's discouraging results regarding the inability to define a constant dispersivity for heterogeneous porous media can perhaps be attributed to the fact that the length scale of their simulation was only approximately ten times their hydraulic conductivity correlation length, and thus not long enough to allow for sufficient spatial averaging of the particle paths. Nevertheless, their effort indicates that at early times (i.e. the first five to ten years) mass transport cannot be described as a Fickian process.

Monte Carlo techniques are advantageous in that they can handle bounded, non-linear systems with relatively large input variances, and that they are easy to understand and implement. However the computational requirements for a Monte Carlo simulation, and particularly a conditional simulation, restrict its use in practical situations. In general, Monte Carlo techniques are probably best suited for checking the assumptions and results of derived moment problems solved using other methods.

### Spectral Techniques

Another technique being used for stochastic analysis of flow and transport in heterogeneous porous media is analytical spectral analysis. Using this method all random variables in the governing differential equation are expanded into a mean term and an assumed small, zero mean stochastic perturbation around the mean. First order mean and perturbation equations are derived from the expanded equation. Solution of the first order mean equation yields information regarding the effective or mean properties of the random fields which can be used in large scale prediction. Solution of the first covariance equation

yields information regarding the variation that can be expected around the predicted mean. Results obtained using spectral solutions to perturbed stochastic equations are generally closed form analytical expressions, in contrast to the numerical results obtained using Monte Carlo simulation.

Bakr et al.(1978), Mizell et al.(1982), Gutjahr et al. (1978) and Gutjahr and Gelhar (1981) all used spectral methods to evaluate problems of 1-, 2-, and 3-dimensional steady state flow in stationary, spatially correlated, random hydraulic fields with known statistics. They found first order analytic expressions for the effective hydraulic conductivity, mean hydraulic head, hydraulic head variance and head covariance functions which explained and verified the qualitative results reported by Freeze (1975), and Smith and Freeze (1979a and 1979b). However, solution of the steady state perturbed flow equation by spectral methods requires that both the mean-removed head and hydraulic conductivity fields be stationary over an infinite domain. For the steady state flow equation this translates into a requirement that the mean fields have at most a linear trend (i.e. the hydraulic gradient must be constant). This assumption restricts the direct application of the results to regions far removed from boundaries or pumping/recharge sites, where perturbations around a linearly trending mean are more likely to be small and have zero mean.

Gelhar et al.(1979), Gelhar and Axness (1983), and Vomvoris and Gelhar (1985) have used spectral analysis to evaluate mass transport in heterogeneous porous media. The earlier work focused on evaluating the additional solute flux term, or macrodispersivity, that appears in the first order mean equation for steady state mass transport due to the

correlation between the flow and concentration perturbations. By assuming stationarity of the mean removed flow and concentration fields, Gelhar and Axness used the spectral representation theorem to evaluate an ensemble macrodispersion term for steady 3-dimensional flow in an anisotropic hydraulic conductivity field with arbitrary orientation of stratification. They found that macroscopic dispersivity increased with increasing hydraulic conductivity variance, hydraulic conductivity correlation length, and mean head gradient, but was independent of local dispersion for the general case.

For the steady state transport equation, stationarity of the head and concentration fields implies that both the mean hydraulic gradient and the mean concentration gradient are constant. The constant mean concentration gradient assumption is clearly a restrictive assumption that cannot be applied rigorously over the scale of a real-life contaminant plume. Therefore it is assumed that the mean concentration is only locally constant, over a scale much smaller than that of the plume. Thus, at any point of interest, the macrodispersivity is obtained using a local linearization of the actual mean concentration field.

Gelhar and Axness (1983) calculated macroscopic dispersivities which were consistent with field observations and numerical Monte Carlo simulations. Use of the coefficient in the mean mass transport equation for prediction would require that 1) the constant mean concentration gradient assumption be relaxed after the macroscopic dispersivities are calculated, and 2) that the concept of ergodicity, which allows ensemble statistics to be applied to single realizations of random fields, be invoked.

Vomvoris and Gelhar (1985) concentrated on evaluating the variance and covariance functions for the concentration field based on the mass transport perturbation equation. Again, they assumed that the mean head and concentration gradients were locally constant to apply the spectral representation theorem. The concentration variance and covariance functions they derived were found to increase with increasing hydraulic conductivity variance, correlation scale and mean hydraulic gradient. In addition, the concentration variance increased with the square of the local mean concentration gradient and was inversely proportional to the local dispersivity.

The concentration variance found by Vomvoris and Gelhar provides valuable insight into the factors which contribute to uncertainty in mass transport prediction, but the direct applicability of the covariance function is unclear. The linearly trending mean concentration field that the covariance analysis is based upon is not the solution of the mean mass transport equation which is used to derive the perturbation equation.

The spectral approach for deriving the moments of the random head and concentration fields is one of the most computationally efficient and elegant methods of derived distribution analysis. Unfortunately, unlike other derived distribution methods, there is no mechanism within the spectral framework for updating knowledge about the particular realization of interest, and thus reducing prediction uncertainty as observations are gathered. For additional applications of spectral methods, the reader is referred to Dagan (1982a and 1982b), Gutjahr et al. (1978), Kitanidis and Vomvoris (1983), and Rubin and Dagan (1987 and 1988).

## Numerical Small-Perturbation Approximations

Numerical small-perturbation approximations (state-space methods) obtain truncated mean and covariance equations for derived random fields by applying the expectation operator directly to the governing stochastic differential equations. The resulting deterministic system of equations is then typically solved numerically in the spatial domain. Numerical solutions of the moment equations allows for consideration of non-stationary random fields, source/sink terms, and boundary conditions which cannot easily be dealt with using the spectral method.

First order state-space solutions for the analysis of uncertainty in hydraulic head predictions was investigated by Sagar (1978), Dettinger and Wilson (1981), Townley (1983), Hoeksema and Kitanidis (1984,1985), McLaughlin (1985), and Graham and McLaughlin (1989a). Sagar and Dettinger and Wilson solved the derived moment problem by first discretizing the governing transient 2-dimensional flow equation in state-space using finite element and finite difference methods, respectively. The solution to the discretized equation was then expanded in a Taylor series about the expected value of all uncertain parameters (i.e., heads, transmissivities, and boundary conditions). First and second order estimates of the first and second moments of the head distribution are then obtained by applying the expectation operator to the appropriately truncated Taylor series expansion of the head solution.

Dettinger and Wilson found that a first order expected value of head was identical to the head field predicted by deterministic methods using the mean hydraulic conductivity. Second order contributions to the expected value however, modified the first order estimate based on

transmissivity uncertainty. The second moment (covariance) of the head field was found to be a function of the transmissivity covariance function, uncertainty in the boundary conditions, and the sensitivity of heads to these parameters. Townley refined the technique demonstrated by Dettinger and Wilson, and showed that results obtained using this method agree with the results Smith and Freeze (1979a and 1979b) obtained using Monte Carlo simulation.

Hoeksema and Kitanidis, in route to a solution of the inverse problem in groundwater modeling, used a slightly different method to derive the mean and covariance of the 2-dimensional steady state head distribution. They expanded both the transmissivity and head into a mean term and a zero-mean perturbation, and inserted these expansions into the continuous governing partial differential equation. Differential equations for the mean head and head perturbations were obtained by assuming there was no trend in the mean hydraulic conductivity and that second order terms (products of perturbations) could be neglected.

Hoeksema and Kitanidis's equation for the head perturbation showed a clear dependence on the mean head gradient and thus the two equations should be solved simultaneously. However, by neglecting the expected value of the products of perturbations in the mean equation, they could solve the equations sequentially -- which was equivalent to the first order approximation of the mean obtained by Dettinger and Wilson. Dettinger and Wilson also showed a dependence of the second moment equations on the solution of the mean condition. However, since the equations were discretized in the first step of their solution technique, the functional form of the dependence was not obvious.

Hoeksema and Kitanidis found the covariance of the head field by discretizing and solving the equation for the head perturbation using finite difference methods, and applying the expectation operator to this solution. Like Dettinger and Wilson, the head covariance was found to depend on the covariance of the transmissivity and the uncertainty regarding the boundary conditions. These methods provide a covariance matrix describing the correlation between the nodes on a discretized grid, rather than a continuous covariance function which the spectral method provides.

McLaughlin (1985) and Graham and McLaughlin (1989a) used a third slightly different method for obtaining the moments of the head and concentration distributions using state-space methods. Like the other methods, concentration, head, and transmissivity were assumed to be random fields, and approximate expressions were developed which relate the unknown head and concentration moments to the moments of transmissivity, and the moments of the boundary and initial conditions. The approach was similar to that of Dettinger and Wilson except that the operators in the governing partial differential equation were expanded in a Taylor series about the means of their arguments before any discretization occurs. The expectation operator was then applied to the expansion and a closed set of continuous first and second moment equations was obtained. A second order equation for the mean was used and therefore the interdependence of the mean and covariance equations was retained throughout the analysis. The importance of the covariance term in the mean equation can be significant if large variations around the calculated mean occur.

Numerical small-perturbation approximations provide a convenient way to solve derived moment problems when the underlying random fields are non-stationary or boundary conditions play an important role in the analysis. The method is better suited to solving transient problems than the spectral approach, and is usually less computationally demanding than Monte Carlo simulation. Since the method can be used in examining a wide range of problems -- including non-homogeneous, non-isotropic, random fields with spatially varying means -- conditioning of the derived moments based on field measurements is conceptually straight forward.

CHAPTER VIII  
CONCLUSIONS AND RECOMMENDATIONS

Conclusions

- 1) Geostatistical analysis of transmissivities and storativities produced the spatial correlation structures, the optimal estimates throughout the District, and the predicted error of these estimates for both variables. Maps were generated showing estimated means and standard deviations of these estimates. Transmissivities and storativities were found to be highly variable, and in areas with few measurements higher estimation standard deviations indicate greater uncertainty.
  
- 2) A comparison of transmissivity estimates obtained through kriging versus estimates obtained through regional flow modeling showed little resemblance between the estimates. However, the accuracy of the flow model estimates are uncertain since they were obtained by trial and error calibration of a regional flow model to estimated steady-state predevelopment conditions.
  
- 3) Site-specific empirical time series models were developed for 21 wells within the District. Temporal water level fluctuations were accurately modeled in 20 of the 21 wells. For these 20 wells, an analysis of model predictions showed that average prediction error standard deviations for the 1-month, 3-month, and 6-month

forecasts were 0.81, 1.36, and 1.68 feet respectively. Actual model errors fell consistently within the model's predicted 95 percent confidence intervals. These time series models should provide accurate short-term, temporal piezometric head forecasts for the particular wells modeled.

- 4) For wells showing no long-term trends of decreasing or increasing water levels, recursion analysis produced expected low water level estimates at several wells for several recurrence intervals.
- 5) If a critical water level is specified for a nontrending well, it is possible to determine the recurrence interval and the potential duration of the critical event. For several wells, excursion analysis was used to estimate the recurrence intervals and durations of groundwater excursions below levels of one and two standard deviations below mean water levels.
- 6) A regional empirical stochastic groundwater model was developed in this study that is capable of predicting transient effects of declining or increasing regional groundwater levels due to temporal fluctuations in recharge. Model parameters were fit from historical data records of piezometric head fluctuations over space and time. This model is capable of forecasting regional groundwater levels in May and September of any year, and it also estimates the accuracy of predictions. Forecasts can be made for any specified lead time; however, the accuracy of predictions improve greatly for lead times less than seven months. Comparison

of regional 1, 3, and 6-month forecasts to observed regional piezometric head levels showed that forecast errors were unbiased (i.e., symmetrically distributed around zero) with standard deviations ranging from 1.5 to 1.8 feet. Thus, this model should be useful for predicting regional water levels in May and September and provide the basis for drought management decisions.

### Recommendations

- 1) The District should include both physically-based and empirically-based stochastic groundwater models in their decision making process. Stochastic models not only produce regional predictions of the behavior of the groundwater system, but also provide an estimate of the accuracy of the prediction.
- 2) To use the empirical stochastic models developed in this study, the District must continue to intensively sample piezometric head levels throughout the District in May and September of each year, but only needs to sample 21 wells on a monthly basis. Since the model coefficients are empirical parameters fit to historic data, estimates of these parameters should be updated on an annual or biannual basis.
- 3) The times series models developed in this study should be extended to include the effects of temporally variable rainfall and pumpage. This would require conversion of the univariate time

series models to multivariate models which include rainfall and pumpage components.

- 4) A method should be developed to permit timely manipulation of pumpage rates so that target groundwater levels can be maintained at particular monitoring stations. Such a method would use the modified times series models described above in an optimal feedforward/feedback control algorithm.
  
- 5) The correlation structure derived from the geostatistical analysis of the transmissivities and storativities should be used in conditional Monte Carlo simulations or first-order perturbation stochastic models to predict best estimates of steady-state or transient hydraulic head distributions throughout the District and the variance in these estimates resulting from input parameter uncertainty. Results from these analyses could be compared to traditional deterministic models where the transmissivities and storativities are estimated by trial and error efforts to match deterministic model predictions to measured heads. Physically-based stochastic models could be used to evaluate effects of future well-field developments on steady-state and transient groundwater levels; simulations would also give the uncertainty in water level predictions associated with particular development scenarios.

- 6) Transmissivities, storativities, and leakances were found to be highly variable; thus, more measurements should be taken in regions where estimates have the greatest uncertainty.
  
- 7) Head observations should be collected simultaneously at all locations where transmissivities and/or storativities have been measured to permit the development of cross-variograms between these variables. If these variables were found to be highly correlated, the cross-variograms could be used with district-wide head measurements to improve regional transmissivity and storativity estimates throughout the District.

## REFERENCES

- Aboufirassi, M., and Marino, M. A. (1983). "Kriging of water levels in the Sauss aquifer, Morocco." Math. Geol., 15(4), 537-551.
- Ahmed, S. and De Marsily, G. (1987). "Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity." Water Resour. Res., 23(9), 1717-1737.
- Bakr, A. A., Gelhar, L. W., Gutjahr, A. L., and MacMillan, J. R. (1978). "Stochastic analysis of spatial variability in subsurface flows, 1, Comparison of one- and three-dimensional flows." Water Resour. Res., 14(2), 263-271.
- Bras, R. L., and Rodríguez-Iturbe, I. (1985). Random functions and hydrology. Addison-Wesley Publishing Company, Inc., Reading, MA.
- Bouwer, H. (1978). Groundwater hydrology. McGraw-Hill, Inc., New York, NY.
- Box, G. E. P., and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day, Inc., Oakland, CA.
- Bush, P. W., and Johnston, R. H. (1988). "Ground-water hydraulics, regional flow, and ground-water development of the Floridan aquifer system in Florida and in parts of Georgia, South Carolina, and Alabama." Regional aquifer systems analysis, U.S. Geological Survey Professional Paper 1403-C. United States Government Printing Office, Washington, DC.
- Chatfield, C. (1989). The analysis of time series: An introduction, second edition. Chapman and Hall, London, UK.
- Chirlin, G. R., and Dagan, G. (1980). "Theoretical head variogram for steady flow in statistically homogeneous aquifers." Water Resour. Res., 16(6), 1001-1015.
- Chirlin, G. R., and E. F. Wood. (1982). "On the relationship between kriging and state estimation." Water Resour. Res., 18(2), 432-438.
- Cooper R. M., and Istok, J. D. (1988a). "Geostatistics applied to groundwater contamination. I: Methodology." J. Environ. Engrg., ASCE, 114(2), 270-286.
- Cooper R. M., and Istok, J. D. (1988b). "Geostatistics applied to groundwater contamination. II: Application." J. Environ. Engrg., ASCE, 114(2), 287-299.

- Cooper R. M., and Istok, J. D. (1988c). "Geostatistics applied to groundwater pollution. III: Global estimates." J. Environ. Engrg., ASCE, 114(4), 915-928.
- Cryer, J. D. (1986). Time series analysis. PWS Publishers, Boston, MA.
- Dagan, G. (1979). "Models of groundwater flow in statistically homogeneous porous formations." Water Resour. Res., 15(1), 47-53.
- Dagan, G. (1982a). "Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 1. Conditional simulation and the direct problem." Water Resour. Res., 18(4), 813-833.
- Dagan, G. (1982b). "Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 2. The solute transport." Water Resour. Res., 18(4), 835-848.
- Dagan, G. (1985). "Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 3. The inverse problem." Water Resour. Res., 21(1), 65-72.
- David, M. (1977). Geostatistical ore reserve estimation. Elsevier, Amsterdam, The Netherlands.
- Delhomme, J. P. (1974). "La cartographie d'une grandeur physique a partir de donnees de differentes qualites." Proc. Montpellier Meet. Int. Ass. Hydrogeol., 10(1), 185-194.
- Delhomme, J. P. (1976). Applications de la theorie des variables regionalisees dans les sciences de l'eau. Ph.D. thesis, Ecole des Mines de Paris, Fontainebleau, France.
- Delhomme, J. P. (1978). "Kriging in the hydrosciences." Advances in Water Resour., 1(5), 251-266.
- Delhomme, J. P. (1979). "Spatial variability and uncertainty in groundwater flow parameters; A geostatistical approach." Water Resour. Res., 15(2), 269-280.
- Dettinger, M. D., and Wilson, J. L. (1981). "First-order analysis of uncertainty in numerical models of groundwater flow, 1, Mathematical development." Water Resour. Res., 17(1), 149-161.
- De Marsily, G. (1978). De l'identification des systemes hydrogeologiques, 1, Synthese, Ph.D. dissertation, Univ. de Pierre et Marie Curie (Paris VI), Paris.
- De Marsily, G. (1986). Quantitative hydrogeology. Academic Press, Orlando, FL.
- Dunlap, L. E., and Spinazola, J. M. (1981). "Interpolating water-table altitudes in West-Central Kansas." U.S. Geol. Surv., Open File Rep., 81-1062.

- Egbert, G. D., and Lettenmaier, D. P. (1986). "Stochastic modeling of the space-time structure of atmospheric chemical deposition." Water Resour. Res., 22(2), 165-179.
- Freeze, R. A. (1975). "A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media." Water Resour. Res., 11(5), 725-741.
- Freeze, R. A., and Cherry, J. A. (1979). Groundwater. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Gambolati, G., and Volpi, G. (1979). "Groundwater contour mapping in venice by stochastic interpolators 1: Theory." Water Resour. Res., 15(2), 281-290.
- Gelb, A. (1974). Applied Optimal Estimation, MIT Press, Cambridge, MA.
- Gelhar, L. W., Gutjahr, A. L., and Naff, R. F. (1979). "Stochastic analysis of macrodispersion in a stratified aquifer." Water Resour. Res., 15(6), 1387-1397.
- Gelhar, L. W. (1982). "Stochastic analysis of flow in heterogeneous porous media." Proc. of the NATO Advance Study Institute on Mechanic of Fluids in Porous Media, Newark, DE.
- Gelhar, L. W., and Axness, C. L. (1983). "Three-dimensional stochastic analysis of macrodispersion in aquifers." Water Resour. Res., 19(1), 161-180.
- Graham, W., and McLaughlin, D. (1989a). "Stochastic analysis of nonstationary subsurface solute transport, 1, Unconditional moments." Water Resour. Res., 25(2), 215-232, 1989a.
- Graham, W., and McLaughlin, D. (1989b). "Stochastic analysis of nonstationary subsurface solute transport, 2, Conditional moments." Water Resour. Res., 25(11), 2331-2355, 1989b.
- Gutjahr, A. L., and Gelhar, L. W. (1981). "Stochastic models of subsurface flow: Infinite versus finite domains and stationarity." Water Resour. Res., 17(2), 337-350.
- Gutjahr, A. L., Gelhar, L. W., Bakr, A. A., and MacMillan, J. R. (1978). "Stochastic analysis of spatial variability in subsurface flows: 2. Evaluation and application." Water Resour. Res., 14(5), 953-959.
- Gutjahr, A. L., and Wilson, J. (1985). "Cokriging an s stochastic differential equations." Proceedings of the International Association of Hydrologic Research, Symposium on Stochastic Approach to Subsurface Flow, Ecole des Mines de France.
- Hoeksema, R. J., and Kitanidis, P. K. (1984). "An application of the geostatistical approach to the inverse problem in two dimensional groundwater modeling." Water Resour. Res., 20(7), 1003-1020.

- Hoeksema, R. J., and Kitanidis, P. K. (1985). "Analysis of the spatial structure of properties of selected aquifers." Water Resour. Res., 21(4), 563-572.
- Houston, J. F. T. (1983). "Ground-water systems simulation by time-series techniques." Ground Water, 21(3), 301-310.
- Huff, M. D., and McKenzie-Arenberg, M. (1990). "Lower St. Johns and St. Marys Ground Water Basin Resource Availability Inventory." St. Johns River Water Management District Technical Publication SJ 90-8. Palatka, FL.
- Johnson, R. A., Frazee, J. M. Jr., and Fenzel, F. W. (1982). "Hydrogeology of the St. Johns River Water Management District." Ground Water in Florida. Proceedings of the First Annual Symposium on Florida Hydrogeology. Northwest Florida Water Management District Public Information Bulletin 82-2. 83-104.
- Journel, A. G., and Huijbregts, Ch. J. (1978). Mining geostatistics. Academic Press Limited, London, UK.
- Kitanidis, P. K., and Vomvoris, E. G. (1983). "A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations." Water Resour. Res., 19(3), 677-690.
- Marquardt, D. W. (1963). "An algorithm for least squares estimation of non-linear parameters." J. Soc. Ind. Appl. Math. 11(431).
- McKenzie-Arenberg, M. (1989). "Volusia Ground Water Basin Resource Availability Inventory." St. Johns River Water Management District Technical Publication SJ 89-4. Palatka, FL.
- McKenzie-Arenberg, M., and Toth, D. J. (1990). "Upper St. Johns Ground Water Basin Resource Availability Inventory." St. Johns River Water Management District Technical Publication SJ 90-10. Palatka, FL.
- McKenzie-Arenberg, M., and Szell, G. (1990). "Middle St. Johns Ground Water Basin Resource Availability Inventory." St. Johns River Water Management District Technical Publication SJ 90-11. Palatka, FL.
- McLaughlin, D. B. (1985). A Distributed State Space Approach for Evaluating the Accuracy of Groundwater Predictions, Ph.D. Thesis, Princeton, Univ., Princeton, NJ.
- McLaughlin, D., and Graham, W. (1986). "Design of cost-effective programs for monitoring groundwater contamination." Intl. Associ. Hydrol. Sci., Publ. 158.

- Miller, J. A. (1986). "Hydrogeologic framework of the Floridan aquifer system in Florida and in parts of Georgia, Alabama, and South Carolina." Regional aquifer systems analysis, U.S. Geological Survey Professional Paper 1403-B. United States Government Printing Office, Washington, DC.
- Mizell, S. A., Gutjahr, A. L., and Gelhar, L. W. (1982). "Stochastic analysis of spatial variability in two-dimensional steady groundwater flow assuming stationary and nonstationary fields." Water Resour. Res., 18(4), 1053-1067.
- Moore, S. F., and McLaughlin, D. B. (1980). "Mapping contaminated soil plumes by kriging." Proc., EPA National Conference on Management of Uncontrolled Hazardous Wastes, Washington, DC, 66-77.
- Neuman, S. P. (1980). "A statistical approach to the inverse problem of aquifer hydrology, 3, Improved solution method and added perspective." Water Resour. Res., 16(2), 331-346.
- Neuman, S. P., Fogg, G. E., and Jacobson, E. A. (1980). "A statistical approach to the inverse problem of aquifer hydrology, 2, Case study." Water Resour. Res., 16(1), 33-58.
- Neuman, S. P., and Jacobson, E. A. (1984). "Analysis of nonintrinsic spatial variability by residual kriging with applications to regional groundwater levels." Mathematical Geology, 16(5), 499-521.
- Neuman, S. P., and Yakowitz, S. (1979). "A statistical approach to the inverse problem of aquifer hydrology. I: Theory." Water Resources Research, 15(4), 845-859.
- Nordin, C. F., and Rosbjerg, D. M. (1970). "Applications of crossing theory in hydrology." Bulletin of the International Association of Scientific Hydrology, 15(1), 27-43.
- Philip, R. D., and Kitanidis, P. K. (1989). "Geostatistical estimation of hydraulic head gradients." Ground Water, 27(6), 855-865.
- Rodríguez-Iturbe, I., and Mejía, J. M. (1974). "The design of rainfall networks in time and space." Water Resour. Res., 10(4), 713-728.
- Rouhani, S. (1986). "Comparative study of ground water mapping techniques." J. of Ground Water, 24(2), 297-216.
- Rouhani, S., and Hall, T. J. (1988a). "Space-time kriging of groundwater data." Geostatistics, Vol. 2, M. Armstrong, ed., Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Rouhani, S., and Hall, T. J. (1988b). "Geostatistical schemes for groundwater sampling." J. Hydrol., 103, 85-102.

- Rouhani, S., Georgakakos, A. P., Kitanidis, P. K., Loaiciga, H. A., Olea, R. A., and Yates, S. R. (1990a). "Review of geostatistics in hydrology. I: Basic concepts." J. of Hydraulic Engrg., ASCE Div., 116(5), 612-632.
- Rouhani, S., Georgakakos, A. P., Kitanidis, P. K., Loaiciga, H. A., Olea, R. A., and Yates, S. R. (1990b). "Review of geostatistics in geohydrology II: Applications." J. of Hydraulic Engrg., ASCE Div., 116(5), 633-658.
- Rouhani, S., and Wackernagel, H. (1990). "Multivariate geostatistical approach to space-time data analysis." Water Resour. Res., 26(4), 585-592.
- Rubin, Y., and Dagan, G. (1987). "Stochastic identification of transmissivity and effective recharge in steady groundwater flow, parts 1 and 2." Water Resour. Res., 23(7), 1185-1200.
- Rubin, Y., and Dagan, G. (1988). "Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 1. Constant head boundary." Water Resour. Res., 24(10), 1689-1697.
- Sagar, B. (1978). "Galerkin finite element method for analyzing flow through random media." Water Resour. Res., 14(6), 1035-1044.
- SAS/ETS® user's guide, version 6, first edition. (1988). SAS Institute Inc., Cary, NC.
- Smith, L., and Freeze, R. A. (1979a). "Stochastic analysis of steady state groundwater flow in a bounded domain, 1, One-dimensional simulations." Water Resour. Res., 15(3), 521-528.
- Smith, L., and Freeze, R. A. (1979b). "Stochastic analysis of steady state groundwater flow in a bounded domain, 2, Two-dimensional simulations." Water Resour. Res., 15(6), 1543-1559.
- Smith, L., and Schwartz, F. W., (1980). "Mass transport 1, A stochastic analysis of macroscopic dispersion." Water Resour. Res. 16(2), 303-313.
- Smith, L., and Schwartz, F. W., (1981a). "Mass transport 2, Analysis of uncertainty in prediction." Water Resour. Res. 17(2), 351-369.
- Smith, L., and Schwartz, F. W., (1981b). "Mass transport 3, Role of hydraulic conductivity measurements." Water Resour. Res. 17(5), 1463-1479.
- Sophocleous, M., Paschetto, J. E., and Olea, R. A. (1982). "Groundwater network design for Northwest Kansas, using the theory of regionalized variables." Ground Water, 20(1), 48-58.
- Townley, L. R. (1983). Numerical models of groundwater flow: Prediction and parameter estimation in the presence of uncertainty. Ph.D. Thesis, MIT, Cambridge, MA.

- Vomvoris, E. G., and Gelhar, L. W. (1985). "Stochastic Analysis of concentration variability in aquifers." Proceedings of the International Association of Hydrologic Research, Symposium on Stochastic Approach to Subsurface Flow, Ecole des Mines de France.
- Wilson, J. L., Kitanidis, P. K., and Dettinger, M. D. (1978). "State and parameter estimation in groundwater models." Proceedings of AGU Chapman Conference, Pittsburgh Pa.
- Yates, S. R., Warrick, A. W., and Myers, D. E. (1986). "Disjunctive kriging II: Examples." Water Resour. Res., 22(5), 623-630.
- Yates, S. R., and Warrick, A. W. (1987). "Estimating soil water content using cokriging." Soil Science Society of America J., 51, 23-30.
- Yates, S. R., and Yates, M. V. (1990). Geostatistics for waste management: A user's manual for the GEOPACK (version 1.0) geostatistical software system. United States Environmental Protection Agency, Ada, OK.

APPENDIX A  
 LATITUDE, LONGITUDE, TRANSMISSIVITY, STORATIVITY,  
 AND LEAKANCE VALUES FOR WELLS PROVIDED BY  
 THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT

Table A.1 Values for shallow aquifer.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
S1	302300	812400	2,700	0.001	-	Duval
S2	295200	812500	6,751	0.2	-	St. Johns
S3	295800	813200	307	0.01	-	
S4	295700	813000	1,210	0.13	-	
S5	295340	812138	1,507	0.000235	-	
S6	290300	821300	3,650	-	-	Marion
S7	290900	820200	695	-	-	
S8	290800	820300	6,150	-	-	
S9	282318	815440	9,795	-	-	Lake
S10	282400	810000	1,782	-	-	Orange
S11	282400	810300	3,075	-	-	
S12	282600	810400	4,278	-	-	
S13	282600	810300	1,484	-	-	
S14	275900	810400	401	-	-	Osceola
S15	281000	810400	2,005	0.0004	-	
S16	275900	803200	1,678	-	-	Brevard
S17	280100	803600	2,003	-	-	
S18	280100	803700	2,691	-	-	
S19	280200	803500	1,845	0.000054	-	
S20	274700	803000	10,745	0.000101	0.000149	Indian River
S21	274600	803000	10,491	0.000157	0.000388	

Table A.2 Values for intermediate aquifer.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
I1	301933	823505	2,276	-	-	Columbia
I2	295700	812900	1,664	0.00012	0.000053	St. Johns
I3	295800	812900	2,288	0.00042	0.0334	
I4	295847	812353	27,485	0.00035	0.001	
I5	290900	820200	361	-	-	Marion
I6	290800	820300	21,390	-	-	
I7	290000	812000	6,484	0.00019	-	Volusia
I8	292200	810900	3,676	0.00016	-	
I9	292400	810900	2,986	0.04	-	
I10	292000	810800	622	0.00112	0.00945	
I11	292200	810900	666	-	-	
I12	292100	810900	1,670	-	-	
I13	282318	815440	11,812	-	-	Lake
I14	282510	810545	2,353	0.001	-	Orange
I15	282531	810543	4,251	0.0003	-	

Table A.3 Values for Upper Floridan aquifer.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
U1	304400	813000	18,717	-	-	Camden
U2	301933	823505	32,995	0.00007	0.000695	Columbia
U3	301000	823700	36,096	0.0008	-	Nassau
U4	303836	812742	30,000	0.00025	0.000053	
U5	300800	813800	107,955	0.00115	-	Duval
U6	300800	813900	19,953	0.000286	-	
U7	295144	813771	7,799	-	-	Clay
U8	300656	814634	87,005	-	-	
U9	300100	814400	30,882	0.0002	-	
U10	293950	812842	87,968	0.0006	0.0187	St. Johns
U11	294053	812952	56,016	0.0006	0.00508	
U12	294752	812905	25,000	0.001	0.00110	
U13	295730	812930	54,011	0.0002	0.00495	
U14	294343	812840	29,011	0.0003	0.000963	
U15	295028	813309	8,703	-	-	
U16	295132	811648	13,035	-	-	
U17	300354	813012	6,818	-	-	

Table A.3--continued.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
U18	300048	812333	1,604	-	-	
U19	294612	812534	15,000	-	-	
U20	294700	812900	23,128	0.00016	0.000201	
U21	294748	812923	38,770	0.00059	-	
U22	295826	812339	22,006	0.00018	1.3x10 <sup>-7</sup>	
U23	295800	813200	42,781	-	-	
U24	295700	813000	85,695	0.0012	-	
U25	295800	811800	17,981	0.000235	-	
U26	294200	821600	2,139	-	-	Alachua
U27	293933	813428	45,989	0.001	0.0160	Putnam
U28	294032	813455	23,997	0.0008	0.0100	
U29	294255	813240	57,513	0.001	0.0214	
U30	294540	813833	17,005	-	-	
U31	293234	814241	41,043	-	-	
U32	294500	814300	36,765	0.00094	0.000241	
U33	294025	813358	36,765	-	-	
U34	293930	813436	48,128	-	-	
U35	293955	813445	36,765	-	-	
U36	294300	820100	106,283	-	0.000829	
U37	292300	813700	16,811	0.000345	0.0112	
U38	292700	812157	36,096	0.00047	0.000695	Flagler
U39	291955	812009	37,433	0.0009	-	
U40	291955	811951	25,401	0.00019	0.00227	
U41	291915	811840	36,765	-	-	
U42	293716	812936	25,936	-	-	
U43	293036	811714	9,398	0.0001	0.000896	
U44	293325	811248	55,481	0.0004	0.000755	
U45	292616	811314	22,995	-	-	
U46	292947	811743	9,265	0.00042	0.000134	
U47	292448	811213	12,005	0.00036	0.000254	
U48	292500	811200	6,610	0.000016	-	
U49	290900	820200	16,043	-	-	Marion
U50	290800	820300	29,412	-	-	
U51	290500	821700	62,166	0.0062	0.00227	
U52	285900	821200	66,845	0.0005	-	
U53	290600	820000	33,422	-	-	
U54	290500	820000	66,845	-	-	
U55	290900	810600	40,775	0.000735	-	Volusia
U56	291200	810200	3,743	0.00023	-	
U57	291100	810400	45,900	0.000145	-	
U58	291100	810300	32,754	0.00022	-	
U59	285900	811700	25,401	-	-	
U60	285900	805500	6,952	0.00029	-	
U61	291948	812855	28,509	0.001	-	
U62	291725	812756	24,499	0.00055	-	
U63	291507	812906	26,003	0.00045	-	

Table A.3--continued.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
U64	291439	812817	22,025	0.0006	-	
U65	291541	812546	8,396	0.00045	-	
U66	290927	812128	23,997	0.00095	-	
U67	290847	812030	6,698	0.0003	-	
U68	290535	812148	88,516	0.0007	-	
U69	291929	812840	8,904	-	-	
U70	291802	812741	15,000	-	-	
U71	291506	812857	21,999	-	-	
U72	291433	812852	16,043	-	-	
U73	291440	812828	42,005	-	-	
U74	291422	812547	4,505	-	-	
U75	291055	812850	18,048	-	-	
U76	290850	812021	7,794	-	-	
U77	290635	812027	160,027	-	-	
U78	290532	812135	37,032	-	-	
U79	291004	811014	12,032	-	-	
U80	292200	810900	1,604	0.00015	0.00134	
U81	292300	810900	4,586	-	-	
U82	290600	810800	5,548	0.000205	-	
U83	290500	810800	4,445	0.00012	-	
U84	285700	805800	10,160	0.000497	0.000127	
U85	291600	810900	41,979	-	-	
U86	291500	810900	21,658	-	-	
U87	292100	810800	1,883	0.000156	0.0401	
U88	285600	811800	5,348	-	-	
U89	282318	815440	12,968	0.00025	0.00201	Lake
U90	282800	815400	39,171	0.013	0.00481	
U91	282200	813800	3,476	-	-	
U92	282200	814900	4,278	-	-	
U93	282700	815800	7,620	-	-	
U94	283200	815400	3,743	-	-	
U95	284900	815500	20,098	0.0119	0.00221	
U96	283800	811400	33,824	-	-	Seminole
U97	284100	811000	1,217	-	-	
U98	284100	811300	13,102	-	-	
U99	284200	811000	29,813	-	-	
U100	284300	811800	42,112	-	-	
U101	284500	811300	9,581	0.000004	-	
U102	284700	811200	25,802	-	-	
U103	284900	811800	17,914	-	-	
U104	284428	810726	4,104	-	-	
U105	284550	810715	16,979	-	-	
U106	284706	810708	1,698	-	-	
U107	284712	810443	3,703	-	-	
U108	284800	811900	13,492	-	-	
U109	284600	812300	152,406	0.0035	0.0100	

Table A.3--continued.

Well Number	Latitude	Longitude	T (ft <sup>2</sup> /d)	S	L (d <sup>-1</sup> )	County
U110	283100	812100	79,713	0.001367	0.0188	Orange
U111	283600	812800	36,230	0.000935	0.0263	
U112	282552	810756	46,791	0.00007	0.000267	
U113	282531	810822	73,529	0.03	0.00134	
U114	282531	810957	72,861	0.00063	0.00535	
U115	283200	812000	574,866	-	0.000012	
U116	282412	810447	508,021	-	-	
U117	282341	810401	550,000	0.0009	-	
U118	282530	810542	209,893	-	-	
U119	282510	810545	758,021	0.00007	-	
U120	281000	814400	14,706	-	-	Polk
U121	281300	814900	5,348	-	-	
U122	281400	813900	84,893	0.012	-	
U123	281400	814000	90,909	0.0018	-	
U124	281400	814300	10,294	-	-	
U125	281400	813400	4,947	-	-	
U126	281500	814900	3,877	-	-	
U127	275901	811215	6,003	-	-	Osceola
U128	280905	812701	6,003	-	-	
U129	281037	810751	18,984	-	-	
U130	281150	810241	2,005	-	-	
U131	281159	811428	37,968	-	-	
U132	281632	805150	6,992	-	-	
U133	281714	810930	83,021	-	-	
U134	282729	811340	25,000	-	-	
U135	281820	805405	2,995	-	-	
U136	281919	805333	4,011	-	-	
U137	281955	813707	6,000	-	-	
U138	281600	813500	1,604	0.011	-	
U139	280100	814400	170,242	-	-	Polk
U140	284700	805100	40,107	0.0008	-	Brevard
U141	275119	804824	10,000	-	-	
U142	275725	804127	5,000	-	-	
U143	275738	805210	8,008	-	-	
U144	275831	805135	8,008	-	-	
U145	280658	804651	9,000	-	-	
U146	280746	805016	5,000	-	-	
U147	280811	805144	8,008	-	-	
U148	280947	805134	2,995	-	-	
U149	275900	803200	24,612	0.00136	-	
U150	274700	802600	19,831	0.001267	-	Indian River
U151	274400	802400	6,150	0.00525	-	
U152	274500	802400	26,719	0.000228	-	

APPENDIX B  
ANNUAL MEAN, STANDARD DEVIATION, MAXIMUM AND MINIMUM  
PIEZOMETRIC HEAD LEVELS FOR 21 WELLS WITHIN  
THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT

Well Number	SJRWMD Well Number	Years of Data	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
1	N-0003	1977-1990	157	-13.68	6.03	7.22	-30.64
2	D-0348	1976-1990	158	45.68	1.89	50.05	41.23
3	D-0160	1945-1990	529	37.90	5.95	52.34	24.65
4	U-0001	1959-1982	284	59.80	3.17	66.51	53.41
5	C-0120	1974-1989	172	83.07	1.63	87.02	79.53
6	P-0172	1976-1990	163	17.28	3.23	21.33	2.84
7	SJ-0104	1959-1989	368	17.06	1.99	22.20	11.68
8	F-0087	1937-1990	643	15.21	1.60	18.21	10.64
9	M-0048	1933-1989	681	48.56	2.21	55.33	43.41
10	M-0013	1975-1990	175	53.04	1.82	57.63	49.12
11	SU-0013	1973-1990	198	29.67	1.12	31.77	25.27
12	V-0101	1951-1990	469	42.29	1.61	46.12	38.55
13	S-0125	1953-1990	444	44.67	3.48	55.19	35.15
14	L-0062	1959-1990	371	100.05	0.95	102.01	97.03
15	OR-0047	1943-1990	557	63.12	5.13	76.88	50.42
16	OR-0007	1961-1990	347	36.32	1.87	40.72	30.78
17	SU-0002	1959-1990	370	91.00	1.60	93.61	85.58
18	OR-0064	1959-1990	370	108.72	1.32	112.17	104.87
19	BR-0202	1955-1989	406	28.42	2.24	33.95	23.01
20	OS-0001	1976-1990	169	43.75	1.53	46.64	38.83
21	PO-0006	1960-1990	359	127.04	1.59	131.16	121.59

APPENDIX C  
 MONTHLY MEAN, STANDARD DEVIATION, MAXIMUM AND MINIMUM  
 PIEZOMETRIC HEAD LEVELS FOR 21 WELLS WITHIN  
 THE ST. JOHNS RIVER WATER MANAGEMENT DISTRICT

Table C.1 Monthly sample statistics for Well 1 (SJRWMD Well N-0003).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	13	-11.72	6.19	-1.80	-23.90
February	13	-14.18	5.12	-8.80	-24.40
March	13	-14.12	5.06	-7.90	-23.10
April	13	-12.45	6.18	-6.30	-22.90
May	13	-15.12	5.92	-3.70	-24.30
June	13	-16.28	4.34	-9.50	-23.80
July	13	-14.15	5.25	-3.00	-21.90
August	13	-15.71	9.19	7.20	-30.60
September	13	-12.83	8.09	6.70	-28.90
October	13	-13.61	4.56	-4.70	-22.20
November	13	-13.69	4.28	-6.20	-21.90
December	13	-9.70	5.73	-1.00	-20.80

Table C.2 Monthly sample statistics for Well 2 (SJRWMD Well D-0348).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	13	44.96	1.37	47.22	43.26
February	13	44.67	1.60	47.54	42.66
March	13	44.33	1.80	47.78	41.97
April	13	44.48	2.05	48.15	41.23
May	13	45.35	2.06	49.03	41.89
June	13	46.30	2.04	49.97	42.61
July	13	46.86	1.92	50.04	43.54
August	13	46.82	1.80	49.91	43.95
September	13	46.45	1.57	49.17	44.21
October	13	46.16	1.56	48.44	43.88
November	14	45.83	1.45	48.04	43.68
December	14	45.56	1.43	48.26	43.64

Table C.3 Monthly sample statistics for Well 3 (SJRWMD Well D-0160).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	44	38.95	5.64	51.10	28.80
February	44	38.93	5.58	51.20	26.50
March	44	38.50	5.66	51.40	26.20
April	44	37.25	6.01	50.70	26.20
May	44	36.46	6.76	51.80	25.10
June	44	36.48	6.49	52.30	24.60
July	44	36.34	6.23	51.30	25.00
August	44	36.86	5.93	49.80	26.40
September	44	37.88	5.68	50.10	27.60
October	44	38.62	5.70	51.00	27.20
November	44	39.02	5.53	51.60	29.80
December	44	38.97	5.56	51.40	29.60

Table C.4 Monthly sample statistics for Well 4 (SJRWMD Well U-0001).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	23	59.49	2.97	65.45	54.39
February	24	60.05	2.94	65.26	54.99
March	24	60.57	3.19	66.45	55.16
April	24	60.68	3.29	66.27	56.07
May	24	60.11	3.16	65.52	55.23
June	24	59.52	3.14	65.84	54.21
July	24	59.31	3.23	65.35	53.60
August	24	59.54	3.31	65.03	53.41
September	24	59.78	3.24	64.88	53.76
October	23	59.80	3.39	65.64	53.63
November	23	59.43	3.27	65.02	53.77
December	23	59.26	3.07	64.72	53.89

Table C.5 Monthly sample statistics for Well 5 (SJRWMD Well C-0120).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	14	82.99	1.49	86.13	80.66
February	14	83.16	1.38	86.30	80.99
March	14	83.38	1.41	86.57	81.08
April	14	83.47	1.68	87.03	81.31
May	14	83.09	1.81	86.61	80.60
June	15	82.78	1.83	85.98	79.93
July	15	82.66	1.87	85.75	79.76
August	15	82.81	1.90	85.75	79.53
September	15	83.11	1.74	85.80	79.64
October	14	83.33	1.57	85.87	80.34
November	14	83.11	1.60	85.71	80.00
December	14	82.96	1.56	85.67	80.06

Table C.6 Monthly sample statistics for Well 6 (SJRWMD Well P-0172).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	14	18.04	4.52	21.13	2.84
February	14	17.86	4.56	20.95	3.18
March	14	15.99	4.02	21.33	5.57
April	14	13.70	2.84	19.72	9.99
May	14	14.62	1.64	17.33	11.75
June	14	16.92	1.36	18.82	14.50
July	13	17.86	1.20	19.58	15.71
August	13	18.46	1.05	19.86	16.52
September	13	18.93	0.96	20.19	17.57
October	13	18.75	1.32	20.57	16.87
November	13	18.91	1.18	20.97	17.13
December	13	17.88	4.57	20.97	3.01

Table C.7 Monthly sample statistics for Well 7 (SJRWMD Well SJ-0104).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	30	17.78	1.52	20.97	15.05
February	31	17.77	1.66	20.93	15.13
March	31	17.26	2.01	20.91	13.51
April	31	15.94	2.30	20.82	12.71
May	31	15.36	2.19	20.74	11.68
June	31	15.99	1.85	20.60	12.45
July	31	16.57	1.78	20.67	12.99
August	31	17.11	1.73	21.18	13.63
September	31	17.64	1.69	21.75	14.71
October	30	17.96	1.76	22.20	14.87
November	30	17.90	1.72	21.83	15.11
December	30	17.74	1.58	21.21	15.04

Table C.8 Monthly sample statistics for Well 8 (SJRWMD Well F-0087).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	53	15.51	1.33	17.90	12.70
February	54	15.52	1.37	17.70	12.70
March	54	15.28	1.46	17.70	12.60
April	54	14.64	1.71	17.50	11.30
May	54	14.20	1.80	17.20	10.90
June	54	14.45	1.73	17.40	10.60
July	54	14.84	1.70	17.60	10.70
August	54	15.17	1.61	17.80	11.90
September	54	15.55	1.51	18.20	12.60
October	53	15.66	1.45	18.00	12.60
November	53	15.62	1.38	17.90	12.70
December	53	15.55	1.33	17.90	12.80

Table C.9 Monthly sample statistics for Well 9 (SJRWMD Well M-0048).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	56	48.49	2.14	52.93	44.08
February	57	48.26	1.99	52.44	43.96
March	57	48.28	1.98	52.12	43.59
April	57	48.38	2.16	53.67	43.47
May	57	48.16	2.06	53.22	43.41
June	57	47.92	1.97	52.49	43.72
July	57	48.07	2.02	52.66	43.88
August	57	48.52	2.18	54.20	44.12
September	57	49.17	2.35	54.77	44.27
October	57	49.42	2.54	55.33	44.53
November	56	49.25	2.43	54.71	44.54
December	56	48.82	2.27	53.78	44.24

Table C.10 Monthly sample statistics for Well 10 (SJRWMD Well M-0013).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	14	52.75	1.71	55.76	49.12
February	14	52.89	1.77	56.31	49.20
March	14	53.02	1.61	55.69	50.19
April	14	53.16	1.69	56.05	50.20
May	14	52.91	1.70	55.82	50.35
June	15	52.89	1.60	55.60	50.04
July	15	53.09	1.99	56.74	49.79
August	15	53.19	2.02	57.05	50.17
September	15	53.41	1.95	57.19	50.25
October	15	53.48	2.04	57.60	49.80
November	15	53.12	1.94	56.91	49.56
December	15	52.84	1.90	56.63	49.17

Table C.11 Monthly sample statistics for Well 11 (SJRWMD Well SU-0013).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	39	29.93	0.81	31.35	28.15
February	40	30.01	0.87	31.30	28.27
March	39	29.92	0.95	31.42	27.68
April	39	29.49	1.11	31.34	27.26
May	39	28.75	1.21	30.79	26.36
June	39	28.72	1.20	31.17	25.64
July	39	29.34	1.23	31.39	25.27
August	39	29.76	1.07	31.47	26.52
September	39	30.12	0.82	31.77	28.21
October	39	30.21	0.89	31.61	27.58
November	39	30.03	0.86	31.29	27.55
December	39	29.91	0.84	31.35	27.27

**Table C.12 Monthly sample statistics for Well 12 (SJRWMD Well V-0101).**

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	17	42.19	1.17	44.28	39.66
February	17	42.51	1.29	44.40	40.18
March	16	42.89	1.57	45.20	39.75
April	16	42.48	1.56	44.97	39.43
May	16	41.82	1.49	44.12	38.83
June	16	41.74	1.50	44.17	39.18
July	16	42.29	1.91	45.57	38.76
August	16	42.49	1.92	45.58	38.66
September	17	42.78	1.88	45.10	38.93
October	17	42.46	1.84	46.12	38.55
November	17	41.79	1.56	44.53	38.62
December	17	41.79	1.39	43.76	38.79

**Table C.13 Monthly sample statistics for Well 13 (SJRWMD Well S-0125).**

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	37	44.99	3.03	50.96	39.01
February	37	44.96	2.88	50.64	38.51
March	37	44.75	3.22	51.71	37.75
April	37	44.03	3.62	52.49	36.46
May	37	43.17	3.62	50.59	36.14
June	37	43.46	3.49	50.10	35.15
July	37	44.22	3.52	52.12	35.78
August	37	44.94	3.58	53.60	36.55
September	37	45.53	3.58	54.79	38.72
October	37	45.56	3.78	55.19	38.73
November	37	45.25	3.43	53.24	38.87
December	37	44.99	3.19	51.82	38.95

Table C.14 Monthly sample statistics for Well 14 (SJRWMD Well L-0062).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	30	100.01	0.89	101.26	97.70
February	31	100.27	0.78	101.51	98.70
March	31	100.36	0.86	101.65	98.43
April	31	99.99	1.02	101.72	97.62
May	31	99.39	1.14	100.91	97.03
June	31	99.61	0.92	101.21	97.86
July	31	100.13	0.75	101.56	98.42
August	31	100.45	0.73	101.98	99.04
September	31	100.66	0.85	101.99	98.87
October	31	100.23	0.94	101.67	98.16
November	31	99.92	0.88	101.23	97.89
December	31	99.84	0.90	101.19	97.31

Table C.15 Monthly sample statistics for Well 15 (SJRWMD Well OR-0047).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	46	63.57	4.90	73.61	53.41
February	46	63.35	4.80	73.27	53.58
March	46	63.25	5.18	76.53	53.59
April	46	61.43	5.40	77.44	52.09
May	46	61.28	5.23	73.51	50.42
June	46	61.50	4.94	72.39	50.60
July	46	62.67	5.09	76.32	50.87
August	47	63.80	5.12	76.76	51.53
September	47	64.73	5.35	76.88	53.29
October	47	64.80	5.66	76.41	52.66
November	47	64.11	5.26	75.95	53.20
December	47	63.70	5.16	75.17	52.68

Table C.16 Monthly sample statistics for Well 16 (SJRWMD Well OR-0007).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	28	36.76	1.46	39.64	33.75
February	29	36.84	1.79	40.72	34.22
March	29	36.69	1.89	39.91	33.37
April	29	35.66	1.97	39.31	32.40
May	29	34.41	1.70	37.86	30.78
June	29	34.96	1.61	38.40	31.47
July	29	35.90	1.66	39.16	31.76
August	29	36.53	1.56	39.67	32.76
September	29	37.08	1.46	40.08	34.47
October	29	37.26	1.67	40.34	34.29
November	29	37.06	1.63	39.81	33.48
December	29	36.79	1.46	39.51	33.78

Table C.17 Monthly sample statistics for Well 17 (SJRWMD Well SU-0002).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	31	91.18	1.53	93.44	87.04
February	30	91.64	1.43	93.50	87.00
March	30	91.66	1.41	93.61	87.33
April	31	90.79	1.41	92.93	87.28
May	31	89.61	1.75	92.23	85.58
June	31	90.00	1.95	93.11	85.69
July	31	91.03	1.39	92.82	87.36
August	31	91.87	1.24	93.61	87.92
September	31	91.95	1.17	93.41	88.23
October	31	91.17	1.25	93.13	87.97
November	31	90.61	1.29	92.84	87.69
December	31	90.64	1.41	92.99	87.46

Table C.18 Monthly sample statistics for Well 18 (SJRWMD Well OR-0064).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	30	108.68	1.15	110.78	106.60
February	30	108.73	1.18	110.85	106.22
March	31	108.74	1.36	111.67	106.18
April	31	108.45	1.61	111.99	105.58
May	31	108.09	1.50	111.41	104.87
June	31	108.28	1.29	111.29	105.22
July	31	108.69	1.30	111.58	105.64
August	31	108.99	1.26	111.57	105.64
September	31	109.26	1.15	112.17	107.23
October	31	109.12	1.25	111.99	106.69
November	31	108.87	1.22	111.38	106.78
December	31	108.68	1.20	111.03	106.45

Table C.19 Monthly sample statistics for Well 19 (SJRWMD Well BR-0202).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	34	28.88	1.85	32.38	25.65
February	34	28.81	2.01	32.43	25.73
March	34	28.77	2.06	32.71	25.56
April	34	27.98	2.34	32.77	24.67
May	34	26.94	2.43	31.83	23.02
June	34	27.18	2.23	32.03	23.25
July	34	27.99	2.23	32.62	23.68
August	34	28.49	2.16	33.42	24.32
September	33	29.06	2.11	33.97	25.44
October	33	29.42	2.11	34.07	25.99
November	34	29.35	2.15	33.66	25.19
December	34	29.01	1.96	33.08	25.86

Table C.20 Monthly sample statistics for Well 20 (SJRWMD Well OS-0001).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	15	44.33	1.13	45.88	42.11
February	14	44.32	1.31	45.97	42.37
March	14	44.27	1.49	46.64	41.63
April	14	43.22	1.65	46.63	40.70
May	14	41.94	1.47	44.29	38.83
June	14	42.28	1.36	44.00	39.83
July	14	43.14	1.43	45.16	40.17
August	14	43.78	1.26	45.58	41.19
September	14	44.46	0.99	45.65	42.80
October	14	44.54	1.18	46.21	42.16
November	14	44.37	1.21	45.84	41.63
December	14	44.37	1.11	45.68	42.30

Table C.21 Monthly sample statistics for Well 21 (SJRWMD Well PO-0006).

Month	Number of Obs.	Mean (ft)	Std Dev (ft)	Maximum (ft)	Minimum (ft)
January	29	127.22	1.21	129.55	124.58
February	30	127.38	1.33	129.49	123.74
March	30	127.17	1.52	129.79	123.75
April	30	126.01	2.03	129.83	121.59
May	30	125.33	1.74	129.19	121.84
June	30	126.32	1.46	129.68	124.06
July	30	127.35	1.45	131.03	124.80
August	30	127.87	1.12	131.15	126.16
September	30	128.18	1.05	131.45	125.81
October	30	127.62	1.38	131.31	124.21
November	30	127.19	1.32	130.57	125.05
December	30	127.14	1.21	129.74	124.66