



# **TECHNICAL UNIVERSITY OF CRETE**

**Department of Mineral Resources Engineering**

**Division of Exploration and Positioning**

**Geostatistics Research Unit (GRU)**

**“Geostatistical Analysis and Space-Time Models  
of Aquifer Levels: Application to Mires  
Hydrological Basin in the Prefecture of Crete”**

**“Γεωστατιστική Ανάλυση και Χωροχρονικά Μοντέλα Στάθμης  
Υδροφορέα: Εφαρμογή στην Υδρολογική Λεκάνη των Μοιρών  
της Περιφέρειας Κρήτης”**

PhD Thesis

**EMMANOUIL A. VAROUCHAKIS**  
Chemical & Process Engineering (B.Sc.)  
Clean Technology (M.Sc.)  
Geotechnology & Environment (M.Sc.)

**SUPERVISOR:**

Professor:  
D.T. HRISTOPULOS

CHANIA 2012





**TECHNICAL UNIVERSITY OF CRETE**

**Department of Mineral Resources Engineering**

**“Geostatistical Analysis and Space-Time Models of  
Aquifer Levels: Application to Mires Hydrological  
Basin in the Prefecture of Crete”**

**“Γεωστατιστική Ανάλυση και Χωροχρονικά Μοντέλα Στάθμης  
Υδροφορέα: Εφαρμογή στην Υδρολογική Λεκάνη των Μοιρών της  
Περιφέρειας Κρήτης”**

PhD Thesis

**EMMANOUIL A. VAROUCHAKIS**  
Chemical & Process Engineering (B.Sc.)  
Clean Technology (M.Sc.)  
Geotechnology & Environment (M.Sc.)



“A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Geotechnology & Environment, Department of Mineral Resource Engineering, Technical University of Crete”

**Examination Committee:**

**Professor Dionissios T. Hristopulos** (Technical University of Crete): Supervisor and President of the examination committee

**Professor George P. Karatzas** (Technical University of Crete): Member of the advisory committee

**Professor Evangelos Paleologos** (Technical University of Crete): Member of the advisory committee

**Professor Zacharias Agioutantis** (Technical University of Crete): Member of the examination committee

**Professor Ioannis Tsanis** (Technical University of Crete): Member of the examination committee

**Assistant Professor Nikolaos Mamassis** (National Technical University): Member of the examination committee

**Assistant Professor Nikolaos Theodosiou** (Aristotle University of Thessaloniki): Member of the examination committee



## Forward

My first contact with geostatistics was in the summer of 2003 when I first met Professor Dionissis Hristopulos. Since then geostatistics became part of my job and of my scientific interests. The preparation of a PhD thesis that could provide novel and useful geostatistical tools for groundwater hydrology was an interesting challenge. There are many studies and publications regarding geostatistical research and its applications to hydrology during the last thirty years. However, there is always room for new applications, tools and methodologies. This PhD thesis aims to contribute to the geostatistical and hydrological sciences with some new ideas for effective data analysis and prediction.

I am grateful to Professor Dionissis Hristopulos for the guidance, understanding and help through the years of our collaboration in order for this PhD thesis to be fulfilled. I am also thankful to Professor George Karatzas for his guidance, advice and support. I would also like to thank Professor Evangelos Palaiologos for the useful advices and suggestions as well as the members of the examination committee for the valuable comments on the thesis topics.

Many thanks go to the members of the Geoenvironmental lab of the Department of Environmental Engineering and the members of the Geostatistics research unit of the Mineral Resources Department of the Technical University of Crete for their support and their friendship all these years.

I owe a special thank you to Ilias Vardavas, Professor of the Physics Department at the University of Crete, who first gave me the opportunity to work in the topic of groundwater hydrology at the Environmental Technology lab of the Foundation for Research and Technology in Heraklion. I would also like to thank the staff members of the Administration of Land Reclamation and the Department of Water Resources Management of the Prefecture of Crete for the Mesara valley data provided.

This PhD thesis is dedicated to my lovely wife Katerina who understands and encourages me all these years, my parents Andreas and Ioanna for their love, endless support and encouragement and last but not least to my brother who was always there for me.

Emmanouil A. Varouchakis



## Table of contents

Forward.....	I
Table of contents.....	III
List of Figures.....	VII
List of Tables.....	XV
Abstract.....	XIX
Περίληψη.....	XXIII
<b>1. Introduction.....</b>	<b>1</b>
1.1 Motivations for this research.....	1
1.2 Objectives.....	2
1.3 Innovation.....	3
1.4 Outline of the thesis.....	4
<b>2. Background and theory.....</b>	<b>5</b>
2.1 Mathematical background.....	6
2.1.1 <i>Random fields</i> .....	7
2.1.2 <i>Basic concepts in random fields</i> .....	8
2.1.3 <i>Mean value</i> .....	9
2.1.4 <i>Variance</i> .....	11
2.1.5 <i>Covariance function</i> .....	11
2.1.6 <i>Statistical homogeneity</i> .....	13
2.1.7 <i>Statistical isotropy</i> .....	13
2.1.8 <i>Spatial dependence</i> .....	14
2.1.9 <i>Semivariogram estimation</i> .....	17
2.1.10 <i>Semivariogram models</i> .....	21
2.1.11 <i>Matérn model</i> .....	22
2.1.12 <i>Spartan model</i> .....	22
2.1.13 <i>Parameters inference</i> .....	26
2.1.14 <i>Spatial estimation</i> .....	27
2.2 Spatial interpolation.....	28
2.3 Overview of geostatistical methodology.....	33
2.4 Interpolation materials and methods.....	34
2.4.1 <i>Inverse Distance Weight</i> .....	35

2.4.2 <i>Minimum Curvature</i> .....	35
2.4.3 <i>Ordinary kriging interpolation</i> .....	36
2.4.4 <i>Universal kriging interpolation</i> .....	38
2.4.5 <i>Kriging with Delaunay triangulation</i> .....	39
2.4.6 <i>Residual Kriging</i> .....	40
2.5 <b>Spatial Model Validation</b> .....	42
2.6 <b>Comparison of interpolation methods</b> .....	43
2.7 <b>Spatiotemporal interpolation</b> .....	44
<b>3. Study area and data exploratory statistical analysis</b> .....	49
3.1 <b>Physical setting and information</b> .....	49
3.2 <b>Hydrogeological setting</b> .....	51
3.3 <b>Hydrological setting</b> .....	55
3.4 <b>Groundwater level Data availability</b> .....	58
3.5 <b>Exploratory statistics</b> .....	60
3.6 <b>Anisotropy estimation</b> .....	63
<b>4. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins-Application to Mires basin</b> .....	65
4.1 <b>Introduction</b> .....	65
4.2 <b>Semivariogram estimation</b> .....	65
4.3 <b>Results and Discussion</b> .....	66
4.3.1 <i>Global cross validation measures</i> .....	66
4.3.2 <i>Lowest-value estimation</i> .....	70
4.3.3 <i>Isolevel contour maps of hydraulic head</i> .....	72
4.3.4 <i>Estimation variance</i> .....	77
4.3.5 <i>Semivariogram validation</i> .....	78
4.3.6 <i>General remarks</i> .....	80
<b>5. Improving kriging of groundwater level data using non-linear normalizing transformations-Application to Mires basin</b> .....	81
5.1 <b>Introduction</b> .....	81
5.2 <b>Box-Cox transformation method</b> .....	82
5.3 <b>Trans-Gaussian Kriging (TGK)</b> .....	83
5.4 <b>Gaussian Anamorphosis (GA)</b> .....	84
5.5 <b>Modified Box-Cox (MBC)</b> .....	85

5.6 Results and discussion .....	86
<b>6. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables .....</b>	<b>101</b>
6.1 Introduction.....	101
6.2 Normalizing transformations .....	105
6.3 Trend Modeling of Hydraulic Head in Mires Basin .....	105
6.3.1 Topographic variables component .....	106
6.3.2 Multiple-well hydraulic head component .....	108
6.4 Interpolation of Hydraulic Head in Mires Basin: Models and Results .....	112
6.4.1 Exploratory statistics .....	112
6.4.2 Geostatistical head models .....	113
6.4.3 Head models without trend .....	113
6.4.4 Head models with trend .....	115
6.4.5 Mapping of Groundwater Level for the Optimal Model .....	120
6.5 Discussion .....	122
<b>7. Stochastic space-time modeling of groundwater level variations in a Mediterranean basin.....</b>	<b>127</b>
7.1 Introduction.....	127
7.2 Purely temporal variation analysis .....	128
7.2.1 Background of Regionalized ARX model of groundwater level.....	129
7.2.2 ARX Model for groundwater level in Mires basin .....	130
7.2.3 Kalman filter identification of ARX model.....	132
7.2.4 Kalman filter adaptation algorithm of ARX model .....	134
7.3 Spatiotemporal geostatistics.....	137
7.3.1 Spatiotemporal two point function.....	138
7.3.2 Spatiotemporal covariance or semivariogram models .....	142
7.3.3 Spatiotemporal models' summary of characteristics.....	147
7.3.4 A Spatiotemporal covariance function derived from a physical law .....	147
7.3.5 Spatiotemporal geostatistical analysis and prediction .....	148
7.3.6 Spatiotemporal prediction of groundwater level data in Mires basin .....	151
7.3.6.1 Separable product space-time semivariogram using Matérn model .....	154
7.3.6.2 Non-separable space-time semivariogram .....	155
7.4 Results.....	155
7.4.1 Purely temporal analysis .....	155

7.4.2 <i>Spatiotemporal analysis</i> .....	159
7.5 Discussion.....	166
<b>8. Conclusions and future work</b> .....	169
8.1 Conclusions.....	169
8.2 Future work and perspectives .....	176
<b>Publications in international journals and conferences</b> .....	179
References.....	181

## List of Figures

Figure 2.1 Presentation of the main axis system (KA1, KA2) in relation to the coordinate system $x, y$ . The ellipsis corresponds to the semivariogram direction (after (Hristopulos 2008)).	16
Figure 2.2 Schematic figure of the region $B(\mathbf{r})$ around the distance vector (Hristopulos 2008).	18
Figure 2.3 Presentation of typical semivariogram characteristics.	20
Figure 2.4 Delaunay Triangulation of monitoring sites in Mires Basin. The vertices of the enclosing triangle (dark color) that contains the estimation point $s_0$ are the first-order neighbors of $s_0$ ; the vertices of the three adjacent triangles (grey color) that do not belong to the enclosing triangle provide the second-order neighbors of $s_0$ .	40
Figure 3.1 Map of Greece and the schematic representation of the Mesara valley and the Mires basin locations on the island of Crete.	50
Figure 3.2 Geological representation of Mires basin.	51
Figure 3.3 E-W cross-section along the Geropotamos river showing simplified geologic formation of the Mire basin (Fytolakis <i>et al.</i> 2005).	53
Figure 3.4 Simplified geological structure of the Mires basin, modified after (Vardavas <i>et al.</i> 1996).	53
Figure 3.5 Hydro-geological representation of Mires basin.	54
Figure 3.6 Annual rainfall in Mesara valley (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).	56
Figure 3.7 Monthly rainfall in Mesara valley (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).	56
Figure 3.8 Annual runoff of Geropotamos of Mesara catchment (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).	57

Figure 3.9 Annual abstraction rate in Mires basin (data provided by the Administration of Land Reclamation of the Prefecture of Crete, the Department of Water Resources Management of the Prefecture of Crete and by Kritsotakis (2010). 58	
Figure 3.10 Topographic map showing the locations of groundwater head measurement in Mires basin along with the corresponding surface elevation and the temporary river (Geropotamos) path. ....	59
Figure 3.11 Mean bi-annual groundwater level in Mires basin. An average of 10 wells until 2002-03 is considered, 6 wells until 2005-06 and 2 wells until 2009-10.....	59
Figure 3.12 Topographic map showing the locations of monitored wells (triangles) in Mires basin along with the corresponding surface elevation and the temporary river path. Green color is used to mark the wells monitored for the period 1981-2003. ....	60
Figure 3.13 Normal probability plot of the basin's groundwater level data for the hydrological year 2002-2003.....	62
Figure 3.14 Experimental directional semivariograms of groundwater level in Mires basin along the four main geographical directions, E-W, N-S, NE and NW.....	63
Figure 4.1 Plot of omnidirectional experimental semivariogram of groundwater level data (stars), the optimal Spartan model (parameter estimates: variance $\hat{\sigma}^2=184$ , characteristic length $\hat{\xi}=0.46$ (normalized units), stiffness coefficient $\hat{\eta}_1=1.12$ ), and the optimal power-law model ( $\hat{c}=538$ , $2\hat{H}=1.31$ ). Numbers of pairs used at each lag distance are also shown on the plot.....	68
Figure 4.2 Plot of omnidirectional semivariogram of residuals (stars) and optimal Spartan model (parameter estimates: variance $\hat{\sigma}^2=169$ , characteristic length $\hat{\xi}=0.75$ (normalized units), stiffness coefficient $\hat{\eta}_1=1.07$ ). Residuals are derived by removing a linear drift. Numbers of pairs used at each lag distance are also shown on the plot.	69
Figure 4.3 First- and second-order neighbors (full black circles) of extreme low value (at location marked by star) of the dataset located using Delaunay triangulation. The circles centered on the estimation point enclose the neighboring points for specific search radius (0.05 to 0.11 corresponding to normalized units) used for OK, DK and IDW for the extreme low value calculation.....	71
Figure 4.4 Isolevel contour map of estimated groundwater level in Mires basin using IDW.....	72

Figure 4.5 Isolevel contour map of estimated groundwater level in Mires basin using MC. ....	73
Figure 4.6 Isolevel contour map of estimated groundwater level in Mires basin using DK with the Spartan semivariogram model.....	73
Figure 4.7 Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using DK with the Spartan semivariogram model. ....	74
Figure 4.8 Isolevel contour map of estimated groundwater level in Mires basin using OK with the Spartan semivariogram model.....	74
Figure 4.9 Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using OK with the Spartan semivariogram model. ....	75
Figure 4.10 Isolevel contour map of estimated groundwater level in Mires basin using UK with the Spartan semivariogram model.....	75
Figure 4.11 Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using UK with the Spartan semivariogram model. ....	76
Figure 4.12 Comparison of groundwater level semivariograms: data (stars), OK estimates using Spartan (SP) semivariogram (circles), along with optimal SP model fits to data (dashed line) and to OK estimates (continuous line). ....	78
Figure 4.13 Comparison of groundwater level semivariograms: data (stars), DK estimates using Spartan (SP) semivariogram (circles), along with optimal SP model fits to data (dashed line) and to DK estimates (continuous line). ....	79
Figure 4.14 Comparison of omnidirectional groundwater level semivariograms of data (stars) and UK estimates (circles). ....	79
Figure 5.1 Plot of omnidirectional experimental semivariogram of transformed (MBC) groundwater level data (stars) and optimal Spartan model fit (SP), (parameter estimates: variance $\hat{\sigma}^2=13.4$ , characteristic length $\hat{\xi}=0.42$ (normalized units), stiffness coefficient $\hat{\eta}_1=0.97$ ). Numbers of pairs used at each lag distance are also shown. ....	87
Figure 5.2 a Distribution of MBC-OK cross-validation errors.....	89
Figure 5.2 b MBC-OK cross-validation estimates and their corresponding errors.....	90
Figure 5.3 Histograms of measured values and MBC-OK cross-validation estimates using seven classes of width 10 masl each, centered at 5, 15, 25, 35, 45, 55, and 65 masl respectively.....	90

Figure 5.4 Isolevel contour map of estimated groundwater level in the Mires basin using MBC-OK (red circles denote location of wells and solid black line the temporary river path). .....	93
Figure 5.5 Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using MBC-OK (red circles denote location of wells and solid black line the temporary river path). .....	94
Figure 5.6 Isolevel contour map of estimated groundwater level in the Mires basin using GA-OK (red circles denote location of wells and solid black line the temporary river path).....	94
Figure 5.7 Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using GA-OK (red circles denote location of wells and solid black line the temporary river path). .....	95
Figure 5.8 Isolevel contour map of estimated groundwater level in the Mires basin using TGK (red circles denote location of wells and solid black line the temporary river path).....	95
Figure 5.9 Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using TGK (red circles denote location of wells and solid black line the temporary river path). .....	96
Figure 5.10 Isolevel contour map of estimated groundwater level in the Mires basin using Box-Cox-OK (red circles denote location of wells and solid black line the temporary river path). .....	96
Figure 5.11 Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using Box-Cox-OK (red circles denote location of wells and solid black line the temporary river path).....	97
Figure 5.12 Isolevel contour map of estimated groundwater level in the Mires basin using OK (red circles denote location of wells and solid black line the temporary river path). .....	97
Figure 5.13 Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using Box-Cox-OK (blue circles denote location of wells and solid black line the temporary river path).....	98
Figure 5.14 Comparison of groundwater level omnidirectional experimental semivariograms: data (stars), and MBC-OK estimates (circles). .....	99

Figure 6.1 Locations of “anchor points” along the river bed (stars), second-order polynomial model of the river (continuous line), well locations (circles) and their projections (crosses) on the model curve of the temporary river.....	107
Figure 6.2 Annual average groundwater level variation in Mires basin for the period covering the years 1981-2003 (circles) and fitted linear trend (broken line). .....	111
Figure 6.3 Sensitivity analysis to determine the optimal hydraulic conductivity ( $K$ ) value used for the calculation of the radius of influence $R$ in (6.9) or (6.10) and subsequently for the hydraulic head (trend), in (6.8). The cross validation measure MAE, is calculated based on RK methodology. Different values of $K$ between the two extremes are investigated while $s_w$ is set equal to 1.85m.....	112
Figure 6.4 Plots of the omnidirectional semivariogram of the residuals (after applying MBC normalization) and the best-fit Spartan semivariogram (SP) model fit ( $\sigma^2=0.62$ , $\xi=0.38$ , $\eta_1=1.51$ ). The residuals are derived in the framework of the T-DEM-UGA-RD model, i.e., by subtracting a trend that accounts for distance from the river curve and surface elevation (in the uniform-gradient approximation). The numbers of pairs used at each lag distance are also shown on the plot. ....	118
Figure 6.5 Plots of the omnidirectional semivariogram of the residuals (after applying MBC normalization) and the best-fit Spartan semivariogram (SP) model fit ( $\sigma^2=9.9$ , $\xi=0.28$ , $\eta_1=0.52$ ). The residuals are derived in the framework of the T-MW model, i.e., by subtracting a trend that accounts for the groundwater level calculated from a multiple wells system operation analytical equation. The numbers of pairs used at each lag distance are also shown on the plot. ....	118
Figure 6.6 Comparison of omnidirectional experimental semivariograms of the groundwater level measurements and of leave-one-out estimates for T-RD and T-RD-DEM-UGA spatial models.....	119
Figure 6.7 Comparison of omnidirectional experimental semivariograms of the groundwater level measurements and of leave-one-out estimates for the T-MW spatial model.....	119
Figure 6.8 Map of estimated groundwater level in the Mires basin using RK-T-RD-DEM-UGA spatial model, adapted on the real basin coordinates and location in the valley.....	120

Figure 6.9 Map of estimated groundwater level standard deviation in the Mires basin using RK-T-RD-DEM-UGA spatial model, adapted on the real basin coordinates and location in the valley. ....	121
Figure 6.10 Map of estimated groundwater level in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley. ....	121
Figure 6.11 Map of estimated groundwater level standard deviation in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley. ....	122
Figure 6.12 Map of estimated groundwater level residuals in the Mires basin using RK-T-RD-DEM-UGA spatial model (interpolation of the residuals obtained from the subtraction of the measured values minus the trend determined from the approximation to ground surface elevation and the distance from the river curve), adapted on the real basin coordinates and location in the valley. ....	123
Figure 6.13 Map of estimated groundwater level trend in the Mires basin using RK-T-RD-DEM-UGA spatial model (trend using the gradient approximation to ground surface elevation and the distance from the river curve), adapted on the real basin coordinates and location in the valley. ....	124
Figure 7.1 Simplified demonstration of equation (7.7) inputs and output variables.	132
Figure 7.2 Representation of the space-time domain and of the space-time search neighborhood (after (Hengl 2007)) .....	150
Figure 7.3 ARX model results using precipitation surplus as exogenous variable. Red crosses denote the predicted values. ....	158
Figure 7.4 ARX model results using precipitation surplus and pumping rate as exogenous variables. Red crosses denote the predicted values. ....	158
Figure 7.5 ARX model results using precipitation surplus and pumping rate as exogenous variables. Red circles denote the calibration results using as initial parameters the optimal calculated from the last step during the validation process, Figure 7.4 (a=0.3756, b=0.0503, d=0.0205, c=0.0049). ....	159
Figure 7.6 Exponentially weighted moving average filter fit on mean bi-annual groundwater level measurements. ....	160
Figure 7.7 Topographic map showing the locations of the 10 monitored wells (triangles) in Mires basin along with the corresponding surface elevation and the	

temporary river path. With green color the wells monitored for the period 2003-2006 are presented. ....	161
Figure 7.8 Space time non-separable semivariogram fit. The upper space limit in real units is equal to 4Km and the time limit 6.5 years.....	162
Figure 7.9 Space-time product semivariogram fit using the Matérn structure. The upper space limit in real units is equal to 4Km and the time limit 6.5 years. ....	162
Figure 7.10 Map of estimated groundwater level in the Mires basin using STRK (7.56), the non-separable space-time semivariogram (7.64) on the residuals and spatiotemporal trend removal (7.59)(wet period of 2003-2004 hydrological year). .	165
Figure 7.11 Map of estimated groundwater level in the Mires basin using STRK (7.56), the non-separable space-time semivariogram (7.64) on the residuals and spatiotemporal trend removal (7.59) (dry period of 2004-2005 hydrological year)..	165
Figure 8.1 Map of estimated groundwater level in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley. ....	174
Figure 8.2 Map of estimated groundwater level in the Mires basin using STRK and the non-separable space-time semivariogram (wet period of 2003-2004 hydrological year) adapted on the real basin coordinates and location in the valley.....	175
Figure 8.3 Map of estimated groundwater level in the Mires basin using STRK and the non-separable space-time semivariogram (dry period of 2004-2005 hydrological year) adapted on the real basin coordinates and location in the valley.....	176
Figure 8.4 Graphical User Interface (GUI) in Matlab <sup>®</sup> for isotropic and anisotropic semivariogram fitting.....	177



## List of Tables

Table 3.1 Statistical measures of the hydraulic head data.  $z_{\min}$ : minimum value;  $z_{0.25}$  first quartile;  $z_{0.50}$  median;  $m_z$  mean;  $z_{0.75}$  third quartile;  $z_{\max}$  maximum value;  $\hat{\sigma}_z$  standard deviation;  $\hat{s}_z$  skewness coefficient;  $\hat{k}_z$  kurtosis coefficient.....60

Table 4.1 Optimal estimates of semivariogram model parameters obtained by a least squares fit to the experimental semivariogram of the data (Columns 2-4). The search radius defines the neighborhood used in the OK predictor (Column 5). The number of first- and second-order neighbors used by DK at each estimation point. ....66

Table 4.2 Optimal estimates of semivariogram model parameters obtained by a least squares fit to the experimental semivariogram of the residuals (Columns 2-4). The search radius defines the neighborhood used in the UK predictor (Column 5).....66

Table 4.3 Cross validation measures (section 2.5) for the stochastic and deterministic interpolation methods investigated. Results obtained with the three “optimal” (in terms of cross validation measures) semivariogram models are presented. The following abbreviations are used: IDW: Inverse distance weighted. MC: minimum curvature. DK: kriging with Delaunay triangulation. OK: Ordinary kriging. UK: Universal kriging. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. MAE: Mean absolute error. MARE: Mean absolute relative error. RMSE: Root mean square error. R: Linear correlation coefficient. Optimal values are emphasized.....67

Table 4.4 OK-SP, UK-SP (SP: Spartan semivariogram) and IDW estimates (masl) of the extreme low value in the dataset using search radii (normalized units). The radius 0.11 leads to a neighborhood similar to DK which generates the most accurate estimate, 29 masl, of the extreme low value). The numbers in parentheses denote the number of Delaunay neighbors present inside the corresponding search radius, while (+) denotes the presence of other neighbors as well (see Figure 4.3). Delaunay neighbors (symbolized with full black circles in Figure 4.3) are the vertices of the enclosing triangle (dark color) and of the three adjacent triangles (grey color).....71

Table 5.1 Normalization results using Box-Cox, Modified Box-Cox (MBC) and Gaussian Anamorphosis (GA) transformations: skewness coefficient $\hat{s}_z$ ; kurtosis coefficient $\hat{k}_z$ .....	86
Table 5.2 Cross-validation results of the spatial models with the optimal semivariograms based on the measures listed in section 2.5. OK: Ordinary Kriging. Box-Cox-OK: Box-Cox transformation followed by OK and back-transformation. MBC-OK: Modified Box-Cox transformation followed by OK and back-transformation. GA-OK: Gaussian Anamorphosis in combination with OK. T GK: Trans-Gaussian Kriging using the Box-Cox transform. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. MAE: Mean Absolute Error. MARE: Mean Absolute Relative Error. RMSE: Root Mean Square Error. R: Linear correlation coefficient. Optimal values are emphasized. ....	88
Table 6.1 Skewness $\hat{s}_z$ and kurtosis $\hat{k}_z$ coefficients of head data and of trend models residuals following Box-Cox normalization (for non-negative values) and modified Box-Cox (MBC). T-DEM model uses the surface elevation as an external variable, the T-DEM-UGA uses the uniform-gradient approximation of the DEM, the T-RD uses the distance from the river curve, the T-RD-DEM-UGA uses a linear combination of the distance from the river curve and the uniform-gradient approximation of the surface elevation and T-MW approximates the trend using <i>Thiem's</i> equation, Eq. (6.8).....	113
Table 6.2 Optimal estimates of semivariogram model parameters obtained by least squares fit to experimental semivariogram. Sill and characteristic length $\xi$ are in normalized units.....	114
Table 6.3 Cross validation results of spatial models with optimal semivariograms, based on measures listed in section 2.5. OK: Ordinary Kriging. Box-Cox & OK: Box-Cox transformation followed by OK and back-transformation. MBC & OK: Modified Box-Cox transformation followed by OK and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.....	114
Table 6.4 Cross validation measures (cf. section 2.5) for spatial T-DEM model with optimal semivariograms: trend using DEM surface elevation with $a = 0, f = 0.37, c = 3.75$ in (6.1). MBC & RK: Residual Kriging with modified Box-Cox	

transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. ....	115
Table 6.5 Cross validation measures (cf. section 2.5) for the spatial T-DEM-UGA model with the optimal semivariograms: trend using the uniform gradient approximation of the surface elevation with $a = 0, f = 0.28, c = 10.43$ in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of the residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. ....	116
Table 6.6 Cross validation measures (cf. section 2.5) for spatial T-RD model with optimal semivariograms: trend using distance from the river curve with $a = 52.90, f = 0, c = 20.90$ in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. ....	116
Table 6.7 Cross validation measures (cf. section 2.5) for T-DEM-UGA-RD model with optimal semivariograms: trend using gradient approximation to ground surface elevation and distance from river curve with $a = 52.07, f = 0.27, c = 2.22$ in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. ....	116
Table 6.8 Cross validation measures (cf. section 2.5) for T-MW model with optimal semivariograms: trend using hydraulic head obtained from multiple wells system operation. MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. $s_w$ : the drawdown at the well face. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. ....	117
Table 7.1 ARX parameters, determined from Kalman filter approach, with Precipitation Surplus data input for each consecutive run. ....	157
Table 7.2 ARX parameters, determined from Kalman filter approach, with Precipitation Surplus & Pumping data inputs for each consecutive run. ....	157
Table 7.3 Temporal Validation Results-Absolute Error (AE), Absolute Relative Error (ARE) - ARX with Precipitation Surplus for each consecutive run. ....	159
Table 7.4 Temporal Validation Results-Absolute Error (AE), Absolute Relative Error (ARE) - ARX with Precipitation Surplus & Pumping for each consecutive run. ....	159

Table 7.5 Absolute Error (AE) of STRK estimates for the wet period of highest groundwater level increase after 2003 (i.e. the wet period 2003-2004). .....	163
Table 7.6 Absolute Error (AE) of STRK estimations for the dry period of highest groundwater level drop (i.e. the dry period 2004-2005). .....	164
Table 7.7 Absolute Error (AE) of STRK estimation error of prediction average using the non-separable semivariogram model (7.64).....	166
Table 8.1 Cross validation measures for the stochastic and deterministic interpolation methods investigated. Results obtained with the “optimal” (in terms of cross validation measures) semivariogram model are presented. Optimal values are emphasized.....	172

## Abstract

The accurate representation of groundwater levels in an aquifer is very important for groundwater modeling and effective groundwater resources management. However, the number and spatial distribution of monitoring sites in a given aquifer are not always sufficient to accurately represent the water table. Predictions of groundwater level at unvisited locations of an aquifer can be obtained by applying geostatistical methods on the available groundwater level data and thus the free surface of the aquifer can be reliably mapped. In sparsely monitored basins, accurate mapping of the spatial variability of groundwater level requires the interpolation of scattered data. This thesis aims to present new modeling tools that help to better monitor and predict the groundwater level in sparsely gauged basins. The specific area of focus is the Mires basin of the Mesara valley in the island of Crete (Greece). The study area is a sparsely sampled basin that has limited groundwater resources which are vital for the area's welfare; spatiotemporal variations of groundwater level are important for developing management and monitoring strategies. Efficient groundwater management in the basin is crucial in light of regional climate change model estimates showing a substantial risk of desertification for Crete. Our goal is to construct accurate spatial and spatiotemporal models of the basin's groundwater level. Therefore, spatial and spatiotemporal models for the accurate representation of the groundwater level variability in already vulnerable areas with low groundwater resources, like Mires basin, need to be developed in order to identify the susceptible locations, to estimate the groundwater level distribution spatially and spatiotemporally and to provide input for potential groundwater resources management plans. The main data used in this research consist of seventy hydraulic head measurements (wet period of 2002-2003 hydrological year) which are unevenly distributed and mostly concentrated along a temporary river and time series performance consisting of biannual groundwater level data from ten boreholes (1981-2003). After the year 2003 observations of a shorter number of wells are available biannually.

This thesis initially presents a comparison of deterministic interpolation methods, i.e., Inverse Distance Weight (IDW) and Minimum Curvature (MC), with stochastic methods, i.e., Ordinary Kriging (OK), Universal Kriging (UK) and Kriging with Delaunay Triangulation (DK). We evaluate the performance of the interpolation

methods with respect to different statistical cross validation measures. The Spartan variogram family is implemented for the first time to hydrological data and is shown to be optimal with respect to the stochastic interpolation methods such as OK, UK, DK and Residual Kriging (RK) applied in this dataset. The three stochastic methods (OK, DK, UK) perform overall better than the deterministic counterparts (IDW, MC). DK, which is herein for the first time applied to hydrological data, yields the most accurate cross validation estimate for the lowest value in the dataset. OK and UK lead to smooth isopleth contours, while DK and IDW generate more edges. The stochastic methods also deliver estimates of prediction uncertainty.

The present research study also investigates the application of non-linear normalizing data transformations in conjunction with Ordinary Kriging (OK) for the accurate prediction of groundwater level spatial variability in a sparsely gauged basin. We investigate three established normalizing methods, Gaussian Anamorphosis, Trans-Gaussian Kriging and Box-Cox to improve the estimation accuracy. The first two are for the first time applied to groundwater level data. All three methods improve the mean absolute prediction error compared with the application of OK to the non-transformed data. In addition, a Modified Box-Cox (MBC) transformation is proposed and applied to normalize the hydraulic heads. MBC in conjunction with OK is found to be the optimal spatial model based on leave-one-out cross-validation. The recently established Spartan semivariogram family provides the optimal model fit to the transformed data.

Trend functions, as previous studies have shown, improve the accuracy of interpolation. Therefore, we propose that the prediction of the hydraulic head spatial variability in Mires basin can be improved by incorporating in the trend function local properties. Firstly the distance of the prediction points from the temporary river crossing the basin is incorporated in the trend function and secondly a component based on the generalized *Thiem's* equation for multiple wells. Residual Kriging is performed based on these two spatial trend models as well as using the novel MBC transformation to normalize the residuals and the flexible Spartan semivariogram family to optimally determine their spatial correlation. Both proposed spatial models' improve significantly the cross validation measures compared to the other Kriging-based methods tested. We also present maps of the groundwater level spatial variability and the estimation variance in Mires basin obtained by means of the optimal spatial models.

We determine that the optimal spatial interpolation approach is based on Residual Kriging using the non-differentiable Spartan semivariogram model for the normalized (MBC) fluctuations. Our proposal is supported by the results of cross validation analysis. The suggested methodology is applicable to other unconfined aquifers as well. The non-differentiable property of the Spartan semivariogram model is interpreted herein as the result of a deposition-removal process that leads to a fractional Brownian motion, fBm-like, behavior of the groundwater level surface.

The overall goal of this dissertation is to use stochastic methods for the monitoring and prediction of the groundwater level in space and time. Following the spatial interpolation, first we model the temporal variation of groundwater level with a discrete time autoregressive exogenous variable model (ARX) model. In this study pumping data are used in addition to precipitation measurements. The ARX model is embedded in a discrete-time Kalman filter to estimate the model parameters and predict the optimal mean annual groundwater level. The ARX model is calibrated for the years 1981 to 2006 and is then used to predict the mean annual groundwater level in the basin for recent years (2007-2010). The predictions are validated with the available annual averages reported by the local authorities.

Secondly, we use a spatiotemporal geostatistical analysis of the groundwater level using space-time Residual Kriging (STRK). The space-time trend is calculated using the product function of the estimated temporal trend from a weighted moving average filter and the spatial trend determined from the closest distance of the measurement locations from the river bed. A space-time experimental semivariogram is determined from the biannual (wet and dry period) groundwater level fluctuations time series between the years 1981 and 2003 at the ten sampling stations. We model the semivariogram with separable and non-separable theoretical spatiotemporal semivariogram functions. STRK is used to predict the groundwater level for selected hydrological periods at each sampling station in the time period (2004-2010) biannually.

Maps of groundwater level predictions and of prediction accuracy are desirable and significant in order to assess the groundwater level spatiotemporal variability, whether observed changes in water-table levels are statistically significant and finally to identify additional locations where further monitoring is needed to increase the accuracy of the maps. All the methodologies and tools presented in this

thesis are implemented by original code developed by the author and run in Matlab<sup>®</sup> programming environment (Matlab v.7.5 on Microsoft Windows XP).

## Περίληψη

Η ακριβής εκτίμηση της στάθμης του υδροφόρου ορίζοντα αποτελεί σημαντικό παράγοντα για ένα ολοκληρωμένο σχέδιο διαχείρισης των υδατικών πόρων μιας υδρολογικής λεκάνης. Η στάθμη των υπόγειων υδάτων σε έναν υδροφορέα ελέγχεται συνήθως με βάση το υδραυλικό ύψος στις θέσεις εγκατεστημένων πιεζόμετρων και ενεργών γεωτρήσεων σε κατάσταση ηρεμίας. Σε πολλές περιπτώσεις δεν παρακολουθούνται όλες οι γεωτρήσεις λόγω κόστους ή παράλειψης των αρμόδιων φορέων, ενώ σε άλλες περιπτώσεις ο αριθμός των γεωτρήσεων και των πιεζομέτρων είναι ανεπαρκής. Επομένως ο αριθμός των διαθέσιμων μετρήσεων είναι συχνά μικρός και αραιά κατανεμημένος στην περιοχή μελέτης με αποτέλεσμα να μην αντιπροσωπεύεται επαρκώς η στάθμη του υδροφορέα. Εκτιμήσεις της στάθμης σε θέσεις όπου δεν υπάρχουν παρατηρήσεις μπορούν να πραγματοποιηθούν με την εφαρμογή γεωστατιστικών μεθόδων στα διαθέσιμα δεδομένα, προκειμένου να χαρτογραφηθεί με ακρίβεια ο υδροφόρος ορίζοντας του υδροφορέα. Βοηθητικές πληροφορίες χρησιμοποιούνται συχνά για να ενισχύσουν τις εκτιμήσεις της στάθμης των υπόγειων νερών, όπως η τάση της φυσικής μεταβολής του επιπέδου των υπόγειων νερών (η οποία προσεγγίζεται συνήθως από πολώνυμα με βάση τις χωρικές συντεταγμένες), η βροχοπτώση και το υψόμετρο. Η χρήση βοηθητικών χωρικών μεταβλητών έχει αποδειχθεί ότι βελτιώνει τις εκτιμήσεις υδραυλικών υψών.

Το νησί της Κρήτης διαθέτει οριακούς υπόγειους υδατικούς πόρους, οι οποίοι χρησιμοποιούνται εκτενώς για γεωργικές δραστηριότητες και ύδρευση. Η κοιλάδα της Μεσαράς, η οποία βρίσκεται στο νότιο τμήμα του περιφερειακού διαμερίσματος Ηρακλείου και καλύπτει μια έκταση 398 km<sup>2</sup>, είναι η μεγαλύτερη και παραγωγικότερη κοιλάδα του νησιού. Η υπερεκμετάλλευση κατά τη διάρκεια των προηγούμενων τριάντα ετών έχει οδηγήσει σε μια δραματική μείωση, πάνω από τριάντα πέντε μέτρα, στη στάθμη των υπόγειων νερών. Οι πιθανές μελλοντικές κλιματολογικές αλλαγές στην περιοχή της Μεσογείου, τα σενάρια πιθανής ερημοποίησης και η εκτενής γεωργική δραστηριότητα προκαλούν έντονη ανησυχία σχετικά με την αιφορία των υδατικών πόρων της περιοχής. Η παρούσα διατριβή εστιάζει στην υδρολογική λεκάνη Μοιρών της κοιλάδας της Μεσσαράς για δύο λόγους: α) διαθεσιμότητα υδρογεωλογικών και υδρολογικών δεδομένων και β) διότι

αποτελείται κυρίως από προσχωματικές αποθέσεις κάτι που εξασφαλίζει σε μεγάλο βαθμό υδρογεωλογική ομοιογένεια.

Η παρούσα διατριβή έχει στόχο να παρουσιάσει καινοτόμες μεθοδολογίες χωρικής και χρονικής γεωστατιστικής ανάλυσης αλλά και μεθόδους-εργαλεία αποσκοπώντας στην εκτίμηση και χαρτογράφηση της χωρικής και χρονικής μεταβλητότητας της στάθμης των υπόγειων νερών της λεκάνης με τη βέλτιστη δυνατή ακρίβεια λαμβάνοντας υπόψιν και την αβεβαιότητα των εκτιμήσεων. Ο κύριος στόχος της έρευνας είναι να αναπτυχθούν χωρικά μοντέλα για τον ακριβή προσδιορισμό της στάθμης, ο υπολογισμός των αντίστοιχων ισοδυναμικών καμπυλών και ο προσδιορισμός των πιο ευάλωτων περιοχών του υδροφορέα. Σε δεύτερο στάδιο, τα μοντέλα επεκτείνονται σε δυναμικές (space-time) καταστάσεις για τη μοντελοποίηση της χωροχρονικής μεταβολής και τη δυνατότητα εκτίμηση της μελλοντικής στάθμης.

Η χωρική εξάρτηση τόσο των δεδομένων στάθμης όσο και των διακυμάνσεων που προκύπτουν από τα μοντέλα τάσης που αναπτύσσονται στην παρούσα διδακτορική διατριβή μελετώνται με την προσαρμογή εμπειρικών ημιβαριογραμμάτων σε γνωστά κλασικά πρότυπα συναρτήσεων, στο πρότυπο συναρτήσεων Matérn και στο μοντέλο Σπαρτιάτικου ημιβαριογράμματος το οποίο χρησιμοποιείται για πρώτη φορά σε υδρολογικά δεδομένα. Τα μοντέλα Σπαρτιάτικου τύπου και Matérn περιλαμβάνουν περισσότερες παραμέτρους σε σχέση με τα κλασικά πρότυπα, γεγονός που ευνοεί στη βέλτιστη προσαρμογή στο πειραματικό ημιβαριόγραμμα. Το Σπαρτιάτικο ημιβαριόγραμμα και η αντίστοιχη συνάρτηση συνδιασποράς έχουν αναπτυχθεί πρόσφατα και αποτελούν μια επιτυχώς εναλλακτική πρόταση στον προσδιορισμό της χωρικής εξάρτησης των δεδομένων.

Αρχικά αξιολογείται η χρήση γνωστών και ευρέως εφαρμοσμένων μεθόδων χωρικής παρεμβολής όπως το κανονικό Kriging (ordinary Kriging-OK), το γενικευμένο Kriging (universal Kriging-UK), η μέθοδος σταθμισμένων αντίστροφων αποστάσεων (inverse distance weight-IDW), η μέθοδος ελάχιστης καμπυλότητας (minimum curvature-MC) και η μέθοδος Kriging με τριγωνοποίηση κατά Delaunay (Kriging with Delaunay Triangulation-DK). Η τελευταία διερευνάται για πρώτη φορά σε υδρολογικά δεδομένα. Μάλιστα εξετάζεται και παρουσιάζεται η εφαρμογή της μεθόδου χρησιμοποιώντας δύο διαφορετικές γειτονίες εκτίμησης που προκύπτουν από τον τρόπο επιλογής των γειτονικών τιμών του εκτιμώμενου σημείου με βάση την πρωτεύουσα και δευτερεύουσα τριγωνοποίηση των γειτονικών τιμών. Η σύγκριση

των συγκεκριμένων πέντε μεθόδων στο ίδιο δείγμα πραγματοποιείται για πρώτη φορά σύμφωνα με τη μέχρι πρόσφατη βιβλιογραφία. Τα αποτελέσματα της σύγκρισης υποδεικνύουν ότι οι στοχαστικές μέθοδοι (OK, DK, UK) λειτουργούν με μεγαλύτερη αποτελεσματικότητα σε σχέση με τις αντίστοιχες προσδιοριστικές (IDW, MC). Επίσης το Σπαρτιάτικο ημιβαριόγραμμα αποδεικνύεται με βάση διαφορετικά στατιστικά μέτρα το βέλτιστο για την εφαρμογή της γεωστατιστικής εκτίμησης. Οι μέθοδοι OK και UK οδηγούν σε πιο ομαλές ισοδυναμικές σε αντίθεση με τις μεθόδους DK και IDW. Η μέθοδος DK η οποία χρησιμοποιείται για πρώτη φορά σε υδρολογικά δεδομένα υπολογίζει ακριβέστερα την ελάχιστη τιμή του δείγματος βάση της μεθόδου διασταυρωμένης επιβεβαίωσης. Οι στοχαστικές μέθοδοι πλεονεκτούν σε σχέση με τις προσδιοριστικές καθώς υπολογίζουν την αβεβαιότητα των εκτιμήσεων η οποία και μπορεί να υποδηλώσει περιοχές της υπό μελέτη λεκάνης όπου περισσότερες μετρήσεις απαιτούνται για γεωστατιστική ανάλυση μεγαλύτερης ακρίβειας.

Η προκαταρκτική γεωστατιστική ανάλυση των δεδομένων υδραυλικών υψών έδειξε ότι αυτά δεν ακολουθούν την κανονική κατανομή ωστόσο δεν είναι απαγορευτική η χρήση τους σε γραμμική γεωστατιστική ανάλυση. Για να δημιουργηθεί όμως ένα χωρικό μοντέλο με τη βέλτιστη δυνατή αποτελεσματικότητα για τα υδραυλικά ύψη εξετάζονται διάφορες μη γραμμικές προσεγγίσεις κανονικοποίησης των δεδομένων σε συνδυασμό με τη μέθοδο εκτίμησης kriging. Η μέθοδος Box-Cox και η μέθοδος της Γκαουσιανής Αναμόρφωσης (Gaussian Anamorphosis-GA) χρησιμοποιούνται για το μετασχηματισμό των δεδομένων στην κανονική (γκαουσιανή) κατανομή πιθανότητας. Προτείνεται επίσης μια νέα τροποποιημένη εκδοχή της μεθόδου Box-Cox (modified Box-Cox) η οποία βασίζεται στο συντελεστή κύρτωσης και ασυμμετρίας της παρατηρούμενης κατανομής. Χρησιμοποιείται ακόμη η μέθοδος Trans-Gaussian Kriging η οποία ενσωματώνει τη συνάρτηση μετασχηματισμού των δεδομένων στον εκτιμητή της χωρικής παρεμβολής. Η χρήση των εν λόγω μη γραμμικών μοντέλων αποτελεί καινοτομία στην υδρολογία. Η μέθοδος modified Box-Cox σε συνδυασμό με τη μέθοδο kriging και το Σπαρτιάτικο ημιβαριόγραμμα αποτελούν το βέλτιστο χωρικό μοντέλο εκτίμησης της στάθμης του υδροφορέα βάση σύγκρισης με τις υπόλοιπες μεθόδους. Παρουσιάζονται χάρτες ισοδυναμικών καμπυλών και αβεβαιότητας εκτιμήσεων με όλες τις στοχαστικές μεθόδους που εξετάστηκαν.

Επίσης προτείνεται και διερευνάται ένα καινοτόμο χωρικό μοντέλο τάσης για τη στάθμη των υπογείων υδάτων που βασίζεται στα τοπικά χαρακτηριστικά της λεκάνης των Μοιρών, και συγκεκριμένα στην απόσταση των γεωτρήσεων από την κοίτη του ποταμού. Το μοντέλο επίσης ενσωματώνει το αντίστοιχο υψόμετρο στη θέση της κάθε γεώτρησης. Μια δεύτερη καινοτόμα πρόταση χωρικού μοντέλου τάσης αποτελεί η χρήση των πολλαπλών πηγαδιών άντλησης στον υδροφορέα των Μοιρών. Η εφαρμογή της εξίσωσης *Thiem* για πολλαπλά πηγάδια άντλησης μπορεί να αποδώσει την τάση της στάθμης της λεκάνης. Η χρήση των προτεινόμενων μεθόδων για τον υπολογισμό της τάσης της στάθμης του υδροφορέα σε συνδυασμό με τη μέθοδο του υπολειμματικού kriging (Residual Kriging) οδηγεί στη βέλτιστη εκτίμηση του υδραυλικού ύψους σε σημεία της λεκάνης όπου δεν υπάρχουν παρατηρήσεις. Η μέθοδος συνοδεύεται από τη χρήση της καινοτόμου μεθόδου κανονικοποίηση δεδομένων modified Box-Cox (για τις διακυμάνσεις) και του μη διαφορίσιμου Σπαρτιάτικου ημιβαριόγραμματος για τον προσδιορισμό της χωρικής εξάρτησης των διακυμάνσεων. Οι προτάσεις μας υποστηρίζονται από τα αποτελέσματα διασταυρωμένης επιβεβαίωσης.

Το Σπαρτιάτικο ημιβαριόγραμμα αποτελεί το βέλτιστο μοντέλο προσδιορισμού της χωρικής εξάρτησης των δεδομένων για κάθε μεθοδολογία που εξετάστηκε στην παρούσα διατριβή. Ωστόσο η συνάρτηση αυτή είναι μη διαφορίσιμη. Η ιδιότητα αυτή ερμηνεύεται ως το αποτέλεσμα μιας διαδικασίας εμπλουτισμού και άντλησης του υδροφορέα η οποία οδηγεί τον υδροφόρο ορίζοντα σε συμπεριφορά κλασματικής κίνησης Brown (fractional Brownian motion).

Η μοντελοποίηση της χρονικής μεταβολής της στάθμης του υδροφορέα πραγματοποιείται με τη χρήση ενός μοντέλου αυτοσυσχέτισης το οποίο ενσωματώνει εξωγενή πληροφορία από μεταβλητές όπως η βροχόπτωση, η παροχή αντλήσεων και η εξατμισοδιαπνοή. Το συγκεκριμένο μοντέλο έχει χρησιμοποιηθεί στην αρχική του μορφή χωρίς την παροχή αντλήσεων. Στην παρούσα διατριβή διερευνάται η χρήση και της μεταβλητής αυτής για πρώτη φορά με επιτυχή αποτελέσματα. Το μοντέλο αυτοπαλινδρόμησης ενσωματώνεται σε ένα διακριτό χρονικά φίλτρο Kalman για την εκτίμηση των παραμέτρων αλλά και για την πρόβλεψη της βέλτιστης στάθμης του υδροφορέα.

Τα κύρια στοιχεία που χρησιμοποιούνται στην παρούσα διατριβή αποτελούνται από εβδομήντα (70) μετρήσεις στάθμης που αφορούν την υγρή περίοδο του υδρολογικού έτους 2002-2003 και από στάθμες 10 γεωτρήσεων που καλύπτουν

το χρονικό διάστημα 1981-2003 σε υγρή και ξηρή περίοδο. Οι μετρήσεις κατανέμονται ανομοιόμορφα και εστιάζονται κατά μήκος του ποταμού Γεροπόταμου που διασχίζει τη λεκάνη. Από το 2003 και μετά, μετρήσεις πραγματοποιήθηκαν μόνο σε μικρό αριθμό επιλεγμένων γεωτρήσεων ενώ από το 2003 δύο τηλεμετρικοί σταθμοί λειτουργούν στην περιοχή παρακολουθώντας τη μεταβολή της στάθμης των υπογείων υδάτων.

Η ακριβής χωρική εκτίμηση της στάθμης σε συνδυασμό με τη μοντελοποίηση της χρονικής μεταβολής δημιουργούν τις συνθήκες για ένα ολοκληρωμένο χωροχρονικό μοντέλο το οποίο θα περιγράψει με ακρίβεια τη στάθμη του υδροφορέα και θα μπορεί να εκτιμά και τη μελλοντική συμπεριφορά του. Για το σκοπό αυτό οι τάσεις της χρονικής μεταβολής της στάθμης μελετώνται και προσδιορίζονται από ένα μοντέλο σταθμισμένου κινούμενου μέσου όρου (weighted moving average). Η χωροχρονική τάση προσδιορίζεται από το συνδυασμό του σταθμισμένου κινούμενου μέσου όρου και της απόστασης των γεωτρήσεων από την κοίτη του ποταμού. Στη συνέχεια η χωροχρονική εξάρτηση των διακυμάνσεων των μετρήσεων υπολογίζεται με τη βοήθεια του χωροχρονικού εμπειρικού ημιβαριογράμματος. Η μοντελοποίηση του πραγματοποιείται με τη χρήση διαχωριζόμενων και μη διαχωριζόμενων χωροχρονικών συναρτήσεων. Η επαλήθευση εκτίμηση και πρόβλεψη στάθμης πραγματοποιείται με την εφαρμογή του χωροχρονικού υπολειμματικού kriging (Residual Kriging). Η χρήση μη διαχωρίσιμων χωροχρονικών συναρτήσεων αποτελεί επίσης καινοτομία στην υδρολογία.

Οι μεθοδολογίες που αναφέρονται παραπάνω δύναται να εφαρμοστούν και σε άλλες υδρολογικές λεκάνες με παρόμοια χαρακτηριστικά όπως αυτά της λεκάνης των Μοιρών. Η υλοποίηση των προτεινόμενων μεθόδων πραγματοποιήθηκε από το συγγραφέα σε πρωτότυπο κώδικα στο προγραμματιστικό περιβάλλον Matlab® ενώ η αξιολόγηση των γεωστατιστικών μεθόδων πραγματοποιείται με τη μέθοδο της διασταυρωμένης επιβεβαίωσης χρησιμοποιώντας διάφορα στατιστικά μέτρα επίδοσης. Πιστεύουμε ότι η παρούσα διδακτορική διατριβή συνεισφέρει στην επιστήμη της γεωστατιστικής αλλά και στην υδρολογική διερεύνηση της περιοχής μελέτης. Τα αποτελέσματα και οι μέθοδοι που παρουσιάζονται μπορούν να χρησιμοποιηθούν σε συνδυασμό με άλλες γεωεπιστήμες για την ολοκληρωμένη διαχείριση των υδατικών πόρων της λεκάνης των Μοιρών και της ευρύτερης περιοχής της Μεσσαράς.



# 1. Introduction

## 1.1 Motivations for this research

During the last thirty years geostatistics has been successfully applied in several environmental and earth sciences disciplines. Geostatistics helps to overcome field data deficiencies such as, sparse and scarce measurements, uncertainty estimation and space-time data combination providing space-time predictions for variables with environmental and economical importance. To date research on space-time dependent variables and geostatistics continues, in order to develop new more efficient space-time methodologies.

This thesis is primarily motivated by the need for accurate interpolation methodologies in order to determine with the highest possible accuracy the spatiotemporal variability of field data, i.e. hydrological data. Therefore below we introduce space and time geostatistical methodologies which we believe that have something new to contribute in geostatistics, e.g.: field data spatial correlation using the Spartan variogram family, kriging-based spatial models using non-linear normalizing data transformations, kriging-based spatial trend models capturing local properties, Kriging with Delaunay triangulation (DK) using second order neighbors applied for the first time in hydrological data, anisotropy estimation using a recently established method named covariance Hessian identity also applied for the first time in hydrological data, spatiotemporal trend calculation using a novel function based on local basin properties and the exponentially weighted moving average filter, spatiotemporal interpolation using a non-separable semivariogram function for the first time in real data, a comparison of well known stochastic and deterministic interpolation methods that in the same dataset has not been applied before.

Secondly, this thesis is motivated by the dramatic decrease in groundwater levels during the last decades in many Mediterranean basins due to overexploitation. Such an example is Mires basin of the Mesara valley in the island of Crete-Greece. In light of this development and the expected adverse effects of climate change on the basin's water resources, accurate spatiotemporal modeling of the groundwater level variation is significant and is needed for two main reasons: a) to identify "vulnerable" locations on the basin where an integrated groundwater resources management plan should focus and b) to provide accurate information for numerical groundwater flow

models, such as the calibration of equipotentials and the representation of initial groundwater level conditions.

## 1.2 Objectives

The main objective of this thesis is to develop and test interpolation methodologies and geostatistical tools for the accurate mapping of groundwater level's spatial variability in sparsely monitored basins. Therefore initially in this thesis the interpolation performance of OK, UK and of the newly established DK with the deterministic methods IDW and MC on the same groundwater level data set is compared. To our knowledge, this is the first application of DK to groundwater level interpolation. The dataset used involves groundwater levels in a sparsely gauged basin. Measuring the relative performance of different interpolators is important for environmental monitoring.

This dissertation investigates the improvement in groundwater level interpolation with OK using non-linear data normalization methodologies. Well-known OK based methodologies are applied, most of them for the first time to groundwater level data. In addition, a novel normalization method based on the Box-Cox transformation, referred to as Modified Box-Cox (MBC) is established and implemented in this study. The (MBC) method, Gaussian Anamorphosis (GA) normalization and Trans-Gaussian Kriging (TGK) are applied for the first time to groundwater level data.

In addition this dissertation introduces auxiliary trend variables based on local features (i.e., a temporary river crossing the basin) and physical laws (i.e., *Thiem's* multiple well equation) to improve the prediction of groundwater level.

Overall, several kriging-based spatial models are investigated, evaluated, and maps of estimated water table elevation and its associated uncertainty are generated by means of the optimal model.

Another objective of this dissertation is to introduce some recently developed geostatistical tools in the hydrological literature. The recently established Spartan semivariogram family is applied herein along with classical semivariogram models to calculate the data spatial dependence. More specifically, the flexible Spartan semivariogram family is applied for the first time to hydrological data and is shown (based on cross validation) to be the optimal model of spatial variability in Mires. The

geometric anisotropy is estimated using the newly established Covariance Hessian Identity.

The final objective of this research work is to use stochastic methods for the spatiotemporal monitoring and prediction of the groundwater level in sparsely gauged basins e.g. Mires basin, located on the island of Crete (Greece). First, we model the temporal variation of groundwater level with a discrete time autoregressive exogenous variable model (ARX) model and then we perform spatiotemporal geostatistical analysis of the groundwater level taking account the space time groundwater level trend using space-time Residual Kriging (STRK).

### **1.3 Innovation**

This research addresses some practical problems of hydrological data geostatistical analysis and contributes to geostatistics, hydrological theory and methodology and to factual information about the hydrology of a region.

The Spartan variogram family is tested and applied herein for the first time to hydrological data. The application and investigation of the variogram's efficiency on real field data is one of primary objectives of this thesis. Furthermore a recently proposed method to estimate Geometric anisotropy is tested in this thesis namely the Covariance Hessian Identity. This method is also applied for the first time in hydrology.

It also examines the use of non-linear transformation of groundwater level data to obtain improved kriging estimates of the water table elevation. The thesis deals with several original ideas: (i) the Modified Box-Cox, Gaussian Anamorphosis and Trans-Gaussian Kriging transformations have not been previously applied to groundwater level data; (ii) the use of the Matern and Spartan models for the semivariogram is novel to groundwater level data; (iii) the application to real data is another feature of the thesis.

This thesis introduces two novel kriging based spatial models, for the groundwater levels accurate representation in sparsely monitored basins. The proposed spatial models include a trend component where auxiliary variables that incorporate specific features of the studied watershed are included. The first model incorporates in the trend the distance of the wells from the river bed in addition to surface elevation, while the second uses a novel trend approach that involves the

groundwater level predicted from a groundwater hydraulics' equation (*Thiem's*) regarding multiple wells system operation.

New approach is also the application of a discrete time autoregressive exogenous variable (ARX) model that uses in advance to the original approach the groundwater abstraction rate in order to model the aquifer's groundwater level temporal variability. In addition new approach is the function applied to determine the groundwater level's spatiotemporal trend. This function involves a weighted moving average filter for the temporal trend and the closest distance of the wells from the river bed for the spatial trend. Moreover the non-separable spatiotemporal function used to model the experimental spatiotemporal semivariogram of fluctuations is applied for the first time in real data providing very good estimates better than the classical spatiotemporal separable product function.

Finally it examines the use of Delaunay triangulation in conjunction with kriging for interpolation of groundwater level in sparsely monitored basins for the first time. The application of Delaunay triangulation with second neighbours is for the first time applied to field data. Moreover it compares IDW, MC, OK, UK and DK in the same dataset. Such a comparison has not been met in the scientific bibliography.

## **1.4 Outline of the thesis**

The remainder of this thesis is organized as follows. In Section 2 background and theory of geostatistical methodologies and applications are reviewed. Section 3 presents relevant information for the study area (Mires basin) and an exploratory statistical analysis of the data set. Section 4 compares stochastic and deterministic methods for mapping groundwater level spatial variability. In Section 5 we introduce non-linear normalizing transformations for improving kriging of groundwater level data. Section 6 focuses on improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. Section 7 presents stochastic tools for the space-time modeling of groundwater level variations. Finally section 8 contains a general discussion of the results and concluding remarks.

## 2. Background and theory

The Mediterranean Sea region is affected by the global climate change. It is expected, according to recent climatic modeling results, to be further affected in the future (Intergovernmental Panel on Climate Change 2007, Tsanis *et al.* 2011). The Intergovernmental Panel on Climate Change (IPCC) (2007) and a series of climate change studies (Giorgi and Lionello 2008, Somot *et al.* 2008, Tsanis *et al.* 2011) report that the Mediterranean is highly sensitive and vulnerable to climate change, with recent simulations estimating substantial drying and warming effects. These predictions represent a serious threat to water resources in the region. According to the IPCC, there will be a global surface temperature increase of 1.1 - 6.4 °C until 2100. The Mediterranean is expected to warm significantly, well above the global average. IPCC projections suggest that annual precipitation throughout most of the Mediterranean will be significantly reduced (fewer precipitation days, significantly drier summers and a higher risk of drought). Evaporation rates are also expected to increase leading to further reduction of aquifer recharge and surface runoff (Intergovernmental Panel on Climate Change 2007, Bates *et al.* 2008, Howard 2011).

The analysis of climate model data for the island of Crete indicates that the extreme events of the last few years will intensify, i.e., precipitation is expected to be less frequent but more intense, the average temperature will increase, while the severity and frequency of droughts will also increase in some regions. The quantitative impact of these changes on water resources can be significant at basin level (Tsanis *et al.* 2011). Mires basin in the Mesara valley of Crete has registered decreasing trends in annual precipitation and groundwater level over the last 30 years (see section 3.3). At the same time, the water demand is increasing because Mesara valley is the most productive agricultural valley of Crete.

Groundwater resources are very important to both humans and the environment. Hence it is essential to understand and control the environmental impact of groundwater overexploitation. The expansion of irrigated agriculture leads in many cases to groundwater overexploitation with serious impact on the water resources budget and the environment. This results in aquifer depletion, water quality degradation, stream flow reduction, and in major losses of habitat and biodiversity. It is necessary for the authorities to set safe limits of groundwater availability and vulnerability and to reconcile the human development with the preservation of nature

(Sophocleous 2002). Therefore, the accurate representation of the groundwater level and its spatiotemporal variability in Mires (as well as in basins with similar characteristics), is an important management tool for identifying vulnerable areas where pumping needs to be controlled or discontinued. The island of Crete generally may be characterized as having marginal groundwater resources, which are extensively used for agricultural activities and for human consumption (Donta *et al.* 2006, Department of Water Resources Management 2009). A characteristic example of an area where the groundwater resources are overexploited is Mesara valley.

Groundwater levels in an aquifer are usually monitored by means of hydraulic head measurements at borehole locations. In many cases, only a subset of the existing boreholes are monitored due to financial constraints or omission by the responsible authorities; in other cases, the number of operating boreholes is inadequate for a global representation of head variability. In both cases, geostatistical methods can help to more accurately visualize the surface of an aquifer. The geostatistical approach allows the reproduction of spatial variability, while it also honors the available observation data. Hence geostatistics is traditionally used to modeling aquifer properties.

## **2.1 Mathematical background**

Geostatistics has been well established and developed during the last three decades and is widely applied in environmental research and technology (Journel and Huijbregts 1978, Isaaks and Srivastava 1989, Christakos 1991b, Deutsch and Journel 1992, Cressie 1993, Goovaerts 1997, Kitanidis 1997, Christakos 2000). Geostatistics is a sub-discipline of spatial statistics. It includes a set of statistical methods that concern random variables with spatial and/or temporal variability (random fields). These variables represent physical quantities with economic or environmental importance. These methods are based on the assumption that the spatiotemporal variability includes a random component which has space-time correlation. Therefore statistical measures such as mean value, variance, standard deviation, spatiotemporal dependence, e.t.c, are used to extract any useful information from the available data (Mamassis 2006). Geostatistics deals with distributions in which the spatial and/or temporal dependence is the primary characteristic. Geostatistical analysis aims to estimate the statistical parameters that determine the spatial and/or temporal distribution and dependence of the relevant variables. This procedure is called

parameter inference. These parameters are used to estimate (interpolate) the variables at desired spatiotemporal locations where no measurements are available (Hristopulos 2003a).

### *2.1.1 Random fields*

Geostatistics is intrinsically connected and based on the mathematical concept of Random fields (RF). As RF it can be considered a set of random variables that describe the spatiotemporal variation of a physical variable size (e.g. hydraulic head, concentration of a pollutant). Contrary to functions that have a specific mathematical expression, e.g.  $f(x) = \cos(x)$ , random fields don't have a specific expression that represents all possible states. Each state is one sample of the field and is characterized by a probability determined by the multidimensional Probability Density Function of the field. Therefore, a random field can be considered as a multidimensional random variable. Due to the interdependence of the physical characteristics in different points of the space, random fields have particular mathematical properties that distinguish them from a set of independent random variables (Hristopulos 2008).

There are various categories of random fields. If the field takes values only from a finite set of numbers it is called discrete field. If the values of the field belong to a continuous interval of real numbers, the field is called continuous field. When variation is defined in a continuous space, such as natural fields, a continuous field is created. On the contrary, when the positions of a grid are defined the field is called lattice field.

Lattice fields are used in computational (e.g. simulation of the distribution of contaminants in groundwater) but also in theoretical studies, because grid symmetry allows the use of efficient numerical methods (e.g. fast Fourier transform). Moreover lattice fields allow benchmarking of different geostatistical methods (Hristopulos 2008).

In practice the measurements represent a finite number of points, the distribution of which does not necessarily have the symmetry of a regular grid. In these cases the network of sampling points is inhomogeneous. The terms disordered lattice and off lattice can be used as well. On such cases, geostatistical methods are needed to operate adequately, considering the limitations of each spatial distribution. If the distribution is off lattice, the evaluation or simulation of

procedure is realized on a gridded background that covers the area of interest (Hristopulos 2006).

The concept of random fields is based on two key terms: randomness and interdependence of values of physical quantities at different points of the space. Randomness characterizes phenomena in which knowledge of a situation with complete accuracy is impossible due to various constraints. Such constraints originate from the variability of different physical quantities in space and the uncertainty due to limited number of measurements. In these cases the result (the value of the phenomenon) is determined via a probability distribution function, which defines the probability of occurrence of each state.

Spatial dependence is a particular feature in random fields and describes the reliance between the values of two different points in the field. The probability distribution of the field embodies correlations between different points, so the probability of observing a value at a point depends on the values in adjacent points (Hristopulos 2008).

### 2.1.2 Basic concepts in random fields

A random field is denoted as  $Z(\mathbf{s})$ , where  $\mathbf{s}$  is a position vector  $\mathbf{s} = (x, y)$ .  $Z(\mathbf{s})$  represents all possible states in the field, while  $z(\mathbf{s})$  denotes the values that correspond to a specific state. Probability Density Function (PDF) of the field is denoted as  $f_Z[z(\mathbf{s})]$ . Index  $Z$  indicates the field, while the argument of the function is the values of the state of the field (e.g. hydraulic head, concentration of pollutants). An example of PDF, which corresponds to a normal distribution random field, is

given by:  $f_Z[z(\mathbf{s})] = \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left[-\frac{(z(\mathbf{s}) - m'_Z(\mathbf{s}))^2}{2\sigma_Z^2}\right]$ , where  $m'_Z(\mathbf{s})$  is the mean value,

$\sigma_Z^2$  the variance and  $\sigma_Z$  the standard deviation of the random field.

Probability Density Function of a random field includes all values in the space where the field is defined. Therefore PDF is common for any number of points. One-dimensional or point PDF describes all possible states in the field, on a specific point. It is possible that the one-dimensional PDF changes from point to point and that happens when the field is inhomogeneous. Proportionally, two-dimensional PDF of the field expresses

the interdependence of possible states of two points, while multidimensional PDF describes the interdependence of all possible situations for  $N$  points (Hristopulos 2008).

Another type of functions that provides information about the properties of a random field is statistical moments. Statistical moments are deterministic functions which represent average values in all possible situations. In practice, usually low order (up to second order) statistical moment, as mean value, dispersion, covariance functions and semivariogram, is useful (Hristopulos 2008).

Spatial random fields (SRF) are random fields that location plays the primal role as the property values are spatially correlated. An SRF state can be decomposed into a deterministic trend  $m_z(\mathbf{s})$  a correlated fluctuation  $Z'_\lambda(\mathbf{s})$ , and an independent random noise term  $e(\mathbf{s})$  so that,  $Z(\mathbf{s}) = Z'_\lambda(\mathbf{s}) + m_z(\mathbf{s}) + e(\mathbf{s})$ . The fluctuation term corresponds to ‘fast variations’ that reveal structure at small scales, which nonetheless exceed a cut-off  $\lambda$ , the trend is often determined from a single available realization. The random noise represents non-resolved inherent variability due to resolution limits, purely random additive noise, or non-systematic measurement errors. The classical approach of SRF’s is based on Gaussian SRF’s (GSRF’s) and various generalizations for non-Gaussian distributions (Wackernagel 2003). The covariance matrix therefore is used to determine the spatial structure for the GSRF’s which is estimated from the distribution of the data in space. Generally SRF’s model spatial correlations of variables and have various applications e.g., in hydrology (Kitanidis 1997), environmental pollutant mapping and risk assessment (Christakos 1991b), mining exploration and reserves estimation (Goovaerts 1997).

### 2.1.3 Mean value

The mean value of a random field is given by:

$$m'_z(\mathbf{s}) = E[Z(\mathbf{s})]. \quad (2.1)$$

$E[Z(\mathbf{s})]$  denotes the mean value, calculated in all states of the field, i.e.

$$E[Z(\mathbf{s})] = \int dz f_z(z; \mathbf{s}) z, \quad (2.2)$$

where  $z$  the values that correspond to a given state. The integral limits depend on the space where field  $Z$  is defined. If the field takes all negative and positive values the integral varies from  $-\infty$  to  $\infty$ . If the field takes only positive values the integral ranges from 0 to  $\infty$ . If it is known that the values of the field are limited to a predetermined interval  $[a,b]$ , the integral is calculated in this interval. In the latter equation it can be noted that the average value may depend on position,  $\mathbf{s}$ , which comes from a possible dependence between the one-dimensional probability density function and the position. Since PDF is not always known in advance, mean value is estimated through the sample using statistical methods. This is the average of all values in the sample, (Hristopulos 2008):  $\hat{m}'_Z(\mathbf{s}) = \frac{1}{N} \sum_{i=1}^N z_i(\mathbf{s})$ . A useful application topic of the mean value is to describe the large-scale trends in a random field. Mean value  $m'_Z(\mathbf{s})$  is defined using reference functions. They can be divided in general and local dependence patterns. In the case of general dependence only one mathematical equation describes the variance in the entire area. This kind of dependence patterns are:

- Linear dependence, e.g.  $m'_Z = m_0 + \mathbf{b} \mathbf{s}$  which expresses the existence of a constant slope
- Polynomial dependence, e.g.  $m'_Z(\mathbf{s}) = m_0 + b_1 s_1 + b_2 s_2$
- Periodic dependence, e.g.  $m'_Z(\mathbf{s}) = m_0 + \sum_{n=1}^N A_n \cos(\mathbf{k}_n \mathbf{s} + \phi_n)$ , where variables  $\mathbf{k}_n$  correspond to spatial frequencies and  $\phi_n$  in phases
- The overlay of two or more patterns, e.g. a polynomial and a periodic,  $m'_Z(\mathbf{s}) = (m_0 + b_1 s_1 + b_2 s_2) + (m_0 + \sum_{n=1}^N A_n \cos(\mathbf{k}_n \mathbf{s} + \phi_n))$ .

In cases where the general dependence patterns are insufficient for the exact determination of the trends, the use of local dependence functions is preferable (e.g. local polynomials). Such type of dependence is used in the model of locally weighed regression (Hristopulos 2008).

### 2.1.4 Variance

Variance in a random field is given by the mean value of the squared fluctuation according to equation,

$$\sigma_Z^2(\mathbf{s}) \equiv E\left[\{Z(\mathbf{s}) - m'_Z(\mathbf{s})\}^2\right] = E\left[\tilde{Z}^2(\mathbf{s})\right]. \quad (2.3)$$

In general, it is possible for the variance to vary from point to point while remaining stable only when the field is statistically homogeneous. The variance fluctuations in a random field mean that the fluctuations in the field change from point to point (Isaaks and Srivastava 1989).

### 2.1.5 Covariance function

Another property which gives useful information for a random field is the centered covariance function (CCF), which is defined as (Isaaks and Srivastava 1989):

$$c_Z(\mathbf{s}_1, \mathbf{s}_2) \equiv E\left[\{Z(\mathbf{s}_1) - m'_Z(\mathbf{s}_1)\}\{Z(\mathbf{s}_2) - m'_Z(\mathbf{s}_2)\}\right]. \quad (2.4)$$

The random field  $\tilde{Z}(\mathbf{s}_1) \equiv Z(\mathbf{s}_1) - m'_Z(\mathbf{s}_1)$  corresponds to the fluctuation in field  $Z(\mathbf{s}_1)$  around the mean value at point  $\mathbf{s}_1$ . The mean value of the fluctuation field is equal to zero,  $E[\tilde{Z}(\mathbf{s}_1)] = 0$ . Based on the previous equations it holds:

$$c_Z(\mathbf{s}_1, \mathbf{s}_2) = E\left[\tilde{Z}(\mathbf{s}_1)\tilde{Z}(\mathbf{s}_2)\right]. \quad (2.5)$$

Specifically, CCF describes quantitatively the dependence of the fluctuations between two different points in the field. When the points of the covariance function coincide, the value is equal to the variance of the field at that point  $c_Z(\mathbf{s}_1, \mathbf{s}_1) = \sigma_Z^2(\mathbf{s}_1)$ . On the contrary, when the distance between two points grows larger, the dependence of the fluctuations is reduced. An example of change of the covariance function with the distance follows: Let the covariance function between two points in a random field be given by the exponential model,  $c_Z(\mathbf{s}_1, \mathbf{s}_2) = \sigma_Z^2 \exp\left(-\frac{\|\mathbf{r}\|}{\xi}\right)$ , where  $\|\mathbf{r}\|$  is the

Euclidian norm of the distance vector between two points and  $\xi$  is the correlation length. If  $\|\mathbf{r}\|=0$  then  $c_z(\mathbf{s}_1, \mathbf{s}_2) = \sigma_z^2$ , while if  $\|\mathbf{r}\| = \xi$  then  $c_z(\mathbf{s}_1, \mathbf{s}_2) = 0.36\sigma_z^2$  (Hristopulos 2008).

In geostatistical analysis the experimentally determined spatial dependence is fitted to an optimal model selected by a set of accepted theoretical functions (e.g. exponential, Gaussian, powerlaw e.t.c). A function is a valid covariance function if and only if it satisfies the following criteria:

$$\sum_{i=1}^N \sum_{j=1}^N a_i a_j c_z(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \quad (2.6)$$

for any real weights  $a$ ,  $i, j = 1, \dots, N$  and any positive integer  $N$ . Acceptance conditions are also necessary for the covariance function. The acceptance conditions are set by Bochner's theorem (Bochner 1959). This is expressed through the power spectral density of the covariance which is given by the Fourier transformation (Press *et al.* 1992) of the covariance function. Power spectral density is defined by the integral:

$$\tilde{c}_z(\mathbf{k}) = \int d\mathbf{r} \exp(-i\mathbf{k} \cdot \mathbf{r}) c_z(\mathbf{r}), \quad (2.7)$$

where  $\mathbf{r}$  is the distance vector between two points,  $\int d\mathbf{r} = \int dx \int dy$  and  $\mathbf{k}$  is the vector of spatial frequency (wavevector). Function  $c_z(\mathbf{r})$  is an accepted covariance function if the three following conditions are applicable:

- 1) If the power spectral density exists  $\tilde{c}_z(\mathbf{k})$  (i.e. if the Fourier transformation of the function exists).
- 2) If  $\tilde{c}_z(\mathbf{k})$  is non-negative throughout the range of frequencies, i.e.  $\tilde{c}_z(\mathbf{k}) \geq 0$  for every  $\mathbf{k}$ .
- 3) If the integral of  $\tilde{c}_z(\mathbf{k})$  throughout the range of frequencies is bounded (i.e. if the variance exists).

In practice, to determine if a function is an acceptable covariance, the Fourier transform of the function needs to be calculated (Hristopulos 2008).

### 2.1.6 Statistical homogeneity

Assumptions that impose constraints on the properties of a random field can lead to a more efficient geostatistical analysis. The most widely used simplifying assumption is statistical homogeneity, which is an extension of the classical definition of homogeneity. A given property is homogenous if the corresponding variable is constant in space. On the contrary, a random field is statistically homogeneous if the mean value is constant,  $m'_Z(\mathbf{s}) = m'_Z$ , covariance function is defined and depended only on the distance vector  $\mathbf{r} = \mathbf{s}_1 - \mathbf{s}_2$  between two points  $c_Z(\mathbf{s}_1, \mathbf{s}_2) = c_Z(\mathbf{r})$  and the variance of the field is also constant. These conditions define also 2<sup>nd</sup> order stationarity.

These conditions define the statistical homogeneity in a weak sense. A random field is statistically homogeneous in a strong sense when the multidimensional PDF for  $N$  points (where  $N$  is any positive integer number) remains unchanged by transformations that alter the location of the points without altering the distances between them. Therefore the concept of statistical homogeneity is that the statistical properties of a random field does not depend on the spatial coordinates of the points, hence the reference system. Practically, statistical homogeneity implies that there are no systematic trends, so the change of the values in the field can be attributed to fluctuations around a constant level equal to the mean value (Hristopulos 2008).

### 2.1.7 Statistical isotropy

Another property that can be useful in geostatistical analysis of a random field is statistical isotropy. A field is statistically isotropic if it is statistically homogenous and at the same time the covariance function depends on the distance (Euclidean distance), but not on the direction of the distance vector  $\mathbf{r}$ . This is important from a practical point of view because it helps in the identification of spatial dependence. If a covariance function is statistically isotropic is by definition statistically homogeneous, but not vice versa.

In the case of statistically isotropic fields the two most important parameters that determine very basic features of covariance function is the variance  $\sigma_z^2 = c_z(0)$  and correlation length  $\xi$ . Variance is a measure of the width of the fluctuations in the field. The correlation length defines the interval in which there is interdependence, which defines the distance within which the field value at one point affects the value at another point (Christakos 1991b).

### 2.1.8 Spatial dependence

There are several ways to measure the spatial dependence. Two of the most commonly used is the semivariogram and the correlation function. Both functions describe the dependence between two points in the statistical sense as both functions refer to pairs of points so their value depends on the distance between these points. The term, in the statistical sense, means that the described dependence emerges as a mean value from a large number of pairs and not a single pair of points (Hristopulos 2008). Correlation function for a random field is equal to the ratio of the covariance function to the variance and is given by the equation,  $\rho_z(\mathbf{r}) = \frac{c_z(\mathbf{r})}{\sigma_z^2}$ , while the semivariogram of a random field is defined by the equation,

$$\gamma_z(\mathbf{s}, \mathbf{r}) = \frac{1}{2} E \left\{ [Z(\mathbf{s} + \mathbf{r}) - Z(\mathbf{s})]^2 \right\}. \quad (2.8)$$

The semivariogram is defined in relation to a pair of points, using the mean squared difference:  $\delta Z(\mathbf{s}; \mathbf{r}) \equiv Z(\mathbf{s} + \mathbf{r}) - Z(\mathbf{s})$ . The difference field  $\delta Z(\mathbf{s}; \mathbf{r})$  is called distance step  $\mathbf{r}$ . If the field  $Z(\mathbf{s})$  is statistically homogeneous the semivariogram is directly related to the covariance function by the equation (Deutsch and Journel 1992, Hristopulos 2008):

$$\gamma_z(\mathbf{r}) = \sigma_z^2 - c_z(\mathbf{r}). \quad (2.9)$$

For statistically homogeneous fields, semivariogram contains the same information as the covariance function. If the difference  $\delta Z(\mathbf{s}; \mathbf{r})$  is statistically homogeneous, the random field  $Z(\mathbf{s})$  is called field with statistically homogeneous differences. In this case the semivariogram  $\gamma_z(\mathbf{r})$  depends solely on the distance  $\mathbf{r}$

between the points and this is a result of statistical homogeneity of the field differences. If the field  $Z(\mathbf{s})$  is statistically homogeneous then the same applies for the differences  $\delta Z(\mathbf{s}; \mathbf{r})$ , but the opposite is not necessarily true (Hristopulos 2008).

The parameters of the semivariogram determine the spatial dependence of the field values at two neighboring points. From the definition of the semivariogram, using the mean square of the differences, it is shown that the semivariogram is semi-positively defined  $\gamma_Z(\mathbf{r}) \geq 0$ . But the reverse is not always the case, as a semi-positive defined function is not necessarily an admissible semivariogram.

In case of a statistically homogeneous field, if the spatial dependence is isotropic, the semivariogram is determined by two parameters: the *sill* and the *correlation length*. The value of the semivariogram for long distances  $\mathbf{r}$  tends asymptotically to a sill equal to the variance  $\sigma_Z^2$  of the random field. This property is based on  $\gamma_Z(\mathbf{r}) = \sigma_Z^2 - c_Z(\mathbf{r})$  and the fact that at large distance the value of the covariance function tends towards zero. The presence of important large distance trends means that the assumption of statistical homogeneity is not valid. Then the semivariogram does not converge towards a balance value, when the distance tends towards infinite (Hristopulos 2008).

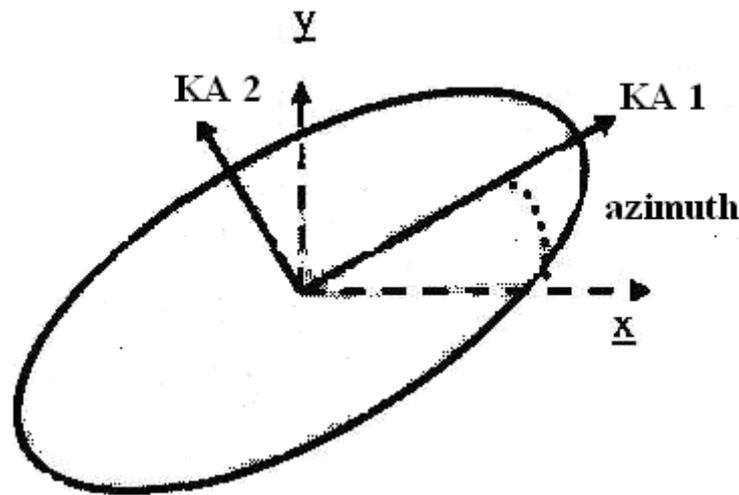
If correlation characteristics vary in different directions in space then the dependence is anisotropic. The two main types of anisotropy that are encountered in practice are geometrical and zone anisotropy. Geometrical anisotropy refers to cases when the semivariogram sill is independent of the direction, but the velocity approaching the sill, depends on the direction (Hohn 1999). In this case the semivariogram is expressed as

function  $\gamma_Z\left(\frac{r_1}{\xi_1}, \dots, \frac{r_d}{\xi_d}\right)$  of non-dimensional distances  $\frac{r_1}{\xi_1}, \dots, \frac{r_d}{\xi_d}$ , where  $\xi_1, \dots, \xi_d$  are the correlation lengths in the corresponding directions.

Zone anisotropy refers to the case where the sill depends on the spatial direction. Then the semivariogram can be expressed as the sum of resultant:  $\gamma_Z(\mathbf{r}) = \gamma_{Z,1}(r) + \gamma_{Z,2}(\hat{\mathbf{r}})$ . In this equation  $\gamma_{Z,1}(r)$ , where  $r = \|\mathbf{r}\|$ , describes an isotropic dependence while  $\gamma_{Z,2}(\hat{\mathbf{r}})$  describes the anisotropic dependence between the sill and the direction of the unit vector  $\hat{\mathbf{r}}$ .

In the case of geometrical anisotropy more than one correlation lengths are required  $\xi_1, \dots, \xi_d$ . Some of them, but not all, may be equal to each other. Therefore additional

parameters are required for the determination of the semivariogram's anisotropy. In a two-dimensional system  $\xi_x$  and  $\xi_y$ , correspond to the correlation lengths along the main axis, the anisotropy parameters are: (1) the ratio  $\rho_{y/x} \equiv \xi_y/\xi_x$ , which is called anisotropy ratio, (2) the orientation angle, which defines the orientation of the main anisotropy axis, in relation to the Cartesian coordinate system.



**Figure 2.1** Presentation of the main axis system (KA1, KA2) in relation to the coordinate system  $x, y$ . The ellipsis corresponds to the semivariogram direction (after (Hristopulos 2008)).

In order to understand the meaning of the orientation angle, the ellipse is defined as the geometrical location of points  $(r_x, r_y)$ , where the value of the semivariogram is constant. The elliptical shape is used since this happens for different semivariogram models, such as exponential and Gaussian anisotropic semivariograms. The orientation angle is the angle between KA 1 axis of the ellipsis with the horizontal axis of the coordinate system (Figure 2.1), (Hristopulos 2008).

The semivariogram generally increases, but not necessarily linearly, with the distance between the points, while on the contrary the correlation function decreases. This is due to the fact that the correlation function describes the dependence between the field values in two different points in space and their dependence decreases in larger distances. On the contrary, the semivariogram measures the difference between field values as a function of their distance. Therefore, semivariogram values increase when the distance increases (Deutsch and Journel 1992).

For statistically homogeneous fields, the two functions are equivalent which means that they have the same information in different form. However there are cases of random fields where semivariogram is a function of the distance between two points only, while the correlation function depends both on the distance and the specific location of the points in space (Hristopulos 2008).

Widely used semivariogram models, which can be also used in practical applications, are the exponential, the Gaussian, the spherical, the powerlaw and the nugget effect. The exponential model characterizes distributions with sharp spatial variations, opponent to the Gaussian model that characterizes more smooth variations. The powerlaw model corresponds to dependence with long distance spatial range and the nugget effect to variations which take place in distances smaller than the resolution that the sample allows. Another way of determining the spatial dependence of a random field, which is presented in this thesis, is that of the method of Spartan variogram family (Hristopulos 2008).

### 2.1.9 Semivariogram estimation

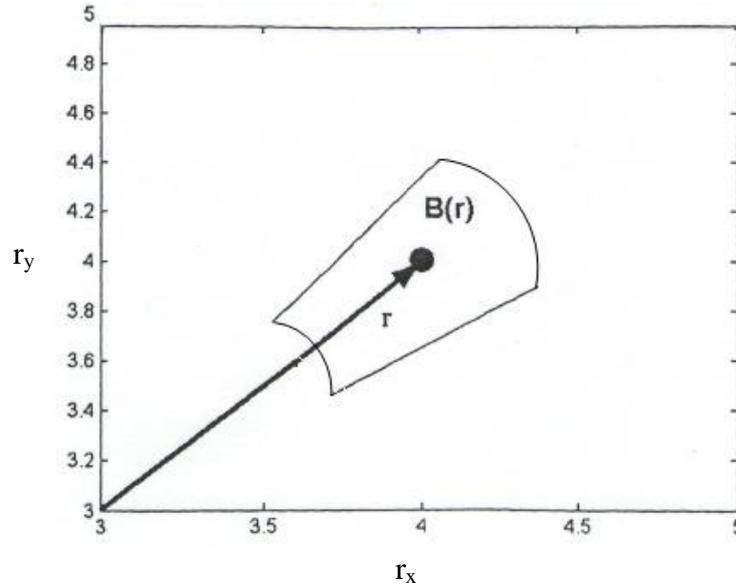
The main mathematical tool in geostatistical modeling is the semivariogram which expresses the spatial dependence between neighboring observations. In the case of geographical distributions and distribution of environmental variables, where the available data are limited to a sole sample, it is attempted to determine an estimation of the real semivariogram through it. This estimation is called sampled or experimental semivariogram and is calculated based on the values of the sample. The Matheron method-of-moments estimator of the semivariogram is given by (Isaaks and Srivastava 1989, Deutsch and Journel 1992):

$$\hat{\gamma}_Z(\mathbf{r}_k) = \frac{1}{2N(\mathbf{r}_k)} \sum_{i,j=1}^{N(\mathbf{r}_k)} \left\{ [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 \right\} \vartheta_{ij}(\mathbf{r}_k), \quad (k = 1, \dots, N_c), \quad (2.10)$$

$$\vartheta_{ij}(\mathbf{r}_k) = \left\langle \begin{array}{l} 1, \mathbf{s}_i - \mathbf{s}_j \in B(\mathbf{r}_k) \\ 0, \text{otherwise} \end{array} \right\rangle$$

- The class function  $\vartheta_{ij}(\mathbf{r}_k)$  defines different classes of distance vectors, choosing the vectors that correspond to a closed region  $B(\mathbf{r}_k)$  (

- Figure 2.2) around vector  $\mathbf{r}_k$
- Variable  $N(\mathbf{r}_k)$  is equal to the number of point pairs inside class  $B(\mathbf{r}_k)$
- The sample semivariogram is defined for a discrete and finite set of distances  $\mathbf{r}_k, (k = 1, \dots, N_c)$  the number of which is equal to the number of classes  $N_c$ .



**Figure 2.2** Schematic figure of the region  $B(\mathbf{r})$  around the distance vector (Hristopulos 2008).

The empirical semivariogram,  $\hat{\gamma}_Z(\mathbf{r}_k)$  is defined as the average square difference of the field values between points separated by the lag vector  $\mathbf{r}_k$ . More precisely, this calculation determines a value for the sample semivariogram for every  $\mathbf{r}_k$ , based on the mean value of differences  $[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2$  in all pairs of points, the distance vector of which belongs in  $B(\mathbf{r}_k)$  region.  $\hat{\gamma}_Z(\mathbf{r}_k)$  is a good estimator of the real  $\gamma_Z(\mathbf{r}_k)$  when the mean value of differences in  $\mathbf{r}_k$  class approaches with accuracy, the mean value  $E[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{r}_k)]^2$  (Hristopulos 2008). The latter is true when the Ergodic assumption applies, which allows the switch between the stochastic and the sample mean. In semivariogram calculation the Ergodic property is valid when the following conditions occur: the field of differences  $Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{r}_k)$  is statistically homogeneous, the number of pairs in each class is large enough so the sample mean of the square difference is determined with good statistical accuracy and the number

of classes is large enough so the dense approach of the semivariogram variations as a function of the distance is allowed. After the sampled semivariogram is calculated, it is adjusted to a theoretical model which allows the calculation of the semivariogram in every distance. This can be achieved using e.g., the least square method, from which the optimal values for parameters  $\zeta$  and  $\sigma_z^2$  of the theoretical model can be calculated. The variable  $\zeta$  denotes the characteristic (correlation) length and  $\sigma_z^2$  the variance (sill) (Hristopulos 2008).

The theoretical model is needed for the estimation of field values in points where measurements are not available. Next, in order to accept the semivariogram and use it for geostatistical analysis, it is tested according to the semivariogram acceptance conditions. A semivariogram is acceptable if it is conditionally negative definite. This means that for any linear coefficients  $\lambda_\alpha$  that satisfies the equation

$$\sum_{\alpha=1}^N \lambda_\alpha = 0 \text{ the following inequality must apply,}$$

$$-\sum_{\alpha=1}^N \sum_{\beta=1}^N \lambda_\alpha \lambda_\beta \gamma_z(s_\alpha - s_\beta) \geq 0, \quad (2.11)$$

for any positive integer  $N$ . For a spatial homogeneous random field is simpler to check the acceptance of a semivariogram or covariance model using the function  $\sigma_z^2 - \gamma_z(\mathbf{r})$ . If the function  $\gamma_z(\mathbf{r})$  describes an acceptable semivariogram then the function  $c_z(\mathbf{r}) = \sigma_z^2 - \gamma_z(\mathbf{r})$  is an acceptable covariance function and vice-versa (Hristopulos 2008).

If anisotropic spatial dependence occurs the semivariogram should be calculated in different directions in space, in order to determine the main direction of the anisotropy. This requires the definition of classes not only according to the range but also according to the direction of the distance vector. Every class has a tolerance ( $2\delta r$ ) in terms of the range, as well as ( $2\delta\phi$ ) in terms of the direction angle of the distance vector, so as every class to include an adequate number of points. The semivariogram is usually calculated in terms of the directions North-South and East-West, while for the angular tolerance the values  $5^\circ$ ,  $10^\circ$ ,  $20^\circ$  and  $45^\circ$  are used (Goovaerts 1997). In this thesis except of the latter approach anisotropy is determined using a newly established method named Covariance Hessian identity (Chorti and Hristopulos 2008). Figure 2.3 presents the characteristics of a typical semivariogram.

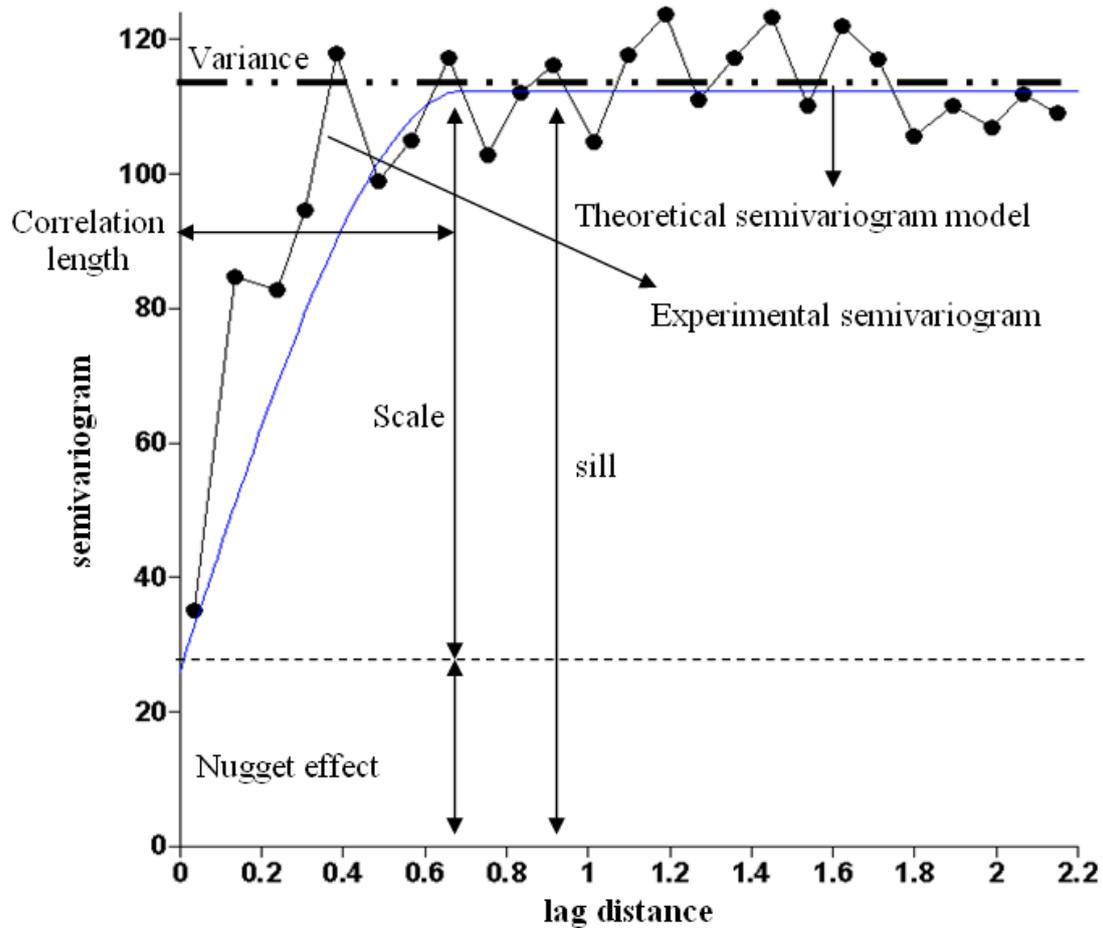


Figure 2.3 Presentation of typical semivariogram characteristics.

- The nugget effect quantifies the variance of the sampling error, as well as the small scale variance, e.g. the spatial variance in distances smaller than the distances between sampling points.
- The sill is the value that approaches asymptotically the experimental semivariogram.
- Scale is the difference between the sill and the nugget effect, and declares the variance of the correlated fluctuations.
- The correlation length is the distance in which the semivariogram almost (e.g. 95-97%) reaches the sill value.
- Variance is the mean squared deviation of every value of the sample from the mean value and is denoted with the horizontal dashed line in the figure.
- The experimental semivariogram represents the classes of pairs along with the corresponding sampled values of the semivariogram.

- The theoretical semivariogram model is a continuous theoretical line which is fitted to the experimental semivariogram.

If there are no distinct anisotropies, the omnidirectional empirical semivariogram  $\hat{\gamma}_Z(\mathbf{r})$ ,  $\mathbf{r} = \mathbf{r}_k$ , is estimated and then fitted to a theoretical model function  $\gamma_Z(\mathbf{r})$  (Deutsch and Journel 1992, Kitanidis 1997).

#### 2.1.10 Semivariogram models

Classical theoretical semivariogram models listed below include the spherical, Gaussian, exponential, power-law and linear functions (Goovaerts 1997, Lantuejoul 2002);  $\sigma_Z^2$  is the variance,  $|\mathbf{r}|$  is the Euclidean norm of the lag vector  $\mathbf{r}$ , and  $\xi$  is the characteristic length.

$$\text{Exponential: } \gamma_Z(\mathbf{r}) = \sigma_Z^2 \left[ 1 - \exp\left(-\frac{|\mathbf{r}|}{\xi}\right) \right] \quad (2.12)$$

$$\text{Gaussian: } \gamma_Z(\mathbf{r}) = \sigma_Z^2 \left[ 1 - \exp\left(-\frac{\mathbf{r}^2}{\xi^2}\right) \right] \quad (2.13)$$

$$\text{Spherical: } \gamma_Z(\mathbf{r}) = \sigma_Z^2 \left[ 1.5 \frac{|\mathbf{r}|}{\xi} - 0.5 \left(\frac{|\mathbf{r}|}{\xi}\right)^3 \right] \theta(\xi - |\mathbf{r}|) \quad (2.14)$$

if  $\xi - |\mathbf{r}| < 0$ ,  $\theta = 0$ , else if  $\xi - |\mathbf{r}| > 0$ ,  $\theta = 1$

$$\text{Power-law: } \gamma_Z(\mathbf{r}) = c |\mathbf{r}|^{2H}, \quad 0 < H < 1 \quad (2.15)$$

$c$  is the coefficient and  $H$  the Hurst exponent.

$$\text{Linear: } \gamma_Z(\mathbf{r}) = c |\mathbf{r}| \quad (2.16)$$

Equations above define the isotropic versions of the models. These involve at most two parameters, i.e., the variance and correlation length for exponential,

Gaussian and spherical models,  $c$  and  $H$  for the power-law model and  $c$  for the linear model. Below we review two semivariogram models that offer increased parameter flexibility.

### 2.1.11 Matérn model

This covariance family includes, in addition to the variance and correlation length a smoothness parameter  $\nu$ , which controls the continuity and differentiability of the random field, and thus also the short-distance behaviour of  $\gamma(\mathbf{r})$ , which has greater impact on interpolation than medium to large distance dependence. The Matérn semivariogram model (Matérn 1960, Stein 1999, Pardo-Iguzquiza and Chica-Olmo 2008) is defined as:

$$\gamma_z(\mathbf{r}) = \sigma_z^2 \left\{ 1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{|\mathbf{r}|}{\xi} \right)^\nu K_\nu \left( \frac{|\mathbf{r}|}{\xi} \right) \right\}, \quad (2.17)$$

where  $\sigma_z^2 > 0$  is the variance,  $\xi > 0$  is the characteristic length,  $\nu > 0$  is the smoothness parameter,  $\Gamma(\cdot)$  is the gamma function, and  $K_\nu(\cdot)$  is the modified Bessel function of the second kind of order  $\nu$ , and  $|\mathbf{r}|$  is the Euclidean norm of vector  $\mathbf{r}$ . For  $\nu=0.5$  the exponential model is recovered, whereas the Gaussian model is obtained at the limit as  $\nu$  tends to infinity. The case  $\nu=1$  was introduced by Whittle (1954). The Matérn model has been applied to different research fields including hydrology e.g., (Rodriguez-Iturbe and Mejia 1974, Zimmermann *et al.* 2008).

### 2.1.12 Spartan model

Spartan Spatial Random Fields (SSRFs) are a recently proposed geostatistical model (Hristopulos 2002, Hristopulos 2003b) with applications in environmental risk assessment (Elogne *et al.* 2008) and atmospheric environment (Žukovič and Hristopulos 2008). SSRFs are generalized Gibbs random fields, equipped with a coarse-graining kernel that acts as a low-pass filter for the fluctuations. The term Spartan indicates parametrically compact model families that involve a small number of parameters. These random fields are defined by means of physically motivated spatial interactions between the field values.

In general, a spatial random field (SRF)  $Z(\mathbf{s})$  representing the measurements can be expressed as:

$$Z(\mathbf{s}) = Z'(\mathbf{s}) + m_z(\mathbf{s}) + e(\mathbf{s}), \quad (2.18)$$

where  $e(\mathbf{s})$  is a zero-mean measurement noise process, assumed to be homogeneous over the domain of interest,  $Z'(\mathbf{s})$  is a correlated fluctuation SRF, and  $m_z(\mathbf{s})$  is a deterministic trend function. The trend is a non-stationary component representing large-scale, deterministic variations, which presumably correspond to the ensemble average of the SRF,

$$m_z(\mathbf{s}) = E[Z'(\mathbf{s})]. \quad (2.19)$$

SSRFs are determined from a probability density function in terms of a spatial random field (SRF)  $Z'(\mathbf{s})$ . The probability density function contains information for spatial dependence. In general, the probability density function SSRF can be expressed with the following equation:

$$f_x[Z'(\mathbf{s})] = \mathcal{Z}^{-1} \exp\{-H[Z'(\mathbf{s})]\}, \quad (2.20)$$

$$\mathcal{Z} = \sum_{Z'(\mathbf{s})} \exp\{-H[Z'(\mathbf{s})]\}, \quad (2.21)$$

is a normalization constant which ensures the basic theorem of probability, (i.e. that the *sum of probabilities* of a SRF is equal to 1).  $H[Z'(\mathbf{s})]$  is an energy functional of spatial dependence which expresses the interdependence of SRF data values  $Z'(\mathbf{s})$  between different locations. Therefore, SSRFs belong in the family of Gibbs random fields (Hristopulos 2003b). The Gibbs property stems from the fact that the joint probability density function of SSRFs is expressed in terms of an energy functional i.e.  $H[Z'(\mathbf{s})]$ . Use of an energy functional containing terms with a clear physical interpretation permits inference of the model parameters based on matching respective sample constraints with their ensemble values (Hristopulos and Elogne 2007). Thus,

the spatial continuity properties can be determined without estimation of the experimental variogram.

The SSRFs provide a new class of generalized covariance functions, which are by construction positive definite for an explicitly specified range of parameter values (Hristopulos 2003b, Hristopulos and Elogne 2007). Fluctuation-gradient-curvature (FGC) SSRFs model have an energy functional that involves the squares of the fluctuations, the gradient and the curvature of the field, see Eq. (2.22) below. This class provides covariance functions with four parameters that give considerable flexibility. The SSRF covariance functions can be used for spatial interpolation with the classical kriging estimators as well as with new spatial predictors (Elogne *et al.* 2008, Hristopulos and Elogne 2009).

Herein we use this new class of covariance function for the first time in groundwater hydrology in association with kriging for spatial interpolation of the groundwater level. For kriging applications, the estimation of the spatial dependence structure (semivariogram or covariance function) is a crucial step.

The isotropic FGC-SSRF functional is given by the following equation:

$$H[Z'(\mathbf{s})] = \frac{1}{2\eta_0\xi^d} \int d\mathbf{s} \left[ \{Z'(\mathbf{s})\}^2 + \eta_1\xi^2 \{\nabla Z'(\mathbf{s})\}^2 + \xi^4 \{\nabla^2 Z'(\mathbf{s})\}^2 \right]. \quad (2.22)$$

The FGC model involves the parameters  $\eta_0$ ,  $\eta_1$ ,  $\xi$  and  $k_m$ . The scale coefficient  $\eta_0$  determines the overall scale of the variance; the scale factor is proportional to the square of the regionalized variable, i.e. the groundwater level, and assumes the variable's units. The shape coefficient  $\eta_1$  is dimensionless and determines the shape of the covariance function in connection with  $\xi$  and  $k_m$ . The characteristic length  $\xi$  has dimensions of length and determines the range of spatial dependence. Finally, the wavevector  $k_m$  has units of inverse length and determines the bandwidth of the covariance spectral density. If the latter is band-limited,  $k_m$  represents the band cutoff and is related to the resolution length scale by means of  $k_m\lambda \approx 1$ .

The SSRF covariance models derived from the above energy functional are determined by the parameters  $\boldsymbol{\theta} = (\eta_0, \eta_1, \xi, k_m)$ . Spartan (SP) covariance and semivariogram functions were introduced in (Hristopulos 2003b) and have been applied to various environmental data sets (Elogne *et al.* 2008, Elogne and

Hristopulos 2008, Hristopulos and Elogne 2009). Herein, we apply this family of functions for the first time in hydrological data.

The SSRF family includes four-parameter functions (Hristopulos 2003b, Hristopulos and Elogne 2007). The Spartan covariance in any dimension  $d$  is expressed using the spectral representation as follows

$$C_Z(\mathbf{r};\boldsymbol{\theta}) = \frac{\eta_0 \xi \|\mathbf{r}\|^{1-d/2}}{(2\pi)^{d/2}} \int_0^{k_m} d\omega \frac{\omega^{d/2} J_{d/2-1}(\|\mathbf{r}\|\omega)}{1 + \eta_1 (\omega\xi)^2 + (\omega\xi)^4}, \quad (2.23)$$

where  $J_{d/2-1}(x)$  is the Bessel function of the first kind of order zero and  $\boldsymbol{\theta} = (\eta_0, \eta_1, \xi, k_m)$  are the model parameters. The Spartan semivariogram is given by  $\gamma_Z(\mathbf{r};\boldsymbol{\theta}) = C_Z(0;\boldsymbol{\theta}) - C_Z(\mathbf{r};\boldsymbol{\theta})$ . The scale parameter  $\eta_0$  determines the variance,  $\xi$  is the characteristic length,  $k_m$  represents the wavenumber cutoff (band limit in Fourier space), and the dimensionless stiffness coefficient  $\eta_1$  determines the shape of the covariance function in connection with  $k_m$  and  $\xi$  (Elogne *et al.* 2008). In  $d=1,3$  explicit expressions for the Spartan covariance are possible at the asymptotic limit  $k_m \rightarrow \infty$  (Hristopulos and Elogne 2007).

The Spartan covariance function of Eq. (2.23) in  $d=3$  dimensions are expressed as follows:

$$C_Z(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} \frac{\eta_0 e^{-h\beta_2}}{2\pi\sqrt{|\eta_1^2 - 4|}} \left[ \frac{\sin(h\beta_1)}{h\beta_1} \right], & \text{for } |\eta_1| < 2, \sigma_z^2 = \frac{\eta_0}{2\pi\sqrt{|\eta_1^2 - 4|}} \\ \frac{\eta_0 e^{-h}}{8\pi}, & \text{for } \eta_1 = 2, \sigma_z^2 = \frac{\eta_0}{8\pi} \\ \frac{\eta_0 (e^{-h\omega_1} - e^{-h\omega_2})}{4\pi(\omega_2 - \omega_1) h\sqrt{|\eta_1^2 - 4|}}, & \text{for } \eta_1 > 2, \sigma_z^2 = \frac{\eta_0}{4\pi\sqrt{|\eta_1^2 - 4|}} \end{cases}, \quad (2.24)$$

$$\omega_{1,2} = (|\eta_1 \mp \Delta|/2)^{1/2}, \quad (2.25)$$

$$\beta_{1,2} = |2 \mp \eta_1|^{1/2}/2. \quad (2.26)$$

In the above  $\Delta = |\eta_1^2 - 4|^{1/2}$ ,  $\omega_{1,2}$  and  $\beta_2$  are dimensionless damping coefficients,  $\beta_1$  is a dimensionless wave-number,  $\xi$  is a characteristic length,  $\|\mathbf{h}\| = \|\mathbf{r}\|/\xi$  is the normalized lag vector,  $h = \|\mathbf{h}\|$  its Euclidean norm and  $\sigma_z^2$  is the variance. The exponential covariance is recovered for  $\eta_1 = 2$ , while for  $|\eta_1| < 2$  the product of the exponential and hole-effect model is obtained. A covariance function that is permissible in *three* spatial dimensions is also permissible in two dimensions (Christakos 1991). Hence, (2.24) can be used in two dimensions, albeit it does not correspond to the FGC-SSRF two-dimensional covariance (Hristopulos and Elogne 2007).

### 2.1.13 Parameters inference

The Spartan parameters can be estimated by fitting the SSRF semivariogram to the empirical semivariogram estimator. A different approach is based on the modified method of moments, in which stochastic constraints are matched with corresponding sample constraints (Elogne *et al.* 2008, Žukovič and Hristopulos 2008, Žukovič and Hristopulos 2009). The constraints are motivated by the terms in the energy functional (2.22); the square of the fluctuations, the square gradient, and the square curvature are used to construct both the sample and the stochastic constraints. The latter approach is not investigated herein, because the focus of this thesis is on kriging interpolation techniques.

There is no universally accepted method for fitting the empirical semivariogram to a theoretical model. For each of the above theoretical models discussed above, we determine the optimal semivariogram parameters using the least squares method. Methods used include least-squares fits, weighted least squares, generalized least squares, maximum likelihood, and even empirical “fitting by eye” (Wackernagel 2003, Olea 2006). We implemented least-squares fitting by means of the «fminsearch» Matlab<sup>®</sup> function which is based on the Nelder-Mead minimization algorithm (Press *et al.* 1992). The selection of the “optimal semivariogram model” is based on the results of leave-one-out cross validation (see section 2.5).

### 2.1.14 Spatial estimation

The determination of the spatial dependence, as well as of the trend and the fluctuations of the field values leads in two basic procedures of geostatistics; the spatial estimation and the simulation. Both procedures help in the representation of a random field in points where not exact values exist, based on available information (e.g. measurements in neighboring points, hydrogeological data). The available information is used in order to impose statistical limitations. Using statistical spatial dependence patterns (semivariograms) the unknown values are defined based on their correlation. The repetitiveness of this procedure in all points of the computational grid allows the mapping of an entire area (Hristopulos 2008).

The simulation process aims to create many of the possible states of the field, which are in accordance with existing statistical restrictions derived by the experimental sample e.g. simulated states with the same mean value, standard deviation and semivariogram with the one calculated using samples. Therefore, the simulation's aims in the creation of many alternative scenarios, which are possible based on existing measurements (Hristopulos 2008).

The term spatial and/or temporal estimate includes all the mathematical procedures that allow the calculation of field values where measurements of a property do not exist. The estimate can be local, if it is referred to a point in space-time or global, if it aims to calculate a characteristic value that describes an entire region. The spatial and/or temporal estimate of a field presupposes the existence of spatial and/or temporal dependence, so that the field value at each point is "influenced" by the neighboring field values. This interdependence allows estimation of a variable where measurements do not exist based on the neighboring measured points. In many cases, the final objective is to estimate the field over a set of points instead of a single one. Various methods of spatial estimation (interpolation) exist that are based on similar principles. The main idea is that the value at the estimation point is given by a linear or nonlinear combination, of the neighboring values. The estimate results from the optimization of a statistical measure, e.g. maximization of probability or minimization of the mean square estimation error. The most popular methods are based on linear interpolation in conjunction with the minimization of the mean square estimate error. This set of methods is known as "kriging" (Goovaerts 1997, Kitanidis 1997, Hristopulos 2003a).

The need of variables estimation at points where no measurements are available is not new. Statistical scientists, mining engineers, oil engineers, hydrologists and geologists who dealt with the problem developed the science of geostatistics. Application areas of geostatistics nowadays include: the analysis of ore deposits (e.g. estimate of extent, depth and quantification of total content) e.g., (Journel and Huijbregts 1978, De-Vitry *et al.* 2010), oceanography (mapping of ocean bed, waves height analysis) e.g., (Özger and Şen 2007), the morphological analysis of natural and technological non-homogeneous (e.g. porous) materials e.g., (Sahimi 2011), the mapping of pollutant concentrations in various environmental means (air, subsoil, surface-underground water resources) e.g., (Goovaerts 1997, Webster and Oliver 2001), the topographic analysis and the geographic information systems (GIS) e.g., (Burrough 2001), the spatiotemporal analysis of rainfall data and of rainfalls in regions with insufficient monitoring stations e.g., (Ly *et al.* 2011), the determination of geological and hydrogeological variables (e.g. subsoil type, hydraulic conductivity, porosity, storativity, evapo-transpiration) e.g., (Kitanidis 1997, Hengl 2007), the environmental and human health risk assessment (e.g. estimate of pollutant concentration, determination of probabilities of exceeding the critical limits) e.g., (Goovaerts 1997, Christakos and Hristopulos 1998), the spatial and/or temporal estimation of hydraulic head of aquifers e.g., (Ahmed 2007).

## 2.2 Spatial interpolation

Geostatistics is based on the work of Kolmogorov (1941) in atmospheric turbulence. He used the structure function (equivalent to the variogram) to represent spatial correlations and to develop optimal interpolation. Later, Matérn developed the family of spatial covariance functions that bear his name (Matérn 1960). His functions are equivalent with those developed by Jowett (1955). The geostatistical method called kriging, the most applied geostatistical method to date, was introduced and established by Krige (Krige 1951, Krige 1966) and Matheron (Matheron 1963) for applications in mining engineering. Since then, kriging has been applied to several other fields of research, such as geology (Davis 1973, Journel and Huijbregts 1978), petroleum engineering (Hohn 1999), hydrogeology (Kitanidis 1997), hydrology, meteorology and soil science (Webster and Oliver 2001, Atkinson and Lloyd 2010). The first application of kriging in groundwater hydrology was by Delhomme (1974). Since then, many studies applied kriging to the interpolation of groundwater levels,

e.g. (Delhomme 1978, Gambolati and Volpi 1979a, b, Sophocleous *et al.* 1982, Aboufirassi and Marino 1983, Pucci and Murashige 1987, Hoeksema *et al.* 1989, Desbarats *et al.* 2002, Ahmadi and Sedghamiz 2007, Kumar 2007, Ahmadi and Sedghamiz 2008, Rivest *et al.* 2008, Nikroo *et al.* 2009).

In general interpolation methods routinely used for groundwater level mapping include deterministic methods such as inverse distance weighting (IDW) (Gambolati and Volpi 1979b, Philip and Watson 1986, Rouhani 1986, Buchanan and Triantafilis 2009, Sun *et al.* 2009) and stochastic methods such as Ordinary kriging (OK) and Universal kriging (UK). Such methods are incorporated in various commercial software packages e.g., mapping software: Arc-View (GIS), Surfer mapping system (Golden software), groundwater modeling software: Visual Modflow, Princeton transport code (PTC), Feflow subsurface flow model.

Deterministic interpolation methods use closed-form mathematical formulas (IDW) or the solution of a linear system of equations (Minimum Curvature) to interpolate the data. The weights assigned to each sample value depend only on the distance between the sample point and the location of the interpolated point. Deterministic methods are categorized as global and local: Global methods use the entire dataset for prediction at each point, while local methods use data in a neighborhood around the interpolation point. Deterministic methods can be either exact or inexact interpolators (Webster and Oliver 2001). Finally, they do not generate measures of estimate uncertainty.

Stochastic methods employ the spatial correlations between values at neighboring points. The most widely used stochastic method is kriging (Krige 1951, Matheron 1963, 1971). The kriging methodology comprises a family of interpolators. The interpolators most commonly used in hydrosciences are Ordinary Kriging (OK) and Universal Kriging (UK). A recently proposed variation of the kriging algorithm is kriging with Delaunay triangulation (DK) (Hessami *et al.* 2001).

Kriging is characterized as the best linear unbiased estimator (BLUE). The kriging estimator is a weighted linear function of the data. The linear weights follow from the unbiasedness constraint (i.e., zero mean estimation error) and the minimum square error condition. The resulting system of linear equations is solved to determine the estimator's weights. The coefficients of the equations depend on the model semivariogram, which is obtained by fitting the empirical semivariogram to theoretical models or by means of the maximum likelihood estimation method

(Kitanidis 1997, Ahmed 2007). The semivariogram measures the degree of spatial correlation as a function of distance and/or direction between data points. The semivariogram determines the kriging weights and therefore controls the quality of the estimates (Mouser *et al.* 2005, Ahmed 2007). If the semivariogram is perfectly known, kriging is the best linear unbiased estimator (BLUE). An advantage of kriging compared to deterministic approaches is that it allows the estimation of the interpolation error at unmeasured points (Deutsch and Journel 1992). In addition, in the absence of a nugget (e.g., measurement error), kriging is an exact interpolator at measurement points (Delhomme 1974, Ahmed 2007). Optimal kriging results are obtained if the probability distribution of the data is normal and stationary in space (spatially homogeneous). Kriging is computationally intensive when applied to large data sets (Webster and Oliver 2001), but the computational complexity is not a problem for sparsely sampled areas.

Ordinary Kriging (OK) bases its estimates at unsampled locations only on the sampled primary variable. OK interpolation is widely used to determine the spatial variability of groundwater levels in hydrological basins e.g., (Olea and Davis 1999, Prakash and Singh 2000, Desbarats *et al.* 2002, Theodossiou and Latinopoulos 2006, Ahmadi and Sedghamiz 2007, Abedini *et al.* 2008, Ahmadi and Sedghamiz 2008, Yang *et al.* 2008, Kholghi and Hosseini 2009, Nikroo *et al.* 2009, Sun *et al.* 2009, Taany *et al.* 2009, Dash *et al.* 2010). OK was also used to predict the piezometric head in West Texas and New Mexico based on implementing clustered piezometric data (Abedini *et al.* 2008). In addition, the design, evaluation and optimization of groundwater level monitoring networks were performed by applying OK (Olea and Davis 1999, Prakash and Singh 2000, Theodossiou and Latinopoulos 2006, Yang *et al.* 2008). Evaluation of the performance and interpolation errors of OK in the estimation of water level elevation can be achieved by means of leave-one-out cross validation (Olea 1999).

OK is not optimal for non-stationary data. The use of a linear drift term improves the accuracy of the interpolated head field if a regional gradient is present (Delhomme 1978, Aboufirassi and Marino 1983). Universal Kriging (UK) also has been used to estimate the groundwater level e.g., (Delhomme 1978, Sophocleous *et al.* 1982, Aboufirassi and Marino 1983, Sophocleous 1983, Pucci and Murashige 1987, Kumar *et al.* 2005, Ahmadi and Sedghamiz 2007, Brus and Heuvelink 2007, Gundogdu and Guney 2007, Kumar 2007, Sun *et al.* 2009). Near extracting or

injecting wells a point logarithmic component is added to the drift to account for the drawdown caused by the pumping well. This approach is applicable if analytical solutions for the aquifer response are available (Tonkin and Larson 2002, Rivest *et al.* 2008).

Auxiliary information can be included in the interpolation as a drift term, usually modeled by polynomial functions of the space coordinates, rainfall, or surface elevation based on a Digital Elevation Model (DEM). The use of auxiliary variables in general improves the accuracy of kriging estimation. Easily measurable secondary variables can also reduce the number of “expensive” observations (Knotters *et al.* 1995). The auxiliary information can be incorporated using the co-Kriging (CoK) method, which utilizes secondary variables in the covariance structure. Various researchers (Hoeksema *et al.* 1989, Deutsch and Journel 1992, Goovaerts 1997) used CoK with ground surface elevation as a secondary variable to construct groundwater level maps that improved the OK predictions. The main disadvantage of CoK is the need to model coregionalisation, which requires the inference of direct and cross covariance functions (Journel and Huijbregts 1978). CoK also becomes cumbersome and time-consuming if many secondary variables are involved (Deutsch and Journel 1992).

Alternatively, Residual Kriging (RK) and Kriging with External Drift (KED), originally described and applied in hydrological problems (Delhomme 1978, Volpi and Gambolati 1978, Gambolati and Volpi 1979a, b), embody secondary information in the drift term. KED and RK are practically equivalent but differ in the methodological steps used (Hengl *et al.* 2003, Hengl 2007). Residual Kriging is also known as Regression Kriging and it was developed and applied in the hydrosociences by Delhomme (1974, 1978) and (Ahmed and De Marsily 1987). Odeh *et al.* (1994, 1995) named it “Regression Kriging”, while (Goovaerts 1999) uses the term Kriging after detrending (Hengl *et al.* 2003).

KED assumes that the expectation of the primary variable is a linear combination of secondary variables (Deutsch and Journel 1992, Wackernagel 2003), while OK assumes the expectation to be constant (Rivest *et al.* 2008). In the case of KED (it has similar methodology to UK), the Kriging covariance matrix of residuals is extended with the auxiliary predictors (Kitanidis 1997, Webster and Oliver 2001). KED was applied for the interpolation of water table elevation by various researchers. Beven and Kirkby (1979) expressed the water table depth as a linear function of the

topographic index. Desbarats et al. (2002) applied KED to the interpolation of water table elevation using two deterministic trend models that include: a) the topographic elevation and b) the topographic index. Rivest et al. (2008) approximated the external drift using numerical solutions for the hydraulic head field obtained by means of finite elements based on a conceptual model that included hydrogeological parameters estimates, geology and boundary conditions. KED and collocated CoK incorporated topography as secondary information in Boezio et al. (2006a, b). Both methodologies improved the quality of the water table elevation maps compared to OK. Another approach combines KED with the regionalized autoregressive exogenous variable (RARX) model with precipitation surplus as the exogenous variable, and with DEM data as secondary variables (Knotters and Bierkens 2002).

Neuman and Jacobsen (Neuman and Jacobson 1984) used RK to estimate the hydraulic head in a catchment by approximating the trend function with space polynomials. RK with rainfall data as secondary variable was also applied to examine the influence of land use/cover change on the temporal and spatial variability of groundwater levels (Moukana and Koike 2008). Nikroo et al. (2009) predicted water table elevation by different (SK, OK and RK) kriging methods and trend functions, including auxiliary information from ground surface elevation and slope as well as draining rates.

Other researches that include kriging interpolation techniques in extended comparison studies regarding different interpolation methods applied to groundwater level data along with other hydrological variables e.g., (Subyani and Sen 1989, Kholghi and Hosseini 2009, Sun *et al.* 2009).

It can be proved mathematically that KED and RK are practically equivalent, although the methodological steps differ (Hengl *et al.* 2003, Hengl *et al.* 2007). The KED estimator is analyzed into a generalized regression of the primary variable with the secondary variables, followed by SK or OK of the regression de-trended residuals; in the Kriging equation system the covariance matrix is extended with the auxiliary predictors. A limitation of KED is the potential instability of the extended matrix if the covariate varies irregularly in space (Goovaerts 1997). In RK the drift model coefficients are first determined by regression, and the residuals are then interpolated using OK and finally added to the drift model. The main advantage of RK over KED is that it explicitly separates the trend estimation from the interpolation of the residuals, thus enabling the use of advanced regression methods (Hengl *et al.* 2003,

Hengl 2007). In addition, RK permits separate interpretation of the interpolated components and straightforward inclusion of multiple sources of external information that may compensate for small sample sizes (Alsamamra *et al.* 2009). For the reasons explained previously, RK is chosen over KED in this thesis.

### 2.3 Overview of geostatistical methodology

In the following, we will assume that the hydraulic head is represented by a spatial random field (SRF), which herein will be in generally denoted by  $Z(\mathbf{s}, \omega)$ , where  $\omega$  is the state index used to denote that  $Z(\mathbf{s}, \omega)$  is a realization from an ensemble of possible states (to be omitted for brevity). The sampled field at the measurement points will be denoted by  $Z(\mathbf{s} \in S)$ , where  $S$  is the set of sampling points with cardinal number  $N$ . The values of the SRF in a given state will be denoted by lower-case letters. The target is to derive estimates,  $\hat{Z}(\mathbf{s} \in P)$  of the head at the prediction set points,  $P$  that lie on a rectangular grid that covers the basin. Therefore  $\mathbf{s}_i, i = 1, \dots, N$  denote the sampling points,  $z(\mathbf{s}_i)$  are the head values (in masl) at these points, and  $\mathbf{s}_0$  denotes an estimation point, which is assumed to lie inside the convex hull of the sampling network. For mapping purposes, it is assumed that  $\mathbf{s}_0$  moves sequentially through all the nodes of the mapping grid.

We examine linear interpolation methods for mapping spatial and/or temporal groundwater level variability. In spatial linear interpolation methods, it holds that

$$\hat{z}(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i z(\mathbf{s}_i), \quad (2.27)$$

where  $\mathbb{S}_0$  is the set of sampling points in the search neighborhood of  $\mathbf{s}_0$ . The neighborhood is empirically chosen so as to optimize the cross validation measures.

For spatial interpolation we initially use two deterministic (IDW, MC) and three stochastic (OK, UK, DK) methods (chapter 4). Then, we use OK (chapter 5) and RK (chapter 6) methods in combination with non-linear normalizing transformations. In the first approach, we apply a normalizing transformation  $g(\cdot)$  to the data. Then, we use OK to predict the transformed field  $Y(\mathbf{s}) = g(Z(\mathbf{s}))$ , and we back-transform the predictions to obtain head estimates. Several methods can be used to handle non-

Gaussianity in the data. We applied the Box-Cox transformation, TGK, GA, and a MBC transform. We review these methods in the following chapters.

In the second approach, we introduce a trend model  $m_z(\mathbf{s})$  that captures local features and physical laws. Since the fluctuation SRF,  $Z'(\mathbf{s}) = Z(\mathbf{s}) - m_z(\mathbf{s})$ , is non-Gaussian, we apply a transformation  $g(\cdot)$  to obtain a normalized SRF,  $Y(\mathbf{s}) = g(Z'(\mathbf{s}))$ , estimate its experimental semivariogram and fit it to theoretical models. Next, we estimate the Gaussian field  $\hat{Y}(\mathbf{s} \in P)$  at the prediction points using OK. Finally, we retrieve head estimates from  $\hat{Y}(\mathbf{s} \in P)$  by applying the back-transformation and adding the trend. We use leave-one-out cross-validation analysis to determine the optimal spatial model and to assess the accuracy of the interpolated head field (Ahmed 2007).

Chapters 4, 5 and 6 are based on the above overview and focus on the spatial interpolation of groundwater level. The overview of spatiotemporal geostatistical methodology is presented in section 7.

In the thesis we opt to keep the interpolation estimates within the convex hull of the sampling points. In principle we can estimate maps over the entire study domain (Figure 3.10); however, this is equivalent to extrapolation. Kriging can be used for extrapolation but the results outside the quadrilateral, determined from the sampling locations boundaries, are often less accurate and subject to higher uncertainty. In addition the semivariogram is determined by the measurements and expresses the spatial dependence of the measured points. In performing extrapolation, we accept that the semivariogram is valid outside the range of measurements. Therefore the estimates inside the quadrilateral are more accurate and precise than those outside.

## 2.4 Interpolation materials and methods

Interpolation is the process of estimating the data values in unvisited locations using known measured data values from neighbor points. The interpolation methods are divided in deterministic and stochastic. Deterministic methods provide no information regarding the possible estimation errors while stochastic methods provide probabilistic estimates (i.e. provide the variance of the estimates). Deterministic interpolation methods assign weights to each sample value depending only on the

distance between the sample point and the location of the interpolated point. On the other hand stochastic or geostatistical methods treat the observations dataset as an arbitrary realization of a stochastic process and employ the spatial correlations between the values at neighboring points in order to distribute the weights. In this section the theoretical background of the deterministic and stochastic interpolation methods used for the groundwater level spatial variability prediction in Mires basin, is explicitly presented.

#### 2.4.1 Inverse Distance Weight

The estimation with the IDW method is given by means of the equation

$$\hat{z}(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \left( \frac{d_{i,0}^{-n}}{\sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} d_{i,0}^{-n}} \right) z(\mathbf{s}_i), \quad (2.28)$$

where  $d_{i,0}$  is the distance between the estimation point and the sampling points, and  $n > 0$  is the power exponent; usually  $n = 2$  is used. IDW assigns larger weights to data closer to the estimation point  $\mathbf{s}_0$  than to more distant points. Higher values of  $n$  increase the impact of values near the interpolated point, while lower values of  $n$  imply more uniform weights. As it follows from (2.28) the weights add up to one. IDW is an exact and convex interpolation method (Hengl *et al.* 2007). In addition it is very fast, straightforward and computationally non-intensive (Webster and Oliver 2001). According to (2.28), as the distance of  $\mathbf{s}_i$  from  $\mathbf{s}_0$  increases, the respective weight is reduced. IDW's disadvantages are the arbitrary choice of the weighting function and the lack of an uncertainty measure (Webster and Oliver 2001).

#### 2.4.2 Minimum Curvature

MC interpolation is based on the minimization of the total square curvature of the surface  $z(\mathbf{s})$ , i.e.,  $\int d\mathbf{s} [\nabla^2 z(\mathbf{s})]^2$  subject to the data constraints. In MC, the interpolated surface can be viewed as a thin linear elastic plate pinned to the data values at the sampling points. The estimate is obtained by solving the biharmonic partial differential equation (Briggs 1974, Sandwell 1987), i.e.,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\left(\frac{\partial^2 z(\mathbf{s})}{\partial x^2} + \frac{\partial^2 z(\mathbf{s})}{\partial y^2}\right) = 0, \quad (2.29)$$

conditioned by the data values  $z(\mathbf{s}_i)$ . The interpolating function  $z(\mathbf{s})$  honours the observed data and tends to a planar surface as the distance between the interpolation point and the observations increases. Typical applications of MC include interpolating hydrocarbon (oil) depths (Cooke *et al.* 1993), interpolation of gravimetric and magnetometric geophysical data for mineral exploration (Mendonca and Silva 1995, Kay and Dimitrakopoulos 2000) and mapping the earth surface (Yilmaz 2007).

The MC method often suffers from oscillations due to the presence of outliers in the data or due to very large gradients. This problem can become important if the dataset is relatively small. The MC interpolation is based on the Green's function  $g_m$  of the biharmonic equation, which satisfies  $\nabla^4 g_m(\mathbf{s} - \mathbf{s}') = \delta(\mathbf{s} - \mathbf{s}')$  where  $\delta(\mathbf{s} - \mathbf{s}')$  is the Dirac delta function. The two dimensional (2D) Green's function is given by  $g_m(d) = d^2(\ln d - 1)$  (Sandwell 1987, Wessel 2009). The MC estimate is then expressed as follows:

$$\hat{z}(\mathbf{s}_0) = \sum_{i=1}^N w_i g_m(d_{i,0}). \quad (2.30)$$

The weights  $w_i$  are determined by solving the following linear system at the  $N$  data locations.

$$z(\mathbf{s}_i) = \sum_{j=1}^N w_j g_m(d_{i,j}), \quad (2.31)$$

where  $j = 1, \dots, N$  and  $d_{i,j}$  the distances between the sample points  $d_{i,j} = |\mathbf{s}_i - \mathbf{s}_j|$ .

### 2.4.3 Ordinary kriging interpolation

The term kriging is used for a suite of interpolation methods that are based on the principles of zero bias and minimum mean square error. Kriging estimates the

value of a process over an entire domain, over a finite-volume block or at a specific point  $\mathbf{s}_0$ . The estimates are formed by means of a linear combination of the data values. The summation is over the entire area or a restricted neighborhood centered at the estimation point. Kriging interpolation method also quantifies the estimation variance, and thus, the precision of the resulting estimates. The commonly used OK method is based on the following equations (Goovaerts 1997, Kitanidis 1997).

The OK method assumes that  $z(\mathbf{s})$  is a random function with a constant but unknown mean. The OK estimate  $\hat{z}(\mathbf{s}_0)$  at  $\mathbf{s}_0$  is calculated based on a weighted sum of the data

$$\hat{z}(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathcal{S}_0\}} \lambda_i z_i(\mathbf{s}_i). \quad (2.32)$$

The weights  $\lambda_i$  in (2.32) are obtained by minimizing the mean square estimation error conditionally on the zero-bias constraint (Cressie 1993), and they depend on the semivariogram model  $\gamma_z(\mathbf{r})$  (Deutsch and Journel 1992).

The kriging weights  $\lambda_i$  follow from the minimization of the mean square error and are given by the following  $(N_0 + 1) \times (N_0 + 1)$  linear system of equations

$$\sum_{\{i: \mathbf{s}_i \in \mathcal{S}_0\}} \lambda_i \gamma_z(\mathbf{s}_i, \mathbf{s}_j) + \mu = \gamma_z(\mathbf{s}_j, \mathbf{s}_0), \quad j = 1, \dots, N_0 \quad (2.33)$$

$$\sum_{\{i: \mathbf{s}_i \in \mathcal{S}_0\}} \lambda_i = 1, \quad (2.34)$$

where  $N_0$  is the number of points within the search neighborhood of  $\mathbf{s}_0$ ,  $\gamma_z(\mathbf{s}_i, \mathbf{s}_j)$  is the semivariogram between two sampled points  $\mathbf{s}_i$  and  $\mathbf{s}_j$ ,  $\gamma_z(\mathbf{s}_j, \mathbf{s}_0)$  the semivariogram between  $\mathbf{s}_j$  and the estimation point  $\mathbf{s}_0$ , and  $\mu$  is the Lagrange multiplier enforcing the no-bias constraint.  $N_0 + 1, j = j, N_0 + 1 = 1$  for  $j = 1, \dots, N_0$ , while  $N_0 + 1, N_0 + 1 = 0$ . Equation (2.34) enforces the zero-bias condition.

Kriging provides not only an estimation of the variable  $z(\mathbf{s}_0)$  but also the corresponding estimation's error variance (associated uncertainty). For ordinary kriging the error variance 1) depends on the semivariogram model; the estimation

precision depend on the complexity of the spatial variability of random field  $Z$  as modeled by the semivariogram, 2) depends on the data configuration and their distances to the location  $z(\mathbf{s}_0)$  being estimated, 3) is independent of data values; for a given semivariogram model, two identical data configurations yield the same variance no matter their values and 4) the error variance is zero at data locations and increases away from the data while reaches a maximum value for extrapolation situation.

The OK estimation variance is defined by,

$$\sigma_E^2(\mathbf{s}_0) = E \left[ \left\{ Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0) \right\}^2 \right],$$

and is given by the following equation, with the Lagrange coefficient  $\mu$  compensating for the uncertainty of the mean value:

$$\sigma_E^2(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i \gamma_z(\mathbf{s}_i, \mathbf{s}_0) + \mu. \quad (2.35)$$

Overall OK variance is termed as the weighted average of semivariograms from the new point  $\mathbf{s}_0$  to all calibration points  $\mathbf{s}_j$ , plus the Lagrange multiplier.

#### 2.4.4 Universal kriging interpolation

In certain cases, the data exhibit a global trend over the study area. It is possible to incorporate in kriging a trend (drift function) modeling the global behavior. The resulting estimation algorithm is known as “Universal kriging” (UK) and was proposed by (Matheron 1969). UK requires the drift function  $m_z(\mathbf{s})$  and the semivariogram of the residuals  $e_z(\mathbf{s})$  (Goovaerts 1997). The trend is usually approximated by linear or higher order polynomials of the space coordinates (Ahmed 2007). The drift function is given by

$$m_z(\mathbf{s}) = \sum_{k=1}^K a_k f_k(\mathbf{s}), \quad (2.36)$$

where  $f_k(\mathbf{s})$  are basis functions and  $a_k$  are the drift coefficients (Goovaerts 1997). The UK estimator of the hydraulic head is expressed as follows:

$$\hat{z}(\mathbf{s}_0) = m_z(\mathbf{s}_0) + \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i e(\mathbf{s}_i) = m_z(\mathbf{s}_0) + \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i [z(\mathbf{s}_i) - m_z(\mathbf{s}_i)], \quad (2.37)$$

where  $\lambda_i$  ( $i = 1, \dots, N_0$ ) are the UK weights,  $e(\mathbf{s}_i)$  is the residual at  $\mathbf{s}_i$  and  $m_z(\mathbf{s}_0)$  is the drift at  $\mathbf{s}_0$ .

The kriging weights are determined by the solution of the following  $(N_0 + K) \times (N_0 + K)$  linear system of equations, where  $N_0$  is the number of points within the search neighborhood of  $\mathbf{s}_0$ ,

$$\sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i \gamma_z(\mathbf{s}_i, \mathbf{s}_j) + \sum_{k=1}^K f_k(\mathbf{s}_j) \mu_k = \gamma_z(\mathbf{s}_j, \mathbf{s}_0), \quad j = 1, \dots, N_0 \quad (2.38)$$

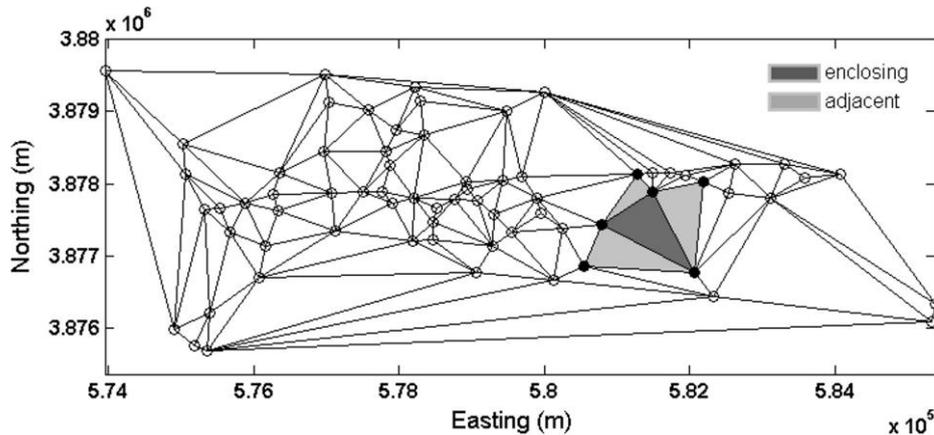
$$\sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i f_k(\mathbf{s}_i) = f_k(\mathbf{s}_0), \quad k = 1, \dots, K, \quad (2.39)$$

where  $\gamma_z(\mathbf{s}_i, \mathbf{s}_j)$  is the semivariogram of the residuals between two sampled points  $\mathbf{s}_i$  and  $\mathbf{s}_j$ ,  $\gamma_z(\mathbf{s}_j, \mathbf{s}_0)$  the semivariogram of the residuals between a sampled point  $\mathbf{s}_j$  and the estimation point  $\mathbf{s}_0$ , and  $\mu_k$  are the Lagrange multipliers for each basis function. The kriging variance is given by the following equation (Goovaerts 1997),

$$\sigma_E^2(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i \gamma_z(\mathbf{s}_i, \mathbf{s}_0) + \sum_{k=1}^K f_k(\mathbf{s}_0) \mu_k. \quad (2.40)$$

#### 2.4.5 Kriging with Delaunay triangulation

Kriging with Delaunay triangulation (DK) uses the Delaunay triangles to determine the search neighborhood  $\mathbb{S}_0$  around the estimation point. The kriging equations in DK are identical to OK (Hessami *et al.* 2001). DK reduces the computational cost of kriging and ensures that the estimate depends only on data in each point's immediate neighborhood.



**Figure 2.4** Delaunay Triangulation of monitoring sites in Mires Basin. The vertices of the enclosing triangle (dark color) that contains the estimation point  $\mathbf{s}_0$  are the first-order neighbors of  $\mathbf{s}_0$ ; the vertices of the three adjacent triangles (grey color) that do not belong to the enclosing triangle provide the second-order neighbors of  $\mathbf{s}_0$ .

The Delaunay triangulation (e.g., Figure 2.4) is the dual graph of the Voronoi diagram for the sampling locations  $\mathbf{s}_i$ ,  $i = 1, \dots, N$ . The latter is a set of polygons  $P_i$ , each of which is centered at  $\mathbf{s}_i$  and contains all the points that are closer to  $\mathbf{s}_i$  than to any other data point. The Delaunay triangulation is formed by drawing line segments between Voronoi vertices if their respective polygons have a common edge (Okabe *et al.* 1992, Mulchrone 2003, Ling *et al.* 2005). The Delaunay triangle containing the estimation point  $\mathbf{s}_0$  is located using the «T-search<sup>1</sup>» function of Matlab<sup>®</sup> (Matlab v.7.5). The vertices of the triangle  $T_0$  containing  $\mathbf{s}_0$  are the first-order neighbors of  $\mathbf{s}_0$ . Second-order neighbors are determined from the vertices of the triangles adjacent to  $T_0$  that do not belong to  $T_0$  (Hessami *et al.* 2001) (see Figure 2.4). The number of second-order neighbors ranges between one and three. If the search neighborhood only includes the first-order neighbors, the CPU time is reduced but the precision of the estimates is lower (Hessami *et al.* 2001).

#### 2.4.6 Residual Kriging

Residual Kriging (RK) combines a trend function with interpolation of the residuals. In RK the estimate is expressed as:

$$\hat{z}(\mathbf{s}_0) = m_z(\mathbf{s}_0) + \hat{z}'(\mathbf{s}_0), \quad (2.41)$$

<sup>1</sup> The tsearch function will be replaced in future Matlab<sup>®</sup> releases by DelaunayTri class

where  $m_z(\mathbf{s}_0)$  is the trend function, and  $\hat{z}'(\mathbf{s}_0)$  is the interpolated residual by means of OK (Rivoirard 2002). Typically, the trend function is modeled as:

$$m_z(\mathbf{s}_0) = \sum_{k=0}^p \beta_k q_k(\mathbf{s}_0); q_k(\mathbf{s}_0) \equiv 1, \quad (2.42)$$

where  $q_k(\mathbf{s}_0)$  are the values of *auxiliary variables* at  $\mathbf{s}_0$ ,  $\beta_k$  are the estimated regression coefficients and  $p$  is the number of auxiliary variables (Draper and H. Smith 1981, Hengl 2007, Hengl *et al.* 2007). Auxiliary variables could include polynomials of the data coordinates (x,y). The regression coefficients are estimated from the sample using ordinary least squares (OLS) or generalized least squares (GLS). However, it has been shown (Kitanidis 1993) that GLS does not confer any significant benefit if the sampling locations are not clustered. The variance of the estimates follows from the equations (Hengl *et al.* 2003, Hengl *et al.* 2007):

$$\sigma^2(\mathbf{s}_0) = \sigma_d^2(\mathbf{s}_0) + \sigma_f^2(\mathbf{s}_0), \quad (2.43)$$

$$\sigma_d^2(\mathbf{s}_0) = \mathbf{q}_0^T (\mathbf{q}^T \boldsymbol{\gamma}_z^{-1} \mathbf{q})^{-1} \mathbf{q}_0, \quad (2.44)$$

$$\sigma_f^2(\mathbf{s}_0) = \sum_{\{i: \mathbf{s}_i \in \mathbb{S}_0\}} \lambda_i \gamma_z'(\mathbf{s}_i, \mathbf{s}_0) + \mu, \quad (2.45)$$

where  $\sigma_d^2(\mathbf{s}_0)$  is the drift prediction variance,  $\mathbf{q}_0$  is the vector of  $(p+1) \times 1$  predictors at the unvisited location,  $\mathbf{q}$  is the matrix of  $(N_0+1) \times (p+1)$  predictors at the sampling points in the search neighborhood,  $\boldsymbol{\gamma}_z'$  is the semivariogram matrix of the  $(N_0+1) \times (N_0+1)$  residuals at the measured locations (neighborhood) and  $\sigma_f^2(\mathbf{s}_0)$  is the kriging (OK) variance of residuals. The terms involved in the drift variance prediction are presented below in vector and matrix form as appropriate:

$$\mathbf{q}_0 = \begin{bmatrix} q_1(\mathbf{s}_0) \\ q_2(\mathbf{s}_0) \\ \vdots \\ q_p(\mathbf{s}_0) \\ 1 \end{bmatrix}$$

$$\mathbf{q} = \begin{bmatrix} q_1(\mathbf{s}_1) & \dots & \dots & q_p(\mathbf{s}_1) & 1 \\ q_1(\mathbf{s}_2) & \dots & \dots & q_p(\mathbf{s}_2) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_1(\mathbf{s}_{N_0}) & \dots & \dots & q_p(\mathbf{s}_{N_0}) & 1 \\ 1 & \dots & \dots & 1 & 1 \end{bmatrix}$$

$$\gamma_{z'} = \begin{bmatrix} \gamma_{z'}(\mathbf{s}_1, \mathbf{s}_1) & \dots & \dots & \gamma_{z'}(\mathbf{s}_1, \mathbf{s}_{N_0}) & 1 \\ \gamma_{z'}(\mathbf{s}_2, \mathbf{s}_1) & \dots & \dots & \gamma_{z'}(\mathbf{s}_1, \mathbf{s}_2) & 1 \\ \vdots & & & \vdots & \vdots \\ \gamma_{z'}(\mathbf{s}_{N_0}, \mathbf{s}_1) & \dots & \dots & \gamma_{z'}(\mathbf{s}_{N_0}, \mathbf{s}_{N_0}) & 1 \\ 1 & \dots & \dots & 1 & 0 \end{bmatrix}$$

## 2.5 Spatial Model Validation

The groundwater level in Mires basin is investigated using the methods that are described above. All the methods are implemented by original code developed and run in the Matlab<sup>®</sup> programming environment (Matlab v.7.5 on Microsoft Windows XP). This approach allows control of the model parameters and straightforward comparison of the results. To avoid numerical instabilities, we normalize the coordinates of the study area in the interval [0, 1].

Given the small size of the Mires data set, we use leave-one-out cross-validation to compare different spatial models. This procedure consists of removing one datum at a time from  $S$  and estimating its value based on the remaining  $N-1$  data. Interpolated values are compared with their measured counterparts using the global performance measures listed below. The “optimal” spatial model is determined based on the performance of statistical metrics that quantify differences between the estimated and true values (Isaaks and Srivastava 1989, Goovaerts 1997, Leuangthong *et al.* 2004, Ahmadi and Sedghamiz 2008). The validation measures defined below are used, where  $z^*(\mathbf{s}_i)$  and  $z(\mathbf{s}_i)$  are, respectively, the estimated and true head values

at point  $\mathbf{s}_i$ . The estimates are obtained by removing  $z(\mathbf{s}_i)$  from the dataset and interpolating the remaining data;  $\overline{z(\mathbf{s}_i)}$  denotes the spatial average of the data and  $\overline{z^*(\mathbf{s}_i)}$  the spatial average of the estimates, while  $N$  is the number of observations.

*Mean absolute error (MAE):*

$$\varepsilon_{\text{MA}} = \frac{1}{N} \sum_{i=1}^N |z^*(\mathbf{s}_i) - z(\mathbf{s}_i)|, \quad (2.46)$$

*Bias:*

$$\varepsilon_{\text{BIAS}} = \frac{1}{N} \sum_{i=1}^N z^*(\mathbf{s}_i) - z(\mathbf{s}_i), \quad (2.47)$$

*Mean absolute relative error (MARE):*

$$\varepsilon_{\text{MAR}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{z^*(\mathbf{s}_i) - z(\mathbf{s}_i)}{z(\mathbf{s}_i)} \right|, \quad (2.48)$$

*Root mean square error (RMSE):*

$$\varepsilon_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N [z^*(\mathbf{s}_i) - z(\mathbf{s}_i)]^2}, \quad (2.49)$$

*Linear Correlation Coefficient:*

$$R = \frac{\sum_{i=1}^N [z(\mathbf{s}_i) - \overline{z(\mathbf{s}_i)}][z^*(\mathbf{s}_i) - \overline{z^*(\mathbf{s}_i)}]}{\sqrt{\sum_{i=1}^N [z(\mathbf{s}_i) - \overline{z(\mathbf{s}_i})]^2} \sqrt{\sum_{i=1}^N [z^*(\mathbf{s}_i) - \overline{z^*(\mathbf{s}_i)}]^2}}, \quad (2.50)$$

## 2.6 Comparison of interpolation methods

There is not a universally optimum interpolation method that can be used for all kind of datasets. Two spatial interpolation comparison exercises have been organized by the Radioactivity Environmental Monitoring (REM) Group of the Joint Research Centre of the European Commission (Dubois 1998, Dubois and Galmarini

2005). These exercises focused on radioactivity monitoring in the European continent and in particular on automatic (i.e., without user involvement) mapping.

Cornford (2005) emphasized the problems of interpreting comparative interpolation studies. First, their results do not admit generalization and are often contradictory. In addition, for a single dataset several or all assessed methods may exhibit similar performance. Hence, the choice of the “optimal” interpolation method is dictated by other factors, such as computational speed, implementation cost, scaling with data size and the ability to make probabilistic predictions (estimates of the prediction error). Van den Boogaart (2005) agrees that comparative studies based on one or two datasets can be misleading, and that a uniformly optimal method for all kinds of datasets does not exist. He points out that the performance and utility of the methods should be assessed in terms of decision making requirements (e.g. concerning outliers, estimation variances) and its adaptability to the complexity of the specific dataset (e.g. sparse data, presence of trends) and not only in terms of mean square errors. Myers (2005) emphasizes the use of clear software standards, common hardware configurations and an extensive set of performance measures to allow the duplication of reported results by others.

In light of the above remarks, we use the same programming environment for all the methods tested so that results are directly comparable. The methods are described in detail including the values of user-defined parameters to allow reproduction of our results by others. The performance of the interpolation methods for the Mires basin dataset is based on cross-validation measures, uncertainty estimation ability, methodological specifications (search neighbourhood, differentiability, contour map effects) and adaptability to the data set statistics (size of data, outliers). The results obtained in this work are useful for mapping groundwater level spatial variability in basins with similar characteristics, and more generally in environmental monitoring applications that involve spatially distributed data (e.g., air pollutants, groundwater quality, soil contaminants, etc.).

## **2.7 Spatiotemporal interpolation**

Space-time analysis and prediction of groundwater level variability in a basin is more important than purely spatial analysis. In scarce and sparsely monitored basins, analysis and prediction becomes more difficult because of limited data availability. In case of temporal-only estimation it is desirable to formulate a

predictive model of groundwater table depth that can incorporate the various physical parameters that determine the groundwater level, such as meteorological data (e.g. precipitation), evapotranspiration, runoff, and water usage. A stochastic model is required due to the considerable uncertainty of certain parameters (e.g., evapotranspiration), the spatial variability of conditions within the basin, the sparse nature of sampling in space-time, and the inadequacy of available measurements (e.g., water usage is estimated based on data from the official boreholes; however, an unspecified number of unregulated boreholes operate in the area). We propose to model the variation of water table depth with a discrete time autoregressive exogenous variable (ARX) model (Knotters and Bierkens 2001, Knotters and Bierkens 2002). The autoregressive discrete-time model will account for the time variability based on the time series of groundwater level, precipitation measurements and abstraction rates. The term exogenous denotes that the model equations incorporate information from other (than the water table depth) variables. The ARX model is embedded in a Kalman filter in order to determine the ARX model parameters according to the system identification procedure (Ljung 1999). This is typically based on the maximum likelihood algorithm to determine the “optimal” parameters. A similar approach using differently trained Artificial Neural Network models was applied to a specific well in the Mesara valley to model the groundwater level temporal variability and then to estimate it during the period 1997-2002 (Daliakopoulos *et al.* 2005, Tsanis *et al.* 2008).

Space-time geostatistical approaches can be used to model the groundwater level variability. In areas with limited spatial and temporal data availability, application of space-time approaches can improve the reliability of predictions by incorporating space-time correlations instead of purely spatial ones; therefore the former approaches involve more parameters (Lee *et al.* 2010).

In Christakos (1991a, b) a theory of spatiotemporal random fields is developed and properties of the most important classes of spatiotemporal fields are examined. The theory is used to describe the correlation structure of space non-homogeneous/time non-stationary processes and to derive optimal estimators for data dispersed simultaneously in space and time. Christakos and Hristopoulos (1998) presented a completed review and new material on Bayesian maximum entropy estimation techniques and space-time random field estimation methods. Later Kolovos *et al.* (2004) presented various methods for constructing space-time

covariance models. These include non-separable (in space and time) covariance models derived from physical laws (i.e., differential equations and dynamic rules), spectral functions, and generalized random fields. It is also shown that non-separability is often a direct result of the physical laws that govern the process. The proposed methods can generate covariance models for homogeneous/stationary as well as for non-homogeneous/non-stationary environmental processes across space and time.

Kyriakidis and Journel (1999) presented an extensive review for space-time geostatistical techniques. The initial approach for space-time geostatistical analysis was to add time as an additional dimension of space (Rouhani and Myers 1990, Kyriakidis and Journel 1999). This approach was implemented using kriging technique developing space-time kriging. Advanced space-time geostatistical approaches were also developed by (Christakos 1991b, Christakos 2000) and (Kyriakidis and Journel 1999) to account for fundamental dependencies in the combined space-time metric (Lee *et al.* 2010).

Bayesian approaches as an alternative to non-Bayesian i.e. kriging were introduced by Christakos (1990, 2000). The Bayesian Maximum Entropy (BME) is a non-linear method that relies on a two-steps procedure that first involves a Maximum Entropy step (the ME part of BME) to obtain a prior distribution and on a Bayesian conditioning rule for the assimilation of secondary information (possibly soft data). BME provides a flexible framework that accounts for the wide variety of available knowledge bases and leads, in general, to optimal non linear space/time estimators. Applications include soil science e.g., (Brus *et al.* 2008), water consumption (Lee and Wentz 2008), environmental health studies e.g., (Christakos and Hristopulos 1998, Yu *et al.* 2009, Kolovos *et al.* 2012), atmospheric environment e.g., (Vyas and Christakos 1997, Christakos and Serre 2000, Yu *et al.* 2011).

A framework for stochastic spatiotemporal modeling has also been presented by Kyriakidis (2001a, b). A data set that is more densely sampled in time than in space can be modeled via a set of spatially correlated time series (Rouhani and T.J. Hall 1989). The time series at each sampled location can be decomposed into a non-stationary deterministic or stochastic trend component and a stationary residual component. The residual time series is then fitted with a covariance model. It is possible to apply this approach to perform spatial interpolation or extrapolation; extending it to a continuous spatial domain by determining temporal covariance

models or time series independently at each fixed location and then regionalising them in space. Time series regionalization involves simulation of the spatiotemporal residual field by generating simulated realizations at any unmonitored location; sequential Gaussian simulation, i.e., autoregression (Kyriakidis and Journel 2001a). Simple kriging (SK) is used for the covariance parameters regionalization. This allows temporal covariance models or time series to be determined at unsampled locations and reduce the computational effort associated with the number of (temporal) covariances. A simulation procedure is also used for the trend regionalization which is typically approximated by periodic and sine and cosine functions in conjunction with multiple regression. The independently simulated trend and residual realizations are then added to produce realizations for the spatiotemporal field. An estimate of the standard deviation of the unknown residual profile at any unmonitored location is also derived via SK. Although this framework has been characterized as powerful, it requires multiple regionalisations, thus time and computational load (Kyriakidis and Journel 2001b).

Space-time Kriging has been applied in geohydrology by Rouhani and T.J. Hall (1989) where intrinsic random functions (polynomial spatiotemporal covariance) for space-time kriging of piezometric data were used. In Rouhani and Myers (1990) potential drawbacks of space-time geostatistical analysis on geohydrological data (piezometric data) are discussed. More recently space-time kriging was applied by Mendoza-Cazares and Herrera-Zamarron (2010) for the estimation of the water level of the Queretaro-Obrajuelo aquifer and Hoogland *et al.* (2010), where the goal of the study was to map the seasonal fluctuation of water-table depths in Dutch nature conservation areas. Furthermore space-time Kriging was used for the design of rainfall networks in time and space (Rodriguez-Iturbe and Mejia 1974) and in a comparison study for estimating runoff time series in ungauged locations (Skøien and Blöschl 2007).

In addition, space-time Kriging has also been used in a wide range of scientific fields and topics such as agriculture (Stein 1998, Heuvelink and Egmond 2010) atmospheric data (De Iaco *et al.* 2002b, Myers 2002, Nunes and Soares 2005), soil science-water content (Snepvangers *et al.* 2003, Jost *et al.* 2005), surface temperature data (Hengl *et al.* 2011) wind data (Gneiting 2002), gama radiation data (Heuvelink and Griffith 2010), epidemiology (Gething *et al.* 2007) and forecasting municipal water demand (Lee *et al.* 2010).

Space-time geostatistical analysis is based on the joint spatial and temporal dependence between observations. There are two ways to represent space-time random variables (Christakos 1991b), a) the full space-time models using separable or non-separable space-time covariance functions;  $Z(\mathbf{s}, t)$ ,  $(\mathbf{s}, t) \in D \times T$ ,  $D \subseteq \mathbb{R}^d$  is the spatial domain ( $d$  is the dimensions) and  $T \subseteq \mathbb{R}$  is the temporal domain; and b) vectors of temporally correlated spatial random fields  $Z(\mathbf{s}, t) = Z_t(\mathbf{s})$ ,  $t = 1 \dots T$ , where  $T$  is the number of temporally correlated SRF or vectors of spatially correlated time series  $Z(\mathbf{s}, t) = Z_s(t)$ ,  $\mathbf{s} = 1 \dots n$ , where  $n$  is the number of locations. The representation depends on the domain density (space or time).

The space-time kriging method employs the first model. The two main tasks of space-time analysis are interpolation and extrapolation. The first refers to estimation of variable values at unmeasured locations inside the spatial extend of the study area, while the latter extends the estimations ahead of the boundaries of the observations in space or time. The main assumption used in interpolation and extrapolation is that the specific patterns extracted from the available data analysis delivers sufficient information to capture the spatiotemporal dynamics of the observed data (Lee *et al.* 2010).

The application of space-time kriging to space-time field data entails practical difficulties, especially for geohydrological data. The most important problem is the construction of valid covariance or semivariogram models in space-time; valid covariance or semivariogram models constructed in the spatial or temporal context are not, in general valid when a valid temporal model with a valid spatial model are combined to produce a spatiotemporal model. Geohydrological data are usually dense in time and sparse in space. This feature is significant since covariances or semivariograms can lead to significantly different levels of reliability in space and time. The kriging estimator is inferior if the data is collected during the wet season and the estimates refer to the dry season. Finally, in space-time kriging applications computational problems may arise for specific sampling patterns with the coefficient matrix in the kriging system. These problems are due, e.g. to insufficient number of sample locations compared to the order of a drift function applied to the data, or scarcity and clustering of sampling locations (Rouhani and Myers 1990).

### 3. Study area and data exploratory statistical analysis

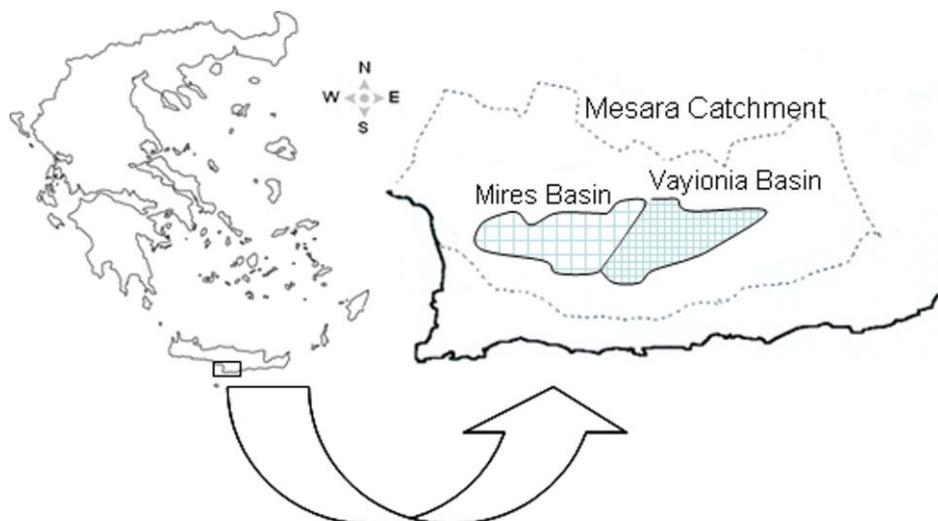
#### 3.1 Physical setting and information

The study area is located on the island of Crete (Greece). Crete has a dry sub-humid Mediterranean climate with long hot and dry summers and relatively humid and cold winters. The island's marginal groundwater resources are extensively used for agricultural activities and human consumption. Although Crete's temperature lies between the 18.5 and 19.0 °C isotherms, it shows considerable variability throughout the island. During the winter period the temperature decreases with increasing altitude, but during the summer period it increases from the coast to the inland areas. The presence of mountains, mainly at the center and the south of the island, significantly affects the climate of different areas. Precipitation is higher in the Northwestern coastal areas and lower in the Southeastern part of the island. The total hours of sunshine in the Southeastern part are more than in the Northwestern part. These differences in climatic parameters create local, quite different microclimates (Department of Water Resources Management 2000, Chartzoulakis and Psarras 2005).

The Mesara valley catchment covers an area of 398 km<sup>2</sup> in the south of the Heraklion prefecture, and it is the largest and most productive valley of the island (Figure 3.1). About 250 km<sup>2</sup> of the total valley area are cultivated. The distribution of the cultivated crops is olives 75%, 10% vines, 2% citrus and 13% vegetables (mostly open field and some greenhouses). Olive trees occupy most areas of flat land while vines are less widespread located mainly to the south east of Mires and in the north east of the catchment. An extensive network of pumping stations has been installed since 1984, turning the dry cultivation of olive trees to drip-irrigated (Donta *et al.* 2006). As a result, productivity has risen at the cost of an alarming drop of the water table. Over-exploitation during the past thirty years has led to a dramatic decrease, in excess of 35 meters, in the groundwater level. Potential future climatic changes in the Mediterranean region, population increase, and extensive agricultural activity generate concern over the sustainability of the water resources. The accurate estimation of the spatial variability of the hydraulic head is important for integrated

groundwater resources management plans that will help reduce the risk of desertification.

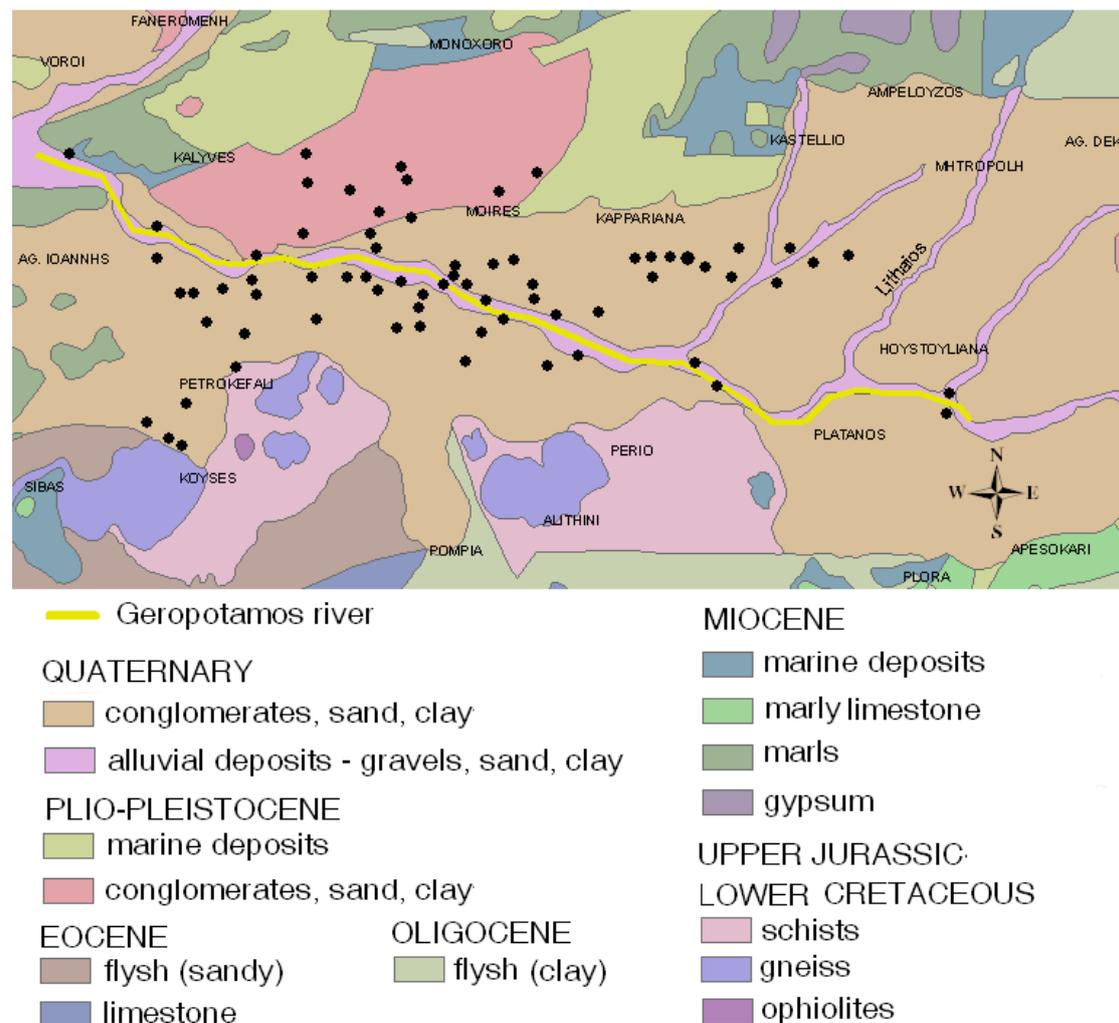
The Mesara Valley catchment comprises an east-west plain. Along the north side of the plain, the mountain altitude varies from 2200 m to 600 m from west to east. Along the southern side, the Asterousia mountain chain rises from 600 m in the west to 1200 m in the east. At the Phaistos constriction in the west, the catchment outlet of the Geropotamos River is at 30 meters above sea level. The catchment area of the northern slopes is 160 km<sup>2</sup> while the southern slopes comprise a catchment area of 126 km<sup>2</sup>. Mainly quaternary alluvial clays, silts, sands and gravels with thickness from a few meters to 100 m or more, cover the plain. The inhomogeneities of the plain deposits give rise to great variations in the hydrogeological conditions even over small distances. The northern slopes are mainly silty-marly Neogene formations while the southern slopes are mainly schist and limestone Mesozoic formations (Donta *et al.* 2006, Kritsotakis and Tsanis 2009, Tsanis and Apostolaki 2009, Kritsotakis 2010). The outlet of the catchment is narrow, confined to a channel cut into an impermeable barrier of lower Tertiary near Phaistos. Topographically, the Mesara catchment is characterized by a flat basin morphology modified by river terraces and alluvial fans. (Peterek and Schwarze 2004). The alluvium basin of Mesara catchment is not a uniform hydrogeological unit, and therefore it is divided into two sub-basins with different hydrogeological properties: the Mires and the Vayionia basins (Kritsotakis and Tsanis 2009), (Figure 3.1). A detailed hydrogeological description of the Mesara valley can be found in Vardavas *et al.* (1996) and Kritsotakis (2010).



**Figure 3.1** Map of Greece and the schematic representation of the Mesara valley and the Mires basin locations on the island of Crete.

### 3.2 Hydrogeological setting

This thesis focuses on the Mires basin of the Mesara valley for two reasons. The first is hydrogeological data availability, and the second that the basin consists mainly of alluvial sediments, providing to a large extent a hydro-geologically uniform study area. The basin has been consistently monitored over the last thirty (30) years for groundwater level variations, rainfall and surface runoff by the Department of Water Resources Management (DWRM) of the region of Crete. The Mires basin is a down-faulted trough with an area of 50.3 km<sup>2</sup>, roughly 14 km long and on average 3 km wide. The trough is filled mainly with Quaternary alluvial sediments which form an inter-bedded sequence of gravels, gravelly sands, sands, silts, silty sands and clays (Kilili-Polychronaki 2001, Donta *et al.* 2006, Kritsotakis 2010). The geological composition of the basin is presented in Figure 3.2.



**Figure 3.2** Geological representation of Mires basin.

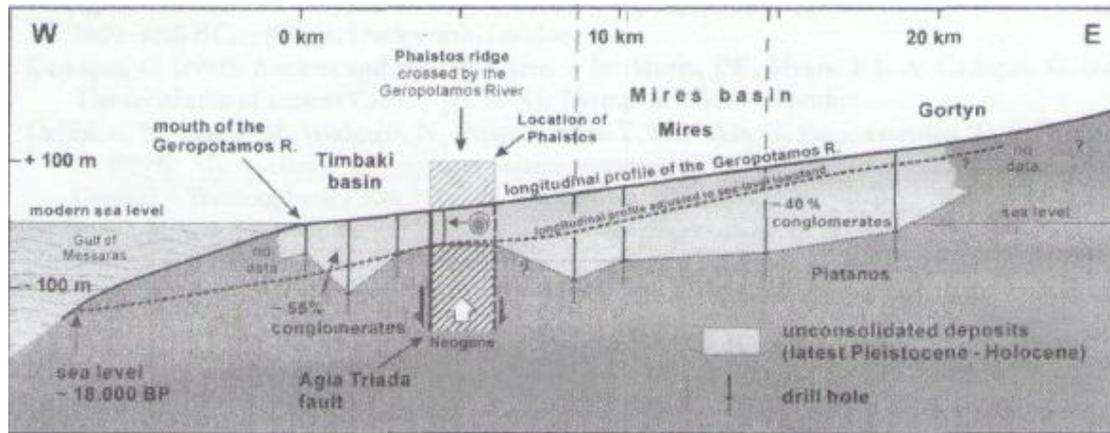
The alluvium aquifer of Mires basin is the most significant of Mesara catchment. Its thickness decreases from the centre of the basin to the south and north. The thickest part is located at the eastern part where Lithaios river crosses Geropotamos river (near Houstoulia). The trough is covered in geological terms mainly with conglomerates, sands and clays of fluvial origin (Figure 3.2). These formations are favor aquifers of high capacity especially at the locations where mainly conglomerates occur.

Borehole logs at the eastern and central part of the basin denote that the conglomerate thickness varies from 75 to 180m (Figure 3.3). The alluvial basin constitution changes from East to West (Figure 3.4). At the Eastern part the deposits are coarser with layers of clay and silt of less than 5m thickness. The surface layer is mainly composed of gravel and sand allowing high infiltration. At the downstream part, west of Mires, the thickness of clay and silt layers increases with the surface layer comprising mainly of clay deposits (Vardavas *et al.* 1996, Kritsotakis and Tsanis 2009).

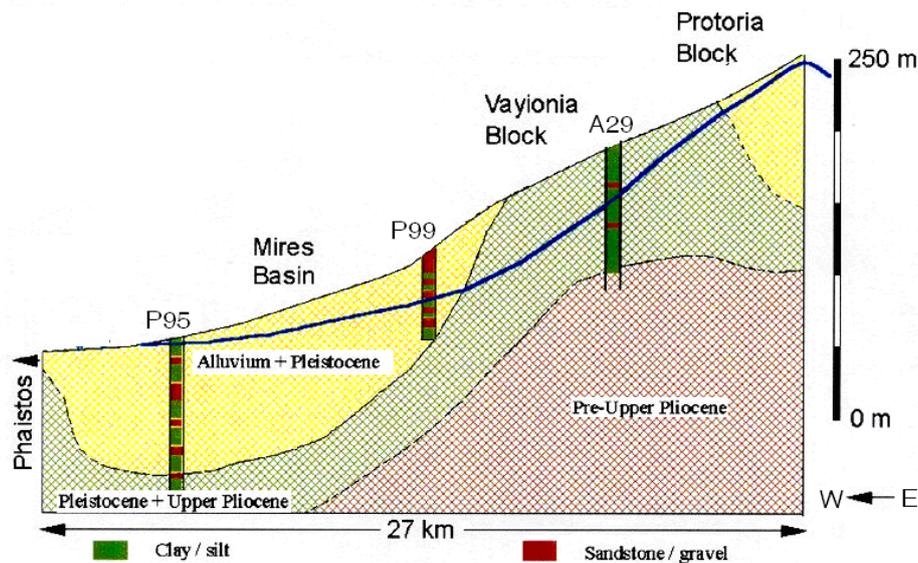
Geophysical surveys have revealed that the flanks of the Mires basin are steep-sided, which may reflect the presence of East-West fractures, or the bank of an erosion channel cut into the underlying lower Pleistocene sediments. Whatever the true nature of the trough boundaries are, extensive fractures parallel to and across the valley are present and have caused the variable thickness of the alluvium throughout the basin. The elevation of the base of the alluvium-lower Pleistocene aquifer ranges from -50 to +100 m, taking the mean sea level (MSL) as reference datum; the lowest points of the elongated trough being to the northeast of Petrokefali and north of Houstoulia. This surface is based on the elevation of the lowest permeable unit identified from borehole logs. Saturated thickness ranges from less than 20 m to over 100 m in the lowest part of the trough (Donta *et al.* 2006).

The aquifer capacity cannot be accurately estimated, as it is part of the unified Mesara catchment hydrological system. The Food and Agriculture organization of United Nations (FAO 1972) estimated the capacity of the aquifer equal to around 86 Mm<sup>3</sup> based on boreholes log data taken that year. The present aquifer capacity is smaller (due to the significant groundwater depletion) and is estimated equal to around 55 Mm<sup>3</sup> (Kritsotakis and Tsanis 2009). The aquifer is supplied during the winter period from the main river that crosses the basin, Geropotamos, and from a secondary, Lithaios, in the eastern part. Surface runoff from the northern and southern

sides of the basin ends up in the basin where a significant percentage supplies the aquifer. Lateral groundwater inflow in the basin occur from Vayionia basin in the eastern part of the basin, while groundwater inflow is likely to also occur from the neogenic formations at the north-northwest part (Kilili-Polychronaki 2001).



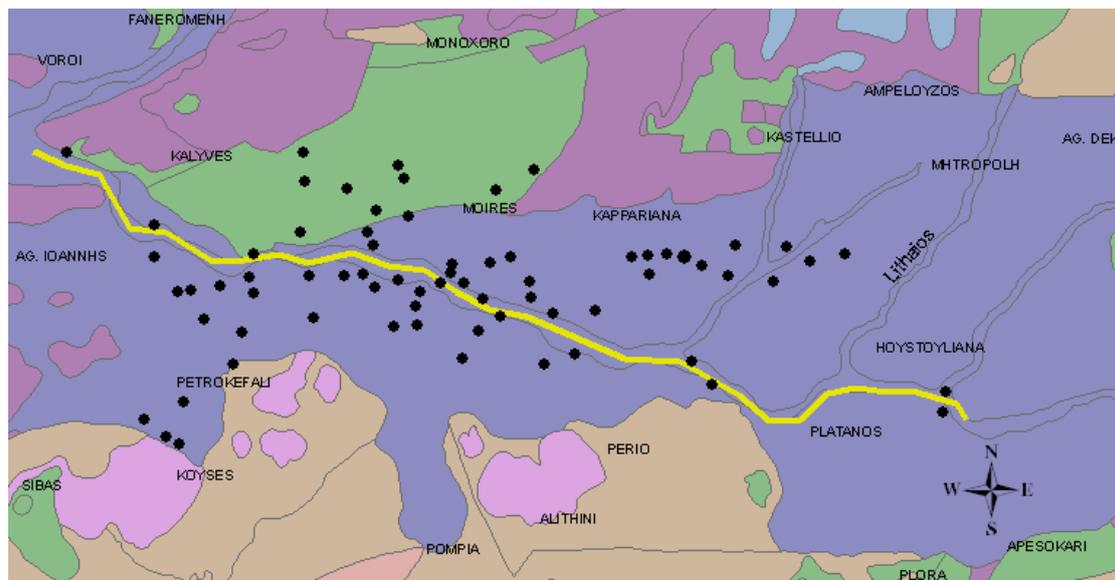
**Figure 3.3** E-W cross-section along the Geropotamos river showing simplified geologic formation of the Mires basin (Fytolakis *et al.* 2005).



**Figure 3.4** Simplified geological structure of the Mires basin, modified after (Vardavas *et al.* 1996).

Groundwater levels in Mesara valley are maximized during March or April with long dropdowns until recharge occurs in winter. The aquifers of Mires and Vayionia basins yielded high discharge rates, as high as  $300 \text{ m}^3/\text{hr}$ , in the early seventies but they are now reduced to about one tenth of that figure. Based on pumping tests, the specific yield mainly ranges between 0.05 and 0.15, but in Mires basin locally it reaches 0.2. The horizontal transmissivity whereas ranges between 0.1 and  $0.01 \text{ m}^2/\text{sec}$  from east to west, while it takes the highest value along the

Geropotamos river bed. At the center of the alluvial aquifer and across the river bed the wells specific discharge is determined 40-100m<sup>3</sup>/h per meter of groundwater level drop. In the eastern part of the basin the rate decreases to 15-40m<sup>3</sup>/h. Lateral groundwater outflow from the valley is small compared with the vertical groundwater outflow. The higher values of hydraulic conductivity are concentrated in the Mires basin, where it varies between 10 and 120 m/day reflecting the presence of several gravel and sand horizons in the alluvial sequence. The least permeable areas are in the Vayionia block, where the scarcity of gravel and sand implies that values are reduced to an average of only 1 m/d. Lower values also occur in the northern side of the Mires basin, where lower Pleistocene rocks similar to those of the Vayionia block are present. The hydraulic properties of the Plio-Pleistocene aquifer are of a magnitude less than that of the alluvial deposits (Donta *et al.* 2006, Kritsotakis 2010).



- Geropotamos river
- impermeable deposits of low to very low hydraulic conductivity
- impermeable or selectively permeable deposits of low to very low hydraulic conductivity
- karst formations of medium to low hydraulic conductivity
- granular mainly alluvial deposits of variable hydraulic conductivity
- miocene and pliocene deposits of medium to low hydraulic conductivity
- granular non alluvial deposits of low to very low hydraulic conductivity
- gypsum

**Figure 3.5** Hydro-geological representation of Mires basin.

The porosity decreases with depth below surface in a range of 0.05 to 0.12 (Croke *et al.* 2000) while the effective porosity which is the percentage of pores (interconnected) that are available for fluid flow (Bear 1979) is determined in Mires basin equal to 0.085 (Kilili-Polychronaki 2001). The rapid groundwater level drop (>35 masl) has affected the hydraulic characteristics of the aquifer. An assessment of the groundwater level depletion impact on the hydraulic properties of the aquifer has shown a decrease of the transmissivity mainly because the thickness of the saturated zone has changed (Kilili-Polychronaki 2001).

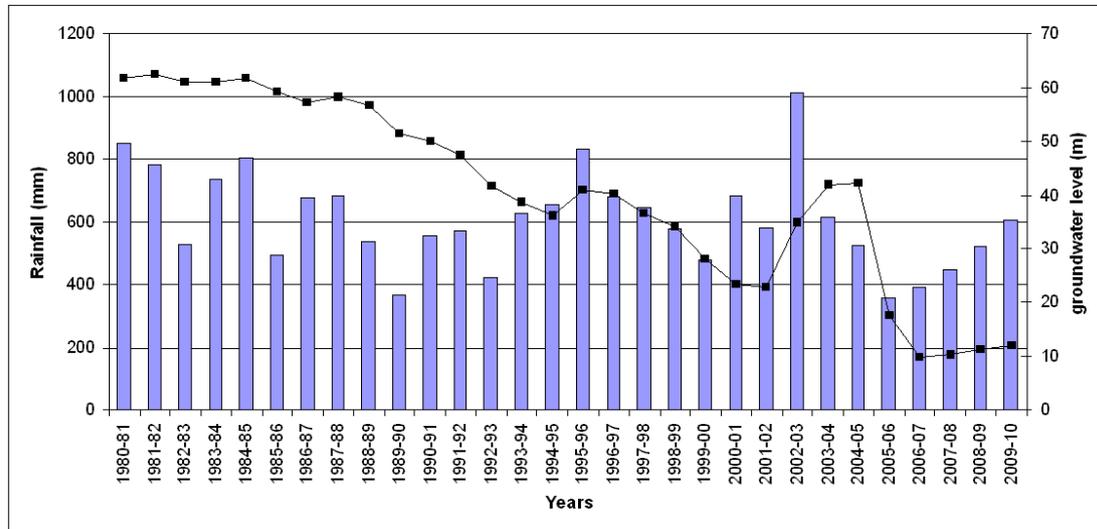
The basin's aquifer is characterized in hydro-geological reports as unconfined (Kilili-Polychronaki 2001, Donta *et al.* 2006), and it is sparsely monitored regarding the groundwater level. Aquifer storage coefficients, obtained from FAO pumping tests, in conjunction with the behavior of the water table (free surface existence) suggests, that although heterogeneous and locally confined, the aquifer behaves at the regional scale as an unconfined unit (Donta *et al.* 2006).

### **3.3 Hydrological setting**

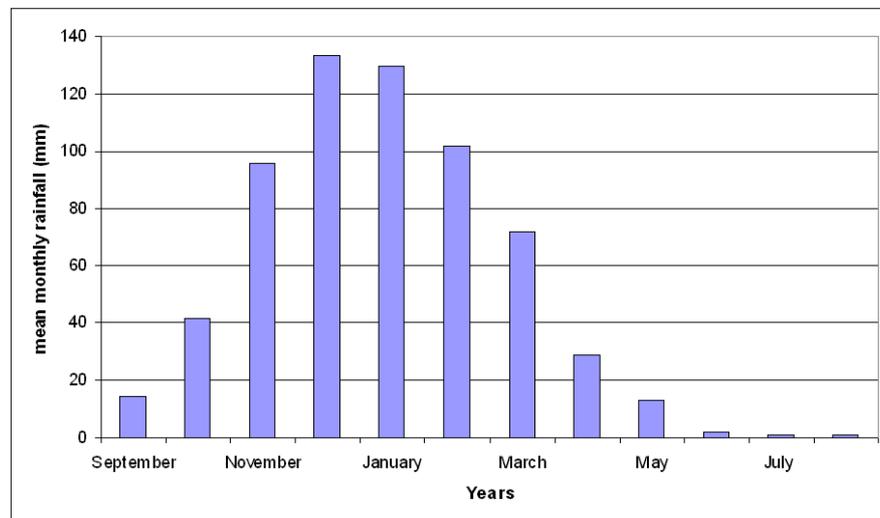
The mean annual rainfall in the Mesara catchment is around 650 mm. About 65% of the rainfall is lost to evapotranspiration and 10% as runoff to the sea, leaving only 25% to recharge the groundwater store (Croke *et al.* 2000). Rainfall increases with the elevation from about 500 mm in the Mesara plain to about 800 mm in the valley slopes, 1100 mm in the Asterousian Mountains and up to 2000 mm in the Idi mountain (Tsanis and Apostolaki 2009). Figure 3.6 presents the mean annual rainfall in Messara for the hydrological years 1981 to 2010. About 40% of the precipitation occurs in the months of December and January, while there is negligible rainfall from June to August (Figure 3.7).

The mean winter temperature is 12°C while the mean summer temperature is 28°C. The relative humidity in the winter is about 70%, while it is about 60% in the summer (Tsanis *et al.* 2011). Pan evaporation is estimated at 1500±300 mm per year while western winds prevail. The potential evaporation is estimated at 1300 mm per year setting the ratio of mean annual rainfall to potential evaporation at about 0.5 implying a dry sub-humid classification according to the (UNCED 1994) definitions (Croke *et al.* 2000, Tsanis and Apostolaki 2009). The plain altitude in Mires is less than 300m although the mean annual rainfall is similar to the catchment average.

Specifically for Mires basin the most recent hydro-geological study (Kilili-Polychronaki 2001) estimates the evapotranspiration at 82%, the infiltration 18% and the surface runoff as negligible.



**Figure 3.6** Annual rainfall in Mesara valley (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).

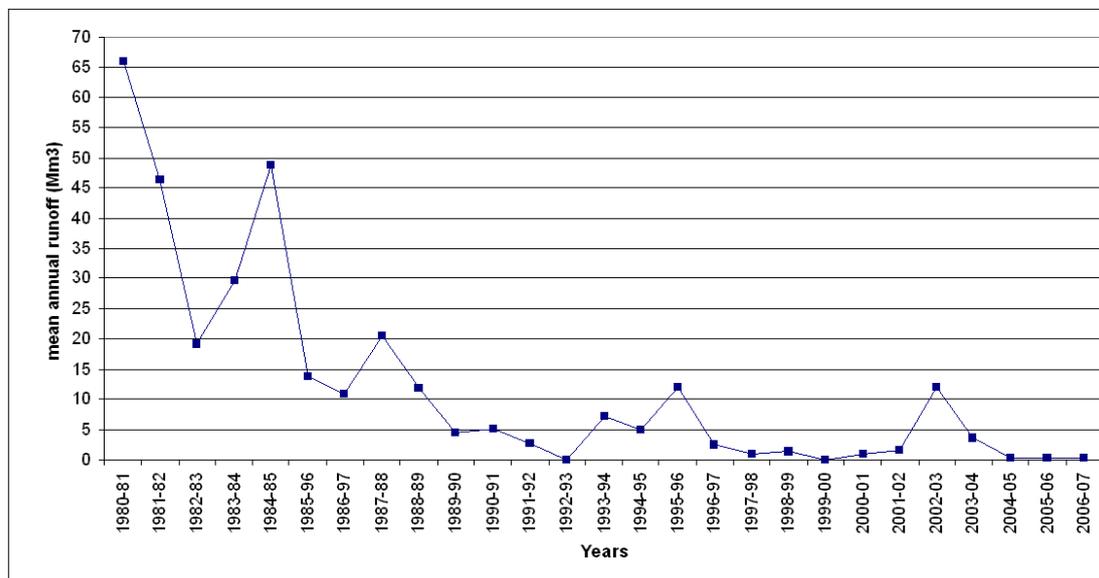


**Figure 3.7** Monthly rainfall in Mesara valley (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).

The main outlet of the catchment, as mentioned above, is Geropotamos river at the Phaistos constriction in the west. In its natural state, the Geropotamos River of the Mesara Valley flowed continuously, and sustained a wetland located near the catchment outlet. The drop in the groundwater level has resulted in the wetland drying up and no flow in the river so during the dry season of the 1989-90 hydrological year as during the whole 1992-1993. Additionally during the hydrological year 1992-1993

there was no river flow out of the valley, while very low flow was measured during the years 1999-2000 and 2004 to 2007 (Figure 3.8).

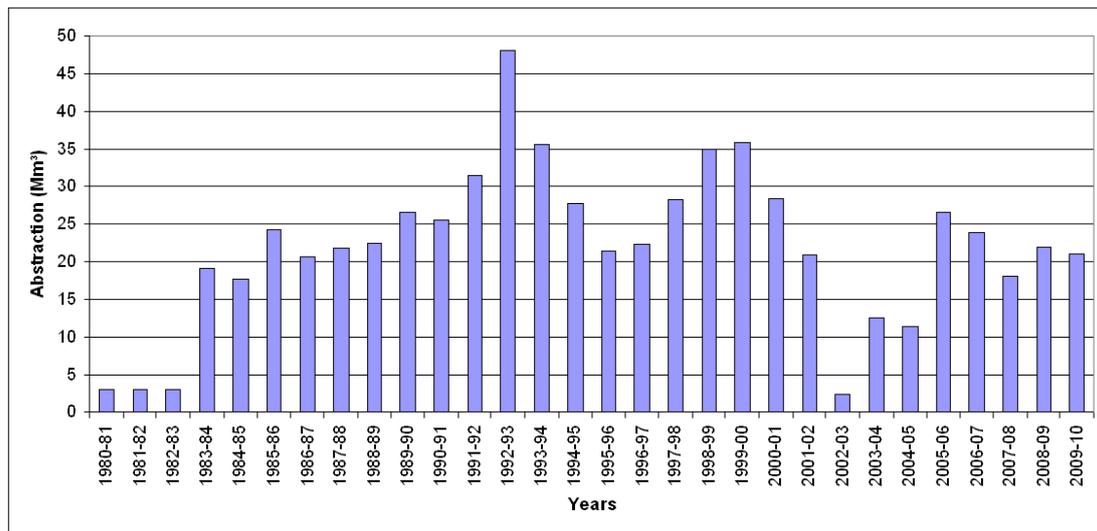
Groundwater abstraction rates in Mires basin varied from 20-400 m<sup>3</sup>/h before the extensive exploitation of the aquifer, while the pumping rates were correlated with the hydrogeological formation (higher in the alluvial formations). The number of boreholes operated at that time in the entire valley was 26. The average pumping rate at the center of the basin across the Geropotamos root was 200 m<sup>3</sup>/h and the groundwater level was less than 10 meters below surface. At the eastern part, near Lithaios, the pumping rates were similar but the depth of the water table varied from 20-40 m<sup>3</sup>/h from south to north.



**Figure 3.8** Annual runoff of Geropotamos of Mesara catchment (data provided by the Administration of Land Reclamation of the Prefecture of Crete and the Department of Water Resources Management of the Prefecture of Crete).

An extensive network of pumping stations was installed on 1984. It is estimated that after the installation of the network around 40Mm<sup>3</sup> (22Mm<sup>3</sup> in Mires basin) on average are being pumped per year. The temporal variability of the annual pumping volume in Mires basin is presented in Figure 3.9. Before the installation of the groundwater irrigation system, less than 10Mm<sup>3</sup> were pumped per year, the average discharge out of the valley was about 20 Mm<sup>3</sup>/yr corresponding to 50 mm of the annual rainfall lost as runoff to the sea. It is estimated that the annual recharge of the groundwater store was about 60 Mm<sup>3</sup>/yr (150 mm) and the evapotranspiration losses were about 160 Mm<sup>3</sup>/yr (400 mm). Nowadays, the surface runoff and the groundwater supplies are decreasing rapidly (Donta *et al.* 2006, Kritsotakis 2010).

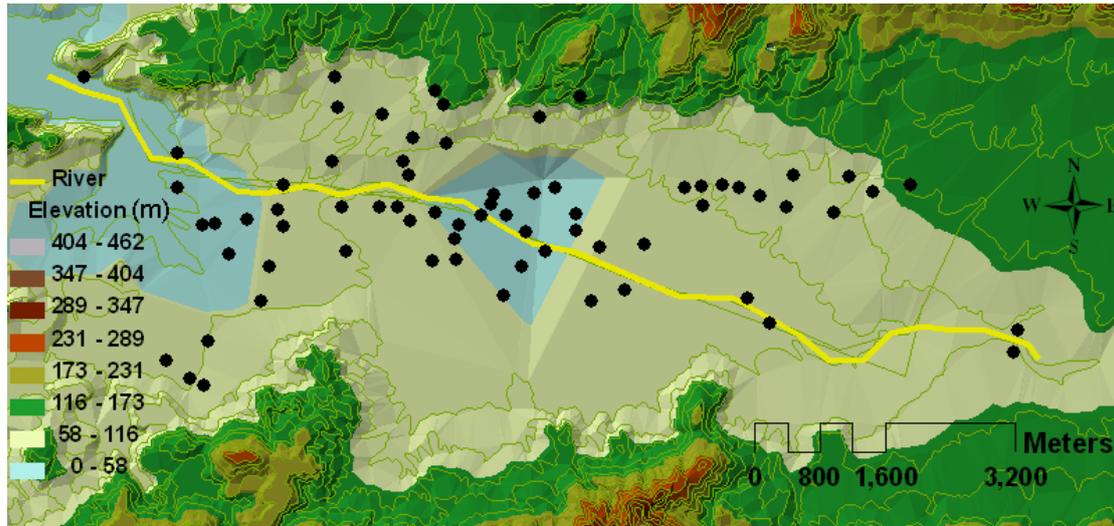
The large number of operating boreholes and overexploitation are responsible for this trend. A recent publication cites that 1400 wells were operated on 2007 in whole Mesara valley (Kritsotakis 2009). In Mires basin, the Greek Institute of Geology and Mineral Exploration (IGME) has registered more than 80, although data are not available for all of them.



**Figure 3.9** Annual abstraction rate in Mires basin (data provided by the Administration of Land Reclamation of the Prefecture of Crete, the Department of Water Resources Management of the Prefecture of Crete and by Kritsotakis (2010).

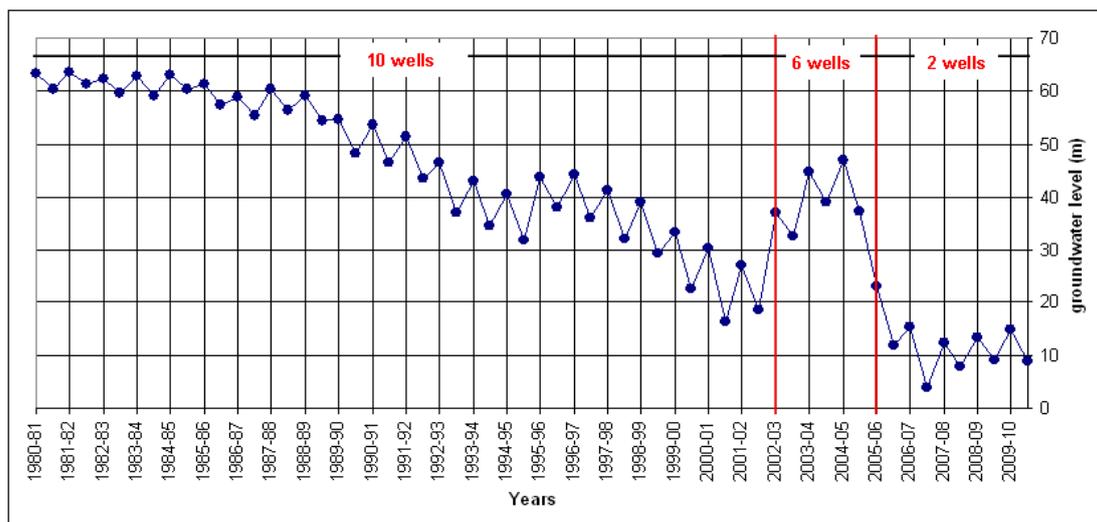
### 3.4 Groundwater level Data availability

The groundwater level monitoring locations in the basin are presented in Figure 3.10, while the groundwater level drop since 1981 has exceeded 35m (Figure 3.11). Our overall goal is to use stochastic methods for the spatiotemporal monitoring and prediction of the groundwater level in the basin. The data used in this thesis comprise seventy (70) hydraulic head measurements, from wells located in Mires basin, for the wet period of the hydrological year 2002-2003 (October – April is the wet period of the hydrological year). This is the only period for which a full set of recorded head values exists. The data have been provided by the Administration of Land Reclamation of the Prefecture of Crete. The measurements are unevenly distributed and mostly concentrated along Geropotamos, a temporary river that crosses the basin (Figure 3.10). The range of hydraulic heads varies from an extremely low value of 9.4 meters above sea level (masl) to 62 masl for the wet period of the year. The head values are even lower during the dry period.



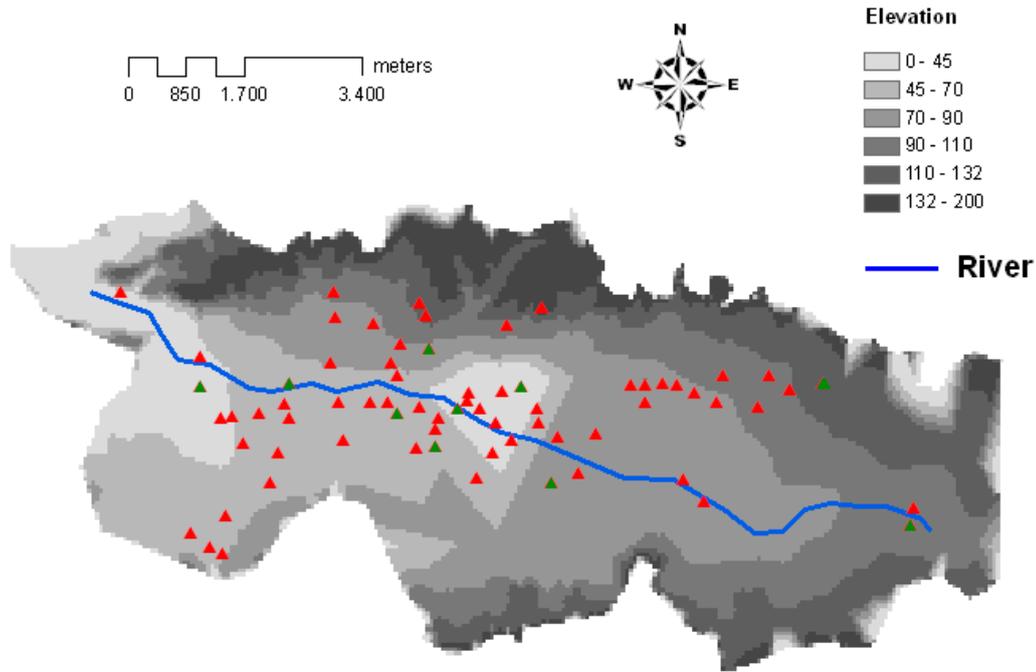
**Figure 3.10** Topographic map showing the locations of groundwater head measurement in Mires basin along with the corresponding surface elevation and the temporary river (Geropotamos) path.

Ten wells (Figure 3.12) were monitored between the years 1981 and 2003, while others were monitored for shorter periods. Since 2003, the regular biannual monitoring of the operating boreholes has been replaced by the continuous monitoring of two telemetric stations placed in two boreholes selected by the DWRM. It has to be mentioned that selective measurements occur at specific wells in the basin on 2003-2006 period biannually at four of the ten set boreholes (leading to six monitored locations). The groundwater level increase during the period 2003-2005 is due to the increased rainfall and the low pumping activity during that period (Figures 3.6 and 3.9). On the other hand the groundwater level trend since 2006 is mainly affected from the small number of the available observations.



**Figure 3.11** Mean bi-annual groundwater level in Mires basin. An average of 10 wells until 2002-03 is considered, 6 wells until 2005-06 and 2 wells until 2009-10.

Average pumping rates for the 70 wells operated in the basin have been also provided from the Administration of Land Reclamation of the Prefecture of Crete. The correlation of the pumping rates with the corresponding groundwater levels is characterized important as it is 68%. The pumping rates vary from 10-170 m<sup>3</sup>/h.



**Figure 3.12** Topographic map showing the locations of monitored wells (triangles) in Mires basin along with the corresponding surface elevation and the temporary river path. Green color is used to mark the wells monitored for the period 1981-2003.

### 3.5 Exploratory statistics

The results of exploratory analysis for the hydraulic head data are shown in Table 3.1. The head data have skewness and kurtosis coefficients equal to  $\hat{s}_z = 0.81$  and  $\hat{k}_z = 2.58$  respectively, implying a mild deviation from Gaussian statistics ( $\hat{s}_z = 0$  and  $\hat{k}_z = 3$  respectively).

**Table 3.1** Statistical measures of the hydraulic head data.  $z_{\min}$ : minimum value;  $z_{0.25}$  first quartile;  $z_{0.50}$  median;  $m_z$  mean;  $z_{0.75}$  third quartile;  $z_{\max}$  maximum value;  $\hat{\sigma}_z$  standard deviation;  $\hat{s}_z$  skewness coefficient;  $\hat{k}_z$  kurtosis coefficient.

$z_{\min}$	$z_{0.25}$	$z_{0.50}$	$m_z$	$z_{0.75}$	$z_{\max}$	$\hat{\sigma}_z$	$\hat{s}_z$	$\hat{k}_z$
9.40 masl	20.50 masl	24.25 masl	32.05 masl	40.2 masl	62.00 masl	12.4 masl	0.81	2.58

The *Skewness* coefficient  $\hat{s}_z$  is estimated based on the third central moment according to the equation below:

$$\hat{s}_z = \frac{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{m}_z)^3}{\hat{\sigma}_z^3}, \quad (3.1)$$

where  $\hat{m}_z$  the sample mean and  $N$  the number of sampling points. The *skewness* coefficient of the normal distribution is equal to zero (Nist/Sematech 2009).

The *Kurtosis* coefficient is specified from the fourth central moment and defined by the following equation:

$$\hat{k}_z = \frac{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{m}_z)^4}{\hat{\sigma}_z^4}. \quad (3.2)$$

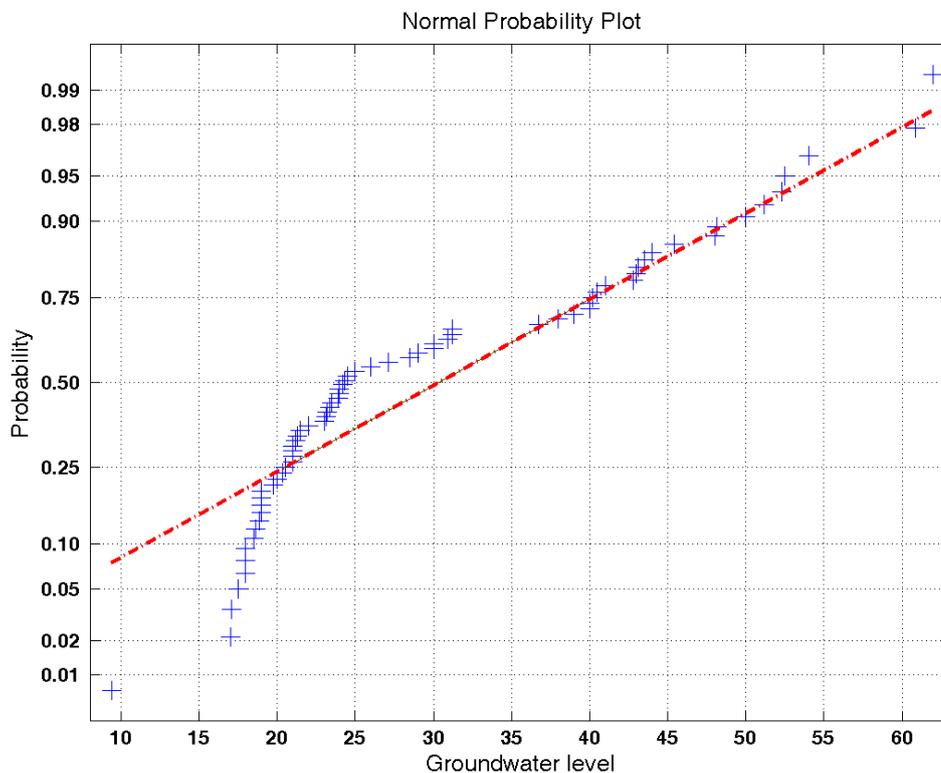
The *Kurtosis* coefficient of the normal distribution is equal to three (Nist/Sematech 2009).

Another way to test if the hydraulic head data follows the normal distribution is the Kolmogorov-Smirnov test (Massey Jr 1951, Kitanidis 1997). The Kolmogorov-Smirnov test is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test). The Kolmogorov-Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution for the one-sample K-S test. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution (in the one-sample case); the distributions considered under the null hypothesis are continuous distributions. The null hypothesis for the Kolmogorov-Smirnov test is that data follows the standard normal distribution. The test is applied comparing the normalized data values according to:

$$\frac{z_i - \hat{m}_z}{\sigma_z}, \quad (3.3)$$

with a normal distribution having mean  $\hat{m}_z=0$  and standard deviation  $\sigma_z=1$ . The test statistic is:  $D = \max |F(z) - G(z)|$ , where  $F(z)$  is the empirical cumulative distribution function (cdf) and  $G(z)$  is the standard normal cdf. If the sample comes from distribution  $F(z)$ , then  $D$  converges to 0 almost surely. The test is implemented in Matlab<sup>®</sup> environment using the function «*kstest*». For this dataset the null hypothesis is rejected at significant levels 5% and of 10%.

This is also confirmed by the normal probability plot presented in Figure 3.13. The purpose of a normal probability plot is to graphically assess whether the data could come from the normal distribution. If the data are normal the graph is linear; otherwise, the graph is curved as herein. In light of OK, it is known to be the optimal estimator if the data follow a multivariate normal distribution (Deutsch and Journel 1992). Therefore a series of normalizing methodologies are assessed in this thesis in order to transform the data closer to the normal distribution.

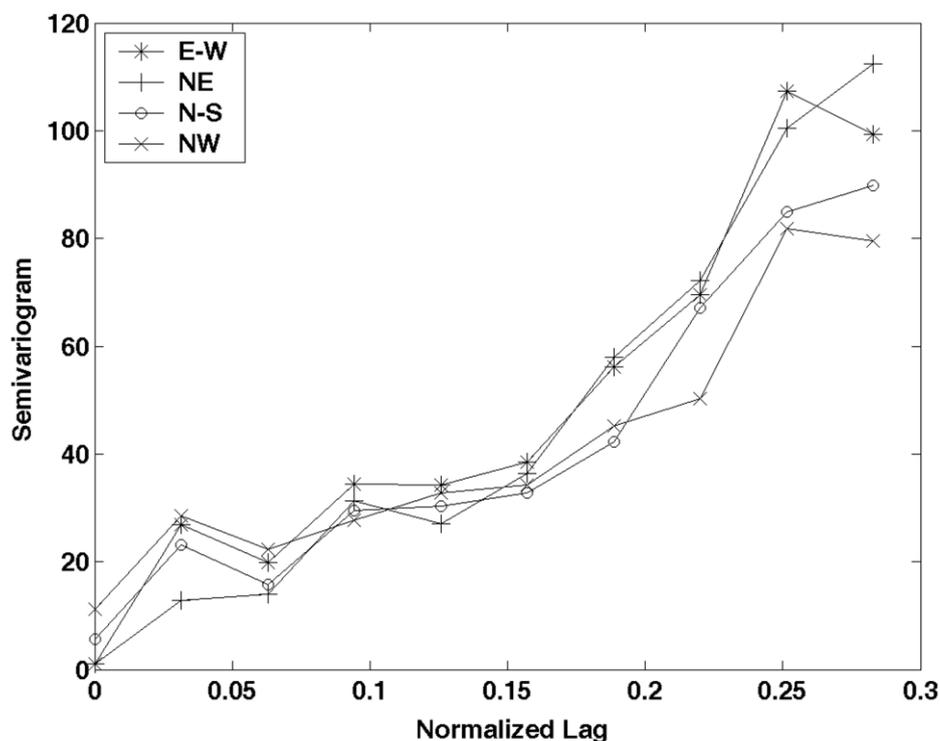


**Figure 3.13** Normal probability plot of the basin's groundwater level data for the hydrological year 2002-2003.

### 3.6 Anisotropy estimation

In the present study a test for geometric anisotropy is performed by comparing directional semivariograms in four main directions (Goovaerts 1997) using an angle tolerance of  $40^\circ$  for the semivariogram estimation. As shown in Figure 3.14, there are no distinct differences among the directional semivariograms. Therefore, the spatial variation of groundwater level is considered to be isotropic (Ahmadi and Sedghamiz 2007).

We also performed a test of geometric anisotropy based on the method of Covariance Hessian Identity (Hristopulos 2002, Chorti and Hristopulos 2008). This method is non-parametric, in the sense that it provides an estimate of the aspect ratio (i.e. the ratio of the two principal correlation lengths) and the orientation of the principal axes, without requiring semivariogram estimation and modeling. The aspect ratio is thus estimated at 0.806, while the short principal axis is rotated by  $6^\circ$  with respect to the E-W direction. The value of 0.806 does not differ significantly from unity. Indeed, the isotropic hypothesis cannot be rejected with 95% confidence for ratios in the range [0.73 – 1.37] using the anisotropy test given in (Spiliopoulos *et al.* 2011, Petrakis and Hristopulos 2012).



**Figure 3.14** Experimental directional semivariograms of groundwater level in Mires basin along the four main geographical directions, E–W, N–S, NE and NW.



## **4. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins-Application to Mires basin<sup>2</sup>**

### **4.1 Introduction**

The accurate mapping of groundwater levels in an aquifer is important for effective management and monitoring decisions. However, the number and spatial distribution of hydraulic head measurements are not always sufficient to accurately represent the groundwater levels in a given aquifer. Estimates at unsampled locations can be obtained by applying geostatistical and deterministic interpolation methods to the available data. This study aims to compare the performance of stochastic versus deterministic methods for mapping groundwater level in areas with sparsely distributed measurements and to specify additional observation locations where denser sampling is needed.

This chapter compares the interpolation performance of OK, UK and DK with the deterministic methods IDW and MC. The dataset used involves groundwater levels in a sparsely gauged basin. Measuring the relative performance of different interpolators is important for environmental monitoring.

### **4.2 Semivariogram estimation**

The omnidirectional empirical semivariogram of the hydraulic head fluctuations (after trend removal in UK) is determined using the method of moments. Anisotropy is not modeled since the directional semivariograms (not shown here) do not exhibit significant anisotropic dependence. The empirical semivariogram is fitted with isotropic classical models such as the exponential, Gaussian, spherical, power-law, and linear models (Deutsch and Journel 1992), the Matérn model (Matérn 1960, Stein 1999, Pardo-Iguzquiza and Chica-Olmo 2008), and the new family of Spartan variograms (Hristopulos 2003b, Hristopulos and Elogne 2007). For each of the above

---

<sup>2</sup> This chapter is an adaptation of a paper published in Environmental Monitoring and Assessment Journal (Varouchakis and Hristopulos, 2012), please see page 179 for details.

theoretical models, we determine the optimal semivariogram parameters that correspond to the stochastic methods tested (Table 4.1 and Table 4.2).

**Table 4.1** Optimal estimates of semivariogram model parameters obtained by a least squares fit to the experimental semivariogram of the data (Columns 2-4). The search radius defines the neighborhood used in the OK predictor (Column 5). The number of first- and second-order neighbors used by DK at each estimation point.

Model	Sill	$\xi$	Other parameters	OK search radius (normalized units)	DK No of neighbors
<b>Exponential</b>	133	0.30	NA	0.38	4 to 6
<b>Gaussian</b>	160	0.28	NA	0.38	4 to 6
<b>Spherical</b>	150	0.63	NA	0.59	4 to 6
<b>Power-law</b>	538	NA	$2H = 1.31$	0.59	4 to 6
<b>Linear</b>	331	NA	NA	0.38	4 to 6
<b>Matérn</b>	440	0.94	$\nu = 0.92$	0.59	4 to 6
<b>Spartan</b>	184	0.46	$\eta_1 = 1.12$	0.46	4 to 6

**Table 4.2** Optimal estimates of semivariogram model parameters obtained by a least squares fit to the experimental semivariogram of the residuals (Columns 2-4). The search radius defines the neighborhood used in the UK predictor (Column 5).

Model	Sill	$\xi$	Other parameters	UK search radius (normalized units)
<b>Exponential</b>	142	0.34	NA	0.38
<b>Gaussian</b>	211	0.35	NA	0.38
<b>Spherical</b>	137	0.69	NA	0.59
<b>Power-law</b>	500	NA	$2H = 1.39$	0.59
<b>Linear</b>	300	NA	NA	0.38
<b>Matérn</b>	236	0.66	$\nu = 0.87$	0.59
<b>Spartan</b>	169	0.75	$\eta_1 = 1.07$	0.59

## 4.3 Results and Discussion

### 4.3.1 Global cross validation measures

Table 4.3 presents the results for the cross-validation measures previously defined for each of the interpolation methods studied. IDW is applied using inverse square distance weights ( $n = 2$ ). This exponent value is widely used in geohydrology and also provides more accurate results for the Mires basin than other values. The

optimum search neighborhood consists of the four closest observation points to the estimation location. The MC method, implemented based on equations (2.30) and (2.31), uses the entire dataset for prediction.

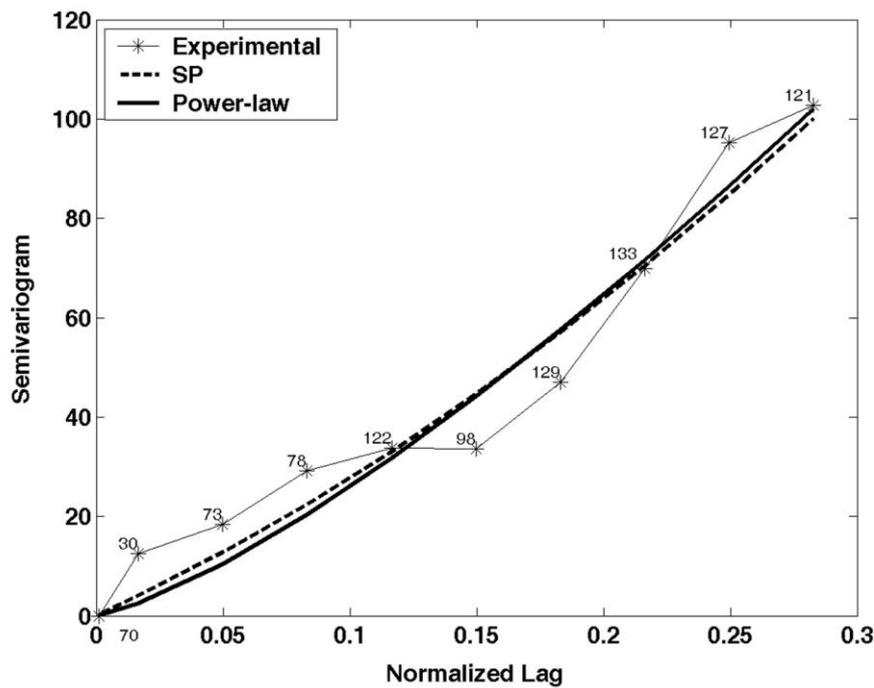
Figure 4.1 shows the empirical semivariogram and its fit with the optimal Spartan and power-law models that provide similar cross validation results for the OK and DK methods. The empirical semivariogram does not approach a sill, which is interpreted as lack of stationarity within the study area. The power-law model is non-stationary, while the Spartan model is stationary but approaches the sill outside the study area. The cross validation measures obtained with the above semivariogram models and with the best-fit Matérn model, which gives slightly inferior results, are shown in Table 4.3. The results obtained with UK using the same semivariogram models are also shown in Table 4.3.

**Table 4.3** Cross validation measures (section 2.5) for the stochastic and deterministic interpolation methods investigated. Results obtained with the three “optimal” (in terms of cross validation measures) semivariogram models are presented. The following abbreviations are used: IDW: Inverse distance weighted. MC: minimum curvature. DK: kriging with Delaunay triangulation. OK: Ordinary kriging. UK: Universal kriging. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram. MAE: Mean absolute error. MARE: Mean absolute relative error. RMSE: Root mean square error. R: Linear correlation coefficient. Optimal values are emphasized

<b>Method</b>	<b>MAE</b> (masl)	<b>BIAS</b> (masl)	<b>MARE</b>	<b>RMSE</b> (masl)	<b>R</b>
<b>IDW</b>	3.45	-0.17	0.15	5.58	0.89
<b>MC</b>	4.01	0.10	0.17	6.18	0.87
<b>DK-SP</b>	3.48	0.10	0.15	5.47	0.89
<b>DK-P</b>	3.48	0.14	0.15	5.52	0.87
<b>DK-M</b>	3.63	-0.08	0.15	5.74	0.89
<b>OK-SP</b>	<b>3.37</b>	<b>0.02</b>	<b>0.14</b>	<b>5.15</b>	<b>0.91</b>
<b>OK-P</b>	3.58	0.07	0.15	5.46	0.90
<b>OK-M</b>	3.80	<b>0.02</b>	0.16	5.84	0.89
<b>UK-SP</b>	3.40	0.13	<b>0.14</b>	5.23	<b>0.91</b>
<b>UK-P</b>	3.50	0.09	0.15	5.54	0.89
<b>UK-M</b>	3.8	0.09	0.15	5.78	0.89

In OK the Spartan semivariogram model (Figure 4.1) gives the most accurate estimates in terms of mean absolute error (MAE), i.e., 3.37 masl, compared to the power-law model which is a close second with 3.58 masl. The Spartan model is also

superior with respect to the other estimation measures (Table 4.3). DK has a MAE of 3.48 masl respectively, for both the Spartan and power-law semivariograms. However, as shown in Table 4.3 the validation measures obtained with the Spartan model are overall slightly better. The Spartan semivariogram is thus used in OK and DK interpolation. For OK a search radius equal to the characteristic length (0.46 in normalized units) yields the best cross validation results (Table 4.1). DK is applied using the first and second-order neighbors of the estimation point (Table 4.1), resulting in higher accuracy.

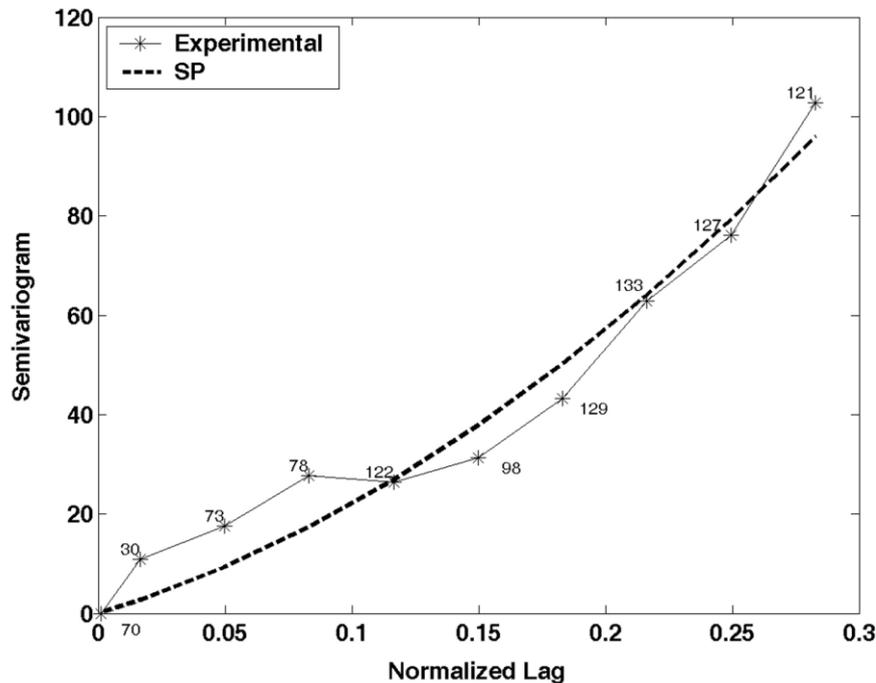


**Figure 4.1** Plot of omnidirectional experimental semivariogram of groundwater level data (stars), the optimal Spartan model (parameter estimates: variance  $\hat{\sigma}^2=184$ , characteristic length  $\hat{\xi}=0.46$ (normalized units), stiffness coefficient  $\hat{\eta}_1=1.12$ ), and the optimal power-law model ( $\hat{c}=538$ ,  $2\hat{H}=1.31$ ). Numbers of pairs used at each lag distance are also shown on the plot.

For the application of UK, the drift is approximated by  $m_z(\mathbf{s}) = k_1x + k_2y + k$ , where  $k_1 = 29.83$ ,  $k_2 = -11.14$ ,  $k = 23.13$  are the drift coefficients (constants) and  $\mathbf{s} = (x, y)$  are the space coordinates of the data. This is followed by a calculation of the semivariogram of the residuals. Leave-one-out cross validation (Table 4.3) shows that the Spartan model (Figure 4.2) delivers the most accurate results with respect to MAE, i.e., 3.40 masl, and performs overall better than the other “near-optimal” semivariogram models; the power-law model comes second with 3.50 masl.

Therefore, the Spartan semivariogram is applied in UK interpolation. The optimum search radius used with the Spartan model is equal to 0.59 (normalized units), which is somewhat shorter than the estimated characteristic length (Table 4.2). The power-law model in both OK and UK is also applied using an optimum search radius equal to 0.59 (Table 4.1, Table 4.2).

The cross validation measures (Table 4.3) show that no method performs significantly better than the others. OK gives uniformly the best results for the mean errors and the correlation coefficient followed, for most measures, by UK, DK and IDW in the order mentioned here. OK has clearly the lowest bias, very close to zero. MC generates a bias similar to DK and UK but lower than IDW; however, its performance is inferior with respect to other validation measures.



**Figure 4.2** Plot of omnidirectional semivariogram of residuals (stars) and optimal Spartan model (parameter estimates: variance  $\hat{\sigma}^2=169$ , characteristic length  $\hat{\xi}=0.75$  (normalized units), stiffness coefficient  $\hat{\eta}_1=1.07$ ). Residuals are derived by removing a linear drift. Numbers of pairs used at each lag distance are also shown on the plot.

The Spartan semivariogram model provides the most accurate cross validation estimates for the three stochastic methods investigated. In terms of MAE, OK-SP gives the most accurate estimate followed by UK-SP and DK-SP. The bias of OK-SP is very close to zero, while it is worse for DK-SP and UK-SP in the order stated. MARE and R are similar for OK-SP and UK-SP but slightly lower for DK-SP.

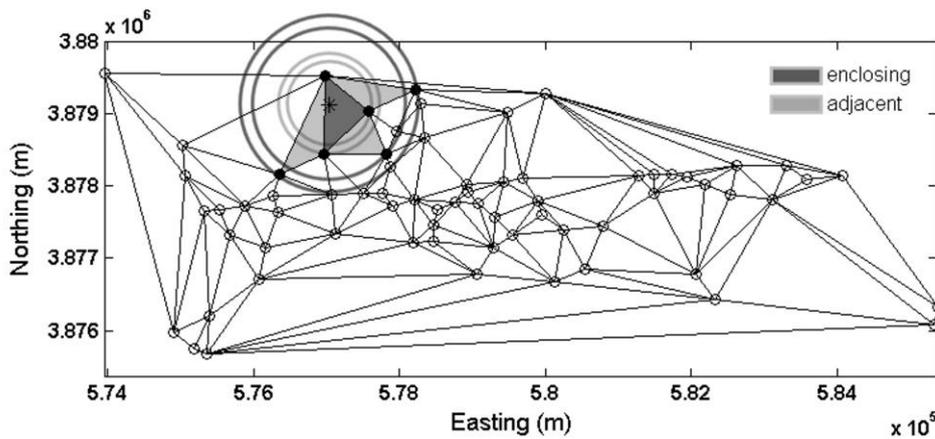
Finally, OK-SP yields the lowest RMSE followed by UK-SP and DK-SP. Overall, OK-SP provides the most accurate estimation measures compared to the other stochastic and deterministic methods.

#### 4.3.2 *Lowest-value estimation*

In addition to the global cross-validation results investigated above a statistic of interest is the estimation accuracy of the lowest groundwater level, i.e., 9.4 masl. DK-SP gives the most accurate estimate, i.e., 29 masl. DK estimates the lowest level 17% more accurately than OK-SP and UK-SP which yield 33 masl, and 17.6% more accurately than IDW, which yields 33.15 masl. In contrast, the highest level of 62 masl is accurately estimated by both the stochastic and deterministic interpolation methods. The superior performance of DK with respect to the lowest-value estimation is due to the local averaging property of DK.

DK is herein applied using both the first-order and second-order neighbors of the estimation point. At the location of the minimum, up to six neighbors are used in DK. The maximum distance from the neighbors (0.11 normalized units) is shorter than the estimated optimal radius for OK-SP and UK-SP interpolation (0.46, 0.59 normalized units, respectively). In order for OK-SP and UK-SP to approach the DK-SP optimal estimate (29 masl), they should be applied with a smaller estimation neighborhood. Using circular neighborhoods, the OK-SP and UK-SP neighbors of the minimum value location do not coincide with the DK-SP neighbors (see Figure 4.3). So, the optimal local neighborhood used by DK-SP is not feasible for OK-SP and UK-SP. IDW yields optimal global cross-validation results if applied with a circular neighborhood that encloses the four nearest neighbors but delivers an estimate of 33.15 masl for the lowest value.

In light of the above, we compare the minimum value estimation by means of OK-SP, UK-SP, and IDW with the same radii (0.05 to 0.13 normalized units, see Table 4.4). Leave-one-out cross validation results are shown in Table 4.4 using the Spartan semivariogram, which provides the most accurate estimates of the minimum compared to the other models tested, for all the methods (DK, OK and UK) and all search radii used (OK and UK). Most of the estimates in Table 4.4 are better than those obtained using “globally optimal” interpolation radii, which are derived by minimizing the mean absolute error over all the points.



**Figure 4.3** First- and second-order neighbors (full black circles) of extreme low value (at location marked by star) of the dataset located using Delaunay triangulation. The circles centered on the estimation point enclose the neighboring points for specific search radius (0.05 to 0.11 corresponding to normalized units) used for OK, DK and IDW for the extreme low value calculation.

**Table 4.4** OK-SP, UK-SP (SP: Spartan semivariogram) and IDW estimates (masl) of the extreme low value in the dataset using search radii (normalized units). The radius 0.11 leads to a neighborhood similar to DK which generates the most accurate estimate, 29 masl, of the extreme low value). The numbers in parentheses denote the number of Delaunay neighbors present inside the corresponding search radius, while (+) denotes the presence of other neighbors as well (see Figure 4.3). Delaunay neighbors (symbolized with full black circles in Figure 4.3) are the vertices of the enclosing triangle (dark color) and of the three adjacent triangles (grey color).

Search radius	0.05 (2) neighbors	0.06 (3) neighbors	0.09 (4+) neighbors	0.11 (5,6+) neighbors	0.13 (5,6+) neighbors
<b>OK-SP</b>	32.61	31.93	32.36	31.50	33.23
<b>UK-SP</b>	32.25	31.71	32.31	31.34	33.15
<b>IDW</b>	33.80	33.23	32.87	31.98	32.27

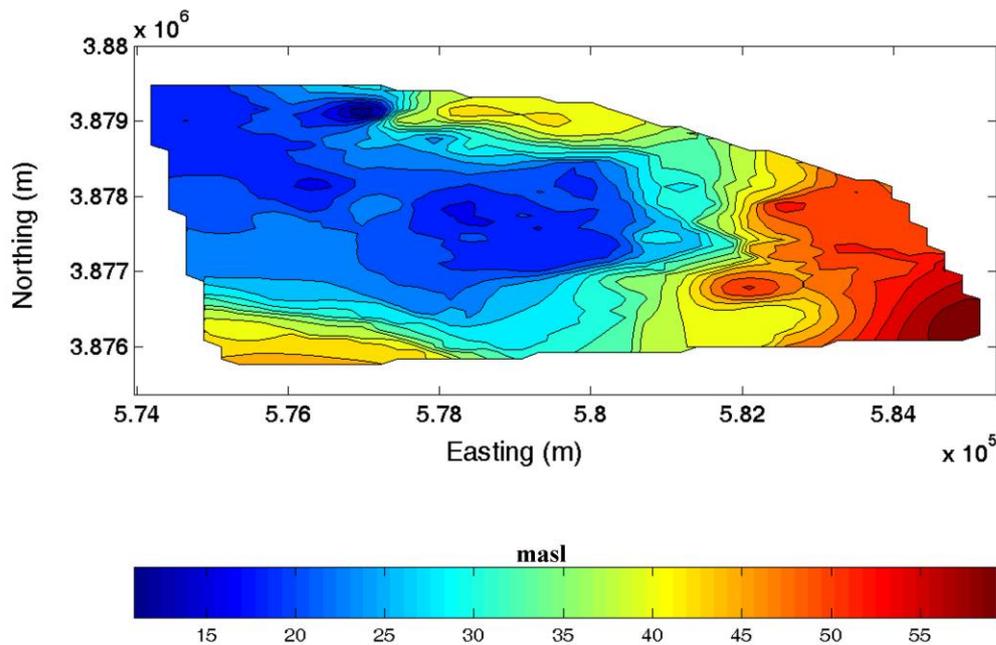
DK-SP estimation based on the first-order neighbors (i.e., the vertices of the triangle enclosing the lowest-value location), provides the same accuracy (31.93 masl) as OK-SP, but inferior than UK-SP (see Table 4.4). By increasing the search radius of OK-SP, UK-SP and IDW the second-order neighbors are progressively included (see Figure 4.3). The cross validation performance of OK-SP, UK-SP and IDW improves (see Table 4.4) approaching the optimal of DK-SP (29 masl) as the neighborhood of the latter is closely matched (i.e., for search radius of 0.11 normalized units). The best estimate is obtained with UK-SP (31.34 masl).

The optimal radius for the lowest value estimation is not generally suitable for OK, UK and IDW interpolation, because it generates search neighborhoods that do not include any neighbors around some of the data points. The smallest search radius

that leads to at least one neighbor for each data point is equal to 0.13 (normalized units). However, this value delivers inferior cross validation measures for OK-SP, UK-SP and IDW compared to the respective optimal radii. A radius of 0.13 (normalized units) provides a better IDW estimate (32.27 masl) of the minimum than the optimal neighborhood (33.15 masl). In contrast, the OK-SP and UK-SP estimates (33.23 and 33.15 masl respectively) are inferior to those obtained with the “globally optimal” interpolation radii (33 masl).

#### 4.3.3 Isolevel contour maps of hydraulic head

Next, we generate isolevel contour maps of the groundwater surface in the basin. We use interpolated values of the hydraulic head on a  $100 \times 100$  grid (actual cell size:  $114 \times 47$  m). Only grid points inside the convex hull (7317 grid points) of the sampling network are given numerical values, to ensure that the interpolated field is based on sufficient information. The contour maps generated are shown in Figure 4.4 to Figure 4.11.



**Figure 4.4** Isolevel contour map of estimated groundwater level in Mires basin using IDW.

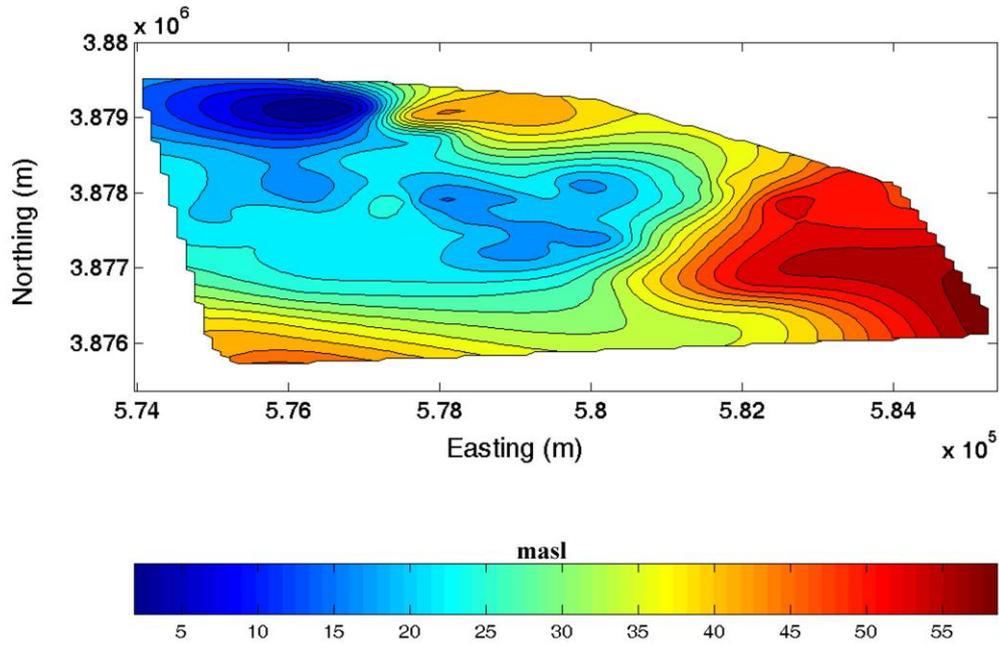


Figure 4.5 Isolevel contour map of estimated groundwater level in Mires basin using MC.

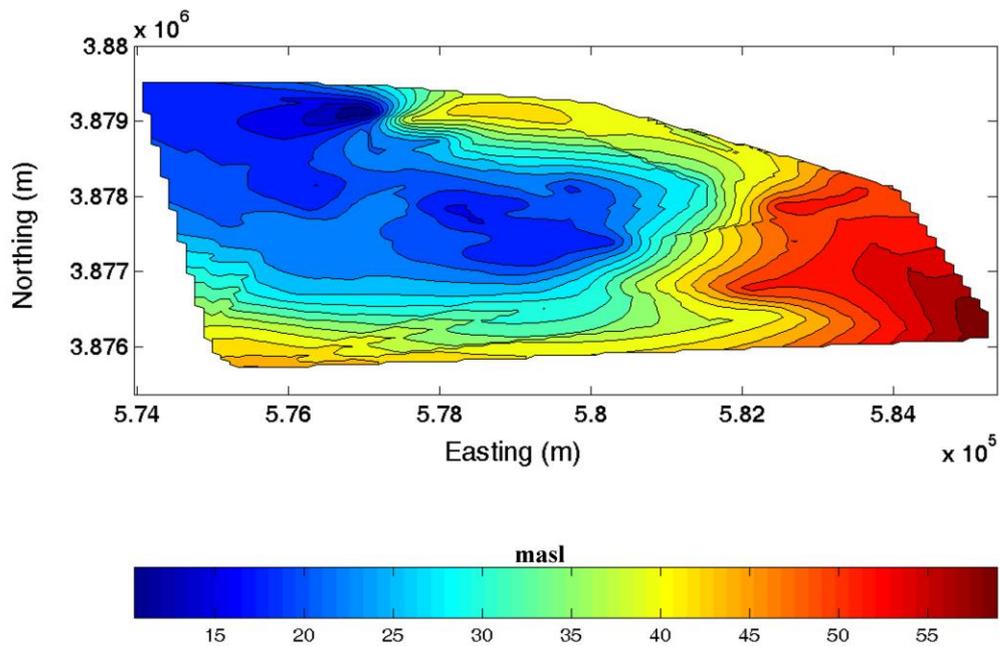
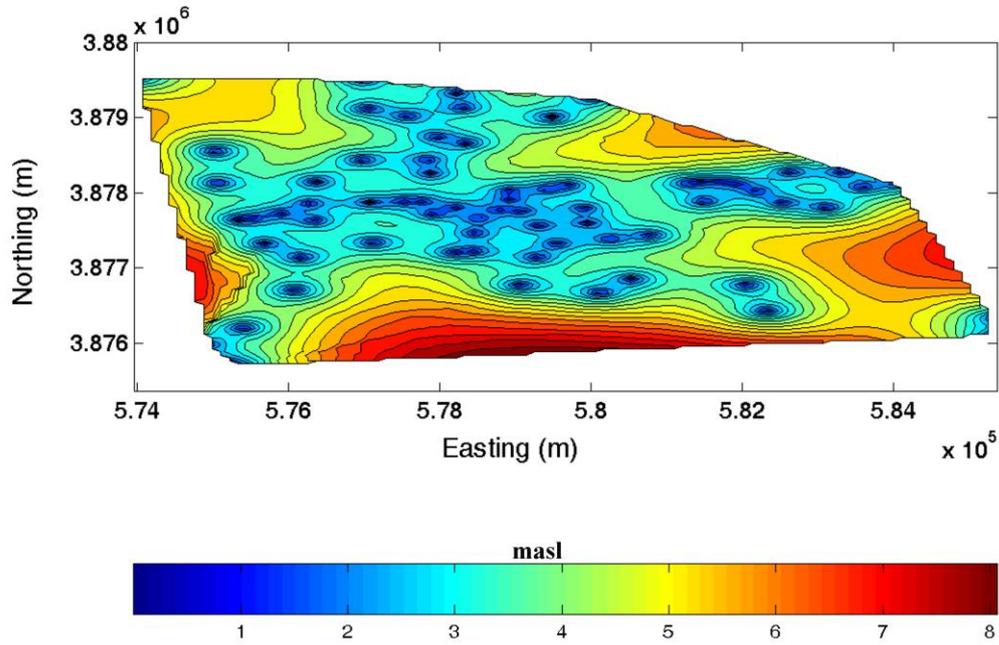
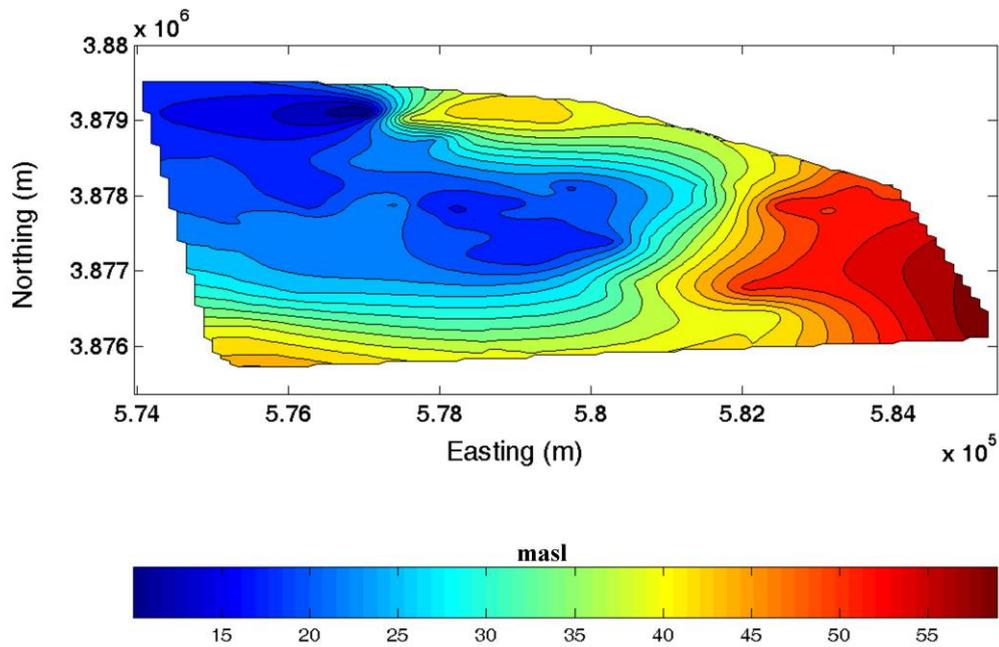


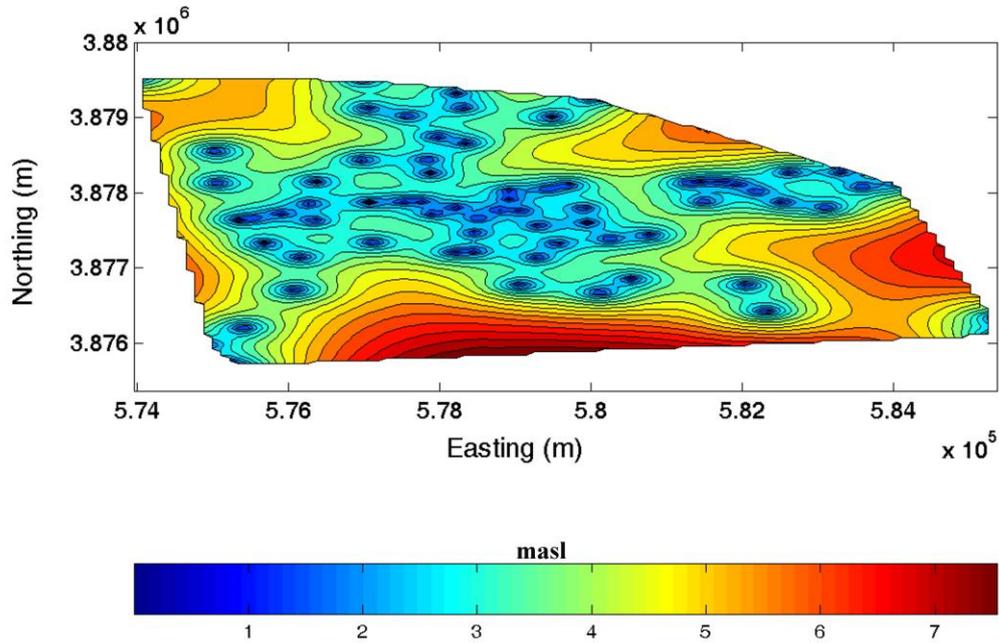
Figure 4.6 Isolevel contour map of estimated groundwater level in Mires basin using DK with the Spartan semivariogram model.



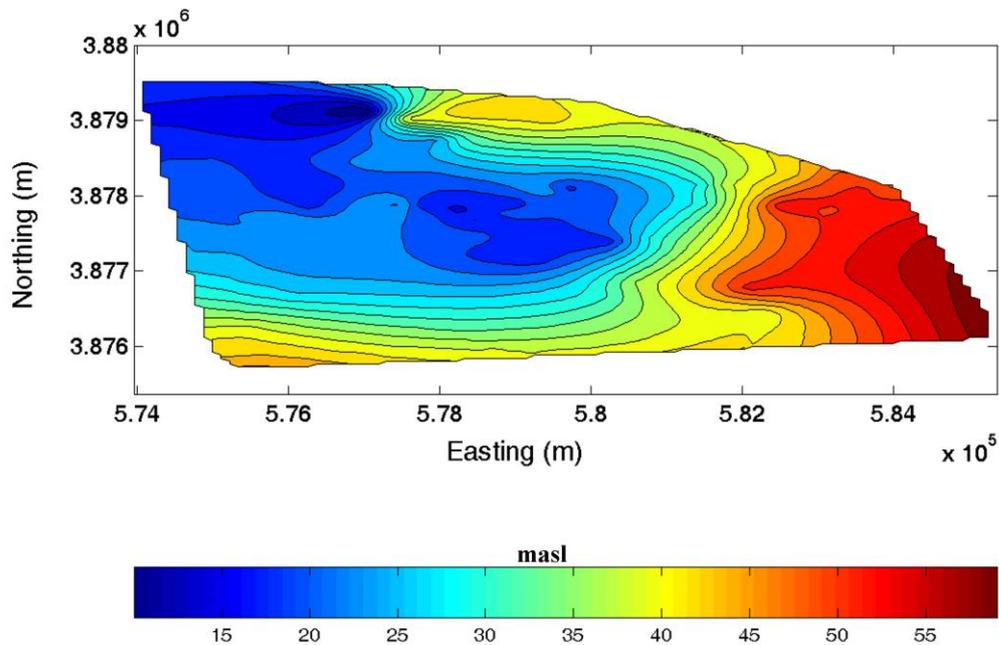
**Figure 4.7** Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using DK with the Spartan semivariogram model.



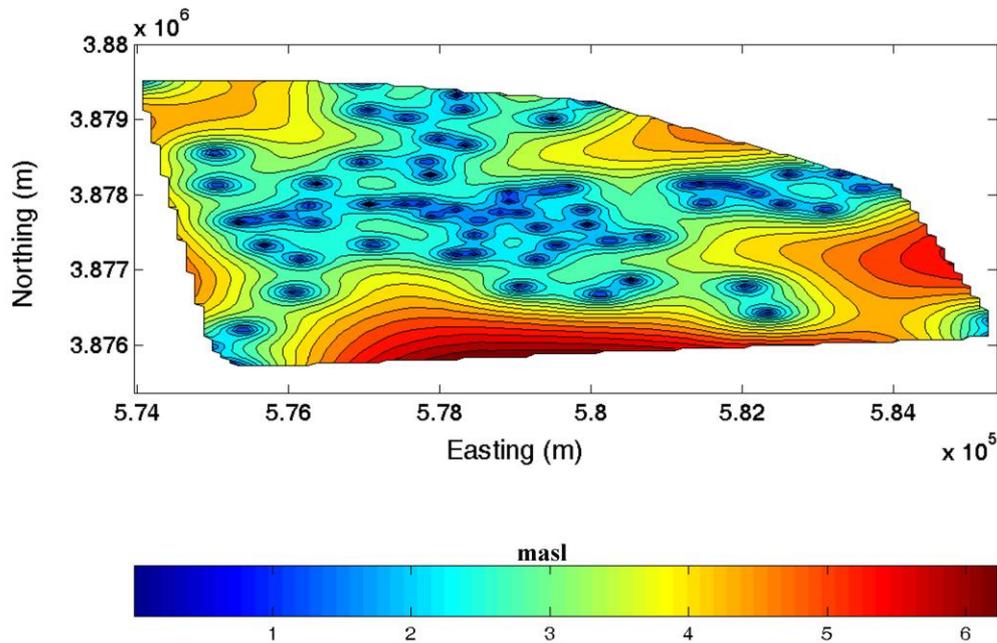
**Figure 4.8** Isolevel contour map of estimated groundwater level in Mires basin using OK with the Spartan semivariogram model.



**Figure 4.9** Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using OK with the Spartan semivariogram model.



**Figure 4.10** Isolevel contour map of estimated groundwater level in Mires basin using UK with the Spartan semivariogram model.



**Figure 4.11** Isolevel contour map of kriging standard deviation for groundwater level in Mires basin using UK with the Spartan semivariogram model.

IDW and DK contours are rougher than those estimated by means of the other methods. This is due to the fact that both methods use a small number of neighbors, leading to considerable variation of the estimates. MC, OK and UK lead to smoother contours. The smoothness of MC contours is due to the assumption of an underlying differentiable function. OK and UK yield very smooth contours, because their estimates are based on observations within a neighborhood defined by the large characteristic length (50% - 75% of the area's extent in normalized units).

OK, DK and UK interpolation maps are derived using the non-differentiable Spartan model. The power-law semivariogram, which is also non-differentiable, gives similar results to the Spartan model. Third best is a non-differentiable Matérn model with smoothness coefficient  $\nu < 1$ . (0.92 and 0.87 for original data and residuals respectively). Similarly, a non-differentiable semivariogram (spherical model), was used for the hydraulic head in a different study (Fasbender *et al.* 2008). We propose an explanation for the non-differentiability of the groundwater level surface. The data reflect the surface formed by the upper boundary of the saturated zone. We suggest that the height of this zone is determined by a deposition-removal process: locally

varying increments of water are added (e.g., due to precipitation) and removed (e.g., due to pumping and evapotranspiration) from the aquifer. Hence, the height at any given time results from the superposition of (both positive and negative) random increments. If the increments are approximately Gaussian, this process is expected to generate a non-stationary fractional Brownian motion (fBm) pattern (Mandelbrot and Van Ness 1968). In surface hydrology, fBm processes have been used to model the level of water reservoirs (Feder 1988). Hence, it is not coincidental that the fBm power-law semivariograms are very close to the best performing model in the OK, DK and UK cross validation procedure. Of course, for a finite-size basin the purely power-law fBm dependence should be truncated by the domain size. The non-differentiability of the groundwater level explains the poor performance of MC, which assumes a differentiable hydraulic head function, in comparison to the other methods.

#### 4.3.4 Estimation variance

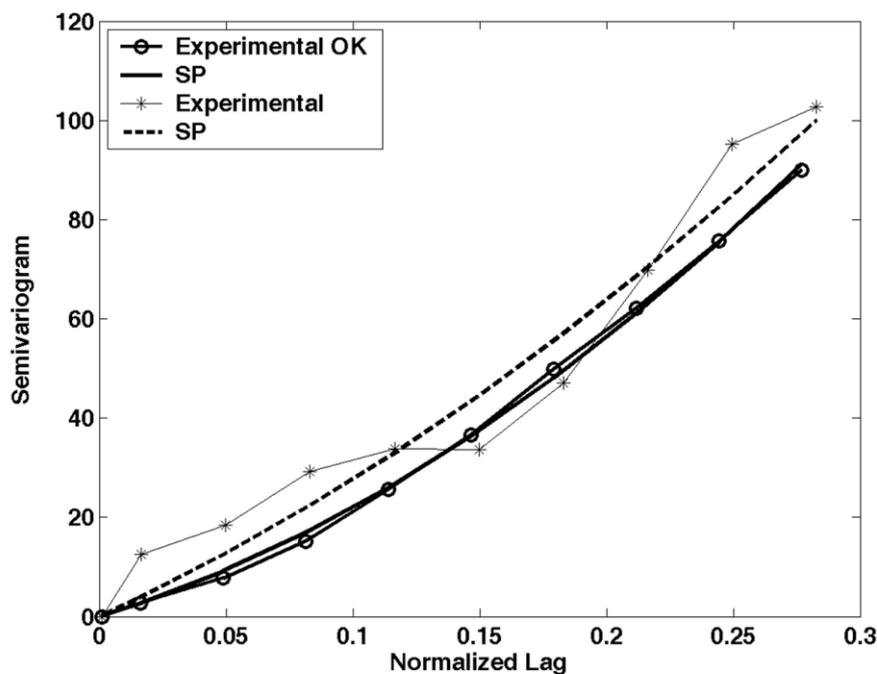
Stochastic interpolation methods quantify the kriging (error) variance, which determines the precision of the estimates. The map of kriging standard deviation (kriging error) can be used to identify locations where the estimates have high uncertainty and further sampling is needed (Prakash and Singh 2000, Fatima 2006, Theodossiou and Latinopoulos 2006, Yang *et al.* 2008).

The error maps (Figure 4.7, Figure 4.9, Figure 4.11) identify the locations of the Mires basin with the largest kriging standard deviation. The south and east borders of the basin can benefit from further sampling according to OK (standard deviation 6–7.5 masl) and UK (standard deviation 5–6.5 masl). DK shows a standard deviation range approximately between 6 and 8 masl at the same locations, but also similar values along the west border. The fact that DK is based only on three to six neighboring points often results in higher kriging variances than OK or UK. In Mires basin, most estimation points have more than six neighbors in their UK and OK search neighborhoods, thus reducing the OK and UK variances with respect to DK. UK delivers the lowest standard deviation, as it includes a linear trend function that reduces variability compared to OK. Interpolation with the Spartan semivariogram model delivers the lowest standard deviation for all three (OK, UK, DK) interpolation methods tested.

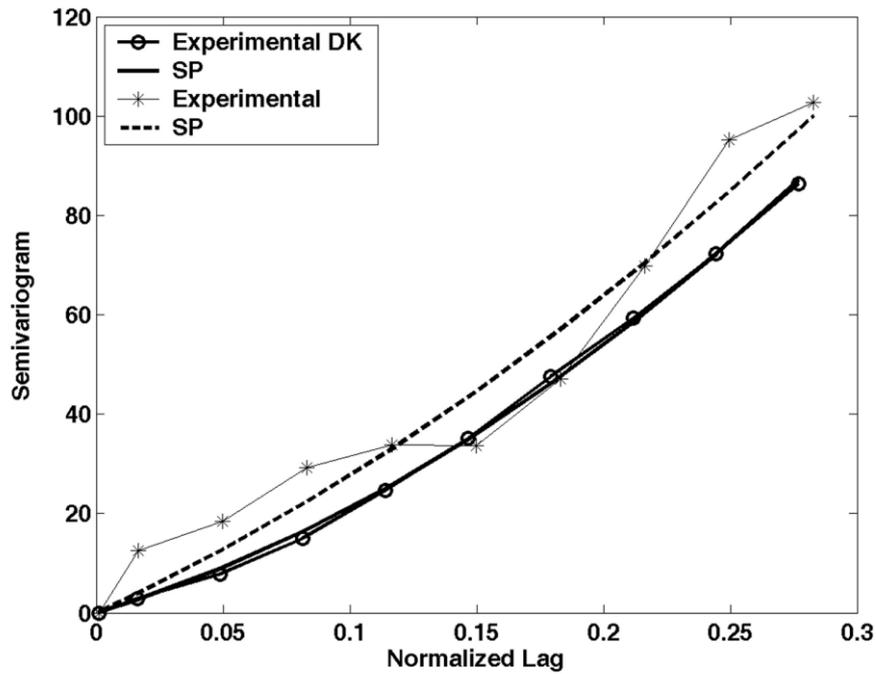
#### 4.3.5 Semivariogram validation

Cross validation studies mostly focus on univariate measures of performance, such as the ones described above. Stochastic interpolation methods also allow comparison of the empirical semivariogram with that obtained from the interpolated values, thus testing the accurate reproduction of spatial continuity by interpolation (Kitanidis 1997, Olea 1999).

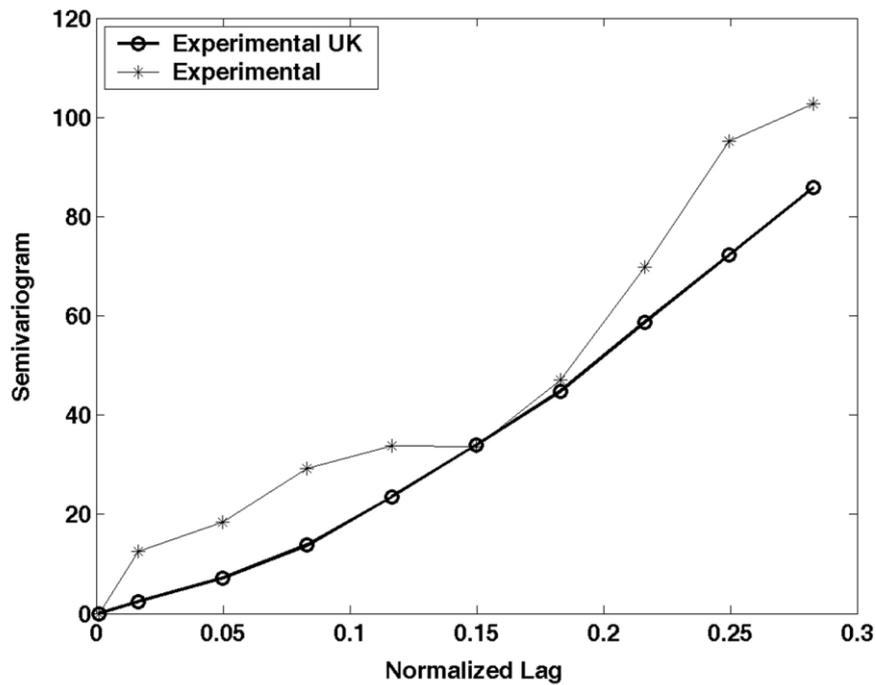
In Figure 4.12 and Figure 4.13, we compare (i) the experimental semivariogram of the observations (ii) the optimal theoretical model (iii) the experimental semivariograms obtained from OK- (Figure 4.12) and DK- (Figure 4.13) interpolation estimates and (iv) the respective optimal models. Figure 4.14 presents the experimental semivariograms of the observations and the UK interpolation estimates. Optimal theoretical models are not presented, because the semivariogram fit is performed for the residuals. In all cases, the semivariogram of the estimates shows very similar structure as that of the data. However, the former exhibit overall lower variability tending to have lower values than the empirical semivariogram of the data. This behaviour reflects the smoothing effect of interpolation.



**Figure 4.12** Comparison of groundwater level semivariograms: data (stars), OK estimates using Spartan (SP) semivariogram (circles), along with optimal SP model fits to data (dashed line) and to OK estimates (continuous line).



**Figure 4.13** Comparison of groundwater level semivariograms: data (stars), DK estimates using Spartan (SP) semivariogram (circles), along with optimal SP model fits to data (dashed line) and to DK estimates (continuous line).



**Figure 4.14** Comparison of omnidirectional groundwater level semivariograms of data (stars) and UK estimates (circles).

#### 4.3.6 General remarks

Stochastic and deterministic methods for the interpolation of groundwater levels have been used in other studies before. Below we briefly describe how this study differs from previous ones.

To our knowledge, this is the first application of DK to groundwater level interpolation. First- and second-order neighbors are used in DK to improve estimation accuracy. In contrast with studies that employ only first-order neighbors, we clearly present and apply the methodology for locating second-order neighbors. This paper also presents the recently developed Spartan semivariogram model for environmental applications. In the present study, this model is shown to be optimal for interpolation. The OK-SP and UK-SP methods employed in this manuscript apply the Spartan semivariogram model for the first time to hydrological data.

We compare three stochastic versus two deterministic methods for mapping groundwater level that have not heretofore been compared on the same data set. The case study in our manuscript investigates the performance of well-known methods with respect to interpolation in a sparsely gauged basin (Mires basin). To our knowledge, the groundwater level in Mires basin has not been modeled with geostatistical methods.

According to the five statistical measures assessed, no method performs extremely better than the others, with OK-SP to lead. The statistical measures can be also assessed by categories. Best absolute error method, lowest bias method, best statistical accuracy (MARE, RMSE) method and method with the higher cross validation correlation coefficient (R). However, the conclusion is the same; the results are not significantly different. Therefore, except of the estimation accuracy the choice of the best performing interpolation method should be also based on the prediction uncertainty, that is an advantage only of the stochastic methods.

The method comparison is based on cross validation measures, which include global statistical quantities, the accuracy of the minimum value estimate, estimation variance and semivariogram reproduction. The impact of search neighborhood effects on the validation results is analyzed in detail. The comparison is conducted in the Matlab<sup>®</sup> programming environment, using code developed by the author, as opposed to commercial software packages. This approach provides increased flexibility and common ground for comparison.

## 5. Improving kriging of groundwater level data using non-linear normalizing transformations-Application to Mires basin<sup>3</sup>

### 5.1 Introduction

Skewed or erratic data can often be made more suitable for geostatistical modeling by appropriate transformation. Such applications can lead to the reliable mapping of the free surface of an aquifer. A normal distribution for the variable under study is desirable in linear geostatistics (Clark and Harper 2000). Even though mild deviations from normality do not cause problems, significant deviations, e.g. due to high skewness and outliers, have an undesirable impact on the semivariogram structure and the kriging estimates (Gringarten and Deutsch 2001, Ouyang *et al.* 2006). OK is well-known to be optimal when the data have a multivariate normal distribution and the true semivariogram is known. Therefore, transformation of data may be required before kriging to normalize the data distribution, suppress outliers, and improve data stationarity (Deutsch and Journel 1992, Armstrong 1998). Then the estimation is performed in the Gaussian domain, before back-transforming the estimates to the original domain. An advantage of the Gaussian distribution is that spatial variability is easier to be modeled, because it reduces effects of extreme values providing more stable semivariograms (Goovaerts 1997, Armstrong 1998, Pardo-Iguzquiza and Dowd 2005). Kriging represents variability only up to the second order moment (covariance), therefore the random field of the transformed variable must be Gaussian to derive unbiased estimates at non-sampled locations (Deutsch and Journel 1992, Goovaerts *et al.* 2005). In practice, multi-normality is invoked as a working hypothesis.

The aim of this work is to investigate the improvement in groundwater level interpolation with OK using non-linear data normalization methodologies. Well-known OK based methodologies are applied and most of them for the first time to groundwater level data. In addition, a novel normalization method based on the Box-

---

<sup>3</sup> This chapter is an adaptation of a paper accepted for publication in Hydrological Sciences Journal (Varouchakis *et al.* 2012), please see page 179 for details.

Cox transformation, termed Modified Box-Cox is established and implemented in this article. Furthermore, the recently established Spartan semivariogram family is applied herein along with classical semivariogram models. The Modified Box-Cox (MBC) method, the Gaussian Anamorphosis (GA) normalization method and the Trans-Gaussian Kriging (TGK) method are implemented for the first time to groundwater level data. Overall, several kriging-based spatial models are investigated, evaluated, and maps of estimated water table elevation and its associated uncertainty are generated by means of the optimal model.

## 5.2 Box-Cox transformation method

The Box-Cox (BC) method (Box and Cox 1964) is widely used to transform hydrological data into approximately Gaussian distributions (Chander *et al.* 1978, Hirsch 1979, Jain and Singh 1986, Salas 1993, Thyer *et al.* 2002). The transform is defined only for positive data values and is defined by means of

$$y := g_{BC}(z; k) = \begin{cases} \frac{(z^k - 1)}{k}, & k \neq 0 \\ \log(z), & k = 0 \end{cases} . \quad (5.1)$$

Given the vector of data observations  $\mathbf{z}^T = (z_1, \dots, z_N)$ , the optimal value of the power exponent  $k$ , which leads to the best agreement of  $\mathbf{y}^T = (g_k(z_1), \dots, g_k(z_N))$  with the Gaussian distribution, can be determined by means of the maximum likelihood estimation method (De Oliveira *et al.* 1997). The power exponent  $k$  is estimated by maximizing the logarithm of the likelihood function:

$$f(\mathbf{z}; k) = -\frac{N}{2} \ln \left[ \sum_{i=1}^N \frac{(g_{BC}(z_i; k) - \bar{g}_{BC}(\mathbf{z}; k))^2}{N} \right] + (k-1) \sum_{i=1}^N \ln(z_i), \quad (5.2)$$

where  $\bar{g}_{BC}(\mathbf{z}; k) = \frac{1}{N} \left[ \sum_{i=1}^N g_{BC}(z_i; k) \right]$  is the arithmetic mean of the transformed data whilst the sum of squares  $\left[ \sum_{i=1}^N \frac{(g_{BC}(z_i; k) - \bar{g}_{BC}(z_i; k))^2}{N} \right]$  denotes the transformed data variance.

### 5.3 Trans-Gaussian Kriging (TGK)

Trans-Gaussian Kriging is more general than the Box-Cox transformation (Cressie 1993, Kozintseva 1999, Schabenberger and Gotway 2005). For a nonlinear normalizing transformation,  $Y(\mathbf{s}) = g(Z(\mathbf{s}))$ , where  $Y(\mathbf{s})$  follows the multivariate Gaussian distribution, assume that  $Z(\mathbf{s}) = \varphi(Y(\mathbf{s}))$ , where  $\varphi(\cdot) = g^{-1}(\cdot)$  is a one-to-one, twice-differentiable function. It is also assumed that  $Y(\mathbf{s})$  is an intrinsically stationary SRF with mean  $m_Y$  and semivariogram  $\gamma_Y(\mathbf{r})$ . For an unknown  $m_Y$ , the OK predictor,  $\hat{Y}_{OK}(\mathbf{s}_0)$ , is used to predict  $Y(\mathbf{s}_0)$ . An estimate of  $Z(\mathbf{s}_0)$  is then given by  $\hat{Z}(\mathbf{s}_0) = \varphi(\hat{Y}_{OK}(\mathbf{s}_0))$ , where  $\varphi(\cdot)$  is the inverse of the transformation function. However, this results in a biased predictor if  $\varphi(\cdot)$  is a nonlinear transformation. A bias-correcting approximation is the trans-Gaussian predictor (Cressie 1993):

$$\hat{Z}(\mathbf{s}_0) = \varphi(\hat{Y}_{OK}(\mathbf{s}_0)) + \frac{\varphi''(\hat{m}_Y)}{2} [\sigma_{OK;Y}^2(\mathbf{s}_0) - 2\mu_Y], \quad (5.3)$$

where  $\hat{m}_Y$  is the OK estimate of  $m_Y$ ,  $\mu_Y$  is the Lagrange multiplier of the OK system,  $\varphi''(\cdot)$  is the second-order derivative of the inverse transformation function, and  $\sigma_{OK;Y}^2(\mathbf{s}_0)$  is the OK variance. If the Box-Cox normalizing transformation (5.1) is used, as herein, the functions  $\varphi(\cdot)$  and  $\varphi''(\cdot)$  have the following form:

$$\varphi(y) = (ky + 1)^{1/k}, \quad (5.4)$$

$$\varphi''(\hat{m}_Y) = (1-k)(\hat{m}_Y k + 1)^{\frac{1}{k}-2}. \quad (5.5)$$

## 5.4 Gaussian Anamorphosis (GA)

This method is based on the transformation of a Gaussian variable  $Y$  into a new variable  $Z$  with an arbitrary distribution by means of  $Z = \Phi_{GA}(Y)$ , where  $\Phi_{GA}(\cdot)$  is the Gaussian anamorphosis transformation. The transformation used in GA involves the following polynomial expansion (Chiles and Delfiner 1999):

$$\Phi_{GA}(Y) = \sum_{i=0}^K \Psi_i H_i(Y), \quad (5.6)$$

where the functions  $H_i(Y)$ ,  $i=0, \dots, K$  are *Hermite Polynomials* and  $\Psi_i$  denote the coefficients of the expansion. The *Hermite polynomials* are defined in terms of the derivatives of the Gaussian density function:

$$H_i(x) = \frac{g^{(i)}(x)}{g(x)}, \quad (5.7)$$

where  $g(x)$  is the zero-mean and unit variance Gaussian density function, i.e.,

$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  and  $g^{(i)}(x)$  is the  $i$ th-order derivative of  $g(x)$ . The *Hermite polynomials* are calculated by means of the following recurrence relation:

$$H_{i+1}(x) = -xH_i(x) - iH_{i-1}(x), \quad i \geq 0. \quad (5.8)$$

Typically, a high polynomial order ( $K=30-100$ ) is used. Model fitting consists of estimating the coefficients  $\Psi_i$ . The normalization of a non-Gaussian variable requires the inversion of the anamorphosis function, by means of  $Y = \Phi_{GA}^{-1}(Z)$ . The geostatistical analysis is performed on the transformed variable  $Y$ , and the estimates are finally back-transformed to the original values through the anamorphosis function (Olea 1999, Wackernagel 2003, Casa and Castrignano 2008).

Practically any function of  $Y$  which is square integrable with respect to the Gaussian density can be expanded in terms of *Hermite polynomials*. The coefficients of the expansion are given by the following equation (Journel and Huijbregts 1978, Wackernagel 2003):

$$\Psi_i = \int_{-\infty}^{\infty} \Phi_{GA}(x) H_i(x) g(x) dx. \quad (5.9)$$

For the field application studied in this thesis the expansion coefficients  $\Psi_i$  are estimated for the linear, polynomial and exponential functions. The function  $\Phi_{GA}$  that gives the best fit to the data is the quadratic function  $\Phi_{GA}(x) = x^2$ . For the quadratic, the integral (5.9) used to estimate  $\Psi_i$  is solved analytically for any *Hermite polynomial* using integration by parts. As an example, for the second-order *Hermite polynomial*, equation (5.9) becomes:

$$\Psi_2 = \frac{1}{\sqrt{2\pi}} \left( \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} dx - \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx \right) = 2. \quad (5.10)$$

In general, the solution of the integral is:

$$a_n = \int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} (n-1)!! = \sqrt{2\pi} (n-1)(n-3)\cdots 3 \times 1, \quad (5.11)$$

for  $n$  even, while  $a_n = 0$  for odd  $n$ . Hence, the corresponding expansion coefficients  $\Psi_i$  vanish for Hermite polynomials of odd order.

### 5.5 Modified Box-Cox (MBC)

This new method focuses on normalizing the skewness and kurtosis coefficients of the data, but it neglects higher-order moments. It is defined by the following function:

$$y := g_{MBC}(z; \mathbf{\kappa}) = \frac{(z - z_{\min} + k_2^2)^{k_1} - 1}{k_1}, \quad \mathbf{\kappa}^T = (k_1, k_2), \quad (5.12)$$

where  $k_1$  is the power exponent and  $k_2$  is an offset parameter. Use of the latter allows negative  $z$  values and so the transformation (5.12) can be applied to fluctuations as well. Parameters  $(k_1, k_2)$  are estimated from the numerical solution of the equations  $\hat{s}_z = 0, \hat{k}_z = 3$ , where  $\hat{s}_z$  and  $\hat{k}_z$  are the sample skewness and kurtosis coefficients respectively,

$$\left( \frac{\hat{m}_z - \tilde{m}_z}{\sigma_z} \right)^2 + (\hat{k}_z - 3)^2 \simeq 0, \quad (5.13)$$

where  $\tilde{m}_z$  is the sample's median. The minimization is performed using the Nelder-Mead simplex optimization method (Nelder and Mead 1965, Press *et al.* 1992).

## 5.6 Results and discussion

Three general approaches are tested for interpolation. The first approach applies OK with the optimal semivariogram function to the original data. The second approach first applies a normalizing transformation (Box-Cox, MBC, GA), then applies OK on the transformed variable, and finally it back-transforms the predictions. The third one employs TGK using the Box-Cox transform. The application of transformation methods improves the data normality as can be seen in Table 5.1.

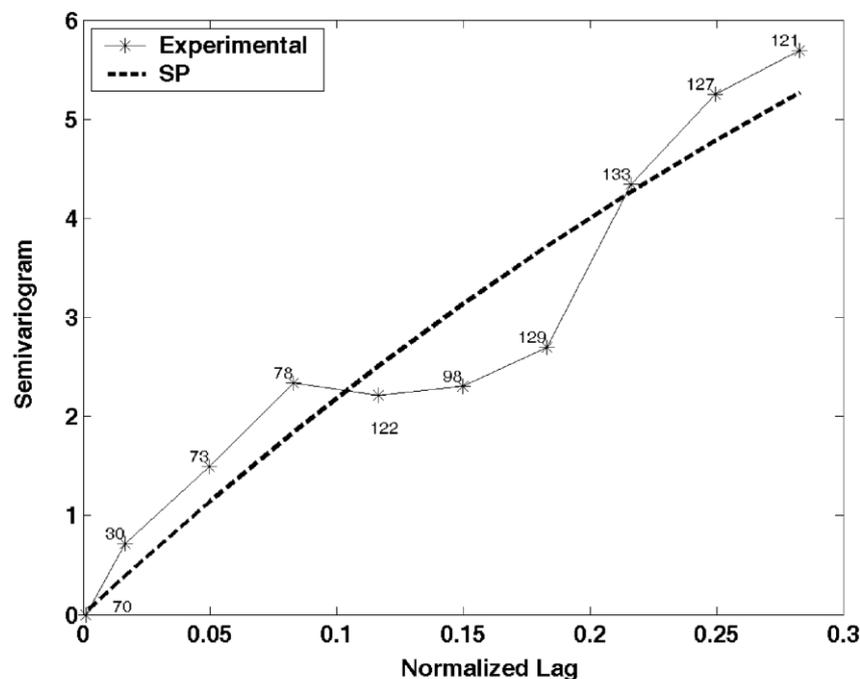
**Table 5.1** Normalization results using Box-Cox, Modified Box-Cox (MBC) and Gaussian Anamorphosis (GA) transformations: skewness coefficient  $\hat{s}_z$ ; kurtosis coefficient  $\hat{k}_z$ .

Method	$\hat{s}_z$	$\hat{k}_z$
Box-Cox	0.01	2.70
MBC	0.13	2.99
GA	0.10	2.87

The results of leave-one-out cross-validation are shown in Table 5.2, which shows that no transformation method is significantly superior. The best semivariogram fit, in terms of cross-validation results, is obtained by means of the

Spartan model for all the kriging-based methodologies assessed (Table 5.2). The power-law semivariogram gives similar results to the Spartan model. Third best is the Matérn model.

The Box-Cox transformation with  $k = -0.22$  and the GA transformation in combination with OK improve the mean absolute prediction error (3.30 masl) compared to OK (3.37 masl). TGK using the Box-Cox transform and the Spartan semivariogram performs best in terms of the Mean Absolute Error (MAE), yielding a value of 3.28 masl. All three methodologies have greater (in absolute value) bias than OK (0.02 masl), and similar value for the other estimation measures. Although MBC-OK ( $k_1 = 0.51, k_2 = -0.0001$ ) with the Spartan semivariogram model (shown in Figure 5.1) has a slightly larger MAE (3.30 masl) than TGK, MBC-OK provides overall the most accurate cross-validation results (Table 5.2). The parameters of the Spartan semivariogram are  $\hat{\sigma}^2 = 13.4$ ,  $\hat{\xi} = 0.42$  (in normalized units), and  $\hat{\eta}_1 = 0.97$ . The MBC-OK method with the above parameters improves the MAE and the Root Mean Square Error (RMSE) compared to OK while its bias, Mean Absolute Relative Error (MARE) and correlation coefficient R are identical to OK.



**Figure 5.1** Plot of omnidirectional experimental semivariogram of transformed (MBC) groundwater level data (stars) and optimal Spartan model fit (SP), (parameter estimates: variance  $\hat{\sigma}^2 = 13.4$ , characteristic length  $\hat{\xi} = 0.42$  (normalized units), stiffness coefficient  $\hat{\eta}_1 = 0.97$ ). Numbers of pairs used at each lag distance are also shown.

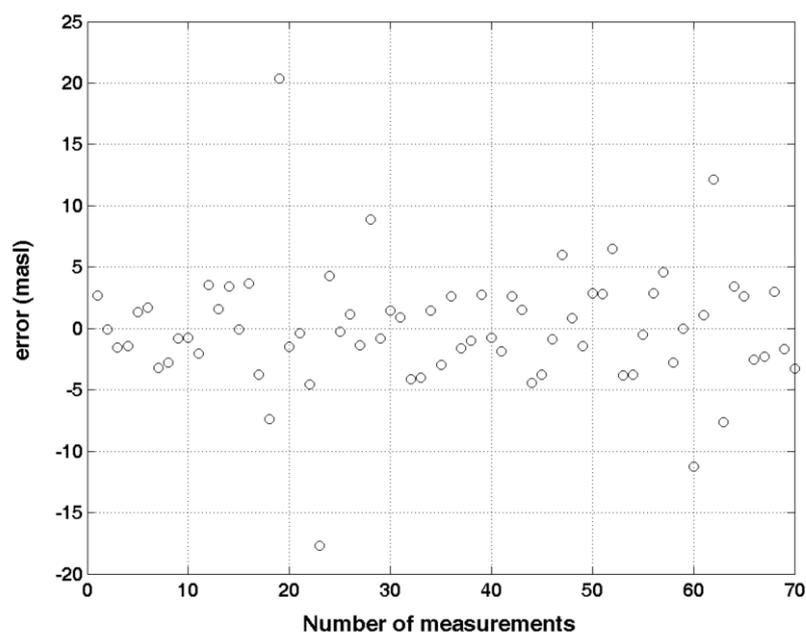
The scale factor  $\eta_0$  of the Spartan semivariogram is equal to 147.20 (dimensionless). This is obtained by equating the semivariogram sill with  $\sigma_z^2$ ; the latter is given by (2.24) for  $|\eta_1| < 2$ , based on the estimate  $\hat{\eta}_1=0.97$ . The scale factor of the transformed data is dimensionless, because the initial data are rendered dimensionless by dividing with an arbitrary constant that has the same units as the data (one meter in the present case study). In the back-transform stage the hydraulic head estimates are multiplied by this constant. If we use a unit constant (as we do herein), these operations have no impact on the number values, they just ensure that the final estimates are in the correct units.

**Table 5.2** Cross-validation results of the spatial models with the optimal semivariograms based on the measures listed in section 2.5. **OK**: Ordinary Kriging. **Box-Cox-OK**: Box-Cox transformation followed by OK and back-transformation. **MBC-OK**: Modified Box-Cox transformation followed by OK and back-transformation. **GA-OK**: Gaussian Anamorphosis in combination with OK. **TGK**: Trans-Gaussian Kriging using the Box-Cox transform. **SP**: Spartan semivariogram. **P**: Power-law semivariogram. **M**: Matérn semivariogram. **MAE**: Mean Absolute Error. **MARE**: Mean Absolute Relative Error. **RMSE**: Root Mean Square Error. **R**: Linear correlation coefficient. Optimal values are emphasized.

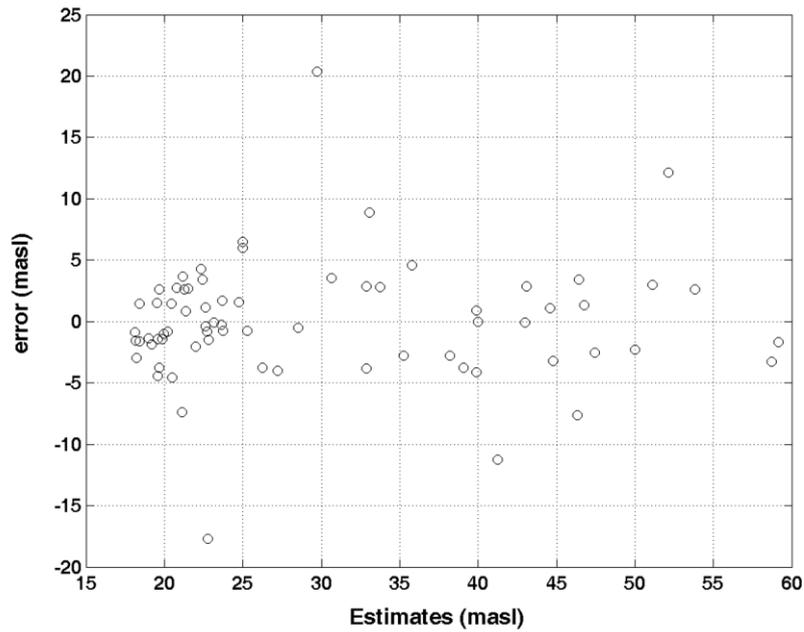
NT Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
<b>OK</b>	SP	3.37	0.02	0.14	5.15	0.91
	P	3.58	0.07	0.15	5.46	0.90
	M	3.80	0.02	0.16	5.84	0.89
<b>Box-Cox-OK</b>	SP	<b>3.30</b>	0.10	0.14	5.14	0.91
	P	3.41	0.09	0.14	5.31	0.90
	M	3.60	-0.30	0.15	5.65	0.89
<b>MBC-OK</b>	SP	<b>3.30</b>	<b>0.02</b>	<b>0.14</b>	<b>5.12</b>	<b>0.91</b>
	P	3.39	0.05	0.14	5.17	0.90
	M	3.60	0.03	0.15	5.54	0.89
<b>GA-OK</b>	SP	<b>3.30</b>	-0.3	0.14	5.14	0.90
	P	3.32	-0.4	0.14	5.21	0.89
	M	3.48	-0.59	0.14	5.54	0.89
<b>TGK</b>	SP	<b>3.28</b>	-0.1	0.14	5.14	0.91
	P	3.35	-0.13	0.14	5.19	0.90
	M	3.43	-0.2	0.14	5.48	0.90

Therefore, the MBC-OK method is optimal for interpolating groundwater levels in the Mires basin. To further support this choice, a series of specialized statistical measures are investigated. The correlation coefficient of the estimates vs. the true values is equal to 0.91 (Table 5.2), while the distribution of errors is symmetric with a low bias equal to 0.02 masl (Figure 5.2 a). The plot of errors vs. estimates (Figure 5.2 b) is centered about zero error, satisfying the “conditional unbiasedness” property. According to (Leuangthong *et al.* 2004), cross-validation that yields such results satisfies the conditions for a model with “good” parameters. Histogram reproduction is another measure for evaluating the spatial model performance (Leuangthong *et al.* 2004).

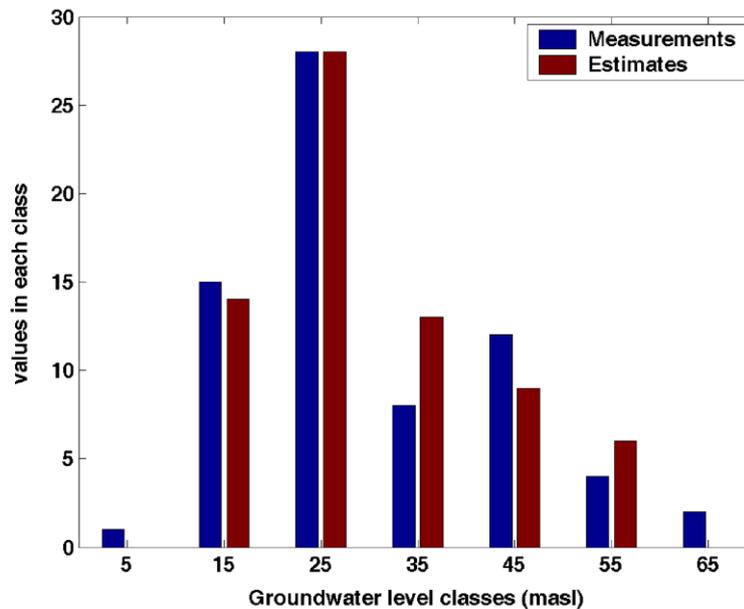
Figure 5.3 shows that the histograms of the measurements and the cross-validated values are overall in good agreement. In general, the spatial model tends to overestimate the lower values while the opposite is true for the higher values. MBC-OK (with the Spartan model) gives the most accurate estimate, i.e., 29.7 masl for the extreme low level in the data set (9.4 masl). Thus it determines the low level with  $\approx 6\%$  higher accuracy than the second best Box-Cox-OK and TGK, which yield  $\approx 31$  masl and  $\approx 17\%$  higher accuracy than the OK and GA-OK, which yield 33 masl. In contrast, the highest level of 62 masl is estimated with higher accuracy  $\approx 60$  masl by all interpolation methodologies tested.



**Figure 5.2 a** Distribution of MBC-OK cross-validation errors.



**Figure 5.2 b** MBC-OK cross-validation estimates and their corresponding errors.



**Figure 5.3** Histograms of measured values and MBC-OK cross-validation estimates using seven classes of width 10 masl each, centered at 5, 15, 25, 35, 45, 55, and 65 masl respectively.

The extreme values in this dataset include the global minimum (9.4 masl), as well as three local extremes that differ significantly from their measured neighbors. These values are estimated with relatively large errors (see Figure 5.2 a,b) due to the significant deviations of the measurements from the values of their nearest neighbours. These errors may be due to the presence of fractures near the measurement-well (in the case of the minimum) or to locally different inter-bedded

sequence of alluvial sediments that can affect the water table head. The highest estimation error (Figure 5.2 a,b) is associated with the minimum value of 9.4 masl, which differs from its neighbors by at least 10 masl.

As shown in the histogram of Figure 5.3, the tails of the measured histogram are wider than those of the estimates. The discrepancy in the lower tail has been explained above. The underestimation of the high values (i.e., the values exceeding 60 masl) is due to the smoothing effect of kriging and the fact that only two such values exist in the dataset. The impact of the extreme values is mostly noticeable in the cross-validation analysis, because the values in question are removed and then estimated from their neighbors. The generated maps, however, are based on all the measurements and thus not affected by the removal of local extreme values.

A series of statistical metrics are used to compare the performance of different interpolation methods. MAE is a linear score, meaning that the errors at all points are equally weighted in the average. On the other hand, RMSE is a quadratic scoring rule, i.e., the errors are squared before averaging, thus leading to relatively higher contributions of larger errors. The difference between the RMSE and MAE increases with the variance of the errors. Herein the difference between the two metrics is small (Table 5.2). The slightly better RMSE obtained with MBC-OK is due to the more accurate estimation of the extreme low value by this method. The bias is the difference between the estimated and the true values at a single point; hence, it can be positive or negative or zero. Unbiased estimation corresponds to zero bias. The sign of the average bias shows if the specific method underestimates (negative bias) or overestimates (positive bias) the data. Low bias errors mean more accurate estimations.

The MARE measures the accuracy of the estimates relatively to the respective measured values; MARE is independent of the units of measurement. The magnitude of MARE for all the normalization methods presented herein is similar and approximately equal to 14%. This estimate is partly due to the overestimation of the minimum value (9.4 masl) by all the methods (see above). The linear correlation coefficient (R) measures the strength of the association between the estimates and the measured values. Values of the coefficient close to 1 imply high association strength between the estimates and the measurements. In the present study, for all the methods the correlation coefficient has high values (ranging from 0.89 to 0.91).

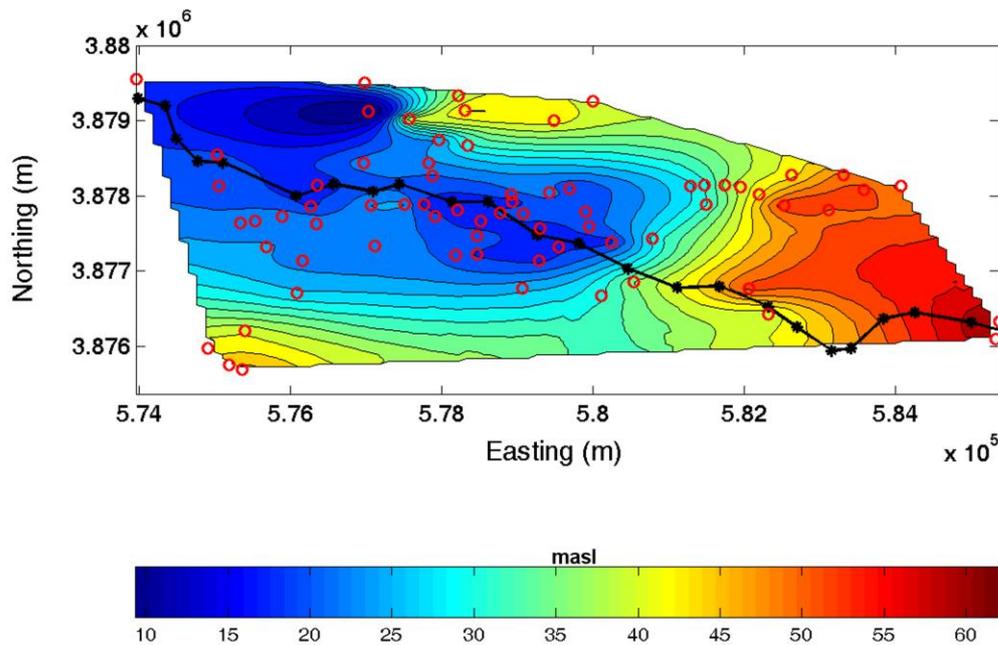
All the error metrics are significant for evaluating the methods' performance, and they are typically used in geostatistical studies (Goovaerts 1997). We give priority to MAE, bias and RMSE which are able to differentiate between the methodology and the semivariogram model used, in contrast with MARE and R that are similar for the majority of methodologies. Since the bias is the most sensitive evaluator of method performance, we use it to select the optimal transformation method (MBC-OK).

The validation metrics presented in Table 5.2 show that the studied methods are practically insensitive to the methodology used to normalize the data. This occurs because the transformation methods used lead to similar values for the skewness and kurtosis of the transformed data (Table 5.1). In addition, the differences between the transformation methods and classical OK are not significant. This is due to the fact that the original dataset has a mild deviation from the normal distribution, and the OK estimator is used for all cases tested following the normalizing transformation. For all the methods tested, however, there are differences in the validation metrics between different semivariogram models. Nevertheless, the validation metrics obtained with different normalizing methods but with the same semivariogram model are similar. This chapter shows that non-linear normalization methods help to improve the performance of kriging estimations (Table 5.2), even for datasets that has mild deviations from the normal distribution.

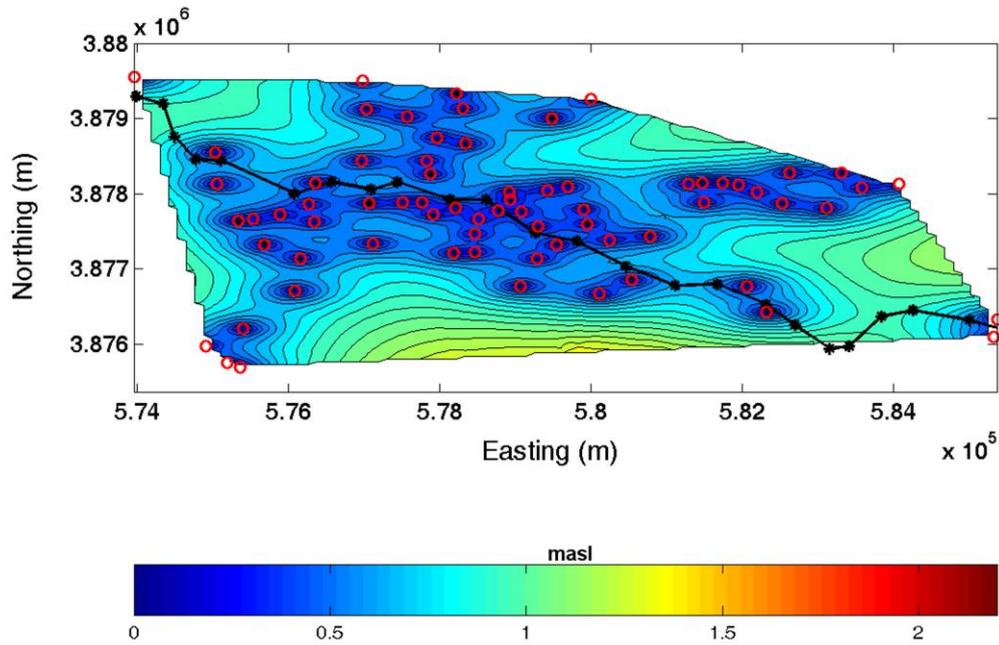
The optimal interpolation map is derived using the MBC-OK with the Spartan model approach on a  $100 \times 100$  grid defined in normalized coordinate space (actual cell size:  $114\text{m} \times 47\text{m}$ ). Estimates are obtained only at points inside the convex hull of the measurement locations (7317 grid points). The contour map in physical space is shown in Figure 5.4. The kriging standard deviation, which represents the uncertainty of the estimates, is shown in Figure 5.5. The optimum search radius used with the Spartan model (determined from leave-one-out cross-validation) is equal to 0.39 (normalized units), which is a little shorter than the determined characteristic length.

Interpolation and error maps for all the normalization methods investigated and for classical OK are constructed with the optimal semivariogram (Spartan); the maps are compared in Figure 5.6 to Figure 5.13. For the specific dataset there are no significant differences between the generated groundwater level maps. This is expected since the estimation measures are similar for all the methods, and especially for those that use a normalization process (Table 5.2). According to the interpolation

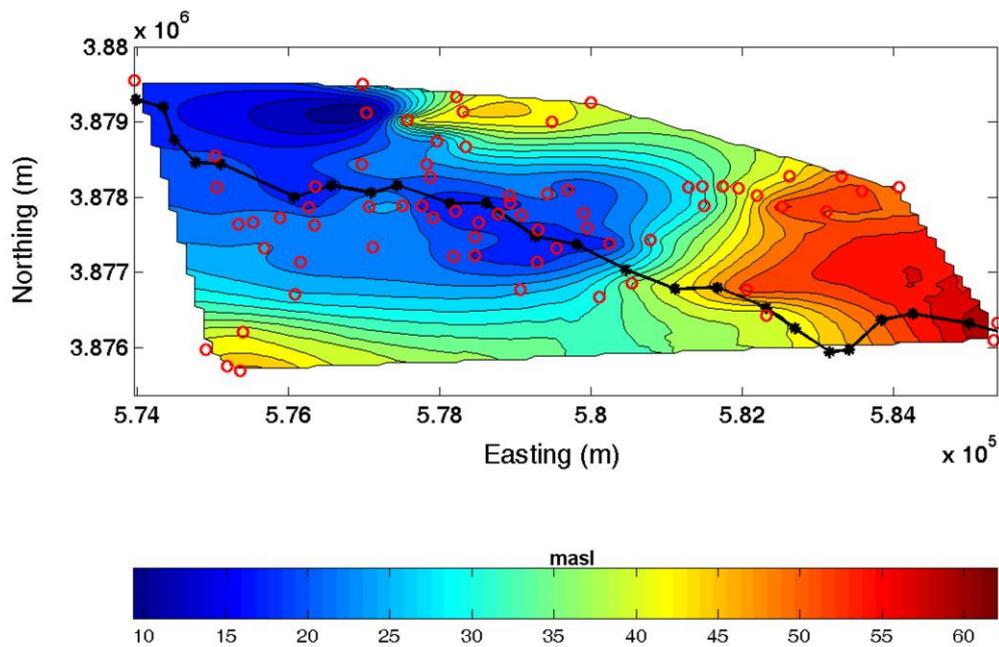
results, MBC-OK delivers a slightly more accurate groundwater level interpolation map compared to the other methods. Specifically, on the grid nodes close to the minimum value, MBC-OK method provides estimates closer to 9.4 masl (9.44 masl) than the other methods (which give estimates larger than 9.6 masl). Another feature observed in the maps is the smoothness of TGK contours compared to the other normalization methods. In advance the standard deviation maps of the methodologies tested (MBC-OK, GA-OK, TGK, Box-Cox-OK) present similar results due to the normalization methods similar performance and of the more efficient semivariogram parameters calculation, but significantly lower of the OK method. MBC-OK method provides the estimates with the lowest standard deviation. This shows that non-linear normalization methods improve in addition to kriging estimations their standard deviation results.



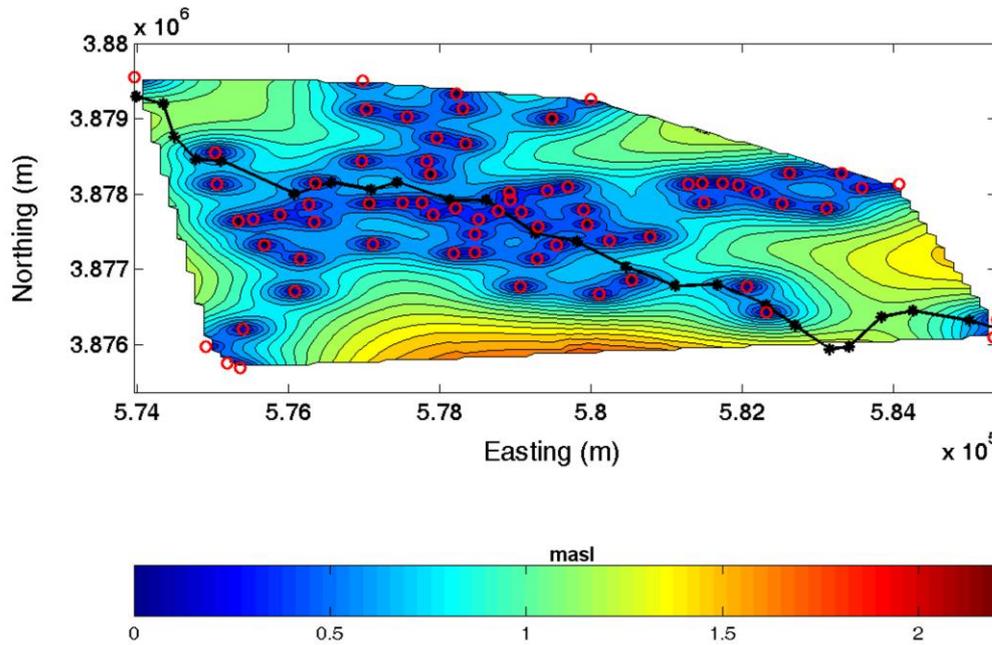
**Figure 5.4** Isolevel contour map of estimated groundwater level in the Mires basin using MBC-OK (red circles denote location of wells and solid black line the temporary river path).



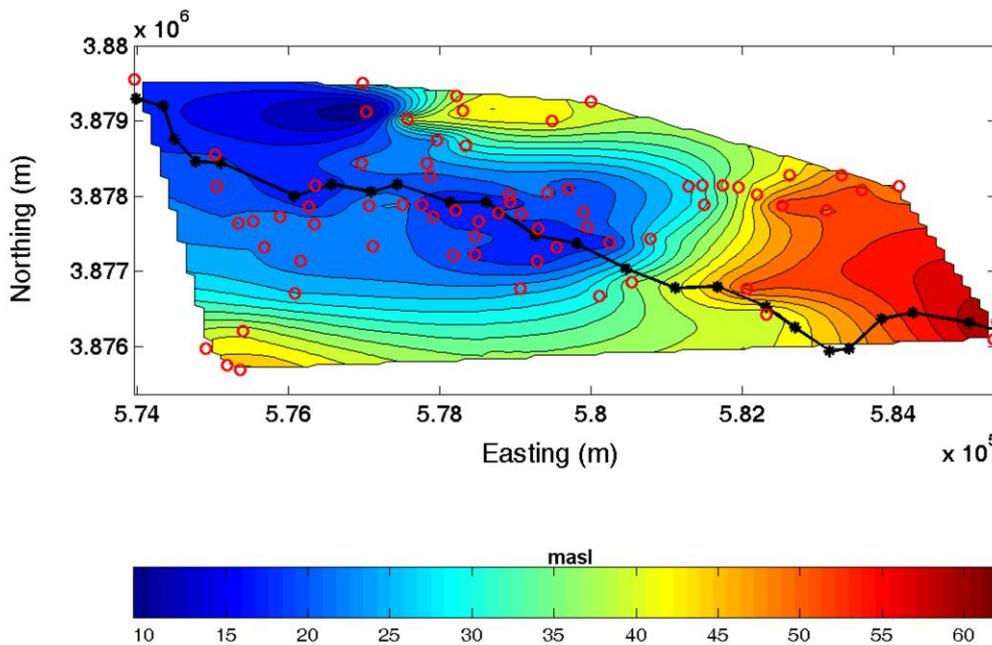
**Figure 5.5** Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using MBC-OK (red circles denote location of wells and solid black line the temporary river path).



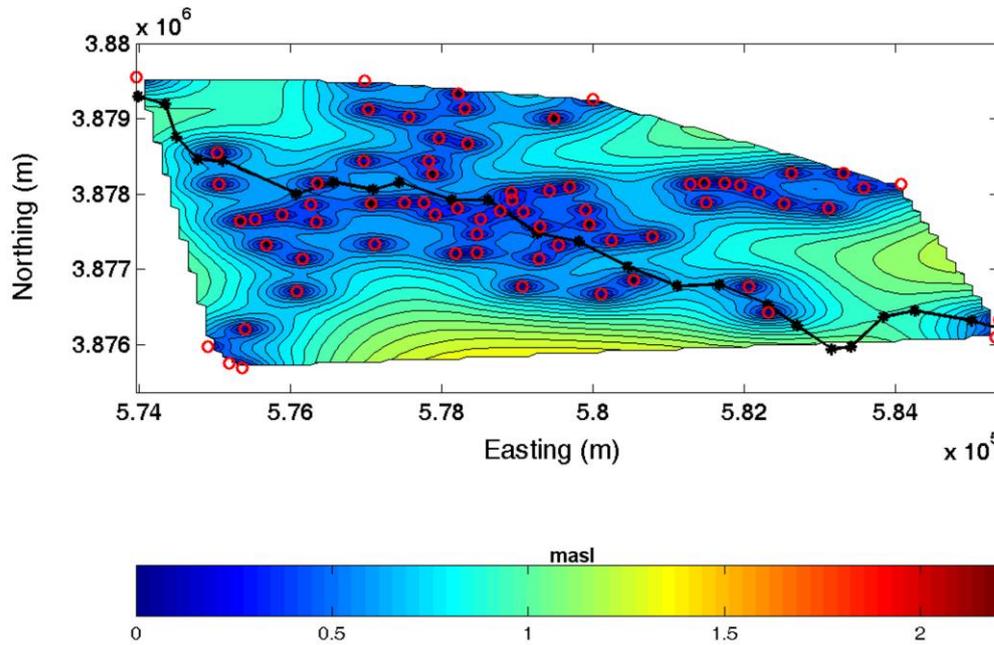
**Figure 5.6** Isolevel contour map of estimated groundwater level in the Mires basin using GA-OK (red circles denote location of wells and solid black line the temporary river path).



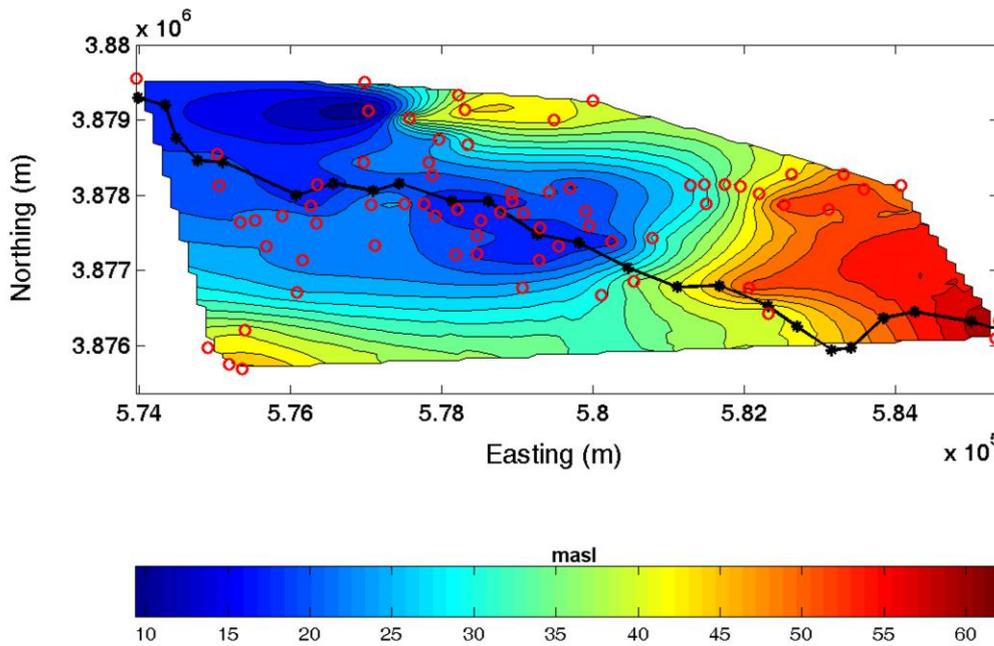
**Figure 5.7** Isopleth contour map of kriging standard deviation for groundwater level in the Mires basin using GA-OK (red circles denote location of wells and solid black line the temporary river path).



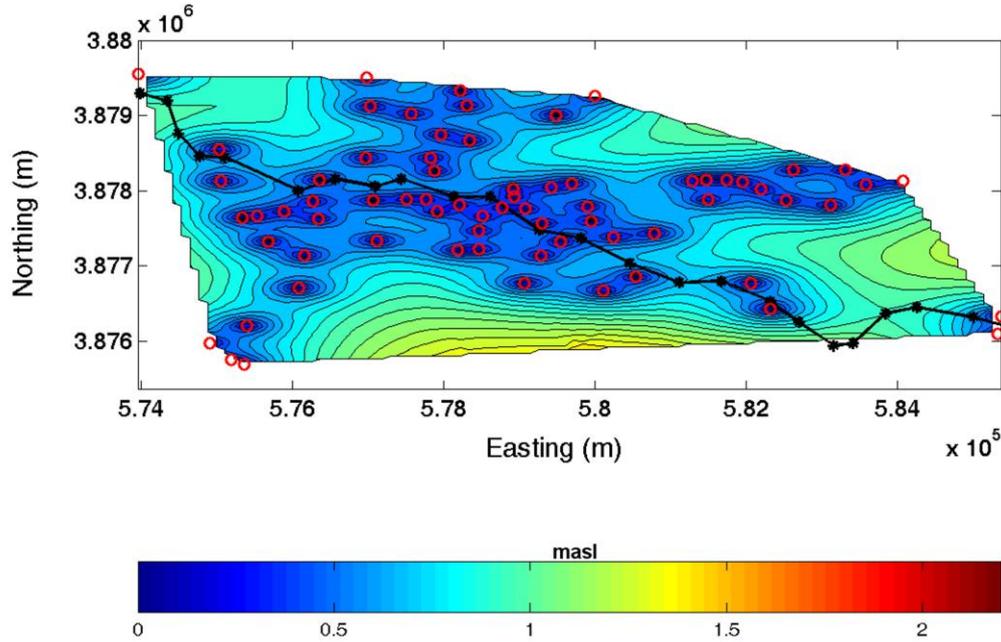
**Figure 5.8** Isopleth contour map of estimated groundwater level in the Mires basin using TGK (red circles denote location of wells and solid black line the temporary river path).



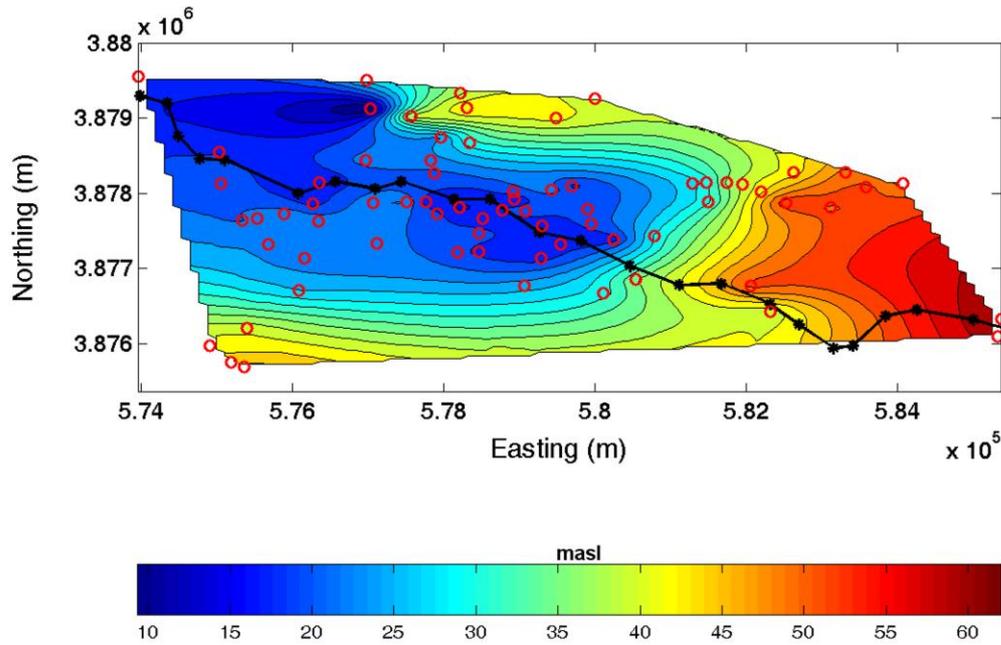
**Figure 5.9** Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using TGK (red circles denote location of wells and solid black line the temporary river path).



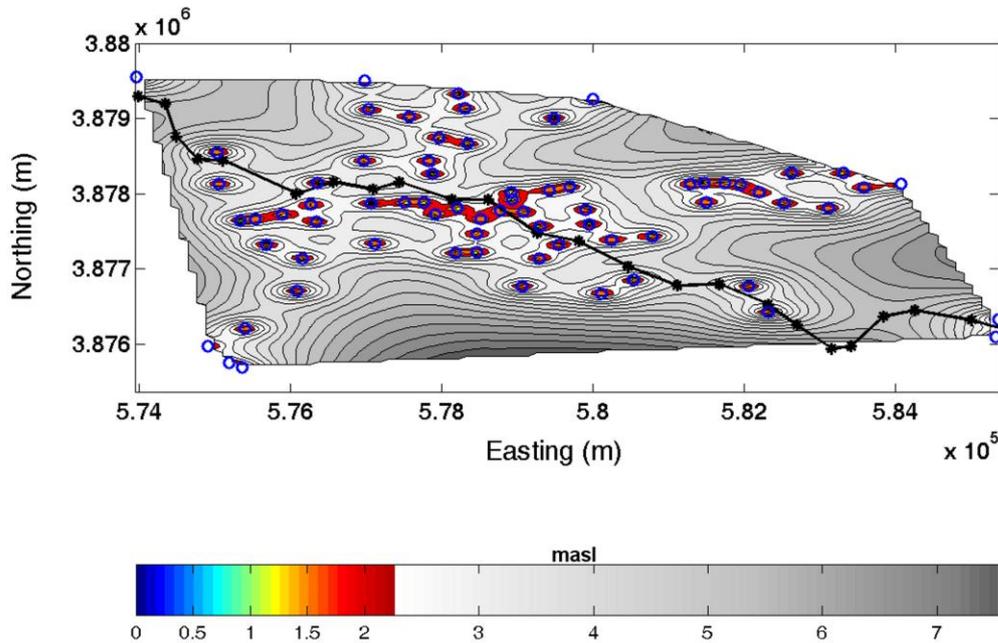
**Figure 5.10** Isolevel contour map of estimated groundwater level in the Mires basin using Box-Cox-OK (red circles denote location of wells and solid black line the temporary river path).



**Figure 5.11** Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using Box-Cox-OK (red circles denote location of wells and solid black line the temporary river path).

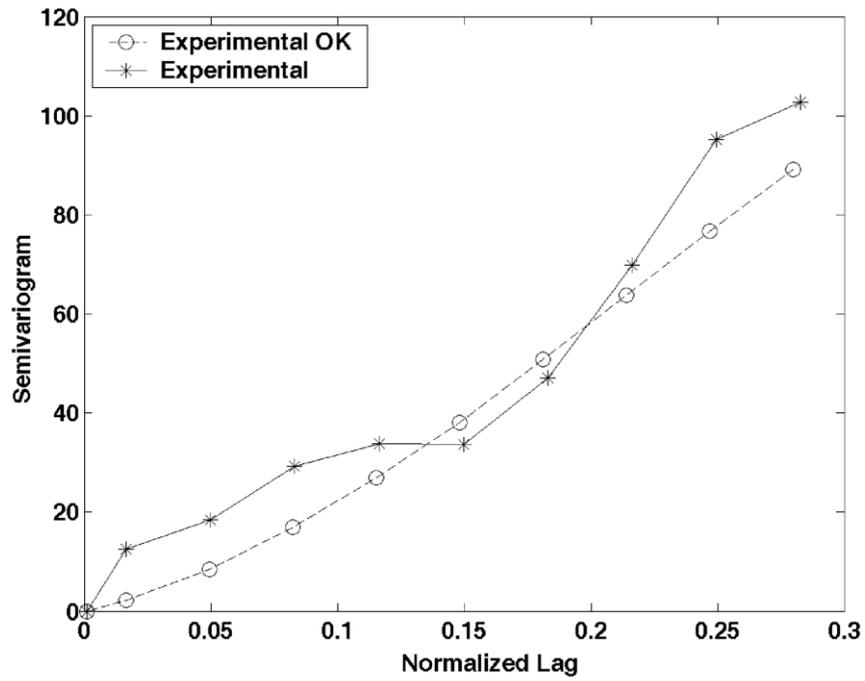


**Figure 5.12** Isolevel contour map of estimated groundwater level in the Mires basin using OK (red circles denote location of wells and solid black line the temporary river path).



**Figure 5.13** Isolevel contour map of kriging standard deviation for groundwater level in the Mires basin using Box-Cox-OK (blue circles denote location of wells and solid black line the temporary river path).

Stochastic interpolation methods such as kriging allow the comparison of the empirical semivariogram of the data with that obtained from the interpolation results, in order to test for the accurate reproduction of the spatial continuity. In Figure 5.14 we compare the experimental semivariogram of the observations and the experimental semivariogram obtained from the MBC-OK interpolation estimates. The semivariogram of the estimates shows very similar structure to that of the data. However, the former exhibits overall lower variability that reflects the smoothing effect of interpolation. The MBC-OK method is optimal with respect to the statistical metrics investigated above, and thus we used it for the interpolation of groundwater levels in the basin. Experience from previous studies has shown that a generally optimal interpolation method does not exist. Therefore, MBC-OK is optimal with respect to the specific case study and no general conclusions for the method's efficiency can be drawn. Hence, the optimal method should be determined for each case individually.



**Figure 5.14** Comparison of groundwater level omnidirectional experimental semivariograms: data (stars), and MBC-OK estimates (circles).



## **6. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables<sup>4</sup>**

### **6.1 Introduction**

In Stochastic hydrology physical laws and analytical solutions are incorporated in stochastic processes in order to predict the value of some variable at non-observed times or at non-observed locations. A physical law is a generalization obtained from the scientific study of the facts of observations. Physical laws are expressed by mathematical equations and govern the evolution of natural processes in space/time. They provide an important source of general knowledge that can interpret the natural process derived from the available data. Any physically based model becomes a stochastic model once its inputs, parameters or outputs are treated as random (Christakos 2000). A series of characteristic examples in stochastic hydrology that incorporated physical laws are presented for completeness below.

A physical law used in hydrology is the Perturbation-based spectral theory, which presumes local statistical homogeneity, and provides generic theoretical results for the head variance, effective conductivity tensor, and macrodispersivity tensor in a field (Gelhar 1986). Another stochastic approach based on the physical law of hydrological balance of an aquifer achieves water table elevations estimation using a regionalised autoregressive exogenous variable (RARX) model with precipitation surplus as the input variable. Classical geostatistics were applied as regionalisation functions (Knotters and Bierkens 2001). In addition a stochastic rainfall-runoff model based on the mass balance of a watershed was constructed with the rainfall excess input to the model to be treated as a stochastic process with a unit-step function. A stochastic differential equation described the relationship between the mean rainfall excess and the mean direct runoff (Lee *et al.* 2001). Finally eco-hydrological processes in water-limited ecosystems are described by simplified, vertically averaged soil moisture models. The principal aim of these models is to understand how the

---

<sup>4</sup> This chapter is an adaptation of a paper accepted for publication in *Advances in Water Resources Journal* (Varouchakis and Hristopoulos, 2012), please see page 179 for details.

main physical processes linking soil, vegetation, and climate impact on the statistical properties of soil moisture. A key component of these models is the stochastic nature of daily rainfall, which is mathematically described as a compound Poisson process with daily rainfall amounts drawn from an exponential distribution (Verma *et al.* 2011).

Classical geostatistics (e.g., kriging estimators) were primarily designed to use hard data (i.e., sets of measurements). These methods according to Christakos (2000) lack to account for important sources of physical knowledge (including physical laws, empirical models, higher-order space/time moments, and uncertain information). A framework that is incorporating various physical knowledge bases into spatiotemporal analysis and mapping is BME introduced by Christakos (1990, 1991b, 2000). The spatiotemporal distribution of most natural variables can be expressed through physical laws which, thus, constitute important aspects of the mapping process. Incorporation of general knowledge in the mapping process can lead to considerable gains in the estimation accuracy. For example an advection model provides valuable information for managing uncertainty in an air quality study. In the BME framework, physical laws are incorporated in the general knowledge by means of an appropriate set of moment equations (Christakos 2000). Another example is the incorporation of the Darcy law in the mapping of water table elevations that leads to predictions that are physically meaningful and more accurate than those provided by classical data analysis. (Serre and Christakos 1999).

Physical laws however have found applications in classical geostatistical modeling, e.g. for the definition of space only or spatiotemporal distance metrics, for the development of covariance functions in order to model the spatiotemporal data dependence and in the approximation of the trend term for spatial geostatistical models.

Metrics are mathematical expressions that define the concept of distance in the space-time continuum. Distance cannot always be defined explicitly in space-time. It is possible to decide using additional information, about the distance between two pairs of points by considering the outcome of a natural process. For example, the distance between two pairs of points in an aquifer regarding a pollutant transport is defined not purely form a geometric property of space-time, but it also depends on the medium's properties. Measuring distance by means of fluid tracer dispersion (dynamics of transport) can lead to very different results than measuring distance by

means of Euclidean distance. If a physical law is known for a natural (e.g. hydrologic) process it can also play a fundamental role in determining the appropriate space/time metric (Christakos and Hristopulos 1998, Christakos *et al.* 2000).

Covariance models are inherently connected to the physical laws governing a process. Natural processes are involved in the transport of water and pollutants from a source through different media. The evolution of natural processes is governed by differential equations that describe the motion of concentrations and fluxes in space and time. The coefficients of these equations represent properties of the media within which the processes take place. Usually, a complete characterization of the coefficients variability is impossible due to measurement errors that lead to uncertainty and because of limited sampling points in space-time. Such processes are represented by means of stochastic partial differential equations (SPDE). Exact solutions of SPDE are not in general available in explicit form. Two approaches are commonly used for the solution of SPDE: The first focuses on obtaining solutions that are valid for specific realizations of the coefficients of the spatiotemporal random field (S/TRF), the second approach focuses on the estimation of stochastic moments. The latter is of interest in this thesis as the integration of physical processes in the covariance function is researched. Moment-based approaches (obtain the mean and the covariance from monitoring data and physical modeling) focus on solving the deterministic equations that govern the stochastic moments of the natural processes represented by the SPDEs. The moment equations are solved explicitly only if the correlation functions of the coefficients satisfy certain symmetry requirements such as homogeneity and isotropy. As moment equations may suffer from the well-known closure problem; it is achieved using truncated perturbation series or non-perturbative approximations (Christakos and Hristopulos 1998).

Covariance models except of describing how the correlations behave in space and time, they can be intrinsically connected to the physical laws governing the process. For example power law correlation indicates the existence of scaling in the system i.e. due to different physical causes which denotes that the system may exhibit critical behavior. Percolation type models of flow and transport in porous media is an example of systems that exhibit critical behavior. Certain processes also, such as fractional Brownian motions (Mandelbrot and Van Ness 1968) are characterized by power law correlations. The apparent irregularity of such processes was shown to derive from the long range nature of the power law correlations among individual

events (Christakos and Hristopulos 1998). Generally power law correlations have been observed in environmental processes (Webster and Oliver 2001).

In addition physically motivated space-time random fields covariance models can be developed inspired by natural models such as the diffusion equation and the invasion percolation model. Invasion percolation describes the displacement of a defending viscous fluid in a porous medium (e.g., oil) by an invading fluid of lower viscosity (e.g., water). Spatiotemporal covariance models are derived so from a physical differential equation such as the diffusion equation as from the invasion percolation model in order to be applied in geostatistical modeling (Christakos and Hristopulos 1998, Kolovos *et al.* 2004).

Furthermore Heuvelink *et al.* (1996) suggested the inclusion of more process-oriented information into the interpolation, thus to incorporate physical laws about the transport of water in the soil, such as the law of continuity and Darcy's law. Therefore (Tonkin and Larson 2002) suggested the following approach. They use a linear drift term to approximate the hydraulic head field trend. However near extracting or injecting wells a point logarithmic component is added to the drift to account for the drawdown caused by the pumping well. This approach is applicable if analytical solutions for the aquifer response are available such as application of Theis method which is based on pumping tests. The above propositions gave the idea of using a physical law that could describe the aquifer behavior and used as the trend term in a spatial trend model for the groundwater level of the Mires basin.

Auxiliary information is often included as a drift term in spatial models in order to improve the accuracy of the estimations by capturing local properties. Usually polynomial functions of space coordinates, rainfall, or surface elevation from a Digital Elevation Map (DEM) are used as secondary information.

We propose therefore that the prediction of the hydraulic head spatial variability in Mires basin can be improved by incorporating in the trend a) the distance from a temporary river crossing the basin and b) a component based on the generalized *Thiem's* equation for multiple wells. In addition we use the flexible Spartan semivariogram family to perform Residual Kriging. Our proposal is supported by the results of cross validation analysis. Our results can be generalized to other unconfined aquifers.

## 6.2 Normalizing transformations

A nonlinear transformation  $g(\cdot)$  is applied to the data if there is deviation from the normal distribution. Normality is desirable for the application of OK (Clark and Harper 2000). OK is the optimal estimator if the data follow a multivariate normal distribution and the true semivariogram is known. Significant deviations from normality, e.g., excessive skewness or presence of outliers, impair the estimation of the semivariogram structure and OK performance (Gringarten and Deutsch 2001, Ouyang *et al.* 2006). Application of a normalizing transformation on the univariate distribution of the data can suppress outliers, improve stationarity and semivariogram stability and reduce the impact of extremes (Deutsch and Journel 1992, Armstrong 1998). In practice, multi-normality is invoked as a working hypothesis.

On the non-detrended head data we applied the Box-Cox transformation (section 5.2) and the new normalization method presented in this thesis; modified Box-Cox transform (section 5.5). On the de-trended head data we applied the modified Box-Cox transform, which can handle negative values.

## 6.3 Trend Modeling of Hydraulic Head in Mires Basin

Below we present the trend models for Mires basin. Following other studies, we first include secondary information in the trend from a Digital Elevation Model (DEM) of the area (Hoeksema *et al.* 1989, Deutsch and Journel 1992, Goovaerts 1997, Desbarats *et al.* 2002, Rivest *et al.* 2008, Nikroo *et al.* 2009). The correlation coefficient of the groundwater level and the ground surface elevation in Mires basin is calculated at 0.65, a value that is characterized as important (Tichy 1993). We also include information about the distance of the estimation point from the temporary river crossing the basin. This is an important auxiliary variable as it is observed that the groundwater level at the measured locations have an increasing trend moving away from the river bed. Finally, the hydraulic head trend is approximated by a term that is based on the multiple well extension of *Thiem's* equation. The motivation of using such a term corresponding to a physical law came from the idea to use an analytical solution that could describe the water table level under pumping conditions. A data set with information about the basin's hydraulic conductivity, the pumping

activity, the aquifer's saturation thickness and the wells radius of influence were used in order to solve locally the proposed analytical equation.

### 6.3.1 Topographic variables component

We introduce in the trend an external variable that represents the closest distance between the sampling locations and the temporary river that crosses Mires basin. The correlation coefficient between the groundwater level and the closest distance from the river is equal to 0.68. This means that the groundwater level is higher away from the river bed than closer to it. The dependence is reasonable considering that the agricultural activity in the area is concentrated along the temporary river. The following expression for the trend of the hydraulic head (in masl) is proposed:

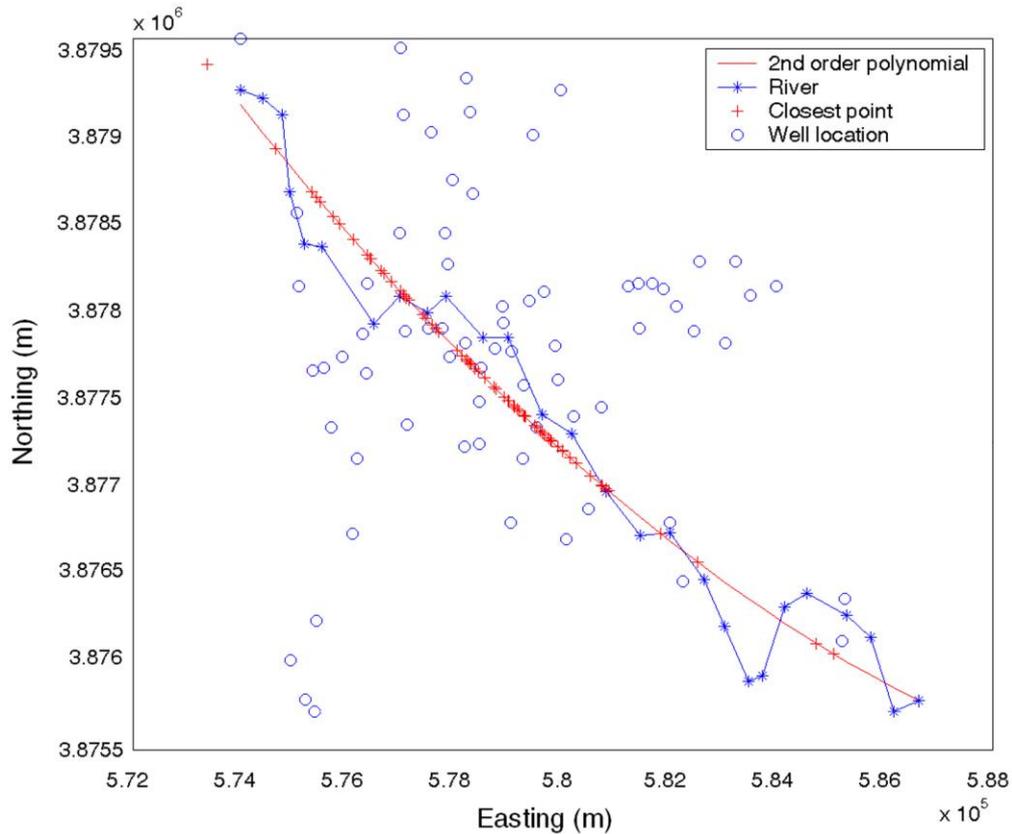
$$m_z(\mathbf{s}) = a d(\mathbf{s}) + f DEM(\mathbf{s}) + c, \quad (6.1)$$

where  $a, f, c$  are linear coefficients,  $d(\mathbf{s})$  is the minimum distance of point  $\mathbf{s}$  from the curve that follows the river bed, and  $DEM(\mathbf{s})$  is the local DEM value. We also use the linear approximation  $m_{DEM}(\mathbf{s}) = \mathbf{g} \cdot \mathbf{s} + c_0$ , where  $m_{DEM}(\mathbf{s})$  is the smoothed topographic elevation,  $\mathbf{g}$  is the uniform gradient, and  $c_0$  the reference elevation at the origin of the coordinate system.

The river bed can be modeled in two dimensions as a curve. Herein we represent the curve by a second-order polynomial,  $y(x) = w_0 + w_1 x + w_2 x^2$ . The coefficients  $w_0, w_1, w_2$  are determined by a least-squares fit of  $y(x)$  to "anchor points" along the river bed (see Figure 6.1).

As it is well-known, the closest distance of any given point from a curve is the perpendicular distance between the point and the tangent at a unique point of the curve. The slope of the straight line with the closest distance from a curve, which is perpendicular to the tangent, is given by  $\alpha = -\frac{1}{\ell(\mathbf{s}_{\min,0})}$ , where  $\ell(\mathbf{s}_{\min,0})$  denotes the tangent's slope. The perpendicular line has a form of  $y = \alpha \cdot x + b$ . The initial point coordinates  $x_0, y_0$  belong to that line as well as the closest to it point of the curve

$x_{\min,0}, y_{\min,0}$ . Substituting these two pairs of coordinates and the  $\alpha$  factor in the perpendicular line equation after elimination of the factor  $b$  an equation of the perpendicular line that contains the two desired pairs of coordinates is produced,  $\ell(\mathbf{s}_{\min,0}) \cdot y_{\min,0} - \ell(\mathbf{s}_{\min,0}) \cdot y_0 + x_{\min,0} - x_0 = 0$ . The tangent's slope is equal to the first derivative of the curve.



**Figure 6.1** Locations of “anchor points” along the river bed (stars), second-order polynomial model of the river (continuous line), well locations (circles) and their projections (crosses) on the model curve of the temporary river.

In general, the distance of a point  $\mathbf{s}_0 = (x_0, y_0)$  from the river curve is given by,

$$d^2(\mathbf{s}_0) = (x_{\min,0} - x_0)^2 + (y_{\min,0} - y_0)^2, \quad (6.2)$$

where  $\mathbf{s}_{\min,0} = (x_{\min,0}, y_{\min,0})$  is the closest point to  $\mathbf{s}_0$  on the river curve. This is determined by solving the following system of equations:

$$y_{\min,0} = w_0 + w_1 x_{\min,0} + w_2 x_{\min,0}^2, \quad (6.3)$$

$$\ell(\mathbf{s}_{\min,0})(y_{\min,0} - y_0) + x_{\min,0} - x_0 = 0, \quad (6.4)$$

where  $\ell(\mathbf{s}_{\min,0}) = w_1 + 2w_2 x_{\min,0}$  is the slope of the line tangent to the river curve at  $\mathbf{s}_0$ . The solution of the nonlinear system (6.3)-(6.4) with respect to  $(x_{\min,0}, y_{\min,0})$ , for every  $x_0, y_0$  depends on the roots of a third order (cubic) function after the substitution of (6.3) into (6.4):

$$\underbrace{2w_2^2}_{a} x_{\min,0}^3 + \underbrace{3w_1w_2}_{b} x_{\min,0}^2 + \underbrace{(2w_2w_0 + w_1^2 - 2w_2y_0 + 1)}_c x_{\min,0} + \underbrace{w_1w_0 - w_1y_0 - x_0}_d = 0. \quad (6.5)$$

Every cubic equation with real coefficients has at least one solution among the real numbers however several possible cases can be distinguished using the discriminant,

$$\Delta = 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2. \quad (6.6)$$

Hence if  $\Delta > 0$ , the equation has three distinct real roots, if  $\Delta = 0$ , the equation has a multiple root and all its roots are real and if  $\Delta < 0$ , the equation has one real root and two non-real complex conjugate roots. The real roots of equation (6.5) with respect to  $(x_{\min,0})$  are then substituted in equation (6.3) to calculate the corresponded  $(y_{\min,0})$ . The point  $\mathbf{s}_{\min,0}$  is then determined by the root, with respect to  $(x_{\min,0}, y_{\min,0})$ , of the nonlinear system (6.3)-(6.4) that minimizes  $d(\mathbf{s}_0)$  given by (6.2).

### 6.3.2 Multiple-well hydraulic head component

Physical laws incorporation in stochastic hydrology aspects and in classical geostatistical modelling reviewed previously gave the idea to introduce a spatial trend model that could incorporate in the trend a physical law that describes the Mires basin aquifer behavior with respect to groundwater level and pumping activity.

Therefore we include in the trend modeling the analytical solution for a system of multiple wells in an unconfined aquifer. This component of the trend is based on *Thiemø*s equation for an unconfined aquifer. The equation describes the relationship

between the steady-state radial inflow into a pumping well and the drawdown (Todd 1959, Charbeneau 2000). It is commonly used to estimate aquifer properties, i.e., hydraulic conductivity or transmissivity (Bear 1979, Butler Jr. 1988, Silliman and Caswell 1998, Balkhair 2006, Altunkaynak 2007) but also for calculating the hydraulic head (Yeo and Lee 2003, Pinder and Celia 2006, Steward and Jin 2006). *Thiem's* equation assumes that the aquifer is homogeneous (Thiem 1906) and in steady-state (Todd 1959, Bear 1979). As it was previously mentioned (section 3.2) to a large extent Mires basin is considered as a hydro-geologically uniform study area, therefore the basin is assumed geologically homogeneous. Steady-state conditions are also assumed as according to observations there is not temporal short term variation of hydraulic head.

*Thiem's* equation for a single pumping well in an unconfined aquifer is provided by the following function:

$$H^2(\mathbf{s}) = H_0^2(\mathbf{s}) + \frac{1}{\pi K} Q \ln\left(\frac{r}{R}\right), \quad r < R. \quad (6.7)$$

However it can be generalized to include the influence of a number of pumping wells as follows (Todd 1959, Bear 1979):

$$H^2(\mathbf{s}) = H_0^2(\mathbf{s}) + \frac{1}{\pi K} \sum_{i=1}^n Q_i \ln\left(\frac{r_i}{R_i}\right), \quad r_i < R_i, \quad i = 1, \dots, n. \quad (6.8)$$

In the above,  $H(\mathbf{s})$  is the estimated hydraulic head,  $H_0(\mathbf{s})$  the initial hydraulic head before abstraction,  $K$  is the hydraulic conductivity,  $n$  is the number of wells ( $i = 1, \dots, n$ ),  $Q$  is the pumping rate,  $r_i = \|\mathbf{s} - \mathbf{s}_i\|$  is the distance of the estimation point from the  $i$ th well, and  $R_i$  is the radius of influence of the  $i$ -th well. The pumping wells contributing in Eq. (6.8) are those whose distance from the estimation point does not exceed their radius of influence.

The average pumping rates ( $\text{m}^3/\text{h}$ ) at the 70 wells of the study are used in order to determine the variable  $Q_i$ . As initial values before abstraction we use the hydraulic head profile of the preceding hydrological year for the same period (April).

Readings from 10 piezometers in the basin are available for that month. The initial hydraulic head is assumed to follow the linear trend  $H_0(\mathbf{s}) = J_1x + J_2y + D$ , where  $J_1, J_2, D$  are linear coefficients obtained by a least squares fit to the 10 available measurements. The hydraulic conductivity in the basin varies from 0.0014 to 0.00014 m/sec with the value  $10^{-4}$  considered to be typical for the basin (Donta *et al.* 2006, Kritsotakis 2010).

Since pumping tests are not available, we determine the radius of influence using empirical equations (Bear 1979, Sen and Al-Somayien 1991, Prakash 2004), subject to available hydrogeological field data, i.e.,

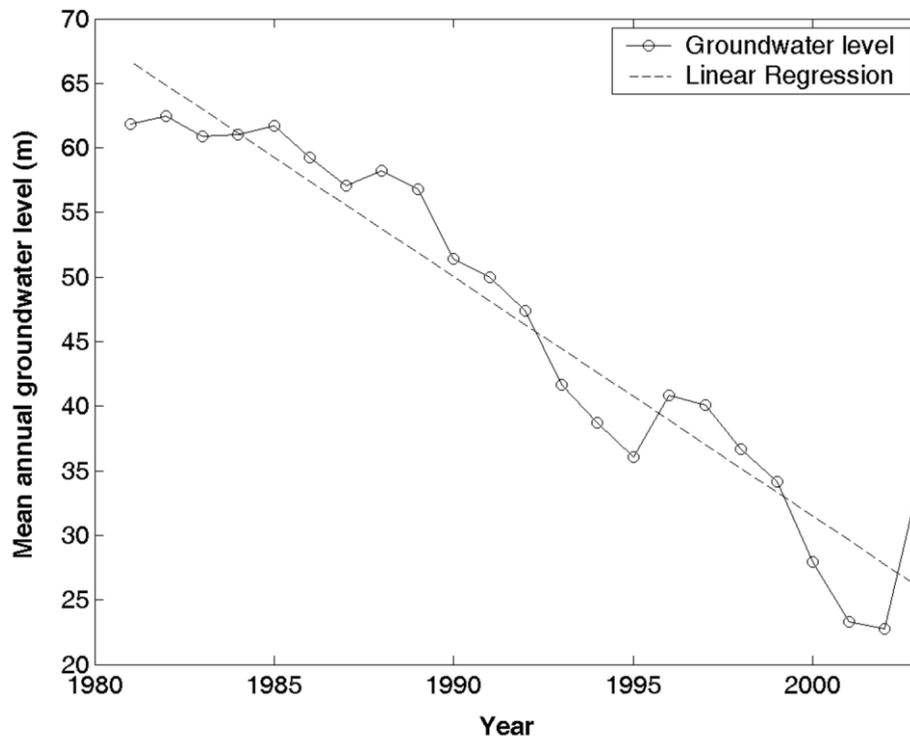
$$R_i = 3000 s_{w,i} \sqrt{K_i}, \quad (6.9)$$

$$R_i = 575 s_{w,i} \sqrt{H_{0,i} K_i}, \quad (6.10)$$

where  $s_w$  is the drawdown at the well face (m),  $K_i$  is the hydraulic conductivity around the pumping well and  $H_{0,i}$  (m) is the initial saturated thickness, i.e., the initial hydraulic head before the abstraction. Since the drawdown and hydraulic conductivity are not known at every well, an effective uniform value is used based on the sensitivity analysis described below. Hence, Eq. (6.9) provides a common radius of influence for all the wells, while Eq. (6.10) leads to  $R_i$  values that depend on the initial saturated thickness of the aquifer.

Linear regression analysis of the mean annual groundwater levels (Figure 6.2) estimates the rate of mean annual level decrease at 1.85m/yr with the 95% confidence interval at [1.60 - 2.10]. The correlation coefficient of the fit to the data is  $R = -0.96$ , implying a strong negative correlation. Analysis based on a groundwater balance model leads to a drop similar with that predicted by the regression. The mean annual recharge in the basin is 14.2  $\text{Mm}^3/\text{yr}$ , as reported by the Department of Agriculture of the Regional Council of Crete, while the mean annual abstraction rate is 22  $\text{Mm}^3/\text{yr}$ . Therefore, the mean absolute volumetric consumption is calculated at 7.8  $\text{Mm}^3/\text{yr}$ . Dividing this value by the surface area of Mires basin (50.3  $\text{km}^2$ ), the loss of groundwater content is estimated at 0.156m/yr. Further, dividing this figure with the average porosity of the basin (8.5%), the annual decrease of groundwater level is

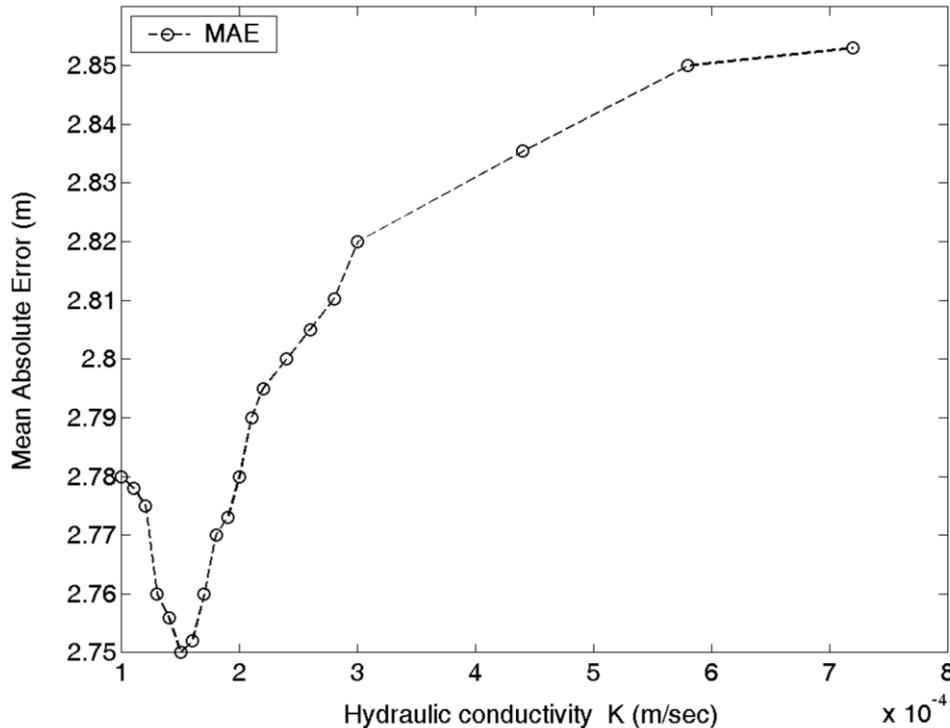
estimated at 1.83 m/yr, in close agreement with the estimate from linear regression (1.85 m/yr). A similar balance analysis (Croke *et al.* 2000) estimated a groundwater level drop of 1.5m/yr for the time period from 1985 until 1995.



**Figure 6.2** Annual average groundwater level variation in Mires basin for the period covering the years 1981-2003 (circles) and fitted linear trend (broken line).

We estimate an optimal “effective” hydraulic conductivity using sensitivity analysis that focuses on the reproduction of the measured head values by means of leave-one-out cross validation and RK. The mean absolute error (MAE) is used as the criterion of performance. In the analysis we use drawdown values in the 95% confidence interval [1.60 – 2.10]. Values of the hydraulic conductivity between the reported minimum and maximum are investigated to determine an effective  $K$  value for the basin. We found that the MAE is primarily sensitive to the hydraulic conductivity. Figure 6.3 shows the dependence of the MAE on the hydraulic conductivity; a clear minimum is obtained for  $K = 0.00015$  m/sec. This value minimizes the MAE for all values of  $s_w$  tested (e.g., Figure 6.3). The lowest MAE is obtained by using  $s_w$  equal to 1.85m. Based on the above estimates and  $H_0(\mathbf{s}) = J_1x + J_2y + D$ , the minimum radius of influence in Eq. (6.10) is

approximately 80 m and the maximum 150 m. Equation (6.9) leads to a uniform value of  $\approx 70$  m.



**Figure 6.3** Sensitivity analysis to determine the optimal hydraulic conductivity ( $K$ ) value used for the calculation of the radius of influence  $R$  in (6.9) or (6.10) and subsequently for the hydraulic head (trend), in (6.8). The cross validation measure MAE, is calculated based on RK methodology. Different values of  $K$  between the two extremes are investigated while  $s_w$  is set equal to 1.85m.

## 6.4 Interpolation of Hydraulic Head in Mires Basin: Models and Results

### 6.4.1 Exploratory statistics

The main statistics of the head data are shown in Table 3.1. The skewness and kurtosis coefficients are equal to  $\hat{s}_z = 0.81$  and  $\hat{k}_z = 2.58$  respectively, implying a mild deviation from Gaussian statistics ( $\hat{s}_z = 0$ ,  $\hat{k}_z = 3$  respectively). Data transformations are used to improve normality as shown in Table 6.1. The residuals of the trend models also display deviations from normality that are reduced by means of the MBC transformation (Table 6.1).

**Table 6.1** Skewness  $\hat{s}_z$  and kurtosis  $\hat{k}_z$  coefficients of head data and of trend models residuals following Box-Cox normalization (for non-negative values) and modified Box-Cox (MBC). T-DEM model uses the surface elevation as an external variable, the T-DEM-UGA uses the uniform-gradient approximation of the DEM, the T-RD uses the distance from the river curve, the T-RD-DEM-UGA uses a linear combination of the distance from the river curve and the uniform-gradient approximation of the surface elevation and T-MW approximates the trend using *Thiem's* equation, Eq. (6.8).

	Method	No Trend approximation	T-DEM	T-DEM-UGA	T-RD	T-DEM-UGA-RD	T-MW
<b>Head data</b>	$\hat{s}_z$	0.81	-	-	-	-	-
	$\hat{k}_z$	2.58	-	-	-	-	-
<b>Residuals</b>	$\hat{s}_z$	-	1.23	0.9	0.7	0.61	0.81
	$\hat{k}_z$	-	5.1	4.4	4.07	4.17	2.47
<b>Box-Cox</b>	$\hat{s}_z$	0.01	N/A	N/A	N/A	N/A	N/A
	$\hat{k}_z$	2.70	N/A	N/A	N/A	N/A	N/A
<b>MBC</b>	$\hat{s}_z$	0.13	0.6	0.54	0.21	0.19	0.20
	$\hat{k}_z$	2.99	3.54	3.47	3.23	3.19	2.91

#### 6.4.2 Geostatistical head models

As reported in Section 2.5, for interpolation we use both models with trend function, to which we refer as (T), and models without trend (NT). Normalizing transformations are used in both cases. For models with trend, the transforms act on the residuals, while for the non-trend models they act on the original data. In the T-case we investigate various trend options: the T-DEM model uses the surface elevation as an external variable, the T-DEM-UGA uses the uniform-gradient approximation of the DEM, the T-RD uses the distance from the river curve, the T-RD-DEM-UGA uses a linear combination of the distance from the river curve and the uniform-gradient approximation of the surface elevation and T-MW approximates the trend using Eq. (6.8).

#### 6.4.3 Head models without trend

Two general approaches are used for interpolation: the first one applies OK with the optimal semivariogram function to the original data. The second applies a normalizing transformation (Table 6.1) followed by OK on the transformed variable, and it finally back-transforms the predictions.

The parameters of the theoretical semivariogram models (Gaussian, Exponential, Linear, Spartan, Matérn, Spherical, and Power-law) obtained by least

squares fitting to the experimental omnidirectional semivariogram of the original hydraulic head data are shown in Table 6.2. The Spartan model gives the best fit in terms of cross validation results of all the NT models tested (Table 6.3). Normalization methods improve OK estimation measures, with the MBC method performing overall slightly better than Box-Cox.

**Table 6.2** Optimal estimates of semivariogram model parameters obtained by least squares fit to experimental semivariogram. Sill and characteristic length  $\xi$  are in normalized units.

Semivariogram model	sill	$\xi$	Other parameters
Matérn	440	0.94	$\nu = 0.92$
Exponential	133	0.30	NA
Spherical	150	0.63	NA
Spartan	184	0.46	$\eta_1 = 1.12$
Gaussian	160	0.28	NA
Power Law	538	NA	$2H = 1.31$
Linear	331	NA	NA

**Table 6.3** Cross validation results of spatial models with optimal semivariograms, based on measures listed in section 2.5. OK: Ordinary Kriging. Box-Cox & OK: Box-Cox transformation followed by OK and back-transformation. MBC & OK: Modified Box-Cox transformation followed by OK and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

NT Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
OK	SP	3.37	0.02	0.14	5.15	0.91
	P	3.58	0.07	0.15	5.46	0.9
	M	3.80	0.02	0.16	5.84	0.89
Box-Cox & OK	SP	3.30	0.10	0.14	5.14	0.91
	P	3.41	0.09	0.14	5.31	0.90
	M	3.60	-0.30	0.15	5.65	0.89
MBC & OK	SP	3.30	0.02	0.14	5.12	0.91
	P	3.39	0.05	0.14	5.17	0.90
	M	3.60	0.03	0.15	5.54	0.89

Note that the Spartan semivariogram is continuous but non-differentiable, implying that the water table level is a non-differentiable function. Similarly, a study focusing on an aquifer in Belgium (Fasbender *et al.* 2008), proposed the spherical semivariogram. The superior performance of non-differential semivariograms versus differentiable models is surprising at first, since the hydraulic head is assumed to be a differentiable function in the saturated zone. However, the water table level corresponds to the surface defining the upper boundary of the saturated zone. One can model the elevation as the result of a deposition-removal process that adds (e.g., due to precipitation) and removes (e.g., due to pumping and evapotranspiration) locally varying increments of water. Such an idealized process is expected to yield a fractional Brownian motion (fBm) (Mandelbrot and Van Ness 1968). In surface hydrology fBm processes have been used as models of reservoir water levels (Feder 1988). The fBm models have power-law semivariograms of the form  $\gamma_z(r) \propto r^{2H}$  with  $0 < H < 1$ . As shown in Table 6.3, the power-law semivariogram with  $H \approx 0.65$  performs closely to the Spartan model.  $H > 0.5$  implies persistent correlations of the water table level values in the basin.

#### 6.4.4 Head models with trend

In the case of spatial models with trend components RK is applied. RK combines a trend function with interpolation of the residuals. The omnidirectional experimental semivariogram is calculated by applying the method of moments (2.10) to the transformed residuals. The MBC transformation is used to normalize the residuals (Table 6.1). The Spartan semivariogram model gives overall the most accurate cross validation results in all cases, while the power-law and the Matérn semivariogram come close (Table 6.4 to Table 6.8).

**Table 6.4** Cross validation measures (cf. section 2.5) for spatial T-DEM model with optimal semivariograms: trend using DEM surface elevation with  $a = 0, f = 0.37, c = 3.75$  in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
MBC & RK	SP	3.32	0.07	0.15	5.20	0.90
	P	3.31	0.03	0.15	5.23	0.90
	M	3.65	0.04	0.16	5.70	0.88

**Table 6.5** Cross validation measures (cf. section 2.5) for the spatial T-DEM-UGA model with the optimal semivariograms: trend using the uniform gradient approximation of the surface elevation with  $a = 0, f = 0.28, c = 10.43$  in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of the residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
<b>MBC &amp; RK</b>	SP	3.21	0.03	0.14	5.08	0.90
	P	3.22	0.09	0.14	5.10	0.90
	M	3.51	0.04	0.15	5.40	0.88

**Table 6.6** Cross validation measures (cf. section 2.5) for spatial T-RD model with optimal semivariograms: trend using distance from the river curve with  $a = 52.90, f = 0, c = 20.90$  in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
<b>MBC &amp; RK</b>	SP	3.11	0.08	0.12	4.86	0.92
	P	3.11	0.11	0.12	4.88	0.91
	M	3.11	-0.08	0.12	4.90	0.92

**Table 6.7** Cross validation measures (cf. section 2.5) for T-DEM-UGA-RD model with optimal semivariograms: trend using gradient approximation to ground surface elevation and distance from river curve with  $a = 52.07, f = 0.27, c = 2.22$  in (6.1). MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
<b>MBC &amp; RK</b>	SP	3.02	0.07	0.12	4.79	0.92
	P	3.02	-0.09	0.12	4.81	0.92
	M	3.01	-0.13	0.12	4.82	0.92

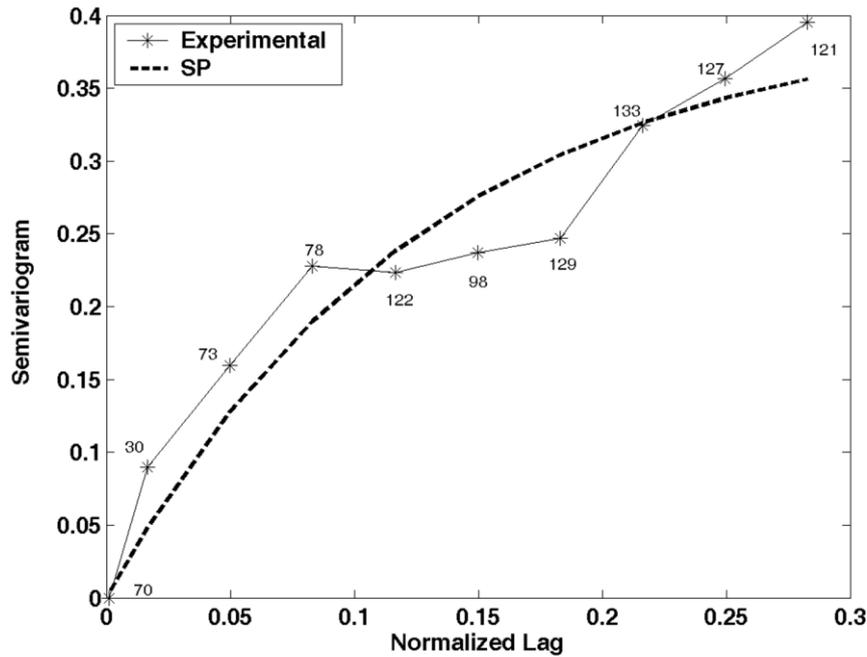
The cross validation results are compared in Table 6.4 to Table 6.8 with the respective ones for the NT models (Table 6.3). Similar performance measures are obtained with the T-DEM model that uses the surface elevation, and the results are improved by the T-DEM-UGA model using the uniform-gradient approximation. The validation measures overall improve by adding the distance from the river curve to the trend model (Table 6.6). Incorporation of the smoothed surface elevation (Table 6.7) can further improve certain validation measures, e.g., the MAE drops from 3.11 masl to 3.02 masl and the RMSE from 4.86 to 4.79 masl. Nevertheless, the bias is similar (0.08 and 0.07 respectively), while the MARE and R remains unchanged.

The T-MW model uses Eq. (6.8) for the trend. Equation (6.10) is used for the calculation of the wells’ radius of influence. We use the model coefficients determined in Section 6.4.2. RK combines the trend with the fluctuation estimate based on OK. Table 6.8 presents the cross validation results for different  $s_w$  values and the three “optimal” semivariogram functions for the transformed residuals. As shown in Table 6.8, the Spartan semivariogram gives overall the most accurate estimates followed closely by the power-law model. The validation measures are further improved, i.e., the MAE drops to 2.75 masl, the RMSE to 4.57 masl, the MARE to 0.11, and  $R$  increases to 0.93. Such a trend determination does not include variance in its estimations, as the coefficients of basis functions, (6.8),  $r$ ,  $R$  are known. Therefore the variance of estimations is only due to the interpolation of the residuals.

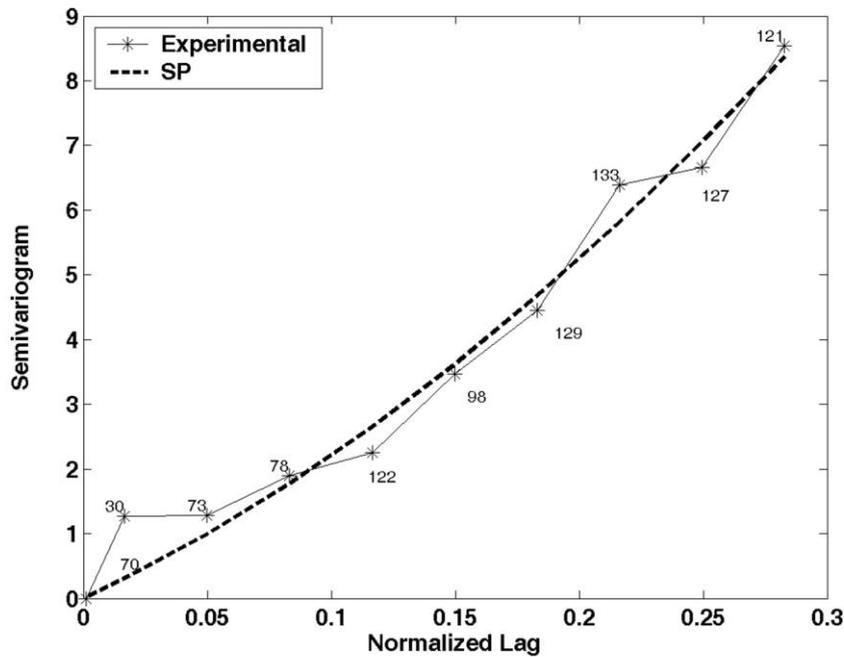
**Table 6.8** Cross validation measures (cf. section 2.5) for T-MW model with optimal semivariograms: trend using hydraulic head obtained from multiple wells system operation. MBC & RK: Residual Kriging with modified Box-Cox transformation of residuals and back-transformation.  $s_w$ : the drawdown at the well face. SP: Spartan semivariogram. P: Power-law semivariogram. M: Matérn semivariogram.

Method	Semi-variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
<b>MBC &amp; RK</b> $s_w = 1.85$ (masl)	SP	<b>2.75</b>	<b>0.07</b>	<b>0.11</b>	<b>4.57</b>	<b>0.93</b>
	P	2.75	0.11	0.11	4.63	0.90
	M	3.00	-0.14	0.12	4.76	0.89
<b>MBC &amp; RK</b> $s_w = 2.10$ (masl)	SP	2.77	0.13	0.11	4.67	0.91
	P	2.81	0.09	0.11	4.7	0.90
	M	3.04	-0.2	0.12	4.8	0.89
<b>MBC &amp; RK</b> $s_w = 1.60$ (masl)	SP	2.80	0.16	0.11	4.72	0.91
	P	2.80	0.12	0.11	4.71	0.90
	M	3.10	-0.13	0.12	4.84	0.89

Based on the analysis above and the cross-validation results, we rank the “optimal spatial models” (that perform similarly) as follows: (1) T-MW, (2) T-DEM-UGA-RD, and (3) T-RD. The semivariograms (experimental and modeled) of the transformed residuals for the T-DEM-UGA-RD and T-MW models are shown in Figure 6.4, Figure 6.5 and the model predictions are based on RK.

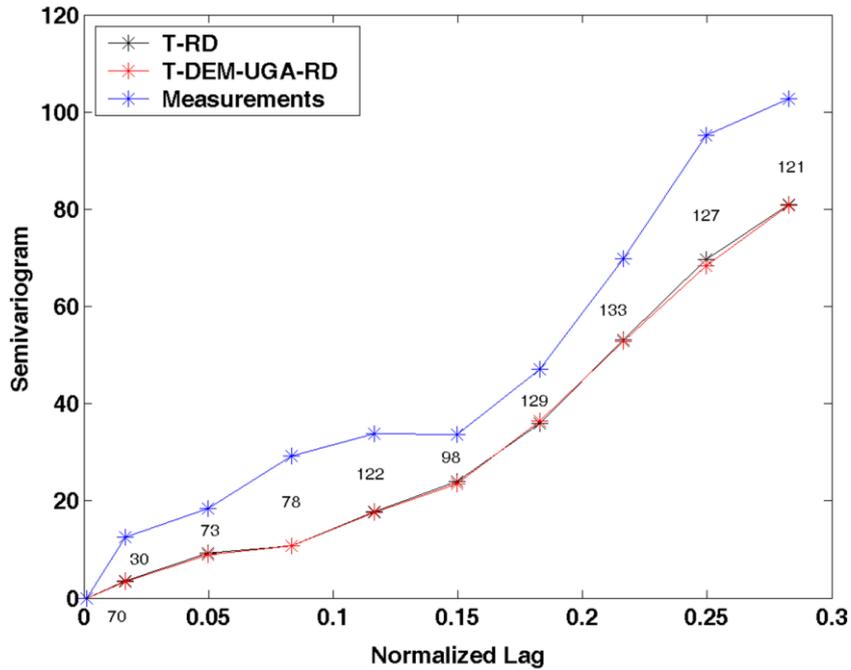


**Figure 6.4** Plots of the omnidirectional semivariogram of the residuals (after applying MBC normalization) and the best-fit Spartan semivariogram (SP) model fit ( $\sigma^2=0.62$ ,  $\xi=0.38$ ,  $\eta_1=1.51$ ). The residuals are derived in the framework of the T-DEM-UGA-RD model, i.e., by subtracting a trend that accounts for distance from the river curve and surface elevation (in the uniform-gradient approximation). The numbers of pairs used at each lag distance are also shown on the plot.

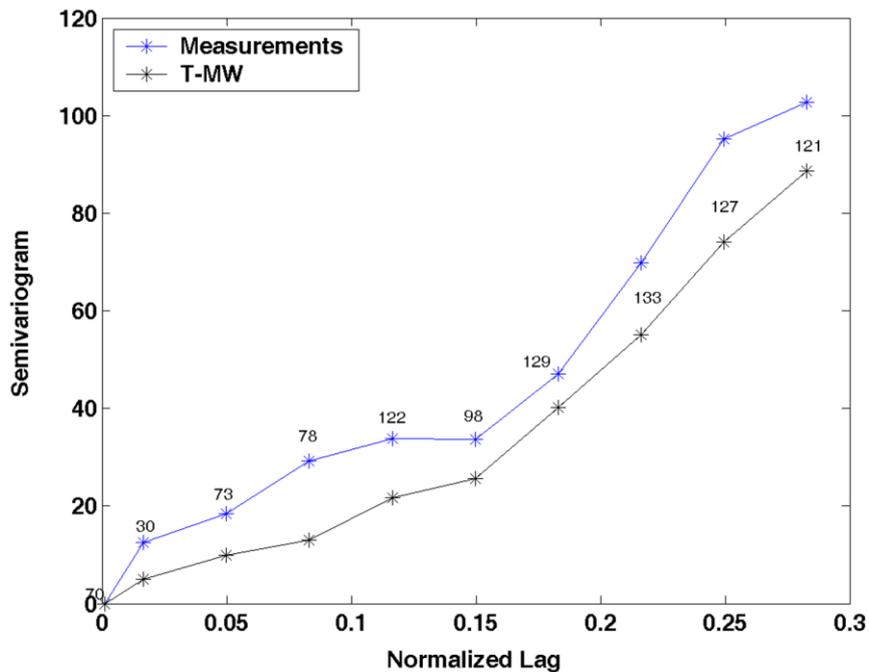


**Figure 6.5** Plots of the omnidirectional semivariogram of the residuals (after applying MBC normalization) and the best-fit Spartan semivariogram (SP) model fit ( $\sigma^2=9.9$ ,  $\xi=0.28$ ,  $\eta_1=0.52$ ). The residuals are derived in the framework of the T-MW model, i.e., by subtracting a trend that accounts for the groundwater level calculated from a multiple wells system operation analytical equation. The numbers of pairs used at each lag distance are also shown on the plot.

Figure 6.6 and Figure 6.7 show the experimental omnidirectional semivariogram of the measurements along with the semivariograms obtained from the leave-one-out estimates of the three top performance models. All four plots show very similar structure of spatial continuity. The semivariograms of the estimates exhibit lower variability reflecting the smoothing effect of the interpolation.



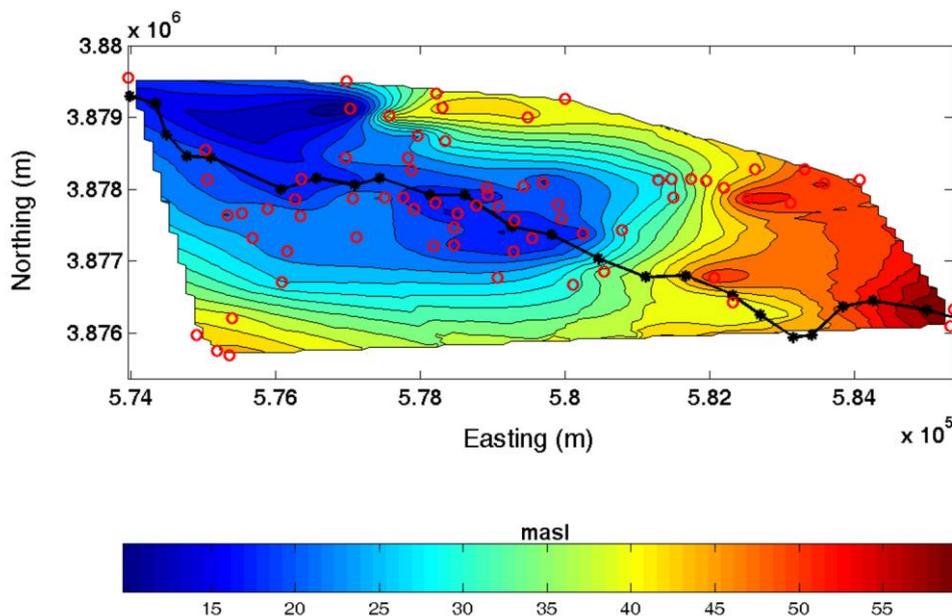
**Figure 6.6** Comparison of omnidirectional experimental semivariograms of the groundwater level measurements and of leave-one-out estimates for T-RD and T-RD-DEM-UGA spatial models.



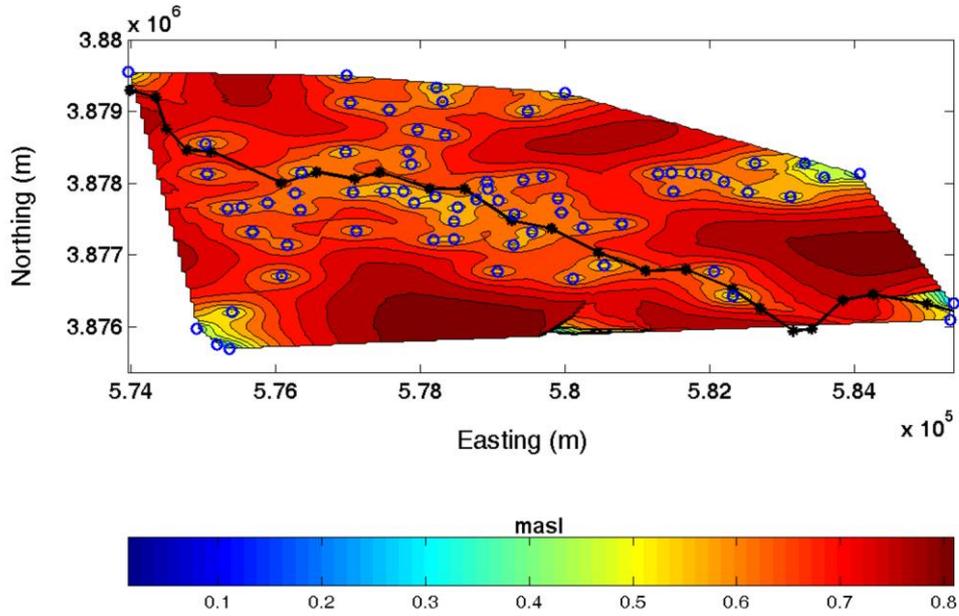
**Figure 6.7** Comparison of omnidirectional experimental semivariograms of the groundwater level measurements and of leave-one-out estimates for the T-MW spatial model.

#### 6.4.5 Mapping of Groundwater Level for the Optimal Model

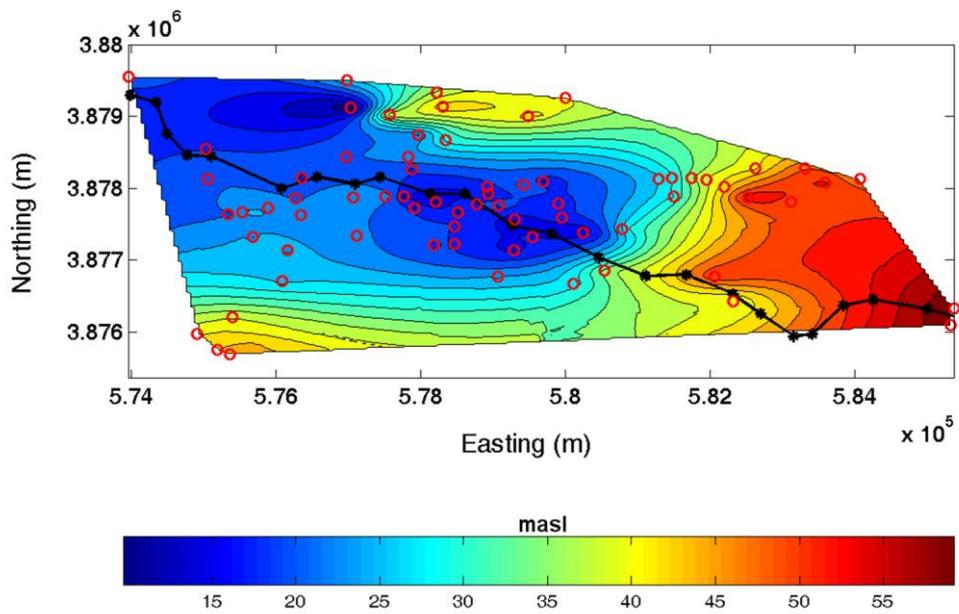
We use the T-DEM-UGA-RD and T-MW models with RK to estimate the groundwater level on a 100x100 grid defined in normalized coordinate space (actual cell size: 114x47 m). Estimates are obtained only at points that lie inside the convex hull of the measurement locations (7317 grid points). The contour maps in physical space are shown in Figure 6.8 and Figure 6.10. The residuals of the T-DEM-UGA-RD model are interpolated using the Spartan semivariogram model (see Figure 6.4) with the following optimal parameter values:  $\sigma^2=0.62$ ,  $\xi=0.38$ ,  $\eta_1=1.51$ . The residuals of the T-MW model are interpolated with the Spartan semivariogram model (see Figure 6.5) with  $\sigma^2=9.9$ ,  $\xi=0.28$ ,  $\eta_1=0.52$ . The optimum search radius used with the Spartan model (determined by the leave-one-out cross validation test) is equal to 0.38 (normalized units) for both models, identical to the estimated  $\xi$  by T-DEM-UGA-RD and slightly greater than the  $\xi$  of T-MW.



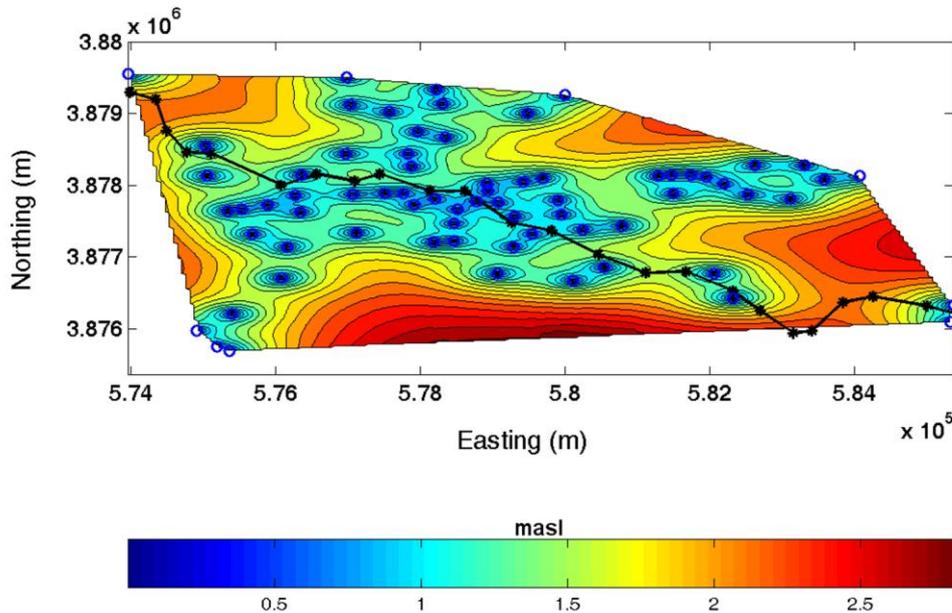
**Figure 6.8** Map of estimated groundwater level in the Mires basin using RK-T-RD-DEM-UGA spatial model, adapted on the real basin coordinates and location in the valley.



**Figure 6.9** Map of estimated groundwater level standard deviation in the Mires basin using RK-T-RD-DEM-UGA spatial model, adapted on the real basin coordinates and location in the valley.



**Figure 6.10** Map of estimated groundwater level in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley.



**Figure 6.11** Map of estimated groundwater level standard deviation in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley.

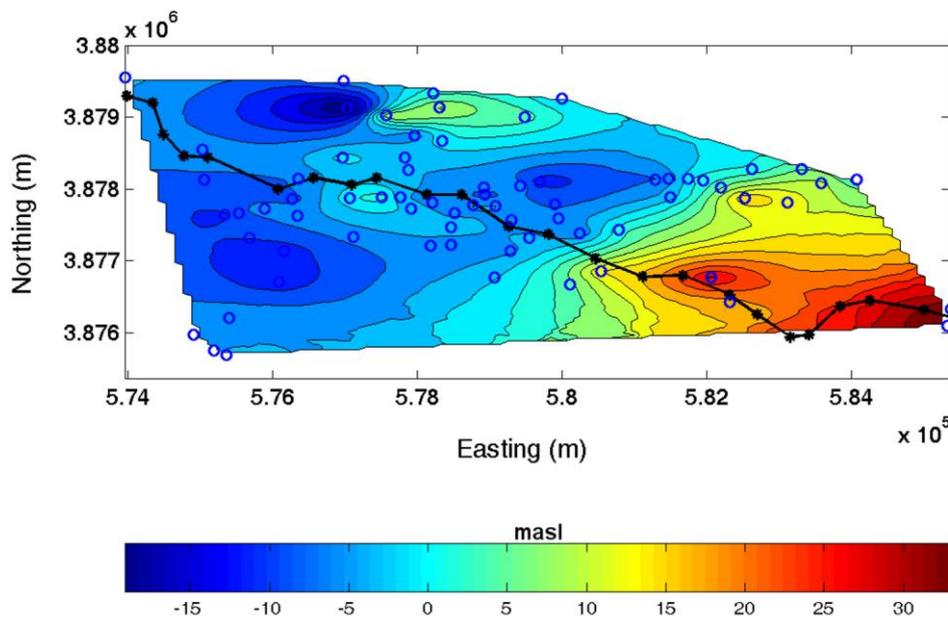
Low groundwater levels (20 masl or less) are observed over a significant fraction of the basin, with the lowest values observed near the northwest end. The kriging standard deviation, which represents the uncertainty of the estimates, is shown in Figure 6.9 and Figure 6.11. The highest values, around 0.8 masl for T-DEM-UGARD model and around 2.8 masl for T-MW model are obtained near the boundaries of the basin, and especially at distant points from the measurement stations.

## 6.5 Discussion

Changes in precipitation and temperature due to predicted climate changes in Crete will substantially affect the island's water resources due to declining recharge rates and increasing abstraction rates. Spatial models for the accurate representation of groundwater level variability in vulnerable areas with low groundwater resources, such as the Mires basin, will help the identification of susceptible locations and potential groundwater resources management plans. In addition, accurate on-grid representation of the basin's groundwater level can be used in numerical models to calibrate the estimated hydraulic head field by providing suitable initial conditions. The calibrated model combined with an optimization method can lead to improved

planning of the abstraction rates that will aim to maintain sustainable groundwater levels in the basin (Garg and Ali 2000, Karterakis *et al.* 2007, Ahlfeld and Hoque 2008).

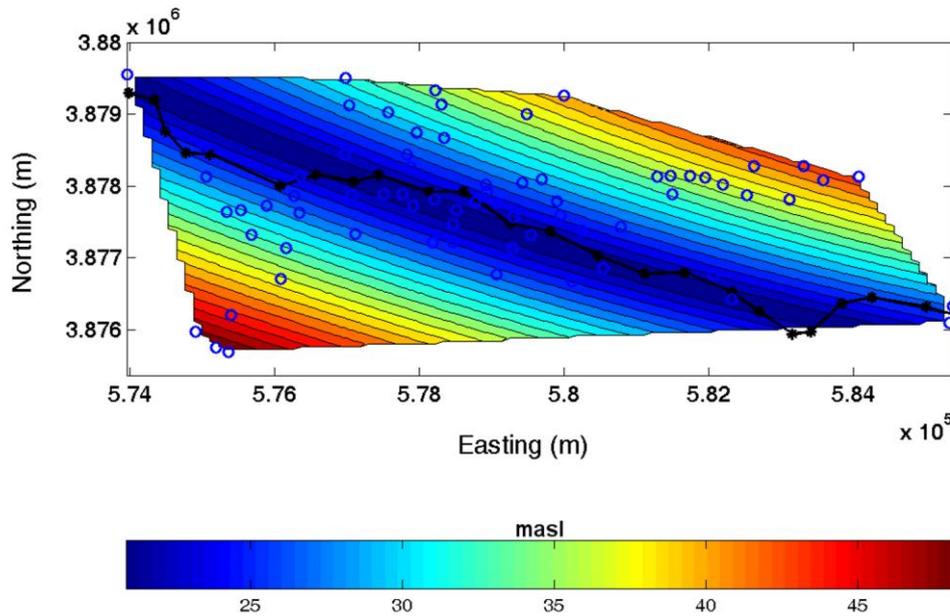
In this work we assess kriging-based methods for groundwater level interpolation (Table 6.3 to Table 6.8) using cross-validation metrics (section 2.7). The most accurate results are obtained with the RK method. We introduce two new trend components that improve the RK performance. The first one combines the closest distance of the sampling stations from the river and a uniform-gradient approximation to ground surface elevation (T-RD-DEM-UGA). The residuals are then normalized using the MBC method. Figure 6.12 and Figure 6.13 present the interpolated residuals and the determined trend distribution in physical space.



**Figure 6.12** Map of estimated groundwater level residuals in the Mires basin using RK-T-RD-DEM-UGA spatial model (interpolation of the residuals obtained from the subtraction of the measured values minus the trend determined from the approximation to ground surface elevation and the distance from the river curve), adapted on the real basin coordinates and location in the valley.

Using only the distance from the river (T-RD) model reduces the mean absolute error (MAE) to 3.11 masl compared to 3.30 masl and higher for the OK-based models. Using the full T-RD-DEM-UGA further reduces the MAE to approximately 3.00 masl. We obtain the most accurate results based on the second trend model that uses the hydraulic head distribution for a system of multiple

operating wells (T-MW). This leads to an MAE of 2.75 masl. On the other hand, the T-MW spatial model leads to higher standard deviation ( $\approx 2.80$  masl) compared T-RD-DEM-UGA (around 0.8 masl). We believe that this is due to the uncertainty in the estimation of the  $s_w$  and  $R$  parameters for the aquifer's wells.



**Figure 6.13** Map of estimated groundwater level trend in the Mires basin using RK-T-RD-DEM-UGA spatial model (trend using the gradient approximation to ground surface elevation and the distance from the river curve), adapted on the real basin coordinates and location in the valley.

T-RD, T-RD-DEM-UGA and T-MW (with the Spartan model) also give the most accurate leave-one-out cross-validation estimate, i.e.,  $\approx 29.0$  (T-RD, T-RD-DEM-UGA) and 27.7 masl (T-MW) at the extreme low level location (9.4 masl) compared to OK which yields 33 masl. In contrast, the highest level of 62 masl is accurately estimated by all interpolation methodologies tested.

The Spartan semivariogram function provides the optimal fit with the empirical semivariogram, with the power-law and the Matérn (with  $\nu = 0.92$  for original data,  $\nu = 0.46$  for residuals of T-RD,  $\nu = 0.34$  for residuals of T-RD-DEM-UGA and  $\nu = 0.48$  for residuals of T-MW) models following closely. The Spartan model is used for interpolation, because it gives slightly better cross validation measures for all methods tested. All of the above semivariogram models are non-differentiable. Fasbender et al. (Fasbender *et al.* 2008) also used the non-differentiable

spherical model for water table elevation. As we mention in section 4.4.3 we attribute the non-differentiability to the water table level being the result of a deposition-removal process. An idealized such process generates an fBm-like random field that has a power-law semivariogram function. The power-law semivariogram is unbounded, while the Spartan and Matérn models reach a finite sill. This difference is not important on the short and intermediate distances that are important for interpolation. From a practical viewpoint, the size of the basin and the deposition-removal increments is finite; hence, it makes sense that the semivariogram reach a sill at some large distance.



## 7. Stochastic space-time modeling of groundwater level variations in a Mediterranean basin

### 7.1 Introduction

Classical geostatistical analysis deals with spatial data variability. Geostatistical modeling however often needs to address variables that change in space as well as in time, such as groundwater level or pollution transport in air or porous media. When data is distributed through time and space, a major advantage is that higher number of data support parameter estimation and prediction. In a statistical context, these data can be considered as random fields spread out in space and evolving in time (space-time random fields-S/TRF). Usually spatiotemporal interpolation is performed by applying the standard kriging algorithms extended in a space-time frame. A historic review of space-time methodology is provided in section 2.3.

In this chapter we use stochastic methods for the spatiotemporal monitoring and prediction of the groundwater level in sparsely gauged basins. Sparsely monitored watersheds are not regularly monitored through space and time and therefore the data availability is not always appropriate for purely spatial or temporal analysis. Nevertheless the combination of the measured data can create a very useful dataset for spatiotemporal modeling and analysis by incorporating spatiotemporal correlations.

Time series of mean annual groundwater level data is available from ten boreholes that were monitored usually biannually in Mires basin (wet (April) and dry (September) period) between the years 1981 and 2003. Since then (2003-2010), data are available from two telemetric stations, (one of which belongs in the set of the 10 boreholes) that operate in the basin. For the time span 2003-2006 biannual measurements at four of the ten boreholes are available (wet period only for 2006). Annual precipitation measurements and pumping rate data are also available for the time period 1981-2010.

First, we model the temporal variation of the mean annual groundwater level in order to assess the aquifer's behavior during the last thirty years with respect to parameters that affect the water table fluctuations (e.g. precipitation, pumping). We use a discrete time autoregressive exogenous variable model (ARX) based on physical

motivation. The term “exogenous” denotes that the model equations incorporate information from auxiliary variables in addition to groundwater level. In this study precipitation measurements and pumping data are used. The ARX model is embedded in a discrete-time Kalman filter to estimate the model parameters and predict the optimal mean annual groundwater level. The ARX model is calibrated for the years 1981 to 2006 and is then used to predict the mean annual groundwater level in the basin for the recent years (2007-2010). The ARX model is calibrated with data up to the year 2006 in order to include the extremes of the aquifer behavior, i.e. the groundwater level increase during 2002-2003 and the decrease in 2005-2006. The predictions are validated with the available annual averages reported by the local authorities.

Secondly, we conduct a spatiotemporal geostatistical analysis of the groundwater level using space-time Residual Kriging (STRK). A space-time experimental semivariogram is determined from the biannual groundwater level time series between the years 1981 and 2003 at the ten sampling stations. We model the empirical semivariogram with separable and non-separable theoretical spatiotemporal semivariogram functions. STRK is used to predict the groundwater level for selected hydrological periods at each sampling station (every six months in the time period 2004-2010). The predictions are validated for the years up to 2006 (wet period). The average of the estimates is compared, for similar periods (2007-2010), to the groundwater level in the basin predicted by ARX and to the values reported by the local authorities based on the average of the two remotely sensed holes.

The ARX estimates are initially characterized by considerable initial fluctuations,  $\sigma_{err} \approx 9.6$  m for 1981-1995, which are then reduced to  $\sigma_{err} \approx 5.8$  m for 1996-2006 (according to equation (7.1)). The optimal non-separable semivariogram function, based on the diffusion equation, delivers significantly more accurate STRK predictions than the separable function (product model). Both ARX and STRK provide satisfactory predictions, but in contrast with ARX STRK also provides spatially distributed estimates.

## 7.2 Purely temporal variation analysis

Groundwater has an important role in the water resources balance of hydrological basins as it replenishes streams or wetlands and is a primary source of drinking and

agricultural water. Groundwater level reduction has implications for the water cycle, because groundwater supplies the base flow in many rivers and supports evapotranspiration in high water table regions (Famiglietti 2008). Examination and modelling of temporal trends of groundwater monitoring wells or of basin averages provide useful information about the aquifer temporal response to different meteorological (e.g. precipitation extremes) or anthropogenic effects (groundwater over-exploitation). Therefore the temporal variation modelling of the groundwater level in sparsely monitored basins provide a useful management tool of the basin's groundwater level that helps to assess the future trends.

### *7.2.1 Background of Regionalized ARX model of groundwater level*

A regionalized discrete time auto-regressive exogenous variable model (RARX) model that relates explicitly the precipitation surplus with water table depth was introduced by Bierkens *et al.* (2001) and Knotters and Bierkens (2001). The RARX-based approach is useful if the data are dense in time and sparse in space; for example, if sufficiently long time series of the water table depth are available at a limited number of locations. The RARX model is a linear time series model, the parameters of which depend on the location (this spatial dependence is referred to as regionalization). At locations where time series of water table depths are available, the RARX parameters are obtained from the model calibration process; which minimizes the error between the measured and the modeled value by adapting the parameter values. The parameters of RARX model are physically motivated. The spatial variability can be handled with classical geostatistics, such as kriging methods (OK, UK). The model parameters can also be estimated at other locations using auxiliary physical information, such as rainfall data and topography. Then, classical geostatistics approaches that incorporate auxiliary information can be used for regionalization (KED, RK) (Knotters and Bierkens 2001).

The RARX model can be combined with the Kalman filter algorithm (Ljung 1999, Knotters and Bierkens 2001). This approach permits (1) recursively determining the model parameters from the available data and (2) predicting water table depths in space and time conditionally on observed water table depths. Optionally, auxiliary information, such as meteorological variables and water usage can be incorporated in the model. The predictions are updated as new measurements of the water table depth are added to the time series. Updating is based on Bayes

theorem. This Bayesian updating property makes the Kalman filter attractive for water management purposes. The method also permits estimating the statistics of water table fluctuations. The accuracy of the estimates of the water table depth forward in time is evaluated as follows: expected depths are estimated for time periods with available observations, by also incorporating meteorological conditions (e.g. precipitation rates) during the monitoring period. The precipitation surplus is originally used as an exogenous variable in the RARX model. The parameters of the geostatistical model are estimated by treating the RARX model as an equation of state within the Kalman filter, and subsequently minimizing the mean square error of the filter innovations (i.e., the difference between the measured and predicted values). The root mean square prediction error is given by the following equation, i.e.,

$$\sigma_{err} = \sqrt{\frac{1}{N} \sum_{t=1}^N [Z_s(t) - \hat{Z}_s(t)]^2}, \quad (7.1)$$

where  $\hat{Z}_s(t)$  is the predicted water table depth,  $Z_s(t)$  is the corresponding measured value, and  $N$  is the length of the time series.

### 7.2.2 ARX Model for groundwater level in Mires basin

The regionalization approach is not applied for Mires basin due to the insufficient number of groundwater level monitoring wells (see section 7.1) that does not allow a reliable estimation of the model spatial variability at unmeasured locations. We therefore model the temporal variation of the mean annual groundwater level of the basin in order to simulate the historic groundwater level annual fluctuation that is determined from data reported by the regional department of water recourses management in Crete and Administration of Land Reclamation of the Prefecture of Crete for the basin (Figure 3.6).

We model the variation with a recursive discrete time auto-regressive exogenous variable model, (ARX) model. The ARX model is embedded in a discrete-time Kalman filter to estimate the model parameters and predict the optimal mean annual groundwater level. The model is defined by equation (7.2), where precipitation surplus was used only as an exogenous variable, proposed by Knotters and Bierkens (2001):

$$\overline{Z(t_k)} - c = a \left( \overline{Z(t_{k-1})} - c \right) + bP(t_k) + \varepsilon(t_k), \quad t_{k-1} = t_k - \Delta t, \quad (7.2)$$

where  $\overline{Z(t_k)}$  is the average groundwater level at time  $t_k$ ,  $\Delta t$  is the time step i.e.  $\Delta t = t_k - t_{k-1}$  and  $P(t_k)$  is the precipitation surplus over  $\Delta t$ .  $P(t_k)$  is defined in terms of the annual precipitation  $\overline{P^*(t_k)}$  averaged over the nearby rainfall stations over the respective time interval minus the mean annual actual evapotranspiration,  $E[\overline{P^*(t_k)}]$ , (see section 7.4.1) i.e.

$$P(t_k) = \overline{P^*(t_k)} - E[\overline{P^*(t_k)}]. \quad (7.3)$$

where,

$$\overline{P^*(t_k)} = \int_{t_k - \Delta t}^{t_k} dt P(t'), \quad (7.4)$$

and  $P(t')$  is the average daily precipitation.

The parameters  $a$ ,  $b$  determine the dynamic response of the water table, and  $c$  is a parameter that determines the average water table depth if  $P(t_k) = 0$ . The variable  $\varepsilon_t$  is a discrete white noise process with the following properties (where  $E[\cdot]$  denotes the expectation operator):

$$E[\varepsilon(t_k)] = 0, \quad (7.5)$$

$$E[\varepsilon(t_k)\varepsilon(u_e)] = \sigma_\varepsilon^2 \delta(t_k, u_e), \quad (7.6)$$

where  $\sigma_\varepsilon^2$  is the error variance and  $\delta(t_k, u_e)$  is the Kronecker delta defined by  $\delta(t_k, u_e) = 1$  if  $u_e = t_k$  and  $\delta(t_k, u_e) = 0$  if  $u_e \neq t_k$ .

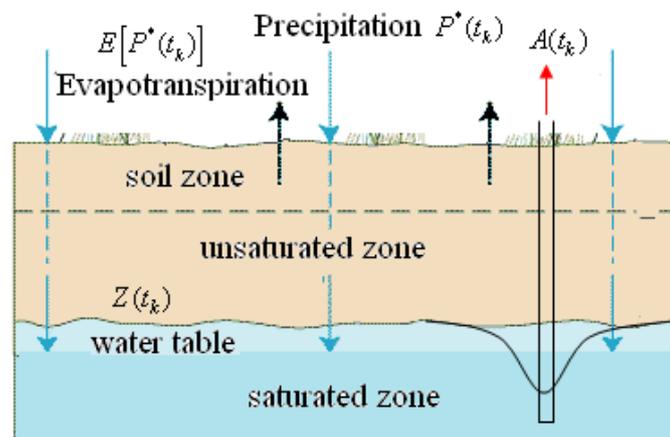
Herein we propose and apply an extension of the original model (7.2) in order to model the temporal variation of the mean annual groundwater level,

$$\overline{Z(t_k)} - c = a \left( \overline{Z(t_{k-1})} - c \right) + bP(t_k) + dA(t_k) + \varepsilon(t_k), \quad (7.7)$$

where  $A(t_k)$  is an estimate of the annual abstraction rate over  $\Delta t$  and  $d$  an additional parameters that determine with  $a$ ,  $b$  and  $c$  the dynamic response of the water table.

The order of the ARX model is defined by the triplet (1,0,0). The first entry denotes that equation (7.7), involves values of the water table depth with a maximum delay equal to one time step. The second and the third entry denote that the inputs, i.e., the precipitation surplus and the abstraction rate contain only one term with no delay.

The equation (7.2) has physical background. It describes the water table fluctuation for discrete time steps if zero surface runoff is assumed and a linear relation between water table depth and drainage from the groundwater zone to the surface water. Other variables that can be considered as exogenous (as the groundwater withdrawal rates herein or varying surface water levels) can also be incorporated in the model in a straightforward manner (Bierkens *et al.* 2001, Knotters and Bierkens 2001). It is assumed that the precipitation surplus  $P(t_k)$  is a global variable, that is, space invariant. This assumption is reasonable for relatively small areas, such as the Mires basin.



**Figure 7.1** Simplified demonstration of equation (7.7) inputs and output variables.

### 7.2.3 Kalman filter identification of ARX model

A Kalman filter comprises two sets of equations: one set predicts the state at the next time step, and the other set updates the predictions using available state measurements. The filter can be applied to any system described by a linear discrete-

time state-space equation given by the following general formulation (Van Geer and Van Der Kloet 1985, Eigbe *et al.* 1998, Ljung 1999):

$$\hat{\mathbf{z}}(t_k) = \mathbf{F}(t_k)\mathbf{z}(t_{k-1}) + \mathbf{B}(t_k)\mathbf{p}(t_k) + \mathbf{w}(t_k), \quad (7.8)$$

where  $\hat{\mathbf{z}}(t_k)$  is the vector of system state variables at time  $t_k$ ,  $\mathbf{F}(t_k)$  is the state transition matrix from time moment  $t_{k-1}$  to time  $t_k$ ,  $\mathbf{B}(t_k)$  is the control matrix that represents the impact of external inputs  $\mathbf{p}(t_k)$  on the state of the system at time  $t_k$ , while  $\mathbf{w}(t_k)$  is a Gaussian noise vector that accounts for random noise (Bierkens *et al.* 2001). The noise properties are defined by,

$$E[\mathbf{w}(t_k)] = 0, \quad (7.9)$$

$$E[\mathbf{w}(t_k)\mathbf{w}(u_e)^T] = \mathbf{Q}(t_k)\delta(t_k, u_e), \quad (7.10)$$

where  $\delta(t_k, u_e)$  is the Kronecker delta defined above in section 7.2.2.  $\mathbf{Q}(t_k)$  is the covariance matrix of estimation errors. For any vector  $\mathbf{w}_i$  multiplied by  $\mathbf{w}_j$  it holds that  $(\mathbf{w}_i \cdot \mathbf{w}_j = \mathbf{Q}_{i,j})$ .

The measurement equation relating the observed state variables and the true state of the system is expressed as (Van Geer and Van Der Kloet 1985, Eigbe *et al.* 1998, Ljung 1999):

$$\mathbf{z}(t_k) = \mathbf{M}(t_k)\hat{\mathbf{z}}(t_k) + \mathbf{v}(t_k), \quad (7.11)$$

where  $\mathbf{z}(t_k)$  is the vector of observed state variables,  $\mathbf{M}(t_k)$  is the observation matrix and  $\mathbf{v}(t_k)$  is the observation noise vector that accounts for measurement errors. If the element in row  $i$  of the vector  $\mathbf{z}(t_k)$  is the observation of the state variable  $\hat{\mathbf{z}}(t_k)$  located in row  $j$ , then the element  $(i,j)$  of  $\mathbf{M}(t_k)$  is set to “1”; all other elements of the row  $\mathbf{M}(t_k)$  are set to “0”. The noise process  $\mathbf{v}(t_k)$  is assumed to be multi-

dimensional Gaussian white random processes. The following equations define the statistical properties of  $\mathbf{v}(t_k)$ :

$$E[\mathbf{v}(t_k)] = \mathbf{0}, \quad (7.12)$$

$$E[\mathbf{v}(t_k)\mathbf{v}(u_e)^T] = \mathbf{R}(t_k)\delta(t_k, u_e), \quad (7.13)$$

where  $\mathbf{R}(t_k)$  is the covariance matrix of the observation errors. It is usually assumed that the observation errors are independent, which means that  $\mathbf{R}(t_k)$  is a diagonal matrix whose elements equal the variances of the corresponding observation errors. The observation errors are also independent of the system noise, i.e.:

$$E[\mathbf{w}(t_k)\mathbf{v}(u_e)^T] = \mathbf{0} \quad \forall (t_k, u_e). \quad (7.14)$$

To execute the estimation procedure the matrices  $\mathbf{F}$ ,  $\mathbf{B}$ ,  $\mathbf{M}$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  must be known. The calculations can be divided into two steps. First, a prediction is made for the state at time  $t_k$ , based on measurements up to time  $t_{k-1}$ . Secondly at time  $t_k$  as the new measurement becomes available the prediction is corrected. This yields the optimal linear estimate for the state at time  $t_k$  based on measurements up to time  $t_k$ . The matrices  $\mathbf{F}$  and  $\mathbf{B}$  are functions of the system parameters, which are not known a priori. Therefore the Kalman filter cannot be directly applied. The following set of equations describes the *Kalman filter adaptation algorithm* that estimates recursively the parameters and the state equation output of an ARX model (Ljung 1999).

#### 7.2.4 Kalman filter adaptation algorithm of ARX model

Linear model structures such as ARX that are equivalent to linear regressions can be expressed as (Ljung 1999, Lanzi *et al.* 2006):

$$z(t_k) = \boldsymbol{\chi}(t_k)^T \boldsymbol{\theta}(t_k) + v(t_k). \quad (7.15)$$

In the above equation,  $\chi(t_k)$  is the gradient of the predicted model output  $z(t_k)$  with respect to the parameter vector and represents the model's *regression vector*,  $\theta(t_k)$  represents the vector of the true parameters (true description of the system) and  $v(t_k)$  is the measurement error innovation. The predicted output is given by:

$$\hat{z}(t_k) = \chi(t_k)^T \hat{\theta}(t_{k-1}) \quad (7.16)$$

where  $\hat{z}(t_k)$  is the prediction of  $z(t_k)$  based on parameters up to  $t_{k-1}$  and  $\chi(t_k)$  is a  $(n \times 1)$  vector of gradient values (regression vector) that represents the gradient of the predicted model output  $\hat{z}(t_k)$  with respect to the parameters  $\hat{\theta}(t_{k-1})$ . Since the true parameters of the system are unknown it is assumed that  $\theta(t_k) = \hat{\theta}(t_{k-1})$ . The estimation algorithm minimize the mean square prediction-error term  $E[(z(t_k) - \hat{z}(t_k))^2]$ , which means that equation (7.16) is solved for all time steps using the parameters  $\hat{\theta}(t_{k-1})$ .

The general recursive parameter identification equation is:

$$\hat{\theta}(t_k) = \hat{\theta}(t_{k-1}) + \mathbf{K}(t_k) [z(t_k) - \hat{z}(t_k)], \quad (7.17)$$

where  $\hat{\theta}(t)$  is a vector of the parameter estimates  $(n \times 1)$  at time  $t_k$ ,  $z(t_k)$  is the observed output at time  $t_k$  and  $\hat{z}(t_k)$  is the prediction of  $z(t_k)$  based on observations up to time  $t_{k-1}$ .  $\mathbf{K}(t_k)$  is the Kalman gain, a  $(n \times 1)$  vector, that determines how much the current prediction error  $z(t_k) - \hat{z}(t_k)$  affects the update of the parameters estimate.

The above formulation assumes that the true system parameters  $\theta(t_k)$  are described by a random walk:

$$\theta(t_k) = \theta(t_{k-1}) + \mathbf{w}_\theta(t_k), \quad (7.18)$$

where  $\mathbf{w}_\theta(t_k)$  is a Gaussian white noise with covariance matrix  $E[\mathbf{w}_\theta(t_k)\mathbf{w}_\theta(t_k)^T] = \mathbf{R}_1(t_k)$ .  $\mathbf{R}_1(t_k)$  is the covariance matrix ( $n \times n$ ) of the parameter changes (error) for each time step  $t_k$ .

The Kalman gain  $\mathbf{K}(t_k)$  is derived based on the Least Mean Square (LMS) parameter estimation algorithm and it has the following general form (Ljung 1999, Lanzi *et al.* 2006),

$$\mathbf{K}(t_k) = \mathbf{J}(t_k)\boldsymbol{\chi}(t_k), \quad (7.19)$$

where  $\mathbf{J}(t_k)$  is given by,

$$\mathbf{J}(t_k) = \frac{\mathbf{P}(t_{k-1})}{R_2(t_k) + \boldsymbol{\chi}(t_k)^T \mathbf{P}(t_{k-1}) \boldsymbol{\chi}(t_k)} \quad (7.20)$$

and

$$S(t_k) = R_2(t_k) + \boldsymbol{\chi}(t_k)^T \mathbf{P}(t_{k-1}) \boldsymbol{\chi}(t_k). \quad (7.21)$$

In the above equations  $\mathbf{J}(t_k)$  is a ( $n \times n$ ) covariance matrix, the ( $n \times n$ ) covariance matrix  $\mathbf{P}(t_{k-1})$  corresponds to parameters estimation error at  $t_{k-1}$ ,  $S(t_k)$  is the residual (innovation) covariance during the parameters update process and  $R_2(t_k)$  is the variance of the innovations  $v(t_k)$  in equation (7.15):  $R_2(t_k) = E[v^2(t_k)]$  (a scalar).

The ( $n \times n$ ) covariance matrix  $\mathbf{P}(t_k)$  represents the parameter error covariance and is updated as follows:

$$\mathbf{P}(t_k) = \mathbf{P}(t_{k-1}) + \mathbf{R}_1(t_k) - \frac{\mathbf{P}(t_{k-1})\boldsymbol{\chi}(t_k)\boldsymbol{\chi}(t_k)^T \mathbf{P}(t_{k-1})}{R_2(t_k) + \boldsymbol{\chi}(t_k)^T \mathbf{P}(t_{k-1}) \boldsymbol{\chi}(t_k)}. \quad (7.22)$$

The Kalman filter algorithm is entirely specified by the sequence of data  $z_t$ , the gradient  $\boldsymbol{\chi}(t_k)$ , the covariance matrix  $R_1$ , the variance of the innovations  $R_2$ , the true parameters  $\boldsymbol{\theta}(t_k)$  or an initial guess and the parameter error covariance matrix  $P_0(t_k)$ . The recursive estimate of the parameters and of the output of an auto-

regressive exogenous variable model, (ARX) model, is implemented in Matlab<sup>®</sup> programming environment using the “*rarx*” function after the appropriate coding of the state and measurement equations.

For this application  $\hat{\boldsymbol{\theta}}(t)$  is the vector of the parameter estimates ( $4 \times 1$ ) ( $a$ ,  $b$ ,  $c$  and  $d$  of the state equation (7.7)) at time  $t_k$ ,  $\mathbf{K}(t_k)$  is a ( $4 \times 1$ ) vector,  $\boldsymbol{\chi}(t_k)$  is a ( $4 \times 1$ ) vector of gradient values (regression vector) corresponding to each variable input involved in the state equation (7.7) ( $\overline{Z(t_{k-1})}$ ,  $P(t_k)$ ,  $A(t_k)$ ), unit value for parameter  $c$ ),  $\mathbf{J}(t_k)$  a ( $4 \times 4$ ) covariance matrix and  $\mathbf{P}(t_k)$  a ( $4 \times 4$ ) covariance matrix that corresponds to parameters estimation error.

### 7.3 Spatiotemporal geostatistics

Spatiotemporal geostatistical models provide a probabilistic framework for data analysis and predictions which is based on the joint spatial and temporal dependence between observations (Kyriakidis and Journel 1999, Fischer and Getis 2010). Initial approaches to spatiotemporal data modeling were based on separable covariance functions, obtained by combining separate spatial and temporal covariance models (Rodriguez-Iturbe and Mejia 1974, Rouhani and Myers 1990, Cressie 1993, Dimitrakopoulos and Luo 1994). The last two decades there is significant development of non-separable covariance functions. These models aim to improve spatiotemporal data modeling and prediction (Cressie and Huang 1999, De Iaco *et al.* 2001, Gneiting 2002, Kolovos *et al.* 2004) by extracting in some case the covariance functions from physical laws such as differential equations and dynamic rules (Christakos and Hristopoulos 1998, Christakos 2000, Gneiting 2002, Kolovos *et al.* 2004).

The main goal of space-time analysis is to model multiple time series of data at spatial locations where a distinct time series is allocated. The time variable is considered as an additional dimension in geostatistical prediction. A spatiotemporal stochastic process can be represented by  $Z(\mathbf{s}, t)$  where the variable of interest of random field  $Z$  is observed at  $N$  space-time coordinates  $(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_N, t_N)$ , while the optimal prediction of the variable in space and time is based on  $Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_N, t_N)$  (Cressie and Huang 1999, Giraldo Henao 2009). S/TRF  $Z(\mathbf{s}, t)$  can be decomposed

into a mean component  $m_z(\mathbf{s}, t)$  modeling the trend and a residual S/TRF component  $Z'(\mathbf{s}, t)$  modeling fluctuations around that trend in both space and time according to the following equation:

$$Z(\mathbf{s}, t) = m_z(\mathbf{s}, t) + Z'(\mathbf{s}, t). \quad (7.23)$$

The trend can be calculated either deterministically and the fluctuation using a stochastic framework such as space-time kriging (Christakos 1991b, Kyriakidis and Journel 1999).

### 7.3.1 Spatiotemporal two point function

Set  $Z(\mathbf{s}, t)$ ,  $(\mathbf{s}, t) \in D \times T$ , a second-order stationary space–time random field.  $D \subseteq \mathbb{R}^d$  is the spatial domain ( $d$  is the space dimensions) and  $T \subseteq \mathbb{R}$  is the temporal domain, with expected value (Myers *et al.* 2002):  $E[Z(\mathbf{s}, t)] = 0$ ,  $\forall (\mathbf{s}, t) \in D \times T$  and covariance function:

$$C_{ST}(\mathbf{r}_s, r_t) = E[Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t) Z(\mathbf{s}_i, t_i)] - E[Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t)] E[Z(\mathbf{s}_i, t_i)], \quad (7.24)$$

where  $\mathbf{r}_s = (\mathbf{s}_i - \mathbf{s}_j)$ ,  $r_t = (t_i - t_j)$ ,  $i, j = 1, \dots, N$ . The covariance function depends only on the lag vector  $\mathbf{r} = (\mathbf{r}_s, r_t)$  and not on location or time, while it must satisfy the positive-definiteness condition in order to be a valid covariance function. Hence, for any  $(\mathbf{s}_i, t_i) \in D \times T$ , any real  $a_i$ ,  $i = 1, \dots, N$  and any positive integer  $N$ ,  $C_{ST}$  must satisfy the following inequality:

$$\sum_{i=1}^N \sum_{j=1}^N a_i a_j C_{ST}(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) > 0^5.$$

If  $E[Z(\mathbf{s}, t)]$  is constant and  $C_{ST}(\mathbf{r}_s, r_t)$  depends only on the lag vector  $\mathbf{r} = (\mathbf{r}_s, r_t)$ :

---

<sup>5</sup> positive-definiteness condition is often presented also as non-negative definiteness condition, i.e.  $\geq 0$

$$\text{Cov}(Z(\mathbf{s}_i, \mathbf{s}_j; t_i, t_j)) = C_{ST}(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) = C_{ST}(\mathbf{r}_s, r_t). \quad (7.25)$$

the S/TRF  $Z(\mathbf{s}, t)$  is characterized as *second-order stationary*. Spatial and spatiotemporal geostatistical prediction methodologies generally rely on stationarity (stationary mean and covariance or semivariogram).

In addition the field is isotropic if ,

$$C_{ST}(\mathbf{r}_s, r_t) = C_{ST}(\|\mathbf{r}_s\|, |r_t|), \quad (7.26)$$

meaning that the covariance function depends only on the length of the lag.

Under the weaker intrinsic stationarity assumption the increment  $(Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t) - Z(\mathbf{s}_i, t_i))$  is second order stationary for every lag vector  $\mathbf{r}_s, r_t$  instead of the random field. Then  $Z(\mathbf{s}, t)$  is called an intrinsic random function and is characterized by:

$$E(Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t) - Z(\mathbf{s}_i, t_i)) = 0, \quad (7.27)$$

and

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \frac{1}{2} \text{var}(Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t) - Z(\mathbf{s}_i, t_i)) \quad (7.28)$$

where the term *var* denotes the variance. The function  $\gamma_{ST}(\mathbf{r}_s, r_t)$  only depends on the lag vector  $\mathbf{r} = (\mathbf{r}_s, r_t)$ . The quantity  $\frac{1}{2} \text{var}(Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t) - Z(\mathbf{s}_i, t_i))$  is called the semi-variance at lag  $\mathbf{r} = (\mathbf{r}_s, r_t)$ .

The random field  $Z(\mathbf{s}, t)$  has an intrinsically stationary semivariogram if it is intrinsically stationary with respect to both the space and the time dimensions. The  $Z(\mathbf{s}, t)$  has a spatially intrinsically stationary semivariogram if the semivariogram depends only on the spatial separation vector  $\mathbf{r}_s$  for every pairs of time instants  $t_i, t_j$  and it has temporally intrinsically stationary semivariogram if depends only on the temporal lag  $r_t$ . Equation (7.28) provides the space-time stationary semivariogram

function (Gneiting *et al.* 2007). Under the stronger assumption of second-order stationarity, the semi-variance is defined as:

$$\gamma_{ST}(\mathbf{r}_s, r_t) = C_{ST}(0,0) - C_{ST}(\mathbf{r}_s, r_t). \quad (7.29)$$

The primary concerns when modeling space–time structures, is to ensure that the chosen model is valid and that the model is suitable for the data. The space-time kriging estimator can be applied if the space-time covariance function satisfies the positive definiteness condition,  $C_{ST} > 0$  explained above (Cressie and Huang 1999). The model’s suitability is ensured by testing a series of available structures on the data. The semivariogram function must be conditionally negative definite to ensure that the space-time kriging equations have a valid unique solution (Myers *et al.* 2002, De Iaco 2010).

Space-time kriging is a well-established method for space-time interpolation (Christakos *et al.* 2001, De Cesare *et al.* 2001). It is however complicated, as the kriging system of equations needs to be solved at the same time for spatial and temporal weights (Skøien and Blöschl 2007). In addition, space-time kriging is data demanding, while often the number of locations where time series of groundwater level data are available is very limited. Also according to Bierkens (2001), space-time kriging may not be appropriate to analyze the change in groundwater level if climate change effects (rainfall shortage, intense rainfall periods, droughts) affect the area of study or changes of land use and surface water management occur. The kriging estimator and the kriging equations have the same form for spatiotemporal problems as for spatial problems. The difference from spatial-only kriging is the covariance modeling. This is because the time component is not an extra dimension that can be used to form a single Euclidean space-time metric. The time axis is by nature different and not necessarily orthogonal to the three spatial axes. The time component has been proved to cause both theoretical and practical problems if it is treated as an additional space dimension (Rouhani and Myers 1990). Hence, in space-time variography the spatial lag,  $\mathbf{r}_s$ , and the temporal lag,  $r_t$ , are treated as independent arguments  $\gamma(\mathbf{r}_s; r_t)$ . Space-time and purely spatial kriging methods were analytically presented and compared on simulated data by Bogaert (1996). He concludes that in the space-time context ordinary space-time kriging (OSTK) is preferable; nevertheless, it requires the

hypothesis of mean and variance homogeneity and is limited to second order stationary random fields.

Two categories of models are used for semivariogram or covariance modeling. The first includes separable models whose covariance function is a combination of a spatial and a temporal covariance function; the second includes non-separable models in which a single function models the spatiotemporal data dependence. Separable models however, suffer from unrealistic assumptions and properties (Snepvangers *et al.* 2003, Hengl *et al.* 2011). Both space-time covariance models are valid according to (De Iaco *et al.* 2001, 2002b) and (Cressie and Huang 1999).

Separable and non-separable covariance functions can describe the random field's spatiotemporal continuity. Separable covariance functions are used if separate spatial and separate temporal covariance functions exist for the data (Gneiting *et al.* 2007),

$$\text{Cov}(Z(\mathbf{s}_i, \mathbf{s}_j; t_i, t_j)) = C_{ST}(\mathbf{r}_s, r_t) = C_{\mathbf{r}_s}(\mathbf{s}_i, \mathbf{s}_j) C_{r_t}(t_i, t_j). \quad (7.30)$$

Separability provides many advantages, such as the simplified representation of the covariance matrix and consequently important computational benefits (Park and Fuentes 2008). The separable covariance models however, in spite of their simplicity are not usually physically motivated. Correlations that have separable spatial and temporal components are particularly useful when the correlations are inferred on the basis of existing data and not when they follow from a physical model (Christakos and Hristopulos 1998). When data (e.g. hydrologic, atmospheric, oceanographic) are influenced by dynamic processes spatiotemporal dependency structures are difficult to be modeled by a separable covariance function. Physically meaningful covariance models can be derived instead, based on environmental data dynamic processes (Christakos 1991b, Christakos and Hristopulos 1998, Gneiting 2002, Kolovos *et al.* 2004). Covariance structures such these are non-separable. Modeling non-separable covariance functions is one of the keys for more reliable prediction in the environmental research fields (Gneiting *et al.* 2007).

The random field  $Z(\mathbf{s}, t)$  has fully symmetric separable covariance if (Gneiting *et al.* 2007, De Iaco 2010):

$$\text{Cov}(Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)) = \text{Cov}(Z(\mathbf{s}_i, t_j), Z(\mathbf{s}_j, t_i)), \quad (7.31)$$

for all  $(\mathbf{s}_i, t_i)$ ,  $(\mathbf{s}_j, t_j)$ . Non-separable covariance structures that are not fully symmetric have been proposed by Gneiting *et al.* (2007) e.g.:

$$C_{ST}(\mathbf{r}_s, r_t) = (1 + r_t)^{-1} \exp\left[\mathbf{r}_s / (1 + r_t)^{\beta/2}\right], \quad 0 \leq \beta \leq 1. \quad (7.32)$$

A significant part in the space-time process is the choice of the semivariogram or covariance model and the estimation of its parameters. Contrary to purely spatial prediction, where a well established set of semivariogram models exists, several spatiotemporal models have been developed for modeling space-time structures (Christakos and Hristopoulos 1998, Kyriakidis and Journel 1999, De Cesare *et al.* 2001, Gething *et al.* 2007). These models involve the product model (Rodriguez-Iturbe and Mejia 1974), the sum model (Rouhani and T.J. Hall 1989), the metric model (Dimitrakopoulos and Luo 1994), the integrated product model (Cressie and Huang 1999, De Iaco *et al.* 2002a), the product–sum model (De Cesare *et al.* 2001, De Cesare *et al.* 2002), the integrated product–sum model (De Iaco *et al.* 2002a, b), Gneiting’s non-separable models (Gneiting 2002, Gneiting *et al.* 2007), a series of non-separable models reviewed in (Kolovos *et al.* 2004) and non-separable models expressed through the spectrum density function instead of the direct covariance function (Porcu *et al.* 2008).

### 7.3.2 Spatiotemporal covariance or semivariogram models

A comprehensive description of some widely used spatiotemporal covariance or semivariogram models follows.

The metric model is given by the following equation (Dimitrakopoulos and Luo 1994):

$$C_{ST}(\mathbf{r}_s, r_t) = C\left(\alpha_1 \|\mathbf{r}_s\|^2 + \alpha_2 |r_t|^2\right), \quad (7.33)$$

or

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \gamma\left(\alpha_1 \|\mathbf{r}_s\|^2 + \alpha_2 |r_t|^2\right), \quad (7.34)$$

where  $\|\mathbf{r}_s\|$  is the Euclidean norm of the spatial lag vector and  $\alpha_1, \alpha_2$  coefficients that weigh relatively the space and time contributions. For this model the spatial and temporal covariances used are of the same type.

Another separable space-time covariance model is the sum model, in which spatial  $C_S(\mathbf{r}_s)$  and temporal  $C_T(r_t)$  covariance functions are added (Rouhani and T.J. Hall 1989):

$$C_{ST}(\mathbf{r}_s, r_t) = C_S(\mathbf{r}_s) + C_T(r_t), \quad (7.35)$$

or

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \gamma_S(\mathbf{r}_s) + \gamma_T(r_t). \quad (7.36)$$

In the above  $C_{ST}$  is the spatiotemporal covariance and  $\gamma_{ST}$  is the spatiotemporal semivariogram respectively. According to Rouhani and Myers (1990), covariance matrices  $C_{ST}(\mathbf{r}_s, r_t)$  of certain configurations of space-time data can be singular. In this case the covariance function is only positive semi-definite  $C_{ST} \geq 0$  (De Iaco 2010). The sum expression therefore is nearly an acceptable model as it only fails the strict definiteness condition. The resulting spatial-temporal form of covariance or semivariogram does not satisfy the strict definiteness conditions for the separate spatial and temporal covariances and the strict conditional negative definiteness condition for the separate spatial and temporal semivariograms (Myers and Journel 1990, Rouhani and Myers 1990, Dimitrakopoulos and Luo 1994). Thus this model is unsatisfactory for optimal prediction (De Iaco 2010).

The product model (Rodriguez-Iturbe and Mejia 1974) belongs to the separate space-time model category and is one of the simplest ways to model a covariance or semivariogram in space-time. The product of a space semivariogram and a time semivariogram is generally not a valid semivariogram; on the other hand, the product of a space covariance and a time covariance leads to a valid model. A semivariogram structure can then be determined by the product covariance model. Valid spatial and

temporal covariance models can be used in the product form below to create spatiotemporal models,

$$C_{ST}(\mathbf{r}_s, r_t) = C_S(\mathbf{r}_s)C_T(r_t). \quad (7.37)$$

If both components  $C_S(\mathbf{r}_s), C_T(r_t)$  are strictly positive definite then  $C_{ST}(\mathbf{r}_s, r_t)$  is strictly positive definite on  $\mathbf{R}^d \times T$ . The covariance equation can be expressed in terms of the semivariogram as:

$$\gamma_{ST}(\mathbf{r}_s, r_t) = C_T(0)\gamma_S(\mathbf{r}_s) + C_S(0)\gamma_T(r_t) - \gamma_S(\mathbf{r}_s)\gamma_T(r_t). \quad (7.38)$$

The product-sum space-time model (De Cesare *et al.* 2001, De Cesare *et al.* 2002) is a generalization of the product and the sum model, while it constitutes the starting point for its integrated product sum versions. It is defined as:

$$C_{ST}(\mathbf{r}_s, r_t) = k_1C_S(\mathbf{r}_s)C_T(r_t) + k_2C_S(\mathbf{r}_s) + k_3C_T(r_t). \quad (7.39)$$

$C_S, C_T$  are purely spatial and temporal covariance models with  $k_1 > 0, k_2 \geq 0, k_3 \geq 0$ . If  $C_S(\mathbf{r}_s)$  and  $C_T(r_t)$  are strictly positive definite, then  $C_{ST}(\mathbf{r}_s, r_t)$  is strictly positive definite on  $\mathbf{R}^d \times T$ . In terms of the semivariogram, the above equation is expressed as:

$$\gamma_{ST}(\mathbf{r}_s, r_t) = (k_1C_S(0) + k_3)\gamma_T(r_t) + (k_1C_T(0) + k_2)\gamma_S(\mathbf{r}_s) - k_1\gamma_S(\mathbf{r}_s)\gamma_T(r_t), \quad (7.40)$$

where  $\gamma_S, \gamma_T$  are purely spatial and temporal semivariogram models.  $C_S(0)$  and  $C_T(0)$  are the sills of the spatial and temporal semivariograms respectively. Each space-time model (sum, product) separately have limitations which their combination does not have. The semivariogram structure can be expressed alternatively as follows (De Iaco *et al.* 2001):

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \gamma_{ST}(\mathbf{r}_s, 0) + \gamma_{ST}(0, r_t) - \mathbf{K}\gamma_{ST}(\mathbf{r}_s, 0)\gamma_{ST}(0, r_t), \quad (7.41)$$

where  $0 < K \leq 1/\max(\text{sill } \gamma_{ST}(\mathbf{r}_s, 0), \text{sill } \gamma_{ST}(0, r_t))$ .

The Cressie-Huang models (Cressie and Huang 1999) are non-separable spatiotemporal stationary covariance functions defined by,

$$C_{ST}(\mathbf{r}_s, r_t) = \int e^{i\mathbf{r}_s^T \boldsymbol{\omega}} \rho(\boldsymbol{\omega}, r_t) k(\boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (7.42)$$

where  $\rho(\boldsymbol{\omega}, \bullet)$  is a continuous auto correlation function  $\forall \boldsymbol{\omega} \in \mathbf{R}^d$  and  $k(\bullet)$  a positive function with  $k(\boldsymbol{\omega}) > 0$  and  $\int k(\boldsymbol{\omega}) d\boldsymbol{\omega} < \infty$ . Bochner's theorem is used to derive non-separable space-time covariance functions of this type.

Gneiting (2002) proposed a wide class of non-separable covariances derived from the following equation:

$$C_{ST}(r_s, r_t) = \frac{\sigma^2}{[\psi(r_t^2)]^{d/2}} \varphi\left(\frac{\|\mathbf{r}_s\|^2}{\psi(r_t^2)}\right), (\mathbf{r}_s, r_t) \in \mathbf{R}^d \times T, \quad (7.43)$$

where  $d$  is the number of spatial dimensions,  $\varphi(\tau)$ ,  $\tau \geq 0$ , is a completely monotone function and  $\psi(\tau)$ ,  $\tau \geq 0$ , is a positive function (i.e. Bernstein function or equivalently a variogram) with a completely monotone derivative. A real and positive function  $f: [0, \infty] \rightarrow [0, \infty]$  is called completely monotone if and only if,  $(-1)^N f^{(N)}(\tau) \geq 0$ , for any positive integer  $N$  (Porcu *et al.* 2006). Examples of such functions are given in Gneiting (2002). The spatial and temporal structures are determined by  $\varphi$  and  $\psi$  respectively. However, contrary to the Cressie-Huang models the Gneiting models do not recall the Bochner's theorem.

A similar approach to the Cressie-Huang models can be formulated for the product (7.37) and product-sum (7.39) constructions. Their integration also gives valid spatiotemporal models (De Iaco *et al.* 2002a, Myers *et al.* 2002) as follows:

$$C_{ST}(\mathbf{r}_s, r_t) = \int_V k C_S(\mathbf{r}_s; a) C_T(r_t; a) d\mu(a), \quad (7.44)$$

where  $k > 0$ ,  $\mu(a)$  is a positive measure on  $U \subseteq \mathbf{R}$ ,  $C_S(\mathbf{r}_s; a), C_T(r_t; a)$  are valid covariance functions in  $D \subseteq \mathbf{R}^d$  and  $T \subseteq \mathbf{R}$  respectively for each  $a \in V \subseteq U$ , and  $C_S(\mathbf{r}_s; a)C_T(r_t; a)$  is integrable with respect to the measure  $\mu$  on  $V$  for all  $\mathbf{r}_s, r_t$ . The integrated product model generates non-separable and non-integrable models.

In terms of the semivariogram structure the above equation is rewritten as:

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \int_V k \left[ C_T(0; a) \gamma_S(\mathbf{r}_s; a) + C_S(0; a) \gamma_T(r_t; a) - \gamma_S(\mathbf{r}_s; a) \gamma_T(r_t; a) \right] d\mu(a) \quad (7.45)$$

Similarly for the product-sum model one obtain:

$$C_{ST}(\mathbf{r}_s, r_t) = \int_V \left[ k_1 C_S(\mathbf{r}_s; a) C_T(r_t; a) + k_2 C_S(\mathbf{r}_s; a) + k_3 C_T(r_t; a) \right] d\mu(a), \quad (7.46)$$

where  $k_1 C_S(\mathbf{r}_s; a) C_T(r_t; a) + k_2 C_S(\mathbf{r}_s; a) + k_3 C_T(r_t; a)$  is integrable with respect to the measure  $\mu$  on  $V$  for all  $\mathbf{r}_s, r_t$  given  $k_1 > 0$ ,  $k_2 \geq 0$ ,  $k_3 \geq 0$ . Separate structures in space and time are used to generate the product-sum model and integrated product-sum models. In addition, the integrated product-sum model (7.46) and the product-sum model (7.39) are non-integrable with respect to  $\mathbf{r}_s$  and  $r_t$  and non-separable. Equation (7.46) can be written in terms of a semivariogram structure as:

$$\gamma_{ST}(\mathbf{r}_s, r_t) = \int_V \left[ (k_1 C_S(0; a) + k_3) \gamma_T(r_t; a) + (k_1 C_T(0; a) + k_2) \gamma_S(\mathbf{r}_s; a) - k_1 \gamma_S(\mathbf{r}_s; a) \gamma_T(r_t; a) \right] d\mu(a), \quad (7.47)$$

where  $C_S(0; a)$  and  $C_T(0; a)$  are the corresponding sill values of the spatial and temporal semivariograms. Both space-time semivariogram structures are valid if  $\gamma_S(\mathbf{r}_s; a)$  and  $\gamma_T(r_t; a)$  are valid spatial and temporal semivariogram models.

### 7.3.3 Spatiotemporal models' summary of characteristics

The metric model in spite of its nice asymptotic features has restrictive assumptions. As previously mentioned the same type of covariances describe the spatial and temporal correlation, they have the same sill (if the model is bounded) and it can be used only for processes, whose space–time correlation is described by a model with geometric anisotropy. Finally, it is the only model that requires a space–time metric (De Iaco 2010).

The product model, the product-sum model, their integrated versions and the sum model are produced by separate space and time functions. The main advantage of such models is their ease of use in modeling and estimation. Because the sum model is separable, anisotropy can be incorporated in the spatial component. The product model is separable and integrable; its integrated version can generate non-separable and non-integrable models. In contrast the product-sum model is non-integrable with respect to  $\mathbf{r}_s$  and  $r_t$ , and it is non-separable as the integrated version of the product-sum model. On the other hand, the Cressie–Huang and Gneiting models are alternative choices to the separable models. However the  $\varphi$  and  $\psi$  functions in the Gneiting model can be chosen so that separable models are obtained. Finally, anisotropic covariance or semivariogram versions of the functions described can be constructed by inserting anisotropy in the spatial component of the semivariogram function (Myers *et al.* 2002, De Iaco 2010).

The space-time semivariogram models described above, except for the metric function, are typically used to model the space-time experimental semivariogram because an arbitrary space-time metric is not required and the fitting process is similar to that for spatial semivariograms (Gething *et al.* 2007, De Iaco 2010).

### 7.3.4 A Spatiotemporal covariance function derived from a physical law

An alternative covariance function suitable for space-time semivariogram modeling is proposed by (Christakos and Hristopulos 1998, Kolovos *et al.* 2004). It is an extension of a non-separable spatiotemporal covariance inspired from the diffusion equation. A covariance model derived from a physical differential equation, such as the diffusion equation is of the form (Christakos 2000):

$$C_{ST}(\mathbf{r}_s, r_t) = (4\alpha\pi|r_t|)^{-n/2} \exp(-\mathbf{r}_s^2 / 4\alpha|r_t|), \quad \alpha > 0. \quad (7.48)$$

An extension of this equation can be obtained following Gneiting's proposition to add constants after the time lag in space-time formulations. (Hristopulos 2002) proposed a similar approach for the spatial-only case. Therefore the equation can be modified to:

$$C_{ST}(\mathbf{r}_s, r_t) \cong (\beta r_t^{2\gamma} + 1)^{-n/2} \exp(-\mathbf{r}_s^2 / (\beta r_t^{2\gamma} + 1)), \quad 0 \leq \beta \leq 1, \quad 0 < \gamma \leq 1, \quad (7.49)$$

where  $\beta$ ,  $\gamma$  are the function's parameters and  $n$  are the dimensions. This covariance class has been used in the area of fluid mechanics (Monin and Yaglom 1975).

### 7.3.5 Spatiotemporal geostatistical analysis and prediction

Under the second-order stationarity hypothesis, the semivariogram and the covariance function are equivalent. For reasons of convenience the semivariogram structure is preferred. The appropriate semivariogram structure (separable or non-separable) is fitted to the experimental spatiotemporal model given by:

$$\hat{\gamma}_Z(\mathbf{r}_s, r_t) = \frac{1}{2N(\mathbf{r}_s, r_t)} \sum_{N(\mathbf{r}_s, r_t)} [Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j + \mathbf{r}_s, t_j + r_t)]^2, \quad (7.50)$$

where  $\mathbf{r}_s = \|\mathbf{s}_i - \mathbf{s}_j\|$ ,  $r_t = |t_i - t_j|$ , and  $N(\mathbf{r}_s, r_t)$  is the number of pairs in  $N(\mathbf{r}_s, r_t)$ . The space-time experimental semivariogram is estimated as half the mean squared difference between data separated by a given spatial and temporal lag  $(\mathbf{r}_s, r_t)$ .

Geostatistical prediction is then achieved using space-time Ordinary Kriging (STOK) (Christakos 1991b, Goovaerts 1997). The STOK estimator with respect to residual data notation is given below:

$$\hat{Z}'(\mathbf{s}_0, t_0) = \sum_{\{i: \mathbf{s}_i, t_i \in \mathbb{S}_0\}} \lambda_i Z'(\mathbf{s}_i, t_i), \quad (7.51)$$

$\hat{Z}'(\mathbf{s}_0, t_0)$  is the unsampled location-time,  $Z'(\mathbf{s}_i, t_i)$  are the sampled location-time neighbors and  $\lambda_i$  are the corresponding space-time kriging weights.

$$\sum_{\{i:\mathbf{s}_i, t_i \in \mathbb{S}_0\}} \lambda_i \gamma_{Z'}(\mathbf{s}_i, \mathbf{s}_j; t_i, t_j) + \mu = \gamma_{Z'}(\mathbf{s}_j, \mathbf{s}_0; t_j, t_0), \quad j=1, \dots, N_0, \quad (7.52)$$

$$\sum_{\{i:\mathbf{s}_i, t_i \in \mathbb{S}_0\}} \lambda_i = 1, \quad (7.53)$$

where  $N_0$  is the number of points within the search neighborhood of  $\mathbf{s}_0$ ,  $\gamma_{Z'}(\mathbf{s}_i, \mathbf{s}_j; t_i, t_j)$  is the semivariogram between two sampled points  $\mathbf{s}_i$  and  $\mathbf{s}_j$  at times  $t_i$  and  $t_j$ ,  $\gamma_{Z'}(\mathbf{s}_j, \mathbf{s}_0; t_j, t_0)$  the semivariogram between  $\mathbf{s}_j, t_j$  and the estimation point  $\mathbf{s}_0, t_0$  and  $\mu$  is the Lagrange multiplier enforcing the zero bias constraint.

The STOK estimation variance is given by the following equation, with the Lagrange coefficient  $\mu$  compensating for the uncertainty of the mean value:

$$\sigma_E^2(\mathbf{s}_0, t_0) = \sum_{\{i:\mathbf{s}_i, t_i \in \mathbb{S}_0\}} \lambda_i \gamma_{Z'}(\mathbf{s}_j, \mathbf{s}_0; t_j, t_0) + \mu. \quad (7.54)$$

The prediction is also described in matrix notation below where the system  $\Gamma \boldsymbol{\lambda} = \mathbf{c}$  is solved to estimate the spatiotemporal weights  $\boldsymbol{\lambda}$ :

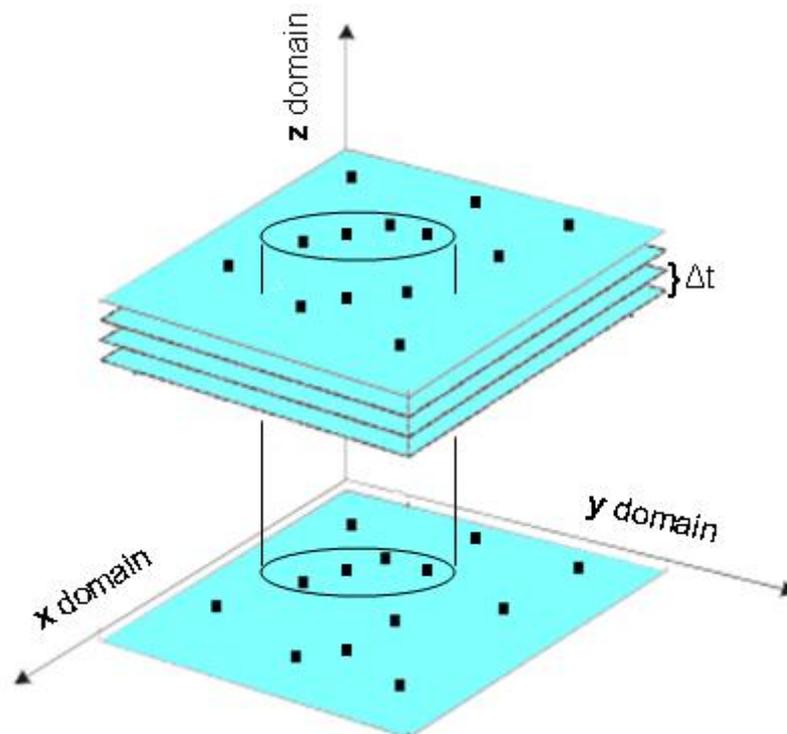
$$\Gamma = \begin{pmatrix} \hat{\gamma}_{Z'}(\mathbf{s}_1 - \mathbf{s}_1, t_1 - t_1) & \cdots & \hat{\gamma}_{Z'}(\mathbf{s}_1 - \mathbf{s}_{N_0}, t_1 - t_{N_0}) & 1 \\ \vdots & \ddots & \vdots & \\ \hat{\gamma}_{Z'}(\mathbf{s}_{N_0} - \mathbf{s}_1, t_{N_0} - t_1) & \cdots & \hat{\gamma}_{Z'}(\mathbf{s}_{N_0} - \mathbf{s}_{N_0}, t_{N_0} - t_{N_0}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}$$

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{N_0} \\ \mu \end{pmatrix}, \quad (7.55)$$

$$\mathbf{c} = \begin{pmatrix} \hat{\gamma}_{z'}(\mathbf{s}_0 - \mathbf{s}_1, t_0 - t_1) \\ \vdots \\ \hat{\gamma}_{z'}(\mathbf{s}_0 - \mathbf{s}_{N_0}, t_0 - t_{N_0}) \\ 1 \end{pmatrix}$$

where  $\Gamma$  is the matrix of the spatiotemporal semivariogram between the observed space-time data locations,  $\lambda$  are the spatiotemporal weights and  $\mathbf{c}$  the matrix of the spatiotemporal semivariogram between the observed space-time data locations and the space-time estimation location.

Space-time predictions are usually based on a space-time neighborhood which encloses observations inside a search radius in space and in time; the search radii depend on the space and time correlation lengths  $\xi_s, \xi_t$  estimated from the semivariogram fitting process. For small datasets the entire dataset is used for predictions. Figure 7.2 presents a schematic representation of the space-time domain and the space-time search neighborhood.



**Figure 7.2** Representation of the space-time domain and of the space-time search neighborhood (after Hengl 2007))

In STRK the estimate of the head, groundwater level, is expressed as:

$$\hat{Z}(\mathbf{s}_0, t_0) = m_z(\mathbf{s}_0, t_0) + \hat{Z}'(\mathbf{s}_0, t_0), \quad (7.56)$$

where  $m_z(\mathbf{s}_0, t_0)$  is the estimated trend function, and  $\hat{Z}'(\mathbf{s}_0, t_0)$  is the interpolated residual by means of STOK (Hengl 2007). The trend function herein is modeled based on equation (7.59).

### 7.3.6 Spatiotemporal prediction of groundwater level data in Mires basin

Mires basin of the Mesara valley is a sparsely monitored basin. Since 1981 where a rapid increase of drip irrigation and increased pumping were started, only 10 wells were consistently monitored biannually until the year 2003. Since then a telemetric network is operating in the area consisting of two monitoring wells while some selected measurements were taken in specific boreholes. The basin is consistently overexploited and the result is a great drawdown of the water table; more than 35m since 1981. The water resources availability in the area and especially the groundwater are encountering great shortage. Therefore the area is of spatially and temporally groundwater analyses need. Using spatiotemporal geostatistics the limited groundwater level dataset can be usefully exploited in order to identify the historic spatiotemporal behavior of the aquifer and to take useful information regarding the space-time data correlations for future predictions.

Space-time geostatistical analysis and predictions are made following the steps denoted below: 1) approximation of the spatio-temporal trend, 2) Space-time semivariogram calculation of the residuals, 3) application of space-time residual kriging (STRK) - STOK for prediction of fluctuations adding the predicted trend at the desired locations and 4) estimation of prediction accuracy. In the following, the above steps are addressed in detail.

As it was stated in section 7.3 (Eq.(7.23)) the random field can be decomposed in trend and fluctuations. Therefore, the initial geostatistical analysis step is to approximate the spatiotemporal trend of the field data. The spatiotemporal trend approximation involves the separate temporal and spatial trend removal. For the first the exponentially weighted moving average filter (Roberts 1959, Pham 2006, Nist/Sematech 2009) is used on the mean bi-annual groundwater level of the 10 available wells. This model is selected because it can provide discrete estimates of

future biannual groundwater level trends based on the previous period trend value and the groundwater level value of the desired estimation time. The equation that describes the model follows:

$$\hat{m}_z(t_i) = \alpha z(t_i) + (1 - \alpha) \hat{m}_z(t_{i-1}), \quad \hat{m}_z(t_1) = z(t_1). \quad (7.57)$$

where  $\hat{m}_z$  is temporal trend and  $z$  the groundwater level measurement and  $0 < \alpha \leq 1$  is the weight of the temporal model.

The spatial trend approximation is based on the closest distance of the wells from the river traversing the basin (see section 6.3.1). This approach using closest distance from the river root as auxiliary information is a new geostatistical tool that proved effective for the purely spatial analysis (section 6.4.4). The river bed can be seen as a curve, therefore we model the river by means of a *river curve*, herein represented by a second-order polynomial,  $y(x) = w_0 + w_1 x + w_2 x^2$ . As it is well-known and analytically presented in section 6.3.1, the closest distance of any given point from a curve is the perpendicular distance between the point and the tangent at a unique point of the curve. The shortest distance of the wells is first calculated and then the function (7.58) is applied to obtain the spatial trend. The spatial trend obtained for the time  $t^*$ , wet period of the hydrological year 2002-2003, when the most reliable measurements were taken is given by,

$$m_z(\mathbf{s}, t^*) = f d(\mathbf{s}) + g, \quad (7.58)$$

and is adopted as the reference year period spatial trend. In the above equation  $m_z$  is the trend,  $d$  is the well's distance from the river bed,  $f$  and  $g$  are coefficients of the linear spatial model.

We present a model for spatiotemporal trend  $m_z(\mathbf{s}, t)$  which is obtained by multiplying the spatial with the temporal trend and dividing by the reference temporal trend:

$$m_z(\mathbf{s}, t) = \frac{\hat{m}_z(t) m_z(\mathbf{s}, t^*)}{\hat{m}_z(t^*)}, \quad (7.59)$$

The first component of the numerator  $\hat{m}_z(t)$  is the temporal trend, which is approximated by applying an exponentially weighted moving average filter (7.57) as discussed above (Roberts 1959). The second term,  $m_z(\mathbf{s}, t^*)$ , is the spatial trend model. The denominator  $\hat{m}_z(t^*)$  denotes the temporal trend for the reference year period of the most reliable measurements (wet period of the hydrological year 2002-2003). This period is the basin's last regularly monitored period. The temporal trend values are divided with the temporal trend of the reference year period, obtained from the same function, to produce coefficients without units. This is necessary in order the spatiotemporal trend to retain units in meters.

Based on Eq. (7.59) it follows that,

$$m_z(\mathbf{s}, t^*) = m_z(\mathbf{s}, t^*), \quad (7.60)$$

confirming that Eq. (7.59) yields the spatial trend model at  $t^*$ . Another property of Eq. (7.59) is that the average spatiotemporal trend  $\overline{m_z(\mathbf{s}, t)}$  for a given time span  $t$  is equal to the average spatial trend of the reference year period  $\overline{m_z(\mathbf{s}, t^*)}$  multiplied by the ratio of temporal trend at time  $t$  over the reference temporal trend.

$$\overline{m_z(\mathbf{s}, t)} = \overline{m_z(\mathbf{s}, t^*)} \times \frac{\hat{m}_z(t)}{\hat{m}_z(t^*)}. \quad (7.61)$$

The spatiotemporal residuals are calculated simply by subtracting the groundwater level of each well at a specific time step from the corresponding spatiotemporal trend. The temporal trend remains constant for every time step while the spatial changes with the distance of each well from the river bed. Then, the space-time Residual Kriging (RK) method (7.56) is applied as it combines a trend function for the data, obtained by an appropriate model, with the interpolation of the residuals.

First, we need to determine the experimental semivariogram of the residuals. Then we model the experimental semivariogram with separable and non-separable theoretical spatiotemporal semivariogram functions. The simplest permissible space-time covariance function is the product model, where the separate spatial and temporal covariances models are simply multiplied while the semivariogram is obtained from the spatiotemporal covariance (7.37). The product space-time semivariogram model does not require the calculation of other parameters than only of the chosen spatial and temporal models (7.38). The product model which can be considered a special case of the product sum model is characteristic, as the whole class, for its flexibility in modeling and in estimation (Gething *et al.* 2007, De Iaco 2010). The Matérn semivariogram model is chosen to simulate the spatial and temporal continuity of the data with the separable product space-time model. The purely spatial geostatistical analysis of groundwater level data in preceding chapters shows that the Matérn describes well the spatial correlation of the observed data. In particular it delivers similar cross validation estimates as the Power-law and the Spartan semivariogram (Hristopulos 2002, Hristopulos and Elogne 2007). Preliminary tests showed that spatiotemporal prediction results with the Power-law model are inferior to those with the Matérn.

### 7.3.6.1 Separable product space-time semivariogram using Matérn model

The separable spatial and temporal Matérn semivariograms are presented below (Matérn 1960):

$$\gamma_z(\mathbf{r}_s) = \sigma_z^2 \left( 1 - \frac{2^{1-\nu_1} \left( \frac{|\mathbf{r}_s|}{\xi_1} \right)^{\nu_1}}{\Gamma(\nu_1)} K_{\nu_1} \left( \frac{|\mathbf{r}_s|}{\xi_1} \right) \right), \quad (7.62)$$

$$\gamma_z(r_t) = \sigma_z^2 \left( 1 - \frac{2^{1-\nu_2} \left( \frac{|r_t|}{\xi_2} \right)^{\nu_2}}{\Gamma(\nu_2)} K_{\nu_2} \left( \frac{|r_t|}{\xi_2} \right) \right), \quad (7.63)$$

$\sigma_z^2 > 0$  is the variance,  $\zeta$  is the range parameter,  $\nu > 0$  is the smoothness parameter,  $\Gamma$  is the gamma function,  $K$  is the modified Bessel function of the second kind of order  $\nu$ ,  $\mathbf{r}_s$  is the space lag vector and  $r_t$  is the time lag. These are inserted in the product space-time semivariogram, (7.38).

### 7.3.6.2 Non-separable space-time semivariogram

The non-separable covariance model obtained from the diffusion equation is also used herein because a) it is a covariance structure that has not been used before in geo-hydrological data and b) it is similar in concept to Gneiting's model that has successfully been applied to wind data in the past. The space-time covariance structure presented in equation (7.49) is used. According to equation (7.29), the resulting semivariogram is:

$$\gamma_z(\mathbf{r}_s, r_t) = \sigma_z^2 \left[ 1 - \left( \beta r_t^{2\gamma} + 1 \right)^{-n/2} \exp\left( -\mathbf{r}_s^2 / \left( \beta r_t^{2\gamma} + 1 \right) \right) \right], \quad (7.64)$$

where  $0 \leq \beta \leq 1$ ,  $0 < \gamma \leq 1$  and  $n$  the number of spatial dimensions.

## 7.4 Results

### 7.4.1 Purely temporal analysis

Temporal groundwater level fluctuations modeling and analysis for the Mires basin is aiming on the representation of the aquifer behavior locally or as a unit through time in conjunction with physical quantities (precipitation, pumping) that affects the aquifer's water table level. Useful information therefore can be extracted regarding correlations of mean annual groundwater level, annual precipitation and annual pumping for future predictions.

The input data for the Mires basin include the time series of mean annual groundwater level, the annual abstraction and the annual precipitation surplus. The mean annual groundwater level is obtained from a spatial average over 10 boreholes in the Mires basin from 1981 to 2003. At each location, usually bi-annual and often bi-monthly data of groundwater level are averaged over each hydrological year (from

October to October). Since 2003 the mean annual groundwater level is derived from two telemetric stations that operate in the basin, while for the period 2003-2006<sup>6</sup> the average also includes biannual measurements of four of the ten set boreholes (leading to an average over six locations) due to selective measurements for this time period. The data used for the model calibration span the interval from the hydrological year 1980-1981 to 2005-2006. The annual average of the groundwater level is used, because only annual reliable values of the precipitation were available at this point.

The ARX model was used by (Bierkens *et al.* 2001) with a time step equal to one day. Here we use a considerably larger time step (one year). Hence we use a coarse-grained version of the original model based on the linearity of the equation. As we show below, the model predictions are in good agreement with the data after the initial period of parameter adaptation.

The groundwater level is predicted consecutively for the hydrological years 2006-2007, 2007-2008, 2008-2009 and 2009-2010. The model is ran separately to predict the groundwater level for each year period as the number of time steps is small (25) for optimal parameters prediction that represent the process accurately. Therefore each time the model is ran; the optimal parameters of the system at the last time step are calculated. Based on these parameters the next period prediction regarding precipitation surplus and pumping estimate is provided. For every validated period the groundwater level is added consecutively in the model in order to test the next period's prediction accuracy. Therefore the system for each tested time period is updating the optimal parameters.

We have added as opposed to the initial model that the precipitation surplus is only used as an exogenous variable (Bierkens *et al.* 2001), a term proportional to the annual abstraction rate (7.7). The model parameters determined using the Kalman filter without and with pumping terms are presented in Table 7.1 and Table 7.2. The average groundwater levels and the annual precipitation and abstraction rates for the above time periods are known. Therefore, the groundwater levels predicted by the model can be validated against the real values as shown in Table 7.3 and Table 7.4. As shown in Figure 7.3 and Figure 7.4 the agreement between predictions and measurements improves as more data are processed. The parameters of the last validation period are considered optimal for the process based on the available data

---

<sup>6</sup> wet period (October to April) only for the hydrological year 2005-2006)

and can be used for future predictions. In order to test their efficiency they are inserted as initial parameters in the ARX model. From Figure 7.5 it can be observed that the model's adaption is faster and better.

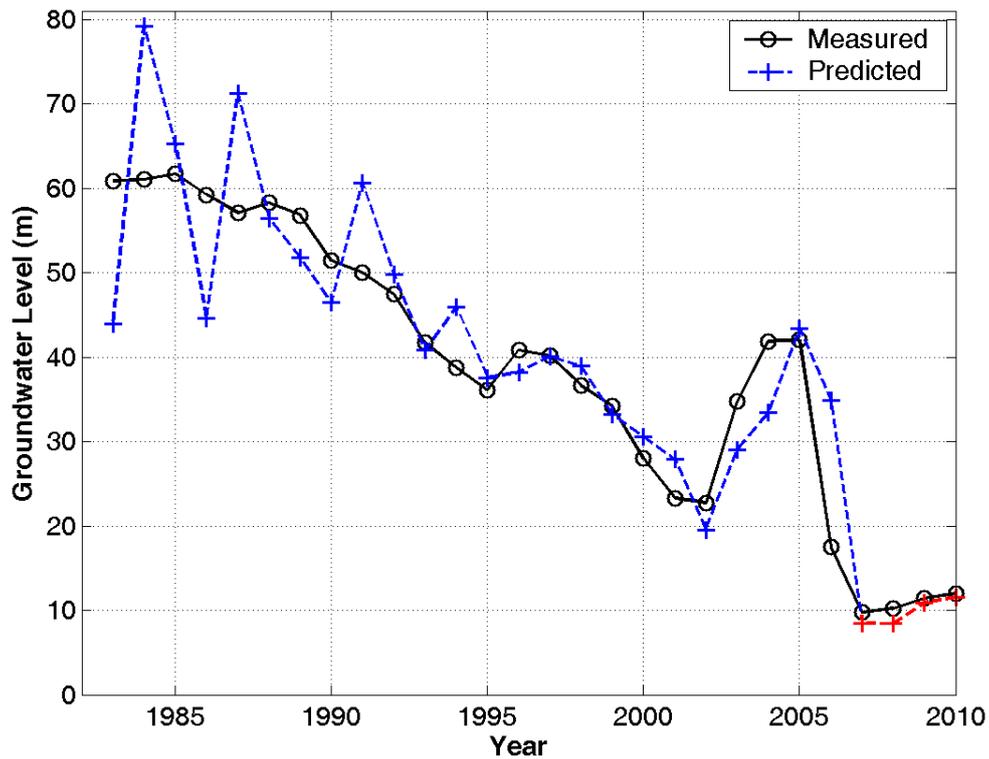
The actual evapotranspiration is a key variable for the model, because it enters in the calculation of the precipitation surplus. According to the regional department of water recourses management (Department of Water Resources Management 2000), the mean annual actual evapotranspiration  $\overline{E[P^*(t_k)]}$  on the island of Crete is 70% of the mean annual precipitation, while in low elevation areas, less than 300 meters above sea level, it is 75%. The mean annual actual evapotranspiration of the Mesara valley is estimated at 65% of the mean annual precipitation (Croke *et al.* 2000). However the Mires basin is only a part of the Mesara catchment, and its elevation is less than 300 meters above sea level. Therefore we consider three different values of mean annual actual evapotranspiration in the model, for which we compare the predicted groundwater levels. The 75% evapotranspiration level leads to the most accurate prediction.

**Table 7.1** ARX parameters, determined from Kalman filter approach, with Precipitation Surplus data input for each consecutive run.

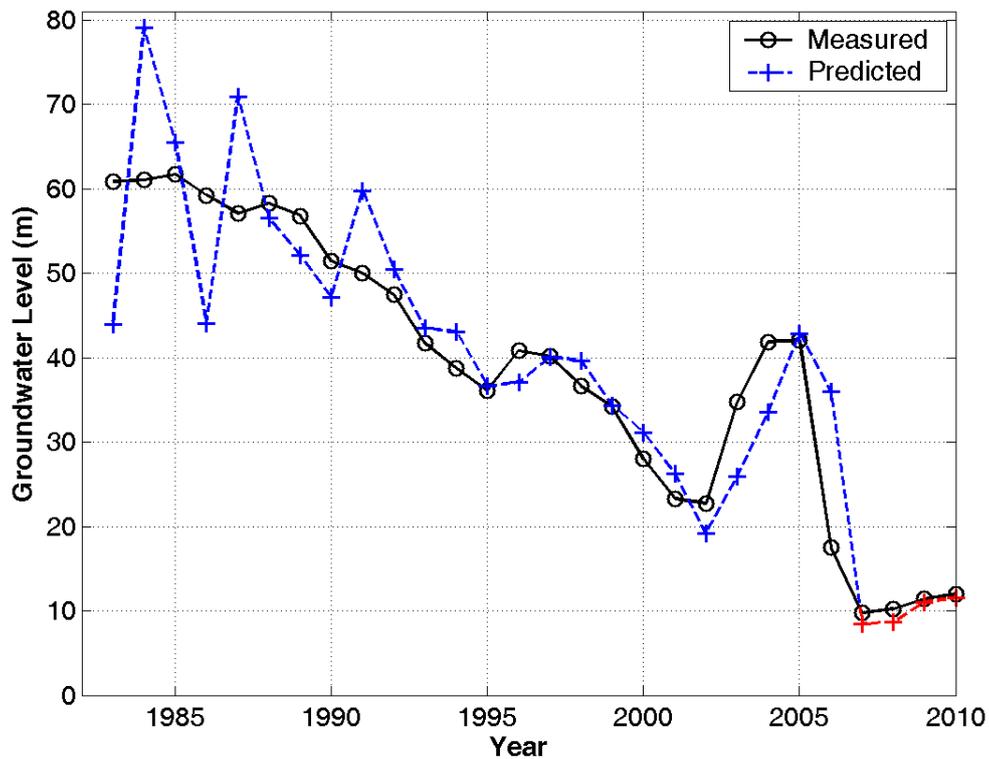
Year	$\alpha$	$b$	$c$
<b>2006-2007</b>	0.3846	0.0181	0.0049
<b>2007-2008</b>	0.3735	0.0328	0.0050
<b>2008-2009</b>	0.3704	0.0458	0.0049
<b>2009-2010</b>	0.3704	0.0490	0.0049

**Table 7.2** ARX parameters, determined from Kalman filter approach, with Precipitation Surplus & Pumping data inputs for each consecutive run.

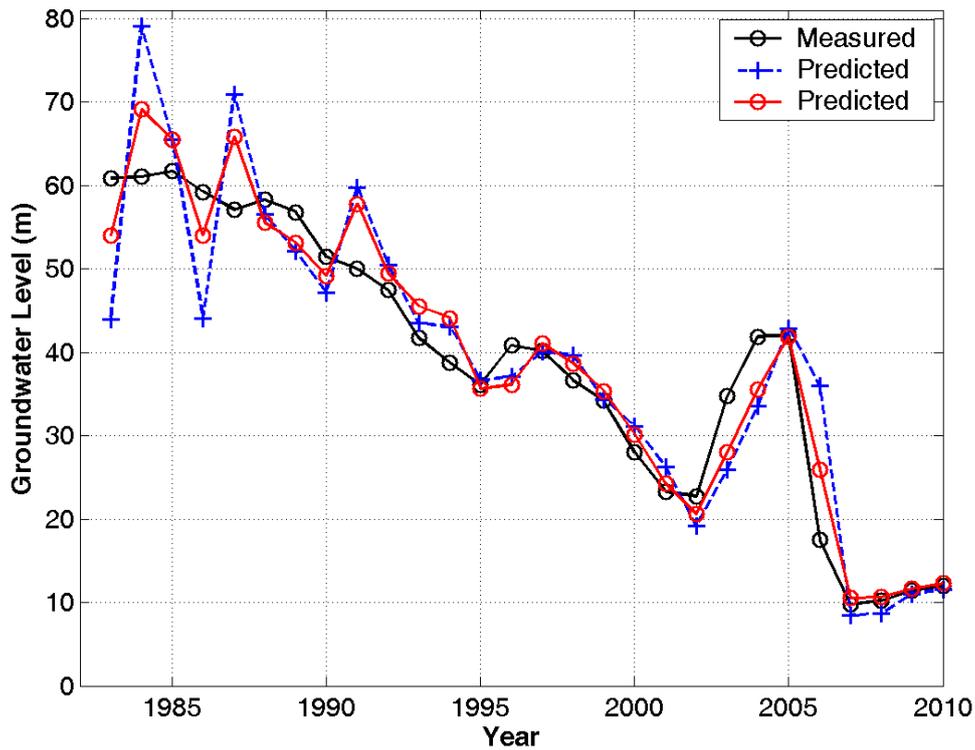
Year	$\alpha$	$b$	$d$	$c$
<b>2006-2007</b>	0.3900	0.0216	-0.0182	0.0049
<b>2007-2008</b>	0.3766	0.0360	-0.0153	0.0050
<b>2008-2009</b>	0.3754	0.0479	-0.0197	0.0049
<b>2009-2010</b>	0.3756	0.0503	-0.0205	0.0049



**Figure 7.3** ARX model results using precipitation surplus as exogenous variable. Red crosses denote the predicted values.



**Figure 7.4** ARX model results using precipitation surplus and pumping rate as exogenous variables. Red crosses denote the predicted values.



**Figure 7.5** ARX model results using precipitation surplus and pumping rate as exogenous variables. Red circles denote the calibration results using as initial parameters the optimal calculated from the last step during the validation process, Figure 7.4 ( $a=0.3756$ ,  $b=0.0503$ ,  $d=0.0205$ ,  $c=0.0049$ ).

**Table 7.3** Temporal Validation Results-Absolute Error (AE), Absolute Relative Error (ARE) - ARX with Precipitation Surplus for each consecutive run.

Year	2006-2007	2007-2008	2008-2009	2009-2010
AE (m)	1.25	1.75	0.51	0.40
ARE	0.125	0.16	0.04	0.02

**Table 7.4** Temporal Validation Results-Absolute Error (AE), Absolute Relative Error (ARE) - ARX with Precipitation Surplus & Pumping for each consecutive run.

Year	2006-2007	2007-2008	2008-2009	2009-2010
AE (m)	1.25	1.62	0.37	0.40
ARE	0.125	0.157	0.03	0.02

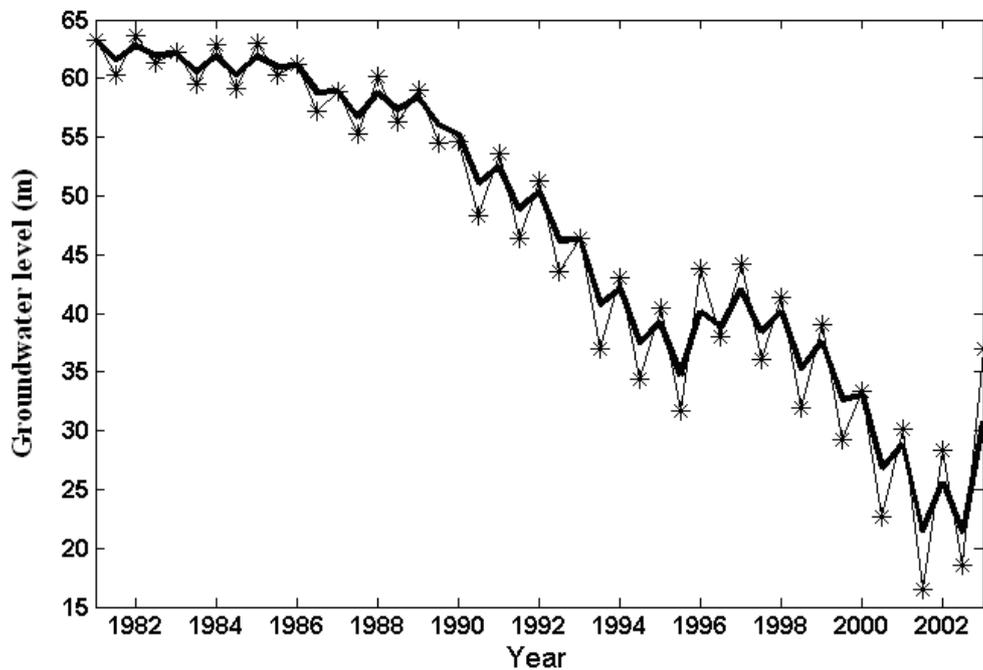
7.4.2 Spatiotemporal analysis

Spatiotemporal geostatistical analysis of Mires basin groundwater level data is applied in order to identify the spatiotemporal behavior of the aquifer since 1981 and to undertake future predictions based on the space-time data correlations.

A spatiotemporal trend model is developed to approximate the S/TRF spatiotemporal trend. Fitting of a spatiotemporal trend provides “trend-free” data (residuals) which yield more stable semivariograms as the semivariogram is reaching

easier a sill and the intrinsic hypothesis is satisfied (stationary mean value and semivariance) (Journel and Huijbregts 1978).

The spatiotemporal trend approximation (7.59) involves the separate estimation of temporal and spatial trend component. For the first the exponentially weighted moving average filter (7.57) is used on the mean bi-annual groundwater level from the 10 available wells. This trend model can provide bi-annual estimates of future groundwater level trends based on the previous period trend value and the groundwater level of the desired period. The trend fit is presented in Figure 7.6 and the optimal weight  $0 < \alpha \leq 1$  of the temporal model is calculated equal to 0.6. This value was determined from the temporal trend fitting process and it is the one that under STRK prediction provided the most accurate results. Then the trend values are divided with the temporal trend of the reference year, obtained from the same function, to produce coefficients without units. This is necessary in order the spatiotemporal trend to retain units in meters. The spatial groundwater level trend approximation involves the closest distance of the wells from the river traversing the basin.

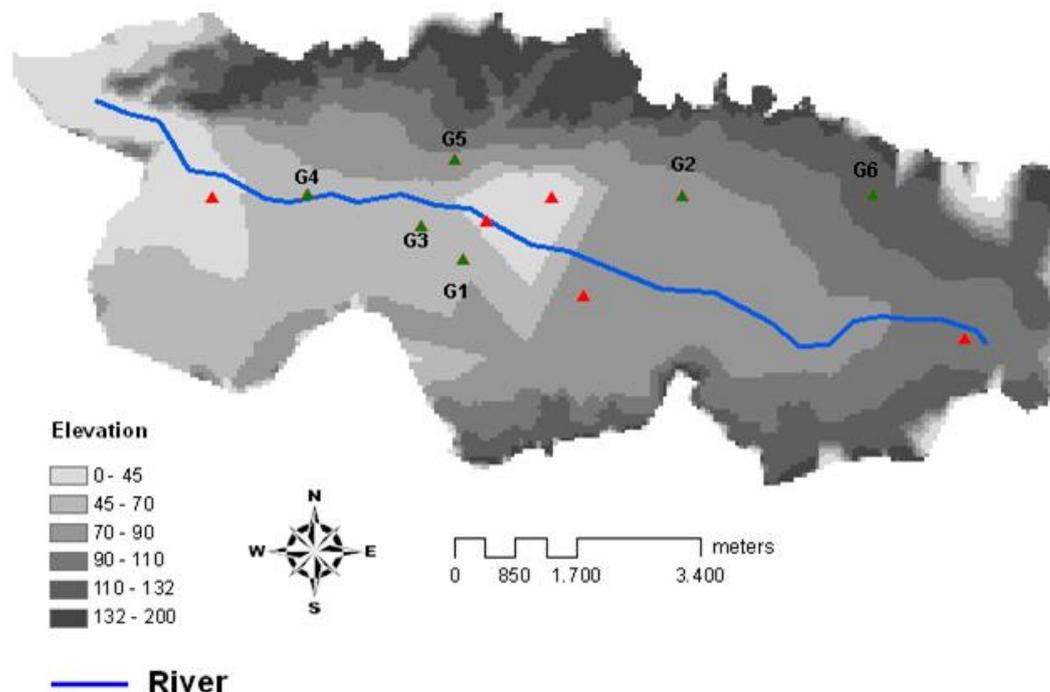


**Figure 7.6** Exponentially weighted moving average filter fit on mean bi-annual groundwater level measurements.

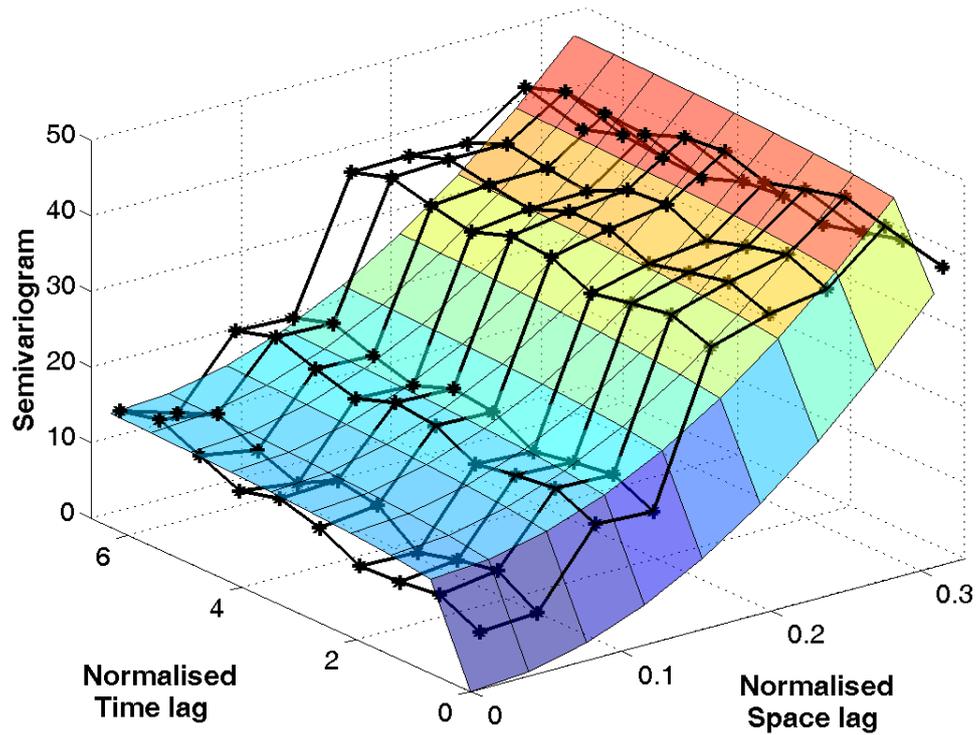
The space-time experimental semivariogram is determined from the biannual (wet and dry period) groundwater level residuals (after trend removal) time series

between the years 1981 and 2003 at the ten sampling stations. Validation of the STRK estimates is performed for six wells, Figure 7.7, where biannual observations (wet and dry period) are available for the period 2003-2004 to 2005-2006 (wet period). The period after the hydrological year 2002-2003 is characterized by significant groundwater level increase in the wet period of 2003-2004 and by considerable groundwater level drop in the dry period of 2004-2005. Therefore we decided to assess the reliability of the space-time model (STRK) and estimated space-time semivariograms at these periods.

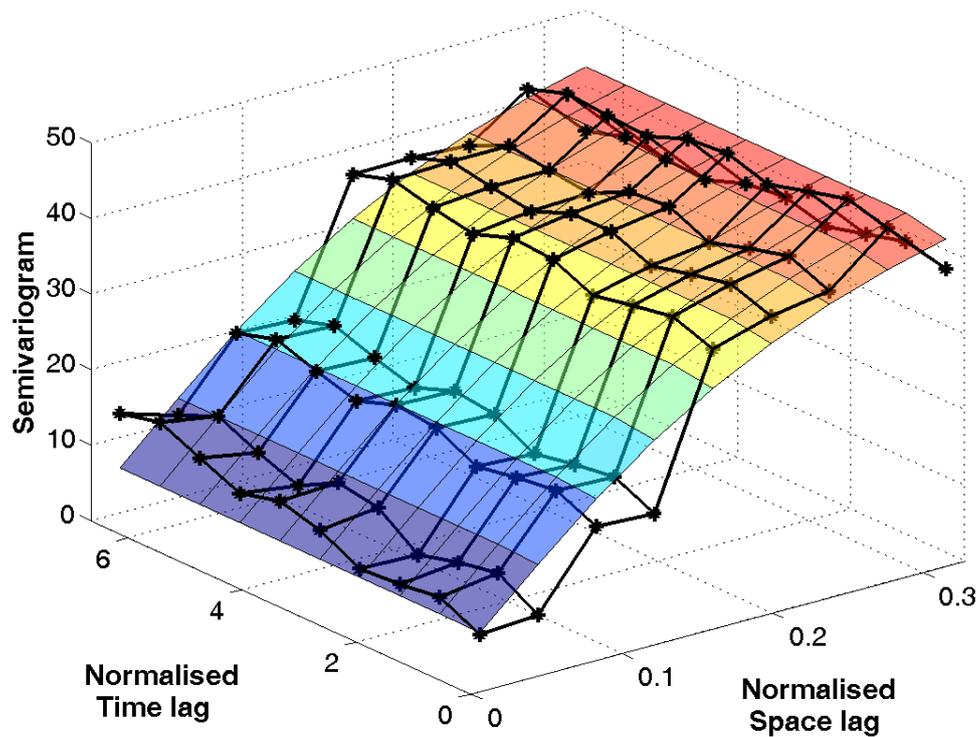
The theoretical space-time semivariogram model fitting on the experimental space-time semivariogram obtained from the observed data residuals is presented in Figure 7.8 and Figure 7.9. The respective parameters for the two semivariogram types are  $\sigma_z^2 = 397$ ,  $\beta = 0.0334$  and  $\gamma = 0.0452$  for the non-separable space-time semivariogram type and  $\sigma_z^2 = 46.57$ ,  $\xi_1 = 0.25$  ( $\approx 3\text{km}$ ),  $\nu_1 = 0.7103$ ,  $\xi_2 = 0.1570$  ( $\approx 2$  months),  $\nu_2 = 1.5138$  and nugget variance  $c = 1.8336$  for the separable type (product model) using the Matérn function. For the latter, the nugget terms is inserted in order to better fit to the experimental semivariogram.



**Figure 7.7** Topographic map showing the locations of the 10 monitored wells (triangles) in Mires basin along with the corresponding surface elevation and the temporary river path. With green color the wells monitored for the period 2003-2006 are presented.



**Figure 7.8** Space time non-separable semivariogram fit. The upper space limit in real units is equal to 4Km and the time limit 6.5 years.



**Figure 7.9** Space-time product semivariogram fit using the Matérn structure. The upper space limit in real units is equal to 4Km and the time limit 6.5 years.

The prediction (extrapolation) involves STRK application using the appropriate semivariogram to estimate the residuals at the specified location and time. The corresponding trend is calculated from equation (7.59) where the spatial trend in a new location involves the calculation of its closest distance and the application of function (7.58). The temporal trend is determined using the adequate period's mean groundwater level applying function (7.57) and dividing over the temporal trend of the reference year. The product of these functions delivers the spatiotemporal trend at the desired location-time, which is then added to the predicted residual to provide the groundwater level estimate.

The validation results for the absolute estimation error (AE)<sup>7</sup> are presented in Table 7.5 and Table 7.6. The first table presents the groundwater levels' estimation error for the wet period of hydrological year 2003-2004 using separable and non-separable semivariogram while the second for the dry period of hydrological year 2004-2005. As it can be seen the non-separable semivariogram delivers more accurate estimates compared to the separable.

**Table 7.5** Absolute Error (AE) of STRK estimates for the wet period of highest groundwater level increase after 2003 (i.e. the wet period 2003-2004).

a) using the non-separable semivariogram model (7.64)

Well No	AE (m)
<b>G1</b>	3.56
<b>G2</b>	4.02
<b>G3</b>	2.72
<b>G4</b>	2.21
<b>G5</b>	2.32
<b>G6</b>	4.32

b) using the separable semivariogram model type (product model) with the Matérn function

Well No	AE (m)
<b>G1</b>	5.95
<b>G2</b>	5.80
<b>G3</b>	6.14
<b>G4</b>	4.87
<b>G5</b>	3.00
<b>G6</b>	5.27

<sup>7</sup> AE error is only used as validation metric because we intent to present the physical quantity of the error regarding predictions.

**Table 7.6** Absolute Error (AE) of STRK estimations for the dry period of highest groundwater level drop (i.e. the dry period 2004-2005).

a) using the non-separable semivariogram model (7.64)

Well No	AE (m)
<b>G1</b>	3.69
<b>G2</b>	6.26
<b>G3</b>	3.57
<b>G4</b>	2.88
<b>G5</b>	3.11
<b>G6</b>	4.90

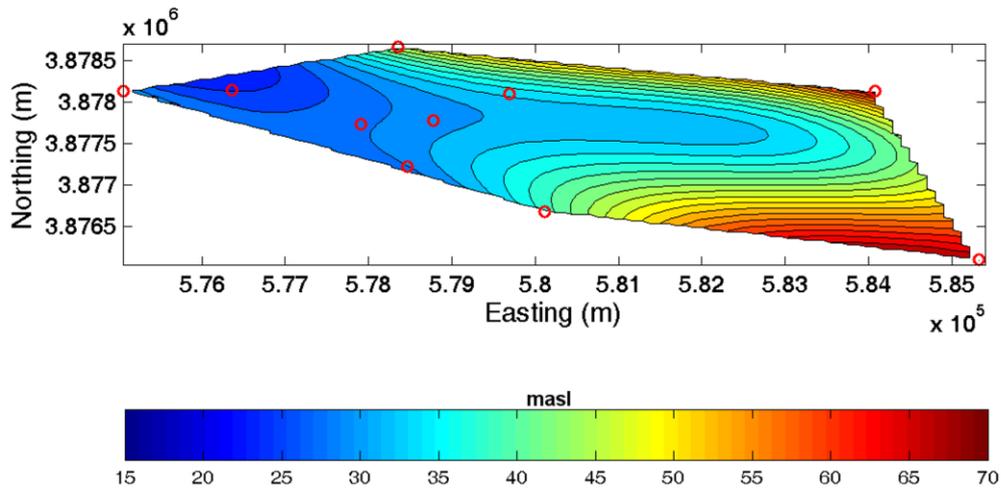
c) using the separable semivariogram model type (product model) with the Matérn function

Well No	AE (m)
<b>G1</b>	6.35
<b>G2</b>	7.87
<b>G3</b>	7.07
<b>G4</b>	2.59
<b>G5</b>	3.76
<b>G6</b>	5.83

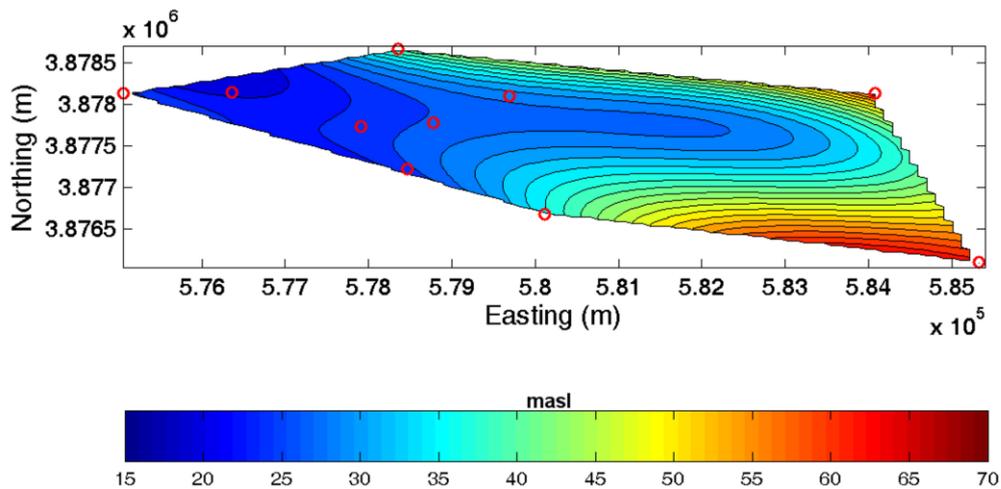
After the hydrological year 2005-2006, the average of the biannual estimates (i.e. the estimated mean annual groundwater level) is compared with the groundwater level in the basin reported by the local authorities (based on the average of the two remotely sensed holes) and with the ARX predictions. The temporal model, as shown above gives very good agreement with the reported values (period 2007-2010). The AE of the estimates obtained with the optimal<sup>8</sup> non-separable space-time semivariogram is presented in Table 7.7.

Interpolation maps are derived using STRK with the optimal non-separable spatiotemporal semivariogram structure for the wet period of 2003-2004 and the dry period of 2004-2005. For these periods, the accuracy of separable and non-separable semivariograms is tested (Table 7.5 and Table 7.6). The contour maps of groundwater level spatial variability in physical space are shown in Figure 7.10 and Figure 7.11. The maps are constructed using estimates only at points inside the convex hull of the measurement locations.

<sup>8</sup> The semivariogram model that delivers the most accurate estimations



**Figure 7.10** Map of estimated groundwater level in the Mires basin using STRK (7.56), the non-separable space-time semivariogram (7.64) on the residuals and spatiotemporal trend removal (7.59) (wet period of 2003-2004 hydrological year).



**Figure 7.11** Map of estimated groundwater level in the Mires basin using STRK (7.56), the non-separable space-time semivariogram (7.64) on the residuals and spatiotemporal trend removal (7.59) (dry period of 2004-2005 hydrological year).

**Table 7.7** Absolute Error (AE) of STRK estimation error of prediction average using the non-separable semivariogram model (7.64)

Year	AE (m)
<b>2005-2006</b>	7.35
<b>2006-2007</b>	4.55
<b>2007-2008</b>	6.75
<b>2008-2009</b>	8.60
<b>2009-2010</b>	9.00

## 7.5 Discussion

The scope of this thesis's chapter is to model spatiotemporally the Mires basin aquifer response since 1981. Reliable modeling provides the ground for spatiotemporal future groundwater level fluctuations prediction with the highest possible accuracy.

The ARX model (7.7) used for the purely temporal mean annual groundwater level modeling delivers satisfactory predictions with low validation errors (Table 7.3, Table 7.4). We include in the ARX model training period the recent extreme groundwater level fluctuations in order to train the model more efficiently. As observed in Figure 7.3 and Figure 7.4 the accuracy of the estimates improves as more data are processed by the method while the recent extremes are captured reasonably by the model. The model is trained for the period 1981-2006 while the period 2007-2010 is used as the validating period. The ARX model is embedded in a discrete-time Kalman filter, trained gradually until the previous year of prediction, calculating reliable model parameters and producing the reliable ARX estimations presented.

According to equation (7.7) the mean annual groundwater level value  $\overline{Z(t_k)}$  estimation at time  $t_k$  depends on the previous year measured groundwater level value  $\overline{Z(t_{k-1})}$  reliable estimations of rainfall and/or abstraction volume for the specified prediction year period  $t_k$  and the estimated parameters that describe the aquifer dynamic response and are calculated during the calibration process. For each value added in the data set the optimal parameters are gradually updated. The validation results show that the predicted value of each year forward from 2006 is very well

estimated and it can be substituted to the dataset as  $\overline{Z(t_{k-1})}$  in order to predict the parameters and the next year's level.

ARX estimates are reliable only for one year forward, as long term or even short term predictions of rainfall intensity involve high uncertainty. In addition the abstraction volume rate is also uncertain, because it depends on anthropogenic activities that cannot be predicted accurately for longer than a year. The tables of results (Table 7.3, Table 7.4) and the figures (Figure 7.3, Figure 7.4) show that the incorporation of the abstraction rate in equation (7.7) improves the estimates; however, the improvement is not significant meaning that the driving variable for accurate estimation is the precipitation surplus. Nevertheless, so Figure 7.3 and Figure 7.4 as mean square prediction error (7.1) show that after initial fluctuations the adaptation of the ARX model that includes in addition to precipitation surplus the abstraction rate is better,  $\sigma_{\text{err}} \approx 7.83$  m, than the one including the precipitation surplus only  $\sigma_{\text{err}} \approx 8.06$  m. Thus, the abstraction rate is significant for the modeling of the groundwater level temporal variability.

A novel goal of this study is to assess the use of a recently proposed covariance function extracted from a physical differential equation such as the diffusion equation in real data; the aim of this test is more efficient spatiotemporal interpolation results. The non-separable space-time covariance structure tested herein delivers better estimates than the classic space-time product covariance function. The space-time semivariogram is calculated for the years 1981 to 2003, from the calculated residuals of the biannual groundwater level data of 10 wells. STRK estimates with non-separable semivariogram model are significantly more accurate for the year of highest groundwater level increase (74%) and for the year of the highest drop (44%). Similarly, the mean annual groundwater level (average of the biannual estimations) for 2006 to 2010 is more accurate using the non-separable space-time semivariogram in the STRK model.

For the hydrological year 2005-2006 only wet period measurements are available. However, because of the unprecedented average level decline in this hydrological year a high estimation error is obtained for this period and for the hydrological year average. A common observation for the prediction averages after 2005-2006 is that beyond the temporal limit (wet period of 2003) of the space-time semivariogram the estimation accuracy decreases consecutively (Table 7.7). This

means that the semivariogram is reliable only for short-time periodical estimations. The spatial correlation length is determined after the semivariogram fitting equal to almost 3km and the temporal length equal to almost 2 months. As shown herein the predictions of the wet period 2003-2004 which is the closer prediction period so to that temporal correlation length as to semivariogram temporal limit were the most accurate compared to the other periods tested. To predict future values reliably more temporal and spatial data are required. However the selected spatiotemporal trend simulation model in conjunction with the non-separable semivariogram function and STRK estimator provide reliable predictions, even for the two periods with the most intense fluctuations, wet period of 2003-2004 and dry period of 2004-2005 respectively. The potential expansion of the data set will increase the model's reliability and credibility. Tests with more space-time semivariogram structures can also provide a more general comparison of non-separable models and separable ones.

The ARX model can be used in combination with a spatiotemporal geostatistical model to provide short-term (bi-annual) future predictions of the groundwater level spatial variability. A seasonal extension of the ARX annual average model can yield reliable estimates of the seasonal groundwater level that can be used in the exponentially weighted average filter to calculate the temporal trend value at a specific future time. Then, after applying equation (7.59), the spatiotemporal trend is determined for the desired location and time. Residuals are estimated for the same location and time using STOK and then added to the trend estimation. Reliable ARX predictions involve mainly reliable future rainfall estimations from a climate model or from a statistical approach. Improvement of both ARX and STRK models predictions should involve more spatiotemporal data on shorter time scales and knowledge of the pumping activity of the unauthorized wells operating in the basin.

## 8. Conclusions and future work

### 8.1 Conclusions

This thesis presents an integrated approach of interpolating groundwater level spatial and/or temporal variability in sparsely monitored basins. The performed research intergrades initial data exploratory analysis, anisotropy estimation methodologies, common and new data normalization transformation techniques, spatial data interpolation using well known and newly established geostatistical methods based on kriging essence, purely temporal groundwater level fluctuations modeling and spatiotemporal geostatistical analysis and interpolation of groundwater level data.

Innovative geostatistical tools and methodologies developed in this thesis improve the accuracy of spatiotemporal interpolation of groundwater level data. The accurate representation of groundwater spatiotemporal variability in a basin is important for management purposes and for sound groundwater modeling. The analytical methodologies and tools introduced in this thesis contribute to applied geostatistical research. We apply these methods and tools to real data. In addition our analyses support the statement that there is not a globally best interpolation method. The performance, efficiency and suitability of the methods depend on the statistical properties of the dataset (e.g. skewness, kurtosis, trends, outliers) and on the specific validation measures discussed in chapter 2.6.

This thesis was initially motivated by the dramatic decrease of groundwater levels in Mesara valley in recent years due to overexploitation. In light of this development and the expected adverse effects of climate change on the basin's water resources, accurate spatial modeling of the groundwater level variation is needed for two reasons: a) to identify "vulnerable" locations where an integrated groundwater resources management plan should focus and b) to provide accurate information for the calibration of numerical groundwater flow models for the basin, e.g. for the representation of initial groundwater level conditions. Modflow code interfaces (Visual, GMS, Groundwater Vistas) for groundwater level modeling integrate in the process geostatistical approaches for data interpolation but without providing a sense of quality of the geostatistical methods interpolation accuracy. For example transient

flow modeling requires accurate initial groundwater level conditions for efficient modeling results.

This thesis presents a comparison of stochastic (Ordinary Kriging-OK, Universal Kriging-UK, Delaunay Kriging-DK) and deterministic (Inverse Distance Weight-IDW, Minimum Curvature-MC) interpolation methods for groundwater level monitoring in sparsely gauged areas. For the hydraulic head data from Mires basin (Crete, Greece), we established that the OK and UK interpolation methods overall perform better with respect to various cross validation measures, while DK and IDW show similar performance. However, no method is significantly superior to the others as the estimation error metrics are similar Table 4.3. The isopleth contours generated by DK and especially by IDW are rough contrary to smooth representations from OK, UK and MC. The stochastic methods provide guidance for the location of additional monitoring sites, based on the values of the kriging variance. Since the size of the Mires basin dataset is relatively small, computational limitations are irrelevant. For large datasets, computational time and memory usage for each method should also be investigated.

The advantages of the stochastic interpolation methods performance corresponds to the assessment of the specific dataset. However as it has also been shown in previous works, the stochastic methods perform generally better than the deterministic. Rough contours are a characteristic of the IDW method as also are the smooth contours delivered by the MC method. DK is expected to have rough contours because it is based on a small number of neighbors. OK and UK generally convey smooth contours. This happens because the estimates are based on correlated observations within a neighborhood.

The three-parameter Spartan semivariogram model is herein applied for the first time to hydrological data and yields the optimal cross validation performance among the investigated models. In addition, it delivers the estimates with the lowest standard deviation. The Spartan model is non-differentiable. We interpret this property as the result of a deposition-removal process that leads to an fBm-like behavior of the groundwater level surface. We also show that DK provides the best cross-validation estimate for the extreme low value, due to the localized nature of DK interpolation.

Subsequently this thesis presents non-linear data normalizing methods for the improvement of kriging groundwater levels. The application field is the Mires basin

on the island of Crete (Greece). Tgk<sup>9</sup>, GA<sup>10</sup>-OK, Box-Cox-OK and MBC<sup>11</sup>-OK, using the Spartan semivariogram model improve the mean absolute estimation error (MAE) compared to OK. For the first three methodologies the other estimation measures considered (RMSE, bias, MARE, R) are similar to those of OK, except for their inferior bias error. However, the MBC-OK also improves the RMSE, delivers the same low bias error and identical MARE and R. Overall, they deliver the most accurate estimation measures compared to the other methodologies tested and overall better than those of OK. The MBC method is applicable to both positive and negative values in contrast to the Box-Cox method that can be applied only to positive values. Normalization methods in general show that they can improve the effectiveness of the kriging interpolation method by reducing the estimation error compared to OK, thus leading to more accurate predictions. In this study, the normalization method (MBC) and the recently proposed spatial semivariogram structure (Spartan) are applied with OK for groundwater level interpolation. They obtain overall the most accurate cross-validation results while; cross-validation estimates and interpolation estimates satisfy a wide range of statistical criteria. The correlation coefficient of the cross-validation estimates vs. the true values is equal to 0.91, the distribution of errors is symmetric with a low bias equal to 0.02 masl, the plot of cross-validation errors vs. estimates is centered about zero error, satisfying the “conditional unbiasedness” property, measurements histogram reproduction from cross validation estimates and experimental semivariogram reproduction from the interpolation estimates.

As this thesis focuses on the spatial analysis of groundwater level in sparsely monitored basins, two novel spatial trend models are proposed for groundwater level interpolation. In addition the Covariance Hessian Identity method for anisotropy estimation is for the first time applied to hydrological data. We establish that besides the interpolation methodologies tested previously two optimal approaches based on Residual Kriging-RK significantly improves groundwater level interpolation. The first method uses a novel trend model that incorporates, in addition to smoothed surface elevation, the shortest distance of the monitoring locations from the temporary river traversing the basin. The second trend model uses an analytical equation specifying

---

<sup>9</sup> Trans-Gaussian Kriging

<sup>10</sup> Gaussian Anamorphosis

<sup>11</sup> Modified Box-Cox

the head for a system of multiple pumping wells. The combination of the proposed trend models and RK implemented with the non-differentiable Spartan semivariogram and the new MBC data normalization method leads to optimal cross validation results. The optimal spatial models significantly improve the cross validation measures compared to other models tested. In particular, RK interpolation with the trend models T-RD-DEM-UGA and T-MW reduces the prediction error of the lowest value (9.4 masl) by 20% and 29%, respectively, compared to the standard OK prediction. At the same time, the highest level of 62 masl is accurately estimated. The Spartan semivariogram provides the most accurate results from all the spatial models investigated in this thesis, while the stochastic methods overall perform better than the deterministic ones. An outline of the methods performance is presented in Table 8.1.

**Table 8.1** Cross validation measures for the stochastic and deterministic interpolation methods investigated. Results obtained with the “optimal” (in terms of cross validation measures) semivariogram model are presented. Optimal values are emphasized.

Method	Semi- variogram	MAE (masl)	BIAS (masl)	MARE	RMSE (masl)	R
MC		4.01	0.10	0.17	6.18	0.87
IDW		3.45	-0.17	0.15	5.58	0.89
DK-SP	SP	3.48	0.10	0.15	5.47	0.89
UK-SP	SP	3.40	0.13	0.14	5.23	0.91
OK-SP	SP	3.37	0.02	0.14	5.15	0.91
Box-Cox-OK	SP	3.30	0.10	0.14	5.14	0.91
MBC-OK	SP	3.30	0.02	0.14	5.12	0.91
GA-OK	SP	3.30	-0.3	0.14	5.14	0.90
TGK	SP	3.28	-0.1	0.14	5.14	0.91
T-DEM MBC & RK	SP	3.32	0.07	0.15	5.20	0.90
T-DEM-UGA MBC & RK	SP	3.21	0.03	0.14	5.08	0.90
T-RD MBC & RK	SP	3.11	0.08	0.12	4.86	0.92
T-DEM-UGA-RD MBC & RK	SP	3.02	0.07	0.12	4.79	0.92
T-MW MBC & RK	SP	<b>2.75</b>	<b>0.07</b>	<b>0.11</b>	<b>4.57</b>	<b>0.93</b>

$s_w = 1.85$  (masl)

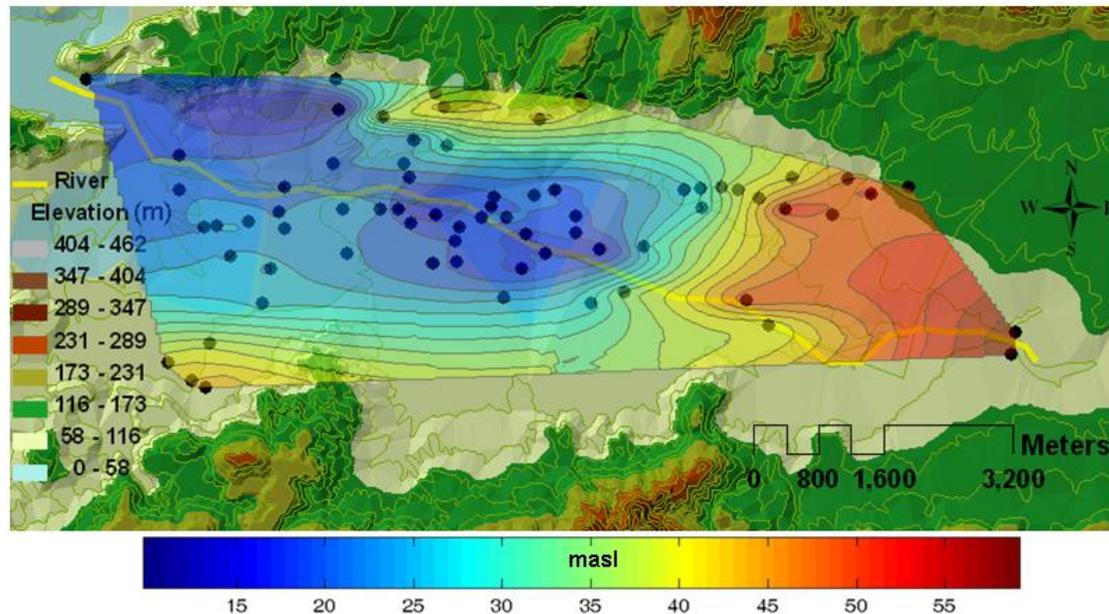
Low Bias and the highest possible statistical accuracy (MARE, RMSE) are the objectives in comparing different methods performance apart from the MAE.

Comparing the deterministic with the stochastic methods the results can be characterized statistically different. The most efficient deterministic method (IDW), delivers clearly inferior statistical measures compared to the most efficient stochastic method (RK-T-MW). Comparing the stochastic methods efficiency from the least accurate to the most accurate consecutively, the results are not clearly statistically different for this sample. Most of the spatial models deliver overall close results. However, an important difference exists between the most accurate stochastic interpolation method (OK) and the most accurate spatial model (RK-T-MW) for this dataset. An accuracy improvement of around 60 cm in MAE, 3% in MARE, around 60 cm in RMSE, similar bias and slightly improved correlation coefficient can be characterized statistically different and significant for groundwater modeling approaches or water resources management plans.

The optimal approach of all the spatial methods tested is based on RK, includes a trend component based on the generalized *Thiem's* equation for multiple wells, and employ as well as two newly established geostatistical tools: a) the flexible Spartan semivariogram family and b) the MBC data normalization transformation. Figure 8.1 presents the most accurate groundwater level spatial variability according to the statistical metrics evaluated.

Finally this thesis also instigates the spatiotemporal and temporal-only modeling of groundwater level in a sparsely monitored basin. We use the ARX time-series model to relate the groundwater level to precipitation surplus and/or the abstraction rate. The ARX temporal model for the groundwater level is embedded in a Kalman filter to estimate the model parameters. After initial considerable fluctuations, the model adapts well with the level's temporal evolution and provides very accurate estimates. Based on the results of the predicted groundwater level, the ARX model estimates become progressively more accurate, as more data are incorporated and the model parameters are recursively refined during each update. The recursive nature of the parameter inference procedure implies that the model becomes more accurate as the length of the time series increases. As shown in Figure 7.3 and Figure 7.4, the groundwater level in the Mires basin has a definite declining trend. The model captures this trend, adapts well with the extremes and accurately predicts the groundwater level for the time periods 2007-2010. A reliable prediction of future groundwater levels for the Mires basin can lead to a scientifically sound management plan for the exploitation of the groundwater resources in the area. Meaning that

different scenarios can be examined regarding pumping activity and precipitation trends and also by taking account the real water needs of the area to control the pumping rates especially in years that groundwater level is expected to drop.

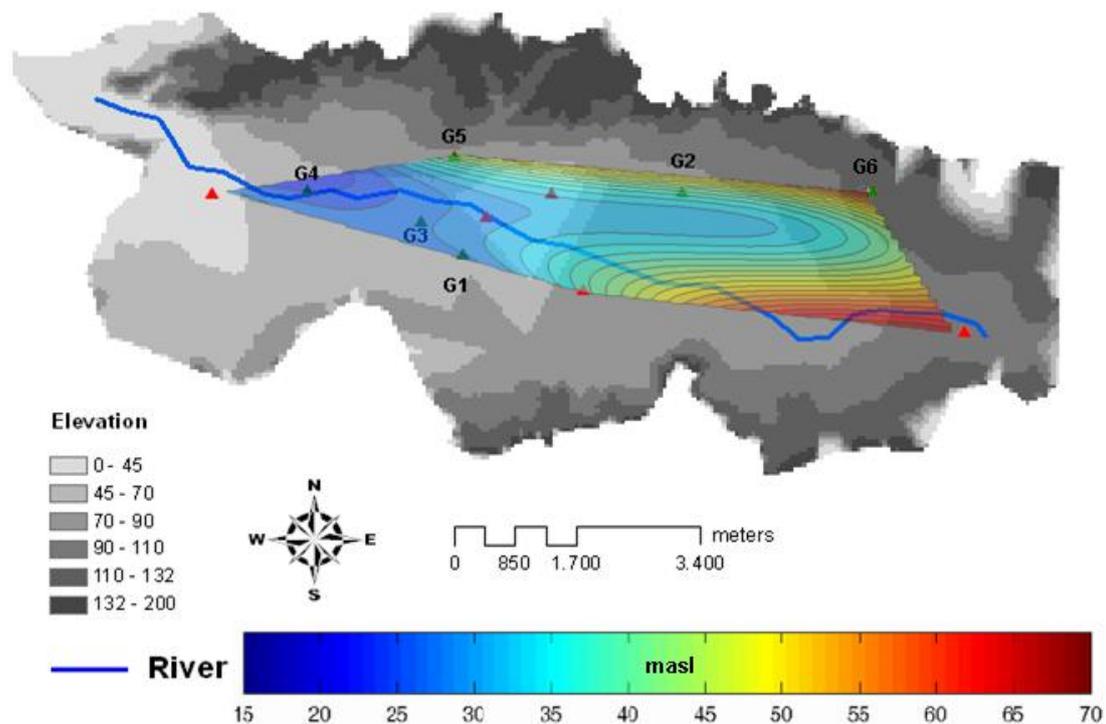


**Figure 8.1** Map of estimated groundwater level in the Mires basin using RK-T-MW spatial model, adapted on the real basin coordinates and location in the valley.

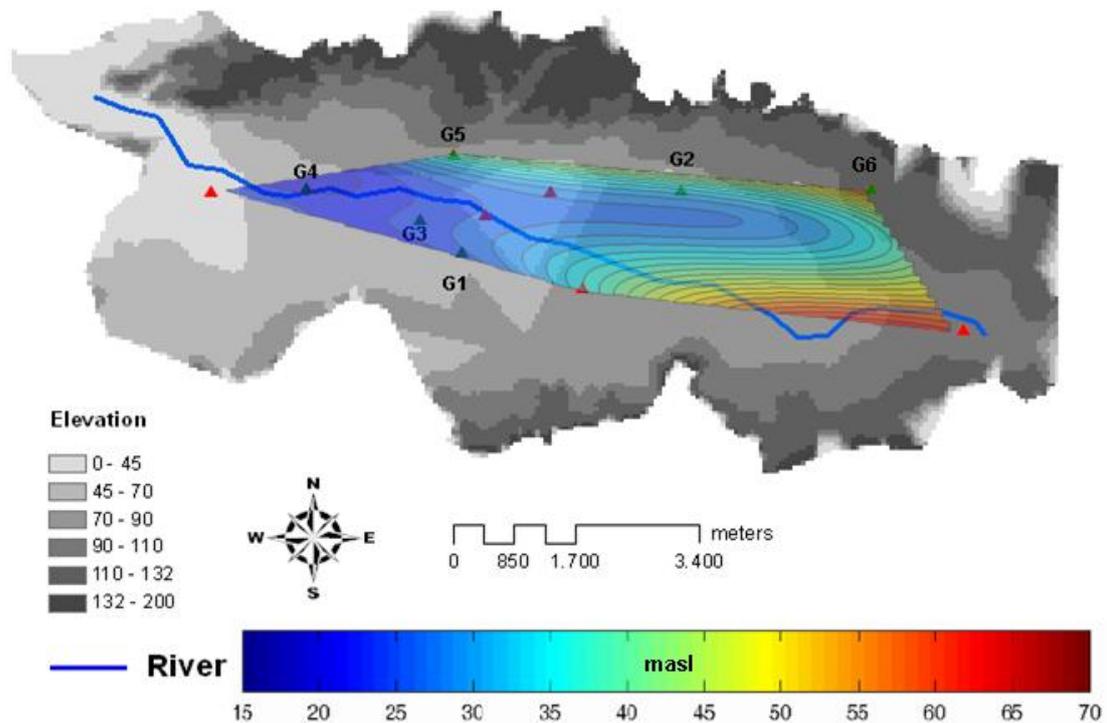
The spatiotemporal groundwater level modeling involves the spatiotemporal trend approximation using an innovative approach. The temporal and the spatial trend are separately determined and then are combined to determine the spatiotemporal trend of the time series data. The combination involves first the multiplication of the temporal and of the spatial trend. Then the trend values are divided with the temporal trend of the reference year to produce coefficients without units which is necessary in order spatiotemporal trend to retain the assessed data units (meters). The product of temporal and spatial trend components is not new; however the use of the exponentially weighted moving average filter for the temporal trend and the use of a spatial component based on the distance from the riverbed are new elements. In addition the spatiotemporal approach involves the application of a spatiotemporal covariance function that is based on the diffusion equation. The non-separable spatiotemporal semivariogram structure obtained fits very well the experimental space-time semivariogram of the residuals. The STRK estimates based on this semivariogram are more accurate than those based on separable, product-type semivariograms. Figure 8.2 and Figure 8.3 present the groundwater level spatial

variability, obtained with the non-separable spatiotemporal semivariogram model, in Mires basin for the hydrological periods 2003-2004 (wet) and 2004-2005 (dry) adapted on the real basin coordinates and topographical location. These two periods are characteristic of the dataset as significant groundwater level increase occurs in the wet period of 2003-2004 and considerable groundwater level drop in the dry period of 2004-2005.

The non-separable semivariogram is shown to provide a reliable alternative in spatiotemporal semivariogram modeling. Another advantage is that it has fewer parameters (three) compared to the product of Matérn functions. Reliable STRK estimates are crucial for groundwater recourses management as they provide information for groundwater level spatiotemporal variability. The potential combination with the ARX model can lead to an integrated approach for stochastic spatiotemporal modeling and prediction of groundwater level in the basin.



**Figure 8.2** Map of estimated groundwater level in the Mires basin using STRK and the non-separable space-time semivariogram (wet period of 2003-2004 hydrological year) adapted on the real basin coordinates and location in the valley.



**Figure 8.3** Map of estimated groundwater level in the Mires basin using STRK and the non-separable space-time semivariogram (dry period of 2004-2005 hydrological year) adapted on the real basin coordinates and location in the valley.

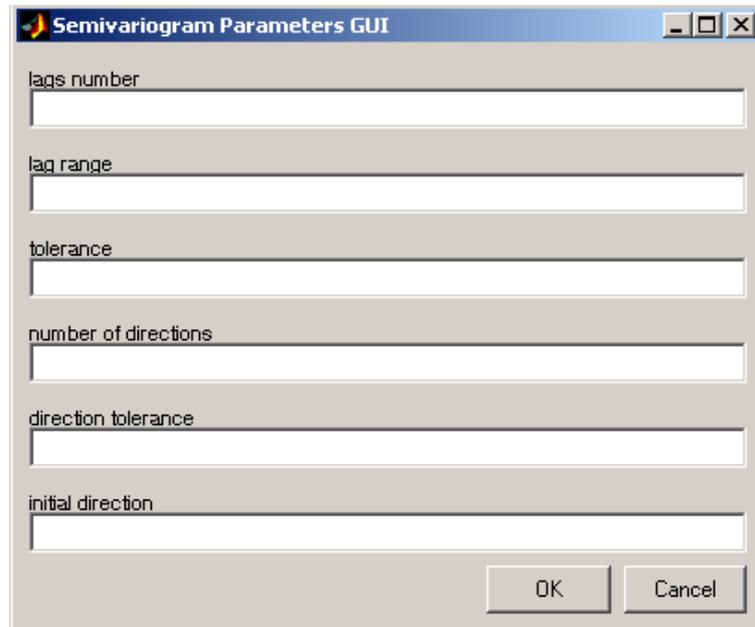
## 8.2 Future work and perspectives

This thesis introduces new geostatistical tools and ideas for space-time mapping which contribute in applied geostatistics. In this section some ideas for future research are presented that emerged after the conclusion of the thesis.

A software tool can be developed based on the code written for this thesis. This is a Graphical User Interface (GUI) in Matlab<sup>®</sup> that can give easy access to all the methodologies introduced and applied in this thesis. A GUI for isotropic and anisotropic semivariogram fitting and parameters calculation has been already developed (Figure 8.4). Thus the application of the complex space-time models could become user friendly and straightforward for non-expert users.

The MBC data normalization method introduced in this thesis is optimal for the specific case study. Application of the method in association with kriging methodology to a wide range of datasets will lead to safer general conclusions for the suggested method's efficiency. Further investigation of GA normalization method on the hydraulic head fluctuations derived by the spatial trend models presented in

chapter 6 may lead to interesting conclusions for the method's efficiency on residuals data that deal with negative values.



**Figure 8.4** Graphical User Interface (GUI) in Matlab<sup>®</sup> for isotropic and anisotropic semivariogram fitting.

Further investigation is needed for the spatiotemporal trend variance due to the complex nature of its calculation. The variance of the spatial trend component is based on the predictors at the prediction location (i.e. shorter distance from the river) but the variance of the exponential weighted moving average filter for the temporal trend and the total trend variance of the spatiotemporal trend function need further investigation. This is necessary in order to calculate STRK estimations variance.

A more extensive comparison of separable and non-separable spatiotemporal semivariograms interpolation efficiency is also necessary. Application of additional separable and non-separable space-time semivariogram functions (Kolovos *et al.* 2004) should be tested. In addition new models based on SSRF's (Hristopulos 2003b) can be developed and verified.

Incorporation of a climate change or statistical stochastic (gamma distribution) model for the precipitation estimations in the ARX model, can lead to improved temporal groundwater level predictions depending on the models' reliability. If groundwater level time series become available for individual wells of the basin, then the ARX model application can be extended for local predictions.

Finally, enrichment of the available dataset with more space-time groundwater level data (additional measurements locations as well as more frequent and recent measurements) could lead to more representative spatiotemporal semivariograms. Consequently, more accurate estimates and reliable predictions of the spatial and temporal distribution of the basin's groundwater level or fluctuations (residuals) can be expected. Additionally, a seasonal extension of the ARX model could provide the spatiotemporal trend model with information on the seasonal average of groundwater level and therefore to spatiotemporal trend predictions.

## Publications in international journals and conferences

### Journal publications of thesis' original aim and scope

- 1) **E.A. Varouchakis** and D. T. Hristopulos, "Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables", *Advances in Water Resources*, doi: <http://dx.doi.org/10.1016/j.advwatres.2012.08.002>, 2012.
- 2) **E. A. Varouchakis**, D. T. Hristopulos and G.P. Karatzas, "Improving kriging of groundwater level data using non-linear normalizing transformations-A field application", *Hydrological Sciences Journal*, 57 (7), 1404-1419, 2012.
- 3) **E. A. Varouchakis** and D. T. Hristopulos, "Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins", *Environmental Monitoring and Assessment*, doi: 10.1007/s10661-012-2527-y, 2012.
- 4) S. N. Elogne, D. T. Hristopulos, and **E. Varouchakis**, "An Application of Spartan Spatial Random Fields in Environmental Mapping: Focus on Automatic Mapping Capabilities", *Stochastic Environmental Research and Risk Assessment (SERRA)*, 22(5), 633–646, 2008.

### Journal publications based on work derived from this thesis

- 5) M. P. Papadopoulou, **E. A. Varouchakis** and G. P. Karatzas, "Terrain Discontinuities Effects in the Regional Flow of a Complex Karstified Aquifer", *Environmental Modeling & Assessment*, 15 (5), 319-328, 2010.
- 6) M. P. Papadopoulou, **E. A. Varouchakis** and G. P. Karatzas, "Simulation of complex aquifer behavior using numerical and geostatistical methodologies", *Desalination*, 237, (1-3), 42-53, 2009.

### Conference publications

- 1) **E. A. Varouchakis**, D. T. Hristopulos, "Stochastic space-time modelling of groundwater level variations in a Mediterranean basin", Geophysical Research Abstracts, Vol. 13, EGU2011-4719, European Geosciences Union, Vienna, 2011.
- 2) **E. A. Varouchakis**, D. T. Hristopulos, and G. P. Karatzas, "A Study of the Groundwater Level Spatial Variability in the Messara Valley of Crete", Geophysical Research Abstracts, Vol. 11, EGU2009-9351-1, European Geosciences Union, Vienna, 2009.

- 3) **E.A. Varouchakis**, D. T. Hristopulos, "Geostatistical Study of Groundwater Level Variability in the Messara Valley of Crete", Geophysical Research Abstracts, Vol. 10, EGU2008-A-00281, European Geosciences Union, Vienna, 2008.
- 4) **E. Varouchakis**, S. Elogne and D. T. Hristopulos "Geostatistical Applications of Spartan Spatial Random Fields in Environmental Mapping", Geophysical Research Abstracts, Vol. 8, 00674, European Geosciences Union, Vienna, 2006.

## References

- Abedini, M. J., Nasser, M. and Ansari, A. 2008. Cluster-based ordinary kriging of piezometric head in West Texas/New Mexico - Testing of hypothesis. *Journal of Hydrology*, 351 (3-4), 360-367.
- Aboufirassi, M. and Marino, M. 1983. Kriging of water levels in the Souss aquifer, Morocco. *Mathematical Geology*, 15 (4), 537-551.
- Ahlfeld, D. P. and Hoque, Y. 2008. Impact of simulation model solver performance on ground water management problems. *Ground Water*, 46 (5), 716-726.
- Ahmadi, S. and Sedghamiz, A. 2007. Geostatistical analysis of spatial and temporal variations of groundwater Level. *Environmental Monitoring and Assessment*, 129 (1), 277-294.
- Ahmadi, S. and Sedghamiz, A. 2008. Application and evaluation of kriging and cokriging methods on groundwater depth mapping. *Environmental Monitoring and Assessment*, 138 (1), 357-368.
- Ahmed, S., 2007. Application of geostatistics in hydrosciences. In: M. Thangarajan ed. *Groundwater*. Netherlands: Springer, 78-111.
- 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.
- Alsamamra, H., Ruiz-Arias, J. A., Pozo-Vazquez, D. and Tovar-Pescador, J. 2009. A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain. *Agricultural and Forest Meteorology*, 149 (8), 1343-1357.
- Altunkaynak, A. 2007. Steady-state groundwater flow model with variable hydraulic conductivity. *Hydrological Sciences Journal*, 52 (1), 221-229.
- Armstrong, M., 1998. *Basic linear geostatistics*. Berlin: Springer Verlag.
- Atkinson, P. M. and Lloyd, C. D., 2010. *Geostatistics for Environmental Applications*. Springer Verlag.
- Balkhair, K. S. 2006. Comments on the article "Determination of groundwater availability in shallow arid region aquifers utilizing GIS technology: A case study in Hada Al-Sham, Western Saudi Arabia" by A. S. El-Hames and K. A. Al-Ahmed (Hydrogeology Journal (2005), Volume 13, No.4, p 640-648, and

- Erratum in Volume 13, No. 4, p 649). *Hydrogeology Journal*, 14 (7), 1380-1382.
- Bates, B. C., Kundzewicz, Z. W., Wu, S. and Palutikof, J. P., 2008. *Climate change and water technical paper VI*. Geneva: IPCC Secretariat.
- Bear, J., 1979. *Hydraulics of groundwater*. New York: McGraw-Hill.
- Beven, K. J. and Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24 (1), 43–69.
- Bierkens, M. F. P. 2001. Spatio-temporal modelling of the soil water balance using a stochastic model and soil profile descriptions. *Geoderma*, 103 (1-2), 27-50.
- Bierkens, M. F. P., Knotters, M. and Hoogland, T. 2001. Space-Time Modeling of Water Table Depth Using a Regionalized Time Series Model and the Kalman Filter. *Water Resour. Res.*, 37 (5), 1277-1290.
- Bochner, S., 1959. *Lectures on Fourier integrals*. Princeton, NJ: Princeton University Press.
- Boezio, M., Costa, J. and Koppe, J. 2006a. Accounting for extensive secondary information to improve watertable mapping. *Natural Resources Research*, 15 (1), 33-48.
- Boezio, M., Costa, J. and Koppe, J. 2006b. Kriging with an external drift versus collocated cokriging for water table mapping. *Applied Earth Science*, Volume 115 (3), 103-112(110).
- Bogaert, P. 1996. Comparison of kriging techniques in a space-time context. *Mathematical Geology*, 28 (1), 73-86.
- Box, G. E. P. and Cox, D. R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society*, Ser. B 26 (2), 211–252.
- Briggs, I. C. 1974. Machine contouring using minimum curvature. *Geophysics*, 39 (1), 39-48.
- Brus, D. J., Bogaert, P. and Heuvelink, G. B. M. 2008. Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science*, 59 (2), 166-177.
- Buchanan, S. and Triantafilis, J. 2009. Mapping water table depth using geophysical and environmental variables. *Ground Water*, 47 (1), 80-96.
- Burrough, P. A. 2001. GIS and geostatistics: Essential partners for spatial analysis. *Environmental and Ecological Statistics*, 8 (4), 361-377.

- Butler Jr., J. J. 1988. Pumping tests in nonuniform aquifers - The radially symmetric case. *Journal of Hydrology*, 101 (1-4), 15-30.
- Casa, R. and Castrignano, A. 2008. Analysis of spatial relationships between soil and crop variables in a durum wheat field using a multivariate geostatistical approach. *European Journal of Agronomy*, 28 (3), 331-342.
- Chander, S., Kumar, A. and Spolia, S. K. 1978. Flood frequency analysis by power transformation. *Journal of the Hydraulics Division: proceedings ASCE*, 104 (11), 1495–1504.
- Charbeneau, R., 2000. *Groundwater Hydraulics and Pollutant Transport*. New York: Prentice Hall.
- Chartzoulakis, K. and Psarras, G. 2005. Global change effects on crop photosynthesis and production in Mediterranean: The case of Crete, Greece. *Agriculture, Ecosystems and Environment*, 106 (2-3 .), 147-157.
- Chiles, J. P. and Delfiner, A., 1999. *Geostatistics (modeling spatial uncertainty)*. New York: Wiley.
- Chorti, A. and Hristopulos, D. T. 2008. Non-parametric identification of anisotropic (Elliptic) correlations in spatially distributed data sets. *IEEE Transactions on Signal Processing*, 56 (10), 4738-4751.
- Christakos, G. 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, 22 (7), 763-777.
- Christakos, G. 1991a. On certain classes of spatiotemporal random fields with applications to space - time data processing. *IEEE Transactions on Systems, Man and Cybernetics*, 21 (4), 861-875.
- Christakos, G., 1991b. *Random field models in earth sciences*. San Diego: Academic press.
- Christakos, G., 2000. *Modern Spatiotemporal Geostatistics*. New York, NY: Oxford University Press.
- Christakos, G., Bogaert, P. and Serre, M. L., 2001. *Temporal GIS: Advanced functions for field-based applications*. Springer Verlag.
- Christakos, G. and Hristopulos, D. T., 1998. *Spatiotemporal environmental health modelling: a tractatus stochasticus*. Springer.
- Christakos, G., Hristopulos, D. T. and Bogaert, P. 2000. On the physical geometry concept at the basis of space/time geostatistical hydrology. *Advances in Water Resources*, 23 (8), 799-810.

- Christakos, G. and Serre, M. L. 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmospheric Environment*, 34 (20), 3393-3406.
- Clark, I. and Harper, W. V., 2000. *Practical Geostatistics 2000*. Columbus, Ohio, USA: Ecosse North America Llc.
- Cooke, R., Mostaghimi, S. and Parker, J. C. 1993. Estimating oil spill characteristics from oil heads in scattered monitoring wells. *Environmental Monitoring and Assessment*, 28 (1), 33-51.
- Cornford, D., 2005. Are comparative studies a waste of time? SIC2004. In: G. Dubois ed. *Automatic mapping algorithms for routine and emergency monitoring data. EUR 21595 EN – Scientific and technical research series, ISBN 92-894-9400-X (pp. 61-70)*. EUR, 2005 ed. Luxembourg: Office for official publications of the European Communities.
- Cressie, N., 1993. *Statistics for spatial data (revised ed.)*. New York: Wiley.
- Cressie, N. and Huang, H. C. 1999. Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. *Journal of the American statistical Association*, 94 (448), 1330-1340.
- Croke, B., Cleridou, N., Kolovos, A., Vardavas, I. and Papamastorakis, J. 2000. Water resources in the desertification-threatened Messara Valley of Crete: estimation of the annual water budget using a rainfall-runoff model. *Environmental Modelling and Software*, 15 (4), 387-402.
- Daliakopoulos, I. N., Coulibaly, P. and Tsanis, I. K. 2005. Groundwater Level Forecasting Using Artificial Neural Networks. *Journal of Hydrology*, 309 (1-4), 229-240.
- Davis, J. C., 1973. *Statistics and data analysis in geology*. New York: Wiley.
- De-Vitry, C., Vann, J. and Arvidson, H. 2010. Multivariate iron ore deposit resource estimation - a practitioner's guide to selecting methods. *Transactions of the Institutions of Mining and Metallurgy, Section B: Applied Earth Science*, 119 (3), 154-165.
- De Cesare, L., Myers, D. E. and Posa, D. 2002. FORTRAN programs for space-time modeling. *Computers and Geosciences*, 28 (2), 205-212.
- De Cesare, L. D., Myers, D. E. and Posa, D. 2001. Estimating and modeling space-time correlation structures. *Statistics & Probability Letters*, 51 (1), 9-14.

- De Iaco, S. 2010. Space-time correlation analysis: A comparative study. *Journal of Applied Statistics*, 37 (6), 1027-1041.
- De Iaco, S., Myers, D. E. and Posa, D. 2001. Space-time analysis using a general product-sum model. *Statistics and Probability Letters*, 52 (1), 21-28.
- De Iaco, S., Myers, D. E. and Posa, D. 2002a. Nonseparable space-time covariance models: Some parametric families. *Mathematical Geology*, 34 (1), 23-42.
- De Iaco, S., Myers, D. E. and Posa, D. 2002b. Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics and Data Analysis*, 41 (2), 311-328.
- De Oliveira, V., Kedem, B. and Short, D. A. 1997. Bayesian prediction of transformed gaussian random fields. *Journal of the American statistical Association*, 92 (440), 1422-1433.
- Delhomme, J. P., La cartographie d'une grandeur physique a partir des donnees de differentes qualities. In: International Association of Hydrogeologists, ed. *Proc. of IAH Congress, Montpellier, France, 1974* Montpellier. France: IAH, 185-194.
- Delhomme, J. P. 1978. Kriging in the hydrosiences. *Advances in Water Resources*, 1 (5), 251-266.
- Department of Water Resources Management, 2000. *Integrated Water Resources Management of Crete(in Greek)*. Heraklion: Prefecture of Crete Planning and Development Section.
- Department of Water Resources Management, 2009. *Groundwater level monitoring of Crete's hydrological basins(in Greek)*. Heraklion: Prefecture of Crete Planning and Development Section.
- Desbarats, A. J., Logan, C. E., Hinton, M. J. and Sharpe, D. R. 2002. On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, 255 (1-4), 25-38.
- Deutsch, C. V. and Journel, A. G., 1992. *GSLIB. Geostatistical software library and user's guide*. New York: Oxford University Press.
- Dimitrakopoulos, R. and Luo, X., 1994. Spatiotemporal modeling: covariances and ordinary kriging systems. In: R. Dimitrakopoulos ed. *Geostatistics for the Next Century*. Dordrecht: Kluwer Academic Publishers, , 88-93.
- Donta, A. A., Lange, M. A. and Herrmann, A., 2006. *Water on Mediterranean islands: Current conditions and prospects for sustainable management*.

- Project No EVK1-CT-2001-00092-Funded by the European Commission, ISBN 3-9808840-7-4. Muenster: Centre for Environment Research (CER), University of Muenster.*
- Draper, N. and H. Smith, 1981. *Applied Regression Analysis*. second ed. ed. New York: Wiley.
- Dubois, G. 1998. Spatial interpolation comparison 97: Foreword and introduction. *Journal of Geographical Information and Decision Analysis*, 2 (2), 1-10.
- Dubois, G. and Galmarini, S., 2005. Spatial interpolation comparison SIC2004: introduction to the exercise and overview on the results. In: G. Dubois ed. *Automatic mapping algorithms for routine and emergency monitoring data. EUR 21595 EN – Scientific and technical research series, ISBN 92-894-9400-X (pp 7-18)*. EUR, 2005 ed.: Luxembourg: Office for Official Publications of the European Commission.
- Eigbe, U., Beck, M. B., Wheater, H. S. and Hirano, F. 1998. Kalman filtering in groundwater flow modelling: Problems and prospects. *Stochastic Hydrology and Hydraulics*, 12 (1), 15-32.
- Elogne, S., Hristopulos, D. and Varouchakis, E. 2008. An application of Spartan spatial random fields in environmental mapping: focus on automatic mapping capabilities. *Stochastic Environmental Research and Risk Assessment*, 22 (5), 633-646.
- Famiglietti, J., 2008. *Groundwater*. Food and Agriculture Organization of the United Nations, Terrestrial ECV 03.
- Fasbender, D., Peeters, L., Bogaert, P. and Dassargues, A. 2008. Bayesian data fusion applied to water table spatial mapping. *Water Resources Research*, 44 (W12422), doi:10.1029/2008WR006921.
- Fatima, Z. 2006. Estimating soil contamination with kriging interpolation method. *American Journal of Applied Sciences*, 3 (6), 1894-1898.
- Feder, J., 1988. *Fractals*. New York, NY: Plenum Press.
- Fischer, M. M. and Getis, A., 2010. *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Verlag.
- Fytolakis, N., Peterek, A. and Schroder, B. 2005. Initial geoarchaeologic investigations on the Holocene coastal configuration near Phaistos/Agia Triada (Mesara Plain, central Crete, Greece. *Z. Geomorph. N.F.*, 137, 111-113.

- Gambolati, G. and Volpi, G. 1979a. A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resources Research*, 15 (3), 625-629.
- Gambolati, G. and Volpi, G. 1979b. Groundwater Contour Mapping in Venice by Stochastic Interpolators 1. Theory. *Water Resources Research*, 15 (2), 281-290.
- Garg, N. K. and Ali, A. 2000. Groundwater management for Lower Indus Basin. *Agricultural Water Management*, 42 (3), 273-290.
- Gelhar, L. W. 1986. Stochastic subsurface hydrology from theory to applications. *Water Resources Research*, 22 (9 Suppl), 135-145.
- Gething, P. W., Atkinson, P. M., Noor, A. M., Gikandi, P. W., Hay, S. I. and Nixon, M. S. 2007. A local space-time kriging approach applied to a national outpatient malaria data set. *Computers & Geosciences Spatial Analysis - Spatial Analysis*, 33 (10), 1337-1350.
- Giraldo Henao, R., 2009. *Geostatistical analysis of functional data*. PhD. Universitat Politècnica de Catalunya.
- Gneiting, T. 2002. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97 (458), 590-600.
- Gneiting, T., Genton, M. G. and Guttorp, P., 2007. *Geostatistical space-time models, stationarity, separability and full symmetry*. Department of Statistics, University of Washington, Technical Report no. 475.
- Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press.
- Goovaerts, P. 1999. Using elevation to aid the geostatistical mapping of rainfall erosivity. *Catena*, 34 (3-4), 227-242.
- Goovaerts, P., Avruskin, G., Meliker, J., Slotnick, M., Jacquez, G. and Nriagu, J. 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resources Research*, 41 (W07013, doi:10.1029/2004WR003705).
- Gringarten, E. and Deutsch, C. V. 2001. Teacher's aide: Variogram interpretation and modeling. *Mathematical Geology*, 33 (2001), 507-534.
- Hengl, T., 2007. *A practical guide to geostatistical mapping of environmental variables*. Luxembourg: Office for Official Publications of the European Communities.

- Hengl, T., Geuvelink, G. B. M. and Stein, A., 2003. *Comparison of kriging with external drift and regression-kriging. Technical note, ITC.*
- Hengl, T., Heuvelink, G. B. M., Perčec Tadić, M. and Pebesma, E. J. 2011. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology*, 1-13.
- Hengl, T., Heuvelink, G. B. M. and Rossiter, D. G. 2007. About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33 (10), 1301-1315.
- Hessami, M., Anctil, F. and Viau, A. A. 2001. Delaunay implementation to improve kriging computing efficiency. *Computers & Geosciences*, 27 (2), 237-240.
- Heuvelink, G. B. M. and Egmond, F. M., 2010. Space–Time Geostatistics for Precision Agriculture: A Case Study of NDVI Mapping for a Dutch Potato Field Geostatistical Applications for Precision Agriculture. *In: M. A. Oliver ed.: Springer Netherlands*, 117-137.
- Heuvelink, G. B. M. and Griffith, D. A. 2010. Space-time geostatistics for geography: A case study of radiation monitoring across parts of Germany. *Geographical Analysis*, 42 (2), 161-179.
- Heuvelink, G. B. M., Musters, P. and Pebesma, E. J., 1996. Spatio-temporal kriging of soil water content. *In: E. Y. Baafi and N. A. Schofield eds. Geostatistics Wollongong. Dordrecht: Kluwer Academic Publishing*, 1020– 1030.
- Hirsch, R. M. 1979. Synthetic hydrology and water supply reliability. *Water Resources Research*, 15 (6), 1603–1615.
- Hoeksema, R. J., Clapp, R. B., Thomas, A. L., Hunley, A. E., Farrow, N. D. and Dearstone, K. C. 1989. Cokriging model for estimation of water table elevation. *Water Resources Research*, 25 (3), 429–438.
- Hohn, M. E., 1999. *Geostatistics and petroleum geology*. Springer.
- Hoogland, T., Heuvelink, G. B. M. and Knotters, M. 2010. Mapping water-table depths over time to assess desiccation of groundwater-dependent ecosystems in the Netherlands. *Wetlands*, 30 (1), 137-147.
- Howard, K. W. F., 2011. Implications of Climate Change on Water Security in the Mediterranean Region. *In: A. Baba, et al. eds. Climate Change and its Effects on Water Resources. Springer Netherlands*, 9-16.

- Hristopulos, D. T. 2002. New anisotropic covariance models and estimation of anisotropic parameters based on the covariance tensor identity. *Stochastic Environmental Research and Risk Assessment*, 16 (1), 43-62.
- Hristopulos, D. T., 2003a. Introduction to Geostatistics-course notes (in Greek). Chania, Crete, Greece: Technical University of Crete, 200.
- Hristopulos, D. T. 2003b. Spartan Gibbs random field models for geostatistical applications. *SIAM Journal on Scientific Computing*, 24 (6), 2125-2162.
- Hristopulos, D. T., 2006. Data analysis-postgraduate course notes (in Greek). Chania, Crete, Greece: Technical University of Crete, 50.
- Hristopulos, D. T., 2008. Applied geostatistics-course notes (in Greek). Chania, Crete, Greece: Technical University of Crete, 200.
- Hristopulos, D. T. and Elogne, S. N. 2007. Analytic properties and covariance functions for a new class of generalized Gibbs random fields. *IEEE Transactions on Information Theory*, 53 (12), 4667-4467.
- Hristopulos, D. T. and Elogne, S. N. 2009. Computationally efficient spatial interpolators based on spartan spatial random fields. *IEEE Transactions on Signal Processing*, 57 (9), 3475-3487.
- Intergovernmental Panel on Climate Change, 2007. *Climate change 2007: synthesis report – summary for policymakers. Contribution of Working Groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Geneva: IPCC.
- Isaaks, E. H. and Srivastava, R. M., 1989. *An introduction to applied geostatistics*. New York: Oxford University Press.
- Jain, D. and Singh, V. P. 1986. A comparison of transformation methods for flood frequency analysis. *Water Resources Bulletin*, 22 (6), 903–912.
- Jost, G., Heuvelink, G. B. M. and Papritz, A. 2005. Analysing the space-time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. *Geoderma*, 128 (3-4 SPEC. ISS.), 258-273.
- Journel, A. G. and Huijbregts, C., 1978. *Mining geostatistics*. New York: Academic Press.
- Jowett, G. H. 1955. Sampling properties of local statistics in stationary stochastic series. *Biometrika*, 42, 160–169.
- Karterakis, S. M., Karatzas, G. P., Nikolos, I. K. and Papadopoulou, M. P. 2007. Application of linear programming and differential evolutionary optimization

- methodologies for the solution of coastal subsurface water management problems subject to environmental criteria. *Journal of Hydrology*, 342 (3-4), 270-282.
- Kay, M. and Dimitrakopoulos, R. 2000. Integrated interpolation methods for geophysical data: Applications to mineral exploration. *Natural Resources Research*, 9 (1), 53-64.
- Kholghi, M. and Hosseini, S. 2009. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. *Environmental Modeling and Assessment*, 14 (6), 729-737.
- Kilili-Polychronaki, A., 2001. *Hydrogeological study of Mires municipallity*. Rethymnon: Institute of Geology and Mineral Exploration.
- Kitanidis, P. 1993. Generalized covariance functions in estimation. *Mathematical Geology*, 25 (5), 525-540.
- Kitanidis, P. K., 1997. *Introduction to geostatistics*. Cambridge: University Press.
- Knotters, M. and Bierkens, M. F. P. 2001. Predicting water table depths in space and time using a regionalised time series model. *Geoderma*, 103 (1-2), 51-77.
- Knotters, M. and Bierkens, M. F. P. 2002. Accuracy of spatio-temporal RARX model predictions of water table depths. *Stochastic Environmental Research and Risk Assessment*, 16 (2), 112-126.
- Knotters, M., Brus, D. J. and Oude Voshaar, J. H. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67 (3-4), 227-246.
- Kolmogorov, A. N. 1941. Interpolirovanie i ekstrapolirovanie statsionarnykh sluchainykh posledovatel' nostei (Interpolated and extrapolated stationary random sequences). *Isvestia AN SSSR, Seriya Matematicheskaya*, 5 (1).
- Kolovos, A., Angulo, J., Modis, K., Papantonopoulos, G., Wang, J.-F. and Christakos, G. 2012. Model-driven development of covariances for spatiotemporal environmental health assessment. *Environmental Monitoring and Assessment*, 1-17.
- Kolovos, A., Christakos, G., Hristopoulos, D. T. and Serre, M. L. 2004. Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources*, 27 (8), 815-830.
- Kozintseva, A., 1999. *Comparison of Three Methods of Spatial Prediction*. Thesis (MSc). University of Maryland.

- Krige, D. G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*. 119-139.
- Krige, D. G. 1966. Two-dimensional weighted moving average trend surfaces for ore evaluation. *Journal of the South African Institute of Mining and Metallurgy*, 66, 13–38.
- Kritsotakis, M., 2010. *Water resources management in Mesara valley of Crete (in Greek: Διαχείριση υδατικών πόρων Μεσαράς Κρήτης)*. PhD. Technical University of Crete.
- Kritsotakis, M. and Tsanis, I. 2009. An integrated approach for sustainable water resources management of Messara basin, Crete, Greece. *European Water*, 27 (28), 15-30.
- Kumar, V. 2007. Optimal contour mapping of groundwater levels using universal kriging-a case study. *Hydrological Sciences*, 52 (5), 1038-1050.
- Kyriakidis, P. and Journel, A. 1999. Geostatistical Space-Time Models: A Review. *Mathematical Geology*, 31 (6), 651-684.
- Kyriakidis, P. C. and Journel, A. G. 2001a. Stochastic modeling of atmospheric pollution: a spatial time-series framework. Part I: methodology. *Atmospheric Environment*, 35 (13), 2331-2337.
- Kyriakidis, P. C. and Journel, A. G. 2001b. Stochastic modeling of atmospheric pollution: a spatial time-series framework. Part II: application to monitoring monthly sulfate deposition over Europe. *Atmospheric Environment*, 35 (13), 2339-2348.
- Lantuejoul, C., 2002. *Geostatistical simulation*. New York: Springer.
- Lanzi, P. L., Loiacono, D., Wilson, S. W. and Goldberg, D. E., Prediction update algorithms for XCSF: RLS, kalman filter, and gain adaptation. In: M. Keijzer, ed. *Genetic and Evolutionary Computation Conference (GECCO)*, 2006: ACM, 1505-1512.
- Lee, C. C., Tan, Y. C., Chen, C. H. and Jim Yeh, T. C. 2001. Stochastic series lumped rainfall-runoff model for a watershed in Taiwan. *Journal of Hydrology*, 249 (1-4), 30-45.
- Lee, S.-J. and Wentz, E. A. 2008. Applying Bayesian Maximum Entropy to extrapolating local-scale water consumption in Maricopa County, Arizona. *Water Resour. Res.*, 44 (1), W01401.

- Lee, S. J., Wentz, E. A. and Gober, P. 2010. Space-time forecasting using soft geostatistics: A case study in forecasting municipal water demand for Phoenix, Arizona. *Stochastic Environmental Research and Risk Assessment*, 24 (2), 283-295.
- Leuangthong, O., McLennan, J. A. and Deutsch, C. V. 2004. Minimum acceptance criteria for geostatistics realizations. *Natural Resources Research*, 13 (3), 131-141.
- Ling, M., Rifai, H. S. and Newell, C. J. 2005. Optimizing groundwater long-term monitoring networks using Delaunay triangulation spatial analysis techniques. *Environmetrics*, 16 (6), 635-657.
- Ljung, L., 1999. *System Identification - Theory for the User*,. Upper Saddle River, N.J: Prentice Hall.
- Ly, S., Charles, C. and Degré, A. 2011. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, 15 (7), 2259-2274.
- Mamassis, N., 2006. Advance Hydrology (in Greek). *Introduction to Geostatistics*. National Technical University of Athens, 100.
- Mandelbrot, B. B. and Van Ness, J. W. 1968. Fractional Brownian Motions, fractional noises and applications. *SIAM Review*, 10 (4), 422-437.
- Massey Jr, F. J. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46 (253), 68-78.
- Matérn, B. 1960. Spatial variation. *Meddelanden fran Statens Skogsforskningsinstitut*, 49 (5), 1-144.
- Matheron, G., 1963. Principles of geostatistics. *Economic Geology*. 1246-1266.
- Matheron, G., 1969. *Le krigeage universel (Universal kriging)*. vol. 1. Ecole des Mines de Paris, Fontainebleau.
- Matheron, G., 1971. *The theory of regionalized variables and its applications*. Fontainebleau, Paris: Ecole Nationale Supérieure des Mines de Paris.
- Mendonca, C. A. and Silva, J. B. C. 1995. Interpolation of potential-field data by equivalent layer and minimum curvature: A comparative analysis. *Geophysics*, 60 (2), 399-407.

- Mendoza-Cazares, E. Y. and Herrera-Zamarron, G. D. 2010. Spatiotemporal estimation of hydraulic head using a single spatiotemporal random function model. *Tecnologia Y Ciencias Del Agua*, 1 (2), 87-111.
- Monin, A. and Yaglom, A., 1975. *Statistical fluid mechanics: mechanics of turbulence, volume 2*. Cambridge Univ Press.
- Moukana, J. A. and Koike, K. 2008. Geostatistical model for correlating declining groundwater levels with changes in land cover detected from analyses of satellite images. *Computers & Geosciences*, 34 (11), 1527-1540.
- Mouser, P. J., Rizzo, D. M., Röling, W. F. M. and Van Breukelen, B. M. 2005. A multivariate statistical approach to spatial representation of groundwater contamination using hydrochemistry and microbial community profiles. *Environmental Science and Technology*, 39 (19), 7551-7559.
- Mulchrone, K. F. 2003. Application of delaunay triangulation to the nearest neighbour method of strain analysis. *Journal of Structural Geology*, 25 (5), 689-702.
- Myers, D. E. 2002. Space-time correlation models and contaminant plumes. *Environmetrics*, 13 (5-6), 535-553.
- Myers, D. E., 2005. Spatial interpolation comparison exercise 2004: a real problem or an academic exercise. In: G. Dubois ed. *Automatic mapping algorithms for routine and emergency monitoring data. EUR 21595 EN – Scientific and technical research series, ISBN 92-894-9400-X (pp. 79-88)*. EUR, 2005 ed. Luxembourg: Office for official publications of the European Communities.
- Myers, D. E., De Iaco, S., Posa, D. and De Cesare, L. 2002. Space-time radial basis functions. *Computers and Mathematics with Applications*, 43 (3-5), 539-549.
- Myers, D. E. and Journel, A. 1990. Variograms with zonal anisotropies and noninvertible kriging systems. *Mathematical Geology*, 22 (7), 779-785.
- Nelder, J. A. and Mead, R. 1965. A simplex method for function minimization. *Computer Journal*, 7 (4), 308-313.
- Neuman, S. and Jacobson, E. 1984. Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Mathematical Geology*, 16 (5), 499-521.
- Nikroo, L., Kompani-Zare, M., Sepaskhah, A. and Fallah Shamsi, S. 2009. Groundwater depth and elevation interpolation by kriging methods in Mohr Basin of Fars province in Iran. *Environmental Monitoring and Assessment*, 166 (1-4), 387-407.

- Nist/Sematech, 2009. *e-Handbook of Statistical Methods* [online]. National Institute of Standards and Technology. Available from: <http://www.itl.nist.gov/div898/handbook/> [Accessed May 2012].
- Nunes, C. and Soares, A. 2005. Geostatistical space-time simulation model for air quality prediction. *Environmetrics*, 16 (4), 393-404.
- Okabe, A., Boots, B. and Sugihara, K., 1992. *Spatial tessellations: Concepts and applications of Voronoi diagrams*. New York: Wiley.
- Olea, R. 2006. A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment*, 20 (5), 307-318.
- Olea, R. A., 1999. *Geostatistics for engineers and earth scientists*. New York: Kluwer Academic Publishers.
- Ouyang, Y., Zhang, J. E. and Ou, L. T. 2006. Temporal and spatial distribution of sediment total organic carbon in an estuary river. *Journal of Environmental Quality*, 35, 93–100.
- Özger, M. and Şen, Z. 2007. Triple diagram method for the prediction of wave height and period. *Ocean Engineering*, 34 (7), 1060-1068.
- Pardo-Iguzquiza, E. and Chica-Olmo, M. 2008. Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation. *Computers & Geosciences*, 34 (9), 1073-1079.
- Pardo-Iguzquiza, E. and Dowd, P. 2005. Empirical maximum likelihood Kriging: The general case. *Mathematical Geology*, 37 (5), 477-492.
- Park, M. S. and Fuentes, M. 2008. Testing lack of symmetry in spatial-temporal processes. *Journal of Statistical Planning and Inference*, 138 (10), 2847-2866.
- Peterek, A. and Schwarze, J. 2004. Architecture and Late Pliocene to recent evolution of outer-arc basins of the Hellenic subduction zone (south-central Crete, Greece). *Journal of Geodynamics*, 38 (1), 19-55.
- Petrakis, M. P. and Hristopulos, D. T. 2012. Elliptical Anisotropy Statistics of Two-Dimensional Differentiable Gaussian Random Fields: Joint Probability Density Function and Confidence Regions. *Arxiv preprint arXiv:1203.5010*.
- Pham, H., 2006. *Springer Handbook of Engineering Statistics*. Springer-Verlag London Limited.
- Philip, G. M. and Watson, D. F. 1986. Automatic interpolation methods for mapping piezometric surfaces. *Automatica*, 22 (6), 753-756.

- Pinder, G. F. and Celia, M. A., 2006. *Subsurface hydrology*. New Jersey: John Wiley & Sons.
- Porcu, E., Gregori, P. and Mateu, J. 2006. Nonseparable stationary anisotropic space-time covariance functions. *Stochastic Environmental Research and Risk Assessment*, 21 (2), 113-122.
- Porcu, E., Mateu, J. and Saura, F. 2008. New classes of covariance and spectral density functions for spatio-temporal modelling. *Stochastic Environmental Research and Risk Assessment*, 22 (suppl.1), 65-79.
- Prakash, A., 2004. *Water resources engineering: handbook of essential methods and design*. Virginia: American Society of Civil Engineers.
- Prakash, M. R. and Singh, V. S. 2000. Network design for groundwater monitoring-a case study. *Environmental Geology*, 39 (6), 628-632.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., 1992. *Numerical Recipes in Fortran*. 2nd ed. ed. New York: Cambridge University Press.
- Pucci, A. A. J. and Murashige, J. A. E. 1987. Applications of universal kriging to an aquifer study in New Jersey. *Ground Water*, 25 (6), 672-678.
- Rivest, M., Marcotte, D. and Pasquier, P. 2008. Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information. *Journal of Hydrology*, 361 (3-4), 349-361.
- Rivoirard, J. 2002. On the Structural Link Between Variables in Kriging with External Drift. *Mathematical Geology*, 34 (7), 797-808.
- Roberts, S. 1959. Control chart tests based on geometric moving averages. *Technometrics*, 239-250.
- Rodriguez-Iturbe, I. and Mejia, M. J. 1974. The design of rainfall networks in time and space. *Water Resources Research*, 10 (4), 713-728.
- Rouhani, S. 1986. Comparative study of ground-water mapping techniques. *Ground Water*, 24 (2), 207-216.
- Rouhani, S. and Myers, D. 1990. Problems in space-time kriging of geohydrological data. *Mathematical Geology*, 22 (5), 611-623.
- Rouhani, S. and T.J. Hall, 1989. Space-time kriging of groundwater data. In: M. Armstrong ed. *Geostatistics*. The Netherlands: Kluwer Academic Publishers, 639-651.
- Sahimi, M., 2011. *Flow and transport in porous media and fractured rock: From classical methods to modern approaches*. Wiley-VCH.

- Salas, J., 1993. Analysis and modeling of hydrologic time series. In: D. Maidment ed. *Handbook of Hydrology*. New York, USA: McGraw-Hill, 19.11–19.72.
- Sandwell, D. 1987. Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter Dta. *Geophysical Research Letters*, 14 (2), 139-142.
- Schabenberger, O. and Gotway, C. A., 2005. *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: CRC Press.
- Sen, Z. and Al-Somayien, M. S. 1991. Some simple management criteria for confined Saq aquifer in Tabuk region, Saudi Arabia. *Water Resources Management*, 5 (2), 161-179.
- Serre, M. L. and Christakos, G. 1999. Modern geostatistics: Computational BME analysis in the light of uncertain physical knowledge - The Equus Beds study. *Stochastic Environmental Research and Risk Assessment*, 13 (1-2), 1-26.
- Silliman, S. E. and Caswell, S. 1998. Observations of measured hydraulic conductivity in two artificial, confined aquifers with boundaries. *Water Resources Research*, 34 (9), 2203-2213.
- Skøien, J. O. and Blöschl, G. 2007. Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, 43 (9).
- Snepvangers, J. J. J. C., Heuvelink, G. B. M. and Huisman, J. A. 2003. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112 (3-4), 253-271.
- Sophocleous, M., Paschetto, J. E. and Olea, R. A. 1982. Ground-water network design for northwest Kansas, using the theory of regionalized variables. *Ground Water*, 20 (1), 48-58.
- Sophocleous, M. A., Understanding the environmental implications of groundwater overexploitation. In: Geological Society of America, ed. *Geological Society of America Annual Meeting*, 2002 Denver: Geological Society of America Groundwater Depletion & Overexploitation: A Global Problem, 159.
- Spiliopoulos, I., Hristopulos, D. T., Petrakis, E. and Chorti, A. 2011. A Multigrid method for the estimation of geometric anisotropy in environmental data from sensor networks. *Computers & Geosciences*, 37 (3), 320-330.
- Stein, A. 1998. Analysis of space-time variability in agriculture and the environment with geostatistics. *Statistica Neerlandica*, 52, 18–41.
- Stein, M. L., 1999. *Interpolation of spatial data. Some theory for kriging*. New York: Springer.

- Steward, D. R. and Jin, W. 2006. Uniform Head in Horizontal and Vertical Wells. *Ground Water*, 44 (1), 86-90.
- Subyani, A. M. and Sen, Z. 1989. Geostatistical modelling of the Wasia aquifer in Central Saudi Arabia. *Journal of Hydrology*, 110 (3-4), 295-314.
- Sun, Y., Kang, S., Li, F. and Zhang, L. 2009. Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. *Environmental Modelling & Software*, 24 (10), 1163-1170.
- Theodossiou, N. and Latinopoulos, P. 2006. Evaluation and optimisation of groundwater observation networks using the kriging methodology. *Environmental Modelling & Software*, 21 (7), 991-1000.
- Thiem, G., 1906. Hydrologische methoden (in German). In: J. M. Gebhardt ed. *Hydrologic methods*. Leipzig, Germany, pp 56.
- Thyer, M., Kuczera, G. and Wang, Q. J. 2002. Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. *Journal of Hydrology*, 265 (1-4), 246-257.
- Tichy, M., 1993. *Applied methods of structural reliability*. Dordrecht: Springer.
- Todd, D. K., 1959. *Groundwater Hydrology*. New York, N.Y: Wiley.
- Tsanis, I. K. and Apostolaki, M. G. 2009. Estimating groundwater withdrawal in poorly gauged agricultural basins. *Water Resources Management*, 23 (6), 1097-1123.
- Tsanis, I. K., Coulibaly, P. and Daliakopoulos, I. N. 2008. Improving Groundwater Level Forecasting with an ANN and Linearly Regressed Projected Precipitation. *Journal of Hydroinformatics*, 10 (4), 317-330.
- Tsanis, I. K., Koutroulis, A. G., Daliakopoulos, I. N. and Jacob, D. 2011. Severe climate-induced water shortage and extremes in Crete. *Climatic Change*, 106 (4), 667-677.
- UNCED, 1994. *United Nations Convention to Combat Desertification in those countries experiencing serious drought and/or desertification, particularly in Africa*. New York, approved June 1994, Paris.
- Van den Boogaart, K. G., 2005. The comparison of one-click mapping procedures for emergency. In: G. Dubois ed. *Automatic mapping algorithms for routine and emergency monitoring data*. EUR 21595 EN – Scientific and technical

- research series, ISBN 92-894-9400-X (pp. 71-78). EUR, 2005 ed. Luxembourg: Office for official publications of the European Communities.*
- Van Geer, F. C. and Van Der Kloet, P. 1985. Two algorithms for parameter estimation in groundwater flow problems. *Journal of Hydrology*, 77 (1-4), 361-378.
- Vardavas, I. M., Bolle, H.-J., Bromley, J., de Bruin, H., Devereux, B. and Prastacos, P., 1996. *An integrated monitoring and modelling study of desertification and climatic change impacts in the Messara Valley of Crete. EU DGXII, Environment and Climate Change Programme, Final Report EV5V-CT94-0466.*
- Verma, P., Yeates, J. and Daly, E. 2011. A stochastic model describing the impact of daily rainfall depth distribution on the soil water balance. *Advances in Water Resources*, 34 (8), 1039-1048.
- Vyas, V. M. and Christakos, G. 1997. Spatiotemporal analysis and mapping of sulfate deposition data over Eastern U.S.A. *Atmospheric Environment*, 31 (21), 3623-3633.
- Wackernagel, H., 2003. *Multivariate Geostatistics: An Introduction with Applications*. 3rd ed. Berlin: Springer.
- Webster, R. and Oliver, M., 2001. *Geostatistics for environmental scientists: Statistics in practice*. Chichester: Wiley.
- Wessel, P. 2009. A general-purpose Green's function-based interpolator. *Computers & Geosciences*, 35 (6), 1247-1254.
- Whittle, P. 1954. On stationary processes in the plane. *Biometrika*, 41, 439-449.
- Yang, F.-g., Cao, S.-y., Xing-nian, L. and Yang, K.-j. 2008. Design of groundwater level monitoring network with ordinary kriging. *Journal of Hydrodynamics, Ser. B*, 20 (3), 339-346.
- Yeo, I. W. and Lee, K.-K. 2003. Analytical solution for arbitrarily located multiwells in an anisotropic homogeneous confined aquifer. *Water Resources Research*, 39 (5), TNN31-TNN35.
- Yilmaz, H. M. 2007. The effect of interpolation methods in surface definition: An experimental study. *Earth Surface Processes and Landforms*, 32 (9), 1346-1361.

- Yu, H. L., Chen, J. C., Christakos, G. and Jerrett, M. 2009. BME estimation of residential exposure to ambient PM10 and ozone at multiple time scales. *Environmental Health Perspectives*, 117 (4), 537-544.
- Yu, H. L., Yang, S. J., Yen, H. J. and Christakos, G. 2011. A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. *Stochastic Environmental Research and Risk Assessment*, 25 (4), 485-494.
- Zimmermann, B., Zehe, E., Hartmann, N. K. and Elsenbeer, H. 2008. Analyzing spatial data: An assessment of assumptions, new methods, and uncertainty using soil hydraulic data. *Water Resources Research*, 44 (10), doi:10.1029/2007WR006604.
- Žukovič, M. and Hristopulos, D. T. 2008. Environmental time series interpolation based on Spartan random processes. *Atmospheric Environment*, 42 (33), 7669-7678.
- Žukovič, M. and Hristopulos, D. T. 2009. The method of normalized correlations: A fast parameter estimation method for random processes and isotropic random fields that focuses on short-range dependence. *Technometrics*, 51 (2), 173-185.