

SPATIAL DATA ANALYSIS FOR MONITORING AND PREDICTION OF
SELECTED WATER QUALITY PARAMETERS IN RESERVOIRS: PORSUK
DAM RESERVOIR CASE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

FİRDES YENİLMEZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ENVIRONMENTAL ENGINEERING

FEBRUARY 2014

Approval of the thesis:

**SPATIAL DATA ANALYSIS FOR MONITORING AND PREDICTION OF
SELECTED WATER QUALITY PARAMETERS IN RESERVOIRS:
PORSUK DAM RESERVOIR CASE**

submitted by **FİRDES YENİLMEZ** in partial fulfillment of the requirements for
the degree of **Doctor of Philosophy in Environmental Engineering**
Department, Middle East Technical University by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Faika Dilek Sanin
Head of Department, **Environmental Engineering**

Assoc. Prof. Dr. Ayşegül Aksoy
Supervisor, **Environmental Engineering Dept., METU**

Prof. Dr. H. Şebnem Düzgün
Co-Supervisor, **Mining Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Kahraman Ünlü
Environmental Engineering Dept., METU

Assoc. Prof. Dr. Ayşegül Aksoy
Environmental Engineering Dept., METU

Prof. Dr. Ayşe Muhammetoğlu
Environmental Engineering Dept., Akdeniz University

Assoc. Prof. Dr. Elçin Kentel
Civil Engineering Dept., METU

Assist. Prof. Dr. Emre Alp
Environmental Engineering Dept., METU

Date: February 13, 2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all the material and results that are not original to this work.

Name, Last name: Firdes Yenilmez

Signature:

ABSTRACT

SPATIAL DATA ANALYSIS FOR MONITORING AND PREDICTION OF SELECTED WATER QUALITY PARAMETERS IN RESERVOIRS: PORSUK DAM RESERVOIR CASE

Yenilmez, Firdes
Ph.D., Department of Environmental Engineering
Supervisor: Assoc. Prof. Dr. Ayşegül Aksoy
Co- Supervisor: Prof. Dr. H. Şebnem DÜZGÜN

February 2014, 217 pages

In the design of a water quality monitoring network, selection of water quality sampling locations is crucial to adequately represent the water quality of the water body when high costs of analyses and field work are taken into account. In this study, a new approach was proposed to identify the representative water quality sampling locations in reservoirs and lakes using geostatistical tools for estimation of spatial distribution of selected water quality parameters. To do so, kernel density estimation (KDE) was coupled with ordinary 2-dimensional kriging (OK) in order to select the representative sampling locations in kriging of dissolved oxygen (DO) concentrations in Porsuk Dam Reservoir (PDR). Field data obtained in August 2010 were used to start the process of sampling point elimination while maintaining the spatial correlation structure of DO. KDE was used as a tool to aid in identification of the sampling locations that would be removed from the sampling network in order to decrease the total number of samples. Accordingly, several networks were generated in which sampling locations were reduced from 65 to 10 in increments of 4 or 5 points at a time based on kernel density maps. DO variograms were constructed and DO values in PDR were kriged. Performance of the networks in DO estimations were evaluated through various error metrics, standard error maps (SEM), and whether the spatial correlation structure was conserved. Results indicated that lower sampling points resulted in loss of

information in regard to spatial correlation structure in DO when more than 30 sampling points were removed from the initial 65. Representativeness of the selected network for specific conductivity (SC) was also checked and confirmed. Furthermore, potential hotspots for DO and SC were also assessed based on landuses in the vicinity of PDR. Then, efficacy of the representative sampling locations selection method was tested against the networks generated by experts. It was shown that the evaluation approach used in this study provided a better sampling network design in which the spatial correlation structure of DO was sustained.

In the second part of the study, three-dimensional (3D) kriging of DO with the 81 sampling points was performed using Stanford Geostatistical Modeling Software (SGeMS). Hence, not only the hotspots at the surface of PDR but also in deeper layers were constituted and evaluated in terms of the inlets of pollution sources. Similar hotspots were obtained both for 2D kriging and 3D kriging of DO for the dataset used in this study. Moreover, 3D distributions of DO, SC and temperature were constituted to determine the location of the thermocline layer. It was identified that the traditional approach of collecting samples from mid depths may cause incomplete characterization and evaluation of water quality since thermocline layer may not coincide with mid-depth.

Keywords: Monitoring network, kernel density estimation, kriging, Porsuk Dam Reservoir, SGeMS

ÖZ

REZERVUARLARDA SEÇİLMİŞ SU KALİTESİ PARAMETRELERİNİ İZLEME VE TAHMİN İÇİN MEKANSAL VERİ ANALİZİ: PORSUK BARAJ GÖLÜ ÖRNEĞİ

Yenilmez, Firdes
Doktora, Çevre Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Ayşegül Aksoy
Ortak Tez Yöneticisi: Prof. Dr. H. Şebnem DÜZGÜN

Şubat 2014, 217 sayfa

Su kalite izleme ağlarının oluşturulmasında, örnekleme noktalarının seçimi su kütlesindeki su kalitesinin doğru belirlenmesi açısından çok önemlidir. Bununla birlikte, analiz ve arazi çalışmalarının yüksek maliyet gerektirmesi bu seçimi daha da önemli hale getirmektedir. Bu çalışmada, seçilmiş su kalitesi parametrelerinin mekansal dağılımının jeostatistiksel yöntemler kullanılarak oluşturulması ile baraj ve göllerde temsili örnekleme noktalarının belirlenmesi için yeni bir yaklaşım geliştirilmiştir. Bu amaçla, kernel yoğunluk tahmini (KDE), iki-boyutlu kriging yöntemi ile birlikte 2010 yılı Ağustos ayında elde edilen çözünmüş oksijen (ÇO) konsantrasyonu verileri kullanılarak Porsuk Baraj Gölü temsili örnekleme noktalarını belirlemek için uygulanmıştır. KDE, örnekleme ağındaki toplam örnekleme sayısını azaltmak için uzaklaştırılacak noktaların belirlenmesi amacıyla kullanılmıştır. Böylece, her seferinde 4 veya 5 örnekleme noktası uzaklaştırılarak farklı örnekleme ağları örnekleme noktaları 65'den 10'a düşürülerek oluşturulmuştur. Örnekleme ağları için ÇO variogramları ve kriging haritaları oluşturulmuştur. Temsili örnekleme noktalarından oluşan ağ, variogram modelleri, kriging haritaları, standart hata haritaları ve verifikasyon sonuçları değerlendirilerek belirlenmiştir. Buna göre, ÇO verileri baz alındığında göldeki mekansal ilişkinin yansıtılabilmesi için 35 örnekleme noktasına ihtiyaç vardır. Seçilen örnekleme ağının, Özgül İletkenlik (Öİ) parametresi için de uygunluğu

kontrol edilmiştir. Bununla birlikte arazi kullanımı göz önüne alınarak ÇO ve Öİ parametrelerinin yüzeyde oluşturduğu kümelenmeler irdelenmiştir. Daha sonra önerilen metodun etkinliği, uzmanlar tarafından seçilen örnekleme ağlarına karşı test edilmiştir. Sonuçlar, çözünmüş oksijen parametresi mekansal korelasyonunu sürdürmesi sebebiyle bu çalışmada önerilen metodun daha iyi olduğunu göstermiştir.

Çalışmanın ikinci bölümünde, Stanford Jeostatistiksel Modelleme yazılımı (SGeMS) kullanılarak 81 noktadan elde edilen ÇO verisi için üç boyutlu kriging uygulanmıştır. Böylece, yalnızca PDR yüzeyindeki değil alt tabakalardaki kümelenmeler de krililik giriş noktaları dikkate alınarak değerlendirilmiştir. İki boyutlu ve üç boyutlu kriging uygulamaları neticesinde elde edilen kriging haritaları karşılaştırılmış ve ÇO için benzer kümeler gözlenmiştir. Son olarak çözünmüş oksijen, iletkenlik ve sıcaklık parametrelerinin üç boyutlu dağılımları termoklin tabakasının belirlenmesi için incelenmiştir. Buna göre geleneksel örnekleme çalışmalarında toplam derinliğin yarısından alınan numunelerin termoklin tabakası ile çakışmayabileceği ve su kalitesinin hatalı değerlendirilmesine sebep olabileceği görülmüştür.

Anahtar Kelimeler: İzleme ağı, Kernel Yoğunluk Tahmini, kriging, Porsuk Baraj Gölü, SGeMS

To My Beloved Mom...

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my supervisor Assoc.Prof. Dr. Ayşegül Aksoy for her valuable guidance, advice, criticism, encouragements and insight throughout the research. Furthermore, I would like to thank my co-supervisor Prof. Dr. H. Sebnem Düzgün for her encouragements, advice and guidance. I have felt fortunate due to study with them.

I would also like to thank to my committee members Prof. Dr. Ayşe Muhammetoğlu, Prof. Dr. Kahraman Ünlü and Assist. Prof. Dr. Emre Alp, Assoc. Prof. Dr. Elçin Kentel, for their encouragements, insightful comments and valuable suggestions.

I would like to extend my gratitude to the financial support from The Scientific and Technological Research Council of Turkey (TUBITAK) BİDEB without whose financial support this study would not be possible.

My special thanks are to Prof. Dr. Bülent Topkaya, Prof. Dr. Ayşe Muhammetoğlu, Prof. Dr. Habib Muhammetoğlu and Prof. Dr. Narin Ünal for their endless support, suggestions, comments and encouragements. I would also like to thank, Assoc. Prof. Dr. Adil N. Godrej, Saurav Kumar, Justin Bartlett and everybody in the Occoquan Watershed Monitoring Laboratory for their friendship help and support which made my stay in USA a pleasant experience.

I would like to thank to the all members of Department of Environmental Engineering of Middle East Technical University, especially Güldane Kalkan, Döndü Çakır, Cemalettin Akın and Bilal Özdemir.

I wish to thank to all my friends, especially ıgdem Alkılıgil Keleş, Fadime Kara Murdoch, Onur Yüzügüllü, Tolga Pilevneli, Deniz G. Tokgöz, Güray Doęan, Murat Varol, Elif Küçük, Gülcan Sarp, Pınar Bostan, Arzu Erener, İpek Koer Güler, Fatma Öztürk, Mihriban Y. Civan, Sema Yurdakul, Hande Bozkurt, Özge Can, Mehmet Hamgöl, Didem Civancık, Nihat Ataman, Mustafa Kemal Emil, Umut Cırık, Selcen Sönmez Ak, Filiz Demircioęlu Karakaş, Tolga Özbilge, Tuba Özdemir, Ebru Yıldız, Muhsine Mısırlıoęlu, and Bilgen Girgin for their support, motivation and encouragement.

Finally, I would like to express my deepest appreciation to my mother Behice Gürbüz Gençaban and my family, especially Mustafa Karaman for their endless support, understanding, encouragement and patience. My lovely mom gave me extra strength to overcome difficulties I faced not only during this study but also throughout my life. Thank God she's my mother.

TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vii
ACKNOWLEDGMENTS	x
TABLE OF CONTENTS	xii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS	xxiii

CHAPTERS

1. INTRODUCTION	1
2. LITERATURE REVIEW & THEORETICAL BACKGROUND	5
2.1 Optimization of Water Quality Sampling Locations	5
2.2 Application of Kriging in Evaluation of Water Quality Parameters within Water Bodies (Rivers, Lakes and Reservoirs)	9
2.3 Stanford Geostatistical Modeling Software (SGeMS)	13
3. STUDY SITE	17
3.1 Porsuk Dam Reservoir and Its Watershed	17
3.1.1 Topography	20
3.1.2 Landuses in the PDR Watershed	24
3.1.3 Water Quality History in the Watershed	29
3.1.4 Current Water Quality of PDR	37
4. METHODOLOGY	43
4.1 Field Work	48
4.2 Kriging	51
4.2.1 Data Transformations Used in Environmental Applications	52
4.2.2 Identifying trends in data	54
4.2.3 Variogram Analysis	56

4.2.4	Ordinary Kriging (OK)	60
4.2.5	Kernel Density Estimation	61
4.2.6	Evaluation	64
5.	RESULTS AND DISCUSSION	67
5.1	2D Kriging.....	67
5.1.1	KDE	67
5.1.2	Normality Check	75
5.1.3	Trend Analysis	81
5.1.4	Variogram Analysis	82
5.1.5	Kriging	89
5.1.5.1	Results for the Networks Selected by Experts.....	110
5.2	3D Kriging.....	113
5.2.1	Normality Check	113
5.2.2	Variogram Analysis	116
5.2.3	Kriging	117
6.	CONCLUSIONS AND RECOMMENDATIONS	125
	REFERENCES.....	129
APPENDICES		
A.	REGULATION ON THE QUALITY OF SURFACE WATER INTENDED FOR THE ABSTRACTION OF POTABLE WATER.....	147
B.	THE TURKISH WATER POLLUTION CONTROL REGULATION.....	149
C.	WATER QUALITY MEASUREMENT RESULTS	151
D.	KDE MAPS.....	157
E.	VARIOGRAMS	161
F.	RELATIONSHIP BETWEEN $\gamma(h)$ OF N00 VARIOGRAM MODEL AND $\gamma(h)$ OF NXX VARIOGRAM MODEL.....	167
G.	OK AND ERROR MAPS OF DO	173
H.	OK AND ERROR MAPS OF SC	191

I. MEASURED VS. PREDICTED GRAPHS	209
CURRICULUM VITAE	213

LIST OF TABLES

TABLES

Table 2.1 Geostatistical softwares with functionalities (modified from the list on http://www.ai-geostats.org/ & Goovaerts, 2010)	15
Table 3.1 Basic characteristics of PDR (Ministry of Environment and Forestry, 2006)	20
Table 3.2 Elevation Ranges within PDR Watershed (AKS, 2010)	21
Table 3.3 Slopes within PDR Watershed (AKS, 2010)	21
Table 3.4 Distribution of land uses in the watershed (AKS, 2010)	24
Table 3.5 The dams and ponds present in PDR Watershed (AKS, 2010)	25
Table 3.6 Industrial area groups within PDR Watershed (AKS, 2010)	26
Table 3.7 Sizes of mining areas within PDR Watershed (AKS, 2010)	29
Table 3.8 Sampling stations and coordinates	38
Table 3.9 General Trophic Classification of Lakes and Reservoirs in Relation to Phosphorus and Nitrogen ^a (Wetzel, 2001) and measured concentrations in PDR	39
Table 5.1 D and P Value for DO and SC calibration data (N00)	79
Table 5.2 DO variogram model parameters for different sampling networks	85
Table 5.3 Relationships between the DO variogram models for N00 and different networks (NXX)	87
Table 5.4 Variogram models and parameters for SC	89
Table 5.5 Validation results for OK of DO	99
Table 5.6 Validation Results for OK of SC	110
Table 5.7 Validation Results for the selections of experts	111
Table 5.8 Variogram model parameters for the networks of experts and N55	113
Table 5.9 D and P value for DO data for 3D-kriging	115
Table 5.10 Variogram Parameters for 3D DO data	116

Table A. 1 Quality Standards By Category 147

Table B. 1 Quality Criteria for Inland Water Resources Classes 149

LIST OF FIGURES

FIGURES

Figure 3.1 Watershed of PDR	18
Figure 3.2 a) Bathymetry of the PDR in 2001, b) Histogram of bathymetry (AKS, 2010)	19
Figure 3.3 Elevations within PDR Watershed (AKS, 2010)	22
Figure 3.4 Slopes within PDR Watershed (AKS, 2010)	23
Figure 3.5 Mining areas and other land uses in PDR Watershed.....	25
Figure 3.6 Sampling Locations in AKS (2010).....	38
Figure 4.1 The methodology employed in selection of sampling points	46
Figure 4.2 The methodology employed in 3D kriging.....	47
Figure 4.3 Porsuk River Inlet to PDR	49
Figure 4.4 Sampling locations in the field study (green circles: calibration data points, red circles: validation data points).....	50
Figure 4.5 An example variogram model illustrating stationary and nonstationary behavior (Houlding 1994; Mohamed and Antia, 1998)	57
Figure 4.6 Frequently Used Variogram Models for Representing Spatial Variability (a: Linear Model, b: Spherical Model, c: Exponential Model) (Houlding, 1994)	59
Figure 4.7 Kernel Density Estimation (Wilson, 2012).....	61
Figure 5.1 Application of 250 m kernel bandwidth for DO.....	69
Figure 5.2 Application of 500 m kernel bandwidth for DO.....	70
Figure 5.3 Application of 750 m kernel bandwidth for DO.....	70
Figure 5.4 Application of 1000 m kernel bandwidth for DO.....	71
Figure 5.5 Application of 1250 m kernel bandwidth for DO.....	71
Figure 5.6 Application of 250 m kernel bandwidth for SC.....	72
Figure 5.7 Application of 500 m kernel bandwidth for SC.....	72
Figure 5.8 Application of 750 m kernel bandwidth for SC.....	73
Figure 5.9 Application of 1000 m kernel bandwidth for SC.....	73

Figure 5.10 Application of 1250 m kernel bandwidth for SC	74
Figure 5.11 Distance based KDE with 1000 m kernel bandwidth	74
Figure 5.12 Histogram and Q-Q plot of DO for N00	76
Figure 5.13 Histogram and Q-Q plot of SC for N00	77
Figure 5.14 Q-Q Plots of SC Data after transformations were applied (a: Logarithmic, b: ArcSin, c: ArcTan)	79
Figure 5.15 Trend Analysis for DO	82
Figure 5.16 Trend Analysis for SC	82
Figure 5.17 Variogram Fitting in R Cran	83
Figure 5.18 DO variograms for all sampling networks	84
Figure 5.19 SC variograms for all sampling networks	88
Figure 5.20 OK of DO for N00 (calibration data set).....	91
Figure 5.21 OK of DO for N04	93
Figure 5.22 SEM of DO for N00	94
Figure 5.23 SEM of DO for N04	95
Figure 5.24 Difference of SEMs of DO for N04 and N00	96
Figure 5.25 Measured versus predicted DO at validation points for N00 and N04	97
Figure 5.26 Measured versus predicted DO at validation points for N00 after selected sampling locations (red circles in Figure 5.25) were removed from the validation dataset	98
Figure 5.27 ASE values for OK of DO for different sampling point numbers	100
Figure 5.28 RMSE values for OK of DO for different sampling point numbers	101
Figure 5.29 RMSSE values for OK of DO for different sampling point numbers	101
Figure 5.30 Relationship between SC and salinity	103
Figure 5.31 OK of SC for N00	104
Figure 5.32 OK of SC for N30	105
Figure 5.33 SEM of SC for N00	106

Figure 5.34 SEM of SC for N30	107
Figure 5.35 Difference of SEMs of SC for N30 and N00.....	108
Figure 5.36 Measured versus predicted SC for N00 and N30	109
Figure 5.37 Measured versus predicted DO at validation points for the networks of experts and N55.....	112
Figure 5.38 Histogram and Q-Q plot for 3D DO data	114
Figure 5.39 The spherical variogram model fitted to the experimental variogram of 3D DO data.....	117
Figure 5.40 X-Z View of 3D temperature data	118
Figure 5.41 X-Z View of 3D DO data	119
Figure 5.42 X-Z View of 3D SC data	119
Figure 5.43 Grid for 3D kriging	121
Figure 5.44 3D Kriging Maps of DO at different depths	121
Figure 5.45 Measured versus predicted DO obtained from 3D OK.....	123
Figure C. 1 Average COD distribution within PDR (AKS, 2010)	151
Figure C. 2 Average BOD ₅ distribution within PDR (AKS, 2010)	152
Figure C. 3 Average DO Saturation distribution within PDR (AKS, 2010).....	152
Figure C. 4 Average DO distribution within PDR (AKS, 2010)	153
Figure C. 5 Average Phosphates distribution within PDR (AKS, 2010).....	153
Figure C. 6 Average Nitrate distribution within PDR (AKS, 2010).....	154
Figure C. 7 Average Kjeldahl Nitrogen distribution within PDR (AKS, 2010).....	154
Figure C. 8 Average Ammonia Nitrogen distribution within PDR (AKS, 2010).....	155
Figure C. 9 Average Chlorophyll-a distribution within PDR (AKS, 2010)	155
Figure C. 10 Average Conductivity distribution within PDR (AKS, 2010).....	156
Figure D. 1 KDE map of DO for each network	157
Figure E. 1 DO variogram for each network	161
Figure E. 2 SC variogram for each network	163
Figure F. 1 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N04 variogram model	167

Figure F. 2 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N10 variogram model.....	168
Figure F. 3 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N15 variogram model.....	168
Figure F. 4 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N20 variogram model.....	169
Figure F. 5 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N25 variogram model.....	169
Figure F. 6 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N30 variogram model.....	170
Figure F. 7 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N35 variogram model.....	170
Figure F. 8 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N40 variogram model.....	171
Figure F. 9 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N45 variogram model.....	171
Figure F. 10 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N50 variogram model.....	172
Figure F. 11 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N55 variogram model.....	172
Figure G. 1 OK of DO using N00 data set	173
Figure G. 2 SEM of DO using N00 data set.....	174
Figure G. 3 OK of DO using N04 data set	174
Figure G. 4 SEM of DO using N04 data set.....	175
Figure G. 5 OK of DO using N10 data set	175
Figure G. 6 SEM of DO using N10 data set.....	176
Figure G. 7 OK of DO using N15 data set	176
Figure G. 8 SEM of DO using N15 data set.....	177
Figure G. 9 OK of DO using N20 data set	177
Figure G. 10 SEM of DO using N20 data set.....	178
Figure G. 11 OK of DO using N25 data set	178

Figure G. 12 SEM of DO using N25 data set	179
Figure G. 13 OK of DO using N30 data set.....	179
Figure G. 14 SEM of DO using N30 data set	180
Figure G. 15 OK of DO using N35 data set.....	180
Figure G. 16 SEM of DO using N35 data set	181
Figure G. 17 OK of DO using N40 data set.....	181
Figure G. 18 SEM of DO using N40 data set	182
Figure G. 19 OK of DO using N45 data set.....	182
Figure G. 20 SEM of DO using N45 data set	183
Figure G. 21 OK of DO using N50 data set.....	183
Figure G. 22 SEM of DO using N50 data set	184
Figure G. 23 OK of DO using N55 data set.....	184
Figure G. 24 SEM of DO using N55 data set	185
Figure G. 25 Difference of SEMs of DO for N04 and N00.....	185
Figure G. 26 Difference of SEMs of DO for N10 and N00.....	186
Figure G. 27 Difference of SEMs of DO for N15 and N00.....	186
Figure G. 28 Difference of SEMs of DO for N20 and N00.....	187
Figure G. 29 Difference of SEMs of DO for N25 and N00.....	187
Figure G. 30 Difference of SEMs of DO for N30 and N00.....	188
Figure G. 31 Difference of SEMs of DO for N35 and N00.....	188
Figure G. 32 Difference of SEMs of DO for N40 and N00.....	189
Figure G. 33 Difference of SEMs of DO for N45 and N00.....	189
Figure G. 34 Difference of SEMs of DO for N50 and N00.....	190
Figure G. 35 Difference of SEMs of DO for N55 and N00.....	190
Figure H. 1 OK of SC using N00 data set.....	191
Figure H. 2 SEM of SC using N00 data set	192
Figure H. 3 OK of SC using N04 data set.....	192
Figure H. 4 SEM of SC using N04 data set	193
Figure H. 5 OK of SC using N10 data set.....	193
Figure H. 6 SEM of SC using N10 data set	194
Figure H. 7 OK of SC using N15 data set.....	194

Figure H. 8 SEM of SC using N15 data set.....	195
Figure H. 9 OK of SC using N20 data set	195
Figure H. 10 SEM of SC using N20 data set.....	196
Figure H. 11 OK of SC using N25 data set	196
Figure H. 12 SEM of SC using N25 data set.....	197
Figure H. 13 OK of SC using N30 data set	197
Figure H. 14 SEM of SC using N30 data set.....	198
Figure H. 15 OK of SC using N35 data set	198
Figure H. 16 SEM of SC using N35 data set.....	199
Figure H. 17 OK of SC using N40 data set	199
Figure H. 18 SEM of SC using N40 data set.....	200
Figure H. 19 OK of SC using N45 data set	200
Figure H. 20 SEM of SC using N45 data set.....	201
Figure H. 21 OK of SC using N50 data set	201
Figure H. 22 SEM of SC using N50 data set.....	202
Figure H. 23 OK of SC using N55 data set	202
Figure H. 24 SEM of SC using N55 data set.....	203
Figure H. 25 Difference of SEMs of SC for N04 and N00	203
Figure H. 26 Difference of SEMs of SC for N10 and N00	204
Figure H. 27 Difference of SEMs of SC for N15 and N00	204
Figure H. 28 Difference of SEMs of SC for N20 and N00	205
Figure H. 29 Difference of SEMs of SC for N25 and N00	205
Figure H. 30 Difference of SEMs of SC for N30 and N00	206
Figure H. 31 Difference of SEMs of SC for N35 and N00	206
Figure H. 32 Difference of SEMs of SC for N40 and N00	207
Figure H. 33 Difference of SEMs of SC for N45 and N00	207
Figure H. 34 Difference of SEMs of SC for N50 and N00	208
Figure H. 35 Difference of SEMs of SC for N55 and N00	208
Figure I. 1 Measured DO vs. Predicted DO graph obtained from OK.....	209
Figure I. 2 Measured SC vs. Predicted SC graph obtained from OK.....	210

LIST OF ABBREVIATIONS

ADL	: Allowable Discharge Limits
ArcSWAT	: Geographic Information System Integrated Soil Water Assessment Tool
ASE	: Average Standard Error
BAT	: Best Achievable Technologies
BOD	: Biochemical Oxygen Demand
C	: Partial Sill
C ₀	: Nugget
C ₀ +C	: Sill
CA	: Cluster Analysis
CBOD	: Carbonaceous Biochemical Oxygen Demand
CEUC	: Coverage Elimination Uniform Cost
Chl-a	: Phytoplankton Pigment
Cl	: Chlorine
COD	: Chemical Oxygen Demand
D	: Kolmogorov-Smirnov Statistic
DA	: Discriminant Analysis
DEM	: Digital Elevation Model
DO	: Dissolved Oxygen
EC	: Electrical Conductivity
FA	: Factor Analysis
GA	: Genetic Algorithm
GIS	: Geographic Information System
GPS	: Global Positioning Systems
HCB	: Hexachlorobenzene
ITU	: Istanbul Technical University
KDE	: Kernel Density Estimation
KFI	: Kütahya Fertilizer Industry
KWWTP	: Kütahya Wastewater Treatment Plant
LRB	: Limpopo River Basin
MBAS	: Surfactant
ME	: Mean Error
MoEF	: Ministry of Environment and Forestry
MRI	: Marmara Scientific and Industrial Research Institute
MSE	: Mean Standardized Error
NH ₃ -N	: Ammonia-nitrogen
NO ₂ -N	: Nitrite-nitrogen
NO ₃ -N	: Nitrate-nitrogen
OECD	: Organization of Economic Co-Operation and Development
OK	: Ordinary Kriging
PCA	: Principal Components Analysis
PCBs	: Polychlorinated Biphenyls

PDR	: Porsuk Dam Reservoir
PFA	: Principal Factor Analysis
PO ₄ -P	: Phosphate-phosphorus
PPS	: Potential Pollution Score
Q	: Flow Rate
QUAL2E	: Enhanced Stream Water Quality Model
R ²	: Coefficient of Determination
RMSE	: Root Mean Square Error to the Standard Deviation of Measured Data
RMSSE	: Root Mean Square Standardized Error
SC	: Specific Conductivity
SEM	: Standard Error Map
SGeMS	: Stanford Geostatistical Modeling Software
SHW	: State Hydraulic Works
SQI	: Sediment Quality Index
SS	: Suspended Solids
STPCSV	: Spatio-temporal Point Cumulative Semivariogram
SWAT	: Soil Water Assessment Tool
T	: Temperature
TDS	: Total Dissolved Solids
TMDL	: Total Maximum Daily Load
TN	: Total Nitrogen
TP	: Total Phosphorus
TSS	: Total Suspended Solids
TUBITAK	: The Scientific and Technological Research Council of Turkey
UC	: Uniform Cost
UNEP	: United Nations Environment Programme
US	: United States
USEPA	: United States Environmental Protection Agency
UTM	: Universal Transverse Mercator
WHO	: World Health Organization
XPR	: X Coordinate in the Defined Projection
YPR	: Y Coordinate in the Defined Projection

CHAPTER 1

INTRODUCTION

Surface waters play a vital role as drinking water supply, transportation media, and for recreation purposes (Nam, 2008). However, the water quality of surface waters is under stress due to domestic and industrial wastewater discharges, agricultural activities, mining activities, etc. Monitoring is important to assess the impacts of these activities.

The design of water quality monitoring networks that constitute of robust and representative sampling locations is essential for better water quality characterization and management (Do et al., 2012). The method used in the design of the monitoring network should be flexible permitting modifications at later stages and recognize the economic constraints (Strobl and Robillard, 2008). Geostatistical tools are beneficial in identifying the representative sampling locations, which can be crucial given the high costs of installment and operation (Chehata et al., 2007).

Although geostatistical applications were originated and applied in mining and petroleum exploration industries, they are now routinely employed in environmental studies (Chehata et al., 2007). Many water quality monitoring networks for surface freshwaters have been randomly designed with inconsistent and illogical design strategies (Strobl and Robillard, 2008). Researches related with this topic point out the necessity of developing water quality monitoring network design procedures which can be generally regarded as valid by most designers (Sanders et al., 1983; Strobl and Robillard, 2008; Nam, 2008). In many water quality network designs, it is aimed to reduce the number of sampling location to reduce monitoring costs (Nam, 2008; Strobl et al., 2006a; Kao et al., 2012; Hedger et al., 2001). Yet, water quality data exhibit spatial correlations. As

a result, care should also be given to sustain the original spatial correlation structure while reducing the number of sampling points to have a representative network.

In lakes and reservoirs, there are several factors like wind action, shape of the lake/reservoir, feeder streams and effluents that cause water quality to vary spatially and temporally. In the guide to design and implement freshwater quality studies and monitoring programs by the United Nations Environment Programme (UNEP) and the World Health Organization (WHO), it is stated that if there is good horizontal mixing, a single station near the center or at the deepest part of the lake will be sufficient to monitor long-term trends. If the lake is large with many narrow bays or contains several deep basins, the number of sampling stations can be calculated from the \log_{10} of the lake area (in km^2) (UNEP/WHO, 1996). Furthermore, it is necessary to take more than one sample to describe water quality in stratified lakes. According to the Methods of Sampling and Analysis guide that supplements the Turkish Water Pollution Control Regulation, minimum five sampling locations should be selected by taking water inlets and outlets, coastal activities, locations of pollution sources, hydrodynamic properties into account to characterize the water quality in reservoirs and lakes. According to the regulation, a network is obtained by dividing the surface of the lake/reservoir into grids in different seasons. Samples are collected at various depths of the sampling locations that are at the corner points of the network. The results of this study were used to determine routine sampling stations.

The European Union Water Framework Directive (2000/60/EC) aims for establishment of a framework for the protection of inland surface waters, coastal waters and groundwater. In the directive, it is mentioned that monitoring sites should be selected in sufficient amounts to assess the magnitude and impact of point sources, diffuse sources and hydromorphological pressures in designing operational monitoring. Artiola et al. (2004) mentioned that costs associated with sampling and analysis, accessibility and sampling time should be taken into

account in the selection of sample location and number of samples. Furthermore, they emphasized the importance of selection of sampling locations randomly. Moreover, the number of samples should be such that the maximum-accepted level of error in the results for statistically-based monitoring plans should be acceptable. In the study, it was also highlighted that samples should be taken throughout the depth in enclosed bodies of water such as lakes and lagoons, especially if thermocline is present. Burden et al. (2002) pointed out the dependence of the number and location of potential sites on purpose of the monitoring program. For example, sampling locations should be at places where nutrients are likely to enter the water body to monitor the impact of point sources. Furthermore, representative sampling sites should provide appropriate spatial information, away from boundary areas (e.g., confluence of streams or rivers), convenient to reach, and safely accessible. Yet, no specific method is declared to identify representative sampling locations. In overall, although the mentioned regulations and guidelines include frameworks for the determination of sampling locations in surfacewaters, it is necessary to propose a new approach to design a water quality monitoring network with representative sampling locations in reservoirs and lakes in which water quality can be governed by thermal and density stratification as well.

The main objective of this study is to propose a new approach to identify the representative water quality sampling locations especially to obtain the spatial water quality distribution through kriging with smaller number of sampling points. Porsuk Dam Reservoir (PDR) is the study site. PDR is a large reservoir located within the borders of Kütahya and Eskişehir provinces. The water quality in PDR is under stress due to point and non-point pollution sources present in the borders of both provinces. The water quality of PDR is important because PDR supplies drinking water for Eskişehir. The current status of the water quality in PDR is in the hypereutrophic-eutrophic range (AKS, 2010) and urgent precautions are required to improve the water quality.

In this study, the combination of Kernel density estimation (KDE) and ordinary 2-dimensional kriging (OK) were used to determine the representative sampling locations using the field data obtained from 81 sampling locations in August 2010. Although there have been many studies to optimize monitoring networks, this study is the first application where KDE and kriging are combined to identify the representative sampling locations in reservoirs while maintaining the spatial correlation structure of given water quality parameters, which is crucial for accurate interpolation in kriging. Available studies do not focus on whether the selected sampling locations are sufficient in reflecting the spatial correlation structure of given water quality parameters. If the spatial correlation structure of a given water quality parameter is not sustained, the design of water quality monitoring network that consist of representative sampling locations for the selected parameter may not be possible. This may result in incorrect estimation of the spatial distribution of water quality parameters which may lead to inappropriate management approaches. In the study, potential hotspots for dissolved oxygen (DO) and specific conductivity (SC) are also evaluated based on landuses in the vicinity of PDR. Moreover, three-dimensional (3D) distributions of DO, SC and temperature were constituted to determine the thermocline layer. In most of the water quality modeling and water quality sampling studies, the depth to thermocline is assumed as fixed throughout the reservoirs and lakes. This situation may cause incorrect characterization and evaluation due to insufficient sampling design. Stanford Geostatistical Modeling Software (SGeMS) was used to perform 3D kriging of DO and compare the hotspots observed at the surface and in deeper layers.

CHAPTER 2

LITERATURE REVIEW & THEORETICAL BACKGROUND

2.1 Optimization of Water Quality Sampling Locations

It is important to develop a solid strategy to design better monitoring networks in surface waters. Although many researches have been conducted to optimize monitoring networks in river systems, researches on how to design and optimize a monitoring network in reservoirs are limited (Nam, 2008).

Do et al. (2012) identified the representative river water quality sampling locations using a water monitoring network design procedure which combined river mixing length, human activities, and geographic information systems (GIS). The importance of each sampling point was determined by a potential pollution score (PPS). Lo et al. (1996) designed the water quality monitoring network of the Keelung River in Northern Taiwan. A steady-state water quality model was used to simulate the water quality of the river. The required monitoring stations were selected using kriging. While 15 stations were required for the average flow condition, 21 stations were needed for the medium/low flow condition. Icaga (2005) used a genetic algorithm (GA) for the optimization of water quality monitoring network of the Gediz River Basin in Turkey. Station combination scores were determined by taking drainage area, population, irrigation area, sample number, observation period and quality observations into account. Then, GA was applied to select the best station combination. The number of stations required was decreased from 33 to 14 by this optimization technique.

Karamouz et al. (2009a) designed a river water quality monitoring network to determine sampling frequencies and location of water quality monitoring stations for the Karoon River in Iran. Firstly, a combination of kriging method and

analytical hierarchy process was used to determine the optimal location of the monitoring stations. In this method, estimation error was used as a criterion to locate the station. Secondly, a combination of GA and a river water quality simulation model was used. According to the results of models, 35 stations were proposed. Chilundo et al. (2008) designed a water quality monitoring network for the Limpopo River Basin (LRB) in Southern Africa. They examined physico-chemical, biological and microbiological characteristics at 23 sampling locations within the basin between November 2006 and January 2007. 16 sampling locations were selected for future sampling by taking temporal changes in pollutant concentrations, existing infrastructures, representativeness of the locations, identified sources of pollution, main water intakes, compliance with water quality standards, and number of contributing tributaries in the basin into account.

Beveridge et al. (2012) optimized water quality monitoring network in Lake Winnipeg with a surface area of 23750 km² located in Canada. They used water isotopes ($\delta^2\text{H}$, $\delta^{18}\text{O}$) data collected in 240 stations in September-October 2009. Two techniques were used to assess the relationships of redundancies between neighboring stations. While Local Moran's I technique was conducted to determine the clusters of stations that were similar or different, kriging technique was used to evaluate the suitability of the sampled network configuration. Good correlations were obtained between observed and predicted values. Cross-validation was applied to evaluate kriging model performance. According to results, a large number of stations were identified as redundant. Stations that were statistically important or redundant could be determined by the combination of techniques proposed in the study. Importance of the evaluation of information provided by an individual station together with clusters of stations was emphasized.

Kazi et al. (2009) applied principal components analysis (PCA) and cluster analysis (CA) to assess the variations in surface water quality of the Manchar

Lake in Pakistan. The lake had a surface area of 233 km². Five sampling stations were grouped into three clusters based on similar water quality characteristics. The need for development of a new optimal sampling design by reducing the number of sampling locations and associated cost was suggested.

Jiménez et al. (2005) developed a methodology to design data collection networks in lakes and reservoirs. Developed methodology was applied to Porce II reservoir in Colombia. Firstly, ELCOM model was used to simulate the hydrodynamics and water quality tracers to overcome the lack of field data. Then, the reservoir was divided into five subdomains based on temperature distributions and one way ANOVA² indices. After that kriging was applied to interpolate and get estimates from the available monitoring networks. Finally, optimization process was performed to determine the optimal monitoring network considering accuracy versus cost.

Varol et al. (2012) examined the spatial and temporal variations of the water qualities in Kralkızı (KDR), Dicle (DDR) and Batman (BDR) dam reservoirs in 2008-2009. The surface areas of Kralkızı, Dicle and Batman dam reservoirs were 57.5 km², 24 km², 49.25 km², respectively. They used CA, PCA, factor analysis (FA) and discriminant analysis (DA). The sampling period (12 months) was grouped into two clusters as wet and dry seasons based on CA. Furthermore, 10 monitoring sites were converted into four clusters based on similarities in the water quality characteristics. The outcome of this study was optimal future monitoring strategies with decreased monitoring frequencies, number of sampling stations and corresponding costs.

Thornton et al. (1982) designed a water quality monitoring program for DeGray Lake with a surface area of 53.4 km² located in Southern Arkansas. Sampling program was constituted based on horizontal and vertical water quality gradients by transect and depth, respectively. Kao et al. (2012) used the uniform cost (UC) and coverage elimination uniform cost (CEUC) models to design an optimum

water quality monitoring network with minimal cost for the Derchi Reservoir in Taiwan which would also help to trace the source of a pollution event. Nam (2008) developed an optimization procedure to determine the best locations and paths to monitor surface water in a two dimensional space. Hydrodynamics and contaminant transport were simulated via a numerical approach. Then, multi-objective GA was proposed to find the optimized locations and paths.

In addition to the studies mentioned above, Panda et al. (2006) applied factor and cluster analysis for effective management of monitoring network in Mahanadi River (India). Harmancioglu and Alpaslan (1992) applied the entropy principle to evaluate the network efficiency and cost-effectiveness of Porsuk River. Ouyang (2005) used PCA and PFA for identification of important surface water quality monitoring stations and parameters in St. Johns River in Florida, USA. Ning and Chang (2004) introduced an optimal expansion strategy of water quality monitoring stations in a river system by fuzzy optimization procedure. Karamouz et al. (2009b) designed on-line river water quality monitoring systems for the Karoon River (Iran) using the entropy theory. Dixon et al. (1999) presented a method to optimize the selection of river sampling sites using GIS, graph theory and simulated annealing algorithm. Park et al. (2006) designed water quality monitoring network in Nakdong River system (Korea) using genetic algorithm. Strobl et al. (2006a, 2006b) proposed a water quality monitoring network design methodology for the selection of critical sampling points using GIS, hydrologic simulation model and fuzzy logic. Telci et al. (2009) proposed a model for the optimal design of monitoring networks in river systems. Erechtkhoukova et al. (2009) used optimization algorithms for the improvement of temporal monitoring design at an existing sampling site.

When the studies and regulations related to the design of monitoring networks in water bodies were examined, it was realized that there was a gap in literature and regulations about the selection of representative sampling locations in lakes and reservoirs. Furthermore, previous studies did not take the spatial correlation of

water quality parameters in lakes and reservoirs into account. This situation motivated this study to be carried out.

2.2 Application of Kriging in Evaluation of Water Quality Parameters within Water Bodies (Rivers, Lakes and Reservoirs)

While there are several studies on the application of kriging on soil samples (Goovaerts, 1998; Goovaerts, 1999; Simbahan et al., 2006), oil reservoirs (Rhea et al., 1994; Jian and Fanhua, 2009; Aggoun et al., 2006), mining sites (Bastante et al., 2008; Ertunc et al., 2013; Bastante et al., 2005), the studies related with the application of kriging technique to evaluate the spatial distribution of water quality parameters in water bodies are sparse. Some applications for various purposes are summarized below.

Little et al. (1997) assessed the accuracy of kriging method to predict water quality variable concentrations in Murrells Inlet estuary in South Carolina. Different simulations were applied using the combinations of distance metrics (Euclidean versus in-water), semivariogram types (spherical versus linear) and model trend components (distance to the inlet mouth; without versus with). Results indicated improvement when in-water distances were used together with a model trend component. The prediction accuracy was not affected significantly by the semivariogram type. They mentioned that it is worth to extend the studies related with the integration of GIS-based network analysis with kriging using in-water distances.

Ouyang et al. (2003) evaluated the spatial distribution of total sediment Hg in the Cedar-Ortega Rivers watershed in north-east Florida using 3D kriging analysis in conjunction with a geostatistical model named ISATIS and assessed the potential risk of Hg to aquatic life based on the Florida Sediment Assessment Guidelines. Mercury data was collected from three sampling depth intervals at 58 locations along the rivers between February 1998 and February 1999. After preliminary

data analysis and constitution of a variogram model, 3D kriging analysis was performed. Results showed that rivers were contaminated with Hg to a sediment depth of 1.0 m. There was a sharp decrease in Hg between sediment depths of 1.0 to 2.0 m. This situation pointed out recent and ongoing Hg contamination in the rivers. 3D kriging estimation plot of Hg concentrations in Cedar River displayed that the maximum depth with Hg concentrations above the limit values was about 1.5 m. Furthermore, evenly distribution of Hg in rivers was the evidence of Hg contamination from multiple sources.

Chehata et al. (2007) adapted, customized and tested data interpretation and visualization software tools with geostatistical capabilities to support the Chesapeake Bay Program to improve water-quality modeling protocols. Following the constitution of a 3D grid, data structure was analyzed using experimental semivariograms. Then, anisotropy features were extracted, 3D ordinary kriging model was fit, 3D maps for water quality data, such as DO and salinity, were constituted at fixed stations. Mapped data in 3D was used as an input to the Bay water quality model which simulated future water quality in the Bay. Furthermore, the interpolated 3D data was separated into three zones to evaluate the upper and lower pycnocline boundary surfaces for a better characterization of the pycnocline layer.

Jakubek and Forsythe (2004) calculated sediment quality index (SQI) for 70 sediment core-sampling locations in Lake Ontario in Canada. Prediction maps for polychlorinated biphenyls (PCBs), mercury, lead, and hexachlorobenzene (HCB) and SQI were created by ordinary kriging technique. Furthermore, cross-validation was performed to evaluate the accuracy of results. While the most successful surfaces were created for mercury and lead, the most inaccurate results were obtained for PCBs. The constituted maps were helpful to explain how stream loading and land use practices affected the distribution of contaminated sediment in Lake Ontario.

Barabas et al. (2001) used geostatistical models to quantify the uncertainty in 2,3,7,8-tetrachlorodibenzo-p-dioxin concentrations in the sediment of Passaic River in New Jersey. After the analysis of 3D variograms and application of indicator kriging, cross-validation was performed to evaluate the results. It was mentioned that results were helpful in decision-making, delineation of contaminated areas and additional sampling needs.

Rathbun (1998) focused on spatial modeling in irregularly shaped regions and applied kriging in Charleston Harbor, an estuary on the coast of South Carolina, USA. It was mentioned that Euclidean distance may not be an appropriate distance metric for spatial analyses of estuaries due to their irregularly shaped non-convex regions. Some points between two sampling locations may lie on land rather than in water. Kriging model was applied to constitute the prediction maps of DO and salinity under both Euclidean and water distance metrics. When the cross-validation results for kriging predictions were compared, it was seen that all models provided acceptable predictions for DO and salinity. Furthermore, some modifications of co-kriging model were suggested to obtain a prediction surface for the joint distribution of DO, salinity and temperature.

Forsythe et al. (2004) investigated mercury and lead concentrations in the sediments of Lake Ontario and Lake Erie. It was observed that three of the four developed models for the 1997-1998 data were statistically valid. Although there had been an overall reduction in mercury and lead concentrations within time, the concentrations still exceeded Canadian sediment quality limit values in some areas. In the study, these areas were evaluated with the landuse and historical data.

Külahcı and Şen (2009) developed a methodology to model the migration of ^{210}Pb radioisotope in Keban Dam, Turkey. They applied spatio-temporal point cumulative semivariograms (STPCSV) using the data measured at 44 stations. The radius of influence maps for a set of desired hours were constituted using kriging method based on STPCSVs.

Büttner et al. (1998) examined the horizontal spatial distributions of 14 parameters measured in 66 sediment samples collected in ML111 which is an acidic mine lake in Germany. While PCA was used to investigate the patterns and similarities between concentrations of different heavy metals, ordinary kriging method was used to evaluate the spatial distribution of the parameters in the sediment. According to results, 66 samples were not enough to decrease the spatial variance distributions to suitable levels.

Villard et al. (2010) used kriging method and conditional stochastic simulations to analyze, map and simulate 3D patterns of temperature, DO and nitrate in the Geneva Lake located in Switzerland and France for the period from 1954 to 2008. Because kriging technique was applied on a long term data, the weighted average of ten variograms was used for the study period. 3D maps were constructed using the data at several depths. Then, spatial and temporal distributions of the parameters were investigated. The movements of cool and warm waters in the lake with respect to time and space were examined. Results showed no global warming effects in the lake.

Hedger et al. (2001) optimized a sampling strategy based on the spatial distribution of chlorophyll-a in two British lakes (Loch Awe and Loch Ness) using geostatistical techniques with remote sensing. They mentioned that different sampling regimes will be required for different lakes and for different times because the spatial variation was specific to lake boundary conditions such as morphometry or wind regime. More accurate estimates with smaller estimation variances were obtained by systematic sampling rather than random sampling.

Zhang et al. (2011) used ordinary kriging to assess the spatial distribution of water quality parameters of Wuliangsuhai Lake in the north of China. The levels of water pollution were investigated. They stated that the eutrophication of Wuliangsuhai Lake was in a severe state and results of this study were significant for the management of the lake.

Copertino et al. (1998) evaluated 3D spatial distributions of temperature, DO, electrical conductivity, pH and redox potential in the Camastra Reservoir in Italy using kriging technique. Results showed the existence of hypolimnetic anoxic phenomena. When temperature and redox potential contour maps were analyzed, a density increase was observed around the spilling tower. This showed that this area in the reservoir was important for environmental problems and evolution of anoxic zones.

Guillemette et al. (2011) used ordinary kriging to evaluate water temperature characteristics at unsampled locations in the Ste-Marguerite River, Canada. Kienel and Kumke (2002) evaluated the spatial variation of diatom assemblages from surface sediments in Lake Lama, Siberia using a combined approach of ordination and geostatistics.

As it was mentioned before KDE and 2D OK were used in combination to determine the representative sampling locations in this study. Neither of the studies mentioned above used the combination of KDE and OK nor did these studies take into account spatial correlation of water quality parameters in lakes and reservoirs.

2.3 Stanford Geostatistical Modeling Software (SGeMS)

Although geostatistical tools were originated in mining industry in late 1950s and early 1960s, it is now routinely employed in many disciplines including environmental sciences (i.e. characterization of contaminated sediments, remote sensing, estimation of fish abundance), meteorology (i.e. spatial and temporal distribution of rainfall and temperature), hydrology (i.e. modeling of subsurface hydraulic conductivity), ecology (i.e. characterization of population dynamics), agriculture (i.e. evaluation of soil properties and crop yields), and health (i.e. patterns of diseases, exceeding of limit values, exposure to pollutants). A lot of geostatistical softwares with different price, operating systems, user-friendliness,

functionalities, graphical and visualization capabilities have been developed due to the increasing popularity of geostatistics (Goovaerts, 2010).

The selection of a geostatistical software should be performed based on the following considerations (Goovaerts, 2010):

- (i) Necessity of access to the source code to enable improvement of in algorithms
- (ii) Characteristics of the data (2D or 3D, spatial, temporal, etc.)
- (i) Type of analysis to be implemented (i.e. description of the spatial pattern, prediction at unsampled locations, incorporation of secondary information, modeling of local or spatial uncertainty).
- (ii) Level of expertise of the user in geostatistics

When available geostatistical softwares were examined (Table 2.1) according to the criteria mentioned above, it was decided to use the Stanford Geostatistical Modeling Software (SGeMS) to conduct the 3D geostatistical analysis in this study. SGeMS supports a plug-in mechanism to let the addition of new geostatistical algorithms or supports (Remy et al., 2009). The software offers a graphical user interface allowing interactive variogram modeling which measures dissimilarity as a function of separation distance and direction and facilitates the visualization of data and results in up to 3D. However, basic GIS capabilities, such as data queries or linked windows, are not available in the software. SGeMS can compute variograms in three directions. The software allows the specification of user-defined interpolation grids. It includes methods to evaluate the uncertainty about unsampled values and probability of exceeding critical values. It has also modules to post-process the set of realizations, create maps of averaged simulated values, compute probability of exceeding critical thresholds or measures of differences among realizations (Goovaerts, 2010).

Table 2.1 Geostatistical softwares with functionalities (modified from the list on <http://www.ai-geostats.org/> & Goovaerts, 2010)

Name	Code	Cost ^a	Reference	Data	V	K	CK	IK	MG	S	G	O
Agromet	C++	F	Bogaert et al. (1995)	2D	X	X	X					
AUTO-IK	Fortran	F	Goovaerts (2009)	2D	X			X				
BMELib	Matlab	F	Christakos et al. (2002)	3D, ST	X	X	X			X		
COSIM	Fortran	F	ai-geostats website	2D						X		
EVS (C-Tech)		H	C Tech Development Corporation	3D	X	X		X			X	
Explostat		F	www.explostat.nl/	2D	X	X	X				X	X
E(Z)-Kriging		F	ai-geostats website	2D	X	X						
Geopack		F	EPA website	2D	X	X	X					
GeoDa		F	https://geodacenter.asu.edu/software	2D								X
GeoXP 1.3		F	http://cran.r-project.org/web/packages/GeoXp/index.html		X							
GMT	C	F	http://gmt.soest.hawaii.edu/	2D								X
GRNN		F	www.unil.ch/igar/page48171_en.html	2D								X
ISIM	C	F	ai-geostats website	3D						X		
IV		F	http://alghalandis.com/?page_id=358	2D	X							
Kriging	C	F	www.nbb.comell.edu/neurobio/land/OldStudentProjects/cs490-94to95/clang/kriging.html#Program	2D		X						
SAGA	C++	F	www.saga-gis.org/en/index.html	2D	X	X					X	X
SGS	C	F	www.ftonn.de/Software/sgs/index.html	2D		X				X		
SpatDesign	Matlab, Octave	F	ai-geostats website	3D	X	X						X
Spherekit	C	F	www.ncgia.ucsb.edu/pubs/spherekit/	2D		X						X
Surfit	C++	F	http://surfit.sourceforge.net/index.html									X
GCOSIM3D/ISIM3D	C	F	Gomez-Hernandez and Srivastava (1990)	3D						X		
Genstat		F, L	Payne et al. (2008)	3D	X	X	X					
GEO-EAS	Fortran	F	Englund and Sparks (1988)	2D	X	X						
GeoR	R	F	Ribeiro and Diggle (2001)	2D	X	X				X		
Geostat Analyst		H	Extension for ArcGIS	2D	X	X	X	X	X	X	X	
Geostatistical Toolbox		F	Froidevaux (1990)	3D	X	X	X					
Geostokos Toolkit		H	ai-geostats website	3D	X	X	X	X	X	X		
GS+		M	Robertson (2008)	2D	X	X	X			X		
GSLIB	Fortran	F	Deutsch and Journel (1998)	3D	X	X	X	X	X	X		
Gstat	C, R	F	Pebesma and Wesseling (1998)	3D	X	X	X			X		
ISATIS (Geovariances)		H	www.geovariances.com	3D	X	X	X	X	X	X	X	
MGstat	Matlab	F	ai-geostats website	3D, ST	X	X						
SADA (UT Knoxville)		F	Spatial analysis and decision assistance	3D	X	X		X			X	X
SAGE 2001		M	Isaaks (1999)	3D	X							
SAS/STAT		H	SAS Institute Inc. (1989)	2D	X	X						
S-GeMS	C++	F	Remy et al. (2008)	3D	X	X	X	X	X	X		
SPRING		F	Camara et al. (1996)	2D	X	X		X			X	X
Space-time routines	Fortran	F	De Cesare et al. (2002)	2D, ST	X	X						
STIS (TerraSeer)		M	AvRuskin et al. (2004)	2D, ST	X	X			X	X	X	
Surfer		M	Golden Software, Inc.	2D	X	X						
Uncert	C	F	Wingle et al. (1999)	3D	X	X				X		
Variowin		F	Pannatier (1996)	2D	X							
VESPER		F	Minasny et al. (2005)	2D	X	X						
WinGslib	Fortran	L	www.statios.com	3D	X	X	X	X	X	X		

Notes: Cost^a: H high, M moderate, L low, F free, V variography, K kriging, CK cokriging, IK indicator kriging, MG multi-Gaussian kriging, S simulation, G GIS interface, O Other estimators (NN, IDW, splines...), ST spatial-temporal

There have been studies in which SGeMS software was used. However, the use of SGeMS on the determination of spatial distributions of water quality parameters within reservoirs is inexistent. Some of the studies related with environmental field are summarized in this part. Basarir et al. (2010) used SGeMS to simulate the uncertainty associated with the spatial variability of the standard penetration test results for a borax stockpile site. Bastante et al. (2008) compared the results of indicator kriging, conditional indicator simulation and multiple-point statistics to estimate useful slate reserves using SGeMS. Escobedo et al. (2010) used SGeMS to analyze semivariograms and estimate kriging parameters to analyze the impact of subtropical urban forests in offsetting carbon emissions from cities. Pardo-Iguzquiza et al. (2011) applied SGeMS to evaluate the morphometric parameters of three-dimensional networks of karst conduits. He et al. (2010a) applied sequence indicator simulation algorithm using SGeMS to model 3D spatial distribution of soil texture under agricultural systems. He et al. (2010b) assessed the anisotropic spatial variability of agricultural soil bulk density in an alluvial plain of north China using the 3D sequential Gaussian simulation in SGeMS. Qu et al. (2013a) evaluated spatial variability of soil total nitrogen contents and uncertainty of deficiency or abundance degrees of soil TN concentrations under different crop types via sequential Gaussian simulation algorithm of SGeMS software. Qu et al. (2013b) simulated spatial distribution of soil Ni concentration by sequential Gaussian simulation algorithm of SGeMS software and used the results to evaluate health and remediation risk costs caused by incorrect soil Ni estimations. Todd et al. (2010) investigated the spatial distribution of SOD values using SGeMS to identify the areas with intense oxygen demand in the Little River Experimental Watershed and addressed the importance of instream swamps in contributing to high DO demand at the watershed scale. Zhou and Xia (2010) modeled soil geochemical background distributions of copper, zinc, lead and cadmium in the Hengshi River watershed located in southern China using SGeMS software.

CHAPTER 3

STUDY SITE

3.1 Porsuk Dam Reservoir and Its Watershed

Porsuk Dam is located at 25 km west of Eskişehir (Figure 3.1). Construction of the dam was started in 1966 and completed in 1972. The dam and its reservoir were designed and constructed to protect Eskişehir from flooding, irrigate Eskişehir and Alpu Plains and supply drinking water to Eskişehir (Ministry of Environment and Forestry, 2006). Drainage and surface area of PDR are about 5104 km² and 24.32 km², respectively. The watershed of PDR is depicted in Figure 3.1. While the minimum water surface elevation within PDR is 850 m, the maximum water surface elevation is 892.85 m in the bathymetry map in 2001 obtained from the General Directorate of State Hydraulic Works (SHW) (Figure 3.2). The basic characteristics of PDR are provided in (Table 3.1).

PDR is mainly fed by Porsuk River, Kunduzlu Stream and waters coming from Kargin Regulator of State Hydraulic Works (Muhammetoğlu et al., 2005). A big portion of the watershed of PDR is sited within the borders of Kütahya and Eskişehir provinces (Sakarya Basin). As a result, the water quality of the reservoir is affected by point and non-point pollution sources present in the borders of both cities. The water quality of PDR is important because it supplies drinking water to Eskişehir.

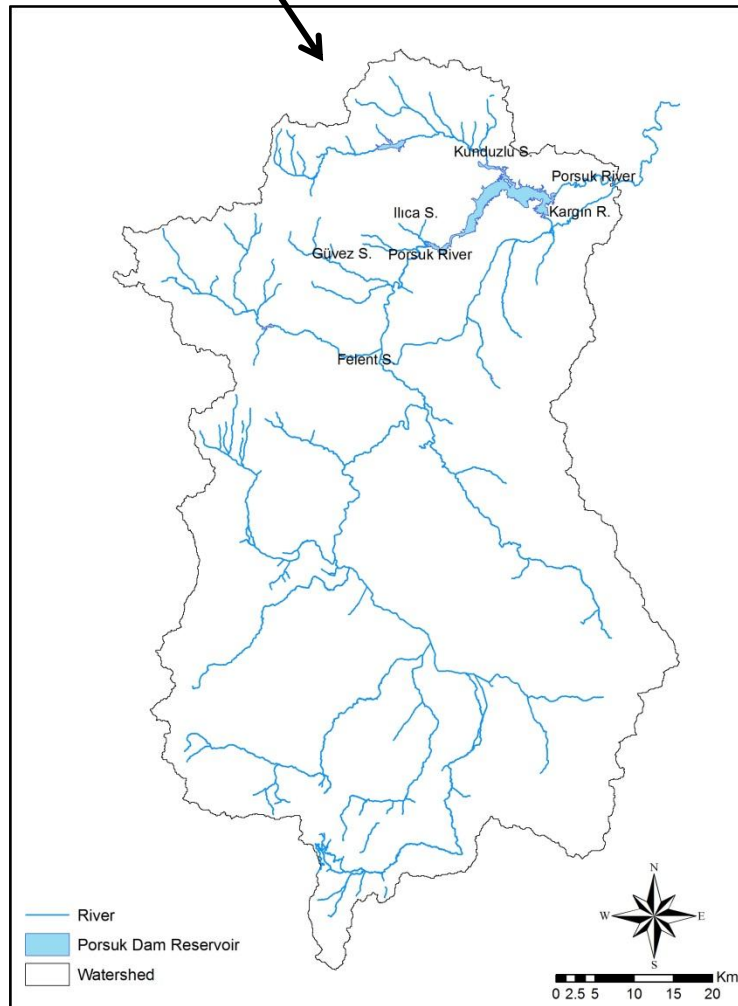
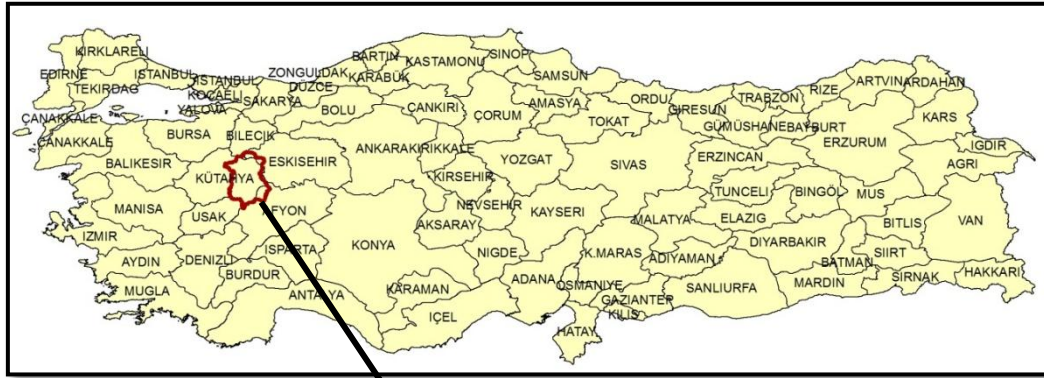


Figure 3.1 Watershed of PDR

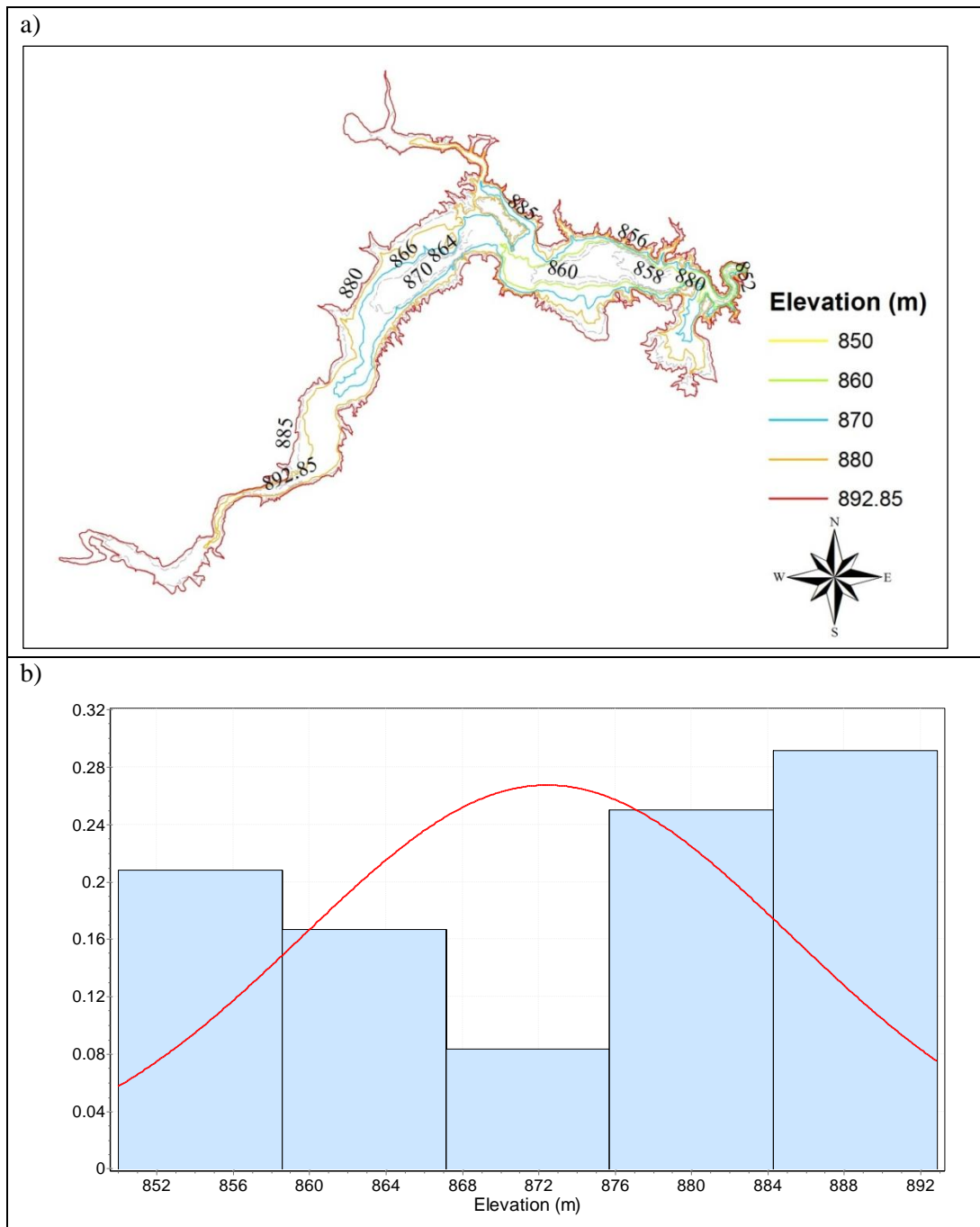


Figure 3.2 a) Bathymetry of the PDR in 2001, b) Histogram of bathymetry (AKS, 2010)

Table 3.1 Basic characteristics of PDR (Ministry of Environment and Forestry, 2006)

Characteristics	Quantities
Dam Volume	224 000 m ³
Crest Elevation	894.35 m
Crest Length	258 m
Height	64.70 m
Maximum Water Elevation	885.20 m
Normal Water Elevation	882.60 m
Maximum Storage Volume	525x10 ⁶ m ³
Maximum Storage Area	27.7 km ²
Active Reservoir Volume	446x10 ⁶ m ³
Dead Reservoir Volume	19x10 ⁶ m ³
Reservoir Drainage Area	5018 km ²
Length of the reservoir	16 km

One of the most current studies relevant to the water quality in the reservoir was conducted under “special rules designation” study (AKS, 2010). In the study, it was mentioned that the large parts of risky areas for surface drainage within the watershed of PDR were located at the north of PDR and at the north border of the watershed. Topography and landuse within the watershed of PDR was important in the evaluation of water quality.

3.1.1 Topography

Topography of a region is important especially in evaluating the impact of diffuse sources. Elevation ranges within the watershed of PDR and total surface areas within each elevation range are provided in Table 3.2 and Figure 3.3. Accordingly, 60.64% of the watershed is within the elevation range of 1000-1200 m. Lower elevation regions (<1000 m) are located around the PDR and covers 11.46 percent of the total area of the PDR Watershed.

Table 3.2 Elevation Ranges within PDR Watershed (AKS, 2010)

Elevation Ranges (m)	Area (km²)	Percent (%)
<1000	584.878	11.46
1000-1200	3095.494	60.64
1200-1400	1104.098	21.63
1400-1600	264.413	5.18
1600-1800	47.002	0.92
1800-2000	7.232	0.14
2000-2200	1.169	0.02
>2200	0.001	0.00
Total	5104.287	100.00

In addition to elevation ranges, slopes within PDR Watershed are given in three slope groups namely flat and low slope areas (0-10 %), medium slope areas (10-20 %), and high slope areas (>20%) (Table 3.3). Areas and percent within total of each slope group can be seen in Table 3.3 and Figure 3.4. The highest portion of the watershed has a slope between 0-10 %.

Table 3.3 Slopes within PDR Watershed (AKS, 2010)

Slope Group (%)	Area (km²)	Percent (%)
0-10	2657	52
10-20	1157	23
>20	1290	25
Total	5104	100

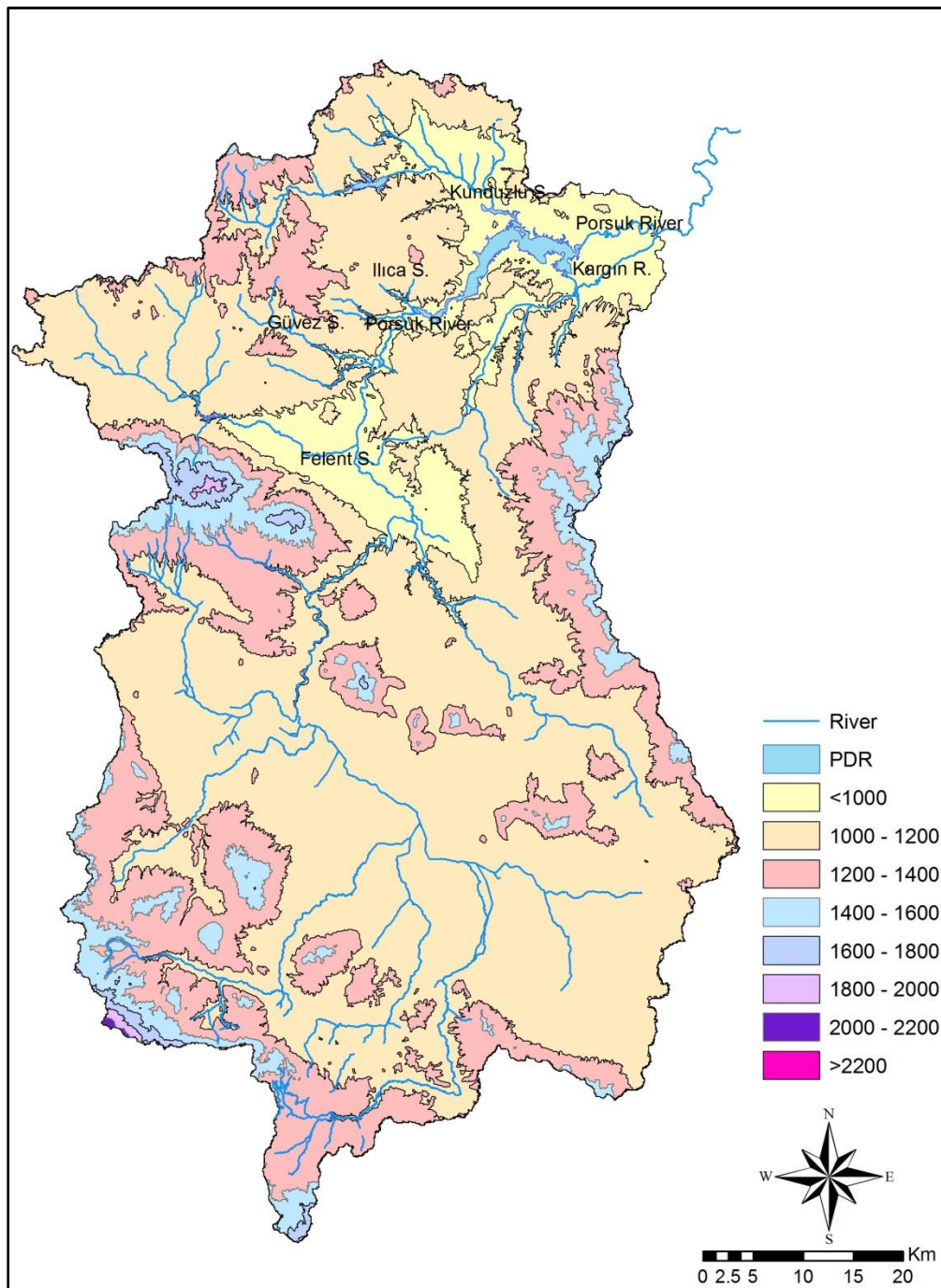


Figure 3.3 Elevations within PDR Watershed (AKS, 2010)

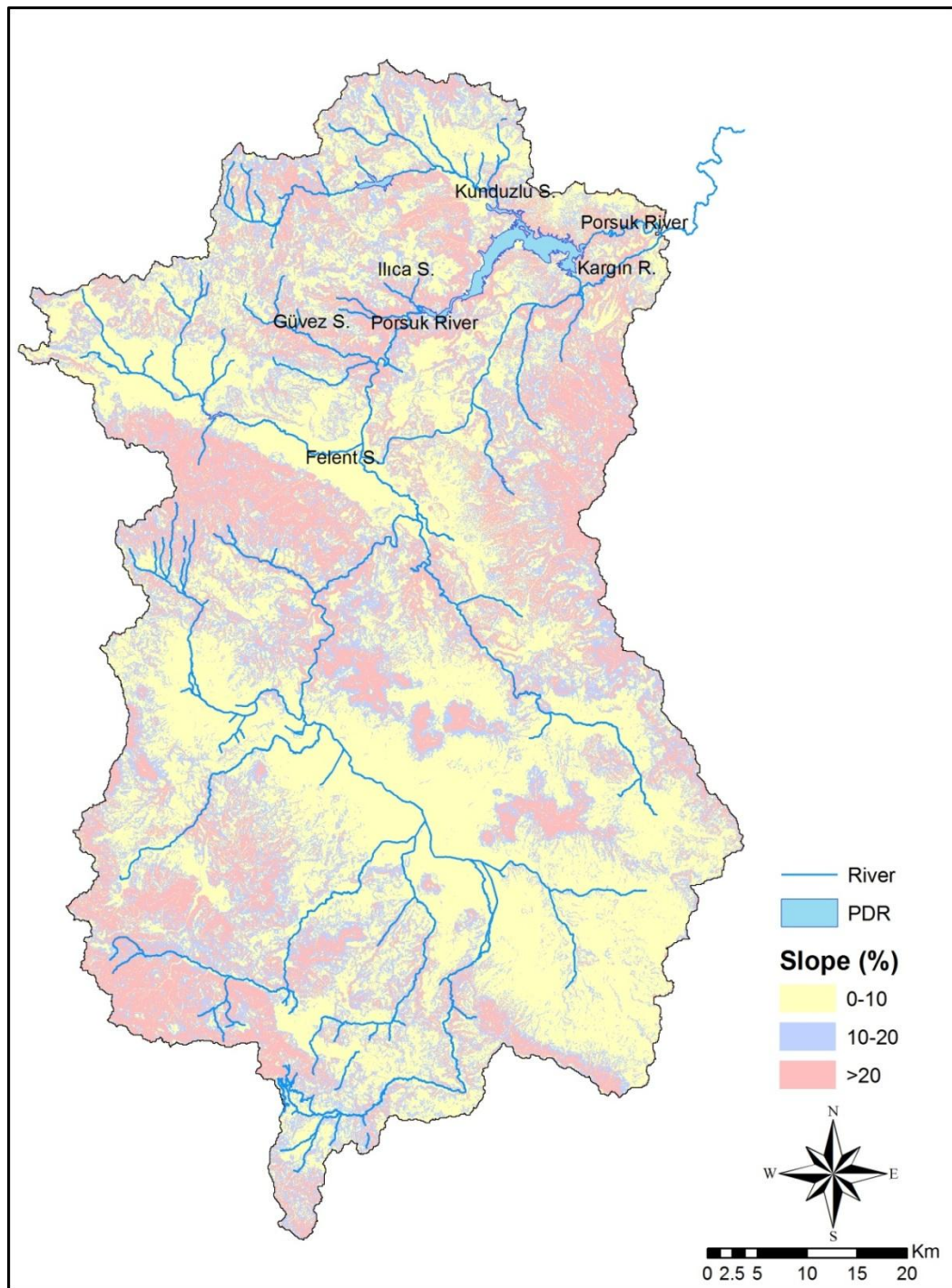


Figure 3.4 Slopes within PDR Watershed (AKS, 2010)

3.1.2 Landuses in the PDR Watershed

As given in Table 3.4 (AKS, 2010), forests and agricultural areas constitute the big portion of the landuses in the PDR Watershed. While forest areas constitute 38% of the watershed, agricultural areas cover 42% of the watershed. Most of the agricultural areas are located on Kütahya, Köprüören, Aslanapa and Altıntaş Plains. There are also large agricultural areas around İhsaniye and Dumlupınar (Figure 3.5). These land uses, their locations and practices applied in these locations are the factors which would impact the pollution load carried to PDR. In the watershed, four dams and seven ponds are present. The names and surface areas are specified in (Table 3.5) (AKS, 2010).

Table 3.4 Distribution of land uses in the watershed (AKS, 2010)

Land Use Type	Area (ha)	%
Forests	192593	37.73
Agricultural Areas	212657	41.66
Water Bodies	2826.3	0.55
Settlements	7089.7	1.39
Industrial Areas	1480	0.29
Mining Areas	3624	0.71
Other Land Uses	90221	17.67
Total	510491	100

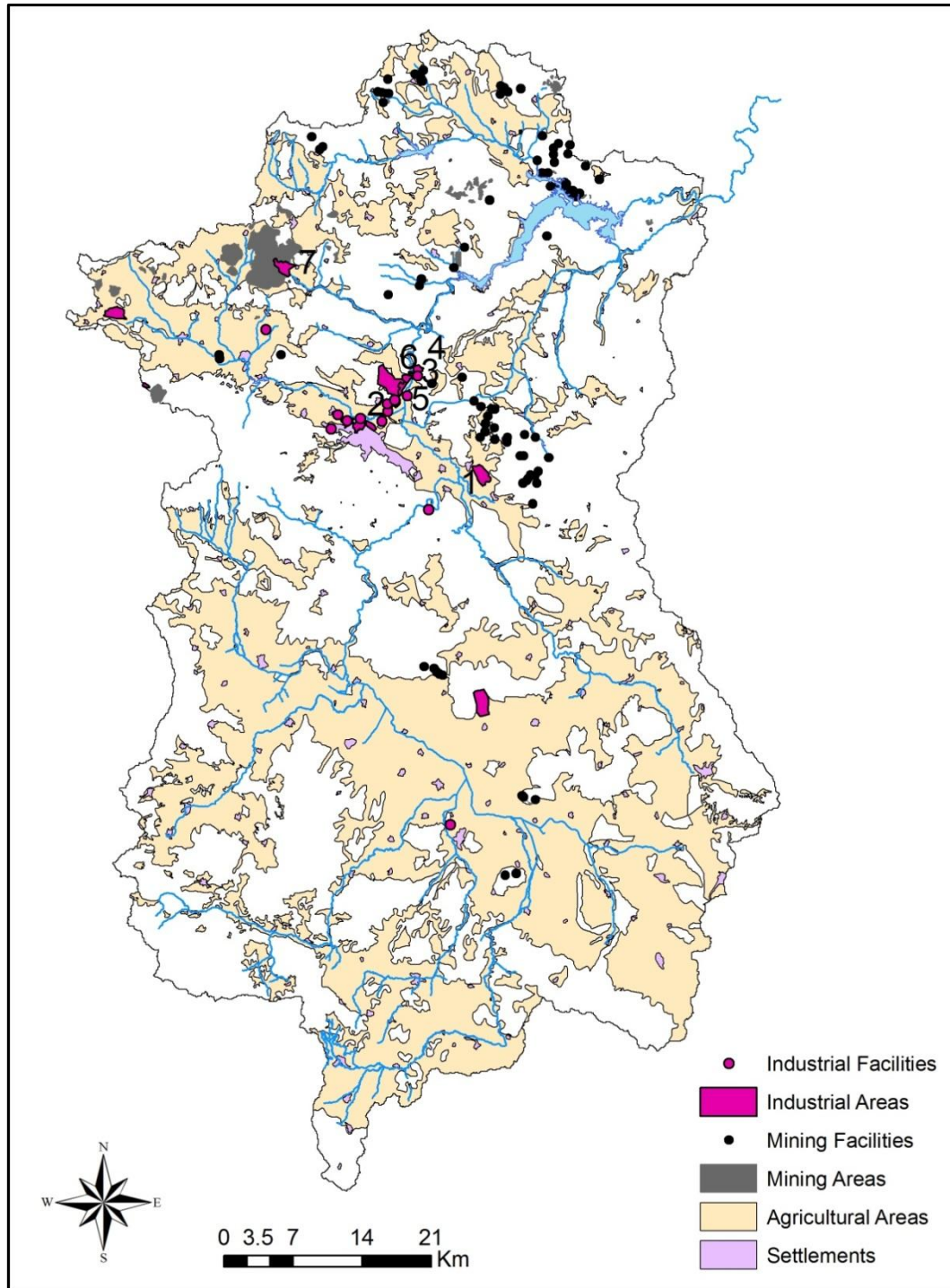


Figure 3.5 Mining areas and other land uses in PDR Watershed

(1: Kütahya Organised Industrial Zone, 2: Kütahya Sugar Refinery, 3: Nitrogen Factory, 4: Kütahya Magnesite Facility, 5: Gural Porcelain Factory, 6: Kütahya Porcelain Factory, 7: Seyitömer Thermal Power Plant)

Table 3.5 The dams and ponds present in PDR Watershed (AKS, 2010)

Dam/Pond	Area (Ha)
Porsuk Dam	2432.6
Aşağıkuzfındık Dam	238.9
Enne Dam	42.5
Söğüt Dam	9.6
Dutluca 1 Pond	8.6
Dutluca 2 Pond	4.1
Dereyalak Pond	3.0
Pullar Pond	15.3
Zafertepe Pond	31.6
Kızılcaköy Pond	13.0
Emre Pond	27.1
Total	2826.3

Table 3.6 Industrial area groups within PDR Watershed (AKS, 2010)

Groups	Area (Ha)
Organized industrial zones	221
Organized industrial zones under construction	608
Individual private industries	470
Thermal power plant	181
Total	1480

Twenty three extensive industrial areas are present in the PDR watershed. These areas can be separated into mainly four groups as organized industrial zones, organized industrial zones in construction, individual industries and thermal power plant. Surface areas covered by these groups are represented in Table 3.6. Some of the major industrial facilities and pollution stresses arising from these facilities are as follows (AKS, 2010):

Kütahya Organised Industrial Zone: Wastewater of this organized industrial zone is sent to Kütahya wastewater treatment plant (KWWTP) which has been declared to have the capacity to handle the wastewater discharge of both the city and the organized industrial zone till 2025. Kütahya Organized Industrial Zone is marked with the number 1 in Figure 3.5.

Kütahya Sugar Refinery: Until 1981, the refinery was discharging its wastewater directly into Felent Stream. In 1981-1994, treatment was applied and the effluent from the refinery's wastewater treatment plant was discharged into Felent Stream. But then, as a result of road construction, the treatment plant of the refinery was demolished and the wastewater has been sent to KWWTP since. However, as KWWTP became out of service due to shock loadings, the refinery has set up its own treatment plant again and started a closed loop operation. The domestic wastewater of the facility has been sent to KWWTP. Industrial wastewater has been discharged into the Felent Stream as treated or untreated depending on the period of the year. The location of Kütahya Sugar Refinery is marked with number 2 in Figure 3.5.

Nitrogen Factory: The factory, founded 47 years ago, produces ammonium nitrate fertilizer, technical ammonium nitrate, concentrated/diluted nitric acid and aqueous ammonia. The wastewater of the factory containing nitrate, nitrite, ammonia salt and suspended solids is kept in sedimentation basins and then discharged into Porsuk River. Nitrogen Factory is depicted with number 3 in Figure 3.5.

Kütahya Magnesite Facility: The facility had a wastewater treatment system working as a closed loop between 1983 and 1995. The slurry coming from the grinding and washing units of the facility was kept in the sedimentation basins gradually. Following sedimentation, the water was recycled back to the system for reuse. After 1995, magnetic separators have been used and industrial wastewater has not been produced. The facility has a wastewater treatment system for

domestic wastewater. The effluent is used to irrigate the garden of the facility. Kütahya Magnesite Facility is located in the place shown with number 4 in Figure 3.5.

Porcelain Factories: Since Güral Porcelain Factory had no wastewater treatment plant before 2005, wastewaters were discharged into Porsuk River following sedimentation. After 2005, they started to operate an industrial wastewater treatment plant and discharged the effluent into Porsuk River. Kütahya Porcelain Factory has its own sedimentation basins and effluent is discharged into Porsuk River after sedimentation. Yet, treatment is insufficient. The construction of a new wastewater treatment unit is in progress. While Güral Porcelain Factory is marked with number 5 in Figure 3.5. Kütahya Porcelain Factory is marked with number 6 in the same figure.

Seyitömer Thermal Power Plant: Following treatment, the effluent of the power plant is discharged into Güvez Stream which merges with Porsuk River at Başdeğirmen region. However, it has been noticed that the wastewater from the ash ponds is being discharged into Güvez Stream when cleaning the ponds. The plant location is shown with the number 7 in Figure 3.5.

The watershed is also active in terms of mining activities. Figure 3.5 depicts the locations of various mining activities within the watershed. As can be seen several mining areas are present in the north of PDR where also higher slopes are observed (Figure 3.4). Types and sizes of the mining activities are given in Table 3.7 (AKS, 2010).

Table 3.7 Sizes of mining areas within PDR Watershed (AKS, 2010)

Type	Area (Ha)
Coal Mine	2933
Magnesite Mine	449
Silver Mine	201
Marble Quarry	30
Dolomite Quarry	5
Chromite Mine	6
Total	3624

3.1.3 Water Quality History in the Watershed

There have been many studies focusing on the pollution sources and the water quality in the watershed of the PDR. Although the study site in this study is limited by the reservoir only, studies relevant to water and sediment pollution in rivers and streams in the watershed of the PDR will be summarized as well since pollution is transported from these sources to PDR.

Water quality monitoring studies in the watershed of PDR started in 1969 by collection of data related to pollution of rivers within the Sakarya Watershed by the Department of Technical and Quality Control of SHW (Bilge, 1997). SHW conducted a study between March 1974 and January 1975 to see the impact of Porsuk and Çarksuyu Rivers on the water quality of Sakarya River. The results of this study were evaluated and reported in 1977 by the same department. It was determined that although Porsuk and Çarksuyu Rivers were polluted by domestic and industrial wastewater discharges, the water quality of Sakarya River was quite good (Bilge, 1997). In 1978, SHW started a pilot project supported by the United Nations Development Program and World Health Organization to examine the water quality of Porsuk River. Samples from 10 observation stations and from wastewaters of industries were collected for 2 years (Bilge, 1997).

In 1981, a project, entitled as “River Basin Management for the Sakarya Basin”, was conducted by the Civil Engineering Department of Istanbul Technical University, with the cooperation of SHW and participation of the Operations Research Department of the Marmara Scientific and Industrial Research Institute under the NATO Science for Stability Program. The goal of the project was to select the most cost-effective development plan for the basin. In the scope of that study, the river basin model was setup and simulations were performed. Simulation results showed that the water quality in the Porsuk River was not suitable for any use. Domestic discharges, the fertilizer plant, and the industrial organized zone were mentioned as the major sources of nutrient pollution. It was mentioned that implementation of the best available technologies (BAT) resulted with a noticeable improvement in all water quality parameters. Overall, the quality of Porsuk River could reach Class II status (slightly polluted water) with the application of control strategies (Bilge, 1997; Tanik et al., 2005).

A project entitled as "Water Quality Studies in Natural and Artificial Lakes and Basins" was started with the contribution of the British Government in 1984. In addition to investigation of physical and chemical water quality parameters within the watershed of PDR, biological and microbiological studies were carried out (Bilge, 1997).

Yücel et al. (1995) examined cadmium, zinc, copper and lead concentrations in soil, water and leaf samples collected from 10 stations located along the Porsuk River and its tributaries. The concentrations of mentioned parameters reached to unacceptable levels. Moreover, they mentioned that Porsuk River should not be used as a drinking and domestic water supply and alternative water supplies should be identified.

Bilge (1997) estimated suspended solids, $\text{NO}_3\text{-N}$, $\text{PO}_4\text{-P}$, chlorophyll-a and light intensity in the Porsuk Dam Reservoir using Landsat TM data. Data was obtained from SHW and samples were collected on the date of imagery. The relationships

between water quality parameters and radiance values in the first four bands of the TM data were determined. Results indicated that PDR acted as a settling pond for suspended solids.

Mazlum et al. (1999) investigated the factors causing variations in the observed quality data at the Ağaçköy water quality monitoring station located on Porsuk River by PCA. For this purpose, water quality parameters including flow rate, temperature, pH, electrical conductivity, suspended solids, M-Al, Cl, NH₃-N, NO₃-N, DO, and BOD measured between 1979 and 1984 were used. The results of PCA showed that the main causes of variations in water quality in that station were small domestic waste discharge, industrial waste discharge, nitrification and seasonal effects.

SHW (2001) worked on a project to prepare a water management plan for the watershed of PDR. It was declared that the trophic status of PDR shifted from eutrophic to hyper-eutrophic state from 1986 to 2001. It was mentioned that the phosphorus load reaching to PDR was higher than the critical phosphorus load for PDR. Similarly, accumulation of nitrogen within PDR was obvious based on the mass balance on nitrogen. Measures that should be applied to control and reduce nutrients loads and improve the water quality of PDR were summarized as follows:

- KWWTP, which was one of the major phosphorus load contributors to PDR, should be operated at full performance,
- The discharge of the wastewaters of Kütahya Sugar Factory to Porsuk River should be prevented,
- The water quality in Güvez Stream, which is mainly affected by wastewater discharges from Seyitömer Thermal Power Plant and settlements, should be controlled,

- Precautions should be taken to reduce the nitrogen load to Porsuk River originating from Kütahya Nitrogen Factory,
- Suspended solids load originating from porcelain factories should be controlled and accumulation should be prevented in PDR.

Çiçek and Koparal (2001) examined the levels of lead, chromium and cadmium both in fish (*Cyprinus carpio* Linnaeus and *Barbus plebejus* Bonaparte) tissues and in the reservoir. While lead and chromium in fish tissues were below detection limits, cadmium was above the limit values set by the Turkish Agricultural Ministry. Water quality in the reservoir was Class I (high quality water) for lead, Class II (slightly polluted water) for chromium, Class III (polluted water) for cadmium with respect to the limit values given in Inland Water Resources Classes of Water Pollution Control Regulation (Çiçek and Koparal, 2001).

Beğenirbaş (2002) studied Cu, Cr, Pb, Hg, As and Cd accumulations in freshwater mussel samples collected from upstream and downstream of the Kütahya section of the Porsuk River. It was observed that the amounts of As, Hg and Pb in viscera of mussels were higher than the limit values with respect to relevant regulations.

Ocak et al. (2002) examined ecotoxicological and morphological effects of the irrigation water withdrawn from the Porsuk River on some agricultural plants. Furthermore, Fe, Zn, Cd, Ni, Pb and Cr contents of the plants were determined. Results were compared with the results of control plants which were irrigated with tap water. Reasonably higher accumulations of Ni and Cr on some plants compared to control plants were observed. Moreover, accumulation of heavy metals was detected in sediment as well.

Akdeniz (2004) evaluated stabilization of the polluted sediment of Porsuk River with additive materials such as fly ash, lime, cement, zeolites and sepiolites. For this purpose, sediment samples were collected from various locations in Porsuk

River. While heavy metals including Pb, Zn, Ni, Cd, Cr and Mn in sediment samples were higher than limit values, Cu and Cd were less. It was mentioned that pollution was especially concentrated at the regions where treatment systems were inadequate and at the regions close to industries.

Kutlu et al. (2004) assessed the potential for contamination by mutagenic substances in the Porsuk River water and sediments with the aid of *Salmonella* mutagenicity assay (Ames test). Results showed the presence of heavy metals in the Porsuk River, mainly in the upstream of Porsuk Dam. This was due to the presence of anthropogenic stresses such as industrial activities and urban drainage.

Özyurt et al. (2004) investigated the pollution problem in the watershed of PDR originating from Kütahya. In this study, discharges from several industrial facilities located in Kütahya were analyzed in 2003 to 2004. Analyzed parameters were pH, BOD, COD, TSS, Pb, Cd, oil and grease, and total phosphorus. Results of the experiments were compared with the Turkish Water Pollution Control Regulation. It was observed that some industries did not comply with the discharge standards set in the regulation. It was concluded that these industries have negative impacts on the pollution load and the conditions for aquatic life in the river.

Muhammetoğlu et al. (2005) assessed the impact of different pollution control scenarios on the water quality of Porsuk River and the dam system to develop reasonable water quality management strategies. Different levels of treatment ranging from conventional treatment to tertiary treatment were evaluated for the major domestic point sources. Considered treatment options for the major industries were based on the allowable discharge limits in the Turkish Water Pollution Control Regulation and the best available technologies. QUAL2E and BATHTUB models were used to simulate the water quality in the river and in the reservoir, respectively. In order to improve the water quality of Porsuk Dam

Reservoir from highly hypertrophic state to eutrophic-mesotrophic state, as well as to improve the water quality of Porsuk River, the followings were suggested:

- The effluent of KWWTP that is discharged to Porsuk River should be diverted,
- For all industries best available technologies (BAT) and 50% reduction in nitrogen and phosphorus loads in the catchment area should be applied.

Koçal (2006) analyzed the eutrophication problem in PDR via modeling. Loads coming from point and non-point sources were determined. GROWA, a water balance model, was used for the determination of loads originating from non-point sources. Total annual evapotranspiration, runoff and groundwater recharge maps were obtained by GROWA. According to results, runoff in the fallowed agricultural areas was around 50-200 mm/year. Therefore, agricultural activities and fertilizer usage in the area constituted significant pollution sources for the reservoir. Koçal (2006) suggested that people should be conscious when using fertilizers to minimize the transport of excess nutrients to Porsuk River and the reservoir.

Orak (2006) assessed the pollution levels in the Porsuk River using fuzzy logic. 2001-2002 data including DO, BOD, COD and $\text{NH}_3\text{-N}$ at seven water quality sampling stations on Porsuk River was obtained from SHW. While the pollution class in upstream of Kutahya was determined as Class I (high quality water) for both years, Class IV (highly polluted) and Class II (slightly polluted water) were observed downstream of industries in 2001 and 2002, respectively. Although the water quality of PDR was Class I in 2001 at the outlet of the dam, the Porsuk River was polluted in Eskisehir and merged with the Sakarya River as Class IV.

Semerçi (2006) collected sediment, water and soil samples from the watershed of Porsuk River to determine geotechnical and chemical properties such as pH, cation exchange capacity, electrical conductivity and total metal amounts in

samples. Results revealed Fe, Mn, Cd, Pb, Ni, Cr and Co and pesticide concentrations in the sediment of Porsuk River exceeded the limit values.

Yuce et al. (2006) examined the pollution of water resources in the Eskişehir Plain and the Porsuk River Basin. The study aimed at investigation of the degree of influence of contaminant sources on the quality of water resources. Six surface water samples, 24 groundwater samples and 12 soil samples were collected and analyzed from May 2001 to October 2001. Results indicated intensive NO₂, NO₃, NH₃, NH₄, phenol, AOX, phosphorus, free chlorine, sulfur, Fe, Al, Pb, Cr, Mn, Cd, and Zn pollution in Porsuk River and groundwater in some parts of the watershed. It was concluded that surface and groundwater in the study area were not suitable to use as a drinking and domestic water supply based on the comparisons with the relevant Turkish, European Union, and World Health Organization standards (Yuce, 2006).

Kavaf and Nalbantcilar (2007) assessed the contamination characteristics of surface water and groundwater in the Kütahya plain and examined the relationship between contaminant sources and quality of waters. Furthermore, potential risk on public health was discussed by comparing with EPA drinking water standards. Ag, Al, As, B, Br, Ca, Cd, Cu, Fe, Hg, K, Li, Mg, Na, Ni, P, Pb, S, U, and Zn were analyzed in a total of 44 spring, river and groundwater samples. According to results, maximum levels were exceeded on the basis of EPA standards. It was mentioned that Porsuk River should not be used for drinking and domestic use.

Öztürk (2007) evaluated the environmental problems related to watershed of Porsuk River and suggested solutions to these problems.

Arslan (2008) identified the variables that are most important in assessing the variations in the water quality of Porsuk River using PCA. The study was performed by eight water quality parameters (flow rate, temperature, DO, BOD, NH₃-N, NO₃-N, NO₂-N, and PO₄-P) measured in 11 water quality sampling

stations located on the Porsuk River. Water quality class was determined as Class-IV both in upstream and downstream of Porsuk Dam.

Büyükerşen and Efelerli (2008) mentioned the factors that negatively affected the water quality of PDR. Major ones were partially treated and untreated wastewaters of Kütahya, wastewaters of Kütahya Nitrogen Factory and Kütahya Sugar Factory, wastes of slaughterhouse, use of fertilizers and pesticides intensively in agriculture. While the water quality of Porsuk River exhibited Class I properties at upstream of Kütahya, it declined to Class III or Class IV at downstream of Kütahya. Similarly, the water quality just at downstream PDR exhibited Class I or Class II properties, but then declined to Class IV after it is polluted by sources in Eskişehir. It is mentioned that all users of Porsuk River must comply with regulations.

Yerel (2010) used multivariate statistical techniques such as PCA, FA and CA to evaluate the surface water quality of Porsuk River. Data used in the study covered 11 parameters measured in 11 observation stations for 5 years. According to the results of PCA, 66.88% of variances in the data set was explained by three factors of which 42.93% originated from anthropogenic sources. In the CA, eleven observation stations were grouped into two clusters based on the similarity of surface water quality parameters. The results showed that urban, industrial and agricultural discharge strongly affected the east part of the region.

Bakış et al. (2011) investigated the pollution levels both in surface water and groundwater within the Porsuk River Watershed. In the study, seasonal surface water and groundwater samples were collected in 2005-2006. Results revealed that the surface water and ground water within the watershed were classified as Class IV on the basis of relevant regulations.

Küçük (2013) conducted a hypothetical assessment to show the relative effects of different agricultural practices on the watershed of PDR using ArcSWAT. Ten

different scenarios with varying fertilizer and irrigation rates, landuse and point source phosphorus load reductions were applied to detect the vulnerable areas for phosphorus transport. 40% decrease in the overall soluble P load to the PDR was obtained when simultaneously 50% and 20% reductions were applied in fertilizer usage and irrigation, respectively.

In none of the studies above, geostatistical tools were used to evaluate the representativeness of the water quality sampling locations. Moreover, hotspots based on water quality parameter distributions were not evaluated in 2D and 3D space based on landuses in the vicinity of PDR. As a result, this study will contribute to the enhancement and protection of the water quality in PDR through establishing the spatial distributions of DO and SC in 2D and 3D to identify the hot spots based on geostatistically representative sampling points.

3.1.4 Current Water Quality of PDR

One of the most current water quality studies is performed under “special rules designation” study for enhancement and protection of the water quality in PDR, focusing mainly on preparation of watershed protection and landuse management plans and definition of the protection zones at where several activities (agricultural, industrial, commercial, etc.) can be banned or restricted (AKS 2010; AKS, 2011). In this study, 11 sampling locations within the PDR were determined to evaluate the water quality in the PDR. Both surface and bottom samples were collected. Sampling locations are depicted in Figure 3.6. While sampling locations 9, 10, 12, 14, 16 and 18 constitute surface sampling locations within PDR, sampling locations 11, 13, 15, 17 and 19 are the bottom sampling locations. Coordinates of all sampling points are given in Table 3.8.

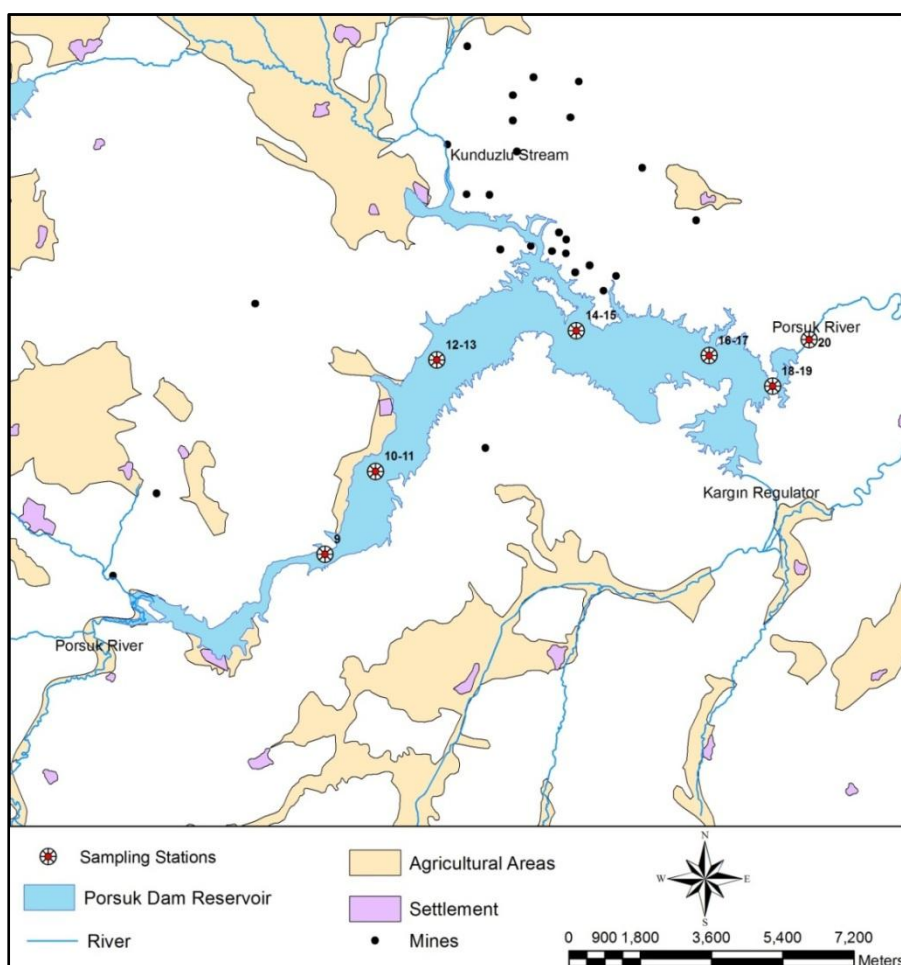


Figure 3.6 Sampling Locations in AKS (2010)

Table 3.8 Sampling stations and coordinates

No	Sampling Station (SS)	Coordinates (UTM ED-50 Zone 36)	
		East	North
9	Southwest PDR	254634	4385905
10-11	West PDR	255924	4388012
12-13	Central PDR	257452	4390804
14-15	Kunduzlu Stream Expansion	261008	4391594
16-17	East PDR	264353	4390962
18-19	Dam Crest Area in PDR	265960	4390198
20	Outlet of PDR	266881	4391337

The general trophic classification (Wetzel, 2001) of PDR is defined in relation to concentrations of phosphorus, nitrogen, phytoplankton pigment (chl-a), and water transparency (Table 3.9). While PDR is in hyper-eutrophic range according to mean total P and Chl-a concentrations, it is in eutrophic range in mean TN concentration. According to this classification, the reservoir is in hyper-eutrophic range.

Table 3.9 General Trophic Classification of Lakes and Reservoirs in Relation to Phosphorus and Nitrogen^a (Wetzel, 2001) and measured concentrations in PDR

Parameter (annual mean values)		Oligotrophic	Mesotrophic	Eutrophic	Hyper- eutrophic	Measured concentrations in PDR
Total Phosphorus (mg m ⁻³)	Mean	8.0	26.7	84.4	-	445
	Range	3.0-17.7	10.9-95.6	16-386	750-1200	
Total Nitrogen (mg m ⁻³)	Mean	661	753	1875	-	1244
	Range	307-1630	361-1387	393-6100	-	
Chlorophyll a (mg m ⁻³) of phytoplankton	Mean	1.7	4.7	14.3	-	21
	Range	0.3-4.5	3-11	3-78	100-150	
Chlorophyll a maxima (mg m ⁻³) (worst case)	Mean	4.2	16.1	42.6	-	80
	Range	1.3-10.6	4.9-49.5	9.5-275	-	
Secchi transparency depth (m)	Mean	9.9	4.2	2.45	-	
	Range	5.4-28.3	1.5-8.1	0.8-7.0	0.4-0.5	

^aBased on data of an international eutrophication program. Trophic status based on the opinions of the experienced investigators. (Modified from Vollenweider, 1979)

An evaluation is made based on several water quality parameters measured in January, April, August and September 2010 (AKS, 2010). Evaluation is performed according to Appendix 1 of the regulation on the quality of surface waters intended for abstraction of potable water (Ek 1, İçmesuyu Elde Edilen veya Elde Edilmesi Planlanan Yüzeysel Suların Kalitesine Dair Yönetmelik). The limit values are given in the Appendix A. Data obtained in January, April, August and September of 2010 was averaged. These measurements for some parameters for each sampling station are presented in Figures C1-C10 in Appendix C. In these figures, green, yellow, orange and red colors map into categories of A1 (drinkable water following physical treatment and disinfection), A2 (Drinkable water following physical and chemical treatment and disinfection), A3 (Drinkable water following physical, chemical and advanced treatment and disinfection) and higher than A3, respectively.

The water quality in PDR falls into Category A1 for color, odor, TSS, nitrate, copper, zinc, boron, cobalt, cadmium, chrome (total), lead, mercury, barium, sulphate, chloride, fluoride, dissolved iron, surfactant (MBAS), fecal coliform, fecal streptococcus, and salmonella parameters. COD decreased as progressed from entrance (SS9) to outlet of PDR (SS18-19) (Figure C. 1). These results are well proportioned with DO concentrations (Figure C. 3 & Figure C. 4). There was also a huge decrease in BOD at the outlet of PDR compared to inlet of PDR with the exception of increase in the station located in the Kunduzlu Stream expansion (SS14-15) (Figure C. 2). Phosphates in all sampling locations within PDR represent higher values than for Category A3 (Appendix A). Therefore, PDR is rich in phosphates (Figure C. 5). Unlike phosphates, nitrate falls into A1 class at all sampling stations (Figure C. 6). There was a declining trend in TKN and $\text{NH}_3\text{-N}$ from the inlet (SS9) to the outlet of PDR (SS18-19). This was also valid for chlorophyll-a (Figure C. 7, Figure C. 8 and Figure C. 9). While TKN observed in at the inlet of PDR was in the range of A2 (SS9), it improved to A1 at the outlet of PDR (SS18-19) (Figure C. 7). Similarly, $\text{NH}_3\text{-N}$ concentration at the inlet of PDR was in the range for A3, it was A2 at the outlet (Figure C. 8). Although

conductivity exhibited A1 property throughout PDR, there should be some factors causing the increase of conductivity between SS10-11 and SS12-13 (Figure C. 10). In general it was shown that besides Porsuk River, other factors such as Kunduzlu Stream, regulator of SHW, Sofça and Sobran villages and mining activities had negative affect on the water quality of PDR.

CHAPTER 4

METHODOLOGY

The principal instrument used in the temporal and spatial management of water resources is the water quality monitoring network (Strobl et al., 2006a). Principal elements of a monitoring plan can be listed as a clear statement of the objectives, definition of intended uses (the present and planned water uses), a complete description of the study area (extend of the area, the present and expected pollution sources, etc.), description of the locations of the sampling sites, a listing of the water quality variables that will be measured, determination of the frequency and timing of sampling, estimation of the resources available to implement the proposed water quality network (UNEP/WHO, 1996). Strobl et al. (2006a) pointed out the absence of a concise methodology to design monitoring networks, especially in selection of sampling stations. Modelers often faced with non-homogeneity (clustering) problem in data obtained from monitoring networks due to preferential sampling or impossibility of reaching certain regions in the water body of concern (Kanevski, 2008). Therefore, the design of water quality monitoring network that constitutes robust and representative sampling locations is an issue in water quality characterization. A good monitoring network design is essential in decision making at watershed scale (Do et al., 2012). Geostatistical tools can aid in better design of the monitoring networks which is the focus of this study. In this study a water quality network is defined by the number of sampling points and their distributions.

In the first part of this study, the steps proposed to design a representative water quality monitoring network (selection of sampling locations) that maintains the actual spatial correlation structure in water quality parameter distribution within reservoirs and lakes using geostatistical tools (2D kriging and Kernel Density Estimation (KDE)) are presented. The spatial correlation structure defines the

distance which two measurements of a variable become spatially correlated/uncorrelated. This information can be beneficial to determine the locations of sampling points. If the actual structure in water quality parameter distribution within reservoirs and lakes is not retained, selected sampling points may not be representative. The methodology followed is depicted in Figure 4.1. Firstly, the surface area of PDR was divided into square grids of 500 m by 500 m because it was assumed to represent actual spatial correlation as it is one of the largest sample set that can be obtained. Then, data collection (water quality measurements) was performed in field at points in the middle of the grids for the field work step of the methodology. After that, measured data was preprocessed and divided into sub-data sets as calibration and validation sets for 2D kriging (Figure 4.1). Details of these steps are given in Section 4.1. Following these steps, exploratory data analysis step was applied. In this step, normality check and transformation, trend analysis and removal were performed on the calibration data consecutively. If the calibration data did not follow normal distribution, necessary transformations were applied. Similarly, if trend was observed in the data, it was removed before kriging. After that variogram analysis was performed which consisted of experimental variogram construction and fitting a variogram model to the experimental variogram. Hence, kriging estimations were made using processed calibration data (Figure 4.1).

Validation data set was used to evaluate the accuracy of kriging estimations through several measures. In the evaluation step, variogram models, kriging maps, kriging error maps, and error metrics were taken into account to decide whether the given network was representative and further point reduction was possible or not. If further point reduction was possible in the given network (a network is defined by the number of sampling points and their distributions), KDE was used to define the areas with similar densities. These areas were used in identifying the potential water quality sampling locations that could be removed from the sampling network in order to decrease the total number of samples. Then based on land use information and locations of surface drainage and point source inlets, the

sampling points that would be eliminated from the network were determined. Several networks were generated to find the best combination. In addition to this methodology, sampling points within PDR were selected by 3 different experts in watershed management and kriging estimations were obtained based on their selection to constitute surface DO maps. As a result comparisons were made between kriged estimations based on sampling points selected by experts and the ones identified with the proposed methodology.

In the second part of this study, 3D kriging was performed to compare the hotspots observed on the surface of PDR with the hotspots in deeper layers. The methodology is depicted in Figure 4.2 as a summary. Data collection was performed with respect to depth in the middle of the square grids of 500 m by 500 m that were also used in the first part of the study. Collected data was preprocessed. Different from the approach used for 2D kriging, data was not divided into calibration and validation sets in 3D kriging application to provide more data for 3D kriging. In the exploratory data analysis, normality check and transformation were applied on the data. After that variogram analysis was performed. Then 3D grid was constructed for kriging and kriging was performed. Cross-validation was used to evaluate the accuracy of 3D kriging estimations. In the following parts, details of the steps proposed in identification of the representative sampling locations for 2D kriging and 3D kriging application methodologies will be explained.

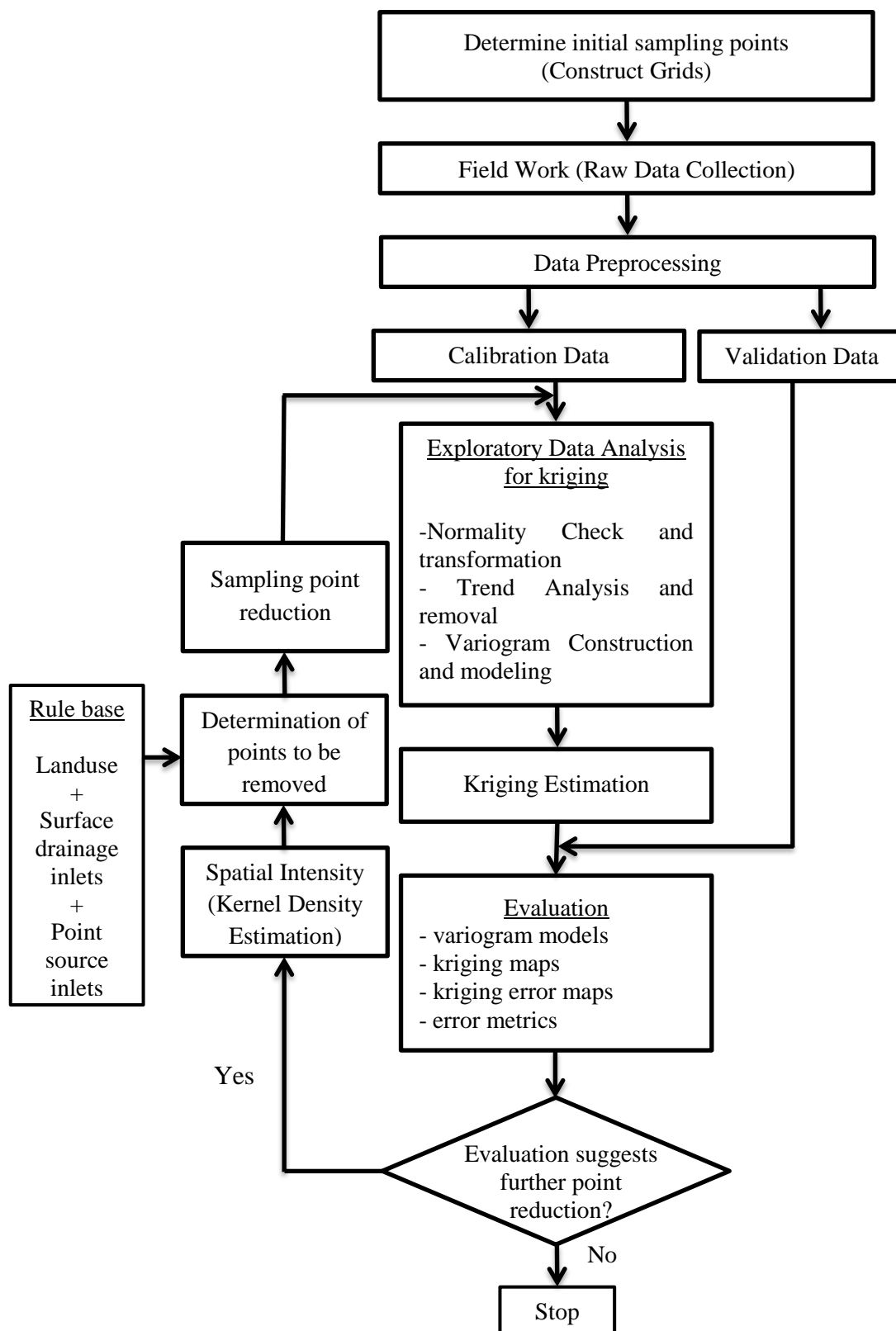


Figure 4.1 The methodology employed in selection of sampling points

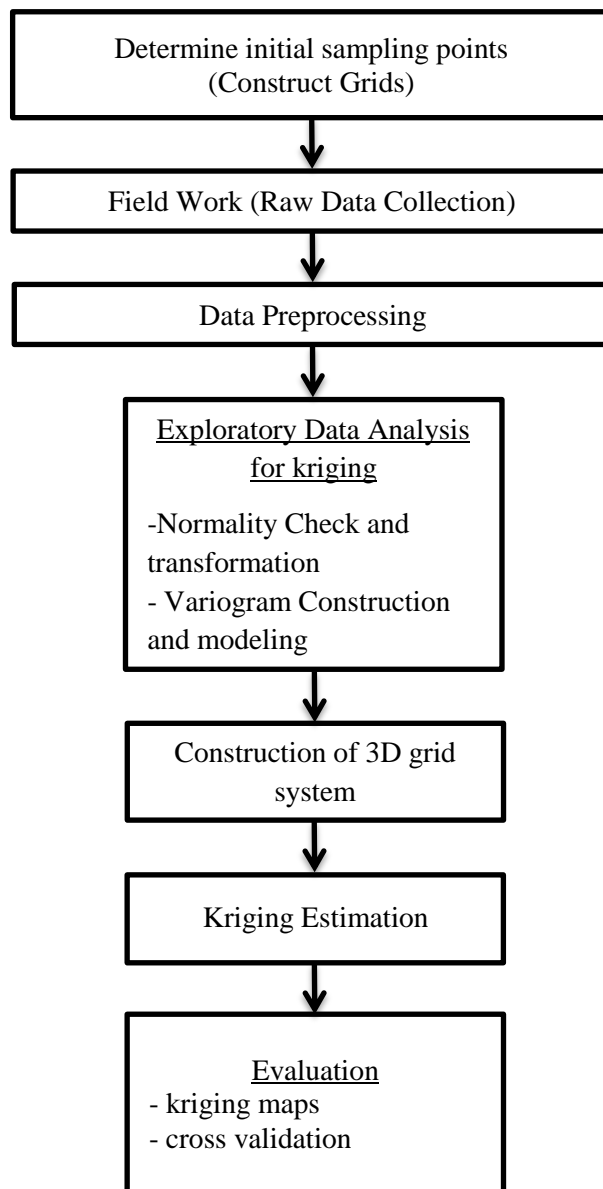


Figure 4.2 The methodology employed in 3D kriging

4.1 Field Work

The maximum distance between monitoring locations depends on several factors such as the size of the water body, fluctuations in water quality parameters, and inlets of point and non-point pollution sources. Industries present in the area as mentioned in Section 3.1.2 directly affect the water quality of Porsuk River which is the main inflow to PDR (Figure 4.3). In previous studies where kriging was employed in lakes and reservoirs, the sampling point densities were in the range of 0.01 – 0.29 points/km² (Beveridge et al., 2012; K lahcı and Ően, 2009). Although higher sampling point densities allow for better representation of the water quality in lakes/reservoirs, economic and feasible constraints should be taken into account in the selection of the density. Therefore, in this study it was deemed that division of the PDR into grids of 500*500 m with approximately 3 points/km² sampling point density was adequate to start the process of elimination of sampling points based on the assumption that it is sufficiently large to represent actual spatial correlation. This value is higher than the sampling densities reported in literature.

A total of 81 sampling locations were obtained corresponding to the sampling density used in this study and sampling points were at the center of 500*500 m grids (Figure 4.1). The coordinates of these locations are determined through ArcGIS and transferred into a Garmin GPS receiver (GPS map 76CX) before conducting the field study. The sampling locations identified earlier were pinned down during the field study which was conducted on 24th and 25th of August 2010. At each sampling location in-situ measurements were made for temperature, DO, salinity and SC along the water depth. While DO is one of the important water quality parameters for aquatic life and one of the primary concerns of limnologists and water resource engineers, SC can be used to evaluate the locations of the point and non-point loads (Soyupak et al., 2003; Karamouz et al., 2009a). For in-situ measurements, a YSI 6600 EDS multiparameter water quality sonde was utilized. The data was automatically recorded while the sonde

descended along the depth at a given sampling point. Data and corresponding depth values were recorded until the sonde hits to the bottom of the lake. Then, the data and corresponding depth values were transferred to a computer for processing in the office. The data recorded on the surface of sampling locations shown in green circles in Figure 4.4 was used as the calibration data set (65 points), while data recorded in red circles was used as the validation data set (16 points) to identify the representative water quality sampling locations. The data recorded along the depth in the sampling locations (81 points) was used to constitute 3D distributions of selected water quality parameters and to perform 3D kriging in this study.



Figure 4.3 Porsuk River Inlet to PDR

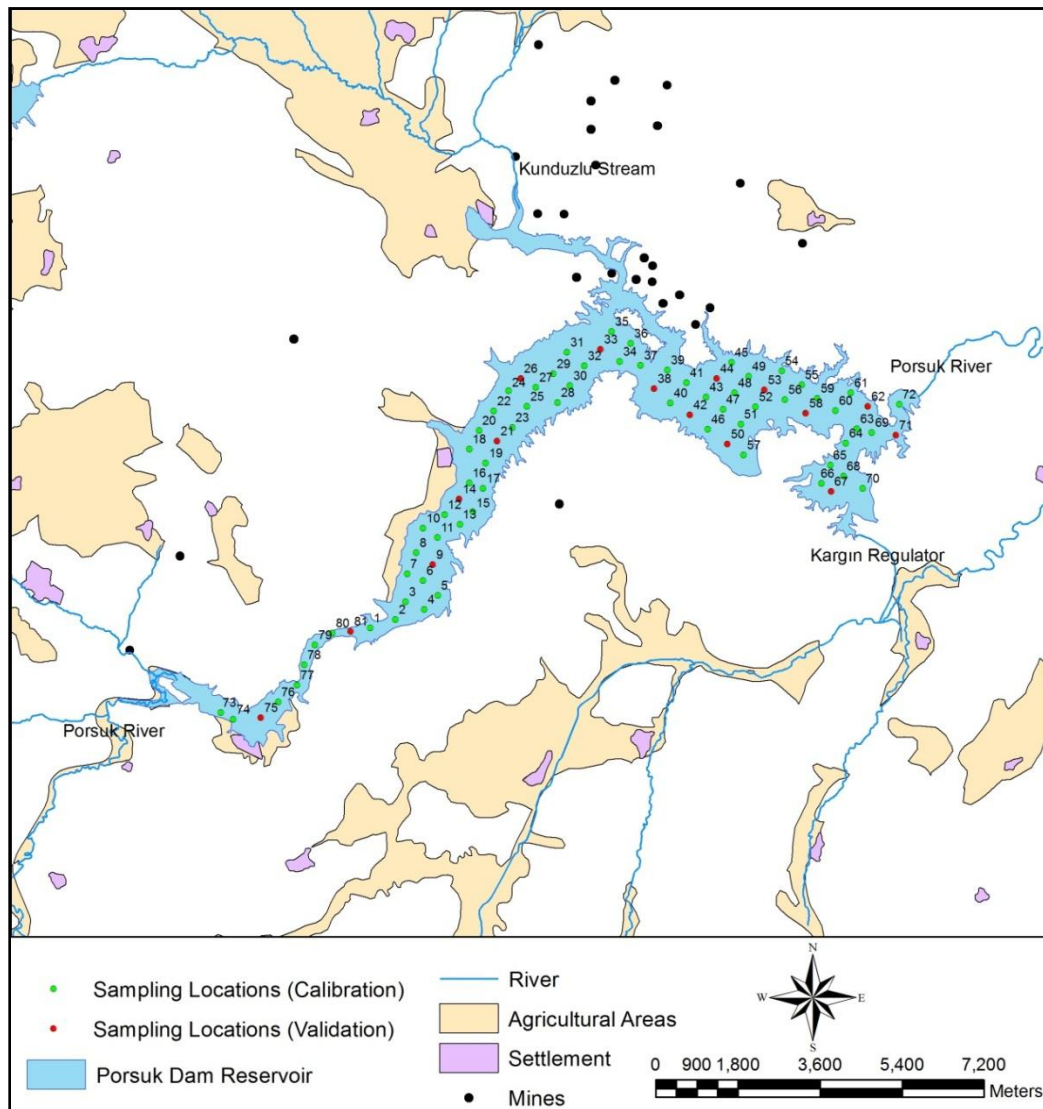


Figure 4.4 Sampling locations in the field study (green circles: calibration data points, red circles: validation data points)

4.2 Kriging

Kriging is a weighted moving average technique to estimate expected values of pollutant concentrations at unmeasured locations. It can be used to assign an estimated value to a particular location or to a block, named as point kriging or block kriging, respectively. Kriged estimate is a weighted combination of the sample values around the point to be estimated. Although other linear unbiased estimators also exist such as inverse distance methods, kriging is the best linear unbiased estimator in which the mean value can be reproduced by the kriging estimate. Moreover, kriging minimizes the estimation error variance (Mohamed and Antia, 1998; Kanevski, 2008).

The kriging family models have common features according to the basic principles of kriging. First, kriging weights do not depend on the variable values as they are defined by a spatial correlation structure described by the covariance or the variogram model. Second, kriging is an exact estimator in which the estimate honors the conditioning data exactly. Third, kriging features a smoothing effect on the estimates - the kriging estimate cannot exceed the data maximum or go below the data minimum. Smoothing is characterized by the variability of the kriging estimates. Fourth, the kriging variance is not higher than the variance of the initial data (Kanevski, 2008).

Kriging analysis of pollutant concentrations consists of five steps (Mohamed and Antia, 1998):

- Determine if the measured pollutant concentrations are additive and normally distributed. If they are not normally distributed, appropriate transformations are required
- If a trend, not adequately portraying the surface for a particular need, exists in data, remove it

- Estimate the spatial correlation of pairs of measured concentrations as a function of the distance and the direction of their separation, i.e., determine the experimental semivariogram
- Perform a structural analysis, i.e., fit a theoretical model to the experimental semivariogram
- Perform kriging (Mohamed and Antia, 1998).

4.2.1 Data Transformations Used in Environmental Applications

Measured data should follow a normal distribution in order to apply kriging technique to evaluate geospatial distribution of the data. A general Q-Q plot (a plot on which the quantiles of measured data versus the quantiles of a standard normal distribution are depicted) can be drawn to check normality. For identical distributions, the Q-Q plot will be a straight line with 45° inclination. If measured data departures from normality, it can be necessary to transform the measured values into a new scale (logarithmic, square root, angular, logit, power, arcsin, tangent, arctan, and cosine) in which the distribution is nearly normal (Webster and Oliver, 2007; Johnston et al., 2001). A normal distribution is not mandatory to obtain prediction maps via Ordinary Kriging (OK). However, kriging is the best predictor among the ones that are formed from weighted averages and all unbiased predictors when data is normally distributed (Johnston et al., 2001). Therefore, it is beneficial to apply transformation to obtain a normal distribution. Following data transformation, further analysis can be performed and, if necessary, results can be transformed back into the original scale later on. Based on literature and variables used in this study, logarithmic, square root, angular, logit, power, arcsin transformations were used for data which does not follow normal distribution based on Q-Q plot. These transformations are explained below.

Logarithmic transformation: The log transformation is very common for environmental data. If the transformation of data z_i ($i=1, 2, \dots, N$) to $\log(z_i)$

exhibit a normal distribution, then the variable is said to have a lognormal distribution (Webster and Oliver, 2007; Şen, 2002).

Square root transformation: Distributions that are strongly positively skewed will often normalize or become symmetric by taking logarithms. Less pronounced positive skewness can be removed by taking square roots (Webster and Oliver, 2007; Şen, 2002).

Angular transformation (ArcSin square root transformation): This is sometimes used for proportions in the range 0 to 1, or 0 to 100 if expressed as percentages. If p is the proportion, then transform will be expressed as $\sin^{-1}\sqrt{p}$ (Webster and Oliver, 2007). The effect of this transformation is to spread the end of the scale while compressing the middle, which can be quite useful for proportion data with positive skewness (McGarigal, 2012).

Logit transformation: If p is a proportion ($0 < p < 1$), then its logit transformation will be $\ln \{(p)/(1-p)\}$. Logit transformation can be quite useful for fitting linear models to sigmoid distributions of proportions (Webster and Oliver, 2007; Armitage et al., 2002).

Power transformation: This transformation can be used to stabilize variance and applied by taking the square or cube of the measured data (Şen, 2002).

ArcSin transformation: This transformation can be quite useful to transform the measured values with negative skewness to a distribution that is nearly normal (Şen, 2002).

Besides above approaches, transformations like tangent, arctan, and cosine can also be applied when the data does not follow a normal distribution. In this study, Q-Q plots were derived using ESRI's ArcGIS 9.3 and Easyfit Professional software. Besides visual inspection of the histogram and Q-Q plot of measured

DO values, Easyfit Professional software was used to statistically confirm if the data follows normal distribution or not. For this purpose, Kolmogorov-Smirnov statistic (D), quantifies the largest vertical difference between empirical cumulative distribution function of sample and theoretical cumulative distribution function, was calculated. The null hypothesis (H_0 : data follow the specified distribution) was tested according to this statistic. If D is smaller than the critical value for a given significance level (α) in standard tables, the null hypothesis was accepted. When H_0 is rejected at all predefined significance levels, the p-value is useful to know at which level it could be accepted. The p-value is calculated based on the test statistic, and denotes the risk to reject H_0 while it is true (Mathwave, 2013). If the data does not follow a normal distribution, transformations mentioned above were applied to find out the most suitable method for transformation of the data to obtain normality.

4.2.2 Identifying trends in data

Another assumption of variogram analysis is that there is no spatially determinant trend present in sample values. Sample values should be independent of geographic coordinates. If a significant trend is present, then it is likely to distort the variogram results to the extent that they may be unintelligible. Alternatively, the determinant trend may overwhelm the natural variability to the extent that the latter is undetectable (Houlding, 1994). If a trend exists in data, the value at a point was consisted of trend component value and fluctuation (residual) at that point (Equation 4.1).

$$z(x_i) = m(x_i) + R(x_i) \dots \dots \dots (\text{Eq. 4.1})$$

Where $z(x_i)$ is the value at point x_i , $m(x_i)$ is trend component value at point x_i and $R(x_i)$ is the fluctuation or residual.

The trend is the nonrandom (deterministic) component of a surface that can be represented by some mathematical formula. If the trend surface does not adequately portray the surface for a particular need, it should be removed and the analysis should be done for data with no trend (residual) (Johnston et al., 2001). The existence of a spatial trend in the data can be investigated by several ways:

- Plotting the variable of interest (z) against x and y coordinates
- Making a contour plot
- Making a wire diagram or other three-dimensional representation
- Fitting a loess (Local Regression) or other regression model
- Testing model significance
- Looking for spatial trends in the residuals (Schuenemeyer and Drew, 2011).

According to these investigations, if trend is flat, linear or quadratic, it represents no trend, first order trend and second order in the data, respectively. While equation 4.2 was used in the case of a first order trend was present, equation 4.3 was used for a second order trend. Then, trends were removed from data and residuals were used in variogram analysis.

$$m_{x_i} = m_{x,y} = a_0 + a_1x + a_2y \dots \dots \dots (\text{Eq. 4.2})$$

$$m_{x_i} = m_{x,y} = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy \dots \dots \dots (\text{Eq. 4.3})$$

Where x and y coordinates of point x_i in x and y directions, respectively.

In this study, the existence of a spatial trend was investigated by plotting DO values against 2-dimensional space (in x and y coordinate system), preparing a contour plot, and fitting a regression model. Trend analysis and removal were performed using ArcGIS 9.3.

4.2.3 Variogram Analysis

Variogram analysis consists of the experimental variogram constructed from the data and the variogram model fitted to the data. The experimental variogram measures the spatial correlation between data points as a function of separation distance and direction. If an experimental variogram is randomly distributed or does not show a pattern of a typical variogram with separation distance, it simply indicates no spatial correlation among data points, and therefore kriging analysis is not taken into account. The experimental variogram is constructed by averaging one-half the difference squared of the values of measured variable over all pairs of observations with the specified separation distance and direction. The variogram ($\gamma(h)$) is calculated as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i + h) - Z(x_i))^2 \dots\dots\dots (\text{Eq. 4.4})$$

where $\gamma(h)$ is the variogram for lag distance h , $N(h)$ is the number of pairs separated by the lag distance h , $Z(x_i)$ and $Z(x_i+h)$ are the values of measured variable at spatial locations i and $i+h$, respectively (Barnes, 2006; He et al., 2010b; Ouyang et al., 2003).

The variogram model is chosen from a set of mathematical functions. The appropriate model is chosen by matching the curve generated by the experimental variogram with the curve of the mathematical function. To account for geometric anisotropy (variable spatial continuity in different directions), separate experimental and model variograms can be constructed for different directions in the data set (Barnes, 2006). The behavior of the variogram at distances comparable to the size of the domain determines whether the function is stationary or not. A function is considered as stationary if it consists of small-scale fluctuations (compared to the size of the domain) compared to some well-defined mean value. For such a function, the variogram should stabilize around a value,

called the sill, as shown in Figure 4.5. The sill is approximately equal to the variance of the data. For a stationary function, the length scale at which the sill is obtained describes the scale at which two measurements of the variable become spatially uncorrelated. This length scale is known as the range, correlation length or radius of neighborhood (Mohamed and Antia, 1998). A nugget is an apparent discontinuity in the experimental variogram near the origin caused by measurement errors or microscale variations. Theoretically, the value of the variogram must be zero at the origin. However, in the presence of a nugget, the variogram does not seem to approach zero at the origin, but rather have some positive value that is significantly larger than zero (Ouyang et al., 2003; Olea, 1991).

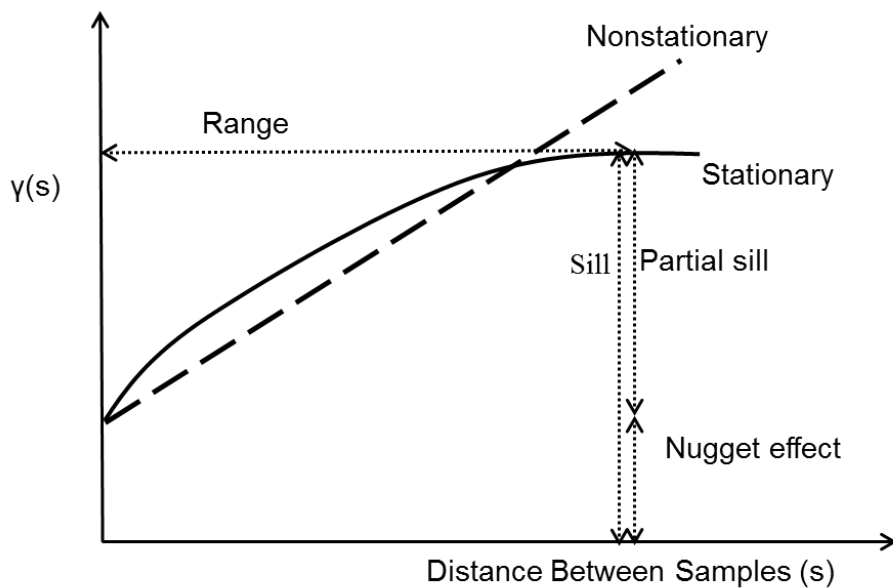


Figure 4.5 An example variogram model illustrating stationary and nonstationary behavior (Houlding 1994; Mohamed and Antia, 1998)

There are many possible variogram models. When combined with a nugget effect, one of the three models is adequate for most data sets: linear, exponential, and

spherical models (Barnes, 2006). Example representations for these three models are shown in Figure 4.6. If the experimental variogram never levels out, then the linear model is usually appropriate. If the experimental variogram levels out, but is "curvy" all the way up, then the exponential model should be considered. If the experimental variogram starts out straight, then bends over sharply and levels out, the spherical model is a good first choice (Barnes, 2006).

Experimental variograms of the DO and SC data were constructed by a script written in R Cran Version 2.14.1. Then, the variogram model was automatically fitted to the data. The script iterated over 19 different variogram models present in the software and picked the one that resulted in the smallest residual sum of squares with respect to the experimental variogram. Then, variogram model type, nugget, sill and range values were obtained from the selected variogram model.

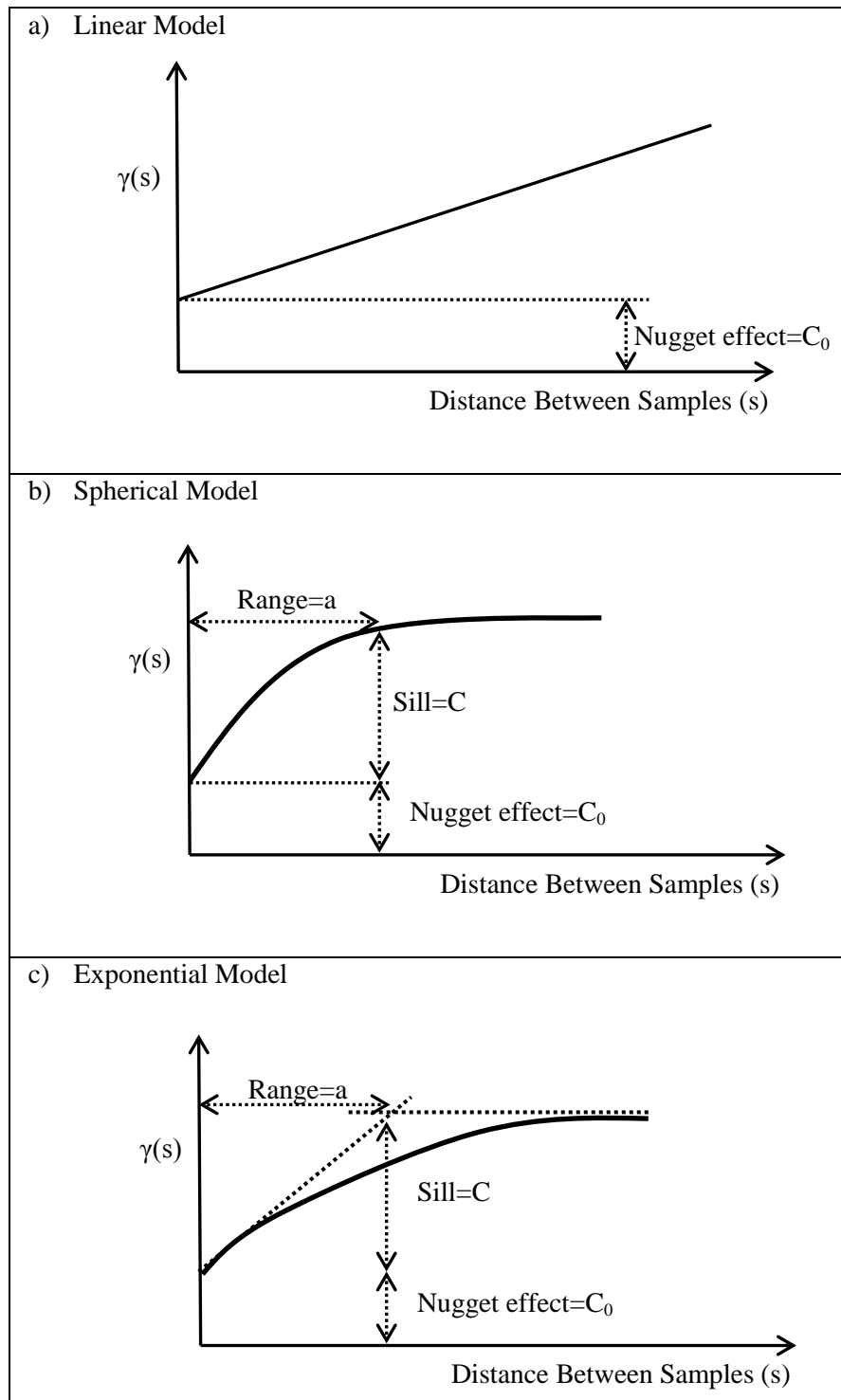


Figure 4.6 Frequently Used Variogram Models for Representing Spatial Variability (a: Linear Model, b: Spherical Model, c: Exponential Model)
(Houlding, 1994)

4.2.4 Ordinary Kriging (OK)

Kriging estimations were made through OK with the generated variogram model (Equation 4.5). Kriging covers a range of least-squares methods of spatial prediction such as ordinary kriging (OK), simple kriging, lognormal kriging, universal kriging, factorial kriging, ordinary co-kriging, indicator kriging, disjunctive kriging, probability kriging and Bayesian kriging. Among these, OK is the most robust and commonly employed method (Webster and Oliver, 2007). OK estimates the values of any measurable parameter at unmeasured locations using a weighted linear combination of available samples and creates a prediction surface. In OK, the mean is assumed to be constant over the field but unknown. The sum of the weights for estimation must add up to 1 to fulfill unbiasedness (Equation 4.6). Thus, the estimated value will be free of systematic error. The minimization of the estimation error variance is also achieved using OK (Ouyang et al., 2003; Schuenemeyer and Drew, 2011; Kanevski, 2008). OK estimator is given as:

$$Z^*(x_o) = \sum_{i=1}^{N(x_o)} w_i(x_o) z(x_i) \dots \dots \dots (\text{Eq. 4.5})$$

$$\sum_{i=1}^{N(x_o)} w_i(x_o) = 1 \dots \dots \dots (\text{Eq. 4.6})$$

Where $Z^*(x_o)$ is the OK estimate at location x_o , and $N(x_o)$ is the number of data from the neighborhood of (x_o) used for the estimation. $Z^*(x_o)$ is obtained as a linear combination of known values $z(x_i)$.

4.2.5 Kernel Density Estimation

KDE is one of the most common nonparametric methods to calculate a magnitude per unit area from point or polyline features in a defined kernel with a constant radius or bandwidth to obtain a continuous density surface. The features within the kernel are weighted according to their distance from the center of a kernel (Figure 4.7). Features near the center have a higher weight to contribute to kernel density estimation. In KDE maps, locations (clusters) with relatively higher or lower densities can be identified easily based on kernel density values (Plug et al., 2011; Kuter et al., 2011; Amatulli et al., 2007; De la Riva et al., 2004; Koutsias et al., 2004; Silverman, 1998; Wang and Wu, 2009).

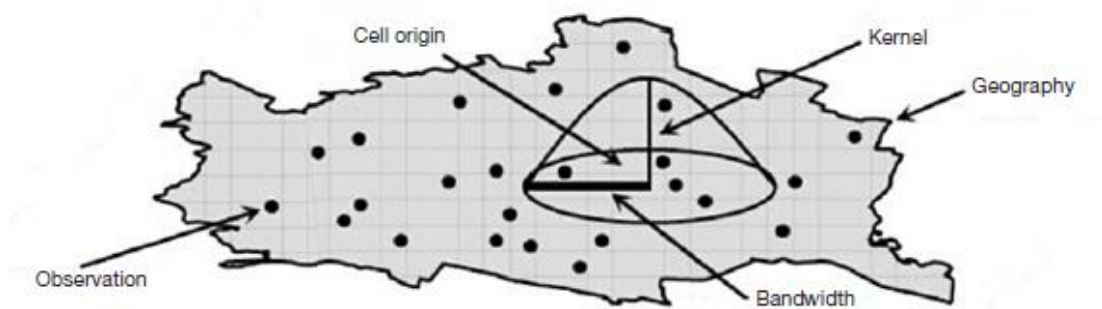


Figure 4.7 Kernel Density Estimation (Wilson, 2012)

KDE can be conducted both based on distance and based on population. KDE based on distance is mathematically defined as (Silverman, 1998):

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left\{\frac{1}{h}(x - X_i)\right\} \dots\dots\dots(\text{Eq. 4.7})$$

where n is the number of point observations, h is the bandwidth, K is the kernel function, x is a vector of coordinates that represent the location where the function

is being estimated, X_i is the vector of coordinates that represent each point observation i and d is the number of dimensions in space.

When a population field is used, KDE is mathematically defined as (Levine, 2010):

$$\hat{f}(x) = \sum_{i=1}^n \left\{ W_i * I_i * \frac{1}{h^2 * 2\pi} * e^{-\left[\frac{(x - X_i)^2}{2 * h^2} \right]} \right\} \dots\dots\dots (\text{Eq. 4.8})$$

where W_i is a weight at the location where the function is being estimated (observations closer to the location are weighted more highly than those farther away; observations outside the specified bandwidth are excluded) and I_i is an intensity (population) at the point location. In this study, selected water quality parameters were used as the population field in the calculation of KDE (Levine, 2010; Wilson 2012).

Selection of bandwidth (h) directly affects the smoothness of density patterns. There is no a straightforward rule to specify an appropriate bandwidth. When a very small bandwidth is selected, the result highlights individual points with inadequate smoothing (spiky). If bandwidth is too large, a smoother density surface will be obtained (Plug et al., 2011; Kuter et al., 2011; Brimicombe, 2010; Amatulli et al., 2007). Often selection of the bandwidth is performed subjectively by visual inspection (trial and error) of several density estimates over a range of bandwidths. Evaluation starts with a large bandwidth and progresses with decreasing the amount of smoothing until fluctuations that are more “random” than “structural” start to appear (Kuter et al., 2011; Wand and Jones, 1995).

In this study, a reduction strategy is proposed to identify the potential water quality sampling locations that can be removed in order to decrease the total sampling points. After the reduction, remaining sampling points can be suggested

for the monitoring program and estimation of the spatial water quality distributions in PDR through kriging (Figure 4.1). For this purpose, KDE was used and maps were constituted. With KDE maps, areas that have similar kernel densities and the sampling points within these areas were identified and put into the same cluster. Then based on land use information and locations of surface drainage and point source inlets, the sampling points that would be eliminated from the network were determined. During sampling point reduction it was made sure that at least one representative point retained within a cluster for kriging purposes in the procedure presented in Figure 4.1. If there is a point source near to a cluster where two sampling points are located, select the point which one is the most downstream as representative point. Furthermore, surface drainage inlets which are identified using topographic maps or hydrology toolbar of ArcGIS should be taken into account in the selection. In a cluster where two sampling points are located, again select the location in the most downstream of surface drainage inlet. The selected point should be away from streams or rivers inflow points and be accessible. By this way, several networks were generated to find the best combination.

A naming convention was used in order to represent different sampling networks with different numbers of sampling points. The “NXX” notation stands for the network that constitute of calibration data with “XX” number of sampling points eliminated. Eliminations were conducted based on KDE maps by reducing the number of sampling points by 4 or 5 at a time starting initially from 65 sampling locations. In this regard, “N04” and “N30” mean that 4 and 30 sampling points were removed from the initial calibration data set of 65 sampling points, respectively. As a result, kriging for N04 and N30 was carried out based on 61 and 35 sampling locations, respectively.

4.2.6 Evaluation

In the evaluation stage (Figure 4.1), several factors were considered to decide whether the given network was representative and conserved the spatial correlation structure. The variogram models for the initial network (N00) and networks with reduced sampling points (NXX) were compared through constructing scatter plots and checking slope and RMSE values to evaluate whether the spatial correlation structure of the initial network was conserved or not based on $\gamma(h)$. Kriged values were compared to validation data to check whether a given network was successful in determining the concentrations at validation points. Measured versus predicted values for validation locations were plotted. Moreover, several error metrics (mean errors, root mean square errors, etc.) were used in evaluations. The total dataset including measurements from 81 sampling locations was divided into two sub-datasets to make evaluation possible. While the data obtained from 80% of the sampling locations was used to calibrate the model, remaining 20% of the locations was selected as the validation dataset in 2D kriging estimations.

KDE maps, river inflow points, landuse should be taken into account in the selection of validation dataset in reservoirs/lakes. Criteria that can be used to compare the methods that are used for prediction of a given parameter are either local or global scale. The pattern of the map produced by each method is the local one. The maps can be compared based on the ability of the methods to reproduce local features by comparison with reference to spatial pattern. On the global scale, the methods can be evaluated on basis of their mean errors (ME), root mean square errors (RMSE), average kriging standard errors (ASE), mean standardized errors (MSE) and root mean square standardized errors (RMSSE) (Wibrin et al., 2006; Johnston et al., 2001). The Prediction Standard Errors quantify the uncertainty for each location in the kriging prediction surface. If data follows a normal distribution, the true value of the surface will be within the interval formed by the predicted value ± 2 times the prediction standard error in the 95 percent of

the time (Johnston et al., 2001). Summary statistics and graphs can be obtained by comparing the predicted value to the actual value from cross-validation or validation. Let $Z(s_i)$ be the predicted value from kriging, $z(s_i)$ be the observed value from validation, $\sigma(s_i)$ be the prediction standard error for location s_i , and n be the number of observations. Then some of the summary statistics are (Johnston et al., 2001):

$$ME = \frac{\sum_{i=1}^n (Z(s_i) - z(s_i))}{n} \dots\dots\dots (\text{Eq. 4.9})$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z(s_i) - z(s_i))^2}{n}} \dots\dots\dots (\text{Eq. 4.10})$$

$$ASE = \frac{\sum_{i=1}^n \sigma(s_i)}{n} \dots\dots\dots (\text{Eq. 4.11})$$

$$MSE = \frac{\sum_{i=1}^n (Z(s_i) - z(s_i)) / \sigma(s_i)}{n} \dots\dots\dots (\text{Eq. 4.12})$$

$$RMSSE = \sqrt{\frac{\sum_{i=1}^n ((Z(s_i) - z(s_i)) / \sigma(s_i))^2}{n}} \dots\dots\dots (\text{Eq. 4.13})$$

ME should be close to 0, RMSE and MSE should be as small as possible, and RMSSE should be close to 1 for a model providing accurate predictions (Johnston et al., 2001). The number of sampling locations in calibration dataset is reduced subsequently until finding the best combination with small errors (in terms of the metrics given above) while maintaining the spatial correlation structure of the calibration dataset. The entire evaluation criteria mentioned above was employed for 2D kriging

In 3D kriging estimation, dataset including measurements from 81 sampling locations was used as collected without dividing the data into subsets. Cross validation was used to identify the degree of accuracy that the semivariogram

parameters and the search neighborhood possess in predicting the unknown locations. In the cross validation procedure, each sample value C at a location x is temporarily removed in turn from the data set and a value C^* at that location is estimated using the remaining $(n-1)$ samples. Then estimated values were compared with the measured ones to evaluate measures of accuracy for the predictions generated using the OK model. The spread of the points should be as close as possible around the linear regression line in the observed versus predicted graph (Ouyang et al., 2003; Chehata et al., 2007; Jakubek and Forsythe, 2004).

CHAPTER 5

RESULTS AND DISCUSSION

5.1 2D Kriging

As it is mentioned before, 2D kriging estimation was conducted to determine the representative water quality sampling locations within PDR at 2D space for selected water quality parameters. As mentioned in section 4.1, raw data obtained from field work was preprocessed and divided into sub-data sets as calibration and validation sets for 2D kriging. Then, the methodology given in Figure 4.1 was followed.

5.1.1 KDE

KDE is the tool which was used to eliminate sampling points based on the methodology proposed in this study and also assess the representative water quality sampling locations in PDR. The analysis also helped to identify the clusters with respect to DO and SC. With population (DO and SC) based KDE, kernel density areas and the sampling points within these areas were identified based on initial network (N00 with 65 sampling points). Sampling point eliminations were conducted based on KDE maps by reducing the number of sampling points by 4 or 5 at a time starting initially from 65 sampling locations if the given network was representative and suggests further point reduction according to variogram models, kriging maps, kriging error maps and error metrics. Then during sampling point reduction it was made sure that some of the points retained within a given area of similar kernel densities in order to keep one or more representative sampling points for kriging purposes.

It has been already mentioned in the methodology part that selection of the bandwidth (h) of the kernel can strongly affect the smoothness of the density patterns. Hence, different bandwidths (250 m, 500 m, 750 m, 1000 m, and 1250 m) were applied in KDE of calibration dataset to find the most suitable bandwidth. While there was too much smoothing in higher bandwidths (>1250 m) due to including more water quality sampling locations in KDE, there are many spikes in lower ones (250 m, 500 m) both for DO and SC (Figure 5.1 - Figure 5.10). Therefore, the most suitable bandwidth was selected as 1000 m that exhibited sufficient detail to observe kernel density areas without obscuring local details both for DO and SC.

Figure 5.4 represents KDE map with 1000 m kernel bandwidth for DO. In the figure, population (DO) based kernel density values which were calculated from Equation 4.8 ranged between 0 and 58. Kernel density areas with zero value mean that there were no sampling locations within kernel bandwidth in these regions to estimate kernel density. However, areas with high kernel values around 58 signify the contribution of more sampling locations with high DO values to the calculation of kernel value at the given locations. Clusters which are areas that have similar values and sampling points within these areas can be easily detected with this bandwidth both for DO and SC (Figure 5.4 & Figure 5.9).

KDE was also performed based on distance with the selected bandwidth. When Figure 5.4 and Figure 5.11 were examined, similar clusters were observed except for a cluster around sampling locations 73 and 74 in the population based KDE. As a result, for incremental reduction of sampling locations, the h value of 1000 m and population based KDE were used. Sampling point eliminations were conducted at each specific kernel density area (cluster). As an example KDE map of DO for calibration data set (N00) was examined and it was seen that sampling locations 47 and 48 were located in the same density area (Figure 5.4). So, at least one of them should be kept as a representative point in the mentioned density area for kriging. Therefore, sampling location 48 was selected as a representative point

and sampling location 47 was removed for N04 (network with 61 sampling points) based on surface drainage inlet locations, point source inlets and landuse. In this way, KDE maps were obtained iteratively for each reduction in the number of water quality sampling locations until the best combination according to variogram models, kriging maps, kriging error maps and error metrics is found (Appendix D). Other sampling point eliminations were performed in a similar way to construct N04. Then KDE map of N04 was generated to determine the sampling locations to be removed for N10. Eliminated sampling points based on KDE at each network are represented in the maps given in Appendix D.

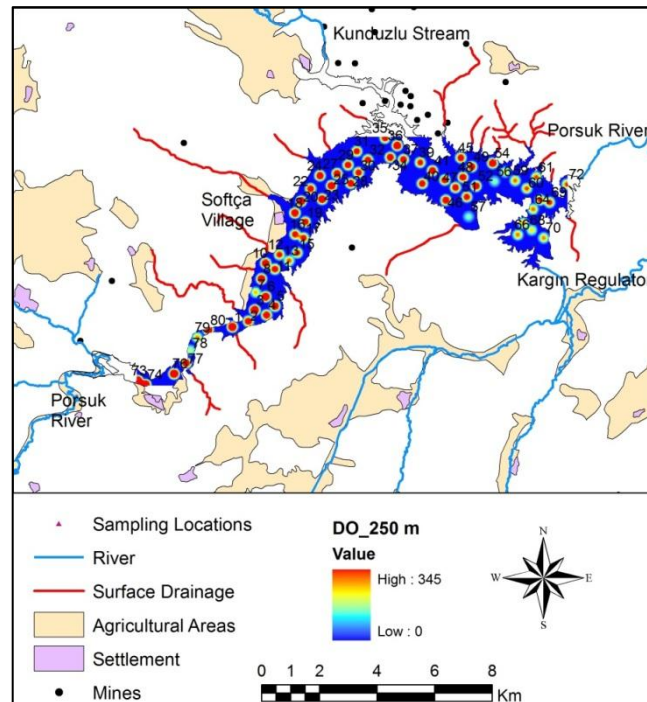


Figure 5.1 Application of 250 m kernel bandwidth for DO

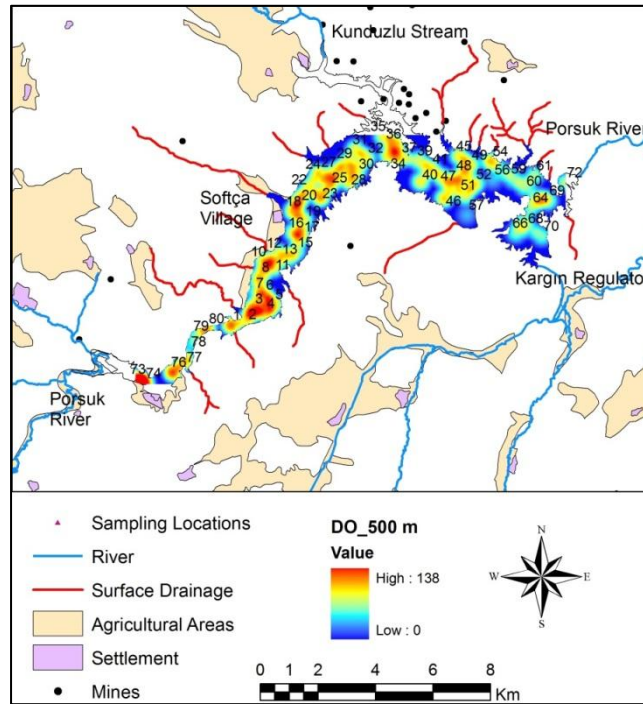


Figure 5.2 Application of 500 m kernel bandwidth for DO

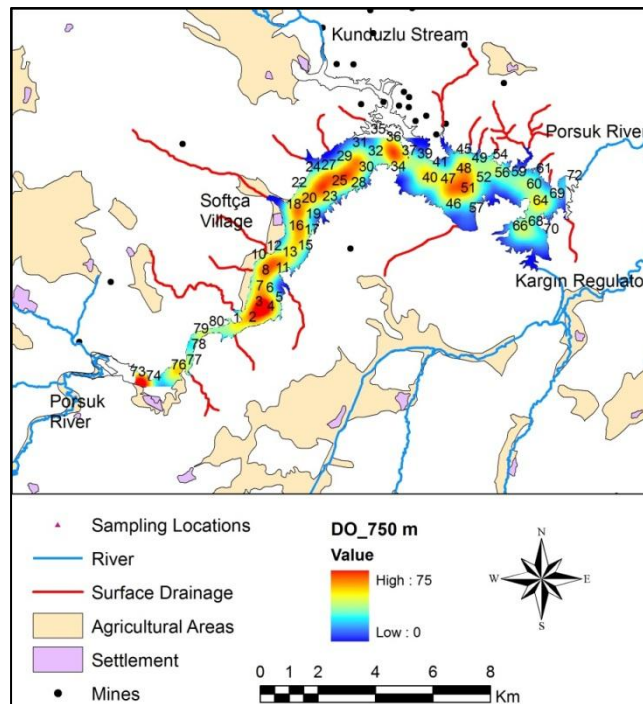


Figure 5.3 Application of 750 m kernel bandwidth for DO

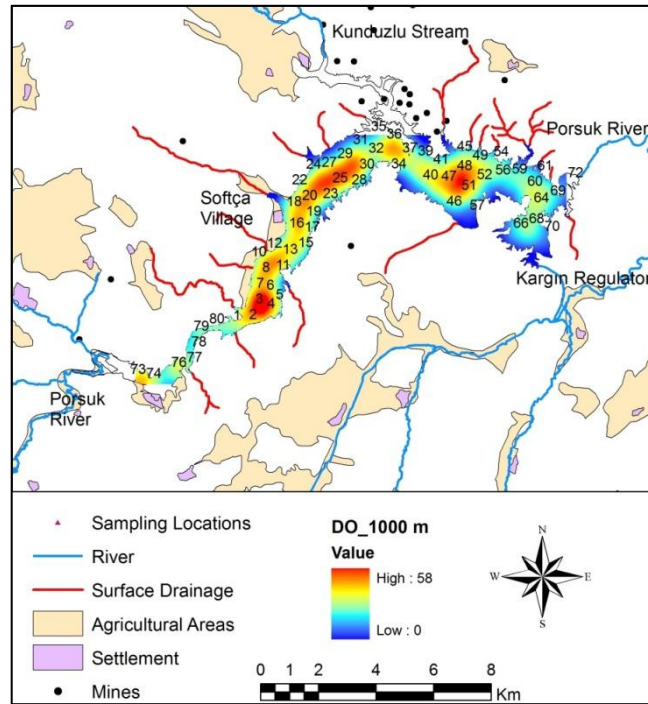


Figure 5.4 Application of 1000 m kernel bandwidth for DO

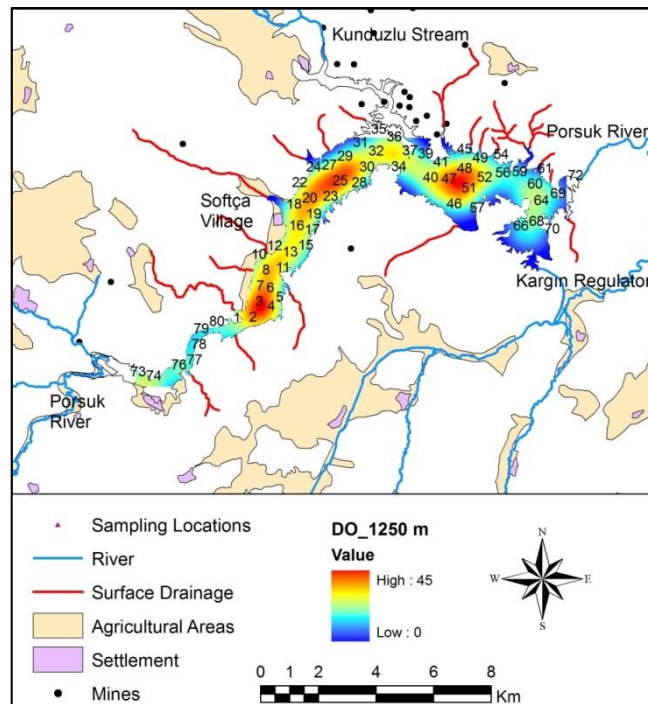


Figure 5.5 Application of 1250 m kernel bandwidth for DO

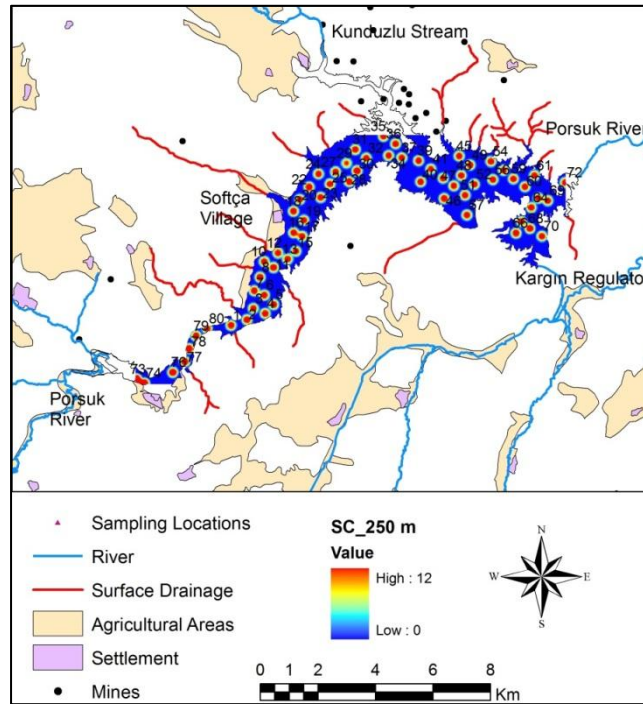


Figure 5.6 Application of 250 m kernel bandwidth for SC

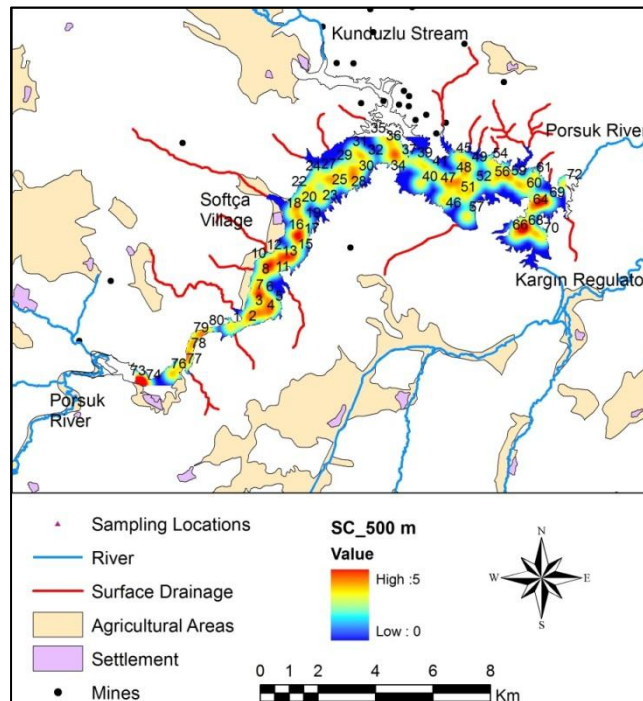


Figure 5.7 Application of 500 m kernel bandwidth for SC

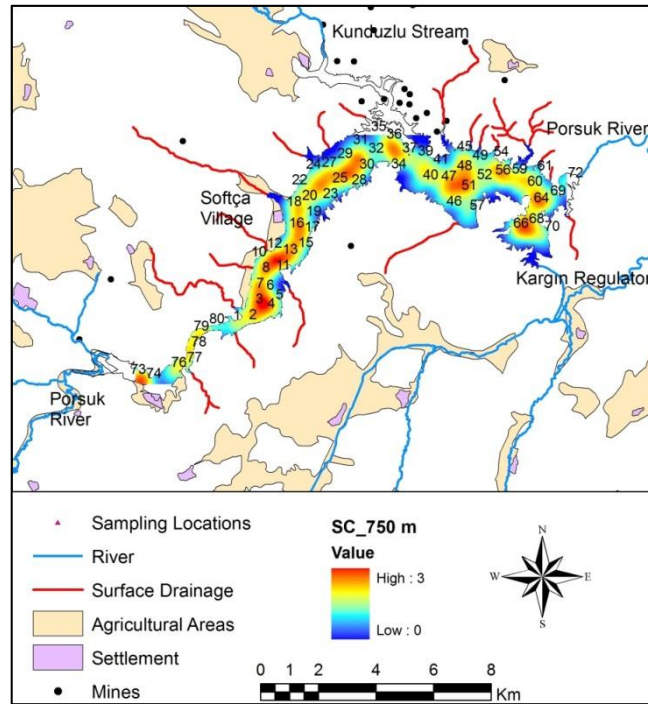


Figure 5.8 Application of 750 m kernel bandwidth for SC

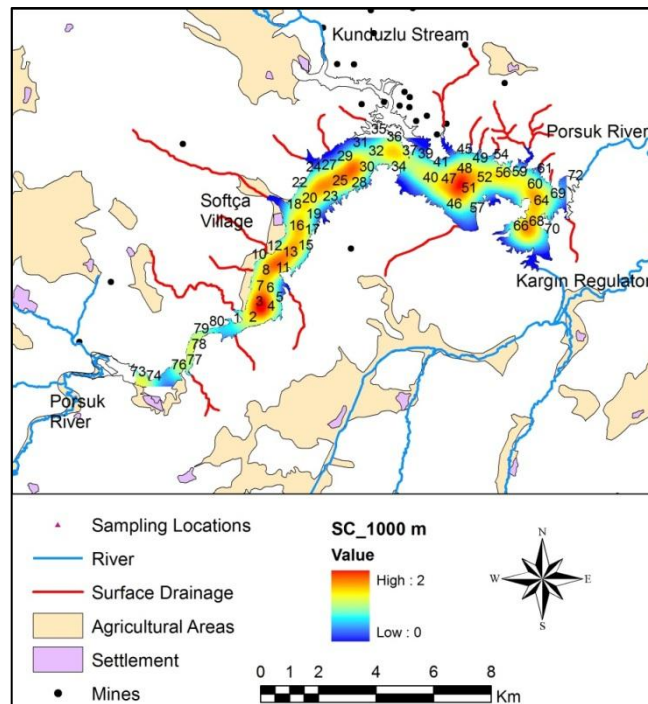


Figure 5.9 Application of 1000 m kernel bandwidth for SC

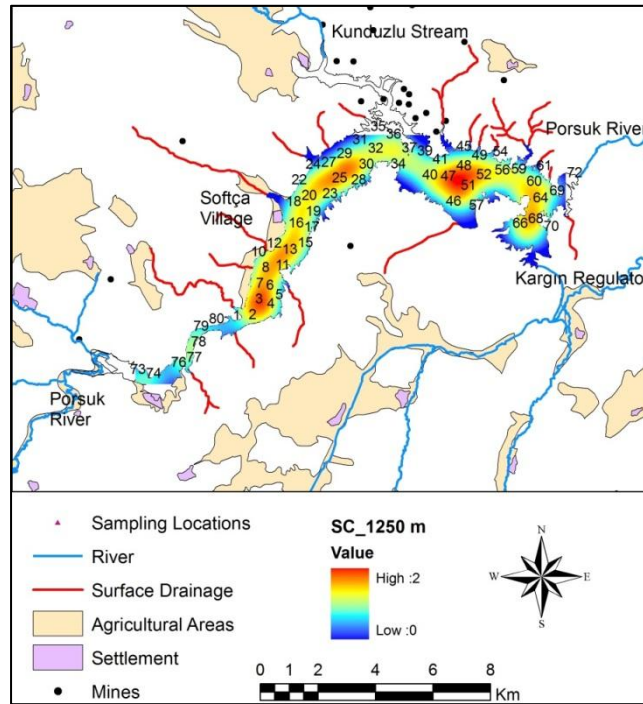


Figure 5.10 Application of 1250 m kernel bandwidth for SC

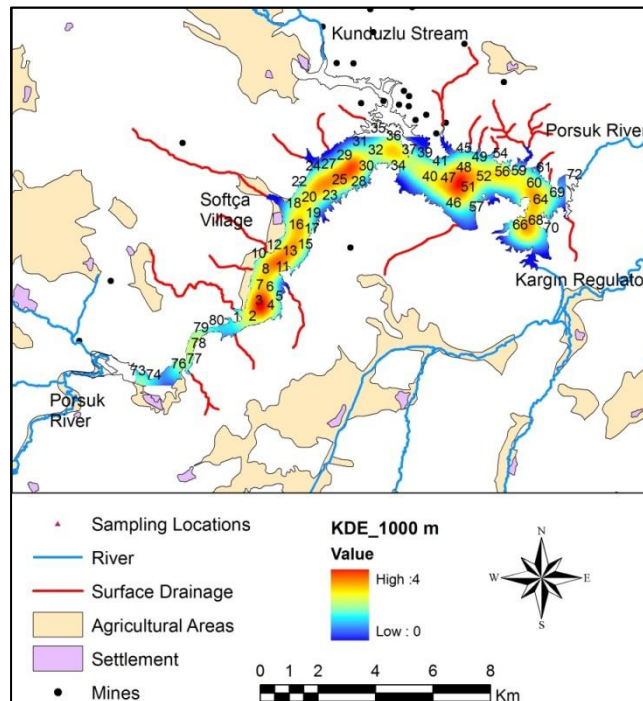
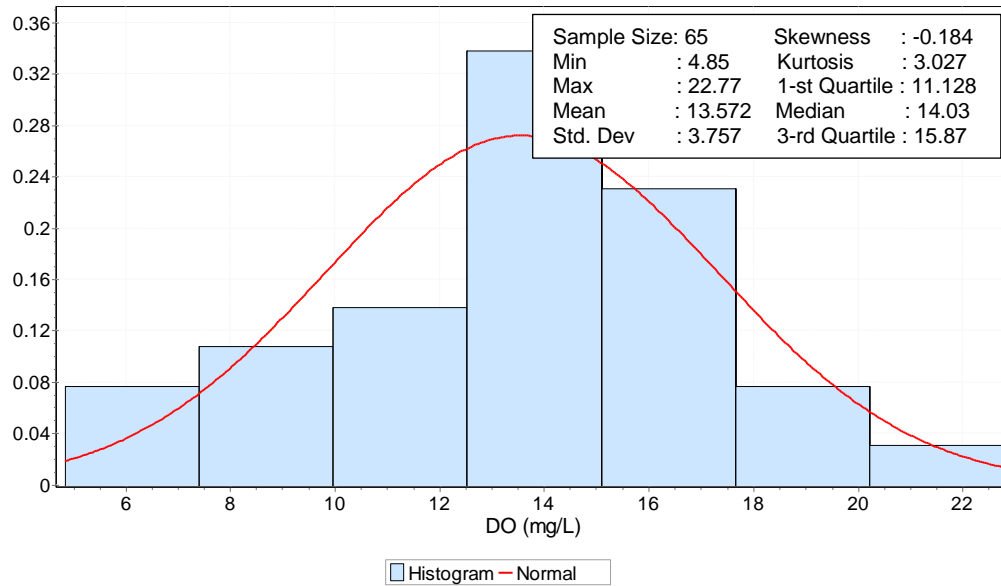


Figure 5.11 Distance based KDE with 1000 m kernel bandwidth

5.1.2 Normality Check

The second step of the approach suggested in this study is data exploration including normality check, trend analysis and variogram analysis. Firstly, normality check was conducted. As it is mentioned before, a normal distribution is not necessary to obtain prediction maps in OK process. However, Johnston et al. (2001) mentioned that kriging is the best predictor among not only predictors that are formed from weighted averages but also among all unbiased ones when data is normally distributed. Therefore, it is advantageous to build the analysis on normally distributed data. For this purpose, histograms and Q-Q plots were prepared to check the normality for DO and SC prior to kriging application. The DO concentrations measured at 81 sampling points in PDR ranged from 4.85 mg/L to 22.77 mg/L. The mean DO concentration was 13.43 mg/L with a standard deviation of 3.64 mg/L. High DO concentrations are a result of local algal abundance and oversaturation. Surface DO concentrations measured at calibration points (65 water quality sampling stations in Figure 4.4), on the other hand, ranged from 4.85 mg/L to 22.78 mg/L with a mean concentration of 13.57 mg/L and a standard deviation of 3.76 mg/L. Furthermore, the SC concentrations were between 0.43 mS/cm and 0.76 mS/cm with a mean concentration of 0.56 mS/cm and a standard deviation of 0.04 mS/cm. When the histogram and Q-Q plot of DO for the initial network (N00 with 65 sampling points) in Figure 5.12 were examined, it was obviously seen that the shape of the histogram (a fairly symmetrical distribution) and Q-Q plot (near to a linear line) were similar to a normal distribution. These were the indicators of normality. However, the shape of histogram and the Q-Q plot of SC for the same network slightly departed from normality (Figure 5.13).

a) Histogram



b) Q-Q Plot

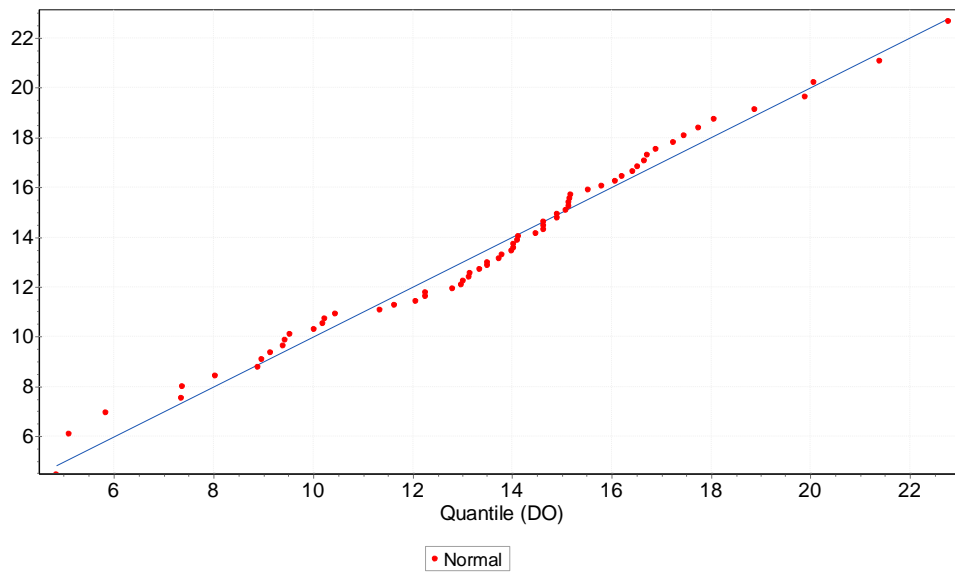
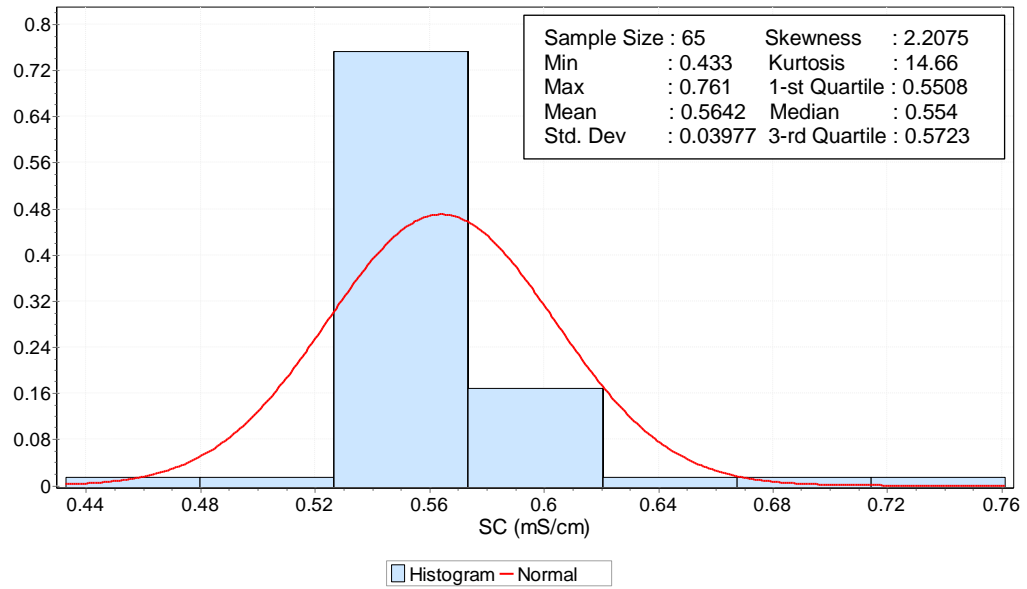


Figure 5.12 Histogram and Q-Q plot of DO for N00

a) Histogram



b) Q-Q Plot

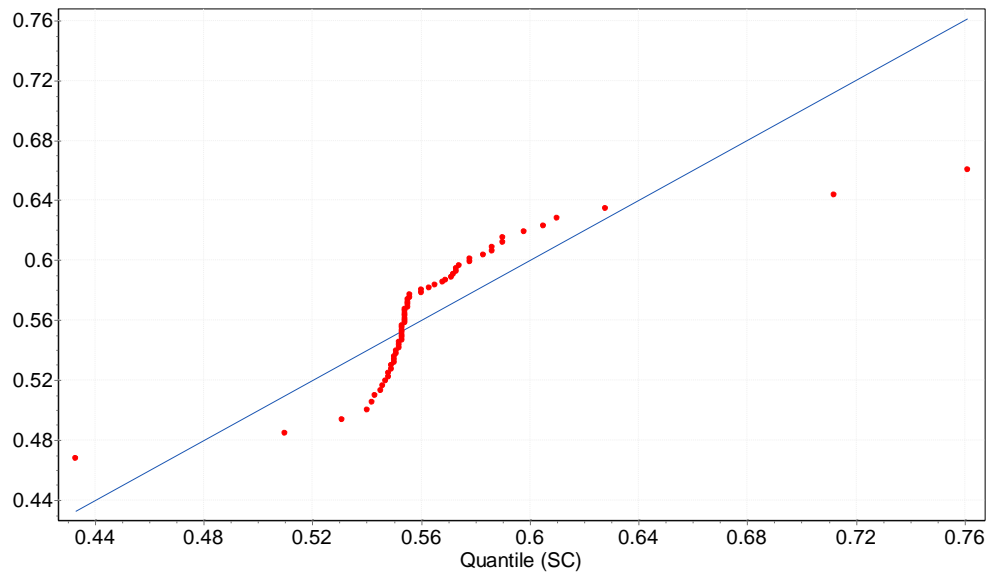


Figure 5.13 Histogram and Q-Q plot of SC for N00

The EasyFit Professional software was also used to statistically confirm whether the data follows a normal distribution. Firstly, D (Kolmogorov Smirnov statistic) was calculated. According to this statistic, the null and the alternative hypotheses were:

- H_0 : the data follow the normal distribution;
- H_A : the data do not follow the normal distribution.

In Table 5.1, D value for DO of N00 network (0.09897) was less than the critical values (0.131-0.199) stated in the standard table for all significance levels. Consequently, the null hypothesis was accepted for all significance levels. As a result, the D value provided additional confirmation about the normality in DO data besides visual inspection and DO data was used in the original (non-transformed) form. However, measured SC data does not follow the normal distribution since D value (0.22683) was greater than critical values (0.131-0.199) for all significance levels. Therefore, p-value was calculated for the SC data. When H_0 is rejected at all predefined significance levels, the p-value is useful to know at which level it could be accepted (Mathwave, 2013). As it was mentioned before the p-value denotes the risk to reject the H_0 while it is true. In Table 5.1, p-value for SC is 0.00202. Accordingly, the risk to reject the H_0 while it is true is 0.202 %. Therefore, it can be said that measured SC data does not follow the normal distribution (H_0 is rejected). Therefore transformations mentioned in section 4.2.1 were applied to obtain a normal distribution. Some of Q-Q plots for SC which were obtained after transformations applied were given in Figure 5.14. As it was seen from the figure that a better fit to normal distribution for SC was not obtained via the transformations when it was compared with Q-Q plot of original data (Figure 5.13). So, it was decided to use the SC data in the original (non-transformed) form.

Table 5.1 D and P Value for DO and SC calibration data (N00)

Sample Size		65				
DO	D-Value	0.09897				
	P-Value	0.51568				
SC	D-Value	0.22683				
	P-Value	0.00202				
α		0.2	0.1	0.05	0.02	0.01
Critical Value		0.131	0.149	0.166	0.185	0.199

a)

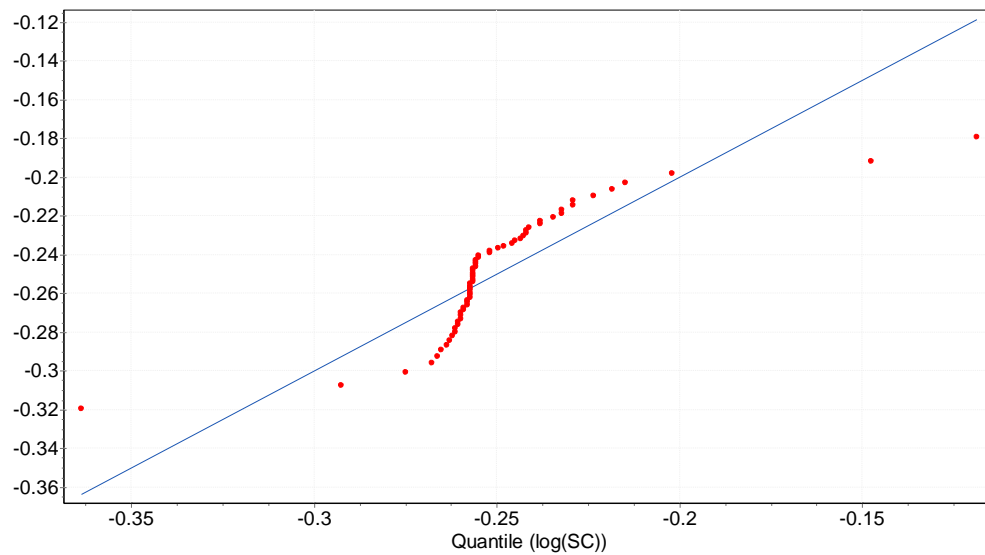
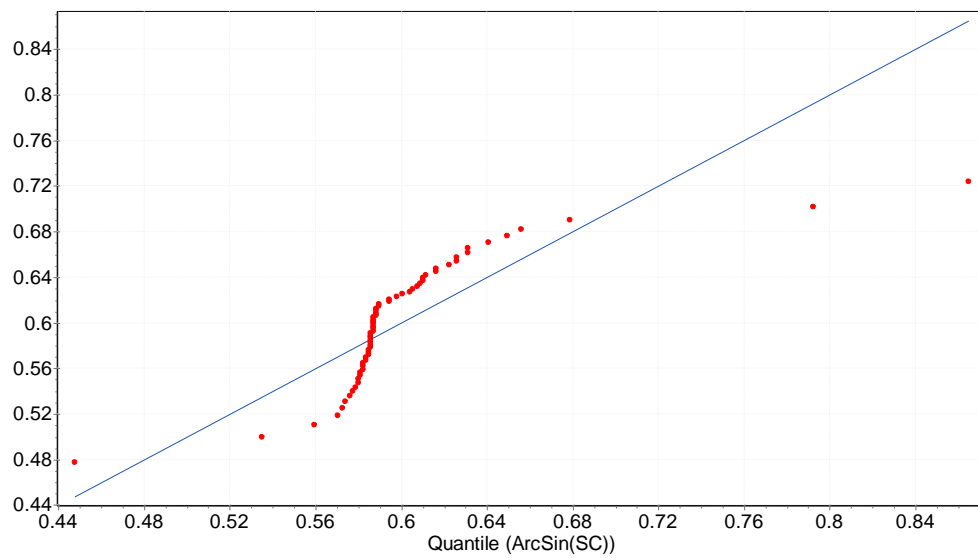


Figure 5.14 Q-Q Plots of SC Data after transformations were applied
(a: Logarithmic, b: ArcSin, c: ArcTan)

b)



c)

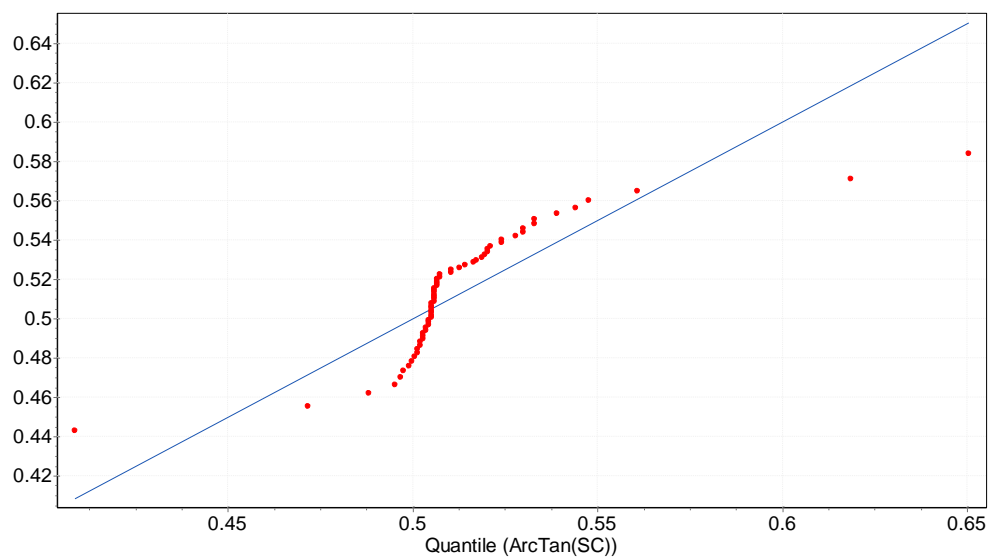


Figure 5.14 Q-Q Plots of SC Data after transformations were applied
(a: Logarithmic, b: ArcSin, c: ArcTan) (Continued)

5.1.3 Trend Analysis

The existence of a spatial trend in the data was also investigated in the data exploration part of the methodology. It can be determined by several ways such as plotting the variable of interest against x and y coordinates, making a contour plot, fitting a regression model (Schuenemeyer and Drew, 2011). The trend analysis plot of DO data for initial network (N00 with 65 sampling points) in 2 dimensional space ($DO(x,y)$) is depicted in Figure 5.15. Each vertical stick in the plot represents the location and the values of DO at each data point along the depth. The points are projected onto perpendicular planes. A best-fit line is drawn through the projected points, which model trends in specific directions. If the line is flat, this would indicate that there is no trend. However, the light green line in the Figure 5.15 starts out with high values and decreases from west to east to a certain distance until it levels out. This demonstrates that the data seems to exhibit first order trend in the east-west direction (Johnston et al., 2001). Moreover, the projection actually exhibits a U-shape in the north-south direction. Because the trend is U-shaped, a second order was used for the global trend (Johnston et al., 2001). The trend seen is possibly caused by the fact that the pollution is high at the entrance part of PDR (inlet of Porsuk River). High nutrient concentrations and low depth stimulate the growth of algal blooms in this region of PDR. In addition, a decreasing trend is observed from the inlet of PDR towards the outlet. Similarly, trend analysis plot of SC for N00 exhibits a U-shape both in the east-west direction and in the north-south direction (Figure 5.16). Trend analysis was performed for each subsequent network (N04, N10, N15, N20, N25, N30, N35, N40, N45, N50, N55) generated by reduced sampling locations. If existent, these trends were removed before kriging application via ArcGIS 9.3 using the approach presented in section 4.2.2. Then, residuals were used in variogram analysis.

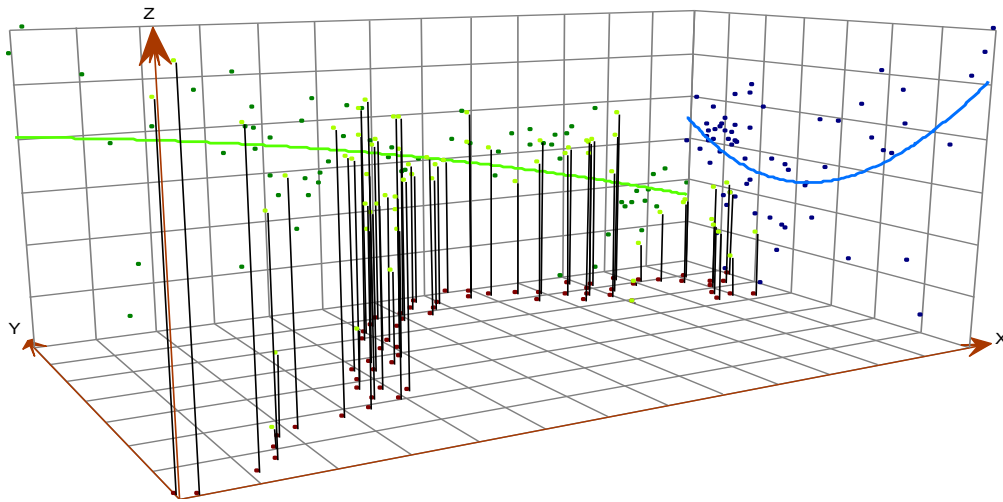


Figure 5.15 Trend Analysis for DO

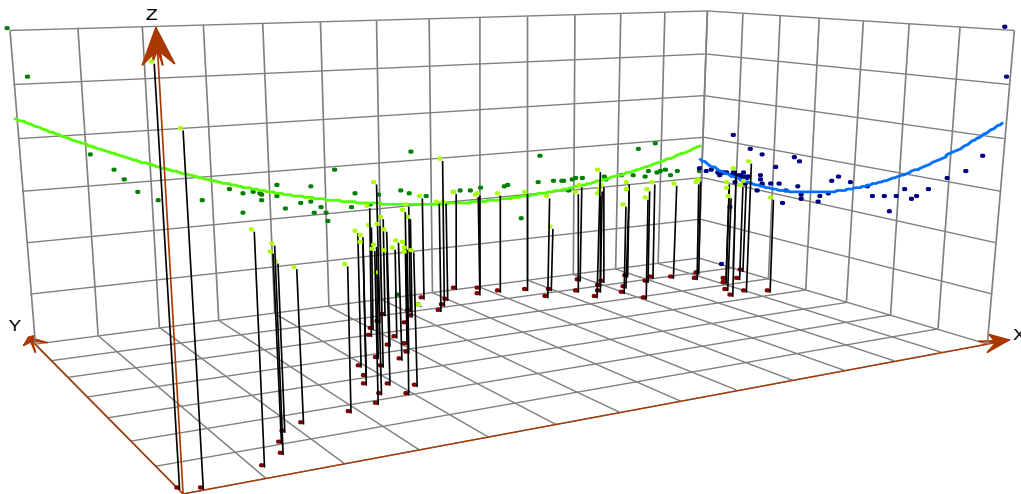


Figure 5.16 Trend Analysis for SC

5.1.4 Variogram Analysis

The last step before kriging estimation is the variogram analysis. As given in the methodology section, variogram analysis consisted of constructing experimental variograms for different sampling networks and fitting the variogram models.

Experimental variograms for DO and SC were created by R Cran Version 2.14.1 and then the variogram models were automatically fitted to the DO and SC data belonging to N00 network (calibration data with 65 sampling points) (Figure 5.17). Since the variogram of DO exhibited a typical variogram pattern (Figure 5.17), it was indicated that DO values in PDR were spatially correlated and kriging could be applied. Moreover, directional dependence was checked in variogram analysis. Variograms constructed in different directions indicated that omni-directional dependence (isotropic).

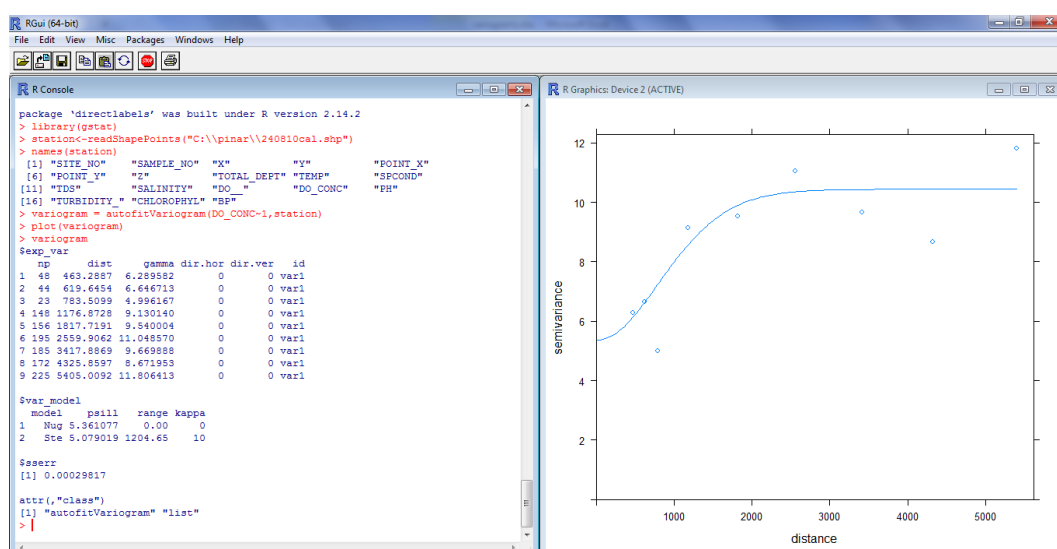


Figure 5.17 Variogram Fitting in R Cran

Variograms were generated for each subsequent sampling network (Appendix E) and it was checked whether the spatial correlation structure of DO was maintained with respect to the one for N00. Figure 5.18 provides the DO variograms for all networks. While x axis shows the distance between samples, y axis corresponds to variogram values which were obtained from fitted variogram models. The figure revealed that the spatial correlation structure of the DO calibration data set was significantly sustained until 30 water quality sampling locations were removed

(N30). The spatial correlation structure for DO was noticeably deteriorated with the elimination of 35 or more sampling locations.

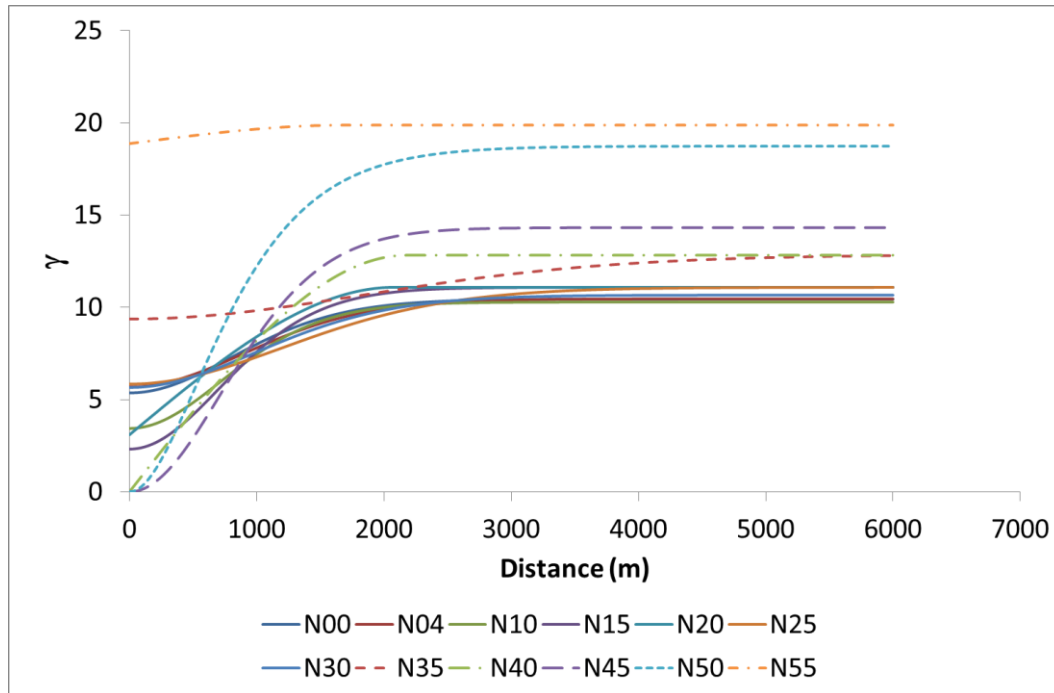


Figure 5.18 DO variograms for all sampling networks

Sill and nugget values were also checked. Sill (the sum of nugget and partial sill) defines the average maximum variance between a set of points (Jimenez et al., 2005). When sill and nugget values were compared with respect to N00 (Table 5.2), percent deviations increased significantly when 35 or more sampling points were removed (more than 75% for nugget and 23% for sill). The highest variance was observed for N55 (252% for nugget and 90% for sill). The deviations for N30 were about 6% and 2% for nugget and sill, respectively. The deviations in range values with the removal of sampling points did not exhibit a pattern. The range value for N30 was 1491 m. According to this result, sample locations separated by distances closer than 1491 m are spatially autocorrelated, whereas locations

farther apart than 1491 m are not. For most networks, non-zero nugget values were obtained. This may be due to microscale variations in DO which may be expected in eutrophic water bodies. In general, there were no significant differences in sill, nugget, and range values depending on the model type (spherical or gaussian) for DO in PDR. According to Table 5.2, the highest maximum variance was observed for N55.

Table 5.2 DO variogram model parameters for different sampling networks

Sampling Network	Number of Sampling Locations	Variogram				
		Model	Nugget (C ₀)	psill (C)	Sill (C ₀ +C)	Range
N00	65	Gaussian	5.36	5.08	10.44	1205
N04	61	Gaussian	5.68	4.77	10.45	1350
N10	55	Gaussian	3.43	6.85	10.28	1093
N15	50	Gaussian	2.31	8.77	11.08	1069
N20	45	Spherical	3.09	8.00	11.09	2081
N25	40	Gaussian	5.84	5.24	11.08	1820
N30	35	Gaussian	5.66	5.00	10.66	1491
N35	30	Gaussian	9.36	3.49	12.85	2793
N40	25	Spherical	0	12.82	12.82	2178
N45	20	Gaussian	0	14.32	14.32	1095
N50	15	Gaussian	0	18.73	18.73	1070
N55	10	Spherical	18.87	1.00	19.87	1680

In order to validate whether the spatial structure in DO was conserved or not for networks with smaller number of sampling points, the variogram models for N00 and others (NXX) were compared through constructing a plot which $\gamma(h)$ of N00 variogram model versus $\gamma(h)$ of NXX variogram model was drawn. All plots

were provided in Appendix F. The slope of the graph and RMSE values within 95% confidence level are given in Table 5.3. If the slope is close to 1, then it can be concluded that the variogram model of a network (NXX) is similar to the variogram model of N00. This can be achieved if the correlation structure in DO is maintained for networks with lower sampling points (NXX). The slope values presented in Table 5.3 are significant at 95% confidence level. As defined by slope and RMSE values, spatial correlation structure is conserved until 35 or more sampling points are removed. The best fit was obtained for N30 with a slope of 0.996 and an RMSE value of 0.269. Starting with N35, slopes deviated from 1.0 and RMSE values increased. The deviation in the slope was less than 6% for networks that maintained the spatial correlation structure. For N35, the deviation was 16% and increased up to 52% (for N55) as more sampling points were removed.

Nugget-to-sill ratio was also used to classify the spatial dependence. Sun et al. (2003) stated that the variable have a strong, moderate and weak spatial dependence if the ratios are less than 25%, between 25 and 75% and higher than 75%, respectively. N00 network represents a moderate spatial dependence for DO with a nugget-to-sill ratio of 51. Similarly, N30 represented a moderate spatial dependence with a ratio of 53. However the ratio increased to 73 for N35 and reached to 95 for N55. The nugget-to-sill ratio shifted from moderate to weak spatial dependence for DO starting with N35.

Table 5.3 Relationships between the DO variogram models for N00 and different networks (NXX)

Sampling Network	Number of Sampling Locations	Slope	RMSE
N04	61	1.003	0.129
N10	55	1.027	0.554
N15	50	0.959	0.960
N20	45	0.942	0.479
N25	40	0.980	0.514
N30	35	0.996	0.249
N35	30	0.837	0.867
N40	25	0.829	1.322
N45	20	0.747	1.716
N50	15	0.571	1.486
N55	10	0.483	1.437

The same procedure was applied for SC as well. Although the spatial correlation structure of the calibration data set for SC was significantly protected with the removal of up to 35 sampling points (N35) as for DO, it was seen from the fluctuations in Figure 5.19 that SC was more sensitive to the variations in spatial correlation structure. As can be seen in Table 5.4, there is no a pattern in the deviations of variogram model parameters with the removal of sampling points. The range value for N30 was 996 m. For the same network SC exhibited a shorter range value with respect to the range value for DO. According to this result, sample locations separated by distances closer than 996 m are spatially autocorrelated.

The information such as type of variogram model, nugget, sill and range values obtained from variogram analysis was used as input values to variogram analysis included in ArcGIS 9.3 software to conduct kriging analysis.

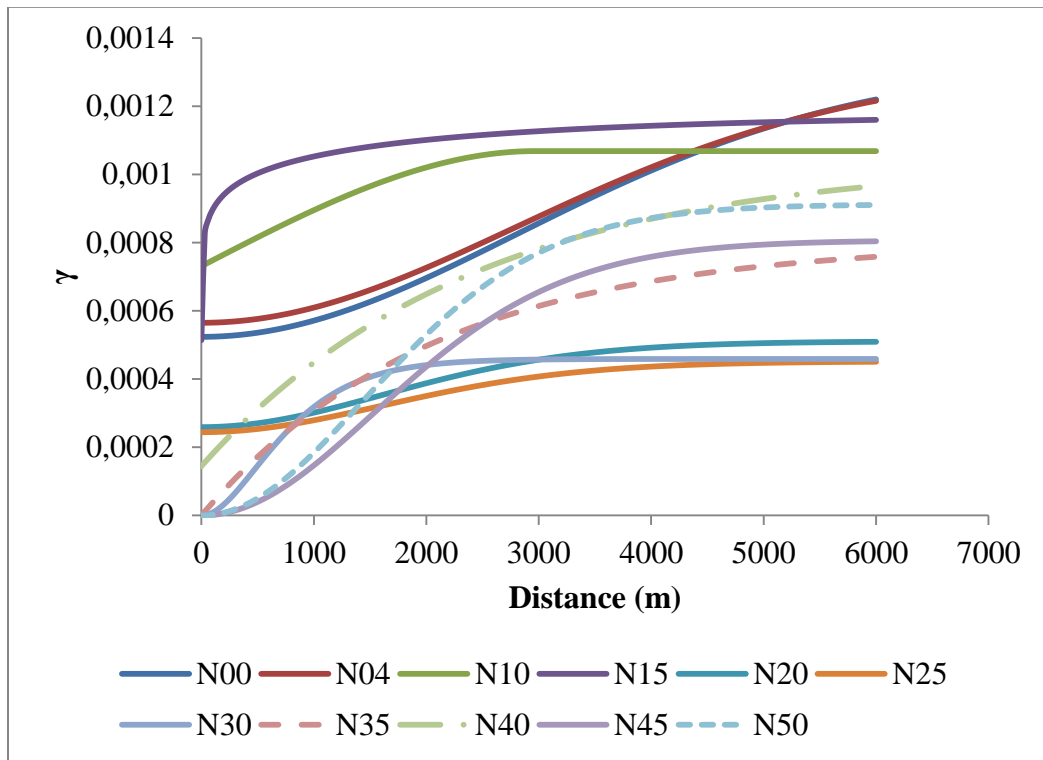


Figure 5.19 SC variograms for all sampling networks

Table 5.4 Variogram models and parameters for SC

Sampling Network	Number of Sampling Locations	Variogram				
		Model	Nugget (C ₀)	psill (C)	Sill (C ₀ +C)	Range
N00	65	Gaussian	0.00052	0.00081	0.00133	4272
N04	61	Gaussian	0.00057	0.00076	0.00133	4272
N10	55	Spherical	0.00073	0.00034	0.00107	2967
N15	50	Gaussian	0.00028	0.00090	0.00118	1717
N20	45	Gaussian	0.00026	0.00025	0.00051	2438
N25	40	Gaussian	0.00024	0.00021	0.00045	2438
N30	35	Gaussian	0	0.00046	0.00046	996
N35	30	Gaussian	0	0.00080	0.00080	2931
N40	25	Exponential	0.00014	0.00090	0.00104	2419
N45	20	Gaussian	0	0.00081	0.00081	2335
N50	15	Gaussian	0	0.00091	0.00091	2204
N55	10	Spherical	0.00089	1	1.00089	1680

5.1.5 Kriging

OK method was used to estimate water quality parameters (surface DO and SC) to determine the potential hotspots (problematic zones in terms of water quality with high/low DO and SC concentrations) and to find representative water quality sampling locations generated for using KDE maps and variogram models as well. Generated variogram model and parameters in the previous section were used as input values to conduct OK. Firstly, kriging maps produced through OK were evaluated by taking the possible pollutant sources into account. When surface DO distribution in PDR given in Figure 5.20 for N00 is examined, three zones with high DO concentrations (shown by black circles) and one zone (shown by red

circle) with low DO concentration were identified. The zones with lowest and highest concentrations can be critical for environmental evaluations. Low DO may hinder ecological activity and adversely impact aquatic species. According to Table 1 of the Turkish Water Pollution Control Regulation (Su Kirliliği Kontrolü Yönetmeliği, Tablo1, Kıtaçi Su Kaynaklarının Sınıflarına Göre Kalite Kriterleri) given in Appendix B, a DO value greater than 8 mg/L corresponds to Class I (high quality water which is suitable to provide potable water only with disinfection, for recreational purposes, for trout production, for animal production and farming needs, and for other purposes). If DO is between 6-8 mg/L and 3-6 mg/L, the water body is classified as Class II (slightly polluted water which is suitable to provide potable water through advance or appropriate purification, for recreational purposes, for fish production other than trout, for irrigation water and for all uses other than Class I) and Class III (polluted water which can be used for providing industrial water following an appropriate purification, excluding industries requiring quality water such as food and textile), respectively. DO less than 3 mg/L indicates Class IV (highly polluted water) quality. Furthermore, the limit values for eutrophication control given in Table 2 of the same regulation states that while the limit value for natural conservation areas is 7.5 mg/L, it is 5 mg/L for use for various purposes. When surface DO concentrations in PDR are examined with respect to the criteria presented in the regulation, it can be concluded that PDR is in good conditions since DO values in PDR are high. Yet, excess DO (oversaturation) can be an indication of poor quality as well which results due to excess algal productivity (eutrophication problem). In fact, high DO values observed in PDR is a result of eutrophication. This was also confirmed by high TP and Chl-a concentrations (above the hyper-eutrophic range) in PDR (Table 3.9). As seen in Figure 5.20, oversaturated DO concentrations around 22 mg/L was observed at the Porsuk River inlet upstream. This is not surprising because Porsuk River transports high loads of nutrients and pollutants to PDR and cause eutrophication. At the inlet, depth is really shallow (around 1 m) in comparison to the rest of the reservoir. This stimulates the growth of algae and results in high chlorophyll-a and DO concentrations.

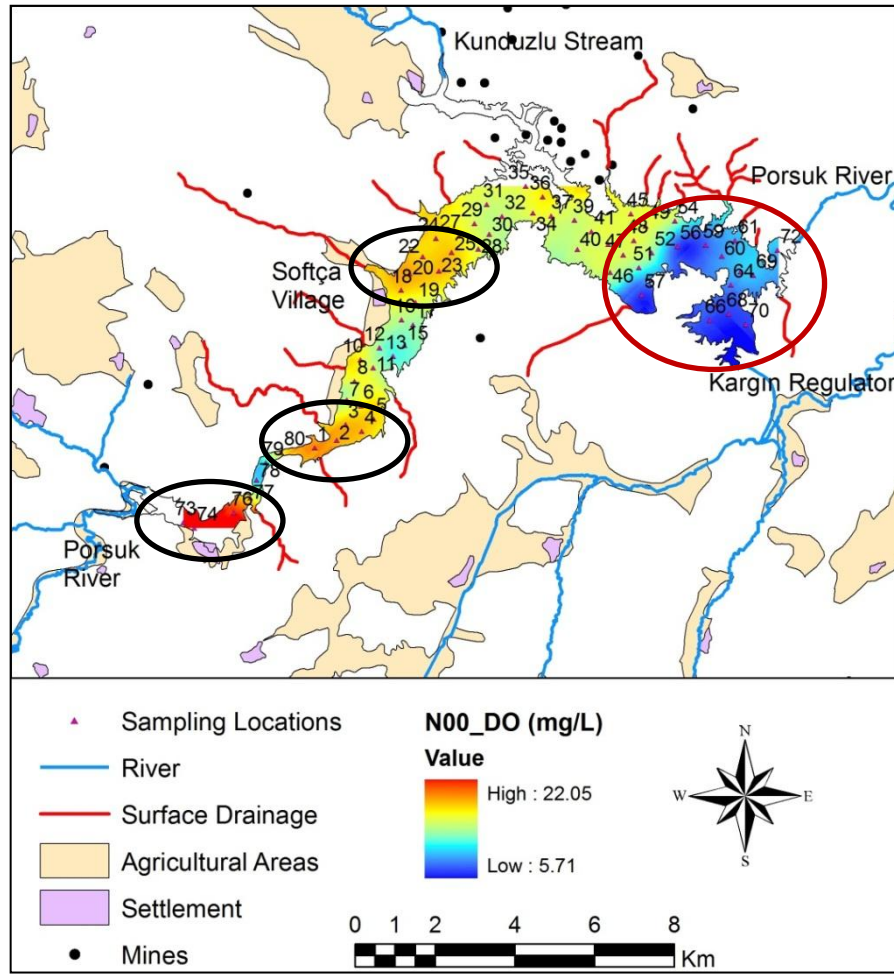


Figure 5.20 OK of DO for N00 (calibration data set)

High DO clusters can also be seen around Softça Village (around sampling location 18) where the wastewater of the village was discharged into the reservoir. The same area was also subject to agricultural runoff from the fields around the village. DO concentrations decrease to acceptable levels in the east of PDR (8-10 mg/L). This can be as a result of deeper depths (up to 32 m) and mechanisms in PDR affecting the fate of pollutants. PDR acts as a treatment medium. It is obvious from the figures provided in Appendix C that the water quality improved from inlet to the outlet. For example, while BOD exhibits higher values than A3 class (drinkable water following physical, chemical and advanced treatment and disinfection) in the inlet, it falls into A1 class (drinkable water following physical

treatment and disinfection) in the outlet (Figure C. 2). Furthermore, algal productivity decreases due to reduction in nutrients (Figure C. 9).

Proceeding the general evaluation of the kriging map derived for N00, kriging maps were also constituted for other networks with reduced number of sampling points (N04, N10, N15, N20, N25, N30, N35, N40, N45, N50 and N55). When the kriging maps provided in Figure 5.20 and Figure 5.21 were compared, it was seen that similar hotspots were obtained for N00 and N04 dataset (removal of 4 water quality sampling locations from the initial calibration dataset – N00). The deviations of minimum, mean and maximum DO concentrations for N04 compared to the prediction surface of N00 were less than 1%. This showed that there was no big difference in terms of predictions. When kriging maps of DO for all networks (Appendix G) were examined, it was noticed that the low and high DO regions identified in N00 were conserved up to the removal of 35 sampling points (N35). However, relatively more smoothing was obtained for N40 to N55. Furthermore, hot spots observed in N00 were less distinctive when more than 35 sampling points were removed. Moreover, the range of DO values in prediction surface increased for N55, indicating overestimation in some regions while underestimations in some other.

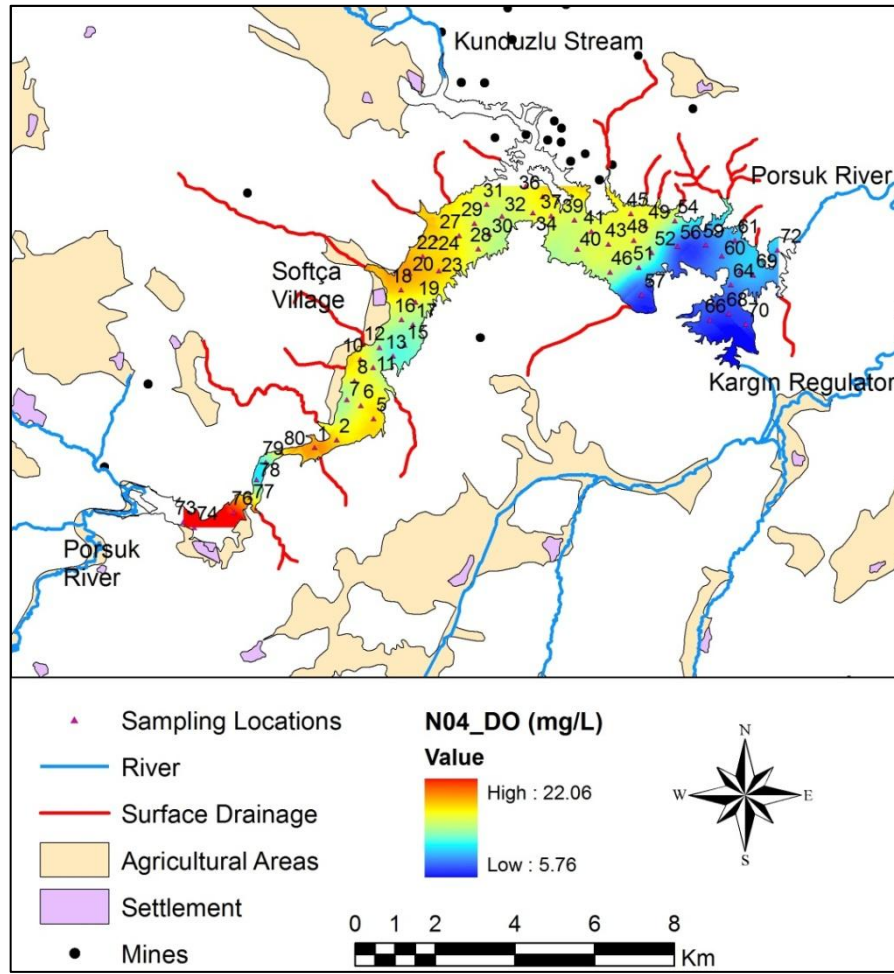


Figure 5.21 OK of DO for N04

Standard error maps were also constituted for all networks based on the validation data set (data obtained from 20% of the 81 sampling locations) to see the errors associated with the kriging predictions (Appendix G). While the highest standard errors (around 3.7 mg/L) in DO were observed in the shoreline of PDR, lowest standard errors (around 2.7 mg/L) were observed at the sampling locations for N00 and N04 networks (Figure 5.22 & Figure 5.23). When SEMs for all networks were examined (Appendix G), it was seen that when up to 30 sampling points were removed, the maximum standard error remained the same (about 3.7 mg/L). For N35 the value slightly increased to 3.8 mg/L. However, as more sampling

points were removed, the maximum standard error in DO values increased (5.0 mg/L for N55). As well as the standard error magnitudes, the total area of high standard errors expanded.

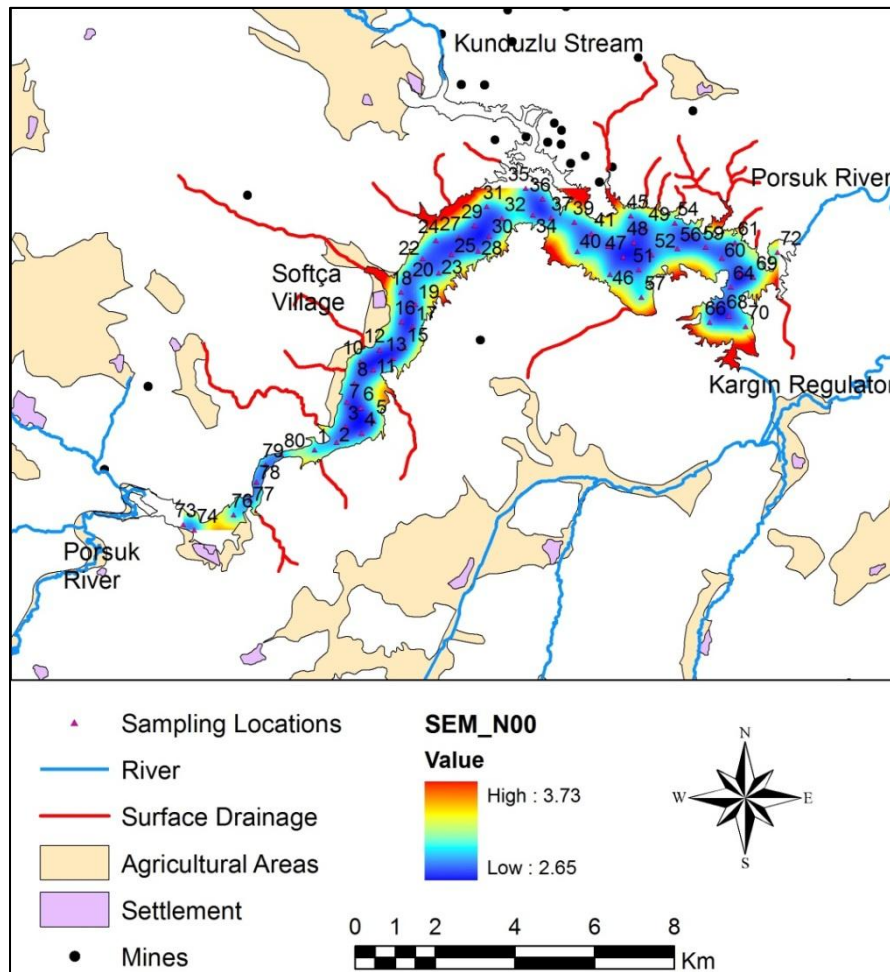


Figure 5.22 SEM of DO for N00

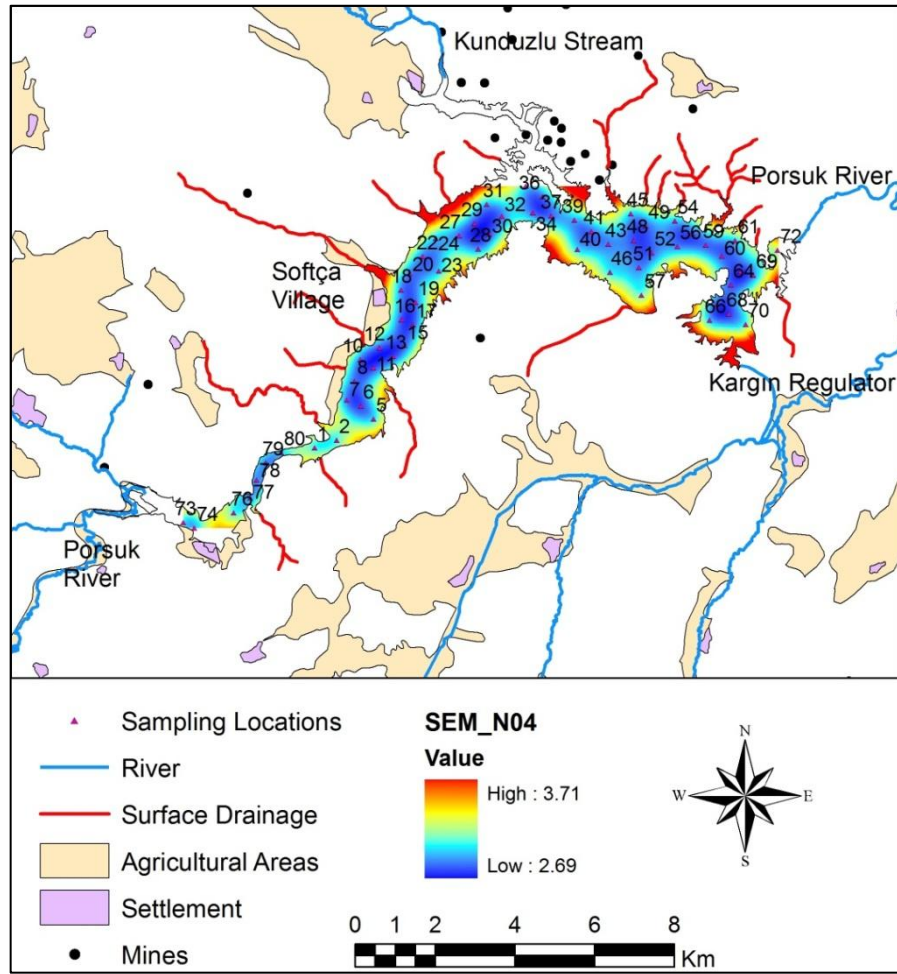


Figure 5.23 SEM of DO for N04

SEM of N00 was also subtracted from SEM of N04 to clearly indicate the locations where standard errors increase or decrease following the elimination of 4 sampling points. Figure 5.24 depicts that while standard errors increased especially in the locations where water quality sampling stations were removed, errors decreased in some parts of the PDR shoreline. Difference of SEMs of DO for N00 and other networks were given in Appendix G (Figures G25-G35).

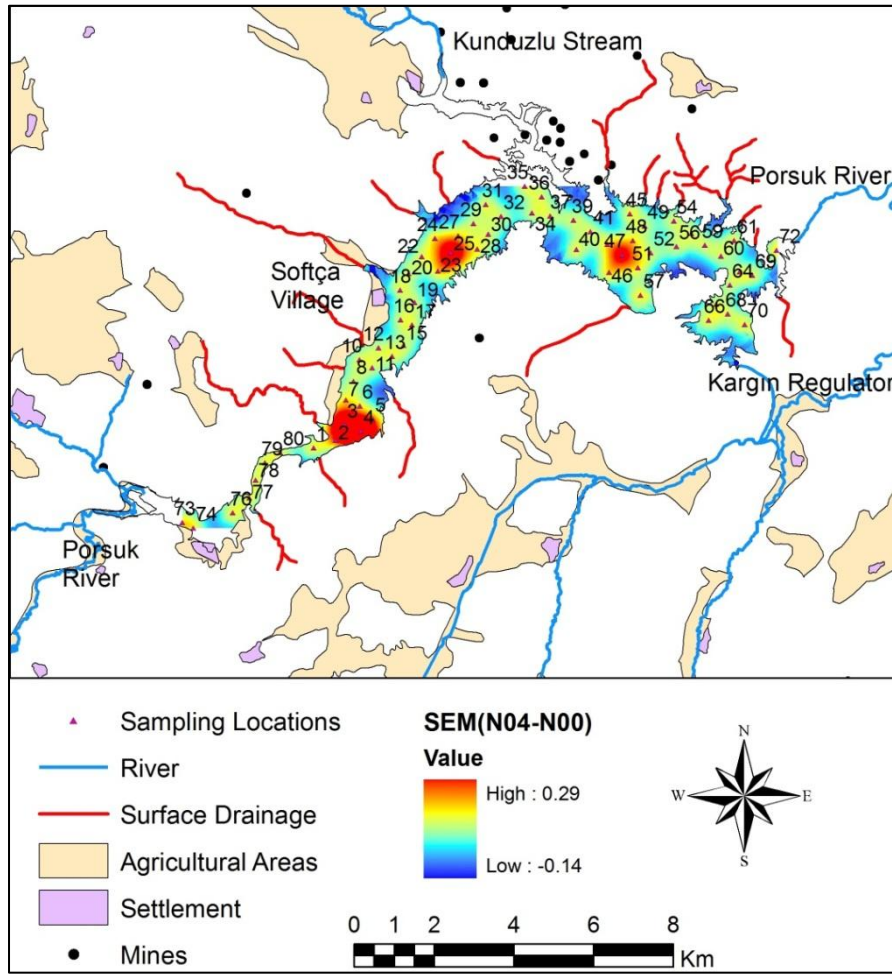


Figure 5.24 Difference of SEMs of DO for N04 and N00

Another way to evaluate the quality of kriging map is to compare the predicted values obtained from kriging map of the network at validation locations with the measured ones. DO concentrations at validation locations were predicted by creating a prediction surface via OK based on measured values at sampling locations of each network. This step was called as calibration. Moreover, the comparison of the predicted values at validation locations with the measured ones constituted validation step. The spread of the points should be as close as possible around the dashed gray line in the observed versus predicted DO graph given in Figure 5.25a and Figure 5.25b. In these figures most of the points are close to the dashed gray line except for two points (sampling locations 75 and 81) shown

within red circles. At these points measured DO concentrations were nearly half of the DO concentrations in the adjacent sampling stations. Sampling location 75 was so near to a village which discharged its domestic wastewater directly into PDR. This may be the reason of lower DO value in the location with respect to its neighbors. The lower DO value in sampling location 81 than the adjacent sampling locations may be a result of measurement error, morphological structure of the reservoir and unidentified pollution sources. High DO values in the adjacent sampling stations caused overestimations in predictions at the mentioned locations. When these two sampling locations were removed from the validation dataset, a better result was obtained (Figure 5.26). However, the mentioned locations were kept in the validation dataset to avoid biased sampling point selection and to see the success of kriging predictions in the presence of the mentioned points. All measured vs. predicted graphs were given in Appendix I.

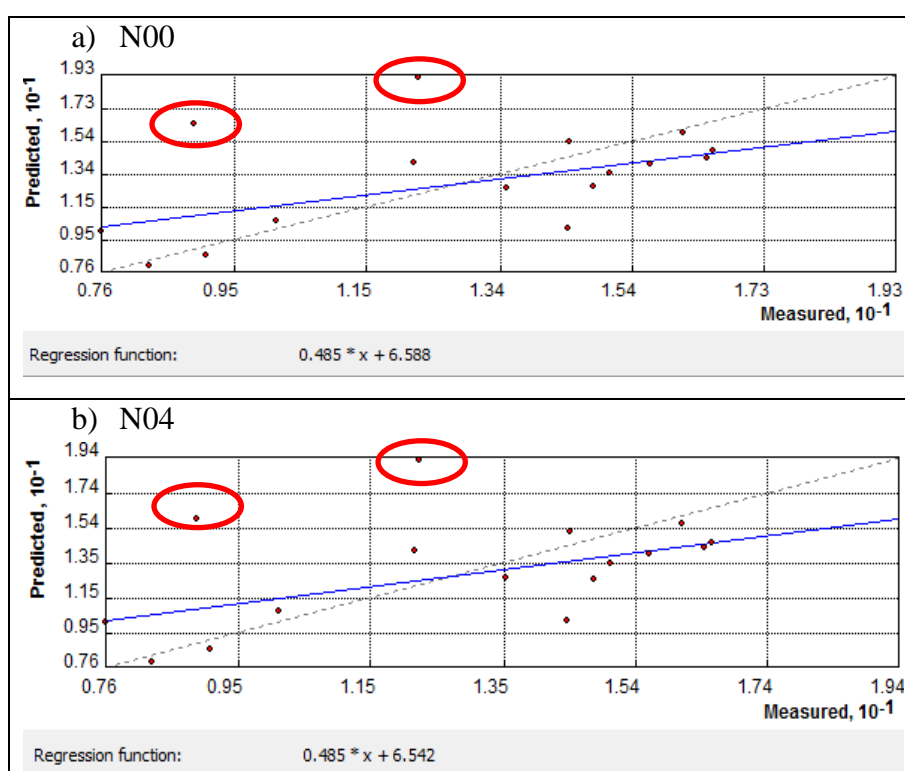


Figure 5.25 Measured versus predicted DO at validation points for N00 and N04

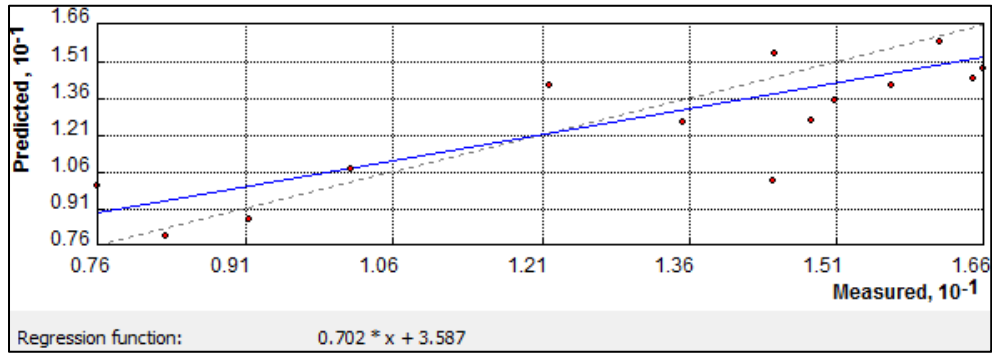


Figure 5.26 Measured versus predicted DO at validation points for N00 after selected sampling locations (red circles in Figure 5.25) were removed from the validation dataset

Summary statistics obtained for validation are given in Table 5.5. As it is mentioned before in Section 4.2.6, ME and MSE values should be close to 0 to indicate accurate predictions. Furthermore, RMSE, RMSSE and ASE values indicate the variability in the predictions with respect to measured values. While RMSE and ASE should be as small as possible, RMSSE should be close to 1 to identify that predictions are close to observations and the prediction standard errors are valid. When Table 5.5 was examined, it was observed that all networks rendered low ME values between 0.2-0.4 except for N40 and N55 with ME values of 0.65 and 1.18, respectively. Figure 5.27 revealed that ASE usually increased as the number of sampling locations decreased, with the exception of N40 (25 sampling locations were used in this network). When measured DO values in the removed sampling locations in N35 to generate N40 were examined, it was seen that most of the values were below the average value for the N35 network. While the average concentration in the removed sampling locations was 10.73 mg/L, the average concentration for N35 was 13.40 mg/L. The removal resulted in an increase in the average DO concentration and a decrease in the DO variance for N40 with respect to N35. Therefore, prediction values were close to the mean of N40 and hence to each other. This situation resulted with a 28% decrease in ASE compared to N35. However, the kriging prediction values at validation locations

were close to the mean, the values diverged from the observation values and caused the increase of ME for N40. A sharp increase (76 %) was observed in ME for N40 compared to N35. This is because ME was calculated based on the differences between the observed and predicted values at validation points.

Table 5.5 Validation results for OK of DO

Sampling Network	Number of Sampling Locations	ME	RMSE	ASE	MSE	RMSSE
N00	65	0.40	3.09	2.94	0.12	1.03
N04	61	0.36	3.06	2.92	0.10	1.02
N10	55	0.36	3.00	2.75	0.091	1.04
N15	50	0.35	3.20	2.72	0.087	1.12
N20	45	0.35	3.13	2.65	0.10	1.14
N25	40	0.25	2.99	2.92	0.06	0.99
N30	35	0.41	2.92	2.99	0.118	0.95
N35	30	0.37	2.96	3.43	0.097	0.85
N40	25	0.65	2.93	2.47	0.24	1.27
N45	20	0.34	3.11	3.31	0.043	0.94
N50	15	0.32	3.43	3.90	0.044	0.87
N55	10	1.18	4.38	4.09	0.21	1.04

As given in Figure 5.28, the lowest RMSE value was 2.92 corresponding to N30 network. When the number of sampling locations was decreased, RMSE of

estimates were increased. If ASE value is greater than RMSE, the variability of the prediction is being overestimated. However, RMSSE values greater than 1 translate to an underestimation in the variability of the predictions. According to validation results given in Table 5.5, while the variability of the predictions is underestimated for N00, N04, N10, N15, N20, N25, N40 and N55 based on the comparisons of ASE and RMSE values, for others (N30, N35, N45 and N50) it is overestimated. However, RMSSE values for all networks, except for N40 with a RMSSE value of 1.27, were close to 1 which showed the success of reduction scenarios in capturing of the variability of the predictions (Figure 5.29).

According to Table 5.5, MSE was close to 0 for some networks such as N45 and N50 compared to N00. This showed that the use of high number of sampling locations in predictions may mess up MSE values. It was obvious from Table 5.5 that different networks were seen as better when different error metrics were taken into account. So, the use of only error metrics in the selection of the suitable networks may not be sufficient.

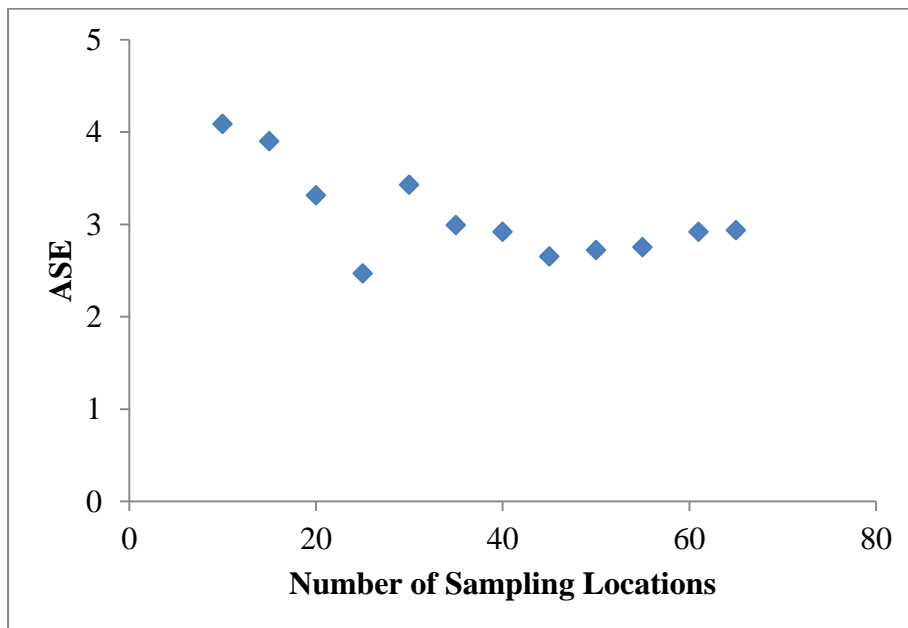


Figure 5.27 ASE values for OK of DO for different sampling point numbers

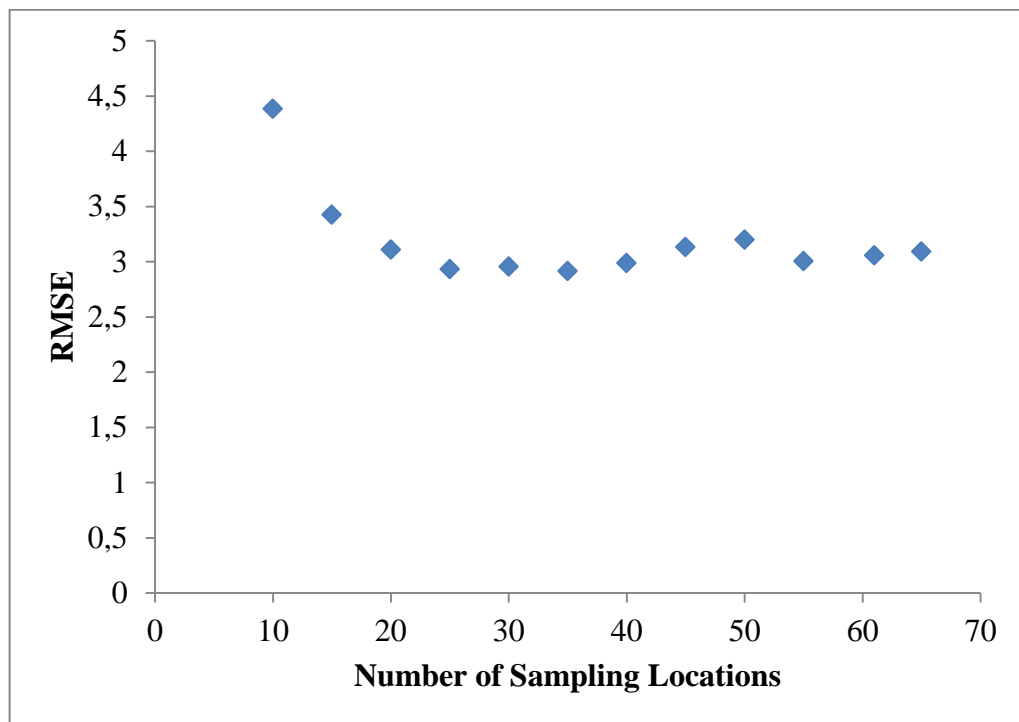


Figure 5.28 RMSE values for OK of DO for different sampling point numbers

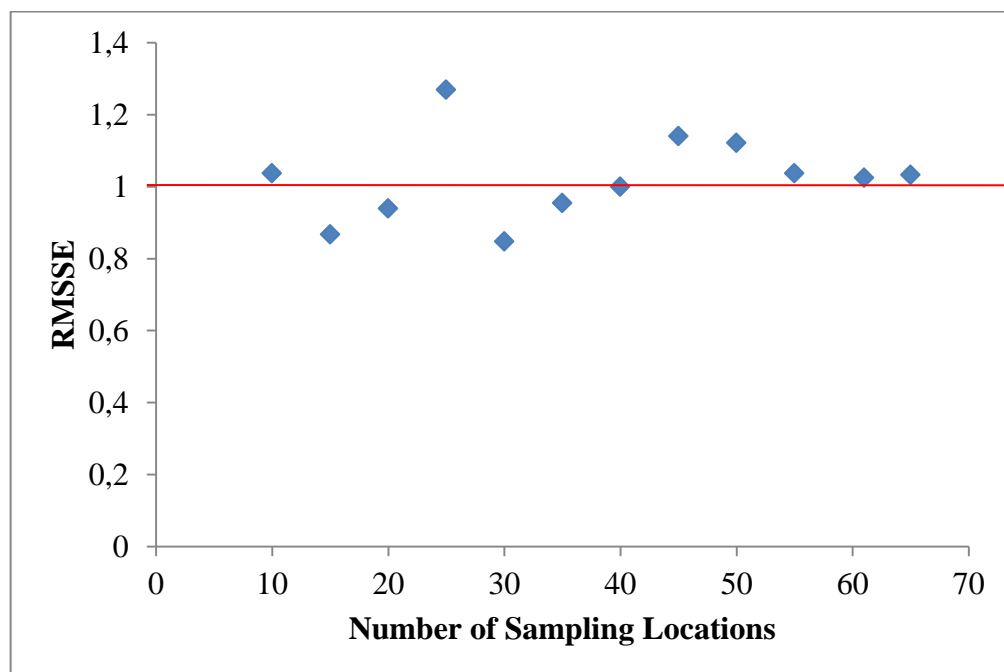


Figure 5.29 RMSSE values for OK of DO for different sampling point numbers

The error metrics, variogram model, kriging map and SEM revealed that 10 water quality sampling locations were not enough to correctly reflect the spatial correlation structure of DO over the surface of PDR. Furthermore, N30 (35 sampling locations) was the best network among other networks with the representative sampling locations for DO, based on error metrics, SEM, kriging map and mainly variogram model. In this network, the spatial correlation structure of DO variogram was significantly conserved with respect to N00. In addition, similar hotspots were observed in N30 and N00. Lowest RMSE value was obtained for this network. On the average, two sampling points were selected per correlation length of N30 (1491 m) based on average pair distance (703 m).

After the selection of representative water quality sampling locations for surface DO in PDR, it was checked whether similar results could be obtained for another water quality parameter for the selected monitoring network. For this purpose salinity and SC were selected as candidates. The relationship between surface SC and salinity in PDR is depicted in Figure 5.30. High correlation was observed ($R^2 = 0.97$) between SC and salinity. Therefore it was decided to use SC only. The surface SC values can differ greatly from location to location depending on the composition of inflowing tributaries, point source pollution such as domestic and industrial wastewater discharges and non-point source pollution such as agricultural run-off.

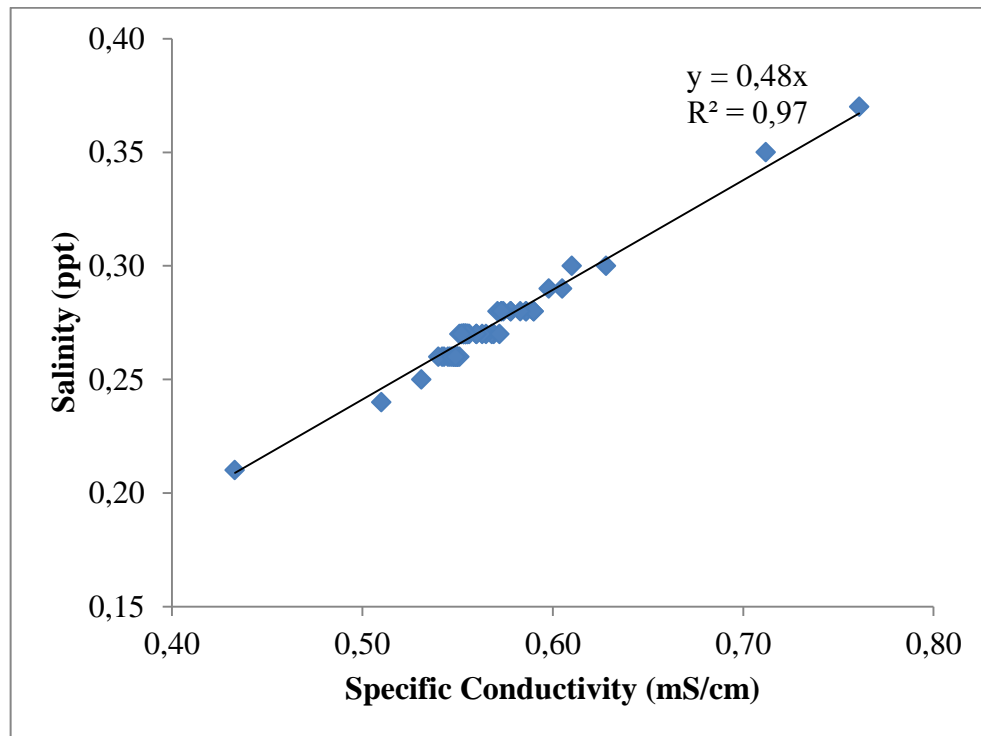


Figure 5.30 Relationship between SC and salinity

As a first step, surface SC distribution was constituted (Figure 5.31) and examined. It was obviously seen that SC at the inlet of the reservoir was noticeably higher than in the rest of PDR. This was mainly due to Porsuk River. SC values in PDR exhibited higher values especially around Softça Village (sampling locations 18 and 20) and in the southeast of PDR where the residual wastewater from agricultural areas are discharged to PDR through the Kargin Regulator intermittently (sampling locations 66, 68 and 70).

Surface SC values in PDR were evaluated according to Appendix 1 of the regulation on the quality of surface water intended for the abstraction of potable water. Even the maximum surface SC value was less than the limit value (1 mS/cm) given in the mentioned regulation. Therefore, PDR was in a better condition in SC compared to DO.

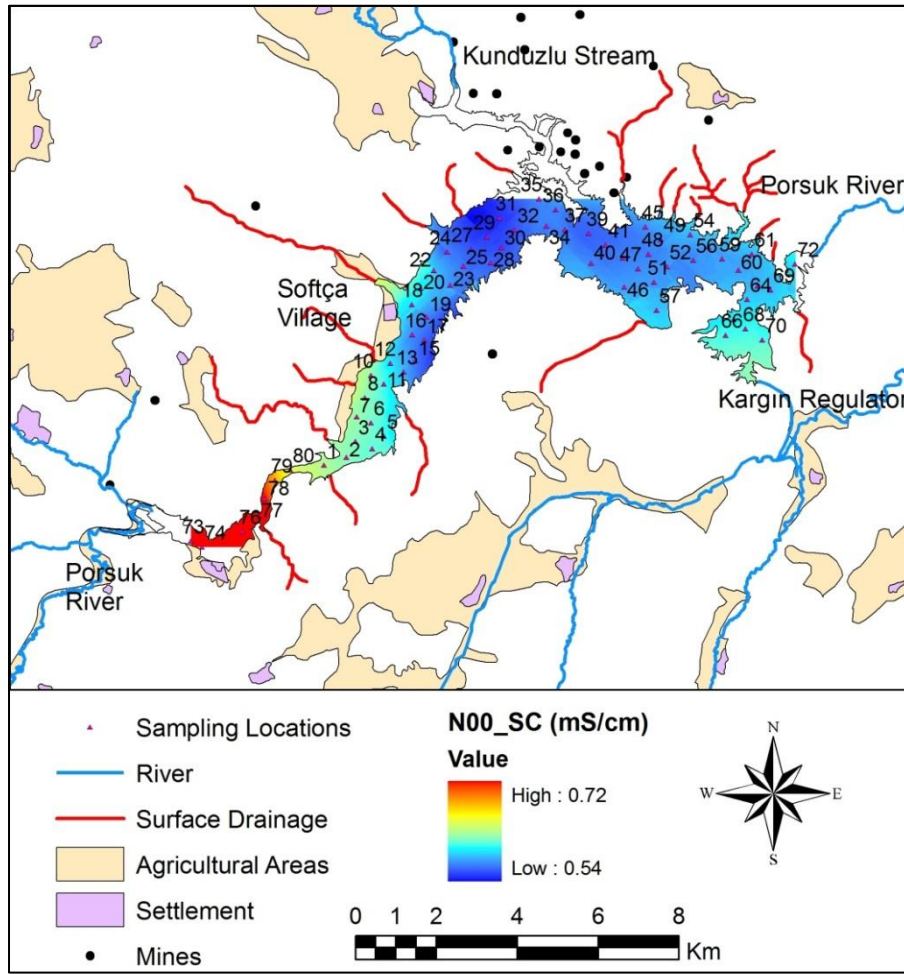


Figure 5.31 OK of SC for N00

The selected network for DO (N30) was also used to constitute the surface SC distribution in PDR (Figure 5.32). In addition to the hotspots observed in Figure 5.31, two additional hotspots with high SC concentrations and one with low SC concentration were detected around sampling locations 28, 46 and 39, correspondingly. The additional hotspot with low SC value around sampling location 39 can be as a result of dilution effect due to inflow of Kunduzlu Stream to PDR. The lower salinity value in Kunduzlu Stream compared to the salinity in PDR was the evidence of lower ionic content with respect to PDR. In Figure 5.33 and Figure 5.34 the highest standard errors (around 0.03 mS/cm) were at the

shoreline of PDR for N00 and N30, respectively. Similar to the results obtained for DO, the lowest standard errors were observed at sampling locations for N00 and N30. Figure 5.35 provides a visual representation of the difference between the standard error maps for N00 and N30. According to Figure 5.35, standard errors decreased noticeably. This may be attributed to the measurement errors in some locations during the field study and improvements in predictions after these locations were removed.

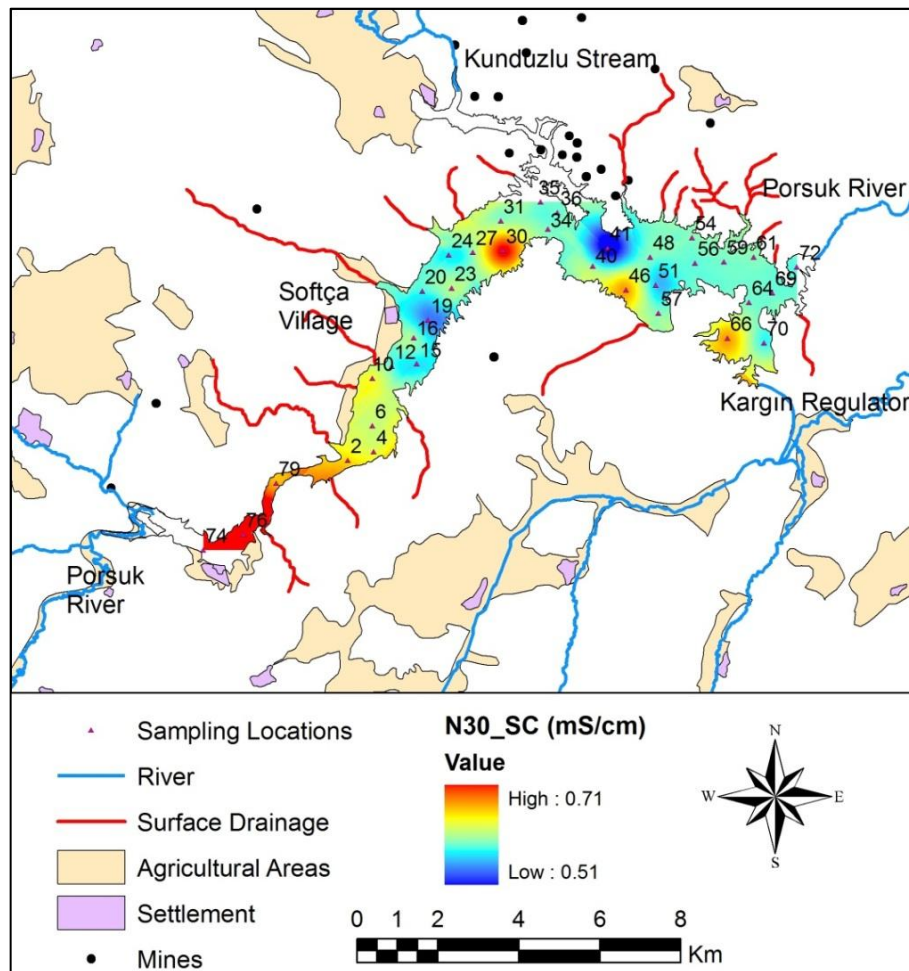


Figure 5.32 OK of SC for N30

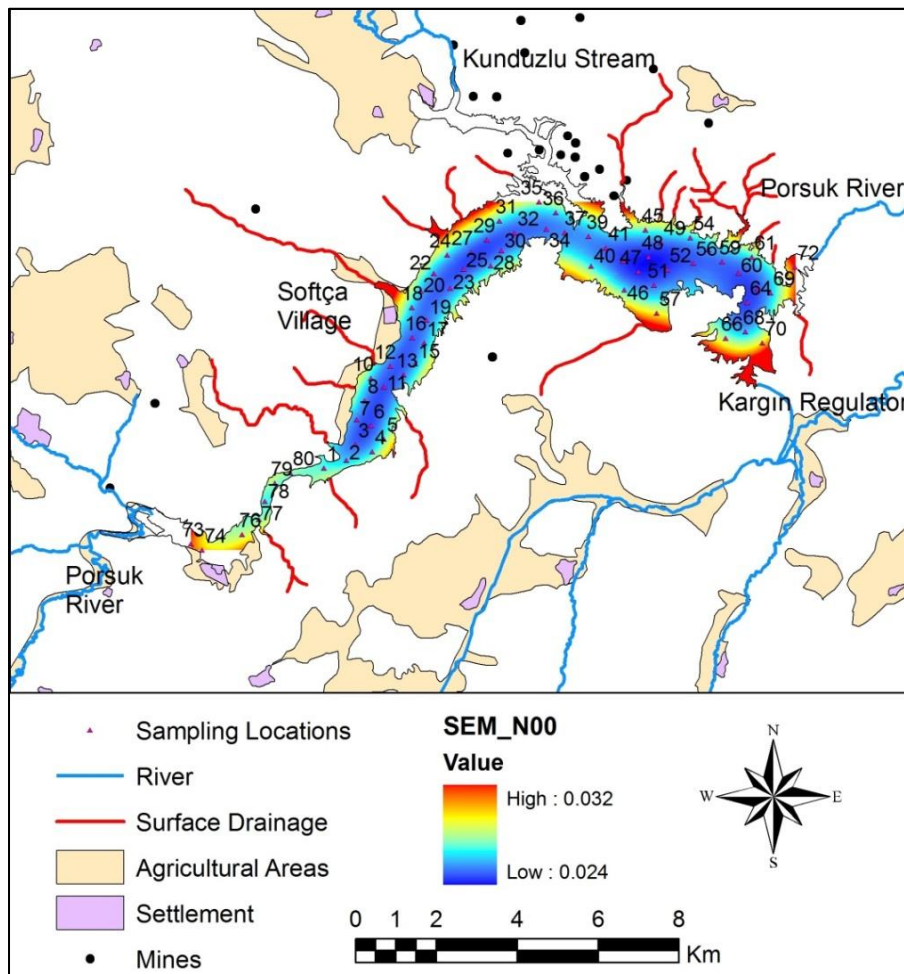


Figure 5.33 SEM of SC for N00

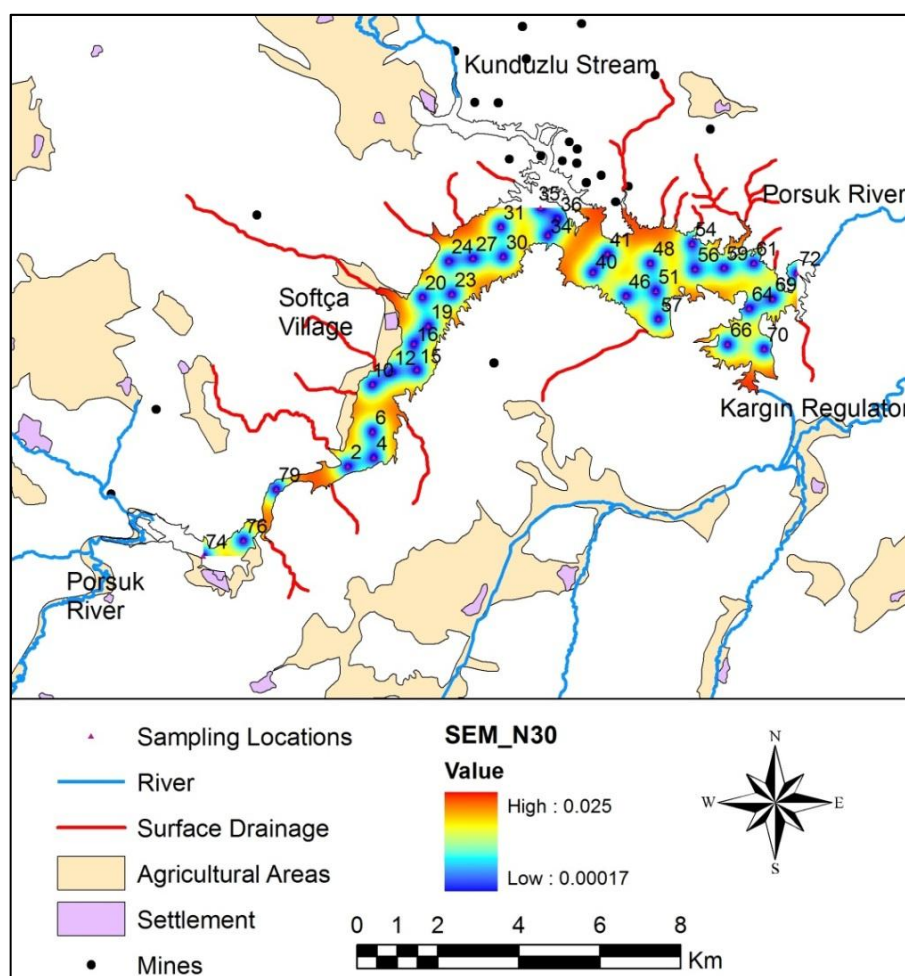


Figure 5.34 SEM of SC for N30

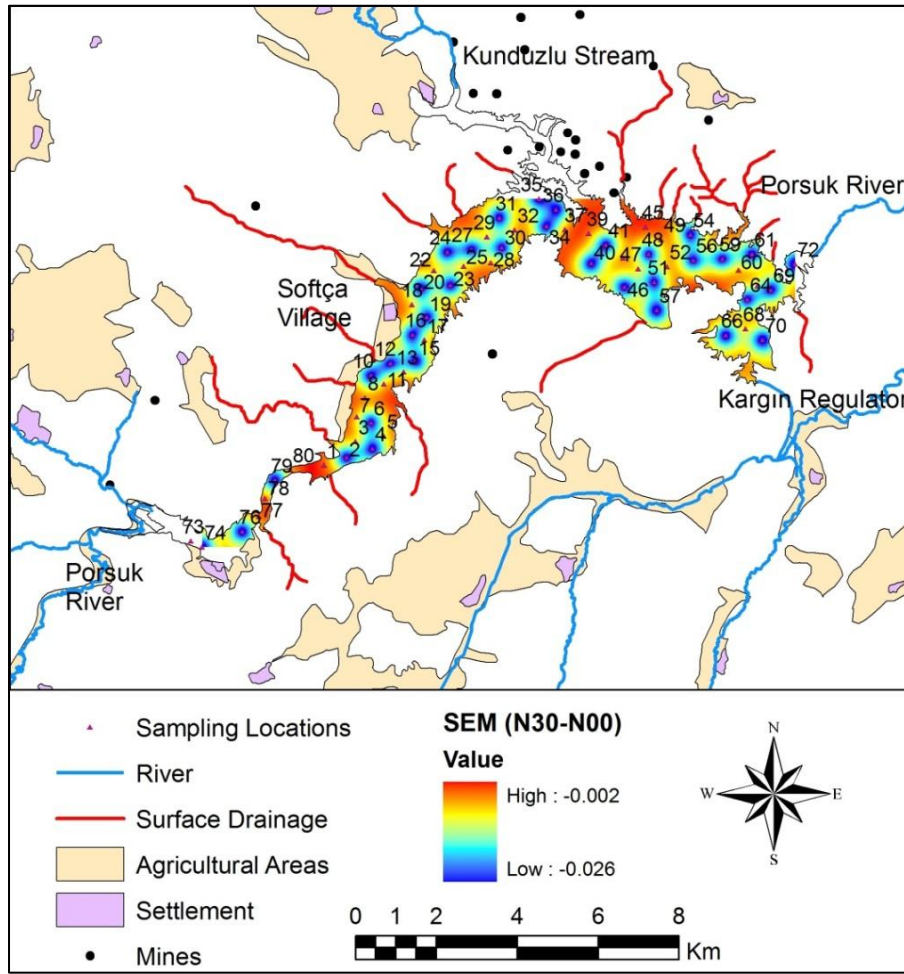
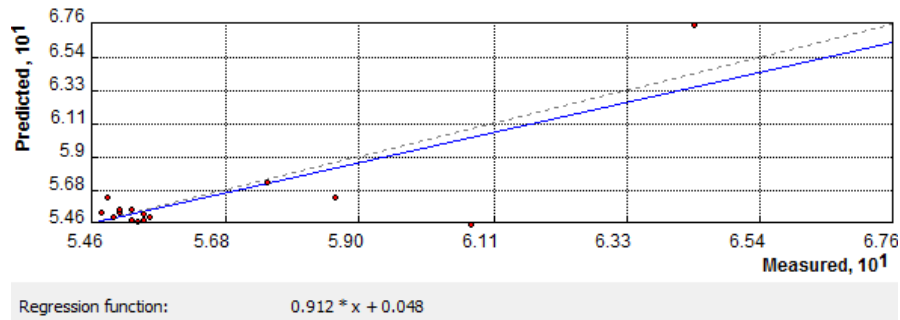


Figure 5.35 Difference of SEMs of SC for N30 and N00

According to Figure 5.36a and Figure 5.36b, which depicted measured versus predicted SC values at validation locations for N00 and N30, respectively, most of the points were close to the dashed gray line (theoretical best). This was a visual evidence of qualified results obtained from OK of SC.

a) N00



b) N30

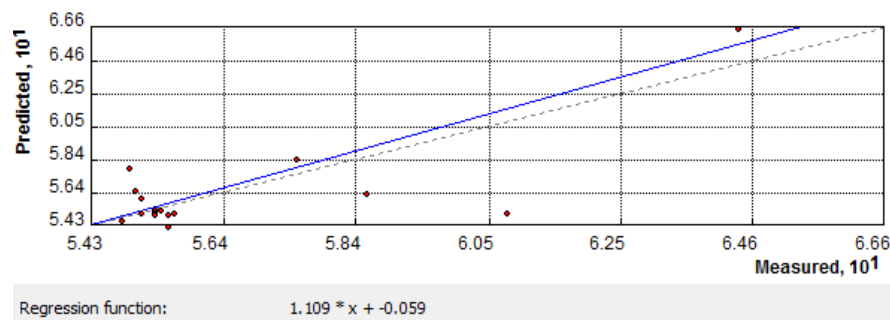


Figure 5.36 Measured versus predicted SC for N00 and N30

Summary statistics for the validation stage are given in Table 5.6. When Table 5.6 was examined, low error values were observed both for N00 and N30 in general. While higher ASE compared to RMSE for N00 translates to an overestimation in the variability of the predictions, lower ASE compared to RMSE for N30 is the indication of underestimation in the variability of the predictions. Although the same RMSE value was exhibited for N30 and N00, ME, ASE and MSE for N30 were lower compared to N00. Although the deviation of RMSSE from 1 for N30 was higher than the one for N00, still RMSSE value for N30 was close to 1 showing the success of the network in capturing the variability of the predictions. In general N30 also generated good results for SC based on error metrics, SEM, variogram model and kriging map. All kriging and standard error (SEM) maps were provided in Appendix G and Appendix H.

Table 5.6 Validation Results for OK of SC

Sampling Network	Number of Sampling Locations	ME	RMSE	ASE	MSE	RMSSE
N00	65	-0.002	0.019	0.025	-0.096	0.750
N30	35	-0.001	0.019	0.017	-0.092	1.377

5.1.5.1 Results for the Networks Selected by Experts

The monitoring network defined by N30 was the best one among other networks with reduced number of sampling points in being representative and preserving the spatial correlation structure in selected water quality parameters through the proposed monitoring network selection approach. In order to test the efficiency of the framework applied in this study, a traditional network design approach was used as well for comparison. Networks were created by three watershed experts. The experts were provided with landuses in the PDR Watershed, inflows and outflows, surface drainage inlets, locations of point and non-point pollution sources, and distributions of initial 65 sampling locations on PDR. Then, the experts were asked to select 10 sampling points which was the lower limit to enable kriging. After the selections of networks by experts, kriging were applied to constitute the surface DO maps. Then results were compared with the one for N55 which was obtained through the approach proposed in this study. First of all, each expert selected different points as the sampling locations among 65 points. Sanders et al. (1983) stated that 100 different designers would develop 100 different sampling network designs for a given water body. According to validation results obtained for these selections (Table 5.7), different networks appeared as better if different error metrics were taken into account. For example, while the network of Expert 3 (NE3) seemed better than the others based on ME and MSE, the selection of Expert 2 (NE2) was better in terms of RMSE and ASE. In overall the selection of Expert 1 (NE1) was poor in accordance to error metric

quantities. N55 and NE3 produced better DO predictions than others. The graphs of measured versus predicted DO values at validation sampling locations for the networks are provided in Figure 5.37. The graphs belonging to N55 and NE3 indicated better results such that predictions were close to the measured ones. The worst case was for NE1.

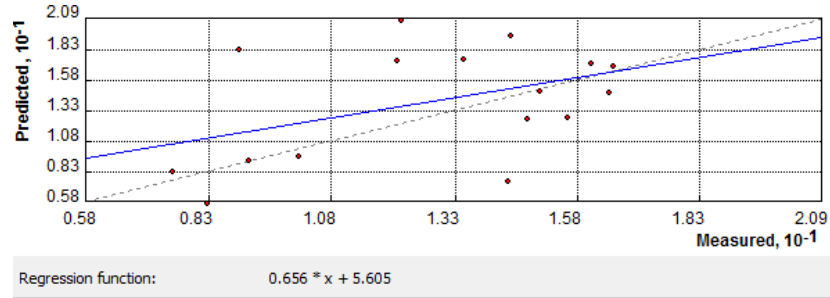
Table 5.7 Validation Results for the selections of experts

Sampling Network	Number of Sampling Locations	ME	RMSE	ASE	MSE	RMSSE
N55	10	1.18	4.38	4.09	0.21	1.04
NE1	10	-4.27	7.32	3.93	-0.99	1.71
NE2	10	1.41	4.05	3.26	0.44	1.25
NE3	10	0.09	4.25	4.54	0.003	0.92

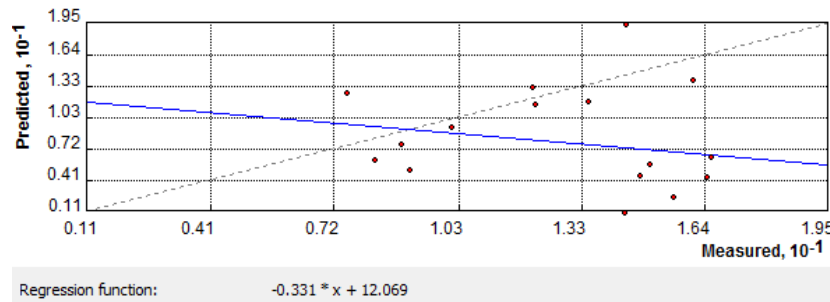
The last evaluation was performed based on variogram parameters. While NE3 was successful based on the evaluations made so far, the spatial correlation structure of the original data set was totally lost in NE3 as can be derived from the variogram parameter values in Table 5.8. The range, nugget, and sill values were unreasonable when the size of the study domain was considered. The range value for NE3 was about 98 km which was more than 6 times longer than the length of the reservoir. This indicated that only error metrics and observed versus predicted maps may not be sufficient in judging the representativeness of the sampling network. It should be noticed that experts considered watershed characteristics (inflows, outflows, pollution sources, landuses, etc.) in the selection of sampling locations to constitute monitoring networks. In the approach, on the other hand, statistically valid sampling locations were determined while still using judgment. When the spatial correlation structure in DO data was considered in addition to

the other evaluation criteria, N55 was the best network. This indicated that proposed strategy for sampling network design was efficient.

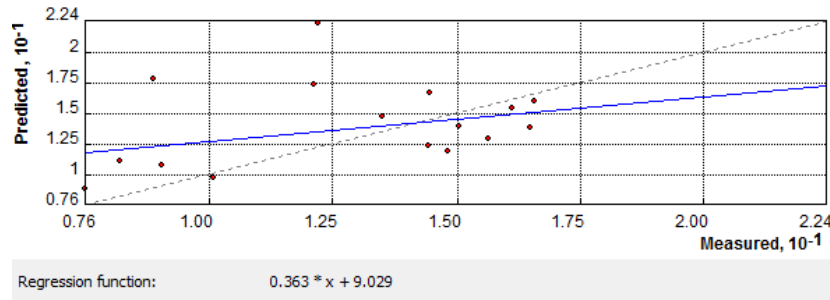
a) N55



b) NE1



c) NE2



d) NE3

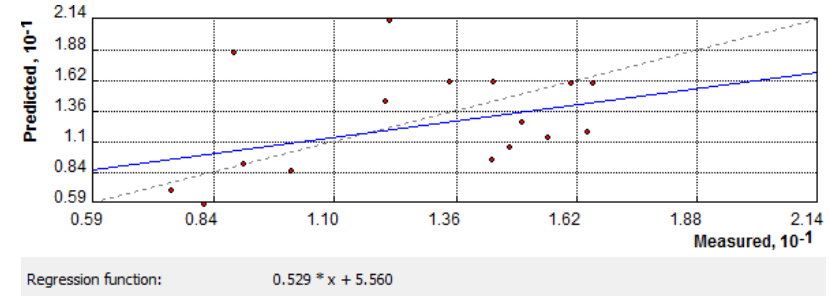


Figure 5.37 Measured versus predicted DO at validation points for the networks of experts and N55

Table 5.8 Variogram model parameters for the networks of experts and N55

Sampling Network	Number of Sampling Locations	Variogram				Range
		Model	Nugget (C_0)	psill (C)	Sill (C_0+C)	
N55	10	Spherical	19	1.00	20	1680
NE1	10	Spherical	17	1.00	18	1710
NE2	10	Exponential	7	2.06	9.06	1674
NE3	10	Exponential	14.89	49.50	64.39	98392

5.2 3D Kriging

In addition to 2D kriging of DO, 3D kriging was also applied to investigate the change of hot spots in deeper layers. Hotspots may change depending on pollution sources (surface drainage and landuse around the PDR) and other factors like groundwater inflow, water abstraction, etc. Therefore, hotspots obtained from 3D kriging could be good indicators of entrance points of pollution sources.

5.2.1 Normality Check

In the first step of 3D kriging application, normality checking was performed. As it is mentioned before a normal distribution is not necessary to obtain prediction maps in OK process. However, Johnston et al. (2001) mentioned that if the data is normally distributed, kriging is the best predictor. Therefore, DO data was statistically analyzed, histogram and Q-Q plot (Figure 5.38) were derived to check normality using ArcGIS software prior to kriging analyses. The DO concentrations measured along the depths (a total of 2090 points) ranged from 0.35 mg/L to 27.33 mg/L with a mean concentration of 9.41 mg/L and a standard deviation of 4.82 mg/L while the surface DO concentrations in PDR measured at 65 water quality sampling stations were between 4.85 mg/L and 22.77 mg/L with a mean concentration of 13.57 mg/L and a standard deviation of 3.76 mg/L. It was

seen that mean DO concentration within PDR decreased when DO concentrations measured along the depths were taken into account.

The shape of the histogram and Q-Q plot of 3D DO data confirmed that DO data exhibited a distribution close to normal distribution. EasyFit Professional software was again used to statistically confirm that data follows the normal distribution. According to Table 5.9, D value was greater than all critical values. So H_0 (the data follow the normal distribution) was rejected. It was seen that the risk to reject the H_0 while it is true was so small based on p-value. This also confirmed that measured DO data in 3D does not follow normal distribution. Transformations explained in methodology section were applied but a better fit to normal distribution for DO was not obtained. Therefore, it was decided to use the original (non-transformed) data in spite of slight deviation from normality based on Q-Q plot.

a) Histogram

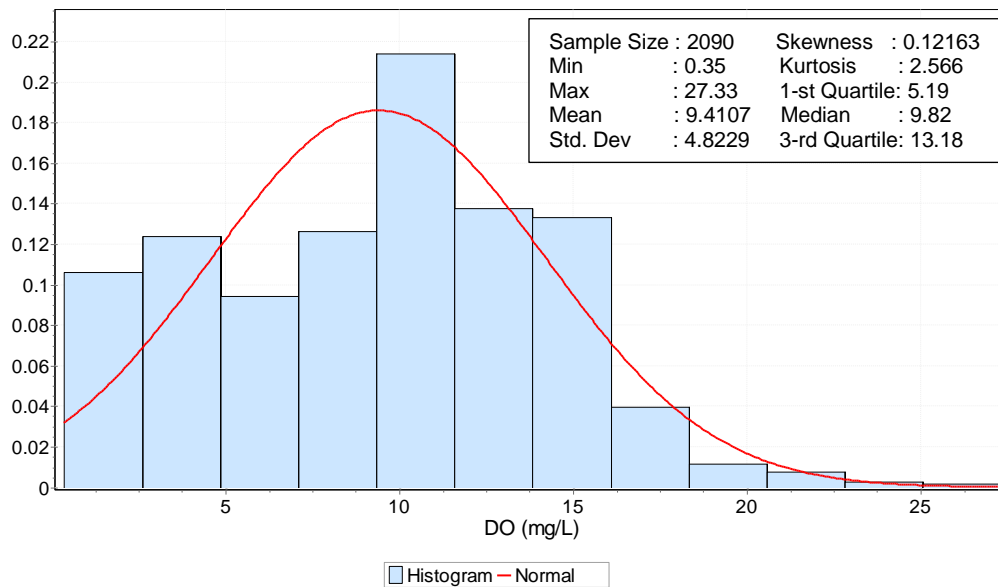


Figure 5.38 Histogram and Q-Q plot for 3D DO data

b) Q-Q Plot

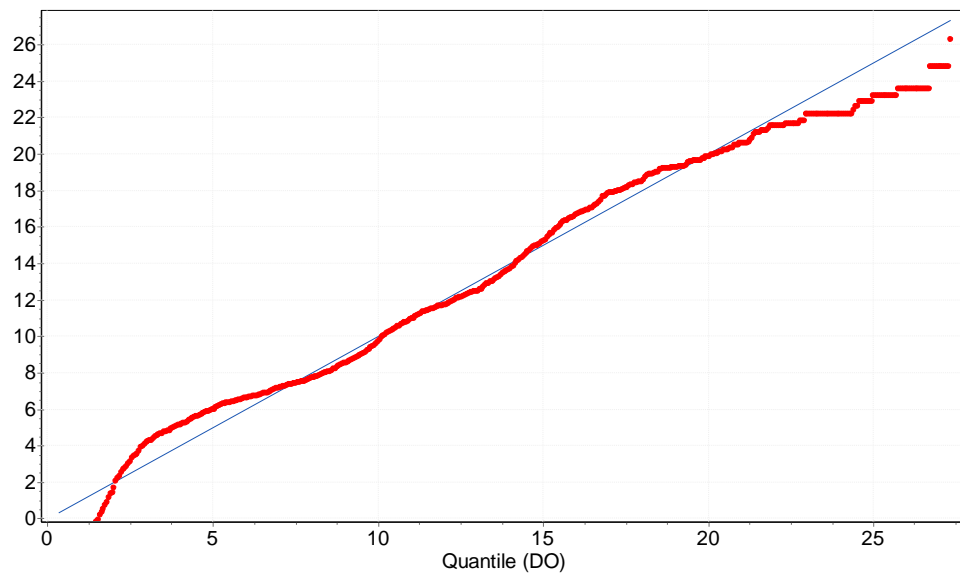


Figure 5.38 Histogram and Q-Q plot for 3D DO data (Continued)

Table 5.9 D and P value for DO data for 3D-kriging

Sample Size		2090				
DO	D-Value	0.06122				
	P-Value	2.9594E-7				
α		0.2	0.1	0.05	0.02	0.01
Critical Value		0.02347	0.02675	0.0297	0.0332	0.03563

5.2.2 Variogram Analysis

Variogram for 3D DO data was constituted using SGeMS. For this purpose, firstly data set was prepared in the appropriate format for SGeMS. SGeMS mainly supports three analytical variogram models and positive linear combination of these models. These are spherical, exponential and gaussian. All these models are permissible in 3D (Remy et al., 2009). As a first step, the experimental semivariogram was constituted to express the spatial correlation of pairs of measured DO concentrations as a function of the distance and the direction of their separation. Omni-directional variogram was calculated since the directional variography is assumed to be isotropic. Afterwards, theoretic model which was suitable for this experimental variogram was determined. The plot and parameters of this variogram model are shown in Table 5.10 and Figure 5.39. The spherical model provided good fit to the experimental variogram of 3D DO data. Nugget effect was not observed in the variogram. A sill of 21.53 and a range of 543.91 were used for DO.

Table 5.10 Variogram Parameters for 3D DO data

Variogram Model	DO
Type	Spherical
Nugget	0
Sill	21.53
Range (m)	543.91

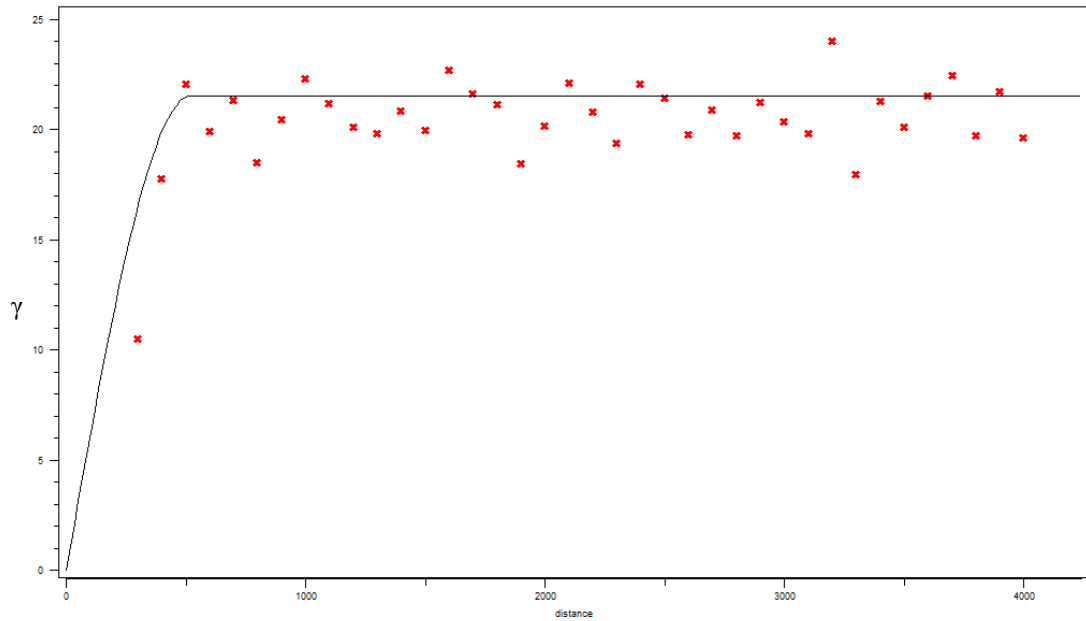


Figure 5.39 The spherical variogram model fitted to the experimental variogram of 3D DO data

5.2.3 Kriging

Firstly distributions of DO, SC and temperature in 3D were constituted using a script written in Matlab. In deep large water bodies, a thermocline layer can be formed due to change of temperature with depth. This layer acts as a physical barrier to exchanges between surface and bottom layers and causes the depletion of DO in deeper parts (Chehata et al., 2007). From Figure 5.40, it was realized that thermocline layer exists approximately between 10 m and 14 m below the surface and exhibits a regular pattern in deep parts. However, DO exhibits a patchy pattern at sampling locations located close to the outlet of PDR (Figure 5.41). A thin thermocline layer is observed in Figure 5.42. In most of the water quality modeling and water quality sampling studies, water samples are collected from half of the total depths at the water quality sampling stations as well as from surface and bottom. For example, the total depth at sampling location 62 is 32 m, one of the water samples would be taken from 16 m below surface. In this

situation, this water sample may not represent the quality of the thermocline layer. This situation may cause incorrect characterization and evaluation. Therefore, temperature, DO and SC profile of the water body of concern should be determined using a sonde that can take measurements along the depth at a sampling point, before water sampling points along the depth are determined.

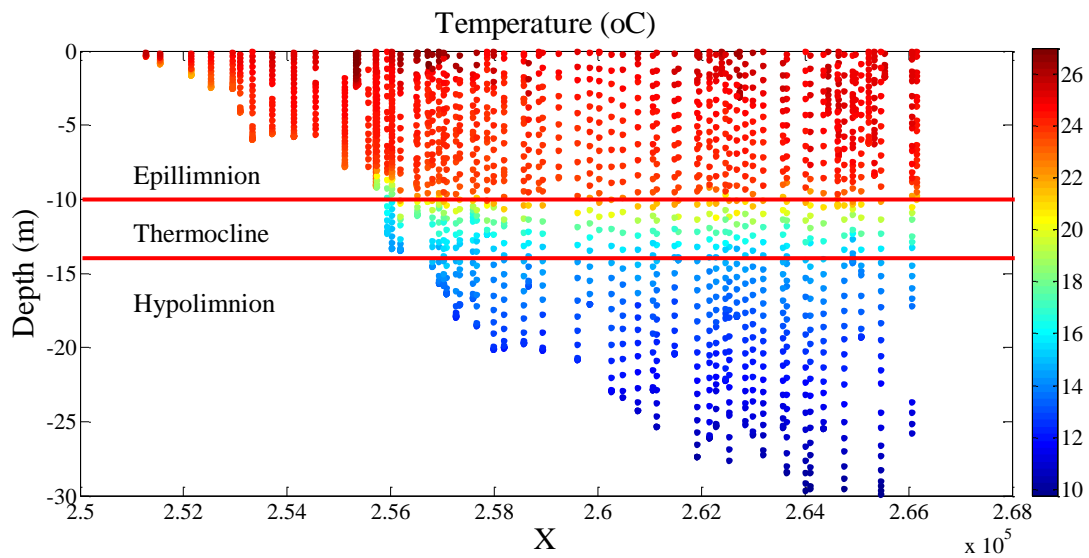


Figure 5.40 X-Z View of 3D temperature data

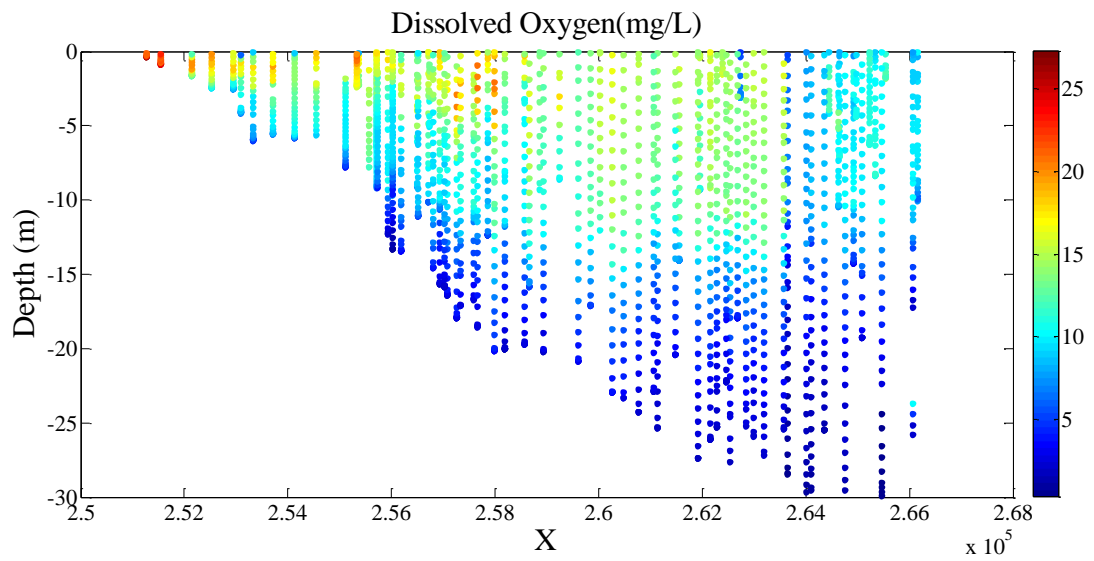


Figure 5.41 X-Z View of 3D DO data

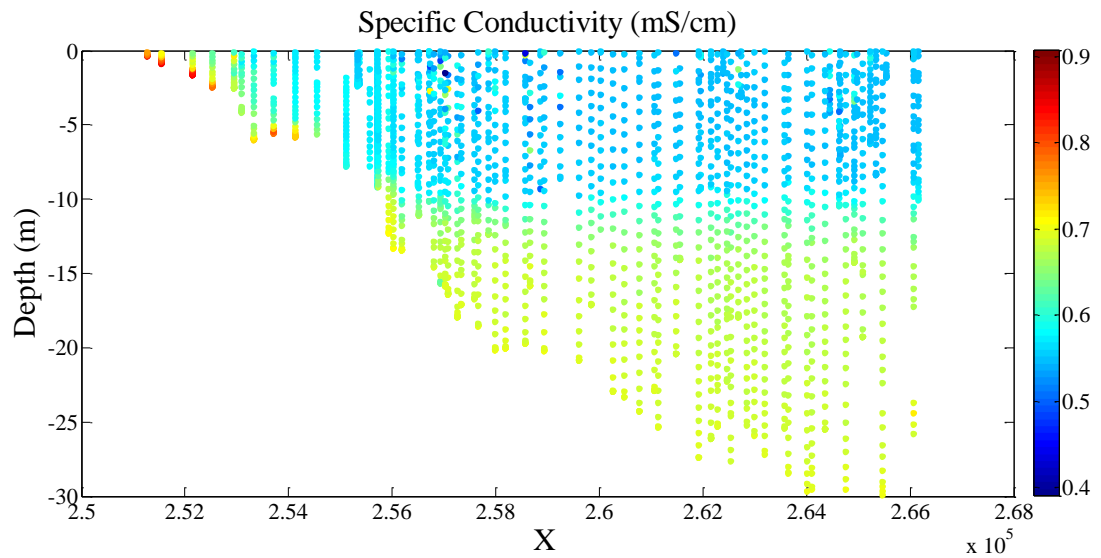


Figure 5.42 X-Z View of 3D SC data

The multiparameter sonde used in data collection is capable of recording measurements down to 30 m below surface level. Following the examination of collected data, it is realized that the effective depth which data could be collected

in the field study conducted in August 2010 was approximately 29 m below surface.

To perform 3D kriging for DO, 3D grid system was constructed in SGeMS (Figure 5.43) and OK was performed. When the distribution of DO for each layer is examined, the lowest DO values (0.35 mg/L) were observed at the bottom (depth of 29 m) (Figure 5.44). This also showed the presence of an anoxic zone at the bottom of the reservoir. There was no significant improvement in DO values from bottom to 15 m below surface. Moreover, estimated values obtained from cross-validation were compared with the measured ones to evaluate measures of accuracy of 3D kriging estimation for DO. The spread of the points was near to the linear regression line in the observed versus predicted graph (Figure 5.45). Correlation coefficient between measured and estimated DO values was 0.94. This is a good correlation and an indication of the accuracy of the predictions. Results obtained from kriging are given with a depth increment of 5 m in Figure 5.44. Figure 5.20 was compared to Figure 5.44 and it was seen that similar hotspots were obtained both for 2D kriging and 3D kriging of DO. Therefore, for the dataset used in this study 2D kriging was sufficient to determine the hotspots for DO. Yet, the results of 3D kriging can be useful to determine the water abstraction points for domestic usage within PDR. According to this, the turquoise zone in the central part of PDR with 8.5 mg/L DO is the most suitable part of PDR for water abstraction. Furthermore, 3D kriging can be valuable for water quality control and modeling studies by producing the data at unsampled locations if desired. Yet, it must be emphasized that the results obtained in this study is based on a field work employed only once in time due to logistic and economic constraints.

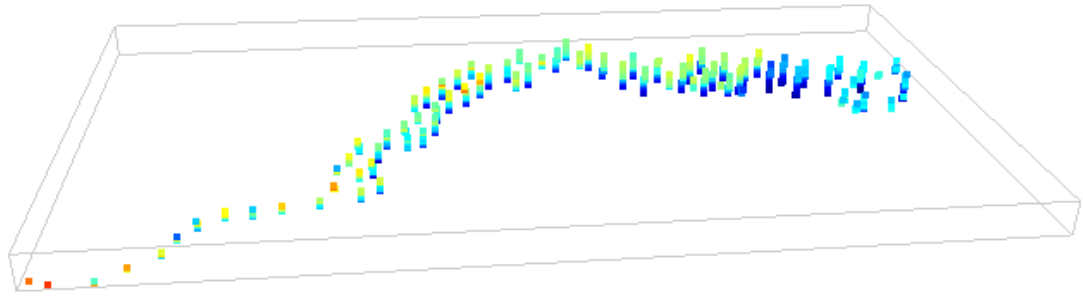


Figure 5.43 Grid for 3D kriging

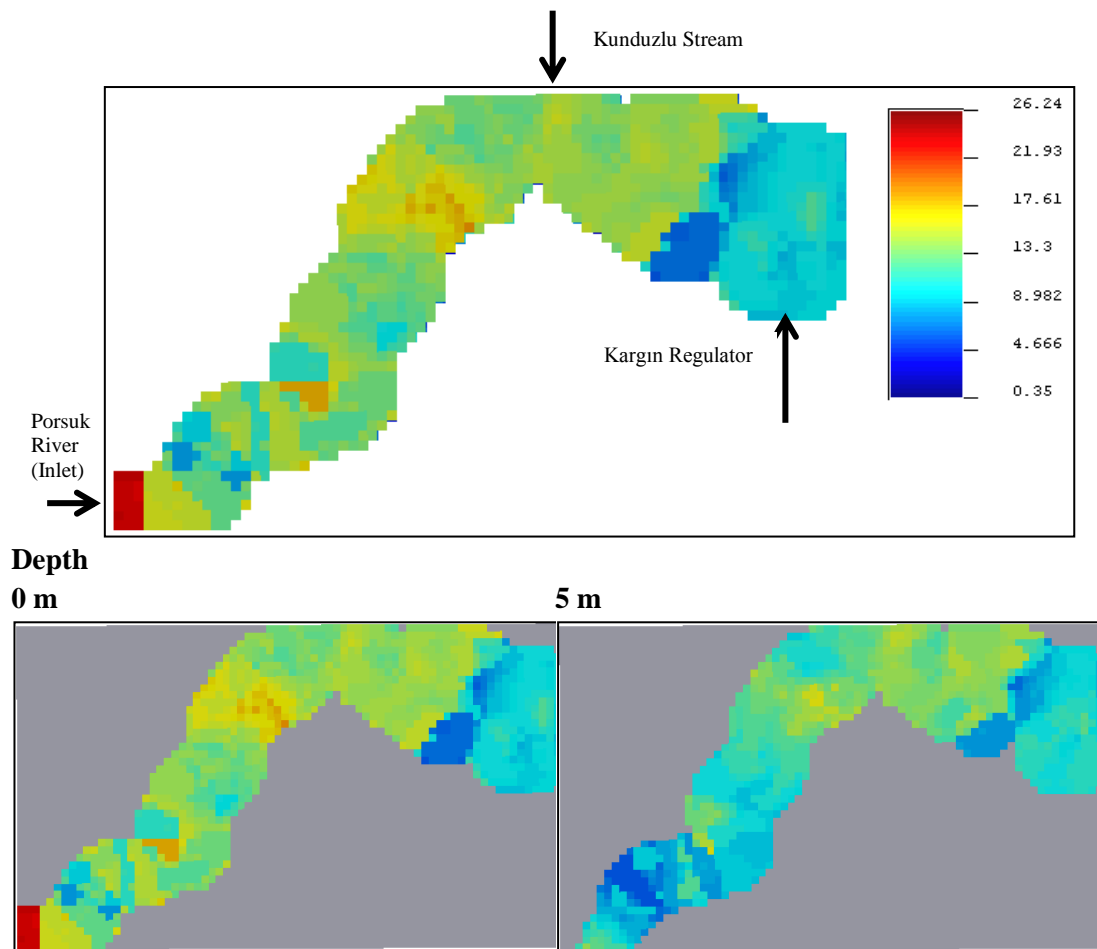


Figure 5.44 3D Kriging Maps of DO at different depths

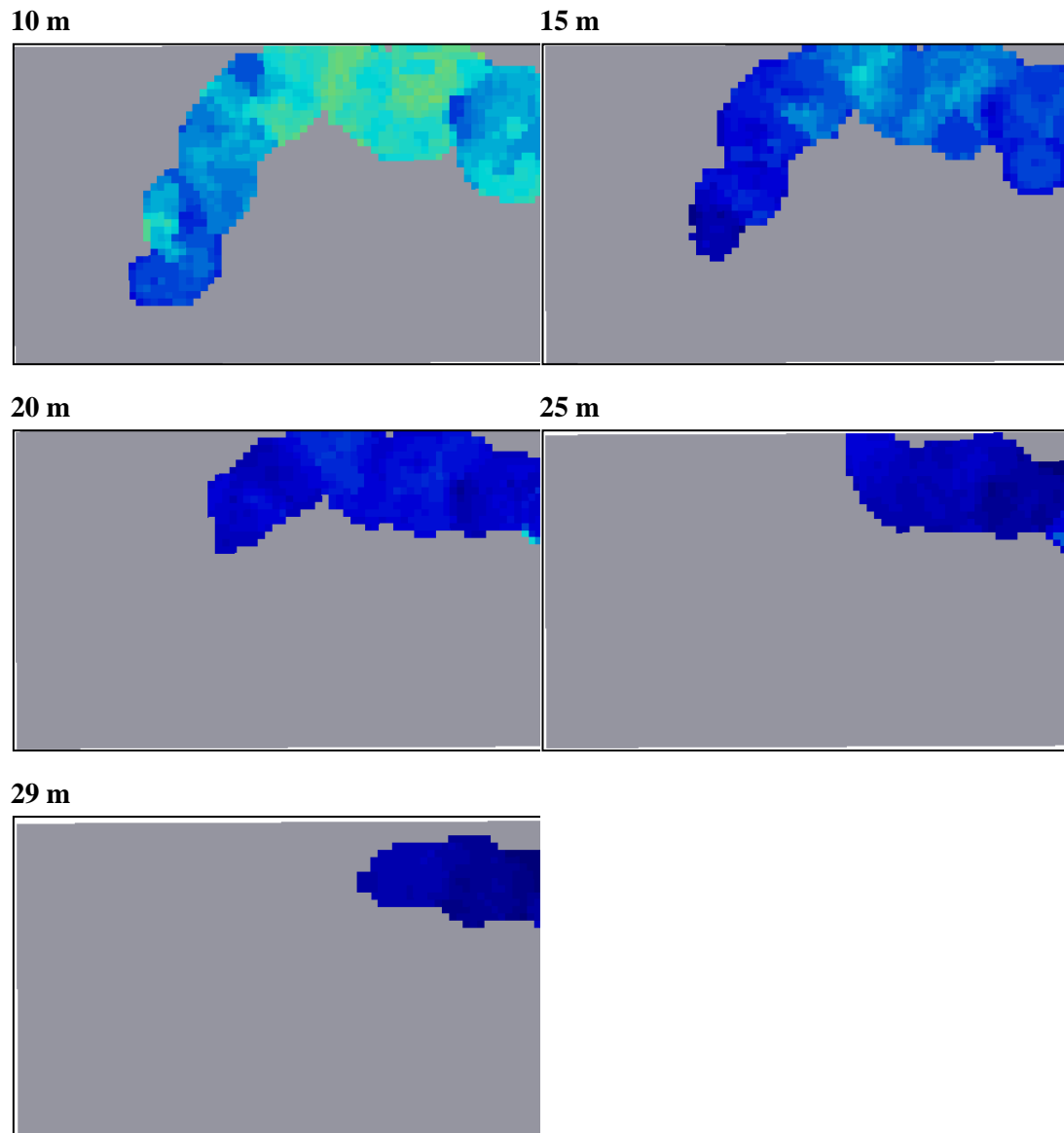


Figure 5.44 3D Kriging Maps of DO at different depths (Continued)

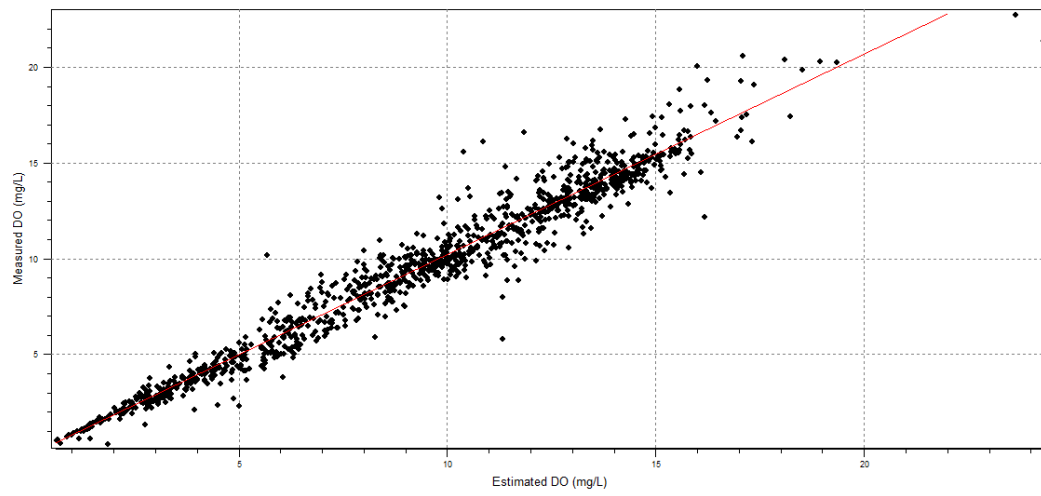


Figure 5.45 Measured versus predicted DO obtained from 3D OK

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

The main objective of this study is to propose an approach to identify the representative water quality sampling locations for the Porsuk Dam Reservoir (PDR) in predicting the spatial distribution of selected water quality parameters with the least number of sampling points possible. Different from the studies in literature, KDE and ordinary 2-dimensional kriging method were used in combination to determine the representative sampling locations.

Constructed kriging maps were evaluated by taking the possible pollutant sources into account. The zones with the lowest and highest concentrations were determined. While high DO clusters were observed at the inlet of PDR upstream and around Softça Village (around sampling location 18), a low DO cluster was observed in the east part of PDR. While the highest standard errors were observed in the shoreline of PDR due to being the furthest locations from sampling points, lowest standard errors were observed at the sampling locations. Furthermore, predicted values at validation locations were compared with the measured ones to evaluate the quality of kriging maps. Generally, observed values were correlated with the measured ones. According to the Methods of Sampling and Analysis guide that supplements the Turkish Water Pollution Control Regulation, a total of five sampling locations are enough to characterize the water quality in reservoirs. However, the variogram model, kriging map, SEM and error metrics revealed that 35 water quality sampling locations were required to correctly reflect the spatial correlation structure of water quality over the surface of PDR for the given case. Conservation of this dependency and structure can be important in better characterization of continuous water quality distribution especially when kriging will be used. Uncertainties in predictions can be decreased as well when the spatial correlation structure in water quality data is conserved. This, on the other

hand, may require high number of samples. It can be argued that high number of sampling locations would be increasing sampling costs. The trade-off between cost and better characterization would always be an issue. However, a representative sampling network design should also sustain spatial dependencies of a given parameter in a water body as well. Geostatistical tools can be helpful for that purpose. Developments in monitoring tools (i.e. ion sensitive probes) and remote sensing techniques may enable measurements at higher number of sampling points with lower costs which would aid in designing representative sampling or monitoring networks.

According to European Union Water Framework Directive (2000/60/EC), good status in all water bodies should be achieved until 2015. If a part of a lake is of a different water quality class compared to the rest of the lake, the lake must be subdivided into different parts because each part will have a different ecological reference situation and a different vulnerability of pressures. This study would be beneficial in the identification of these parts and taking decisions for management.

The efficiency of the monitoring network selection approach was tested using the monitoring stations defined by three watershed management experts. Results showed that rather than subjective sampling location selection, a systematic approach in designing the network would aid in better estimation of the distribution of a given water quality parameter over the domain. It is likely that the procedure suggested in this study could be used for other lakes and reservoirs efficiently. It was also observed that the selected KDE bandwidth and range value for N30 are not exactly the same, but they have the same order of magnitude. Furthermore, 3D distributions of DO, SC and temperature were constituted to determine the location of the thermocline layer. It was seen that the traditional approach of collecting samples at mid-depths at the water quality sampling stations may not represent the thermocline or average concentrations. 3D kriging was also applied to investigate the change of hot spots in deeper layers. In general 3D kriging maps were beneficial to observe the spatial change of water quality

within deeper layers, decide on the water abstraction points for domestic usage and produce data at unsampled locations.

To improve the findings, some recommendations for future studies are as follows;

- application of the suggested approach in this study for other reservoirs and lakes would be beneficial.
- the size of the grids which is used in the determination of initial sampling points should be changed depending on size of the water body.
- this study should be repeated for other seasons as well to see whether the same points are selected as representative sampling locations or not and to determine the seasonal variations in hot spots. Because PDR is a dynamic system in nature, water quality parameters within PDR fluctuate over time and space. So, temporal and spatial variations should be considered as well in the selection of representative water quality sampling locations within dynamic systems like PDR.
- the water quality parameters tested can go beyond DO and SC due to the eutrophication problem in PDR. In this study, only probe measurements were performed due to economic constraints. Inclusion of other water quality parameters such as Chl-a, NO₂, NO₃, PO₄ and other available nutrients may be possible.
- the impact of different variogram models and kriging models on results can be evaluated
- uncertainties caused by data, variogram model, kriging model, trend removal, etc. should be examined.

- remote sensing can be included in the approach for estimating the spatial distribution of selected water quality parameters which may also aid in reducing the number of sampling stations.
- in general, the concentration prediction maps produced by kriging can considerably smoother than those produced by simulated annealing. Therefore, this study can be repeated by simulated annealing to better identify local details.

REFERENCES

- Aggoun, R. C., Tiab, D., & Owayed, J. F. (2006). Characterization of flow units in shaly sand reservoirs—Hassi R'mel Oil Rim, Algeria. *Journal of Petroleum Science and Engineering*, 50 (3-4), 211-226.
- Akdeniz, G. (2004). Kirlenmiş Porsuk sedimentlerinin ilave malzemeler ile iyileştirilmesi [Engineering properties and stabilization of contaminated Porsuk River sediments], MSc. Thesis, Anadolu University.
- AKS Consulting. (2010). İçme ve kullanma suyu kaynağı olarak Porsuk Barajı'nın korunması ve özel hüküm belirlenmesi çalışması mevcut durum raporu. Çevre ve Şehircilik Bakanlığı, Ankara.
- AKS Consulting. (2011). İçme ve kullanma suyu kaynağı olarak Porsuk Barajı'nın korunması ve özel hüküm belirlenmesi çalışması nihai rapor. Çevre ve Şehircilik Bakanlığı, Ankara.
- Amatulli, G., Perez-Cabello, F., & De la Riva, J. (2007). Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling*, 200, 321-333.
- Armitage, P., Berry, G., & Matthews, J. N. S. (2002). Statistical Methods in Medical Research. ISBN: 0-632-05257-0, Blackwell.
- Arslan, O. (2008). Su kalitesi verilerinin CBS ile çok değişkenli istatistik analizi (Porsuk Çayı örneği). *Jeodezi, Jeoinformasyon ve Arazi Yönetimi Dergisi*, 99.

- Artiola, J., Pepper, I. L., & Brusseau, M. L. (2004). Environmental Monitoring and Characterization. ISBN: 0120644770, Elsevier Science and Technology.
- Bakış, R., Koyuncu, H., Özkan, A., Banar, M., Yılmaz, G., & Yörükoğulları, E. (2011). An investigation of the surface and groundwater pollution level in the Porsuk Basin. *Anadolu University Journal of Science and Technology-A*, 12(2), 75-89.
- Barabás, N., Goovaerts, P., & Adriaens, P. (2001). Geostatistical Assessment and Validation of Uncertainty for Three-Dimensional Dioxin Data from Sediments in an Estuarine River. *Environmental science & technology*, 35(16), 3294–3301.
- Barnes, R. (2006). Variogram Tutorial, Golden Software Inc.
- Basarir, H., Kumral, M., Karpuz, C., & Tutluoglu, L. (2010). Geostatistical modeling of spatial variability of SPT data for a borax stockpile site. *Engineering Geology*, 114(3-4), 154–163. doi:10.1016/j.enggeo.2010.04.012
- Bastante, F. G., Ordóñez, C., Taboada, J., & Matías, J. M. (2008). Comparison of indicator kriging, conditional indicator simulation and multiple-point statistics used to model slate deposits. *Engineering Geology*, 98(1-2), 50–59. doi:10.1016/j.enggeo.2008.01.006.
- Bastante, F. G., Taboada, J., Alejano, L. R., & Ordóñez, C. (2005). Evaluation of the resources of a slate deposit using indicator kriging. *Engineering Geology*, 81 (4), 407-418.
- Beğenirbaş, A.S.C. (2002). Porsuk Çayı (Kütahya Bölümü)’ndaki Tatlısu Midyesi (Unio sp.)’nde Bazı Ağır Metallerin Araştırılması. MSc. Thesis, Anadolu University.

- Beveridge, D., St-Hilaire, A., Ouarda, T. B. M. J., Khalil, B., Conly, F. M., Wassenaar, L. I., & Ritson-Bennett, E. (2012). A geostatistical approach to optimize water quality monitoring networks in large lakes: Application to Lake Winnipeg, *Journal of Great Lakes Research*, 38, 174–182. doi:10.1016/j.jglr.2012.01.004.
- Bilge, F. (1997). Porsuk Baraj Gölü su kalitesi izleme çalışmalarında uzaktan algılama ve coğrafi bilgi sistemlerinin kullanılması. MSc. Thesis, Anadolu University.
- Brimicombe, A. (2010). GIS, Environmental Modeling and Engineering. ISBN: 978-1-4398-0870-2, CRC Press, Taylor & Francis Group, USA.
- Burden, F. R., McKelvie, I., Forstner, U., & Guenther, A. (2002). Environmental Monitoring Handbook. ISBN: 978-0071351768, McGraw-Hill.
- Büttner, O., Becker, A., Kellner, S., Kuehn, B., Wendt-Potthoff, K., Zachmann, D. W., & Freise, K. (1998). Geostatistical Analysis of Surface Sediments in an Acidic Mining Lake. *Water, Air, and Soil Pollution*, 108, 297–316.
- Büyükerşen, Y., & Efelerli, S. S. (2008). Porsuk Havzası su yönetimi ve Eskişehir örneği. TMMOB Su Politikaları Kongresi, 451-459.
- Chehata, M., Jasinski, D., Monteith, M. C., & Samuels, W. B. (2007). Mapping Three-Dimensional Water-Quality Data in the Chesapeake Bay Using Geostatistics. *Journal of the American Water Resources Association*, 43(3), 813–828. doi:10.1111/j.1752-1688.2007.00065.x.

- Chilundo, M., Kelderman, P., O' keeffe, J. H. (2008). Design of a water quality monitoring network for the Limpopo River Basin in Mozambique. *Physics and Chemistry of the Earth*, 33 (8-13), 655–665. doi:10.1016/j.pce.2008.06.055.
- Copertino, V. A., Molino, B., & Telesca, V. (1998). Spatial and Temporal Evolution of Water Quality in Reservoirs. *Physics and Chemistry of the Earth*, 23(4), 475–478. doi:10.1016/S0079-1946(98)00057-3.
- Çiçek, A., & Koparal, A. S. (2001). Porsuk Baraj Gölü'nde yaşayan *Cyprinus carpio* ve *Barbus plebejus*'da kurşun, krom ve kadmiyum seviyeleri. *Ekoloji Çevre Dergisi*, 10(39), 3-6.
- De la Riva, J., Perez-Cabello, F., Lana-Renault, N., & Koutsias, N. (2004). Mapping wildfire occurunce at regional scale. *Remote Sensing of Environment*, 92, 363-369.
- Dixon, W., Smyth, G. K., & Chiswell, B. (1999). Optimized Selection of River Sampling Sites. *Water Resources*, 33(4), 971-978.
- Do, H. T., Lo, S.-L., Chiueh, P.-T., & Phan Thi, L. A. (2012). Design of sampling locations for mountainous river monitoring. *Environmental Modelling & Software*, 27-28, 62–70. doi:10.1016/j.envsoft.2011.09.007.
- Erechtchoukova, M. G., Chen, S.Y., & Khaiter, P. A. (2009). Application of Optimization Algorithms for the Improvement of Water Quality Monitoring Systems, Information Technologies in Environmental Engineering, *Environmental Science and Engineering*, 176-188. doi: 10.1007/978-3-540-88351-7_13.

- Ertunc, G., Tercan, A. E., Hindistan, M. A., Ünver, B., Ünal, S., Atalay, F., & Kılhoğlu, S. Y. (2013). Geostatistical estimation of coal quality variables by using covariance matching constrained kriging. *International Journal of Coal Geology*, 112, 14-25.
- Escobedo, F., Varela, S., Zhao, M., Wagner, J. E., & Zipperer, W. (2010). Analyzing the efficacy of subtropical urban forests in offsetting carbon emissions from cities. *Environmental Science & Policy*, 13(5), 362–372. doi:10.1016/j.envsci.2010.03.009.
- Forsythe, K. W., Dennis, M., & Marvin, C. H. (2004). Comparison of Mercury and Lead Sediment Concentrations in Lake Ontario (1968–1998) and Lake Erie (1971–1997/98) using a GIS-Based Kriging Approach. *Water Qual. Res. J. Canada*, 39(3), 190–206.
- Goovaerts, P. (1998). Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of Soils*, 27 (4):315–334.
- Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89 (1-2), 1-45.
- Goovaerts, P. (2010). Geostatistical Software. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis*. Springer Berlin Heidelberg, 125–134. doi:10.1007/978-3-642-03647-7_8.
- Guillemette, N., St-hilaire, A., Ouarda, T. B. M. J., & Bergeron, N. (2011). Statistical Tools for Thermal Regime Characterization at Segment River Scale: Case Study of The Ste-Marguerite River. *River Research Applications*, 27, 1058–1071. doi:10.1002/rra.1411.

- Harmancioglu, N. B., & Alpaslan, N. (1992). Water Quality Monitoring Network Design: A Problem of Multi-Objective Decision Making. *Water Resources Bulletin*, American Water Resources Association, 28 (1),179-192.
- He, Y., Hu, K. L., Chen, D. L., Suter, H. C., Li, Y., Li, B. G., Yuan, X. Y., & Huang, Y. F. (2010a). Three dimensional spatial distribution modeling of soil texture under agricultural systems using a sequence indicator simulation algorithm. *Computers and Electronics in Agriculture*, 715, 524–531. doi:10.1016/j.compag.2009.06.012.
- He, Y., Hu, K. L., Huang, Y. F., Li, B. G., & Chen, D. L. (2010b). Analysis of the anisotropic spatial variability and three-dimensional computer simulation of agricultural soil bulk density in an alluvial plain of north China. *Mathematical and Computer Modelling*, 51(11-12), 1351–1356. doi:10.1016/j.mcm.2009.11.011.
- Hedger, R. D., Atkinson, P. M., & Malthus, T. J. (2001). Optimizing sampling strategies for estimating mean water quality in lakes using geostatistical techniques with remote sensing. *Lakes and Reservoirs: Research and Management*, 6(4), 279–288. doi:10.1046/j.1440-1770.2001.00159.x.
- Houlding, S. W. (1994). 3D Geoscience Modeling, Computer Techniques for Geological Characterization. ISBN:0-387-58015-8, Springer-Verlag Berlin Heidelberg.
- Icaga, Y. (2005). Genetic Algorithm Usage In Water Quality Monitoring Networks Optimization in Gediz (Turkey) River Basin, *Environmental Monitoring and Assessment*, 108 (1-3): 261–277. doi:10.1007/s10661-005-4328-z.

- Jakubek, D. J., & Forsythe, K. W. (2004). A GIS-based Kriging Approach for Assessing Lake Ontario Sediment Contamination. *The Great Lakes Geographer*, 11(1), 1–14.
- Jian, W., & Fanhua, L. (2009). Prediction of oil-bearing single sandbody by 3D geological modeling combined with seismic inversion. *Petroleum Exploration and Development*, 36(5), 623–627.
- Jiménez, N., Toro, F. M., Vélez, J. I., & Aguirre, N. (2005). A methodology for the design of quasi-optimal monitoring networks for lakes and reservoirs. *Journal of Hydroinformatics*, 7(2), 105-116.
- Johnston, K., Ver Hoef, J. M., Krivoruchko, K., Lucas, N. (2001). Using ArcGIS Geostatistical Analyst. ESRI, USA. Retrieved from http://www.ci.uri.edu/projects/geostats/Using_ArcGIS_Geostat_Anal_Tutor.pdf, Last Access: 13-September-2013.
- Kanevski, M. (2008). Advanced Mapping of Environmental Data, Geostatistics, Machine Learning and Bayesian Maximum Entropy. ISBN: 978-1-84821-060-8, ISTE Ltd and John Wiley& Sons Inc.
- Kao, J. J., Li, P. H., & Hu, W. S. (2012). Optimization models for siting water quality monitoring stations in a catchment, *Environ Monitoring Assessment*, 184, 43–52. doi:10.1007/s10661-011-1945-6.
- Karamouz, M., Kerachian, R., Akhbari, M., & Hafez, B. (2009a). Design of River Water Quality Monitoring Networks: A Case Study, *Environmental Modeling & Assessment*, 14(6) , 705-714. doi:10.1007/s10666-008-9172-4.

- Karamouz, M., Nokhandan, A., Kerachian, R., & Maksimovic, C. (2009b). Design of online river water quality monitoring systems using the entropy theory: a case study. *Environmental Monitoring and Assessment*, 155, 63-81. doi: 10.1007/s10661-008-0418-z.
- Kavaf, N., & Nalbantcilar, M. T. (2007). Assessment of contamination characteristics in waters of the Kütahya Plain, Turkey. *Clean*, 35(6), 585–593. doi:10.1002/clen.200700104.
- Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi H. I., Sarfraz, R. A., Baig, A., & Shah, A. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, 72, 301– 309. doi:10.1016/j.ecoenv.2008.02.024.
- Kienel, U., & Kumke, T. (2002). Combining ordination techniques and geostatistics to determine the patterns of diatom distributions at Lake Lama, Central Siberia. *Journal of Paleolimnology*, 28, 181–194.
- Koçal, M. (2006). Investigation of water quality in Porsuk Dam Lake using a mathematical model. MSc. Thesis, Gebze Institute of Technology, Gebze.
- Koutsias, N., Kalabokidis, K. D., & Allgöwer, B. (2004). Fire Occurrence Patterns at Landscape Level: Beyond Positional Accuracy of Ignition Points with Kernel Density Estimation Methods. *Natural Resource Modeling*, 17(4), 359-375.
- Kuter, N., Yenilmez, F., & Kuter, S. (2011). Forest Fire Risk Mapping by Kernel Density Estimation. *Croatian Journal of Forest Engineering*, 32(2), 599–610.

- Kutlu, M., Aydogan, G., Susuz, F., & Özata, A. (2004). The Salmonella mutagenicity of water and sediments from the Porsuk River in Turkey. *Environmental Toxicology and Pharmacology*, 17(2), 111-116.
- Küçük, E. (2013). Investigation of Non-Point Source Pollution Potential in the Watershed of Porsuk Dam Reservoir. MSc. Thesis, Middle East Technical University, Department of Environmental Engineering.
- Külahci, F., & Sen, Z. (2009). Spatio-temporal modeling of ^{210}Pb transportation in lake environments. *Journal of Hazardous Materials*, 165(1-3), 525–532. doi:10.1016/j.jhazmat.2008.10.026.
- Levine, N. (2010). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3), Part III: Spatial Modeling, Chapter 8: Kernel Density Estimation. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.
- Little, L. S., Edwards, D., & Porter, D. E. (1997). Kriging in estuaries: as the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology*, 213, 1–11.
- Lo, S. L., Kuo, J. T., & Wang, S.M. (1996). Water quality monitoring network design of Keelung River, Northern Taiwan, *Water Science and Technology*, 34(12), 49–57. doi:S0273-1223(96)00853-0.
- Mathwave, (2013). Mathwave Data Analysis and Simulation. Retrieved from http://www.mathwave.com/help/easyfit/html/analyses/goodness_of_fit/kolmogorov-smirnov.html, Last Access: 15-August-2013.

- Mazlum, N., Özer, A., & Mazlum, S. (1999). Interpretation of Water Quality Data by Principal Components Analysis. *Turkish Journal of Engineering and Environmental Science*, 23, 19-26.
- McGarigal, K. (2012). The Landscape Ecology Lab at the University of Massachusetts in Amherst. Analysis of Environmental Data Lecture Notes. *Analysis of Environmental Data. Conceptual Foundations: Data Exploration, Screening & Adjustments*. Retrieved from <http://www.umass.edu/landeco/teaching/ecodata/schedule/exploratory.pdf>, Last Access: 20-January-2012.
- Ministry of Environment and Forestry (2006). Environmental Status Report of Eskişehir, 26-27.
- Mohamed, A. M. O., & Antia, H. E. (1998). Geoenvironmental Engineering. ISBN: 0-444-89847-6, Elsevier Science B. V.
- Muhammetoglu, A., Muhammetoglu, H., Oktas, S., Ozgokcen, L., & Soyupak, S. (2005). Impact Assessment of Different Management Scenarios on Water Quality of Porsuk River and Dam System – Turkey. *Water Resources Management*, 19(2), 199–210. doi:10.1007/s11269-005-3473-z.
- Nam, K. (2008). Optimization of Paths and Locations of Water Quality Monitoring Systems in Surface Water Environments, PhD Thesis, School of Civil and Environmental Engineering, Georgia Institute of Technology.
- Ning, S. K., & Chang, N. B. (2004). Optimal expansion of water quality monitoring network by Fuzzy optimization approach. *Environmental Monitoring and Assessment*, 91 (1), 145-170.

- Ocak, A., Çiçek, A., Zeytinoğlu, H., & Mercangöz, A. (2002). Porsuk Çayı suyunun bazı tarım bitkileri üzerindeki ekotoksikolojik etkileri. *Ekoloji Çevre Dergisi*, 11(45), 9-13.
- Olea, R. A. (1991). Geostatistical Glossary and Multilingual Dictionary. Oxford University Press, New York.
- Orak, E. (2006). Water quality modelling with fuzzy logic in Porsuk Creek. MSc. Thesis, Gebze Institute of Technology, Gebze.
- Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, 39, 2621-2635.
- Ouyang, Y., Higman, J., Campbell, D., & Davis, J. (2003). Three-Dimensional Kriging Analysis of Sediment Mercury Distribution: A Case Study, *Journal of the American Water Resources Association*, 39(3), 689–702.
- Öztürk, R. (2007). Porsuk Creek Environmental Problems and Watershed Management Suggestions in Solutions to These Problems. MSc. Thesis, Çukurova University, Department of Landscape Architecture.
- Özyurt, M. S., Dayıoğlu, H., Bingöl, N., & Yamık, A. (2004). Porsuk Baraj Havzası'nın Kütahya kökenli kirlilik problemi. *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 6, 43-52.
- Panda, U. C., Sundaray, S. K., Rath, P., Nayak, B. B., & Bhatta, D. (2006). Application of factor and cluster analysis for characterization of river and estuarine water systems – A case study: Mahanadi River (India). *Journal of Hydrology*, 331(3-4), 434– 445.

- Pardo-Iguzquiza, E., Durán-Valsero, J. J., & Rodríguez-Galiano, V. (2011). Morphometric analysis of three-dimensional networks of karst conduits. *Geomorphology*, 132(1-2), 17–28. doi:10.1016/j.geomorph.2011.04.030.
- Park, S.-Y., Choi, J. H., Wang, S., & Park, S. S. (2006). Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modelling*, 199 (3), 289-297.
- Plug, C., Xia, J. C., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident; analysis and prevention*, 43(6), 1937–1946. doi:10.1016/j.aap.2011.05.007.
- Qu, M., Li, W., & Zhang, C. (2013a). Assessing the spatial uncertainty in soil nitrogen mapping through stochastic simulations with categorical land use information. *Ecological Informatics*, 16, 1–9. doi:10.1016/j.ecoinf.2013.04.001.
- Qu, M., Li, W., & Zhang, C. (2013b). Assessing the risk costs in delineating soil nickel contamination using sequential Gaussian simulation and transfer functions. *Ecological Informatics*, 13, 99–105. doi:10.1016/j.ecoinf.2012.06.005.
- Rathbun, S. L. (1998). Spatial Modelling in Irregularly Shaped Regions: Kriging Estuaries. *Environmetrics*, 9(2), 109–129.
- Remy, N., Boucher, A., & Wu, J. (2009). Applied Geostatistics with SGeMS: A Users' Guide. Cambridge University Press.
- Rhea, L., Person, M., De Marsily, G., Ledoux, E., Galli, A. (1994). Geostatistical Models of Secondary Oil Migration Within Heterogeneous Carrier Beds: A Theoretical Example. *AAPG Bulletin*, 78(11), 1679-1691.

- Sanders, T. G., Ward, R. C., Loftis, J. C., Steele, T. D., Adrian, D. D., & Yevjevich, V. (1983). Design of Networks for Monitoring Water Quality. Water Resources Publications, LLC.
- Schuenemeyer, J. H., & Drew, L. J. (2011). Statistics for Earth and Environmental Scientists. ISBN:978-0-470-58469-9, John Wiley & Sons Inc., Hoboken, New Jersey.
- Semerci, H. (2006). Porsuk Sedimentlerinin geoteknik ve kimyasal özelliklerinin belirlenmesi. MSc. Thesis, Anadolu University, Eskişehir.
- SHW. (1975). Eskişehir and İnönü Plains hydrogeological investigation report. Report, No: 40, Ankara.
- SHW. (1980). Protection of inland water quality-Porsuk River pilot project. TUR//777019 Final Report, State Hydraulic Works, Ankara.
- SHW. (2001). Management plan for Porsuk Watershed. Final Report, State Hydraulic Works, Ankara.
- Silverman, B. W. (1998). Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, USA.
- Simbahan, G. C., Dobermann, A., Goovaerts, P., Ping, J., & Haddix, M. L. (2006). Fine-resolution mapping of soil organic carbon based on multivariate secondary data. *Geoderma*, 132 (3-4), 471-489.

- Soyupak, S., Karaer, F., Gürbüz, H., Kivrak, E., Sentürk, E., & Yazici, A. (2003). A neural network-based approach for calculating dissolved oxygen profiles in reservoirs. *Neural Computing & Applications*, 12(3-4), 166–172. doi:10.1007/s00521-003-0378-8.
- Strobl, R., Robillard, P., Shannon, R., Day, R., & McDonnell, A. (2006a). A water quality monitoring network design methodology for the selection of critical sampling points: Part I. *Environmental Monitoring and Assessment*, 112 (1), 137-158.
- Strobl, R. O., Robillard, P. D., Day, R. L., Shannon, R. D., & McDonnell, A. J. (2006b). A water quality monitoring network design methodology for the selection of critical sampling points: Part II. *Environmental Monitoring and Assessment*, 122, 319–334.
- Strobl, R. O., & Robillard, P. D. (2008). Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management*, 87 (4), 639-648.
- Sun, B., Zhou, S., & Zhao, Q. (2003). Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. *Geoderma*, 115(1-2), 85–99. doi:10.1016/S0016-7061(03)00078-8.
- Şen, Z. (2002). İstatistik Veri İşleme Yöntemleri (Hidroloji ve Meteoroloji). ISBN:975-97034-7-5, Su Vakfı Yayınları.
- Tanik, A., Avilés, A., Clavero, V., & Niell, F. X. (2005). Watershed Modeling. In: Russo, R. C. ed. Modeling nutrient loads and response in river and estuary systems. NATO/CCMS Pilot Study Final Report, Report No: 271, 90-107.

- Telci, I. T., Nam, K., Guan, J., & Aral, M. (2009). Optimal water quality monitoring network design for river systems. *Journal of Environmental Management*, 90, 2987–2998.
- Thornton, K. W., Kennedy, R. H., Magoun, A. D., & Saul, G. E. (1982). Reservoir Water Quality Sampling Design. *Water Resources Bulletin, American Water Resources Association*, 18 (3), 471-480.
- Todd, M. J., Lowrance, R. R., Goovaerts, P., Vellidis, G., & Pringle, C. M. (2010). Geostatistical modeling of the spatial distribution of sediment oxygen demand within a Coastal Plain blackwater watershed. *Geoderma*, 159(1-2), 53–62. doi:10.1016/j.geoderma.2010.06.015.
- UNEP/WHO, 1996. Water Quality Monitoring – A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes, Edited by Jamie Bartram and Richard Balance, ISBN:0 419 22320 7 (Hbk) 0 419 21730 4 (Pbk).
- Varol, M., Gökot, B., Bekleyen, A., & Şen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena*, 92, 11-21. doi:10.1016/j.catena.2011.11.013.
- Villard, R., Maignan, M., Kanevski, M., Rapin, F., & Klein, A. (2010). 3D mapping and simulation of Geneva Lake environmental data. *Geophysical Research Abstracts*.
- Wand, M. P., & Jones, M. C. (1995). Kernel Smoothing. Chapman & Hall, UK.
- Wang, J., & Wu, J. (2009). Occurrence and potential risks of harmful algal blooms in the East China Sea. *The Science of the Total Environment*, 407(13), 4012–4021. doi:10.1016/j.scitotenv.2009.02.040.

- Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley&Sons, Ltd, Second Edition.
- Wetzel, R. G. (2001). *Limnology, Lake and River Ecosystems*, Third Edition, Elsevier Science.
- Wibrin, M. A., Bogaert, P., & Fasbender, D. (2006). Combining categorical and continuous spatial information within the Bayesian maximum entropy paradigm. *Stochastic Environmental Research and Risk Assessment*, 20(6), 423-433.
- Wilson, R. (2012). Using Dual Kernel Density Estimation To Examine Changes in Voucher Density Over Time. *Cityscape: A Journal of Policy Development and Research*, 14(3), 225–234.
- Yerel, S. (2010). Water quality assessment of Porsuk River, Turkey. *E-Journal of Chemistry*, 7(2), 593-599.
- Yuce, G. (2006). The vulnerability of groundwater dependent ecosystems: a study on the Porsuk River Basin (Turkey) as a typical example. *Groundwater and Ecosystems*, 295-310.
- Yuce, G., Pinarbasi, A., Ozelik, S., & Ugurluoglu, D. (2006). Soil and water pollution derived from anthropogenic activities in the Porsuk River Basin, Turkey. *Environmental Geology*, 49(3), 359-375. doi: 10.1007/s00254-005-0072-5.
- Yücel, E., Doğan, F., & Öztürk, M. (1995). Porsuk Çayında ağır metal kirlilik düzeyleri ve halk sağlığı ilişkisi. *Ekoloji Çevre Dergisi*, 17, 29-32.

Zhang, C., Jia, K., & Tong, W. (2011). The application of ordinary kriging in the study of water eutrophication of Wuliangsuhai Lake. *Inner Mongolian Environmental Sciences*, 7. Retrieved from <http://www.bfhjzz.com/gongzuodongtai/398/>, Last Access: 16-May-2012.

Zhou, X., & Xia, B. (2010). Defining and modeling the soil geochemical background of heavy metals from the Hengshi River watershed (southern China): integrating EDA, stochastic simulation and magnetic parameters. *Journal of hazardous materials*, 180(1-3), 542–551. doi:10.1016/j.jhazmat.2010.04.068.

APPENDIX A

REGULATION ON THE QUALITY OF SURFACE WATER INTENDED FOR THE ABSTRACTION OF POTABLE WATER

(İçmesuyu Elde Edilen veya Elde Edilmesi Planlanan Yüzeysel Suların Kalitesine Dair Yönetmelik (79/869/AB ile değişik 75/440/AB) (Ministry of Environment and Forestry, 2010))

Table A. 1 Quality Standards By Category

	Parameters		A1 K	A1 Z	A2 K	A2 Z	A3 K	A3 Z
1	pH		6,5- 8,5		5,5-9		5,5-9	
2	Color (after simple filtration)	mg/l Pt scale	10	20 (I)	50	100 (I)		
3	Total Suspended Solids	mg/l SS	25					
4	Temperature	°C	22	25 (I)	22	25 (I)	22	25 (I)
5	Conductivity	@ 20 °C µs/cm	1000		1000		1000	
6	Odor	(dilution factor @ 25 °C)	3		10		20	
7*	Nitrates	mg/l NO ₃	25	50 (I)		50 (I)		50 (I)
8 ¹	Fluorides	mg/l F	0,7-1	1,5	0,7-1,7		0,7-1,7	
9	Total Decomposable Organic Chlorine	mg/l CI						
10*	Dissolved Ferrous	mg/l Fe	0,1	0,3	1	2	1	
11*	Manganese	mg/l Mn	0,05		0,1		1	
12	Copper	mg/l Cu	0,02	0,05 (I)	0,05		1	
13	Zinc	mg/l Zn	0,5	3	1	5	1	5
14	Boron	mg/l B	1		1		1	
15	Beryllium	mg/l Be						
16	Cobalt	mg/l Co						
17	Nickel	mg/l Ni						
18	Vanadium	mg/l V						
19	Arsenic	mg/l As	0,01	0,05		0,05	0,05	0,1
20	Cadmium	mg/l Cd	0,001	0,005	0,001	0,005	0,001	0,005
21	Total Chrome	mg/l Cr		0,05		0,05		0,05
22	Lead	mg/l Pb		0,05		0,05		0,05
23	Selenium	mg/l Se		0,01		0,01		0,01
24	Mercury	mg/l Hg	0,0005	0,001	0,0005	0,001	0,0005	0,001
25	Barium	mg/l Ba		0,1		1		1
26	Cyanide	mg/l Cn		0,05		0,05		0,05
27	Sulphur	mg/l SO ₄	150	250	150	250 (I)	150	250 (I)

Table A. 1 Quality Standards By Category (Continued)

28	Chlorine	mg/l Cl	200		200		200	
29	Surfactants (react with methylene blue)	mg/l (laurilsulfat)	0,2		0,2		0,5	
30* ¹	Phosphates	mg/l P ₂ O ₅	0,4		0,7		0,7	
31	Phenols (Phenol Index) Para nitroaniline 4 aminoantipyrine	mg/l C ₆ H ₅ OH		0,001	0,001	0,005	0,01	0,1
32	Dissolved or emulsified hydrocarbons (after the separation with petroleum ether)	mg/l		0,05		0,2	0,5	1
33	Polycyclic aromatic hydrocarbons	mg/l		0,0002		0,0002		0,001
34	Total Pesticide (Parathion,BHC,dieldrin)	mg/l		0,001		0,0025		0,005
35*	Chemical Oxygen Demand (COD)	mg/l O ₂					30	
36*	DO Saturation Ratio	% O ₂	>70		>50		>30	
37*	Biochemical Oxygen Demand (BOD ₅) (Without nitrification @ 20 °C)	mg/l O ₂	<3		<5		<7	
38	Kjeldahl Nitrogen (except for NO ₃)	mg/l N	1		2		3	
39	Ammonia Nitrogen (NH ₃ -N)	mg /l N	0,05		1	1,5	2	4(I)
40	Decomposable substances with chloroform	mg/l SEC	0,1		0,2		0,5	
41	Total organic carbon	mg/l C						
42	Remaining organic carbon TOC after flocculation and membrane filtration (5μ)	mg/l C						
43	Total Coliforms @ 37 °C	/100 ml	50		5.000		50.000	
44	Fecal Coliforms	/100 ml	20		2.000		20.000	
45	Fecal streptococci	/100 ml	20		1.000		10.000	
46	Salmonella	5.000 ml	No		No			

A1: Drinkable water following physical treatment and disinfection, A2: Drinkable water following physical and chemical treatment and disinfection, A3: Drinkable water following physical, chemical and advanced treatment and disinfection

¹ The highest average annual temperature of these values for a given limit values. (top and bottom)
Z = compulsory, K = guide, I= exceptional climatic or geographic conditions, * See paragraph (d) of Article 12

APPENDIX B

THE TURKISH WATER POLLUTION CONTROL REGULATION (Ministry of Environment and Forestry, 2010)

Table B. 1 Quality Criteria for Inland Water Resources Classes

WATER QUALITY PARAMETERS	WATER QUALITY CLASSES			
	I	II	III	IV
A) Physical and inorganic chemical parameters				
1) Temperature (°C)	25	25	30	> 30
2) pH	6.5-8.5	6.5-8.5	6.0-9.0	except 6.0-9.0
3) Dissolved oxygen (mg O ₂ /L) ^a	8	6	3	< 3
4) Oxygen saturation (%) ^a	90	70	40	< 40
5) Chloride ion (mg Cl ⁻ /L)	25	200	400 ^b	> 400
6) Sulphate ion (mg SO ₄ ⁼ /L)	200	200	400	> 400
7) Ammonium nitrogen (mg NH ₄ ⁺ -N/L)	0.2 ^c	1 ^c	2 ^c	> 2
8) Nitrite nitrogen (mg NO ₂ ⁻ -N/L)	0.002	0.01	0.05	> 0.05
9) Nitrate nitrogen (mg NO ₃ ⁻ -N/L)	5	10	20	> 20
10) Total phosphorus (mg P/L)	0.02	0.16	0.65	> 0.65
11) Total dissolved solid (mg/L)	500	1500	5000	> 5000
12) Color (Pt-Co unit)	5	50	300	> 300
13) Sodium (mg Na ⁺ /L)	125	125	250	> 250
B) Organic parameters				
1) Chemical oxygen demand (COD) (mg/L)	25	50	70	> 70
2) Biological oxygen demand (BOD) (mg/L)	4	8	20	> 20
3) Total organic carbon (mg/L)	5	8	12	> 12
4) Total kjeldahl nitrogen (mg/L)	0.5	1.5	5	> 5
5) Oil and grease (mg/L)	0.02	0.3	0.5	> 0.5
6) Surfactants (react with methylene blue) (MBAS) (mg/L)	0.05	0.2	1	> 1.5
7) Phenolic compounds (volatile) (mg/L)	0.002	0.01	0.1	> 0.1
8) Mineral oils and derivatives (mg/L)	0.02	0.1	0.5	> 0.5
9) Total pesticides (mg/L)	0.001	0.01	0.1	> 0.1
C) Inorganic pollution parameters ^d				
1) Mercury (µg Hg/L)	0.1	0.5	2	> 2
2) Cadmium (µg Cd/L)	3	5	10	> 10
3) Lead (µg Pb/L)	10	20	50	> 50
4) Arsenic (µg As/L)	20	50	100	> 100
5) Copper (µg Cu/L)	20	50	200	> 200
6) Chrome (total) (µg Cr/L)	20	50	200	> 200
7) Chrome (µg Cr ⁺⁶ /L)	less than detection limit	20	50	> 50
8) Cobalt (µg Co/L)	10	20	200	> 200
9) Nickel (µg Ni/L)	20	50	200	> 200
10) Zinc (µg Zn/L)	200	500	2000	> 2000
11) Cyanide (total) (µg CN/L)	10	50	100	> 100
12) Fluoride (µg F ⁻ /L)	1000	1500	2000	> 2000
13) Free chlorine (µg Cl ₂ /L)	10	10	50	> 50
14) Sulphur (µg S ⁼ /L)	2	2	10	> 10

Table B. 1 Quality Criteria for Inland Water Resources Classes (Continued)

15) Iron ($\mu\text{g Fe/L}$)	300	1000	5000	> 5000
16) Manganese ($\mu\text{g Mn/L}$)	100	500	3000	> 3000
17) Boron ($\mu\text{g B/L}$)	1000 ^e	1000 ^e	1000 ^e	> 1000
18) Selenium ($\mu\text{g Se/L}$)	10	10	20	> 20
19) Barium ($\mu\text{g Ba/L}$)	1000	2000	2000	> 2000
20) Aluminum (mg Al/L)	0.3	0.3	1	> 1
21) Radioactivity (pCi/L)				
alfa-activity	1	10	10	> 10
beta-activity	10	100	100	> 100
D) Bacteriologic parameters				
1) Fecal coliform(MPN/100 mL)	10	200	2000	> 2000
2) Total coliform (MPN/100 mL)	100	20000	100000	> 100000

(a) It is sufficient that only one of the parameters of the saturation concentration or percentage is satisfied.

(b) It might be necessary to reduce the concentration limit for watering plants that are sensitive to chlorine.

(c) Free ammonium nitrogen concentration should not exceed $0.02 \text{ mg NH}_4^+\text{N/L}$ depending on pH value.

(d) The criteria in this group give the total concentration of the chemical types that constitute the parameters.

(e) It might be necessary to reduce the criteria down to $300 \mu\text{g/L}$ for watering plants that are sensitive to boron.

Class I: high quality water, Class II: slightly polluted water, Class III: Polluted water, Class IV: highly polluted water

APPENDIX C

WATER QUALITY MEASUREMENT RESULTS

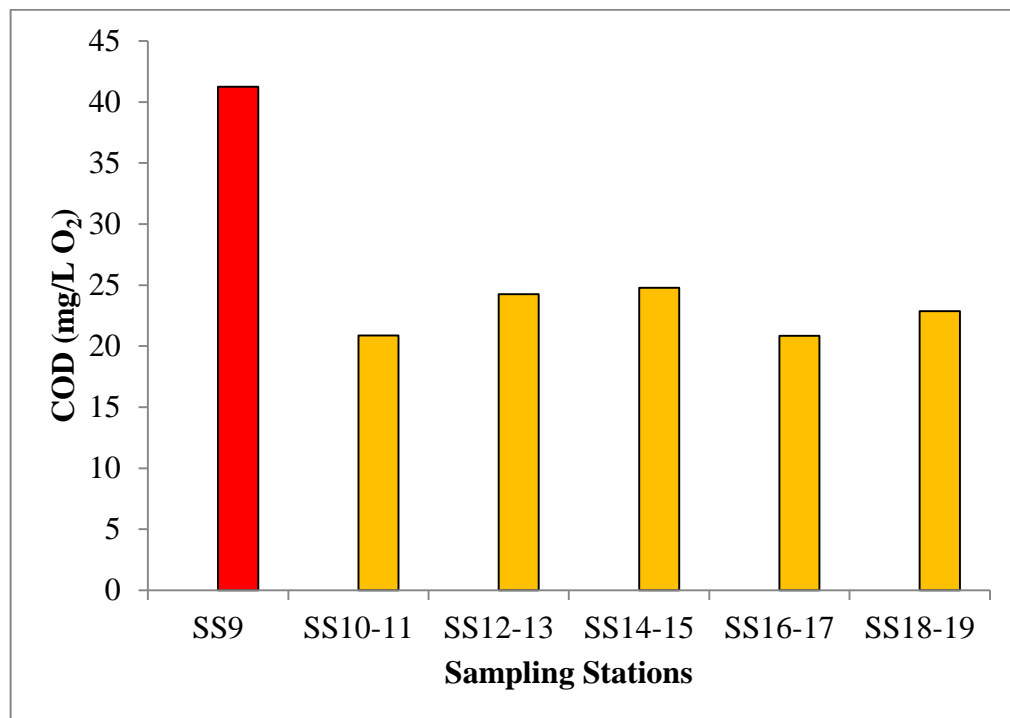


Figure C. 1 Average COD distribution within PDR (AKS, 2010)

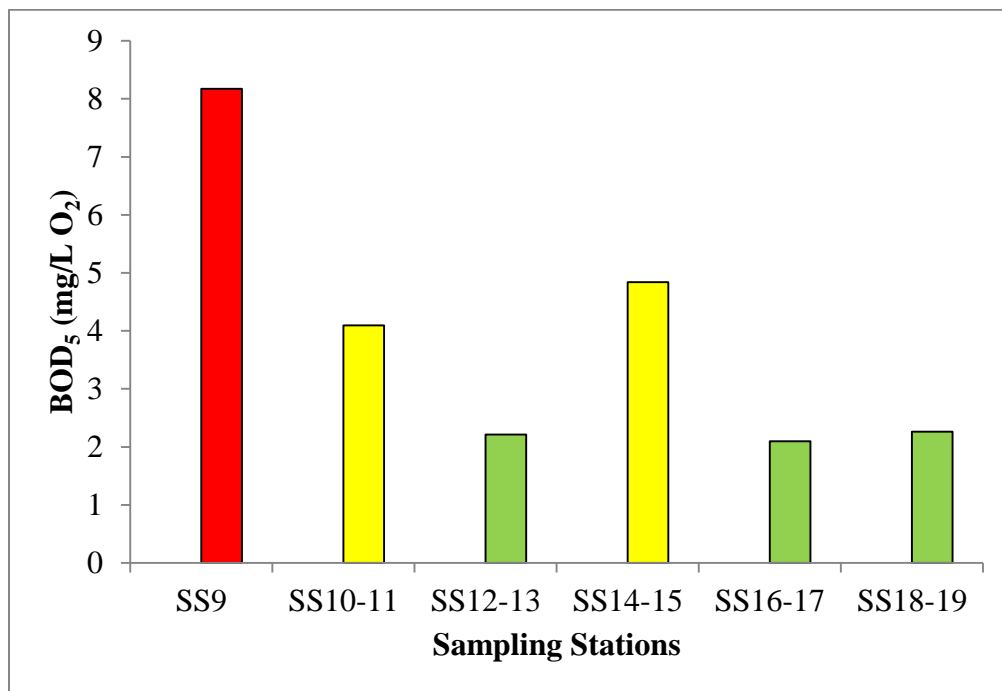


Figure C. 2 Average BOD₅ distribution within PDR (AKS, 2010)

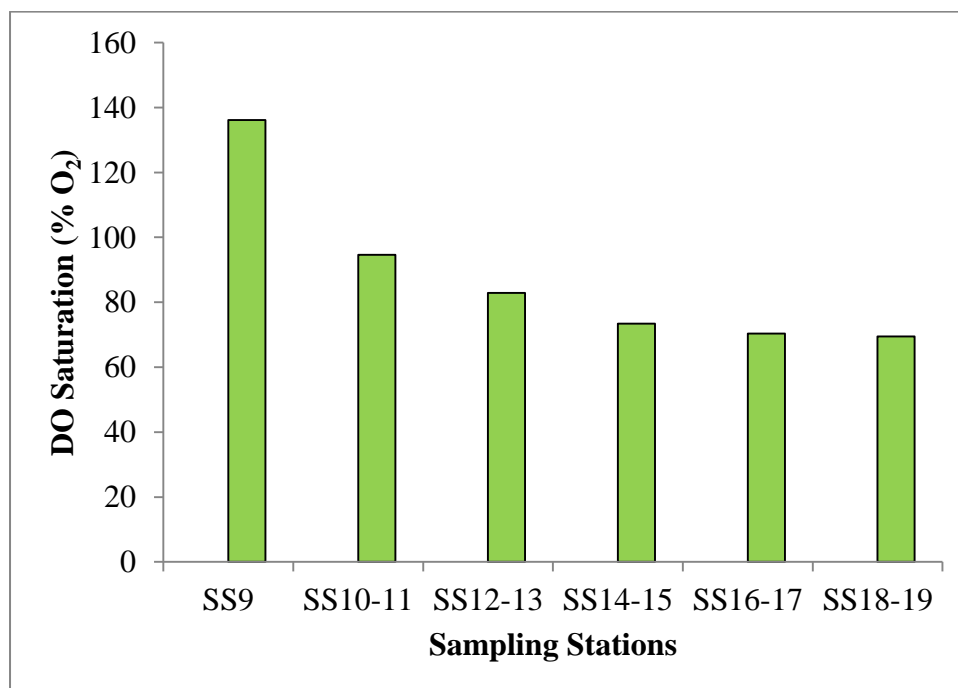


Figure C. 3 Average DO Saturation distribution within PDR (AKS, 2010)

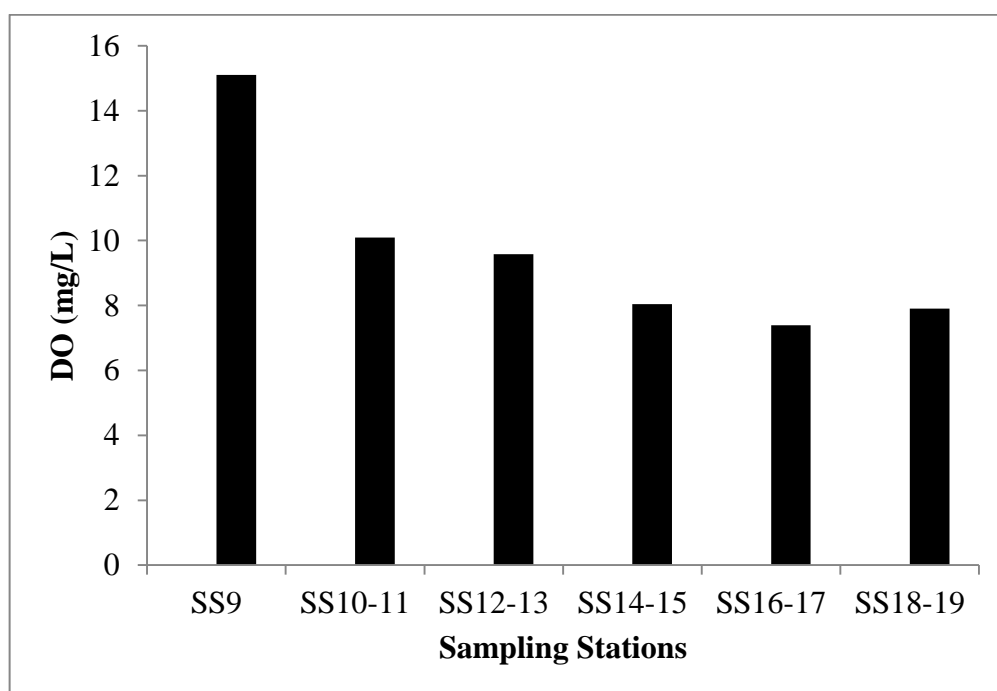


Figure C. 4 Average DO distribution within PDR (AKS, 2010)

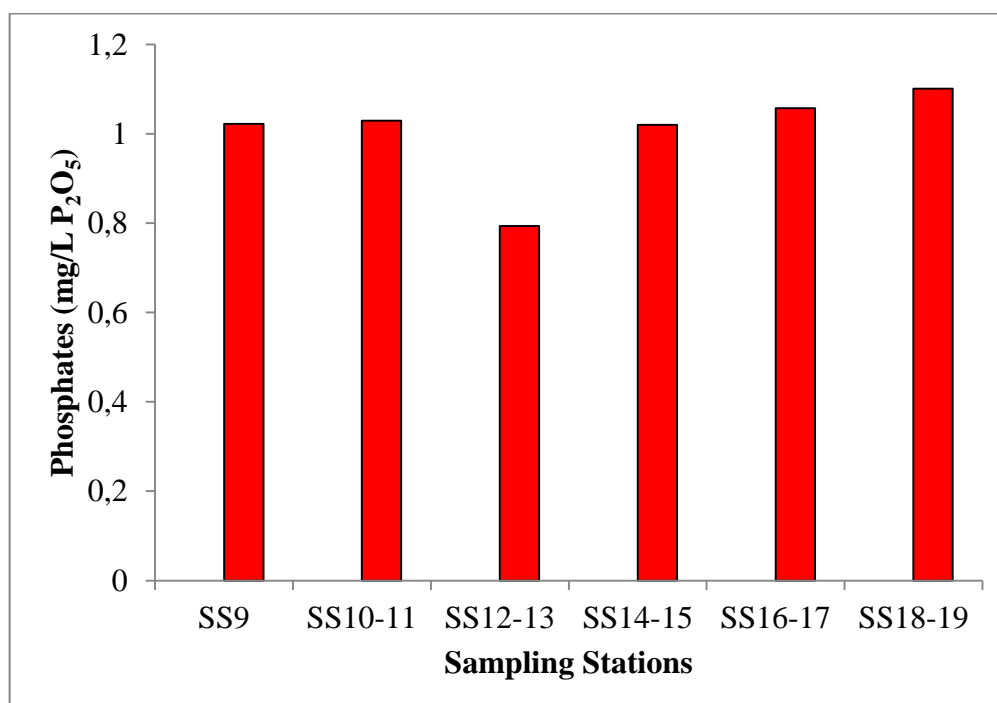


Figure C. 5 Average Phosphates distribution within PDR (AKS, 2010)

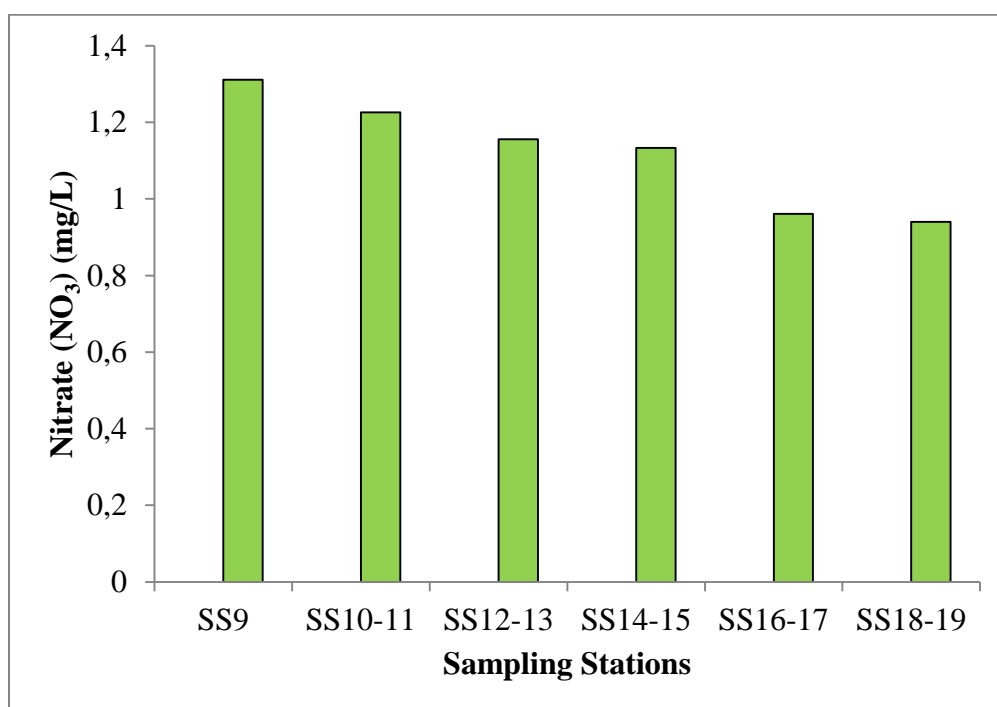


Figure C. 6 Average Nitrate distribution within PDR (AKS, 2010)

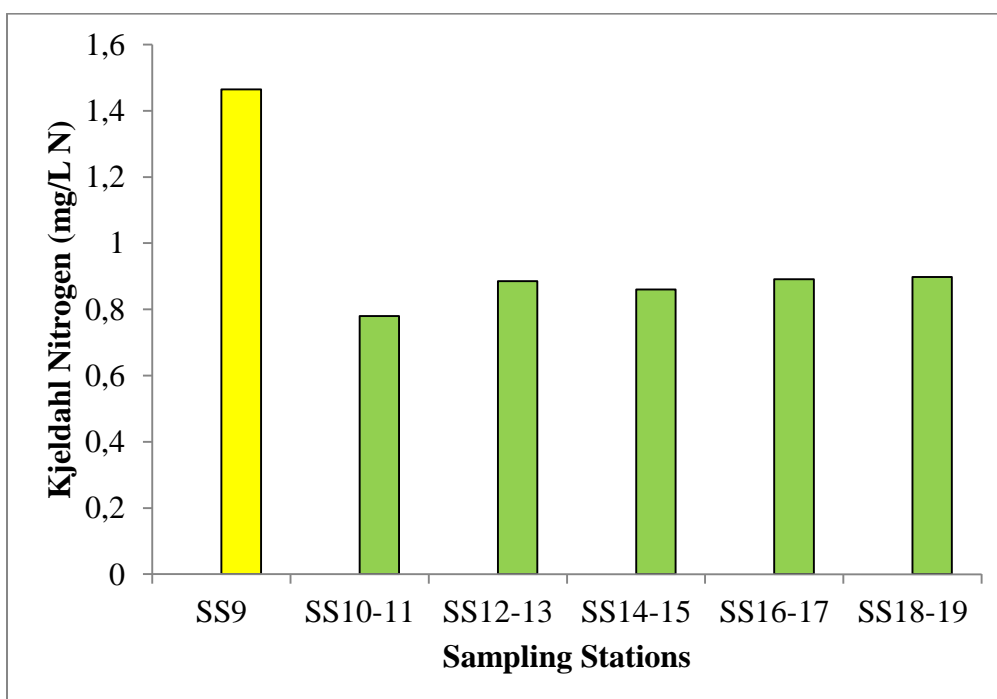


Figure C. 7 Average Kjeldahl Nitrogen distribution within PDR (AKS, 2010)

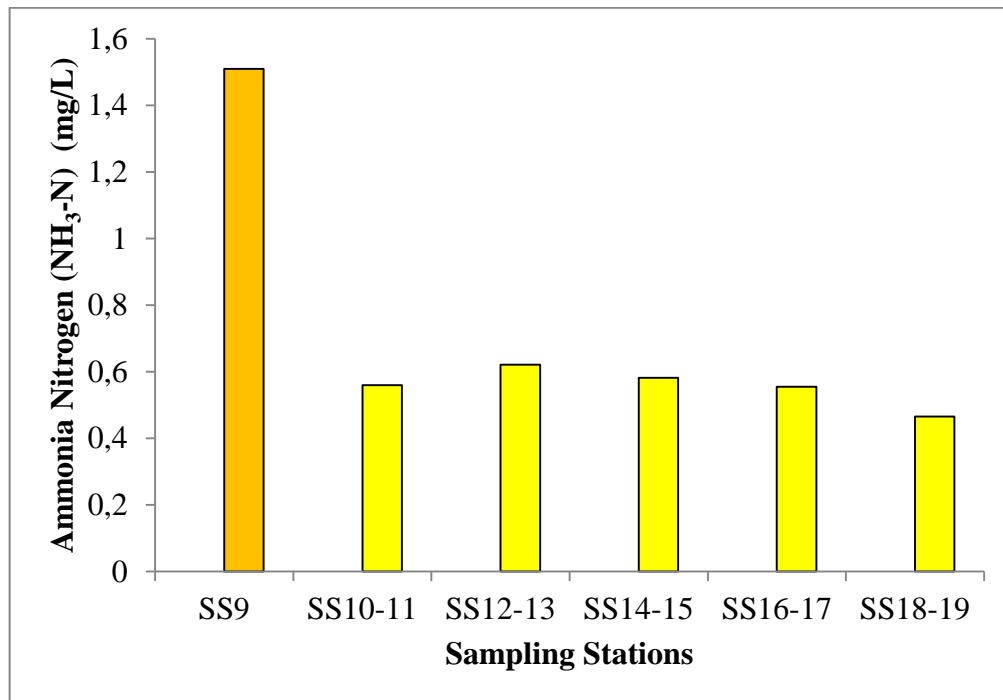


Figure C. 8 Average Ammonia Nitrogen distribution within PDR (AKS, 2010)

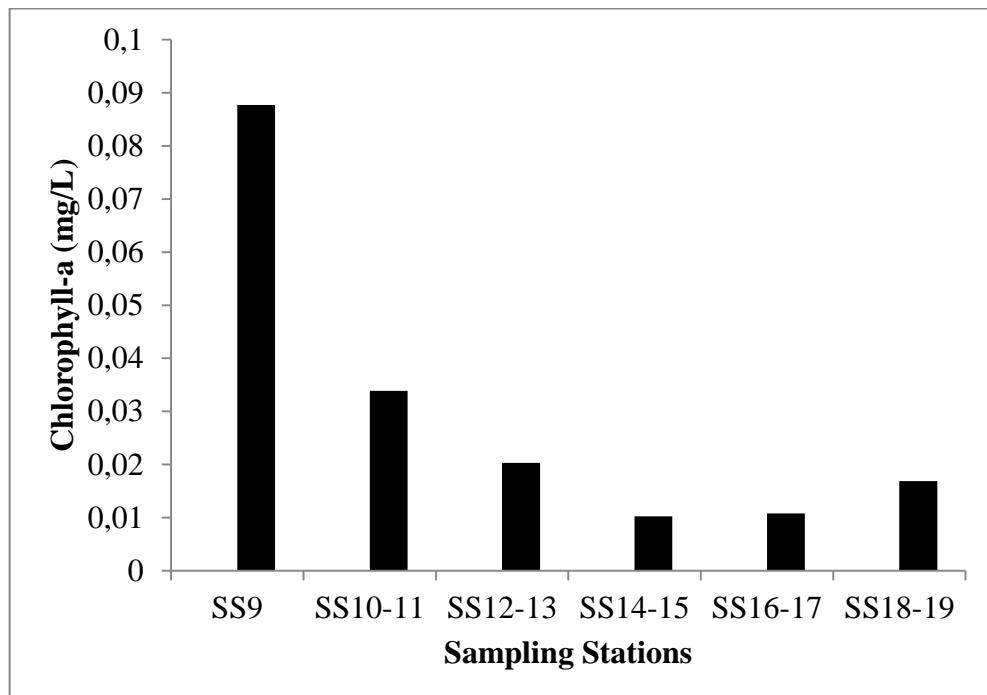


Figure C. 9 Average Chlorophyll-a distribution within PDR (AKS, 2010)

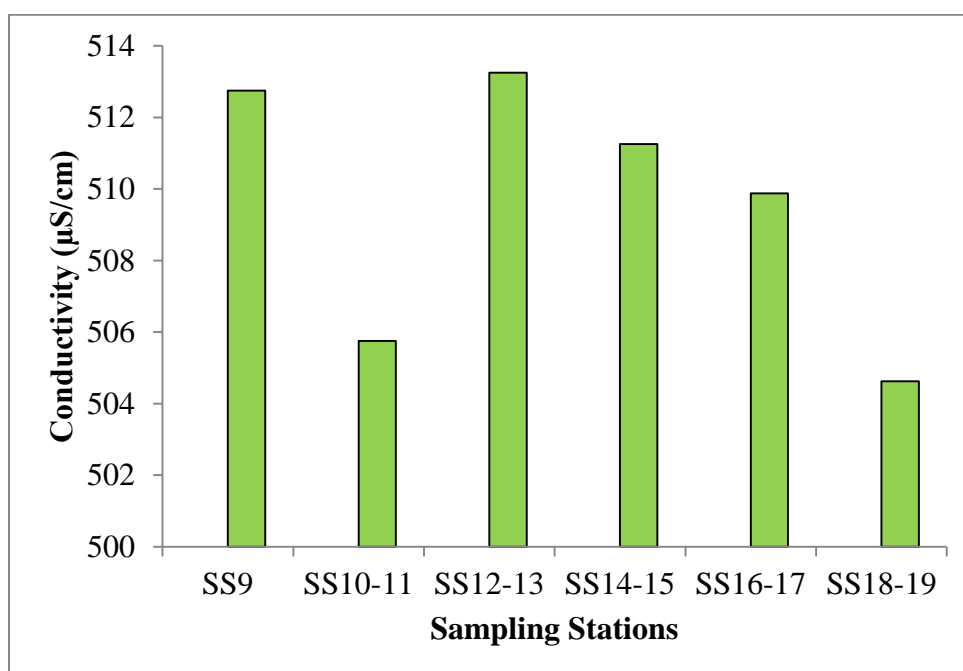


Figure C. 10 Average Conductivity distribution within PDR (AKS, 2010)

APPENDIX D

KDE MAPS

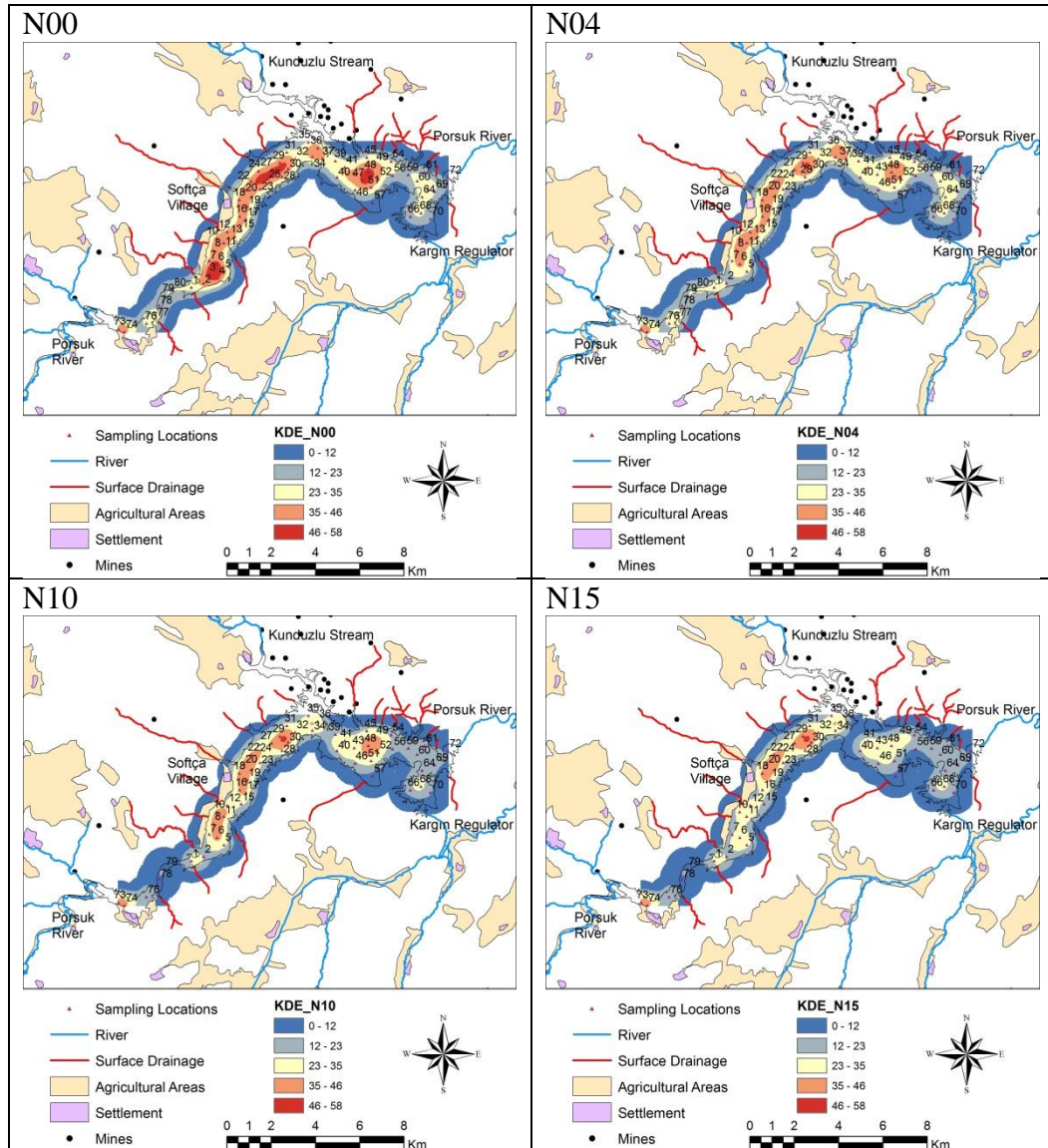


Figure D. 1 KDE map of DO for each network

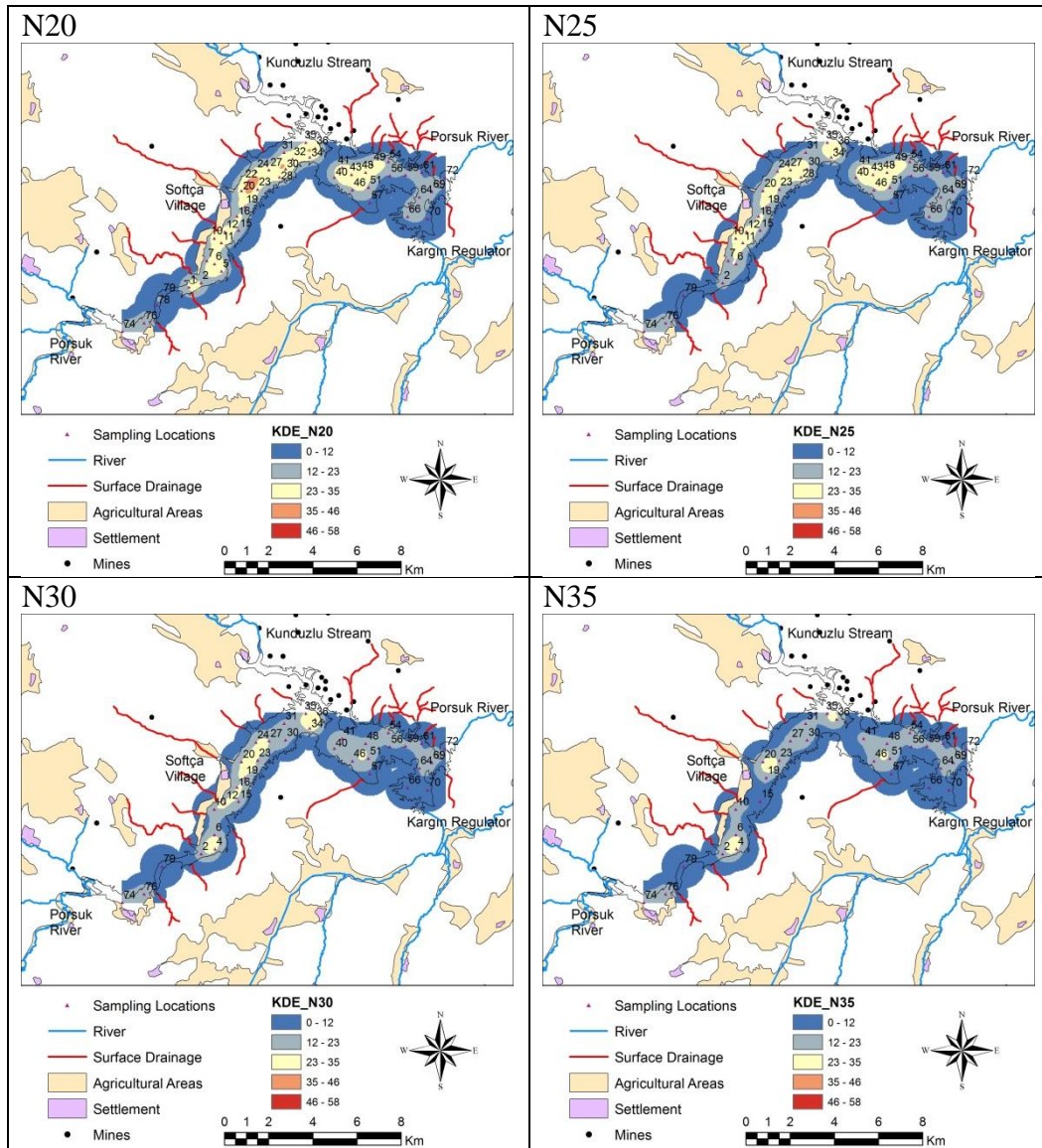


Figure D. 1 KDE map of DO for each network (Continued)

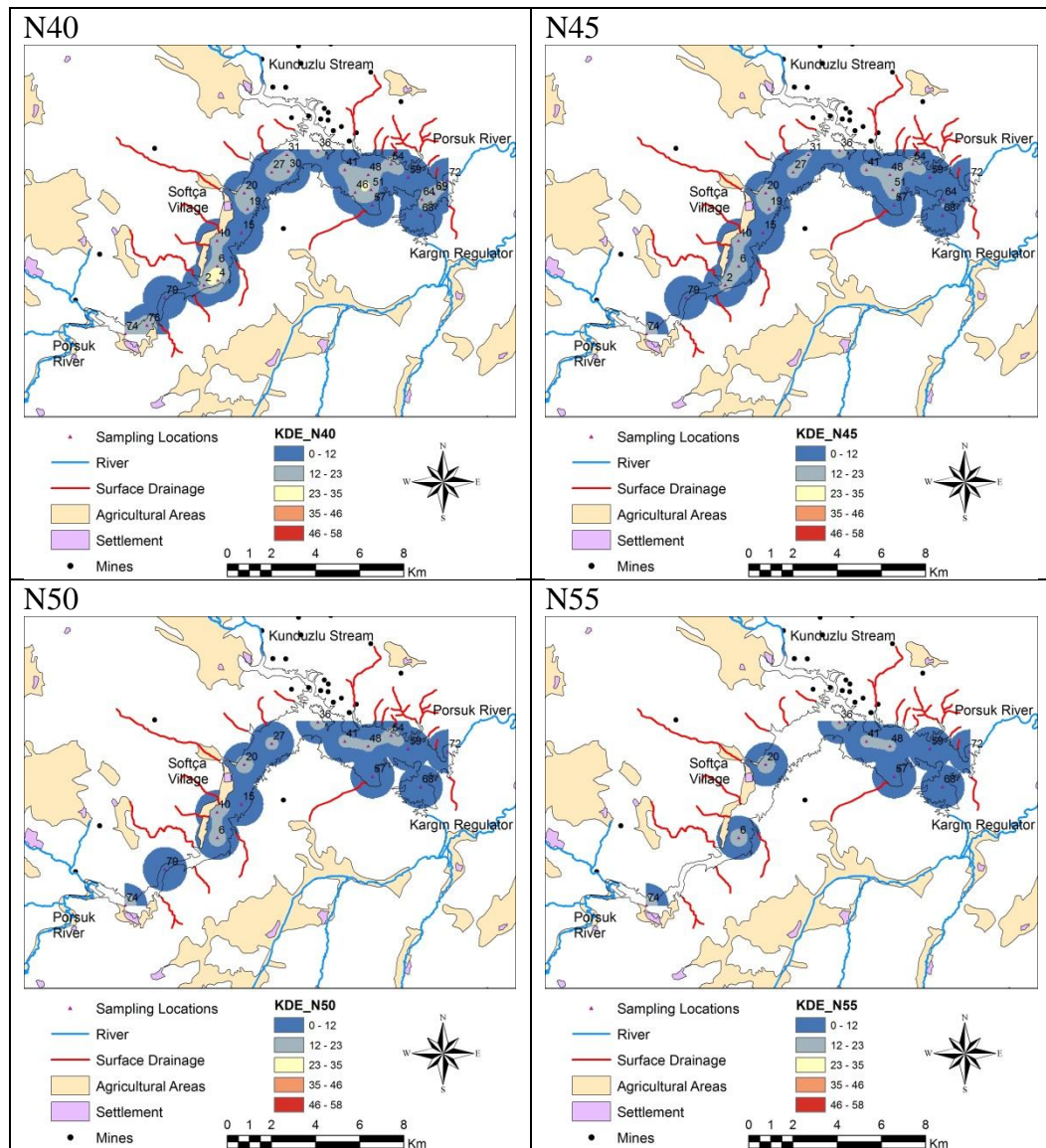


Figure D. 1 KDE map of DO for each network (Continued)

APPENDIX E

VARIOGRAMS

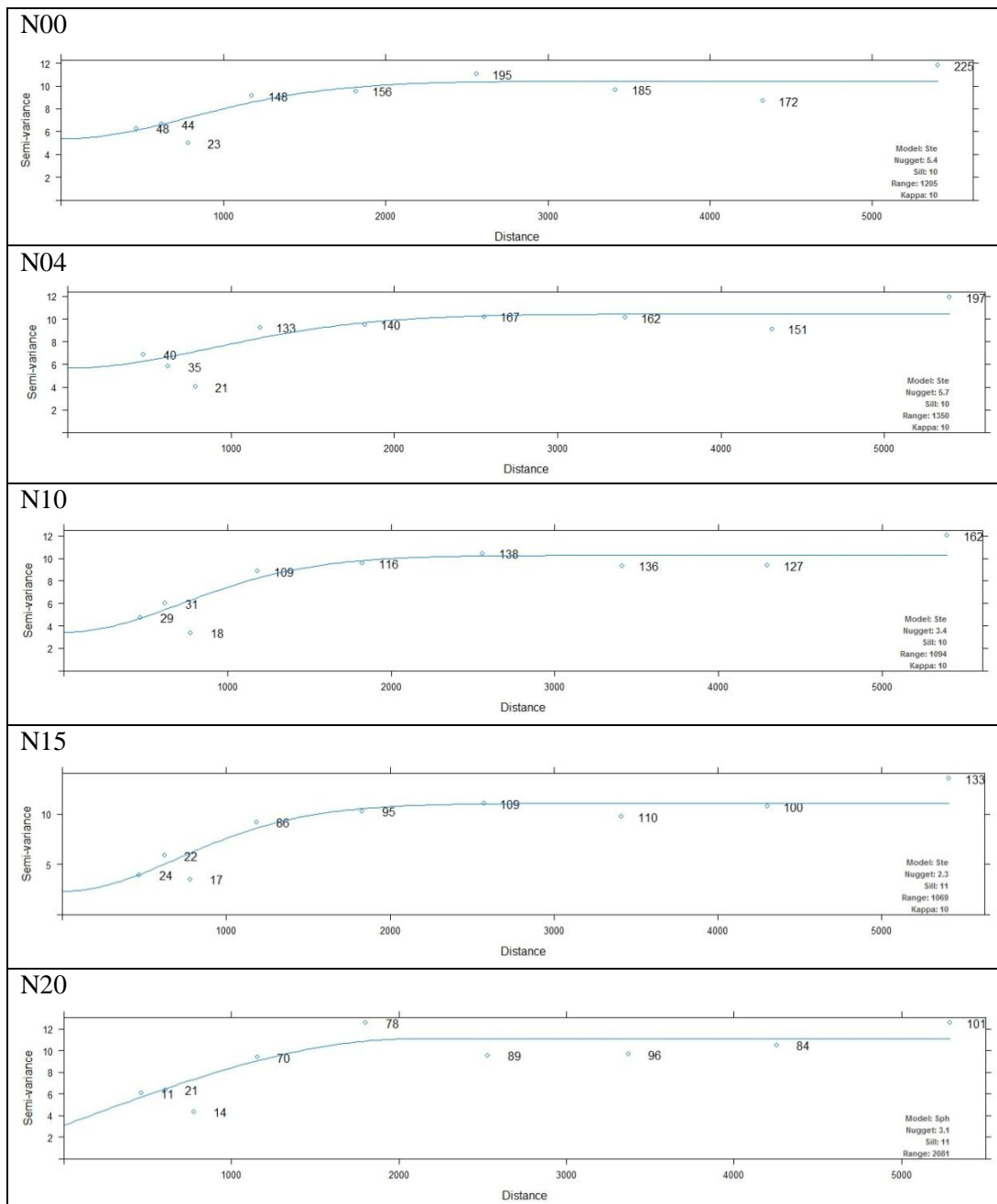


Figure E. 1 DO variogram for each network

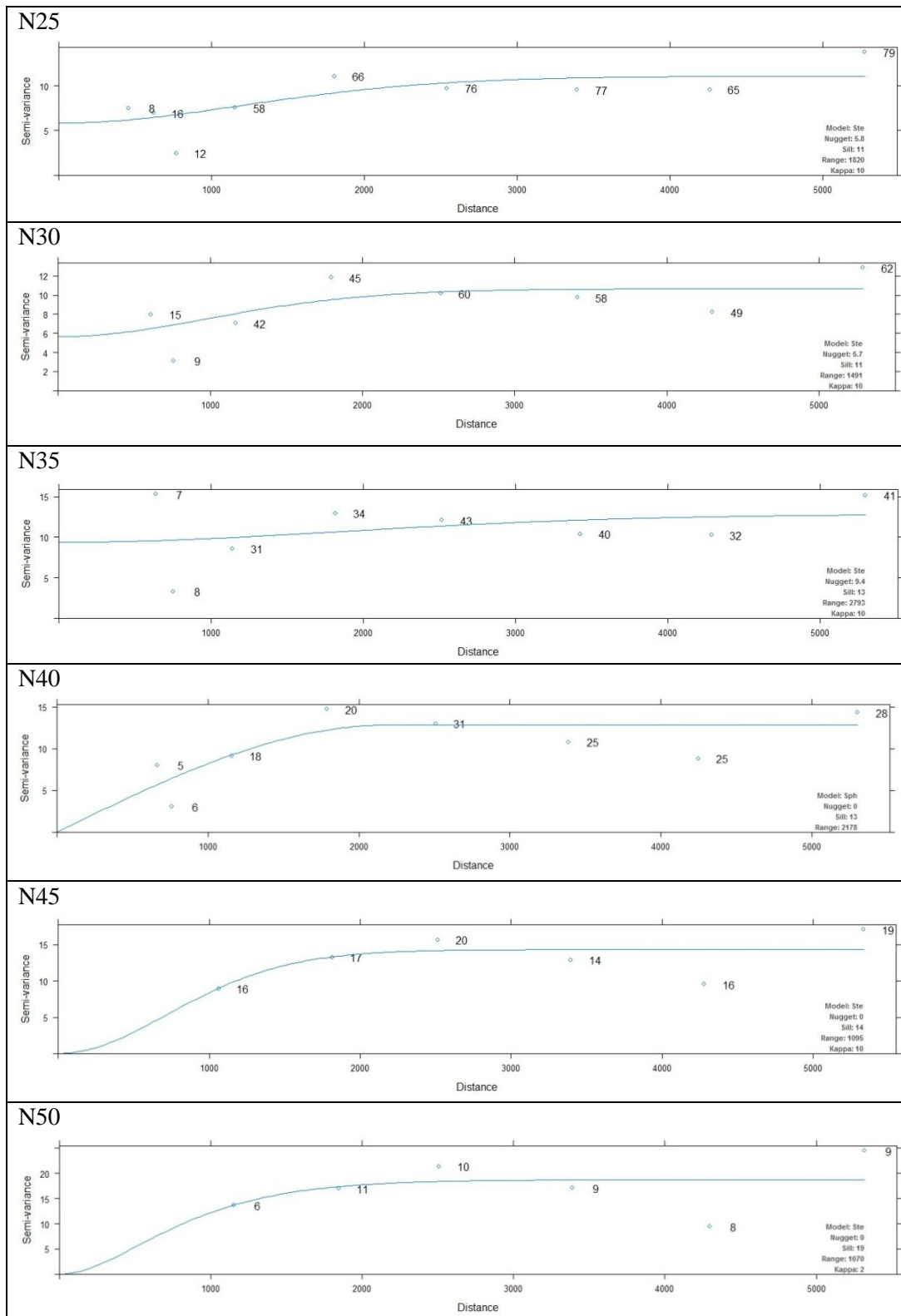


Figure E. 1 DO variogram for each network (Continued)

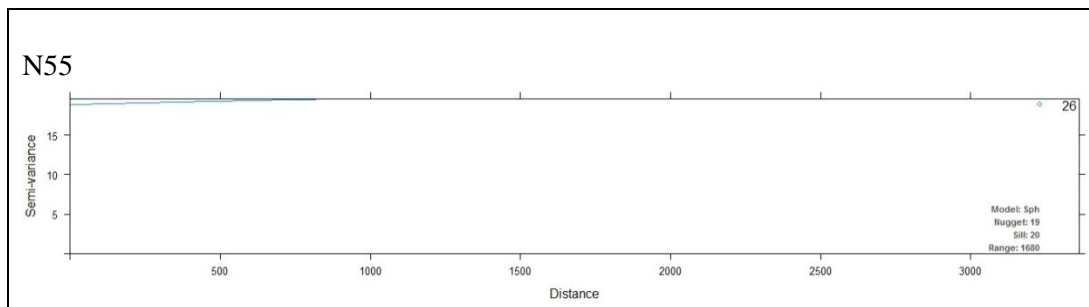


Figure E. 1 DO variogram for each network (Continued)

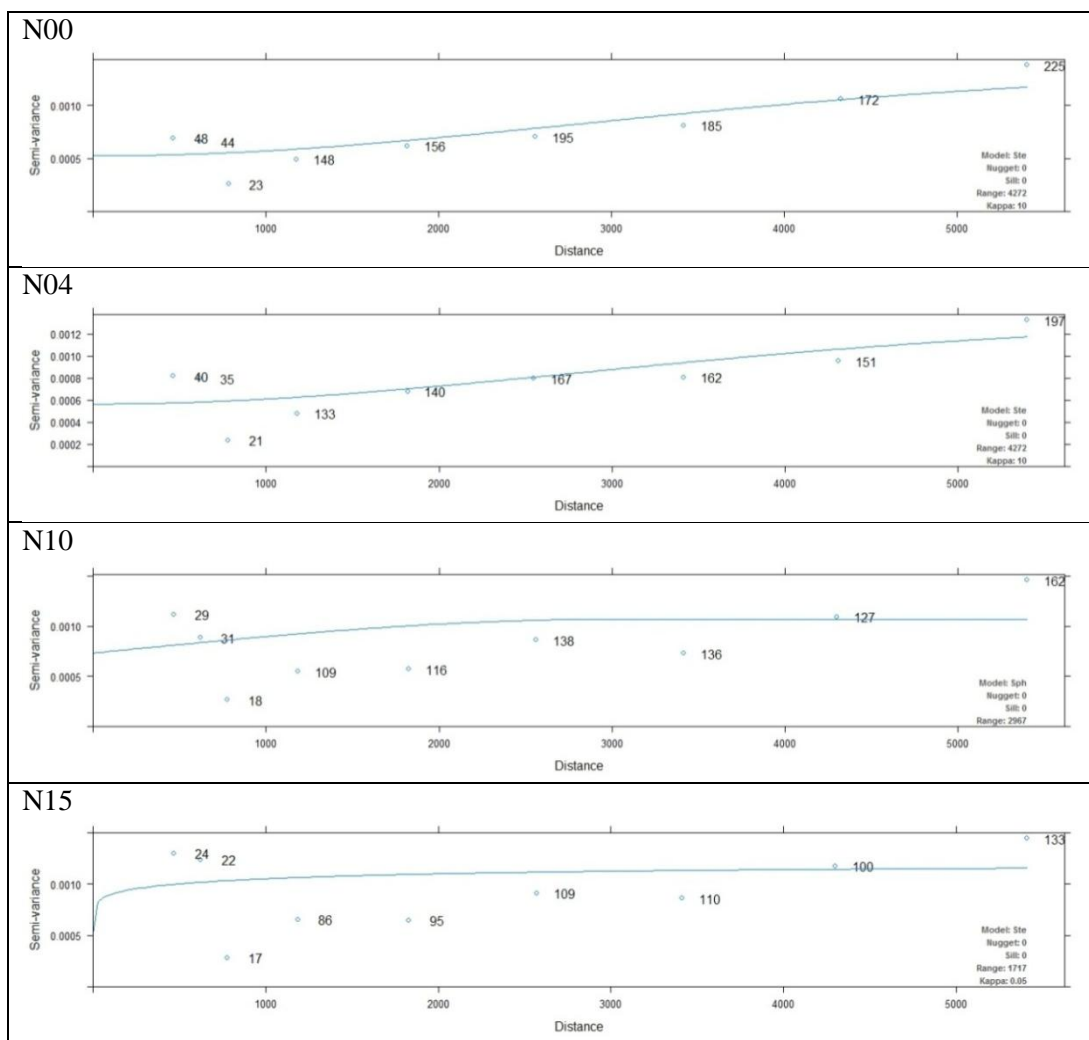


Figure E. 2 SC variogram for each network

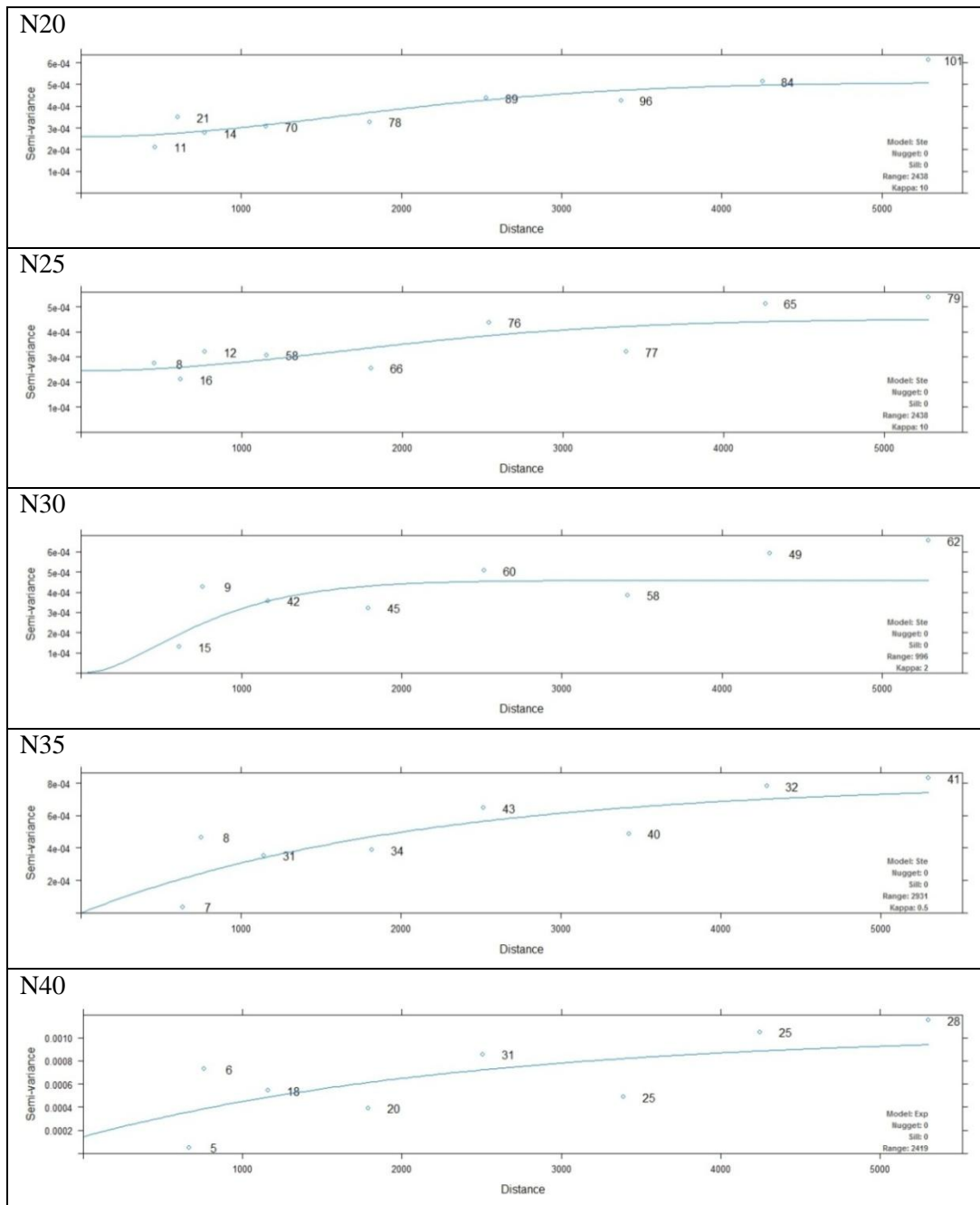


Figure E. 2 SC variogram for each network (Continued)

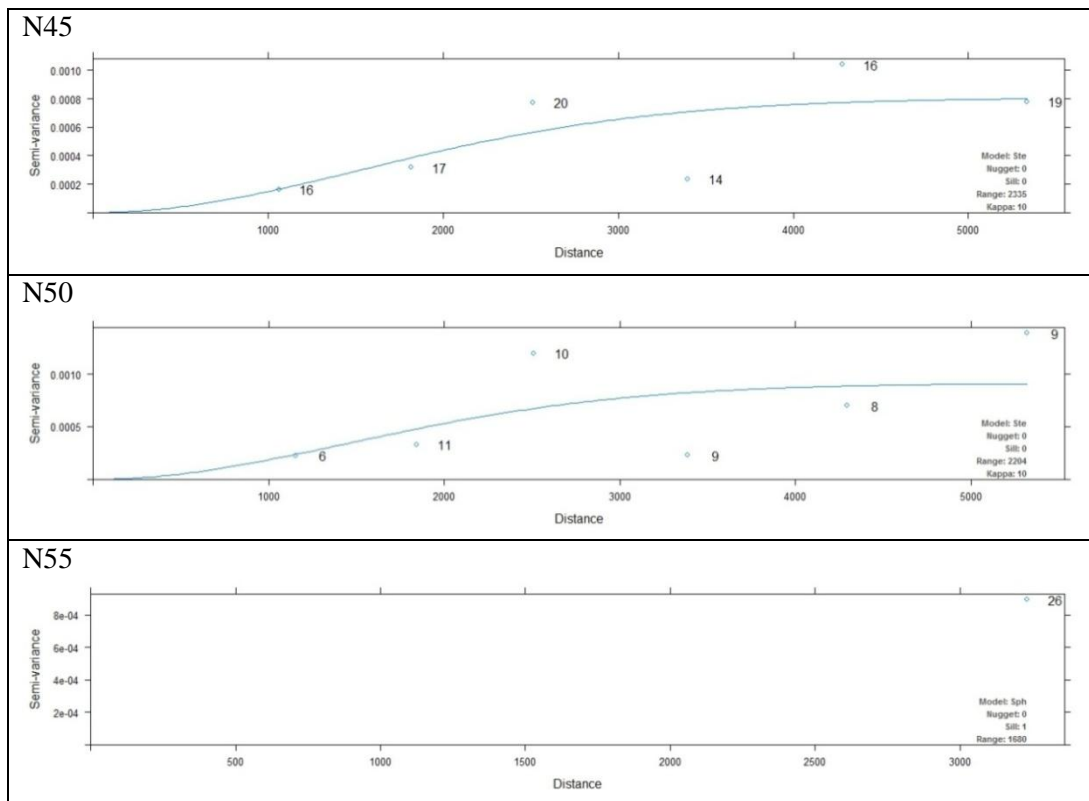


Figure E. 2 SC variogram for each network (Continued)

APPENDIX F

RELATIONSHIP BETWEEN $\gamma(h)$ OF N00 VARIOGRAM MODEL AND $\gamma(h)$ OF NXX VARIOGRAM MODEL

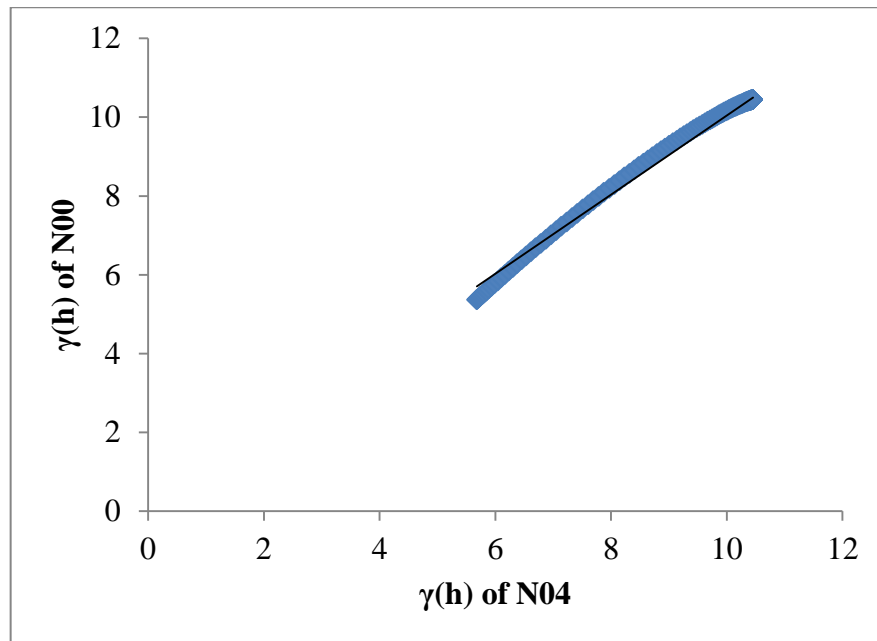


Figure F. 1 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N04 variogram model

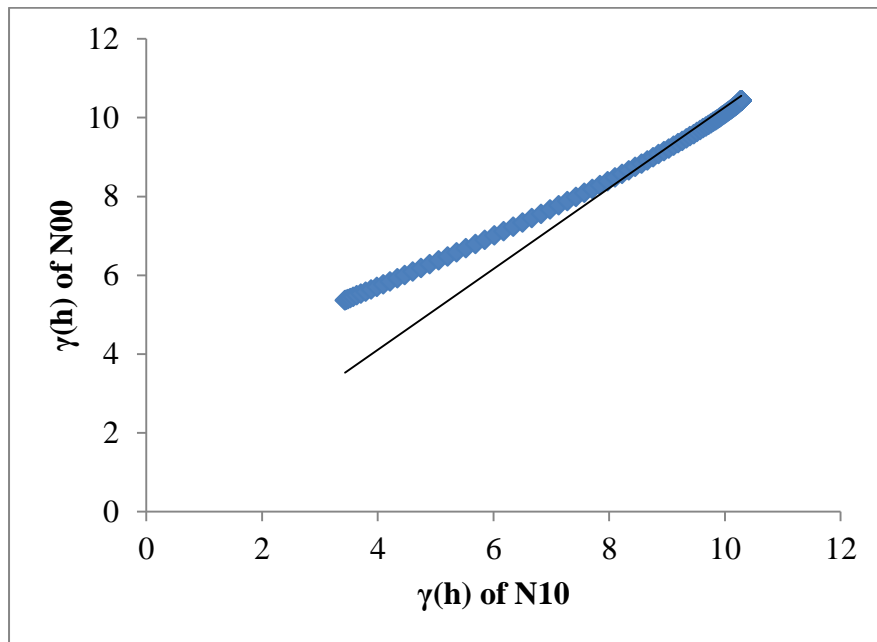


Figure F. 2 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N10 variogram model

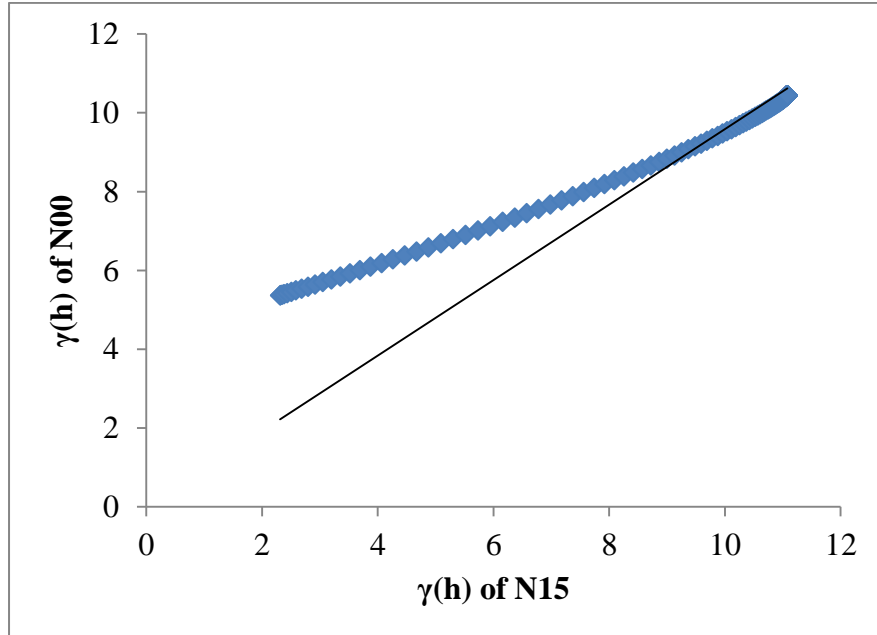


Figure F. 3 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N15 variogram model

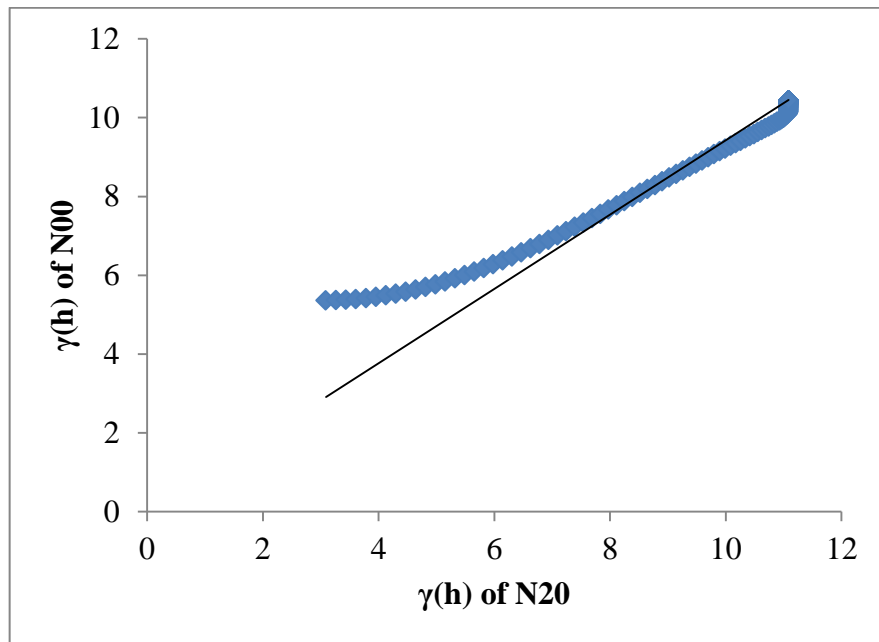


Figure F. 4 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N20 variogram model

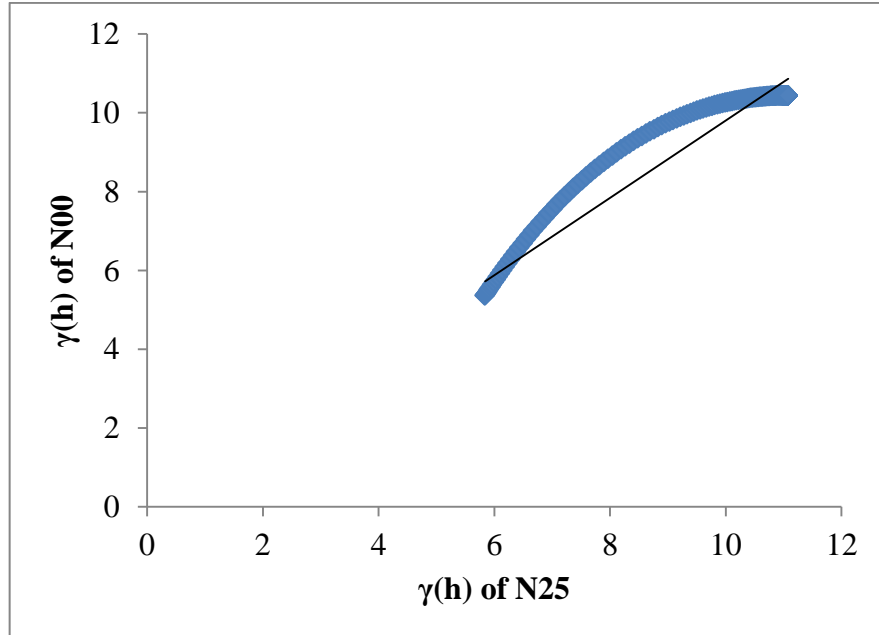


Figure F. 5 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N25 variogram model

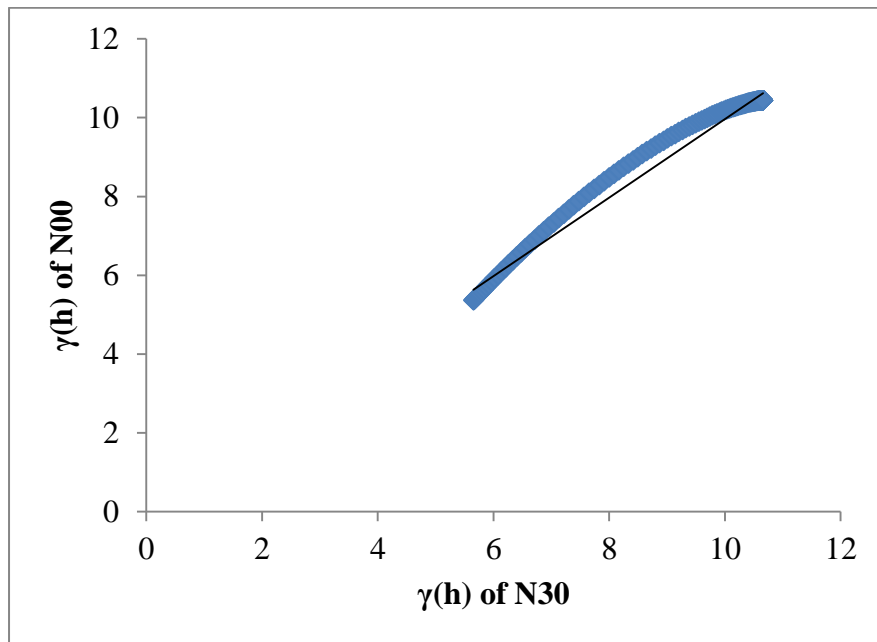


Figure F. 6 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N30 variogram model

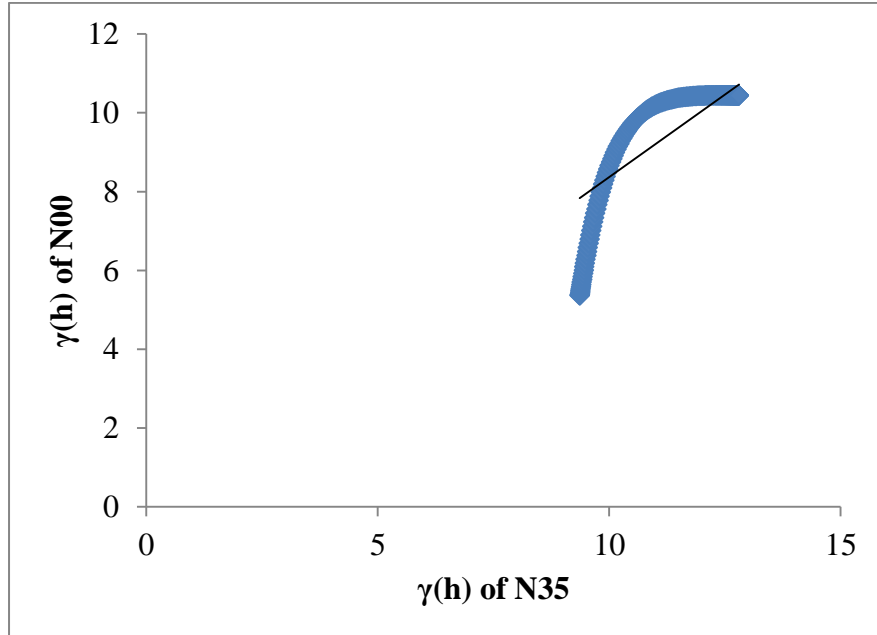


Figure F. 7 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N35 variogram model

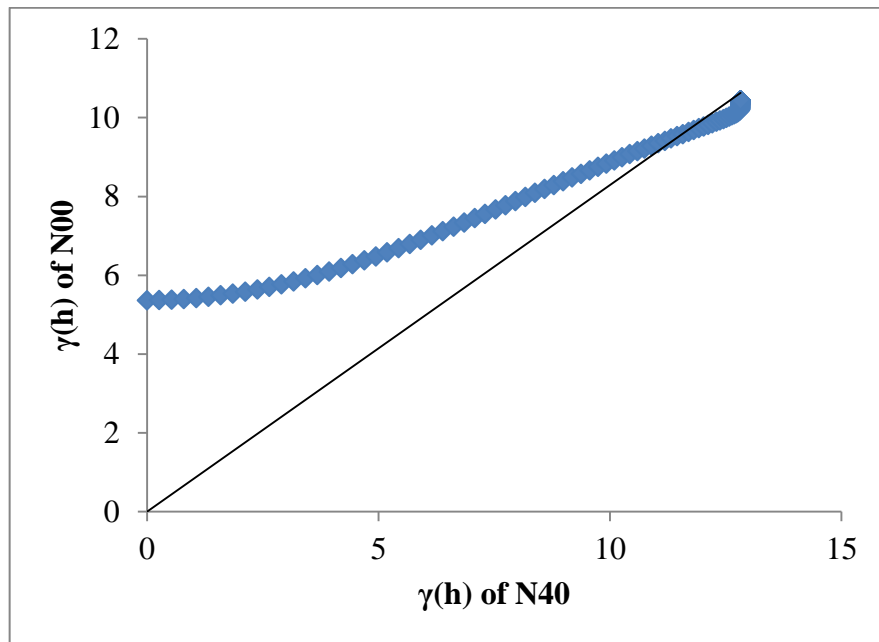


Figure F. 8 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N40 variogram model

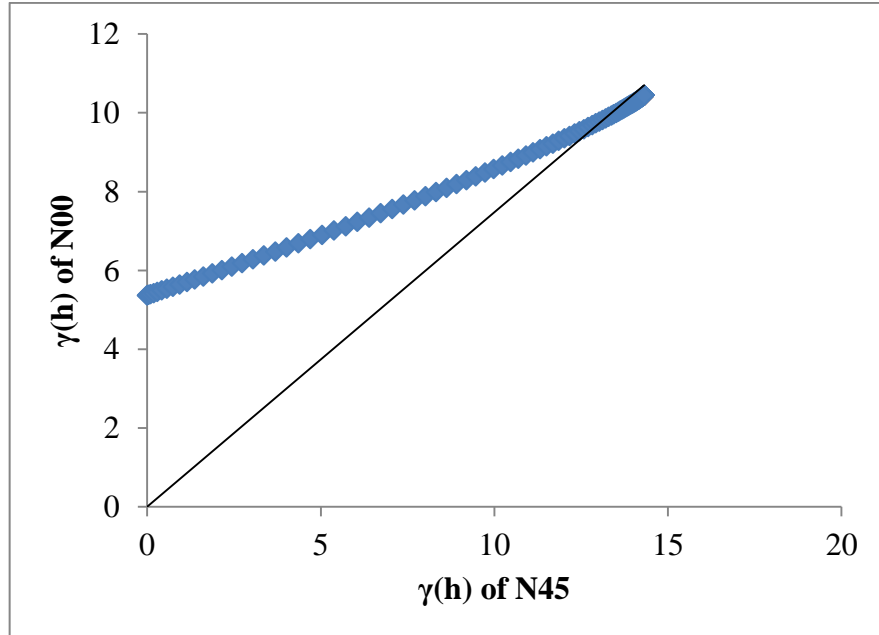


Figure F. 9 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N45 variogram model

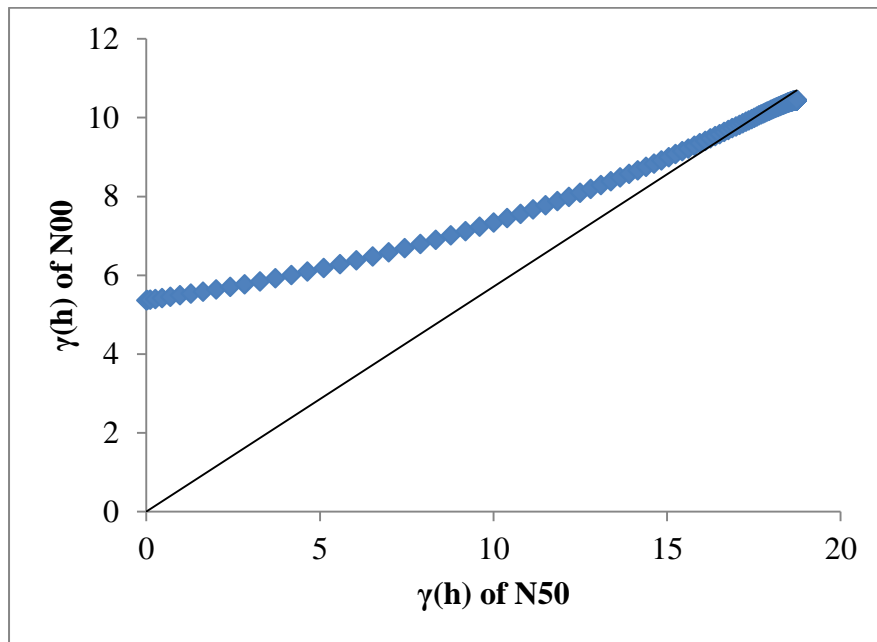


Figure F. 10 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N50 variogram model

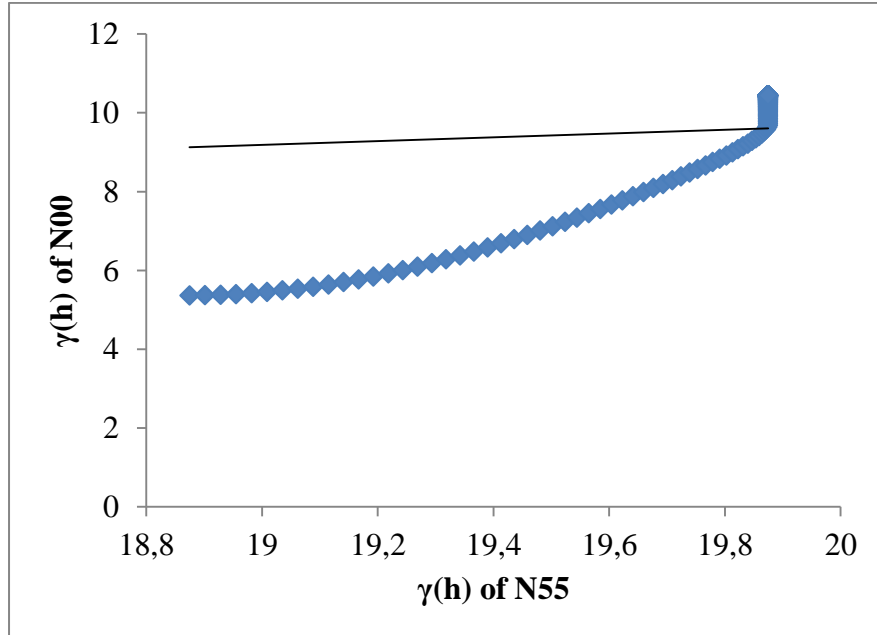


Figure F. 11 Relationship between $\gamma(h)$ of N00 variogram model and $\gamma(h)$ of N55 variogram model

APPENDIX G

OK AND ERROR MAPS OF DO

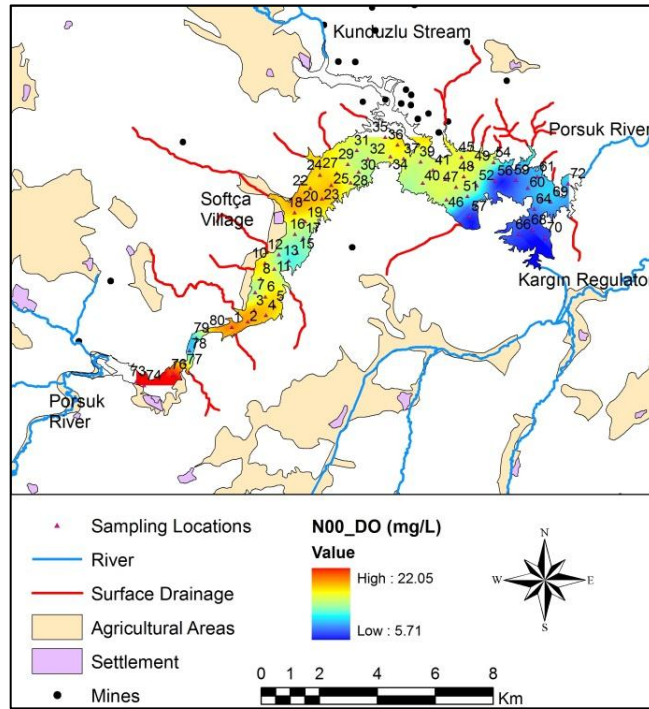


Figure G. 1 OK of DO using N00 data set

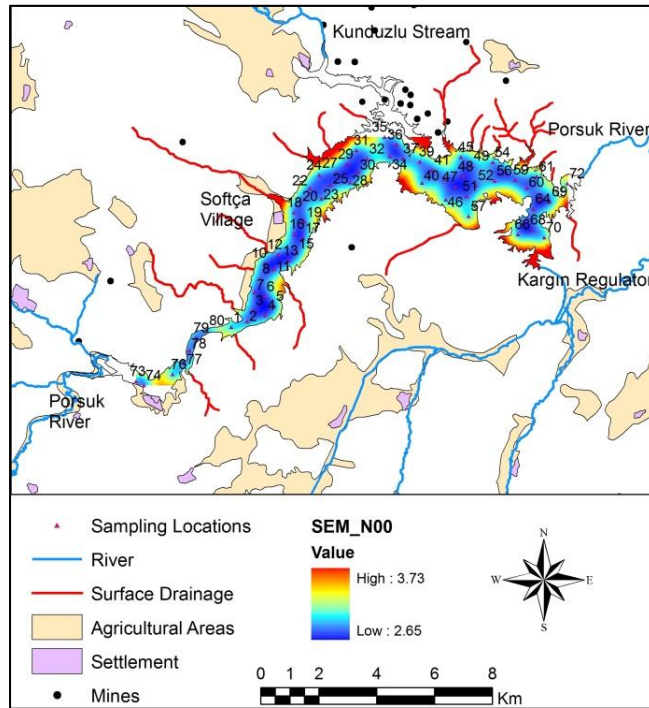


Figure G. 2 SEM of DO using N00 data set

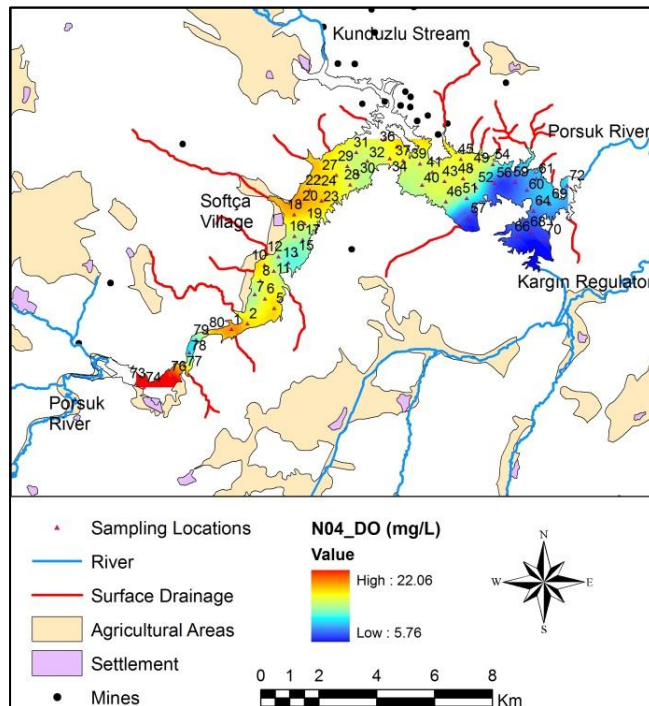


Figure G. 3 OK of DO using N04 data set

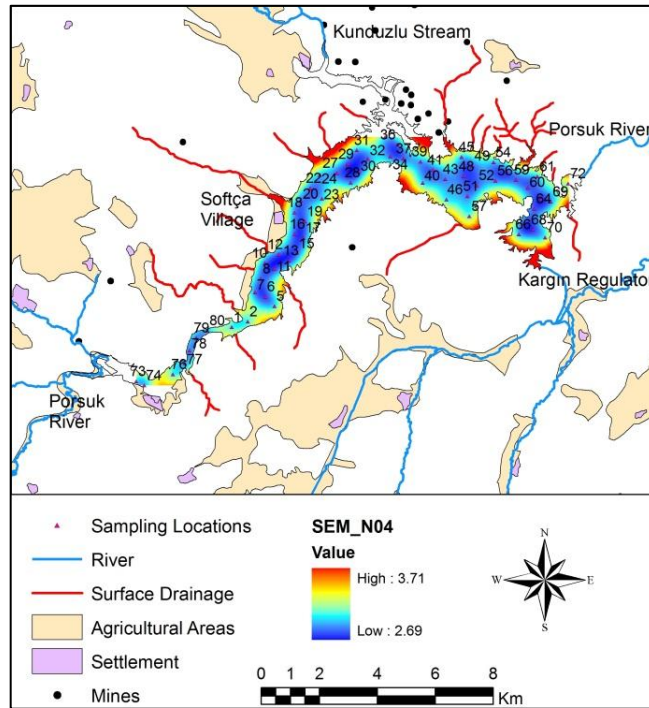


Figure G. 4 SEM of DO using N04 data set

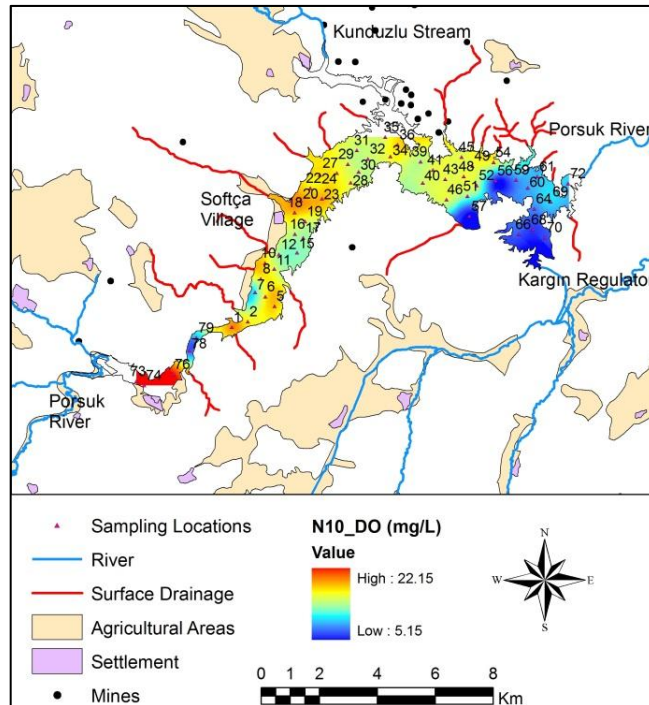


Figure G. 5 OK of DO using N10 data set

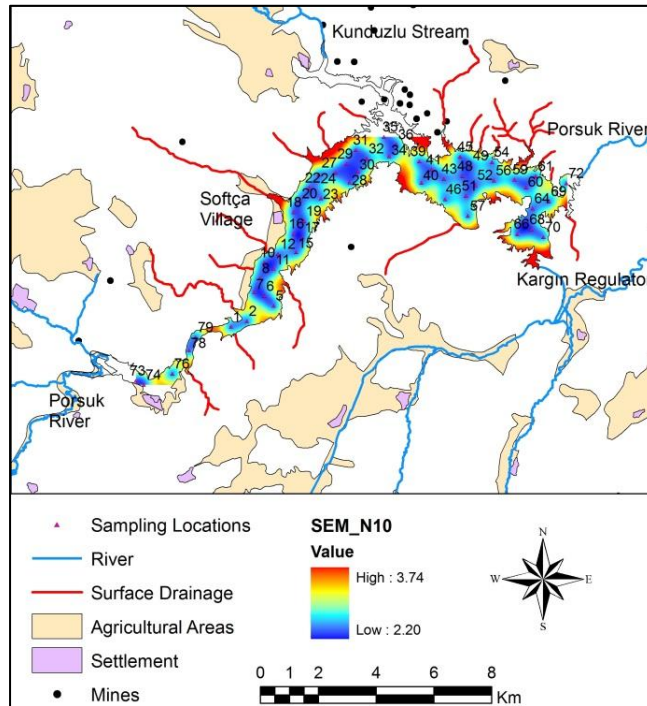


Figure G. 6 SEM of DO using N10 data set

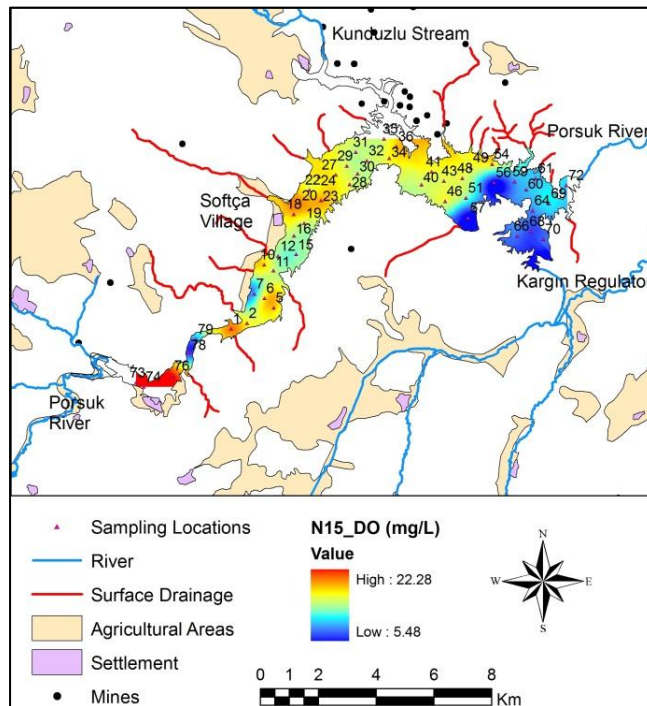


Figure G. 7 OK of DO using N15 data set

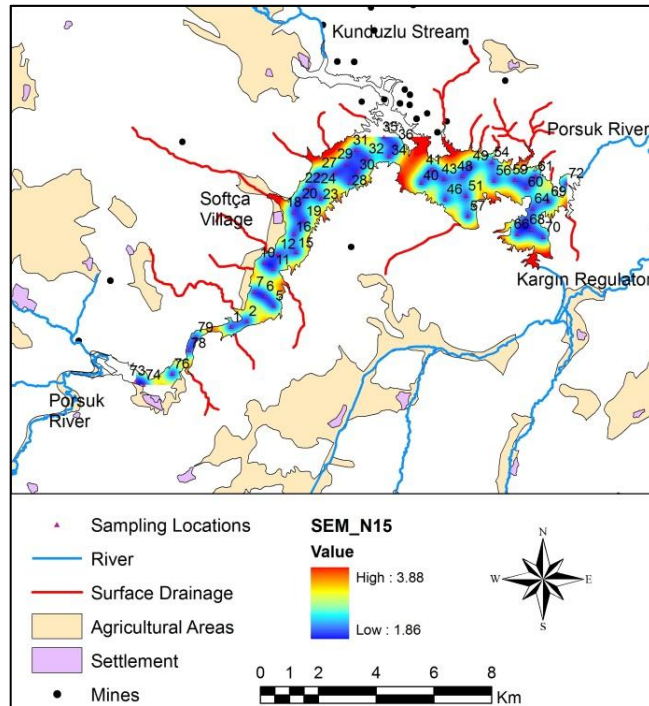


Figure G. 8 SEM of DO using N15 data set

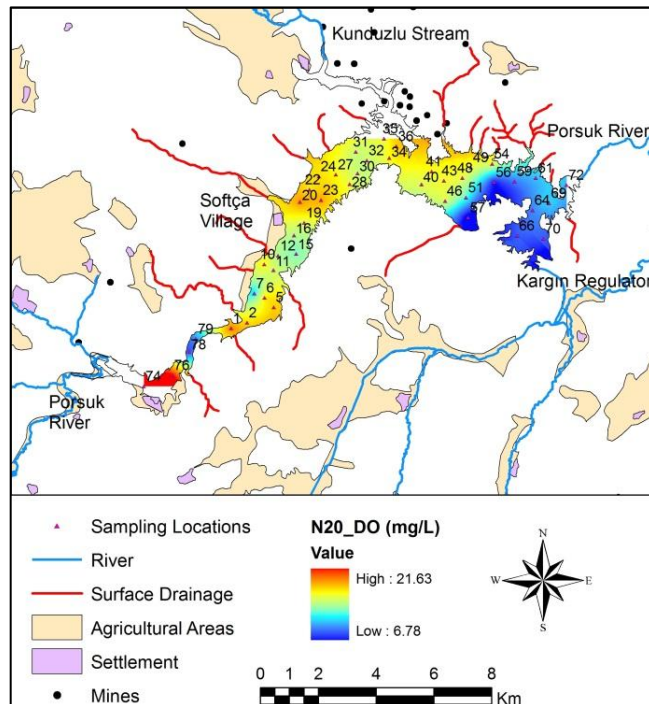


Figure G. 9 OK of DO using N20 data set

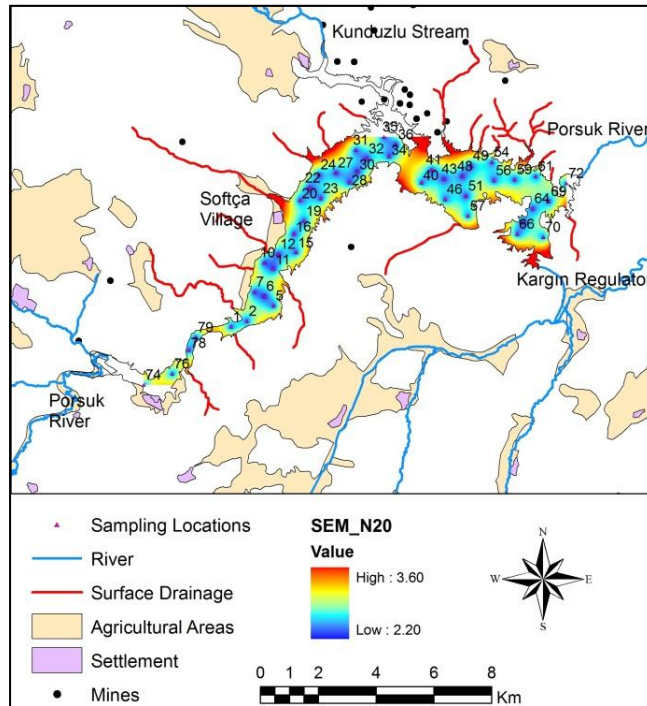


Figure G. 10 SEM of DO using N20 data set

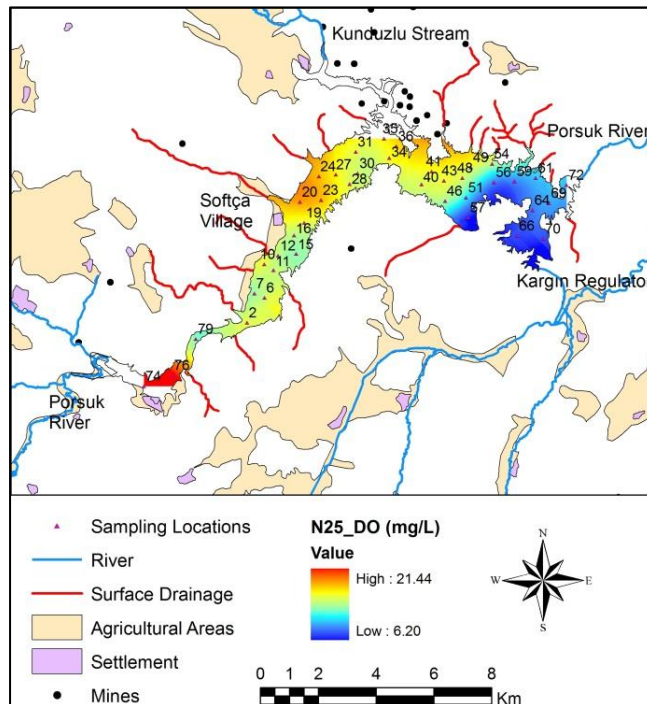


Figure G. 11 OK of DO using N25 data set

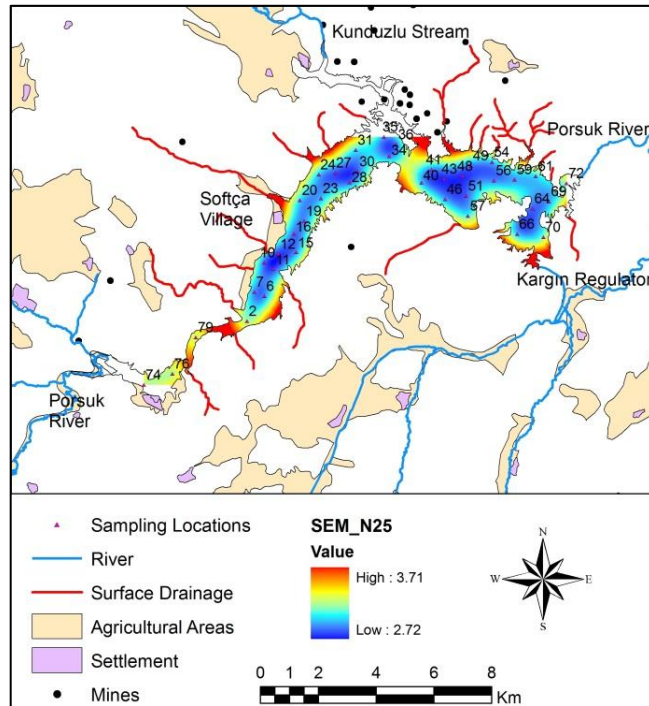


Figure G. 12 SEM of DO using N25 data set

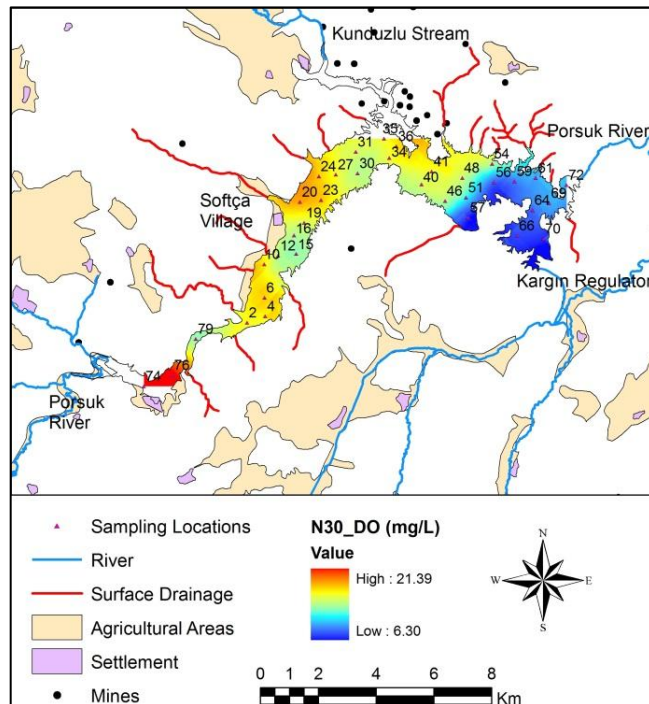


Figure G. 13 OK of DO using N30 data set

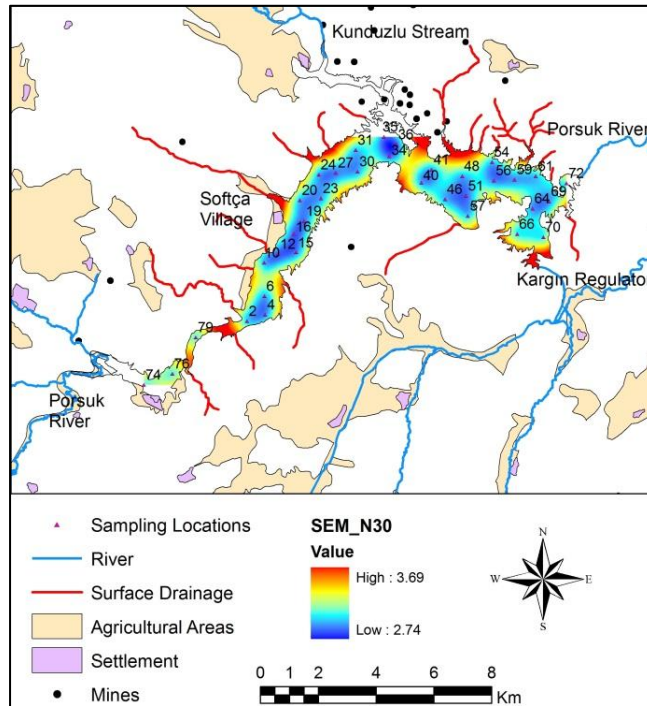


Figure G. 14 SEM of DO using N30 data set

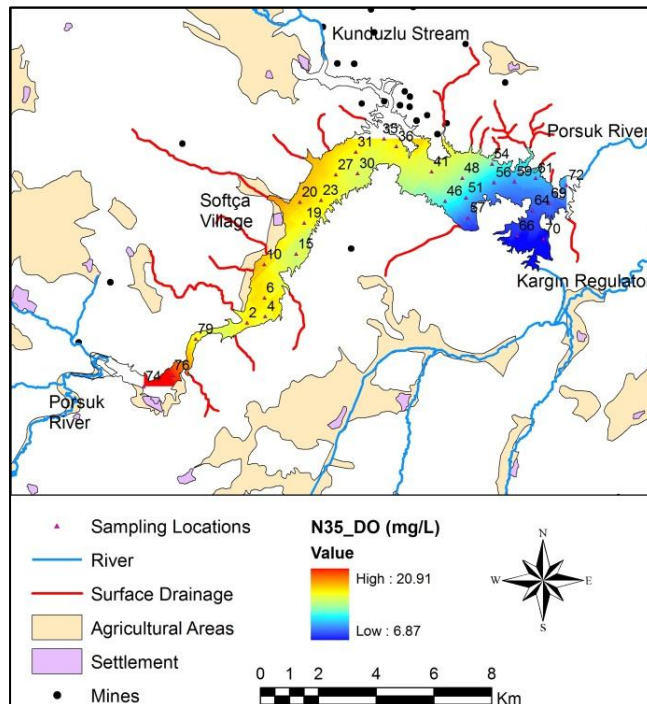


Figure G. 15 OK of DO using N35 data set

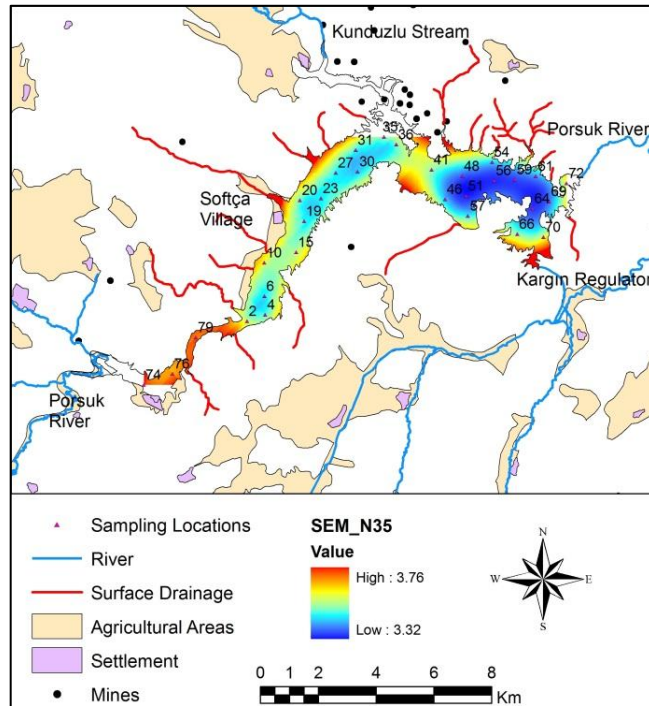


Figure G. 16 SEM of DO using N35 data set

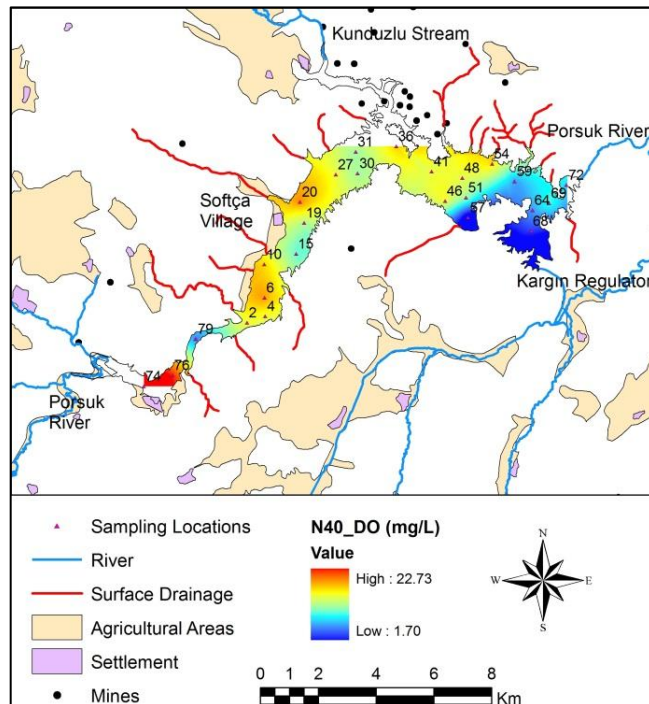


Figure G. 17 OK of DO using N40 data set

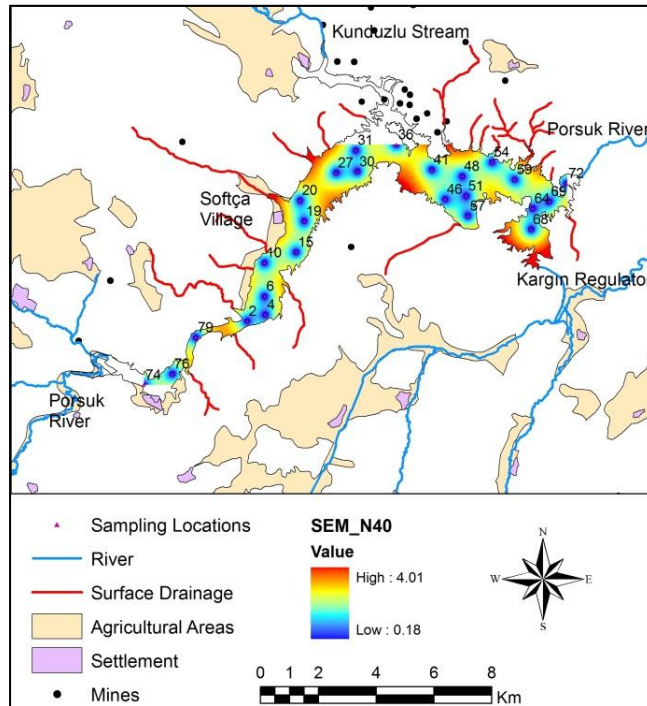


Figure G. 18 SEM of DO using N40 data set

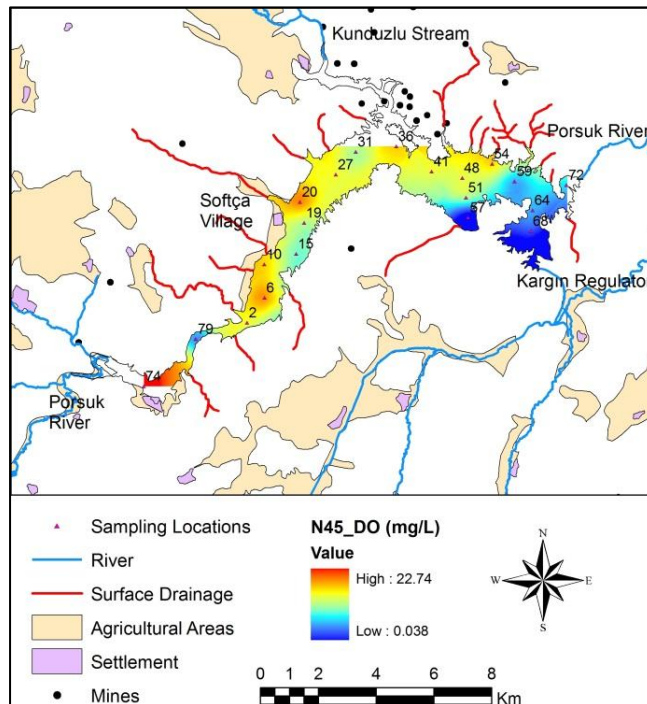


Figure G. 19 OK of DO using N45 data set

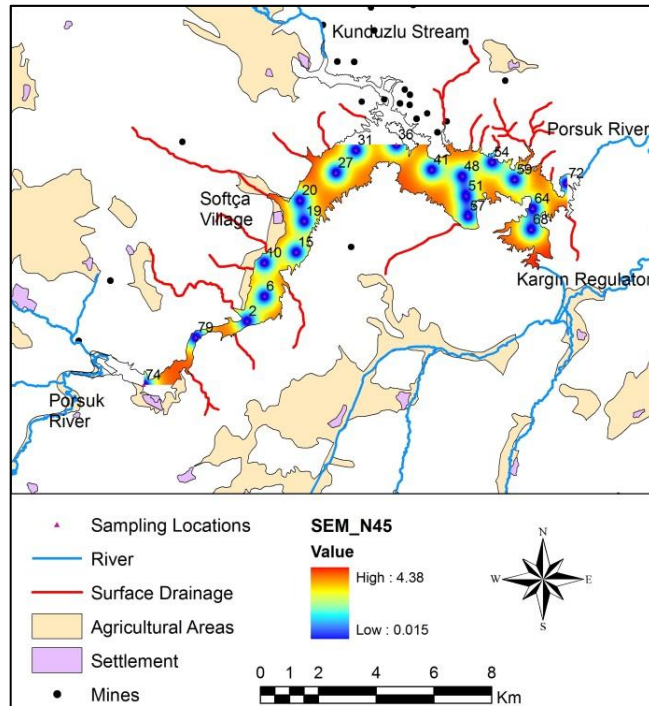


Figure G. 20 SEM of DO using N45 data set

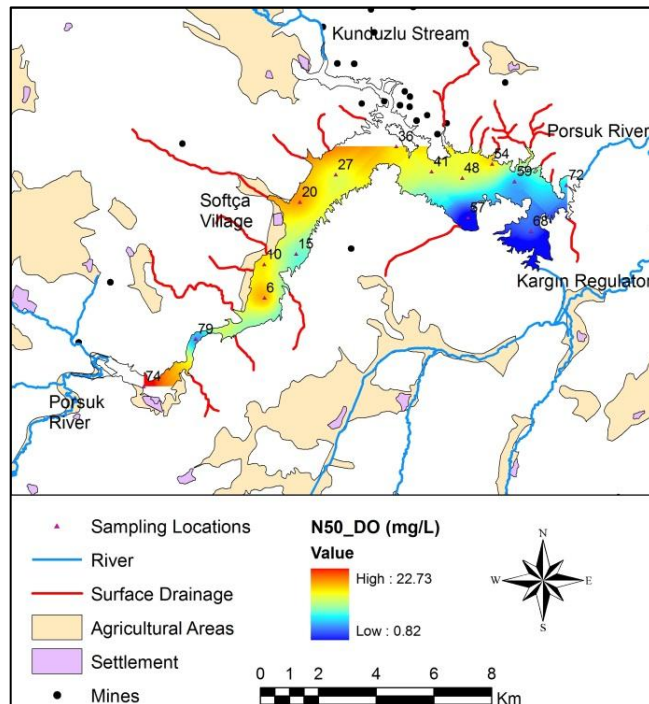


Figure G. 21 OK of DO using N50 data set

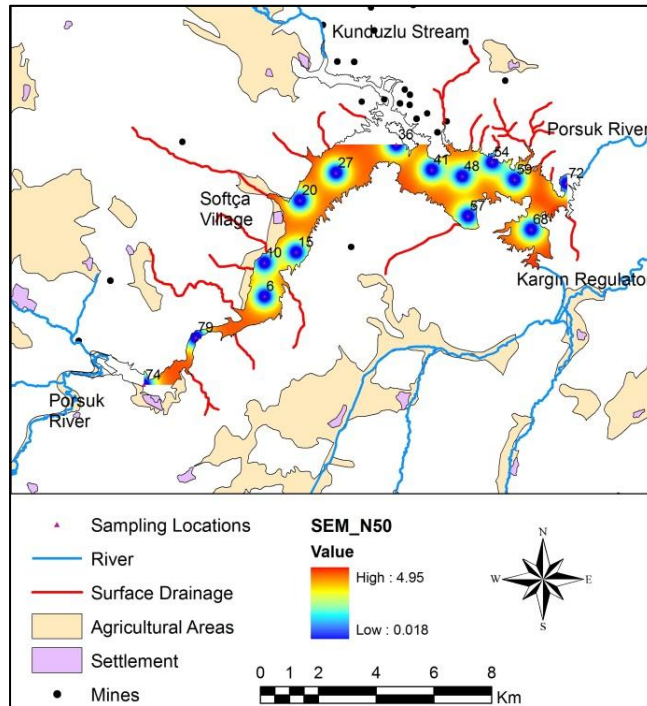


Figure G. 22 SEM of DO using N50 data set

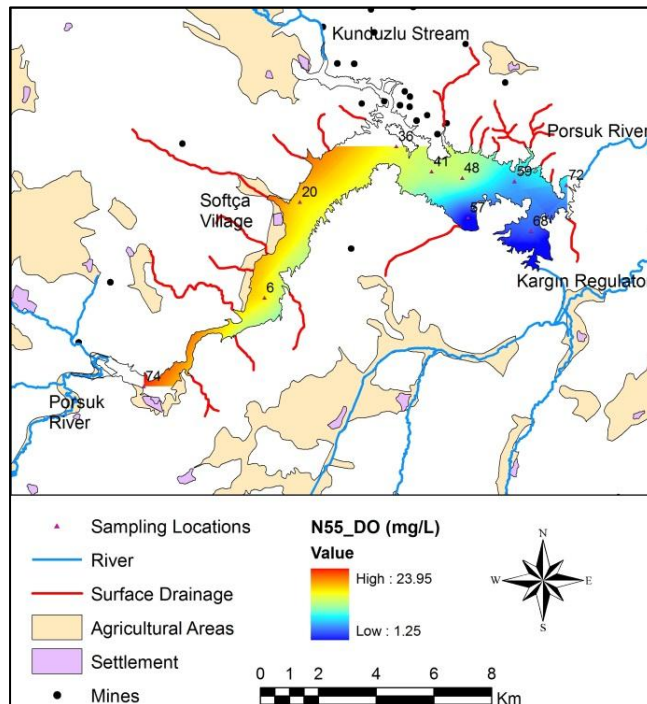


Figure G. 23 OK of DO using N55 data set

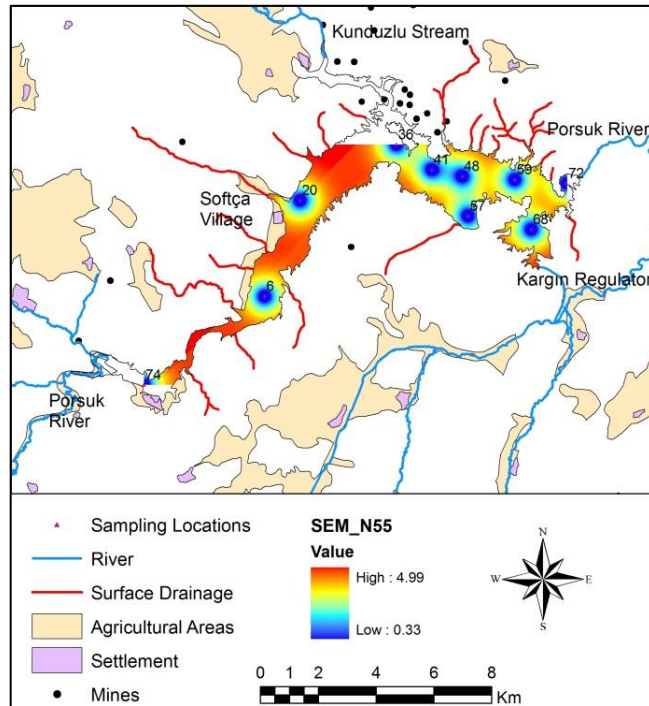


Figure G. 24 SEM of DO using N55 data set

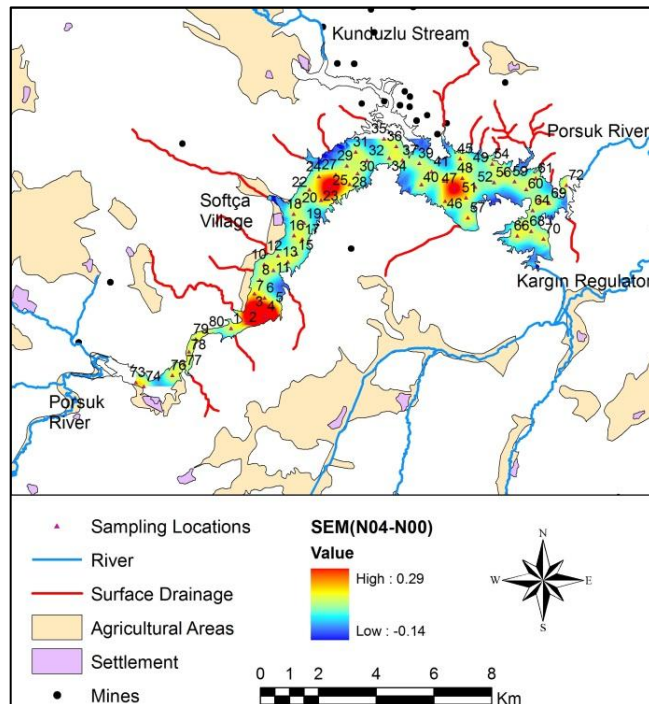


Figure G. 25 Difference of SEMs of DO for N04 and N00

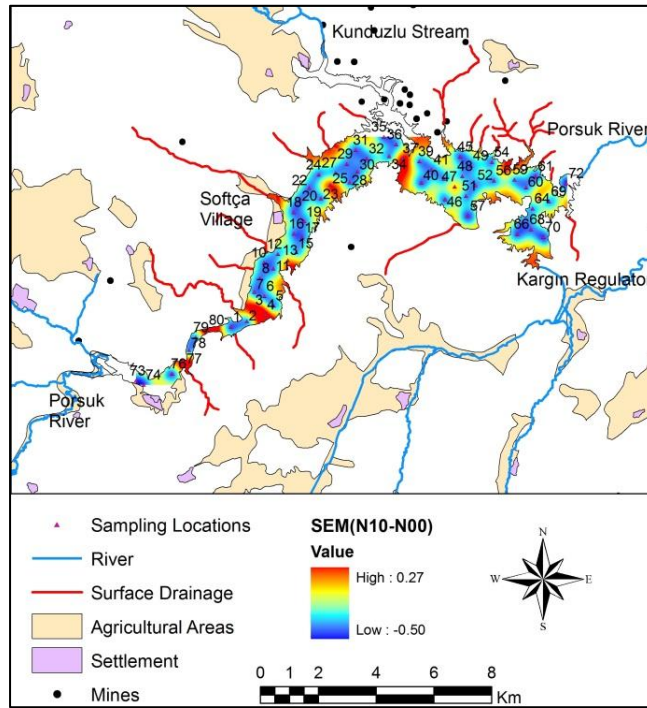


Figure G. 26 Difference of SEMs of DO for N10 and N00

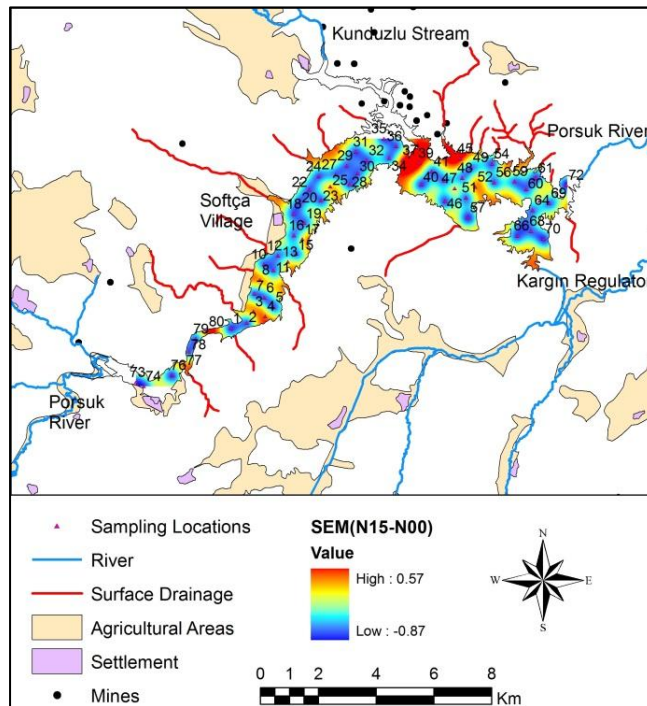


Figure G. 27 Difference of SEMs of DO for N15 and N00

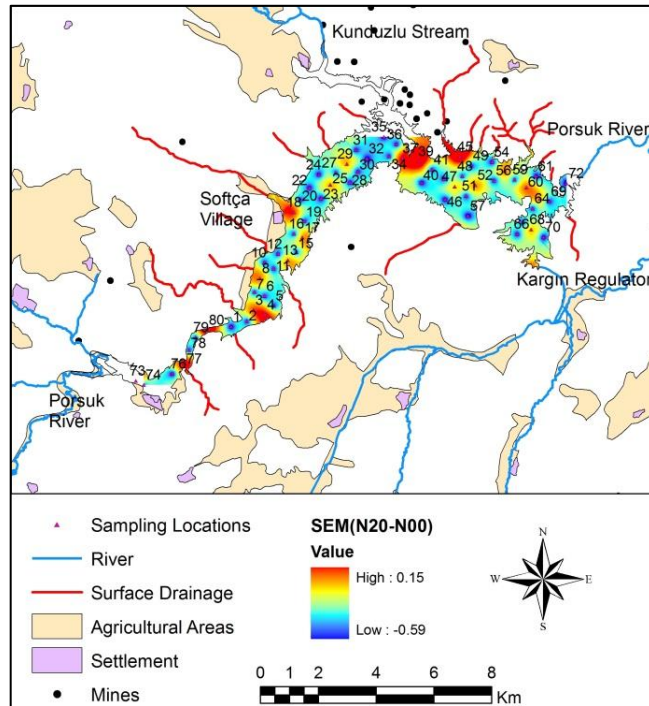


Figure G. 28 Difference of SEMs of DO for N20 and N00

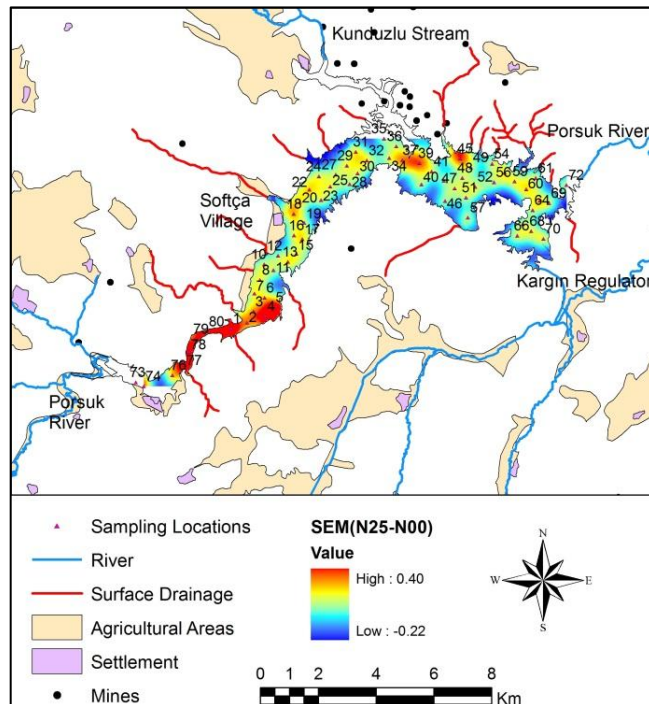


Figure G. 29 Difference of SEMs of DO for N25 and N00

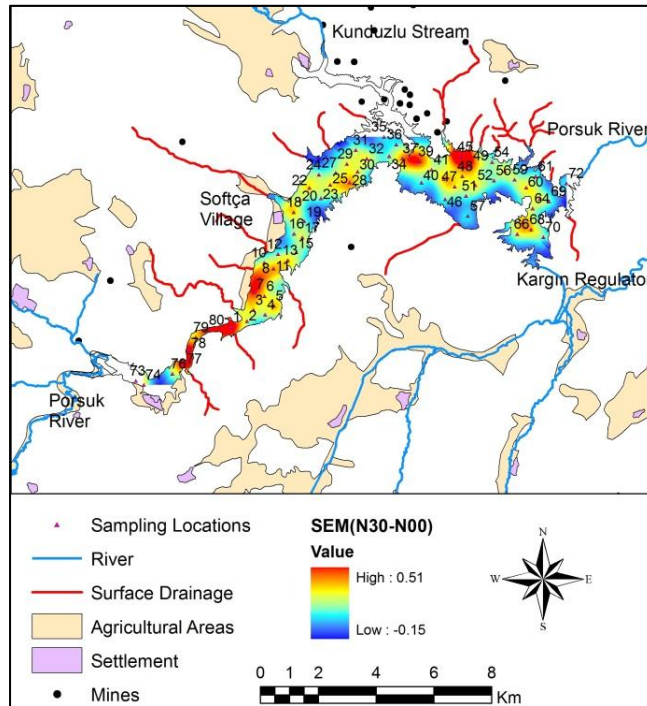


Figure G. 30 Difference of SEMs of DO for N30 and N00

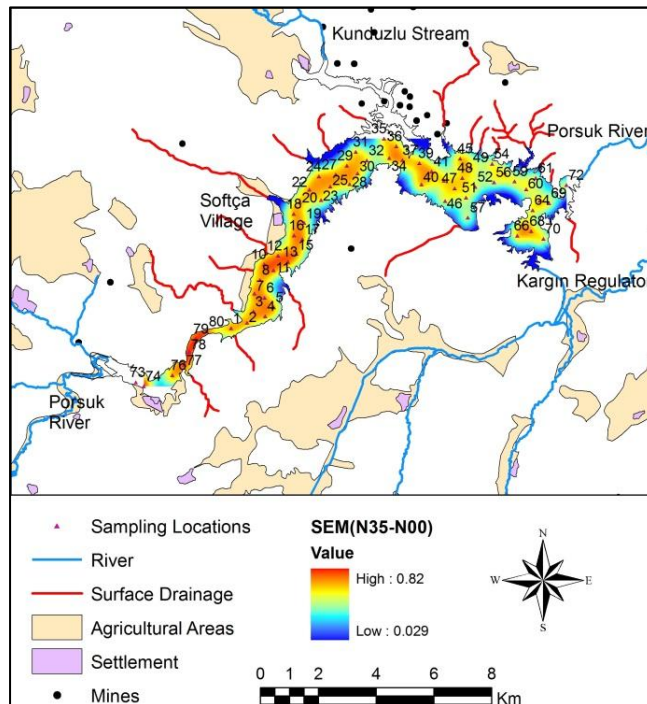


Figure G. 31 Difference of SEMs of DO for N35 and N00

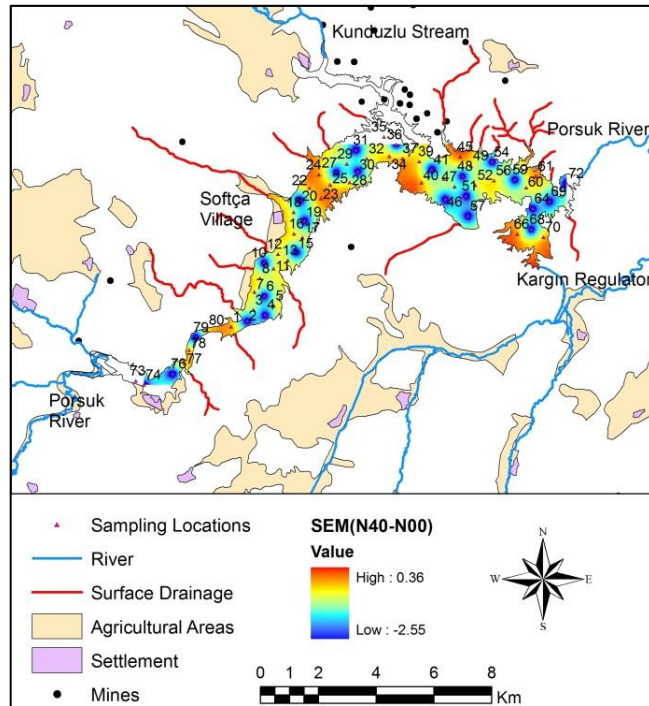


Figure G. 32 Difference of SEMs of DO for N40 and N00

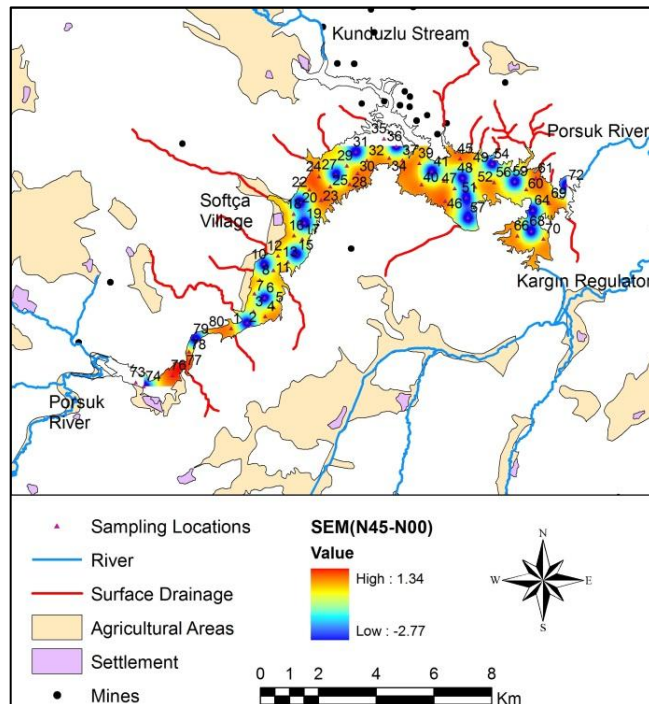


Figure G. 33 Difference of SEMs of DO for N45 and N00

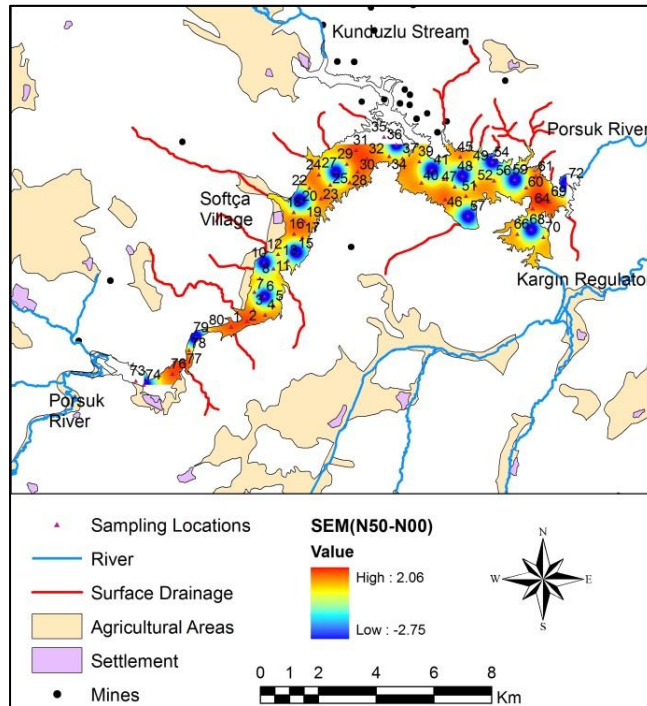


Figure G. 34 Difference of SEMs of DO for N50 and N00

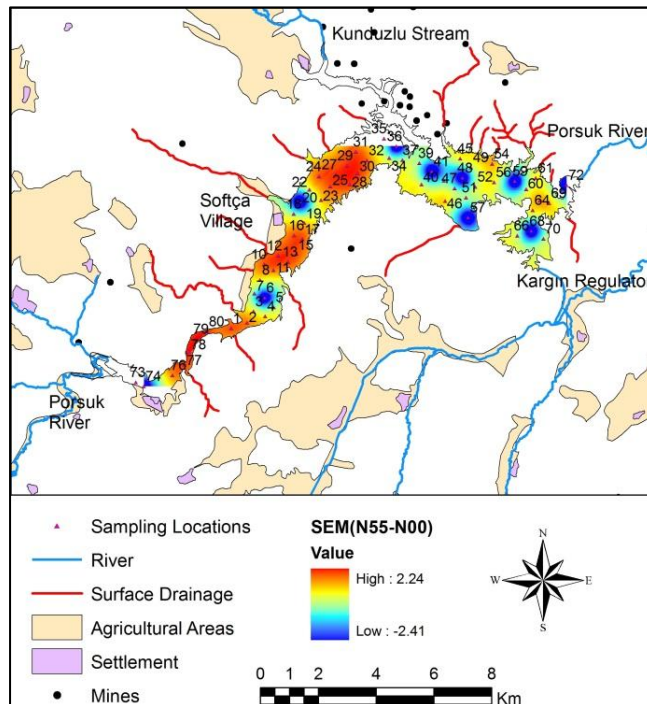


Figure G. 35 Difference of SEMs of DO for N55 and N00

APPENDIX H

OK AND ERROR MAPS OF SC

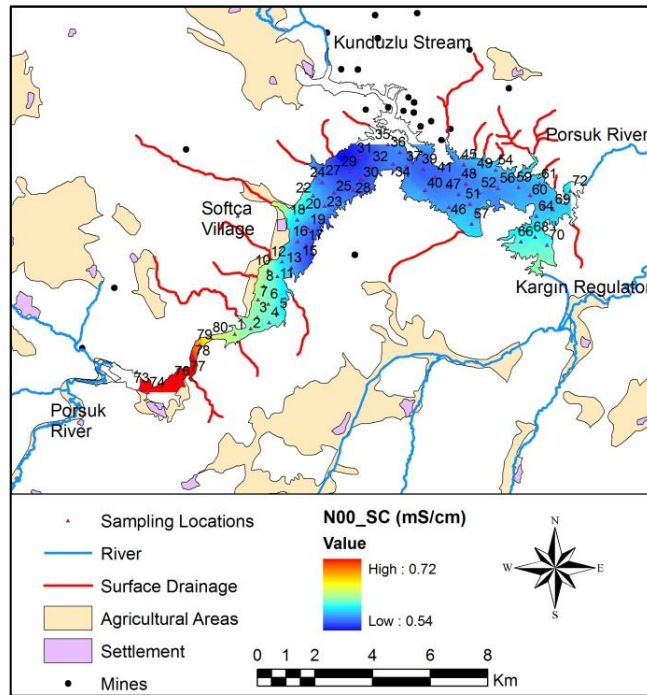


Figure H. 1 OK of SC using N00 data set

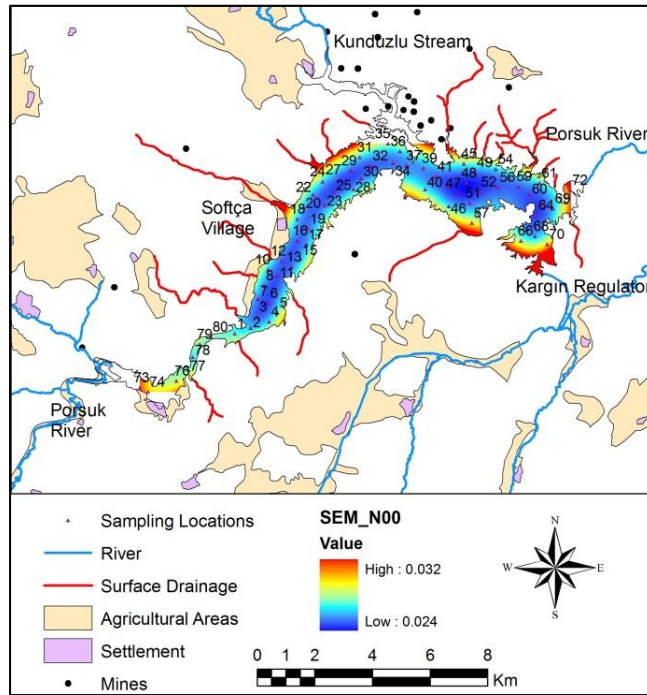


Figure H. 2 SEM of SC using N00 data set

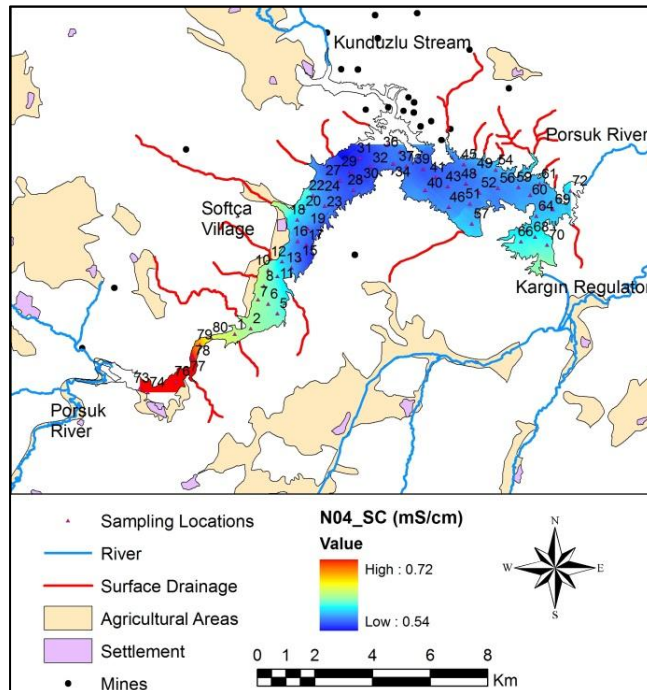


Figure H. 3 OK of SC using N04 data set

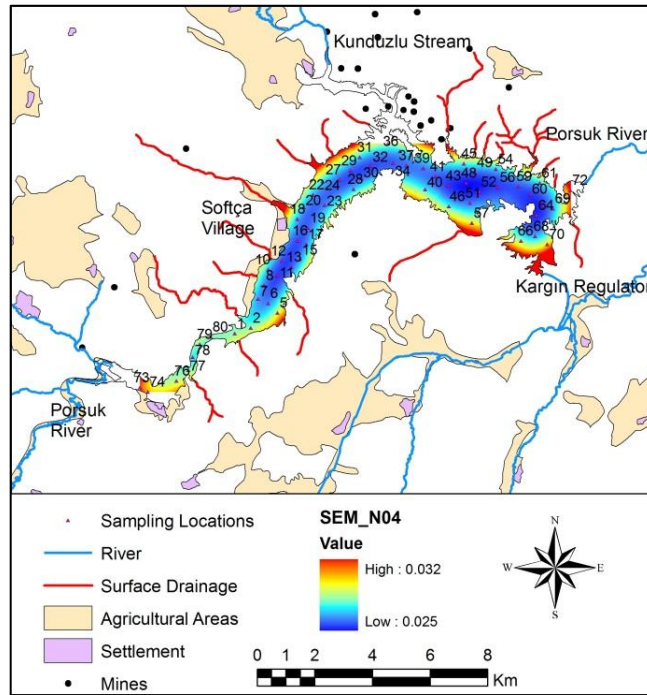


Figure H. 4 SEM of SC using N04 data set

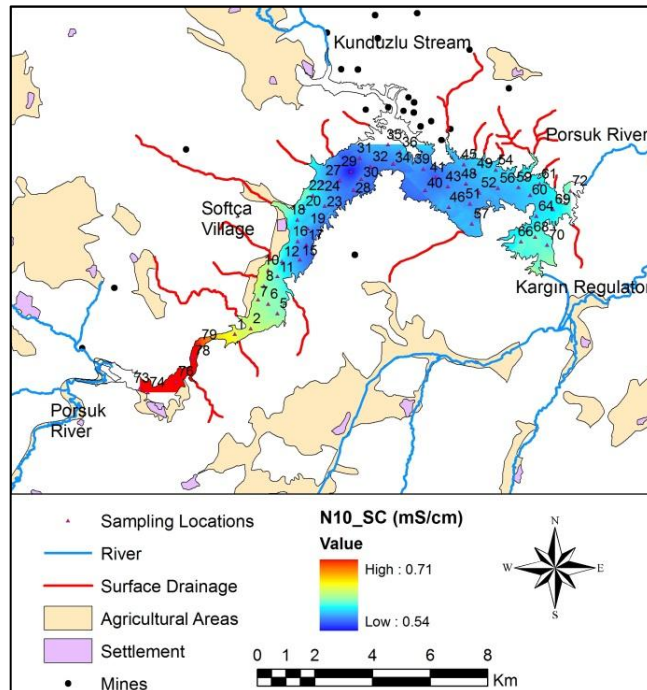


Figure H. 5 OK of SC using N10 data set

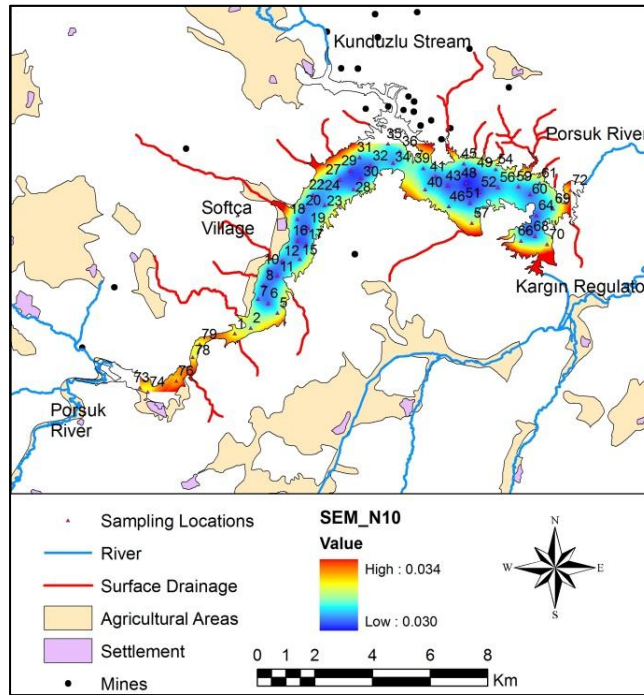


Figure H. 6 SEM of SC using N10 data set

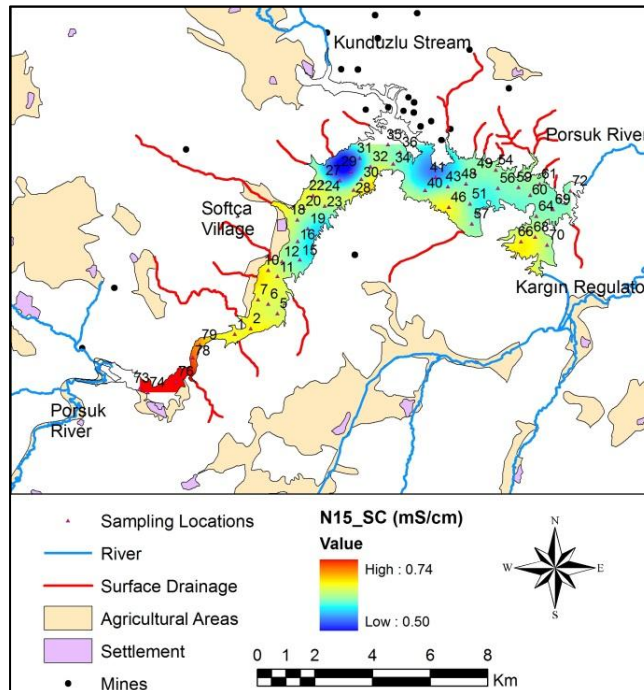


Figure H. 7 OK of SC using N15 data set

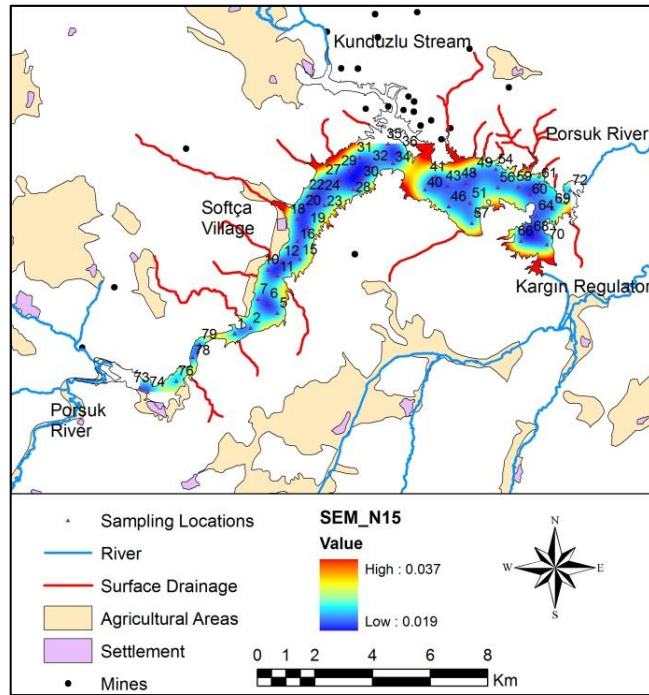


Figure H. 8 SEM of SC using N15 data set

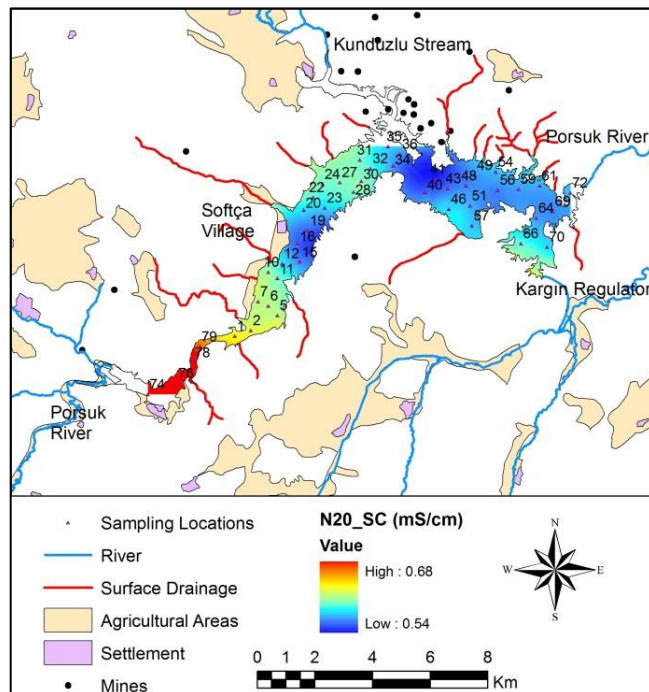


Figure H. 9 OK of SC using N20 data set

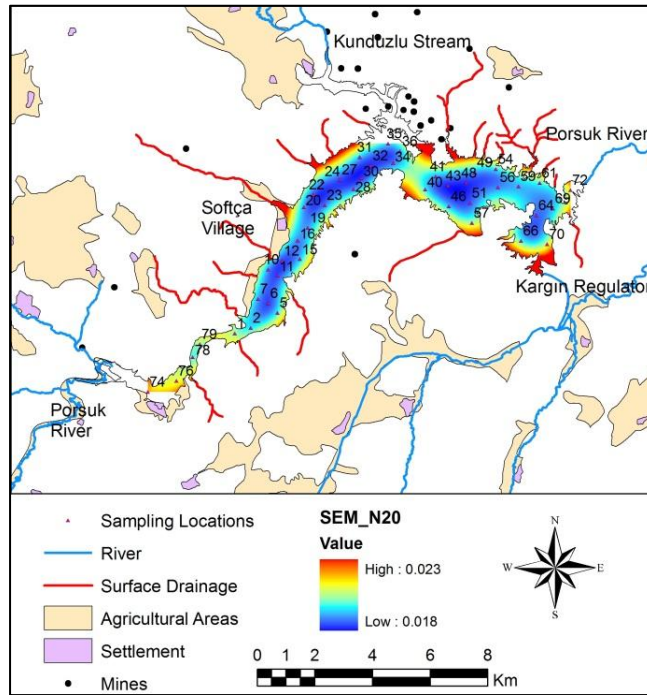


Figure H. 10 SEM of SC using N20 data set

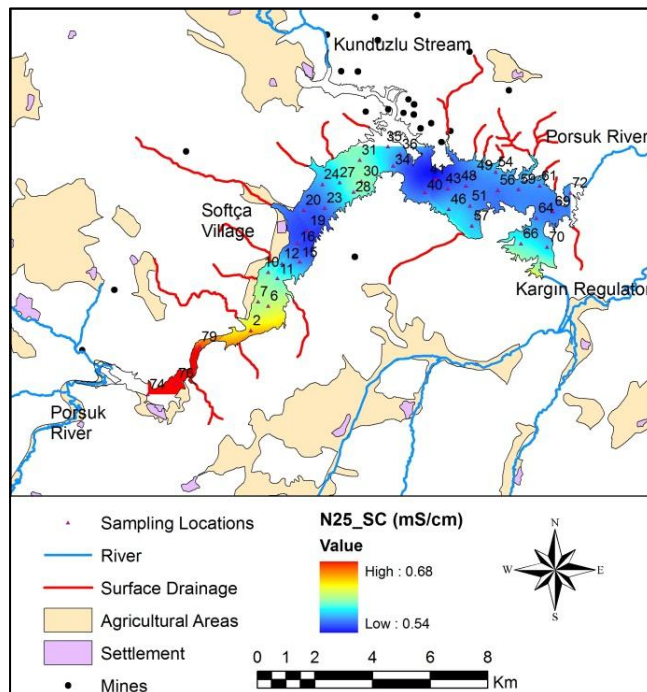


Figure H. 11 OK of SC using N25 data set

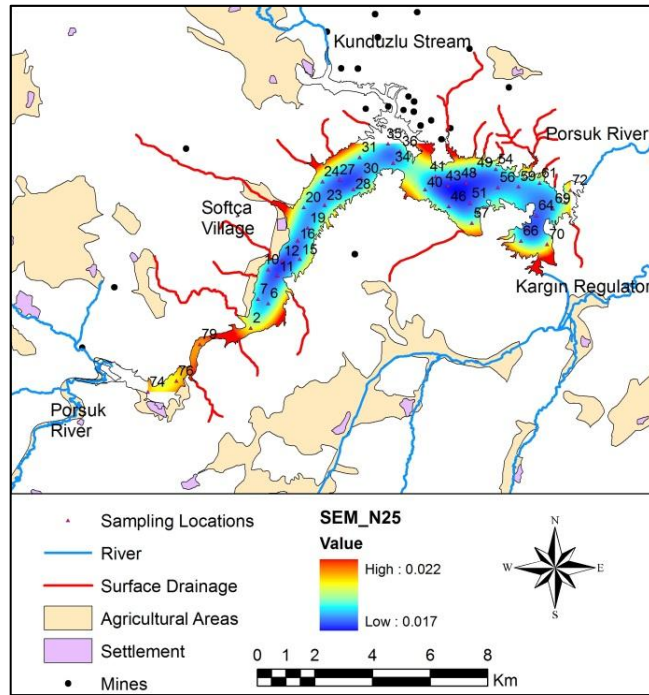


Figure H. 12 SEM of SC using N25 data set

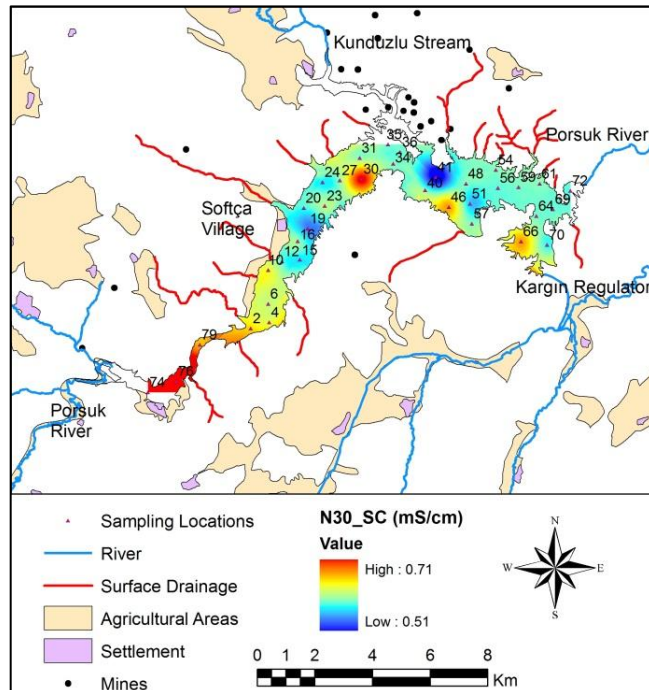


Figure H. 13 OK of SC using N30 data set

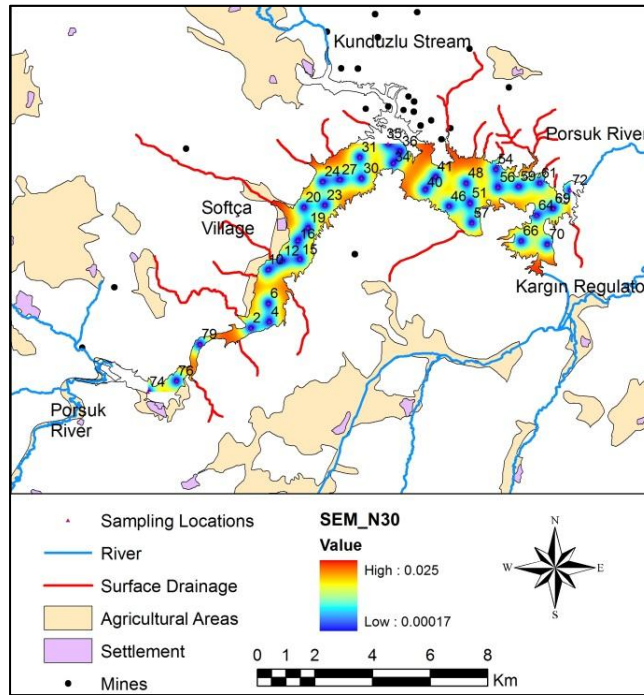


Figure H. 14 SEM of SC using N30 data set

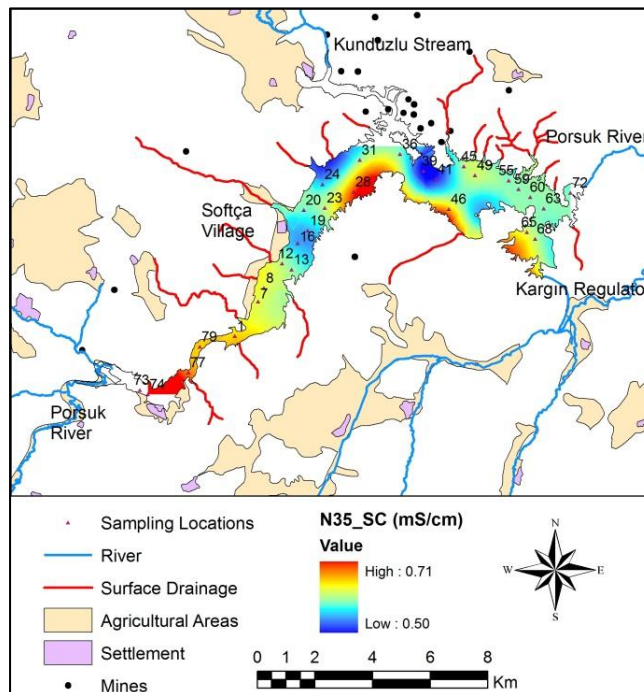


Figure H. 15 OK of SC using N35 data set

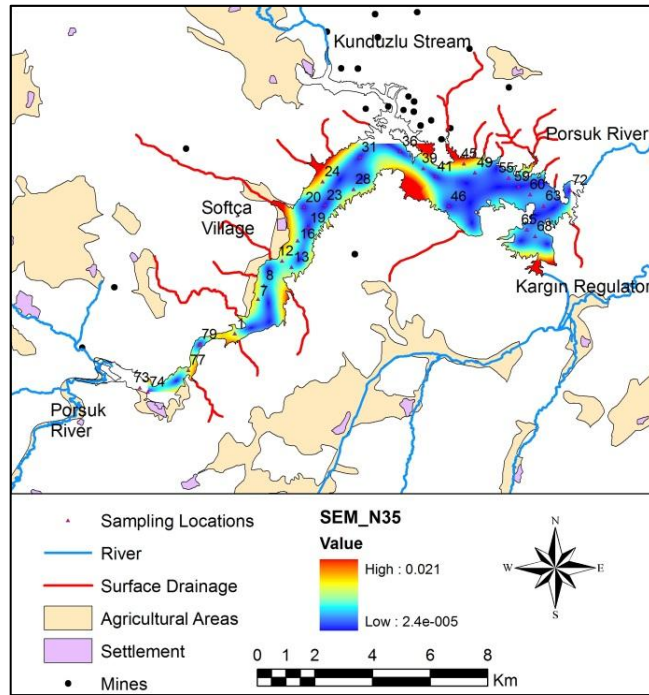


Figure H. 16 SEM of SC using N35 data set

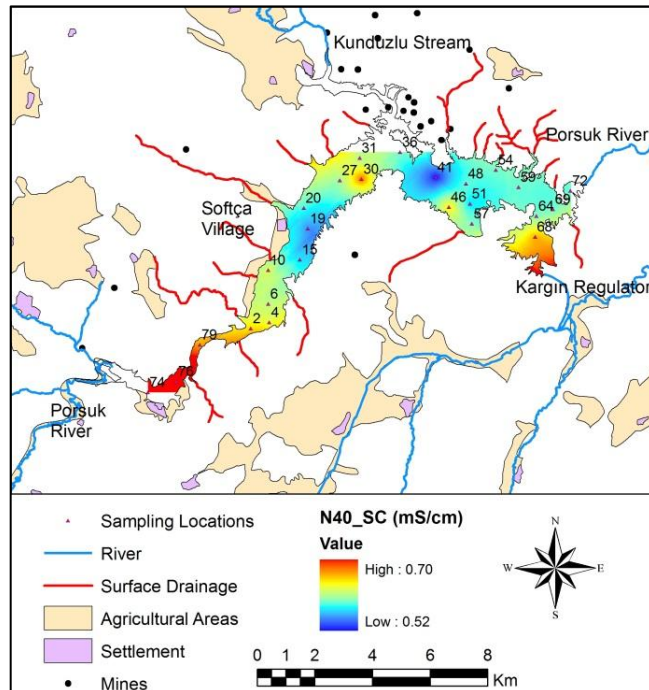


Figure H. 17 OK of SC using N40 data set

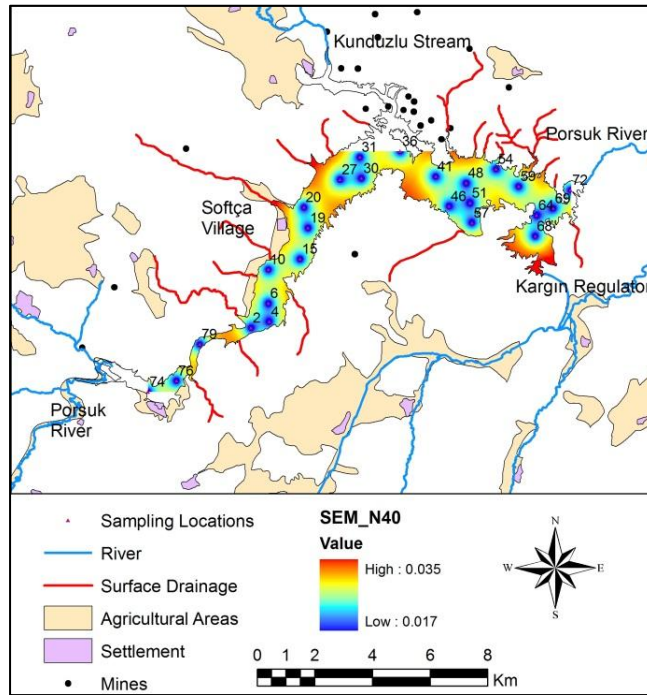


Figure H. 18 SEM of SC using N40 data set

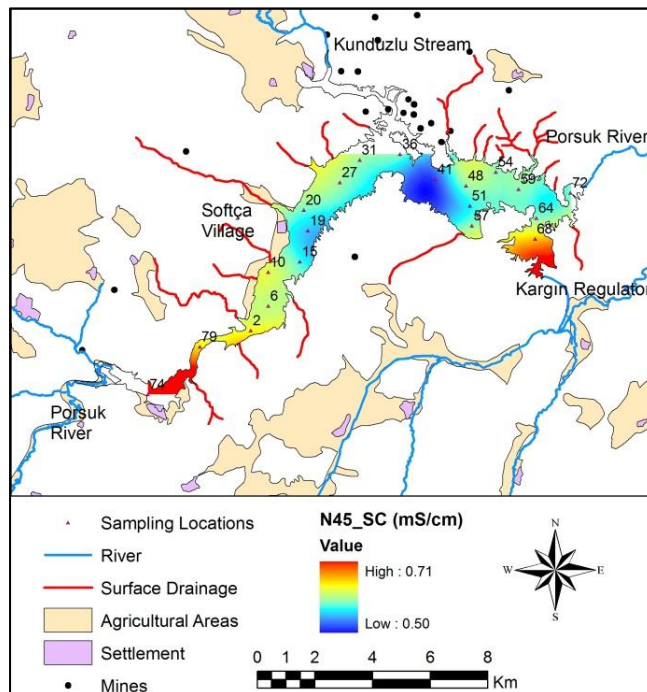


Figure H. 19 OK of SC using N45 data set

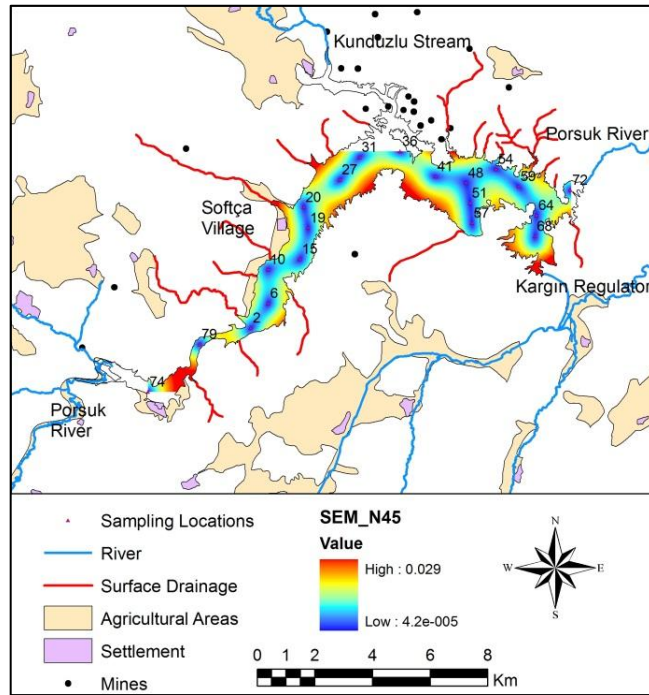


Figure H. 20 SEM of SC using N45 data set

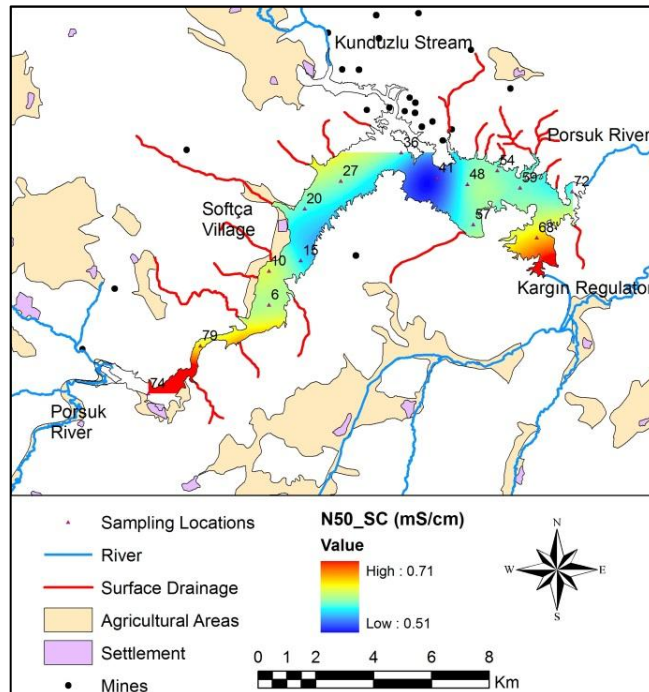


Figure H. 21 OK of SC using N50 data set

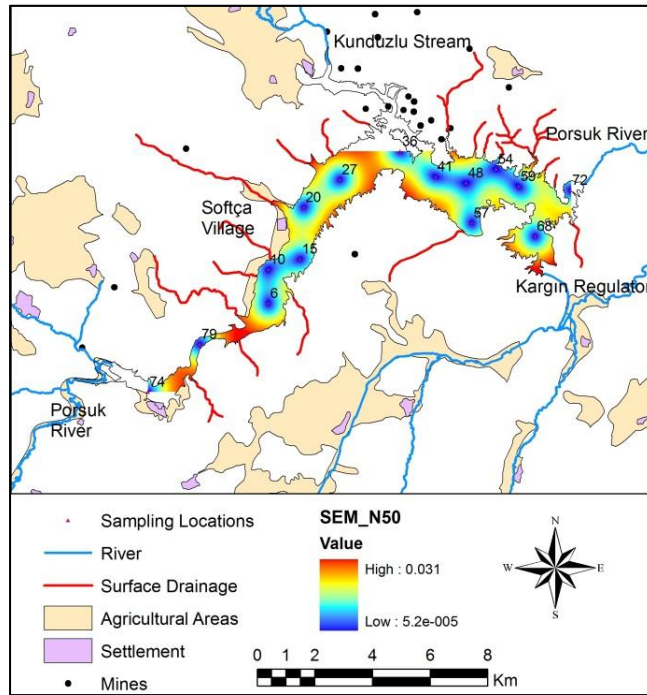


Figure H. 22 SEM of SC using N50 data set

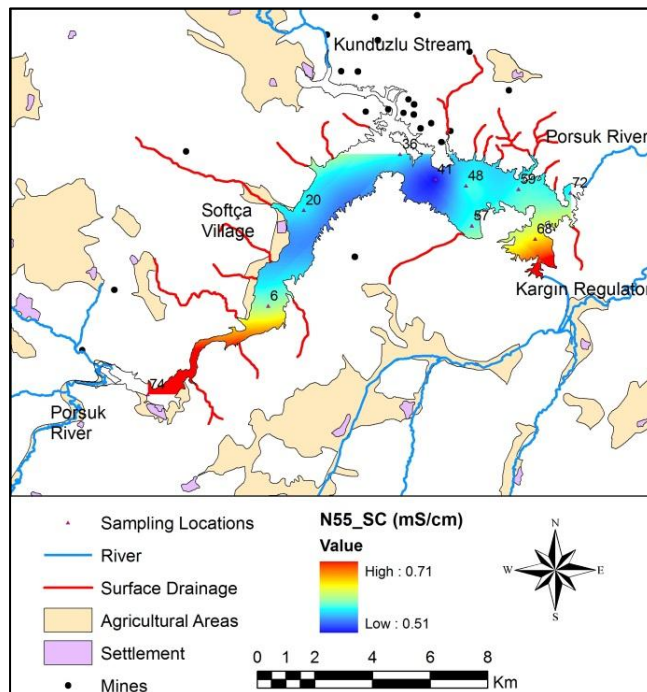


Figure H. 23 OK of SC using N55 data set

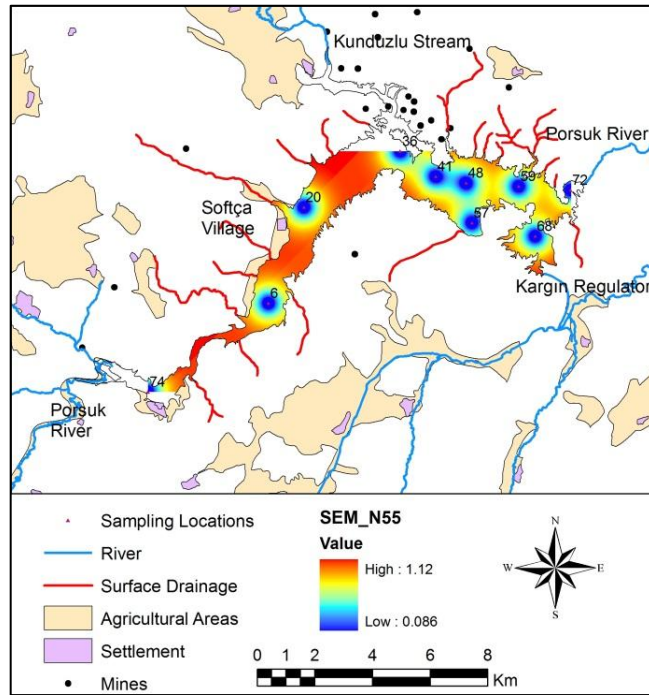


Figure H. 24 SEM of SC using N55 data set

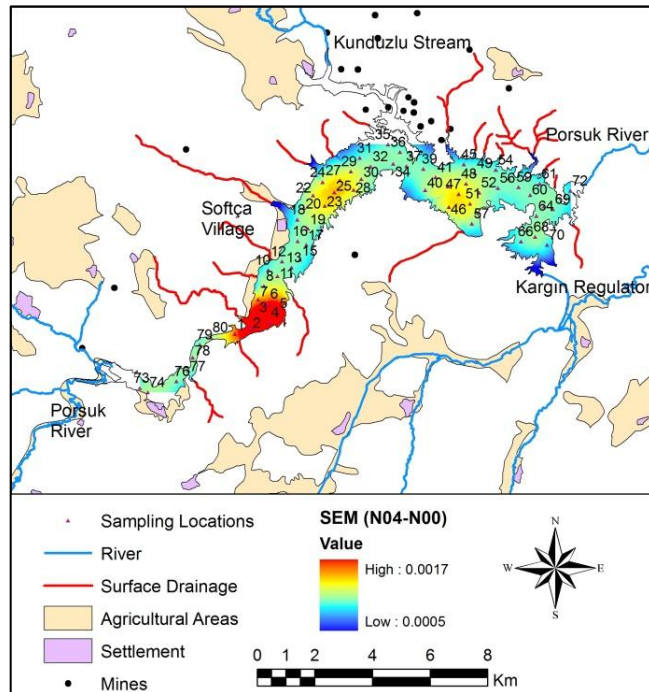


Figure H. 25 Difference of SEMs of SC for N04 and N00

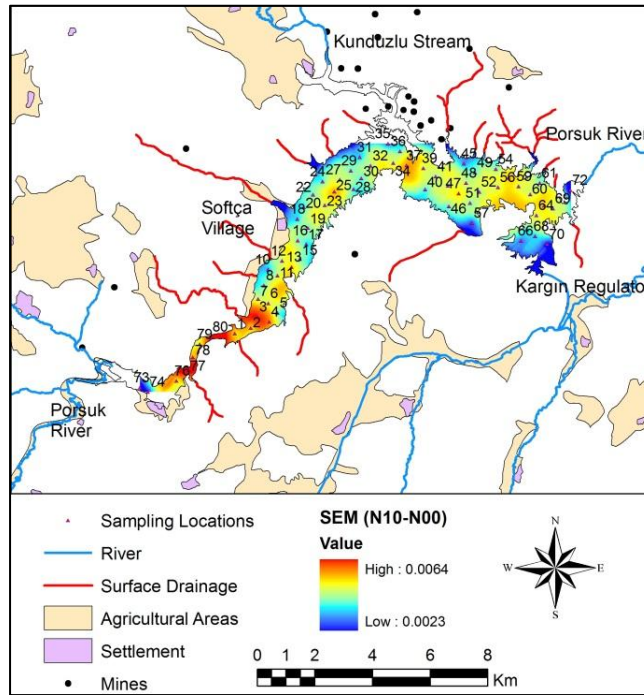


Figure H. 26 Difference of SEMs of SC for N10 and N00

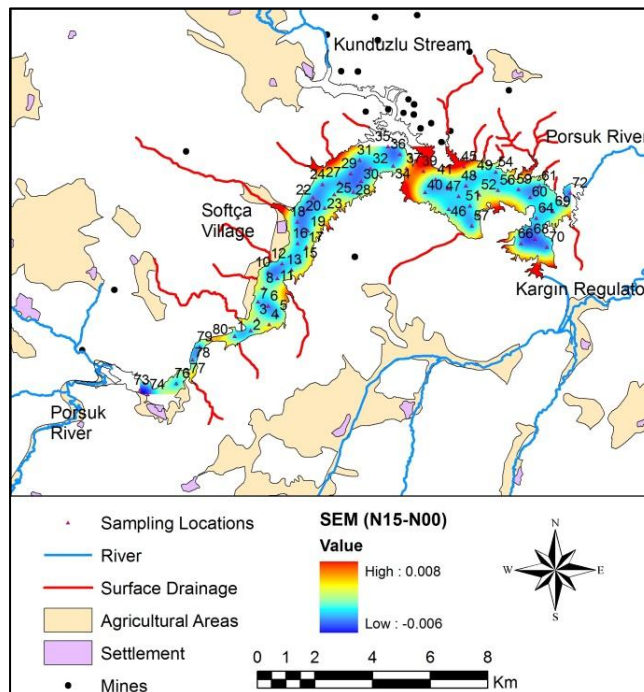


Figure H. 27 Difference of SEMs of SC for N15 and N00

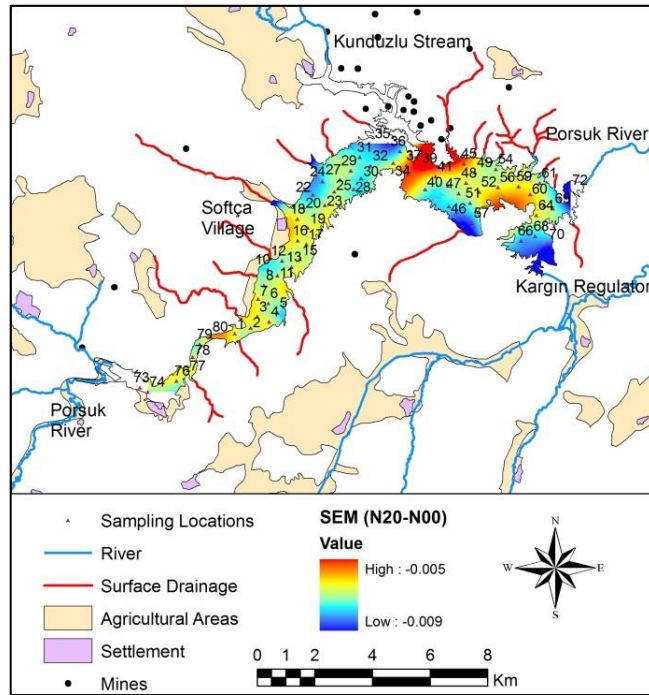


Figure H. 28 Difference of SEMs of SC for N20 and N00

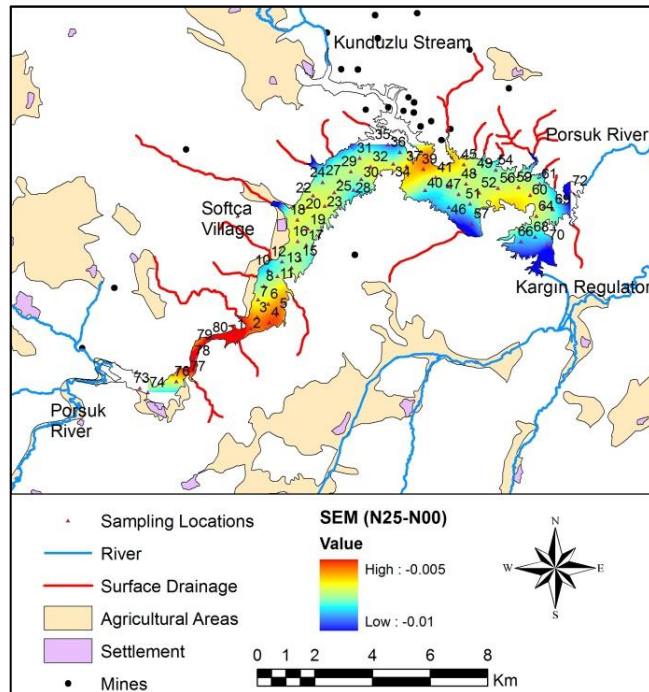


Figure H. 29 Difference of SEMs of SC for N25 and N00

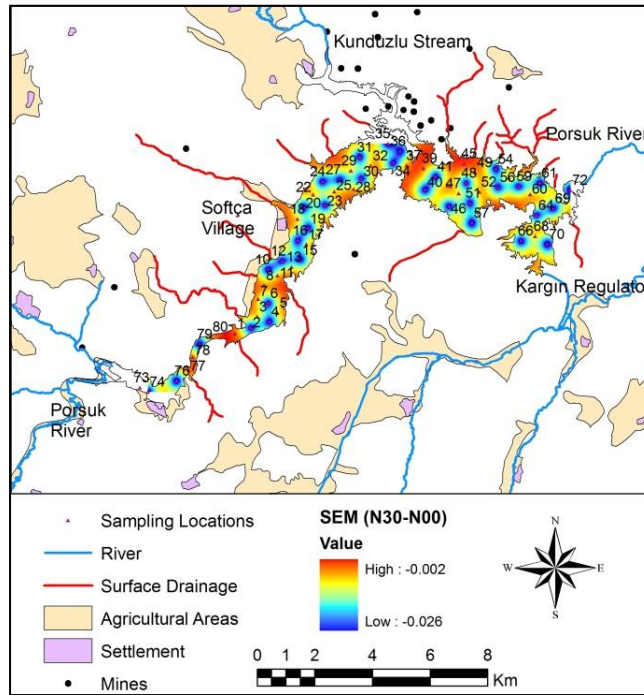


Figure H. 30 Difference of SEMs of SC for N30 and N00

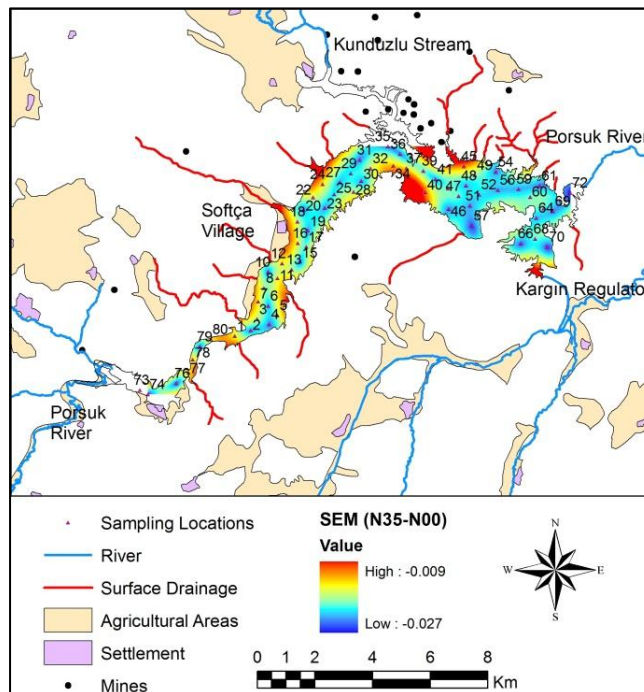


Figure H. 31 Difference of SEMs of SC for N35 and N00

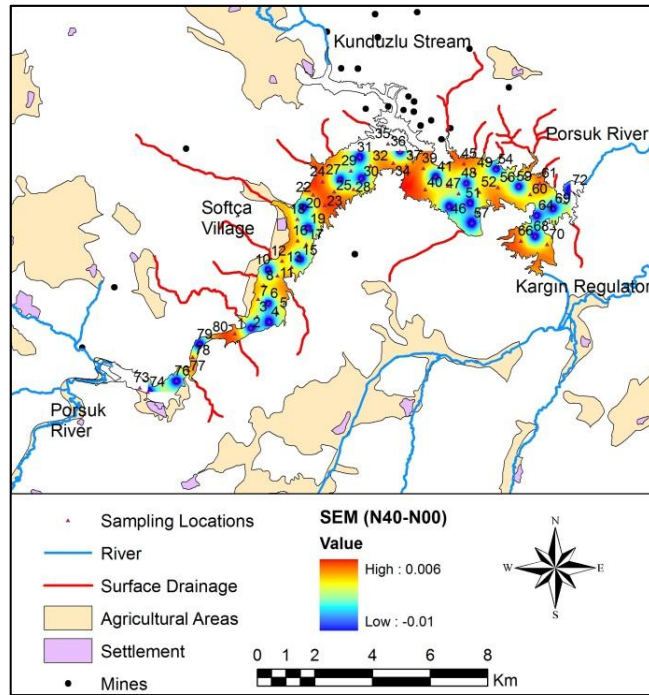


Figure H. 32 Difference of SEMs of SC for N40 and N00

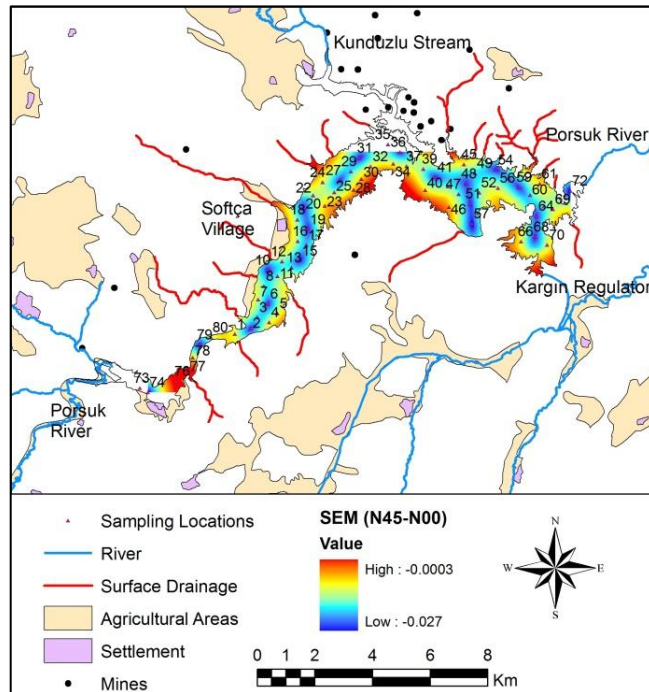


Figure H. 33 Difference of SEMs of SC for N45 and N00

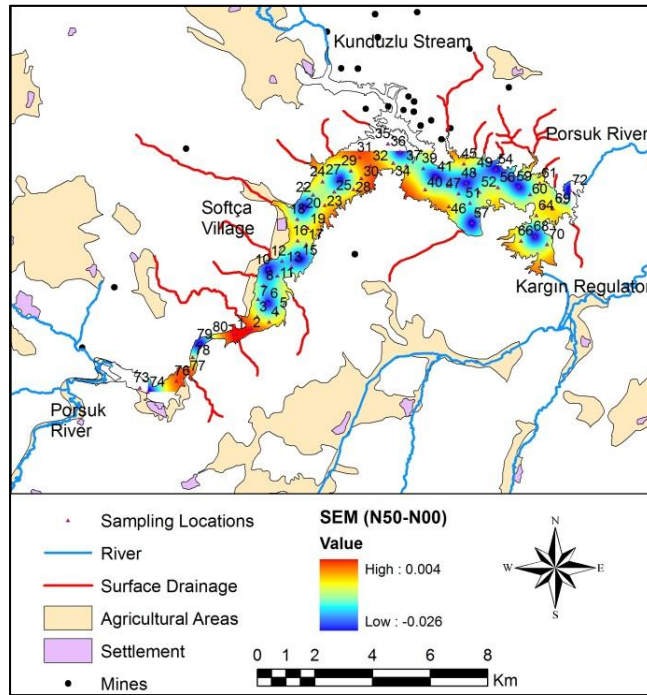


Figure H. 34 Difference of SEMs of SC for N50 and N00

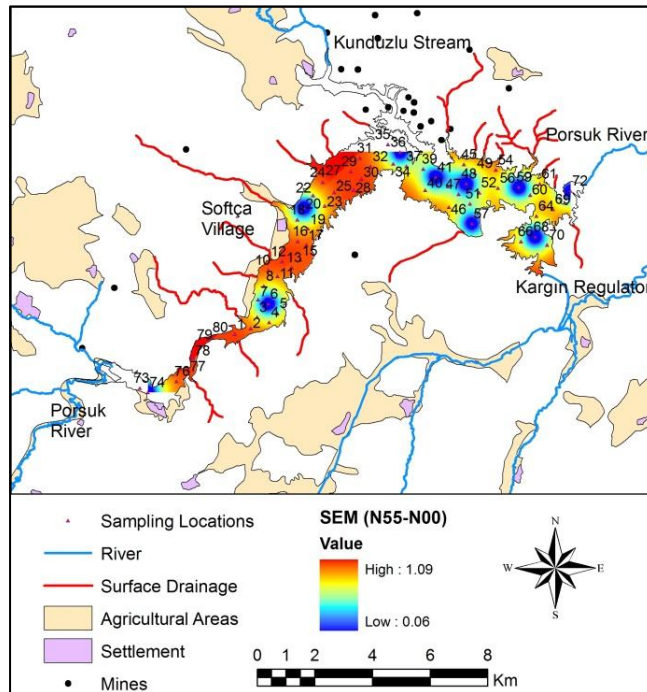


Figure H. 35 Difference of SEMs of SC for N55 and N00

APPENDIX I

MEASURED VS. PREDICTED GRAPHS

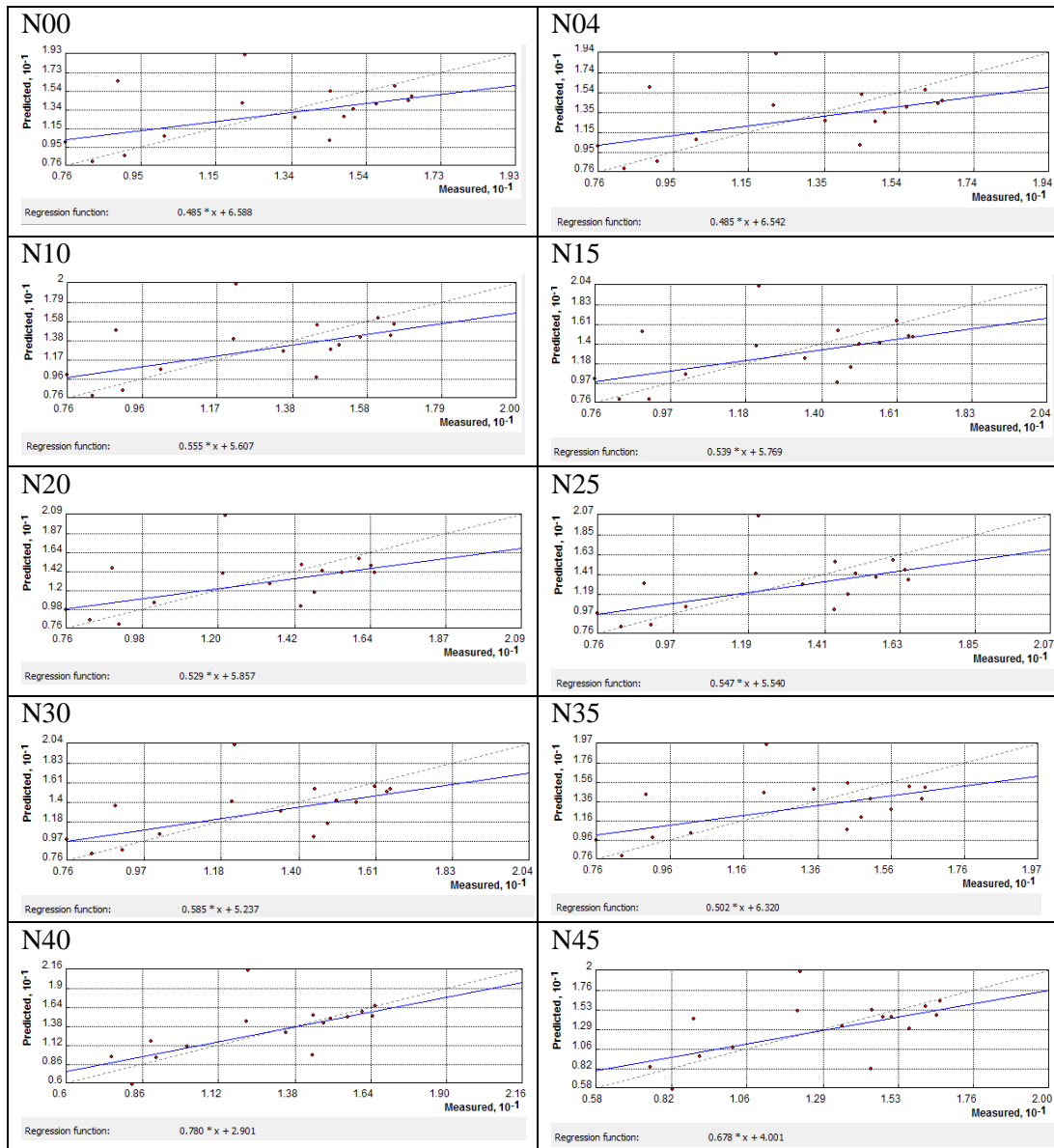


Figure I. 1 Measured DO vs. Predicted DO graph obtained from OK

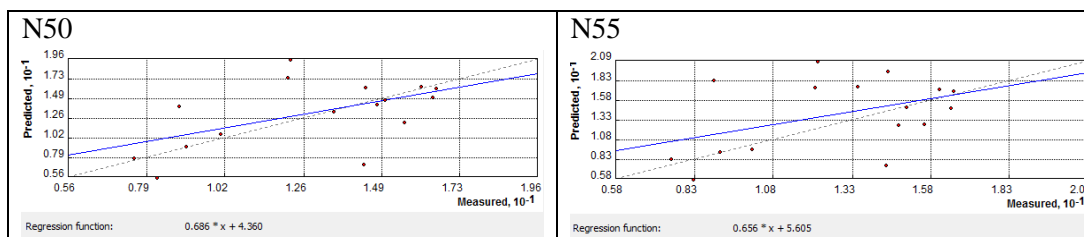


Figure I. 1 Measured DO vs. Predicted DO graph obtained from OK (Continued)

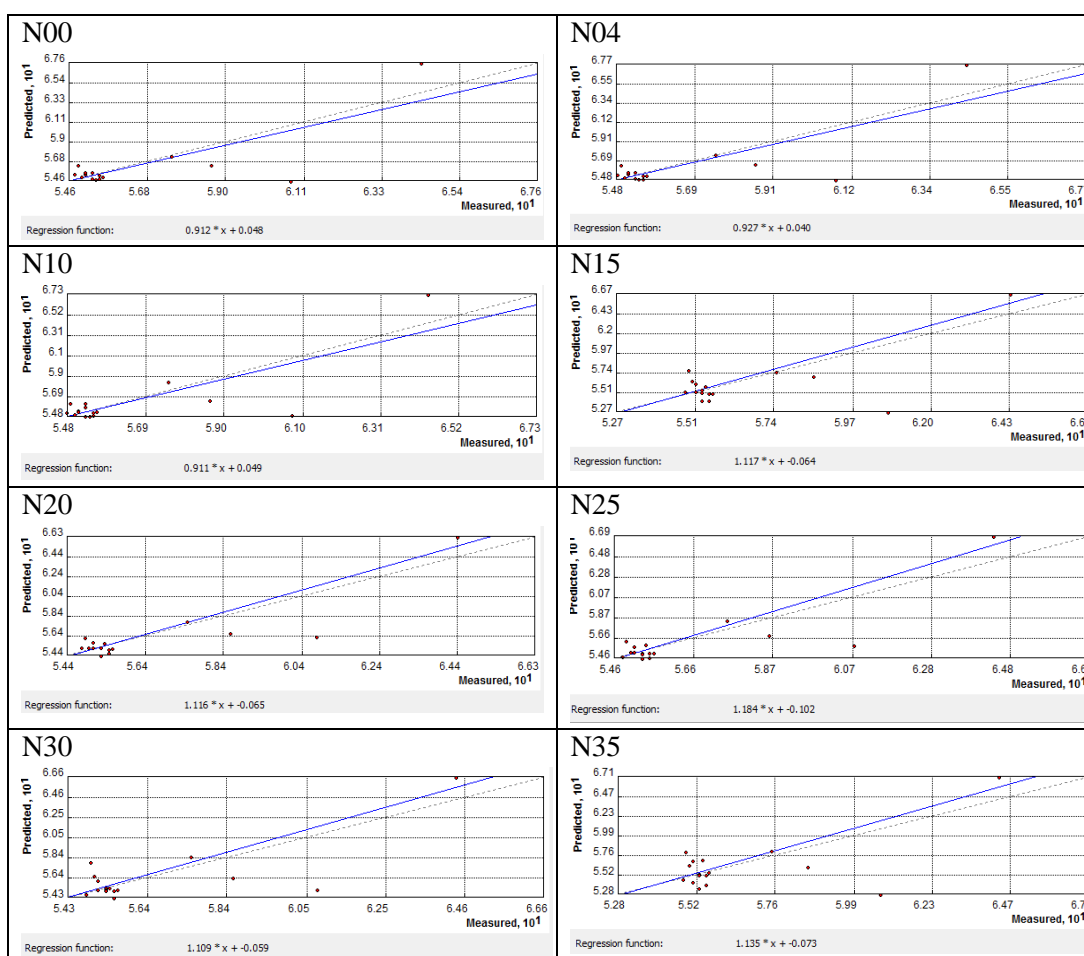


Figure I. 2 Measured SC vs. Predicted SC graph obtained from OK

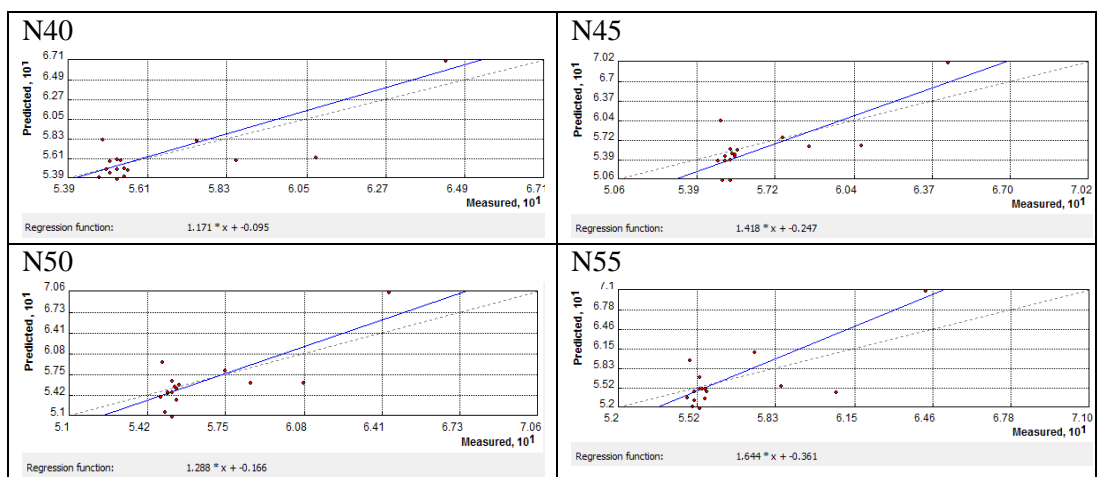


Figure I. 2 Measured SC vs. Predicted SC graph obtained from OK (Continued)

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Yenilmez, Firdes

Nationality: Turkish (T.C.)

Date and Place of Birth: 18 November 1980, Germany

email: firdesyenilmez@hotmail.com



1. Education

Degree	Area	University	Year
M.Sc.	Environmental Engineering	Middle East Technical University	2007
B.Sc.	Environmental Engineering	Akdeniz University	2002

2. Academic Experience

Title	Area	University	Year
Res. Asst.	Environmental Engineering	Middle East Technical University	2004-2014
Visiting Scholar	Civil and Environmental Engineering	Virginia Polytechnic Institute and State University	2012

3. Projects

	Title	Organization	Position	Duration
1	Determination of Best Representative Sampling Locations in Porsuk Dam Reservoir	METU Research Fund - No: BAP-08-11-DPT-2002K120510	Researcher	2010-2012
2	Protection and Special Rule Declaration Project For Porsuk Dam Reservoir As A Drinking and Potable Water	Ministry of Forestry and Water Affairs	Consultant	2009-2012

3	Preparation of Mine Closure Plan at an abandoned coal mine site (Ovacik) in Turkey with the aid of GIS and RS	METU Research Fund - No:BAP-07.02.2009.07	Researcher	2009-2012
4	TMDL development for Uluabat Lake Watershed	TUBITAK- Project No:103Y122	Researcher	2003-2005
5	Modeling of the water quality in Uluabat Lake	METU Research Fund - No:BAP-08-11-DPT-2002K120510	Researcher	2005-2007

4. Publications

4.1. Article, review, technical note, letter to the editor, book critic and other publications in journals indexed by SCI:

Yenilmez, F., Düzgün, S., Aksoy, A (under review). An evaluation of potential sampling locations in a reservoir with emphasis on conserved spatial correlation structure.

Golge, M., **Yenilmez, F.**, Aksoy, A. (2013). Development of pollution indices for the middle section of the Lower Seyhan Basin (Turkey). *Ecological Indicators*, 29, 6-17.

Yenilmez, F., Aksoy, A (2013). Comparison of Phosphorus Reduction Alternatives in Control of Nutrient Concentrations in Lake Uluabat (Bursa, Turkey): Partial Versus Full Sediment Dredging. *Limnologica-Ecology and Management of Inland Waters*, 43 (1), 1-9.

Yenilmez, F., Kuter, N., Emil, M. K., Aksoy, A. (2011). Evaluation of pollution levels at an abandoned coal mine site in Turkey with the aid of GIS. *International Journal of Coal Geology*, 86 (1), 12-19.

Yenilmez, F., Keskin, F., Aksoy, A. (2011). Water quality trend analysis in Eymir Lake, Ankara. *Physics and Chemistry of the Earth, Parts A/B/C*, 36 (5-6), 135-140.

Kuter, N., **Yenilmez, F.**, Kuter, S. (2011). Forest Fire Risk Mapping by Kernel Density Estimation. *Croatian Journal of Forest Engineering*, 32 (2), 599-610.

4.2. Article, review, technical note, letter to the editor, book critic and other publications in journals indexed by non-SCI:

Yenilmez, F., Keskin F., Aksoy, A. (2010). Meteorolojik Koşulların Mogan Gölü Su Kalitesine Etkileri. *BİBAD (Research Journal of Biology Sciences)*, 3 (2), 33-38.

Keskin F., **Yenilmez, F.**, Çolak M., Yavuzer İ., Düzgün H. Ş. (2011). Analysis of Traffic Incidents in METU Campus. *Procedia Social and Behavioral Sciences*, 19, 61-70.

4.3. Presentations in international scientific meetings and published in proceedings:

Kuter, N., **Yenilmez, F.**, Kuter, S. Evaluation of Forest Fires with GIS: Case Study at METU Campus, Ecology and Forests for Public Health, 18-20 September 2009, Soria Moria Hotel and Conference Centre, Oslo, Norway, p.70.

Yenilmez, F., Keskin, F., Aksoy, A. Water Quality Trend Analysis in Lake Eymir, Ankara, Regional Process of The 5th World Water Forum, Regional Meeting on Water in the Mediterranean Basin, 09-11 October 2008, Near East University, Lefkosa, TRNC, p.65.

Muhammetoğlu, A., **Yenilmez, F.**, Muhammetoğlu, H. Fate of agricultural nitrogen in the soil and groundwater of Kumluca-Turkey, 12th International Symposium on Environmental Pollution and its Impact on Life in the Mediterranean Region, Antalya, Turkey, October 4-8, 2003, p.29.

4.4. Presentations in international scientific meetings:

Aksoy, A., **Yenilmez, F.**, Düzgün, S. Determination of geostatistically representative sampling locations in Porsuk Dam Reservoir (Turkey), AGU Fall Meeting, 9-13 December 2013, San Francisco, USA.

Yenilmez, F., Aksoy, A. Evaluation of Water Quality Within the Watershed of Porsuk Dam Reservoir, 17th International Symposium on Environmental Pollution and its Impact on Life in the Mediterranean Region, September 28 – October 1, 2013, Istanbul, Turkey.

Yenilmez, F., Aksoy, A. Assesment of Long Term Pollution Loads in Porsuk River, 17th International Symposium on Environmental Pollution and its Impact on Life in the Mediterranean Region, September 28 – October 1, 2013, Istanbul, Turkey.

Yenilmez, F., Aksoy, A. Determination of the vulnerable areas for diffuse pollution transport within the watershed of the Porsuk Dam Reservoir using ArcGIS, IPWE 6th International Perspective on Water Resources & The Environment, January 7-9, 2013, Izmir, Turkey.

Yenilmez F., Keskin F., Aksoy A. Spatial and Temporal Analysis of Water Quality Parameters In Mogan Lake, GEOMED 2010 (The 2nd International Geography Symposium), Kemer-Antalya-Turkey, June 2-5, 2010.

Keskin, F., **Yenilmez, F.**, Çolak, M., Yavuzer, İ., Düzgün, H.Ş. Analyzing of Traffic Incidents In METU, GEOMED 2010 (The 2nd International Geography Symposium), Kemer-Antalya-Turkey, June 2-5, 2010.

Yenilmez, F., Keskin, F., Aksoy, A. Impact of Meteorological Conditions on the Water Quality in Eymir Lake, Turkey, 5th World Water Forum, 16-22 March 2009, İstanbul, Turkey.

4.5. Presentations in national scientific meetings and published in proceedings:

Yenilmez, F., Keskin, F., Aksoy, A. Mogan Gölü su kalite parametrelerinin mekansal ve zamansal değerlendirilmesi, JEOUZAL 2010 Sempozyumu, 4-5 Kasım 2010, MTA, Ankara, p.29.

Yenilmez, F., Keskin, F., Aksoy, A. Çevre Koşullarının Mogan Gölü Su Kalitesine Etkileri, IX. Ulusal Ekoloji ve Çevre Kongresi, 7-10 Ekim 2009, Ürgüp, Nevşehir, p.21.

Yenilmez, F., Keskin, F., Aksoy, A. Çevre Koşullarının Eymir Gölü Su Kalitesine Etkileri, 8. Ulusal Çevre Mühendisliği Kongresi, 12-14 Kasım 2009, Antalya.

Yenilmez, F., Aksoy, A. Uluabat Gölü Su Kalitesinin WASP 7.2 Modeli Kullanılarak Değerlendirilmesi, 7.Ulusal Çevre Mühendisliği Kongresi, 24-27 Ekim 2007, İzmir, E.Ü. Atatürk Kültür Merkezi, pp.56-62.

5. Prizes and Awards (Scholarships, scientific, encouragement awards)

- TUBITAK BİDEB 2214- International Research Scholarship, 2012.
- TUBITAK BİDEB 2211- National Ph.D. Scholarship, 2007-2012.
- Academic Performance Award, METU, Graduate School of Natural and Applied Sciences, 2009.
- 1st Highest Degree in Graduation, Akdeniz University, Department of Environmental Engineering, 2002.

6. National or International Activities (Congress, workshop organization etc.)

Organizing Committee Support Staff, “Facing Sludge Diversities: Challenges, Risks, and Opportunities”, IWA International Water Association, Middle East Technical University, Dokuz Eylul University, 28-30 March 2007, Antalya, Turkey.

7. Work Experience

Position	Institution/Organization/Firm	Year
Student Assistant	Akdeniz University, Department of Environmental Engineering	September 2001-June 2003
Engineering Intern	ISKI Water Quality Control Directorate Istanbul, Turkey	July-August 2001
Engineering Intern	Erbay Environmental Engineering. Inc., Antalya, Turkey	July-August 1999

8. Research Interests

Water Quality Modeling and Management
Spatial Data Analysis and GIS Applications
Watershed Management