

DEVELOPING REGIONAL FLOW DURATION CURVES AND  
EVALUATING THE PERFORMANCES IN THE UNGAUGED BASINS

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

YAPRAK KOCATEPE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF MASTER OF SCIENCE  
IN  
CIVIL ENGINEERING

FEBRUARY 2011

Approval of the thesis:

**DEVELOPING REGIONAL FLOW DURATION CURVES AND  
EVALUATING THE PERFORMANCES IN THE UNGAUGED  
BASINS**

submitted by **YAPRAK KOCATEPE** in partial fulfillment of the requirements for the degree of Master in Science in **Civil Engineering Department, Middle East Technical University** by,

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

\_\_\_\_\_

Prof. Dr. Güney Özcebe  
Head of Department, **Civil Engineering**

\_\_\_\_\_

Assoc. Prof. Dr. Zuhall Akyürek  
Supervisor, **Civil Engineering Department, METU**

\_\_\_\_\_

**Examining Committee Members:**

Prof.Dr. Melih Yanmaz  
Civil Engineering Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Zuhall Akyürek  
Civil Engineering Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Burcu Altan Sakarya  
Civil Engineering Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Nurunnisa Usul  
Civil Engineering Dept., METU

\_\_\_\_\_

Dr. Koray Yılmaz  
Geological Engineering Dept., METU

\_\_\_\_\_

**Date:** 28.01.2011

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

Name, Last name : Yaprak Kocatepe

Signature :

# ABSTRACT

## DEVELOPING REGIONAL FLOW DURATION CURVES AND EVALUATING THE PERFORMANCES IN UNGAUGED BASIN

Kocatepe, Yaprak  
M.Sc., Department of Civil Engineering  
Supervisor: Assoc. Prof. Dr. Sevda Zuhall Aky rek

February 2011, 282 pages

A flow duration curve (FDC) defines the relation between the flow amount of any time (daily, yearly, or another time) and its frequency. Moreover, FDCs are used in many water resources projects. However, the ungauged basins or limited amount of gauging in a basin is a common problem. Therefore, regional FDCs are needed to be developed in ungauged basins. Oltu basin has been chosen as the study area, which is located in the north-eastern part of Turkey in  oruh Basin. Two parametric approaches and a statistical approach have been applied to develop regional flow duration curves (FDCs) in Oltu Basin. Parametric approaches cover two different models, namely Model Kocatepe, which is a five parameter model depending on the regression analysis between discharge having certain probability of occurrences and geomorphologic and climatic factors; Model Quimpo, which is a two parameter model proposed by Quimpo. Lognormal distribution has been used in the statistical approach. Several performance indices have been evaluated to decide on if the model dependable or not. As a result of these analysis, it is concluded that, Model Quimpo gives good results in small basins, whereas, Model Kocatepe is effective in large areas. Statistical approach is not an appropriate method to use while regionalizing FDCs in Oltu basin.

The analysis performed for short-term duration has revealed that 5-years record lengths of discharges are enough to develop a dependable FDC compared to regional FDC. The validation results and the performance indices are presented with the analysis results.

**Keywords:** Regional Flow Duration Curves, performance indices, ungauged basins

# ÖZ

## **BÖLGESEL DEBİ SÜREKLİLİK EĞRİLERİNİN GELİŞTİRİLMESİ VE ÖLÇÜMÜ OLMAYAN HAVZALARDA PERFORMANSLARININ DEĞERLENDİRİLMESİ**

Kocatepe, Yaprak  
Yüksek lisans, İnşaat Mühendisliği Bölümü  
Tez yöneticisi: Doç.Dr. Sevda Zuhal Akyürek

Şubat 2011, 282 sayfa

Debi süreklilik eğrileri (DSEler) bir akarsudaki herhangi bir zaman aralığındaki (günlük, yıllık ya da başka bir zaman için) akım miktarlarının görülme sıklığıyla ilişkisini tanımlar. Buna ek olarak, DSEler su kaynakları projelerinde sıklıkla kullanılmaktadır. Fakat ölçümü olmayan havzalar ya da kısıtlı ölçümü olan havzalar sıklıkla karşılaşılan bir problemdir. Bu bağlamda ölçümü olmayan havzalarda debi süreklilik eğrilerinin geliştirilmesi gerekmektedir. Türkiye'nin kuzeydoğusunda Çoruh havzasında yer alan Oltu havzası çalışma alanı olarak belirlenmiştir. Bölgesel debi süreklilik eğrilerinin geliştirilmesinde iki parametrik yaklaşım ve bir istatistiksel yaklaşım kullanılmıştır. Parametrik yaklaşım iki farklı modeli içermektedir. Bunlardan ilki olan Model Kocatepe, beş parametrelili bir model olup, belirli aşılma olasılıklarına denk gelen akım değerlerinin jeomorfolojik ve iklimsel parametrelerle regresyon analizine dayanmaktadır. İkinci parametrik yaklaşım olan Model Quimpo, Quimpo tarafından önerilen iki parametrelili bir modeldir. İstatistiksel yaklaşımda ise lognormal dağılım kullanılmıştır. Birçok performans indeksi, modelin uygunluğunu doğrulamak için değerlendirilmiştir. Bu analizlerin sonucunda, Model Kocatepe'nin büyük alanlı havzalarda, Model Quimpo'nun ise küçük alanlı havzalarda iyi sonuçlar verdiği gözlemlenmiştir. İstatistiksel yaklaşımın ise bölgesel DSElerin geliştirilmesinde uygun bir yaklaşım olmadığı gözlemlenmiştir.

Kısa dönem sürekliliği için yapılan analizlerde 5-yıllık veri uzunluğunun, güvenilir bölgesel DSElerin geliştirilmesi için yeterli olduğu bulunmuştur. Doğrulama sonuçları ve performans indeksleri analiz sonuçlarıyla birlikte sunulmuştur.

**Anahtar Kelimeler:** Bölgesel Debi Süreklilik Eğrileri, performans indeksleri, ölçümü olmayan havzalar

# ACKNOWLEDGEMENTS

I am grateful to my family, especially to my mother, for supporting me for my whole life and giving me their endless love. I am proud of to be your daughter.

I owe special thanks to my supervisor Assoc. Prof. Dr. Zuhale Aky rek for letting me have the opportunity to work with her in this study.

I am also quite thankful to Serdar S rer for being a co-supervisor and a friend to me. Without his guidance and help, I would be lost in this study.

I would like to thank my lovely friends  a da  Dedeo lu, Er an Erdo an and Mustafa  er i, for their support, patience and endless love. I wish to be with you for rest of my life.

I want to thank my beloved friends, H. Alper  zcan, Kemal Songur, Pelin Asena, and Melis Tutuk for their encouragements, supports, motivations and love. They were like a family to me. Thanks for being with me, not only in the lovely times but also in the hard times.

Finally I owe a great thanks to Emircan  elik for being nearby me during this study, encouraging me, and for his patience and love.



Dedicated to My Mother

# TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZ .....	vi
ACKNOWLEDGEMENTS .....	viii
TABLE OF CONTENTS.....	x
LIST OF TABLES .....	xiii
LIST OF FIGURES .....	xvi
LIST ABBREVIATIONS AND SYMBOLS .....	xix
CHAPTERS	
1 INTRODUCTION .....	1
1.1. Organization Of The Thesis .....	3
2 LITERATURE SURVEY AND METHODOLOGY .....	4
2.1. Flow Duration Curves.....	4
2.1.1.2. Ranking Method.....	5
2.2. Literature Survey .....	6
2.3. Methodology .....	9
2.3.1. Regionalization of FDCs.....	11
2.3.2. Performance Indices.....	12
3 STUDY AREA AND DATA.....	15
3.1. Study Area .....	15
3.2. Data .....	15
3.2.1. Hydrologic Data.....	17
3.2.2. Meteorological Data.....	18
3.2.3. Topographic Data.....	18
3.2.4. Soil and Land Use Data .....	19
3.3. Data Preprocessing.....	22
3.3.1. Hydrologic Data Preprocessing .....	22
3.3.2. Topographic Data Preprocessing .....	24
3.3.3. Soil and Land Use Data Preprocessing .....	25
3.4. Data Processing.....	26

3.4.1.	Perimeter (P) and Drainage Area (A).....	26
3.4.2.	Length of Main River (LMR) .....	28
3.4.3.	Elevation Parameters (Hmin, Hmean, Hmax, BR) .....	28
3.4.4.	Slope Parameter (S) .....	29
3.4.5.	Aspect Parameter (ASPCT.) .....	30
3.4.6.	Mean Annual Temperature Parameter (T) .....	30
3.4.7.	Mean Annual Precipitation Parameter (MAP) .....	30
3.4.8.	Soil and Land Use Parameter (CN).....	30
3.4.9.	Mean Annual Discharge (Q) .....	30
4	HETEROGENEITY AND SEASONALITY ANALYSIS.....	31
4.1.	Heterogeneity Analysis .....	31
4.2.	Seasonality Analysis .....	34
4.2.1.	Seasonality Analysis Methodology .....	35
5	REGIONALIZATION OF FDACS.....	52
5.1.	Descriptive Statistics of the Parameters .....	53
5.1.1.	Perimeter (P) .....	58
5.1.2.	Area (A) .....	58
5.1.3.	Length of Main River (LMR) .....	60
5.1.4.	Maximum Elevation (Hmax) .....	62
5.1.5.	Minimum Elevation (Hmin).....	64
5.1.6.	Minimum Elevation (Hmin).....	65
5.1.7.	Mean Elevation (Hmean) .....	67
5.1.8.	Basin Relief (BR).....	68
5.1.9.	Slope (S).....	70
5.1.10.	Aspect (ASPCT.) .....	71
5.1.11.	Mean Annual Precipitation (MAP) .....	73
5.1.12.	Mean Annual Temperature (T) .....	75
5.1.13.	Curve Number (CN).....	77
5.1.14.	Mean Annual Discharge (Q) .....	79
5.2.	Correlation Analysis .....	81
5.2.1.	Kendall's $\tau$ .....	82
5.2.2.	Pearson's $r$ .....	82
5.2.3.	Interpretation of the Results of Correlation Matrix.....	83
5.2.3.1.	The computed Kendall's Tau ( $\tau$ ) values for 13 parameters .....	84
5.2.3.2.	The computed Pearson's ( $r$ ) values for 13 parameters .....	87
5.3.	Parametric Approach.....	88
5.4.	Jack Knife Cross Validation .....	94

5.5.	Regional Model Of Oltu Basin .....	117
5.6.	Quimpo's Parametric Approach .....	122
5.6.1.	Regional Index, c .....	123
5.6.2.	Area Analysis, QA .....	124
5.7.	Statistical Approach .....	128
6	MODEL VALIDATION .....	139
6.1.	Model Kocatepe Validation .....	140
6.2.	Model Quimpo Validation .....	141
6.3.	Model Comparison For HEPPs in Oltu Basin.....	147
6.4.	Model Validation for Short Samples .....	155
6.4.1.	Model Kocatepe Short-Term Duration Validation.....	156
6.4.2.	Model Quimpo Short-Term Duration Validation .....	160
7	SUMMARY AND CONCLUSIONS .....	163
7.1.	Summary of the Results .....	163
7.2.	Conclusions.....	165
7.3.	Recommendations.....	167
	REFERENCES .....	168
	APPENDICES	
A	DATA .....	172
B	HYDROLOGIC DATA PREPROCESSING.....	176
C	TOPOGRAPHIC DATA PREPROCESSING .....	244
D	DATA PROCESSING .....	250
E	HETEROGENEITY AND SEASONALITY ANALYSIS .....	252
F	DESCRIPTIVE STATISTICS OF PARAMETERS .....	274
G	SOFTWARE EASY FIT.....	281

# LIST OF TABLES

## TABLES

Table 3.1. The streamgauges and the in record range of the streamgauges in the study area	17
Table 3.2. Subbasins of Oltu Basin and their parameters obtained data processing	27
Table 4.1. Seasonality indices of stations for 5% probability of exceedance	37
Table 4.2. Similarity matrix for the subbasins for 5% probability of exceedance	42
Table 4.3. Seasonality indices of stations for 10% probability of exceedance	43
Table 4.4. Similarity matrix for the subbasins for 10% probability of exceedance	46
Table 4.5. Seasonality indices of stations for 20% probability of exceedance	47
Table 4.6. Similarity matrix for the subbasins for 20% probability of exceedance	49
Table 4.7. Seasonality indices of stations for 30% probability of exceedance	50
Table 5.1. Minimum, average and maximum values of climatic and geomorphologic characteristics of 9 basins used in the regression analysis	55
Table 5.2. Descriptive statistics of “Perimeter”	57
Table 5.3. Descriptive statistics of “Area”	59
Table 5.4. Descriptive statistics of “Length of Main River”	61
Table 5.5. Descriptive statistics of “maximum elevation”	63
Table 5.5. Descriptive statistics of “maximum elevation”	65
Table 5.6. Descriptive statistics of “Minimum Elevation”	67
Table 5.8. Descriptive statistics of “Basin Relief”	70
Table 5.9. Descriptive statistics of “Slope”	73
Table 5.11. Descriptive statistics of “Mean Annual Precipitation”	75
Table 5.12. Descriptive statistics of “temperature”	77
Table 5.13. Descriptive statistics of “Curve Number”	79
Table 5.14. Descriptive statistics of “Mean Annual Discharge”	80
Table 5.15. Kendall’s $\tau$ correlation matrix	84
Table 5.16. The Pearson’s $r$ correlation matrix	87
Table 5.17. Flow measurement stations and the record lengths	89
Table 5.18. The flow equations of the study of Algancı et al. (2009) and the $R^2$ values	90
Table 5.19. Nonlinear models developed by Castellarin et al. (2007)	90
Table 5.20. Linear models regressed and the performance indices	92
Table 5.21. Nonlinear models regressed and the performance indices	93
Table 5.22. Stations and corresponding empirical $q_j$ discharges in $m^3/s$	96
Table 5.23. Equations of $q_j$ discharges obtained by omitting DSI-2336	97

Table 5.24. Empirical and predicted discharges and relative errors for station DSI-2336 ....	98
Table 5.25. Performance indices calculated for DSI-2336 .....	98
Table 5.26. Equations of $q_j$ discharges obtained by omitting DSI-2335 .....	99
Table 5.27. Empirical and predicted discharges and relative errors for station DSI-2335 ....	99
Table 5.28. Performance indices calculated for DSI-2335 .....	100
Table 5.29. Equations of $q_j$ discharges obtained by omitting DSI-2324 .....	101
Table 5.30. Empirical and predicted discharges and relative errors for station DSI-2324 ..	102
Table 5.31. Performance indices calculated for DSI-2324 .....	102
Table 5.32. Equations of $q_j$ discharges obtained by omitting DSI-2323 .....	103
Table 5.33. Empirical and predicted discharges and relative errors for station DSI-2323 ..	104
Table 5.34. Performance indices calculated for DSI-2323 .....	104
Table 5.35. Equations of $q_j$ discharges obtained by omitting DSI-2337 .....	105
Table 5.36. Empirical and predicted discharges and relative errors for station DSI-2337 ..	106
Table 5.37. Performance indices calculated for DSI-2337 .....	106
Table 5.39. Empirical and predicted discharges, and relative errors for station DSI-2339 .	107
Table 5.40. Performance indices calculated for DSI-2339 .....	107
Table 5.41. Equations of $q_j$ discharges obtained by omitting EIE-2323 .....	108
Table 5.42. Empirical and predicted discharges and relative errors for station EIE-2323...	109
Table 5.43. Performance indices calculated for EIE-2323 .....	109
Table 5.44. Equations of $q_j$ discharges obtained by omitting EIE-2325 .....	110
Table 5.45. Empirical and predicted discharges and relative errors for station EIE-2325...	111
Table 5.46. Performance indices calculated for EIE-2325 .....	111
Table 5.47. Equations of $q_j$ discharges obtained by omitting EIE-2329.....	112
Table 5.48. Empirical and predicted discharges and relative errors for station EIE-2329...	113
Table 5.49. Performance indices calculated for EIE-2329 .....	113
Table 5.50. Jack-knife indices of performance results for each station .....	115
Table 5.51. Jack-knife indices of performance results for each duration .....	116
Table 5.52. Equations of $q_j$ discharges obtained for Model Kocatepe .....	118
Table 5.53. Comparison Table of the index of performances between jack-knifed and Kocatepe results for the stations .....	119
Table 5.54. Index of performances of jack-knifed cross-validation results and Model Kocatepe results for each duration.....	120
Table 5.55. $Q_A$ and $c$ values for each basin for Quimpo's et al. regional model .....	123
Table 5.56. Basins included in area regression analysis for Quimpo approach.....	125
Table 5.57. Basins included in area regression analysis for Quimpo approach.....	126
Table 5.58. Comparison of cases for Quimpo et al. area equation .....	126

Table 5.59. Index of performances for regional Model Quimpo .....	127
Table 5.60. Quimpo et al. regional model indices of perform. results for each duration ....	128
Table 5.61. Lognormal parameter estimators for modeling stations .....	129
Table 5.62. $\mu$ models and performance indices.....	131
Table 5.63. $\sigma$ models and performance indices .....	132
Table 5.64. Calculated and regionalized basin estimators .....	134
Table 5.65. Performance indices of all stations for statistical approach .....	137
Table 5.66. Overall evaluation of parametric and statistical approaches.....	138
Table 6.1. Validation stations and the basin parameters .....	139
Table 6.2. Performance indices of validation stations for Model Kocatepe .....	140
Table 6.3. Kocatepe model indices of performance results for each duration .....	141
Table 6.4. Validation stations and the parameters of Quimpo Model.....	142
Table 6.5. Quimpo model indices of performance results for each duration .....	143
Table 6.6. Comparison of validation results for parametric approaches.....	143
Table 6.7. HEPPs and the parameters used for validation .....	148
Table 6.8. HEPP characteristics .....	149
Table 6.9. Turbine types and range of heads (m).....	149
Table 6.9. HEPP parameters for Model Kocatepe .....	150
Table 6.10. $q_j$ discharges of HEPP 17 for both approaches .....	151
Table 6.11. Designed and calculated installed power of HEPPs .....	153
Table 6.10. (continued) .....	154
Table 6.13. 1 year record length performance indices .....	157
Table 6.14. 2 years record length performance indices .....	158
Table 6.15. 5 years record length performance indices .....	159
Table 6.17. 2 years record length performance indices .....	161
Table 6.18. 5 years record length performance indices .....	162

# LIST OF FIGURES

## FIGURES

Figure 3.1. SRTM view of Çoruh Basin and the boundaries of Oltu Basin with the basin streamgauges available in the basin .....	16
Figure 3.2. 45 elevation map sheets belonging to Oltu basin .....	19
Figure 3.3. Soil map of Oltu Basin .....	20
Figure 3.4. CORINE land cover/use map of Oltu basin .....	20
Figure 3.5. CORINE land cover/use classes .....	21
Figure 3.6. Daily FDC of station EIE-2329 (Class Interval Method) .....	23
Figure 3.7. Daily FDC of station EIE-2329 (Ranking Method) .....	23
Figure 3.8. Monthly FDC of station EIE-2329 .....	23
Figure 3.9. Yearly FDC of station EIE-2329 (year 1982).....	23
Figure 3.10. CN raster data of Çoruh Basin.....	25
Figure 3.11. Hypsometric curve of Oltu Basin .....	29
Figure 4.5. Seasonality space of station DSI-2321 .....	38
Figure 4.7. Seasonality space of station DSI-2323 .....	38
Figure 4.6. Seasonality space of station DSI-2322 .....	38
Figure 4.9. Seasonality space of station DSI-2333 .....	39
Figure 4.8. Seasonality space of station DSI-2324 .....	38
Figure 4.11. Seasonality space of station DSI-2336 .....	39
Figure 4.10. Seasonality space of station DSI-2335 .....	39
Figure 4.12. Seasonality space of station DSI-2337 .....	39
Figure 4.13. Seasonality space of station DSI-2338 .....	40
Figure 4.14. Seasonality space of station DSI-2339 .....	40
Figure 4.15. Stations on seasonality space for 5% probability of exceedance.....	41
Figure 4.16. Stations on seasonality space for 10% probability of exceedance.....	44
Figure 4.17. Stations on seasonality space for 20% probability of exceedance.....	48
Figure 4.18. Stations on seasonality space for 30% probability of exceedance.....	51
Figure 5.1. Scatter plot of “Perimeter” against “Mean Annual Discharge” .....	56
Figure 5.2. Scatter plot of “Area” against “Mean Annual Discharge” .....	58
Figure 5.3. Scatter plot of “Length of Main River” against “Mean Annual Discharge” .....	60



Figure 5.4. Scatter plot of “Maximum Elevation” against “Mean Annual Discharge” .....	62
Figure 5.5. Scatter plot of “Minimum Elevation” against “Mean Annual Discharge” .....	64
Figure 5.5. Scatter plot of “Minimum Elevation” against “Mean Annual Discharge” .....	66
Figure 5.6. Scatter plot of “Mean Elevation” against “Mean Annual Discharge” .....	68
Figure 5.7. Scatter plot of “Basin Relief” against “Mean Annual Discharge” .....	69
Figure 5.8. Scatter plot of “Slope” against “Mean Annual Discharge” .....	71
Figure 5.9. Scatter plot of “Aspect” against “Mean Annual Discharge” .....	72
Figure 5.10. Scatter plot of “M.A. Precipitation” against “Mean Annual Discharge” .....	74
Figure 5.11. Scatter plot of “Temperature” against “Mean Annual Discharge in” .....	76
Figure 5.12. Scatter plot of “Curve Number” against “Mean Annual Discharge” .....	78
Figure 5.13. Jack-knifed relative errors distributed through stations.....	117
Figure 5.14. Model Kocatepe relative errors distributed through station .....	121
Figure 5.15. Regionalized values of “c” .....	124
Figure 5.16. Comparison of observed FDC and Regionalized FDC at station DSI-2336 ...	135
Figure 5.17. Comparison of observed FDC and Regionalized FDC at station DSI-2335 ...	135
Figure 5.18. Comparison of observed FDC and Regionalized FDC at station DSI-2324 ...	135
Figure 5.19. Comparison of observed FDC and Regionalized FDC at station DSI-2323 ...	135
Figure 5.20. Comparison of observed FDC and Regionalized FDC at station DSI-2337 ...	135
Figure 5.21. Comparison of observed FDC and Regionalized FDC at station DSI-2339 ...	135
Figure 5.22. Comparison of observed FDC and Regionalized FDC at station EIE-2323....	136
Figure 5.23. Comparison of observed FDC and Regionalized FDC at station EIE-2325....	136
Figure 5.24. Comparison of observed FDC and Regionalized FDC at station EIE-2329....	136
Figure 5.25. Comparison of observed FDC and Regionalized FDC at station DSI-2313 ...	136
Figure 5.26. Comparison of observed FDC and Regionalized FDC at station DSI-2321 ...	136
Figure 5.27. Comparison of observed FDC and Regionalized FDC at station DSI-2322 ...	136
Figure 5.28. Comparison of observed FDC and Regionalized FDC at station DSI-2323 ...	137
Figure 5.29. Comparison of observed FDC and Regionalized FDC at station DSI-2338 ...	137
Figure 6.1. Empiric and regional FDCS for DSI-2313 region.....	144
Figure 6.2. Empiric and regional FDCS for DSI-2321 region.....	144
Figure 6.3. Empiric and regional FDCS for DSI-2333 region.....	145
Figure 6.4. Empiric and regional FDCS for DSI-2322 region.....	145
Figure 6.5. Empiric and regional FDCS for DSI-2338 region.....	146
Figure 6.6. HEPPs used for validation of regional model.....	147
Figure 6.7. FDC of HEPP 17 developed with Model Kocatepe .....	150
Figure 6.8. FDC of HEPP 18 developed with Model Kocatepe .....	152

Figure 6.9. FDC of HEPP 19 developed with Model Kocatepe .....	152
Figure 6.10. FDC of HEPP 40 developed with Model Kocatepe .....	152
Figure 6.11. FDC of HEPP 39 developed with Model Kocatepe .....	152
Figure 6.12. Empirical and modeled FDCs for EIE-2323 1 year sub-sample .....	156
Figure 6.13. Empirical and modeled FDCs for EIE-2323 2 years sub-sample .....	156
Figure 6.14. Empirical and modeled FDCs for EIE-2323 5 years sub-sample .....	156
Figure 6.15. Empirical and modeled FDCs for DSI-2339 1 year sub-sample .....	160
Figure 6.16. Empirical and modeled FDCs for DSI-2339 2 years sub-sample .....	160
Figure 6.17. Empirical and modeled FDCs for DSI-2339 5 years sub-sample .....	161
Figure A.1. Raw data of DSI daily flow discharges .....	172
Figure A.2. Raw data of EIE daily flow discharges .....	173
Figure A.3. Temperature raster map used as the climatic data in the study .....	173
Figure A.4. Precipitation raster map used as the climatic data in the study .....	173
Figure A.5. G48-d2 map section raster image .....	174
Figure A.6. H47-d2 map section vector image of elevation data .....	175
Figure B.1. EIE-2323 arranged data .....	176
Figure B.2 – Figure B.223 Daily, Monthly and Yearly FDCs of Stations .....	177-243
Figure C.1. H48-d1 map section TIN image .....	244
Figure C.2. G48-d3 map section DEM image .....	245
Figure C.3. Catchment boundaries and drainage network developed 1k stream .....	247
Figure C.4. Catchment boundaries and drainage network developed by 5k stream .....	247
Figure C.5. Catchment boundaries and drainage network developed by 10k stream .....	248
Figure C.6. Catchment boundaries and drainage network developed by 15k stream .....	248
Figure C.7. Catchment boundaries and drainage network developed by 20k stream .....	249
Figure C.8. Agree DEM generated with drainage network 1k .....	249
Figure D.1 - Figure D.2 Slope image of Oltu basin created from DEM and SRTM .....	250
Figure D.3 - Figure D.4 Aspect image of Oltu basin created from DEM and SRTM .....	251
Figure E.1. 3 homogeneous clusters obtained from station data .....	252
Figure E.2 - Figure E.15 Seasonality spaces of stations for 10 % duration .....	256
Figure E.16 - Figure E.29 Seasonality spaces of stations for 20 % duration .....	262
Figure E.30 - Figure E.43 Seasonality spaces of stations for 30 % duration .....	268
Figure F.1- Figure F.26 Histogram and Normal Lines of parameters .....	274-280

# LIST ABBREVIATIONS AND SYMBOLS

## ABBREVIATIONS AND SYMBOLS

A	; Area of basin ( $\text{km}^2$ )
ASPCT.	; Aspect of basin ( $^\circ$ )
BR	; Basin Relief
CN	; Curv Number, permeability index of basin (varies between 70 – 100)
DSI	; Devlet Su İşleri (General Directorate of State Hydraulic Works)
E	; Nash-Sutcliffe efficiency index
EIE	; Elektrik İşleri Etüt İdaresi (General Directorate of Electrical Power Resources Survey and Development Administration)
$E_j$	; Jackknifed Nash-Sutcliffe efficiency index
$E_s$	; Performance index
FDC	; Flow Duration Curve
HEPP	; Hydroelectric Power Plant
HGK	; Harita Genel Komutanlığı (General Command of Mapping)
$H_{\max}$	; Maximum elevation above sea level (m)
$H_{\text{mean}}$	; Mean elevation above sea level (m)
$H_{\min}$	; Minimum elevation above sea level (m)
LMR	; Length of Main River (km)
MAP	; Mean Annual Precipitation (km)
N	; Number of gauged basins
$N_D^*$	; Number of durations, $j$ , satisfying the condition $q_{s,j} \geq q^*$
$N_D$	; Number of durations used to develop modeled FDCs
P	; Perimeter of basin (km)
$P_1$	; 25 <sup>th</sup> percentile, first quartile
$P_2$	; 50 <sup>th</sup> percentile, second quartile
$P_3$	; 75 <sup>th</sup> percentile, third quartile
Q	; Mean annual discharge ( $\text{m}^3/\text{s}$ )
$q^*$	; Minimum discharge value for feasible utilization
$q_j$	; Empirical daily discharge for duration $j$ ( $\text{m}^3/\text{s}$ )
$q_{s,j}$	; Empirical daily discharge for station $s$ and duration $j$ ( $\text{m}^3/\text{s}$ )
$\hat{q}_{s,j}$	; Predicted daily discharges for station $s$ and duration $j$ ( $\text{m}^3/\text{s}$ )

$\bar{q}_{s,j}$	; Mean daily empirical discharge for station $s$ and duration $j$ ( $\text{m}^3/\text{s}$ )
RMSE	; Root Mean Square Error
RRMSE	; Relative Root Mean Square Error
$S$	; Slope of basin (%)
$s$	; Gauged stations
SPSS	; Statistical Package for the Social Sciences
$T$	; Temperature of basin ( $^{\circ}\text{C}$ )
$w_j$	; A coefficient
WRL	; Water Resources Laboratory
$\varepsilon_{s,j}$	; Relative error for basin $s$ and duration $j$
$\bar{\varepsilon}$	; Average mean relative error
$\bar{\varepsilon}_s$	; Mean relative error for station $s$
$\bar{\varepsilon}_j$	; Mean relative error for duration $j$
$\sigma_{\varepsilon,s}$	; Mean relative errors standard deviation
$\bar{\sigma}$	; Average of mean relative errors standard deviations

# CHAPTER 1

## INTRODUCTION

In water resources planning and management, a *flow duration curve* (FDC) is a reasonable and widely used tool for many kinds of applications; such as, hydropower generations, irrigation system designs, sedimentation problems and integrated basin management including pollution and erosion. Flow duration curve of a river is simply a brief summary of the magnitude of the discharge during its period of record (Quimpo, et al, 1983). However, obtaining FDCs for ungauged sites is still a challenging work. An ungauged site is simply where measured flows of a stream do not exist or are inadequate (quantitatively and/or qualitatively). To be able to control and take advantages of water resources for vital needs, namely energy, irrigation and water supply needs, FDCs for ungauged basins have to be prevailed.

Like in many countries in the world, Turkey is in need a considerable amount of energy. Moreover, importance of renewable energy resources increases day by day and it is a well-known fact that water resources of Turkey have not been used remarkably. More than half of the water resources are not in use of any kind of energy generation (Yanmaz, et al, 2007). Furthermore, there exist many ungauged sites where a hydroelectric power plant is appropriate to construct. In other words, the potential of these sites for energy generation is unfortunately unknown. In light of these, it can be said that, determination of regional flow duration curves is essential to generate highest amount of energy in ungauged basin.

A flow duration curve characterizes the relation between the daily, weekly, monthly and yearly (or another time) flow amount and its frequency in any particular station along a stream. It is an important tool that shows all possible amounts of flows in a river as in terms of percentage of time flow exceeded. This information is important for increasing the performance and sustainability of the energy generation.

Until today, for the ungauged sites (where flow data are absent, inadequate or inappropriate), FDCs have been constructed by using the data from the same basin or the adjacent basins. The most widely used methods while relating the stations are regression analysis or using the drainage area ratios. Regression analysis is only applicable if a certain level of data are available; otherwise, it is not possible to make a regression analysis (Yanık and Avcı, 2005). On the other method, drainage-area ratio method, is a rough approach for such an important issue.

Moreover, despite the fact that many studies have been performed about the subject of regionalization of FDCs, it is still an uncommon subject to study. Especially, in most studies the validation of the proposed regional model is not available.

In this study, several regional models of FDCs for daily streamflows have been developed in Oltu Basin, which is a subbasin of Çoruh Basin. Besides this, the performances of all regional models have been analyzed by *Jack-knife Cross-Validation Method* and several performance indices have been calculated.

Two main purposes of this study are:

- Developing regional FDCs through different approaches and evaluating the performances of these models with the application of jack-knife cross-validation method, and making a final decision for the appropriate regional FDC model.
- Evaluating the proper data duration in obtaining a dependable FDC by comparing the performance of regional FDCs with the empirical FDCs, for long durations (record lengths  $\geq 10$  years) and short durations (1, 2 and 5 years record lengths).

## 1.1. Organization of the Thesis

In Chapter 1, an *Introduction* of the thesis is available. Chapter 2 explains the *Methodology* used in the study and shows examples from *Literature* about regionalization of flow duration curves. In Chapter 3 the *Study Area and Data* have been described. Chapter 4 includes the *Heterogeneity Analysis*, which is used to explain the heterogeneous distribution of the subbasins; and the *Seasonality Analysis*, which accounts for defining the streamflow season and the source of the flow, which is due to either precipitation or snowmelt. In Chapter 5, *Regionalization of FDCs* (briefly, correlation analysis, parametric approaches, statistical approach, jack-knife cross-validation, regional model development) has been covered. Chapter 6 includes *Validation* results of the study, for ungauged basins, existing Hydroelectric Power Plants (HEPPs), and short-term durations. *Summary of Results* and *Conclusions* are in Chapter 7.

## CHAPTER 2

# LITERATURE SURVEY AND METHODOLOGY

### 2.1. Flow Duration Curves

As previously defined, a flow duration curve (FDC) of a river is simply a brief summary of the magnitude characteristics of the discharge during its period of record (Quimpo et al., 1983). Furthermore, it is presented with a cumulative distribution function of streamflows at a station. The earliest use of FDC's is attributed to Clements Herschel and dates back to 1880 FDCs (Foster, 1934). FDCs are being used in many fields of hydrology and water resources; such as, hydropower, irrigation systems planning and design, water supply. In addition to these, they are also used in water-quality management problems and river pollution problems. Moreover, Quimpo (1983) stated that FDCs provide information promptly; it is beneficial to construct FDCs for all candidate sites for small hydropower development.

#### 2.1.1. Construction of Flow Duration Curves

*Class Interval Method* and *Ranking Method* are two methods used in developing the FDCs.



#### 2.1.1.1. Class Interval Method

This is a widely used method for frequency analysis in statistics, especially if the data amount is large. The main idea behind the frequency analysis is grouping the data in intervals. Firstly, the number of intervals is decided. Number of intervals ( $k$ ) can be found by using the following equations (McCuen, 1993):

$$\sqrt{n} = k \quad (\text{Equation 2.1})$$

$$k = 1 + 3.33 \log_{10} n \quad (\text{Equation 2.2})$$

$$k = \frac{r \times n^{1/3}}{2 \times iqr}, \text{ } iqr = Q_3 - Q_1, r = x_{\max} - x_{\min} \quad (\text{Equation 2.3})$$

Where “ $n$ ” is the sample size, **iqr** means inter-quartile range,  $x_{\max}$  and  $x_{\min}$  are the maximum and minimum values of the data set.

Secondly, the data grouped into  $k$  groups and the number of observations that occur in a particular predefined interval are counted. Then cumulative number of occurrences is calculated and these values are converted to percentages. This gives the non-exceedance of the data, and the probability of exceedance should be obtained. By simply subtracting the non-exceedance values from the 100 %, probability of exceedance values are obtained. Finally, upper class limit versus probability of exceedance values are drawn and the FDCs are constructed.

#### 2.1.1.2. Ranking Method

In this method the observed discharges  $q_i$ ,  $i = 1, 2, \dots, N$ , are ranked and an ordered data set is obtained.  $q_1$  is the largest observation and  $q_N$  is the smallest one in the data set. The ordered,  $q_i$ , observations are plotted against their corresponding plotting position, probability of exceedance. For this study, the plotting position is the corresponding duration  $D_i$  that is dimensionless.

Generally, the Weibull plotting position is used, which is

$$p_i = P(Q > q_i) = \frac{i}{N+1} \quad (\text{Equation 2.4})$$

Ranking method and plotting position approaches are dependable methodologies while constructing FDCs, and they are used in many studies, for example Castellarin et al. (2007).

## 2.2. Literature Survey

Ungauged basins or limited amount of gauging in a basin is a common problem all over the world. Furthermore, it is a well-known fact that management of water resources is a vital task for each country. These facts led to the generating several formulations and procedures to be able to regionalize the flow duration curves of ungauged sites and for the sites with limited amount of discharge data.

Three approaches are being applied to regionalize the FDCs of a basin. These are statistical approaches, parametric approaches and graphical approaches.

In statistical approaches, a suitable probability distribution, such as normal distribution or lognormal distribution, is chosen and the parameters of the statistical distribution are estimated. Then a regional-regression analysis is performed and a regional model is developed for predicting distribution parameters, in the light of the several basin characteristics, such as climatic, meteorologic and geomorphologic characteristics.

Fennessey and Vogel (1990) proposed a two-parameter lognormal frequency distribution for daily discharges over the interval  $0.50 \leq p \leq 0.99$  for ungauged basins of Massachusetts.

The exceedance probability is as follows,

$$p = 1 - (2\pi)^{-1/2} \int_{-\infty}^{z_p} \exp(-\frac{1}{2} t^2) dt \quad (\text{Equation 2.5})$$

$$p = g(q_p | \mu, \sigma) \quad (\text{Equation 2.6})$$

$$z_p = [\ln q_p - \mu] / \sigma \quad (\text{Equation 2.7})$$

Similarly, Singh, Mishra and Chowdhary (2001) used a normal distribution to represent the 10-day streamflow series in the Himalayan region.

Parametric approaches define analytical relationships between basin parameters and FDCs. There are many models and procedures proposed as a parametric approach. One of the oldest approaches is recommended by Quimpo et al. (1983) for the ungauged basins in Philippines to evaluate the FDCs of small hydropower sites. A two-parameter exponential function is defined as in Equation 2.8

$$Q(D) = Q_A \exp(-cD) \quad (\text{Equation 2.8})$$

$Q_A$  and  $c$  are the parameters of the equation, and  $Q_A$  depends on drainage area and  $c$  stands for the climatic regimes.  $D$  is the duration in the FDC .

Similarly, Mimikou and Kaemaki (1990) performed a study in Greece to regionalize the hydrologic parameters. They used a third order polynomial equation for the regionalization. The parameters included in the model are, mean annual precipitation ( $P$ , mm), drainage area ( $A$ , km<sup>2</sup>), hypsometric fall ( $H$ , m) and main channel length ( $L$ , km) and mean annual flood  $Q$  (m<sup>3</sup>/s) of daily extremes. The following equations (Equation 2.9 – Equation 2.13) are for the regional model developed by Mimikou and Kaemaki (1990).

$$Q(D) = a - bD + cD^2 - dD^3 \quad (\text{Equation 2.9})$$

where

$$a = 0.0011 P^{0.526} A^{0.608} H^{0.007} L^{0.253} \quad (\text{Equation 2.10})$$

$$b = 0.053 P^{0.522} A^{0.684} H^{-0.181} L^{0.278} \quad (\text{Equation 2.11})$$

$$c = 0.010 P^{0.708} A^{0.952} H^{-0.315} L^{0.073} \quad (\text{Equation 2.12})$$

$$d = 4.215 \times 10^{-6} P^{7.157} A^{1.637} H^{-0.053} L^{-0.687} \quad (\text{Equation 2.13})$$

$D$  in Equation 2.9 is used for duration in FDC.

Franchini and Suppo (1996) proposed a parametric approach for estimating daily FDCS in Italy. They developed a model which describes the lower portion of FDCS's of daily streamflows (i.e.,  $D \geq 0.3$ ). The adopted equation is:

$$Q(D) = c + a(1.0 - D)^b \quad (\text{Equation 2.14})$$

Where,  $a$ ,  $b$  and  $c > 0$

$b$  is associated with physical characteristics of basin (e.g., imperviousness, size, etc.)

$a$  and  $c$  are associated with the climate and land use, and  $D$  is for duration in FDC.

Castellarin, et al, (2004) developed regional FDCs of eastern central Italy and adopted a validation for performance analysis. Nonlinear equations were proposed for the quantiles,  $Q_{30}$ ,  $Q_{70}$ ,  $Q_{90}$  and  $Q_{95}$ . Moreover, uncertainty of short samples of streamflow data was also analyzed.

Yu, et al, (2002), made a parametric analysis with various geomorphoclimatic indices in Taiwan and regionalize FDCs for daily discharges,  $Q_p$ , where  $p = 10, 20, \dots, 90\%$ . Unlike the others, Yu et al. (2002) evaluate the performance of the approach.

Yanık and Avcı (2005) studied the Black-Sea region in Turkey. Hierarchical and nonhierarchical cluster analysis methods were used to develop regional flow duration curves in the interval of 30%-100%.

The third approach while regionalizing the FDCs is the graphical approach. Moreover, *standardized curves* are one of the tools in graphical approach to the regionalization. FDCs of a gauged basin are standardized by dividing the empirical FDCs by an index flow. Then any FDC of an ungauged site is the product of this standardized FDC by using the index flow (Castellarin, et al, 2004).

In their study, Smakhtin, et al, (1997), consider their study area as a single homogeneous region and they normalized the observed FDCs by a long-term mean daily flow and averaged the normalized ordinates of the curves for ungauged sites.

Smakhtin et al. (1997) mentioned, “Methods for regional estimation of floods, low-flow indices, flow duration curves and low-flow frequency curves are described in a number of sources with examples from all over the world”. For example, in the study of “Development of regionalization procedures using a multi-model approach for flow in an ungauged catchment”, Goswami, et al, (2007) explored rainfall-runoff models and combination techniques. As topographic data, they used the mean altitude, the altitude at the highest point and the altitude at the outlet.

Moreover, seven years of daily rainfall, evaporation and discharge data of 12 catchments were used. Furthermore, climatologically, the humid seaboard climatic zone, the Mediterranean zone, the semi-continental zone, and intermediate climatic zones were used.

Some of the simulation models used in the study are; Parametric simple linear model, non-parametric simple linear model, linearly varying gain factor model, artificial neural network model and soil moisture accounting and routing model. Nash-Sutcliffe efficiency index was used for assessing the performances of regional models.

Mwakalila (2003) involved in a similar study in southwestern Tanzania to obtain a set of parameters to regionalize the hydrologic responses of catchments. He used the Geographical Information Systems with the physical indices like topography, geology, climate and land use. The identified data-based mechanistic (DBM) model structure was used in the study to

model the hydrologic responses. Finally, a multiple regression analysis was performed and several linear equations were developed. The DBM model was calibrated with six parameters (Mwakalila, 2003). These six parameters were derived by using the topographical indices, geology and soil index, climate indices and vegetation and cover indices. One of the DBM equation is as follows:

$$Q_{s(t)} = aQ_{s(t-1)} + bU_{(t)} \quad (\text{Equation 2.15})$$

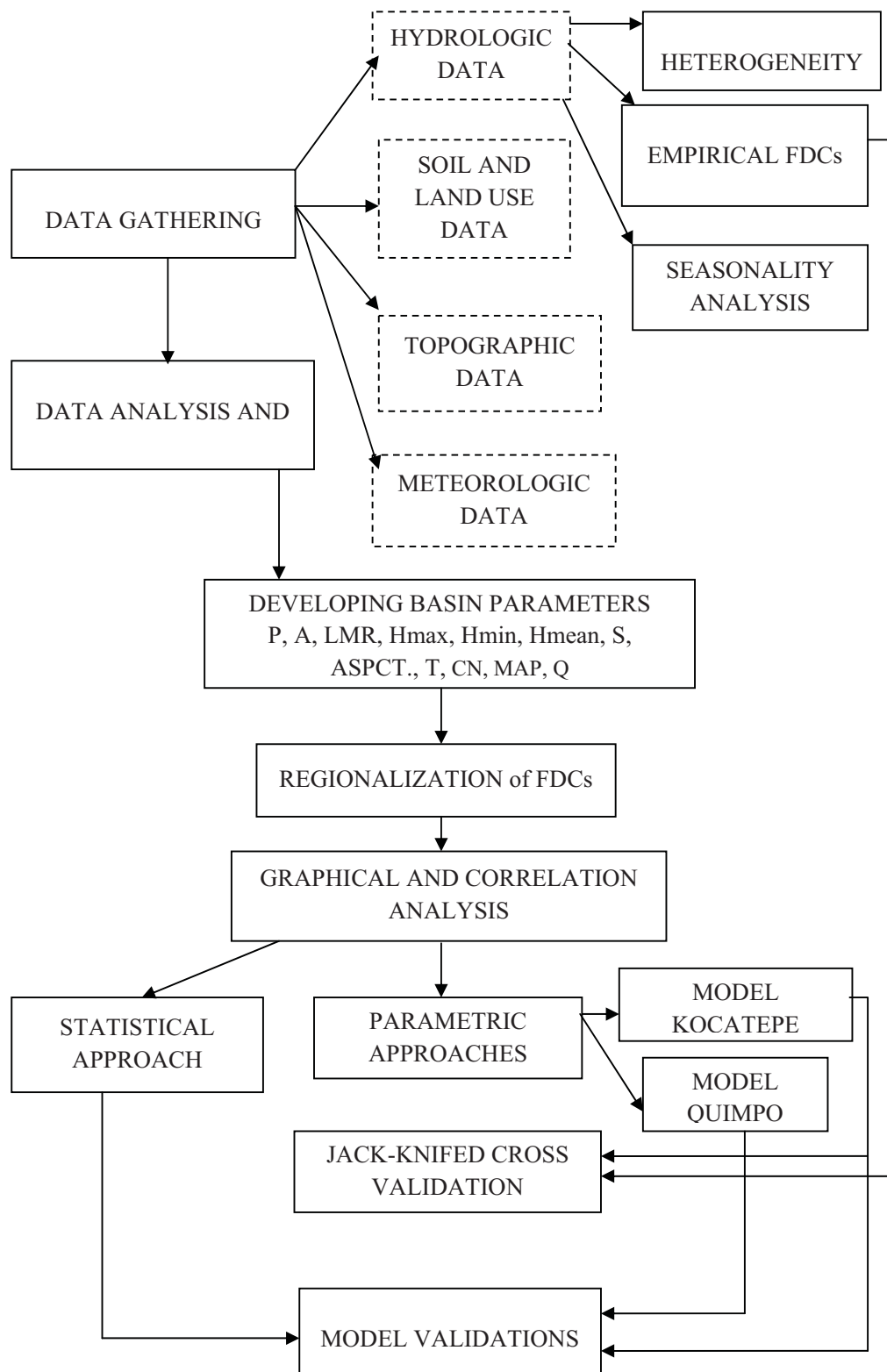
The six parameters in DBM model are:

- c, controls the proportion of rainfall which contributes to ground water storage as recharge
- m, describes an effective storage capacity of the catchment
- $Q_0$ , represents the discharge when recession flow commences
- n, express the non-linearity of surface runoff generation
- a, controls the mean residence time of surface runoff generation
- b, scales the differences in total volumes of input and output.

Sanborn and Bledsoe (2006) studied on predicting streamflow regime metrics for ungauged streams in Colorado, Washington and Oregon. The methodology that they used is stratifying the flow regimes of gauged sites, classifying the regimes of ungauged streams and developing models for predicting metrics for ungauged streams. Physical and climatic drainage basin characteristics used to compute flow regimes characterized by eighty-four streamflow metrics (Sanborn and Bledsoe, 2006). Multiple regression models developed as the predictive models for each flow regime type. The coefficient of determination,  $R^2$ , was used as the conservative estimate of model fit.

### 2.3. Methodology

The motivation of the study is the hydropower generation and the aim is evaluation of the regionalized daily FDCs in the ungauged basins. All durations of a FDC are considered. Parametric and statistical approaches are used to estimate regional FDC. The steps that followed in the study are summarized in Figure 2.1. In Figure 2.1, Drainage Area (A), Perimeter of basin (P), Length of Main River (LMR), Minimum Elevation ( $H_{\min}$ ), Mean Elevation ( $H_{\text{mean}}$ ), Maximum Elevation ( $H_{\max}$ ), Basin Relief (BR), Slope (S), Aspect (ASPCT.), Mean Annual Precipitation (MAP), Curve Number (CN) and Temperature (T) are stated.



### 2.3.1. Regionalization of FDCs

First step is the “data gathering” as in many researches, which is one of the most challenging parts in this study. General Directorate of State Hydraulic Works (DSI) and General Directorate of Electrical Power Resources Survey and Development Administration (EIE) are the hydrologic data suppliers, and General Command of Mapping (Harita Genel Komutanlığı, HGK) has been contacted for the topographic data. Meteorologic and climatic data have been gathered from the other studies available in Water Resources Laboratory.

The second step is the “data analysis and management”, which is as difficult as, “data gathering”. In this step, briefly, all data available are organized and prepared for further analysis. After organizing the data, *heterogeneity test* and *seasonality analysis* have been explored with the hydrologic data. *Heterogeneity test* is to identify if the basins are different in flow characteristics or not. In heterogeneous basins, basic approach which is interpolation according to area, are inadequate. Moreover, *seasonality analysis* is to understand the runoffs’ sources, namely, snow melting or spring precipitations. With the hydrologic data, empirical FDCs have been constructed, which is one of the essential analyses in the study.

After organizing available data, basin parameters have been extracted by using geographic information system (GIS) software, ArcGIS. The parameters have been chosen according to the literature study and the available data.

The fourth step deals with the parameters explored in the previous step, “graphical analysis” has been performed to learn the behavior of the parameters, the underlying statistical distribution, etc. However, the critical analysis is the “correlation analysis”, which shows the relation between parameters and the importance of each one of them for the discharges.

After this step, elimination is performed between the parameters to get the most satisfactory and explanatory ones to use in “model development”.

“Regional model development” is the core step in the analysis. In this step, a regional model has been developed for Oltu Basin, by trying several linear and nonlinear models. The indices of performances have been evaluated for each model and the most reasonable one has been selected. Then jack-knifed cross-validation has been performed to see the performance and robustness of the model.

As an extension of this step, another parametric approach and a statistical approach have been tested on the data to be able to compare the performance of the regional model, and create an alternative path.

As the final step, all of the approaches have been validated. Moreover, the regional model of Oltu Basin has been validated in terms of available HEPP projects and short-term duration response.

### 2.3.2. Performance Indices

Several statistical indices have been used to compare the empirical, modeled and jack-knifed FDCs. Castellarin et al. (2004) suggests the following first seven of nine indices for a non-linear regression performance study. These are:

1. Relative error,  $\epsilon_{s,j}$ , for basin  $s$  and duration  $j$ ;

$$\epsilon_{s,j} = \frac{\hat{q}_{s,j} - q_{s,j}}{q_{s,j}} \quad (\text{Equation 2.15})$$

where  $q_{s,j}$  and  $\hat{q}_{s,j}$  stands for empirical and modeled –or- jack-knifed daily discharges of duration  $j$ .

2. Mean relative error,  $\bar{\epsilon}_s$ , for basin  $s$  and the standard deviation of it,  $\sigma_{\epsilon,s}$ ;

$$\bar{\epsilon}_s = \frac{1}{N_D^*} \sum_{j=1}^{N_D} w_j \epsilon_{s,j} \quad (\text{Equation 2.16})$$

$$\sigma_{\epsilon,s} = \sqrt{\frac{1}{N_D^* - 1} \sum_{j=1}^{N_D} w_j (\epsilon_{s,j} - \bar{\epsilon}_s)^2} \quad (\text{Equation 2.17})$$

where;

$N_D$ ; number of durations used for the comparison,

$N_D^*$ ; number of durations  $j$  which belong to  $q^*$ ,

$q^*$  ; a limit for the discharge of water resources. Using the discharge below this limit is not feasible and may cause environmental problems.

$w_j$  ; a coefficient.

The relations between these variables are simply:

- $N_D^* \leq N_D$  for condition  $q_{s,j} \geq q^*$
- $w_j = 1$  , if  $q_{s,j} \geq q^*$ ;  $w_j = 0$  , otherwise.

$q^*$  is the minimum streamflow value (i.e.,  $q^* = 0.1 \text{ m}^3/\text{s}$ ) below which the utilization of the water resource is unfeasible or impacts significantly on the environment (Castellarin et al., 2004).



3. Average of  $\bar{\epsilon}_s$  and  $\sigma_s$ ,  $\bar{\epsilon}$  and  $\bar{\sigma}$ ;

$$\bar{\epsilon} = \frac{1}{N} \sum_{i=1}^N \bar{\epsilon}_s \quad (\text{Equation 2.18})$$

$$\bar{\sigma} = \frac{1}{N} \sum_{i=1}^N \bar{\sigma}_s \quad (\text{Equation 2.19})$$

These indices (Equation 2.18) and (Equation 2.19) give the information of overall performance of the regional model for the study region.

4. The mean, median and the percentiles 10 (P1), 25 (P2), 75 (P3) and 90 (P4), (as error bands of median) of N relative errors  $\epsilon_{i,j}$ . These indices then plotted against duration. The resultant graph helps to evaluate the uncertainty of the jack-knifed FDC's of the regional model.

5. Mean relative error for duration j,  $\bar{\epsilon}_j$ ;

$$\bar{\epsilon}_j = \frac{1}{N^*} \sum_{i=1}^N w_i \epsilon_{i,j} \quad (\text{Equation 2.20})$$

where;

- $N^* \leq N$  for condition  $q_{s,j} \geq q^*$
- $w_j = 1$ , if  $q_{s,j} \geq q^*$ ;  $w_j = 0$ , otherwise.

6. Performance index,  $E_s$ ;

$$E_s = 1 - \frac{\sum_{j=1}^{N_D} (\hat{q}_{s,j} - q_{s,j})^2}{\sum_{j=1}^{N_D} (q_{s,j} - \sum_{j=1}^{N_D} (q_{s,j}))^2} \quad (\text{Equation 2.21})$$

The performance index,  $E_s$ , takes values between 1 and  $-\infty$ . 1 means perfect fit, whereas 0 indicates that the model's performance as efficient as the mean regional value (Castiglioni, et al., 1970). Moreover, the  $E_s$  is very similar to the index Nash-Sutcliffe (Nash and Sutcliffe, 1970) efficiency criterion which is also used as an index of performance in this study.

7. Root mean square error, **RMSE**;

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N_D} (q_{s,j} - \hat{q}_{s,j})^2}{N_D - 1}} \quad (\text{Equation 2.22})$$

RMSE has the same units with the data and an unbiased estimator. The smaller the Mean Squared Error, the closer the fit is to the data.

8. Relative root mean square error, **RRMSE** ( Castiglioni, et al, 2009);

$$\text{RRMSE} = \sqrt{\frac{1}{N_D} \sum_{j=1}^{N_D} \left( \frac{q_{s,j} - \hat{q}_{s,j}}{q_{s,j}} \right)^2} \quad (\text{Equation 2.23})$$

9. Nash-Sutcliffe efficiency criterion, **E** ( Castiglioni et al., 2009)(Nash and Sutcliffe, 1970);

$$E = 1 - \frac{\sum_{j=1}^{N_D} (q_{s,j} - \hat{q}_{s,j})^2}{\sum_{j=1}^{N_D} (q_{s,j} - \bar{q}_{s,j})^2} \quad (\text{Equation 2.24})$$

## CHAPTER 3

### STUDY AREA AND DATA

#### 3.1. Study Area

Oltu Basin, which is a subbasin of Çoruh Basin, was selected as the study area. It has drainage area of 6857.6 km<sup>2</sup> and fifteen subbasin. For these fifteen subbasins daily discharge series observed between 1963 and 2005 are available. The gauging stations are operated by General Directorate of State Hydraulic Works (DSI) and General Directorate of Electrical Power Resources Survey and Development Administration (EIE). Oltu Basin lies at the Northeast of Turkey, within 40.16° to 41° North latitudes, 41.58° to 42.66° East longitudes (see Figure 3.1)

Oltu basin has been chosen as the study area, since the gauging in the basin is insufficient and many HEPP projects exist in this area. In addition to this the flows are not regulated in the basin.

#### 3.2. Data

In this study, during development, validation and evaluation stages of regional flow duration curves; topographic, hydrologic, meteorological, geologic data and soil and land use data have been used. In this chapter the data used in the study and the parameters obtained from these data are explained.

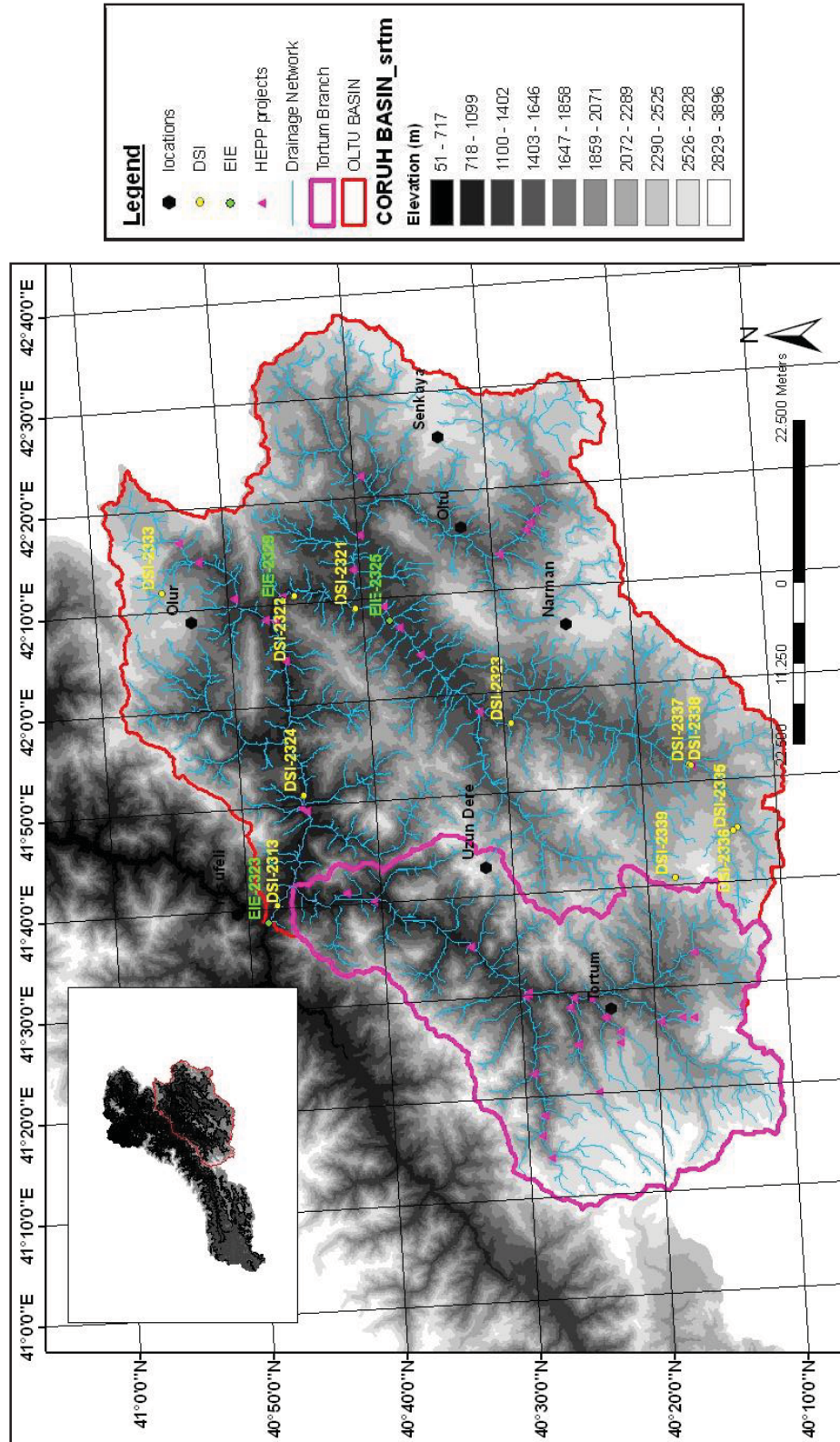


Figure 3.1. SRTM view of Çoruh Basin and the boundaries of Oltu Basin with the basin streamgauges available in the basin

### 3.2.1. Hydrologic Data

Daily flow discharges, which do not exceed two days covering the long periods, have been purchased from General Directorate of State Hydraulic Works (DSI) and General Directorate of Electrical Power Resources Survey and Development Administration (EIE). These data have been used for the purpose of obtaining FDCs of the subbasin, developing regional FDCs and validating the regional model.

The streamgauges and the record range of the streamgauges that satisfy the conditions stated above are presented in Table 3.1. The locations of these stations on Digital Elevation Model (DEM) of the basin are given in Figure 3.1. Furthermore, the raw data taken from DSI and EIE are available in Appendix A (see Figure A.1 and Figure A.2).

Table 3.1. The streamgauges and the in record range of the streamgauges in the study area

Station	Record Range	Record Duration (yrs)	Area (km <sup>2</sup> )
EIE-2323	1963, 1965-2003, 2005	41	6857.6
EIE-2325	1974-1989, 1991-2005	31	1762.0
EIE-2329	1982-2005	24	3538.8
DSI-2313	1967, 1972	2	6933.0
DSI-2321	1969,1971	2	1822.0
DSI-2322	1969-1972	4	3522.0
DSI-2323	1979-1994, 1998-1999	17	1118.3
DSI-2324	1976-1980, 1985-1996, 1998	17	4693.6
DSI-2333	1985, 1986-1988, 1993	4	43.4
DSI-2335	1991-1997, 1999	12	70.9
DSI-2336	1991-2000	10	55.5
DSI-2337	1991-2000	15	207.6
DSI-2338	1991-1996	3	72.7
DSI-2339	1990, 1994-1999	11	13.5

The HEPP projects, which are on application stage or those received licenses, have been obtained from DSI. These data would be used for validation of the installed capacity of projects by using regionalized FDCs. In Figure 3.1, the locations of the HEPP projects are depicted.

### 3.2.2. Meteorological Data

Mean annual temperature values of Turkey are presented in Appendix A, in Figure A.3, as a raster image. The raster map has been taken from another project conducted in WRL (Soytekin, 2010). Mean values are changing between 4°C and 19°C. Moreover, Oltu Basin is located at the Northeast of Turkey and the temperature values are low in this part of Turkey.

Mean annual precipitation data have been obtained by the similar way with Temperature data (Soytekin, 2010). The spatial distribution of temperature is given in Appendix A in Figure A.4.

### 3.2.3. Topographic Data

1:25000 scaled, 45 elevation map sheets belonging to Oltu basin (Figure 3.2) have been obtained from General Command of Mapping (Harita Genel Komutanlığı, HGK). The details of data taken from HGK are as follows:

- **Raster map:** Raster image that is generated by scanning and coordinating of 1:25000 scaled topographic maps. 45 items belongs to Oltu basins have been taken in GEOTIFF format from HGK. The map section, G48-d2 is in Appendix A, in Figure A.5.

Projection: UTM

Datum: ED-50

Format: GEOTIFF

- **Vector map:** The details and the information relating to these details, which are 1:25000 scaled topographic maps, are kept in digital format. These details are represented with point, line and area or with text type. Moreover, the coordinates and the details of these types are available on attribute tables. 45 items of YUKPAF and KARTO-25 vector maps belonging to Oltu basin have been taken in ARC/INFO coverage format. The vector map of H47-d2 is given in Appendix A in Figure A.6.

KARTO-25 Layers: Elevation (ele\_l, ele\_p), facilities (uti\_p, uti\_l, uti\_a), settlement (pop\_p, pop\_l, pop\_a), YUKPAF-25

Projection: UTM

Datum: ED-50

Format: ARC/INFO Coverage

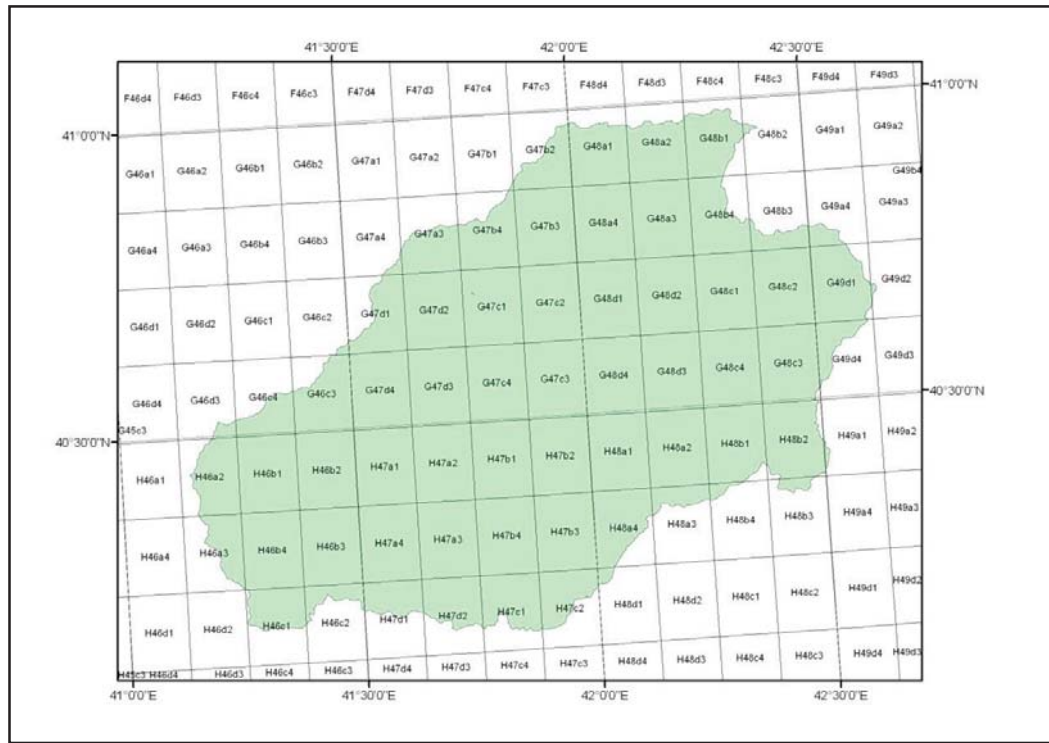


Figure 3.2. 45 elevation map sheets belonging to Oltu basin

#### 3.2.4. Soil and Land Use Data

1:100000 scaled soil map and CORINE land cover information have been obtained for Oltu basin. The soil map and CORINE land use /cover maps of Oltu Basin are presented in Figure 3.3, Figure 3.4 and Figure 3.5.



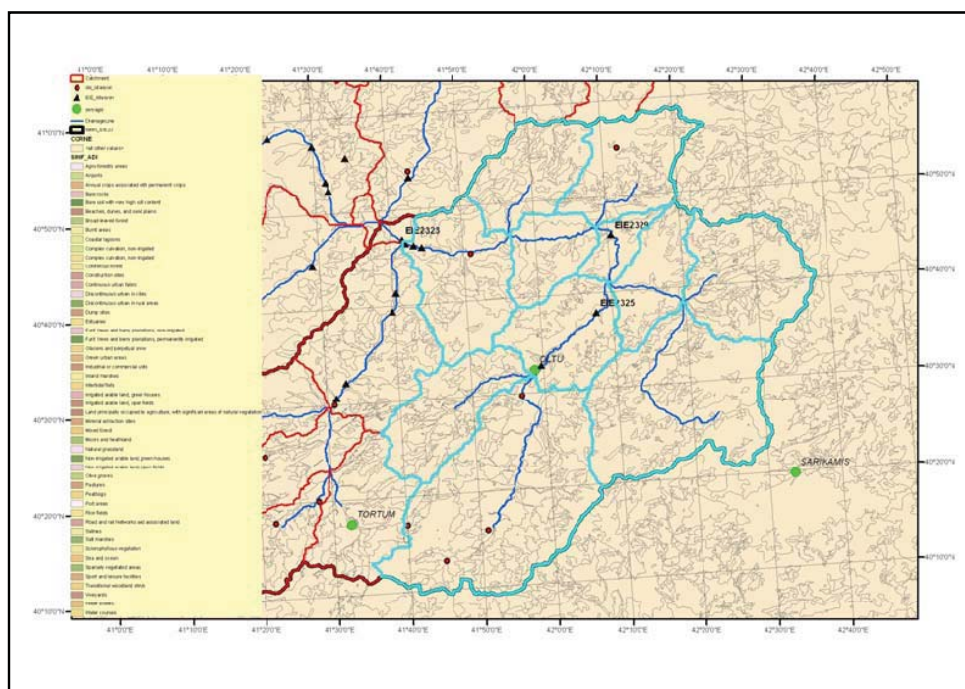
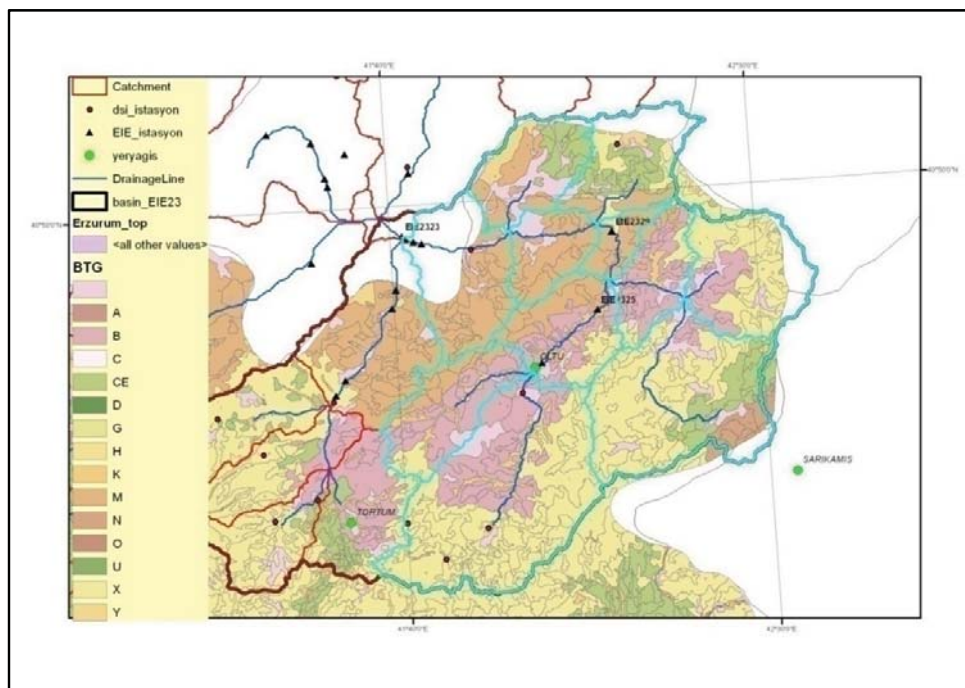






Figure 3.5. CORINE land cover/use classes

### 3.3. Data Preprocessing

Before using these raw data in the analyses, they have been processed and prepared to use in the software programs, such as Excel, ArcGIS, SPSS, etc. The following sections introduce these initial processes.

#### 3.3.1. Hydrologic Data Preprocessing

The raw hydrologic data, which are presented in Appendix A (see Figure A.1 and A.2), have been arranged by using Excel (see Appendix B, Figure B.1) and these data were prepared to be used in several analyses. The leading analysis for the hydrologic data is the construction of FDCs.

As explained in the Methodology chapter, two methods have been applied to construct flow duration curves (FDCs) on a chosen station, namely *class interval method* and *ranking method*. Then, the method, which gives reasonable results, has been chosen to construct FDCs of every station. EIE-2329 has been chosen to apply both of the methods, since the data belong to long years and are continuous. Figures 3.6 and Figure 3.7 show the FDCs of EIE-2329 developed by applying *class interval method* and *ranking method*, respectively.

Although both graphs in Figure 3.6 and Figure 3.7 show the same trend, the accumulation of small discharges in the *class interval method* causes a rough approach to the FDC. *Class interval method* does not allow us to see the real effects of small discharges, which are greater in number, and accept that the different values cause the same effects in that specified class. For the study of regionalizing FDCs, this approach is not realistic. By considering these facts, *Ranking Method* is applied for the rest of the stations.

After deciding on the method that would be used, daily, monthly and yearly FDCs have been constructed for each DSI and EIE station. Figure 3.8 and Figure 3.9 show the monthly and yearly (year 1982), FDCs of station EIE-2329, respectively. The rest of the FDCs of each station are available in Appendix B.

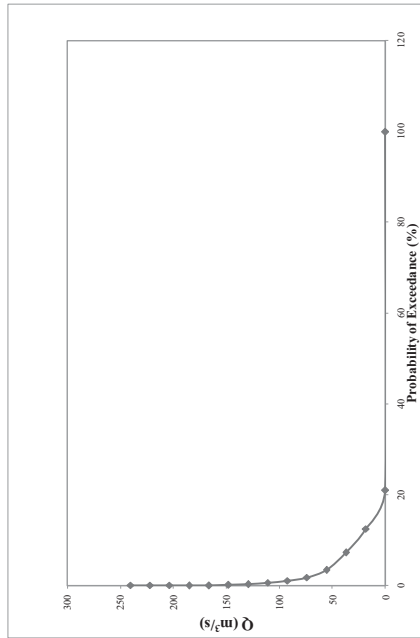


Figure 3.6. Daily FDC of station EIE-2329 (Class Interval Method)

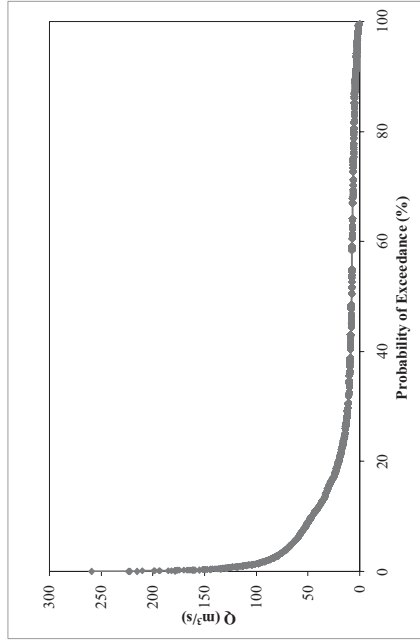


Figure 3.7. Daily FDC of station EIE-2329 (Ranking Method)

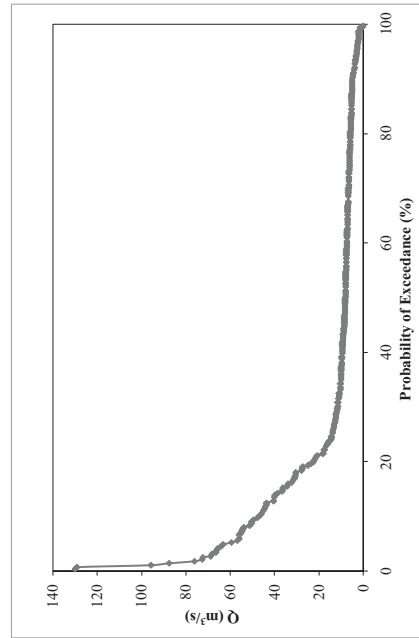


Figure 3.8. Monthly FDC of station EIE-2329

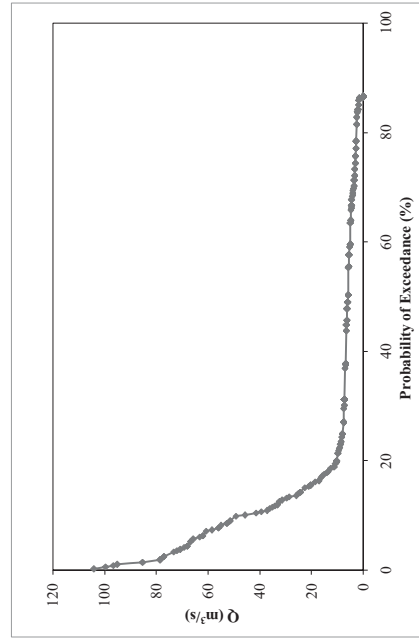


Figure 3.9. Yearly FDC of station EIE-2329 (year 1982)

### 3.3.2. Topographic Data Preprocessing

After purchasing the topographic data from HGK, the vector and raster maps have been examined to get defects if exist any. Moreover, it is realized that some of the digital map sections, which carry the elevation data, are misconfigured. The faulty map sheets are; G48-a1, G48-a2, G-48-b4, G48-c4, G49-a4, G49-d4, H47-a2, H47-b1, H47-b2, H47-b3, H47-b4, H47-c1, H47-d2, H48-a1, H48-a2, H48-a4, H48-b1 and H48-d1. Generally, the elevations are miswritten or incorrectly written. Moreover, H48-b1, H48-b2, H48-b3, H48-b4 and H49-a1 map sheets' projections are different than UTM/ED-50.

Firstly, all these errors have been corrected in ArcMAP and ERDAS Imagine 9.1, and the corrected digital data have been obtained. Secondly, by using ArcGIS, Triangular Irregular Networks (TINs) have been created for all map sections, which are used to represent the surface morphology or topography (see Figure C.1 in Appendix C). TINs created from digital data have been converted into corresponding Digital Elevation Models (see Figure C.2 in Appendix C). The DEMs have been used to quantify the characteristics of the land surfaces. In this study, DEMs have been merged to get Oltu Basin's DEM with 10 m resolution. This final DEM have been used to develop drainage line (flow direction process), catchment boundaries, Hypsometric Curves of subcatchments and other parameters of subbasins.

Catchment and drainage lines have been created as vector data in ArcHydro in ArcGIS. The steps of obtaining catchment and drainage line are listed in Appendix C.

One of the most important steps while processing catchment boundaries is to decide on stream value (step 4, Appendix C) while defining streams. It is generally taken as 10000 (10k). However, in this study, to get most detailed and processable catchment boundaries; 1k, 5k, 10k, 15k and 20k have been developed. The stream defined by 1k (1000 cells) would result in a larger number of subcatchments, but 20k (20000 cells) resulted few initial subcatchments. The resultant catchments and drainage networks can be seen in Appendix C in Figure C.3 – Figure C.7. As can be seen in the figures catchments and drainage networks defined by 1k are denser than the ones defined by 10k or 20k.

By considering these facts, the drainage network generated by 1k had been used as the drainage network of the Oltu Basin and the catchment defined by 20k had been used to construct subbasins under streamgauge stations of DSI and EIE.

After generating drainage line (1k), drainage lines have been used to obtain “Agree DEM” of Oltu Basin (Appendix C, Figure C.8).

Since topographic map of Tortum subbasin was not obtained from HGK , the DEM of the basin was obtained from Shuttle Radar Topography Mission (SRTM).

### 3.3.3. Soil and Land Use Data Preprocessing

To be able to represent the soil and land use/cover data in the analysis easily, CN values have been produced from soil map and CORINE land use/cover map and presented in Figure 3.10.

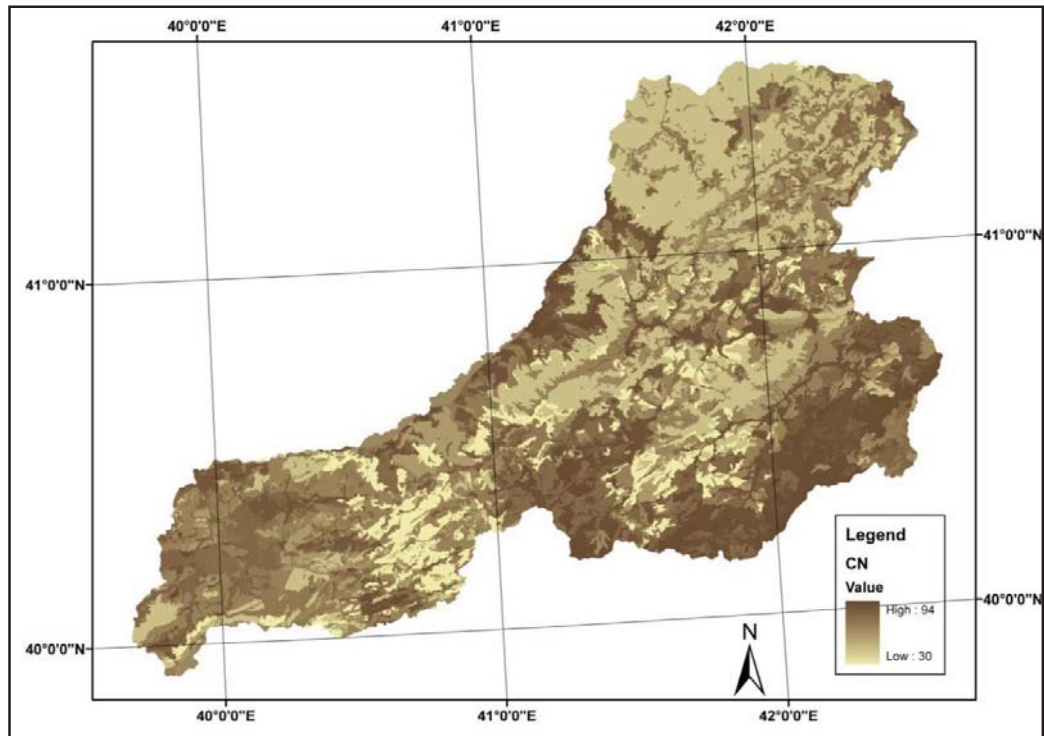


Figure 3.10. CN raster data of Çoruh Basin

### 3.4. Data Processing

As explained in the Literature section, there are many studies related with the topic of “regionalization of flow duration curves in ungauged basins”. In these studies, many different parameters have been used to regionalize FDCs in these ungauged basins. For example, Algancı (2009) and Castellarin et al. (2004) used area, perimeter, and elevation (minimum, average and maximum elevations), slope, aspect, precipitation, temperature and land use data. In addition to these Castellarin et al. (2004) used main channel length, mean annual evapotranspiration and mean annual net precipitation. Moreover, Mimikou and Kaemaki (1990) used mean annual precipitation, drainage area, hypsometric fall and length of main river course to model the ungauged basin in the Northwestern of Greece.

By considering the literature and data available for this study, it has been decided to obtain and use the parameters: Drainage Area (A), Perimeter of basin (P), Length of Main River (LMR), Minimum Elevation ( $H_{\min}$ ), Mean Elevation ( $H_{\text{mean}}$ ), Maximum Elevation ( $H_{\max}$ ), Basin Relief (BR), Slope (S), Aspect (ASPCT.), Mean Annual Precipitation (MAP), Curve Number (CN) and Temperature (T). These 12 parameters have been defined and generated by using available topographic, hydrologic, meteorological, soil and land use data.

#### 3.4.1. Perimeter (P) and Drainage Area (A)

The subcatchments of stations DSI-2313, DSI-2321, DSI-2322, DSI-2323, DSI-2324, DSI-2333, DSI-2335, DSI-2336, DSI-2337, DSI-2338, DSI-2339, EIE-2329, EIE-2325 and EIE-2323 have been produced in the data preprocessing; it is an easy task to obtain area and perimeter values of these basins since they are represented as polygons in ArcGIS. By using ArcGIS, the polygon areas ( $\text{km}^2$ ) and perimeters (km) have been calculated. The values are presented in Table 3.2 (first and second column, respectively). The values calculated in GIS medium of parameter area have been compared to the ones taken from DSI and EIE. According to this comparison, it is observed that the values are close to each other, which means the data operations handled in ArcGIS are meaningful.

Table 3.2. Subbasins of Oltu Basin and their parameters obtained data processing

Station	Perimeter (km)	Area (km <sup>2</sup> )	LMR (m)	H <sub>max</sub> (m)	H <sub>min</sub> (m)	H <sub>mean</sub> (m)	BR (m)	Slope (%)	Aspect (°)	MAP (mm)	CN	T (°C)	Q (m <sup>3</sup> /s)
DSI-2313	576.4	6967.9	158.4	3222.0	551.0	2020.1	1469.1	39.8	181.6	563.3	72.2	7.5	37.6
DSI-2321	310.2	1851.2	91.0	2990.0	1055.2	2022.6	967.4	17.6	175.2	521.1	73.1	7.3	5.5
DSI-2336	50.3	49.2	13.4	2755.6	1980.0	2344.5	364.5	15.4	175.6	423.6	86.6	6.8	0.6
DSI-2335	58.6	68.4	12.8	2813.8	1980.0	2371.9	391.9	17.6	170.8	430.3	89.0	6.9	0.6
DSI-2333	40.6	47.2	0.0	2837.2	2140.0	2411.9	271.9	9.5	161.0	669.6	73.5	8.0	0.8
DSI-2324	535.1	4537.5	139.7	3120.0	750.0	2006.1	1256.1	18.5	180.8	549.2	74.3	7.1	23.3
DSI-2323	225.9	1094.2	59.1	2928.0	1055.2	2122.7	1067.5	15.6	175.0	481.4	80.4	7.0	5.5
DSI-2322	444.6	3501.0	101.5	3120.0	1020.0	2049.0	1029.0	17.7	180.7	514.4	76.7	6.8	14.7
DSI-2337	104.0	197.0	27.1	2813.8	1630.0	2245.6	615.6	16.4	172.2	434.1	87.7	6.8	1.5
DSI-2338	55.9	72.8	16.3	2613.6	1157.4	2264.1	1106.7	14.5	187.0	444.0	84.9	6.8	0.6
DSI-2339	17.1	10.6	2.1	2758.5	2510.9	2624.8	113.9	27.3	196.3	441.0	86.2	7.1	0.2
EIE-2323	576.8	6978.7	161.2	3222.0	520.0	2033.1	1513.1	30.2	181.5	563.5	72.2	7.5	34.3
EIE-2325	301.4	1785.6	83.5	2990.0	1055.2	2032.3	977.1	17.5	175.5	518.5	73.5	7.3	7.1
EIE-2329	442.6	3537.1	104.5	3120.0	1010.0	2043.9	1033.9	17.8	181.0	515.2	76.5	6.8	16.5

### 3.4.2. Length of Main River (LMR)

From the drainage network generated before, the main rivers have been isolated and the lengths have been determined. Furthermore, the lengths of main rivers LMR are tabulated in Table 3.2.

### 3.4.3. Elevation Parameters ( $H_{\min}$ , $H_{\text{mean}}$ , $H_{\max}$ , BR)

DEM generated with 45 map sections of Oltu Basin have been used to determine the Minimum and Maximum elevations of basins. Zonal statistics is the main tool of ArcGIS while determining similar characteristics. The table shows the value, count, area, minimum value and maximum value, range of data, mean value, standard deviation and summation of the values. The values correspond to elevations for this analysis. Although deriving the maximum and minimum values is an easy process, deriving the mean elevation is a little more complicated. To derive the mean elevations, firstly the *Hypsometric Curves* should be obtained. Then, by using *Hypsometric Curves*, the mean elevations could be derived. In this study, by using CalHypso (Pérez-Peña, et al, 2009) hypsometric curves for all basins have been constructed. Figure 3.11 shows the hypsometric curve of Oltu Basin. The elevation parameters are also in Table 3.2. (Maximum Elevation  $H_{\max}$  (m), Minimum Elevation  $H_{\min}$  (m), Mean Elevation  $H_{\text{mean}}$  (m)).

The formula to get the  $H_{\text{mean}}$  from a hypsometric curve is:

$$H_{\text{mean}} = (H_{\max} - H_{\min}) \times \left(\frac{h}{H}\right)^{\frac{a}{A}} + H_{\min} \quad (\text{Equation 3.1})$$

Vivoni, et al,(2008) described the hypsometric distribution as “the relative height ( $h/H$ ) versus the relative area ( $a/A$ ), where  $a$  is the area of basin above height  $h$ ,  $A$  is the total basin area,  $h$  is the height above the basin outlet, and  $H$  is the total relief of the basin”.



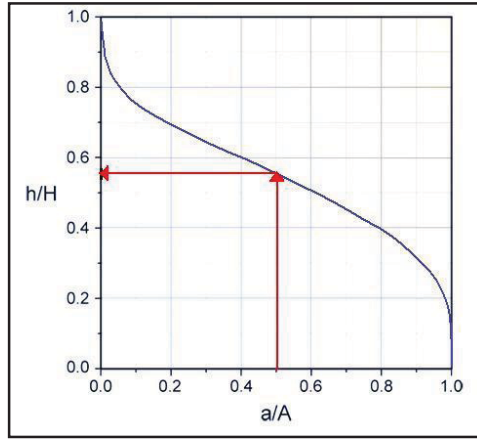


Figure 3.11. Hypsometric curve of Oltu Basin

The step, which is executed to get the  $H_{\text{mean}}$ , is an essential step to calculate the *basin relief* values of subbasins. Basin relief is a basin parameter that is used in many models developed for the basins, such as regional models of FDCs in the basins (Castellarin, et al, 2007) (Castellarin et al., 2004).

The following equation gives the basin relief (BR) parameters in a basin:

$$BR = H_{\text{mean}} - H_{\text{min}} \quad (\text{Equation 3.2})$$

The results are summarized in Table 3.2.

### Slope Parameter (S)

To be able to identify the slope features of the basins, a slope map (as a raster data set) is produced from the DEM of the Oltu Basin. Actually two slope maps have been created; one from the DEM produced by the 45 map sections and does not include Tortum Branch, the other one is DEM created for Oltu Basin from the SRTM image with a 10 m resolution, plus DEM produced before. The slope map is presented in Appendix D, Figure D.1.

After slope maps have been created, by using zonal statistics average slope value for each basin has been determined as a percent rise. Slope values (%) are available in Table 3.2.

#### **3.4.5. Aspect Parameter (ASPCT.)**

The steps performed while deriving slope values have been repeated for parameter Aspect. Again two aspect maps have been created; one for DEM data and one for SRTM plus DEM (for Tortum Branch, EIE-2323 and DSI-2313) data (see Figures D.3 and D.4, Appendix D). By using zonal statistics average aspect values ( $^{\circ}$ ), have been obtained for each basin (see Table 3.2).

#### **3.4.6. Mean Annual Temperature Parameter (T)**

The temperature map presented in Figure A.3, in Appendix A, has been used to derive the average temperature values in the subcatchments of Oltu Basin by using ArcGIS zonal statistics tools. The mean annual temperature T ( $^{\circ}\text{C}$ ) is given in Table 3.2.

#### **3.4.7. Mean Annual Precipitation Parameter (MAP)**

The precipitation map presented in Figure A.4, Appendix A, has been used to derive the average precipitation values in the subcatchments of Oltu Basin by using ArcGIS zonal statistics tools. The mean annual precipitation MAP (mm) is given in Table 3.2.

#### **3.4.8. Soil and Land Use Parameter (CN)**

The Curve Number (CN), which represents the raster image presented in Figure 3.10 has been used to develop zonal statistics for the subcatchments. The results of the zonal statistics analysis are depicted in Table 3.2.

#### **3.4.9. Mean Annual Discharge (Q)**

The mean annual discharge values (Q,  $\text{m}^3/\text{s}$ ) are also presented in Table 3.2.

## CHAPTER 4

# HETEROGENEITY AND SEASONALITY ANALYSIS

### 4.1. Heterogeneity Analysis

For any kind of regional analysis (frequency analysis, model development etc.) in hydrology, a heterogeneity (homogeneity) analysis is needed. It is also necessary to implement a heterogeneity analysis to develop a regional model in Oltu basin. If the subbasins show different flow characteristics, in other words if they are not distributed homogeneously, developing a regional model is reasonable. However, if the distribution of the subbasins in the basin is homogeneous, a complex model is not needed, a simple linear approach would be sufficient.

When the annual discharges in Table 3.2 of the subbasins and the other characteristics have been examined by inspection, it is observed that the subbasins show different characteristics. However, a heterogeneity test helps us to make it a certain conclusion. For this reason, Hosking and Wallis heterogeneity test (A. Castellarin, Burn, and Brath, 2008) have been applied to the discharge data of subbasins.

Hosking and Wallis heterogeneity test is a widely used homogeneity test in hydrology and its effectiveness has been tested in some studies recently (Viglione, Laio, and Claps, 2007).

It is a statistical test to analyze the homogeneity of a group of basins. Three measures of dispersions, which are obtained from L-moments, are used as the statistical measures (Hosking, 1990).

Castellarin (2008) summarizes the test as follows:

The basic idea behind the Hosking and Wallis heterogeneity test is to measure the sample variability of the L-moment ratios and compare it to the variation that would be estimated in

a homogeneous group. The mean value and the standard deviation of the dispersion measures, namely  $\mu_{V_k}$  and  $\sigma_{V_k}$ , are evaluated by repeated simulations, by generating homogeneous groups of basins having the same record length. Hosking and Wallis (1997) recommended using a four- parameter kappa distribution.

The equations of the measure of dispersions and the test statistics are:

1. A measure of dispersion for the L-Cv (L moment of coefficient of variation)

$$V_1 = \frac{\sum_{i=1}^R n_i (t_{2(i)} - \bar{t}_2)^2}{\sum_{i=1}^R n_i} \quad (\text{Equation 4.1})$$

2. A measure of dispersion for both the L-Cv and the L-Cs (L moment of coefficient of skewness) coefficients in the L-Cv – L-Cs space

$$V_2 = \frac{\sum_{i=1}^R n_i [(t_{2(i)} - \bar{t}_2)^2 + (t_{3(i)} - \bar{t}_3)^2]^{1/2}}{\sum_{i=1}^R n_i} \quad (\text{Equation 4.2})$$

3. A measure of dispersion for both the L-Cs and the L-Kurtosis coefficients in the L-Cs – L-Kurtosis Space

$$V_3 = \frac{\sum_{i=1}^R n_i [(t_{3(i)} - \bar{t}_3)^2 + (t_{4(i)} - \bar{t}_4)^2]^{1/2}}{\sum_{i=1}^R n_i} \quad (\text{Equation 4.3})$$

where  $\bar{t}_2$ ,  $\bar{t}_3$  and  $\bar{t}_4$  are the group mean of L-Cv, L-Cs and L-Kurtosis, respectively;  $t_{2(i)}$ ,  $t_{3(i)}$  and  $t_{4(i)}$ , and  $n_i$  are the values of L-Cv, L-Cs and L-Kurtosis and the sample size for site I; R is the number of sites in the pooling group.

4. The heterogeneity measures  $H_k$ ;

$$H_k = \frac{V_k - \mu_{V_k}}{\sigma_{V_k}}; \text{ for } k = 1, 2, 3. \quad (\text{Equation 4.4})$$

Hosking and Wallis suggested the evaluation criteria as, if:

$H_k < 1$  , acceptably homogeneous

$1 \ll H_k < 2$  , possibly heterogeneous

$H_k \gg 2$  , definitely heterogeneous

Moreover, generally  $H_1$  is used as the criterion to decide, however in this study, all three measures have been calculated to be on the safe side.

The procedure for this test can be summarized as follows:

1. The discharge data, which belong to fourteen (14) different stations (DSI and EIE stations), have been pooled into homogeneous groups by applying Cluster Analysis in software STATISTICA. Hundred (100) simulations have been generated.
2. The clustered data are then converted into Excel files.
3. A software developed by Hosking for The L-moments calculation and Legendre Polynomial have been obtained through internet (Hosking et al., 1997), and they have been used to develop a program to get four L-moments in MATLAB.
4. The program developed in MATLAB has been used to get the L-moments of the homogeneous clusters and the values have been written in different Excel files.
5. Finally, the Equations 4.1, 4.2, 4.3 and 4.4 have been applied to the data. In brief, hundred sets of  $V_1$ ,  $V_2$  and  $V_3$  have been calculated, and then the mean and standard deviations have been calculated. As the last step,  $H_1$ ,  $H_2$  and  $H_3$  have been computed.

For a single heterogeneity test, see Appendix E Heterogeneity Test.

The results are:

- $H_1 = 1.12$
- $H_2 = 0.94$
- $H_3 = 0.92$

As mentioned before,  $H_1$  is a sufficient measure and it is between the interval [1,2), and accepted as *possibly heterogeneous*.  $H_2$  and  $H_3$  are close to 1, which is reasonable to accept them as heterogeneous.

To sum up, the  $H_1$ ,  $H_2$  and  $H_3$  are in the interval of “possibly heterogeneous”, and the distribution of the subbasins have been accepted as heterogeneous and it is relevant to generate a regional model.

## 4.2. Seasonality Analysis

Seasonality indices are used greatly in hydrologic and water resources designs. Cunderlink and Burn (2001) studied hydro-climatological controls on floods to be able to estimate flood frequencies. McCuen (2003) recommended a method to make seasonal flood frequency analysis for gauged and ungauged sites and made a regionalization; Burn (1997) used catchment similarity to obtain regional flood estimates by using seasonality measures; and Castellarin, Burn and Brath (2000) evaluate the effectiveness on several measures (seasonality, size of rainfall extremes and permeability) for flood frequency analysis. All these studies are concentrated on flood issues. However, Önöz and Albostan (2007) proposed that seasonality analysis is an important tool while launching water resources policies. Moreover, they claimed that high flows play an important role while planning and operating hydroelectric power plants (HEPP), and seasonality measures can be used to analyze these high flows.

Seasonality analysis explains several questions in a basin. These are:

- the dates of high and low flows which are important from the floods and droughts point of views
- the regularity of the flows which is a necessary information while operating a HEPP or a dam for irrigation or water supply
- amount of installed capacity of power of a HEPP or a dam
- physical, geographic and meteorologic similarities between basins; basins similar from this point of view give the same responses to the flows come at the same time zones (Önöz and Albostan, 2007)
- Sustainable development policies

#### 4.2.1. Seasonality Analysis Methodology

Önöz and Albostan (2007) stated that on a FDC, discharges up to 5% probability of exceedance are extremely high but they occur in a small period of time. Moreover, this small period can occur in a regular basis. These discharges are used for flood analysis but they may also be used to find out the installed capacity of a power plant. The discharges up to 5% probability of exceedance will be analyzed from seasonality manner. Since the subbasin areas are different from each other, FDCs with specific discharges are used and this makes easy to compare basins' similarities.

Mardia (1972) proposed the seasonality index, which is based on two parameters,  $\theta$ , and  $r$ , which are calculated from the Julian dates of all days of the observation period when discharges are equal or below  $Q_{95}$ , by means of circular statistics.

The parameter  $\theta$  is the mean day of occurrence, measured in radians. It takes values between 0 and  $2\pi$ . "0" value is for 1 January;  $\pi/2$  relates to 1 April,  $\pi$  relates to 1 June and  $3\pi/2$  is for 1 October. The date of occurrence of the discharge  $i$ , can be written in angular units as converting the Julian date of occurrence into an angular measure through;

$$\theta_i = (Julian\ Date)_i \times \frac{2\pi}{365} \quad (\text{Equation 4.5})$$

Each date of occurrence can be represented in polar coordinates as a vector with a unit magnitude and a direction given by Equation 4.5. This allows the determination of the x and y coordinates of the mean of a sample of n dates of occurrence as (Castellarin, et al, 2001).

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^n \cos(\theta_i) \quad (\text{Equation 4.6})$$

$$\bar{y} = \frac{1}{n} \times \sum_{i=1}^n \sin(\theta_i) \quad (\text{Equation 4.7})$$

The direction,  $\bar{\theta}$ ; along with the magnitude,  $r$ , of the vector representing this point in polar coordinates can then be obtained by

$$\bar{\theta} = \arctan\left(\frac{\bar{y}}{\bar{x}}\right) \quad 1^{\text{st}} \text{ and } 4^{\text{th}} \text{ quadrants: } x > 0 \quad (\text{Equation 4.8})$$

$$\bar{\theta} = \arctan\left(\frac{\bar{y}}{\bar{x}}\right) + \pi \quad 2^{\text{nd}} \text{ and } 3^{\text{rd}} \text{ quadrants: } x < 0 \quad (\text{Equation 4.8})$$

The mean day of occurrence is obtained by back transforming the mean angle to a Julian date:

$$\text{MD} = \bar{\theta} \times \frac{365}{2\pi} \quad (\text{Equation 4.9})$$

The basins having similar MD values can show similar hydrologic characteristics (Önöz and Albostan, 2007)

The parameter  $r$  is the mean resultant of days of occurrence, which is a dimensionless measure of the variability of flow seasonality.  $r$  range from zero to unity, with  $r = 1$  indicates strong seasonality, which means flow events occurred on exactly the same day of the year, and  $r = 0$  indicating no seasonality, which means high flow events are uniformly distributed over the year.

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (\text{Equation 4.10})$$

The Euclidian distance between the basins is a good indicator to observe the differences between the basins.

$$d_s^{ij} = \sqrt{(\bar{x}_i^2 - \bar{x}_j^2)^2 + (\bar{y}_i^2 - \bar{y}_j^2)^2} \quad (\text{Equation 4.10})$$

The smaller value the  $d_s^{ij}$  takes, the more hydrologic similarity is expected between the basins.

In this analysis, the daily flow discharges taken from the DSI and EIE stations have been used. For each subbasin, the days on which discharge was smaller than  $Q_{95}$  have been extracted over the period of record from the specific FDCs and have been transformed into Julian dates. Then,  $\bar{x}$  and  $\bar{y}$ ,  $\bar{\theta}$ , MD,  $\bar{r}$  and  $d_s^{ij}$  values have been calculated respectively. The calculated indices are in Table 4.1.

The most important index,  $\bar{r}$ , has values close to 1, which indicates strong seasonality. In other words, all high flow values show up in the same period.



To be able to visualize the MD values on seasonality space, they have been plotted according to the  $\bar{x}$  and  $\bar{y}$ . Figure 4.1 – Figure 4.14 show these seasonality spaces.

Table 4.1. Seasonality indices of stations for 5% probability of exceedance

Station	Probability of Exceedance	$\bar{x}$	$\bar{y}$	$\bar{\theta}$	MD	$\bar{r}$
EIE-2323	5%	-0.651	0.711	2.31	134	0.96
EIE-2325	5%	-0.562	0.771	2.20	128	0.95
EIE-2329	5%	-0.586	0.765	2.22	129	0.96
DSI-2313	5%	-0.654	0.716	2.31	134	0.97
DSI-2321	5%	-0.494	0.711	2.18	126	0.87
DSI-2322	5%	-0.582	0.762	2.22	129	0.96
DSI-2323	5%	-0.532	0.794	2.16	126	0.96
DSI-2324	5%	-0.599	0.758	2.24	130	0.97
DSI-2333	5%	-0.585	0.768	2.22	129	0.97
DSI-2335	5%	-0.591	0.768	2.23	129	0.97
DSI-2336	5%	-0.654	0.729	2.30	134	0.98
DSI-2337	5%	-0.588	0.770	2.22	129	0.97
DSI-2338	5%	-0.500	0.815	2.12	123	0.96
DSI-2339	5%	-0.585	0.673	2.29	133	0.89

If the dates of flow events (MD column in Table 4.1) and the physical characteristics of the stations (Table 3.1) are examined, it is observed that there is no relation between the station characteristics and the flow events. MD column in Table 4.1 shows also about the average dates of flow events. Generally, the high flows (flows  $< Q_{95}$ ) in Oltu basin occurred in May. To be more specific, seasonality spaces should be examined.

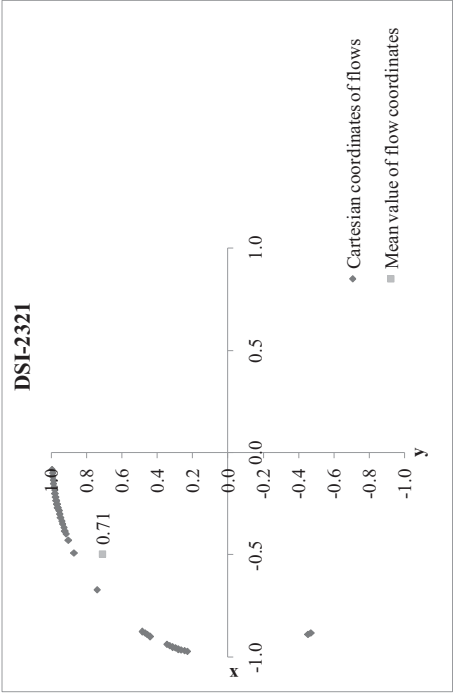


Figure 4.5. Seasonality space of station DSI-2321

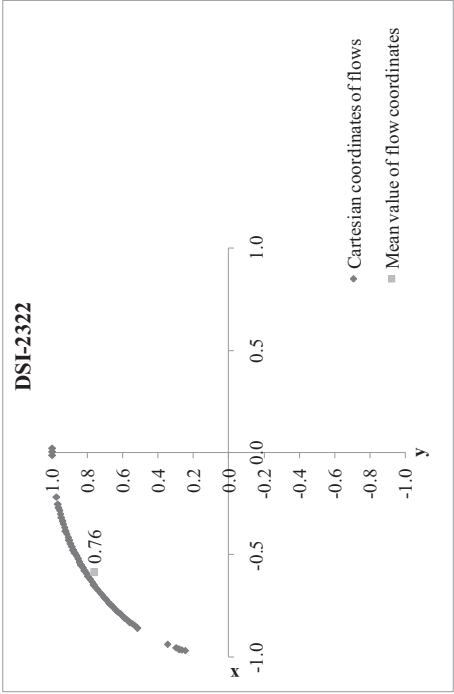


Figure 4.6. Seasonality space of station DSI-2322

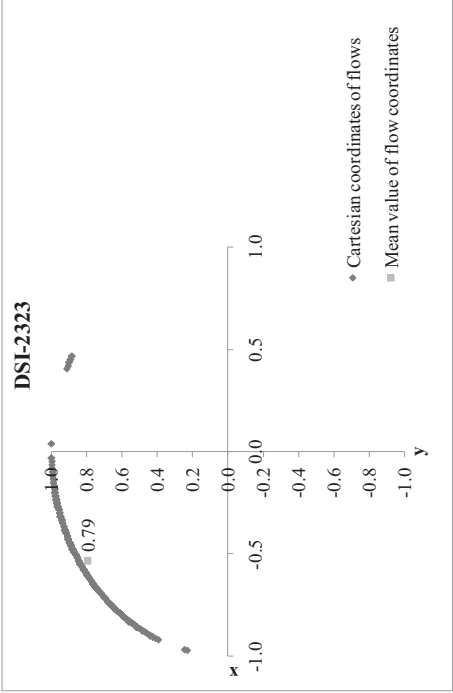


Figure 4.7. Seasonality space of station DSI-2323

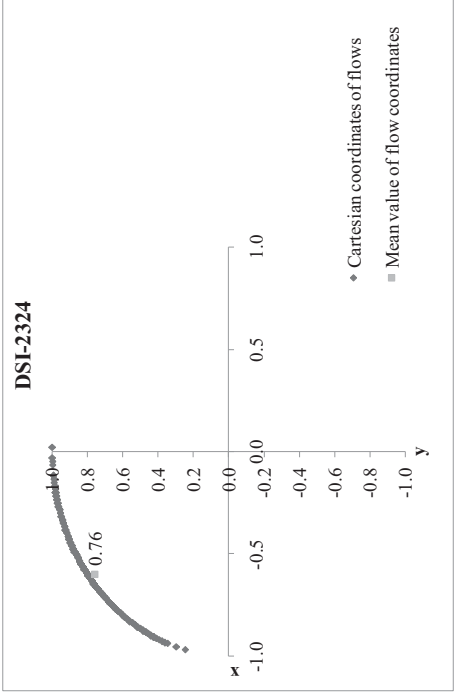


Figure 4.8. Seasonality space of station DSI-2324

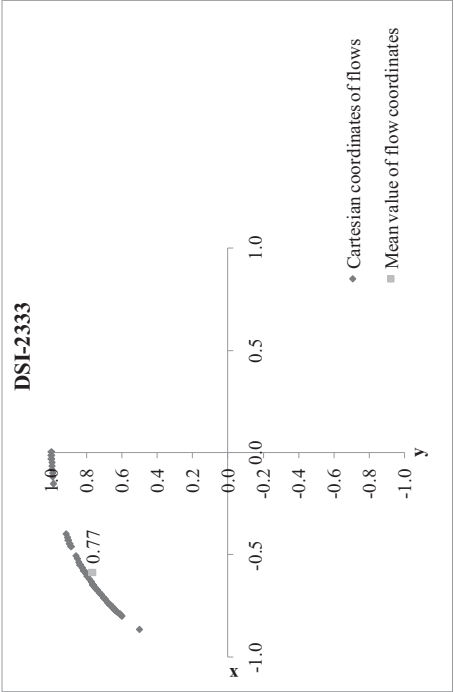


Figure 4.9. Seasonality space of station DSI-2333

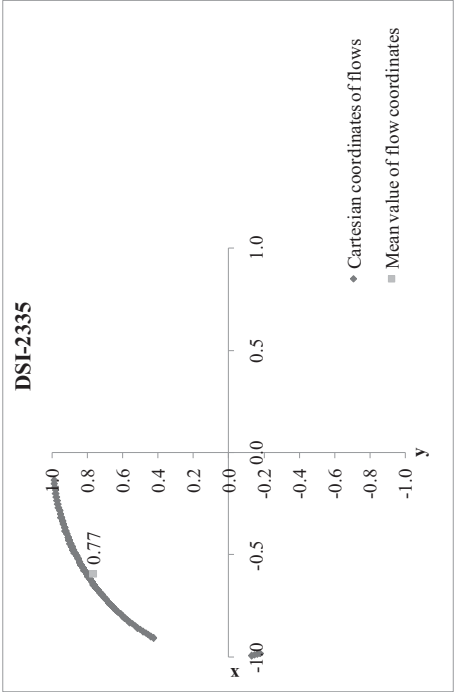


Figure 4.10. Seasonality space of station DSI-2335

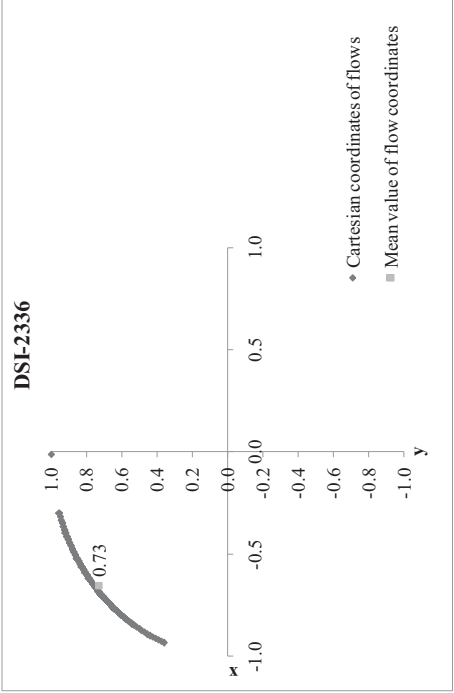


Figure 4.11. Seasonality space of station DSI-2336

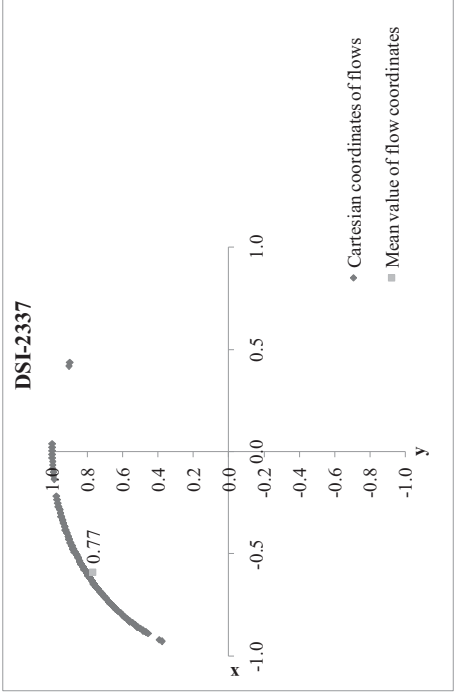


Figure 4.12. Seasonality space of station DSI-2337

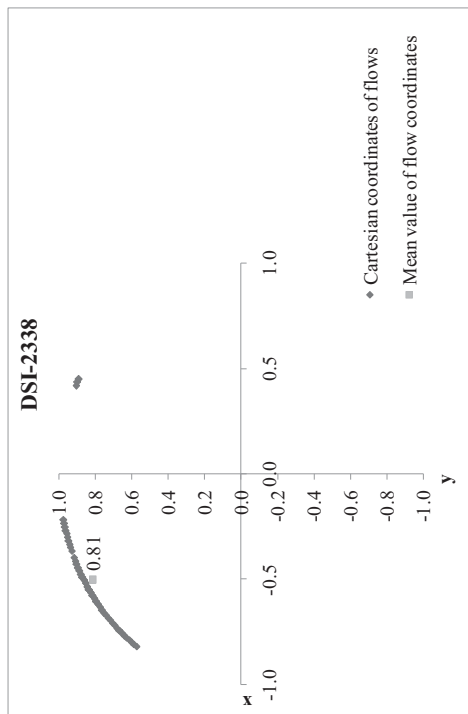


Figure 4.13. Seasonality space of station DSI-2338

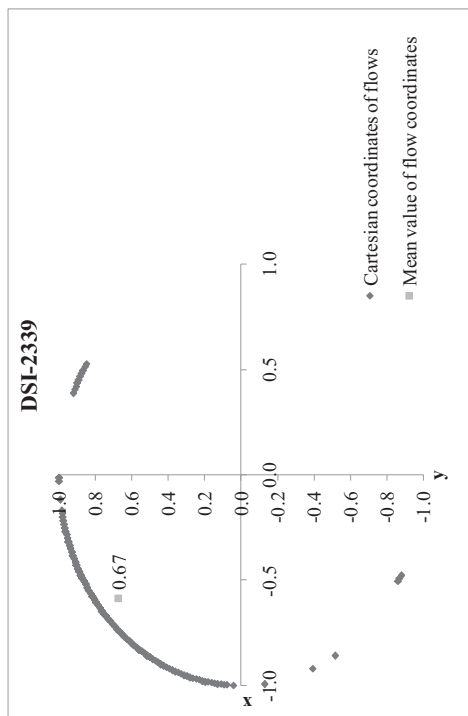


Figure 4.14. Seasonality space of station DSI-2339

Figure 4.1 – Figure 4.14 display the seasonality analysis results for discharges up to 5% probability of exceedance. 1<sup>st</sup> quadrant represents winter and 2<sup>nd</sup> quadrant represents spring, and so on. The figures can be summarized as all discharges up to 5% duration are due to the snow melting. There are only a few flow periods caused by rainfall, and they happened mostly in autumn (3<sup>rd</sup> quadrant) and some of them occurred in winter (EIE-2323, EIE-2329, DSI-2323, DSI-2337, DSI-2338 and DSI-2339).

Average of  $\bar{r}_i$  values change between 0.67 (DSI-2339) and 0.81 (DSI-2338), which means strong seasonality exists in the basin for 5% probability of exceedance. Figure 4.15. also proves this claim. The stations are closely located in the seasonality space.

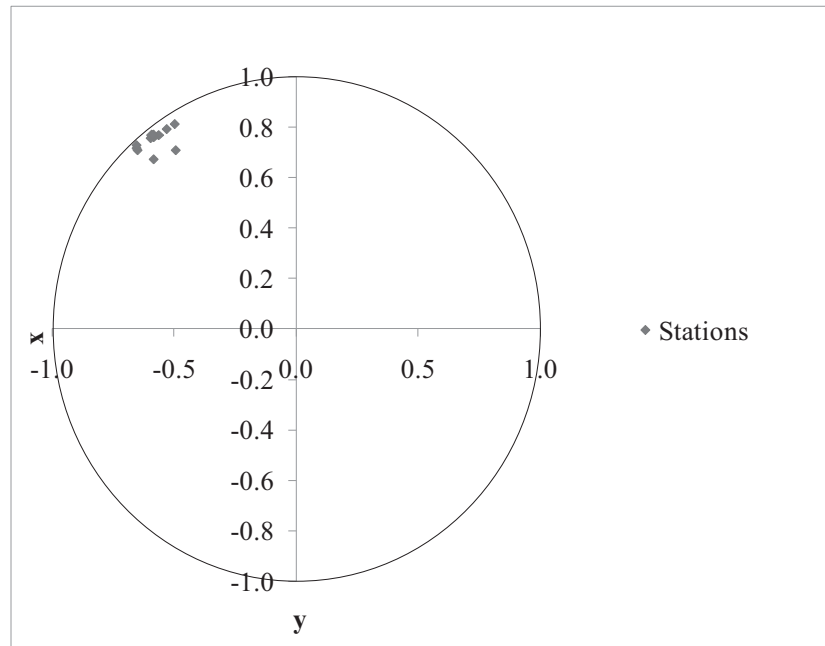


Figure 4.15. Stations on seasonality space for 5% probability of exceedance

As a final analysis, the  $d_s^{ij}$  values, namely the similarity measures, are calculated. The similarity matrix is given in Table 4.2. When the  $d_s^{ij}$  values are examined for probability of exceedance of 5%, it is observed that the DSI-2321, DSI-2313 and DSI-2338 stations are most unlikely station with the rest of the stations.

Table 4.2. Similarity matrix for the subbasins for 5% probability of exceedance

Basin	EIE-2323	EIE-2325	EIE-2329	DSI-2313	DSI-2321	DSI-2322	DSI-2323	DSI-2324	DSI-2333	DSI-2335	DSI-2336	DSI-2337	DSI-2338	DSI-2339
EIE-2323				✓									X	
EIE-2325	X			X		✓								
EIE-2329					X				✓	✓		✓		
DSI-2313	✓												X	
DSI-2321		✓		X							X			
DSI-2322			✓		X				✓					
DSI-2323	X	✓		X										
DSI-2324			✓		X					✓			X	
DSI-2333			✓		X	✓				✓		✓		
DSI-2335				X					✓			✓		
DSI-2336				✓									X	
DSI-2337			✓		X				✓	✓				
DSI-2338	X						✓						X	
DSI-2339	✓												X	

Euclid distances are available in Appendix E, Table E.2.

For the seasonality analysis applied for 5% of probability of exceedance, it is shown that the flow events up to this point are due to snowmelt. This analysis is applied for 10%, 20% and 30% probability exceedance also to see the different behavior of seasonality.

- **10% Probability of Exceedance Seasonality Analysis**

Fourteen stations have been used (from DSI and EIE) as in the previous analysis. The same steps have been performed. For each subbasin, the days on which discharge was smaller than  $Q_{90}$  were extracted over the period of record from the specific FDCs and transformed into Julian dates. Then,  $\bar{x}$  and  $\bar{y}$ ,  $\bar{\theta}$ , MD,  $\bar{r}$  and  $d_s^{ij}$  values are calculated respectively. The calculated indices are given in Table 4.3.

Table 4.3. Seasonality indices of stations for 10% probability of exceedance

Station	Probability of Exceedance	$\bar{x}$	$\bar{y}$	$\bar{\theta}$	MD	$\bar{r}$
EIE-2323	10%	-0.654	0.670	-0.80	136	0.94
EIE-2325	10%	-0.562	0.742	-0.92	129	0.93
EIE-2329	10%	-0.585	0.741	-0.90	130	0.94
DSI-2313	10%	-0.654	0.705	-0.82	135	0.96
DSI-2321	10%	-0.540	0.664	-0.89	131	0.86
DSI-2322	10%	-0.545	0.765	-0.95	127	0.94
DSI-2323	10%	-0.529	0.772	-0.97	126	0.94
DSI-2324	10%	-0.587	0.745	-0.90	130	0.95
DSI-2333	10%	-0.590	0.758	-0.91	130	0.96
DSI-2335	10%	-0.623	0.718	-0.86	133	0.95
DSI-2336	10%	-0.633	0.726	-0.85	133	0.96
DSI-2337	10%	-0.605	0.742	-0.89	131	0.96
DSI-2338	10%	-0.492	0.816	-1.03	123	0.95
DSI-2339	10%	-0.499	0.594	-0.87	132	0.78

Mean day of occurrences (MD column, Table 4.3) indicates that the flow events happened in spring, similar to the flows for 5% probability of exceedance. In other words, snowmelt is the main reason of the discharges for 10% probability of exceedance. For this analysis, seasonality spaces of stations have been placed in Appendix E (Figure E.2 – Figure E.15).

Figure E.2 – Figure E.15 are displaying the seasonality analysis results for discharges up to 10% probability of exceedance. By increasing the duration to 10% probability of exceedance, some changes have been observed. Station DSI-2339 has showed the more remarkable change; if Figure E.15 is examined it is seen that, at the 10% probability of exceedance, the flows are due to both the snowmelt and precipitation. The flow events draw a  $\frac{3}{4}$  circle in the seasonality space, which means precipitation is encountered in spring and autumn. Similarly, a change is noticed at station EIE-2323, autumn precipitations are available. The rest of the stations show a similar trend as 5% probability of exceedance. Although, the mean value of flow coordinates have been started to decrease, which means the seasonality has been started to spread in space.

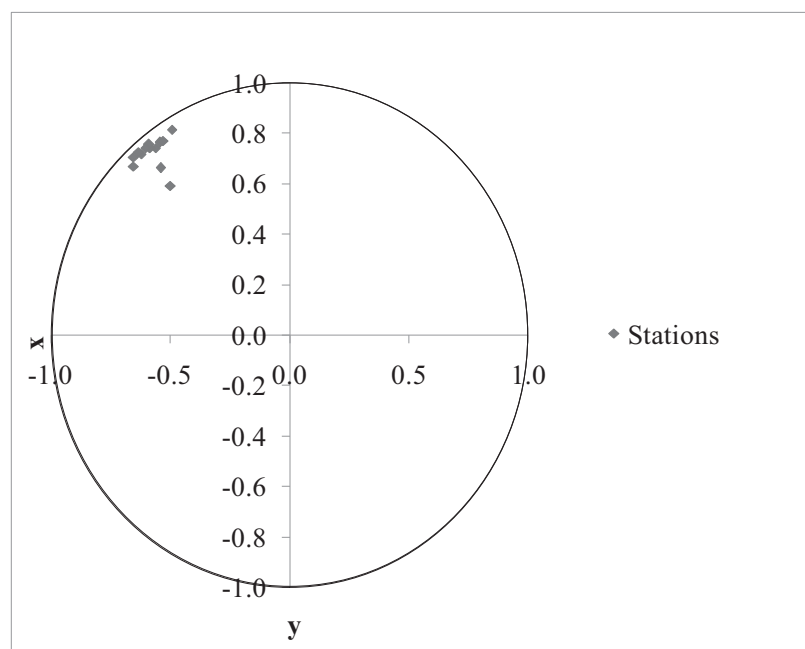


Figure 4.16. Stations on seasonality space for 10% probability of exceedance



As represented in Figure 4.16, the closely clustered stations in Figure 4.15 have been started to scatter on the seasonality space. It is because of; as probability of exceedance is increased, the regularity of flow tends to decrease.

Euclid distances of stations,  $d_s^{ij}$  values, are calculated for this exceedance. The similarity matrix has been tabulated in Table 4.4, and the necessary tables are available in Appendix E (Table E.3).

When the results are examined, it is observed that the DSI-2338 and DSI-2339 stations are most unlikely stations with the rest of the stations.

Table 4.4. Similarity matrix for the subbasins for 10% probability of exceedance

Basin	EIE-2323	EIE-2325	EIE-2329	DSI-2313	DSI-2321	DSI-2322	DSI-2323	DSI-2324	DSI-2333	DSI-2335	DSI-2336	DSI-2337	DSI-2338	DSI-2339
EIE-2323				✓									x	
EIE-2325			✓											x
EIE-2329								✓						x
DSI-2313											✓		x	
DSI-2321													x	✓
DSI-2322	x						✓							
DSI-2323						✓								x
DSI-2324									✓					x
DSI-2333								✓						x
DSI-2335											✓			x
DSI-2336										✓				x
DSI-2337			✓											x
DSI-2338	x						✓							
DSI-2339					✓								x	

- **20% Probability of Exceedance Seasonality Analysis**

For 20% probability of exceedance seasonality analysis, the days on which discharge was smaller than  $Q_{80}$  were extracted over the period of record from the specific FDCs and transformed into Julian dates. Then,  $\bar{x}$  and  $\bar{y}$ ,  $\bar{\theta}$ , MD,  $\bar{r}$  and  $d_s^{ij}$  values have been calculated respectively. The mean values of calculated indices of each station are given in Table 4.5.

Table 4.5. Seasonality indices of stations for 20% probability of exceedance

Station	Probability of Exceedance	$\bar{x}$	$\bar{y}$	$\bar{\theta}$	MD	$\bar{r}$
EIE-2323	20%	-0.654	0.604	2.40	139	0.89
EIE-2325	20%	-0.525	0.642	2.26	131	0.83
EIE-2329	20%	-0.577	0.652	2.30	133	0.87
DSI-2313	20%	-0.722	0.549	2.49	145	0.91
DSI-2321	20%	-0.042	0.470	1.66	96	0.47
DSI-2322	20%	-0.128	0.220	2.10	122	0.25
DSI-2323	20%	-0.530	0.692	2.22	129	0.87
DSI-2324	20%	-0.575	0.683	2.27	132	0.89
DSI-2333	20%	-0.659	0.670	2.35	136	0.94
DSI-2335	20%	-0.598	0.635	2.33	135	0.87
DSI-2336	20%	-0.665	0.640	2.38	138	0.92
DSI-2337	20%	-0.621	0.665	2.32	135	0.91
DSI-2338	20%	-0.537	0.704	2.22	129	0.88
DSI-2339	20%	-0.411	0.439	2.32	135	0.60

MD values tell us that the flow dates have been widening to the early April (day 96 belongs to 6 April) and late May (day 145 belongs to 25 May). The  $\bar{r}$  values have been started to decrease since irregular flows started to be observed.

Figure E.16 – Figure E.29 are displaying the seasonality analysis results for discharges up to 20% probability of exceedance. Except stations DSI-2313, DSI-2333, DSI-2336, DSI-2337 and DSI-2338, all stations show flow events in all seasons including summer. Because of this scattering, the mean r-values have been decreased drastically in many stations.

Figure 4.17 shows that the station cloud is moving downwards, and more stations are scattered around unit seasonality circle. It is due to the diversity of the flow events in time and it is an expected condition.

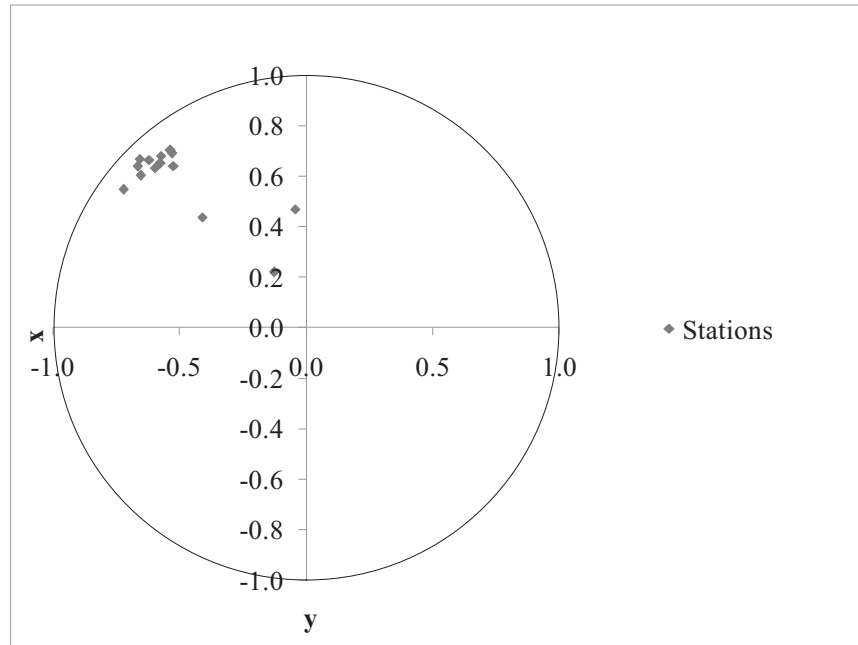


Figure 4.17. Stations on seasonality space for 20% probability of exceedance

Similarity measures,  $d_s^{ij}$ , are calculated for 20% probability of exceedance. As the value of  $d_s^{ij}$  gets smaller, the similarity of the basins gets stronger. The similarity matrix is shown below give an idea about the similarities of basins and the Euclid distance values are provided in Appendix E (see Table E.4).

Table 4.6. Similarity matrix for the subbasins for 20% probability of exceedance

Basin	EIE-2323	EIE-2325	EIE-2329	DSI-2313	DSI-2321	DSI-2322	DSI-2323	DSI-2324	DSI-2333	DSI-2335	DSI-2336	DSI-2337	DSI-2338	DSI-2339
EIE-2323						x					✓			
EIE-2325						x	✓							
EIE-2329						x						✓		
DSI-2313	✓					x								
DSI-2321				x		✓								
DSI-2322					✓				x					
DSI-2323						x		✓						
DSI-2324			✓			x								
DSI-2333						x					✓			
DSI-2335			✓			x								
DSI-2336						x			✓					
DSI-2337						x				✓				
DSI-2338						x	✓							
DSI-2339		✓			x									

- **30% Probability of Exceedance Seasonality Analysis**

For 30% probability of exceedance seasonality analysis, the days on which discharge were smaller than  $Q_{70}$  were extracted over the period of record from the specific FDCs and transformed into Julian dates. Then,  $\bar{x}$  and  $\bar{y}$ ,  $\bar{\theta}$ , MD,  $\bar{r}$  and  $d_s^{ij}$  values are calculated respectively. The mean values of calculated indices of each station are in Table 4.7.

Table 4.7. Seasonality indices of stations for 30% probability of exceedance

Station	Probability of Exceedance	$\bar{x}$	$\bar{y}$	$\bar{\theta}$	MD	$\bar{r}$
EIE-2323	30%	-0.602	0.464	2.48	144	0.76
EIE-2325	30%	-0.363	0.433	2.27	132	0.56
EIE-2329	30%	-0.462	0.466	2.35	137	0.66
DSI-2313	30%	-0.736	0.353	2.69	157	0.82
DSI-2321	30%	0.130	0.367	1.23	72	0.39
DSI-2322	30%	-0.172	0.385	1.99	116	0.42
DSI-2323	30%	-0.410	0.493	2.26	131	0.64
DSI-2324	30%	-0.457	0.510	2.30	134	0.68
DSI-2333	30%	-0.664	0.504	2.49	145	0.83
DSI-2335	30%	-0.460	0.463	2.35	137	0.65
DSI-2336	30%	-0.693	0.451	2.56	149	0.83
DSI-2337	30%	-0.486	0.545	2.30	134	0.73
DSI-2338	30%	-0.533	0.583	2.31	134	0.79
DSI-2339	30%	-0.407	0.391	2.38	138	0.56

The flow events have spread to the March (day 72 belongs to 13 March) and June (day 157 belongs to 6 June) (see Figure 4.18). Since probability of exceedance is high, the widening of flow events makes sense. Because of the same reasons the  $\bar{r}$  values have been also decreased, and got far from unity.

In this analysis, it is realized that the record length of the stations is important from seasonality analysis point of view. This can be explained by looking at stations EIE-2323 and DSI-2313 (see Figures E.30 and E.33). The seasonality space of EIE-2323 completes a perfect circle whereas spring and autumn flow events are observed for DSI-2313.

If the study area is considered, it is seen that EIE-2323 and DSI-2313 are very close to each other. Both of the stations are on the exit of the Oltu Basin. Therefore, for any seasonality analysis, record length of streamgauges should be taken into account.

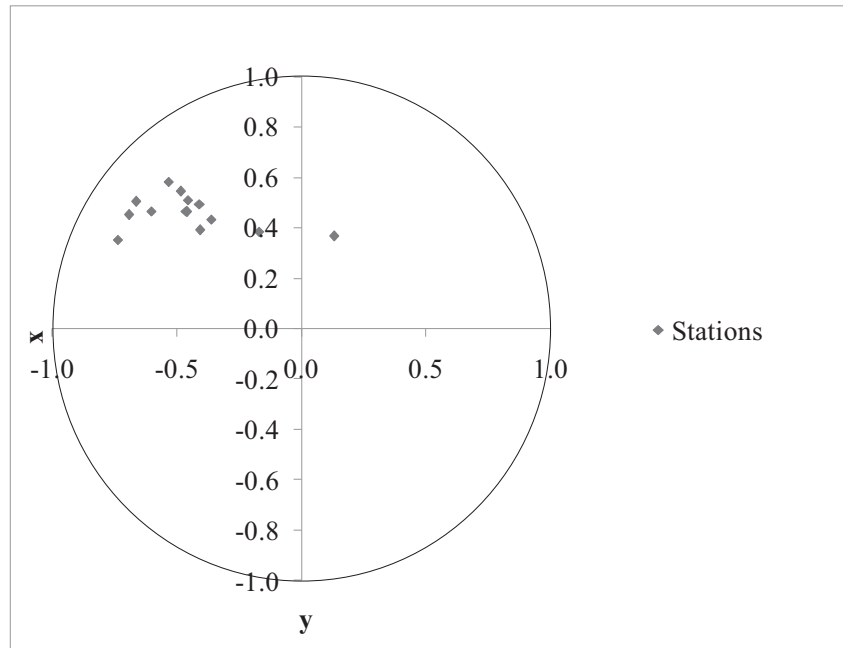


Figure 4.18. Stations on seasonality space for 30% probability of exceedance

Stations have been diverged from each other and r-values get smaller in seasonality space, so they come closer to the X and Y axes.

## CHAPTER 5

### REGIONALIZATION OF FDCS

As explained in Methodology chapter, (Section 2.3), in the regionalization of FDCs of Oltu Basin, parametric and statistical approaches have been considered and applied. Perimeter, area, length of main river, minimum elevation, mean elevation, maximum elevation, basin relief, slope, aspect, mean annual precipitation, mean annual temperature, CN parameters have been used as basin and climatological parameters in the analysis. Before regionalization analysis, in order to investigate the strength of relation between predictor (basin parameters suggested in the previous chapter) and criterion variable (discharge parameter) a *Correlation Analysis* is needed.

Moreover, the correlation analysis and the scatter plots of predictor variables against criterion variable is unfortunately the only tool while selecting the *Principle Parameters* (parameters that represent the flow behavior best), because the sample size (data gathered) is less than the sample size that would be needed in a parameter selection analysis, such as Principle Component Analysis. In other words, degrees of freedom are insufficient to make a Principle Component Analysis, a Stepwise Regression Analysis or a Principle Component Regression Analysis.



It is suggested that to use minimum 15 samples for each predictor variables (Hohenberger, 2009), which means minimum 180 samples (station data) should be necessary for this analysis. However, only 14 samples (station data) are available for this case. In brief, correlation analysis, scatter diagrams, theoretical literature and background knowledge has been used to eliminate useless predictors. After this elimination process, statistical and parametric approaches have been applied to develop a regional model for FDCs of Oltu Basin.

## 5.1. Descriptive Statistics of the Parameters

In many statistics courses and books, it is suggested to analyze descriptive statistics before any analysis. The histograms and statistics such as mean, median, skewness are the basic descriptive statistics. The histograms give information about the shape of the general distribution such as normal or Gumbel and the symmetry of this distribution (or the skewness). McCuen (1993) noted:

A decay function shape occurs when the values below the mean are (1) more numerous than those above the mean and (2) closer to the mean. These characteristics are important in determining where models developed with the data are applicable; a model should be used cautiously outside the range of the data used to calibrate it.

In addition to information obtained by histograms, descriptive statistics also give basic information about data and enables managing large sets of data. Moreover, the graphical analysis helps us to identify extreme events, form the relationship between any two variables and observe the type of this relation (McCuen, 1993). A systematic manner should be observed between variables to study correlation between those variables. However, there is a drawback of *Graphical Analysis*. If the sample data range is smaller than the expected data range of population, it may cause an unstable relation between variables. This directly affects the model and the predicted values. Thus, the data range should be examined carefully and the model range should be set to avoid faulty predictions.

The *Statistical Package for the Social Sciences* (SPSS) is used as the statistical tool to get the descriptive statistics and to draw the graphs of each descriptive parameter (P, A, LMR, etc.) against criterion parameter (Q).

In addition to descriptive statistics analysis performed by SPSS, another commercially available program EasyFit–Distribution Fitting Software is used to analyze the distributions of parameters. It is important to know the distribution of any data to understand its nature. Especially for hydrologic variables, normal distribution is not possible, but in many statistical analyses, normal distribution is accepted as the valid distribution.

The program automatically fits 55 probability distributions (Appendix G) to the data and ranks the distribution models from best to worst. Moreover, the interactive figures, namely probability density function (PDF), cumulative distribution function (CDF), survival function, hazard function (failure rate), cumulative hazard function, P-P plot, Q-Q plot and probability difference graphs are available for any type of study. Furthermore, Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared tests are used as the goodness of fit (GOFs) tests. These tests help to decide on the models, which best fit the data in the preferred confidence interval ("EasyFit - distribution fitting software", 2011 ).

In this study, all the 55 probability distributions are tested, ranked and examined by taking into account the GOF Test Anderson-Darling, since;

- Kolmogorov-Smirnov tends to be more sensitive near the center of the distribution than at the tails. In the study, most of the parameters seem to be tailed.
- Chi-Squared test is used for the binned data and the test statistics depend on how the data binned ("NIST/SEMATECH e-handbook of statistical methods", 2010).

In the following sections, the descriptive statistics of descriptor parameters have been explored. The histograms of the parameters and the PDF graph of the data are given in Appendix F.

Before that average, minimum and maximum values of the parameters have been explored and shown in Table 5.1.

Table 5.1. Minimum, average and maximum values of climatic and geomorphologic characteristics of 9 basins used in the regression analysis

	P	A	LMR	H <sub>max</sub>	H <sub>min</sub>	H <sub>mean</sub>	BR	S	ASPCT.	MAP	CN	T	Q	Duration
	km	km <sup>2</sup>	m	m	m	m	m	%	°	mm	-	°C	m <sup>3</sup> /s	yr
Minimum	17.1	10.6	2.1	2755.6	520.0	2006.1	113.9	15.4	170.8	423.6	72.2	6.8	0.2	10.0
Average	361.9	3553.2	96.6	3042.7	1167.3	2157.2	994.3	21.8	180.6	511.4	77.3	7.1	17.1	27.8
Maximum	576.8	6978.7	161.2	3222.0	2510.9	2624.8	1513.1	30.2	196.3	563.5	89.0	7.5	34.3	41.0

### 5.1.1. Perimeter (P)

Computing descriptive statistics without a scatter plot causes misinterpretations of data. The scatter plot of perimeter (P) against mean annual discharge (Q) (Figure 5.1) shows that there is a strong nonlinear relation between these variables. Probably the correlation coefficient will be high. Moreover, there are no outliers in the range of the data, which means the mean would not be affected and the mean and median could be used to the balance point of data (Helsel and Hirsh, 1992). The range of the data is quite large to make a projection beyond the range of sample data and there would not be any stability problems to worry about (McCuen, 1993).

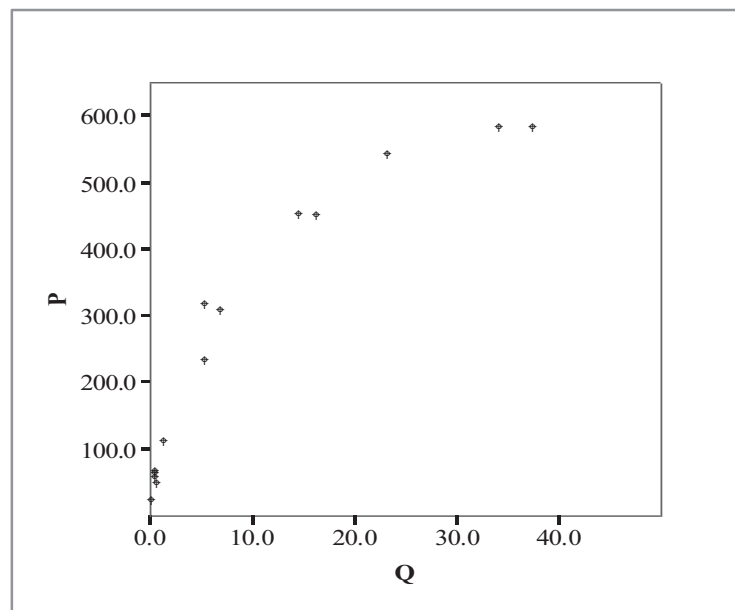


Figure 5.1. Scatter plot of “Perimeter (km)” against “Mean Annual Discharge ( $\text{m}^3/\text{s}$ )”

When Table 5.2 is examined, it is seen that the mean and the median, which are the measures of central tendency, are close. These statistics give an idea about the population mean. It is mentioned earlier that there are no outliers in the relation of P and Q, which means the observations influence the mean value almost the same (Helsel and Hirsh, 1992). Standard deviation and the variance are high, which are the measures of spread. Moreover,

the data slightly skewed to the right of mean with the value of 0.6. When the kurtosis is examined a flat distribution is expected due to the negative sign where  $K = -1.7$ . Furthermore, since the value is small it could be concluded that there is a flat distribution near the mean value.

Table 5.2. Descriptive statistics of “Perimeter”

Perimeter (km)		
N (sample size)	Valid	14
	Missing	0
Mean		267.1
Std. Error of Mean		57.7
Median		263.7
Mode		17.1
Std. Deviation		215.9
Variance		46609.2
Skewness		0.3
Std. Error of Skewness		0.6
Kurtosis		-1.7
Std. Error of Kurtosis		1.2
Range		559.7
Minimum		17.1
Maximum		576.8
Percentiles	25	54.5
	50	263.7
	75	467.2

In many statistical analysis and tests a normal distribution is which for the data, this is also valid for water resources data. However, in reality the data are neither symmetric nor normally distributed. In other words, while doing analysis this should be kept in mind as a misleading factor. When Figure F.1 (Appendix F) is inspected, it could be seen that the descriptive statistics are rational for the data. P is not normally distributed, it is slightly skewed to the right, and the data is flat near the mean value and the mean value is representative for the data.

The program EasyFit selected the best distribution model (PDF) as Generalized Pareto according to the Anderson-Darling GOF test (see Figure F.2, Appendix F). Moreover,

normal distribution is in the tenth order in the ranking. If equilibrium is found in the distribution of the "small" to the "large" Generalized Pareto is preferred (Weisstein, 2010).

### 5.1.2. Area (A)

The scatter plot, A ( $\text{km}^2$ ) against Q ( $\text{m}^3/\text{s}$ ), shows a strong linear relationship. There are no outliers, but the range of data is high. There are many values under  $100 \text{ km}^2$  and a few above  $6000 \text{ km}^2$ , in other words there is a small clustering for small values in the data set. This may cause under/over estimates for the developed model and some restrictions may be needed for clustered part.

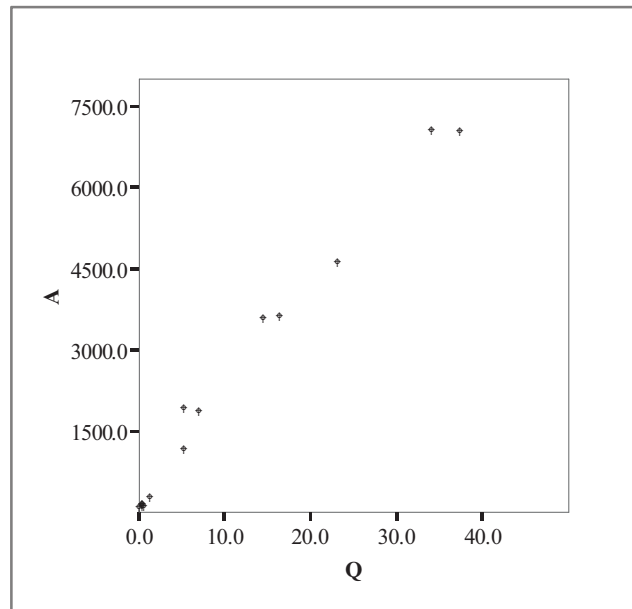


Figure 5.2. Scatter plot of “Area ( $\text{km}^2$ )” against “Mean Annual Discharge  $\text{m}^3/\text{s}$ ”

For the descriptive statistics of Area, Table 5.3, mean and the median are not close and very different from each other. The descriptive statistics mean is easily influenced by the large values, as in this case. Median will be more representative since it is resistant to change in value of observations and outliers (Helsel and Hirsh, 1992).

The measure of variations, variance and standard deviation are high as expected from the scatter plot analysis. The data is skewed positively as in P, but this time it is less peaked near the mean.

Table 5.3. Descriptive statistics of “Area”

Area (km <sup>2</sup> )		
N (sample size)	Valid	14
	Missing	0
Mean		2192.8
Std. Error of Mean		676.0
Median		1439.9
Mode		10.6
Std. Deviation		2529.3
Variance		6397206.4
Skewness		1.0
Std. Error of Skewness		0.6
Kurtosis		-0.2
Std. Error of Kurtosis		1.2
Range		6968.1
Minimum		10.6
Maximum		6978.7
Percentiles	25	63.6
	50	1439.9
	75	3787.2

The histogram of parameter Area seems that it is not normally distributed (see Figure F.3). The interpretations of the descriptive statistics are meaningful. According to the fitting distribution analysis, the most appropriate distribution came out to be as Dagum distribution (Figure F.4). Moreover, Lognormal and Normal distributions are ranked in the 16<sup>th</sup> and 23<sup>th</sup> order, respectively.

### 5.1.3. Length of Main River (LMR)

The relationship between Length of Main River (km) and Discharge ( $\text{m}^3/\text{s}$ ) is strong and positive nonlinear relationship (see Figure 5.3). There are some impurities disturbing the perfect relationship but it would not create a problem during modeling process. Since small catchments are more than the larger ones, and shorter river lengths exist more than the longer ones as expected. The range seems reasonable for any type of modeling.

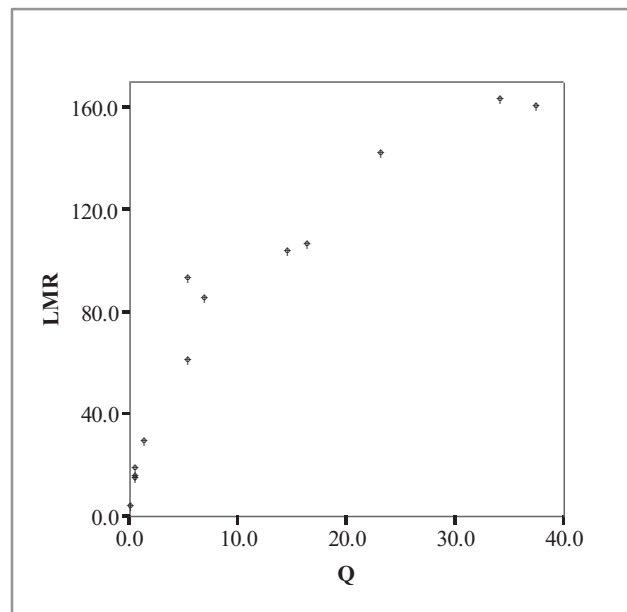


Figure 5.3. Scatter plot of “Length of Main River (km)” against “Mean Annual Discharge ( $\text{m}^3/\text{s}$ )”

The central tendency measures mean and median are close, it is expected since there is no extreme cases or outliers in data set (see Figure 5.3, no outliers were observed). Since the range of the basins is large ( $10 \text{ km}^2 - 7000 \text{ km}^2$ ) the range of any parameter is large. Therefore, standard deviation and variance for the parameters are high. The data again slightly skewed to the right but it could be accepted as approximately symmetric since the value is between -0.5 and 0.5. Furthermore, since the standard error of skewness is between -2 and 2, any conclusions about population data could not be obtained; it could be symmetric or skewed in either ways (Brown, 2010). Moreover, near the mean value, a peak



is expected and since the kurtosis below the critical number is 3. A normal distribution has a kurtosis exactly 3, for the kurtosis value -1.4, a lower and broader central peak should be existed in the distribution (Brown, 2010). Interpretation of standard error of kurtosis is similar to interpretation of standard error of skewness. Between -2 and 2 any information could not be inferred about the population.

Table 5.4. Descriptive statistics of “Length of Main River”

LMR (km)		
N (sample size)	Valid	14
	Missing	0
Mean		69.3
Std. Error of Mean		15.7
Median		71.3
Mode		0.0
Std. Deviation		58.6
Variance		3433.0
Skewness		0.3
Std. Error of Skewness		0.6
Kurtosis		-1.4
Std. Error of Kurtosis		1.2
Range		161.2
Minimum		0.0
Maximum		161.2
Percentiles	25	13.2
	50	71.3
	75	113.3

The histogram has the peak at the first class interval. For the best description of the data, probability density function analysis results are examined. It is suggested to use Johnson SB probability distribution for the Length of Main River data set (Figure F.6).

#### 5.1.4 Maximum Elevation ( $H_{\max}$ )

There is a strong positive nonlinear relationship between Maximum Elevation (m),  $H_{\max}$  (m), and Discharge ( $Q$ ,  $\text{m}^3/\text{s}$ ) (see Figure 5.4). The range seems quite reasonable since the study area is in a mountainous area.

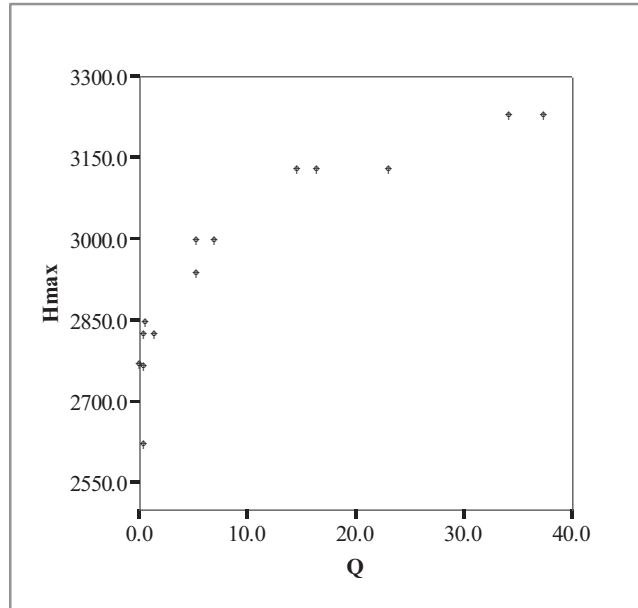


Figure 5.4. Scatter plot of “Maximum Elevation (m)” against “Mean Annual Discharge ( $\text{m}^3/\text{s}$ )”

Median and the mean values are approximately the same, so mean value could be used as a representative value for  $H_{\max}$ . Standard deviation and variance follow the same pattern with other basin parameters. The data skewed to the left, in other words the left tail is a bit longer. Since the skewness value is again between -0.5 and 0.5, it is nearly symmetric. To get a conclusion about not for this data set but for a population, the standard error of skewness should be checked. The value is 0.6, which is below 2 and that does not give any significant t information about the population’s distribution. In addition to skewness, kurtosis is negative (flat distribution) and it is not possible to get any predictions about population kurtosis by using standard error of kurtosis.

Table 5.5. Descriptive statistics of “maximum elevation”

$H_{\max}$ (m)		
N (sample size)	Valid	14
	Missing	0
Mean		2950.3
Std. Error of Mean		51.1
Median		2959.0
Mode		3120.0
Std. Deviation		191.3
Variance		36597.9
Skewness		-0.1
Std. Error of Skewness		0.6
Kurtosis		-1.1
Std. Error of Kurtosis		1.2
Range		608.4
Minimum		2613.6
Maximum		3222.0
Percentiles	25	2800.0
	50	2959.0
	75	3120.0

As shown in the Figure F.8 the suitable distribution for the Maximum Elevation is Johnson SB as for LMR. Normal distribution is ranked in the 7<sup>th</sup> order.

### 5.1.5. Minimum Elevation ( $H_{\min}$ )

The Minimum Elevation ( $H_{\min}$ , m) and  $Q$  ( $\text{m}^3/\text{s}$ ) has a negative, almost strong and nonlinear relationship. No outliers are detected by inspection of the scatter plot. However, there are some disturbances, which do not let a strong relationship. Moreover, the range is too wide and it will affect the central tendency measures.

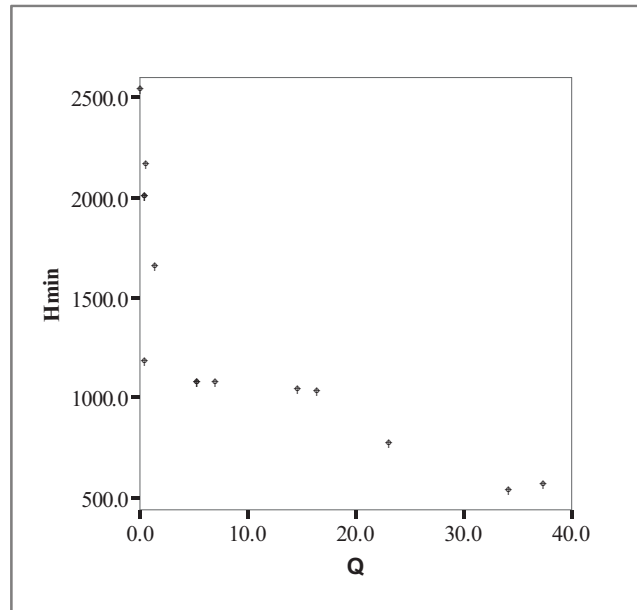


Figure 5.5. Scatter plot of “Minimum Elevation in m” against “Mean Annual Discharge in  $\text{m}^3/\text{s}$ ”

Median and the mean values are approximately the same, so mean value could be used as a representative value for  $H_{\max}$ . Standard deviation and variance follow the same pattern with other basin parameters. The data skewed to the left, in other words the left tail is a bit longer. Since the skewness value is again between -0.5 and 0.5, it is nearly symmetric. To get a conclusion about not for this data set but for a population, the standard error of skewness should be checked. The value is 0.6, which is below 2 and that does not give any significant t information about the populations' distribution.

In addition to skewness, kurtosis is negative (flat distribution) and it is not possible to get any predictions about population kurtosis by using standard error of kurtosis.

Table 5.5. Descriptive statistics of “maximum elevation”

$H_{\max}$ (m)		
N (sample size)	Valid	14
	Missing	0
Mean		2950.3
Std. Error of Mean		51.1
Median		2959.0
Mode		3120.0
Std. Deviation		191.3
Variance		36597.9
Skewness		-0.1
Std. Error of Skewness		0.6
Kurtosis		-1.1
Std. Error of Kurtosis		1.2
Range		608.4
Minimum		2613.6
Maximum		3222.0
Percentiles	25	2800.0
	50	2959.0
	75	3120.0

As shown in the Figure F.7 the suitable distribution for the Maximum Elevation is Johnson SB as for LMR. Normal distribution is ranked in the 7<sup>th</sup> order.

#### 5.1.6. Minimum Elevation ( $H_{\min}$ )

The Minimum Elevation ( $H_{\min}$ , m) and  $Q$  ( $\text{m}^3/\text{s}$ ) has a negative, almost strong and nonlinear relationship. No outliers are detected by inspection of the scatter plot. However, there are some disturbances, which do not let a strong relationship. Moreover, the range is too wide and it will affect the central tendency measures.

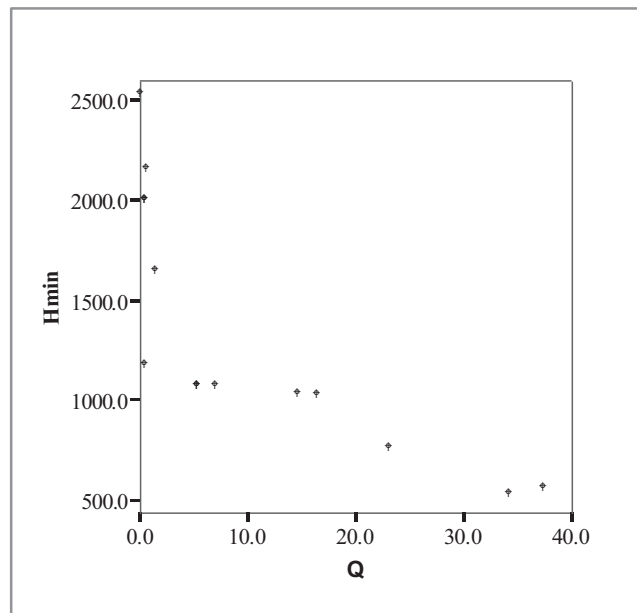


Figure 5.5. Scatter plot of “Minimum Elevation” against “Mean Annual Discharge”

As expected, the mean and median are not close and it is logical to use the resistant measure, namely, the median as a central value. In addition to this, the standard error of the mean for  $H_{\min}$  is higher compared to the one for  $H_{\max}$ . Skewness is positive and between 0.5 and 1, which means a moderately skewed data, a long right tail is expected (Brown, 2010). On the other hand, the standard error of skewness is insufficient to get a conclusion for population data. The kurtosis statistics are similar to the one for  $H_{\max}$ .

Figure F.9 shows that, as estimated before, the distribution has a long right tail. The peak is close to median value. Moreover, the goodness of fit test, Anderson-Darling, is the appropriate test for the tailed data as mentioned at the beginning. The distribution that explains the behavior of data is presented in Figure F.10. Anderson-Darling recommends Generalized Pareto for the distribution, but the 3<sup>rd</sup> suitable distribution is Log-Pearson and the 10<sup>th</sup> one is Log Normal.

Table 5.6. Descriptive statistics of “Minimum Elevation”

$H_{\min}$ (m)		
N (sample size)	Valid	14
	Missing	0
Mean		1315.4
Std. Error of Mean		166.4
Median		1055.2
Mode		1055.2
Std. Deviation		622.5
Variance		387445.4
Skewness		0.6
Std. Error of Skewness		0.6
Kurtosis		-0.8
Std. Error of Kurtosis		1.2
Range		1990.9
Minimum		520.0
Maximum		2510.9
Percentiles	25	945.0
	50	1055.2
	75	1980.0

#### 5.1.7. Mean Elevation ( $H_{\text{mean}}$ )

Mean Elevation ( $H_{\text{mean}}$ , m) and  $Q$  ( $\text{m}^3/\text{s}$ ) has a nonlinear relationship but it is not perfect or strong. There are some disturbances in this relationship. Nevertheless, the range is not wide, which minimizes standard error of the mean. A higher value is inspected, point A in Figure 5.6, but it could not be an outlier, but it could be an extreme point, since it follows the pattern.

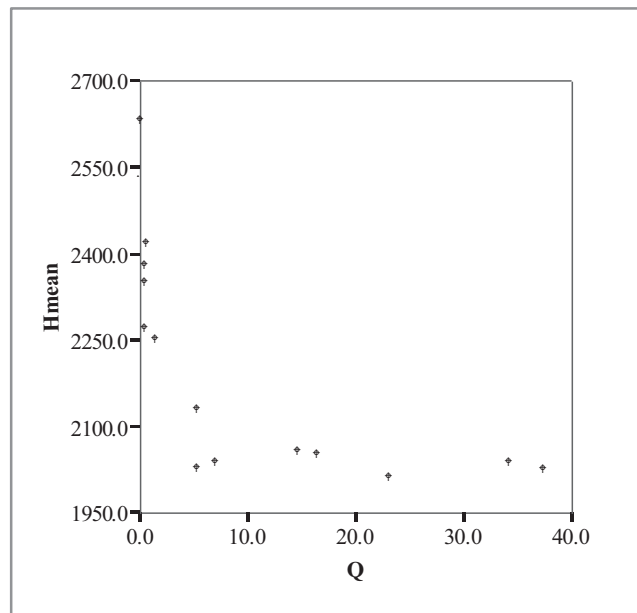


Figure 5.6. Scatter plot of “Mean Elevation (m)” against “Mean Annual Discharge in m³/s”

Median is appropriate to use as a balance value. Skewness tells the tail of distribution is to the right and long, the data is not symmetric. Kurtosis tells that there is no sharp peak near the mean value.

The proposed best distribution is Fatigue Life (3P) according to the Anderson-Darling GOF test. This distribution is very similar to the Lognormal (3P), the second appropriate one in the analysis.

#### 5.1.8. Basin Relief (BR)

The scatter plot (Figure 5.7) shows a positive, nonlinear relationship between the variables. However, this relationship is not as strong as the previous ones. The range is so wide; this should be kept in mind during modeling process.



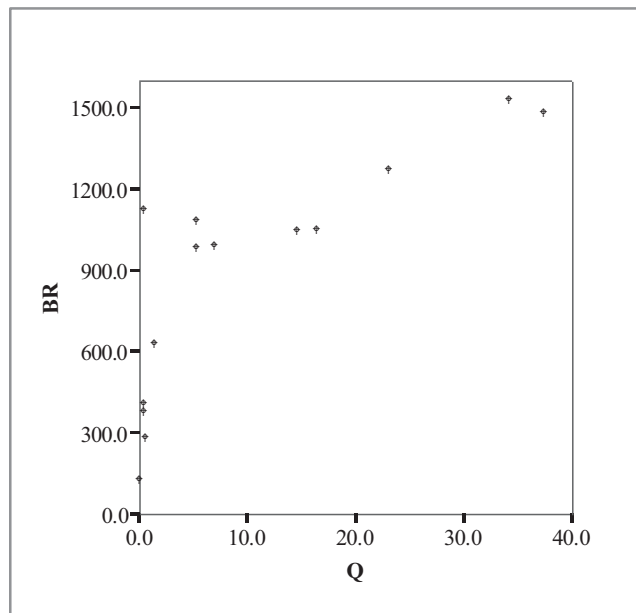


Figure 5.7. Scatter plot of “Basin Relief (m)” against “Mean Annual Discharge (m<sup>3</sup>/s)”

The range of the data had been affected the mean value, so it is not healthy to use mean as a representative value. The median value will be appropriate. Skewness is negative and under 0.5, which means the distribution skewed to the left and it is almost symmetric. Moreover, the negative kurtosis implies that relatively flat distribution and a peak near the central value.

Johnson SB is recommended distribution according to the Anderson-Darling. Furthermore, Normal, Lognormal and Uniform distributions are placed in the first ten-distribution list.

Table 5.8. Descriptive statistics of “Basin Relief”

BR (m)		
N (sample size)	Valid	14
	Missing	0
Mean		869.8
Std. Error of Mean		118.8
Median		1003.1
Mode		113.9
Std. Deviation		444.4
Variance		197532.4
Skewness		-0.3
Std. Error of Skewness		0.6
Kurtosis		-1.0
Std. Error of Kurtosis		1.2
Range		1399.2
Minimum		113.9
Maximum		1513.1
Percentiles	25	385.1
	50	1003.1
	75	1144.0

#### 5.1.9. Slope (S)

Slope has nonlinear relationship, which is not so clear, but needs attention during correlation analysis. Point B (Figure 5.8) is an outlier since it does not follow the trend of other observations. However, it should not be removed from the data set (Helsel and Hirsh, 1992).

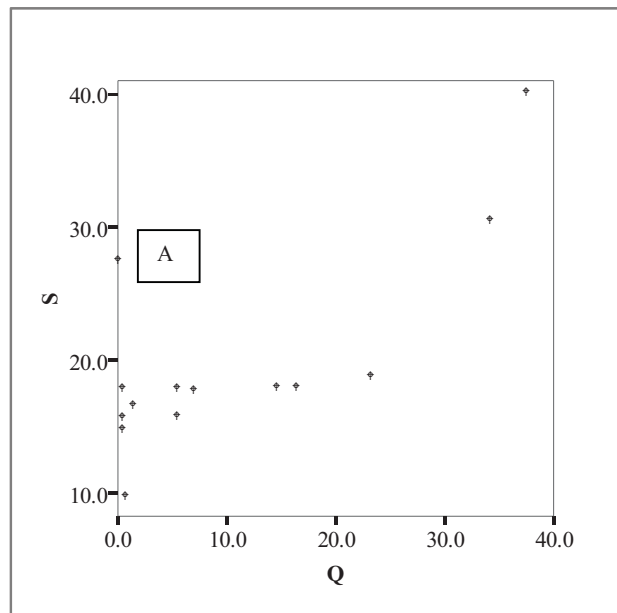


Figure 5.8. Scatter plot of “Slope (%)” against “Mean Annual Discharge (m<sup>3</sup>/s)”

Mean value is suitable to use as a balance point for Slope data. Skewness is very high compared to other variables skewness values. If the skewness is higher than 1 it is accepted as highly skewed data (Brown, 2010). Kurtosis is close to 3, which is the number for normal distribution, but probably there exist a peak and the other parts are flat.

Wakeby probability distribution has been selected as the best appropriate distribution by program Easy-Fit. Lognormal distribution is listed in the 24<sup>th</sup> order in the list, and normal distribution is ranked in the 39<sup>th</sup> order.

#### 5.1.10. Aspect (ASPCT.)

Without considering the points C and D, a nonlinear relationship, between the values 160 and 180, could be proposed. However, this relation seems not very significant. Therefore, the correlation coefficients will judge the importance of the variable.

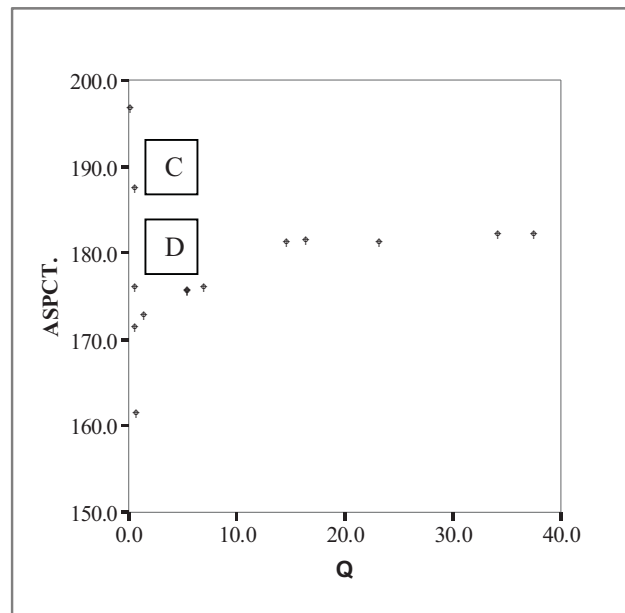


Figure 5.9. Scatter plot of “Aspect in °” against “Mean Annual Discharge in m<sup>3</sup>/s”

Mean and median values come out to be the same, which is unexpected because of the outlier values C and D (see Figure 5.9). Thus, the histogram plays an important role to define the behavior of the data set. Skewness is very small which means the Aspect data is nearly symmetric. The positive kurtosis implies a peak near mean value.

As supposed, there is a peak near the mean value and distribution is very close to normal distribution (see Figure F.17), but the difference between frequencies is too high. Hypersecant probability distribution is proposed as the best-fitted distribution for Aspect data. Lognormal distribution is placed in the 11<sup>th</sup> order.

Table 5.9. Descriptive statistics of “Slope”

Slope (%)		
N (sample size)	Valid	14
	Missing	0
Mean		19.7
Std. Error of Mean		2.1
Median		17.6
Mode		9.5
Std. Deviation		7.7
Variance		59.4
Skewness		1.6
Std. Error of Skewness		0.6
Kurtosis		2.8
Std. Error of Kurtosis		1.2
Range		30.3
Minimum		9.5
Maximum		39.8
Percentiles	25	15.5
	50	17.6
	75	20.7

#### 5.1.11. Mean Annual Precipitation (MAP)

Excluding the point E, there is an obvious positive nonlinear relationship between Mean Annual Precipitation (MAP, mm) and Discharge. Point E, in Figure 5.10, could be an extreme event since it is from a natural phenomenon. Therefore, it would not be excluded from data set as suggested in literature.

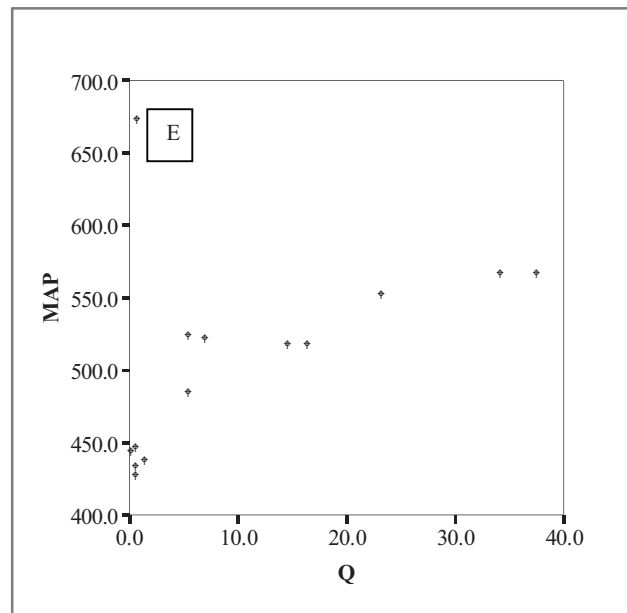


Figure 5.10. Scatter plot of “Mean Annual Precipitation (mm) ” against “Mean Annual Discharge ( $\text{m}^3/\text{s}$ )”

Mean value is close enough to the median value to use as a central tendency measure. Skewness is very close to the critical value 1, therefore it could be accepted as a highly skewed data. The data skewed to the right, right tail is longer. The positive kurtosis number indicates a peak value in the distribution.

The distribution has two peaks and histogram is tailed (see Figure F.19). For Mean Annual Precipitation the best probability distribution is Wakeby probability distribution (see Figure F.20). Generalized Pareto and Johnson SB are the 2<sup>nd</sup> and the 3<sup>rd</sup> best distributions. Moreover, lognormal and normal distributions are placed in the 11<sup>th</sup> and 15<sup>th</sup> order.

Table 5.11. Descriptive statistics of “Mean Annual Precipitation”

Mean Annual Precipitation (mm)		
N (sample size)	Valid	14
	Missing	0
Mean		504.9
Std. Error of Mean		18.5
Median		514.8
Mode		423.6
Std. Deviation		69.1
Variance		4777.7
Skewness		0.9
Std. Error of Skewness		0.6
Kurtosis		0.9
Std. Error of Kurtosis		1.2
Range		246.0
Minimum		423.6
Maximum		669.6
Percentiles	25	439.3
	50	514.8
	75	552.8

#### 5.1.12. Mean Annual Temperature (T)

Any significant relationship can be inferred from the scatter plot of Temperature (T, °C) with respect to Discharge. In other words, there is no relationship.

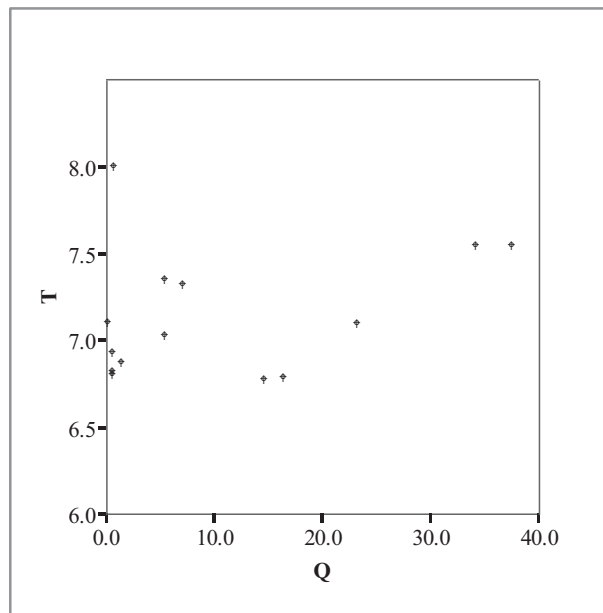


Figure 5.11. Scatter plot of “Temperature (°C)” against “Mean Annual Discharge (m³/s)”

Mean and median are good representatives for the temperature data (see Table 5.11). Standard deviation and variance are small. Data is skewed to the right and has a peak.

Lognormal distribution seems quite appropriate for Temperature data according to Figure F.21, but the confidence intervals are significant to make a conclusion. Johnson SB fits the parameter Temperature’s nature as it does in many parameters analyzed before (see Figure F.22).



Table 5.12. Descriptive statistics of “temperature”

Temperature (°C)		
N (sample size)	Valid	14
	Missing	0
Mean		7.1
Std. Error of Mean		0.1
Median		7.0
Mode		7.5
Std. Deviation		0.4
Variance		0.1
Skewness		1.1
Std. Error of Skewness		0.6
Kurtosis		0.6
Std. Error of Kurtosis		1.2
Range		1.2
Minimum		6.8
Maximum		8.0
Percentiles	25	6.8
	50	7.0
	75	7.4

#### 5.1.13. Curve Number (CN)

The Curve Number and Discharge has a strong negative, nonlinear relationship without considering the *Cluster F*.

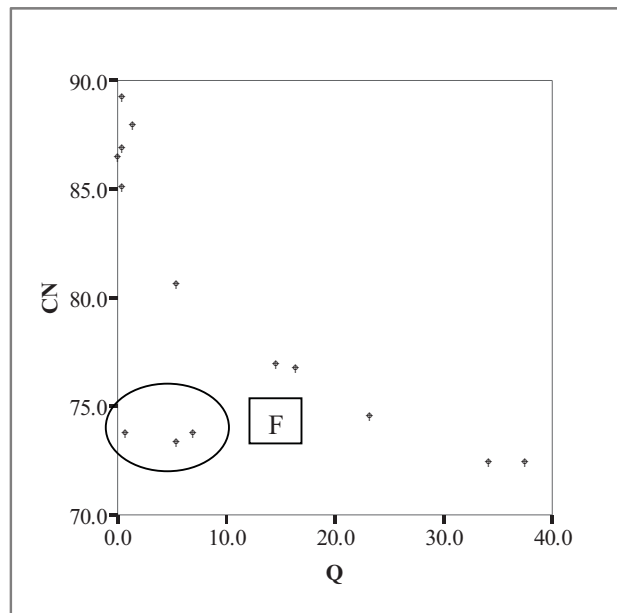


Figure 5.12. Scatter plot of “Curve Number” against “Mean Annual Discharge ( $\text{m}^3/\text{s}$ )”

Mean value is proper to represent the central tendency of data. Data is slightly skewed to the right but not symmetric, and it is flat near the mean value. The range seems small but it is the natural tendency of curve number, as mentioned before it takes values between 0 and 100. The value 100 corresponds to full surface runoff.

The *Cluster F* (in Figure 5.12) creates the peaked part. The rest of the distribution is flat. The probability density function (Figure F. 24) is Wakeby. Normal distribution is placed in the 23<sup>th</sup> order in the list.

Table 5.13. Descriptive statistics of “Curve Number”

CN		
N (sample size)	Valid	14
	Missing	0
Mean		79.1
Std. Error of Mean		1.7
Median		76.6
Mode		72.2
Std. Deviation		6.5
Variance		41.7
Skewness		0.4
Std. Error of Skewness		0.6
Kurtosis		-1.7
Std. Error of Kurtosis		1.2
Range		16.8
Minimum		72.2
Maximum		89.0
Percentiles	25	73.4
	50	76.6
	75	86.3

#### 5.1.14. Mean Annual Discharge (Q)

Mean and median are very different in values, so it is appropriate to use median as the balance value. The range, standard deviation and variance are very large, and this explains the contradiction between the mean and the median. The data is skewed to the right and has a small peak (rest is flat).

Table 5.14. Descriptive statistics of “Mean Annual Discharge”

Discharge (m <sup>3</sup> /s)		
N (sample size)	Valid	14
	Missing	0
Mean		10.6
Std. Error of Mean		3.4
Median		5.5
Mode		0.6
Std. Deviation		12.9
Variance		165.8
Skewness		1.2
Std. Error of Skewness		0.6
Kurtosis		0.3
Std. Error of Kurtosis		1.2
Range		37.4
Minimum		0.2
Maximum		37.6
Percentiles	25	0.6
	50	5.5
	75	18.2

Figure F.25 shows that there is an accumulation in the first class interval, namely between 0 and 10. This causes the peak, and the rest of histogram is nearly uniform but this peak will change distribution from uniform to something more different. Fatigue Life (1<sup>st</sup> place, Figure F.26), Weibull (2<sup>nd</sup> place), Johnson SB (4<sup>th</sup> place) and Lognormal (9<sup>th</sup> place) are the best options for the probability distribution functions.

## 5.2. Correlation Analysis

Correlation analysis is performed to measure the strength of relationship between two or more random variables or measured data values. The correlated data may be either linear or nonlinear. If the dependent variable increases or decreases as the independent variable increases a *monotonic correlation* exists (Helsel and Hirsh, 1992). The scatter plots of predictor variables vs criterion variable are explored and it can be said that almost all are correlated, except T and ASPCT. (ASPCT. is not clear), are monotonic correlations.

There are three coefficients of correlation; these are Kendall's Tau ( $\tau$ ), Spearman's Rho ( $\rho$ ) and Pearson's  $r$ . The Kendall and Spearman are nonparametric correlations and Pearson is the parametric correlation. The method nonparametric is used when the distribution parameters are not known, such as mean or standard deviation, which describe the behavior of variable ("Kendall's Tau and Spearman's rank correlation coefficient", 2010). It is suggested to use nonparametric methods if the sample size is small. Moreover, Kendall and Spearman coefficients are resistant to outliers and Pearson's  $r$  is a measure of linear correlation (Helsel and Hirsh, 1992). Because of these reasons, in the correlation analysis Kendall's method is used despite the fact that the underlying distributions are covered by using EasyFit as summarized in the previous section. However, since Pearson's  $r$  is very familiar from many studies, Pearson's method is also applied and these two methods are compared to be on the safe side. Before giving the results of the analysis, brief information is explained on Kendall's Tau and Pearson's  $r$  correlation measures.

### 5.2.1. Kendall's Tau ( $\tau$ )

Kendall's Tau is based on the ranks of the data and this makes the method resistant to effect of small number of outliers or extremes (Helsel and Hirsh, 1992). Moreover, Helsel and Hirsch (1992) stated that it is suitable for variables, which shows skewness around the general associations. It is an important statement for this study, since the parameters of this study are generally skewed.

One of the significant characteristics of the Kendall's method is the values  $\tau$  takes. Similar to the Pearson's coefficient  $r$ ,  $\tau$  takes the values between -1 and 1. However, the strong coefficients of  $r$  as 0.9 or above correspond to  $\tau$  values of about 0.7 or above. This difference is due to the scale difference (Helsel and Hirsh, 1992).

The coefficient  $\tau$  is calculated as:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (\text{Equation 5.1})$$

Where:

$n_c$ , number of concordant

$n_d$ , Number of discordant

$\frac{1}{2}n(n-1)$ , Total number of possible pairing of observations

### 5.2.2. Pearson's $r$

This correlation coefficient is the most widely used measure in many analyses. It is also known as the linear correlation coefficient. This feature makes the analysis more risky since any low correlation coefficient may be due to a nonlinear relationship.

Pearson's  $r$  is sensitive to the outliers since it is computed using the normal distribution parameters, namely the mean and the standard deviation. Thus, for the untransformed hydrologic variables, which are generally skewed, it is not suitable to use  $r$  as a descriptor (Helsel and Hirsh, 1992).

The coefficient  $r$  is calculated as (Yu, Yang, and Wang, 2002):

$$r = \frac{\text{cov}(X,Y)}{\sqrt{S(X)S(Y)}} \quad (\text{Equation 5.2})$$

where

$$\text{cov}(x,y) = \frac{\sum_{i,j} X_i Y_j f_{ij} - (\sum_{i=1}^R X_i r_i)(\sum_{j=1}^C Y_j c_j)}{W} \quad (\text{Equation 5.3})$$

$$S(X) = \frac{\sum_{i=1}^R X_i^2 r_i - (\sum_{i=1}^R X_i r_i)^2}{W} \quad (\text{Equation 5.4})$$

$$S(Y) = \frac{\sum_{j=1}^C Y_j^2 c_j - (\sum_{j=1}^C Y_j c_j)^2}{W} \quad (\text{Equation 5.5})$$

Where

$x$  and  $y$ , the variables

$W = n - 1$

$S(X)$  and  $S(Y)$ , standard deviations of the two variables

(Equation 5.6)

### 5.2.3. Interpretation of the Results of Correlation Matrix

**r and  $\tau$**  : Indicates strength and direction ( $\pm$ ) of the correlation. As **r** and  **$\tau$**  values get higher the correlation between the variables get stronger. **r** values equal to or above 0.9 indicate a strong relationship, whereas for the  **$\tau$**  values the critical number corresponds 0.7,  **$\tau$**  values above this value indicate a strong relationship.

**Statistical Hypotheses:** Every **r** and  **$\tau$**  value (sample statistics) is a representation of actual correlation value in the population,  $\rho$ . When **r** and  **$\tau$**  get higher, more assurance is gotten that there really is a correlation. The following hypotheses proposed and one of them must be true.

$$H_0: \rho=0 \quad (\text{There is no actual correlation})$$

$$H_A: \rho \neq 0 \quad (\text{There is a correlation})$$

**p-value (Sig. 2-tailed):** The significant test value for a two tailed test. Smaller p-value is better. 5 %, two tailed confidence intervals are used during correlation analysis. If  $p \leq 0.05$ ,  $H_0$  is rejected, which means there is a true relationship.

The correlation matrices have been examined for Kendall's  $\tau$  and Pearson's **r** in the following sections in Table 5.15 and Table 5.16.

### 5.2.3.1. The computed Kendall's Tau ( $\tau$ ) values for 13 parameters

Table 5.15. Kendall's  $\tau$  correlation matrix

PARAMETERS	P	A	LMR	H <sub>max</sub>	H <sub>min</sub>	H <sub>mean</sub>	BR	S	ASPCT.	MAP	CN	T	Q
$\tau$	1	0.956	0.912	0.807	-0.910	-0.736	0.714	0.604	0.297	0.516	-0.516	0.155	<b>0.843</b>
Sig. (2-tailed)	-	1.9E-06	5.5E-06	8.6E-05	8.0E-06	2.4E-04	3.7E-04	0.003	0.139	0.010	0.010	0.443	3.6E-05
$\tau$	0.956	1	0.956	0.785	-0.955	-0.780	0.758	0.604	0.341	0.560	-0.560	0.155	<b>0.865</b>
Sig. (2-tailed)	-	1.9E-06	1.4E-04	1.4E-04	2.8E-06	1.0E-04	1.6E-04	2.6E-03	0.090	0.005	0.005	0.443	2.2E-05
$\tau$	0.912	0.956	1	0.739	-0.933	-0.780	0.714	0.604	0.385	0.516	-0.560	0.110	<b>0.843</b>
Sig. (2-tailed)	5.5E-06	1.9E-06	-	3.3E-04	4.7E-06	1.0E-04	3.7E-04	0.003	0.055	0.010	0.005	0.584	3.6E-05
$\tau$	0.807	0.785	0.739	1	-0.744	-0.580	0.557	0.603	0.262	0.557	-0.557	0.263	<b>0.872</b>
Sig. (2-tailed)	8.6E-05	1.4E-04	3.3E-04	-	0.000	0.005	0.007	0.003	0.203	0.007	0.007	0.203	3.0E-05
$\tau$	-0.910	-0.955	-0.933	-0.744	1	0.753	-0.820	-0.551	-0.371	-0.551	0.573	-0.090	<b>-0.862</b>
Sig. (2-tailed)	8.0E-06	2.8E-06	4.7E-06	3.7E-04	-	2.2E-04	5.7E-05	0.007	0.069	0.007	0.005	0.659	3.1E-05
$\tau$	-0.736	-0.780	-0.780	-0.580	0.753	1	-0.538	-0.429	-0.209	-0.516	0.560	-0.133	<b>-0.708</b>
Sig. (2-tailed)	2.4E-04	1.0E-04	1.0E-04	4.8E-03	2.2E-04	-	7.3E-03	3.3E-02	0.298	0.010	0.005	0.511	5.1E-04
$\tau$	0.714	0.758	0.714	0.557	-0.820	-0.538	1	0.363	0.407	0.363	-0.363	0.088	<b>0.686</b>
Sig. (2-tailed)	3.7E-04	1.6E-04	3.7E-04	6.7E-03	5.7E-05	7.3E-03	-	0.071	0.043	0.071	0.071	0.661	0.001
$\tau$	0.604	0.604	0.604	0.603	-0.551	-0.429	0.363	1	0.473	0.297	-0.297	0.221	<b>0.551</b>
Sig. (2-tailed)	0.003	0.003	0.003	0.003	0.007	0.033	0.071	-	0.019	0.139	0.139	0.273	0.007
$\tau$	0.297	0.341	0.385	0.262	-0.371	-0.209	0.407	0.473	1	0.077	-0.165	-0.044	0.303
Sig. (2-tailed)	0.139	0.090	0.055	0.203	0.069	0.298	0.043	0.019	-	0.702	0.412	0.826	0.137
$\tau$	0.516	0.560	0.516	0.557	-0.551	-0.516	0.363	0.297	0.077	1	-0.824	0.552	<b>0.551</b>
Sig. (2-tailed)	0.010	0.005	0.010	0.007	0.007	0.010	0.071	0.139	0.702	-	4.0E-05	6.1E-03	0.007
$\tau$	-0.516	-0.560	-0.560	-0.557	0.573	0.560	-0.363	-0.297	-0.165	-0.824	1	-0.464	<b>-0.573</b>
Sig. (2-tailed)	1.0E-02	5.2E-03	5.2E-03	6.7E-03	4.9E-03	5.2E-03	7.1E-02	1.4E-01	4.1E-01	4.0E-05	-	0.021	0.005
$\tau$	0.155	0.155	0.110	0.263	-0.090	-0.133	0.088	0.221	-0.044	0.552	-0.464	1	0.181
Sig. (2-tailed)	0.443	0.443	0.584	0.203	0.659	0.511	0.661	0.273	0.826	6.1E-03	0.021	-	0.377
$\tau$	0.843	0.865	0.843	0.872	-0.862	-0.708	0.686	0.551	0.303	0.551	-0.573	0.181	1
Sig. (2-tailed)	3.6E-05	2.2E-05	3.6E-05	3.0E-05	3.1E-05	5.1E-04	7.7E-04	6.9E-03	0.137	0.007	4.9E-03	0.377	-



When the last column of the Table 5.15 is examined the correlation of the independent parameters with the dependent parameter, Q, can be seen. The interpretations of the correlation matrix are:

- It is mentioned before that the strong coefficients of  $r$  as 0.9 or above corresponding to  $\tau$  values of about 0.7 or above. The P, A, LMR,  $H_{\max}$ ,  $H_{\min}$  and  $H_{\text{mean}}$  are all above 0.7 which means a high correlation. However, when the correlations between these parameters are analyzed, it is observed that they are highly intercorrelated. The high correlation between P and A is an expected situation, since they are both length measures of the same basin. Therefore, while developing the model one of them should be taken as the length parameter, this would be probably the A, since it is more meaningful than P and in the literature A is the most widely used one. The high correlations between LMR and the length measures of the basins, namely A and P, make also sense. As the basin gets bigger, the length of the channel gets longer.
- The elevation parameters,  $H_{\max}$ ,  $H_{\min}$  and  $H_{\text{mean}}$  are somehow intercorrelated and this should be taken into account. Moreover, BR, which is a derivation parameter from elevation parameters give lower correlation with A and P, than the others do. Furthermore, the correlation between BR and Q is slightly less than 0.7, which means it could be a good representative for the elevations' contribution to the model.
- Slope parameter gives a fair correlation number than the ASPCT. does. Moreover, the criterion for p-value is vanished in the ASPCT. correlation, it equals to 0.137 and it is strongly larger than 0.05. ASPCT. should not be used in the model as a descriptive parameter.
- The MAP, which is one of the climatic parameters, gives also not a strong correlation but a good correlation value. It should be included in the model.
- The terrain parameter CN, gives a negative medium correlation. It is an important parameter since it represents the permeability of the soil.
- Temperature, T, another climatic characteristic's correlation is very small and the p-value (confidence interval test statistic) is very large. It is not appropriate to use in the model.

The discussions above are the first interpretations of a correlation matrix. Some models will be developed according to some combinations, and the outcomes will be discussed in the light of these interpretations. The Kendall method is used as the correlation method since it is more convenient for the nonlinear relations and the scatter plots show nonlinear relations. However, to see the differences and to be on the safe side Pearson method's outcome matrix has been also analyzed.

### 5.2.3.2. The computed Pearson's (r) values for 13 parameters

Table 5.16. The Pearson's r correlation matrix

PARAMETERS	P	A	LMR	H <sub>max</sub>	H <sub>min</sub>	H <sub>mean</sub>	BR	S	ASPCT.	MAP	CN	T	Q
<b>P</b>	R	1	0.956	0.992	0.959	-0.879	-0.856	0.858	0.549	0.183	-0.760	0.169	<b>0.924</b>
	Sig. (2-tailed)	-	9.8E-08	4.0E-12	5.9E-08	3.4E-05	9.4E-05	8.7E-05	4.2E-02	5.3E-01	1.6E-03	5.6E-01	2.3E-06
<b>A</b>	R	0.956	1	0.960	0.921	-0.811	-0.719	0.822	0.714	0.234	-0.711	0.283	<b>0.992</b>
	Sig. (2-tailed)	9.8E-08	-	5.3E-08	2.9E-06	4.3E-04	3.7E-03	3.1E-04	4.1E-03	0.421	0.004	0.327	3.6E-12
<b>LMR</b>	R	0.992	0.960	1	0.941	-0.897	-0.861	0.880	0.592	0.196	-0.756	0.201	<b>0.931</b>
	Sig. (2-tailed)	4.0E-12	5.3E-08	-	5.3E-07	1.4E-05	7.8E-05	3.3E-05	0.026	0.501	0.002	0.491	1.4E-06
<b>H<sub>max</sub></b>	R	0.959	0.921	0.941	1	-0.755	-0.779	0.717	0.542	0.042	-0.782	0.283	<b>0.888</b>
	Sig. (2-tailed)	5.9E-08	2.9E-06	5.3E-07	-	0.002	0.001	0.004	0.045	0.888	0.001	0.328	2.3E-05
<b>H<sub>min</sub></b>	R	-0.879	-0.811	-0.897	-0.755	1	0.942	-0.989	-0.377	-0.128	0.669	-0.064	<b>-0.779</b>
	Sig. (2-tailed)	3.4E-05	4.3E-04	1.4E-05	1.8E-03	-	4.8E-07	2.2E-11	0.184	0.663	0.009	0.828	0.001
<b>H<sub>mean</sub></b>	R	-0.856	-0.719	-0.861	-0.779	0.942	1	-0.883	-0.200	0.069	-0.340	-0.032	<b>-0.666</b>
	Sig. (2-tailed)	9.4E-05	3.7E-03	7.8E-05	1.0E-03	4.8E-07	-	2.9E-05	4.9E-01	0.814	0.005	0.913	0.009
<b>BR</b>	R	0.858	0.822	0.880	0.717	-0.989	-0.883	1	0.440	0.209	-0.632	0.075	<b>0.801</b>
	Sig. (2-tailed)	8.7E-05	3.1E-04	3.3E-05	3.9E-03	2.2E-11	2.9E-05	-	0.115	0.473	0.015	0.798	0.001
<b>S</b>	R	0.549	0.714	0.592	0.542	-0.377	-0.200	0.440	1	0.522	-0.274	0.245	<b>0.747</b>
	Sig. (2-tailed)	0.042	0.004	0.026	0.045	0.184	0.492	0.115	-	0.056	0.343	0.398	0.002
<b>ASPCT.</b>	R	0.183	0.234	0.196	0.042	-0.128	0.069	0.209	0.522	1	-0.380	-0.367	0.232
	Sig. (2-tailed)	0.530	0.421	0.501	0.888	0.663	0.814	0.473	0.056	-	0.180	0.196	0.425
<b>MAP</b>	R	0.464	0.481	0.437	0.548	-0.323	-0.340	0.304	0.091	-0.380	1	0.824	<b>0.474</b>
	Sig. (2-tailed)	0.095	0.081	0.118	0.043	0.260	0.235	0.290	0.757	0.180	-	2.9E-04	0.087
<b>CN</b>	R	-0.760	-0.711	-0.756	-0.782	0.669	0.698	-0.632	-0.274	0.128	-0.853	1	<b>-0.668</b>
	Sig. (2-tailed)	1.6E-03	4.3E-03	1.8E-03	9.5E-04	9.0E-03	5.5E-03	1.5E-02	3.4E-01	6.6E-01	1.0E-04	-	0.009
<b>T</b>	R	0.169	0.283	0.201	0.283	-0.064	-0.032	0.075	0.245	-0.367	0.824	1	0.299
	Sig. (2-tailed)	0.563	0.327	0.491	0.328	0.828	0.913	0.798	0.398	0.196	2.9E-04	-	0.299
<b>Q</b>	R	0.924	0.992	0.931	0.888	-0.779	-0.666	0.801	0.747	0.232	0.474	-0.668	<b>1</b>
	Sig. (2-tailed)	2.3E-06	3.6E-12	1.4E-06	2.3E-05	1.0E-03	9.3E-03	5.8E-04	2.1E-03	0.425	0.087	9.1E-03	0.299

Table 5.16 shows almost the same results with Kendall's matrix. The P, A, LMR,  $H_{\max}$  and  $H_{\min}$  give strong correlations and  $H_{\text{mean}}$  give relatively small correlation with Q as mentioned in the discussions of the scatter plots. Different from the Kendall's correlation matrix the p-value of MAP is slightly larger than 0.05 and it seems acceptable. However the correlation of MAP is small, but it may due to the nonlinear relationship between MAP and Q, which is observed in Figure 5.10. Since Pearson is not good at representing the nonlinear relations, it is an expected result.

In brief, both of the correlation coefficients and the significance tests give similar results. P, A, LMR,  $H_{\max}$ ,  $H_{\min}$  and  $H_{\text{mean}}$  give high correlations. BR and S show good correlations too however, ASPCT. and T are poorly correlated. CN and MAP are questionable but will be beneficial to use to represent the climatic and soil characteristics.

### 5.3. Parametric Approach

In developing the regional model regression analysis, both linear and nonlinear multiple regression analysis has been performed by using SPSS. The program uses the method of least squares to perform the regression. The regression analysis in the SPSS contains several regression types, namely, Linear R. (Regression), Curve Estimation R., Binary Logistic R., Multinomial Logistic R., Ordinal R., Probit R., and Nonlinear R., Weight Estimation R., Optimal Scaling, 2-Stage Least Squares Regression.

To decide on the regional model, linear and nonlinear analyses have been performed. The linear regression analysis contains the inputs as follow:

- The dependent variable (Q), the independent variables (P, A, LMR,  $H_{\max}$ ,  $H_{\min}$ ,  $H_{\text{mean}}$ , BR, S, ASPCT., MAP, CN and T)
- Choosing one or more of the methods of regression, which are Enter R., Stepwise R., Remove R., Backward and Forward Regressions: In this analysis, *Enter method* is used due to the degree of freedom problem between parameters and the samples.
- Statistics, Plots, Save and Options subcommands can be specialized in line according to the needs.

In the nonlinear analysis the default algorithm, which is for unconstrained models, Levenberg-Marquardt method is used as the iteration method. It is possible to enter new values for maximum iterations, Sum-of-squares convergence and Parameter convergence could be changed.

The inputs of the nonlinear regression analysis are:

- The dependent variable (Q)
- The model expression with the parameters needed. For this case, the nonlinear expression from the study of Castellarin et al. (2004) and Algancı et al. (2009), which are similar to each other, are used.
- Inserting the starting values of the parameters for iteration is also needed. The starting values are taken from the models developed in the literature and tried to be chosen that are reasonable and, if possible, close to the expected final solution. At the end, most of them converged to an interval.

Before starting the regression analysis, five out of fourteen stations are selected to be used in the validation analysis. Therefore, only nine stations have been used in the model development by regression analysis. The stations are chosen according to the record duration period (yr) of the station. (Castellarin et al., 2007) proposed that for developing long-term flow-duration curves five years of observed streamflows are sufficient. Then it is decided to remove stations less than five years record duration. Table 5.16 shows the stations and record lengths of them. The stations marked with grey color (DSI-2313, DSI-2321, DSI-2333, DSI-2322 and DSI-2338) are under five years observation and eliminated from the data set to be used as validation stations. Moreover, Table 5.17 shows the basins used while developing parametric approaches. Grey shaded lines have been eliminated.

Table 5.17. Flow measurement stations and the record lengths

Station	Q (m <sup>3</sup> /s)	Duration (yr)	Area (km <sup>2</sup> )
DSI-2313	37.59	2*	6967.93
DSI-2321	5.47	2	1851.19
DSI-2336	0.57	10	49.25
DSI-2335	0.59	12	68.37
DSI-2333	0.84	4	47.20
DSI-2324	23.31	17	4537.55
DSI-2323	5.47	17	1094.21
DSI-2322	14.70	4	3501.01
DSI-2337	1.45	15	197.04
DSI-2338	0.64	3	72.83
DSI-2339	0.19	11	10.62
EIE-2323	34.30	41	6978.73
EIE-2325	7.10	31	1785.57
EIE-2329	16.50	24	3537.10

(\*) Grey shaded ones are for validation

As mentioned before, there are several studies about the subject of regionalization of flow duration curves. For example, Algancı et al. (2009) suggested two different functions and four different equations for Solaklı Basin and used  $R^2$  as the index of performance to decide on the appropriate model. The models have been developed with the parameters Area (A, km<sup>2</sup>), Height of Drop (H, m), Mean Precipitation (P, mm) and Mean Annual Discharge (Q, m<sup>3</sup>/s). Table 5.18 shows the equations and the  $R^2$  values.

Table 5.18. The flow equations of the study of Algancı et al. (2009) and the  $R^2$  values

Function	Type	Equation	$R^2$
Q= f(A,P)	Linear	$Q = -7.843 + (0.025A) + (0.0071P)$	0.893
	Nonlinear	$Q = (10^{-4.598}) * (A^{0.907}) * (P^{1.057})$	0.932
Q= f(A,P,H)	Linear	$Q = 18.692 + (0.0132A) + (0.0105P) - (0.0128H)$	0.904
	Nonlinear	$Q = (10^{5.748}) * (A^{0.469}) * (P^{1.337}) * (H^{-3.053})$	0.937

Similarly Castellarin et al. (2004) proposed predictive models for the empirical flood quantiles Q (D) (i.e., Q<sub>30</sub>, Q<sub>70</sub>, Q<sub>90</sub> and Q<sub>95</sub>) of the form:

$$Q(D) = A_0 X_1^{A_1} X_2^{A_2} \dots X_n^{A_n} + v \quad (\text{Equation 5.7})$$

$X_i$  stands for the parameters which are Area (A, km<sup>2</sup>), Length of Main Channel (L, km), Mean Annual Precipitation (MAP, mm) and Basin Relief ( $\Delta H = H_{\text{mean}} - H_{\text{min}}$ , m) and  $v$  is the error term. Table 5.19 demonstrates the proposed models. E stands for the Nash Sutcliffe efficiency criterion explained in the Index of Performance section in Chapter 2.

Table 5.19. Nonlinear models developed by Castellarin et al. (2007)

Function	Equation	E
Q= f(A,LMR,MAP,BR)	$Q_{30} = 1.154 * E-3 * (A^{0.795}) * (L^{0.317}) * (MAP^{1.234}) * (\Delta H^{0.426}) + \eta$	0.965
	$Q_{70} = 1.764 * E-6 * (A^{1.267}) * (L^{-0.016}) * (MAP^{0.62}) * (\Delta H^{1.109}) + \eta'$	0.979
	$Q_{90} = 2.053 * E-7 * (A^{1.326}) * (L^{-0.16}) * (MAP^{0.045}) * (\Delta H^{1.278}) + \eta''$	0.973
	$Q_{95} = 1.315 * E-7 * (A^{1.427}) * (L^{-0.194}) * (MAP^{-0.159}) * (\Delta H^{1.273}) + \eta'''$	0.974

In this study, the exact descriptive variables are not decided yet due to degrees of freedom issues. To decide the parameters and the model (i.e., linear or nonlinear) several regression analyses have been performed. The models of Castellarin et al. (2004) and Algancı's et.al (2009) models have been used as a guide. They have been used as guide while proposing the model and the coefficients of the equations in Table 5.18 and Table 5.19 have been used as the initial values, which are necessary to insert in software.

Table 5.20 gives a list of cases applied in regression analysis. The parameters used in the equation and the performance indices are available. Moreover, when these performance indices are examined it is seen that the models developed parallel to the study of Algancı et al. (2009), namely Cases 8, 9 and 10, are not as good as the first two cases, Case 1 and Case 2.

In brief, Case 1 and Case 2 seem to be the best equations for the linear models and the descriptive variables A, LMR, BR, S, MAP and CN seem to be suitable to use in the model. However, the nonlinear models should also be analyzed to make a final decision.

Table 5.21 illustrates the nonlinear models. The performance index, E (Nash-Sutcliffe efficiency criterion), for each model is higher than the linear models' E, which is the first outcome of the analysis of the Table 5.21. The last three cases, Case 6 and Case 7 give the best results. Actually, Case 6 seems perfect, E and R are the highest, and the rest of the indices should be minimum in value as in this case. However, a good model is not only the one, which has the best index values but also the one includes minimum number of descriptive parameters to represent the natural phenomenon. Case 6 includes six parameters although Case 7 includes five, but the Case 7 is as good as Case 6, which means using S as a parameter do not contribute much explanation to the model. Therefore, it will be appropriate to use Case 7 as the representative model of the discharge. As a final word, the parameters A, LMR, BR, MAP and CN are the explanatory parameters of the nonlinear model:

$$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6} \quad (\text{Equation 5.8})$$

$$Q = 9.93 \times 10^{-7} (A^{0.68}) (LMR^{0.964}) (MAP^{-1.516}) (BR^{0.423}) (CN^{3.025}) \quad (\text{Equation 5.9})$$

Equation 5.8 is the regional model of the Oltu Basin. The next step is deriving the coefficients  $b_1, b_2, \dots, b_6$ . The regression coefficients can be seen in Equation 5.9, which are representative for the regional model and help to see the order of the coefficients.

Table 5.20. Linear models regressed and the performance indices

Case	Equation	Reliability Indices					
		E	R	$\varepsilon_g$	$\sigma_{\varepsilon,s}$	RRMSE	RMSE
<b>1</b> Eqn. Std. Eqn.*	Q= f(A,LMR,BR,S,MAP,CN)						
	Q= -29.65 + 0.004A + 0.071LMR - 0.001BR + 0.085S - 0.002MAP + 0.325CN	0.995	0.997	-0.363	0.679	0.735	0.867
	Q= 0.825A + 0.344LMR - 0.019BR + 0.037S - 0.008MAP + 0.181CN						
<b>2</b> Eqn. Std. Eqn.	Q= f(A,LMR,S,MAP,CN)						
	Q= -32.785 + 0.004A + 0.059LMR + 0.076S + 0.004MAP + 0.335CN	0.997	0.999	-0.025	0.486	0.459	0.640
	Q= 0.851A + 0.282LMR + 0.033S + 0.017MAP + 0.186CN						
<b>3</b> Eqn. Std. Eqn.	Q= f(A,LMR,S,MAP)						
	Q= 28.535 + 0.004A + 0.111LMR + 0.174S - 0.076MAP	0.990	0.995	0.038	0.360	0.342	1.226
	Q= 0.761A + 0.533LMR + 0.076S - 0.341MAP						
<b>4</b> Eqn. Std. Eqn.	Q= f(A,LMR,MAP)						
	Q= 15.218 + 0.005A + 0.123LMR - 0.035MAP	0.438	0.662	1.222	1.110	1.609	9.128
	Q= 1.035A + 0.113LMR - 0.158MAP						
<b>5</b> Eqn. Std. Eqn.	Q= f(A,MAP,S)						
	Q= 10.267 + 0.005A - 0.022MAP - 0.021S	0.989	0.994	-0.029	0.350	0.332	1.297
	Q= 1.095A - 0.100MAP - 0.009S						
<b>6</b> Eqn. Std. Eqn.	Q= f(LMR,BR,MAP,S,CN)						
	Q= 36.133 - 0.151MAP + 0.55S + 0.357LMR - 0.003BR + 0.174CN	0.992	0.996	-0.499	0.937	1.015	1.062
	Q= -0.672MAP + 0.241S + 1.721LMR - 0.098BR + 0.097CN						
<b>7</b> Eqn. Std. Eqn.	Q= f(A,H <sub>max</sub> ,LMR,MAP,S,CN)						
	Q= -19.848 + 0.006A - 0.015H <sub>max</sub> + 0.003LMR + 0.063MAP - 0.095S + 0.407CN	0.993	0.997	-0.972	1.829	1.979	1.002
	Q= 1.136A - 0.211H <sub>max</sub> + 0.017LMR + 0.282MAP - 0.042S + 0.226CN						
<b>8</b> Eqn. Std. Eqn.	Q= f(A,MAP)						
	Q= 9.196 + 0.005A - 0.021MAP	0.988	0.994	-0.209	0.331	0.376	1.331
	Q= 1.084A - 0.093MAP						
<b>9</b> Eqn. Std. Eqn.	Q= f(A,MAP,H <sub>mean</sub> )						
	Q= 10.667 + 0.005A - 0.022MAP + 0.0004H <sub>mean</sub>	0.983	0.991	1.913	3.020	3.431	1.601
	Q= 1.086A - 0.099MAP - 0.005H <sub>mean</sub>						
<b>10</b> Eqn. Std. Eqn.	Q= f(A,MAP,H <sub>max</sub> )						
	Q= 22.61 + 0.006A - 0.009MAP - 0.007H <sub>max</sub>	0.989	0.994	-0.926	1.421	1.628	1.280
	Q= 1.126A - 0.039MAP - 0.096H <sub>max</sub>						
<b>11</b> Eqn. Std. Eqn.	Q= f(P,A,LMR,H <sub>max</sub> ,BR,S,MAP,CN)						
	Q= 14.796 + 0.025P + 0.006A - 0.108LMR - 0.027H <sub>max</sub> + 0.003BR - 0.076S + 0.072MAP + 0.353CN	0.993	0.997	0.991	1.457	1.694	1.009
	Q= 0.443P + 1.250A - 0.519LMR - 0.391H <sub>max</sub> + 0.101BR - 0.033S + 0.319MAP + 0.196CN						

\* Standardized coefficients are used.



Table 5.21. Nonlinear models regressed and the performance indices

Case	Equation	Reliability Indices				
		E	R	$\varepsilon_s$	$\sigma_{e,s}$	RMSE
1 Eqn.	$Q=f(A,MAP)$ $Q=7.84 \times 10^{-5} (A^{0.992}) (MAP^{0.666})$	0.99656	0.99828	-0.248	0.332	0.714
2 Eqn.	$Q=f(A,H_{max},MAP)$ $Q=1 \times 10^7 (A^{1.202}) (MAP^{0.144}) (H_{max}^{-3.031})$	0.83985	0.91643	-0.518	0.278	4.872
3 Eqn.	$Q=f(A,H_{mean},MAP)$ $Q=1 \times 10^5 (A^{1.042}) (MAP^{0.006}) (H_{mean}^{-2.299})$	0.99689	0.99844	-0.303	0.377	0.679
4 Eqn.	$Q=f(A,LMR,MAP,BR)$ $Q=1.098 (A^{0.736}) (LMR^{0.722}) (MAP^{1.192}) (BR^{0.111})$	0.99783	0.99891	-0.291	0.385	0.568
5 Eqn.	$Q=f(A,LMR,MAP,BR,S)$ $Q=0.039 (A^{1.109}) (LMR^{-0.044}) (MAP^{1.283}) (BR^{0.938}) (S^{-0.457})$	0.99882	0.99941	-0.360	0.441	0.418
6 Eqn.	$Q=f(A,LMR,MAP,BR,S,CN)$ $Q=3.43 \times 10^{-11} (A^{0.447}) (LMR^{1.446}) (MAP^{0.769}) (BR^{-0.056}) (S^{0.318}) (CN^{4.796})$	0.99983	0.99991	-0.210	0.345	0.160
7 Eqn.	$Q=f(A,LMR,MAP,BR,CN)$ $Q=9.93 \times 10^{-7} (A^{0.668}) (LMR^{0.964}) (MAP^{1.516}) (BR^{0.423}) (CN^{3.025})$	0.99950	0.99975	-0.292	0.392	0.271

## 5.4. Jack-Knife Cross-Validation

As explained in the performance indices section, performance assessment is the most significant issue after developing the regional model. Jack-knife cross-validation method is a composition of two basic statistical techniques to get a statistical inference about the model. They are about re-using and re-sampling of the data. In brief, Jack-knife method is used for bias reduction for an estimator and cross-validation is the method used for evaluating the error rate of a predictive model. Both methods have a common step that omits the data points at a one time. However, they are completely different. Cross-validation method does not include any estimated statistics that is extended as in jackknife method. Cross-validation is to use how well the fitted model predicts the eliminated point and averages the prediction errors to get a final prediction error (Efron, 1982).

The purposes of applying Jack-knife and cross-validation are (Castellarin et al., 2007):

- calculating the uncertainty of FDCs predicted for the ungaged basins
- comparing the performance of the regional FDCs predicted through the developed regional model (Kocatepe model) and the statistical methods

The procedure of jackknife cross-validation can be summarized as follows (Castellarin et al., 2004):

1.  $N$  streamgauges, which are selected before in the regression analysis part, are used;
2. the streamgauge,  $s$ , is removed;
3. Model parameterisation for the parametric approach and determination of the regional FDCs are performed by considering the streamflow data and the geomorphoclimatic characteristics of the remaining  $N - 1$  gauged sites;
4. Using the regional model identified at section 5.3 the FDC for station  $s$  is estimated;
5. Steps 2-4 are repeated  $N - 1$  times, considering in turn one of the remaining streamgauges, providing important indications on the robustness and the regional FDC model for the entire basin.

Before applying the procedure, empirical discharges for durations  $j$  have been extracted from the FDCs constructed in Chapter 3 (see also Appendix B). In section 3.3, while constructing the FDCs Weibull plotting position is used and each  $q_{(i)}$  against its corresponding duration  $D_{(i)}$  is plotted as FDC. To be able to apply the jackknife cross-validation procedure several durations,  $j$ , have been chosen from these plotted FDCs (Appendix B, daily FDCs of stations). The selection criteria include:

- hydropower engineering motivation, means durations between 0.3 and 0.99 (Castellarin et al., 2004)
- more frequent durations at the top end of the FDC since it is hard to model (Castellarin et al., 2004)

In the light of these criteria 12 durations,  $j$ , are selected as: 2, 5, 8, 10, 15, 20, 30, 50, 70, 90, 95 and 98; corresponds to  $D_{(i)}$ : 0.02, 0.05, 0.08, 0.1, 0.15, 0.2, 0.3, 0.5, 0.7, 0.9, 0.95 and 0.98.

Table 5.22 represents the streamgauges selected for regression analysis in section 5.3, and corresponding twelve,  $q_j$  ( $q_2, q_5, \dots, q_{98}$ ) values.

After extraction of the  $q_j$  discharges jackknife cross-validation procedure is applied all 9 stations presented in Table 5.22. The following sections contain the nonlinear equations for each  $q_j$  after nonlinear regression analysis applied according to the regional model developed, observed,  $q_{s,j}$ , and the predicted discharges  $\hat{q}_{s,j}$  and indices of performances for each subset (the station  $s$ , removed from the data set).

Table 5.22. Stations and corresponding empirical  $q_j$  discharges in  $m^3/s$

Stations	$q_2$	$q_5$	$q_8$	$q_{10}$	$q_{15}$	$q_{20}$	$q_{30}$	$q_{50}$	$q_{70}$	$q_{90}$	$q_{95}$	$q_{98}$
DSI-2336	4.11	2.56	1.98	1.68	1.12	0.68	0.30	0.15	0.12	0.07	0.03	0.03
DSI-2335	4.19	2.80	1.98	1.62	1.07	0.76	0.49	0.12	0.07	0.04	0.04	0.02
DSI-2324	121.00	86.00	69.00	61.00	45.89	31.60	16.50	11.80	9.11	5.39	3.28	2.03
DSI-2323	31.40	22.90	17.55	15.20	10.60	6.80	3.72	2.58	1.77	0.97	0.67	0.44
DSI-2337	9.34	6.74	5.20	4.25	2.76	1.91	1.02	0.46	0.30	0.19	0.13	0.10
DSI-2339	1.20	0.65	0.43	0.38	0.31	0.22	0.15	0.09	0.05	0.05	0.02	0.01
EIE-2323	168.00	116.30	92.40	83.00	60.95	43.35	26.85	18.25	14.65	11.35	9.91	7.80
EIE-2325	37.00	25.45	20.40	17.50	12.00	8.12	5.20	3.90	3.06	1.23	0.56	0.30
EIE-2329	88.47	65.00	52.43	46.16	30.96	19.95	11.15	7.64	6.14	3.68	2.44	1.46

In Table 5.23., the representative equations obtained by applying nonlinear regression are given. In the regression, station DSI-2336 has been omitted and remaining stations have been used for regression, but this time the dependent variable was not mean annual discharge,  $Q$ , it is  $q_j$ , which has been taken from the empirical FDCs. By using these equations, predicted discharges,  $\hat{q}_{s,j}$ , for station  $s$  and duration  $j$  can be obtained. The empirical and the predicted discharges have been used to calculate several performance indices.

Table 5.23. Equations of  $q_j$  discharges obtained by omitting DSI-2336

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$9.45 \times 10^{-8} (A^{0.61})(LMR^{1.098})(BR^{0.366})(MAP^{-1.641})(CN^{4.212})$
$q_5$	$9.94 \times 10^{-10} (A^{0.654})(LMR^{1.104})(BR^{0.269})(MAP^{-1.606})(CN^{5.208})$
$q_8$	$1.74 \times 10^{-9} (A^{0.667})(LMR^{1.149})(BR^{0.136})(MAP^{-1.575})(CN^{5.126})$
$q_{10}$	$1.1 \times 10^{-9} (A^{0.702})(LMR^{1.130})(BR^{0.183})(MAP^{-1.666})(CN^{5.212})$
$q_{15}$	$2.55 \times 10^{-11} (A^{0.414})(LMR^{1.748})(BR^{0.367})(MAP^{-1.843})(CN^{5.827})$
$q_{20}$	$2.71 \times 10^{-11} (A^{0.34})(LMR^{1.829})(BR^{0.528})(MAP^{-1.646})(CN^{5.224})$
$q_{30}$	$1.12 \times 10^{-6} (A^{0.797})(LMR^{0.486})(BR^{0.705})(MAP^{-0.64})(CN^{1.485})$
$q_{50}$	$6.57 \times 10^{-5} (A^{0.588})(LMR^{0.888})(BR^{0.541})(MAP^{-0.777})(CN^{0.882})$
$q_{70}$	$6.01 \times 10^{-5} (A^{0.839})(LMR^{0.463})(BR^{0.261})(MAP^{-0.051})(CN^{0.091})$
$q_{90}$	$1.79 \times 10^{-6} (A^{1.453})(LMR^{-0.225})(BR^{1.326})(MAP^{-1.977})(CN^{1.580})$
$q_{95}$	$2.81 \times 10^{-10} (A^{2.4})(LMR^{-2.034})(BR^{2.655})(MAP^{-2.508})(CN^{2.298})$
$q_{98}$	$6.4 \times 10^{-5} (A^{2.671})(LMR^{-2.424})(BR^{3.874})(MAP^{-5.371})(CN^{1.414})$

Table 5.24 is giving information about empirical and predicted discharges. When Table 5.24 is examined, it is seen that the predicted discharges highly underestimated. Moreover, the relative errors of each station is high, they should be smaller to represent a good fit.

Table 5.24. Empirical and predicted discharges and relative errors for station DSI-2336

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	4.110	1.077	-0.738
$q_5$	5	2.560	0.810	-0.683
$q_8$	8	1.978	0.640	-0.676
$q_{10}$	10	1.680	0.493	-0.706
$q_{15}$	15	1.120	0.292	-0.740
$q_{20}$	20	0.680	0.165	-0.757
$q_{30}$	30	0.300	0.089	0.705
$q_{50}$	50	0.150	0.074	-0.510
$q_{70}$	70	0.121	0.027	-0.777
$q_{90}$	90	0.070	0.005	-0.933
$q_{95}$	95	0.032	7.7E-04	-0.976
$q_{98}$	98	0.029	1.4E-04	-0.995

Cross-validated performance index in Table 5.25,  $E_j$  equals to 0.100, state that the predicted values are not good. RMSE and RRMSE are overall evaluations of the station, and furthermore they tell the model is poor for the basin represented by the station DSI-2336. As a reminder, the area of the subbasin is 49.2 km<sup>2</sup>.

Table 5.25. Performance indices calculated for DSI-2336

Performance Index	Value
$E_j (*)$	0.100
<b>R</b>	0.317
$\bar{\varepsilon}_s$	-0.699
$\sigma_{\varepsilon,s}$	0.078
$E_s$	0.988
<b>RRMSE</b>	0.778
<b>RMSE</b>	1.224

\* $E_j$ : Nash-Sutcliffe efficiency criterion for Jack-knife Cross-validation, j holds for jack-knives' initials

- **DSI-2335**

The jackknifed equations for duration  $j$  at the streamgauge DSI-2335 are given in Table 5.26. The upstream of the streamgauge DSI-2335 is a small catchment like DSI-2336 its area is 68.4 km<sup>2</sup>. The coefficients of the equations are very similar for these two stations.

Table 5.26. Equations of  $q_j$  discharges obtained by omitting DSI-2335

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$8.08 \times 10^{-9} (A^{0.626})(LMR^{1.037})(BR^{0.348})(MAP^{-1.295})(CN^{4.347})$
$q_5$	$5.91 \times 10^{-7} (A^{0.606})(LMR^{1.281})(BR^{0.310})(MAP^{-2.496})(CN^{4.852})$
$q_8$	$1.06 \times 10^{-9} (A^{0.669})(LMR^{1.138})(BR^{0.131})(MAP^{-1.502})(CN^{5.150})$
$q_{10}$	$8.37 \times 10^{-10} (A^{0.703})(LMR^{1.125})(BR^{0.179})(MAP^{-1.623})(CN^{5.223})$
$q_{15}$	$4.23 \times 10^{-8} (A^{0.359})(LMR^{1.951})(BR^{0.416})(MAP^{-2.878})(CN^{5.416})$
$q_{20}$	$1.69 \times 10^{-9} (A^{0.310})(LMR^{1.939})(BR^{0.552})(MAP^{-2.208})(CN^{4.981})$
$q_{30}$	$3.66 \times 10^{-7} (A^{0.804})(LMR^{0.462})(BR^{0.696})(MAP^{-0.489})(CN^{1.554})$
$q_{50}$	$1.12 \times 10^{-5} (A^{0.600})(LMR^{0.841})(BR^{0.530})(MAP^{-0.533})(CN^{0.982})$
$q_{70}$	$1.78 \times 10^{-5} (A^{0.847})(LMR^{0.430})(BR^{0.254})(MAP^{0.220})(CN^{0.158})$
$q_{90}$	$7.49 \times 10^{-9} (A^{1.492})(LMR^{-0.370})(BR^{1.292})(MAP^{-1.221})(CN^{1.891})$
$q_{95}$	$0.00019 (A^{2.297})(LMR^{-1.625})(BR^{2.775})(MAP^{-4.575})(CN^{1.739})$
$q_{98}$	$5.22 \times 10^{-12} (A^{2.789})(LMR^{-2.876})(BR^{3.759})(MAP^{-3.040})(CN^{2.266})$

The predicted discharges like the coefficients of the equations are underestimated at station DSI-2335 as in DSI-2336. Moreover, the underestimate values cause high relative errors, which can be seen at the last column in Table 5.27. The performance indices show very similar pattern with station DSI-2336.  $E_j$  is very small to represent a good fit. However, the performance index is not that bad, it is above the critical value 0.75, which means fair to good fit. It can be interpreted, as, although the equations give underestimated results, the performances of the equations are reasonable.

Table 5.27. Empirical and predicted discharges and relative errors for station DSI-2335

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	4.190	1.479	-0.647
$q_5$	5	2.800	0.978	-0.651
$q_8$	8	1.980	0.863	-0.564
$q_{10}$	10	1.620	0.794	-0.510
$q_{15}$	15	1.070	0.318	-0.703
$q_{20}$	20	0.759	0.186	-0.755
$q_{30}$	30	0.488	0.125	-0.744
$q_{50}$	50	0.124	0.093	-0.254
$q_{70}$	70	0.072	0.067	-0.068
$q_{90}$	90	0.044	0.011	-0.760
$q_{95}$	95	0.041	0.0017	-0.959
$q_{98}$	98	0.023	0.0006	-0.971

Table 5.28. Performance indices calculated for DSI-2335

Performance Index	Value
$E_j$	0.297
$R$	0.545
$\bar{\varepsilon}_s$	-0.604
$\sigma_{\varepsilon,s}$	0.164
$E_s$	0.990
RRMSE	0.680
RMSE	1.113

- **DSI-2324**

The basin at the upstream of the streamgauge DSI-2324 is the second largest basin, the area its equals to 4537.55 km<sup>2</sup>. When equations and coefficients in Table 5.29 is studied, it is observed that, the order of the coefficient  $b_1$  is different than for DSI-2336 and DSI-2335. The size of the parameters is the main cause of this difference, since A and LMR values are high of station DSI-2324.



Table 5.29. Equations of  $q_j$  discharges obtained by omitting DSI-2324

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{-3.025})$
$q_j$	Equation of $q_j$
$q_2$	$3.84 \times 10^{-13} (A^{0.745})(LMR^{0.623})(BR^{0.273})(MAP^{0.315})(CN^{4.661})$
$q_5$	$1.26 \times 10^{-9} (A^{0.974})(LMR^{0.258})(BR^{0.3120})(MAP^{-0.776})(CN^{4.194})$
$q_8$	$1.26 \times 10^{-13} (A^{0.907})(LMR^{0.430})(BR^{0.088})(MAP^{0.254})(CN^{5.084})$
$q_{10}$	$1.28 \times 10^{-13} (A^{0.985})(LMR^{0.308})(BR^{0.148})(MAP^{0.193})(CN^{5.027})$
$q_{15}$	$1.38 \times 10^{-14} (A^{0.902})(LMR^{0.417})(BR^{0.389})(MAP^{0.266})(CN^{4.998})$
$q_{20}$	$4.206 \times 10^{-15} (A^{0.714})(LMR^{0.776})(BR^{0.513})(MAP^{0.381})(CN^{4.776})$
$q_{30}$	$1.77 \times 10^{-12} (A^{0.678})(LMR^{0.665})(BR^{0.564})(MAP^{0.805})(CN^{2.742})$
$q_{50}$	$2.52 \times 10^{-12} (A^{0.463})(LMR^{1.068})(BR^{0.382})(MAP^{1.029})(CN^{2.514})$
$q_{70}$	$1.24 \times 10^{-11} (A^{0.448})(LMR^{1.336})(BR^{0.064})(MAP^{1.058})(CN^{2.306})$
$q_{90}$	$3.27 \times 10^{-16} (A^{1.079})(LMR^{0.521})(BR^{1.052})(MAP^{0.073})(CN^{4.141})$
$q_{95}$	$5.46 \times 10^{-17} (A^{1.403})(LMR^{0.093})(BR^{1.844})(MAP^{-0.612})(CN^{4.025})$
$q_{98}$	$1.17 \times 10^{-18} (A^{0.964})(LMR^{1.513})(BR^{1.344})(MAP^{0.468})(CN^{3.344})$

The predicted discharges are very close to the observed discharges for this station as it can be seen in Table 5.29. Therefore, the relative errors get smaller. There are overestimated results as well as underestimated ones. The jackknifed Nash-Sutcliffe efficiency criterion  $E_j$ , square root of coefficient determination  $R$ , and the performance index  $E_s$  are high enough to claim a good fit.

Furthermore, mean relative error  $\bar{e}_s$  and the standard deviation of it  $\sigma_{e,s}$ , and the mean square errors are small enough to support this claim. In brief, the regional model for the basin presents good results.

Table 5.30. Empirical and predicted discharges and relative errors for station DSI-2324

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	121.000	118.744	-0.019
$q_5$	5	86.000	80.000	-0.070
$q_8$	8	69.000	65.541	-0.050
$q_{10}$	10	61.000	57.875	-0.051
$q_{15}$	15	45.890	41.504	-0.096
$q_{20}$	20	31.600	29.428	-0.069
$q_{30}$	30	16.500	17.272	0.047
$q_{50}$	50	11.800	12.366	0.048
$q_{70}$	70	9.110	10.212	0.121
$q_{90}$	90	5.390	6.104	0.132
$q_{95}$	95	3.280	4.321	0.317
$q_{98}$	98	2.030	3.490	0.719

Table 5.31. Performance indices calculated for DSI-2324

Performance Index	Value
$E_j$	0.994
$R$	0.997
$\bar{\varepsilon}_s$	0.086
$\sigma_{\varepsilon,s}$	0.232
$E_s$	0.9999
<b>RRMSE</b>	0.238
<b>RMSE</b>	2.902

- **DSI-2323**

The basin that gives its surface water to station DSI-2323 is one of the largest basins, in the study area being equals to 1094.21 km<sup>2</sup>. Therefore, the coefficients of the equations for  $q_j$  discharges are very similar to the coefficients of the model for station DSI-2324.

Table 5.32. Equations of  $q_j$  discharges obtained by omitting DSI-2323

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$2.3 \times 10^{-13} (A^{0.677})(LMR^{0.650})(BR^{0.430})(MAP^{0.170})(CN^{4.839})$
$q_5$	$1.01 \times 10^{-12} (A^{0.767})(LMR^{0.732})(BR^{0.006})(MAP^{0.198})(CN^{4.805})$
$q_8$	$8.96 \times 10^{-12} (A^{0.799})(LMR^{0.804})(BR^{-0.239})(MAP^{0.246})(CN^{4.438})$
$q_{10}$	$2.76 \times 10^{-12} (A^{0.839})(LMR^{0.780})(BR^{-0.203})(MAP^{0.233})(CN^{4.592})$
$q_{15}$	$1.57 \times 10^{-14} (A^{0.566})(LMR^{1.381})(BR^{-0.061})(MAP^{0.261})(CN^{5.294})$
$q_{20}$	$2.34 \times 10^{-14} (A^{0.483})(LMR^{1.465})(BR^{0.130})(MAP^{0.379})(CN^{4.692})$
$q_{30}$	$2.79 \times 10^{-13} (A^{0.780})(LMR^{0.087})(BR^{1.131})(MAP^{0.595})(CN^{2.993})$
$q_{50}$	$1.07 \times 10^{-8} (A^{0.635})(LMR^{0.659})(BR^{0.551})(MAP^{0.314})(CN^{1.464})$
$q_{70}$	$5.08 \times 10^{-7} (A^{0.813})(LMR^{0.372})(BR^{0.472})(MAP^{0.224})(CN^{0.752})$
$q_{90}$	$1.5 \times 10^{-12} A^{1.603} (LMR^{-0.669})(BR^{1.039})(MAP^{0.474})(CN^{1.927})$
$q_{95}$	$9.55 \times 10^{-14} (A^{2.678})(LMR^{-2.373})(BR^{1.742})(MAP^{0.310})(CN^{1.383})$
$q_{98}$	$1.15 \times 10^{-15} (A^{3.122})(LMR^{-3.289})(BR^{2.639})(MAP^{0.340})(CN^{0.949})$

If Table 5.33 is explored it is observed that the estimated and observed discharges are not very different from each other. There are under and over estimates as expected. Some of the relative errors are questionable due to being quantitatively large. However, Table 5.34 ends these questions, since performance indices are fair and reasonable to prove a good fit.

Table 5.33. Empirical and predicted discharges and relative errors for station DSI-2323

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	31.400	35.322	0.125
$q_5$	5	22.900	21.665	-0.054
$q_8$	8	17.550	15.710	-0.105
$q_{10}$	10	15.200	13.532	-0.110
$q_{15}$	15	10.600	9.206	-0.131
$q_{20}$	20	6.800	6.054	-0.110
$q_{30}$	30	3.715	4.937	0.329
$q_{50}$	50	2.580	2.672	0.036
$q_{70}$	70	1.770	1.989	0. 24
$q_{90}$	90	0.970	0.895	-0.078
$q_{95}$	95	0.672	0.453	-0.326
$q_{98}$	98	0.440	0.271	-0.384

Table 5.34. Performance indices calculated for DSI-2323

Performance Index	Value
$E_j$	0.976
$R$	0.988
$\bar{\varepsilon}_s$	-0.057
$\sigma_{\varepsilon,s}$	0.195
$E_s$	0.99973
RRMSE	0.195
RMSE	1.5 3

- **DSI-2337**

It is one of the medium-sized basins in the study area. The area of the basin is 197.04 km<sup>2</sup>. Some of the coefficients display significant changes as it can be seen in Table 5.35. The order of  $q_{70}$  and  $q_{95}$  are quite larger than other  $b_1$  values. This difference is due to the characteristics of the basin and the FDC.

Table 5.35. Equations of  $q_j$  discharges obtained by omitting DSI-2337

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$9.38 \times 10^{-9} (A^{0.475})(LMR^{1.101})(BR^{0.997})(MAP^{-2.574})(CN^{5.329})$
$q_5$	$1.83 \times 10^{-10} (A^{0.595})(LMR^{1.027})(BR^{0.577})(MAP^{-1.823})(CN^{5.610})$
$q_8$	$8.28 \times 10^{-11} (A^{0.597})(LMR^{1.073})(BR^{0.514})(MAP^{-1.834})(CN^{5.807})$
$q_{10}$	$1.62 \times 10^{-9} (A^{0.648})(LMR^{1.094})(BR^{0.413})(MAP^{-1.973})(CN^{5.335})$
$q_{15}$	$1.37 \times 10^{-11} (A^{0.391})(LMR^{1.683})(BR^{0.486})(MAP^{-1.825})(CN^{5.866})$
$q_{20}$	$2.93 \times 10^{-12} (A^{0.283})(LMR^{1.738})(BR^{0.818})(MAP^{-1.744})(CN^{5.618})$
$q_{30}$	$3.97 \times 10^{-10} (A^{0.605})(LMR^{0.406})(BR^{1.664})(MAP^{-1.540})(CN^{3.526})$
$q_{50}$	$3.56 \times 10^{-5} (A^{0.536})(LMR^{0.908})(BR^{0.764})(MAP^{-1.157})(CN^{1.291})$
$q_{70}$	$0.00011 (A^{0.753})(LMR^{0.532})(BR^{0.590})(MAP^{-0.702})(CN^{0.591})$
$q_{90}$	$1.4 \times 10^{-6} (A^{1.551})(LMR^{-0.331})(BR^{0.948})(MAP^{-1.065})(CN^{0.858})$
$q_{95}$	$0.00018 (A^{1.448})(LMR^{-1.056})(BR^{6.103})(MAP^{-11.441})(CN^{7.294})$
$q_{98}$	$6.17 \times 10^{-6} (A^{2.921})(LMR^{-2.668})(BR^{2.941})(MAP^{-3.08})(CN^{-0.063})$

Positive relative errors indicate the overestimated results, and the negative ones indicate underestimates. All of the performance indices in Table 5.37 are ideal for a developed model.

Table 5.36. Empirical and predicted discharges and relative errors for station DSI-2337

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	9.340	9.674	0.036
$q_5$	5	6.740	6.316	-0.063
$q_8$	8	5.200	5.064	-0.026
$q_{10}$	10	4.250	3.782	-0.110
$q_{15}$	15	2.760	.426	-0.121
$q_{20}$	20	1.910	1.599	-0.163
$q_{30}$	30	1.020	0.998	-0.022
$q_{50}$	50	0.455	0.468	0.028
$q_{70}$	70	0.300	0.306	0.020
$q_{90}$	90	0.188	0.054	-0.712
$q_{95}$	95	0.125	0.121	-0.031
$q_{98}$	98	0.096	0.004	-0.956

Table 5.37. Performance indices calculated for DSI-2337

Performance Index	Value
$E_j$	0.993
$R$	0.996
$\bar{\varepsilon}_s$	-0.106
$\sigma_{\varepsilon,s}$	0.211
$E_s$	0.99991
<b>RRMSE</b>	0.352
<b>RMSE</b>	0.264

- **DSI-2339**

The basin DSI-2339 is the smallest basin in this study it is only 10.62 km<sup>2</sup>. This area size probably effects the regional model that is developed before. The tables with the discharges and the performance indices would give better inferences.

The modeled discharges are all underestimated with a high relative error as shown in Table 5.39. As described in the Index of Performances section in the study,  $E_j$  values vary between 1 and  $-\infty$ . For this station,  $E_j$  comes out as negative, which means it is a poor fit for this station. RMSE, RRMSE and mean relative error  $\bar{\epsilon}_s$  are large and support the idea of poor fit.

Table 5.39. Empirical and predicted discharges, and relative errors for station DSI-2339

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\epsilon_{s,j}$
$q_2$	2	1.199	0.026	-0.979
$q_5$	5	0.652	0.028	-0.958
$q_8$	8	0.430	0.022	-0.949
$q_{10}$	10	0.383	0.016	-0.959
$q_{15}$	15	0.310	3.6E-03	-0.988
$q_{20}$	20	0.220	1.7E-03	-0.992
$q_{30}$	30	0.153	5.1E-03	-0.967
$q_{50}$	50	0.085	3.1E-03	-0.963
$q_{70}$	70	0.054	4.3E-03	-0.920
$q_{90}$	90	0.049	1.3E-04	-0.997
$q_{95}$	95	0.023	3.5E-05	-0.998
$q_{98}$	98	0.011	1.9E-06	-1.000

Table 5.40. Performance indices calculated for DSI-2339

Performance Index	Value
$E_j$	-0.708
$R$	N.C
$\bar{\epsilon}_s$	-0.970
$\sigma_{e,s}$	0.017
$E_s$	0.977
RRMSE	0.973
RMSE	0.452

- **EIE-2323**

EIE-2323 is located at the outlet of Oltu Basin; therefore, it has the largest area, 6978.73 km<sup>2</sup>. The equations and the index of performances carry an important role in these analyses. The coefficients do not give much idea about the conformity of the regional model. They are similar to the coefficients of models of stations DSI-2324 and DSI-2323. The errors and other criteria are the main tool to make a decision.

Table 5.41. Equations of  $q_j$  discharges obtained by omitting EIE-2323

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$1.052 \times 10^{-11} (A^{0.253})(LMR^{1.981})(BR^{-0.339})(MAP^{-1.136})(CN^{6.439})$
$q_5$	$8.65 \times 10^{-12} (A^{0.354})(LMR^{1.867})(BR^{-0.278})(MAP^{-1.533})(CN^{6.821})$
$q_8$	$6.917 \times 10^{-11} (A^{0.365})(LMR^{1.926})(BR^{-0.390})(MAP^{-1.682})(CN^{6.602})$
$q_{10}$	$1.91 \times 10^{-11} (A^{0.410})(LMR^{1.869})(BR^{-0.335})(MAP^{-1.654})(CN^{6.714})$
$q_{15}$	$1.083 \times 10^{-12} (A^{0.224})(LMR^{2.212})(BR^{0.012})(MAP^{-1.711})(CN^{6.793})$
$q_{20}$	$9.484 \times 10^{-13} (A^{0.04})(LMR^{2.570})(BR^{0.007})(MAP^{-1.654})(CN^{6.612})$
$q_{30}$	$5.107 \times 10^{-10} (A^{0.154})(LMR^{2.123})(BR^{-0.336})(MAP^{-0.917})(CN^{4.786})$
$q_{50}$	$1.284 \times 10^{-5} (A^{0.453})(LMR^{1.229})(BR^{0.327})(MAP^{-0.836})(CN^{1.574})$
$q_{70}$	$2.492 \times 10^{-5} (A^{0.633})(LMR^{0.995})(BR^{-0.037})(MAP^{-0.239})(CN^{1.007})$
$q_{90}$	$6.346 \times 10^{-12} (A^{1.150})(LMR^{0.436})(BR^{0.706})(MAP^{-0.911})(CN^{3.793})$
$q_{95}$	$7.921 \times 10^{-15} (A^{1.827})(LMR^{-0.705})(BR^{1.662})(MAP^{-1.972})(CN^{5.184})$
$q_{98}$	$6.344 \times 10^{-17} (A^{1.791})(LMR^{-0.688})(BR^{1.901})(MAP^{-1.868})(CN^{5.695})$

All of the discharges,  $q_j$ , are underestimated (see Table 5.42) but Table 5.43 supplies the information that the model is a good fit.  $E_j$  is not equal to 1 (a perfect fit) but it is 0.933 and this is a good value to indicate a good fit. Square root of coefficient of determination agrees with this idea with the value 0.966. However, root mean square error is quite large to be a good fit.



This is due to the larger discharges compared to discharges of other basins. When the equation of the RMSE is inspected, it is seen that the main part of the equation is the difference between the empirical and the predicted discharges. The difference gets greater as the basin gets larger.

Table 5.42. Empirical and predicted discharges and relative errors for station EIE-2323

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	168.00	135.036	-0.196
$q_5$	5	116.30	98.778	-0.151
$q_8$	8	92.40	78.927	-0.146
$q_{10}$	10	83.00	70.044	-0.156
$q_{15}$	15	60.95	54.125	-0.112
$q_{20}$	20	43.35	36.669	-0.154
$q_{30}$	30	26.85	19.177	-0.286
$q_{50}$	50	18.25	16.907	-0.074
$q_{70}$	70	14.65	13.246	-0.096
$q_{90}$	90	11.35	9.414	-0.171
$q_{95}$	95	9.91	7.239	-0.270
$q_{98}$	98	7.80	4.550	-0.417

Table 5.43. Performance indices calculated for EIE-2323

Performance Index	Value
$E_j$	0.933
$R$	0.966
$\bar{\varepsilon}_s$	-0.186
$\sigma_{\varepsilon,s}$	0.096
$E_s$	0.999
RRMSE	0.207
RMSE	13.206

- **EIE-2325**

EIE-2325 covers a large area as in the case DSI-2323. The  $b_1, b_2, \dots, b_6$  are considerably different from each other when EIE-2325 is cross-validated.

Table 5.44. Equations of  $q_j$  discharges obtained by omitting EIE-2325

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Equation of $q_j$
$q_2$	$0.002 (A^{0.612})(LMR^{0.134})(BR^{-0.290})(MAP^{2.142})(CN^{-1.474})$
$q_5$	$0.001 (A^{0.629})(LMR^{0.295})(BR^{-0.334})(MAP^{1.401})(CN^{-0.479})$
$q_8$	$0.002 (A^{0.640})(LMR^{0.388})(BR^{-0.43})(MAP^{1.242})(CN^{-0.320})$
$q_{10}$	$0.001 (A^{0.676})(LMR^{0.319})(BR^{-0.420})(MAP^{1.352})(CN^{-0.473})$
$q_{15}$	$0.0004 (A^{0.386})(LMR^{0.787})(BR^{-0.346})(MAP^{1.718})(CN^{-0.894})$
$q_{20}$	$2.01 \times 10^{-5} (A^{0.339})(LMR^{0.640})(BR^{-0.287})(MAP^{2.954})(CN^{-1.935})$
$q_{30}$	$0.00018 (A^{0.820})(LMR^{-0.537})(BR^{0.071})(MAP^{3.532})(CN^{-3.626})$
$q_{50}$	$7.28 \times 10^{-5} (A^{0.623})(LMR^{0.313})(BR^{0.206})(MAP^{1.621})(CN^{-1.507})$
$q_{70}$	$0.00011 (A^{0.869})(LMR^{-0.101})(BR^{-0.071})(MAP^{2.391})(CN^{-2.342})$
$q_{90}$	$2.33 \times 10^{-5} (A^{1.521})(LMR^{-1.782})(BR^{0.421})(MAP^{4.451})(CN^{-5.279})$
$q_{95}$	$1.36 \times 10^{-6} (A^{2.439})(LMR^{-3.786})(BR^{1.596})(MAP^{4.571})(CN^{-6.353})$
$q_{98}$	$1.97 \times 10^{-7} (A^{2.838})(LMR^{-4.927})(BR^{2.508})(MAP^{5.221})(CN^{-7.948})$

Discharges of EIE-2325 are highly overestimated station. The predicted discharges differ significantly from the observed ones.  $E_j$  takes an average value, which is not a good result. Rest of the values are also not good to be at least a fair fit.

Table 5.45. Empirical and predicted discharges and relative errors for station EIE-2325

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	37.000	55.675	0.505
$q_5$	5	25.450	33.348	0.310
$q_8$	8	20.400	32.315	0.584
$q_{10}$	10	17.500	22.033	0.259
$q_{15}$	15	12.000	21.212	0.768
$q_{20}$	20	8.120	15.330	0.888
$q_{30}$	30	5.200	8.383	0.612
$q_{50}$	50	3.900	4.938	0.266
$q_{70}$	70	3.060	3.810	0.245
$q_{90}$	90	1.230	2.392	0.944
$q_{95}$	95	0.564	1.298	1.301
$q_{98}$	98	0.296	0.783	1.644

Table 5.46. Performance indices calculated for EIE-2325

Performance Index	Value
$E_j$	0.514
$R$	0.717
$\bar{\varepsilon}_s$	0.694
$\sigma_{\varepsilon,s}$	0.44
$E_s$	0.995
RRMSE	0.813
RMSE	8.115

- **EIE-2329**

EIE-2329 has a large area, which equals to 3537.10 km<sup>2</sup>. The coefficients of the equation are alike with these of the equations for EIE-2323 or DSI-2324.

Table 5.47. Equations of qj discharges obtained by omitting EIE-2329

<b>Regional Model</b>	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
<b>Representative Equation</b>	$Q = 9.93 \times 10^{-7} (A^{0.68})(LMR^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
<b>q<sub>j</sub></b>	<b>Equation of q<sub>j</sub></b>
q <sub>2</sub>	$5.5 \times 10^{-9} (A^{0.494})(LMR^{1.187})(BR^{0.489})(MAP^{-1.186})(CN^{4.128})$
q <sub>5</sub>	$1.7 \times 10^{-10} (A^{0.779})(LMR^{0.897})(BR^{0.130})(MAP^{-1.321})(CN^{5.421})$
q <sub>8</sub>	$1.4 \times 10^{-9} (A^{0.616})(LMR^{1.198})(BR^{0.184})(MAP^{-1.491})(CN^{5.014})$
q <sub>10</sub>	$1.4 \times 10^{-10} (A^{0.735})(LMR^{1.040})(BR^{0.145})(MAP^{-1.337})(CN^{5.313})$
q <sub>15</sub>	$3.5 \times 10^{-11} (A^{0.411})(LMR^{1.738})(BR^{0.360})(MAP^{-1.799})(CN^{5.720})$
q <sub>20</sub>	$3.1 \times 10^{-7} (A^{0.720})(LMR^{1.484})(BR^{0.103})(MAP^{-2.940})(CN^{5.307})$
q <sub>30</sub>	$3.2 \times 10^{-8} (A^{2.213})(LMR^{-1.440})(BR^{-0.819})(MAP^{-0.200})(CN^{3.634})$
q <sub>50</sub>	$2.2 \times 10^{-5} (A^{1.947})(LMR^{-0.909})(BR^{-0.924})(MAP^{-0.705})(CN^{2.867})$
q <sub>70</sub>	$5.9 \times 10^{-5} (A^{1.628})(LMR^{-0.572})(BR^{-0.590})(MAP^{0.011})(CN^{1.208})$
q <sub>90</sub>	$1.8 \times 10^{-6} (A^{1.572})(LMR^{-0.382})(BR^{1.197})(MAP^{-1.979})(CN^{1.743})$
q <sub>95</sub>	$1.8 \times 10^{-13} (A^{1.981})(LMR^{-1.569})(BR^{3.144})(MAP^{-1.678})(CN^{2.264})$
q <sub>98</sub>	$5.1 \times 10^{-10} (A^{2.543})(LMR^{-2.687})(BR^{3.886})(MAP^{-2.61})(CN^{0.624})$

EIE-2329 gives better results than the station DSI-2325. Predicted and observed discharges are quite close; therefore, the relative errors are small. Furthermore, each entry in Table 5.48 point outs a perfect fit and appropriate model.

Table 5.48. Empirical and predicted discharges and relative errors for station EIE-2329

$q_j$	Duration (%)	$q_{s,j}$	$\hat{q}_{s,j}$	$\varepsilon_{s,j}$
$q_2$	2	88.470	84.443	-0.046
$q_5$	5	65.000	67.867	0.044
$q_8$	8	52.433	51.258	-0.022
$q_{10}$	10	46.158	46.451	0.006
$q_{15}$	15	30.957	30.819	-0.004
$q_{20}$	20	19.950	23.642	0.185
$q_{30}$	30	11.150	19.205	.722
$q_{50}$	50	7.643	12.925	0.691
$q_{70}$	70	6.138	8.343	0.359
$q_{90}$	90	3.680	3.864	0.050
$q_{95}$	95	2.440	2.047	-0.161
$q_{98}$	98	1.460	1.317	-0.098

Table 5.49. Performance indices calculated for EIE-2329

Performance Index	Value
$E_j$	0.985
$R$	0.992
$\bar{\varepsilon}_s$	0.144
$\sigma_{\varepsilon,s}$	0.295
$E_s$	0.9998
RRMSE	0.317
RMSE	3.534

Up to the present, all jackknifed cross-validated stations (DSI-2336, DSI-2335...etc.) are explored. Table 5.50 is the summary table for these stations, which includes areas and indices of performances of each.

As Table 5.50 shows, as the area gets bigger the performance and the goodness of fit results get better. The most problematic results arise at the station DSI-2339, which is the smallest one. In summary, up to now it is observed that the area of the basin plays an important role while implementing the regional model.

Furthermore, the jack-knife cross-validation of models in the form of Equation 5.8 produced  $E$ ,  $R$ ,  $RMSE$ ,  $RRMSE$  and  $\bar{\epsilon}_j$  values are calculated for each duration,  $j$ , and these are displayed in Table 5.51.

Table 5.50. Jack-knife indices of performance results for each station

	DSI-2336	DSI-2335	DSI-2324	DSI-2323	DSI-2337	DSI-2339	EIE-2323	EIE-2325	EIE-2329
<b>Area (km<sup>2</sup>)</b>	49.25	68.37	4537.55	1094.21	197.04	10.62	6978.73	1785.57	3537.10
<b>E<sub>j</sub></b>	0.100	0.297	0.994	0.976	0.993	-0.708	0.933	0.514	0.985
<b>R</b>	0.317	0.545	0.997	0.988	0.996	N.C	0.966	0.717	0.992
<b><math>\bar{\epsilon}_s</math></b>	-0.699	-0.604	0.086	-0.057	-0.106	-0.970	-0.186	0.694	0.144
<b><math>\sigma_{\epsilon,s}</math></b>	0.078	0.164	0.232	0.195	0.211	0.017	0.096	0.442	0.295
<b>E<sub>s</sub></b>	0.988	0.990	0.9999	0.99973	0.99991	0.977	0.999	0.995	0.9998
<b>RRMSE</b>	0.778	0.680	0.238	0.195	0.352	0.973	0.207	0.813	0.317
<b>RMSE</b>	1.224	1.113	2.902	1.573	0.264	0.452	13.206	8.115	3.534

Table 5.51. Jack-knife indices of performance results for each duration

$q_j$	<b>E</b>	<b>R</b>	<b>RMSE</b>	<b>RRMSE</b>	$\bar{\epsilon}_j$
$q_2$	0.939	0.969	13.648	0.498	-0.218
$q_5$	0.946	0.973	7.264	0.464	-0.253
$q_8$	0.890	0.943	6.552	0.478	-0.217
$q_{10}$	0.882	0.939	5.043	0.447	-0.260
$q_{15}$	-0.434	N.C	4.389	0.544	-0.236
$q_{20}$	0.540	0.735	3.812	0.579	-0.214
$q_{30}$	0.915	0.957	4.126	0.584	-0.112
$q_{50}$	0.988	0.994	1.972	0.449	0.029
$q_{70}$	0.997	0.999	1.041	0.432	-0.001
$q_{90}$	0.998	0.999	0.842	0.658	0.028
$q_{95}$	0.998	0.999	1.058	0.735	0.139
$q_{98}$	0.997	0.999	1.274	0.907	0.293

The worst results of E and R are for  $q_{15}$  and  $q_{20}$ . However, the average of mean relative errors of  $q_{15}$  and  $q_{20}$ ,  $\bar{\epsilon}_j$ , are not so different from others. Besides, from the hydropower engineering point of view durations between  $q_{30}$  and  $q_{95}$ , [ $q_{30}$ ,  $q_{95}$ ] are the most important ones. In addition, the RMSE of  $q_2$  is the largest but the other performance indices represent a good fit. If these indices are considered as a whole, it is seen that the regional model proposed is appropriate for Oltu Basin.

The relative errors obtained by comparing the jack-knifed FDCs and empirical FDCs are depicted in Figure 5.13. In this Figure, P1 (10%), P2 (25%), P3 (75%) and P4 (90%) are percentiles of mean relative errors.



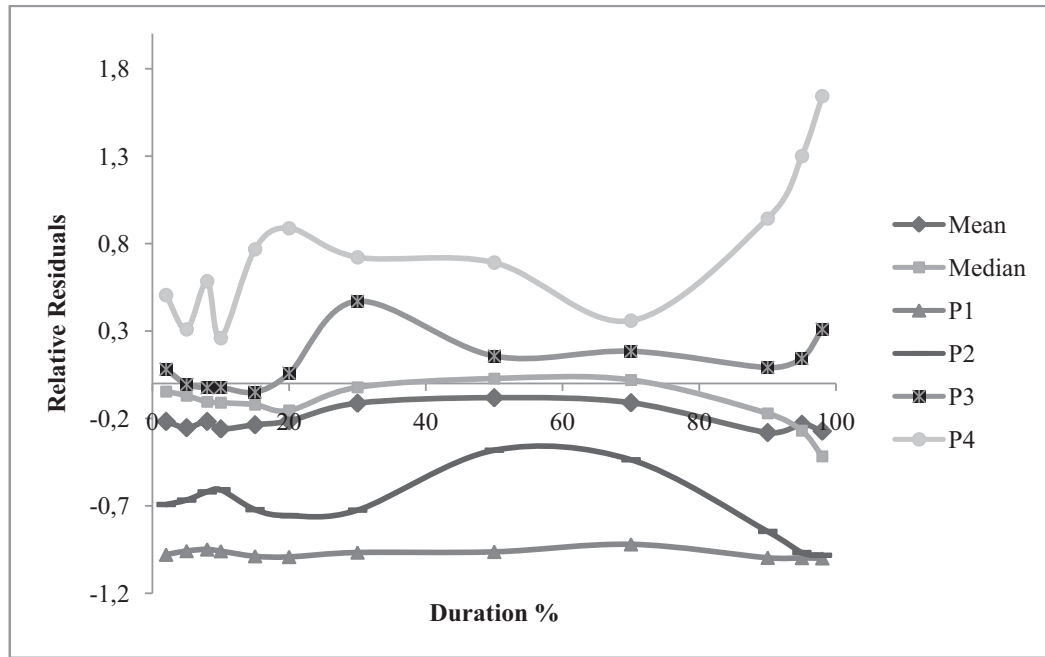


Figure 5.13. Jack-knifed relative errors distributed through stations  
(\*) P1 = 10%, P2 = 25%, P3 = 75%, P4 = 90%

## 5.5. Regional Model Of Oltu Basin

In the previous section jackknife cross-validation method has been applied and the goodness of fit of the model “ $Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$ ” has been checked according to performance indices. Since the results of the goodness of fit evaluations are reasonable and acceptable, a model has been developed for each  $q_j$ , as in the previous section and it has been named as model **Kocatepe**.

The model generating procedure is similar to the nonlinear regression analysis performed at the beginning of this chapter. The only differences are:

- The leading equation is known and it is directly applied (see Equation 5.9)
- Instead of mean annual discharge ( $Q$ ), empirical daily discharges for duration  $j$  ( $q_j$ ) are used as the dependent variable and then regression analysis is run.
- All nine stations have been included in the regression analysis

Table 5.52 is similar to the ones displayed in jack-knife cross-validation section. It depicts the equations for each  $q_j$  discharge.

Table 5.52. Equations of  $q_j$  discharges obtained for Model Kocatepe

Regional Model	$Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$
Representative Equation	$Q = 9.93 \times 10^{-7} (A^{0.68})(L^{0.964})(MAP^{-1.516})(BR^{0.423})(CN^{3.025})$
$q_j$	Model Kocatepe
$q_2$	$8.56 \times 10^{-7} (A^{0.593})(LMR^{1.149})(BR^{0.379})(MAP^{-1.913})(CN^{4.055})$
$q_5$	$3.36 \times 10^{-10} (A^{0.660})(LMR^{1.068})(BR^{0.260})(MAP^{-1.427})(CN^{5.241})$
$q_8$	$1.75 \times 10^{-9} (A^{0.666})(LMR^{1.141})(BR^{0.134})(MAP^{-1.543})(CN^{5.093})$
$q_{10}$	$5.51 \times 10^{-10} (A^{0.706})(LMR^{1.105})(BR^{0.177})(MAP^{-1.543})(CN^{5.224})$
$q_{15}$	$2.51 \times 10^{-11} (A^{0.414})(LMR^{1.738})(BR^{0.363})(MAP^{-1.802})(CN^{5.789})$
$q_{20}$	$1.77 \times 10^{-11} (A^{0.342})(LMR^{1.813})(BR^{0.525})(MAP^{-1.569})(CN^{5.229})$
$q_{30}$	$1.27 \times 10^{-7} (A^{0.812})(LMR^{0.426})(BR^{0.692})(MAP^{-0.331})(CN^{1.601})$
$q_{50}$	$2.68 \times 10^{-5} (A^{0.594})(LMR^{0.863})(BR^{0.536})(MAP^{-0.651})(CN^{0.931})$
$q_{70}$	$5.01 \times 10^{-6} (A^{0.857})(LMR^{0.395})(BR^{0.245})(MAP^{0.4})(CN^{0.226})$
$q_{90}$	$3.33 \times 10^{-5} (A^{1.432})(LMR^{-0.149})(BR^{1.344})(MAP^{-2.375})(CN^{1.407})$
$q_{95}$	$3.93 \times 10^{-10} (A^{2.398})(LMR^{-2.030})(BR^{2.653})(MAP^{-2.531})(CN^{2.256})$
$q_{98}$	$6.35 \times 10^{-5} (A^{2.671})(LMR^{-2.424})(BR^{3.874})(MAP^{-5.369})(CN^{1.413})$

Table 5.53 is a summary of an index of performance table as done in the jack-knife cross-validation analysis. The results for Model Kocatepe seem better but to see the differences, using a comparison table would be more useful. Therefore, Table 5.53 has been formed.

Table 5.53 is a simple comparison table. The grey colored rows are for the jack-knifed cross-validated results and the others are for the developed regional model, namely for Model Kocatepe. The comparison table tells important things. Firstly, the  $E$  values are greater than the  $E_j$  values, except for station DSI-2337. However, the other indices do not show a significant change, so it is an acceptable decrease for the Nash-Sutcliffe efficiency criterion. Larger  $E$  values mean the model works ideally for the basin. Similar to the Nash-Sutcliffe efficiency criterion, other indices are more meaningful for the regional model.

Secondly, the effect of area becomes more visible. As area gets larger, the  $E$  values get larger for the basins and the  $\bar{E}_S$ ,  $\sigma_{E,S}$ , **RRMSE** and **RMSE** get smaller as they should be.

Table 5.53. Comparison Table of the index of performances between jack-knifed and Kocatepe results for the stations

	DSI-2336	DSI-2335	DSI-2324	DSI-2323	DSI-2337	DSI-2339	EIE-2323	EIE-2325	EIE-2329
<b>Area</b> <b>(km<sup>2</sup>)</b>	49.25	68.37	4537.55	1094.21	197.04	10.62	6978.73	1785.57	3537.10
<b>E<sub>j</sub></b>	0.100	0.297	0.994	0.976	0.993	-0.708	0.933	0.514	0.985
<b>E</b>	0.108	0.477	0.99991	0.998	0.881	-0.700	0.99994	0.9994	0.99993
<b>R<sub>j</sub></b>	0.317	0.545	0.997	0.988	0.996	N.C	0.966	0.717	0.992
<b>R</b>	0.328	0.690	0.99996	0.999	0.939	N.C	0.99997	0.9997	0.99997
<b>(<math>\bar{\epsilon}_s</math>)<sub>j</sub></b>	-0.699	-0.604	0.086	-0.057	-0.106	-0.970	-0.186	0.694	0.144
<b><math>\bar{\epsilon}_s</math></b>	-0.672	-0.598	0.002	0.018	-0.332	-0.969	3.9E-04	0.011	0.001
<b>(<math>\sigma_{\epsilon,s}</math>)<sub>j</sub></b>	0.078	0.164	0.232	0.195	0.211	0.017	0.096	0.442	0.295
<b><math>\sigma_{\epsilon,s}</math></b>	0.082	0.153	0.005	0.013	0.230	0.016	0.004	0.009	0.004
<b>(E<sub>s</sub>)<sub>j</sub></b>	0.988	0.990	0.9999	0.99973	0.99991	0.977	0.999	0.995	0.9998
<b>E<sub>s</sub></b>	0.988	0.9903	0.9999991	0.99998	0.99852	0.977	0.9999994	0.9999939	0.9999992
<b>(RRMSE)<sub>j</sub></b>	0.778	0.680	0.238	0.195	0.352	0.973	0.207	0.813	0.317
<b>RRMSE</b>	0.759	0.678	0.005	0.022	0.465	0.971	0.004	0.013	0.004
<b>(RMSE)<sub>j</sub></b>	1.224	1.113	2.902	1.573	0.264	0.452	13.206	8.115	3.534
<b>RMSE</b>	1.219	1.116	0.360	0.428	1.055	0.451	0.400	0.274	0.240

\*E<sub>j</sub>: Nash-Sutcliffe efficiency criterion for Jack-knife Cross-validation

Table 5.54. Index of performances of jack-knifed cross-validation results and Model Kocatepe results for each duration

$q_j$	$E_j$	$E$	$R_j$	$R$	$(RRMSE)_j$	$RRMSE$	$(RMSE)_j$	$RMSE$	$(\bar{e}_j)_j$	$\bar{e}_j$
$q_2$	0.939	0.99885	0.969	0.99942	13.648	0.474	0.498	1.881	-0.218	-0.287
$q_5$	0.946	0.99875	0.973	0.99938	7.264	0.445	0.464	1.107	-0.253	-0.272
$q_8$	0.890	0.99811	0.943	0.99905	6.552	0.437	0.478	0.859	-0.217	-0.259
$q_{10}$	0.882	0.99772	0.939	0.99886	5.043	0.446	0.447	0.701	-0.260	-0.266
$q_{15}$	-0.434	0.98475	N.C	0.99235	4.389	0.470	0.544	0.454	-0.236	-0.284
$q_{20}$	0.540	0.99595	0.735	0.99797	3.812	0.494	0.579	0.358	-0.214	-0.308
$q_{30}$	0.915	0.99972	0.957	0.99986	4.126	0.490	0.584	0.238	-0.112	-0.310
$q_{50}$	0.988	0.9999928	0.994	0.9999964	1.972	0.374	0.449	0.048	0.029	-0.107
$q_{70}$	0.997	0.9999969	0.999	0.9999984	1.041	0.361	0.432	0.035	-0.001	-0.102
$q_{90}$	0.998	0.9999930	0.999	0.9999965	0.842	0.569	0.658	0.057	0.028	-0.110
$q_{95}$	0.998	0.9999963	0.999	0.9999982	1.058	0.628	0.735	0.043	0.139	-0.135
$q_{98}$	0.997	0.9999976	0.999	0.9999988	1.274	0.650	0.907	0.036	0.293	0.006

A comparison table similar to the one Table 5.53 has been constructed for durations ( $j = 2, 5, 8, \dots, 98$ ) this time (see Table 5.54). The grey colored columns are for the jack-knifed cross-validated results and the other ones are for Model Kocatepe. The trend for the durations is alike to the stations, Nash-Sutcliffe efficiencies get larger, and RRMSE and RMSE get smaller. The model Kocatepe works efficiently also for the FDCs.

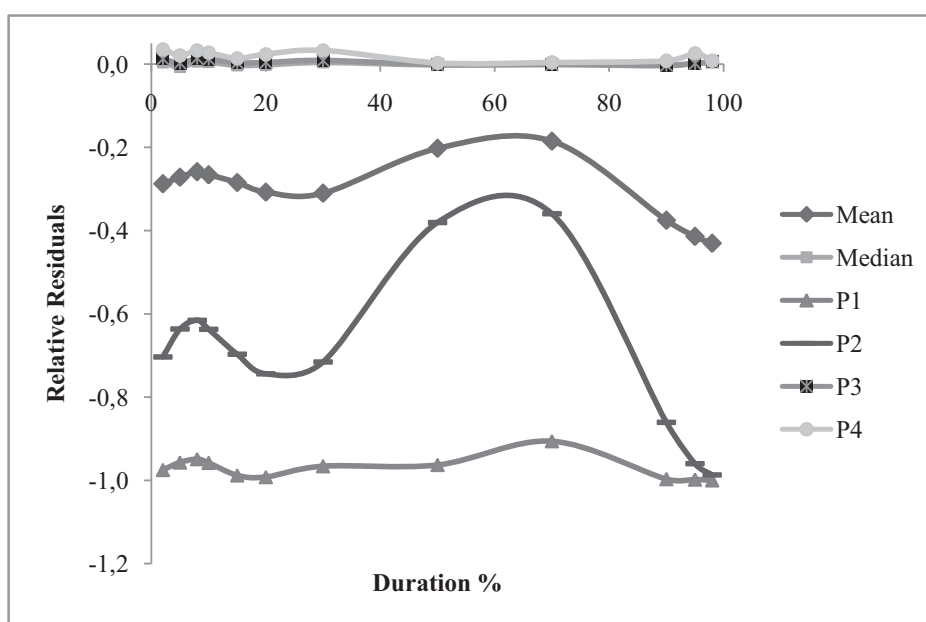


Figure 5.14. Model Kocatepe relative errors distributed through station

## 5.6. Quimpo's Parametric Approach

In the literature survey chapter, several regional parametric approaches have been explained. One of the oldest and widely used approaches proposed by Quimpo et al. (1983) is also explained in this section. In this section, the parametric model suggested by Quimpo et al. (1983) is applied to Oltu Basin data to see:

- how good it fits to the data of a basin in Turkey
- a different model application
- the difference between the models since one of them completely depend on area and climatic regimes only, and the Model Kocatepe uses three more parameters in addition to area and climatic regimes.

The motivation of Quimpo et al. (1983) is the hydroelectric power generation in the Philippines, since a hydropower development program is initiated in the country. However, there are similar problems to obtain water availability data as in Turkey. Therefore, a regionalization method is generated (Quimpo et al., 1983).

To define the tail behavior of a FDC some mathematical models are used. Two of them are:

- Negative Exponential Type  
$$F(\chi_i) = A \exp(\lambda \chi_i) \quad (\text{Equation 5.10})$$

- Power Type  
$$F(\chi_i) = a \chi_i^\beta \quad (\text{Equation 5.11})$$

In the light of these two equations, Quimpo et al. (1983) suggested a regional parametric model, which is:

$$Q = Q_A \exp(-cD) \quad (\text{Equation 5.12})$$

$$\text{and } Q_A = pA^{-m} \quad (\text{Equation 5.13})$$

where;

- $Q$  ; discharge per unit area of the basin,  $(\text{m}^3/\text{s})/\text{km}^2$
- $D$  ; percent of time that  $Q$  has been exceeded during the period of record
- $c$  ; constant, regionalization parameter, vary regularly around country along the lines of its climatic properties
- $Q_A$  ; constant, depends on the basin area
- $A$  ; drainage area,  $\text{km}^2$
- $p, m$ ; constants

For Oltu Basin, regionalized parametric approach proposed by Quimpo et al. (1983) has been applied. It is a familiar procedure to apply. The Equation has been put into a nonlinear regression implementation in SPSS. Q values (discharge per unit area) and durations used in Model Kocatepe (12 different durations; 0.02, 0.05, 0.05,..., 0.95, 0.98) have been regressed and after several iterations the values of constants  $Q_A$  and  $c$  have been reached for each basin (Table 5.55).

Table 5.55.  $Q_A$  and  $c$  values for each basin for Quimpo's et al. regional model

Basin	Area (km <sup>2</sup> )	$Q_A$	$c$
DSI-2336	49.25	0.098	10.605
DSI-2335	68.37	0.073	10.665
DSI-2324	4537.55	0.029	7.369
DSI-2323	1094.21	0.033	8.372
DSI-2337	197.04	0.056	9.111
DSI-2339	10.62	0.134	12.558
EIE-2323	6978.73	0.026	7.371
EIE-2325	1785.57	0.028	7.845
EIE-2329	3537.1	0.023	8.294

#### 5.6.1. Regional Index, $c$

Quimpo et al. (1983) stated that the regional parameter for their model is the index  $c$ . The  $c$  values obtained in the previous analysis are interpolated between contour lines. A similar spatial variation analysis is performed for the  $c$  values obtained for Oltu basin. The  $c$  values obtained in the previous analysis (see Table 5.55) have been interpolated between contour lines by using ArcGIS. The resulting map is shown in Figure 5.15.

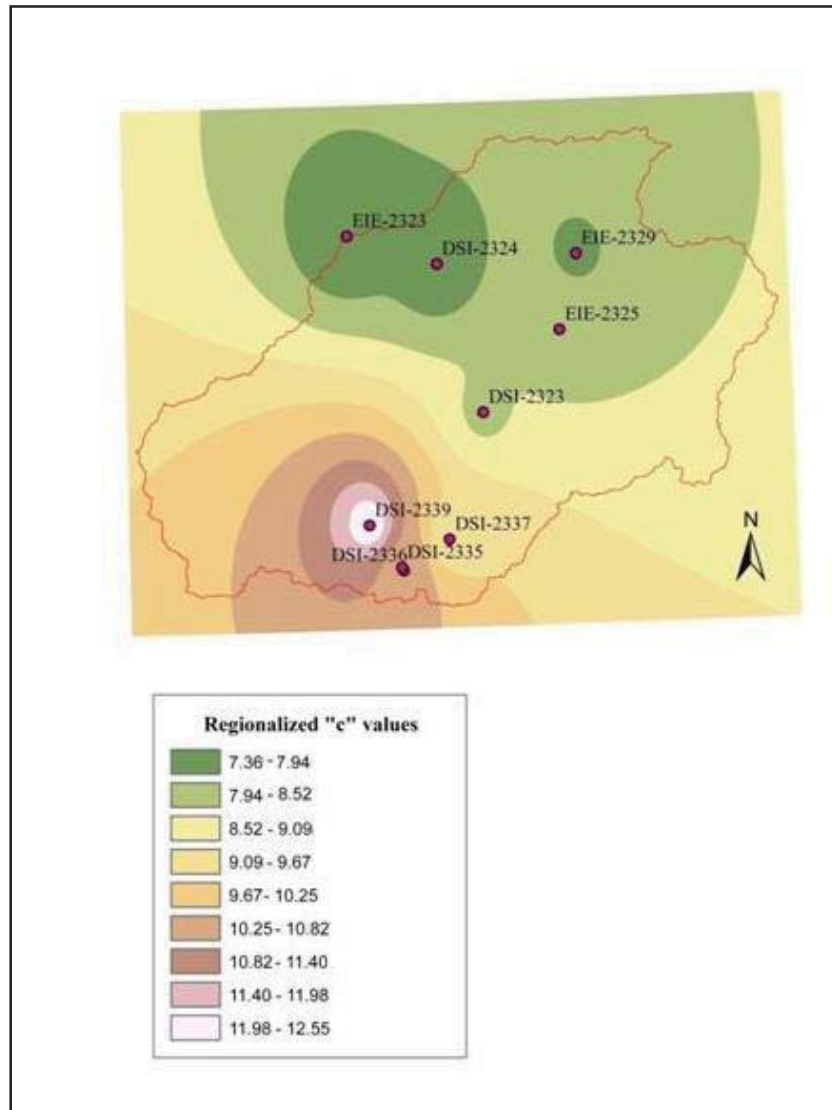


Figure 5.15. Regionalized values of “c”

### 5.6.2. Area Analysis, $Q_A$

According to the Equation 5.14, there is a consequential relation between the constant  $Q_A$  and basin area  $A$ . Quimpo et al. (1983) made a regression analysis for the drainage areas greater than  $100 \text{ km}^2$ . The reasons of the limitation  $100 \text{ km}^2$  are; short record lengths and high variability of data from small basins. After regression analysis, values found as

$$p = 4.514 \text{ and } m = 0.447 \text{ with a correlation coefficient, } R = 0.961.$$



For this study a similar nonlinear regression analysis has been performed. However, this time two regression analyses have been performed. The first of the regressions has been performed with all watersheds including the ones below 100 km<sup>2</sup>, and the second one has been done with in only the ones larger than 100 km<sup>2</sup>.

**1<sup>st</sup> nonlinear multiple regression:** All basins are included. Table 5.56 shows the basins included and the values of  $Q_A$ ,  $c$ ,  $Q_{A,p}$ ,  $p$  and  $m$ .

Table 5.56. Basins included in area regression analysis for Quimpo approach

Station	Area (km <sup>2</sup> )	$Q_A$	$c$	$Q_{A,p}$
DSI-2336	49.25	0.098	10.605	0.087
DSI-2335	68.37	0.073	10.665	0.079
DSI-2324	4537.55	0.029	7.369	0.025
DSI-2323	1094.21	0.033	8.372	0.036
DSI-2337	197.04	0.056	9.111	0.059
DSI-2339	10.62	0.134	12.558	0.133
EIE-2325	1785.57	0.023	8.294	0.032
EIE-2329	3537.10	0.028	7.845	0.026
EIE-2323	6978.73	0.026	7.371	0.022
<b>p</b>	0.257			
<b>m</b>	0.279			

For equation  $Q_A = pA^{-m}$ ,  $p$  and  $m$  constants come out as 0.257 and 0.279, respectively.

**2<sup>nd</sup> nonlinear multiple regression:** Stations smaller than 100 km<sup>2</sup> are not included. Therefore, DSI-2336, DSI-2335 and DSI-2339 are excluded from the data set.

For this case, the constants of Equation 5.13 are 0.238 and 0.271 as can be seen in Table 5.57.

Table 5.57. Basins included in area regression analysis for Quimpo approach

Station	Area (km <sup>2</sup> )	Q <sub>A</sub>	c	Q <sub>A,p</sub>
DSI-2324	4537.55	0.029	7.369	0.024
DSI-2323	1094.21	0.033	8.372	0.036
DSI-2337	197.04	0.056	9.111	0.057
EIE-2323	6978.73	0.026	7.371	0.022
EIE-2329	3537.10	0.028	7.845	0.026
EIE-2325	1785.57	0.023	8.294	0.031
<b>p</b>	0.238			
<b>m</b>	0.271			

To decide on which case is better Nash-Sutcliffe efficiency criterion, coefficient of determination R, relative root mean square error RRMSE, root mean square error RMSE, mean relative error and standard deviation of it have been calculated and summarized in Table 5.58.

Table 5.58. Comparison of cases for Quimpo et al. area equation

Performance Index	Case 1	Case 2
<b>E</b>	0.974	0.820
<b>R</b>	0.987	0.906
<b>RRMSE</b>	0.161	0.180
<b>RMSE</b>	0.006	0.005
$\bar{\epsilon}_s$	0.012	0.008
$\sigma_{\epsilon,s}$	0.170	0.197

Table 5.58 allows us to see the fact that Case 1 is better than Case 2. E and R are higher and left of the indices are smaller than Case 2. It is appropriate to choose the constants from Case 1. As a result, the equation becomes:

$$Q_A = 0.257 (A^{-0.279}) \quad (\text{Equation 5.14})$$

After that, as done in all statistical analysis the performance indices have been calculated for each basin and duration (Table 5.59). The following Tables 5.59 and 5.60 summarize these results.

When the results of performance analysis inspected (see Table 5.59) it is noticed that, the regional model of Quimpo et al. (1983) works better in small basins compared to Model Kocatepe. For example, the E and R-values of basins with area under 75 km<sup>2</sup> are greater than for the ones above 75 km<sup>2</sup>. However, the RMSE, RRMSE and mean relative errors are relatively greater than the ones for Model Kocatepe (see Table 5.53) for large basins. In brief, it can be concluded that, Model Kocatepe is better for large areas, whereas Quimpo et al. approach is better for small ones.

Table 5.59. Index of performances for regional Model Quimpo

	Area (km <sup>2</sup> )	E	R	$\bar{\epsilon}_s$	$\sigma_{\epsilon,s}$	RRMSE	RMSE
<b>DSI-2336</b>	49.25	0.989	0.994	-0.256	0.392	0.632	0.136
<b>DSI-2335</b>	68.37	0.989	0.994	-0.191	0.330	0.641	0.136
<b>DSI-2324</b>	4537.55	0.982	0.991	-0.385	0.465	0.589	5.172
<b>DSI-2323</b>	1094.21	0.990	0.995	-0.405	0.474	0.608	1.028
<b>DSI-2337</b>	197.04	0.995	0.998	-0.360	0.453	0.609	0.210
<b>DSI-2339</b>	10.62	0.942	0.971	-0.172	0.367	0.704	0.083
<b>EIE-2323</b>	6978.73	0.965	0.982	-0.401	0.475	0.606	9.533
<b>EIE-2325</b>	1785.57	0.978	0.989	-0.415	0.476	0.617	1.722
<b>EIE-2329</b>	3537.10	0.988	0.994	-0.389	0.475	0.598	3.094

Table 5.60. Quimpo et al. regional model indices of performance results for each duration

<b>q<sub>j</sub></b>	<b>E</b>	<b>R</b>	<b>RMSE</b>	<b>RRMSE</b>
q <sub>2</sub>	0.992	0.996	4.990	0.051
q <sub>5</sub>	0.985	0.992	3.894	0.082
q <sub>8</sub>	0.973	0.986	3.260	0.088
q <sub>10</sub>	0.989	0.995	1.532	0.038
q <sub>15</sub>	0.948	0.973	0.839	0.112
q <sub>20</sub>	0.975	0.987	0.896	0.185
q <sub>30</sub>	0.963	0.981	2.736	0.398
q <sub>50</sub>	0.879	0.938	6.192	0.802
q <sub>70</sub>	0.902	0.950	6.116	0.950
q <sub>90</sub>	0.955	0.977	4.555	0.984
q <sub>95</sub>	0.973	0.986	3.731	0.983
q <sub>98</sub>	0.985	0.993	2.846	0.979

To conclude this section, the regional model proposed by Quimpo et al. is an appropriate model to use in small basins. The model is evaluated for Oltu Basin, the equations and regional indices are obtained. They will be used in “Validation” chapter to see their performances.

## 5.7. Statistical Approach

Statistical approach is one of the methods used in regionalization of FDCs. The main idea behind the statistical approach is to fit a cumulative probability density function to the observed flow series and estimate the distribution parameters by designating regional parameters, such as topographic, climatic and geomorphologic parameters, to the distribution estimators.

Fennessey and Vogel (1990) proposed that a two-parameter lognormal function is a good approximation to the lower half ( $0.5 \leq p \leq 0.99$ ) of daily flow-duration curves in Massachusetts. This proposal seems to be rational to apply for FDCs of Oltu, since for many natural phenomena lognormal distribution is accepted instead of a normal distribution.

Assuming the lognormal distribution is valid for daily discharges (Q), the probability of exceedance p can be defined with the Equations 2.5, 2.6 and 2.7.

$z_p$  is the  $p^{\text{th}}$  percentile of a zero mean,  $\mu$  and  $\sigma$  are the mean and variance of the natural logarithms of daily discharges (Fennessey and Vogel, 1990).

Firstly, the estimators  $\mu$  and  $\sigma$  have been obtained for daily discharges belong to  $0.5 \leq p \leq 0.99$  in the FDCs. The estimators have been calculated for all the stations (modeling stations) by using distribution software EasyFit.

Table 5.61. Lognormal parameter estimators for modeling stations

Stations	Area (km <sup>2</sup> )	LMR (km)	BR (m)	MAP (mm)	CN	Q (m <sup>3</sup> /s)	$\mu$	$\sigma$
DSI-2336	49.25	13.37	364.54	423.61	86.62	0.6	-2.37	0.47
DSI-2335	68.37	12.80	391.90	430.30	88.99	0.6	-2.72	0.42
DSI-2324	4537.55	139.73	1256.10	549.25	74.29	23.3	2.01	0.45
DSI-2323	1094.21	59.13	1067.50	481.40	80.38	5.5	0.37	0.45
DSI-2337	197.04	27.10	615.60	434.05	87.69	1.5	-1.38	0.37
DSI-2339	10.62	2.07	113.90	441.00	86.21	0.2	-3.09	0.45
EIE-2323	6978.73	161.22	1513.12	563.52	72.19	34.3	2.61	0.19
EIE-2325	1785.57	83.53	977.10	518.47	73.49	7.1	0.82	0.42
EIE-2329	3537.10	104.49	1033.90	515.20	76.50	16.5	1.61	0.43

The next step is the regionalization of these estimators. Two regional predictive models for  $\hat{\mu}$  and  $\hat{\sigma}$  would be developed. The technique used in the parametric model development would be used for statistical approach. The basin parameters (A, LMR, BR, MAP and CN) have been used in the regression as an independent variable against the dependent variables  $\mu$  and  $\sigma$ , instead of  $q_j$  discharges or mean annual discharges, Q.

After the regression analysis, two predictive models would be reached as the regional statistical models.

Fennessey and Vogel (1990) recommended the models:

$$\hat{\mu} = 1.0088 \ln(A) + \epsilon \quad (\text{Equation 5.18})$$

$$R^2 = 0.998, \sigma_{\epsilon} = 0.1343$$

$$\hat{\sigma} = 1.10 + \frac{271}{H} + \eta \quad (\text{Equation 5.19})$$

$$R^2 = 0.720, \sigma_{\eta} = 0.2013$$

The equations 5.18 and 5.19 indicate that, the area (A) and the basin relief (H) are good indicators for Massachusetts.

Moreover, Castellarin et al. (2004) considered the statistical models are in following the form:

$$\hat{\theta} = A_0 + A_1 \ln(X_1) + A_2 \ln(X_2) + \dots + A_n \ln(X_n) + \vartheta$$

and the equations:

$$\hat{\mu} = -11.255 + 1.117 \ln(A) + 0.808 \ln(\Delta H) + 0.108 \ln(MANP) + \vartheta' \quad (\text{Equation 5.20})$$

E = 0.968

$$\hat{\sigma} = 3.155 - 1.238 \ln(P) - 0.419 \ln(\Delta H) + 0.209 \ln(MANP) + \vartheta'' \quad (\text{Equation 5.21})$$

E = 0.528

$\hat{\mu}$  and  $\hat{\sigma}$  are in  $\ln(\text{m}^3/\text{s})$ , area (A) in  $\text{km}^2$ , basin relief ( $\Delta H$ ) in m, mean annual net precipitation (MANP) in mm and permeable portion of basin area (P) in %.

In addition to these two approaches, several linear and nonlinear models have been tried and these are listed in Table 5.62 and in Table 5.63 with corresponding performance indices.

Table 5.62 displays the trial models for the lognormal distribution estimator,  $\hat{\mu}$ . First equation is the application of the model proposed by Fennessey and Vogel (1990). However, the results are poor for Oltu basin in this case. The bolded equations (row 6 and row 11) seem to be reasonable. The equation at row 6 is a simple linear model and row 11 is the implementation of a natural logarithm type equation, like suggested by Castellarin et al. (2004). If the variance table is inspected (see Table 5.63) the equation which uses the same parameters with the one for mean ( $\hat{\mu}$ ), equation at row 6, shows a good fit for the variance,  $\hat{\sigma}$ , although, the performance indices of the one with natural logarithms (see Table 5.66, row 11) are inadequate.

Table 5.62.  $\hat{\mu}$  models and performance indices

Models developed for $\hat{\mu}$		E	R	RMSE	RRMSE
$\hat{\mu} = 0.069 \ln(A)$		0.034	0.185	2.147	0.877
$\hat{\mu} = 0.014 (A^{1.395}) (L^{1.419}) (BR^{-3.5}) (MAP^{2.527}) (CN^{-1.097})$		0.353	0.594	1.757	0.720
$\hat{\mu} = -1.831 + 0.00785(A)$		-111.880	N.C	23.216	12.230
$\hat{\mu} = -3.097 - 0.000617 (A) + 0.61 (LMR)$		-534.890	N.C	50.584	40.393
$\hat{\mu} = -3.526 - 0.00041 (A) + 0.39 (LMR) + 0.002 (BR)$		-220.309	N.C	32.507	26.149
$\hat{\mu} = -5.345 - 0.00037 (A) + 0.033 (LMR) + 0.002 (BR) + 0.004 (MAP)$		<b>0.981</b>	<b>0.990</b>	<b>0.303</b>	<b>0.189</b>
$\hat{\mu} = 13.474 - 0.000282 (A) + 0.036 (LMR) + 0.02 (BR) - 0.016 (MAP) - 0.116 (CN)$		-65.225	N.C	17.782	20.164
$\hat{\mu} = -5.983 + 0.928 \ln (A)$		0.968	0.984	0.393	0.203
$\hat{\mu} = -6.167 + 1.577 \ln (A) - 1.066 \ln (LMR)$		0.985	0.992	0.268	0.175
$\hat{\mu} = -1.776 + 1.451 \ln (A) - 0.313 \ln (LMR) - 0.975 \ln (BR)$		0.987	0.994	0.247	0.133
$\hat{\mu} = -32.878 + 0.943 \ln (A) - 0.201 \ln (LMR) - 0.302 \ln (BR) + 4.773 \ln (MAP)$		<b>0.990</b>	<b>0.995</b>	<b>0.218</b>	<b>0.148</b>
$\hat{\mu} = -20.037 + 0.972 \ln (A) - 0.273 \ln (LMR) - 0.247 \ln (BR) + 3.650 \ln (MAP) - 1.409 \ln (CN)$		0.990	0.995	0.217	0.152
$\hat{\mu} = -32.772 + 0.880 \ln (A) - 0.488 \ln (BR) + 4.890 \ln (MAP)$		0.989	0.995	0.225	0.202

Table 5.63.  $\hat{\sigma}$  models and performance indices

	<b>Models developed for <math>\hat{\sigma}</math></b>	<b>E</b>	<b>R</b>	<b>RMSE</b>	<b>RRMSE</b>
$\hat{\sigma} = 0.384 + 9.818/\text{BR}$		0.079	0.281	0.082	0.362
$\hat{\sigma} = 2167.771 (A^{0.017}) (L^{0.381}) (\text{BR}^{-0.652}) (\text{MAP}^{-1.195}) (\text{CN}^{0.351})$		0.284	0.533	0.072	0.287
$\hat{\sigma} = 0.453 - 2.3 \times 10^{-5}(A)$		0.467	0.683	0.062	0.213
$\hat{\sigma} = 0.404 - 7.7 \times 10^{-5}(A) + 0.002 (\text{LMR})$		0.500	0.707	0.060	0.132
$\hat{\sigma} = 0.462 - 0.0001(A) + 0.005 (\text{LMR}) - 0.00025 (\text{BR})$		0.721	0.849	0.045	0.099
<b><math>\hat{\sigma} = 0.590 - 0.000107 (A) + 0.006 (\text{LMR}) - 0.000258 (\text{BR}) - 0.0003 (\text{MAP})</math></b>		<b>0.718</b>	<b>0.847</b>	<b>0.045</b>	<b>0.141</b>
$\hat{\sigma} = 0.401 - 0.000108 (A) + 0.006 (\text{LMR}) - 0.000257 (\text{BR}) - 9.8 \times 10^{-5} (\text{MAP}) + 0.001 (\text{CN})$		0.739	0.860	0.044	0.128
$\hat{\sigma} = 0.504 - 0.016 \ln(A)$		0.189	0.435	0.077	0.319
$\hat{\sigma} = 0.510 - 0.036 \ln (A) + 0.033 \ln (\text{LMR})$		0.205	0.453	0.076	0.314
$\hat{\sigma} = 1.253 - 0.057 \ln (A) + 0.160 \ln (\text{LMR}) - 0.165 \ln (\text{BR})$		0.242	0.492	0.074	0.304
<b><math>\hat{\sigma} = 7.93 + 0.052 \ln (A) + 0.136 \ln (\text{LMR}) - 0.309 \ln (\text{BR}) - 1.025 \ln (\text{MAP})</math></b>		<b>0.326</b>	<b>0.571</b>	<b>0.070</b>	<b>0.264</b>
$\hat{\sigma} = 16.435 + 0.071 \ln (A) + 0.088 \ln (\text{LMR}) - 0.273 \ln (\text{BR}) - 1.768 \ln (\text{MAP}) - 0.933 \ln (\text{CN})$		0.358	0.598	0.068	0.254



In brief, the preferable equations of the  $\hat{\mu}$  are;

- $\hat{\mu} = -5.345 - 0.00037 (A) + 0.033 (LMR) + 0.002 (BR) + 0.004 (MAP)$   
E = 0.981, R = 0.990, RMSE = 0.303, RRMSE = 0.189
- $\hat{\mu} = -32.878 + 0.943 \ln (A) - 0.201 \ln (LMR) - 0.302 \ln (BR) + 4.773 \ln (MAP)$   
E = 0.990, R = 0.995, RMSE = 0.218, RRMSE = 0.148

More importantly, instead of the high performance indices, the parameters included in the equations have been decided on equations as final choices. They include reasonable number of explanatory variables and their performance indices are higher than the ones with more explanatory variables.

The preferable equations of the  $\hat{\sigma}$  are;

- $\hat{\sigma} = 0.590 - 0.000107 (A) + 0.006 (LMR) - 0.000258 (BR) - 0.0003 (MAP)$   
E = 0.718, R = 0.847, RMSE = 0.045, RRMSE = 0.141
- $\hat{\sigma} = 7.93 + 0.052 \ln (A) + 0.136 \ln (LMR) - 0.309 \ln (BR) - 1.025 \ln (MAP)$   
E = 0.326, R = 0.571, RMSE = 0.070, RRMSE = 0.264

These equations are chosen since they are parallel to the ones chosen for  $\hat{\mu}$ , and due to the reason of adequate number of parameters. However, the second one in the form of natural logarithm is a poor model. Therefore, the equations of the lognormal statistical approach estimators are;

$$\hat{\mu} = -5.345 - 0.00037 (A) + 0.033 (LMR) + 0.002 (BR) + 0.004 (MAP) \quad (\text{Equation 5.23})$$

$$E = 0.981, R = 0.990, RMSE = 0.303, RRMSE = 0.189$$

and

$$\hat{\sigma} = 0.590 - 0.000107 (A) + 0.006 (LMR) - 0.000258 (BR) - 0.0003 (MAP) \quad (\text{Equation 5.24})$$

$$E = 0.718, R = 0.847, RMSE = 0.045, RRMSE = 0.141$$

The resulting estimates  $\hat{\mu}$  and  $\hat{\sigma}$  are summarized in Table 5.64.

Furthermore, these equations have to be inverted to obtain a direct estimate of the  $p^{\text{th}}$  quantiles of average daily discharges  $q_p$ . The following equation is being used for this inversion (Fennessey and Vogel, 1990):

$$q_p = \exp (\hat{\mu} + z_p \hat{\sigma}) \quad (\text{Equation 5.25})$$

and Tukey (1960) proposed an approximation to  $z_p$  which is

$$z_p = 4.91[(1-p)^{0.14} - p^{0.14}] \quad (\text{Equation 5.26})$$

By using Equation 5.25 and Equation 5.26 the estimated flow duration curves have been obtained using the parameters  $\hat{\mu}$  and  $\hat{\sigma}$ . The lower part resultant FDCs and the empiric FDCs of every station in the basin are shown in Figure 16 – Figure 29.

Table 5.64. Calculated and regionalized basin estimators

Stations	Area (km <sup>2</sup> )	$\mu$	$\hat{\mu}$	$\sigma$	$\hat{\sigma}$
DSI-2336	49.25	-2.37	-2.50	0.47	0.44
DSI-2335	68.37	-2.72	-2.44	0.42	0.43
DSI-2324	4537.55	2.01	2.30	0.45	0.45
DSI-2323	1094.21	0.37	0.26	0.45	0.41
DSI-2337	197.04	-1.38	-1.56	0.37	0.44
DSI-2339	10.62	-3.09	-3.29	0.45	0.44
EIE-2323	6978.73	2.61	2.67	0.19	0.25
EIE-2325	1785.57	0.82	0.78	0.42	0.49
EIE-2329	3537.10	1.61	0.92	0.43	0.42
DSI-2313	6967.93	2.84	2.50	0.18	0.25
DSI-2321	1851.19	-0.12	0.99	0.82	0.53
DSI-2322	3501.01	1.56	0.83	0.33	0.40
DSI-2333	47.20	-2.79	-1.86	0.47	0.36
DSI-2338	72.83	-3.71	-0.84	0.52	0.26

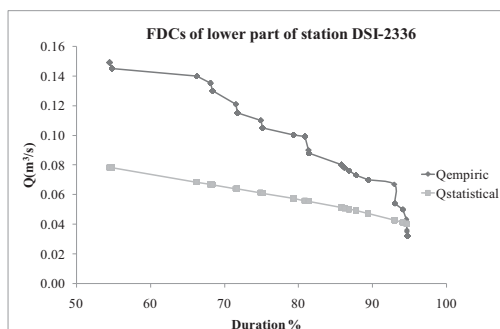


Figure 5.16. Comparison of observed FDC and Regionalized FDC at station DSI-2336

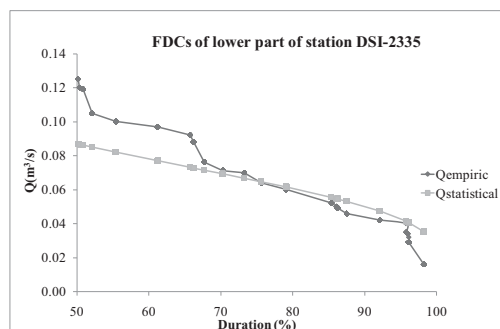


Figure 5.17. Comparison of observed FDC and Regionalized FDC at station DSI-2335

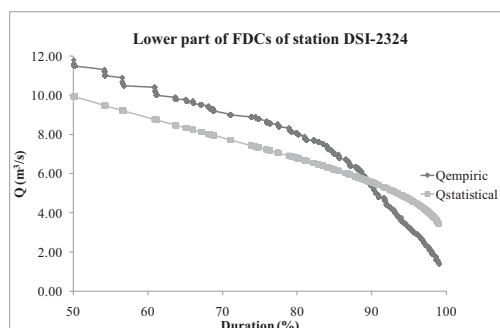


Figure 5.18. Comparison of observed FDC and Regionalized FDC at station DSI-2324

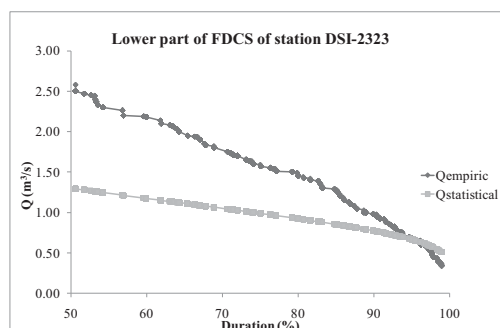


Figure 5.19. Comparison of observed FDC and Regionalized FDC at station DSI-2323

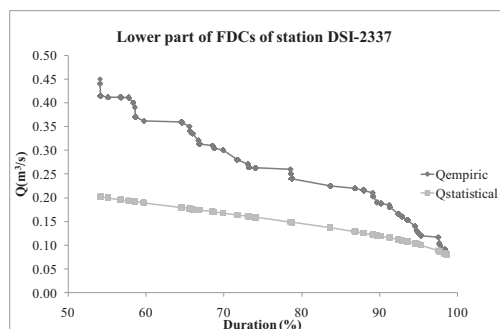


Figure 5.20. Comparison of observed FDC and Regionalized FDC at station DSI-2337

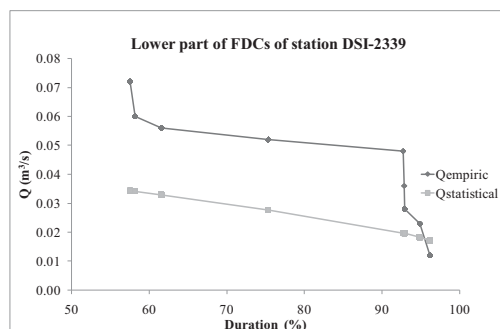


Figure 5.21. Comparison of observed FDC and Regionalized FDC at station DSI-2339

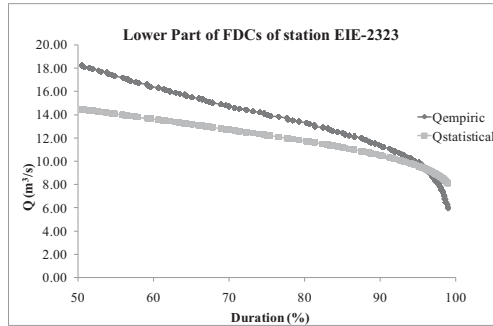


Figure 5.22. Comparison of observed FDC and Regionalized FDC at station EIE-2323

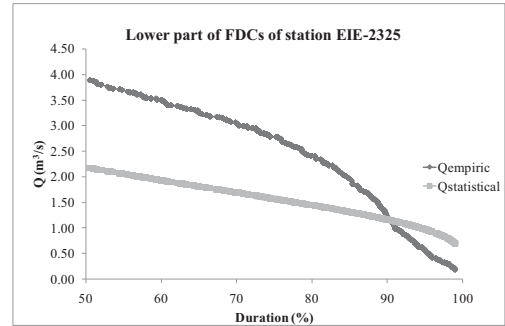


Figure 5.23. Comparison of observed FDC and Regionalized FDC at station EIE-2325

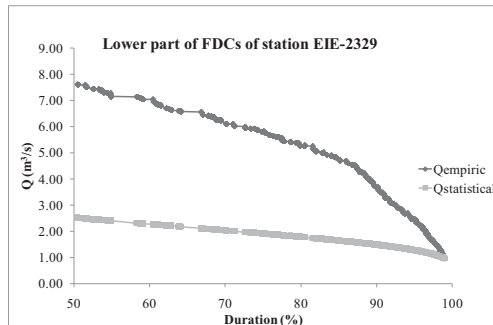


Figure 5.24. Comparison of observed FDC and Regionalized FDC at station EIE-2329

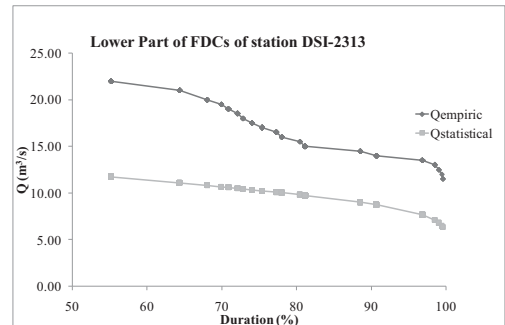


Figure 5.25. Comparison of observed FDC and Regionalized FDC at station DSI-2313

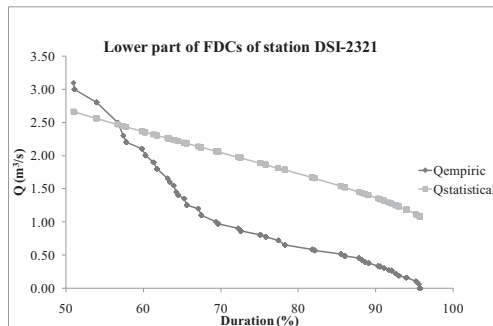


Figure 5.26. Comparison of observed FDC and Regionalized FDC at station DSI-2321

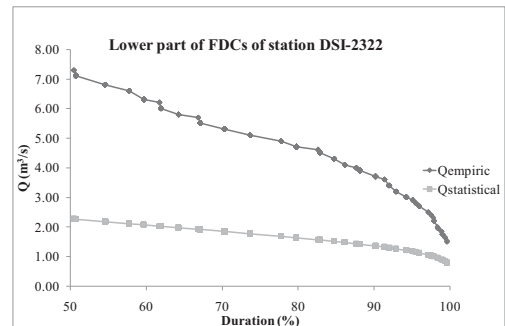


Figure 5.27. Comparison of observed FDC and Regionalized FDC at station DSI-2322

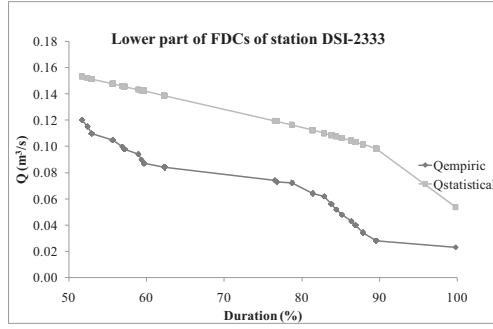


Figure 5.28. Comparison of observed FDC and Regionalized FDC at station DSI-2323

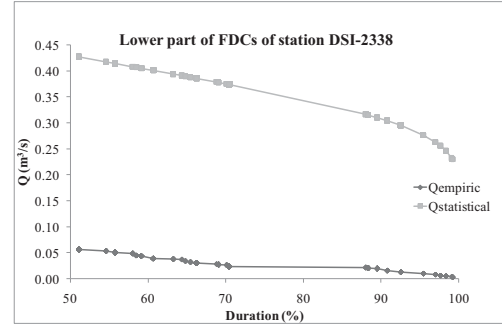


Figure 5.29. Comparison of observed FDC and Regionalized FDC at station DSI-2338

The statistical approach based on the theory of the lognormal probability distribution is seemed to be inappropriate for Oltu Basin. Nearly, all the FDCs are highly underestimated, except DSI-2321 and DSI-2333, as can be seen on previous figures. In order to be sure about this fact, performance indices should be checked (see Table 5.65).

Table 5.65. Performance indices of all stations for statistical approach

	Area (km <sup>2</sup> )	E	R	$\epsilon_s$	$\sigma_{\epsilon,s}$	RRMSE	RMSE
<b>DSI-2336</b>	49.25	-0.681	Not Calculated N.C	-0.459	0.025	0.406	0.049
<b>DSI-2335</b>	68.37	0.725	0.852	-0.201	0.037	0.241	0.013
<b>DSI-2324</b>	4537.55	0.763	0.874	-0.022	0.274	0.275	1.290
<b>DSI-2323</b>	1094.21	-0.501	N.C	-0.321	0.183	0.369	0.707
<b>DSI-2337</b>	197.04	-1.123	N.C	-0.404	0.090	0.410	0.127
<b>DSI-2339</b>	10.62	-1.685	N.C	N.C	N.C	0.036	0.023
<b>EIE-2323</b>	6978.73	0.384	0.619	-0.115	0.075	0.137	2.096
<b>EIE-2325</b>	1785.57	-0.163	N.C	-0.174	0.572	0.598	1.170
<b>EIE-2329</b>	3537.10	-3.826	N.C	-0.624	0.105	0.632	3.748
<b>DSI-2313</b>	6967.93	-4.575	N.C	-0.414	0.046	0.416	7.632
<b>DSI-2321</b>	1851.19	-0.042	N.C	1.619	1.618	2.470	0.919
<b>DSI-2322</b>	3501.01	-4.963	N.C	-0.624	0.037	0.948	3.432
<b>DSI-2333</b>	47.20	-3.140	N.C	0.370	0.375	1.401	0.056
<b>DSI-2338</b>	72.83	-591.438	N.C	N.C	N.C	19.195	0.338

It is an unfortunate conclusion, which is derived by looking at Table 5.65, is that the statistical approach suggested is dreadful except for the basins DSI-2335 and DSI-2324. E values are negative, R-values could not be calculated and relative errors are generally high. Therefore, by inspecting FDC figures and the performance table it is concluded that statistical approach based on theory of the lognormal probability distribution is not applicable for Oltu basin.

After deciding on the lognormal cumulative probability distribution function (CPDF) is not appropriate as a statistical approach, several probability distributions are tried on the data to have a final word.

As in many statistical analyses containing distribution fitting, EasyFit has been used as the statistical tool. For the lower part of the FDCs of stations, 65 different distributions have been applied and ranked in the order according to the goodness of fit tests (GOF) at several confidence intervals. Anderson-Darling has chosen as the GOF statistics as in the parameter analysis in the previous parts of this chapter. In Appendix F, the proposed probability distribution functions for the daily FDCs' lower part are available.

Table 5.66 shows overall evaluation of the approaches applied in this chapter. It can be concluded that Model Kocatepe and Model Quimpo give the best for regionalization of FDCs.

Table 5.66. Overall evaluation of parametric and statistical approaches

<b>Approach</b>	<b>E</b>	<b>R</b>	<b><math>\bar{e}_s</math></b>	<b><math>\sigma_{e,s}</math></b>	<b>RRMSE</b>	<b>RMSE</b>
<b>Kocatepe</b>	0.640	0.870	-0.282	0.057	0.325	0.616
<b>Quimpo</b>	0.980	0.990	-0.330	0.434	0.623	2.346
<b>Statistical</b>	-43.590	0.782	-0.114	0.286	1.967	1.543

## CHAPTER 6

### MODEL VALIDATION

In the “Regionalization of FDCs” chapter (Chapter 5), some of the stations (basins) are separated from data set to use in validation (see Table 6.1). The procedure for validation can be summarized as follow:

- Model Kocatepe and Regional Model of Quimpo et al. (1983) will be applied to the stations in Table 6.1.
- The performance indices will be calculated for stations and durations as done in jack-knife cross-validation section
- The empiric, Model Kocatepe and Model Quimpo FDCs will be plotted on the same graph

Table 6.1. Validation stations and the basin parameters

Station	Q (m <sup>3</sup> /s)	Duration (yr)	Area (km <sup>2</sup> )	LMR (km)	BR (m)	MAP (mm)	CN
DSI-2313	37.6	2	6967.93	158.39	1469.10	563.30	72.21
DSI-2321	5.5	2	1851.19	91.00	967.40	521.10	73.13
DSI-2333	0.8	4	47.20	8.48	271.90	669.58	73.54
DSI-2322	14.7	4	3501.01	101.51	1029.00	514.39	76.70
DSI-2338	0.6	3	72.83	16.33	1106.68	444.03	84.86

## 6.1. Model Kocatepe Validation

The equations used in Model Kocatepe are shown in Table 5.52. The  $q_j$  discharges of stations DS-2313, DSI-2321, DSI-2333, DSI-2322 and DSI-2338 have been computed by using the equations in the list. Then, they have been compared with the observed discharges and goodness of fit statistics has been studied to confirm the performance and robustness of the model (see Table 6.2).

According to Table 6.2 it can be concluded that Model Kocatepe works perfectly for large basins. However, for basins areas under  $75 \text{ km}^2$  high deviations are observed. E and R-values tend to decrease (sometimes cannot be computed); relative error indices tend to increase, which means error gets higher.

Table 6.2. Performance indices of validation stations for Model Kocatepe

Station	Area ( $\text{km}^2$ )	E	R	RMSE	RRMSE	$\bar{\epsilon}_s$	$\sigma_{e,s}$
<b>DSI-2313</b>	6967.93	0.976	0.988	7.011	0.211	-0.153	0.152
<b>DSI-2321</b>	1851.19	0.885	0.941	3.606	1.54	0.835	1.361
<b>DSI-2333</b>	47.20	-0.536	N.C	3.311	0.874	-0.936	0.087
<b>DSI-2322</b>	3501.01	0.758	0.871	9.932	0.221	0.093	0.207
<b>DSI-2338</b>	72.83	0.379	0.616	1.366	1.291	-0.286	0.810

When the goodness of fit tests and relative error statistics examined for Model Kocatepe (see Table 6.3), they seem to be reasonable and good. The only annoyance is due to the duration  $q_{20}$ . However, a similar disturbance was observed in the Jack-knife cross-validation section. Therefore, it seems to be ignorable.



Table 6.3. Kocatepe model indices of performance results for each duration

$q_i$	<b>E</b>	<b>R</b>	<b>RMSE</b>	<b>RRMSE</b>
$q_2$	0.932	0.966	13.954	0.541
$q_5$	0.868	0.932	10.167	0.549
$q_8$	0.888	0.942	7.113	0.555
$q_{10}$	0.874	0.935	6.440	0.555
$q_{15}$	0.287	0.536	5.140	0.523
$q_{20}$	-0.963	N.C	4.418	0.497
$q_{30}$	0.932	0.966	3.480	0.813
$q_{50}$	0.981	0.990	2.559	1.065
$q_{70}$	0.984	0.992	2.608	1.738
$q_{90}$	0.995	0.997	1.701	1.326
$q_{95}$	0.992	0.996	2.092	1.719
$q_{98}$	0.985	0.993	2.954	0.944

## 6.2. Model Quimpo Validation

This subsection of the chapter is one of the most important parts, since validation is the most crucial analysis in a model development study. The same procedure, which is performed for Model Kocatepe, will be applied for Model Quimpo. The model will be used to evaluate the FDCs of validation basins DS-2313, DSI-2321, DSI-2333, DSI-2322 and DSI-2338, The equations used for Model Quimpo are listed below:

$$Q_A = 0.257 (A^{-0.279}) \quad (\text{Equation 5.14})$$

As a reminder:

- D is the percent of time that Q has been exceeded during period of record
- A is the basin area ( $\text{km}^2$ )
- c, regional parameter, see Figure 5.15.
- E, R, RMSE, RRMSE,  $\bar{\epsilon}_s$  and  $\sigma_{\epsilon,s}$  are goodness of fit statistics

The only challenging step is to get regionalized c values from the map. To do this “zonal statistics” analysis has been performed in ArcGIS. The basin shape files are used to specify the basin borders and the zonal statistics have been run. The mean values of the statistics have been taken as the “c” values.  $Q_A$  values have been calculated from Equation 5.15.

The essential parameters are listed for each basin are listed in Table 6.4.

Table 6.4. Validation stations and the parameters of Quimpo Model

Station	Area (km <sup>2</sup> )	Q <sub>A</sub>	c
<b>DSI-2313</b>	6967.93	0.022	9.01
<b>DSI-2321</b>	1851.19	0.032	9.41
<b>DSI-2333</b>	47.20	0.088	8.15
<b>DSI-2322</b>	3501.01	0.026	8.98
<b>DSI-2338</b>	72.83	0.078	9.73
<b>p</b>	0.257		
<b>m</b>	0.279		

Table 6.5 is the same table developed for Model Kocatepe; the durations are evaluated from the performance index point of view. However, Model Quimpo gives deficient results for these durations, compared to Model Kocatepe.  $q_{15}$  and  $q_{20}$  are erroneous as in the Model Kocatepe, and the relative errors are quite large compared to Model Kocatepe.

A comparison Table is the best way to figure out which of the models work best. According to the comparison Table, (see Table 6.6) the explicit inference is Model Kocatepe works better with large basins; on the contrary, Model Quimpo works better with small basins. Nash-Sutcliffe efficiency criterion, coefficient of determination and the relative errors point out the same conclusion.  $E_{Kocatepe}$  and  $R_{Kocatepe}$  are greater for large basins;  $E_{Quimpo}$  and  $R_{Quimpo}$  take higher values for small basins. It can be concluded that, for basins like Oltu in Turkey, it is appropriate to use both of the models according to the size of the sub basin area. If basin size is smaller than 75 km<sup>2</sup> Model Quimpo should be used, otherwise Model Kocatepe is convenient to use.

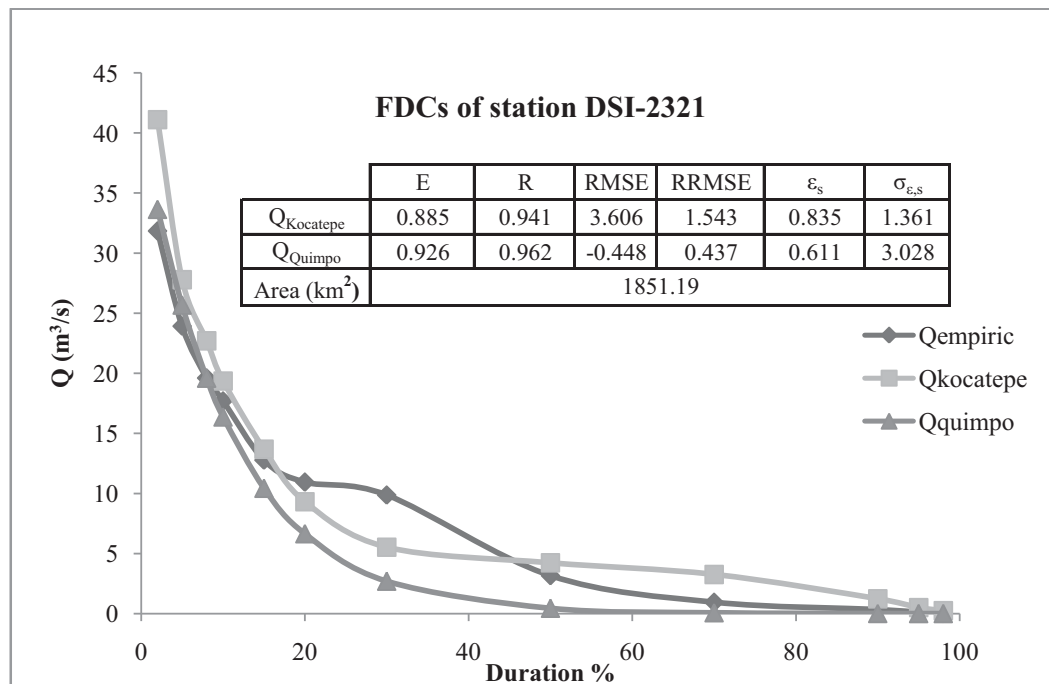
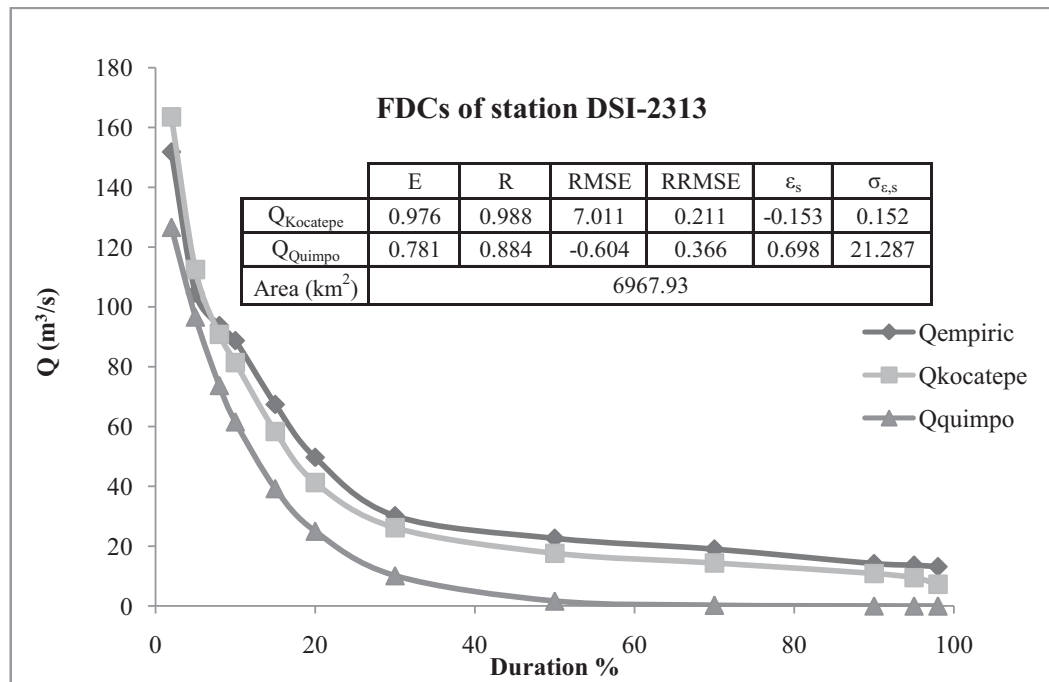
The following figures are for the FDCs developed for validation stations. It is more understandable which model is best for the basin considered.

Table 6.5. Quimpo model indices of performance results for each duration

$q_j$	<b>E</b>	<b>R</b>	<b>RMSE</b>	<b>RRMSE</b>
$q_2$	0.929	0.964	14.291	0.310
$q_5$	0.922	0.960	7.798	0.299
$q_8$	0.753	0.868	10.541	0.280
$q_{10}$	0.433	0.658	13.652	0.289
$q_{15}$	-4.391	N.C	14.132	0.252
$q_{20}$	-15.624	N.C	12.857	0.353
$q_{30}$	0.270	0.520	11.438	1.111
$q_{50}$	0.646	0.804	11.050	0.805
$q_{70}$	0.782	0.884	9.734	1.002
$q_{90}$	0.903	0.950	7.297	1.075
$q_{95}$	0.915	0.956	6.962	1.075
$q_{98}$	0.925	0.962	6.651	1.114

Table 6.6. Comparison of validation results for parametric approaches

<b>Station</b>	<b>DSI-2313</b>	<b>DSI-2321</b>	<b>DSI-2333</b>	<b>DSI-2322</b>	<b>DSI-2338</b>
<b>Area (km<sup>2</sup>)</b>	6967.93	1851.19	47.20	3501.01	72.83
<b>E<sub>Kocatepe</sub></b>	0.976	0.885	-0.536	0.758	0.379
<b>E<sub>Quimpo</sub></b>	0.781	0.926	0.571	0.872	0.984
<b>R<sub>Kocatepe</sub></b>	0.988	0.941	N.C	0.871	0.616
<b>R<sub>Quimpo</sub></b>	0.884	0.962	0.755	0.934	0.992
<b>RMSE<sub>Kocatepe</sub></b>	7.011	3.606	3.311	9.932	1.366
<b>RMSE<sub>Quimpo</sub></b>	21.287	3.028	1.751	7.226	0.217
<b>RRMSE<sub>Kocatepe</sub></b>	0.211	1.543	0.874	0.221	1.291
<b>RRMSE<sub>Quimpo</sub></b>	0.698	0.611	0.647	0.658	0.735
<b><math>\bar{\epsilon}_s</math> Kocatepe</b>	-0.153	0.835	-0.936	0.093	-0.286
<b><math>\bar{\epsilon}_s</math> Quimpo</b>	-0.604	-0.448	-0.213	-0.418	0.179
<b><math>\sigma_{\epsilon,s}</math> Kocatepe</b>	0.152	1.361	0.087	0.207	0.810
<b><math>\sigma_{\epsilon,s}</math> Quimpo</b>	0.366	0.437	0.449	0.531	0.700



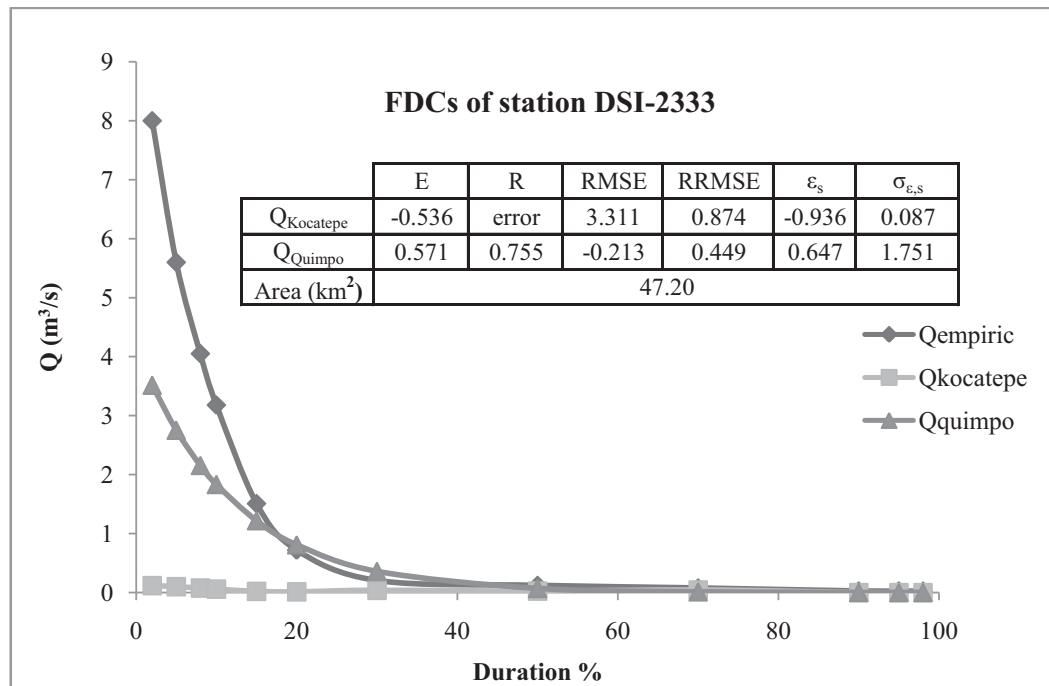


Figure 6.3. Empiric and regional FDCS for DSI-2333 region

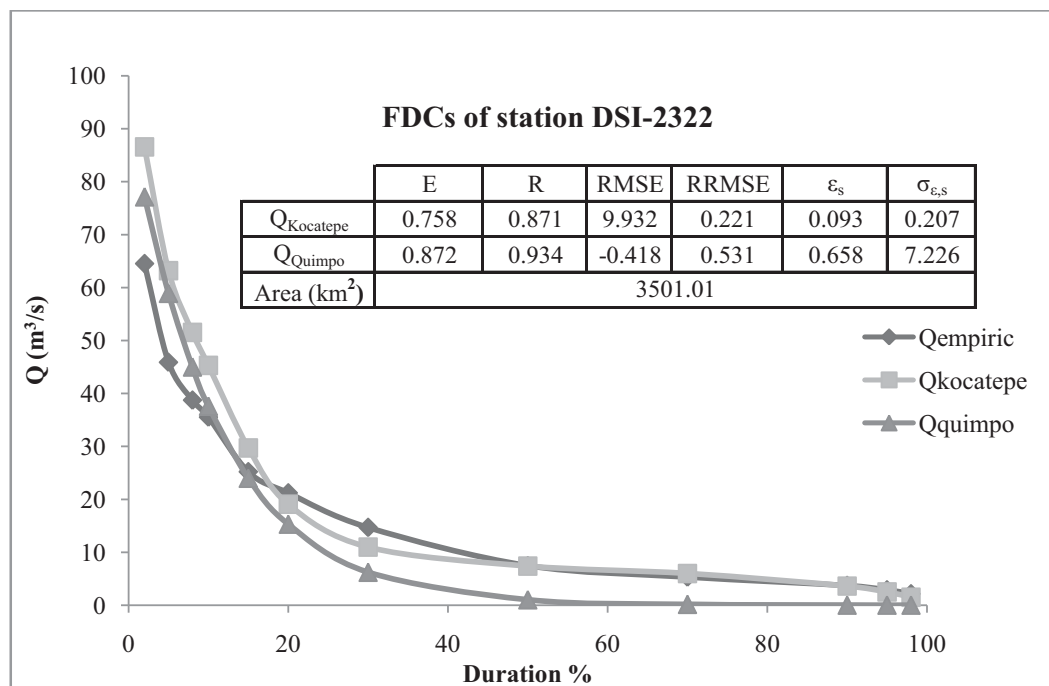
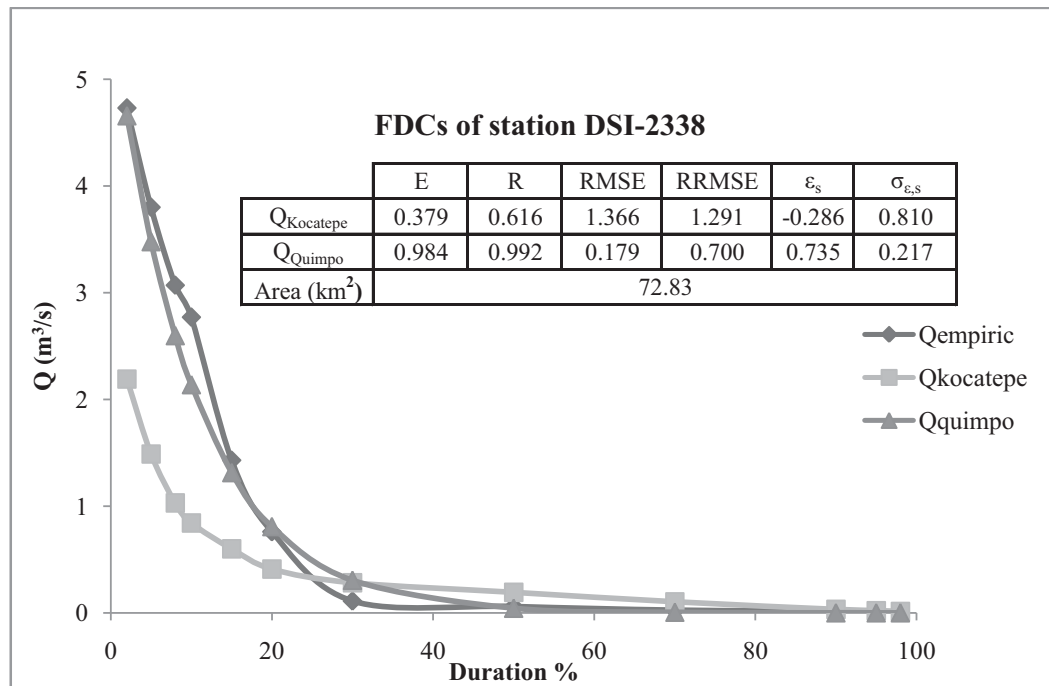


Figure 6.4. Empiric and regional FDCS for DSI-2322 region



### 6.3. Model Comparison For HEPPs In Oltu Basin

To be able to get more conclusions about model performance, a validation is performed with the existing HEPPs in the Oltu Basin. As mentioned in the Data Analysis chapter, 43 HEPP locations have been provided from DSI. The corporations applied for the construction of these HEPPs are generally private sector and the projects are on the feasibility stage. Only 20 of the HEPPs' installed power is available, and only 5 of the HEPPs' net head is available. Therefore, these 5 HEPPs have been chosen for validation procedure. The HEPPs are named by the numbers available on the attribute table. Figure 6.6 displays the HEPPs, and corresponding catchments and main channels. HEPPs 17, 40 and 39 are on the Oltu Branch and 18 and 19 are on Tortum Branch.

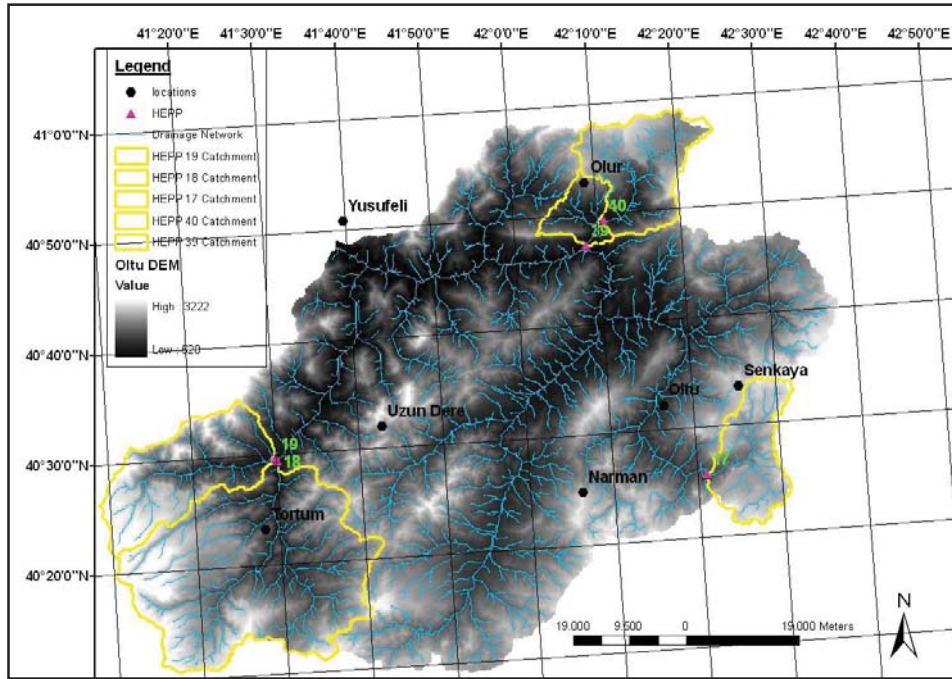


Figure 6.6. HEPPs used for validation of regional model

During the validation process for HEPPs, firstly, the catchment and the main channels have been obtained. Then, by using slope, aspect, CN, temperature, precipitation and elevation maps, the parameters, which are necessary for the model have been derived. Table 6.7 displays the HEPPs and the parameter values of them.

Table 6.7. HEPPs and the parameters used for validation

HEPP	17	18	19	40	39
Perimeter (km)	103.21	213.00	128.10	113.94	129.14
Area (km <sup>2</sup> )	224.88	1008.52	374.08	289.84	381.43
LMR (m)	24.68	46.10	38.02	28.90	33.21
H <sub>max</sub> (m)	3095.00	32222.00	3221.00	2839.00	2839.00
H <sub>min</sub> (m)	1760.00	1158.00	1188.00	1100.00	980.00
H <sub>mean</sub> (m)	2417.87	2202.53	2289.20	2109.55	1999.33
BR (m)	657.87	1044.53	1101.20	1009.55	1019.33
Slope (%)	11.69	38.39	46.13	14.24	15.47
Aspect	184.24	52.46	49.52	177.97	175.72
MAP (mm)	456.76	486.10	628.28	624.45	639.27
CN	77.70	72.55	78.62	69.91	69.47
T (°C)	5.18	7.52	8.69	7.43	7.62

In this analysis there are no discharge values to compare, only installed power values are available. Therefore, for each HEPP, turbine types, turbine units, hydraulic efficiency due to head losses  $\eta_H$ , turbine efficiency  $\eta_T$ , generator efficiency  $\eta_G$ , transformer efficiency  $\eta_{Tr}$  and overall efficiency of hydropower plant  $\eta$  have been assumed. The assumptions have been made with the help of the feasibility report of Yıldırım HEPP located in the Black Sea region.

The equations used for the analysis are:

$$P = \gamma \eta Q H \quad (\text{Equation 6.1})$$

where

P, installed power (MW)

$\gamma$ , specific weight of water (9.81 kN/m<sup>3</sup>)

Q, discharge (m<sup>3</sup>/s)

H, Net Head (m)

$$\eta = \eta_H \cdot \eta_G \cdot \eta_T \cdot \eta_{Tr} \quad (\text{Equation 6.2})$$



Table 6.8. HEPP characteristics

<b>HEPP</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>40</b>	<b>39</b>
<b>Net Head (m)</b>	500.00	62.63	247.00	133.00	103.00
<b>Installed power (MW)</b>	8.50	5.40	4.00	2.45	0.43
<b>Unit number</b>	5	3	2	2	1
<b>Turbine type</b>	Pelton	Francis	Francis	Francis	Francis
$\eta_H$	0.95	0.95	0.95	0.95	0.95
$\eta_G$	0.98	0.98	0.98	0.98	0.98
$\eta_T$	0.89	0.93	0.93	0.93	0.93
$\eta_{T\text{-system}}$	0.82	0.90	0.93	0.93	0.93
$\eta_{Tr}$	0.99	0.99	0.99	0.99	0.99
$\eta$	0.76	0.83	0.86	0.86	.86

Table 6.8 displays the HEPP characteristics. Net head and installed power values have been obtained from DSI the rest of the inputs have been assumed. However, turbine types have been determined according to the Layman's guidebook (Penche, 1998). Layman's guidebook uses the net head, range of discharges through the turbine, rotational speed, cavitation problems and the cost as the decision criteria. Since this is a preliminary study, only the "net head" has been used to decide on turbine types. The criterion is as follows:

Table 6.9. Turbine types and range of heads (m)

<b>Turbine type</b>	<b>Head range in meters</b>
Kaplan and Propeller	$2 < H < 40$
Francis	$10 < H < 350$
Pelton	$50 < H < 1300$

The regional model on the Equation 5.8 ( $Q = b_1 A^{b_2} LMR^{b_3} MAP^{b_4} BR^{b_5} CN^{b_6}$ ) and the equations of Table 5.52 have been used to obtain the  $q_j$  discharges and the FDCs of the HEPPs. Table 6.9 displays the main parameters of HEPPs, which are used in Model Kocatepe.

Table 6.9. HEPP parameters for Model Kocatepe

HEPP	Area (km <sup>2</sup> )	LMR (km)	BR (m)	MAP (mm)	CN
17	224.88	24.68	657.87	456.76	77.70
18	1008.52	46.10	1044.53	486.10	72.55
19	374.08	38.02	1101.20	628.28	78.62
40	289.84	28.90	1009.55	624.45	69.91
39	381.43	33.21	1019.33	639.27	69.47

By using the parameters in Table 6.9 FDC of HEPPs have been developed. These FDCs are available in Figure 6.7 – Figure 6.11.

Moreover, in Figure 6.7, the FDC of HEPP 17 developed with area interpolation method is available. In this method, the areas of gauged and ungauged sites are proportioned. Then, discharge values are transferred to the ungauged site. For HEPP 17, DSI-2321 is the available gauged site. The daily FDC of the station DSI-2321 has been used. Finally, daily FDC of HEPP 17 has been generated. Table 6.10 shows the  $q_j$  values of HEPP 17. It is obvious that the area proportioning approach makes rough estimates for FDCs.

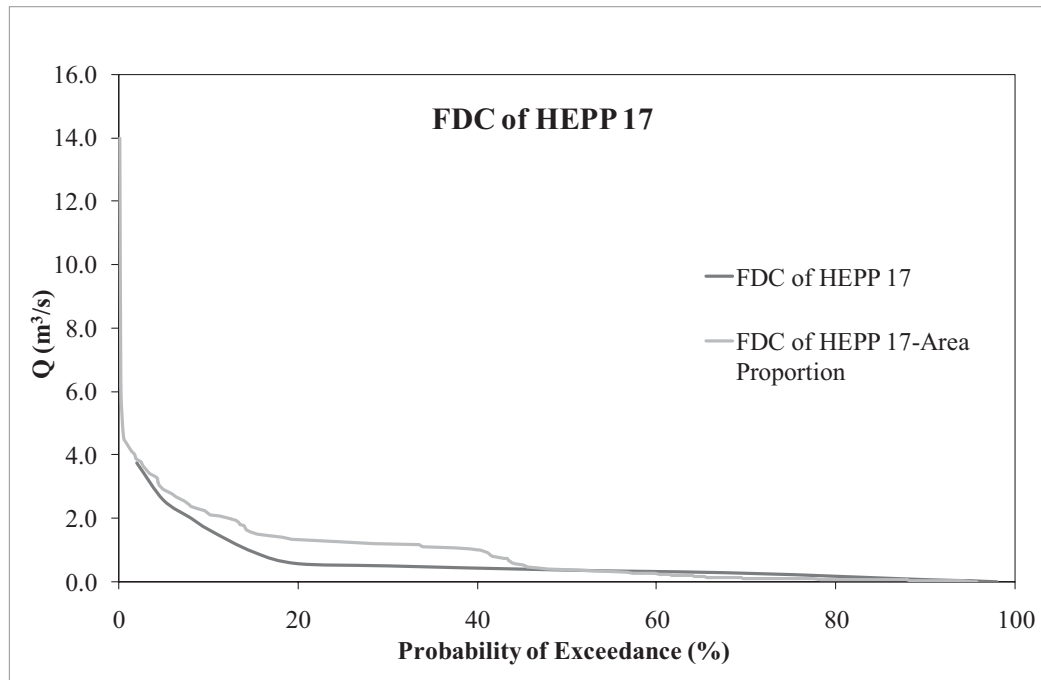


Figure 6.7. FDC of HEPP 17 developed with Model Kocatepe

Table 6.10.  $q_j$  discharges of HEPP 17 for both approaches

$q_j$	Area Proportion Approach	Regional Model Kocatepe
$q_{10}$	2.15	1.624
$q_{20}$	1.32	0.586
$q_{30}$	1.19	0.505
$q_{40}$	1.01	0.490
$q_{50}$	0.38	0.368
$q_{60}$	0.25	0.300
$q_{70}$	0.117	0.280
$q_{80}$	0.07	0.100
$q_{90}$	0.042	0.065

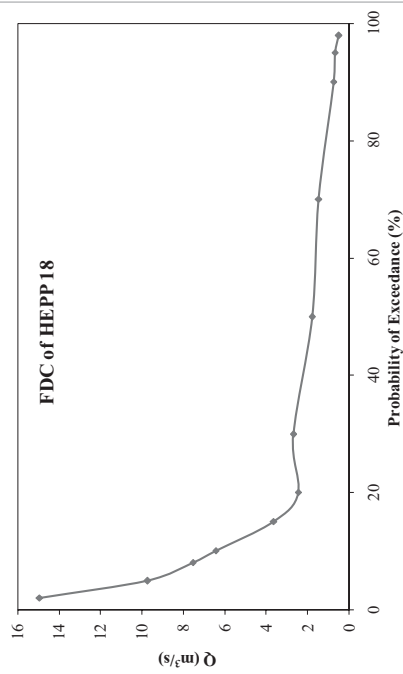


Figure 6.8. FDC of HEPP 18 developed with Model Kocatepe

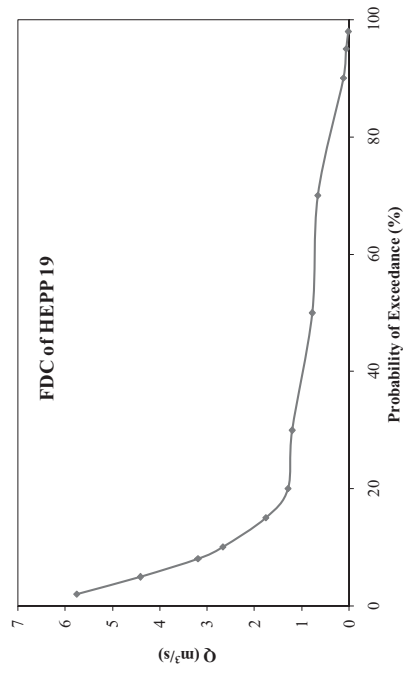


Figure 6.9. FDC of HEPP 19 developed with Model Kocatepe

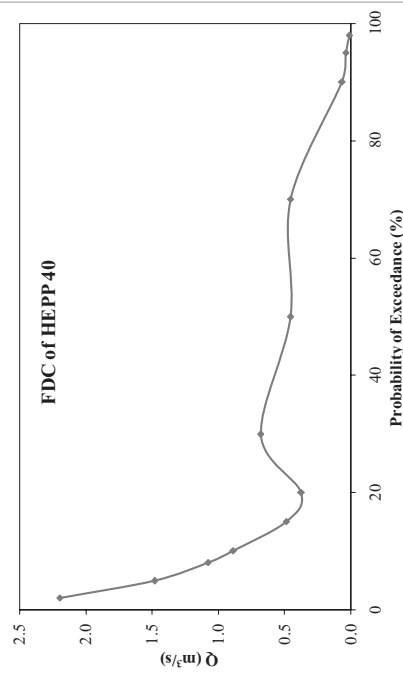


Figure 6.10. FDC of HEPP 40 developed with Model Kocatepe

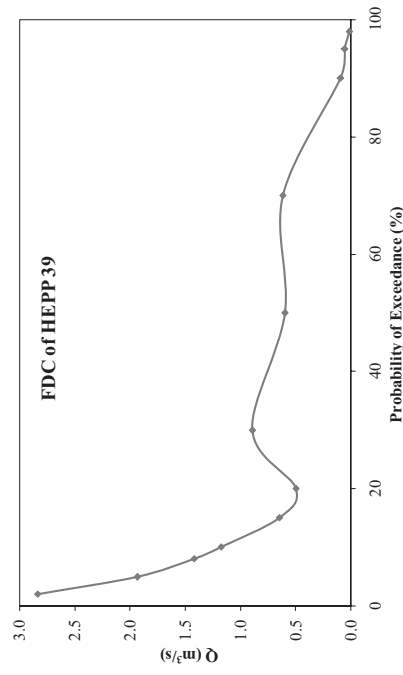


Figure 6.11. FDC of HEPP 39 developed with Model Kocatepe

Table 6.11. Designed and calculated installed power of HEPPs

HEPP	q <sub>2</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	3.74	500.00	0.76	13.86	8.50
18	14.94	62.63	0.83	7.62	5.40
19	5.76	247.00	0.86	11.95	4.00
40	2.20	133.00	0.86	2.46	2.45
39	2.84	103.00	0.86	2.46	0.43
HEPP	q <sub>5</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	1.10	500.00	0.76	9.54	8.50
18	4.32	62.63	0.83	4.96	5.40
19	1.84	247.00	0.86	9.16	4.00
40	0.59	133.00	0.86	1.66	2.45
39	0.79	103.00	0.86	1.67	0.43
HEPP	q <sub>8</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	1.99	500.00	0.76	7.39	8.50
18	7.53	62.63	0.83	3.84	5.40
19	3.19	247.00	0.86	6.62	4.00
40	1.08	133.00	0.86	1.21	2.45
39	1.42	103.00	0.86	1.23	0.43
HEPP	q <sub>10</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	1.62	500.00	0.76	6.02	8.50
18	6.44	62.63	0.83	3.28	5.40
19	2.67	247.00	0.86	5.55	4.00
40	0.89	133.00	0.86	0.99	2.45
39	1.17	103.00	0.86	1.02	0.43
HEPP	q <sub>15</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.93	500.00	0.76	3.44	8.50
18	3.63	62.63	0.83	1.85	5.40
19	1.76	247.00	0.86	3.66	4.00
40	0.49	133.00	0.86	0.55	2.45
39	0.65	103.00	0.86	0.56	0.43
HEPP	q <sub>20</sub>	Net Head (m)	η	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.59	500.00	0.76	2.17	8.50
18	2.46	62.63	0.83	1.25	5.40
19	1.29	247.00	0.86	2.68	4.00
40	0.38	133.00	0.86	0.42	2.45
39	0.50	103.00	0.86	0.43	0.43

Table 6.10. (continued)

HEPP	$q_{30}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.50	500.00	0.76	1.87	8.50
18	2.69	62.63	0.83	1.37	5.40
19	1.20	247.00	0.86	2.50	4.00
40	0.68	133.00	0.86	0.76	2.45
39	0.89	103.00	0.86	0.77	0.43
HEPP	$q_{50}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.37	500.00	0.76	1.36	8.50
18	1.78	62.63	0.83	0.91	5.40
19	0.78	247.00	0.86	1.63	4.00
40	0.46	133.00	0.86	0.51	2.45
39	0.60	103.00	0.86	0.52	0.43
HEPP	$q_{70}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	I Installed power Designed (MW)
17	0.28	500.00	0.76	1.04	8.50
18	1.47	62.63	0.83	0.75	5.40
19	0.66	247.00	0.86	1.38	4.00
40	0.46	133.00	0.86	0.51	2.45
39	0.61	103.00	0.86	0.53	0.43
HEPP	$q_{90}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.07	500.00	0.76	0.24	8.50
18	0.74	62.63	0.83	0.38	5.40
19	0.12	247.00	0.86	0.25	4.00
40	0.07	133.00	0.86	0.07	2.45
39	0.09	103.00	0.86	0.08	0.43
HEPP	$q_{95}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.03	500.00	0.76	0.10	8.50
18	0.67	62.63	0.83	0.34	5.40
19	0.07	247.00	0.86	0.14	4.00
40	0.04	133.00	0.86	0.04	2.45
39	0.05	103.00	0.86	0.05	0.43
HEPP	$q_{98}$	Net Head (m)	$\eta$	Installed power Calculated (MW)	Installed power Designed (MW)
17	0.01	500.00	0.76	0.04	8.50
18	0.49	62.63	0.83	0.25	5.40
19	0.02	247.00	0.86	0.04	4.00
40	0.01	133.00	0.86	0.01	2.45
39	0.02	103.00	0.86	0.01	0.43

When the results of the Table 6.11 are examined, it is observed that the discharges  $q_5$ ,  $q_8$  and  $q_{10}$  give consistent results. Moreover, the discharges belong to 5%, 8% and 10% probability of exceedance are widely used in the design of the HEPPs in Turkey. In brief, Model Kocatepe is a reasonable model to use in the designs.

#### **6.4. Model Validation For Short Samples**

It is an important issue to analyze the sensitivity of the regional model to the sample length, since short record lengths are as common as ungaged basins. In this analysis, 8 basins (and corresponding stations) have been considered with at least 12 years of daily discharges (see Table 6.12). The short-term duration validation has been performed for both Model Kocatepe and Model Quimpo. The steps in the analysis are as follows:

- As short durations ( $l$  years); 1, 2 and 5 year sub-samples have been extracted from the complete records (see Table 6.12).
- The empirical FDCs for each sub-sample and modeled ones have been generated and compared as performed in section Model Kocatepe and Model Quimpo.
- For Model Kocatepe, performance indices, which are defined before, have been computed and then averaged to get the representative for the entire region for 17 sub-samples (see Tables 6.13, 6.14, and 6.15).
- For Model Quimpo, performance indices, which are defined before, have been computed and then averaged to get the representative for the entire region for 8 sub-samples (see Tables 6.16, 6.17, and 6.18).
- Empirical FDCs for short durations (for  $l = 1, 2$  and 5 years) and the developed FDCs (Model Kocatepe and Quimpo) has been constructed.

### 6.4.1. Model Kocatepe Short-Term Duration Validation

Figure 6.12 – Figure 6.14 present the empirical and modeled FDCs of EIE-2323 for 1, 2 and 5 years durations. The rest of the FDCs of other station show a similar pattern, therefore only EIE-2323 FDCs are presented. The FDCs (Figure 6.12 – Figure 6.14) show that, the discharges of 1 and 2 years durations have been underestimated when modeled, and overestimated in 5 years duration.

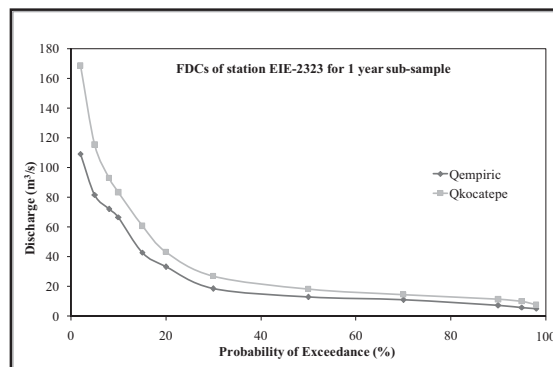


Figure 6.12. Empirical and modeled FDCs for EIE-2323 1 year sub-sample (1985, 1989)

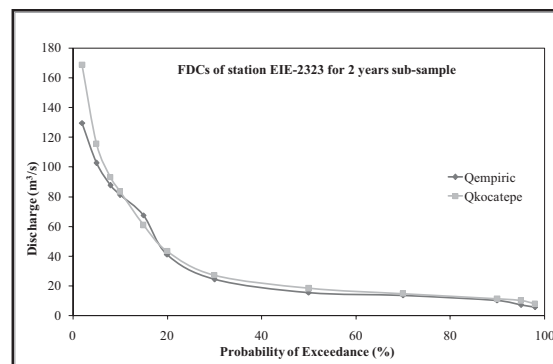


Figure 6.13. Empirical and modeled FDCs for EIE-2323 2 years sub-sample (1985, 1989)

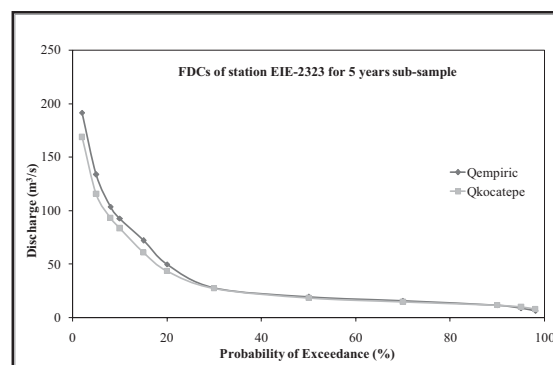


Figure 6.14. Empirical and modeled FDCs for EIE-2323 5 years sub-sample (1985, 1989)



The performance indices in Table 6.13 confirm this preliminary approach.

Table 6.13. 1 year record length performance indices

Station	Sample	E	R	$\bar{E}_s$	$\sigma_{E,s}$	$E_s$	RRMSE	RMSE
DSI-2323	1	0.379	0.616	1.346	1.008	0.9886	1.656	6.042
DSI-2324	1	0.596	0.772	0.676	0.489	0.9952	0.8224	17.797
	2	0.932	0.966	-0.070	0.162	0.9994	0.170	10.020
DSI-2335	1	.670	0.819	-0.564	0.089	0.9942	0.698	0.663
DSI-2337	1	0.668	0.817	-0.257	0.621	0.9957	0.648	2.371
	2	-1.397	N.C.*	0.294	0.688	0.9810	0.729	1.601
DSI-2339	1	-0.647	N.C.	-0.962	0.018	0.9785	-0.647	0.319
EIE-2323	1	0.774	0.880	-0.022	0.222	0.9984	0.213	17.933
	2	-2.552	N.C.	0.370	0.526	0.9816	0.625	39.273
	3	0.584	0.764	0.433	0.131	0.9961	0.450	23.270
	4	0.555	0.745	0.170	0.191	0.9966	0.249	23.204
EIE-2325	1	0.348	0.590	0.251	0.356	0.9943	0.423	5.928
	2	0.909	0.953	-0.194	0.363	0.9988	0.398	4.718
	3	-2.989	N.C.	0.463	0.755	0.9697	0.859	9.340
EIE-2329	1	0.964	0.982	0.328	0.361	0.9995	0.409	5.565
	2	0.947	0.973	-0.089	0.221	0.9995	0.230	5.863
	3	-5.113	N.C.	0.809	0.655	0.9518	1.024	25.162
<b>AVG.</b>		-0.257	0.823	0.175	0.403	0.9893	0.527	11.710

(\*) N.C. : Could Not Be Calculated

The calculations for the average of the efficiency indices for all sub-samples, which are for identifying the representative number for the entire study region, have been resulted in a negative value. E varies between 1 (perfect fit) and  $-\infty$ . For 1-year duration analysis, the fit of model is poor (see Table 6.13). However, the rest of the performance indices are not that bad to reject the model application completely.

2-years duration sub-samples performance indices (Table 6.14) are better than the 1-year duration sub-samples results. The most significant improvement can be observed in the Nash-Sutcliffe efficiency criterion, E. It has been grown in number and become a positive number. RMSE and RRMSE have been decreased, and this result supports the idea of better fit for 2-years durations. However, when the Figure 6.13 is examined, it is observed that discharges are still underestimated.

Table 6.14. 2 years record length performance indices

Station	Sample	E	R	$\bar{\epsilon}_s$	$\sigma_{\epsilon,s}$	$E_s$	RRMSE	RMSE
DSI-2323	1	-0.609	N.C.	1.035	0.569	0.9991	1.170	7.127
DSI-2324	1	0.863	0.929	0.334	0.176	0.9978	0.3739	13.8485
	2	0.944	0.972	-0.202	0.149	0.9991	0.248	13.925
DSI-2335	1	0.520	0.721	-0.639	0.070	0.9913	0.644	1.006
DSI-2337	1	0.756	0.870	-0.323	0.435	0.9969	0.528	1.802
	2	0.921	0.960	-0.265	0.187	0.9990	0.435	0.811
DSI-2339	1	-0.868	N.C.	-0.923	0.229	0.9758	0.957	2.885
EIE-2323	1	0.944	0.972	-0.037	0.145	0.9995	0.144	10.386
	2	0.984	0.992	0.060	0.094	0.9998	0.109	6.308
	3	0.913	0.956	0.149	0.149	0.9992	0.206	12.880
	4	0.835	0.914	0.169	0.141	0.9984	0.216	16.384
EIE-2325	1	0.097	0.312	0.468	0.298	0.9916	0.548	6.600
	2	0.974	0.987	0.182	0.359	0.9984	0.389	2.211
	3	0.534	0.731	0.171	0.362	0.9951	0.387	5.420
EIE-2329	1	0.687	0.829	0.807	0.788	0.9954	1.004	12.838
	2	0.856	0.925	-0.223	0.186	0.9981	0.285	15.678
	3	0.914	0.956	0.153	0.168	0.9990	0.222	7.227
<b>AVG.</b>		0.604	0.868	0.054	0.265	0.9961	0.463	8.079

The performance indices have been improved for the 5-years duration (Table 6.15). The E value comes closer to the 1, the perfect fit. The performance index ( $E_s$ ) has been increased, and the relative errors have been dropped, which indicates the good fit. However, Figure 6.14, the empirical and modeled FDCs of station EIE-2323, shows an overestimation of discharges.

Table 6.15. 5 years record length performance indices

Station	Sample	E	R	$\bar{E}_s$	$\sigma_{E,s}$	$E_s$	RRMSE	RMSE
DSI-2323	1	0.988	0.994	0.143	0.336	0.9999	0.352	1.083
DSI-2324	1	0.987	0.993	0.132	0.124	0.9999	0.178	4.128
	2	0.998	0.999	0.002	0.057	1.0000	0.055	1.798
DSI-2335	1	0.229	0.478	-0.658	0.083	0.9894	0.692	1.260
DSI-2337	1	0.768	0.876	-0.315	0.264	0.9969	0.470	1.804
	2	0.907	0.952	-0.370	0.259	0.9990	0.500	0.857
DSI-2339	1	0.907	0.952	-0.370	0.259	0.9990	0.500	0.857
EIE-2323	1	0.878	0.937	-0.180	0.079	0.9986	0.195	25.376
	2	0.983	0.992	0.008	0.089	0.9998	0.086	6.184
	3	0.970	0.985	-0.037	0.124	0.9997	0.124	10.505
	4	0.932	0.965	0.122	0.079	0.9993	0.144	11.525
EIE-2325	1	0.994	0.997	-0.051	0.099	0.9999	0.108	0.945
	2	0.956	0.978	-0.143	0.102	0.9995	0.173	3.003
	3	0.937	0.968	0.133	0.136	0.9992	0.186	2.621
EIE-2329	1	0.976	0.988	0.233	0.337	0.9997	0.380	4.266
	2	0.970	0.985	-0.038	0.166	0.9996	0.163	5.544
	3	0.993	0.997	-0.096	0.138	0.9999	0.164	2.260
<b>AVG.</b>		0.904	0.943	-0.087	0.161	0.9988	0.263	4.942

Figure 6.12 – Figure 6.14 show an underestimation of the FDC's for short durations and overestimation for long durations. It is an expected result, since in long-term durations it is more possible to encounter with dry or wet periods of the record period. On the contrary, for short durations, extreme events may not be included (floods or droughts). In brief, it can be concluded that, the model underestimates for short durations and over estimates for long durations.

#### 6.4.2. Model Quimpo Short-Term Duration Validation

Figure 6.15 – Figure 6.17 present the empirical and modeled FDCs of DSI-2339 for 1, 2 and 5 years durations. In Figure 6.15, for 1-year durations, there is an overestimation of the discharge values up to 30 % probability of exceedance, and then the estimations are being started underestimated from 30 % duration. Similarly, for 2 years durations, there is an overestimation up to 15 % probability of exceedance, and then the estimations are being started underestimated from 15 % duration. Moreover, for 5 years durations, discharges of FDCs are underestimated.

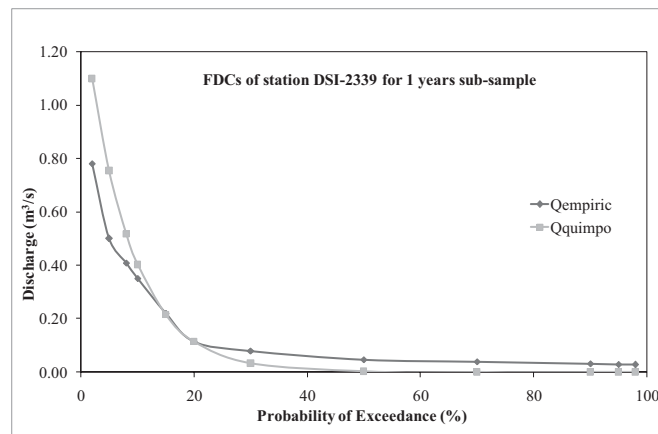


Figure 6.15. Empirical and modeled FDCs for DSI-2339 1 year sub-sample

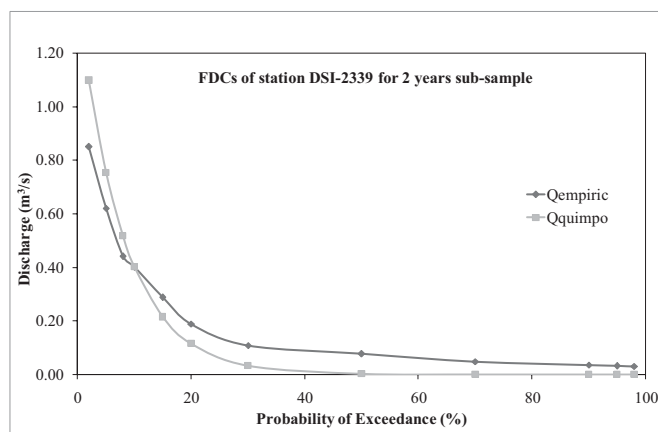


Figure 6.16. Empirical and modeled FDCs for DSI-2339 2 years sub-sample

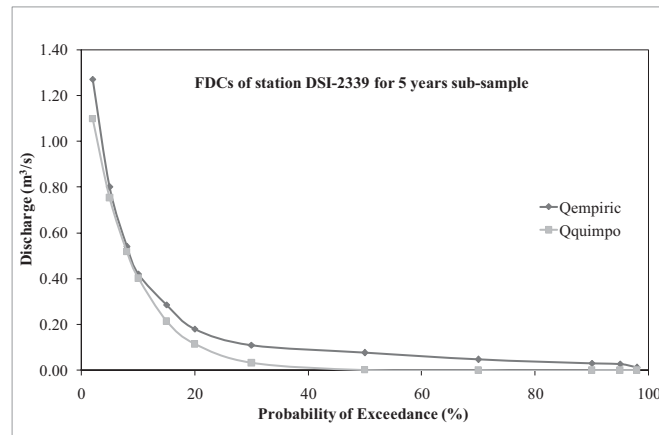


Figure 6.17. Empirical and modeled FDCs for DSI-2339 5 years sub-sample

When the average values of the performance indices in Table 6.16 are examined, it is observed that the E value is so small to be treated as a good result. However, if the E values in each line are examined, it is noticed that the subbasins with small areas (i.e. DSI-2337, DSI-2339), show adequate results for a suitable model. In other words, Model Quimpo works fine with small basins for 1-year durations, but for the entire region (Oltu Basin), it seems to be not appropriate to use with 1-year data. In addition to these, average R-value is close to 1.

The average E value, and the E values of each stations have been increased significantly in Table 17 . Average R-value has been also increased and has been gotten close to 1. Average RMSE has been decreased for 2 years durations compared to the RMSE value for 1-year duration.

Table 6.17. 2 years record length performance indices

Station	E	R	$\bar{\epsilon}_s$	$\sigma_{\epsilon,s}$	RRMSE	RMSE	$E_s$
DSI-2323	-1.325	N.C.	0.237	1.045	1.004	8.351	0.9725
DSI-2324	0.958	0.979	-0.179	0.651	0.648	10.024	0.9996
DSI-2335	0.943	0.979	-0.015	0.242	0.624	0.292	0.9993
DSI-2337	0.971	0.986	-0.438	0.406	0.587	0.607	0.9997
DSI-2339	0.858	0.927	-0.093	0.366	0.694	0.101	0.9984
EIE-2323	0.934	0.966	-0.474	0.427	0.625	11.118	0.9995
EIE-2325	0.794	0.891	-0.065	0.757	0.740	6.113	0.9972
EIE-2329	0.973	0.986	-0.364	0.593	0.611	3.991	0.9997
<b>AVG.</b>	<b>0.638</b>	<b>0.959</b>	<b>-0.174</b>	<b>0.561</b>	<b>0.692</b>	<b>5.074</b>	<b>0.996</b>

Although the discharges have been underestimated, the Nash-Sutcliffe efficiency criterion (E), and the R-value have been increased for the entire region for 5 years duration as can be seen in Table 18. Similar to the Model Kocatepe, 5-year duration could be taken as the minimum record length to use in the models.

Table 6.18. 5 years record length performance indices

Station	E	R	$\bar{\epsilon}_s$	$\sigma_{\epsilon,s}$	RRMSE	RMSE	$E_s$
DSI-2323	0.941	0.970	-0.309	0.558	0.617	2.383	0.9993
DSI-2324	0.965	0.982	-0.442	0.417	0.596	6.802	0.9996
DSI-2335	0.985	0.993	-0.197	0.357	0.650	0.175	0.9998
DSI-2337	0.979	0.989	-0.374	0.416	0.594	0.536	0.9997
DSI-2339	0.965	0.982	-0.228	0.242	0.685	0.072	0.9995
EIE-2323	0.691	0.831	-0.589	0.321	0.664	39.373	0.9966
EIE-2325	0.913	0.955	-0.268	0.567	0.605	4.129	0.9990
EIE-2329	0.967	0.983	-0.480	0.426	0.630	4.926	0.9997
<b>AVG.</b>	0.926	0.961	-0.361	0.413	0.630	7.299	0.999

## **CHAPTER 7**

### **SUMMARY AND CONCLUSIONS**

#### **7.1. Summary of the Results**

In many of the water resources planning and management projects, such as irrigation, hydropower generation and water supply, FDCs play an important role. To be able to construct a FDC of a basin, streamflow data are needed. However, ungauged basins are frequently encountered in many of the studies. In the ungauged basins, the streamflow data are inadequate quantitatively and/or qualitatively.

In the framework of this study, developments of regional models of FDCs for the ungauged basins, and validation of the developed regional models have been discussed. Oltu basin has been studied.

In order to develop a model for Oltu basin several data have been gathered and processed. Hydrologic data (daily streamflows) have been purchased from DSI and EIE; topographic data have been purchased from HGK; meteorological data, geologic data and soil and land use data have been gathered from other studies performed in WRL.

The data have been processed and homogeneity and seasonality analysis have been performed. Heterogeneity analysis has been performed to see if the distributions of the subbasins homogeneous or not. Analysis has been resulted that the subbasins of Oltu basin are heterogeneously distributed. Furthermore, seasonality analysis has been done to observe the seasonality of high discharges (discharges belong to probability of exceedance 5 % - probability of exceedance 30 %).

At the end of the seasonality analysis, it is observed that up to 30 % probability of exceedance, snowmelt is the main reason of high discharges, and high discharges are encountered in the spring. At 30 % probability of exceedance, high discharges are encountered in all seasonality space (in spring, summer, autumn and winter) and effect precipitation is observed.

By using the data processed, several basin parameters have been developed. These are; perimeter (km), area ( $\text{km}^2$ ), length of main river (km), minimum elevation (m), mean elevation (m), maximum elevation (m), basin relief (m), slope (%), aspect ( $^\circ$ ), mean annual precipitation (mm), mean annual temperature ( $^\circ\text{C}$ ), and CN and mean annual discharge ( $\text{m}^3/\text{s}$ ). These parameters have been chosen according to the use in literature and availability.

Two approaches have been used to regionalize FDCs in Oltu basin, namely, parametric approaches and statistical approaches. Moreover, two different parametric approaches have been applied with the available data, namely Model Kocatepe and Model Quimpo. Five streamgauge data have been removed from the data set to use in Validation. The remaining nine stations have been used in these analyses. Firstly, a correlation analysis has been performed, and the reasonable parameters that would be used in model development have been selected. After this elimination with the correlation analysis, multiple regression analyses have been implemented to develop the model and to eliminate the parameters. Since, degrees of freedom problems elimination of parameters could not be done with principal component analysis or stepwise regression analysis.

With the results of the multiple regression analysis a regional model with describing parameters have been chosen (Equation 5.8 and Equation 5.9). This primitive model has been validated through stations with jack-knife cross-validation method for the discharges correspond to durations  $j$  ( $j = 2\%, 5\%, 8\%, \dots, 95\%, 98\%$  probability of exceedance).

The performance indices were good then the Model Kocatepe has been developed for the Oltu Basin. Moreover, the performance indices of nine stations have been calculated.

After this, Model Quimpo has been applied to the nine station data. The coefficients  $Q_A$  and  $c$  of the Equation 5.12 have been defined. Then the performance indices have been calculated again for Model Quimpo.



Thereafter, statistical approach has been applied to the parameters given in Equation 5.8. Lognormal distribution has been chosen as the probability distribution. Parameters of the statistical distribution have been estimated. Then a regional-regression analysis has been performed for the lower part of the FDCs and a regional model is developed for predicting distribution parameters. Equation 5.25 and Equation 5.26 stand for the lognormal distribution parameters. Afterwards, performance indices for the statistical approach have been calculated. The results are not good. Lognormal distribution (statistical approach) is not successful while developing regional FDCs in Oltu basin.

Since the parametric approach's performance indices were good for both models, Model Kocatepe and Model Quimpo, validation of these models have been generated with unused data (5 stations removed before). In addition to these, the validation process has been carried on the existing HEPPs in Oltu Basin and short-term durations. Short-term duration's validation is an important issue to decide on the sufficient record length to use the regional models. Model Kocatepe and Model Quimpo have been tested on short-term durations,  $l = 1, 2$  and 5 years. Finally, performance indices of these validations have been calculated.

## 7.2. Conclusions

For Oltu basin, with two parametric approaches (Model Kocatepe and Model Quimpo) and with a statistical approach (Lognormal Distribution), regional FDCs have been developed. Moreover, the performance indices for both models have been calculated. The necessary validations have been generated.

Consequently, three important outcomes have been reached after these analyses. Firstly, Model Kocatepe and Model Quimpo (parametric approaches while regionalizing the FDCs) estimate the FDCs appropriately. However, drainage area is an important criterion to be taken into account while selecting the convenient method. As drainage area gets bigger, Model Kocatepe has started to give results that are more dependable.

In addition to these, as drainage area gets smaller Model Quimpo has started to give results that are more dependable. This result could be supported by Table 5.52 (performance indices of Model Kocatepe developed during regionalizing process) and by Table 5.58, (performance indices of Model Quimpo developed during regionalizing process). In Table 5.52, the performance indices E and R are high for large basins, such as DSI-2324, DSI-2323 and EIE-2323; and the indices  $\bar{\epsilon}_s$ ,  $\sigma_{\epsilon,s}$  and RMSE are small compared to other basins. In Table 5.52, the performance of indices of small basins, such as DSI-2339 ( $A = 10.62 \text{ km}^2$ ) and DSI-2335 ( $A = 68.37 \text{ km}^2$ ) are poor.

Whereas, in Table 5.58 the performance indices are good for the basins DSI-2339 and DSI-2335 (relatively small drainage areas compared to DSI-2323 and DSI-2324).

Similar results have been obtained for the validation stations (DSI-2313, DSI-2338, DSI-2321, DSI-2333 and DSI-2322). DSI-2313 ( $A = 6967.93 \text{ km}^2$ ), DSI-2322 ( $A = 3501.01 \text{ km}^2$ ) and DSI-2321 ( $A = 1851.19 \text{ km}^2$ ) shows higher E and R-values and smaller  $\bar{\epsilon}_s$ ,  $\sigma_{\epsilon,s}$  and RMSE values for Model Kocatepe; DSI-2333 ( $A = 47.20 \text{ km}^2$ ) and DSI-2338 ( $A = 72.83 \text{ km}^2$ ) shows higher E and R-values and smaller  $\bar{\epsilon}_s$ ,  $\sigma_{\epsilon,s}$  and RMSE values for Model Quimpo.

From these relations it can be concluded that,  $75 \text{ km}^2$  drainage area can be the upper boundary to use Model Quimpo. The limitation is set according to the station DSI-2338 ( $A = 72.83 \text{ km}^2$ ). Between  $75 \text{ km}^2$  and  $200 \text{ km}^2$  both methods, Model Quimpo and Model Kocatepe, can be used. The boundary  $200 \text{ km}^2$  is set according to the DSI-2337. Over  $200 \text{ km}^2$  Model Kocatepe should be used. In brief:

- $A \leq 75 \text{ km}^2$ , Model Quimpo
- $75 \text{ km}^2 \leq A \leq 200 \text{ km}^2$ , Model Quimpo or Model Kocatepe
- $A \geq 200 \text{ km}^2$ , Model Kocatepe

Moreover, the results of short-term duration analysis in sections 6.4.1 (for Model Kocatepe) and section 6.4.2 (for Model Quimpo) have been revealed that minimum 5 years record length for discharge data is appropriate to use (see Table 6.14 and Table 6.17 and check the performance indices). However, if Table 6.13 and 6.16 (for 2 years record length) are examined, data belongs to 2 years record length is also sufficient.

### **7.3. Recommendations**

While processing all the analysis it has been realized that using 1/25000 map sheets are not necessary. Because correcting and processing of these 10 m contour interval map sheets cause a loss of time. ASTER DEMs, which are commercially available, with the resolution 30 m are satisfactory to use in fact.

Moreover, in this study the effect of snow has not been included in the models, although snowmelt is the main reason of high discharges. In a future study snow can be included in the models. According to the available data, more indices can be generated to develop different models.

Finally, the proposed models can be tested in different basins to see its success.

## REFERENCES

- Algancı, U., Coşkun, H. G., Eriş, E., Ağırlioğlu, N., Cıgızoğlu, K., Yılmaz, L., and Toprak, Z. F. (2009). Akım ölçümleri olmayan akarsu havzalarında hidroelektrik potansiyelin belirlenmesine yönelik uzaktan algılama ve CBS ile hidrolik modelleme [hydrologic modelling with remote sensing and GIS determination of hydroelectric potential in ungaged water basins]. Paper presented at the 12. *Türkiye Harita Bilimsel Ve Teknik Kurultayı*, Ankara.
- Brezinski, C. (1991). Implementing the jackknife. *Applied Mathematics and Computation*, 42(2), 111-119. doi:10.1016/0096-3003(91)90047-Q
- Brown, S. (2010). *Measures of shape: Skewness and kurtosis*. Retrieved 26 August, 2010, from <http://www.tc3.edu/instruct/sbrown/stat/shape.htm>
- Burges, S. J., and James, L. D. (1982). Selection, calibration, and testing of hydrologic models. In C. T. Haan (Ed.), *Hydrologic modeling of small watersheds* ( pp. 437-472). Michigan,USA: American Society of Agricultural Engineers.
- Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202(1-4), 212-230. doi:10.1016/S0022-1694(97)00068-1
- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A. (2004). Regional flow-duration curves: Reliability for ungauged basins. *Advances in Water Resources*, 27(10), 953-965. doi:10.1016/j.advwatres.2004.08.005
- Castellarin, A., Burn, D. H., and Brath, A. (2001). Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology*, 241(3-4), 270-285. doi:10.1016/S0022-1694(00)00383-8
- Castellarin, A., Burn, D. H., and Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, 360(1-4), 67-76. doi:10.1016/j.jhydrol.2008.07.014
- Castellarin, A., Camorani, G., and Brath, A. (2007). Predicting annual and long-term flow-duration curves in ungauged basins. *Advances in Water Resources*, 30(4), 937-953. doi:10.1016/j.advwatres.2006.08.006
- Castellarin, A., Vogel, R. M., and Matalas, N. C. (2007). Multivariate probabilistic regional envelopes of extreme floods. *Journal of Hydrology*, 336(3-4), 376-390 . doi:10.1016/j.jhydrol.2007.01.007
- Castiglioni, S., Castellarin, A., and Montanari, A. (2009). Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. *Journal of Hydrology*, 378(3-4), 272-280. doi:10.1016/j.jhydrol.2009.09.032

- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A. (2010). Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. *Advances in Water Resources, In Press, Corrected Proof* doi:10.1016/j.advwatres.2010.04.009
- Cunderlik, J. M., and Burn, D. H. (2001). The use of flood regime information in regional flood frequency analysis. *Hydrological Sciences Journal*, 47(1), 77-92. doi:10.1080/02626660209492909
- EasyFit – *Distribution fitting software*, Retrieved 8/25/2010, from <http://www.mathwave.com>
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- ESRI. (1999). *ArcGIS*. USA:
- Fennessey, N., and Vogel, R. M. (1990). Regional flow-duration curves for ungauged sites in massachusetts. *Journal of Water Resources Planning and Management*, 116(4), 530-549. doi:10.1061/(ASCE)0733-9496(1990)116:4(530)
- Foster, H. A. (1934). Duration curves. *Trans. ASCE*, (99), 1213-1267.
- Franchini, M., and Suppo, M. (1996). Regional analysis of flow duration curves for a limestone region. *Water Resour Manage*, 10, 199-218.
- Gong, G., and Efron, B. (1983). A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician*, 37(1), 36-48. Retrieved from <http://www.jstor.org/stable/2685844>
- Goswami, M., O'Connor, K. M., and Bhattarai, K. P. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology*, 333(2-4), 517-531. doi:10.1016/j.jhydrol.2006.09.018
- Helsel, D. R., and Hirsh, R. M. (1992). *Statistical methods in water resources*. New York: Elsevier.
- The histogram*. Retrieved 8/12/2010, 2010, from
- Hohenberger, A. (2009). *Regression*. Retrieved August/12, 2010, from [www.ii.metu.edu.tr/~hohenberger/statistics/Chapter\\_5.ppt](http://www.ii.metu.edu.tr/~hohenberger/statistics/Chapter_5.ppt)
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1), 105-120. Retrieved from <http://www.jstor.org/stable/2345653>
- Hosking, J. R. M., and Wallis, J. R. *Regional frequency analysis: An approach based on L-moments* Cambridge ; Cambridge University Press, 1997.

- Kendall's tau and spearman's rank correlation coefficient | statistics solutions Retrieved 8/27/2010, 2010, from <http://www.statisticssolutions.com/methods-chapter/statistical-tests/kendall-spearman-rank-correlation-coefficient/>
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Vol. 9, No. 70*, 209-219.
- Mardia, K. V., and Jupp, P. E. (2008). *Directional statistics*. London: John Wiley and Sons, Inc. doi:10.1002/9780470316979
- McCuen, R. (1993). In Goodwin B., Hays M. and Papanikolaou S. (Eds.), *Microcomputer applications in statistical hydrology* (1st ed.). USA: Prentice Hall.
- McCuen, R. H., and Beighley, R. E. (2003). Seasonal flow frequency analysis. *Journal of Hydrology*, 279(1-4), 43-56. doi:10.1016/S0022-1694(03)00154-9
- Mimikou, M. (1990). Regional analysis of hydrological variables in greece. *Regionalization in Hydrology, IAHS Publication*, 191, 195-201.
- Mwakalila, S. (2003). Estimation of stream flows of ungauged catchments for river basin management. *Physics and Chemistry of the Earth, Parts A/B/C*, 28(20-27), 935-942. doi:10.1016/j.pce.2003.08.039
- Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282-290. doi:10.1016/0022-1694(70)90255-6
- NIST/SEMATECH e-handbook of statistical methods. (2010). Retrieved August 25, 2010, retrieved from <http://www.itl.nist.gov/div898/handbook/>
- Önöz, B., and Albostan, A. (2007). Hidroelektrik santral planlama ve işletmesinde yüksek akımların mevsimselliğinin belirlenmesi. *TMMOB VI. Enerji Sempozyumu - Küresel Enerji Politikaları Ve Türkiye Gerçeği*, Ankara. 196-206.
- Penche, C. (1998). *Layman's guidebook on how to develop a small hydro site* (2nd ed.) Comission of the European Communities. Retrieved from <http://europa.eu.int/en/comm/dg17/dg17home.htm>, Retrieved 19 December, 2010
- Pérez-Peña, J. V., Azañón, J. M., and Azor, A. (2009). CalHypso: An ArcGIS extension to calculate hypsometric curves and their statistical moments. applications to drainage basin analysis in SE spain. *Comput.Geosci.*, 35(6), 1214-1223. doi:10.1016/j.cageo.2008.06.006
- Quimpo, R. G., Alejandrino, A. A., and McNally, T. A. (1983). Regionalized flow duration for Philippines. *Journal of Water Resources Planning and Management*, 109(4), 320-330. doi:10.1061/(ASCE)0733-9496(1983)109:4(320)
- Sanborn, S. C., and Bledsoe, B. P. (2006). Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *Journal of Hydrology*, 325(1-4), 241-261. doi:10.1016/j.jhydrol.2005.10.018

- Smakhtin, V. Y., Hughes, D. A., and Creuse-Naudin, N. E. (1997). Regionalization of daily flow characteristics in part of the eastern cape, south africa. [Régionalisation des caractéristiques des débits journaliers dans une région du Cap Oriental (Afrique du Sud)] *Hydrological Sciences Journal*, 42(6), 919.
- Statistics for biomedical informatics, spring 2007* Retrieved 9/26/2010, 2010, from <http://episun7.med.utah.edu/~alun/teach/stats/>
- Soytekin, Arzu (2010). Evaluating the use of satellite-based precipitation estimates for discharge estimation in ungauged basins (Master's Thesis). Retrieved from <http://etd.lib.metu.edu.tr/upload/12612542/index.pdf>
- Viglione, A., Laio, F., and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3), W03428. doi:10.1029/v2006WR005095
- Vivoni, E. R., Di Benedetto, F., Grimaldi, S., and Eltahir, E. A. B. (2008). Hypsometric control on surface and subsurface runoff. *WATER RESOURCES RESEARCH*, 44 doi:10.1029/2008WR006931
- Weisstein, E. W. (2010). *Pareto distribution*. Retrieved August 25, 2010, from <http://mathworld.wolfram.com/ParetoDistribution.html>
- Wikipedia contributors. *Mean squared error*, 2010, from
- Yanık, B., and Avcı, İ. (2005 Ekim, Bölgesel debi süreklilik eğrilerinin elde edilmesi [determination of regional flow duration curves]. *İtüdergisi/d - Mühendislik Serisi*, 4(5), 19-30.
- Yanmaz, M.,Dr., Bozkuş, Z.,Dr., Akyürek, Z.,Dr., and Kentel, E.,Dr. (2007). *Nehir santrallerinin yer seçimi için havza hidrolojik-hidrolik modelinin kurulması: Çoruh havzası örneği [developing hydrologic-hydrolic basin model to select the site location for the hydroelectric power plants: Çoruh basin example]*. (Technical Report METU).
- Yu, P., Yang, T., and Wang, Y. (2002). Uncertainty analysis of regional flow duration curves. *Journal of Water Resources Planning and Management*, 128(6), 424-430. doi:10.1061/(ASCE)0733-9496(2002)128:6(424)

# APPENDIX A

## DATA

23-21 - Halı Defteri											
Dünya Düzeyi Gözetim Yarımları											
İSTASYON ADI = İSTASYON NO =23-21 YILI=1970											
GÜN/AY	EKM	KSM	APR	OKT	SBT	MRT	NSN	MYS	HZR	TMZ	AGS
1	1.8	1.25	0.77	0.57	1.6	3	10	0	1.1	0.65	2.5
2	2.2	1.00	0.86	0.57	1.8	1.9	9.6	0	1.1	0.97	1.4
3	3	1.60	0.86	0.57	1.8	4.7	9.6	0	0.97	1.4	1.15
4	2.8	1.15	0.65	0.57	2.2	4.3	9.6	0	0.38	1.25	1.1
5	1	0.97	0.86	0.57	2.2	1.3	10.6	0	0.12	0.97	0.50
6	5.1	1.1	0.97	0.57	2	2.8	24	0	0.32	0.86	2.2
7	3.9	0.97	0.86	0.57	2	1.8	34	0	1.1	0.38	3
8	4.7	0.97	0.65	0.57	2	3	34	0	1.8	0.27	2.8
9	1.1	0.86	0.65	0.57	2	1.6	21	0	1.4	0.27	2.2
10	3	0.86	0.65	0.57	3	3.9	10	0	0.86	0.27	1.15
11	1.1	0.97	0.65	0.48	3.6	3.9	17.5	0	0.57	0.32	0.97
12	1.6	0.97	0.57	0.48	3.6	1.9	11	0	0.48	0.48	0.50
13	3.3	0.86	0.57	0.48	3.3	4.7	10	0	0.38	1.25	0.97
14	2.3	0.86	0.57	0.48	2.8	8.3	21	0	0.38	1.8	0.86
15	1.8	0.86	0.77	0.48	2.2	1.5	24	0	0.38	3.9	0.42
16	2.2	0.77	0.77	0.48	2.2	1.3	36	0	1.8	0.22	0.52
17	1.1	0.86	0.77	0.48	2.2	1.6	36	0	1.8	12.5	0.65
18	0.97	0.86	0.65	0.48	2.2	3.9	33	0	1.4	6.6	0.65
19	0.77	0.86	0.77	0.48	1.3	6.6	15	0	1.75	3.3	1.1
20	0.42	0.97	0.86	0.48	3.3	5.5	15.5	0	1.25	2.8	1.4
21	0.32	0.77	0.86	0.57	3.3	4.7	8.3	0	1.8	1.2	2.8
22	0.65	0.65	0.97	0.57	3.3	4.3	7.7	0	0.48	1.6	2.5
23	1.25	0.65	1.1	0.57	3.3	1.6	7	0	0.38	1.4	2.5
24	2.2	0.65	0.97	0.57	3.6	3.3	6.6	0	0.38	2.8	1.4
25	3	0.65	0.77	0.57	3	3	6.6	0	0.38	4.7	1.15
26	1.3	0.65	0.77	0.57	2	3	6.6	0	0.38	7.7	0.97
27	2.8	0.65	0.77	0.57	2	3.6	6.6	0	0.52	10	0.86
28	2.2	0.57	0.65	0.57	1.2	4.7	6.6	0	0.65	14.5	0.77
29	2	0.57	0.65	0.65	//	6	6.6	0	0.97	17.5	1.15
30	2	0.65	0.65	0.65	//	6.6	6.6	0	0.66	5.1	4.1
31	1.4	//	0.65	0.65	//	10	//	0	3.3	3.9	
İSTASYON ADI = İSTASYON NO =23-21 YILI=1972											
GÜN/AY	EKM	KSM	APR	OKT	SBT	MRT	NSN	MYS	HZR	TMZ	AGS
1	0.2	10	10	11	11	8.9	16	26	19.5	3.1	2.1
2	8.2	9.2	11.5	11	11	10	16.5	22	32	2.8	2.1
3	8.5	10	15	11	10.5	10	17	19.5	30	1.65	2.1
4	0.9	10.5	17.5	11	10.5	9.2	16.5	18.5	28	1.2	2.1
5	9.2	11	17.5	11	10.5	9.2	17	17.5	35	0.72	1.9

23-39 - KARAPINAR DERESİ -KARAPINAR											
Dünya Düzeyi Gözetim Yarımları											
1964											
GÜN/AY	EKM	KSM	APR	OKT	SBT	MRT	NSN	MYS	HZR	TMZ	AGS
1	0.056	0.056	0.056	0.056	0.056	1.63	0.023	2.32	0.011	0.22	0.22
2	0.056	0.056	0.056	0.056	0.056	1.63	0.023	1.63	0.011	0.22	0.22
3	0.056	0.056	0.056	0.056	0.056	1.13	0.023	1.13	0.011	0.22	0.22
4	0.056	0.056	0.056	0.056	0.056	0.98	0.023	1.13	0.023	0.22	0.22
5	0.056	0.056	0.056	0.056	0.056	1.1	0.023	0.46	0.115	0.22	0.22
6	0.056	0.056	0.056	0.056	0.056	5.9	0.023	0.385	0.115	0.22	0.22
7	0.056	0.056	0.056	0.056	0.056	2.12	0.023	0.22	0.385	0.22	0.22
8	0.056	0.056	0.056	0.056	0.056	0.385	0.023	0.385	0.385	0.22	0.22
9	0.056	0.056	0.056	0.056	0.056	0.115	0.056	0.385	0.385	0.22	0.22
10	0.056	0.056	0.056	0.056	0.056	0.115	0.023	0.385	0.385	0.22	0.22
11	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
12	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
13	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
14	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
15	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
16	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
17	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
18	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
19	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
20	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
21	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
22	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
23	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
24	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
25	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
26	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
27	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
28	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
29	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
30	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22
31	0.056	0.056	0.056	0.056	0.056	0.056	0.385	0.385	0.385	0.22	0.22

Figure A.1. Raw data of DSI daily flow discharges





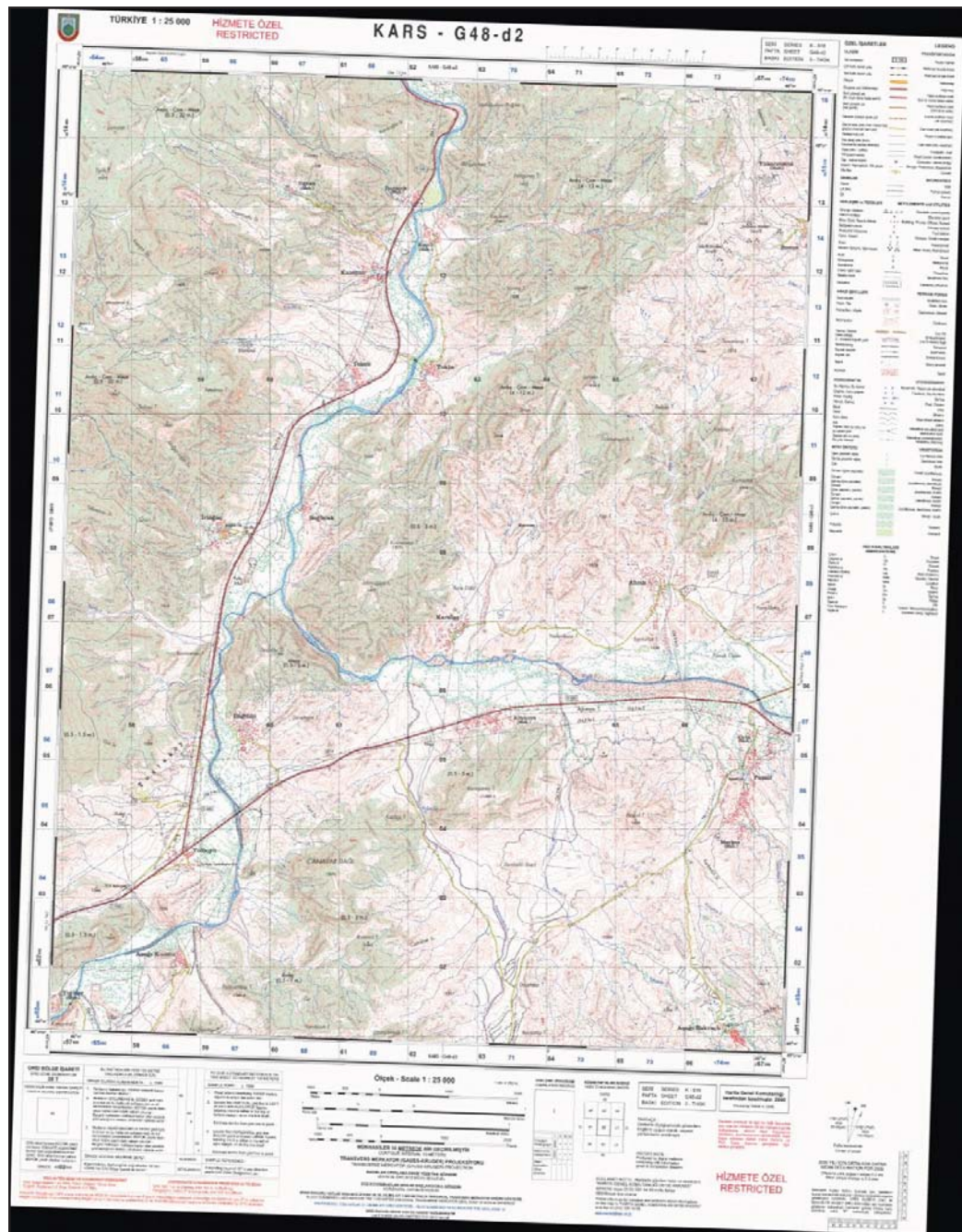


Figure A.5. G48-d2 map section raster image



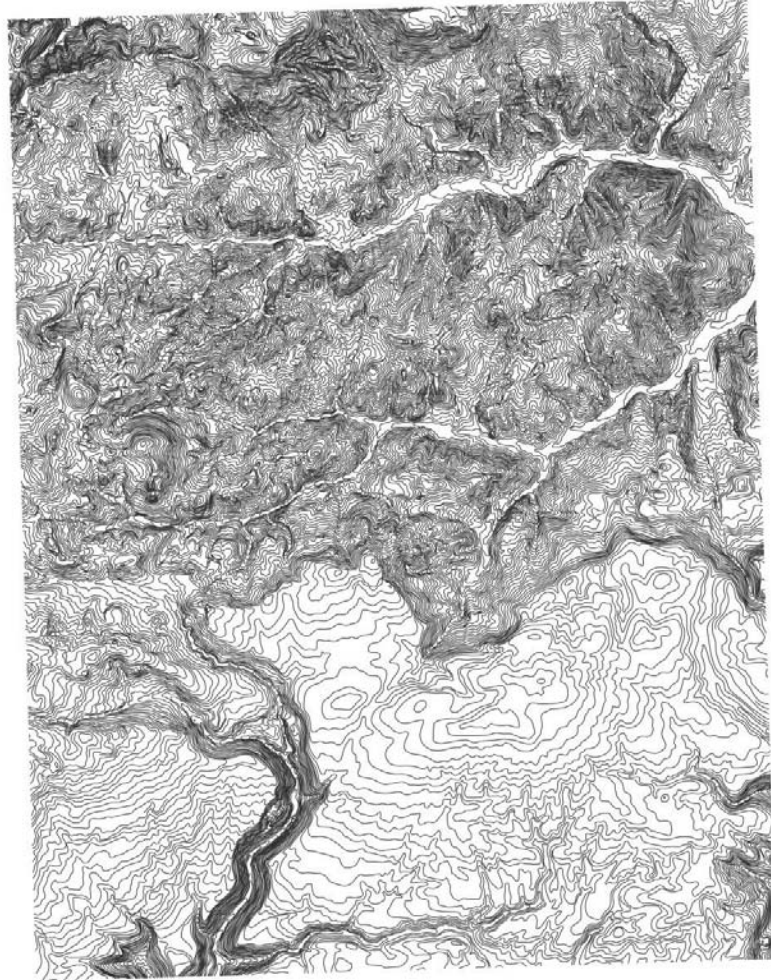
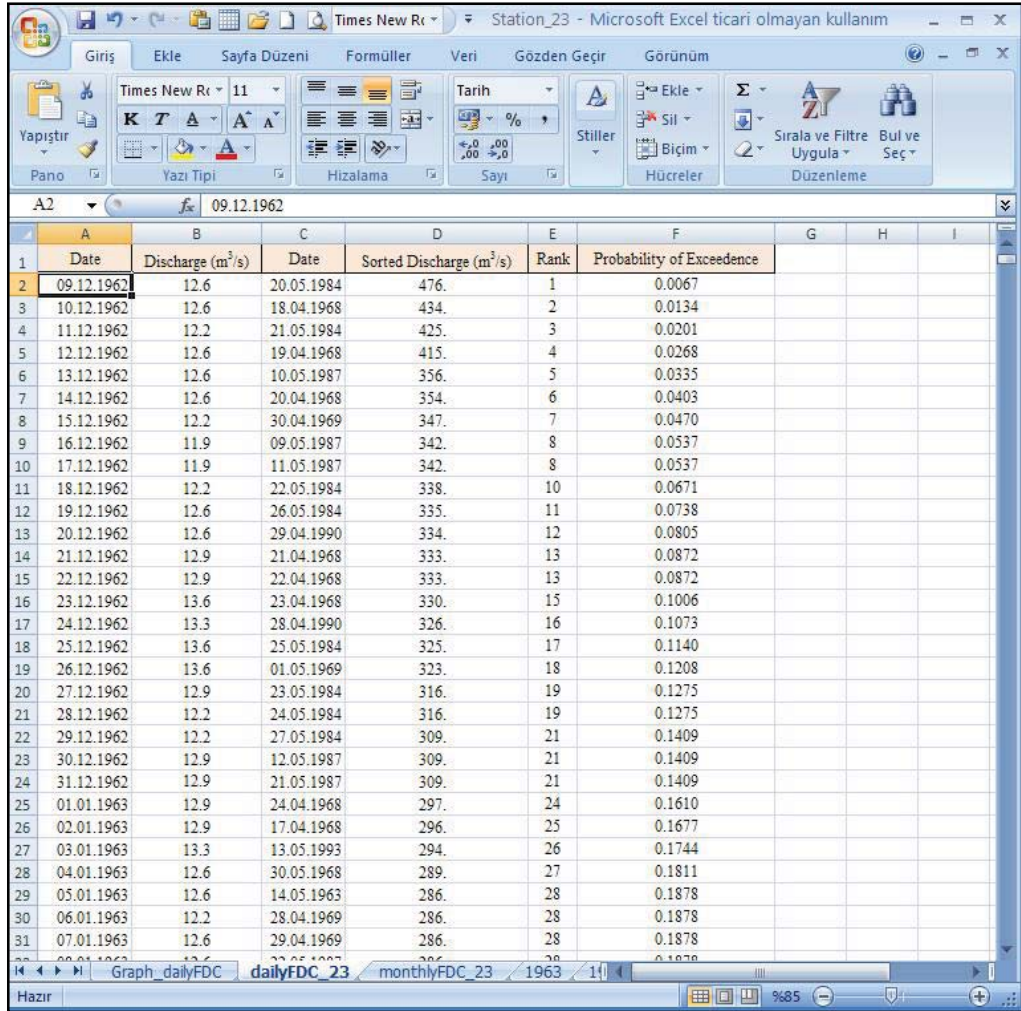


Figure A.6. H47-d2 map section vector image of elevation data

## APPENDIX B

### HYDROLOGIC DATA PREPROCESSING



	A	B	C	D	E	F	G	H	I
	Date	Discharge (m³/s)	Date	Sorted Discharge (m³/s)	Rank	Probability of Exceedence			
2	09.12.1962	12.6	20.05.1984	476.	1	0.0067			
3	10.12.1962	12.6	18.04.1968	434.	2	0.0134			
4	11.12.1962	12.2	21.05.1984	425.	3	0.0201			
5	12.12.1962	12.6	19.04.1968	415.	4	0.0268			
6	13.12.1962	12.6	10.05.1987	356.	5	0.0335			
7	14.12.1962	12.6	20.04.1968	354.	6	0.0403			
8	15.12.1962	12.2	30.04.1969	347.	7	0.0470			
9	16.12.1962	11.9	09.05.1987	342.	8	0.0537			
10	17.12.1962	11.9	11.05.1987	342.	8	0.0537			
11	18.12.1962	12.2	22.05.1984	338.	10	0.0671			
12	19.12.1962	12.6	26.05.1984	335.	11	0.0738			
13	20.12.1962	12.6	29.04.1990	334.	12	0.0805			
14	21.12.1962	12.9	21.04.1968	333.	13	0.0872			
15	22.12.1962	12.9	22.04.1968	333.	13	0.0872			
16	23.12.1962	13.6	23.04.1968	330.	15	0.1006			
17	24.12.1962	13.3	28.04.1990	326.	16	0.1073			
18	25.12.1962	13.6	25.05.1984	325.	17	0.1140			
19	26.12.1962	13.6	01.05.1969	323.	18	0.1208			
20	27.12.1962	12.9	23.05.1984	316.	19	0.1275			
21	28.12.1962	12.2	24.05.1984	316.	19	0.1275			
22	29.12.1962	12.2	27.05.1984	309.	21	0.1409			
23	30.12.1962	12.9	12.05.1987	309.	21	0.1409			
24	31.12.1962	12.9	21.05.1987	309.	21	0.1409			
25	01.01.1963	12.9	24.04.1968	297.	24	0.1610			
26	02.01.1963	12.9	17.04.1968	296.	25	0.1677			
27	03.01.1963	13.3	13.05.1993	294.	26	0.1744			
28	04.01.1963	12.6	30.05.1968	289.	27	0.1811			
29	05.01.1963	12.6	14.05.1963	286.	28	0.1878			
30	06.01.1963	12.2	28.04.1969	286.	28	0.1878			
31	07.01.1963	12.6	29.04.1969	286.	28	0.1878			

Figure B.1. EIE-2323 arranged data

- **EIE-2329**

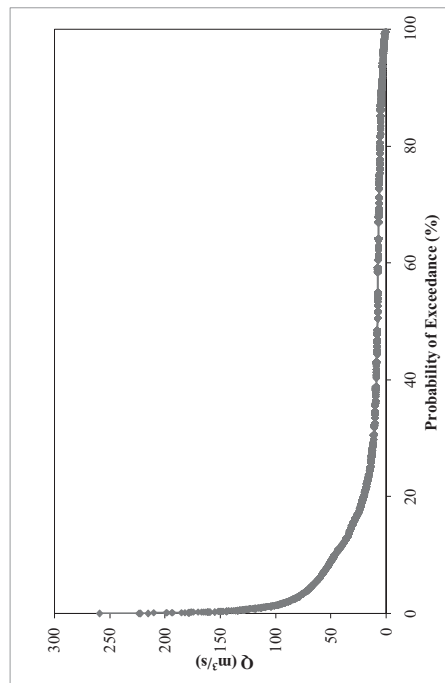


Figure B.2. Daily FDC of EIE-2329

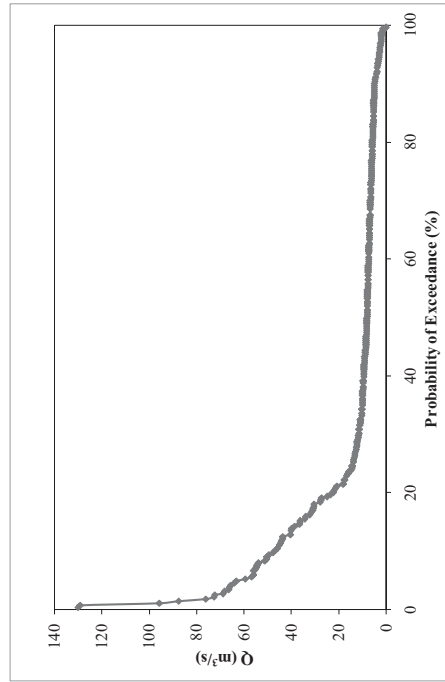


Figure B.3.. Monthly FDC of EIE-2329

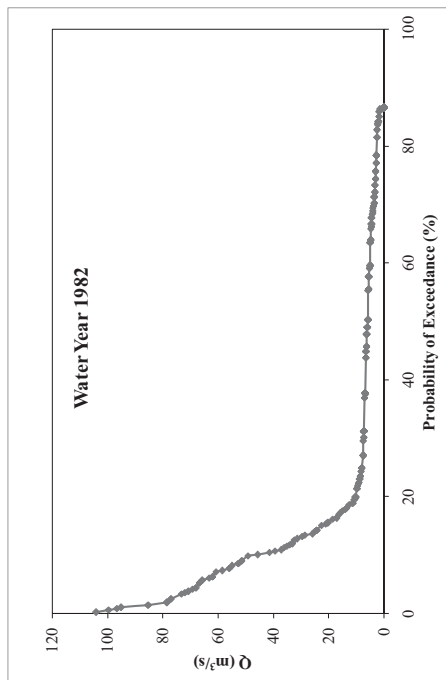


Figure B.4. EIE-2329 1982 water year yearly FDC

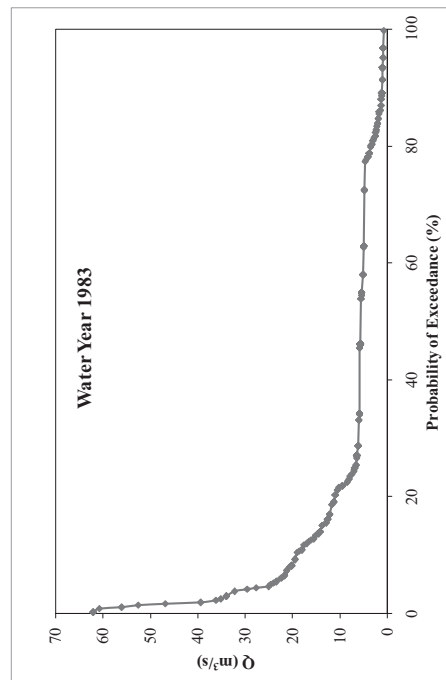


Figure B.5. EIE-2329 1983 water year yearly FDC

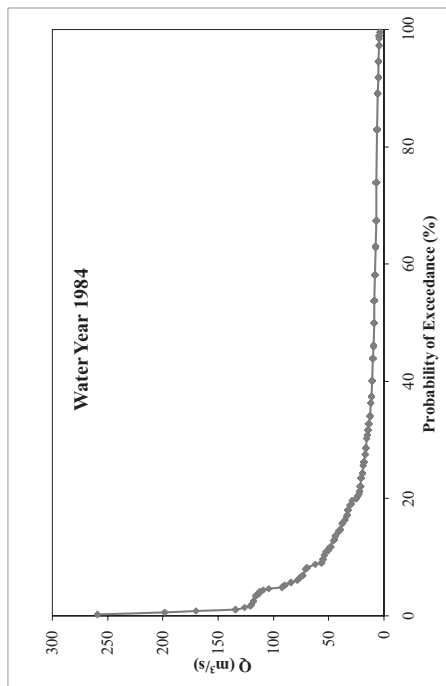


Figure B.6. EIE-2329 1984 water year yearly FDC

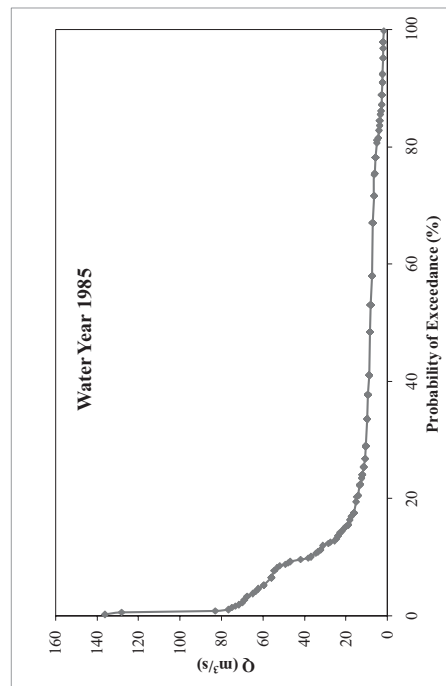


Figure B.7. EIE-2329 1985 water year yearly FDC

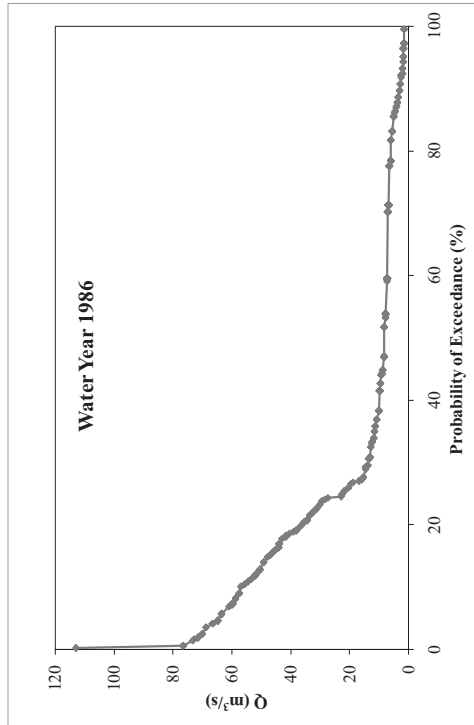


Figure B.8. EIE-2329 1986 water year yearly FDC

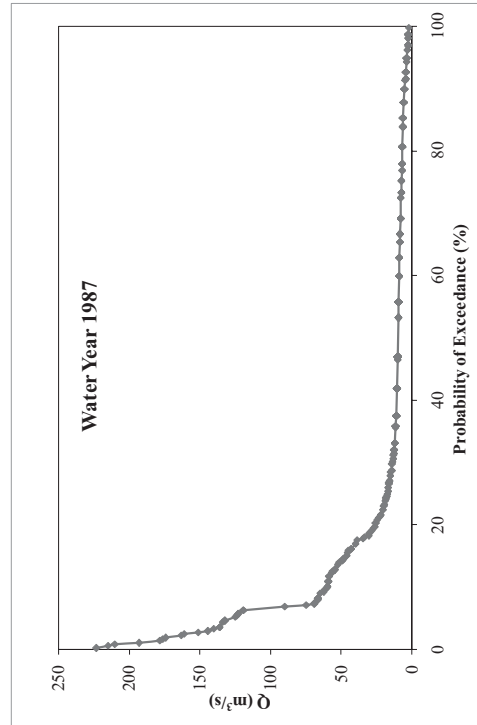


Figure B.9. EIE-2329 1987 water year yearly FDC

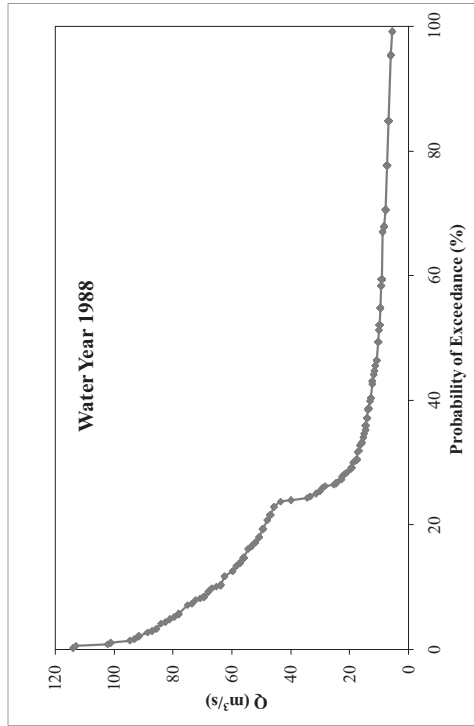


Figure B.10. EIE-2329 1988 water year yearly FDC

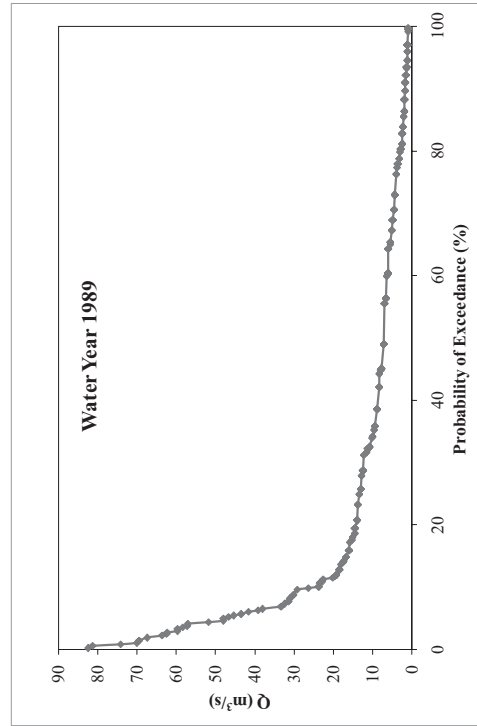


Figure B.11. EIE-2329 1989 water year yearly FDC

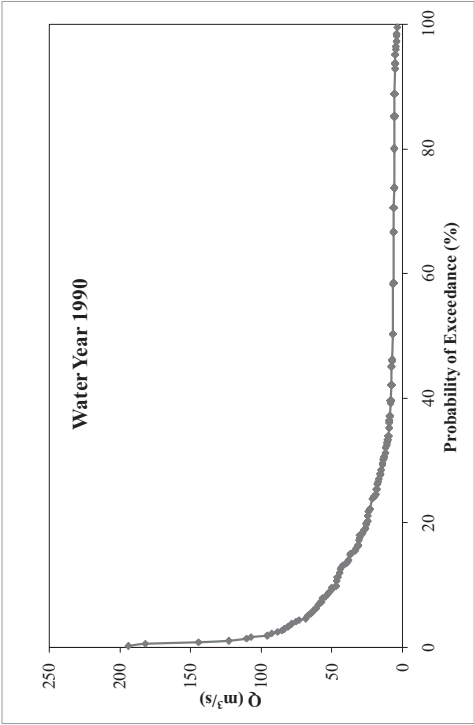


Figure B.12. EIE-2329 1990 water year yearly FDC

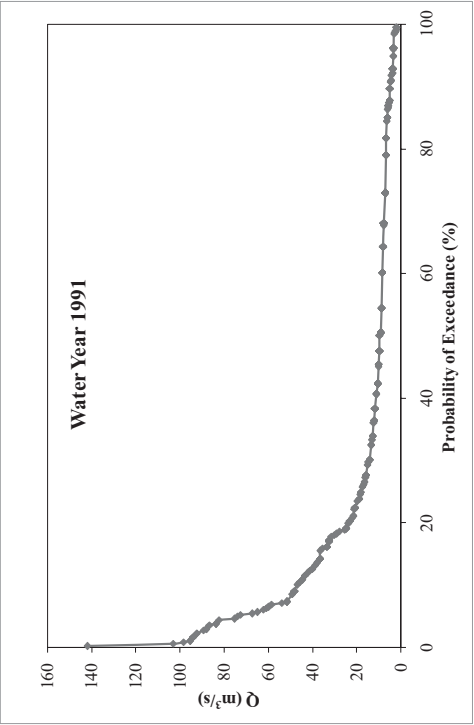


Figure B.13. EIE-2329 1991 water year yearly FDC

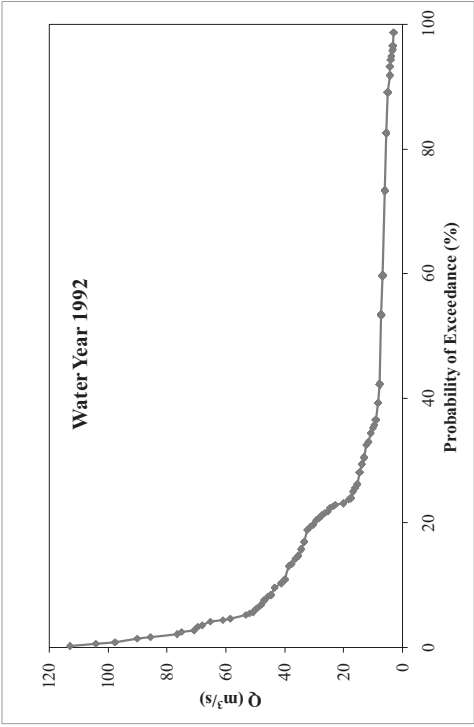


Figure B.14. EIE-2329 1992 water year yearly FDC

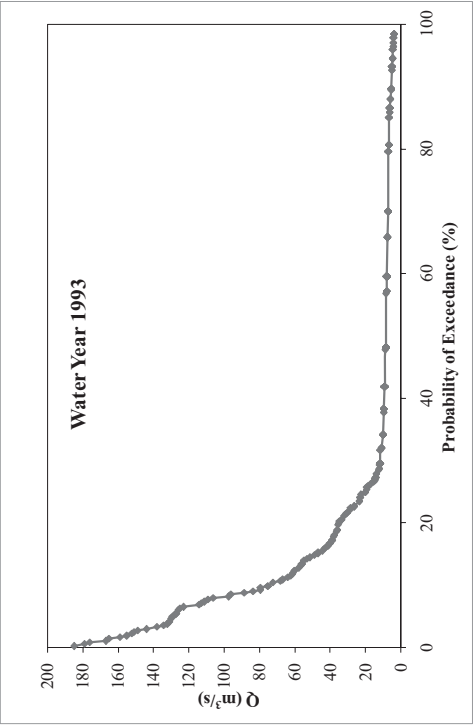


Figure B.15. EIE-2329 1993 water year yearly FDC



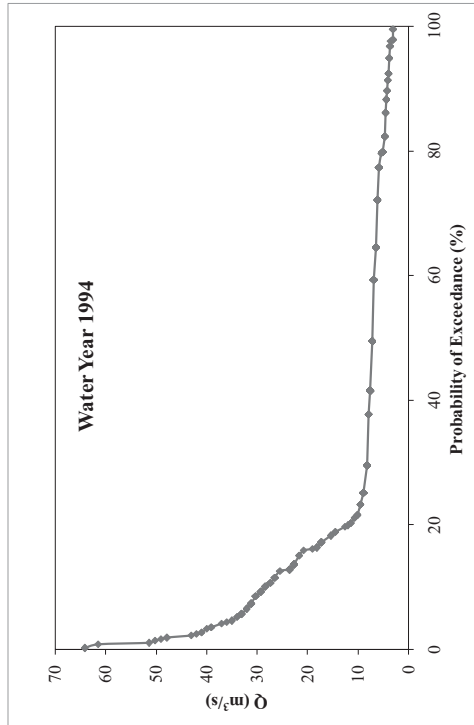


Figure B.16. EIE-2329 1994 water year yearly FDC

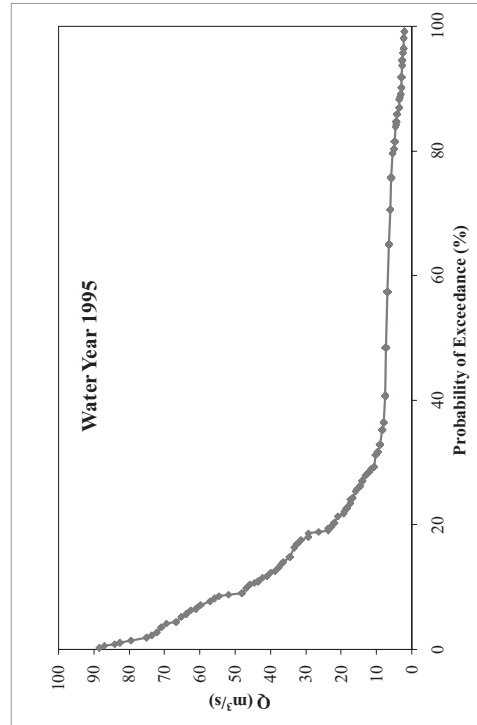


Figure B.17. EIE-2329 1995 water year yearly FDC

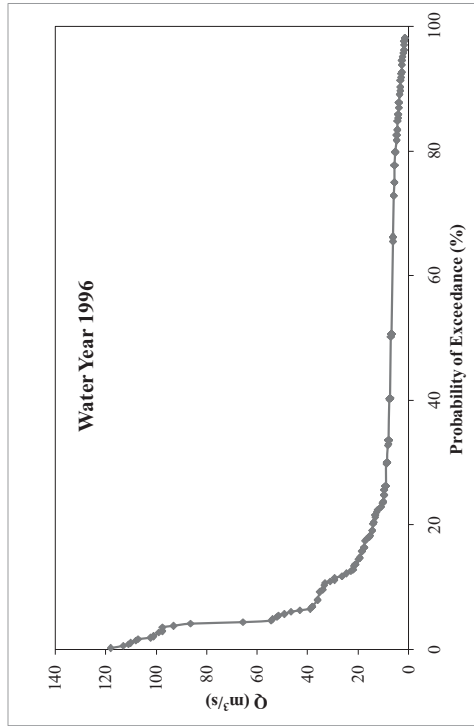


Figure B.18. EIE-2329 1996 water year yearly FDC

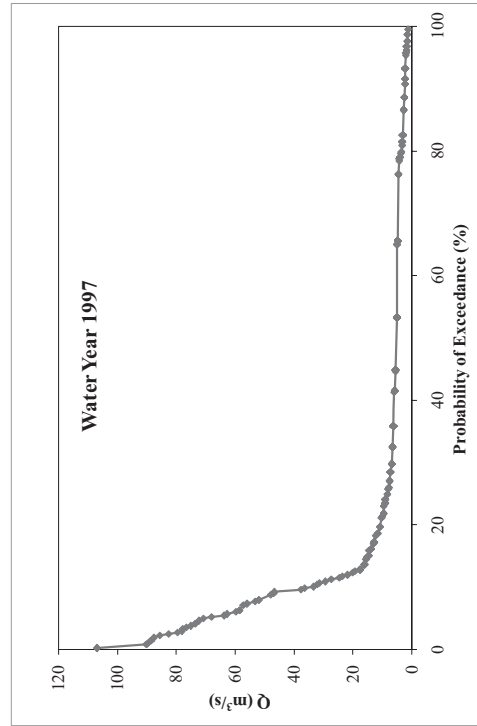


Figure B.19. EIE-2329 1997 water year yearly FDC

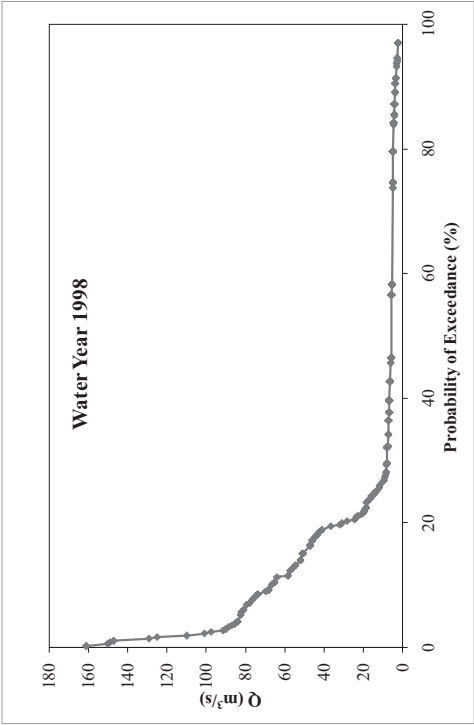


Figure B.20. EIE-2329 1998 water year yearly FDC

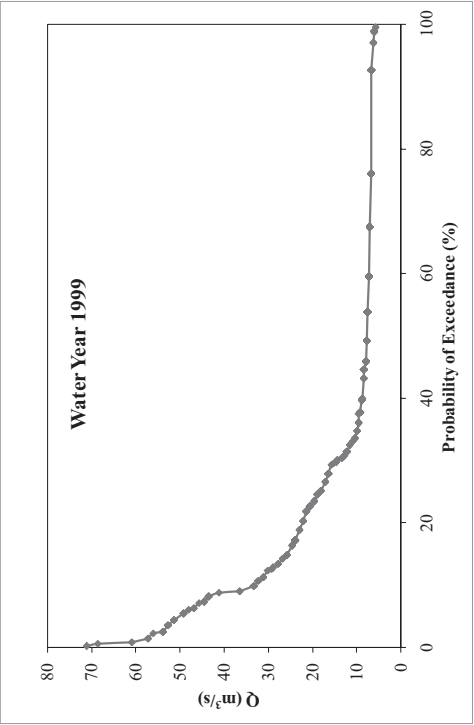


Figure B.21. EIE-2329 1999 water year yearly FDC

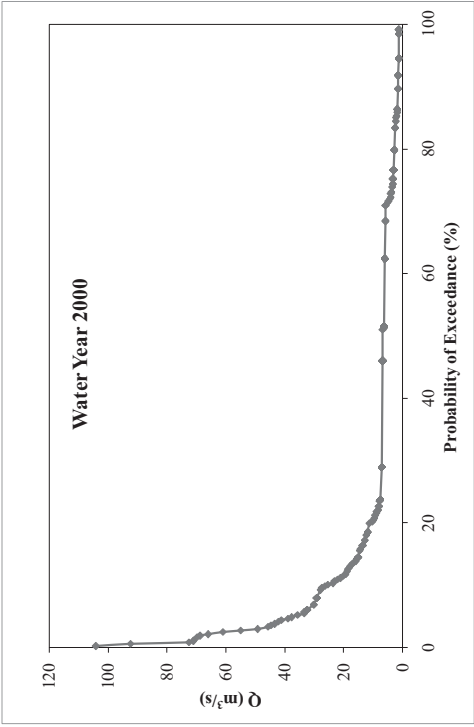


Figure B.22. EIE-2329 2000 water year yearly FDC

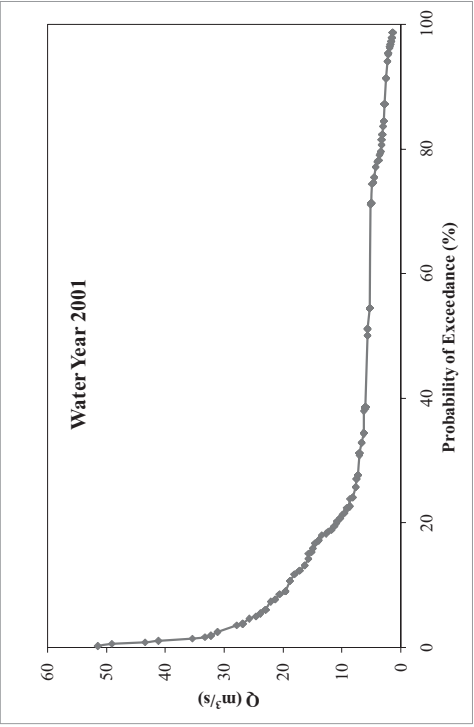


Figure B.23. EIE-2329 2001 water year yearly FDC

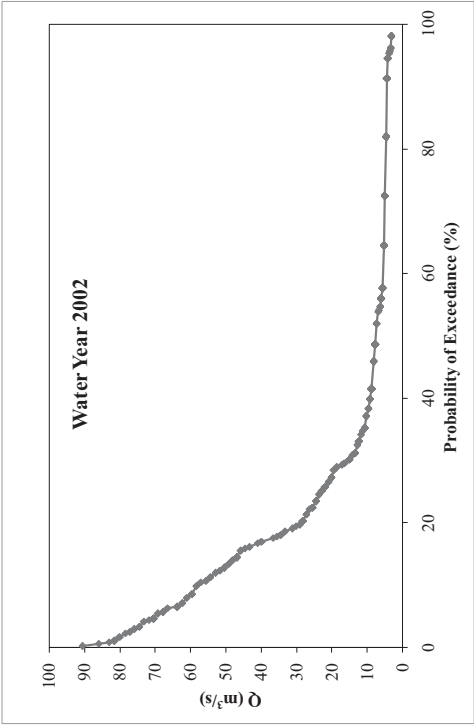


Figure B.24. EIE-2329 2002 water year yearly FDC

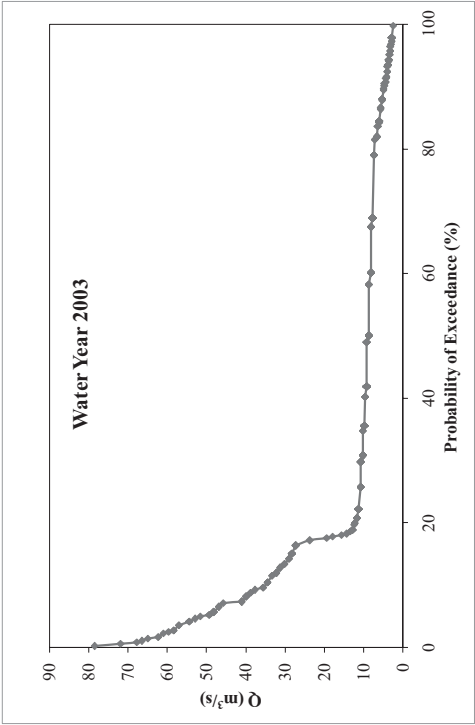


Figure B.25. EIE-2329 2003 water year yearly FDC

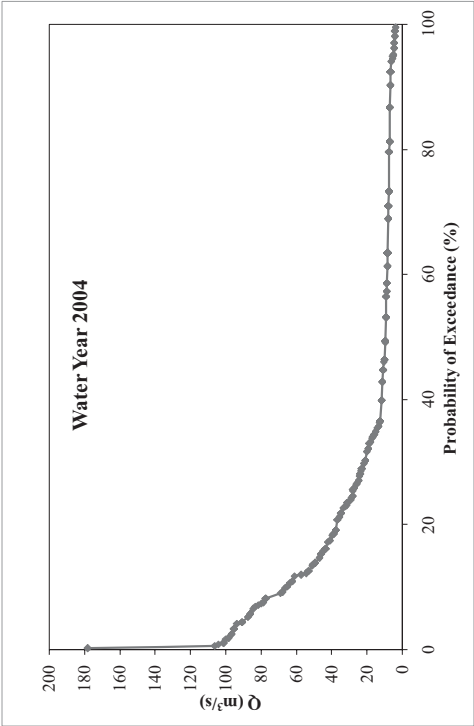


Figure B.26. EIE-2329 2004 water year yearly FDC

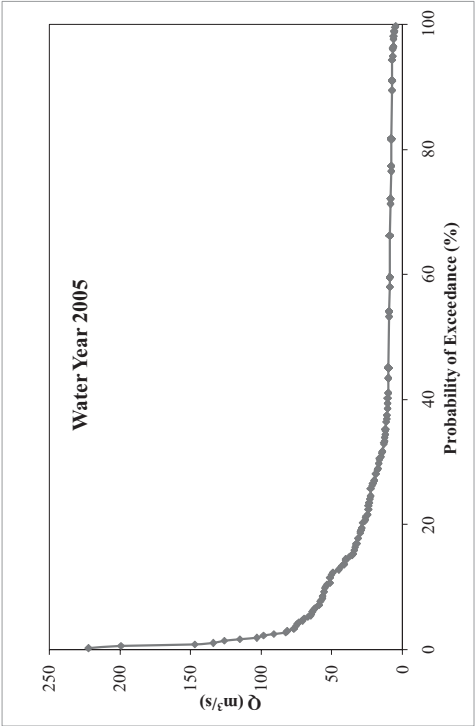


Figure B.27. EIE-2329 2005 water year yearly FDC

- **EIE-2323**

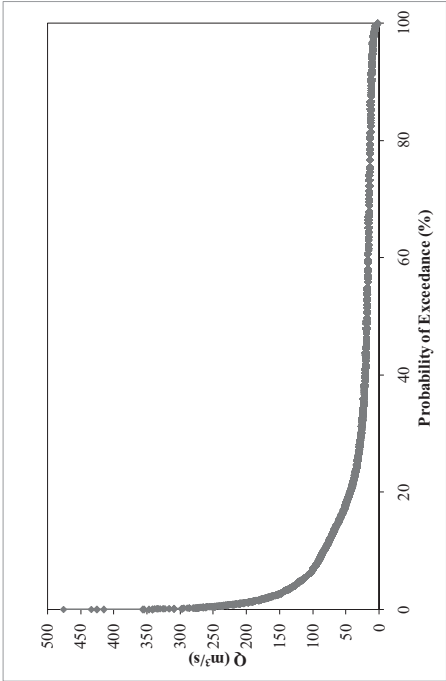


Figure B.28. Daily FDC of EIE-2323

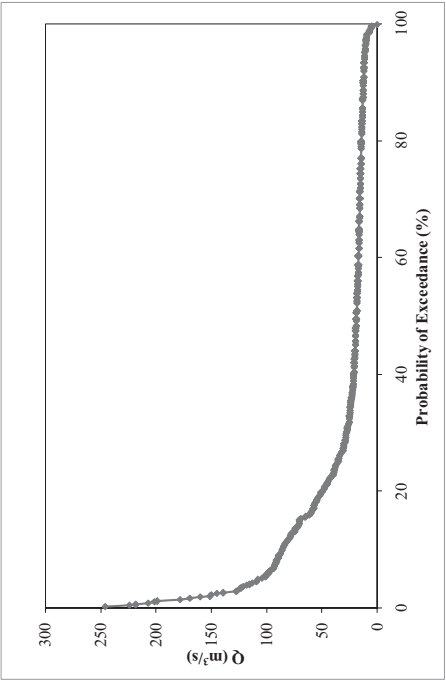


Figure B.29. Monthly FDC of EIE-2323

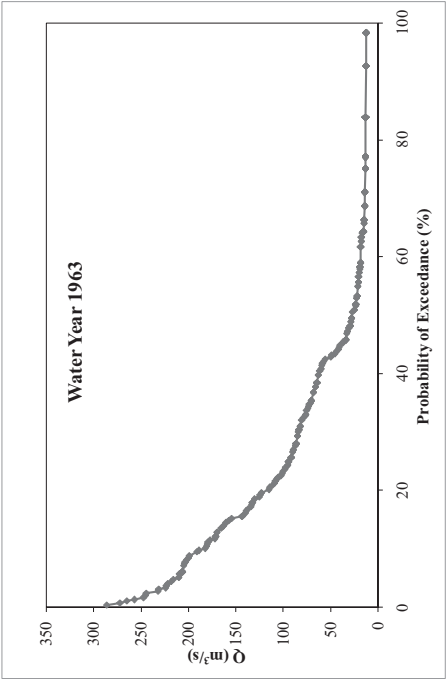


Figure B.30. EIE-2323 1963 water year yearly FDC

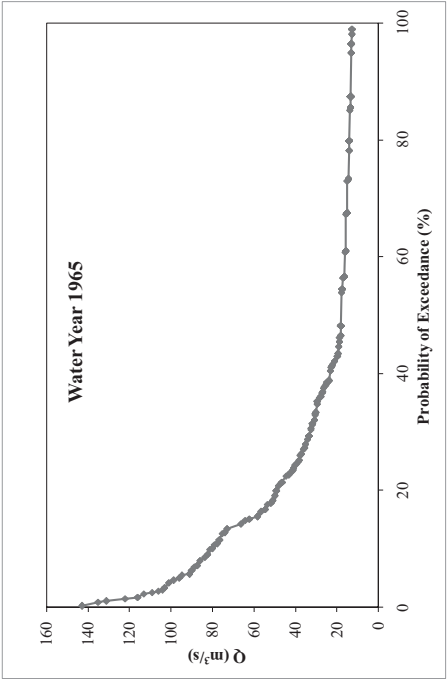


Figure B.31. EIE-2323 1965 water year yearly FDC

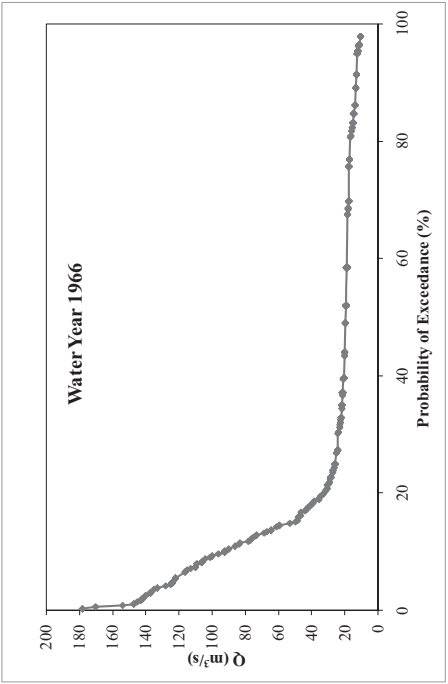


Figure B.32. EIE-2323 1966 water year yearly FDC

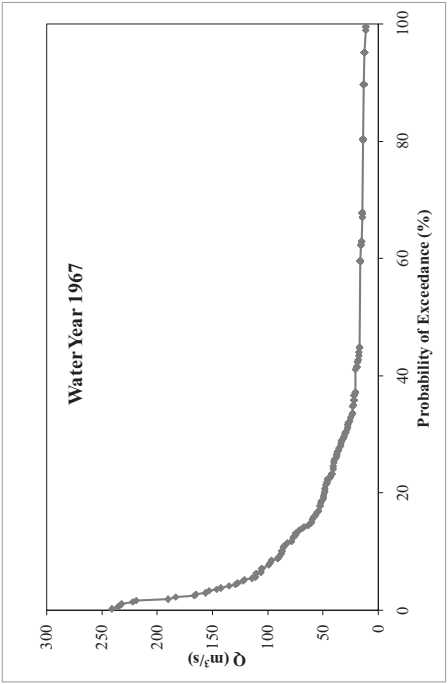


Figure B.33. EIE-2323 1967 water year yearly FDC

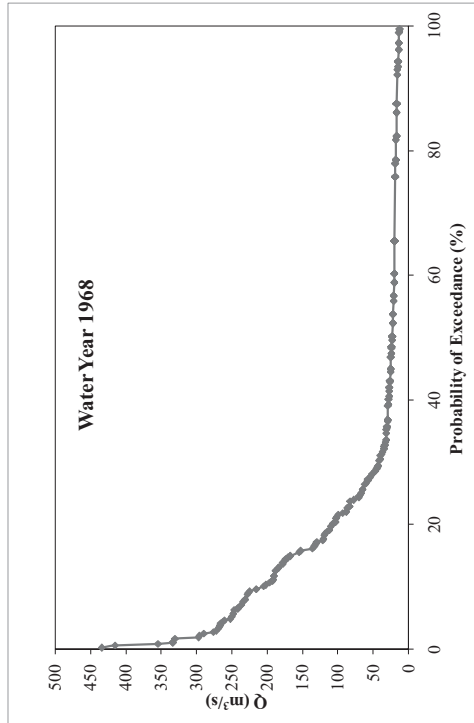


Figure B.34. EIE-2323 1968 water year yearly FDC

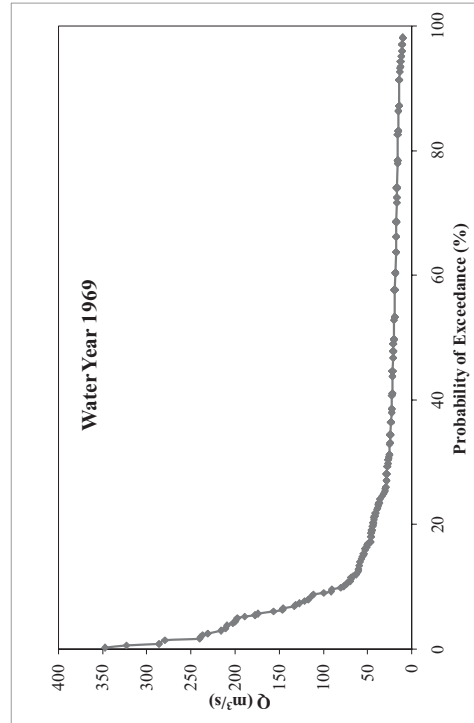


Figure B.35. EIE-2323 1969 water year yearly FDC

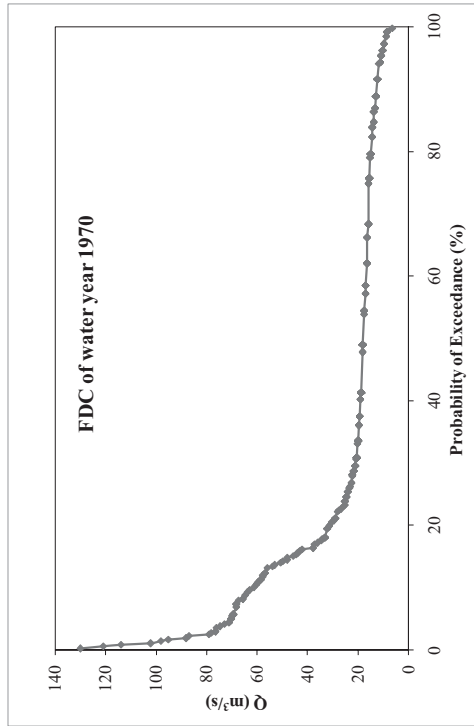


Figure B.36. EIE-2323 1970 water year yearly FDC

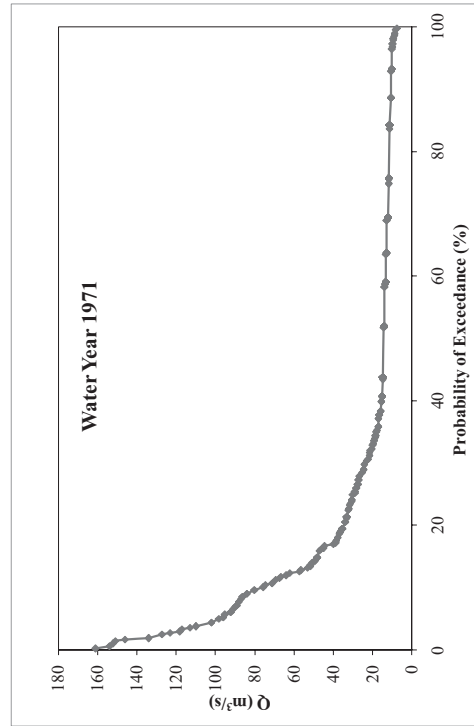


Figure B.37. EIE-2323 1971 water year yearly FDC

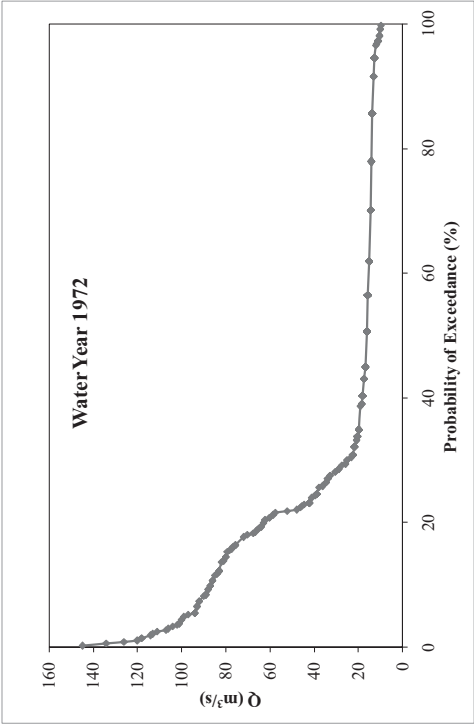


Figure B.38. EIE-2323 1972 water year yearly FDC

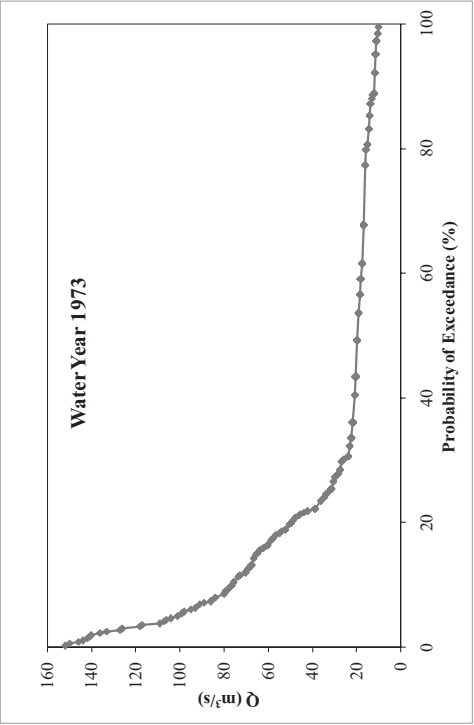


Figure B.39. EIE-2323 1973 water year yearly FDC

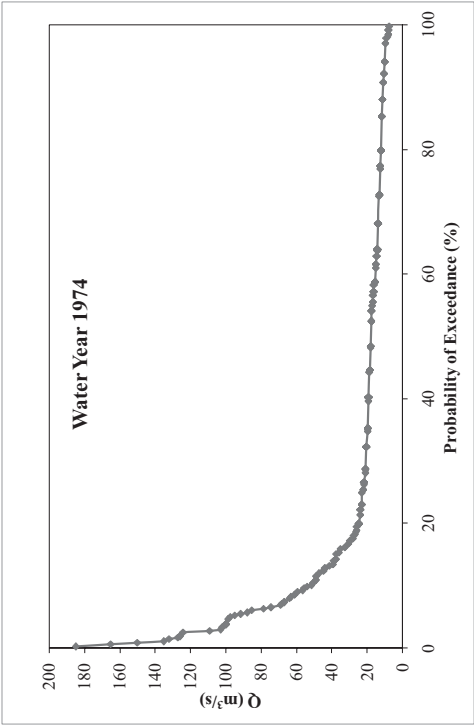


Figure B.40. EIE-2323 1974 water year yearly FDC

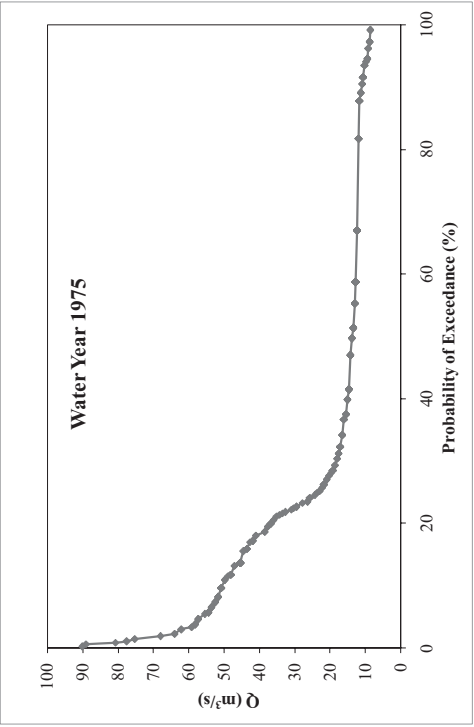


Figure B.41. EIE-2323 1975 water year yearly FDC

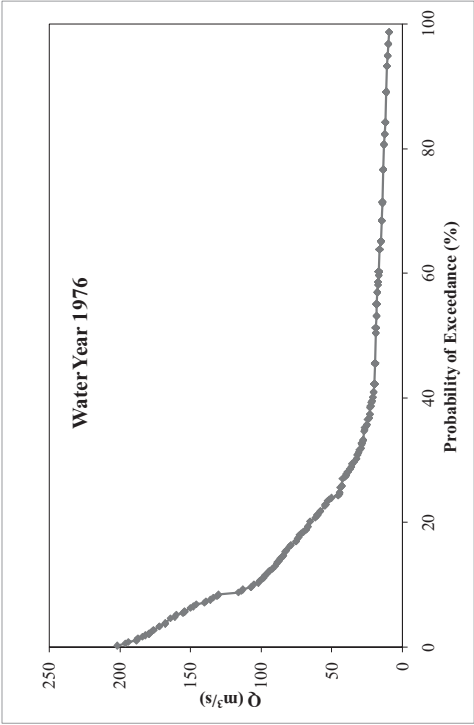


Figure B.42. EIE-2323 1976 water year yearly FDC

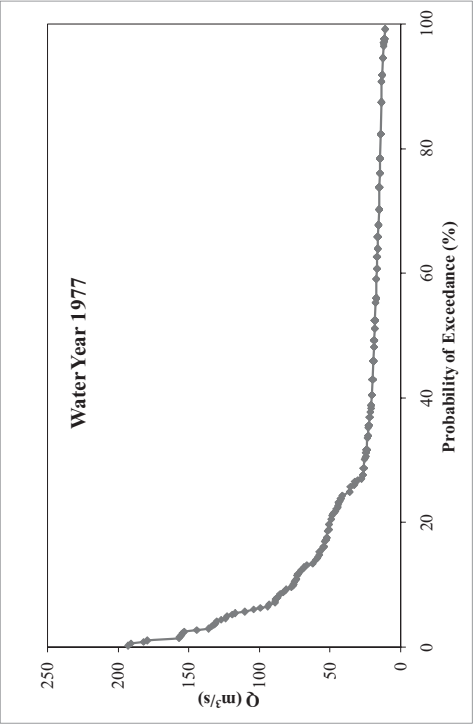


Figure B.43. EIE-2323 1977 water year yearly FDC

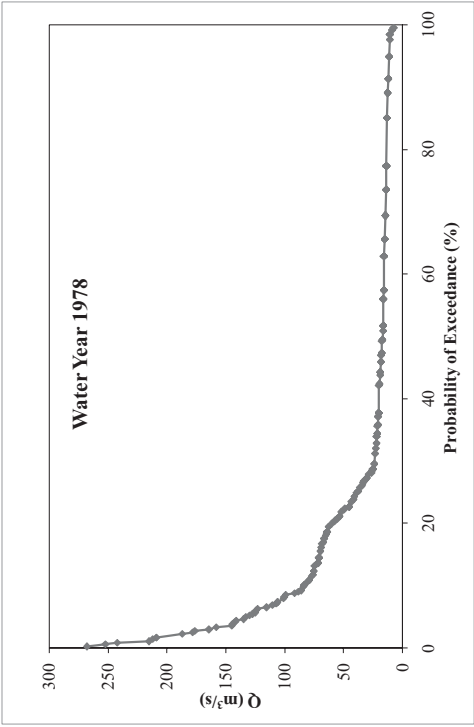


Figure B.44. EIE-2323 1978 water year yearly FDC

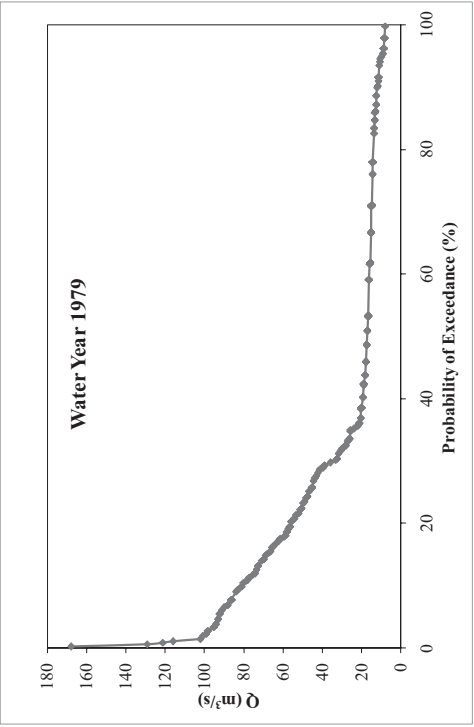


Figure B.45. EIE-2323 1979 water year yearly FDC



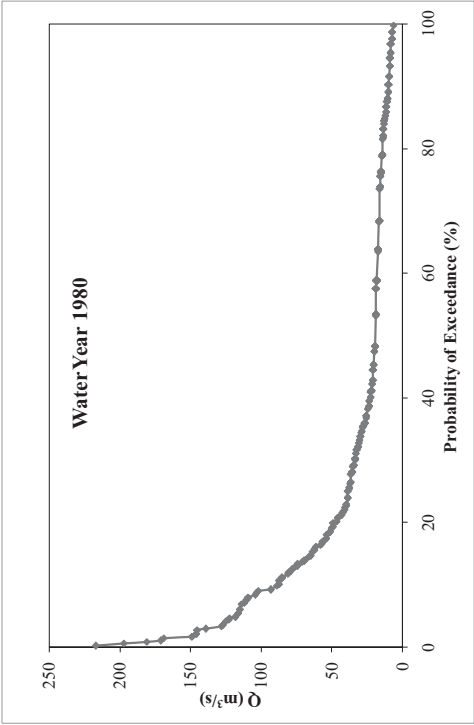


Figure B.46. EIE-2323 1980 water year yearly FDC

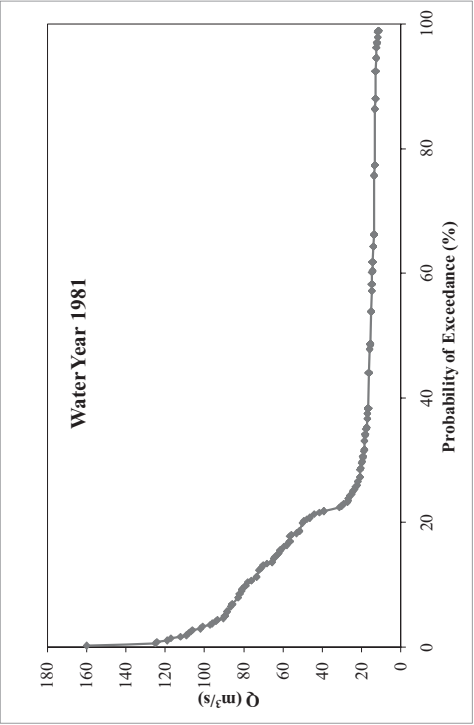


Figure B.47. EIE-2323 1981 water year yearly FDC

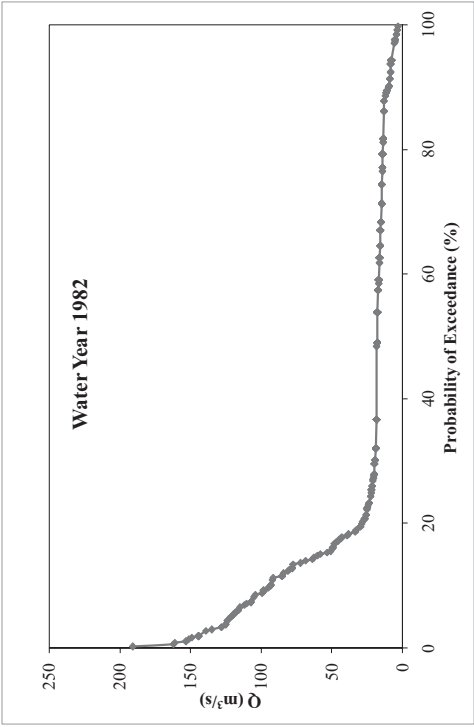


Figure B.48. EIE-2323 1982 water year yearly FDC

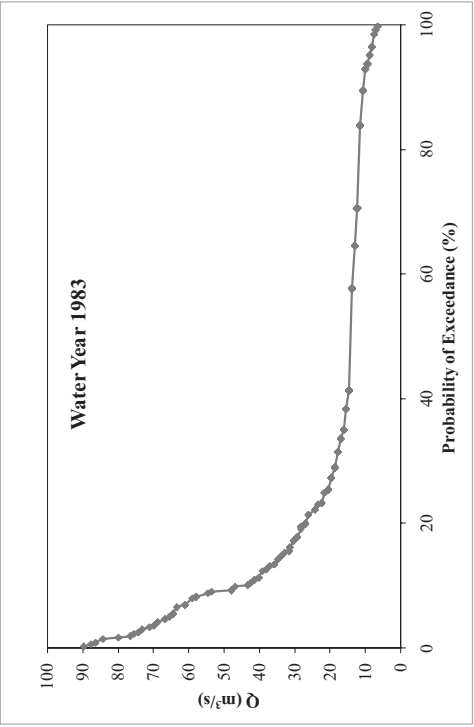


Figure B.49. EIE-2323 1983 water year yearly FDC

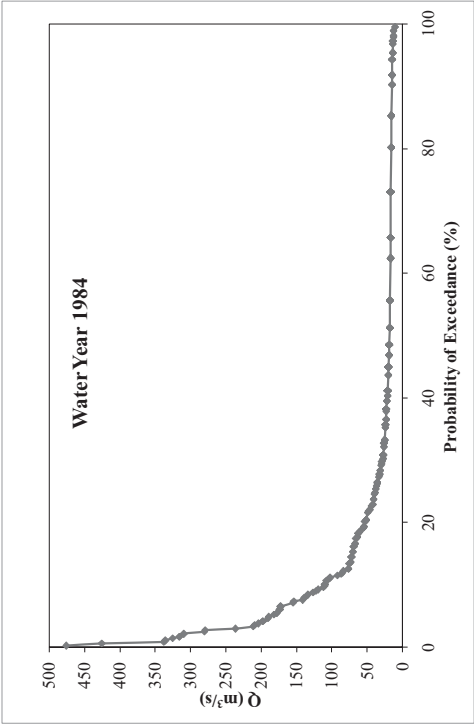


Figure B.50. EIE-2323 1984 water year yearly FDC

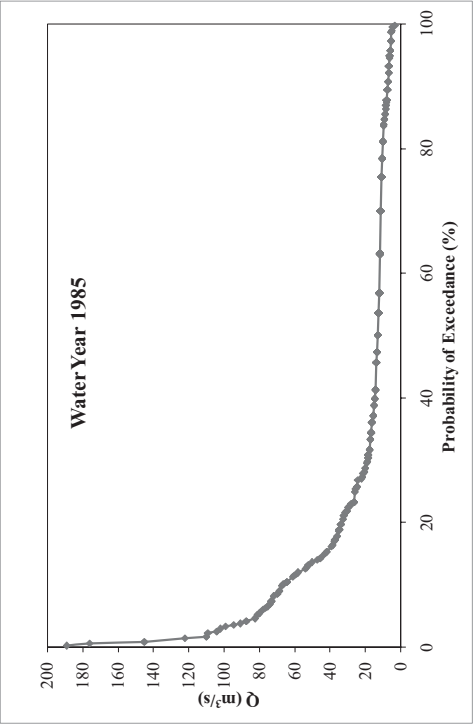


Figure B.51. EIE-2323 1985 water year yearly FDC

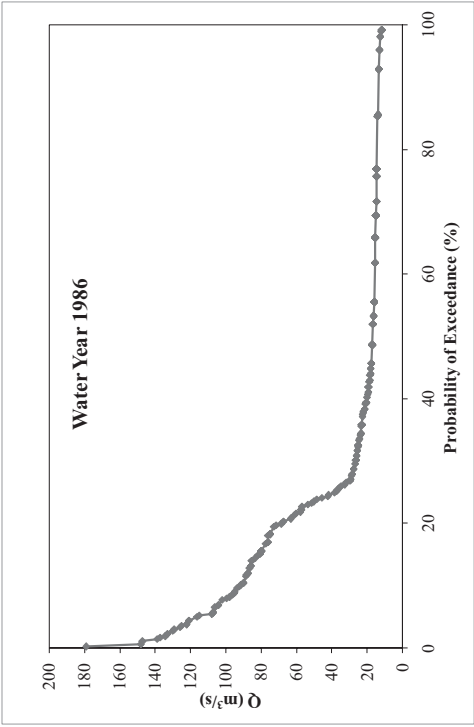


Figure B.52. EIE-2323 1986 water year yearly FDC

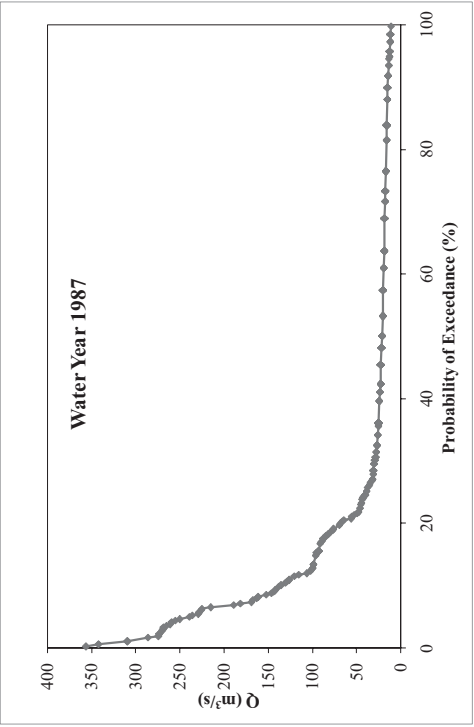


Figure B.53. EIE-2323 1987 water year yearly FDC

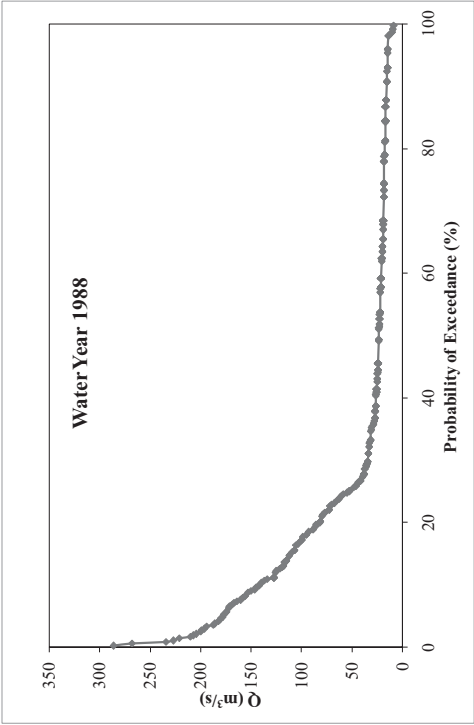


Figure B.54. EIE-2323 1988 water year yearly FDC

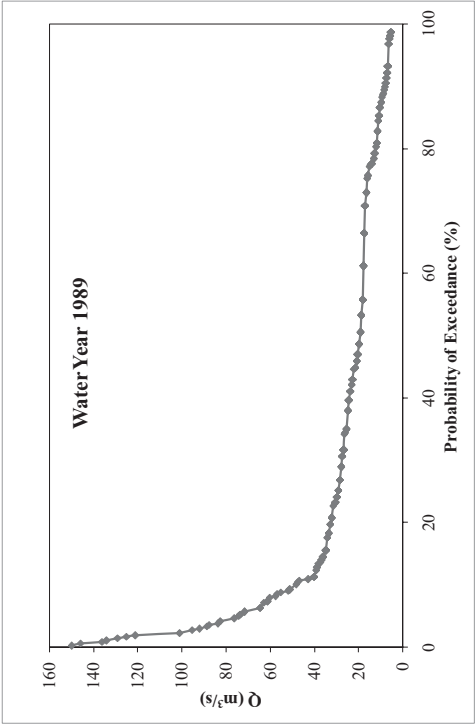


Figure B.55. EIE-2323 1989 water year yearly FDC

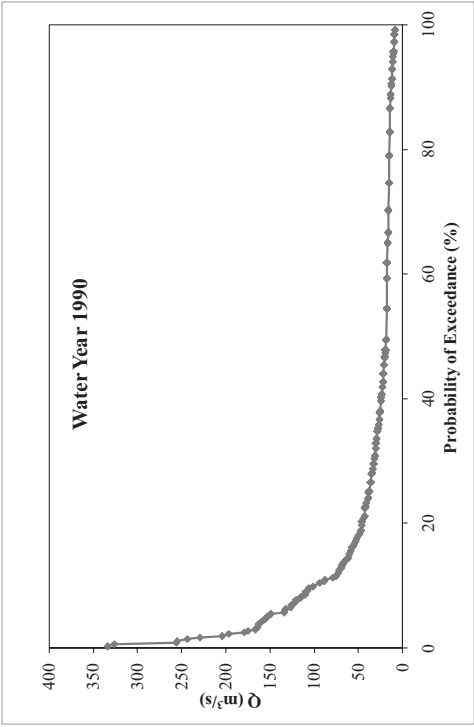


Figure B.56. EIE-2323 1990 water year yearly FDC

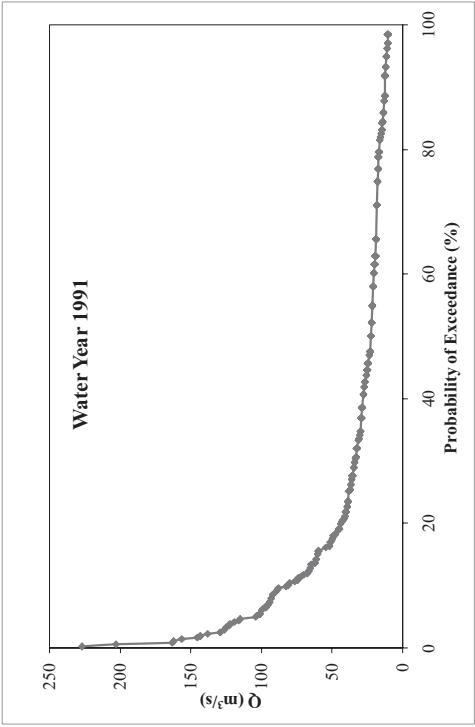


Figure B.57. EIE-2323 1991 water year yearly FDC

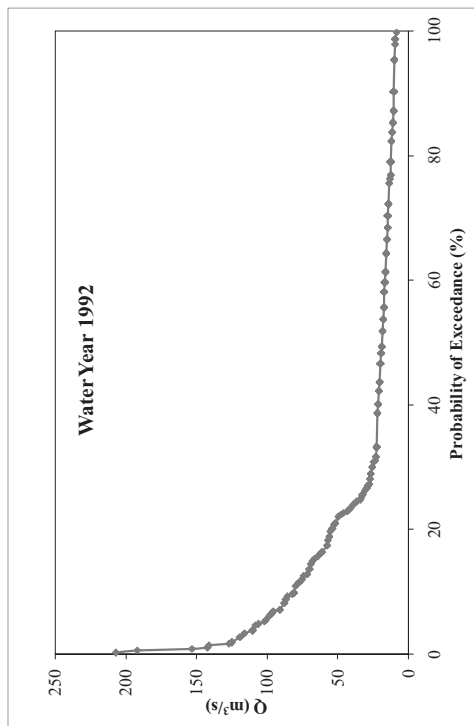


Figure B.58. EIE-2323 1992 water year yearly FDC

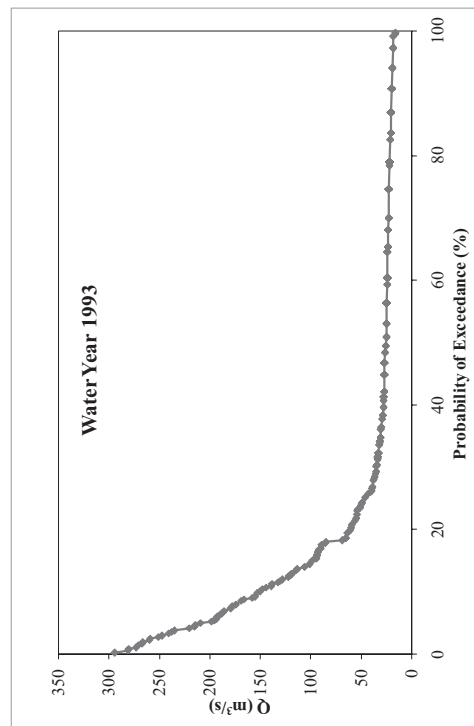


Figure B.59. EIE-2323 1993 water year yearly FDC

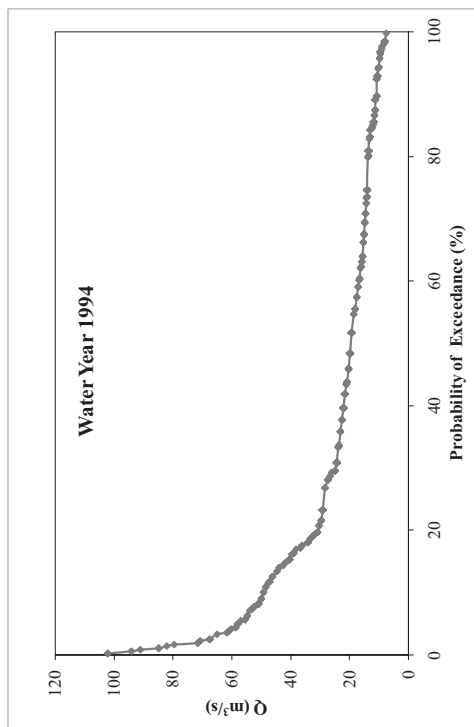


Figure B.60. EIE-2323 1994 water year yearly FDC

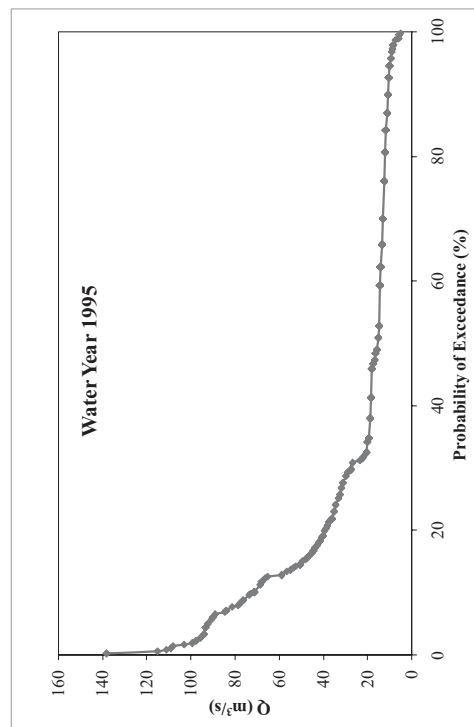


Figure B.61. EIE-2323 1995 water year yearly FDC

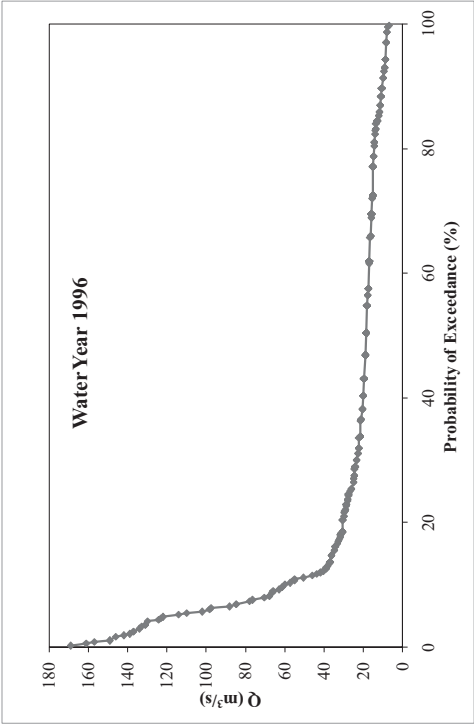


Figure B.62. EIE-2323 1996 water year yearly FDC

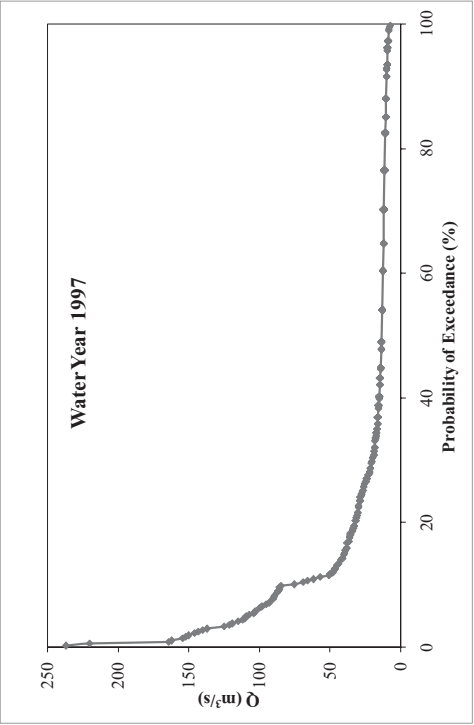


Figure B.63. EIE-2323 1997 water year yearly FDC

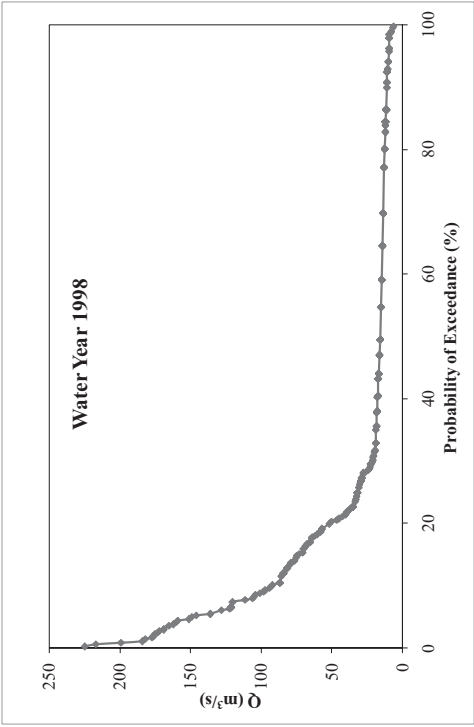


Figure B.64. EIE-2323 1998 water year yearly FDC

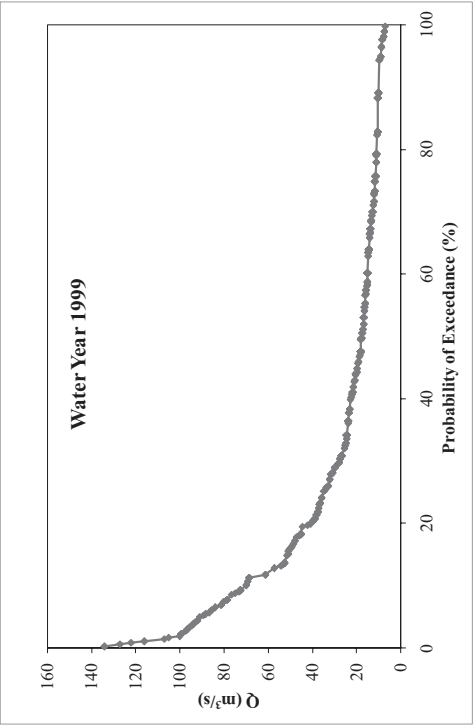


Figure B.65. EIE-2323 1999 water year yearly FDC

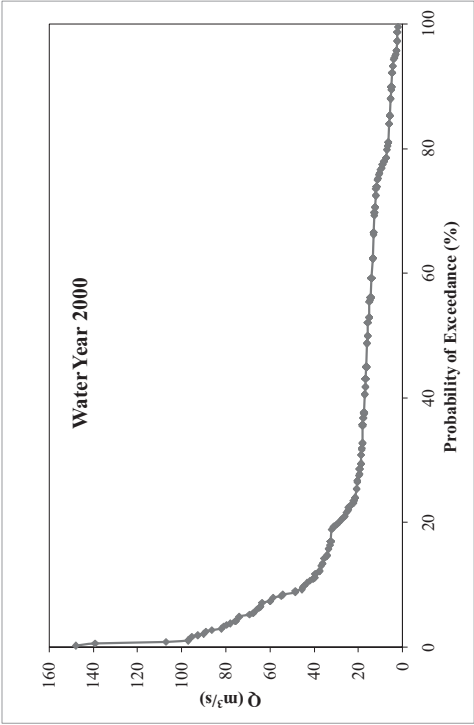


Figure B.66. EIE-2323 2000 water year yearly FDC

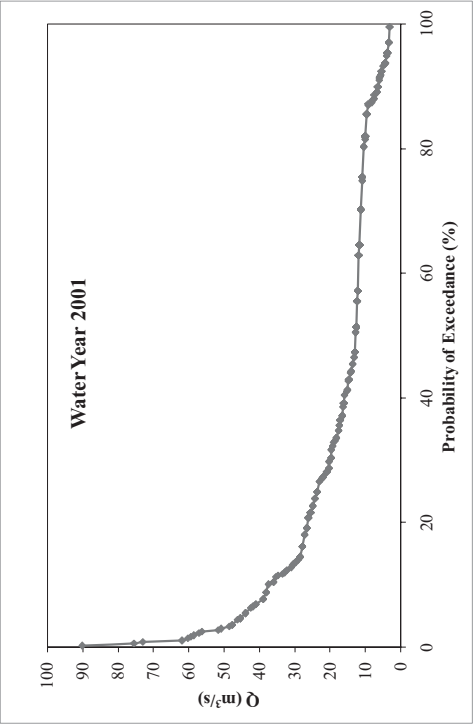


Figure B.67. EIE-2323 2001 water year yearly FDC

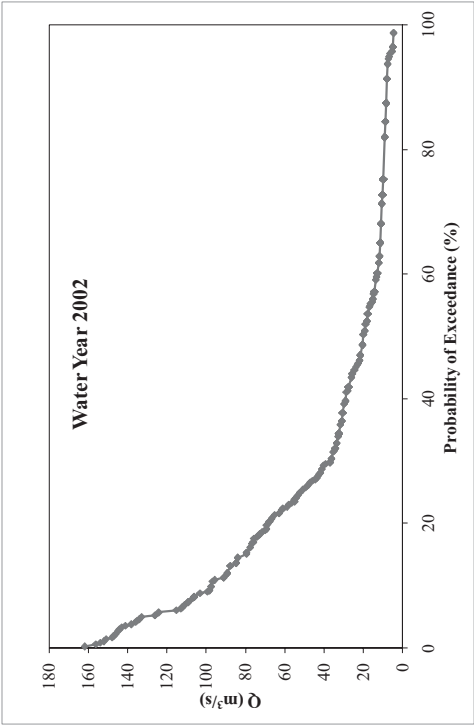


Figure B.68. EIE-2323 2002 water year yearly FDC

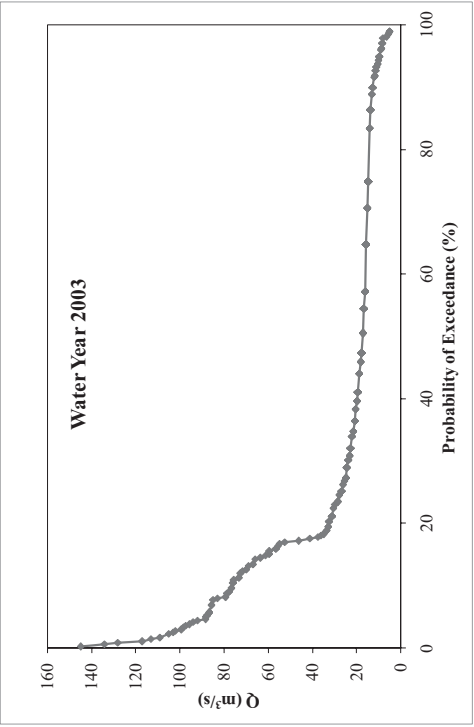


Figure B.69. EIE-2323 2003 water year yearly FDC

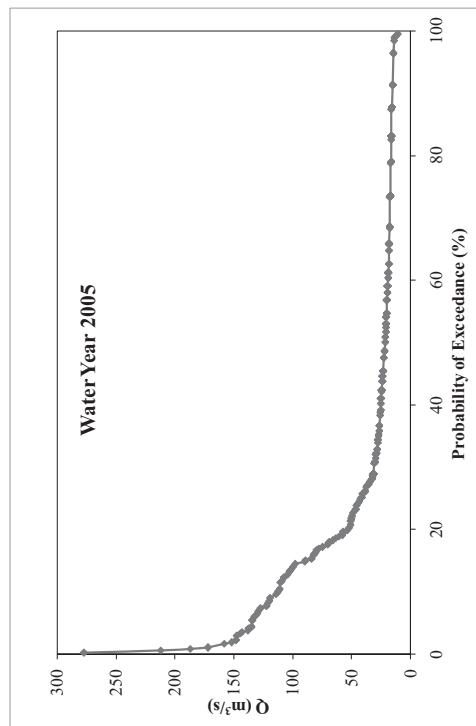


Figure B.70. EIE-2323 2004 water year yearly FDC

- **EIE-2325**

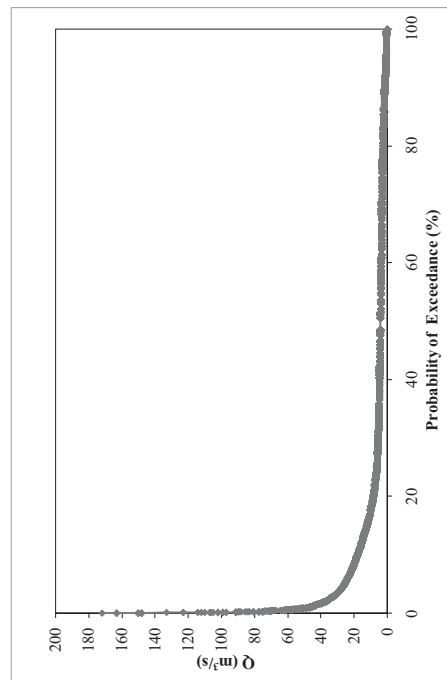


Figure B.71. Daily FDC of EIE-2325

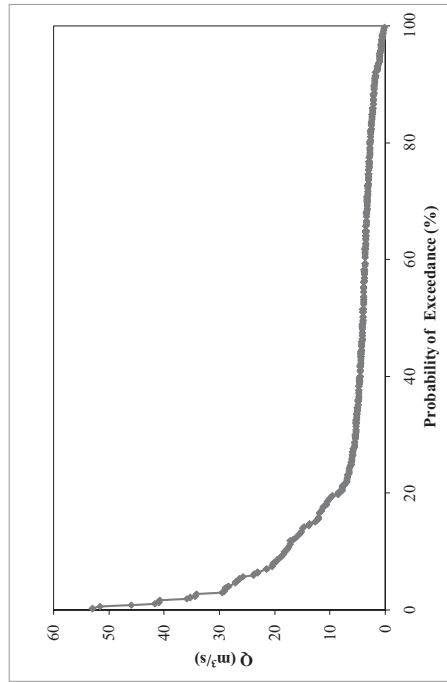


Figure B.72. Monthly FDC of EIE-232



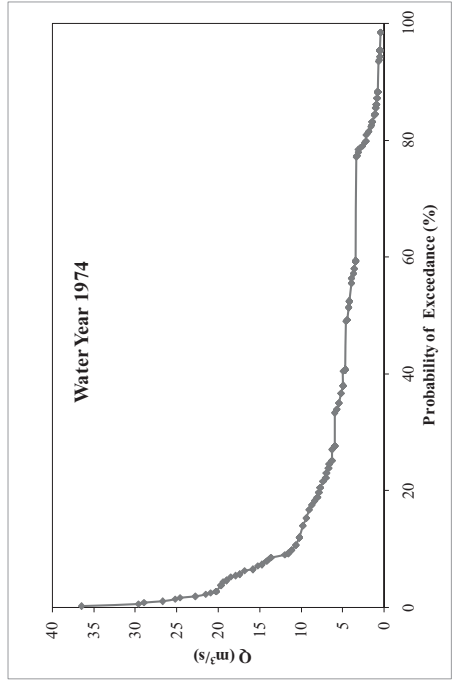


Figure B.73. EIE-2325 1974 water year yearly FDC

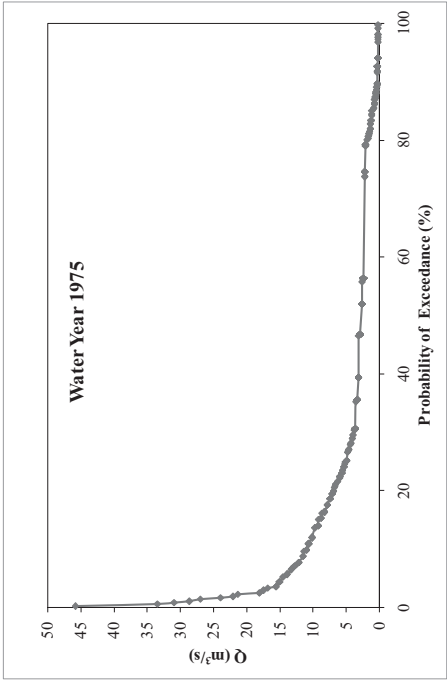


Figure B.74. EIE-2325 1975 water year yearly FDC

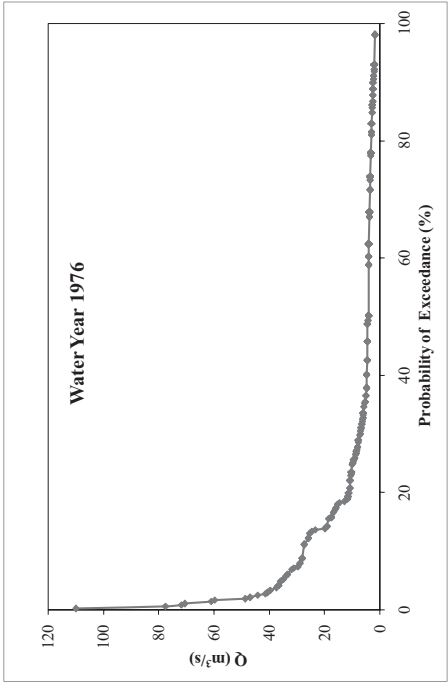


Figure B.75. EIE-2325 1976 water year yearly FDC

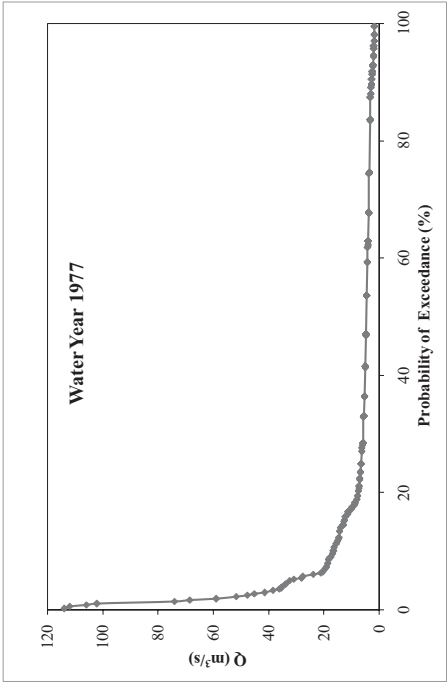


Figure B.76. EIE-2325 1977 water year yearly FDC

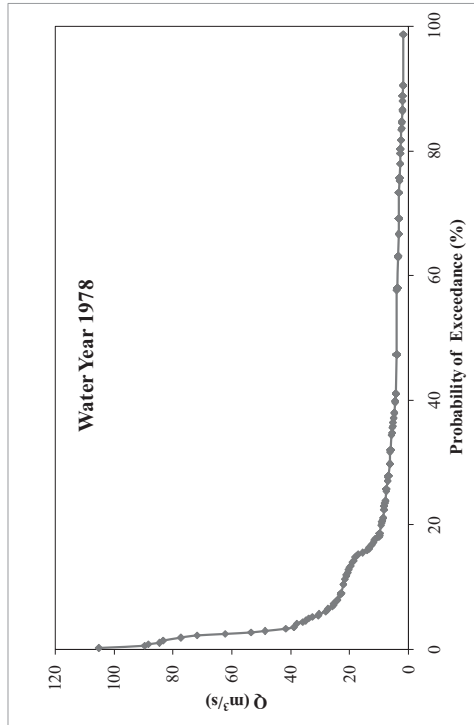


Figure B.77. EIE-2325 1978 water year yearly FDC

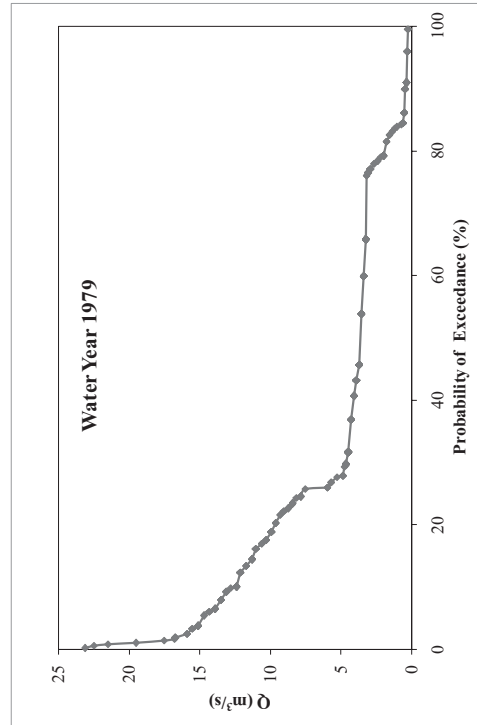


Figure B.78. EIE-2325 1979 water year yearly FDC

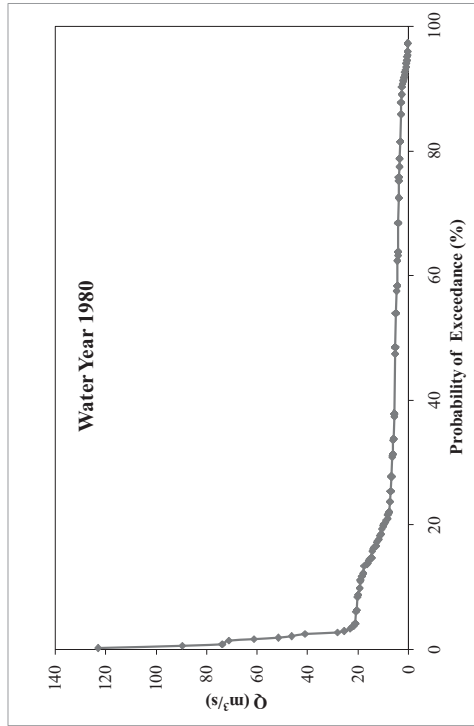


Figure B.79. EIE-2325 1980 water year yearly FDC

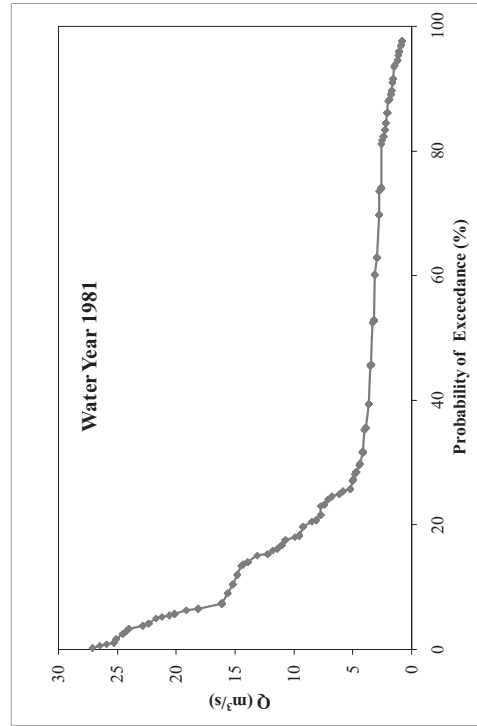


Figure B.80. EIE-2325 1981 water year yearly FDC

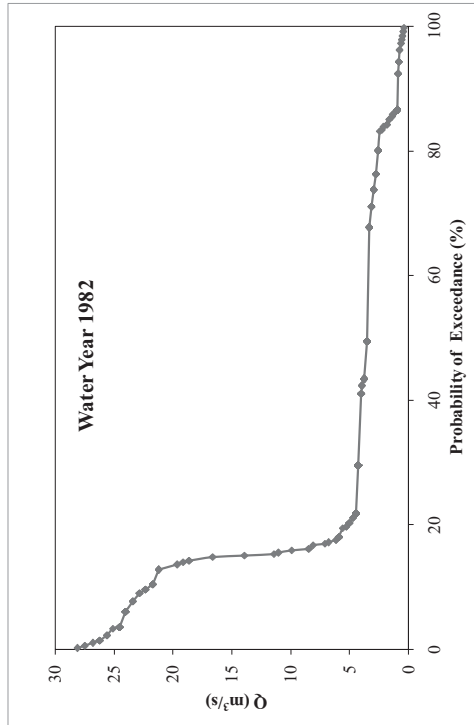


Figure B.81. EIE-2325 1982 water year yearly FDC

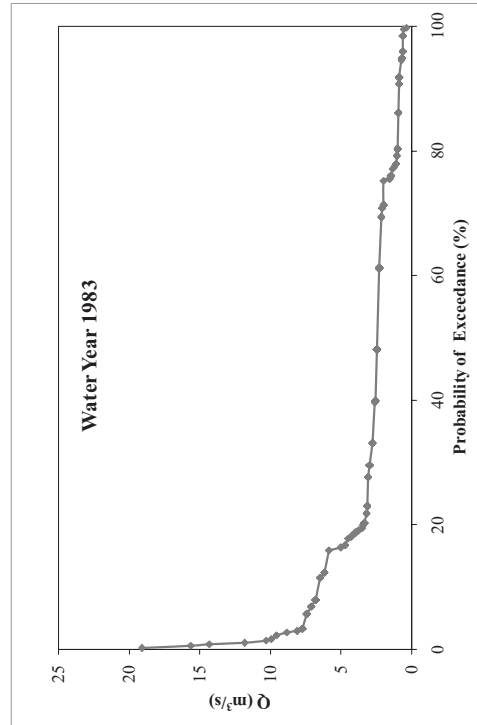


Figure B.82. EIE-2325 1983 water year yearly FDC

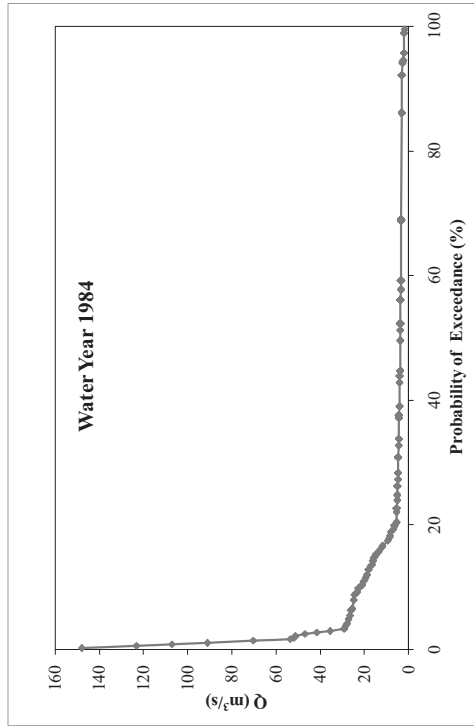


Figure B.83. EIE-2325 1984 water year yearly FDC

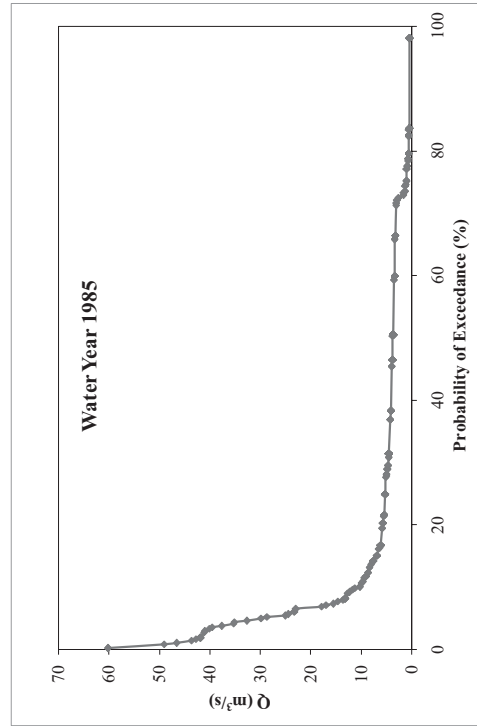


Figure B.84. EIE-2325 1985 water year yearly FDC

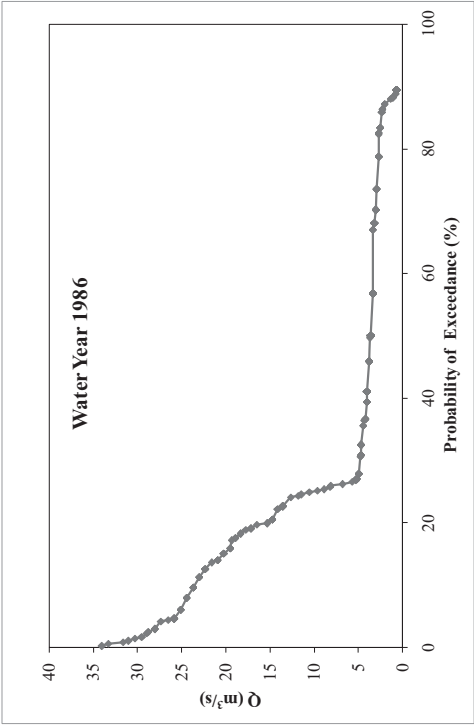


Figure B.85. EIE-2325 1986 water year yearly FDC

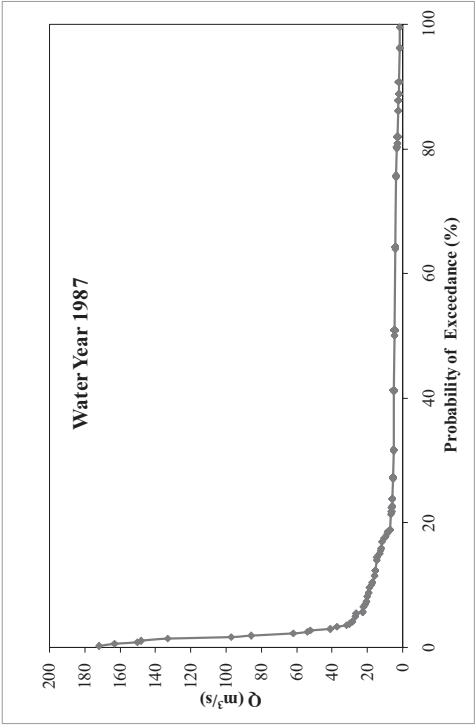


Figure B.86. EIE-2325 1987 water year yearly FDC

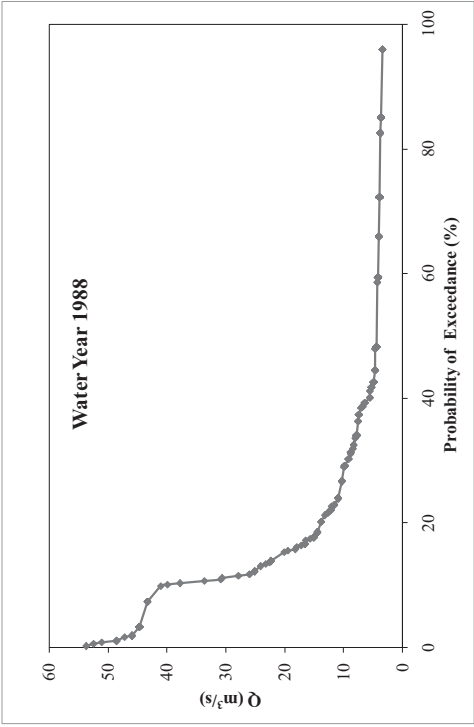


Figure B.87. EIE-2325 1988 water year yearly FDC

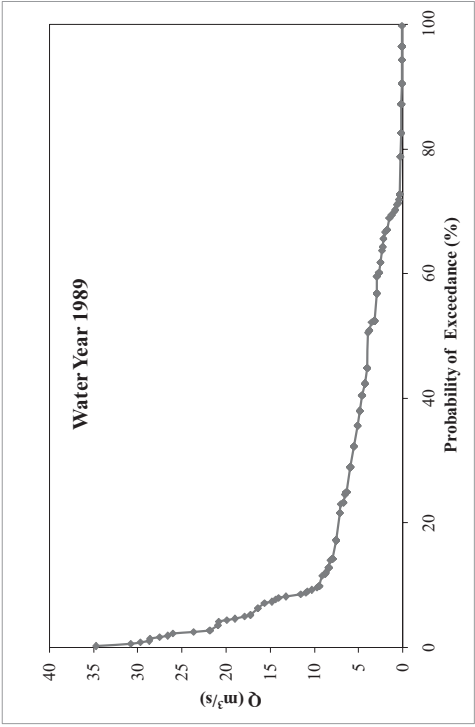


Figure B.88. EIE-2325 1989 water year yearly FDC

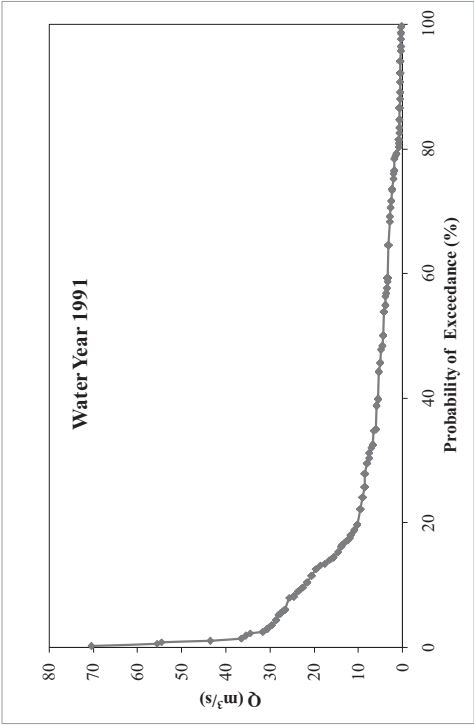


Figure B.89. EIE-2325 1991 water year yearly FDC

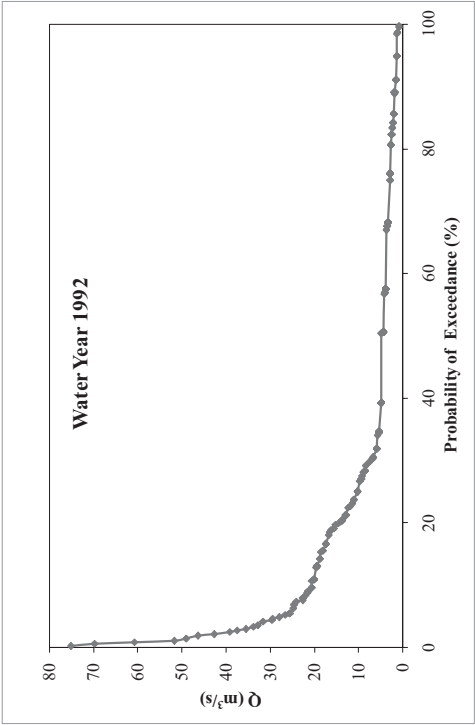


Figure B.90. EIE-2325 1992 water year yearly FDC

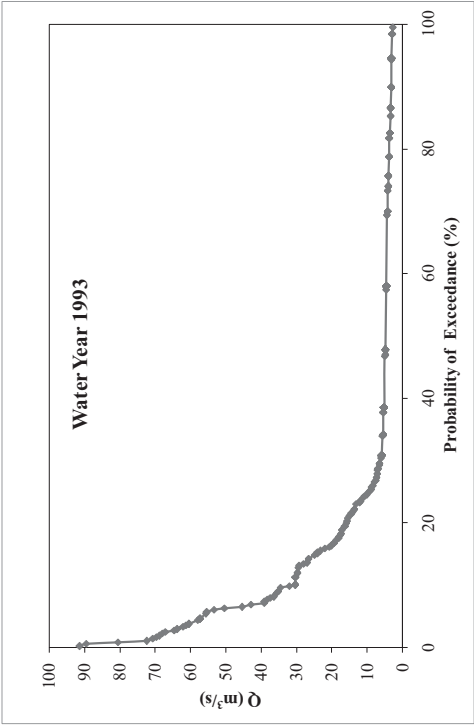


Figure B.91. EIE-2325 1993 water year yearly FDC

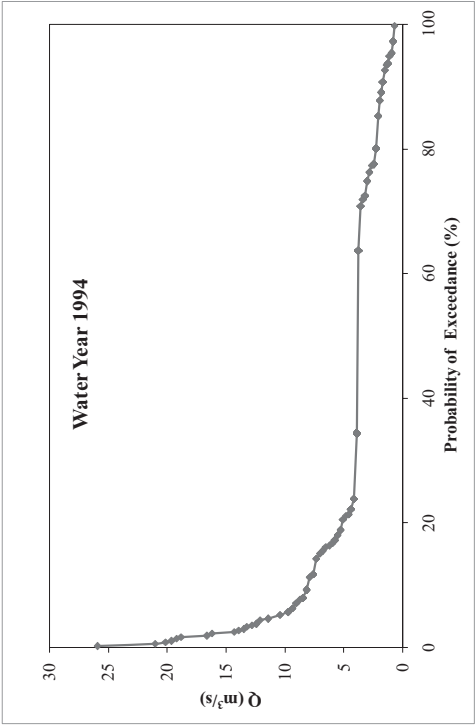


Figure B.92. EIE-2325 1994 water year yearly FDC

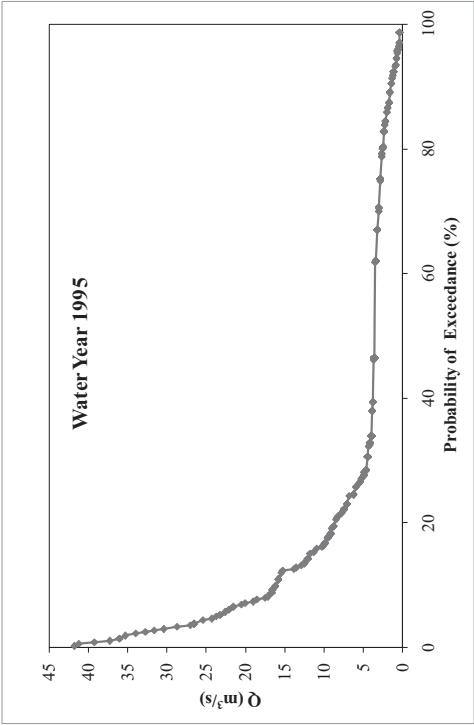


Figure B.93. EIE-2325 1995 water year yearly FDC

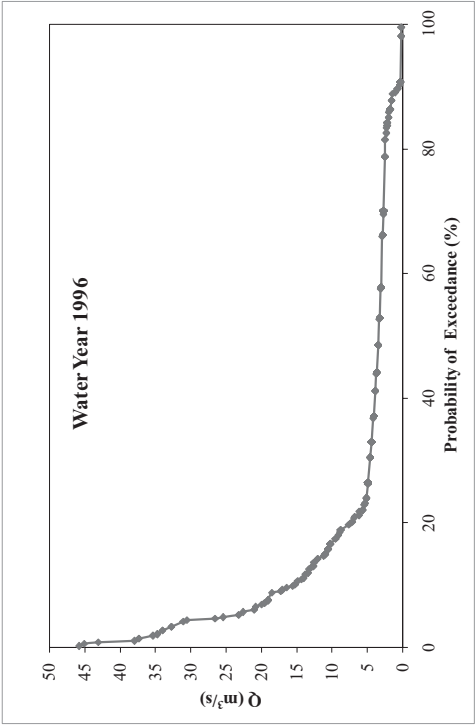


Figure B.94. EIE-2325 1996 water year yearly FDC

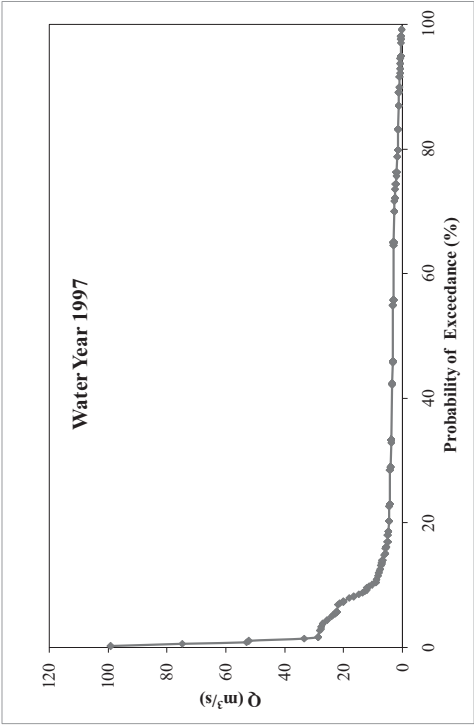


Figure B.95. EIE-2325 1997 water year yearly FDC

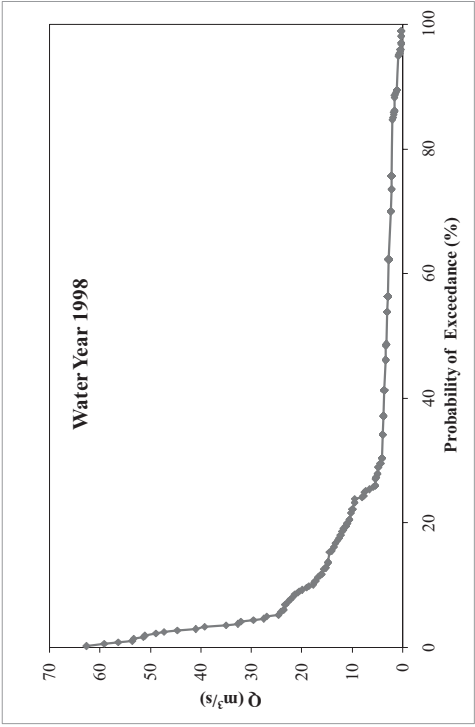


Figure B.96. EIE-2325 1998 water year yearly FDC

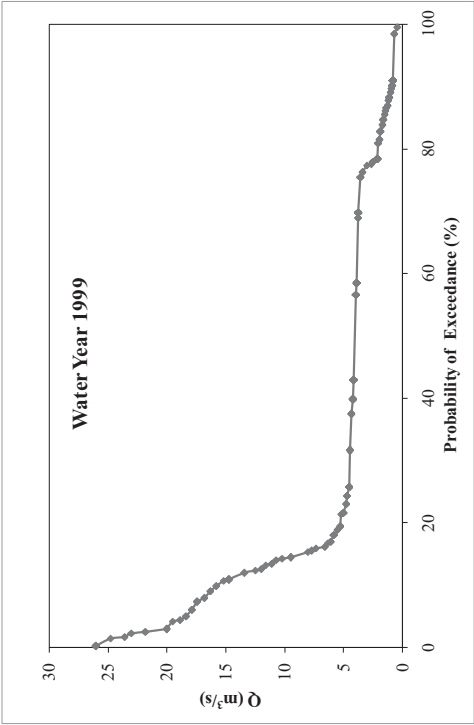


Figure B.97. EIE-2325 1999 water year yearly FDC

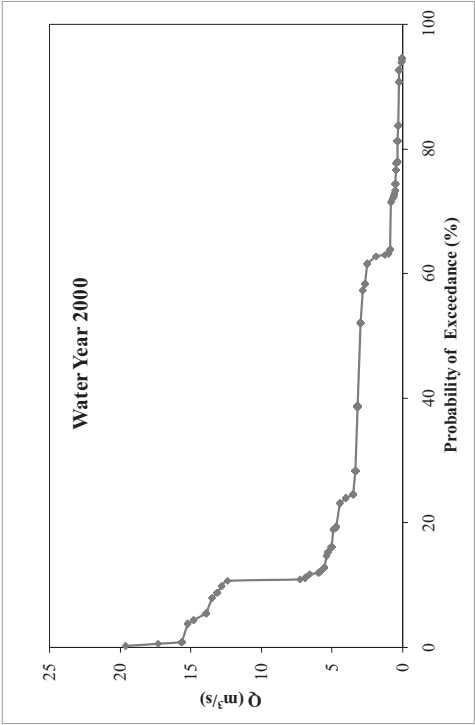


Figure B.98. EIE-2325 2000 water year yearly FDC

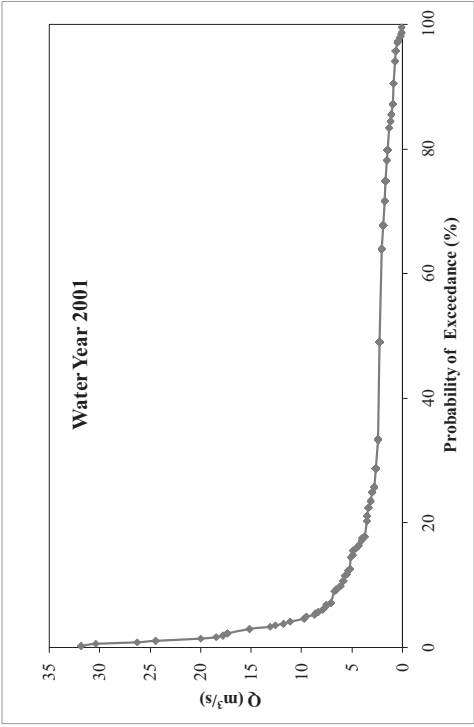


Figure B.99. EIE-2325 2001 water year yearly FDC

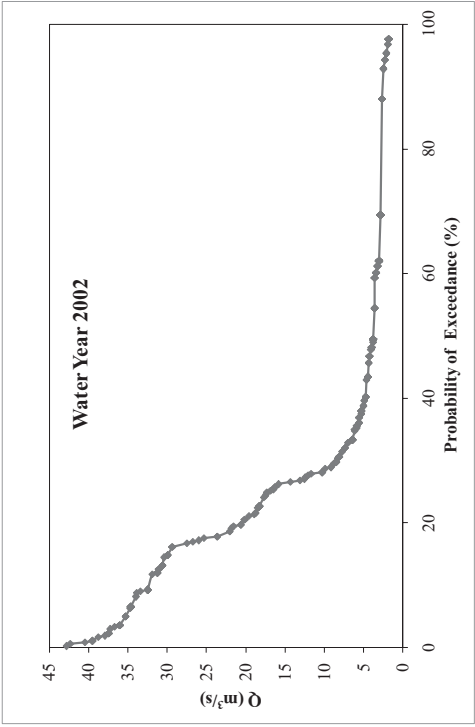


Figure B.100. EIE-2325 2002 water year yearly FDC

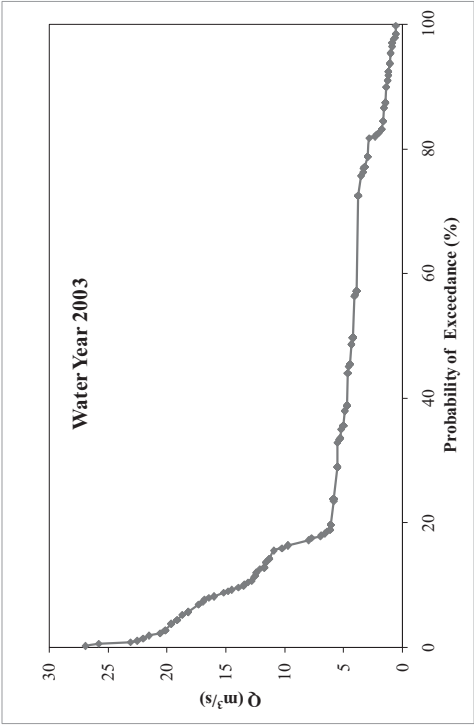


Figure B.101. EIE-2325 2003 water year yearly FDC

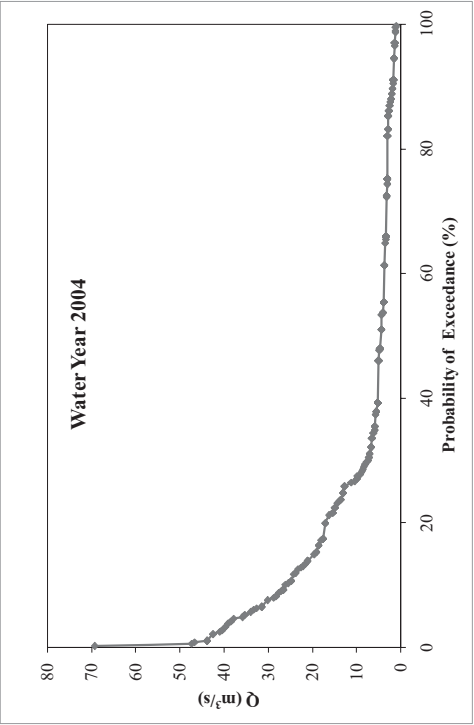


Figure B.102. EIE-2325 2004 water year yearly FDC

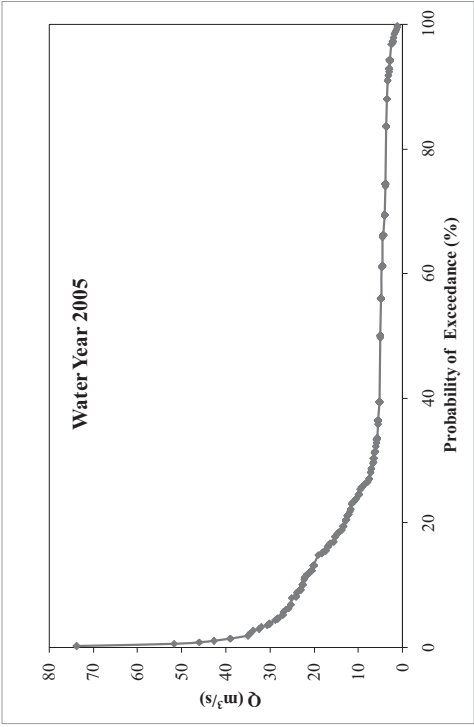


Figure B.103. EIE-2325 2005 water year yearly FDC



- **DSI-2313**

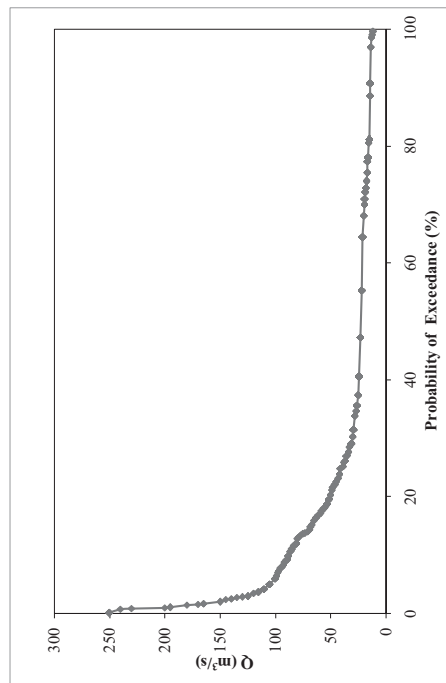


Figure B.104. Daily FDC of DSI-2313

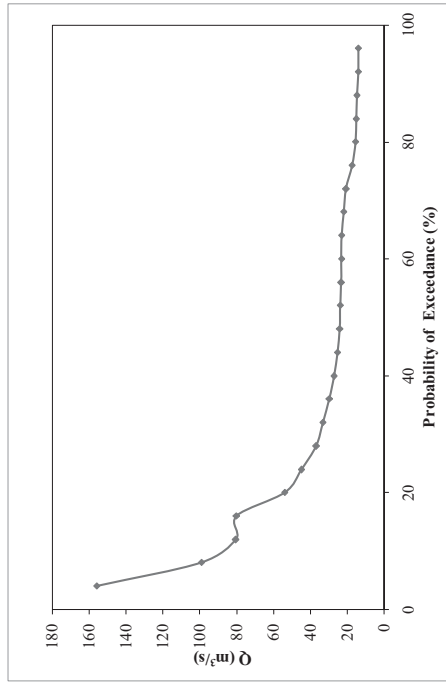


Figure B.105. Monthly FDC of DSI-2313

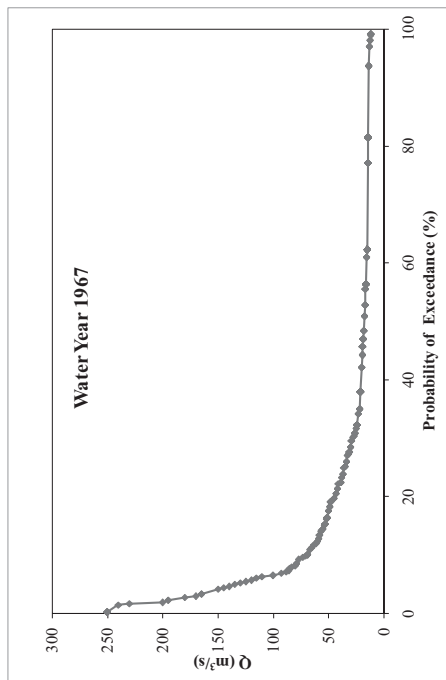


Figure B.106. DSI-2313 1967 water year yearly FDC

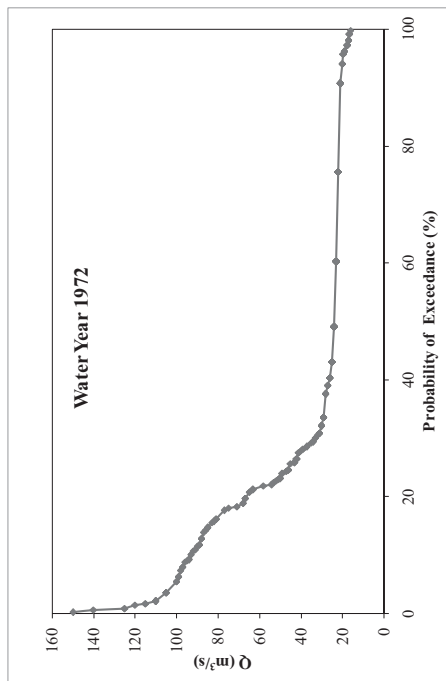


Figure B.107. DSI-2313 1972 water year yearly FDC

- **DSI-2321**

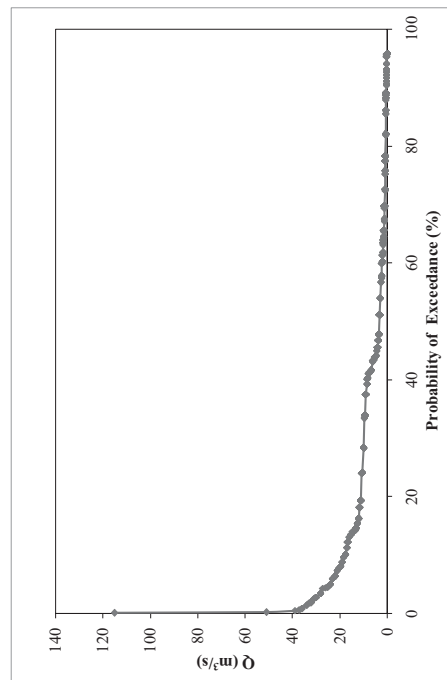


Figure B.108. Daily FDC of DSI-2321

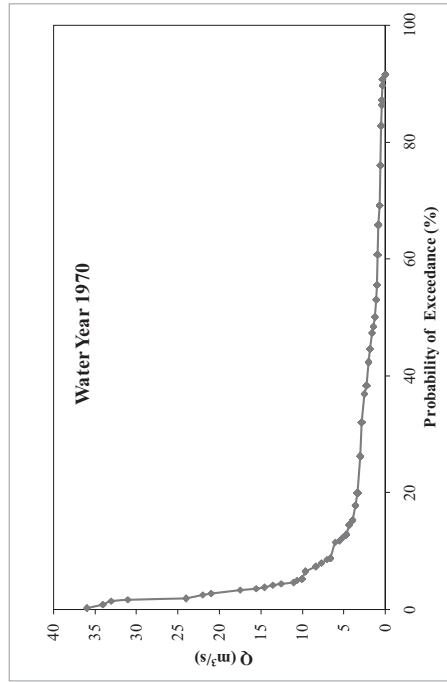


Figure B.109. DSI-2321 1970 water year yearly FDC

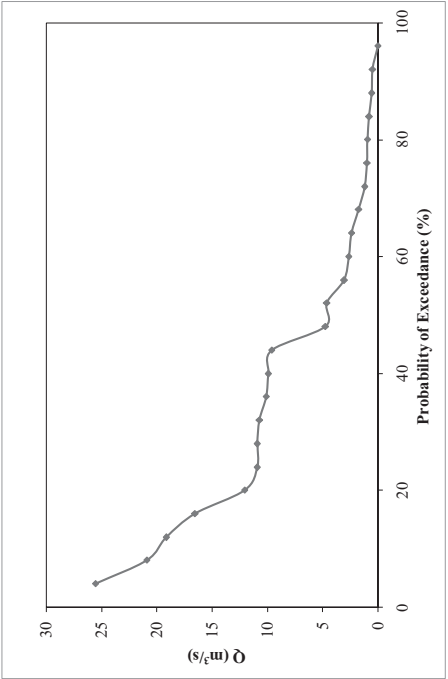


Figure B.110. Monthly FDC of DSI-2321

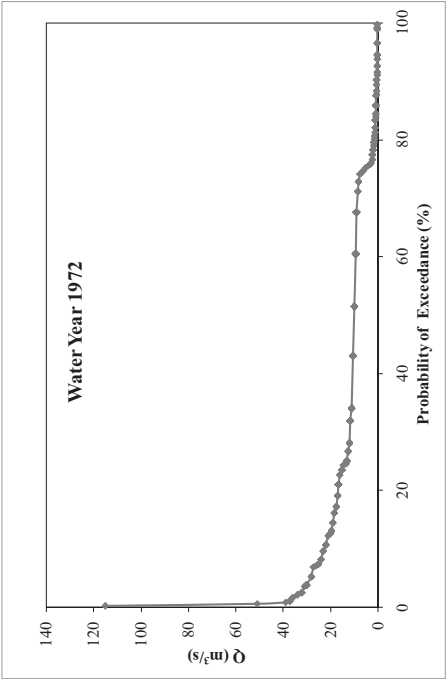


Figure B.111. DSI-2321 1972 water year yearly FDC

- **DSI-2322**

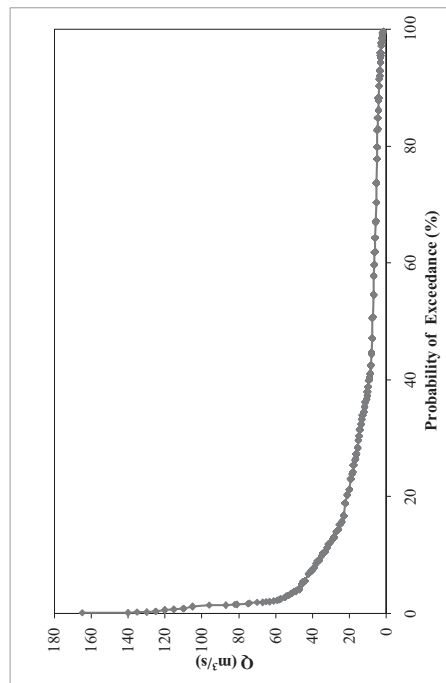


Figure B.112. Daily FDC of DSI-2322

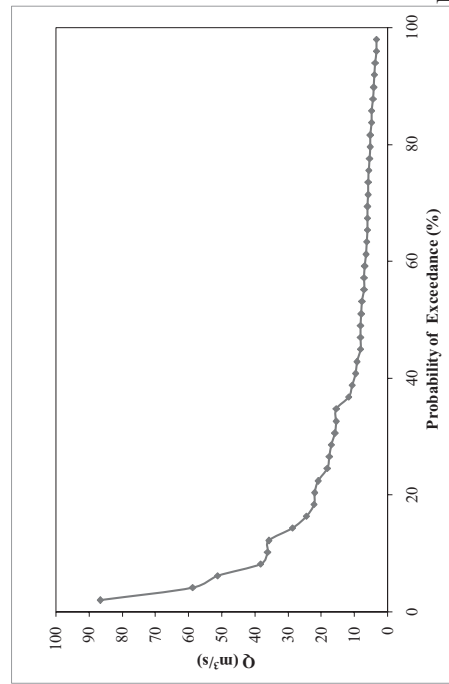


Figure  
B.113. Monthly FDC of DSI-2322

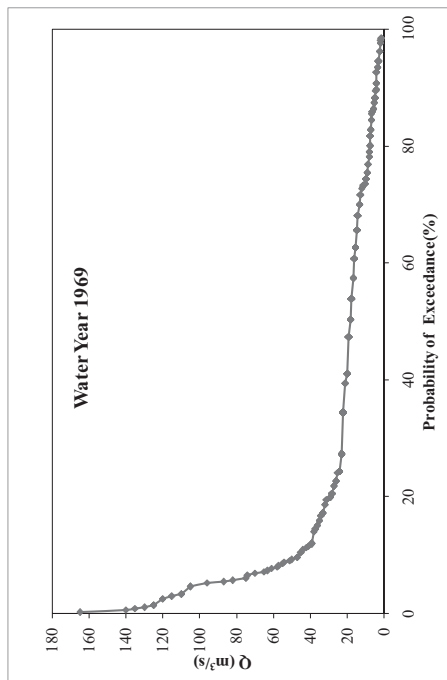


Figure B.114. DSI-2322 1969 water year yearly FDC

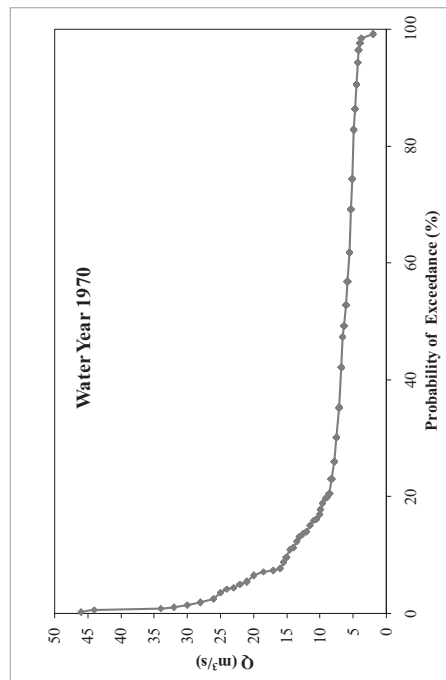


Figure B.115. DSI-2322 1970 water year yearly FDC

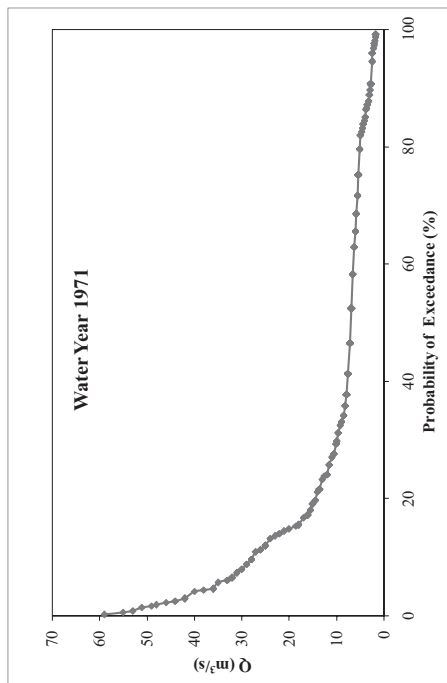


Figure B.116. DSI-2322 1971 water year yearly FDC

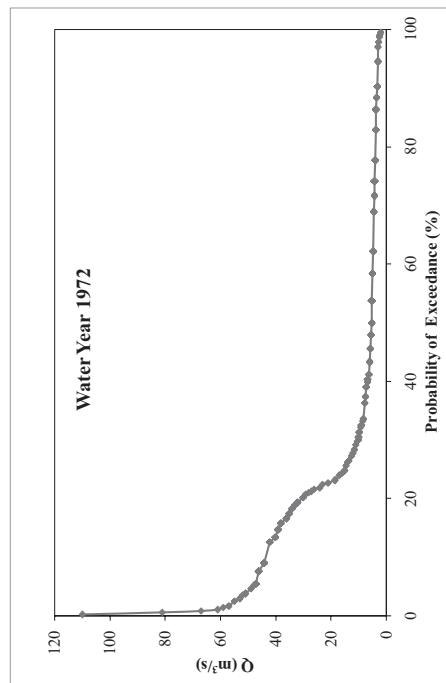


Figure B.117. DSI-2322 1972 water year yearly FDC

- **DSI-2323**

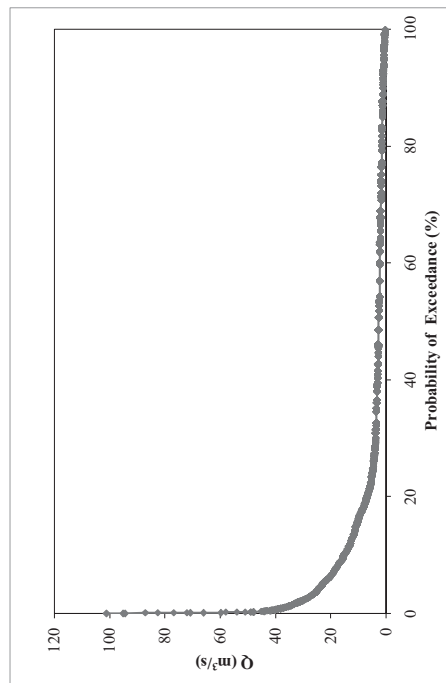


Figure B.118. Daily FDC of DSI-2323

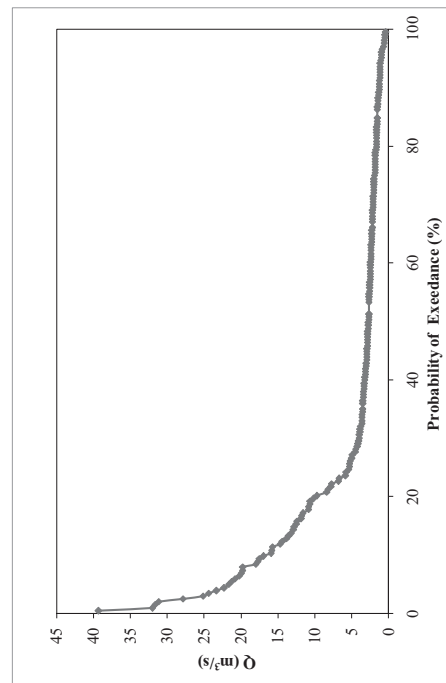


Figure B.119. Monthly FDC of DSI-2323

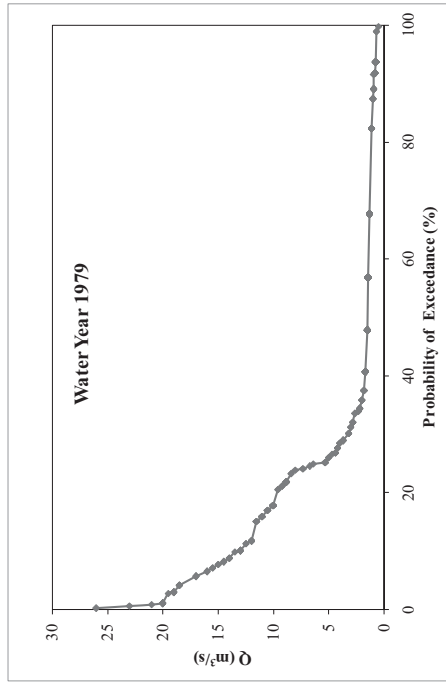


Figure B.120. DSI-2323 1979 water year yearly FDC

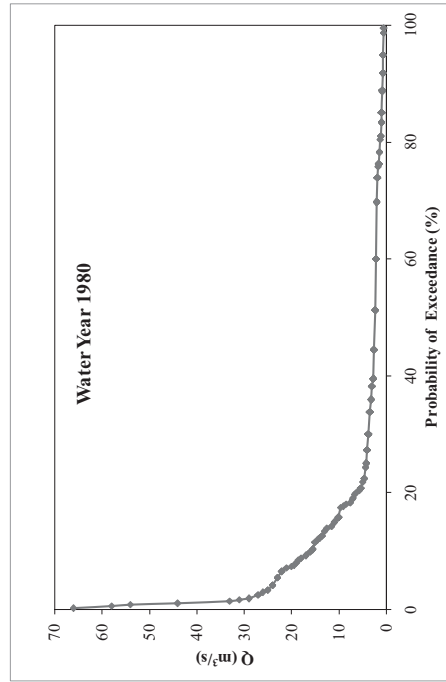


Figure B.121. DSI-2323 1980 water year yearly FDC

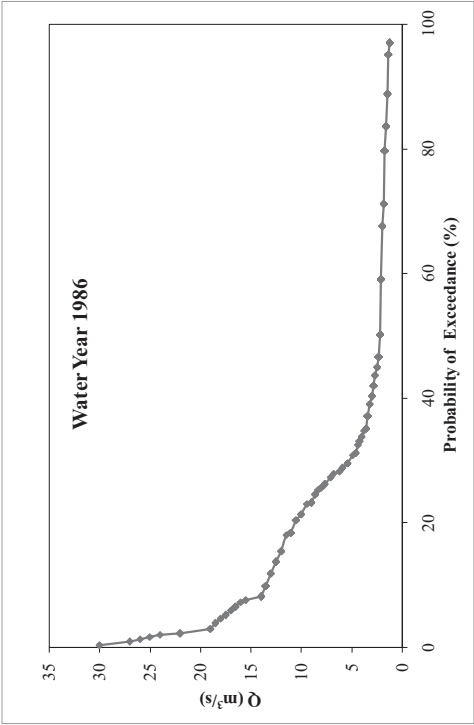


Figure B.122. DSI-2323 1986 water year yearly FDC

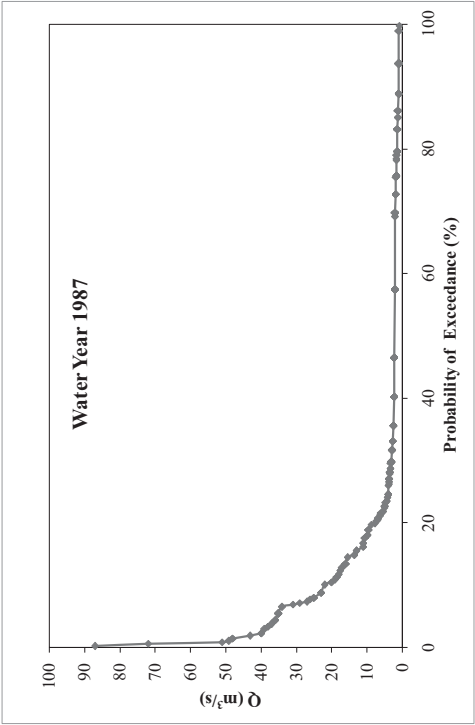


Figure B.123. DSI-2323 1987 water year yearly FDC

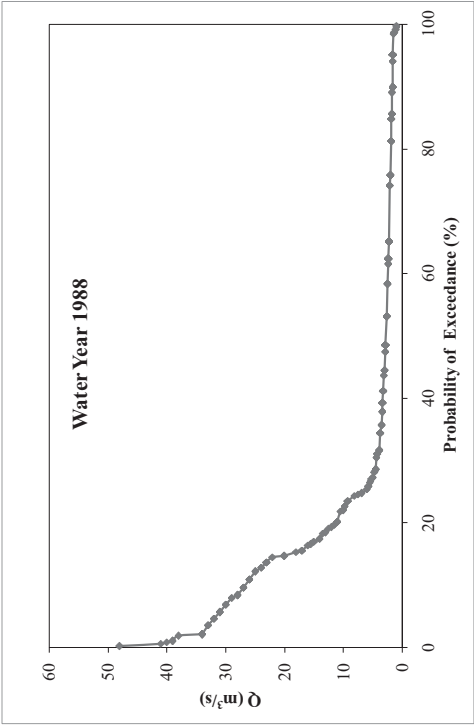


Figure B.124. DSI-2323 1988 water year yearly FDC

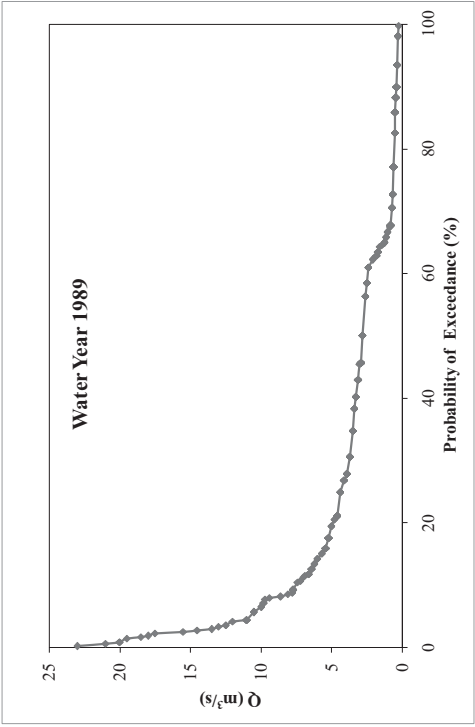


Figure B.125. DSI-2323 1989 water year yearly FDC



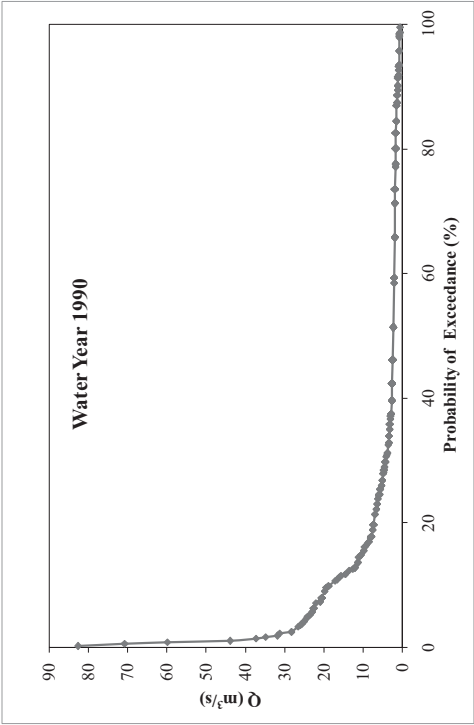


Figure B.126. DSI-2323 1990 water year yearly FDC

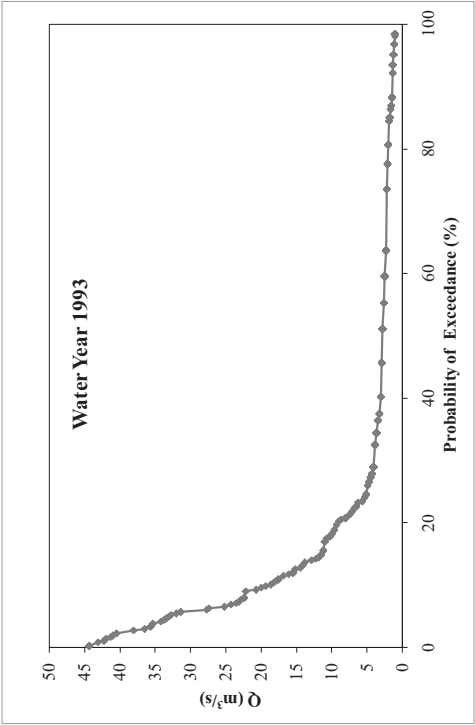


Figure B.127. DSI-2323 1993 water year yearly FDC

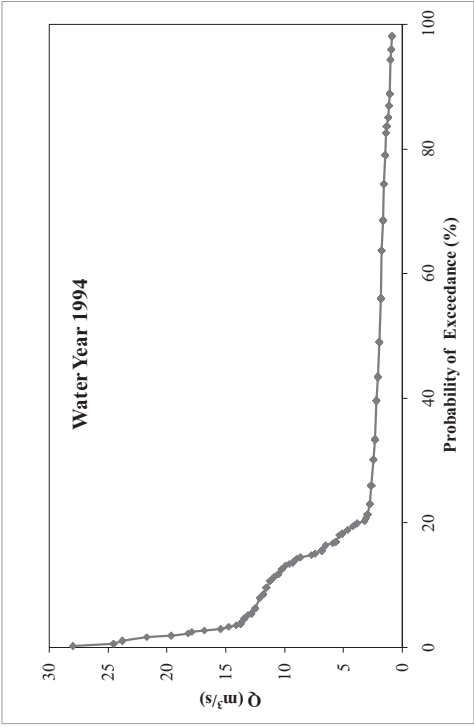


Figure B.128. DSI-2323 1994 water year yearly FDC

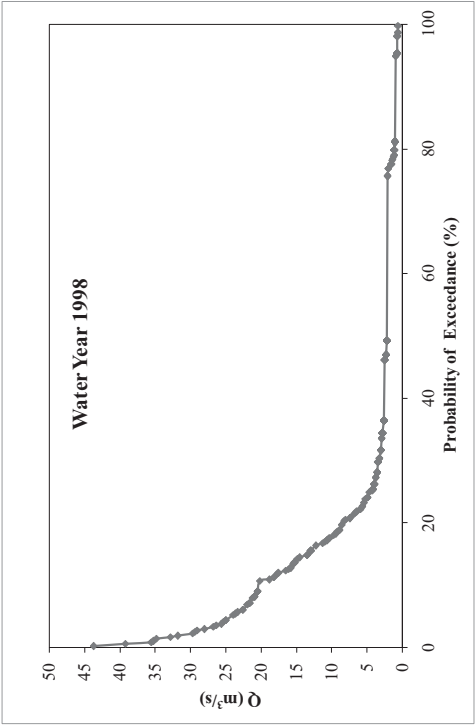


Figure B.129. DSI-2323 1998 water year yearly FDC

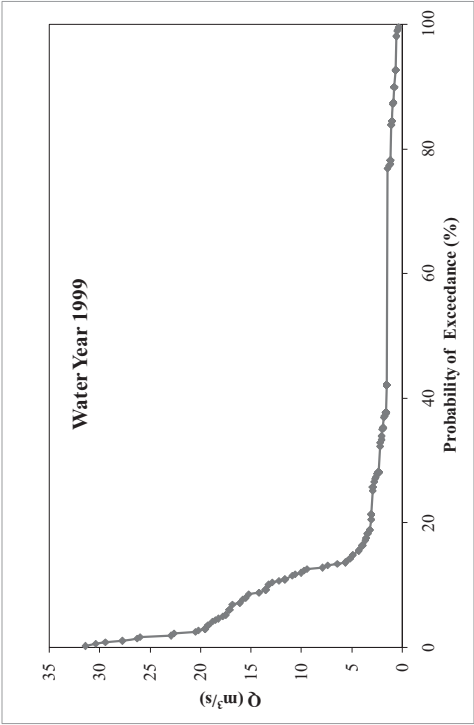


Figure B.130. DSI-2323 1999 water year yearly FDC

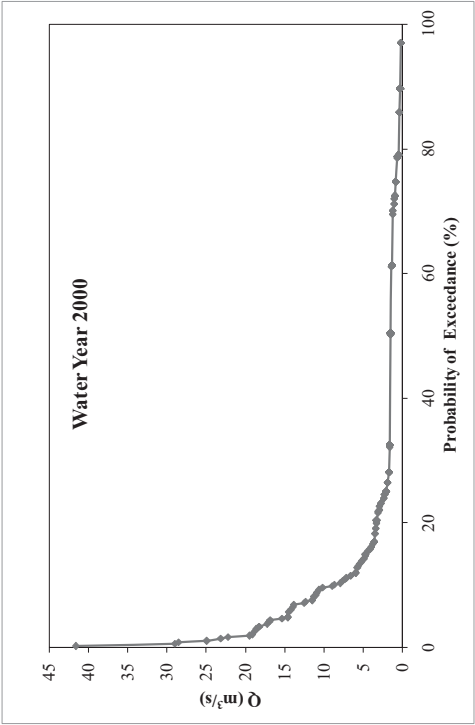


Figure B.131. DSI-2323 2000 water year yearly FDC

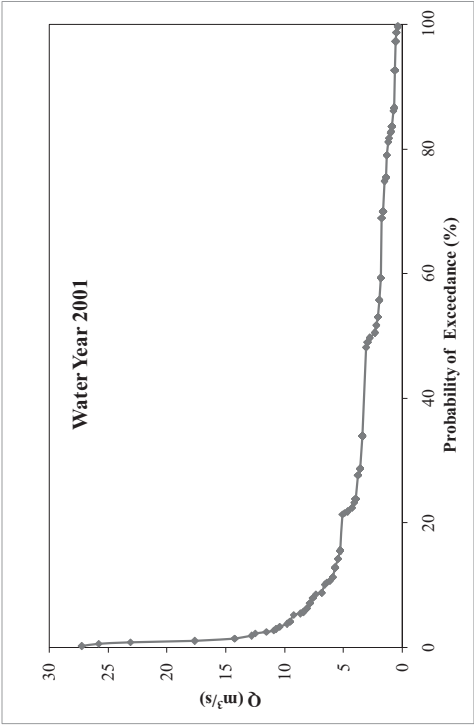


Figure B.132. DSI-2323 2001 water year yearly FDC

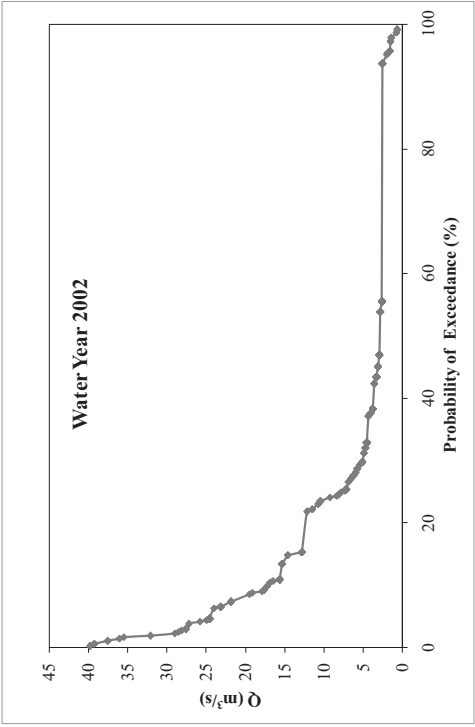


Figure B.133. DSI-2323 2002 water year yearly FDC

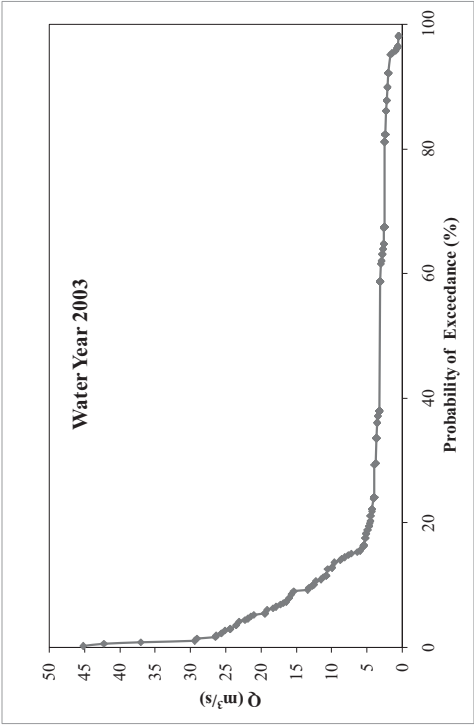


Figure B.134. DSI-2323 2003 water year yearly FDC

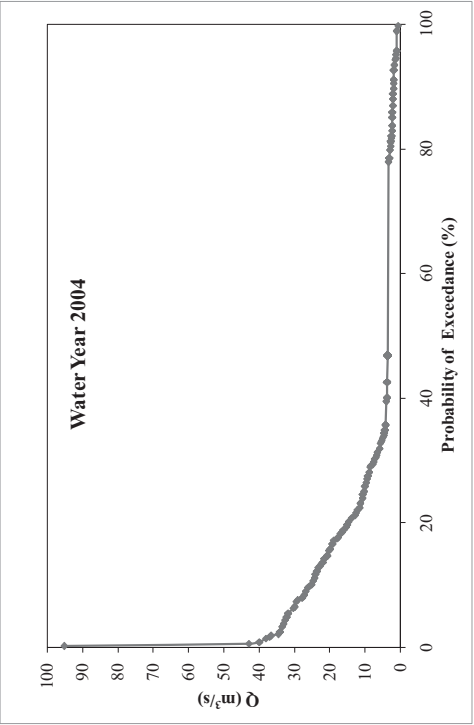


Figure B.135. DSI-2323 2004 water year yearly FDC

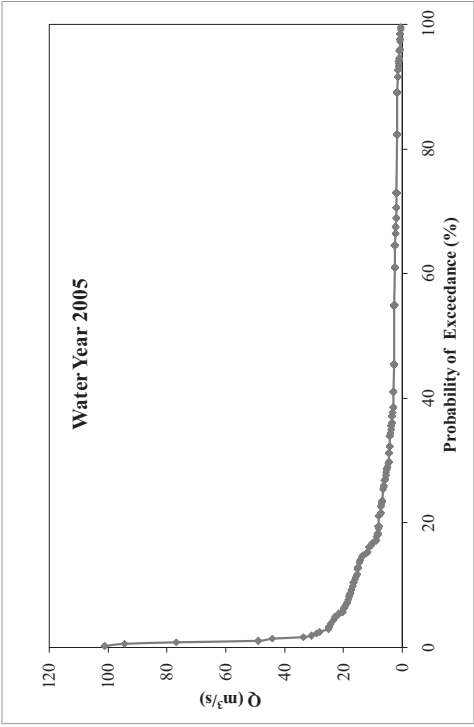


Figure B.136. DSI-2323 2005 water year yearly FDC

- **DSI-2324**

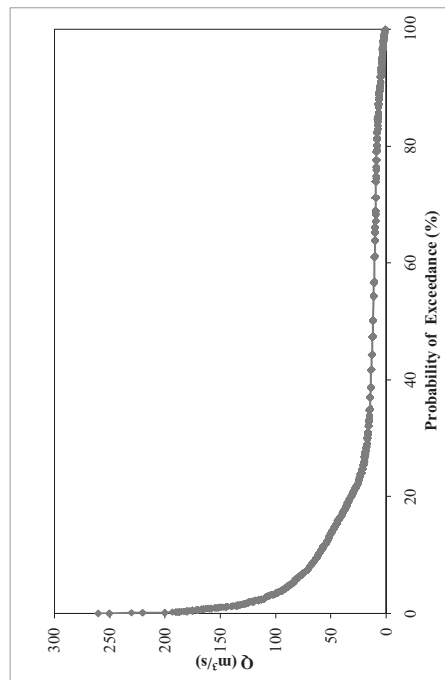


Figure B.137. Daily FDC of DSI-2324

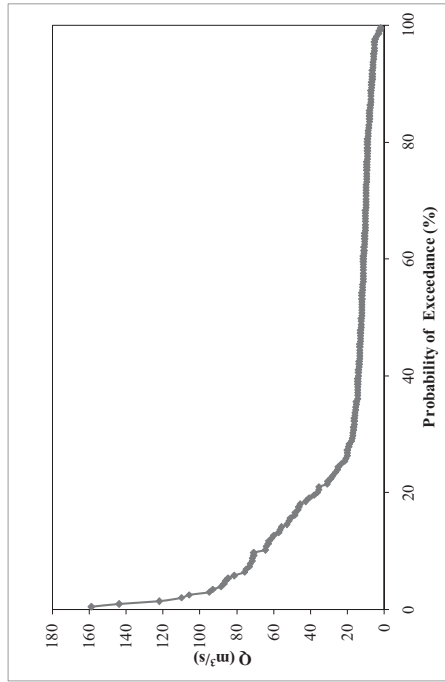


Figure B.138. Monthly FDC of DSI-2324

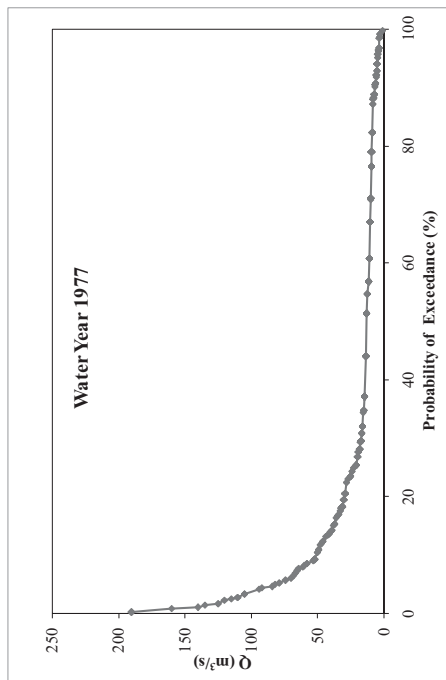


Figure B.139. DSI-2324 1977 water year yearly FDC

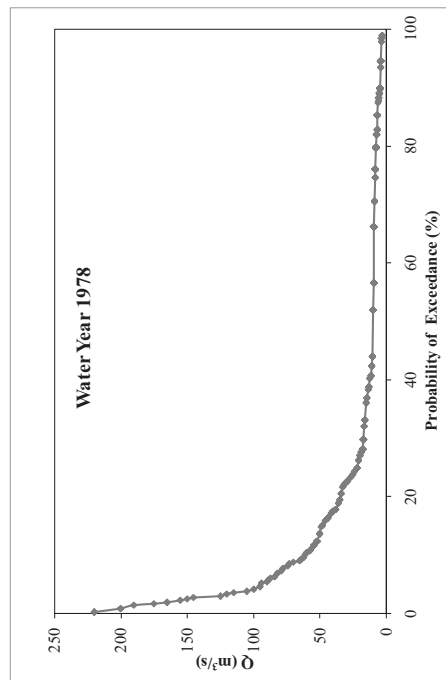


Figure B.140. DSI-2324 1978 water year yearly FDC

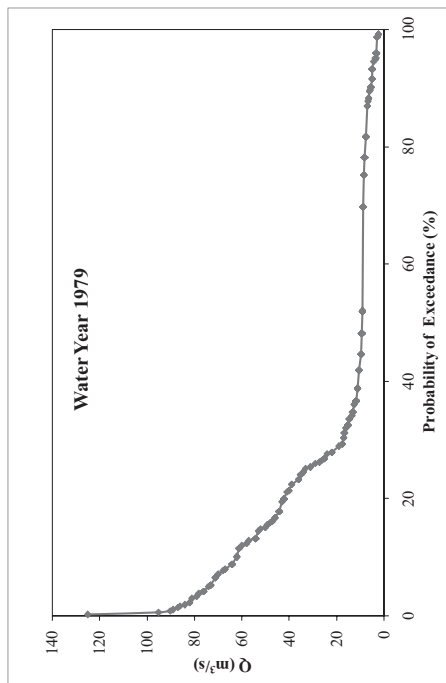


Figure B.141. DSI-2324 1979 water year yearly FDC

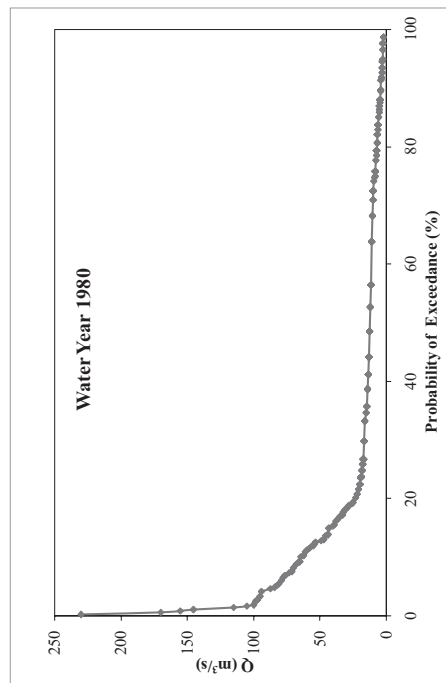


Figure B.142. DSI-2324 1980 water year yearly FDC

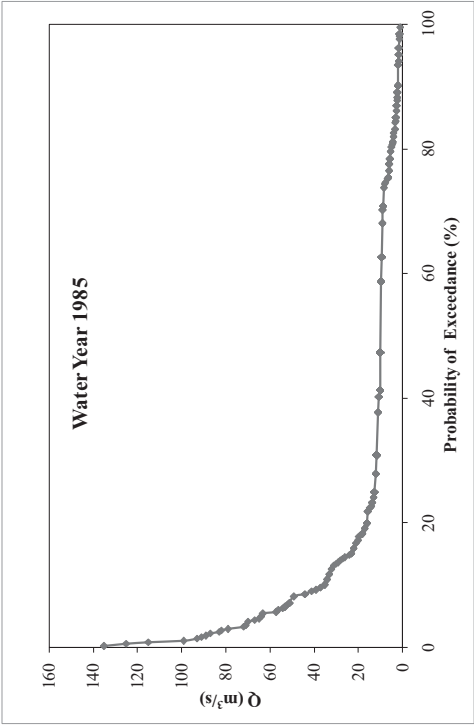


Figure B.143. DSI-2324 1985 water year yearly FDC

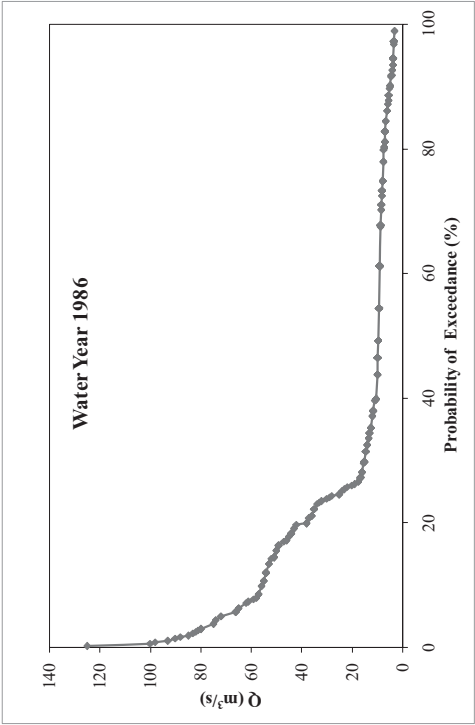


Figure B.144. DSI-2324 1986 water year yearly FDC

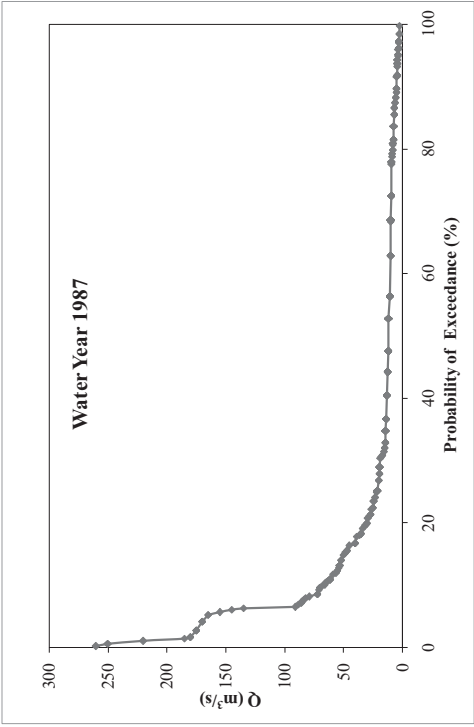


Figure B.145. DSI-2324 1987 water year yearly FDC

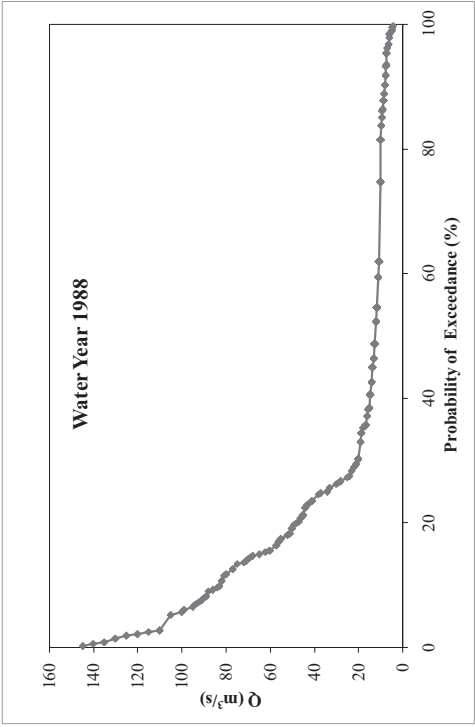


Figure B.146. DSI-2324 1988 water year yearly FDC

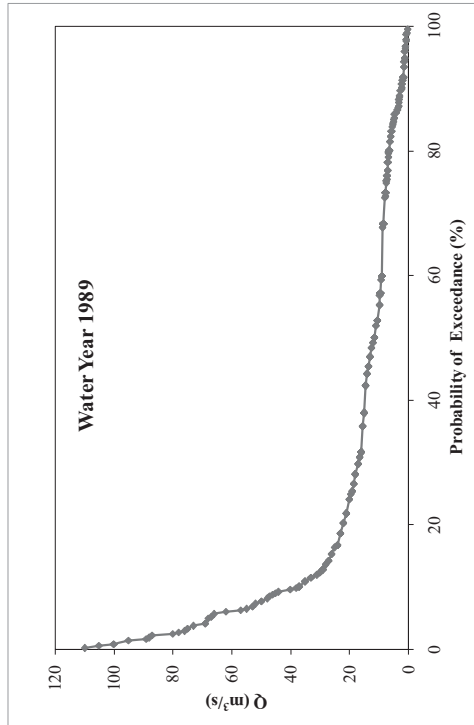


Figure B.147. DSI-2324 1989 water year yearly FDC

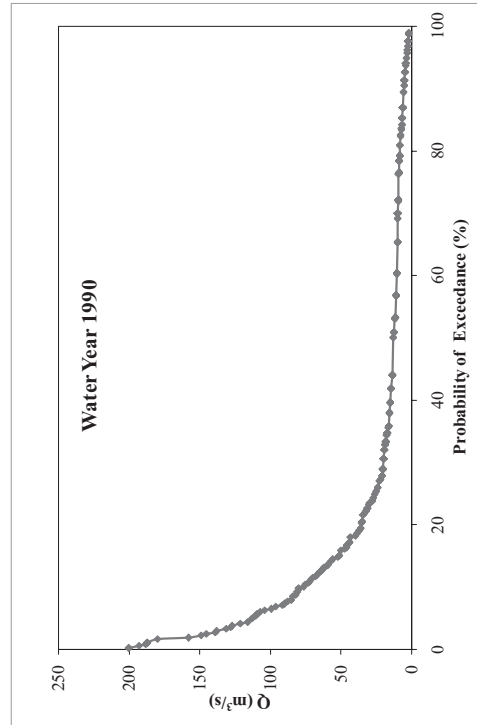


Figure B.148. DSI-2324 1990 water year yearly FDC

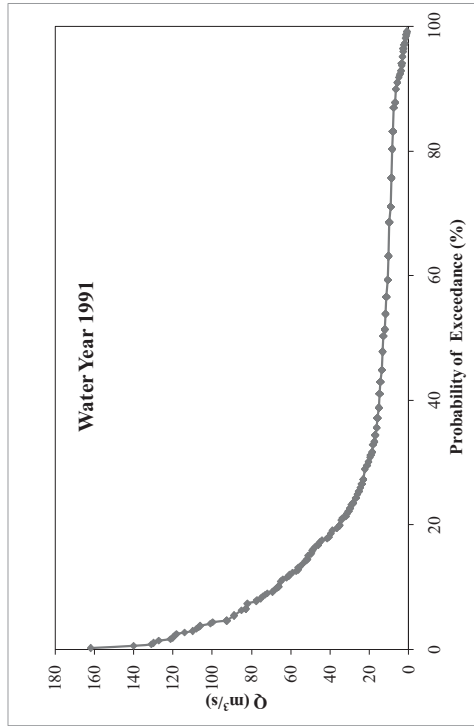


Figure B.149. DSI-2324 1991 water year yearly FDC

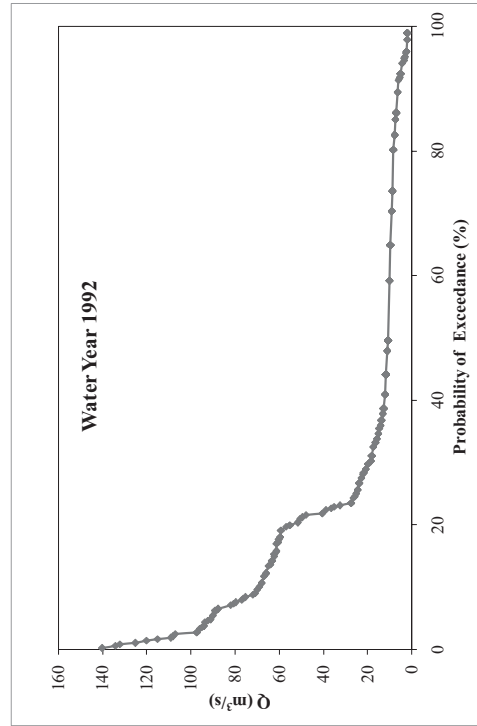


Figure B.150. DSI-2324 1992 water year yearly FDC

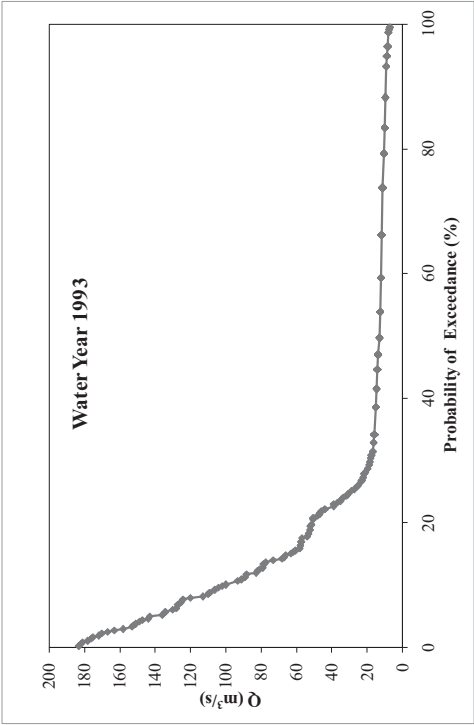


Figure B.151. DSI-2324 1993 water year yearly FDC

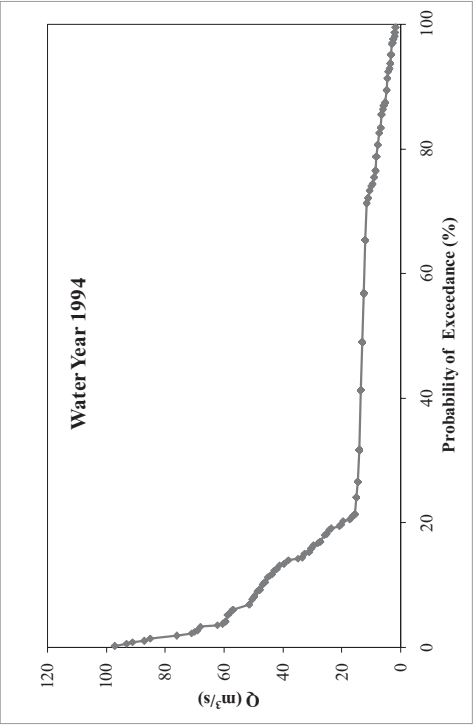


Figure B.152. DSI-2324 1994 water year yearly FDC

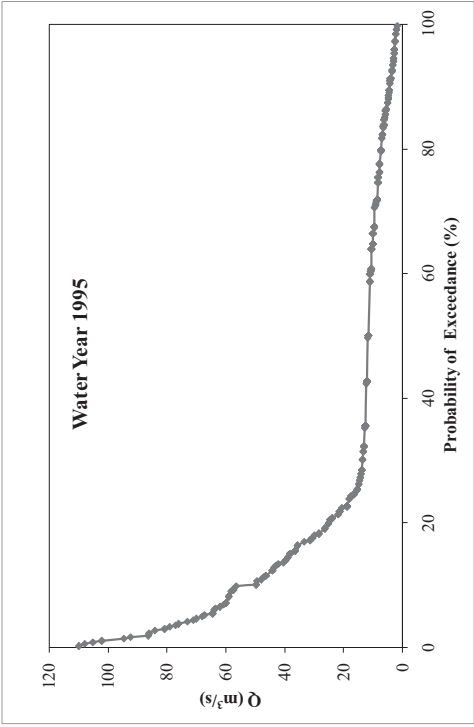


Figure B.153. DSI-2324 1995 water year yearly FDC

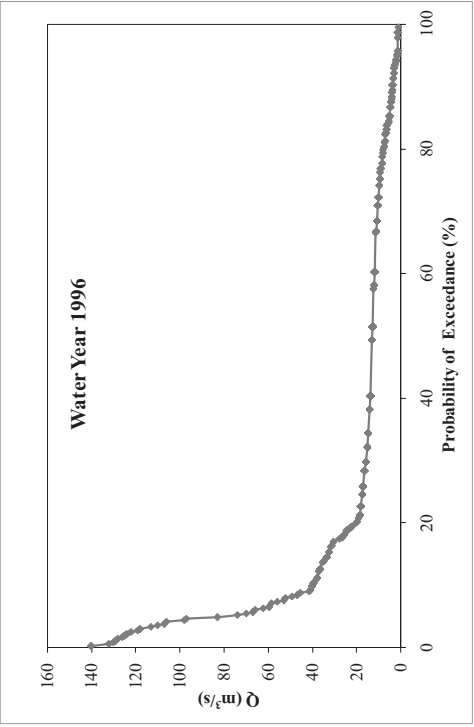


Figure B.154. DSI-2324 1996 water year yearly FDC



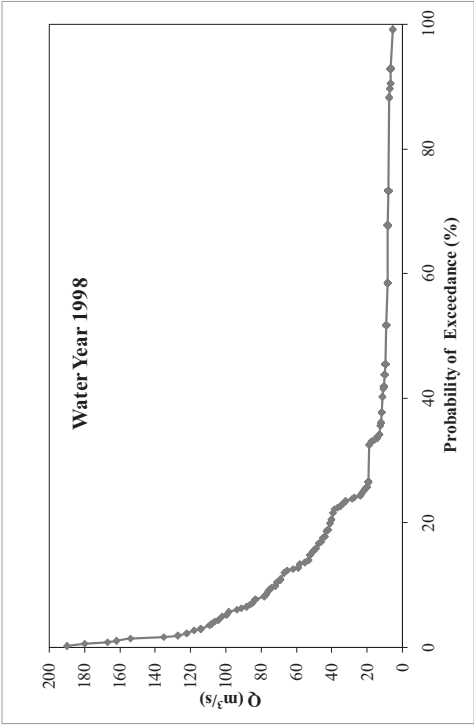


Figure B.155. DSI-2324 1998 water year yearly FDC

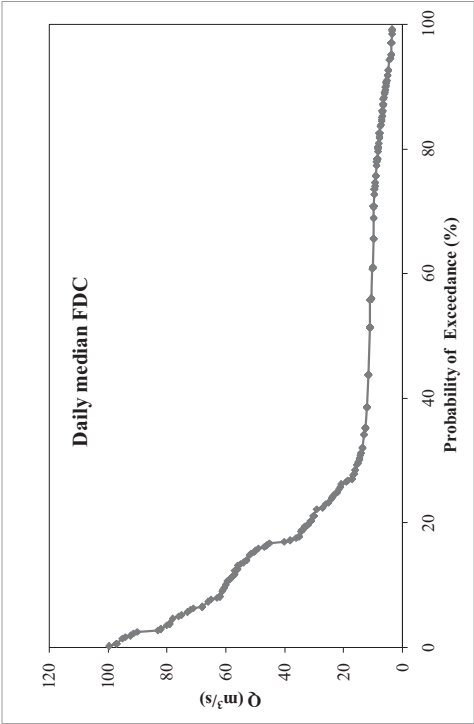


Figure B.156. Daily median FDC of DSI-2324

- **DSI-2333**

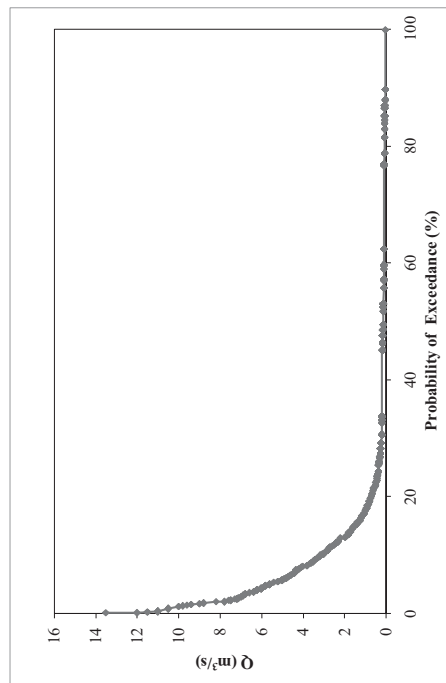


Figure B.157. Daily FDC of DSI-2333

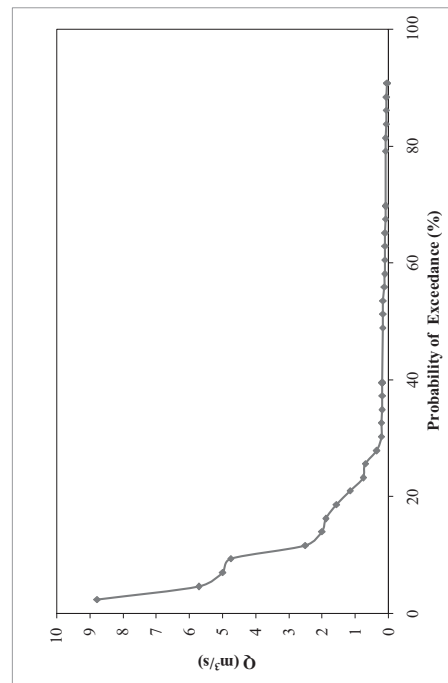


Figure B.158. Monthly FDC of DSI-2333

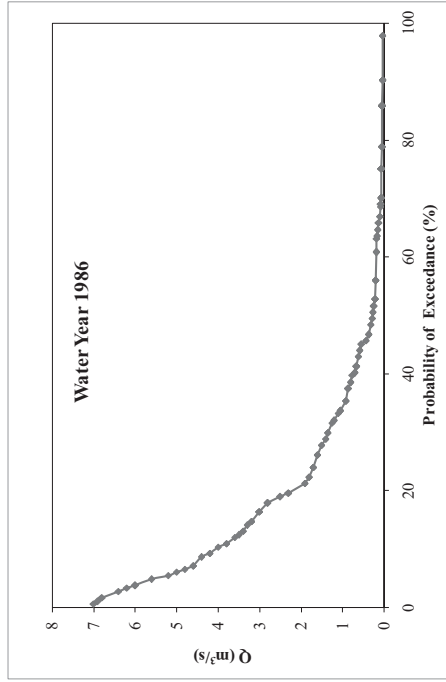


Figure B.159. DSI-2333 1986 water year yearly FDC

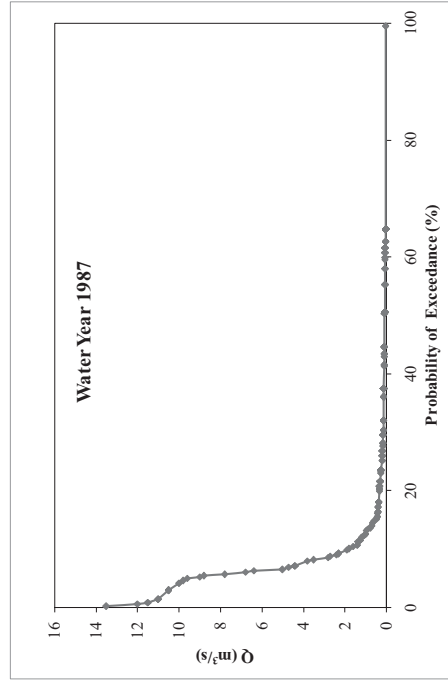


Figure B.160. DSI-2333 1987 water year yearly FDC

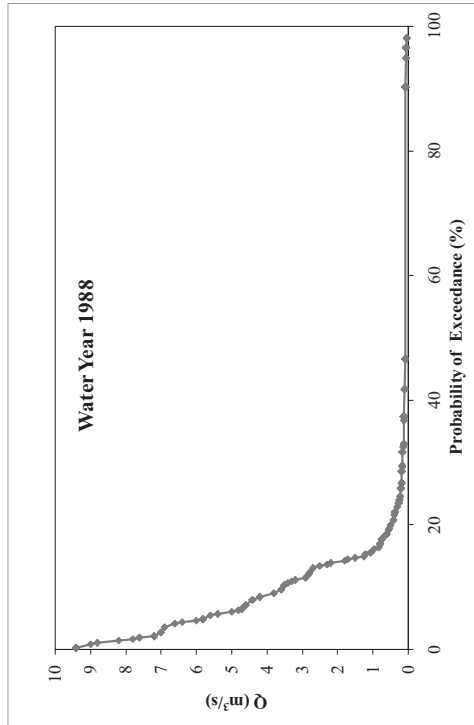


Figure B.161. DSI-2333 1988 water year yearly FDC

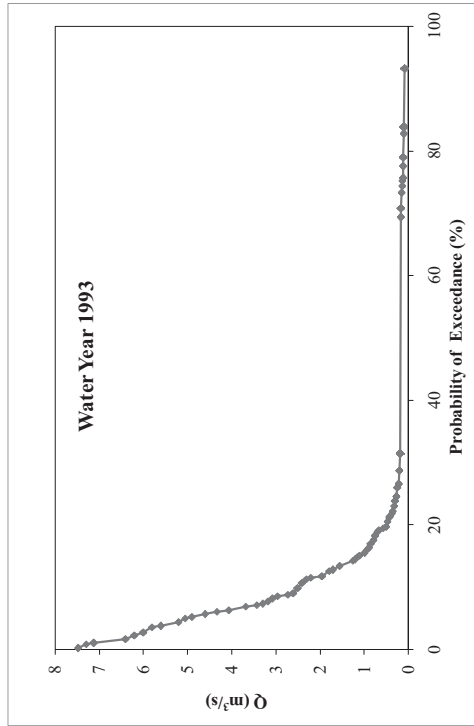


Figure B.162. DSI-2333 1993 water year yearly FDC

- **DSI-2335**

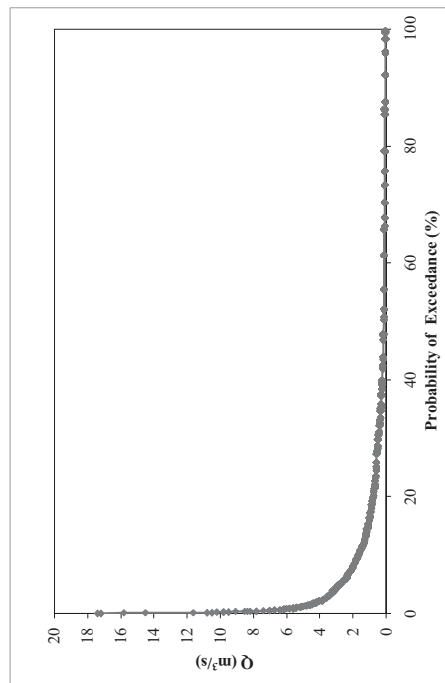


Figure B.163. Daily FDC of DSI-2335

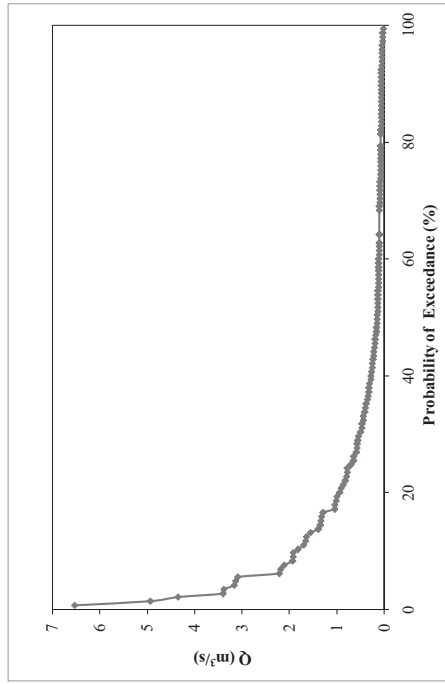


Figure B.164. Monthly FDC of DSI-2335

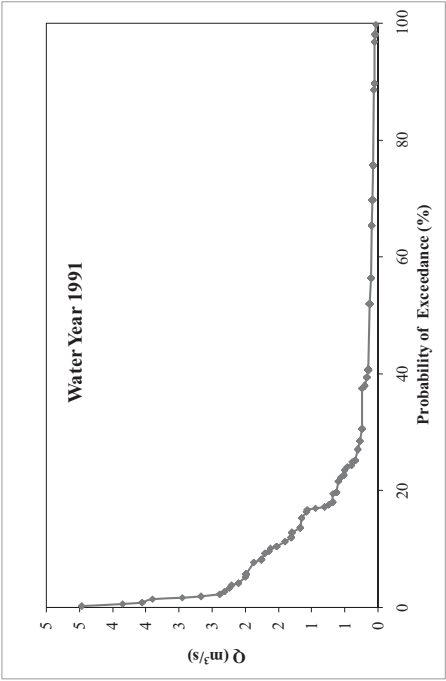


Figure B.165. DSI-2335 1991 water year yearly FDC

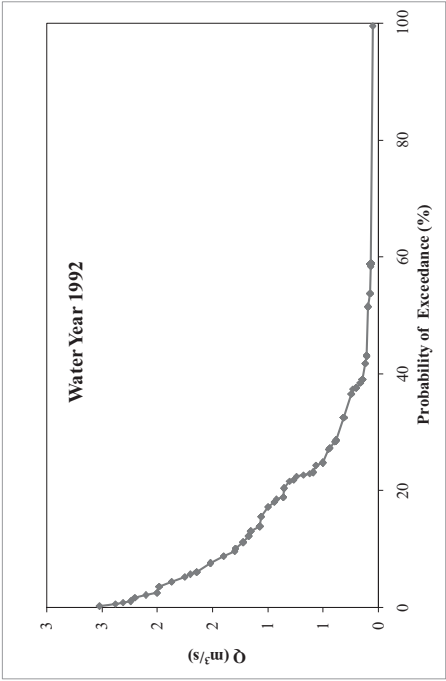


Figure B.166. DSI-2335 1992 water year yearly FDC

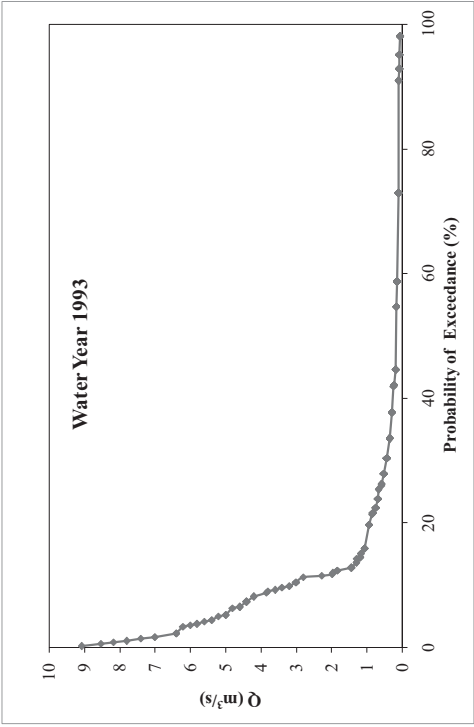


Figure B.167. DSI-2335 1993 water year yearly FDC

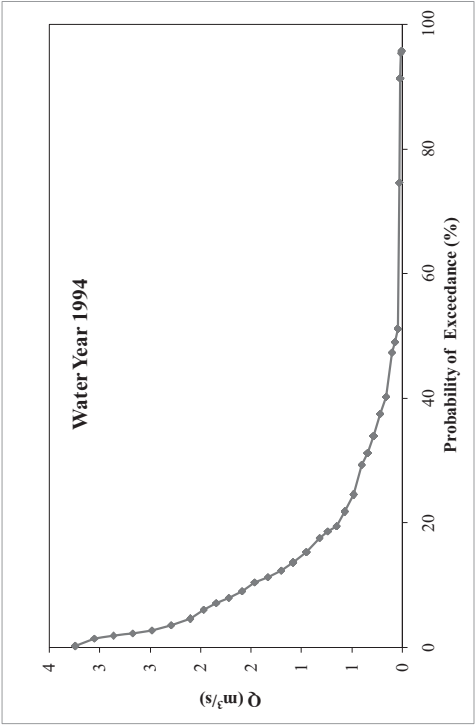


Figure B.168. DSI-2335 1994 water year yearly FDC

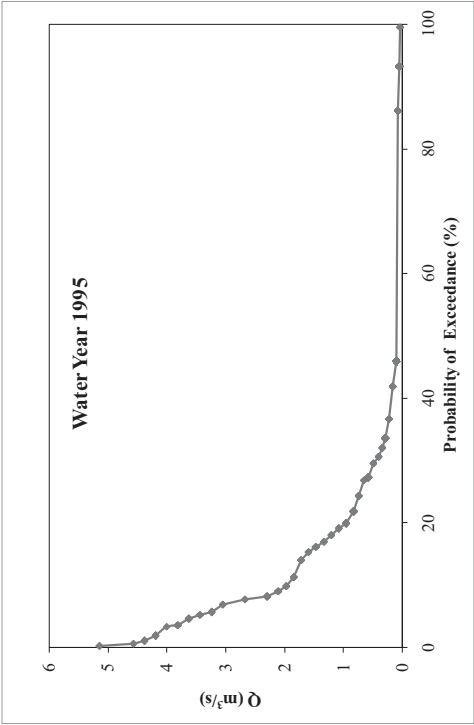


Figure B.169. DSI-2335 1995 water year yearly FDC

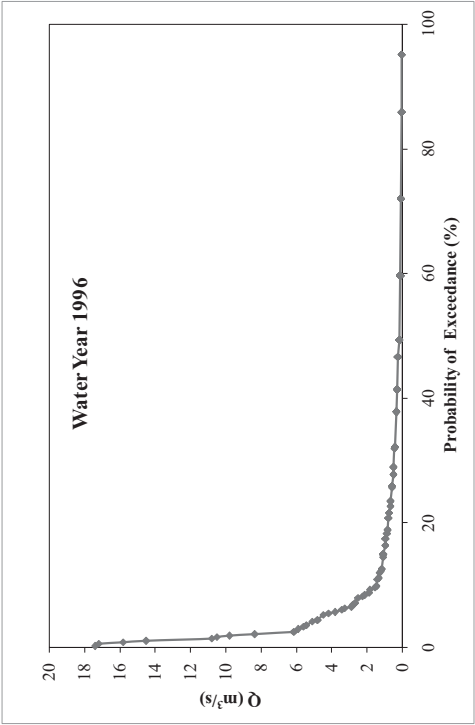


Figure B.170. DSI-2335 1996 water year yearly FDC

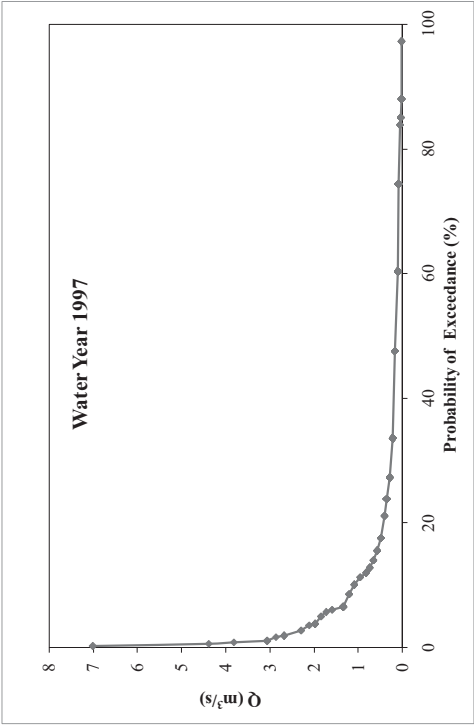


Figure B.171. DSI-2335 1997 water year yearly FDC

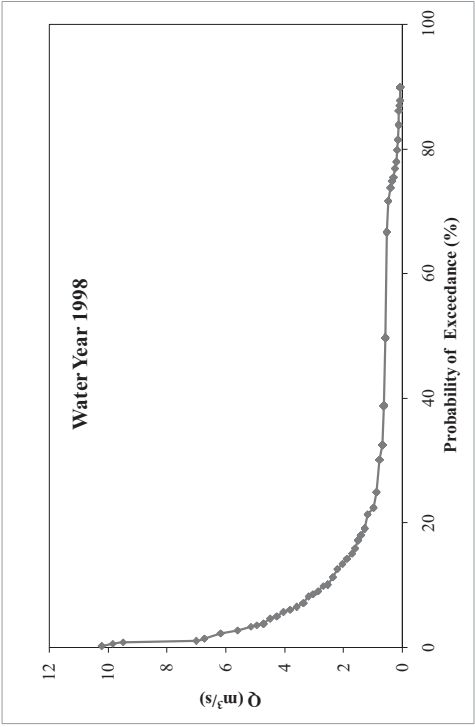


Figure B.172. DSI-2335 1998 water year yearly FDC

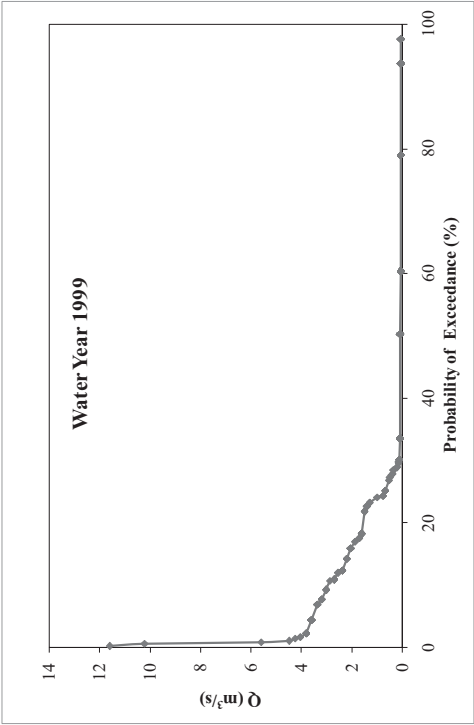


Figure B.173. DSI-2335 1999 water year yearly FDC

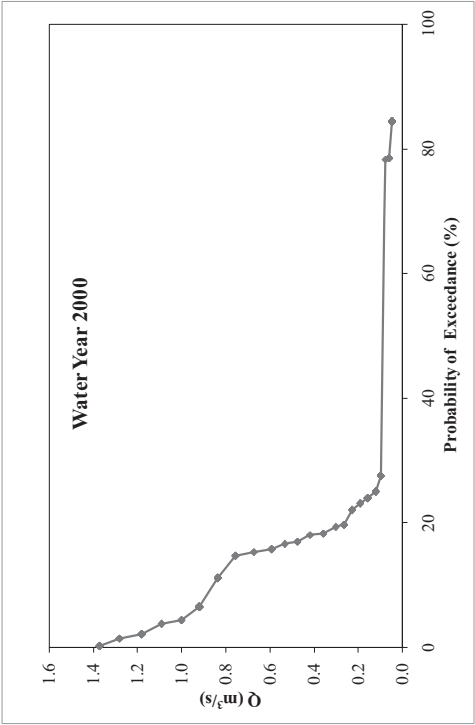


Figure B.174. DSI-2335 2000 water year yearly FDC

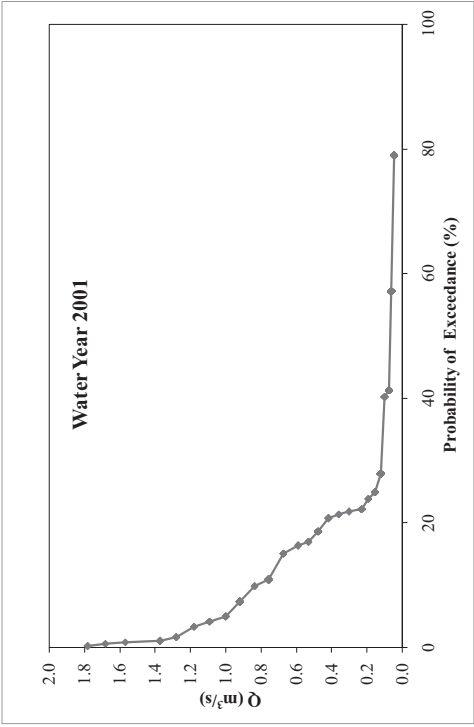


Figure B.175. DSI-2335 2001 water year yearly FDC

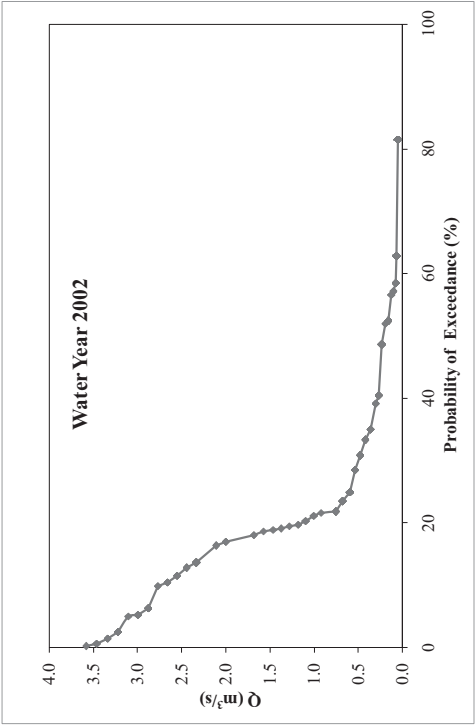


Figure B.176. DSI-2335 2002 water year yearly FDC

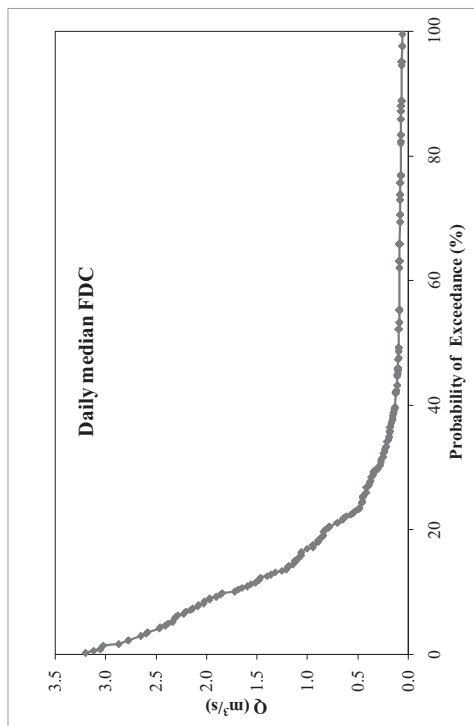


Figure B.177. Daily median FDC of DSI-2335



- DSI-2336

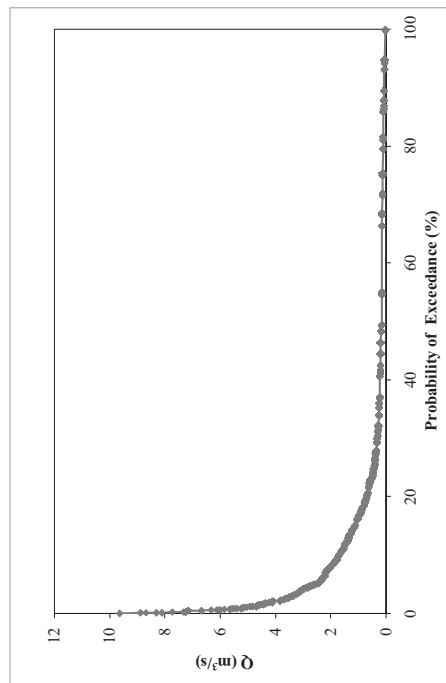


Figure B.178. Daily FDC of DSI-2336

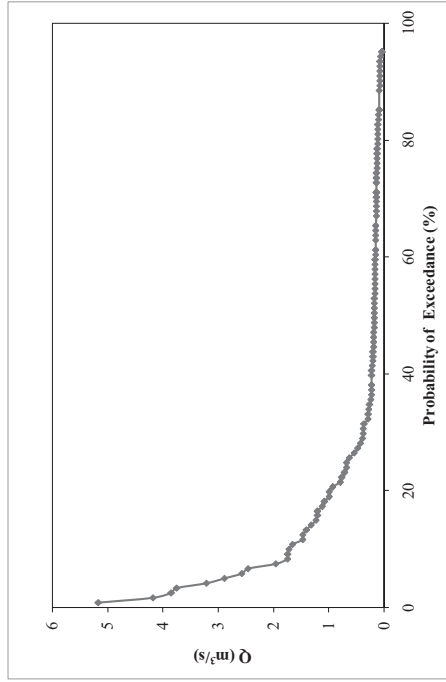


Figure B.179. Monthly FDC of DSI-2336

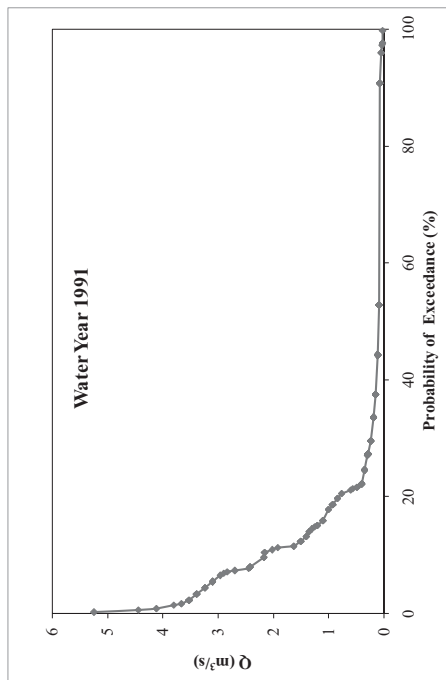


Figure B.180. DSI-2336 1991 water year yearly FDC

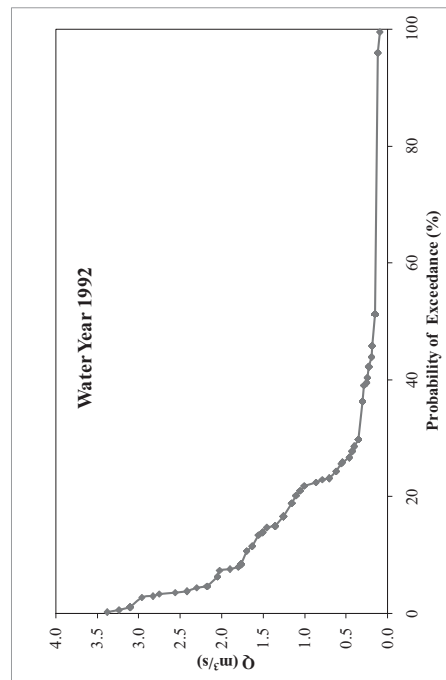


Figure B.181. DSI-2336 1992 water year yearly FDC

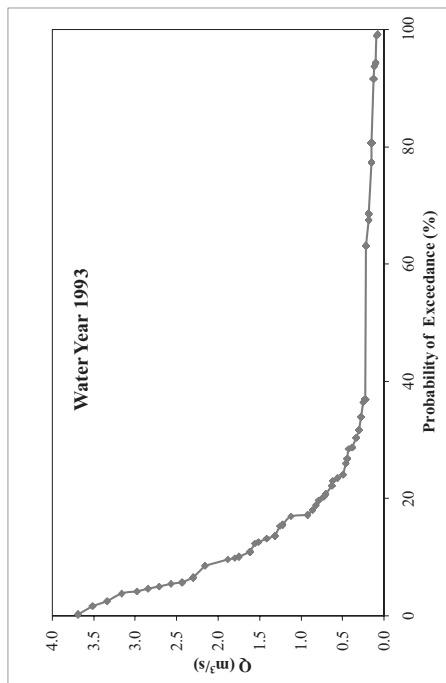


Figure B.182. DSI-2336 1993 water year yearly FDC

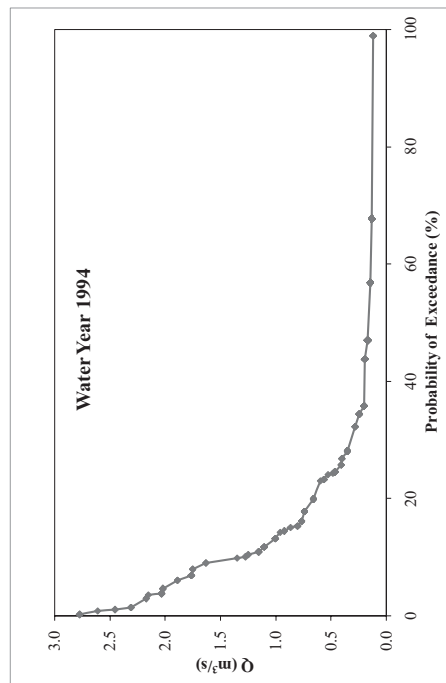


Figure B.183. DSI-2336 1994 water year yearly FDC

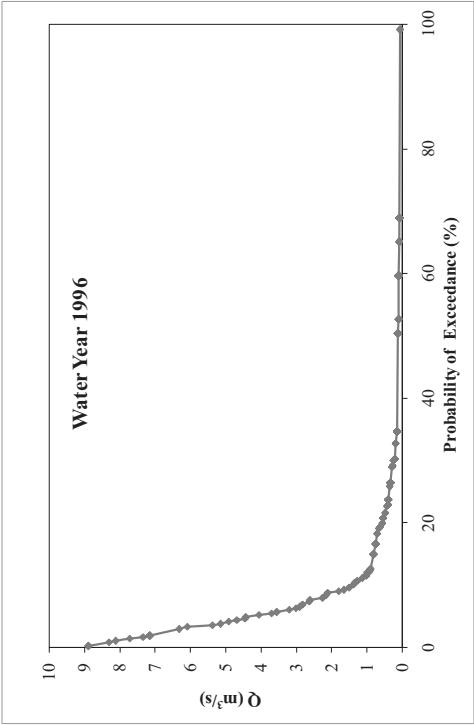


Figure B.184. DSI-2336 1996 water year yearly FDC

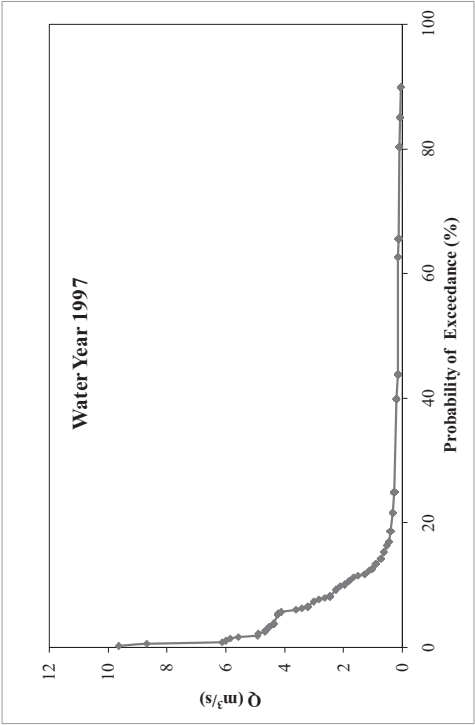


Figure B.185. DSI-2336 1997 water year yearly FDC

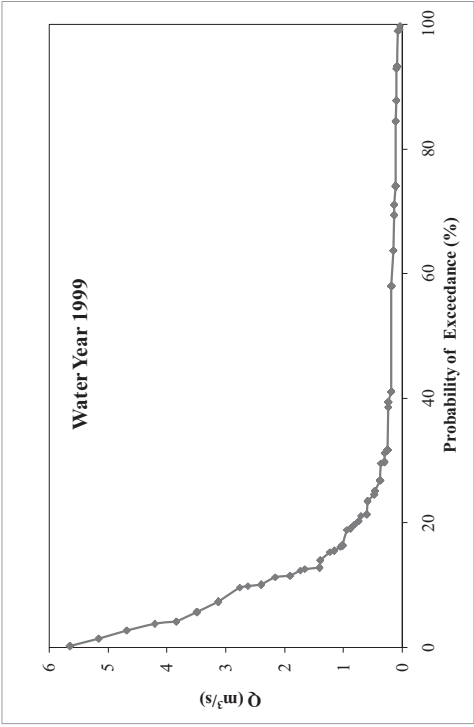


Figure B.186. DSI-2336 1999 water year yearly FDC

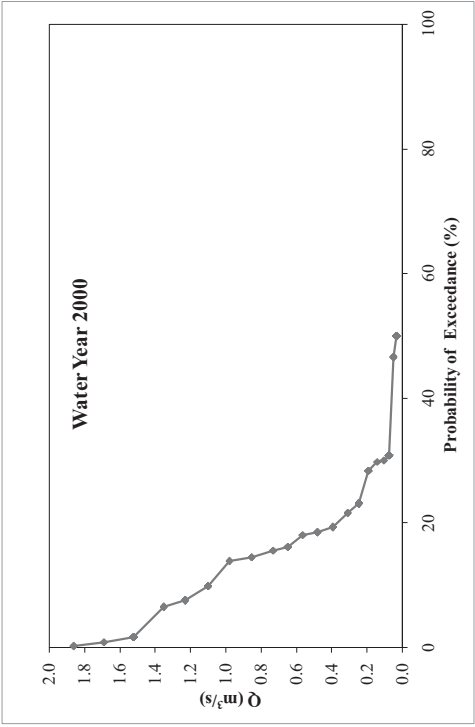


Figure B.187. DSI-2336 2000 water year yearly FDC

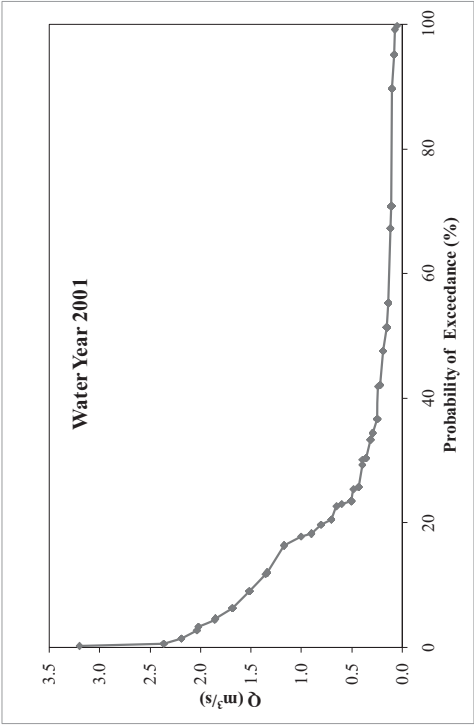


Figure B.188. DSI-2336 2001 water year yearly FDC

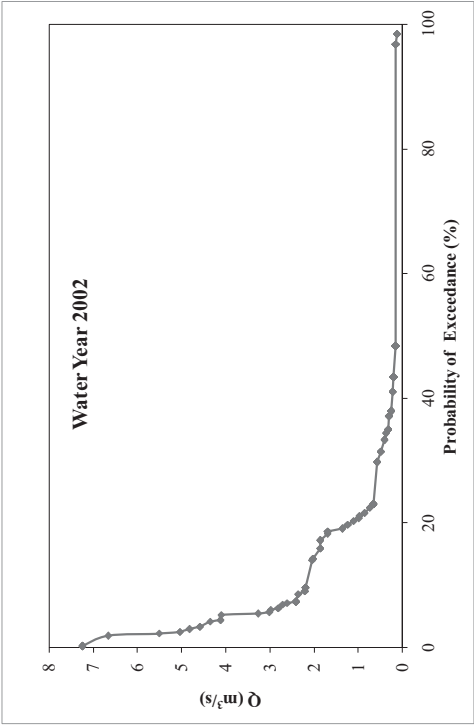


Figure B.189. DSI-2336 2002 water year yearly FDC

- DSI-2337

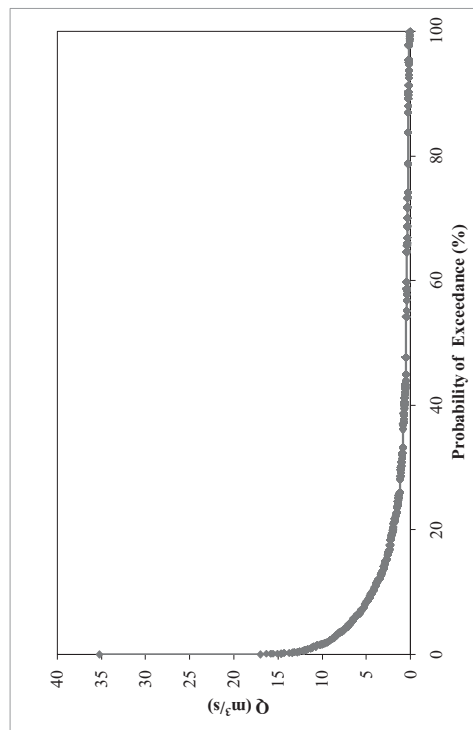


Figure B.190. Daily FDC of DSI-2337

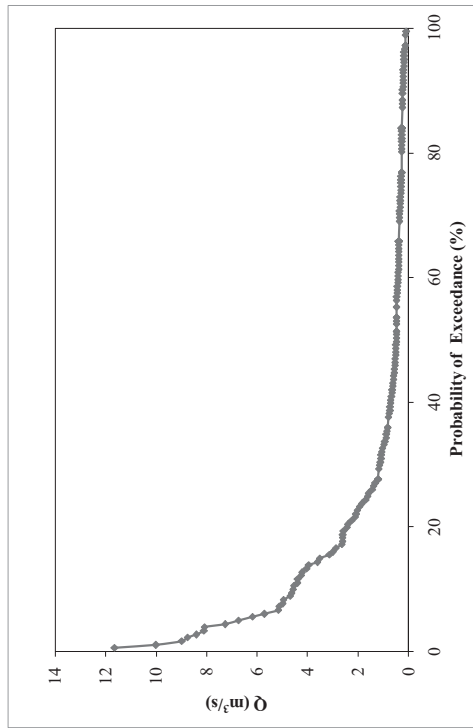


Figure B.191. Monthly FDC of DSI-2337

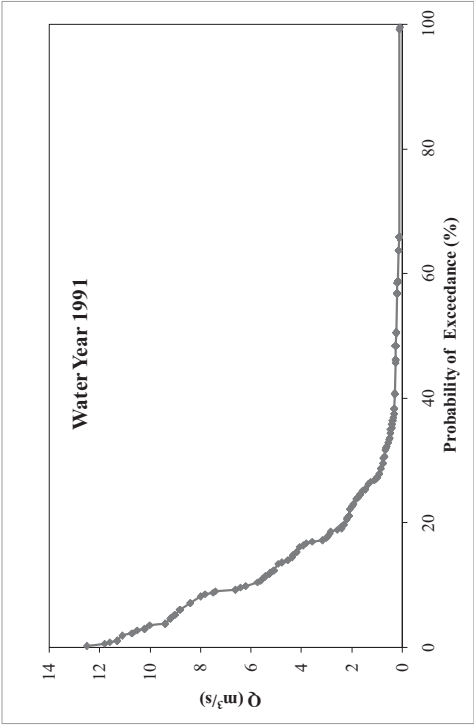


Figure B.192. DSI-2337 1991 water year yearly FDC

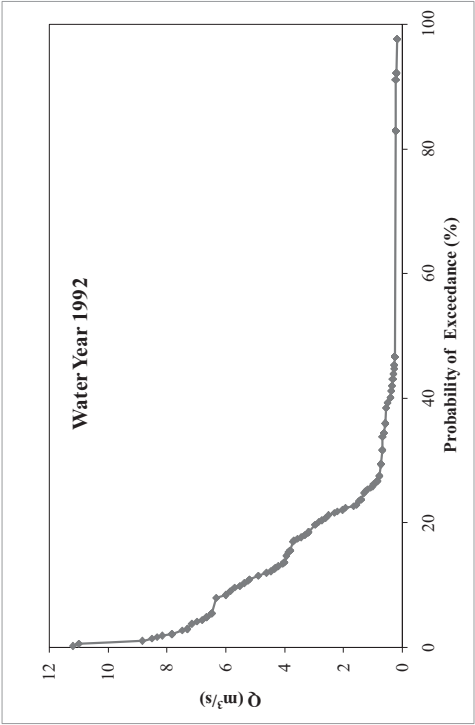


Figure B.193. DSI-2337 1992 water year yearly FDC

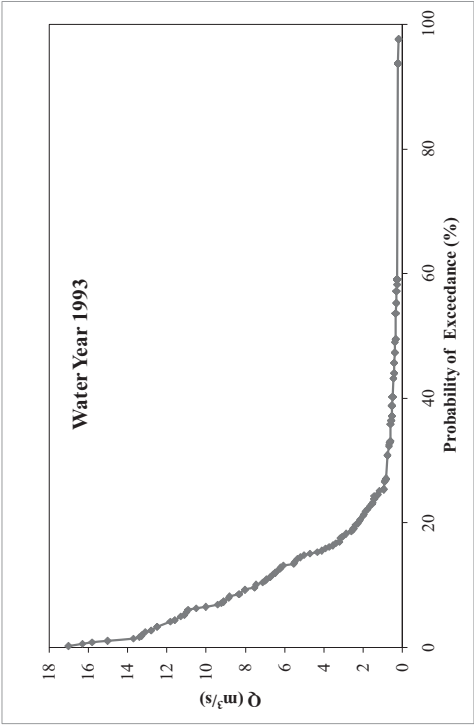


Figure B.194. DSI-2337 1993 water year yearly FDC

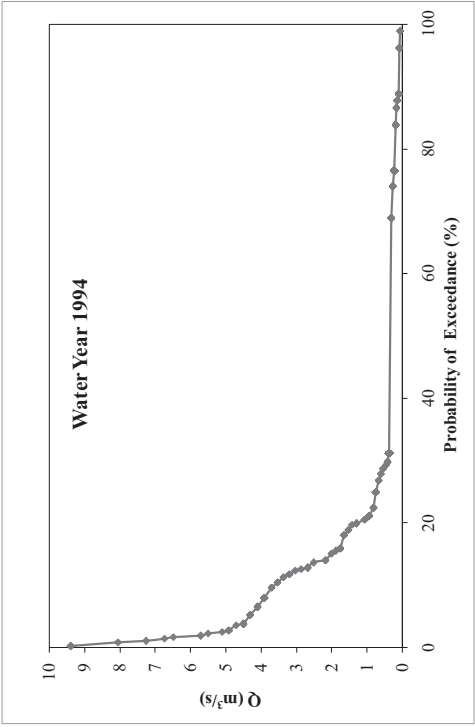


Figure B.195. DSI-2337 1994 water year yearly FDC

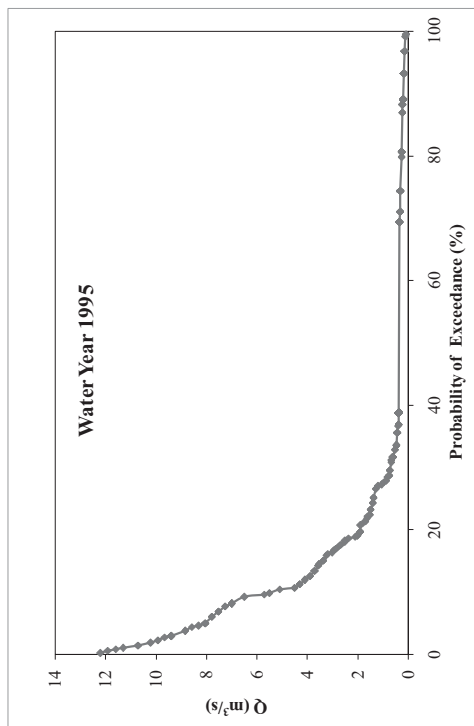


Figure B.196. DSI-2337 1995 water year yearly FDC

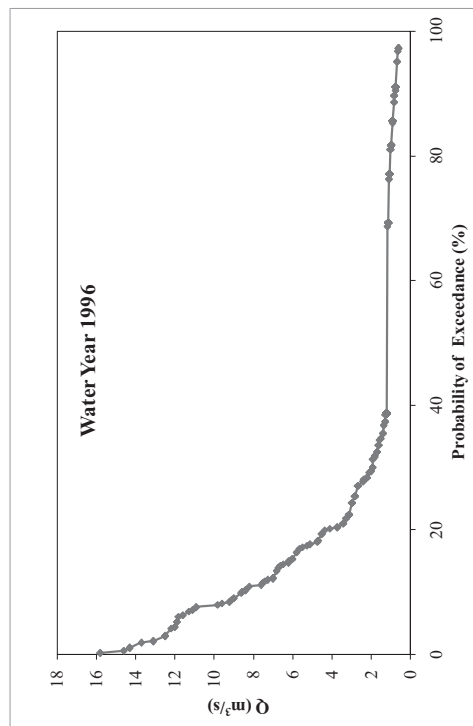


Figure B.197. DSI-2337 1996 water year yearly FDC

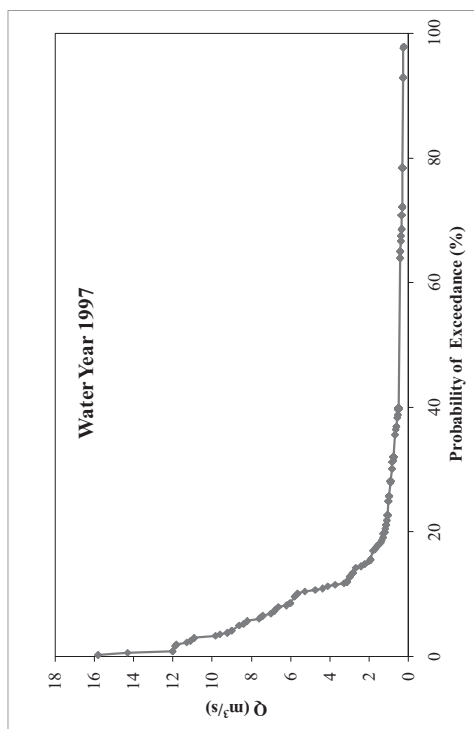


Figure B.198. DSI-2337 1997 water year yearly FDC

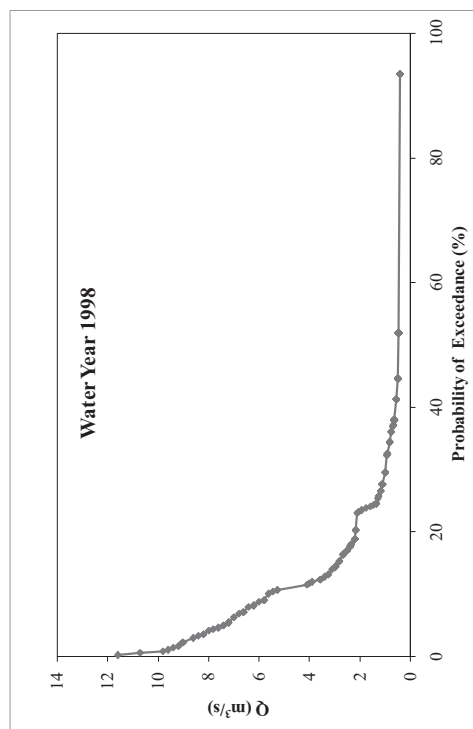


Figure B.199. DSI-2337 1998 water year yearly FDC

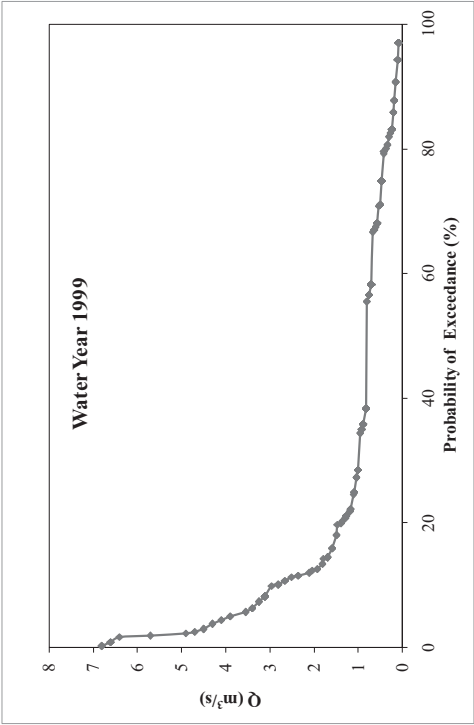


Figure B.200. DSI-2337 1999 water year yearly FDC

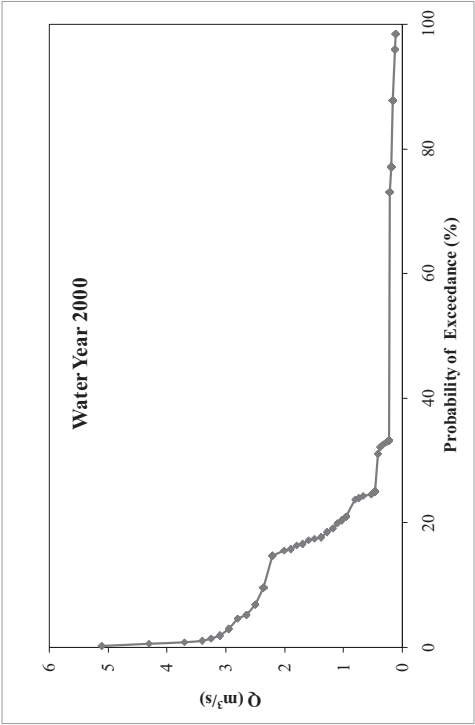


Figure B.201. DSI-2337 2000 water year yearly FDC

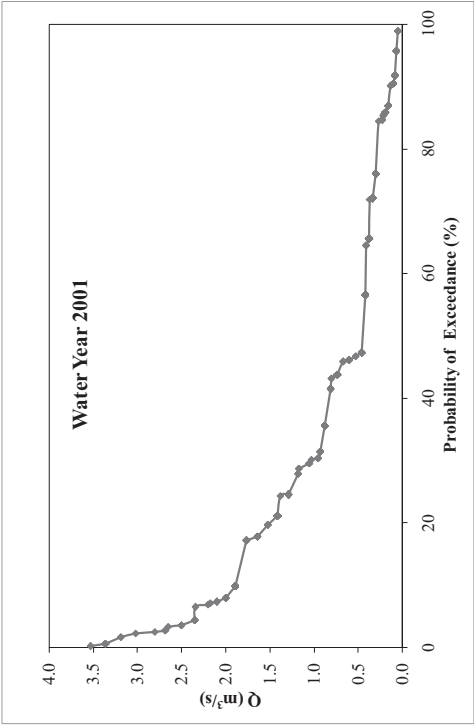


Figure B.202. DSI-2337 2001 water year yearly FDC

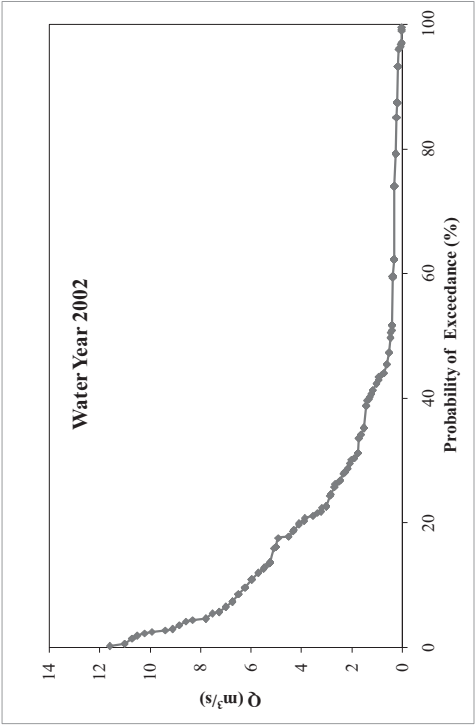


Figure B.203. DSI-2337 2002 water year yearly FDC



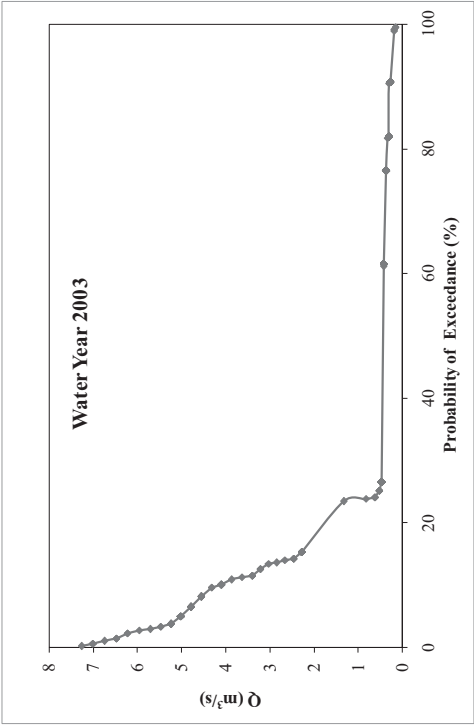


Figure B.204. DSI-2337 2003 water year yearly FDC

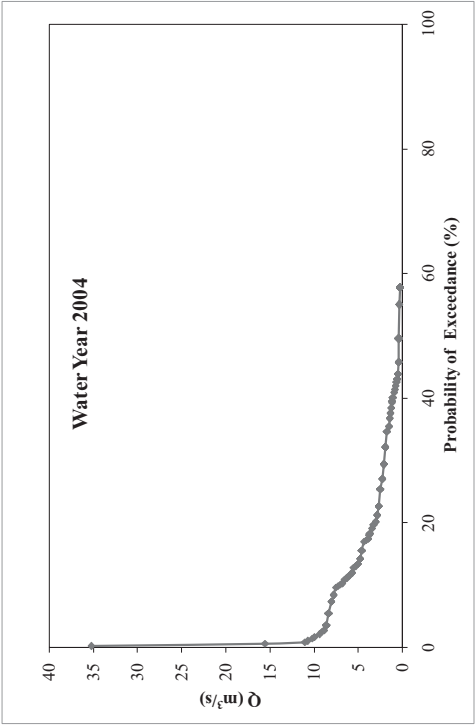


Figure B.205. DSI-2337 2004 water year yearly FDC

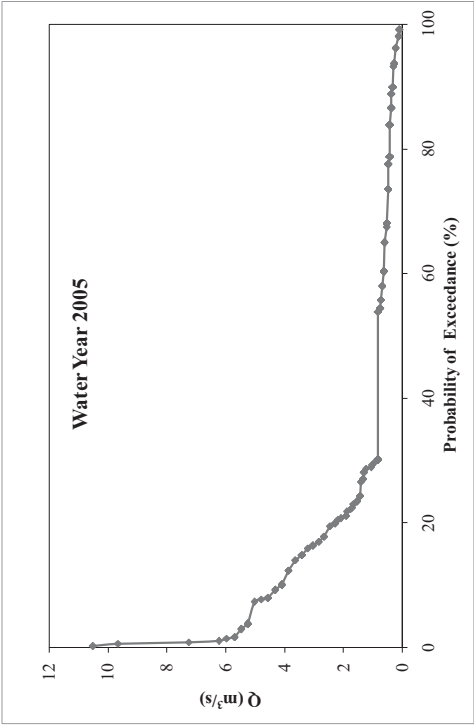


Figure B.206. DSI-2337 2005 water year yearly FDC

- **DSI-2338**

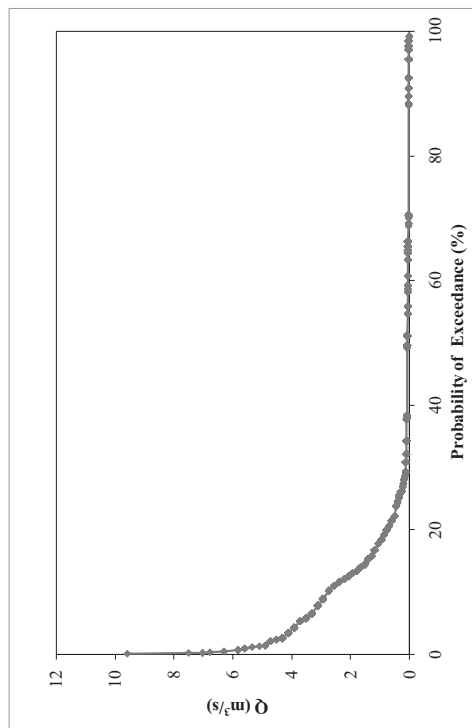


Figure B.207. Daily FDC of DSI-2338

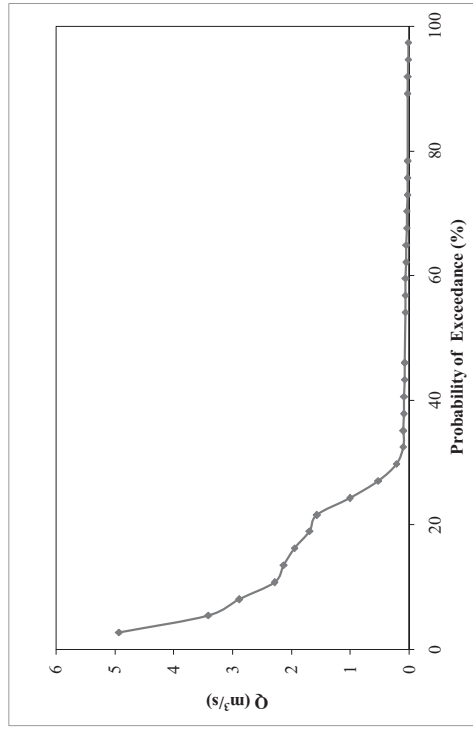


Figure B.208. Monthly FDC of DSI-2338

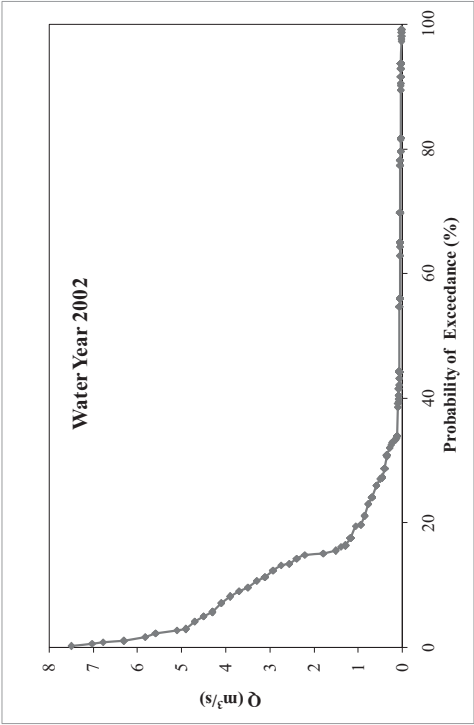


Figure B.209. DSI-2338 2002 water year yearly FDC

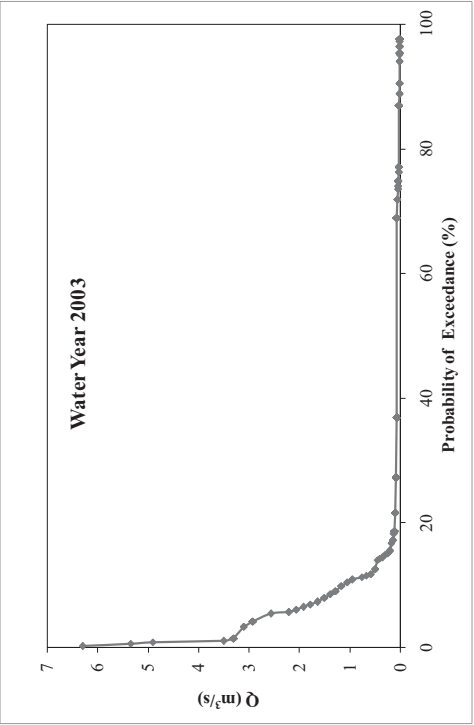


Figure B.210. DSI-2338 2003 water year yearly FDC

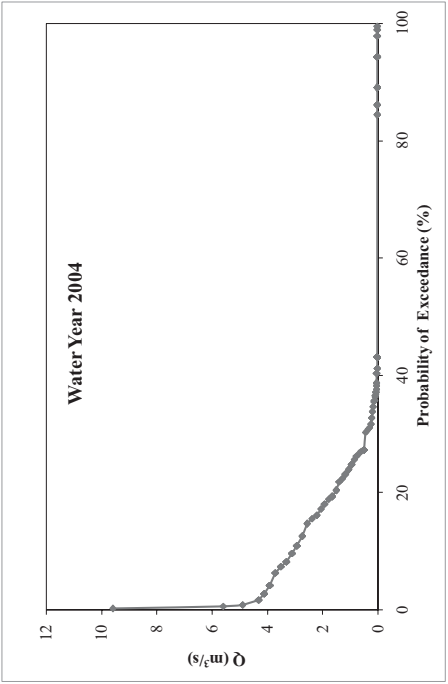


Figure B.211. DSI-2338 2004 water year yearly FDC

- DSI-2339

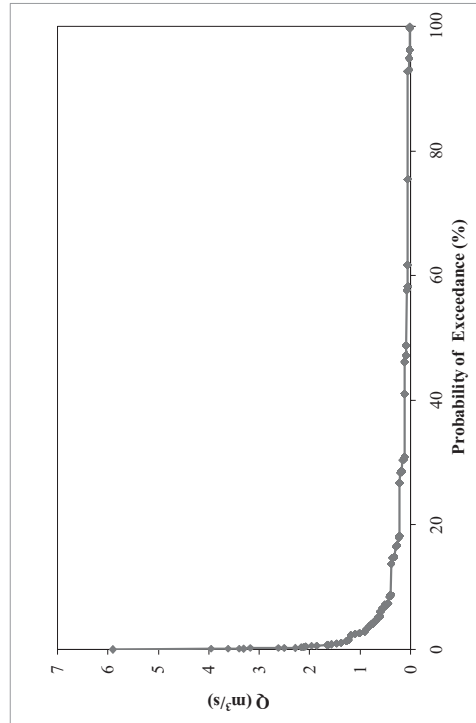


Figure B212. Daily FDC of DSI-2339

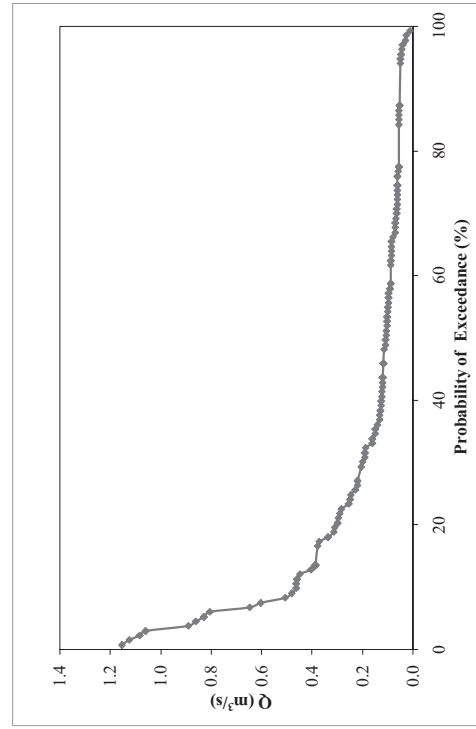


Figure B.213. Monthly FDC of DSI-2339

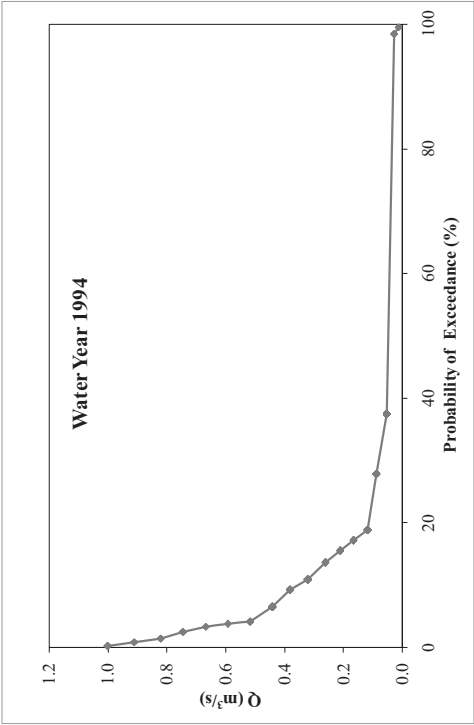


Figure B.214. DSI-2339 1994 water year yearly FDC

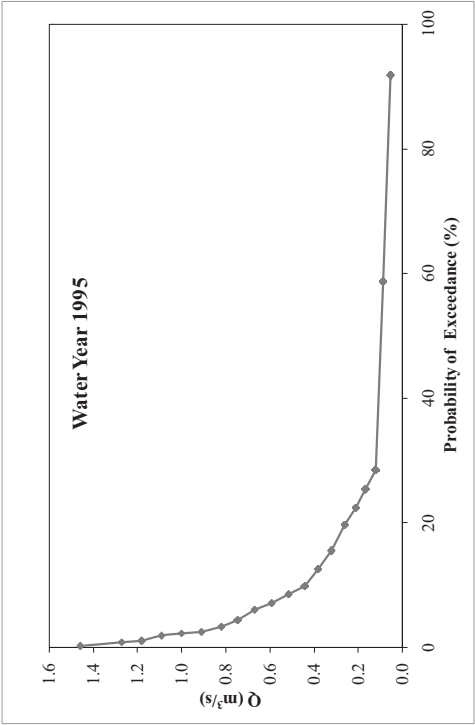


Figure B.215. DSI-2339 1995 water year yearly FDC

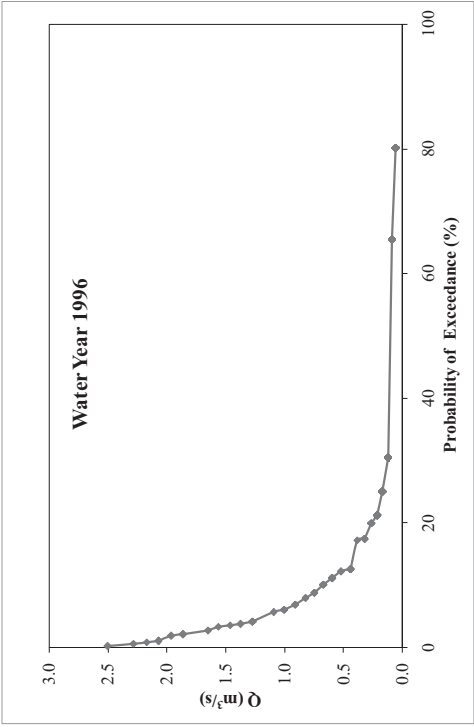


Figure B.216. DSI-2339 1996 water year yearly FDC

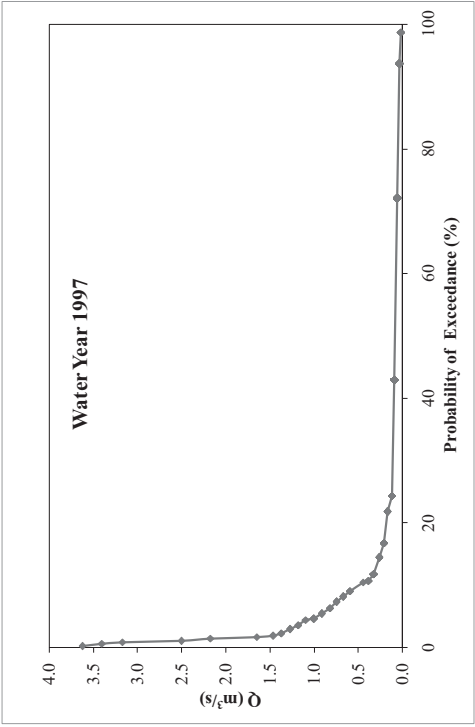


Figure B.217. DSI-2339 1997 water year yearly FDC

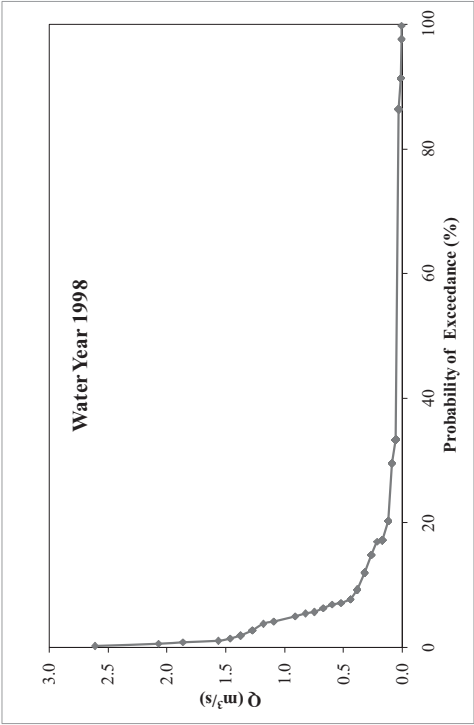


Figure B.218. DSI-2339 1998 water year yearly FDC

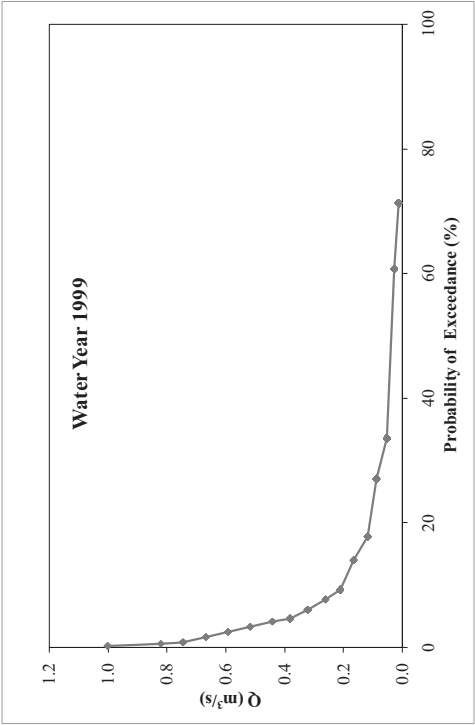


Figure B.219. DSI-2339 1999 water year yearly FDC

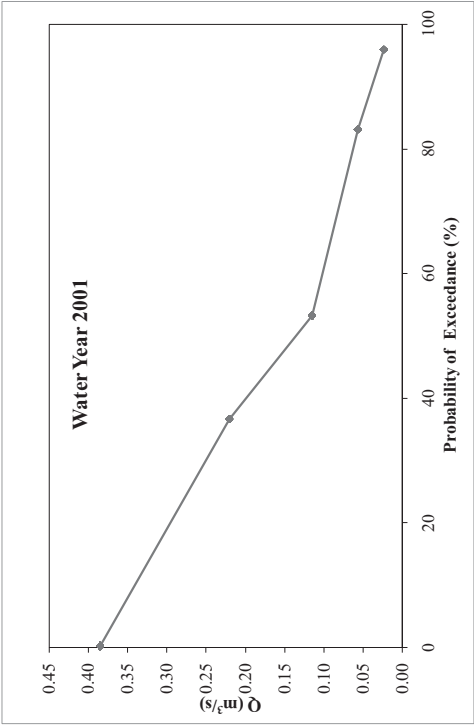


Figure B.220. DSI-2339 2001 water year yearly FDC

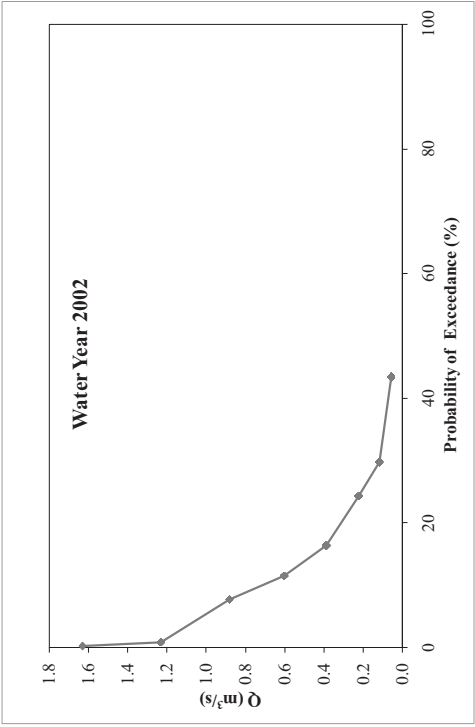


Figure B.221. DSI-2339 2002 water year yearly FDC

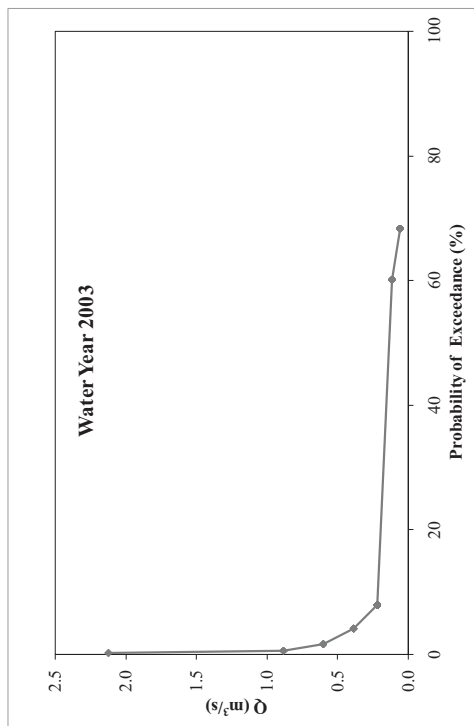


Figure B.222. DSI-2339 2003 water year yearly FDC

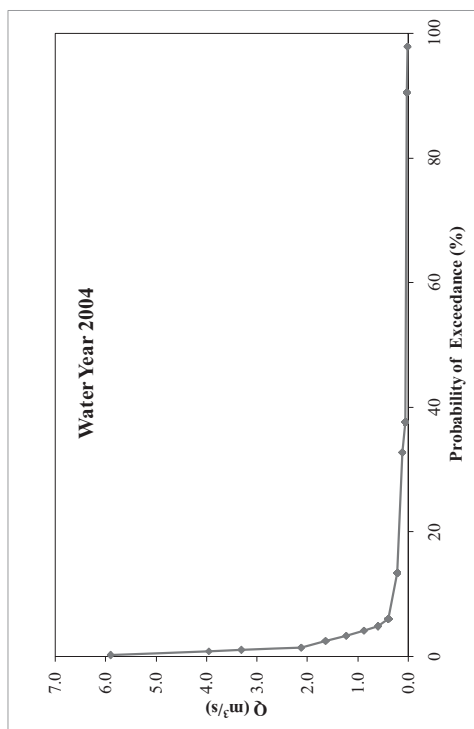


Figure B.223. DSI-2339 2004 water year yearly FDC

## APPENDIX C

### TOPOGRAPHIC DATA PREPROCESSING

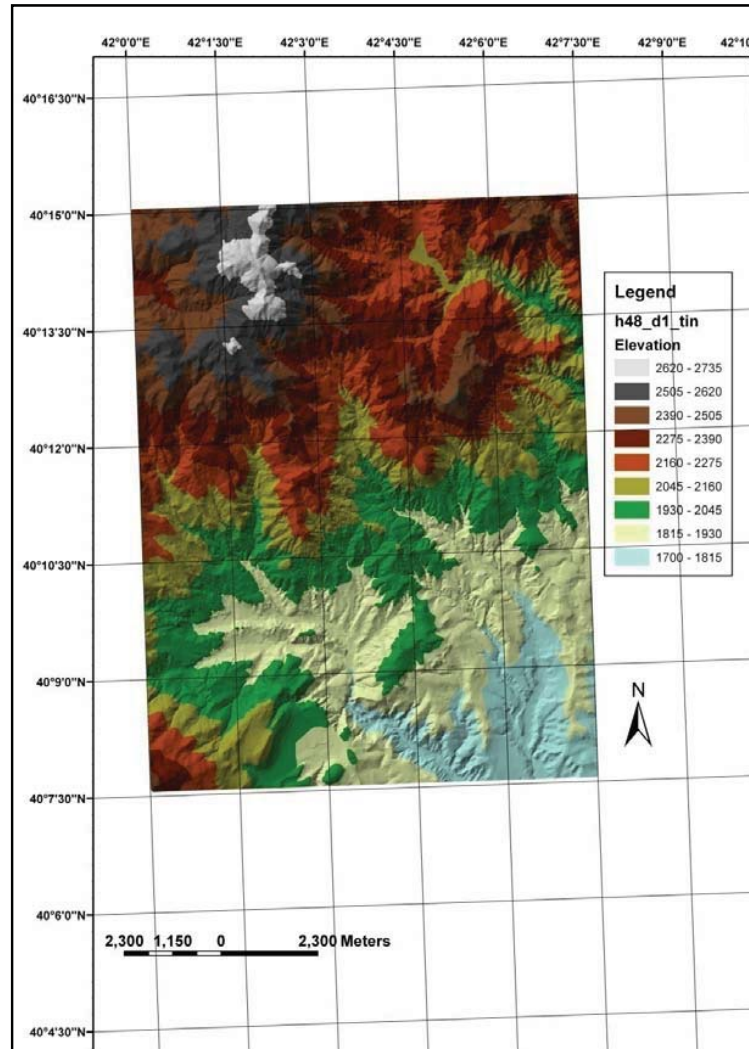


Figure C.1. H48-d1 map section TIN image



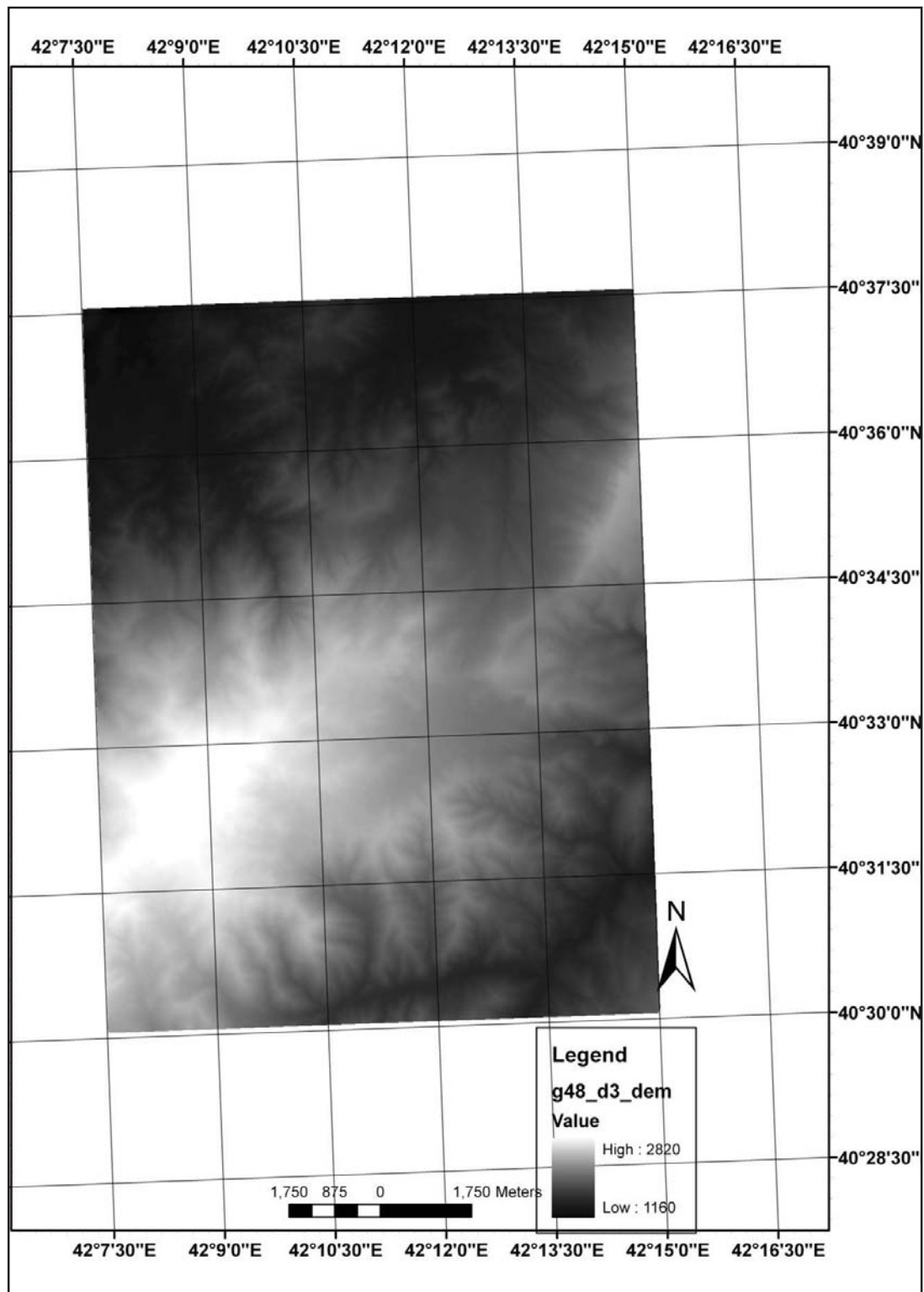


Figure C.2. G48-d3 map section DEM image

- Steps implemented while creating the subbasin boundaries and drainage line network
  1. ArcHydro (AH), Terrain Preprocessing (TP) → DEM Manipulation → Fill Sinks → Filled DEM
  2. Filled DEM, AH → TP → Flow Direction
  3. Flow Direction, AH → TP → Flow Accumulation
  4. Flow Accumulation, AH → TP → Stream Definition → Stream (10000), 10k
  5. Stream, AH → TP → Stream Segmentation → Stream Link
  6. Stream Link, AH → TP → Catchment Grid Delineation → Cat
  7. Cat, AH → TP → Catchment Polygon Processing → Catchment Polygons (Catchment Boundaries)
  8. AH → TP → Drainage Line Processing → Drainage Line

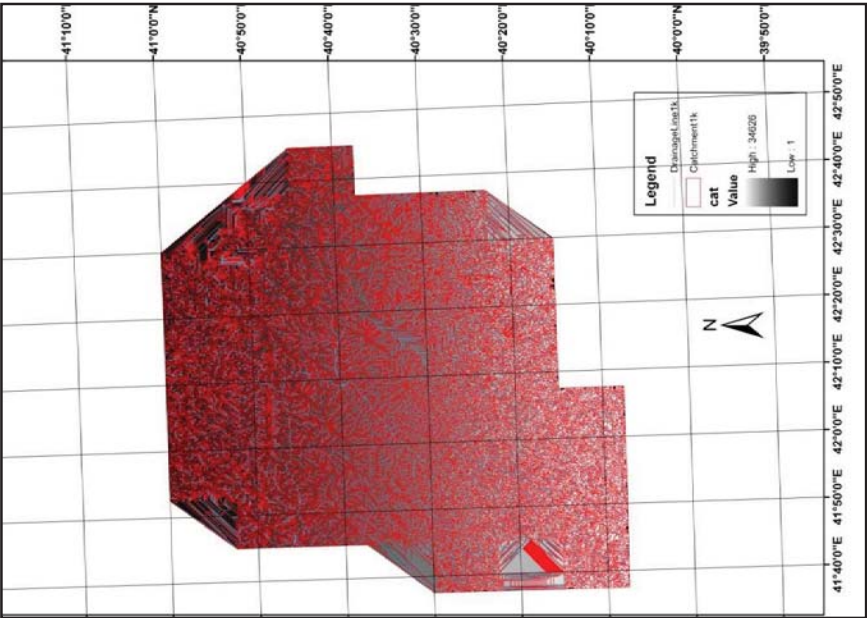


Figure C.3. Catchment boundaries and drainage network developed  
1k stream

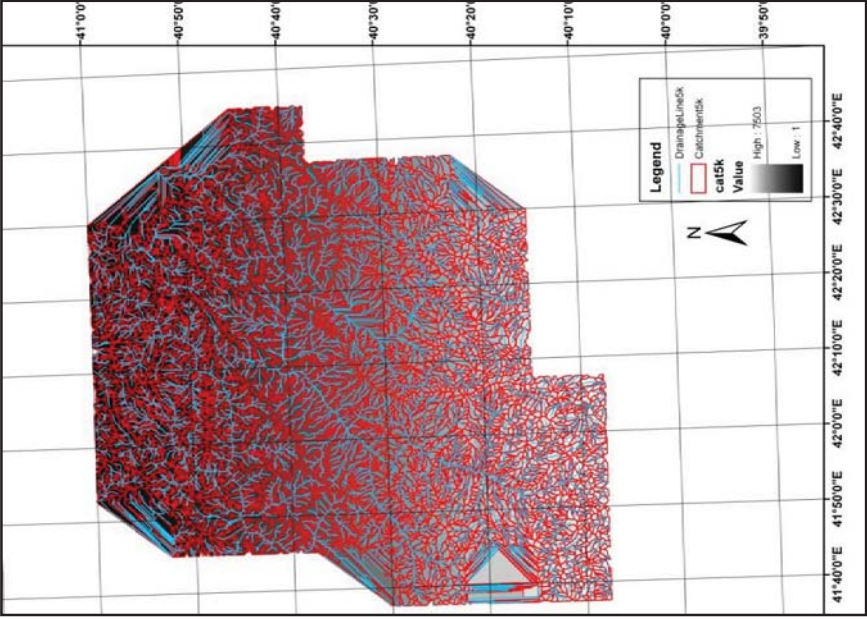


Figure C.4. Catchment boundaries and drainage network developed  
by 5k stream

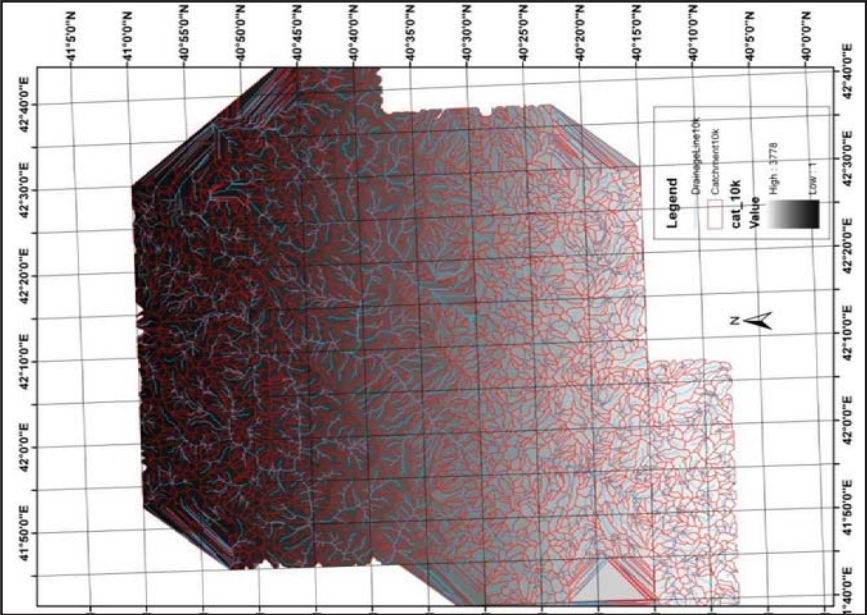


Figure C.5. Catchment boundaries and drainage network developed by 10k stream

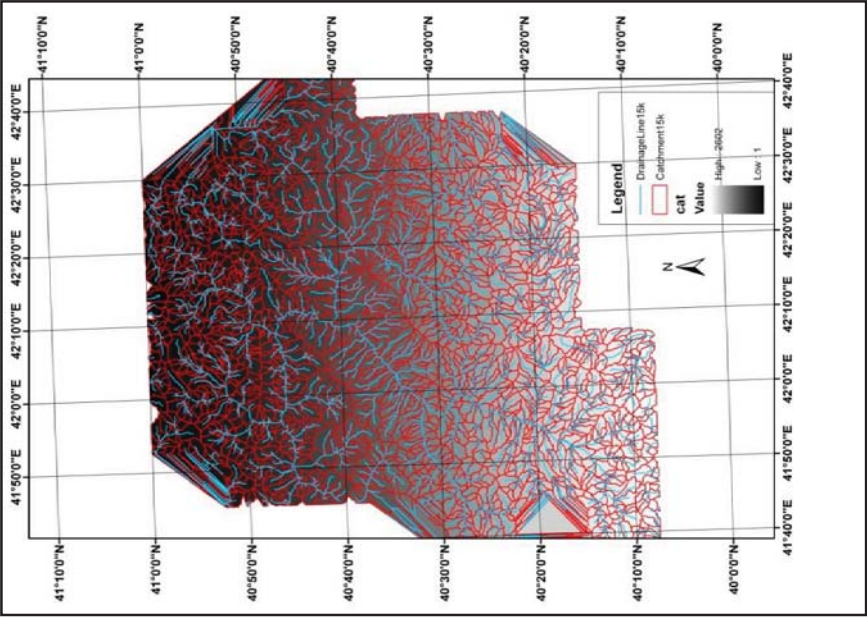


Figure C.6. Catchment boundaries and drainage network developed by 15k stream

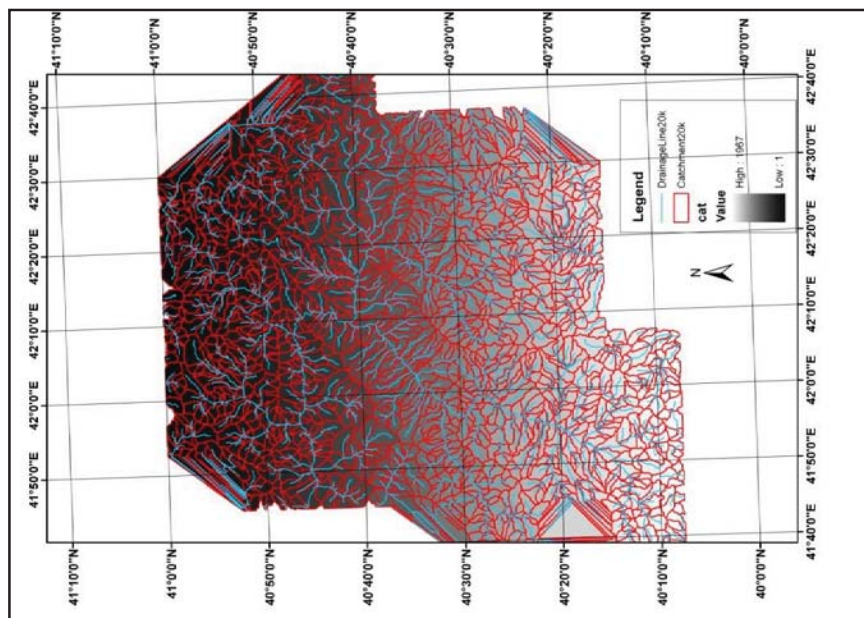


Figure C.7. Catchment boundaries and drainage network developed by 20k stream

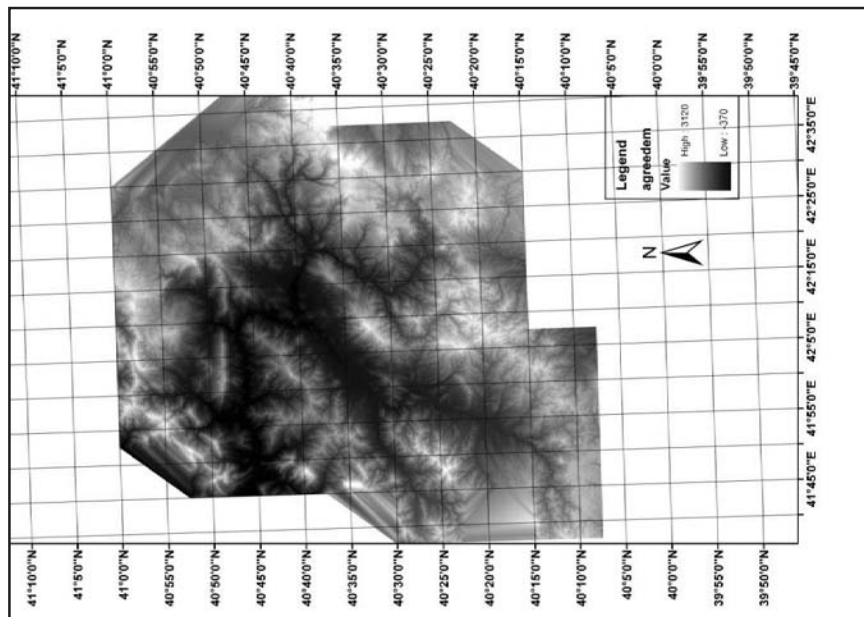


Figure C.8. Agree DEM generated with drainage network 1k



## APPENDIX D:

### DATA PROCESSING

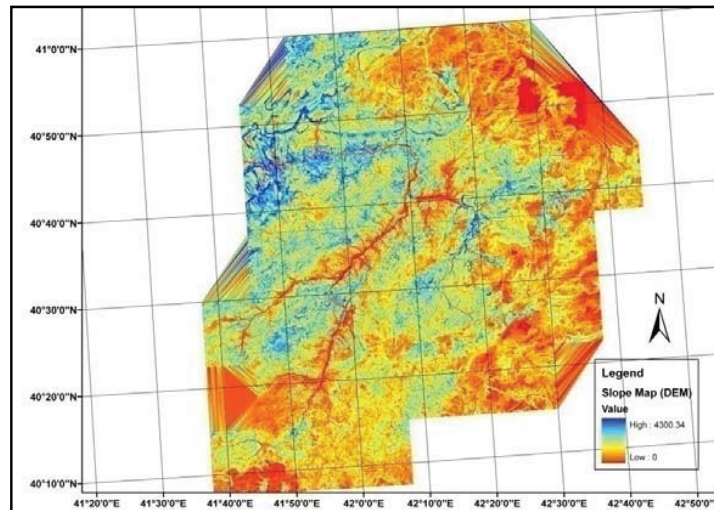


Figure D.1. Slope image of Oltu basin created from DEM

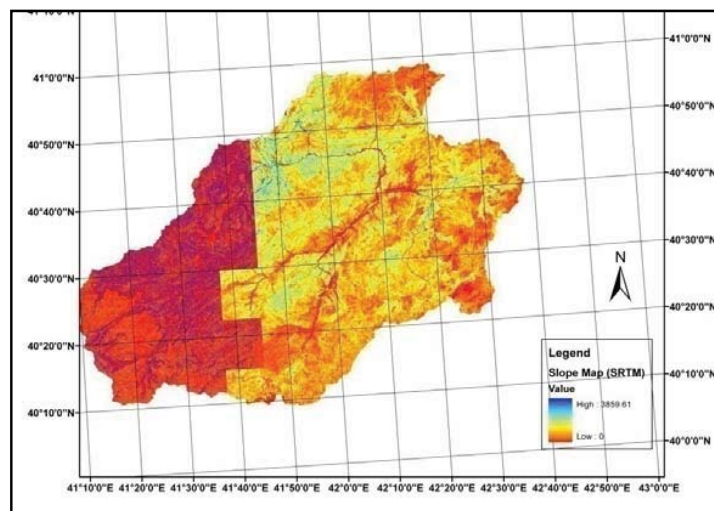


Figure D.2. Slope image of Oltu basin created from SRTM

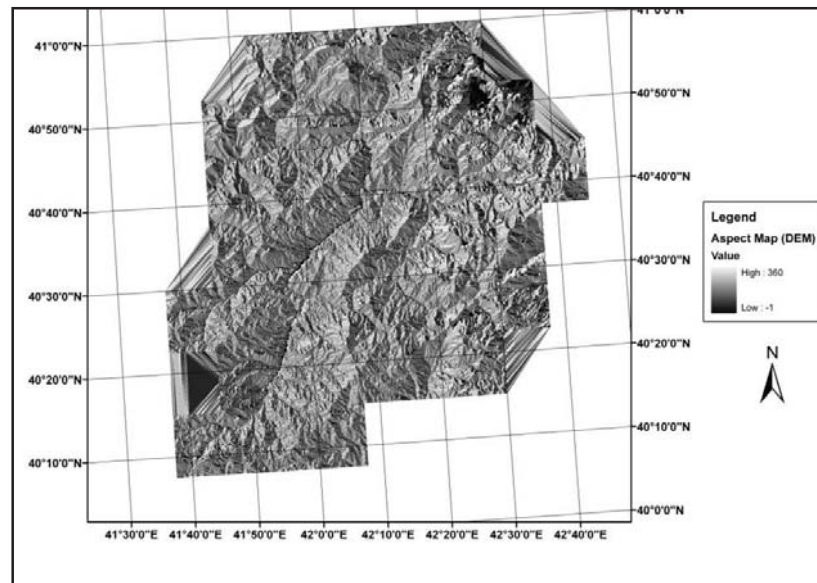


Figure D.3. Aspect image of Oltu basin created from DE

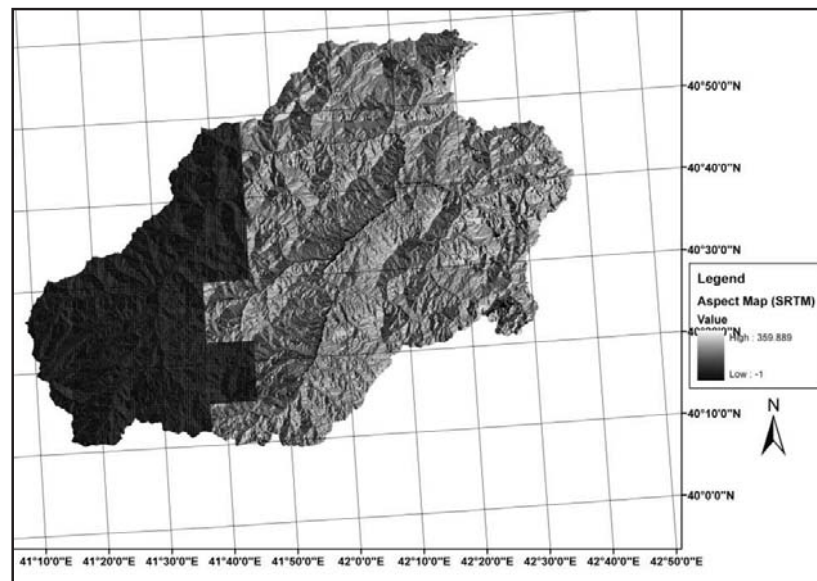


Figure D.4. Aspect image of Oltu basin created from SRTM

# APPENDIX E

## HETEROGENEITY AND SEASONALITY ANALYSIS

### 1. HETEROGENEITY ANALYSIS

In this example, daily discharge data of all 14 stations have been used in the cluster analysis. The example includes that only 3 homogeneous clusters are obtained from the analysis. Figure E.1 shows these homogeneous clusters. Rest of the computations have been performed according to these sub-clusters. Table E.1 presents the results table of this special example, after implementing the MatLAB program and Excel formulations.

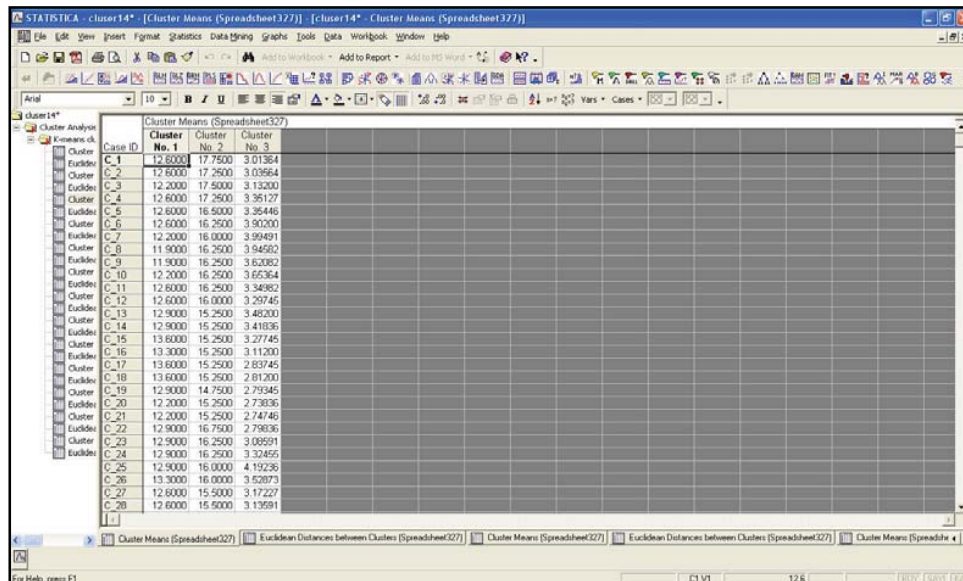


Figure E.1. 3 homogeneous clusters obtained from station data



Table E.1. L- Moments and  $V_1$ ,  $V_2$  and  $V_3$  table for the case given above

	<b>L1</b>	<b>L2</b>	<b>L3</b>	<b>L4</b>	<b>L-Cv</b>	<b>L-skewness</b>	<b>L-kurtosis</b>
Cluster1	44.94	22.85	12.20	5.53	0.51	0.53	0.24
Cluster2	30.70	13.30	7.70	4.33	0.43	0.58	0.33
Cluster3	4.41	1.93	0.95	0.63	0.44	0.49	0.33
				mean	0.46	0.53	0.30

	$n_i$	$t_2$	$D^2$	$n_i D$
Cluster1	731	0.51	0.0024	1.72
Cluster2	731	0.43	0.0007	0.52
Cluster3	731	0.44	0.0005	0.35
$\Sigma$	2193		$\Sigma$	2.59
	<b>V1</b>	<b>0.001182</b>		

	$n_i$	$t_2$	$t_3$	K
Cluster1	731	0.51	0.53	35.49
Cluster2	731	0.43	0.58	37.71
Cluster3	731	0.44	0.49	35.47
$\Sigma$	2193		$\Sigma$	108.67
	<b>V2</b>	<b>0.049554</b>		

	$n_i$	$t_3$	$t_4$	K
Cluster1	731	0.53	0.24	40.71
Cluster2	731	0.58	0.33	38.04
Cluster3	731	0.49	0.33	37.76
$\Sigma$	2193		$\Sigma$	116.51
	<b>V3</b>	<b>0.053129</b>		

## 2. SEASONALITY ANALYSIS

- 5% Probability of Exceedance

Table E.2. Basin Euclid distances for 5% probability of exceedance

i = EIE-2323	d <sub>23-i</sub>	i = EIE-2325	d <sub>25-i</sub>	i = EIE-2329	d <sub>29-i</sub>	i = DSI-2313	d <sub>13-i</sub>	i = DSI-2321	d <sub>21-i</sub>
j = EIE-2325	0.107	j = EIE-2329	0.025	j = DSI-2313	0.084	j = DSI-2321	0.161	j = DSI-2322	0.102
j = EIE-2329	0.085	j = DSI-2313	0.107	j = DSI-2321	0.107	j = DSI-2322	0.085	j = DSI-2323	0.091
j = DSI-2313	0.006	j = DSI-2321	0.090	j = DSI-2322	0.005	j = DSI-2323	0.145	j = DSI-2324	0.115
j = DSI-2321	0.157	j = DSI-2322	0.022	j = DSI-2323	0.061	j = DSI-2324	0.070	j = DSI-2333	0.108
j = DSI-2322	0.086	j = DSI-2323	0.038	j = DSI-2324	0.014	j = DSI-2333	0.086	j = DSI-2335	0.113
j = DSI-2323	0.145	j = DSI-2324	0.039	j = DSI-2333	0.003	j = DSI-2335	0.082	j = DSI-2336	0.161
j = DSI-2324	0.071	j = DSI-2333	0.023	j = DSI-2335	0.005	j = DSI-2336	0.013	j = DSI-2337	0.111
j = DSI-2333	0.087	j = DSI-2335	0.029	j = DSI-2336	0.077	j = DSI-2337	0.085	j = DSI-2338	0.104
j = DSI-2335	0.083	j = DSI-2336	0.101	j = DSI-2337	0.005	j = DSI-2338	0.183	j = DSI-2339	0.099
j = DSI-2336	0.019	j = DSI-2337	0.026	j = DSI-2338	0.099	j = DSI-2339	0.082	j = DSI-2313	0.161
j = DSI-2337	0.086	j = DSI-2338	0.076	j = DSI-2339	0.092	j = EIE-2323	0.006	j = EIE-2329	0.107
j = DSI-2338	0.183	j = DSI-2339	0.100	j = EIE-2323	0.085	j = EIE-2325	0.107	j = EIE-2325	0.090
j = DSI-2339	0.076	j = EIE-2323	0.107	j = EIE-2325	0.025	j = EIE-2329	0.084	j = EIE-2323	0.157

Table E.2. (continued)

i = DSI-2322	d <sub>22,i</sub>	i = DSI-2323	d <sub>23,i</sub>	i = DSI-2324	d <sub>24,i</sub>	i = DSI-2333	d <sub>33,i</sub>	i = DSI-2335	d <sub>35,i</sub>
j = DSI-2323	0.059	j = DSI-2324	0.075	j = DSI-2333	0.017	j = DSI-2335	0.006	j = DSI-2336	0.074
j = DSI-2324	0.017	j = DSI-2333	0.059	j = DSI-2335	0.013	j = DSI-2336	0.079	j = DSI-2337	0.003
j = DSI-2333	0.006	j = DSI-2335	0.064	j = DSI-2336	0.062	j = DSI-2337	0.003	j = DSI-2338	0.102
j = DSI-2335	0.010	j = DSI-2336	0.138	j = DSI-2337	0.016	j = DSI-2338	0.097	j = DSI-2339	0.095
j = DSI-2336	0.079	j = DSI-2337	0.061	j = DSI-2338	0.114	j = DSI-2339	0.095	j = DSI-2333	0.006
j = DSI-2337	0.009	j = DSI-2338	0.039	j = DSI-2339	0.086	j = DSI-2324	0.017	j = DSI-2324	0.013
j = DSI-2338	0.098	j = DSI-2339	0.132	j = DSI-2323	0.075	j = DSI-2323	0.059	j = DSI-2323	0.064
j = DSI-2339	0.089	j = DSI-2322	0.059	j = DSI-2322	0.017	j = DSI-2322	0.006	j = DSI-2322	0.010
j = DSI-2321	0.102	j = DSI-2321	0.091	j = DSI-2321	0.115	j = DSI-2321	0.108	j = DSI-2321	0.113
j = DSI-2313	0.085	j = DSI-2313	0.145	j = DSI-2313	0.070	j = DSI-2313	0.086	j = DSI-2313	0.082
j = EIE-2329	0.005	j = EIE-2329	0.061	j = EIE-2329	0.014	j = EIE-2329	0.003	j = EIE-2329	0.005
j = EIE-2325	0.022	j = EIE-2325	0.038	j = EIE-2325	0.039	j = EIE-2325	0.023	j = EIE-2325	0.029
j = EIE-2323	0.086	j = EIE-2323	0.145	j = EIE-2323	0.071	j = EIE-2323	0.087	j = EIE-2323	0.083

i = DSI-2335	d <sub>35,i</sub>	i = DSI-2337	d <sub>37,i</sub>	i = DSI-2338	d <sub>38,i</sub>	i = DSI-2339	d <sub>39,i</sub>
j = DSI-2337	0.077	j = DSI-2338	0.099	j = DSI-2339	0.165	j = DSI-2338	0.165
j = DSI-2338	0.176	j = DSI-2339	0.096	j = DSI-2337	0.099	j = DSI-2337	0.096
j = DSI-2339	0.089	j = DSI-2336	0.077	j = DSI-2336	0.176	j = DSI-2336	0.089
j = DSI-2335	0.074	j = DSI-2335	0.003	j = DSI-2335	0.102	j = DSI-2335	0.095
j = DSI-2333	0.079	j = DSI-2333	0.003	j = DSI-2333	0.097	j = DSI-2333	0.095
j = DSI-2324	0.062	j = DSI-2324	0.016	j = DSI-2324	0.114	j = DSI-2324	0.086
j = DSI-2323	0.138	j = DSI-2323	0.061	j = DSI-2323	0.039	j = DSI-2323	0.132
j = DSI-2322	0.079	j = DSI-2322	0.009	j = DSI-2322	0.098	j = DSI-2322	0.089
j = DSI-2321	0.161	j = DSI-2321	0.111	j = DSI-2321	0.104	j = DSI-2321	0.099
j = DSI-2313	0.013	j = DSI-2313	0.085	j = DSI-2313	0.183	j = DSI-2313	0.082
j = EIE-2329	0.077	j = EIE-2329	0.005	j = EIE-2329	0.099	j = EIE-2329	0.092
j = EIE-2325	0.101	j = EIE-2325	0.026	j = EIE-2325	0.076	j = EIE-2325	0.100
j = EIE-2323	0.019	j = EIE-2323	0.086	j = EIE-2323	0.183	j = EIE-2323	0.076

- 10% Probability of Exceedance

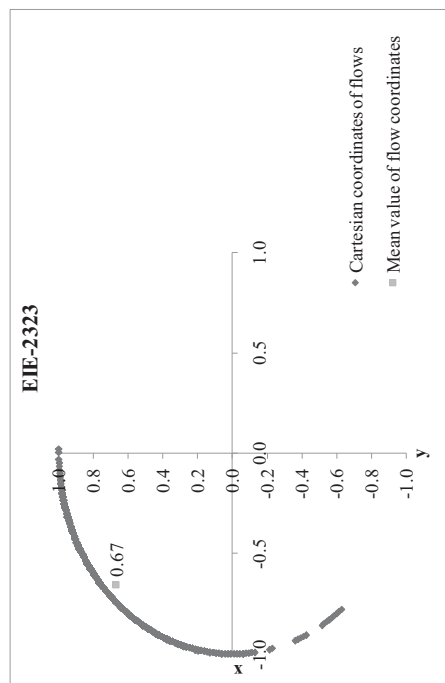


Figure E.2. Seasonality space of station EIE-2323

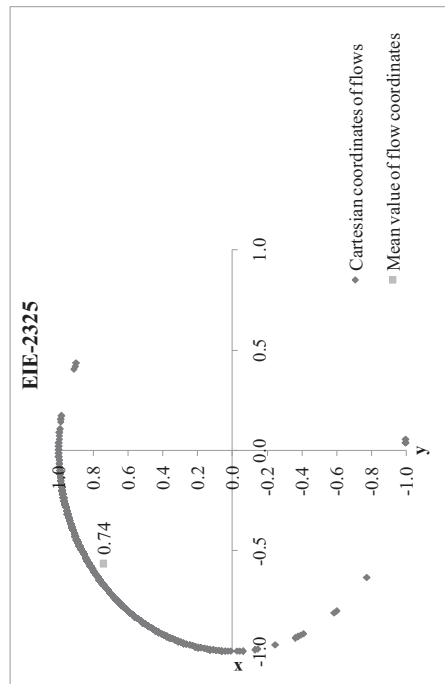


Figure E.3. Seasonality space of station EIE-2325

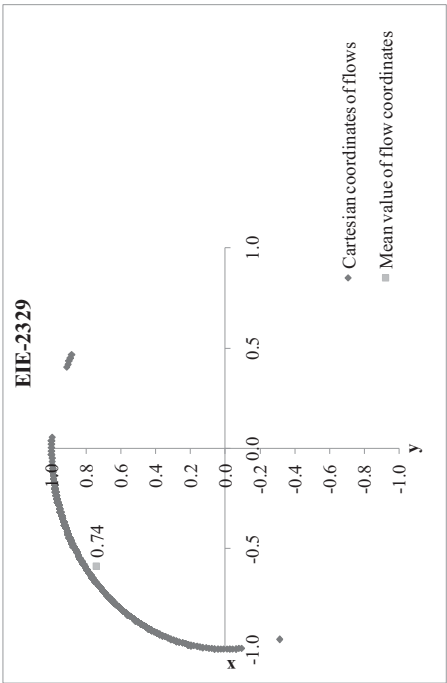


Figure E.4 Seasonality space of station EIE-2329

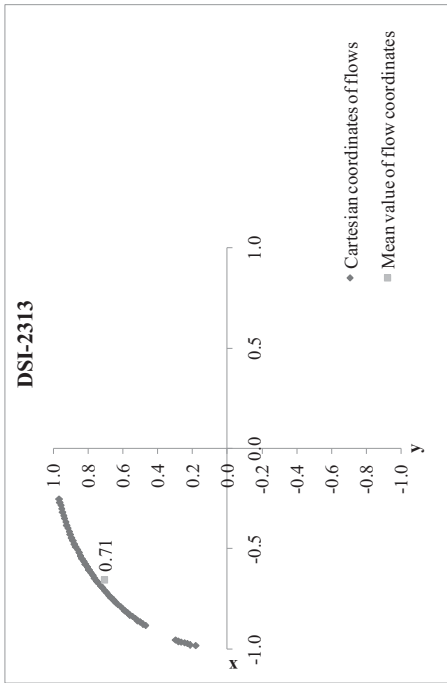


Figure E.5. Seasonality space of station DSI-2313

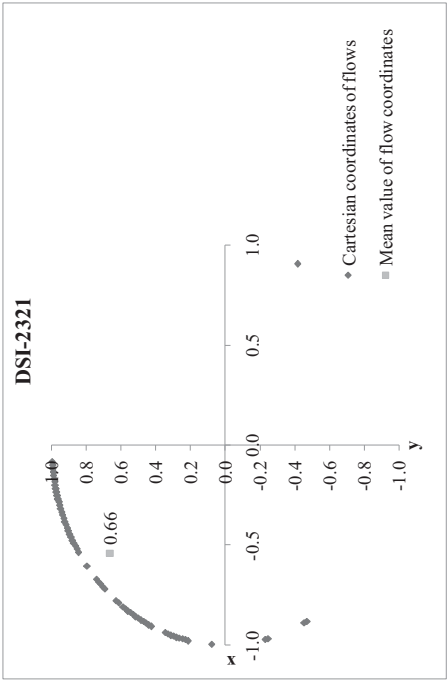


Figure E.6. Seasonality space of station DSI-2321

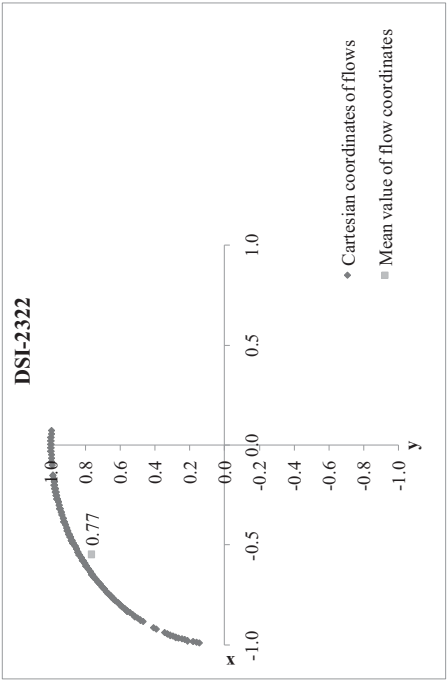


Figure E.7. Seasonality space of station DSI-2322

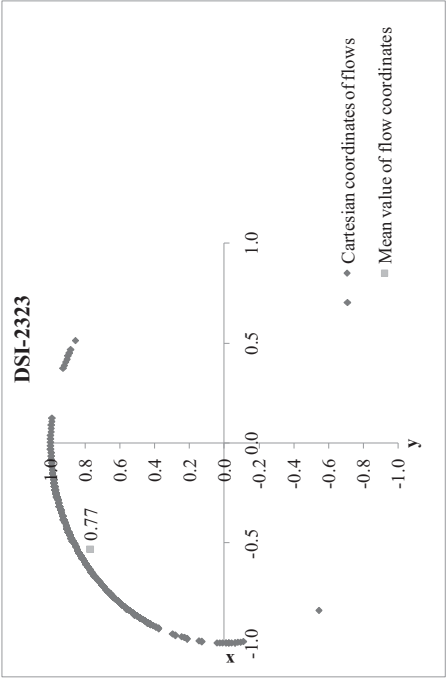


Figure E.8. Seasonality space of station DSI-2323

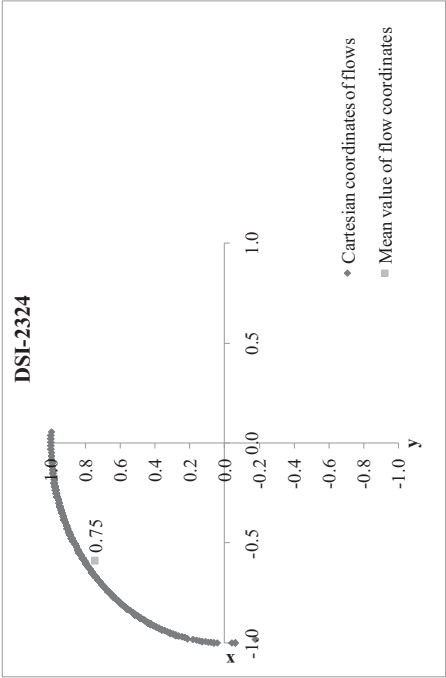


Figure E.9. Seasonality space of station DSI-2324

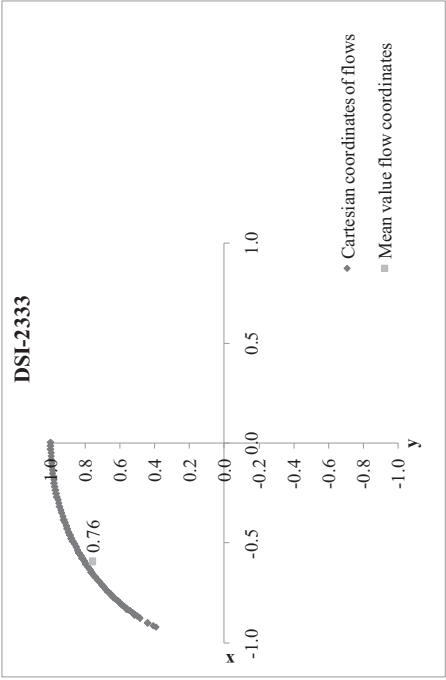


Figure E.10. Seasonality space of station DSI-2333

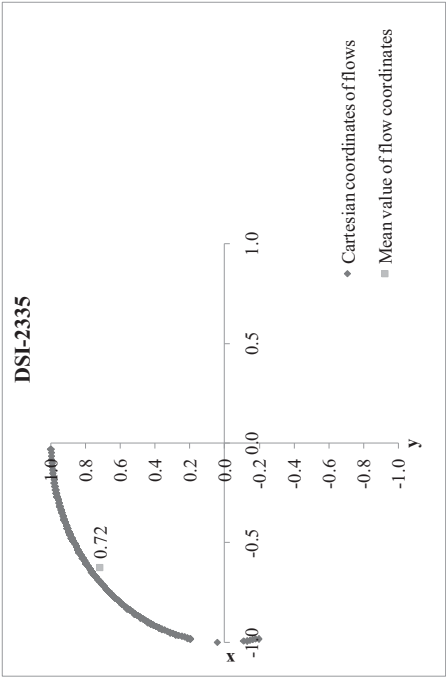


Figure E.11. Seasonality space of station DSI-2335

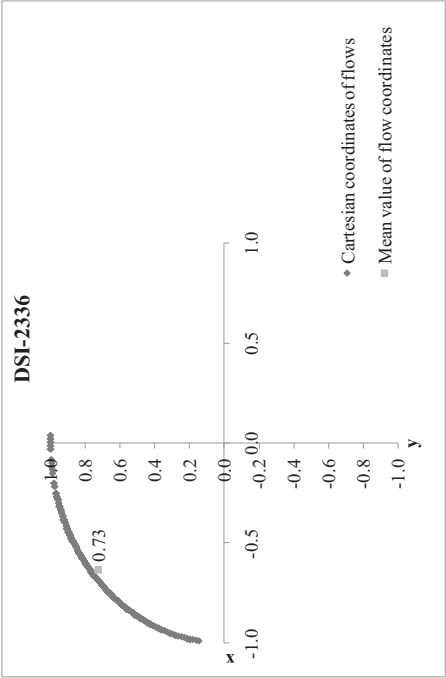


Figure E.12. Seasonality space of station DSI-2336

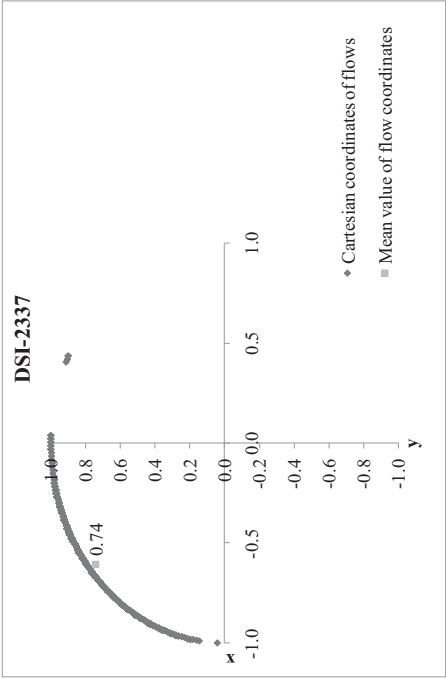


Figure E.13. Seasonality space of station DSI-2337

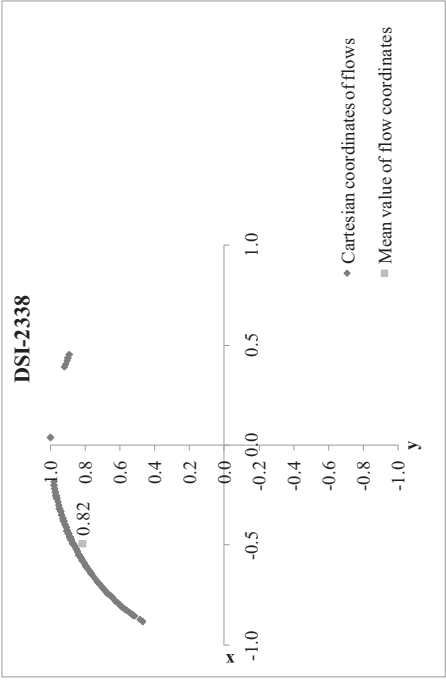


Figure E.14. Seasonality space of station DSI-2338

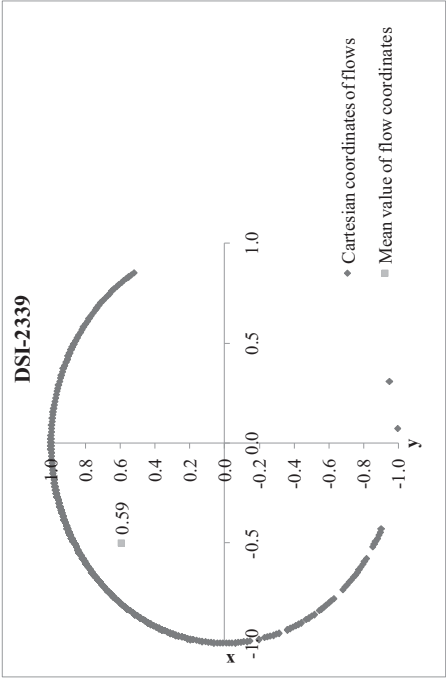


Figure E.15. Seasonality space of station DSI-2339

Table E.3. Basin Euclid distances for 10% probability of exceedance

<b>i = EIE-2323</b>	<b>d<sub>23-i</sub></b>	<b>i = EIE-2325</b>	<b>d<sub>25-i</sub></b>	<b>i = EIE-2329</b>	<b>d<sub>29-i</sub></b>	<b>i = DSI-2313</b>	<b>d<sub>13-i</sub></b>	<b>i = DSI-2321</b>	<b>d<sub>21-i</sub></b>
j = EIE-2325	0.117	j = EIE-2329	0.023	j = DSI-2313	0.078	j = DSI-2321	0.121	j = DSI-2322	0.101
j = EIE-2329	0.099	j = DSI-2313	0.099	j = DSI-2321	0.090	j = DSI-2322	0.125	j = DSI-2323	0.108
j = DSI-2313	0.035	j = DSI-2321	0.081	j = DSI-2322	0.047	j = DSI-2323	0.142	j = DSI-2324	0.094
j = DSI-2321	0.115	j = DSI-2322	0.029	j = DSI-2323	0.064	j = DSI-2324	0.079	j = DSI-2333	0.106
j = DSI-2322	0.145	j = DSI-2323	0.045	j = DSI-2324	0.004	j = DSI-2333	0.083	j = DSI-2335	0.099
j = DSI-2323	0.161	j = DSI-2324	0.025	j = DSI-2333	0.017	j = DSI-2335	0.034	j = DSI-2336	0.112
j = DSI-2324	0.101	j = DSI-2333	0.032	j = DSI-2335	0.044	j = DSI-2336	0.030	j = DSI-2337	0.101
j = DSI-2333	0.109	j = DSI-2335	0.065	j = DSI-2336	0.050	j = DSI-2337	0.061	j = DSI-2338	0.159
j = DSI-2335	0.058	j = DSI-2336	0.073	j = DSI-2337	0.020	j = DSI-2338	0.196	j = DSI-2339	0.081
j = DSI-2336	0.060	j = DSI-2337	0.043	j = DSI-2338	0.119	j = DSI-2339	0.191	j = DSI-2313	0.121
j = DSI-2337	0.087	j = DSI-2338	0.101	j = DSI-2339	0.171	j = EIE-2323	0.035	j = EIE-2329	0.090
j = DSI-2338	0.218	j = DSI-2339	0.161	j = EIE-2323	0.099	j = EIE-2325	0.099	j = EIE-2325	0.081
j = DSI-2339	0.173	j = EIE-2323	0.117	j = EIE-2325	0.023	j = EIE-2329	0.078	j = EIE-2323	0.115



Table E.3. (continued)

<b>i = DSI-2322</b>	<b>d<sub>22-i</sub></b>	<b>i = DSI-2323</b>	<b>d<sub>23-i</sub></b>	<b>i = DSI-2324</b>	<b>d<sub>24-i</sub></b>	<b>i = DSI-2333</b>	<b>d<sub>33-i</sub></b>	<b>i = DSI-2335</b>	<b>d<sub>35-i</sub></b>
j = DSI-2323	0.017	j = DSI-2324	0.063	j = DSI-2333	0.013	j = DSI-2335	0.051	j = DSI-2336	0.013
j = DSI-2324	0.046	j = DSI-2333	0.062	j = DSI-2335	0.045	j = DSI-2336	0.054	j = DSI-2337	0.029
j = DSI-2333	0.045	j = DSI-2335	0.108	j = DSI-2336	0.050	j = DSI-2337	0.022	j = DSI-2338	0.163
j = DSI-2335	0.091	j = DSI-2336	0.114	j = DSI-2337	0.019	j = DSI-2338	0.113	j = DSI-2339	0.175
j = DSI-2336	0.096	j = DSI-2337	0.082	j = DSI-2338	0.118	j = DSI-2339	0.187	j = DSI-2333	0.051
j = DSI-2337	0.065	j = DSI-2338	0.057	j = DSI-2339	0.175	j = DSI-2324	0.013	j = DSI-2324	0.045
j = DSI-2338	0.073	j = DSI-2339	0.181	j = DSI-2323	0.063	j = DSI-2323	0.062	j = DSI-2323	0.108
j = DSI-2339	0.177	j = DSI-2322	0.017	j = DSI-2322	0.046	j = DSI-2322	0.045	j = DSI-2322	0.091
j = DSI-2321	0.101	j = DSI-2321	0.108	j = DSI-2321	0.094	j = DSI-2321	0.106	j = DSI-2321	0.099
j = DSI-2313	0.125	j = DSI-2313	0.142	j = DSI-2313	0.079	j = DSI-2313	0.083	j = DSI-2313	0.034
j = EIE-2329	0.047	j = EIE-2329	0.064	j = EIE-2329	0.004	j = EIE-2329	0.017	j = EIE-2329	0.044
j = EIE-2325	0.029	j = EIE-2325	0.045	j = EIE-2325	0.025	j = EIE-2325	0.032	j = EIE-2325	0.065
j = EIE-2323	0.145	j = EIE-2323	0.161	j = EIE-2323	0.101	j = EIE-2323	0.109	j = EIE-2323	0.058

<b>i = DSI-2335</b>	<b>d<sub>36-i</sub></b>	<b>i = DSI-2337</b>	<b>d<sub>37-i</sub></b>	<b>i = DSI-2338</b>	<b>d<sub>38-i</sub></b>	<b>i = DSI-2339</b>	<b>d<sub>39-i</sub></b>
j = DSI-2337	0.032	j = DSI-2338	0.135	j = DSI-2339	0.222	j = DSI-2338	0.222
j = DSI-2338	0.167	j = DSI-2339	0.182	j = DSI-2337	0.135	j = DSI-2337	0.182
j = DSI-2339	0.188	j = DSI-2336	0.032	j = DSI-2336	0.167	j = DSI-2336	0.188
j = DSI-2335	0.013	j = DSI-2335	0.029	j = DSI-2335	0.163	j = DSI-2335	0.175
j = DSI-2333	0.054	j = DSI-2333	0.022	j = DSI-2333	0.113	j = DSI-2333	0.187
j = DSI-2324	0.050	j = DSI-2324	0.019	j = DSI-2324	0.118	j = DSI-2324	0.175
j = DSI-2323	0.114	j = DSI-2323	0.082	j = DSI-2323	0.057	j = DSI-2323	0.181
j = DSI-2322	0.096	j = DSI-2322	0.065	j = DSI-2322	0.073	j = DSI-2322	0.177
j = DSI-2321	0.112	j = DSI-2321	0.101	j = DSI-2321	0.159	j = DSI-2321	0.081
j = DSI-2313	0.030	j = DSI-2313	0.061	j = DSI-2313	0.196	j = DSI-2313	0.191
j = EIE-2329	0.050	j = EIE-2329	0.020	j = EIE-2329	0.119	j = EIE-2329	0.171
j = EIE-2325	0.073	j = EIE-2325	0.043	j = EIE-2325	0.101	j = EIE-2325	0.161
j = EIE-2323	0.060	j = EIE-2323	0.087	j = EIE-2323	0.218	j = EIE-2323	0.173

- 20% Probability of Exceedance

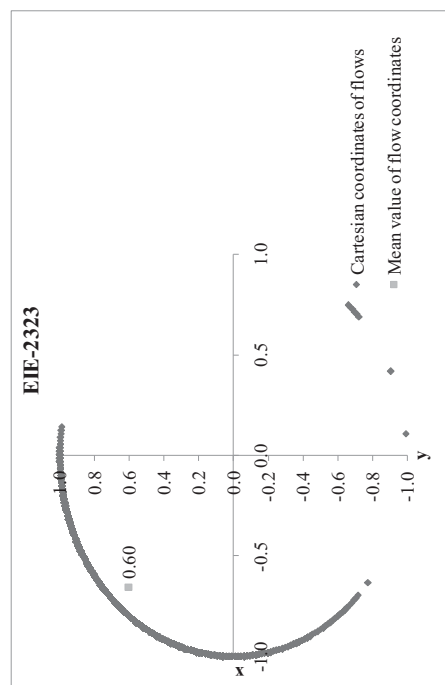


Figure E.16. Seasonality space of station EIE-2323

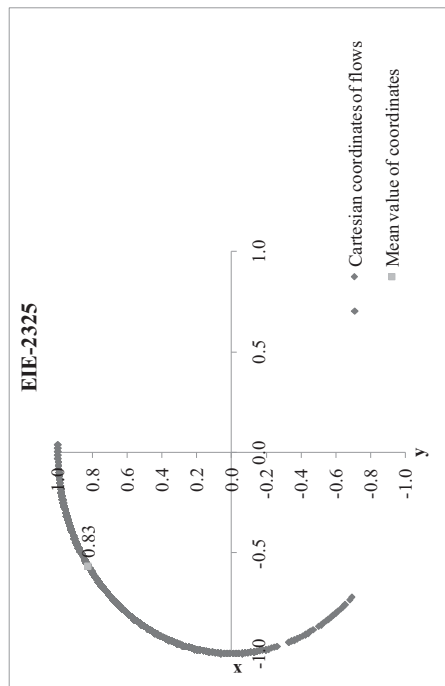


Figure E.17. Seasonality space of station EIE-2325

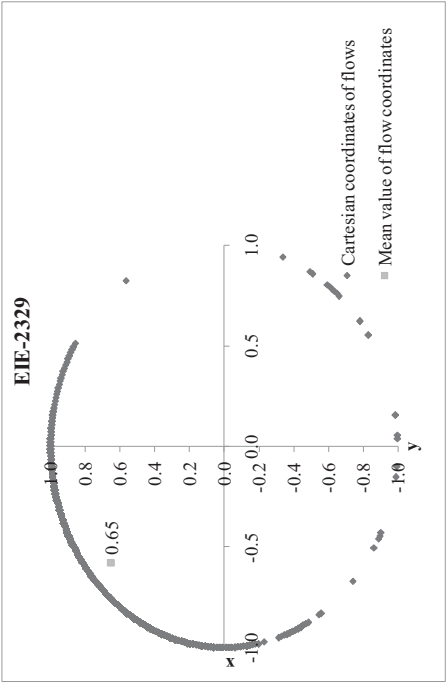


Figure E.18. Seasonality space of station EIE-2329

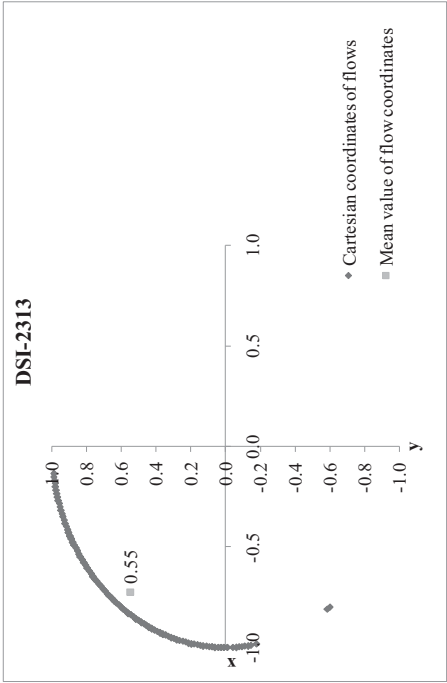


Figure E.19. Seasonality space of station DSI-2313

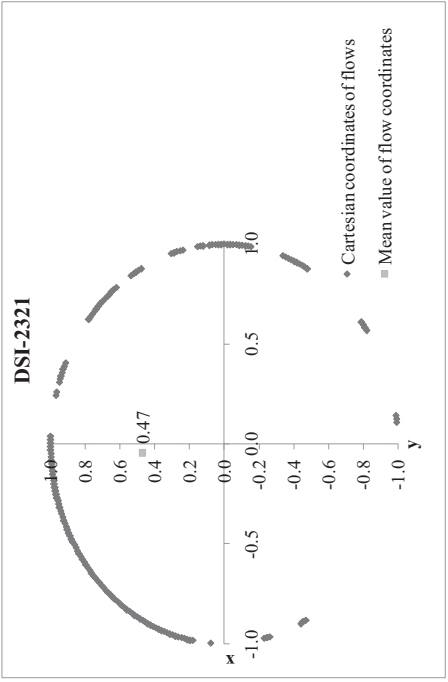


Figure E.20. Seasonality space of station DSI-2321

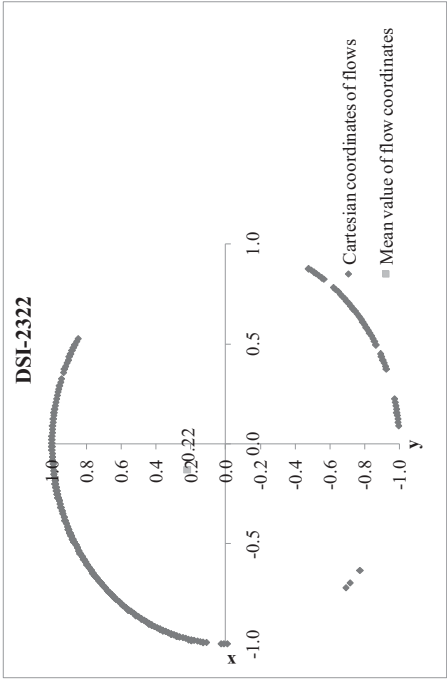


Figure E.21. Seasonality space of station DSI-2322

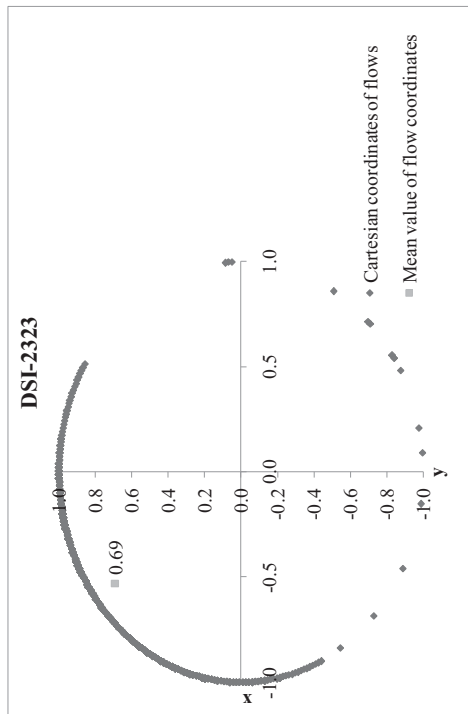


Figure E.22. Seasonality space of station DSI-2323

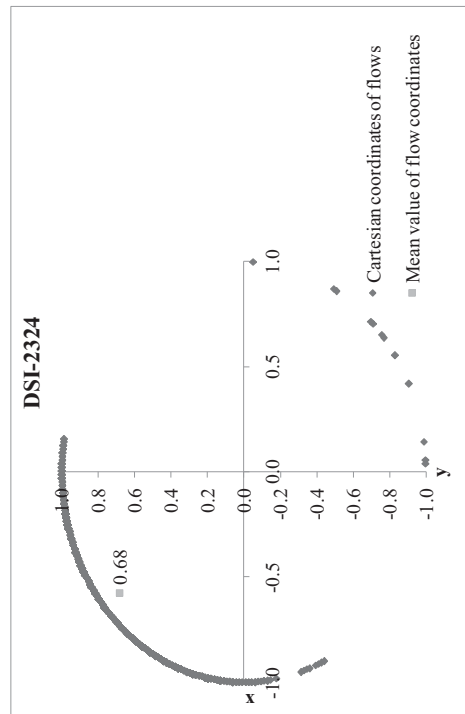


Figure E.23. Seasonality space of station DSI-2324

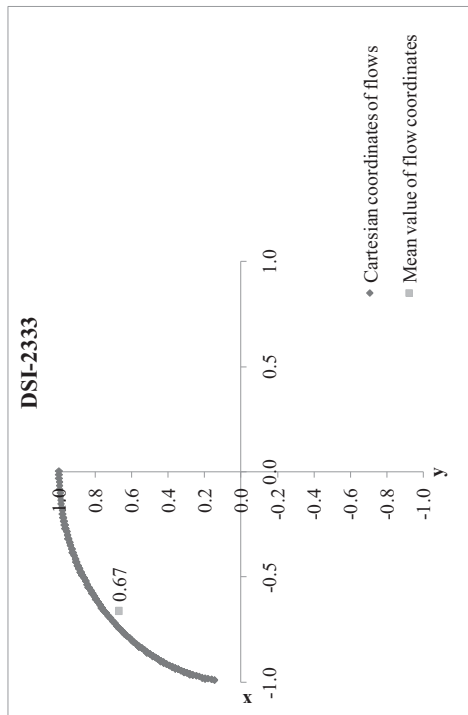


Figure E.24. Seasonality space of station DSI-2333

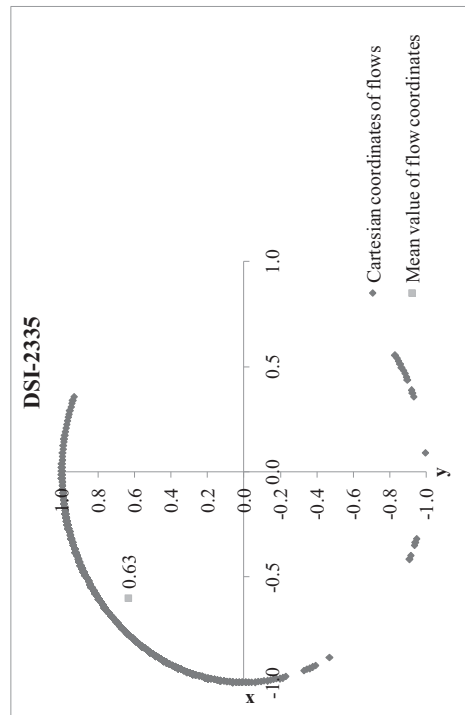


Figure E.25. Seasonality space of station DSI-2335

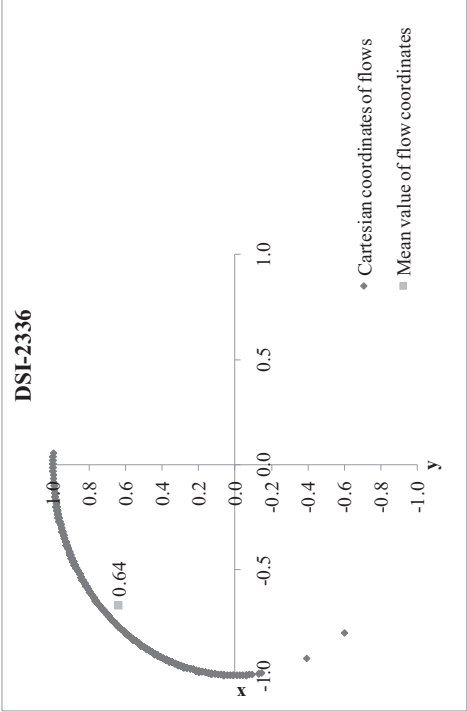


Figure E.26. Seasonality space of station DSI-2336

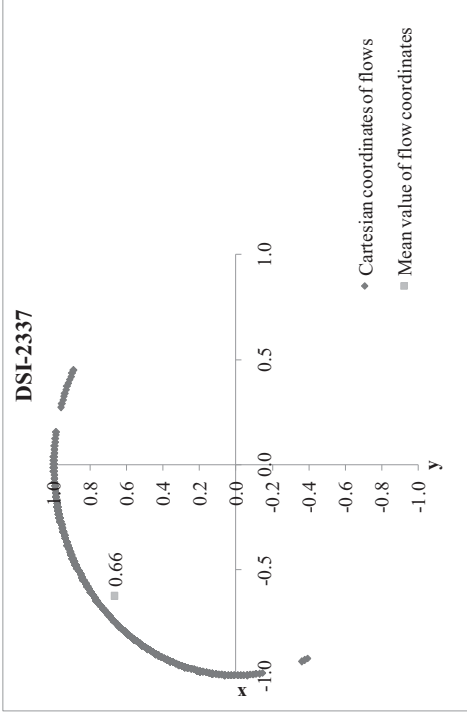


Figure E.27. Seasonality space of station DSI-2337

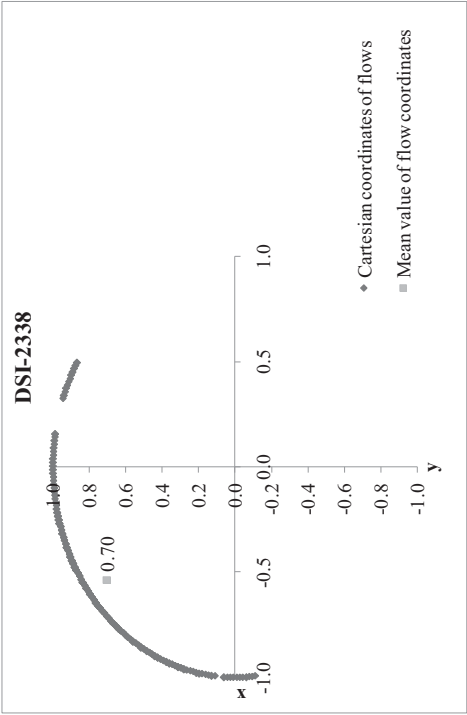


Figure E.28. Seasonality space of station DSI-2338

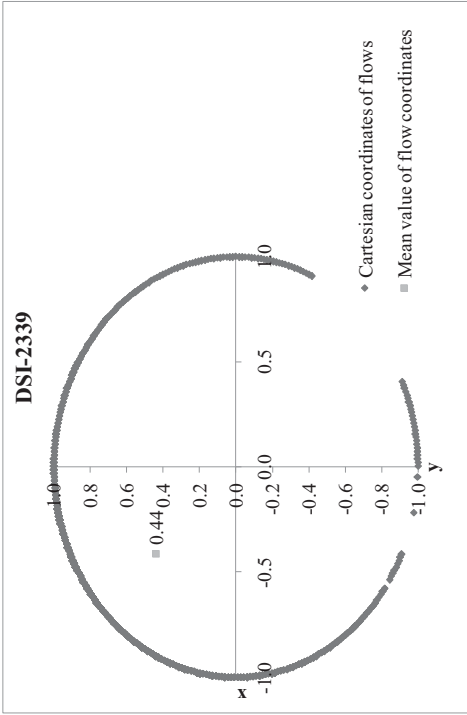


Figure E.29. Seasonality space of station DSI-2339

Table E.4. Basin Euclid distances for 20% probability of exceedance

i = EIE-2323	<b>d<sub>23-j</sub></b>	i = EIE-2325	<b>d<sub>25-j</sub></b>	i = EIE-2329	<b>d<sub>29-j</sub></b>	i = DSI-2313	<b>d<sub>13-j</sub></b>	i = DSI-2321	<b>d<sub>21-j</sub></b>
j = EIE-2325	0.135	j = EIE-2329	0.053	j = DSI-2313	0.178	j = DSI-2321	0.684	j = DSI-2322	0.264
j = EIE-2329	0.091	j = DSI-2313	0.218	j = DSI-2321	0.564	j = DSI-2322	0.679	j = DSI-2323	0.535
j = DSI-2313	0.088	j = DSI-2321	0.512	j = DSI-2322	0.623	j = DSI-2323	0.240	j = DSI-2324	0.573
j = DSI-2321	0.626	j = DSI-2322	0.579	j = DSI-2323	0.062	j = DSI-2324	0.200	j = DSI-2333	0.649
j = DSI-2322	0.652	j = DSI-2323	0.050	j = DSI-2324	0.032	j = DSI-2333	0.137	j = DSI-2335	0.579
j = DSI-2323	0.152	j = DSI-2324	0.065	j = DSI-2333	0.085	j = DSI-2335	0.151	j = DSI-2336	0.646
j = DSI-2324	0.112	j = DSI-2333	0.137	j = DSI-2335	0.027	j = DSI-2336	0.108	j = DSI-2337	0.610
j = DSI-2333	0.067	j = DSI-2335	0.073	j = DSI-2336	0.089	j = DSI-2337	0.154	j = DSI-2338	0.547
j = DSI-2335	0.064	j = DSI-2336	0.141	j = DSI-2337	0.046	j = DSI-2338	0.241	j = DSI-2339	0.370
j = DSI-2336	0.038	j = DSI-2337	0.098	j = DSI-2338	0.066	j = DSI-2339	0.330	j = DSI-2313	0.684
j = DSI-2337	0.069	j = DSI-2338	0.063	j = DSI-2339	0.270	j = EIE-2323	0.088	j = EIE-2329	0.564
j = DSI-2338	0.154	j = DSI-2339	0.233	j = EIE-2323	0.091	j = EIE-2325	0.218	j = EIE-2325	0.512
j = DSI-2339	0.294	j = EIE-2323	0.135	j = EIE-2325	0.053	j = EIE-2329	0.178	j = EIE-2323	0.626

Table E.4. (continued)

<b>i = DSI-2322</b>	<b>d<sub>22-i</sub></b>	<b>i = DSI-2323</b>	<b>d<sub>23-i</sub></b>	<b>i = DSI-2324</b>	<b>d<sub>24-i</sub></b>	<b>i = DSI-2333</b>	<b>d<sub>33-i</sub></b>	<b>i = DSI-2335</b>	<b>d<sub>35-i</sub></b>
j = DSI-2323	0.620	j = DSI-2324	0.046	j = DSI-2333	0.086	j = DSI-2335	0.071	j = DSI-2336	0.068
j = DSI-2324	0.644	j = DSI-2333	0.131	j = DSI-2335	0.054	j = DSI-2336	0.031	j = DSI-2337	0.038
j = DSI-2333	0.697	j = DSI-2335	0.089	j = DSI-2336	0.100	j = DSI-2337	0.039	j = DSI-2338	0.092
j = DSI-2335	0.627	j = DSI-2336	0.145	j = DSI-2337	0.050	j = DSI-2338	0.127	j = DSI-2339	0.270
j = DSI-2336	0.683	j = DSI-2337	0.095	j = DSI-2338	0.043	j = DSI-2339	0.340	j = DSI-2333	0.071
j = DSI-2337	0.664	j = DSI-2338	0.014	j = DSI-2339	0.294	j = DSI-2324	0.086	j = DSI-2324	0.054
j = DSI-2338	0.633	j = DSI-2339	0.279	j = DSI-2323	0.046	j = DSI-2323	0.131	j = DSI-2323	0.089
j = DSI-2339	0.358	j = DSI-2322	0.620	j = DSI-2322	0.644	j = DSI-2322	0.697	j = DSI-2322	0.627
j = DSI-2321	0.264	j = DSI-2321	0.535	j = DSI-2321	0.573	j = DSI-2321	0.649	j = DSI-2321	0.579
j = DSI-2313	0.679	j = DSI-2313	0.240	j = DSI-2313	0.200	j = DSI-2313	0.137	j = DSI-2313	0.151
j = EIE-2329	0.623	j = EIE-2329	0.062	j = EIE-2329	0.032	j = EIE-2329	0.085	j = EIE-2329	0.027
j = EIE-2325	0.579	j = EIE-2325	0.050	j = EIE-2325	0.065	j = EIE-2325	0.137	j = EIE-2325	0.073
j = EIE-2323	0.652	j = EIE-2323	0.152	j = EIE-2323	0.112	j = EIE-2323	0.067	j = EIE-2323	0.064

<b>i = DSI-2335</b>	<b>d<sub>36-i</sub></b>	<b>i = DSI-2337</b>	<b>d<sub>37-i</sub></b>	<b>i = DSI-2338</b>	<b>d<sub>38-i</sub></b>	<b>i = DSI-2339</b>	<b>d<sub>39-i</sub></b>
j = DSI-2337	0.051	j = DSI-2338	0.092	j = DSI-2339	0.293	j = DSI-2338	0.293
j = DSI-2338	0.143	j = DSI-2339	0.308	j = DSI-2337	0.092	j = DSI-2337	0.308
j = DSI-2339	0.325	j = DSI-2336	0.051	j = DSI-2336	0.143	j = DSI-2336	0.325
j = DSI-2335	0.068	j = DSI-2335	0.038	j = DSI-2335	0.092	j = DSI-2335	0.270
j = DSI-2333	0.031	j = DSI-2333	0.039	j = DSI-2333	0.127	j = DSI-2333	0.340
j = DSI-2324	0.100	j = DSI-2324	0.050	j = DSI-2324	0.043	j = DSI-2324	0.294
j = DSI-2323	0.145	j = DSI-2323	0.095	j = DSI-2323	0.014	j = DSI-2323	0.279
j = DSI-2322	0.683	j = DSI-2322	0.664	j = DSI-2322	0.633	j = DSI-2322	0.358
j = DSI-2321	0.646	j = DSI-2321	0.610	j = DSI-2321	0.547	j = DSI-2321	0.370
j = DSI-2313	0.108	j = DSI-2313	0.154	j = DSI-2313	0.241	j = DSI-2313	0.330
j = EIE-2329	0.089	j = EIE-2329	0.046	j = EIE-2329	0.066	j = EIE-2329	0.270
j = EIE-2325	0.141	j = EIE-2325	0.098	j = EIE-2325	0.063	j = EIE-2325	0.233
j = EIE-2323	0.038	j = EIE-2323	0.069	j = EIE-2323	0.154	j = EIE-2323	0.294

- 30% Probability of Exceedance

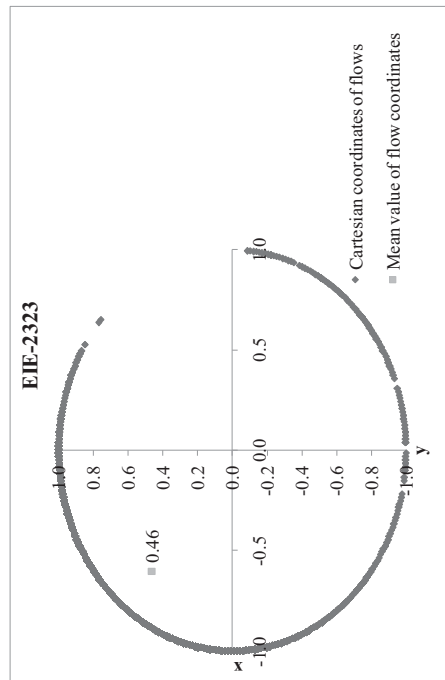


Figure E.30. Seasonality space of station EIE-2323

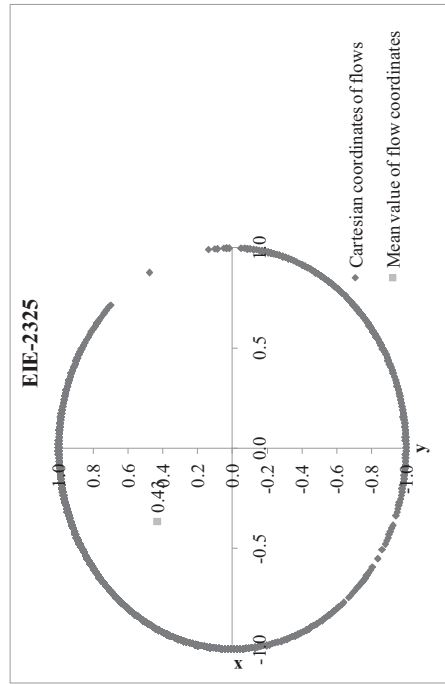


Figure E.31. Seasonality space of station EIE-2325



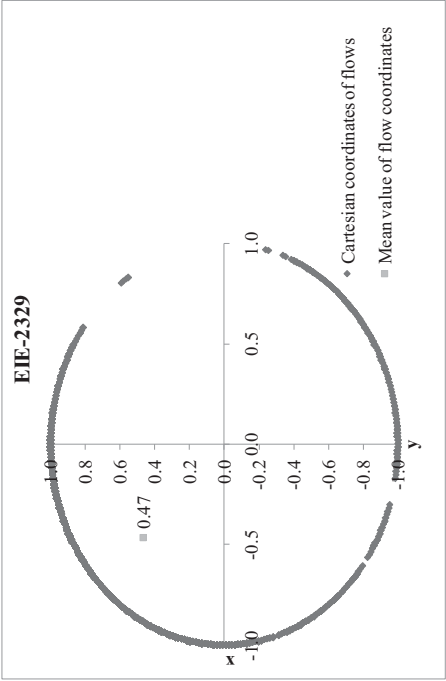


Figure E.32. Seasonality space of station EIE-2329

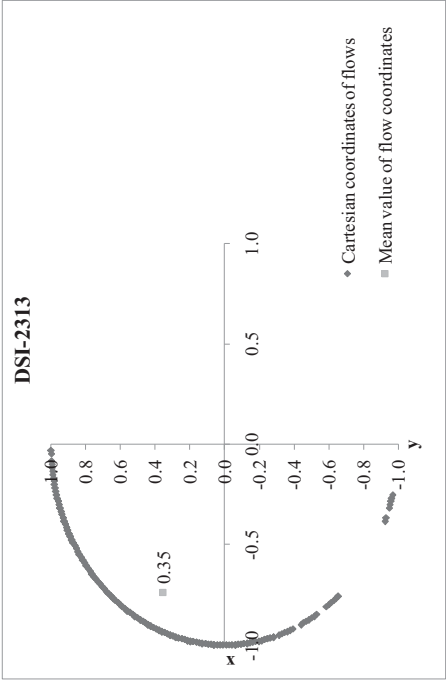


Figure E.33. Seasonality space of station DSI-2313

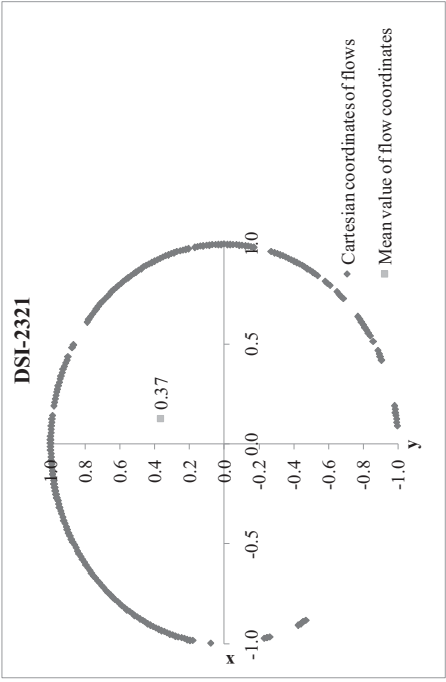


Figure E.34. Seasonality space of station DSI-2321

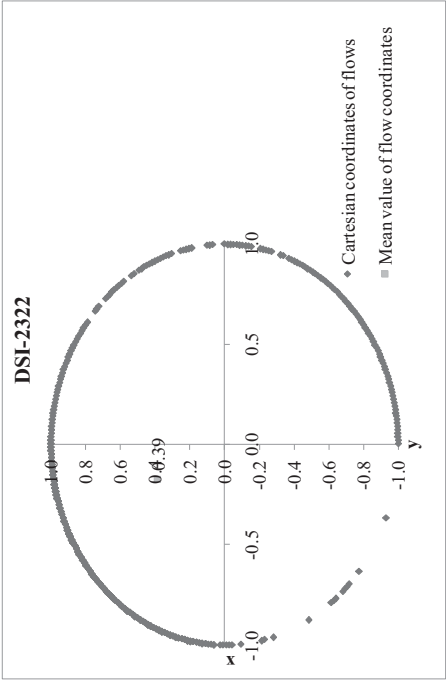


Figure E.35. Seasonality space of station DSI-2322

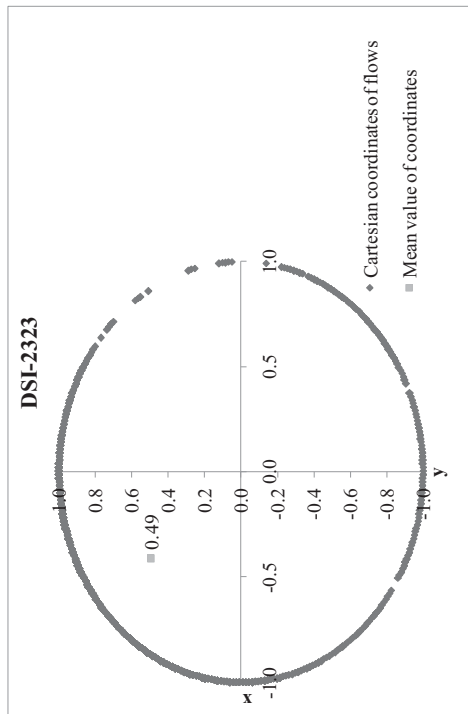


Figure E.36. Seasonality space of station DSI-2323

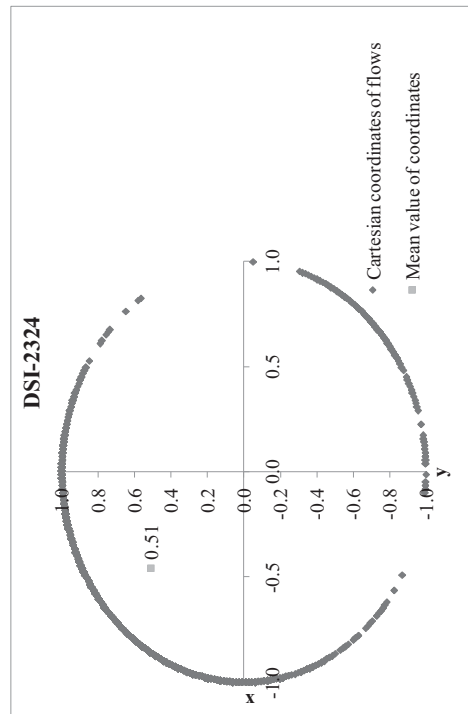


Figure E.37. Seasonality space of station DSI-2324

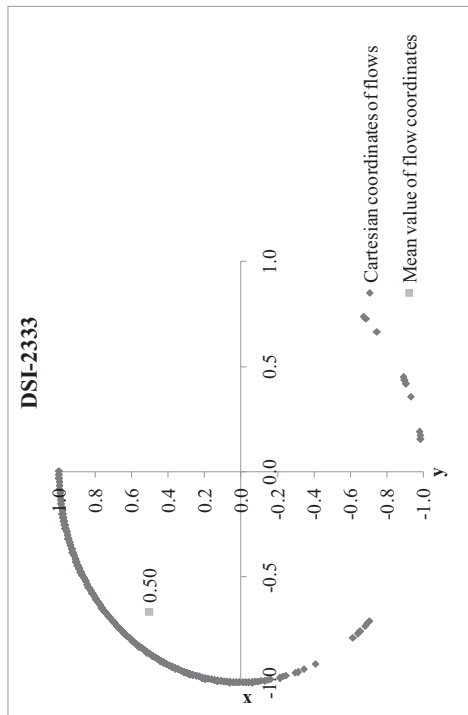


Figure E.38. Seasonality space of station DSI-2333

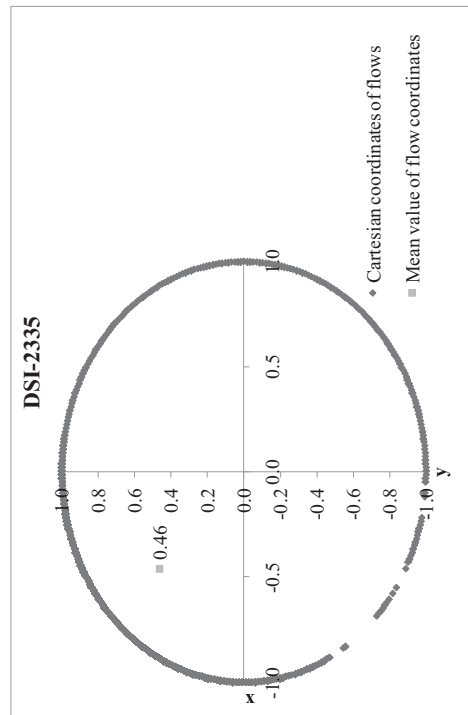


Figure E.39. Seasonality space of station DSI-2335

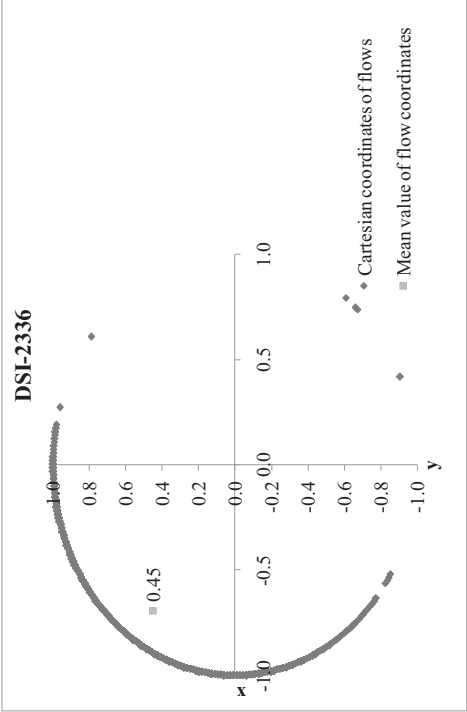


Figure E.40. Seasonality space of station DSI-2336

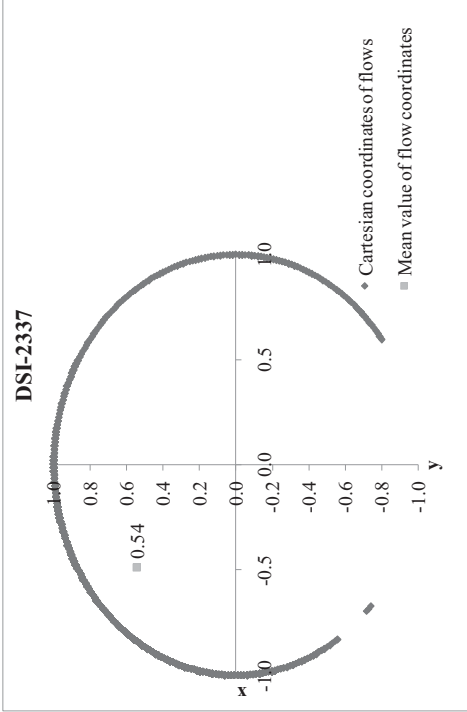


Figure E.41. Seasonality space of station DSI-2337

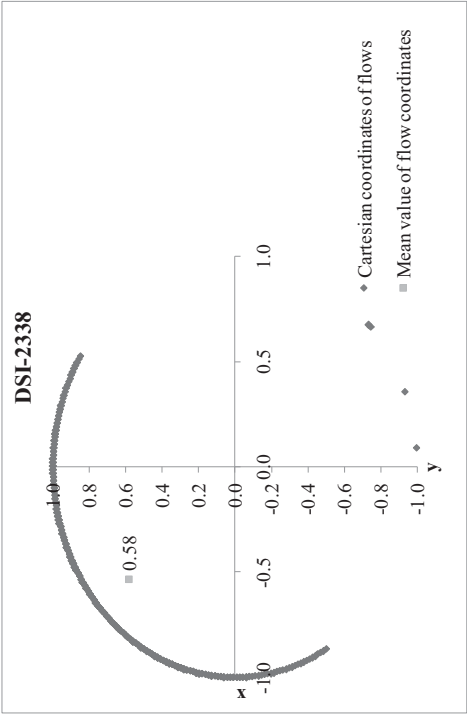


Figure E.42. Seasonality space of station DSI-2338

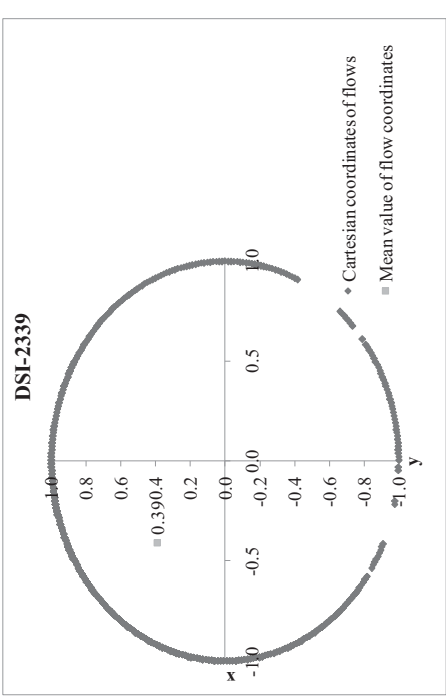


Figure E.43. Seasonality space of station DSI-2339

Table E.5. Basin Euclid distances for 30% probability of exceedance

<b>i = EIE-2323</b>	<b>d<sub>23-j</sub></b>	<b>i = EIE-2325</b>	<b>d<sub>25-j</sub></b>	<b>i = EIE-2329</b>	<b>d<sub>29-j</sub></b>	<b>i = DSI-2313</b>	<b>d<sub>13-j</sub></b>	<b>i = DSI-2321</b>	<b>d<sub>21-j</sub></b>
j = EIE-2325	0.241	j = EIE-2329	0.105	j = DSI-2313	0.296	j = DSI-2321	0.866	j = DSI-2322	0.302
j = EIE-2329	0.139	j = DSI-2313	0.382	j = DSI-2321	0.600	j = DSI-2322	0.566	j = DSI-2323	0.554
j = DSI-2313	0.175	j = DSI-2321	0.497	j = DSI-2322	0.302	j = DSI-2323	0.356	j = DSI-2324	0.604
j = DSI-2321	0.738	j = DSI-2322	0.197	j = DSI-2323	0.060	j = DSI-2324	0.321	j = DSI-2333	0.805
j = DSI-2322	0.438	j = DSI-2323	0.077	j = DSI-2324	0.045	j = DSI-2333	0.168	j = DSI-2335	0.598
j = DSI-2323	0.195	j = DSI-2324	0.122	j = DSI-2333	0.205	j = DSI-2335	0.297	j = DSI-2336	0.826
j = DSI-2324	0.152	j = DSI-2333	0.310	j = DSI-2335	0.004	j = DSI-2336	0.107	j = DSI-2337	0.641
j = DSI-2333	0.074	j = DSI-2335	0.102	j = DSI-2336	0.231	j = DSI-2337	0.315	j = DSI-2338	0.697
j = DSI-2335	0.142	j = DSI-2336	0.330	j = DSI-2337	0.082	j = DSI-2338	0.307	j = DSI-2339	0.538
j = DSI-2336	0.092	j = DSI-2337	0.167	j = DSI-2338	0.137	j = DSI-2339	0.331	j = DSI-2313	0.866
j = DSI-2337	0.141	j = DSI-2338	0.227	j = DSI-2339	0.094	j = EIE-2323	0.175	j = EIE-2329	0.600
j = DSI-2338	0.138	j = DSI-2339	0.061	j = EIE-2323	0.139	j = EIE-2325	0.382	j = EIE-2325	0.497
j = DSI-2339	0.208	j = EIE-2323	0.241	j = EIE-2325	0.105	j = EIE-2329	0.296	j = EIE-2323	0.738

Table E.5. (continued)

<b>i = DSI-2322</b>	<b>d<sub>22-j</sub></b>	<b>i = DSI-2323</b>	<b>d<sub>23-j</sub></b>	<b>i = DSI-2324</b>	<b>d<sub>24-j</sub></b>	<b>i = DSI-2333</b>	<b>d<sub>33-j</sub></b>	<b>i = DSI-2335</b>	<b>d<sub>35-j</sub></b>
j = DSI-2323	0.261	j = DSI-2324	0.050	j = DSI-2333	0.207	j = DSI-2335	0.208	j = DSI-2336	0.233
j = DSI-2324	0.311	j = DSI-2333	0.255	j = DSI-2335	0.047	j = DSI-2336	0.061	j = DSI-2337	0.085
j = DSI-2333	0.506	j = DSI-2335	0.059	j = DSI-2336	0.243	j = DSI-2337	0.182	j = DSI-2338	0.140
j = DSI-2335	0.299	j = DSI-2336	0.286	j = DSI-2337	0.045	j = DSI-2338	0.153	j = DSI-2339	0.090
j = DSI-2336	0.525	j = DSI-2337	0.092	j = DSI-2338	0.105	j = DSI-2339	0.280	j = DSI-2333	0.208
j = DSI-2337	0.352	j = DSI-2338	0.152	j = DSI-2339	0.130	j = DSI-2324	0.207	j = DSI-2324	0.047
j = DSI-2338	0.412	j = DSI-2339	0.103	j = DSI-2323	0.050	j = DSI-2323	0.255	j = DSI-2323	0.059
j = DSI-2339	0.236	j = DSI-2322	0.261	j = DSI-2322	0.311	j = DSI-2322	0.506	j = DSI-2322	0.299
j = DSI-2321	0.302	j = DSI-2321	0.554	j = DSI-2321	0.604	j = DSI-2321	0.805	j = DSI-2321	0.598
j = DSI-2313	0.566	j = DSI-2313	0.356	j = DSI-2313	0.321	j = DSI-2313	0.168	j = DSI-2313	0.297
j = EIE-2329	0.302	j = EIE-2329	0.060	j = EIE-2329	0.045	j = EIE-2329	0.205	j = EIE-2329	0.004
j = EIE-2325	0.197	j = EIE-2325	0.077	j = EIE-2325	0.122	j = EIE-2325	0.310	j = EIE-2325	0.102
j = EIE-2323	0.438	j = EIE-2323	0.195	j = EIE-2323	0.152	j = EIE-2323	0.074	j = EIE-2323	0.142

<b>i = DSI-2335</b>	<b>d<sub>35-j</sub></b>	<b>i = DSI-2337</b>	<b>d<sub>37-j</sub></b>	<b>i = DSI-2338</b>	<b>d<sub>38-j</sub></b>	<b>i = DSI-2339</b>	<b>d<sub>39-j</sub></b>
j = DSI-2337	0.227	j = DSI-2338	0.061	j = DSI-2339	0.230	j = DSI-2338	0.230
j = DSI-2338	0.208	j = DSI-2339	0.173	j = DSI-2337	0.061	j = DSI-2337	0.173
j = DSI-2339	0.291	j = DSI-2336	0.227	j = DSI-2336	0.208	j = DSI-2336	0.291
j = DSI-2335	0.233	j = DSI-2335	0.085	j = DSI-2335	0.140	j = DSI-2335	0.090
j = DSI-2333	0.061	j = DSI-2333	0.182	j = DSI-2333	0.153	j = DSI-2333	0.280
j = DSI-2324	0.243	j = DSI-2324	0.045	j = DSI-2324	0.105	j = DSI-2324	0.130
j = DSI-2323	0.286	j = DSI-2323	0.092	j = DSI-2323	0.152	j = DSI-2323	0.103
j = DSI-2322	0.525	j = DSI-2322	0.352	j = DSI-2322	0.412	j = DSI-2322	0.236
j = DSI-2321	0.826	j = DSI-2321	0.641	j = DSI-2321	0.697	j = DSI-2321	0.538
j = DSI-2313	0.107	j = DSI-2313	0.315	j = DSI-2313	0.307	j = DSI-2313	0.331
j = EIE-2329	0.231	j = EIE-2329	0.082	j = EIE-2329	0.137	j = EIE-2329	0.094
j = EIE-2325	0.330	j = EIE-2325	0.167	j = EIE-2325	0.227	j = EIE-2325	0.061
j = EIE-2323	0.092	j = EIE-2323	0.141	j = EIE-2323	0.138	j = EIE-2323	0.208

- PERIMETER (P)

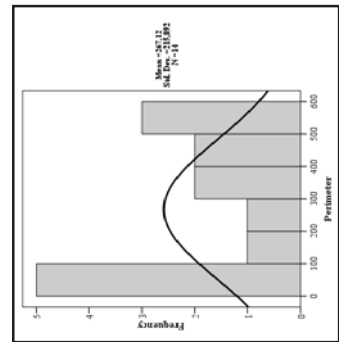


Figure F.1. Histogram and normal line of perimeter

- AREA (A)

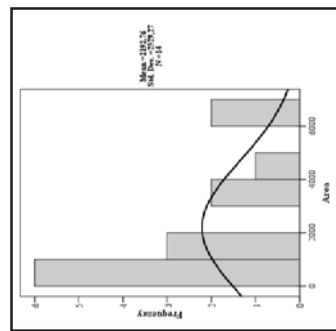


Figure F.3. Histogram and normal line of “area”

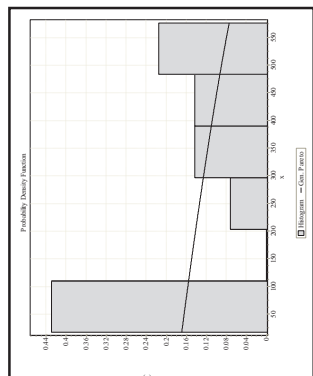


Figure F.2. Histogram and probability distribution function of “perimeter”

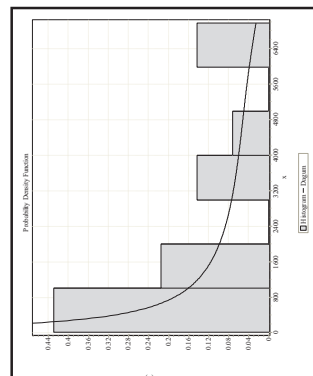


Figure F.4. Probability distribution function of “area”

# APPENDIX F

## DESCRIPTIVE STATISTICS

- **LENGTH OF MAIN RIVER (LMR)**

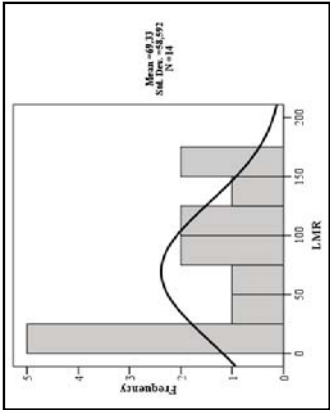


Figure F.5. Histogram and normal line of “main river length”

- **MAXIMUM ELEVATION (H<sub>max</sub>)**

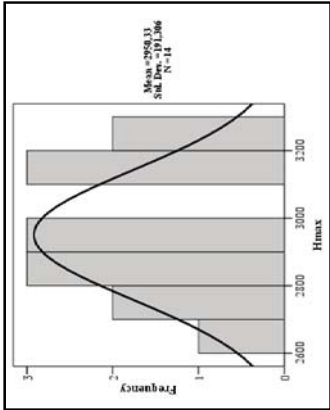


Figure F.7. Histogram and normal line of “maximum elevation”

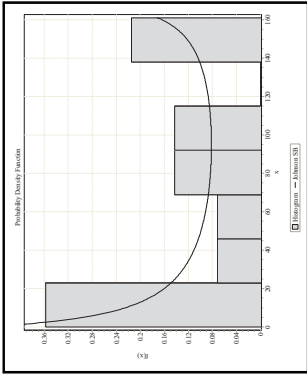


Figure F.6. Probability distribution function of “main river length”

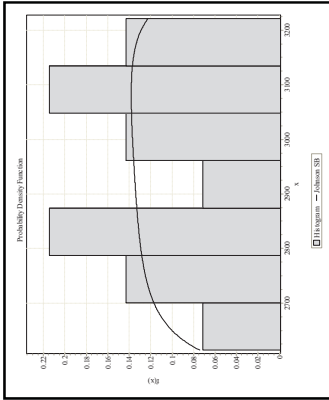


Figure F.8. Probability distribution function of “maximum elevation”

- **MINIMUM ELEVATION ( $H_{\min}$ )**

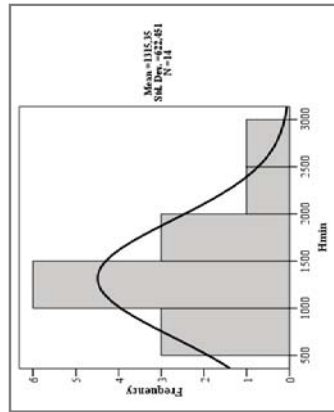


Figure F.9. Histogram and normal line of “minimum elevation”

- **MEAN ELEVATION ( $H_{\text{mean}}$ )**

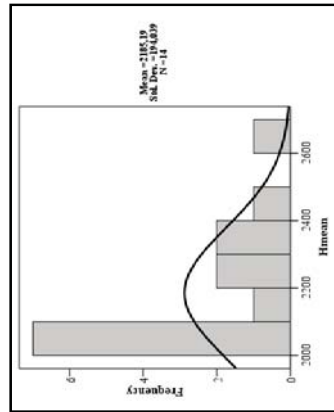


Figure F.11. Histogram and normal line of “mean elevation”

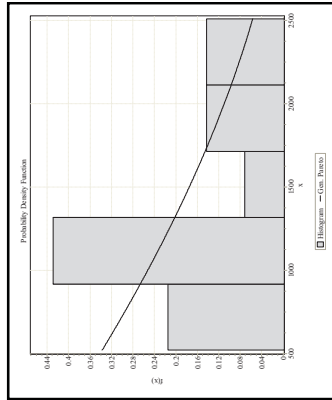


Figure F.10. Probability distribution function of “minimum elevation”

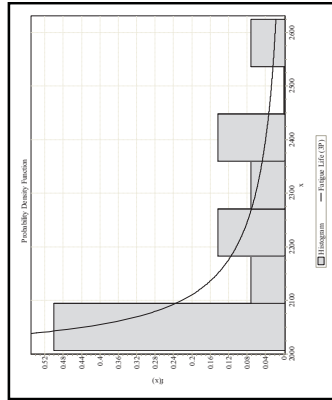


Figure F.12. Probability distribution function of “mean elevation”



- **BASIN RELIEF (BR)**

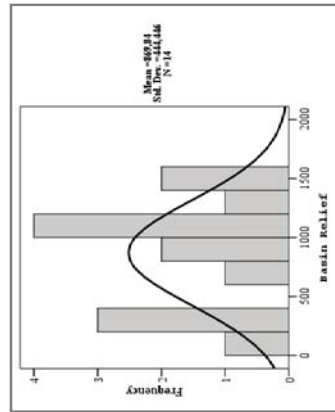


Figure F.13. Histogram and normal line of “basin relief”

- **SLOPE (S)**

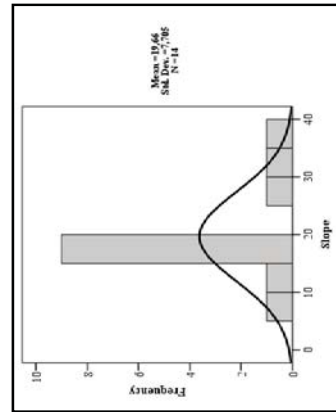


Figure F.15. Histogram and normal line of “slope”

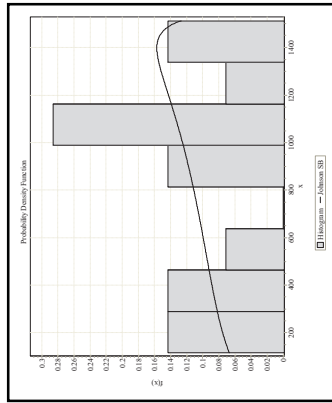


Figure F.14. Probability distribution function of “basin relief”

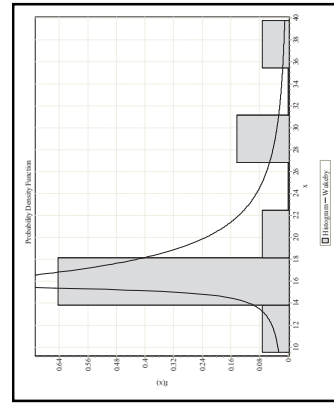


Figure F.16. Probability distribution function of “slope”

- **ASPECT (ASPCT.)**

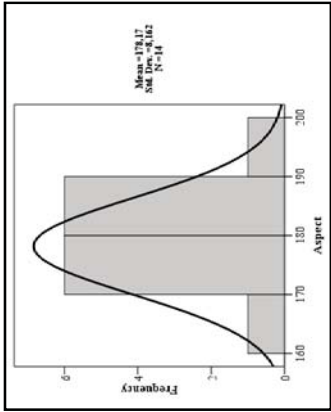


Figure F.17. Histogram and normal line of “aspect”

- **MEAN ANNUAL PRECIPITATION (MAP)**

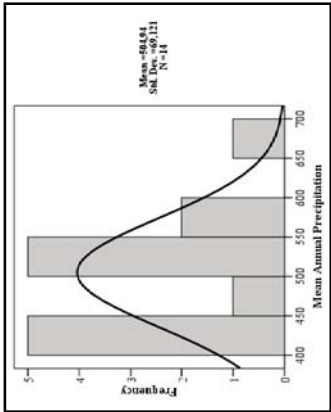


Figure F.19. Histogram and normal line of “mean annual precipitation”

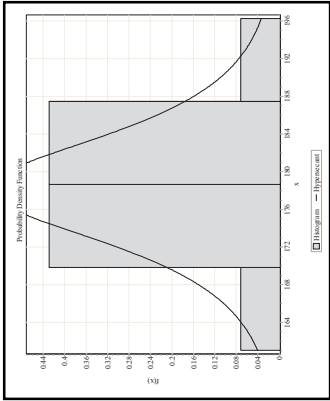


Figure F.18. Probability distribution function of “aspect”

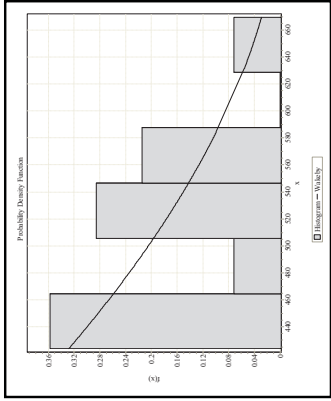


Figure F.20. Probability distribution function of “mean annual precipitation”

- **MEAN ANNUAL TEMPERATURE (T)**

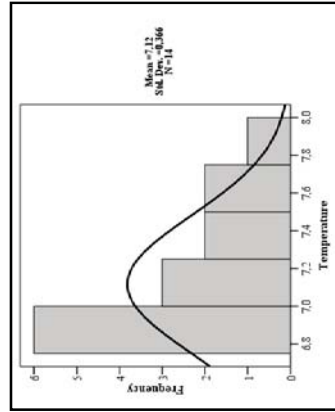


Figure F.21. Histogram and normal line of “temperature”

- **CURVE NUMBER (CN)**

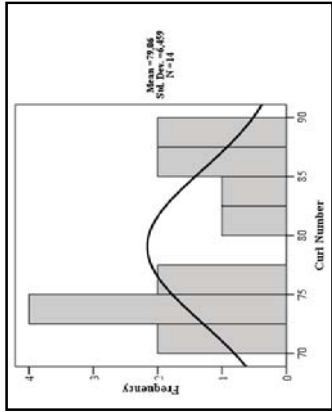


Figure F.23. Histogram and normal line of “curve number”

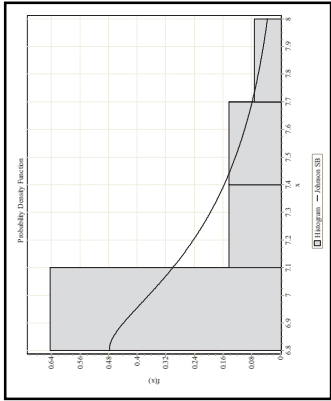


Figure F.22.. Probability distribution function of “temperature”

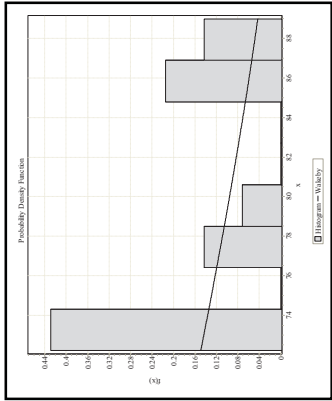


Figure F.24. Probability distribution function of “curve number”

- **MEAN ANNUAL DISCHARGE (Q)**

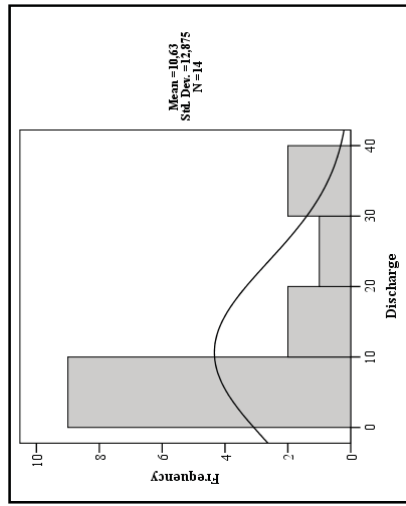


Figure F.25. Histogram and normal line of “Mean Annual Discharge”

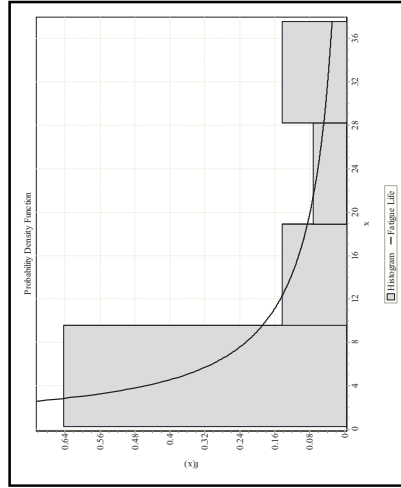


Figure F.26. Probability distribution function of “Mean Annual Discharge”

# APPENDIX G

## SOFTWARE EASY-FIT

### Supported Distributions by Easy-Fit Distribution Fitting Software

- Bernoulli
- Beta
- Binomial
- Burr
- Cauchy
- Chi-Squared
- Dagum
- Discrete Uniform
- Erlang
- Error
- Error Function
- Exponential
- F
- Fatigue Life
- Frechet
- Gamma
- Generalized Extreme Value
- Generalized Gamma
- Generalized Logistic
- Generalized Pareto
- Geometric
- Gumbel Max
- Gumbel Min
- Hyperbolic Secant
- Hypergeometric
- Inverse Gaussian
- Johnson SB
- Johnson SU
- Kumaraswamy
- Laplace
- Levy
- Logarithmic
- Logistic
- Log-Gamma
- Log-Logistic
- Log-Pearson 3 (LP3)
- Lognormal
- Negative Binomial
- Nakagami
- Normal
- Pareto
- Pareto 2 (Lomax)
- Pearson 5
- Pearson 6
- Pert
- Poisson
- Phased Bi-Exponential
- Phased Bi-Weibull
- Power Function
- Rayleigh
- Reciprocal
- Rice
- Student's t
- Triangular
- Uniform
- Wakeby
- Weibull

Many distributions are available in two versions. For example, both two-parameter and three-parameter Weibull distributions are supported. In addition, seven advanced distributions are available:

- Generalized Extreme Value
- Generalized Logistic
- Generalized Pareto
- Log-Pearson 3 (LP3)
- Phased Bi-Exponential
- Phased Bi-Weibull
- Wakeby

The use of advanced distributions for data analysis essentially increases the validity of models, which, in turn, leads to better decisions.

The following discrete distributions are supported:

- Bernoulli
- Binomial
- Discrete Uniform
- Geometric
- Hypergeometric
- Logarithmic
- Negative Binomial
- Poisson