

DISSERTATION

THE EFFECT OF PARAMETER UNCERTAINTY  
IN STOCHASTIC STREAMFLOW SIMULATION

Submitted by

Dong-Jin Lee

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2009

UMI Number: 3401018

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3401018

Copyright 2010 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



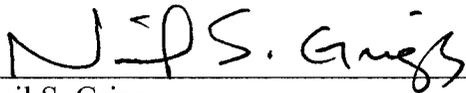
ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

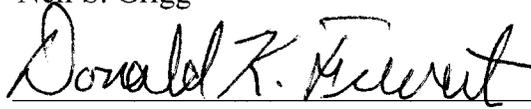
COLORADO STATE UNIVERSITY

September 8, 2009

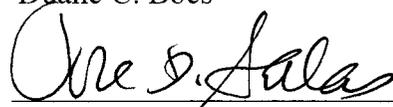
WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY DONG JIN LEE ENTITLED THE EFFECT OF PARAMETER UNCERTAINTY IN STOCHASTIC STREAMFLOW SIMULATION BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate work

  
\_\_\_\_\_  
Neil S. Grigg

  
\_\_\_\_\_  
Donald K. Frevert

  
\_\_\_\_\_  
Duane C. Boes

  
\_\_\_\_\_  
Advisor (Jose D. Salas)

  
\_\_\_\_\_  
Department Head (Luis Garcia)

## ABSTRACT OF DISSERTATION

### THE EFFECT OF PARAMETER UNCERTAINTY IN STOCHASTIC STREAMFLOW SIMULATION

Hydrologic time series simulation based on a stochastic model is intended to obtain a set of equally likely hydrologic sequences that could possibly occur in the future and might be useful for determining the uncertainty of decision variables such as the storage capacity of a reservoir. Since stochastic models generally hinge on parameters that are estimated based on a limited historical sample, the model parameters become uncertain and so are any decision variables that are derived from the generated samples. The main objective of this study is to propose and analyze methods for quantifying the effect of parameter uncertainty of the models that are used in the generation of synthetic streamflow series. As a way of quantifying parameter uncertainty of a stochastic model, asymptotic and Bayesian approaches have been implemented and their performances compared through extensive simulation experiments. Alternative streamflow simulation techniques have been utilized with parameter uncertainty incorporated such as stochastic models of annual streamflows at single and multiple sites as well as temporal and spatial disaggregation models. The impact of parameter uncertainty is shown to increase the variability of generated flow statistics and resultant design related variables, which is visible even with a relatively large sample size, e.g. sample size of 200. The Bayesian

approach produces larger variability of generated statistics for small sample sizes than the asymptotic approach, and the difference between the two approaches is more evident for the case of generation of streamflows with high serial correlations. The effect of parameter uncertainty within disaggregation models is not as significant on the first and second moments of disaggregated flows as the effect of parameter uncertainty of the models that generate the input variables; whereas the effect of parameter uncertainty of disaggregation models results in more variability of month-to-month, month-to-annual, and cross correlations than those induced by the uncertainty of the model parameters of input variables.

Dong-Jin Lee  
Department of Civil and Environmental Engineering  
Colorado State University  
Fort Collins, CO 80523  
Fall 2009

## ACKNOWLEDGEMENTS

The author expresses his appreciations to the living God who makes the impossible the possible.

The author expresses his sincere gratitude to his advisor, Dr. Jose D. Salas, Professor of Civil and Environmental Engineering for his continuous guidance and encouragement during the author's work on this research. Great gratitude is due to his co-advisor, Dr. Duane C. Boes, Emeritus Professor of Statistics for his continuously invariable contribution and comment to this research. Special thanks and appreciations are expressed to the author's committee members: Dr. Neil S. Grigg, Professor of Civil and Environmental Engineering, and Dr. Donald K. Frevert, U.S. Bureau of Reclamation for their assistance and suggestions.

The acknowledgement is due to the Civil and Environmental Engineering Department Borland Graduate Student Scholarship and the author also expresses thanks to the U.S. Bureau of Reclamation project "Development of Stochastic Hydrology for the Colorado River System" for the technical support throughout his studies.

The author wishes to express his sincere gratitude to his former advisor, Dr. Jun-Haeng Heo, Professor of Civil and Environmental Engineering, Yonsei University, Korea,

for his encouragement and advice. Special appreciations are extended to Professors of Hydro lab in Civil and Environmental Engineering, Yonsei University, Korea: Dr. Woncheol Cho and Dr. Sung-Uk Choi for their encouragements.

The author wishes to express the greatest gratitude to his parents: Chang-Woo Lee and Eun-Hee Yoon for their continuous supports mentally and physically during the author's life. Deepest gratitude is due to his wife, Jung Woon Shin for encouragement, endurance, comprehension, and advice.

Finally, this dissertation is dedicated to my parents.

## TABLE OF CONTENTS

### Chapter I

INTRODUCTION .....	1
1.1 General Remarks.....	1
1.2 Problem Definition and Brief Literature Review.....	2
1.3 Research Objectives.....	6
1.4 Chapter Organization .....	8
References.....	10

### Chapter II

UNCERTAINTY ANALYSIS FOR SYNTHETIC STREAMFLOW GENERATION....	13
2.1 Introduction.....	14
2.2 Stochastic Model and Parameter Uncertainty .....	15
2.2.1 Asymptotic analysis .....	17
2.2.2 Bayesian analysis .....	21
2.2.3 Bootstrap .....	23
2.3 Theoretical Example of Parameter Uncertainty Effect.....	25
2.4 Simulation of Synthetic Streamflows with Parameter Uncertainty .....	28
2.4.1 Sampling of parameter estimates .....	29
2.4.2 Parameter uncertainty effects on basic statistics.....	32
2.5 Parameter Uncertainty Effect on Design Variables.....	34
2.5.1 Uncertainty effect on storage capacity with different sample sizes in two demand level options .....	34
2.5.2 Uncertainty effect on storage capacity for different demand level.....	36
2.5.3 Uncertainty effect on storage capacity with reliability .....	37
2.5.4 Drought related Statistics .....	39

2.5.5 Brief summary of parameter uncertainty effect on design variables .....	40
2.6 Comparative Analysis with Nonparametric Technique .....	41
2.7 Concluding Remarks.....	42
References.....	59
Appendix 2.A: Additional Figures and Tables.....	62

### Chapter III

UNCERTAINTY CONSIDERATION IN TEMPORAL DISAGGREGATION MODEL .....	86
3.1 Introduction.....	87
3.2 Generation of monthly synthetic streamflows using disaggregation models .....	88
3.2.1 Traditional temporal disaggregation models.....	88
3.3 Theoretical consideration of parameter uncertainties .....	93
3.3.1 Asymptotic approach .....	93
3.3.2 Bayesian inference .....	95
3.4 Uncertainty incorporation into synthetic streamflows.....	97
3.4.1 Simulation experiments .....	97
3.4.2 Preliminary analysis of disaggregation models and uncertainty considerations .	99
3.4.3 Uncertainty effect of temporal disaggregation model parameters combined with historical annual flows.....	101
3.4.4 Uncertainty effect of parameter uncertainty of annual streamflow generation and temporal disaggregation.....	103
3.4.5 Brief summary of parameter uncertainty effect on design variables .....	107
3.5 Summary and Conclusions .....	108
References.....	120
Appendix 3.A: Derivation of the asymptotic variance-covariance matrix of Lane's condensed temporal disaggregation scheme.....	122
Appendix 3.B: Derivation of the asymptotic variance-covariance matrix of Lane's condensed temporal disaggregation scheme.....	127
Appendix 3.C: Additional Figures and Tables.....	131

## Chapter IV

PARAMETER UNCERTAINTY IN SPATIAL DISAGGREGATION .....	153
4.1 Introduction.....	154
4.2 Parameter estimation in spatial disaggregation.....	158
4.2.1 Traditional parameter estimation .....	158
4.2.2 Maximum likelihood estimation.....	159
4.3 Parameter uncertainty consideration.....	163
4.3.1 Asymptotic distribution of the parameters.....	163
4.3.2 Bayesian inference on parameters of the spatial disaggregation .....	167
4.4 Application to Colorado River System .....	170
4.4.1 Parameter uncertainty effect based on historical key-station flows.....	170
4.4.2 Parameter uncertainty effects combining with uncertainty of input variables...	174
4.4.3 Parameter uncertainty effects based on Lane’s condensed model .....	179
4.5 Summary and Concluding Remarks .....	180
References.....	197
Appendix 4.A: Additional Figures and Tables.....	199

## Chapter V

UNCERTAINTY CONSIDERATION IN MULTIVARIATE ANNUAL STREAMFLOWS GENERATION .....	221
5.1 Introduction.....	222
5.2 Multivariate Autoregressive Model .....	226
5.2.1 Maximum likelihood estimators and asymptotic distributions.....	227
5.2.2 Parameter uncertainty incorporation by using Bayesian framework.....	229
5.3. Application to Colorado River Basin.....	232
5.3.1 Parameter uncertainty effect .....	232
5.3.2 Comparative analysis of univariate and multivariate generation.....	239
5.3.3 Parameter uncertainty associated with transformation .....	241
5.4. Summary and Conclusion.....	243

References.....	256
Appendix 5.A: Additional Figures and Tables.....	258

## Chapter VI

### PROPORTIONAL DISAGGREGATION MODEL AND PARAMETER

UNCERTAINTY.....	296
6.1 Introduction.....	297
6.2 Proportional disaggregation model.....	299
6.2.1 Model description .....	299
6.2.2 Parameter estimation.....	301
6.3 Application to temporal disaggregation.....	306
6.3.1 Comparative analysis of different estimators .....	306
6.3.2 Model evaluation based on sample statistics .....	307
6.3.3 Parameter uncertainty incorporation.....	308
6.4 Summary and Conclusion.....	309
Reference .....	322
Appendix 6.A: Additional Figures and Tables.....	324

## Chapter VII

CONCLUSIONS, CONTRIBUTIONS, AND RECOMMENDATIONS.....	335
7.1 Summary and Conclusions .....	335
7.2 Contributions and recommendations .....	340

## **Chapter I**

### **INTRODUCTION**

#### **1.1 General Remarks**

Stochastic modeling of time series has been widely used in the planning and management of water resource systems (Salas et al., 1980; Loucks et al., 1981; Bras and Rodriguez-Iturbe, 1985; Grygier and Stedinger, 1990; Hipel and McLeod, 1994; Salas et al., 2000). Two important utilizations of stochastic modeling in hydrology are to forecast and to simulate a hydrologic time series. The purpose of the simulation is to obtain a set of stochastically equivalent series of observations that could possibly happen in the future, which might be useful to determine the expected capacity of hydrologic structures with certain reliability. A number of mathematical stochastic models have been extensively suggested regarding stochastic simulation of hydrologic processes: autoregressive (AR), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) models, disaggregation models, fractional Gaussian noise models, broken line models, and shifting level models, etc (Salas et al., 1980). Three steps are usually required for formal model procedures: model identification, model estimation, and model diagnostic checking. Obtaining efficient parameter estimates might be crucial after the specific stochastic model is identified for a hydrologic system

of interest. It is common that inadequacy of fit may simply arise from inefficient fitting and not from an inadequate model (Box and Jenkins, 1970). Several parameter estimation techniques are available to obtain the optimal estimates of parameters: method of moments estimators, maximum likelihood estimators, and least square estimators. However, the precision of parameter estimates arise from the limited information of available historical data sets, even though the best estimator could be applied. If the available historical data is sufficiently long (e.g. hundreds of years), the model parameters could be estimated with good precision; the synthetic samples produced from the model would reflect the expected variability of the process under consideration; and consequently the expected variability of the design variables could be obtained from them (e.g. the size of the needed storage capacity for a reservoir). However, the usual lengths of historical streamflow records are short, which means that the model parameters are uncertain. Consequently, the variability of the design variables may be uncertain beyond what is expected.

## **1.2 Problem Definition and Brief Literature Review**

Vicens et al. (1975) argued that the uncertainties in water resources can be classified into two different uncertainties, which are natural uncertainty developing from the assumed random nature of hydrological process and informational uncertainty developing from the limited information regarding the true nature of the process. The sampling errors, which could be defined by the discrepancy between the parameter estimates based on historical samples and true population parameters, have been

illustrated to impose a significant effect on practical water resource design problems such as impact on estimates of the reliability with which reservoirs of different capacities can meet a specific demand (Stedinger and Taylor, 1982). However, the conventional approaches of reservoir determination procedures do not have the abilities to exhibit these parameter uncertainties. Therefore, the incorporation of uncertainties of the sampling errors could result in a better design than the traditional approaches (Vicens et al., 1975; Wood, 1978; Valdes et al., 1977; Klemes et al., 1981; Salas et al., 1980; Grygier and Stedinger, 1990).

Theoretically, this parameter uncertainty can be quantified and expressed in terms of the sampling distribution. Based on the asymptotic behavior of the maximum likelihood estimators, Box and Jenkins (1970) derived the large sample variance-covariance matrix of maximum likelihood estimators (MLE) for univariate autoregressive moving average model by using the information matrix to interpret the uncertainty of parameter estimators, which enables one to define an approximate distribution of parameter estimates for a sufficiently large sample size. McLeod and Hipel (1978) developed the WASIM3 (Waterloo Simulation Procedure 3) algorithm for the univariate ARMA model based on the large sample theory so as to explain how parameter uncertainty is incorporated into reservoir design and affects the storage capacity-reliability relationships for reservoir systems. The incorporation of parameter uncertainty into the streamflow generation is also available for method of moments estimators (MOME). The idea is that once parameter estimates are expressed as functions of sample moments, relevant asymptotic distributions could be given by central limit theory. For instance, Vecchia et al. (1983) proposed the asymptotic distributions of

MOME of the first order periodic AR model based on the relationship with asymptotic distributions of seasonal autocorrelations. The application of multivariate stochastic models has been useful in the planning and management of water resource systems of several sites with interdependence (Matalas, 1967; Pegram and James, 1972; O'Connell, 1974; Salas and Pegram, 1977; Lettenmaier, 1980; Salas et al., 1980; Loucks et al, 1981; Camacho et al, 1985, 1987; Salas et al., 1985; Stedinger, et al., 1985a). By introducing the usefulness of a contemporaneous autoregressive moving average model (CARMA) as a subset of the general multivariate ARMA model in the generation of the multivariate streamflows, Camacho et al. (1987) examined the large sample properties of MLE of the CARMA model and derived asymptotic variances of MLE to compare the efficiency of joint estimates with those in the univariate case.

Parameter uncertainty issues could also be dealt within a Bayesian framework. For a Bayesian approach to parameter uncertainty, there is no major difference between univariate autoregressive models and univariate regression models (see Zeller, 1971; Vicens et al., 1975; Stedinger and Taylor, 1982), likewise for the multivariate cases of each of the models (Valdes et al., 1977). Based on the Bayesian regression model, Vicens et al. (1975) proposed incorporation of parameter uncertainty in the synthetic streamflow generation scheme for the univariate annual AR model by deriving a Bayesian predictive probability density function of streamflows. A multivariate extension of that by Vicens et al. (1975) is available in the literature based on Bayesian multivariate regression and a multivariate AR model (Valdes et al., 1977). However, these previous works have inadequacy since those algorithms are not a possible realization of the underlying stochastic process and only approximate the relationship

between parameter uncertainty and synthetic flows to be generated (Davis, 1977; McLeod and Hipel, 1978; Stedinger and Taylor, 1982). In addition to parameter uncertainty regarding the annual streamflow generation, Stedinger and Taylor (1982) presented the incorporation procedure of parameter uncertainty of the first order annual univariate AR process into monthly streamflow generation based on the Bayesian technique. They also examined the impact on estimates of monthly reservoir system reliability, which supported the previous suggestions by Klemes et al. (1981) and Burges and Lettenmaier (1981) that parameter uncertainty might be more important than the choice of generation models. More recently, Thyer et al. (2002) applied a numerical method to incorporate parameter uncertainty into streamflow generation based on the first order AR model in conjunction with the Box-Cox transformation, where the Bayesian Markov Chain Monte Carlo method was employed to evaluate parameter uncertainty.

Since the model proposed by Valencia and Schaake (1973), several temporal or spatial disaggregation models have been proposed and widely used for their simplicity in parameter estimation and for better performance in preserving statistical characteristics of both levels of flows (annual/seasonal or key-sites/sub-sites) as an alternative to direct modeling of seasonal or multivariate streamflows (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Tao and Delleur, 1976; Lane, 1979; Stedinger and Vogel, 1984; Stedinger et al, 1985a, 1985b; Grygier and Steinger, 1990; Lane and Frevert, 1990; Santos and Salas, 1992). Applying the normal regression model enables one to incorporate parameter uncertainty into the disaggregation model in the Bayesian framework, which has been found either in theoretical studies (Zeller, 1971; Box and Tiao, 1973), or in the application to water resource systems planning based on the

temporal disaggregation model (Stedinger et al., 1985b). Parameter uncertainty consideration might improve the ability to specify how a given hydrologic system is likely to perform in the future (Stedinger et al., 1985b). On the other hand, Grygier and Stedinger (1990) proposed that the parameter uncertainty effect in spatial disaggregation is thought to be relatively small compared with that of the annual or seasonal basin flows at a single site, and they also proposed that the annual to seasonal disaggregation model should be used for incorporating parameter uncertainty into the complex multivariate annual to seasonal disaggregation. However, extensive exploration of the effect of parameter uncertainty incorporation into spatial disaggregation has not been clearly made.

### **1.3 Research Objectives**

The general objective of this research is to develop and to propose uncertainty quantification procedures that incorporate parameter uncertainty of mathematical models into the generation of a synthetic streamflow series in a multi-sites channel system. Specific objectives that will be considered in this research are:

- (1) To systematically analyze the effect of parameter uncertainty of a single univariate stochastic model for generating streamflow data based on traditional asymptotic and Bayesian approaches. The nonparametric approach based on bootstrapping will also be utilized for the comparison of parametric consideration.
- (2) To expand the parameter uncertainty incorporation procedure of the single site

generation to seasonal flow generation based on temporal disaggregation. In order to consider parameter uncertainty, asymptotic distributions will be derived and the comparative analysis with the conventional Bayesian approach will be given.

- (3) To develop the parameter uncertainty incorporation technique for synthetic flows at multiple sites by using spatial disaggregation model. Based on the multivariate regression concept, the asymptotic distribution and Bayesian posterior distribution of parameters of spatial disaggregation models will be derived as a way of incorporating parameter uncertainty into the flow generation, and the uncertainty impacts on statistical properties and related design variables will be investigated and compared.
- (4) To enhance parameter uncertainty incorporation for multivariate synthetic streamflows generated by using the general multivariate autoregressive model. The traditional Bayesian regression concept for the first multivariate autoregressive model will be expanded to the general multivariate autoregressive model and comparative analysis with the conventional asymptotic approach will be presented.
- (5) To present a simple disaggregation model based on the proportionality between annual and seasonal flows (key and sub stations flows), which does not require the normality constraint and the correspondingly required adjustment procedure.

## 1.4 Chapter Organization

This research consists of five stand alone chapters, of which each chapter has its own abstract, application and conclusion. In Chapter 2, parameter uncertainty incorporation into a lag-1 autoregressive model, a commonly used stochastic model for generating synthetic streamflow data, will be presented. The impact of each type of parameter uncertainty of the model will be investigated based on statistical properties and related design variables, and the performance of conventional approaches will be compared. In Chapter 3, the expansion of parameter uncertainty consideration into a temporal disaggregation model will be provided. Two temporal condensed disaggregation models, LAST (Lane, 1979; Lane and Frevert, 1990) and SPC (Stedinger et al., 1985), will be employed and parameter uncertainty regarding those models will be taken into account by using asymptotic theory and Bayesian posterior distributions. In Chapter 4, the parameter uncertainty effect will be discussed in the case of generated streamflows based on the spatial disaggregation model. Asymptotic and Bayesian posterior distributions of parameters of the simple spatial disaggregation model (Valencia-Schaake model) will be derived. The impact of parameter uncertainty and also the performance of two approaches will be examined based on flow related statistics. Chapter 5 will present the uncertainty incorporation to the multivariate generation case. The General multivariate autoregressive model will be implemented and the uncertainty of parameters will be investigated based on both asymptotic and Bayesian approaches. Finally, Chapter 6 will report a simple disaggregation method which is based on the proportionality between two levels of flows. Throughout the chapters, annual and

monthly streamflows data sets at several sites in the Colorado River Basin, as well as annual and monthly streamflows at the St. Lawrence River will be used to show how parameter uncertainty can be brought into a practical simulation study.

## References

- Box, G.E.P. and G.C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc., N.Y.
- Box, G. E. P. and G.W. Jenkins (1976). *Time Series Analysis Forecasting and Control*, revised edition, Holden-Day, San Francisco.
- Bras, R.L. and I. Rodriguez-Iturbe (1985). *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass..
- Burges, S.J. and D.P. Lettenmaier (1981). Reliability measures for water supply reservoirs and the significance of long-term persistence, *Proc. Int. Symp. Real-Time Operation of Hydrosystems*, Waterloo, Ont., Canada.
- Camacho, F., A.I. McLeod, and K.W. Hipel (1985). Contemporaneous autoregressive-moving average (CARMA) modeling in water resources, *Water Resources Bulletin*, 21(4), pp. 709-720.
- Camacho, F., A.I. McLeod, and K.W. Hipel (1987). Multivariate contemporaneous ARMA model with hydrological applications, *Stochastic Hydrology and Hydraulics*, 1, pp. 141-154.
- Davis, D.R. (1977). Comment on 'Bayesian generation of synthetic streamflows' by G.I. Vicens, I. Rodriguez-Iturbe, and J.C. Schaake, Jr., *Water Resources Research*, 13(5), pp. 853-854.
- Grygier, J.C. and J.R. Stedinger (1990). *SPIGOT, A synthetic Streamflow Generation Software Package, technical description*, version 2.5, School of Civil and Environmental Engineering, Cornell University, Ithaca, N. Y.
- Hipel, K.W. and A.I. McLeod (1994). *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier.
- Klemes, V., R. Srikanthan and T. A. McMahon (1981). Long-memory flow models in reservoir analysis: What is their practical value?, *Water Resources Research*, 17(3), pp. 737-751.
- Lane, W.L. (1979). *Applied Stochastic Techniques, User's Manual*, Bureau of Reclamation, Engineering and Research Center, Denver, Co.
- Lane, W.L. and D.K. Frevert (1990). *Applied Stochastic Techniques: User's Manual*, personal computer version 5.2, Earth Sciences Division, Bureau of Reclamation, U.S. Department of the Interior, Denver, Colo.

- Lettenmaier, D.P. (1980). Parameter estimation for multivariate streamflow synthesis, *Proc of Joint Automatic Control Conf.*, San Francisco, Paper FA6-D.
- Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981). *Water Resources Planning and Analysis*, Prentice Hall, Englewood Cliffs, New Jersey.
- Matalas, N.C. (1967). Mathematical assessment of synthetic hydrology, *Water Resources Research*, 3(4), pp. 937-945.
- McLeod, A.I. and K.W. Hipel (1978). Simulation procedures for Box-Jenkins models, *Water Resources Research*, 14(5), pp. 969-975.
- Mejia, J.M. and J. Roussell (1976). Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), pp. 185-186.
- O'Connell, P.E. (1974). *Stochastic Modeling of Long-Term Persistence in Streamflow Sequences*, Ph.D Dissertation, Civil Engineering Dept., Imperial College of Science and Technology, London.
- Pegram, G.G.S. and W. James (1972). Multilag multivariate autoregressive model for the generation of operational hydrology, *Water Resources Research*, 8(4), pp. 1074-1076.
- Salas, J.D. and G.G.S. Pegram (1977). A seasonal multivariate multilag autoregressive model in hydrology, *Proceeding of the Third International Symposium of Theoretical and Applied Hydrology*, Colorado State University, Fort Collins, Colorado.
- Salas, J.D., G. Tabios, and P. Bartolini (1985) Approaches to multivariate modeling of water resources time series, *Water Resources Bulletin*, 21(4), pp. 683-708.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.
- Salas, J.D., N. Saada, C.H. Chung, W.L. Lane, and D.K. Frevert (2000). Stochastic Analysis, Modeling, and Simulation (SAMS) version 2000 User's manual, *Technical report No. 10*, Computing Hydrology Laboratory, Water Resources, Hydrologic and Environmental Sciences, Engineering Research Center, Fort Collins, Colorado.
- Santos, E.G. and J.D. Salas (1992). Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, 118(5), pp. 765-784.
- Stedinger, J.R. and M.R. Taylor (1982). Synthetic streamflow generation: 2. Effect of parameter uncertainty, *Water Resources Research*, 18(4), pp. 919-924.
- Stedinger, J.R. and R.M. Vogel (1984). Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(1), pp. 47-56.

- Stedinger, J.R., D.P. Lettenmaier, and R.M. Vogel (1985a). Multisite ARMA(1,1) and disaggregation models for annual streamflow generation, *Water Resources Research*, 21(4), pp. 497-509.
- Stedinger, J.R., D. Pei, and T. Cohn (1985b). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resources Research*, 21(5), pp. 665-675.
- Tao, P.C. and J.W. Delleur (1976). Multistation, multiyear synthesis of hydrologic time series by disaggregation, *Water Resources Research*, 12(7), pp. 1303-1312.
- Thyer, M., G. Kuczera, and Q.J. Wang (2002). Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *Journal of Hydrology*, 265, pp. 246-257.
- Valdes, J.B., I. Rodriguez-Iturbe, and G.J. Vicens (1977). Bayesian generation of synthetic streamflows; 2. The multivariate case, *Water Resources Research*, 13(2), pp. 291-295.
- Valencia, D. and J.C. Schaake, Jr. (1973). Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), pp. 580-585, 1973.
- Vecchia, Jr. A.V., J.T. Obeysekera, J.D. Salas, and D.C. Boes (1983). Aggregation and estimation for low order periodic ARMA models, *Water Resources Research*, 19(5), pp. 1297-1306.
- Vicens, G.J., I. Rodriguez-Iturbe, and J.C. Schaake (1975). Bayesian generation of synthetic streamflows, *Water Resources Research*, 11(6), pp. 827-838.
- Wood, E.F. (1978). Analysing hydrologic uncertainty and its impact upon decisionmaking in water resources, *Advances in Water Resources*, 1(5), pp. 299-305.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., New York.

## Chapter II

# UNCERTAINTY ANALYSIS FOR SYNTHETIC STREAMFLOW GENERATION

**Abstract:** Synthetic streamflow generation has been widely used in hydrology and water resources for a number of practical problems such as determining the capacity of a reservoir and assessing the long-term behavior of an existing reservoir. Synthetic streamflows can be obtained using parametric and non-parametric approaches. The former assumes that a certain mathematical model describes the stochastic behavior of the underlying process, e.g. streamflow. The mathematical model hinges on a number of parameters that must be estimated from historical data. However, the usual lengths of historical streamflow records are short, which means that the model parameters are uncertain and consequently, the variability of the design variables may be uncertain beyond what is expected. A number of approaches have been proposed in literature to tackle the problem of parameter uncertainty in simple stochastic models. In the paper described herein, we take two different approaches based on the asymptotic theory and Bayesian framework for an AR(1) model and investigate in some detail the effect of the uncertainty in one or more parameters on the design variables such as the reservoir size and reliability. A non-parametric approach based on bootstrap has been also

implemented for the purpose of the performance comparison. Our analysis has been conducted based on simulation studies. In the first, part theoretical behavior of the synthetic design variable has been illustrated for a wide range of parameters. Furthermore, this paper also has included examples to illustrate the applicability of the concepts obtained in the study. As a result, parameter uncertainty shows considerable effects on the variability of generated design variables even if the sample size is equal to 100. Bayesian analysis provides more variability of generated design variables than asymptotic analysis for smaller sample size, and it also concludes that less variability and quantiles of design variables would be expected when using a nonparametric approach rather than a parametric one.

## **2.1 Introduction**

Uncertainties in water resources commonly arise from the random nature of the hydrological process and from the limited information (data) that is available regarding the true nature of the underlying process. Since the parameters of stochastic models are estimated using the limited historical records, these estimates are uncertain quantities. These uncertainties in the parameters of stochastic models are translated into the uncertainty of decision variables of planning and management of water resources. Perhaps the simplest example may be the case of designing a flood related structure for a 100-yr flood. The flood frequency distribution is an expression of the natural uncertainty of the underlying extreme floods, but the magnitude of a specific flood

quantile, e.g. the 100-yr flood quantile, is uncertain. It is well known that such uncertainty is commonly expressed by determining the confidence limits of the population quantile (e.g. Stedinger et al., 1993).

Likewise, conventional approaches for designing the capacity of a reservoir generally consider the effect of the natural uncertainty of streamflows. For example this is done by building a stochastic model and simulating synthetic flow records from which the frequency distribution of the needed reservoir storage capacity can be obtained. However, as is the case for the flood protection design problem illustrated above, the stochastic model of streamflows has a parameter set that is uncertain because of the limited data available and as a result, the distribution of decision variables related to them, e.g. reservoir capacity, is also uncertain. Although this issue has been recognized in the past and some procedures have been suggested (e.g. Vicens et al., 1975; Wood, 1978; Valdes et al., 1977; McLeod and Hipel, 1978; Salas et al., 1980; Klemes et al., 1981; Grygier and Stedinger, 1990), unfortunately the problem remains perhaps because of its complexity and the lack of understanding of the many factors involved.

In this paper we will report results of a systematic study of the effect of parameter uncertainty of a lag-1 autoregressive model, which is a commonly used stochastic model for generating synthetic streamflow data, on the size of the storage capacity of a reservoir.

## **2.2 Stochastic Model and Parameter Uncertainty**

The uncertainties of parameter estimates can be statistically quantified in terms of their probability distribution functions. To tackle this problem, two different

approaches are currently available; asymptotic analysis and Bayesian inference. In the asymptotic analysis, the approximate distribution of a parameter estimate is derived based upon large sample theory that the estimates will converge into their exact values when the sample size is large enough. Box and Jenkins (1970) derived the large sample variance-covariance matrix of parameter estimators (MLE) for univariate autoregressive moving average model by using the information matrix to interpret the uncertainty of parameter estimators, which enables one to define an approximate distribution of parameter estimates for sufficient large sample size. By introducing the usefulness of a contemporaneous autoregressive moving average model (CARMA) as a subset of the general multivariate autoregressive moving average model (MARMA) model for the generation of the multivariate streamflows, Camacho et al. (1987) examined the large sample properties of MLE of the CARMA model and derived asymptotic variances of MLE to compare the efficiency of joint estimates with those in the univariate case.

In the Bayesian framework, posterior distributions of parameter estimates are allowed to measure their uncertainties, and by deriving the posterior from the proper prior, both the natural and the parameter uncertainties can be evaluated. Using Bayesian analysis, Vicens et al. (1975) and Valdes et al. (1977) investigated parameter uncertainty effects on univariate and multivariate annual synthetic streamflow generation and showed that it could give better reliability under uncertainty conditions.

McLeod and Hipel (1978) suggested the computer algorithm, which is able to incorporate parameter uncertainty into streamflow generation by sampling parameter estimates from the asymptotically derived posterior distributions. Based on the lag-1 autoregressive model, AR(1), the parameter uncertainty impact on the reservoir system

was discussed by Stedinger and Taylor (1982), which showed that the uncertainty effects might be more significant than model choice to improve the reservoir system reliability. In this study, the incorporation of parameter uncertainty into the synthetic streamflow generation will be systematically examined based on both asymptotic and Bayesian methods, and the performance of the two methods will be compared. Basically, annual streamflows in a single site will be generated by using the AR(1) model, and MLE will be used for parameter estimation. Sampled parameter estimates from derived distributions, based on either the asymptotic distribution or Bayesian posterior distribution, will be used as a way of incorporating the parameter uncertainty into the synthetic streamflows, which will be able to generate a different synthetic streamflow set in each simulated trace. Two different historical streamflow sets, with different serial correlations will be chosen for simulative analysis, and basic statistics of simulated streamflows and resultant design variables, such as storage and drought related statistics, will be used to investigate the parameter uncertainty effect. The effect of each uncertain parameter estimate on simulated flows and design variables will be tested throughout the simulation analysis, and the most significant factor and its effect will be examined. Moreover, a bootstrapping technique will be implemented for comparison with the parametric methods.

### **2.2.1 Asymptotic analysis**

Assume that the underlying annual streamflows, denoted by  $Y_t$ , is stationary and normally distributed with the mean  $\mu$  and variance  $\sigma^2$ , which has an autoregressive

autocorrelation with the constant parameter. The first order autoregressive model, AR(1) representing the variable  $Y_t = T(X_t)$  is generally expressed by:

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \varepsilon_t, \quad (2.1)$$

where  $T(\cdot)$  is a transformation function to obtain normalized flow  $Y_t$  from the original streamflow  $X_t$ ,  $\phi$  is the autoregressive coefficient that satisfies the causal condition in the range of  $-1 < \phi < 1$ , and  $\varepsilon_t$  is the time independent innovation term with zero mean and variance  $\sigma_\varepsilon^2$ . The first step in applying the AR(1) model is to estimate the parameters from the historical streamflows data. It is well known that MLE of AR(1) parameters are given by (Box and Jenkins, 1976):

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n Y_t, \quad (2.2)$$

$$\hat{\phi} = \frac{D_{12}}{D_{22}}, \quad (2.3)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-1} (D_{11} - \hat{\phi} D_{12}), \quad (2.4)$$

where  $D_{ij} = \frac{n}{n+2-i-j} \sum_{l=0}^{n+1-i-j} Y'_{i+l} Y'_{j+l}$  for  $Y'_t = Y_t - \hat{\mu}$ ,  $t = 1, 2, \dots, n$  and  $n$  represents sample size.

Since  $Y_t$  is stationary with mean  $\mu$  and auto-covariance function  $\gamma(\cdot)$ ,

$E(\hat{\mu}) = \mu$  and  $E(\hat{\mu} - \mu)^2 = \frac{1}{n} \sum_{h=-n}^n \left(1 - \frac{|k|}{n}\right) \gamma(k)$ . If  $\gamma(n) \rightarrow 0$  as  $n \rightarrow \infty$ , then

$\frac{1}{n} \sum_{h=-n}^n \left(1 - \frac{|k|}{n}\right) \gamma(k)$  converges to zero, so that  $\hat{\mu}$  converges in mean square to  $\mu$ .

Also, as  $n \rightarrow \infty$ ,  $\text{var}(\hat{\mu}) = E(\hat{\mu} - \mu)^2 \rightarrow 0$  if  $\gamma(n) \rightarrow 0$  and  $nE(\hat{\mu} - \mu)^2 \rightarrow \sum_{k=-\infty}^{\infty} \gamma(k)$  if

$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$  (Brockwell and Davis, 1991). Thus, the asymptotic distribution of the

sample mean  $\hat{\mu}$  for AR(1) model can be given by:

$$\hat{\mu} \sim AN\left(\mu, \frac{\sigma^2}{n} \left(\frac{1+\rho}{1-\rho}\right)\right), \quad (2.5)$$

where  $\sigma^2$  is the variance of series  $Y_t$ ,  $\rho$  is the lag-1 correlation coefficient of series  $Y_t$ , and ' $\sim AN$ ' means asymptotically normal distributed.

$$L(\phi, \sigma_\varepsilon, \mu | \mathbf{y}) = (2\pi\sigma_\varepsilon^2)^{-n/2} |1 - \phi^2|^{1/2} \exp\left[\frac{-S(\phi, \mu)}{2\sigma_\varepsilon^2}\right], \quad (2.6)$$

where the sum of squares  $S(\phi, \mu) = \sum_{t=1}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$

is an observation vector. Then, the log-likelihood function is:

$$LL(\phi, \sigma_\varepsilon, \mu | \mathbf{y}) = -n \log(\sigma_\varepsilon) + \frac{1}{2} \log(1 - \phi^2) - \frac{S(\phi, \mu)}{2\sigma_\varepsilon^2}. \quad (2.7)$$

The asymptotic variance-covariance matrix of MLE can be given by the inverse of the information matrix of which element is defined by negative second derivative of the log-likelihood function (2.7). Furthermore, if the log-likelihood is approximate quadratic and the maximum is not close to a boundary, even in a moderate sample size, adequate approximations to the variances and covariances of the estimates can be determined by the asymptotic variance-covariance matrix. The second derivatives with respect to  $\phi$  is given by (Box and Jenkins, 1976):

$$\frac{\partial^2 LL}{\partial \phi^2} \approx -\frac{3n}{n-2} \frac{D_{22}}{\sigma_\varepsilon^2}, \quad (2.8)$$

and second derivatives with respect to other parameters are:

$$\frac{\partial^2 LL}{\partial \mu^2} = -\frac{1}{2\sigma_\varepsilon^2} \frac{\partial^2 S(\phi, \mu)}{\partial \mu^2} = -\frac{n}{\sigma_\varepsilon^2} (1-\phi)^2, \quad (2.9)$$

$$\frac{\partial^2 LL}{\partial \sigma_\varepsilon^2} = \frac{n}{\sigma_\varepsilon^2} - \frac{3S(\phi, \mu)}{\sigma_\varepsilon^4}, \quad (2.10)$$

$$\frac{\partial^2 LL}{\partial \mu \partial \phi} = \sum_{t=1}^n (Y_t - \mu) - 2(\phi - 1) \sum_{t=1}^n (Y_{t-1} - \mu), \quad (2.11)$$

$$\frac{\partial^2 LL}{\partial \sigma_\varepsilon \partial \mu} = \frac{2(\phi - 1)}{\sigma_\varepsilon^3} \sum_{t=1}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)], \quad (2.12)$$

$$\frac{\partial^2 LL}{\partial \sigma_\varepsilon \partial \phi} = -\frac{1}{\sigma_\varepsilon^3} \sum_{t=1}^n [(Y_t - \mu)(Y_{t-1} - \mu) - \phi(Y_{t-1} - \mu)^2]. \quad (2.13)$$

From (2.8), the inverse matrix associated with  $\phi$  follows (Box and Jenkins, 1976):

$$I(\phi) = E \left[ -\frac{\partial^2 LL}{\partial \phi^2} \right] \approx \frac{n}{(1-\phi^2)}, \quad (2.14)$$

and the information matrix  $I(\mu)$  and  $I(\sigma_\varepsilon)$  can be simply derived from (2.9) and (2.10):

$$I(\mu) = E \left[ -\frac{\partial^2 LL}{\partial \mu^2} \right] = \frac{n(1-\phi)^2}{\sigma_\varepsilon^2}, \quad (2.15)$$

$$I(\sigma_\varepsilon) = E \left[ -\frac{\partial^2 LL}{\partial \sigma_\varepsilon^2} \right] = \frac{2n}{\sigma_\varepsilon^2}. \quad (2.16)$$

Using derived asymptotic variances of the MLE of  $\hat{\phi}$ ,  $\hat{\sigma}_\varepsilon^2$ , and  $\hat{\mu}$  by taking inverse of information matrices and then by Taylor's theorem for  $\hat{\sigma}_\varepsilon^2$ , the asymptotic

distributions of MLEs are as follows

$$\hat{\phi} \sim AN\left(\phi, \frac{1}{n}(1-\phi^2)\right), \quad (2.17)$$

$$\hat{\sigma}_\varepsilon^2 \sim AN\left(\sigma_\varepsilon^2, \frac{2\sigma_\varepsilon^4}{n}\right), \quad (2.18)$$

$$\hat{\mu} \sim AN\left(\mu, \frac{1}{n} \frac{\sigma_\varepsilon^2}{(1-\phi)^2}\right), \quad (2.19)$$

where  $\hat{\phi}$ ,  $\hat{\sigma}_\varepsilon^2$ , and  $\hat{\mu}$  are shown to be asymptotically uncorrelated to one another

since  $E\left[-\frac{\partial^2 LL}{\partial \mu \partial \phi}\right] = 0$ ,  $E\left[-\frac{\partial^2 LL}{\partial \sigma_\varepsilon \partial \mu}\right] = 0$ ,  $E\left[-\frac{\partial^2 LL}{\partial \sigma_\varepsilon \partial \phi}\right] = 0$  by the stationary assumption

for large sample  $n$ . It is simply found that the asymptotic distribution of  $\hat{\mu}$  in (2.19)

is equivalent with (2.5) since  $\sigma_\varepsilon^2 = \sigma^2(1-\phi^2)$  and  $\phi = \rho$  from the property of AR(1).

### 2.2.2 Bayesian analysis

If the AR(1) model of (2.1) is rewritten using observations  $\mathbf{y}$ , then:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2.20)$$

where  $\alpha = \mu(1-\phi)$ ,  $\beta = \phi$ ,  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ,  $-\infty < \alpha, \beta < \infty$ ,  $0 < \sigma_\varepsilon < \infty$ , and  $\sim N$  denotes normally distributed. Zellner (1971) showed how the Bayesian inference could be applied in the analysis of time series data as follows (also see Box and Tiao, 1973; Stedinger and Taylor, 1982). Assume that the information about  $\alpha$ ,  $\beta$ , and  $\sigma_\varepsilon^2$  are diffused and independent of each other, then it gives the joint prior distribution of the parameters by:

$$p(\alpha, \beta, \sigma_\varepsilon) \propto \frac{1}{\sigma_\varepsilon}, \quad (2.21)$$

with  $-\infty < \alpha < \infty$ ,  $-\infty < \beta < \infty$ , and  $0 < \sigma_\varepsilon < \infty$ . And the likelihood function is

$$L(\alpha, \beta, \sigma_\varepsilon | \mathbf{y}, y_0) \propto \frac{1}{\sigma_\varepsilon^n} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (y_t - \alpha - \beta y_{t-1})^2\right], \quad (2.22)$$

where  $y_0$  denotes an assumed initial value of the series. Combining (2.21) and (2.22),

the joint posterior distribution for the parameters can be derived as

$$p(\alpha, \beta, \sigma_\varepsilon | \mathbf{y}, y_0) \propto \frac{1}{\sigma_\varepsilon^{n+1}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (y_t - \alpha - \beta y_{t-1})^2\right] \quad (2.23)$$

$$\propto \frac{1}{\sigma_\varepsilon^{n+1}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \left[ (n-2)\hat{\sigma}_\varepsilon^2 + n(\alpha - \hat{\alpha})^2 + n(\beta - \hat{\beta})^2 \sum_{t=1}^n y_{t-1}^2 + 2(\alpha - \hat{\alpha})(\beta - \hat{\beta}) \sum_{t=1}^n y_{t-1} \right]\right\},$$

where the mean square error  $\hat{\sigma}_\varepsilon^2$  is give by:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_{t=1}^n (y_t - \hat{\alpha} - \hat{\beta} y_{t-1})^2.$$

Let  $\theta = (\alpha, \beta)'$  and from (2.23) it is seen that the conditional probability density function of  $\theta$  given  $\sigma_\varepsilon$ ,  $p(\theta | \mathbf{y}, y_0, \sigma_\varepsilon)$  is the bivariate normal distribution as (Zellner, 1971):

$$p(\theta | \mathbf{y}, y_0, \sigma_\varepsilon) \sim N_2(\hat{\theta}, V_{\theta|\sigma_\varepsilon}), \quad (2.24)$$

$$\text{where } \hat{\theta} = (\hat{\alpha}, \hat{\beta})' = \begin{bmatrix} n & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix}, \quad V_{\theta|\sigma_\varepsilon} = \sigma_\varepsilon^2 \begin{bmatrix} n & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1}$$

with the  $\sum$  extending from  $t=1$  to  $t=n$ . However, since  $\sigma_\varepsilon$  is not known, the actual posterior distribution of  $\theta$  is obtained by integrating (2.17) with respect to  $\sigma_\varepsilon$ ,

thus the marginal posterior distribution of  $\theta$  is (Zellner, 1971):

$$p(\theta | \mathbf{y}, y_0) = \int_0^{\infty} p(\theta, \sigma_\varepsilon | \mathbf{y}, y_0) d\sigma_\varepsilon$$

$$\propto \left[ (n-2)\hat{\sigma}_\varepsilon^2 + n(\alpha - \hat{\alpha})^2 + n(\beta - \hat{\beta})^2 \sum_{t=1}^n y_{t-1}^2 + 2(\alpha - \hat{\alpha})(\beta - \hat{\beta}) \sum_{t=1}^n y_{t-1} \right]^{-n/2}, \quad (2.25)$$

which is the equivalent to a bivariate Student  $t$  distribution with mean  $\hat{\theta} = (\hat{\alpha}, \hat{\beta})'$ , and the covariance matrix of  $V_\theta$  with  $n-2$  degree of freedom.

As for the marginal posterior distribution for parameter  $\sigma_\varepsilon$ , it can be obtained by integrating (2.23) with respect to  $\theta$  as (Zellner, 1971):

$$p(\sigma_\varepsilon | \mathbf{y}, y_0) \propto \frac{1}{\sigma_\varepsilon^{n-1}} \exp\left(-\frac{(n-2)\hat{\sigma}_\varepsilon^2}{2\sigma_\varepsilon^2}\right) \text{ for } 0 < \sigma_\varepsilon < \infty, \quad (2.26)$$

which is equivalent with an inverted gamma probability distribution with mean and variance as:

$$E(\sigma_\varepsilon) = \hat{\sigma}_\varepsilon \left(\frac{n-2}{2}\right)^{1/4} \Gamma\left(\frac{n-3}{2}\right) / \Gamma\left(\frac{n-2}{2}\right), \quad \text{Var}(\sigma_\varepsilon) = \frac{n-2}{n-4} \hat{\sigma}_\varepsilon^2 - [E(\sigma_\varepsilon)]^2,$$

where the posterior probability distribution function of the ratio of the known residual total sum of square  $(n-2)\hat{\sigma}_\varepsilon^2$  to the unknown residual variance  $\sigma_\varepsilon^2$  is the standard chi-squared distribution with  $(n-2)$  degree of freedom as  $(n-2)\hat{\sigma}_\varepsilon^2 / \sigma_\varepsilon^2 \sim \chi_{n-2}^2$ . Then, the actual variance of  $\sigma_\varepsilon^2$  could be calculated from  $\sigma_\varepsilon^2 = (n-2)\hat{\sigma}_\varepsilon^2 / \chi_{n-2}^2$  (Box and Tiao, 1973).

### 2.2.3 Bootstrap

Since the bootstrap was introduced by Elfon (1979), this resampling technique

has been widely used in streamflow generation as an alternative to the parametric approach because of its simplicity and efficiency, which might have the ability to reproduce sample statistics without any transformation and selecting of the stochastic model. The empirical probability distribution function of any statistic of concern is generated by resampling either directly the historical data (Vogel and Shallcross, 1996; Lall and Sharma, 1996) or the residuals from a fitted stochastic model (Cover and Unny, 1986, Pereira et al., 1984; Oliveira et al., 1988; Tasker and Dunne, 1997; Srinivas and Srinivasan, 2000, 2005).

By using the bootstrap technique, parameter uncertainty can be incorporated into the synthetic streamflow generation: Cover and Unny (1986) applied the bootstrapping technique to the residuals calculated from a fitted ARMA model, in which different parameters were estimated from generated streamflows by resampling residuals at random with replacement. By minimizing the conditional likelihood function updated by resampled residuals in each realization, new parameter sets are recursively obtained. This procedure was extended to the multivariate periodic streamflow case by Tasker and Dunne (1997). After fitting the model which is assumed appropriate into a standardized historical sample, multi-site residuals are extracted with non-overlapping one year (12 consecutive months) residual blocks, which are contemporaneous across sites. Then, new monthly streamflows are generated based on the assumed model by using resampled residuals and consequently, different parameter sets are generated.

In this analysis, a simple resampling technique, as in Takser and Dunne (1997), will be used to incorporate the parameter uncertainty effect. From given residuals obtained by fitting the AR(1) model into normalized historical streamflow data, a new

residual series with the length of larger than historical flows ( $100+n$  will be used in this analysis to eliminate the initial value effect) are generated by randomly resampling with replacement and then those are plugged into (2.1) to generate a new streamflow series. Different parameter sets are then estimated from the new generated streamflows, which correspond to uncertainty incorporated parameter estimates. The comparative analysis with parametric approaches will be given at the end of this paper.

### 2.3 Theoretical Example of Parameter Uncertainty Effect

To evaluate the effect of parameter uncertainty we examine six cases: (a) no uncertainty is considered, i.e. all parameters are assumed as constant values, (called natural uncertainty); (b) only uncertainty of  $\mu$  is considered; (c) only uncertainty of  $\sigma_\varepsilon^2$  is considered; (d) the uncertainties of  $\mu$  and  $\phi$  are considered; and (e) the uncertainties of all parameters  $\mu$ ,  $\phi$ , and  $\sigma_\varepsilon^2$  are considered.

As a preliminary analysis, the effect of uncertainty in terms of the coefficient of variation  $\eta_X$  and lag-1 serial correlation  $\phi_X$  of the original (non-transformed) series is theoretically evaluated with different values of  $\eta_X$  and  $\phi_X$ ;  $\eta_X = 0.1-2.0$  with an increment of 0.05 and  $\phi_X = 0-0.9$  with an increment of 0.02. Log-transformation without a location parameter is used to prevent complex values in the original domain caused by simulated negatives in the transformed domain for several  $\eta_X$  and  $\phi_X$ 's, when some transformations are applied; e.g. Box-Cox or power transformation. Simulation experiments are conducted assuming  $\mu_X = 10$  and the AR(1) model

parameters of normalized series  $Y_t$  are estimated for the assumed values of  $\eta_X$  and  $\phi_X$ . Four sample sizes will be considered:  $n=25, 50, 75, 100$ , and uncertainties of the parameters  $\mu$ ,  $\phi$ , and  $\sigma_\varepsilon^2$  will be quantified by sampling from different asymptotic distributions in Cases (a)–(e). Based on 10,000 different sets of sampled parameter estimates, 10,000 different synthetic streamflow series are obtained. For each set of synthetic streamflows, a storage capacity is calculated using the sequent peak algorithm as (Loucks et al., 1981):

$$S_t = \max(0, S_{t-1} + D_t - X_t) \quad , \quad t = 1, \dots, N_d, \quad (2.27)$$

where  $D_t$  = water demand,  $X_t$  = reservoir inflow,  $N_d$  = planning horizon, and  $S_0 = 0$ . Then the storage capacity becomes:

$$S_c = \max(S_0, S_1, \dots, S_{N_d}). \quad (2.28)$$

For the purpose of this study we assume that the demand level  $D_t$  is constant as mean annual flows (MAF). However, we considered two options: (i) the historical sample mean, which is assumed to be fixed for all generated samples, this option is labeled *FM*; (ii) the sample means obtained from each generated sample is utilized as the demands, labeled *SM*.

Figure 2.1 illustrates the expected value of storage capacity obtained over the specified ranges of values of  $\eta_X$  and  $\phi_X$  for a sample size of  $n=100$  and demand option *SM*. For  $\eta_X \leq 0.5$ , similar patterns could be observed for the various cases analyzed. Overall, the expected value of storage capacity increases as  $\phi_X$  increases or  $\eta_X$  increases; also the effect of  $\eta_X$  on the expected storage capacity is larger than that of  $\phi_X$ . Figure 2.1 (a)–(e) shows that the effect of the uncertainty of  $\sigma_\varepsilon^2$  seems to be

less important that the effect of the uncertainty of  $\mu$  or  $\phi$ . Remarkably, the effect of the uncertainty of  $\phi$  seems very significant.

Figure 2.1 (d) and (e) illustrate the cases where the effect of  $\phi$  is included. The figure suggests that the effect of the uncertainty of  $\phi$  dominates over the effect of the other two when high values of  $\eta_x$  (say in the range  $\eta_x=1.0-2.0$ ) and high values of  $\phi_x$  (say  $\phi_x>0.5$ ) are combined, in which very large values of the storage capacity may occur. This could occur for example if  $\hat{\phi}$  is sampled from  $\hat{\phi} \sim N(0.8, 0.06^2)$  when  $\phi_x=0.7$ ,  $\eta_x=1.5$ ,  $n=100$ . Sampling from this distribution may lead to values of  $\phi_x$  close to 1 (correspondingly  $\phi$  close to 1), which in turn may produce very large values of streamflows, a large value of the sample mean, and consequently a very large value of the storage capacity. Note the very large values of the storage capacity obtained for Cases (d) and (e). On the other hand, this possibility of very large values of the storage capacity are not found for the other cases, i.e. (b) and (c), where the uncertainties of  $\mu$  or  $\sigma_\varepsilon^2$  are considered. The foregoing results correspond to the demand option *SM*. On the other hand, extremely large storage capacities shown in *SM* are not found for the demand option *FM* (refer Figure 2.A1 in Appendix). However, the storage capacity increases with larger variation and larger serial correlation as similar in *SM*. Except extremely large storage capacities in some ranges (high value of  $\eta_x$  s and high value of  $\phi_x$  s), *FM* produces larger expected values of storage capacities over the whole ranges of  $\eta_x$  and  $\phi_x$ , when compared with *SM*. As expected, Case (e), which includes the effect of the uncertainties in all parameters, gives the largest storage capacities as compared with other cases.

Similar patterns could be found for the generated standard deviations of storage capacities over the different values of  $\eta_x$  and  $\phi_x$  (see Figures 2.A2-2.A3 in the Appendix). However, *SM* produces larger standard deviations of storage capacities over the whole range of  $\eta_x$  and  $\phi_x$ , when compared with *FM*, except extremely large storage capacities in some ranges (high value of  $\eta_x$ s and high value of  $\phi_x$ s).

## 2.4 Simulation of Synthetic Streamflows with Parameter Uncertainty

Two annual streamflow series, one with the high lag-1 serial correlation and the other with the low serial correlation have been chosen to study the effect of parameter uncertainty on synthetic streamflows. The annual flow at Lee's Ferry in the Colorado River Basin has been analyzed where  $\eta_x=0.28$ ,  $\hat{\rho}_x=0.28$ , the skewness coefficient of 0.14, and the kurtosis coefficient of 2.40. Additionally, the annual flow of the St. Lawrence River at Cornwall, Ontario near Massena, NY, has been selected. It has the following characteristics:  $\eta_x=0.09$ ,  $\hat{\rho}_x=0.73$ , the skewness coefficient of -0.28, and the kurtosis coefficient of 2.74.

For both historical sample sets, normalizations have been applied by using power transformations. 10,000 different streamflow sets with the length of historical sample size (98 for Lee's Ferry and 59 for St. Lawrence River) have been generated based on the AR(1) model with different sample sizes:  $n=25, 50, 75, 100$ . In each generation, it was assumed that population parameters are equal to those from historical annual streamflow data. Two different population parameter sets are possible because of different estimators in asymptotic and Bayesian analysis. For comparison, approximate MLEs

adopted in the asymptotic method are implemented as population parameters in Bayesian analysis as well instead of least square estimators.

#### 2.4.1 Sampling of parameter estimates

First, the asymptotic distributions of parameter estimates will be given in the analysis. A measure of relative efficiency of simulation experiments might be useful, which is defined by the ratio of standard error (square root of variance) of generated parameter estimates from asymptotic distributions relative to the theoretically defined asymptotic standard error of corresponding parameter estimates. Since the large number of traces is generated (10,000 traces), the simulated standard error would be expected to be closer to the theoretical one. However, the variances of parameter estimates calculated from simulations are less than theoretically derived ones for moderate or small sample sizes ( $n=25,50$ ) (see Table 2.A1 in Appendix). This non-consistency could be due to restricted ranges of parameters associated with real streamflow generation such that (i)  $0 < \hat{\phi} < 1$ , (ii) generated original streamflows  $X_i^G = T^{-1}(Y_i) > 0$  where  $T^{-1}$  means a back-transformation function and  $Y_i$  denotes generated streamflow in the normal domain. For example, for  $n = 25$  and Case (e),  $P(\hat{\phi} < 0) = 7.4\%$  for Lee's Ferry [ $\hat{\phi} \sim N(0.277, 0.192^2)$ ] and  $P(\hat{\phi} > 1) = 2.3\%$  for the St. Lawrence River [ $\hat{\phi} \sim N(0.720, 0.139^2)$ ]. Probability of having negative streamflows in back-transformation to the original domain is not explicitly given, but it could be numerically counted during simulation: e.g.  $P(X_i^G < 0) = 2.6, 1.3, 1.0, 0.8\%$  with  $n=25, 50, 75, 100$  for Lee's

Ferry data. However, simulated standard errors are shown to be not significantly affected by the limited parameter space for both cases with relatively large sample sizes  $n \geq 75$  and therefore, a theoretical asymptotic distribution might be applicable without being affected by the limited parameter space.

For Bayesian analysis, simulated standard errors of sampled parameter estimates using conditional posterior distribution of (2.24) and (2.26) have been compared with those from asymptotic distribution by using the ratio of calculated standard deviations of parameter, which are illustrated in Table 2.1. Non-informative prior distributions are assumed in the derivation of the posterior distributions of parameters, but during the simulation parameter estimates generated from their distributions are truncated and redistributed when those are not on assumed parameter spaces:  $\hat{\sigma}_\epsilon^2 > 0$ ,  $\hat{\mu} > 0$ , and  $0 < \hat{\phi} < 1$ . Note that  $\hat{\alpha}$  and  $\hat{\phi}(=\hat{\beta})$  are jointly sampled from the Bayesian posterior distribution (bivariate normal). Generated  $\hat{\alpha}$ s are expected to be linearly correlated with generated  $\hat{\phi}$ s (say  $\hat{\alpha} = k\hat{\phi}$ ) of which the relationship could be explained by the off-diagonal element in the variance-covariance matrix of (2.24). Thus, the nonlinear relationship of  $\hat{\mu}$  and  $\hat{\alpha}$  in  $\hat{\mu} = \hat{\alpha} / (1 - \hat{\phi}) = k\hat{\phi} / (1 - \hat{\phi})$  results in an extremely large value of  $\hat{\mu}$  when  $\hat{\phi}$  is getting closer to 1: for example, (1) Case (e),  $\hat{\mu} = 1.22 \times 10^{12}$  for  $\hat{\sigma}_\epsilon^2 = 1.23 \times 10^{17}$ ,  $\hat{\phi} = 0.9996$ ,  $\hat{\alpha} = 4.55 \times 10^7$  (2) Case (d),  $\hat{\mu} = 1.99 \times 10^{10}$  for  $\hat{\sigma}_\epsilon^2 = 7.17 \times 10^{16}$ ,  $\hat{\phi} = 0.999$ ,  $\hat{\alpha} = 1.93 \times 10^7$  [compare  $\hat{\mu} = 2.36 \times 10^9$  for  $\hat{\sigma}_\epsilon^2 = 7.17 \times 10^{16}$ ,  $\hat{\phi} = 0.72$ ,  $\hat{\alpha} = 0.66 \times 10^9$  in Case (a)] for the St. Lawrence River with  $n = 25$ . This effect of high  $\hat{\phi}$  will still be effective for relative moderate sample size ( $n = 50$ ) for the

St. Lawrence River since there exist more chances of having  $\hat{\phi}$  close to 1 due to the high historical parameter estimates ( $\hat{\phi}=0.72$ ). But, in practical hydrologic streamflow generation this very high  $\hat{\phi}$  seems unrealistic. A choice of the appropriate upper bound of  $\hat{\phi}$  to eliminate the unrealistic high  $\hat{\mu}$  can enable one to assure the applicability in real streamflow generation. Thus, another simulation experiment of parameter estimates in a newly defined range of  $\hat{\phi}: 0 \leq \hat{\phi} \leq 0.99$  (assume that  $\hat{\phi}=0.99$  is large enough in the real streamflow generation) is taken into account. For this newly defined range of  $\hat{\phi}$ , the efficiency of  $\hat{\mu}$  is shown to reduce to comparable values when  $n=25$  and 50.

Comparison of asymptotic and Bayesian methods in terms of efficiency shows similar variability (within 10% difference) of  $\hat{\phi}$  for all sample sizes and  $\hat{\sigma}_\epsilon^2$  for  $n \geq 50$  in both low and high serially correlated streamflows sets. When the Bayesian method is used, larger variability of generated  $\hat{\mu}$  is notable in Cases (d) and (e), which are shown for the range of  $n \leq 50$  for Lee's Ferry and all applied sample sizes for the St. Lawrence River. Note that the similar variability (within 10% difference) of  $\hat{\mu}$  between two methods is illustrated in Case (b). As discussed before, this might be caused by the nonlinear relationship of  $\hat{\mu}$  and  $\hat{\phi}$  (especially for high  $\hat{\phi}$ , which is shown in the last two columns in Table 2.1). Proper choice of possible range of  $\hat{\phi}$  in the prior distribution would reduce the unexpected variability of  $\hat{\mu}$ , but in reality it seems hard to define the possible range of  $\hat{\phi}$  because it is uncertain. Alternatively, employing a reliability concept of uncertain parameter estimates might be a useful way to avoid using

the unrealistic, extremely high parameter estimates since the more chance of having extremely larger  $\hat{\phi}$  would be expected for the large number of simulations (e.g. 10,000 in this study). Reliability is associated with quantiles of simulated variables calculated from a large number of traces. This will be discussed later.

Moreover, as shown in Table 2.1, the usage of conditional posterior (Normal) and more realistic posterior ( $t$ ) distributions show the similar efficiency for  $n \geq 50$ , while the conditional posterior distribution shows the larger variability of  $\hat{\mu}$  when  $n=25$ . It might be probable to use either the conditional posterior distribution or the more realistic distribution for  $n \geq 50$ . Without loss of generality, the  $t$  distribution will be used as the posterior distribution for further analysis of the parameter uncertainty effect examination.

#### **2.4.2 Parameter uncertainty effects on basic statistics**

Parameter uncertainty effects on the basic statistics of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from synthetic streamflows of different sample sizes  $n$  will be examined based on both asymptotic distribution and Bayesian posterior distribution, which are listed in Table 2.2. Increased variabilities of distributions of generated means, standard deviations, and lag-1 serial correlations are shown for all sample sizes, which is explained by the effect of parameter uncertainty. However, this is not the case for generated skewness coefficients. In order to compare parameter uncertainty effects for two different historical data, the ratio of increased variability by parameter uncertainty in Cases (b)-(e) relative to natural variability of the

general simulation procedure might be useful.

Generated synthetic streamflows with the uncertain  $\hat{\mu}$  (Cases (b),(d),(e)) show increased variability in the generated mean, but not much in the generated standard deviation, skewness, and lag-1 serial correlation. The uncertainty effect of  $\hat{\phi}$  (Cases (d), (e)) leads to increased variability of the generated lag-1 serial correlation. For the calculated variances of generated streamflows, both uncertain  $\hat{\sigma}_\varepsilon^2$  and  $\hat{\phi}$  make an important impact. In particular  $\hat{\sigma}_\varepsilon^2$  seems more effective in low correlated data, while  $\hat{\phi}$  is more significant in high serially correlated data (for more, see Figures 2.A4-2.A7 in Appendix).

As generally expected, the uncertainty effect would be reduced as the sample size  $n$  increases; for example, for Lee's Ferry and Case (e) when using the asymptotic method, 137% increased standard error of the generated mean in  $n=25$  decreases to 44% in  $n=100$  and a 94% increase in the generated lag-1 correlation in  $n=25$  reduces to 38% in  $n=100$ . Practically,  $n=100$  might be considered as a large sample size in the streamflow generation, but the simulated result shows that the parameter uncertainty effect is still notable in generated mean, standard deviation, and lag-1 correlation in that large sample size. Overall, larger values of ratio can be found for the St. Lawrence River compared with the case of Lee's Ferry. When the sample size is relatively larger ( $n \geq 75$ ), little difference of calculated standard deviation of generated mean and standard deviation might be expected between Bayesian and asymptotic analyses, while larger variability of generated mean is mostly expected by Bayesian analysis for a relatively smaller sample size ( $n \leq 50$ ).

## 2.5 Parameter Uncertainty Effect on Design Variables

### 2.5.1 Uncertainty effect on storage capacity with different sample sizes in two demand level options

Based on two different categories of demand levels, *FM* and *SM*, parameter uncertainty effects incorporated by asymptotic and Bayesian methods on the simulated storage capacities  $S_C^*$  are shown in Figure 2.2 and 2.3, where  $S_C^*$  are calculated from synthetic streamflows with the planning horizon  $N_d$  assumed as the historical sample size (98 for Lee's Ferry, 59 for St. Lawrence River, respectively) with different sample size  $n$ . When *FM* is used, the effect of uncertain  $\hat{\mu}$  on  $S_C^*$  is shown quite significant in all sample sizes. Mean parameter uncertainty expands the interdecile range between 25% and 75% quantiles. It also results in very high 90 and 99% quantiles of generated storage capacities, which is still notable even in a relatively large sample size  $n=100$ . The magnitude of the storage capacity is generally based on the deficit of generated streamflows from the specified demand level. Thus, if new generated streamflows are distributed with an averaged value less than the historical mean, a larger amount of the storage capacity would be expected than the historical one. Therefore, incorporation of the uncertain  $\hat{\mu}$  into the simulation results in the upward shift of the distribution, which is caused by the large number of generated high storage capacities. On the other hand, uncertain  $\hat{\phi}$  affects the increased variability of  $S_C^*$

significantly in option *SM*. Overall, the increased magnitude of  $S_C^*$  by parameter uncertainty is not very notable in option *SM* than in *FM*.

As in generated basic statistics, Bayesian analysis shows similar variability of  $S_C^*$  as asymptotic analysis when  $n \geq 75$  for Lee's Ferry data. The calculated upper quantiles of  $S_C^*$  (especially 90% or higher quantiles) from Bayesian analysis are larger than ones from the asymptotic method in small sample size ( $n \leq 50$ ). The difference between asymptotic and Bayesian methods is notable in Cases (d) and (e) with option *FM*, which again results from the different variability of  $\hat{\mu}$  for a small sample size.

A comparison by using relative bias (*RBIAS*) and relative root mean square error (*RRMSE*) of generated storage capacities  $Cv(S_C^*)$  based on two methods is given in Table 2.3. *RRMSE* is defined as the square root of mean square error (*MSE*) relative to the certain quantity of concern (historical storage capacity  $S_C^H$  herein). *MSE* is defined as  $MSE(S_C^*) = Var(S_C^*) + (E(S_C^*) - S_C^H)^2$ , where  $E(S_C^*) - S_C^H$  is the bias of  $S_C^*$ . *RBIAS* and *RRMSE* are defined as  $RBIAS = (E(S_C^*) - S_C^H) / S_C^H$  and  $RRMSE = \sqrt{MSE} / S_C^H$ . For Lee's Ferry with option *FM*, both asymptotic and Bayesian methods give almost similar *RRMSE* for all sample sizes in Case (b), while the Bayesian method shows larger *RRMSE* in Cases (d) and (e) for  $n \leq 50$ . The Bayesian method shows larger *RRMSE* (greater than 10% difference between two methods) even if  $n=75$  for the St. Lawrence River data, which results from the effect of uncertain  $\hat{\phi}$  and  $\hat{\mu}$  on the generated mean which is still considerable for larger  $n$  as discussed before. However, when  $n=100$ , the difference is getting into the range less than 10%. For option *SM*, the two methods

show little difference in variation of  $S_C^*$ . Additionally, note that at least 35% more increased  $RRMSE$  still exists in Case (e) even when  $n=100$  for both low and high serially correlated data regardless of asymptotic and Bayesian analyses.

### 2.5.2 Uncertainty effect on storage capacity for different demand level

The behavior of a water supply system can be classified by using the Hurst's standardized net inflow parameter  $m$ , defined by (Hurst, 1951):

$$m = \frac{1 - \alpha}{\eta_X}, \quad (2.29)$$

where  $\alpha$  is the annual demand given by the ratio of demand to mean annual flows (MAF) and  $\eta_X$  is the coefficient of variation of annual streamflows. This index gives a measure of the degree to which the designed storage capacity depends on monthly or annual storage requirements (Vogel et al., 1999). Literature suggested that as long as  $0 < m \leq 1$ , the system will be behave like an over-year system, while the system is dominated by within-year behavior if  $m > 1$  (Vogel and Stedinger, 1987). Based on this,  $\alpha > 0.71$  or  $\alpha > 0.91$  will be respectively required to ensure the over-year regulation of the assumed reservoir for two different historical annual streamflows of Lee's Ferry in the Colorado River Basin ( $\eta_X = 0.29$ ) and the St. Lawrence River ( $\eta_X = 0.09$ ). For small  $\alpha$  (e.g.  $\alpha = 0.5$ ), the generation of  $S_C$  would be allowed in the monthly scale instead of the annual, thus  $\alpha = 1, 0.9, 0.8, 0.7$  at the Lee's Ferry station and  $\alpha = 1, 0.975, 0.95, 0.925$  at the St. Lawrence River are assumed as fractions of MAF in demand levels. Generally, the distribution of  $S_C^*$  shows the upward shift from the historical  $S_C$  for a

given  $\alpha$ , but in some cases of  $\alpha \approx 1$  in low correlated streamflows,  $S_C^*$  can give the positive bias for both options of *FM* and *SM* (refer to Figure 2.A8 in Appendix). Comparison between *FM* and *SM* will be given by using *RBIAS* and *RRMSE*. Compared with *SM*, *FM* gives an upward shift of *RBIAS* and larger *RRMSE* over different demand levels. Since *RRMSE* tends to increase as  $\alpha$  decreases, much variability of storage capacity will be expected when the parameter uncertainty is incorporated into the simulation with smaller fraction of MAF as a demand level in the calculation of the storage capacity.

It is generally expected that the variability of  $S_C^*$  decreases as  $\alpha$  decreases (see Figures 2.A9 and 2.A10 in Appendix). Table 2.4 illustrates that the uncertain  $\hat{\mu}$  and  $\hat{\phi}$  are more significant on the variability of  $S_C^*$  than  $\hat{\sigma}_\epsilon$  regardless of different demand levels. However, taking  $\alpha$  by a smaller value makes the mean uncertainty effect become less significant. The effect of uncertain  $\hat{\phi}$  gets more significant as  $\alpha$  decreases, which could be judged by *RBIAS* and *RRMSE* in Table 2.4. To be notable, the difference in the variability of  $S_C^*$  between asymptotic and Bayesian analyses becomes more with smaller  $\alpha$  in option *FM*.

### 2.5.3 Uncertainty effect on storage capacity with reliability

When the generated  $S_C$  is used to determine the size of the reservoir of concern, it is required to assure the failure-free operation over  $N_d$  design period with the probability  $p$ . Steady state reliability  $q$  of selected storage capacity  $\hat{S}_C$  can be

defined as the probability  $q = 1 - p$  with which the reservoir is able to carry the yield for next year without failure. Empirical distributions of  $S_C^*$  in Cases (a)-(e) are plotted, which are obtained from the 10,000 synthetic flow traces with 100 years of design period for different sample sizes, two demand options *FM* and *SM*, and  $\alpha = 1$  based on both the Colorado River and the St. Lawrence River (Figures 2.4-2.5, 2.A11-2.A12).

For a given reliability expressed by a quantile of  $S_C^*$ , parameter uncertainty incorporation increases the magnitude of  $S_C^*$ , and in option *FM*, this effect is still visible even in a relatively large sample size, say  $n=100$ . For smaller  $n$ , the uncertainty effect is remarkably found even in relatively small reliability; for example, consider the quantiles of  $S_C^*$  for a reliability 90% ( $q=0.9$ ) based on Lee's Ferry data with  $n=50$  and *FM*. Increased magnitude of  $S_C^*$  by uncertain parameters is given by 9.8MAF for the asymptotic method (Case (e)), which is a 44% increment compared with the case of natural uncertainty (Case (a), 6.8MAF). That is, in order to yield the fail-free operation in the next 10 years of the design period, storage capacities might be required more than at least 3MAF as a result of parameter uncertainty (for more, see Tables 2.5 and 2.A2). When Bayesian analysis is used, additional increase of  $S_C^*$  compared with asymptotic analysis can be expected for small sample size in option *FM*, which is caused by the effect of uncertain  $\hat{\mu}$  combining with uncertain  $\hat{\phi}$ . This difference between asymptotic and Bayesian methods is much distinguishable at a greater degree for smaller reliability for smaller  $n$  and for the low serial correlated sample. In high correlated data, the Bayesian method still shows the large increment of  $S_C^*$  by uncertain  $\hat{\phi}$  when  $n=100$  even for a small reliability (e.g.  $q \approx 0.9$ ). As previously discussed, *SM* shows

only the effect of uncertain  $\hat{\phi}$  on  $S_C^*$ , and for high correlated data this effect still seems significant for larger reliability ( $q \geq 0.99$ ) even if  $n$  becomes larger.

Uncertainty effect associated with quantiles is additionally examined with the different demand levels (Figures 2.A13-2.A15 in Appendix). For *FM*, significant parameter uncertainty effects can be shown in lower reliability when the demand level is closer to the historical mean, but those become visible in higher reliabilities as the demand level decreases. On the other hand, a consistent pattern of distributions of  $S_C^*$  over different demand levels is shown for *SM*.

#### **2.5.4 Drought related Statistics**

As another tool for evaluating parameter uncertainty effect, drought related statistics will be utilized. For a given demand level (assumed as MAF), the deficit occurs when annual streamflows  $Y_t < \text{MAF}$  during one or more years until  $Y_t > \text{MAF}$  again, which can be defined by its magnitude, by its duration, and by its intensity (magnitude divided by corresponding duration). Among the number of deficits in a given streamflows, the maximum deficit magnitude, length and intensity in a given sample are referred to the critical drought length  $cdl$ , critical maximum magnitude  $cdm$ , and critical drought intensity  $cdi$  (Salas et al., 1980).

For two different annual streamflow data, 10,000 different traces of synthetic streamflows with the length of the historical flows of each station have been generated by incorporating parameter uncertainties based on different sample sizes similar in the case of storage capacity. After this, critical drought statistics are calculated and compared

between two demand options (*FM*, *SM*) and between asymptotic and Bayesian method in Figures 2.6-2.7 and 2.A17-2.A20.

It can be shown that parameter uncertainty effects on simulated *cdm* and *cdl* seem very similar, as in the case of storage capacity. When *FM* is used, uncertain  $\hat{\mu}$  makes an important impact on upper quantiles of *cdm*\* and *cdl*\*, while the uncertainty of  $\hat{\phi}$  significantly affects on them for the case of *SM*. This effect seems notable in small *n* for Lee's Ferry but still visible even in *n*=100 for the St. Lawrence River. Also, the Bayesian method shows the higher upper quantiles of *cdm*\* and *cdl*\* by uncertain  $\hat{\phi}$  for moderate or smaller *n* (say *n* ≤ 50) for both Lee's Ferry and St. Lawrence River. However, uncertain parameter estimates which are significant on *cdi*\* seem to be dependent with the sample characteristics: e.g.  $\hat{\sigma}_\epsilon^2$  for Lee's Ferry and  $\hat{\phi}$  for the St. Lawrence River. For Lee's Ferry, the effect of  $\hat{\sigma}_\epsilon^2$  is observable in small *n* but the increased *cdi*\* is not quite significant compared with *cdm*\* and *cdl*\* for the whole.

### 2.5.5 Brief summary of parameter uncertainty effect on design variables

Table 2.6 illustrates the brief summary of the parameter uncertainty effect on synthetic storage capacity and critical drought magnitude obtained from generated annual streamflows at Lee's Ferry and at the St. Lawrence River for option *FM*. For simplicity, Case (a) (no parameter uncertainty, NU) and Case (e) (all parameter uncertainty PU) are numerically compared. For example, compared with NU, PU shows increased expected

value of the storage capacity with a range of from about 17% up to 72% depending on different demand levels and sample serial correlations. The parameter uncertainty effect is more visible for standard deviation and higher quantiles of storage capacity; e.g. about a 280% increase of the 99% quantile of the storage capacity for the case of the St. Lawrence River when using 90% MAF demand level. Overall, less increase by the parameter uncertainty effect is illustrated in the generated critical drought magnitude compared with the generated storage capacity, except that a similar pattern of increased statistics and quantiles by parameter uncertainty over different demand levels and sample serial correlations could be found for both design variables.

## **2.6 Comparative Analysis with Nonparametric Technique**

For two different sets with different sample sizes ( $n=50, 98$  for Lee's Ferry and  $n=30,59$  for St. Lawrence River), synthetic streamflows with the length of historical sample have been generated with the parameter uncertainty incorporated based on natural uncertainty (NU), asymptotic (AS), Bayesian (BA), and bootstrapping (BS). The BS does not enable one to incorporate the uncertainty of each parameter estimates separately because of its structure. For this reason, the uncertainty of all parameter estimates are taken into account (Case (e)) in the generation based on asymptotic and Bayesian methods. Some distortions of generated parameter estimates are found in BS compared with AS and BA since estimated  $\hat{\phi}$  in BS is based not on the statistical structure but only on resampled residuals of the historical sample (see Figure 2.A21 in the Appendix). Less variability of generated basic statistics (mean and standard deviation of synthetic

streamflows) in method BS is notable for both Lee's Ferry and the St. Lawrence River, especially in the case of simulating the larger design period than given sample sizes as shown in Figure 2.8. By plotting generated design variables of the storage capacity, critical drought magnitude, length, and intensity with different reliabilities (see Figure 2.9) the difference among AS, BY, BS has been investigated. BS shows similar performance as in the case of natural uncertainty, and it is shown to be incapable of demonstrating the parameter uncertainty effect on generated design variables.

## 2.7 Concluding Remarks

Overall, the uncertainty effect of mean parameter  $\hat{\mu}$  and AR(1) parameter  $\hat{\phi}$  make a significant impact on the generated streamflow statistics and related design variables. An uncertain  $\hat{\mu}$  shows significant effect on the generated mean and does an uncertain  $\hat{\phi}$  on the generated serial correlation. Storage and drought related design variables have been employed as design variables on which the sample mean uncertainty causes enlargement of the variability, particularly increasing the upper quantiles when the historical mean is used as a demand level. When using the simulated mean as an alternative of a demand level, increased variability of the design variable is affected by the uncertain AR(1) parameter  $\hat{\phi}$ . Distinguishable effect on design variables by uncertain parameters is still visible even if the sample size is 100 in option *FM*. The parameter uncertainty effect is related with the different fraction of demand levels: the mean uncertainty effect becomes less significant, while uncertainty of  $\hat{\phi}$  becomes more

significant for smaller demand levels.

Asymptotic and Bayesian show similar variability when incorporating the uncertain  $\hat{\mu}$  only. The Bayesian method causes the increased variability of the design variable when combining uncertain  $\hat{\mu}$  with uncertain  $\hat{\phi}$ , as in Cases (d) and (e). This difference is notable for the relatively small sample size for Lee's Ferry, but for all sample sizes for the St. Lawrence River, which might resulted from the different historical sample serial correlations. Since  $\hat{\mu}$  is actually correlated with  $\hat{\phi}$ , the Bayesian method seems more applicable in reality, but unrealistic  $\hat{\mu}$  might be occasionally expected when  $\hat{\phi}$  close to 1 is generated in the Bayesian method. In calculation of storage and drought related statistics, this high value of  $\hat{\mu}$  might not be significant when using the historical sample as a demand since the deficit of synthetic streamflows is mostly of concern in the calculation of those variables. However, the appropriate choice of the parameter space of  $\hat{\phi}$  can give a better performance in simulation. Asymptotic analysis for  $n \geq 100$  is recommended for incorporating parameter uncertainty into the generation of streamflows. Using a bootstrap as an alternative for incorporating the parameter uncertainty shows similar variability as natural uncertainty and the notable parameter uncertainty effect on design variables would be less expected by the bootstrap technique compared with a parametric approach.

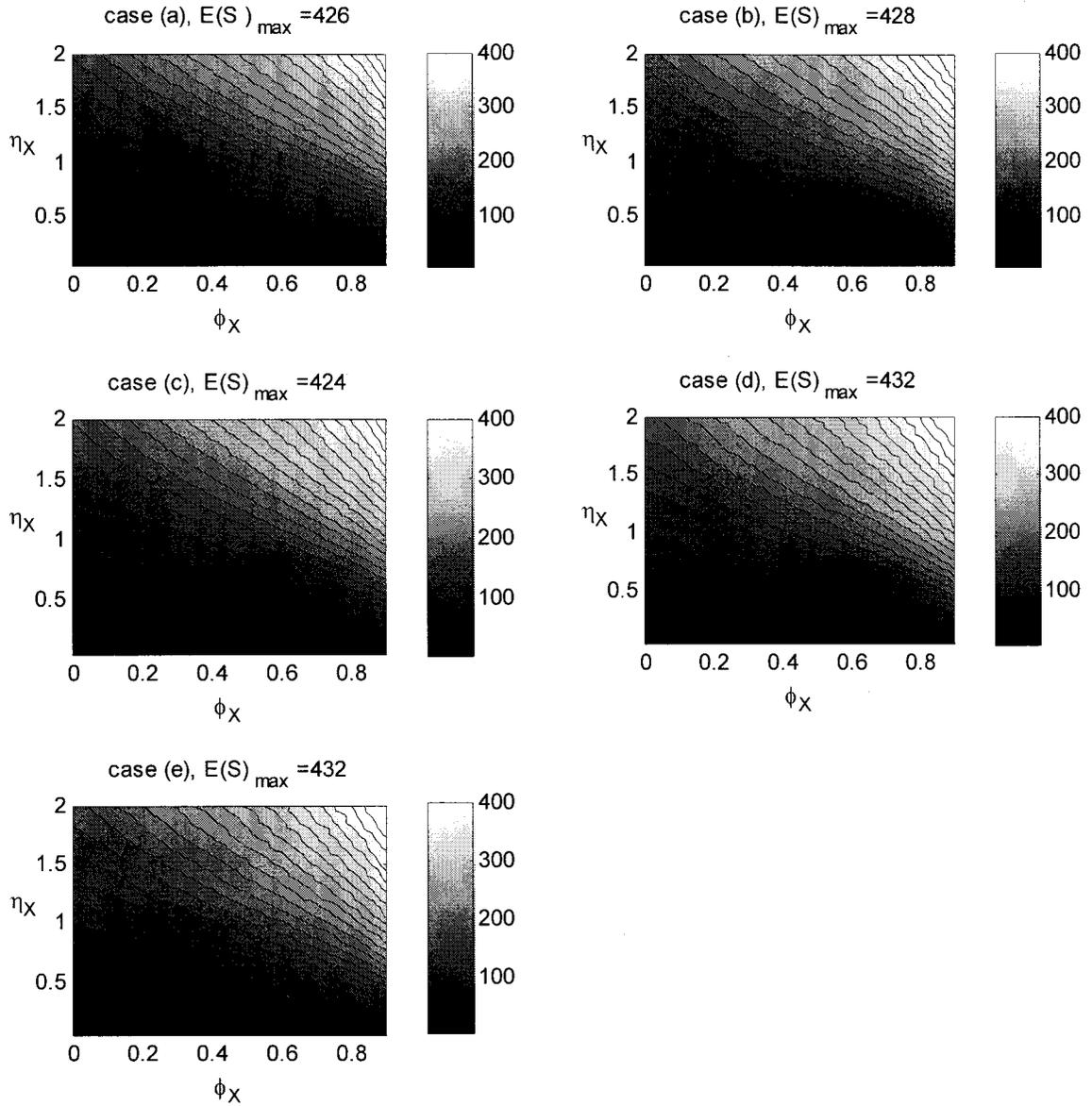


Figure 2.1: Expected values of generated storage capacities calculated from 10,000 different traces simulated from theoretically assumed streamflows with different coefficient of variation  $\eta_X$  (0.1-2) and lag-1 serial correlation  $\phi_X$  (0-0.9). Historical mean ( $FM$ ) is used as a demand level in the calculation of storage capacity by using SPA with demand level  $\alpha = 100\%FM$ . Calculated maximum expected values are (a) 426, (b) 428, (c) 424, (d) 432, and (e) 432.

Table 2.1: Efficiency  $Eff(\cdot) = BSTD(\cdot) / ASTD(\cdot)$ : ratio of simulated standard errors of sampled parameter estimates using Bayesian distribution,  $BSTD(\cdot)$  relative to ones using asymptotic distribution,  $ASTD(\cdot)$  calculated from normalized annual streamflows of Lee's Ferry and St. Lawrence River.

Case	Eff( $\hat{\phi}$ )		Eff( $\hat{\sigma}_\epsilon^2$ )		Eff( $\hat{\mu}$ )			maximum $\hat{\phi}$		
	(d)	(e)	(c)	(e)	(b)	(d)	(e)	(d)	(e)	
<i>n</i>	Lee's Ferry in the Colorado River Basin, <i>Conditional posterior distribution</i> ( $0 \leq \hat{\phi} < 1$ )									
25	1.01	1.01	1.11	1.06	0.99	1.36	1.22	0.99999	0.91	
50	1.02	1.01	1.06	1.05	1.00	1.07	1.09	0.78	0.88	
75	1.00	1.00	1.03	1.04	0.98	1.03	1.05	0.67	0.77	
100	0.99	0.99	1.02	1.03	1.00	1.04	1.03	0.69	0.63	
<i>n</i>	Lee's Ferry in the Colorado River Basin, <i>Conditional posterior distribution</i> ( $0 < \hat{\phi} \leq 0.99$ )									
25	1.01	1.01	1.11	1.06	0.99	1.36	1.22	0.98	0.91	
50	1.02	1.01	1.06	1.05	1.00	1.07	1.09	0.78	0.88	
75	1.00	1.00	1.03	1.04	0.98	1.03	1.05	0.67	0.77	
100	0.99	0.99	1.02	1.03	1.00	1.04	1.03	0.69	0.63	
<i>n</i>	Lee's Ferry in the Colorado River Basin, <i>Actual posterior distribution</i> ( $0 \leq \hat{\phi} \leq 0.99$ )									
25	1.02	1.03	1.10	1.11	1.01	1.41	1.54	0.97	0.98	
50	1.03	1.02	1.07	1.06	1.02	1.12	1.08	0.81	0.83	
75	1.00	1.00	1.04	1.04	1.01	1.04	1.04	0.72	0.71	
100	1.00	0.99	1.03	1.01	1.00	1.04	1.03	0.69	0.62	
<i>n</i>	St. Lawrence River, <i>Conditional posterior distribution</i> ( $0 \leq \hat{\phi} < 1$ )									
25	0.93	0.92	1.15	1.12	0.99	23.98	63.41	0.999	0.9999	
50	0.97	0.98	1.06	1.07	1.00	1.55	7.02	0.997	0.999	
75	0.98	0.99	1.04	1.05	1.02	1.24	1.21	0.99	0.97	
100	0.98	0.99	1.03	1.02	1.00	1.12	1.13	0.97	0.95	
<i>n</i>	St. Lawrence River, <i>Conditional posterior distribution</i> ( $0 \leq \hat{\phi} \leq 0.99$ )									
25	0.91	0.93	1.13	1.11	1.00	1.84	1.78	0.99	0.99	
50	0.96	0.98	1.06	1.05	1.00	1.48	1.48	0.99	0.99	
75	0.98	0.99	1.06	1.04	1.00	1.20	1.22	0.98	0.98	
100	0.98	1.00	1.03	1.02	1.00	1.15	1.14	0.98	0.97	
<i>n</i>	St. Lawrence River, <i>Actual posterior distribution</i> ( $0 \leq \hat{\phi} \leq 0.99$ )									
25	0.95	0.93	1.15	1.12	1.00	2.22	1.94	0.99	0.99	
50	0.98	0.98	1.06	1.04	1.03	1.48	1.47	0.99	0.99	
75	0.99	0.99	1.05	1.04	1.03	1.23	1.23	0.99	0.99	
100	1.00	0.99	1.02	1.03	1.00	1.16	1.14	0.98	0.96	

Note:  $\hat{\phi}=0.277$ ,  $\hat{\sigma}_\epsilon^2=122,768$ ,  $\hat{\mu}=549,101$  for Lee's Ferry,  $\hat{\phi}=0.720$ ,  $\hat{\sigma}_\epsilon^2=267,848,289$ ,  $\hat{\mu}=2,324,977,641$  for St. Lawrence River

Table 2.2: Ratio of standard deviation of generated basic statistics in cases with parameter uncertainty incorporated relative to natural uncertainty case.

$n$		Asymptotic				Bayesian			
		(b)	(c)	(d)	(e)	(b)	(c)	(d)	(e)
Lee's Ferry in the Colorado River Basin									
mean	25	2.26	1.01	2.36	2.37	2.32	1.06	3.29	3.74
	50	1.74	1.01	1.79	1.77	1.74	1.03	1.93	1.85
	75	1.54	1.00	1.57	1.59	1.55	1.01	1.60	1.60
	100	1.42	0.99	1.43	1.44	1.42	0.99	1.47	1.46
Std	25	1.02	2.05	1.52	2.39	1.02	2.25	1.82	2.78
	50	1.03	1.65	1.23	1.76	1.01	1.70	1.23	1.84
	75	1.02	1.48	1.15	1.57	1.01	1.49	1.12	1.59
	100	0.99	1.33	1.09	1.41	1.01	1.37	1.09	1.43
skew	25	1.01	1.00	1.03	1.03	1.00	1.00	1.02	1.05
	50	1.00	1.00	1.00	1.02	1.02	1.02	1.03	1.04
	75	0.99	0.98	1.01	1.00	0.99	0.99	1.00	1.00
	100	1.00	0.99	1.00	1.00	1.00	1.01	1.02	1.02
Lag-1 corr.	25	0.99	0.99	1.93	1.94	1.01	1.00	1.94	1.96
	50	1.02	1.01	1.66	1.62	1.01	1.00	1.66	1.65
	75	1.00	0.99	1.47	1.47	0.99	1.00	1.48	1.48
	100	1.01	1.00	1.37	1.38	1.00	0.99	1.38	1.35
St. Lawrence River									
mean	25	2.21	0.99	2.64	2.73	2.33	1.07	3.96	3.84
	50	1.77	1.02	2.00	2.01	1.77	1.03	2.35	2.38
	75	1.55	1.00	1.68	1.78	1.54	1.01	1.87	1.87
	100	1.42	1.01	1.51	1.53	1.43	1.01	1.66	1.62
Std	25	1.06	1.52	2.47	2.77	1.05	1.63	2.24	2.65
	50	1.05	1.31	1.95	2.07	1.04	1.36	1.86	2.07
	75	1.02	1.18	1.59	1.78	1.03	1.23	1.63	1.79
	100	1.02	1.14	1.41	1.54	1.04	1.18	1.48	1.59
skew	25	1.00	1.01	1.05	1.04	1.02	1.03	1.05	1.07
	50	1.01	1.00	1.03	1.05	1.00	1.00	1.03	1.05
	75	0.99	0.99	1.00	1.01	1.03	1.01	1.04	1.03
	100	0.99	1.03	1.02	1.01	1.00	1.01	1.01	1.02
Lag-1 corr.	25	1.01	1.00	1.90	1.91	0.99	0.98	1.91	1.89
	50	1.00	1.01	1.58	1.60	1.02	1.01	1.58	1.57
	75	1.01	0.99	1.39	1.39	0.99	1.00	1.40	1.41
	100	0.98	0.98	1.29	1.31	1.00	1.00	1.32	1.33

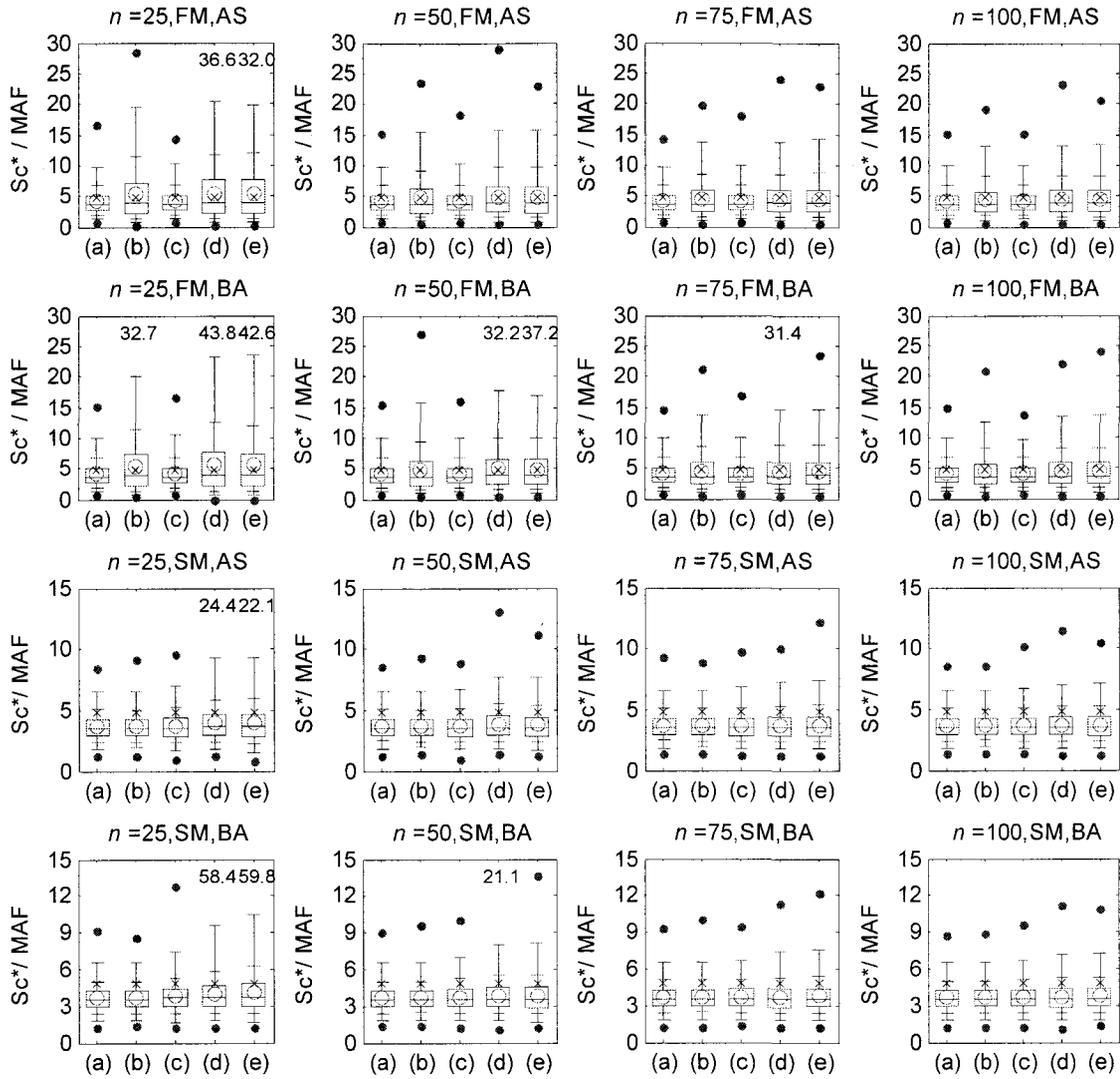


Figure 2.2: Distribution of generated storage capacities  $Sc^*$  scaled by mean annual flows (MAF) for different sample sizes, at Lee's Ferry in the Colorado River Basin, where FM: fixed mean, SM: simulated mean, AS: asymptotic analysis, BA: Bayesian analysis,  $n$  denotes sample size, and (a)-(e) in each subplot represent different parameter uncertainties considerations (design period  $N_d=98$ ). In box plots, the upper, middle and lower line in the box represent 75, 50, 25% quantiles and from the box the whiskers extend to 90,99 and 10,1% quantiles in the upper and lower sides. Two dots outside box represent the maximum and minimum values, 'O' represents the averaged value, and 'X' represents historical storage capacity. The maximum value which is larger than the vertical axis limit is expressed by the number.

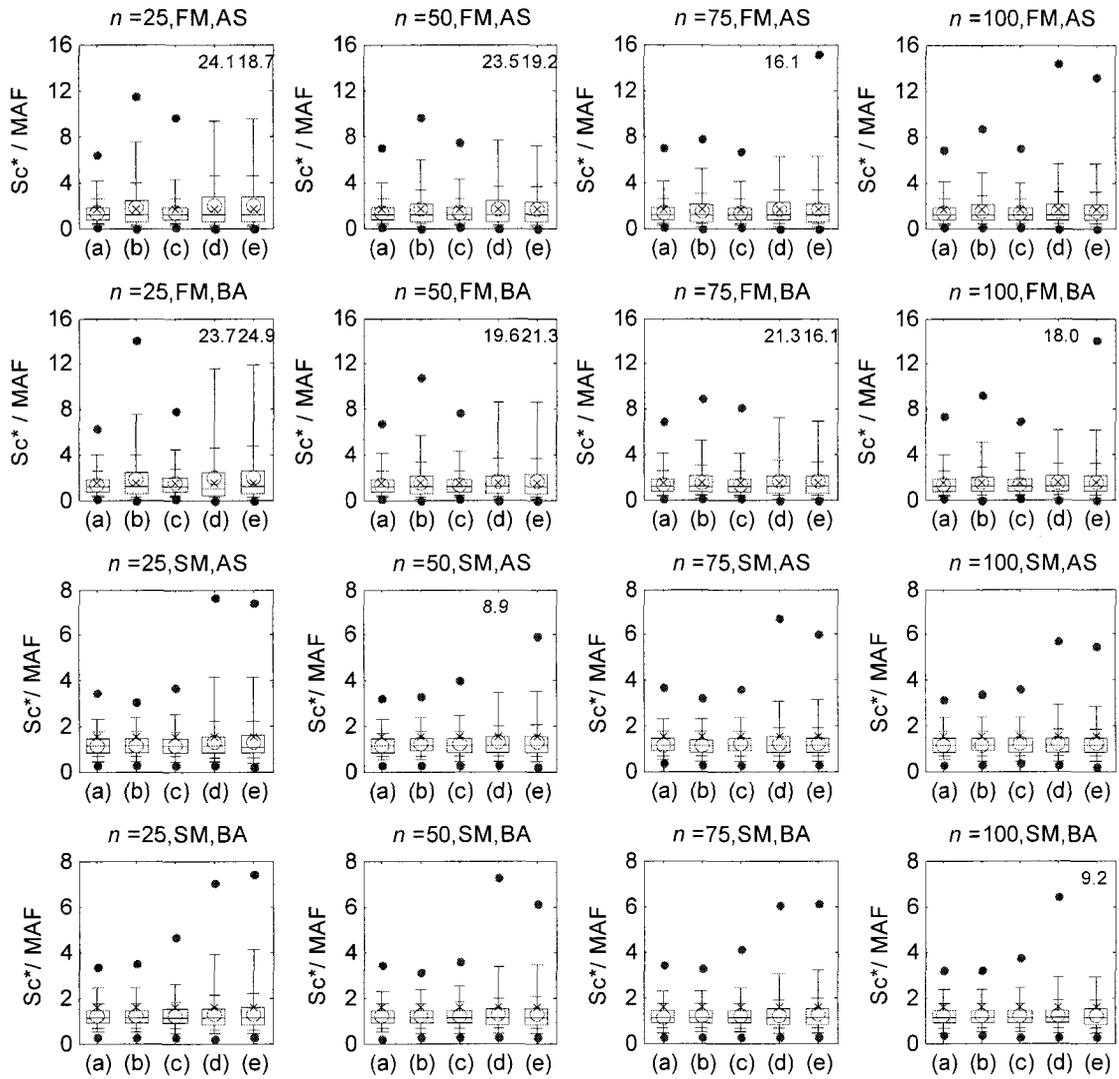


Figure 2.3: Distribution of generated storage capacities  $Sc^*$  scaled by mean annual flows (MAF) for different sample sizes, at St. Lawrence River (design period  $N_d=59$ ).

Table 2.3: Relative BIAS and relative RMSE of generated storage capacities for different sample sizes (demand level  $\alpha=100\%$ ).

	$n$	RBIAS (%)					RRMSE (%)				
		(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
Lee's Ferry in the Colorado River Basin ( $Nd=98$ )											
<i>AS,FM</i>	25	-15	9	-15	14	15	42	89	44	94	94
	50	-15	-2	-15	1	1	42	68	44	71	70
	75	-15	-5	-15	-5	-4	43	60	43	60	62
	100	-14	-7	-15	-6	-6	43	56	43	57	58
<i>BA,FM</i>	25	-15	11	-14	19	16	43	92	44	105	104
	50	-15	-2	-14	4	2	43	69	43	77	74
	75	-15	-7	-15	-4	-4	43	60	43	63	63
	100	-14	-8	-15	-7	-7	42	56	43	58	58
<i>AS,SM</i>	25	-24	-24	-24	-17	-17	32	32	34	37	38
	50	-23	-24	-24	-21	-21	32	32	33	34	34
	75	-23	-23	-23	-22	-22	32	32	32	33	34
	100	-23	-23	-23	-22	-22	32	32	32	33	33
<i>BA,SM</i>	25	-24	-24	-21	-16	-14	32	32	33	40	43
	50	-24	-23	-22	-20	-20	32	32	32	34	34
	75	-24	-23	-23	-22	-21	32	32	32	33	33
	100	-23	-24	-23	-22	-22	32	32	32	33	33
St. Lawrence River ( $Nd = 59$ )											
<i>AS,FM</i>	25	-12	11	-13	23	24	58	106	60	135	133
	50	-14	0	-13	9	8	57	84	59	104	101
	75	-13	-5	-12	3	1	57	74	58	88	88
	100	-13	-6	-13	-3	-2	57	69	58	79	79
<i>BA,FM</i>	25	-12	12	-8	22	29	57	108	61	153	160
	50	-12	0	-11	8	9	57	84	60	112	115
	75	-14	-5	-12	4	4	58	75	58	97	97
	100	-13	-7	-12	-1	-1	57	70	57	86	85
<i>AS,SM</i>	25	-25	-25	-25	-17	-16	37	37	38	51	53
	50	-25	-25	-25	-19	-19	37	36	38	45	45
	75	-25	-25	-25	-21	-22	36	37	37	41	42
	100	-25	-25	-24	-22	-23	37	37	37	40	40
<i>BA,SM</i>	25	-25	-25	-22	-19	-16	36	37	37	48	51
	50	-24	-25	-24	-20	-19	36	37	37	44	44
	75	-25	-25	-24	-21	-20	36	36	37	41	42
	100	-25	-25	-24	-22	-22	36	37	37	40	40

Table 2.4: Relative BIAS and relative RMSE of generated storage capacities for different demand levels ( $n=50$ ).

	$\alpha$	RBIAS (%)					RRMSE (%)				
		(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
Lee's Ferry in the Colorado River Basin ( $Nd=98$ )											
AS,FM	1.0	-15	-2	-15	1	1	42	68	44	71	70
	0.9	56	73	56	86	83	88	136	92	159	153
	0.8	45	51	45	64	62	76	95	82	125	121
	0.7	65	69	65	83	82	99	111	105	144	144
BA,FM	1.0	-15	-2	-14	4	2	43	69	43	77	74
	0.9	56	75	60	91	90	88	141	95	180	173
	0.8	45	53	49	68	69	75	100	85	143	142
	0.7	65	70	71	87	92	98	114	111	157	168
AS,SM	1.0	-23	-24	-24	-21	-21	32	32	33	34	34
	0.9	49	48	48	58	56	71	69	74	91	91
	0.8	42	40	42	51	50	67	66	73	90	91
	0.7	62	61	62	74	73	92	91	99	118	121
BA,SM	1.0	-24	-23	-22	-20	-20	32	32	32	34	34
	0.9	48	49	53	59	61	70	71	79	93	97
	0.8	41	41	46	52	56	66	68	78	93	100
	0.7	62	62	69	75	81	91	93	106	121	134
St. Lawrence River ( $Nd = 59$ )											
AS,FM	1.0	-13	-6	-13	-3	-2	57	69	58	79	79
	0.975	-20	-13	-20	-8	-8	61	73	62	88	88
	0.950	-29	-23	-28	-16	-16	64	74	65	94	93
	0.925	-27	-20	-25	-11	-10	72	82	74	113	112
BA,FM	1.0	-13	-7	-12	-1	-1	57	70	57	86	85
	0.975	-20	-14	-19	-5	-6	61	75	62	99	96
	0.950	-29	-23	-27	-12	-13	63	76	65	108	104
	0.925	-27	-20	-24	-6	-6	70	84	73	136	128
AS,SM	1.0	-25	-25	-24	-22	-23	37	37	37	40	40
	0.975	-31	-31	-30	-27	-27	45	45	45	49	50
	0.950	-38	-37	-37	-33	-33	52	52	53	58	58
	0.925	-36	-34	-34	-29	-29	58	58	59	68	68
BA,SM	1.0	-25	-25	-24	-22	-22	36	37	37	40	40
	0.975	-31	-31	-30	-27	-26	45	45	45	50	50
	0.950	-38	-37	-36	-32	-32	52	52	52	58	59
	0.925	-35	-35	-33	-28	-27	57	58	59	69	71

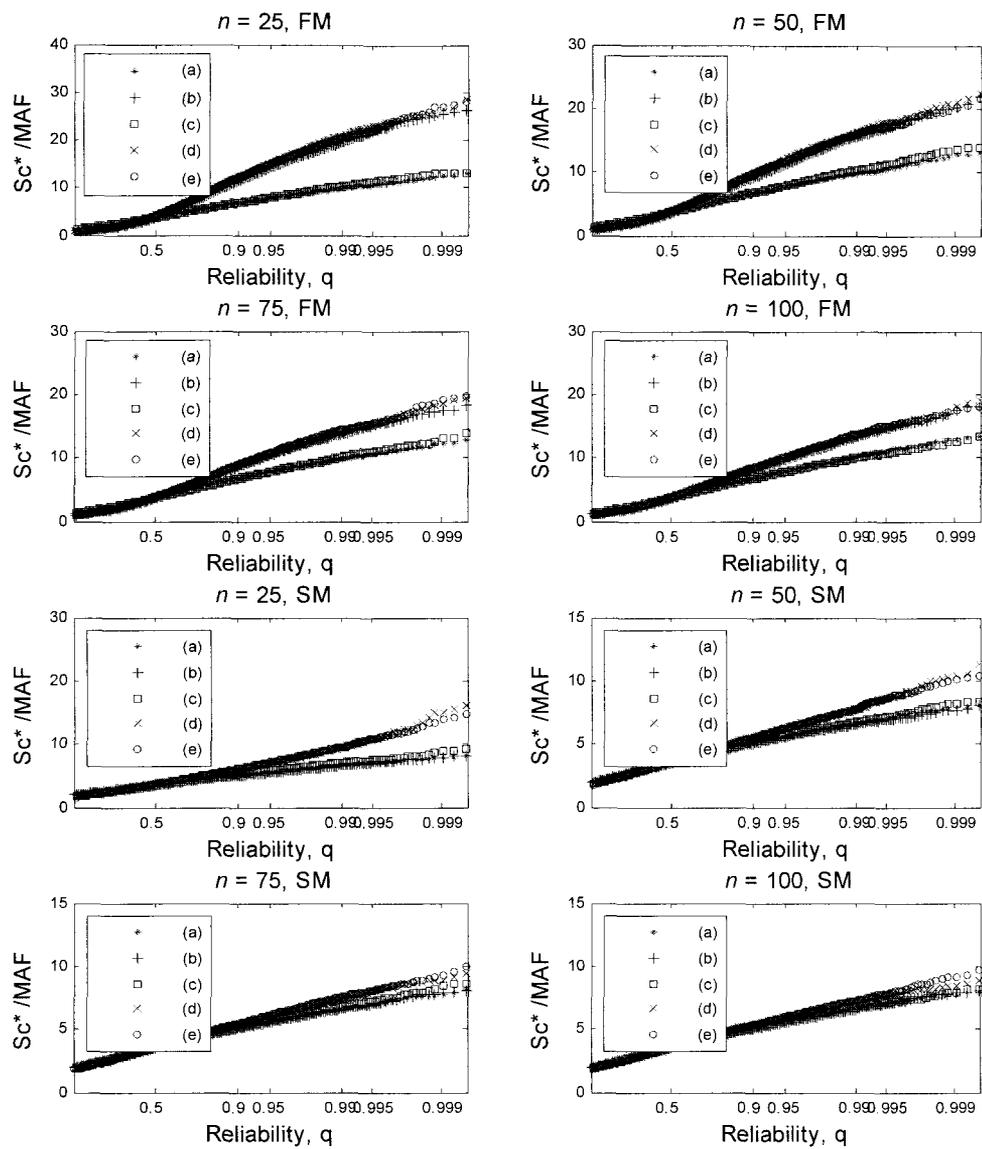


Figure 2.4: Quantile plots of simulated storage capacities  $Sc^*$  for different sample sizes where reliability denotes nonexceedance probability associated with the corresponding quantile (Lee's Ferry in the Colorado River Basin, asymptotic analysis,  $N_d=100$ ).

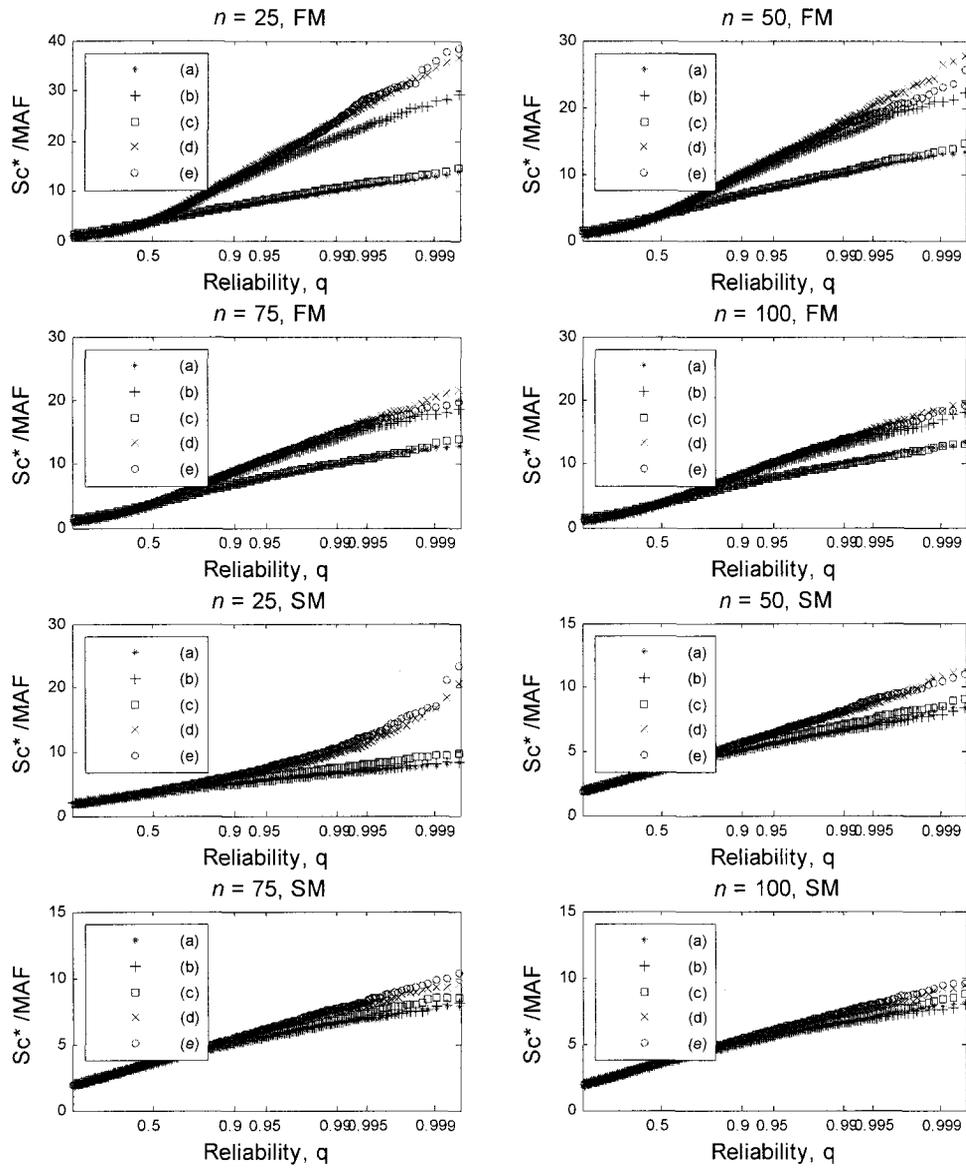


Figure 2.5: Quantile plots of simulated storage capacities  $Sc^*$  for different sample sizes where reliability denotes nonexceedance probability associated with the corresponding quantile (Lee's Ferry in the Colorado River Basin, Bayesian analysis,  $N_d=100$ ).

Table 2.5: Quantiles of  $S_C^*$  scaled by MAF for different reliability  $q$  (Lee's Ferry in the Colorado River Basin,  $N_d=98$ ).

$n$	case	Asymptotic					Bayesian						
		$q=0.5$	0.9	0.95	0.99	0.995	0.999	$q=0.5$	0.9	0.95	0.99	0.995	0.999
25 FM	(a)	3.7	6.8	7.9	9.9	10.6	12.5	3.7	6.8	7.8	10.1	11.0	12.9
	(b)	3.8	11.6	14.4	19.6	21.8	25.6	3.8	11.7	14.5	20.5	22.9	27.3
	(c)	3.7	6.8	7.9	10.3	11.2	13.0	3.8	6.9	8.1	10.5	11.6	13.4
	(d)	4.0	12.1	14.9	20.7	22.8	26.4	4.0	12.7	16.2	23.7	27.1	34.1
	(e)	4.1	12.1	15.0	20.2	21.8	26.9	3.9	12.1	15.4	23.9	28.4	35.7
50 FM	(a)	3.7	6.8	7.8	9.9	10.5	12.7	3.7	6.9	7.8	10.1	11.1	12.9
	(b)	3.7	9.3	11.4	15.6	17.0	20.2	3.7	9.4	11.5	15.9	17.7	21.0
	(c)	3.7	6.8	7.9	10.3	11.3	13.5	3.7	6.9	8.0	10.2	11.3	13.5
	(d)	3.9	9.7	11.9	16.2	17.8	21.0	3.9	10.1	12.5	18.0	20.3	26.2
	(e)	3.9	9.8	11.8	15.9	17.1	20.0	3.9	9.9	12.1	17.3	19.0	23.1
75 FM	(a)	3.7	6.8	7.9	9.9	10.7	12.2	3.7	6.8	7.9	10.0	10.9	12.5
	(b)	3.8	8.8	10.5	13.8	15.2	17.4	3.7	8.5	10.3	13.8	15.5	17.8
	(c)	3.7	6.7	7.8	10.1	11.0	12.9	3.7	6.8	8.0	10.0	10.8	13.1
	(d)	3.8	8.8	10.5	14.0	15.1	18.5	3.7	8.9	10.8	14.9	16.6	20.5
	(e)	3.8	8.9	10.7	14.3	15.3	18.8	3.8	8.9	10.8	14.8	16.2	18.9
100 FM	(a)	3.8	6.8	7.9	10.1	10.9	12.8	3.8	6.8	7.8	10.0	10.9	12.6
	(b)	3.8	8.3	9.9	13.3	14.8	17.3	3.7	8.3	9.8	12.9	14.1	16.7
	(c)	3.7	6.8	7.8	10.0	10.7	12.5	3.7	6.8	7.8	9.8	10.7	12.8
	(d)	3.9	8.3	10.1	13.5	14.7	17.5	3.8	8.4	10.1	13.6	15.3	18.4
	(e)	3.8	8.4	10.1	13.6	15.0	17.4	3.8	8.3	10.2	13.8	15.0	18.2
25 SM	(a)	3.6	5.1	5.6	6.6	7.0	7.8	3.6	5.1	5.6	6.6	7.0	7.8
	(b)	3.6	5.1	5.6	6.7	7.2	7.9	3.6	5.1	5.6	6.7	7.1	8.1
	(c)	3.5	5.3	5.8	7.1	7.4	8.8	3.6	5.4	6.0	7.4	8.0	9.4
	(d)	3.7	6.0	6.9	9.5	10.7	14.9	3.7	6.0	7.0	9.8	11.2	16.6
	(e)	3.7	6.1	7.2	9.5	10.8	13.9	3.8	6.3	7.5	10.5	12.3	16.7
50 SM	(a)	3.6	5.2	5.7	6.6	7.0	7.8	3.6	5.1	5.6	6.6	7.1	8.3
	(b)	3.6	5.1	5.6	6.6	6.9	7.7	3.6	5.1	5.6	6.7	7.0	7.8
	(c)	3.6	5.2	5.7	6.7	7.2	8.2	3.6	5.3	5.8	7.0	7.5	8.5
	(d)	3.6	5.6	6.2	7.7	8.6	10.3	3.6	5.6	6.4	8.0	8.6	10.8
	(e)	3.6	5.5	6.3	7.7	8.6	10.0	3.6	5.7	6.4	8.1	9.0	10.4
75 SM	(a)	3.6	5.1	5.6	6.6	6.9	7.7	3.6	5.1	5.6	6.6	6.9	7.7
	(b)	3.6	5.1	5.6	6.6	7.0	7.9	3.6	5.1	5.6	6.6	7.0	7.8
	(c)	3.6	5.2	5.7	6.8	7.2	8.4	3.6	5.2	5.7	6.8	7.4	8.4
	(d)	3.6	5.4	6.0	7.3	8.0	8.9	3.6	5.3	6.0	7.5	8.0	9.2
	(e)	3.6	5.4	6.1	7.6	8.1	9.3	3.6	5.5	6.2	7.7	8.2	9.7
100 SM	(a)	3.6	5.1	5.6	6.7	7.1	7.8	3.6	5.1	5.6	6.6	6.9	7.9
	(b)	3.6	5.1	5.6	6.6	7.1	7.9	3.6	5.1	5.6	6.6	6.9	7.5
	(c)	3.6	5.2	5.7	6.7	7.1	8.0	3.6	5.2	5.7	6.8	7.2	8.4
	(d)	3.6	5.4	5.9	7.1	7.7	8.4	3.6	5.3	5.9	7.2	7.6	9.1
	(e)	3.6	5.3	6.0	7.2	7.8	9.1	3.6	5.4	6.0	7.4	8.0	9.4

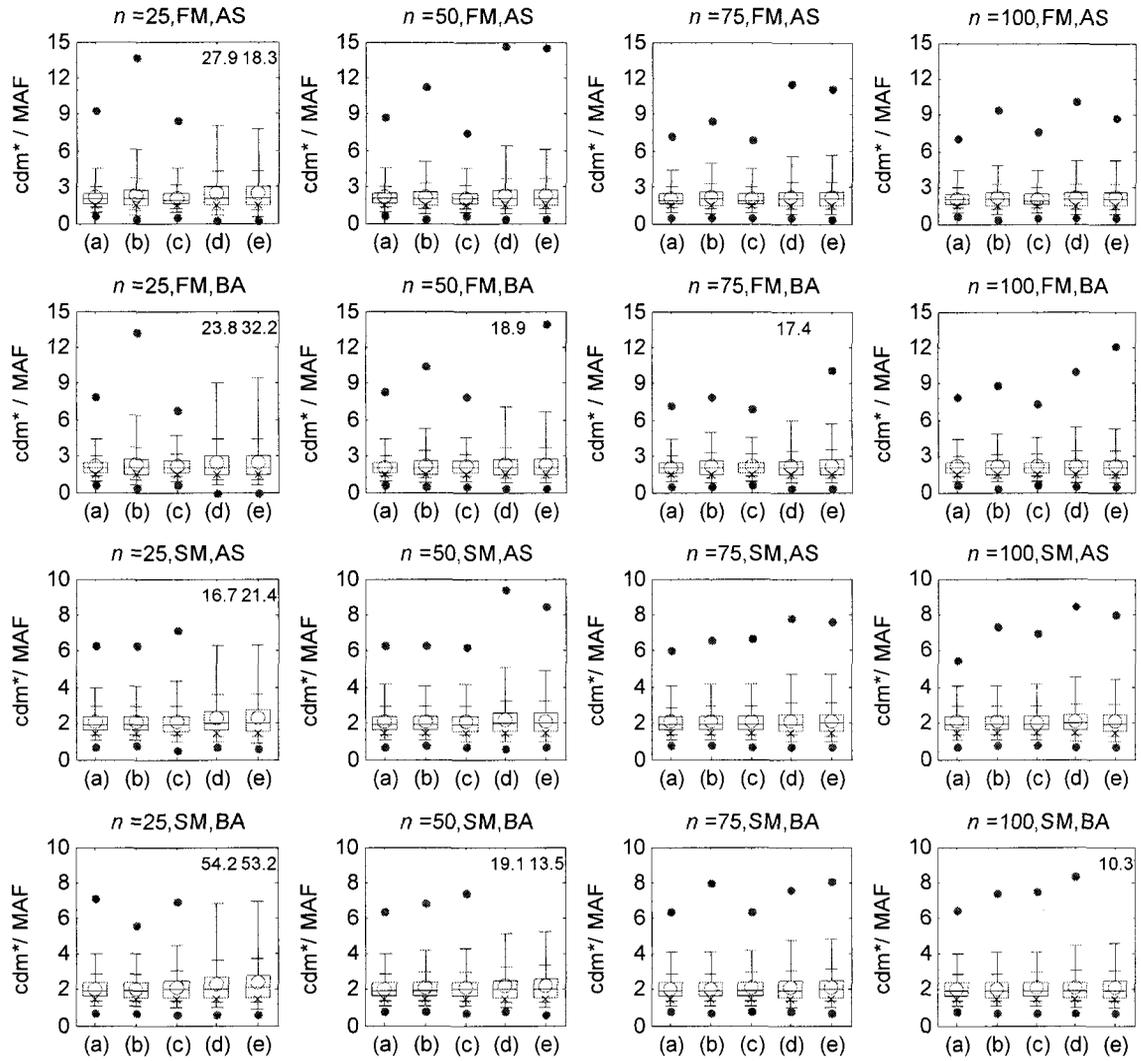


Figure 2.6: Distribution of generated critical drought magnitude  $cdm^*$  for different sample sizes, at Lee's Ferry in the Colorado River Basin ( $N_d=98$ ).

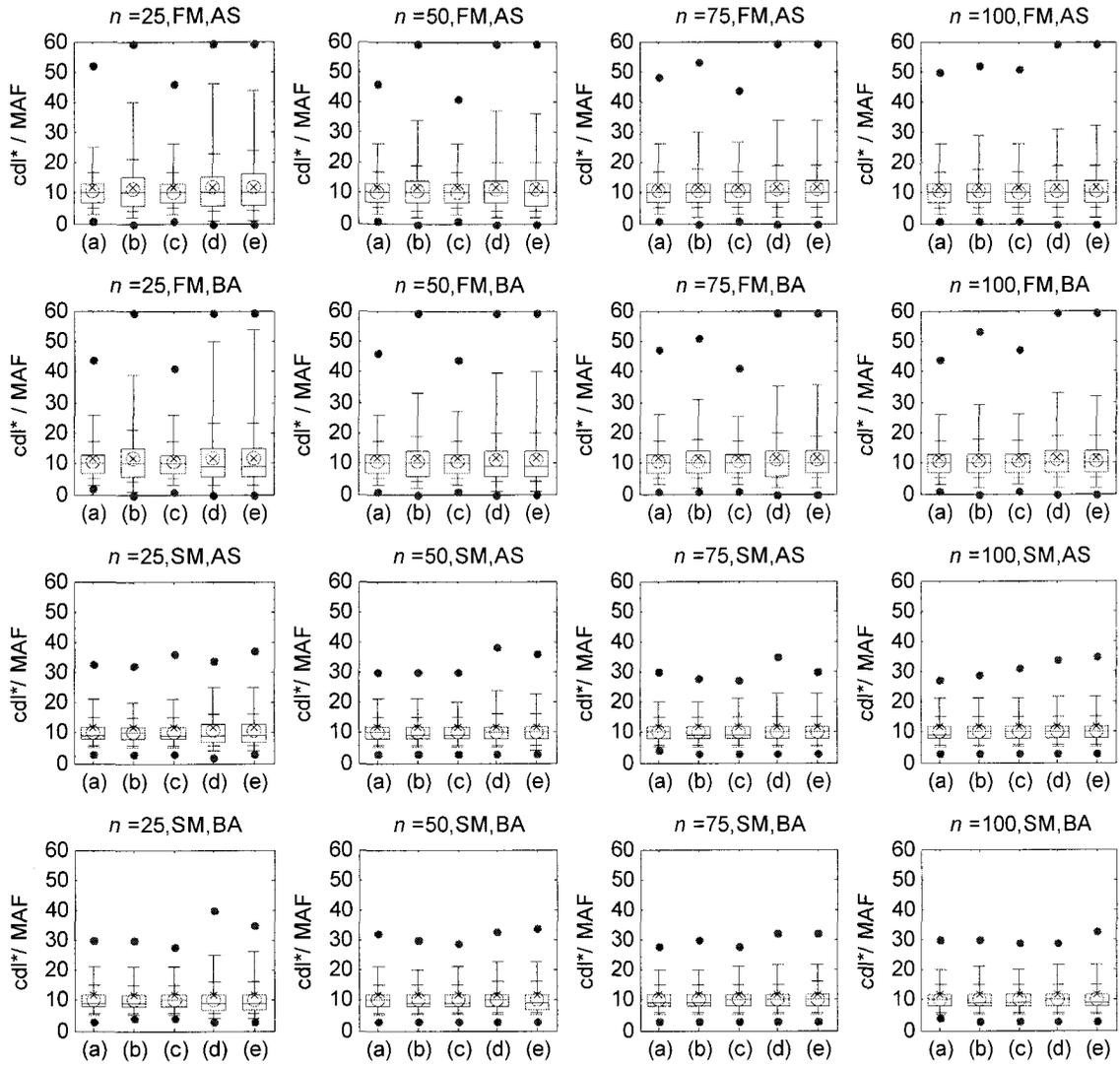


Figure 2.7: Distribution of generated critical drought length  $cdl^*$  for different sample sizes, at St. Lawrence River ( $N_d=59$ ).

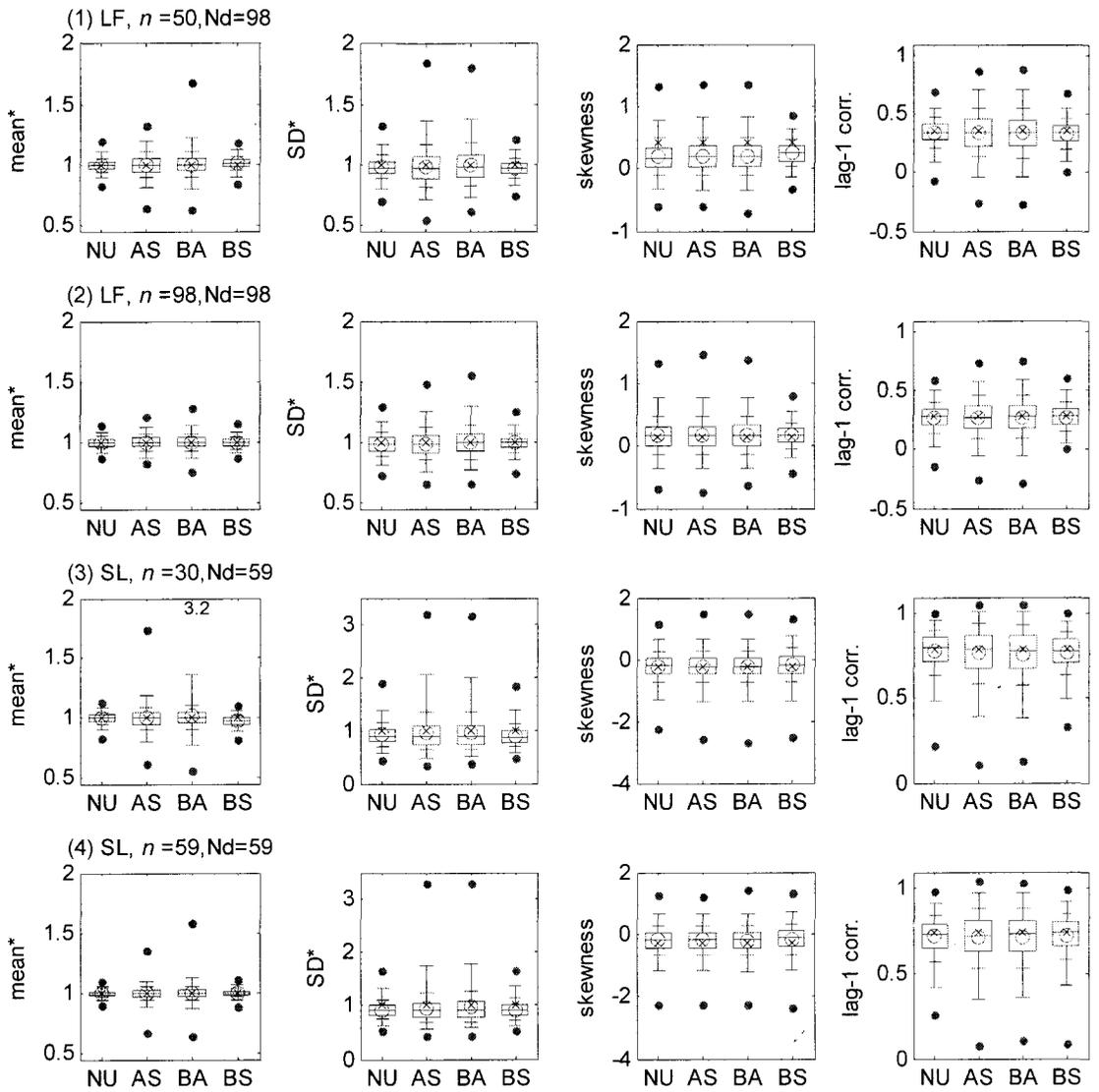


Figure 2.8: Distributions of mean, standard deviation (SD), skewness, and lag-1 serial correlation coefficients based on 10,000 synthetic annual streamflows where NU, AS, BA, BS represents natural uncertainty, asymptotic, Bayesian, and bootstrap, respectively and LF, SL represents Lee's Ferry in the Colorado River Basin and St. Lawrence River, respectively. In AS and BA, uncertainties of all parameters are incorporated into simulation.

Table 2.6: Example of generated storage capacity and critical drought magnitude (scaled by MAF) ( $n=50$ ,  $N_d=100$ , Bayesian analysis, FM)

Demand level NU/PU	Lee's Ferry in the Colorado River Basin				St. Lawrence River			
	MAF		0.8MAF		MAF		0.9MAF	
	NU	PU	NU	PU	NU	PU	NU	PU
Storage capacity								
mean	4.08	4.89	0.93	1.09	1.36	1.69	0.23	0.39
		19.8%		17.1%		24.6%		71.7%
SD	1.93	3.53	0.39	0.80	0.86	1.77	0.23	0.82
		82.4%		104.3%		106.7%		259.3%
$q_{0.9}$	6.78	9.82	1.43	1.93	2.54	3.69	0.52	0.86
		44.9%		34.8%		45.6%		65.7%
$q_{0.95}$	7.73	11.95	1.67	2.44	3.03	5.00	0.69	1.48
		54.6%		46.1%		64.7%		115.2%
$q_{0.99}$	9.93	16.99	2.17	3.98	4.04	8.59	1.05	3.97
		71.1%		83.8%		112.6%		278.5%
Critical drought magnitude								
mean	2.10	2.30	0.84	0.94	0.97	1.21	0.21	0.34
		9.4%		12.0%		23.8%		61.6%
SD	0.74	1.21	0.33	0.56	0.60	1.42	0.21	0.68
		63.4%		69.9%		135.9%		227.6%
$q_{0.9}$	3.06	3.75	1.27	1.61	1.77	2.45	0.47	0.75
		22.2%		26.5%		38.2%		58.4%
$q_{0.95}$	3.51	4.54	1.46	1.97	2.15	3.54	0.63	1.23
		29.3%		35.3%		64.7%		96.3%
$q_{0.99}$	4.39	6.66	1.91	2.96	3.01	7.39	0.99	3.15
		51.6%		55.3%		145.2%		219.0%

Note NU: no parameter uncertainty considered (natural uncertainty), PU: uncertainty of all parameters incorporated case, SD: standard deviation,  $q_{0.9}$ ,  $q_{0.95}$ ,  $q_{0.99}$  mean 90%, 95%, and 99% quantile, respectively. Value (%) in the column of PU represents the ratio of increased storage capacity(or critical drought magnitude) in PU with respect to NU. (parameter uncertainty effect)

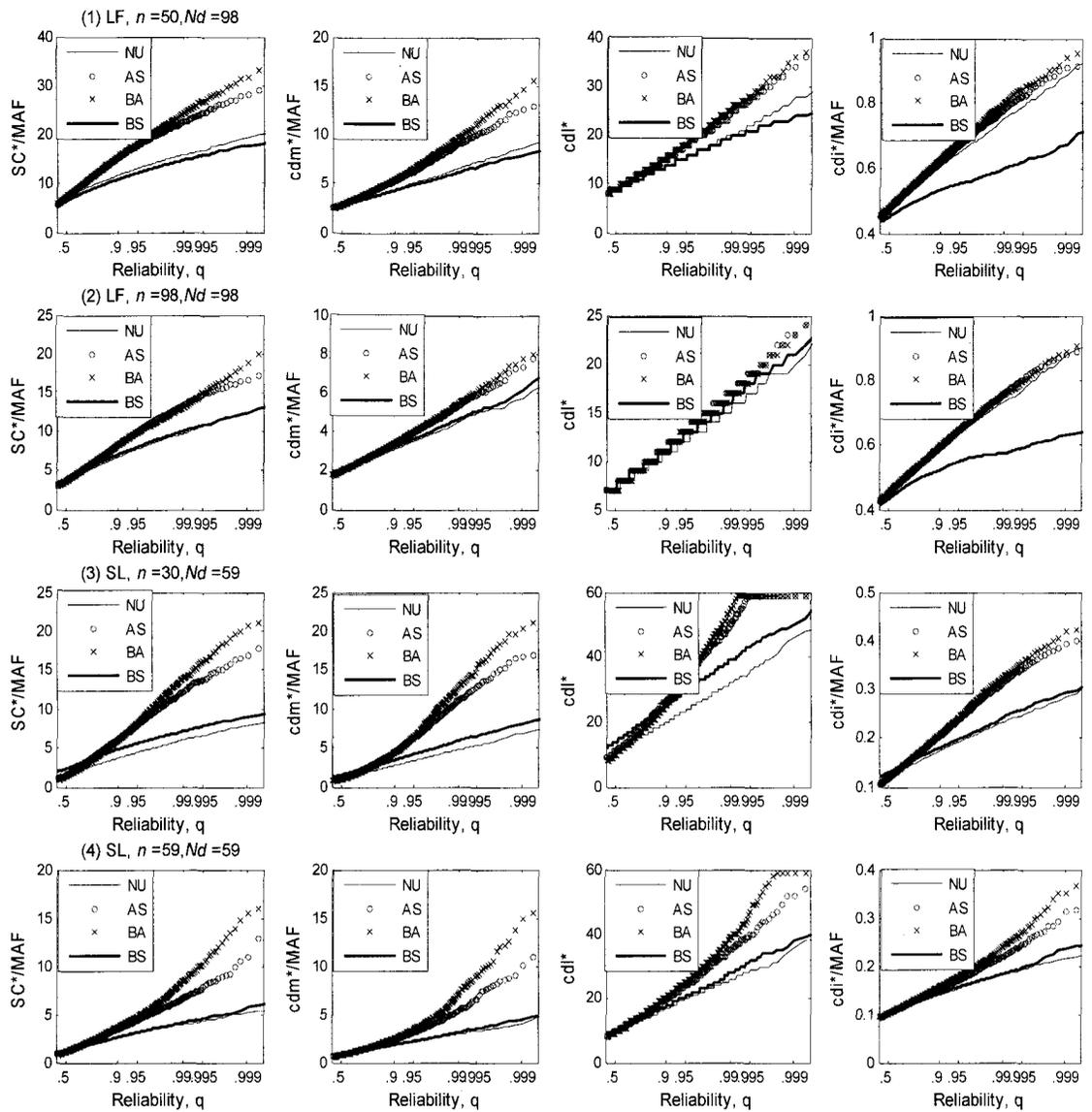


Figure 2.9: Empirical distribution of generated storage capacities  $Sc^*$ , critical drought magnitude  $cdm^*$ , critical drought length  $cdl^*$ , critical drought intensity  $cdi^*$  for different sample sizes in option *FM*. In AS and BA, uncertainties of all parameters are incorporated into simulation.

## References

- Box, G.E.P. and G.W. Jenkins (1970) *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Box, G.E.P. and G.C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc., N.Y.
- Box, G. E. P. and G. W. Jenkins, (1976) *Time Series Analysis Forecasting and Control*, revised edition, Holden-Day, San Francisco.
- Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods*, 2<sup>nd</sup> edition, Springer-Verlag, N.Y.
- Camacho, F., A.I. McLeod, and K.W. Hipel (1987). Multivariate contemporaneous ARMA model with hydrological applications, *Stochastic Hydrology and Hydraulics*, 1, pp. 141-154.
- Cover, K.A. and T.E. Unny (1986). Application of computer intensive statistics to parameter uncertainty in streamflow synthesis, *Water Resources Bulletin*, 22(3), pp. 495-507.
- Elfon, B. (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics*, 7(1), pp.1-26.
- Grygier, J.C. and J.R. Stedinger, (1990). *SPIGOT, A synthetic Streamflow Generation Software Package, technical description*, version 2.5, School of Civil and Environmental Engineering, Cornell University, Ithaca, N. Y.
- Hurst, H.E. (1951). Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, 116, pp. 770-799.
- Klemes, V., R. Srikanthan and T. A. McMahon (1981). Long-memory flow models in reservoir analysis: What is their practical value?, *Water Resources Research*, 17(3), pp. 737-751.
- Lall, U. and A. Sharma (1996). A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resources Research*, 32(3), pp. 679-693.
- Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981). *Water Resources Planning and Analysis*, Prentice Hall, Englewood Cliffs, New Jersey.
- McLeod, A.I. and K.W. Hipel (1978). Simulation procedures for Box-Jenkins models,

*Water Resources Research*, 14(5), pp. 969-975.

- Oliveira, G.C., J. Kelman, M.V.F. Pereira and J.R. Stedinger (1988). Representation of Spatial Cross-Correlation in a Seasonal Streamflow Model, *Water Resource Research*, 24(4), 781-785.
- Pereira, M.V.F., G.C. Oliveira, C.C.G. Costa, and J. Kelman (1984). Stochastic streamflow models for hydroelectric systems, *Water Resources Research*, 20(3), pp. 379-390.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.
- Srinivas, V.V. and K. Srinivasan (2000). Post-blackening approach for modeling dependent annual streamflows, *Journal of Hydrology*, 230 (1-2), pp. 86-126.
- Srinivas, V.V. and K. Srinivasan (2005). Hybrid matched-block bootstrap for stochastic simulation of multiseason streamflows, *Journal of Hydrology*, 329 (1-2), pp. 1-15.
- Stedinger, J. R. and M. R. Taylor (1982). Synthetic streamflow generation, Part 2. Parameter uncertainty, *Water Resources Research*, 18(4), pp. 919-924.
- Stedinger, J.R., R.M. Vogel, and E. Foufoula-Georgiou, Frequency analysis of extreme events, in *Handbook of Hydrology*, chapter 18, edited by D.A. Maidment, McGraw-Hill, New York, NY, 1993.
- Tasker, G.D. and P. Dunne (1997). Bootstrap position analysis for forecasting low flow frequency, *Journal of Water Resources Planning and Management*, ASCE, 123(6), pp. 359-367.
- Valdes, J.B., I. Rodriguez-Iturbe, and G.J. Vicens (1977). Bayesian generation of synthetic streamflows; 2. The multivariate case, *Water Resources Research*, 13(2), pp. 291-295.
- Valencia, D. and J.C. Schaake, Jr. (1973). Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), pp. 580-585, 1973.
- Vogel, R.M. and A.L. Shallcross (1996). The moving blocks bootstrap versus parametric time series models, *Water Resources Research*, 32(6), pp. 1875-1882.
- Vogel, R.M. and J.R. Stedinger (1987). Generalized storage-reliability-yield relationships, *Journal of Hydrology*, 89, pp. 303-327.
- Vogel, R.M., M. Lane, R.S. Ravindiran, and P. Kirshen (1999). Storage Reservoir Behavior in the United States, *Journal of Water Resources Planning and Management*, ASCE, 125(5).

Wood, E. F. (1978). Analysing hydrologic uncertainty and its impact upon decisionmaking in water resources, *Advances in Water Resources*, 1(5), pp. 299-305.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., New York.

## Appendix 2.A: Additional Figures and Tables

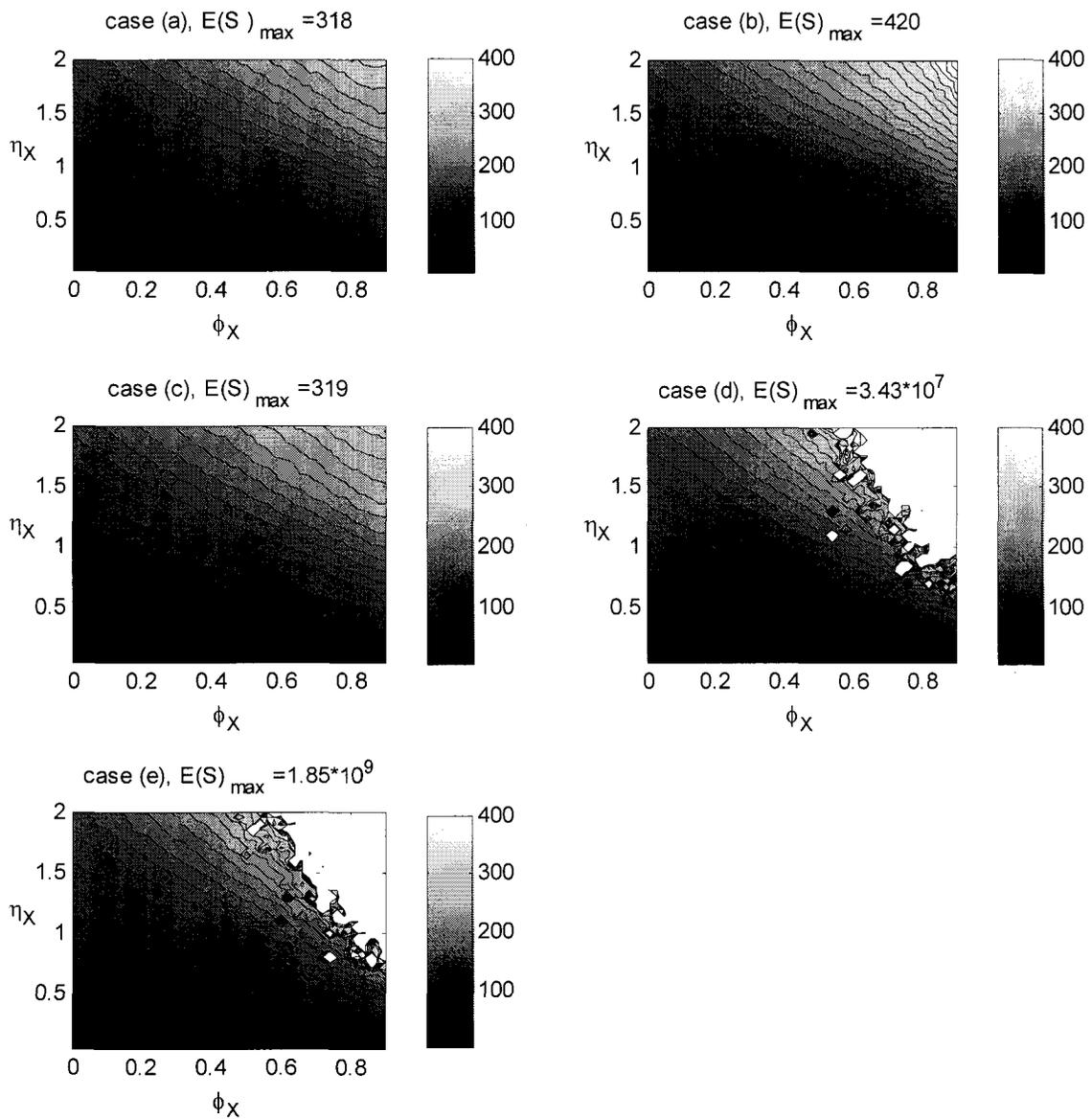


Figure 2.A1: Expected values of generated storage capacities calculated from 10,000 different traces simulated from theoretically assumed streamflows with different coefficient of variation  $\eta_X$  (0.1-2) and lag-1 serial correlation  $\phi_X$  (0-0.9). 100% generated mean (averaged over each generated flow set,  $SM$ ) is used as a demand level in the calculation of storage capacity based on SPA. Calculated maximum expected values are (a) 318, (b) 420, (c) 319, (d)  $3.43 \times 10^7$ , and (e)  $1.85 \times 10^9$ .

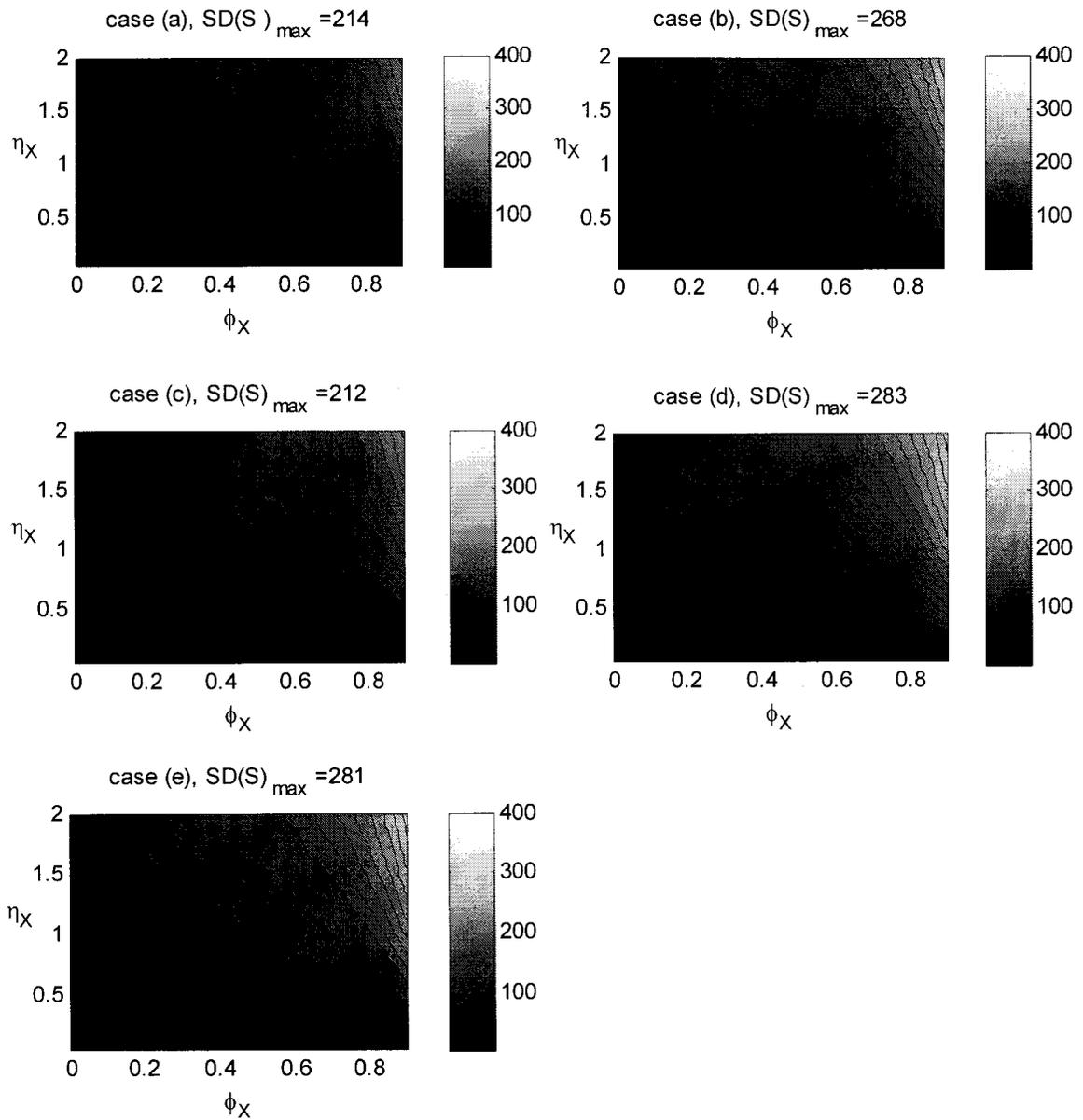


Figure 2.A2: Standard deviations of generated storage capacities calculated from 10,000 different traces simulated from theoretically assumed streamflows with different coefficient of variation  $\eta_X$  (0.1-2) and lag-1 serial correlation  $\phi_X$  (0-0.9). 100% Historical mean ( $FM$ ) is used as a demand level in the calculation of storage capacity based on SPA. Calculated maximum expected values are (a) 214, (b) 268, (c) 212, (d) 283, and (e) 281.

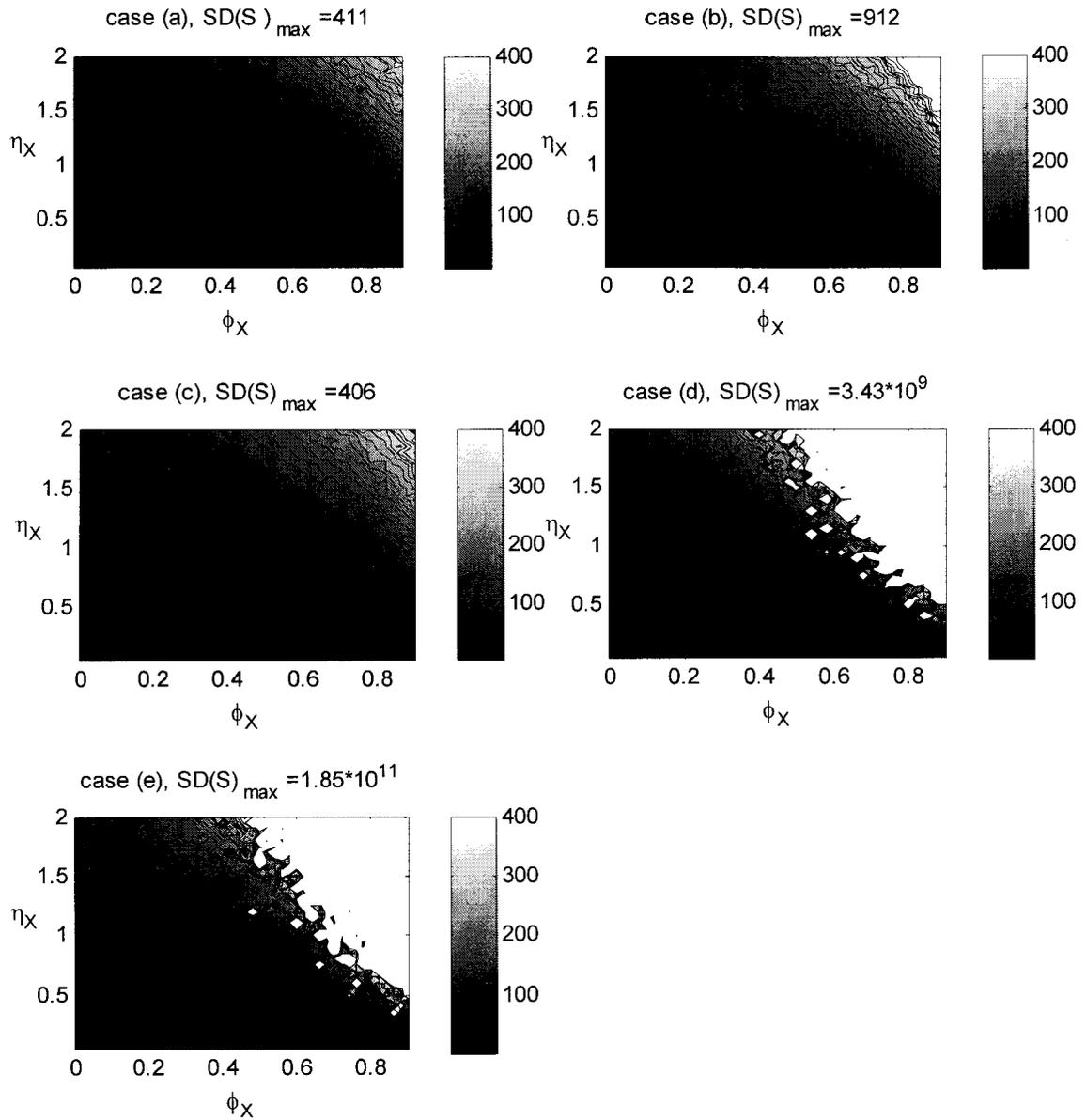


Figure 2.A3: Standard deviations of generated storage capacities calculated from 10,000 different traces simulated from theoretically assumed streamflows with different assumed coefficient of variation  $\eta_x$  (0.1-2) and lag-1 serial correlation  $\phi_x$  (0-0.9). 100% generated mean (averaged over each generated flow set,  $SM$ ) is used as a demand level in the calculation of storage capacity based on SPA. Calculated maximum expected values are (a) 411, (b) 912, (c) 406, (d)  $3.43 \times 10^9$ , and (e)  $1.85 \times 10^{11}$ .

Table 2.A1: Efficiency of simulated standard error (*ASTD*) of parameter estimates sampled from the asymptotic distribution relative to theoretical asymptotic standard error (*TASTD*) defined by  $Eff(\cdot) = ASTD(\cdot) / TASTD(\cdot)$ .

Lee's Ferry: $\hat{\phi}=0.277$ , $\hat{\sigma}_\epsilon^2=122,768$ , $\hat{\mu}=549,101$			
Asymptotic variance	$TASTD(\hat{\phi})$	$TASTD(\hat{\sigma}_\epsilon^2)$	$TASTD(\hat{\mu})$
$n = 25$	0.192	17,362	34,072
50	0.136	12,277	24,092
75	0.111	10,024	19,671
100	0.096	8,681	17,036
Simulated variance	$ASTD(\hat{\phi})$	$ASTD(\hat{\sigma}_\epsilon^2)$	$ASTD(\hat{\mu})$
$n = 25$	0.163	17287	34,369
50	0.126	12236	24,330
75	0.107	9967	19,538
100	0.096	8656	17,025
Efficiency	$Eff(\hat{\phi})$	$Eff(\hat{\sigma}_\epsilon^2)$	$Eff(\hat{\mu})$
$n = 25$	0.85	1.00	1.01
50	0.93	1.00	1.01
75	0.97	0.99	0.99
100	1.00	1.00	1.00
St. Lawrence River: $\hat{\phi}=0.720$ , $\hat{\sigma}_\epsilon^2=267,848,289$ , $\hat{\mu}=2,324,977,641$			
Asymptotic variance	$TASTD(\hat{\phi})$	$TASTD(\hat{\sigma}_\epsilon^2)$	$TASTD(\hat{\mu})$
$n = 25$	0.139	37,879,468	198,632,953
50	0.098	26,784,829	140,454,708
75	0.080	21,869,721	114,680,789
100	0.069	18,939,734	99,316,477
Simulated variance	$ASTD(\hat{\phi})$	$ASTD(\hat{\sigma}_\epsilon^2)$	$ASTD(\hat{\mu})$
$n = 25$	0.129	37,634,980	197,715,202
50	0.097	27,178,879	141,399,777
75	0.080	21,641,456	114,572,395
100	0.069	18,988,777	99,686,224
Efficiency	$Eff(\hat{\phi})$	$Eff(\hat{\sigma}_\epsilon^2)$	$Eff(\hat{\mu})$
$n = 25$	0.93	0.99	1.00
50	0.99	1.01	1.01
75	1.00	0.99	1.00
100	1.00	1.00	1.00

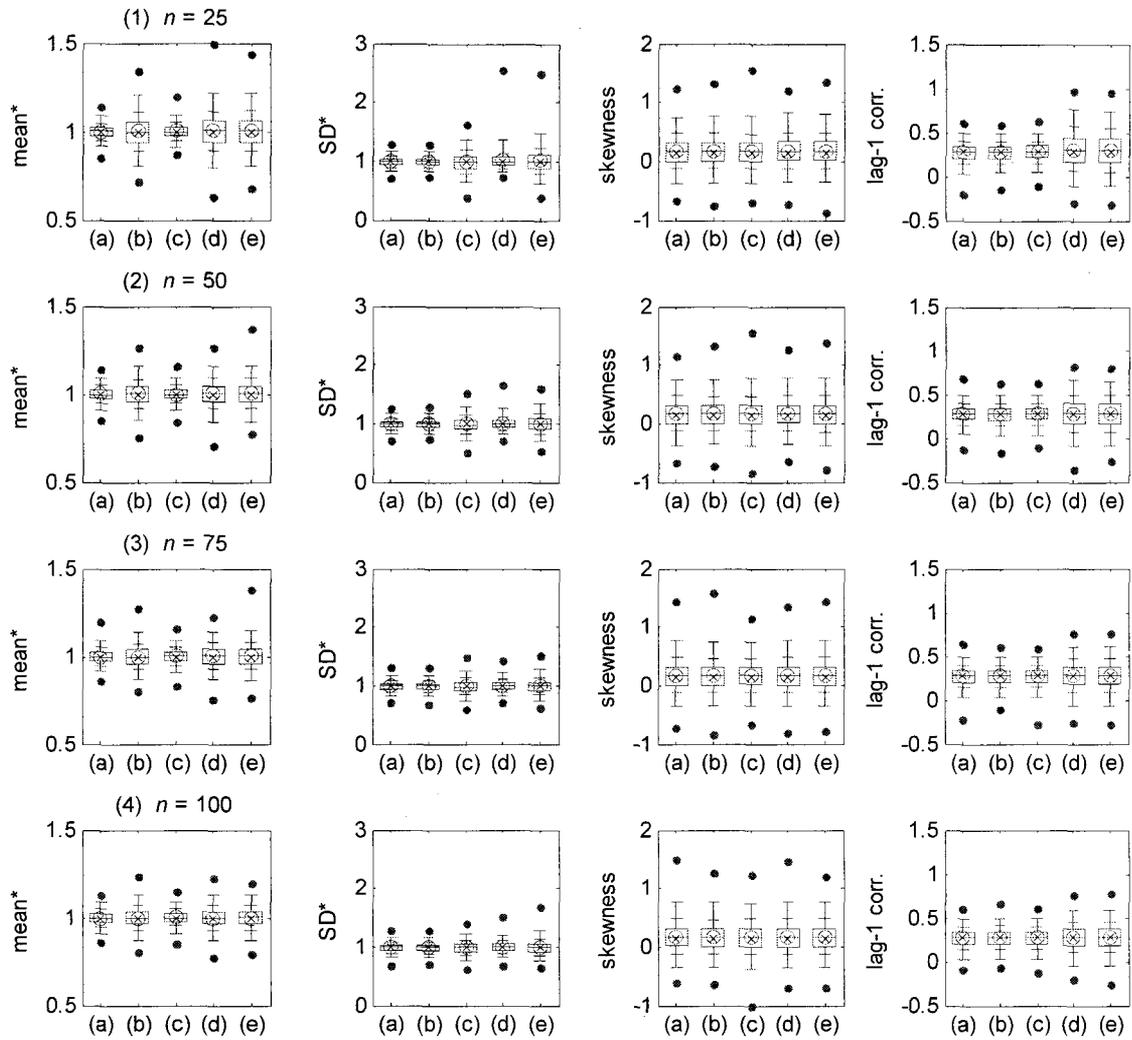


Figure 2.A4: Distributions of mean, standard deviation (SD), skewness, and lag-1 serial correlation coefficients based on 10,000 synthetic annual streamflow series at Lee's Ferry in the Colorado River Basin (asymptotic analysis).



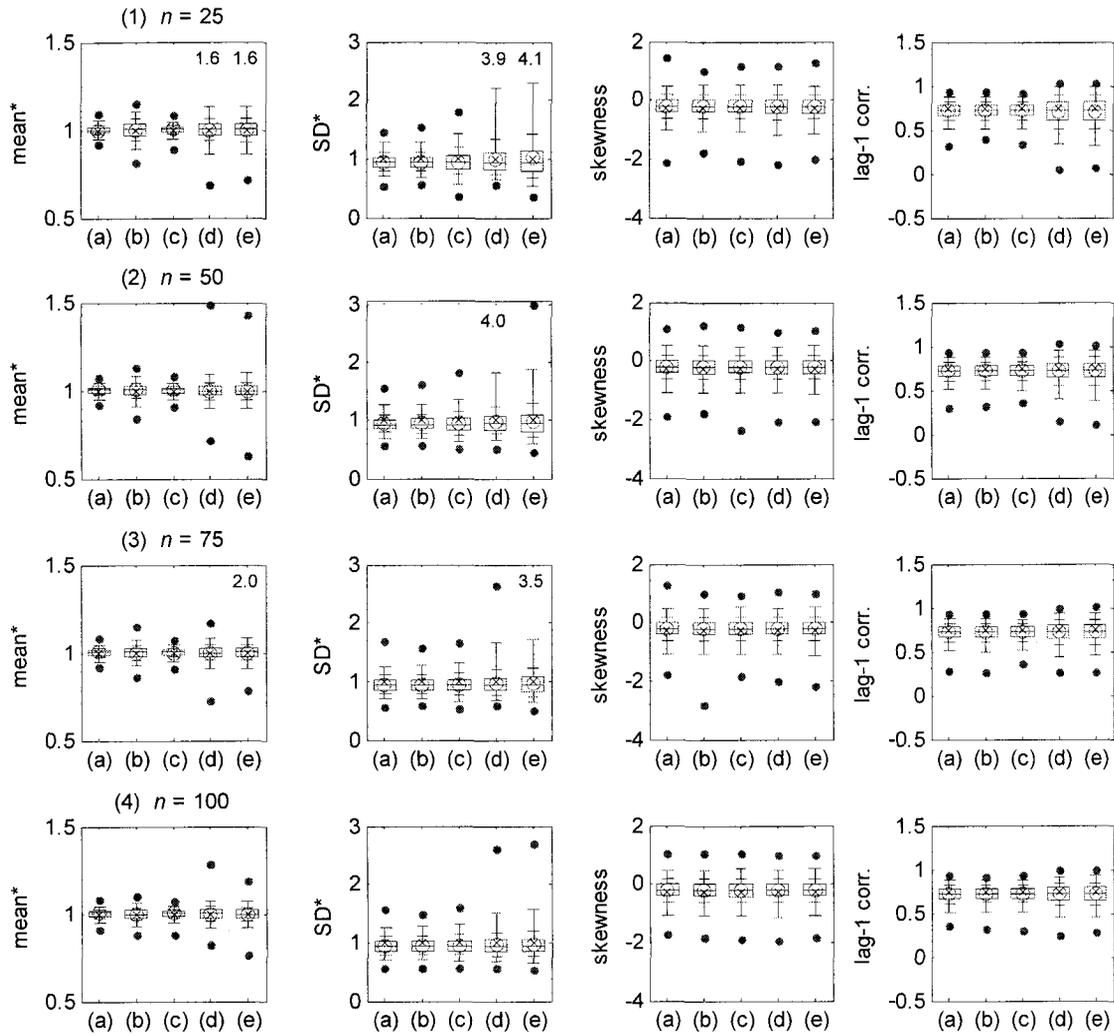


Figure 2.A6: Distributions of mean, standard deviation (SD), skewness, and lag-1 serial correlation coefficients based on 10,000 synthetic annual streamflow series at St. Lawrence River (asymptotic analysis).

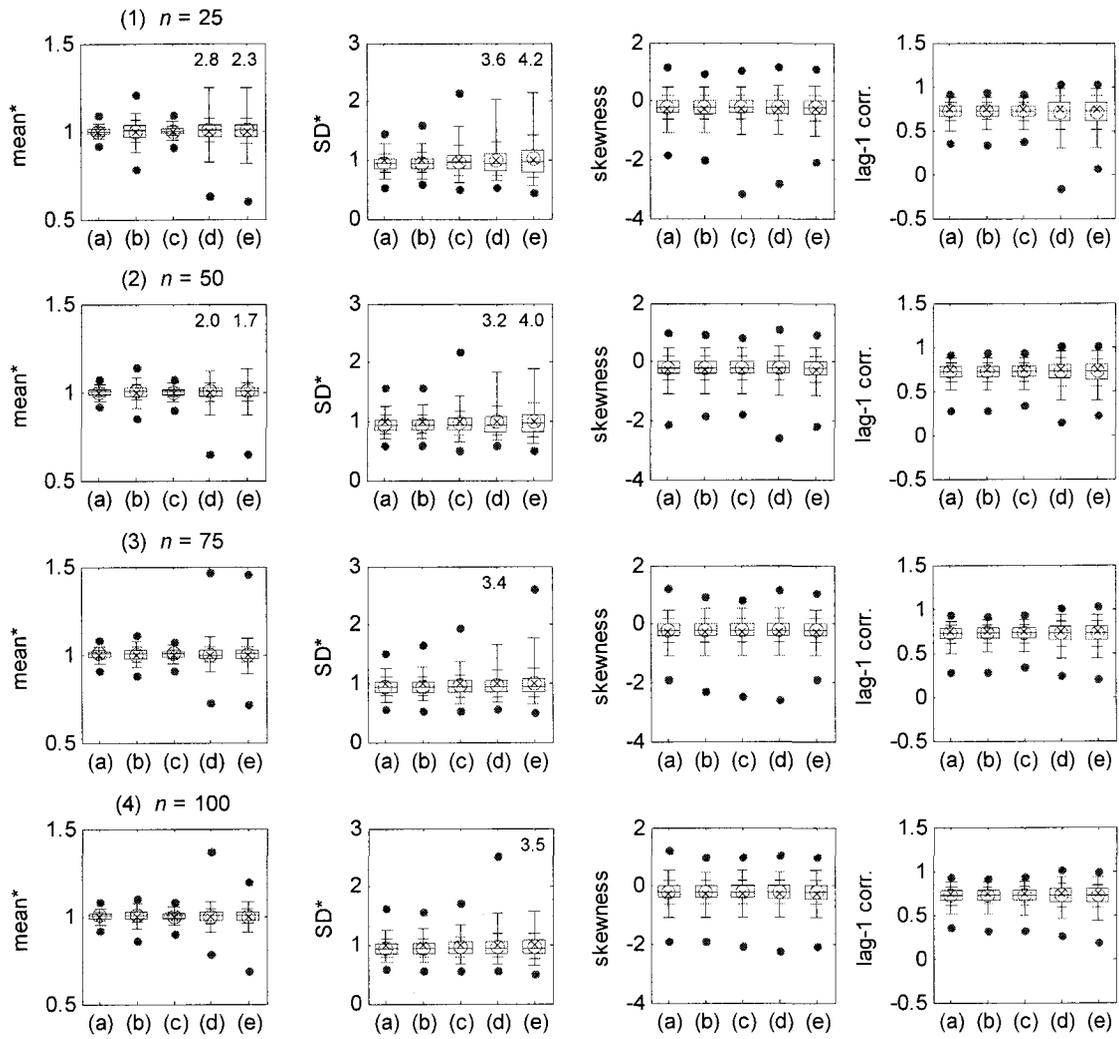
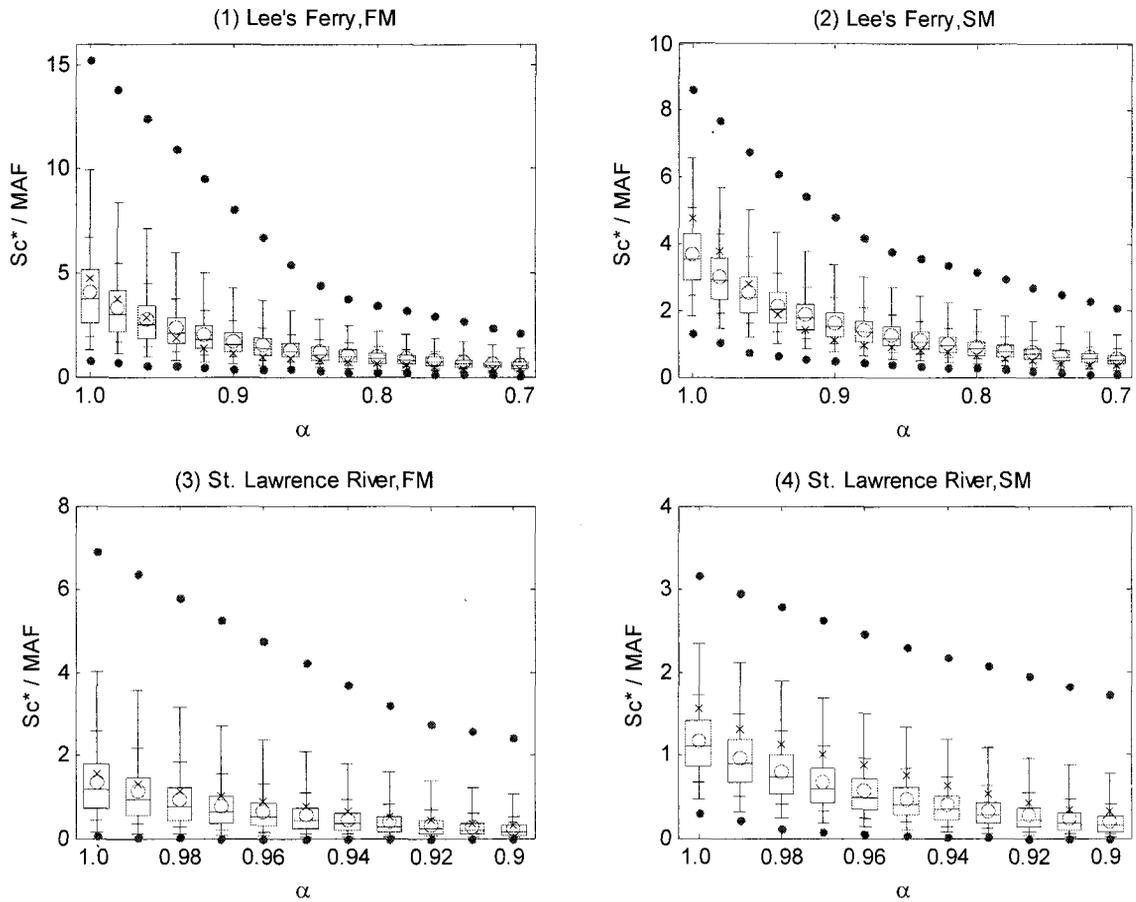


Figure 2.A7: Distributions of mean, standard deviation (SD), skewness, and lag-1 serial correlation coefficients based on 10,000 synthetic annual streamflow series at St. Lawrence River (Bayesian analysis).



RBIAS and RRMSE of  $s_C^*$  with different  $\alpha$  (Lee's Ferry)

$\alpha$		1	0.98	0.96	0.94	0.92	0.9	0.88	0.86	0.84	0.82	0.8	0.78	0.76	0.74	0.72	0.7
RBIAS	FM	-14	-11	-2	23	43	56	55	49	45	43	44	47	54	66	71	64
(%)	SM	-23	-20	-10	14	34	48	49	44	40	39	40	44	51	63	68	61
RRMSE	FM	43	44	46	61	77	88	87	80	76	75	75	79	86	99	104	98
(%)	SM	32	32	31	41	57	70	71	67	65	65	66	71	78	91	97	91

RBIAS and RRMSE of  $s_C^*$  with different  $\alpha$  (St. Lawrence River)

$\alpha$		1	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.92	0.91	0.9
RBIAS	FM	-13	-15	-17	-23	-27	-29	-29	-27	-25	-24	-29
(%)	SM	-25	-27	-29	-33	-36	-38	-38	-37	-34	-34	-38
RRMSE	FM	57	59	61	62	63	64	66	70	74	78	80
(%)	SM	37	40	43	47	50	52	55	57	59	62	65

Figure 2.A8: Distributions of generated storage capacities  $S_C^*$  calculated using different demand levels  $\alpha$  ( $n=100$ ,  $N_d=98$  for Lee's Ferry, and  $N_d=59$  for St. Lawrence River)

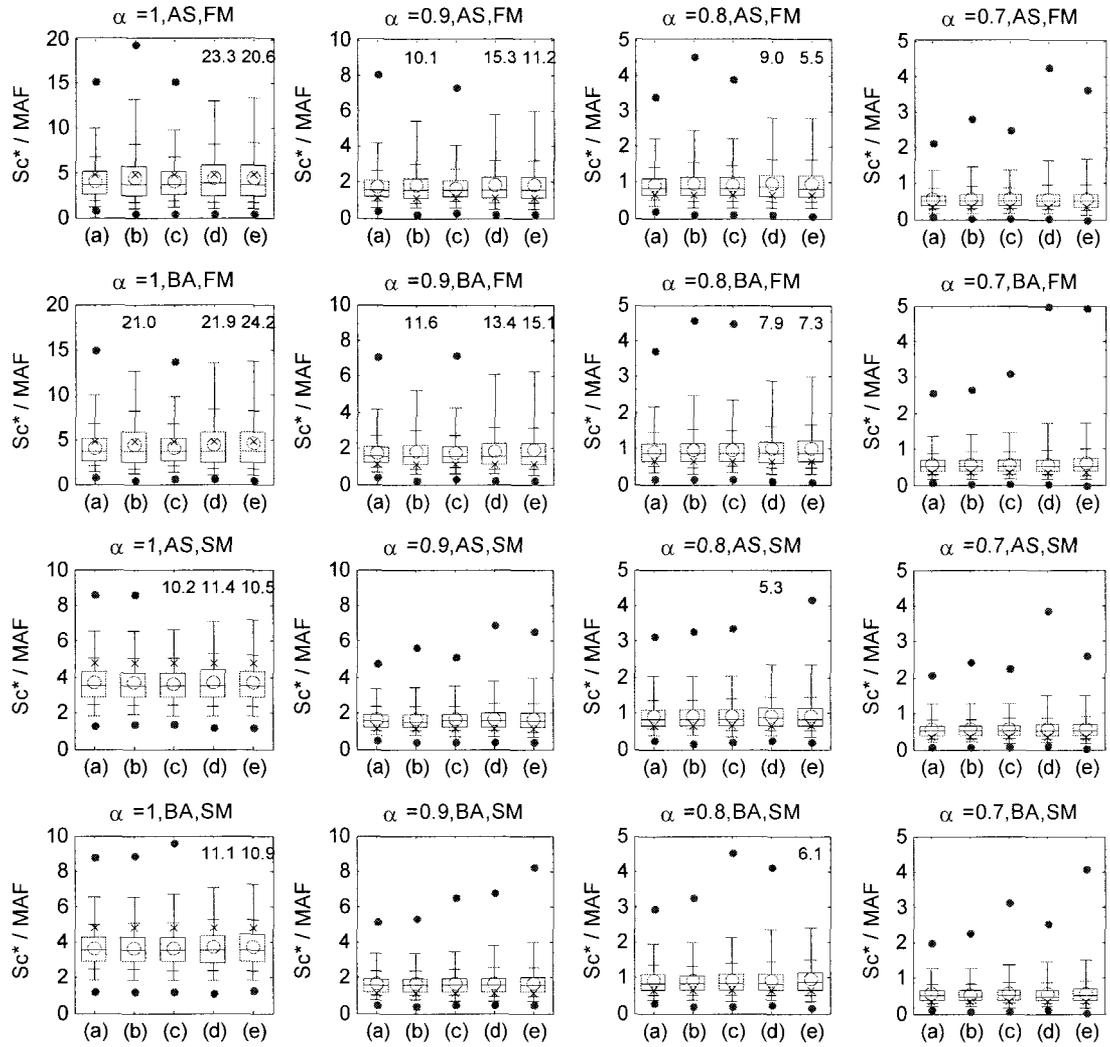


Figure 2.A9: Distribution of generated storage capacities  $Sc^*$  for different demand levels ( $\alpha$  MAF) at Lee's Ferry in the Colorado River Basin ( $n=100, N_d=98$ ).

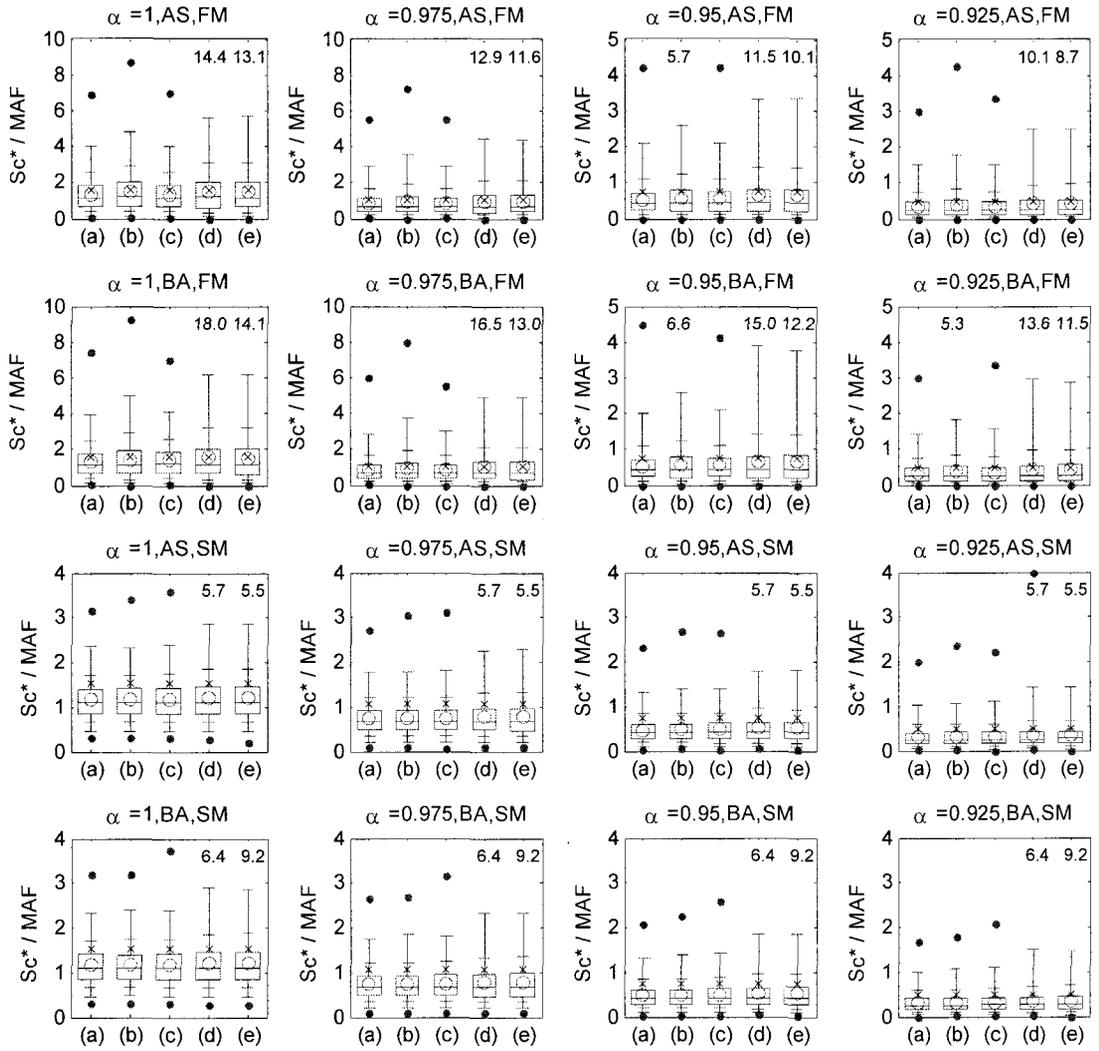


Figure 2.A10: Distribution of generated storage capacities  $Sc^*$  for different demand levels ( $\alpha$  MAF) at St. Lawrence River ( $n=100, N_d=59$ ).

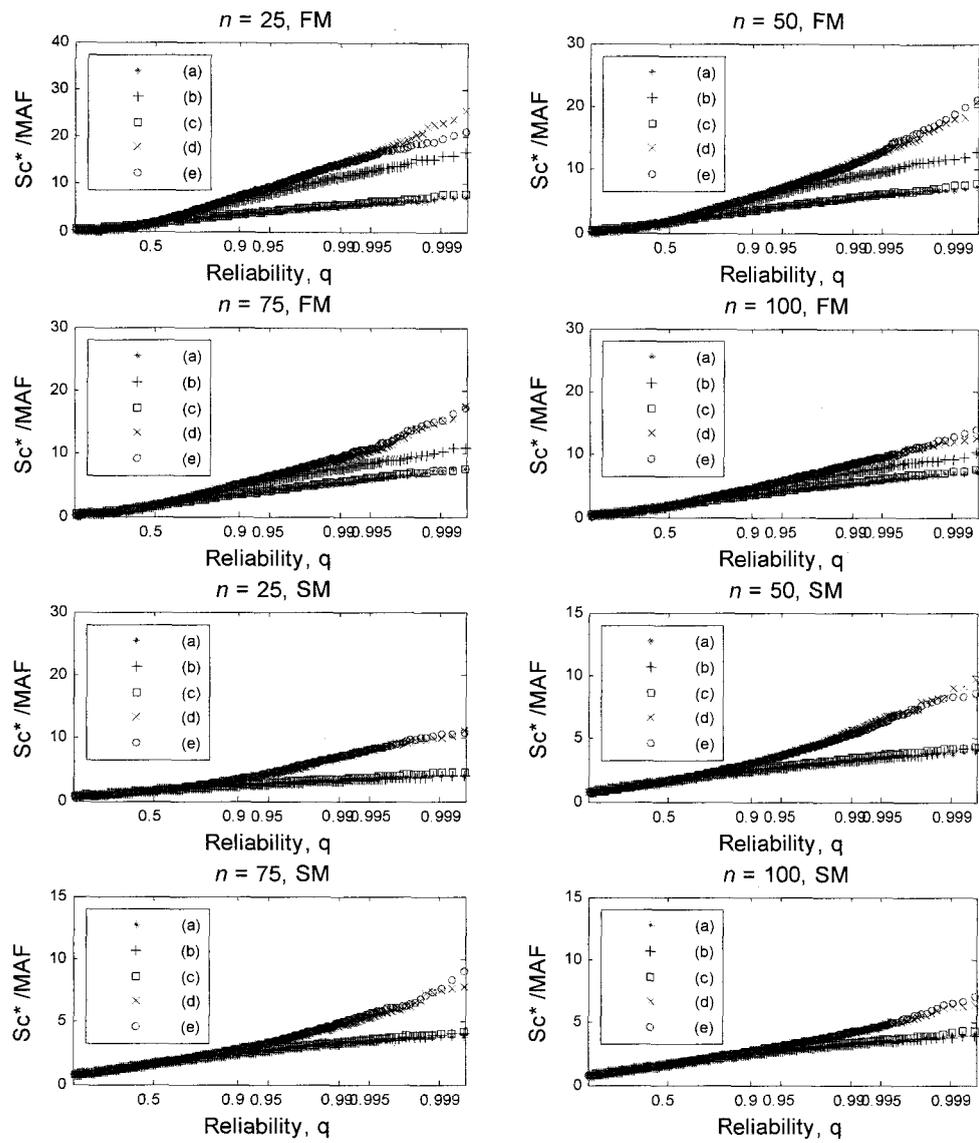


Figure 2.A11: Simulated storage capacities  $Sc^*$  for different sample sizes associated with reliabilities (St. Lawrence River, asymptotic analysis,  $N_d=100$ ).

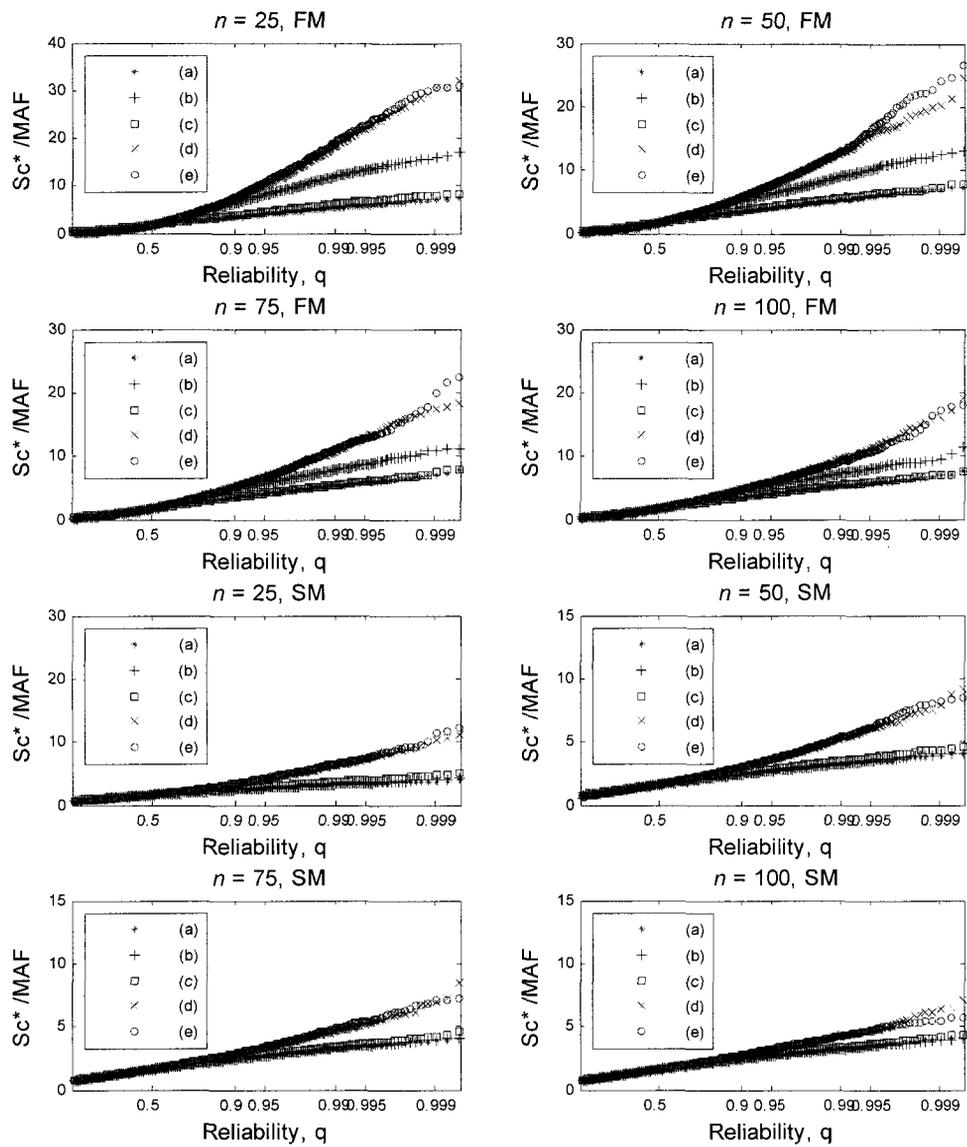


Figure 2.A12: Simulated storage capacities  $Sc^*$  for different sample sizes associated with reliabilities (St. Lawrence River, Bayesian analysis,  $N_d=100$ ).

Table 2.A2: Quantiles of  $S_c^*$  scaled by MAF for different reliability  $q$  (St. Lawrence River,  $N_d=100$ ).

$n$	case	Asymptotic						Bayesian					
		$q=0.5$	0.9	0.95	0.99	0.995	0.999	$q=0.5$	0.9	0.95	0.99	0.995	0.999
25 FM	(a)	1.7	3.5	4.1	5.4	5.9	7.0	1.7	3.5	4.1	5.3	5.8	7.1
	(b)	1.7	6.3	8.0	11.5	12.8	15.9	1.7	6.4	8.2	12.0	13.3	15.7
	(c)	1.7	3.5	4.2	5.6	6.3	7.8	1.7	3.7	4.5	6.2	6.8	8.0
	(d)	1.9	7.1	9.1	14.1	16.3	22.7	1.6	7.2	10.1	18.4	22.1	30.3
	(e)	1.9	7.3	9.3	14.5	16.1	19.5	1.7	7.4	10.7	19.0	22.6	30.5
50 FM	(a)	1.7	3.5	4.1	5.4	6.0	6.8	1.7	3.5	4.1	5.3	5.9	7.1
	(b)	1.7	5.1	6.4	8.9	9.7	11.8	1.7	5.0	6.3	9.1	10.2	12.4
	(c)	1.7	3.5	4.2	5.6	6.1	7.4	1.7	3.6	4.3	5.7	6.2	7.3
	(d)	1.9	5.6	7.1	11.3	12.9	17.7	1.7	5.7	7.6	12.8	15.5	20.1
	(e)	1.8	5.6	7.1	11.0	13.0	18.4	1.7	5.7	7.7	13.0	16.5	23.4
75 FM	(a)	1.7	3.4	4.1	5.3	6.0	7.3	1.7	3.5	4.1	5.4	5.8	7.0
	(b)	1.7	4.6	5.7	7.7	8.5	10.2	1.7	4.6	5.7	7.8	8.7	10.9
	(c)	1.7	3.5	4.1	5.5	6.2	7.3	1.7	3.5	4.2	5.5	6.0	7.4
	(d)	1.8	5.0	6.3	8.9	10.5	15.1	1.7	5.2	6.7	10.8	12.9	17.5
	(e)	1.7	5.1	6.4	9.3	10.8	15.1	1.7	5.1	6.7	11.1	12.8	18.9
100 FM	(a)	1.7	3.5	4.1	5.3	5.8	6.9	1.7	3.5	4.1	5.3	5.7	6.9
	(b)	1.7	4.3	5.2	7.2	7.8	9.2	1.7	4.3	5.2	7.3	8.1	9.5
	(c)	1.7	3.5	4.1	5.5	6.1	7.1	1.7	3.5	4.1	5.5	5.9	7.0
	(d)	1.7	4.5	5.6	8.5	9.4	12.0	1.7	4.7	5.9	9.5	11.3	16.1
	(e)	1.7	4.6	5.7	8.5	9.6	12.6	1.7	4.7	5.9	9.1	10.7	17.3
25 SM	(a)	1.6	2.5	2.7	3.3	3.5	3.9	1.6	2.5	2.8	3.3	3.5	3.8
	(b)	1.6	2.5	2.7	3.3	3.5	3.9	1.6	2.5	2.8	3.3	3.5	4.0
	(c)	1.6	2.5	2.8	3.5	3.8	4.4	1.7	2.6	3.0	3.7	4.0	4.7
	(d)	1.6	3.3	4.2	7.0	8.2	10.1	1.6	3.2	4.1	6.3	7.4	10.1
	(e)	1.6	3.4	4.4	7.1	8.2	10.5	1.6	3.4	4.3	6.5	7.4	10.9
50 SM	(a)	1.6	2.4	2.7	3.3	3.5	3.8	1.6	2.5	2.7	3.3	3.5	3.9
	(b)	1.6	2.5	2.7	3.3	3.5	4.0	1.6	2.5	2.7	3.3	3.5	4.0
	(c)	1.6	2.5	2.8	3.4	3.7	4.2	1.6	2.5	2.8	3.5	3.8	4.4
	(d)	1.7	3.0	3.7	5.6	6.5	8.8	1.6	3.0	3.6	5.3	6.1	7.9
	(e)	1.6	3.0	3.6	5.4	6.3	8.3	1.6	3.0	3.6	5.4	6.1	8.1
75 SM	(a)	1.6	2.5	2.7	3.3	3.5	3.8	1.6	2.5	2.7	3.3	3.5	3.9
	(b)	1.6	2.4	2.7	3.3	3.5	4.0	1.6	2.4	2.7	3.4	3.5	4.0
	(c)	1.6	2.5	2.8	3.3	3.6	4.0	1.6	2.5	2.8	3.4	3.7	4.2
	(d)	1.7	2.8	3.3	4.7	5.3	7.5	1.6	2.8	3.3	4.6	5.2	6.9
	(e)	1.6	2.8	3.3	4.9	5.7	7.6	1.6	2.8	3.4	4.8	5.4	7.0
100 SM	(a)	1.6	2.5	2.7	3.3	3.5	3.9	1.6	2.4	2.7	3.3	3.5	3.9
	(b)	1.6	2.5	2.7	3.3	3.5	3.8	1.6	2.5	2.7	3.3	3.5	3.9
	(c)	1.6	2.5	2.8	3.4	3.6	4.1	1.6	2.5	2.8	3.3	3.6	4.2
	(d)	1.6	2.7	3.1	4.2	4.7	6.2	1.6	2.7	3.1	4.3	4.7	6.4
	(e)	1.6	2.7	3.1	4.3	4.7	6.5	1.6	2.7	3.2	4.2	4.7	5.4

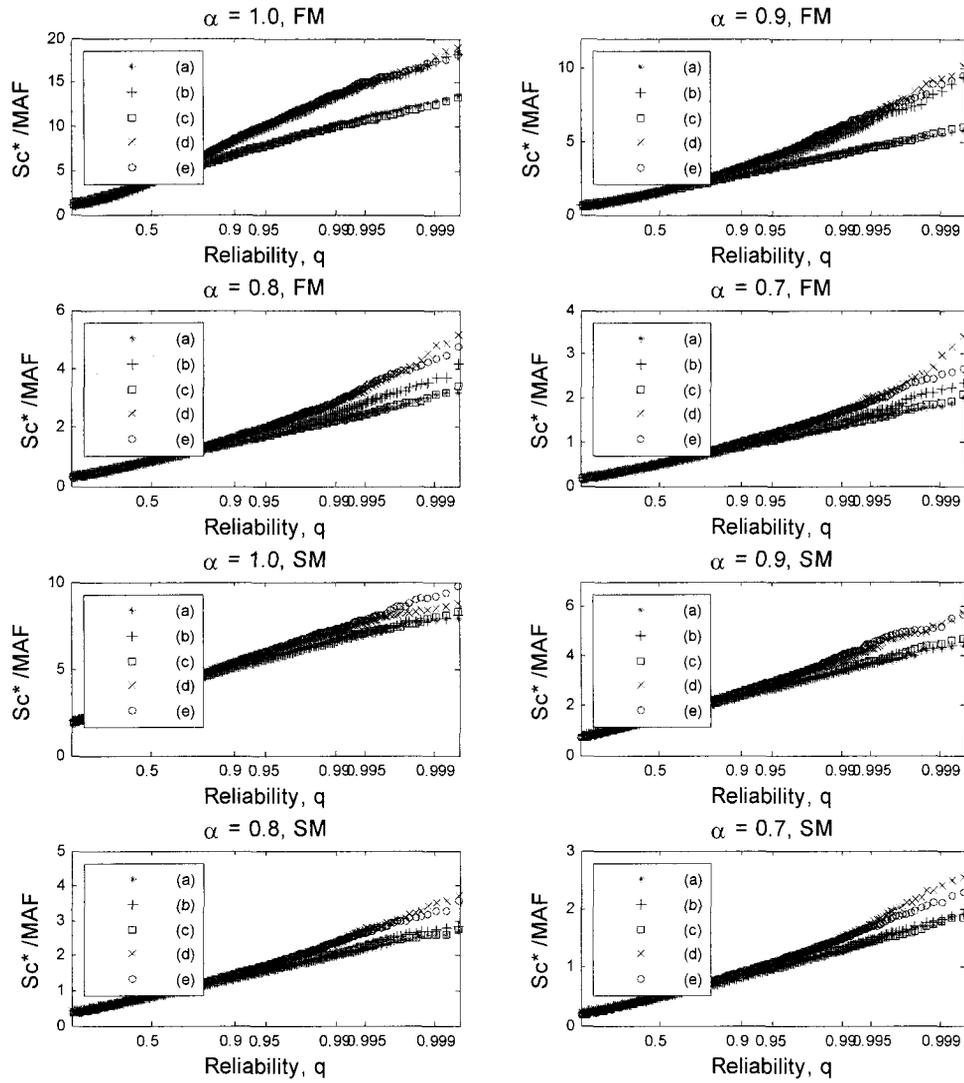


Figure 2.A13: Simulated storage capacities  $Sc^*$  for different demand levels associated with reliabilities (Lee's Ferry in the Colorado River Basin, asymptotic analysis,  $n=100$ ,  $N_d=100$ ).

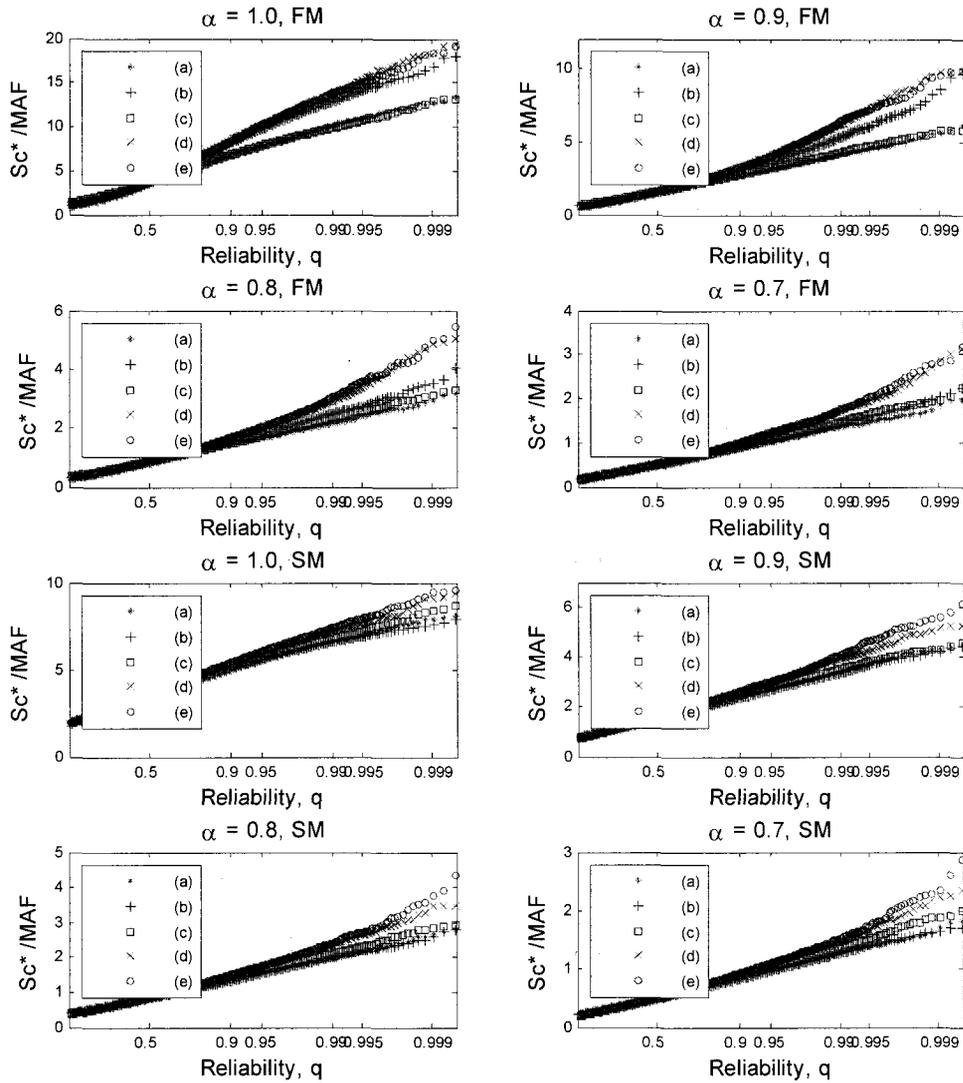


Figure 2.A14: Simulated storage capacities  $Sc^*$  for different demand levels associated with reliabilities (Lee's Ferry in the Colorado River Basin, Bayesian analysis,  $n=100$ ,  $N_d=100$ ).

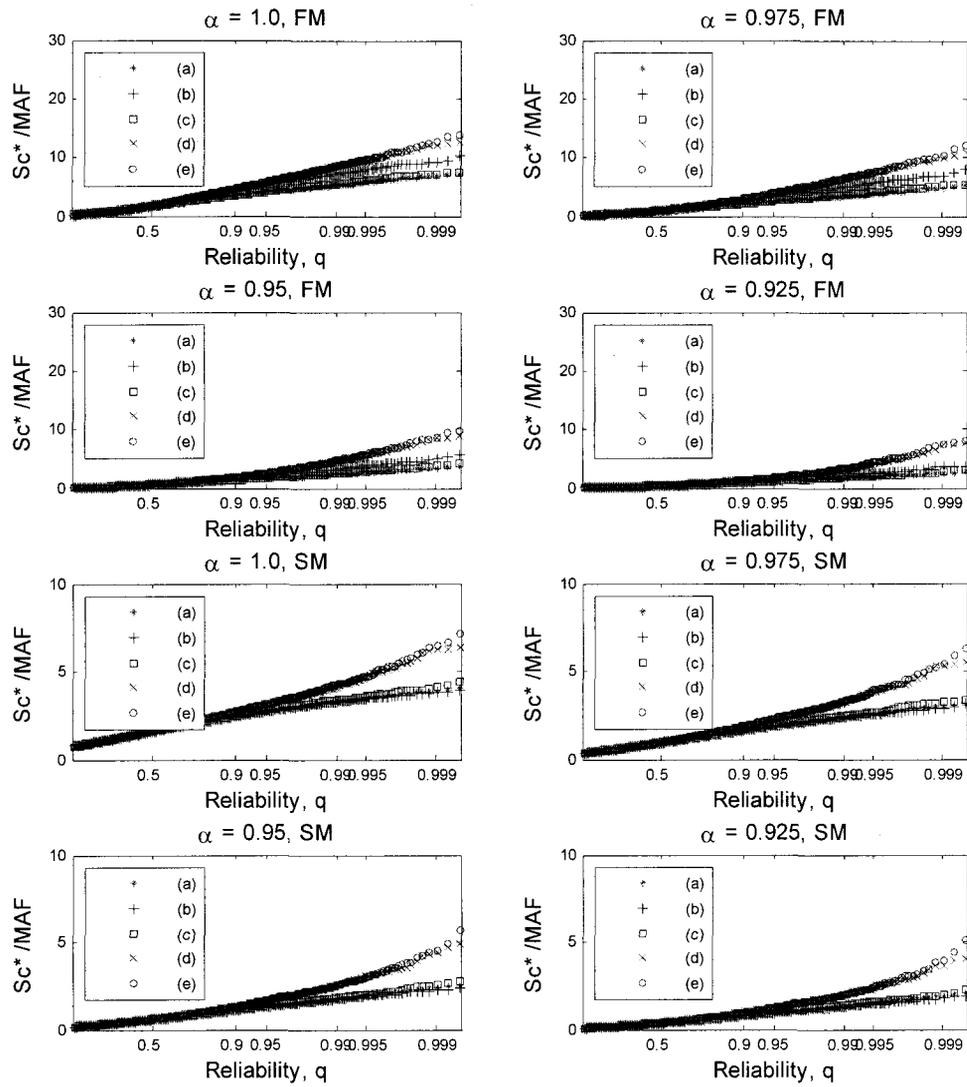


Figure 2.A15: Simulated storage capacities  $Sc^*$  for different demand levels associated with reliabilities (St. Lawrence River, asymptotic analysis,  $n=100$ ,  $N_d=100$ ).

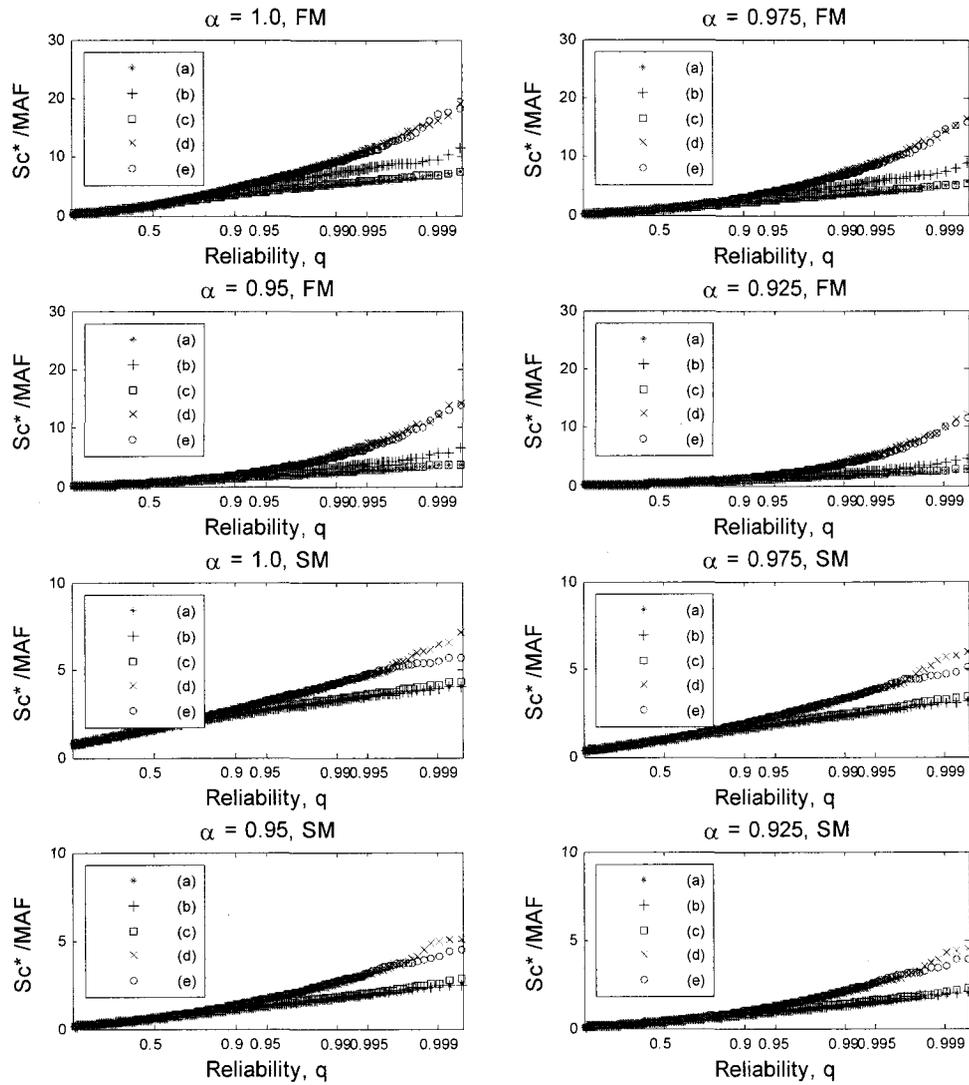


Figure 2.A16: Simulated storage capacities  $Sc^*$  for different demand levels associated with reliabilities (St. Lawrence River, Bayesian analysis,  $n=100$ ,  $N_d=100$ ).

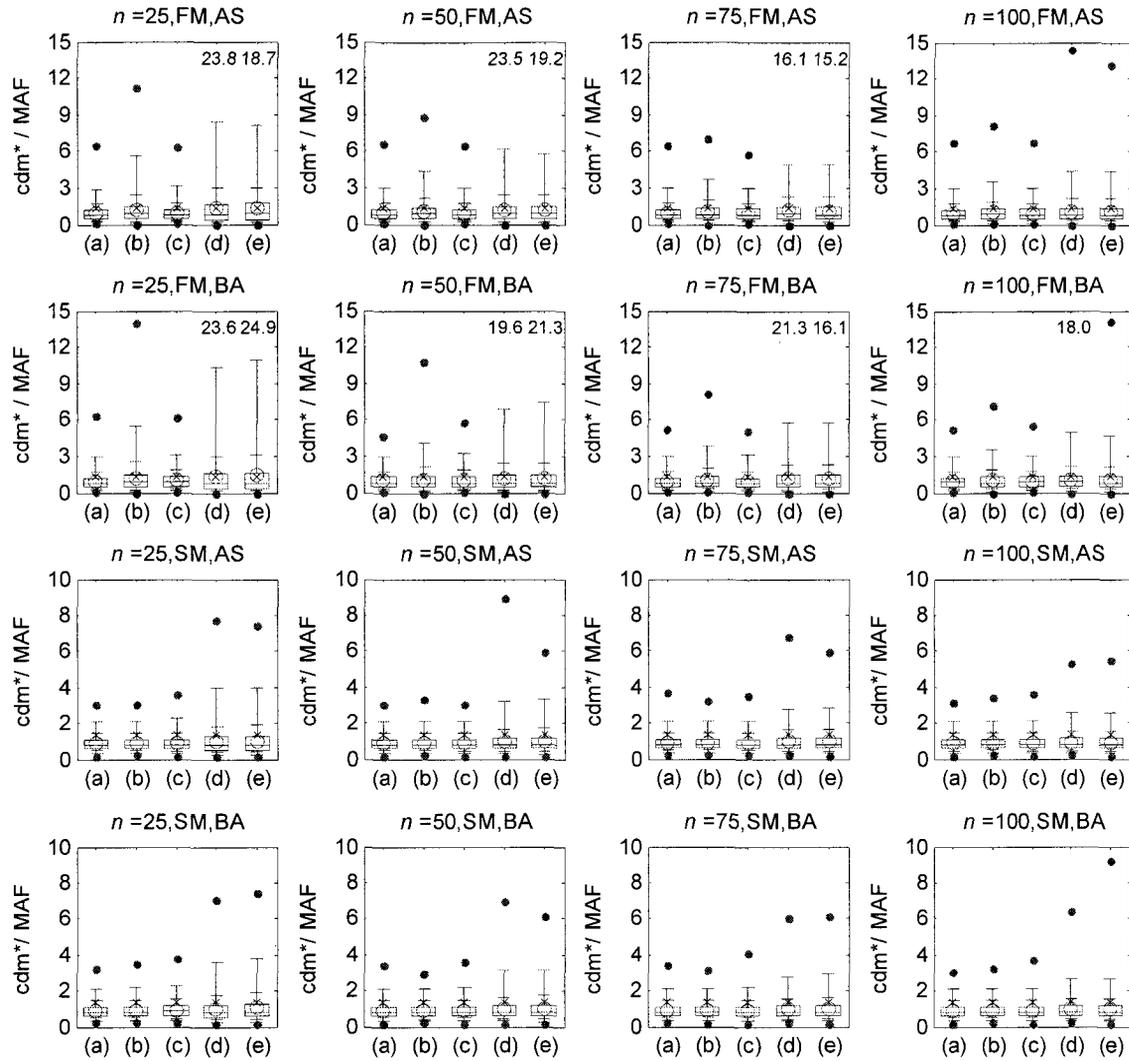


Figure 2.A17: Distribution of generated critical drought magnitude  $cdm^*$  for different sample sizes at St. Lawrence River ( $N_d=59$ ).

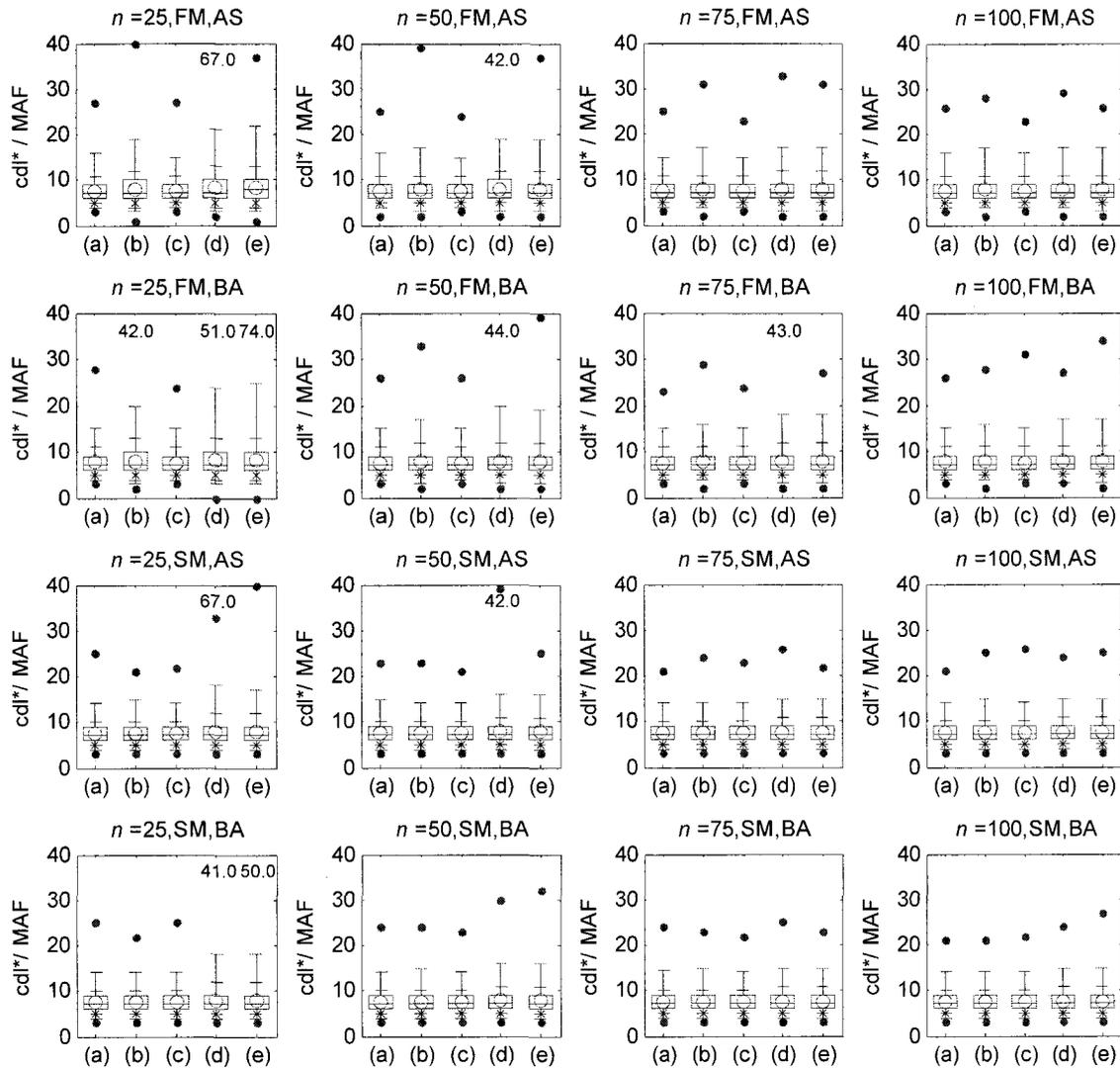


Figure 2.A18: Distribution of generated critical drought length  $cdl^*$  for different sample sizes, at Lee's Ferry in Colorado River Basin ( $N_d=98$ ).

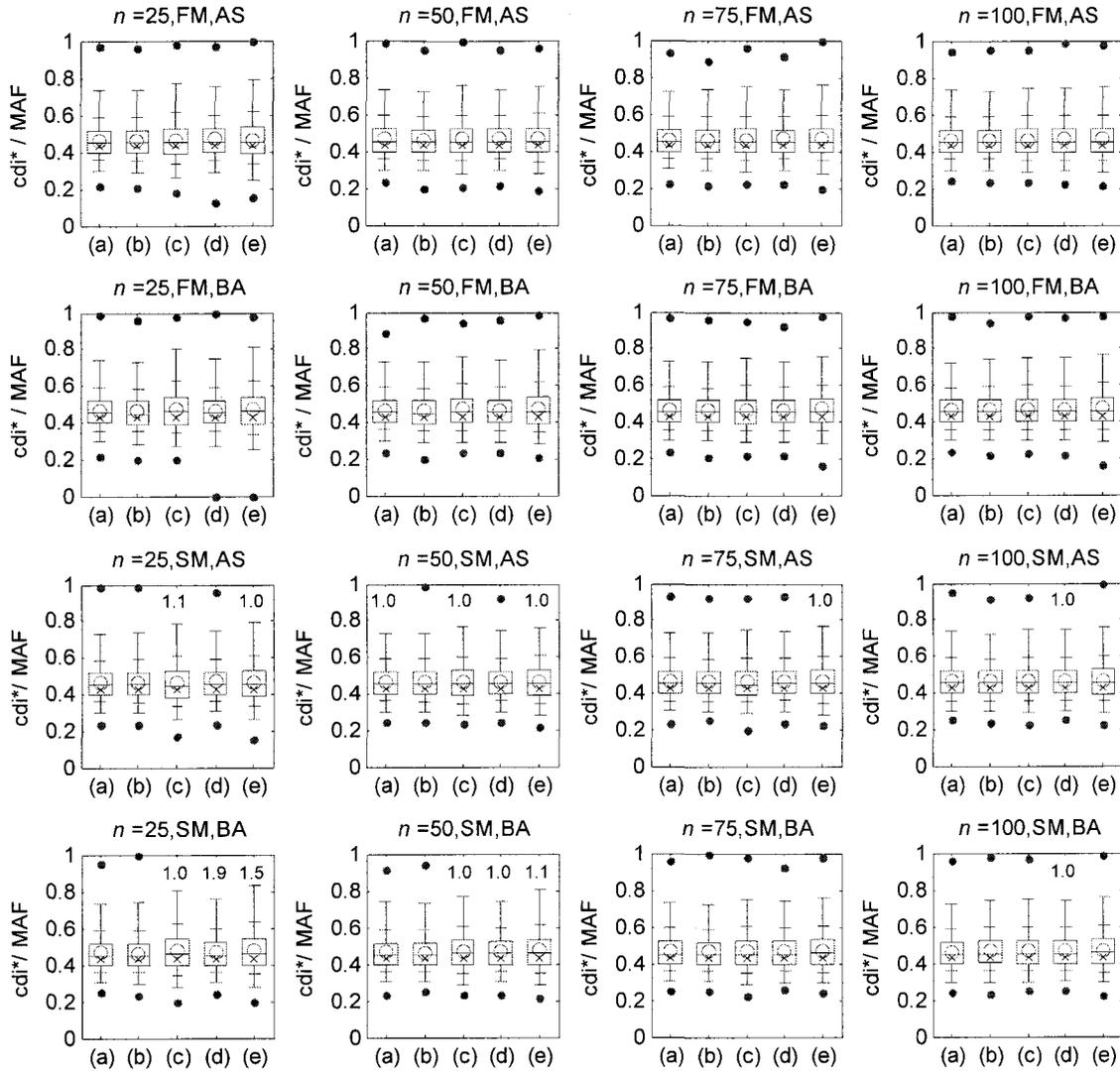


Figure 2.A19: Distribution of generated critical drought intensity  $cdi^*$  for different sample sizes, at Lee's Ferry in the Colorado River Basin ( $N_d=98$ ).

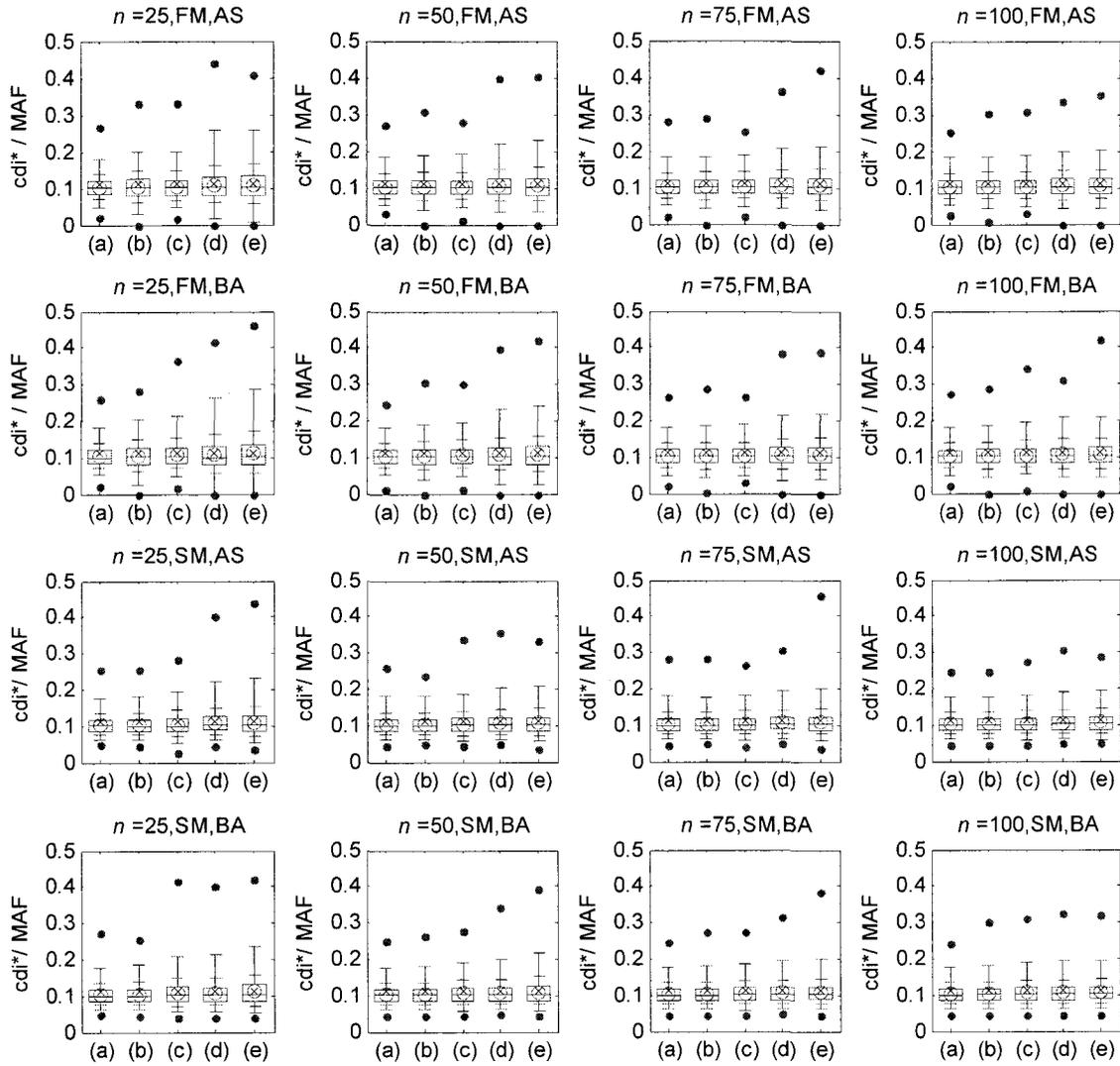


Figure 2.A20: Distribution of generated critical drought intensity  $cdi^*$  for different sample sizes, at St. Lawrence River ( $N_d=59$ ).

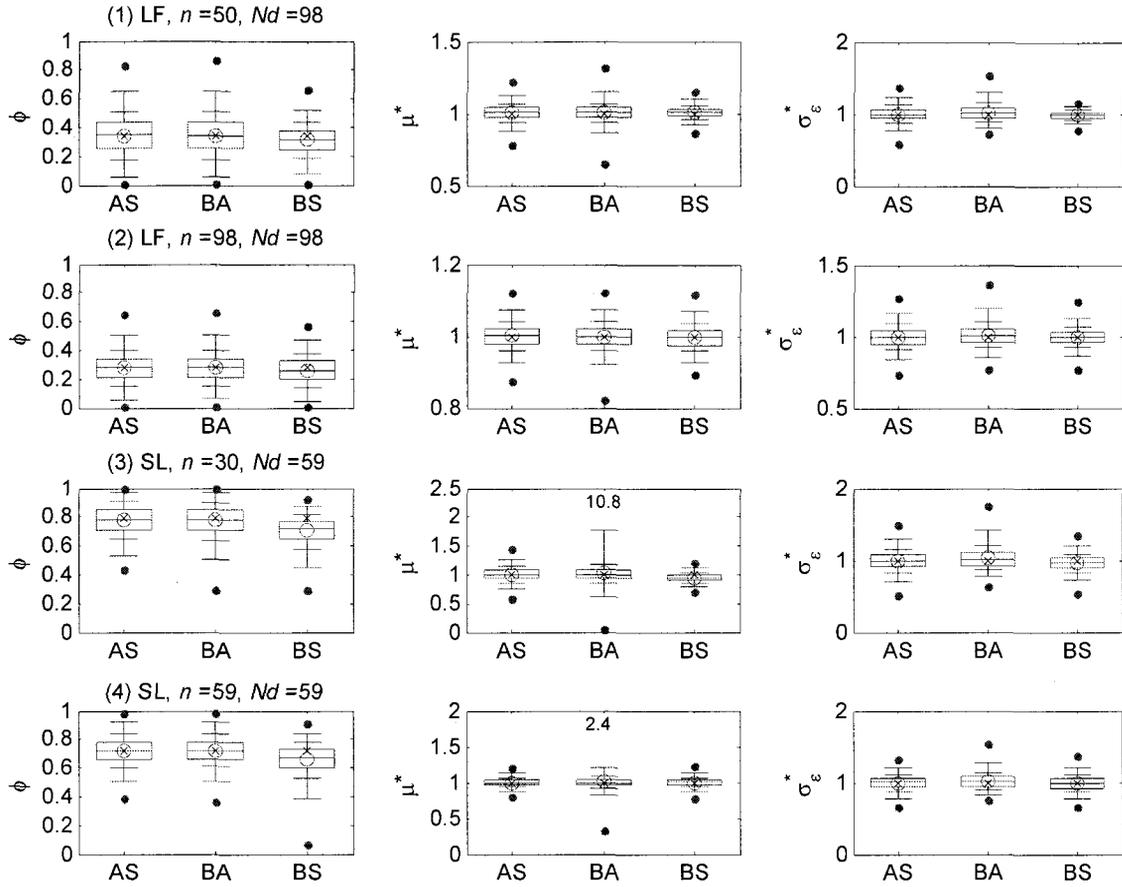


Figure 2.A21: Distribution of parameter estimates based on 3 different uncertainty incorporation methods: asymptotic analysis (AS), Bayesian analysis (BA), and bootstrapping (BS), where subscript \* means ‘scaled by historical value’.

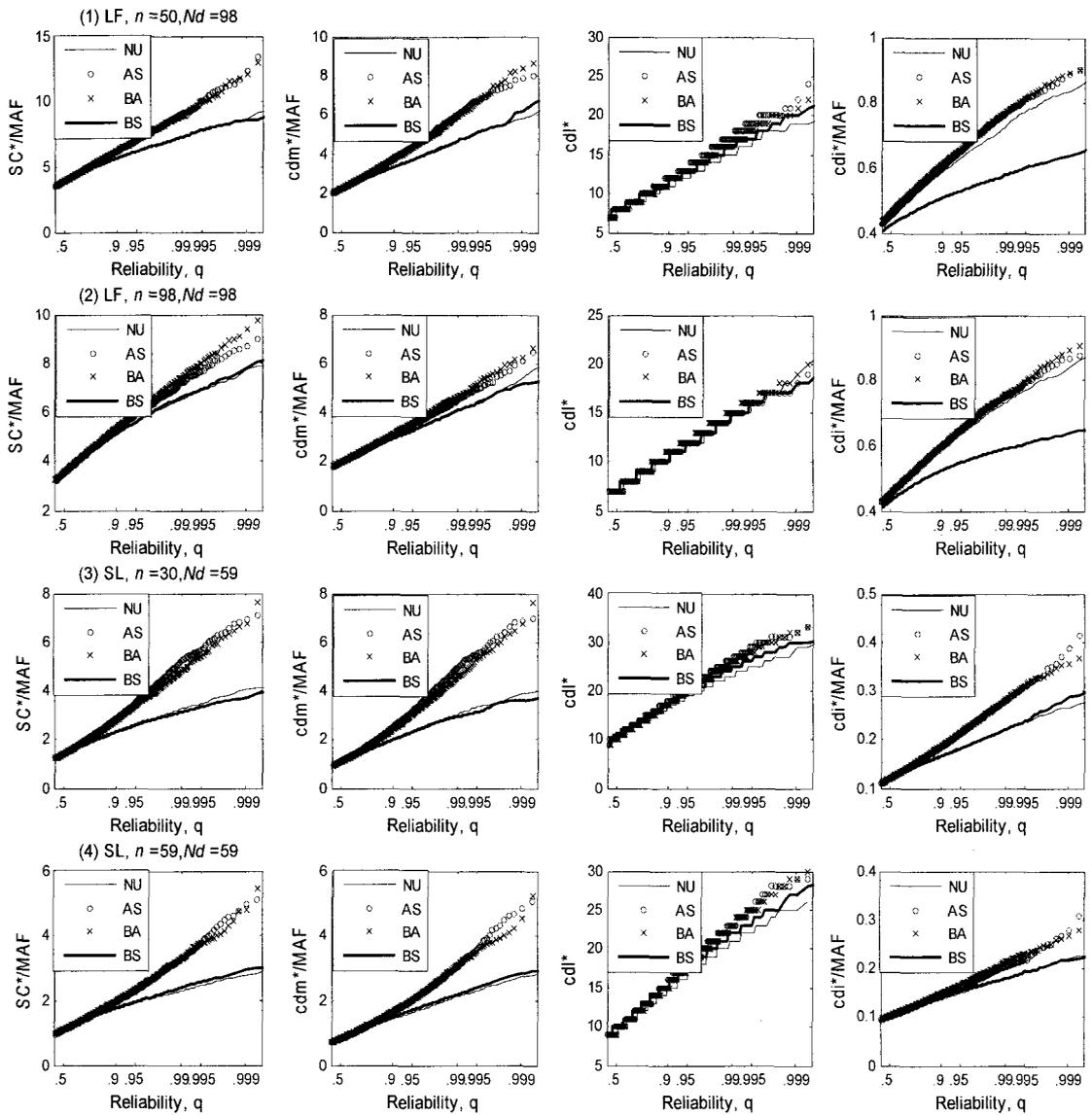


Figure 2.A22: Empirical distribution of generated storage capacities  $Sc^*$ , critical drought magnitude  $cdm^*$ , critical drought length  $cdl^*$ , and critical drought intensity  $cdi^*$  for different sample sizes in option *SM*.

## Chapter III

### UNCERTAINTY CONSIDERATION IN TEMPORAL DISAGGREGATION MODEL

**Abstract:** Synthetic monthly streamflows will be generated based on parametric generation model. Since the available historical information is limited, estimated parameters of the generation model have uncertainty as a consequence, thus this parameter uncertainty will translate into generated monthly flows. In this study, this parameter uncertainty effect will be statistically quantified and the effect on generated flows will be evaluated from the simulation experiment. As a way of quantifying the parameter uncertainty, sampling distributions are theoretically derived based on the limiting property of parameter estimates and Bayesian inference. Simulation experiments of monthly streamflows are performed with different parameter sets of which uncertainty is incorporated. A coupled AR(1) annual generation model and temporal disaggregation model will be employed for the generation. For generated monthly streamflows sets, the impact of parameter uncertainty is examined by inspecting monthly statistics and design variables related to them. Uncertainty effects which have arisen from annual and monthly generation stages will be compared respectively or together based on several different combinations of generation options. The parameter

uncertainty in disaggregation increases the variability of generated monthly statistics depending on different sample sizes, but it is shown to be not as significant for generated design variables as the natural uncertainty in generated annual flows. The parameter uncertainty which arises from the annual generation step propagates into monthly streamflows through the disaggregation and affects generated design variables of storage capacity and critical drought indices most significantly.

### **3.1 Introduction**

As an alternative of using direct monthly streamflow generation techniques, like the periodic autoregressive moving average model, several temporal disaggregation models have been proposed in modeling monthly streamflows generation for the purpose of preserving statistical characteristics of both annual and seasonal streamflows at a site of interest. The full temporal disaggregation model, of which structure is intended to reproduce the cross relationships simultaneously between annual flows and all seasonal flows, might result in a huge number of parameters; e.g. 156 parameters for the Valencia Schaake temporal disaggregation model and 168 parameters for the Mejia-Rousselle model are required in generating monthly streamflows from the annual for only a single site. As a substitute, the condensed disaggregation model has been proposed to resolve the precision problem related with the number of parameters (Lane, 1979 (LAST); Stedinger et al., 1985 (SPC)). The condensed disaggregation model (Lane, 1979) needs only 36 parameters by considering season by season, and this staged disaggregation

scheme has been known to be considerably efficient and adaptable in real synthetic streamflow modeling. However, even though an optimal disaggregation model structure could be determined, a precision problem would not be guaranteed due to the shortage of historical streamflows data. A disaggregation model has uncertain parameters and consequently, design variables related to them are uncertain. In this study, uncertainty of parameters of the temporal disaggregation model is theoretically taken into account through the Bayesian inference and the large sample theory. As in a previous work, Stedinger et al. (1985) suggested the procedure to account for parameter uncertainty regarding their model in a simple form within a Bayesian framework. In our study, we derive the asymptotic variance-covariance matrix of parameter estimates based on the maximum likelihood estimation. Also, alternative Bayesian posterior distribution is implemented to investigate the uncertainty effect for smaller sample size. We assess the proposed uncertainty analysis techniques by simulating synthetic streamflows and comparing a number of storage and drought related statistics. The applicability of the suggested methods using actual streamflow data of the Colorado River and the St. Lawrence River will be presented.

## **3.2 Generation of monthly synthetic streamflows using disaggregation models**

### **3.2.1 Traditional temporal disaggregation models**

Disaggregation schemes were basically suggested for the purpose of

conservation of the statistical properties at more than one level of aggregation (Harms and Campbell, 1967) with the purpose of preserving the statistical properties at both original (key) series and sub-series levels by a linear dependence model. It could allow for reducing the number of parameters without loss of expected properties in the generated data, as well as more flexibility in generation techniques (Salas et al., 2000).

Valencia and Schaake (1973) proposed a theoretically well established disaggregation model (named as VS model throughout this paper), and Mejia and Rousselle (1976) introduced an extension of the VS model which mainly dealt with the temporal disaggregation of annual flow to seasonal flow (MS model). Since then, various approaches have been implemented by Lane (1979), Salas et al. (1980), Stedinger and Pei (1982), Stedinger and Vogel (1984), Grygier and Stedinger (1988, 1990), Santos and Salas (1992), Lane and Frevert (1990), and Salas et al. (2000). Recently, the disaggregation technique has been widely suggested as the efficient one in modeling a hydrologic time series in a complicated multi-site system.

The temporal disaggregation model by Valencia and Schaake (1973) is first given by:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{u}, \quad (3.1)$$

where  $\mathbf{Y}$  is the seasonal streamflows being generated (low level),  $\mathbf{X}$  is the annual streamflows (high level),  $\mathbf{u}$  is the error terms, and  $\mathbf{A}, \mathbf{B}$  are parameter matrices. Also,  $\mathbf{u}$  is assumed as independent with  $\mathbf{X}$  and multivariate normally distributed with zero mean and unit diagonal matrix as the variance-covariance matrix. By choosing the proper estimates of parameter matrices, the correlation structures between low-level, high-level flows are preserved, respectively, and between low-level and high-level

streamflows as well. A transformation would be required to ensure that  $\mathbf{X}$  and  $\mathbf{Y}$  are normally distributed (Valencia and Schaake, 1973). For generating monthly streamflows from the annual based on the VS scheme, it can not be assured to explicitly preserve correlations of the consecutive months between different years; e.g. the first month in the current year and the last month in the previous year. In order to overcome this drawback, modified scheme was proposed by linking the past with the current value being disaggregated as (Mejia and Rousselle, 1976):

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{u} + \mathbf{C}\mathbf{Z}, \quad (3.2)$$

where the additional term  $\mathbf{Z}$  is the column matrix containing as many seasonal values from the previous year as are desired, and  $\mathbf{C}$  is the parameter matrix.

### 3.2.2 Condensed temporal disaggregation models

Even though disaggregation methods could explain the cross correlation of all seasonal flows, a serious drawback of the disaggregation methods arises from the fact that it needs too many parameters coming from reproducing many cross correlations. For this reason the condensed or staged (stepped or cascaded) disaggregation models have been suggested that could decrease the size of parameter matrices involved (correspondingly the large number of parameters) and make those models efficient in computation work (Curry and Bras, 1978; Salas et al., 1980; Loucks et al., 1981; Bras and Rodriguez-Iturbe, 1985; Stedinger et al., 1985; Santos and Salas, 1992). For example, to reduce the number of parameters Curry and Bras (1978) applied staged disaggregation in which annual data is disaggregated into quarterly data and then into

monthly data. Also, in order to come up with how efficiently the number of parameters in the seasonal disaggregation process could be reduced, Lane (1979) proposed the simplification scheme (known as the condensed model) in the computer package of LAST (Lane's Applied Stochastic Techniques) based on the classic disaggregation models by MR (1976) as:

$$Y_{v,\tau} = \alpha_{\tau} + \beta_{\tau} Y_{v,\tau-1} + \gamma_{\tau} X_{\nu} + V_{v,\tau} , \quad (3.3)$$

where  $\tau$  denotes the different season of concern and a total of 36 parameters are needed for  $\tau=12$ .  $X_{\nu}$  is the high level flows (annual flows) in year  $\nu$ ,  $Y_{v,\tau}$  and  $Y_{v,\tau-1}$  are the low level flows (monthly flows) for season  $\tau$  and  $\tau-1$  in year  $\nu$ .  $V_{v,\tau}$  refers to the innovation term, which is independent and normally distributed with zero mean and variance of  $\sigma_{\tau}$  for each season  $\tau$ . Annual and monthly variables  $X_{\nu}$ ,  $Y_{v,\tau}$ , and  $Y_{v,\tau-1}$  are assumed to follow normal distribution with mean zero. Lane (1982) discussed that even though the statistical relationship between  $Y_t$  and  $X_t$ , between  $Y_t$  itself, and between  $Y_t$  and  $Y_{t-1}$  are preserved, this is not the case between  $X_t$  and  $Y_{t-1}$  since it is not represented by the disaggregation model structure. For this reason, adjusted moments explaining this relationship were suggested, which could allow moments to be precisely conserved in the generated sequences, although the moments are not equivalent to the historical ones but to the adjusted moments. However, this model has a drawback: since the low-levels, i.e. seasonal streamflows, are not generated simultaneously, these are not summed exactly to the high-level series, i.e. annual streamflows. That discrepancy can be resolved by recalculating the annual values or by introducing the adjustment procedure make generated seasonal values exactly added to

annual ones (Salas et al., 1980; Lane, 1979; Stedinger and Vogel, 1984; Stedinger et al., 1985; Lane and Frevert, 1990). Lane and Frevert (1990) proposed three methods of adjusting seasonal flows equal to annual flow.

Based on the log normality of real streamflows, Stedinger, et al. (1985) proposed another temporal disaggregation scheme (SPC) by adding an extra term in LAST so as to conserve the additivity of seasonal flows, which is given by:

$$Y_{v,1} = \alpha_1 + \beta_1 X_v + V_{v,1} \quad \text{for } \tau = 1, \quad (3.4a)$$

$$Y_{v,\tau} = \alpha_\tau + \beta_\tau X_v + \gamma_\tau Y_{v,\tau-1} + \delta_\tau \left[ \sum_{s=1}^{\tau-1} w_s Y_{v,s} \right] + V_{v,\tau} \quad \text{for } \tau \geq 2, \quad (3.4b)$$

where  $\alpha_t, \beta_t, \gamma_t, \delta_t$  are parameters and  $V_{v,\tau}$  is independent normally distributed residual term with zero mean and variance  $\sigma_\tau$ .  $Y_{v,\tau}, X_v$  are the normalized low-level and high-level streamflows as shown in VS and MR schemes, of which  $Y_{v,\tau}, X_v$  are transformed from the original series to be normally distributed. The weighting factor  $w_s$  is defined as the derivatives of untransformed monthly flows with transformed monthly flows,  $dQ_{v,\tau}/dY_{v,\tau}$  evaluated at the expected value of the annual flow. For example, the weight factor  $w_s$  is given by  $w_s = \exp(\mu_{Y_\tau} + \sigma_{Y_\tau}^2/2)$  for three parameter lognormal transformation where  $\mu_{Y_\tau}, \sigma_{Y_\tau}^2$  are the mean and variance of  $Y_{v,\tau}$  (Grygier and Stedinger, 1990). To ensure that the sum of generated monthly streamflows is equal to actual annual streamflows an additional adjustment process would be required as a final step.

As another way of condensed temporal disaggregation, Santos and Salas (1992) proposed the stepwise disaggregation procedure (STEP model) assuming the normality of

real streamflows. It can be simply shown that SPC and STEP have the identical structure in case that (3.4) is applied into streamflows without transformation (letting  $w_s=1$ ).

### 3.3 Theoretical consideration of parameter uncertainties

#### 3.3.1 Asymptotic approach

Rewrite SPC temporal disaggregation scheme of Eq. (3.4) as a regression form for  $\tau \geq 2$ :

$$\mathbf{Y}_\tau = \mathbf{X}_\tau \boldsymbol{\theta}_\tau + \mathbf{V}_\tau, \quad (3.5)$$

where  $\mathbf{Y}_\tau = (Y_{1,\tau}, Y_{2,\tau}, \dots, Y_{n,\tau})'$ ,  $\mathbf{V}_\tau = (V_{1,\tau}, V_{2,\tau}, \dots, V_{n,\tau})'$ ,  $\boldsymbol{\theta}_\tau = (\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau)'$ ,

$$\mathbf{X}_\tau = \begin{bmatrix} 1 & Y_{1,\tau-1} & X_1 & Z_{1,\tau-1} \\ 1 & Y_{2,\tau-1} & X_2 & Z_{2,\tau-1} \\ & \vdots & & \\ 1 & Y_{n,\tau-1} & X_n & Z_{n,\tau-1} \end{bmatrix},$$

with  $Z_{v,\tau-1} = \sum_{s=1}^{\tau-1} w_s Y_{v,s}$  and  $n$  means sample size. The maximum likelihood

estimators (MLE) of  $\boldsymbol{\theta}_\tau = (\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau)'$  and the variance of the error vector  $\mathbf{V}_\tau$ ,  $\sigma_\tau^2$

are given by, respectively:

$$\hat{\boldsymbol{\theta}}_\tau = (\mathbf{X}'_\tau \mathbf{X}_\tau)^{-1} \mathbf{X}'_\tau \mathbf{Y}_\tau, \quad (3.6)$$

$$\hat{\sigma}_\tau^2 = \frac{(\mathbf{Y}_\tau - \mathbf{X}_\tau \hat{\boldsymbol{\theta}}_\tau)' (\mathbf{Y}_\tau - \mathbf{X}_\tau \hat{\boldsymbol{\theta}}_\tau)}{n - k}, \quad (3.7)$$

where  $k$  is the rank of  $\mathbf{X}_\tau$ . The  $4 \times 4$  matrix  $E(-I_{ij}) = I(\theta_\tau, \sigma_\tau)$  is referred to as an information matrix for parameter  $\theta_\tau$  and  $\sigma_\tau$  where  $I_{ij}$  is defined by the 2<sup>nd</sup> derivatives of log-likelihood function. For a large sample, the variance-covariance matrix  $V(\hat{\theta}_\tau, \hat{\sigma}_\tau^2)$  for MLE of  $(\hat{\theta}_\tau, \hat{\sigma}_\tau^2)$  can be approximately given by the inverse of the information matrix. Hence, the asymptotic distribution of MLE in the SPC temporal disaggregation model is given by a multivariate normal distribution with mean vector  $(\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau, \sigma_\tau)'$  and symmetric variance-covariance matrix as (Appendix 3.A1):

$$V(\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\gamma}_\tau, \hat{\delta}_\tau, \hat{\sigma}_\tau^2) = \frac{\sigma_\tau^2}{n} \begin{bmatrix} 1 & \hat{\mu}_{Y_{\tau-1}} & \hat{\mu}_X & \hat{\mu}_{Z_{\tau-1}} & 0 \\ \hat{\sigma}_{Y_{\tau-1}}^2 + \hat{\mu}_{Y_{\tau-1}}^2 & \hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}} & \hat{\sigma}_{Z_{\tau-1}Y_{\tau-1}} + \hat{\mu}_{Z_{\tau-1}} \hat{\mu}_{Y_{\tau-1}} & 0 & 0 \\ \hat{\sigma}_X^2 + \hat{\mu}_X^2 & & \hat{\sigma}_{XZ_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Z_{\tau-1}} & 0 & 0 \\ \text{symm.} & & & \hat{\sigma}_{Z_{\tau-1}}^2 + \hat{\mu}_{Z_{\tau-1}}^2 & 0 \\ & & & & 2 \end{bmatrix}^{-1} \quad (3.8)$$

Thus, two separate asymptotic distributions can be defined for  $\hat{\theta}_\tau$  and  $\hat{\sigma}_\tau^2$  as:

$$\hat{\theta}_\tau \sim MVN\left(\theta_\tau, \frac{\sigma_\tau^2}{n} E(\mathbf{X}'_\tau \mathbf{X}_\tau)^{-1}\right), \quad (3.9)$$

$$\hat{\sigma}_\tau^2 \sim N\left(\sigma_\tau^2, \frac{\sigma_\tau^2}{2n}\right). \quad (3.10)$$

In a similar way, asymptotic variance-covariance matrix of parameter estimates for LAST model can be given as (Appendix 3.A2):

$$V(\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\gamma}_\tau, \hat{\sigma}_\tau) = \frac{\sigma_\tau^2}{n} \begin{bmatrix} 1 & & & & \\ & \hat{\mu}_{Y_{\tau-1}} & & \hat{\mu}_X & \\ & \hat{\sigma}_{Y_{\tau-1}}^2 + \hat{\mu}_{Y_{\tau-1}}^2 & \hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}} & & \\ & & \hat{\sigma}_X^2 + \hat{\mu}_X^2 & & \\ \text{symm.} & & & & 2 \end{bmatrix}^{-1}. \quad (3.11)$$

### 3.3.2 Bayesian inference

As noted in the previous section, the temporal disaggregation model can be translated into the form of normal multiple regression with one fixed parameter of innovation term. Applying this normal regression model enables one to incorporate parameter uncertainties into the temporal disaggregation schemes in the Bayesian framework, which has been suggested in theoretical Bayesian studies (Zeller, 1974; Box and Tiao, 1973) and in the application to stochastic hydrology (Stedinger et al., 1985). Consider the relationship between annual and seasonal streamflow series season by season, separately, and assume that prior distributions of parameter sets of temporal disaggregation scheme are independent (Box and Tiao, 1973; Zeller, 1974; Stedinger et al., 1985). Introducing the Bayesian framework shows that parameter uncertainties can be explained in terms of posterior distributions. From (3.5), the likelihood function of  $\mathbf{Y}_\tau$  given  $\mathbf{X}_\tau$ ,  $\theta_\tau$ ,  $\sigma_\tau$  follows as

$$\begin{aligned} L(\mathbf{Y}_\tau; \mathbf{X}_\tau, \theta_\tau, \sigma_\tau) &= \frac{1}{(2\pi\sigma_\tau^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_\tau^2}(\mathbf{Y}_\tau - \mathbf{X}_\tau\theta_\tau)'(\mathbf{Y}_\tau - \mathbf{X}_\tau\theta_\tau)\right] \\ &\propto \frac{1}{\sigma_\tau^n} \exp\left[-\frac{1}{2\sigma_\tau^2}\left(d\hat{\sigma}_\tau^2 + (\theta_\tau - \hat{\theta}_\tau)' \mathbf{X}_\tau' \mathbf{X}_\tau (\theta_\tau - \hat{\theta}_\tau)\right)\right], \end{aligned} \quad (3.12)$$

where mean square error  $\hat{\sigma}_\tau^2 = d^{-1}(\mathbf{Y}_\tau - \mathbf{X}_\tau\hat{\theta}_\tau)'(\mathbf{Y}_\tau - \mathbf{X}_\tau\hat{\theta}_\tau)$  are sufficient statistics

(Zeller, 1974) and  $\eta = n - k$  is degree of freedom. Under the assumption that the information about the parameters is diffusive and the prior of each parameter element of  $\theta_\tau$  is independent of one another, the joint prior distribution yields:

$$p(\theta_\tau, \sigma_\tau) \propto \frac{1}{\sigma_\tau}, -\infty < \alpha_i, \beta_i, \gamma_i, \delta_i < \infty, 0 < \sigma_\tau < \infty, \quad (3.13)$$

and then the joint posterior distribution is derived as:

$$p(\theta_\tau, \sigma_\tau | \mathbf{Y}_\tau, \mathbf{X}_\tau) \propto \frac{1}{\sigma_\tau^{n+1}} \exp \left[ -\frac{1}{2\sigma_\tau^2} \left( \eta \hat{\sigma}_\tau^2 + (\theta_\tau - \hat{\theta}_\tau)' \mathbf{X}_\tau' \mathbf{X}_\tau (\theta_\tau - \hat{\theta}_\tau) \right) \right] \quad (3.14)$$

The marginal distribution of  $\sigma_\tau^2$  can be obtained by integrating (3.7) with respect to  $\theta_\tau$ , and it gives the form of inverted gamma distribution as (Zeller, 1974):

$$p(\sigma_\tau | \mathbf{Y}_\tau, \mathbf{X}_\tau) \propto \frac{1}{\sigma_\tau^{n-k+1}} \exp \left[ -\frac{\eta \hat{\sigma}_\tau^2}{2\sigma_\tau^2} \right], \quad (3.15)$$

which implies  $\eta \hat{\sigma}_\tau^2 / \sigma_\tau^2$  follows a  $\chi^2$  - distribution with  $\eta$  degree of freedom so that  $\sigma_\tau$  is distributed as inverse  $\chi^2$  - distribution. The posterior distribution of parameter matrix  $\theta_\tau$  can be taken into consideration for two possible cases; one is that the standard deviation  $\sigma_\tau$  is assumed known, and the other is unknown  $\sigma_\tau$ . First, when  $\sigma_\tau$  is assumed to be known the conditional posterior distribution of  $\theta_\tau$  given  $\sigma_\tau$  is the multivariate normal distribution with mean  $\hat{\theta}_\tau$  and variance-covariance matrix  $\sigma_\tau^2 (\mathbf{X}_\tau' \mathbf{X}_\tau)^{-1}$ . On the other hand, for the case of unknown  $\sigma_\tau$ , the joint posterior distribution of  $\theta_\tau$  is given by the multivariate  $t$  distribution with mean  $\hat{\theta}_\tau$ , variance-covariance matrix  $\hat{\sigma}_\tau^2 (\mathbf{X}_\tau' \mathbf{X}_\tau)^{-1}$  with  $\eta$  degree of freedom (Box and Tiao, 1973). For

known  $\sigma_\tau$  :

$$p(\theta_\tau | \mathbf{Y}_\tau, \mathbf{X}_\tau, \sigma_\tau) = \frac{|\mathbf{X}'_\tau \mathbf{X}_\tau|^{1/2}}{(2\pi)^{k/2} \sigma_\tau^k} \exp\left[-\frac{1}{2\sigma_\tau^2} (\theta_\tau - \hat{\theta}_\tau)' \mathbf{X}'_\tau \mathbf{X}_\tau (\theta_\tau - \hat{\theta}_\tau)\right], \quad (3.16)$$

and for unknown  $\sigma_\tau$  :

$$p(\theta_\tau | \mathbf{Y}_\tau, \mathbf{X}_\tau) = \frac{\Gamma[(\eta + k)/2] |\mathbf{X}'_\tau \mathbf{X}_\tau|^{1/2} \hat{\sigma}_\tau^{-k}}{[\Gamma(1/2)]^k \Gamma(\eta/2) \eta^{k/2}} \left[1 + \frac{(\theta_\tau - \hat{\theta}_\tau)' \mathbf{X}'_\tau \mathbf{X}_\tau (\theta_\tau - \hat{\theta}_\tau)}{\eta \hat{\sigma}_\tau^2}\right]^{-(\eta+k)/2} \quad (3.17)$$

It is well known that  $t$  distribution will become equivalent to normal distribution when the sample size is large enough and (3.16) has been employed to define the posterior distribution in the previous literature (Stedinger and Taylor, 1982; Stedinger et al., 1985). Comparison of (3.9) and (3.16) reveals that Bayesian posterior distribution is equivalent to the derived asymptotic distribution. That is, using (3.17) instead of (3.16) seems more appropriate in the Bayesian sense, and (3.17) will be used as the underlying posterior distribution in this study.

### 3.4 Uncertainty incorporation into synthetic streamflows

#### 3.4.1 Simulation experiments

Two monthly streamflow sets with distinct statistical characteristics were selected: Lee's Ferry in the Colorado River Basin and the St. Lawrence River (see Tables 3.1 and 3.2). Over 60% annual flows are from only three months (May, June, and July) in the Colorado River, while almost similar monthly flows are contributed to total annual flows in the St. Lawrence River. A significant difference in month-to-month serial

correlations and also the difference in variations in each monthly flow are remarkable for these two sites. In addition, significantly different variations in aggregated monthly flows (annual flows) are notable.

As in the traditional simulation, different traces of annual flows (5000 traces will be generated) are first generated with the same size as the historical flows (98 for Lee's Ferry and 59 for St. Lawrence River), and then those generated annual flows are disaggregated into monthly flows by using SPC and LAST temporal disaggregation models. The normality of flows is examined by using a skewness test with a 5% significant level for two historical flows series. Log-transformations are applied to both annual and monthly steamflows of Lee's Ferry flows; log-transformation with a location parameter for February, May and June flows and without a location parameter for the rest of the months and the annual. It is shown that the normality assumption cannot be rejected through the skewness test so that historical flows of the St. Lawrence River are directly applied to the monthly flow generations. Generated monthly flows are not exactly summed up to the annual because of the transformation, thus a adjustment process is required to make the sum of generated monthly flows equal to the annual. Several adjustment methods have been proposed in previous literature, among which the proportional scheme might distort marginal distributions the least and not generate negative flows negative (Grygier and Stedinger, 1988). Note that the SPC model applied to non-transformed data as in the St. Lawrence River has the equivalent form as Salas and Santos' model (1988).

In order to consider the parameter uncertainty, 5000 different parameter estimates sets of the AR(1) model for annual generation and those of SPC, LAST disaggregation

models are first sampled from their distributions, which are already defined by asymptotic and Bayesian approaches. In each generation, one of 5000 different parameter sets is substituted for the parameter set estimated from the historical flows.

### 3.4.2 Preliminary analysis of disaggregation models and uncertainty considerations

Before stepping forward to compare uncertainty effects with the different combinations of annual and monthly parameter uncertainty considerations, a comparative analysis of SPC and LAST disaggregation models has been taken into account based on both asymptotic and Bayesian approaches. Different sample sizes are considered to examine their related impacts since parameter uncertainties are closely related to the available sample size. Parameter estimates calculated from the historical flows are assumed to be population values, by which the expected values of parameter estimates are set as fixed ones, and then new parameter estimates are sampled from distributions of parameters by changing the variances corresponding to sample sizes of  $n=25, 50, 75, 100$ .

Storage and drought related statistics are implemented to inspect the simulated synthetic flow characteristics. For storage related statistics, a storage capacity is calculated using the sequent peak algorithm (Loucks et al., 1981):

$$S_t = \max(0, S_{t-1} + D_t - X_t) \quad , \quad t = 1, \dots, N_d, \quad (3.18)$$

where  $D_t$  = demand level,  $X_t$  = reservoir inflow,  $N_d$  = planning horizon, and  $S_0 = 0$ .

Then the storage capacity becomes:

$$S_c = \max(S_0, S_1, \dots, S_{N_d}). \quad (3.19)$$

As another tool for evaluating parameter uncertainty effect, drought related statistics will be utilized. For a given demand level (for example, mean monthly flows, MMF), the deficit occurs when annual streamflows  $Y_t < \text{MMF}$  during one or more months until  $Y_t > \text{MMF}$  again, which can be defined by its magnitude, by its duration, and by its intensity (magnitude divided by corresponding duration). Among the number of deficits in a given streamflow set, the maximum deficit magnitude, length and intensity in a given sample are referred to the critical drought length  $cdl$ , critical maximum magnitude  $cdm$ , and critical drought intensity  $cdi$  (Salas et al., 1980).

After generating 5000 different sets of monthly streamflows, synthetic storage capacities and critical indices are calculated and compared for different sample sizes, disaggregation models, and different uncertainty consideration approaches (Figures 3.1-3.2). A demand level is assumed equal to mean monthly flows (MMF) for both Lee's Ferry and the St. Lawrence River. The effect of different thresholds will be discussed in detail later in this chapter.

Increased variabilities of storage capacities by parameter uncertainties are shown in Figures 3.1 and 3.2, and those parameter uncertainty effects are more significant for smaller sample size. It is notable that those variabilities still can be found even in large sample size (e.g.  $n=100$ ) for both Lee's Ferry and the St. Lawrence River. Almost similar variabilities of generated storage capacities are expected for SPC and LAST models. The Bayesian approach results in a little increased variabilities of storage capacities (99% quantiles) than asymptotic distribution in smaller sample sizes (e.g.  $n \leq 50$ ), but it seems not significant in Lee's Ferry flows. However, the difference in variabilities of generated storage capacities between two approaches is shown to be

remarkable for the St. Lawrence River. Overall, upward shifts of storage capacities are shown as a parameter uncertainty effect. Distribution plots of generated critical drought indices are listed in Appendix 3.C1-3.C6. As in the case of the storage capacity, the Bayesian approach is shown to produce much variability in smaller sample size, especially for the St. Lawrence River. Also, SPC and LAST models show almost similar distributions for both sites and parameter uncertainties result upward shifts of drought indices. Tables 3.C1 and 3.C2 in Appendix illustrate coefficients of variation of generated storage capacities and drought indices. Consequently, parameter uncertainty effects on reservoir and drought related statistics are not to be ignored even in relatively large sample sizes ( $n \geq 75$ ), and their effects might be more considerable, which is affected by the statistical characteristics of data sets of concern in the synthetic streamflow generation. For further study, the SPC temporal disaggregation model and the Bayesian approach will be assumed as underlying schemes based on the comparative analysis result demonstrated in this section.

### **3.4.3 Uncertainty effect of temporal disaggregation model parameters combined with historical annual flows**

In a traditional disaggregation procedure, simulated annual streamflows are generally in use as input variables to generate monthly streamflows. Even though the parameter uncertainty effect is not incorporated in the stage of the annual flow generation, the error term which is implicated in the annual flow generation model might produce a variety of values not certain ones. These will be called by natural uncertainty in this

study and this natural uncertainty of the annual flows is usually embedded into the synthetic monthly flows in the general disaggregation procedure.

In order to eliminate the natural uncertainty effect in the stage of annual flows generation and to examine the uncertainty effect of disaggregation parameters themselves, historical annual flows are assumed as an input in the disaggregation stage instead of simulated annual flows. Assuming that given historical sample size of each site is large enough; estimated parameters from the historical flows data are used as the real parameter set. By changing the sample size, parameter estimates of the disaggregation model are sampled from posterior distributions and then those are assigned to new ones. That is, the expected values of estimated regression parameters are assumed to be the same as the historical one regardless of the sample size, and variances of parameter estimates are proportional to the historical ones (Stedinger et. al, 1985). Note that applied sample sizes is:  $n = 25, 50, 75, 98$  for Lee's Ferry and  $n = 15, 30, 45, 59$  for the St. Lawrence River data.

Table 3.3 represents increased quantiles of generated storage capacities by uncertain parameter estimates of the SPC temporal disaggregation model. It is shown that the mean and high quantiles ( $q \geq 75\%$ ) of generated storage capacities are not significantly influenced by disaggregation parameter uncertainties for Lee's Ferry, while significant increases of those estimates are found for the St. Lawrence River. For example, in 99% quantile of generated storage capacities, 3.7, 1.6, 0.6, 1.0% increase with a demand level of MMF and 8.1, 3.5, 1.5, 1.1% increase with a 80% MMF demand for Lee's Ferry, while 21.4, 12.0, 6.5, 5.4% increase with 95% MMF demand level and 50.1, 25.4, 12.4, 11.0% increase with 85% MMF demand level for the St. Lawrence River.

To be remarkable, increased standard deviation of storage capacities is visible for both statistically different steamflow sets.

#### **3.4.4 Uncertainty effect of parameter uncertainty of annual streamflow generation and temporal disaggregation**

Four different considerations of parameter uncertainty are utilized in order to examine the combined effect of parameter uncertainty in both levels of annual and monthly flow generations. As in the traditional monthly streamflow generation procedure based on the temporal disaggregation model, monthly streamflows are simply disaggregated from generated annual flows and set to Case 1. In this case, parameter uncertainty is not incorporated into both annual flow generation and temporal disaggregation (no parameter uncertainty). That is, only natural uncertainty from the generation models (annual or disaggregation) will be considered. In Case 2, parameter uncertainty of the temporal disaggregation model will be incorporated. Annual flows are generated without parameter uncertainty, and then those are disaggregated into monthly streamflows with parameter uncertainty incorporated. Parameter uncertainty in annual flow generation will be taken into account in Cases 3 and 4. Generated annual flows with parameter uncertainty are simply disaggregated into monthly streamflows without consideration of disaggregation parameters in Case 3. In Case 4, parameter uncertainty in both annual flow generation and temporal disaggregation will be applied. The four different combinations of natural and parameter uncertainty in the annual flow generation and temporal disaggregation are summarized as:

Cases	Annual flow generation	Temporal disaggregation
Case 1	NU	NU
Case 2	NU	NU + PU
Case 3	NU + PU	NU
Case 4	NU + PU	NU + PU

For each case, basic statistics of synthetic monthly flows such as mean, variance, month-to-month correlation and correlation between monthly to annual flows are evaluated. Also, as observed before, the distributions of the storage capacity and drought indices based on the generated monthly streamflows will be compared, which enables one to allow for the parameter uncertainty effect translated into design variables.

Figures 3.3 and 3.4 show monthly means, standard deviations, month-to-month serial correlations and month-to-annual correlations calculated from generated monthly flows for four different Cases assuming sample size is equal to 50. The parameter uncertainty in the annual generation (Cases 3 and 4) result in much variability in the mean and standard deviation of generated monthly streamflows compared with Cases 1 and 2. A log-transformation was applied in the generation of Lee's Ferry streamflows, thus simulated monthly flows were adjusted with respect to the annual flows, while the adjustment was not applied to the St. Lawrence River case. By this adjustment procedure, the proportionality of monthly flows to the annual is shown to be well preserved by the disaggregation model even with the parameter uncertainty incorporated into the model. The parameter uncertainty in the temporal disaggregation (Cases 2 and 4) exhibits increased month-to-month serial correlation and month-to-annual correlation

for each month compared with natural uncertainty in the disaggregation model (Case 1 and 3).

In the previous section, the uncertainty effect of monthly disaggregation models were shown to be visible and dependent on the available sample sizes. However, it is notable that there seems little difference in the distributions of storage capacities between Cases 1 and 2. Based on this, it can be demonstrated that the effect of parameter uncertainty in the monthly flows disaggregation on generated storage capacities is not as significant as one of natural uncertainty in the annual streamflow generation (Figures 3.5 and 3.6). That might be related to the disaggregation model structure that the proportionality of monthly flows to the annual is intended to be preserved. Even though the parameter uncertainty of the disaggregation model is translated into generated monthly flows, the mean and variance of generated monthly flows were not significantly increased by the parameter uncertainty as shown in Figures 3.3 and 3.4. Much variability of storage capacities is expected as shown in Cases 3 and 4, which might conclude that the parameter uncertainty in the annual generation is dominant over one in the disaggregation in the sense of the storage related statistics.

For the Lee's Ferry generation (Figure 3.5), there could be found the difference of variabilities of storage capacities in Cases 3 and 4 for smaller sample sizes ( $n \leq 50$ ), which is distinguishable with the case of the St. Lawrence River generation (Figure 3.6). For simple monthly flow generation (Case 1), generated storage capacities show larger variability in Lee's Ferry case than those in the St. Lawrence River, which might result from higher variation of historical monthly streamflows within a year. This statistical characteristics plays a role in increasing much variability of storage capacities if the

parameter uncertainty in the disaggregation would be combined with parameter uncertainty incorporated synthetic annual flows (compare Cases 3 and 4 for Lee's Ferry generation in Figure 3.5) when available sample size is relatively small ( $n \leq 50$ ).

The comparable results might be found in distributions of generated critical drought indices (see Figures 3.C7-3.C12 in Appendix). The parameter uncertainty in the disaggregation step is not as significant as the natural uncertainty in input variables (annual flows), and the parameter uncertainty in input variables affects most considerably on synthetic drought related statistics.

When determining storage and drought related statistics, different options of demand levels and planning horizons might be available. Assuming that the available sample size is equal to 50 for both Lee's Ferry and the St. Lawrence River, different demand levels are implemented to examine associated parameter uncertainty effects (100, 90, 80, 70% demand levels for Lee's Ferry and 100, 95, 09, 85% for St. Lawrence River), and distributions of generated storage capacities and drought indices are plotted in Figures 3.C13-3.C20 in the Appendix. Comparing distributions for four Cases shows the similar patterns for parameter uncertainty considerations over applied demand levels. Since different demand levels result in different variabilities of storage capacities and drought indices in the case without parameter uncertainty (Case 1), the increased variability of design variables by parameter uncertainty is expected to be different corresponding to different demand levels.

### **3.4.5 Brief summary of parameter uncertainty effect on design variables**

Table 3.4 gives a brief summary of the parameter uncertainty effect on synthetic storage capacity and critical drought magnitude obtained from generated monthly streamflows at Lee's Ferry and at the St. Lawrence River. For simplicity, Case 1 in which no parameter uncertainty is considered either in annual streamflow generation or in temporal disaggregation and Case 4 where parameter uncertainty is incorporated in both annual generation and in temporal disaggregation are numerically compared. Relative to Case 1, Case 4 shows increased expected value of the storage capacity with a range of from about 10% up to 66% depending on different demand levels and sample serial correlations; e.g. increased expected value is shown up to 66% for the St. Lawrence River with 90% MAF as a demand level. Moreover, the increased standard deviation of storage capacity by parameter uncertainty is shown to be about 73%-375% depending on different sample serial correlations and assumed demand levels. The parameter uncertainty effect is also visible for higher quantiles of storage capacity, e.g. about a 200% increase of the 99% quantile of the storage capacity for the case of the St. Lawrence River with 90% MAF demand level. Overall, less increase by the parameter uncertainty effect is reported in the generated critical drought magnitude compared with the generated storage capacity, except that similar pattern of increased statistics and quantiles by parameter uncertainty over different demand levels and sample serial correlations could be found for both design variables.

### 3.5 Summary and Conclusions

Parameter uncertainty effect on synthetic monthly streamflows has been examined in this paper using the simple AR(1) annual streamflows generation model and LAST and SPC temporal disaggregation models. Based on regression analysis, asymptotic distribution and more precise Bayesian posterior distribution of parameter estimates were theoretically derived and then parameter uncertainties were considered by sampling parameter estimates from derived distributions. Uncertainty effects were compared in terms of basic statistics as well as storage and drought related statistics which were calculated from generated monthly streamflows. Two historical flows with different statistical characteristics were implemented to investigate related effects. It is notable that even for large sample size ( $n=100$ ), parameter uncertainty could be found in increased variabilities of design variables.

LAST and SPC models show similar variabilities of generated storage and drought related statistics in terms of parameter uncertainty. The Bayesian approach shows larger variabilities of design variables for smaller sample size ( $n \leq 50$ ) than asymptotic distribution and the difference between two approaches is more visible for the St. Lawrence River case.

Parameter uncertainty in the disaggregation model does not affect greatly on the mean and standard deviation of generated monthly flows but on the month-to-month serial correlation and month-to-annual correlation. The effect of natural uncertainty in annual streamflow generation is more significant than parameter uncertainty in temporal disaggregation. This is closely associated with the model structure that the sum of

monthly flows would be equal to the annual and the proportionality of monthly flows would be preserved. Parameter uncertainty in annual flow generation is propagated into simulated monthly streamflows through the temporal disaggregation and parameter uncertainty causes significant effect on related storage and drought statistics.

Table 3.1: Basic statistics of annual and monthly streamflows (Lee's Ferry in Colorado River)

	Mean (ac-ft)	S. Dev. (ac-ft)	Coeff. Of Variation	Skewness Coeff.	Kurtosis Coeff.	lag-1 correlation	proportion to annual(%)
Historical							
Oct	580893	272006	0.47	1.64	6.64	0.54	3.9
Nov	480821	141531	0.29	1.21	4.83	0.76	3.2
Dec	382530	95858	0.25	1.22	4.92	0.83	2.5
Jan	356611	78632	0.22	0.59	3.06	0.70	2.4
Feb	393776	97576	0.25	1.42	5.74	0.55	2.6
Mar	645201	211390	0.33	1.08	4.81	0.48	4.3
Apr	1199946	512460	0.43	0.96	4.22	0.47	8.0
May	3037199	1146760	0.38	0.27	2.65	0.59	20.1
Jun	4054340	1572353	0.39	0.43	2.89	0.63	26.9
Jul	2190444	1012249	0.46	1.13	4.33	0.83	14.5
Aug	1083174	423971	0.39	0.95	3.20	0.78	7.2
Sep	671371	309698	0.46	1.95	8.21	0.64	4.5
Annual	15076307	4365301	0.29	0.14	2.40	0.28	100.0
Log-transformed series							
Oct	13.18	0.42	0.03	0.38	3.04	0.66	
Nov	13.04	0.28	0.02	0.20	3.89	0.78	
Dec	12.83	0.23	0.02	0.44	3.37	0.79	
Jan	12.76	0.22	0.02	0.04	2.77	0.69	
Feb	12.20	0.42	0.03	0.10	2.98	0.55	
Mar	13.33	0.31	0.02	0.09	2.94	0.55	
Apr	13.91	0.43	0.03	-0.11	2.53	0.57	
May	16.34	0.09	0.01	0.06	2.50	0.67	
Jun	16.20	0.14	0.01	0.07	2.68	0.64	
Jul	14.50	0.45	0.03	-0.06	2.74	0.87	
Aug	13.83	0.37	0.03	0.24	2.41	0.83	
Sep	13.33	0.39	0.03	0.54	3.31	0.66	
Annual	18.34	0.05	0.003	0.04	2.39	0.28	

Table 3.2: Basic statistics of annual and monthly streamflows (St. Lawrence River)

	Mean (ac-ft)	S. Dev. (ac-ft)	Coeff. Of Variation	Skewness Coeff.	Kurtosis Coeff.	lag-1 correlation	proportion to annual(%)
Oct	3997	357	0.09	-0.37	2.74	0.98	8.3
Nov	3790	345	0.09	-0.38	2.87	0.98	7.9
Dec	3871	350	0.09	-0.36	3.03	0.96	8.1
Jan	3765	380	0.10	-0.11	2.82	0.89	7.8
Feb	3352	348	0.10	-0.15	2.91	0.93	7.0
Mar	3884	397	0.10	0.15	2.59	0.91	8.1
Apr	4100	368	0.09	0.09	2.83	0.92	8.5
May	4376	405	0.09	0.07	2.53	0.96	9.1
Jun	4278	413	0.10	0.01	2.28	0.97	8.9
Jul	4382	420	0.10	-0.13	2.33	0.98	9.1
Aug	4261	403	0.09	-0.24	2.49	0.99	8.9
Sep	3984	370	0.09	-0.31	2.58	0.99	8.3
Annual	48042	4157	0.09	-0.28	2.74	0.75	100.0

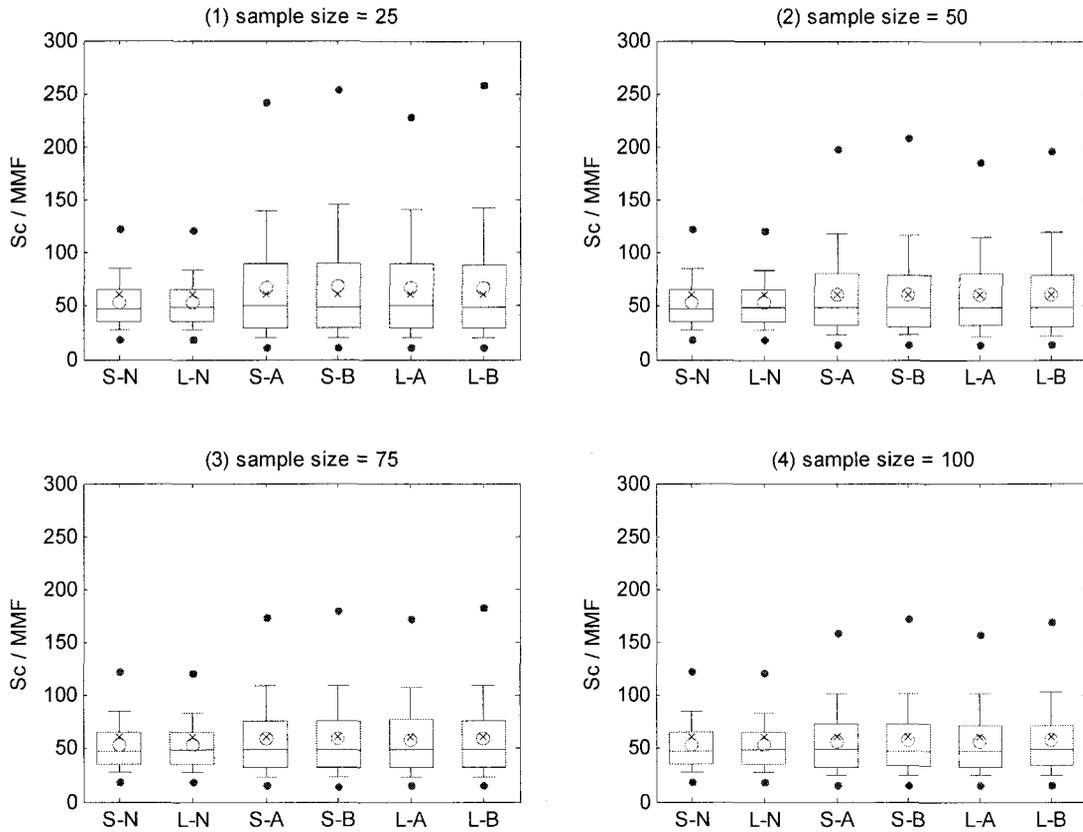


Figure 3.1: Distributions of storage capacities for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: no parameter uncertainty. Storage capacities are calculated with the demand level equal to MMF and the planning horizon of 98 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at Lee's Ferry in Colorado River. The upper, middle and lower line in the box means 75, 50, 25% quantile, respectively, and from the box the whisker extends to 90, 10% quantile for each side. Two dots outside box mean 99% and 1% quantile values. 'X' means historical storage capacity.

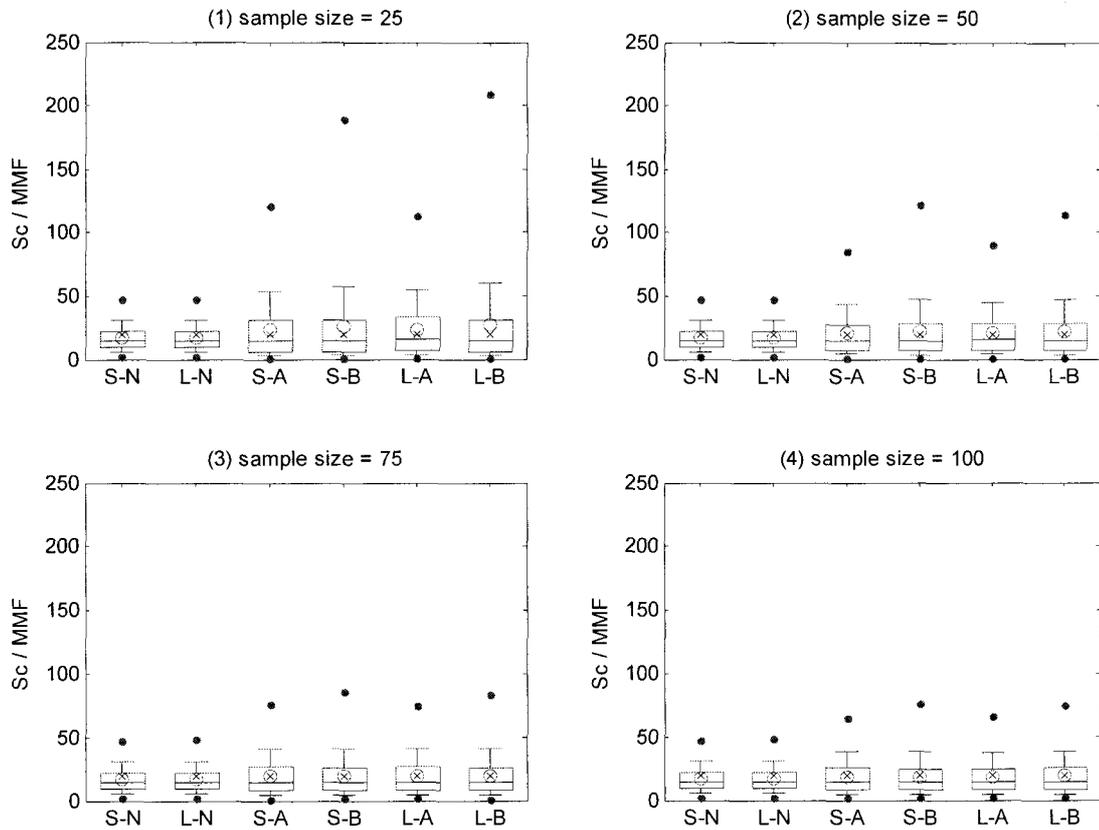


Figure 3.2: Distributions of storage capacities for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Storage capacities are calculated with the demand level equal to MMF and the planning horizon of 59 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at St. Lawrence River. The upper, middle and lower line in the box means 75, 50, 25% quantile, respectively, and from the box the whisker extends to 90, 10% quantile for each side. Two dots outside box mean 99% and 1% quantile values. 'X' means historical storage capacity.

Table 3.3: Increased quantiles(%) of generated storage capacities by parameter uncertainty in SPC disaggregation model: Storage capacities are calculated from synthetic monthly flows disaggregated with parameter uncertainty from historical annual flows for different sample sizes. (Bayesian approach)

Lee's Ferry in Colorado River (planning horizon = 25 years)								
	Demand level = 100% MMF				Demand level = 60% MMF			
	sample size				sample size			
	25	50	75	98	25	50	75	98
min	0.1	1.9	2.5	5.3	-8.1	-3.2	-4.9	0.9
1% quantile	-0.4	-1.1	-0.1	-0.1	-2.2	-0.8	-0.5	-0.7
10% quantile	-0.3	-0.6	-0.4	-0.7	0.0	0.1	-0.4	-0.1
25% quantile	0.2	-0.2	-0.2	-0.4	1.6	0.6	0.2	0.4
median	0.9	0.1	0.1	0.1	2.9	1.1	0.8	0.6
75% quantile	1.3	0.5	0.4	0.3	4.3	1.8	1.5	1.2
90% quantile	2.2	0.8	0.6	0.6	6.4	2.6	1.9	1.4
99% quantile	3.7	1.6	0.6	1.0	8.1	3.5	1.5	1.1
max	8.4	4.4	-0.5	2.9	12.4	12.4	2.3	9.0
mean	1.0	0.2	0.2	0.1	3.3	1.3	0.8	0.7
SD	15.0	8.4	4.7	6.2	28.9	11.8	8.0	7.5
Coeff. of variation	13.9	8.2	4.6	6.1	24.9	10.4	7.2	6.7

St. Lawrence River (planning horizon = 15 years)								
	Demand level = 95% MMF				Demand level = 85% MMF			
	sample size				sample size			
	25	50	75	98	25	50	75	98
min	1.1	8.6	1.8	-2.1	-34.5	-52.4	-52.4	-33.3
1% quantile	1.8	-0.8	0.5	1.8	-4.5	-0.1	-1.5	-1.1
10% quantile	1.8	0.5	-0.2	-0.6	4.3	0.1	1.1	1.0
25% quantile	2.6	1.3	0.6	-0.5	9.4	2.6	1.2	1.8
median	5.4	1.6	1.3	0.7	18.6	8.2	3.7	3.2
75% quantile	7.7	2.8	1.7	0.8	29.5	11.0	5.4	4.6
90% quantile	10.9	3.7	2.8	1.7	36.1	16.2	9.4	5.8
99% quantile	21.4	11.9	6.5	5.4	50.1	25.4	12.4	11.0
max	39.6	17.1	3.7	1.9	73.6	35.1	7.6	-7.1
mean	6.7	2.6	1.6	0.9	23.8	10.1	5.0	4.0
SD	28.6	13.6	7.6	5.5	59.3	29.7	14.2	8.8
Coeff. of variation	20.5	10.8	5.9	4.6	28.6	17.8	8.7	4.7

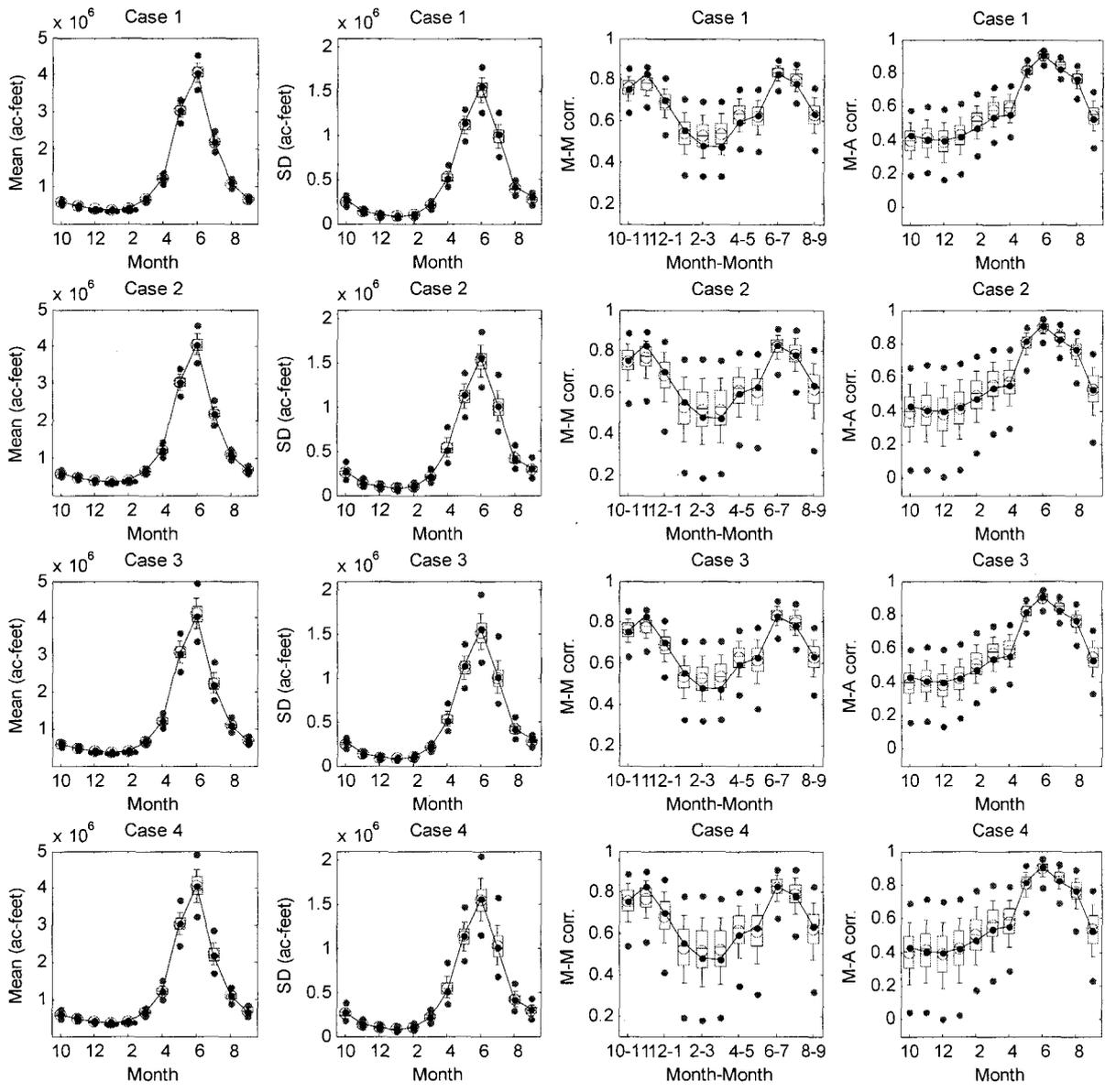


Figure 3.3: Basic statistics of monthly streamflows disaggregated using SPC temporal disaggregation model from annual flows generated using AR(1). Parameter uncertainty is incorporated using Bayesian posterior distribution and available sample size is assumed to be 50 (Lee's Ferry in Colorado River)

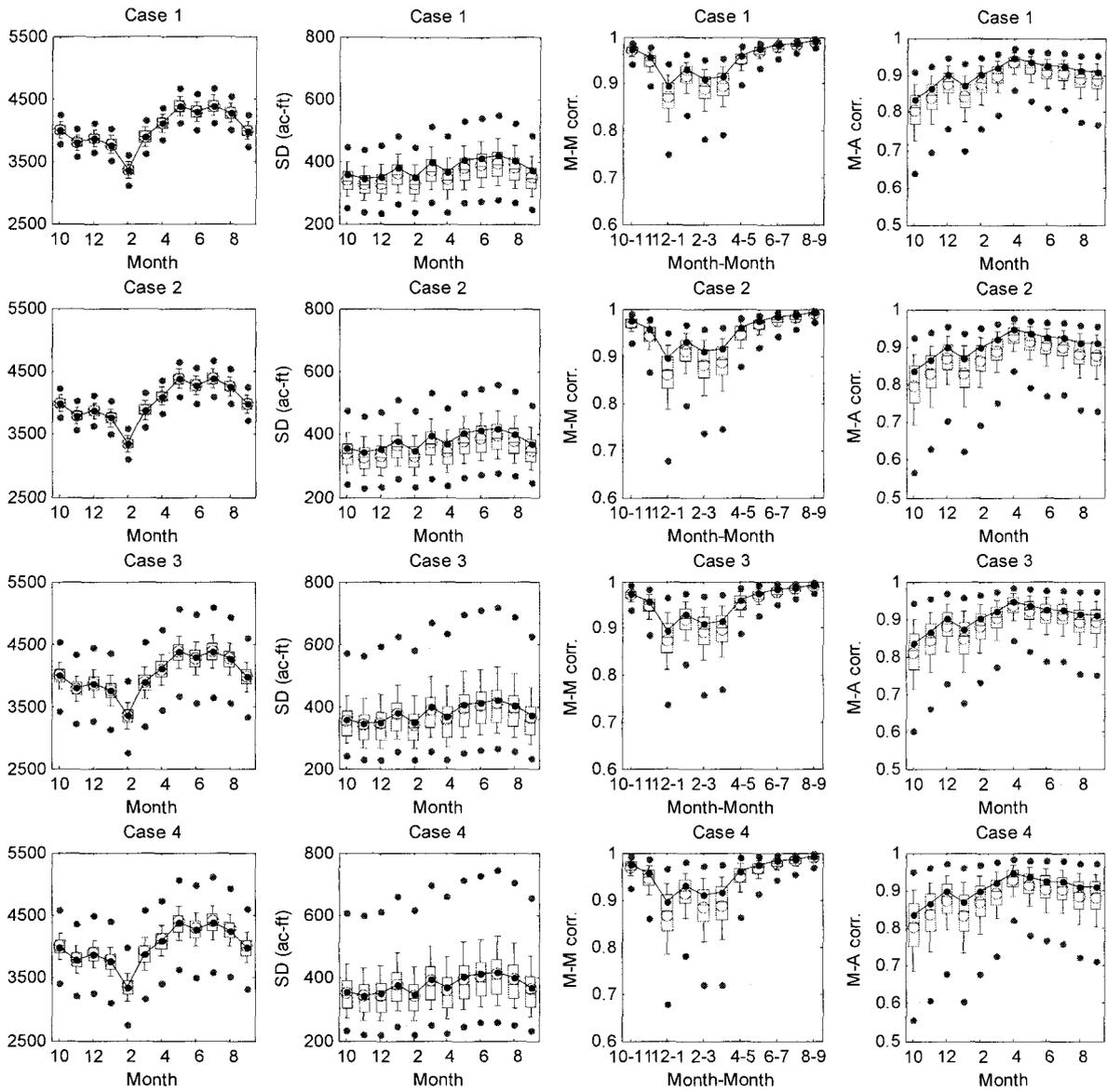


Figure 3.4: Basic statistics of monthly streamflows disaggregated using SPC temporal disaggregation model from annual flows generated using AR(1). Parameter uncertainty is incorporated using Bayesian posterior distribution and available sample size is assumed to be 50 (St. Lawrence River)

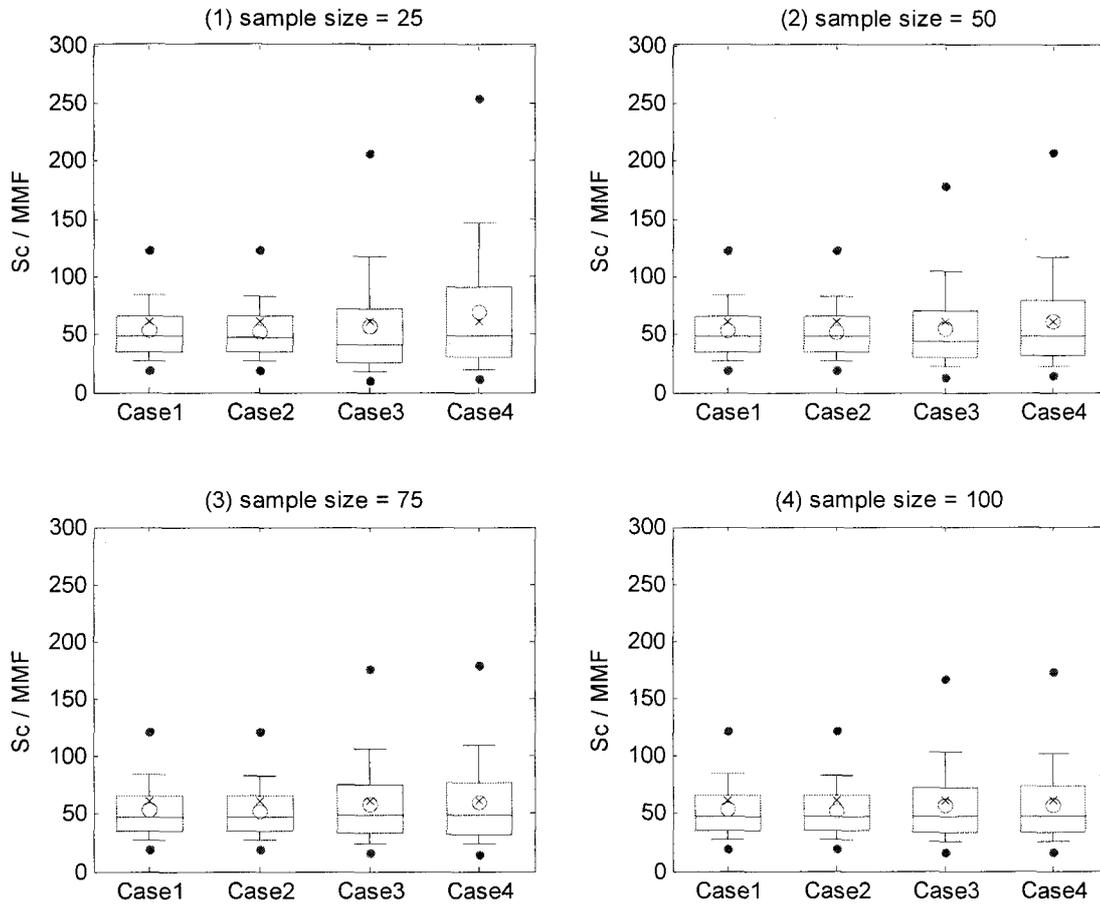


Figure 3.5: Distributions of storage capacities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River)

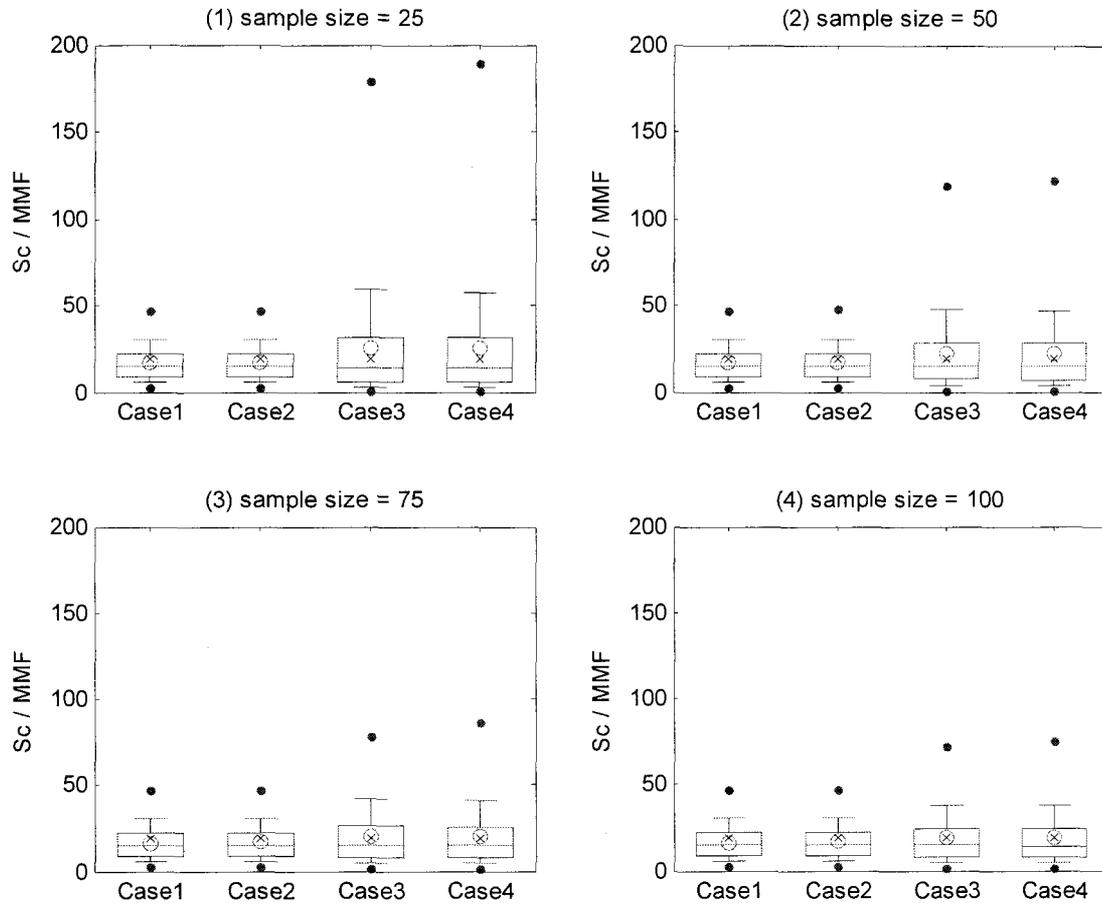


Figure 3.6: Distributions of storage capacities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, St. Lawrence River)

Table 3.4: Example of generated storage capacity and critical drought magnitude (scaled by MMF) ( $n=50$ , Bayesian analysis)

Demand level	Lee's Ferry in the Colorado River Basin ( $N_d=98$ )				St. Lawrence River ( $N_d=59$ )			
	MMF		0.8MMF		MMF		0.9MMF	
	Case 1	Case 4	Case 1	Case 4	Case 1	Case 4	Case 1	Case 4
Storage capacity								
mean	52.8	61.5	14.2	15.6	16.6	20.6	2.5	4.2
		16.4%		9.8%		23.7%		66.0%
SD	23.3	40.4	4.8	8.3	10.1	21.0	2.3	10.7
		73.0%		74.3%		106.9%		374.8%
$q_{0.9}$	84.7	118.3	20.5	25.5	30.6	43.9	5.4	9.1
		39.7%		24.6%		43.5%		70.4%
$q_{0.95}$	98.0	140.6	23.3	31.3	36.7	57.6	7.1	14.0
		43.5%		34.0%		56.9%		97.1%
$q_{0.99}$	123.1	197.4	29.3	46.4	46.9	84.4	10.8	33.0
		60.4%		58.4%		79.8%		204.4%
Critical drought magnitude								
mean	12.6	13.2	7.8	8.2	6.5	8.4	1.5	2.5
		4.7%		5.6%		30.5%		66.7%
SD	4.3	5.8	3.1	4.0	4.2	14.1	1.3	8.9
		35.1%		29.7%		234.0%		596.2%
$q_{0.9}$	16.5	20.7	11.4	12.5	12.1	16.9	2.9	4.6
		25.3%		9.1%		40.3%		56.6%
$q_{0.95}$	21.0	23.9	12.4	15.5	14.8	23.3	4.0	7.4
		13.4%		24.3%		57.1%		85.3%
$q_{0.99}$	24.3	32.3	17.0	21.4	20.1	50.8	6.5	18.0
		32.6%		25.6%		152.3%		178.6%

Note Case 1: no parameter uncertainty considered (natural uncertainty), Case 4: parameter uncertainty incorporated both in annual flow generation and in temporal disaggregation, SD: standard deviation,  $q_{0.9}$ ,  $q_{0.95}$ ,  $q_{0.99}$  means 90%, 95%, and 99% quantile, respectively. Value (%) in the column of Case 4 represents the ratio of increased storage capacity(or critical drought magnitude) in Case 4 with respect to Case 1. (parameter uncertainty effect)

## References

- Box, G.E.P. and G.C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc., N.Y.
- Bras, R.L. and I. Rodreigues-Iturbe (1985). *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass..
- Curry, K. D. and R.L. Bras (1978). *Theory and Applications of the Multivariate Broken Line, Disaggregation and Monthly Autoregressive Generators to the Nile River*, MIT Report 78-5, Cambridge, Mass.
- Grygier, J.C. and J.R. Stedinger (1988). Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research*, 24(10), pp. 1574-1584.
- Grygier, J.C. and J.R. Stedinger (1990). *SPIGOT, A synthetic Streamflow Generation Software Package, technical description*, version 2.5, School of Civil and Environmental Engineering, Cornell University, Ithaca, N. Y.
- Harms, A.A. and T.H. Campbell (1976). An extension to the Thomas-Fiering model for the sequential generation for streamflow, *Water Resources Research*, 3(3), pp. 653-661.
- Mejia, J.M. and J. Roussel (1976). Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), pp. 185-186.
- Lane, W.L. (1979). *Applied Stochastic Techniques, User's Manual*, Bureau of Reclamation, Engineering and Research Center, Denver, Co.
- Lane, W.L. (1982). Corrected parameter estimates for disaggregation schemes, in *Statistical Analysis of References R - 41 Rainfall and Runoff*, V. P. Singh, ed., Water Resources Publications, Littleton, Colo.
- Lane, W.L. and D.K. Frevert (1990). *Applied Stochastic Techniques: User's Manual*, personal computer version 5.2, Earth Sciences Division, Bureau of Reclamation, U.S. Department of the Interior, Denver, Colo.
- Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981). *Water Resources Planning and Analysis*, Prentice Hall, Englewood Cliffs, New Jersey.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.

- Salas, J.D., N. Saada, C.H. Chung, W.L. Lane, and D.K. Frevert (2000). Stochastic Analysis, Modeling, and Simulation (SAMS) version 2000 User's manual, *Technical report No. 10*, Computing Hydrology Laboratory, Water Resources, Hydrologic and Environmental Sciences, Engineering Research Center, Fort Collins, Colorado.
- Santos, E.G. and J.D. Salas (1994). Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, 118(5), pp. 765-784.
- Stedinger, J.R. and D. Pei (1982). An annual-monthly streamflows model for incorporating parameter uncertainty into reservoir simulation, *Time Series Methods in Hydroscience*, A.H. El-Shaarawi and S.R. Esterby, eds., Elsevier, New York, N.Y., pp 520-529.
- Stedinger, J.R. and R.M. Vogel (1984). Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(1), pp. 47-56.
- Stedinger, J.R., D. Pei, and T. Cohn (1985). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resources Research*, 21(5), pp. 665-675.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., New York.

### Appendix 3.A: Derivation of the asymptotic variance-covariance matrix of Lane's condensed temporal disaggregation scheme

Recall SPC temporal disaggregation scheme for each season  $\tau$  ( $\tau=2, \dots, \omega$ ) as

$$Y_{v,\tau} = \alpha_\tau + \beta_\tau Y_{v,\tau-1} + \gamma_\tau X_v + \delta_\tau \left[ \sum_{s=1}^{\tau-1} w_s Y_{v,s} \right] + V_{v,\tau} \quad (3.A1)$$

Assume that the seasonal random variable  $Y_{v,\tau}, Y_{v,\tau-1}$  for season  $\tau, \tau-1$  are normally distributed with the mean  $\mu_{Y_\tau}, \mu_{Y_{\tau-1}}$  and variance  $\sigma_{Y_\tau}^2, \sigma_{Y_{\tau-1}}^2$ , respectively and the annual random variable  $X_v$  is normally distributed with mean  $\mu_X$  and variance  $\sigma_X^2$ .

The seasonal cumulative term  $\sum_{s=1}^{\tau-1} w_s Y_{v,s} = Z_{v,\tau-1}$  is assumed as normally distributed with

mean  $\mu_{Z_{\tau-1}} = \sum_{s=1}^{\tau-1} w_s \mu_{Y_s}$  and variance  $\sigma_{Z_{\tau-1}}^2$ , and  $V_{v,\tau}$  is the error term with zero mean

and variance  $\sigma_\tau^2$ . Let  $\boldsymbol{\theta}_\tau = (\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau)$ ,  $\mathbf{V}_\tau = (V_{1,\tau}, V_{2,\tau}, \dots, V_{n,\tau})'$  and then from

(3.A1) the likelihood function  $L(\cdot) = L(\mathbf{V}_\tau; \boldsymbol{\theta}_\tau)$  and the resultant log-likelihood function

$LL(\cdot) = LL(\mathbf{V}_\tau; \boldsymbol{\theta}_\tau)$  are

$$L(\cdot) = \frac{1}{(2\pi)^{n/2} \sigma_\tau^n} \exp \left[ -\frac{1}{2\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})^2 \right]$$

$$LL(\cdot) = -\frac{n}{2} \log(2\pi) - n \log(\sigma_\tau) - \frac{1}{2\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})^2$$

1st derivative of log-likelihood  $LL(\cdot)$  with respect to each parameter is given as, respectively,

$$\frac{\partial LL(\cdot)}{\partial \alpha_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})$$

$$\frac{\partial LL(\cdot)}{\partial \beta_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) Y_{v,\tau-1}$$

$$\frac{\partial LL(\cdot)}{\partial \gamma_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) X_v$$

$$\frac{\partial LL(\cdot)}{\partial \delta_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) Z_{v,\tau-1}$$

$$\frac{\partial LL(\cdot)}{\partial \sigma_\tau} = -\frac{n}{\sigma_\tau} + \frac{1}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})^2$$

And 2<sup>nd</sup> derivatives with respect to parameters yield

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau^2} = -\frac{n}{\sigma_\tau^2}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau^2} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau^2} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \delta_\tau^2} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1}^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \sigma_\tau^2} = \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \beta_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \gamma_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \delta_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1})$$

$$\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \gamma_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v Y_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \delta_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1} Y_{v,\tau-1}$$

$$\frac{\partial LL(\cdot)}{\partial \beta_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) Y_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \delta_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1} X_v$$

$$\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) X_v$$

$$\frac{\partial^2 LL(\cdot)}{\partial \delta_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) Z_{v,\tau-1}$$

Taking expectation for negative 2<sup>nd</sup> derivatives over  $V_{v,\tau}$  using observations of

$Y_{v,\tau-1}$ ,  $X_v$ ,  $Z_{v,\tau}$  yields

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau^2}\right) = \frac{n}{\sigma_\tau^2}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau^2}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}^2\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Y_{v,\tau-1}^2) = \frac{n}{\sigma_\tau^2} [\hat{\sigma}_{Y_t}^2 + \hat{\mu}_{Y_{t-1}}^2]$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau^2}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v^2\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_v^2) = \frac{n}{\sigma_\tau^2} [\hat{\sigma}_X^2 + \hat{\mu}_X^2]$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \delta_\tau^2}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1}^2\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Z_{v,\tau-1}^2) = \frac{n}{\sigma_\tau^2} [\hat{\sigma}_{Z_{\tau-1}}^2 + \hat{\mu}_{Z_{\tau-1}}^2]$$

$$\begin{aligned} E\left(-\frac{\partial^2 LL(\cdot)}{\partial \sigma_\tau^2}\right) &= E\left(\frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v - \delta Z_{v,\tau-1})^2\right) \\ &= \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n E(\varepsilon_{v,\tau}^2) = \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} n\sigma_\tau^2 = \frac{2n}{\sigma_\tau^2} \end{aligned}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \beta_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Y_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} \hat{\mu}_{Y_{\tau-1}}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \gamma_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_v) = \frac{n}{\sigma_\tau^2} \hat{\mu}_X$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \delta_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Z_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} \hat{\mu}_Z$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \gamma_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v Y_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_v Y_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} (\hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}})$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \delta_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1} Y_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Z_{v,\tau-1} Y_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} (\hat{\sigma}_{Z_{\tau-1}Y_{\tau-1}} + \hat{\mu}_{Z_{\tau-1}} \hat{\mu}_{Y_{\tau-1}})$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \delta_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Z_{v,\tau-1} X_v\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Z_{v,\tau-1} X_v) = \frac{n}{\sigma_\tau^2} (\hat{\sigma}_{Z_{\tau-1}X} + \hat{\mu}_{Z_{\tau-1}} \hat{\mu}_X)$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v - \delta Z_{v,\tau-1})\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau}) = 0$$

$$E\left(-\frac{\partial LL(\cdot)}{\partial \beta_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v - \delta Z_{v,\tau-1}) Y_{v,\tau-1}\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau} Y_{v,\tau-1}) = 0$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v - \delta Z_{v,\tau-1}) X_v\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau} X_v) = 0$$

$$\begin{aligned}
E\left(-\frac{\partial^2 LL(\cdot)}{\partial \delta_\tau \partial \sigma_\tau}\right) &= E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v - \delta_\tau Z_{v,\tau-1}) Z_{v,\tau-1}\right) \\
&= \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau} Z_{v,\tau-1}) = 0
\end{aligned}$$

Hence, asymptotic variance-covariance matrix of ML estimates of parameter sets

$V(\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau, \sigma_\tau)$  can be given by

$$V(\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau, \sigma_\tau) = I(\alpha_\tau, \beta_\tau, \gamma_\tau, \delta_\tau, \sigma_\tau)^{-1}$$

$$\begin{aligned}
&= \begin{bmatrix} E\left(-\frac{\partial^2 L}{\partial \alpha_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \beta_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \gamma_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \delta_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \sigma_\tau}\right) \\ E\left(-\frac{\partial^2 L}{\partial \beta_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \beta_\tau \partial \gamma_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \beta_\tau \partial \delta_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \beta_\tau \partial \sigma_\tau}\right) & \\ E\left(-\frac{\partial^2 L}{\partial \gamma_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \gamma_\tau \partial \delta_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \gamma_\tau \partial \sigma_\tau}\right) & & \\ \text{symm.} & & E\left(-\frac{\partial^2 L}{\partial \delta_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \delta_\tau \partial \sigma_\tau}\right) & \\ & & & E\left(-\frac{\partial^2 L}{\partial \sigma_\tau^2}\right) & \end{bmatrix}^{-1} \\
&= \frac{\sigma_\tau^2}{n} \begin{bmatrix} 1 & \hat{\mu}_{Y_{\tau-1}} & \hat{\mu}_X & \hat{\mu}_{Z_{\tau-1}} & 0 \\ \hat{\sigma}_{Y_{\tau-1}}^2 + \hat{\mu}_{Y_{\tau-1}}^2 & \hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}} & \hat{\sigma}_{Z_{\tau-1}Y_{\tau-1}} + \hat{\mu}_{Z_{\tau-1}} \hat{\mu}_{Y_{\tau-1}} & 0 & 0 \\ & \hat{\sigma}_X^2 + \hat{\mu}_X^2 & \hat{\sigma}_{XZ_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Z_{\tau-1}} & 0 & 0 \\ \text{symm.} & & \hat{\sigma}_{Z_{\tau-1}}^2 + \hat{\mu}_{Z_{\tau-1}}^2 & 0 & 0 \\ & & & & 2 \end{bmatrix}^{-1} \tag{3.A2}
\end{aligned}$$

### Appendix 3.B: Derivation of the asymptotic variance-covariance matrix of Lane's condensed temporal disaggregation scheme

Lane's condensed temporal disaggregation scheme can be expressed as a regression form for each season  $\tau$  as

$$Y_{\nu,\tau} = \alpha_{\tau} + \beta_{\tau} Y_{\nu,\tau-1} + \gamma_{\tau} X_{\nu} + V_{\nu,\tau} \quad (3.B1)$$

where,  $\nu$  is the annual index,  $\tau$  is seasonal one ( $\tau = 1, \dots, \omega$ ), and  $\omega$  is the number of seasons. Since Eq. (3.B1) is defined for each season, there are  $\omega$  equations can be considered among total  $\omega$  seasons.  $X_{\nu}$  denotes annual observation series in year  $\nu$  with sample mean  $\hat{\mu}_X$  and variance  $\hat{\sigma}_X^2$ ,  $Y_{\nu,\tau-1}$  is monthly observation series in a season  $\tau-1$  with sample mean  $\hat{\mu}_{Y_{\tau-1}}$  and variance  $\hat{\sigma}_{Y_{\tau-1}}^2$ , and  $V_{\nu,\tau}$  is the error term with zero mean and variance  $\sigma_{\tau}^2$ . Since  $V_{\nu,\tau}$  is normally distributed, monthly streamflows in year  $\nu$  and season  $\tau$   $Y_{\nu,\tau}$  follows a normal distribution with mean  $\alpha_{\tau} X_{\nu} + \beta_{\tau} Y_{\nu,\tau-1} + \gamma_{\tau} X_{\nu}$  and variance  $\sigma_{\tau}^2$ .

Let  $\boldsymbol{\theta}_{\tau} = (\alpha_{\tau}, \beta_{\tau}, \gamma_{\tau})$ ,  $\mathbf{Y}_{\tau} = (Y_{1,\tau}, Y_{2,\tau}, \dots, Y_{n,\tau})'$  and then from Eq.(3.B1) the likelihood function  $L(\cdot) = L(\mathbf{Y}_{\tau}; \boldsymbol{\theta}_{\tau})$  and the resultant log-likelihood function  $LL(\cdot) = LL(\mathbf{Y}_{\tau}; \boldsymbol{\theta}_{\tau})$  are

$$L(\cdot) = \frac{1}{(2\pi)^{n/2} \sigma_{\tau}^n} \exp \left[ -\frac{1}{2\sigma_{\tau}^2} \sum_{\nu=1}^n (Y_{\nu,\tau} - \alpha_{\tau} - \beta_{\tau} Y_{\nu,\tau-1} - \gamma_{\tau} X_{\nu})^2 \right]$$

$$LL(\cdot) = -\frac{n}{2} \log(2\pi) - n \log(\sigma_{\tau}) - \frac{1}{2\sigma_{\tau}^2} \sum_{\nu=1}^n (Y_{\nu,\tau} - \alpha_{\tau} - \beta_{\tau} Y_{\nu,\tau-1} - \gamma_{\tau} X_{\nu})^2$$

1st derivative of log-likelihood  $LL(\cdot)$  with respect to each parameter is given as, respectively,

$$\frac{\partial LL(\cdot)}{\partial \alpha_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v)$$

$$\frac{\partial LL(\cdot)}{\partial \beta_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v) Y_{v,\tau-1}$$

$$\frac{\partial LL(\cdot)}{\partial \gamma_\tau} = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v) X_v$$

$$\frac{\partial LL(\cdot)}{\partial \sigma_\tau} = -\frac{n}{\sigma_\tau} + \frac{1}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v)^2$$

And 2<sup>nd</sup> derivatives with respect to parameters yield

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau^2} = -\frac{n}{\sigma_\tau^2}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau^2} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau^2} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \sigma_\tau^2} = \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v)^2$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \beta_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \gamma_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v$$

$$\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v)$$

$$\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \gamma_\tau} = -\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v Y_{v,\tau-1}$$

$$\frac{\partial LL(\cdot)}{\partial \beta_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v) Y_{v,\tau-1}$$

$$\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \sigma_\tau} = -\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha_\tau - \beta_\tau Y_{v,\tau-1} - \gamma_\tau X_v) X_v$$

Taking expectation for negative 2<sup>nd</sup> derivatives over  $Y_{v,\tau}$  using observations yields

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau^2}\right) = \frac{n}{\sigma_\tau^2}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau^2}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}^2\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Y_{v,\tau-1}^2) = \frac{n}{\sigma_\tau^2} [\hat{\sigma}_{Y_\tau}^2 + \hat{\mu}_{Y_{\tau-1}}^2]$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau^2}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_\tau^2\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_\tau^2) = \frac{n}{\sigma_\tau^2} [\hat{\sigma}_X^2 + \hat{\mu}_X^2]$$

$$\begin{aligned} E\left(-\frac{\partial^2 LL(\cdot)}{\partial \sigma_\tau^2}\right) &= E\left(\frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v)^2\right) \\ &= \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} \sum_{v=1}^n E(\varepsilon_{v,\tau}^2) = \frac{n}{\sigma_\tau^2} - \frac{3}{\sigma_\tau^4} n \sigma_\tau^2 = \frac{2n}{\sigma_\tau^2} \end{aligned}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \beta_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n Y_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(Y_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} \hat{\mu}_{Y_{\tau-1}}$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \gamma_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_v) = \frac{n}{\sigma_\tau^2} \hat{\mu}_X$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \gamma_\tau}\right) = E\left(\frac{1}{\sigma_\tau^2} \sum_{v=1}^n X_v Y_{v,\tau-1}\right) = \frac{1}{\sigma_\tau^2} \sum_{v=1}^n E(X_v Y_{v,\tau-1}) = \frac{n}{\sigma_\tau^2} (\hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}})$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \alpha_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v)\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau}) = 0$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \beta_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v) Y_{v,\tau-1}\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau} Y_{v,\tau-1}) = 0$$

$$E\left(-\frac{\partial^2 LL(\cdot)}{\partial \gamma_\tau \partial \sigma_\tau}\right) = E\left(\frac{2}{\sigma_\tau^3} \sum_{v=1}^n (Y_{v,\tau} - \alpha - \beta Y_{v,\tau-1} - \gamma X_v) X_v\right) = \frac{2}{\sigma_\tau^3} \sum_{v=1}^n E(V_{v,\tau} X_v) = 0$$

Hence, asymptotic variance-covariance matrix of ML estimates of parameter sets

$V(\alpha_\tau, \beta_\tau, \gamma_\tau, \sigma_\tau)$  can be given by

$$V(\alpha_\tau, \beta_\tau, \gamma_\tau, \sigma_\tau) = I(\alpha_\tau, \beta_\tau, \gamma_\tau, \sigma_\tau)^{-1}$$

$$= \begin{bmatrix} E\left(-\frac{\partial^2 L}{\partial \alpha_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \beta_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \gamma_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \alpha_\tau \partial \sigma_\tau}\right) \\ & E\left(-\frac{\partial^2 L}{\partial \beta_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \beta_\tau \partial \gamma_\tau}\right) & E\left(-\frac{\partial^2 L}{\partial \beta_\tau \partial \sigma_\tau}\right) \\ & & E\left(-\frac{\partial^2 L}{\partial \gamma_\tau^2}\right) & E\left(-\frac{\partial^2 L}{\partial \gamma_\tau \partial \sigma_\tau}\right) \\ & & & E\left(-\frac{\partial^2 L}{\partial \sigma_\tau^2}\right) \end{bmatrix}^{-1}$$

*symm.*

$$= \frac{\sigma_\tau^2}{n} \begin{bmatrix} 1 & \hat{\mu}_{Y_{\tau-1}} & \hat{\mu}_X & 0 \\ \hat{\sigma}_{Y_{\tau-1}}^2 + \hat{\mu}_{Y_{\tau-1}}^2 & \hat{\sigma}_{XY_{\tau-1}} + \hat{\mu}_X \hat{\mu}_{Y_{\tau-1}} & 0 & 0 \\ \hat{\sigma}_X^2 + \hat{\mu}_X^2 & & 0 & 0 \\ \text{symm.} & & & 2 \end{bmatrix}^{-1} \quad (3.B2)$$

### Appendix 3.C: Additional Figures and Tables

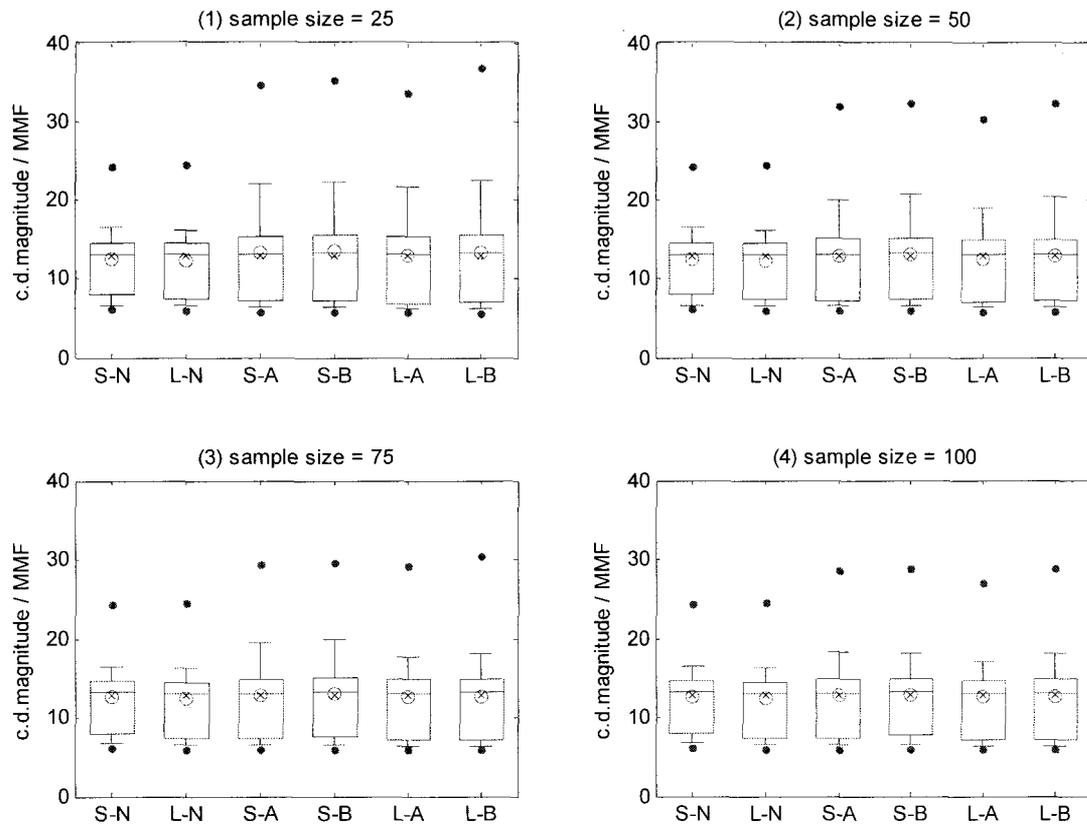


Figure 3.C1: Distributions of critical drought magnitudes for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought magnitudes are calculated with the demand level equal to MMF and the planning horizon of 98 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at Lee's Ferry in Colorado River. The upper, middle and lower line in the box means 75, 50, 25% quantile, respectively, and from the box the whisker extends to 90, 10% quantile for each side. Two dots outside box mean 99% and 1% quantile values.

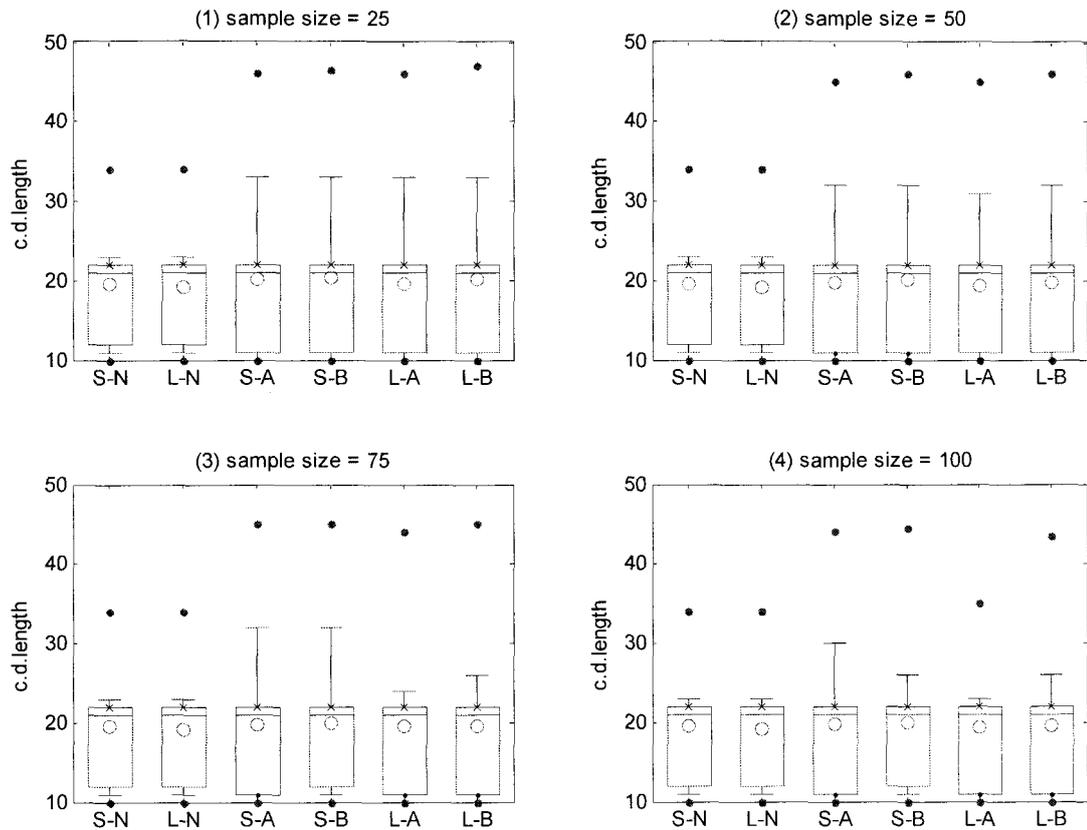


Figure 3.C2: Distributions of critical drought lengths for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought lengths are calculated with the demand level equal to MMF and the planning horizon of 98 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at Lee's Ferry in Colorado River.

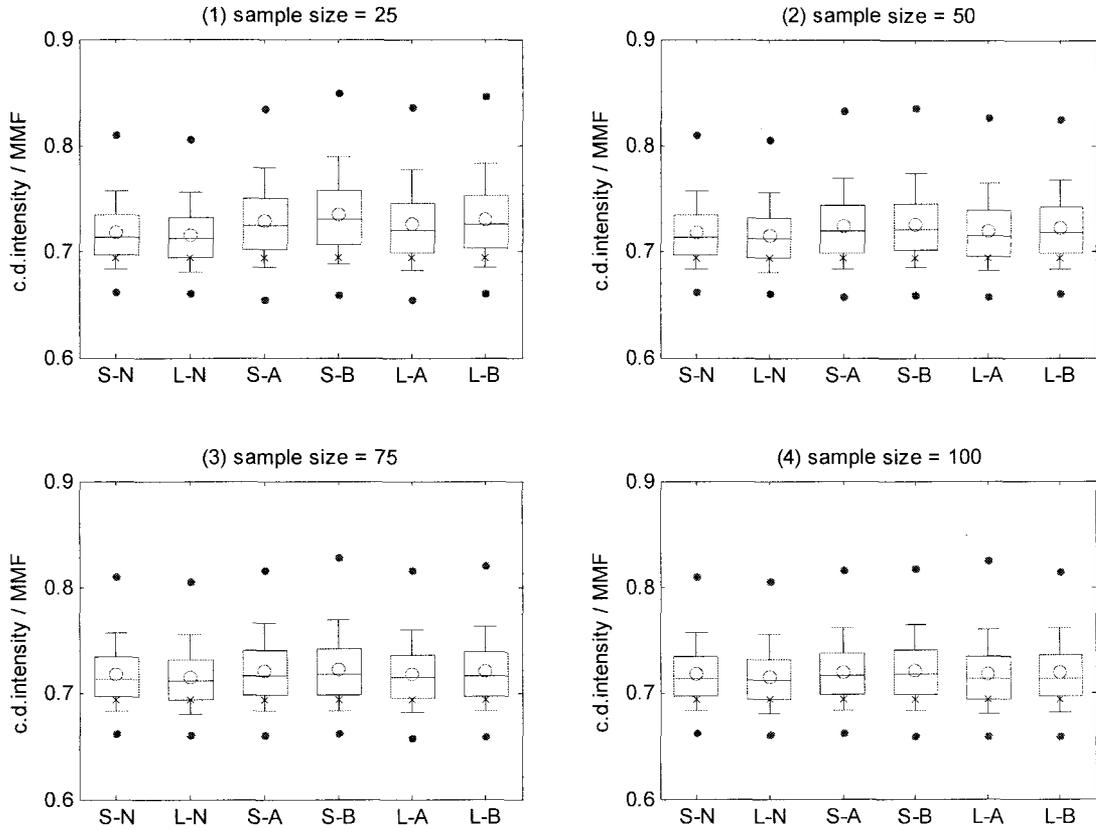


Figure 3.C3: Distributions of critical drought intensities for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought intensities are calculated with the demand level equal to MMF and the planning horizon of 98 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at Lee's Ferry in Colorado River.

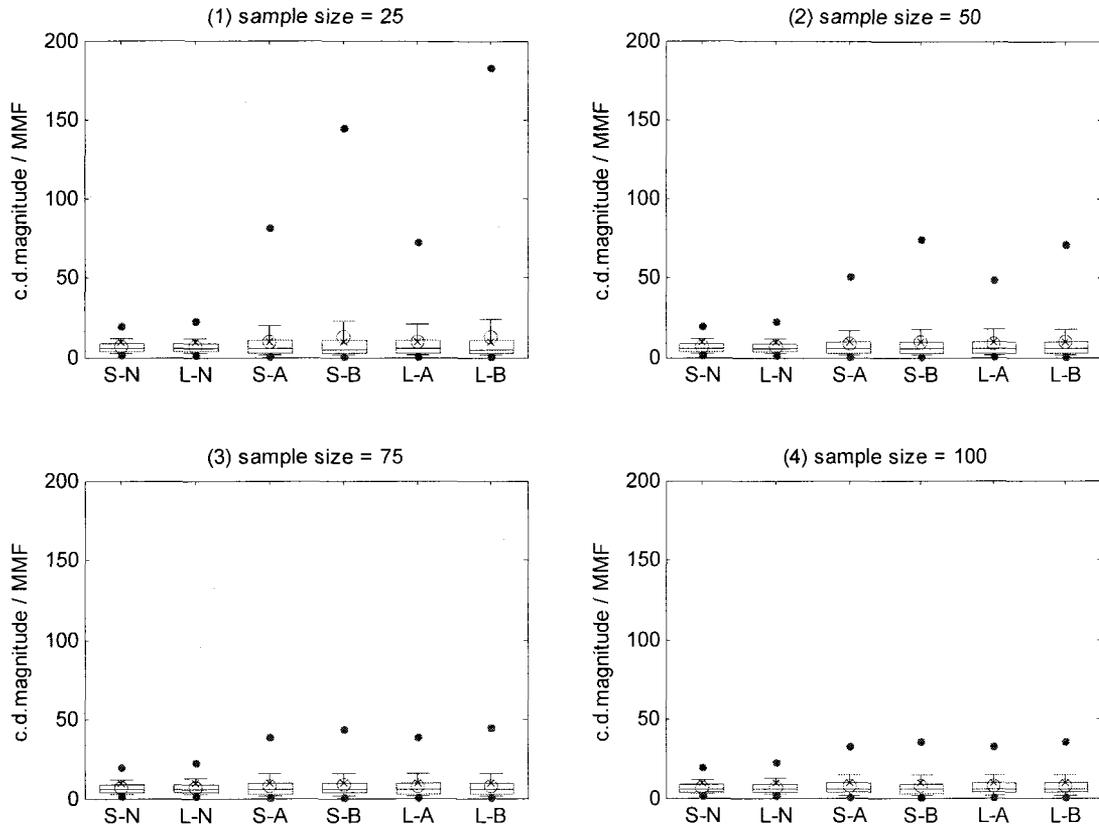


Figure 3.C4: Distributions of critical drought magnitudes for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought magnitudes are calculated with the demand level equal to MMF and the planning horizon of 59 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at St. Lawrence River.

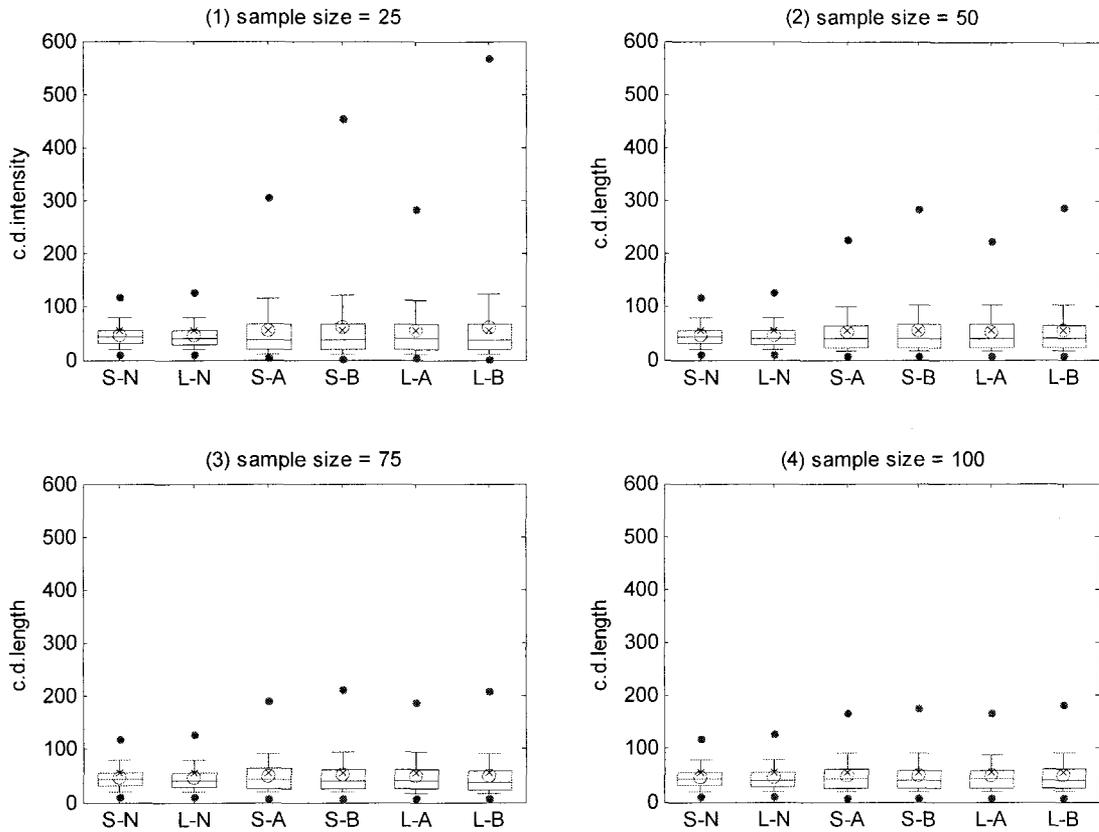


Figure 3.C5: Distributions of critical drought lengths for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought lengths are calculated with the demand level equal to MMF and the planning horizon of 59 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at St. Lawrence River.

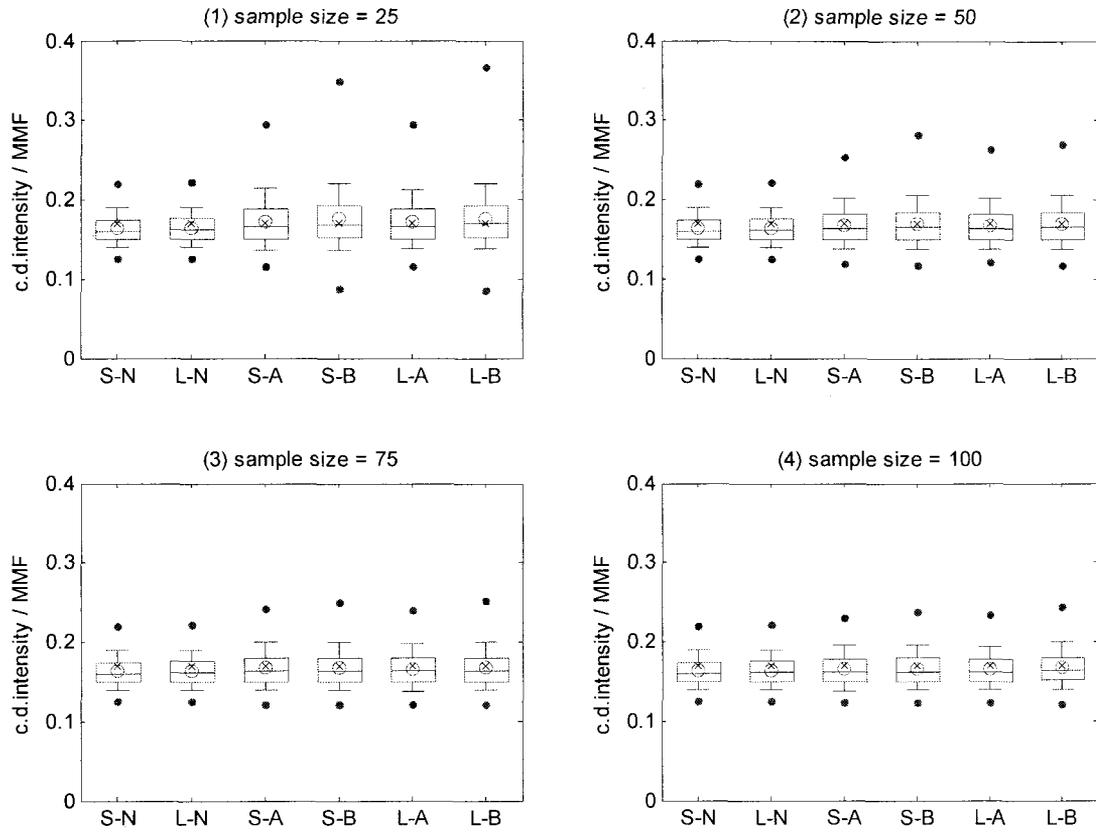


Figure 3.C6: Distributions of critical drought intensities for different disaggregation models (SPC and LAST) and different parameter uncertainty consideration schemes in the temporal disaggregation (Asymptotic and Bayesian approaches) where S: SPC, L: LAST, A: Asymptotic approach, B: Bayesian approach, and N: natural uncertainty. Critical drought intensities are calculated with the demand level equal to MMF and the planning horizon of 59 years based on synthetic monthly streamflows disaggregated from simulated annual flows with parameter uncertainty incorporated as well using historical flows at St. Lawrence River.

Table 3.C1: Coefficient of variations of simulated reservoir statistics (Lee's Ferry in Colorado River, demand level = MMF)

sample size	S-N	L-N	S-A	S-B	L-A	L-B
Storage capacity						
25	0.44	0.43	0.76	0.80	0.75	0.82
50	0.44	0.43	0.66	0.68	0.64	0.67
75	0.44	0.43	0.61	0.62	0.60	0.62
98	0.44	0.43	0.56	0.57	0.57	0.58
Critical drought magnitude						
25	0.34	0.35	0.48	0.49	0.49	0.51
50	0.34	0.35	0.43	0.44	0.43	0.45
75	0.34	0.35	0.41	0.40	0.41	0.41
98	0.34	0.35	0.39	0.39	0.39	0.41
Critical drought length						
25	0.32	0.32	0.43	0.43	0.43	0.45
50	0.32	0.32	0.39	0.39	0.38	0.40
75	0.32	0.32	0.37	0.36	0.36	0.37
98	0.32	0.32	0.36	0.35	0.35	0.36
Critical drought intensity						
25	0.04	0.04	0.05	0.06	0.05	0.05
50	0.04	0.04	0.05	0.05	0.05	0.05
75	0.04	0.04	0.05	0.05	0.05	0.05
98	0.04	0.04	0.04	0.05	0.05	0.05

Table 3.C2: Coefficient of variations of simulated reservoir statistics (St. Lawrence River, demand level = MMF)

sample size	S-N	L-N	S-A	S-B	L-A	L-B
Storage capacity						
15	0.61	0.62	1.12	1.43	1.07	1.48
30	0.61	0.62	1.02	1.11	1.04	1.16
45	0.61	0.62	0.88	0.94	0.83	0.97
59	0.61	0.62	0.77	0.87	0.78	0.83
Critical drought magnitude						
15	0.65	0.69	1.75	2.44	1.67	2.57
30	0.65	0.69	1.67	1.75	1.79	1.96
45	0.65	0.69	1.26	1.43	1.04	1.55
59	0.65	0.69	0.92	1.13	0.96	1.06
Critical drought length						
15	0.52	0.54	1.06	1.35	1.03	1.41
30	0.52	0.54	0.91	1.03	0.93	1.06
45	0.52	0.54	0.80	0.87	0.75	0.86
59	0.52	0.54	0.68	0.74	0.69	0.75
Critical drought intensity						
15	0.12	0.12	0.21	0.26	0.21	0.27
30	0.12	0.12	0.18	0.19	0.18	0.20
45	0.12	0.12	0.15	0.16	0.15	0.16
59	0.12	0.12	0.14	0.15	0.14	0.15

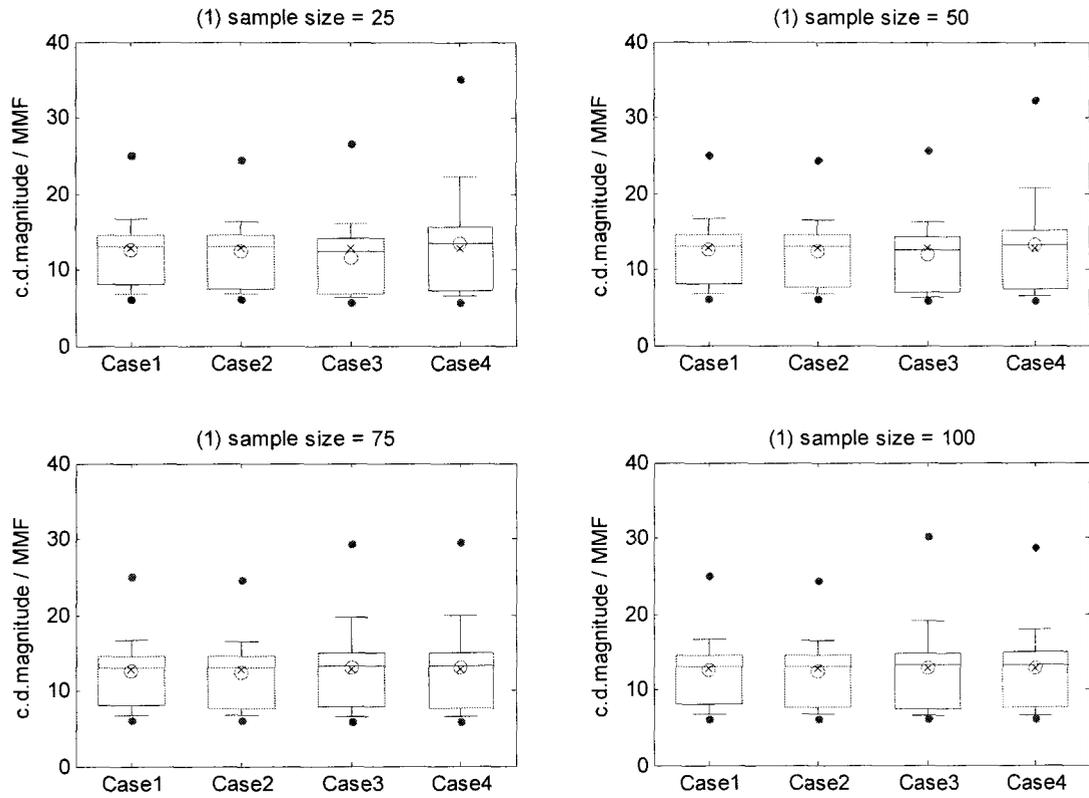


Figure 3.C7: Distributions of critical drought magnitudes for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River)

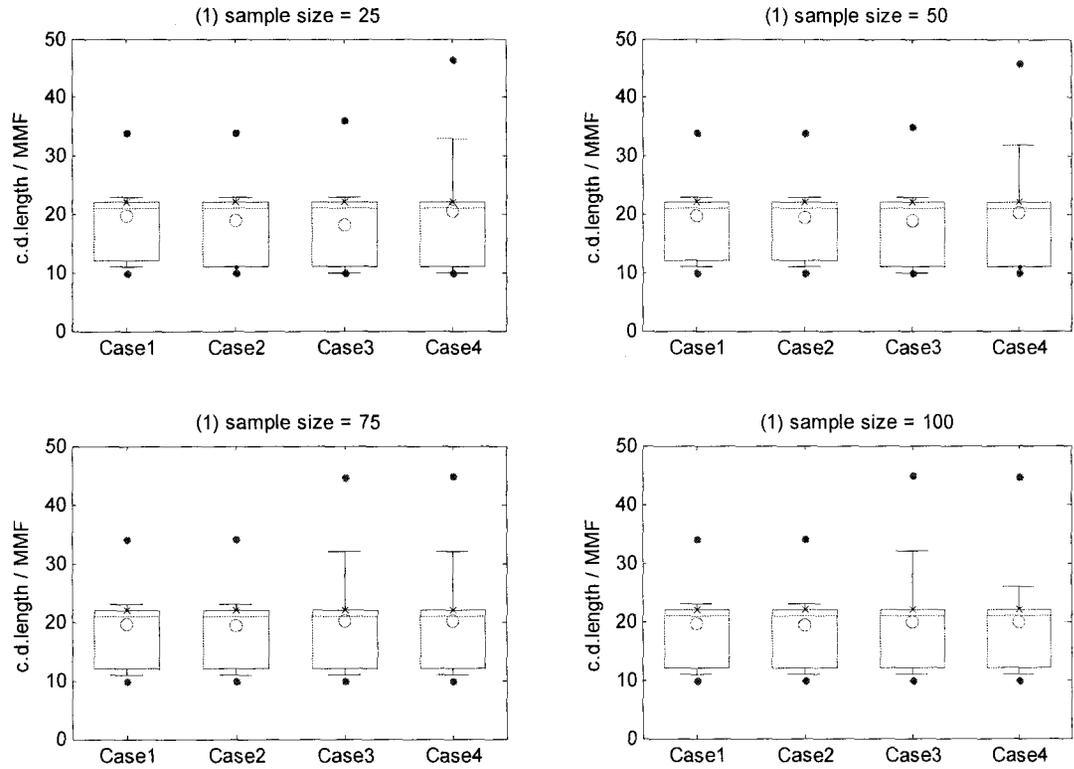


Figure 3.C8: Distributions of critical drought lengths for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River)

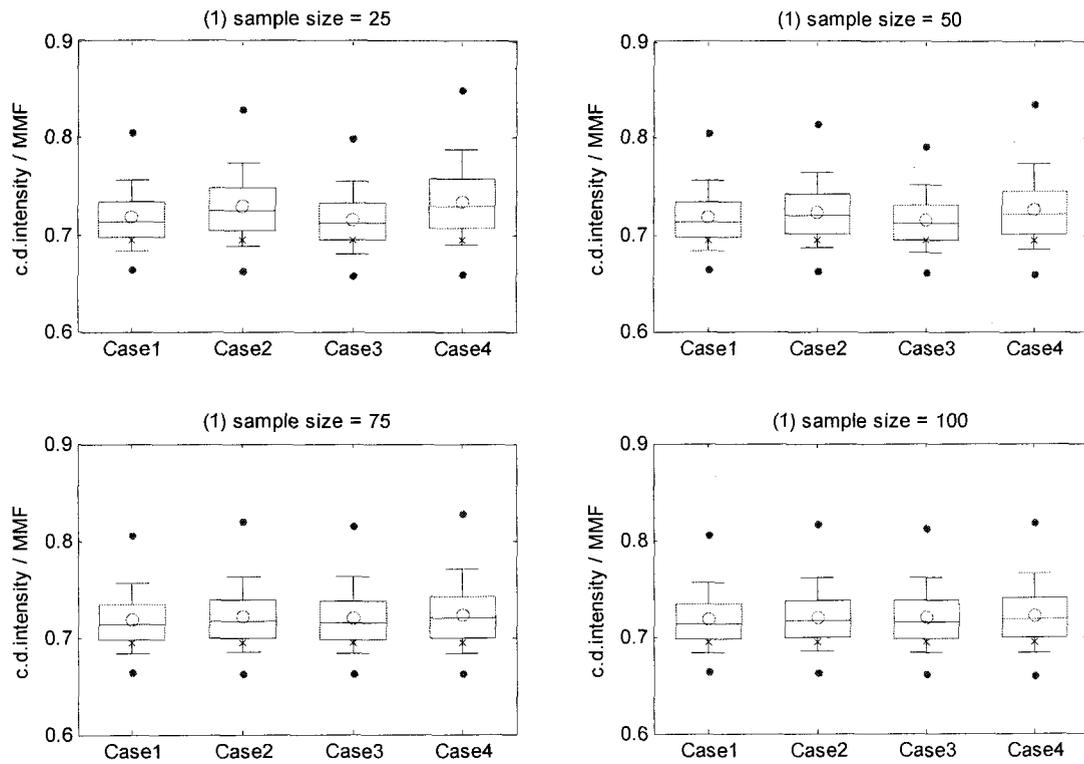


Figure 3.C9: Distributions of critical drought intensities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, St. Lawrence River)

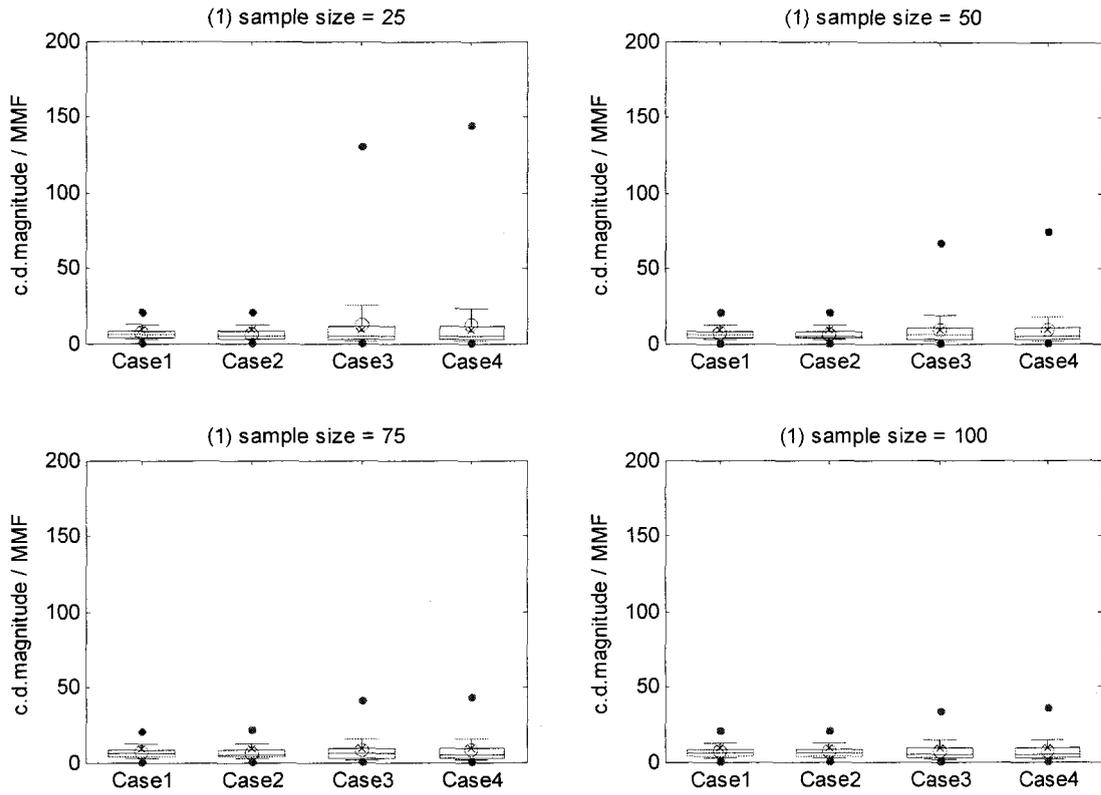


Figure 3.C10: Distributions of critical drought magnitudes for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 98 years (SPC, Bayesian approaches, St. Lawrence River)

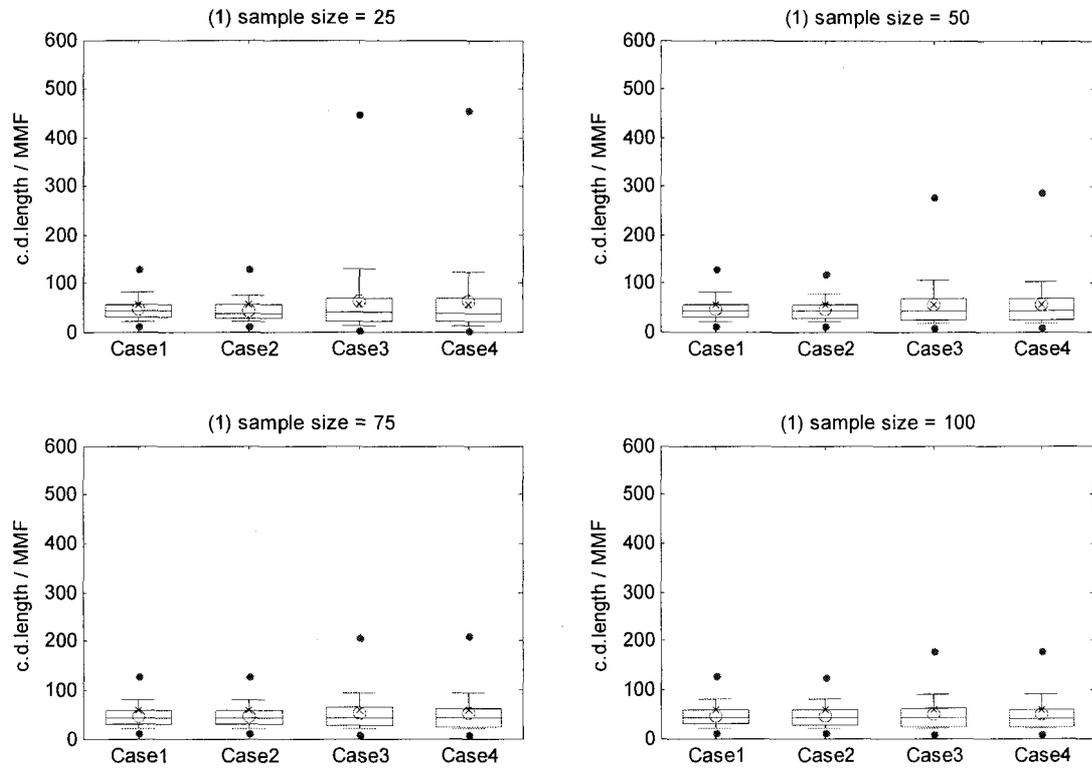


Figure 3.C11: Distributions of critical drought lengths for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River)

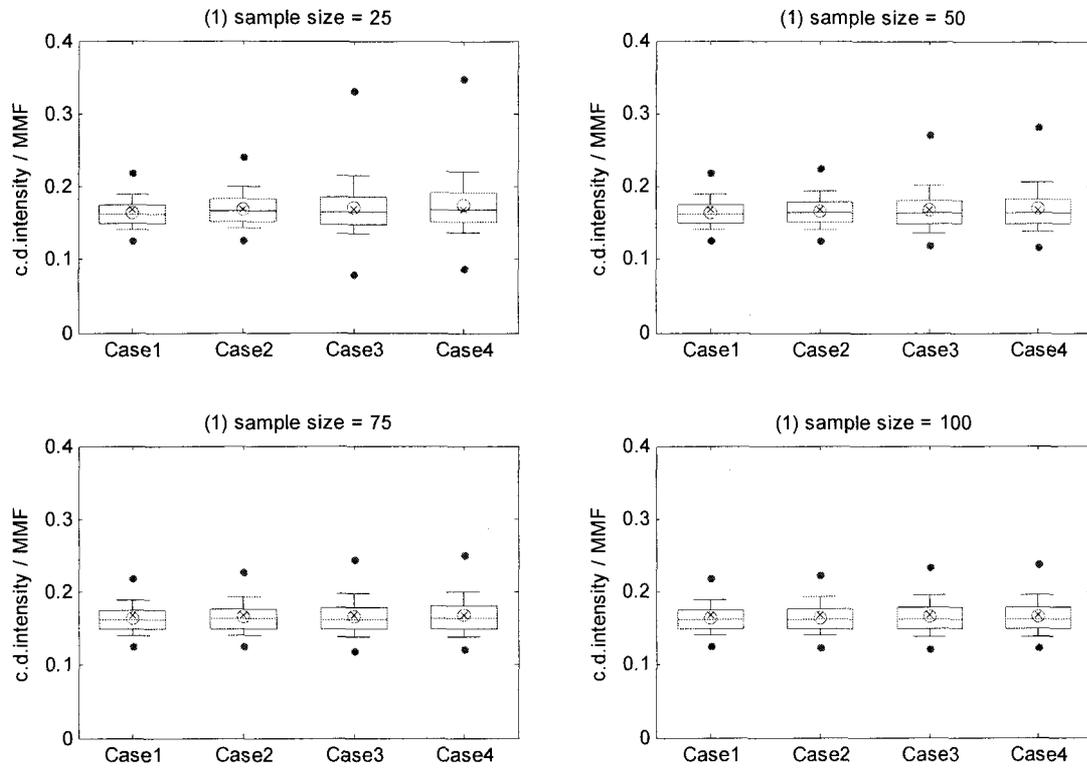


Figure 3.C12: Distributions of critical drought intensities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the demand level of MMF and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River)

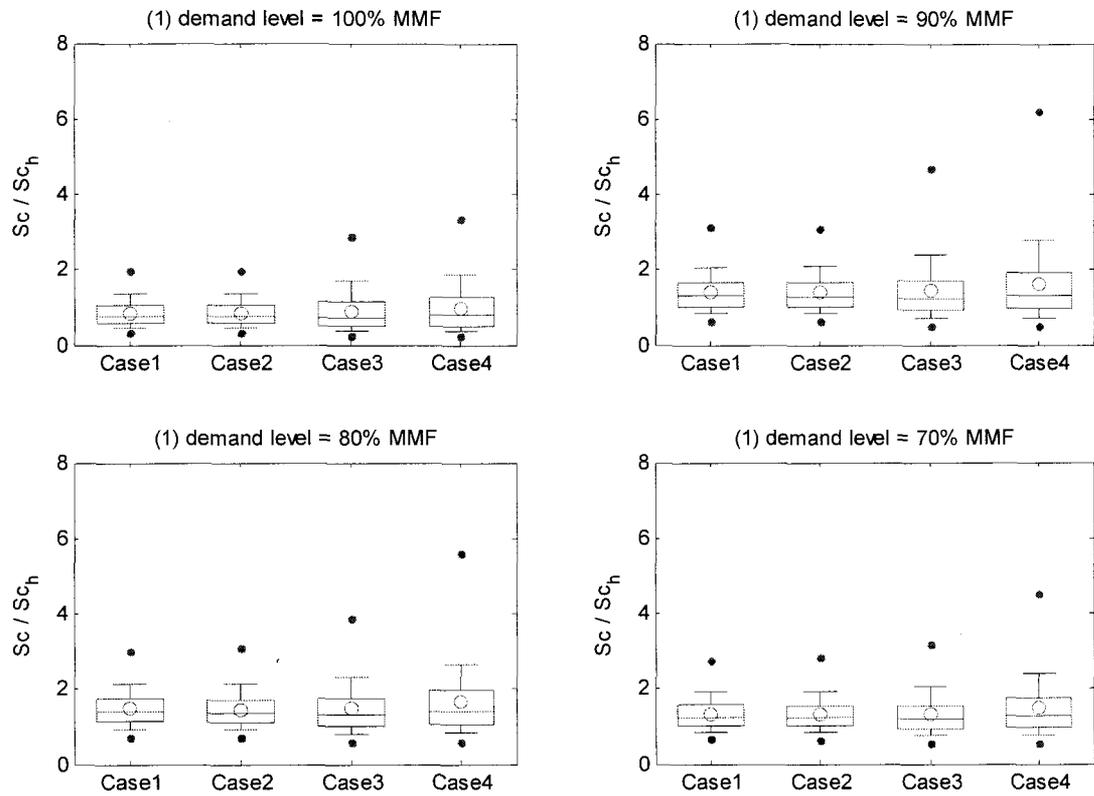


Figure 3.C13: Distributions of storage capacities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River). Subscript "h" means the historical.

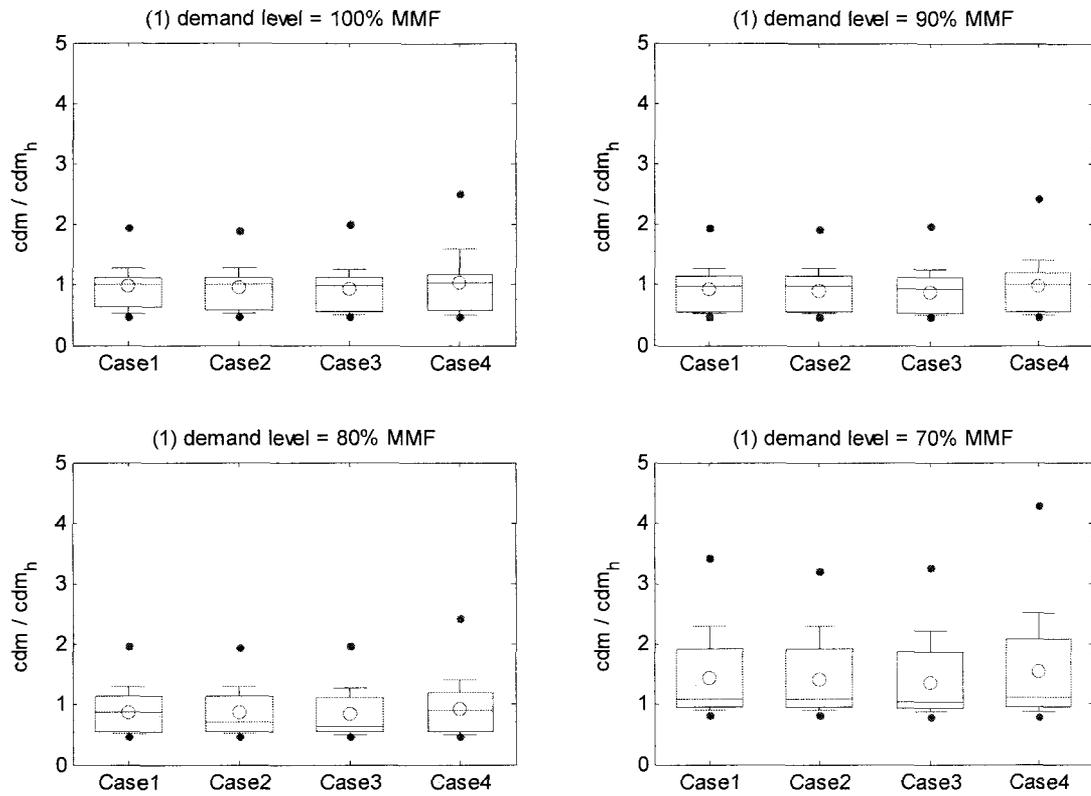


Figure 3.C14: Distributions of critical drought magnitudes for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River). Subscript "h" means the historical.

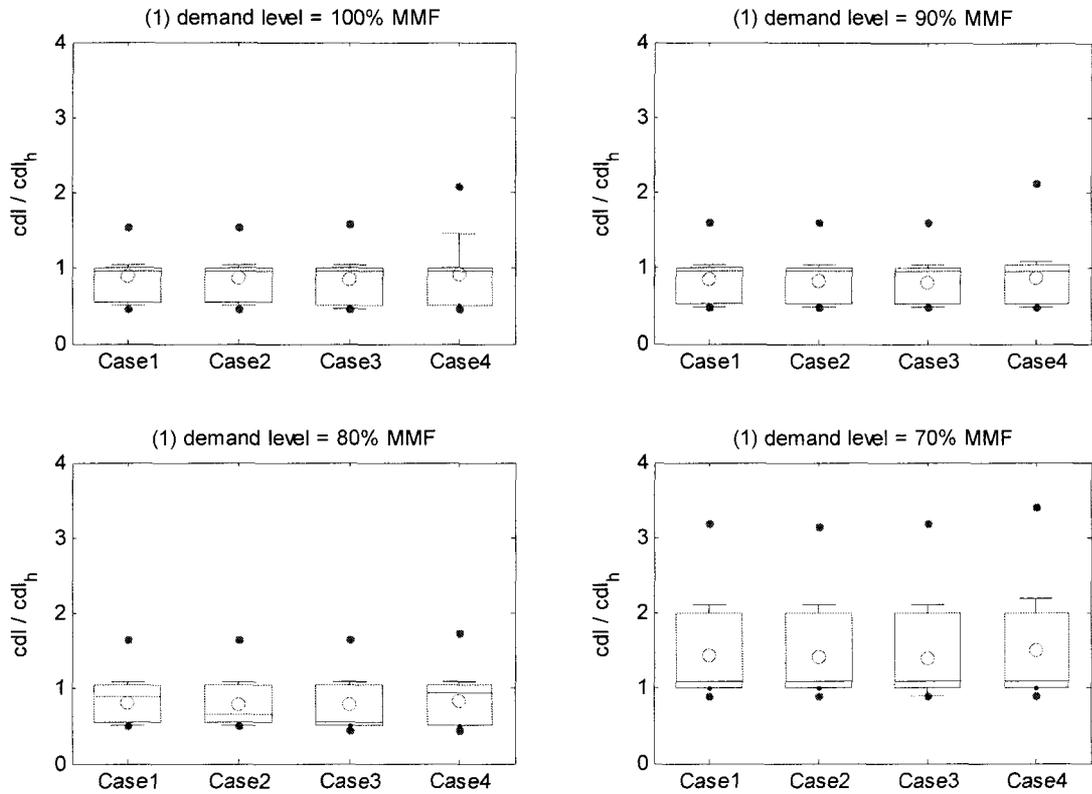


Figure 3.C15: Distributions of critical drought lengths for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 98 years (SPC, Bayesian approaches, Lee's Ferry in Colorado River). Subscript "h" means the historical.

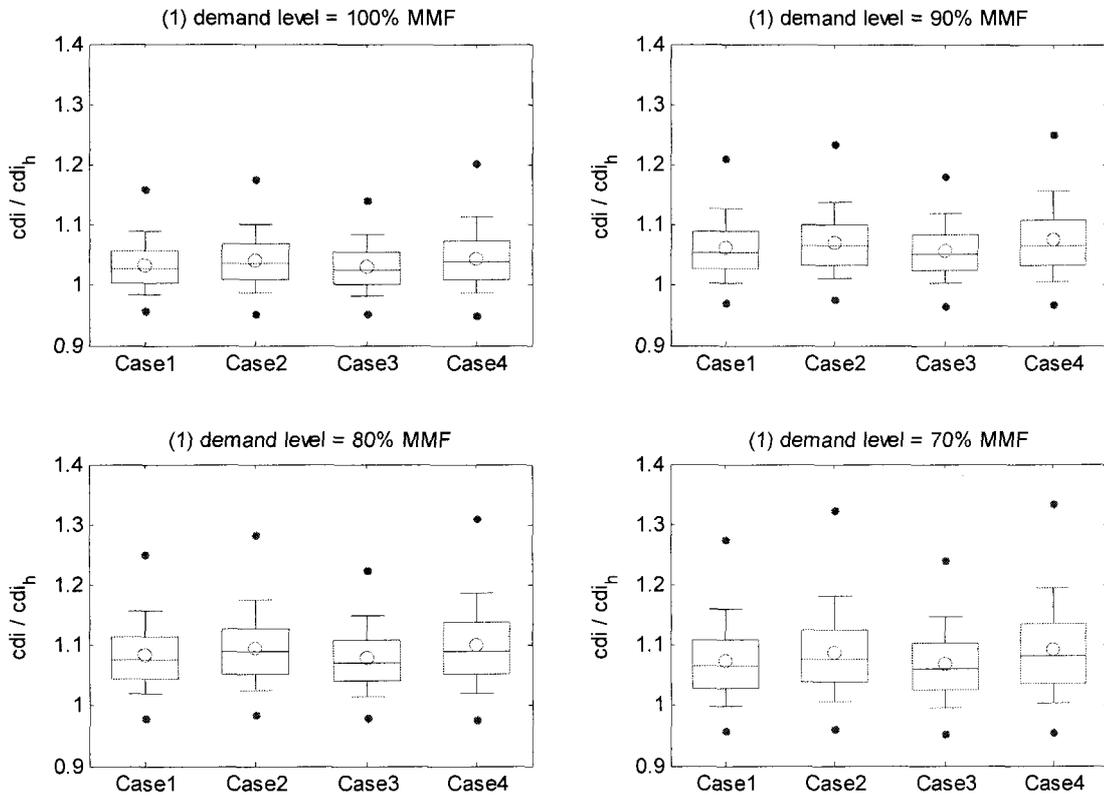


Figure 3.C16: Distributions of critical drought intensities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 98 years (SPC, Bayesian approaches, St. Lawrence River). Subscript “h” means the historical.

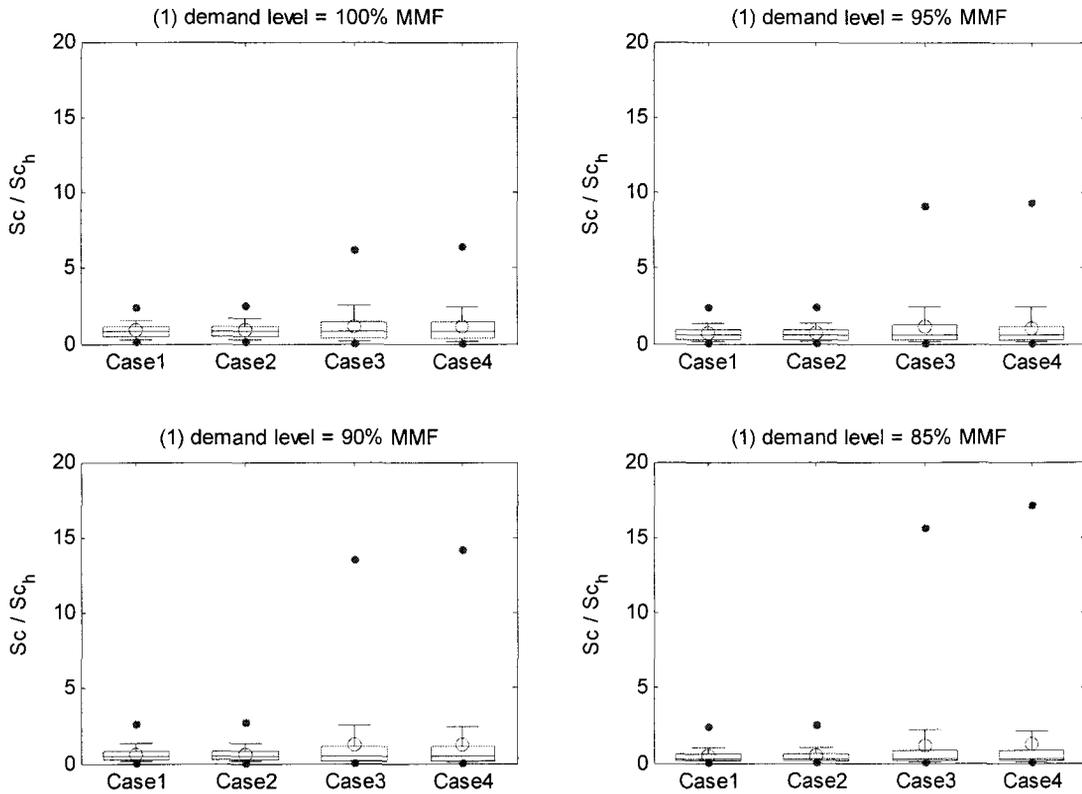


Figure 3.C17: Distributions of storage capacities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River). Subscript “h” means the historical.

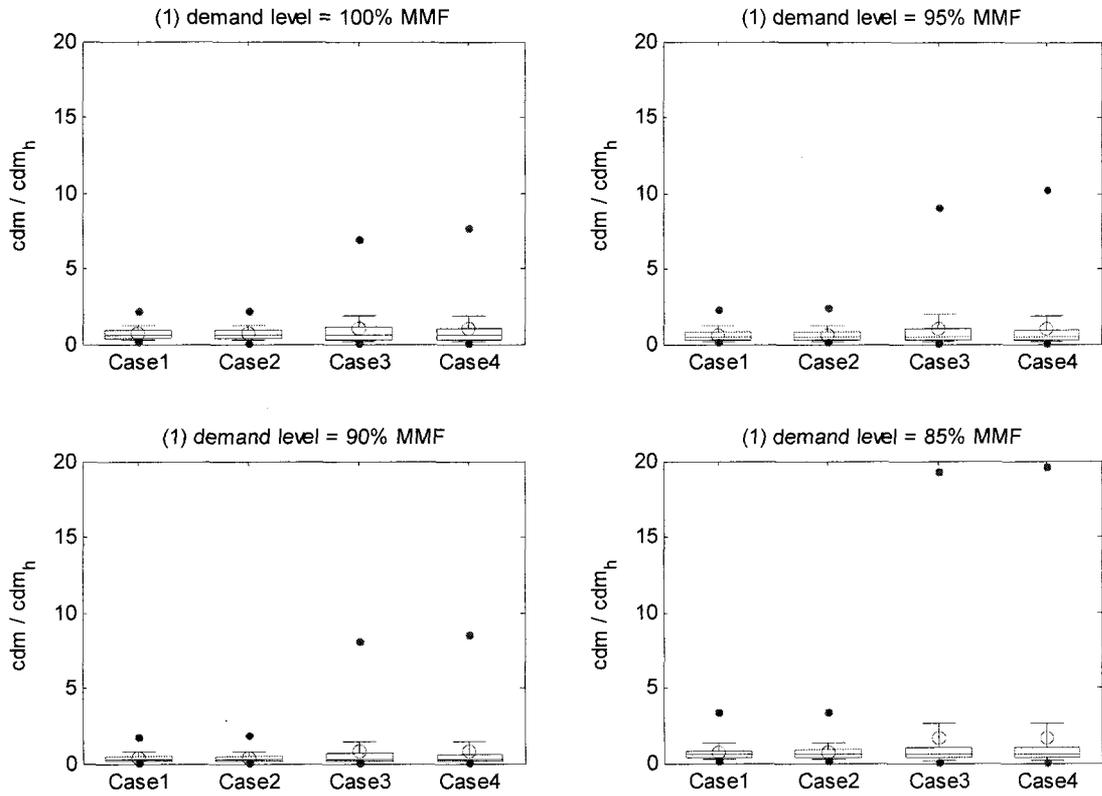


Figure 3.C18: Distributions of critical drought magnitudes for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River). Subscript “h” means the historical.

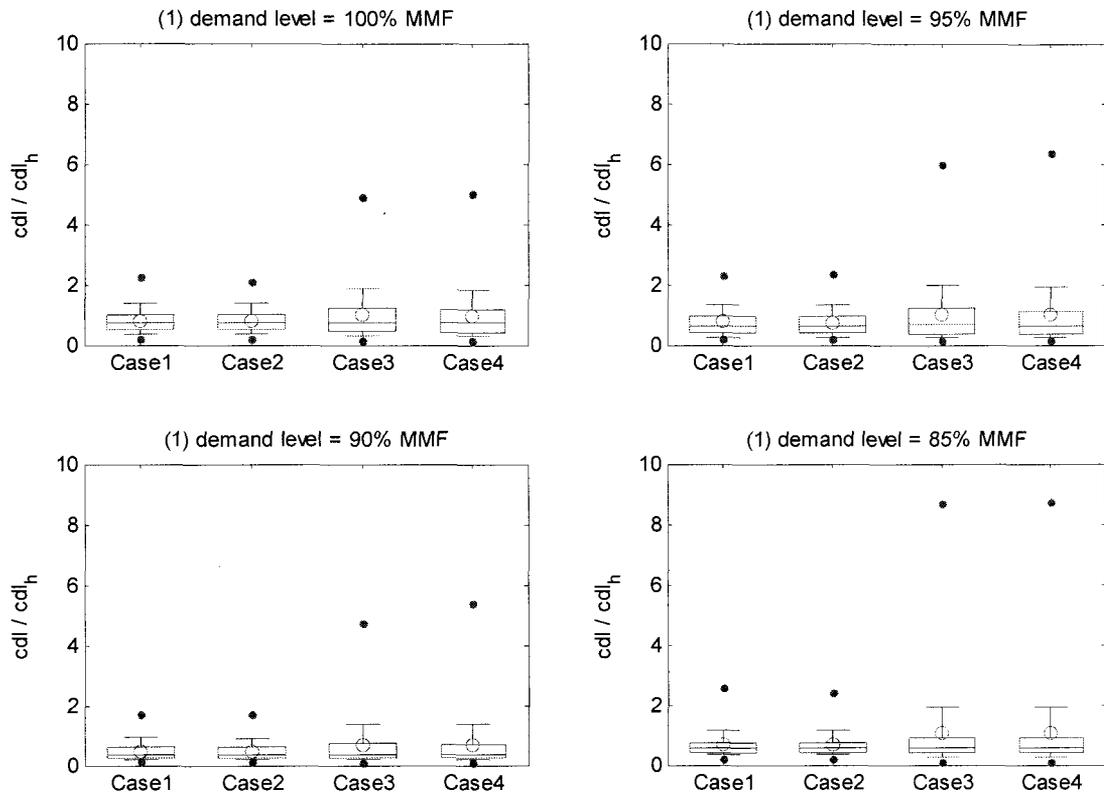


Figure 3.C19: Distributions of critical drought lengths for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River). Subscript “h” means the historical.

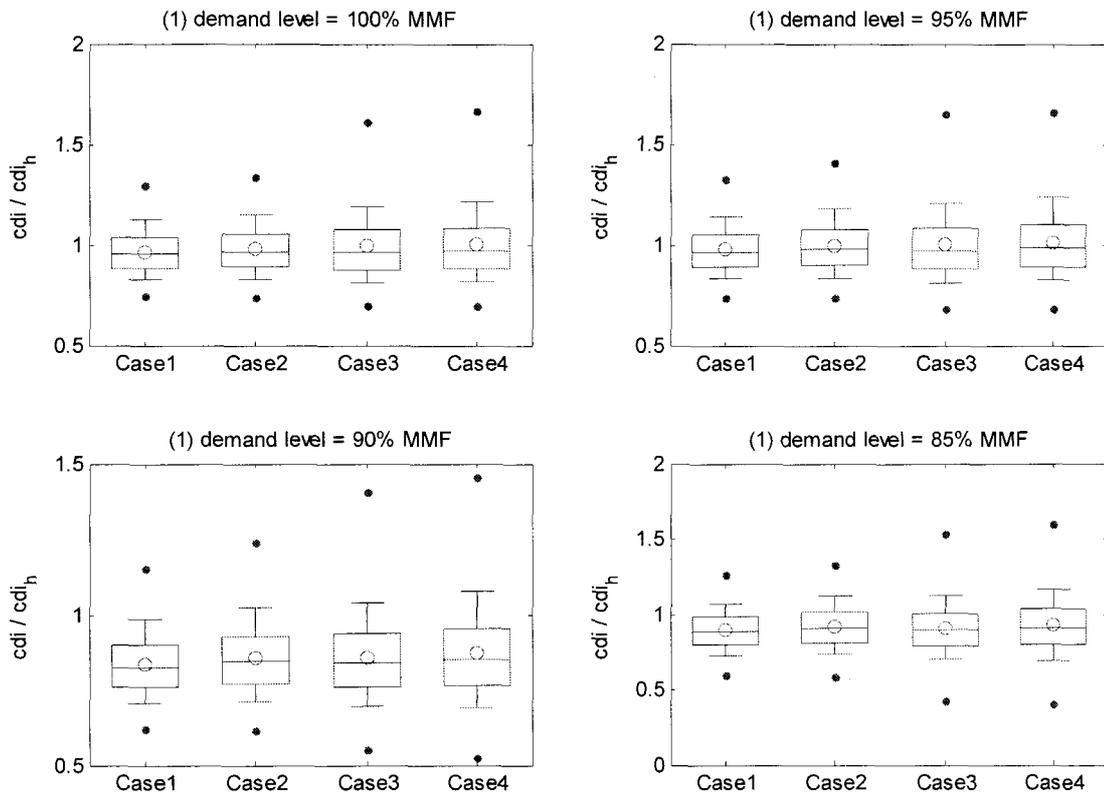


Figure 3.C20: Distributions of critical drought intensities for different uncertainty consideration cases in annual-monthly flow generation. Storage capacities are calculated with the sample size of 50 years and the planning horizon of 59 years (SPC, Bayesian approaches, St. Lawrence River). Subscript “h” means the historical.

## Chapter IV

### PARAMETER UNCERTAINTY IN SPATIAL DISAGGREGATION

**Abstract:** Parameter uncertainty considerations might be useful when using synthetically generated streamflows in the design or management of hydrologic structures. Parameter uncertainty is quantified in two ways; one using the asymptotic distribution of the parameter estimates and the other using Bayesian inference. A simple Valencia-Schaake disaggregation model is implemented to generate synthetic streamflows at sub-stations from the given key-station flows. In order to incorporate parameter uncertainty into the data generation, parameter sets of the disaggregation model are replaced with those sampled from the derived asymptotic distribution or from the Bayesian posterior distribution. Parameter uncertainty of given input variables of the disaggregation model is also taken into account and combined effects of two different kinds of parameter uncertainty are examined as well. As an example, 5000 different streamflow sets are generated for different sample sizes based on historical annual streamflow data at three sites in the Colorado River Basin, and parameter uncertainty effects are evaluated and compared for different scenarios based on generated statistics, as well as reservoir and drought related statistics. As a result, parameter uncertainty shows considerable effects on the variability of calculated reservoir and drought statistics especially for relatively

small sample sizes (equal or less than 50), and it is still visible for sample size of 100 that has been traditionally thought of as a large sample size. The parameter uncertainty effect associated with input variables (key-station flows) of a disaggregation model on generated streamflows is shown to be dominant over that from the spatial disaggregation model. Moreover, the effect of parameter uncertainty of disaggregation models is also reported when combined with parameter uncertainty in the input variables generation.

## 4.1 Introduction

The disaggregation model has been proposed as an alternative to the direct stochastic model because of its simple structure and the good ability to preserve relevant statistical properties of interest. Several disaggregation models and their applications are available in the literature based on either the parametric or non-parametric sense (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Tao and Delleur, 1979; Lane, 1979; Salas et al., 1980; Todini, 1980; Stedinger and Vogel, 1984; Grygier and Stedinger, 1988, 1990; Oliveira et al., 1988; Salas et al., 2000; Koutsoylannis, 2001; Kumar et al., 2001). For example, Salas et al. (1980) applied the disaggregation model in the spatial case by disaggregating the total annual precipitation over an area into several sub-areas and showed that the model performed well in preserving statistical properties.

The Valencia and Schaake (VS) model applied to spatial disaggregation may be written as: (Valencia and Schaake, 1973)

$$\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{B} \mathbf{E}, \quad (4.1)$$

where  $\mathbf{Y}$  is the dependent vector of the sub-station flows to be generated and  $\mathbf{X}$  represents the key-station flows.  $\mathbf{X}$  and  $\mathbf{Y}$  are assumed to be normally distributed with zero means.  $\mathbf{E}$  is an error term with a zero mean vector and a unit variance-covariance matrix.  $\mathbf{A}$  and  $\mathbf{B}$  are the parameter matrices that may enable preserving cross correlations between key- and sub-stations, as well as correlations among sub-stations.

In LAST package, Lane (1979) proposed the spatial disaggregation model which is identical to the temporal disaggregation model by the Mejia and Rousselle model (1976) where the difference with the temporal model is the definition of variables in the model. The Mejia- Rousselle's scheme can be given as:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{E} + \mathbf{C}\mathbf{Z}. \quad (4.2)$$

Instead of defining  $\mathbf{X}, \mathbf{Y}$  as normally distributed annual and seasonal random variables,  $\mathbf{X}, \mathbf{Y}$  are treated as key-station and sub-station streamflows to be generated, and  $\mathbf{Z}$  is a column matrix of the sub-station annual flows in the previous year. Similar to the temporal disaggregation model, this spatial model could preserve cross correlations over key-stations, over sub-stations, between key-stations and sub-stations, as well as lag-1 correlations among sub-stations.

These spatial disaggregation models could be staged for data generation of the complex real river systems (Grygier and Stedinger, 1990; Salas et. al, 2000). Depending on the basin of interest and the underlying problem, key-stations may be one or several sites downstream, or they may be index stations that are a group of sub-stations. Generated streamflows at key-stations are disaggregated into sub-stations, and then the sub-station flows are further disaggregated into sub-substations and so on, finally to the

tributary. Grygier and Stedinger (1990) developed the SPIGOT package which is intended to simulate a multivariate synthetic series based on the staged disaggregation. As a first step, seasonal disaggregation of aggregated annual flows into aggregated seasonal flows or into individual seasonal flows at key stations is performed, and then it is extended to seasonal flows at sub-stations using the spatial disaggregation model proposed by Stedinger and Vogel (1984). Grygier and Stedinger (1988) claimed that this approach might be able to reproduce concurrent cross correlations of seasonal flows rather than the LAST model, but not for annual flows at the various key stations, explicitly.

Based on Valencia-Schaake's model, Stedinger and Vogel (1984) suggested the spatial disaggregation model be expressed in a linear regression form with correlated residual terms. In their model, the relationship between monthly flows at sub-stations and key-stations can be expressed in terms of aggregated monthly flows and serially correlated innovations:

$$Y_t = A + B Z_t + W_t, \quad (4.3)$$

where  $Y_t$  is the normally transformed monthly flows at the sub-stations and  $Z_t$  is the aggregated normally transformed monthly flows at key-stations in the current year  $t$ . Also,  $A$  and  $B$  are the parameter matrices, and the residual term  $W_t$  is given in the autoregressive form as:

$$W_t = C W_{t-1} + V_t, \quad (4.4)$$

where  $V_t$  is the independent normally distributed residuals with zero mean and the specific spatial variance-covariance matrix that could explain the month-month

correlations in transformed monthly flows at each site. Parameter matrix  $C$  is a diagonal matrix, which suggests that only lag-1 auto-correlations could be reproduced in generated annual streamflows at sub-stations.

The condensed temporal disaggregation model might be useful for of a spatial disaggregation case in the complex system (Salas et al., 2000). For example, Lane's condensed temporal disaggregation model can be written for the spatial disaggregation of annual flows as: (Lane, 1979)

$$Y_t^{(i)} = A^{(i)} X + B^{(i)} E_t^{(i)} + C^{(i)} Y_{t-1}^{(i)}, \quad (4.5)$$

in which  $Y_t^{(i)}, X, Y_{t-1}^{(i)}, E_t^{(i)}, A^{(i)}, B^{(i)}$ , and  $C^{(i)}$  are defined in a way similar to Eq.(4.2).

The subscript  $t$  indices the current year and the superscript  $(i)$  denotes the corresponding site  $i$ . After generating flows at each site, an adjustment procedure is required to ensure the additivity constraints due to the separate site by site generations and the normality assumption of flows.

As with the temporal disaggregation case which was discussed in the previous chapter, the uncertainty issue regarding estimated parameters of disaggregation models also arises because the available historical sample is of limited size. Grygier and Stedinger (1990) suspected that the parameter uncertainty effect in the spatial disaggregation to be relatively small compared with that of the annual or seasonal basin flows at a single site, and they proposed that the annual to seasonal disaggregation model should be used for incorporating parameter uncertainty into the complex multivariate annual to seasonal disaggregation. However, extensive exploration of the effect of parameter uncertainty incorporation into the spatial disaggregation has not been clearly made.

In this chapter, a systematical comparison of parameter uncertainty incorporation into the spatial disaggregation model is provided. The simple spatial disaggregation model by Valencia and Schaake (1973) is assumed as the underlying model for the streamflows generation. Based on the multiple regression frameworks, asymptotic distribution and Bayesian posterior distribution for spatial disaggregation model parameters are theoretically derived in order to quantify parameter uncertainty. Several sites in the Colorado River Basin are chosen for the streamflows simulation with parameter uncertainty incorporated, and the impacts of parameter uncertainty are investigated in terms of statistical properties of generated flows, as well as related design variables. In addition, the condensed disaggregation model by Lane (1979) is applied for the sake of comparison where site-by-site flow generations are first performed by the condensed model and generated sub-site flows are adjusted to come up with the key-station flows.

## **4.2 Parameter estimation in spatial disaggregation**

### **4.2.1 Traditional parameter estimation**

Generally, the method of moments estimates of the VS model are given from the sample moments associated with historical annual flows at key-stations and sub-stations by:

$$\hat{\mathbf{A}} = \mathbf{S}'_{XY} \mathbf{S}^{-1}_{XX} \quad (4.6)$$

$$\hat{\mathbf{B}} \hat{\mathbf{B}}' = \mathbf{S}_{YY} - \hat{\mathbf{A}} \mathbf{S}_{XY} \quad (4.7)$$

where  $p$  is the number of sub-stations,  $q$  is the number of key-stations,  $n$  is the sample size,  $\mathbf{S}_{XX}$  is the  $q \times q$  historical sample covariance matrix among key-stations,  $\mathbf{S}_{YY}$  is the  $p \times p$  historical sample covariance matrix among sub-stations,  $\mathbf{S}_{XY}$  is the  $q \times p$  historical sample covariance matrix between key-stations and substations, and  $\mathbf{S}'_{XY}$  is the transpose of  $\mathbf{S}_{XY}$ .  $\mathbf{S}_{XX}$ ,  $\mathbf{S}_{YY}$ , and  $\mathbf{S}_{XY}$  are expressed by:

$$\mathbf{S}_{XX} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{x}_j \mathbf{x}'_j, \quad \mathbf{S}_{YY} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{y}_j \mathbf{y}'_j, \quad \mathbf{S}_{XY} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{x}_j \mathbf{y}'_j, \quad (4.8)$$

where  $p \times 1$  vector  $\mathbf{y}_j$  denotes concurrent sub-station flows at year  $j$  with zero means and  $q \times 1$  vector  $\mathbf{x}_j$  is concurrent key-station flows at year  $j$  with zero means.

Assuming that  $\mathbf{B}$  is a low triangular matrix, the matrix equation  $\hat{\mathbf{B}}\hat{\mathbf{B}}' = \mathbf{D}$  can be solved by using the square root method, which requires that the variance-covariance matrix of error term  $\mathbf{D}$  is a positive semi-definite matrix (Salas et al., 1980).

#### 4.2.2 Maximum likelihood estimation

In the basic VS model, key- and sub-sites flows  $\mathbf{Y}$  and  $\mathbf{X}$  are required to have the zero means. In order to allow for parameter uncertainty of the historical mean through the model, the VS model of Eq.(4.1) is rewritten with the sample mean of historical flows added into the model as:

$$\mathbf{Y} = \mathbf{X}\Theta + \mathbf{V}, \quad (4.9)$$

where  $\mathbf{Y}$ ,  $\mathbf{X}$ ,  $\Theta$ ,  $\mathbf{V}$  have the dimension of  $n \times p$ ,  $n \times (q+1)$ ,  $(q+1) \times p$ , and  $n \times p$ , respectively. A sub-station flows matrix is defined by  $\mathbf{Y} \equiv (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_j, \dots, \mathbf{y}_n)'$  where

$\mathbf{y}_j$  is a  $1 \times p$  vector defined by  $\mathbf{y}_j = (y_j^{(1)}, y_j^{(2)}, \dots, y_j^{(p)})$  for year  $j$  and the superscript indexes the sub-station flows. Also,  $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n)'$  is the newly defined key station flows matrix (actually, this  $\mathbf{X}$  is different from Eq.(4.1)), where  $1 \times (q+1)$  dimensional vector  $\mathbf{x}_j = (1, x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(q)})$  shows key-station flows for year  $j$  (Note again that in this model key-station and sub-station flows  $\mathbf{X}$  and  $\mathbf{Y}$  are not required to have zero means as was the case in Eq.(4.1)). An unknown parameter  $\Theta$  is given by  $\Theta \equiv (\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_p)$ , where  $\theta_j = (\alpha_j, \beta_j^{(1)}, \beta_j^{(2)}, \dots, \beta_j^{(q)})'$  and a residual matrix is given by  $\mathbf{V} \equiv (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n)'$ , where  $\mathbf{v}_j = (v_j^{(1)}, v_j^{(2)}, \dots, v_j^{(p)})'$  is a  $1 \times p$  random vector. It is assumed that  $\mathbf{v}_j$  is independent and normally distributed with  $1 \times p$  dimensional zero means and the positive definite variance-covariance matrix  $\text{var}(\mathbf{v}_j) = \Sigma_{p \times p}$ :

$$\mathbf{v} \sim MVN_p(0, \Sigma_{p \times p}). \quad (4.10)$$

Then, the likelihood function of  $\mathbf{v}$  given  $\mathbf{X}, \mathbf{Y}$ ,  $L(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma)$  is easily given by:

$$L(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma) = \prod_{i=1}^n f(\mathbf{v}_i) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \sum_{i=1}^n \mathbf{v}_i \Sigma^{-1} \mathbf{v}_i'\right].$$

From the properties of a trace of a matrix, it follows that  $\sum_{i=1}^n \mathbf{v}_i \Sigma^{-1} \mathbf{v}_i' = \text{tr}\left(\Sigma^{-1} \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i'\right) = \text{tr}(\Sigma^{-1} \mathbf{V}'\mathbf{V})$ , thus

$$L(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{V}'\mathbf{V}\right], \quad (4.11)$$

and the log-likelihood function  $LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma)$  is derived by:

$$LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma) = \log L(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log|\Sigma| - \frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{V} \quad (4.12)$$

where the first derivative with respect to  $\Theta$  follows:

$$\begin{aligned} \frac{\partial LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma)}{\partial \Theta} &= \frac{\partial}{\partial \Theta} \left( -\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{V} \right) = -\frac{1}{2} \frac{\partial}{\partial \Theta} \text{tr} \left( \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\Theta)' (\mathbf{Y} - \mathbf{X}\Theta) \right) \\ &= -\frac{1}{2} \frac{\partial}{\partial \Theta} \text{tr} \left( \Sigma^{-1} \mathbf{Y}' \mathbf{Y} - \Sigma^{-1} \mathbf{Y}' \mathbf{X}\Theta - \Sigma^{-1} \Theta' \mathbf{X}' \mathbf{Y} + \Sigma^{-1} \Theta' \mathbf{X}' \mathbf{X}\Theta \right). \end{aligned} \quad (4.13)$$

A derivative of matrix  $\mathbf{Q}_{r \times s}$  with respect to  $\mathbf{P}_{r \times j}$  is defined by  $\frac{d\mathbf{Q}}{d\mathbf{P}} = \frac{\partial}{\partial \mathbf{P}} \otimes \mathbf{Q}$  so that

the  $(i, j)$  element of the derivative is shown as:

$$\left[ \frac{d\mathbf{Q}}{d\mathbf{P}} \right]_{i,j} = \begin{pmatrix} \frac{\partial q_{11}}{\partial p_{ij}} & \dots & \frac{\partial q_{1s}}{\partial p_{ij}} \\ \frac{\partial p_{ij}}{\partial p_{ij}} & \ddots & \frac{\partial p_{ij}}{\partial p_{ij}} \\ \vdots & & \vdots \\ \frac{\partial q_{r1}}{\partial p_{ij}} & & \frac{\partial q_{rs}}{\partial p_{ij}} \\ \frac{\partial p_{ij}}{\partial p_{ij}} & & \frac{\partial p_{ij}}{\partial p_{ij}} \end{pmatrix}.$$

Using the properties of derivatives regarding a scalar function of a matrix for derivative terms in Eq.(4.13) gives:

$$\frac{\partial}{\partial \Theta} \text{tr} \left( \Sigma^{-1} \mathbf{Y}' \mathbf{X}\Theta \right) = \mathbf{X}' \mathbf{Y} \Sigma^{-1},$$

$$\frac{\partial}{\partial \Theta} \text{tr} \left( \Sigma^{-1} \Theta' \mathbf{X}' \mathbf{Y} \right) = \mathbf{X}' \mathbf{Y} \Sigma^{-1},$$

$$\frac{\partial}{\partial \Theta} \text{tr} \left( \Sigma^{-1} \Theta' \mathbf{X}' \mathbf{X}\Theta \right) = \frac{\partial}{\partial \Theta} \text{tr} \left( \Theta \Sigma^{-1} \Theta' \mathbf{X}' \mathbf{X} \right) = \mathbf{X}' \mathbf{X}\Theta \Theta^{-1} + \mathbf{X}' \mathbf{X} \Theta (\Sigma^{-1})' = 2\mathbf{X}' \mathbf{X}\Theta \Sigma^{-1}.$$

Thus, Eq.(4.13) is simplified as:

$$\frac{\partial LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma)}{\partial \Theta} = \mathbf{X}' \mathbf{Y} \Sigma^{-1} - \mathbf{X}' \mathbf{X}\Theta \Sigma^{-1}. \quad (4.14)$$

Maximum likelihood estimators of  $\Theta$  can be derived by setting Eq.(4.14) equal to zero, which yields:

$$\hat{\Theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}, \quad (4.15)$$

which is equivalent to the general least square estimators. Let  $\Lambda \equiv \Sigma^{-1}$  for convenience in derivation, and then differentiating Eq.(4.12) with respect to  $\Lambda$  gives:

$$\begin{aligned} \frac{\partial LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}, \Theta, \Sigma)}{\partial \Lambda} &= \frac{n}{2} \frac{\partial}{\partial \Lambda} (\log |\Lambda^{-1}|) - \frac{1}{2} \frac{\partial}{\partial \Lambda} (\text{tr} \Lambda \mathbf{V}'\mathbf{V}) \\ &= \frac{n}{2|\Lambda|} (2|\Lambda|\Lambda^{-1} - |\Lambda| \text{diag}(\Lambda^{-1})) - \frac{1}{2} (2\mathbf{U}'\mathbf{U} - \text{diag}(\mathbf{V}'\mathbf{V})), \end{aligned} \quad (4.16)$$

where  $\text{diag}(\cdot)$  is defined by the vector consisting of the elements of the main diagonal of a square matrix. Since  $\text{diag}(\Lambda^{-1}) = \text{diag}(\Sigma) = n^{-1} \text{diag}(\hat{\mathbf{V}}'\hat{\mathbf{V}})$ , setting the above equation equal to zero shows:

$$\frac{\partial LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}, \Theta, \Sigma)}{\partial \Lambda} = n\Lambda^{-1} - \hat{\mathbf{V}}'\hat{\mathbf{V}} = 0 \quad (4.17)$$

and thus maximum likelihood estimates of  $\Sigma$  becomes:

$$\hat{\Sigma} = \hat{\Lambda}^{-1} = \frac{1}{n} \hat{\mathbf{V}}'\hat{\mathbf{V}}. \quad (4.18)$$

In practical consideration of the spatial disaggregation model, when  $\mathbf{y}_j$  and  $\mathbf{x}_j$  are normally distributed and the sum of  $\mathbf{y}_j$  is expected to be  $\mathbf{x}_j$ ,  $\mathbf{D}$  becomes the singular matrix, which might not enable one to incorporate the parameter uncertainty. Thus, a constraint that  $\mathbf{D}$  should be positive definite would be required for the consideration of parameter uncertainty. To avoid the singularity of the variance-covariance matrix in the actual application, transformation of real flows will be implemented and the corresponding adjustment procedure will be followed to obtain the generated flows through the disaggregation model.

### 4.3 Parameter uncertainty consideration

In order to quantify the uncertainty of parameters in the spatial disaggregation model, sampling distributions were derived using large sample theory and Bayesian inference. As addressed before, two possible parameter estimators of VS disaggregation models are available, and those estimates might be different. For parameter uncertainty consideration, a maximum likelihood estimator (MLE) will be used instead of a method of moments estimator because MLE (least square estimator) is asymptotically efficient.

#### 4.3.1 Asymptotic distribution of the parameters

A  $p \times p$  dimensional matrix  $\mathbf{W}$  is distributed as a  $p$ -dimensional central Wishart distribution with scale matrix  $\mathbf{\Sigma}$  and  $n_d$  degrees of freedom, i.e.,  $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n_d)$ , if and only if  $\mathbf{W} = \mathbf{U}\mathbf{U}'$  for some matrix  $\mathbf{U} \equiv (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_d})'$  where  $p \times 1$  random vector  $\mathbf{u} \sim MVN_p(0, \mathbf{\Sigma}_{p \times p})$  (Wishart, 1928). It is generally regarded as the multivariate extension of the chi square distribution. The expectation matrix  $E(\mathbf{W})$  and the dispersion matrix  $D(\mathbf{W})$  which is the variance-covariance matrix of vectorized  $\mathbf{W}$  are given by (Kollo and Rosen, 2005):

$$E(\mathbf{W}) = E(\mathbf{u}\mathbf{u}') = n_d \mathbf{\Sigma}_{p \times p} \quad (4.19)$$

$$D(\mathbf{W}) = E[\text{vec}(\mathbf{u})\text{vec}'(\mathbf{u})] = n_d (\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) (\mathbf{\Sigma}_{p \times p} \otimes \mathbf{\Sigma}_{p \times p}) \quad (4.20)$$

where,  $\text{vec}(\cdot)$  is a vectorization operator from  $R^{p \times q}$  to  $R^{pq \times 1}$  and  $\mathbf{K}_{p,p}$  is defined as

the commutation matrix of the  $p^2 \times p^2$  dimension consisting of  $p \times p$  blocks which has the  $p \times p$  sub matrix with the element of  $k_{j,i} = 1$  or otherwise 0 in an  $(i, j)$  block.

For example,

$$\mathbf{K}_{p \times p} = \begin{matrix} & & p & & \cdots & & p & & \\ & & & & & & & & \\ p & & \left[ \begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & & 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & & & & 0 & 0 & \cdots & 0 \\ & & & & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & & 0 & 0 & 0 & 1 \end{array} \right] & & & & & & & & \\ & & & & & & & & \\ p & & & & & & & & \end{matrix}$$

From Eqs.(4.9) and (4.18), an unbiased variance-covariance matrix estimator of

$\Sigma$  is given by  $\hat{\Sigma} = (n - n_d)^{-1} \sum_{i=1}^n \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i'$  where  $n_d = n - (q + 1)$ ; further  $\hat{\Sigma}$  converges in

probability to  $\Sigma$ . Also, by the central limit theorem,  $\sqrt{n}(\text{vec}(\hat{\Sigma} - \Sigma))$  converges in

distribution to a  $p^2$  dimensional multivariate normal distribution with zero mean vector

and variance-covariance matrix  $D(\Sigma) = (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\Sigma_{p \times p} \otimes \Sigma_{p \times p})$  (See Kollo and

Rosen, 2005), i.e.,

$$\sqrt{n} \text{vec}(\hat{\Sigma}_{p \times p} - \Sigma_{p \times p}) \xrightarrow{D} MVN_{p^2} \left( \mathbf{0}_{p^2}, (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\Sigma_{p \times p} \otimes \Sigma_{p \times p}) \right) \quad (4.21)$$

Hence, the asymptotic distribution of the vectorized sample variance-covariance matrix is

given by:

$$\text{vec}(\hat{\Sigma}) \sim AMVN_{p^2} \left( \text{vec}(\Sigma_{p \times p}), \frac{1}{n} (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\Sigma_{p \times p} \otimes \Sigma_{p \times p}) \right), \quad (4.22)$$

where *AMVN* means an asymptotic multivariate distribution.

Recall Eq.(4.14) as:

$$\frac{\partial LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \boldsymbol{\Theta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Theta}} = \mathbf{X}'\mathbf{Y} \boldsymbol{\Sigma}^{-1} - \mathbf{X}'\mathbf{X}\boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1}.$$

Thus, the second derivative of the loglikelihood function is given by:

$$\begin{aligned} & \frac{\partial^2 LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \boldsymbol{\Theta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Theta}^2} \\ &= -\frac{\partial}{\partial \boldsymbol{\Theta}} \mathbf{X}'\mathbf{X}\boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1} = -\boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X})' = -\boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X}) \end{aligned} \quad (4.23)$$

The information matrix for parameter  $\boldsymbol{\Theta}$  can be derived by taking the expectation of the negative of the second derivative shown above, of which the inverse could be defined as the Cramer Rao lower bound *CRLB*, that is:

$$CRLB(\hat{\boldsymbol{\Theta}}) = E \left[ -\frac{\partial^2 LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \boldsymbol{\Theta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Theta}^2} \right]^{-1} = (\boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X}))^{-1} = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}$$

The *CRLB* is a lower bound for unbiased estimators and gives the variance-covariance matrix of the maximum likelihood estimators for large samples. Thus, the asymptotic distribution of the maximum likelihood parameter estimator  $\hat{\boldsymbol{\Theta}}$  becomes:

$$vec(\hat{\boldsymbol{\Theta}}) \sim AMVN_{q \times p} \left( vec(\boldsymbol{\Theta}), \boldsymbol{\Sigma}_{p \times p} \otimes (\mathbf{X}'\mathbf{X})_{q \times q}^{-1} \right). \quad (4.24)$$

For example, the relationship of derived asymptotic distributions of the spatial disaggregation parameter estimates with asymptotic distributions of a temporal disaggregation model (which were provided in the previous chapter) will be briefly examined. Recall the SPC spatial disaggregation model in the multivariate regression form as:

$$\mathbf{Y}_{n \times 1}^\tau = \mathbf{X}_{n \times 4} \boldsymbol{\Theta}_{4 \times 1}^\tau + \mathbf{V}_{n \times 1}^\tau, \quad \text{for each season } \tau = 2, \dots, 12,$$

where the subscript denotes the dimension of corresponding matrix,  $\mathbf{Y}_{n \times 1}^\tau$

$$= [y_{1,\tau}, y_{2,\tau}, \dots, y_{n,\tau}]', \quad \mathbf{V}_{n \times 1}^\tau = [v_{1,\tau}, v_{2,\tau}, \dots, v_{n,\tau}], \quad \Theta_{4 \times 1}^\tau = (\alpha_t, \beta_t, \gamma_t, \delta_t)', \text{ and}$$

$$\mathbf{X}_{n \times 4} = \begin{bmatrix} 1 & y_{1,\tau-1} & x_1 & \sum_{s=1}^{t-1} w_s y_{1,s} \\ 1 & y_{2,\tau-1} & x_2 & \sum_{s=1}^{t-1} w_s y_{2,s} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{n,\tau-1} & x_n & \sum_{s=1}^{t-1} w_s y_{n,s} \end{bmatrix} = \begin{bmatrix} 1 & y_{1,\tau-1} & x_1 & z_1 \\ 1 & y_{2,\tau-1} & x_2 & z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{n,\tau-1} & x_n & z_n \end{bmatrix}.$$

Since  $p=1$ , the asymptotic distribution of  $\hat{\sigma}_v = \frac{1}{n-4} \sum_{i=1}^n v_{i,\tau} v'_{i,\tau}$  directly follows from

Eq.(4.22) as:

$$\hat{\sigma}_v^2 \sim AN \left( \sigma_v^2, \frac{2}{n} \sigma_v^4 \right), \quad (4.25)$$

where  $AN$  denotes asymptotic normal distribution. Also, from Eq.(4.24) the asymptotic distribution of a parameter matrix  $\Theta$  is given by:

$$\hat{\Theta}^\tau \sim AMVN_4 \left( \Theta^\tau, \sigma_v^2 (\mathbf{X}'\mathbf{X})^{-1} \right) = AMVN_4 \left( \Theta^\tau, \frac{\sigma_v^2}{n^2} \mathbf{H} \right), \quad (4.26)$$

where

$$\mathbf{H} = \begin{pmatrix} 1 & \mu_{Y(\tau-1)} & \mu_{X(\tau)} & \mu_{Z(\tau)} \\ \sigma_{Y(\tau-1)}^2 + \mu_{Y(\tau-1)}^2 & \sigma_{X(\tau)Y(\tau-1)} + \mu_{X(\tau)}\mu_{Y(\tau-1)} & \sigma_{Z(\tau)Y(\tau-1)} + \mu_{Z(\tau)}\mu_{Y(\tau-1)} \\ \text{symm.} & \sigma_{X(\tau)}^2 + \mu_{X(\tau)}^2 & \sigma_{X(\tau)Z(\tau)} + \mu_{X(\tau)}\mu_{Z(\tau)} \\ & & \sigma_{Z(\tau)}^2 + \mu_{Z(\tau)}^2 \end{pmatrix}^{-1}.$$

Both Eq.(4.25) and Eq.(4.26) are equivalent to the asymptotic distributions of parameter estimates of the SPC temporal disaggregation model given in the previous chapter.

### 4.3.2 Bayesian inference on parameters of the spatial disaggregation

Recall Eq.(4.9) as:

$$\mathbf{Y} = \mathbf{X}\Theta + \mathbf{V} ,$$

where rows of  $\mathbf{V}$  are independently distributed, each with an  $p$  dimensional zero vector and variance–covariance matrix  $\Sigma$ . Under this assumption, the probability density function for  $\mathbf{Y}$  given  $\mathbf{X}, \Theta$  and  $\Sigma$  is:

$$p(\mathbf{Y} | \mathbf{X}, \Theta, \Sigma) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}\left(\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\Theta)'(\mathbf{Y} - \mathbf{X}\Theta)\right)\right] \quad (4.27)$$

where  $(\mathbf{Y} - \mathbf{X}\Theta)'(\mathbf{Y} - \mathbf{X}\Theta)$  can be newly defined by using sample statistics and least square estimates of parameters as  $(\mathbf{Y} - \mathbf{X}\Theta)'(\mathbf{Y} - \mathbf{X}\Theta) = \mathbf{M} + (\Theta - \hat{\Theta})' \mathbf{X}' \mathbf{X} (\Theta - \hat{\Theta})$ , where the least square estimates matrix  $\mathbf{M} \equiv (\mathbf{Y} - \mathbf{X}\hat{\Theta})'(\mathbf{Y} - \mathbf{X}\hat{\Theta})$  and  $\hat{\Theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ .

Then, the likelihood function for  $\Theta$  and  $\Sigma$  follows:

$$L(\Theta, \Sigma | \mathbf{X}, \mathbf{Y}) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}(\mathbf{M}\Sigma^{-1}) - \frac{1}{2} \text{tr}\left((\Theta - \hat{\Theta})' \mathbf{X}' \mathbf{X} (\Theta - \hat{\Theta})\Sigma^{-1}\right)\right]. \quad (4.28)$$

The invariance theory that probability statements on observable random variables should remain invariant under changes in the parameterization of the problem (Jeffery, 1961) can give one an idea about the prior distribution of a parameters sets  $p(\Theta, \Sigma)$ , such that  $p(\Theta)$  is a constant and  $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$ . Since the prior distributions of  $\Theta$  and  $\Sigma$  are assumed to be independent, the joint prior distribution of parameters is obtained as a diffuse prior probability density function by:

$$p(\Theta, \Sigma) = p(\Theta) p(\Sigma) \propto |\Sigma|^{-(p+1)/2} \quad (4.29)$$

Combining Eq.(4.28) and Eq.(4.29) yields the joint posterior pdf of  $\Theta$  and  $\Sigma$  as:

$$\begin{aligned} p(\Theta, \Sigma | \mathbf{X}, \mathbf{Y}) &\propto |\Sigma|^{-(n+p+1)/2} \exp\left[-\frac{1}{2}tr(\mathbf{M}\Sigma^{-1}) - \frac{1}{2}tr((\Theta - \hat{\Theta})' \mathbf{X}'\mathbf{X}(\Theta - \hat{\Theta})\Sigma^{-1})\right] \\ &= |\Sigma|^{-(n+p+1)/2} \times \exp\left[-\frac{1}{2}tr(\mathbf{M}\Sigma^{-1}) - \frac{1}{2}(vec\Theta - vec\hat{\Theta})'(\Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1})^{-1}(vec\Theta - vec\hat{\Theta})\right], \end{aligned} \quad (4.30)$$

where  $vec\Theta$  is a vectorized form of the parameter matrix  $\Theta$ ,  $vec\Theta = vec(\Theta) = (\theta'_1, \theta'_2, \dots, \theta'_p)'$ . From  $p(\Theta, \Sigma | \mathbf{X}, \mathbf{Y}) = p(\Theta | \mathbf{X}, \mathbf{Y}, \Sigma)p(\Sigma | \mathbf{X}, \mathbf{Y})$ ,

$$p(\Sigma | \mathbf{X}, \mathbf{Y}) \propto |\Sigma|^{-(n+p-q+1)/2} \exp\left[-\frac{1}{2}tr(\mathbf{M}\Sigma^{-1})\right] \quad (4.31)$$

$$p(\Theta | \mathbf{X}, \mathbf{Y}, \Sigma) \propto |\Sigma|^{-q/2} \exp\left[-\frac{1}{2}(vec\Theta - vec\hat{\Theta})'(\Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1})^{-1}(vec\Theta - vec\hat{\Theta})\right]. \quad (4.32)$$

Since Eq.(4.31) is the form of inverse central Wishart distribution, then the posterior distribution of  $\Sigma$  given  $\mathbf{X}, \mathbf{Y}$  can be given by:

$$p(\Sigma | \mathbf{X}, \mathbf{Y}) \sim IW(\mathbf{M}, \nu), \quad (4.33)$$

where  $\nu = n - (p + q) + 1$  is the degree of freedom and  $IW$  denotes inverse Wishart distribution. Also, from Eq.(4.32), the conditional posterior pdf of  $\Theta$  given  $\mathbf{X}, \mathbf{Y}, \Sigma$  is a multivariate normal distribution with mean  $vec\hat{\Theta}$  and covariance matrix  $\Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1}$  expressed by:

$$p(vec\Theta | \mathbf{X}, \mathbf{Y}, \Sigma) \sim MVN_{q \times p}\left(vec\hat{\Theta}, \Sigma_{p \times p} \otimes (\mathbf{X}'\mathbf{X})_{q \times q}^{-1}\right) \quad (4.34)$$

which has the equivalent form as the asymptotic distribution of Eq.(4.24). Note that in large samples the posterior distribution of parameters  $\Theta$  will be approximately normal

with mean  $\hat{\Theta}$ , the maximum likelihood estimates, and variance-covariance matrix

$$E \left[ - \frac{\partial^2 LL(\mathbf{v} | \mathbf{X}, \mathbf{Y}; \Theta, \Sigma)}{\partial \Theta^2} \right]_{\Theta = \hat{\Theta}}^{-1}$$

(Zellner, 1971). The assumption that  $\Sigma$  is replaced by  $\hat{\Sigma}$  is partially justified if the sample size is large enough realizing that  $\hat{\Sigma}$  converge in probability  $\Sigma$ , and thus Eq.(4.34) can be regarded as the asymptotic posterior distribution of  $\Theta$  rather than a fixed sample size posterior distribution. When  $\Sigma$  is unknown, the fixed sample size posterior distribution of  $\Theta$  needs to be derived from Eq.(4.30). Using  $\Sigma^{-1}$  instead of  $\Sigma$  in Eq.(4.30) yields:

$$p(\Theta, \Sigma^{-1} | \mathbf{X}, \mathbf{Y}) \propto |\Sigma^{-1}|^{(n-(p+1))/2} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{M}\Sigma^{-1}) - \frac{1}{2} \text{tr}((\Theta - \hat{\Theta})' \mathbf{X}' \mathbf{X} (\Theta - \hat{\Theta}) \Sigma^{-1}) \right], \quad (4.35)$$

and by integrating (4.34) with respect to  $\Sigma^{-1}$  gives (Zellner, 1971)

$$p(\Theta | \mathbf{X}, \mathbf{Y}) \propto |\mathbf{M} + (\Theta - \hat{\Theta})' \mathbf{X}' \mathbf{X} (\Theta - \hat{\Theta})|^{n/2}, \quad (4.36)$$

which is thought of as a matrix-variate extension of the  $t$  distribution (Box and Tiao, 1973) or a generalized multivariate  $t$  distribution (Zellner, 1971). It is not possible to express the distribution of  $\Theta$  as a multivariate  $t$  distribution, but it can be simplified to a multivariate  $t$  distribution when either  $p$  or  $q$  is equal to 1.

## 4.4 Application to Colorado River System

### 4.4.1 Parameter uncertainty effect based on historical key-station flows

As a way of examining uncertainty effects on the synthetic annual streamflows regarding a spatial disaggregation model, a simulation experiment will be performed using historical annual streamflows data sets in the Colorado River Basin: one key-station and three sub-stations with the period of 1905-2002 (USBR, 2007). Figure 4.1 shows the locations of adopted stations in the Colorado River Basin, where the key station (key) is located at the Colorado River main stream above Cisco, Utah (USGS site number: 09180500, site 8), and three sub-stations are: sub-station 1 at Colorado River main stream above Cameo, Colorado (09095500, site 2), sub-station 2 at Gunnison River above Grand Junction, Colorado (09152500, site 6), and sub-station 3 at Dolores River near Cisco, Utah (09180000, site 7). Table 4.1 illustrates that the averaged proportion of each historical sub-station flow to the historical key-station flow is 53%, 35%, and 12%, respectively, and calculated cross correlations exhibit close relationships among either the key-station and sub-stations or sub-stations themselves. A skewness test of historical flows in all stations was performed with a 5% confidence level to ensure the normality assumption for the VS disaggregation model, and power transformations were applied to all four stations. ML parameter estimates of the VS disaggregation model are obtained as:

$$\hat{\Theta} = \begin{bmatrix} 2224 & -24705 & 36 \\ 0.10 & 2.47 & 0.01 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} 168497 & -1121408 & -35385 \\ -1121408 & 30333242 & 174491 \\ -35385 & 174491 & 11441 \end{bmatrix},$$

which shows that sub-station 1 flows is negatively correlated with those of other sub-stations. Based on these parameter estimates, 5,000 different sets of synthetic annual flows in the sub-stations with the same sample size as historical flows will be disaggregated from the key-station flows.

In a traditional spatial disaggregation procedure, key station flows which are used as an input variable are separately generated outside the disaggregation structure. Several stochastic models are available for generating input variables as in the usual case of annual streamflows generation. It is general that generated annual flows will have the variability caused by the error term in the generation model even without parameter uncertainty incorporated, and this variability is called by a natural uncertainty effect. Once the generated annual flows are applied into the disaggregation model, this natural uncertainty will probably propagate into sub station flows. It has been already shown in the previous chapter that the natural uncertainty effect of the annual flow generation on basic statistics of generated monthly flows (means, standard deviations) and related design variables are larger than the temporal disaggregation parameter uncertainty effect. In order to eliminate any uncertainty effect from input variables and to report the inherent uncertainty effect of the spatial disaggregation model, input variables are initially assumed as historical flows in the key station.

Derived asymptotic and posterior distributions are employed to quantify the parameters uncertainty of the spatial disaggregation. In each generation (out of total 5,000 generations), new parameter estimates of the VS model are sampled from the two different distributions and those will be substituted for original parameters estimated from the historical sample. Different sample sizes;  $n=25, 50, 75, 98$  are assumed to be

available to investigate the related uncertainty effect. The given historical sample size  $n = 98$  is assumed to be large enough and parameter estimates based on the historical sample are set as the true parameters. That is, expected values of parameter estimates of  $\Theta$  and  $\Sigma$  are equivalent to historical parameter estimates regardless of sample sizes. In addition, the sum of squares of input variables  $\mathbf{X}'\mathbf{X}$  is assumed to be proportional to historical  $\mathbf{X}'\mathbf{X}$  for different sample sizes (Stedinger, et. al, 1985). Another parameter sets with the zero variances (historical parameter estimates) are referred to “ignoring the parameter uncertainty”, which will be represented by  $n = 98^*$ . As discussed before, the conditional posterior distribution of parameter estimates  $\hat{\Theta}$  in Eq.(4.34) has the equivalent form as the asymptotic distribution of Eq.(4.24). Note that the posterior distribution of  $\Sigma$  is an inverse Wishart distribution, which is used as the multivariate normal distribution in the asymptotic approach.

As illustrated in Figures 4.2 and 4.3 (Figure 4.A1 and 4.A2 in the Appendix), parameter uncertainty results in increased variability of calculated means, standard deviations, and cross correlations of generated annual flows at sub-stations. Increased variability of generated statistics is getting less as the sample size increases, but the parameter uncertainty effect can still be found even when the sample size is equal to 98. However, it seems that parameter uncertainty has little effect on the generated skewness and lag-1 serial correlations. For the generated mean and standard deviation, the Bayesian approach results in wider variability (even not much) than the asymptotic approach for a small sample size (e.g.,  $n < 50$ ). Note that Bayesian posterior distribution shows positive bias of generated standard deviations and negative bias of generated cross correlations for all sub-stations especially for small sample sizes ( $n \leq 50$ ).

As the sample size is getting relatively larger, distributions of generated standard deviations and cross correlations based on the Bayesian approach are getting closer to those from the asymptotic distribution, but still there is a little difference, even when  $n = 98$ . This difference may be caused by the basic assumption of an asymptotic approach which requires large enough sample sizes. A sample size of at least 100 would be required when the asymptotic distribution is applied for parameter uncertainty consideration in the spatial disaggregation as an alternative to Bayesian distribution.

The impact of parameter uncertainty on the statistical properties of generated flows would affect the related design variables which are usually employed as a design criteria for the practical planning and management of water resources. Based on synthetic flows at each sub-station, storage related statistics (storage capacity) and drought indices (critical drought magnitude, length and intensity) are calculated and compared. For more information of the utilized design variables, see previous chapters. In the calculation of the storage capacity based on the sequent peak algorithm, demand levels of reservoirs are assumed 100% of the mean annual flows (MAF), the design period  $N_d$  is assumed equal to the historical sample size of 98, and the reservoirs are assumed to be initially full. Also, 100% MAF is assumed to be the threshold level for critical drought statistics. It is not surprising that generated storage capacities and drought indices with parameter uncertainty incorporated would be closely affected by the sample sizes, in particular when sample sizes are equal or less than 50. Compared with a parameter uncertainty ignored case ( $n = 98^*$ ), increased variations of the storage capacities and drought indices are still visible when the sample size is relatively large ( $n = 98$ ). Asymptotic and Bayesian approaches give almost similar variability of storage

capacity distributions for sample size greater than 50 (See Figure 4.4, Figures 4.A3 and 4.A4 in the Appendix). For  $n=25$ , the Bayesian approach shows larger upper quantiles (i.e. 90, 99% quantiles) of storage capacities than the asymptotic approach. A similar pattern with generated storage capacities could be reported for the generated critical drought magnitudes and intensities, while this pattern is hard to judge for generated critical drought lengths. Based on simulated results of sample statistics and design variables associated with parameter uncertainty, Bayesian posterior distributions of parameter estimates will be used in interpreting the parameter uncertainty effect in the further analysis.

#### **4.4.2 Parameter uncertainty effects combining with uncertainty of input variables**

As discussed in the previous chapter, generated annual flows in key-stations are significantly affected by the parameter uncertainty of the annual generation model and this is related with limited sample size of available historical steamflow data. Through spatial disaggregation, uncertainty effects of annual generation model parameters will be translated into generated sub-station flows. In this section, different combinations will be categorized to investigate the combined parameter uncertainty effect of the disaggregation model on generated sub-station flows with generated key station flows that have parameter uncertainty incorporated. The four unique combinations of natural and parameter uncertainty in key-station generation and spatial disaggregation are summarized in the following:

Cases	Key-station generation	Spatial disaggregation
C1	NU	NU
C2	NU	NU + PU
C3	NU + PU	NU
C4	NU + PU	NU + PU

*Note* NU : natural uncertainty, NU+PU : natural uncertainty + parameter uncertainty

5000 different sets of key-station annual flows are first generated with the same size of historical data by using the first order autoregressive model, AR(1), which are then spatially disaggregated into sub-station flows. In order to have parameter uncertainties embedded, different parameter sets are sampled from Bayesian posterior distribution in the stage of key-station generation or spatial disaggregation in each generation step (depending on cases) for different sample sizes: 25, 50, 75, 100, and 200. Parameter uncertainty effects for the four different cases are compared with each other based on statistical properties of generated flows, as well as calculated design variables.

As shown in Figure 4.5 (Figures 4.A5 and 4.A6 in the Appendix), C2 and C4 show similar variability of generated means with C1 and C3, respectively, from which it can be thought that parameter uncertainty in the disaggregation procedure does not significantly affect the generated means of sub-station flows. However, parameter uncertainty of the disaggregation model was visible when historical key-station flows were used as input variables in the disaggregation step as shown in the previous Figures, thus it is supposed that the natural uncertainty of input variables (annual flow generation) is more dominant over parameter uncertainty of the disaggregation model for generated means of sub-station flows. A comparison between C3 with C1 (or a comparison between C4 with C2) demonstrates that the increased variability of generated key-station

flows with parameter uncertainty incorporated visibly propagates into means of sub-station flows through the spatial disaggregation causing the increased variability of those means. Quantile plots of generated standard deviations and generated lag-1 serial correlations of sub-station flows show a similar pattern with generated means. As the sample size increases, uncertainty effect decreases, but an increased variability of standard deviations can still be found even when the sample size is equal to 200. Note that the historical lag-1 serial correlations were well reproduced through simple VS spatial disaggregation, even though they are not to be preserved by the model structure.

However, a different pattern can be found in generated cross correlations such that the parameter uncertainty effect on the increased variability of the spatial disaggregation model is greater than the parameter uncertainty of the key-station flow generation. (See Figures 4.6 and 4.7). Parameter uncertainty of the disaggregation model shows the negative bias of generated cross correlations even when the sample size is 100, and this will be reduced for a larger sample (see example of  $n=200$ ).

The quantile plots of storage capacities and drought indices for four different cases are illustrated in Figure 4.8 for sub-station1 with the demand level of 70%MAF. (For different sub-stations and demand levels, see Figures 4.A7 through 4.A11 in the Appendix.) Distributions of generated storage capacities seem close to those of critical drought magnitudes, and this is because those two statistics are based on total deficits of generated streamflows. Little difference of calculated quantiles of storage capacities and drought magnitudes between C1 and C2 (or between C3 and C4) is noted, while increased quantiles of those design variables in C3 compared with C1 (or C4 compared with C2) are notable. The parameter uncertainty effect in C3 and C4 could also be

expected for generated critical drought lengths, but quantiles of generated critical drought intensities are not significantly influenced by parameter uncertainties.

Again, as similar in generated means, standard deviations, and lag-1 correlations, parameter uncertainty effects on generated design variables regarding the key-station flows generation are dominant over those from the spatial disaggregation stage. The variability of generated storage capacities and drought indices is closely related with generated means, standard deviations, and lag-1 correlations of generated flows. That is, the larger variability of generated means, standard deviations, and lag-1 correlations of generated flows can make a considerable impact on the distributions of storage capacities, drought magnitudes and lengths. On the other hand, if other statistics which are more dependent on cross-correlations of several sites in a region were to be required for the design criteria in the multivariate sense, parameter uncertainty of the disaggregation model would be a significant factor affecting those statistics.

Different demand levels could be also associated with the parameter uncertainty effects which are shown in Figures 4.9 (and also Figures 4.A12 and 4.A13 in the Appendix). Furthermore, increased upper quantiles are expected for C4 when compared with C3, which could be explained by the combined effect of parameter uncertainty of the spatial disaggregation model with parameter uncertainty of the key-station flow generation. Table 4.2 provides the numerical comparison of the quantile estimates of generated storage capacities (see Table 4.A1 through 4.A3 for generated design variables). Overall, the parameter uncertainty effect of the key-station generation model is much more visible for upper quantiles, means, and standard deviations (coefficients of variations) of generated storage capacities. However, the effect of parameter uncertainty

of the spatial disaggregation models also seems notable when combined with the effect of the annual generation uncertainty. For example, 4.03, 4.31, 6.67, 7.46 of 99% quantiles of generated rescaled storage capacities (scaled by historical one) are provided for substation 1 for the sample size of 50, which correspond with a 7% (C2), 66% (C3), and 85% (C4) increases compared with C1. A significant difference between C3 and C4 may be reported when compared with C1 and C2, which is thought of as the aforementioned combined effects of parameter uncertainties from both key-station flow generation and spatial disaggregation. Besides, more remarkable increases of quantiles for higher quantiles are expected by parameter uncertainty; e.g., maximum estimates, which are 131% (C2), 201% (C3), and 473% (C4) increases compared with C1. In a practical sense, empirical quantiles estimates of design variables can be conceptually associated with reliability and 99% reliability corresponds to the 100 year return period magnitude. If one would need to design the reservoir based on storage capacity with more than 99% reliability, parameter uncertainty would be more effective on improving the precision. For sample size equal to 200, the parameter uncertainty effect of generated key-station flows for the upper quantiles of generated storage capacities is getting negligible, but it is still visible on maximum estimates. Similar patterns may be noted in the cases of generated critical drought magnitudes and lengths, but not much more visible than generated storage capacities. In the example of generated critical drought intensities, the parameter uncertainty effect is hard to judge, which might be related to the definition of the drought intensity (defined as the ratio of the critical drought magnitude to the critical drought length).

Table 4.3 gives a brief summary of the parameter uncertainty effect on synthetic

storage capacity and critical drought magnitude calculated from generated annual streamflows at sub-station 1 (Site 2). For simplicity, C1 where no parameter uncertainty is considered either in key-station generation or in spatial disaggregation and C4 where parameter uncertainty is incorporated in both annual generation and in temporal disaggregation are numerically compared. Compared with C1, C4 shows about a 20% increased expected value of storage capacity. Moreover, C4 demonstrates the increased standard deviations of storage capacities of about 85% and 115% for demand levels of 100% MAF and 80% MAF, respectively. The parameter uncertainty shows the increased quantiles with a range from about 32% to 83%, and more increased quantiles of storage capacity would be expected by parameter uncertainty especially for the larger quantile. Overall, less increase by the parameter uncertainty effect is demonstrated in the generated critical drought magnitude compared with the generated storage capacity, except that similar pattern of increased statistics and quantiles by parameter uncertainty over different demand levels and sample serial correlations are reported for both design variables.

#### **4.4.3 Parameter uncertainty effects based on Lane's condensed model**

Lane's condensed temporal disaggregation model (Lane, 1979) is additionally applied in the streamflow generation for the purpose of comparison with the VS model. Slightly different than Lane's original condensed model, an additional term is included so as to consider the parameter uncertainty of mean estimates, and the ML estimators are used for the parameter estimation. After site-by-site generations were initially

employed, adjustment procedures were taken for generated flows at each site to come up with the given key-site flows (exactly to match with the proportion of sub-station flows to key-station flows). Statistical properties are calculated based on generated flows and illustrated in Figure 4.10 (4.A15 through 4.A17). Compared with Figure 4.5 for the case of the VS model, Lane's condensed (LC) model produces similar variability of generated means, standard deviations, skewness coefficients, and lag-1 serial correlations. The LC model does not show any ability in preserving the historical cross correlation, which is due to the model structure. An additional term for the LC model or an adjustment to handle the discrepancy of cross correlation might be required. Basically in C1, generated storage capacities and critical drought magnitudes based on the LC model have less variability than the VS model, whereas the LC model shows greater variability of generated critical drought length (Figure 4.11 and Figures 4.A18-4.A19 in the Appendix). Correspondingly, the less variability of generated storage capacities and drought magnitudes could be expected in the LC model when parameter uncertainty is incorporated (C2 through C4).

## **4.5 Summary and Concluding Remarks**

Synthetic streamflows at multi-sites were generated by using the VS spatial disaggregation model with parameter uncertainty incorporated. Based on the multivariate regression concept among the key-station and sub-station flows, asymptotic distribution and posterior distribution of parameter estimates of the VS model were theoretically derived. Basic statistics, synthetic storage capacities and critical drought

indices were calculated from generated flows and compared with different sample sizes to evaluate related uncertainty effects. The conclusions are as follows:

Two approaches of asymptotic and Bayesian were able to explain the variability resulted from parameter uncertainty similarly in most sample statistics, an exception being cross-correlations. A little wider variability and upward bias of generated storage and drought related statistics were reported when using the Bayesian distribution for small sample size. The asymptotic and Bayesian approaches showed similar variability of generated means, standard deviations, and design variables once the sample size is larger than 50; little difference between the two approaches was expected for the sample size of 98. Thus, a sample size of at least 100 would be required if the asymptotic distribution approach is utilized for the parameter uncertainty consideration. That is, for the Bayesian approach, the effect of non-informative prior, which results in more variability of parameters than the exact prior, becomes negligible and sample information might be enough to interpret the variability of uncertain parameters.

The effect of parameter uncertainty of the disaggregation procedure was not as significant on the generated means, standard deviations, and lag-1 serial correlations as the natural uncertainty of given input variables. Likewise, the effect of the parameter uncertainty of the disaggregation model on the utilized storage capacity and critical drought indices was also not as significant as the natural uncertainty of input variables, which results from the fact that the variability of design variables of concern are closely related with that of means, variances, and serial correlations of generated streamflows. However, cross correlations of generated flows were much influenced by parameter uncertainty of the disaggregation. If any statistics, which are based on cross correlations,

are to be required in the practical design problem, parameter uncertainty of the disaggregation model would become more significant.

When the parameter uncertainty of input variables was incorporated into the disaggregation, increased variabilities of key-stations was shown to propagate into the variabilities of means, standard deviations, and lag-1 serial correlations of generated sub-station flows, more significantly for smaller sample sizes. Thus, uncertain parameters regarding input variables might cause important consequences on the related determination of reservoir size as illustrated by increased variabilities of storage capacities and critical drought indices. In this case, parameter uncertainty of the disaggregation model might result in an additional increase over those by parameter uncertainty regarding input variables for small sample size.

The parameter uncertainty still exhibits its effect on the variability of generated storage capacities and drought indices even when the sample size is equal to 100. A sample size of 100 does not seem large enough to neglect parameter uncertainty. However, the determination of what constitutes an adequate sample size depends on how much variability of design variables would be tolerated in the practical sense. In most applications for real streamflow generation, the available sample size is usually less than 100. Therefore, the incorporation of parameter uncertainty issues into the streamflow simulation is of importance, and thus the precision and reliability of applications of generated streamflows will be improved.

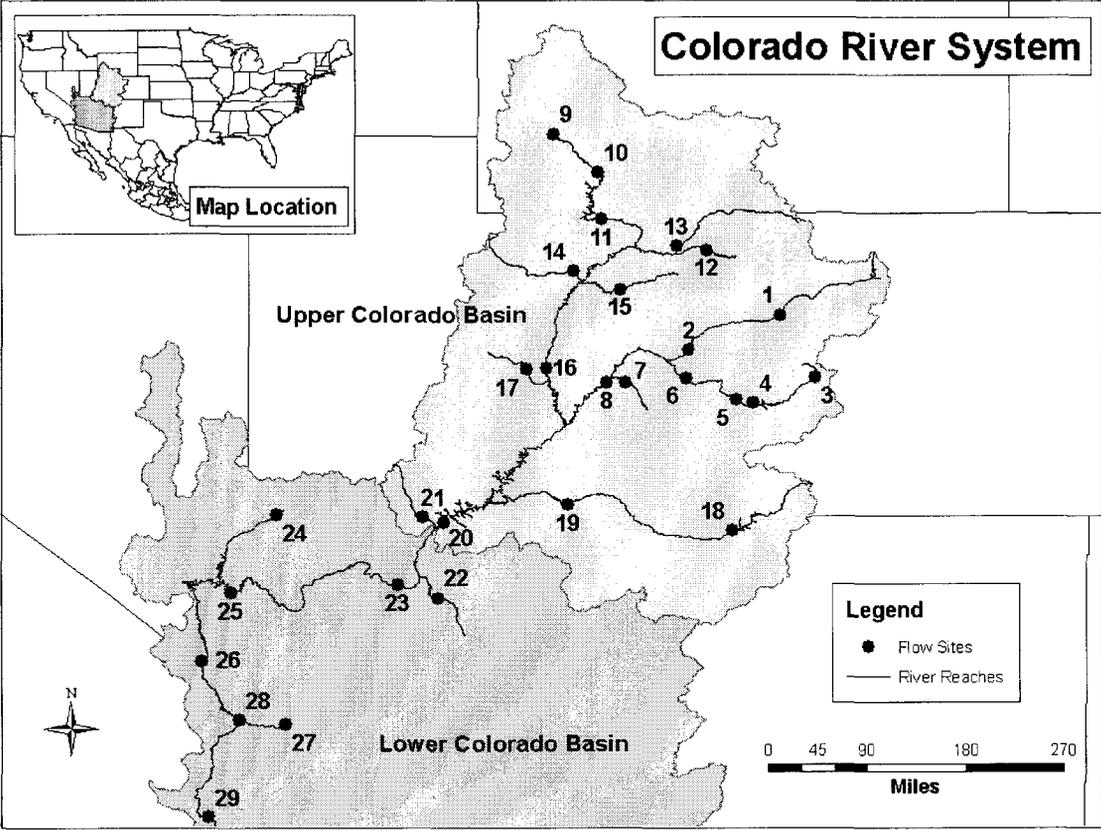


Figure 4.1: Colorado River Basin gauging stations (USBR, 2007)

Table 4.1: Basic statistics of historical streamflows at chosen key-station and sub-stations in Colorado River Basin

	Key(site 8)	Sub1(site 2)	Sub2(site 6)	Sub3(site 7)
Mean (ac-ft)	6824630	3580294	2355220	813287
SD (ac-ft)	1965948	934273	724565	360718
Coeff. of Variation	0.29	0.26	0.31	0.44
Skewness	0.20	0.25	0.14	0.44
Lag-1 corr.	0.29	0.26	0.26	0.23
Proportion to Key-station		52.5%	34.5%	11.9%
Cross correlation	Key(site 8)	Sub1(site 2)	Sub2(site 6)	Sub3(site 7)
Key(site 8)	1.00	0.95	0.98	0.84
Sub1(site 2)	0.95	1.00	0.91	0.67
Sub2(site 6)	0.98	0.91	1.00	0.85
Sub3(site 7)	0.84	0.67	0.85	1.00

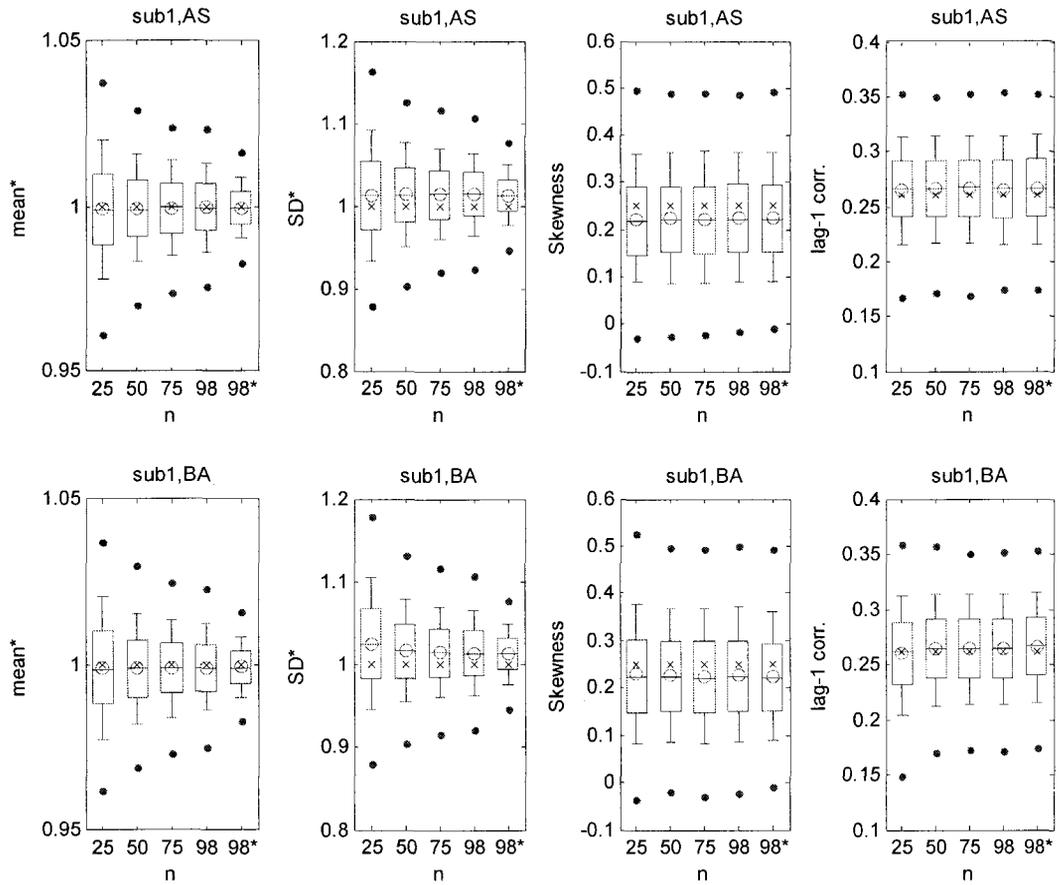
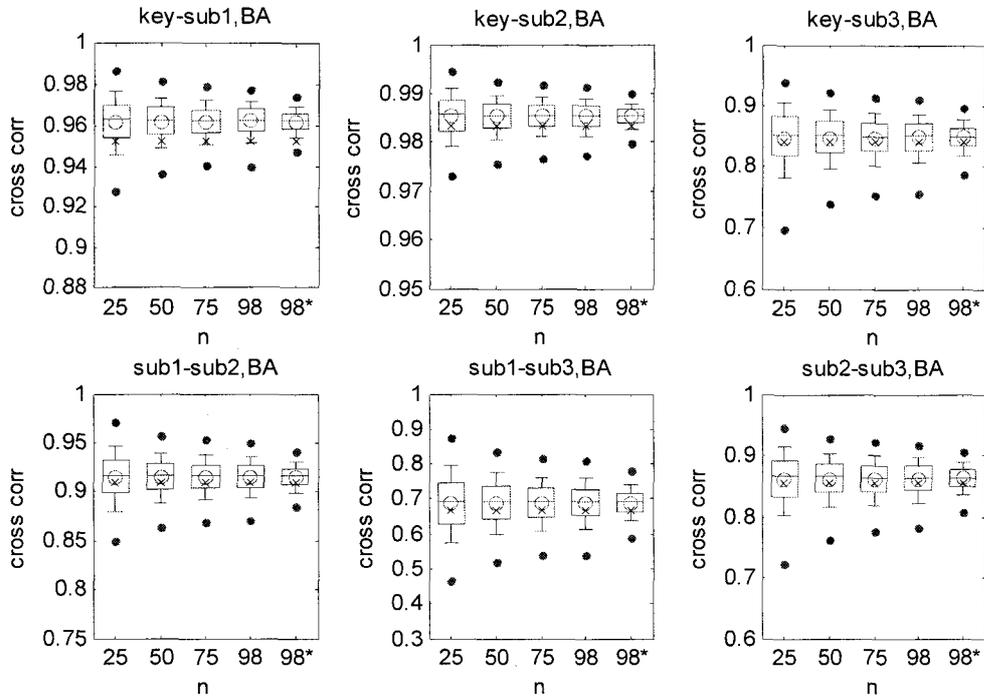
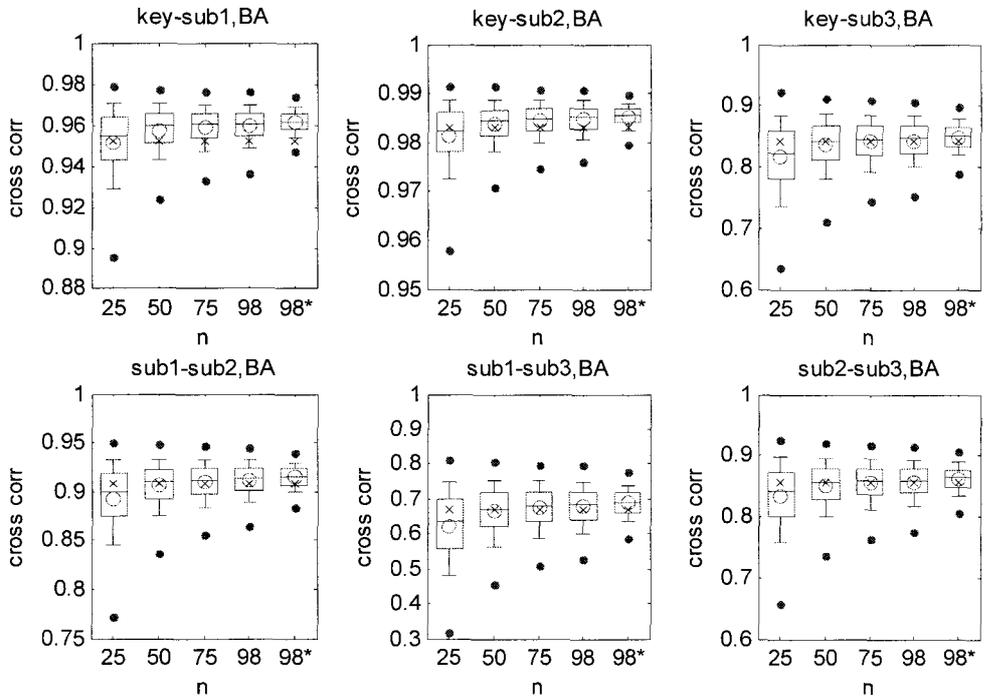


Figure 4.2: Quantile plots of basic statistics calculated from synthetic flows at sub-station 1(site 2) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.



(1) Asymptotic approach



(2) Bayesian approach

Figure 4.3: Quantile plots of calculated cross correlations between historical key-station flows and synthetic sub-station flows generated from historical key-station using VS disaggregation model with parameter uncertainty incorporated where AS means asymptotic approach, BA means Bayesian approach, 'X' denotes the historical cross correlation, sub1 means sub-station 1 (site 2), sub2 means sub-station 2 (site 6), and sub3 means sub-station 3 (site 7).

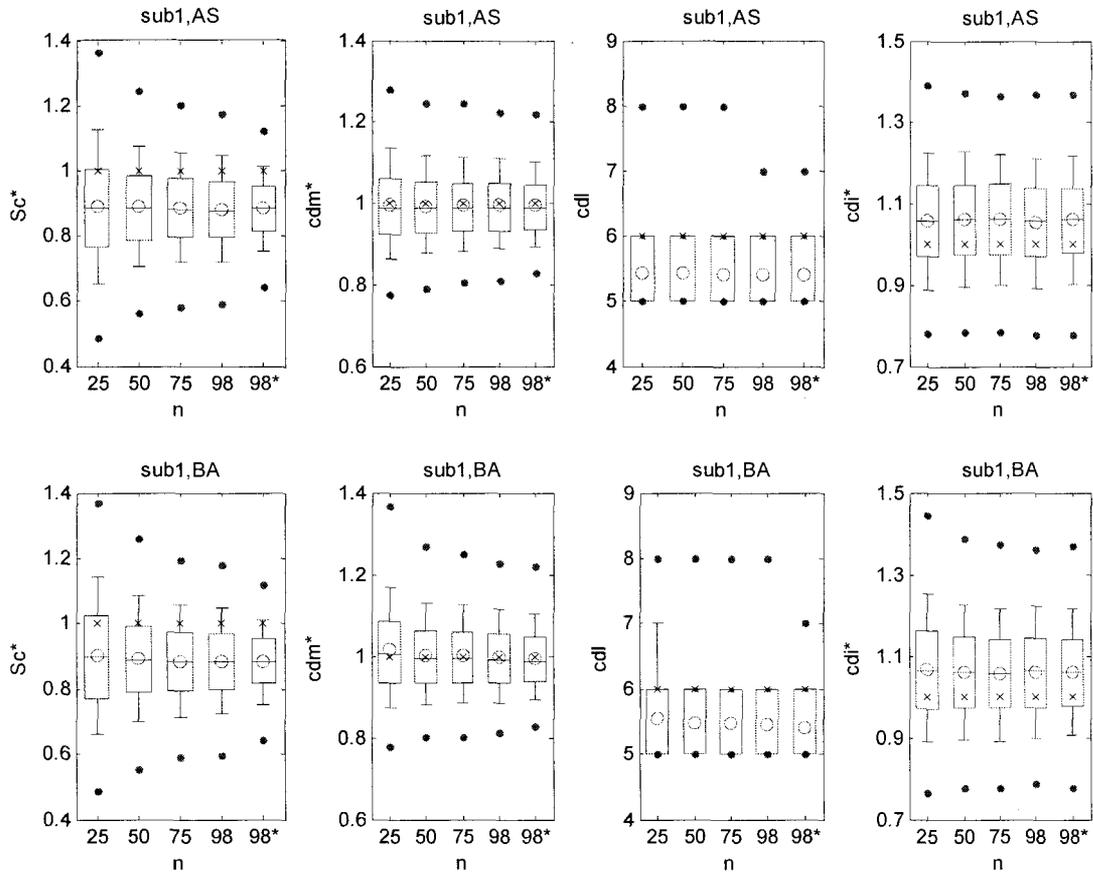


Figure 4.4: Quantile plots of storage capacities ( $Sc$ ), critical drought magnitudes ( $cdm$ ), critical drought lengths ( $cdl$ ), and critical drought intensities ( $cdi$ ) calculated from synthetic flows at sub-station 1(site 2) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics. The demand level was assumed as 100% MAF at each site.

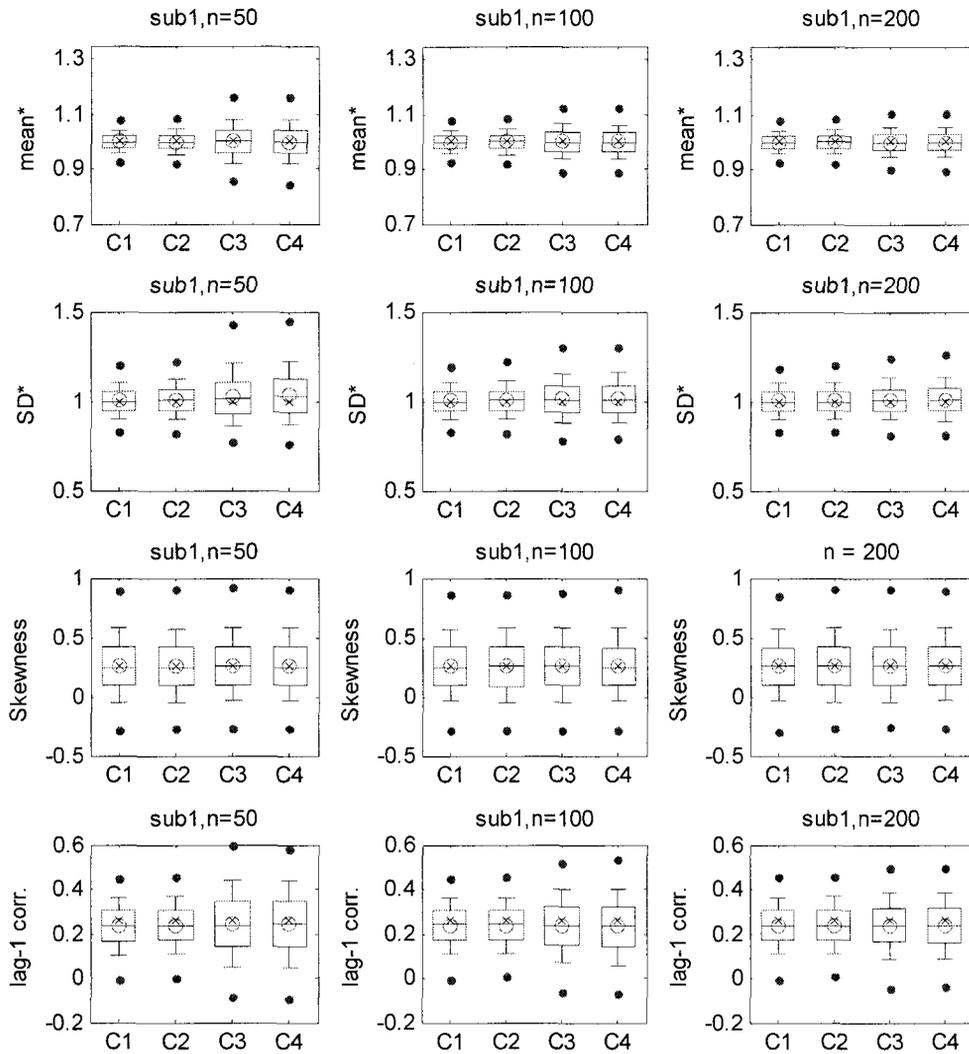


Figure 4.5: Quantile plots of basic statistics calculated from synthetic flows at sub-station 1(site 2) disaggregated (VS model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

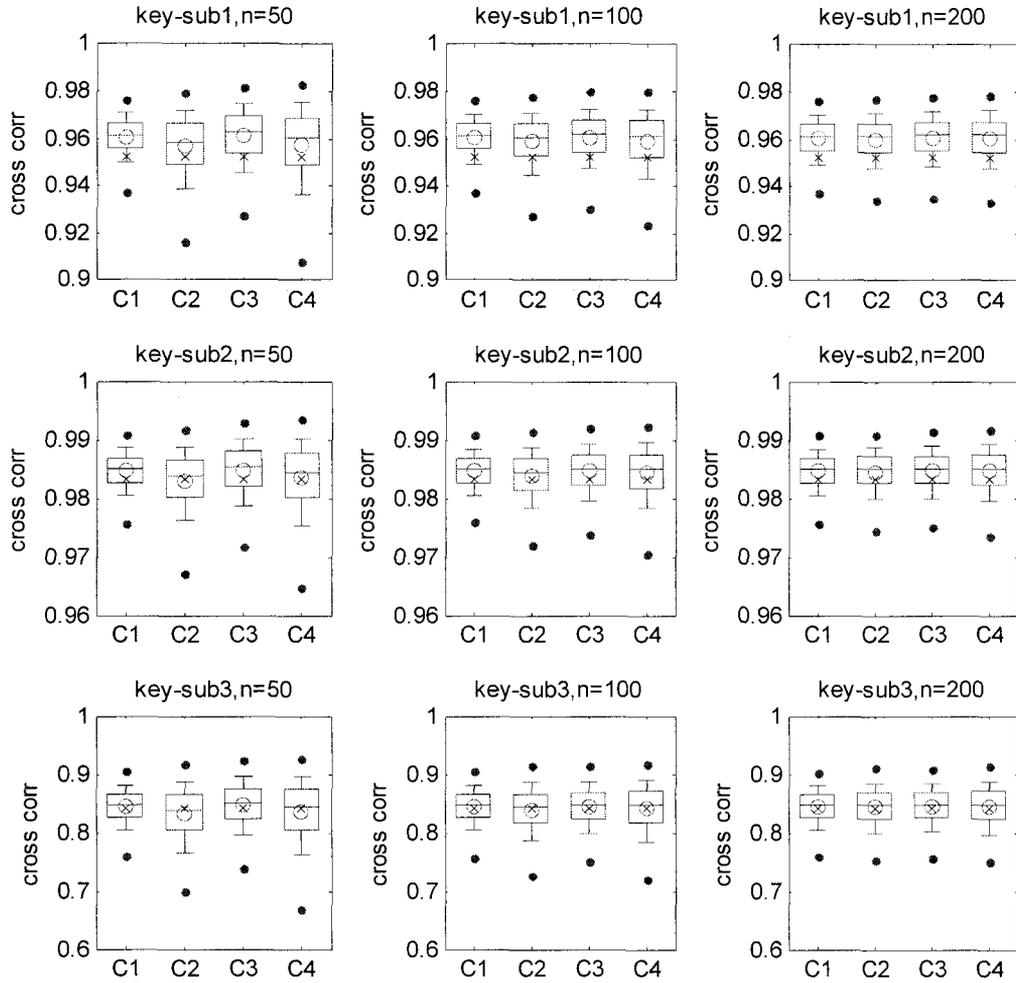


Figure 4.6: Quantile plots of calculated cross correlations between generated key-station flows (AR(1) model) and disaggregated sub-station flows (VS model) with parameter uncertainty incorporated (Bayesian approach), where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size, sub1 means sub-station 1(site 2), sub2 means sub-station 2 (site 6), and sub3 means sub-station 3 (site 7).

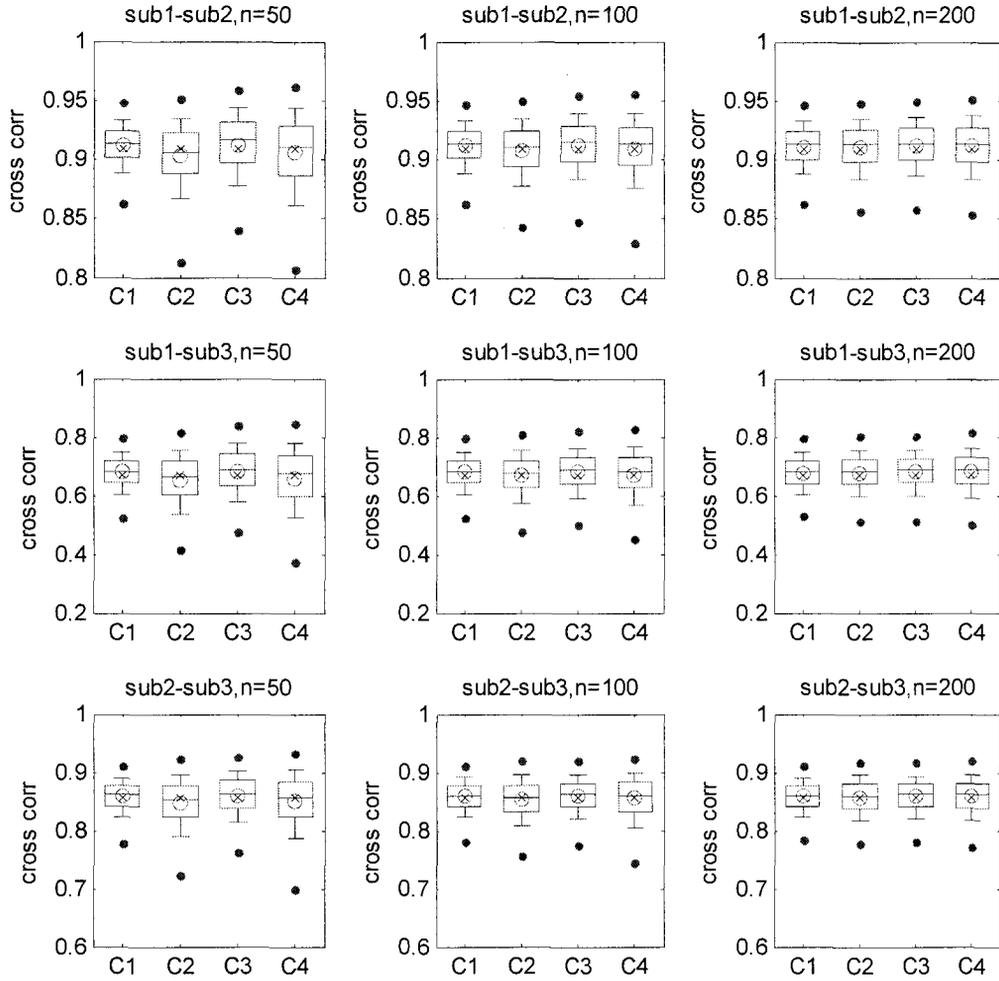


Figure 4.7: Quantile plots of calculated cross correlations between sub-station flows disaggregated (VS model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach), where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size, sub1 means sub-station 1(site 2), sub2 means sub-station 2 (site 6), and sub3 means sub-station 3 (site 7).

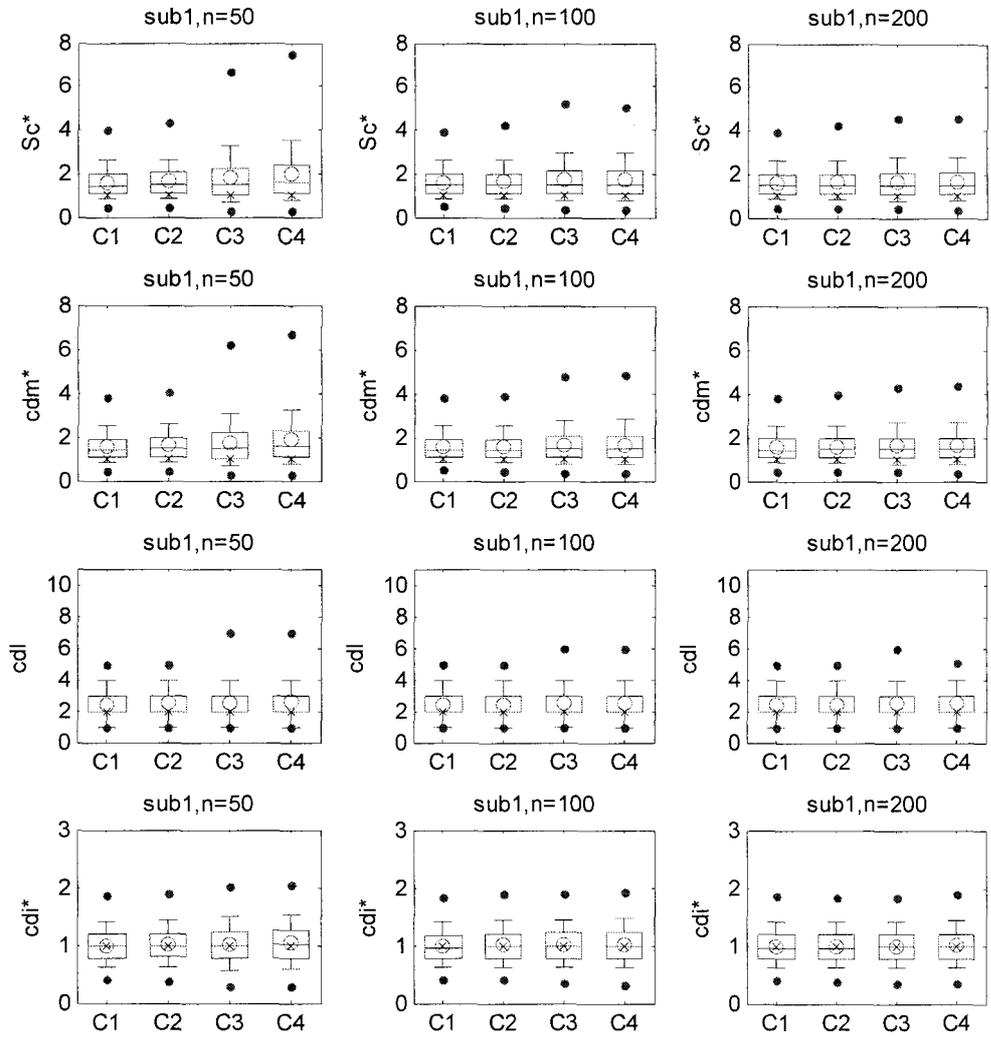


Figure 4.8: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 1(site 2) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

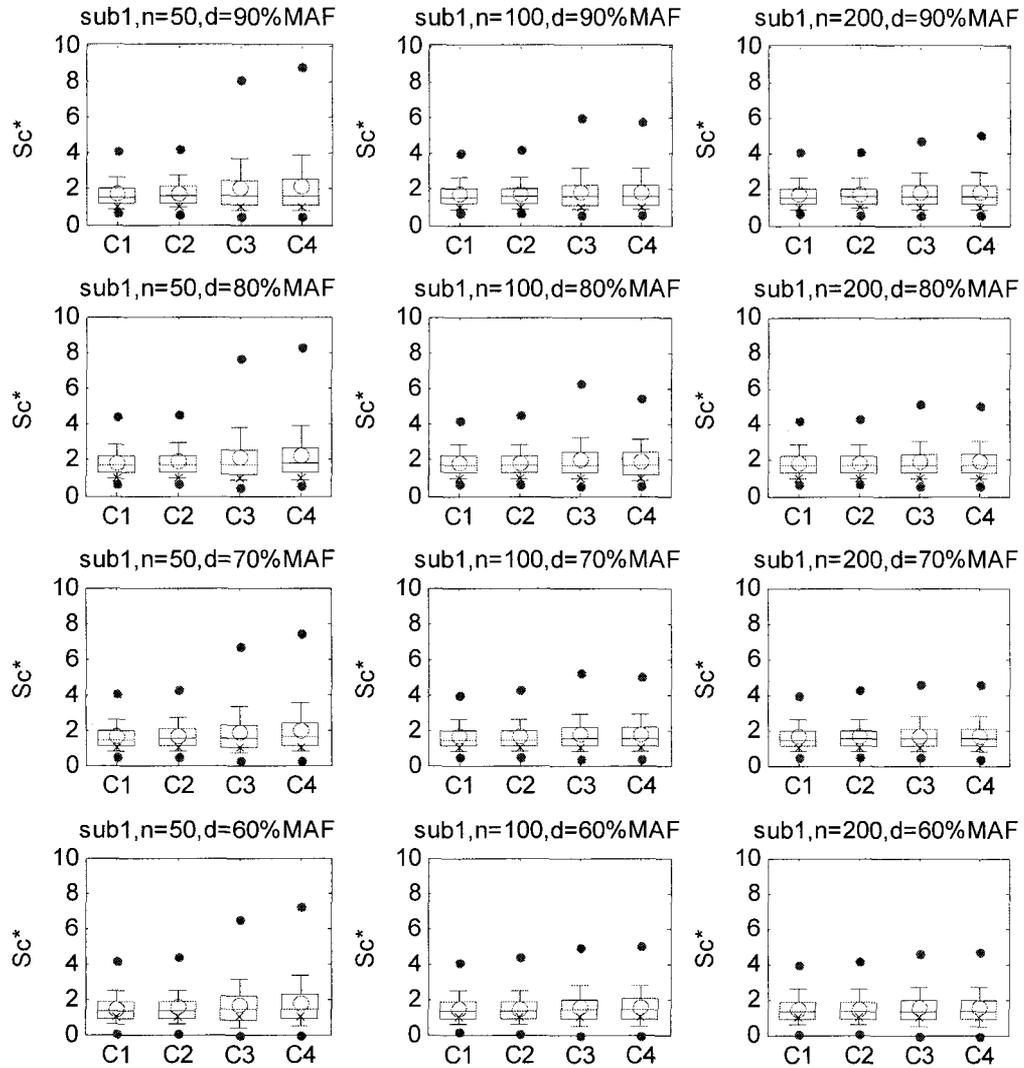


Figure 4.9: Quantile plots of storage capacity ( $Sc$ ) calculated from synthetic streamflows at sub-station 1(site 2) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) for different demand levels. C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size.

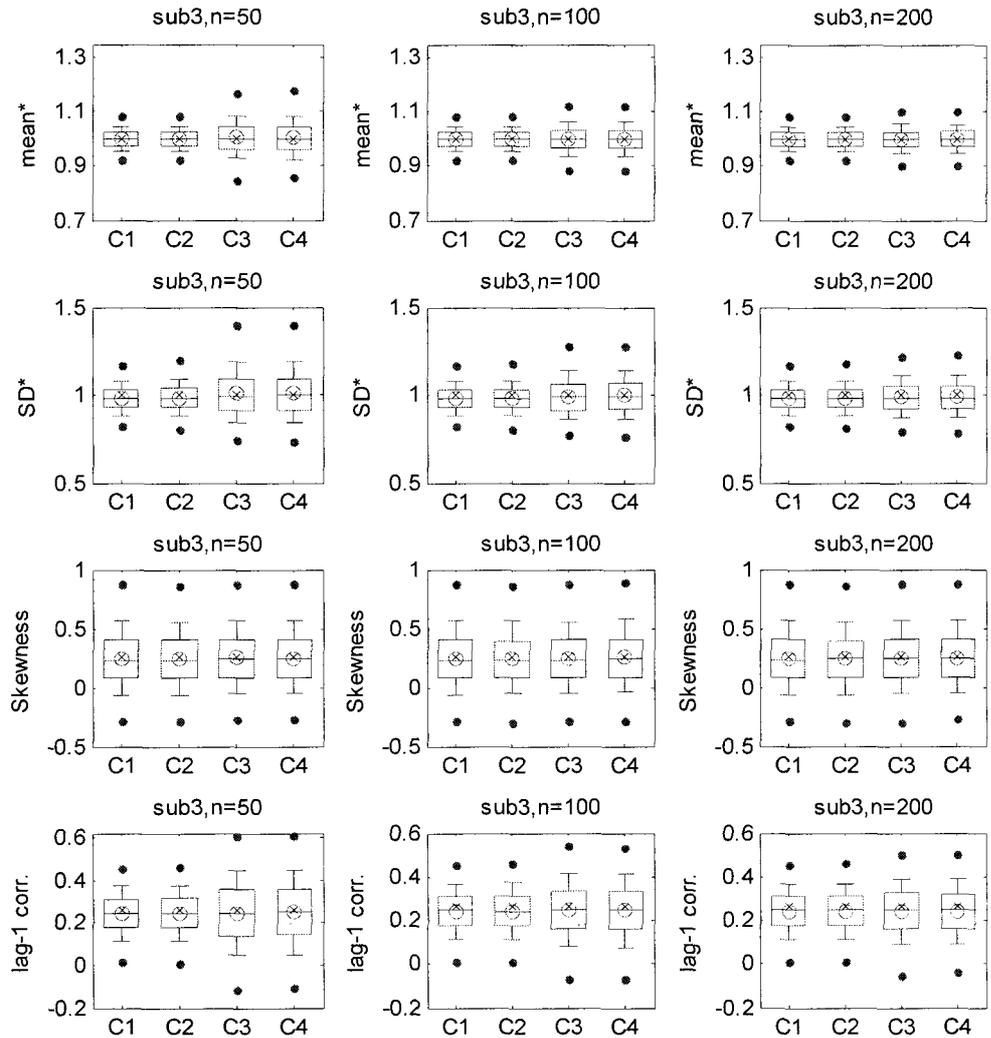


Figure 4.10: Quantile plots of basic statistics calculated from synthetic flows at substation 1(site 2) disaggregated (LC model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4,respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

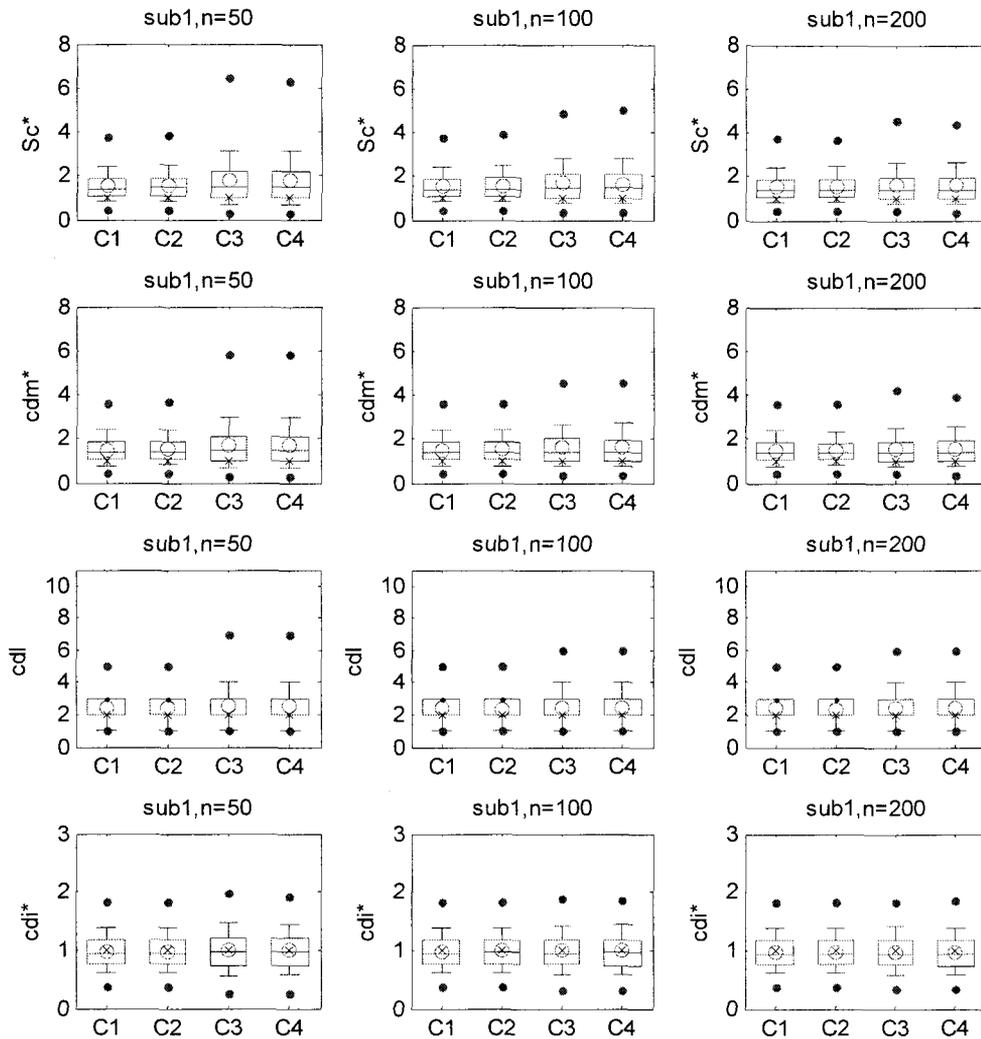


Figure 4.11: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 1(site 2) which are disaggregated (LC model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

Table 4.2: Quantile estimates of generated storage capacities (rescaled by dividing the historical storage capacity) calculated from generated sub-station flows disaggregated from generated annual flows by using AR(1) and VS disaggregation models for 4 different uncertainty considerations. (Bayesian approach, demand level = 70% MAF)

sub-station 1	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.20	0.07	0.00	0.00	0.19	0.00	0.01	0.01
1% quantile	0.48	0.47	0.29	0.31	0.48	0.46	0.42	0.39
10% quantile	0.83	0.84	0.70	0.75	0.83	0.82	0.80	0.80
25% quantile	1.08	1.10	1.03	1.08	1.08	1.10	1.08	1.09
Median	1.45	1.49	1.52	1.59	1.46	1.48	1.48	1.50
75% quantile	1.98	2.02	2.25	2.39	1.99	1.97	2.08	2.06
90% quantile	2.61	2.66	3.31	3.48	2.60	2.62	2.77	2.82
99% quantile	4.03	4.31	6.67	7.46	3.94	4.28	4.58	4.56
Maximum	6.23	8.16	12.54	29.44	6.38	7.70	7.09	9.81
Mean	1.61	1.65	1.83	1.96	1.61	1.63	1.68	1.68
Standard Deviation	0.74	0.79	1.24	1.52	0.74	0.77	0.85	0.88
Coeff. of variation	0.46	0.48	0.68	0.77	0.46	0.47	0.51	0.52
sub-station 2	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.20	0.19	0.00	0.00	0.31	0.23	0.24	0.03
1% quantile	0.57	0.57	0.38	0.42	0.58	0.57	0.54	0.53
10% quantile	0.92	0.91	0.78	0.82	0.91	0.92	0.91	0.91
25% quantile	1.18	1.19	1.11	1.16	1.19	1.19	1.19	1.18
Median	1.58	1.60	1.64	1.71	1.59	1.60	1.62	1.60
75% quantile	2.12	2.14	2.44	2.52	2.11	2.12	2.22	2.22
90% quantile	2.72	2.75	3.57	3.77	2.72	2.76	2.94	3.01
99% quantile	4.06	4.30	7.08	7.62	4.11	4.24	4.83	4.90
Maximum	6.58	6.97	14.02	30.63	6.15	7.20	7.65	10.60
Mean	1.72	1.74	1.98	2.09	1.73	1.74	1.81	1.82
Standard Deviation	0.74	0.77	1.34	1.59	0.75	0.77	0.89	0.93
Coeff. of variation	0.43	0.44	0.68	0.76	0.43	0.44	0.49	0.51
sub-station 3	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.50	0.35	0.22	0.10	0.37	0.38	0.26	0.26
1% quantile	0.62	0.58	0.52	0.50	0.63	0.63	0.61	0.60
10% quantile	0.89	0.86	0.80	0.81	0.90	0.89	0.88	0.87
25% quantile	1.12	1.11	1.06	1.08	1.10	1.12	1.12	1.10
Median	1.42	1.46	1.47	1.51	1.42	1.44	1.45	1.45
75% quantile	1.84	1.91	2.11	2.17	1.83	1.87	1.90	1.93
90% quantile	2.31	2.41	2.91	3.04	2.28	2.40	2.50	2.49
99% quantile	3.31	3.65	5.34	5.89	3.38	3.63	3.86	4.05
Maximum	4.84	7.19	8.68	20.24	4.39	7.22	5.97	6.82
Mean	1.53	1.58	1.72	1.80	1.53	1.56	1.60	1.60
Standard Deviation	0.58	0.66	0.98	1.18	0.58	0.64	0.69	0.71
Coeff. of variation	0.38	0.42	0.57	0.66	0.38	0.41	0.43	0.45

Table 4.3: Example of generated storage capacity and critical drought magnitude (scaled by MAF) ( $n=50$ ,  $N_d=98$ , Bayesian analysis, sub-station 1 (Site 2))

Demand level	MMF		0.8MMF	
Case	C1	C4	C1	C4
Storage capacity				
mean	3.8	4.6	0.7	0.9
		21.0%		18.7%
SD	1.8	3.4	0.3	0.7
		85.3%		115.3%
$q_{0.9}$	6.3	9.1	1.2	1.5
		45.4%		32.3%
$q_{0.95}$	7.3	11.2	1.4	2.0
		52.3%		42.8%
$q_{0.99}$	9.2	15.9	1.8	3.3
		73.8%		83.1%
Critical drought magnitude				
mean	1.9	2.1	0.7	0.8
		10.0%		13.6%
SD	0.7	1.1	0.3	0.5
		65.5%		67.3%
$q_{0.9}$	2.8	3.4	1.0	1.3
		21.8%		24.8%
$q_{0.95}$	3.2	4.1	1.2	1.7
		29.8%		37.7%
$q_{0.99}$	4.2	6.4	1.6	2.5
		32.6%		25.6%

*Note* C1: no parameter uncertainty considered (natural uncertainty), C4: parameter uncertainty incorporated both in key-station flow generation and in spatial disaggregation, SD: standard deviation,  $q_{0.9}$ ,  $q_{0.95}$ ,  $q_{0.99}$  means 90%, 95%, and 99% quantile, respectively. Value (%) in the column of C4 represents the ratio of increased storage capacity(or critical drought magnitude) in C4 with respect to C1. (parameter uncertainty effect)

## References

- Box, G.E.P. and G.C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc., N.Y.
- Grygier, J.C. and J.R. Stedinger (1988). Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research*, 24(10), pp. 1574-1584.
- Grygier, J.C. and J.R. Stedinger (1990). *SPIGOT, A synthetic Streamflow Generation Software Package, technical description, version 2.5*, School of Civil and Environmental Engineering, Cornell University, Ithaca, N. Y.
- Jeffreys, H. (1961). *Theory of Probability*, 3<sup>rd</sup> Edition, Oxford; Clarendon.
- Kollo, T. and Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Springer.
- Mejia, J.M. and J. Roussell (1976). Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), pp. 185-186.
- Lane, W.L. (1979). *Applied Stochastic Techniques, User's Manual*, Bureau of Reclamation, Engineering and Research Center, Denver, Co.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.
- Salas, J.D., N. Saada, C.H. Chung, W.L. Lane, and D.K. Frevert (2000). Stochastic Analysis, Modeling, and Simulation (SAMS) version 2000 User's manual, *Technical report No. 10*, Computing Hydrology Laboratory, Water Resources, Hydrologic and Environmental Sciences, Engineering Research Center, Fort Collins, Colorado.
- Stedinger, J.R. and R.M. Vogel (1984). Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(1), pp. 47-56.
- Stedinger, J. R., D. Pei, and T. Cohn (1985b). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resources Research*, 21(5), pp. 665-675.
- USBR, Department of Interior (2007). *Development of Stochastic Hydrology for the Colorado River System*.
- Valencia, D. and J.C. Schaake, Jr. (1973). Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), pp. 580-585, 1973.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population, *Biometrika*, 20A, pp. 32-52.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., New York.

## Appendix 4.A: Additional Figures and Tables

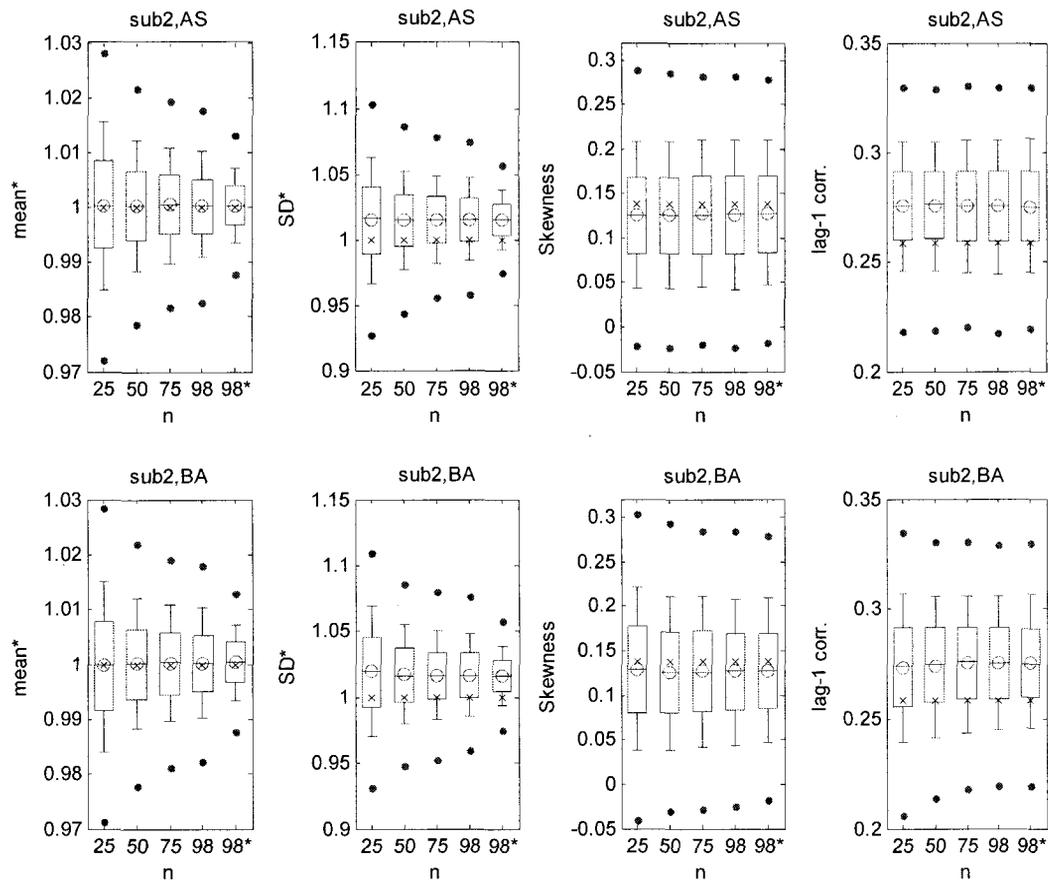


Figure 4.A1: Quantile plots of basic statistics calculated from synthetic flows at substation 2(site 6) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

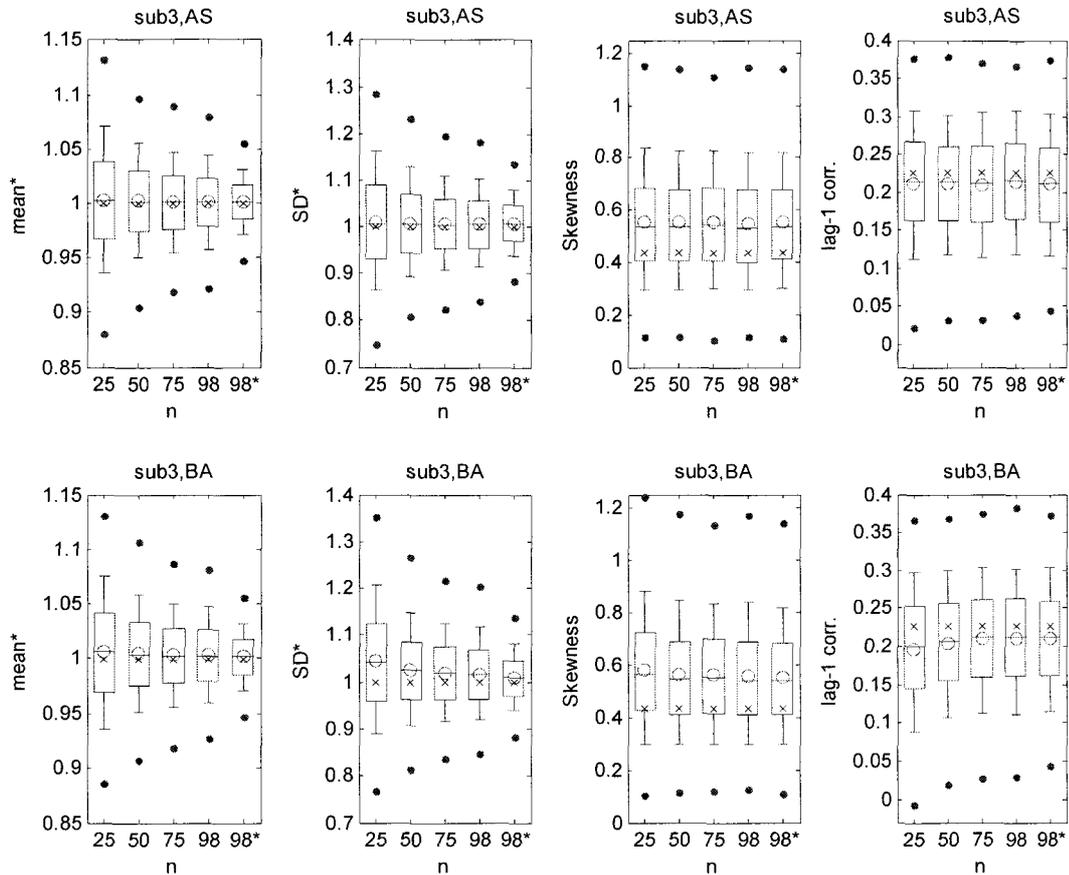


Figure 4.A2: Quantile plots of basic statistics calculated from synthetic flows at substation 3(site 7) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

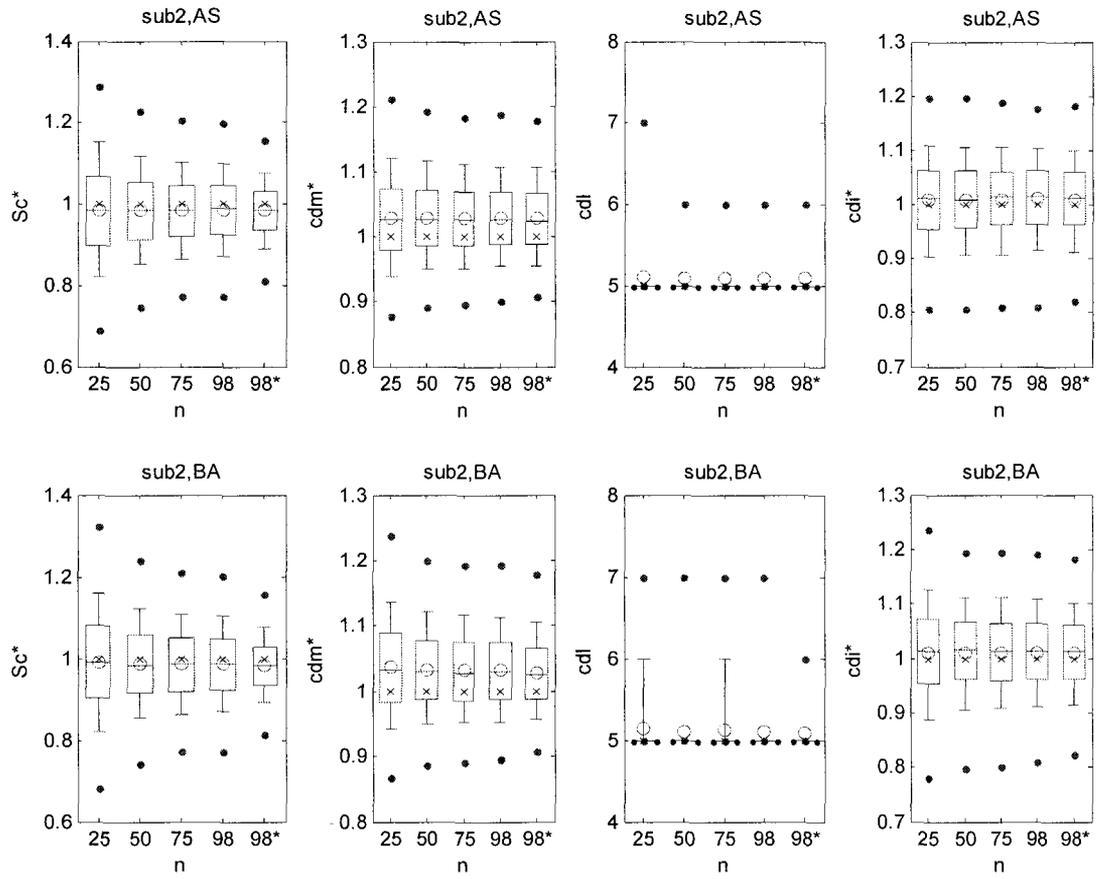


Figure 4.A3: Quantile plots of storage capacities ( $Sc^*$ ), critical drought magnitudes ( $cdm^*$ ), critical drought lengths ( $cdl$ ), and critical drought intensities ( $cdi^*$ ) calculated from synthetic flows at sub-station 2(site 6) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics. The demand level was assumed as 100% MAF at each site.

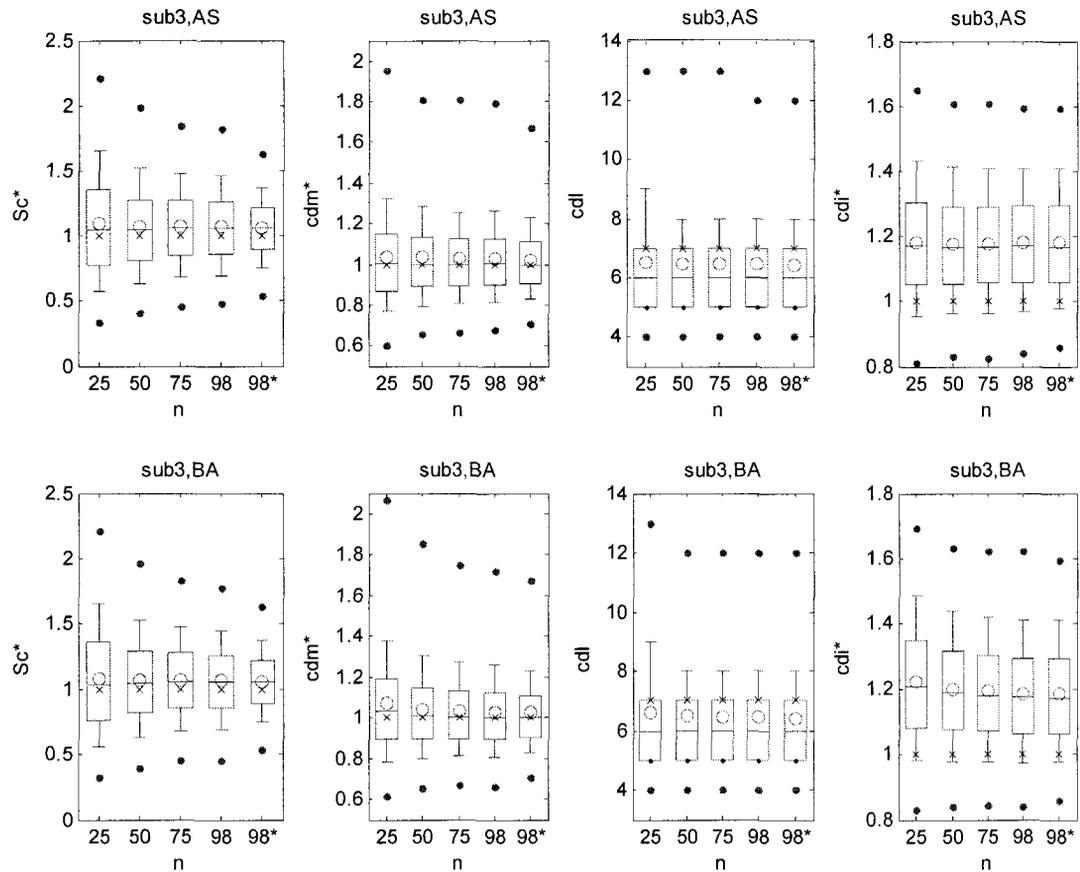


Figure 4.A4: Quantile plots of storage capacities ( $Sc$ ), critical drought magnitudes ( $cdm$ ), critical drought lengths ( $cdl$ ), and critical drought intensities ( $cdi$ ) calculated from synthetic flows at sub-station 3(site 7) generated from historical key-station flows using VS disaggregation model with parameter uncertainty incorporated where the superscript \* means “scaled by historical statistics”, AS means asymptotic approach, and BA means Bayesian approach. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics. The demand level was assumed as 100% MAF at each site.

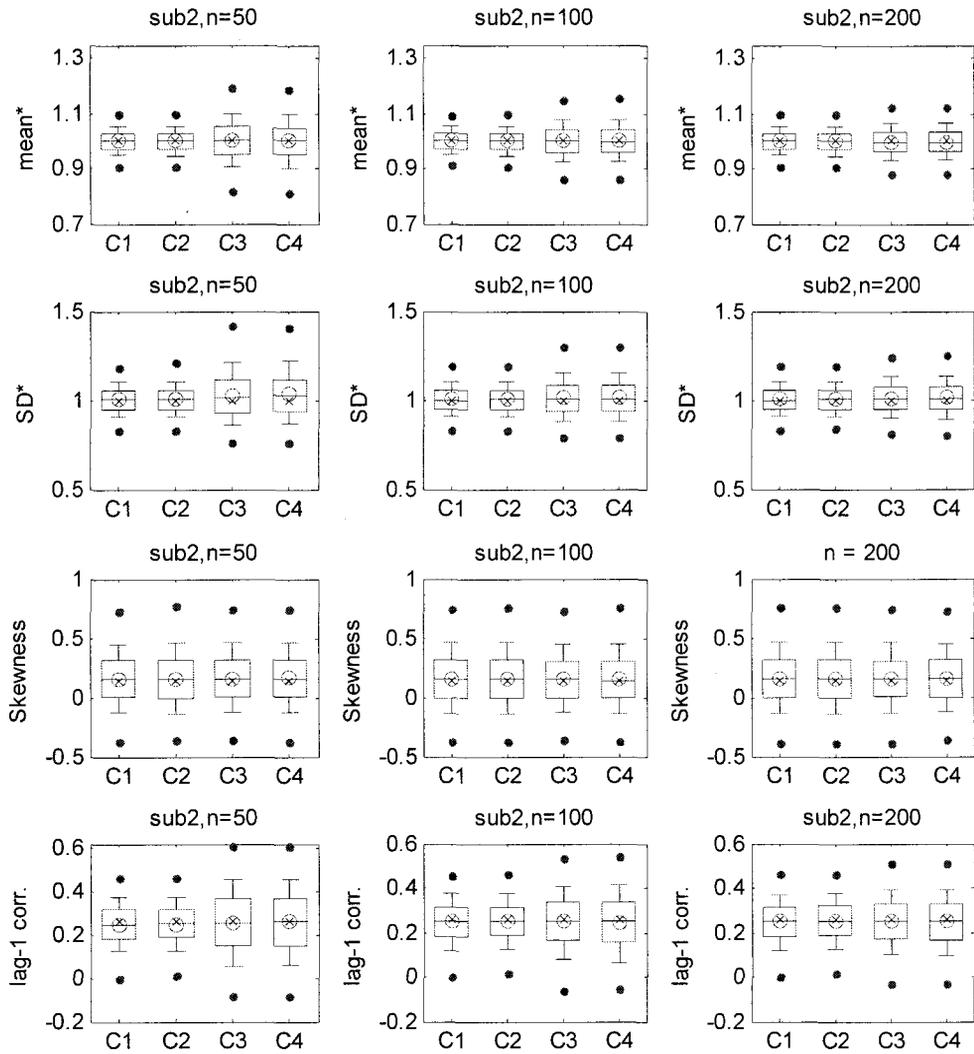


Figure 4.A5: Quantile plots of basic statistics calculated from synthetic flows at substation 2(site 6) disaggregated (VS model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4,respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

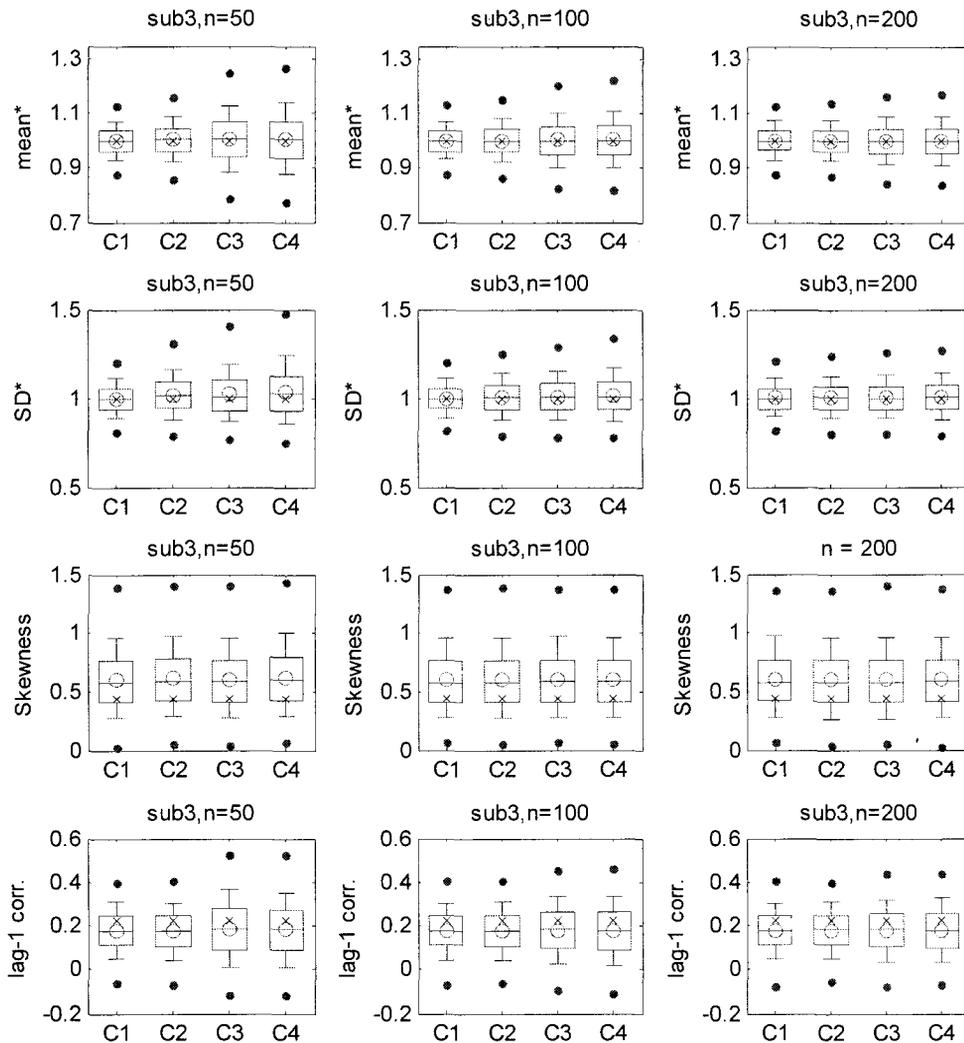


Figure 4.A6: Quantile plots of basic statistics calculated from synthetic flows at substation 3(site 7) disaggregated (VS model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4,respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

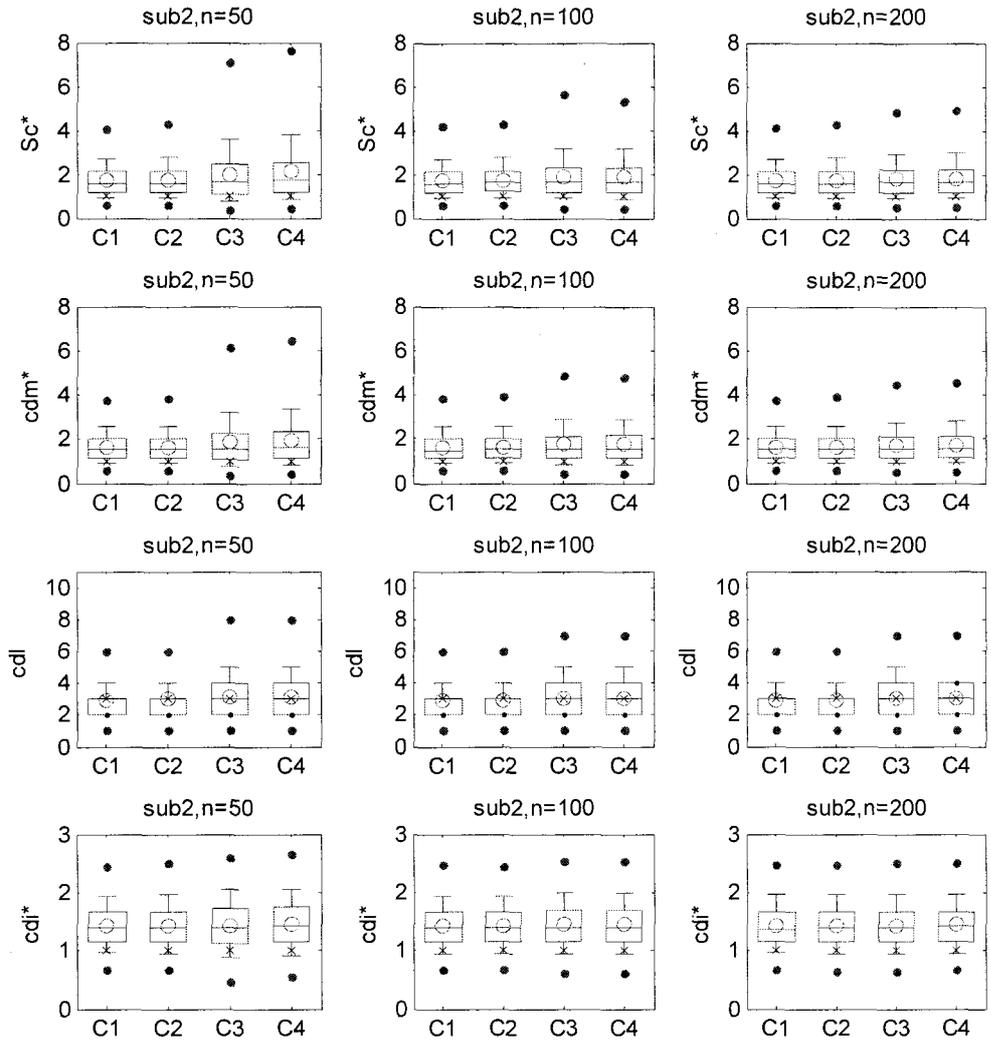


Figure 4.A7: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 2(site 6) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

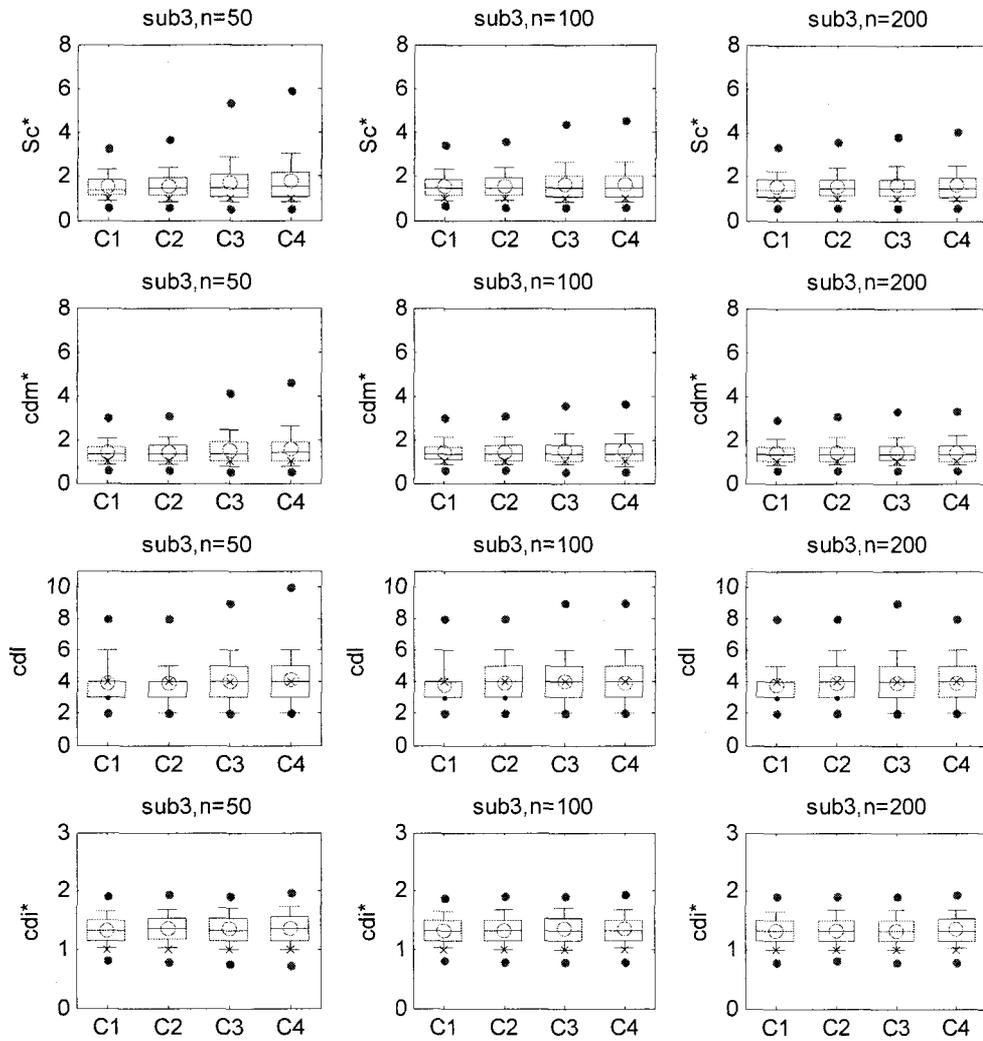


Figure 4.A8: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 3(site 7) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

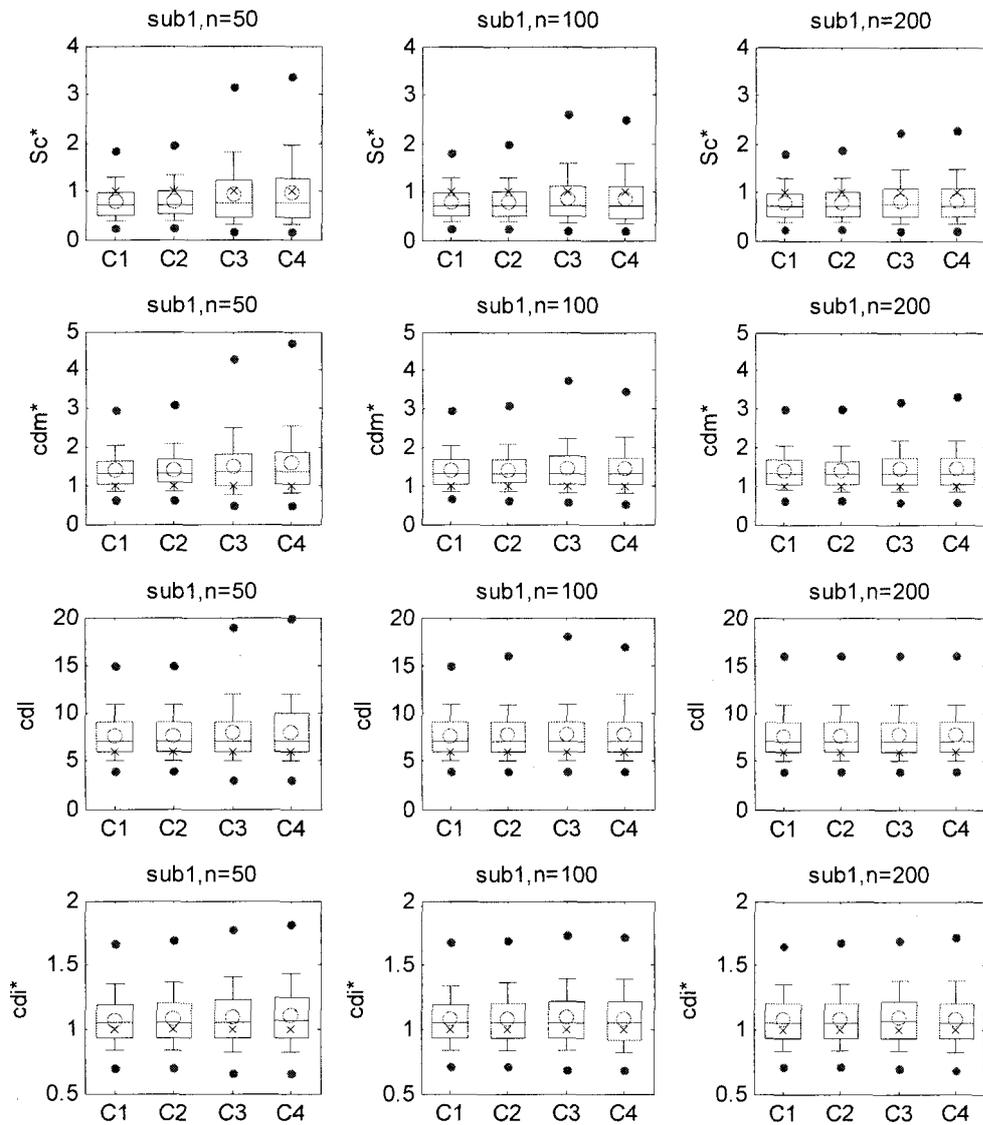


Figure 4.A9: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 1(site 2) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 100% MAF at each site..

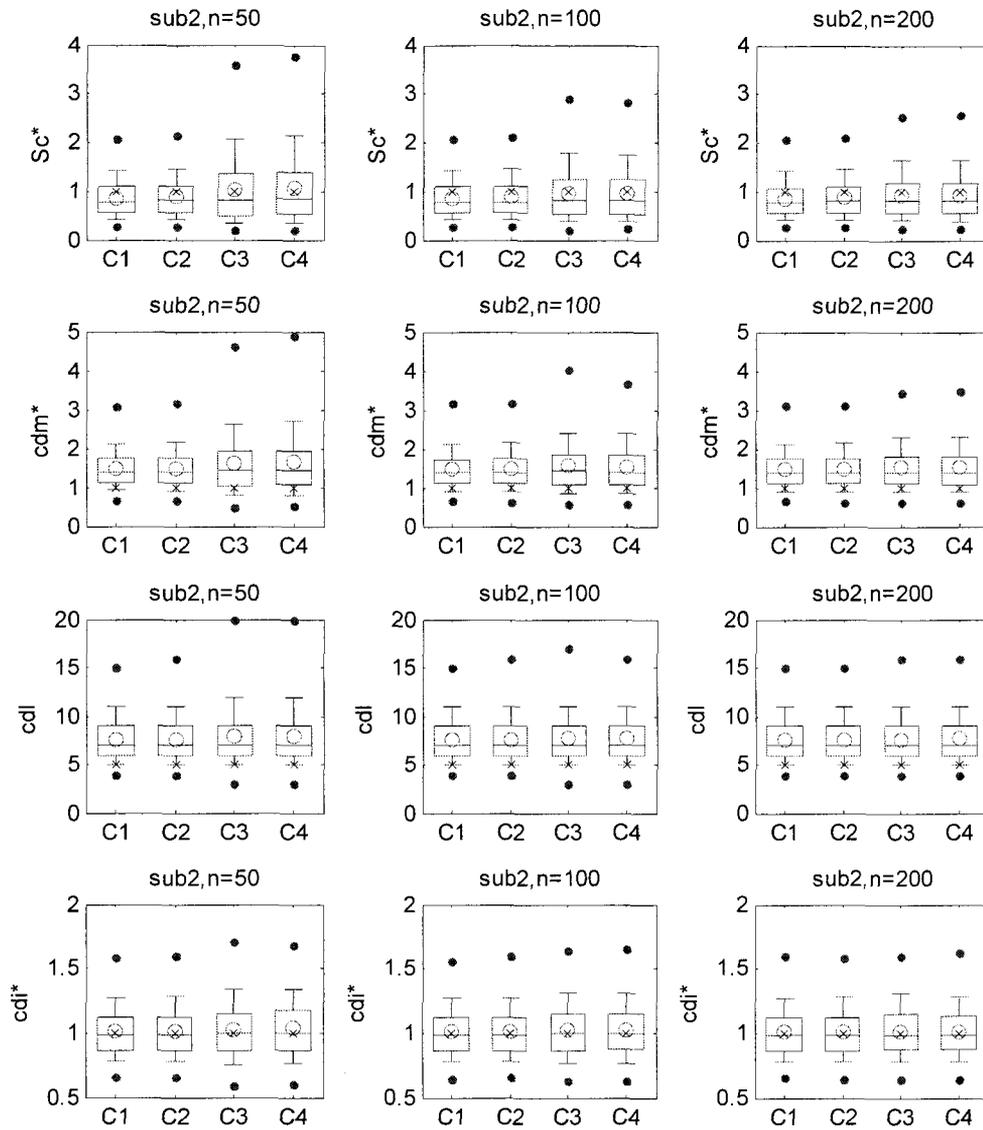


Figure 4.A10: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 2(site 6) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 100% MAF at each site.

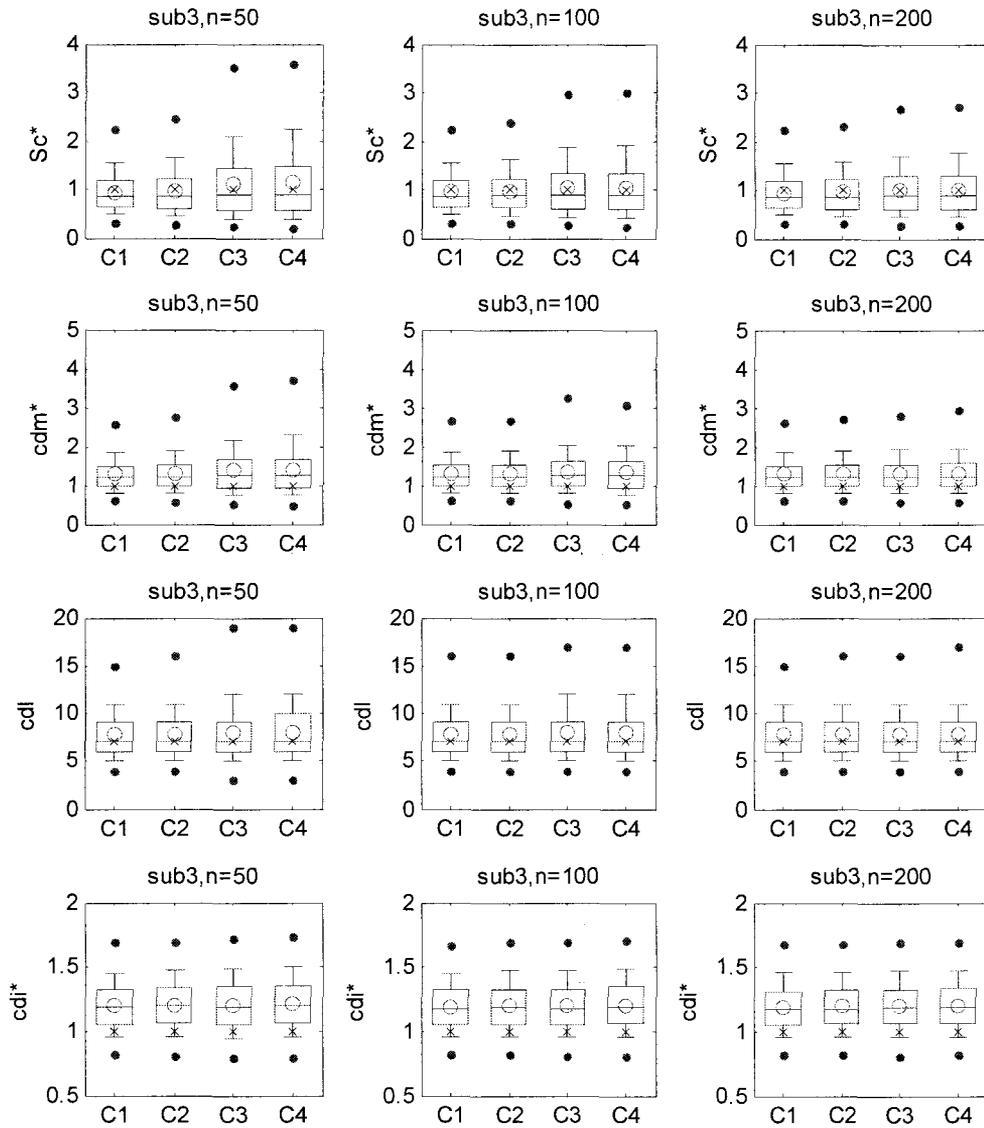


Figure 4.A11: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 3(site 7) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 100% MAF at each site.

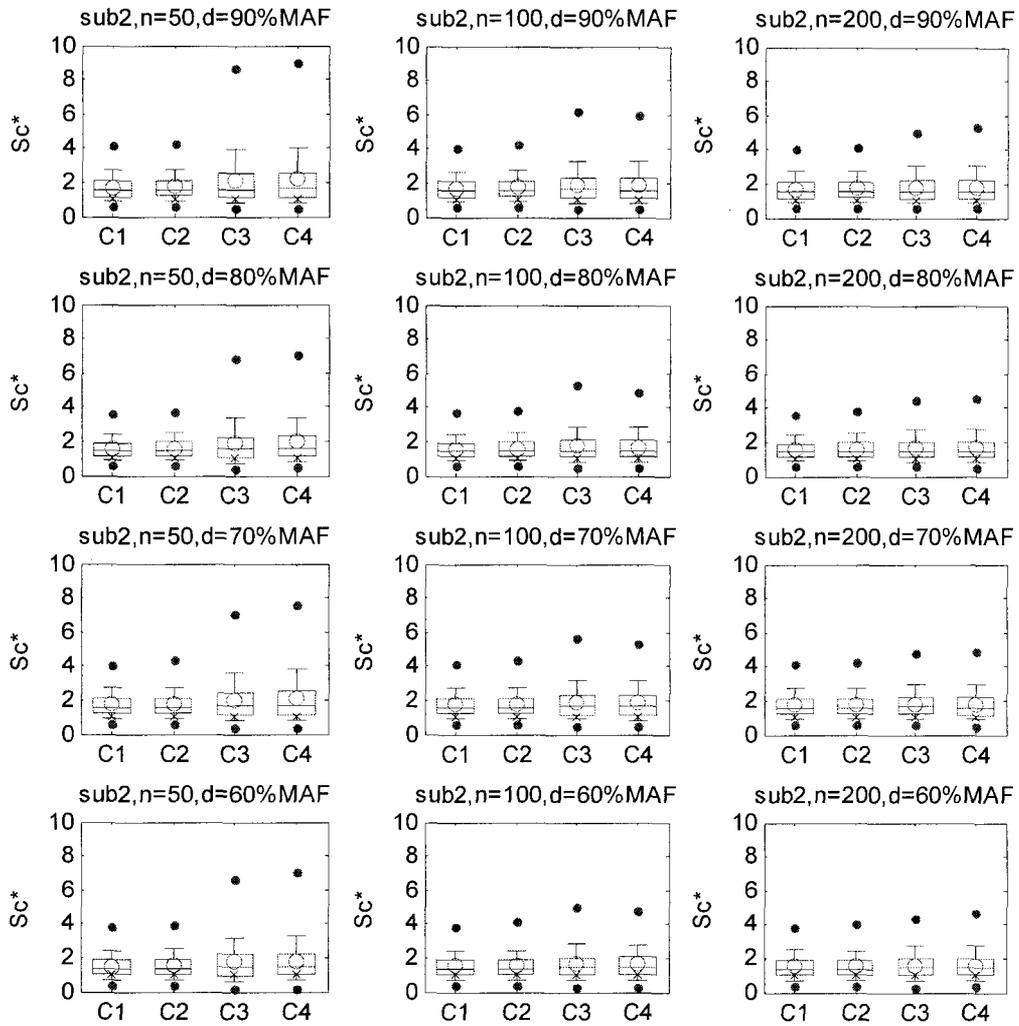


Figure 4.A12: Quantile plots of storage capacity ( $Sc$ ) calculated from synthetic streamflows at sub-station 2(site 6) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) for different demand levels. C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size.

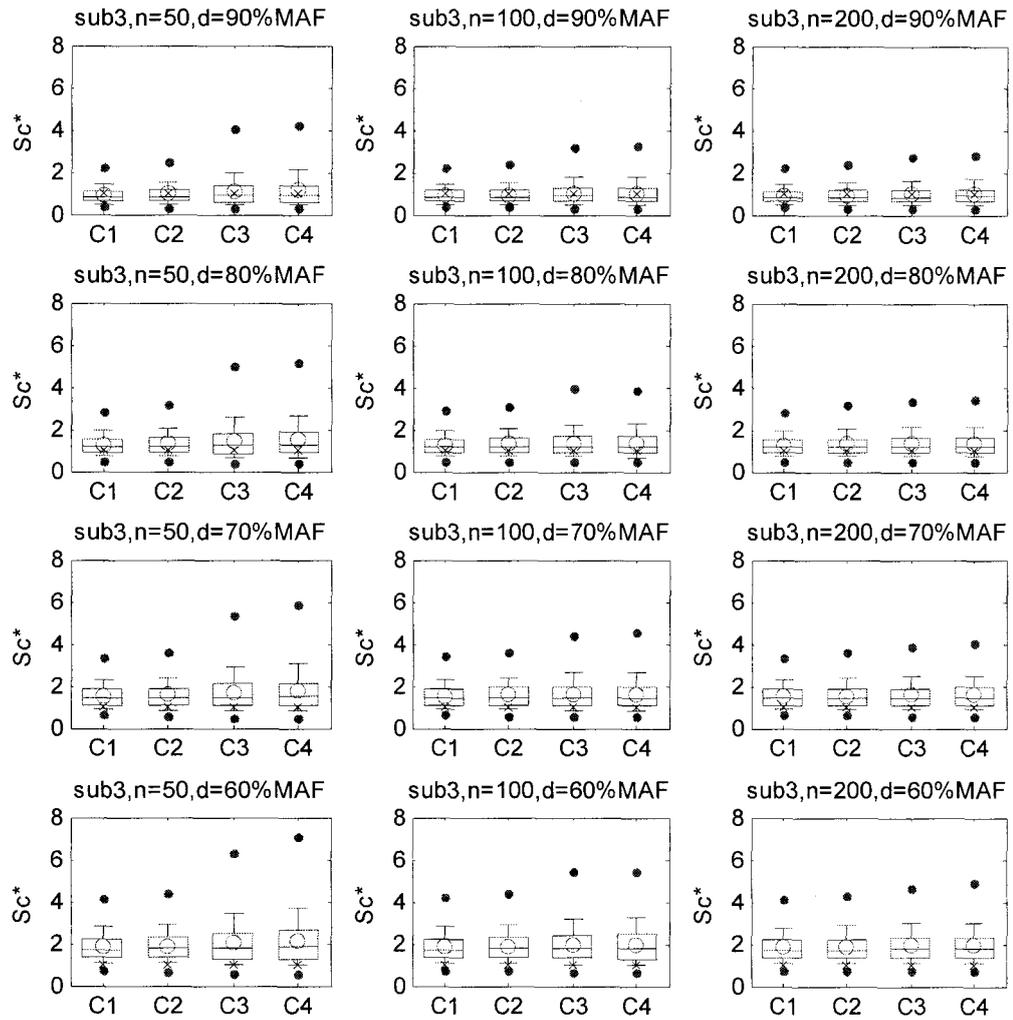


Figure 4.A13: Quantile plots of storage capacity ( $Sc$ ) calculated from synthetic streamflows at sub-station 3(site 7) which are disaggregated (VS model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) for different demand levels. C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size.

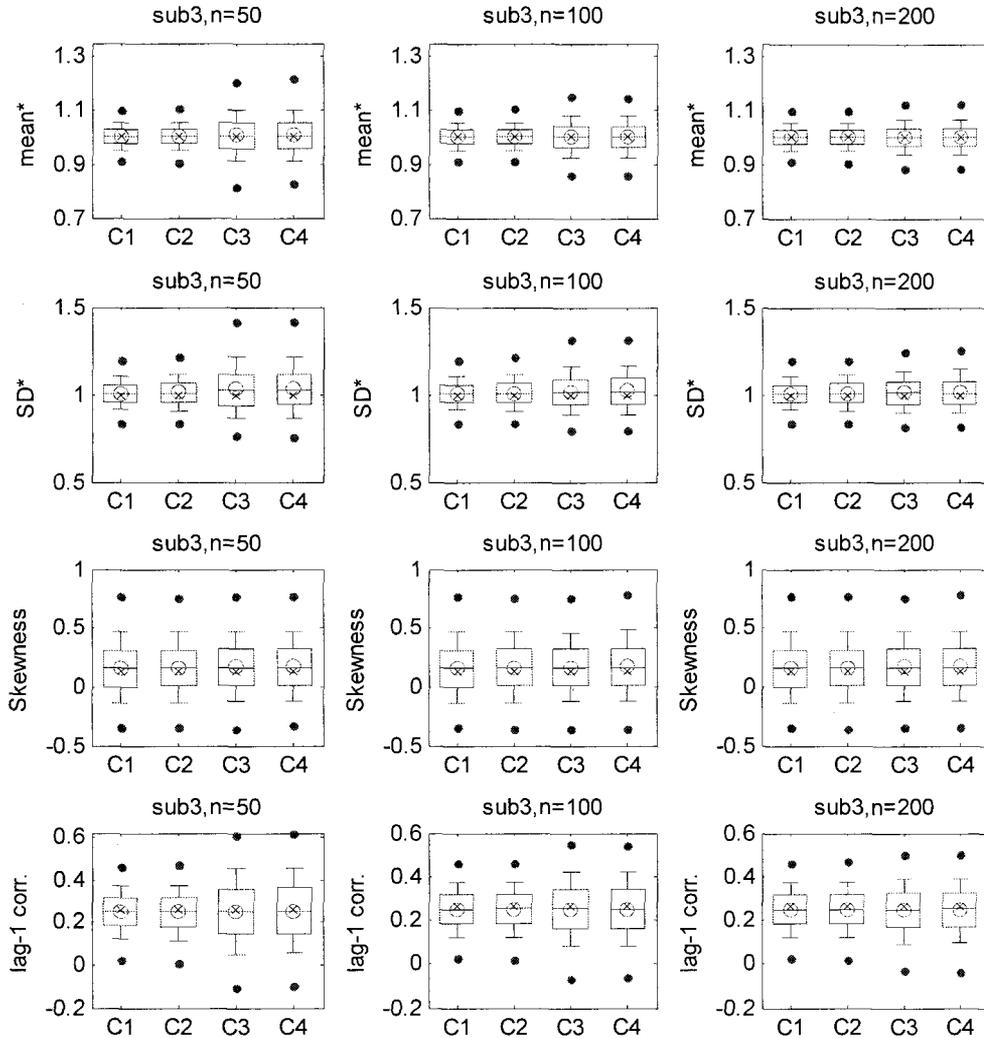


Figure 4.A14: Quantile plots of basic statistics calculated from synthetic flows at sub-station 2(site 6) disaggregated (LC model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

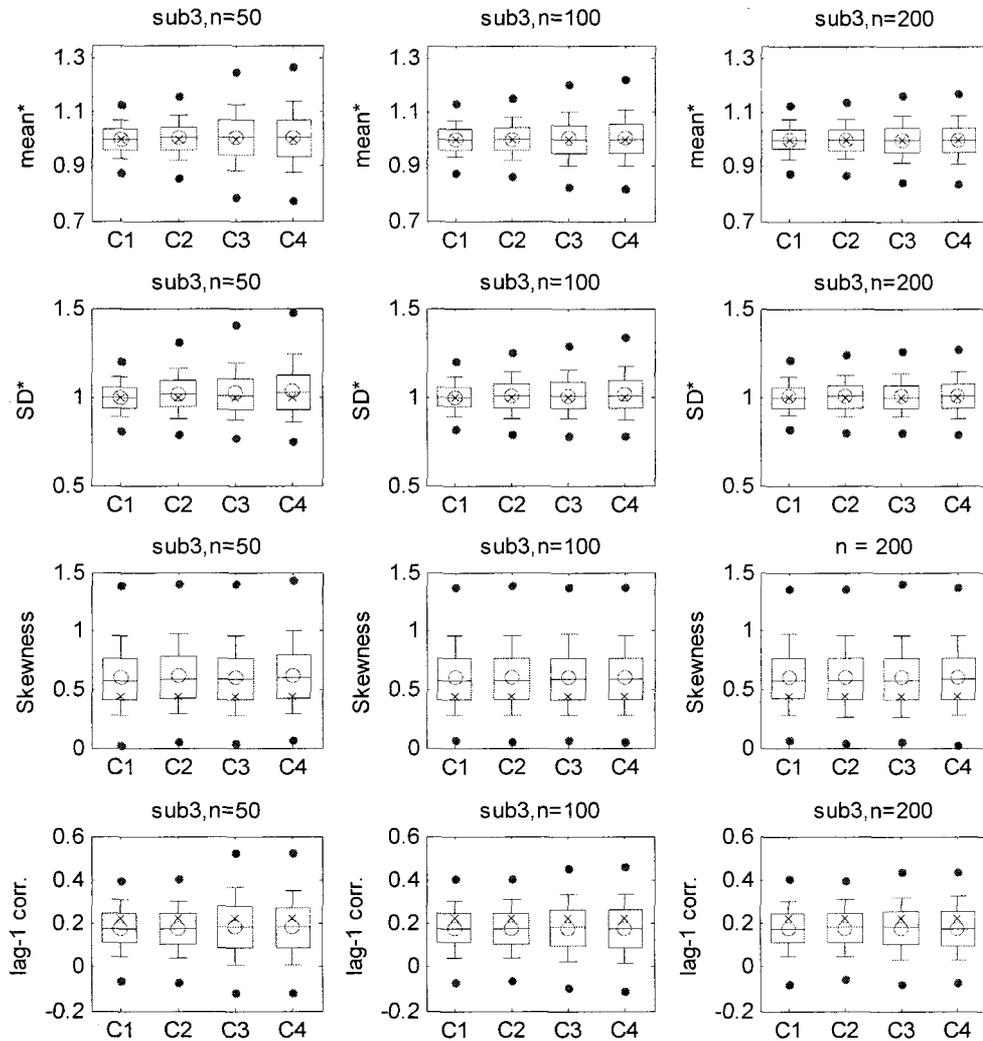


Figure 4.A15: Quantile plots of basic statistics calculated from synthetic flows at sub-station 3(site 7) disaggregated (LC model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach) where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, respectively, and the superscript \* means “scaled by historical statistics”. The upper, middle and lower line in the box are 75, 50, 25% quantiles, the whisker extends from the box to 90, 10% quantile for each side, two dots outside box mean 99, 1% quantile values, and ‘X’ denotes historical statistics.

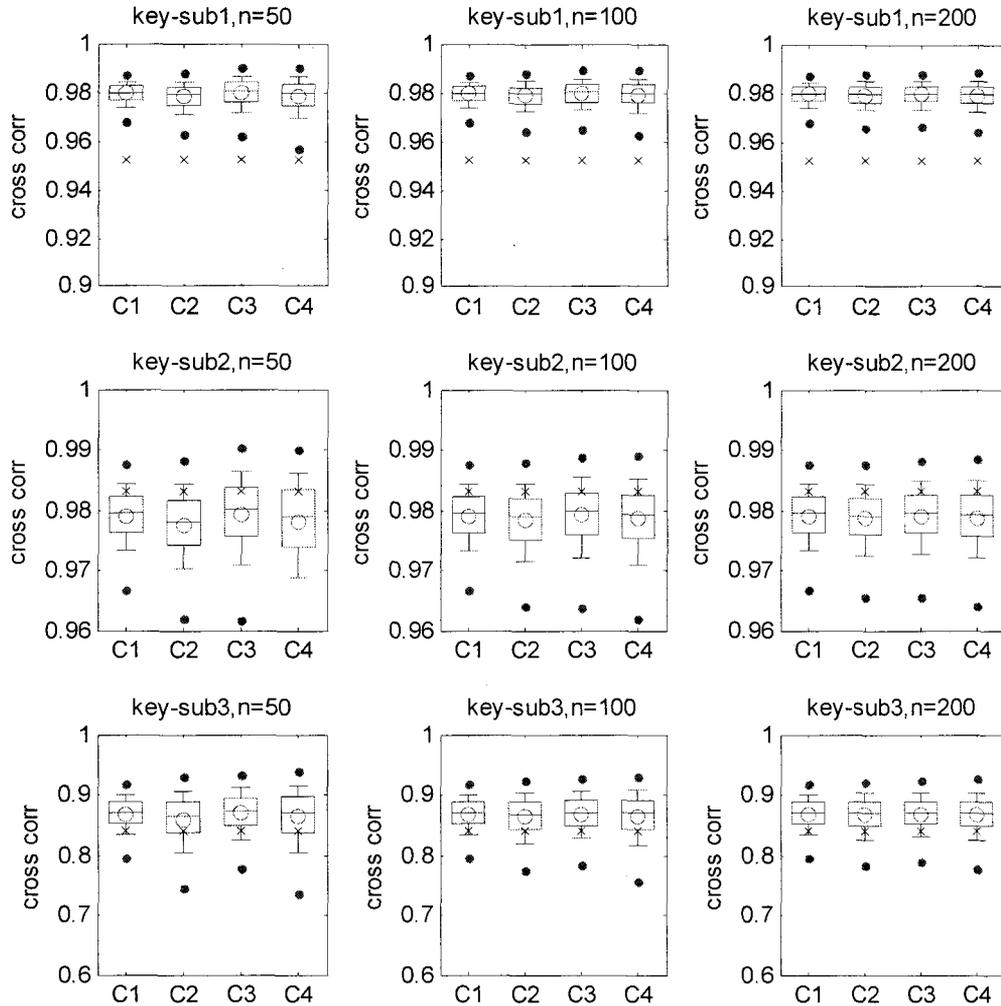


Figure 4.A16: Quantile plots of calculated cross correlations between generated key-station flows (AR(1) model) and disaggregated sub-station flows (LC model) with parameter uncertainty incorporated (Bayesian approach), where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size, sub1 means sub-station 1(site 2), sub2 means sub-station 2 (site 6), and sub3 means sub-station 3 (site 7).

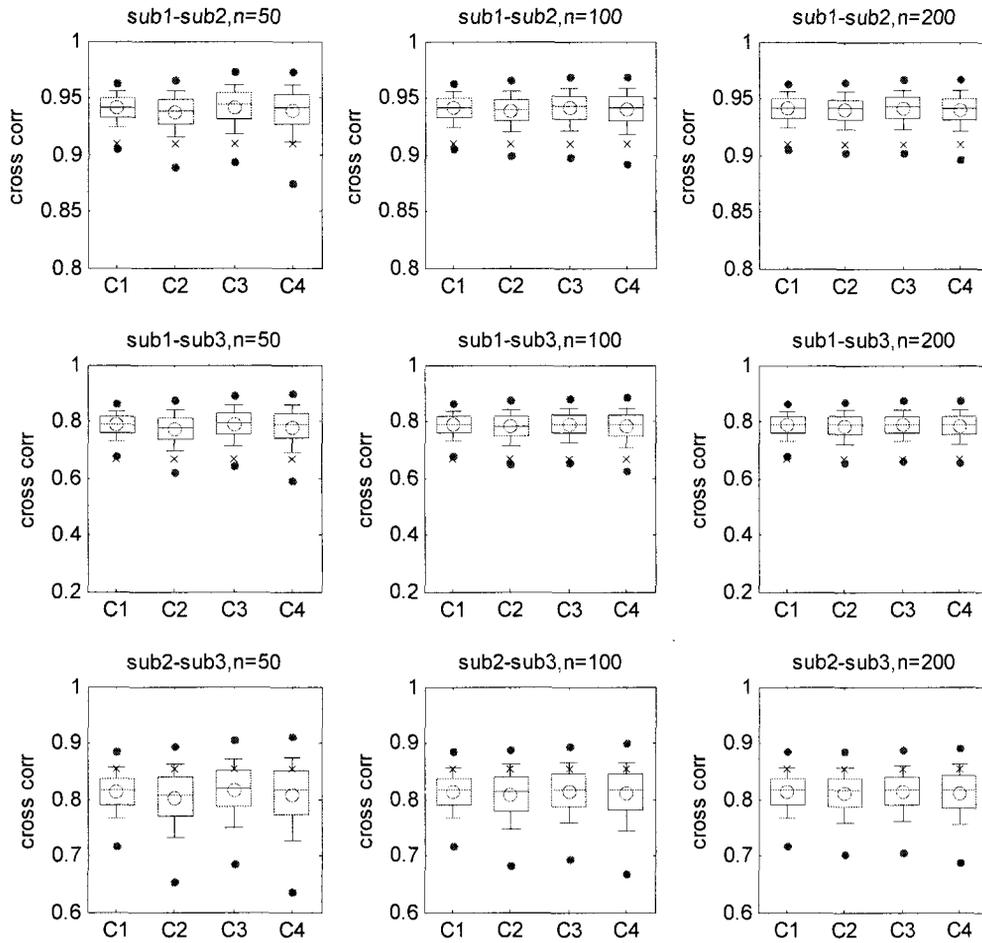


Figure 4.A17: Quantile plots of calculated cross correlations between sub-station flows disaggregated (LC model) from generated key-station flows (AR(1) model) with parameter uncertainty incorporated (Bayesian approach), where C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size, sub1 means sub-station 1(site 2), sub2 means sub-station 2 (site 6), and sub3 means sub-station 3 (site 7).

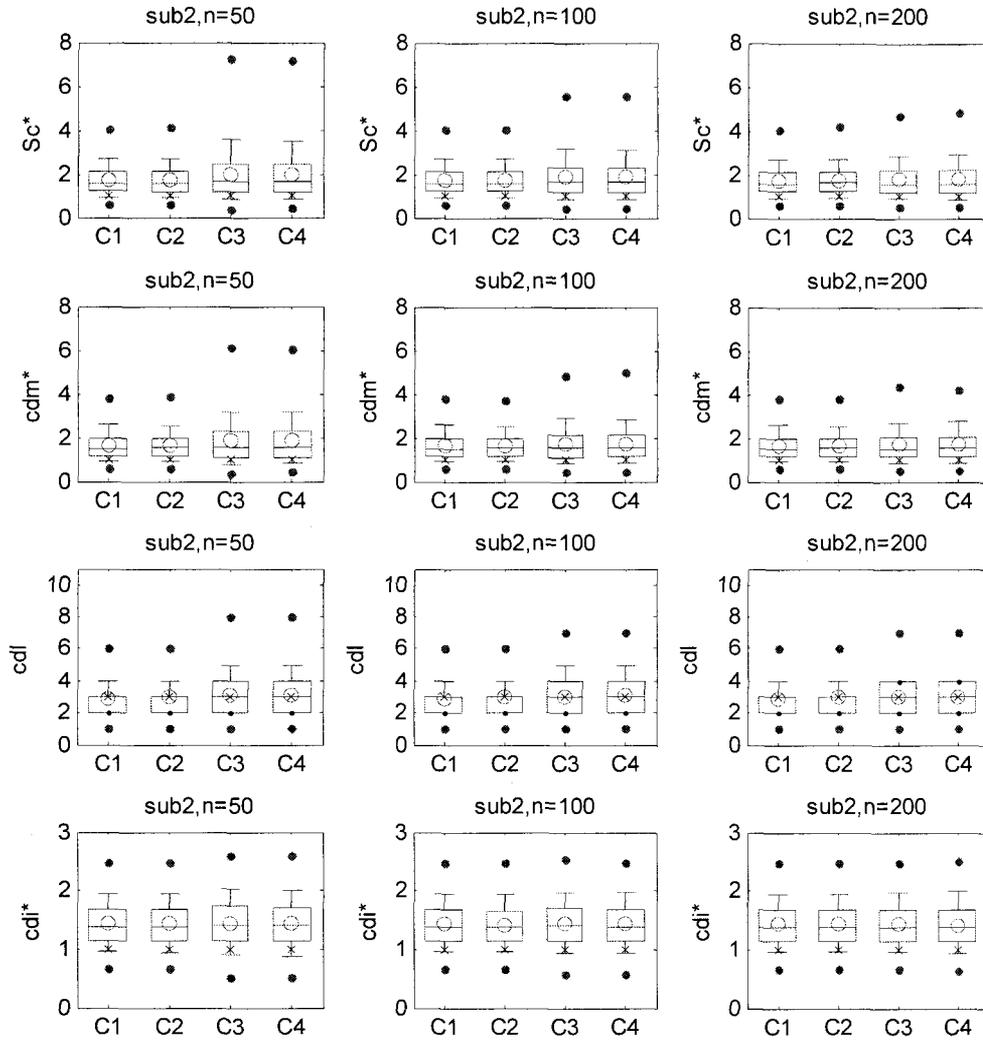


Figure 4.A18: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 2(site 6) which are disaggregated (LC model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

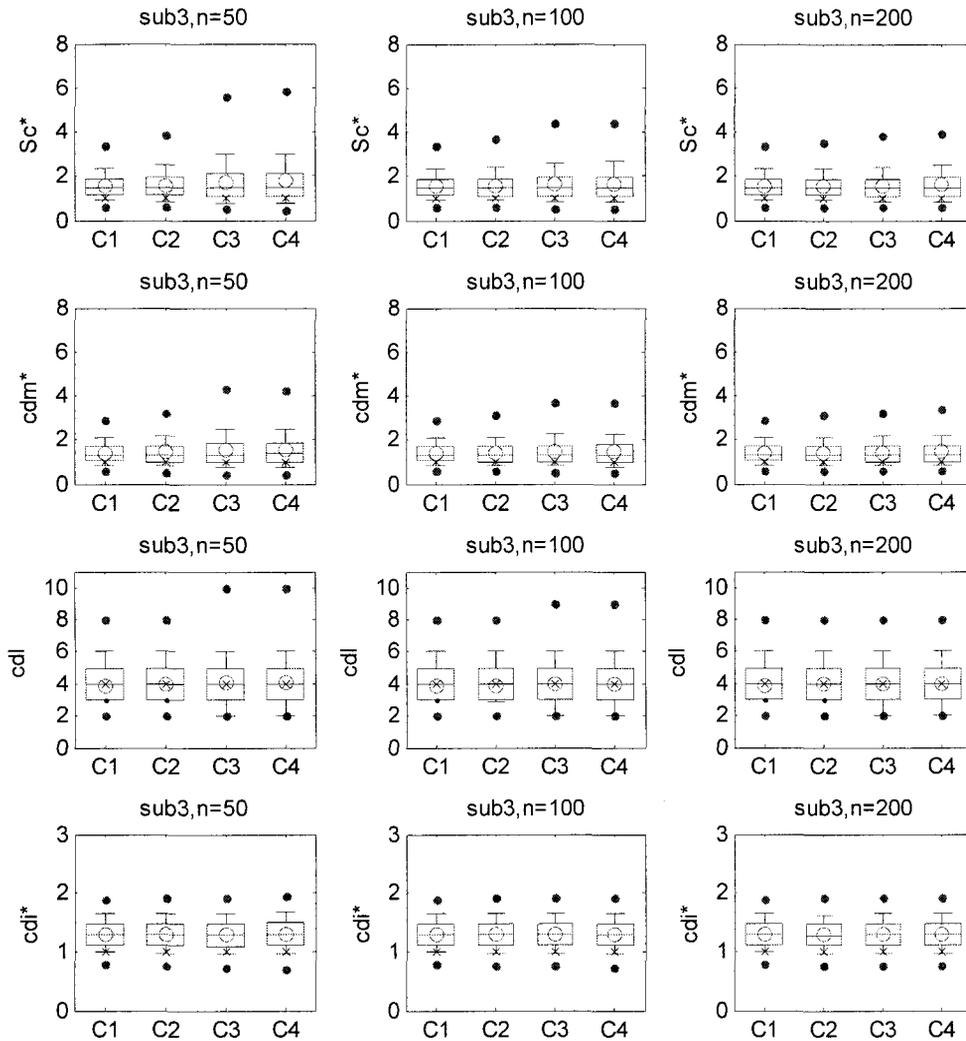


Figure 4.A19: Quantile plots of storage capacity ( $Sc$ ), critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) calculated from synthetic streamflows at sub-station 3(site 7) which are disaggregated (LC model) from generated flows at key-station (AR(1) model) with parameter uncertainty incorporated (Bayesian approach). C1,C2,C3,C4 denote case 1, case 2, case 3, case 4, 'X' denotes the historical cross correlation,  $n$  is assumed sample size. The demand level was assumed as 70% MAF at each site.

Table 4.A1: Quantile estimates of generated critical drought magnitudes (rescaled by dividing the historical critical drought magnitude) calculated from generated sub-station flows disaggregated from generated annual flows by using AR(1) and VS disaggregation models for 4 different uncertainty considerations. (Bayesian approach, demand level = 70%MAF)

sub-station 1	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.20	0.07	0.00	0.00	0.19	0.00	0.01	0.01
1% quantile	0.47	0.47	0.29	0.31	0.48	0.46	0.42	0.39
10% quantile	0.82	0.83	0.70	0.74	0.82	0.81	0.80	0.79
25% quantile	1.07	1.08	1.02	1.07	1.06	1.08	1.07	1.07
Median	1.42	1.45	1.47	1.55	1.43	1.45	1.44	1.46
75% quantile	1.90	1.96	2.16	2.27	1.93	1.92	1.99	1.99
90% quantile	2.50	2.56	3.10	3.23	2.50	2.51	2.68	2.67
99% quantile	3.79	4.02	6.17	6.68	3.80	3.98	4.25	4.34
Maximum	6.23	7.60	10.52	14.04	6.38	7.22	7.09	7.76
Mean	1.56	1.60	1.75	1.85	1.57	1.58	1.62	1.62
Standard Deviation	0.70	0.74	1.13	1.24	0.70	0.72	0.80	0.81
Coeff. of variation	0.45	0.46	0.65	0.67	0.45	0.46	0.49	0.50
sub-station 2	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.20	0.19	0.00	0.00	0.31	0.23	0.24	0.03
1% quantile	0.56	0.57	0.38	0.42	0.57	0.56	0.54	0.53
10% quantile	0.90	0.89	0.77	0.81	0.90	0.90	0.89	0.89
25% quantile	1.15	1.15	1.09	1.13	1.14	1.15	1.15	1.14
Median	1.51	1.54	1.56	1.63	1.51	1.53	1.54	1.53
75% quantile	2.01	2.00	2.29	2.35	2.00	2.01	2.10	2.09
90% quantile	2.56	2.60	3.22	3.38	2.58	2.56	2.75	2.79
99% quantile	3.76	3.86	6.15	6.45	3.76	3.88	4.50	4.57
Maximum	5.63	6.61	10.70	16.56	5.75	7.07	7.65	7.99
Mean	1.64	1.66	1.84	1.91	1.64	1.65	1.72	1.72
Standard Deviation	0.69	0.71	1.15	1.21	0.69	0.71	0.81	0.83
Coeff. of variation	0.42	0.43	0.62	0.63	0.42	0.43	0.47	0.49
sub-station 3	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.44	0.35	0.22	0.10	0.37	0.38	0.26	0.26
1% quantile	0.59	0.58	0.51	0.49	0.61	0.59	0.57	0.59
10% quantile	0.84	0.82	0.76	0.78	0.84	0.84	0.82	0.83
25% quantile	1.04	1.04	0.99	1.00	1.03	1.04	1.04	1.03
Median	1.30	1.32	1.34	1.38	1.31	1.31	1.33	1.32
75% quantile	1.64	1.70	1.85	1.91	1.65	1.67	1.73	1.72
90% quantile	2.07	2.12	2.44	2.57	2.05	2.13	2.16	2.21
99% quantile	3.04	3.06	4.14	4.62	2.91	3.08	3.34	3.31
Maximum	4.59	5.64	7.72	10.16	4.26	5.49	5.22	6.82
Mean	1.39	1.42	1.51	1.57	1.39	1.41	1.44	1.44
Standard Deviation	0.51	0.54	0.75	0.83	0.49	0.54	0.58	0.59
Coeff. of variation	0.36	0.38	0.49	0.53	0.35	0.38	0.40	0.41

Table 4.A2: Quantile estimates of generated critical drought lengths calculated from generated sub-station flows disaggregated from generated annual flows by using AR(1) and VS disaggregation models for 4 different uncertainty considerations. (Bayesian approach, demand level = 70%MAF)

sub-station 1	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	1	1	0	0	1	0	1	1
1% quantile	1	1	1	1	1	1	1	1
10% quantile	1	1	1	1	1	1	1	1
25% quantile	2	2	2	2	2	2	2	2
Median	2	2	2	2	2	2	2	2
75% quantile	3	3	3	3	3	3	3	3
90% quantile	4	4	4	4	4	4	4	4
99% quantile	5	5	7	7	5	5	6	5.01
Maximum	8	10	12	13	8	7	9	8
Mean	2.4	2.5	2.5	2.7	2.4	2.4	2.5	2.5
Standard Deviation	0.9	1.0	1.3	1.3	0.9	0.9	1.0	1.0
Coeff. of variation	0.38	0.39	0.49	0.51	0.37	0.38	0.41	0.41
sub-station 2	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	1	1	0	0	1	1	1	1
1% quantile	1	1	1	1	1	1	1	1
10% quantile	2	2	2	2	2	2	2	2
25% quantile	2	2	2	2	2	2	2	2
Median	3	3	3	3	3	3	3	3
75% quantile	3	3	4	4	3	3	4	4
90% quantile	4	4	5	5	4	4	5	4
99% quantile	6	6	8	8	6	6	7	7
Maximum	9	12	15	16	9	10	10	11
Mean	2.9	3.0	3.1	3.2	2.9	3.0	3.0	3.0
Standard Deviation	1.0	1.1	1.5	1.5	1.0	1.1	1.2	1.2
Coeff. of variation	0.35	0.36	0.47	0.48	0.35	0.36	0.39	0.39
sub-station 3	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	1	1	1	1	1	1	1	1
1% quantile	2	2	2	2	2	2	2	2
10% quantile	3	2	2	2	3	3	2	2
25% quantile	3	3	3	3	3	3	3	3
Median	4	4	4	4	4	4	4	4
75% quantile	4	4	5	5	4	5	5	5
90% quantile	6	5	6	6	5	6	6	6
99% quantile	8	8	9	10	8	8	9	8
Maximum	14	14	14	20	11	12	13	13
Mean	3.9	3.9	4.0	4.1	3.8	3.9	3.9	3.9
Standard Deviation	1.3	1.3	1.6	1.7	1.2	1.3	1.4	1.4
Coeff. of variation	0.34	0.34	0.40	0.43	0.32	0.34	0.36	0.36

Table 4.A3: Quantile estimates of generated critical drought intensities (rescaled by dividing the historical critical drought intensity) calculated from generated sub-station flows disaggregated from generated annual flows by using AR(1) and VS disaggregation models for 4 different uncertainty considerations. (Bayesian approach, demand level = 70%MAF)

sub-station 1	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.16	0.07	0.00	0.00	0.19	0.00	0.01	0.01
1% quantile	0.40	0.38	0.27	0.28	0.40	0.38	0.35	0.34
10% quantile	0.63	0.64	0.57	0.59	0.62	0.62	0.62	0.62
25% quantile	0.78	0.79	0.76	0.78	0.78	0.77	0.78	0.78
Median	0.97	0.98	0.98	1.01	0.97	0.96	0.98	0.99
75% quantile	1.19	1.20	1.23	1.27	1.19	1.19	1.19	1.21
90% quantile	1.42	1.43	1.49	1.54	1.41	1.41	1.42	1.44
99% quantile	1.86	1.88	2.01	2.05	1.85	1.85	1.84	1.89
Maximum	2.49	2.49	2.35	2.74	2.48	2.44	2.46	2.68
Mean	1.00	1.01	1.01	1.04	1.00	1.00	1.00	1.01
Standard Deviation	0.31	0.32	0.36	0.38	0.31	0.31	0.32	0.33
Coeff. of variation	0.31	0.31	0.36	0.36	0.31	0.31	0.32	0.32
sub-station 2	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.31	0.29	0.00	0.00	0.34	0.29	0.36	0.05
1% quantile	0.68	0.68	0.47	0.55	0.68	0.65	0.63	0.66
10% quantile	0.96	0.95	0.90	0.92	0.97	0.95	0.95	0.96
25% quantile	1.14	1.14	1.13	1.14	1.14	1.15	1.15	1.16
Median	1.38	1.39	1.39	1.43	1.38	1.40	1.40	1.41
75% quantile	1.65	1.67	1.71	1.75	1.66	1.67	1.68	1.68
90% quantile	1.93	1.95	2.04	2.06	1.95	1.96	1.97	1.97
99% quantile	2.46	2.50	2.59	2.66	2.46	2.47	2.50	2.51
Maximum	2.87	2.84	2.86	2.87	2.86	2.86	2.86	2.86
Mean	1.42	1.43	1.44	1.46	1.42	1.43	1.43	1.44
Standard Deviation	0.39	0.40	0.45	0.45	0.39	0.39	0.40	0.40
Coeff. of variation	0.27	0.28	0.31	0.31	0.27	0.27	0.28	0.28
sub-station 3	n=50				n=200			
	C1	C2	C3	C4	C1	C2	C3	C4
Minimum	0.51	0.45	0.44	0.21	0.55	0.55	0.52	0.45
1% quantile	0.81	0.78	0.75	0.74	0.80	0.81	0.80	0.79
10% quantile	1.01	1.01	0.99	1.00	1.00	1.01	1.01	1.01
25% quantile	1.15	1.16	1.15	1.15	1.14	1.15	1.15	1.15
Median	1.31	1.34	1.33	1.35	1.31	1.32	1.33	1.33
75% quantile	1.50	1.52	1.53	1.56	1.50	1.50	1.51	1.53
90% quantile	1.66	1.69	1.69	1.73	1.66	1.66	1.67	1.68
99% quantile	1.92	1.93	1.93	1.97	1.92	1.92	1.92	1.93
Maximum	2.05	2.05	2.05	2.06	2.06	2.05	2.05	2.06
Mean	1.33	1.34	1.34	1.35	1.33	1.33	1.33	1.34
Standard Deviation	0.25	0.26	0.27	0.28	0.25	0.25	0.25	0.26
Coeff. of variation	0.19	0.19	0.20	0.21	0.19	0.19	0.19	0.19

## Chapter V

### UNCERTAINTY CONSIDERATION IN MULTIVARIATE ANNUAL STREAMFLOWS GENERATION

**Abstract:** Multivariate ARMA models have been widely used in hydrologic streamflows simulation and forecasting for their better performance over univariate models when sites of concern are statistically closely related. In addition to model uncertainty encountered when choosing a well fitted model for application to historical records, parameter uncertainty issues arise. A quantification of parameter uncertainty and its incorporation into the generation of synthetic streamflows are expected to improve the performance and reliability in both planning and management of water resources. In this study, parameter uncertainty associated with multivariate generation model will be considered by using asymptotic theory and Bayesian framework. A multivariate AR model is used for the underlying model and a streamflows generation will be performed with parameter uncertainty incorporated for sites on the Colorado River basin. The parameter uncertainty effect has been shown to result in generated flows with higher variance than historical flows (Valdes, et al, 1977). Basic statistics as well as storage and drought related statistics are calculated to evaluate the parameter uncertainty effect. This effect enlarges the variability of generated statistics, even with a sample size is 100. Upward

bias of generated design variables are also expected which could be explained by the parameter uncertainty effect. The Bayesian approach differs from the asymptotic approach because it shows larger variability of generated design variables, especially in upper quantiles for sample sizes equal to or smaller than 100. Different transformations are applied to the data in order to evaluate the skewness effect. A proper elimination of skewness of real data for the case of the multivariate generation with parameter uncertainty incorporated might be able to eliminate any unexpected increases in the variability of generated design variables, which are not found in the univariate generation.

## **5.1 Introduction**

The application of multivariate stochastic models has been useful in the planning and management of water resources systems of several sites interdependently. Since the work of Fiering (1964), a number of multivariate generation models have been presented for streamflows simulation and forecasting in hydrology, e.g., multivariate autoregressive moving average (ARMA), multiple regression, and disaggregation models (Fiering, 1964; Matalas, 1967; Pegram and James, 1972; Valencia and Shaake, 1973; O'Connell, 1974; Mejia and Rouselle, 1976; Salas and Pegram, 1977; Lettenmaier, 1980; Salas et al., 1980; Camacho et al, 1985).

Multivariate ARMA models can be classified into three categories depending on the correlation structure of the ARMA model; full (general) multivariate ARMA models, transfer function ARMA models, and contemporaneous ARMA models. In particular, a

contemporaneous ARMA model with a structure using a diagonal parameter matrix has been advocated in hydrology as an alternative to the general multivariate ARMA model because of its simplistic structure (Salas et al, 1985). In any case, several parameter estimation methods for multivariate ARMA models are available and may be applied (Matalas, 1967; Pegram and James, 1972; O'Connell, 1974; Salas and Pegram, 1977; Lettenmaier, 1980; Salas et al., 1980; Camacho et al, 1985; Stedinger, et al., 1985). For example, Stedinger et al. (1985) compared several parameter estimation techniques for the contemporaneous ARMA(1,1) and proposed that maximum likelihood (ML) estimators of autoregressive moving average parameters based the univariate estimation and method of moments (MOM) estimators of residuals using model parameters behave the best.

Even though a specific multivariate model has been well suited for synthetic streamflows generation as in the case of the univariate streamflows generation described in the previous chapter, parameter uncertainty remains an issue due to a small historical sample size. To address the issue of parameter uncertainty, Camacho, et el. (1987) developed asymptotic distributions of ML estimators of parameter matrix for the case of the contemporaneous ARMA and compared the asymptotic variances of parameter estimates between on the univariate estimation and on the joint estimation. They proposed that even though those two ML estimators are asymptotically efficient jointly obtained ML estimators would give the smaller asymptotic variance-covariance matrix. Use of joint ML estimators might be therefore more efficient in this sense and an approximation procedure to get ML estimators corresponding to joint estimation has been developed based on the approximation (Camacho, et el., 1987). Compared to

contemporaneous ARMA, parameter estimation of general multivariate ARMA models, especially those based on ML estimators, is not easy when higher order models are required. However, once parameter estimates are obtained the asymptotic distributions of ML estimators may be applied (Lütkepohl, 1993).

Additionally, parameter uncertainty has been dealt with in the Bayesian framework. Based on the Bayesian multivariate regression model (Zellner, 1971), an evaluation of parameter uncertainty of the general multivariate AR model through the generation of annual streamflows was performed (Valdes et al., 1977). They applied Bayesian posterior distributions as a way of incorporating parameter uncertainty and proposed that parameter uncertainty produces synthetic streamflows with higher variances when compared with the historical records.

In this study, a parameter uncertainty issue concerning the multivariate generation case will be discussed. Asymptotic distributions of ML estimators will be utilized and their performance will be compared with that from the Bayesian framework. Instead of the general multivariate ARMA model, a multivariate AR model will be used as the underlying model for its simplicity in obtaining ML estimators. Furthermore note that the traditional Bayesian approach will be restricted on the multivariate AR model because of using the multivariate regression structure.

In the work of Vicens et al. (1975) a conjugate prior has been compared with a non-informative prior, which showed that a proper conjugate prior could reduce the parameter uncertainty effect. However, the true prior might not be the conjugate prior, so using the conjugate prior might lead to an incorrect posterior and affect the reliability of the simulation performance when the available sample size is small. Therefore, a

non-informative prior seems more appropriate than a conjugate prior. Previous chapters showed that derived posterior distributions of either AR parameters for the univariate AR models or disaggregation parameters have similarity with asymptotic distributions when a non-informative prior is used, which has been verified by structurally a correct and asymptotically exact Bayesian method (McLeod and Hipel, 1978; Stedinger and Taylor, 1982). However, a redistribution of AR parameters sampled from a posterior distribution with the non-informative prior would be required to meet the stationary condition of the models since the non-informative prior of AR parameters is assumed in the range of from minus infinity to infinity. This modification of the parameter space will be applied to the model generation, which will be termed by a modified posterior distribution. It is also probable that using the uniform prior within the parameter space might be the better approach for the unknown prior but this is not considered in this study since the posterior distribution could not only be explicitly defined but also numerical calculation would be required.

A multivariate AR model is assumed to generate annual streamflows at three different sites in the Colorado River basin, where the parameter uncertainty effect on the jointly simulated streamflows will be examined. In order to incorporate the uncertainty of parameter estimates into the generation of streamflows, new sets of parameter estimates will be sampled from their asymptotic or posterior distributions in each generation and then substituted for the original parameter estimates obtained from the historical sample. Distributions of generated basic statistics, storage related statistics, and drought related statistics will be compared. Additionally, a comparison of generated statistics based on both univariate and multivariate generations will be conducted to

investigate any additional parameter uncertainty effects associated with the extension of the univariate model to the multivariate case. Then the skewness effect combining with the parameter uncertainty incorporation into the multivariate generation will be discussed by evaluating different transformations of real data.

## 5.2 Multivariate Autoregressive Model

A general multivariate autoregressive process is given by (Matalas, 1967)

$$\Phi(B)(\mathbf{Z}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t \quad (5.1)$$

where  $\mathbf{Z}_t = (z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(m)})'$  is the  $m \times 1$  vector with mean of  $\boldsymbol{\mu} = (\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(m)})'$  and  $z_t^{(i)}$  for  $t=1, 2, \dots, n$ ,  $i=1, 2, \dots, m$  denotes flows sequences at site  $i$  where  $n$  is sample size and  $m$  is the number of sites.  $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \dots, \varepsilon_t^{(m)})'$  is the  $m \times 1$  independent normal random vectors with zero mean vector and  $m \times m$  variance-covariance matrix  $\Delta$  where  $(i, j)$  element of  $\Delta$  is denoted by  $\sigma_{ij}$ . If  $\sigma_{ij} = 0$  for  $i \neq j$ , Equation (5.1) has the same form as  $m$  independent univariate AR model.  $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 \dots - \Phi_p B^p$  is the AR operator of order  $p$  where  $B$  is backward shift operator. In Equation (5.1), it is assumed that the zeros of polynomial equations  $\Phi(B) = 0$  lie outside the unit circle such that the model is stationary.

Generally, contemporaneous AR model results if each  $m \times m$  matrix  $\Phi_1, \Phi_2, \dots, \Phi_p$  is considered to be a diagonal (Camacho et al., 1985; Salas et. al, 1985). Thus, a contemporaneous AR model can be thought of a collection of  $m$  univariate AR

models with contemporaneously correlated innovations. Note that if  $\Phi_1, \Phi_2, \dots, \Phi_p$  are upper (lower) diagonal matrices, then Equation (5.1) can be said to be classified as a transfer function AR model, and if  $\Phi_1, \Phi_2, \dots, \Phi_p$  are full matrices, then Equation (5.1) is a vector AR model or full multivariate model (Salas, et. al, 1985).

### 5.2.1 Maximum likelihood estimators and asymptotic distributions

Defining:

$$\mathbf{Y} = [\mathbf{Z}_1 - \boldsymbol{\mu}, \mathbf{Z}_2 - \boldsymbol{\mu}, \dots, \mathbf{Z}_t - \boldsymbol{\mu}, \dots, \mathbf{Z}_n - \boldsymbol{\mu}] \quad (m \times n)$$

$$\mathbf{Z}_t^0 = \begin{bmatrix} \mathbf{Z}_t - \boldsymbol{\mu} \\ \vdots \\ \mathbf{Z}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix} \quad (mp \times 1)$$

$$\mathbf{X} = [\mathbf{Z}_0^0, \mathbf{Z}_1^0, \dots, \mathbf{Z}_t^0, \dots, \mathbf{Z}_{n-1}^0] \quad (mp \times n)$$

$$\mathbf{A} = [\Phi_1, \Phi_2, \dots, \Phi_p] \quad (m \times mp)$$

$$\mathbf{U} = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_t, \dots, \boldsymbol{\varepsilon}_n] \quad (m \times n)$$

where  $\mathbf{Z}_t = (z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(m)})'$ ,  $\boldsymbol{\mu} = (\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(m)})'$ , and  $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \dots, \varepsilon_t^{(m)})'$  is the  $m \times 1$  independent normal random vectors with zero mean vector and  $m \times m$  variance-covariance matrix  $\boldsymbol{\Sigma}$ . Equation (5.1) can be rewritten more simply as

$$\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{U} \quad (5.2a)$$

or, equivalently,

$$\text{vec}(\mathbf{Y}) = (\mathbf{X}' \otimes \mathbf{I}_m) \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{U}) \quad (5.2b)$$

where  $\text{vec}(\cdot)$  represents the vectorization operator; i.e.  $\text{vec}(\mathbf{U})$  represents  $mn \times 1$

column vector formed by stacking the columns of the  $m \times n$  matrix of  $\mathbf{U}$ ,  $vec(\mathbf{U})_{(i-1)n+j} = U_{ij}$  for  $i=1,2,\dots,m$ ,  $j=1,2,\dots,n$ , and  $\mathbf{I}_m$  is  $m \times m$  identity matrix. The likelihood function can be easily driven from Equation (5.2a) and the maximum likelihood estimators of parameter matrices  $\boldsymbol{\mu}$ ,  $\mathbf{A}$ , and  $\boldsymbol{\Sigma}$  are given by solving normal equations as follows: (Lütkepohl, 1993)

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \left( \mathbf{I}_m - \sum_{i=1}^p \hat{\boldsymbol{\Phi}}_i \right)^{-1} \sum_{t=1}^n \left( \tilde{\mathbf{z}}_t - \sum_{i=1}^p \hat{\boldsymbol{\Phi}}_i \tilde{\mathbf{z}}_{t-1} \right) \quad (5.3)$$

$$vec(\hat{\mathbf{A}}) = ((\tilde{\mathbf{X}}\tilde{\mathbf{X}}')^{-1} \tilde{\mathbf{X}} \otimes \mathbf{I}_m) (vec(\tilde{\mathbf{Y}}) - vec(\hat{\boldsymbol{\mu}})) \quad (5.4)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} (\tilde{\mathbf{Y}} - \hat{\mathbf{A}}\tilde{\mathbf{X}})(\tilde{\mathbf{Y}} - \hat{\mathbf{A}}\tilde{\mathbf{X}})' \quad (5.5)$$

where  $\tilde{\mathbf{Y}}$ ,  $\tilde{\mathbf{X}}$ ,  $\tilde{\mathbf{z}}_t$  are observed values of  $\mathbf{Y}$ ,  $\mathbf{X}$ ,  $\mathbf{z}_t$ . An initial value of  $\hat{\mathbf{A}}$  can be set to the least square estimator:  $\hat{\mathbf{A}} = \tilde{\mathbf{Y}}\tilde{\mathbf{X}}'(\tilde{\mathbf{X}}\tilde{\mathbf{X}}')^{-1}$  or method of moments estimators, and iterations are repeated until convergence is obtained to give the MLE of  $\boldsymbol{\mu}$ ,  $\mathbf{A}$ .

Asymptotic distributions of  $\hat{\boldsymbol{\mu}}$ ,  $\hat{\mathbf{A}}$ , and  $\hat{\boldsymbol{\Sigma}}$  are given as follows (Lütkepohl, 1993)

$$vec(\hat{\mathbf{A}}) \sim AMVN(vec(\mathbf{A}), \frac{1}{n} \Gamma_{\mathbf{X}}(0)^{-1} \otimes \boldsymbol{\Sigma}) \quad (5.6)$$

$$\hat{\boldsymbol{\mu}} \sim AMVN\left(\boldsymbol{\mu}, \frac{1}{n} (\mathbf{I}_m - \sum_{i=1}^p \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Sigma} (\mathbf{I}_m - \sum_{i=1}^p \boldsymbol{\Phi}_i)^{\prime -1}\right) \quad (5.7)$$

$$vech(\hat{\boldsymbol{\Sigma}}) \sim AMVN\left(vech(\boldsymbol{\Sigma}), \frac{2}{n} \mathbf{D}_m^+ (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \mathbf{D}_m^{\prime +}\right) \quad (5.8)$$

where  $\sim AMVN$  means ‘asymptotically follows multivariate normal distribution’,  $\Gamma_{\mathbf{X}}(0)$  is the variance-covariance matrix of  $\mathbf{X}$ :  $\Gamma_{\mathbf{X}}(0) = \mathbf{X}\mathbf{X}'/n$ .  $\mathbf{D}_m^+ = (\mathbf{D}_m' \mathbf{D}_m)^{-1} \mathbf{D}_m'$

where  $\mathbf{D}_m$  is a duplication matrix so that  $\text{vec}(\Delta) = \mathbf{D}_m \text{vech}(\Delta)$  where vector operators of  $\text{vec}(\cdot)$  and  $\text{vech}(\cdot)$  are defined:  $\text{vec}(\mathbf{A})$  is  $mn \times 1$  column vector formed by stacking the columns of the  $m \times n$  matrix of  $\mathbf{A}$ ,  $\text{vec}(\mathbf{A})_{(j-1)m+i} = \mathbf{A}_{ij}$  for  $i=1,2,\dots,m$ ,  $j=1,2,\dots,n$  and  $\text{vech}(\mathbf{B})$  stacks the elements on and below the main diagonal of a square matrix  $\mathbf{B}$ . Thus,  $\text{vec}(\Delta)$  is  $m^2$  dimensional vector,  $\text{vech}(\Delta)$  is  $m(m+1)/2$  dimensional vector, and  $\mathbf{D}_m$  is  $m^2 \times m(m+1)/2$  matrix. For example,  $\mathbf{D}_2$  and  $\mathbf{D}_2^+$  associated with  $(2 \times 2)$   $\Delta$  are given by

$$\mathbf{D}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{D}_2^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

### 5.2.2 Parameter uncertainty incorporation by using Bayesian framework

As an alternative, parameter uncertainty of multivariate AR model can be quantified by using posterior distributions of parameter estimates in Bayesian multivariate regression framework (Vales, et al., 1977). After switching the columns and rows, Equation (5.2a) could be rewritten as

$$\mathbf{Z} = \mathbf{W}\mathbf{B} + \mathbf{V} \quad (5.9)$$

with

$$\mathbf{Z}_{n \times m} = (\mathbf{z}'_1, \mathbf{z}'_2, \dots, \mathbf{z}'_n)' = \begin{bmatrix} z_1^{(1)} & z_1^{(2)} & \dots & z_1^{(m)} \\ z_2^{(1)} & z_2^{(2)} & \dots & z_2^{(m)} \\ \vdots & \vdots & \dots & \vdots \\ z_n^{(1)} & z_n^{(2)} & \dots & z_n^{(m)} \end{bmatrix}$$

$$\begin{aligned}
\mathbf{W}_{n \times (mp+1)} &= \begin{bmatrix} 1 & \mathbf{z}'_0 & \mathbf{z}'_{-1} & \cdots & \mathbf{z}'_{-p+1} \\ 1 & \mathbf{z}'_1 & \mathbf{z}'_0 & \cdots & \mathbf{z}'_{-p+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{z}'_{n-1} & \mathbf{z}'_{n-2} & \cdots & \mathbf{z}'_{-p+n} \end{bmatrix} \\
&= \begin{bmatrix} 1 & z_0^{(1)} & z_0^{(2)} & \cdots & z_0^{(m)} & z_{-1}^{(1)} & z_{-1}^{(2)} & \cdots & z_{-1}^{(m)} & \cdots & z_{-p+1}^{(1)} & z_{-p+1}^{(2)} & \cdots & z_{-p+1}^{(m)} \\ 1 & z_1^{(1)} & z_1^{(2)} & \cdots & z_1^{(m)} & z_0^{(1)} & z_0^{(2)} & \cdots & z_0^{(m)} & \cdots & z_{-p+2}^{(1)} & z_{-p+2}^{(2)} & \cdots & z_{-p+2}^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n-1}^{(1)} & z_{n-1}^{(2)} & \cdots & z_{n-1}^{(m)} & z_{n-2}^{(1)} & z_{n-2}^{(2)} & \cdots & z_{n-2}^{(m)} & \cdots & z_{-p+n}^{(1)} & z_{-p+n}^{(2)} & \cdots & z_{-p+n}^{(m)} \end{bmatrix} \\
\mathbf{B}_{(mp+1) \times m} &= \begin{bmatrix} \Phi'_0 \\ \vdots \\ \Phi'_1 \\ \vdots \\ \Phi'_2 \\ \vdots \\ \vdots \\ \Phi'_p \end{bmatrix} = \begin{bmatrix} \phi_0^{(1)} & \phi_0^{(2)} & \cdots & \phi_0^{(m)} \\ \phi_1^{(11)} & \phi_1^{(21)} & \cdots & \phi_1^{(m1)} \\ \phi_1^{(12)} & \phi_1^{(22)} & \cdots & \phi_1^{(m2)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^{(1m)} & \phi_1^{(2m)} & \cdots & \phi_1^{(mm)} \\ \phi_2^{(11)} & \phi_2^{(21)} & \cdots & \phi_2^{(m1)} \\ \phi_2^{(12)} & \phi_2^{(22)} & \cdots & \phi_2^{(m2)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_2^{(1m)} & \phi_2^{(2m)} & \cdots & \phi_2^{(mm)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_p^{(11)} & \phi_p^{(21)} & \cdots & \phi_p^{(m1)} \\ \phi_p^{(12)} & \phi_p^{(22)} & \cdots & \phi_p^{(m2)} \\ \vdots & \vdots & \cdots & \vdots \\ \phi_p^{(1m)} & \phi_p^{(2m)} & \cdots & \phi_p^{(mm)} \end{bmatrix} \\
\mathbf{V}_{n \times m} = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n)' &= \begin{bmatrix} u_1^{(1)} & u_1^{(2)} & \cdots & u_1^{(m)} \\ u_2^{(1)} & u_2^{(2)} & \cdots & u_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ u_n^{(1)} & u_n^{(2)} & \cdots & u_n^{(m)} \end{bmatrix}
\end{aligned}$$

where  $n$  means sample size,  $m$  means the number of sites, and  $p$  is the order of multivariate AR model of concern.  $\Phi_0 = (\phi_0^{(1)}, \phi_0^{(2)}, \dots, \phi_0^{(m)})$  is the function of AR parameters and mean terms:  $\Phi_0 = (\mathbf{I}_m - \sum_{i=1}^p \Phi_i) \boldsymbol{\mu}$ .  $\mathbf{Z}$  and  $\mathbf{W}$  are assumed to be

normally distributed where each row of the innovation matrix  $\mathbf{V}$  is assumed to be independently multivariate normal distributed with zero mean and variance-covariance matrix  $\Sigma$ . Under this assumption, the joint likelihood function of  $\mathbf{B}$  and  $\Sigma$  given  $\mathbf{Z}, \mathbf{W}$  is

$$L(\mathbf{B}, \Sigma | \mathbf{Z}, \mathbf{W}) \propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{S}\Sigma^{-1}) - \frac{1}{2} \text{tr}((\mathbf{B} - \hat{\mathbf{B}})' \mathbf{W}' \mathbf{W} (\mathbf{B} - \hat{\mathbf{B}}) \Sigma^{-1}) \right] \quad (5.11)$$

where  $\mathbf{S} \equiv (\mathbf{Z} - \mathbf{W}\hat{\mathbf{B}})'(\mathbf{Z} - \mathbf{W}\hat{\mathbf{B}})$  is a matrix proportional to the sample innovation covariance matrix. Assuming that the priors of parameters are diffusive and independent such that  $p(\mathbf{B}) = \text{constant}$ ,  $p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$ , and  $p(\mathbf{B}, \Sigma) \propto |\Sigma|^{-(m+1)/2}$ , the posterior distribution of  $\Sigma$  given  $\mathbf{Z}, \mathbf{W}$  are derived by (Zellner, 1971)

$$p(\Sigma | \mathbf{Z}, \mathbf{W}) \sim IW(\mathbf{S}, \nu) \quad (5.12)$$

where  $\nu = n - mp + 1$  is the degree of freedom and  $IW$  denotes the inverse Wishart distribution. Also, the posterior distribution of  $\mathbf{B}$  is given as the generalized multivariate  $t$  distribution:

$$p(\mathbf{B} | \mathbf{X}, \mathbf{Y}) \propto \left| \mathbf{S} + (\mathbf{B} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B} - \hat{\mathbf{B}}) \right|^{n/2} \quad (5.13)$$

which is the form a generalized multivariate  $t$  distribution which has the property that any row or column vector of the matrix will be distributed as a multivariate  $t$  distribution (Zellner, 1971). Thus, it does not appear to be appropriate to apply Equation (5.13) to general multivariate AR model with orders greater than 1. In the literature by Valdes et al. (1977), this posterior distribution of the multivariate  $t$  distribution was implemented for only multivariate AR(1) model. On the other hand, when  $\Sigma$  is assumed to be known the conditional posterior pdf of  $\text{vec}(\mathbf{B})$  given

$\mathbf{Z}, \mathbf{W}, \Sigma$  is a multivariate normal distribution with mean  $vec(\hat{\mathbf{B}})$  and variance covariance matrix  $\Sigma \otimes (\mathbf{W}'\mathbf{W})^{-1}$  expressed by (Zellner, 1971)

$$p(vec(\mathbf{B}) | \mathbf{Z}, \mathbf{W}, \Sigma) \sim MVN \left( vec(\hat{\mathbf{B}}), \Sigma_{m \times m} \otimes (\mathbf{W}'\mathbf{W})^{-1} \right) \quad (5.14)$$

Note that  $vec(\mathbf{B})$  is a  $m^2(p+1)$  dimensional vector and  $\Sigma \otimes (\mathbf{W}'\mathbf{W})^{-1}$  has the dimension of  $m^2(p+1) \times m^2(p+1)$ . For more information regarding derivation of posterior distributions in the Bayesian framework, see Chapter 5 or Zellner (1971), Valdes, et al. (1977).

### 5.3. Application to Colorado River Basin

#### 5.3.1 Parameter uncertainty effect

As a simple example, consider the application of the 1<sup>st</sup> order trivariate autoregressive model TAR(1) to the generation of annual streamflows sequences. Three sites in Colorado River Basin which are available from USBR (2007) for years from 1905 to 2002 (sample size is 98) are selected and the chosen sites are (1) Colorado River main stream above Cisco, UT (USGS site number: 9180500, site 8), (2) Green River, UT (USGS site number: 9315000, site 16), and (3) San Juan River above Navajo, UT (USGS site number: 9379500, site 19). Site locations are shown in Figure 5.1 and main statistics of historical data are presented in Table 5.1. Site 8 shows the largest mean annual flows (MAF) among the selected sites and MAFs of site 16 and site 19 correspond to 79% and 40% to MAF of site 8, respectively. Similar coefficients of variation may be

seen at each site that ranges from 0.29 to 0.41. It is also seen that the particularly close relationship between site 8 and site 16 as well as between site 8 and site 19 is due to a geomorphologic characteristics; i.e., estimated sample cross correlations are  $\hat{\rho}$ (site 8, site 16)= 0.855,  $\hat{\rho}$ (site 16, site 19)=0.766, and  $\hat{\rho}$ (site 8, site 19)=0.562. A skewness test was performed to check the normality of three historical data sets, which suggested that the normality assumption cannot be rejected with 5% significance interval for all data sequences. Based on the these results for the skewness test, a transformation procedure was not utilized for applying the TVR(1) model.

Initial values required for the ML estimation was given by method of moments estimators based on the Yule-Walker equation assuming that the initial observation vector  $\mathbf{z}_0$  is equal to the last year observation vector  $\mathbf{z}_n$ , which yields  $\tilde{\phi}_{11}=0.333$ ,  $\tilde{\phi}_{12} = -0.070$ ,  $\tilde{\phi}_{13} = 0.017$ ,  $\tilde{\phi}_{21}=0.033$ ,  $\tilde{\phi}_{22} = -0.234$ ,  $\tilde{\phi}_{23} = 0.125$ ,  $\tilde{\phi}_{31}=0.089$ ,  $\tilde{\phi}_{32} = -0.030$ , and  $\tilde{\phi}_{33} = -0.004$ . Then, from Equations (5.3)-(5.5), ML estimates of TAR(1) are obtained by:

$$\hat{\mathbf{A}} = \hat{\Phi}_1 = \begin{bmatrix} \hat{\phi}_{11} & \hat{\phi}_{12} & \hat{\phi}_{13} \\ \hat{\phi}_{21} & \hat{\phi}_{22} & \hat{\phi}_{23} \\ \hat{\phi}_{31} & \hat{\phi}_{32} & \hat{\phi}_{33} \end{bmatrix} = \begin{bmatrix} 0.343 & -0.082 & -0.005 \\ 0.043 & 0.223 & 0.103 \\ 0.092 & -0.034 & -0.012 \end{bmatrix} \quad (5.15)$$

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 6.824 \times 10^6 \\ 5.414 \times 10^6 \\ 2.139 \times 10^6 \end{bmatrix} \text{ (acre-feet)} \quad (5.16)$$

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \hat{\sigma}_{23} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_{33} \end{bmatrix} = \begin{bmatrix} 3.516 \times 10^{12} & 2.487 \times 10^{12} & 1.245 \times 10^{12} \\ & 2.421 \times 10^{12} & 0.754 \times 10^{12} \\ \text{symm.} & & 0.755 \times 10^{12} \end{bmatrix} \text{ (acre-feet)}^2 \quad (5.17)$$

It is notable that estimated AR parameters associated with two different sites seem very

small, therefore the instant causality between sites might be of major concern. As a model identification technique, Hipel and McLeod (1994) proposed the residual cross correlation function (CCF) where the residual at each site is calculated after fitting the univariate model. They suggested that applied time series might have the instant Granger causality in terms of predictability and contemporaneous model might be appropriate if the residual CCF is not significantly different from zero, respectively, for instances where the lag is not equal to zero. The calculated residual CCF (not provided in Figure) has been shown to be not significantly different zero for the lag greater than 0, thus a contemporaneous AR(1) model might be applicable for the present application. However, a general TAR(1) model will be used since there is no special reason to stick to the contemporaneous AR model with loss of generosity.

Five thousand synthetic annual streamflows sequences are generated with the same size of historical flows ( $n=98$ ) for each site. Aforementioned asymptotic and posterior distributions are then used to quantify parameter uncertainty. Different parameter sets are obtained from asymptotic and posterior distributions; then those sampled parameters were substituted for historical estimates in each generation. Clearly, the variance term in either asymptotic or posterior distribution is a function of the available sample size as well as the true parameter sets. Through the generation historical parameter estimates are assumed as true parameters and available sample sizes are set to 50, 100, and 200 in order to examine the associated sample size effects associated with parameter uncertainty. In order to resolve the unrealistic range of non-informative prior (from  $-\infty$  to  $\infty$ ), the stationary condition of TAR parameters that the zeros of  $\Phi(B) = 0$  that lie outside the unit circle will be taken into account.

Parameters sampled from posterior distribution that lie outside the parameter space will be discarded, at which point another sampling from posterior distributions will be performed.

First, as a way of interpreting the effect from incorporating parameter uncertainty into the streamflows generation, basic statistics of generated streamflows were calculated and compared. Figures 5.2-5.3 and Figures 5.A1-5.A2 in Appendix illustrate quantile distributions of 5000 generated means, standard deviations, skewness coefficients, lag-1 serial correlations, and cross correlations. Each set of statistics is separately calculated in a trace of generated streamflows set for 3 different scenarios: (1) NU: natural uncertainty (no parameter uncertainty) (2) AS: asymptotic approach, and (3) BA: Bayesian approach. In Box plots of provided Figures, two upper, middle, and lower lines in the box are 75, 50, 25% quantile, respectively, the whisker from the box means 90, 10% quantile for each side, two dots outside box mean 99%, 1% quantile values, and 'X' means historical storage capacity. As expected, the parameter uncertainty effect on simulated statistics increases the variability especially when sample size is relative small ( $n = 50$ ). Increased variability of generated statistics can be still visible when  $n = 100$  (even when  $n = 200$ ) and therefore sample size of 100 might not be enough to neglect the effect of parameter uncertainty into the streamflows generation. Note that this is not the case of generated skewness coefficients since all calculated skewness show similar distributions with all sample sizes and sites.

A comparison between AS and BA in terms of expected values of generated statistics yielded little difference in the expected generated mean, lag-1 correlation coefficients, and cross correlations between simulated and historical samples. However,

increased upward bias of generated standard deviation for BA is notable for  $n = 50$ , which coincided with the result by Vales et al. (1977). This positive bias of generated standard deviations could not be found for AS, which might have been due to using the multivariate normal distribution in AS instead of the Wishart distribution in BA for the variance-covariance parameter matrix. Furthermore, BA shows a slightly wider variability of generated means for  $n=50$ , but AS and BA showed similar variability of distributions for all examined statistics once the available sample size is equal to or greater than 100.

Also performed was a comparison based on design variables of storage related statistics (storage capacity based on sequence peak algorithm) and drought related statistics (critical drought indices) by assuming that the design period  $N_d$  is equal to the historical sample size of 98. For more information regarding the definition and calculation of these statistics, see the previous chapter or Salas et al. (1980). Different fractions of MAF are employed as a threshold in the calculations of design variables where demand levels are assumed to be 100%, 80%, 60% MAF, and the same record length with historical flows is assumed to be a design period. Figure 5.4 represents quantiles of synthetic storage capacities ( $Sc$ ) and critical drought magnitude ( $cdm$ ), critical drought length ( $cdl$ ), and critical drought intensity ( $cdi$ ) with  $d = 60\%$  MAF calculated from generated annual streamflows for site 8.

Parameter uncertainty induces different patterns of estimated quantiles of design variables compared to the case of NU. Expected values of generated  $Sc$ ,  $cdm$ ,  $cdl$ , and  $cdi$  are shown to be greater than those calculated from the historical samples, particularly when the available sample is small ( $n=50$ ). A notable effect of parameter

uncertainty is that it results in increased upper quantiles generated for all of the design variables. Parameter uncertainty effects are still expected when  $n=100$  (even  $n=200$ ). Similar patterns of parameter uncertainty effects on quantiles of design variables are expected for different demand levels and sites. See quantile plots in Figures 5.A3-5.A10 Appendix.

Compared with AS, BA produces larger expected values and upper quantile estimates of  $Sc$  and  $cdm$ , especially for  $n=50$ , but this difference is inversely proportional to the available sample size. However, a difference remains between AS and BA in upper quantiles of design variables when  $n=100$ . As mentioned before, this difference results from using different parameter distributions of variance-covariance matrix of error terms in TAR(1) model.

Table 5.2 shows the related relative bias ( $RBIAS$ ) and relative root mean square errors ( $RRMSE$ ) of design variables for all three sites with  $d=100\%$  MAF (see the previous chapter for the definition of  $RBIAS$  and  $RRMSE$ ). The  $RBIAS$  and  $RRMSE$  are greater than the historical ones, which could be the effect of parameter uncertainty, and this effect is shown to decrease as sample size increases. More significant effects of parameter uncertainty on  $Sc$  and  $cdm$  are visible rather than  $cdl$  and  $cdi$ . The following example will be helpful to evaluate the difference between two approaches of AS and BA, which is demonstrated in Table 5.3. Consider  $RRMSE$  of generated storage capacities for the case with the demand level ( $d$ ) of 80% MAF at site 8. Note that  $RRMSE$  is the function of variance and square root of  $RBIAS$ . It is shown that BA (258% of  $RRMSE$ ) gives 68% more  $RRMSE$  than AS (190% of  $RRMSE$ ) when  $n=50$ , but that difference between AS and BA is getting to 8% of

*RRMSE* when  $n = 200$  (129% of *RRMSE* for BA, 121% of *RRMSE* for AS). It is clear that the reduction of *RRMSE* for  $Sc$  with larger sample size is also valid for  $cdm$ ,  $cdl$ , and  $cdi$

Additionally, the parameter uncertainty effect is also considered with the different demand levels. The previous chapter shows that patterns of bias (upward or downward) and magnitude of variability of generated design variables can be different depending on the assumed demand levels, even if parameter uncertainty is not incorporated. In the present application, it seems that a smaller demand level causes larger bias and larger variability of generated design variables. For this reason, the comparison of AS and BA in terms of the ratio relative to NU is appropriate. The difference of *RRMSE* between AS and BA in terms of the ratio relative to NU is shown to be dependent of different demand levels: e.g., for  $n=50$  the difference in increased *RRMSE* of AS and BA relative to NU regarding  $Sc$  is 7% (51%-44%) for  $d=100\%$  MAF, 15% (58%-43%) for  $d=80\%$  MAF, and 17% (56-39%) for  $d=60\%$  MAF, which means that the BA reveals more variability of  $Sc$  for a smaller  $d$ .

Another example is given for *RRMSE* of  $cdm$ , which shows the similar pattern with  $Sc$ . Compared with 120% in NU for  $d=60\%$  and  $n=50$ , AS shows 191% of *RRMSE* and BA shows 250% of *RRMSE* which corresponds to 37% increased *RRMSE* for AS and 52% increased *RRMSE* for BA compared with NU (15% difference in increased *RRMSE* between AS and BA). However, the difference in increased *RRMSE* between AS and BA reduces to 8% when  $n=200$ . Also, different demand levels shows the range of *RRMSE* between AS and BA relative to NU by 10-15%. Similar patterns between  $Sc$  and  $cdm$  are to be expected for the reason

that those two statistics are conceptually related with the largest deficit over the design period. Similar patterns may be found of increased *RRMSE* by parameter uncertainty in *cdl* and *cdi*, but the magnitude of *RRMSE* which ranges 13%-34% for *cdl* and 33%-56% for *cdi* is not as much as in the case of *Sc* and *cdm*. See Tables 5.A1-5.A3 in the Appendix for more information regarding calculated *RBIAS* and *RRMSE* of generated design variables for all sites.

To summarize the comparative analysis of AS and BA for different sample sizes and demand levels in terms of *RRMSE* of generated design variables, BA results the larger *RRMSE* (up to 11%) than AS when  $n=100$  but BA will be closer to AS when  $n \approx 200$ . That is, at least  $n \geq 200$  would be required to ensure the applicability of the asymptotic approach and use of the Bayesian approach for  $n < 200$  would be recommended.

### **5.3.2 Comparative analysis of univariate and multivariate generation**

In order to evaluate the parameter uncertainty effect associated with the appropriate model selection, an additional simple univariate generation is utilized and a comparison will be provided. Parameter uncertainty effect on the main statistics and design variables based on TAR(1) model generation will be compared with the case of the univariate generation where streamflows of two sites are independently generated by the univariate AR(1) model. Depending on parameter uncertainty, consideration for four different cases for models are exemplified, which are (1) NUu: univariate generation without parameter uncertainty (natural uncertainty of model), (2) BAu: univariate

generation with parameter uncertainty incorporated based on Bayesian posterior distribution, (3) NUM: trivariate generation without parameter uncertainty, and (4) BAM: trivariate generation with parameter uncertainty incorporated based on Bayesian posterior distribution.

Basic statistics for each of the available sample sizes in the 4 unique cases are shown in Figures 5.A11-5.A13 in Appendix. Almost identical distributions of generated statistics are expected between univariate and multivariate generations for the cases without parameter uncertainty, which can be supported by similar estimates of lag-1 serial correlation coefficients at each site when taking either the univariate and trivariate estimation as shown in Table 5.1. However, increases in the expected value and variability of generated standard deviation are notable for small sample size ( $n=50$ ) in the trivariate case when parameter uncertainty is included in the generation of streamflows. Figure 5.5 illustrates the quantile distributions of generated design variables with  $d=60\%$  MAF for site 16 (see Figures 5.A14-5.A24 in the Appendix for other sites and demand levels). Note that the uncertainty incorporation effect with TAR(1) model is more significant on increased bias and upper quantiles of design variables than one in the univariate model. It is probably associated with the increased number of parameters or other factors. For example, currently there are 18 parameters for trivariate AR(1) model compared with 9 parameters for two univariate AR(1) model, and uncertainty associated with more parameters might result an increase in the variability of design variables. This increase of variability will probably be more dominant for smaller demand levels. Calculated *RBIAS* and *RRMSE* of generated design variables would be helpful in recognizing the differences between the four cases

as was discussed in this section. See Tables 5.A4-5.A6 in the Appendix for calculated *RBIAS* and *RRMSE* for generated design variables.

### **5.3.3 Parameter uncertainty associated with transformation**

Even though the initial normality assumption of historical streamflows at all three sites in the present application would be acceptable according to a skewness applied with 5% significance levels, in reality the sample skewness coefficients that range between 0.21-0.34 show some skewness in the historical sample. Practically, most historical streamflows (even annual) are not exactly normally distributed (they have some skewness) and some transformations might be required to treat this non-zero skewness to ensure the normality assumption that is fundamentally required for our models. A power transformation is selected from several possible transformation methods and employed to the present historical streamflows sequences at all three sites (non-transformation and power transformation will be referred to Case 1 and 2, respectively). A different parameter uncertainty effect might be expected due to the different treatment of the normality of data and will be discussed in this section. Case 2 produces similar distributions of basic statistics (as observed in Case 1 and Case 2) and demonstrates superior performance in preserving the historical skewness in the generated streamflows (See Figures 5.A25-5.A27). Figure 5.6 demonstrates quantile distributions of generated design variables with  $d=60\%$  MAF for site 16 in Case 2 (see Figures 5.A28-5.A29). Additionally, see the Appendix for other sources of information. A comparison based on Figures 5.5 and 5.6 illustrates Case 2 yields similar patterns as in Case 1 with regard

to parameter uncertainty effects on quantiles of design variables for the univariate generation. Overall, Case 2 produces less variability of the quantile distributions than that which is found in Case 1 without parameter uncertainty. Therefore, the elimination of skewness of historical data would reduce the natural uncertainty of the streamflows generation. Case 2 would be expected to have less variability of design variables than Case 1 when parameter uncertainty is incorporated, which could be a consequence of the reduced natural uncertainty. In both Case 1 and Case 2, the trivariate generation shows the similar variability of quantile distributions of design variables with the univariate generation when parameter uncertainty is not incorporated. However, when parameter uncertainty is taken into account, the reduced variability of quantile estimates for the trivariate generation could be expected in Case 2, but not in Case 1. This difference between two Cases could be explained by manipulating the historical skewness and consequent impacts in the trivariate generation. Since the parameter uncertainty consideration is theoretically based on the normal assumption of sample, the existing skewness of historical data (even if that skewness is not shown to be significantly different from zero by the skewness test) could be in some cases more effective on quantile distributions of design variables; moreover, its impact might be boosted with the incorporation of parameter uncertainty. In such an example, allowing for the elimination of the historical skewness by transformation reduces the associated parameter uncertainty effect in the case of multivariate generation.

Table 5.4 gives a brief summary of the parameter uncertainty effect on the storage capacity and critical drought magnitude calculated from synthetic annual streamflows at site 8 generated by using trivariate AR(1) model with power

transformation applied. Generated mean, standard deviation, and quantiles of design variables are shown for the Cases of NUm (without parameter uncertainty) and BAm (with parameter uncertainty). Compared with NUm, BAm shows about 25% and 31% increased expected values of generated storage capacities and about 93% and 139% increased standard deviations of generated storage capacities for demand levels of 100% MAF and 80% MAF, respectively. The parameter uncertainty shows the increased quantiles with a range from about 45% to 116%, and more increased quantiles of storage capacity would be expected by parameter uncertainty especially for the larger quantile. Overall, less increase by the parameter uncertainty effect is demonstrated in the generated critical drought magnitude compared with the generated storage capacity, except that similar pattern of increased statistics and quantiles by parameter uncertainty over different demand levels and sample serial correlations are reported for both design variables.

#### **5.4. Summary and Conclusion**

Parameter uncertainty has been embedded into the multivariate generation based on the multivariate AR model by using asymptotic and posterior distributions of maximum likelihood estimators. Increased variance of generated streamflows at each site has been found, which would be associated with the effect of parameter uncertainty as discussed by Valdes, et al. (1977). Compared to traditional generation without parameter uncertainty incorporation, parameter uncertainty creates an increase in variability of quantile distributions of basic statistics. Furthermore, design variables of

storage capacity and critical drought indices have been utilized to evaluate the parameter uncertainty effect in streamflows generation and thus increases expected values of design variables and increases the variability of quantile distributions of design variables as well. The parameter uncertainty effect remains visible when sample size is equal to 100, and also with a sample size of 200. Therefore, when employing the multivariate generation into the real hydrologic application using a small sample size, the incorporation of parameter uncertainty into the streamflows generation might be useful to improve the reliability of design variables.

A comparative analysis of the asymptotic and Bayesian approaches has been performed based on an analysis using generated basic statistics and design variables. Compared with the asymptotic approach, the Bayesian approach produces synthetic streamflows with bigger standard deviations as well as associated design variables with upward shifted expected values and larger variability, which is more visible when the small sample size is equal or less than 100 and is of a smaller demand level. When the sample size approaches 200, the results from Bayesian approach become identical to those from asymptotic approach.

In general, sample sizes that are large enough for the posterior distribution will be defined by only the sample structure rather than it depending on its prior. Indeed, an applied posterior distribution of AR parameters that is derived from the non-informative prior has been shown to be equivalent to an asymptotic distribution. On the other hand, different distributions of variance-covariance for error terms of TAR(1) models based on asymptotic and Bayesian approaches are available which result in an increased standard deviation of generated streamflows and increased variability in quantile distributions of

design variables.

Theoretically, more precise posterior distributions can be obtained if the true prior of parameters is available, but this is not the case unless the exact prior is given. Thus, the modified posterior distribution with bounded parameter spaces seems appropriate for incorporation of parameter uncertainty into the streamflows generation. Use of the modified posterior distribution would be preferred for the real hydrologic planning and design over asymptotic distribution since there would be significant variability generated by the Bayesian approach as it is applied to a small sample size. Such an approach to ensure the reliability of streamflows generation might not be negligible.

Skewness of real data might significantly effect the variability of design variables generated from the multivariate model with parameter uncertainty incorporated. In which case, a proper elimination of skewness of the real data might be useful in removing any unexpected increases in the variability of generated design variables.

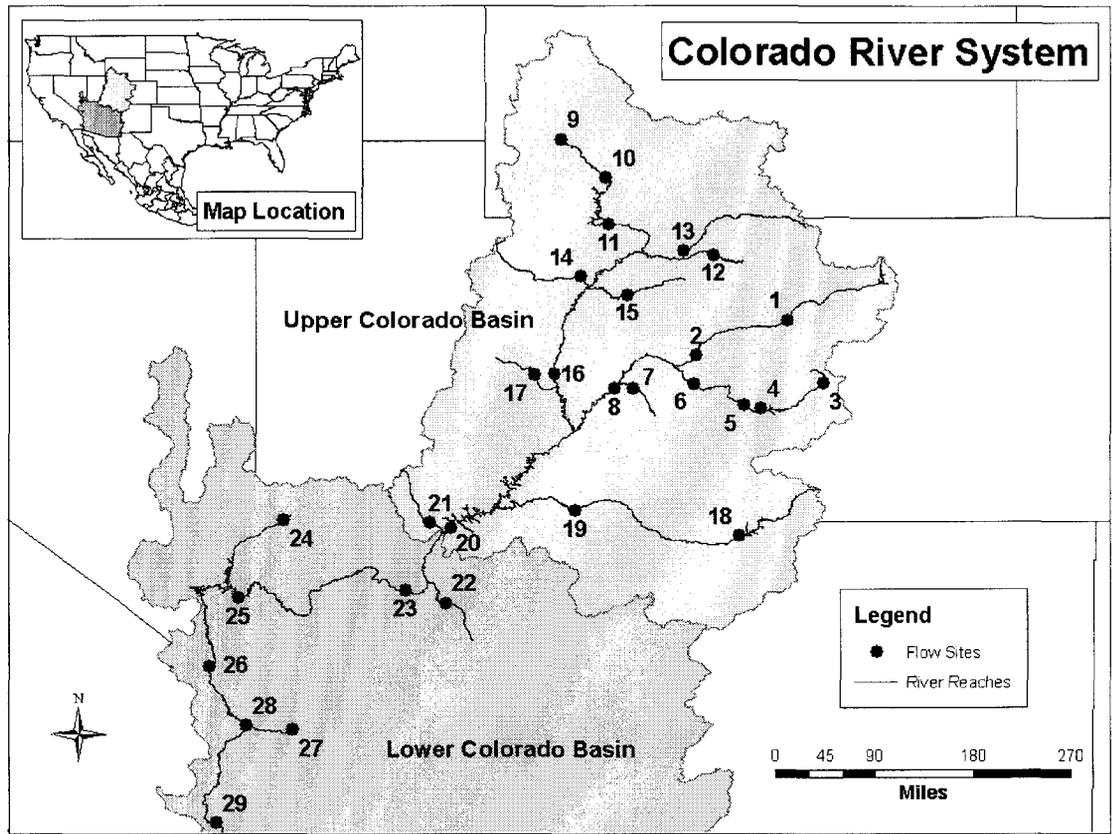


Figure 5.1: Colorado River basin gauging stations (USBR, 2007)

Table 5.1 Basic statistics of applied annual streamflows data in Colorado River basin

	Site 8	Site 16	Site 19
Sample mean (ac-ft)	6,824,630	5,415,641	2,148,418
Sample standard deviation (ac-ft)	1,965,947	1,642,136	883,149
Sample coefficient of variation	0.29	0.30	0.41
Sample skewness coefficient	0.21	0.34	0.31
Sample kurtosis coefficient	2.63	2.87	2.38
Sample lag-1 correlation	0.288	0.306	0.116
Estimated AR parameters (based on univariate ML estimator)	0.287	0.304	0.115
Estimated lag-1 correlation (based on TAR(1))	0.280	0.295	0.109
<hr/>			
Sample cross correlation	Site 8	Site 16	Site 19
Site 8	1	0.855	0.766
Site 16	0.855	1	0.562
Site 19	0.766	0.562	1

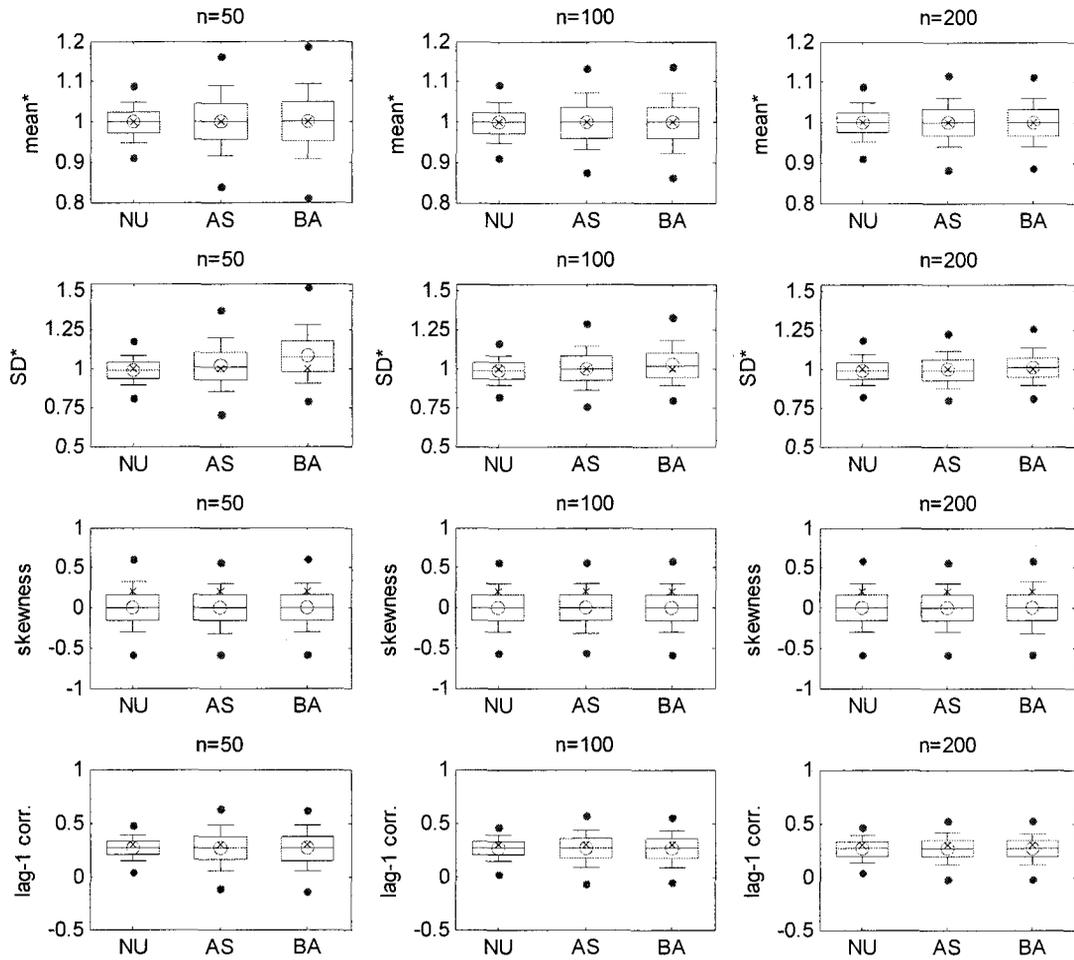


Figure 5.2: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 8) with parameter uncertainty incorporated where 'NU', 'AS', BA' mean different parameter uncertainty consideration: 'NU' means natural uncertainty, 'AS' means using asymptotic distributions, and 'BA' means using posterior distribution. Superscript \* notes "scaled by historical statistics" and 'X' denotes historical statistics.

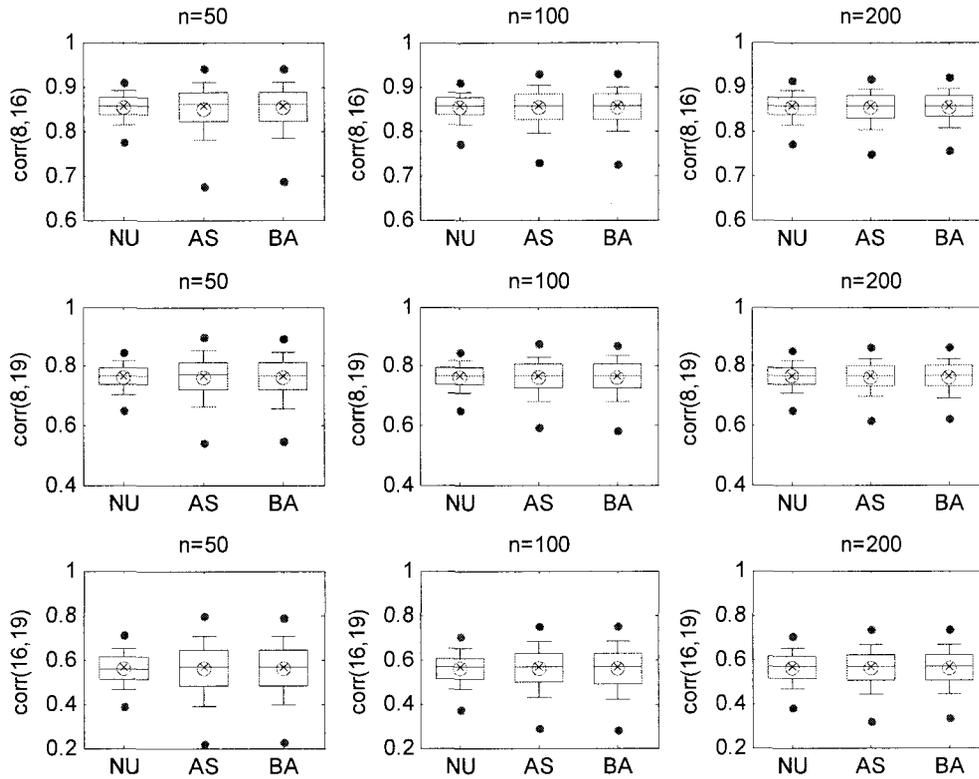


Figure 5.3: Distributions of cross correlations calculated from 5000 different generated annual streamflows sets with parameter uncertainty incorporated where 'NU', 'AS', 'BA' mean different parameter uncertainty consideration: 'NU' means natural uncertainty, 'AS' means using asymptotic distributions, and 'BA' means using posterior distribution.

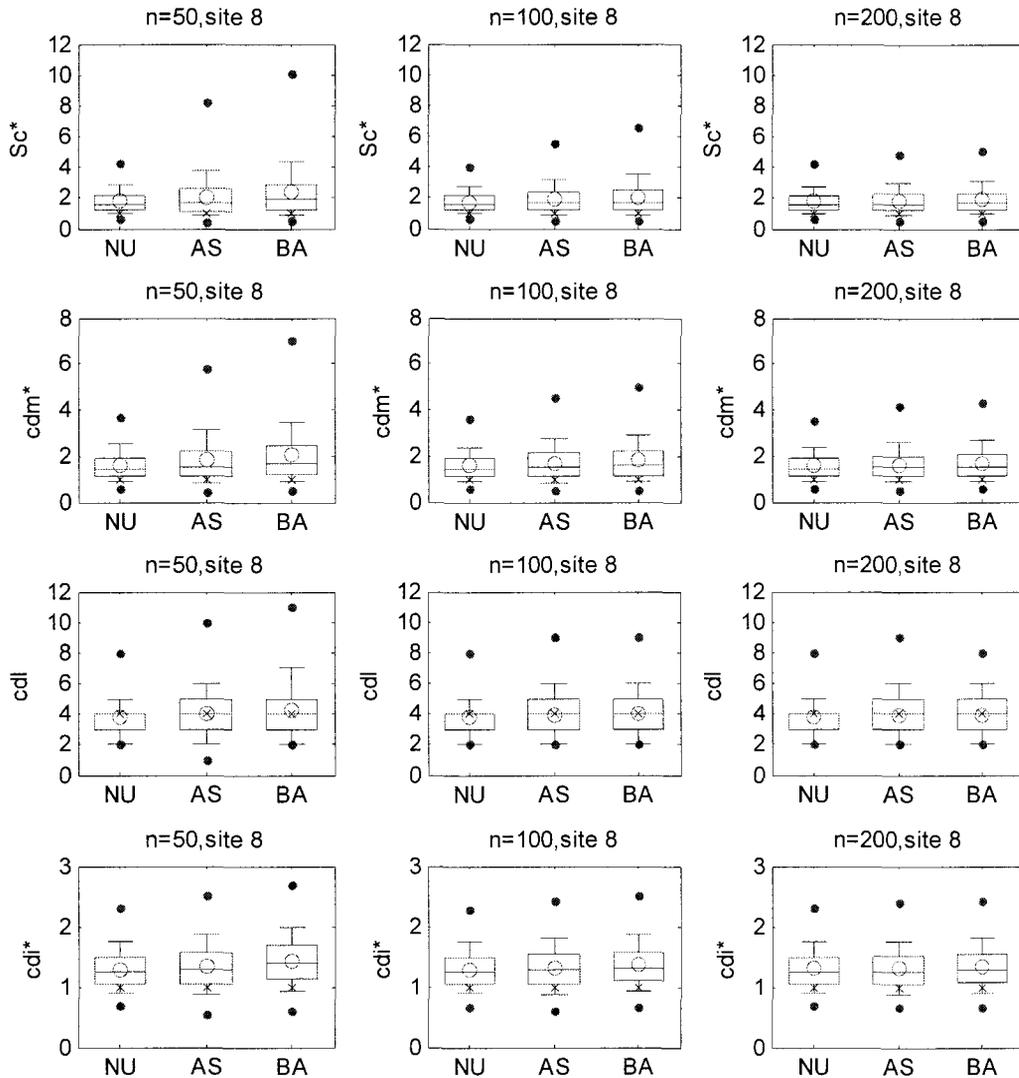


Figure 5.4: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 8 given demand level of 80% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

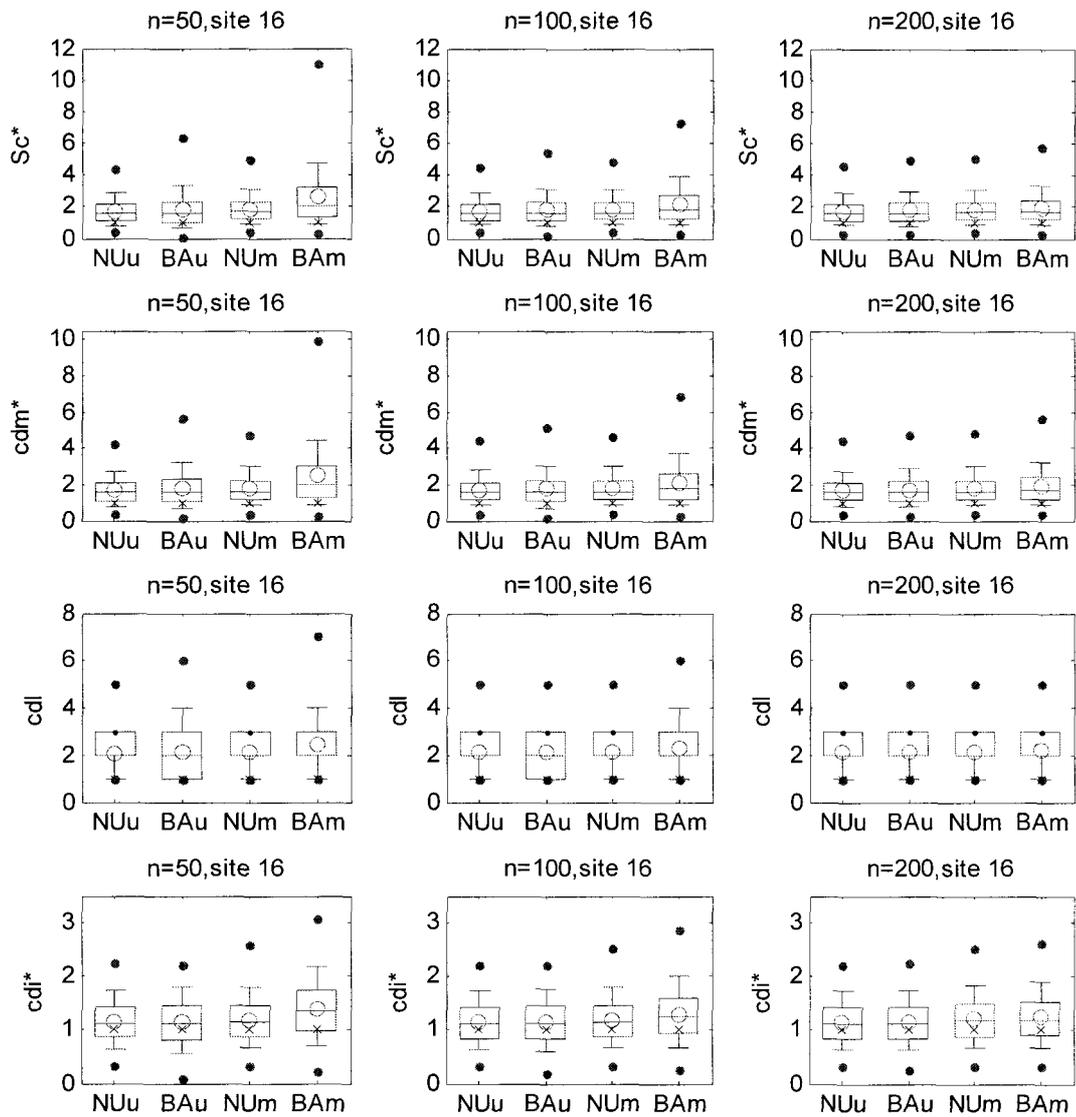


Figure 5.5: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

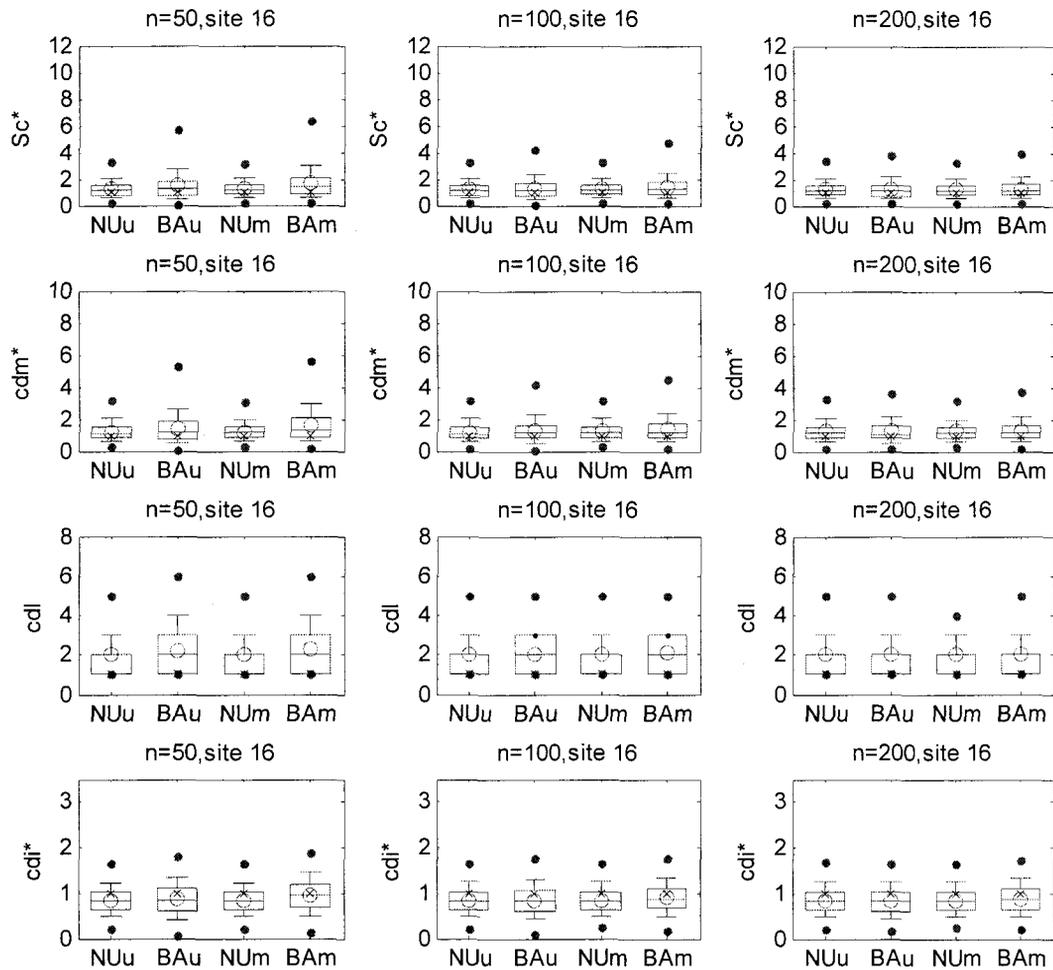


Figure 5.6: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

Table 5.2: *RBIAS* and *RRMSE* of generated design variables based on different parameter uncertainty consideration methods (demand level = 100%)

Sample size	Site 8			Site 16			Site 19			
	NU	AS	BA	NU	AS	BA	NU	AS	BA	
Storage Capacity, <i>Sc</i>										
Rbias	50	-10	8	14	-3	17	22	-24	-10	-7
(%)	100	-11	-1	2	-3	7	11	-24	-17	-18
	200	-11	-7	-6	-4	1	2	-24	-21	-23
rrmse	50	43	77	87	46	86	95	43	63	68
(%)	100	43	60	64	46	67	70	43	52	53
	200	44	53	52	46	56	57	42	48	47
Critical drought magnitude, <i>cdm</i>										
Rbias	50	50	66	76	40	55	63	27	37	45
(%)	100	49	58	63	39	46	51	26	31	33
	200	50	52	55	39	42	44	26	28	28
rrmse	50	73	109	129	63	101	121	50	75	89
(%)	100	72	91	97	63	81	86	49	61	65
	200	73	80	84	63	71	74	49	55	55
Critical drought length, <i>cdl</i>										
rbias	50	48	55	55	26	31	30	-5	-2	-3
(%)	100	49	53	53	26	28	29	-6	-4	-5
	200	50	50	50	25	27	27	-6	-5	-7
rrmse	50	67	84	85	47	63	65	30	37	40
(%)	100	68	77	77	47	55	56	30	33	34
	200	69	71	72	47	50	51	30	32	31
Critical drought intensity, <i>cdi</i>										
rbias	50	6	9	16	26	29	38	34	38	46
(%)	100	6	7	11	25	27	31	35	36	40
	200	6	6	8	26	26	28	35	35	37
rrmse	50	23	28	34	37	42	52	45	52	60
(%)	100	23	26	28	37	40	44	46	48	53
	200	24	24	26	38	38	40	46	48	49

Table 5.3: Example of comparison of  $RRMSE$  (%) of generated design variables based on different parameter uncertainty consideration methods (site 8)

sample size	Demand level, $d$								
	100% MAF			80% MAF			60% MAF		
	NU	AS	BA	NU	AS	BA	NU	AS	BA
<b>Storage Capacity, <math>S_c</math></b>									
50	43	77 (44)	87 (51)	109	190 (43)	258 (58)	125	205 (39)	285 (56)
100	43	60 (28)	64 (33)	105	140 (25)	160 (34)	118	156 (24)	181 (35)
200	44	53 (17)	52 (15)	107	121 (12)	129 (17)	119	135 (12)	146 (18)
<b>Critical drought magnitude, <math>cdm</math></b>									
50	73	109 (33)	129 (43)	89	139 (36)	177 (50)	120	191 (37)	250 (52)
100	72	91 (21)	97 (26)	85	110 (23)	123 (31)	115	150 (23)	171 (33)
200	73	80 (9)	84 (13)	86	97 (11)	103 (17)	116	129 (10)	141 (18)
<b>Critical drought length, <math>cdl</math></b>									
50	67	84 (20)	85 (21)	33	46 (28)	50 (34)	127	159 (20)	185 (31)
100	68	77 (12)	77 (12)	32	38 (16)	39 (18)	125	143 (13)	151 (17)
200	69	71 (3)	72 (4)	33	36 (8)	36 (8)	125	132 (5)	138 (9)
<b>Critical drought intensity, <math>cdi</math></b>									
50	23	28 (18)	34 (32)	45	52 (13)	62 (27)	56	68 (18)	80 (30)
100	23	26 (12)	28 (18)	45	50 (10)	53 (15)	55	63 (13)	66 (17)
200	24	24 (0)	26 (8)	46	47 (2)	50 (8)	56	59 (5)	62 (10)

Note value in the parenthesis below calculated  $RRMSE$  denotes the ratio(%) of  $RRMSE$  based on corresponding approach relative to  $RRMSE$  based on NU.

Table 5.4: Example of generated storage capacity and critical drought magnitude (scaled by MAF) ( $n=50$ ,  $N_d=98$ , Bayesian analysis, site 8, power transform applied)

Demand level Case	MAF		0.8MAF	
	NUm	BAm	NUm	BAm
Storage capacity				
mean	4.0	5.0 24.5%	0.9	1.2 30.5%
SD	1.9	3.6 92.6%	0.4	0.9 138.8%
$q_{0.9}$	6.6	9.9 50.4%	1.4	2.0 45.3%
$q_{0.95}$	7.6	12.0 59.3%	1.6	2.5 61.1%
$q_{0.99}$	9.7	17.0 74.1%	2.1	4.4 116.0%
Critical drought magnitude				
mean	2.1	2.4 15.5%	0.8	1.0 23.4%
SD	0.7	1.3 79.6%	0.3	0.6 84.6%
$q_{0.9}$	3.0	3.9 27.9%	1.2	1.7 37.6%
$q_{0.95}$	3.4	4.7 38.4%	1.4	2.0 45.6%
$q_{0.99}$	4.4	7.0 60.3%	1.8	3.2 73.5%

*Note* NUm: no parameter uncertainty considered (natural uncertainty), BAm: parameter uncertainty incorporated to trivariate AR(1) model (Bayesian), SD: standard deviation,  $q_{0.9}$ ,  $q_{0.95}$ ,  $q_{0.99}$  means 90%, 95%, and 99% quantile, respectively. Value (%) in the column of BAm represents the ratio of increased storage capacity(or critical drought magnitude in BAm with respect to NUm. (parameter uncertainty effect)

## References

- Camacho, F., A.I. McLeod, and K.W. Hipel (1985). Contemporaneous autoregressive-moving average (CARMA) modeling in water resources, *Water Resources Bulletin* 21(4), pp. 709-720.
- Camacho, F., A.I. McLeod, and K.W. Hipel (1987). Multivariate contemporaneous ARMA model with hydrological applications, *Stochastic Hydrology and Hydraulics*, 1, pp. 141-154.
- Fiering, M.B. (1964). Multivariate technique for synthetic hydrology, *Journal of Hydraulics Division, ASCE* 90(HY5), pp. 844-850.
- Hipel, K.W. and A.I. McLeod (1994). *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier.
- Lettenmaier, D.P. (1980). Parameter estimation for multivariate streamflow synthetics, *Proceeding of Joint Automatic COntrol Conference, San Francisco, Paper FA6-D*.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd edition, Springer-Verlag, Berlin-Heidelberg, Germany.
- Matalas, N.C. (1967). Mathematical assessment of synthetic hydrology, *Water Resources Research*, 3(4), pp. 931-945.
- McLeod, A.I. and K.W. Hipel (1978). Simulation procedures for Box-Jenkins models, , *Water Resources Research*, 14(5), pp. 969-975.
- Mejia, J.M. and J. Roussell (1976). Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), pp. 185-186.
- O'Connell, P.E. (1974). *Stochastic Modeling of Long-Term Persistence in Streamflow Sequences*, Ph.D Dissertation, Civil Engineering Dept., Imperial College of Science and Technology, London.
- Pegram, G.G.S. and W. James (1972). Multilag multivariate autoregressive model for the generation of operational hydrology, *Water Resources Research*, 8(4), pp. 1074-1076.
- Salas, J.D. and G.G.S. Pegram (1977). A seasonal multivariate multilag autoregressive model in hydrology, *Proceeding of the Third International Symposium of Theoretical and Applied Hydrology*, Colorado State University, Fort Collins, Colorado.

- Salas, J.D., J.W. Deller, V. Yevjevich, and W. L. Lane (1980). Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado.
- Salas, J.D., G. Tabios, and P. Bartolini (1985) Approaches to multivariate modeling of water resources time series, Water Resources Bulletin, 21(4), pp. 683-708.
- Stedinger, J.R. and M.R. Taylor (1982). Synthetic streamflow generation: 2. Effect of parameter uncertainty, Water Resources Research, 18(4), pp. 919-924.
- Stedinger, J.R., D.P. Lettenmaier, and R.M. Vogel (1985). Multisite ARMA(1,1) and disaggregation models for annual streamflow generation, Water Resources Research, 21(4), pp. 497-509.
- USBR, Department of Interior (2007). Development of Stochastic Hydrology for the Colorado River System.
- Valdes, J.B., I. Rodriguez-Iturbe, and G.J. Vicens (1977). Bayesian generation of synthetic streamflows; 2. The multivariate case, Water Resources Research, 13(2), pp. 291-295.
- Valencia, D. and J.C. Schaake, Jr. (1973). Disaggregation processes in stochastic hydrology, Water Resources Research, 9(3), pp. 580-585, 1973.
- Vicens, G.J., I. Rodriguez-Iturbe, and J.C. Schaake (1975). Bayesian generation of synthetic streamflows, Water Resources Research, 11(6), pp. 827-838.
- Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.

## Appendix 5.A: Additional Figures and Tables

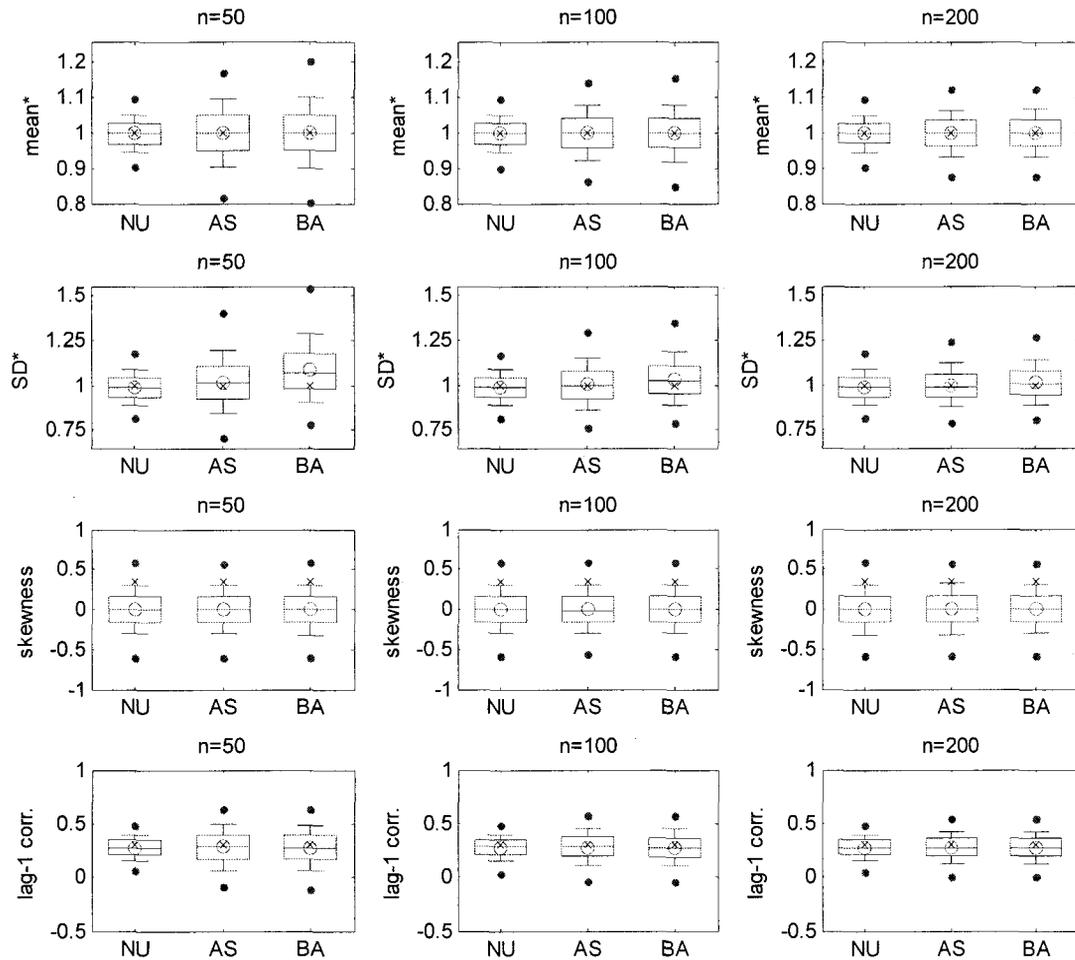


Figure 5.A1: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 16) with parameter uncertainty incorporated where 'NU', 'AS', BA' mean different parameter uncertainty consideration: 'NU' means natural uncertainty, 'AS' means using asymptotic distributions, and 'BA' means using posterior distribution. Superscript \* notes "scaled by historical statistics" and 'X' denotes historical statistics.

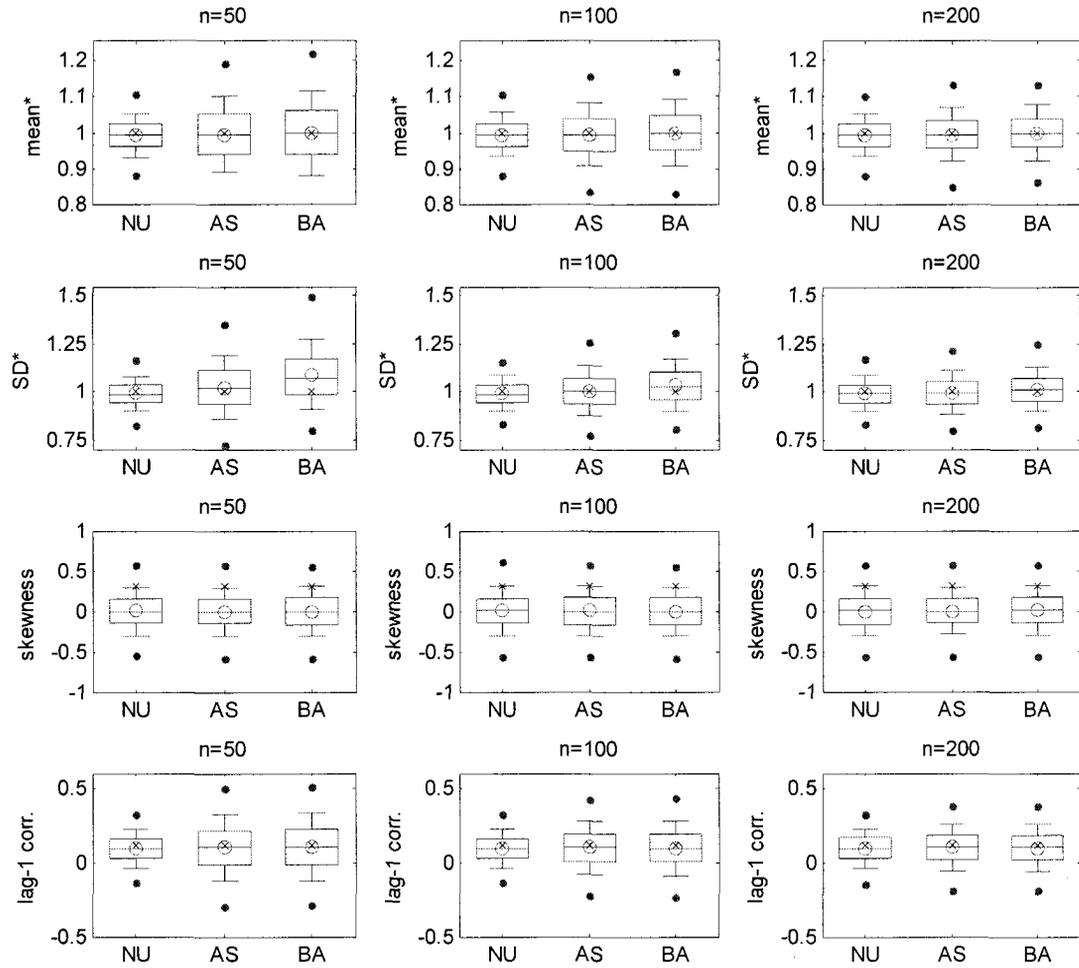


Figure 5.A2: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 19) with parameter uncertainty incorporated where 'NU', 'AS', BA' mean different parameter uncertainty consideration: 'NU' means natural uncertainty, 'AS' means using asymptotic distributions, and 'BA' means using posterior distribution. Superscript \* notes "scaled by historical statistics" and 'X' denotes historical statistics.

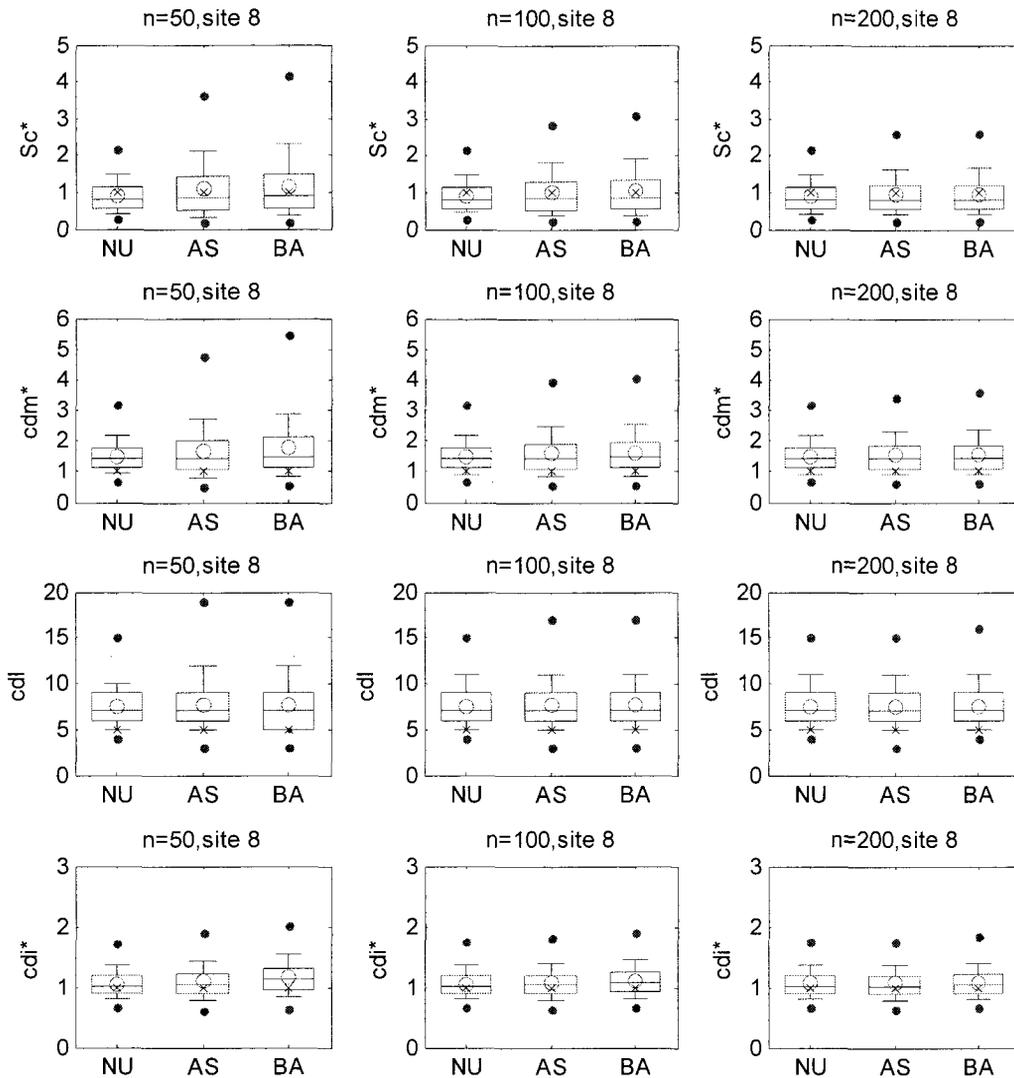


Figure 5.A3: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 8 given demand level of 100% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

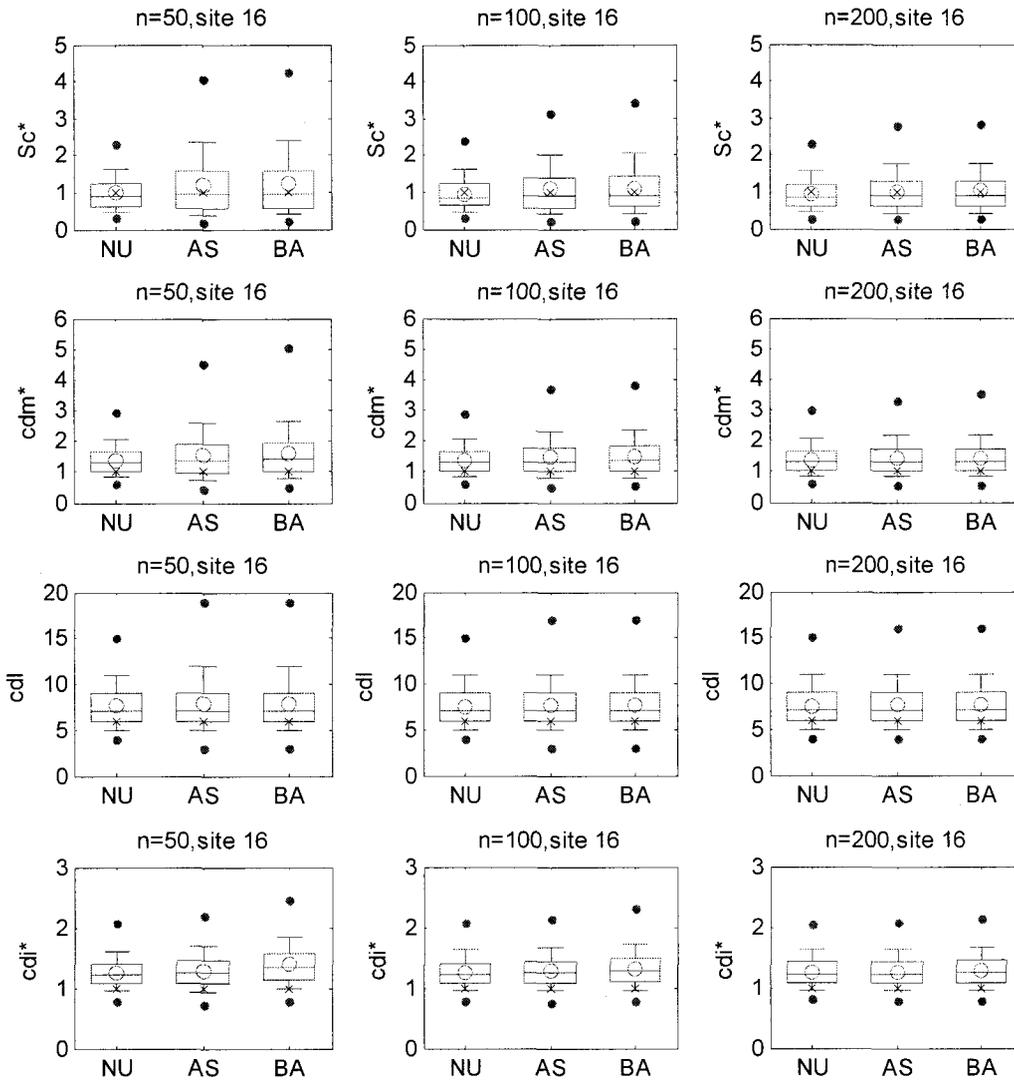


Figure 5.A4: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 100% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

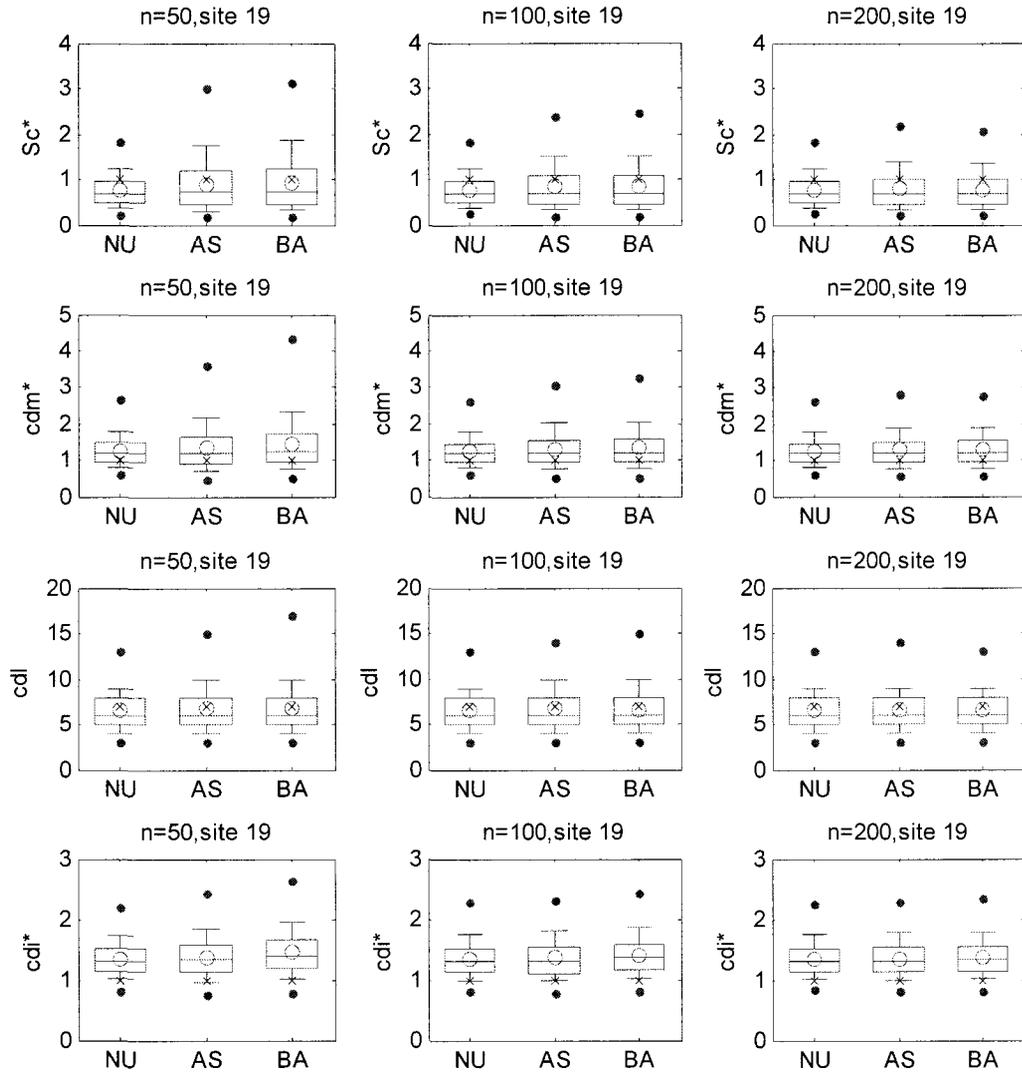


Figure 5.A5: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 100% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

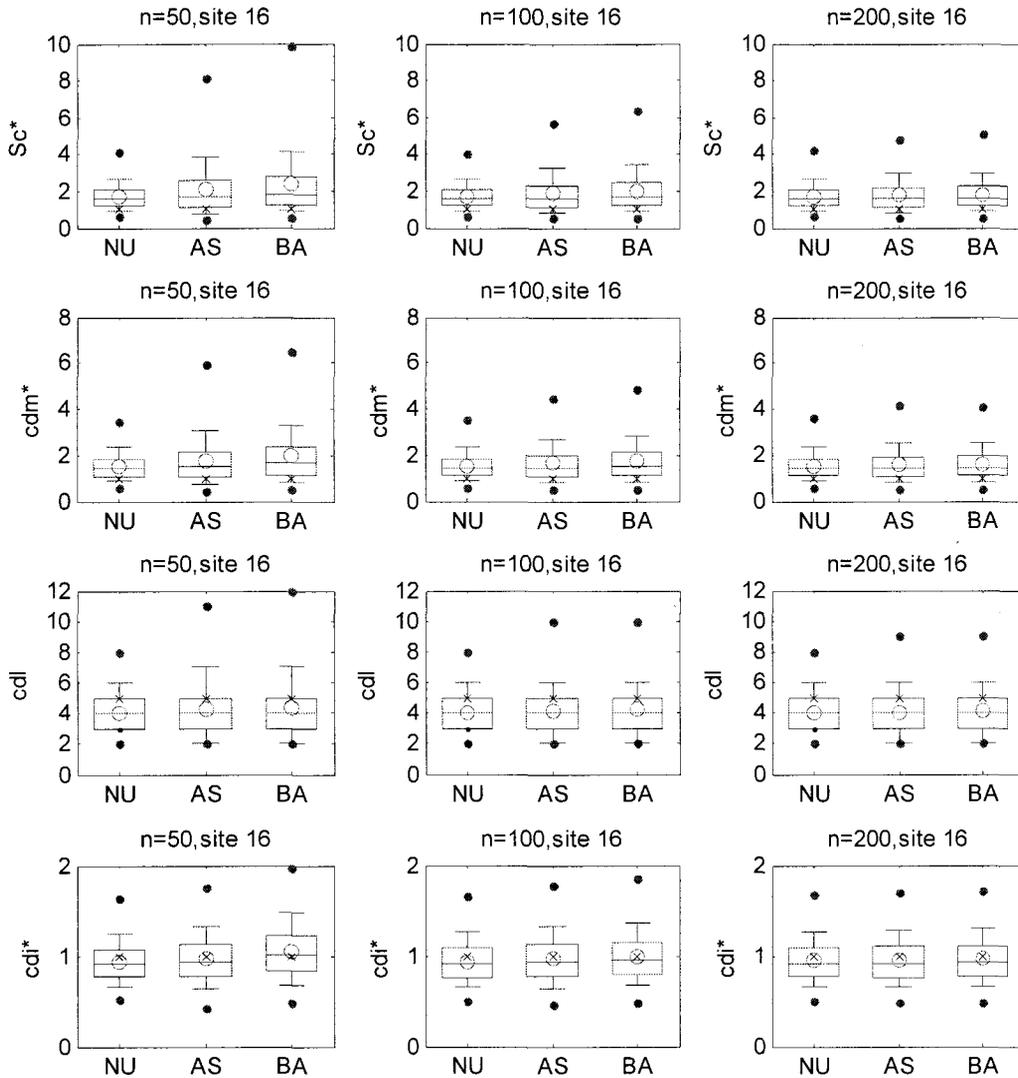


Figure 5.A6: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 80% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

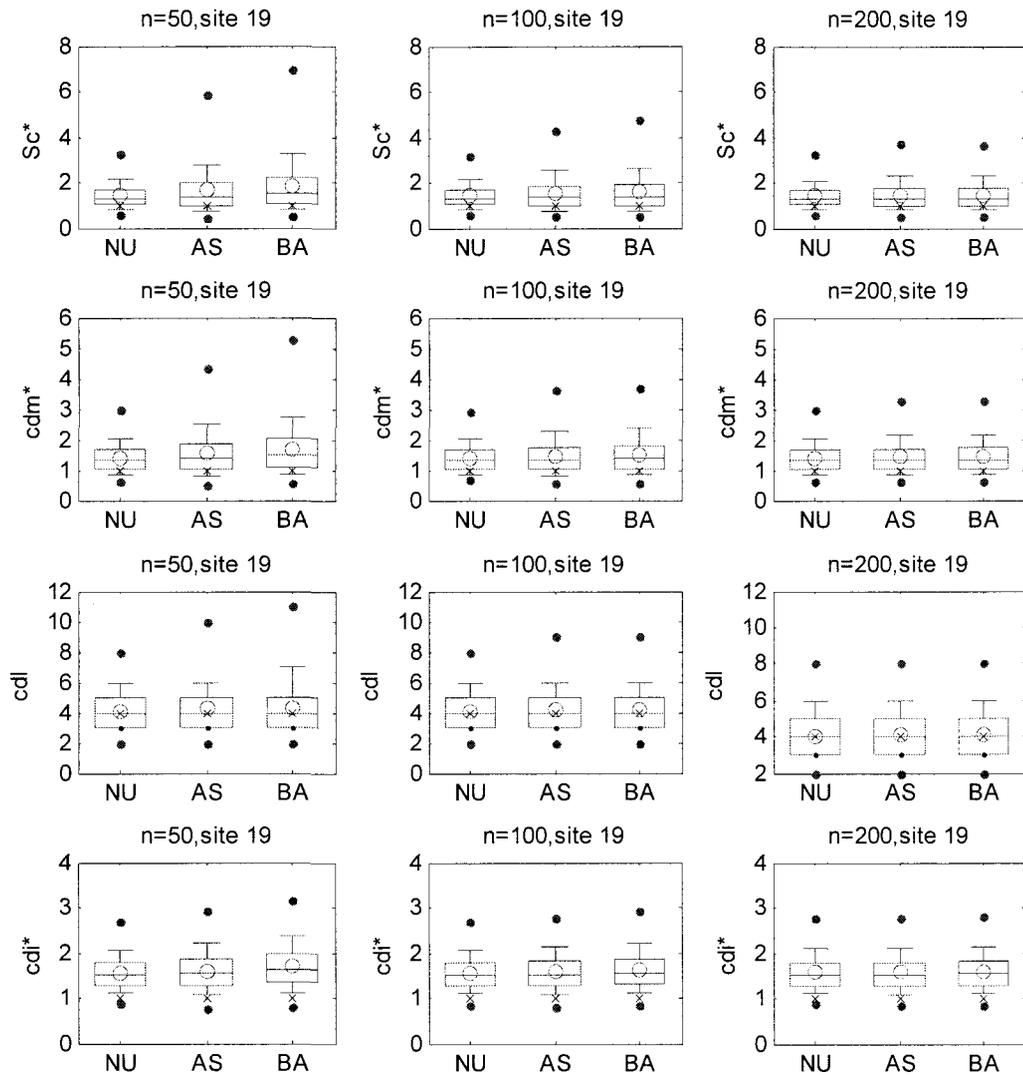


Figure 5.A7: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 80% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

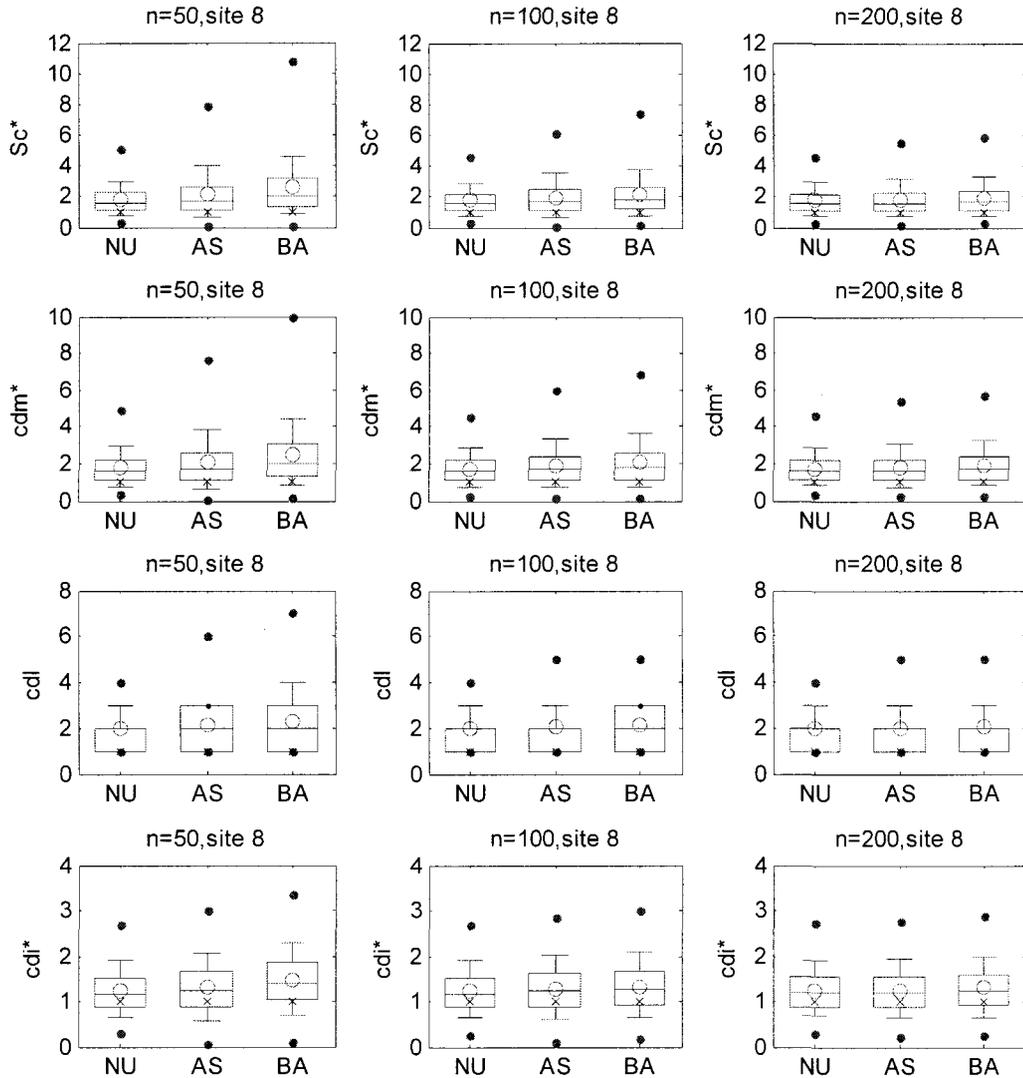


Figure 5.A8: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 8 given demand level of 60% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

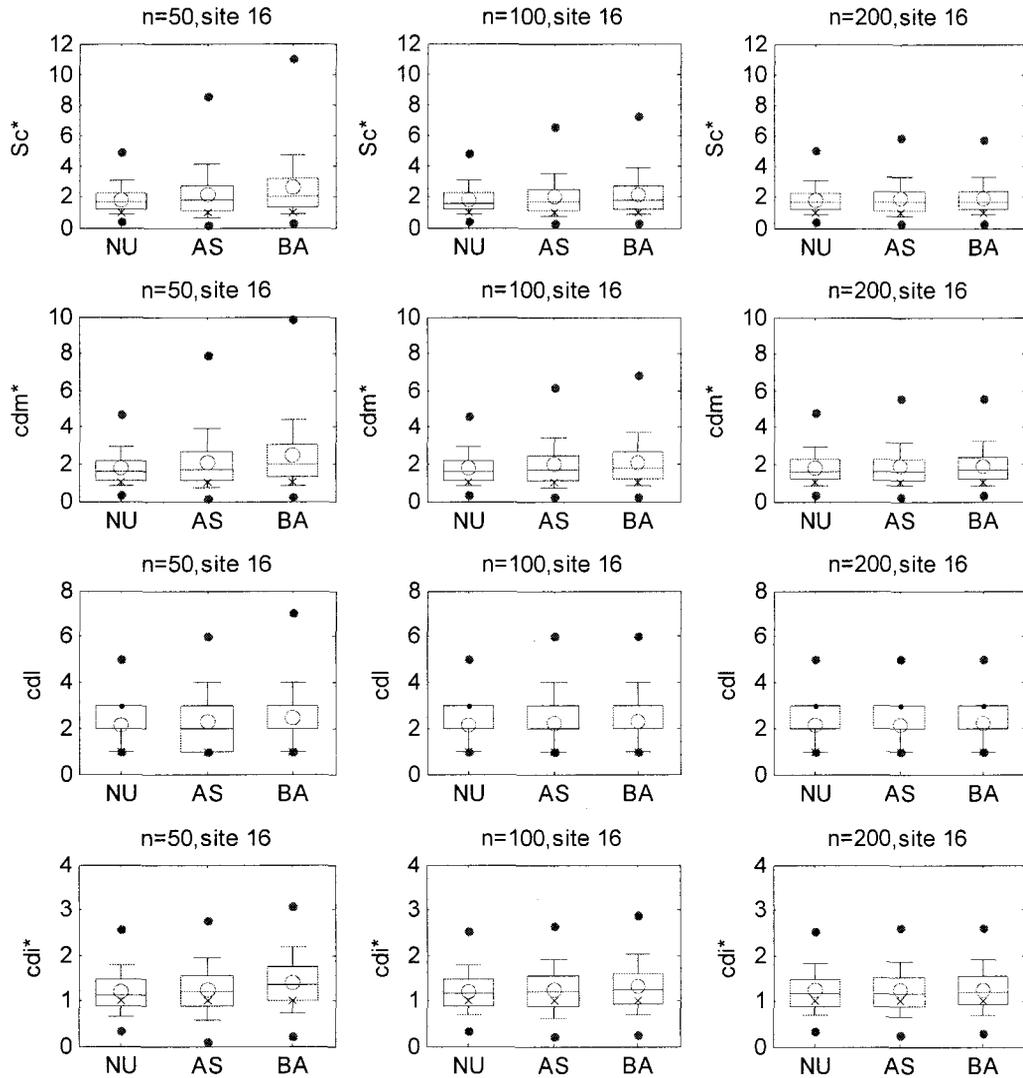


Figure 5.A9: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 60% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

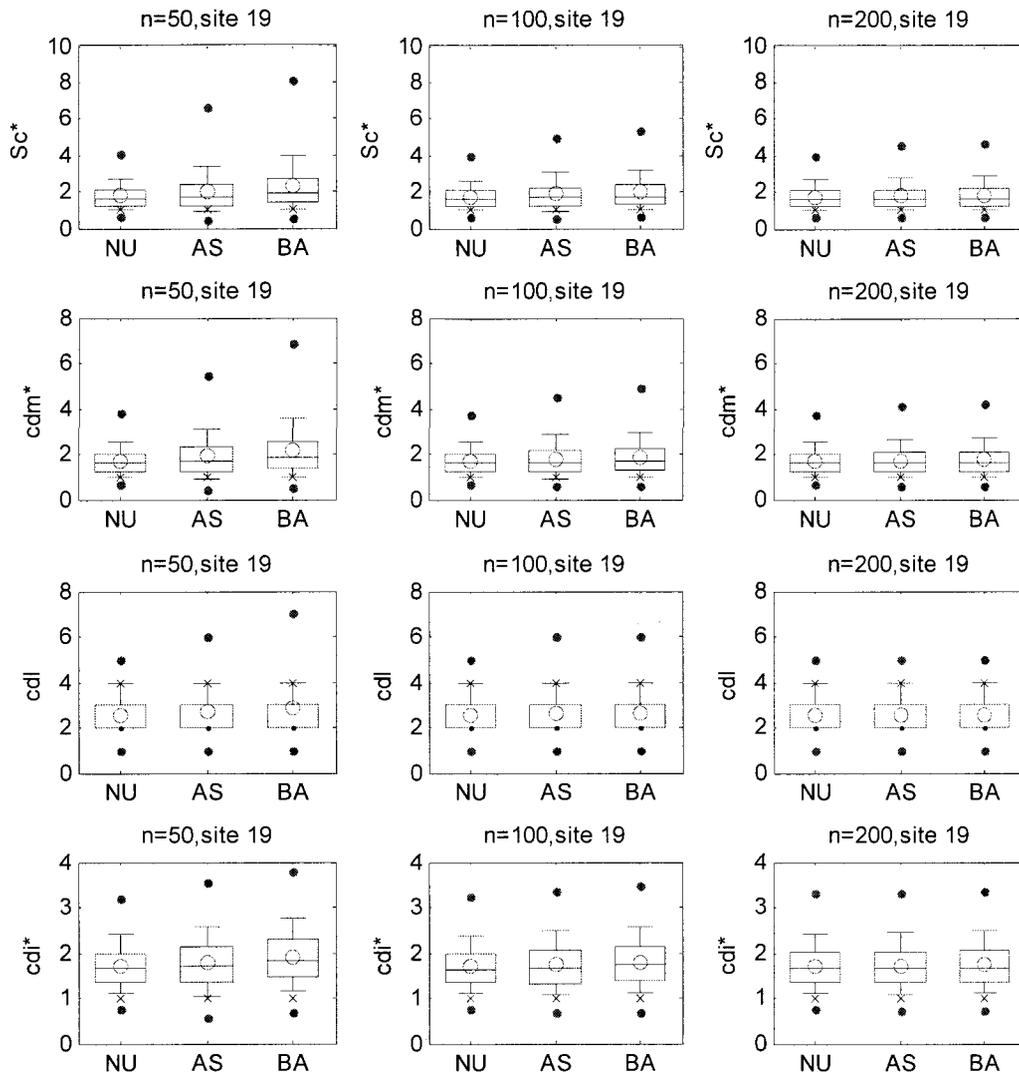


Figure 5.A10: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 60% MAF. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

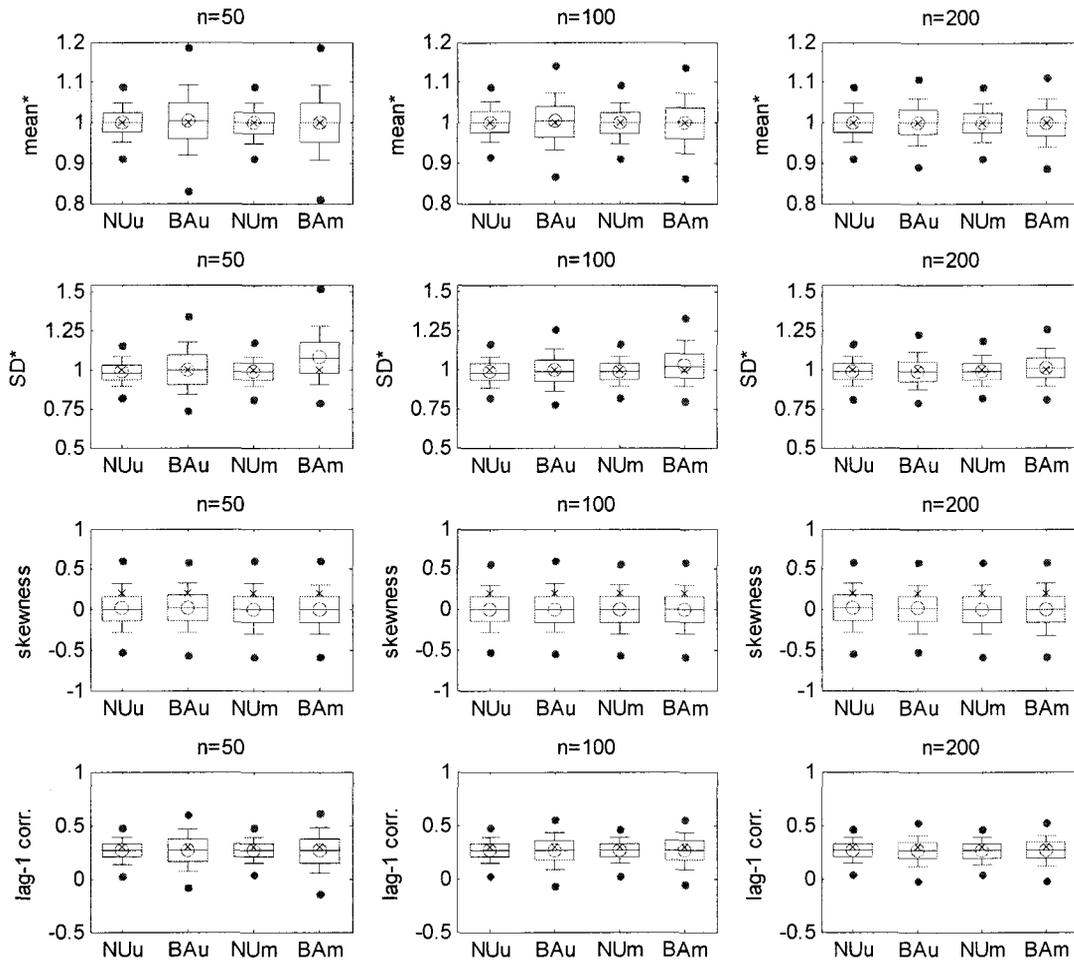


Figure 5.A11: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 8) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

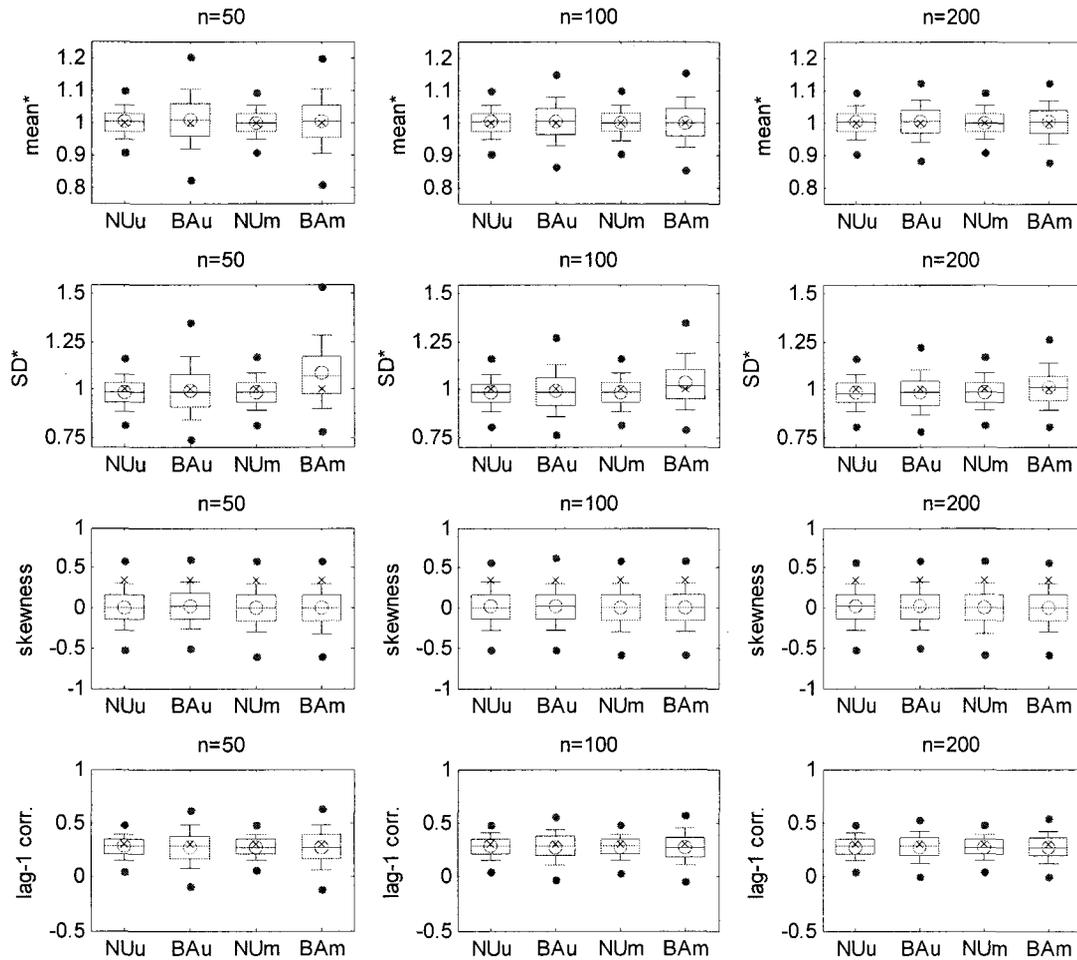


Figure 5.A12: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 16) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

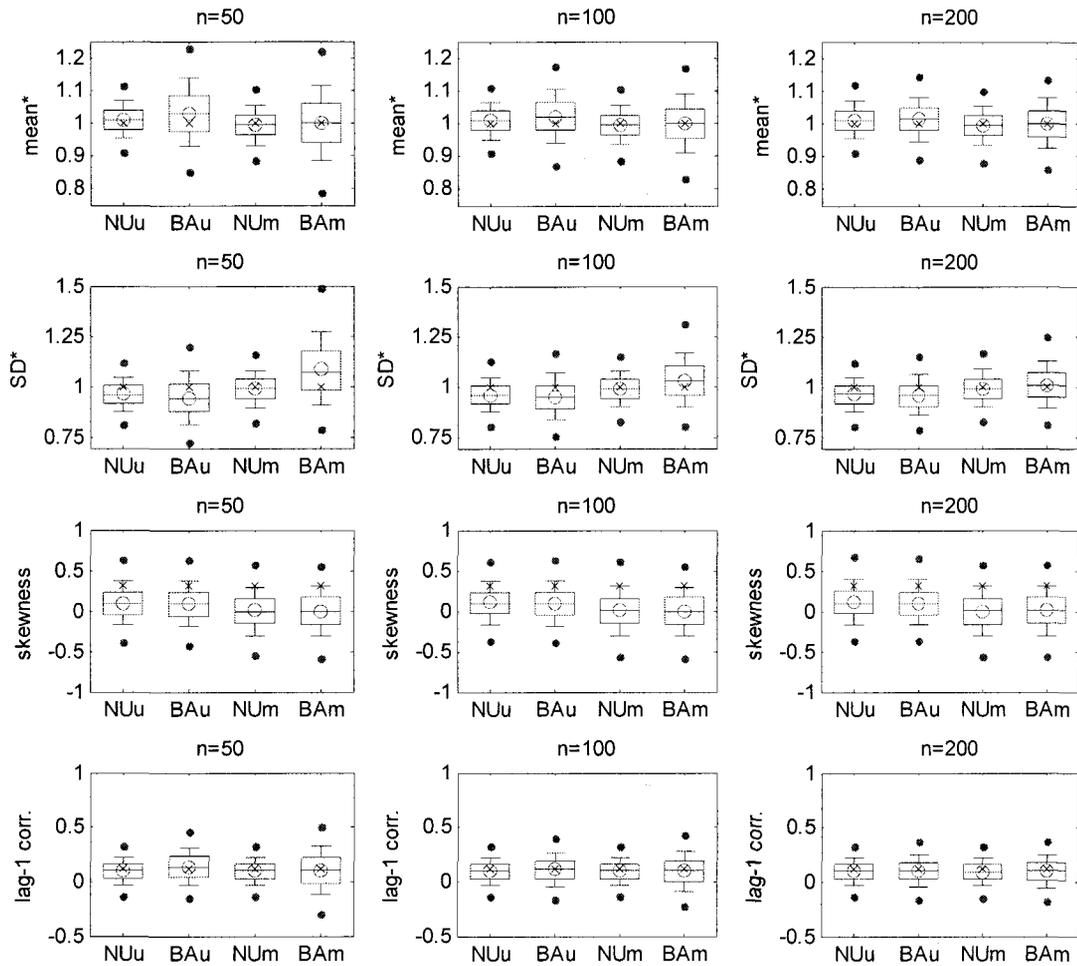


Figure 5.A13: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 19) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

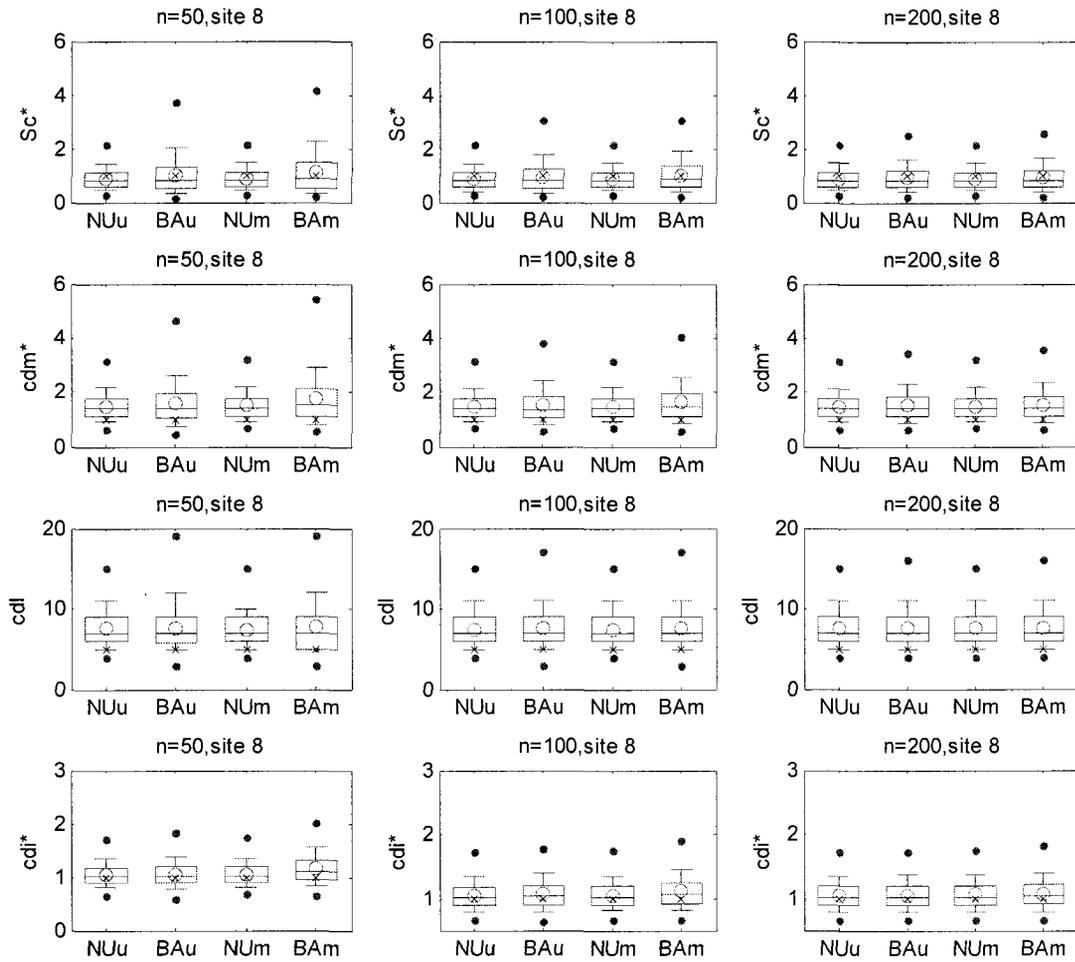


Figure 5.A14: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 8 given demand level of 100% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

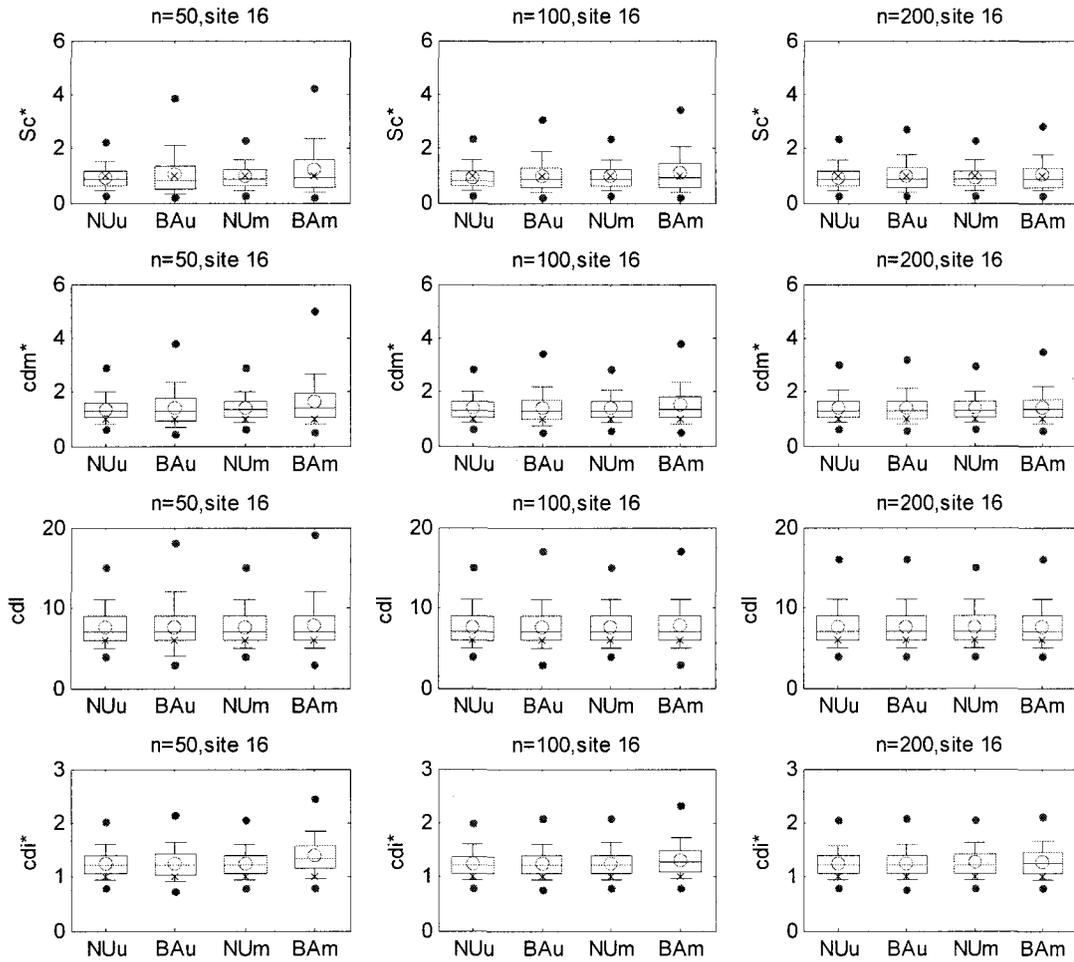


Figure 5.A15: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 100% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

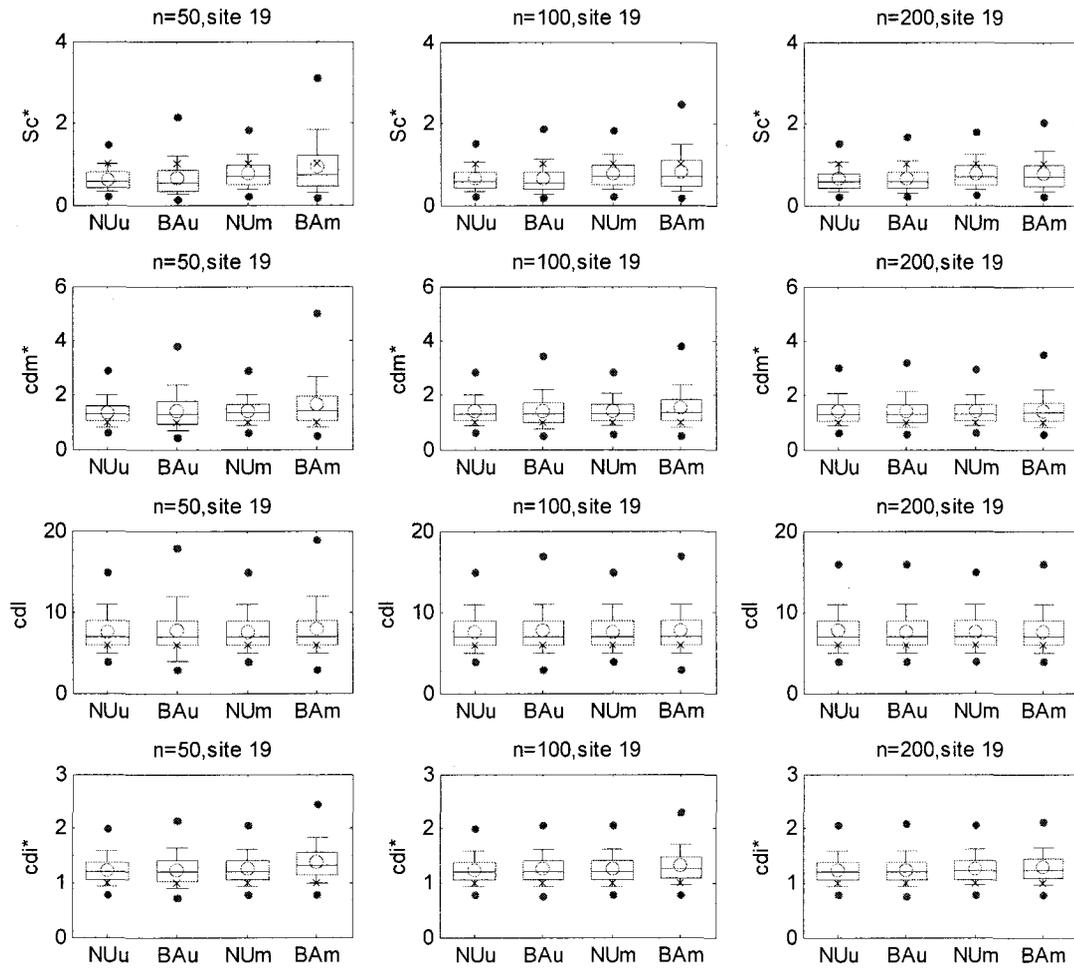


Figure 5.A16: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 100% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

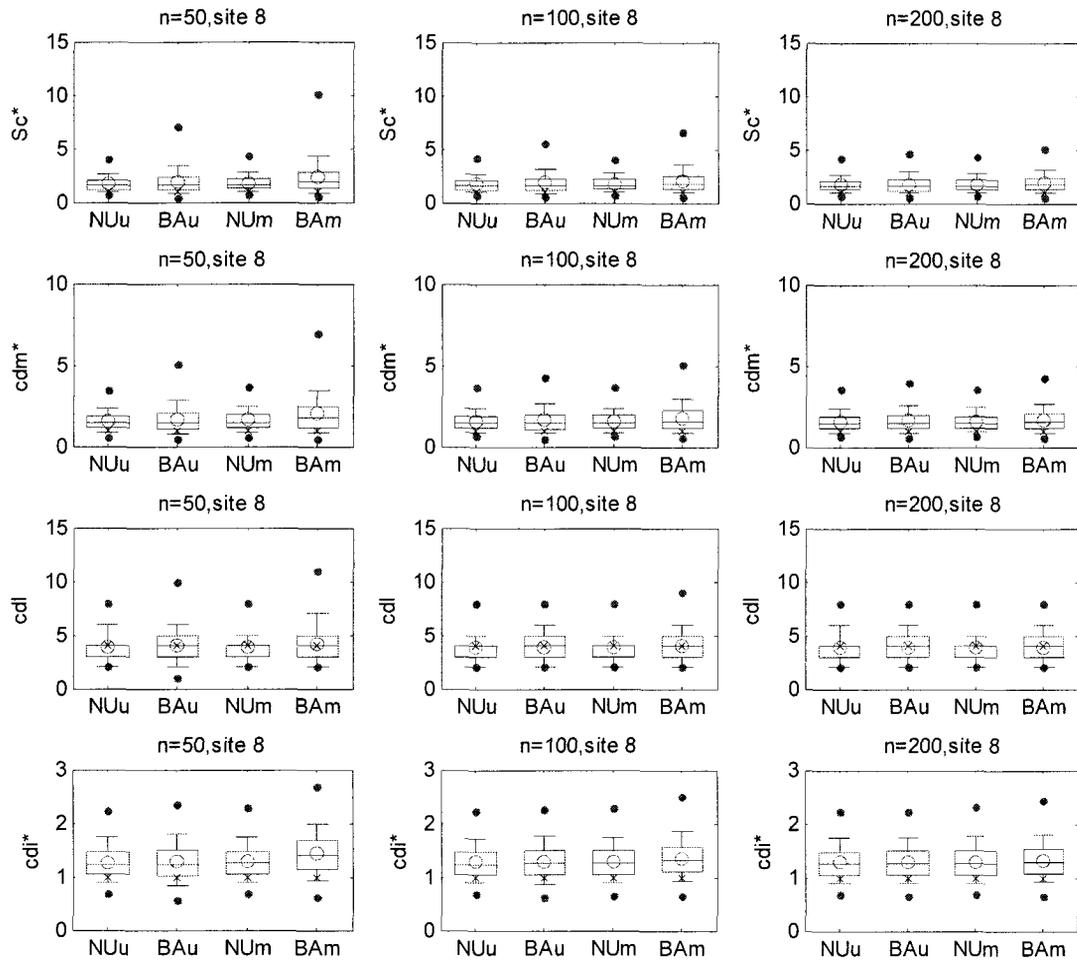


Figure 5.A17: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 8 given demand level of 80% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

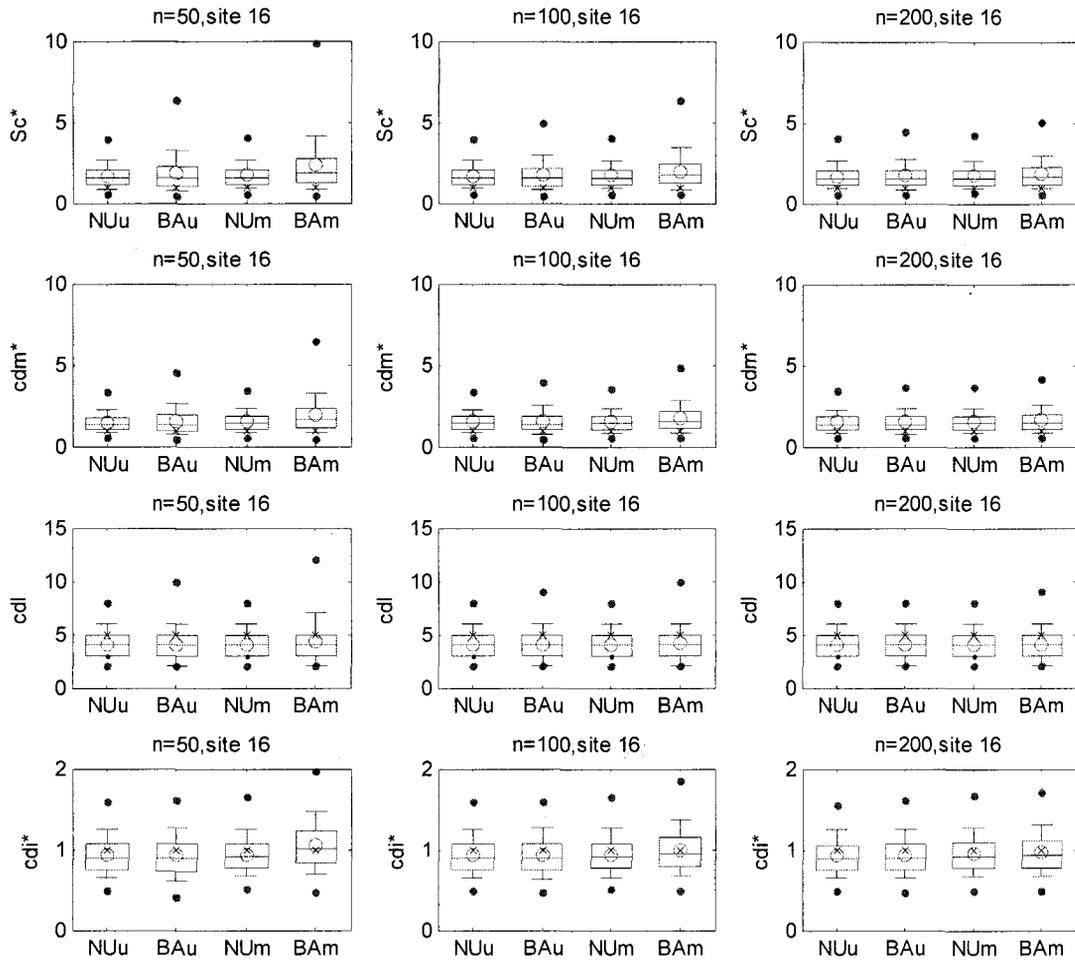


Figure 5.A18: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude (cdm), length (cdl), intensity (cdi) calculated from generated annual streamflows at site 16 given demand level of 80% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

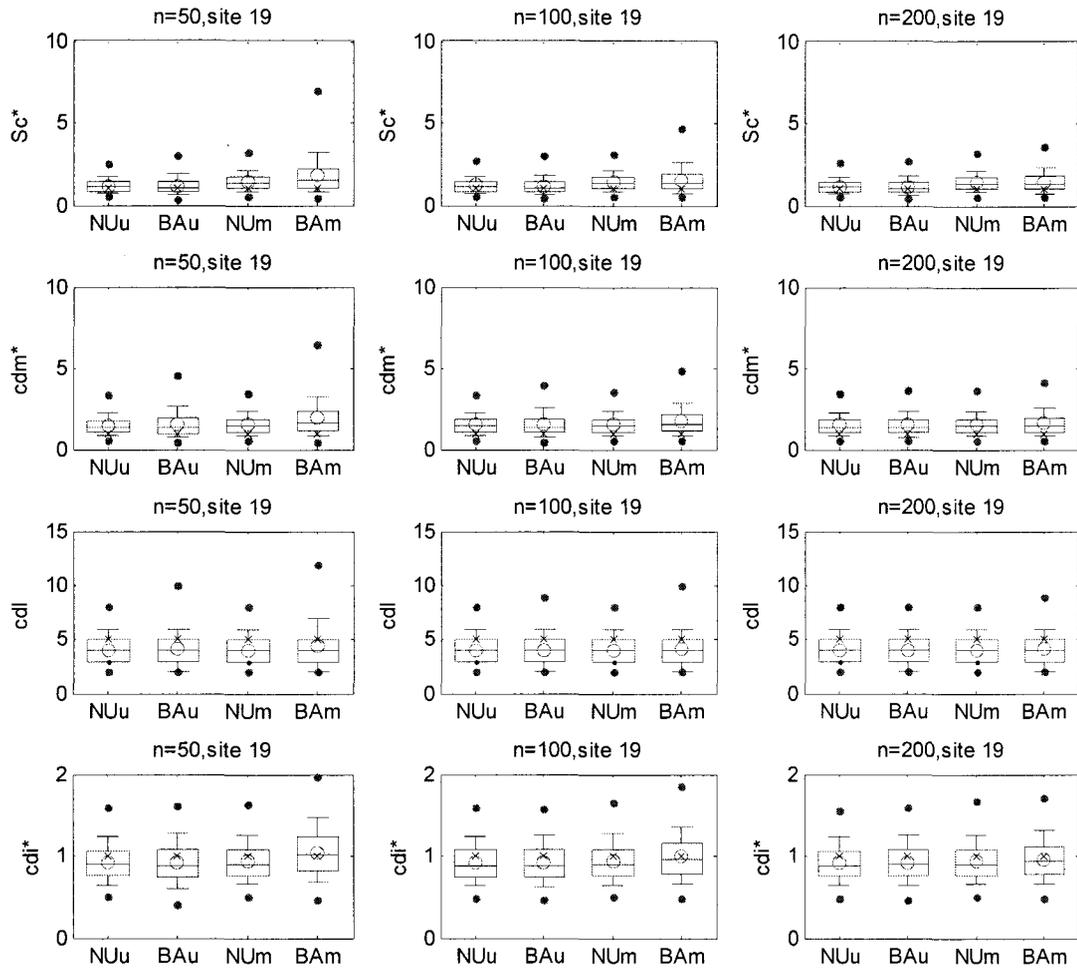


Figure 5.A19: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 80% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

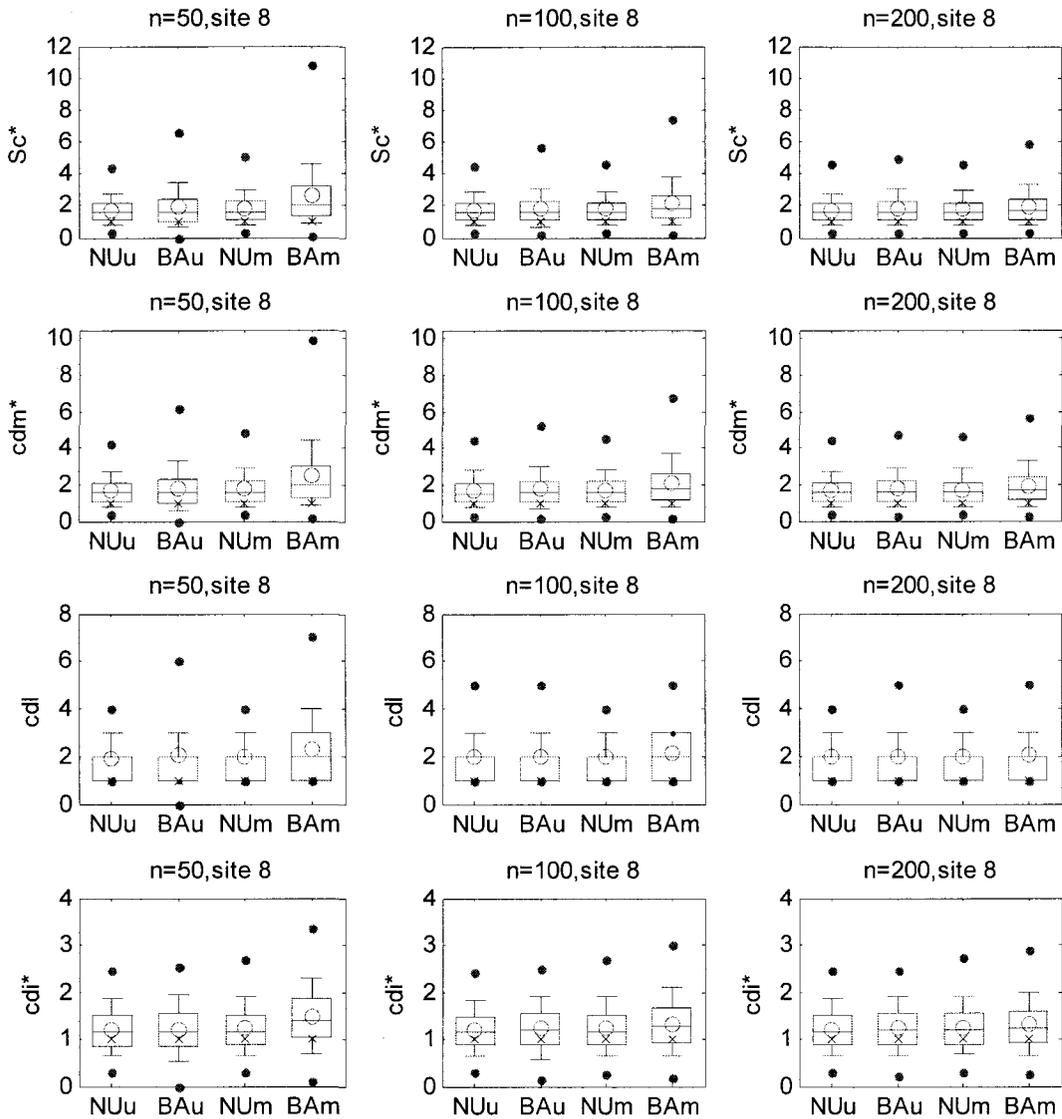


Figure 5.A20: Distributions of storage capacity ( $Sc^*$ ), critical drought indices: magnitude (cdm), length (cdl), intensity (cdi) calculated from generated annual streamflows at site 8 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

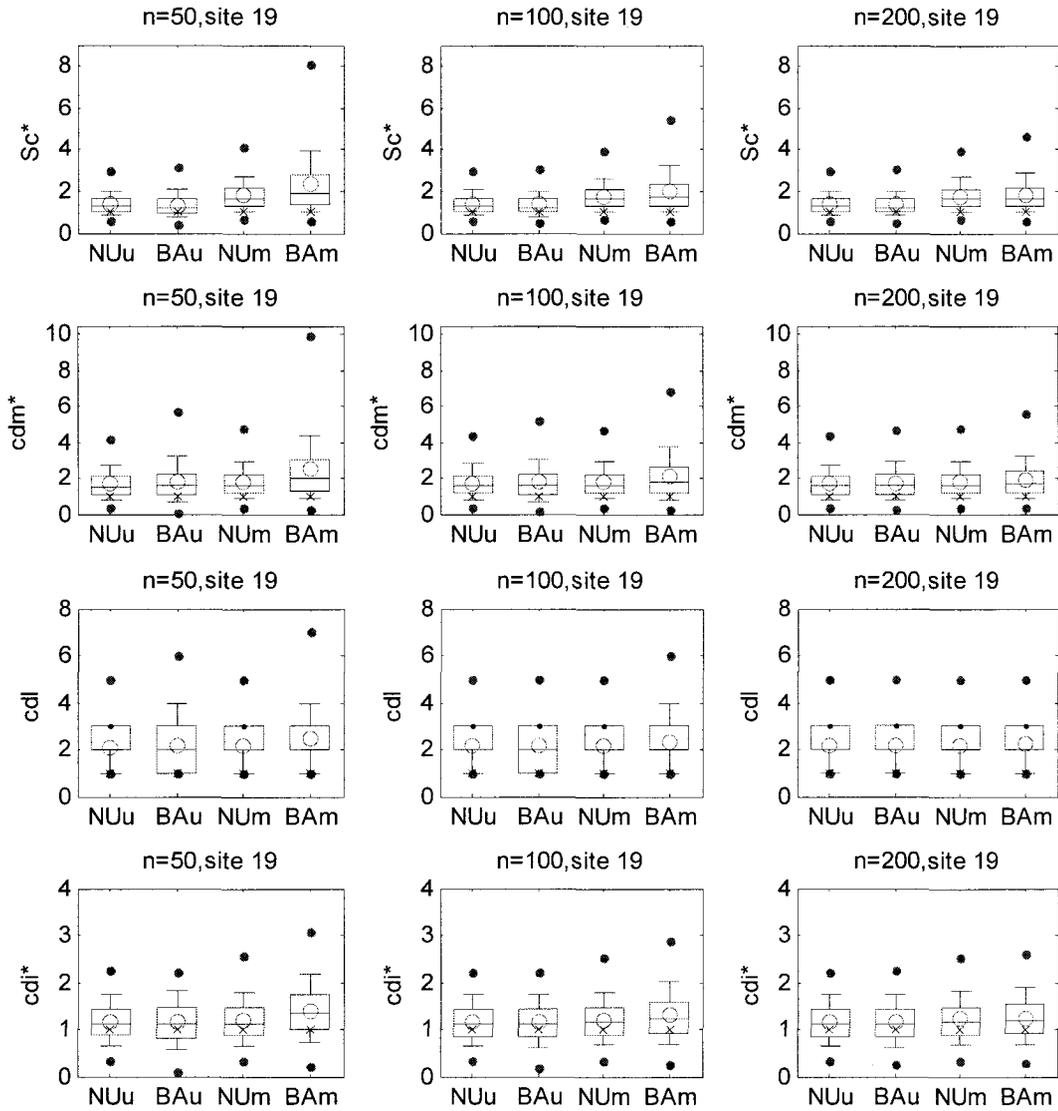


Figure 5.A21: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude (cdm), length (cdl), intensity (cdi) calculated from generated annual streamflows at site 19 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘Num’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

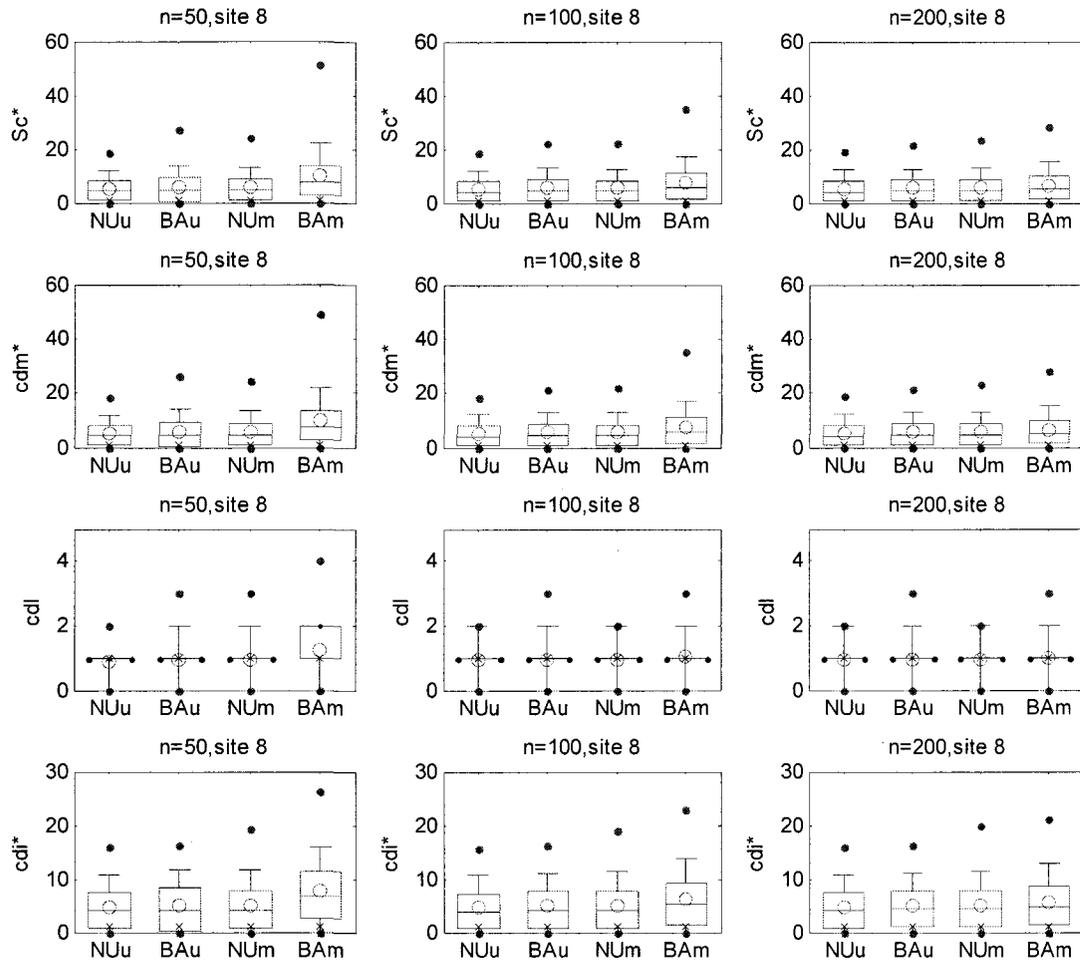


Figure 5.A22: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude (cdm), length (cdl), intensity (cdi) calculated from generated annual streamflows at site 8 given demand level of 40% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

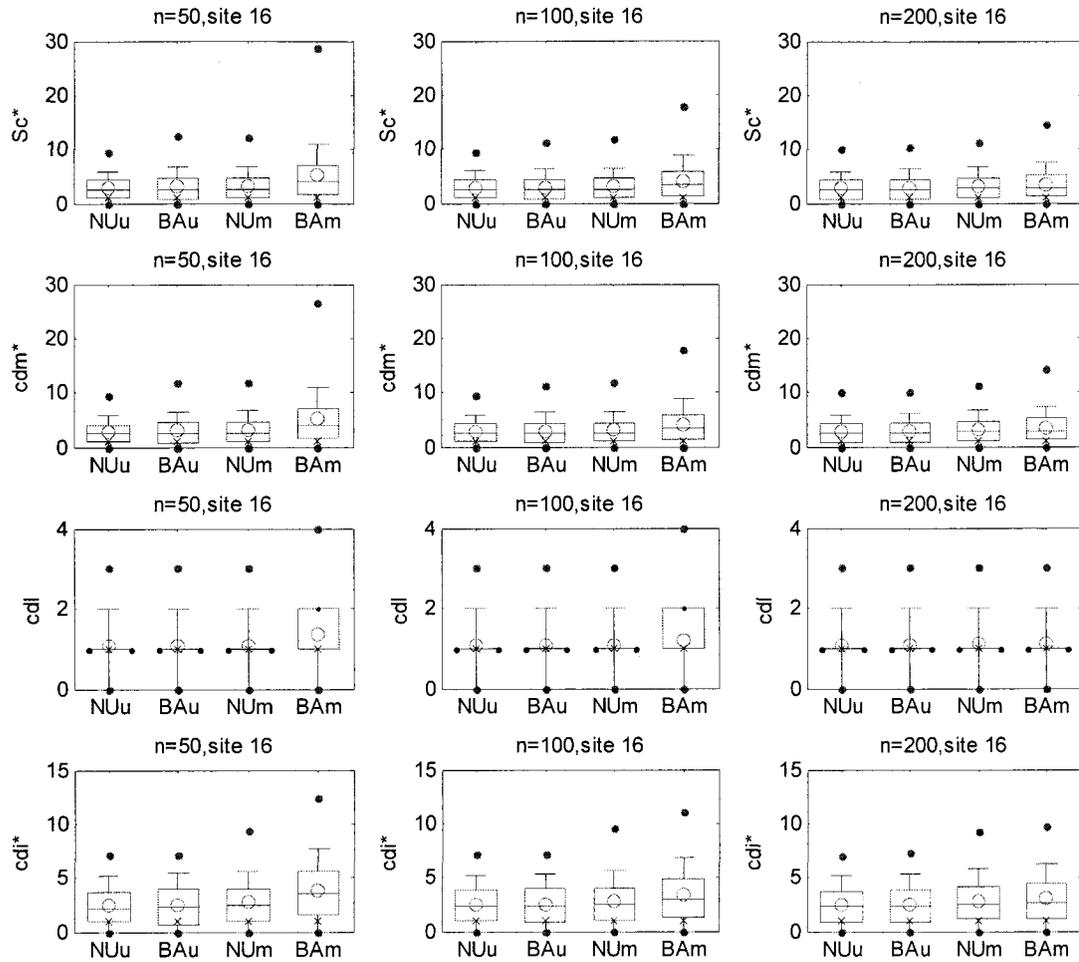


Figure 5.A23: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 16 given demand level of 40% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

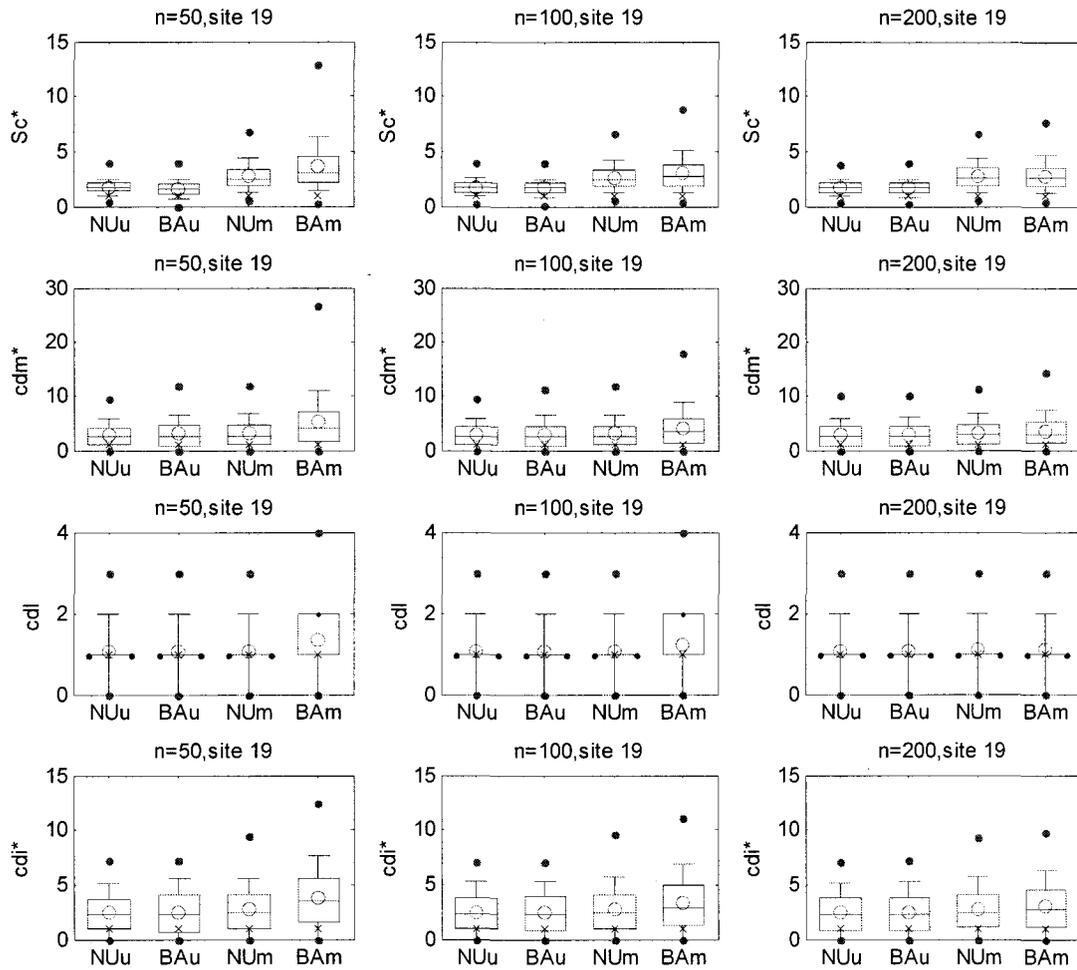


Figure 5.A24: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 40% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics.

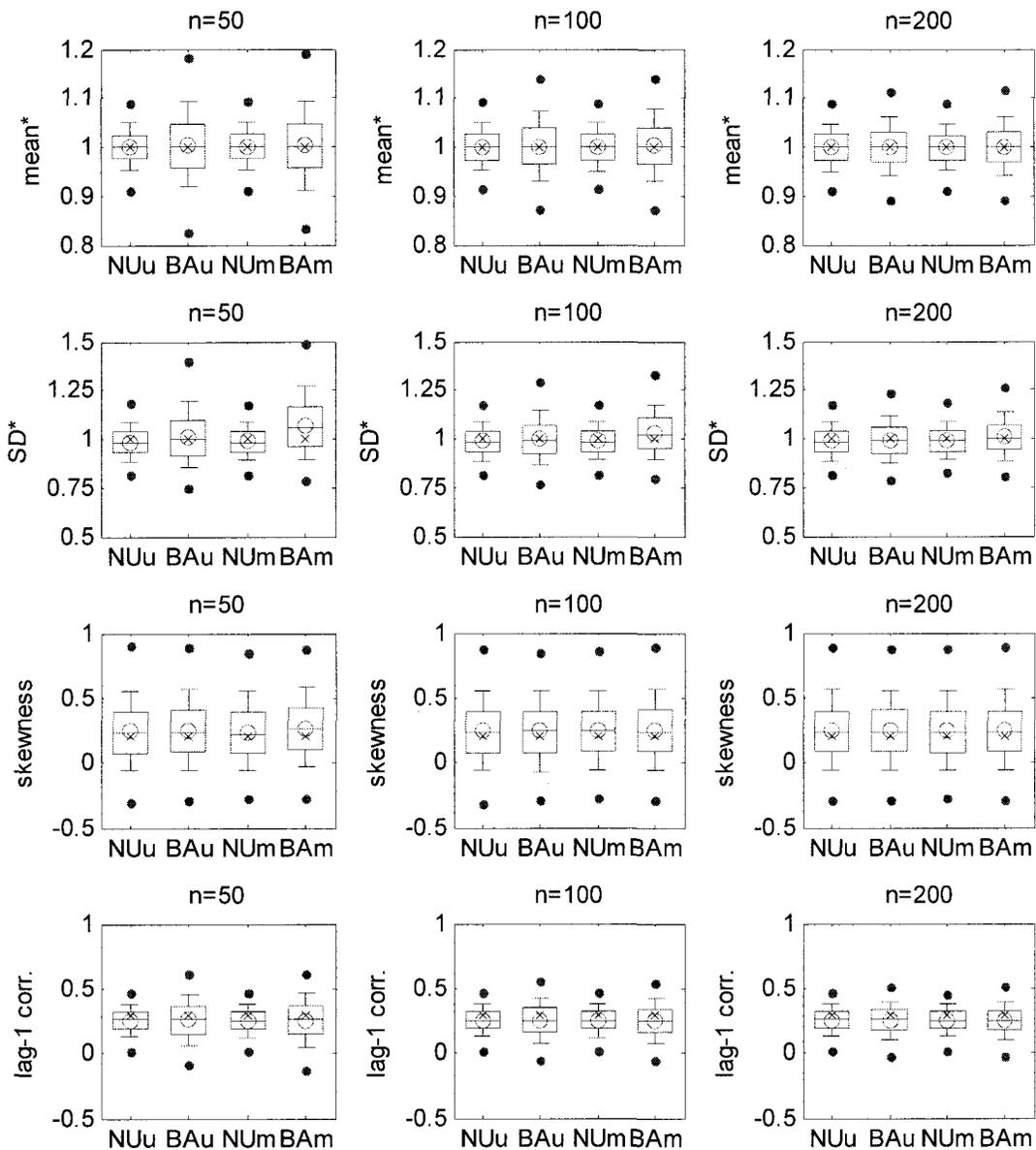


Figure 5.A25: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 8) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

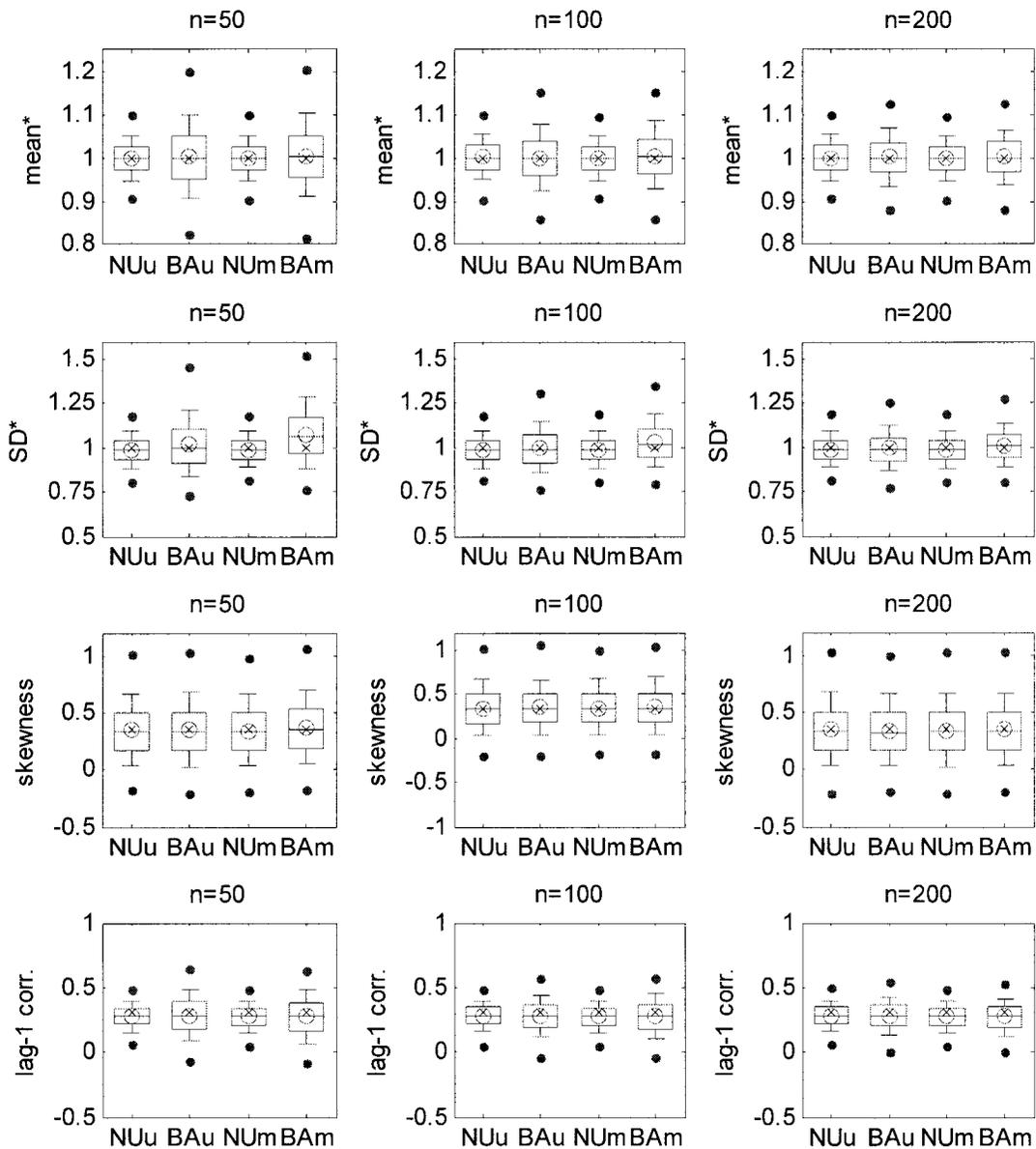


Figure 5.A26: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 16) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

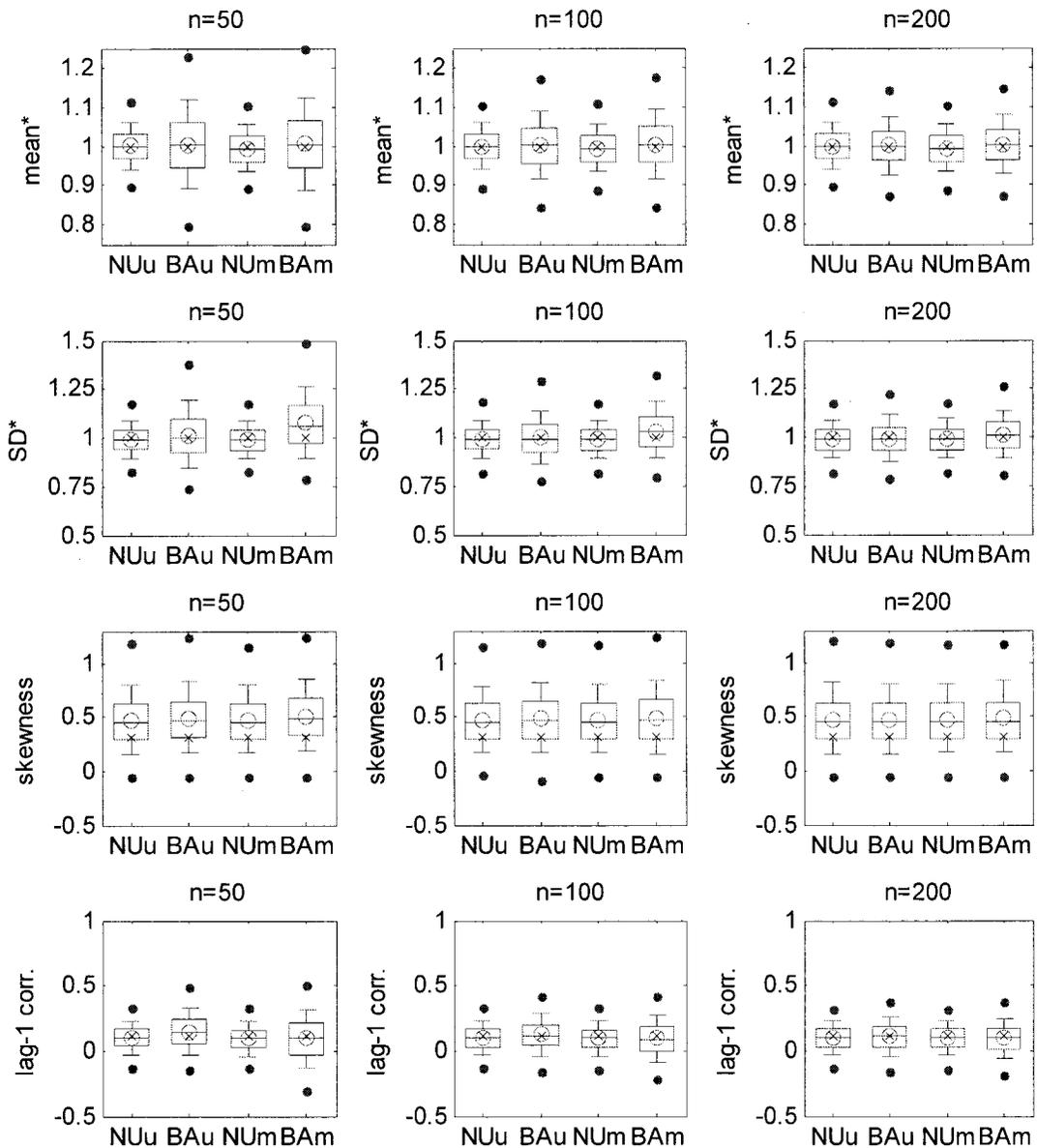


Figure 5.A27: Distributions of mean, standard deviation, skewness coefficient, and lag-1 serial correlation calculated from 5000 different generated annual streamflows sets (site 19) with parameter uncertainty incorporated where: ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

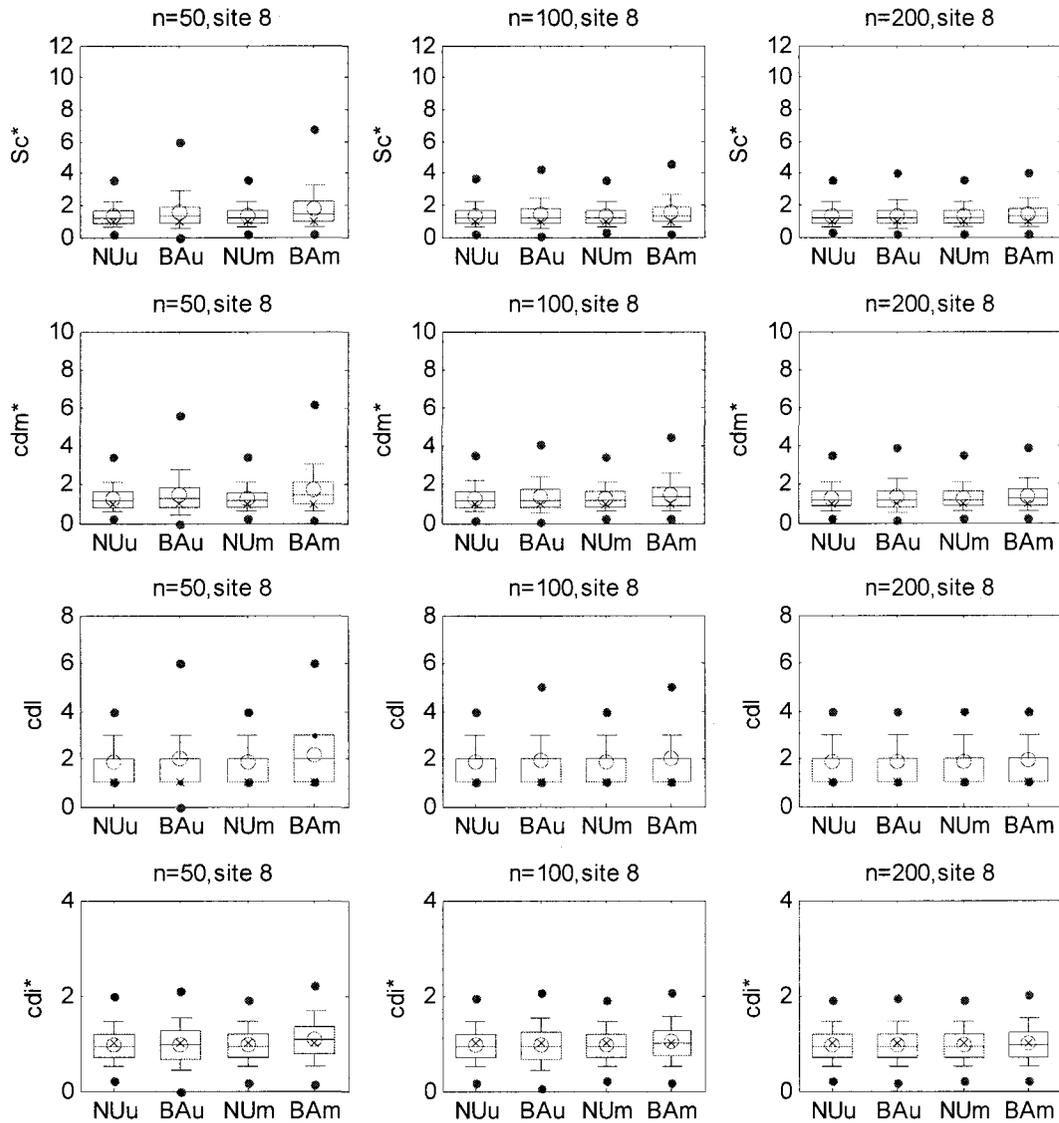


Figure 5.A28: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude (cdm), length (cdl), intensity (cdi) calculated from generated annual streamflows at site 8 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

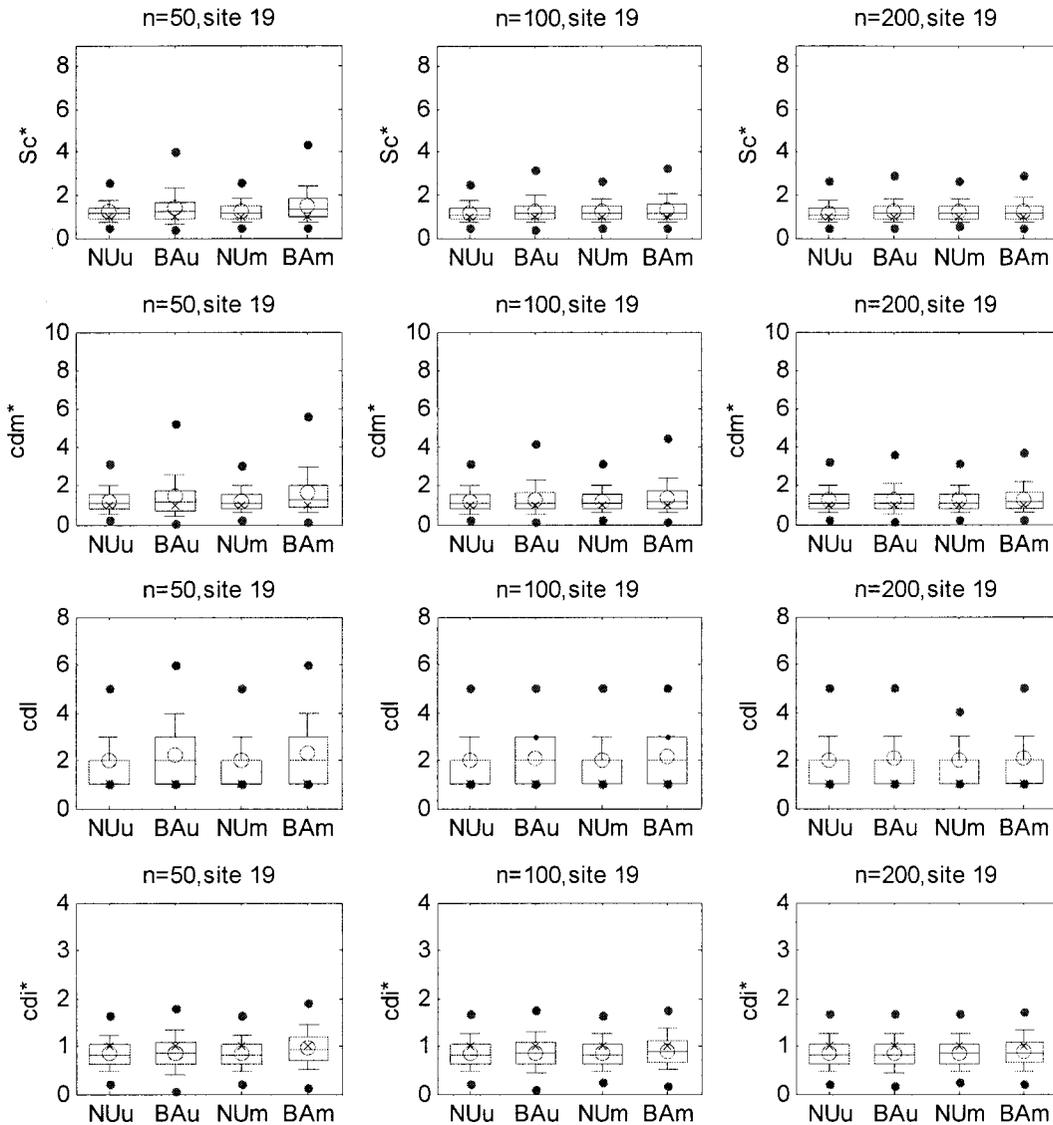


Figure 5.A29: Distributions of storage capacity ( $Sc$ ), critical drought indices: magnitude ( $cdm$ ), length ( $cdl$ ), intensity ( $cdi$ ) calculated from generated annual streamflows at site 19 given demand level of 60% MAF. ‘NUu’ means natural uncertainty based on univariate generation, ‘BAu’ means using posterior distribution based on univariate generation, ‘NUm’ means natural uncertainty based on multivariate generation, and ‘BAm’ means using posterior distribution based on multivariate generation. Superscript \* notes “scaled by historical statistics” and ‘X’ denotes historical statistics. Transformation applied.

Table 5.A1: Rbias and rmse of generated design variables based on different parameter uncertainty consideration methods (site 8)

sample size		d=100% MAF			d=80% MAF			d=60% MAF		
		NU	AS	BA	NU	AS	BA	NU	AS	BA
Storage capacity										
rbias	50	-10	8	14	77	110	142	78	113	157
(%)	100	-11	-1	2	75	91	105	75	95	111
	200	-11	-7	-6	75	81	88	76	82	93
rmse	50	43	77	87	109	190	258	125	205	285
(%)	100	43	60	64	105	140	160	118	156	181
	200	44	53	52	107	121	129	119	135	146
Critical drought magnitude										
rbias	50	50	66	76	60	81	103	75	107	145
(%)	100	49	58	63	58	70	79	73	91	105
	200	50	52	55	58	62	68	73	79	89
rmse	50	73	109	129	89	139	177	120	191	250
(%)	100	72	91	97	85	110	123	115	150	171
	200	73	80	84	86	97	103	116	129	141
Critical drought length										
rbias	50	48	55	55	-4	2	5	96	111	130
(%)	100	49	53	53	-5	-1	0	96	104	112
	200	50	50	50	-5	-3	-2	94	98	103
rmse	50	67	84	85	33	46	50	127	159	185
(%)	100	68	77	77	32	38	39	125	143	151
	200	69	71	72	33	36	36	125	132	138
Critical drought intensity										
rbias	50	6	9	16	29	33	44	24	30	46
(%)	100	6	7	11	29	32	36	23	28	33
	200	6	6	8	30	30	33	25	25	30
rmse	50	23	28	34	45	52	62	56	68	80
(%)	100	23	26	28	45	50	53	55	63	66
	200	24	24	26	46	47	50	56	59	62

Table 5.A2: Rbias and rmse of generated design variables based on different parameter uncertainty consideration methods (site 16)

sample size		d=100% MAF			d=80% MAF			d=60% MAF		
		NU	AS	BA	NU	AS	BA	NU	AS	BA
Storage capacity										
rbias	50	-3	17	22	72	108	134	82	121	163
(%)	100	-3	7	11	71	88	100	81	101	118
	200	-4	1	2	71	78	83	84	91	96
rrmse	50	46	86	95	104	191	266	125	213	333
(%)	100	46	67	70	102	141	155	123	165	188
	200	46	56	57	103	120	124	127	144	147
Critical drought magnitude										
rbias	50	40	55	63	54	77	95	79	112	148
(%)	100	39	46	51	53	64	73	77	96	111
	200	39	42	44	54	58	61	80	86	91
rrmse	50	63	101	121	82	135	172	121	194	268
(%)	100	63	81	86	82	107	117	118	156	178
	200	63	71	74	83	94	95	121	136	140
Critical drought length										
rbias	50	26	31	30	-20	-14	-12	114	131	147
(%)	100	26	28	29	-20	-17	-16	113	123	130
	200	25	27	27	-20	-19	-18	113	116	120
rrmse	50	47	63	65	33	41	43	144	179	203
(%)	100	47	55	56	34	37	37	143	162	171
	200	47	50	51	34	35	35	142	149	153
Critical drought intensity										
rbias	50	26	29	38	-6	-4	5	19	24	40
(%)	100	25	27	31	-6	-4	-1	20	23	30
	200	26	26	28	-6	-5	-3	22	22	25
rrmse	50	37	42	52	25	28	32	50	59	72
(%)	100	37	40	44	26	28	29	50	56	62
	200	38	38	40	25	26	26	51	53	55

Table 5.A3: Rbias and rmse of generated design variables based on different parameter uncertainty consideration methods (site 19)

sample size		d=100% MAF			d=80% MAF			d=60% MAF		
		NU	AS	BA	NU	AS	BA	NU	AS	BA
Storage capacity										
rbias	50	-24	-10	-7	43	66	87	77	101	131
(%)	100	-24	-17	-18	42	54	58	76	87	97
	200	-24	-21	-23	42	46	47	76	79	84
rmse	50	43	63	68	71	128	170	104	162	208
(%)	100	43	52	53	69	96	104	102	126	140
	200	42	48	47	69	81	82	103	112	118
Critical drought magnitude										
rbias	50	27	37	45	40	55	71	70	89	115
(%)	100	26	31	33	40	47	51	69	78	87
	200	26	28	28	40	42	45	70	72	76
rmse	50	50	75	89	64	96	119	95	137	172
(%)	100	49	61	65	63	78	84	94	114	125
	200	49	55	55	62	70	72	95	103	106
Critical drought length										
rbias	50	-5	-2	-3	1	6	9	-36	-33	-29
(%)	100	-6	-4	-5	2	4	3	-36	-34	-34
	200	-6	-5	-7	1	2	2	-36	-36	-35
rmse	50	30	37	40	32	43	47	42	44	43
(%)	100	30	33	34	32	37	38	42	43	43
	200	30	32	31	32	34	35	42	43	43
Critical drought intensity										
rbias	50	34	38	46	56	60	70	71	77	91
(%)	100	35	36	40	56	57	63	70	73	80
	200	35	35	37	57	57	59	72	72	75
rmse	50	45	52	60	68	76	86	88	99	113
(%)	100	46	48	53	68	71	77	87	92	100
	200	46	48	49	70	70	72	90	91	93

Table 5.A4: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 8)

Sample size		100%				80%			
		NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm
Storage Capacity									
rbias	50	-12	2	-10	14	73	94	77	142
(%)	100	-13	-4	-11	2	73	82	75	105
	200	-11	-8	-11	-6	73	80	75	88
rrmse	50	43	75	43	87	104	162	109	258
(%)	100	43	62	43	64	104	133	105	160
	200	43	51	44	52	103	118	107	129
Critical drought magnitude									
rbias	50	49	58	50	76	56	68	60	103
(%)	100	48	53	49	63	56	61	58	79
	200	49	53	50	55	56	61	58	68
rrmse	50	72	101	73	129	84	119	89	177
(%)	100	72	87	72	97	84	102	85	123
	200	71	80	73	84	84	94	86	103
Critical drought length									
rbias	50	50	53	48	55	-4	-1	-4	5
(%)	100	49	50	49	53	-5	-4	-5	0
	200	50	51	50	50	-4	-3	-5	-2
rrmse	50	69	83	67	85	33	44	33	50
(%)	100	69	75	68	77	33	38	32	39
	200	68	72	69	72	33	36	33	36
Critical drought intensity									
rbias	50	5	6	6	16	27	28	29	44
(%)	100	5	7	6	11	28	29	29	36
	200	5	6	6	8	28	29	30	33
rrmse	50	23	26	23	34	43	47	45	62
(%)	100	23	25	23	28	43	46	45	53
	200	23	24	24	26	44	45	46	50

Table 5.A4: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 8) -Continued

Sample size	60%				40%				
	NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm	
Storage Capacity									
rbias	50	69	87	78	157	413	482	466	904
(%)	100	70	79	75	111	410	455	447	647
	200	70	79	76	93	416	454	459	558
rmse	50	109	158	125	285	631	776	737	1552
(%)	100	112	138	118	181	635	716	702	1018
	200	111	126	119	146	640	692	713	866
Critical drought magnitude									
rbias	50	67	81	75	145	412	479	466	891
(%)	100	68	75	73	105	409	452	446	642
	200	68	75	73	89	415	452	458	555
rmse	50	105	146	120	250	630	769	735	1470
(%)	100	109	130	115	171	632	709	700	1007
	200	108	120	116	141	638	689	709	861
Critical drought length									
rbias	50	93	104	96	130	-9	-5	-6	22
(%)	100	94	98	96	112	-8	-7	-7	5
	200	95	99	94	103	-7	-5	-6	0
rmse	50	124	152	127	185	57	73	59	91
(%)	100	126	138	125	151	58	65	58	71
	200	125	133	125	138	57	60	57	62
Critical drought intensity									
rbias	50	21	20	24	46	369	393	405	660
(%)	100	20	22	23	33	360	388	395	518
	200	21	23	25	30	369	392	405	469
rmse	50	52	58	56	80	556	603	622	913
(%)	100	51	56	55	66	549	584	609	760
	200	52	54	56	62	557	582	618	696

Table 5.A5: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 16)

Sample size		100%				80%			
		NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm
Storage Capacity									
rbias	50	-6	7	-3	22	67	83	72	134
(%)	100	-6	2	-3	11	68	75	71	100
	200	-5	-1	-4	2	68	71	71	83
rmse	50	45	80	46	95	98	149	104	266
(%)	100	45	64	46	70	99	122	102	155
	200	45	56	46	57	99	112	103	124
Critical drought magnitude									
rbias	50	38	43	40	63	49	57	54	95
(%)	100	38	41	39	51	51	54	53	73
	200	39	39	39	44	50	52	54	61
rmse	50	62	85	63	121	77	105	82	172
(%)	100	62	73	63	86	78	92	82	117
	200	64	70	63	74	79	86	83	95
Critical drought length									
rbias	50	26	28	26	30	-20	-18	-20	-12
(%)	100	26	27	26	29	-20	-19	-20	-16
	200	27	26	25	27	-20	-20	-20	-18
rmse	50	48	59	47	65	34	40	33	43
(%)	100	48	54	47	56	33	37	34	37
	200	50	51	47	51	34	36	34	35
Critical drought intensity									
rbias	50	24	25	26	38	-7	-8	-6	5
(%)	100	24	25	25	31	-8	-8	-6	-1
	200	24	24	26	28	-8	-7	-6	-3
rmse	50	36	39	37	52	25	27	25	32
(%)	100	36	37	37	44	25	26	26	29
	200	36	36	38	40	25	25	25	26

Table 5.A5: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 16) –Continued

Sample size		60%				40%			
		NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm
Storage Capacity									
rbias	50	72	84	82	163	180	203	214	436
(%)	100	74	80	81	118	183	194	211	318
	200	73	77	84	96	179	187	221	258
rmse	50	111	150	125	333	292	351	351	843
(%)	100	114	134	123	188	295	325	344	521
	200	113	124	127	147	293	312	350	409
Critical drought magnitude									
rbias	50	68	78	79	148	179	201	214	428
(%)	100	71	75	77	111	182	193	210	316
	200	69	73	80	91	178	186	220	257
rmse	50	106	138	121	268	290	345	350	806
(%)	100	110	125	118	178	294	323	342	513
	200	108	119	121	140	292	309	347	406
Critical drought length									
rbias	50	109	117	114	147	6	7	8	35
(%)	100	112	115	113	130	6	7	7	20
	200	112	113	113	120	5	5	9	12
rmse	50	139	159	144	203	58	74	60	106
(%)	100	143	153	143	171	59	69	59	77
	200	142	148	142	153	59	65	60	64
Critical drought intensity									
rbias	50	16	14	19	40	146	151	173	286
(%)	100	16	15	20	30	149	150	173	232
	200	15	15	22	25	145	149	179	203
rmse	50	45	50	50	72	238	255	280	408
(%)	100	45	47	50	62	240	247	280	349
	200	44	46	51	55	237	246	283	312

Table 5.A6: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 19)

Sample size		100%				80%			
		NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm
Storage Capacity									
rbias	50	-37	-35	-24	-7	18	18	43	87
(%)	100	-35	-37	-24	-18	20	18	42	58
	200	-36	-36	-24	-23	19	18	42	47
rmse	50	47	56	43	68	46	59	71	170
(%)	100	46	52	43	53	49	55	69	104
	200	46	49	42	47	48	50	69	82
Critical drought magnitude									
rbias	50	13	12	27	45	20	18	40	71
(%)	100	14	12	26	33	21	19	40	51
	200	14	12	26	28	20	19	40	45
rmse	50	38	47	50	89	44	51	64	119
(%)	100	39	43	49	65	46	48	63	84
	200	39	40	49	55	44	46	62	72
Critical drought length									
rbias	50	-9	-9	-5	-3	-4	-4	1	9
(%)	100	-9	-10	-6	-5	-3	-4	2	3
	200	-9	-10	-6	-7	-3	-4	1	2
rmse	50	29	34	30	40	30	35	32	47
(%)	100	29	31	30	34	31	33	32	38
	200	29	31	30	31	31	31	32	35
Critical drought intensity									
rbias	50	24	18	34	46	36	29	56	70
(%)	100	24	21	35	40	36	33	56	63
	200	24	23	35	37	36	35	57	59
rmse	50	33	30	45	60	44	39	68	86
(%)	100	33	31	46	53	44	42	68	77
	200	33	32	46	49	43	43	70	72

Table 5.A6: Rbias and rmse of generated design variables based on univariate and multivariate estimations (site 19) -Continued

Sample size		60%				40%			
		NUu	BAu	NUm	BAm	NUu	BAu	NUm	BAm
Storage Capacity									
rbias	50	37	32	77	131	77	62	170	263
(%)	100	37	34	76	97	76	69	167	206
	200	35	34	76	84	74	72	171	182
rrmse	50	61	66	104	208	104	101	213	367
(%)	100	62	63	102	140	104	101	209	269
	200	59	61	103	118	100	100	213	233
Critical drought magnitude									
rbias	50	33	28	70	115	76	61	168	254
(%)	100	33	30	69	87	76	68	165	202
	200	31	31	70	76	73	71	170	179
rrmse	50	57	60	95	172	102	98	211	348
(%)	100	57	57	94	125	103	100	207	262
	200	54	57	95	106	99	99	211	229
Critical drought length									
rbias	50	-40	-41	-36	-29	45	40	58	83
(%)	100	-40	-41	-36	-34	46	43	57	68
	200	-40	-40	-36	-35	44	44	57	62
rrmse	50	45	47	42	43	75	76	88	125
(%)	100	45	46	42	43	77	75	86	104
	200	45	45	42	43	74	75	87	95
Critical drought intensity									
rbias	50	37	28	71	91	53	39	126	168
(%)	100	37	32	70	80	53	45	124	143
	200	36	34	72	75	52	49	130	133
rrmse	50	46	43	88	113	71	67	158	209
(%)	100	46	44	87	100	71	69	157	181
	200	46	45	90	93	70	69	163	168

## Chapter VI

### PROPORTIONAL DISAGGREGATION MODEL AND PARAMETER UNCERTAINTY

**Abstract:** In the past decades, disaggregation models have been used for synthetic generation of hydrologic streamflows as an alternative to multivariate and periodic stochastic models. In order to overcome the drawbacks associated with the large number of parameters and/or non-normally distributed flows in the traditional disaggregation model of Valencia and Schaake (1973), condensed disaggregation models have been utilized and have shown good performance in most cases of hydrologic application. However, in some of the condensed models, the generated monthly flows do not sum to their corresponding annual flow, requiring an adjustment. A simple temporal disaggregation model (called the PD model), which avoids such adjustment, is presented. It is based on the proportionality between the annual and seasonal flows; and the seasonal proportions are modeled as Dirichlet random variables. The parameters of the Dirichlet distribution are estimated by the method of moments and the method of maximum likelihood. Using streamflow data sets of the Colorado and St. Lawrence Rivers and simulations based on these data sets, the usefulness of the PD model is assessed. The PD model with the estimation methods employed did quite well in

preserving the seasonal means but not in preserving other statistics. It has only  $m$  parameters, where  $m$  is the number of seasons, which does not give it great flexibility.

## 6.1 Introduction

Since the model proposed by Valencia and Schaake (1973) became acceptable, disaggregation models have become widely used in the simulation of the hydrologic processes (Lane and Frevert, 1990; Salas et al, 1980; Grygier and Stedinger, 1990). Several disaggregation models are available for both temporal and spatial cases (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Tao and Delleur, 1976; Lane, 1979; Stedinger and Vogel, 1984; Stedinger et al, 1985; Santos and Salas, 1985; Grygier and Steinger, 1990; Lane and Frevert, 1990). Compared with traditional stochastic multivariate or periodic models, the advantage of disaggregation models is the preservation of statistical properties at more than one level, i.e. reproduction of the statistics at both the annual level and the monthly level or at both the key site and the sub sites.

Traditional disaggregation models require a large number of parameters, e.g., for temporal disaggregation of annual flows into monthly flows the Valencia and Schaake model requires 156 parameters, which led to condensed or staged disaggregation models (Lane, 1979; Stedinger et al, 1985; Santos and Salas, 1985). However, there are some drawbacks for these disaggregation models. For example, consider the condensed temporal disaggregation model by Lane (1979). Since all seasonal flows are not

generated jointly, the sum of generated flows is not equal to the given annual flows. An adjustment of generated seasonal flows should be required as a post process to treat the discrepancies between generated annual and seasonal flows. Several adjustments have been proposed in the literature (Lane, 1979; Stedinger and Vogel, 1984; Pereira et al., 1984; Grygier and Stedinger, 1988). Grygier and Stedinger (1988) compared the performance of the adjustment procedures based on the SPC disaggregation model (Stedinger et al., 1985) and discussed that use of adjustment results in the distortion of the marginal distribution and the proportional procedure would give the least distortion among several methods. This distortion of marginal distribution is common to all the other disaggregation models if the transformed data are modeled rather than real-space streamflows. This would be more severe for the temporal disaggregation than spatial disaggregation because seasonal flows are seldom normally distributed (Salas et al., 1980).

To avoid this problem of adjustment, a simple disaggregation model is proposed, which will be called by the proportional disaggregation (PD) model throughout this paper. This model simply partitions the annual flow (which can be modeled in a variety of ways and does not have to be normally distributed) into monthly flows in the temporal case, or, partitions the key-site flow into sub-site flows in the spatial case. The partitioning is accomplished by assuming a distribution for the monthly proportions; a Dirichlet distribution is used. This differs from those traditional disaggregation models that represent these proportions as centering parameters plus error terms, often assumed as multivariate normals. In the next section, model description will be given and different parameter estimators are compared. Additionally, parameter uncertainty will be

discussed based on the theoretically derived asymptotic distribution of parameter estimators and its effect will be examined through the real application of temporal disaggregation for the Colorado River and St. Lawrence River. Basic statistics and design variables, such as storage capacity and critical drought indices (magnitude, length, intensity), will be employed for comparison purposes. Spatial disaggregation is not illustrated.

## 6.2 Proportional disaggregation model

### 6.2.1 Model description

Let  $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$  be the annual flows series and  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_j, \dots, \mathbf{Y}_m)$  the seasonal flows series vector to be generated where  $\mathbf{Y}_j = (y_{1,j}, \dots, y_{i,j}, \dots, y_{n,j})^T$  is the  $j$ th seasonal flow vector consisting of flow  $y_{ij}$  at year  $i$  and season  $j$ ,  $m$  is number of season, and  $n$  is the sample size. ( $\mathbf{X}$  and  $\mathbf{Y}$  could be similarly assigned to key-sites and sub-sites in the spatial disaggregation case). A simple proportional disaggregation (PD) model can be given as:

$$\mathbf{Y} = \mathbf{X}\mathbf{P} \quad (6.1)$$

where  $\mathbf{P} = [P_1, P_2, \dots, P_m]$  is  $1 \times m$  random vector whose  $j$ th element  $P_j$  represents the proportion the  $j$ th each seasonal flow is of the annual flow. Realizing that  $0 < P_j < 1$

for all  $j$  and  $\sum_{j=1}^m P_j = 1$ , the  $m$  different proportions of  $P_1, P_2, \dots, P_m$  are assumed to

follow the Dirichlet (or multivariate beta) distribution (Fielitz and Myers, 1975) as:

$$f_{P_1, P_2, \dots, P_m}(P_1, P_2, \dots, P_m; \theta_1, \theta_2, \dots, \theta_m) = \frac{\Gamma\left(\sum_{i=1}^m \theta_i\right)}{\prod_{i=1}^m \Gamma(\theta_i)} \prod_{i=1}^m P_i^{\theta_i} \prod_{i=1}^m I_{(0,1)}(P_i) I_{\{1\}}\left(\sum_{i=1}^m P_i\right) \quad (6.2)$$

where,  $\theta_1, \theta_2, \dots, \theta_m > 0$ . Note again  $\sum_{j=1}^m P_j = 1$ . Marginally, the  $P_j$ 's are beta

distributed. The mean and variance of  $j$ th proportion  $P_j, j = 1, 2, \dots, m$  are given by:

$$\mu_j = E(P_j) = \frac{\theta_j}{\sum_{i=1}^m \theta_i}, \quad \text{var}(P_j) = \frac{\theta_j \left( \sum_{i=1}^m \theta_i - \theta_j \right)}{\left( \sum_{i=1}^m \theta_i \right)^2 \left( \sum_{i=1}^m \theta_i + 1 \right)}, \quad (6.3)$$

and the expectation of a cross product with specified powers is given by:

$$\begin{aligned} \mu_{r_1, r_2, \dots, r_m} &= E(P_1^{r_1} P_2^{r_2} \dots P_m^{r_m}) = \int P_1^{r_1} P_2^{r_2} \dots P_m^{r_m} f_{P_1, P_2, \dots, P_m}(P_1, P_2, \dots, P_m) dp_1, dp_2, \dots, dp_m \\ &= \frac{\Gamma\left(\sum_{i=1}^m \theta_i\right) \prod_{i=1}^m \Gamma(\theta_i + r_i)}{\Gamma\left(\sum_{i=1}^m (\theta_i + r_i)\right) \prod_{i=1}^m \Gamma(\theta_i)}. \end{aligned} \quad (6.4)$$

Using that  $\Gamma(\theta_j) = \theta_j \Gamma(\theta_j - 1)$  and letting  $\sum_{i=1}^m \theta_i = \Sigma \theta$  for simplicity, the covariance

between  $k$ th and  $l$ th proportions with orders  $r_k$  and  $r_l$ ,  $\text{cov}(P_k^{r_k} P_l^{r_l})$  can be derived

using:

$$E(P_k^{r_k} P_l^{r_l}) = \frac{\Gamma\left(\sum_{i=1}^m \theta_i\right)}{\Gamma\left(\sum_{i=1}^m (\theta_i + r_k + r_l)\right)} \frac{\Gamma(\theta_k + r_k) \Gamma(\theta_l + r_l)}{\Gamma(\theta_k) \Gamma(\theta_l)} = \frac{\prod_{i=0}^{r_k-1} (\theta_k + i) \prod_{i=0}^{r_l-1} (\theta_l + i)}{\prod_{i=0}^{r_k+r_l-1} (\Sigma \theta + i)} \quad (6.5)$$

$$E(P_k^{r_k}) = \frac{\Gamma\left(\sum_{i=1}^m \theta_i\right)}{\Gamma\left(\sum_{i=1}^m (\theta_i + r_k)\right)} \frac{\Gamma(\theta_k + r_k)}{\Gamma(\theta_k)} = \frac{\prod_{i=0}^{r_k-1} (\theta_k + i)}{\prod_{i=0}^{r_k-1} (\Sigma\theta + i)} \quad (6.6a)$$

$$E(P_l^{r_l}) = \frac{\Gamma\left(\sum_{i=1}^m \theta_i\right)}{\Gamma\left(\sum_{i=1}^m (\theta_i + r_l)\right)} \frac{\Gamma(\theta_l + r_l)}{\Gamma(\theta_l)} = \frac{\prod_{i=0}^{r_l-1} (\theta_l + i)}{\prod_{i=0}^{r_l-1} (\Sigma\theta + i)}. \quad (6.6b)$$

Thus,

$$\begin{aligned} \text{cov}[P_k^{r_k}, P_l^{r_l}] &= E(P_k^{r_k} P_l^{r_l}) - E(P_k^{r_k}) E(P_l^{r_l}) \\ &= \frac{\prod_{i=0}^{r_k-1} (\theta_k + i) \prod_{i=0}^{r_l-1} (\theta_l + i)}{\prod_{i=0}^{r_k+r_l-1} (\Sigma\theta + i)} - \frac{\prod_{i=0}^{r_k-1} (\theta_k + i) \prod_{i=0}^{r_l-1} (\theta_l + i)}{\prod_{i=0}^{r_k-1} (\Sigma\theta + i) \prod_{i=0}^{r_l-1} (\Sigma\theta + i)} \\ &= \prod_{i=0}^{r_k-1} (\theta_k + i) \prod_{i=0}^{r_l-1} (\theta_l + i) \left[ \left( \prod_{i=0}^{r_k+r_l-1} (\Sigma\theta + i) \right)^{-1} - \left( \prod_{i=0}^{r_l-1} (\Sigma\theta + i) \right)^{-1} \right]. \end{aligned} \quad (6.7)$$

which for  $r_k = r_l = 1$ , reduces to

$$\text{cov}[P_k, P_l] = -\theta_k \theta_l (\Sigma\theta + 1)^{-1} \quad (6.8)$$

which illustrates that all seasonal proportions are negatively correlated.

## 6.2.2 Parameter estimation

*Scheme 1- MOME based on all first and second order moments; MOM1*

Several different methods of moment estimators (MOME) are available based on

moments of concern. Since  $(P_1, P_2, \dots, P_m)$  is multivariate beta distributed, the marginal distribution of the proportion associated with season  $j$ ,  $P_j$  is given by the beta distribution as:

$$P_j \sim \text{beta}(\theta_j, \Sigma\theta - \theta_j) \quad (6.9)$$

for  $j = 1, 2, \dots, m$  and, therefore, the first and second raw moments noted by  $\mu_{1,j}$  and  $\mu_{2,j}$  for  $P_j$ ,  $j = 1, 2, \dots, m$  are:

$$\mu_{1,j} = E(P_j) = \frac{\theta_j}{\Sigma\theta} \quad (6.10)$$

$$\begin{aligned} \mu_{2,j} = E(P_j^2) &= \int_0^1 \frac{1}{B(\theta_j, \Sigma\theta - \theta_j)} p_j^2 p_j^{\theta_j-1} (1-p_j)^{\Sigma\theta-\theta_j-1} dp_j \\ &= \frac{B(\theta_j + 2, \Sigma\theta - \theta_j)}{B(\theta_j, \Sigma\theta - \theta_j)} = \frac{\theta_j(\theta_j + 1)}{\Sigma\theta(\Sigma\theta + 1)}. \end{aligned} \quad (6.11)$$

Let  $M_{1,j}$  and  $M_{2,j}$  be the corresponding sample moments, i.e.,  $M_{1,j} = \frac{1}{n} \sum_{i=1}^n \frac{Y_{i,j}}{X_i}$  and

$M_{2,j} = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_{i,j}}{X_i} \right)^2$ . Equating the respective sample moments to their corresponding

population moments yields the moment equations  $M_{1,j} = \frac{\theta_j}{\Sigma\theta}$  and  $\frac{M_{2,j}}{M_{1,j}} = \frac{\theta_j + 1}{\Sigma\theta + 1}$ , the

solution of which gives the method of moments estimators (MOME)  $\hat{\theta}_j$ :

$$\hat{\theta}_{j,MOM1} = M_{1,j} \frac{M_{1,j} - M_{2,j}}{M_{2,j} - M_{1,j}^2}, \quad \text{for } j = 1, 2, \dots, m. \quad (6.12)$$

*Scheme 2- MOME based on all first order moments and the second order moment of season one; MOM2*

Solving the moment equations,  $M_{1,j} = \mu_{1,j} = \frac{\theta_j}{\Sigma\theta}$ , for  $j=1,2,\dots,m$  and

$M_{2,1} = \frac{\theta_1(\theta_1+1)}{\Sigma\theta(\Sigma\theta+1)}$  yields:

$$\hat{\theta}_{j,MOM2} = M_{1,j} \frac{M_{1,1} - M_{2,1}}{M_{2,1} - M_{1,1}^2}. \quad (6.13)$$

This MOM scheme has been used in Fielitz and Myers (1975) and Narayanan (1990).

Note that  $\hat{\theta}_{j,MOM1} = \hat{\theta}_{j,MOM2}$  for  $j=1$ . Also, under this scheme the MOME of  $\mu_{1,j}$  is

$$M_{1,j} \text{ since the MOME of } \mu_{1,j} = \frac{\hat{\theta}_j}{\Sigma\hat{\theta}} = \frac{M_{1,j} \left( \frac{M_{1,1} - M_{2,1}}{M_{2,1} - M_{1,1}^2} \right)}{\sum_{i=1}^m M_{1,i} \left( \frac{M_{1,1} - M_{2,1}}{M_{2,1} - M_{1,1}^2} \right)} = M_{1,j} \text{ noting } \sum_{i=1}^m M_{1,i} = 1.$$

*Scheme 3- MOME based on all first and second order moments and an average; MOM3*

The MOME of  $\theta_j$  in Eq.(6.13) uses the second order moment of only season one, and could be expected to do well for that season at the expense of other seasons.

The following estimation corrects such possible prejudice by replacing  $\frac{M_{1,1} - M_{2,1}}{M_{2,1} - M_{1,1}^2}$  by

$$\text{the average } \frac{1}{m} \sum_{i=1}^m \frac{M_{1,i} - M_{2,i}}{M_{2,i} - M_{1,i}^2} :$$

$$\hat{\theta}_{j,MOM3} = M_{1,j} \frac{1}{m} \sum_{i=1}^m \frac{M_{1,i} - M_{2,i}}{M_{2,i} - M_{1,i}^2} \quad (6.14)$$

Note that the MOME of  $\mu_{1,j}$  is  $M_{1,j}$  as it was for scheme 2.

### Maximum likelihood estimators (MLE)

From the probability density function of Dirichlet distribution of Eq.(6.2), the likelihood function can be given by:

$$\log f(\mathbf{p}; \boldsymbol{\theta}) = n \log \Gamma(\Sigma \theta) - n \sum_{j=1}^m \log \Gamma(\theta_j) + \sum_{j=1}^m \left( (\theta_j - 1) \log \sum_{i=1}^n p_{ij} \right) \quad (6.15)$$

where  $\mathbf{p} = (p_1, p_2, \dots, p_m)'$  of which  $p_j = (p_{ij})$  is the  $1 \times n$  sample vector for  $i = 1, 2, \dots, n$  and  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)'$ . Then the first derivatives of  $\log f(\mathbf{p}; \boldsymbol{\theta})$  with respect to  $\theta_j$  become:

$$\frac{\partial \log f(\mathbf{p}; \boldsymbol{\theta})}{\partial \theta_j} = n \Psi(\Sigma \theta) - n \Psi(\theta_j) - \log \sum_{i=1}^n p_{ij} \quad (6.16)$$

where  $\Psi(x) = d \log \Gamma(x) / dx$  is the digamma function. The maximum likelihood estimates (MLE) can be obtained by setting Eq.(6.16) equal to 0. However, since such equation does not have a closed solution, one could find the MLE by using a numerical method like Newton-Raphson. Alternatively, Fisher's scoring method (Narayanan, 1990) yields estimators that have the same asymptotic optimality as MLEs. Fisher's scoring method is given as:

$$\begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_m \end{pmatrix}_k = \begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_m \end{pmatrix}_{k-1} + \begin{pmatrix} Var(\hat{p}_1) & \cdots & Cov(\hat{p}_1, \hat{p}_m) \\ \vdots & \ddots & \vdots \\ Cov(\hat{p}_m, \hat{p}_1) & \cdots & Var(\hat{p}_m) \end{pmatrix}_{k-1} + \begin{pmatrix} g_1(\hat{\mathbf{p}}) \\ \vdots \\ g_m(\hat{\mathbf{p}}) \end{pmatrix}_{k-1} \quad (6.17)$$

where subscript  $k$  denotes the iteration,  $\hat{\mathbf{p}}_0 = (\hat{p}_1, \dots, \hat{p}_m)'$  is the vector of the initial estimates, and the components of the gradient vector are  $g_j(\mathbf{p}) = \partial \log f(\mathbf{p}; \boldsymbol{\theta}) / \partial \theta_j$ ,  $j = 1, 2, \dots, m$ .

The second derivatives of  $\log f(\mathbf{p}; \boldsymbol{\theta})$  with respect to  $\theta_j, \theta_k$  is simply derived from Eq.(6.15) as:

$$\begin{aligned} \frac{\partial^2 \log f(\mathbf{p}; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} &= n\Psi'(\Sigma\theta) \quad \text{for } j \neq k \\ \frac{\partial^2 \log f(\mathbf{p}; \boldsymbol{\theta})}{\partial \theta_j^2} &= n\Psi'(\Sigma\theta) - n\Psi'(\theta_j). \end{aligned} \quad (6.18)$$

The information matrix  $\mathbf{I}(\boldsymbol{\theta})$  of which  $j, k$  th element  $I(\theta_j, \theta_k)$  consists of the expected value of a negative second derivative of the log-likelihood function shows the minimal variance of the parameter estimates, which is defined by:

$$I(\theta_j, \theta_k) = \left[ E \left( - \frac{\partial^2 \log f(\mathbf{p}; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right) \right]_{\theta=E(\boldsymbol{\theta})} = \begin{cases} n\Psi'(\Sigma\theta) & j \neq k \\ n\Psi'(\Sigma\theta) - n\Psi'(\theta_j) & j = k \end{cases}. \quad (6.19)$$

Therefore, the asymptotic variance of MLE  $\hat{\boldsymbol{\theta}}$  is given by the inverse of information as:

$$Avar(\hat{\boldsymbol{\theta}}) = \mathbf{I}(\boldsymbol{\theta})^{-1} = \frac{1}{n} \begin{bmatrix} \Psi'(\Sigma\theta) - \Psi'(\theta_1) & \Psi'(\Sigma\theta) & \dots & \Psi'(\Sigma\theta) \\ \Psi'(\Sigma\theta) & \Psi'(\Sigma\theta) - \Psi'(\theta_1) & \dots & \Psi'(\Sigma\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi'(\Sigma\theta) & \Psi'(\Sigma\theta) & \dots & \Psi'(\Sigma\theta) - \Psi'(\theta_m) \end{bmatrix}^{-1}$$

and the asymptotic distribution of MLE  $\hat{\boldsymbol{\theta}}$  is:

$$\hat{\boldsymbol{\theta}}_{MLE} \sim AMVN(E(\boldsymbol{\theta}), Avar(\hat{\boldsymbol{\theta}})) \quad (6.20)$$

where  $\sim AMVN$  means ‘‘asymptotically distributed as multivariate normal’’.  $\mathbf{I}(\boldsymbol{\theta})^{-1}$

can be simplified by  $\mathbf{I}(\boldsymbol{\theta})^{-1} = \mathbf{D} + \gamma \mathbf{a} \mathbf{a}'$  (Narayanan, 1990) where  $\mathbf{D}$  is a diagonal

matrix given by  $\mathbf{D} = \text{diag}(n^{-1}\Psi'(\theta_1)^{-1}, n^{-1}\Psi'(\theta_2)^{-1}, \dots, n^{-1}\Psi'(\theta_m)^{-1})$ ,  $\mathbf{a}$  is a  $m \times 1$  vector denoted by  $\mathbf{a} = (\Psi'(\theta_1)^{-1}, \Psi'(\theta_2)^{-1}, \dots, \Psi'(\theta_m)^{-1})$ , and  $\gamma$  is a scalar given by  $\gamma = n\Psi'(\Sigma\theta) \left( 1 - \Psi'(\Sigma\theta) \sum_{i=1}^m \frac{1}{\Psi'(\theta_m)} \right)^{-1}$ . Estimators obtained by using the method of scoring have the same asymptotic distribution as MLEs given in Eq.(6.20).

## 6.3 Application to temporal disaggregation

### 6.3.1 Comparative analysis of different estimators

Different parameter estimators based on MOM and ML are compared by using simulations for two different streamflows data sets at Lee's Ferry in Colorado River Basin and St. Lawrence River. Table 6.1 illustrate that over 70% of annual flows are contributed by the three months of May, June, and July at Lee's Ferry, while the monthly flow proportions are more similar on the St. Lawrence River. Based on assumed true parameter sets, 5000 different proportion sets are generated and parameters are re-estimated by using the three different MOMEs, as well as the corresponding three different "MLEs" (Here MLEs in quotes refer to estimators obtained by the method of scoring) obtained by using the three different MOM estimates as the initial estimates. Figures 6.1 through 6.4 represent the distributions of the means and standard deviations regarding six different estimates of seasonal proportions. A comparison based on the variability of generated statistics shows that MOM3 reproduces the means and variances of proportions with the smallest variability among three different MOMEs. It is also

illustrated that MOM3 could reproduce the mean and variances at each season with similar variability as the “MLEs”. Similar distributions of generated means and variances are expected for all “MLEs”, which says that the “MLEs” are not significantly dependent on initial estimates as expected. This difference between different parameter estimators is similarly found in the case of the St. Lawrence River. Note that MOME results show negative bias of generated variances compared with “MLE”. This results from the small sample size ( $n=59$ ) and it becomes smaller when the related small sample size is larger (see Figure 6.A1 in Appendix). Thus, preference would be given to “MLE” as the appropriate parameter estimation procedure for proportional disaggregation models. However, the MOM3 scheme does almost as well as the “MLE” scheme. Further, the asymptotic distribution of the MOM3 can be found using the central limit theorem and the delta method and can be expressed in closed form, as opposed to the asymptotic theory associated with “MLEs” used to get the asymptotic distribution of the “MLE” estimators.

### **6.3.2 Model evaluation based on sample statistics**

The evaluation of the PD model was performed by examining the generated sample statistics for the above two streamflows sets. As in traditional disaggregation, the annual flows are generated outside the disaggregation model where AR(1) models are applied for both annual flows generations where historical flows are normalized through power transformations. Proportions at different seasons are calculated from the given historical sample and the Dirichlet distribution is fitted to the seasonal proportions when

parameters are estimated by using MLE3. 5000 different annual flows sets were first generated with the same size as the historical annual flows, and in each generation annual flows were disaggregated into monthly flows by using seasonal proportions generated from the Dirichlet distribution with the same size of the historical flows. Main statistics at each season were calculated from generated seasonal flows, and quantile distributions of those statistics are provided in Figure 6.5 and Figure 6A.2 in the Appendix. Mean statistics are shown to be fairly well reproduced, but it is not the case for other statistics in both sites. Because of its simplicity (only 12 parameters), the PD model shows the lack of ability to capture higher order sample moments even though the first moment could be implicitly preserved through the simulation. One feature of the PD model that illustrates inflexibility is the following. If  $\theta_i < \theta_j$ , then mean flow for season  $i$  is less than the mean flow for season  $j$ ; further, it can be shown that if  $\theta_i < \theta_j$ , then the variance of the flow for season  $i$  is less than the variance of the flow for season  $j$ .

### **6.3.3 Parameter uncertainty incorporation**

The effect of parameter uncertainty in the PD model was compared by combining with the uncertainty effect in annual flow generations. The similar simulation procedure in the previous section has been applied with additional consideration of four different cases based on parameter uncertainty: (1) AnMu; (2)AnMu; (3)AuMn; and (4)AuMu where ‘A’ means annual, ‘M’ means monthly disaggregation, ‘n’ means natural uncertainty (no parameter uncertainty), and ‘u’ means parameter uncertainty. For example, in AnMu monthly flows are disaggregated with

parameter uncertainty incorporated from generated annual flows with only natural uncertainty incorporated. Parameter uncertainty effects are closely related with the available sample size, which were examined through the simulation by using different available samples sizes assumed as:  $n=25, 50, 100,$  and  $200$ . In each generation, new parameter sets were sampled from the derived asymptotic distribution and substituted for the original parameter estimates. Figure 6.6 and Figure 6.A3-6.A7 in the Appendix provide basic statistics calculated from generated monthly streamflows. Also, design variables, e.g. storage capacity, critical drought magnitude, critical drought length, and critical drought intensity, were calculated based on generated monthly streamflows. Figures 6.7-6.10 and Figures 6.A8-6.A11 in the Appendix illustrate the quantile estimates of design variables for different sample sizes and demand levels for the Colorado River and the St. Lawrence River. The parameter uncertainty effect in the disaggregation model seems very small (actually hard to distinguish) compared to the parameter uncertainty effect in the annual flow generation on the generated design variables, which is consistent with the traditional disaggregation models in the previous chapter.

## **6.4 Summary and Conclusion**

A simple disaggregation model was developed based on Dirichlet distribution for the purpose of overcoming the discrepancy caused by the normality limitation in the application to real-space flows. The model structure is very simple and only 12 parameters are required for the case of the annual-monthly temporal disaggregation when compared with the traditional Valencia-Schaake model (156 parameters) and even the

condensed Lane's model (36 parameters). From the comparison of possible different parameter estimators, a maximum likelihood estimator shows the smallest variability of parameter estimates.

The application to the real data sets demonstrates that only mean statistics could be well preserved through the simulation, not for higher order sample moments. However, the PD model preserves the design variables well regardless of demand levels, which is not always the case in traditional disaggregation models. Associated with the parameter uncertainty effect, the uncertain annual flows play the more significant role in the variability of generated monthly flows rather than the uncertainty of the PD disaggregation model. Unfortunately, this PD model does not show enough flexibility. That is, model verification would be needed in advance to the actual application. If the PD model is the true model then it performs well and is useful in addressing parameter uncertainty considerations.

Table 6.1: Basic statistics of annual and monthly streamflows (Lee's Ferry in Colorado River)

	Mean (ac-ft)	S. Dev. (ac-ft)	Cv	skewness	lag-1 corr.	proportion to annual
<i>Lee's Ferry in Colorado River</i>						
Oct	580893	272006	0.47	1.64	0.54	3.9
Nov	480821	141531	0.29	1.21	0.76	3.2
Dec	382530	95858	0.25	1.22	0.83	2.5
Jan	356611	78632	0.22	0.59	0.70	2.4
Feb	393776	97576	0.25	1.42	0.55	2.6
Mar	645201	211390	0.33	1.08	0.48	4.3
Apr	1199946	512460	0.43	0.96	0.47	8.0
May	3037199	1146760	0.38	0.27	0.59	20.1
Jun	4054340	1572353	0.39	0.43	0.63	26.9
Jul	2190444	1012249	0.46	1.13	0.83	14.5
Aug	1083174	423971	0.39	0.95	0.78	7.2
Sep	671371	309698	0.46	1.95	0.64	4.5
Annual (real)	15076307	4365301	0.29	0.14	0.28	
Annual (transformed)	549100	128720	0.23	0.01	0.28	
<i>St. Lawrence River</i>						
Oct	3997	357	0.09	-0.37	0.98	8.3
Nov	3790	345	0.09	-0.38	0.98	7.9
Dec	3871	350	0.09	-0.36	0.96	8.1
Jan	3765	380	0.10	-0.11	0.89	7.8
Feb	3352	348	0.10	-0.15	0.93	7.0
Mar	3884	397	0.10	0.15	0.91	8.1
Apr	4100	368	0.09	0.09	0.92	8.5
May	4376	405	0.09	0.07	0.96	9.1
Jun	4278	413	0.10	0.01	0.97	8.9
Jul	4382	420	0.10	-0.13	0.98	9.1
Aug	4261	403	0.09	-0.24	0.99	8.9
Sep	3984	370	0.09	-0.31	0.99	8.3
Annual (real)	48042	4157	0.09	-0.28	0.75	
Annual (transformed)	2.325X10 <sup>9</sup>	3.95X10 <sup>8</sup>	0.17	-0.07	0.73	

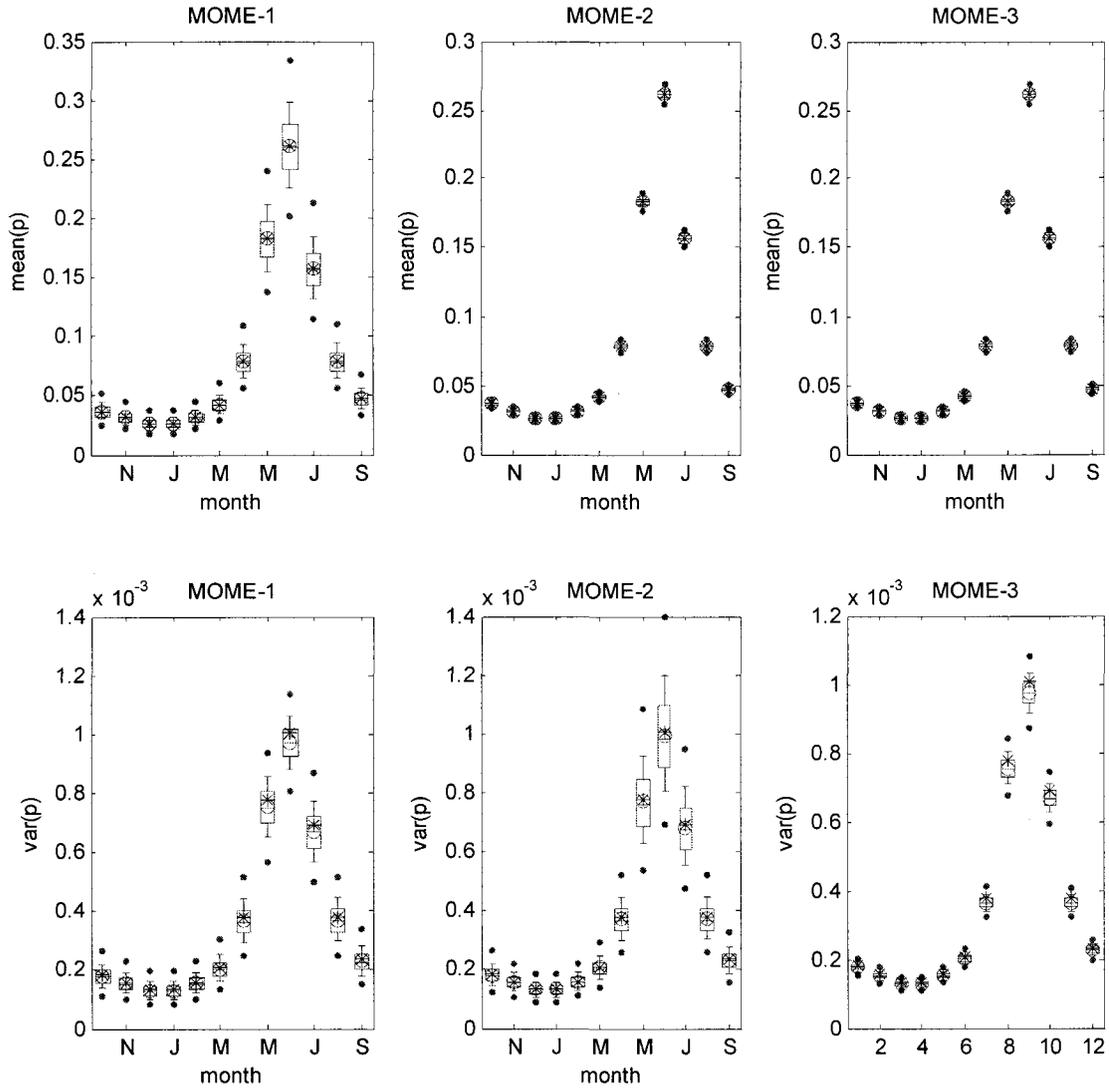


Figure 6.1: Distributions of means and standard deviations of proportions. (MOME, Lee's Ferry in Colorado River basin) where '\*' denotes the mean and standard deviation calculated from assumed true parameters.

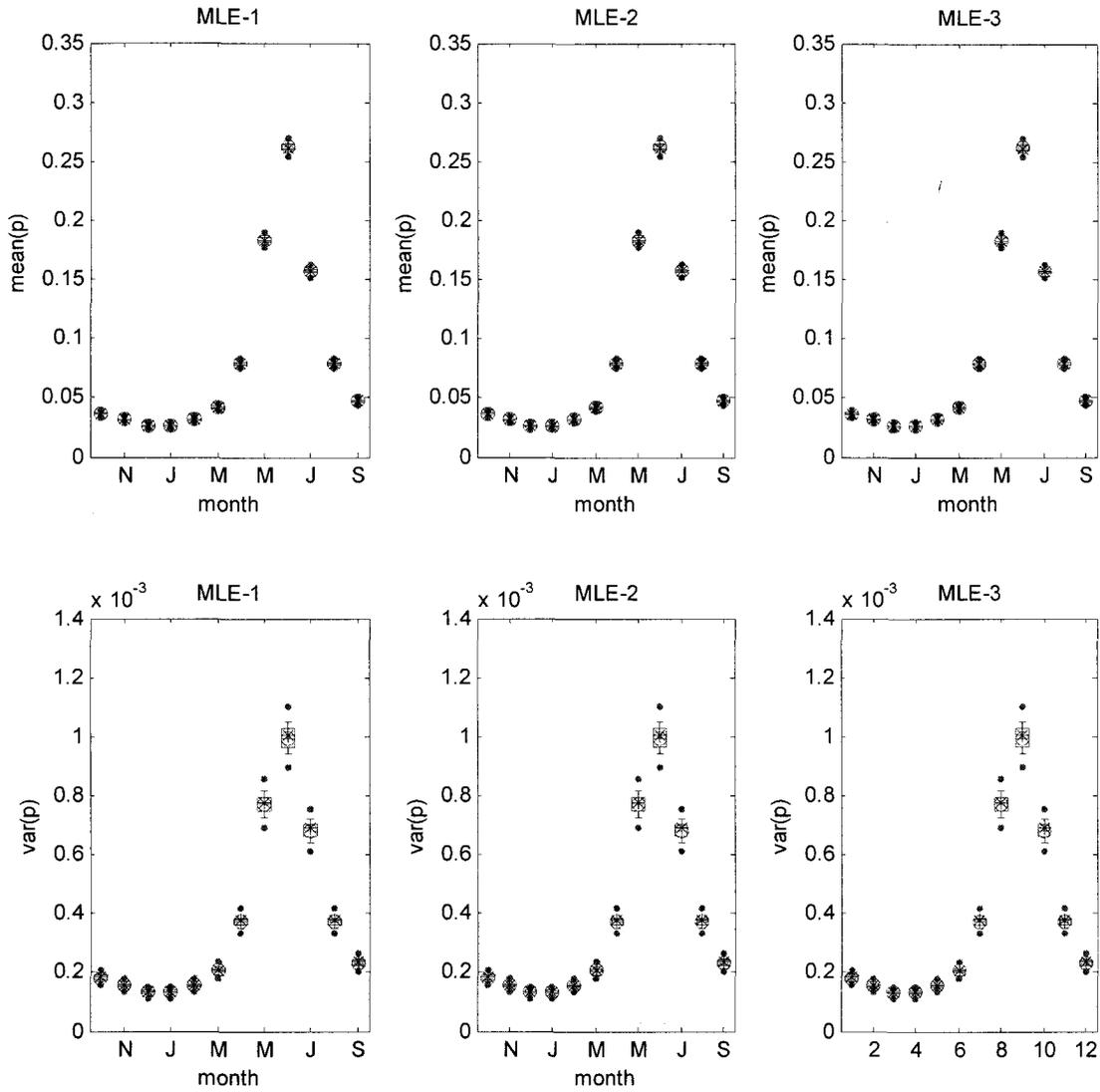


Figure 6.2: Distributions of means and standard deviations of proportions. (MLE, Lee's Ferry in Colorado River basin) where '\*' denotes the mean and standard deviation calculated from assumed true parameters.

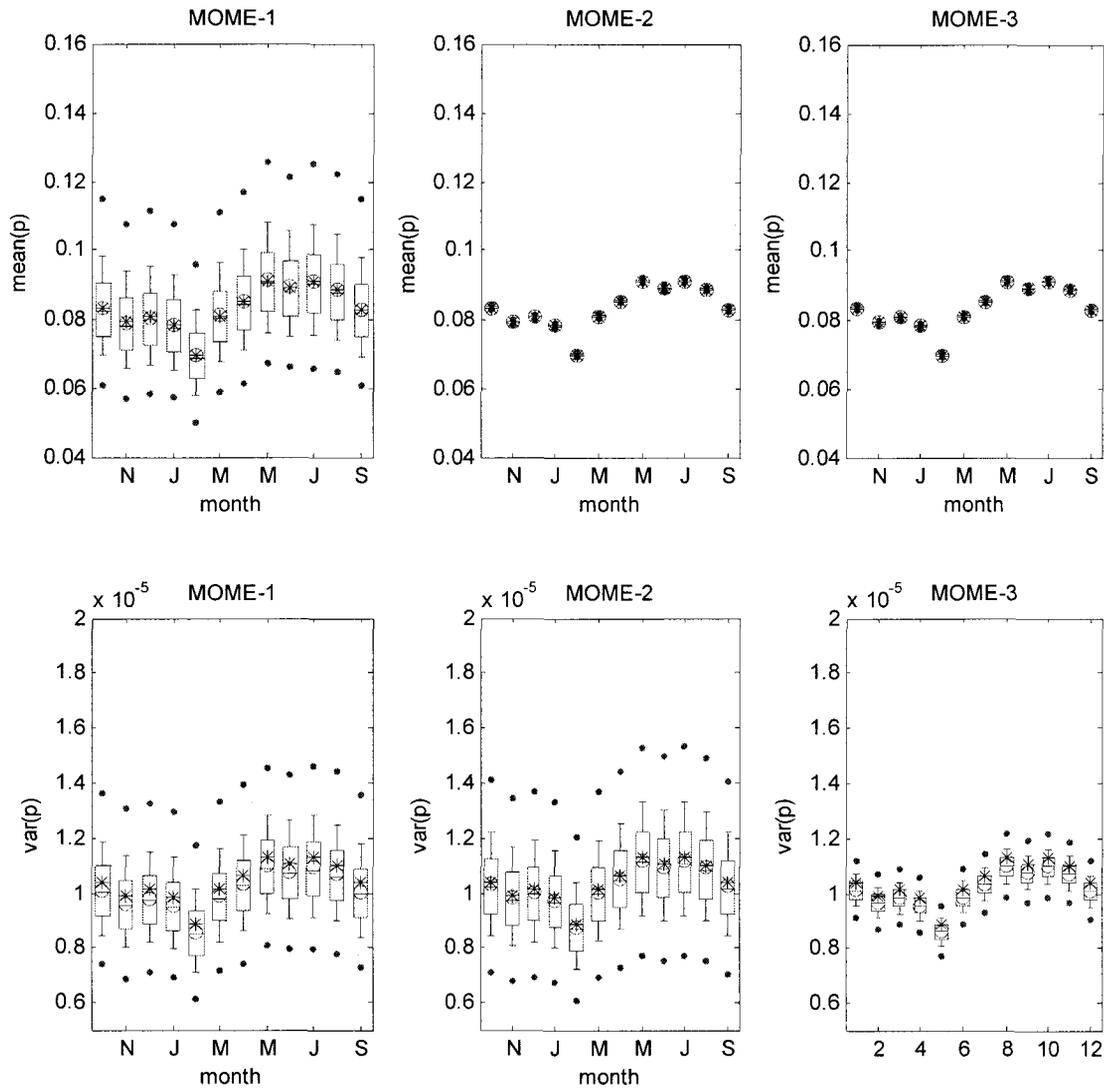


Figure 6.3: Distributions of means and standard deviations of proportions. (MOMe, St. Lawrence River) where ‘\*’ denotes the mean and standard deviation calculated from assumed true parameters.

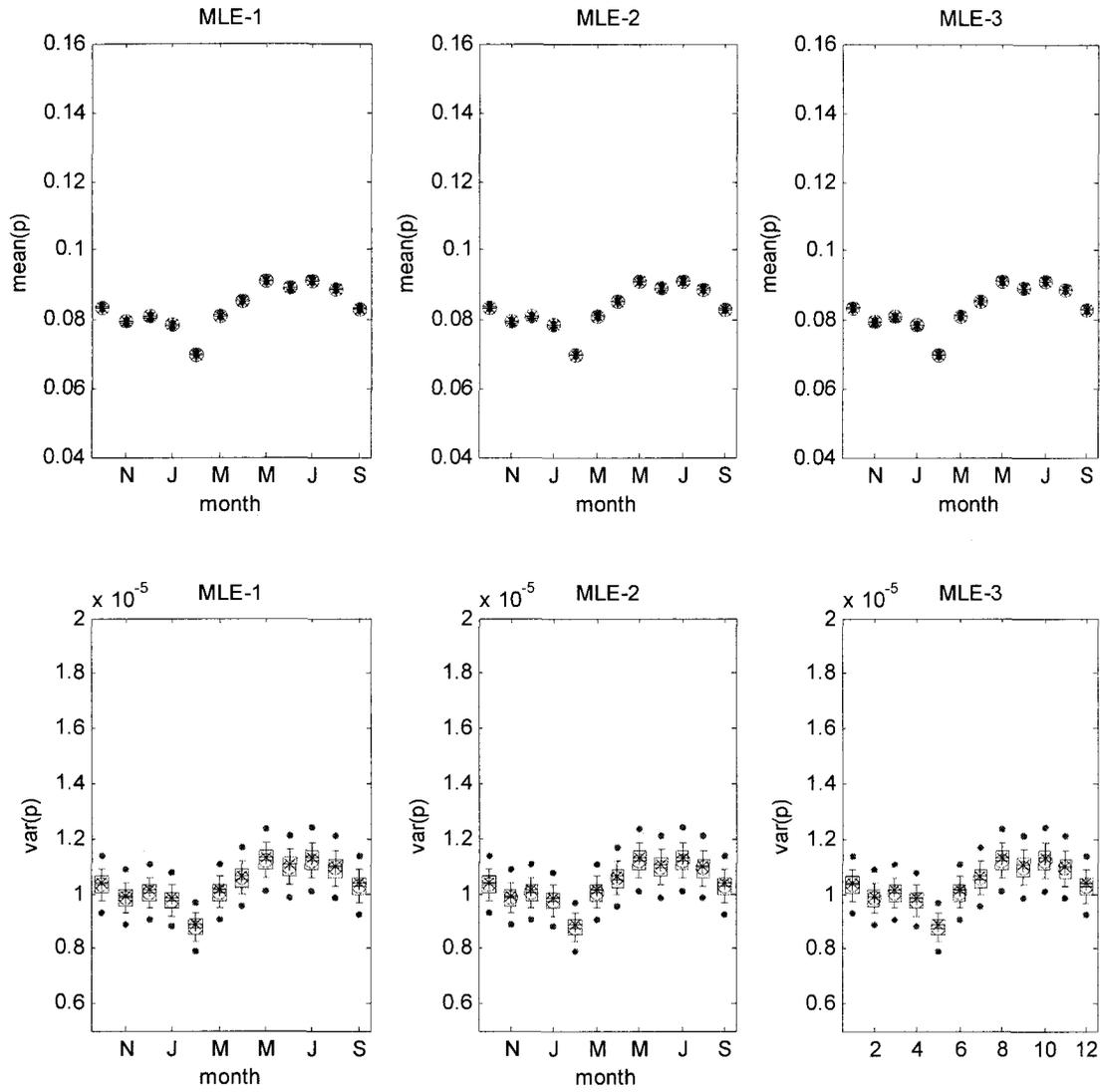


Figure 6.4: Distributions of means and standard deviations of proportions. (MLE, St. Lawrence River) where ‘\*’ denotes the mean and standard deviation calculated from assumed true parameters.

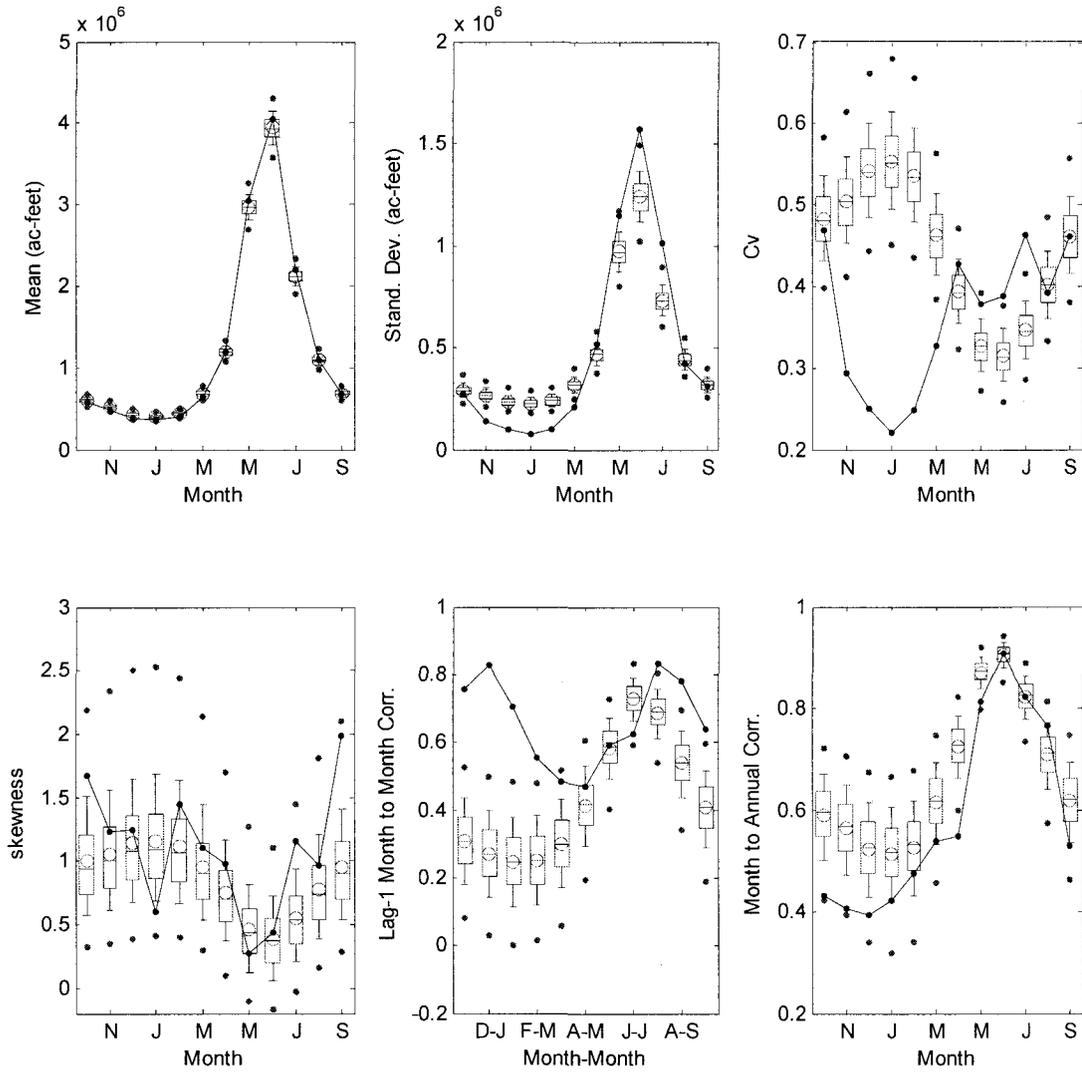


Figure 6.5: Basic statistics of generated monthly streamflows (AnMn, Lee's Ferry in Colorado River basin)

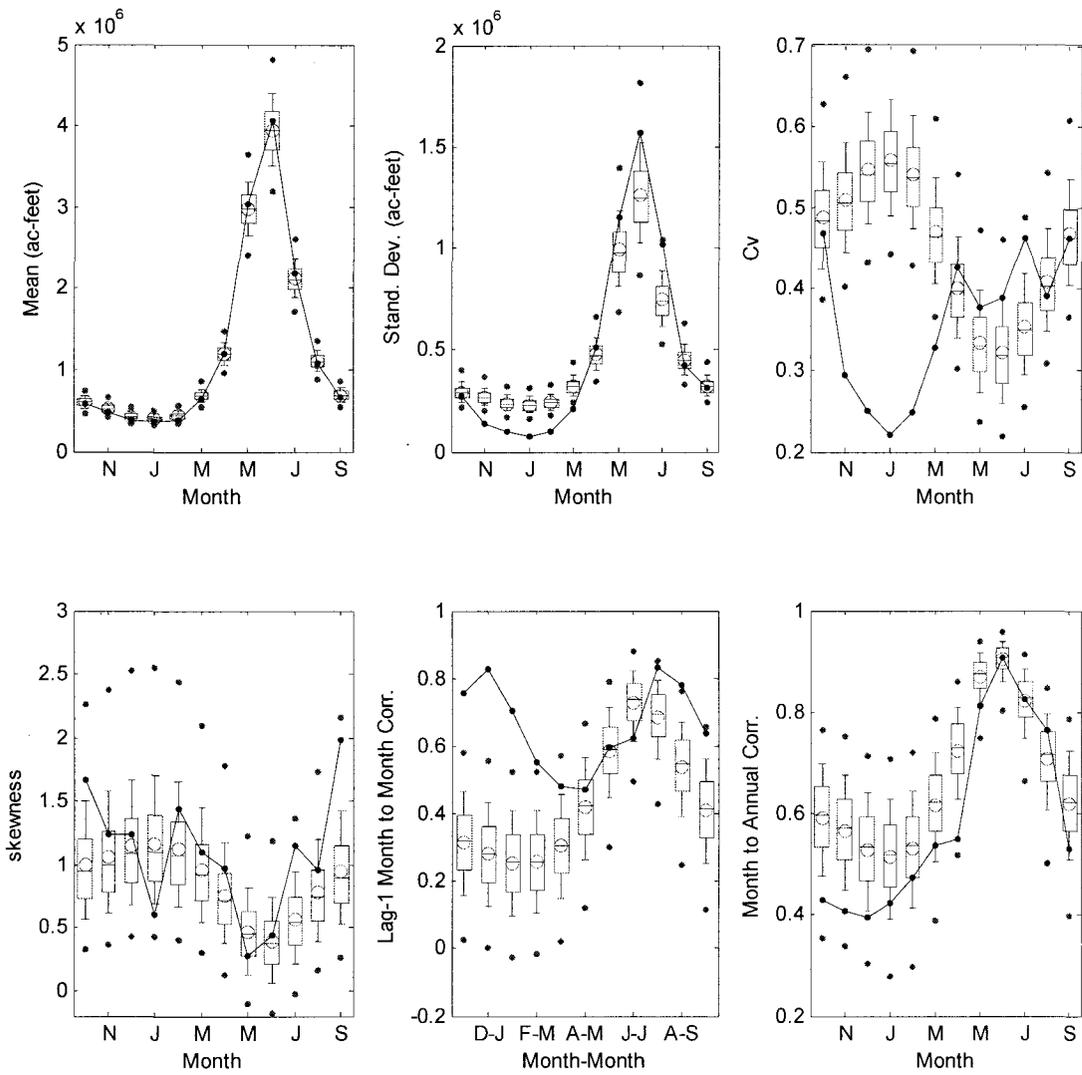


Figure 6.6: Basic statistics of generated monthly streamflows (AnMu,  $n=25$ , Lee's Ferry in Colorado River basin)

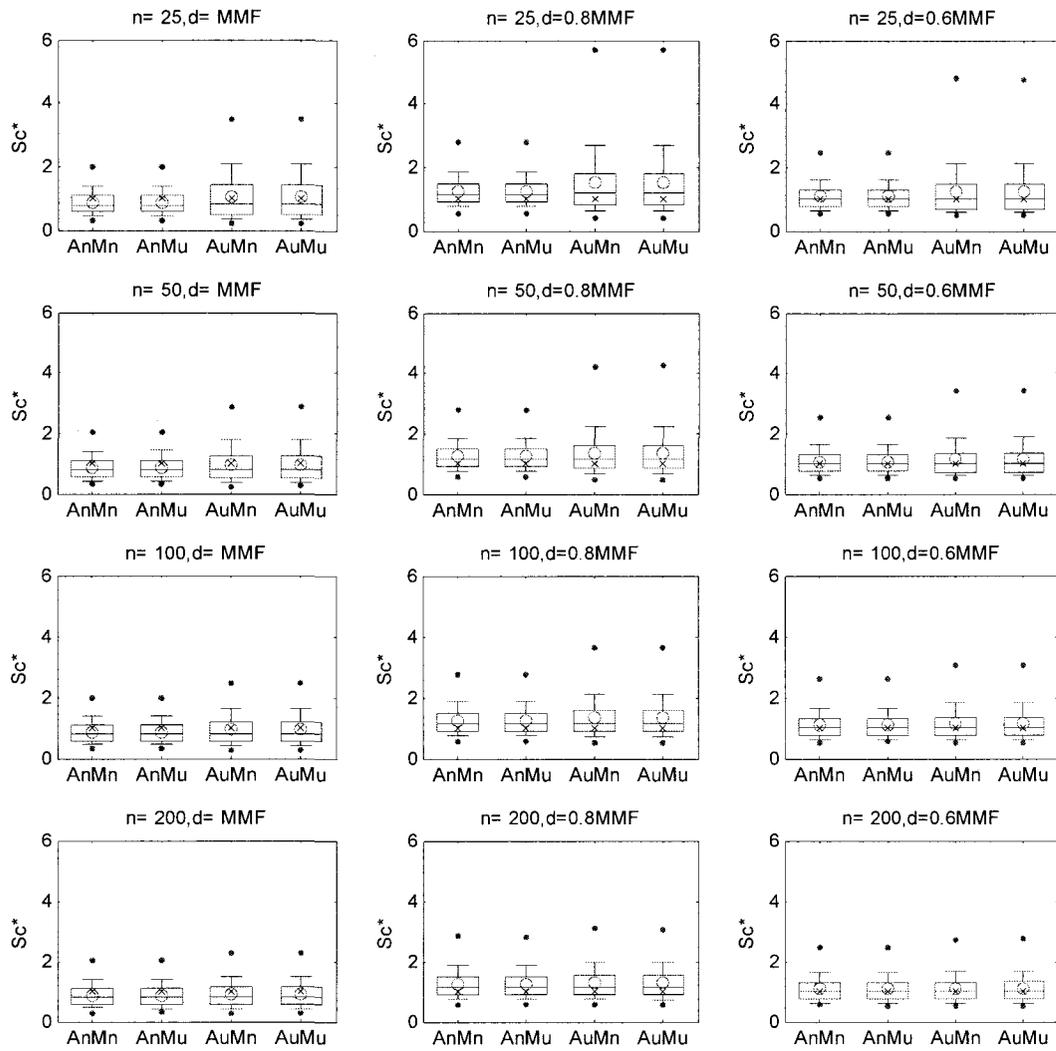


Figure 6.7: Quantile estimates of generated storage capacities  $Sc^*$  (Lee's Ferry in Colorado River basin) where \* denotes scaled by historical storage capacity.

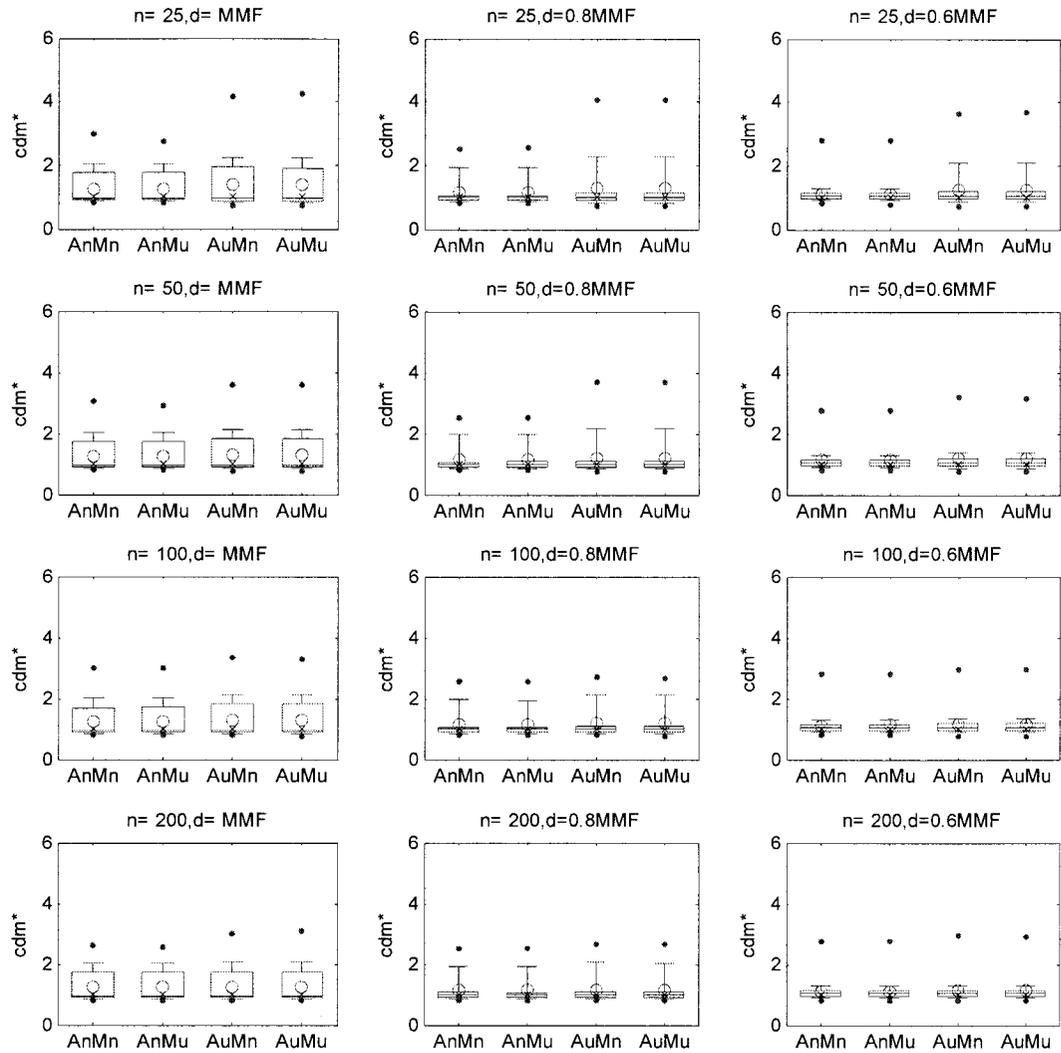


Figure 6.8: Quantile estimates of generated critical drought magnitude  $cdm^*$  (Lee's Ferry in Colorado River basin) where \* denotes scaled by historical storage capacity.

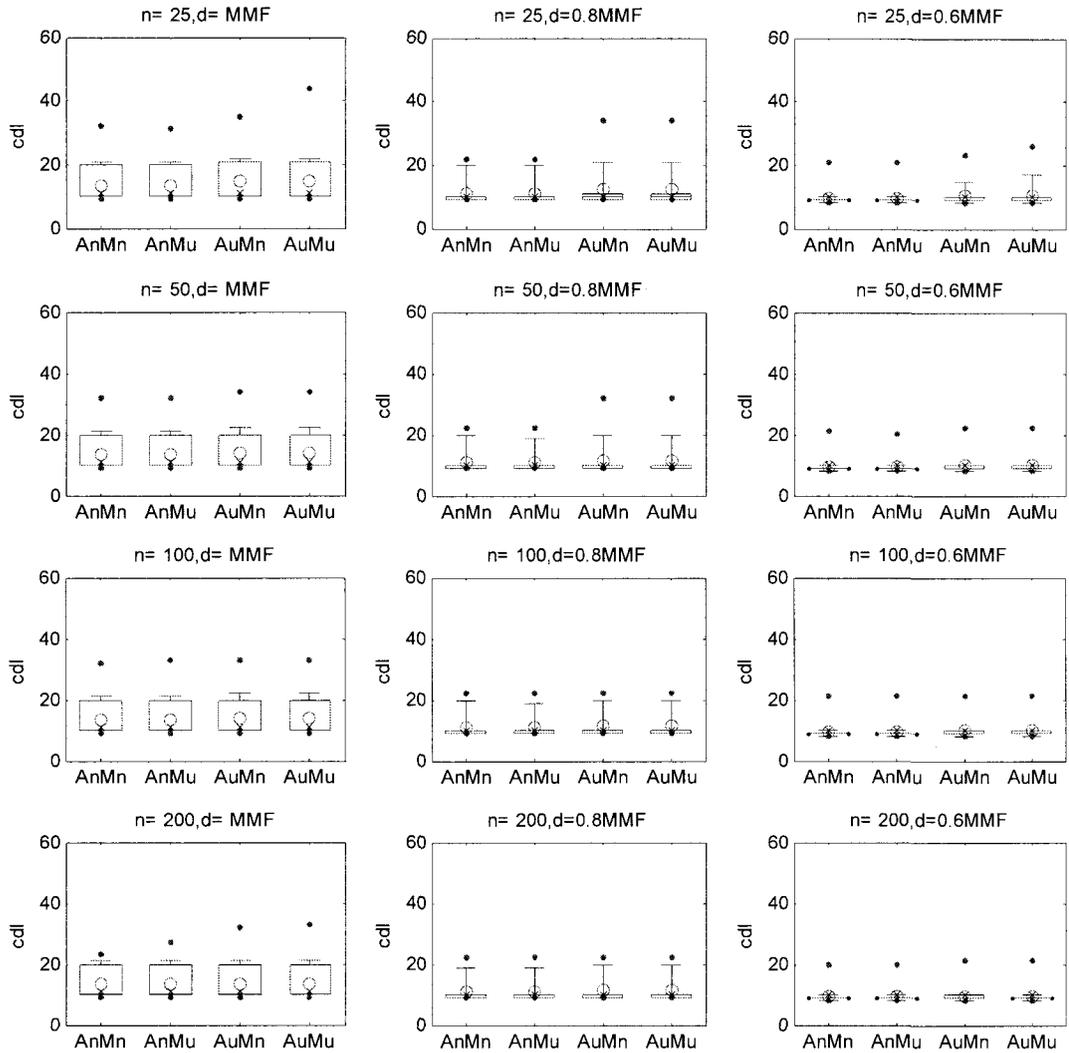


Figure 6.9: Quantile estimates of generated critical drought magnitude  $cdl$  (Lee's Ferry in Colorado River basin).

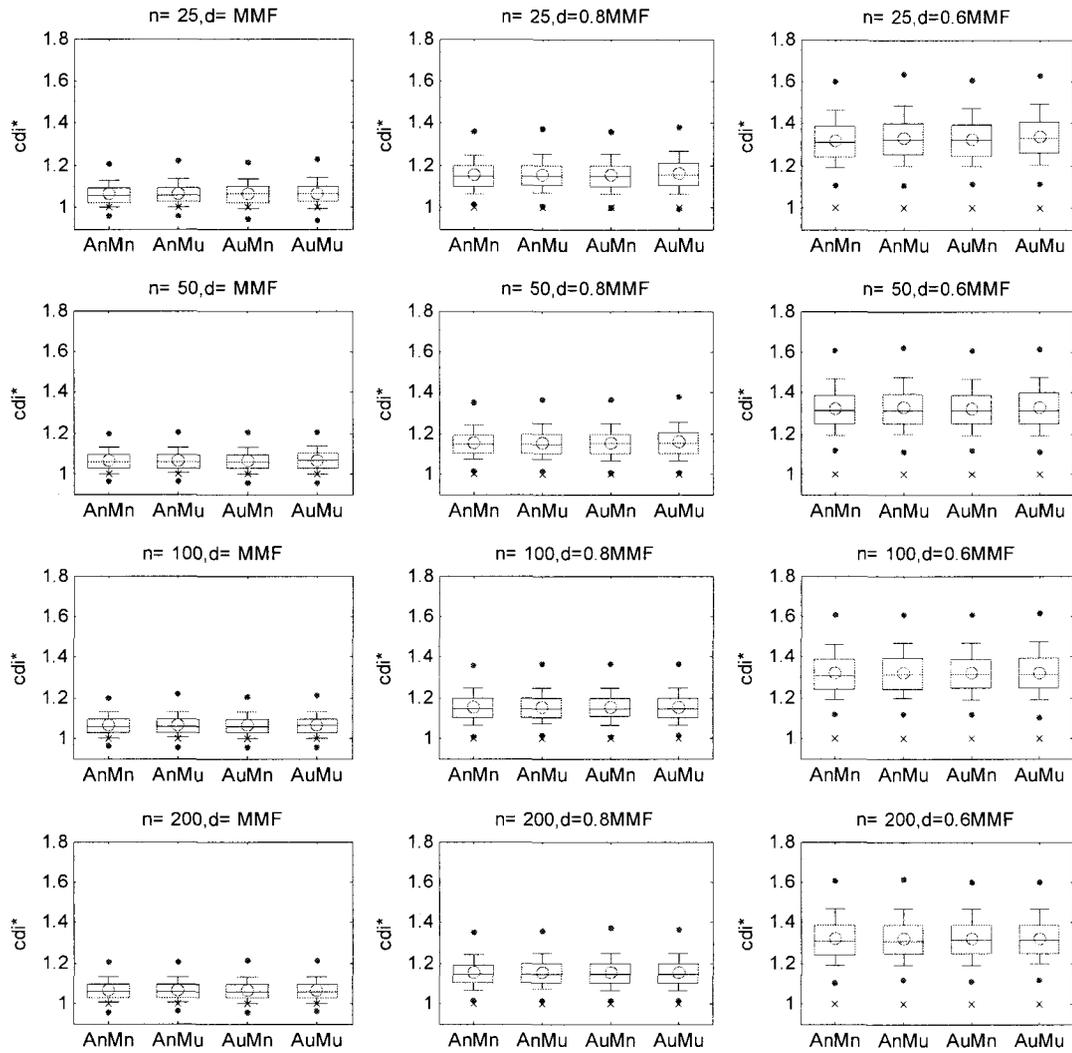


Figure 6.10: Quantile estimates of generated critical drought magnitude  $cdi^*$  (Lee's Ferry in Colorado River basin) where \* denotes scaled by historical storage capacity.

## Reference

- Fielitz, B.D. and Myers, B.L. (1975). Estimation of parameters in the beta distribution, *Decision Sciences*, 6, pp. 1-13.
- Grygier, J.C. and J.R. Stedinger (1988). Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research*, 24(10), pp. 1574-1584.
- Grygier, J.C. and J.R. Stedinger (1990). SPIGOT, A synthetic Streamflow Generation Software Package, technical description, version 2.5, School of Civil and Environmental Engineering, Cornell University, Ithaca, N. Y.
- Lane, W.L. (1979). *Applied Stochastic Techniques, User's Manual*, Bureau of Reclamation, Engineering and Research Center, Denver, Co.
- Lane, W.L. and D.K. Frevert (1990). *Applied Stochastic Techniques: User's Manual, personal computer version 5.2*, Earth Sciences Division, Bureau of Reclamation, U.S. Department of the Interior, Denver, Colo.
- Mejia, J.M. and J. Roussell (1976). Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), pp. 185-186.
- Narayanan, A. (1991). Algorithm as 266: Maximum likelihood estimation of the parameters of the Dirichlet distribution, *Applied Statistics*, 40, pp. 365-374.
- Pereira, M.V.F., G.C. Olivera, C.C.G. Costa, and J. Kelman (1984). Stochastic streamflow models for hydroelectric systems, *Water Resources Research*, 20(3), pp. 379-390.
- Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.
- Santos, E.G. and J.D. Salas (1994). Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, 118(5), pp. 765-784.
- Stedinger, J.R. and R.M. Vogel (1984). Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(1), pp. 47-56.
- Stedinger, J.R., D. Pei, and T. Cohn (1985). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resources Research*, 21(5), pp. 665-675.
- Tao, P.C. and J.W. Delleur (1976). Multistation, multiyear synthesis of hydrologic time

series by disaggregation, *Water Resources Research*, 12(7), pp. 1303-1312.

Valencia, D. and J.C. Schaake, Jr. (1973). Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), pp. 580-585, 1973.

## Appendix 6.A: Additional Figures and Tables

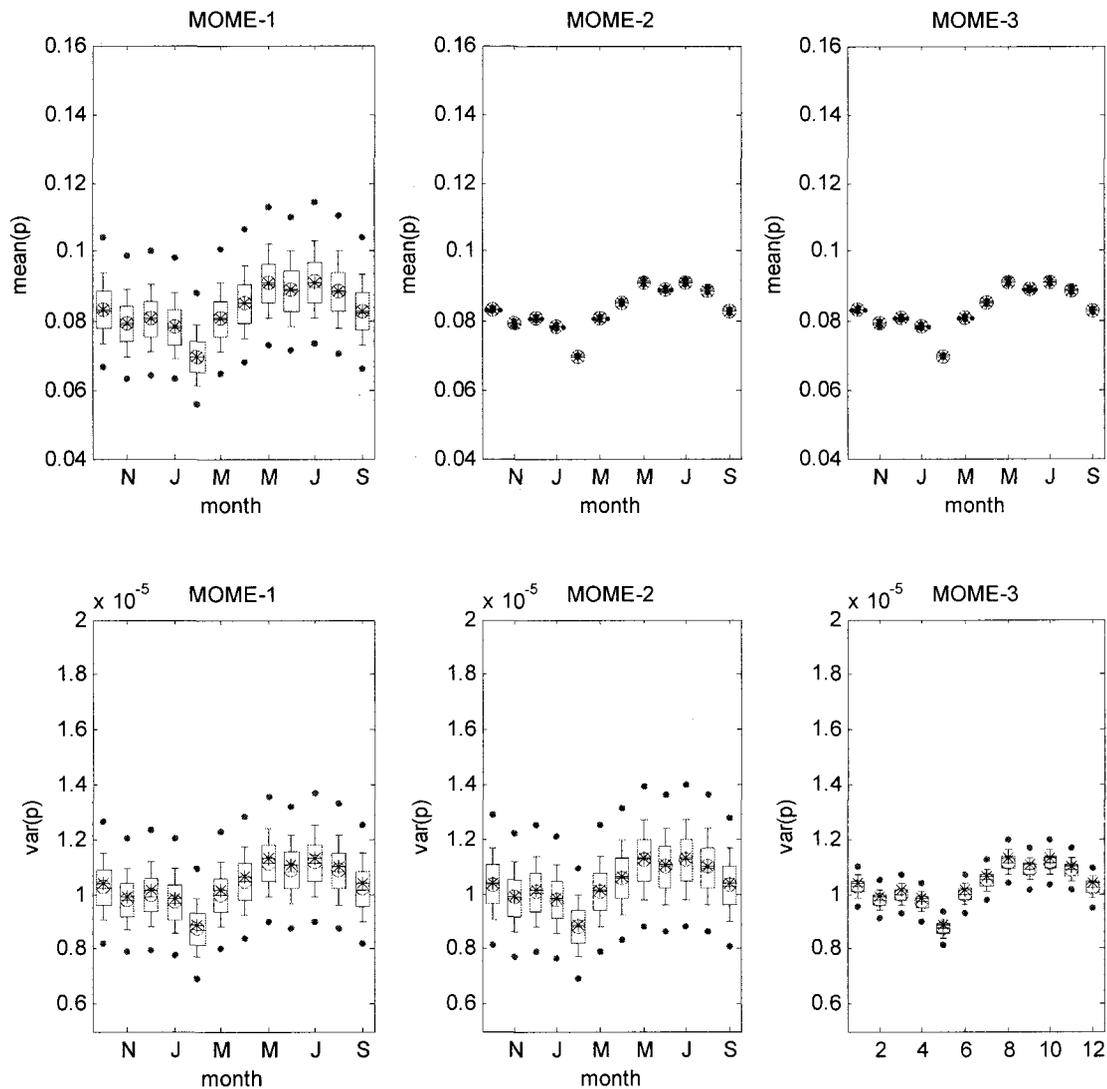


Figure 6.A1: Distributions of means and standard deviations of proportions. (MOME, St. Lawrence River) where ‘\*’ denotes the mean and standard deviation calculated from assumed true parameters.

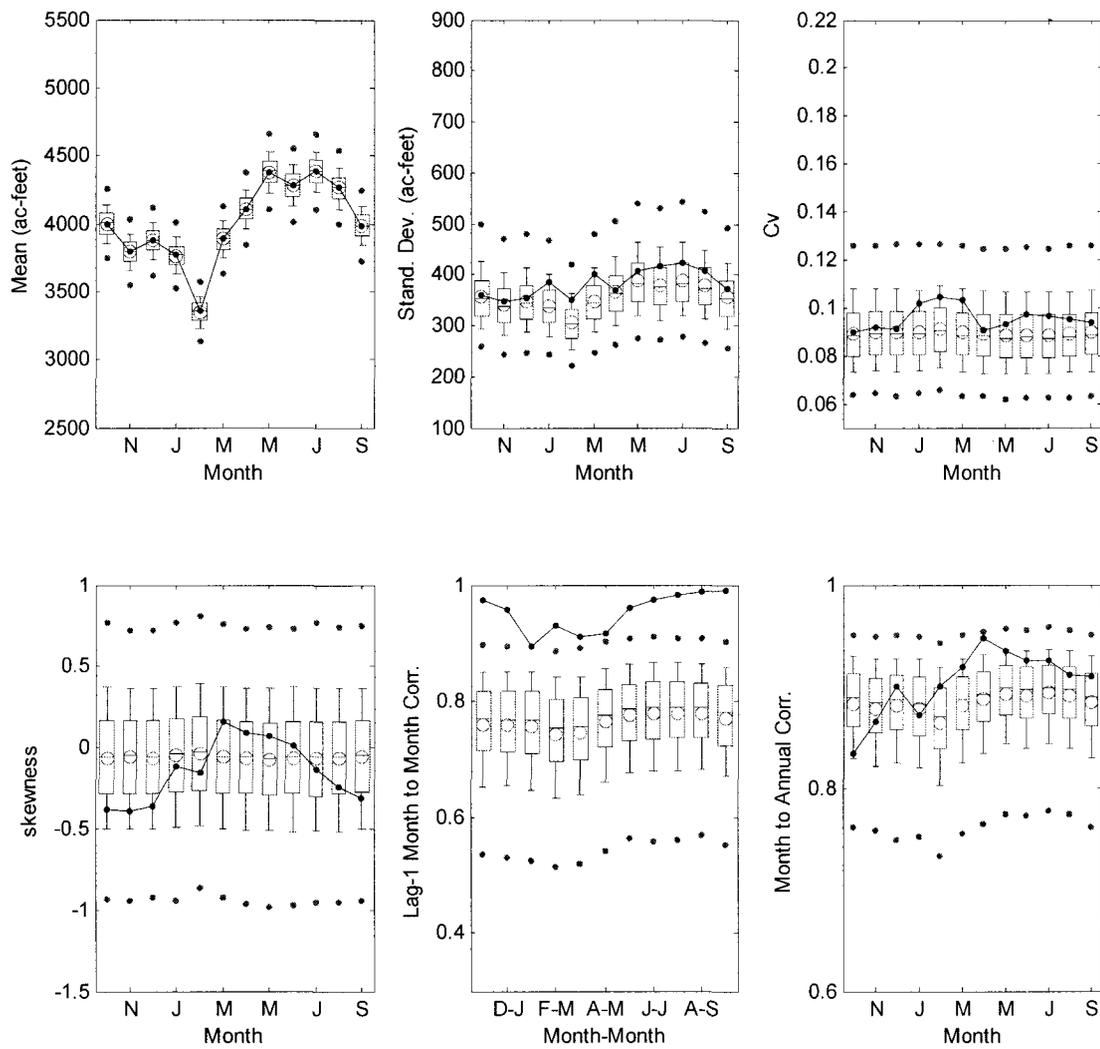


Figure 6.A2: Basic statistics of generated monthly streamflows (St. Lawrence River)

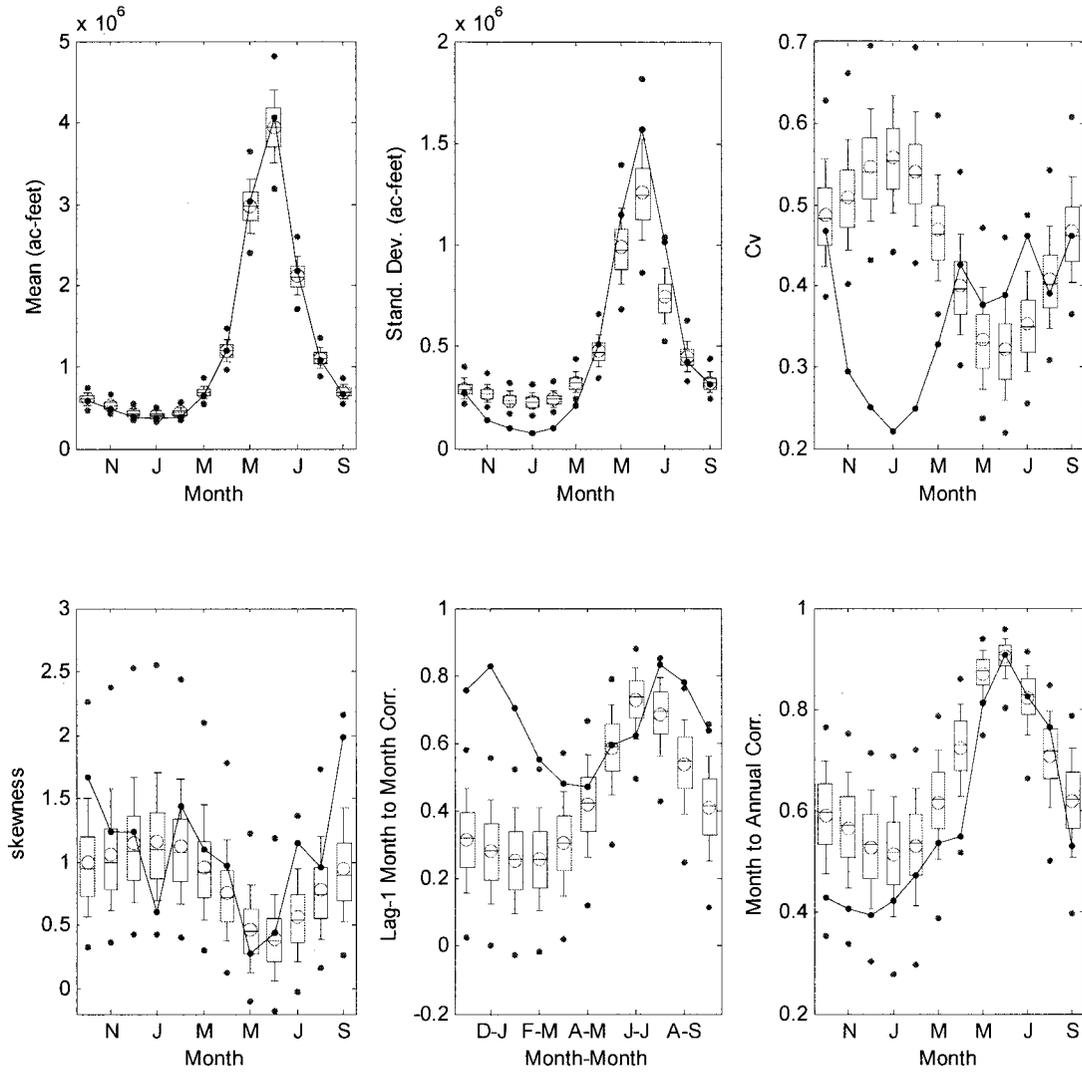


Figure 6.A3: Basic statistics of generated monthly streamflows (AuMn,  $n=25$ , Lee's Ferry in Colorado River basin)

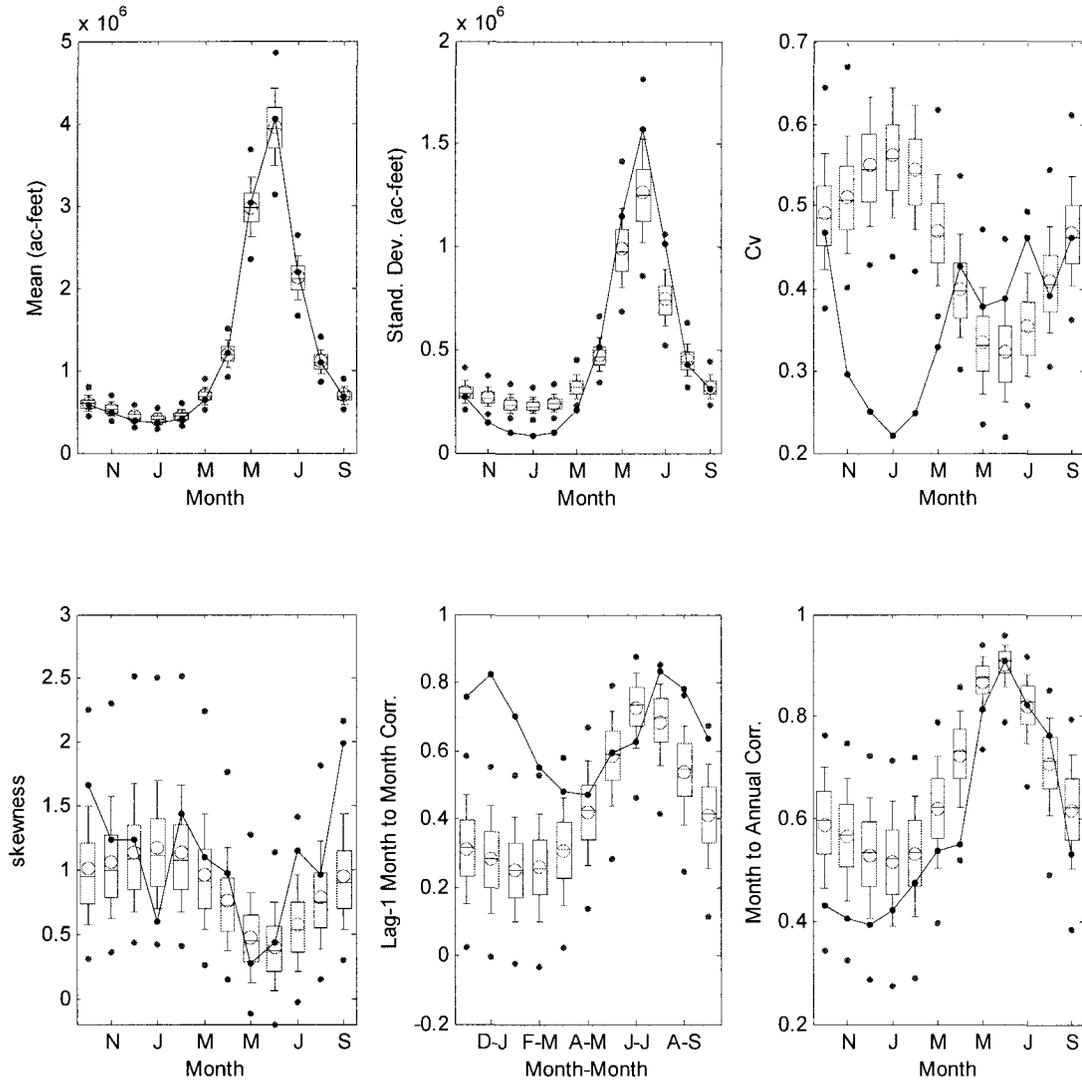


Figure 6.A4: Basic statistics of generated monthly streamflows (AuMu,  $n=25$ , Lee's Ferry in Colorado River basin)

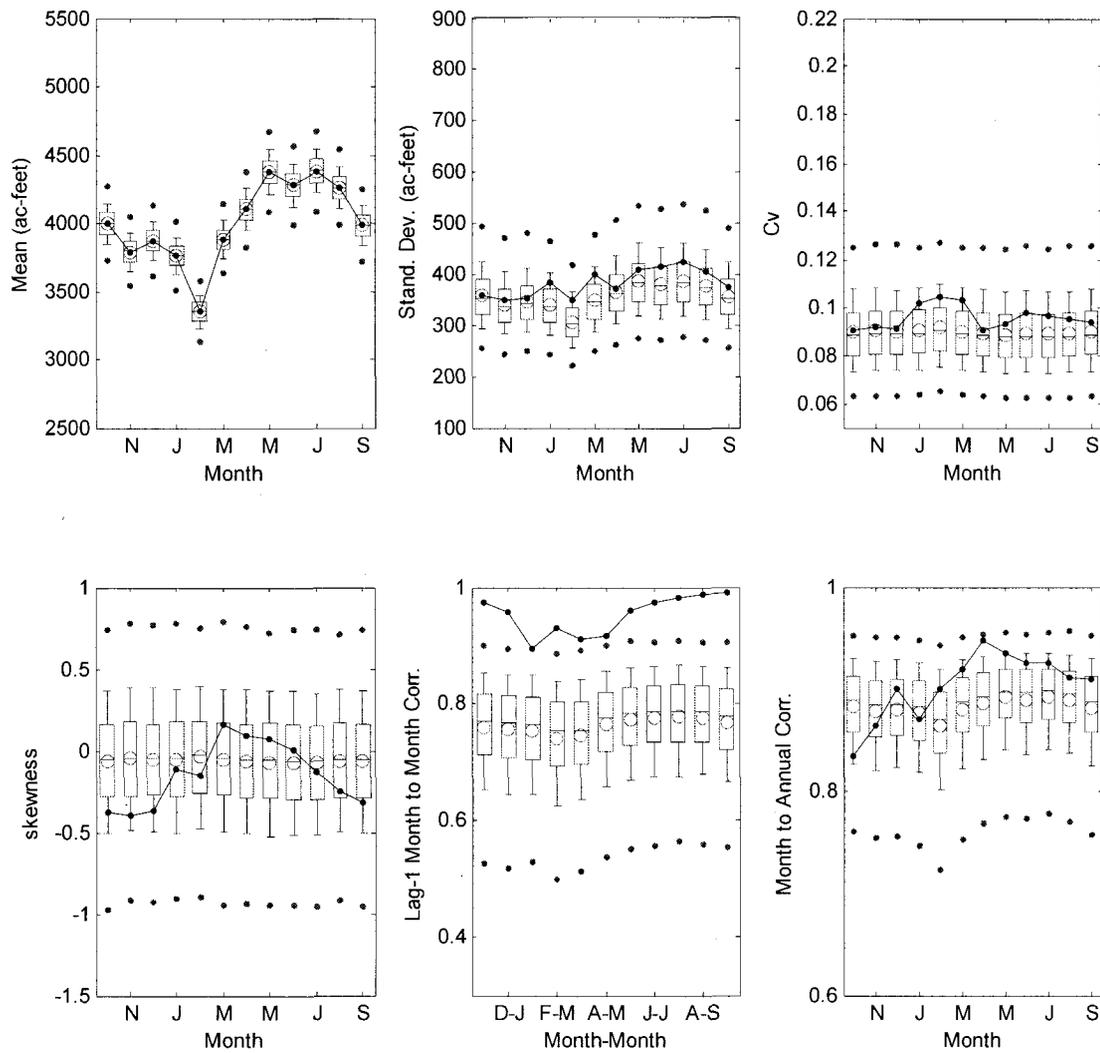


Figure 6.A5: Basic statistics of generated monthly streamflows (AnMu,  $n=25$ , St. Lawrence River)

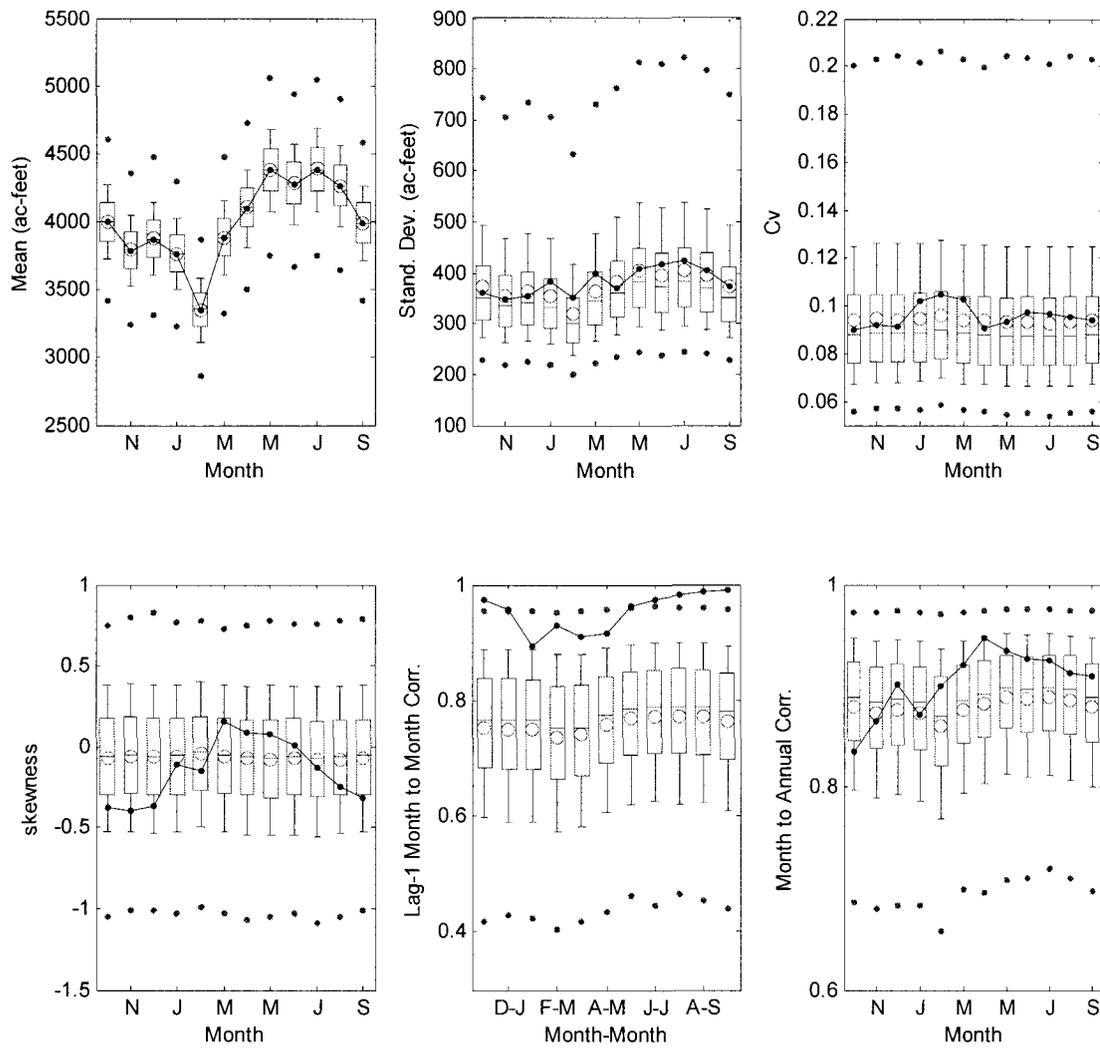


Figure 6.A6: Basic statistics of generated monthly streamflows (AuMn,  $n=25$ , St. Lawrence River)

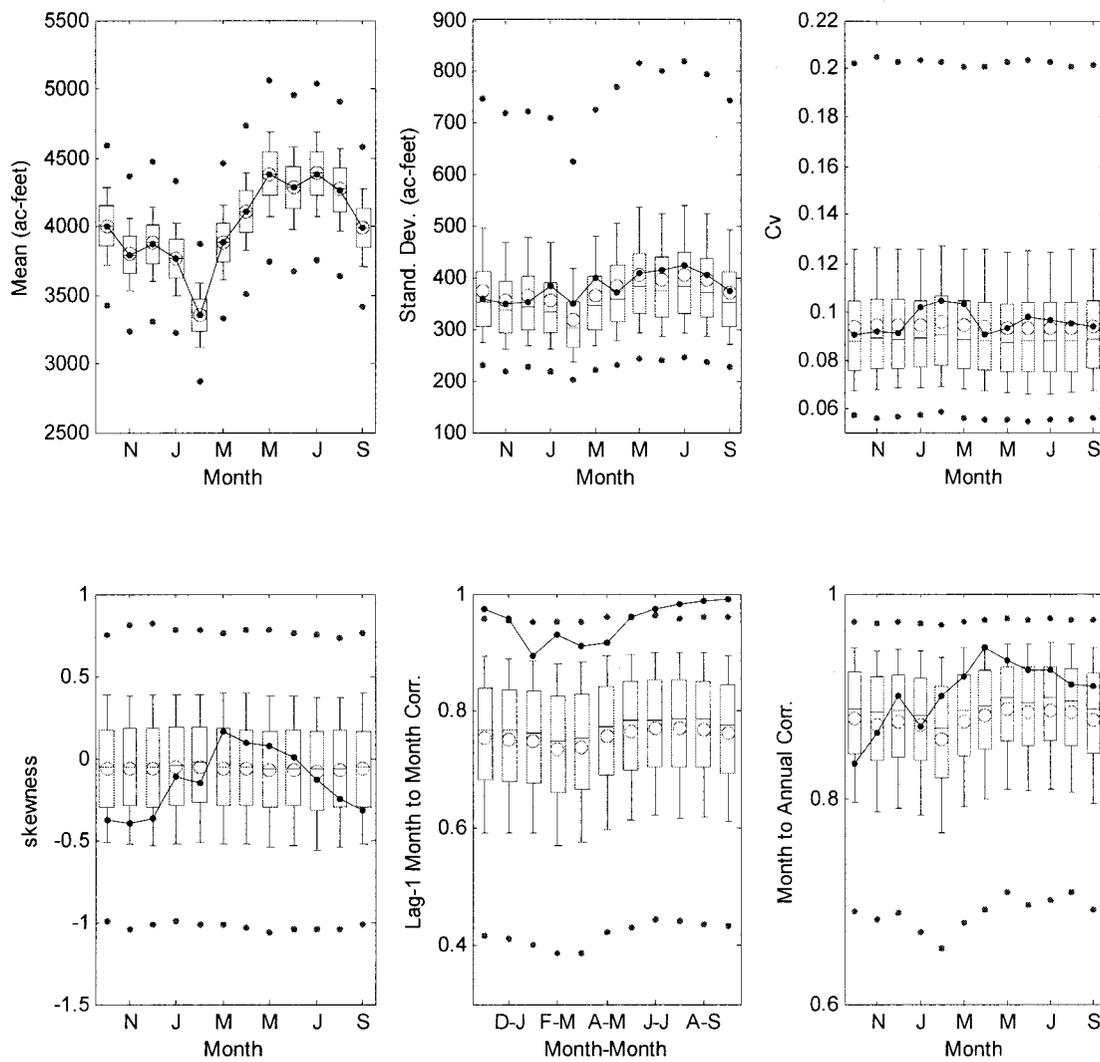


Figure 6.A7: Basic statistics of generated monthly streamflows (AuMu,  $n=25$ , St. Lawrence River)

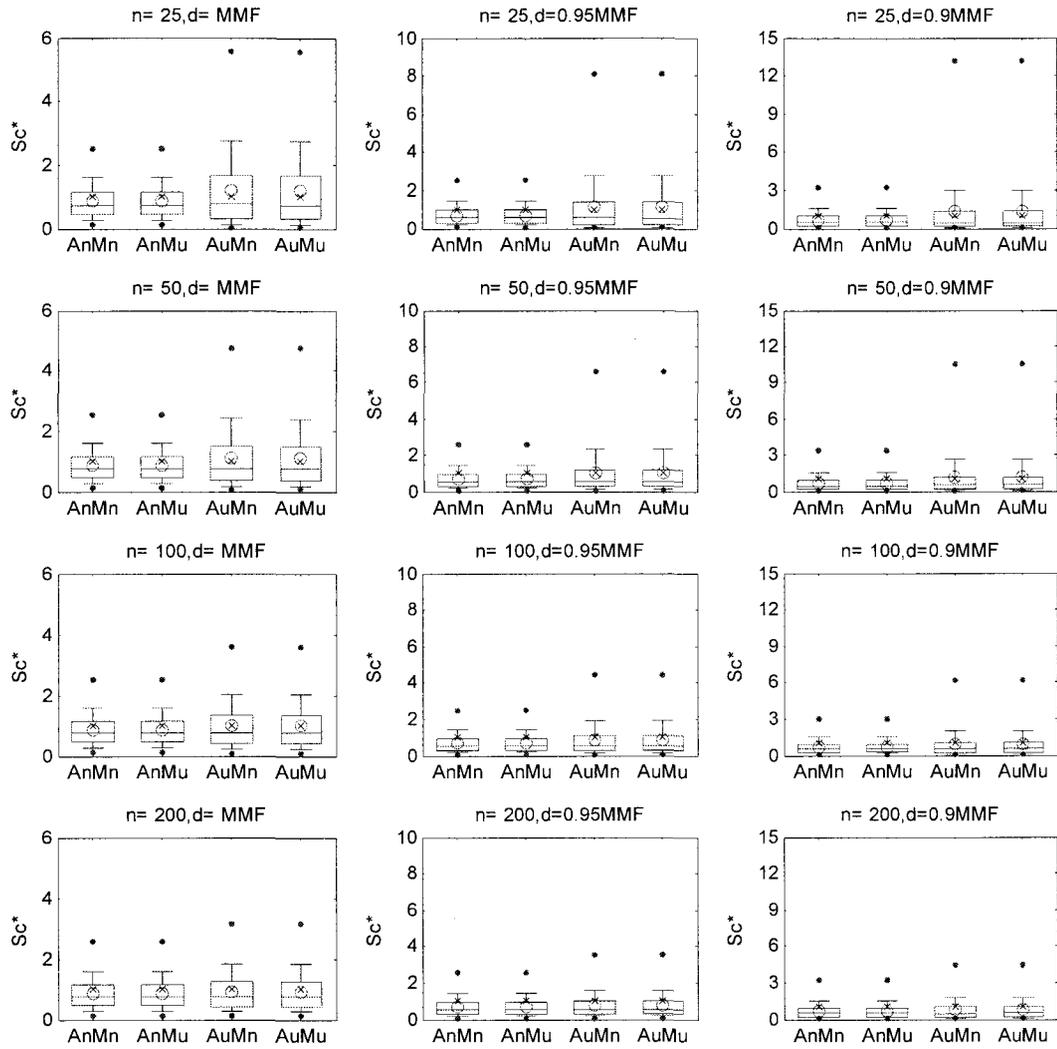


Figure 6.A8: Quantile estimates of generated storage capacities  $Sc^*$  (St. Lawrence River) where \* denotes scaled by historical storage capacity.

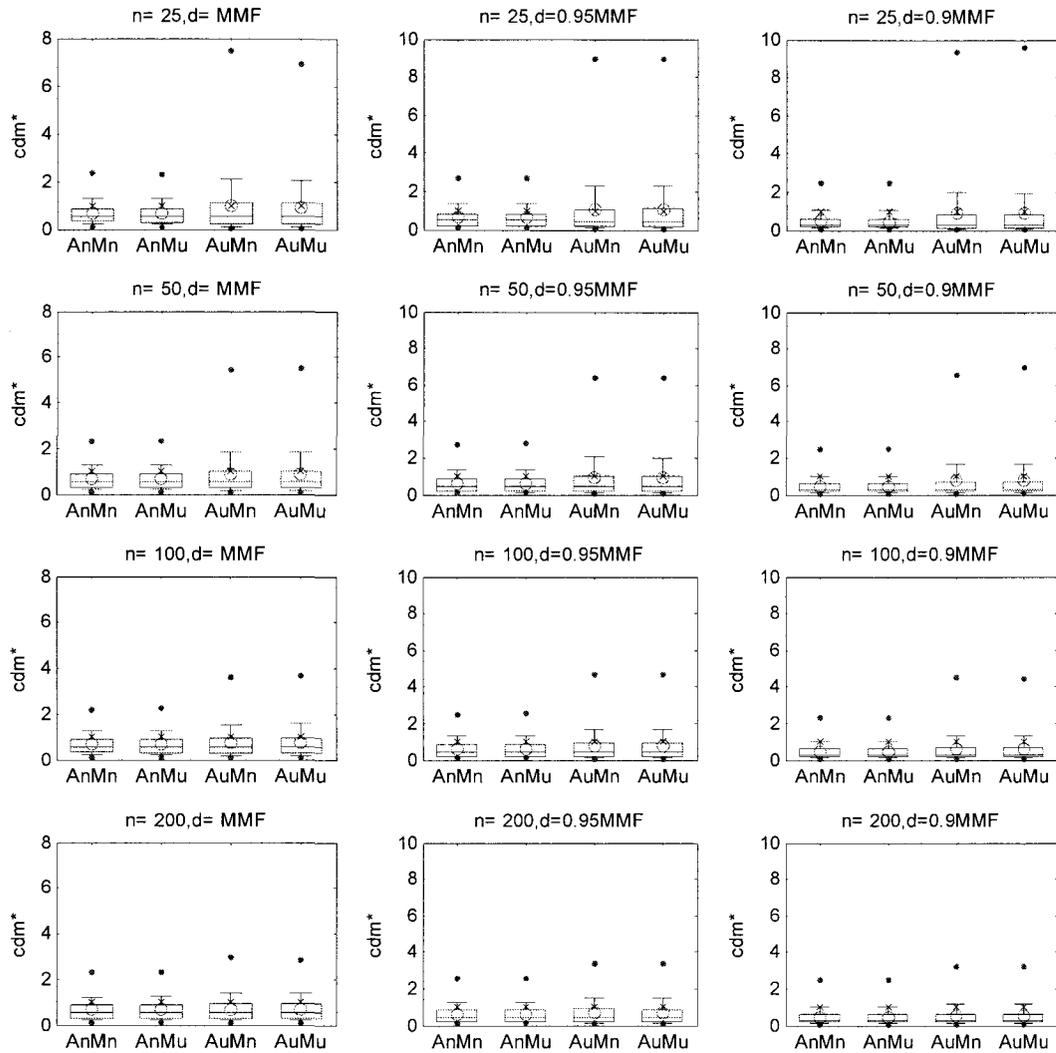


Figure 6.A9: Quantile estimates of generated critical drought magnitude  $cdm^*$  (St. Lawrence River) where \* denotes scaled by historical storage capacity.

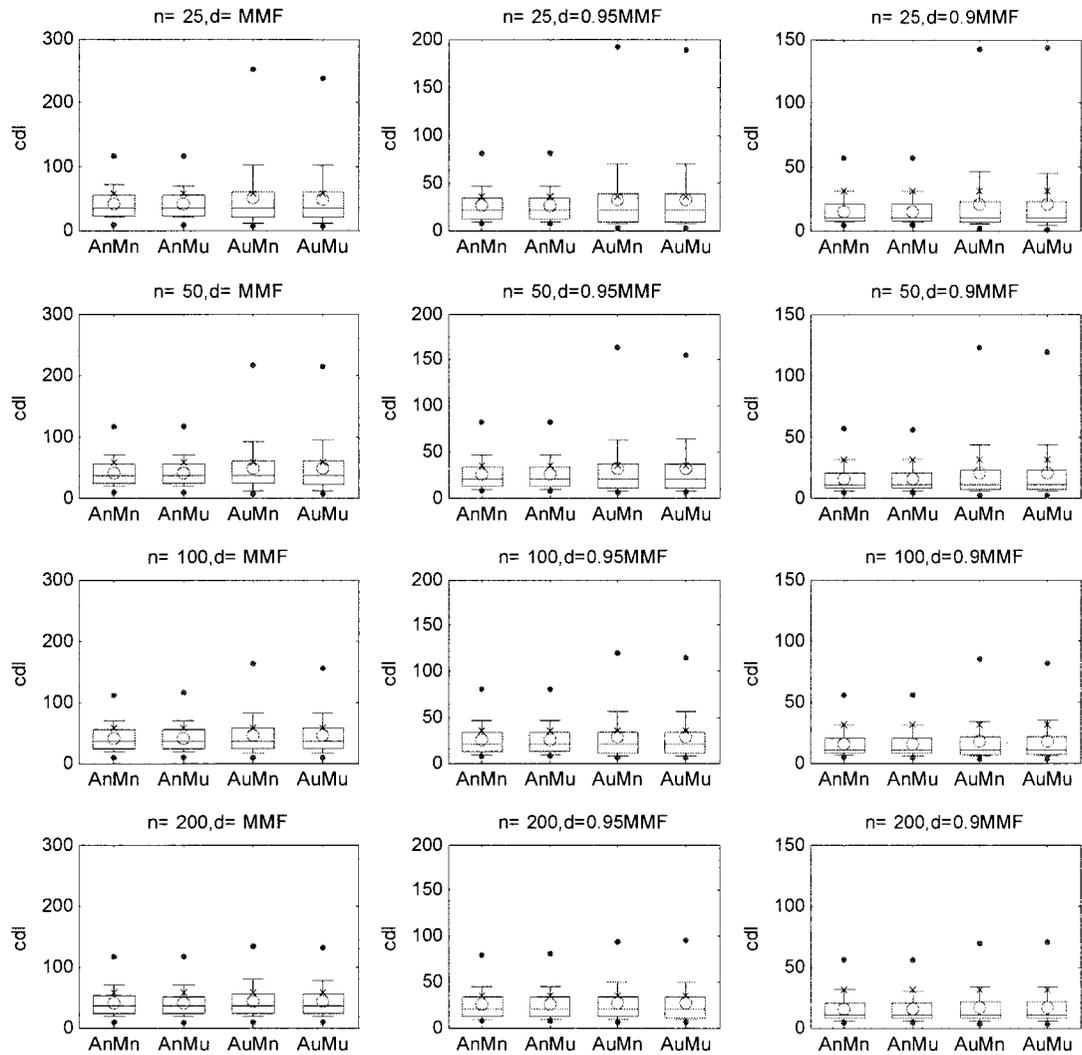


Figure 6.A10: Quantile estimates of generated critical drought magnitude *cdl* (St. Lawrence River).

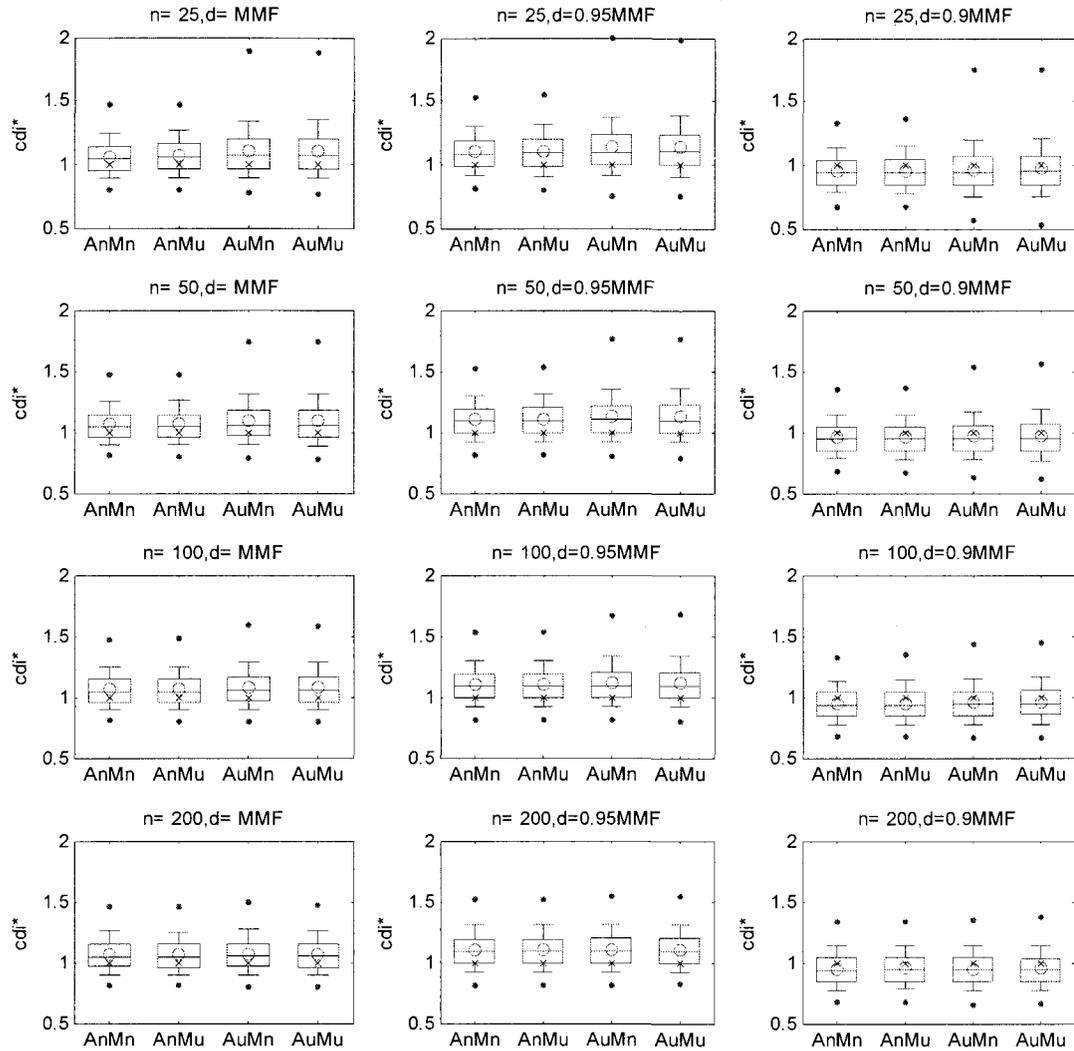


Figure 6.A11: Quantile estimates of generated critical drought magnitude  $cdi^*$  (St. Lawrence River) where \* denotes scaled by historical storage capacity.

## **Chapter VII**

### **CONCLUSIONS, CONTRIBUTIONS, AND RECOMMENDATIONS**

#### **7.1 Summary and Conclusions**

The limited sample size of historical flows causes the uncertainty of parameter estimates of the streamflow generation models, which consequently affects behaviors of the synthetic streamflows and related design variables required for the design and planning of real hydrologic systems. In this study, the impact of parameter uncertainty was taken into account based on different streamflow simulation techniques. For comprehensive remarks and conclusions for the different methods and procedures, refer to corresponding chapters. In summary, several conclusions obtained from this study are as follows:

- (1) Annual streamflow generation: Overall, the uncertainty effect of the mean parameter and the AR(1) parameter makes a significant impact on the generated streamflow statistics and related design variables. The uncertain mean

parameter shows the significant effect on the generated mean and uncertain AR(1) parameter affects on the generated serial correlation. Two different demand level options, such as fixed mean (*FM*) and simulated mean (*SM*) were employed for calculating design variables where the uncertainty of the mean parameter and the AR(1) parameter was shown to be dominant, respectively. Distinguishable effect on design variables by uncertain parameters is still visible, even though the sample size is 100 in option *FM*. The parameter uncertainty effect is related with the different fraction of demand levels: the mean uncertainty effect becomes less significant, while the uncertainty of the AR(1) parameter becomes more significant for smaller demand levels for adopted application. This difference between asymptotic and Bayesian approaches is notable for all applied different sample sizes in the St. Lawrence River, which has a long term memory, but not for the case in the Colorado River Basin. When the sample size is greater than 100, asymptotic analysis might be applicable for incorporating parameter uncertainty into the generation of annual streamflows. Using a bootstrap as an alternative for incorporating the parameter uncertainty showed similar variability as natural uncertainty, thus the parameter uncertainty effect is difficult to evaluate with the bootstrap technique.

- (2) Temporal disaggregation model: It is notable that even for large sample size ( $n=100$ ), the parameter uncertainty could be found in increased variabilities of design variables. LAST and SPC show almost similar variabilities of generated storage and drought related statistics in terms of parameter uncertainty. The

Bayesian approach shows larger variabilities of design variables for a smaller sample size ( $n \leq 50$ ) than asymptotic distribution and the difference between two approaches is more visible for the St. Lawrence River case. Parameter uncertainty in the disaggregation model affects little on the mean and standard deviation of generated monthly flows but more on the month-to-month serial correlation and month-to-annual correlation. The effect of natural uncertainty in annual streamflow generation is more significant than parameter uncertainty in the temporal disaggregation model, which is closely associated with the model structure that the sum of monthly flows would be equal to the annual flow, and the proportionality of monthly flows would be preserved. Parameter uncertainty in the annual flow generation is propagated into simulated monthly streamflows through the disaggregation stage and generates significant effect on calculated storage and drought statistics by increasing their variability.

- (3) Spatial disaggregation model: Two approaches, asymptotic and Bayesian, were able to explain the variability resulted from parameter uncertainty similarly in most sample statistics, an exception being cross-correlations. A little wider variability and upward bias of generated storage and drought related statistics were reported when using the Bayesian distribution for small sample size. A sample size of at least 100 would be required if the asymptotic distribution is utilized for the parameter uncertainty consideration. That is, for the Bayesian approach, the effect of non-informative prior, which results in much variability of parameters than the exact prior, becomes negligible, and sample information

might be enough to interpret the variability of uncertain parameters. The effect of parameter uncertainty of the disaggregation procedure was not as significant on the generated means, standard deviations, and lag-1 serial correlations as the natural uncertainty of given input variables. Likewise, the effect of the parameter uncertainty of the disaggregation model on the utilized storage capacity and critical drought indices was also not as significant as the natural uncertainty of input variables. However, cross correlations of generated flows were much influenced by parameter uncertainty of the disaggregation. If any statistics, which are based on cross correlations, are to be required in the practical design problem, parameter uncertainty of the disaggregation model would become more significant. When the parameter uncertainty of input variables was incorporated into the disaggregation, increased variabilities of key-stations was shown to propagate into the generated sub-station flows, more significantly for smaller sample sizes. Thus, uncertain parameters regarding input variables might cause important consequences on the related determination of reservoir size as illustrated by increased variabilities of storage capacities and critical drought indices. In this case, parameter uncertainty of the disaggregation model results in the additional increase over those by parameter uncertainty regarding input variables. The parameter uncertainty still exhibits its effect on the variability of generated storage capacities and drought indices even when the sample size is equal to 100.

(4) Multivariate AR model: Compared to traditional generation without parameter

uncertainty incorporation, parameter uncertainty creates an increase in variability of quantile distributions of basic statistics and increases the variability of quantile distributions of design variables as well. Compared with the asymptotic approach, the Bayesian approach reproduces synthetic streamflows with bigger standard deviations, as well as associated design variables with upward shifted expected values and larger variability, which are more visible when the small sample size is equal or less than 100 and is of a smaller demand level. When the sample size approaches 200, the results from the Bayesian approach become identical to those from the asymptotic approach. A skewness of real data might be significantly effective on the variability of design variables generated from the multivariate model with parameter uncertainty incorporated. In this case, a proper elimination of skewness of the real data might be useful in removing any unexpected increases in the variability of generated design variables.

- (5) Proportional disaggregation model: A simple disaggregation model was developed based on Dirichlet distribution for the purpose of overcoming the discrepancy caused by the normality limitation in the application to real-space flows. The model structure is very simple, and only 12 parameters are required for the case of the annual-monthly temporal disaggregation when compared with the traditional Valencia-Schaake model (156 parameters) and even the condensed Lane's model (36 parameters). From the comparison of possible different parameter estimators, a maximum likelihood estimator shows the smallest variability of parameter estimates. The application to the real data sets

demonstrates that only mean statistics could be well preserved through the simulation, not for higher order sample moments. However, the PD model preserves the design variables well regardless of demand levels, which could be compared with traditional disaggregation models. Associated with the parameter uncertainty effect, the uncertain annual flows play the more significant role in the variability of generated monthly flows rather than the uncertainty of the PD disaggregation model. Unfortunately, it is supposed that this PD model could not show enough flexibility eligible to the generation of streamflows with the different statistical characteristics. That is, model verification would be needed in advance of the actual application. The uncertainty consideration of parameters could not resolve the flexibility of the model since the parameter uncertainty shows the tight dispersion range of parameter variability.

## **7.2 Contributions and recommendations**

In this study, a sample size of 100 was shown not to be large enough to justify the effect of parameter uncertainty. The parameter uncertainty effect remains visible when sample size is equal to 100, and even to a sample size of 200. However, the determination of what constitutes an adequate sample size depends on how much variability of design variables can be tolerated in a practical sense. In most applications for real streamflow generation, the available sample size is usually less than 100. The incorporation of parameter uncertainty issues into the streamflow simulation is of importance, and thus the precision and reliability of generated streamflows will be

improved.

Generally in the Bayesian framework, the posterior will be explained by only the sample structure rather than depending on its prior when sample sizes are large enough. Indeed, an applied posterior distribution of parameters that is derived from the non-informative prior has been shown to be equivalent to an asymptotic distribution when the sample size is larger than 100. On the other hand, different distributions of variance-covariance for error terms of univariate or multivariate stochastic models based on asymptotic and Bayesian approaches are available which result in an increased standard deviation of generated streamflows and increased variability in quantile distributions of design variables. Theoretically, more precise posterior distributions can be obtained if the proper prior of parameters are available, but this is not the case unless the exact prior is given. Thus, modified posterior distributions with bounded parameter spaces seem appropriate for incorporation of parameter uncertainty into the streamflows generation. Use of the modified posterior distribution would be preferred for the real hydrologic planning and design over asymptotic distribution since there would be significant variability generated by the Bayesian approach as it is applied to a small sample size. Such an approach to ensure the reliability of streamflow generation might not be negligible.

The uncertainty of parameters of temporal and spatial disaggregation models has been shown to not be as significant on the first and second moments of disaggregated flows as the uncertainty of input variables, i.e., parameter uncertainty regarding generated annual flow or key-site flow, which was shown to be closely related with the parameter uncertainty effect on design variables. However, it was notable that parameter

uncertainty of disaggregation models affects the month-to-month and month-to-annual correlations (temporal case) or cross correlations (spatial case) greater than parameter uncertainty of input variables. If the comprehensive consideration of those correlations is to be required more seriously in the synthetic streamflows based on disaggregation models, parameter uncertainty might result in much reliability.

Recommendations for further studies are:

- (1) The parameter uncertainty has been shown the variability of simulated design variables. Developing and applying the design criteria to tolerate this variability would be helpful in real design and planning of water resource systems.
- (2) In this study, theoretical distributions of parameters were employed to quantify the uncertainty and incorporate into the streamflow generation. Comparison with the numerical method such as one based on the Markov Chain Monte Carlo method, would give greater insight for investigating the parameter uncertainty in the hydrologic simulation.
- (3) This study showed that the parameter uncertainty effect could be closely related with the assumed transformation procedure (associated skewness effect of real data). More studies are required for the related uncertainty effect of transformation besides parameter uncertainty to improve the ability and applicability of streamflow generation.
- (4) The parameter uncertainty effect has been comprehensively examined throughout the different generation models, especially traditional disaggregation models. In spite of the good performance of conventional disaggregation models, adjustment

techniques are still important to relieve the normality assumption of applied streamflows in the model, which would distort the marginal distribution of real space flows. A proportional disaggregation model was suggested as an alternative in the last chapter, but the model flexibility was not satisfied because its short number of parameters was not enough to preserve statistics of interest. Expansion of that model to a more general case would be recommended.