

LONG-TERM SALINITY PREDICTION WITH UNCERTAINTY ANALYSIS:
APPLICATION FOR COLORADO RIVER ABOVE GLENWOOD SPRINGS, CO

by

JAMES ROGER PRAIRIE

B.S., State University of New York College of Environmental Science and Forestry,

1993

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Master of Science

Department of Civil, Architectural, and Environmental Engineering

2002

This thesis entitled:
Long-Term Salinity Prediction with Uncertainty Analysis: Application
for Colorado River Above Glenwood Springs, Colorado
written by James Roger Prairie
has been approved for the Department of
Civil, Architectural, and Environmental Engineering

Balaji Rajagopalan

Edith Zagona

Date _____

The final copy of this thesis has been examined by the
signators, and we find that both the content and the form
meet acceptable presentation standards of scholarly work in
the above mentioned discipline.

Prairie, James (M.S., Civil, Architectural, and Environmental Engineering)

Long-Term Salinity Prediction with Uncertainty Analysis: Application for Colorado River Above Glenwood Springs, Colorado

Thesis directed by Professor Balaji Rajagopalan

Understanding the variability of salinity in the Colorado River basin is key to monitoring the water quality and also to honoring Minute No. 242 of the International Boundary and Water Commission. To this end, a modeling framework is presented that is capable of capturing any type of observed variability in salinity, is flexible, is portable, and is easily implemented at the 29 gauges throughout the basin. The need for such a framework was motivated by the inadequacies of the existing CRSS model, which consistently overpredicts the salt mass. The framework consists of three pieces: (i) a stochastic streamflow generator, (ii) a statistical salt model to compute natural salt mass and (iii) a simulation module to estimate the historical salt mass.

The framework works as follows:

Ensembles of streamflows (natural flows) are generated from the streamflow generator and are then passed through the statistical salt model to estimate the natural salt. Finally, the ensembles are then passed through the simulation module to compute the historical salt mass and concentration.

The streamflow generator and the salt model are modified nonparametric techniques, K-nearest neighbor method. We offer comparisons to traditional alternatives and find marked improvement with the nonparametric methodology.

The utility of the model is demonstrated in generating salinity scenarios at Glenwood Springs on the Colorado River. Climate information (such as the state of ENSO etc.) relevant to streamflows can be easily incorporated into this framework, generating streamflows consistent with the climate state.

Dedication

For my parents, my friends, and myself.

Acknowledgments

I would like to acknowledge the support and advising from Terry Fulp, Balaji Rajagopalan, and Edie Zagona. Their encouragement and wisdom were always welcome. The U.S. Bureau of Reclamation and Dave Trueman for their support with funding and freedom to explore new ideas. The members of the Salinity Control Forum Technical Modeling Committee for their insights and expertise. The Center for Advanced Decision Support in Water and Environmental Systems for use of its facilities and software. I would also like to thank Annemarie Curran for her open heart and love through the completion of this work.

Contents

1 Problem Definition	1
Introduction	1
Motivation	1
Description of the Problem	3
Research Tasks	5
Case Study	6
Literature Review	8
Methodology	8
Water Quality Models with Uncertainty	8
Basin-Scale Water Quality Models for the Colorado River Basin	11
Modeling Salt Processes	14
Modeling Methods for Uncertainty	15
Background For the Colorado River Basin and Salinity Modeling	17
The River Basin	17
Salinity in the Basin	19
Laws of River Relating to Water Quality	20
Previous Modeling Efforts	22
Research Tasks Performed	24
Research Issues	24
Identifying Sources of Uncertainty and Determining How to Model Them	24
Representation of Stochastic Hydrology	27
Separation of Natural Flow from Gauged Flow	30
Method of Modeling Salinity Concentrations	32

Issues of Scale in Human-Induced Salt Loading	32
Overall Modeling Approach and Representation of Results	34
Model Integration.	35
Case Study - Colorado River	36
Interconnection for Modeling System	37
Model Validation and Metric of Success	38
Uncertainty in Predictions	40
Policy Modeling	40
Summary	41
2 Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow	43
Introduction	43
Background	45
Index Sequential Method.	46
Periodic Auto Regressive Method.	49
Traditional K Nearest Neighbor Method.	51
Modified K Nearest Neighbor Method	54
Model Evaluation.	57
Test Ensembles	57
Model Evaluation Criteria	59
Results	59
Climate Link	82
Discussion and Conclusions	87
3 Statistical Nonparametric Model for Natural Salt Estimation	90
Salinity Standards on the Colorado River	90
Salinity Sources and Remediation.	91
Previous Modeling Efforts.	93
Statistical Nonparametric Natural Salt Model.	97
Local Regression (Nonparametric Regression).	97
Residual resampling.	102
Model Evaluation.	104

Test Ensembles	104
Model Evaluation Criteria	105
Results	106
Discussion and Conclusions	113
4 Historic and Future Salt Concentration Modeled in RiverWare	114
Introduction	114
Existing CRSS	116
Historical Verification	116
Stochastic Planning Runs	119
Modified CRSS	121
Modified Stochastic Natural Flow Model	122
Statistical Nonparametric Natural Salt Model	127
Simulation Model	136
Simulated Policy	147
Comparison of the Existing CRSS and the Modified CRSS	151
Stochastic Planning Run Simulations	159
Discussion and Conclusions	164
5 Conclusion and Recommendations for Future Work	167
Conclusions	167
Modified Nonparametric Stochastic Flow Model	167
Statistical Nonparametric Natural Salt Model	169
CRSS Simulation Model	169
Future Work	170
References	172
Appendix A Glossary of Terms	179
Appendix B CRSS Inputs, Outputs, and Sources of Data	181
Identify Primary Inputs, Outputs, and the Sources of the Data for Colorado River Simulation System	181
Generating the Natural Flow Database	182

Appendix C Sample Splus Code	185
Sample Splus Modeling Code	185
Sample Splus Data Plotting Code	192

Tables

1.	Uncertainty in Water Quality	26
2.	Uncertainty in Water Quantity	26

Figures

1. Sources Contribution to Total Salinity (U.S. Department of Interior, 2001). . . . 7
2. Map of Colorado River basin. 18
3. Locations of Title II Salinity Reduction Projects and Studies (U.S. Department of Interior, 2001). 21
4. Illustration of Index Sequential Modeling adapted from Ouarda et al., 1997. . . 24
5. Detailed outline of the case study drainage area and stream gauge. 36
6. Flowchart depicting interconnection of modeling system. 39
7. Schematic of the ISM (adapted from Ouarda et al., 1997). The synthetic hydrologies, each 25 years in length, are shown below the original 90-year time series. The additional 24 years used for wraparound are shown in shading. 47
8. Nonlinear local regression fit to March natural flows dependent on February natural flows is depicted by the solid line ($\alpha = 0.3$). A least square fit is shown with the dotted line. 56
9. (a) A time series of annual water year natural flow and (b) monthly natural flow from water year 1906 to 1995 for USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). The time series exhibit a high rate of variability both annually and monthly. 58
10. The ISM technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistic with boxplots. Boxplots show the interquartile range of an ensemble in the box and the vertical dashed lines (“whiskers”) approximate the 5% and 95% range of the ensemble. The historic statistic for the fitting period (water year 1906 to 1995) is shown with the solid line and the solid circle. When the historic statistic falls in the box, the statistic is preserved by the technique. The upper graph shows the mean, while the lower graph shows the standard deviation. Both the month-

- ly and annual statistics are preserved. 61
11. The PAR technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistic with boxplots (see Figure 10 for explanation). The upper graph shows the mean, while the lower graph shows the standard deviation. Both the monthly and annual mean are preserved. The PAR model overestimates the standard deviation monthly and annually. 62
 12. The modified KNN technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistics with boxplots. (See Figure 10 for explanation.) The upper graph shows the mean, while the lower graph shows the standard deviation. Both the monthly and annual statistics are preserved. 63
 13. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the ISM. The lower graph shows the coefficient of skewness. ISM preserved both statistics annually and monthly.. . . . 65
 14. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the PAR model. The lower graph shows the coefficient of skewness. The lag(1) correlation is preserved monthly for most months in the interquantile range. The coefficient of skewness is not preserved in the interquantile range for either the monthly or annual time step. 66
 15. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the K-NN. The lower graph shows the coefficient of skewness. K-NN preserves both statistics monthly. The annual time step is preserved on the fringes of the interquantile range. 67
 16. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the ISM. The lower graph shows the minimum value. The ISM exactly reproduces both statistics. 68
 17. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the PAR(1). The lower graph shows the minimum value. The PAR(1) technique tends to overestimate the maximum and minimum both monthly and annually. 69
 18. The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the K-NN. The lower graph shows the minimum value. The K-NN technique slightly underestimates certain months maximum but preserves the other months and annual statistics well. 70
 19. The upper graph shows the PDFs for ISM for January, which approximates a Gaussian distribution. The ISM can only reproduce the historic probabil-

- ity density function because every simulation includes all the historic time series; however each simulation is sequentially shifted. The lower graph shows the PDFs for February, which is skewed towards the low flows. 72
20. The upper graph shows the PDFs for PAR(1)for January, which approximates a Gaussian distribution. The lower graph shows the PDFs for February, which is skewed towards the low flows. This parametric technique cannot preserve the skewed distribution in the interquantile range of the simulations. 73
21. The upper graph shows the PDFs for K-NN for January, which approximates a Gaussian distribution. The lower graph shows the PDFs for February, which is skewed towards the low flows. Both distributions are preserved by the K-NN technique. 74
22. The upper graph shows the PDFs for ISM for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The ISM exactly recreates the PDF, again as stated in Figure 19. 75
23. The upper graph shows the PDFs for PAR for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The PAR technique can only approximate the annual PDF. Annual flow is not directly modeled. 76
24. The upper graph shows the PDFs for K-NN for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The K-NN technique can preserve the skewness and tail behavior best, but can only approximate the annual PDF, although the K-NN technique can reproduce the frequency of higher flows better than the parametric technique. 77
25. Bivariate probability density function for May/June estimated using an S-plus library from (a) historic values, and a single realization from 100 simulations using the (b) PAR(1), and (c) K-NN model. The nonparametric technique can preserve the historic bivariate PDF's non-Gaussian features, while the parametric technique can only approximate a bell-shaped distribution. 79
26. This figure shows boxplots of storage, surplus, and drought statistics for the nonparametric ISM model's annual values. The drought statistics include LD (longest drought), MD (maximum drought). The surplus statistic include LS (longest surplus), MS (maximum surplus). All the values are exactly preserved because the entire historic record is resampled for each of the 100 simulations. 80

27. This figure shows boxplots of storage, surplus, and drought statistics for the parametric PAR(1) model's annual values (see Figure 26 for explanation). The parametric model preserves the drought statistics are preserved slightly better. The parametric PAR(1) technique cannot preserve the storage statistic within the whiskers. 81
28. This figure shows boxplots of storage, surplus, and drought statistics for the nonparametric K-NN models annual values (see Figure 26 for explanation). The model preserves the storage statistic within the whiskers. 82
29. Both plots are anomaly composites for the geopotential height at 500 mb. The upper plot displays the composite of low-flow years. The lower plot displays the composite of high-flow years. The low-flow years show strong high pressure over the western United States and strong low-pressure over the northern Pacific. 85
30. This figure shows an anomaly composite for high-minus-low-flow years at the geopotential height at 500 mb. The plot shows a strong negative difference over the western United States and positive difference of the northern Pacific. 86
31. Flowchart depicting interconnection of existing CRSS modeling system. 96
32. This figure shows several data fitting techniques for data generated by sine function with noise. The solid line is a local linear regression with $\alpha = 0.2$, calculated by the LOCFIT algorithm. The dashed rectangle depicts the neighbors chosen around point x_t to develop the local linear regression. A local least square linear regression on the points in the rectangle is indicated by a heavy solid line. 100
33. This figure shows nonparametric regression for natural salt dependent on natural flow. Alpha equals 0.95. 102
34. This figure shows nonparametric regression for natural salt dependent on natural flow. Alpha equals 0.95. 103
35. This figure shows the relationship for natural salt dependent on natural flow for water years 1941 to 1985. The dashed line shows the USGS salt model relationship. The solid circles show SNNSM, and the solid lines show the local linear regression. For April (upper graph) the USGS slightly underestimates the results taken directly from the natural flow and salt estimates. For June (lower graph) the USGS relationship predicts a higher salt mass compared to the SNNSM. Both LOCFIT's have an alpha of 0.95. 107
36. The open squares show the annual natural salt mass, summed from monthly natural salt, for the USGS salt relationship, and the dashed line shows a

least squares fit. The solid circles show the data to find the SNNSM relationship, and the solid line shows a LOCFIT through the data. The USGS salt model gives higher natural salt mass compared with the SNNSM. 108

- 37. The upper graph compares the time series for natural salt generated from the USGS relationship and the modified K-NN method with the historic natural flow from 1941 to 1995. The lower graph compares the time series for simulated historic salt using the natural salt generated from the USGS relationship and the modified K-NN method with the observed historic salt from 1941 to 1995. 110
- 38. The upper graph shows the annual natural salt mass time series from the modified K-NN method with 5 percent and 95 percent confidence intervals. The heavy solid line shows the calculated natural salt mass. The dot-dashed line shows the natural salt mass taken directly from the local regression. The 5 percent and 95 percent confidence shows the variation around the line direct from the regression that occurs from the residual resampling. The lower graph removes the calculated natural salt and adds the natural salt determined from the USGS relationship. It is evident that the USGS relationship determines the natural salt mass to be higher than the natural salt mass we determined with our technique. 112
- 39. Line diagram of the CRSS simulation model and data for historic verification. 118
- 40. Line diagram of the CRSS simulation model and data for stochastic planning runs. 121
- 41. The solid line shows the PDFs for the calculated natural flow from 1941 to 1995. The boxplots describe the interquantile range (IQR) and whiskers for approximately 5 percent and 95 percent from the PDFs of the 100 synthetic natural flow traces. The plot represents a single run of the KNN flow model. The run preserved the calculated natural flow in the interquantile range over all flows. 124
- 42. The upper graph shows the PDF for K-NN synthetic natural flow. The lower graph shows ISM synthetic natural flow. Both are based on calculated natural flow from 1906 to 1995. In both graphs the solid line represents the PDF for the calculated natural flow from 1941 to 1995, while the dotted line represents the calculated natural flow from 1906 to 1995. The boxplots represent the PDFs of 100 simulated K-NN and 90 simulated ISM synthetic natural flows. The K-NN model simulated a wider range of flows. 126
- 43. The solid line shows the PDFs for the calculated natural salt from 1941 to 1995. The boxplots describe the interquantile range (IQR) and whiskers for approximately 5 percent and 95 percent of the 100 synthetic natural salt

- traces. The upper graph uses a monthly statistical nonparametric salt model. The lower graph uses annual statistical nonparametric salt model, preserving the tail of the calculated natural salt PDF. 129
44. The upper graph shows the PDF for the modified CRSS synthetic natural salt, while the lower graph shows the existing CRSS synthetic natural salt mass. It is evident that the USGS regressions used in the existing CRSS overestimate the solid line that shows the calculated natural salt mass PDF. . 131
45. The upper graph shows the PDF for synthetic natural salt concentration from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the synthetic natural salt concentration from a single annual regression to calculate synthetic natural salt mass. Using the single annual regression allowed the simulation to preserve the calculated natural salt concentration (solid line) in the interquantile range. 133
46. The upper graph shows the PDF for the modified CRSS synthetic natural salt concentration. The lower graph shows the existing CRSS synthetic natural salt concentration. The existing CRSS overestimated the calculated natural salt concentration (solid line), while the modified CRSS slightly underestimated the calculated natural salt concentration. 135
47. The PDF for observed historic flow, solid line, with boxplots of the simulated historic flow from the modified CRSS. The modified CRSS simulations have a subdued peak and elongated tails compared to observed historic flows. 138
48. The upper graph shows the PDF for simulated historic salt mass from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the simulated historic salt mass from a single annual regression to calculate synthetic natural salt mass. Using the single annual regression allowed the simulation to preserve the observed data in the interquantile range. 139
49. The upper graph shows the PDF for simulated historic salt concentration from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the simulated historic salt concentration from a single annual regression to calculate synthetic natural salt mass. Neither preserved the observed data in the interquantile range. 141
50. The regression relationship between total annual depletion as a function of the calculated natural flow. The solid line shows a LOCFIT through the data points. There is wide scatter around the regression indicating a wide variance around the regression. 143
51. The upper graph shows the PDF for historic flow, while the lower graph

- shows the PDF for the historic salt concentration. Both graphs used the regression between total annual depletion as a function of calculated natural flow without resampling. Both historic flow and salt concentration are preserved better incorporating the regression. 144
52. The upper graph shows the PDF for historic flow, while the lower graph shows the PDF for the historic salt concentration. Both graphs used the regression between total annual depletion as a function of calculated natural flow with resampling. Both historic flow and salt concentration are preserved best incorporating the regression with the resampling. 146
53. The upper graph shows the number of times a theoretical salt mass target of 650,000 tons was exceeded. The lower graph shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated. 148
54. The upper graph shows the number of times a theoretical salt mass target of 650,000 tons was exceeded. The lower graph shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated. 150
55. This figure shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated. Incorporating the regression from total depletions as a function of calculated natural flow with residual resampling allowed the simulations to preserve the observed number of violations. 151
56. The upper graph shows the PDFs for historic flow from the modified CRSS. The lower graph shows the PDFs for historic flow from the existing CRSS. 153
57. The upper graph shows the PDFs for historic salt mass from the modified CRSS. The lower graph shows the PDFs for historic salt mass from the existing CRSS. 154
58. Boxplots of the simulated historic salt mass from each of the 100 simulations. The modified CRSS generated lower salt mass and a wider range of values when compared to the existing CRSS results. The existing CRSS could preserve the median of the observed historic salt mass (solid circles) in the interquantile range. 155
59. The upper graph shows the PDFs for historic salt concentration from the modified CRSS. The lower graph shows the PDFs for historic salt concentration from the existing CRSS. 156
60. The upper graph shows the number of times each modeling system's results

- exceeded a tons standard of 650,000 tons. The lower graph shows the number of times the modified and existing CRSS results violated a salt standard of 350 mg/L. The solid circles show the number of times the observed historic record exceeded the standards. The boxplots show results of 100 simulations from the modified CRSS and 90 simulations from the existing CRSS. 158
61. The upper graph shows the PDF created from the median of boxplots of probability at a given projected future flow generated from the modified and existing CRSS. The lower graph shows the projected future salt mass from the modified and existing CRSS. The shift in the projected future salt mass PDF shows that the existing CRSS generated greater salt mass. 160
62. The upper graph shows the boxplots of the median from projected future salt mass. The modified CRSS generated lower salt mass and a wider range of values than the existing CRSS. The lower graph shows the PDF created from the median of boxplots of probability at a given projected future salt concentration generated from the existing and modified CRSS. The slight shift in the PDFs shows that the existing CRSS generated higher salt concentration. 162
63. The upper graph shows the number of times the modified and existing CRSS results exceeded a tons standard of 750,000 tons, while the lower graph shows number of times a salt standard of 650mg/L was exceeded and, therefore, violated during water year 2002 to 2062. For both graphs the existing CRSS results exceeded the modified CRSS results. 163

Chapter 1

Problem Definition

1.1 Introduction

1.1.1 Motivation

Rivers in arid or semi-arid climates that provide significant amounts of irrigation water tend to have salinity problems. The problems arise when large portions of the average flow are diverted for irrigation. Salts enter a river when surface and groundwater flows mobilize the salts present in soils, rocks, and geologic formations underlying the soils. Diversion for irrigation and return flows to the river increase the surface and groundwater flows, mobilizing additional salts. We use the term *salt loading* to refer to the introduction of salt from natural (precipitation) and human-induced (irrigation) sources. Additionally, evaporation and evapotranspiration typically remove water but not salt from the river basin, effectively increasing the salt concentration in the remaining water. Reservoir storage results in increased evaporation from a river basin. Diversions remove water and salt from a river for irrigation. Crop evapotranspire part of the water applied to them, but the salts are, for the most part, conserved. To prevent deterioration of farm land resulting from salt buildup, the remaining flow from the diversion is used to flush the salt deposited from the diversion in return flow to the river. The return flow will have a higher salinity concentration because only water, not salt, is removed during irrigation.

As water flows downstream, the water quality generally deteriorates because diversions remove water and return less water with higher salinity concentrations than naturally occur in the river. As water passes through reservoirs, evaporation increases the overall salinity of the outflow from the reservoir. To provide downstream users with water of adequate quality for irrigation and other consumptive use, costly projects must be implemented to maintain salinity concentrations at acceptable levels for downstream use.

Predictive models help basin planners to do the following:

- predict the effects of natural hydrologic variability on future salinity levels,
- predict the effects of future land use changes on future salinity levels,
- predict the effects of water quality improvements plans on future salinity levels (both to evaluate benefit/cost of proposed plans and to predict beneficial effects of completed plans),
- assess the likelihood of exceeding standards,
- analyze how operating policies and/or water allocation may need to change to mitigate effects of exceeding standards, especially in critical low-flow periods.

Many water quality modeling projects have not included confidence levels (risks, uncertainties). Recently, it has been recognized that uncertainty analysis is an important part of a model output because it quantifies how well models represent observed behavior and if the theory included in models provides meaningful and useful results (Beck, 1987). Incorporating uncertainty analysis in a model presents many advantages not available in deterministic models. Advantages of uncertainty modeling include the following:

- understanding and quantifying uncertainty of input variables,

- understanding and quantifying uncertainty in model parameters,
- determining the *confidence* of the predictions of future salinity levels,
- allowing policy makers to identify risks and uncertainties in policy analysis,
- targeting future research to reduce highly uncertain data in the model.

No basin-scale models of water quality and salinity exist that can meet these requirements, and this research intends to address that problem. We develop a methodology that meets these requirements.

1.1.2 Description of the Problem

We need to address several technical issues to provide a predictive long-term, basin-scale water quality model that can meet the requirements described previously. We must identify all sources of uncertainty in the basin model and quantify them in relation to each other.

Uncertainty can generally be categorized as either *random* or *knowledge* (Tung, 1996). Random uncertainty is due to the inherent unpredictability of an event. For example, hydrologic variability is a random event that will always be present and cannot be reduced or eliminated. Knowledge uncertainty is due to a lack of complete understanding about modeled algorithms, parameters, and data uncertainties. Additional sources of uncertainty include uncertainty in model processes for flow and salinity and error in measured values. Knowledge uncertainty can be reduced with improved data collection or additional research to further understand model processes.

This research on the Colorado River has identified hydrologic variability as the largest source of uncertainty in a long-term, basin-scale model in semi-arid climates. It is critical to incorporate the effects of hydrologic variability in such a model because hydrologic variability will be key to how well we predict the system response.

In long-term planning models, the historic record is typically used to generate

stochastic hydrology. But historic gauged flows include the effects of human development, such as reservoir storage and diversions. To use the historic record to generate stochastic hydrology for future predictions, we must separate the changes in flow resulting from human development from changes in flow resulting from natural variability (stochastic hydrology).

Likewise, we can use the historic salinity record to predict future salinity. Again, the historic salinity represents both natural and human-induced sources of salts. Separating the gauged salinity into natural and human-induced sources is more challenging than determining the natural flow from the gauged flow because salt loading is not directly measured, unlike diversions and reservoir regulations. Hence, natural and human-induced salt loadings estimates are more uncertain than flow estimates. Appendix A, "Glossary of Terms," provides definitions for variables developed from the separation of historic flows and the results of the modeling, which is described in this document.

When developing a long-term basin planning model, model developers typically make simplifying assumptions to reduce the complexity of the physical processes to be consistent with the long time frame and large computational time steps. In this study, we assume that salt is a conservative substance, and reservoirs and reaches between gauging points are completely mixed systems.

A highly detailed model results in highly uncertain results because the model requires many input data, and each piece of data has an associated uncertainty. To find an acceptable certainty for a model, scales for the aggregation and disaggregation of measured data must be developed. Scale relates to how data is represented in the model. To predict land use changes in the future, diversion and salt loading data must be spatially aggregated to a scale that reflects changing land use. Disaggregating the data at too small a level may lead to an unacceptably large uncertainty.

We researched two of the most common methods for modeling uncertainty: (1)

first order second moment (FOSM) and (2) Monte Carlo simulation. Long-term, basin-scale models using large data sets are highly uncertain and nonlinear. FOSM modeling is typically not a good approach for this type of uncertainty analysis, whereas Monte Carlo simulation has been successfully used in such applications. With the computing power available today, performing repeated simulations is not as pressing an issue as in the past. For these reasons, using the Monte Carlo simulation for uncertainty analysis is preferred for long-term, basin-scale modeling.

1.1.3 Research Tasks

This research performed the following tasks as ordered:

- 1) Analyzed the research issues described previously by researching previous work and modeling efforts with a literature review.
- 2) Selected methodologies reflecting mutual compatibilities of modeling approaches and uncertainties that are appropriate to the long-term, basin-scale predictions. Integrated methodologies in a simulation model.
- 3) Applied the selected and developed methodologies to a case study of the Colorado River basin. Developed a modeling system specific to the data available in the basin. The case study includes the following:
 - a) validation of the predictive modeling system with historical data;
 - b) application of the modeling system to make predictions from now until 2062 for flow and salinity levels with uncertainty;
 - c) discussion of operational policy alternatives in the event that critical low flows cause a sharp increase in salinity concentrations in the lower Colorado River basin;
 - d) demonstration of policy modeling by simulating salt mass and salt concentration violation statistics with uncertainty.

1.1.4 Case Study

The Colorado River basin has a semi-arid climate; it receives the lowest average precipitation, 1.15 in/yr (National Research Council, 1968), of any major river basin in the United States. One of the primary uses of Colorado River water is for irrigated agriculture. Salinity levels in the basin must be maintained at a level that will not adversely affect irrigated crops. Salinity standards, including numeric salinity criteria are required at key points in the Lower Basin by the EPA and must also be maintained in the future. A treaty between the United States and Mexico ensures that the salinity of the Colorado River water received by Mexico is not $115 \text{ ppm} \pm 30 \text{ ppm}$ above the annual salinity at Imperial Dam. These pressures make salinity the most important water quality parameter in the Colorado River basin. Other basins have similar characteristics; for example, the Rio Grande Basin has a semi-arid climate and high irrigation demands. The findings in this research could be applied to these basins in the future.

It is estimated that nearly half of the salinity present in the Colorado River is from natural sources; the remaining half is from human-induced sources that include irrigated agriculture, reservoir evaporation, and municipal and industrial sources (U.S. Department of Interior, 2001). Figure 1 shows the sources of salinity. Irrigated agriculture is the highest contributor of human-induced salts in the Colorado River. The U.S. Bureau of Reclamation (USBR) oversees extensive efforts costing millions of dollars per year to reduce the human-induced salt entering the Colorado River.

To better understand how uses and projects upstream in the system will impact the salinity downstream, models have been developed to predict impacts throughout the basin. Two previous modeling efforts used uncertainty analysis on a basin-scale for long-term policy analysis in the Colorado River basin (Malone et al., 1979 and Lee, 1989). Neither model incorporates the uncertainties in natural flow and salt together in the model. One model looks at the uncertainties in flow due to natural variability, and the other looks at the uncertainty in determining parameters used to model salt concen-

trations. Flow and salt are strongly dependent on each other; therefore, incorporating uncertainties for both will provide a broader view of the uncertainties inherent in a model and how the individual uncertainties together impact the model's predictions.

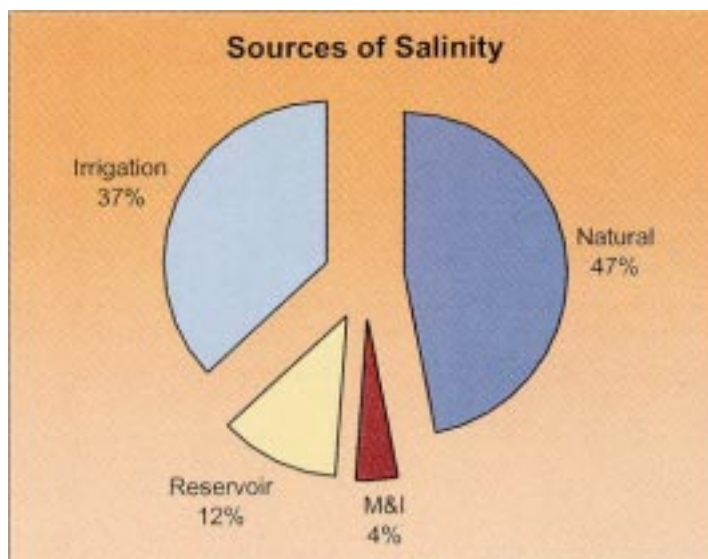


Figure 1: Sources Contribution to Total Salinity (U.S. Department of Interior, 2001).

The USBR developed a highly detailed modeling system of the Colorado River basin called the Colorado River Simulation System (CRSS). This modeling system allows more flexibility than past efforts and includes the operational rules by which USBR regulates its reservoirs. Data for the model is gathered from many sources, including the USBR, U.S. Geological Society (USGS), Department of Agriculture, and U.S. Census Bureau. Data is retrieved from gauging stations and extensive research studies. Using extensive data, the USBR intends to be able to produce more accurate model results. An outline for CRSS input, output, and sources of data are provided in Appendix B, "CRSS Inputs, Outputs, and Sources of Data." When extensive databases such as those used for CRSS are incorporated in models, those using the results of the model may easily conclude that the use of extensive data will provide a more accurate model. This conclusion is not always true, and caution must be used if the uncertainty of the data used in the model is not understood (Beck, 1987).

To prevent this false confidence in a model's results, an uncertainty analysis can be implemented to better convey how well the input data and methodologies in the model allow the model to predict the response of the system. The existing CRSS uses Index Sequential Modeling (ISM) to represent uncertainties due to stochastic hydrology. However, this technique constrains magnitude and duration of events to those of the historical record.

Currently, the existing CRSS uses relationships developed by the USGS to estimate natural salt associated with stochastic hydrology. These relationships do not incorporate the uncertainty of the relationship between natural flow and natural salt. The modified CRSS presented in this research directly addresses both these issues.

1.2 Literature Review

1.2.1 Methodology

We performed a literature search to locate articles pertaining to the general topic of long-term, basin-scale, one-dimensional surface water quality models.

Many articles are related to this research, including salinity modeling at a sub-basin scale (i.e., a field scale) and basin-scale modeling, but they include groundwater interaction. For this research area, sufficient data for the inclusion of basin-scale groundwater modeling is not readily available and would require a significant effort to include in a model. A discussion of papers relevant to our research follows.

1.2.2 Water Quality Models with Uncertainty

As stated previously, the primary research objective is to develop a long-term, predictive water quality model with uncertainties. For variable time series data, such as natural flow or natural salt loading, using stochastic modeling can allow natural variability of the data to be incorporated into a simulation model.

Two articles discuss uncertainty analysis, in general, in water quality modeling. Neither review discusses specific models nor model scales. They only provide a

view of the problems of water quality modeling, uncertainty, and the means to apply uncertainty in general water quality modeling. Following is a brief summary of each article and a statement of their significance to this research.

Beck (1987) discusses the use of models to develop scientific theories describing the behavior of complex environmental systems. Beck emphasizes the need to consider uncertainty in models and not limit oneself to deterministic models. Those using models must also understand that complex models describing the past in great detail may not be able to provide predictions in the same detail. The model's output must convey uncertainties in the prediction model.

He identifies four problem areas in water quality modeling associated with uncertainty: (1) uncertainty about model structure, (2) modeling parameters, (3) future system behavior, and (4) designs of experiments or monitoring programs intended to reduce the uncertainty associated with any of the previous areas.

A predominant focus of the review is model identifiability. Beck emphasizes the importance of using the appropriate model for your system. He discusses techniques for model identification, including the Hornberger-Spear-Young (HSY) algorithm and recursive parameter estimation. The HSY algorithm chooses model parameters that produce acceptable model results. A description using recursive parameter estimation to develop state variables explains one means to develop state variables and/or model parameters.

The article emphasizes the importance of uncertainty analysis when modeling large, data intensive, systems. Simplifications are needed to model the complexities at a basin-scale for planning time frames. Uncertainties must be quantified to understand the accuracies of the prediction. Beck supports the need for models such as the one we propose to develop and explains pitfalls that should be avoided during model development. As we develop methodologies for the modified model, Beck's discussion of model identifiability will be important. Uncertainty analysis should uncover the uncer-

tainties present in the model resulting from the simplification of salt processes and flow systems. The modified CRSS will incorporate the best available information from the extensive research of the basin. The modified CRSS will have the flexibility to be improved easily as new research advances theories that will be important to the basin-scale modeling system.

Tung (1996) reviews the application of uncertainty analysis in water quality modeling. Tung identifies two types of uncertainties: (1) uncertainty due to inherent randomness of an event and (2) uncertainties associated with a lack of complete knowledge about model processes, parameters, and data uncertainties.

Natural variability is a type of uncertainty that is always present in a natural system, and it cannot be reduced or eliminated. This is termed *random* (stochastic) uncertainty. Uncertainty due to a lack of complete understanding of the accuracy of inputs to the model or the processes included in the model is termed *knowledge* uncertainty. Knowledge uncertainty can be reduced through increased sampling or by improving the measurement of input variables for the model. Research can improve the mathematical representation of physical processes used in the model as the processes are understood in greater detail.

Tung identifies three measurements of uncertainty: (1) the probability density function of the variable or parameter subject to uncertainty, (2) a reliability domain (i.e., confidence interval), and (3) a mean and variance for the variable or parameter of interest. With the confidence interval, multiple variables or parameters cannot be combined to find the overall confidence interval for the system. When using a mean and variance for the variable or parameter of interest, the mean and variance for multiple variables can be combined to determine the overall system's mean or variance.

Tung's discussion of uncertainty analysis techniques includes analytical and approximation techniques. Analytical techniques are usually only applicable to fairly simple models because a complete probability density function (PDF) must be avail-

able for the input variables. Approximation techniques are more applicable to complicated engineering problems in which the exact PDF is difficult or impossible to determine. Approximation techniques include first-order variance estimation, first-order second moment, probabilistic point estimation methods, and Monte-Carlo simulation. Tung briefly describes the interrelationship between sensitivity analysis and uncertainty analysis and states “the former is used to analyze the internal mathematical responses of model outputs as affected by changes in model inputs/parameters, whereas the latter is used to analyze the stochasticity of the model through these relationships.” Both are important to uncertainty analysis.

Tung’s article provides a basic guide for types of uncertainty that can be modeled, methods to incorporate uncertainty in a model, means to propagate uncertainty to a model’s results, and a discussion of uncertainty analysis versus sensitivity analysis. This research considers the many alternatives presented when incorporating uncertainty in the modified CRSS. Tung provides a broad view of uncertainty in water quality analysis and provides an overview of how uncertainty may be applied in the modified CRSS.

1.2.3 Basin-Scale Water Quality Models for the Colorado River Basin

Two articles present uncertainty models of the Colorado River basin and provide a point for this research to move forward from published work. The articles describe water quality models that look at the Colorado River basin and predict salinity, including a measurement of the uncertainty of the prediction. A brief summary of each article is followed by a discussion of how this research will build on these models.

The Utah Water Research Laboratory (Malone et al., 1979) developed a basin-scale, stochastic steady state model for water quality throughout the Colorado River basin using an existing model, SALT. The SALT model incorporates stochastic analysis by allowing in-stream salinity and the agricultural base leaching factor to be input

as random variables. The agricultural base leaching factor represents an empirical value depicting the tons of salt removed from the soil per acre-foot of water flow through the soil matrix. Typically, the water flow represents return flows from irrigation. SALT determines the expected value and the variance for the natural salt load. The program SALTEZ, developed from the SALT program, additionally models means, variances, and skewness resulting from stochastic inputs of salinity and the agricultural base leaching factor. These additions allow the propagation of uncertainty from multiple model inputs using first order second moment analysis through the model to the model outputs.

This research considers uncertainties only with salt loadings and the base leaching factor and not with stochastic hydrology. The salt loading term in the model is the mean of the historic loadings from 1940 to 1974. Uncertainties in the mean value are related to natural and human-induced variability in flow and salinity over the time period. The uncertainty also includes measurement and calculation errors because the natural variability could not be separated from measurement and calculation errors. Uncertainty associated with the base leaching factor includes both modeling uncertainty (from the inexactitude of the mathematical representation of a physical phenomena) and uncertainty in the estimation of the base leaching factor from available data. The deviation from a best fit regression curve represents uncertainties, with basin leaching factors determined from field studies in a previous report (URWL, 1975).

Results from the uncertainty analysis show the variance due to the change in total gauged salt load contributes the most to the uncertainty, followed by the variance due to the basin leaching factor, then uncertainty in the total salt loading due to changes in projected diversions. The variance for the change in total salt loading remains constant for all simulations, while the variance due to the estimation of the basin leaching factor and projected diversions varies in each model run, depending on the agricultural parameters for the run. The standard deviation for the total salt load at

Imperial Dam varies from 6 to 10 percent of the total. This model does not consider the uncertainty due to natural variability of the flow regime, which would strongly influence the model's output uncertainty.

Lee et al. (1993) developed a model primarily for economic policy analysis that considers natural flow variability in determining of salt concentration. Lee developed a set of differential equations to describe the flow of total salts in the Colorado River basin. Again, a steady state model is applied to simplify the number of equations required for the model. Based on the Quality of Water Progress Report 18 (USBR, 1997) the equations simplify the system to include only 64 percent of the 981,570 tons of salt targeted by the Colorado River basin Salinity Control Act. The model is used to estimate the probability distribution of water quality improvement resulting from specific reductions in salt load or improvements due to different return flow salinity concentrations. Lee also models replacement of return flows from diversion with better quality water. Replacing the return flows dilutes the salinity in the river, because the replacement water is of better quality than the water in the river. Lee does not consider uncertainties associated with salt load.

Shortcomings of the simplified model are that it models only 64 percent of the salt targeted by the Salinity Control Act and it does not consider water entitlements in the final results. The final modified CRSS will account for all the salt in the system, and it will include water entitlements through the use of rules that represent the entitlements in the RiverWare modeling environment.

Neither the Lee nor the UWRL model incorporate the uncertainties in natural flow and salt together. Each modeled only one of the two uncertainties. This research incorporates uncertainties for both flow and salt in the modified CRSS. Because flow and salt are strongly dependent on each other, incorporating uncertainties for both will provide a broader view of the uncertainties inherent in the system and how the individual uncertainties impact the modeling system's predictions.

1.2.4 Modeling Salt Processes

The literature reviewed discusses various methods to determine the salt addition or removal from sites. Riley and Jurinak (1979) proposed a concept to explain salt production in a natural watershed. They use the term *baseline salinity* to represent the natural release of salt from a watershed basin due to hydro-geochemical weathering within a basin. Data shows that the salt mass from a basin is relatively constant, which means the natural baseline salinity is relatively constant.

Riley proposes two assumptions to develop his methodology. First, he uses measured data to show that, generally, the amount of salt leached from the land is highest when it is first irrigated; it decreases as irrigation continues to what he terms an *agricultural base salinity*, which is constant. He proposes that this concept can be applied to any basin once the land has been irrigated for many years, but does not provide an exact time frame. Second, once the agricultural base salinity is reached, the salt loading is due only to a combination of the base weathering rate of the soil profile and underlying geologic formation. He further proposes that when both of these assumptions are made, the removal of salt is directly proportional to the quantity of water passing through the soil profile. A relationship between irrigation efficiency and salt loading can be developed using the stated assumptions. Irrigation efficiency is a function of the water leaching through the soil profile.

Riley's research shows that soils have a base leaching factor that can be determined from readily available information. He develops a procedure based on mass balance to determine the effects of irrigation management on the salt input into a river. This procedure includes irrigation efficiency to better understand the effects of irrigation practices. The procedure determines the salt load contribution from agriculture as a function of the change in salinity and flow over the reach, evapotranspiration in the reach, and water diverted for agriculture in the reach. (A reach is a section of river where gauged or calculated data is available for flow and salinity at the upper and low-

ermost points in the reach.) Once the salinity contributed by agriculture is calculated, the salinity from natural sources is calculated as the change in salinity over a reach, minus the salinity caused by agriculture.

The model developed by Malone, discussed earlier, uses this method to separate natural and agricultural salinity. Malone attempts to use a mathematical technique to develop the natural and agricultural base leaching factors that Riley found by mass balance. When the factors are calculated using soils data for the basin, a regression reveals a weak relationship between the mass balance method and the mathematical technique. Malone determines the uncertainty of using this mathematical technique to describe salt loading from a soil profile, with the residuals from the regression. He states that other interpretations of Riley's theory will produce different uncertainties. One reason for the weak correlation can be the use of large-scale geology maps to develop the basin leaching factors for each subbasin. Much more detailed maps are required to more accurately represent the underlying rock and formations for each basin. Still, the classifications for the basin geology may not be at a scale that would provide adequate correlations.

This research identifies a method of modeling salt loading that is appropriate to the basin scale, yet yields acceptable levels of uncertainty. Iorns et al. (1965) also used a mass balance technique to determine natural and human-induced salt loads. These studies are important sources of data and methodologies for developing the salt processes for the modified modeling system.

1.2.5 Modeling Methods for Uncertainty

Additional articles discuss general approaches for modeling uncertainty of water quality parameters. These articles explain the application of uncertainty analysis by Monte Carlo simulation. Monte Carlo simulation has the advantage of providing useful information regarding the uncertainty of model outputs when a system is highly uncertain. Hession et al. (1996) use a two-phase uncertainty analysis to present a

determination of uncertainty for an ecological risk assessment. The two-phase uncertainty analysis determines knowledge uncertainty and stochastic uncertainty for the effect of excess phosphorous on the trophic state of a lake. The stochastic model parameters include phosphorous inputs and multiple parameters required by the model.

Hession separates knowledge and stochastic uncertainty so that the effects of each are shown individually. Stochastic uncertainty is due to inherent randomness of natural process, such as precipitation or temperature. This type of uncertainty cannot be reduced but should be recognized. Conversely, knowledge uncertainty can be reduced by further research regarding the parameter. Knowledge uncertainty can be used to guide future research and reduce the model's uncertainty in an effective manner.

Hession's model presents the output uncertainty in two phases. First, the model performs simulations varying only the stochastic uncertainty parameters. The results are shown in a complimentary cumulative density function (CCDF). This defines the probability of exceedance for various levels of phosphorous. Next, the model performs multiple simulations again, changing the knowledge uncertainty parameters before the multiple stochastic uncertainty runs. The multiple knowledge uncertainty runs produce multiple CCDFs. The range of the CCDFs show the knowledge uncertainty in the model.

Griffen (1995) uses Monte Carlo simulation to find the uncertainty associated with achieving a No Net Increase Policy for non-point source loading of nitrogen based on the uncertainty of nitrogen export coefficients and Best Management Plan effectiveness.

These models discuss methods to model uncertainty that are ideal for the Colorado River basin. Incorporating both random and knowledge uncertainty appears to be a possible technique to incorporate in one model natural variability of both flow and

salt, uncertainties in the measurement of input data, and the simplification of complex processes. Incorporating the uncertainty for both flow and salt in one model will be an important addition to the current ability to model salt process in river basins.

1.3 Background For the Colorado River Basin and Salinity Modeling

1.3.1 The River Basin

The research presented develops a long-term stochastic water quality modeling system for a subbasin of the Colorado River basin. In the Colorado River basin, salinity prediction and control modeling is necessary to meet treaty obligations between the United States and Mexico and to maintain national water quality standards required by legislation. The need to maintain salinity standards while concurrently allowing the development of unallocated water has made modeling salinity throughout the basin critical for long-term basin planning. The river drains an approximately 244,000-square-mile basin and is 1,440 miles long. The basin includes parts of Wyoming, Colorado, Utah, New Mexico, Arizona, Nevada, California, and a small part of Mexico. Figure 2 is a map outlining the Colorado River basin and the line that divides the basin into the upper and lower basins.

The Colorado River basin has a low average precipitation over the basin area compared to other basins throughout the United States. River basins such as the Columbia, Mississippi, or Delaware have much more precipitation over their basin area resulting in orders of magnitude greater flow out of the basin. Colorado River basin also has the added pressure of exporting large quantities of water out of the basin for neighboring areas that require additional water.

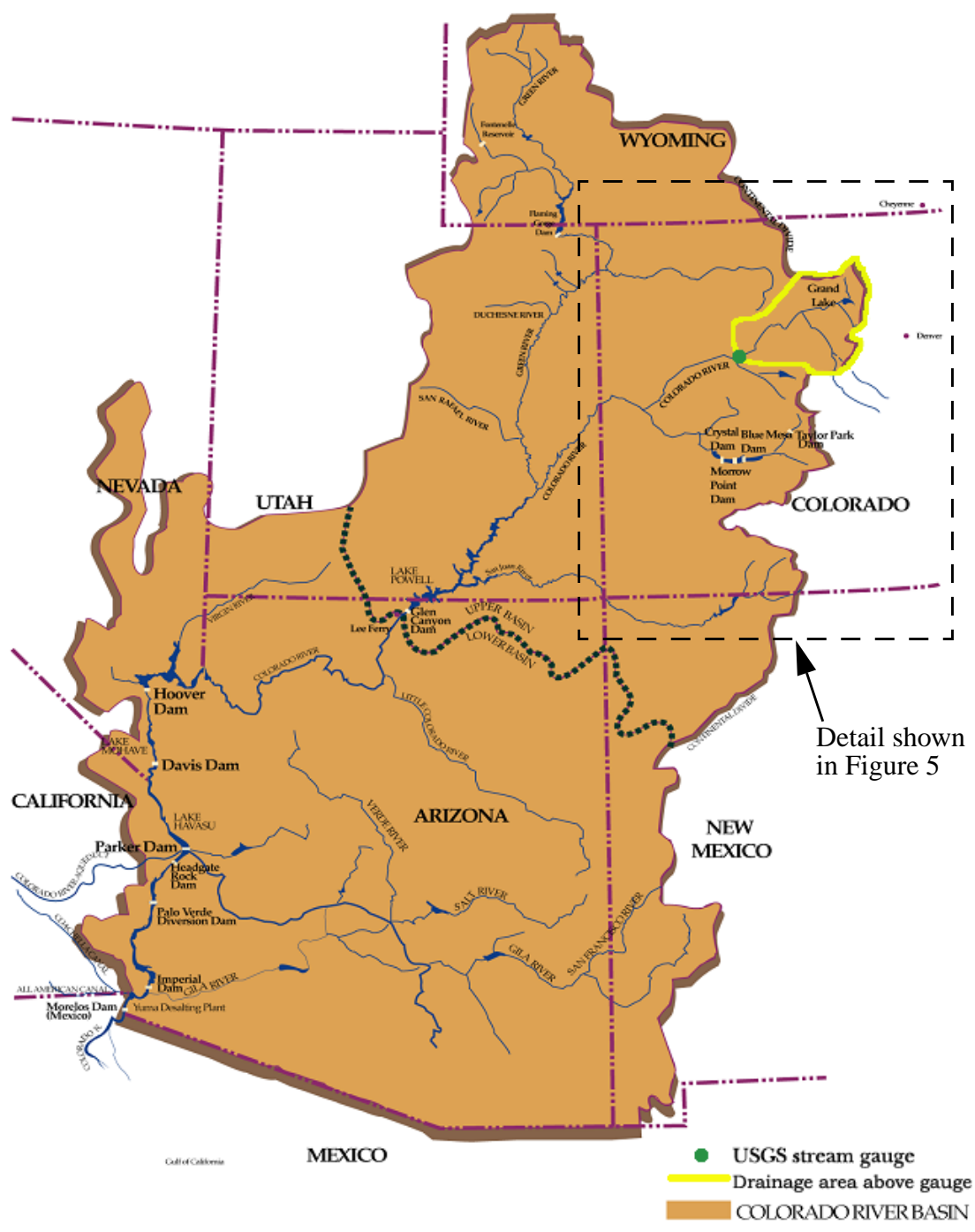


Figure 2: Map of Colorado River basin.

1.3.2 Salinity in the Basin

Natural and human-induced salinity result from point and non-point sources. Natural point sources that have been identified include seeps and saline springs. Some springs originate from deep geological formations containing brackish water. Natural, non-point sources of salinity generally originate from the weathering and dissolution of underlying rocks or soils overlaying the rocks.

The underlying rocks strongly characterize the natural salinity contributed by each subbasin. Areas of the basin include underlying rocks that are relatively resistant to dissolution by water. These rocks include igneous and metamorphic rock types. These rocks weather at a very slow rate and contribute small amounts of dissolved solids (total salts). The upper reaches of the Colorado, Green, and San Juan Rivers characterize these types of subbasins, which are typical in the high mountain areas. The middle and lower reaches of these subbasins contain sedimentary rock, which dissolves more readily in water and, therefore, contributes greater amounts of dissolved solids to the river. Soils in these regions have chemical dissolution characteristics similar to the underlying parent rocks. Therefore, soils over the sedimentary rock formations will tend to contribute greater amounts of dissolved solids. These areas characterize the valleys and lower lying plains, which are the predominant areas for agriculture.

Agriculture increases salinity concentration through two processes: (1) salt concentration and (2) salt loading. The salt concentration process is a result of evapotranspiration from crops, which results in the concentration of dissolved solids in the remaining return flow. Evapotranspiration from crops consumes water but leaves salts behind in the soil. Return flows to the river from the diversion typically contain the same salt mass present in the diversion water, but have less water, hence, higher concentrations of salt. Additionally, reservoirs concentrate salt during evaporation, i.e., water is lost from the reservoir and salt is conserved.

Salt loading occurs when salt is added (from deposits) or removed (by leaching) from salts present in the soil as water is transported through the soil. The water can be introduced into the soil from human-induced sources, such as irrigation practices, or from natural sources, such as precipitation. Irrigation practices increase the flow through soils, which increases the total salt loading from previous natural salt loading levels.

Federal agencies fund extensive efforts, costing millions of dollars per year, to reduce the human-induced and natural loading salt to the Colorado River. Figure 3 shows the extent of salinity projects and studies throughout the basin. USBR has off-farm projects designed to reduce irrigation return flows by lining irrigation canals and laterals. Additional projects not related to agriculture include capping saline springs from abandoned gas and oil exploratory wells. The U.S. Department of Agriculture has implemented on-farm salinity reduction, helping farmers to improve irrigation practices that increase irrigation efficiency, thereby reducing return flows. The Bureau of Land Management has implemented programs to reduce erosion on public lands to limit the release of dissolved solids. As of 1998, salinity control projects, with a reported total cost of approximately \$426 million, had removed an estimated 634,000 tons of salt from the river. An additional \$170 million is projected to remove an additional 390,000 tons in order to maintain current salinity standards (USBR, 1999).

1.3.3 Laws of River Relating to Water Quality

Since 1922, legislation, court decrees, and agreements have been established to form the operational policies of the Colorado River. These are documented in a single volume by Nathanson (1978) and are known as the “Law of the River.” These policies are concerned mainly with water quantity. Problems with salinity did not start to be discussed until 1961. At that time, the Mexican government strongly objected to the quality of the water the country was receiving. The average annual salinity of water delivered to Mexico in 1962 was 1,500 mg/L. This water was not suitable for irrigation

and was adversely affecting crops. The rise in salinity in the water delivered to Mexico was caused by two independent factors. Deliveries were reduced because water was being held back in Lake Mead in preparation for filling Lake Powell and, at the same time, highly saline groundwater from the Wellton Mohawk irrigation district was being pumped into the Colorado River above the Mexico boundary. These highly saline waters greatly deteriorated the quality of the Colorado River water.

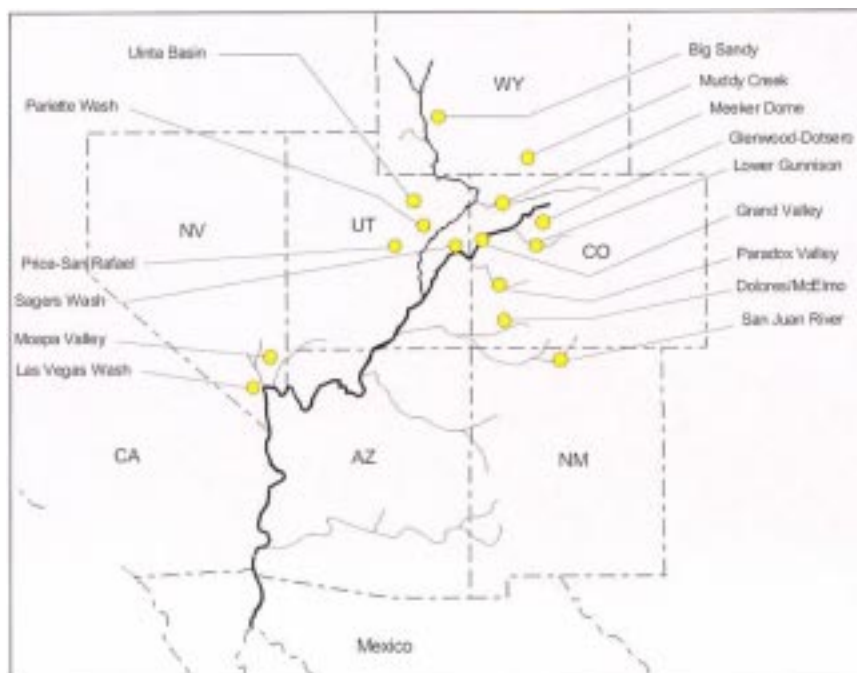


Figure 3: Locations of Title II Salinity Reduction Projects and Studies (U.S. Department of Interior, 2001).

In response to Mexico's concerns and after years of negotiations, Minute No. 242 of the International Boundary and Water Commission dated August 30, 1973, was signed. Minute No. 242 stipulated that water delivered to Mexico have an average salinity of no more than $115 \text{ ppm} \pm 30 \text{ ppm}$ above the annual salinity at Imperial Dam.

Subsequently, the Colorado River Basin Salinity Control Act of 1974 provided measures to ensure the United States could meet its obligation to Mexico under Minute No. 242. The Act authorized construction of a desalting plant and additional

salinity control projects.

Minute No. 242 sets a variable standard for the Mexico delivery, but does not set numerical water quality standards at any fixed points in the basin. Numerical standards resulted from separate U.S. legislation that set policy regarding water quality. The Federal Water Pollution Control Act Amendments of 1972 required development of water quality standards across the nation. The EPA interpreted the new legislation as a requirement to set fixed point numerical standards for salinity levels in the Colorado River basin. The Salinity Control Forum was founded by the basin states (Wyoming, Colorado, Nevada, Utah, Arizona, New Mexico, and California) to help develop the numerical salinity standards, including numeric salinity criteria required by the new legislation.

The following numeric salinity criteria were set in 1975 requiring maintenance of a flow-weighted average total dissolved solids concentration:

- 723 mg/L below Hoover Dam
- 747 mg/L below Parker Dam
- 879 mg/L at Imperial Dam

The standards were developed from the 1972 average annual salinity concentrations at each location and are currently unchanged (USBR, 1999 and Lee, 1989).

1.3.4 Previous Modeling Efforts

The existing Colorado River Simulation System was developed in the late 1970's in response to a need for a modeling system that could simulate operations for various hydrologic and demand sequences. It evaluates how proposed development occurring high in the basin might impact locations downstream from the development. CRSS includes a simulation model of the entire Colorado River system. It also includes a stochastic natural flow model to generate future stochastic flows and a salt regression model that estimates natural salinity associated with natural flows. The sim-

ulation model, termed the Colorado River Simulation Model (CRSM), has recently been re-implemented in RiverWare and is still used for operational planning.

A technique called index sequential modeling (ISM) develops risk-based estimates for CRSM results. A study by Ouarda et al. (1997) assesses the validity of ISM by analyzing synthesized streamflow developed from stochastic techniques compared with ISM. The study explains, "ISM utilizes a synthetic hydrologic database constructed from a series of overlapping short-term inflow sequences extracted directly from the historical record." Figure 4 illustrates the method in which T is the total number of historical years on record, N is the number of years to be modeled, and k is the generating index determined by the modeler. Indexing defines the separation in the years of each successive sequence. A total of $\frac{N \times T}{k}$ years or 250 years of synthesized flow is generated for the values given in the Figure 4.

Using CRSS, the study found that with indexing of 1 to 2 years, the index sequential modeling matches well with the stochastic properties of statistically generated hydrologies. Currently, the hydrologic database contains historical data for flow and salt mass from water years 1906 to 1995 for 29 inflow points included in CRSM, which are used by ISM.

A shortcoming of the method is that it requires an extensive historical record to produce synthetic streamflow of adequate length. Additionally, there is no comparison for salinity modeling in the system. The validity of ISM for risk-based estimates of salinity projections has not been verified. Using ISM to produce a synthetic salinity database should be verified to correspond well with stochastically generated salinity data before ISM is accepted as an adequate method for producing a synthetic salinity database.

Salinity predictions need to provide the probability of critical low-flow periods. These periods are of interest when trying to predict periods of increased salinity concentrations. ISM cannot provide this needed analysis, but times series analysis can be

used to find the probability of low-flow periods.

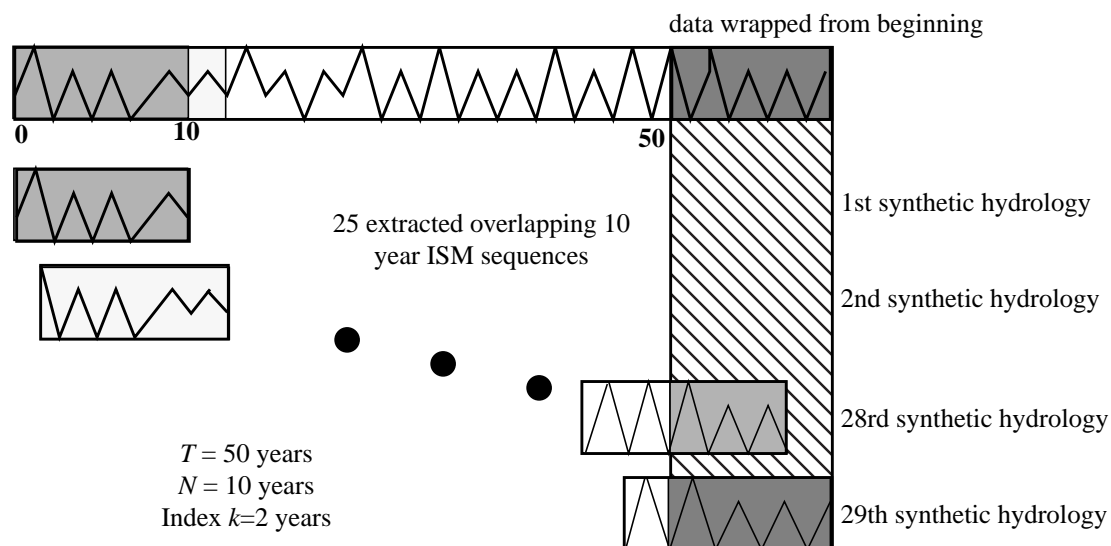


Figure 4: Illustration of Index Sequential Modeling adapted from Ouarda et al., 1997.

1.4 Research Tasks Performed

This research identified several specific technical issues as a result of the literature review and an evaluation of the shortcomings of current prediction models. These are methodological issues relating to modeling uncertainty in a long-term, basin-scale salinity planning model. This research addressed these issues in order to define the combination of methodologies, which together form an effective modeling system with quantifiable uncertainties. These issues are described briefly in the following sections.

1.4.1 Research Issues

1.4.1.1 Identifying Sources of Uncertainty and Determining How to Model Them

The first task of this research was to identify the important sources of uncertainty in predicting future salinity levels in a basin, then to determine how these uncer-

tainties should be represented in the modeling system. Table 1 and Table 2 list all the possible sources of uncertainty for both water quantity and salinity in their relative order of magnitude. The modeling system does not include some of these sources of uncertainty because they have small relative magnitudes of uncertainty and/or the uncertainties would not have a significant effect on the results of the basin-scale prediction modeling system. This research identified the following sources of uncertainty as the most important in a long-term planning model:

- hydrologic variability
- estimation of natural salt loading given natural flow
- estimation of historical salt loading from human-induced sources
- estimation of changes in salt loading due to future land use changes

We found that other sources of uncertainty, such as those contributed by measurement errors and modeling processes, contribute a minimal source of uncertainty compared to sources such as natural variability. Therefore this research excluded uncertainty resulting from measurement error.

Table 1: Uncertainty in Water Quality



	parameter	magnitude (source)
greatest uncertainty  least uncertainty	natural salt load -natural variability vs. measurement error	(USGS)
	human-induced salt load -salinity pickup	
	extension of salt load record -from salinity-flow regression relationship	(USGS)
	modeling process (theory vs. observation) -fully mixed reservoir -effects of stratification -bank storage relation to natural flow	(Hendrick, 1972) (USBR, 1987 p79)
	conservative substance -concentrating effect -no precipitation in reservoirs	(Liebermann et al., 1988 p90)
	EC or TDS measurement	(USGS)

Table 2: Uncertainty in Water Quantity

	parameter	magnitude (source)
greatest uncertainty  least uncertainty	natural flow -natural variability vs. measurement error	(USGS)
	consumptive use - upper basin - irrigated land estimates - Blaney Criddle method	(Jensen et al., 1990)
	unmeasured returns - lower basin	(Owen-Joyce, 1987)
	extension of streamflow record -from correlated gauges	(USGS)
	modeling process (theory vs. observation) - elevation to volume table - elevation to area table - CRSS bank storage - CRSS evaporation	(Carron, 2000) (Carron, 2000) (USBR) (USBR)
	measurement error - USGS stream gauge - pool elevation - precipitation	+/- 5% (USGS) +/- 0.1 feet (USBR) (USGS)(NOAA)

As discussed previously, uncertainty can generally be separated into *random* (stochastic) and *knowledge* uncertainty (Tung, 1996). Natural variability represents random uncertainty. The natural variability of hydrology cannot be reduced but must be included in uncertainty analysis. The uncertainty in measured data and the simplification of physical process for modeling purposes represents knowledge uncertainty. For example, uncertainty can originate from errors when measuring a value, including streamflow or conductivity. Models represent physical processes with simplified mathematical relationships. Using these simplified mathematical relationships to represent complex physical processes introduces uncertainty in a model. For example, the physical process of salt loading will have an error associated with the simplified mathematical relationships used to represent it in a model. If salt loading is not related to flow, the variations in the loading due to changes in flow will not be modeled. We know salt loading is related to flow; therefore, this simplification should be developed around a relationship between salt loading and flow to reduce error in this simplification. Knowledge uncertainty is due to the current level of knowledge we have about the measured data or the physical processes. Knowledge uncertainty can be reduced with improved data collection methods and additional research.

1.4.1.2 Representation of Stochastic Hydrology

The greatest influence on salinity in the river basin is the stochastic nature of hydrology resulting from natural variability. Hence, a useful salinity planning model must incorporate the natural variability of flow to provide useful predictions. In the stochastic salinity model previously discussed, Malone et al. (1979) used the stochastic properties of the historical record to incorporate the uncertainty in the change in salt loading for each reach. He did not use these stochastic properties to incorporate natural variability in flow because a relationship between flow and salt was not developed for the model.

In the existing CRSS, ISM generates synthetic streamflow sequences that

exactly match the historical record, shifted in time. ISM only produces synthetic streamflow sequences that have occurred in the past. Hydrologic events that have not occurred in the historical period we recorded cannot be captured in either uncertainty analysis.

Of particular interest for salinity modeling are sustained low-flow periods (droughts). During these periods, we would expect increases in salt concentrations throughout the system. Being able to incorporate the uncertainty of hydrologic events that are possible, but have not occurred historically, extended the modified model beyond past modeling efforts on the Colorado River and allowed the model to include critical low-flow analysis not included in current uncertainty studies.

Time series analysis techniques provide methods to generate synthetic hydrologic events that have not occurred historically, but are possible given the statistical properties of the historic data. These properties include basic statistics, such as mean, standard deviation, skewness coefficient, coefficient of variation, and maximum and minimum values. The analysis can find correlations between multiple stream gauges and specific statistics related to drought, flood, and storage.

There are two key categories of time series analysis and modeling techniques: parametric and nonparametric technique. Parametric techniques include auto regressive (AR), auto regressive moving average (ARMA), and periodic auto regressive (PAR) models. Most parametric models assume the time series is distributed in a Gaussian (normal) distribution. If the time series is not in a Gaussian distribution, the time series must be transformed. A hydrologic time series is typically transformed with a log or power transformation. The transformed data can then be used in the parametric model. After the model generates synthetic hydrologies, the hydrologies must be back-transformed. The data is back-transformed by multiplying the time series by the inverse log or power transformation. The back-transformed data is not guaranteed to preserve the basic statistics of the transformed data (Sharma et al., 1997; Salas,

1985; Bras and Rodriguez-Iturbe, 1985).

Many software packages can help to develop parametric models. Stochastic Analysis, Modeling, and Simulation (SAMS) (Salas et al., 2000) is a software package designed for parametric time series analysis of hydrologic data. SAMS can help to analyze the statistical properties of the historic flows and then fit the streamflow data to a model that can be used to synthesize streamflow sequences. It can generate synthetic streamflow with longer periods of drought than the original sequence. By generating many sequences, it is possible to determine the probability of droughts of a particular length.

Nonparametric techniques use splines, kernel functions, nearest neighbor methods, and orthogonal series methods to perform statistical estimation (Silverman, 1986). Nonparametric statistical estimation is performed by resampling the historic time series. The techniques are nonparametric because they do not require estimating any parameters for the model from the historic time series. Additionally, the model does not require an assumption of the time series probability density function. Therefore, nonparametric techniques preserve higher order statistics, like the probability density function and bivariate probability density function when they are not Gaussian or nonlinear.

Another advantage of nonparametric techniques are that they can easily be driven by multiple variables. The variables can include precipitation or climate information. Using climate information with a nonparametric technique allows the model to capture shifts and trends related to climate (Rajagopalan and Lall, 1999). Conversely, incorporating this information in a parametric model can be involved and difficult.

The current technique to generate stochastic hydrology, ISM, is a simple nonparametric technique. Recently developed nonparametric techniques eliminate some of the problems that are found in synthetic streamflows generated with ISM, such as

producing only values and time series sequences that are seen in the historic record. Nonparametric techniques, such as the modified k-nearest neighbor algorithm developed in this research, eliminate many of the drawbacks of ISM, providing an alternate nonparametric technique.

1.4.1.3 Separation of Natural Flow from Gauged Flow

Past modeling efforts raised the issue of separating natural flow from gauged flow to estimate the future inflows to the system. The separation of natural flow from gauged flow forces the separation of natural salt mass from gauged salt mass. While measured data is available to support the separation of flow, there is little to no measured data available to support the separation of salt mass.

In an ideal modeling scenario, we would have gauges at the top of all tributaries in a river basin. These *ideal* gauges would have no human development upstream; therefore, the gauged data would represent the *natural* flow into the river. With a historical record of the natural flow, the record could be used directly to predict future inflow into the system. In reality, we do not have gauges above human development for all tributaries. Most of the historical gauged data includes the effects of human development, which poses a problem when attempting to use this data to predict future inflow into the system. If the variations are not considered, predictions will consider the natural and human-induced changes in flow. When modeling future variations in human development, we do not want past variations influencing the prediction. The human-induced variations in flow must be accounted for in the historical record. Those variations can be accounted for in one of two ways. The historical record either must be adjusted to represent a single time of development (i.e. the development in the basin in 2000), or the effects of human development must be removed from the historical record. CRSS removes the effects of human development from the historical record to produce a natural flow database that is used for predictive modeling.

Human development, including diversions, reservoir regulation, exports, and

imports, impacts the flow. Gauged data with human development upstream of the gauge includes natural and human-induced effects on flow. Because human development on many rivers began before flow gauging, natural flow cannot be determined directly. Natural flow can only be calculated by removing the effects of human development from gauged flows.

The same is true for calculating gauged salinity concentrations. The concentrations represent natural and human-induced sources of salt loading into the river. If natural flow is separated from gauged flow, it is also necessary to separate natural salt mass from gauged salt mass to predict future salinity levels. Gauged salt mass cannot be used with the separated natural flow for prediction. The gauged salt mass must also have the salt (mass) loading resulting from human-induced sources removed. Neither natural nor human-induced salt loading can be directly measured. Typically, a mass balance technique is used to separate salt loadings.

In addition to requiring a natural salt loading, future land uses must have an associated human-induced salt loading to predict their effects on future salinity concentrations. A critical area for this research is finding an appropriate method to model natural and human-induced salt loading with the available data.

When modeling uncertainty, trying to reduce gauged data to represent natural and human-induced sources of uncertainty introduces additional uncertainty in the model. For example, if we use natural flow, we cannot directly find the uncertainty of the natural flow. The natural flow uncertainty can only be developed from uncertainty in the historical gauge measurements and all the measured and unmeasured effects of human development. Therefore, the uncertainty in gauged flow is much less than in natural flow. Disaggregating the gauged flow into natural and human-induced flow introduces additional uncertainty in the model.

This investigation did not find an alternate technique to model future salinity levels without using a model that separates historic flow and salt to natural and human-

induced flow and salt. For this reason, we found it necessary to use natural flow and salt to drive the modified CRSS model. The analysis of the existing stochastic technique to generate natural flow, ISM, and two alternate techniques is detailed in Chapter 2, “Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow.”

1.4.1.4 Method of Modeling Salinity Concentrations

The literature review found standard simplifications of physical processes when modeling salinity at a long-term, basin-scale. The modified CRSS simulation model uses the simplifications that were implemented with the existing CRSS simulation model. The simplification includes modeling salinity as a conservative substance. Modeling dissolution and precipitation of salts becomes cumbersome and impractical at a basin-scale. An extensive amount of data is required to include chemical reactions. This type of modeling is usually reserved for field-scale models in which acquiring the required data is possible. Therefore, reservoirs were modeled as completely mixed systems. As stated previously, research has generally shown that variations in salt concentration entering a reservoir are greatly reduced when salt leaves the reservoirs (Helsel and Hirsch, 1992). Modeling reservoirs as a completely mixed system adequately represents the reservoir’s physical salinity process for the long-term, basin-scale.

We developed the process to describe natural salt load in the modified CRSS from a relationship between calculated natural flow and calculated natural salt. The technique includes a means to incorporate the uncertainty in the relationship that generates the natural salt time series. The proposed technique is introduced and explained further in Chapter 3, “Statistical Nonparametric Model for Natural Salt Estimation.”

1.4.1.5 Issues of Scale in Human-Induced Salt Loading

Salt loading from land use needs to be represented at an accurate and meaning-

ful scale. Using a small scale introduces many uncertainties. Large scale (up to a point) is more accurate, but may not satisfy the needs of the model. To have a meaningful uncertainty analysis, an appropriate scale for each set of data must be found. For example, the model for the Colorado River must be able to model future changes in land use. Thus, we needed a scale for diversions that reflects the scale of land use.

We needed a methodology to predict changes in salt load with changes in land use. For example, diversions needed a scale to allow different land uses in the CRSS simulation model. Salt loading values needed a scale to complement the scale of the diversion information. Each land use will result in a different salt loading from the use. To predict the effects of changing land use on salinity concentrations, we needed to quantify the change in salt load when land use changes. For example, if land that historically was used for agriculture is proposed to be set fallow, we needed to quantify the reduction in salt loading from the land since it will no longer be used for agriculture.

After collecting the data available describing salt loading into the basin from human-induced sources, salt loading can be understood at a very basic level. Collection of salt loading from human-induced source is difficult to quantify because salt enters the basin from many diffused sources. Mass balance techniques are used in studies such as Iorns et al. (1965). Iorns provided the most detailed salt loading study performed within the Colorado River basin. Smaller scale studies provided additional salt loading information for specific regions of the basin (USBR, 1974). Because of the lack of available data, a limited choice of scales was available. Currently, the basin is broken down into regions between gauges. Generally, agriculture contributes a single constant salt loading between each gauge. Additionally, exports remove the salt that is present in water transported from the river basin. Most exports occur near the basin boundary and are high in the basin. Generally, the water high in the basin is of excellent quality; in the Colorado River basin the salinity of this water is generally less

than 100 mg/L annually. This information directed the modeling of human-induced salt addition and removal.

1.4.1.6 Overall Modeling Approach and Representation of Results

We considered two types of uncertainty propagation for this model: Monte Carlo simulation and first- order second moment (FOSM) techniques. Both techniques have advantages and disadvantages that we considered in deciding the appropriate method for the CRSS simulation model.

Monte Carlo simulation uses multiple simulations to directly build the probability density function for the output variables. A probability density function is first approximated for the input variables, and the method then approximates the distribution of the output variables. Monte Carlo simulation presents output uncertainty very effectively with highly uncertain systems or when the system responds in a non-linear fashion (LaVenue et al., 1989). The major disadvantage of the Monte Carlo simulation is the added computer time needed to run numerous (100-500) simulations to get an output distribution with minimal sample error.

In contrast, FOSM uses the mean and variance of the input variables and any correlations between the variables to determine the statistical properties of the model output. This technique mathematically propagates the statistical properties of the input variables to the output variables. The advantage of FOSM is that highly complex processes can be modeled because only one simulation is required to achieve results. The disadvantage is that a highly nonlinear system is not represented well, and highly uncertain systems may not provide meaningful output uncertainty (LaVenue et al., 1989).

On the basis of this analysis, Monte Carlo simulation is the most effective method to model the system. The primary need in this research was the ability to model a highly uncertain, nonlinear system. Computing time was not a critical issue with available computing facilities.

Possible methods to report or show uncertainty to a modeler include:

- mean and variance
- boxplots
- confidence interval, i.e., 95 percent confidence bands

Mean and variance provide basic statistical properties of the output value. Boxplots graphically display the median, interquartile range, and the approximated the 5 percent and 95 percent confidence range of a series of data. Confidence intervals provide bounding lines around the mean that specify the probability a value will be between them. Typically, the lower bounding line represents the 5 percent probability a value will be below the line, and the upper bounding line represents the 95 percent probability a value will be below the line.

These different methods convey different information regarding the model's uncertainty. We considered the availability of these multiple methods because each is useful to modelers, depending on the type of analysis they are trying to perform. Examples of each reporting method, along with others, are included throughout the thesis.

1.4.2 Model Integration

The methodologies selected to model flow and salinity are dependent on the characteristics of the specific basin and the data available for the basin. The final representation of each of the sources of uncertainty in the prediction model depended on 1) the magnitude of the contribution to the uncertainty of the final prediction, 2) the significance of the contribution in the prediction, in terms of intended purpose of the model, 3) how the processes and uncertainties can most effectively be represented in view of available data, and 4) compatibility with the overall modeling approach.

The model results included random uncertainty resulting from hydrologic variability and knowledge uncertainty resulting from measurement error and simplifica-

tion of complex processes for modeling purposes. As discussed previously, a disaggregation of gauged flow and salinity was needed to provide future predictions of salinity concentrations in the basin. The scale for the CRSS simulation model was defined by the data available for quantifying human-induced flow and salt.

1.4.3 Case Study - Colorado River

A case study tested a modified CRSS. The case study modeled a portion of the Colorado River basin. Modeling methods and data limitations restrict the case study to a single gauge within the Colorado River basin. Figure 5 shows a detail of the single gauge 09072500 (Colorado River near Glenwood Springs, CO) and the corresponding drainage area that were modeled in this research. Working with a single gauge allowed investigation of multiple methods that generate stochastic hydrology representing natural flow.

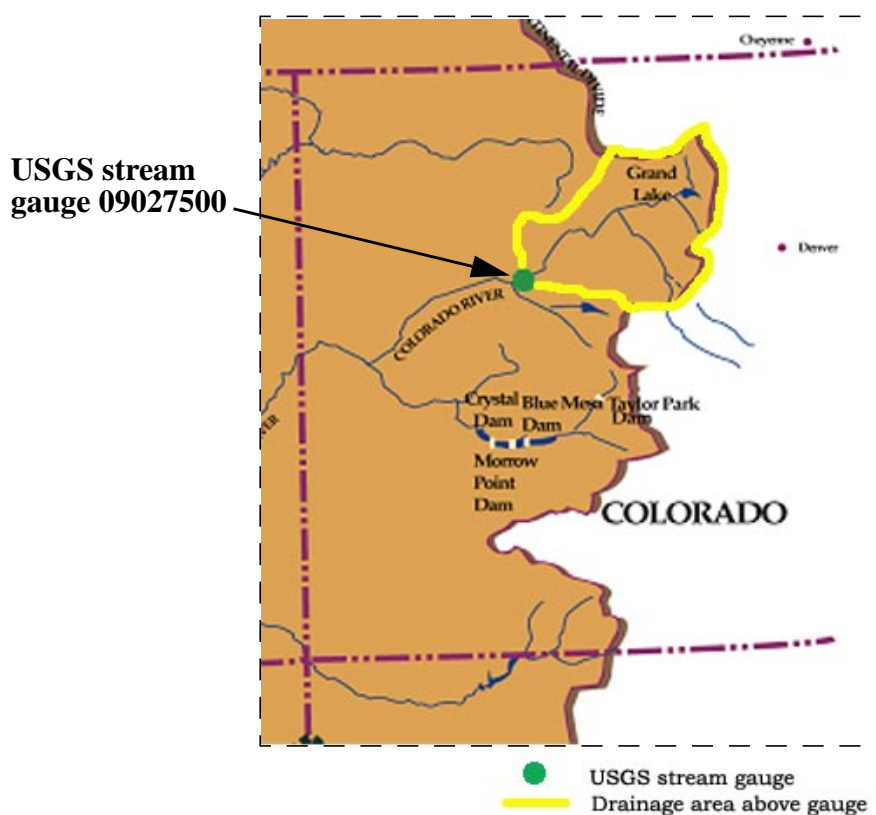


Figure 5: Detailed outline of the case study drainage area and stream gauge.

1.4.3.1 Interconnection for Modeling System

The modified CRSS developed in this research was composed of three primary models. The first model was used to generate synthetic hydrology. The synthetic hydrology represented a time series of hydrology that is statistically possible given the data fit to the model. The incorporation of climate data to condition the natural flows is briefly explored. The second model used the synthetic hydrology to calculate the natural salt mass as a function of the natural flow. The output from the first model (natural flow), along with the output from the second model (natural salt mass), were input to the final model. The final model was developed in the RiverWare river basin modeling environment. RiverWare allows varying spatial scales and flexibility to expand the model further as more details are learned about the physical system. A monthly time step was used to facilitate use of the existing set of river policy available with the existing CRSS simulation model.

Together, these three models simulated the historic flow, salt mass, and salt concentration at USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). Figure 6 depicts the interconnection of the modeling techniques.

Investigating the existing CRSS models and data has been ongoing since August 1999. A series of six status reports presented to the Colorado River Salinity Control Forum describe the investigation in detail. The information gained from these investigations provided data and knowledge used to develop a modified nonparametric K-NN model for generating streamflow. The development of this model and comparison to the existing model and an equivalent parametric model are presented in Chapter 2, “Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow.” A discussion of conditioning the generation of natural flows on climate information is included in a discussion concluding Chapter 2. A statistical nonparametric local linear regression with K-NN residual resampling for estimating salt mass is developed and compared to the existing salt model in Chapter 3, “Statistical Nonparametric

Model for Natural Salt Estimation.” The results from the models to generate synthetic natural flow and an associated natural salt were then taken into the CRSS simulation model. The CRSS simulation model modeled both historic and future flow, salt mass, and salt concentration at USGS stream gauge 09072500. The modified modeling system results are compared with the existing modeling system in Chapter 4, “Historic and Future Salt Concentration Modeled in RiverWare.” which discusses the development and results from the modified CRSS.

The final chapter, Chapter 5, “Conclusion and Recommendations for Future Work,” provides a summary of the result from each model within the modified modeling system and describes future work that could develop from this research.

1.4.3.2 Model Validation and Metric of Success

We performed model validation to ensure the model performed as intended. We also wanted to ensure the model reproduced observed data. If a model is able to reproduce observed data well, there is more confidence the model can produce appropriate projected results.

The model was validated over the time period of water years 1941 to 1995. Each model in the modified modeling system was validated at the completion of its development. The final CRSS simulation model ensured that all the models in the system could together model historic flow, salt mass, and salt concentration to an acceptable degree.

The first model generated the stochastic natural flow. To validate the parametric (PAR) and nonparametric (K-NN) model, each generated an ensemble of 100 simulations, and the probability density functions (PDF) from each ensemble were plotted against the historic PDF to ensure each technique was able to preserve the PDF.

Interconnection of Research Topics

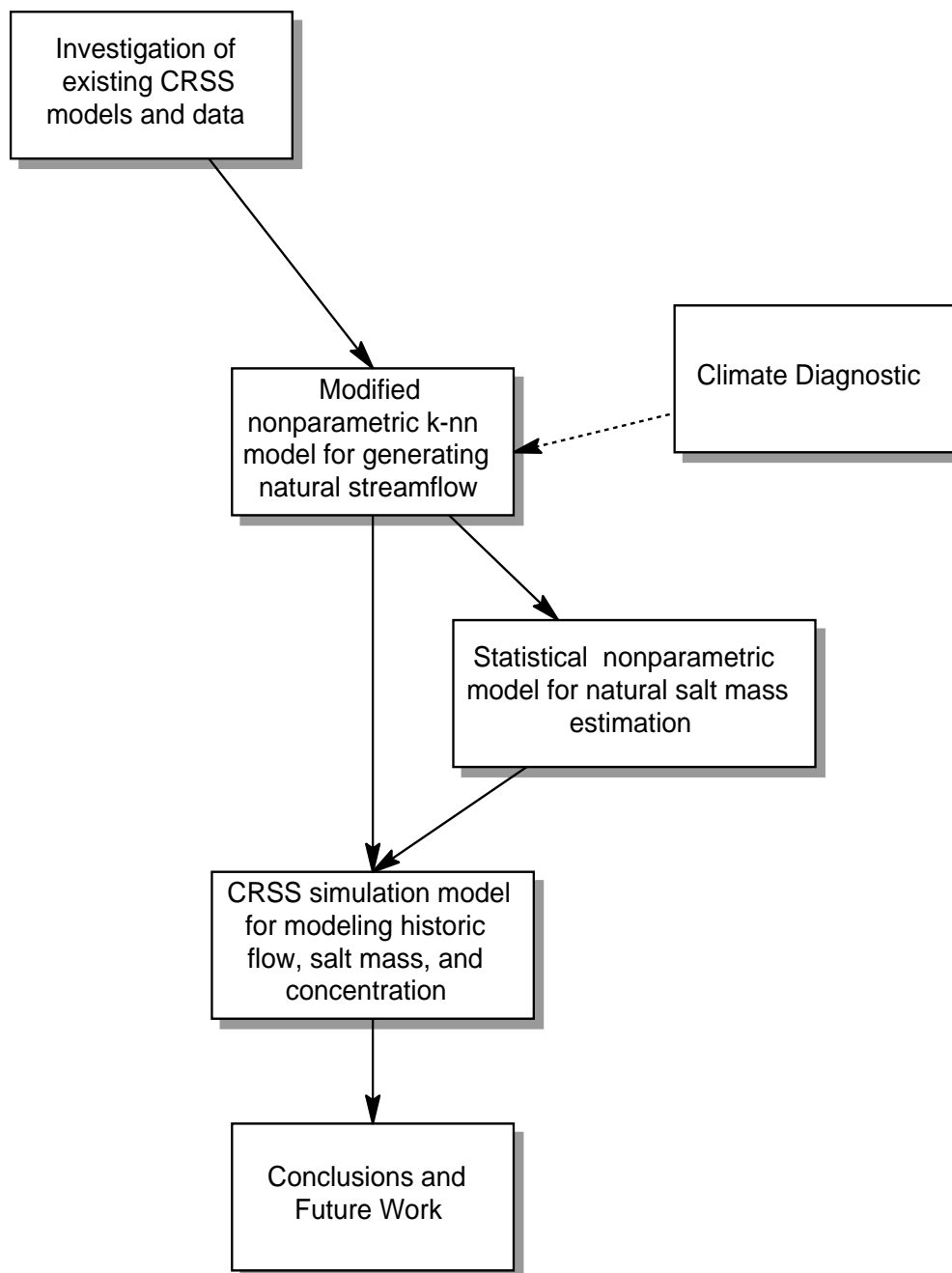


Figure 6: Flowchart depicting interconnection of modeling system.

The second model generated natural salt mass associated with synthetic natural flow. This model was first validated by using the model to generate a natural salt mass for a historic period. Historic flow and salt data were both available from water years 1941 to 1995. The natural flow and salt data fit the regression from water years 1941 to 1985. The fitted salt model was then used to generate natural salt mass from water years 1941 to 1995. The model's ability to reproduce the period it was fit to (water years 1941-1985) and an independent period (water years 1986-1995) is checked. The results from the statistical nonparametric natural salt model and the existing USGS natural salt model were also compared.

The third model generated the historic flow and salt mass given the results from the first two models. To validate this model, the historic period from 1941 to 1995 again was generated with the existing CRSS model and the modified CRSS. The PDF for simulated historic and natural flow, salt mass, and salt concentration were each compared with the PDF for the observed historic data. We deemed the modeling system appropriate if the probability density function was preserved.

1.4.3.3 Uncertainty in Predictions

A simulation for future predictions was performed. The most recent predictions for demands in the basin were input in the CRSS simulation model, and the model was used to predict salinity levels with uncertainty until 2062. This model run demonstrated use of the modeling system and uncertainty analysis for future prediction. The results were compared from the existing CRSS and the modified CRSS that were developed in this research. The results from the comparison were analyzed and discussed.

1.4.3.4 Policy Modeling

Using the modified CRSS will allow more informed policy decisions to be made. To demonstrate the applicability and advantages of the modified CRSS, scenar-

ios considering violations of a fictional salinity concentration and salt mass standard above USGS stream gauge 09072500 were simulated. The uncertainty analysis informs policy makers how well the system is understood and where focused research could improve the understanding.

Violation statistics were generated that display the number of times a standard is violated with uncertainty for the statistic. For the time period 1941 to 1995, the number of simulated violations was compared to the historic number of violations.

For the future simulation, the number of violations simulated by the new modeling system was compared to the existing modeling system.

In the future, when the new modeling system is expanded to a basin-scale model, the influence on river basin policy analysis and decisions can further be evaluated. Policy analysis might include changes in operational and basin planning policy. Reservoir releases designed to dilute high salinity concentrations downstream of a reservoir would be a possible operational policy scenario. Planning policy can analyze the impacts of new water quality improvement projects or changes in land use. With uncertainty analysis, the scenarios could incorporate the uncertainty of the project's effectiveness and also look at how multiple projects throughout the basin will affect salinity levels at points of interest (i.e., where salinity standards are established).

Operational policy was modeled with the rule language available in RiverWare. Planning policy was modeled by changing water uses or demand levels entered into the model as inputs.

1.5 Summary

The basic problem this research addresses is the long-term prediction of salinity with uncertainty at a basin scale. We expand on existing techniques employed for risk assessment, such as index sequential modeling. We found Monte Carlo simulation to be a readily accepted technique for uncertainty modeling in water resources and

very effective in highly uncertain systems such as the Colorado River basin. The uncertainty analysis will advance and consolidate the current information regarding the uncertainty of inputs for models of the Colorado River basin and the uncertainty inherent in the methods used to model a system as complex as the input and transport of salt at a basin-scale.

The modified CRSS provides modelers and policy makers with a preliminary step toward creating a tool to predict salinity at a basin-scale. Confidence was applied to the outputs from the model and the uncertainty of inputs was prioritized. Research can focus on reducing the uncertainties that exhibit the greatest influence on the model.

Chapter 2

Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow

2.1 Introduction

Long-term operational and planning studies in a river basin require the ability to predict streamflow variability (McMahon, 1996). Typically, this ability involves developing a stochastic streamflow model to generate synthetic sequences of streamflow. The generated sequences preserve the historic statistics (mean, standard deviation, lag(1) correlation, and skewness coefficient) and higher order statistics depending upon the model. These models work on the premise that the statistics of the historical flows are likely to occur in the future, i.e., the stationary assumption.

In the Colorado River basin, modeling natural flow variability is important because it is the largest source of uncertainty, as we identified in Chapter 1, “Problem Definition.” Investigating and improving the generation of stochastic flow improves modeling uncertainty in the Colorado River basin. The current technique for generating stochastic natural flow for the Colorado River used by the Colorado River Simulation System (CRSS) simulation model has some limitations. Improving the technique for generating synthetic streamflow sequences would allow basin policy managers to fully understand the effects of natural streamflow variability on future streamflow sequences. We also would like to provide basin managers with a modeling technique that will allow future scenarios of drought or surplus events to be developed in a statis-

tically accepted method. Further, recent studies show a link between climate and extreme streamflow events. We investigated relationships between climate and streamflows in the upper Colorado River mainstem, which can condition streamflow generation on climate data.

One of the first steps of our research was to investigate monthly streamflow data in the Colorado River basin, a required input to the CRSS to model operations, for various hydrologic and demand sequences. We studied the Colorado River basin streamflow at a single gauge, USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). The probability density function (PDF) for monthly natural streamflow from water years 1906 to 1995 exhibits non-Gaussian features that vary from month to month. Generally, the natural streamflow data is skewed towards the low flows, with an extended tail in the high flows. The annual PDF is slightly skewed towards the low flows, with a concentrated point at the higher flows.

The most effective stochastic flow generation model should preserve the skewness and extended tail behavior seen in the natural flow data. Preserving the tails is of particular interest because the tails exhibit the probability of extreme low or high flow. During periods of extreme low flows, salinity has historically increased in concentration. Modeling these periods to our best ability provided confidence in the results of the entire modeling system. The technique should be capable of generating synthetic streamflow values and sequences that have not occurred in the historic record, but are statistically possible given the historic record's statistics. Generating values above or below the maximum and minimum of the historic record should also be possible.

We investigated three different techniques for generating synthetic streamflow: (1) a nonparametric index sequential method (ISM) that is currently used by the U.S. Bureau of Reclamation (USBR) to generate synthetic hydrology for the CRSS, (2) a parametric periodic auto regressive (PAR) model that is traditional and widely used, and (3) a nonparametric k-nearest neighbor (K-NN) model developed in this study.

Our motivation for this investigation was to identify alternatives to the currently used ISM, which has limited ability to generate a variety of synthetic streamflow sequences. We first present a background of stochastic models followed by a description of the three models and results.

2.2 Background

Stochastic streamflow models were traditionally developed in an auto regressive (AR) framework and subsequently ARMA and PAR models (Shama et al., 1997). These are also referred to as parametric models because they involve selecting an *appropriate* model and *fitting* parameters to it. Parametric models assume that the time series is normally (Gaussian) distributed (Salas, 1985). More often than not, streamflows are not Gaussian distributions, thereby violating this assumption. To address this, the data is *transformed* to a Gaussian distribution using a log or power transformation before fitting a parametric model to the transformed data (Shama et al., 1997). The synthetic sequences generated from the model are back-transformed into the original space. This process of fitting the model on the transformed data and then back transforming it often does not guarantee the preservation of statistics (Sharma et al., 1997; Salas, 1985; Bras, 1985; Benjamin, 1970).

The parametric models generally preserve the mean, variance, and auto correlations (depending on the order of the model). As the time series is transformed to a Gaussian distribution (or near Gaussian) by appropriate transformation, the skewness is preserved to the extent the transformations are good.

Parametric models require estimating multiple model parameters, depending on the type of model. Considerable uncertainty can exist in the estimation, depending on the length of the historical data, which adds to variability in the simulations. Furthermore, because they are restricted to a Gaussian framework (Gaussian distribution assumption), parametric models cannot reproduce non-Gaussian features such as

heavily skewed distributions or bimodal distributions that may be present in the historical data (Sharma, et al., 1997; Lall and Sharma, 1996), which is one of the major drawbacks of the parametric model. Such features are not uncommon in Colorado River basin flow data (Lall and Sharma, 1996), as we show later in this chapter when we compare the methods on the Glenwood Springs streamflow gauge.

Nonparametric models have been developed to address these drawbacks of the parametric models. The simplest nonparametric model is the ISM, which involves selecting chunks of historic data. For example, if we have 100 years of historical data, without wraparound we can generate 80 sequences of 20-year lengths, 70 sequences of 30-year lengths, etc. The advantages are that it is simple and easy to implement, assumption free, and can reproduce the entire distributional properties of the historic data: the mean, variance, auto-correlation, etc. The main disadvantage is that only historically observed sequences can be generated.

Recently developed nonparametric models, (Lall, 1995; Lall and Sharma, 1996; Tarboton et al., 1998) have tried to address the problems of ISM and parametric models. A few types of nonparametric models exist for streamflow generation: kernel based (Sharma et al., 1999), nearest neighbor based (Lall and Sharma, 1996), and hybrid parametric/nonparametric models (Srinivas and Srinivasan, 2001).

In effect, the nonparametric models estimate the marginal and conditional probability density functions locally and simulate sequences from them. They are assumption free and can model any shape of the density function.

2.3 Index Sequential Method

USBR uses CRSS to simulate flows in the Colorado River basin. Natural flows are a required input to drive the model when simulating historic flows. Natural flows currently are generated using the ISM. Colorado River basin policymakers have been using the ISM for more than a decade to develop risk-based estimates of important

system output.

The ISM is a nonparametric stochastic technique that generates synthetic hydrologies by sequentially *block bootstrapping* the historic time series. Block bootstrapping is a technique in which a block of the historic time series is resampled as a synthetic trace. For example, our historic time series for stream gauge 09072500 is 90 years in length, from water years 1906 to 1995. To model 25 years into the future, the method extracts a 25-year block from the historic streamflow record then shifts one year forward and extracts 25 years again, repeating the process 90 times. When the end of the historic record is reached, the record is continued from the beginning of the time series. A schematic of the technique is shown in Figure 7. The intent is that every year to be simulated sees all the hydrologies of the historic record. Consequently, the simulated sequences have the same distributional properties as the historic data.

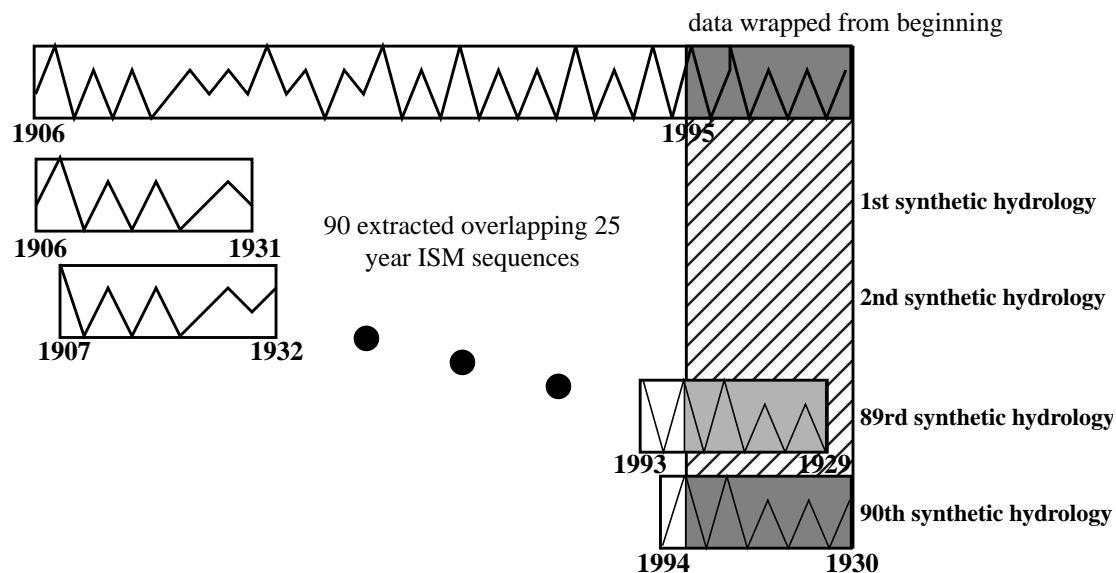


Figure 7: Schematic of the ISM (adapted from Ouarda et al., 1997). The synthetic hydrologies, each 25 years in length, are shown below the original 90-year time series. The additional 24 years used for wraparound are shown in shading.

Ouarda (1997) compared the nonparametric ISM to a traditional parametric autoregressive method to determine how well the two modeling techniques allowed project dependable capacity to be estimated. The study found using ISM, synthetically generated sequences statistically corresponded acceptably with sequences generated from an AR(1) model. Because power generation was of interest, the study looked at the cumulative density function for total energy out and found that the two techniques compared well at the lower and middle sections which are critical to energy production modeling, of the function.

Kendall and Dracup (1991) also compared ISM and an AR(1) model. Again, an annual model was developed to compare the two techniques. Their study compared reservoir storage capacity at Lake Mead and Lake Powell. They also found that the two techniques generated sequences that are not significantly different. Generally, they found that flows generated by the AR(1) model result in slightly higher storage levels than the ISM. They also stated that the AR(1) model has a tendency to underestimate the occurrence of severe droughts. Further, the ISM did not perform well when the tails of the distribution were of interest. They found at 90 percent exceedance and greater, the AR(1) model produced lower flows. They recommended considering AR(1) models if tails of the distribution are important.

Neither of these studies explored higher order statistics, such as the PDF of the generated synthetic flows or the bivariate probability of the current month's flow dependent on the previous month's flow. Higher order flow statistics were important in our work because of our interest in a stochastic techniques performance during extreme events such as high or low flows. We needed to be able to simulate severe droughts to accurately simulate the effects of low flow periods on salinity levels. These events are described in the tails of higher order statistics, such as the PDF of the generated synthetic flows, an area in which the ISM did not perform as well as a parametric model.

Advantages of this technique, which have led to its widespread use, are that it is simple to implement and that it does not require the modeler to estimate parameters (Shama et al., 1997). Because it reproduces the historical data, it preserves all of the statistics, along with the probability density function.

A limitation of this technique is that it cannot generate synthetic hydrologies or sequences that did not occur in the past record. Consequently, it cannot create droughts or surpluses of lengths different from those that have occurred in the past record.

2.4 Periodic Auto Regressive Method

The parametric periodic auto regressive (PAR) model is a traditional and accepted method to model synthetic streamflow. The PAR model is also termed a seasonal auto regressive model and is distinguished from other auto regressive methods in that it explicitly models seasonality, such as the case with streamflow.

The general equation for a PAR model of order ρ is given by

$$y_{\vartheta,\tau} = \mu_{\tau} + \sum_{j=1}^{\rho} \Phi_{j,\tau}(y_{\vartheta,\tau-j} - \mu_{\tau-j}) + \varepsilon_{\vartheta,\tau} \quad \text{Eq. 2.1}$$

where: y is the streamflow process,

ϑ is the year. τ is the season,

μ_{τ} is the mean of the process in season τ streamflow,

Φ is the auto regressive parameter,

$\varepsilon_{\vartheta,\tau}$ is the uncorrelated normally distributed noise term with mean 0 and variance $\sigma^2(\varepsilon)$.

The season could represent months or another subset of a year. In our investigation, we developed a lag(1) monthly model or a model of order $\rho = 1$ with $\tau = 12$. This model can be written as

$$y_{\vartheta,\tau} = \mu_{\tau} + \Phi_{1,\tau}(y_{\vartheta,\tau-1} - \mu_{\tau-1}) + \varepsilon_{\vartheta,\tau} \quad \text{Eq. 2.2}$$

which means each season represents a month, and the current month flow is dependent linearly on the previous month's flow. The model parameters $\Phi, \sigma^2(\epsilon)$ are estimated for each month from the data. With $\rho = 1$ and $\tau = 12$ (representing 12 seasons), 12 estimates of $\Phi_{p,\tau}, \mu$, and $\sigma^2(\epsilon)$ have to be computed, requiring estimation of 36 parameters. Method of Moments, approximating Least Squares, or Yule-Walker equations are used to estimate the auto regressive parameters (Salas, 1985; Bras, 1985).

The PAR method is a parametric method. Traditionally, stochastic models were developed with parametric models, such as AR, ARMA and PAR models. The PAR parametric method preserves the mean, standard deviation, and lag(1) correlation. This parametric technique is only able to preserve a Gaussian probability density function. An additional step is required to preserve the skewness coefficient of a time series. Colorado River monthly hydrologic data has months that do not exhibit a Gaussian distribution. The months must be transformed to change the shape of the PDF to approximate a Gaussian distribution. We transformed the Colorado River data with a log transformation using the equation $y = \ln(x + a)$ where: y is the transformed process, x is the original process, and a is estimated to bring the skewness coefficient towards zero.

Transformations can be used to approximate a time series skewness coefficient; this additional step is required for the parametric method but not for nonparametric methods. After generating the synthetic hydrology with the parametric model, the synthetic hydrology must be "back-transformed" by the inverse log transformation. The back-transformed data is not guaranteed to preserve the basic statistics of the transformed data (Sharma, 1997; Salas, 1985; Bras, 1985; Benjamin, 1970).

A software package developed at Colorado State University, "Stochastic Analysis, Modeling, and Simulation" (SAMS) (Salas, 2000) was used to develop the PAR(1) model. SAMS transforms the flow data to approximate a Gaussian distribution by the transformation process described earlier. The transformed data is then fitted to a

PAR(1) model using least squares to solve the auto regressive parameters. The fitted model is then used to generate synthetic sequences. Using SAMS facilitates transforming the data that is used to fit the parametric model and reduces the time required to fit parameters to the model.

The simulation process for a month, February for example, is

$$y_{\text{Feb}} = \mu_{\text{Feb}} + (y_{\text{Jan}} - \mu_{\text{Jan}}) + \sum (N(0, \sigma^2(\epsilon))) \quad \text{Eq. 2.3}$$

Using generated January flows, we generated a normal random variable with mean 0 variance $\sigma^2(\epsilon)$ and from Eq. 2.3 generated the value for February. The process is repeated for as many simulations as needed.

The advantages to using a parametric model like PAR are that (1) it can simulate values and sequences not seen in the past, (2) much past work exists that uses these methods for various simulations, and (3) a fully developed theoretical background exists. The main disadvantages are that the data must be transformed to a Gaussian distribution to satisfy the assumption of the model and that ϵ is generated from a normal distribution; hence, any values from $-\infty$ to $+\infty$ can be simulated, which can generate unrealistic values.

2.5 Traditional K Nearest Neighbor Method

Recent developments in nonparametric models have tried to address some of the concerns statisticians have with using the index sequential method and parametric methods. The concerns with the index sequential method are that it cannot create values or time series sequences that did not occur in the past. One of the goals of stochastic modeling is to explore values and sequences that did not occur in the past but could occur based on the statistics of the historic record. Parametric methods cannot preserve the non-Gaussian features of a PDF.

Recently developed nonparametric techniques include kernel based (Shama et

al., 1997), and K-NN based resampling (Lall, 1995). As mentioned earlier, the non-parametric techniques approximate the conditional and marginal densities of a time series and simulate from these. The simulation problem can be assumed to be a conditional simulation problem, i.e., a lag(1) model can be thought of as simulation from a conditional PDF. For example, a lag(1) model can be expressed as simulation from the conditional PDF

$$f(y_t | y_{t-1}) \quad \text{Eq. 2.4}$$

where y is a time series of flows.

The parametric models, such as PAR, essentially simulate from a Gaussian conditional distribution because they assume the data is normally distributed, as described in the previous section. The nonparametric models approximate the conditional PDF shown in Eq. 2.4, from the data. Thereby, they have the ability to reproduce any arbitrary PDF structure without any prior assumptions, unlike the parametric counterpart.

Lall and Sharma (1996) introduced a k-nearest neighbor (K-NN) model, distinguishing it from the ISM. Unlike ISM, the K-NN model does not use a block of the historic time series for each generated time series. Instead, the K-NN model resamples one month at a time, generating *sequences* of synthetic flow time series not seen in the historic record. To find successive months, the k-nearest neighbors to the current flow are found. The Euclidian distance between each neighbor and the current flow are calculated. The distances are then weighted, with the closest neighbor receiving the greatest weight. A discrete kernel is used to resample one of the k weighted neighbors. The resampled neighbor then becomes the current month flow. The process then moves up to the next month and is repeated. Rajagopalan and Lall (1999) demonstrated the K-NN technique in developing a daily precipitation model. The daily precipitation model uses six weather variables in a K-NN framework to choose an appropriate precipita-

tion.

One of the drawbacks of the K-NN technique is that the synthetic time series do not have values that differ from the historic time series. Nonparametric models that use a kernel density estimators alleviate this problem (Shama et al., 1997). However, the kernel based methods have problems in the tails of the distribution, creating unrealistic values like the parametric techniques do, but to a lesser degree. Also, this method can get unwieldy in higher dimensions (Lall, 1995).

To further address these problems, Srinivas and Srinivasan (2001) recently developed a stochastic model that incorporates the strengths of parametric and nonparametric models in a single technique. They *prewhitened* the streamflow time series with a periodic autoregressive model that removes the dependence in the historic flow sequence, then use a nonparametric moving block bootstrap to resample the prewhitened streamflow. The prewhitened streamflow smoothed the historic time series and filled in values between the historic data. This smoothing allows the technique to generate streamflows not seen in the historic record while preserving non-Gaussian features in the historic streamflow. The method performs well but still requires extensive steps to generate streamflows.

We adopted the K-NN model because it is simpler, effective, and more flexible than the kernel based models (Lall and Sharma, 1996; Rajagopalan and Lall, 1999). We modified this model to address some of the drawbacks of the K-NN model. The modification developed and tested here is one suggested by Lall and Sharma (1996).

The K-NN model developed by Lall and Sharma (1996) is a nearest neighbor bootstrap because the neighbors are computed to the flow at the current time step and one of them is *selected* from a weight function that gives more weight to the nearest neighbor and less weight to the farthest neighbor. The successor to the *selected* neighbor becomes the flow in the next time step. The algorithm is described in detail in Lall and Sharma (1996) and also in Rajagopalan and Lall (1999). This algorithm is akin to

approximating the conditional PDF (Eq. 2.4) and simulating from it. The advantage is that the traces generated are not exactly the same as historical traces, as for ISM. However, values not seen in the historical record cannot be generated, which is one of the main drawbacks. The modified version of the model, described in the next section, eliminates this concern.

2.6 Modified K Nearest Neighbor Method

The K-NN algorithm developed in our work is an improvement on traditional K-NN techniques. To keep the modeling simple but to allow creation of values that were not seen in the historic time series, we developed a technique discussed in the conclusion of Lall and Sharma (1996) and Rajagopalan and Lall (1999). They suggest a variation of the traditional K-NN model that addresses the inability to recreate values that had not occurred in the past. The techniques also reduced the problems of the kernel estimator while remaining simple to implement.

The modified K-NN technique develops a regression relationship between each month's flows and saves the residuals from the regression. The succeeding month's flow is first calculated from the appropriate regression, then the k-nearest flows to the flow from the regression are found. A residual from the k-nearest flows is resampled and added to the flow from the regression to perturb the regression. This scheme will allow the K-NN method to perturb the historic data within its representative neighborhood and allow extrapolation beyond the sample.

We describe the modified K-NN algorithm for a lag(1) model, indicating the current month's flow is solely dependent on the previous month's flow. Figure 8 shows the scatter plot of February and March streamflows. The solid line shows a nonparametric fit through the scatter. The nonparametric fit is a locally weighted regression scheme (Loader, 1999; Rajagopalan and Lall, 1997). Fit at any given point is based on a local polynomial fitted to the k nearest neighbors. The number of neighbors k is

determined with cross validation. The details are presented in Chapter 3, “Statistical Nonparametric Model for Natural Salt Estimation.” The algorithm starts by fitting a nonparametric fit for each month dependent on the previous month:

$$y_t = f(y_{t-1}) + e_t \quad \text{Eq. 2.5}$$

1. The residuals from the fit are saved.
2. Once we have the value of the flow for the current month y_{t-1} , we estimate the mean flow of the next month y_t^* from Eq. 2.5.
3. We compare k-nearest neighbors to y_{t-1} (these are shown in big circles in Figure 8).
4. We then select a neighbor from the weight function:

$$\frac{y_j}{\left(\sum_{j=1}^k y_j \right)}. \quad \text{Eq. 2.6}$$

This weight function gives more weight to the nearest neighbor and less weight to the farthest neighbor.

5. The residual corresponding to the selected e_t^* neighbor that was computed and saved in Eq. 2.5 above is added to the mean estimate y_t^* . The simulated value for the next time step becomes, $y_t^* + e_t^*$.
6. The process is repeated for each month.

Lall and Sharma (1996) suggested both an objective criteria based on generalized cross validation and a heuristic scheme to select a k, the number of nearest neighbors. They mentioned that the heuristic scheme works well in almost all the cases for $1 \leq p \leq 6$ and $N \geq 100$, and we adopted the same scheme here, where p is the dimension of the model and N is the number of data points in our monthly lag(1) ($p = 1$) model, it is the number of years of data. The heuristic scheme is $k = \sqrt{N}$.

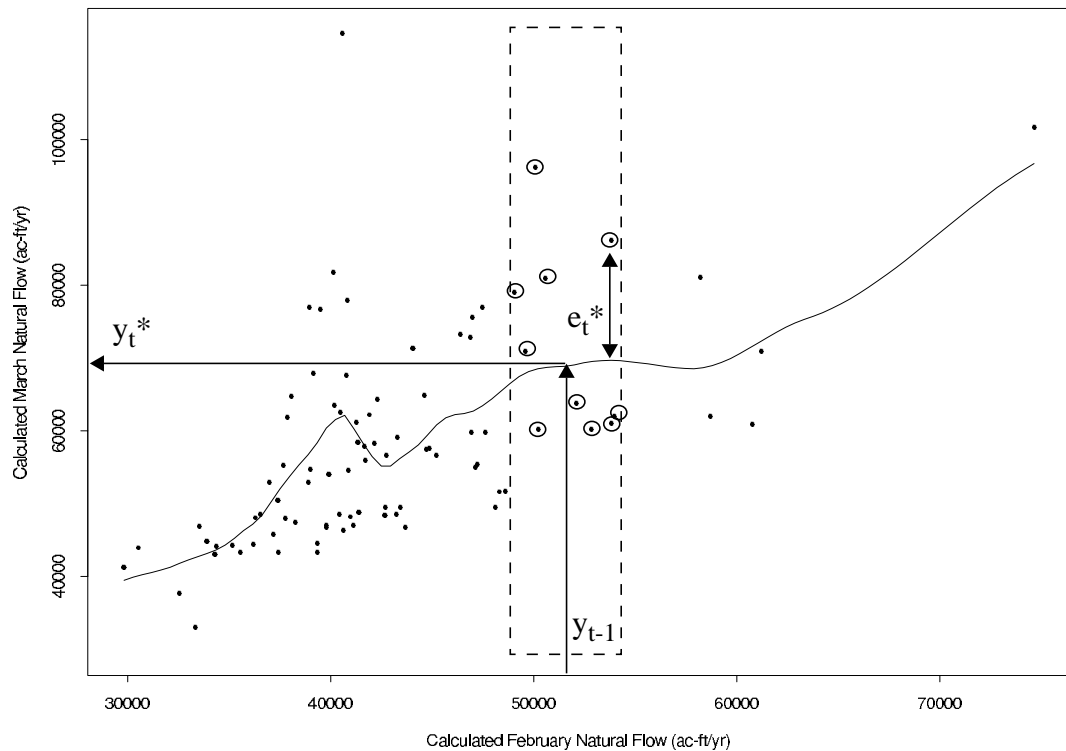


Figure 8: Nonlinear local regression fit to March natural flows dependent on February natural flows is depicted by the solid line ($\alpha = 0.3$). A least square fit is shown with the dotted line.

The modified version of the K-NN model generates values not seen in the historic record and also has the ability to generate extreme values not seen in the history. Further, it retains the basic capability of the traditional K-NN methodology of reproducing PDF structure. The K-NN model of lag(1) will reproduce all the basic statistics, lag(1) autocorrelation, and also the PDF structure, which is a significant improvement over the PAR model.

This nonparametric method is easy to implement, with few parameters to estimate, and makes no assumption about the underlying model. A further advantage of the K-NN framework is that it incorporates multiple variables in a stochastic model more easily than parametric techniques do. This ability allows nonparametric K-NN techniques to be conditioned on additional variables, such as climate variables. Conditioning on climate can allow a model to capture interannual variability caused by cli-

mate. These features of the K-NN framework make the technique a flexible and easier framework to work with than the parametric techniques.

2.7 Model Evaluation

We compared the three models by applying them to the natural streamflow at USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). The monthly flow data was available for the period 1906 to 1995. We generated 100 simulations from the modified K-NN model and 90 simulations from the ISM, each of the same length as the 90-year historical data. We computed a suite of statistics from the simulation and compare them to statistics from the historical data.

2.7.1 Test Ensembles

The three stochastic flow models were applied to natural flows at USGS stream gauge 09072500. Natural flows were calculated from historic gauge records by removing anthropogenic effects such as consumptive use, reservoir regulation, imports, and exports. USBR calculated the monthly natural flows from October 1905 through September 1995. The monthly natural flow mean is 178,000 acre-feet/month, and standard deviation is 217,000. The 90 years of aggregated annual natural flows from water year 1906 to 1995 has a mean of 2,132,000 acre-feet/year and standard deviation of 550,000. High variability in both annual and monthly flows can be seen in Figure 9.

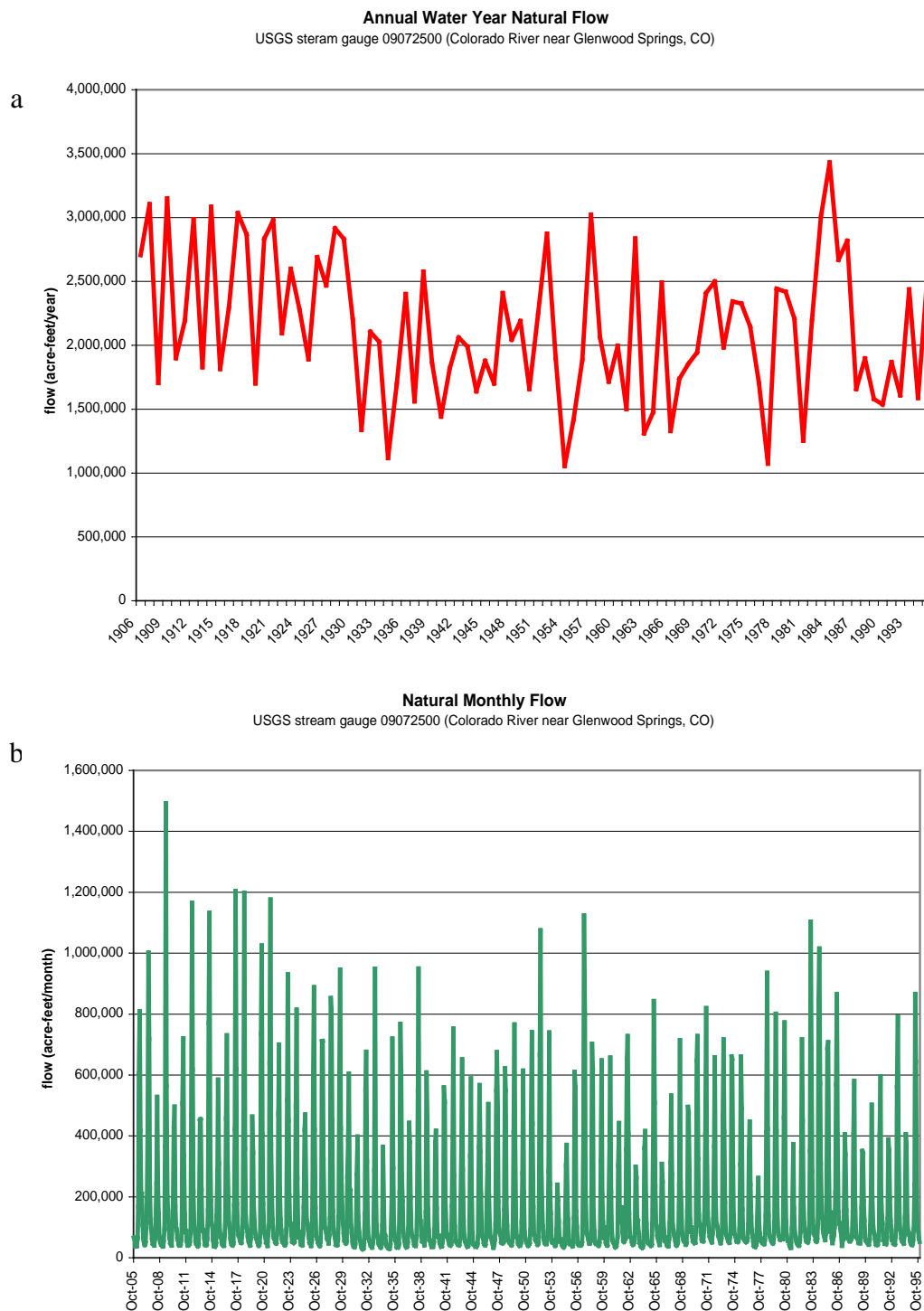


Figure 9: (a) A time series of annual water year natural flow and (b) monthly natural flow from water year 1906 to 1995 for USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). The time series exhibit a high rate of variability both annually and monthly.

2.7.2 Model Evaluation Criteria

The three modeling techniques were each compared to the 1906-1995 historic flow data and then compared to each according to the following criteria.

The first criterion addresses how well the generated ensembles preserved the basic statistics, mean, standard deviation, lag(1) correlation, coefficient of skewness, and maximum and minimum of the historic data. A model that preserved all the statistics well is deemed appropriate to model the original time series.

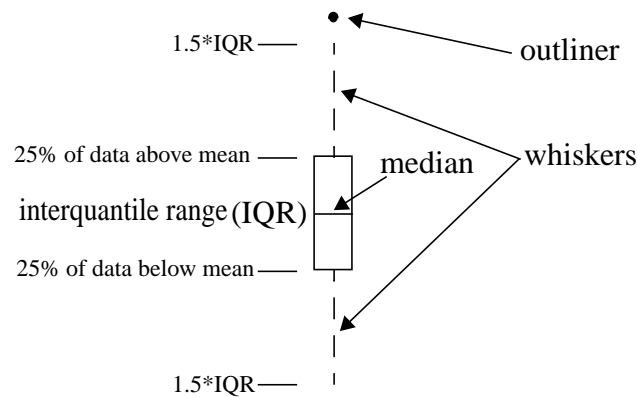
The second criterion addressed how well the modeling techniques preserved the higher order statistics of the original data. Preserving higher order statistics ensures the modeling technique's ability to project variability and be realistic. These abilities can be checked by how well the modeling technique reproduces the probability density function (PDF). We also looked at the bivariate probability density function, which shows how well a technique can preserve the probability of choosing the current month's flow based on the magnitude of flow during the previous month (i.e., conditional PDF).

The last criterion addressed drought, surplus, and storage statistics. Drought, surplus, and storage statistics are important in river basins for reservoir operation. These statistics allow river basin managers to understand the characteristics of extreme events. Neither the parametric nor nonparametric model explicitly models these statistics. The longest drought statistic shows the number of consecutive years the flow is below the median flow. The maximum drought statistic is the maximum volume of water during a drought. Surplus statistics are the opposite. The storage statistics (rescaled range and hurst coefficient) capture long-range memory in the time series. All of these were calculated for each technique.

2.8 Results

The statistics from simulation ensembles and the historic data are shown as boxplots. The boxplots display the interquartile range (IQR) and whiskers extending

to $1.5 * IQR$ for the PDFs of the 100 synthetic natural flow traces. The interquartile range indicates the range for 50 percent of the data around the mean. The horizontal line inside the IQR depicts the median of the data. The whiskers approximate the 5 percent and 95 percent confidence for the traces. Data beyond the whiskers ($1.5 * IQR$) are termed outliers and indicated by a solid circle. An example boxplot is given:



Historic data is shown as a solid circle with a solid line connecting each month.

As expected, all the models preserved well the mean and standard deviation for ISM (Figure 10), PAR (Figure 11), and K-NN (Figure 12). Note that the ISM exactly reproduced the statistics with no variation around the historic statistic, because the ISM uses the historic record *exactly* to develop the ensembles. All the statistics throughout our analysis exhibit this trait. For the remainder of our analysis, we only discuss the comparison with the PAR(1) and K-NN model.

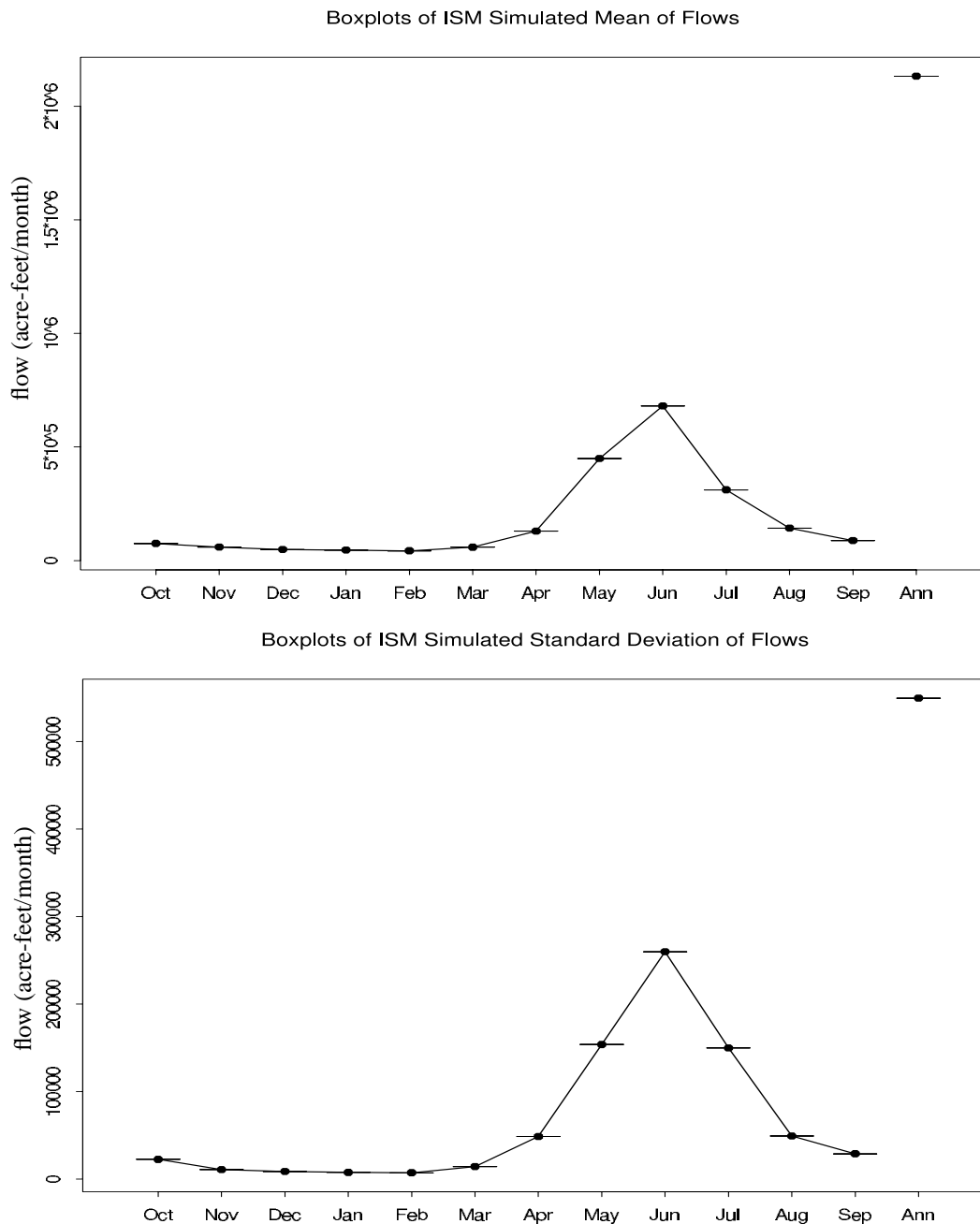


Figure 10: The ISM technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistic with boxplots. Boxplots show the interquartile range of an ensemble in the box and the vertical dashed lines (“whiskers”) approximate the 5% and 95% range of the ensemble. The historic statistic for the fitting period (water year 1906 to 1995) is shown with the solid line and the solid circle. When the historic statistic falls in the box, the statistic is preserved by the technique. The upper graph shows the mean, while the lower graph shows the standard deviation. Both the monthly and annual statistics are preserved.

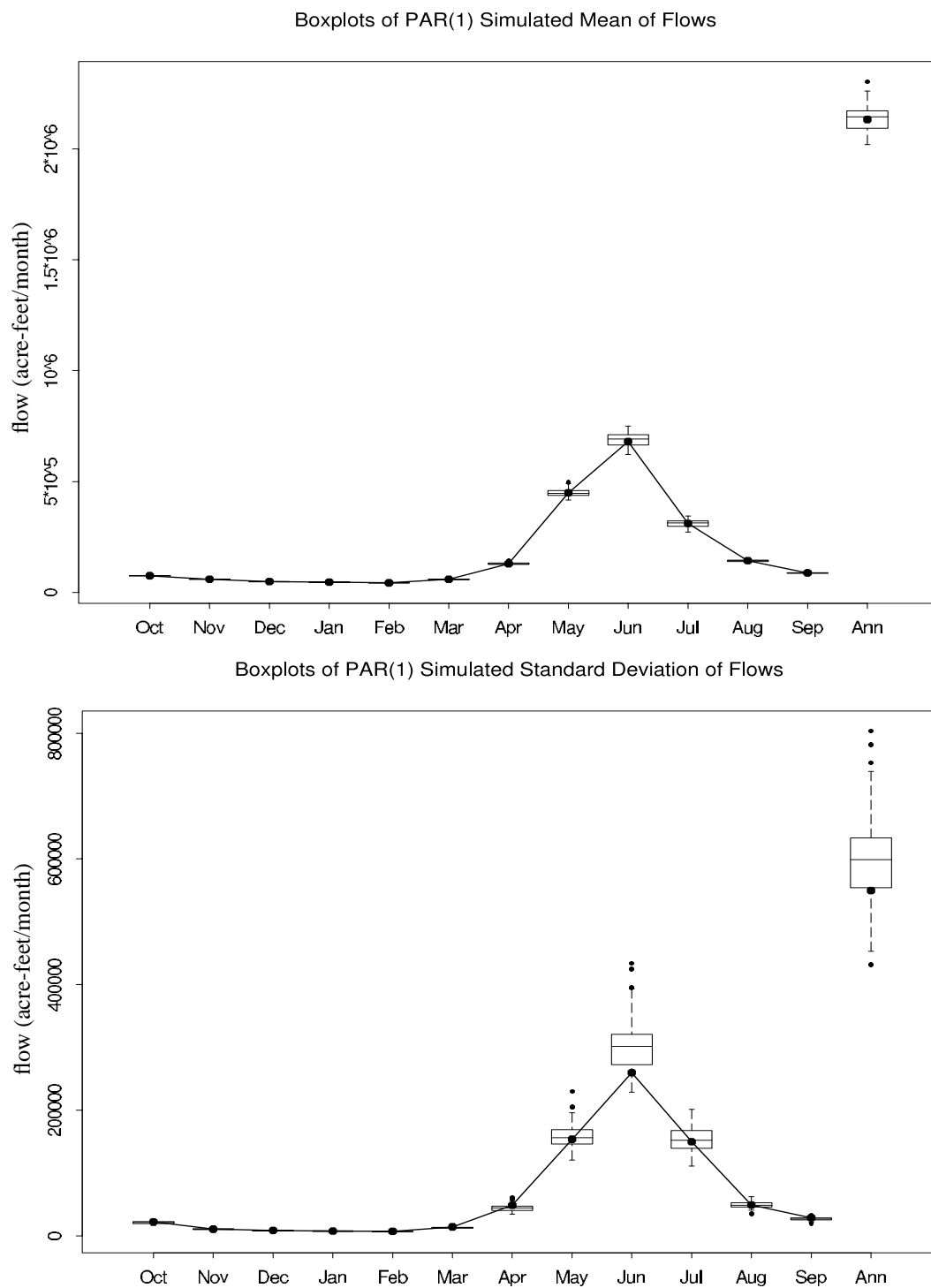


Figure 11: The PAR technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistic with boxplots (see Figure 10 for explanation). The upper graph shows the mean, while the lower graph shows the standard deviation. Both the monthly and annual mean are preserved. The PAR model overestimates the standard deviation monthly and annually.

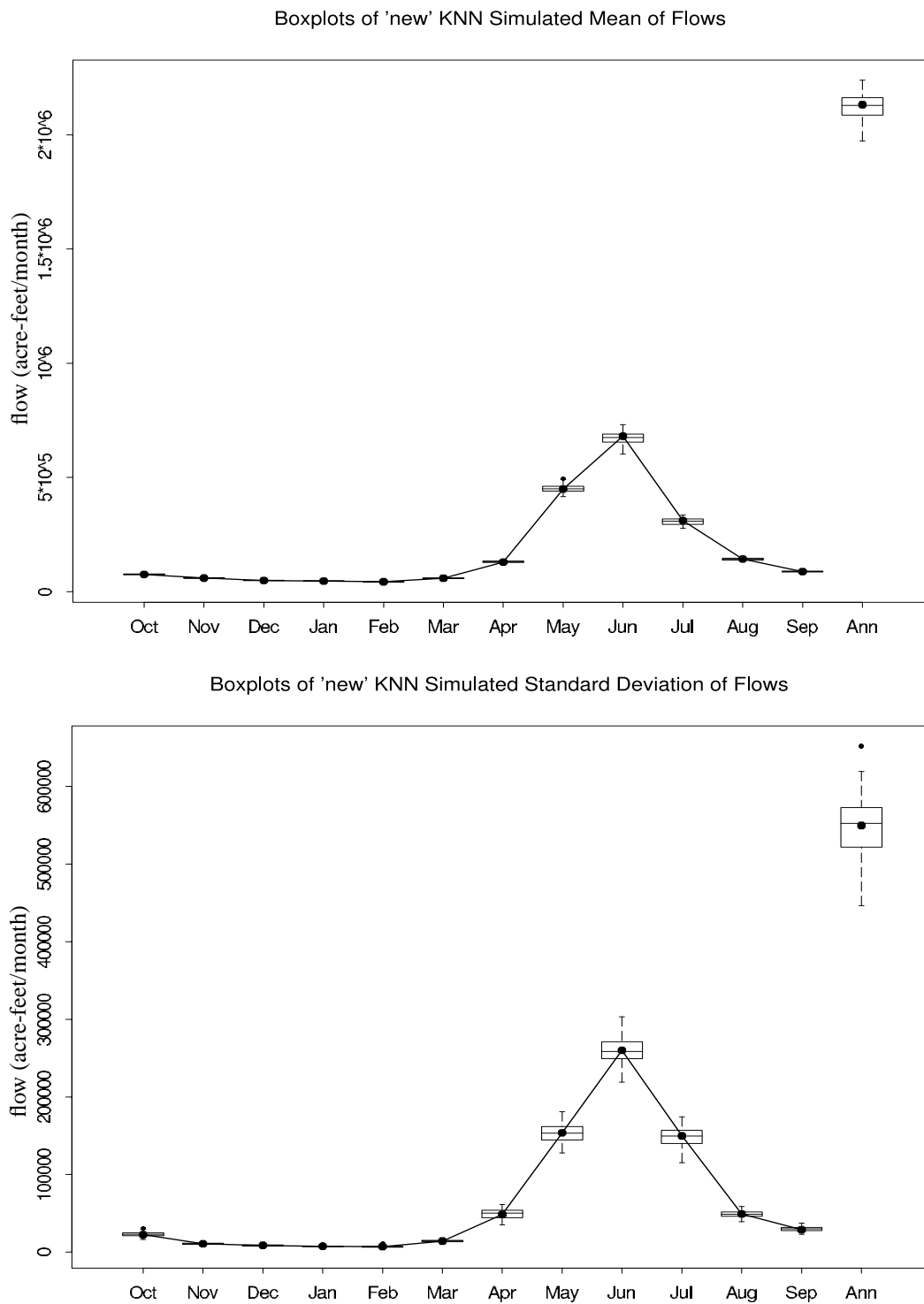


Figure 12: The modified KNN technique generates an ensemble of simulations. The ensemble of simulations are compared to historic statistics with boxplots. (See Figure 10 for explanation.) The upper graph shows the mean, while the lower graph shows the standard deviation. Both the monthly and annual statistics are preserved.

The modified K-NN model and PAR(1) models preserved well the lag(1) correlation for ISM (Figure 13), PAR (Figure 14), and K-NN (Figure 15). PAR(1) did not preserve the skew coefficients because this depends on how well the transformation is able to approximate a Gaussian distribution. The modified K-NN model preserved the coefficient of skewness very well. The modified K-NN models preserved the maximum and the minimum for ISM (Figure 16), PAR (Figure 17), and K-NN (Figure 18) flows well. The parametric PAR(1) model overestimated the maximum and minimum values because PAR(1) can simulate unrealistic values, as mentioned earlier. The modified K-NN generated maximum and minimum values that are not seen in the historical data; this is a significant aspect of the modified K-NN.

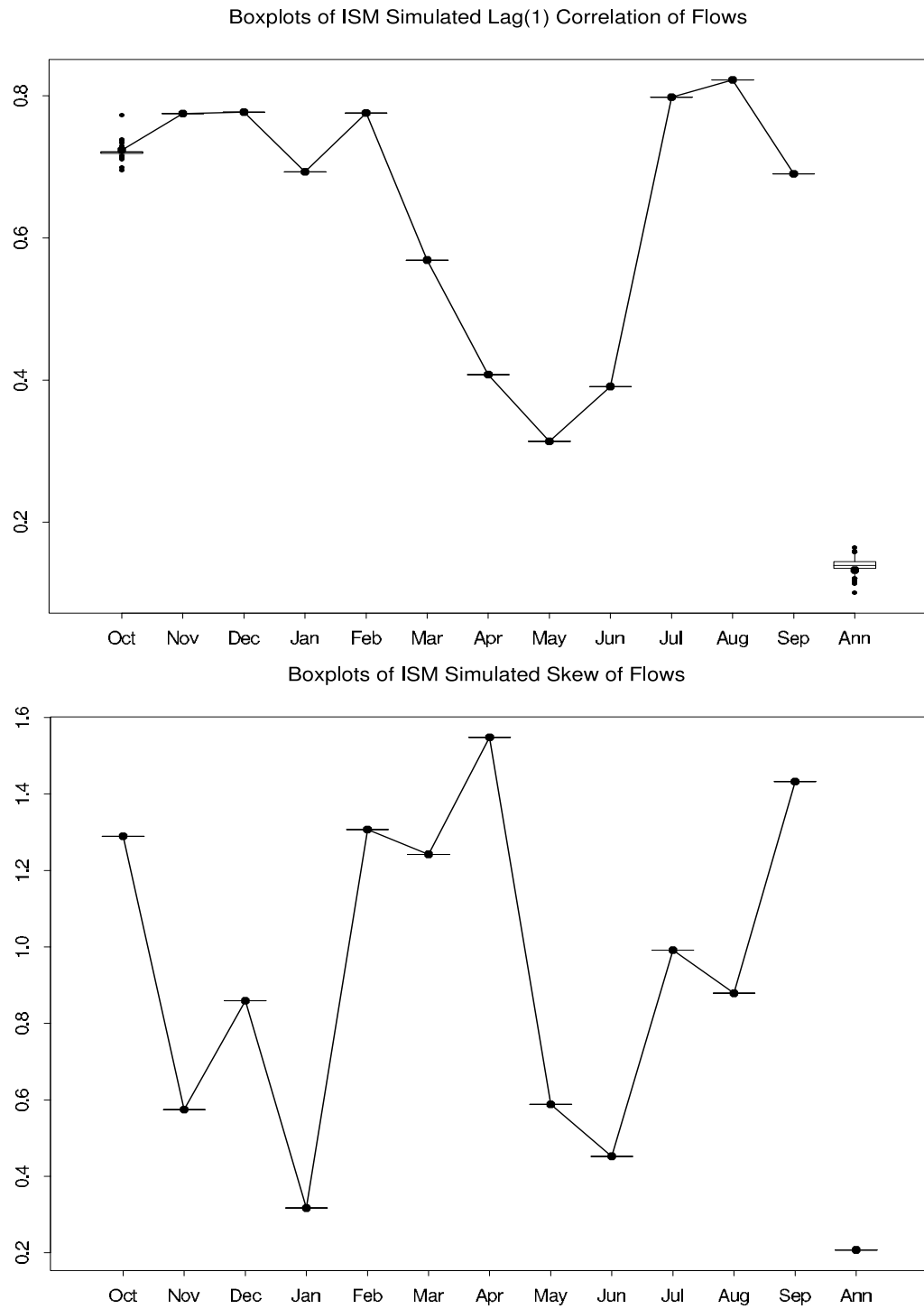


Figure 13: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the ISM. The lower graph shows the coefficient of skewness. ISM preserved both statistics annually and monthly.

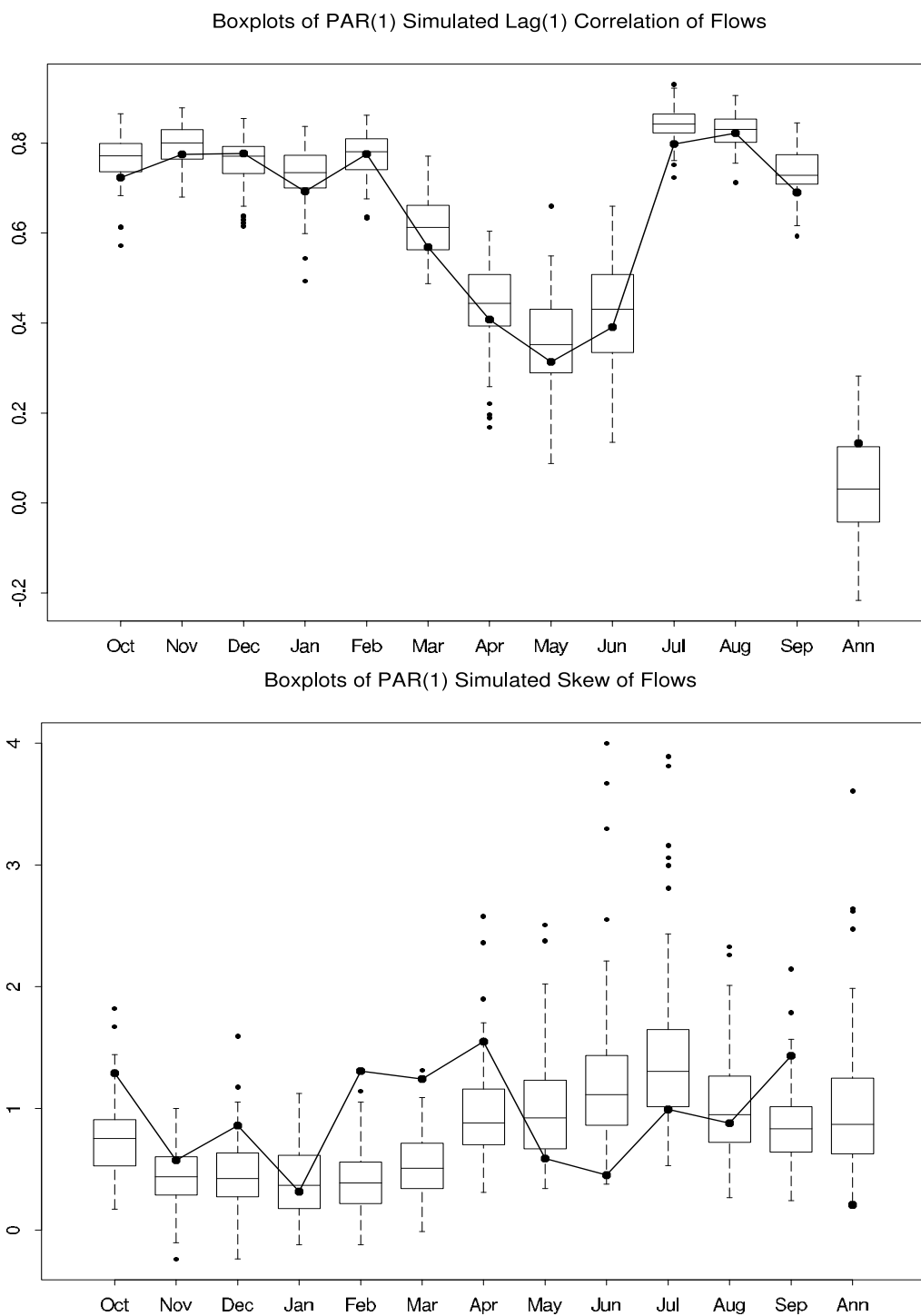


Figure 14: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the PAR model. The lower graph shows the coefficient of skewness. The lag(1) correlation is preserved monthly for most months in the interquartile range. The coefficient of skewness is not preserved in the interquartile range for either the monthly or annual time step.

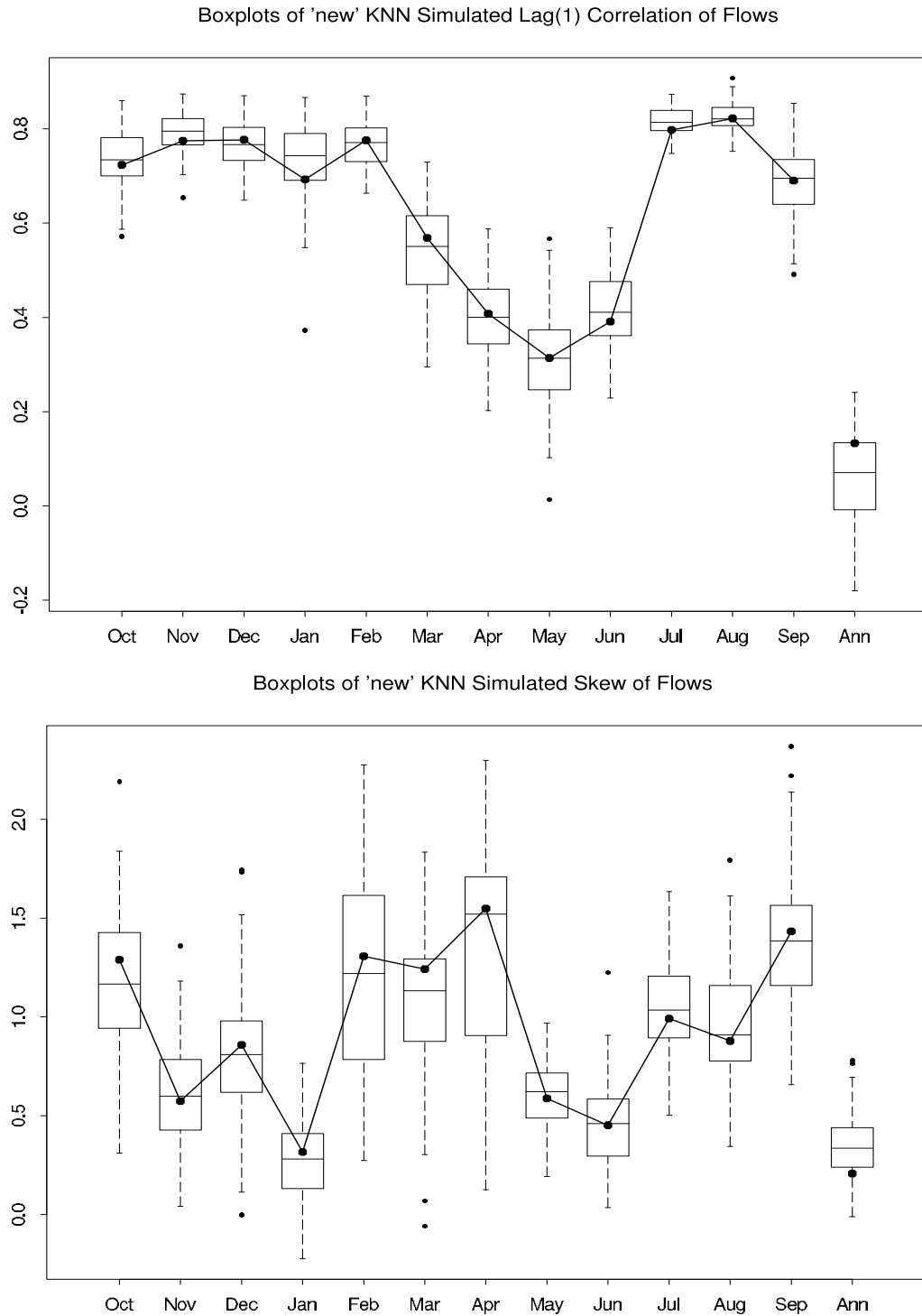


Figure 15: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) correlation of the K-NN. The lower graph shows the coefficient of skewness. K-NN preserves both statistics monthly. The annual time step is preserved on the fringes of the interquartile range.

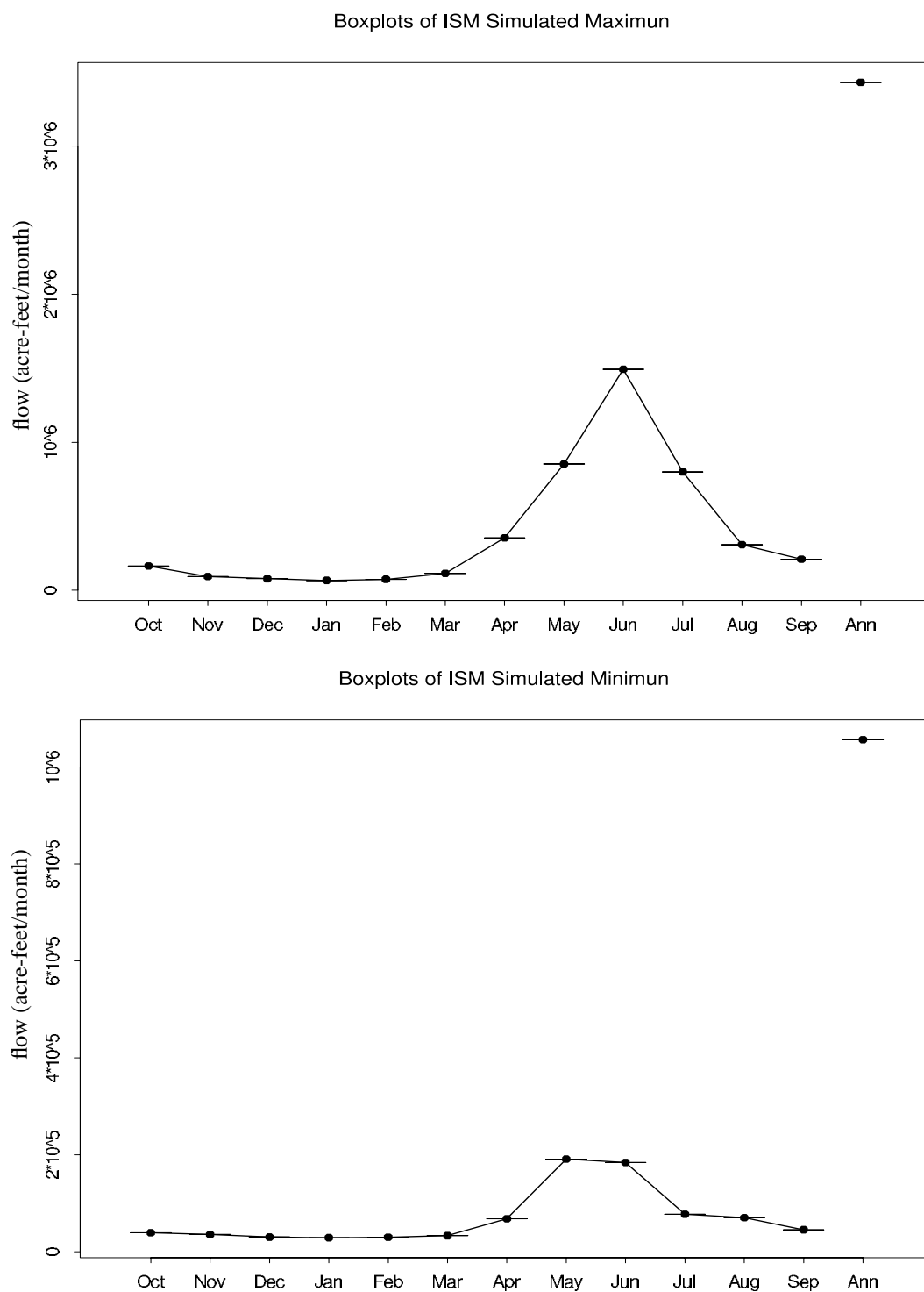


Figure 16: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the ISM. The lower graph shows the minimum value. The ISM exactly reproduces both statistics.

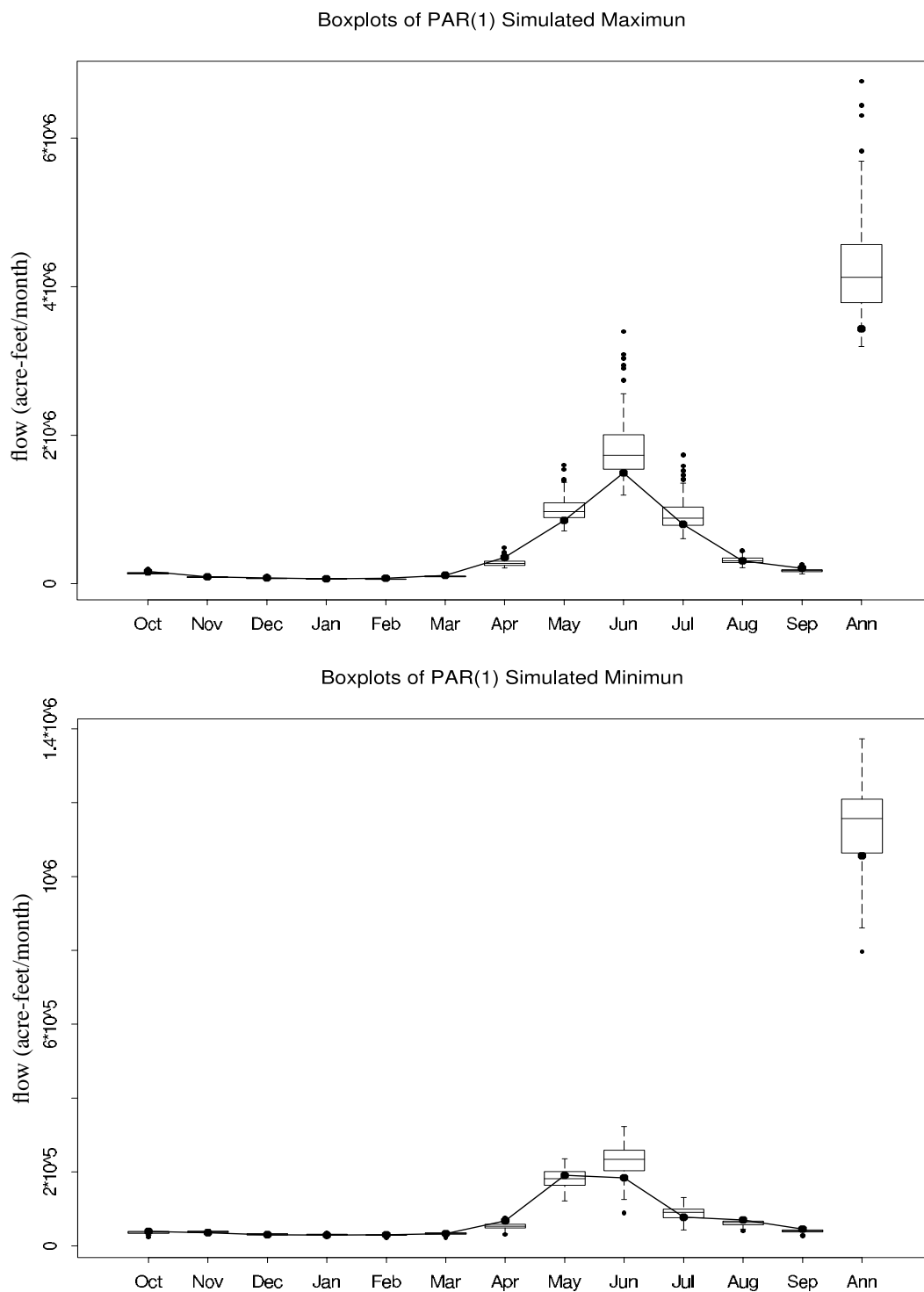


Figure 17: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the PAR(1). The lower graph shows the minimum value. The PAR(1) technique tends to overestimate the maximum and minimum both monthly and annually.

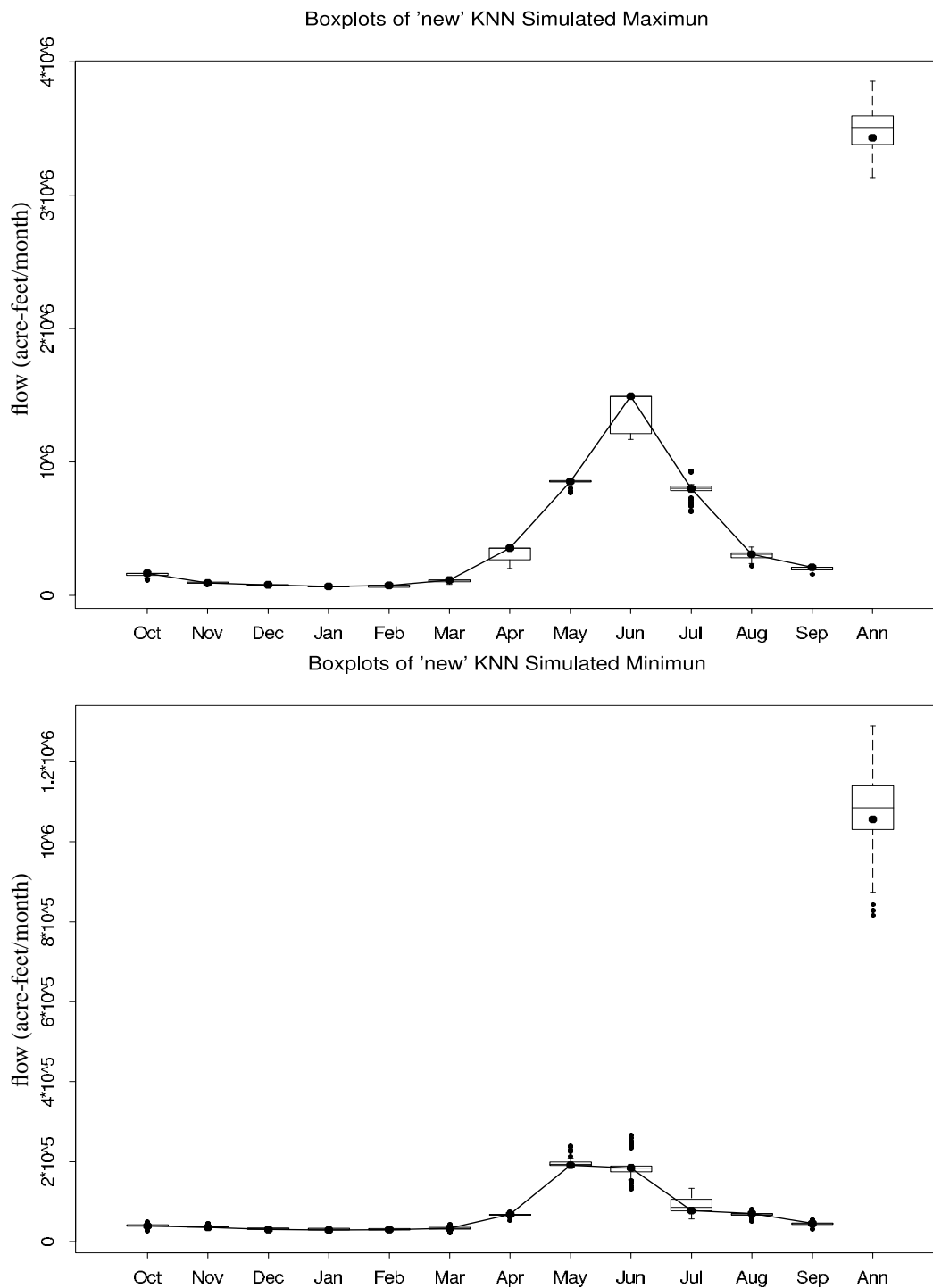


Figure 18: The upper graph shows boxplots (see Figure 10 for explanation) of the lag(1) maximum value of the K-NN. The lower graph shows the minimum value. The K-NN technique slightly underestimates certain months maximum but preserves the other months and annual statistics well.

The higher order statistics continued to show the strengths of nonparametric techniques. Figures 19- 21 show select PDFs for the month of January. In the upper graph, a Gaussian distribution approximates the historical data quite well. The simulations from the PAR(1) and K-NN reproduced this fairly well. The month of February for ISM (Figure 19), PAR (Figure 20), and K-NN (Figure 21), in the lower graph, exhibits a skewed distribution. The parametric PAR(1) tended to reproduce a Gaussian equivalent of the original probability density function, while the nonparametric K-NN model was able to preserve the non-Gaussian structure. Similar observations can be seen for the PDF of September flows (upper graph) and annual flows (lower graph) for ISM (Figure 22), PAR (Figure 23), and K-NN (Figure 24).

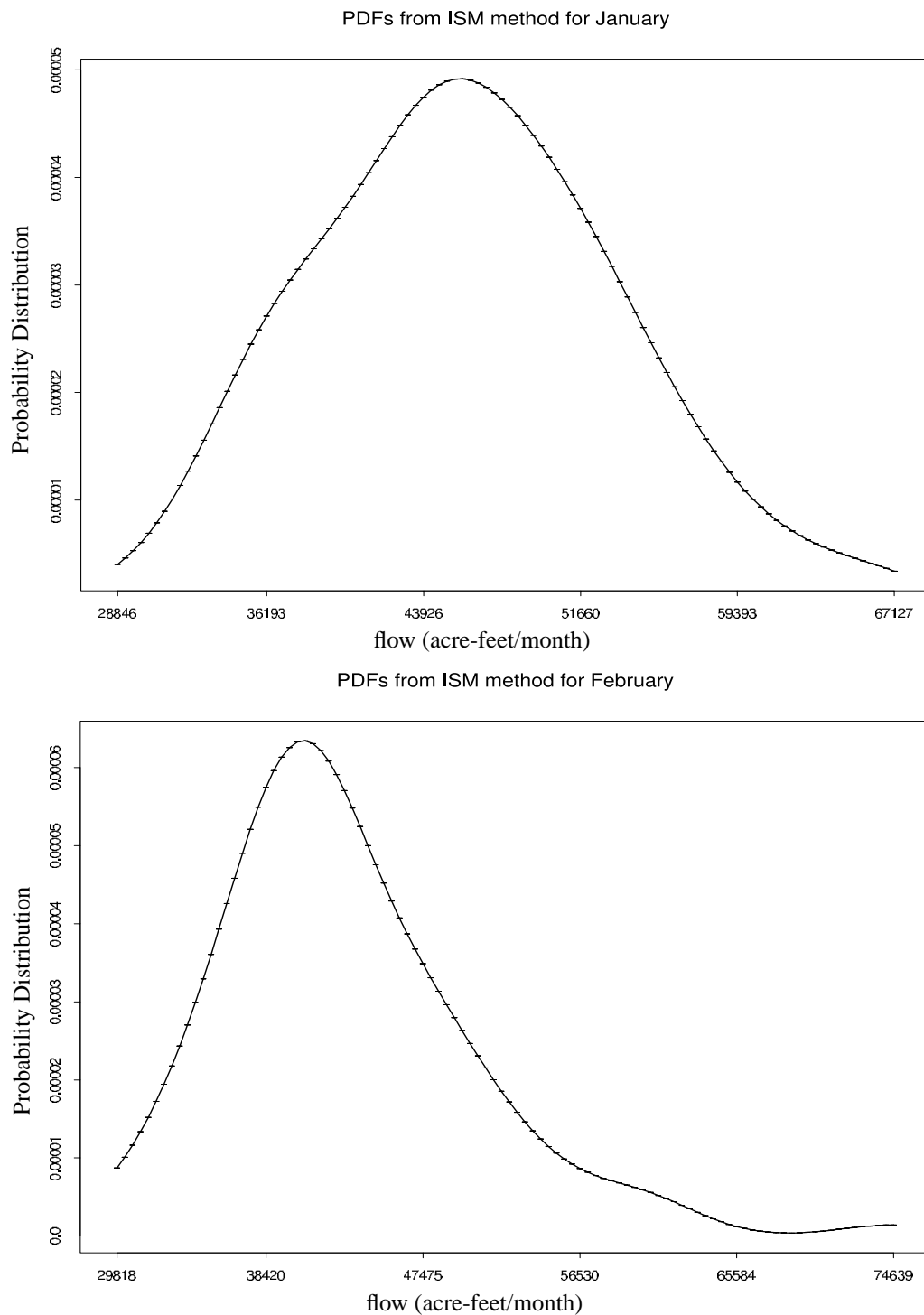


Figure 19: The upper graph shows the PDFs for ISM for January, which approximates a Gaussian distribution. The ISM can only reproduce the historic probability density function because every simulation includes all the historic time series; however each simulation is sequentially shifted. The lower graph shows the PDFs for February, which is skewed towards the low flows.

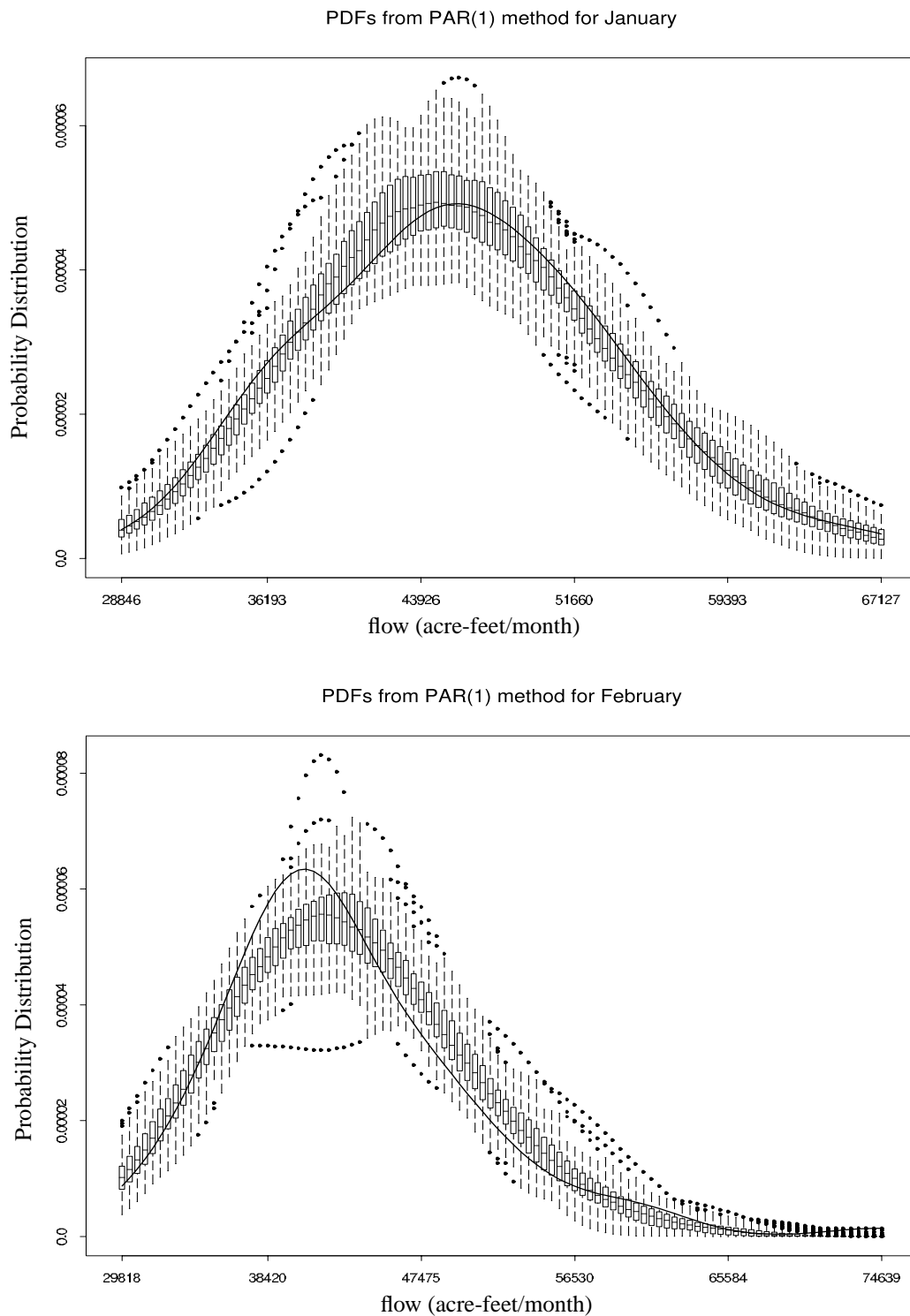


Figure 20: The upper graph shows the PDFs for PAR(1) for January, which approximates a Gaussian distribution. The lower graph shows the PDFs for February, which is skewed towards the low flows. This parametric technique cannot preserve the skewed distribution in the interquartile range of the simulations.

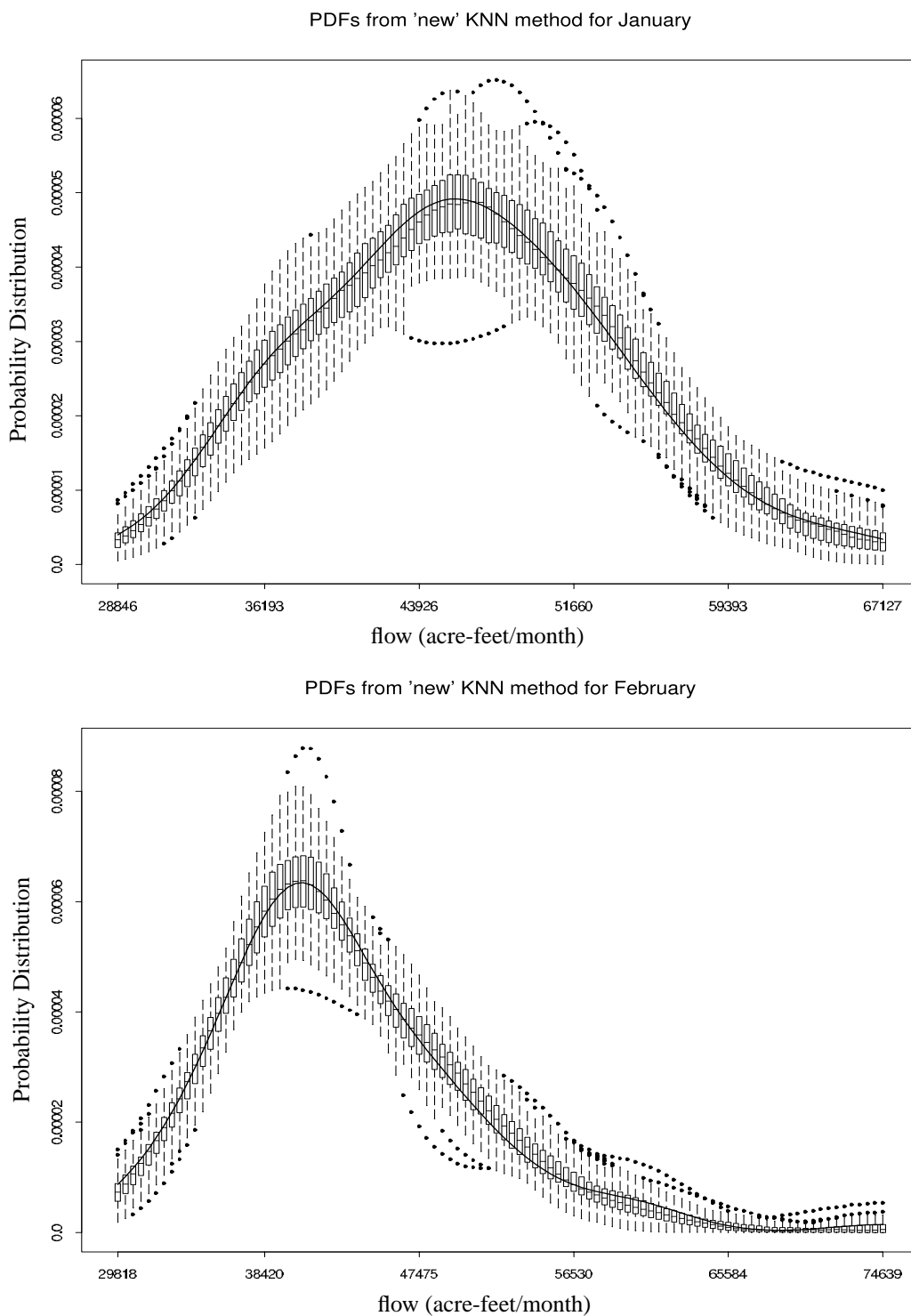


Figure 21: The upper graph shows the PDFs for K-NN for January, which approximates a Gaussian distribution. The lower graph shows the PDFs for February, which is skewed towards the low flows. Both distributions are preserved by the K-NN technique.

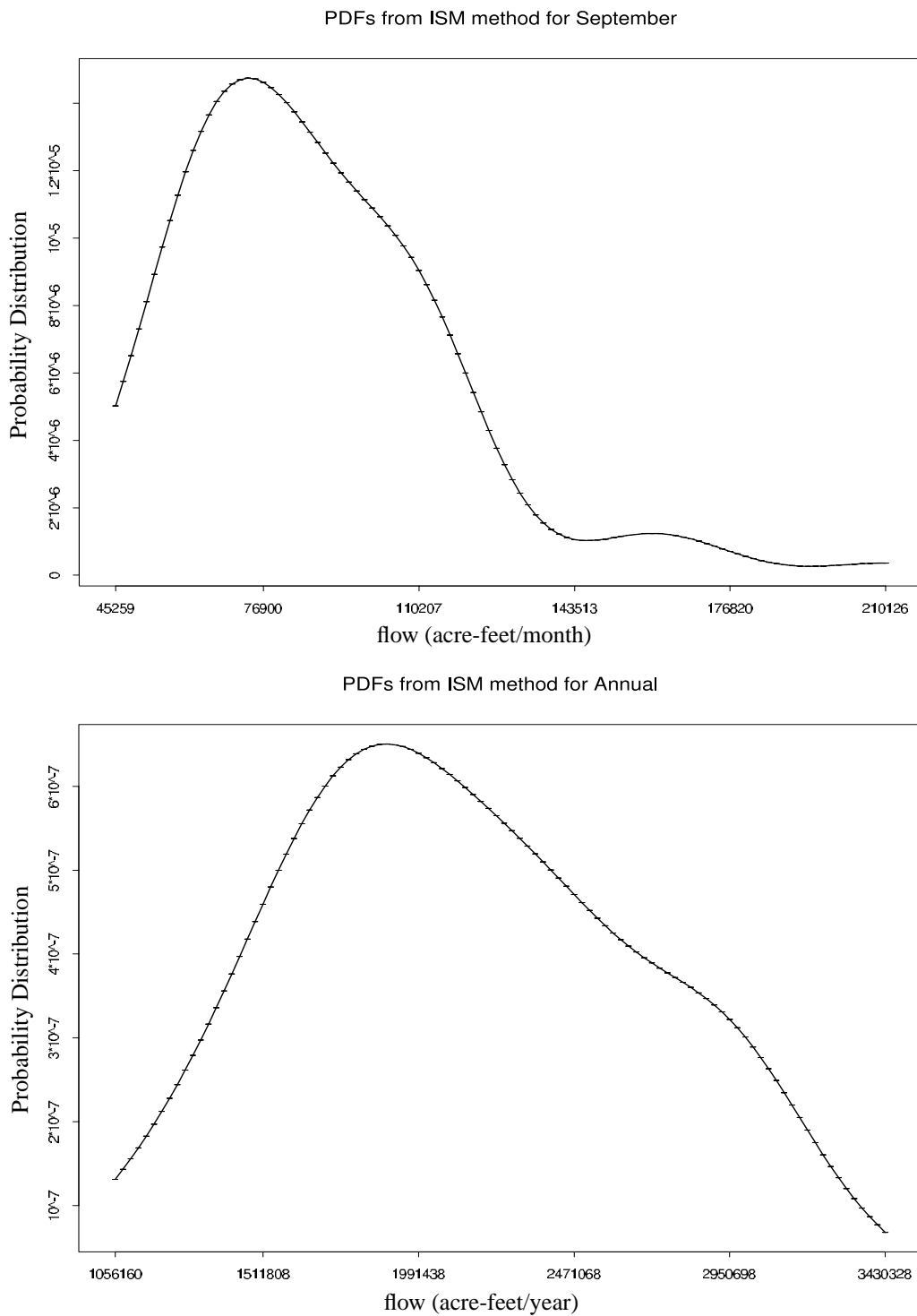


Figure 22: The upper graph shows the PDFs for ISM for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The ISM exactly recreates the PDF, again as stated in Figure 19.

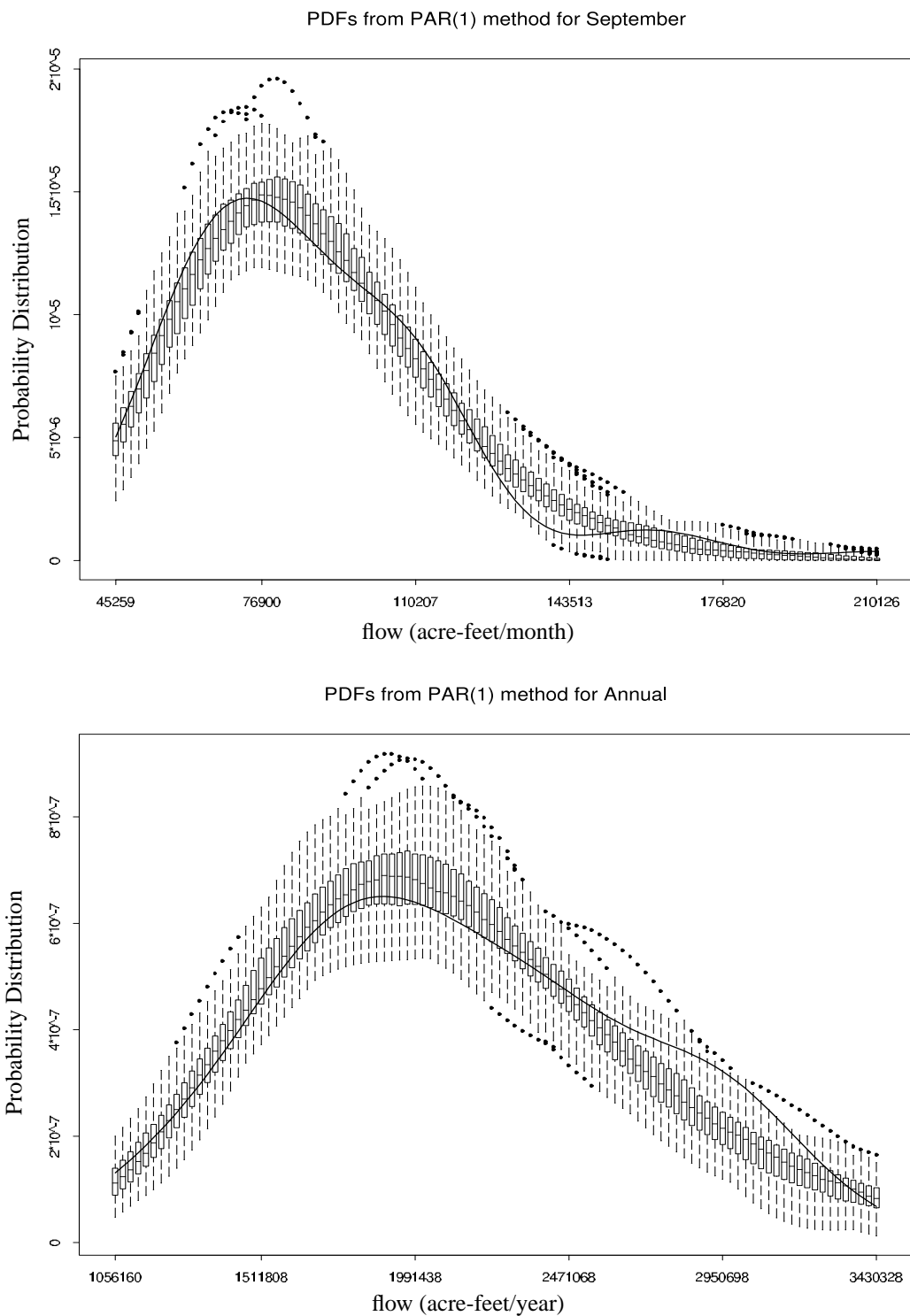


Figure 23: The upper graph shows the PDFs for PAR for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The PAR technique can only approximate the annual PDF. Annual flow is not directly modeled.

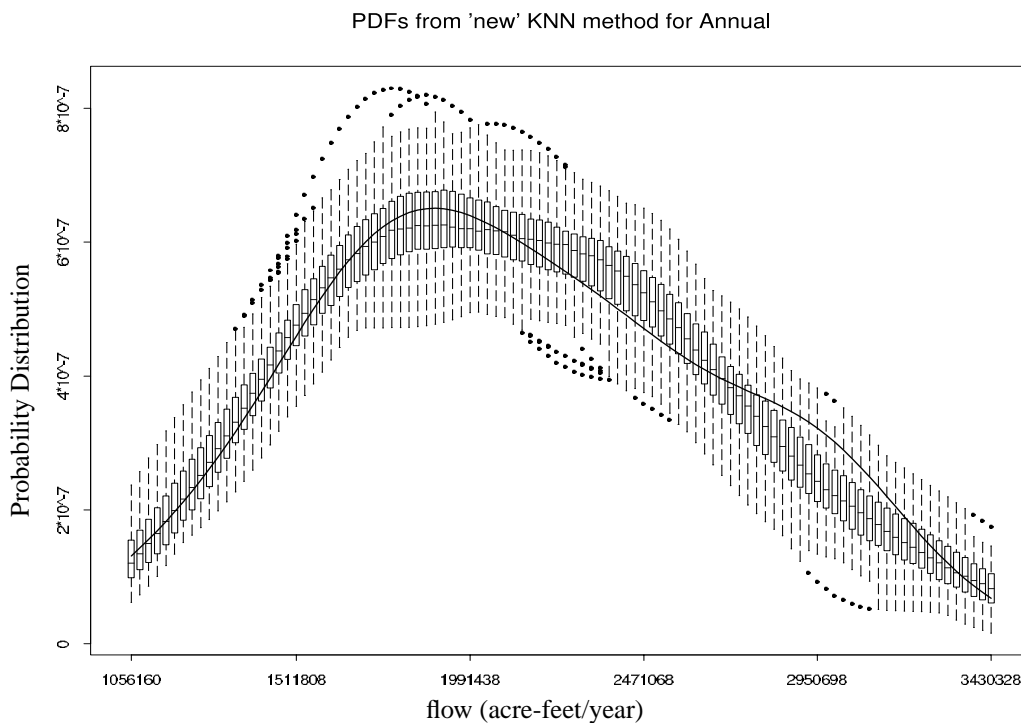
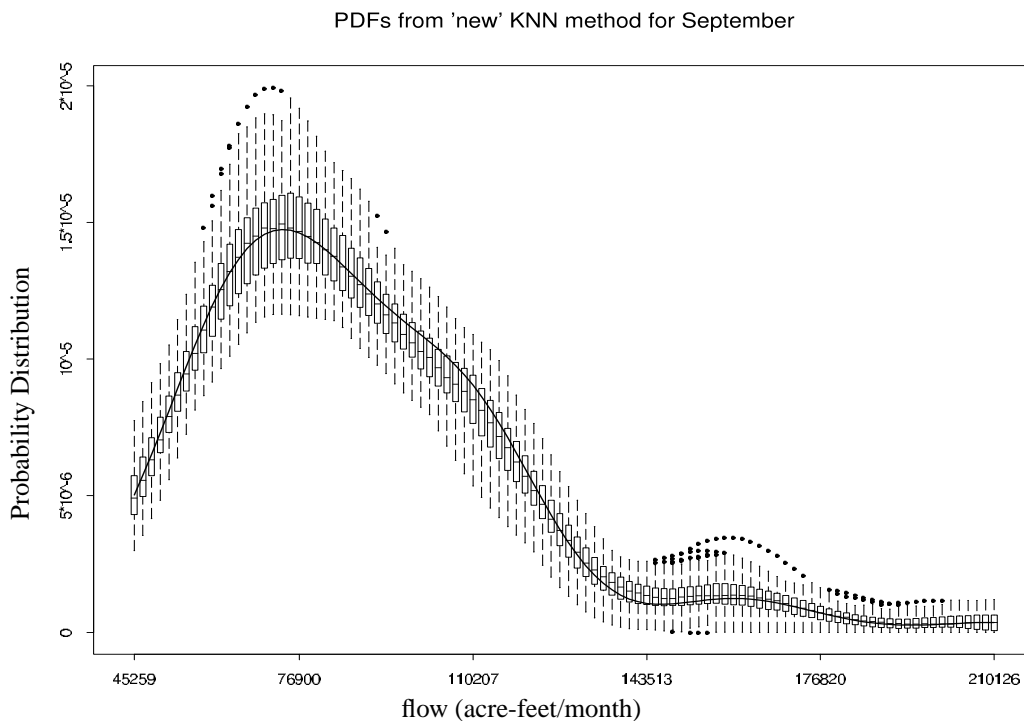


Figure 24: The upper graph shows the PDFs for K-NN for September, which is skewed to the lower flows with a long tail in the higher flows. The lower graph shows the PDFs for annual flows. The K-NN technique can preserve the skewness and tail behavior best, but can only approximate the annual PDF, although the K-NN technique can reproduce the frequency of higher flows better than the parametric technique.

The bivariate PDF (Figure 25) was computed for the months May and June from the historic time series and for one of the simulations from the PAR(1) and K-NN models. The historic time series show nonlinear features, indicating a non-Gaussian relationship in the transition of flow between months May and June. The PAR(1) model was only able to recreate a Gaussian bell-shaped distribution, as expected, while the K-NN model preserved the nonlinear features seen in the bivariate probability density function of the historic values.

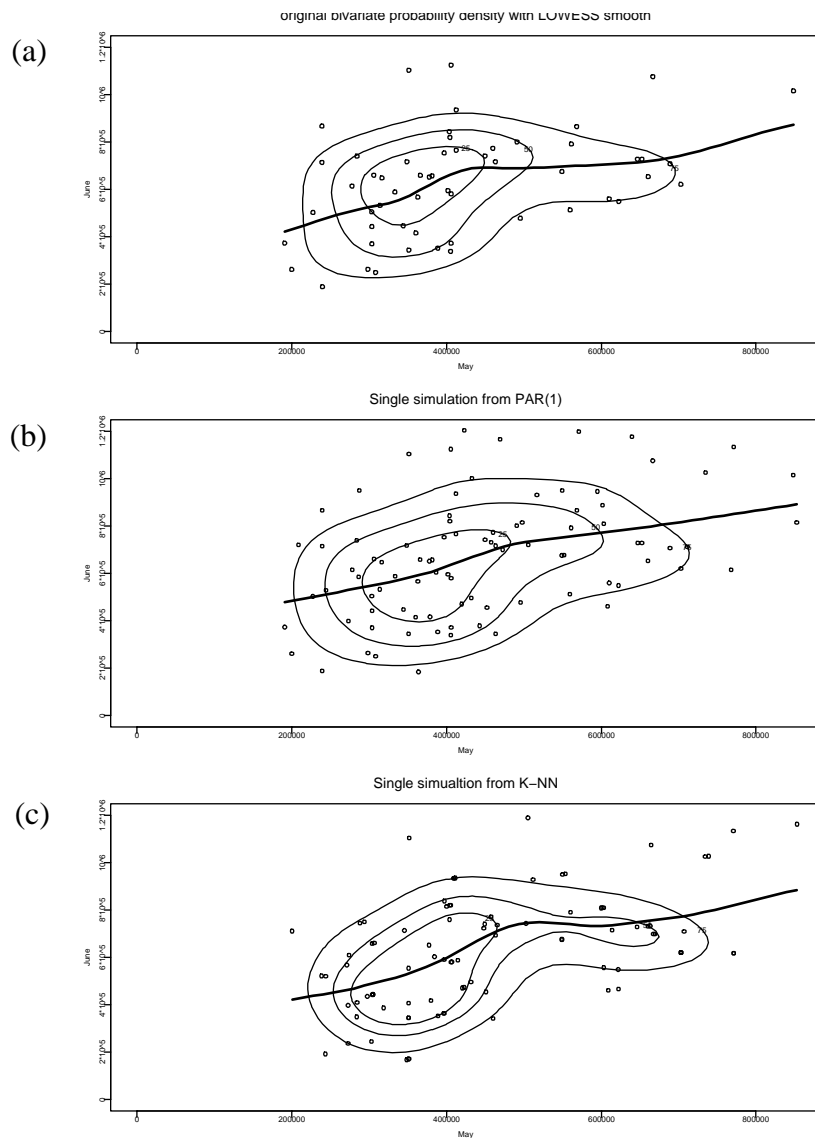


Figure 25: Bivariate probability density function for May/June estimated using an S-plus library from (a) historic values, and a single realization from 100 simulations using the (b) PAR(1), and (c) K-NN model. The nonparametric technique can preserve the historic bivariate PDF's non-Gaussian features, while the parametric technique can only approximate a bell-shaped distribution.

The final statistics compared the drought, surplus, and storage statistics. Both the PAR(1) and K-NN preserved the longest drought within the upper interquartile range. The parametric model preserved the maximum drought better. The nonparametric model tended to underestimate the maximum drought. This underestimation means

the nonparametric model generated lower drought values than the historic record, indicating that the technique produced more conservative drought estimates than the parametric model. Both the K-NN and PAR(1) models preserved the surplus statistics (longest surplus and maximum surplus) well, ISM (Figure 26), PAR (Figure 27), and K-NN (Figure 28). Both the parametric and nonparametric techniques overestimated the storage statistic (rescaled range and hurst coefficient). The nonparametric model preserved the storage statistics within the whiskers, while the parametric model could not preserve either storage statistic within the whiskers.

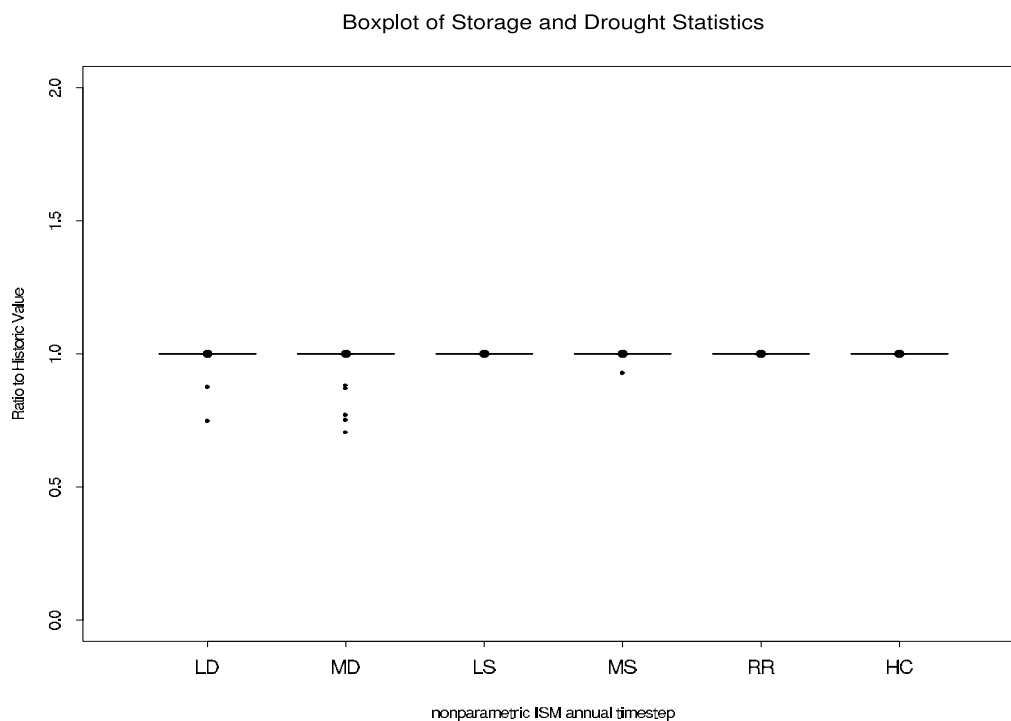


Figure 26: This figure shows boxplots of storage, surplus, and drought statistics for the nonparametric ISM model's annual values. The drought statistics include LD (longest drought), MD (maximum drought). The surplus statistics include LS (longest surplus), MS (maximum surplus). All the values are exactly preserved because the entire historic record is resampled for each of the 100 simulations.

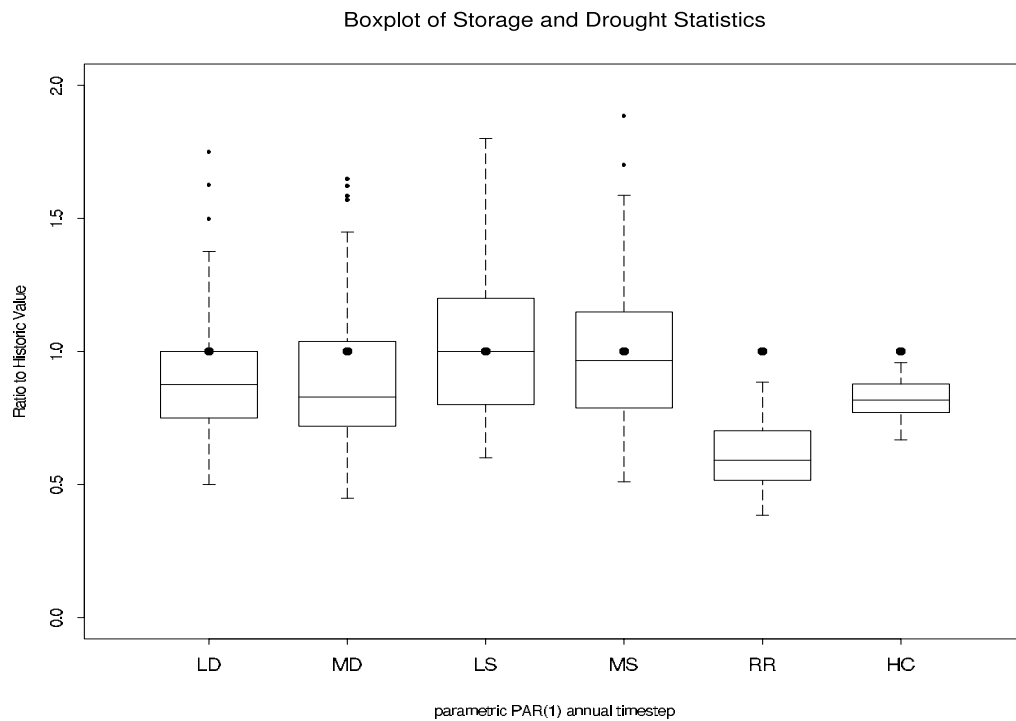


Figure 27: This figure shows boxplots of storage, surplus, and drought statistics for the parametric PAR(1) model's annual values (see Figure 26 for explanation). The parametric model preserves the drought statistics are preserved slightly better. The parametric PAR(1) technique cannot preserve the storage statistic within the whis-

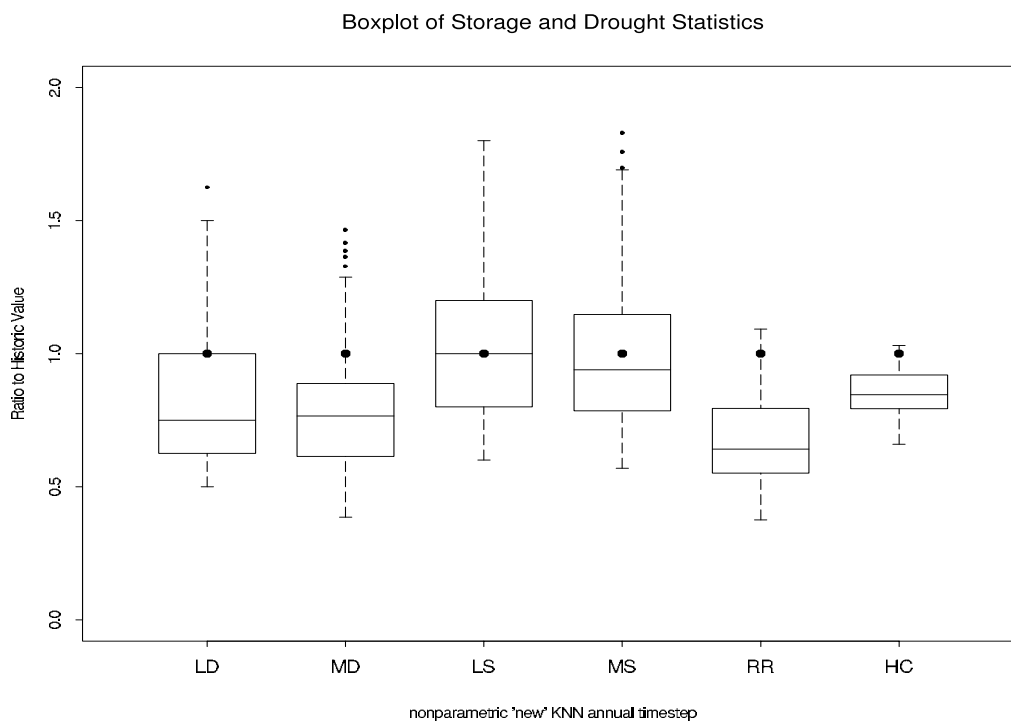


Figure 28: This figure shows boxplots of storage, surplus, and drought statistics for the nonparametric K-NN models annual values (see Figure 26 for explanation). The model preserves the storage statistic within the whiskers.

2.9 Climate Link

Large-scale climate in the western United States modulates streamflows (Cayan and Webb, 1992). Relationships between large-scale climate and streamflow vary in strength throughout the Colorado River basin. Strong relationships have been documented in the southern basin, while weaker relationships have been identified in the upper Colorado River mainstem (Cayan et al., 1999; McCabe and Dettinger, 1999). We explored establishing links between large-scale climate patterns and monthly streamflow in the upper Colorado River mainstem.

Streamflows generated by the K-NN technique can be conditioned on climate variables. To condition the flows on climate, we needed to find an index that exhibits a correlation with the streamflows. To develop such an index, we explored the relation-

ship between natural flow and climate indicators in the upper Colorado River mainstem.

We performed the analysis on annual natural flow at USGS gauge 090163500, Colorado River at Colorado/Utah state line from 1909 to 1996, provided by the Colorado Water Conservation Board, and three snow water equivalent gauges for the region upstream of USGS gauge 09163500, which have a correlation of 0.72 with the USGS stream gauge. Both the snow water equivalent and stream gauge are first correlated with an ENSO index. We found that the correlations not strong enough to directly condition the streamflow with the ENSO index, which is not surprising because previous research has shown the upper Colorado River mainstem has weak correlations with ENSO (Cayan et al., 1999).

Another technique to develop an index first generates composites from the streamflow data. Composites explore climate patterns for various flow regimes, such as high and low flows. Exploring various sections of flow can uncover nonlinear relationships between flow and large-scale climate. We looked at the average climate pattern for three different flow regimes: periods of high-, normal-, and low-flow. The separate flow regimes are developed using salinity concentration that is modeled at an upstream gauge (0909550: Colorado River near Cameo). The flow regimes represent thresholds that can be chosen by many methods. The regimes are usually chosen to try to find unique patterns in climate for particular flows. Our thresholds were based on the behavior of the CRSS salinity model. The CRSS model overpredicted salt concentration during the low-flow years 1977 and 1981. During the high-flow years of 1983 to 1989 and 1995, salt mass was underpredicted when compared to normal years. At this gauge, we determined the low-flow period as flows below 3,500,000 ac-ft/year, normal-flow between 3,500,000 ac-ft/year and 7,500,000 ac-ft/year, and high-flows above 7,500,000 ac-ft/year. This criteria allowed us to categorizes flows that are over or underpredicting salt relative to the average flow regime.

We were interested in looking at the climate patterns for each flow regime and determining if there is a significant difference in the average climate pattern for each flow regime. We examined multiple indicators for climate to verify that they are all consistent with the pattern we uncover. We examined the following climate indicators: sea surface temperature, sea level pressure, geopotential heights at 500 mb level, vector winds at 1000 mb level, outgoing long wave radiation, velocity potential, and divergence. Many of these indicators are complementary, and we used them to verify the accuracy of our analysis. We performed this verification by finding the average climate pattern for the high-, low-, and high-minus-low-flow regimes. We identified a particular climate pattern for each flow regime. We used these patterns in the next step, in which the patterns helped to generate natural flow sequences.

From this analysis, we were able to identify climate patterns that typically produce the high- and low-flow regimes. To incorporate climate information in our model, we needed to develop a relationship between a climate indicator and natural flow. The geopotential height at the 500mb level climate indicator showed the strongest relationship to natural streamflow. Figure 29 show the composite plots for the geopotential height at the 500mb level. The upper plots shows the composite for the low-flow years. A high-pressure system over the western United States can be seen, which tends to bring in cold dry air over the upper Colorado River basin and, consequently, causes a reduced snow cover and lower flows. The lower plot shows the high-flow years in which the pattern is weaker than that seen in the low-flow years; this also highlights the asymmetry of the relationship. Figure 30 shows the composite of high-minus-low-flow years (difference of the two plots in Figure 29). The signal is highlighted in the difference plot.

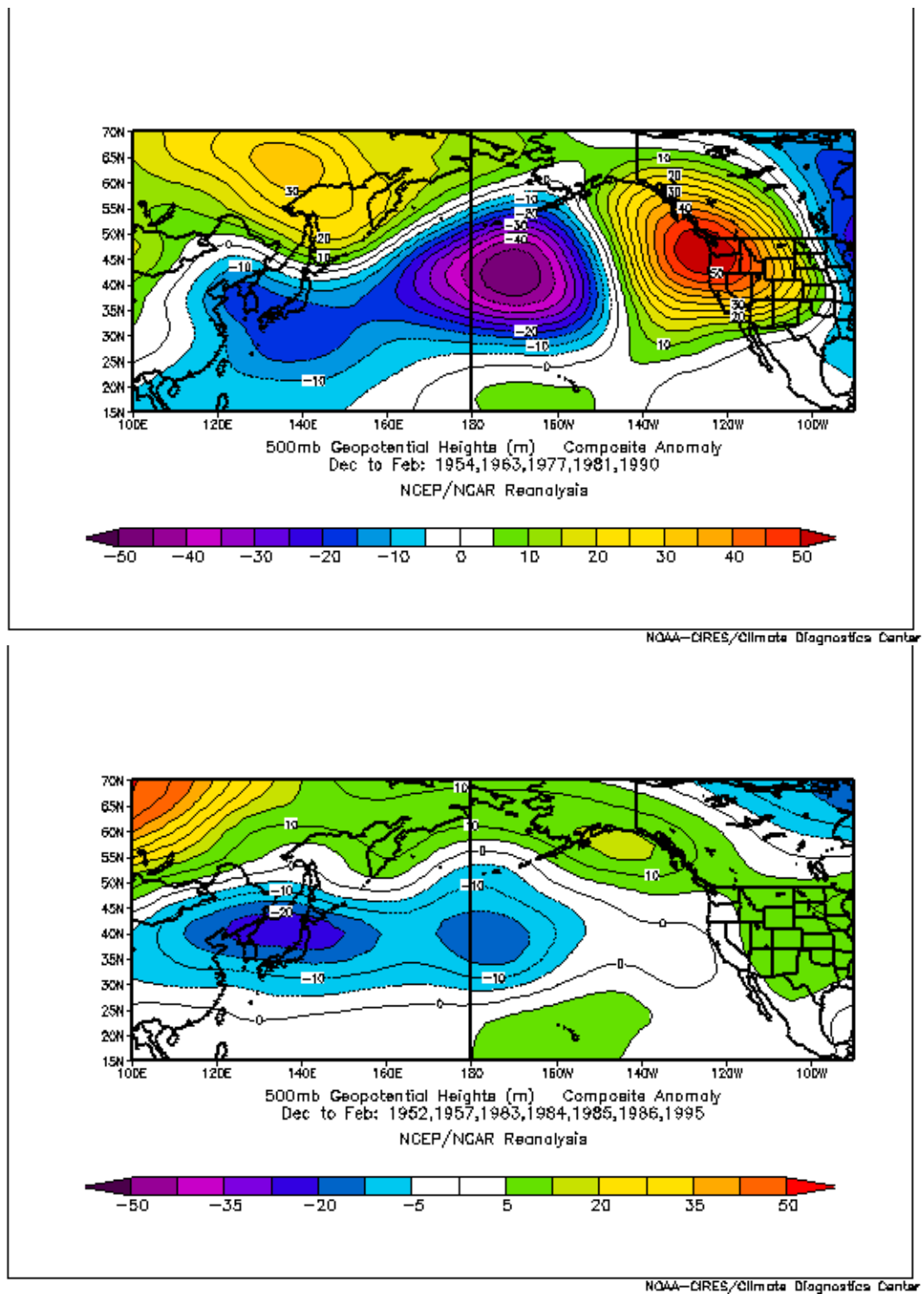


Figure 29: Both plots are anomaly composites for the geopotential height at 500 mb. The upper plot displays the composite of low-flow years. The lower plot displays the composite of high-flow years. The low-flow years show strong high pressure over the western United States and strong low-pressure over the northern Pacific.

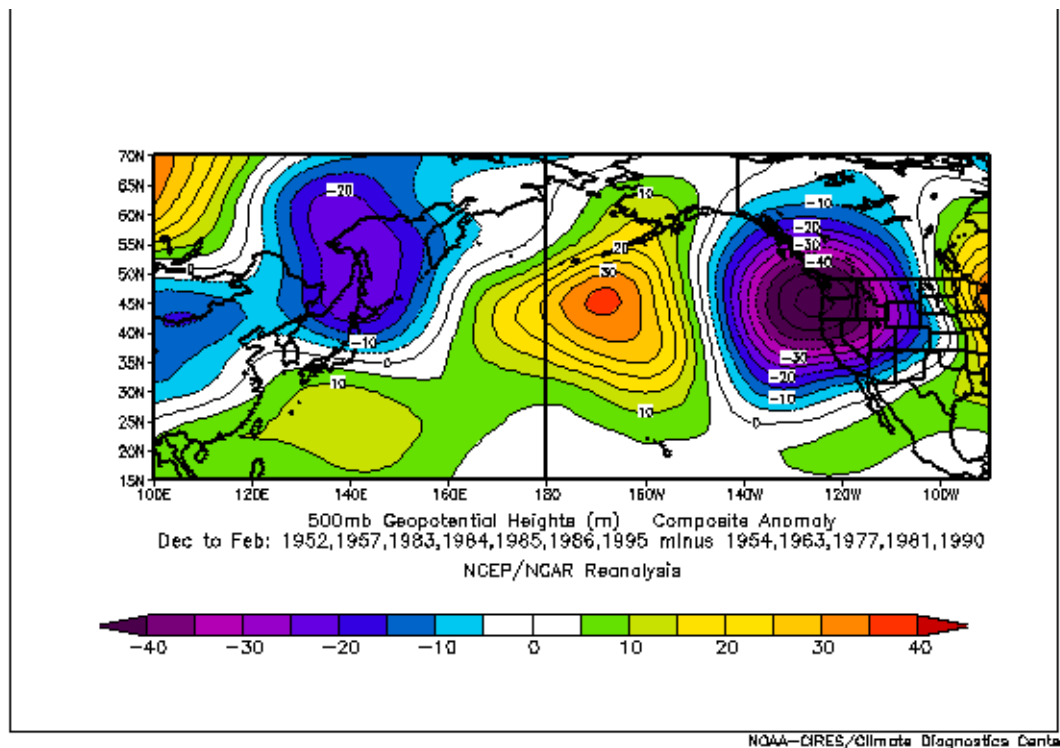


Figure 30: This figure shows an anomaly composite for high-minus-low-flow years at the geopotential height at 500 mb. The plot shows a strong negative difference over the western United States and positive difference of the northern Pacific.

To use this data in developing a climate index, we performed a principal component (PC) analysis on the high-minus-low geopotential height and examine the results. We found the spatial structure for the first PC accounts for 33 percent of the variance and matched the pattern that is found in the geopotential heights relation to flow. A scatterplot of the natural flow dependent on the first PC (PC(1)) revealed a weak relationship during low and high flows. Unfortunately, we found the relationship is not strong enough to condition the flow on climate. Further research needs to be performed to find an index that provides a stronger relationship between the index and the natural flow in the upper Colorado River mainstem.

2.10 Discussion and Conclusions

We compared three stochastic hydrology modeling techniques, one of which is ISM, USBR's currently used technique. We looked at how well each technique preserved basic statistics and the PDF for the time series they modeled.

We compared the ISM to a PAR(1) and modified KNN technique, uncovering the limitations of the ISM and PAR(1) techniques. The ISM is limited because it cannot generate synthetic hydrologies or sequences that did not occur in the past record. Consequently, droughts or surpluses of lengths different from those that have occurred in the past record cannot be created.

The main disadvantages of the PAR(1) technique are (1) that the data has to be transformed to a Gaussian distribution to satisfy the assumption of the model and (2) that ϵ is generated from a normal distribution; hence any values from $-\infty$ to $+\infty$ can be simulated, which can result in generating unrealistic values.

With a traditional K-NN technique, the advantage is that the traces generated are not exactly the same as historical traces, as in the case of ISM. However, values not seen in the historical record cannot be generated, which is seen as one of the main drawbacks. The modified version of the traditional K-NN technique eliminated this concern and the drawbacks found with the ISM and PAR(1) technique.

The modified version of the K-NN model generated values not seen in the historic record, and it also has the ability to generate extreme values not seen in the history. Further, it retains the basic capability of the traditional K-NN methodology of reproducing PDF structure.

We compared the generated synthetic ensemble time series to the historic natural flow record for water year 1906 to 1995 to determine which technique best preserves the historic statistics. Specifically, all methods preserve the mean monthly and annually. The nonparametric methods best preserve the standard deviation. The parametric method tended to overestimate the standard deviation both monthly and annu-

ally. One of the drawbacks of the ISM was evident when comparing statistics. The statistics from the ISM exactly matched the historic statistic because the technique uses the direct historic record. We could not achieve any variability in the statistics, which limits the method when generating synthetic streamflows. One of the goals of modeling with stochastic hydrology is to explore future streamflows that are statistically possible given the historic record. With no variability around the historic statistics, the method is incapable of providing as wide a variety of flows as the PAR(1) or K-NN techniques.

The nonparametric models best preserve the lag(1) correlation. Again, the K-NN nonparametric model displayed a wider range of possible values monthly and annually while the ISM technique could only produce the exact historic statistic. Because the modified K-NN method is able to produce values that did not historically occur, it is not limited to reproducing the historic record like the ISM technique. One disadvantage of past nonparametric models is that they were limited to recreating numbers that occurred in the past record. Parametric models are able to reproduce values that did not occur in the past, which is why the parametric PAR technique is also able to reproduce a wider range of flows. The parametric PAR technique can produce the wider range of flows because the PDF is not limited on the tails; therefore, the model can reproduce extreme values. The unbounded tails in the parametric distribution allow the parametric model to generate unrealistic values. The K-NN technique is limited to generating the most extreme values in the historic record, plus the most extreme residual. The ISM is limited to the most extreme historic values, reducing the variability the ISM technique can model.

The ISM is also limited because it requires an extensive record. The historic record used for the model was 90 years, only allowing the ISM to generate 90 simulations. The simulations of the PAR and K-NN techniques were not limited to the number of years in historic record. We were able to produce 100 simulations for each

technique, which allowed us to converge on an actual estimate of the statistic of interest with more confidence. We were also able to apply the PAR and modified K-NN techniques to a wider range of datasets; we were not limited to datasets with extensive records.

Comparing the PDFs for the three techniques showed the strengths of the modified K-NN method, such as preserving the coefficient of skewness and non-Gaussian features. Incorporating a disaggregation scheme would allow the modified K-NN method to better preserve the annual PDF along with the monthly PDFs.

Another strength of the modified K-NN technique is that the flows it generates can be conditioned on a separate variable, such as climate, without much difficulty. Conditioning on a climate variable, such as the PNA index, would allow the modified K-NN technique to incorporate interannual variability in the generated synthetic streamflows. Incorporating interannual variability would allow the model to vary from the historic record based on future predictions of changing climate. Adding such a framework to a parametric technique is much more complicated than with a nonparametric K-NN technique.

For these reasons, we see the modified K-NN technique as easier to implement than the parametric PAR(1) technique and more flexible than either the PAR(1) or ISM techniques.

Chapter 3

Statistical Nonparametric Model for Natural Salt Estimation

3.1 Salinity Standards on the Colorado River

The most important water quality parameter in the Colorado River basin is total dissolved solids. Total dissolved solids is an accepted measure of water salinity. High levels of salinity in the Colorado River basin are damaging to agricultural, municipal, and industrial water users.

Salinity of the Colorado River became an important issue when the Mexican government strongly objected to the quality of the water Mexico was receiving. The average annual salinity of water delivered to Mexico in 1962 was 1,500 mg/L. This water was not suitable for irrigation and was adversely affecting crops. Two independent factors caused the rise in the salinity of the water delivered to Mexico. Deliveries were reduced because water was being held back in Lake Mead in preparation for filling Lake Powell and, at the same time, highly saline groundwater from the Wellton Mohawk irrigation district was being pumped into the Colorado River above the Mexico boundary. These highly saline waters greatly deteriorated the quality of the Colorado River water.

In response to Mexico's concerns and after years of negotiations, Minute No. 242 of the International Boundary and Water Commission (IBWC) dated August 30, 1973, was signed. Minute No. 242 stipulated that water delivered to Mexico have an

average salinity of no more than 115 ppm \pm 30 ppm above the average annual salinity at Imperial Dam.

Subsequently, the Colorado River Basin Salinity Control Act of 1974 was passed to ensure that the United States could meet its obligation to Mexico under Minute No. 242. The Act authorized construction of a desalting plant and additional salinity control projects in the upper basin.

Minute No. 242 sets a variable standard for the Mexico delivery, but does not set numerical water quality standards at any fixed points in the basin. Numerical standards resulted from separate U.S. legislation that set policy regarding water quality. The Federal Water Pollution Control Act Amendments of 1972 required development of water quality standards across the nation. The EPA interpreted the new legislation as a requirement to set fixed point numerical standards for salinity levels in the Colorado River basin. The basin states (Wyoming, Colorado, Nevada, Utah, Arizona, New Mexico, and California) found the Salinity Control Forum to help develop numerical salinity standards, including numeric salinity criteria required by the new legislation.

The following numeric salinity criteria were set in 1975 requiring maintenance of a flow-weighted average total dissolved solids concentration:

- 723 mg/L below Hoover Dam
- 747 mg/L below Parker Dam
- 879 mg/L at Imperial Dam

The standards were developed from the 1972 average annual salinity concentrations at each location and are currently unchanged (USBR, 1999 and Lee, 1989). To maintain these salinity standards, we must understand the sources for increased salinity.

3.2 Salinity Sources and Remediation

Natural and human-induced salinity result from point and non-point sources.

Natural point sources that have been identified include seeps and saline springs. Some springs originate from deep geological formations containing brackish water. Natural non-point sources of salinity generally originate from the weathering and dissolution of underlying rocks or soils overlaying the rocks.

The underlying rocks strongly characterize the natural salinity contributed by each subbasin. In some areas of the basin, the underlying rocks are relatively resistant to dissolution by water. Igneous and metamorphic rock types, for example weather at a very slow rate and contribute small amounts of dissolved solids (total salts). High mountain areas, such as the upper reaches of the Colorado, Green, and San Juan Rivers, are characterized by these types of formations. The middle and lower reaches of these subbasins contain sedimentary rock, which dissolves more readily in water and, therefore, contributes greater amounts of dissolved solids to the river. Soils in these regions have chemical dissolution characteristics similar to the underlying parent rocks. Valleys and lower lying plains, which are the predominant areas for agriculture, are characterized by these types of soil.

Agriculture increases salinity concentration through two processes: (1) salt concentration and (2) salt loading. The salt concentration process is a result of evapotranspiration from crops which consume water but leave salts behind in the soil. Return flows to the river from the diversion typically contain the same salt mass present in the diversion water but with less water, hence, higher concentration of salt. Additionally, reservoirs concentrate salt by evaporation, when water is lost from the reservoir but salt is conserved.

Salt loading occurs when salt is added or removed from the river when water transported through soil leaches or deposits salts present in the soil. The water can be introduced into the soil from human-induced sources, such as irrigation practices, or from natural sources, such as precipitation. Irrigation practices increase the flow through soils, which increases the total salt loading from previous natural salt loading

levels.

Federal agencies fund extensive efforts, costing millions of dollars per year, to reduce the human-induced and natural loading salt to the Colorado River. U.S. Bureau of Reclamation (USBR) has off-farm salinity control projects designed to reduce irrigation return flows by lining irrigation canals and laterals. Additional salinity control projects not related to agriculture include capping saline springs from abandoned gas and oil exploratory wells. The U.S. Department of Agriculture has implemented on-farm salinity reduction, thereby helping farmers to improve irrigation practices that increase irrigation efficiency, reducing return flows. The Bureau of Land Management has implemented programs to reduce erosion on public lands to limit the release of dissolved solids. As of 1998, salinity control projects, with a reported total cost of approximately \$426 million, had removed an estimated 634,000 tons of salt from the river. An additional \$170 million is projected to remove an additional 390,000 tons in order to maintain current salinity standards. (USBR, 1999).

3.3 Previous Modeling Efforts

In 1996, USBR performed salt-routing studies to estimate future salinity levels under normal hydrologic conditions given three scenarios: (1) without any salinity control projects, (2) with salinity control projects currently in place, (3) with additional salinity control projects to meet the numeric salinity standards in 2015 (U.S. Department of Interior, 2001). A modeling system developed by the USBR assisted basin planners project these future salinity control needs.

The modeling system, Colorado River Simulation System (CRSS) (USBR, 1987), simulates future salinity levels, considering changing human-induced water use in the basin. CRSS was developed in the late 1970's in response to a need for a modeling system that could simulate operations for various hydrologic and demand sequences. It allowed evaluation of impacts of proposed development high in the basin

on locations downstream.

CRSS includes various components of the physical mechanism that adds salt to the Colorado River basin. Agricultural diversions in the CRSS simulation model deplete water, but conserve salt mass, simulating the concentrating effect. In addition, the agricultural diversions model agricultural salt loading, termed *salinity pickup*.

Limited data is available describing salinity pickup throughout the basin. One extensive study (USBR, 1983 (1)) explains how salinity pickup was calculated in the Grand Valley using a mass balance of salt. The value for salinity pickup in the Grand Valley is an average salinity pickup from 1952 to 1980. The report states that the human-induced salinity pickup for the Grand Valley averages 580,000 tons per year \pm 90,000 tons with 95 percent confidence. To explain this large variance, the authors of this report attempt to correlate the salinity pickup with the annual flow of the Colorado River near Cameo, the annual flow of the Dolores River near Cisco, the annual salt loading on the Dolores, mean annual temperature, and annual precipitation. But salinity pickup could not be correlated with any of these. The report goes on to suggest that variations in salinity pickup cannot be due to changes in irrigation practices because variations in practices could not account for the magnitude of annual variations the data showed. The report concludes, “The variance from the mean is due to the inability to measure inflow and outflow with sufficient accuracy. Even with the adjustment for unaccounted flow, the variation about the mean salt load is significantly large.”

In addition to agricultural salt loading, CRSS must use estimates of natural salt loading. Natural salt loading contributes an estimated 47 percent of the total salinity in the Colorado River basin (U.S. Department of Interior, 2001). Natural flow is calculated, described in Chapter 2, “Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow,” by removing the human-induced effects on flow from observed historic flow. Human-induced effects include agricultural consumptive use, exports, and reservoir regulation, all of which are measured or can be estimated. Natu-

ral salt can be calculated by removing the human-induced effects on salt from observed historic salt. Unfortunately, human-induced effects on salt are much more difficult to quantify than effects on flow because salt from agricultural salinity pickup originates from many diffuse sources that are difficult to measure.

The USGS developed the current technique to estimate natural salt entering the river using historic flow and salt data from 1941 to 1983, along with 12 predevelopment months that assist calibration (Mueller and Osen, 1988). USGS used a multiple linear regression to develop an expression for historic salt as a function of historic flow and several human-induced development variables, such as consumptive use, irrigated acres, diversions, net reservoir releases, or a summation of all human induced adjustments (diversions + consumptive use \pm reservoir regulation). The regressions (one for each gauge) were calibrated to equalize the variance of the residuals. Then, each development variable was assigned a value of zero, and the historic flows were replaced by calculated natural flows provided by the USBR. The resulting relationship was proposed to relate natural salt to natural flow (Mueller and Osen, 1988).

CRSS requires an estimate of the natural dissolved solids as an input to the CRSS simulation model, along with the natural flow estimate associated with the natural dissolved solids estimate. Figure 31 displays the interconnection of the existing CRSS components, including a stochastic natural flow model to generate future stochastic flows, a salt regression model that estimates natural salinity associated with natural flows, and the CRSS simulation model. The stochastic natural flow model used for many years is the index sequential method (ISM). The CRSS simulation model calculates simulated historic salt mass by adding the natural salt mass to the salinity picked up by agriculture and subtracting the salt that leaves with water exported from the basin.

Interconnection of existing CRSS Components

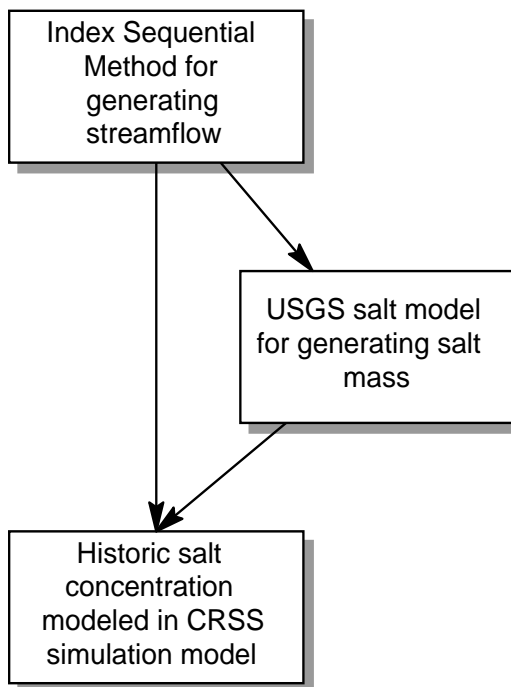


Figure 31: Flowchart depicting interconnection of existing CRSS modeling system.

The existing CRSS overpredicts the historical salt mass at USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO), from 1970 to 1990, by an average 140,000 tons/year. The overprediction could result from

- salinity pickup from agriculture being too high,
- natural salt loading being too high.

From 1970 to 1990, the historic salt mass in the river passing gauge 09072500 averaged 570,300 tons/year. The relationship proposed by the USGS estimates average annual natural salt mass of 583,000 tons/year. For CRSS to simulate the historic salt mass, the human-induced salinity pickup sources would need to *remove* salt from the river. Current estimates, as reflected in CRSS, are that human-induced sources contrib-

ute 137,000 tons/year from agriculture salinity pickup and exports remove an average 44,000 tons/year. The estimate for salinity pickup by agriculture is developed from an extensive study that quantified estimates of natural and human-induced salt (Iorns et al., 1965). The report estimates that, in 1957, natural sources contributed 516,200 tons/year, and human-induced sources contributed 138,881 tons/year from agriculture salinity pickup and removed 15,881 ton/year by exports above Glenwood Springs. These values were adjusted for current basin conditions then input in the CRSS simulation model.

Using these numbers, if human-induced sources contributed no salt above gauge 09072500, the existing CRSS would still overpredict salt mass. The Iorns report indicates that the human-induced sources of salinity are not removing salt, but are adding significant amounts. These findings point to an overestimation of natural salt by the USGS model. To correct or refine the USGS model would require a reanalysis of the detailed data on which the regressions were based. However, this data is not available. Therefore, we propose a new technique to relate natural flow to natural salt that is more accurate and for which the uncertainty can be quantified.

3.4 Statistical Nonparametric Natural Salt Model

To address the problems of overprediction by the existing model technique, we developed a new modeling approach to estimate natural salt mass given natural flow. The model is a nonparametric statistical model, based on a local polynomial method, that fits a function to relate natural salt to natural flow. The uncertainty in the estimates of the natural salt from this model is incorporated via a k-nearest neighbor (K-NN) bootstrap technique.

3.4.1 Local Regression (Nonparametric Regression)

Parametric techniques generally are limited to fitting a linear relationship to the function f , i.e., linear regression. Furthermore, parametric techniques fit an equation

(linear or nonlinear) to the entire data, which restricts the ability to capture non-linearities in the data. In addition, hypothesis testing requires normality assumption of the error term, thus further restricting the model (Helsel and Hirsch, 1992).

A relationship between a dependant variable (y) and a set of independent variables (x) taken in the form of

$$y = f(x) + e_t \quad \text{Eq. 3.7}$$

where e_t is the error term.

Nonparametric methods, on the other hand, fit the function f locally and make no prior assumption about the functional form, i.e., linear, quadratic. Thereby, they have greater flexibility in capturing any arbitrary relationship. Several nonparametric methods exist: kernel based, splines, local polynomials. For a detailed description of these methods and comparisons, refer to Owosina (1992) and references therein. We adopted a local regression scheme that has been shown to be easy to implement and effective (Rajagopalan, and Lall, 1998, and Loader, 1999). The method and the algorithm are described through the following example.

The synthetic dataset was generated from the following equation:

$$y_t = \sin(x_t) - 0.2x_t^2 + e_t \quad \text{Eq. 3.8}$$

where e_t is a mean zero, variance 0.1, Gaussian random variable adding noise to the function. The function is shown in Figure 32, where x_t is a sequence of 100 points from $0 \rightarrow 2\pi$. The solid circles indicate points generated from Eq. 3.8. The long-dashed line shows the true function (i.e., $\sin(x_t) - 0.2x_t^2$) without the noise term. The short-dashed line is a linear regression fit to the data, and the medium-dashed line is a quadratic regression fit. It is evident that neither of these regressions is able to capture

the nonlinear sinusoidal function, the true underlying function.

Local regression, shown by the solid line, is able to fit this data well. We used the library, LOCFIT, designed for use with Splus (Loader, 1999). The local regression technique, as the name suggests, performs local regression at each point of estimate.

We explain this methodology, with reference to Figure 32, as follows:

- (i) Let us assume that we want to estimate the function at x_t .
- (ii) A neighborhood is defined around x_t . The size of the neighborhood is $\alpha \times n$, where α is a parameter between 0 and 1. Bigger α indicates more smoothing. (For example, for $\alpha = 1$ and a local linear fit; it is the same as the parametric linear regression).
- (iii) For the neighbors captured in the neighborhood (shown in the dashed rectangles) a regression of order p is fit. Typically, a linear fit works very well (shown as the heavy solid line within the neighborhood).
- (iv) The fitted regression is used to estimate at x_t .
- (v) This is repeated at all points where we need the estimate.

The advantages to this method are apparent. Because we fit local regressions there is great flexibility in modeling any structure that might be present in the data (linear and nonlinear). The neighborhood size (α) provides the amount of smoothing. Furthermore, this approach is assumption free, unlike the parametric models that require data to be Gaussian distributed.

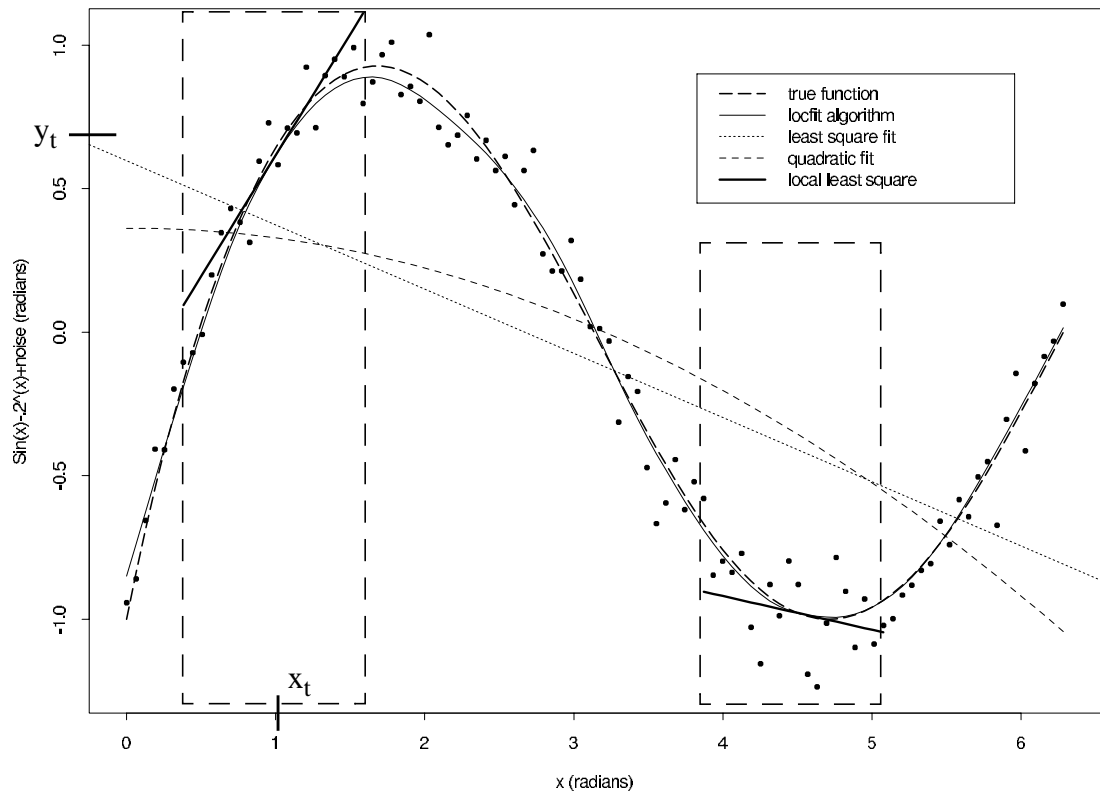


Figure 32: This figure shows several data fitting techniques for data generated by sine function with noise. The solid line is a local linear regression with $\alpha = 0.2$, calculated by the LOCFIT algorithm. The dashed rectangle depicts the neighbors chosen around point x_t to develop the local linear regression. A local least square linear regression on the points in the rectangle is indicated by a heavy solid line.

Alpha can be chosen by various methods; typically a function is chosen (i.e., in linear regression it is the least square minus the sum of the squares of the difference between the observed and estimated values). However, the performance of a fitted model in terms of predicting future values is not assured. Therefore, fitting the model over predictive sum of squares instead of least squares would be very effective. One method is the cross-validation (CV) in which a point is dropped and it is predicted by fitting a model to the rest of the data.

$$CV(\alpha) = \frac{1}{n} \sum_{i=1}^n (Y_i - y_i)^2 \quad \text{Eq. 3.9}$$

where y_{-i} is the estimate at x_i by dropping x_i, y_i from the fit. This approach chooses different α s and selects the one that minimize Eq. 3.9. When a dataset is small Loader (1999) recommends using CV; therefore, we used this technique to find α .

The generalized cross validation (GCV) function, on the other hand, is a good estimate of the predictive error when the dataset is larger. The GCV score function provides an estimate of the CV. The GCV uses all points from a neighborhood, then estimates y_i with various neighborhood sizes and finds the neighborhood size that gives that lowest score. The neighborhood size with the lowest GCV score determines the appropriate alpha.

The GCV score function for a local estimate of $\hat{\mu}$ is given as

$$\text{GCV}(\alpha) = n \frac{\sum_{i=1}^n (Y_i - y_i)^2}{(n - v_1)^2} \quad \text{Eq. 3.10}$$

where: v_1 is the fitted degrees of freedom,

n is the sample size,

$Y_i - y_i$ is the residual.

This local regression technique is used to fit the natural salt as a function of natural flow. This modeling technique uses the human-induced salt loading data used in the current CRSS model to calculate natural salt mass. Calculated natural salt mass is “back-calculated” from the observed historic salt mass by adding the mass from agricultural salinity pickup and removing the mass that leaves with exports. Using the calculated natural salt, a direct relationship between natural flow and salt is found. We developed 12 relationships, one for each month, for years 1941 to 1985. Using the mass added by agriculture and removed by exports from the CRSS model allows the statistical nonparametric natural salt model results to easily be incorporated in the model.

Figure 33 shows a scatter plot of the calculated natural salt and natural flow for

the month of April for the years 1941 to 1985. The dotted line represents a parametric least squares linear regression; it fits the regression using the entire dataset at once and cannot capture nonlinearities of the data distribution. The nonparametric local regression, shown by the solid line, captures nonlinear features of the calculated natural flow and natural salt relationship using an alpha of 0.95. Rajagopalan and Lall (1998) describe similar nonparametric regression in which they use a locally weighted polynomial regression to estimate spatial data describing precipitation.

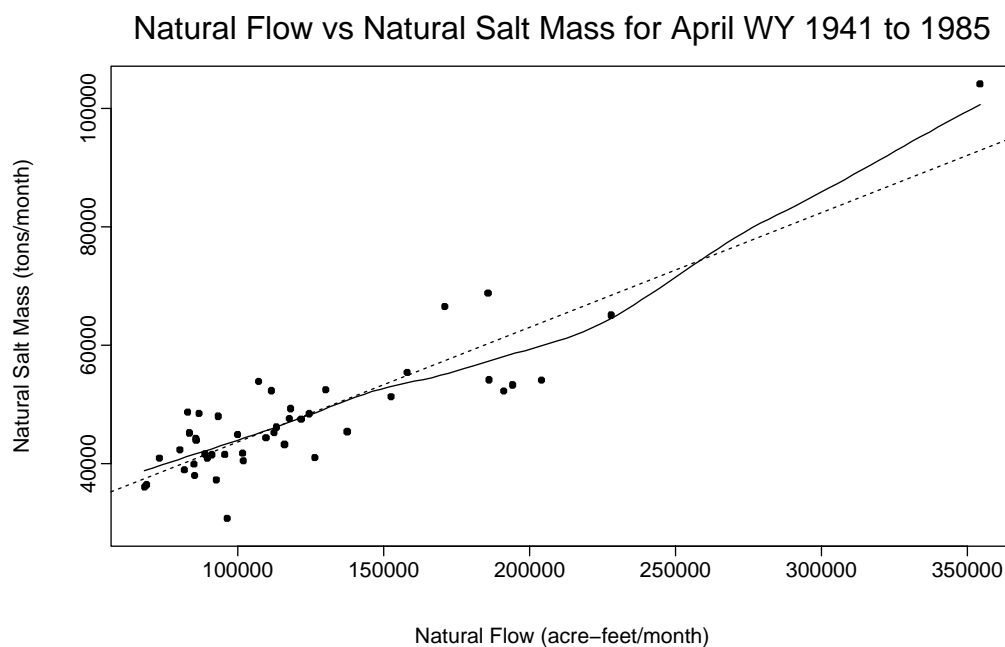


Figure 33: This figure shows nonparametric regression for natural salt dependent on natural flow. Alpha equals 0.95.

We developed a K-NN residual resampling technique to quantify the uncertainty of the estimates from the local regression method. This technique is similar to the modified K-NN technique described in Chapter 2.

3.4.2 Residual resampling

The residuals from the regression indicate the uncertainty of the regression. To

quantify the uncertainty of estimates for natural salt from the regression, we implemented a method to resample the residuals. The data is widely scattered around the regression, indicating large variance. To quantify the variance in the data generated from the regression relationship, we used a K-NN technique as described in Chapter 2. For example, in Figure 34, the stochastic flow model generated a natural flow x_t . We first found the mean x_t associated salt mass y_t from the regression. Next, we found the k nearest neighbors, using a Euclidean distance, shown as circles. The k -nearest neighbors are outlined with the dashed rectangle. We resampled one of the residuals, e_t^* , using a weighted kernel and added it to the mean salt mass estimate y_t . We repeated this technique for each value of natural flow developed from the stochastic natural flow model. When we applied this method to 100 ensembles of simulated natural flow, the resulting 100 ensembles of simulated natural salt reflected the variance around the regression of the calculated natural salt as a function of natural flow.

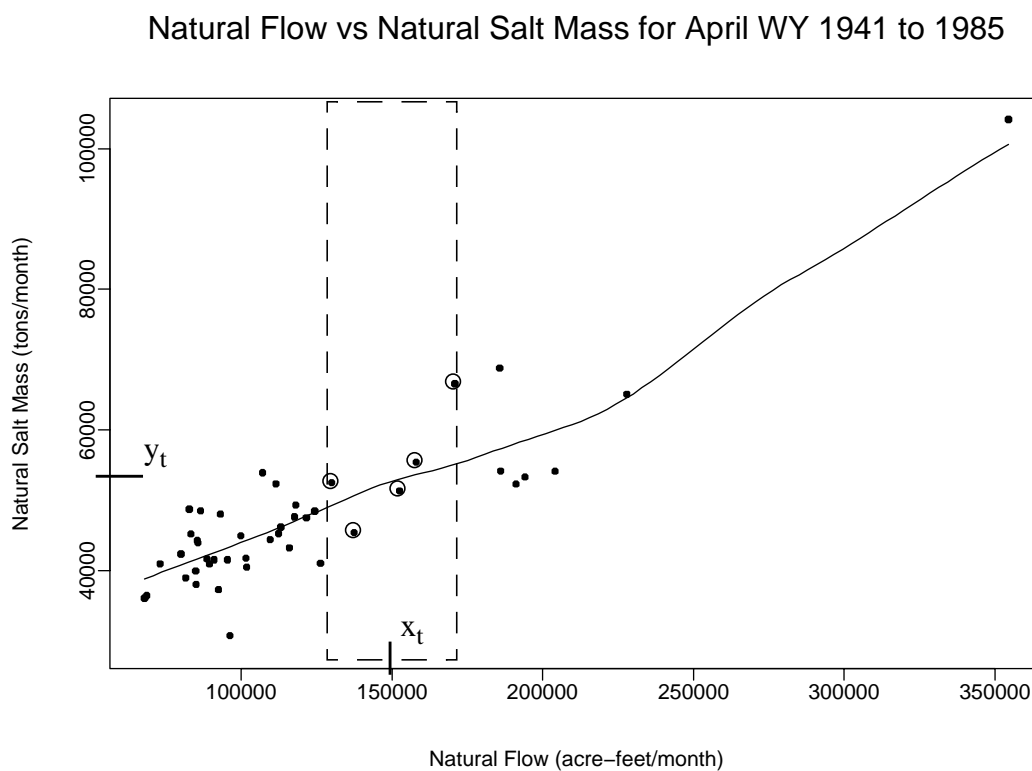


Figure 34: This figure shows nonparametric regression for natural salt dependent on natural flow. Alpha equals 0.95.

3.5 Model Evaluation

We used the statistical nonparametric natural salt model (SNNSM) to generate natural salt mass for given values of natural flow. We compared the results to results from the USGS salt model and the calculated natural salt. We evaluated each technique on how well it reproduced the calculated natural salt, and we discussed the differences.

3.5.1 Test Ensembles

The SNNSM takes natural flow data and generates an estimated value of natural salt. We used calculated natural flow data from water years 1941 to 1985 and the calculated natural salt data from 1941 to 1985 above USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO) to develop the regressions for the SNNSM. The calculated natural flows are the observed historic flows minus the total human-induced consumptive use. We calculated the calculated natural salt from historic gauged data and salt load data from CRSS as

$$\begin{aligned} \text{calculated natural salt} = & \text{observed historic salt} \\ & + \text{salt with water exported out of the basin} \\ & - \text{salinity pickup from agriculture (values} \\ & \quad \text{based on CRSS)} \end{aligned}$$

The salt removed by exports and the salt added by agriculture for the period 1941 to 1985 were taken from the data used to drive the CRSS model. In the CRSS model, agriculture annually adds 137,000 tons of salt above gauge 09072500. A constant salinity pickup is a fair assumption because agricultural consumptive use was basically constant from 1941 to 1995. The exports remove a constant concentration of 100 mg/L. The tons removed by exports vary with flow, according to the relationship between flow and salt mass.

We found the regression relationships for each monthly time series, i.e., we developed 12 relationships between calculated natural flow and salt mass. Developing these relationships was necessary because the current CRSS model simulates river policy at a monthly timestep. Enabling the model to accept an annual timestep is beyond

the scope of this research.

We developed confidence intervals around the estimated salt mass, which was calculated directly from the regression, by running 100 synthetic traces of monthly simulated natural flow through the statistical nonparametric natural salt model, incorporating residual resampling to generate 100 synthetic traces of monthly simulated natural salt mass. We summed the simulations to an annual time step and then calculated the 5 percent and 95 percent quantiles of the data, finding the confidence intervals.

3.5.2 Model Evaluation Criteria

We compared results from the SNNNSM to results from the USGS salt model for water years 1941 to 1995. We evaluated each model's performance on a monthly and annual time scale. We derived the annual time series by summing the water year months, October through September. We compared three aspects for each salt model.

First, we compared the regressions developed with the SNNNSM to the USGS salt model.

Second, we used both model regressions to generate estimated natural salt mass from 1941 to 1995 given the same estimated natural flows. We calculated estimated historic salt from the estimated natural salt mass for each salt model. We calculated estimated historic salt mass as

$$\begin{aligned} \text{estimated historic salt} &= \text{estimated natural salt} \\ &\quad - \text{salt from exports} \\ &\quad + \text{salt from agricultural salinity pickup} \end{aligned}$$

We evaluated each technique on how well it matched the observed historic salt mass.

Third, we added the confidence interval developed from the residual resampling technique in the SNNNSM and plotted the estimated natural salt and the estimated natural salt from the USGS salt model. We evaluated each technique on how well it matched the calculated natural salt. Of interest was if the SNNNSM could capture the

calculated natural salt within the confidence intervals.

3.6 Results

Figure 35 shows the local linear regression relationship developed using our SNNSM and the existing USGS salt model for April and June. The solid circles show the data points for calculated natural flow versus calculated natural salt. The dotted line shows the relationship developed by the USGS salt model. It is a power curve relationship developed for the month of April, from the USGS salt model technique, to find estimated natural salt. The solid lines shows the SNNSM relationship. It shows a local linear regression of the solid circles. The local linear regression is able to capture the nonlinear characteristics of the data. For April, the USGS salt model relationship underestimated the calculated salt mass when compared to the SNNSM. The lower graph compares the same relationships for the month of June, which displays data with a relatively linear relationship. June is a higher flow and salt mass month. During this month, the USGS relationship estimated higher salt mass than the SNNSM. The wider scatter of the data points indicates there is more variability around the relationship. The USGS relationship has no method to incorporate the variance in the data around the regression, while the SNNSM can incorporate this variance by resampling the residuals from the regression when generating salt mass.

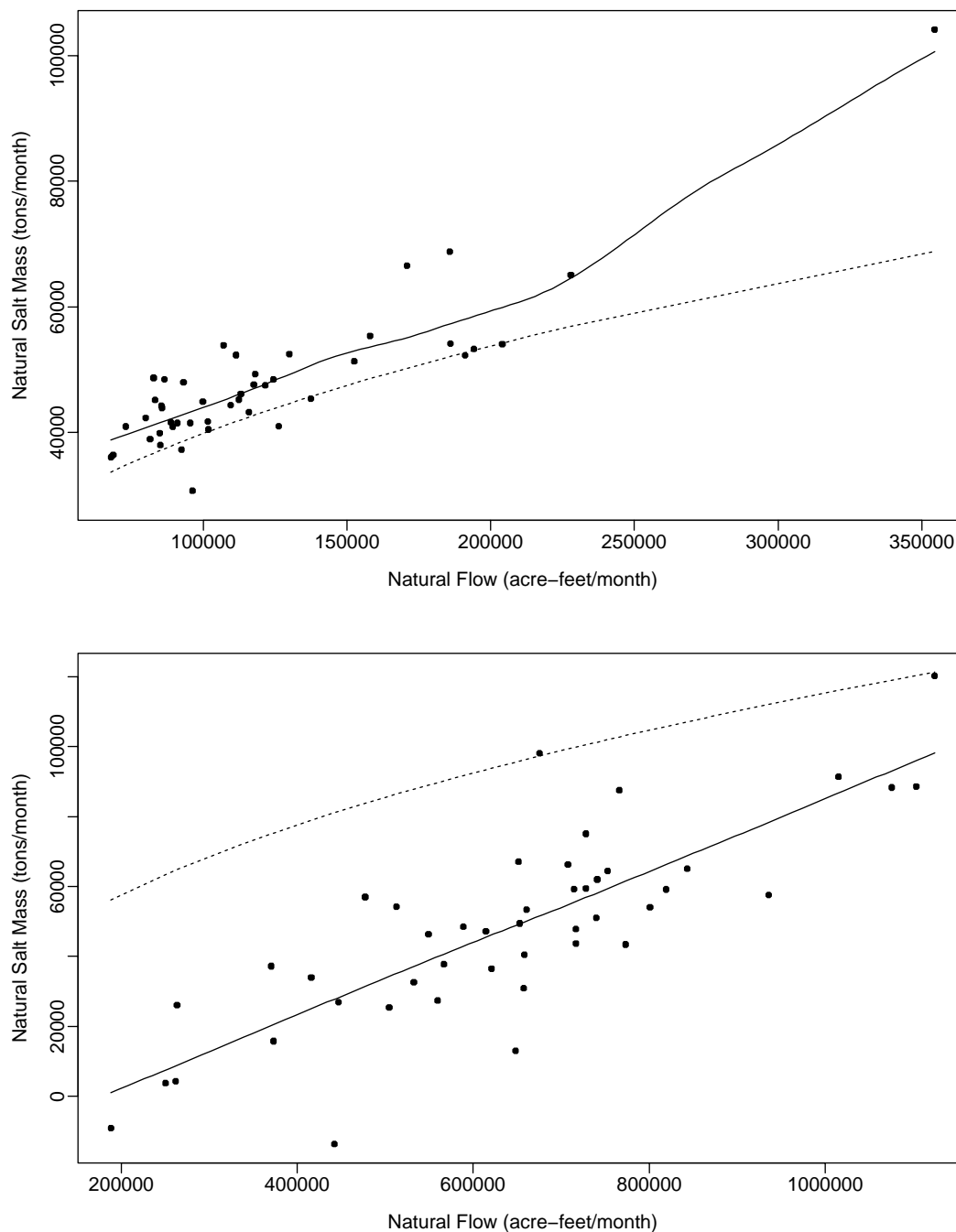


Figure 35: This figure shows the relationship for natural salt dependent on natural flow for water years 1941 to 1985. The dashed line shows the USGS salt model relationship. The solid circles show SNNSM, and the solid lines show the local linear regression. For April (upper graph) the USGS slightly under-estimates the results taken directly from the natural flow and salt estimates. For June (lower graph) the USGS relationship predicts a higher salt mass compared to the SNNSM. Both LOCFIT's have an alpha of 0.95.

Figure 36 compares the regressions of annual values for calculated natural flow and salt. The annual values are a summation of the monthly values, showing the USGS salt model estimates higher natural salt mass than the SNNSM. At the higher flows, both regressions give a similar salt mass for a given high flow while at lower flows, the USGS salt model gives a significantly higher salt mass for a given low flow.

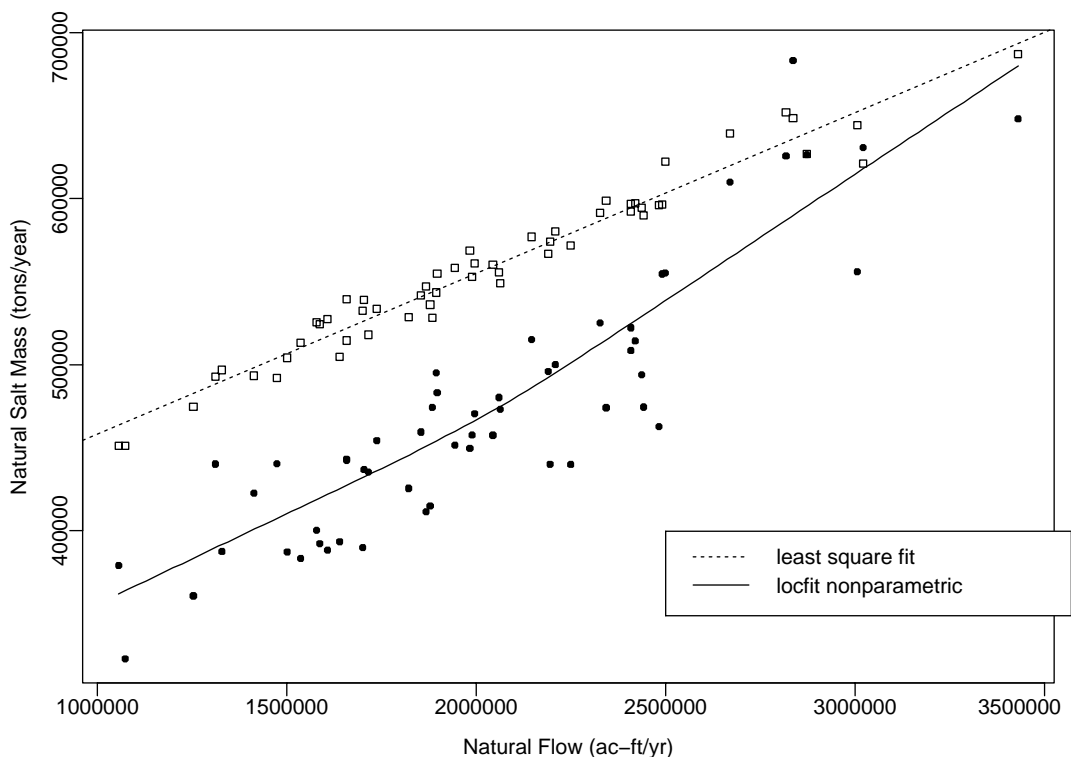


Figure 36: The open squares show the annual natural salt mass, summed from monthly natural salt, for the USGS salt relationship, and the dashed line shows a least squares fit. The solid circles show the data to find the SNNSM relationship, and the solid line shows a LOCFIT through the data. The USGS salt model gives higher natural salt mass compared with the SNNSM.

The graphs of Figure 37 display time series generated using both the new proposed regression of calculated natural salt as a function of natural flow and the regression proposed by the USGS. We used a single calculated natural flow time series from water years 1941 to 1995 to generate estimated natural salt given the calculated natural flow. The estimated natural salt was generated at a monthly time step. We summed the

monthly data to annual to remove the seasonal signal from the time series. The dot-dashed line shows the local linear regression without the residual resampling. The double-dot-dashed line is the result from the USGS salt model. In the upper graph, it can be seen that the USGS salt model estimates a higher estimated natural salt mass than the SNNSM. The lower graph shows the estimated historic salt using the estimated natural salt generated from both the SNNSM and the USGS salt model. The heavy solid line is the observed historic salt mass. As shown, the SNNSM tracked the observed historic salt mass more accurately and tracked the salt mass well during low flow periods.

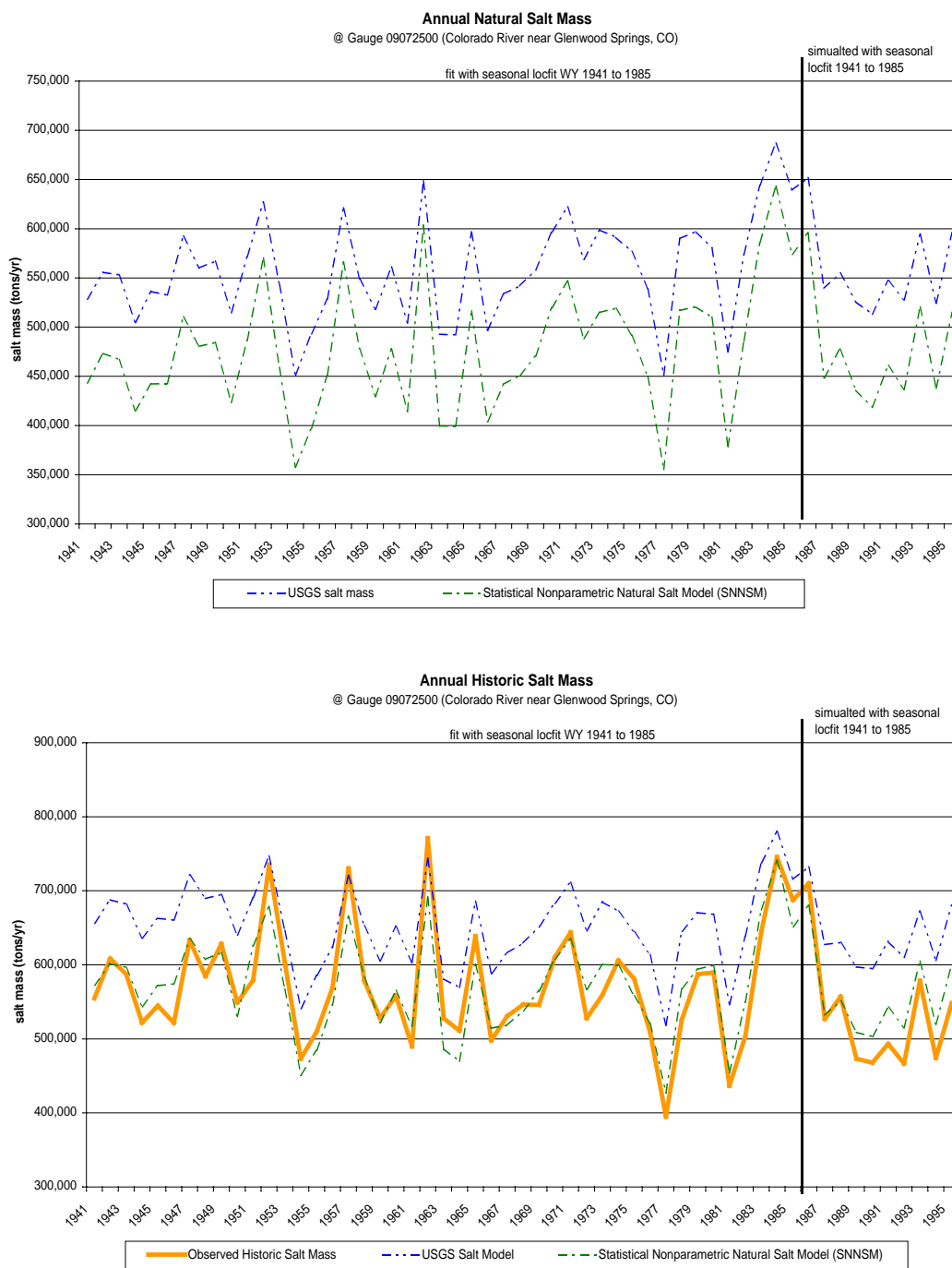


Figure 37: The upper graph compares the time series for natural salt generated from the USGS relationship and the modified K-NN method with the historic natural flow from 1941 to 1995. The lower graph compares the time series for simulated historic salt using the natural salt generated from the USGS relationship and the modified K-NN method with the observed historic salt from 1941 to 1995.

Figure 38 shows the effects of adding the residual resampling step into the SNNSM. With the residual resampling, we were able to incorporate the variance of our regression when generating simulated natural salt given an ensemble of simulated natural flow. The dot-dashed line labeled “Statistical Nonparametric Natural Salt Model (SNNSM)” shows the estimated natural salt from the model if only the regression were used and no resampling were performed. Adding resampling perturbs the result from the regression based on the variance of our regression. To show the effects of the perturbation, we developed confidence intervals (dashed line) around the Statistical Nonparametric Natural Salt Model line, as described in Section 3.5.1, “Test Ensembles.” By incorporating the confidence intervals, we were able to capture the variability of the estimated natural salt from the regression during the period 1941 to 1985. After 1985, the estimated natural salt falls slightly below the 5 percent confidence interval. This indicates that, for given natural flow, the natural salt is lower after 1985 than before 1985. Including the natural flow and salt data after 1985 when developing our SNNSM will allow our confidence intervals to capture the salt after 1985.

The heavy double dot-dashed line shows the natural salt from the USGS regression. The USGS relationship is outside the upper confidence intervals for all but one year. The SNNSM rarely generated values as high as the USGS regression.

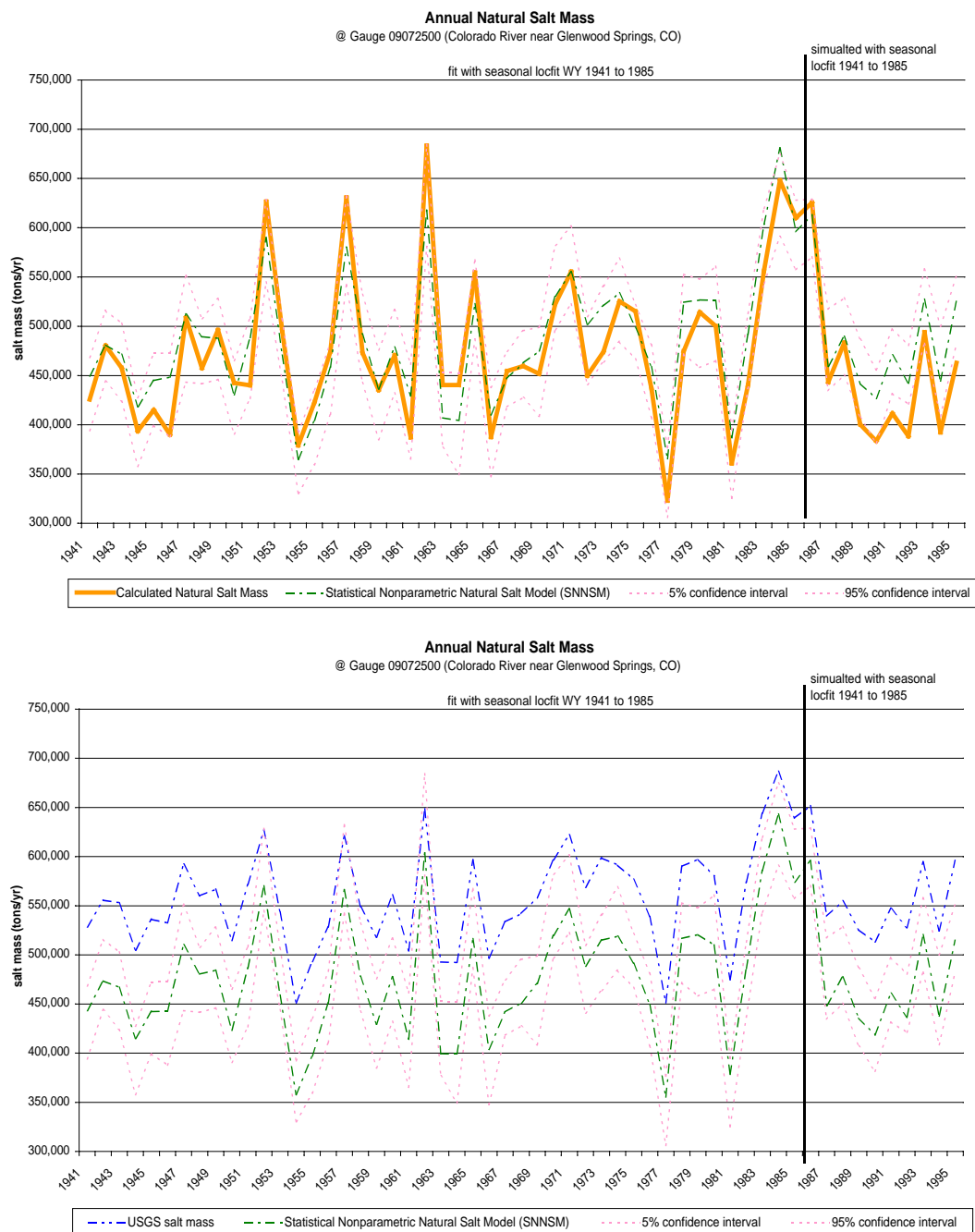


Figure 38: The upper graph shows the annual natural salt mass time series from the modified K-NN method with 5 percent and 95 percent confidence intervals. The heavy solid line shows the calculated natural salt mass. The dot-dashed line shows the natural salt mass taken directly from the local regression. The 5 percent and 95 percent confidence shows the variation around the line direct from the regression that occurs from the residual resampling. The lower graph removes the calculated natural salt and adds the natural salt determined from the USGS relationship. It is evident that the USGS relationship determines the natural salt mass to be higher than the natural salt mass we determined with our technique.

3.7 Discussion and Conclusions

The current salt model proposed a relationship for calculated natural salt mass as a function of calculated natural flow with a weighted least-square regression fit to a model of observed historic salt mass as a function of observed historic flow and several development variables, which represent the human-induced water resource development throughout the basin.

We outlined a technique to calculate the natural salt based on the way CRSS simulates estimated historic salt mass. The estimated natural salt and the human-induced salt determined the estimated historic salt mass. The observed historic salt for water years 1941 to 1995 was compared to the estimated historic salt generated from the USGS salt model and found to be 15 percent, or 86,000 tons, greater than the annual average observed historic salt mass.

Using the calculated natural flow and salt mass, we created a relationship between calculated natural flow and salt mass. Using the regression developed from the relationship and the calculated natural flow, we were able to reduce the difference between the annual average observed historic salt mass and the estimated historic salt mass to 0.8 percent, or 4,000 tons. Further, we incorporated the residual resampling technique in the SNNSM to enable the model to preserve the variance around the regression of the calculated natural salt as a function of calculated natural flow.

The resampling technique allowed a greater variety of natural salt values to be generated for a given natural flow. These techniques allowed the SNNSM to reproduce the calculated natural salt from 1941 to 1985 within confidence intervals created from 100 simulations of estimated natural salt mass.

Chapter 4

Historic and Future Salt Concentration Modeled in RiverWare

4.1 Introduction

Salt concentration is important in the Colorado River basin because it is a water quality parameter regulated by federal water quality standards. Federal water quality standards were set as a result of the Federal Water Pollution Control Act Amendments of 1972. The Amendments, interpreted by the Environmental Protection Agency, require numerical standards for salinity levels in the Colorado River basin. Modeling studies that predict long-term salinity levels under various operational scenarios and salinity control projects facilitate planning and operating the river to meet water quality standards. Operational and planning policies in the Colorado River basin are complicated because the basin is governed by many laws, statutes, and court decrees. To capture these complex policies in a modeling tool, the U.S. Bureau of Reclamation (USBR) developed the Colorado River Simulation System (CRSS) (U.S. Department of Interior, 2001), a series of computer models for the entire basin.

Developed in the early 1970's, one of the purposes of CRSS is to conduct long-term operational and planning studies that allow managers to understand the effects of future development on salinity throughout the Colorado River basin.

CRSS includes a simulation model of the entire Colorado River system. It also includes a stochastic natural flow model to generate future stochastic flows and a salt

regression model that estimates natural salinity associated with natural flows. The stochastic natural flow model used for many years is the index sequential method (ISM). ISM is limited to generating flows and flow sequences that have occurred historically, which limits the ability of planning studies to consider flows that are statistically possible but have not occurred. The salt regression model consists of a series of 12 monthly regressions of natural salt mass as a function of natural flow that were developed by the USGS (Mueller and Osen, 1988), as discussed in Chapter 3. These two models provide the natural flow and natural salt input data for the CRSS simulation model.

To ensure the CRSS simulation model is calibrated, it is periodically used to simulate a historic period and the results are compared to the observed historic record. If discrepancies are found, the model is not simulating the historic period correctly. Recently, the historic runs have indicated that the simulation system overpredicts salt throughout the basin, as discussed in Chapter 3, “Statistical Nonparametric Model for Natural Salt Estimation.” We found that the salt entering and leaving the river from human-induced sources appeared reasonable, based on Iorns et al. (1965) report. Thus, we concluded that the 12 monthly natural salt regressions developed by the USGS overestimate natural salt when the results from the regressions are summed to annual natural salt.

To improve CRSS performance, newly developed models replaced both the stochastic natural flow model and the natural salt regression model. In this chapter we bring these new models together in a simulation model to demonstrate that they more accurately reflect variability by improving flow variability prediction and quantification of risks, fully reflecting future salt variability by improving the relationship between natural flow and salt, and reproducing historical results. These developments result in an improved analysis of future salinity and quantified risks.

We performed our study on the upper mainstem of the Colorado River, at

USGS stream gauge 09072500 Colorado River near Glenwood Springs, CO. We chose the upper Colorado River mainstem because this part of the basin contributed more than 51 percent of the total annual historic salt load seen in the outflow from Lake Powell from 1941 to 1990. Gauge 09072500 also exhibited an overprediction of historic salt mass in the calibration runs from 1970 to 1990 by 20 percent. One of our primary goals was to develop techniques to correct the overprediction. While investigating the cause and means to correct the overprediction, we intended to ensure our solutions were portable and easily implemented at the 28 remaining gauges throughout the basin.

4.2 Existing CRSS

CRSS is used to simulate flow and salt over a historic period to verify that the model is calibrated. It is also used to simulate future periods and model proposed development and changing operational and planning policy. The model uses different input data for different simulated time periods. The CRSS simulation model runs at a monthly time step, as required by the operational rule set.

4.2.1 Historical Verification

When CRSS is used to verify a historic period the CRSS simulation model is populated with historic data. The *calculated natural flow*, available from 1906 to 1995 at a monthly time step, is obtained as follows:

$$\begin{aligned} \text{calculated natural flow} = & \text{observed historic flow} \\ & + \text{agricultural consumptive use} \\ & + \text{exports} \\ & + \text{municipal and industrial consumptive use} \\ & \pm \text{effects of off-stream reservoir regulation} \end{aligned}$$

The primary sources of data include the observed historic flow from USGS stream gauge records, agricultural consumptive use estimated by the USBR using the Blaney Criddle method, exports from USGS stream gauge records, off-stream reser-

voir regulation from USBR records for off-stream reservoirs run by the USBR or an independent manager (USBR, 1983(2); USBR, 1985; USBR, 1987; USBR, 1992).

USGS stream gauge 09072500 recorded river flows from water year 1899 to 1966. Since 1966, flow has been determined at the gauge as the difference between flows at USGS stream gauge 09085000 (Roaring Fork River at Glenwood Springs, CO) and USGS stream gauge 09085100 (Colorado River below Glenwood Springs, CO). A different gauge performs periodic conductivity measurements. The conductivity measurements from USGS water quality gauge 09071100, from 1942 to present are used to calculate total dissolved solids. In the Colorado River basin, total dissolved solids is accepted as a measure of salt concentration. Salt mass is calculated by multiplying flow volume and salt concentration, then using the appropriate conversion factor. The salt gauge was extended back to 1941 as described in Mueller and Osen, (1988). After this work, a historic flow time series from 1906 to 1995 and a historic salt time series from 1941 to 1995 were available. Together, these time series represent the observed historic flow and salt mass in the Colorado River near Glenwood Springs, Colorado, at a monthly time step.

A series of 12 regressions, developed by the USGS, were used to compute the *estimated natural salt mass* associated with natural flows. The regressions were developed using observed historic flows and salt and several development variables (Mueller and Osen, 1988).

Figure 39 depicts the CRSS simulation model's water and salt balance with a line diagram of the basin above USGS stream gauge 09072500. The inputs are at the top of the diagram: calculated natural flow, as described previously, and the associated estimated natural salt mass from the USGS salt model. The model routes these inputs through the river reach above USGS gauge 09072500, where the historic monthly depletions from agriculture, exports, municipal and industrial uses, and the effects of off-stream reservoir regulation are removed from the river reach. Reservoir regulation

describes the total monthly change in surface storage, surface evaporation, and bank storage (as appropriate) in reservoirs above USGS gauge 09072500. Salt mass is added with agricultural returns and removed with exports. The left side of the diagram shows how simulated historic flow is related to calculated natural flow. The right side shows how simulated historic salt mass is related to estimated natural salt mass. The CRSS simulation model simulates these processes and produces as output historic gauge flows and associated salt mass. The result of a historical verification run is shown at the end of Chapter 3, “Statistical Nonparametric Model for Natural Salt Estimation.”

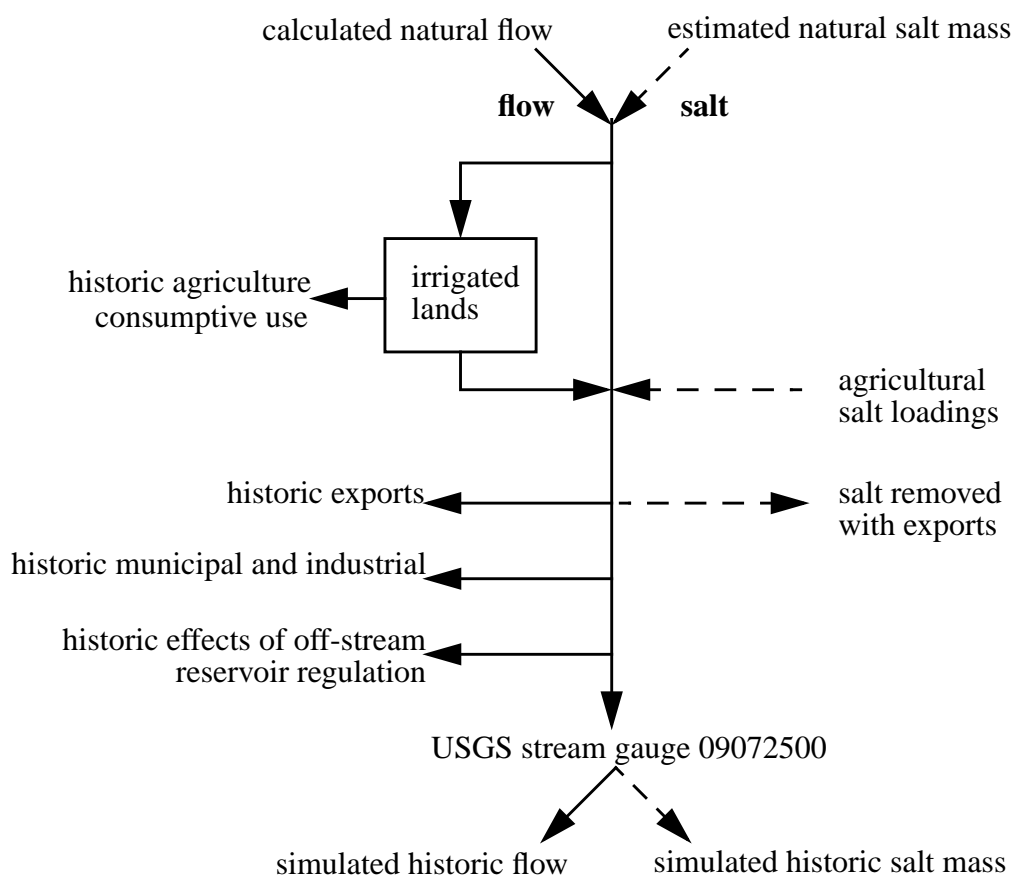


Figure 39: Line diagram of the CRSS simulation model and data for historic verification.

4.2.2 Stochastic Planning Runs

Long-term operational and planning studies are also conducted using CRSS. For these runs, the CRSS simulation model uses natural flows generated by the ISM. The ISM resamples the 90 years of historic calculated natural flows, generating 90 individual time series sequences of *synthetic natural flows*. The CRSS simulation model uses the 90 synthetic natural flow time series sequences, or traces, to produce 90 simulated results, which are then used to generate statistical probabilities of various events.

The USBR has used the index sequential technique to generate synthetic hydrology for the CRSS since the inception of the CRSS in the 1970's. Various studies have found that the ISM generates "statistically faithful" synthetic streamflow sequences (Kendall and Dracup, 1991; Ouarda et al., 1997). The Colorado River System is well suited to using the index sequential modeling system because of the extensive historic time series in the basin (water year 1906 to 1995).

Two attributes of the ISM that detract from its usefulness are that it requires an extensive time series and that it cannot generate values or hydrologic sequences that have not occurred in the original dataset. Thus, a hydrologic sequence such as a period of drought or surplus that has not been seen in the past could not be generated.

The simulation model also requires the *synthetic natural salt mass* associated with each synthetic natural flow sequence. Again, the series of 12 regressions, developed by the USGS, are used to compute the associated synthetic natural salt mass for a given synthetic natural flow. Thus, 90 individual synthetic natural salt traces are found using the regressions, one for each synthetic natural flow trace.

Figure 40 depicts the CRSS simulation model's water and salt balance, with a line diagram of the basin above USGS stream gauge 09072500 when the simulation model simulates future flow and salt mass. The inputs are at the top of the diagram: synthetic natural flow, from the stochastic natural flow model and associated synthetic

natural salt mass from the USGS salt model. The model routes these inputs through the river reach above USGS gauge 09072500, where the projected future monthly depletions from agriculture, exports, municipal and industrial uses are removed from the river reach. Projected salt mass is added with agricultural returns and removed with exports.

The left side of the diagram shows how simulated future flow is related to synthetic natural flow. The right side shows how simulated future salt mass is related to synthetic natural salt mass. The simulation model performs Monte Carlo simulations by running each synthetic natural flow and associated salt time series. The existing CRSS has 90 synthetic natural flow and associated salt time series that are each run through the simulation model, one at a time, calculating 90 simulated historic or future flow and associated salt time series. Together, the 90 simulations of simulated historic or future flow, salt, and concentration can be used to approximate the PDF for the observed historic or predicted future flow, salt, and concentration.

To verify that the stochastic planning runs simulate accurate traces of simulated flows and associated salt mass, the CRSS simulation model was populated with historic monthly depletions, instead of projected future depletions. Then, PDFs of the observed historic flow and salt mass were compared to the simulated historic flow and salt mass from the CRSS simulation model. If the simulation statistics preserved the observed historic flow and salt mass statistics, the CRSS is deemed verified.

We performed stochastic planning runs with both historic depletions to verify the simulation runs and projected future depletions, to compare the existing CRSS, explained previously, and the modified CRSS, explained next.

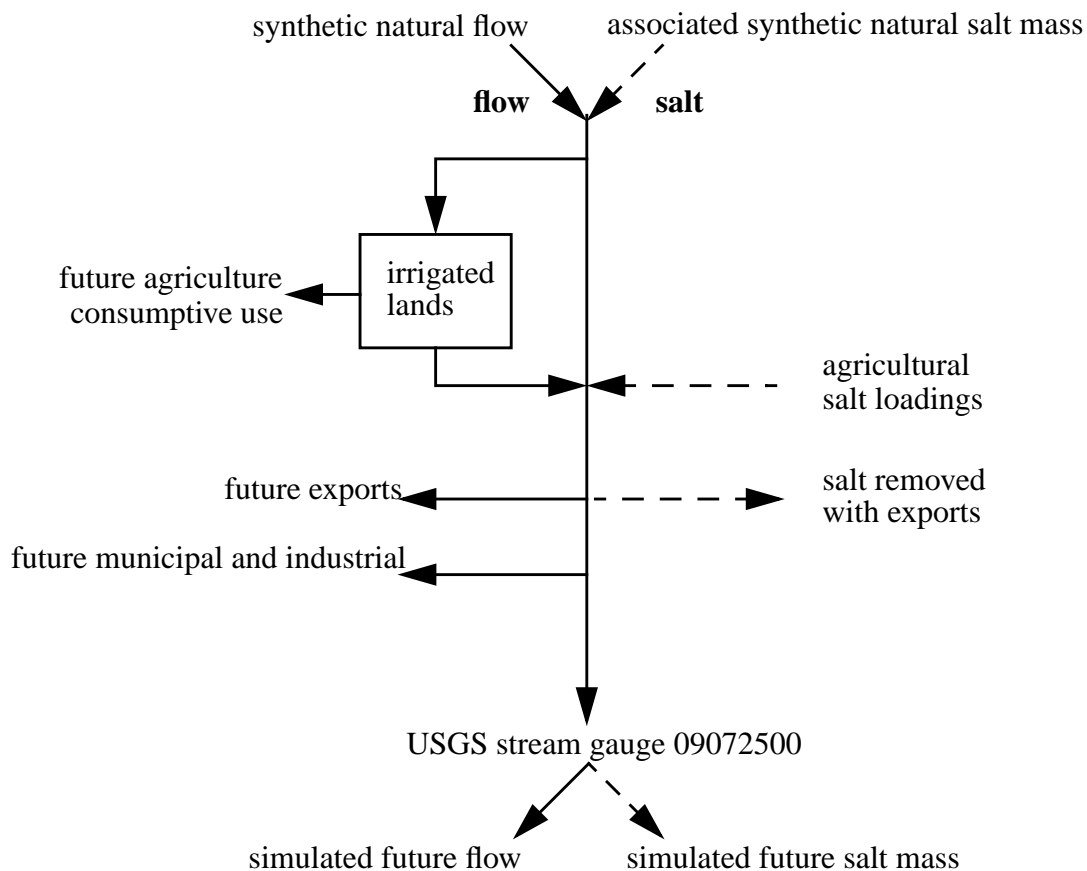


Figure 40: Line diagram of the CRSS simulation model and data for stochastic planning runs.

4.3 Modified CRSS

To address the limitations of the existing CRSS modeling system, we developed a modified stochastic natural flow model and statistical nonparametric natural salt model. A description of these models follows. We used both models to generate data for the CRSS simulation model. We compared the results from the simulation model to results from simulations using the ISM natural flow and USGS natural salt regression model described previously.

4.3.1 Modified Stochastic Natural Flow Model

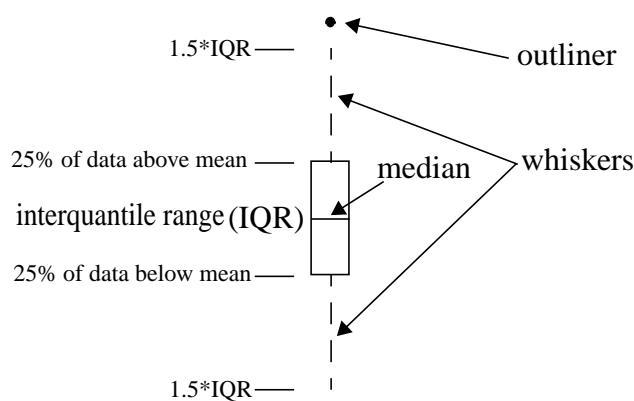
As stated many times, a drawback of ISM is that it cannot generate synthetic sequences that did not occur in the past record. An alternate nonparametric method, K-NN, eliminates this drawback. Traditional K-NN methods resample values from the historic time series, one at a time (Lall, 1995). Because samples are not taken as blocks, as in the ISM, this technique is able to produce time series sequences that did not occur in the historic data. However, values are limited to those in the data set. We further developed the traditional K-NN technique to provide the ability to create values not seen in the historic record. Our modified K-NN method developed a regression relationship between successive months' flows and saved the residuals from the regression. The succeeding month's flow was first calculated from the appropriate regression. Then, the k-nearest flows to the flow from the regression were found. A residual from the k-nearest flows was resampled and added to the flow from the regression to produce a new value. This scheme allowed the K-NN method to perturb the historic data within its representative neighborhood and allows extrapolation beyond the sample, while maintaining the residuals of the data. The method is described in detail in Chapter 2, "Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow."

Like the ISM, the modified K-NN has the advantage of nonparametric models; there is no need to transform the data to fit an assumed probability density function. It has the further advantage of being able to generate synthetic time series containing numbers and sequences that have not occurred in the past, but are "statistically faithful" to the original time series.

Chapter 2, "Modified Nonparametric K-NN Model for Generating Stochastic Natural Streamflow," demonstrated that the generated synthetic time series of natural flow PDFs preserved the PDF of the historic data. We used a PDF with boxplots of the annual synthetic natural flow to validate the modified stochastic natural flow model.

The monthly time series were summed to an annual time series because the salinity standard is based on an annual time step.

We used the K-NN flow model to generate one hundred 55-year natural flow traces based on the monthly calculated natural flows from 1941 to 1995. We chose the 1941-1995 period because of the availability of both observed historic flow and salt data to verify the model results. Figure 41 shows a plot of the annual calculated natural flow PDFs from 1941 to 1995 as the solid line. The boxplots display the interquartile range (IQR) and whiskers extending to $1.5 * IQR$ for the PDFs of the 100 synthetic natural flow traces. The interquartile range (the box) indicates the range for 50 percent of the data around the mean. The horizontal line inside the IQR depicts the median of the data. The whiskers approximate the 5 percent and 95 percent confidence for the traces. Data beyond the whiskers ($1.5 * IQR$) are termed outliers and indicated by a solid circle. An example boxplot is given:



Each plot in Figure 41 shows that the calculated natural flow PDF fell within the interquartile range of the boxplots. Hence, the model generated synthetic data that is statistically consistent with the historic data.

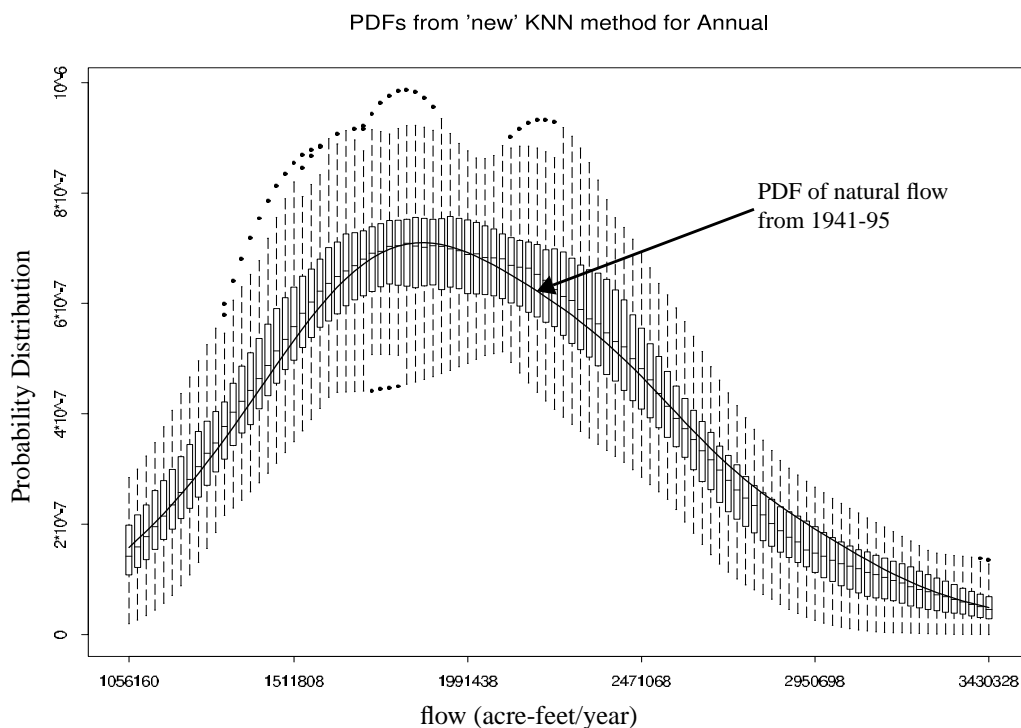


Figure 41: The solid line shows the PDFs for the calculated natural flow from 1941 to 1995. The boxplots describe the interquartile range (IQR) and whiskers for approximately 5 percent and 95 percent from the PDFs of the 100 synthetic natural flow traces. The plot represents a single run of the KNN flow model. The run preserved the calculated natural flow in the interquartile range over all flows.

Unlike the salinity data, the flow data was not limited to the time period 1941 to 1995. Calculated natural flow data was available from 1906 to 1995 at this gauge. Using the full time period increased the sample size for the stochastic model, increasing the certainty of the model results. In Figure 42, the solid lines show the PDF for the calculated natural flow from 1941 to 1995, while the dotted line represents the PDF for the calculated natural flow from 1906 to 1995. Note that the PDF for the longer time period is different from the PDF for the shorter period. When the longer period was used to generate synthetic natural flows, the boxplots from the 100 simulations follow the PDF for the calculated natural flow from 1906 to 1995, not the PDF for the shorter period.

The boxplots in the upper graph represent the 100 synthetic natural flow PDFs

from the modified K-NN stochastic natural flow model, and, in the lower graph, 100 synthetic natural flow PDFs generated from the ISM. ISM performs best with an extensive historic record, therefore, we compared ISM to the modified KNN flow model using the full record, allowing ISM to produce its best results. Using the time period 1906 to 1995, ISM generated 90 simulations of synthetic natural flow.

The main disadvantage of ISM is that it cannot generate synthetic data that is statistically possible but has not occurred in history. As a result, ISM produced a narrower range of probabilities at each given flow than the modified K-NN flow model. This narrower range is indicated by the longer boxplots produced by the modified K-NN flow model, which can generate synthetic stochastic data that is statistically possible but has not occurred in history.

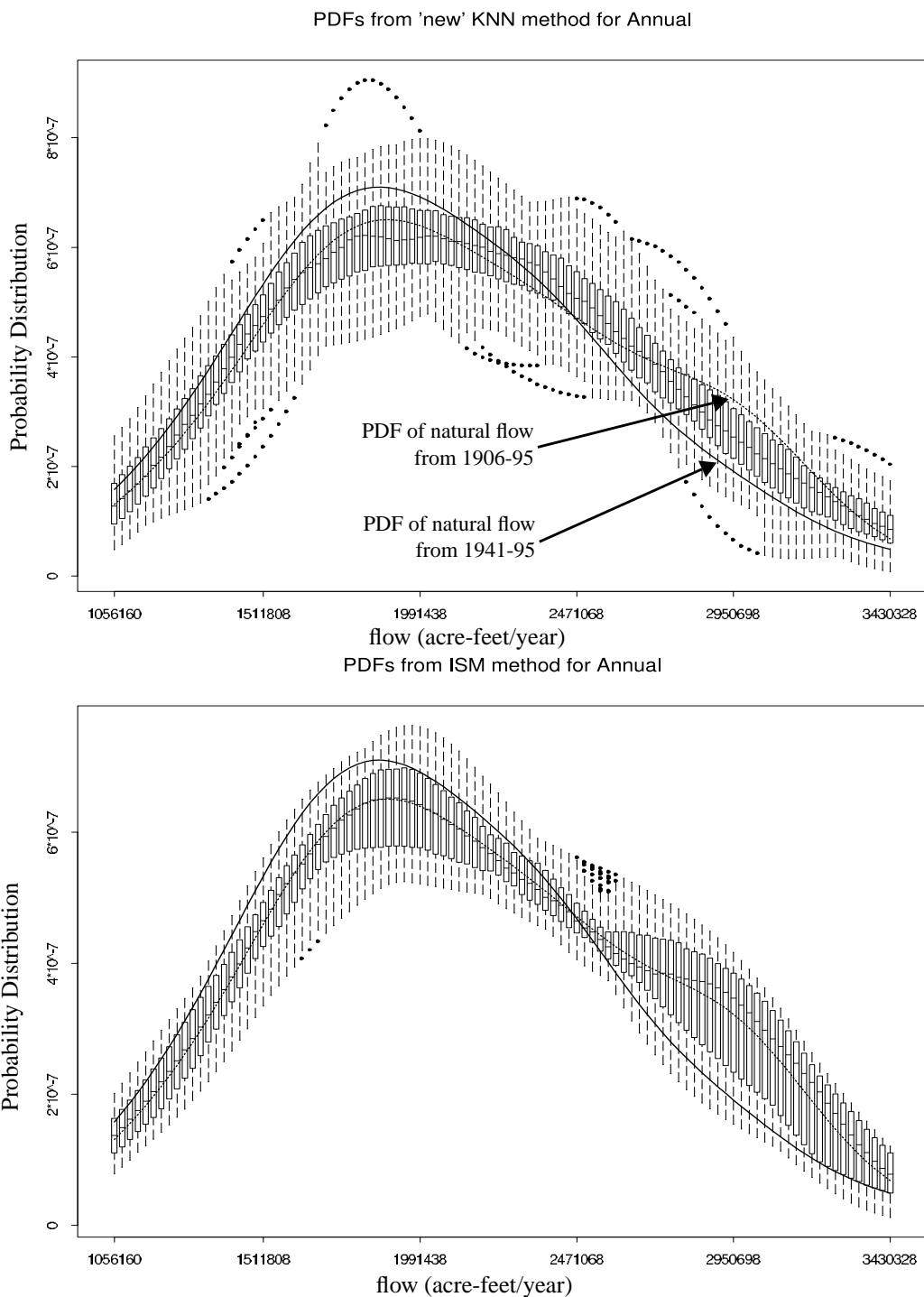


Figure 42: The upper graph shows the PDF for K-NN synthetic natural flow. The lower graph shows ISM synthetic natural flow. Both are based on calculated natural flow from 1906 to 1995. In both graphs the solid line represents the PDF for the calculated natural flow from 1941 to 1995, while the dotted line represents the calculated natural flow from 1906 to 1995. The boxplots represent the PDFs of 100 simulated K-NN and 90 simulated ISM synthetic natural flows. The K-NN model simulated a wider range of flows.

4.3.2 Statistical Nonparametric Natural Salt Model

To replace the USGS model, we developed a statistical nonparametric natural salt model that computes a natural salt mass given a natural flow, either historic or generated by the K-NN flow model. The computation is achieved using a nonparametric local regression fit to a scatter plot of calculated natural flow versus calculated natural salt, both from 1941 to 1995. The calculated natural salt is calculated from historic gauged data and salt load data from the CRSS simulation model as follows:

$$\begin{aligned} \text{calculated natural salt} = & \text{observed historic salt} \\ & + \text{salt with water exported out of the basin} \\ & - \text{salinity pickup from agriculture (values} \\ & \quad \text{based on CRSS)} \end{aligned}$$

The value computed from the local regression is perturbed with a local residual chosen by a K-NN technique that resamples a residual from the nonparametric regression.

A detailed description of the statistical nonparametric natural salt model and its subsequent validation are in Chapter 3, “Statistical Nonparametric Model for Natural Salt Estimation.” To validate the statistical nonparametric natural salt model, we split the estimated natural salt time series into fitting and subsequent future projection periods. The fitting period was from water year 1941 to 1985. The validation period was from water year 1986 to 1995.

The upper graph in Figure 43 shows the annual calculated natural salt mass PDF from 1941 to 1995 as a solid line. The boxplots represent the PDFs of the 100 simulations of monthly synthetic natural salt mass, summed to annual. The statistical nonparametric natural salt model, using synthetic natural flows, generated the monthly synthetic natural salt mass. The calculated natural salt mass did not fall within the interquantile range of the synthetic natural salt mass. We hypothesized that this is a result of certain monthly regression data that have outliers, weakening the regression relationship. To verify our hypothesis, we developed a regression of annual calculated

natural flow as a function of annual calculated salt mass. We used the regression to generate the annual synthetic natural salt as a function of the annual synthetic natural flow, where the monthly synthetic natural flows were summed to annual. The lower graph in Figure 43 shows the PDFs of using the annual regression, rather than summing the results from 12 monthly regressions. The interquartile range of the simulation boxplots encompassed the PDF of the calculated salt mass over the entire range.

These findings show that an annual regression model best preserves the annual natural salt PDF. Unfortunately, CRSS requires entering natural flow and salt data at a *monthly* time step to accommodate the operational rule set. Work to move the operational ruleset to an annual time step is being considered. When the work is complete an annual regression salt model will best preserve the natural salt mass PDF, avoiding the summation of monthly salt mass values.

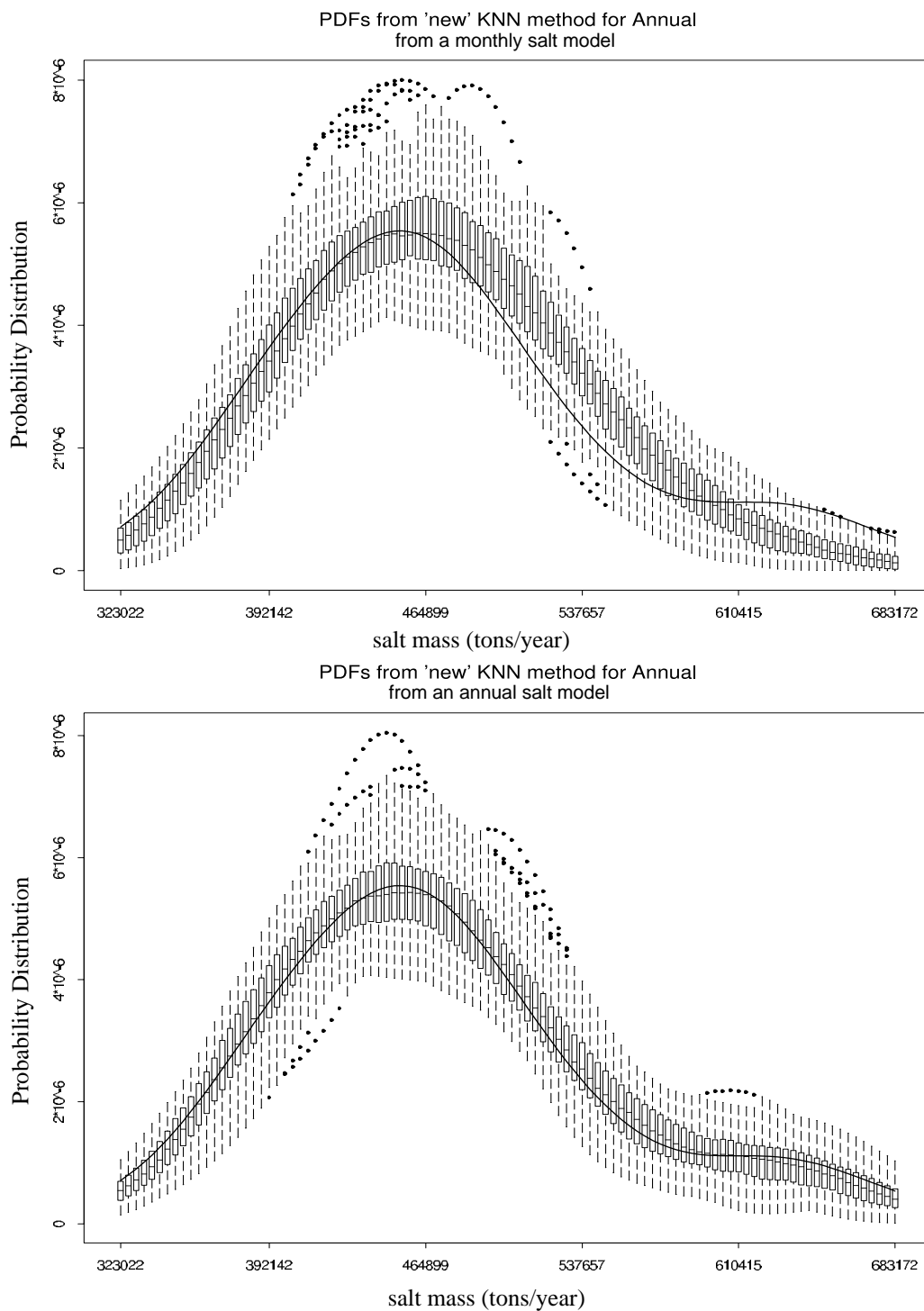


Figure 43: The solid line shows the PDFs for the calculated natural salt from 1941 to 1995. The boxplots describe the interquartile range (IQR) and whiskers for approximately 5 percent and 95 percent of the 100 synthetic natural salt traces. The upper graph uses a monthly statistical nonparametric salt model. The lower graph uses annual statistical nonparametric salt model, preserving the tail of the calculated natural salt PDF.

The statistical nonparametric natural salt model was fitted with the calculated natural salt mass data, from 1941 to 1995. Salt mass data before 1941 was not available. This regression relationship can be applied to estimate (find) natural salt mass for dates outside the 1941 to 1995 range. The statistical nonparametric natural salt mass model was used to compute natural salt associated with one hundred 90-year traces of synthetic natural flows based on the 1906 to 1995 calculated natural flows. In the upper graph of Figure 44, the solid line shows the PDF for calculated natural salt from 1941 to 1995. (A PDF for calculated natural salt from 1906 to 1995 was not available because there is no historic salt mass data before 1941.) The interquartile range of the 100 synthetic natural salt mass traces did not include the calculated natural salt mass, shown as a solid line, over the entire range. This resulted from using synthetic natural flows based on calculated natural flows from a longer time period than the period of the calculated natural salt mass.

The lower graph shows the results using the ISM to generate 90 traces of synthetic natural flows and USGS regressions to find the associated salt mass. The solid line shows a significant shift in the peak of the boxplots compared to the PDF for calculated natural salt mass from 1941 to 1995 (solid line). This shift reveals that the USGS regression gives a higher salt mass for a given flow than our statistical nonparametric natural salt model.

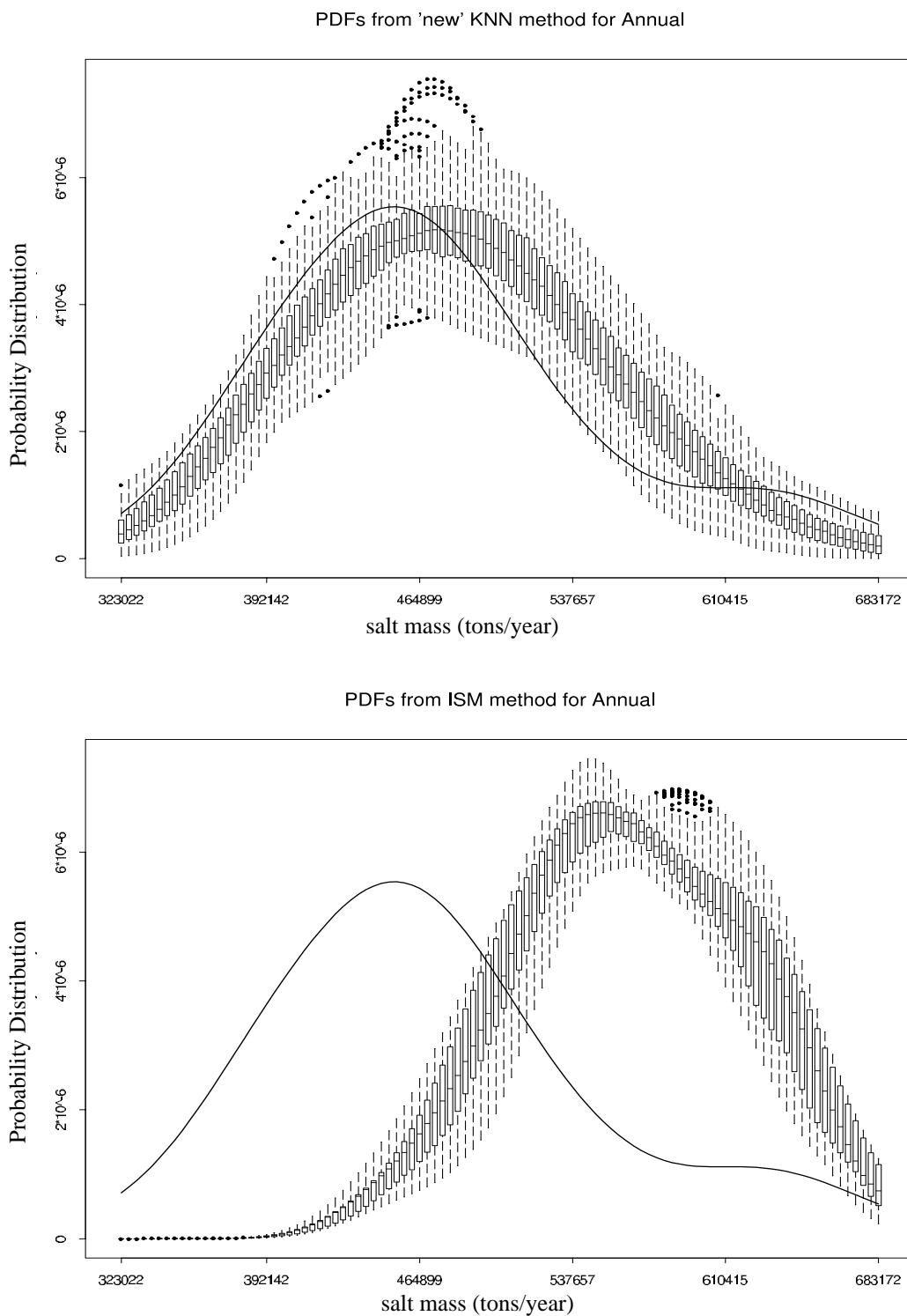


Figure 44: The upper graph shows the PDF for the modified CRSS synthetic natural salt, while the lower graph shows the existing CRSS synthetic natural salt mass. It is evident that the USGS regressions used in the existing CRSS overestimate the solid line that shows the calculated natural salt mass PDF.

Although salt mass is typically the modeled value of interest, concentration is the regulated value and the measured value. Thus, the model should be validated for concentration. Salt concentration is a function of both flow volume and salt mass given by:

$$\text{salt concentration (mg/L)} = \frac{\text{salt mass (tons)} \times 735.29}{\text{flow volume (ac-ft)}} \quad \text{Eq. 4.11}$$

where 735.29 is a conversion factor that converts tons/acre-foot to mg/L. In Figure 45, the upper graph displays the *calculated natural salt concentration* PDF as the solid line, with boxplots from PDFs of 100 synthetic natural concentration traces. We found calculated natural salt concentration using equation 4.11 with the calculated natural salt and natural flow. We found the 100 *synthetic natural salt concentration* traces using equation 4.11 and the 100 monthly synthetic natural flow and salt mass traces described previously. We were not able to preserve the calculated natural salt concentration PDF in the interquartile range over all flows. Again, we hypothesized that this is a result of summing the results from the monthly salt regressions. In the lower graph, using a single annual salt regression preserved the calculated natural salt PDF in the interquartile range of the 100 synthetic natural salt concentration PDFs. This result further indicates the importance of developing a simulation model that can simulate natural flow and salt data at an annual time step.

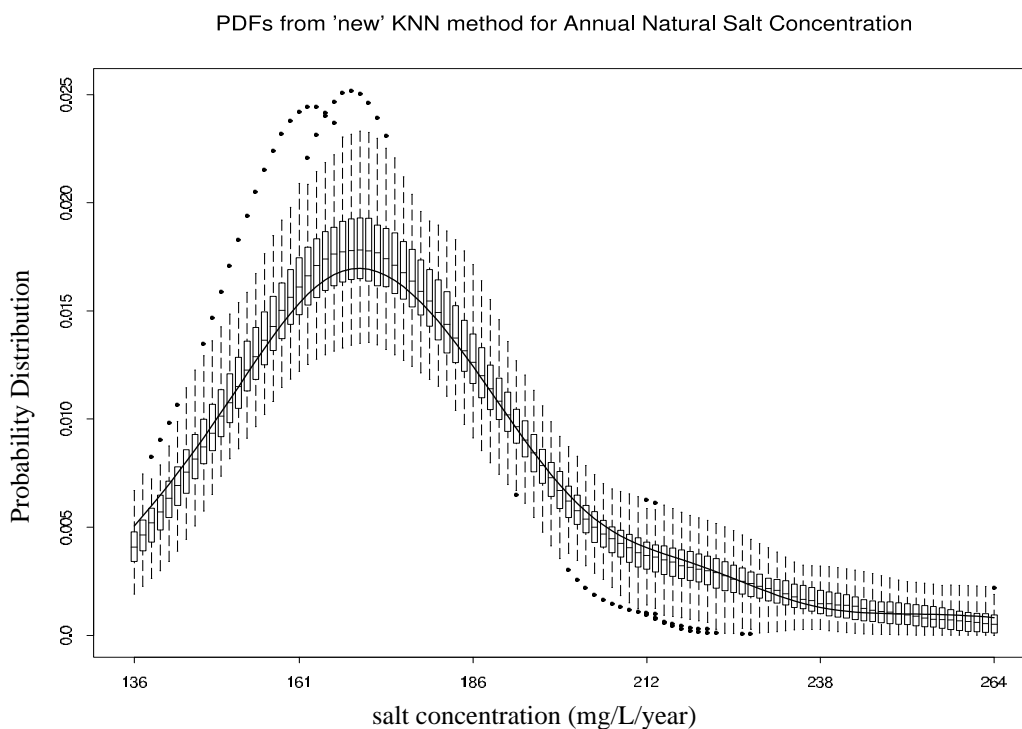
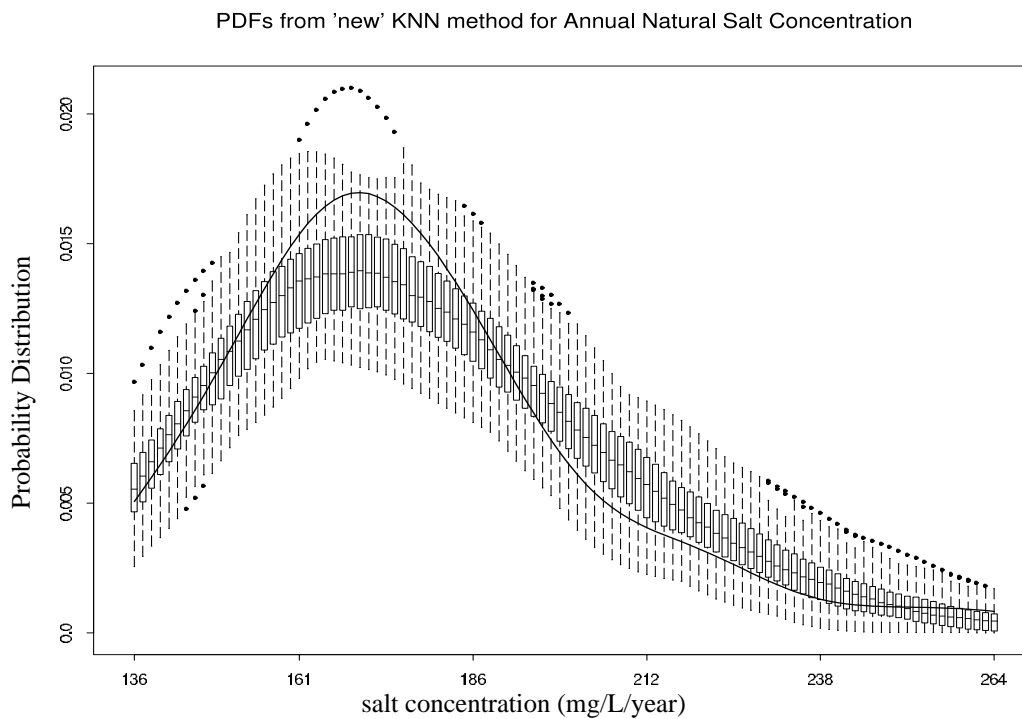


Figure 45: The upper graph shows the PDF for synthetic natural salt concentration from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the synthetic natural salt concentration from a single annual regression to calculate synthetic natural salt mass. Using the single annual regression allowed the simulation to preserve the calculated natural salt concentration (solid line) in the interquantile range.

Using 100 synthetic natural flow traces based on the 1906 to 1995 calculated natural flow and the associated synthetic natural salt mass from the statistical nonparametric natural salt model, we calculated 100 synthetic natural salt concentration traces. Figure 46 shows the calculated natural salt concentration PDF from 1941 to 1995 as a solid line, with the boxplots of the 100 synthetic natural salt concentration PDFs.

The natural flows from 1906 to 1995 were generally higher than flows from 1941 to 1995, causing the synthetic salt concentration to be lower than the calculated natural salt concentration from 1941 to 1995. The lower graph of Figure 46 shows the results using the existing CRSS models to generate 90 synthetic natural salt concentration traces. Figure 46 shows that the boxplots from the PDFs of 90 synthetic traces peaked at a higher salt mass than the calculated natural salt concentration PDF, the solid line, from 1941 to 1995. Again, this higher salt mass occurred because the synthetic natural salt mass from the existing CRSS model is greater than the calculated natural salt mass, as shown in Figure 44.

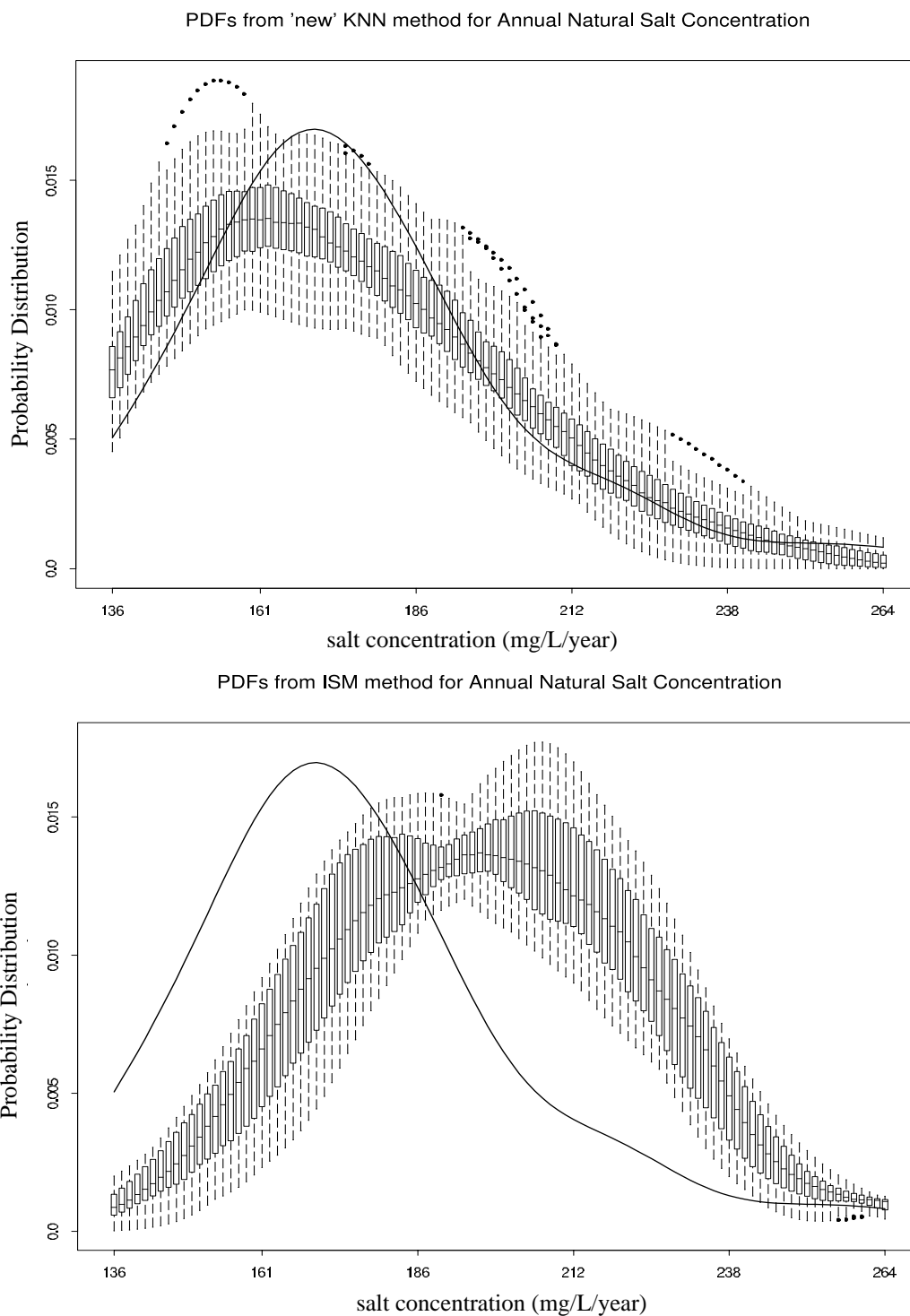


Figure 46: The upper graph shows the PDF for the modified CRSS synthetic natural salt concentration. The lower graph shows the existing CRSS synthetic natural salt concentration. The existing CRSS overestimated the calculated natural salt concentration (solid line), while the modified CRSS slightly underestimated the calculated natural salt concentration.

The code for both the modified stochastic natural flow model and the statistical nonparametric natural salt model are included in Appendix C, “Sample Splus Code.”

4.3.3 Simulation Model

The final step in our modeling system used the CRSS simulation model as implemented in RiverWare (Zagona, 2001). The USBR re-implemented the CRSS simulation model in RiverWare to simplify changing operational policies.

To test the modified stochastic nonparametric natural flow model, we used a segment of the CRSS simulation model that includes USGS gauge 09072500 (Colorado River near Glenwood Springs, CO). The modified stochastic natural flow model and statistical nonparametric natural salt model generated inputs for the CRSS simulation model. The inputs included calculated or synthetic natural flow and associated natural salt mass. Additionally, human-induced depletions and salt loading were entered in the CRSS simulation model, as explained in Section 4.2, “Existing CRSS.” Total depletions are a sum of depletions from agriculture, municipal and industrial sources, exports, and reservoir regulation. Human-induced salt results from agricultural salinity pickup and the salt removed with exports. We concluded that reservoir regulation has minimal effect on salt at an annual scale; therefore, reservoir regulation did not model salt. The time period simulated by the model dictated the depletion and salt loading data used in the run. For example, if the run simulated the historic time period from 1941 to 1995, the model used the depletions and human-induced salt loadings from the 1941 to 1995 historic record. If the model simulated the future time period of 2002 to 2062, it used projected depletions and salt loadings.

CRSS performs Monte Carlo simulations by running each trace of synthetic natural flow and associated salt time series through the CRSS simulation model. Our modified CRSS used 100 synthetic natural flow traces and associated salt mass traces that were each run through the simulation model, one at a time, calculating 100 simu-

lated historic or future flows and associated simulated salt mass and concentration time series. These simulations could be used to approximate the PDF for the observed historic or predicted future flow, salt, and concentration.

We first used the CRSS simulation model to validate the modified CRSS system by showing that it could preserve the observed PDF for historic flow, salt mass, and concentration from 1941 to 1995. The inputs were 100 traces of synthetic natural flow and associated salt mass traces generated with the modified stochastic natural flow model, using the 1941 to 1995 calculated natural flow, and the statistical nonparametric natural salt model, using the synthetic natural flow traces generated from the modified stochastic natural flow model. The total depletions were based on 1941 to 1995 historic records.

The results were compared to the observed historic PDF, as discussed previously. Figure 47 displays the observed historic flow PDF from 1941 to 1995 as a solid line, with boxplots describing the PDFs of the 100 simulations of simulated historic flow from the simulation model. The shape of the observed historic flow PDF was preserved well; however, the peak of the PDFs from the simulated historic flows was subdued and the tails were elongated.

Figure 48, upper graph, displays the observed historic salt mass from 1941 to 1995 as a solid line, with boxplots describing the PDFs of the 100 simulations of simulated historic salt mass. The simulated historic salt mass could not preserve the observed historic salt mass in the interquantile range across all flows. Again, we hypothesized that this was a result of summing the results from 12 monthly natural salt regression models to calculate an annual salt mass.

Subsequently, we created a simulation model equivalent to the CRSS simulation model in S-Plus programming language. S-Plus is a statistical programming language that simplifies designing and modeling statistical models. The S-Plus simulation model allowed the synthetic natural flow and associated salt mass traces to be input

and modeled at an annual time step. Modeling at an annual time step was possible for this model because no major river reservoirs occur in the case study area. Major river reservoir operating policy is expressed at a monthly time step. With the S-Plus simulation model, the statistical nonparametric natural salt model used a single annual regression to compute natural salt mass associated with natural flow. The lower graph of Figure 48 shows the results of using the single annual regression rather than summing the results from 12 monthly regression. The simulated historic salt mass PDFs, shown with the boxplots, preserved the observed historic salt mass PDF in the inter-quantile range over all flows.

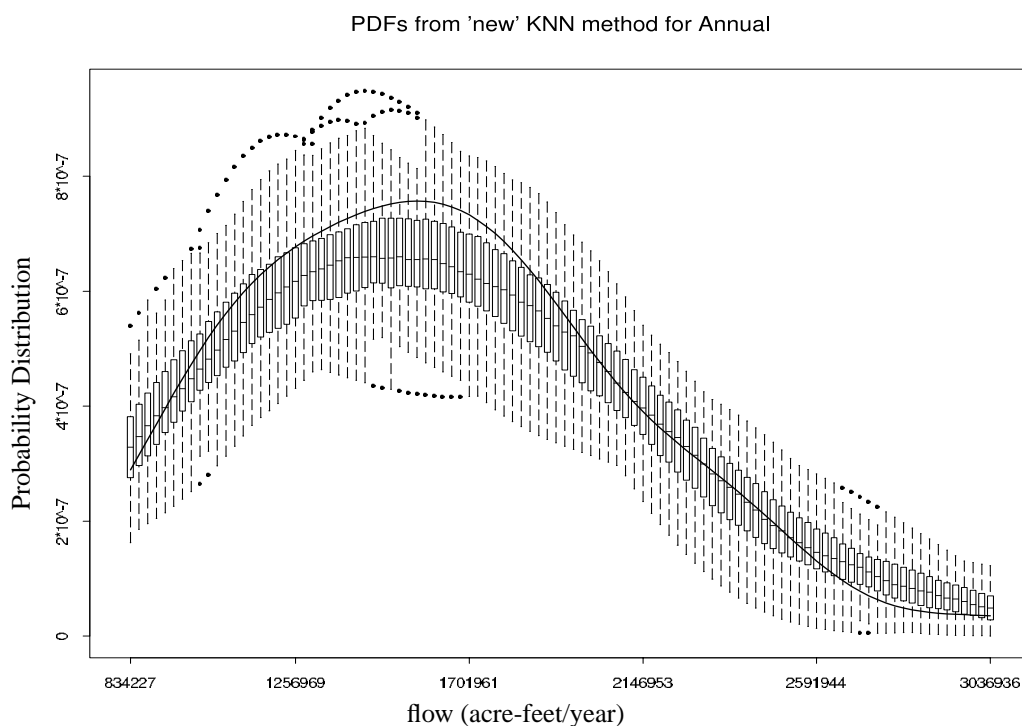


Figure 47: The PDF for observed historic flow, solid line, with boxplots of the simulated historic flow from the modified CRSS. The modified CRSS simulations have a subdued peak and elongated tails compared to observed historic flows.

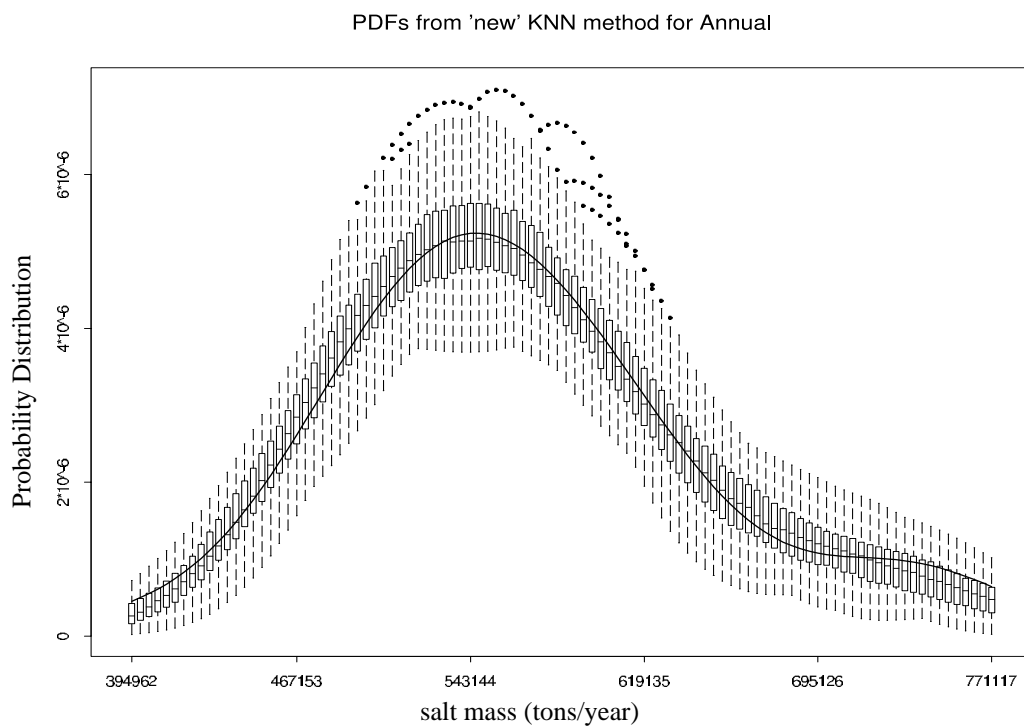
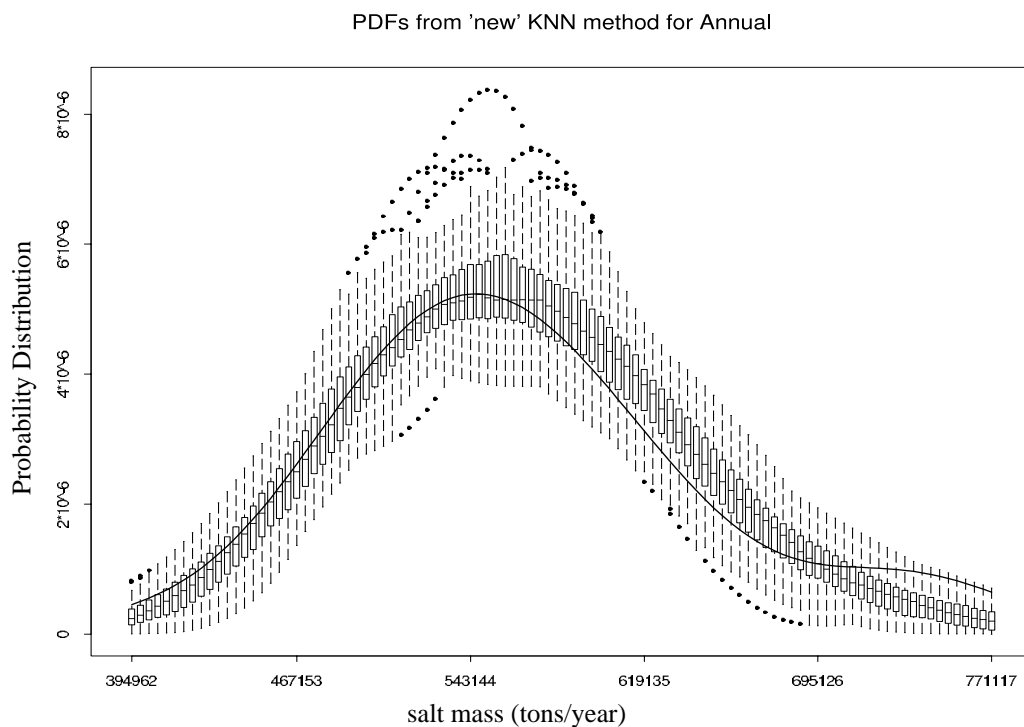


Figure 48: The upper graph shows the PDF for simulated historic salt mass from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the simulated historic salt mass from a single annual regression to calculate synthetic natural salt mass. Using the single annual regression allowed the simulation to preserve the observed data in the interquartile range.

Figure 49, upper graph, displays the observed historic salt concentration PDF from 1941 to 1995 as a solid line, with boxplots describing the PDFs of the 100 simulated historic salt concentration traces representing 1941 to 1995. The simulated historic salt concentration could not preserve the observed historic salt concentration in the whiskers of the boxplots for all salt concentrations. To verify if this was a result of summing the salt mass from the 12 monthly regressions to derive annual salt mass values, we used the S-Plus simulation model results to calculate the annual simulated historic salt concentration. As seen in Figure 49, lower graph, the simulation still could not preserve the observed historic salt concentration in the whisker of the boxplots for all salt concentrations. Both the RiverWare and S-Plus simulation models simulated a much wider range of salt concentrations than observed concentrations.

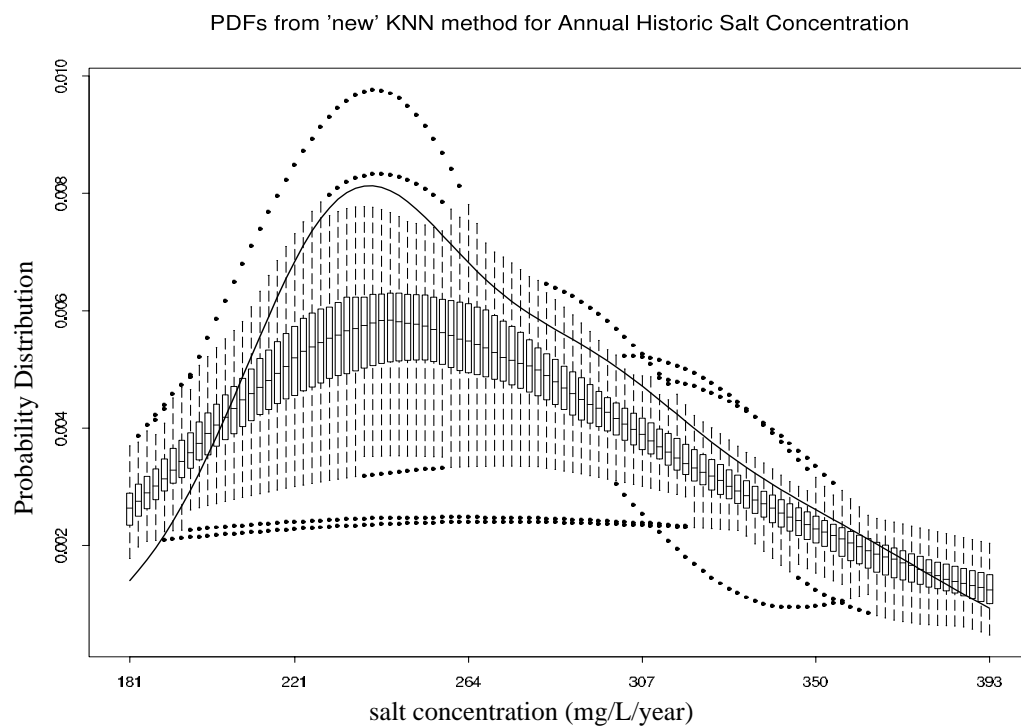
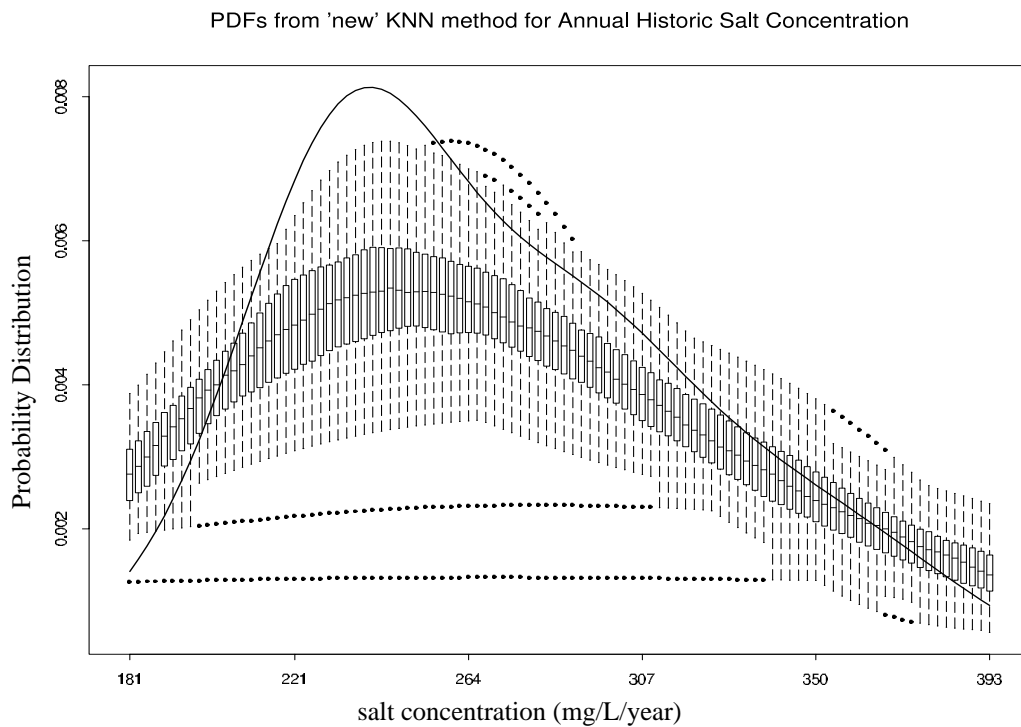


Figure 49: The upper graph shows the PDF for simulated historic salt concentration from 12 monthly regressions to calculate synthetic natural salt mass. The lower graph shows the simulated historic salt concentration from a single annual regression to calculate synthetic natural salt mass. Neither preserved the observed data in the inter-quartile range.

To explain why the observed historic salt concentration PDF could not be preserved, we investigated how the simulated historic salt concentration is calculated. Simulated historic salt concentration is a function of simulated historic flow and simulated historic salt mass. From the previous figure (Figure 48), we see that the annual salt regression model preserved the historic salt mass well, but the simulated historic flow shows a subdued peak in the PDF, indicating a wider range of flows than the observed historic flow. The wider range of flows produced a wider range of concentrations, as indicated in Figure 49. It is apparent that reproducing historical flows is important in reproducing salt concentration. Simulated historic flows are computed by removing total depletions from the synthetic natural flow. Because the natural flows were well reproduced (Figure 41), it seems the depletions might have influenced the historic flow PDFs, which we investigated further.

We assumed that depletions are constant, but Figure 50 shows that the level of depletions in any given year is a function of the flows for that year. Our first attempt to model simulated historic flow did not consider this relationship. The modified stochastic natural flow model generated 100 synthetic natural flow traces based on 1941 to 1995 calculated natural flow. For each year, the 100 simulations represented a variety of natural flows. To find simulated historic flow, the same historic depletions from 1941 to 1995 were removed from every trace of synthetic natural flow, no matter what flow was generated for each trace.

A more accurate model would consider the total depletion as a function of the natural flow. Figure 50 shows the relationship between total depletions as a function of calculated natural flow from 1941 to 1995. The plot shows a tendency for total depletions to increase with increased natural flow. The scatter of the points shows much variability around the regression.

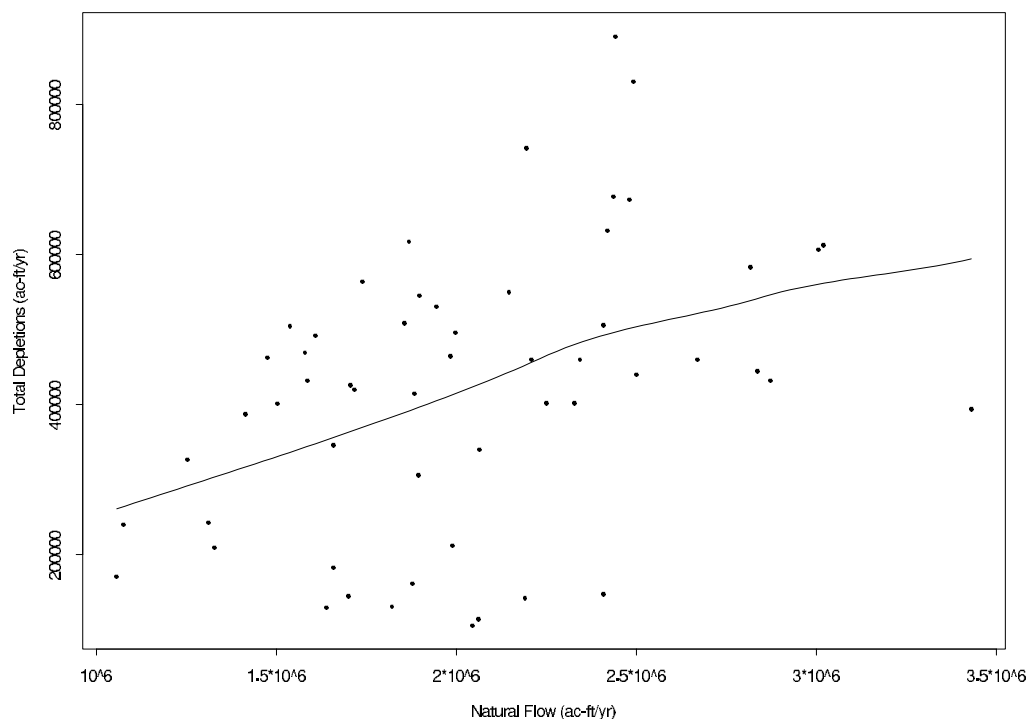


Figure 50: The regression relationship between total annual depletion as a function of the calculated natural flow. The solid line shows a LOCFIT through the data points. There is wide scatter around the regression indicating a wide variance around the regression.

To incorporate the relationship between calculated natural flow and total depletions, we used depletions as a function of natural flow based on the regression shown in Figure 50. We did this by computing a total depletion from the regression for each synthetic natural flow value and removing the appropriate total depletion. We first performed this by calculating the total depletion directly from the regression. The upper graph in Figure 51 shows the observed historic flow PDF (as a solid line) was not preserved in the interquantile range for all the PDFs for the simulated historic flows (as boxplots). The lower graph shows similar results for the simulated historic salt concentration.

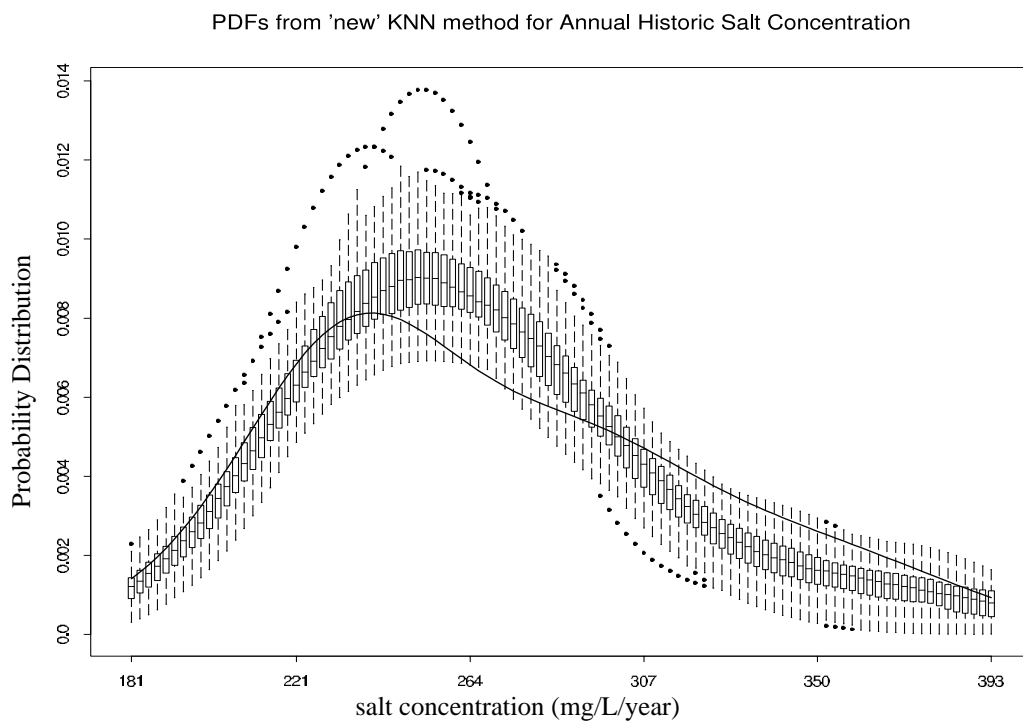
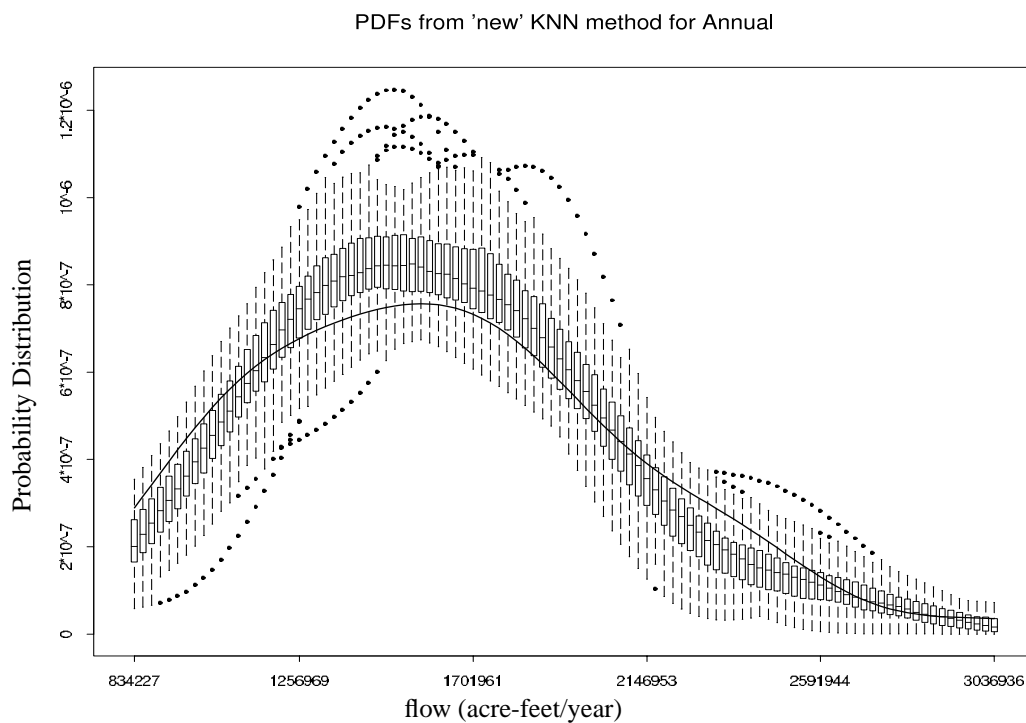


Figure 51: The upper graph shows the PDF for historic flow, while the lower graph shows the PDF for the historic salt concentration. Both graphs used the regression between total annual depletion as a function of calculated natural flow without resampling. Both historic flow and salt concentration are preserved better incorporating the regression.

Adding a K-NN technique to resample the residuals from the neighboring points around the regression point and adding the residuals to the total depletion improved the model results. In Figure 52, upper graph, the simulated historic flow PDFs preserved the observed historic flow PDF in the interquantile range over most flows. The simulated historic salt concentration PDFs, shown in the lower graph, still produced more lower concentrations than were seen historically, but to a lesser degree than when the total depletions were found directly from the regression with no resampling. Generally, we were able to preserve most of the observed historic salt concentration PDFs in the interquantile range of the boxplots for the simulated historic salt concentration PDFs.

This technique, which improves the simulation results when compared to the observed historic salt concentration, cannot be used in the CRSS simulation model until a new set of operational policy rules is developed that can evaluate policy at an annual time step.

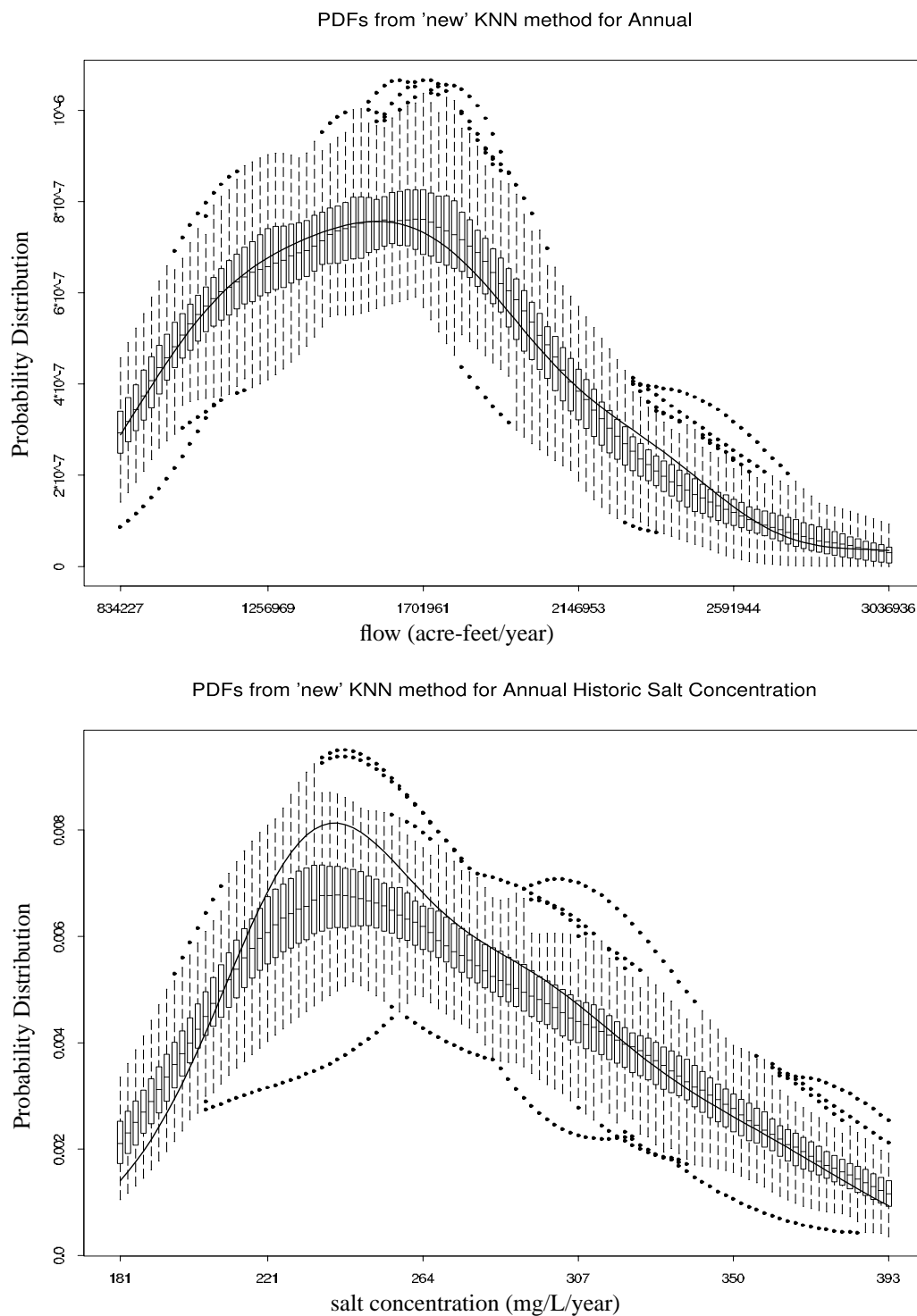


Figure 52: The upper graph shows the PDF for historic flow, while the lower graph shows the PDF for the historic salt concentration. Both graphs used the regression between total annual depletion as a function of calculated natural flow with resampling. Both historic flow and salt concentration are preserved best incorporating the regression with the resampling.

4.3.4 Simulated Policy

To gain an understanding of how using the modified CRSS model could influence policy decisions, we tested policies using some fictional salt mass and concentration standards. As discussed in the opening of this chapter, salinity concentration standards are mandated at three locations in the lower basin of the Colorado River. Our segment of the CRSS simulation model did not include these locations, but developing fictional standards in the modeled segment facilitated comparing the performance of the existing and modified CRSS. We developed standards that occur in the tails of the PDF, where extreme events, such as high salt mass or concentration, occur. For the fictional salt mass standard, we determined the number of times 650,000 tons or more of salt occurred at the simulated gauge. For the observed historic salt mass, this standard was exceeded 6 times. Figure 53 shows the number of times the fictional standards were exceeded by the simulation model that was run at a monthly time step and summed to annual values. The upper graph shows a boxplot for the number of times 100 simulated historic salt mass traces from the simulation model exceeded the fictional standard. The solid circle indicates the number of times the observed historic salt mass exceeded the standard. The simulation model preserved the fictional salt mass standard in the interquartile range. The lower graph shows the number of times a salt concentration standard of 350 mg/L was exceeded and, therefore, violated. The solid circle indicates that the observed historic salt concentration was exceeded 4 times. The simulation model overestimated the number of times the standard was exceeded, as indicated by the interquartile range being well above the number of times the observed historic salinity concentration exceeded the fictional salinity concentration standard.

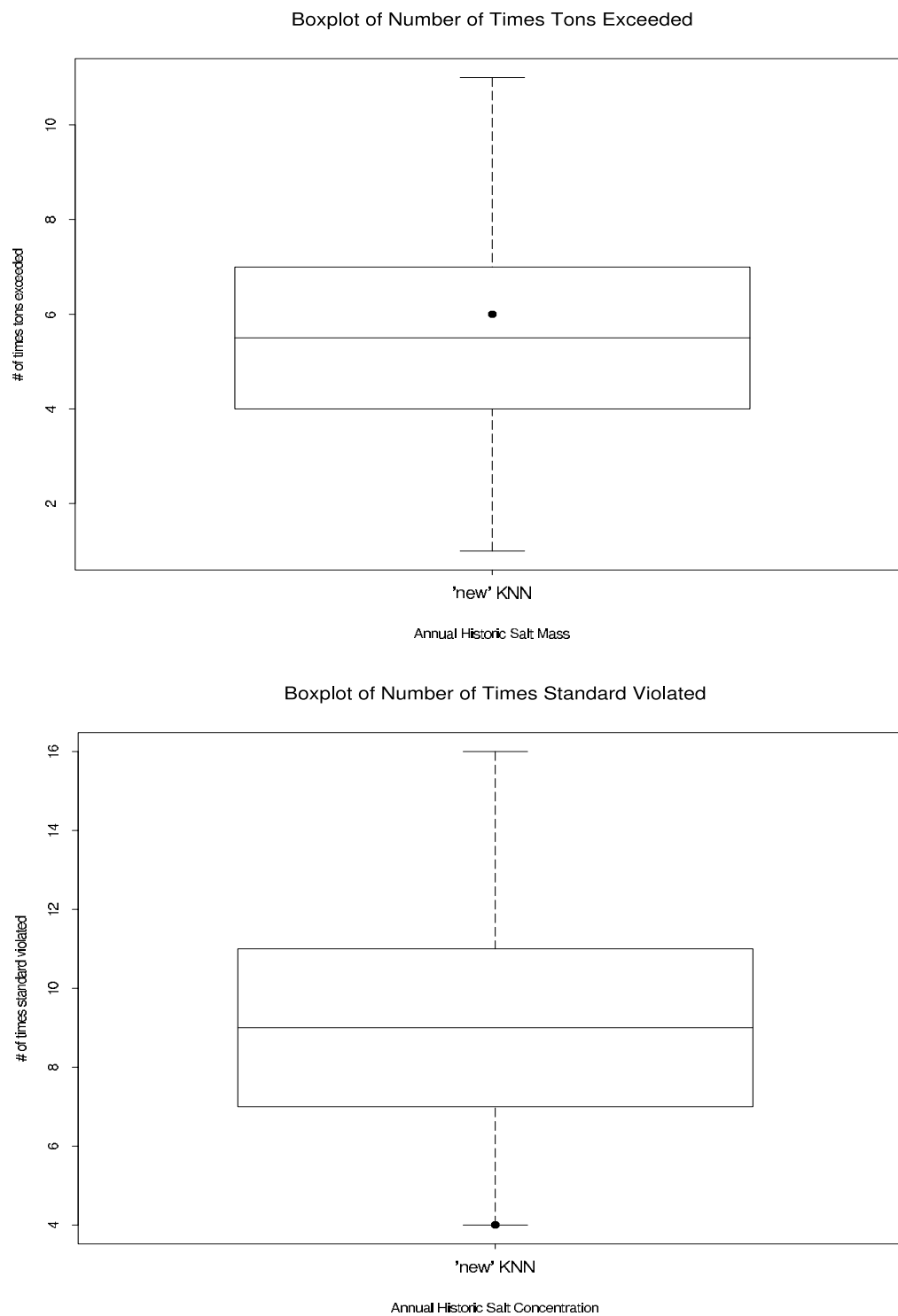


Figure 53: The upper graph shows the number of times a theoretical salt mass target of 650,000 tons was exceeded. The lower graph shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated.

Figure 54 shows how well the annual time step model estimates the number of times the fictional standards were exceeded and, therefore, violated. In the upper plot, the number of times the observed historic salt mass exceeded the standard was within the interquartile range of the simulation model. In the lower plot, the model still overestimated the number of times the observed historic salt concentration standard was violated.

Figure 55 shows the number of times the salinity concentration standard was violated by the annual time step simulation model after the addition of the relationship between calculated natural flow and total depletions with resampling. In this case, the number of violations from the observed historic salt concentration was within the interquartile range of the simulation model results. This is consistent with the PDFs from the simulation model, which did the best job of preserving the PDF of the historic salt concentration.

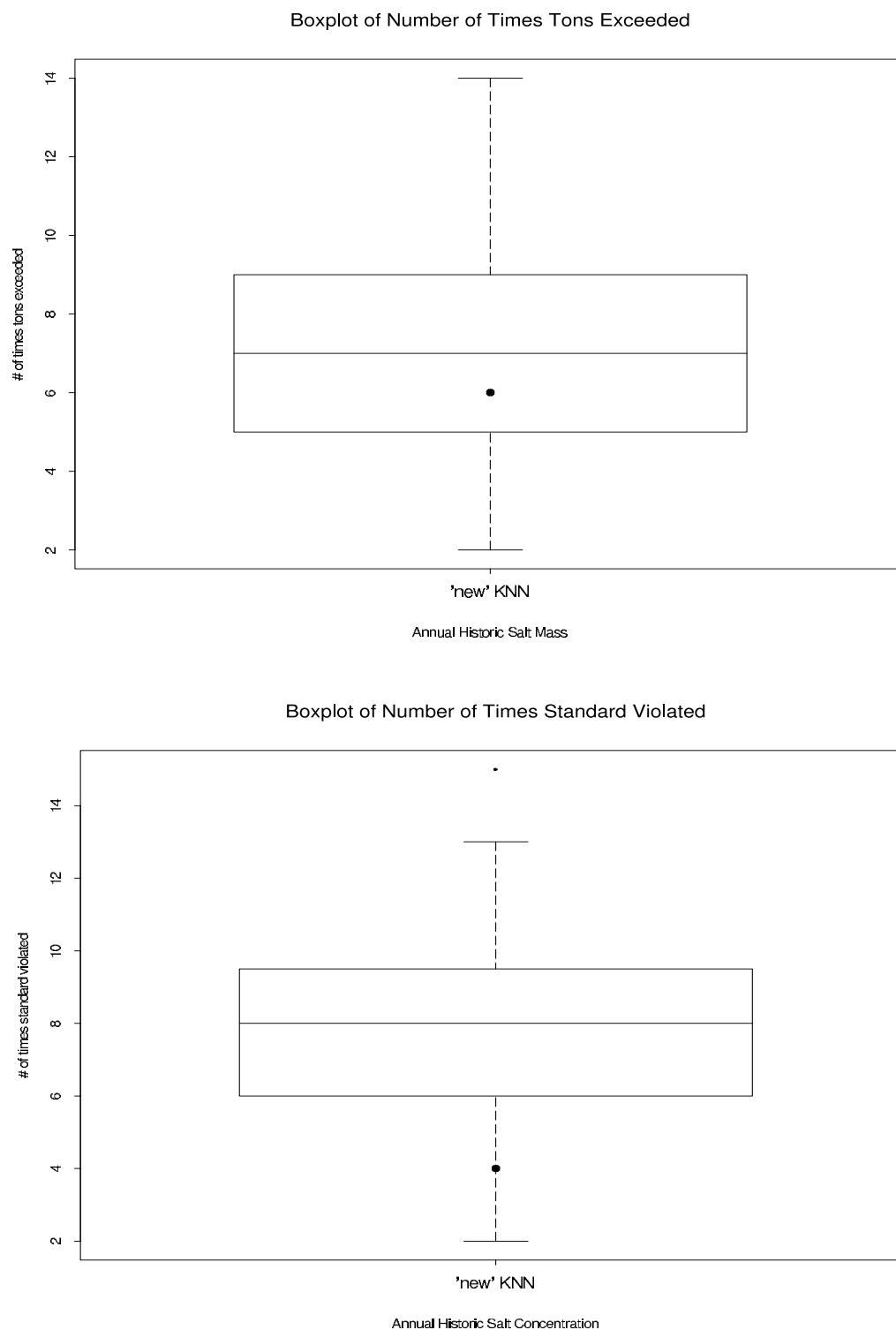


Figure 54: The upper graph shows the number of times a theoretical salt mass target of 650,000 tons was exceeded. The lower graph shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated.

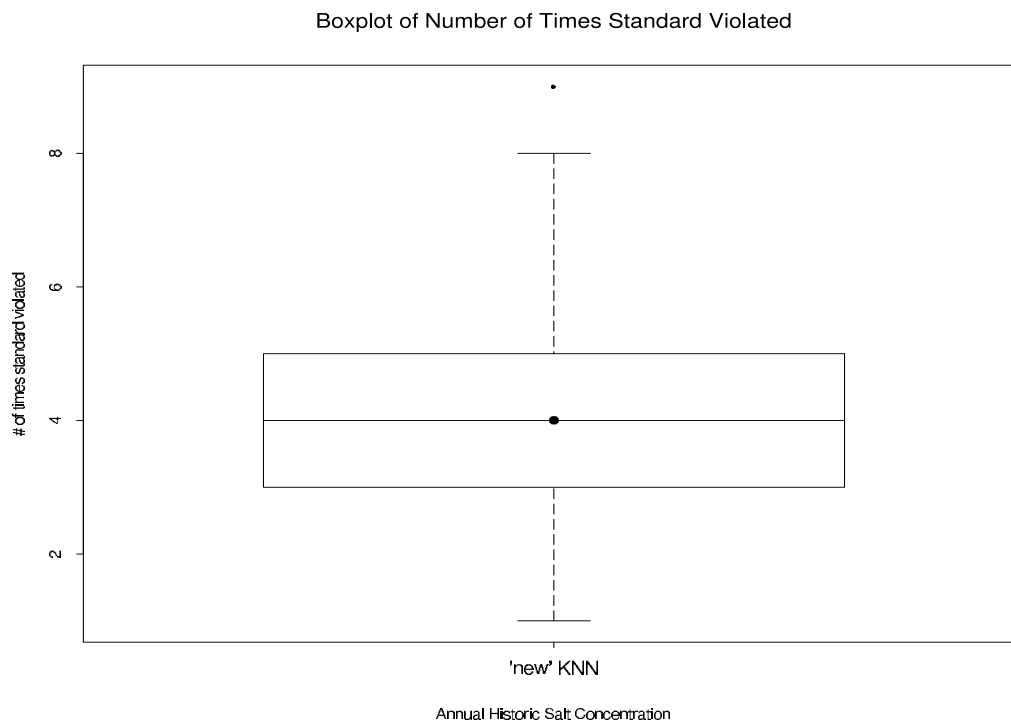


Figure 55: This figure shows the number of times a theoretical salt concentration standard over 350 mg/L was exceeded and, therefore, violated. Incorporating the regression from total depletions as a function of calculated natural flow with residual resampling allowed the simulations to preserve the observed number of violations.

4.3.5 Comparison of the Existing CRSS and the Modified CRSS

The existing CRSS and the modified CRSS were compared by generating the synthetic natural flow traces using the entire calculated natural flow time series. The ISM flow model, in the existing CRSS, performs best using the entire calculated natural flow time series.

We developed the next figures from the results of the monthly time step simulation model. The synthetic natural flow traces of 55 years in length were generated based on 1906 to 1995 calculated natural flow. The associated simulated synthetic natural salt mass was computed using the statistical nonparametric natural salt model, based on calculated natural flow and salt mass from 1941 to 1995. The total depletions were based on historic records from 1941 to 1995. With these inputs, the simulation

model approximated the observed historic flow, salt mass, and concentration from 1941 to 1995 using the most extensive data available.

Figure 56 shows the PDF for observed historic flow from 1941 to 1995 as the solid line. The boxplots describe the PDFs of the 100 traces of simulated historic flow produced by the simulation model. The upper graph shows the results of the modified CRSS, while the lower graph shows the results of the existing CRSS. Neither set of models reproduced the observed PDF in the interquartile range. The existing CRSS could not preserve the observed PDF within the whiskers for all flows, but the modified CRSS preserved the observed PDF. Generating synthetic natural flows based on 1906 to 1995 calculated natural flows did not allow the simulation model to preserve the observed PDF for the time period 1941 to 1995 in the interquartile range. From the plots, it is evident that there are higher calculated natural flows in the years before 1941 because the PDFs from the 100 simulations are skewed towards the higher flows. From our previous analysis, we can also attribute the high PDFs of the simulations to not using a relationship between calculated natural flow and total depletions, which finds a depletion associated with each synthetic natural flow.

A significant difference between the modified and existing CRSS can be seen for the observed historic salt mass, as shown in Figure 57. The upper graph shows results for the modified CRSS, while the lower graph shows the existing CRSS. Because the synthetic natural flows, based on 1906 to 1995 data, are high, natural salt concentrations are low. Figure 58 compares the simulated historic salt mass values generated by the 100 simulations of the modified CRSS and existing CRSS. The graph shows that the existing CRSS generated higher historic salt mass in a narrower range of values than the modified CRSS, which is evident because the median, shown as the horizontal line within the interquartile range, is higher for the existing CRSS than for the modified CRSS. Also, the whisker for the existing CRSS spans a narrower range of values than for the modified CRSS.

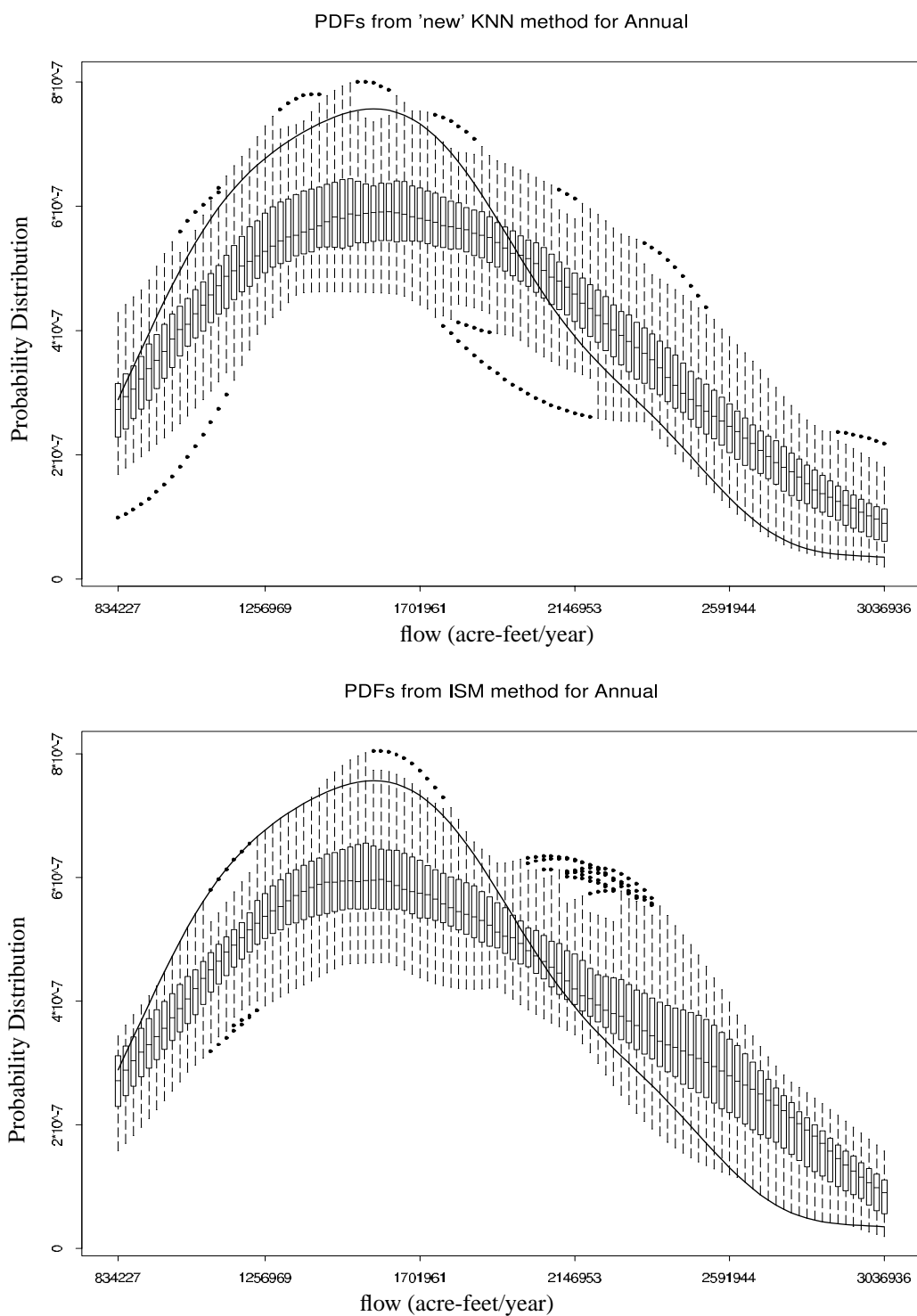


Figure 56: The upper graph shows the PDFs for historic flow from the modified CRSS. The lower graph shows the PDFs for historic flow from the existing CRSS.

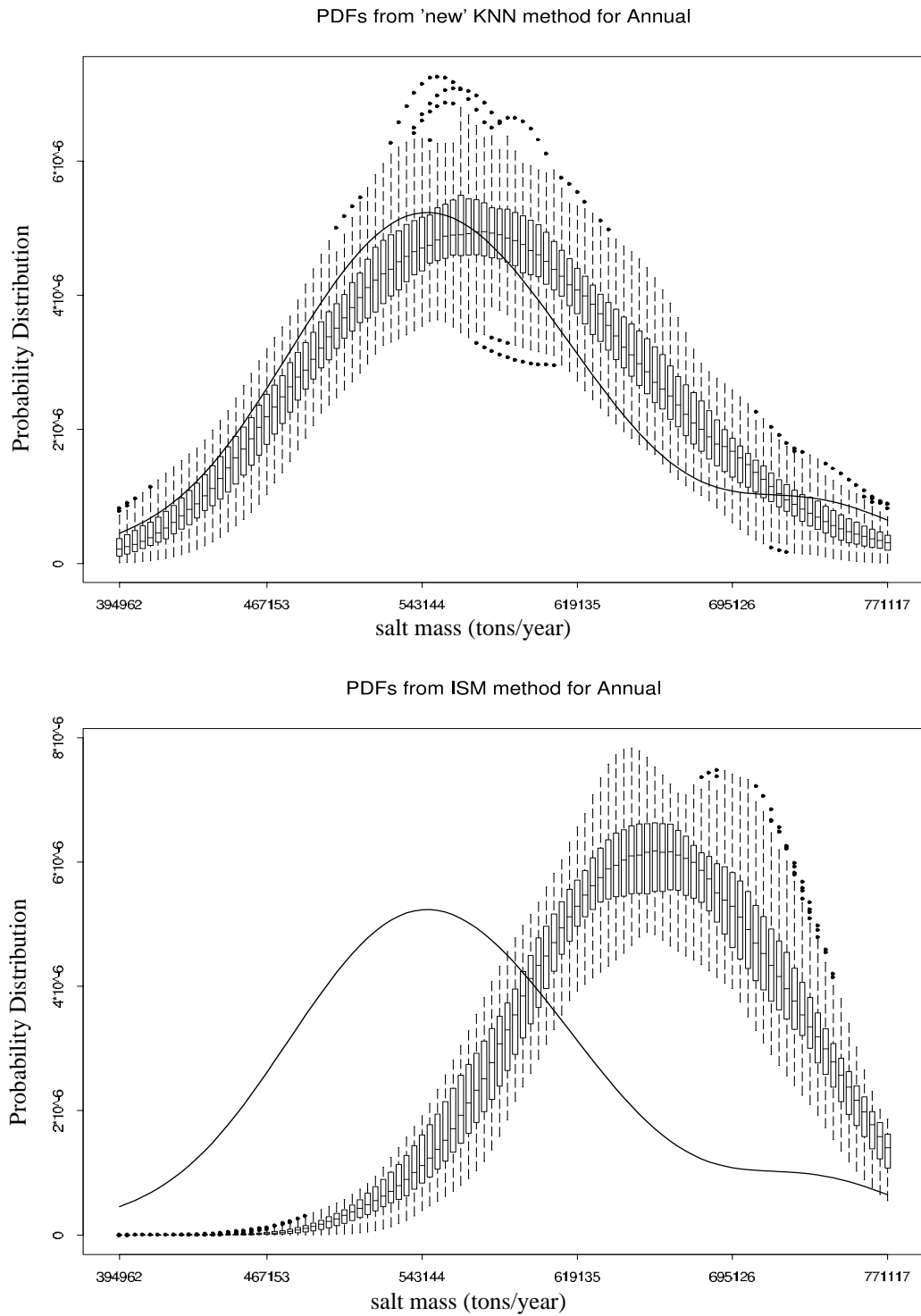


Figure 57: The upper graph shows the PDFs for historic salt mass from the modified CRSS. The lower graph shows the PDFs for historic salt mass from the existing

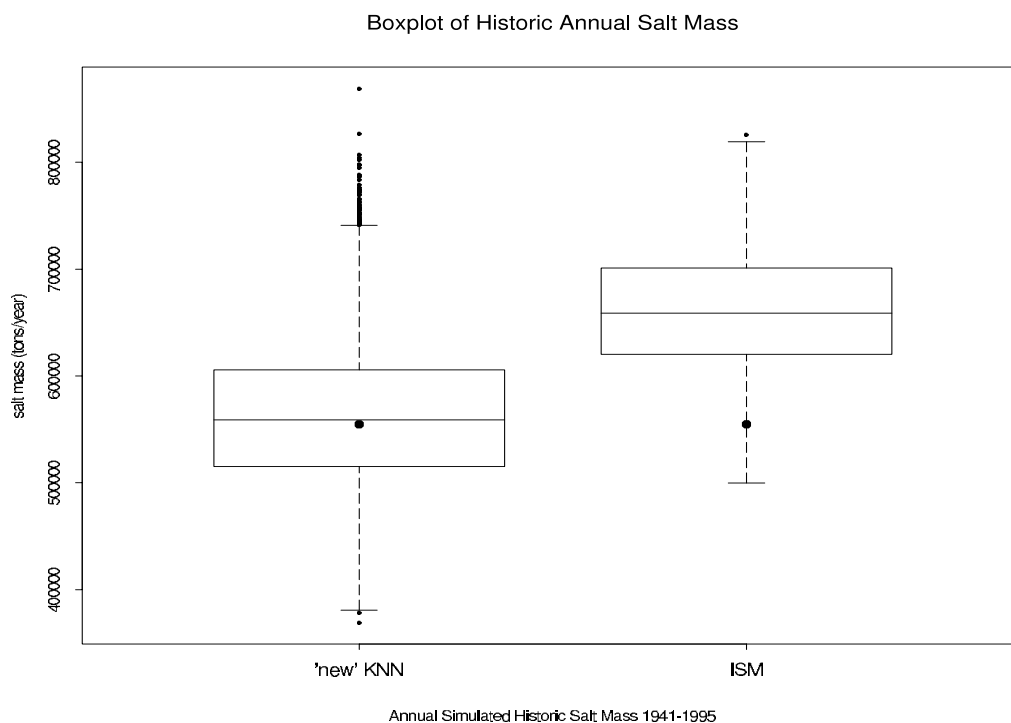


Figure 58: Boxplots of the simulated historic salt mass from each of the 100 simulations. The modified CRSS generated lower salt mass and a wider range of values when compared to the existing CRSS results. The existing CRSS could preserve the median of the observed historic salt mass (solid circles) in the interquartile range.

Figure 59 shows the annual historic salt concentration PDF for the modified CRSS in the upper graph and the existing CRSS in the lower graph. The existing CRSS overestimated the observed historic salt concentration PDF, seen by the shift in the peak of the simulated historic salt concentration PDFs. This overestimate is a result of the overestimate seen in the PDFs of the historic salt mass. The modified CRSS slightly underestimated the observed historic salt concentration PDF because the simulated historic flow was greater than the observed historic flow, diluting the simulated historic salt concentration.

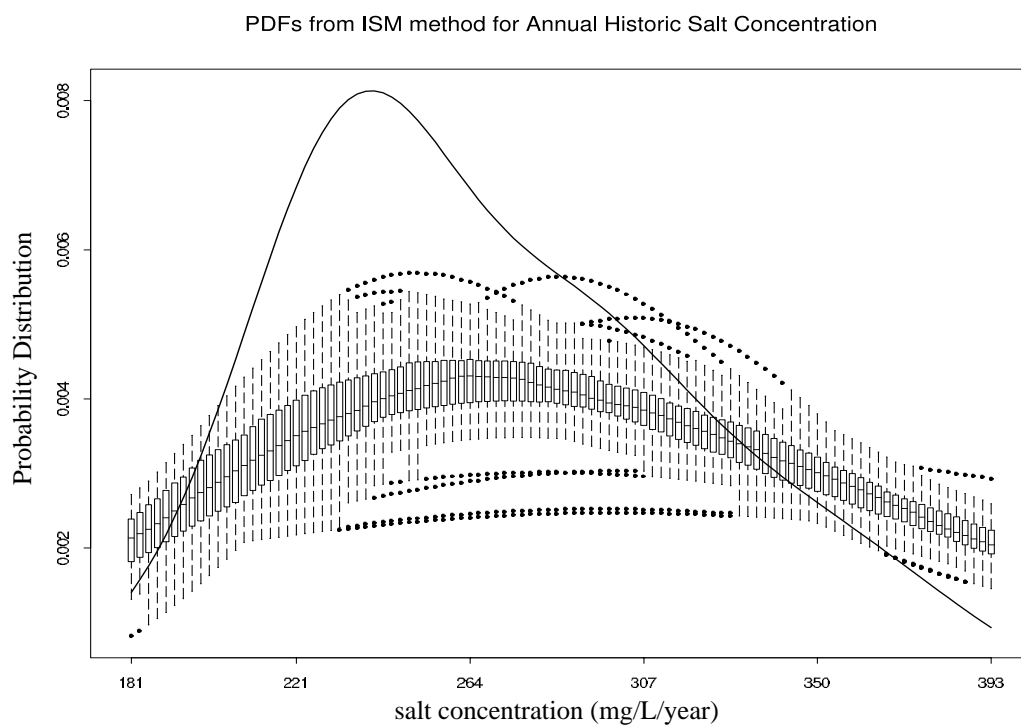
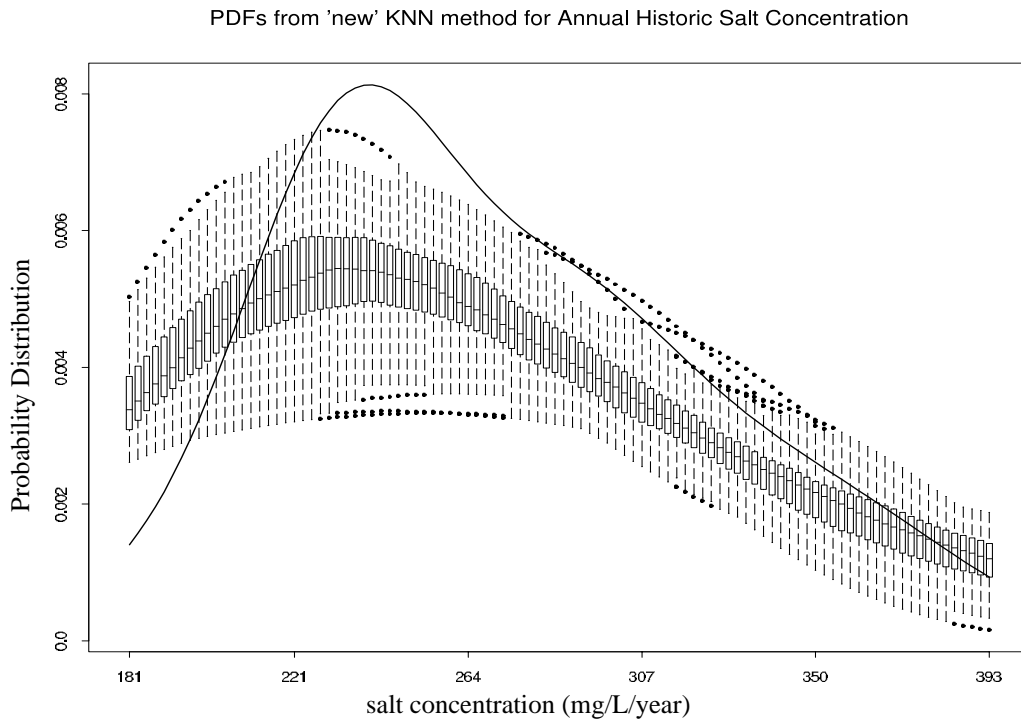


Figure 59: The upper graph shows the PDFs for historic salt concentration from the modified CRSS. The lower graph shows the PDFs for historic salt concentration from the existing CRSS.

Figure 60 shows the results from the simulated policy analysis. These plots compare the results from the modified CRSS to the existing CRSS. The upper graph shows the number of times each CRSS generates results that exceeds a standard of 650,000 tons. The existing CRSS results (boxplots) greatly overestimated the number of times the tons standard was exceeded compared to the observed historic record (solid circle), while the modified CRSS results preserved the number of times the tons standard was exceeded compared to the observed historic record.

The lower graph shows the number of times a salinity concentration standard of 350 mg/L was exceeded. The existing CRSS results greatly overestimated the number of times the salinity concentration standard was exceeded compared to the observed historic record. The modified CRSS results also overestimated the number of times the salinity concentration standard was exceeded compared to the number of times the observed historic record violated the standard, however the number of times the observed historic record violated the standard was preserved within the whiskers of the boxplot. The overestimation seen in the modified CRSS results could be removed if the simulation model were run at an annual time step and the proposed relationship for total depletions as a function of calculated natural flow were incorporated, as discussed when the modified CRSS used the synthetic natural flow based on 1941 to 1995 calculated natural flows.

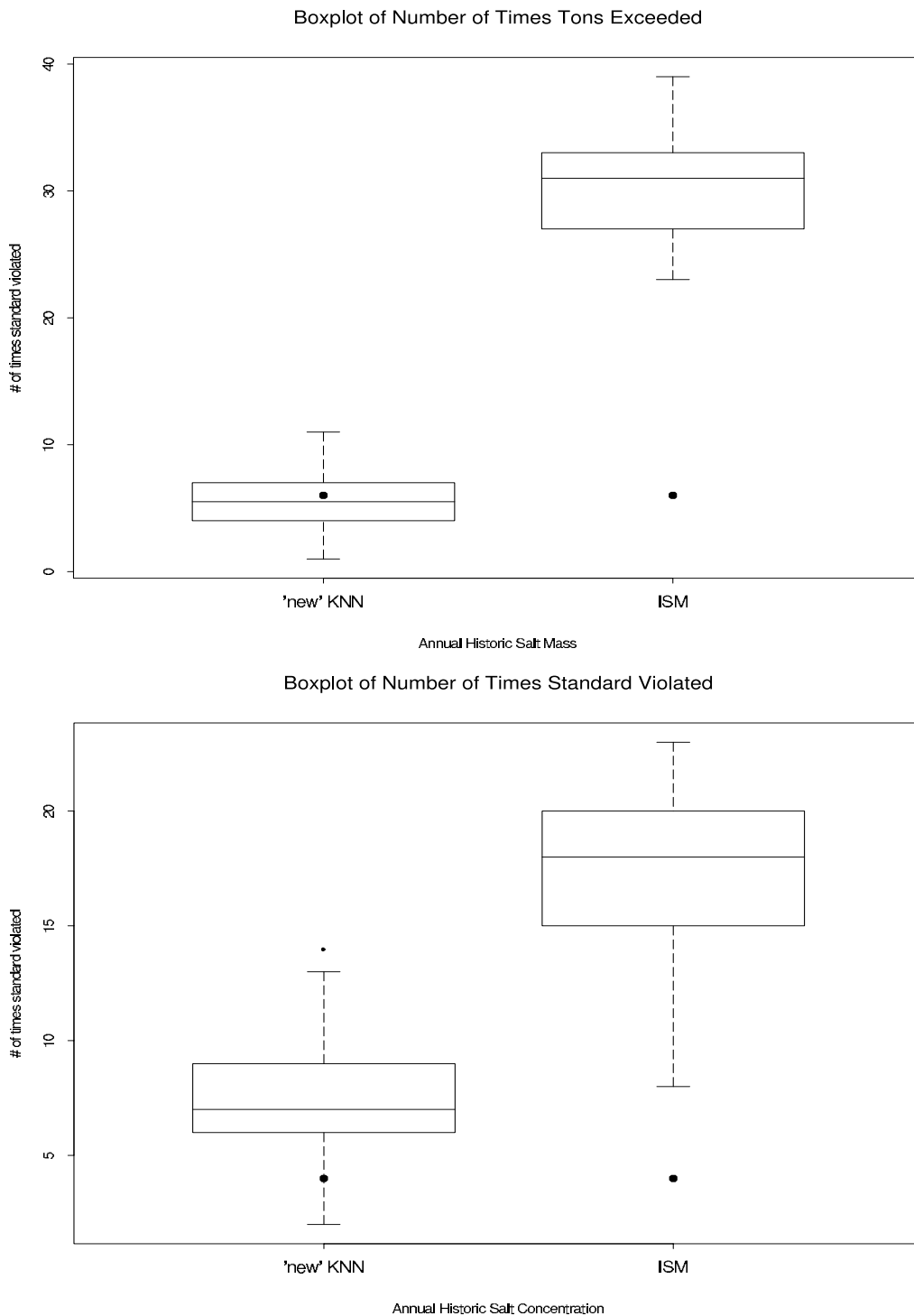


Figure 60: The upper graph shows the number of times each modeling system's results exceeded a tons standard of 650,000 tons. The lower graph shows the number of times the modified and existing CRSS results violated a salt standard of 350 mg/L. The solid circles show the number of times the observed historic record exceeded the standards. The boxplots show results of 100 simulations from the modified CRSS and 90 simulations from the existing CRSS.

4.3.6 Stochastic Planning Run Simulations

Next, the existing CRSS and the modified CRSS were used to simulate stochastic planning runs. The runs approximated the projected future flow, salt mass, and concentration from 2002 to 2061 using the most extensive data available. Using the entire calculated historic flow time series allowed the existing CRSS and the modified CRSS to be compared.

Stochastic planning runs were developed from the results of the monthly time step simulation model. A synthetic natural flow time series of 60 years length was generated based on 1906 to 1995 calculated natural flow. The associated synthetic natural salt mass was computed using the statistical nonparametric natural salt model. The total depletions were projected future depletions from 2002 to 2061. Running the simulation for 60 years ensured the simulation was run until the future depletions reached full development.

The upper graph in Figure 61 shows the PDF created from the median for boxplots of probability at a given projected future flow. Both systems generated projected future flow similarly. The lower graph shows the PDF created from the median for boxplots of probability at a given projected future salt mass. The modified CRSS produced lower salt mass than the existing CRSS.

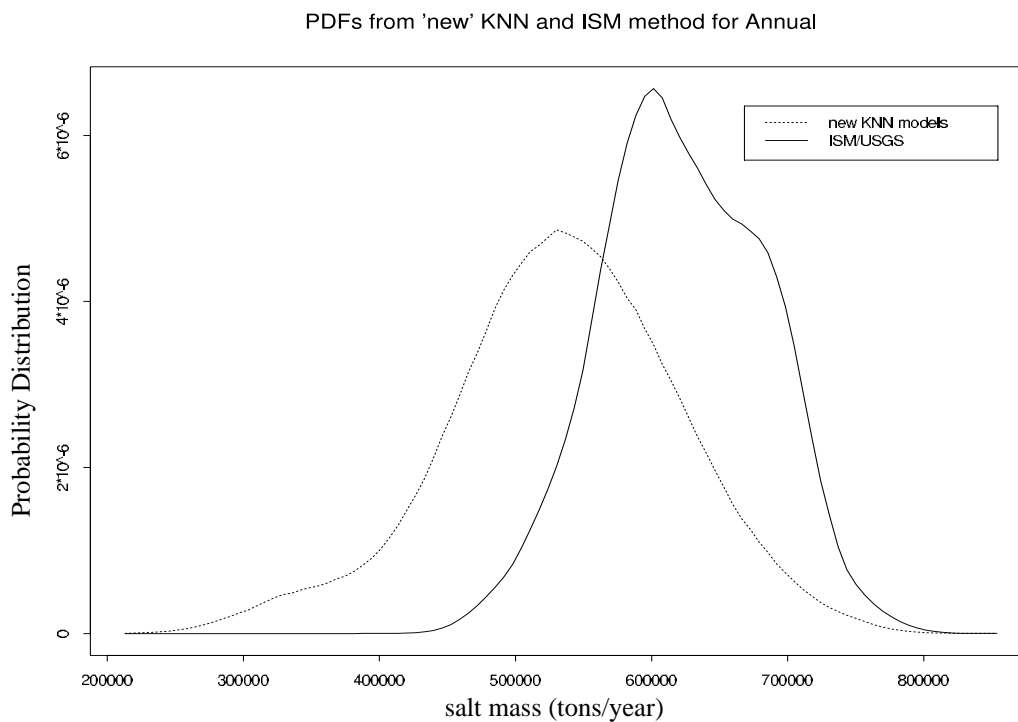
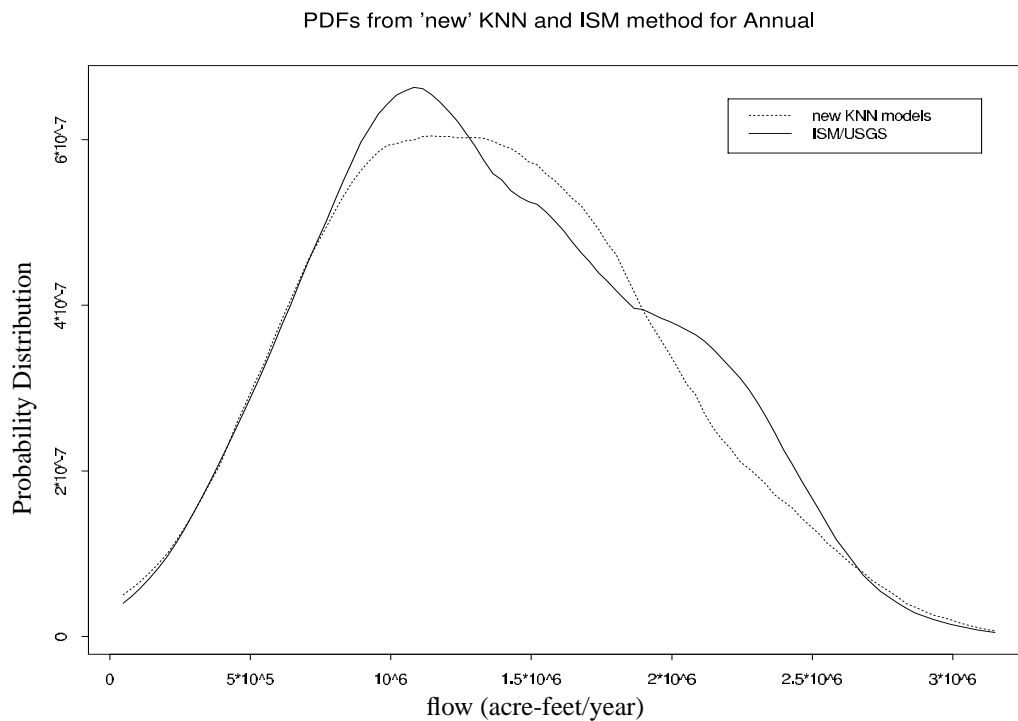
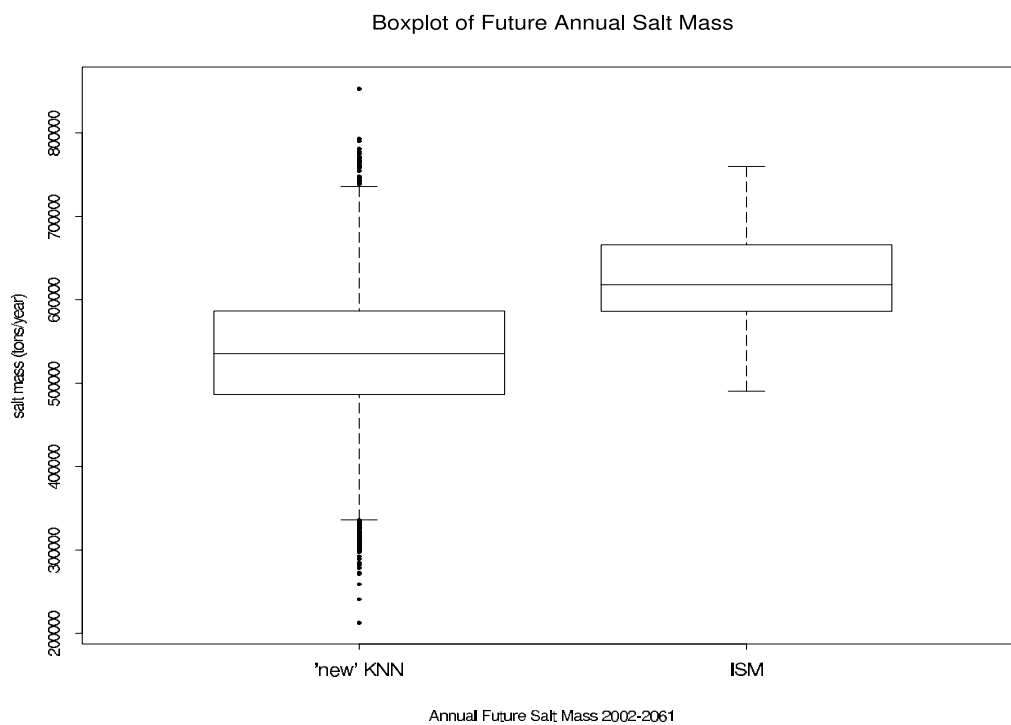


Figure 61: The upper graph shows the PDF created from the median of boxplots of probability at a given projected future flow generated from the modified and existing CRSS. The lower graph shows the projected future salt mass from the modified and existing CRSS. The shift in the projected future salt mass PDF shows that the existing CRSS generated greater salt mass.

Boxplots showing all the simulations, in the upper graph of Figure 62, again show that the modified CRSS produced lower salt masses than the existing CRSS. The lower graph shows the PDF created from the median for boxplots of probability at a given projected future salt concentration for each system. The difference between the two systems is shown; again the modified CRSS simulated lower projected future salt concentrations than the existing CRSS.

Figure 63 reiterates these results when the two modeling systems are compared with the simulated policy analysis. The upper graph shows the number of times a tons standard of 750,000 tons was exceeded and the lower graph shows the number of times at salinity standard of 600 mg/L was exceeded and, therefore, violated. Again, the estimates of the existing CRSS exceeded the estimates of the modified CRSS.



PDFs from 'new' KNN and ISM method for Annual Historic Salt Concentration

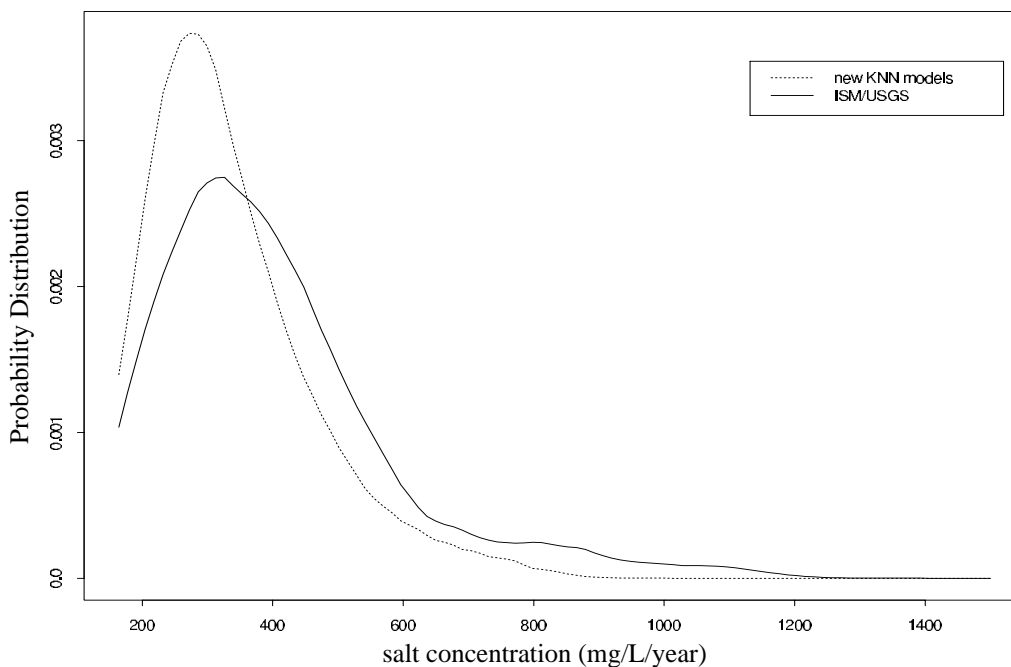


Figure 62: The upper graph shows the boxplots of the median from projected future salt mass. The modified CRSS generated lower salt mass and a wider range of values than the existing CRSS. The lower graph shows the PDF created from the median of boxplots of probability at a given projected future salt concentration generated from the existing and modified CRSS. The slight shift in the PDFs shows that the existing CRSS generated higher salt concentration.

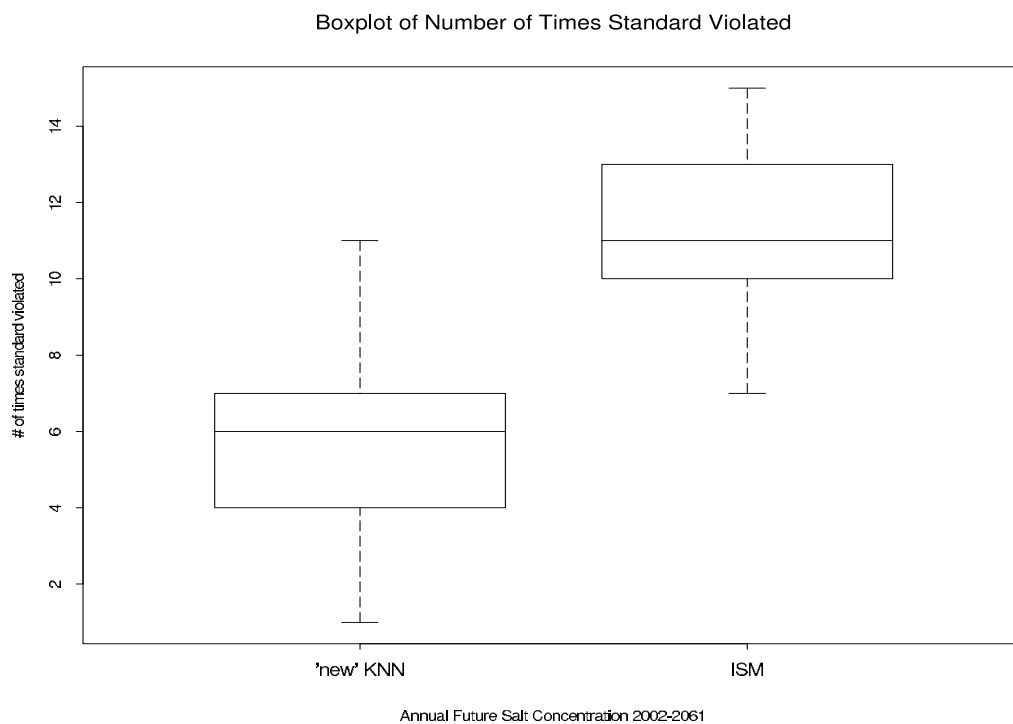
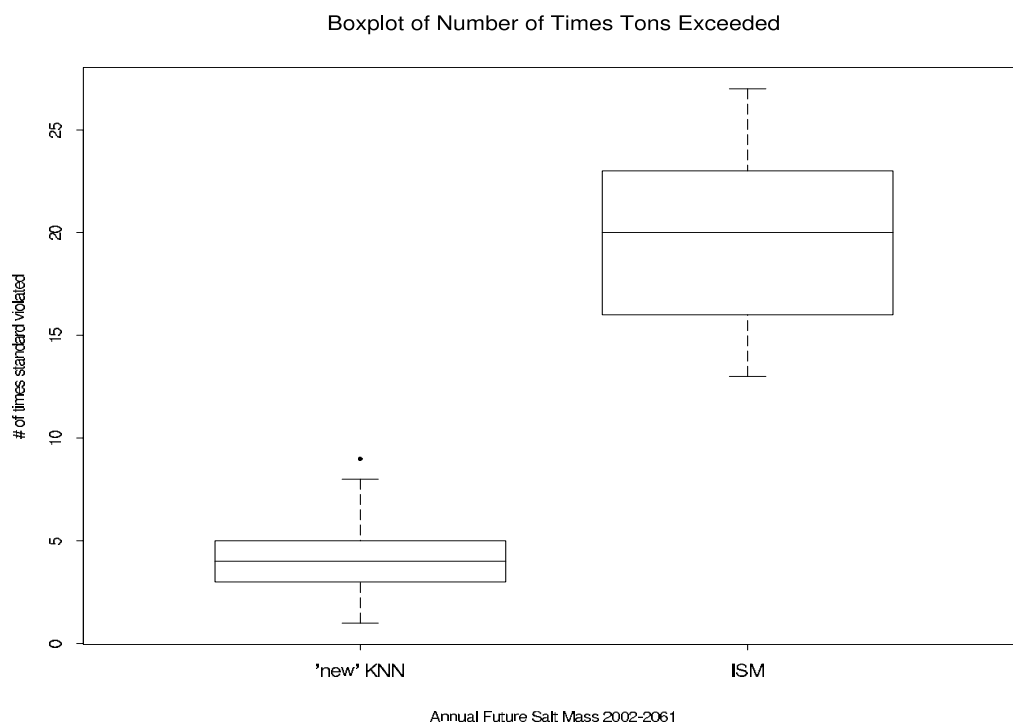


Figure 63: The upper graph shows the number of times the modified and existing CRSS results exceeded a tons standard of 750,000 tons, while the lower graph shows number of times a salt standard of 650mg/L was exceeded and, therefore, violated during water year 2002 to 2062. For both graphs the existing CRSS results exceeded the modified CRSS results.

4.4 Discussion and Conclusions

We outlined and developed a modified CRSS modeling system to simulate long-term historic flow, salt mass, and concentration at a single gauge in the upper Colorado River mainstem. The modified modeling system consisted of two modified models. The first model was a nonparametric K-NN model with resampling that generated synthetic natural flow. The second model used the generated natural flow and computed an associated natural salt mass from a nonparametric local linear regression with K-NN resampling. A simulation model used the generated synthetic natural flow and salt mass to perform Monte Carlo simulations, which model historic flow, salt mass, and concentration with uncertainty.

We validated our modified modeling system by ensuring the modeling system reproduced all the distributional properties, i.e., the PDF of the observed historic record. Reproducing the PDF of the historic record ensured the mean, standard deviation, and skewness were all preserved. Validation was performed by using the model to generate synthetic streamflows for the period 1941 to 1995. We found we were able to preserve the PDF for the natural flow and salt, but we produced a greater range of historic flow values. This greater range of historic flow values caused our historic salt concentration to produce a greater range of values than observed. Upon investigating this problem, we found that the relationship between natural flow and total depletions must be preserved to preserve the correct distribution of historic flows. Once we incorporated the relationship between natural flow and total depletions, we were able to preserve the historic flow and salt concentration PDF.

We used two time scales to develop a monthly and annual nonparametric local linear regression of natural salt mass as a function of natural flow with K-NN residual resampling model. At a monthly time scale, 12 regressions were developed - one for each month. To calculate the annual natural salt the result from the 12 regressions were summed to an annual value. Using this method, we lost the ability to preserve the tails

of the observed annual distribution. For policy analysis, the annual time step is more important than the monthly time step. A monthly time step is only used to facilitate the use of the “rules” to simulate river basin policy. If the simulation model were run at an annual time step, an annual statistical nonparametric natural salt model would generate better results. We showed that using a single annual natural salt mass as a function of natural flow regression to generate annual natural salt mass best allowed the model to preserve the observed annual natural salt mass PDF. Further, incorporating the relationship between total annual depletions as a function of annual natural flow in addition to an annual statistical nonparametric natural salt model allowed us to preserve the number of times a fictional salinity standard was violated.

After completing our validation, we used the modeling system at the monthly time step, accommodating the simulation model’s requirement to model operational policy rules. We first used hydrologies from 1906 to 1995 to generate natural flows for the period 1941 to 1995. We found that the flows before 1941 generally were higher than flows after 1941. Thus, the generated natural flows were higher than observed flows from 1941 to 1995, which caused the model to generate higher salt mass than observed and a wider range of historic salt concentration. The historic salt concentration was skewed towards lower concentrations because the shift in natural flows towards higher flows was larger than the shift in natural salt towards higher salt. The additional natural flow diluted salt mass, lowering the salinity concentration.

We compared the existing CRSS to the modified CRSS using the entire natural flow dataset from 1906 to 1995. The comparison showed that the modified CRSS simulated much lower salt mass and concentrations than the existing CRSS.

The final future simulations gave similar results. Projected future flow, salt mass, and concentration were simulated from 2002 to 2061. We compared the modified and existing CRSS, finding again that the modified CRSS, simulated lower future projected salt mass (88,000 tons lower) and concentration (57 mg/L lower) than the

existing CRSS.

Another strength of the modified CRSS that was not discussed previously is that the nonparametric K-NN stochastic flow model can be easily used to generate high- or low-flow scenarios to test policies under adverse conditions. To add this ability to the K-NN framework, the flow database that is used for resampling could be broken in thirds. The first third would be high flows, the second normal flows, and the third low flows. To generate a low-flow scenario, the K-NN model would generate a future streamflow time series by resampling only from the low flows in the database, thereby generating a synthetic time series with a sustained drought. This could also be performed by resampling from only the high flows to generate the time series.

Our modified modeling system generated historic flow and salt mass that matched the observed time period from which they were generated. We showed that our modified CRSS generated lower historic salt mass and salt concentration than the existing CRSS. We developed the new modeling system from data used to run the simulation model. By using the data from the simulation model, we kept our methods to develop data to drive the simulation model consistent with the data used in the simulation model. Keeping the data consistent allowed the modified CRSS to reproduce historical results and to fully reflect variability by improving the relationship between natural flow and salt mass.

Chapter 5

Conclusion and Recommendations for Future Work

5.1 Conclusions

This research developed a modified modeling system to simulate flow, salt mass, and concentration at USGS gauge 09072500. The modified modeling system was built on three interconnected models: (1) a nonparametric stochastic natural flow model, (2) a statistical nonparametric natural salt mass model, and (3) a river basin simulation model, the CRSS simulation model.

The modified modeling system improved the accuracy of salinity modeling when compared to the existing modeling system used in CRSS. The two systems were compared at a single gauge in the upper Colorado River mainstem, USGS stream gauge 09072500 (Colorado River near Glenwood Springs, Colorado) and its associated water quality gauge 09071100 (Colorado River near Glenwood Springs, Colorado). The modified modeling system was validated over the time period 1941 to 1995, when both historic flow and salt data were available. The validation step demonstrated that the modified modeling system preserves basic and higher order statistics for the observed data. A discussion of each of the three interconnected models follows.

5.1.1 Modified Nonparametric Stochastic Flow Model

This task compared three techniques to generate stochastic flows. The existing technique, nonparametric index sequential method (ISM), can only reproduce values

for stream flow that occurred in the past. The variability of the generated stream flows is also more limited than the other two techniques. These other techniques, a parametric periodic auto regressive (PAR) model and a nonparametric modified k-nearest neighbor (K-NN) model, develop stochastic stream flows values and sequences with more variety because they generate values that did not occur in the historic record, but are “statistically possible.”

The parametric PAR technique requires the modeled data to fit a Gaussian (normal) distribution. If this requirement is not met, the data needs to be transformed to fit a Gaussian distribution. Transformation is a time consuming task that does not guarantee appropriate results. Further, the parametric technique cannot preserve PDFs with non-Gaussian features, which the Colorado River streamflow data exhibit.

Conversely, the nonparametric technique does not require the data to fit a Gaussian distribution. The technique improves current K-NN techniques by incorporating a method to resample residuals from a regression, therefore allowing the modified K-NN technique to generate values not seen in the historic record. Traditional K-NN techniques resample directly from the historic dataset when generating a synthetic time series. The modified K-NN technique is able to preserve both Gaussian and non-Gaussian PDFs and does not require the modeled data to be transformed to a Gaussian PDF when using the modified K-NN technique. The modified K-NN technique is found to apply to a larger variety of data than the PAR technique.

The modified K-NN model was easy to implement and provided a flexible framework. The flexible framework allowed the K-NN technique to generate synthetic streamflows that simulated sustained low-flow or high-flow period. It was also easy to condition the flow data on an independent time series, such as climate indices. Climate indexes have shown a strong correlation in many river basins, i.e., San Juan River. When flows are conditioned on climate, climate forecasts can drive future simulations of streamflow.

5.1.2 Statistical Nonparametric Natural Salt Model

The existing natural salt model was developed from weighted least-square linear regressions based on historic flow, salt mass, and multiple development variables, including agricultural consumptive use, reservoir regulation, and exports. The model generates a higher natural salt mass than the new natural salt regression model. The existing natural salt model does not incorporate any of the data for modeling salt in the CRSS simulation model. To maintain consistency between the CRSS simulation model and the statistical nonparametric natural salt model, the data from the CRSS simulation model was used to develop data for the statistical nonparametric natural salt model. The new salt model is based on a regression between natural salt mass as a function of natural flow. The addition of a K-NN resampling algorithm allows the new salt model to incorporate the variance of the regression in the results from the model.

5.1.3 CRSS Simulation Model

The CRSS simulation model uses the ensembles generated by the modified stochastic natural flow model and the statistical nonparametric natural salt model to drive a model of USGS stream gauge 09072500 (Colorado River near Glenwood Springs, CO). The modified stochastic natural flow model incorporates the uncertainty of flows based on the observed historic flow. The statistical nonparametric natural salt model incorporates the uncertainty (variance) in the natural flow versus natural salt regression. Together, these three models allow the river basin model to incorporate the uncertainties in both the flow and salt together in a single river basin model. The CRSS simulation model simulates the flow, salt mass, and salt concentration with uncertainty at the stream gauge.

In the CRSS simulation model, the existing techniques to simulate future stream flow, salt mass, and salt concentration were compared to our modified techniques. The existing techniques use the ISM and a weighted least-square regression to find the natural salt mass as a function of natural flow. The modified techniques use a

modified K-NN stochastic model with residual resampling and a nonparametric local linear regression with K-NN residual resampling. The modified techniques produced lower salt mass and salt concentration values than the existing techniques. Both techniques generated flows that produced a similar median flow, although the modified techniques generated a wider variety of flows around the median.

5.2 Future Work

Many projects could further develop this research. Incorporating a temporal and spatial disaggregation technique could allow the modeling system to preserve both an annual and monthly time scale and extend the model to additional gauges throughout the Colorado River basin. A temporal disaggregation model would preserve basic statistics at both the monthly and annual time scale. If the operational policy *rules* were written to function at the annual time step, a temporal disaggregation model would not be needed and an annual time step simulation model could be built.

Spatial disaggregation techniques could extend this framework to additional gauges and preserve the correlation structure among the gauges. Once the modified techniques have been extended to a basin-scale, their performance in predicting salt concentration at the locations of the salinity standards could be investigated.

With the extended modeling framework, the impacts on operational and planning policy could be comprehensively studied. Considering the results of this research, the modified modeling system can be expected to simulate lower salt mass and concentration at the salinity standard locations.

Finally, we found that salt loading from human-induced sources is not well understood. Limited data allows little flexibility in designing a technique to model human-induced salt loading. Efforts to better understand salt loading by humans would improve our confidence in modeling human-induced salt loading. Further efforts could explore the relationship between salt loading and land use and develop a

method to model salt loading based on land use. Currently, data does not support a mathematical relationship between salt loading and land use. Instead, estimates of salt loading for each reach in the existing CRSS model are taken from several salinity studies.

Continued research efforts to understand highly uncertain data and methodologies more effectively conveys uncertainty in the modeling system's results, therefore improving the policymaker's decision support system.

References

(Beck, 1987)

Beck, M.B., 1987. *Water Quality Modeling: A Review of the Analysis of Uncertainty*. Water Resources Research, Vol. 23 No. 8, pp 1393-1442.

(Benjamin and Cornell, 1970)

Benjamin, J.R. and Cornell, C.A., 1970. Probability, Statistics, and Decision for Civil Engineers. McGraw-Hill Companies Inc., United States of America

(Bras and Rodriguez-Iturbe, 1985)

Bras, R.L. and Rodriguez-Iturbe, I., 1985. Random Functions and Hydrology. Addison-Wesley Publishing, Reading, Massachusetts.

(Carron, 2000)

Carron, J., 2000. *Simulation and Optimization of Unsteady Flow and Water Temperature in Reservoir-Regulated Rivers*. Dissertation (Ph.D.). University of Colorado, Boulder Colorado.

(Cayan et al., 1999)

Cayan, D.R., Redmond, K.T., Riddle, L.G., 1999. *ENSO and Hydrologic Extremes in the Western United States*. Journal of Climate, Vol. 12 pp 2881-2893.

(Cayan and Webb, 1992)

Cayan, D. and Webb, R. 1992. El Nino/Southern Oscillation and Streamflow in the Western United States. In: El Nino. Henry F. Diaz and Vera Markgraf (Editors). Cambridge University Press, Cambridge, Great Britain, pp. 29-68.

(Chapra, 1997)

Chapra, Steven C. 1997. Surface Water-Quality Modeling. The McGraw-Hill Companies Inc., United States of America.

(Griffin, 1995)

Griffin, C.B. 1995. *Uncertainty analysis of BMP effectiveness for controlling nitrogen from urban nonpoint sources*. Water Resources Bulletin, Vol

31 No. 6pp 1041-1050.

(Helsel and Hirsch, 1992)

Helsel, D.R. and Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier, Reston, Virginia.

(Hendrick, 1972)

Hendrick, J.D., 1972. *Techniques for modeling reservoir salinity*. Dissertation (Ph.D.). Colorado State University, Fort Collins.

(Hession et al., 1996)

Hession, W. C., Storm, D.E., Haan, C.T., Burks, S.L., Matlock, M.D. 1996. *A Watershed-Level Ecological Risk Assessment Methodology*. Water Resources Bulletin, Vol 32 No. 5 pp 1039-1054.

(Iorns et al., 1965)

Iorns, W.V., C.H. Hembree, and G.L. Oakland. 1965. *Water Resources of the Upper Colorado River Basin - Technical Report*. U.S. Geological Survey, Professional Paper 441.

(Jensen et al., 1990)

Jensen, M.E., Burman, R.D. Allen, R.G., 1990. Evapotranspiration and Irrigation Water Requirements. American Society of Civil Engineers, New York, New York.

(Kendall and Dracup 1991)

Kendall, D.B. and Dracup, J.A., 1991. *A Comparison of Index-Sequential and AR(1) Generated Hydrologic Sequences*. Journal of Hydrology, Vol. 122 pp 335-352.

(Labadie et al., 1990)

Labadie, J.W., Fontane, D.G., Salas, J.D., Ouarda, T. 1990. *Validation studies on indexed Sequential Modeling for the Colorado River Basin*, Department of Civil Engineering, Colorado State University, Ft. Collins, Colorado.

(Lall, 1995)

Lall, U. 1995. *Recent Advances in Nonparametric Function Estimation: Hydraulic Applications*. Reviews of Geophysics, Vol. 33 Part 2, Suppl. S, pp 1093-1102.

(Lall and Sharma, 1996)

Lall, U. and Sharma, A., 1996. *A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series*. Water Resources Research, Vol. 32 No. 3, pp 679-693.

(LaVenue et al., 1989)

LaVenue, M., Andrews, R.W., Ramarao, B.S. 1989. *Groundwater*

Travel Time Uncertainty Analysis Using Sensitivity Derivatives. Water Resources Research, Vol. 25 No. 7, pp 1551-1566.

(Lee, 1989)

Lee, D. 1989. *Salinity in the Colorado River Basin: A dynamic modeling approach to policy analysis*. Dissertation (Ph.D). University of California, Davis.

(Lee et al., 1993)

Lee, D., Howitt, R., Marino, M. 1993. *A Stochastic Model of River Water Quality: Application to Salinity in the Colorado River*. Water Resources Research, Vol. 29 No. 12, pp 3917-3923.

(Liebermann et al., 1988)

Liebermann, T.D., Mueller, D.K., Kircher, J.E. Choquette, A.F. 1988. *Characteristics and Trends of Streamflow and Dissolved Solids in the Upper Colorado River Basin, Arizona, Colorado, New Mexico, Utah, and Wyoming*. U.S. Geological Survey in cooperation with the U.S. Bureau of Reclamation. Report 87-568.

(Loader, 1999)

Loader, C., 1999. Local Regression and Likelihood. Springer. New York.

(Malone et al., 1979)

Malone, R.F., Bowles, D.S., Grenney, W.J., Windham, M.P. 1979. *Stochastic Analysis of Water Quality*. Utah Water Research Laboratory, Utah State University, Logan, Utah.

(McCabe and Dettinger, 1999)

McCabe, G.J and Dettinger, M.D., 1999. *Decadal Variations in the Strength of ENSO Teleconnections with Precipitation in the Western United States*. International Journal of Climatology, Vol. 19 pp. 1399-1410.

(McMahon et al., 1996)

McMahon, T.A., Pretto, P.B, Chew, F.H.S., 1996. Synthetic Hydrology - Where's the Hydrology. In: Stochastic Hydraulics '96. Kevin s. Tickle, Ian C. Goulter, Chengcheo Xii, Salch A. Wasimi, and Francois Douchart (Editors). A.A. Balkema, Rotterdam, Netherlands, pp. 3-14.

(Mueller and Osen, 1988)

Mueller, D.K. and Osen, L.L., 1988. *Estimation of Natural Dissolved-Solids for the Upper Colorado River Basin*. U.S. Geological Survey, Water Resources Investigation Report 87-4069.

(Mueller and Liebermann, 1988)

Mueller, D.K. and Liebermann, T.D., 1988. *Extension of Streamflow*

and Dissolved-solids records at selected sites in the Colorado river basin, Arizona, California, Colorado, Nevada, New Mexico, Utah, and Wyoming 1940-1983. U.S. Geological Survey in cooperation with the U.S. Bureau of Reclamation. Report 87-4203

(Nash and Gleick, 1993)

Nash, L. and Gleick, P., 1993. *The Colorado River Basin and Climatic Change, The sensitivity of Streamflow and Water Supply to Variations in Temperature and Prediction*, U.S. Environmental Protection Agency, Report EPA 230-R-93-009.

(Nathanson, 1978)

Nathanson, M.N. 1978. *Updating The Hoover Dam Documents*. United States Department of the Interior, Bureau of Reclamation, United States Government Printing Office, Denver, Colorado.

(National Research Council, 1968)

National Research Council, 1968. Water and Choice in the Colorado Basin. An Example of alternatives in water management. National Academy of Sciences, Washington, D.C.

(Nordlund and Liebermann, 1990)

Nordlund, B.D. and Liebermann, T.B., 1990. *Estimates of Dissolved Solids and Major Dissolved Constituents for 70 streamflow-gaging stations in the Upper Colorado River Basin, Arizona, Colorado, New Mexico, Utah, and Wyoming*. U.S. Geological Survey in cooperation with the U.S. Bureau of Reclamation. Report 87-547.

(Ouarda et al., 1997)

Ouarda, T., Labadie, J.W., Fontane, D.G., 1990. *Index Sequential Hydrologic Modeling For Hydropower Capacity Estimation*, Journal of the American Water Resources Association, Vol. 33 No. 6. pp 1337-1349.

(Owen-Joyce, 1987)

Owen-Joyce, S.J., 1987. *Estimates of Average Annual Tributary Inflow to the Lower Colorado River, Hoover Dam to Mexico*. U.S. Geological Survey Water Resources Investigation Report 87-4078.

(Owosina, 1992)

Owosina, A., 1992. *Methods for assessing the space and time variability of ground water data*. Thesis (M.S.). Utah State University, Logan, UT.

(Piechota et al., 1997)

Piechota, T.C., Dracup, J.A., Fovell, R.G., 1997. *Western US Streamflow and Atmospheric Circulation Patterns during El Nino-Southern Oscillation*. Journal of Hydrology, Vol. 201 pp 230-248.

(Rajagopalan and Lall, 1998)

Rajagopalan, B. and Lall, U., 1998. *Locally Weighted Polynomial Estimation of Spatial Precipitation*. Journal of Geographic Information and Decision Analysis, Vol. 2, No. 3, pp 48-57.

(Rajagopalan and Lall, 1999)

Rajagopalan, B. and Lall, U., 1999. *A K-Nearest-Neighbor Simulator for Daily Precipitation and Other Weather Variables*. Water Resources Research, Vol. 35, Vol. 10, pp 3089-3101.

(Riley and Jurinak, 1979)

Riley, J.P. and Jurinak, J.J., 1979. *Irrigation Management for River-Salinity Control*. Journal of the irrigation and drainage division, ASCE Vol. 105, No. IR4, pp 419-432.

(Salas, 1985)

Salas, J.D., 1985. Analysis and Modeling of Hydrologic Time Series. In: Handbook of Hydrology. David R. Maidment (Editor). McGraw-Hill, New York. pp. 19.1-19.72.

(Salas et al., 2000)

Salas, J.D., Chung, C., Lane, W.L., Frevert, D.K., 2000. Stochastic Analysis, Modeling, and Simulation (SAMS) Version 2000 - User's Manual. Department of Civil Engineering, Hydrologic Science and Engineering Program, Colorado State University, Fort Collins, Colorado.

(Shama et al., 1997)

Shamas, A., Tarboton, D.G., Lall, U., 1997. *Streamflow Simulation: A Nonparametric Approach*. Water Resources Research, Vol. 33 No. 2 pp 291-308.

(Srinivas and Srinivasan, 2001)

Srinivas, V.V. and Srinivasan, K., 2001. *A Hybrid Stochastic Model For Multiseason Streamflow Simulation*. Water Resources Research, Vol. 37 No. 10 pp 2537-2549.

(Tarboton et al., 1998)

Tarboton, D.G., Sharma, A., Lall, U., 1998. *Disaggregation procedures for Stochastic Hydrology Based on Nonparametric Density Estimation*. Water Resources Research, Vol. 34, Vol. 1, pp 107-119.

(Taylor, 1982)

Taylor, J.R., 1982. An Introduction to Error Analysis - The Study of Uncertainties in Physical Measurements. University Science Books. United States of America.

(Tung, 1996)

Tung, Y., 1996. Uncertainty analysis in water resources engineering. In: Stochastic Hydraulics '96. Kevin s. Tickle, Ian C. Goulter, Chengcheo Xii, Salch A. Wasimi, and Francois Douchart (Editors). A.A. Balkema, Rotterdam, Netherlands.

(URWL, 1975)

Utah Water Research Laboratory. 1975. *Colorado River regional assessment study*. Utah Water Research Laboratory. Utah State University, Logan, Utah.

(USBR, 1974)

U.S. Bureau of Reclamation. January 1974. *Colorado River Water Quality Improvement Program; Status Report*. U.S. Department of Interior, Bureau of Reclamation.

(USBR, 1983 (1))

U.S. Bureau of Reclamation. January 1983. *Draft Report - Grand Valley Salt Pickup Calculations*. U.S. Department of Interior, Bureau of Reclamation. Grand Junction, Colorado.

(USBR, 1983 (2))

U.S. Bureau of Reclamation. June 1983. *Draft Report - Colorado River Simulation System Hydrology Data Base - Upper Colorado Region*. U.S. Department of Interior, Bureau of Reclamation.

(USBR, 1985)

U.S. Bureau of Reclamation. May 1985. *Colorado River Simulation System - Hydrology Data Base, Lower Colorado Region, Lees Ferry to Imperial Dam*. U.S. Department of Interior, Bureau of Reclamation.

(USBR, 1987)

U.S. Bureau of Reclamation. May 1987. *Colorado River Simulation System, System Overview*, U.S. Department of Interior, Bureau of Reclamation Denver, Colorado.

(USBR, 1992)

U.S. Bureau of Reclamation. March 1992. *Colorado River Simulation System - Hydrologic Flow and Salt Data Base for the Lower Colorado Region, Lees Ferry to Imperial Dam*. U.S. Department of Interior, Bureau of Reclamation

(USBR, 1997)

U.S. Bureau of Reclamation. January 1997. *Quality of Water Colorado River Basin - Progress Report No. 18*, Technical Communications Group, Technical Services Center, Bureau of Reclamation Denver, Colorado.

(USBR, 1999)

U.S. Bureau of Reclamation. January 1999. *Quality of Water Colorado River Basin - Progress Report No. 19*, Technical Communications Group, Technical Services Center, Bureau of Reclamation Denver, Colorado.

(U.S. Department of Interior, 2001)

U.S. Department of Interior, 2001. Quality of Water Colorado: Progress Report 20 [WWW Document]. URL <http://www.uc.usbr.gov/progact/salinity/index.html> (visited 2001, October 15).

(Wahba and Wold, 1975)

Wahba, G. and Wold, S., 1975. *A Completely Automatic French Curve*. Communications in Statistics. Vol. 4, pp 1-17.

(Zagona et al., 2001)

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., and Goranflo, H.M., 2001. *RiverWare: Generalized Tool For Complex Reservoir System Modeling*. Journal of the American Water Resources Association. Vol. 37 No. 4 pp 913-929.

Appendix A

Glossary of Terms

observed historic flow - from gauged historic records

observed historic salt mass - from gauged historic records

total depletion = sum (agriculture consumptive use + exports \pm reservoir regulation)

human-induced salt mass = sum (salt leaving with export water + agricultural salinity pickup)

calculated natural flow = observed historic flow - total depletions

calculated natural salt mass = observed historic salt mass - human-induced salt mass

estimated natural salt mass - natural salt mass estimated for the corresponding calculated natural flow from a natural salt model over a historic period

calculated natural salt concentration (mg/L) = calculated natural salt mass (tons) * 735.29 / calculated natural flow volume (ac-ft)

calculated historic flow = calculated natural flow - total depletions

calculated historic salt mass = calculated natural salt mass - human-induced salt mass

synthetic natural flow (100 synthetic traces) - generated natural flows from the stochastic natural flow model

associated synthetic natural salt mass (100 synthetic traces) - estimates of natural salt corresponding to synthetic natural flows

synthetic natural concentration (mg/L) = associated synthetic natural salt mass (tons) * 735.29 / synthetic natural flow volume (ac-ft)

simulated historic flow (100 synthetic traces) = simulated natural flow - total depletions

simulated historic salt mass (100 synthetic traces) = simulated natural salt mass - human induced salt mass

simulated historic concentration (mg/L) = simulated historic salt mass (tons) * 735.29 / simulated historic flow volume (ac-ft)

modified stochastic natural flow model (MSNFM)

statistical nonparametric natural salt model (SNNSM)

index sequential method flow model (ISM flow model)

US Geological Survey salt model (USGS salt model)

Appendix B

CRSS Inputs, Outputs, and Sources of Data

2.1 Identify Primary Inputs, Outputs, and the Sources of the Data for Colorado River Simulation System

- a) The inputs and outputs for a projection run are as follows (USBR, 1987)

Inputs:

- (1) Natural flow, calculated as described in following section (b)
- (2) Demands, these are input into SMDID (Simulation Model Demand Input Data) generation program. Withdrawals and depletions are specified in the database. The values for each can be held constant or they can be trended, varying either by a step or linear trend to meet a certain value a number of years in the future. All values are entered annually and distributed for each diversion point according to a monthly distribution. Sources of demands for the Upper Basin are typically taken from the Projected Water Supply and Depletions - Upper Colorado River. For the Lower Basin, demands are determined from the Consumptive Use of Diversions from the Main Stem report. Diversions are determined either from historical consumptive use to diversion ratios, maximum diversion capacity to maximum allowable consumptive use ratios, or an assumed diversion to consumptive use ratio of approximately 167 percent when no better data is available.
- (3) Initial reservoir elevation or storage, taken from USGS or USBR gauges
- (4) Natural salt load input as concentration, calculated as described in following section (b)

For the Upper Basin, developed from historical flow and salt load relationships (Mueller and Osen, 1988). For the Lower Basin, this was deviated from a mass balance for salt load within each reach of the lower basin.

- (5) Salinity Pickup, input as concentration in the SMDID generation program. Determined from project data or estimated (Iorns et al., 1965) for the Upper Basin. Included for only two projects in the Lower Basin. Salt pickup values for Lower Basin obtained from project data. The remaining salt pickup in the Lower Basin is assumed to be captured in the salt gains and losses from the hydrology data base.

Concentration assumed constant in the current model. Why not use salt load? This does not represent salt load dependent on flow.

- (6) Initial reservoir concentration, taken from outflow salt concentration of initial month for model run. Determined from USGS measurements, estimates, or extensions of existing data. (Nordlund and Liebermann, 1990; Mueller and Osen, 1988).

- Outputs:**
- (7) Flows and Salt Concentration at all modeled diversions and reservoirs in the Colorado River basin

2.2 Generating the Natural Flow Database

In order to present all gauged information at the same conditions, a time must be chosen to adjust all gauged data. In the CRSS model, the common level of development chosen was “natural” level. Natural flow represents the streamflows that occur if humans had not used any water from the basin. Natural flow is developed by subtracting human-induced effects (i.e., reservoir regulation, depletion, exports, municipal and industrial uses) from the historical gauge data. The uncertainty in the natural flow and salt would be determined from the uncertainty in the data used to calculate the natural flow database.

- b) The inputs and outputs for generation of natural flow data are as follows:

- Inputs:**
- (1) Inputs and sources for the Upper Basin natural flow (USBR, 1983 (2))

- (a) Historical Flow, USGS gauged data
- (b) Consumptive Use, modified Blaney-Criddle method
- (c) Reservoir Regulation
 - i) monthly change in surface storage, measured
 - ii) monthly change in bank storage, 10% of change in surface storage
 - iii) monthly evaporation, determined from monthly average surface area x monthly evaporation rate
- (d) Exports, USGS and irrigation districts, assumed measured
- (e) M and I Uses, determined from annual consumptive powerplant use, annual value is divided by 12 to get monthly
- (f) Imports, USGS and irrigation districts, assumed measured
- (g) Incidental depletions, account for stock pond evaporation, fish and wildlife uses, etc. Generally only annual totals, measured and calculated

Outputs:

- (2) Outputs total natural flow at specific stream gaging points

Inputs:

- (3) Inputs and sources for the Upper Basin natural salinity (USBR, 1987)
 - (a) The determination of natural salt load has evolved from different attempts to determine natural salt load given measured historical data. Currently salt load is derived from a relationship developed by the USGS taking the form of a regression equation $C = aQ^b$; C is concentration in mg/L, Q is flow in ft^3/s , and a and b are empirical constants. The relationship was developed from a weighted least-squares regression from a model of historical salt load as a function of historical streamflow and several variables representing development. Development variables include upstream adjustments to stream flow, consumptive use, diversions, and irrigated acreage

Outputs:

- (4) Outputs total natural salinity at specific stream gaging locations

- (a) Total natural salt concentration at specific stream gauging points
- Inputs:** (5) Inputs and sources for Lower Basin natural flow (USBR, 1985; USBR, 1992)
 - (a) Historical outflows from reservoirs, USGS gauged data
 - (b) Diversion and measured returns, Decree Accounting Records
 - (c) Phreatophyte consumptive use, Blaney Criddle
 - (d) Unmeasured returns, (Owen-Joyce, 1987)
- Outputs:** (6) Outputs natural flow data for specific actual and intervening flow stations
- Inputs:** (7) Inputs and sources for Lower Basin natural salinity (USBR, 1985; USBR, 1992)
 - (a) Historical outflow salt load. Determined from USGS measurements, estimates, or extensions of existing data (Nordlund and Liebermann, 1990; Mueller and Osen, 1988).
 - (b) Diversion and measured returns salt load, Determined from USGS measurements, estimates, or extensions of existing data (Nordlund and Liebermann, 1990; Mueller and Osen, 1988).
 - (c) Unmeasured returns (Owen-Joyce, 1987)
- Outputs:** (8) Outputs natural salinity for specific actual and intervening flow stations.

Appendix C

Sample Splus Code

3.1 Sample Splus Modeling Code

```
#####  
#NP modified k-nn stochastic flow model  
#  
#models natural flow for gauge 09072500 from water year 1906 to 1995  
#100 simulations each 90 years long are created.  
#####  
{  
WYmonflow_matrix(BCNatFlow[1:1080,1], ncol=12, byrow=T) #10/1905 to  
9/1995  
KNNtnflow_matrix(0,1080,100)  
  
annlee_1:90  
for(i in 1:90)annlee[i]_sum(WYmonflow[i,1:12])  
x_annlee  
  
#fit lowess and get the residuals series..  
resids_matrix(0,nrow=89,ncol=12)  
alpha_seq(0.2,0.8,by=0.05)  
alpha1_seq(0.2,0.8,by=0.05)  
alpha2_c(alpha, alpha1)  
for(i in 1:12){  
i1_i-1  
if(i == 1){  
zz_gcvplot(WYmonflow[2:90,1]~WYmonflow[1:89,12], alpha=alpha, deg=1,  
ev="cross")  
#z1_gcvplot(WYmonflow[2:90,1]~WYmonflow[1:89,12],alpha=alpha,deg=2)  
#z2_order(c(zz$values,z1$values))  
z2_order(zz$values)  
deg1_1
```

```

#if(z2[1] > 13)deg1_2
zz_locfit(WYmonflow[2:90,1]~WYmonflow[1:89,12], alpha=alpha2[z2[1]],
deg=deg1)
x11_zz
resids[,i]_residuals(zz)} else
{
zz_gcvplot(WYmonflow[1:89,i]~WYmonflow[1:89,i1], alpha=alpha, deg=1,
ev="cross")
#z1_gcvplot(WYmonflow[1:89,i]~WYmonflow[1:89,i1], alpha=alpha, deg=2)
#z2_order(c(zz$values,z1$values))
z2_order(zz$values)
deg1_1
#if(z2[1] > 13)deg1_2
zz_locfit(WYmonflow[1:89,i]~WYmonflow[1:89,i1], alpha=alpha2[z2[1]],
deg=deg1)
if(i == 2)x2l_zz
if(i == 3)x3l_zz
if(i == 4)x4l_zz
if(i == 5)x5l_zz
if(i == 6)x6l_zz
if(i == 7)x7l_zz
if(i == 8)x8l_zz
if(i == 9)x9l_zz
if(i == 10)x10l_zz
if(i == 11)x11l_zz
if(i == 12)x12l_zz
resids[,i]_residuals(zz)}
}

```

```

armean_matrix(0,101,13)
arstdev_matrix(0,101,13)
arcor_matrix(0,101,13)
arskw_matrix(0,101,13)
mondiffs_matrix(0,101,347)
armax_matrix(0,101,13)
armin_matrix(0,101,13)

```

```

### drought stats..

```

```

mxsp_1:101
mxdef_1:101
maxs_1:101
maxd_1:101

```

```

index_1:90

```

```

kk_sqrt(89)
kk_round(kk)
W_1:kk
W_1/W
W_W/sum(W)
W_cumsum(W)

frac_1
th_frac*mean(array(t(WYmonflow[1:90,1:12])))
th_quantile(array(t(WYmonflow[1:90,1:12])),0.5)

#   Get the first year..

for(k in 1:100){
xsim_1:1080
i_round(runif(1,1,89))
xsim[1]_WYmonflow[i,1]
xprev_xsim[1]
for(i in 2:1080){
j_i %% 12
if(j == 0)j_12
j1_j-1
if(j == 1){xx_abs(xprev-WYmonflow[1:89,12])}
else
{xx_abs(xprev-WYmonflow[1:89,j1])}

xz_order(xx)
xz_xz[1:kk]
xx_runif(1,0,1)
xy_c(xx,W)
xx_rank(xy)
i1_xz[xx[1]]

if(j == 1)xm_predict(x1l,xprev)
if(j == 2)xm_predict(x2l,xprev)
if(j == 3)xm_predict(x3l,xprev)
if(j == 4)xm_predict(x4l,xprev)
if(j == 5)xm_predict(x5l,xprev)
if(j == 6)xm_predict(x6l,xprev)
if(j == 7)xm_predict(x7l,xprev)
if(j == 8)xm_predict(x8l,xprev)
if(j == 9)xm_predict(x9l,xprev)
if(j == 10)xm_predict(x10l,xprev)
if(j == 11)xm_predict(x11l,xprev)
if(j == 12)xm_predict(x12l,xprev)

```

```

xsim[i]_xm + resids[i1,j]
#print(c(i,j,xprev,xm,resids[i1,j]))

xprev_xsim[i]}

simdismon_matrix(xsim,nrow=12)
simdismon_t(simdismon)

KNNtnflow[,k]_xsim

}
#*****
#Calculate historic flow
#
#*****
cu41to95_scan(file="cu41t095.txt")

KNNthistflow_matrix(99,660,100)
for (i in 1:100){
#   KNNhistflow[,i]_KNNtnflow[1:660,i]-cu41to95[1:660] #10/1941 to 9/
1995
    KNNthistflow[,i]_KNNtnflow[1:660,i]-cu41to95[1:660] #10/1941 to 9/
1995
}
ISMhistflow_matrix(99,660,100)
for (i in 1:100){
    ISMhistflow[,i]_KNNtnflow[1:660,i]-cu41to95[1:660] #10/1941 to 9/
1995
}

#*****
#Regression for natural flow and "back calculated" natural salt

#for monthly data
#fit lowess and get the residuals series..
#
#Required Input
BCNatFlow_matrix(scan("0725RebuiltNatFlow.txt"),ncol=1,byrow=T)
monBCNatFlow_matrix(BCNatFlow[1:1080,], ncol=12, byrow=T)

rebuiltNatSalt_USGSglen[1:660,4]+rebuiltExportMass[421:1080]-rebuilt-
MonRtnmass[421:1080] #10/1940 to 9/1995
rebuiltNatFlow_matrix(scan("0725RebuiltNatFlow.txt"),ncol=1,byrow=T)

```

```
mon0725Nflow_matrix(BCNatFlow[421:960], ncol=12, byrow=T) #10/1940
to 9/1985
mon0725Hmass_matrix(rebuiltNatSalt[1:540], ncol=12, byrow=T) #10/1940
to 9/1985
```

```
mon0725Nflow_matrix(BCNatFlow[421:1080], ncol=12, byrow=T) #10/1940
to 9/1995
mon0725Hmass_matrix(rebuiltNatSalt[1:660], ncol=12, byrow=T) #10/1940
to 9/1995
```

```
#####
resids_matrix(0,nrow=55,ncol=12)
alpha_seq(0.2,0.95,by=0.05)
```

```
for(i in 1:12){
#zz_gcvplot(mon0725Hmass[,i]~mon0725Nflow[,i], alpha=alpha, deg=1,
ev="cross")
zz_gcvplot(mon0725Hmass[,i]~mon0725Nflow[,i], alpha=alpha,deg=1)
z2_order(zz$values)
deg1_1
zz_locfit(mon0725Hmass[,i]~mon0725Nflow[,i], alpha=alpha[z2[1]],
deg=deg1)
```

```
if(i == 1)x1s_zz
if(i == 2)x2s_zz
if(i == 3)x3s_zz
if(i == 4)x4s_zz
if(i == 5)x5s_zz
if(i == 6)x6s_zz
if(i == 7)x7s_zz
if(i == 8)x8s_zz
if(i == 9)x9s_zz
if(i == 10)x10s_zz
if(i == 11)x11s_zz
if(i == 12)x12s_zz
resids[,i]_residuals(zz)
}
```

```
#####
# validate improved salt model
#####
```

```
ISMflow_matrix(BCNatFlow[421:1080,], ncol=12, byrow=T) #10/1940 to 9/
1995
```

```
ISMsalt40to95_matrix(0,55,12)
```

```

data_(ISMflow)
#data2_monResReg

ISMsalt40to95[,1]_predict(x1s,data[,1])
ISMsalt40to95[,2]_predict(x2s,data[,2])
ISMsalt40to95[,3]_predict(x3s,data[,3])
ISMsalt40to95[,4]_predict(x4s,data[,4])
ISMsalt40to95[,5]_predict(x5s,data[,5])
ISMsalt40to95[,6]_predict(x6s,data[,6])
ISMsalt40to95[,7]_predict(x7s,data[,7])
ISMsalt40to95[,8]_predict(x8s,data[,8])
ISMsalt40to95[,9]_predict(x9s,data[,9])
ISMsalt40to95[,10]_predict(x10s,data[,10])
ISMsalt40to95[,11]_predict(x11s,data[,11])
ISMsalt40to95[,12]_predict(x12s,data[,12])

write(t(ISMsalt40to95),file="ISMsalt40to95.txt",ncol=1)

#####
#Improved Natural Salt Model
#
#####
#stats_array(99, dim=c(660,5,100))
KNNVhnsalt_matrix(0,660,100)

index_1:45

kk_sqrt(45)
kk_round(kk)
kk_12
W_1:kk
W_1/W
W_W/sum(W)
W_cumsum(W)

#   Get the first year..

for(k in 1:100){
#WYmonflow_matrix(KNNhflow[1:348,k], ncol=12, byrow=T) #10/1940 to
9/1969
#WYmonflow_matrix(USGSglen[1:348,3], ncol=12, byrow=T) #10/1940 to 9/
1969
#WYmonflow_matrix(BCNatFlow[421:1080,1], ncol=12, byrow=T) #10/1940
to 9/1995
Modelflow_matrix(BCNatFlow[421:1080,1], ncol=1, byrow=T) #10/1940 to
9/1995

```



```
WYmonflow_matrix(BCNatFlow[421:960,1], ncol=12, byrow=T) #10/1940 to
9/1985
```

```
xsalt_1:660
```

```
for(i in 1:660){
#xprev_KNNtnflow[i,k]
xprev_Modelflow[i,1]
j_i %% 12
if(j == 0)j_12
j1_j
xx_abs(xprev-WYmonflow[,j1])

xz_order(xx)
xz_xz[1:kk]
xx_runif(1,0,1)
xy_c(xx,W)
xx_rank(xy)
i1_xz[xx[1]]

if(j == 1)xm_predict(x1s,xprev)
if(j == 2)xm_predict(x2s,xprev)
if(j == 3)xm_predict(x3s,xprev)
if(j == 4)xm_predict(x4s,xprev)
if(j == 5)xm_predict(x5s,xprev)
if(j == 6)xm_predict(x6s,xprev)
if(j == 7)xm_predict(x7s,xprev)
if(j == 8)xm_predict(x8s,xprev)
if(j == 9)xm_predict(x9s,xprev)
if(j == 10)xm_predict(x10s,xprev)
if(j == 11)xm_predict(x11s,xprev)
if(j == 12)xm_predict(x12s,xprev)

#cat(i1)
xsalt[i]_xm + resids[i1,j]
#xsalt[i]_xm
stats[i,,k]_c(i,j,xprev,xm,resids[i1,j])

}

simdismon_matrix(xsim,nrow=12)
simdismon_t(simdismon)

#KNNhsalt[,k]_xsalt
```

```

KNNVhnsalt[,k]_xsalt
cat(k)
}

#write(t(KNNhsalt),file="KNNhsalt.txt",ncol=100)
#write(t(KNNnsalt),file="KNNnsalt.txt",ncol=100)
write(t(stats),file="stats.txt",ncol=5)

#####
#Calculate Historic Salt Mass
#
#####
rebuildExportMass_scan(file="RebuildExportMass.txt") #9/1905 to 12/1995
rebuildMonRtnmass_scan(file="0725monRtnmass.txt") #9/1905 to 12/1995

KNNthsalt_matrix(99,660,100)
for (i in 1:100){
  KNNthsalt[,i]_KNNthsalt[1:660,i] - rebuildExportMass[421:1080] +
  ebuiltMonRtnmass[421:1080] #10/1970 to 9/1995

}

ISMhsalt_matrix(99,660,90)
for (i in 1:90){
  ISMhsalt[,i]_ISMNsalt[1:660,i] - rebuildExportMass[421:1080] + rebuild-
  MonRtnmass[421:1080] #10/1970 to 9/1995

}

SAMSNflowMatrix_matrix(99,1080,100)
for(k in 1:100){
  for(i in 1:1080){
    j_i %% 12
    l_i %% 90
    if(j == 0)j_12
    if(l == 0)l_90
    SAMSNflowMatrix[i,k]_SAMSNflow[j,l,k]
  }
  cat(k)
}

```

3.2 Sample Splus Data Plotting Code

```

#####
#Calculates basic statistical data then produces boxplots of the statistics

```

```

data_array(matrix(KNNHistflows),c(12,85,100))

tdata_array(99, dim=c(85,12,100))
for(i in 1:100){
tdata[:,i]_t(data[:,i])
}

#calculates statistical parameters

sammean_matrix(99,100,13)
samstdev_matrix(99,100,13)
samvar_matrix(99,100,13)
samcor_matrix(99,100,13)
samskw_matrix(99,100,13)

for(i in 1:100){
  for(j in 1:12){
    sammean[i,j]_mean(tdata[:,j,i])
    samstdev[i,j]_stdev(tdata[:,j,i])
    samvar[i,j]_var(tdata[:,j,i])
    samskw[i,j]_sum((tdata[:,j,i]-sammean[i,j])^3)
    samskw[i,j]_samskw[i,j]/85
    samskw[i,j]_samskw[i,j]/samstdev[i,j]^3
  }

  for(j in 2:12) {
    samcor[i,j]_cor(tdata[:,j,i],tdata[:,j-1,i])
  }
samcor[i,1]_cor(tdata[1:74,12,i],tdata[2:85,1,i])

samann_1:85
  for(l in 1:85){
    samann[l]_(sum(tdata[l,i]))
  }
  sammean[i,13]_mean(samann)
  samstdev[i,13]_stdev(samann)
  samskw[i,13]_sum((samann-sammean[i,13])^3)
  samskw[i,13]_samskw[i,13]/85
  samskw[i,13]_samskw[i,13]/samstdev[i,13]^3
  samcor[i,13]_cor(samann[1:84],samann[2:85])
}

monbox_function(data)
{

#data is the matrix containing the monthly flows - 77 rows and 13 columns

```

```
months_c("Oct","Nov","Dec","Jan","Feb","Mar","Apr","May","Jun","Jul","Aug",
"Sep","Ann"
)
```

```
xs_1:13
zz_boxplot(split(t(data),xs),plot=F,cex=1.0)
zz$names_rep("",length(zz$names))
z1_bxp(zz,ylim=range(data),xlab="",ylab="",style.bxp="old",cex=1.25)
axis(1,at=z1,labels=months,cex=1.25)
points(z1,obs,pch=16,cex=1.25)
lines(z1[1:12],obs[1:12],lty=1)
#title(main="Boxplots of monthly mean",cex=1.0)
}
```

```
monbox(sammean)
title(main="Boxplots of Simulated Mean of Flows PAR(1)",cex=1.0)
dev.copy(postscript, file="sammean.ps")
dev.off()
dev.off()
monbox(samstdev)
title(main="Boxplots of Simulated Standard Deviation of Flows PAR(1)",
,cex=1.0)
dev.copy(pscript, file="samstdev.ps")
dev.off()
dev.off()
monbox(samcor)
title(main="Boxplots of Simulated Lag(1) Correlation of Flows PAR(1)",
,cex=1.0)
dev.copy(pscript, file="samcor.ps")
dev.off()
dev.off()
obs_obsskw
monbox(samskw)
title(main="Boxplots of Simulated Skew of Flows PAR(1)",cex=1.0)
dev.copy(pscript, file="samskw.ps")
dev.off()
dev.off()
```

```
monbox(ismmean)
title(main="Boxplots of Simulated Mean of Flows ISM",cex=1.0)
dev.copy(postscript, file="ismmean.ps")
dev.off()
dev.off()
monbox(ismstdev)
title(main="Boxplots of Simulated Standard Deviation of Flows ISM",
,cex=1.0)
```

```

dev.copy(pscript, file="ismstdev.ps")
dev.off()
dev.off()
monbox(ismcor)
title(main="Boxplots of Simulated Lag(1) Correlation of Flows ISM
",cex=1.0)
dev.copy(pscript, file="ismcor.ps")
dev.off()
dev.off()
obs_obsskw
monbox(ismskw)
title(main="Boxplots of Simulated Skew of Flows ISM ",cex=1.0)
dev.copy(pscript, file="ismskw.ps")
dev.off()
dev.off()

#*****
#produces pdf with 100 simulations compared to history

SAMS data
#11 is August shows bivariate

h_hnorm(nat1805[,1])
zx_sm.density(nat1805[,1],h)
points_zx$eval.points
hden_zx$estimate

samdensity0_matrix(0,100,100)
for(i in 1:100){
  h_hnorm(tdata[,1,i])
  zy_sm.density(tdata[,1,i],h,eval.points=points)
  samdensity0[,i]_zy$estimate
}

#*****
#routine to produce and graph boxplots for 100 pdf's

pdfbox_function(data,points)
{
#data is the matrix containing the monthly flows - 85 rows and 13 columns

xs_1:100
zz_boxplot(split((data),xs),plot=F,cex=1.0)
zz$names_rep("",length(zz$names))
z1_bxp(zz,xlim=c(0,714),ylim=range(data),xlab="flow (acre-feet/

```

```

month)",ylab="Probability Density",style.bxp="old",cex=1)
#points(z1,hden,cex=1.25)
lines(z1[1:100],hden,lty=1,xaxt="s",lwd=2)
evalloc_points
ntot_length(z1)
n1mid_as.integer(ntot/4)
n2mid_as.integer(ntot/2)
n3mid_as.integer(ntot*3/4)
z2_1:5
z2[1]_z1[1]
z2[2]_z1[n1mid]
z2[3]_z1[n2mid]
z2[4]_z1[n3mid]
z2[5]_z1[ntot]
n1_1:5
n1[1]_0
n1[2]_evalloc[n1mid]
n1[3]_evalloc[n2mid]
n1[4]_evalloc[n3mid]
n1[5]_evalloc[ntot]
n1_round(n1,digit=0)
n1_as.character(n1)
axis(1,at=z2, label=n1)
#title(main="SAMS AR(1) for October",cex=1.0)
}

#title(main="SAMS AR(1) for August",cex=1.0)

#title(main="K-NN nonparametric for October",cex=1.0)

#title(main="K-NN nonparametric for August",cex=1.0)

pdfbox(samdensity0)
title(main="PDF's from SAMS AR(1) method for October",cex=1.0)

pdfbox(xdensity1)
title(main="PAR(1) parametric for annual totals",cex=1.0)

```