INTEGRATED WATER RESOURCES MANAGEMENT IN A CHANGING CLIMATE: WATER QUALITY, QUANTITY, ALTERNATIVE, AND POLICY MAKING

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of Doctor of Philosophy

with a

Major in Biological & Agricultural Engineering

in the

College of Graduate Studies

University of Idaho

by

JungJin Kim

Major Professor: Jae Hyeon Ryu, Ph.D.

Committee Members: John Abatzoglou, Ph.D.; Jan Boll, Ph.D.; Robert Heinse, Ph.D.

Department Administrator: Ching-An Peng, Ph.D.

August 2017

Authorization to Submit Dissertation

This dissertation of JungJin Kim, submitted for the degree of Doctor of Philosophy with a Major in Biological & Agricultural Engineering and titled "INTEGRATED WATER RESOURCES MANAGEMENT IN A CHANGING CLIMATE: WATER QUALITY, QUANTITY, ALTERNATIVE, AND POLICY MAKING," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
Jae Hyeon Ryu, Ph.D.

Committee Members: _____ Date: _____
John Abatzoglou, Ph.D.

_____ Date: _____
Jan Boll, Ph.D.

_____ Date: _____
Robert Heinse, Ph.D.

Department Administrator: _____ Date: _____
Ching-An Peng, Ph.D.

Abstract

Land Use and Land Change (LULC) and climate variability are significant elements of the integrated water resources management to deal with water quality and quantity at the rural and urban interface. Adaptive strategies to develop sustainable water resources, therefore, are necessary to evaluate the water resources system in a changing global environment. In this dissertation, a series of analytical processes are implemented. First, gap filling processes are conducted by identifying a threshold of missing levels using daily precipitation series. The result indicates that about 15 % missing level of data is plausible to construct daily precipitation series for further hydrological analysis when the Gamma Distribution Function (GDF) is used as an estimation method. It is expected that such a finding can contribute to gap filling guidelines in the field, especially for water managers and hydrologists to take advantage of skillful estimates for missing daily precipitation data. Specifically, a gamma distribution function with statistical correlation (GSC) coupled with cluster analysis (CA) is used to estimate daily precipitation records and the result shows that GSC/CA outperformed other gap filling methods when rain and no-rain conditions are applied in the study area. Additionally, the dissertation identifies how Hydrological Simulation Program-Fortran (HSPF) model can improve performance depending on different sizes of watershed discretization levels within rainfall-runoff modeling settings. All simulations at different discretization levels above approximately 23% of the basin size resulted in satisfactory performances. However, the modeling performances are limited when the catchment size reaches below 8.18% of the basin size, regardless of automatic calibration efforts. The result indicates that basin discretization at finer scales does not necessarily improve HSPF simulation results with Next-Generation Radar (NEXRAD) inputs.

Computer parallelism and spatio-temporal analysis is another avenue in this dissertation in the sense that the proposed method can advance hydrological simulations using HSPF along with different calibration scenarios. Thus, the result indicates that computer parallelism could save computation time up to 90%, while simulation improvement is achieved by 81%. This finding, therefore, will provide useful insights for hydrologists to design and set up their hydrological modeling exercises in a changing climate. As part of this

dissertation, the evaluation of potential changes in water quality and quantity associated hydrologic changes in response to climate and LULC changes is also investigated. For example, HSPF model based on future LULC changes associated with climate scenarios was applied to generate climate-induced streamflow and to evaluate water quality in the Boise River Watershed (BRW). The result shows that the combined impact of LULC changes and climate variability on the BRW is inevitable, but seasonal variations in streamflow and water quality are primarily noticeable. This finding may provide useful information to develop sustainable water resources management when both water quality and quantity is an issue at the snow dominated watershed in a changing climate.

Lastly, under the circumstance: increasing concerns on water quality associated with LULC changes and climate variability, identification of critical hot spots (CHSs) and the implementation of mitigation activities using low impact development (LID)/Best Management Practices (BMPs) is a critical exercise to improve water quality at the BRW. Based on preliminary environmental analysis using different methods, load per sub-area index (LPSAI) is selected as the most cost-effective method because it can reduce the average pollutant loads at the watershed outlet while minimizing cost.

Acknowledgements

It would not have been possible to write this dissertation without the help and support of the kind people around me, to only some of whom it is possible to give a particular mention here.

Most of all, I would like to express the deepest appreciation to my major professor, Dr. Jae Ryu. This research and dissertation would not have been possible without his advice, expertise, encouragement, and persistent help. Since 2011 he has supervised my research and has truly improved my academic life. I greatly appreciate the time we spend discussing parallelizing calibration, climate change, and the development of optimization watershed management tool in general and the freedom he gave me to develop all my ideas. His guidance was fundamental throughout all my Ph.D. program. I would like to thank the other members of my dissertation committee members: Dr. John T. Abatzoglou, Dr. Jan Boll, and Dr. Robert Heinse. All of them are an inspiration for my future career as an engineer and their insight, feedback, and advice was essential for my dissertation. Finally, I would like to voice my gratitude to all my friends and colleagues in South Korea and the University of Idaho who have been supportive my academic life.

Table of Contents

Authorization to Submit Dissertation ..... **ii**

Abstract ..... **iii**

Acknowledgements ..... **v**

Table of Contents ..... **vi**

Lists of Tables ..... **xii**

Lists of Figures ..... **xvii**

CHAPTER 1. INTRODUCTION **1**

1.1. Overview **1**

1.2. Research objectives **5**

1.2.1. Climate data analysis 5

1.2.2. Hydrologic simulations 5

1.2.3. LULC and climate outlook 6

1.2.4. Optimal selection for water management alternatives 6

1.3. Outline of the dissertation **6**

CHAPTER 2. QUANTIFYING A THRESHOLD OF MISSING VALUES FOR GAP FILLING PROCESSES IN DAILY PRECIPITATION SERIES **9**

2.1. Introduction **9**

2.2. Methodology **11**

2.2.1. Autoregressive (AR) Model 11

2.2.2. Sampling Techniques 13

2.2.3. Gamma Distribution Function (GDF) 15

2.2.4. Performance measure....17

2.3. Site Description and Implementation....18

2.4. Results....20

CHAPTER 3. A HEURISTIC GAP FILLING METHOD FOR DAILY PRECIPITATION SERIES....24

3.1. Introduction....24

3.2. Study area....27

3.3. Methodology....29

3.3.1. Cluster Analysis (CA) method....33

3.3.2. Gauge Mean Estimator (GME)....34

3.3.3. Inverse Distance weighting (IDW) method....34

3.3.4. Ordinary Kriging (OK) method....35

3.3.5. Gamma distribution function with statistical correlation (GSC)....36

3.3.6. Performance measures criteria....37

3.3.7. Performance measures....38

3.4. Results....39

CHAPTER 4. THRESHOLD OF BASIN DISCRETIZATION LEVELS FOR HSPF SIMULATIONS WITH NEXRAD INPUTS....49

4.1. Introduction....49

4.2. Study area....52

4.3. Methodology....54

4.3.1. Input data....54

4.3.2. Hydrologic Simulation Program-Fortran (HSPF)....58

4.3.3. Hydrological modeling processes....62
4.3.4. Model calibration....63
4.3.5. Calibration Statistics....65
4.4. Results....66
4.4.1. Evaluating model performance at different catchment sizes....68
4.4.2. Identification of basin discretization threshold....74
CHAPTER 5. IMPROVING HYDROLOGICAL SIMULATIONS USING HSPF IN COMPUTER PARALLELISM....77
5.1. Introduction....77
5.2. Study area....79
5.3. Material and methods....81
5.3.1. Linux Cluster System (LCS)....81
5.3.2. System setup in sLCS....83
5.3.3. Time Series PROCessor (TSPROC)....84
5.3.4. BEO-Parameter ESTimation (BEOPEST)....84
5.3.5. HSPF model input datasets....86
5.3.5.1. Climate datasets....86
5.3.5.2. Observed streamflow and water quality data....89
5.3.5.3. Geological dataset....90
5.3.6. Streamflow calibration scenarios....92
5.3.7. Water quality calibration....95
5.3.8. Parallel performance....96
5.3.9. HSPF performance....97

5.4. Results.....98

5.4.1. Parallel Performance.....98

5.4.2. Model performance by calibration scenarios.....104

5.4.3. Results of calibrated and validated streamflow using SCO 4.....106

5.4.4. Results of calibrated sediment and water quality using SCO 4.....113

CHAPTER 6. IMPACTS OF CLIMATE AND LAND USE CHANGES ON STREAMFLOW AND WATER QUALITY.....117

6.1. Introduction.....117

6.2. Materials and methods.....120

6.2.1. Future Land use and land cover (LULC).....120

6.2.2. Global Climate Models (GCMs) from the fifth phase of the Climate Model Intercomparison Project (CMIP5).....123

6.2.2.1. Optimal GCM selections.....126

6.2.2.2. Statistical downscaling.....127

6.2.2.2.1. Statistical Downscaling and Bias correction (SDBC) method.....129

6.2.2.2.2. Temporal downscaling.....131

6.2.2.3. Potential Evapotranspiration (PET) of SDBC GCMs.....133

6.3. Results.....134

6.3.1. Future LULC variation.....134

6.3.2. Impacts of future LULC changes.....138

6.3.2.1. Streamflow variation.....138

6.3.2.2 Sediment load variation.....138

6.3.2.3. TN load variation.....139

6.3.2.4. TP load variation.....140

6.3.3. The selection of optimal GCMs 150

6.3.4. The performance evaluation of SDBC GCMs 152

6.3.5. Future precipitation and temperature generated by three SDBC GCMs 157

6.3.6. Simulated streamflow evaluation for three SDBC GCMs 161

6.3.7. Impacts of changes in climate and LULC. 165

6.3.7.1. Maximum streamflow variation 165

3.3.7.2. Mean streamflow variation 166

6.3.7.3. Minimum streamflow variation 168

6.3.7.4. Sediment load variation 168

6.3.7.5. TP load variation 169

6.3.7.6. TN load variation 171

6.3.7.7. Water temperature variation 172

CHAPTER 7. DEVELOPMENT OF AN OPTIMIZATION WATERSHED MANAGEMENT TOOL 201

7.1. Introduction 201

7.2. Model input data 204

7.3. Optimization Watershed Management Tool (OWMT) 204

7.3.1. Spatial targeting methods 204

7.3.1.1. Concentration Impact Index (CII) 205

7.3.1.2. Load Impact Index (LII) 205

7.3.1.3. Load per Sub-basin area index (LPSAI) 205

7.3.2. LID/BMP implementation plans 206

7.3.3. MCDA development for optimal LID/BMP selection 206

7.3.3.1. Analytical Hierarchy Process (AHP) 207

7.3.3.1.1. Consistency Index (CI) 208

7.3.3.1.2. Priorities derivation using approximate method 209

7.3.4. Total cost of LID/BMP installation 209

7.4. Results.....210

7.4.1. The location prioritization for spatial targeting methods.....210

7.4.1.1. CII method 210

7.4.1.2. LII method 211

7.4.1.3. LPSAI method.....211

7.4.2. The best LID/BMP selections at CHSs using AHP method 216

7.4.3. Evaluation of LID/BMP effectiveness.....220

CHAPTER 8. SUMMARY AND CONCLUSIONS 225

8.1. Quantifying a threshold of missing values for gap filling processes in daily precipitation 225

8.2. A heuristic gap filling method for daily precipitation series 226

8.3. Threshold of basin discretization levels for HSPF simulations with NEXRAD inputs.....227

8.4. Improving hydrological simulation using HSPF in computer parallelism.....229

8.5. Impacts of climate and land use changes on streamflow and water quality 231

8.6. Development of an optimization watershed management tool.....232

References.....235

Appendix A: Installation and configuration of DRBL.....255

Lists of Tables

Table 2.1. R^2 values from four sampling windows (whole, front, middle, and rear) at different missing levels (5%, 10%, 15%, 16%, 17%, and 18%). 22

Table 2.2. Total number of RDI and p values from four sampling windows (whole, front, middle, and rear) at different missing levels (5%, 10%, 15%, 16%, 17%, and 18%).....23

Table 3.1. The selected weather stations with less than 15% missing level of daily precipitation series (January 1, 1982 - May 31 2013). 32

Table 3.2. The result of CA in the study area. 40

Table 3.3. The performance measure of the gap filling methods with and without CA. 42

Table 3.4. The average performance variations for four gap filling methods with and without CA.....46

Table 3.5. Skill scores for the gap filling methods with CA.....48

Table 4.1. USGS stream gauges in the DMIP 2.....53

Table 4.2. Land Use classification (km^2) for DMIP2 basins. 57

Table 4.3. Initial parameter values and ranges of HSPF. 65

Table 4.4. Calibrated and Uncalibrated Statistics during October 1, 1998–September 30, 2000. 67

Table 4.5. R and NS before and after Calibration at Different Catchment Scales in the Tahlequah Drainage.....72

Table 4.6. R and NS before and after Calibration at Different Catchment Scales in the Tahlequah Drainage.....73

Table 4.7. R and NS before and after Calibration at Different Catchment Scales in the Elk River Drainage. 73

Table 4.8. R and NS before and after Calibration at Different Catchment Scales in the Baron Fork Drainage....73

Table 5.1. Constant values for dew point calculation....88

Table 5.2. Initial values and parameters range for HSPF model. 94

Table 5.3. Calibration model parameters of sediment erosion and solids wash-off. 95

Table 5.4. Calibration model parameters of nutrients, DO, and water temperature. 96

Table 5.5. Total calibration time, speedup, and efficiency in using the different numbers of core computing processes in parallel calibration. 102

Table 5.6. The model performance comparisons of calibrated streamflow by calibration scenarios....105

Table 5.7. General calibration targets of HSPF application (Donigian, 2000). 107

Table 5.8. The results of model performance for the calibrated and validated streamflow at calibration target points in the BRW. 107

Table 5.9. Model performance of calibrated water quality. 113

Table 6.1. Information of IPCC SRES scenarios (Sohl et al., 2012a)....122

Table 6.2. The information of total GCMs. 125

Table 6.3. LULC classification and variation for 2011, 2080, and 2100 LULC. 136

Table 6.4. Changes in monthly maximum streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition. 141

Table 6.5. Changes in monthly mean streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition. 142

Table 6.6. Changes in monthly minimum streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition. 143

Table 6.7. Changes in monthly sediment load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.....144

Table 6.8. Changes in monthly TN load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.....145

Table 6.9. Changes in monthly TP load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.....146

Table 6.10. The GCMs ranked according to weighting values between seasonal precipitation and temperature using normalized HSS, FSS, and RMSE scores.....151

Table 6.11. The performance comparisons for monthly SDBC GCMs. 153

Table 6.12. Future precipitation variations under RCP2.6, RCP4.5, and RPC8.5 climate change in three future periods relative to the baseline period (1981-2005). 159

Table 6.13. Future temperature variations under RCP2.6, RCP4.5, and RPC8.5 climate change in three future periods relative to the baseline period (1979-2005). 160

Table 6.14. Comparison of mean streamflow for three SDBC GCMs at all calibration target points. 161

Table 6.15. Variations of monthly maximum streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).....173

Table 6.16. Variations of monthly mean streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).....174

Table 6.17. Variations of monthly minimum streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).....175

Table 6.18. Variations of monthly sediment load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005). 176

Table 6.19. Variations of monthly TP load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005). 177

Table 6.20. Variations of monthly TN load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005). 178

Table 6.21. Variations of monthly mean water temperature by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).....179

Table 7.1. Weight vectors of criteria properties by application scenarios.....206

Table 7.2. Scale of relative important (adapted from Saaty, 1980) 208

Table 7.3. Random indices from Saaty (1980). 208

Table 7.4. Construction costs and annual maintenance costs of LID/BMPs (Brown and Schueler, 1997; County, 2014; Dhalla and Zimmer, 2010; NCDENR, 2007; USEPA, 1999) 210

Table 7.5. The classification ranges for the priority areas and the selected sub-watersheds by spatial targeting methods. 212

Table 7.6. LID/BMPs implementation area based on spatial target methods.....212

Table 7.7. Pairwise comparison matrix created for sub-watershed 78 based on sediment reduction for CII method at sub-watershed level. 217

Table 7.8. Pairwise comparison matrix created for sub-watershed 78 based on TP reduction for CII method for CII method at sub-watershed level. 217

Table 7.9. Pairwise comparison matrix created for sub-watershed 78 based on TN reduction for CII method at sub-watershed level. 217

Table 7.10. Weight vector calculation of LID/BMPs for sediment, TN, and TP reduction for CII method at sub-watershed 78.....218

Table 7.11. Weight vector calculation of LID/BMPs for total cost and area condition. 218

Table 7.12. Decision matrix of LID/BMPs for all criteria developed for sub-watershed 78. 218

Table 7.13. The selected individual LID/BMPs based on each application scenario at CHSs using CII method. 219

Table 7.14. The selected individual LID/BMPs based on each application scenario at CHSs using LII method. 219

Table 7.15. The selected individual LID/BMPs based on each application scenario at CHSs using LPSAI method. 220

Table 7.16. Total pollutant load reduction per unit area and total cost per unit pollutant load reduction for LID/BMPs implementations. 222

Lists of Figures

Figure 2.1. Workflow to identify a threshold of gap filling for daily precipitation data. 11

Figure 2.2. An example of random sampling at 5 % missing level. 14

Figure 2.3. Three different sampling windows (front, middle, and rear section) at 5 % missing level. 15

Figure 2.4. Statistical correction procedure of precipitation data (mm) for missing values. ...16

Figure 2.5. Location of the study area at Ada County in Idaho. 20

Figure 2.6. The boxplot of R^2 values from 1,000 realizations within four sampling windows (whole, front, middle, and rear) at different missing levels (5, 10, 15, 16, 17, and 18 %).....22

Figure 3.1. Weather stations for daily precipitation with different missing levels in the study area. 28

Figure 3.2. A flowchart of the proposed study.....31

Figure 3.3. The result of step-by-step CA application. 40

Figure 3.4. The boxplot of performance measures with and without CA for gap filling methods. 45

Figure 3.5. The boxplot of skill scores for the gap filling methods. 47

Figure 4.1. Stdy area of DMIP 2 and gage points. 54

Figure 4.2. Soil texture classification for the study area. 57

Figure 4.3. Structure chart for PERLND module (Bicknell et al., 2001). 59

Figure 4.4. Structure chart for IMRLND module (Bicknell et al., 2001). 60

Figure 4.5. Structure chart for RCHRES module (Bicknell et al., 2001). 61

Figure 4.6. Basin delineation of the study basins, including Elk, Illinois, and Baron Fork Rivers.....62

Figure 4.7. Hydrologic simulations for the study basins (October 1, 1998–September 30, 2000): (a) Illinois River at Tahlequah; (b) Illinois River at Siloam Springs; (c) Baron Fork River; (d) Elk River. 68

Figure 4.8. Examples of the first four discretization levels at different catchment sizes: (a) discretization level 1; (b) discretization level 2; (c) discretization 71

Figure 4.9. R and NS values at different basin scales below 22.63% of the drainage area. 75

Figure 4.10. R and NS values at different basin scales above 22.63% of the drainage area. ...76

Figure 5.1. The flowchart for hydrologic model calibration and application. 79

Figure 5.2. The Boise River Watershed (BRW) as the study area. 81

Figure 5.3. The configuration of the small Linux cluster system (sLCS). 83

Figure 5.4. Schematic of computer parallelism with BEOPEST in HSPF model. 86

Figure 5.5. Scheme of creating WDM file from NLDAS-2 using R-Parallel and SARA Time series Utility. 89

Figure 5.6. Geospatial datasets.....91

Figure 5.7. Model calibration scenarios by calibration locations and model parameter sets. ..93

Figure 5.8. Performance measures, including total calibration time (PT), the percentage of time reduction (PP), Speedup (PS), and Efficiency (PE) in computer parallelism. 100

Figure 5.9. Performance measures, including Correlation coefficient (R), Nash-Sutcliffe coefficient (NS), Root mean square error (RMSE), Mean Square Error (MAE), and Percentage of Bias (PBIAS) from HSPF simulations. 105

Figure 5.10. Hydrological simulations from HSPF at calibration target points (1-6) in the BRW. 108

Figure 5.11. Scatter plots of calibration results for calibration target points (1-6) in the BRW. 111

Figure 5.12. Scatter plots of validation results for calibration target points (1-6) in the BRW. 112

Figure 5.13. Hydrographic of calibrated daily sediment (a), TN (b), TP (c), DO (d), and water temperature (f). 114

Figure 6.1. The flowchart for application of future LULC, climate change, and the combined impact of climate and LULC changes using HSPF model. 120

Figure 6.2. Schematic of general future LULC modeling based on IPCC SRES from 1992 to 2100 (Sohl et al., 2014).....121

Figure 6.3. The flowchart of creating HSPF model weather data for future climate using optimal GCMs. 124

Figure 6.4. Simple temporal disaggregation method.....133

Figure 6.5. Spatial distribution of future LULC changes. 135

Figure 6.6. 2080 and 2100 LULC variations under A1B, A2, and B1 scenarios relative to 2011 LULC. 137

Figure 6.7. Changes in monthly maximum streamflow (a), mean streamflow (b), minimum streamflow (c), sediment load (d), TN load (e), and TP load (f) in the 2080 LULC and 2100 LULC under three scenarios three scenarios. 147

Figure 6.8. Comparison of historical 25-year monthly climate data (1981- 2005) for NLDAS-2 and SDBC GCMs. 154

Figure 6.9. Comparison of monthly mean streamflow between NLDAS and SDBC GCMs at all calibration target points for the baseline period (1981-2005). 162
Figure 6.10. Comparison of monthly maximum streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 180
Figure 6.11. Comparison of monthly maximum streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 181
Figure 6.12. Comparison of monthly maximum streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 182
Figure 6.13. Comparison of monthly mean streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 183
Figure 6.14. Comparison of monthly mean streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 184
Figure 6.15. Comparison of monthly mean streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 185
Figure 6.16. Comparison of monthly minimum streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 186

Figure 6.17. Comparison of monthly minimum streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 187
Figure 6.18. Comparison of monthly minimum streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 188
Figure 6.19. Comparison of monthly sediment load for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 189
Figure 6.20. Comparison of monthly sediment load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).....190
Figure 6.21. Comparison of monthly sediment load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 191
Figure 6.22. Comparison of monthly TP load for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).....192
Figure 6.23. Comparison of monthly TP load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 193

Figure 6.24. Comparison of monthly TP load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 194

Figure 6.25. Comparison of monthly TN load for bcc-scm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 195

Figure 6.26. Comparison of monthly TN load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 196

Figure 6.27. Comparison of monthly TN load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 197

Figure 6.28. Comparison of monthly mean water temperature for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 198

Figure 6.29. Comparison of monthly mean water temperature for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 199

Figure 6.30. Comparison of monthly mean water temperature for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3). 200

Figure 7.1. Flowchart of LID/BMPs application at CHS using OWMT. 204

Figure 7.2. Normalized priority areas for CII method using sediment, TN, and TP concentration. 213

Figure 7.3. Normalized priority areas for LII method using sediment, TN, and TP loads. 214

Figure 7.4. Normalized priority areas for LPSAI method using sediment, TN, and TP loads per each sub-watershed. 215

Figure 7.5. Pollutant load reduction per unit area and total cost per unit pollutant load reduction for different targeting methods with priority consideration scenarios. 223

Figure A.1. The first step of DRBL installation. 256

Figure A.2. Displaying option for slave nodes. 256

Figure A.3. Upgrading the operating system. 257

Figure A.4. Two options of the kernel selection. 257

Figure A.5. Assigning the names of DMS domain, NIS, and slave nodes. 258

Figure A.6. The results of internet connection (eth0) and network connection (wlan0) between master and slave nodes. 259

Figure A.7. Selecting the DRBL single system image and not clonezilla mode for the setting up the Beowulf cluster. 260

Figure A.8. Boot mode selection and root password setting. 261

Figure A.9. The mode setting for all slave nodes. 262

Figure A.10. Final setting for the mode. 262

CHAPTER 1. INTRODUCTION

1.1. Overview

An integrated water resources management (IWRM) associated with water quality and quantity is a critical exercise to promote environmental stewardship in the rural and urban interface. Water pollution induced by point source (PS) and nonpoint source (NPS) is a significant challenge for urban sustainability in the 21st century. PS pollutants have been managed by sewage disposal facilities and drainage system, but NPS pollutions are not fully understood in the sense that the heterogeneous discharge distribution and embedded uncertainty can often limit the ability to monitor, predict, and manage such pollutants, especially during intermittent rainfall. NPS load is mainly responsible for the negative effects of waterways associated with water pollution and degradation of ecological functions by a concern with secondary environmental issues such as eutrophication on waterways.

As US.EPA (2009) indicated that the approximately 60 percent of NPS load is contributed from agricultural areas, NPS_resulted in significant water concerns across the States (US.EPA, 2005a). Additionally, urbanization and climate changes are other drivers contributing to regional water quality and quantity issues. Since Land Use and Land Cover (LULC) changes driven by urbanization and economic development are inevitable, hydrologic characteristics and NPS concerns are of great interest in water resources management in the rural-urban interface, such as the Boise River Watershed (BRW). Urbanization increases impervious land as opposed to pervious land segments, resulting in baseflow reduction and flow augmentation during peak flows, frequent flash flood, stream degradation, and soil erosion. Especially, soil erosion and NPS loads by abrupt surface runoff in the impervious

land can be the reason for degradation of watershed hydrology and water quality standard in many waterways in the United States (US). Several studies have been conducted to investigate the variation of water quality and quantity for urbanization effects (Ding et al., 2015; Jarvie et al., 2002; Sliva and Williams, 2001; Tu and Xia, 2006). However, most studies using future LULC products are not clearly formulated to address regional water resources conditions because the existing information is too coarse to characterize land use attributes (Hurtt et al., 2011; Strengers et al., 2004) and the limited LULC classes (US.EPA, 2005b) within modeling settings. Sohl et al. (2014) developed annual future LULC data with high spatial resolution (250 m by 250 m) using the FOREcasting SCEnarios of Land-use change (FORE-SCE) model. The advantage of annual future LULC data is to provide the ability for analyzing carbon and greenhouse gas fluxes, and to improve the simulation of forestry activity as it classifies 12 land use categories. Multiple emission scenarios for the variations of future LULC data, therefore, are critical to assessing uncertainties embedded in the future prediction of landscape changes. Note that this LULC can also be utilized to simulate hydrologic change, ecological processes, biodiversity, and water variability in regional watersheds.

Climate change is another challenge for the water resources management since it in hydrologic context drives various negative effects on the water quality and quantity. According to the Intergovernmental Panel on Climate Change (IPCC) report (2013), mean annual temperature of the United States (US) will increase within the range of 1.2 to 4.5 °C by 2100. The increase in mean air temperature is expected to impact of snow accumulation and snow melting timing as a result of change in seasonal streamflow and precipitation pattern in mountainous areas. For example, the increase in winter and spring air temperature

influences the precipitation falling like the rain versus snow. Also, this results in increasing snowmelt runoff in spring and early summer, and declining the streamflow of the remainder of the year. Consequently, such changes will become more problematic of future water shortage and management in the watersheds. Numerous studies have been conducted to identify and analyze the variation of water quality impacts from climate change (Marshall and Randhir, 2008; Shamir et al., 2015; Tong et al., 2012; Wilson and Weng, 2011). However, few studies have been conducted for future climate change associated with potential LULC change induced by urbanization. An appropriate adaptation strategy for sustainable water management will be needed to identify the system vulnerability across a range of future conditions using spatial-temporal analysis in the regional watershed. Thus, several Global Climate Models (GCMs) should be assessed along with various emission scenarios (RCP26, RCP45, and RCP85) from the Coupled Model Intercomparison Project Phase 5 (CMIP5). It is necessary to define which GCMs can simulate best for the climate of a particular region, and hence hydroclimatic conditions. A further investigation will be made through this research activity to cross-validate the best hydroclimate scenarios representing local conditions in the study area.

The assessment of long-term hydrologic impacts associated with climate and LULC changes will be useful information to manage and improve water quality as well. To improve hydrological simulations, therefore, we will consider several elements. The observed environmental data series (e.g. daily precipitation) are affected by the performance of hydrologic models, but incomplete precipitation data are prevalent and inevitable due to erroneous reality associated with the deformation of the wind field, clogging, and electronical malfunction at the weather station especially in complex terrain. Therefore, missing data

within the proper missing levels are often filled with adequate values through good estimation processes for further environmental modeling and applications. The threshold of basin discretization will be determined to improve overall modeling performances that represents a proper catchment delineation level for their own applications. For high-resolution data and the application of many model parameters, computer parallelism will be used to improve hydrological simulation, thereby minimizing simulation/calibration efforts and computational costs.

Another addition to improving water resources management is to develop a decision support tool using system-based optimizing watershed management tool. It is important to identify 'hotspots' associated with problematic water quality and quantity, and to select workable Low Impact Development (LID) / Best Management Practices (BMPs) in the study area using multi-criteria approaches. Zhao et al. (2013) noted that there are challenges to identify, target, and alleviate NPS pollution loads due to their inherent randomness. For the decision makers and/or watershed manager, it is significant to identify vulnerable areas (hotspots) with concentrating management practices (Zhou and Gao, 2011). Much previous research reported the decision support tool is critical to manage natural resources (McIntosh et al., 2011; Panagopoulos et al., 2012), policy support (Van Delden et al., 2011), and urban water management (Aulinas et al., 2011; Price, 2011). Throughout preliminary data analysis processes, including gap filling and rainfall estimation, hydrological simulations are made to be routed into a decision support tool to evaluate best management alternatives ultimately for water policy makers and local stakeholders.

1.2. Research objectives

1.2.1. Climate data analysis

- Develop the control guideline of missing levels in an analytical framework
- Identify a proper level of missing values by enhancing the existing estimation method for gap filling
- Develop a heuristic gap filling method for improving the existing interpolation methods.
- Evaluate and the estimation accuracy for gap filling methods along with cluster analysis, and
- Compare the estimation performance of gap filling methods based on statistical measure and skill scores.

1.2.2. Hydrologic simulations

- Incorporate high-performance computing and complex geophysical modeling framework.
- Verity how the model responds to independent data periods using cross-validation techniques.
- Determine the proper catchment size for optimal basin discretization for optimal basins scale modeling
- Improve hydrologic model performance to better characterize impacts of climate change and urbanization on local hydrology,
- Enhance model calibration processes using computer parallelism, thereby minimizing efforts and operational costs, and
- Evaluate system performances based on parallelized parameter sets.

1.2.3. LULC and climate outlook

- Select optimal GCMs and evaluate the better represent local hydroclimate consequences in the study area,
- Compare and evaluate the existing downscaling methods and suggest guidelines to minimize uncertainties,
- Evaluate the variation of future LULC and climate in the study area related on,
- Determine the impacts of water quality and quantity by urbanization and climate change, and
- Evaluate the effects of the water quality and quantity by considering future LULC condition coupled with climate change.

1.2.4. Optimal selection for water management alternatives

- Develop system-based Optimizing Watershed Management Tool (OWMT),
- Determine and define the hotspots by urbanization and climate change based on water quality and geographic information,
- Improve water quality and quantity by applying LID/BMPs at hotspots, and
- Evaluate management alternatives to provide useful insights for the stakeholder group.

1.3. Outline of the dissertation

This dissertation is organized as follows. **Chapter 2** presents the proper missing level of precipitation data as a threshold for the existing gap filling process in daily precipitation. Autoregressive (AR) model is used to generate time series of daily precipitation for gap filling processes. Gamma distribution function (GDF) is explained and used along with statistical tests and analysis to determine a threshold of missing values. **Chapter 3** addresses a heuristic

gap filling method termed “gamma distribution function with statistical correlation (GSC) coupled with cluster analysis (CA). The proposed method is discussed along with mathematical formula of the existing gap filling methods, such as Gauge Mean Estimator (GME), Inverse Distance Weight (IDW), and Ordinary Kriging (OK) followed by performance measures using skill scores. **Chapter 4** includes threshold of basin discretization levels for Hydrological simulation FORTRAN-Program (HSPF). Hydrological model procedures such as basin delineation and calibration efforts are discussed along with fundamental statistics to evaluate model performance. A comparative analysis of model performance at different catchment sizes is discussed to determine the threshold of the basin discretization levels. **Chapter 5** presents HSPF simulation in computer parallelism. In this chapter, the originality of calibration strategies with computer parallelism in a Linux cluster is highlighted. We also describe the construction of the cluster framework and compiling process in parallel computing platforms. Parallelized calibration and different calibration scenarios are discussed to evaluate computer parallelism and HSPF model performance. **Chapter 6** addresses the variations of water quality and quantity in a changing climate and land use. In particular, future LULC products provided by USGS model outputs are applied to evaluate the impacts of LULC change on the study area. Three GCMs are selected based on skill scores (two GCMs) and literature review (one GCM). The selected GCMs are then evaluated and downscaled using statistical downscaling and bias correction (SDBC) to create finer resolution datasets to be used as HSPF’s inputs. Outcomes from SDBC coupled with GCMs are then compared with the North American Land Data Assimilation System (NLDAS-2) to justify the utilization of hydrologic model application. A variation in water quality and quantity driven by the combined impact of climate and future LULC changes is

also evaluated using the calibrated HSPF model. **Chapter 7** includes multi-criteria decision making processes based on streamflow and water quality simulations to find the critical hotspots for problematic locations. For the selected hotspots, optimal LDI/BMP methods are chosen to better manage water quality and quantity at the study area. Finally, **Chapter 8** summarizes the findings of the dissertation along with recommendations for future work.

CHAPTER 2. QUANTIFYING A THRESHOLD OF MISSING VALUES FOR GAP FILLING PROCESSES IN DAILY PRECIPITATION SERIES

"Quantifying a threshold of missing values for gap filling processes in daily precipitation series." *Water Resources Management*, vol.29 issue 11, 2015, pp 4173-4184.

2.1. Introduction

A gap filling method is typically used to construct a complete dataset with no missing value for daily precipitation. The daily precipitation is retrieved from local weather stations managed by federal, state, and local agencies, but often those datasets are incomplete due to missing values associated with systematic errors and/or random errors. Systematic error is incurred by the deformation of the wind field at gage location, where systematic undercatching is prevalent because of wind effects (Calder, 1993; Sevruk, 1996), whereas random errors are created by other physical disturbances, such as the clogging of the gage outflow and/or other mechanical and electronical malfunctions (Sevruk, 1996).

From hydrological perspectives, a complete set of valid precipitation data is critical to simulate streamflow in rainfall-runoff models for various water resources decision making (e.g., reservoir operations, drought mitigation, and flood control). Hydrologists often estimate missing values using mathematical and statistical methods, such as the arithmetic mean (Mn), Inverse Distance Weighting (IDW), Regression-based analysis (RA) methods, Kriging Estimation (KE) Method, and Gamma distribution function (GDF) to fill gaps in missing values (Hubbard et al., 2005; Linsley et al., 1982; Mair and Fares, 2011; Teegavarapu et al., 2009 and 2011; Teegavarapu, 2014a and 2014b; Westerberg et al., 2010). The IDW method, in particular, has been commonly used to estimate missing values in geographical sciences

and hydrology because the result from IDW method shows good performances in filling missing values, especially for spatially dense networks (Ahrens, 2006; Dirks et al., 1998; Garcia et al., 2008). Recent studies, however, indicate that the IDW method has a limitation in the sense that it tends to overestimate the number of rainy days or the existence of negative autocorrelation between the nearest weather stations (Garcia et al., 2008; Simolo et al., 2010). Alternatively, GDF method is now applied widely to overcome the limitation of the existing estimation methods, including Mn, IDW, and RA (Hasan and Croke, 2013; Simolo et al., 2010). Hasan and Croke (2013) showed that the Poisson-gamma method can better generate precipitation data and percentage of rainy, and no-rain days in comparison with observed precipitation data along with other estimation methods. Simolo et al. (2010) also observed that GDF method shows an accurate determination of rainfall occurrence and improves the reconstruction of intense precipitation events. One of key questions on gap filling methods mentioned above is what missing level of precipitation dataset would be critical for further hydrological processes and modeling efforts, yet little studies have been done to identify such a threshold of missing values during gap filling processes. Thus, knowing proper missing level of precipitation data as a threshold for the existing estimation method is valuable information for hydrologists to improve modeling performances as well as hydrological prediction. The main objective of the paper, therefore, is to identify a proper level of missing values (threshold) by enhancing the existing estimation method for gap filling in daily precipitation series based on the heuristic control guideline of missing levels in an analytical framework. The proposed method would be valuable assets in the field of hydrology and hydroclimatorogy in a changing climate.

2.2. Methodology

2.2.1. Autoregressive (AR) Model

Figure 2.1 illustrates a schematic of workflow for gap filling tasks and evaluation processes used in this study.

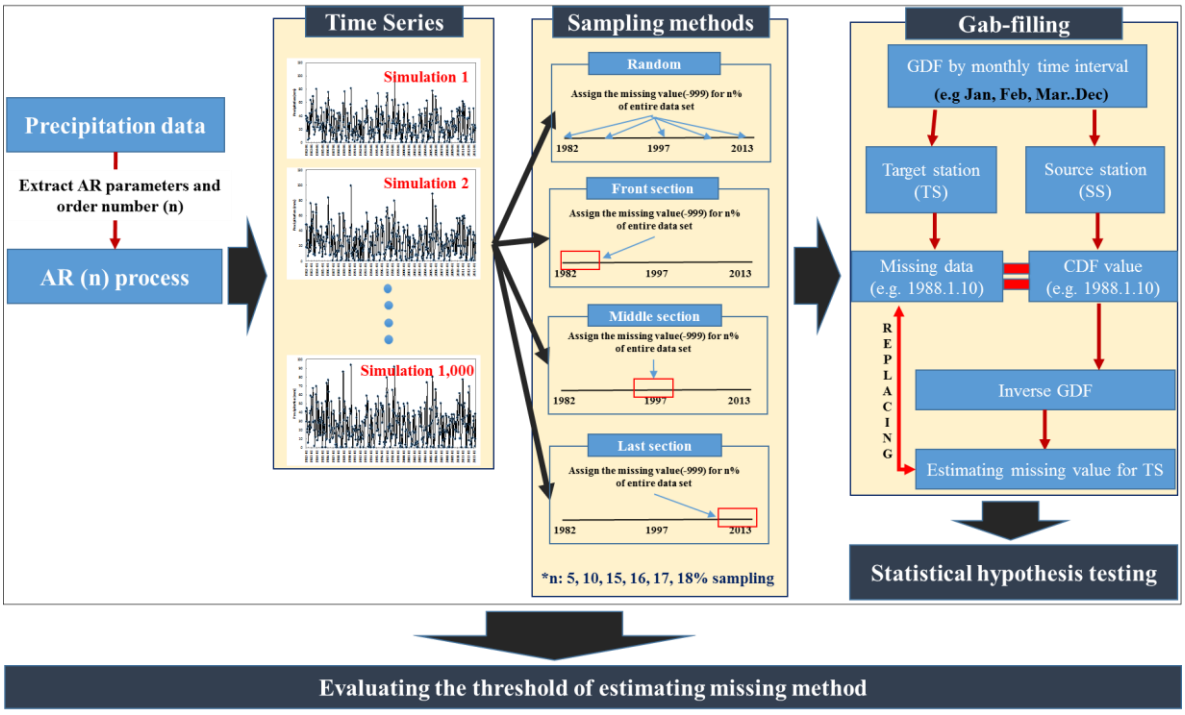


Figure 2.1. Workflow to identify a threshold of gap filling for daily precipitation data.

An autoregressive (AR) model with parameter n, AR (n), is first used to construct daily precipitation datasets. AR (n) model is a linear regression model and widely used for hydrological modeling and forecasting using weather forcing, such as precipitation and temperature (Chattopadhyay et al., 2011; Goswami and O'Connor, 2007; Hipel and McLeod, 1994; Salas et al., 1980). The typical form of AR (n) model with lead time j is given by:

$$Y_t = \sum_{=1}^{n} \emptyset_j \left(Y_{t-j}\right) + \varepsilon_t \qquad (2.1)$$

where, Y_t is the time dependent series (variable) at time t, $\emptyset_j$ is autoregressive parameter, and ε_t is the independent random variables that is normally distributed with mean zero and variance σ^2.

The Akaike Information Criterion (AIC) is employed to estimate the optimal order number (n) of AR model based on the information criterion by employing the relationship between the average mean and Kullback-Leibler condition (Akaike, 1973). The AIC value is computed as:

$$\text{AIC (n, n} *) = N(1 + \log(2\pi)) + Nlog(s^2) + 2n \tag{2.2}$$

where, s^2 is the estimated variance of the residuals resulted from the fitted model, n is order of AR model, and N is the number of data samples. The notation, n* is maximum number of order application to calculate AIC (Ing et al., 2010)

For diagnostic purposes, the Ljung-Box test is used to determine whether or not the data series can be considered independent and identically distributed sequences (iid) (Ljung and Box, 1978). Null hypothesis is defined as the sample data are iid. The Ljung-Box Q-statistic is defined by equation. (2.3) and (2.4).

$$\text{Q} = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \tag{2.3}$$

$$r_j = \frac{\sum_{t=j+1}(a_t - \bar{a})(a_{t-j} - \bar{a})}{\sum_{t=1}^{n}(a_t - \bar{a})^2} \tag{2.4}$$

where, n is sample size, m is the number of autocorrelations included in the statistic, r_j is the squared sample autocorrelation of residual series of the model at time lag j, a_t is the residual series, and $\bar{a}$ is the average of the residuals. Under the null hypothesis, since Q is a

approximately distributed as the sum of squares of the independent random variables from standard normal distribution, such values are chi-squared distributed with h degree of freedom. The iid hypothesis can be rejected at the level alpha if the value of Q exceeds the (1-alpha) quantile of the chi-squared distribution with h degree of freedom.

2.2.2. Sampling Techniques

Once time series of daily precipitation data are generated from the AR model, the selected subsets of the data series are considered missing values. To assign such missing values (e.g. −999.9) for the selected subset of the data series, four sampling windows; 1) random samples from whole dataset, and 2) three different sampling windows, including front, middle, and rear section are selected at 5 10, 15, 16, 17, and 18 % missing levels. Sampling from different missing levels within the selected windows is then carried out using random number generators. Consequently, this approach results in equal probability of being selected for the sample dataset replicating iid. Random sampling with respect to each different missing level (5, 10, 15, 16, 17, and 18 %) runs 1,000 times so that 1,000 precipitation realizations can be generated by the AR model during the study periods (e.g., January 1982–May 2013). Note that the selected missing levels are determined based on preliminary research findings indicating that the existence of the threshold is within 10–20 % missing levels. Thus, the preliminary research indicates that above 20 % missing level is too high to achieve performance targets on hydrological simulation based on Root Mean Square Error (RMSE) and Pearson Correlation Coefficient, whereas hydrological simulation with below the 10 % missing value of daily precipitation outperforms against that with the 10 % above. Respectively, F**igure 2.2 and 2.3** illustrate the process of random sampling from whole data

range and three different sampling windows at 5 % missing level. If sampling is taken from whole data range at 5 % missing level during the study periods, approximately 5 out of 100 data points are replaced by missing values (−999.9).

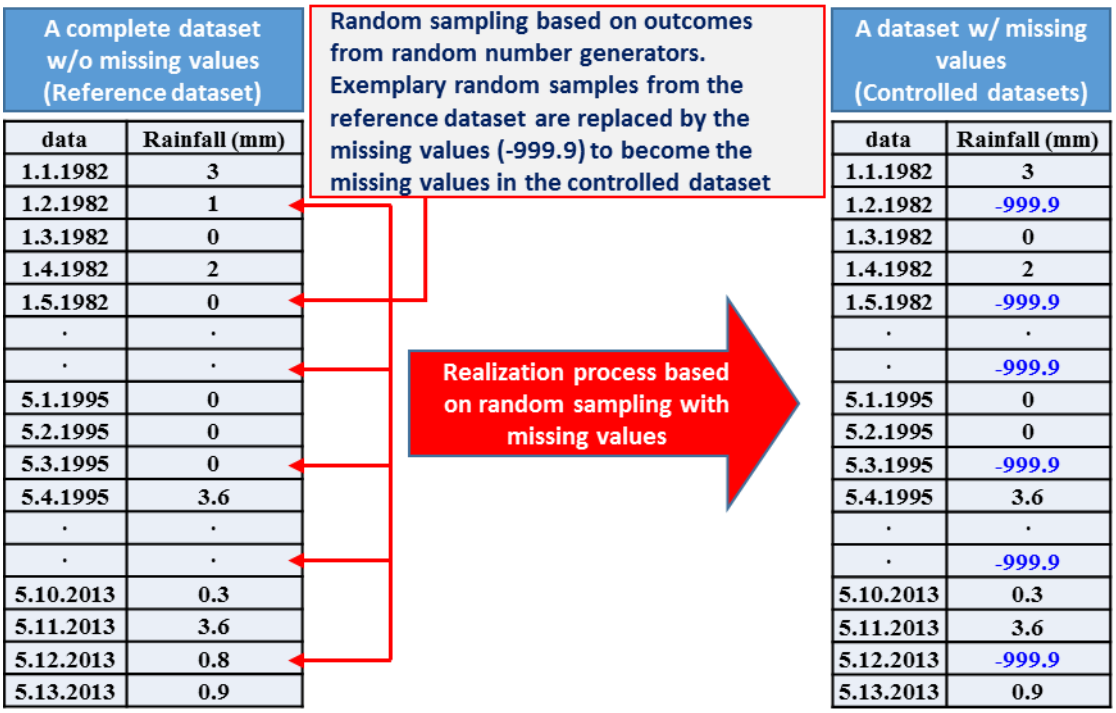


Figure 2.2. An example of random sampling at 5 % missing level.

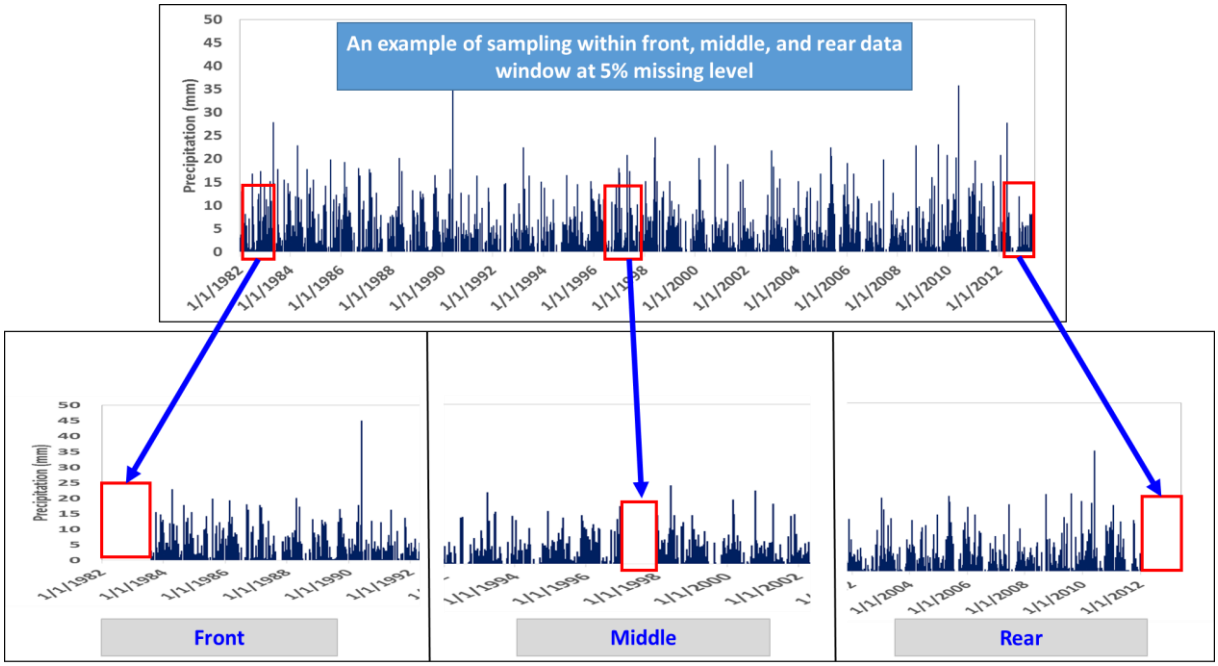


Figure 2.3. Three different sampling windows (front, middle, and rear section) at 5 % missing level.

2.2.3. Gamma Distribution Function (GDF)

The GDF is known as a proper theoretical distribution function to represent precipitation in the sense that no negative value of precipitation can be plotted. The GDF can emulate actual rainfall distribution for extreme cases although the typical rainfall may not be large. It provides some degree of flexibility in the shape of the distribution function from data range (e.g., exponential-decay forms for shape value), so this flexibility allows fitting any number of rainfall regimes to GDF. The accuracy of the GDF performance can be measured by comparing the cumulative distribution function of the estimated Gamma and empirical distributions (Wilks, 1995). The GDF has three different types, including, one-, two-, and three-parameter GDFs. The one-parameter GDF, however, is limited to describe hydrologic phenomena due to inflexibility of fitting to frequency distribution of hydrologic variables

(Aksoy, 2000). The three-parameter GDFs would be good but no significant advantage of its property over the two-parameter GDF has been found (Markovic, 1965). Therefore, a two-parameter GDF was used for gap filling in daily precipitation series. **Figure 2.4** shows a diagram of the estimating procedure for daily precipitation using the GDF. First, the cumulative distribution function (CDF) was defined for both reference station (RS) and target station (TS) from Jan 1982 to May 2013, for example, and then inverse CDF of the gamma function was used to estimate missing values using equations in Appendix. For instance, suppose that TS has the missing value (−999.9) in January 15. 2002 and the value of RS corresponding to the same date was recorded as 15.2 mm having about 99.48 % probability in gamma domain. The missing value of TS is then replaced by an arbitrary value having the same probability of 99.48 % in its own gamma domain, which is 16.47 mm for this particular case (See **Figure 2.4**). This implies that the missing value at TS is statistically adjusted by 8 % from RS through this estimation scheme also used in Ryu et al. (2009).

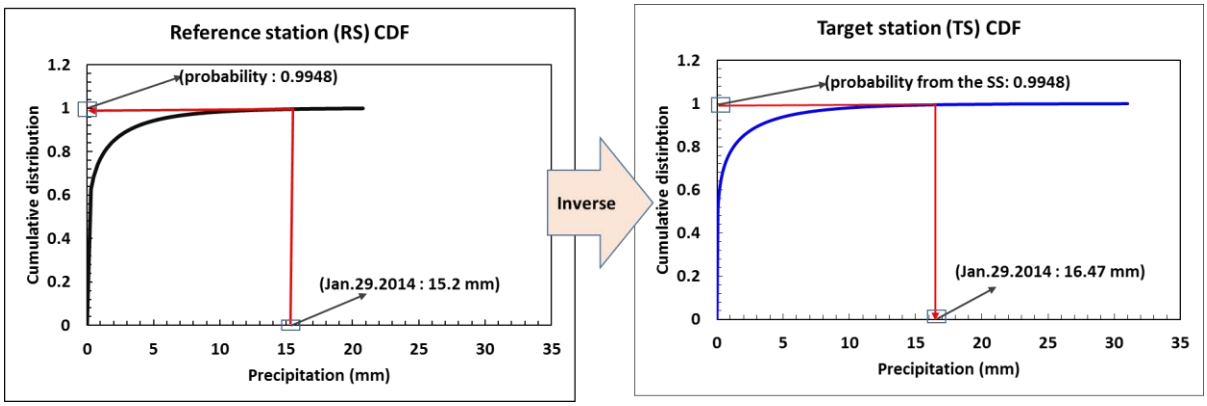


Figure 2.4. Statistical correction procedure of precipitation data (mm) for missing values.

2.2.4. Performance measure

Coefficient of determination (R^2) (equation 2.5), which is one of typical performance measures, was utilized to evaluate how GDF well performs to estimate missing values against original precipitation data. R^2 is the degree of colinearity between the filled datasets after sampling and the observed precipitation datasets. It ranges from 0.0 to 1.0. The closer to 1 indicates better performance of filling the missing data. A statistical hypothesis test is conducted to determine if the selected threshold of the gap filling method is suitable for further data analysis. A statistical hypothesis test is a systematic approach to verify how well the sample can represent the population along with statistical properties, such as mean and standard deviation. The critical level of a hypothesis can determine whether or not the alternative hypothesis is rejected in favor of the null hypothesis. To implement this hypothesis test, we developed two scenarios; the first scenario defines that R^2 value should be greater than 0.8, while the second scenario designates that R^2 value of the lower missing levels should be greater than that of higher missing levels at TS. For example, R^2 value of 10 % missing level should be greater than or equal to that of 15 % missing level. Likewise, R^2 of 5 % should be greater than R^2 of 10 % and above. Next, the R^2 difference index (RDI) was calculated for the second scenarios. Basically, the RDI is a measure to verify if the second scenario (e.g. RDI=all R^2 values of 10 % missing data minus all R^2 values of 15 % missing data) is satisfactory. The RDI can have positive if R^2 value of lower missing level is higher that R^2 value of higher missing level, or negative values in reverse condition, or zero for no difference. Therefore, if the RDI value between datasets with missing values at 15 and 16 % level has a negative value, we can imply that a threshold of missing level should be at 15 % missing level by knowing all RDI values below 15 % missing level are positive. Note that the

null hypothesis can be rejected with the 95 % of confident level if the calculated p-value is less than 0.05; otherwise there is no evidence to reject null hypothesis.

For this study, null hypothesis represents that the selected missing level is inadequate to apply the gap filling technique if R^2 values are less than 0.80 and the RDI is negative ($Hn_1 < 0.80$, $Hn_2 < 0.$); alternative hypothesis states that the proposed missing level is suitable for the gap filling method used in this study ($Ha_1 \geq 0.80$, $Ha_2 \geq 0$). The next step is then to calculate the test statistic from R^2 values and the RDI using z-test using equation (2.6) (Sprinthal, 2011). Finally, a proper threshold value is determined based on p-value from z-test statistic.

$$R^2 = \left(\frac{\frac{1}{N}\times\sum_{i=1}^{N}\left(P_{Qi}-\bar{P}_{Qi}\right)\times\left(P_{Si}-\bar{P}_{Si}\right)}{\sqrt{\frac{N\times\sum_{i=1}^{N}P_{Qi}^2-\left(\sum_{i=1}^{N}P_{Q1}\right)^2}{N\times(N-1)}}\times\sqrt{\frac{N\times\sum_{i=1}^{N}P_{Si}^2-\left(\sum_{i=1}^{N}P_{S1}\right)^2}{N\times(N-1)}}}\right)^2 \quad (2.5)$$

$$z = \frac{\bar{x}-u}{s/\sqrt{N}} \quad (2.6)$$

where, P_{Qi} is the observed precipitation data at time step i, and P_{Si} is the estimated precipitation data. $\bar{P}_{Qi}$ and $\bar{P}_{Si}$ are the mean of the observed and estimated precipitation data, respectively. N is sample size. $\bar{x}$ is sample mean, u is population mean, and s is standard deviation of population.

2.3. Site Description and Implementation

A weather station (COOP ID: 101022) near the Boise Airport in Ada county, Idaho was selected for the study area (See **Figure 2.5**). The climate characteristics of the study area indicate a semi-arid climate with four distinct seasons. Summer is hot and dry with high reaching 38 °C (100 °F). Winter is moderately cold with average temperature about −0.7 °C

(30.7 °F) in December. Sparing and fall is mild. The average annual precipitation is about 660 mm (26.02 in). A decent amount of hydroclimate information and environmental variables, including precipitation, temperature, wind speed, soil moisture, and other energy flux is available at the station that can be used to characterize weather extreme and hydrological modeling. Reliable datasets less than 1 % missing are available and ready to use for further analysis. Daily precipitation data series at the station from January 1, 1981–May 30, 2013are well recorded relatively in comparison with other stations nearby, including Swan Falls Power House (COOP ID:108928) and Boise Lucky Peak Dam (COOP ID: 101018), which are both more than 1 % missing values of the study period. It appears that data missing levels of the weather stations at Swan Falls and Lucky Peak Dam are 1.39 and 15.15 %, respectively. As described above, four sampling windows (whole, front, middle, and rear) at multiple missing levels (5, 10, 15, 16, 17, and 18 %) are taken into account to generate daily precipitation realizations using an AR model. AR (23) model was selected because the sample value of Q are 23.78 with m=23 from the equeation (3) and (4), which is less than 35.17 with the 95 percentile of chi-squared with 23° of freedom from the Chi-square table (Goldberg, 2000). This result suggests that the iid hypothesis is accepted at confidence level 0.05.

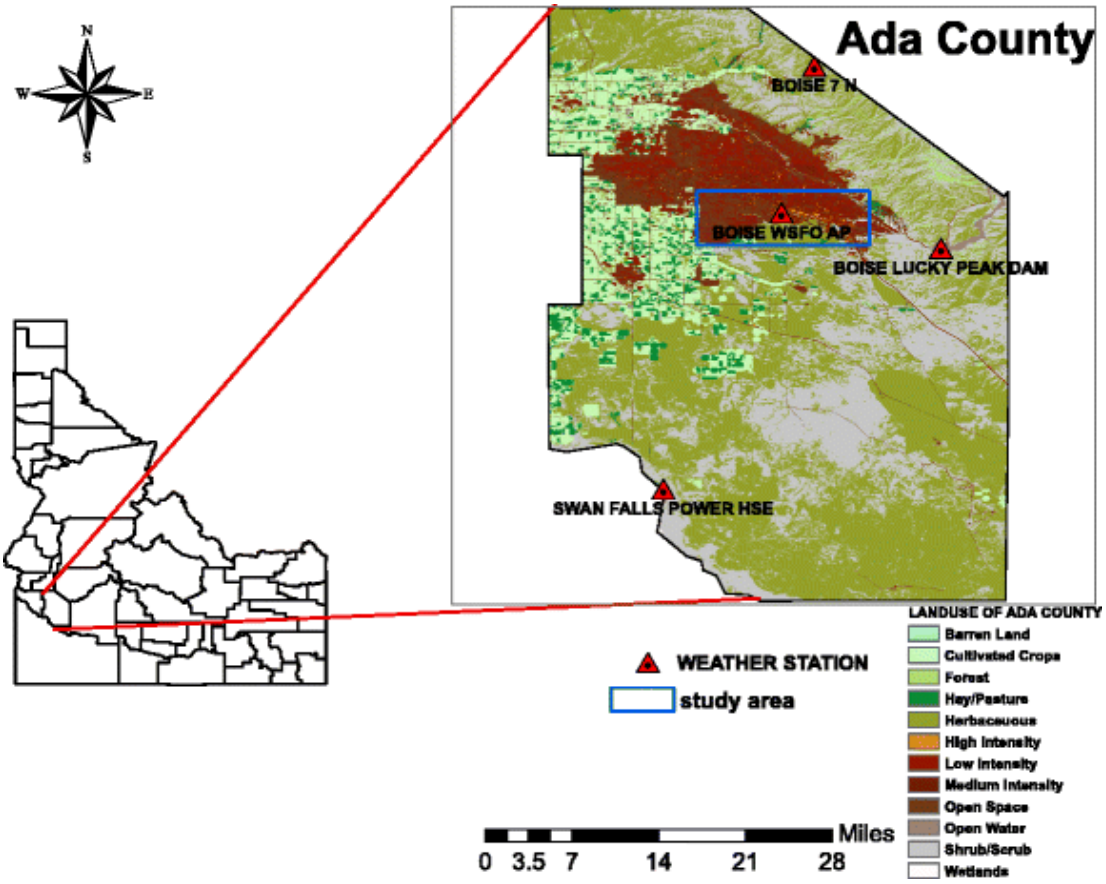


Figure 2.5. Location of the study area at Ada County in Idaho.

2.4. Results

Figure 2.6 shows box plots for R^2 values from four sampling windows (whole, front, middle, and rear) at different missing levels (5, 10, 15, 16, 17, and 18 %). The average R^2 values above 0.85 are listed in **Table 2.1**. Overall, there is variation of R2 values but higher R^2 values are observed for whole sampling window as opposed to the sectional sampling windows (front, middle, and rear). It appears that the sectional sampling windows provide similar results as shown in **Figure 2.6**. However, the R^2 values decrease as the missing level increases. This implies that the proportion of missing datasets clearly affects the robustness of the interpolation method for the missing values. Nonetheless, all R^2 values are higher than

0.80 and all Pvalues support the alternative hypothesis, regardless of the sampling windows (e.g., whole, front, middle, and rear) and missing value levels (e.g., 5, 10, 15, 16, 17, and 18 %). **Table 2.2** shows a total number of RDI values and the p-values from four sampling windows (whole, front, middle, and rear) at different missing levels (5, 10, 15, 16, 17, and 18 %). As shown in **Table 2.2**, all RDIs are greater than and equal to zero from different sampling windows except for whole sampling windows at different missing levels. P-values less than and equal to 15 % missing level are lower than 0.05, while that of above missing level 15 % (e.g., 15–16 %) is greater than 0.05 based on samples from the whole data window. Note that numerous negative values (164 out of 1,000: 16.4 % of dataset) were calculated using samples taken from the whole data window, where sampling takes place at 16 %.

Consequently, the null hypnosis is rejected because p-values of RDIs within 5–15 % are less than 0.05 so that 15 % missing level can be a threshold of gap filling to construct daily precipitation series using the GDF method. Note that the number of zero RDIs is almost 2–3 times more than the number of positive values at 15–16 % within rear sampling window. The similar study has been implemented in daily precipitation datasets (January 1, 1981- May 30, 2013) at Lincoln, Nebraska to verify that the selected threshold is reliably situated at 15 % missing levels. As expected, the result shows that R^2 values gradually decrease as the missing level increases until it reaches to 15 % missing level within different sampling windows (whole, front, middle, and rear). Although R^2 values decrease as the missing level increases within multiple sampling windows, R^2 values above 16 % missing level for whole sampling window fluctuate (Not shown in this paper) so that 15 % missing level can be a threshold for gap filling in daily precipitation series. A similar result was drawn from RDIs as well at

Lincoln airport weather station. Thus, no positive RDIs was obtained when the missing level becomes 16 %. Therefore, we conclude that the missing level of daily precipitation at less than and equal to 15 % is plausible for further hydrological processes regardless of sampling locations (e.g., whole, front, middle, and rear section of the data).

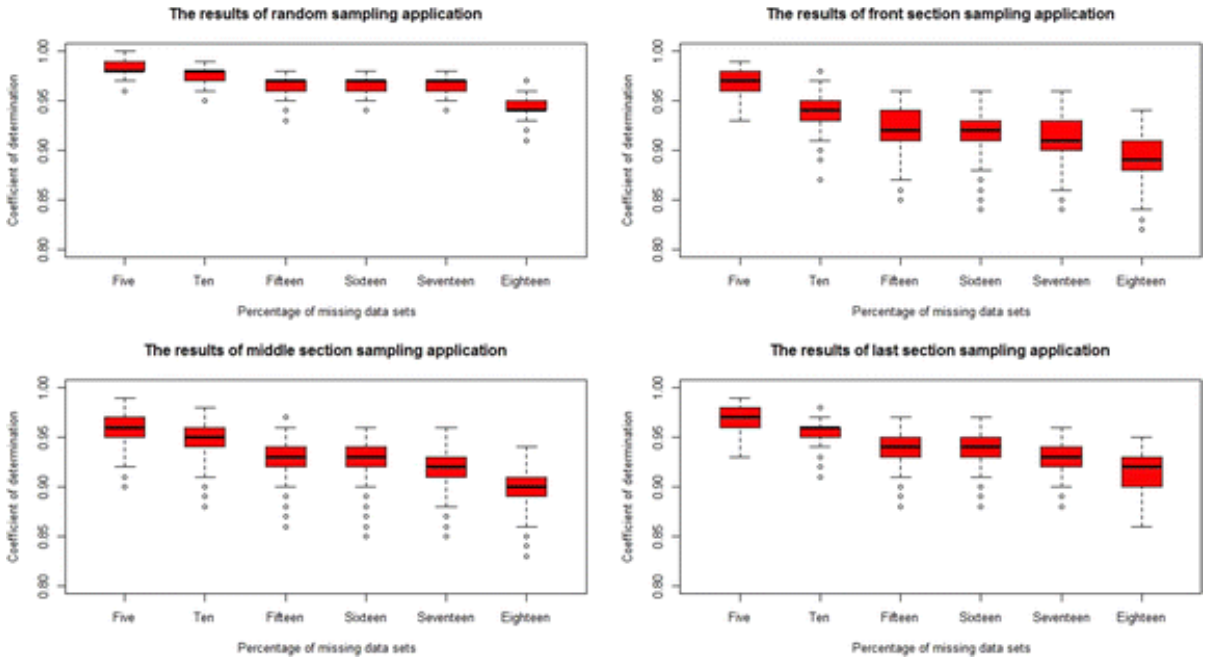


Figure 2.6. The boxplot of R^2 values from 1,000 realizations within four sampling windows (*whole*, *front*, *middle*, and *rear*) at different missing levels (5, 10, 15, 16, 17, and 18 %).

Table 2.1. R^2 values from four sampling windows (whole, front, middle, and rear) at different missing levels (5%, 10%, 15%, 16%, 17%, and 18%).

	Average R^2 values of different missing levels					
	5%	10%	15%	16%	17%	18%
Whole	0.98	0.97	0.96	0.96	0.96	0.94
Front	0.97	0.94	0.92	0.92	0.91	0.90
Middle	0.97	0.95	0.94	0.94	0.93	0.91
Rear	0.96	0.95	0.93	0.93	0.92	0.90

Table 2.2. Total number of RDI and p values from four sampling windows (whole, front, middle, and rear) at different missing levels (5%, 10%, 15%, 16%, 17%, and 18%).

		Number of RDI values (positive, zero, and negative) at different missing levels				
		5-10 %	10-15%	15-16%	16-17%	17-18%
Whole	RDI > 0	651	662	332	308	984
	RDI = 0	341	320	504	528	16
	RDI < 0	8	18	164	164	-
	p value	0.002	0.002	0.241	0.293	-
Front	RDI > 0	1,000	998	457	440	999
	RDI = 0	-	2	543	560	1
	RDI < 0	-	-	-	-	-
	p value	0	0	0.020	0.024	0
Middle	RDI > 0	938	986	519	549	1,000
	RDI = 0	62	14	481	451	-
	RDI < 0	-	-	-	-	-
	p value	0	0	0.010	0.007	0
Rear	RDI > 0	924	995	186	323	1,000
	RDI = 0	76	5	814	677	-
	RDI < 0	-	-	-	-	-
	p value	0	0	0.173	0.067	0

CHAPTER 3. A HEURISTIC GAP FILLING METHOD FOR DAILY PRECIPITATION SERIES

"A heuristic gap filling method for daily precipitation series." *Water Resources Management*, vol.30, issue 7, 2016, pp 2275-2294.

3.1. Introduction

Precipitation is a critical environmental forcing element for hydrologic modeling, agricultural operation and management, and ecosystem modeling. Generally speaking, the performance of hydrologic models relies on the observed environmental data series, such as daily precipitation, yet incomplete precipitation data are prevalent and inevitable due to erroneous reality associated with the systematic errors (e.g. the deformation of the wind field) or random errors (e.g. clogging, electronical malfunction) at the weather station especially in complex terrain. Therefore, missing data are often filled with adequate values through good estimation processes for further environmental modeling and applications.

Historically, many hydrologists have utilized statistical and interpolation methods to estimate the missing values in precipitation records. For example, Gauge Mean Estimator (GME)(McCuen, 1998), Inverse Distance Weighting (IDW) (Lu and Wong, 2008; Teegavarapu and Chandramouli, 2005), and Ordinary Kriging (OK) (Adhikary et al., 2015; Mair and Fares, 2011) are commonly used for gap filling since they are simple and easy to apply. However, these contemporary gap filling methods have a limitation in the sense that they tend to underestimate precipitation magnitudes and to overestimate the number of rainy days in general. Also, the existence of negative autocorrelation between the nearest weather stations is reported at several applications (Garcia et al., 2006; Simolo et al., 2010). Over the

several decades, various gap filling methods have been developed and modified to improve the estimation accuracy and to overcome the limitation of the existing gap filling techniques. For example, Xia et al. (1999) used thin-plate splines, closest station method and multiple linear regression techniques for climatological data estimation. Later, a novel estimation method for missing precipitation, such as artificial neural networks (ANNs) (Teegavarapu and Chandramouli, 2005; Teegavarapu, 2007) and optimal function forms using genetic algorithms with mathematic operators (Teegavarapu et al., 2009) were introduced. These studies indicate that the existing interpolation methods, including GME, IDW, and OK can be improved to develop a better estimation technic for gap filling. For example, as Teegavarapu and Chandramouli (2005) pointed out, the coefficient of correlation weighting (CCWM) method can improve the performance of IDW method by employing gauge-to-gauge correlation relationship, thereby Westerberg et al (2010) utilized CCWM method to estimate the missing precipitation data. Later, Teegavarapu (2012) has applied optimization methods to select rain gauges in the objective matter and delineated optimal clusters using quadrant-based method. Grid-based interpolation method of precipitation using geospatial information was also presented by Teegavaraput et al. (2012).

More recently, by improving the existing interpolation methods, Teegavarapu (2014a) proposed the bias-correction method similar to quantile mapping, while a variant of equi-distant quantile matching and a new optimal single nest estimator scheme are combined. Here, the bias correction method is promising in the sense that it can provide better serially complete precipitation time series than the uncorrected missing estimation method being used. Teegavarapu (2014b) also reported that optimal proximity-based imputation, K-nearest

neighbor classification and K-means cluster methods are plausible to execute gap filling tasks for daily precipitation records.

Although several improvement or new gap filling methods have been developed and proposed, it is still inconclusive to determine the best suitable technique for the precipitation estimation in the field because the estimation performance is depending on geophysical characteristics of the weather station especially in complex terrain. Thus, the geographical conditions, the density of weather stations, and the precipitation pattern from the nearest weather stations as Source Stations (SSs) often dominate the key parameter space during the estimation processes. Additionally, the existing gap filling methods do not necessarily to eliminate the negative autocorrelation of seasonal precipitation patterns embedded in both Source Stations (SSs) and Target Stations (TSs). Thus, a seasonal variability between SS and TS, especially during the summer and winter season is inevitable due to geophysical characteristics associated with distance and altitude between weather stations. Therefore, the combined method proposed here would be a tangible practice in hydrology to identify effective gap filling techniques associated with geophysical characteristics for daily precipitation series.

We propose a method termed "**g**amma distribution function with **s**tatistical **c**orrelation (GSC) coupled with **c**luster **a**nalysis (CA) (hereafter, GSCCA)" to improve the estimation of missing daily precipitation data. Basically, the cluster analysis (CA) represents effectiveness of data summarization and compression, and it divides nearest neighbors efficiently from the large amount of data series encompassing precipitation patterns and geophysical characteristics, such as elevation, latitude, and longitude. The CA first delineates respective clusters and GSC is then used as an estimation tool for the daily precipitation

within the designated clusters by minimizing errors possibly stemmed from the other clusters, which is a difficulty of the existing estimation methods (e.g., GME, IDW, OK).

As Hasan and Croke (2013) showed, the Poisson-gamma method might be a feasible solution to generate the likelihood of rainy and no-rain days using the observed precipitation data. Simolo et al. (2010) also advocated that the gamma-based method is relatively accurate to determine rainfall occurrence and to improve the reconstruction of intense precipitation events. Even though, the estimation accuracy for GSC may be influenced geographical factors and weather station network density associated with TS. It may underestimate for the extremely large precipitation event when the neighboring stations are not close to TS. We, therefore, utilized the gamma distribution function as a statistical framework and CA is employed to act as the screening process for grouping weather stations based on geophysical information (e.g., longitude and latitude). The performance measures with and without CA are then reported and compared to verify how the CA can contribute to the gap filling processes for daily precipitation series in the study area.

3.2. Study area

The state of Idaho is selected as the study area, where more than 150 weather stations administered by National Climate Data Center (NCDC) are situated. For the gap filling process, the weather stations having less than 15% missing level are considered and selected suggested by **chapter 2**. As a subset, 116 stations, therefore, are selected for this study (**Figure 3.1**).

Idaho landscapes are very diverse. The highest elevation is Borah Peak at 3,860 m in the Lost River Range north of Mackay and the lowest elevation is about 216 m in Lewiston.

The climate in Idaho is variable as well. The western region of Idaho located about 420 km from the Pacific Ocean is affected by the maritime climate in humid and wet winter so that the temperature of the northern state is not too cold as much as expected. In the eastern Idaho, by contrast, the seasonal temperature is more extreme along with higher precipitation variability dominated by semi-arid climate, which is reverse pattern. In general, hot summer day is moderated by low relative humidity and cool evening across the state due to the significant diurnal temperature difference. Four major rivers, including Snake, Clark Fork, Clearwater, and Salmon River provide water resources across the state. Due to the distinct seasonal climatic difference and water resources concerns during the growing season, hydrological processes with complete data series without missing values are of great interest to advance water resources research activities, including but not limited to water quality modeling, agricultural water management, and water-energy-food nexus in the mountain west.

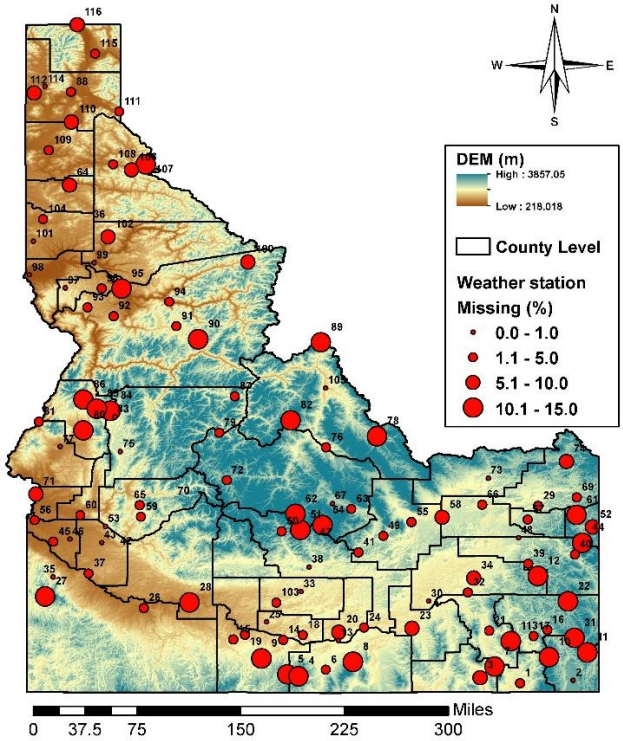


Figure 3.1. Weather stations for daily precipitation with different missing levels in the study area.

3.3. Methodology

Figure 3.2 shows a flowchart of the proposed study. First, the estimation process begins with CA to sort out a regional outlay. To do this, three key elements, such as 1) geographical information (latitude and longitude), 2) elevation (altitude), and 3) seasonal variability (precipitation patterns) are used to define clusters associated with them. Specifically, a total of 116 weather stations are divided into 3 cluster groups based on geographical information (Step 1). Next, elevation information (altitude) is incorporated to delineate two more cluster groups within the pre-defined clusters (Step 2). Once these two steps are completed, the CA further differentiates the clusters by seasonal precipitation patterns (Spring: March - May, Summer: June – August, Fall: September – November, Winter: December - February)(Step 3). Therefore, total 12 cluster groups can be created using the systematic CA. Thus, 3 clusters are created based on geographical information (latitude and longitude) and then each cluster is separated by elevation (altitude) followed by 2 more clusters determined by precipitation patterns associated with four seasons; 12 clusters = (3 clusters from Step 1 x 2 clusters from Step 2) x 2 clusters from Step3. The main reason to apply geophysical information first is not only because that precipitation patterns in the mountain west is very sensitive to geographical information (e.g., proximity between SSs and TSs) but also it is dominated by orographic effects (e.g., altitude difference between SSs and TSs)(Barry and Chorley, 1987). These consequences are observed across the state in the sense that precipitation patterns in eastern Idaho and western Idaho are significantly different due to the influence of the maritime climate. For example, the average amount of precipitation in eastern Idaho is greater than in the southern Idaho. Peaks on the average seasonal precipitation shows at all higher elevation. Seasonal precipitation distribution in

northern and western Idaho occur the maximum precipitation in winter and the minimum precipitation in summer. In the eastern Idaho, maximum precipitation indicates in summer and minimum precipitation shows in winter. In addition, for the correlation between precipitation and elevation, average amount of precipitation for four seasons is increased at higher elevation. Especially, the precipitation of winter and spring seasons represents distinct increasing from lower elevation to higher elevation. As such, the geophysical characteristic of the weather stations has been more weighted than precipitation patterns when the CA is executed for this particular study. By doing so, the CA can further categorize homogenous regions induced by orographic effects and followed by seasonal precipitation patters to minimize negative autocorrelation between TSs and SSs.

For example, once the date (02/02/2012) is selected to fill gaps at Target Station (TS) in Cluster Group 1 (CG1), the average monthly precipitation for February is calculated at TS using the information available at SSs in CG1. Thus, the monthly precipitation by February at SSs is compared to that at TS to identify the best match SS and TS using positive correlation relationship (e.g. correlation coefficient values for February precipitation were calculated from the SSs and best match SS was selected to have higher correlation coefficient values). Next, the cumulative distribution function (CDF) value is computed for the selected day of the month (e.g. 02/02/2012) at the best matched SS. The calculated CDF value is then applied with the inverse gamma function for the same day of the month (e.g. 02/02/2012) to estimate missing precipitation values at TS. These processes are repeated until to estimate all missing precipitation values at TSs in Cluster Groups. Finally, the estimated results are compared with the other existing gap filling methods (e.g. GME, IDW, OK) based on performance measure criteria and skill score described later. Note that the other existing methods have been

evaluated with and without CA for gap filling processes. The period, January 1, 1982 - May 31, 2013 is used for this study and more detailed information for the selected TSs is summarized in **Table 3.1**.

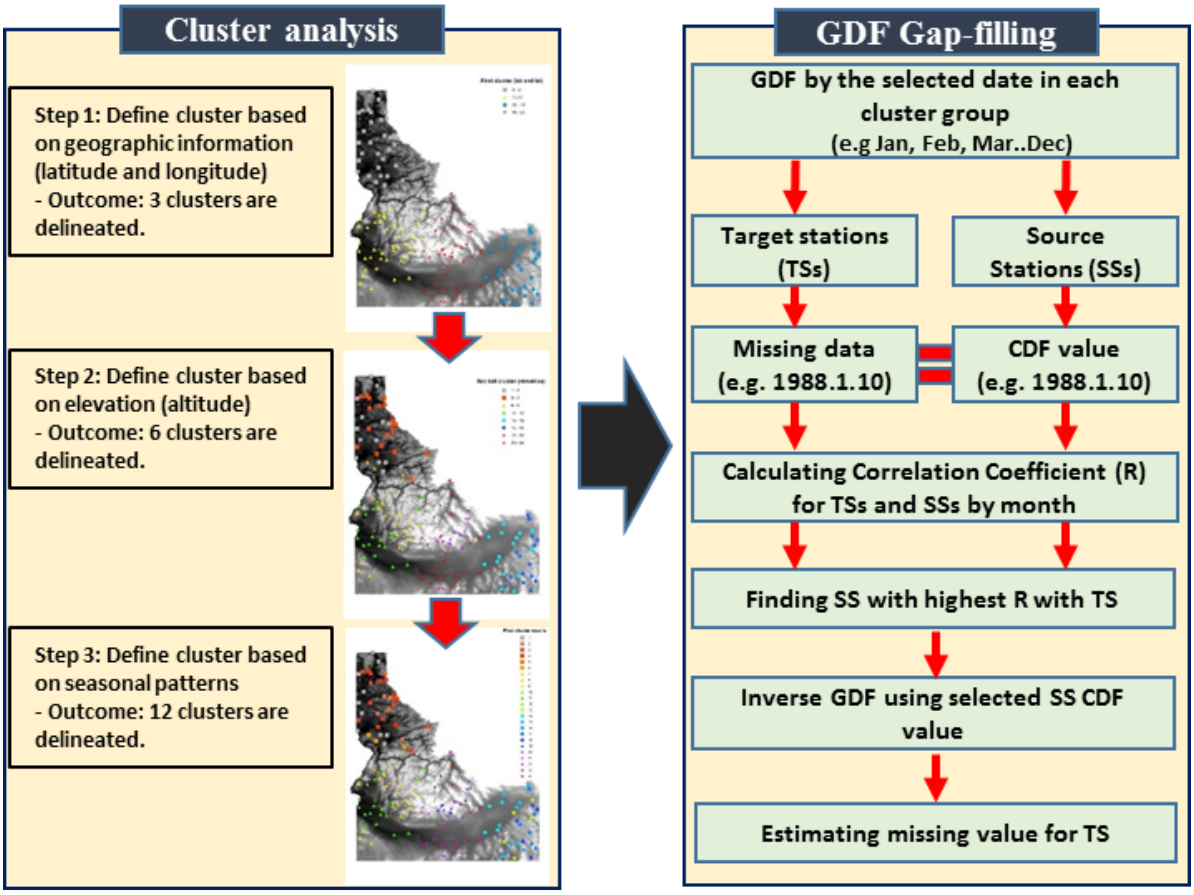


Figure 3.2. A flowchart of the proposed study.

Table 3.1. The selected weather stations with less than 15% missing level of daily precipitation series (January 1, 1982 - May 31 2013).

Station number	Station name	Latitude	Longitude	Elevation (m)	Missing level* (%)
2	LIFTON PUMPING STN	42.1231	-111.3139	1,806	0.25
13	BURLEY FAA AP	42.5417	-113.7661	1,262	0.09
25	JEROME	42.7325	-114.5192	1,140	0.62
30	ABERDEEN EXP STN	42.9536	-112.8253	1,342	0.93
33	RICHFIELD	43.0528	-114.1582	1,305	0.38
35	REYNOLDS	43.2064	-116.7494	1,198	0.56
38	PICABO	43.3111	-114.0742	1,472	0.59
43	BOISE WSFO AP	43.5667	-116.2406	857	0.10
46	NAMPA SUGAR FACTORY	43.6039	-116.5753	753	0.49
48	IDAHO FALLS 16SE	43.6176	-111.8840	1,487	0.41
53	BOISE 7 N	43.7383	-116.2022	1,184	0.26
67	CHILLY BARTON FLAT	43.912772	-113.8289	1,908	0.77
73	DUBOIS EXP STATION	44.2436	-112.2006	1,661	0.27
75	CASCASE 1 NW	44.5228	-116.0481	1,492	0.50
77	CAMBRIDGE	44.5733	-116.6753	807	0.50
83	MC CALL	44.8872	-116.1047	1,531	0.42
88	SANDPOINT KSPT	45.1875	-113.9008	1,198	0.71
97	WINCHESTER 1 ESE	46.2381	-116.6233	1,210	0.52
98	LEWISTON W.B AP	46.3750	-116.9981	442	0.12
99	DWORSHAK FISH HATCHERY	46.5022	-116.3217	303	0.21
101	MOSCOW	46.7281	-116.9558	811	0.32
114	PRIEST RIVER EXP STN	48.3511	-116.8353	723	0.43

* Missing level indicates the actual missing percentage of daily precipitation from January 1, 1982 to May 31, 2013.

3.3.1. Cluster Analysis (CA) method

The cluster analysis (CA) is commonly used to assimilate raw data for further data analysis. For example, it has been widely used in the field of social sciences (Breiger et al., 1975), biology (Eisen et al., 1998; Pan et al., 2002), statistics (Hansen and Jaumard, 1997; Hruschka and Ebecken, 2003), climate investigation (Bunkers et al., 1996; Lund and Li, 2010; Turkes and Tatli, 2011), pattern recognition (Pedrycz, 1990; Gardner, 1991), information retrieval (Maarek et al., 1991; Faloutsos and Oard, 1998), machine learning (Chmielewski and Grzymala-Busse, 1996), and data mining (Chen et al., 1996; Kumar et al., 2012). The CA is a technique to provide useful insights for the analyst to present an overall statistical structure with fewer critical clusters, which is a subset of all of the variable clusters contained in the original data. Thus, a key process of CA is to classify multivariate classes into several cluster groups so that similarity between classes within a group should be large and similarity between groups should be small. In this study, K-means cluster technique (MacQueen, 1967) is first applied to define cluster groups for the selected weather stations based on geographic information (latitude and longitude), elevation (altitude), and seasonal precipitation patterns. The K-means is a simple cluster method to calculate a user-specified number of clusters and to find the mean of a group point using equation (3.1). The minimum sum of the squared error (SEE) defined by equation (3.2) is used to determine the cluster by seeking n-dimensional space repeatedly.

$$\mathrm{SSE} = \sum_{j=1}^{K} \sum_{n \in S_j} | x_n - u_i |^2 \tag{3.1}$$

$$u_i = \frac{1}{m} \sum_{x \in S_j} x \tag{3.2}$$

where, K is disjoint subsets, x_n is a vector representing the nth data point, u_i is the geometric centroid of the data points in S_j, and m is the number of the data points in S_j.

3.3.2. Gauge Mean Estimator (GME)

The Gauge Mean Estimator (GME) is to use an arithmetic average of all SSs. It is a special case of Inverse Distance Weighting (IDW) and it is a similar method being used for the average precipitation estimation method presented by McCuen (1998). The estimation of missing precipitation is given by

$$X_m = \frac{\sum_{i=1}^{i=n} \frac{N_m}{N_i} X_i}{n} \tag{3.3}$$

where, X_m is estimated value at the TS_m, n is the number of stations, and X_i is the observed value at SS_i. N_m and N_i are the average monthly precipitation at TS_m and SS_i, respectively.

3.3.3. Inverse Distance weighting (IDW) method

As proposed by Simanton and Osborn (1980), the Inverse Distance Weighting (IDW) is used to estimate the missing precipitation values and it is implemented by Watson and Philip (1985). IDW has been commonly used to fill the missing values. It is also widely used many water resources applications (ASCE, 1996). For the precipitation estimation at TSs, IDW provides weighting values inversely depending on the distance between TS and SS. The missing values at TSs can be computed by equation (3.4).

$$X_m = \frac{\sum_{i=1}^{n} X_i d_{mi}^{-k}}{\sum_{i=1}^{n} d_{mi}^{-k}} \tag{3.4}$$

where, m is the selected TS, X_m is the estimated value at the TS_m, n is the number of stations, X_i is the observed value at SS I, d_{mi} is the distance from the station i to station m, and k is referred to as friction distance (Vieux, 2001) that ranges from 1.0 to 6.0. In this study, k = 2 is used (Teegavarapu et al., 2009).

3.3.4. Ordinary Kriging (OK) method

The Ordinary Kriging (OK) is the standard approach for surface interpolation method based on scalar measurement at different locations (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989). OK is spatially dependent variance (Vieux, 2001). The degree of spatial dependence in OK method can be determined using a semivariogram. The weights of OK are based on the distance between SS and TS as well as. The equations for OK method to estimate missing values include equation (3.5) – equation (3.7).

$$X_m = \sum_{i=1}^{n} \delta \, X_i \tag{3.5}$$

$$\delta = \tau^{-1}\gamma \tag{3.6}$$

$$\gamma(\mathrm{d}) = \frac{1}{2\,n\,(d)} \sum_{dij} \left(X_i - X_j\right)^2 \tag{3.7}$$

where, δ is the weight obtained from the fitted simivariogram, τ is the gamm matrix, which is the model semivariance for all sampled pairs, $\gamma(\mathrm{d})$ is the semivariance which is defined over observations $X_i$and X_j lagged successively by distance d, and n(d) is the number of distinct pairs in n(d). X_i and X_j are data values at spatial location I and j, respectively.

3.3.5. Gamma distribution function with statistical correlation (GSC)

The standard gamma distribution function (GDF) is known as a proper theoretical distribution function to represent precipitation series in the sense that no negative value of precipitation can be plotted. GDF can emulate actual rainfall distribution for extreme cases although the typical rainfall may not be large. It provides some degree of flexibility in the shape of the distribution function from data ranges (e.g., exponential-decay forms for shape value), so this flexibility allows fitting any number of rainfall regimes to GDF. The accuracy of the estimated GDF can be measured by comparing the cumulative distribution function of the estimated gamma and empirical distributions (Wilks, 1995). In this study, to estimate daily missing precipitation, every month precipitation for the selected day and month between TS and SSs are investigated to apply GDF and compared to select the best SS using the correlation coefficient (R). This process is automatically processed to select the best SS for every month and to fill the missing precipitation from January to December in each cluster group. The statistical correlation is finally applied to improve the estimation at TSs for the specific month. Inverse GDF is then used to estimate missing values using equation (3.8) – equation (3.12)**.**

$$F(x_m \mid \alpha, \beta) = \int_0^{x_m} f(x_m \mid \alpha, \beta) \tag{3.8}$$

$$f(x_m \mid \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} {x_m}^{\alpha-1} e^{-x_m/\beta} \; ; x_m \geq 0 \tag{3.9}$$

$$\Gamma(\alpha) = \int_0^{\infty} {x_m}^{\alpha-1} e^{-x_m} dx \tag{3.10}$$

$$\alpha = (\frac{\overline{x_m}}{s})^2 \; , \qquad \beta = \frac{{S_m}^2}{\overline{x_m}} \tag{3.11}$$

$$x_m = F^{-1}(f(x_m \mid \alpha, \beta) \mid \alpha, \beta) \times \frac{a_{m,t}}{a_{m,s}} \quad (3.12)$$

where, F(x) is CDF of gamma function, f(x) is the probability density function of the gamma distribution, m is the specific month (e.x 1-12), a_t is mean precipitation of total m month at TS, a_s is mean precipitation of total m month at SS. x is daily precipitation at selected day and month, $\bar{x}$ is the mean daily precipitation of m months, and s is the standard deviation of daily precipitation from m month. α is the shape parameter, β is the scale parameter, and Γ is the gamma function.

3.3.6. Performance measures criteria

Error measures such as Squared Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (R) are widely used for the performance comparison of gap filling methods. RMSE calculates a measure difference between simulated and observed values. Theses individual differences are residuals. RMSE provides to aggregate them into a single measure of predictive power. MAE is a quantity measurement to see how close simulated values are to the eventual results. Several studies (Ahrens, 2006; Chang, 2009; Davis, 2002; Kanevski and Maignan, 2004; Lloyd, 2005) have indicated that RMSE and MAE are the best indicators to compare interpolation methods for data estimation. R represents the strength and direction of a linear relationship between the simulated and observed values. R=1 indicates a perfect linear relationship between observed and simulated results, while R=0 represents that there is no linear relationship between the variables. RMSE, MAE, and R are calculated using equation (3.13) – equation (3.15).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_{Si} - P_{Qi})^2}{N}} \tag{3.13}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left| P_{Si} - P_{Qi}\right| \tag{3.14}$$

$$R = \frac{\frac{1}{N}\times\sum_{i=1}^{N}\left(P_{Qi} - \bar{P}_{Qi}\right)\times\left(P_{Si} - \bar{P}_{Si}\right)}{\sqrt{\frac{N\times\sum_{i=1}^{N}P_{Qi}^2 - \left(\sum_{i=1}^{N}P_{Q1}\right)^2}{N\times(N-1)}} \times \sqrt{\frac{N\times\sum_{i=1}^{N}P_{Si}^2 - (\sum_{i=1}^{N}P_{S1})^2}{N\times(N-1)}}} \tag{3.15}$$

where, P_{Qi} is the observed precipitation data at time step I, P_{Si} is the estimated precipitation data, $\bar{P}_{Qi}$ and $\bar{P}_{Si}$ are the mean of the observed and estimated precipitation data, respectively. N is the number of sample sizes.

3.3.7. Performance measures

Performance measures using Heidke Skill Score (HSS), Peirce Skill Score (PSS), and Critical Success Index (CSI) are conducted to verify how the estimated values represent precipitation realization in daily time step (Wilks, 1995). HSS is calculated based on counts from contingency table representing rain and no rain conditions, while CSI evaluates how the estimated rain events represent the observed rain events. PSS indicates the accuracy of the estimated data by a random process. It assumes that both the estimated and observed rain events have the same marginal distribution during data normalization processes. The equations of three skill scores are given by:

$$HSS = \frac{2\times(C_{00}\times C_{11} - C_{01}\times C_{10})}{(C_{00} + C_{10})\times(C_{10} + C_{11}) + (C_{00} + C_{01})\times(C_{01} + C_{11})} \tag{3.16}$$

$$CSI = \frac{C_{11}}{C_{11} + C_{10} + C_{01}} \tag{3.17}$$

$$PSS = \frac{C_{00} \times C_{11} - C_{01} \times C_{10}}{(C_{00} + C_{01}) \times (C_{10} + C_{11})} \quad (3.18)$$

where, C_{00} is the number of no rain events for same day of simulated and observed data, C_{01} is the number of no rain event of observed data, but rain event of simulated data for the same day, C_{10} is the number of rain event of observed data, but no rain event of simulated data for same day, and C_{11} is the number of rain event of simulated and observed data for same day. Positive skill score values mean that they are considered to indicate a minimal level of acceptable performance for the simulated results.

3.4. Results

The results of CA application is shown in **Table 3.2** and **Figure 3.3** as they show the classified 12 cluster groups associated with weather stations in tabular and map format, respectively. The Cluster Group 1 (CG1) includes total 8 weather stations (See **Table 3.2**) and the average elevation is 535 m, while the Cluster Group 8 (CG8) includes 4 weather stations and the highest elevation is 2,378 m. Based on seasonal precipitation patterns, the Cluster Group 4 (CG4) is identified and the highest average seasonal precipitation for each season within CG4 is 363.98 mm, 292.10 mm, 225.55, and 435.10 mm for spring, summer, fall, and winter, respectively. The lowest seasonal precipitation is found at the Cluster Group 6 (CG6)-78.99 mm for spring and 87.38 mm for winter and the Cluster Group 12 (CG12)-65.28 mm for summer and 42.16 mm for fall.

Table 3.2. The result of CA in the study area.

luster group	Gauge station number	Average precipitation (mm)				Average Elevation (m)
		Spring	Summer	Fall	Winter	
1	95, 98, 99, 109, 110, 112, 115, 116	133.86	117.35	83.06	180.85	535
2	91, 92, 93, 96, 97, 101, 104	171.70	168.66	105.92	175.77	1,042
3	94, 100, 102, 103, 105, 106, 108, 111, 113, 114	243.08	184.91	145.29	355.09	766
4	90, 107	363.98	292.10	225.55	435.10	1,617
5	2, 17, 22, 39, 40, 41, 49, 52, 61, 63, 69, 73, 78	90.68	109.47	84.84	96.01	1,748
6	1, 3, 6, 13, 20, 21, 23, 24, 30, 32, 34, 48, 55, 57, 58, 64, 66	78.99	89.92	62.99	87.38	1,396
7	7, 11, 16, 29, 36, 44, 74	198.88	167.64	129.54	250.70	2,058
8	8, 10, 12, 68	289.81	190.75	142.24	325.88	2,378
9	4, 5, 31, 47, 51, 54, 62, 70, 82, 89	226.31	171.70	132.08	267.72	2,189
10	59, 65, 75, 77, 80, 83, 84, 85, 86	206.25	130.81	92.71	270.26	1,310
11	9, 33, 38, 50, 67, 72, 76, 79	81.53	80.77	55.37	96.77	1,584
12	14, 15, 18, 19, 25, 26, 27, 28, 35, 37, 42, 43, 46, 53, 56, 60, 71, 81, 87, 88	85.60	65.28	42.16	98.04	951

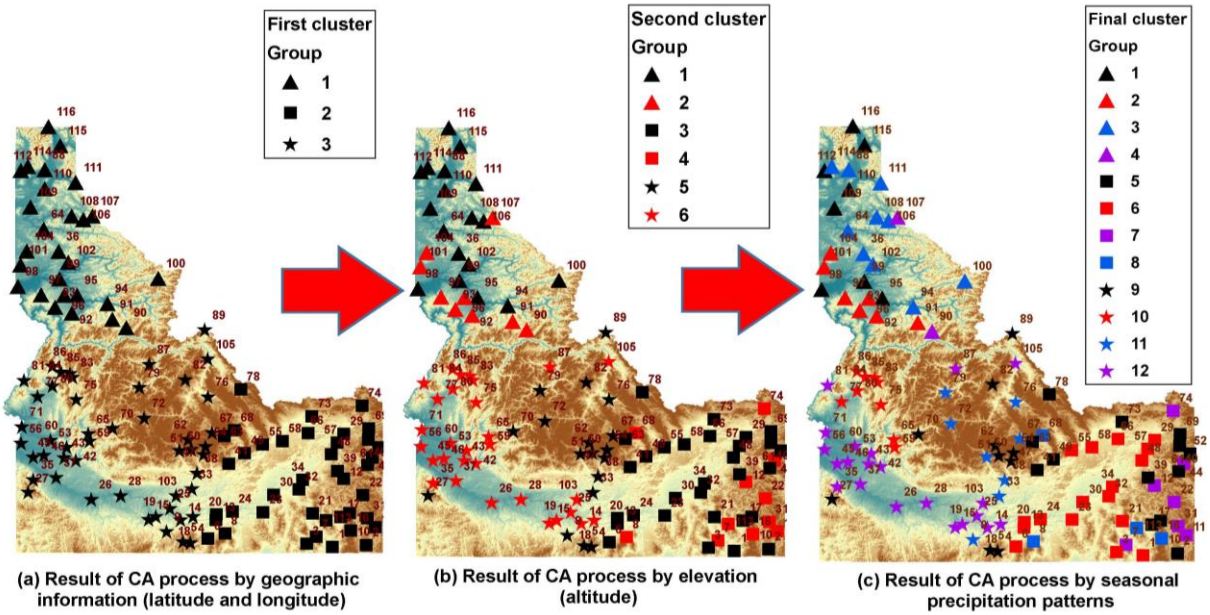


Figure 3.3. The result of step-by-step CA application.

Table 3.3 shows how the gap filling method performs to estimate missing values at TS based on the evaluation criteria, such as R, RMSE, and MAE. The highest and lowest R value is 0.92 and 0.12 at the station number 88 and 75, respectively when GSC and OK method are used (See **Table 3.3**). When CA is employed for gap filling processes, the R value is higher at most stations except few stations. The R values of GSC, in particular, are higher than that of other methods (GME, IDW, OK) being used. The results of RMSE also show the similar result in the sense that the lowest RMSE values are observed at most stations, when CA is employed, except the stations, 73 and 75. The lowest and highest RMSE values are observed at Station 43 and Station 75 when GSC and IDW method are employed, respectively. The result clearly shows that each gap filling method performs very well when CA method is integrated regardless of the gap filling method used. As shown in **Table 3.3**, the typical ranges of RMSE for gap filling methods are 1.54 - 4.27 for GSC, 1.65 - 4.22 for GME, 1.69 - 4.96 for IDW, and 1.61 - 20.11 for OK. The MAE results further justify that gap filling methods with CA can reduce MAE at most stations. Note that the lowest and highest of MAE values are observed 0.52 and 4.63 at Station 67 and Station 114, respectively when GSC and OK are employed with CA.

Table 3.3. The performance measure of the gap filling methods with and without CA.

Station number	Gap filling method	R		RMSE		MAE	
		No CA	CA	No CA	CA	No CA	CA
2	GSC	0.35	0.40	2.20	2.16	0.75	0.74
	GME	0.30	0.37	2.17	2.12	0.97	0.85
	IDW	0.23	0.33	2.72	2.37	1.51	1.09
	OK	0.09	0.36	9.37	5.34	2.99	2.12
13	GSC	0.40	0.55	2.16	1.85	0.71	0.64
	GME	0.55	0.57	1.78	1.74	0.72	0.67
	IDW	0.53	0.55	2.02	1.79	1.14	0.76
	OK	0.38	0.32	3.35	3.62	1.44	1.10
25	GSC	0.57	0.58	1.92	1.90	0.61	0.60
	GME	0.58	0.58	1.85	1.83	0.74	0.70
	IDW	0.48	0.55	2.18	1.87	1.19	0.81
	OK	0.23	0.45	11.37	2.49	4.42	0.93
30	GSC	0.65	0.68	1.70	1.64	0.54	0.54
	GME	0.50	0.59	1.94	1.81	0.78	0.66
	IDW	0.42	0.54	2.34	1.96	1.33	0.77
	OK	0.68	0.67	1.76	1.78	0.66	0.66
33	GSC	0.63	0.70	1.95	1.79	0.60	0.56
	GME	0.63	0.68	2.02	1.85	0.77	0.66
	IDW	0.53	0.55	2.23	2.09	1.20	0.77
	OK	0.20	0.46	7.41	2.55	2.58	0.88
35	GSC	0.36	0.60	1.98	1.64	0.68	0.61
	GME	0.53	0.59	1.73	1.65	0.70	0.61
	IDW	0.51	0.57	1.99	1.69	1.11	0.72
	OK	0.20	0.41	7.64	2.21	3.26	0.90
38	GSC	0.48	0.67	3.08	2.40	0.90	0.78
	GME	0.59	0.64	2.71	2.57	1.03	0.87
	IDW	0.50	0.61	2.83	2.63	1.38	0.88
	OK	0.11	0.37	21.22	3.07	5.62	1.14
43	GSC	0.50	0.83	2.16	1.54	0.77	0.52
	GME	0.56	0.59	1.94	1.89	0.84	0.77
	IDW	0.55	0.62	2.14	1.82	1.20	0.81
	OK	0.09	0.80	37.85	1.61	4.79	0.66
46	GSC	0.64	0.72	1.92	1.73	0.62	0.60
	GME	0.62	0.72	2.02	1.82	0.80	0.65
	IDW	0.56	0.56	2.18	2.18	1.22	1.22
	OK	0.39	0.65	6.66	2.00	2.63	0.79
48	GSC	0.44	0.47	3.92	4.27	1.33	1.57
	GME	0.57	0.60	2.98	2.87	1.18	1.07
	IDW	0.51	0.62	3.04	2.89	1.43	1.03
	OK	0.18	0.44	5.69	3.93	2.10	1.65
53	GSC	0.70	0.80	3.32	3.03	1.23	1.05
	GME	0.66	0.70	2.75	2.55	1.19	1.03
	IDW	0.63	0.70	2.77	2.74	1.32	1.06
	OK	0.32	0.80	7.56	2.22	3.28	0.86

Station number	Gap filling method	R		RMSE		MAE	
		No CA	CA	No CA	CA	No CA	CA
67	GSC	0.34	0.56	1.97	1.81	0.54	0.52
	GME	0.46	0.46	1.79	1.80	0.61	0.55
	IDW	0.32	0.36	2.33	2.07	1.27	0.75
	OK	0.11	0.37	34.31	2.31	15.33	0.83
73	GSC	0.58	0.60	2.66	2.20	0.88	0.79
	GME	***0.61***	***0.54***	***2.20***	***2.26***	0.91	0.87
	IDW	***0.47***	***0.28***	***2.45***	***4.08***	1.26	1.03
	OK	0.32	0.46	5.85	3.55	2.36	1.45
75	GSC	0.58	0.62	3.31	3.18	1.35	1.31
	GME	***0.69***	***0.67***	2.95	2.94	1.31	1.20
	IDW	***0.66***	***0.46***	***3.04***	***4.46***	1.38	1.29
	OK	***0.17***	***0.12***	27.24	20.11	5.79	2.47
77	GSC	0.58	0.60	3.85	3.56	1.42	1.33
	GME	0.63	0.69	3.46	3.21	1.47	1.22
	IDW	0.64	0.69	3.45	3.16	1.56	1.27
	OK	0.40	0.49	27.35	4.65	12.16	1.88
83	GSC	0.60	0.67	3.90	3.23	1.63	1.42
	GME	0.65	0.77	3.23	2.76	1.48	1.16
	IDW	0.68	0.77	3.24	2.73	1.49	1.15
	OK	0.56	0.55	5.39	4.13	2.74	1.88
88	GSC	0.60	0.92	8.97	2.71	1.21	0.75
	GME	0.78	0.79	5.37	4.22	0.87	0.92
	IDW	0.34	0.70	6.24	4.96	1.37	1.87
	OK	0.05	0.67	177.43	5.16	36.07	2.20
97	GSC	0.71	0.79	2.83	2.45	1.11	1.04
	GME	0.47	0.78	3.49	2.68	1.59	1.07
	IDW	0.54	0.78	3.36	2.48	1.52	1.10
	OK	0.75	0.76	2.64	2.59	1.07	1.09
98	GSC	0.43	0.49	2.18	2.11	0.82	0.82
	GME	0.40	0.41	2.14	2.15	0.99	0.93
	IDW	0.45	0.38	2.27	2.52	1.26	1.23
	OK	0.22	0.23	20.64	3.80	6.79	1.63
99	GSC	0.36	0.49	5.76	3.75	2.41	1.51
	GME	0.51	0.53	3.23	3.19	1.60	1.47
	IDW	0.54	0.57	3.20	3.10	1.42	1.51
	OK	0.39	0.34	20.32	10.29	8.97	4.21
101	GSC	0.60	0.70	3.68	3.29	1.53	1.40
	GME	0.51	0.70	3.77	3.18	1.78	1.39
	IDW	0.60	0.66	3.63	3.35	1.64	1.45
	OK	0.25	0.31	7.38	6.44	3.49	2.79
114	GSC	0.57	0.62	4.15	3.76	1.65	1.67
	GME	0.36	0.44	4.42	4.13	2.29	2.03
	IDW	0.45	0.47	4.08	4.00	1.96	1.97
	OK	0.42	0.42	6.06	8.91	2.98	4.63

Note: Shaded areas indicate the performance decreases when CA is employed.

Figure 3.4 illustrates the variation of the performance improvement of each gap filling method at TSs based on R, RMASE, and MAE. It clearly shows that CA can improve overall performance when the gap filling methods are implemented. **Table 3.4** summarizes the performance of four different gap filling methods with and without CA method. It is noteworthy that R values increase significantly when CA is integrated in the gap filling methods, while GSC outperforms the rest of the methods. The similar results are also observed when RMSE and MAE are used for evaluation. The OK method, in particular, shows a significant difference between CA and non-CA integration into the gap filling process by showing highest improvement at 72% in MAE.

In general, the performance of the existing methods (GME, IDW, OK) for gap filling processes is often dominated by the closed stations and the density of their positions. In this study, CA is used to minimize collinearity between stations by segregating them by latitude, longitude, and the altitude. By doing so, characteristics of microclimate and orographic effect can be incorporated into gap filling processes and consequently more reasonable estimation for daily precipitation can be achieved. Thus, as weather patterns in eastern Idaho and western Idaho are significantly different due to maritime climate as discussed before, CA approach prior to gap filling process makes more sense to characterize daily precipitation based on 3 dimensional dominance (latitude, longitude, altitude) rather than 2 dimensional plane (distance between stations).

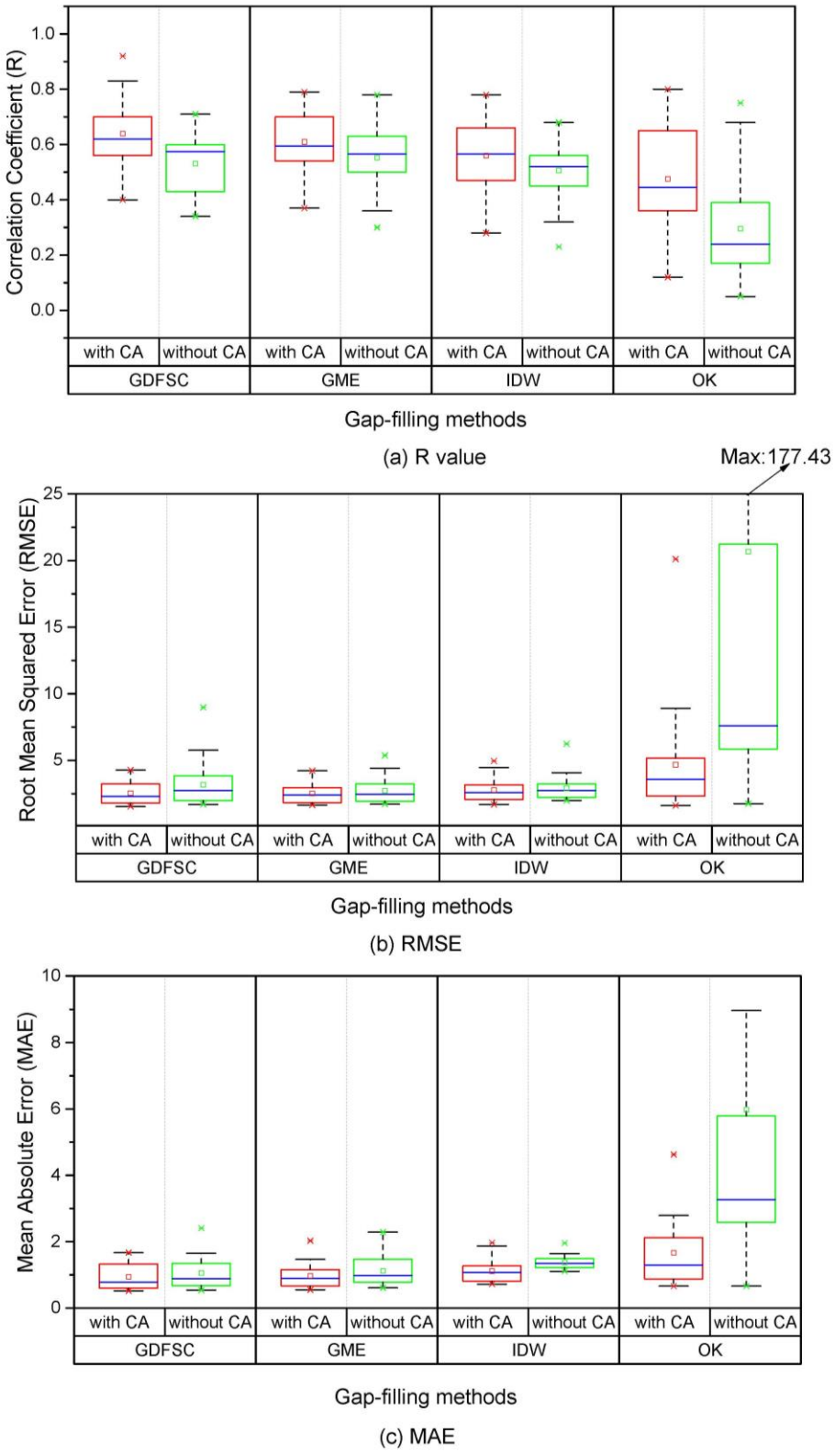


Figure 3.4. The boxplot of performance measures with and without CA for gap filling methods.

Table 3.4. The average performance variations for four gap filling methods with and without CA.

Methods	Mean performance variation								
	R			RMSE			MAE		
	No CA	CA	Increasing (%)	No CA	CA	Decreasing (%)	No CA	CA	Decreasing (%)
GSC	0.53	0.64	17.19	3.16	2.55	19.30	1.06	0.94	11.32
GME	0.55	0.61	9.84	2.72	2.51	7.72	1.12	0.97	13.39
IDW	0.51	0.56	8.93	2.90	2.77	4.48	1.37	1.12	18.25
OK	0.30	0.48	37.50	20.66	6.47	68.68	5.98	1.67	72.07

After evaluating effectiveness of CA for the gap filling methods, skill scores indices are also calculated to evaluate the estimation accuracy for rain and no rain conditions. **Table 3.5** shows skill scores (HSS, CSI, and PSS) of the gap filling methods with CA and **Figure 3.5** illustrates skill scores in the boxplot. As shown in **Table 3.5**, GSC outperforms the other methods at most stations, except few stations, such as Stations 33, 38, 67, 75, and 99. Interestingly, all these stations have less than 1% missing level (See **Figure 3.1**) so that overall skill scores are high, which mean that no additional benefit from gap filling processes are achieved. Nonetheless, GSC outperforms the other methods.

In general, the highest HSS and SCI are recorded when GSC is implemented at most stations except Station 99, where IDW outperforms the other method in terms of the highest HSS and SCI. Additionally, although PSS results using OK method better performs slightly than the other methods based on PSS at few Stations 33, 38, 67, 75, and 77, GSC outperforms the other methods at most stations. The GME method, by contrast, has the lowest average values, 0.32, 0.40, and 0.45 for HSS, CSI, and PSS, respectively in comparison with that of three other gap filling methods. For IDW method, in particular, the skill scores are calculated secondly higher values, but the variation between skill scores is also very high; the average HSS and PSS are 0.28 and 0.43, but the lowest HSS and PSS are 0.02 and 0.04, respectively.

Thus, overall skill scores from GME, IDW, and OK method with CA are similar each other when HSS, CSI, and PSS are computed. Based on the preliminary result along with skill scores, it appears that GSC is a promising method to fill missing values for daily precipitation series especially under rain and no rain conditions. Additionally, higher estimation accuracy is observed when GSC is employed to estimate the missing precipitation data at most stations. Similar results were also found by Hason and Croke (2013) and Simolo et al. (2010). These results confirmed by the analysis of the number of rain and no rain conditions. Gamma distribution method well performs to estimate non-zero rainfall amount than IDW, GME, and OK. However, more improved results were obtained using CA with GSC, IDW, GME, and OK method.

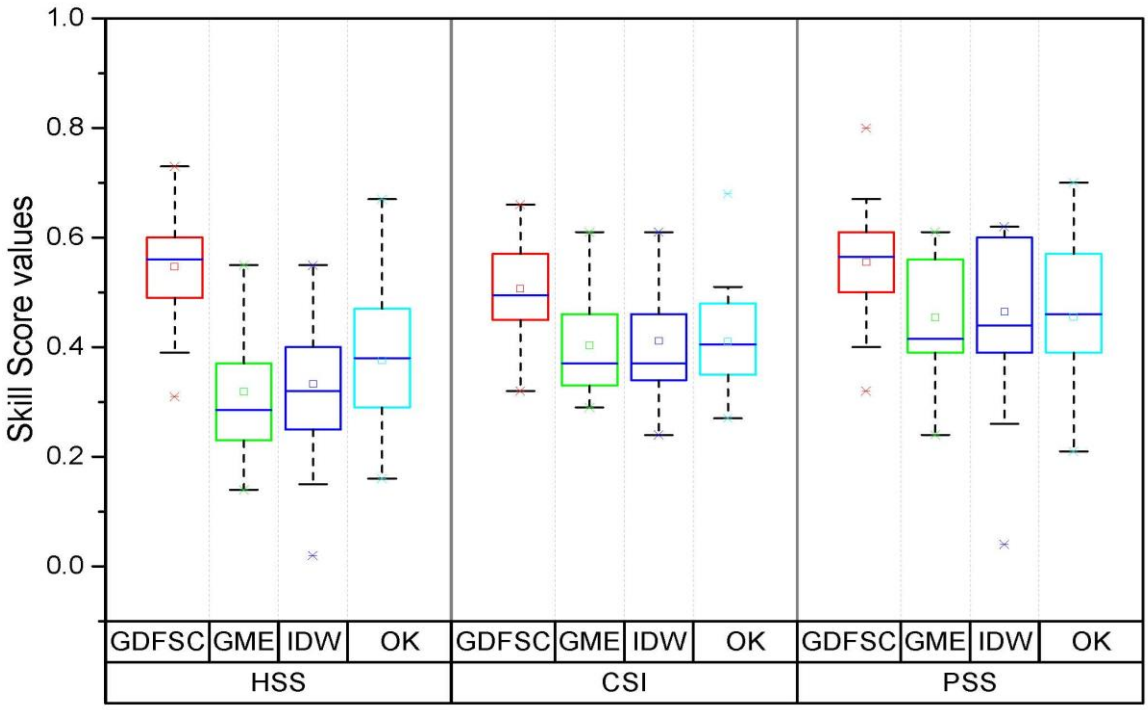


Figure 3.5. The boxplot of skill scores for the gap filling methods.

Table 3.5. Skill scores for the gap filling methods with CA.

Number Station	HSS				CSI				PSS			
	GSC	GME	IDW	OK	GSC	GME	IDW	OK	GSC	GME	IDW	OK
2	0.54	0.25	0.25	0.34	0.49	0.35	0.35	0.36	0.57	0.39	0.39	0.40
13	0.58	0.23	0.27	0.30	0.51	0.33	0.35	0.32	0.57	0.38	0.44	0.33
25	0.58	0.23	0.23	0.35	0.50	0.32	0.32	0.37	0.56	0.39	0.39	0.48
30	0.63	0.20	0.25	0.42	0.56	0.29	0.32	0.42	0.67	0.37	0.44	0.61
33	0.56	0.37	0.38	0.45	0.47	0.37	0.37	0.41	***0.54***	***0.60***	***0.60***	***0.61***
35	0.42	0.23	0.23	0.21	0.37	0.32	0.33	0.29	0.40	0.39	0.39	0.29
38	0.56	0.36	0.36	0.27	0.47	0.35	0.36	0.28	***0.59***	***0.60***	***0.61***	0.39
43	0.73	0.24	0.25	0.47	0.66	0.34	0.34	0.46	0.80	0.41	0.42	0.63
46	0.64	0.21	0.02	0.29	0.57	0.33	0.24	0.35	0.66	0.36	0.04	0.42
48	0.47	0.28	0.35	0.33	0.43	0.39	0.43	0.39	0.44	0.40	0.48	0.39
53	0.67	0.31	0.32	0.47	0.61	0.43	0.43	0.50	0.63	0.43	0.43	0.57
67	0.53	0.36	0.37	0.46	0.46	0.37	0.37	0.42	***0.55***	***0.56***	***0.57***	***0.58***
73	0.49	0.27	0.28	0.30	0.45	0.36	0.36	0.35	0.50	0.42	0.43	0.39
75	0.56	0.48	0.55	0.47	***0.55***	0.54	***0.58***	0.51	***0.56***	***0.56***	***0.62***	0.50
77	0.52	0.34	0.40	0.41	0.49	0.40	0.43	0.42	0.60	0.51	0.56	0.48
83	0.60	0.48	0.53	0.47	0.59	0.55	0.58	0.50	0.61	0.57	0.62	0.49
88	0.31	0.14	0.15	0.16	0.32	0.29	0.29	0.27	0.32	0.24	0.26	0.21
97	0.68	0.55	0.55	0.67	0.66	0.61	0.61	0.68	0.67	0.61	0.61	0.70
98	0.39	0.29	0.32	0.27	0.41	0.40	0.41	0.36	0.43	0.40	0.43	0.33
99	***0.42***	***0.43***	***0.46***	***0.42***	***0.44***	***0.52***	***0.54***	***0.48***	***0.40***	***0.50***	***0.53***	***0.44***
101	0.60	0.48	0.53	0.47	0.59	0.55	0.58	0.50	0.61	0.57	0.62	0.49
114	0.55	0.28	0.28	0.27	0.56	0.46	0.46	0.40	0.55	0.34	0.35	0.29
Mean	0.55	0.32	0.33	0.38	0.51	0.40	0.41	0.41	0.56	0.45	0.47	0.46

Note: Shaded areas indicate that the existing methods outperform GSC.

CHAPTER 4. THRESHOLD OF BASIN DISCRETIZATION LEVELS FOR HSPF SIMULATIONS WITH NEXRAD INPUTS

"Threshold of basin discretization levels for HSPF simulations with NEXRAD inputs."

Journal of Hydrologic Engineering, vol.19, issue 7, 2014, pp 1401-1411.

4.1. Introduction

A rainfall-runoff model is a typical tool to characterize the dynamics of air, soil, and water by means of hydrological simulations, such as streamflow. This type of model plays a critical role in predicting hydrological events (e.g., flood or drought) for effective water management and long-term planning. In terms of model structure, computational difficulty, and basin delineation processes, a hydrologic model can be characterized as a lumped or distributed model. A lumped model can be defined as a hydrologic model that considers the basin as one unit; therefore, the set of various hydrologic parameters is global. Thus, the lumped model is applied to simulate streamflow at the basin scale, in which physical representations of soil, land use, vegetation coverage, and climate forcing (e.g., precipitation and temperature) are averaged and uniformly applied to parameterize hydrological processes (Smith et al., 2004). Over the last few decades, the lumped model has been dominantly used for many water management exercises, such as flood risk analysis and streamflow forecasting (Bell and Moore, 1998; Michaud and Sorooshian, 1994), whereas the distributed model has gradually continued to draw attention over time, owing to improved computer technology and enriched environmental data sets at high spatial and temporal resolution (Koren et al., 1999; Ryu, 2009; Wood, 1995). In recent years, the distributed model has become a popular choice in the field of hydrology because it allows high-resolution data to be incorporated into high-

performance computing and complex geophysical modeling framework (Ryu, 2009). The lumped model can be defined as a hydrologic model that divides the basin into a number of smaller units called catchments; each catchment has its own set of parameters, whereas the semidistributed model is between the lumped model and the distributed model. The ability to simulate the spatial hydrologic variability across the basin further encourages many hydrologists to apply the distributed model for their own applications (Carpenter and Georgakakos, 2004; Smith et al., 1995). Additionally, finer scale geospatial data (e.g., 2–3 m resolution), such as North America Land Data Assimilation System (LDAS) data (Mitchell et al., 2004), are still evolving to characterize the spatial and temporal variability of water and energy cycles, which is critical to improve understanding of the dynamics between land, surface, and atmosphere. Climate change and variability is another contributing factor that facilitates applications of hydrological models in a distributed fashion because local-specific spatial variability of hydrological consequences induced by uncertain future climates is of great interest in many basins around the world. In January 2000, the Hydrology Laboratory (HL) of the National Oceanic and Atmospheric Administration/National Weather Service (NOAA/NWS) initiated the Distributed Model Intercomparision Project (DMIP) to improve river and streamflow forecasts (Smith et al., 2004). A fundamental element in this effort is to incorporate next generation radar (NEXRAD) multisensory precipitation products into hydrological simulation and forecasting frameworks. Smith et al. (2004) and related publications provide additional discussions about a broader spectrum of model development and comparisons to improve streamflow forecasting (Lerat et al., 2012; Smith et al., 2012). Although many research studies discuss the comparisons between distributed models and lumped models during the course of DMIP, few have provided practical guidelines to

determine how basin discretization can improve the overall model performance. The term of discretization here is analogous to the level of basin delineation that can be made from basin to catchment scales with respect to spatial resolution and parameter sets. Identifying such information is challenging in the sense that a distributed model does not necessarily outperform a lumped model. For example, prior to DMIP, Refsgaard and Knudsen (1996) applied three different models, including a lumped conceptual modeling system [Nedbor-Afstromings Model (NAM)], a distributed, physically based system [An integrated catchment modeling by DHI Software (MIKE SHE)], and an intermediate approach [An integrated Water Balance Model (WATBAL)], to three catchments in Zimbabwe ranging from 254 to 1,090 km^2. However, no noticeable performance among these models was observed, although the distributed model performed marginally better for some cases in which no calibration was allowed (Refsgaard and Knudsen, 1996). Boyle et al. (2001) reported performance improvements associated with the application of distributed models, moving from a lumped model to a semidistributed model, but they indicated that no additional gain could be achieved at the selected catchments in the Blue River basin (1,227 km^2). It is possible that their experiment for a fairly similar catchment size, approximately 157 km^2 on average, did not necessarily justify the benefits of spatial details on different catchment sizes. Later, Ajami et al. (2004) conducted an experiment with a slightly different focus to explore the spatial details and complexity that can improve the accuracy of streamflow forecasts during DMIP periods. Specifically, they explored how different combinations of the spatial complexity of input, model structure, and parameters can contribute to improving overall calibration results. They concluded that only marginal effects on the calibration results were attained, even when the spatial complexity was increased by distributing the parameters along with precipitation and

moving from a lumped to a distributed calibration strategy (Ajami et al., 2004). Although many previous studies indicate that the added complexity when transitioning from a lumped to a distributed representation of a basin results in performance improvement after calibration at the level of current operational standards, little research focuses on how spatially downsized catchment scales can contribute to improving overall modeling performances at a certain threshold of basic discretization. Thus, no operational guideline is available for hydrologists to determine the threshold of basin discretization that represents a proper catchment delineation level for their own applications. As such, the goal of **chapter 4** is to provide useful insights for the hydrologic community to support better decisions by determining the proper catchment size for optimal basin scale modeling using HSPF. Having reasonable criteria for basin discretization is an important element of the modeling process, to reflect some aspect of the prediction accuracy of the calibrated model. A threshold suggested by this chapter will be also useful during basin delineation processes because the model performance is constrained by the catchment size, thereby resulting in cost and time savings.

4.2. Study area

Two river basins, the Illinois River basin and the Elk River basin, are selected for this study. As shown in **Figure 4.1**, the Illinois River basin lies to the south of the Elk River basin: the upstream end is located at the USGS gauge station at Savoy, Arkansas (AR) and the downstream end at the USGS gauge at Tahlequah, Oklahoma (OK). The area drained by the river reach between the two stations totals 2,025 km^2 and the primary stem length of the Illinois River is approximately 95 km. The catchment sizes of the Illinois River used in this study vary from 37 km^2 at Springtown to 2,484 km^2, which is the total drainage area of the

Illinois River Basin at Tahlequah (**Figure 4.1** and **Table 4.1**). The drainage areas of the Baron Fork and Siloam Springs catchments, part of the Illinois River basin, are 795 and 1,489 km^2, respectively. The Elk River basin lies to the north of Illinois River and drains to the USGS gage at Tiff City, Missouri (MO). The drainage area and mainstream length of this basin are 2,258 and 56.6 km, respectively. The largest and smallest catchment sizes used in this study are 2,258 km^2 at Tahlequah and 37 km^2 at Springtown. Both basins are selected for DMIP not only because high-quality NEXRAD precipitation data are available but also because there are no complicated water activities, such as upstream diversion, dam operation, or snow, that can compromise natural streamflow (Ryu, 2009). Data from 15 USGS stream gauge stations listed in **Table 4.1** are compiled and used for hydrologic simulations. Both the Elk and Illinois River basins have several streamflow gauge stations, including outlets and interior points to be used for model calibration.

Table 4.1. USGS stream gauges in the DMIP 2.

No	USGS No	Name	DMIP-2 ID	Latitude (^{o})	Longitude (^{o})	Area (km^2)
1	7196500	Illinois River near Tahlequah OK	talo2	35.923	-94.924	2,484
2	7197000	Baron Fork at Eldon OK	eldo2	35.921	-94.839	795
3	7196973	Peacheater Creek at Christie OK	peach	35.955	-94.696	65
4	7196000	Flint Creek near Kansas OK	knso2	36.186	-94.707	285
5	7195500	Illinois River near Watts OK	wtto2	36.130	-94.572	1,645
6	7194800	Illinois River at Savoy AR	savoy	36.103	-94.344	433
7	7189000	Elk river near Tiff City Mo	tifm7	36.631	-94.587	2,258
8	7188653	Big Sugar Creek near Powell MO	powel	36.616	-94.182	365
9	7188885	Indian Creek near Lanagan MO	lanag	36.599	-94.450	619
10	7194880	Osage Creek near Cave Springs AR	caves	36.281	-94.228	90
11	7195000	Osage Creek near Elm Springs AR	elmsp	36.222	-94.289	337
12	7195430	Illinois river South of Siloam Springs AR	sloa4	36.109	-94.534	1,489
13	7195800	Flint Creek at Springtown AR	sprin	36.256	-94.434	37
14	7195865	Sager Creek near West Siloam Springs Ok	wsilo	36.202	-94.605	49
15	7196900	Baron Fork at Dutch Mills AR	dutch	35.880	-94.487	105

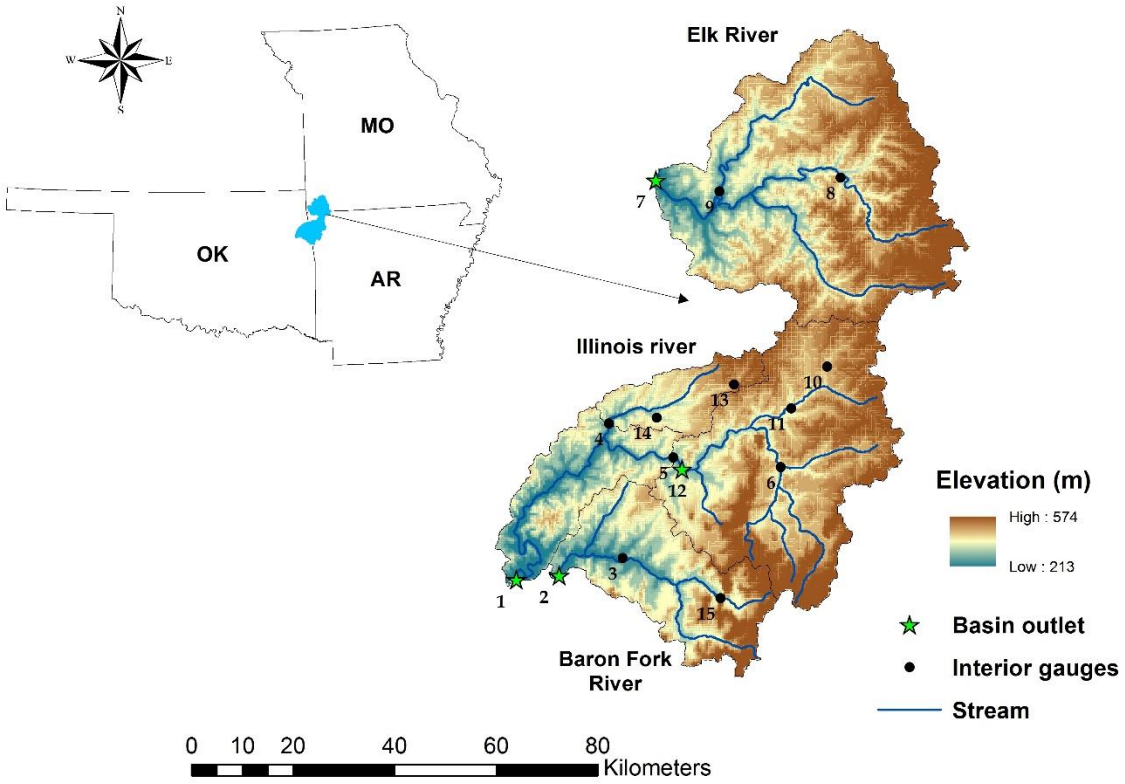


Figure 4.1. Stdy area of DMIP 2 and gage points.

4.3. Methodology

4.3.1. Input data

The high-quality NEXRAD rainfall estimates provided by the National Weather Service (NWS) were used in this study. The NEXRAD precipitation data have a Hydrologic Rainfall Analysis Project (HRAP) grid of 4 by 4 km resolution and are stored in binary file format (e.g., xmrg format) at polar stereographic projection (Greene and Hudlow, 1982). The literature (Fulton et al., 1998; Reed and Maidment, 1999; Seo and Breidenbach, 2002; Wang et al., 2000; Zhang et al., 2011) provides additional information regarding NEXRAD implementation, application, product processing, and algorithms. Although the original cell size (4 by 4 km spatial resolution) of NEXRAD precipitation was considered for application

to this hydrologic modeling framework, 8 by 8 km spatial resolution was chosen because additional climate forcing data sets, including wind speed, temperature, and dew point available from North American Regional Reanalysis (NARR) data in 32 by 32 km grid cells, were needed to compute potential evapotranspiration (PET). The NEXRAD rainfall data in 8 by 8 km resolution were averaged in each catchment by using mean area precipitation (MAP) and applied to HSPF as an input. The data extraction techniques for NEXRAD rainfall estimates and NARR data set assimilation are described in the work of Ryu (2009).

It is hypothesized that a hydrologic model with finer spatial resolution may provide performance gains over such a model in lumped fashion at certain levels. The specific requirement for this experiment includes that the model should have reliable and objective procedures for parameterization, using all climate forcing (e.g., NEXRAD rainfall estimates, PET) and other hydrological data (e.g., basin slope, soil, land cover) in gridded format. To incorporate PET into hydrological simulations using HSPF, Jensen's method (Jensen and Haise, 1963) is typically used, but the Penman-Monteith method (Monteith, 1965) was selected for the current application for the following reasons: the Penman-Monteith equation expresses the evaporating surface as a single "big leaf" (Raupach and Finnigan, 1988) with the atmospheric physics, influenced by the crop canopy architecture, and single-surface resistance. Furthermore, PET using the Penman-Monteith equation has shown better performance results than other PET methods (Chiew et al., 1995; Jensen et al., 1990) and is also recommended by the Food and Agricultural Organization of the United Nations (FAO) (Allen et al. 1998) as a standard method for modeling PET. The Penman-Monteith PET (mm=day) is defined as

$$PET = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{4.1}$$

where Rn is net radiation at the crop surface (MJ=m^2=day); G is soil heat flux density (MJ=m^2=day); T is mean daily air temperature (°C); u_2 is wind speed at 2 m height (m=s); es is saturation vapor pressure (kPa); ea is actual vapor pressure (kPa); Δ is slope vapor pressure curve (kPa=°C); γ is psychrometric constant (kPa=°C).

For PET computation, climate forcing data, including wind speed, dew point, and temperature, were first retrieved from the NARR data depository, and then solar radiation data were derived from equations by using temperature differences provided by the literature (Allen et al., 1998). For spatial adjustment between NEXRAD and NARR grid sizes, a built-in spatial interpolation scheme in the Grid Analysis and Display System (GrADS) software was used. Finally, all required data sets were compiled into watershed data management (WDM) to leverage the input and output processes of HSPF. For topographic relief mapping, basin delineation, and flow direction computation, a digital elevation model (DEM) provided by NWS was used. This DEM provides 15 arc-second (0.004166°, resolution approximately 90 m) intervals, which are resampled from resolution of 1 arc-second DEM (7.5 min, 30 m resolution). The stream network for hydrologic simulations was generated by using National Hydrography Dataset (NHD) via a basin delineation process. For the land use data set, this study used the National Land Cover Dataset (NLCD), which has 76 land cover classifications, but were further reclassified to represent discernible land use characteristics within HSPF modeling blocks. Five land use classifications, including urban or built-up land, agricultural land, forest land, water, and barren land, were incorporated into the model to parameterize the hydrological processes associated with infiltration rate, interflow, and groundwater percolation. Soil classification map and land use information for the study area are shown in **Figure 4.2** and **Table 4.2**, respectively.

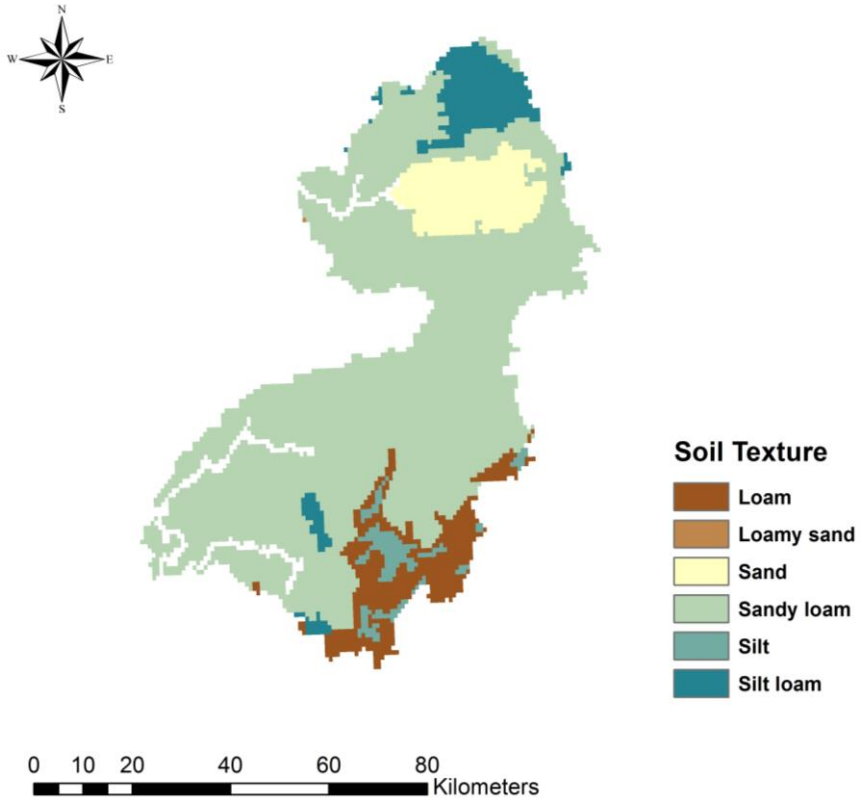


Figure 4.2. Soil texture classification for the study area.

Table 4.2. Land Use classification (km^2) for DMIP2 basins.

Land Use	Elk River	Baron Fork River	Illinois River-Tahlequah	Illinois river-Siloam spring
Total area (km^2)	2,258 (100 %)	795 (100 %)	2,484 (100 %)	1,489 (100 %)
Urban or built-up Land (km^2)	43.13 (1.91 %)	10.89 (1.37 %)	114.26 (4.60 %)	90.08 (6.05 %)
Agricultural Land (km^2)	1,041.39 (46.12 %)	368.40 (46.34 %)	1,488.66 (59.93 %)	997.03 (66.96 %)
Forest (km^2)	1,161.29 (51.43 %)	415.71 (52.29 %)	896.97 (36.11 %)	396.67 (26.64 %)
Barren land (km^2)	10.16 (0.45 %)	-	1.99 (0.08 %)	1.34 (0.09 %)
Water (km^2)	2.03 (0.09 %)	-	6.96 (0.28 %)	3.87 (0.26 %)

4.3.2. Hydrologic Simulation Program-Fortran (HSPF)

HSPF is a process-based, river basin-scale, and semi-distributed model. This model is developed to forecast the impact of land management or climate change on streamflow and water quality in the large and complex watershed. HSPF is integrated model including climatology, hydrologic soil, land use, nutrients, bacteria, and pathogens. The main function of HSPF model is derived from following four process; (1) Hydrocomp Simulation Programming(HSP) (Hydrocomp, Inc., 1976), (2) Nonpoint Source(NPS) Model (Donigian and Crawford, 1976), (3) Agricultural Runoff Management (ARM) model (Donigian, 1977), (4) Sediment and Radionuclides Transport (SERATRA) (Onishi and Wise, 1982). HSPF model is to simulate water quality and quantity efficiencies at various basin scales and locations (e.g. urban, agricultural, mountain area) which have been applied in many international research studies based on higher model performance. HSPF model consists of main three modules (PERLND, IMPLND, and RCHRES) and optional Utility module. Each of modules has different water quality, hydrologic parameters, and state variables (Bicknell et al., 2001). PERLND module simulates the water quality and quantity processes on a pervious land segment (**Figure 4.3**). The primary components of PERLND module simulate snow and ice (SNOW), water budget (PWATER), sediment (SEDMNT), water quality constituents (PQIAL and RQIAL). The auxiliary of PERNLD module modifies air temperature (ATEMP) and simulate soil temperature (PSTEMP).

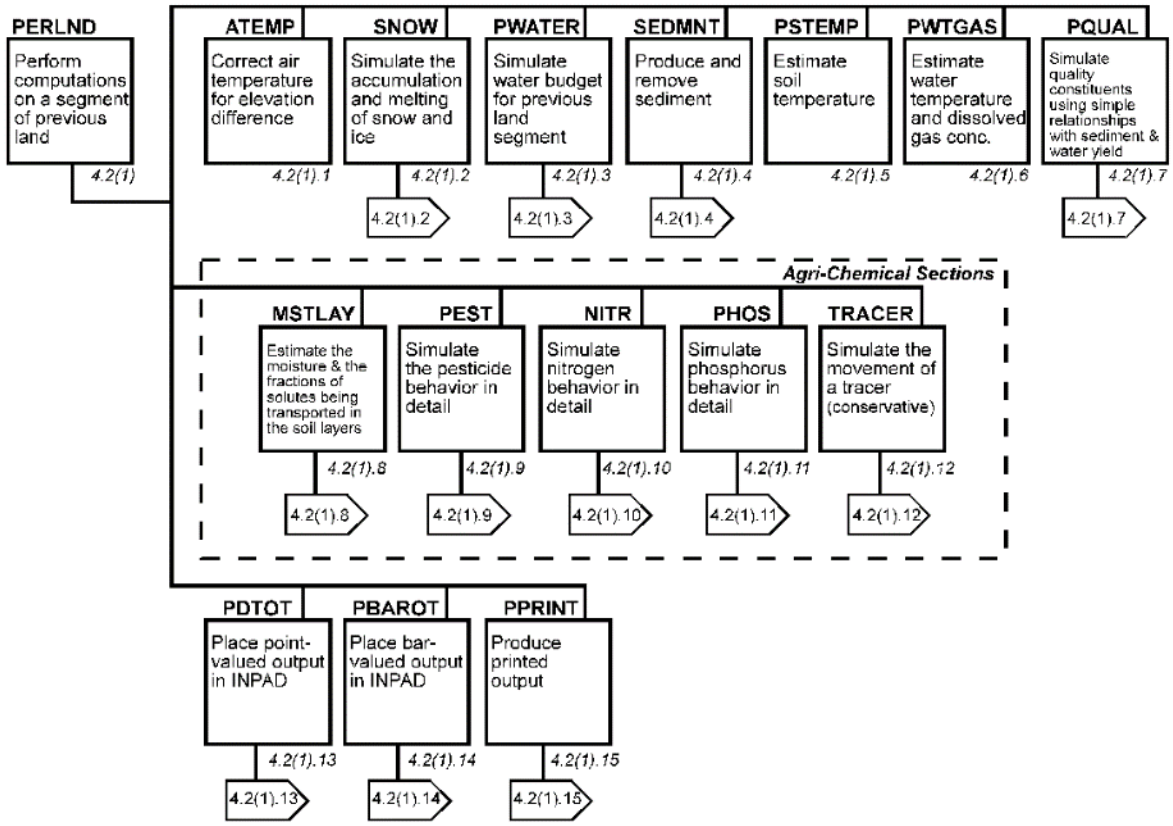


Figure 4.3. Structure chart for PERLND module (Bicknell et al., 2001).

IMPLND module indicates an impervious land segment processes. The impervious land occurs little or no infiltration and no subsurface flows. The structure and functions of IMPLND are given in **Figure 4.4**. Most sections of IMPLND module are similar to the corresponding in the PERLND module. For example, IWATER is similar to PWATER in PERLND module, SOLODS is similar to SEDMNT, IWTGAS is similar to PWTGAS, and IQUAL is similar PQUAL. Functions of SNOW and ATEMP sections are applied to pervious or impervious segment because both modules share their results. However, IMPLND module is less complex due to they do not consider infiltration and subsurface flows.

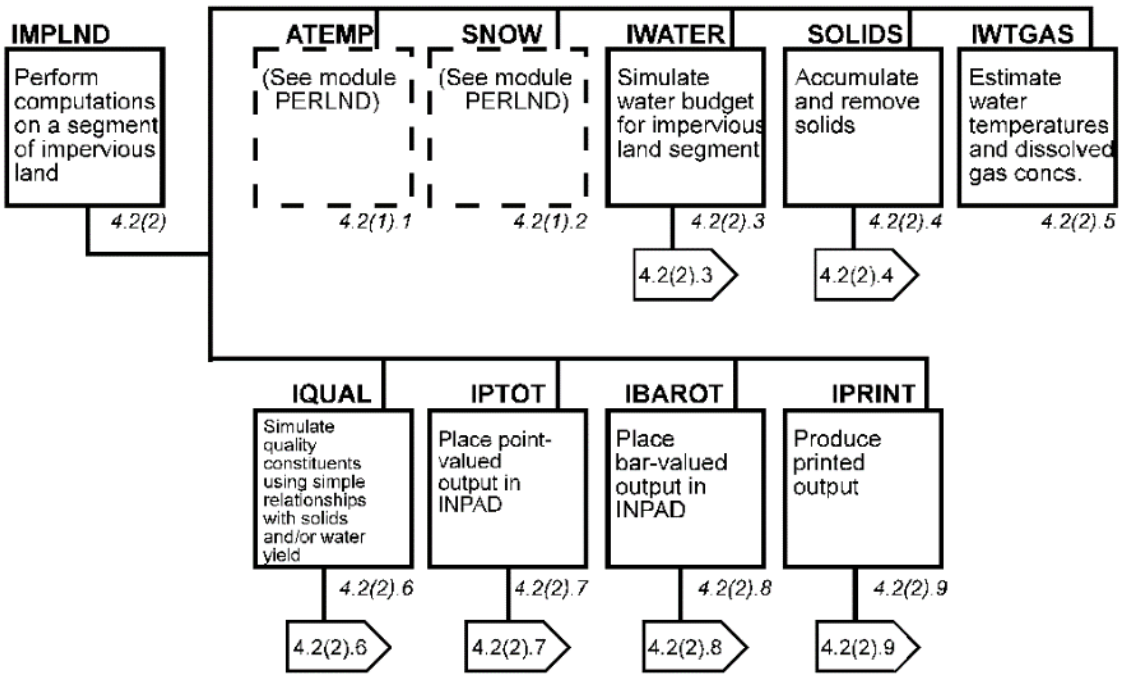


Figure 4.4. Structure chart for IMRLND module (Bicknell et al., 2001).

RCHRES module simulates several processing in a single reach of open, closed channel, or a completely mixed lake. It assumes flow is unidirectional. **Figure 4.5** shows **t**he function and major subdivisions of RCHRES module. To simulate water quality, behavior of conservative constituents(CONS), heat exchange and water temperature (HTRCH), behavior of inorganic sediment (SEDTRN), behavior of a generalized quality constituents(GQUAL), and behavior of constituents involved in biochemical transformations (RQUAL) sections should be active, advection of entrained constituents (ADCALC) section must be active. The water temperature is simulated if RQUAL is active. Within a module section, simulation of physical processes (longitudinal advection, sinking, benthal release) is always performed before simulation of biochemical processes.

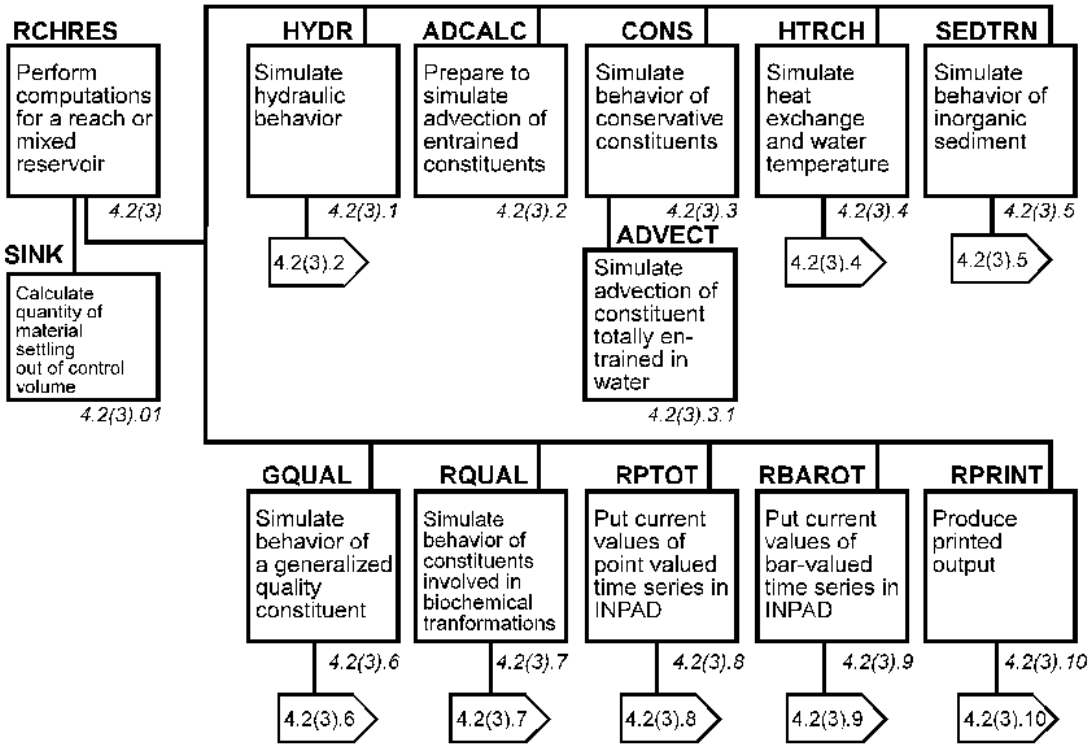


Figure 4.5. Structure chart for RCHRES module (Bicknell et al., 2001).

HSPF model provides optional Utility module for BMPRAC function. BMPRAC can simulate the impacts of Best Management Practices (BMP) by applying removal fraction to each constituent being modeled. A single instance of BMPRAC deals with the transfer of all mass loads from any pervious land in PERLND and impervious land in IMPLND to a single RCHRES. The same fractions are applied for each land use. Separate BMPRAC is required for each land segment to RCHRES connection due to the expected removal fractions are to vary by land use. The general equations for a removal are as follows:

$$\text{REMOVE} = \text{FRAC} * \text{INPUT} \tag{4.2}$$

$$\text{OUTPUT} = \text{INPUT} - \text{REMOVE} \tag{4.3}$$

where: REMOVE = mass removed; FRAC = fraction to remove; INPUT = inflow mass; OUTPUT = outflow mass.

4.3.3. Hydrological modeling processes

Stream network delineation was first conducted to determine the number of catchments and catchment sizes based on geographic information derived from DEM, land use, and channel profiles. As shown in **Figure 4.6**, a total of 15 calibration points was added to generate hydrographs. The study basins were delineated to have three, ten, and three catchments situated right above gauging stations for Elk, Illinois, and Baron Fork drainages, respectively. The drainage area for Baron Fork and Siloam Springs is part of the Illinois River basin.

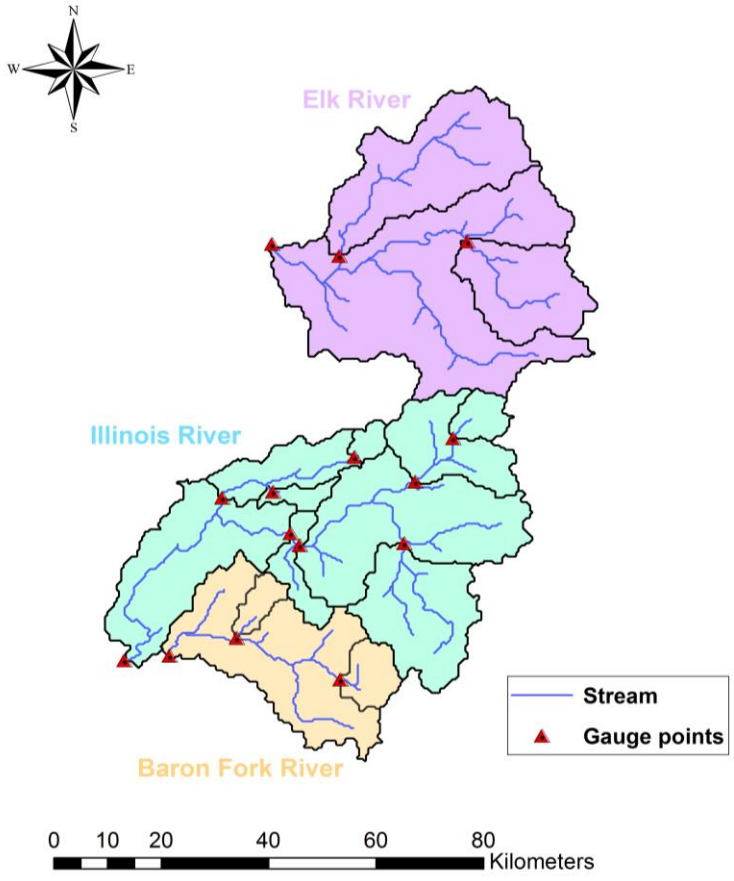


Figure 4.6. Basin delineation of the study basins, including Elk, Illinois, and Baron Fork Rivers

4.3.4. Model calibration

A total of six hydrologic parameters were identified as the most important parameters, based on a literature review for rainfall-runoff simulations using HSPF. Brief descriptions and typical ranges of these parameters are listed in **Table 4.3**. The set of entire model calibration parameters is not included here because such a parameter set varies and changes as calibration points move along with the stream channel network. Thus, the procedures of the calibration were conducted in a progressively upstream fashion, whereas a set of interior points that are not hydrologically linked represented the outlet that was left uncalibrated. Necessary effort has been made for model calibration and validation over the study periods by using the Parameter Estimation (PEST) software. The calibration procedure used in this study from downstream to upstream can be termed "bottom-up calibration" or "out-to-in calibration," which is a reverse calibration procedure, such as "top-down calibration" or "in-to-out calibration" (Ryu, 2009). Cross-validation techniques were adopted to verify how the model responds to independent data periods. Thus, instead of a split-sample calibration and validation experiment, several time segments (January 1, 1999–September 30, 1999, and October 1, 1999–September 30, 2000), which are subsets of simulation periods (October 1998 to September 30, 2000) have been selected for model evaluation processes owing to the data availability of NEXRAD in the short time simulation period. Thus, when the first-time segment is calibrated, the second time segment is set to a target window for validation and vice versa. Some of the first months (e.g., October–December 1998), however, were not used for cross-validation because this was utilized as a warm-up period to stabilize the simulation runs for HSPF. PEST was used for automatic calibration. PEST is a model independent parameter estimation, so it can be applied to many other modeling environments (Doherty and

Skahill, 2006; Fienen et al., 2009; Goegebeur and Pauwels, 2007; Kim et al., 2007). PEST recognizes the position of the model input file and the parameter set for calibration from a template file (e.g., .tpl). The objective function build in PEST is intended to minimize the summed weighted squared differences between volumes calculated on the basis of simulated and observed flows (not shown in this paper) over the entirety of the simulation periods (Doherty and Johnston, 2003). PEST uses the Gauss-Marquardt-Levenberg (GML) method (Marquardt, 1963), which is a nonlinear estimation technique, to minimize systematic errors between observed and simulated flows. The strength of the GML is reducing calibration times because it has fewer running times than other calibration methods, such as manual calibration and the shuffled complex evolution method developed by the Univ. of Arizona (Doherty and Johnston, 2003). PEST can provide a linear-based approximation of the uncertainties associated with parameters, which can be represented by the degree of correlation between them. This approximation is a reflection of the inherent nonuniqueness of the inverse problem byproduct of GML simulation in minimizing the user-specified objective function (Kim et al., 2007). TSPROC, the primary component of PEST, plays a key role in the calibrating processes because it can prepare the necessary files to formulate the objective functions for PEST (Doherty and Johnston, 2003). The multicomponent objective function entailed several observations that can easily be constructed by using TSPROC. TSPROC is also used for postprocessing, which is required to generate outputs from PEST calibration.

Table 4.3. Initial parameter values and ranges of HSPF.

Parameter	Definition	Units	Initial value	Value used in literature	Range of values	
					Typical [d]	Possible [d]
AGWRC	Base groundwater recession rate	None	0.98	0.95[a], 0.98[b], 0.99[c]	0.92-0.99	0.82(0.001)-0.999
INFILT	Infiltration rate	mm/h	4.06	2.03[a], 1.78-15.49[b], 1.78[c]	0.25-6.35	0.025-12.7(2540)
IRC	Interflow recession parameter	1/day	0.5	0.4[a], 0.6[b]	0.5-0.7	0.1-0.9
NSUR	Manning's roughness for overland flow	None	0.2	-	0.03-0.1	0.01-0.3(1.0)
SLSUR	Slope of overland flow plane	None	0.001	-	0.30-1.52	0.0001-4.57(304.8)
UZSN	Upper zone nominal soil moisture storage	mm	28.7	12.7[a], 17.53[b], 12.7[c]	2.54-25.4	1.27-50.8(254.0)

[a]Doherty and Johnston (2003).
[b]Kim et al. (2007).
[c]Im et al. (2003).
[d]Note that typical and possible values are from BASINS technical note 6, and values in parentheses are taken from the HSPF parameter section through WinHSPF graphical user interface (GUI).

4.3.5. Calibration Statistics

Three typical performance measures, including root-mean square error (RMSE), correlation coefficient (R), and Nash-Sutcliffe coefficient (NS) were used to evaluate how the HSPF model performs streamflow simulations against the observed flows in this study. RMSE is an absolute error measuring quantifying error with regard to the variable units. Thus, RMSE provides the SD of the model prediction error. R is the degree of colinearity between the simulated and observed flows. It ranges from 0.0 to 1.0; higher values indicate better agreement between the simulated and observed data. NS is the percentage of the observed variance and determines the efficiency criterion for model verification (Nash and Sutcliffe, 1970). It ranges from minus infinity to 1.0, with higher values indicating better agreement between observed and simulated data. If the NS value is greater than zero, the

model is deemed to be a better system simulation than the mean of observed data. R, NS, and RMSE are calculated by equation (4.4) - (4.6), respectively

$$R = \frac{\frac{1}{N} \times \sum_{i=1}^{N} (Q_{Qi} - \bar{Q}_{Qi}) \times (Q_{Si} - \bar{Q}_{Si})}{\sqrt{\frac{N \times \sum_{i=1}^{N} Q_{Qi}^2 - \left(\sum_{i=1}^{N} Q_{Q1}\right)^2}{N \times (N-1)}} \times \sqrt{\frac{N \times \sum_{i=1}^{N} Q_{Si}^2 - \left(\sum_{i=1}^{N} Q_{S1}\right)^2}{N \times (N-1)}}} \quad (4.4)$$

$$NS = 1.0 - \left[\frac{\sum_{i=1}^{N} (Q_{si} - Q_{Qi})^2}{\sum_{i=1}^{N} (Q_{Qi} - \bar{Q}_{Qi})^2}\right] \quad (4.5)$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (Q_{Qi} - Q_{si})^2\right]^{0.5} \quad (4.6)$$

where, Q_{Qi} and Q_{Si} are observed and simulated streamflow at time step, respectively. $\bar{Q}_{Qi}$ and $\bar{Q}_{Si}$ are mean observed and simulated streamflow for the simulation period. N is total number of values within the simulation period.

4.4. Results

Statistical analysis was first performed to evaluate the suitability of the model and the performance of calibrated and uncalibrated results for Baron Fork, Illinois, and Elk River drainage using PEST. Uncalibrated simulations were performed to observe the valuable gains against the level of calibration efforts (Smith et al., 2004). **Table 4.4** shows the statistical results of the HSPF model in Baron Fork, Illinois, and Elk River over the study period. The results show that the calibrated simulation outperformed the uncalibrated simulation over the study periods. The simulated flow indicates how the model represents the observed flow. After the calibration process, HSPF accurately simulates average flows that are similar to the observed average flows at Siloam Springs, Tahlequah, Baron Fork, and Elk catchments (**Figure 4.7**). Based on statistics (e.g., RMSE, R, and NS), the performance of the model after

calibration shows a noticeable difference from uncalibrated flows at Siloam Springs, Tahlequah, and Elk catchments. As expected, the model performs better after calibration at Baron Fork and Tahlequah catchment. The Elk catchment, however, displays a lower NS value, which is 0.28 because it still has a distinct flow difference between observed and calibrated average flow, even if the calibrated HSPF model decreased the average flow volume. **Figure 4.7 (a–d)** show hydrograph comparisons for the calibrated and uncalibrated simulation results at Illinois River, Baron Fork, and Elk River over the study period. These figures show hydrograph comparisons between simulated streamflow with and without calibration against the observed streamflow. At Tahlequah [**Figure 4.7(a)**] and Baron Fork catchments [**Figure 4.7(c)**], for example, the timing and magnitude of the peaks of the calibrated flow accurately match the observed flow.

Table 4.4. Calibrated and Uncalibrated Statistics during October 1, 1998–September 30, 2000.

	Baron Fork river		Illinois river				Elk river	
			Siloam		Tahlequah			
	Uncal	Cal	Uncal	Cal	Uncal	Cal	Uncal	Cal
Observed average flow (m^3/s)	10.70	10.70	19.21	19.21	32.00	32.00	21.44	21.44
Simulated average flow (m^3/s)	13.42	10.50	31.42	23.68	47.06	27.01	34.80	32.89
R	0.90	**0.94**	0.77	**0.83**	0.79	**0.85**	0.64	0.79
NS	0.82	**0.85**	0.44	0.53	0.55	**0.72**	0.17	0.28
RMSE	11.75	10.43	31.00	28.47	43.04	33.91	37.02	34.57

Note: Bold values indicate satisfactory calibration results, with R and NS greater than 0.8 and 0.7, respectively.

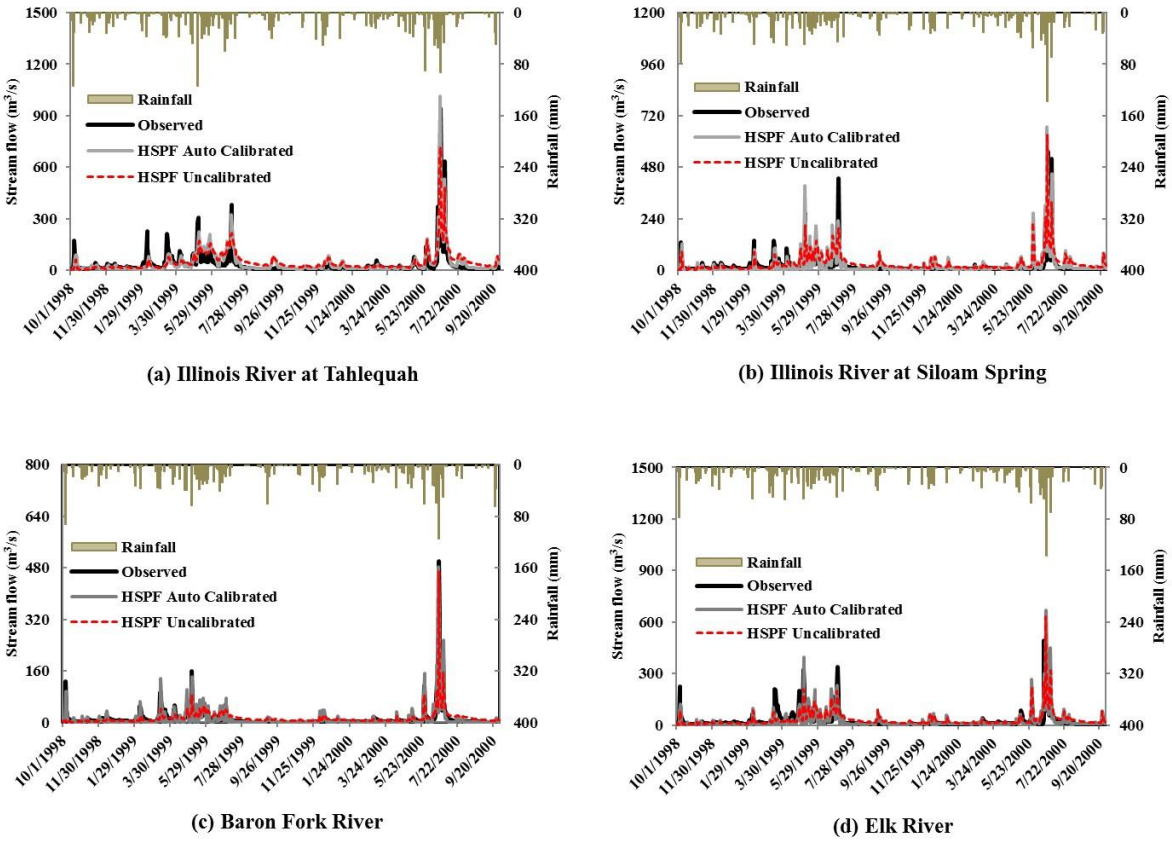


Figure 4.7. Hydrologic simulations for the study basins (October 1, 1998–September 30, 2000): (a) Illinois River at Tahlequah; (b) Illinois River at Siloam Springs; (c) Baron Fork River; (d) Elk River.

4.4.1. Evaluating model performance at different catchment sizes

R and NS were used to evaluate how HSPF performs as the catchment size changes. **Figure 4.8** illustrates how the first four discretization levels associated with different catchment sizes can be determined by decreasing gradually from downstream to upstream. Tributaries are gradually changed as calibration target points by the order of decreasing catchment areas, but the overall calibration results are reported at the previously used calibration points (CPs). First, the Tahlequah drainage, the largest basin in the study area, was investigated to evaluate how the model performance can be improved as the catchment sizes

decrease. As indicated by the values of R listed in **Table 4.5**, all calibrated simulations outperformed uncalibrated simulations, regardless of catchment sizes. R values gradually increase at CPs until the catchment size becomes 285 km^2, which is 11.47% of the Tahlequah drainage area. For example, the R value (0.87) at CP12 is higher than that (0.83) at CP5 within 1,489 km^2. However, when the catchment size decreases to 90 km^2, R values start to decrease at CP10 (R = 0 .65) within 90 km^2 (3.62% of the Tahlequah drainage area), although CP10 was selected as the calibration target point. Similar evidence was found at smaller catchment sizes (e.g., 49 and 37 km^2). This implies that such small catchment sizes do not contribute to overall model performance, based on R values in Tahlequah. A similar result was observed for NS values. As shown in **Table 4.5**, NS values within parentheses for the catchment areas of 2,484, 1,645, and 1,489 km^2 corresponding to CP1, CP5, and CP12, respectively, increased after model calibration (bold values in **Table 4.5**). However, for catchment sizes smaller than 1,489 km^2, NS values before and after calibration of the model vary depending upon catchment sizes and CPs. When the model was calibrated for the catchment area of 433 km^2 (17.43% of the Tahlequah drainage area), for example, the NS value increased at CP12 (NS = 0 .15), then decreased at Calibration Target Point 6 (NS = 0 .10). This pattern was also observed in smaller catchment sizes (e.g., 90 and 49 km^2). It may be claimed that the NS value is too low after calibration, but the primary focus of this study is to investigate the variability of model performances at different catchment scales, rather than calibration improvement. Low values of NS may be also the result of model bias produced by the calibration, with bias resulting either from differences in magnitude or time offset for the time-dependent model (McCuen et al., 2006), particularly in the smaller basins below 23% of the total drainage area. Thus, the smaller catchments (e.g., 37 and 49 km^2) do not experience

significant improvements by applying the calibration procedure. The NS values also indicate some different aspects of the results because they can determine the relative magnitude of the residual variance in comparison with the observed data. Therefore, the smaller catchment areas that simulate lower flow volume than large catchment areas have lower NS values and may be less effective for calibration of the model using the observed streamflow at multiple outlets. In addition to Tahlequah, the Siloam Springs, Elk River, and Baron Fork drainage areas were investigated to evaluate how the performance of the model depends on different catchment sizes. As shown in **Tables 4.6–8**, all calibrated simulation outperformed uncalibrated simulation, regardless of the catchment sizes and CPs. NS values in **Table 4.6** indicate that HSPF performs quite well for the catchment areas of 1,489, 433, and 337 km^2, which are 100, 29.08, and 22.63%, respectively, of the Siloam Springs drainage because NS values are greater after calibration. However, no calibration gain was observed at a catchment size of 90 km^2, which is 6.04% of the Siloam Springs drainage. As a result, there is a calibration gain when the catchment size is greater than 23% of the Siloam Springs drainage, but no gain was reported when further discretization was made.

The Elk River drainage areas were also examined to evaluate variations in the model performance depending upon the catchment size. As shown in **Table 4.7**, all calibrated simulations outperformed uncalibrated simulations, regardless of the catchment size and CPs. It appears that R and NS values increase gradually as the catchment size decreases at all CPs. The final examination was undergone for the Baron Fork drainage. Based on R values in **Table 4.8**, not surprisingly, the model performance after calibration outperformed uncalibrated results, regardless of the catchment size and CPs. Unlike R values, however, the NS results show different behavior. Thus, the highest NS values were observed at CP2 after

calibration, regardless of the catchment size from 65 to 795 km^2, via 105 km^2. However, NS values within the catchment areas of 105 and 65 km^2 presented poorer results than uncalibrated results at CP3 and CP15, which are, 13.21 and 8.18%, respectively, of the Baron Fork drainage. This result implies that CP3 and CP15, which are small catchment areas of Baron Fork, have no effect on improving the model and indicates limitations in the model calibration, especially for interior locations.

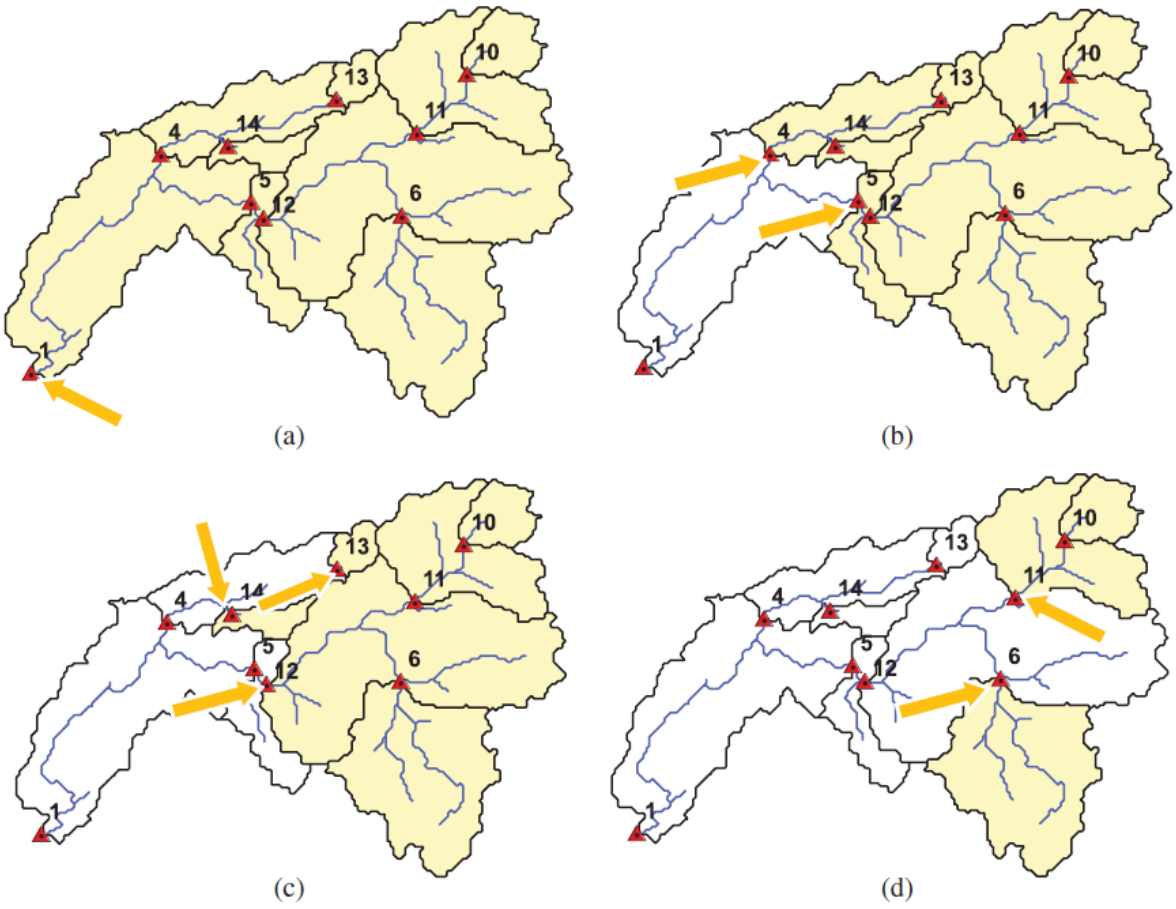


Figure 4.8. Examples of the first four discretization levels at different catchment sizes: (a) discretization level 1; (b) discretization level 2; (c) discretization level 3; (d) discretization level 4.

Table 4.5. R and NS before and after Calibration at Different Catchment Scales in the Tahlequah Drainage.

Area (km^2)	Uncal	CPs									Area Ratio (%)
		1	5	12	6	11	10	4	14	13	
2,484	**0.79 (0.55)**	**0.85 (0.72)**	-	-	-	-	-	-	-	-	100
1,645	**0.78 (0.60)**	**0.86 (0.65)**	**0.87 (0.75)[a]**	-	-	-	-	-	-	-	66.22
1,489	**0.75 (0.55)**	**0.82 (0.61)**	**0.83 (0.65)**	**0.87 (0.73)[a]**	-	-	-	-	-	-	59.94
433	0.48 (0.04)	0.50 (0.00)	0.59 (0.15)	0.66 (0.15)	0.67 (0.10)[a]	-	-	-	-	-	17.43
337	0.49 (-0.05)	0.52 (-0.07)	0.58 (0.01)	0.66 (0.01)	-	0.76 (0.04)[a]	-	-	-	-	13.57
285	0.61 (-0.02)	0.67 (-0.04)	-	-	-	-	-	0.81 (0.05)[a]	-	-	11.47
90	0.36 (-0.19)	0.37 (-0.20)	0.45 (-0.17)	0.50 (0.02)	-	0.70 (-0.16)	0.65 (-0.15)[a]	-	-	-	3.62
49	0.34 (-0.22)	0.34 (-0.22)	-	-	-	-	-	0.63 (-0.20)	0.48 (-0.21)[a]	-	1.97
37	0.39 (-0.22)	0.41 (-0.23)	-	-	-	-	-	0.74 (-0.21)	-	0.71 (0.20)[a]	1.49

Note: Dashes indicate no value because calibration was not applied to the catchment. The values within parentheses represent NS. The bold fonts indicate both R and NS increased after model calibration.

[a]Calibration target points.

Table 4.6. R and NS before and after Calibration at Different Catchment Scales in the Tahlequah Drainage.

Area (km^2)	Uncal	CPs				Area Rate (%)
		12	6	11	10	
1,489	0.77 (0.41)	0.83 (0.53)[a]	-	-	-	100
433	0.52 (0.19)	0.59 (0.24)	0.65 (0.33)[a]	-	-	29.08
337	0.56 (0.10)	0.60 (0.13)	-	0.80 (0.24)[a]	-	22.63
90	0.45 (-0.12)	0.46 (-0.11)	-	0.67 (-0.07)	0.74 (-0.07)[a]	6.04

Note: Dashes indicate no value because calibration was not applied to the catchment. The values within parentheses represent NS.
[a]Calibration target points.

Table 4.7. R and NS before and after Calibration at Different Catchment Scales in the Elk River Drainage.

Area (km^2)	Uncal	CPs			Area Rate (%)
		7	9	8	
2,258	0.64	0.79	-	-	100
619	0.55	0.63	0.67	-	27.41
365	0.42	0.53	-	0.77	16.16

Note: Dashes indicate no value because calibration was not applied to the catchment. The values within parentheses represent NS.
[a]Calibration target points.

Table 4.8. R and NS before and after Calibration at Different Catchment Scales in the Baron Fork Drainage.

Area (km^2)	Uncal	CPs			Area Rate (%)
		2	3	15	
795	0.90 (0.82)	0.94 (0.85)[a]	-	-	100
105	0.76 (0.07)	0.84 (0.12)	0.85 (0.03)[a]	-	13.21
65	0.56 (0.18)	0.70 (0.21)	-	0.73 (0.09)[a]	8.18

Note: Dashes indicate no value because calibration was not applied to the catchment. The values within parentheses represent NS.
[a]Calibration target points.

4.4.2. Identification of basin discretization threshold

Another reason for choosing 23% as the threshold was justified by evaluating the overall model performances. The study area was divided into two groups associated with the different catchment sizes constrained by R and NS values. The first group includes catchment sizes below 23% of the drainage area, such as drainage areas of 433 km^2 (17.43%) in Tahlequah, 90 km^2 (6.04%) in Siloam Springs, and 65 km^2 (8.18%) in Baron Fork; the results are shown in **Figure 4.9**. The second group is the counterpart of the first group: the catchment size is above 23% of the drainage area [e.g., 1,489 km^2 (59.94%) in Tahlequah, 433 km^2 (29.08%) in Siloam Springs, and 619 km^2 (27.41%) in Elk River]; the results are shown in **Figure 4.10**. Several of the interior points are the same in both the Tahlequah and Siloam Springs drainages, but they are classified into different catchment ratios by each drainage area. Thus, the interior points for the Tahlequah and Siloam Springs drainage are labeled as "(1)" and "(2)," respectively, in both **Figure 4.9 and 4.10**. As shown in **Figure 4.9**, which illustrates model performances below 23% of the drainage area, NS and R values vary (e.g., decrease or increase) depending upon CPs at different catchment scales. For example, both NS and R values increased at knso2 (1) and elmsp (1), whereas those values fluctuate at wslio (1) and caves (1). All calibration target points in **Figure 4.9** are included in the catchment size, which is below 23% of the drainage area. However, **Figure 4.10** shows that NS and R values all increased with decreased catchment area, regardless of calibration target points, when the catchment size is above 23% of the drainage area. These results imply that after calibration, a noticeable gain can be achieved when the catchment size is above 23% of the drainage area.

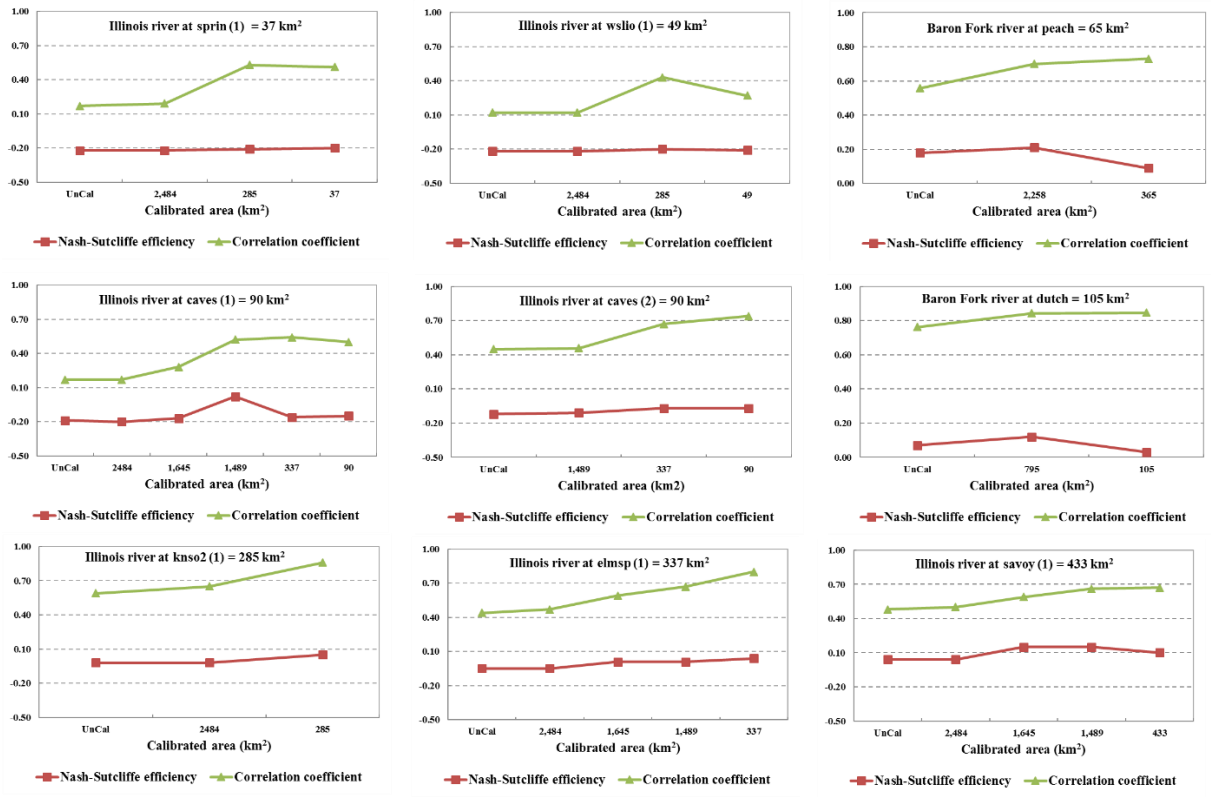


Figure 4.9. R and NS values at different basin scales below 22.63% of the drainage area.

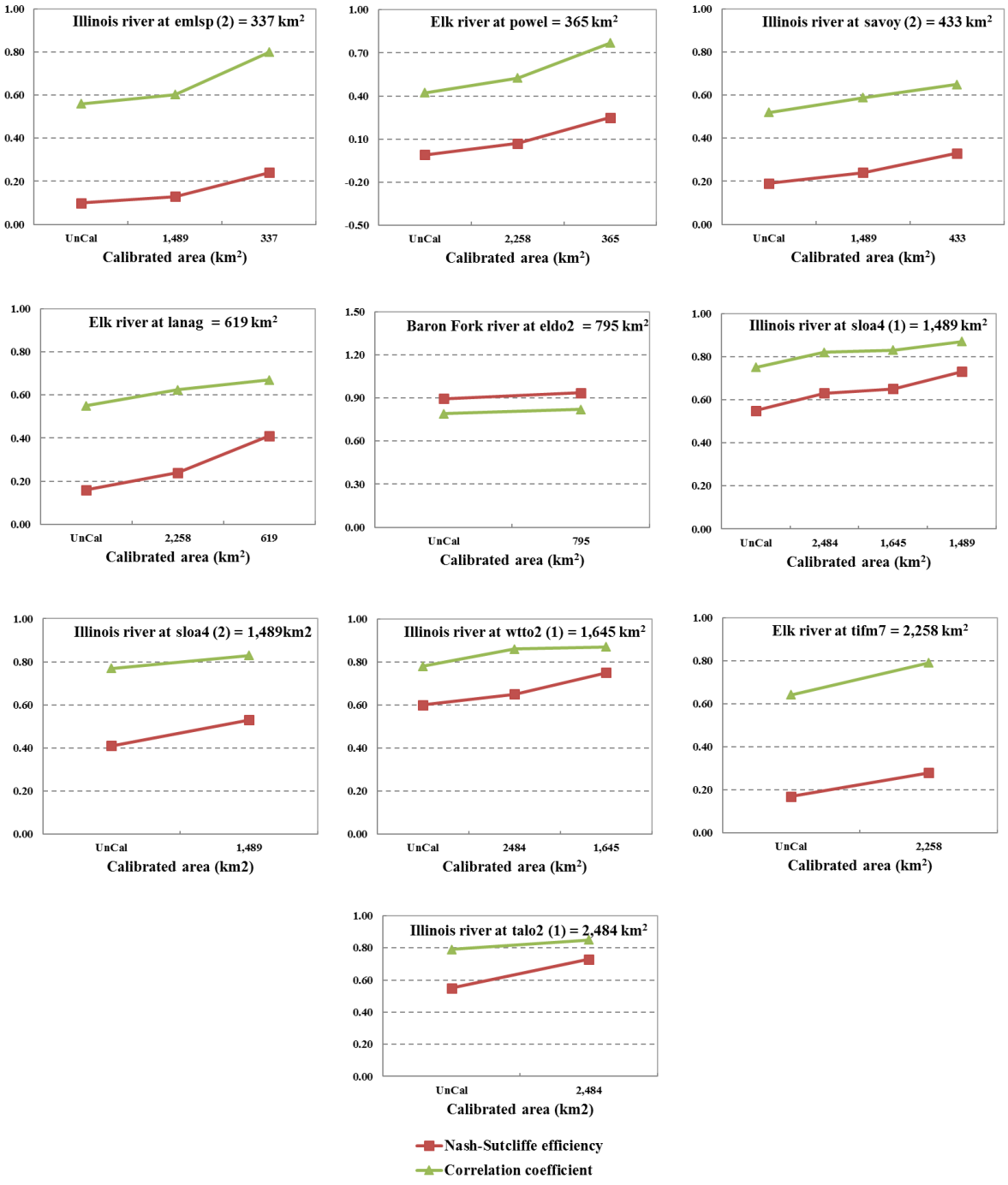


Figure 4.10. R and NS values at different basin scales above 22.63% of the drainage area.

CHAPTER 5. IMPROVING HYDROLOGICAL SIMULATIONS USING HSPF IN COMPUTER PARALLELISM

5.1. Introduction

A hydrologic model is commonly used to simulate real-world problems in many water-related fields, including hydrological, ecological, biological, and environmental studies (Borah and Bera, 2003; Howarth et al., 1996; Wang et al., 2007; Wu and Xu, 2006). Recent advances in data-intensive products, such as **N**orth **A**merican **L**and **D**ata **A**ssimilation system (NLDAS) and **NEX**t Generation **RAD**ar (NEXRAD) enable hydrologists to better characterize hydrological processes at higher spatial and temporal scales (Lerat et al., 2012; Nan et al., 2010; Smith et al., 2012). Yet, it is still challenging hydrologists to calibrate their models using these data- intensive inputs.

Due to insufficient observed datasets, computer simulation is a typical exercise to characterize hydrological processes and to enhance hydrological realizations based on physical and conceptual definite parameters. In general, hydrologists utilize 7-9 key parameters to calibrate their models for efficient simulations associated with cost and time (Goncu and Albek, 2010; Gallagher and Doherty, 2007; Kim and Ryu, 2013; Seong et al., 2015; Xu et al., 2007; Xie and Lian, 2013). Model performances, however, are constrained by the number of parameter set used, which doesn't necessarily ensure that the selected model performs best to characterize hydrological processes in the study area. Therefore, computer parallelism is a way to improve simulation performances by incorporating as many parameters as possible into hydrological modeling settings. With the development of the parallel computing technique, computational modeling has been rapidly advanced (Jordi and

Wang, 2012; Wang and Shen, 2012; Zhao et al., 2013). Although computer speed and capacity improves over time, the model calibration time is still challenging for many practitioners (Zhang et al., 2009).

There are two typical approaches to parallelize hydrological simulations. First, a parallel algorithm with parallel threads (e.g., multiple cores) within a single computer is one approach (Glotic et al., 2014; Li et al., 2011; Li et al., 2014; Wu et al., 2013; Zhang et al., 2013). The other approach is to implement a parallel algorithm in connection with multiple machines (Kalyanapu et al., 2011; Lecca et al., 2011). Although several studies have been conducted for parallelizing model calibrations to reduce computational time and effort with multiple threads in a single machine (Gorgan et al., 2012; Rouholahnejad et al., 2012; Yalew et al., 2013), few studies focus on implementing computer parallelism in multiple machines associated with cluster-based computing architecture to improve model performances. Thus, computer parallelism on cluster-based framework has not been fully implemented to find optimal parameters for hydrological simulations, especially Hydrological Simulation Program-Fortran (HSPF) modeling settings. Therefore, the research explores how computer parallelism can be implemented to measure the enhancement of hydrological simulations using HSPF so that hydrologists can apply it to their own simulation studies.

Figure 5.1 shows a flowchart of computer parallelism to calibrate HSPF model in a Linux cluster framework. A small Linux cluster system (sLCS) using Message Passing Interface (MPI) was first developed and then computer parallelism associated with four different calibration scenarios were implemented to improve model performances. The BEOPEST, a special version of PEST (a model-independent parameter optimization program) (Doherty and Skahill, 2006) was used to calibrate HSPF coupled with multiple cores in sLCS.

Computer parallelism was then evaluated by comparing model results based on calibration time and efficiency. Lastly, we conclude that hydrologic simulations using BEOPEST and HSPF in sLCS environment are a way to improve model performances, especially when many parameters are used for model calibration exercises. In this study, parallelizing calibration was only applied for streamflow parameters because the observed water quality data (e.g. TN, sediment, TP, water temperature, DO) was insufficient to determine optimal parameters using automatic model calibration. Manual calibration was used for the sediment and water quality parameters.

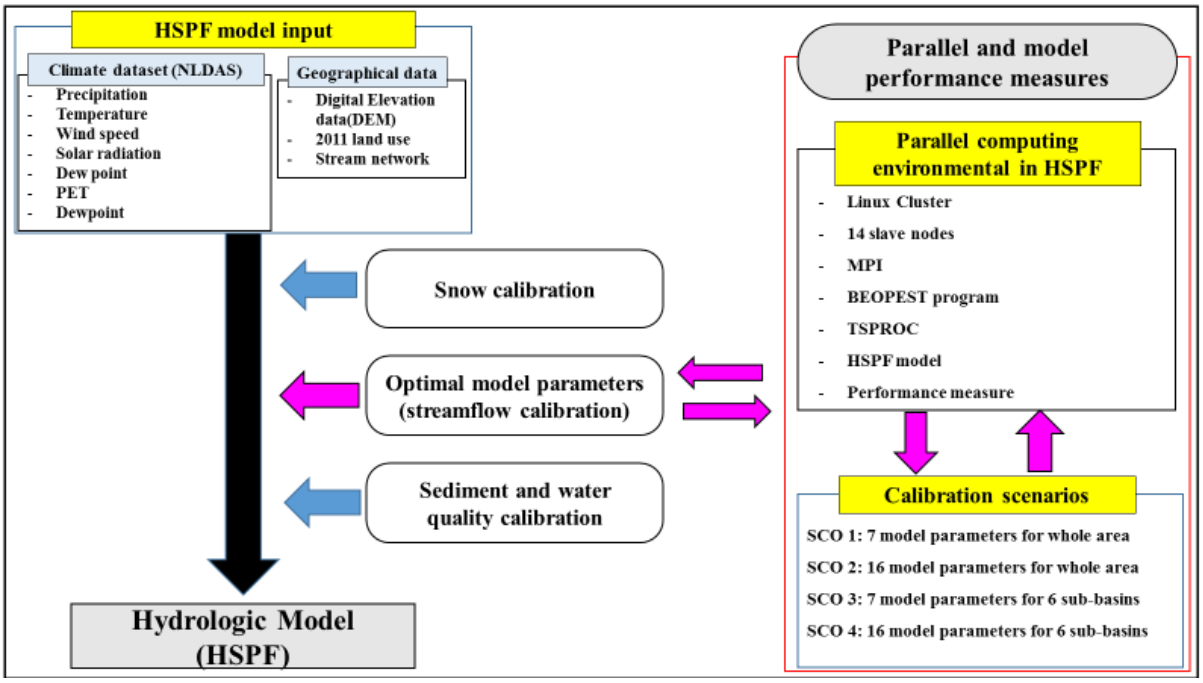


Figure 5.1. The flowchart for hydrologic model calibration and application.

5.2. Study area

In this research, the Boise River Watershed (BRW) was selected as the study area (**Figure 5.2**). As a tributary of the Snake River Watershed (SRW), the BRW plays a key role of providing water supply and lifeblood of Treasure Valley, but it contributes a significant

water quality problem to the lower Snake River. Major cities including Boise, Nampa, Meridian, and Caldwell are situated within the lower Boise River watershed (LBRW). The BRW is 10,619 km^2 with a mainstream length of 164 km stretch and flows into the Snake River near Parma. More than 40 percent of Idaho's population live in this watershed. Water quality issues driven by urbanization are also important within the LBRW. Reservoir operation related to climate change is one of the main concerns in the upper Boise River watershed (UBRW). The main physiographic characteristic of the BRW is a greater proportion of precipitation falling as snow at higher elevations. It becomes the cause of predictably high flows during the snow melting in spring and early summer. Also, localized flooding can arise in winter or spring by warm rains on snow and intense rainfall from convective storms during the warmer months. Therefore, climate change will highly affect streamflow variability in the UBRW.

The LBRW has many challenges to control and manage water supply, quality, and quantity by urbanization. The indication of water quality degradation has increased around in the major urban area. Urbanization rate has increased from 2.06 % in 1995 to 5.40% of the BRW in 2011 based on the information proved by USGS/National Land Cover Database (NLCD). This trend will continue over time due to ongoing and future urbanization in a changing climate.

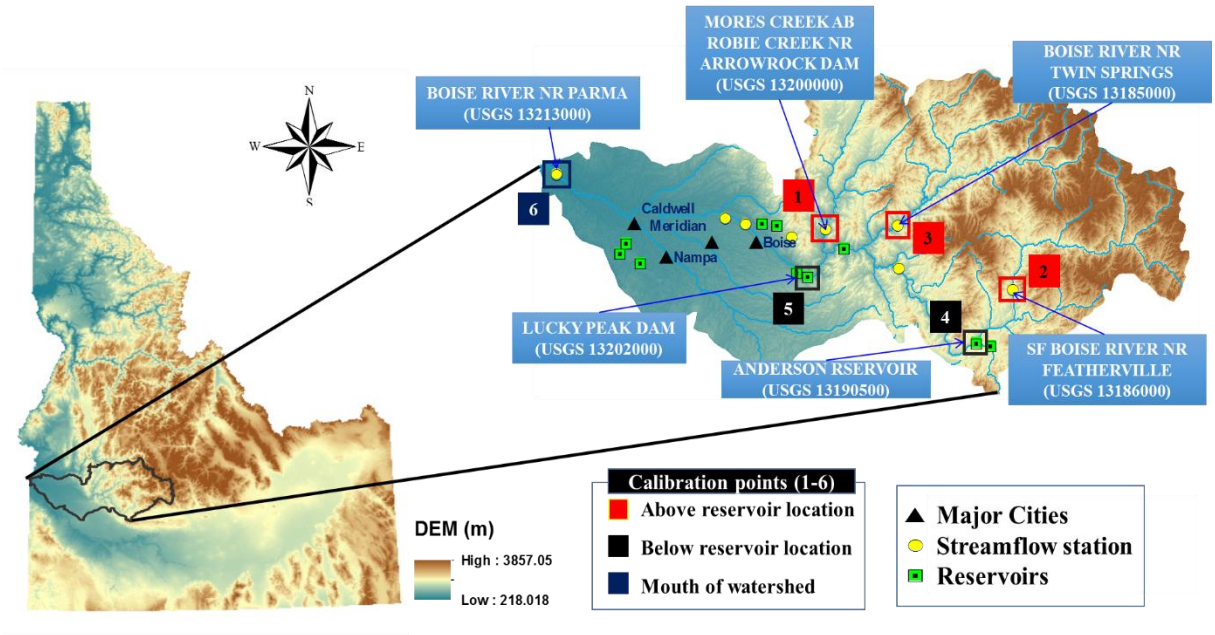


Figure 5.2. The Boise River Watershed (BRW) as the study area.

5.3. Material and methods

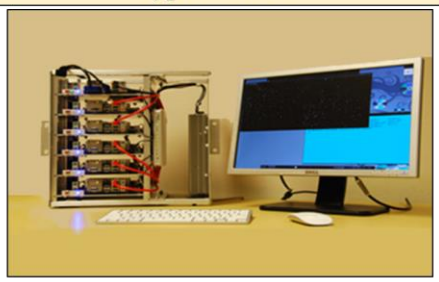
5.3.1. Linux Cluster System (LCS)

Parallel computing or processing is the ability to deal with multiple operations or tasks simultaneously. In parallel processing, more than one CPU or core is used to execute a program or multiple computational threads. Parallel processing makes the program run faster due to more CPUs or cores are running it. Cluster frames have greater computing power than a single large machine at a minimum cost. In this study, a small Linux cluster system (sLCS) is designed and built using a multimode Beowulf, which is a portable computer cluster in various computer architectures (Brown, 2004). Beowulf-style is a local memory machine by message via a network of master and slave clusters. It is designed relatively lower cost with high bandwidths and low latencies using fast Ethernet network.

Beowulf cluster can be utilized to develop the additional tools and parallelization of procedures because it is well supported, available, reliable, scalable, and distributed with open

source code. sLCS consists of inter-connected computers used as unified computer resources as a type of distributed computer system. It is widely used research coding test and central computational resources. sLCS can support Open Multi-Processing (OpenMP), Message Passing Interface (MPI), and Compute Unified Device Architecture (CUDA) parallelism. A main advantage of sLCS is easy to use and the cost-effective to build high-performance computing in small research groups in the sense that it costs less than $3,000 (e.g., 6 x VIA CN10000 with 2 GHZ CPU, 1GB of RAM, 500GB of the hard disk, 1Gbps Ethernet card). A typical sLCS is composed of 1 master and 5 slave nodes that are controlled by the master node and linked each other via local network protocols (e.g., TCP/IP). For this study, 14 cores (4 cores in the master node, 2 cores each × 5 slave nodes = total 14 cores) were used to implement computer parallelism for hydrological simulations. However, typical sLCS has several disadvantages to utilize parallelizing calibration in this study. For example, it is challenging to evaluate parallel performance efficiently for 14 cores because some cores of the master node should be used to run the operational system (OS) and necessary software programs. Another problem is the provided OS, which is Bootable Cluster CD (BCCD), is required an amount of computational knowledge and expertise. Also, it is not user-friendly interface and has lower compatibility with other software. In this study, we modified sLCS to apply parallelism for hydrologic model calibration (**Figure 5.3**). A laptop was used as the master node, while the typical sLCS was connected as slave nodes via an Ethernet cable. Primary roles of the master node include: 1) to use resources for running software, 2) to exchange model parameters with slave nodes, and 3) save and display simulation results. Note that Ubuntu 64-bit version was used to evaluate hydrological simulations using HSPF in computer parallelism.

Typical sLCS



Name	Content	Unit
OS	BCCD: Bootable Cluster CD 64 bit (Linux)	1
CPU	Intel Atom Dual Core (D525/510)	6
Core	Master node (4 cores)	4
	5 slave nodes (2 cores)	10
Ram	2 GB	6
Hard Drive	2.5" SATA 160GB	1
Software compatibility : Low		

Modify

Modified sLCS

Name	Content	Unit
OS	Ubuntu 64 bit (Linux)	1
CPU	Intel i-7 2.3GHz	1
	Intel Atom Dual Core (D525/510)	6
Core	Master node (8 cores)	8
	1 slave node (4 cores), 5 slave nodes (2 cores)	14
Ram	8 Gb (laptop)	1
	2 GB	6
Hard drive	300 GB SSD	1
Software compatibility: High		

Figure 5.3. The configuration of the small Linux cluster system (sLCS).

5.3.2. System setup in sLCS

A diskless sLCS in Beowulf system was developed in Ubuntu Operating System (OS). Slave nodes with no hard disk were connected via an ethernet network hub to the master node that can control, supervise, and monitor other slave nodes. The MPI library was used to coordinate multiple processes in distributed memory environment. For communication protocols, Secure Shell (SSH) method was used along with Ubuntu-based Diskless Remote Boot system (UDRB) to manage cluster nodes. Thus, UDRB installed in master node provides diskless environmental for slave nodes, accessing local hardware. A wireless network (wifi) was used for the master node to access the internet, while all other connections between the master and slave nodes are linked by network cards. The download and installation methods of DRBL are shown in **Appendix A** based on Ubuntu 16.04 version.

5.3.3. **T**ime **S**eries **PROC**essor (TSPROC)

A tool known as a general Time-Series Processor (TSPROC) is an interface to assist seamless data exchanges between input and output for optimal parameterization in hydrologic simulations. Basically, TSPROC generates the key input file for **P**arameter **EST**imation (PEST) program (which minimizes model biases formulated in user-specified objective function). To fully implement TSPROC in sLCS, a compilation of TSPROC is also required because the current version of TSPROC is compiled for Windows only.

5.3.4. **BEO**-**P**arameter **EST**imation (BEOPEST)

PEST, the model-independent nonlinear parameter estimation and optimization tool developed by Doherty and Skahill (2006) was used to assist data interpretation, model calibration, and predictive analysis. PEST uses a recursive gradient-based optimization technique, linearizing the nonlinear problem by iteratively computing the Jacobian matrix of sensitivities of model observations to parameters. The parameter estimation in PEST is accomplished using the Gauss-Marquardt-Levenberg algorithm (GML) to minimize the user-defined objective function (e.g., minimization of root mean squares between simulated and observed values). Thus, the BEOPEST is a tool to mitigate the computation burden and implement parallelism in PEST (Schreüder, 2009). Two communication protocols, such as Transmission Control Protocol/ Internet Protocol (TCP/IP) and MPI are commonly used, but BEOPEST-MPI method was utilized to run HSPF through data exchange between the master and the associated slave nodes because MPI is more suitable for diskless Linux cluster environment. As library sources, OPENMPI library was installed to compile parallel code fully workable in sLCS. Since BEOPEST in sLCS is a cost-effective approach, but powerful,

it is highly recommended to execute model calibration in computer parallelism with affordable costs at small research groups.

Figure 5.4 illustrates how computer parallelism is implemented in HSPF calibration processes. First, BEOPEST communicates between the master and slave nodes on sLCS using MPI protocol and then the master node reads the essential information required for running and optimization from the BEOPEST by simply sending model parameters to the slave nodes for simulations. Next, each slave node can create HSFP simulation files such as a template, instruction, and TSPROC control files from the model parameter sets as instructed by the master node based on the cores of each slave node and sends the simulation results back to the master node. These updated model parameters are then routed into the next BEOPEST iteration. BEOPEST iterations continue until the objective function difference is no longer improved as opposed to the previous value (optimal minimization is achieved). A computation speed during model calibration is then demonstrated and recorded depending upon the number of cores as well as calibration scenarios.

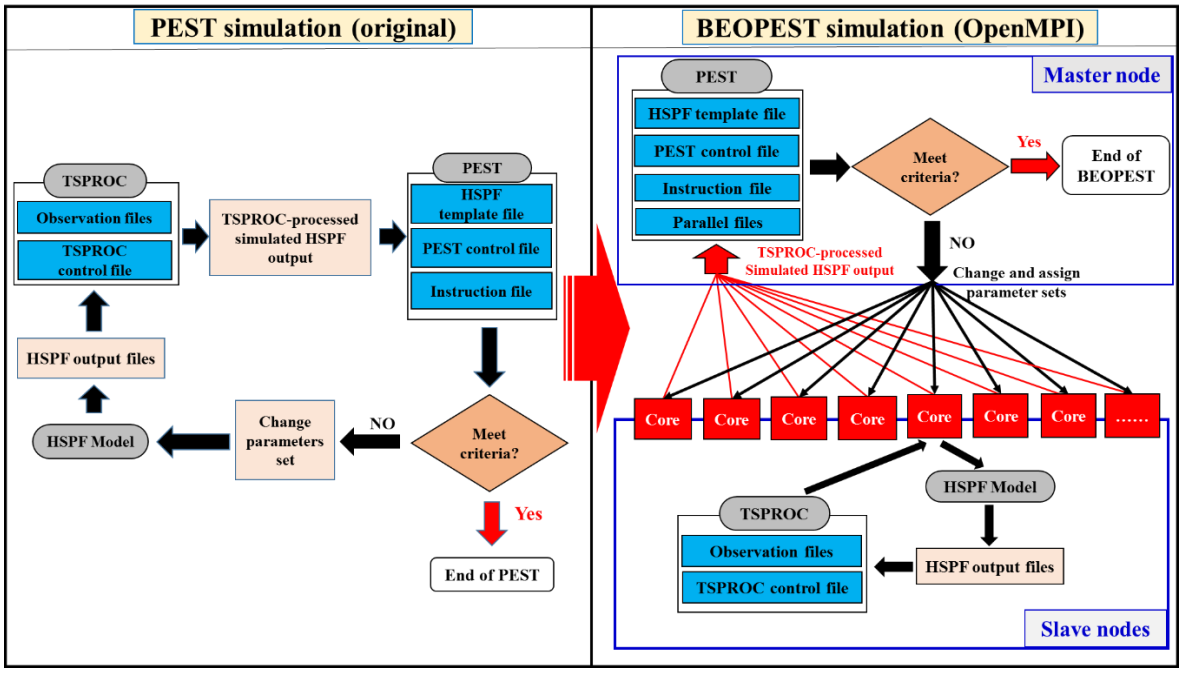


Figure 5.4. Schematic of computer parallelism with BEOPEST in HSPF model.

5.3.5. HSPF model input datasets

5.3.5.1. Climate datasets

HSPF model requires climate datasets (e.g. precipitation, temperature, potential evapotranspiration (PET), wind speed, dew point, solar radiation, and cloud cover) to simulate streamflow and water quality. Phase 2 of the North American Land Data Assimilation System (NLDAS-2) data was used as model input because it provides precipitation, temperature, downward solar radiation, downward longwave radiation, wind speed, specific humidity, surface pressure, and PET. NLDAS-2 is in 1/8th-degree grid spacing and generates real time from 01 Jan 1979 to present. The temporal resolution is provided from hourly to monthly. NLDAS-2 is initially developed to provide reliable initial land surface states and improve weather prediction by a multi-institution partnership, which are National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Prediction

Environmental Modeling Center (NCEP/EMC), National Aeronautics and Space Administration Goddard Space Flight Center (NASA GSFC), NOAA National Weather Service Office of Hydrologic Development (NWS/OHD), NOAA National Environmental Satellite Data and Information Service Office of Research and Allocations (NESDIS/ORA), Princeton University, Rutgers University, the University of Washington, and the University of Maryland (Cosgrove et al., 2003). NLDAS-2 is derived from the North American Regional Reanalysis (NARR) expect precipitation. NARR data is 32 km by 32 km spatial resolution and 3-hourly temporal intervals. These climate variables are bilinearly interpolated to the 12 km by 12 km spatial resolution and disaggregated to 1 hourly temporal frequency by required NLDAS 2 using linear temporal interpolation algorithms. The vertical different between NARR and NLDAS-2 of terrain elevation vertically adjust surface pressure, surface downward longwave radiation, air temperature, and specific humidity data. For the NLDAS-2 precipitation, it is derived from the Climate Prediction Center (CPC) unified gauge-based precipitation analysis with monthly Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al., 1994) adjustments for orographic impacts on precipitation. In areas where these data are unavailable, NARR precipitation is used instead. NLDAS has been evaluated for the improvement of hydrologic model and indicates well agreement results for observation for meteorological variables by several studies (Consgrove et al., 2003; Pinker et al., 2003; Luo et al., 2003; Xia et al., 2012a and 2012b; Xia et al., 2014). NLDAS-2 does not provide dew point and cloud cover. Therefore, an empirical equation was used to calculate the dew point (Oyj, 2013) and cloud cover data was derived from Climate Data Online (CDO) provided by National Climatic Data Center (NCDC) in the BRW.

$$T_{dew} = \frac{T}{\left[\frac{m}{log_{10}(\frac{P_{ws}}{A})} - 1\right]} \quad (5.1)$$

$$P_{ws} = A \times 10^{\left[\frac{m \times T}{T + T_n}\right]} \quad (5.2)$$

where, T_{dew} is calculated dew point (℃), T is air temperature (℃), A, m, and T_n are constants see **Table 5.1**. P_{ws} is water vapor saturation pressure over water (hPa).

Table 5.1. Constant values for dew point calculation.

Temperature range (℃)	A	m	T_n	Max error (%)
- 20 – 50	6.116441	7.591386	240.7263	0.083
50 – 100	6.004918	7.337936	229.3975	0.017
100 – 150	5.856548	7.277310	225.1033	0.003
150 – 200	6.002859	7.290361	227.1704	0.007
200 – 350	9.980622	7.388931	263.1239	0.395

The derived data from NLDAS-2 were then converted to Watershed Data Management (WDM) format to be used for HSPF as inputs. However, there are few issues to convert data from NLDAS-2 to WDM using conventional tools, which requires significant time and effort for all 112 grid points in the BRW. Since the existing WDM Utility tool cannot import a large volume of forcing data (roughly about 30 MB for a single file), we developed an R script to extract forcing data from NLDAS-2 to WDM file. The SARA Time Series Utility was then used to create a complete set of WDM file **(Figure 5.5)**.

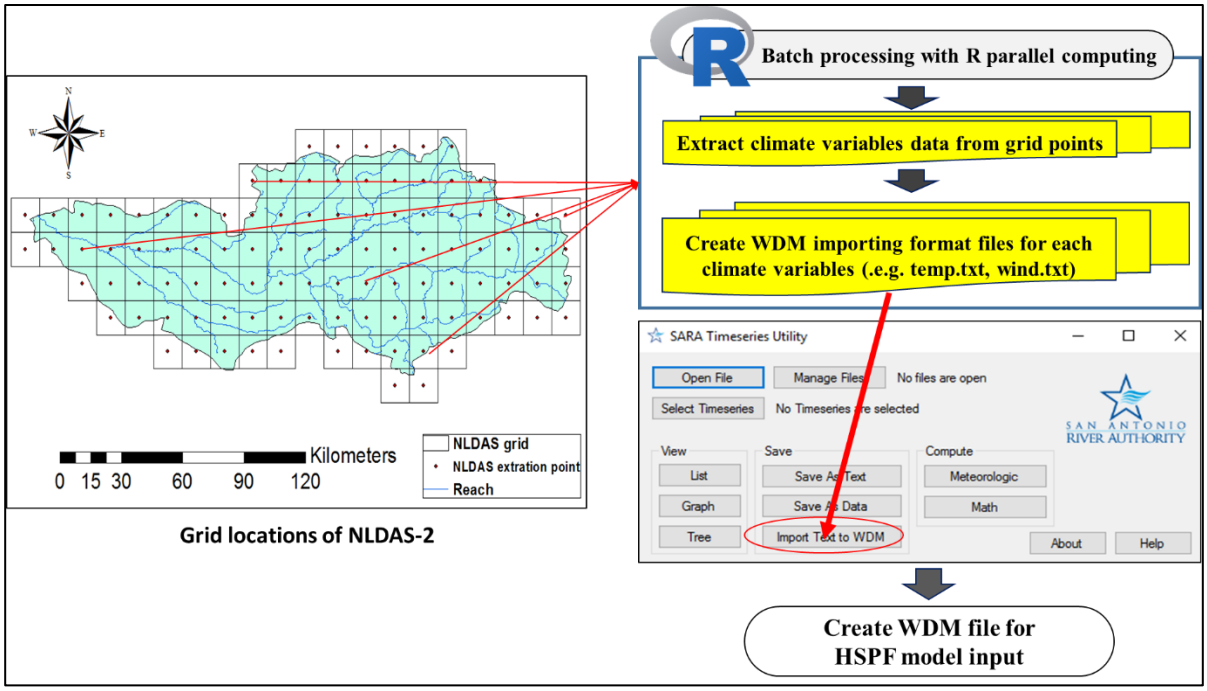


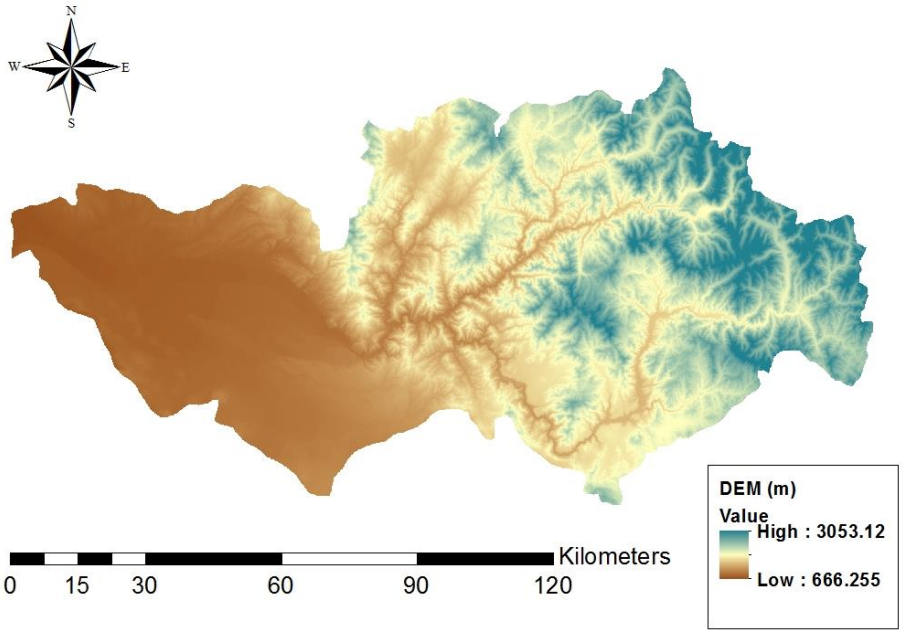
Figure 5.5. Scheme of creating WDM file from NLDAS-2 using R-Parallel and SARA Time series Utility.

5.3.5.2. Observed streamflow and water quality data

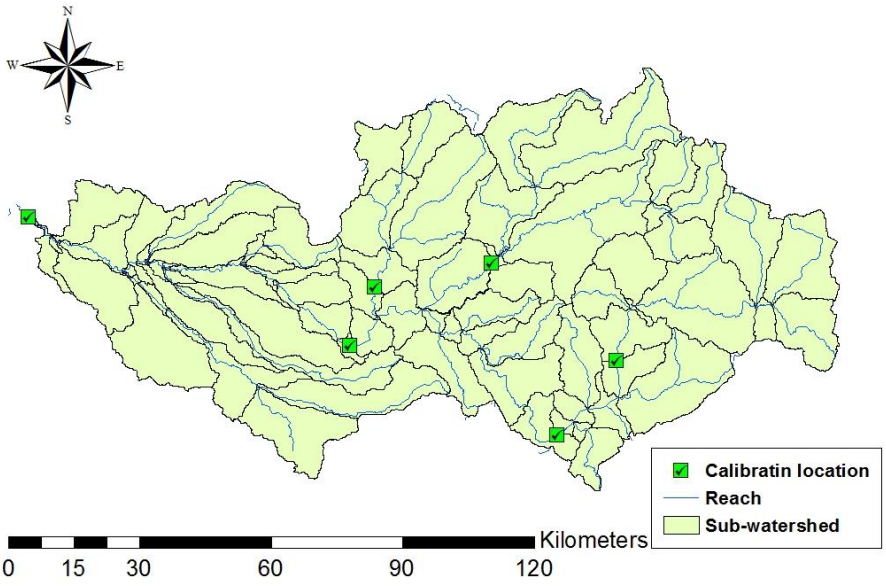
In the procedures of streamflow calibration and validation of HSPF model, we utilized observed streamflow from USGS gages 1320000 (calibration target point 1), 13186000 (calibration target point 2), 13185000 (calibration target point 3), 13190500 (calibration target point 4), 13202000 (calibration target point 5), and 1321995 (calibration target point 6). Water quality calibration of HSPF model employed observed water quality data at the watershed outlet. However, due to the insufficient observed water quality data on a daily interval (sediment, TN, TP, water temperature, and DO), a manual calibration procedure was applied to determine optimal model parameters for daily sediment, water temperature, DO, TN, and TP simulations.

5.3.5.3. Geological dataset

30 m resolution interval of digital elevation model (DEM) provided by US Geological Survey (USGS) was used for topographic relief mapping, watershed delineations, and flow direction computing using automatic watershed delineation in BASIN 4.1 program (**Figure 5.6 (a)**). The stream network was derived depending on DEM or USEPA stream reach file. Sub-basins outputs were easily defined via stream definition processes by assigning a minimum and maximum threshold of the sub-watershed area. National Hydrography Dataset (NHD) was used to develop stream routing for HSPF model since it was applied to integrated a detailed stream network in high resolution scale (1:100,000) which is required by various hydrology models (**Figure 5.6 (b)**). It is important to determine the details of stream network and the number of delineated sub-basins when observed streamflow data are available at interior location for "blind" simulation to determine the model performance of interior simulation when calibration is made at the mouth of the watershed. Total six observed streamflow points were selected to calibrate inlet locations in the BRW: 3 points are above the reservoirs, 2 points are below the reservoirs, and 1 point is at the watershed outlet (**Figure 5.2**). As environmental background data, LULC data was assigned to perform a more detailed assessment of LULC conditions and hydrologic cycles in the study area. LULC coverage, which is a polygon shape in vector format, indicates boundaries associated with LULC classification including urban, agricultural land, forest land, water/wetland, shrubland, grassland, and barren/mining land (**Figure 5.6 (c)**).

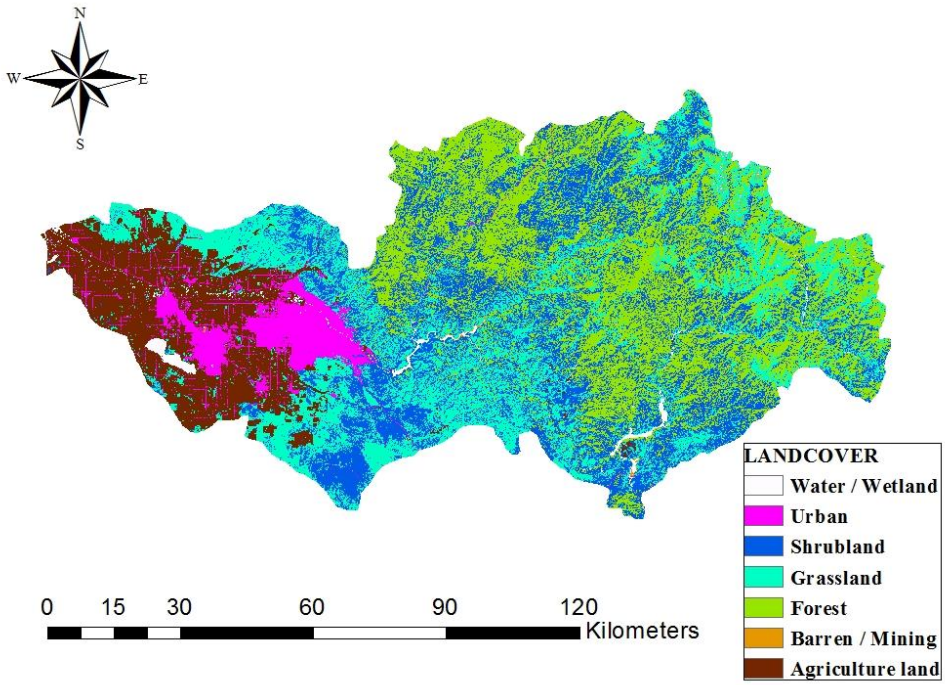


(a) DEM (30 m by 30 m)



(b) Stream reach and observed streamflow points

Figure 5.6. Geospatial datasets.



(c) 2011 LULC

Figure 5.6. Geospatial datasets (Cont.).

5.3.6. Streamflow calibration scenarios

Table 5.2 lists parameter name, unit, initial value, and ranges for HSPF calibrations. Four different calibration scenarios were considered and applied during the simulation period, January 1, 1999 to December 29, 2015. First two years (January 1, 1999 – December 31, 2000) were used for a spin-up period to stabilize HSPF. Specifically, Scenarios (SCOs) 1 and 2 are designed to calibrate the whole basin with different parameter sets. Thus, SCO 1 used 7 model parameters provided from **chapter 4** and SCO 2 used 16 model parameters. SCO 3 and SCO 4, however, were designed to calibrate the model with different parameter sets for 6 different sub-basins as shown in **Figure 5.7**. Thus, SCO 3 used 7 model parameters for 6 sub-basins (total 42 model parameters = standard 7 model parameters × 6 sub-basins) and SCO 4 used 16 model parameters for 6 sub-basins (total 96 model parameters = 16 model parameters

× 6 sub-basins). For each calibration scenario, computer parallelism was applied to evaluate its performances using BEOPEST in sLCS.

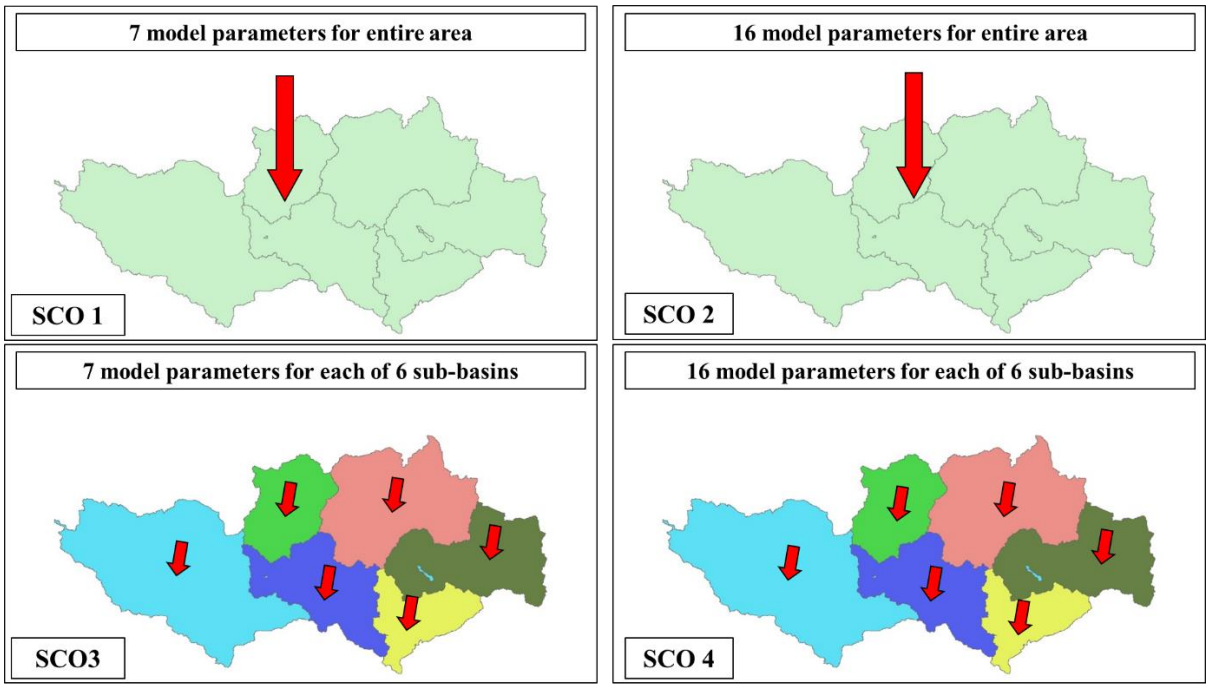


Figure 5.7. Model calibration scenarios by calibration locations and model parameter sets.

Table 5.2. Initial values and parameters range for HSPF model.

Parameter	Definition	Units	Initial value	Range of values	
				Typical	Possible
AGWETP	Fraction of remaining potential evapotranspiration from active groundwater	None	0	0.0 – 0.05	0 – 1.0
GWRC*	***Base groundwater recession rate***	***None***	***0.98***	***0.92 - 0.99***	***0.82 - 0.999***
BASETP*	***Fraction of potential evapotranspiration from baseflow***	***None***	***0.02***	***0.0 - 0.05***	***0 - 1.0***
CEPSC	Interception storage capacity	Inch	0.10	0.03 – 0.20	0.01 – 0.40
DEEPFR	Fraction of groundwater inflow to deep recharge	None	0.1	0.0 - 0.2	0.0 - 1.0
INFILT*	***Infiltration rate***	***in/hr***	***0.16***	***0.01 – 0.25***	***0.001 – 0.50***
INTFW	Interflow inflow parameter	None	2.0	1.0 - 3.0	1.0 – 10.0
IRC*	***Interflow recession parameter***	***1/day***	***0.5***	***0.5 - 0.7***	***0.1 - 0.9***
KVARY	Variable groundwater recession flow	1/inch	0	0.0 – 3.0	0.0 – 5.0
LZETP	Lower zone evapotranspiration parameter	None	0	0.0 - 0.7	0.1 - 0.9
LSUR	Length of the assumed overland flow	Feet	500	200 - 500	100 – 1,000
LZSN*	***Lower zone nominal soil moisture storage***	***Inch***	***6, 6.5***	***3.0 – 8.0***	***2.0 – 15.0***
NSUR	Manning's roughness for overland flow	None	0.2	0.03 - 0.1	0.01 - 1.0
SLSUR*	***Slope of overland flow plane***	***None***	***0.001***	***0.30 - 1.52***	***0.0001 - 304.8***
UZSN*	***Upper zone nominal soil moisture storage***	***Inch***	***1.128***	***0.10 – 1.0***	***0.05 – 2.0***
INFEXP	Exponent in infiltration equation	none	2.0	2.0 – 2.0	1.0 – 3.0

* indicates the model parameters that were used chapter 4.

5.3.7. Water quality calibration

For water quality, manual calibration was applied due to insufficient observed daily data in the BRW. The calibration parameters for water quality were selected based on BASINS Technical Note (US.EPA, 2000). **Table 5.3** shows the model parameters of sediment erosion and solids wash-off to calibrate sediment simulation using the HSPF model. **Table 5.4** indicates the model parameters of water quality constituents for TN, TP, DO, and water temperature.

Table 5.3. Calibration model parameters of sediment erosion and solids wash-off.

Parameter	Definition	Units	Initial value	Range of values	
				Typical	Possible
SMPF	Management Practice (P) factor from USLE	none	1.0	0.0 – 1.0	0.0 – 1.0
KRER	Coefficient in the soil detachment equation	complex	0.0	0.15 – 0.45	0.05 – 0.75
KSER	Coefficient in the sediment washoff equation	complex	0.0	0.5 – 5.0	0.1 – 10.0
KGER	Coefficient in soil matrix scour equation	complex	0.0	0.0 – 0.5	0.0 – 10.0
KEIM	Coefficient in the solids washoff equation	complex	0.0	0.5 – 5.0	0.1 – 10.0
ACCSDP	Solid accumulation rate on the land surface	ton/ha/day	0.0	0.0 – 0.0023	0.0 – 0.035

Table 5.4. Calibration model parameters of nutrients, DO, and water temperature.

Parameter	Definition	Units	Initial value	Typical Range of values
KNO320	The nitrate denitrification rate at 20 degrees C	hr^{-1}	0.05	0.001 – 0.4
REAK	The empirical constant in the equation used to calculate the reaeration coefficient	hr^{-1}	1.0	0.2 – 2.0
KBOD20	The unit BOD decay rate at 20 degrees C	hr^{-1}	0.02	0.00004 – 0.04
KODSET	The rate of BOD setting	m/hr	0.0	0.00012 – 0.015
MALGR	The maximum unit algal growth rate for phytoplankton	hr^{-1}	0.3	0.008 – 0.3
PHYSET	The rate of phytoplankton setting	m/hr	0.0	0.00031 – 0.17
CFSAEX	The correction factor for solar radiation	none	0.5	0.001 – 2.0
KATRAD	The long-wave radiation coefficient	none	6.5	1.0 – 20.0

5.3.8. Parallel performance

To evaluate model performances in computer parallelism, performance criteria, including parallel program runtime, the size of parallel task, run time, time reduction, speedup, efficiency, scalability, and more can be considered. However, only program run time (PT), time reduction rate (PP), speedup (PS) and efficiency (PE) were used for this study at our convenience. Note that speedup (PS) is defined as the degree of true time reduction between a serial computation and parallel computation and this measure indicates the relative improvement of model performance during executions. A notation of PS is proposed by Amdahl's law to compute the theoretical maximum speedup when multiple nodes are used. It is denoted as:

$$PS = \frac{T_S}{T_P} \quad (5.3)$$

where, PS is speedup. T_s is execution time of a serial application on a processor. T_p is execution time of parallel application on p processors.

The parallel efficiency (PE) is another way to measure the effectiveness of processor computation. Under an ideal condition in computer parallelism, PS should be equal to all the cores used with maximum efficiency, which is 1. Although PS varies depending upon the number of cores used, PE should be in between 0 and 1 in real-world applications due to the interference of physical components associated with load balancing, lack of hardware capacity, network connection, and other physical constraints, if any. The parallel efficiency (PE) is denoted as:

$$PE = \frac{S}{P} \quad (5.4)$$

where, s = a processor, p = multiple processors

5.3.9. HSPF performance

Five typical performance measures, including Correlation coefficient (R), Nach-Sutcliffe coefficient (NS), Root mean square error (RMSE), Mean Square Error (MAE), and Percentage of Bias (PBIAS) were used to evaluate how HSPF model simulate streamflow against observed streamflow in the BRW. The definition of R, NS, and RMSE indicate in 4.3.5 section from **chapter 4**. MAE is a quantity measurement how close simulated values are to the eventual results. Lower values of RMSE and MAE show better model performance and zero value indicates a perfect fit. PBIAS calculates the average tendency of the simulated values to be larger or smaller than observed counterparts (Gupta et al., 1999). Zero is the

optimal model performance. A positive value shows underestimation bias and a negative value indicates overestimated bias for simulated results. R, NS, RMSE, MAE, and PBIAS are calculated by equation (4.4) - (4.6) and equation (5.5) - (5.6), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | Q_{Oi} - Q_{Si}| \tag{5.5}$$

$$PBIAS = \frac{\sum_{i=1}^{N}(Q_{Oi} - Q_{Si})}{\sum_{i=1}^{N} QY_{Oi}} \times 100 \tag{5.6}$$

where, Q_{Qi} and Q_{Si} are observed and simulated streamflow at a time step, respectively. $\bar{Q}_{Qi}$ and $\bar{Q}_{Si}$ are mean observed and simulated streamflow for the simulation period. N is total number of values within the simulation period.

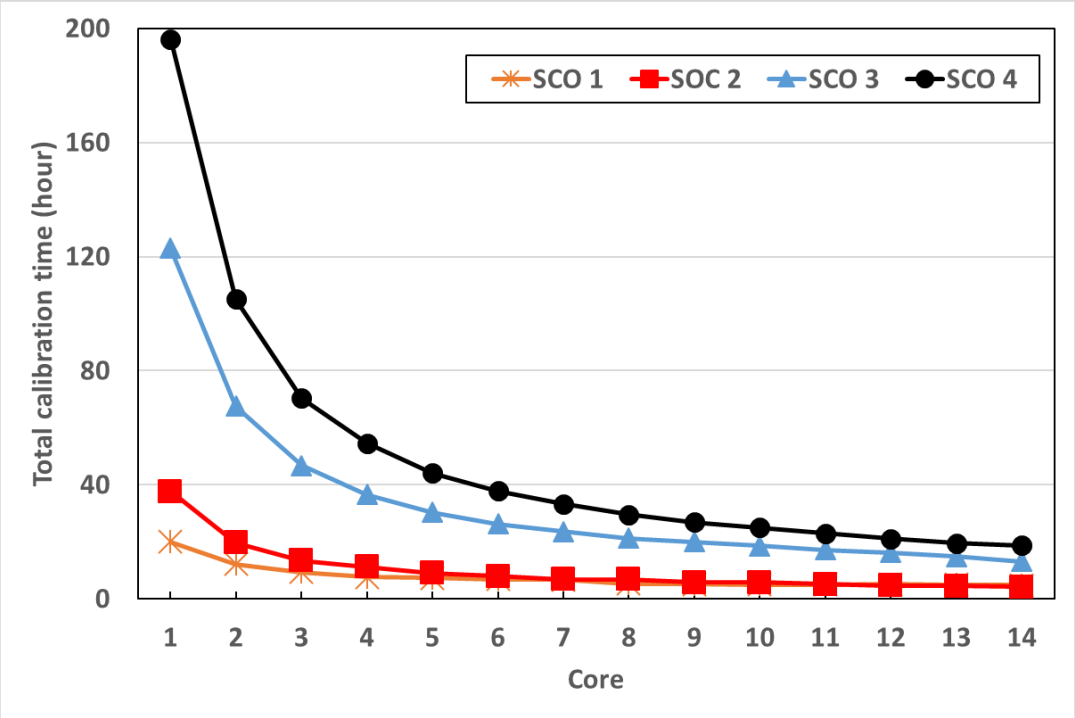
5.4. Results

5.4.1. Parallel Performance

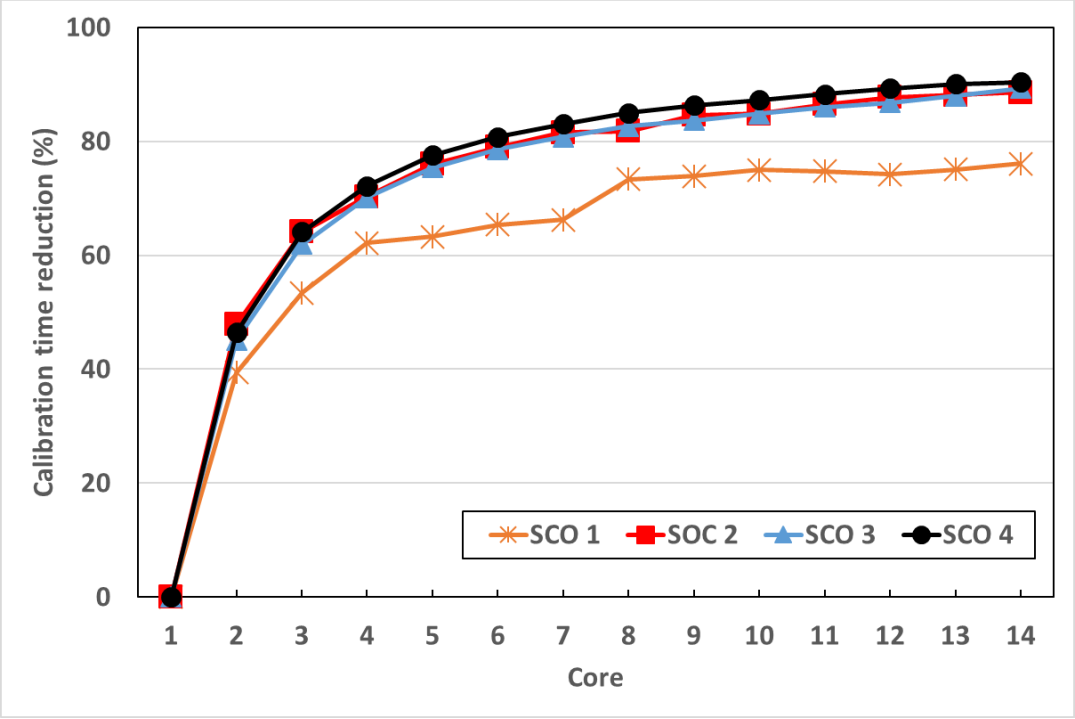
Parallel performances in sLCS were evaluated based on four different HSPF calibration scenarios using BEOPEST. **Figure 5.8** and **Table 5.5** show the results of total calibration time (PT), the percentage of time reduction (PP), speedup (PS), and efficiency (PE) by the number of core processes with four different calibration scenarios. Obviously, PT decreased as the number of cores increases. Note that PT of SCO 4 was about 10 times longer than that of SCO 1 when a single core is used with 7 parameters. However, when model calibrations were conducted using 2 to 8 cores, PT gradually decreased until no distinct improvement was observed at 9 cores above (See **Figure 5.8 (a)**). PP also shows the similar

pattern in the sense that calibration with multiple cores can have time-saving advantages as seen in **Figure 5.8 (b)**. Thus, the reduction rate of the total calibration time (PP) was achieved for 76%, 89%, 89%, and 90% from SCO 1, SCO 2, SCO 3, and SCO 4, respectively.

For PS of parallelizing calibration, SCO 1 didn't gain many benefits as opposed to other scenario cases, while PS values of SCO 2 and SCO 3 gradually increased as the number of cores increases. SCO 4 resulted in the highest speedup value in this analysis. It implies that the loss of speedup is due to the limitation of network connection and hard disk capacity. Theoretically, if the number of parallel jobs is set, each core files simultaneously read and are written to the hard disk via a network. For this reason, the speedup would not reach up to 14 even if 14 cores are fully used due to network constraints. Similarly, PE is most likely less than ideal value, which is one because of the system overhead issues associated with physical constraints (e.g., network bandwidth and/or throughput between cores). The results showed that SCO 1 was the lowest efficiency and SCO 4 was the highest PE. Overall, SCO 4 was the best parallel performance as opposed to other calibration scenarios regardless of the number of cores. The result implies that BEOPEST in sLCS works well especially when many hydrological parameters need to be calibrated at multiple sub-basins.

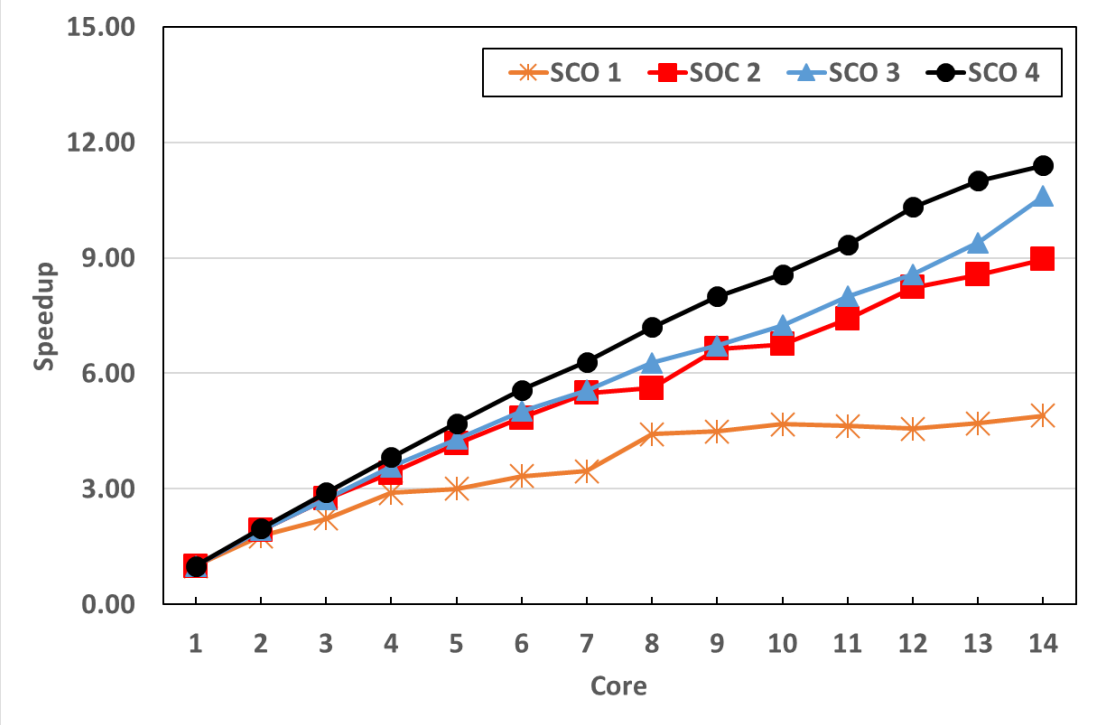


(a) Total calibration time (PT)

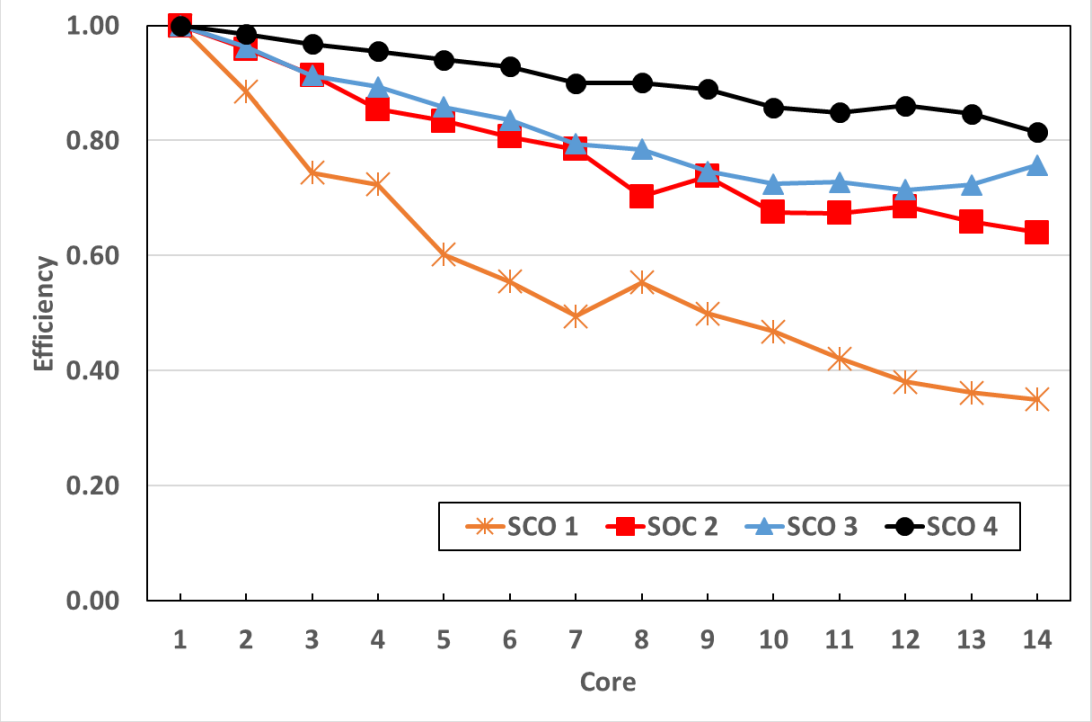


(b) The percentage of time reduction (PP)

Figure 5.8. Performance measures, including total calibration time (PT), the percentage of time reduction (PP), Speedup (PS), and Efficiency (PE) in computer parallelism.



(c) Parallel speedup (PS)



(d) Parallel efficiency (PE)

Figure 5.8. Performance measures, including total calibration time (PT), the percentage of time reduction (PP), Speedup (PS), and Efficiency (PE) in computer parallelism (Cont.).

Table 5.5. Total calibration time, speedup, and efficiency in using the different numbers of core computing processes in parallel calibration.

Calibration Scenarios	Core	Total hours of calibration time (PT)	% Reduction (PP)	Speedup (PS)	Efficiency (PE)
SCO 1	1	19.98	-	1.00	1.00
SCO 2		37.62	-	1.00	1.00
SCO 3		122.88	-	1.00	1.00
SCO 4		196.27	-	1.00	1.00
SCO 1	2	12.11	39.37	1.77	0.88
SCO 2		19.62	47.84	1.92	0.96
SCO 3		67.49	45.08	1.92	0.96
SCO 4		105.06	46.47	1.97	0.99
SCO 1	3	9.30	53.43	2.23	0.74
SCO 2		13.49	64.14	2.74	0.91
SCO 3		46.77	61.94	2.74	0.91
SCO 4		70.37	64.15	2.90	0.97
SCO 1	4	7.56	62.13	2.90	0.72
SCO 2		11.17	70.32	3.42	0.85
SCO 3		36.65	70.18	3.58	0.89
SCO 4		54.57	72.20	3.82	0.96
SCO 1	5	7.33	63.28	3.01	0.60
SCO 2		9.02	76.03	4.17	0.83
SCO 3		30.28	75.36	4.29	0.86
SCO 4		44.09	77.54	4.70	0.94
SCO 1	6	6.92	65.36	3.33	0.55
SCO 2		7.90	78.99	4.84	0.81
SCO 3		26.33	78.57	5.01	0.84
SCO 4		37.76	80.76	5.57	0.93
SCO 1	7	6.74	66.26	3.46	0.49
SCO 2		6.94	81.57	5.49	0.78
SCO 3		23.67	80.74	5.56	0.79
SCO 4		33.27	83.05	6.30	0.90

Calibration Scenarios	Core	Total hours of calibration time (PT)	% Reduction (PP)	Speedup (PS)	Efficiency (PE)
SCO 1	8	5.32	73.37	4.42	0.55
SCO 2		6.86	81.75	5.62	0.70
SCO 3		21.28	82.68	6.28	0.78
SCO 4		29.48	84.98	7.21	0.90
SCO 1	9	5.22	73.89	4.50	0.50
SCO 2		5.79	84.61	6.64	0.74
SCO 3		20.06	83.68	6.72	0.75
SCO 4		26.81	86.34	8.00	0.89
SCO 1	10	5.00	74.98	4.68	0.47
SCO 2		5.69	84.87	6.76	0.68
SCO 3		18.59	84.87	7.25	0.72
SCO 4		24.93	87.30	8.58	0.86
SCO 1	11	5.04	74.78	4.63	0.42
SCO 2		5.08	86.51	7.41	0.67
SCO 3		17.18	86.02	8.00	0.73
SCO 4		22.97	88.30	9.34	0.85
SCO 1	12	5.16	74.18	4.57	0.38
SCO 2		4.63	87.70	8.23	0.69
SCO 3		16.25	86.77	8.57	0.71
SCO 4		21.02	89.29	10.33	0.86
SCO 1	13	4.98	75.06	4.71	0.36
SCO 2		4.48	88.10	8.56	0.66
SCO 3		14.82	87.94	9.40	0.72
SCO 4		19.50	90.06	11.00	0.85
SCO 1	14	4.78	76.09	4.90	0.35
SCO 2		4.28	88.63	8.96	0.64
SCO 3		13.14	89.30	10.60	0.76
SCO 4		18.82	90.41	11.40	0.81

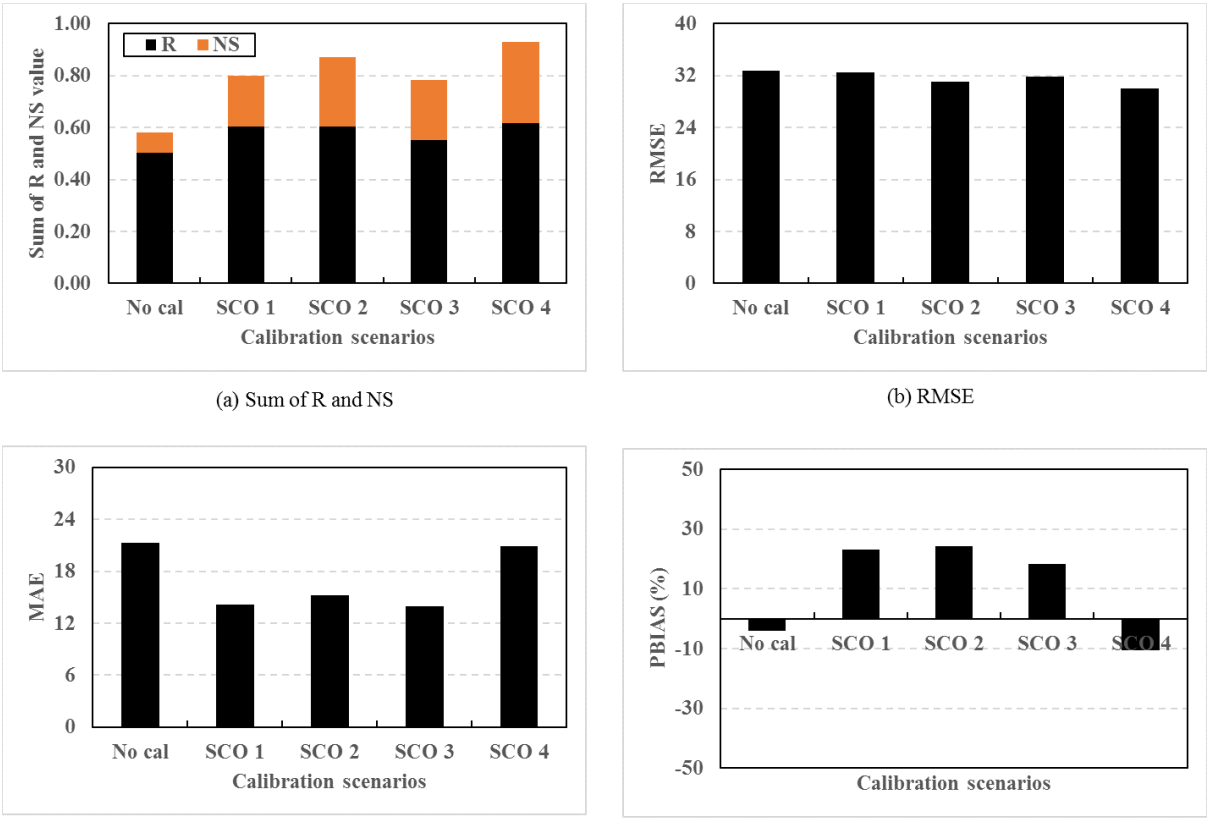
5.4.2. Model performance by calibration scenarios

In addition to computer parallelism aspect, HSPF performances were reported to evaluate how streamflow realization was well made associated with historical data at the mouth of the study area (Calibration target point 6). **Figure 5.9** and **Table 5.6** show the comparison of the model performance criteria for four calibration scenarios. As expected, HSPF performances after calibration improved against no calibration option. SCO 1 and SCO 2 were first compared to see how the different number of parameters affected the model performance. The result showed that NS of SCO 2 was higher than that of SCO 1, but RMSE and PBIAS of SCO 2 were lower than SCO 1, while R of SCO 1 and SCO 2 remained the same. It seems that SCO 2 was more affected by the volume variation of streamflow driven by more parameters used.

Overall, SCO 4 was the best with higher NS and R, and lower RMSE and PBIAS than any other scenarios. Even though MAE indicated higher value than other calibration scenarios, SCO 4 was selected for further calibration analysis because R, NS, and PBIAS are the most important performance measures to calibrate interior calibration target points (1-5) within the BRW.

Table 5.6. The model performance comparisons of calibrated streamflow by calibration scenarios.

Statistics	No Cal	SCO 1	SCO 2	SCO 3	SCO 4
Mean streamflow (m^3/sec)	27.09	24.89	36.36	26.82	36.36
R	0.50	0.60	0.60	0.55	0.62
NS	0.08	0.20	0.27	0.23	0.31
RMSE	32.79	32.53	31.07	31.80	30.03
MAE	21.31	14.13	15.22	13.96	20.94
PBIAS (%)	-4.17	23.04	24.25	18.38	-10.65



(a) Sum of R and NS

(b) RMSE

(c) MAE

(d) PBIAS

Figure 5.9. Performance measures, including Correlation coefficient (R), Nash-Sutcliffe coefficient (NS), Root mean square error (RMSE), Mean Square Error (MAE), and Percentage of Bias (PBIAS) from HSPF simulations.

5.4.3. Results of calibrated and validated streamflow using SCO 4

Table 5.8 indicates the statistical results of calibration and validation for all calibration target points (CTPs) using HSPF model. **Figure 5.10 - 5.12** show hydrograph and scatter plots comparisons for the calibrated and validated streamflow results based on SCO 4. The results indicate adequate calibration and validation performance over the simulation period. The timing of peak flows and the magnitude of peaks well matched between the simulated and observed flows at CTPs 1, 2, 3 and 4 during the calibration period. However, the magnitude of peaks at CTPs 5 and 6 showed somewhat different results due to reservoir diversion nearby. NS values between the simulated and observed streamflow at daily time step during calibration and validation are reported 0.36 to 0.78 and 0.33 to 0.71, respectively, while R values are 0.68 to 0.89 and 0.69 to 0.85.

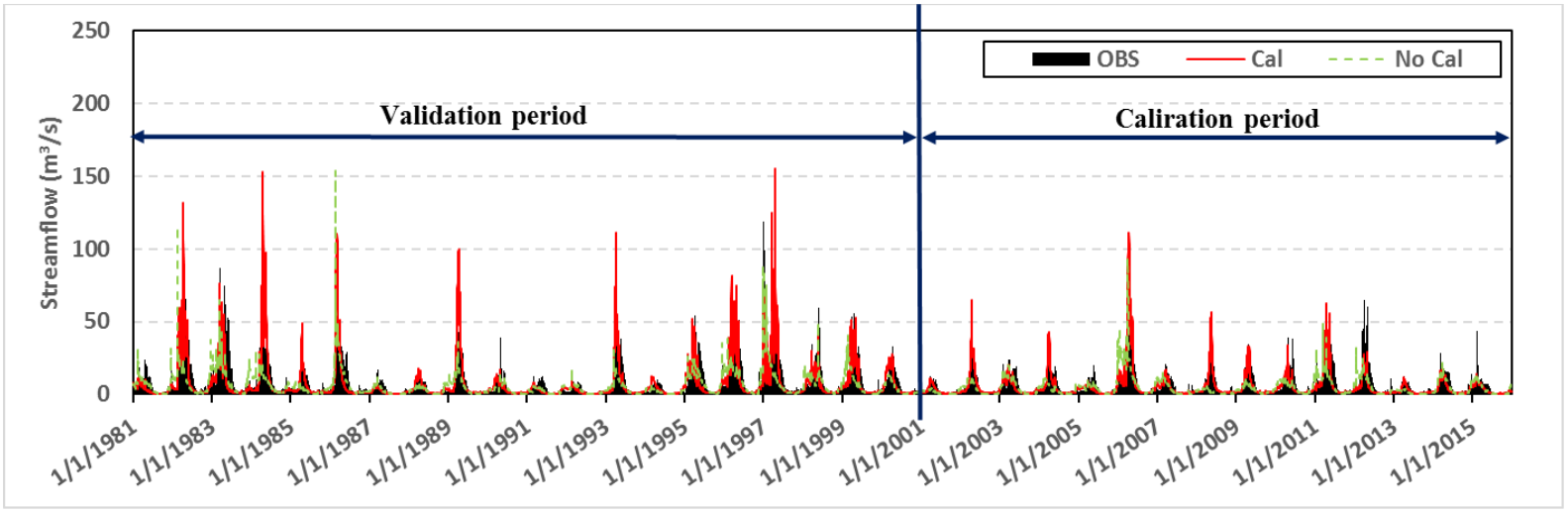
Based on **Table 5.7 and 5.8**, PBIAS of TPs 3, 4, and 6 indicated very good model performances with ranges from 5.49 % to 8.64 %, and that of CTPs 1 and 2 were good model performance with 11.11% to 14.94% during calibration. During the validation period, PBIAS of CTPs 1, 3, and 4 were very good performance with a range from -2.75% to 9.00% and CTP 2 was fair performance with 17.09%. PBIAS of calibration CTPs 5 and 6, however, were poor performance because CTP 5 was located below the large reservoir. Note that no reservoir rule curve and irrigation return flows is incorporated into HSPF model because the main objective of this research focuses on computer parallelism.

Table 5.7. General calibration targets of HSPF application (Donigian, 2000).

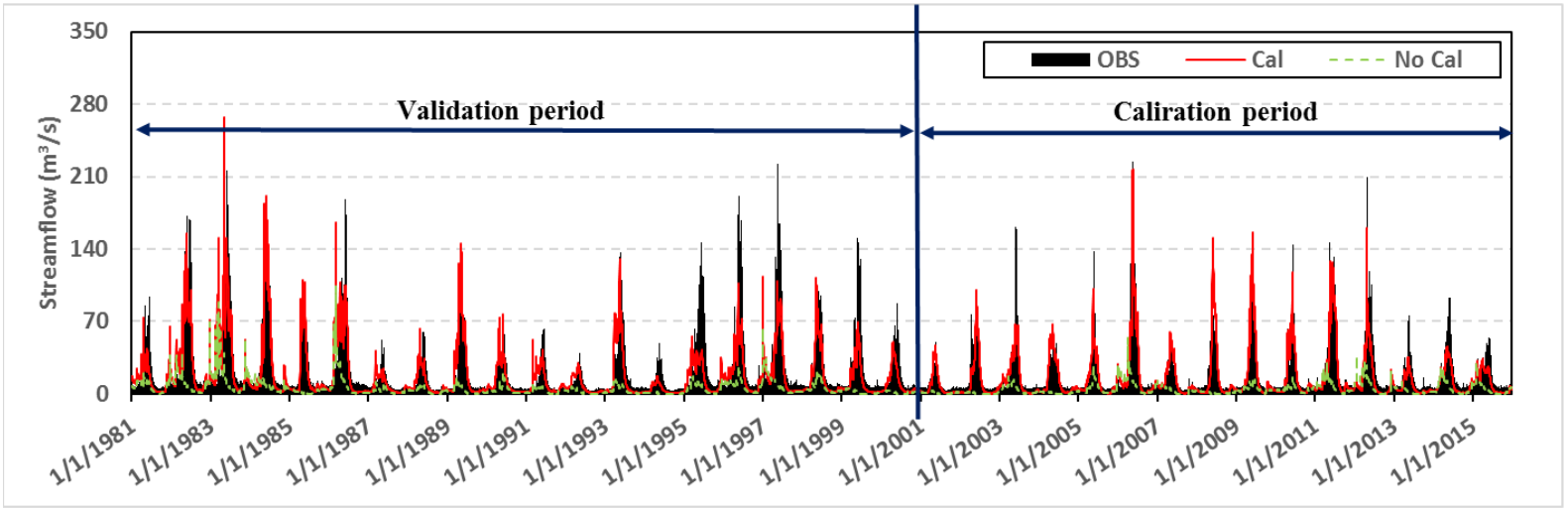
Component	PBIAS (%)			
	Very Good	Good	Fair	Poor
Hydrology/Flow	< ±10	±10 - ±15	±15 - ±25	> ±25
Sediment	< ±20	±20 - ±30	±30 - ±45	> ±45
Water Temperature	< ±7	±8 - ±12	±13 - ±18	> ±18
Water Quality/Nutrients	< ±15	±15 - ±25	±25 - ±35	> ±35

Table 5.8. The results of model performance for the calibrated and validated streamflow at calibration target points in the BRW.

Calibration target points (CTPs)		Mean streamflow (m^3/s)	Evaluation statistic				
			R	NS	RMSE	MAE	PBIAS (%)
1	Cal	5.01 (5.89)	0.87	0.74	4.28	2.17	14.94
	Val	7.29 (7.99)	0.84	0.55	7.63	3.68	8.76
2	Cal	15.17 (17.06)	0.88	0.76	11.27	6.67	11.11
	Val	17.46 (21.06)	0.85	0.71	16.00	8.86	17.09
3	Cal	28.86 (30.98)	0.89	0.78	17.16	9.63	6.86
	Val	34.98 (34.05)	0.87	0.69	22.36	12.14	-2.75
4	Cal	20.45 (21.64)	0.73	0.46	15.57	8.61	5.49
	Val	24.94 (27.41)	0.69	0.33	22.01	13.24	9.00
5	Cal	44.46 (61.94)	0.74	0.45	43.61	31.67	28.21
	Val	54.59 (78.27)	0.75	0.45	53.19	40.07	30.25
6	Cal	28.25 (30.92)	0.62	0.36	27.29	17.28	8.64
	Val	35.39 (49.17)	0.70	0.43	41.44	26.44	28.03

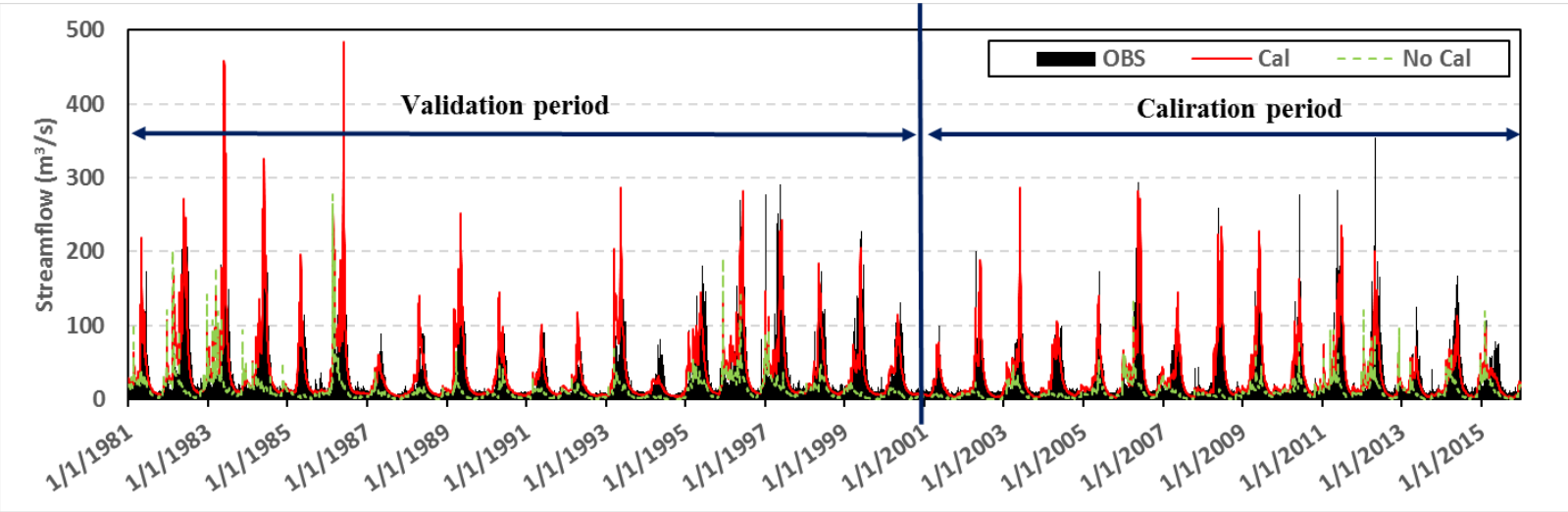


(a) Calibration target point 1

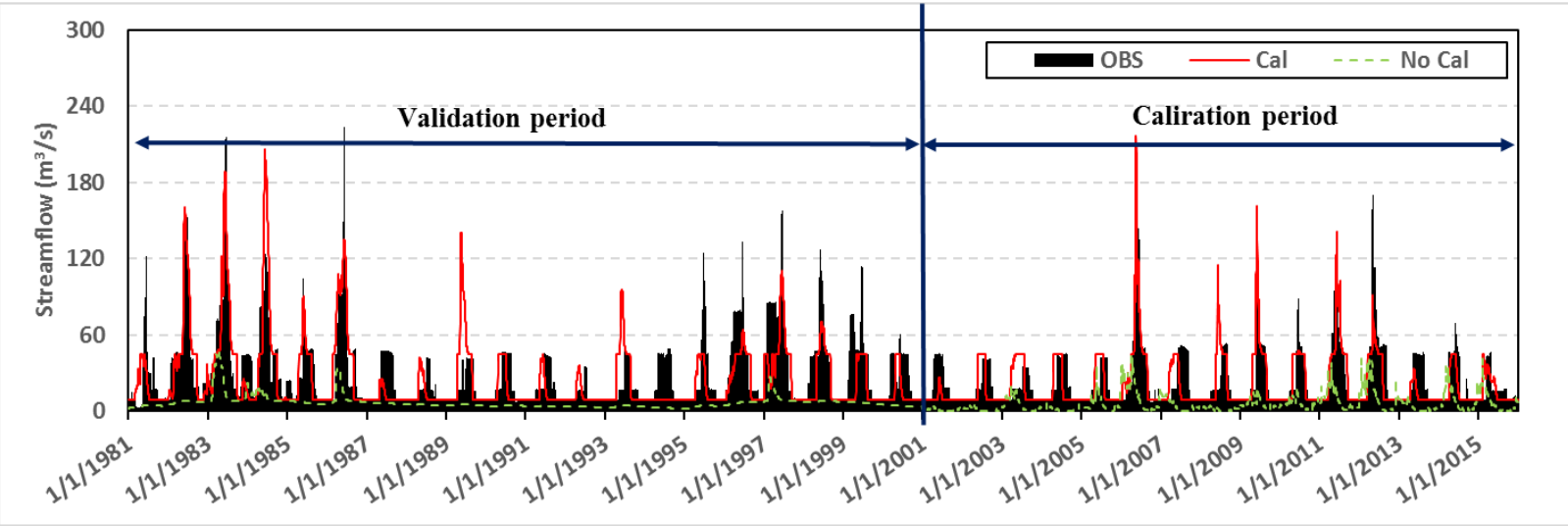


(b) Calibration target point 2

Figure 5.10. Hydrological simulations from HSPF at calibration target points (1-6) in the BRW.

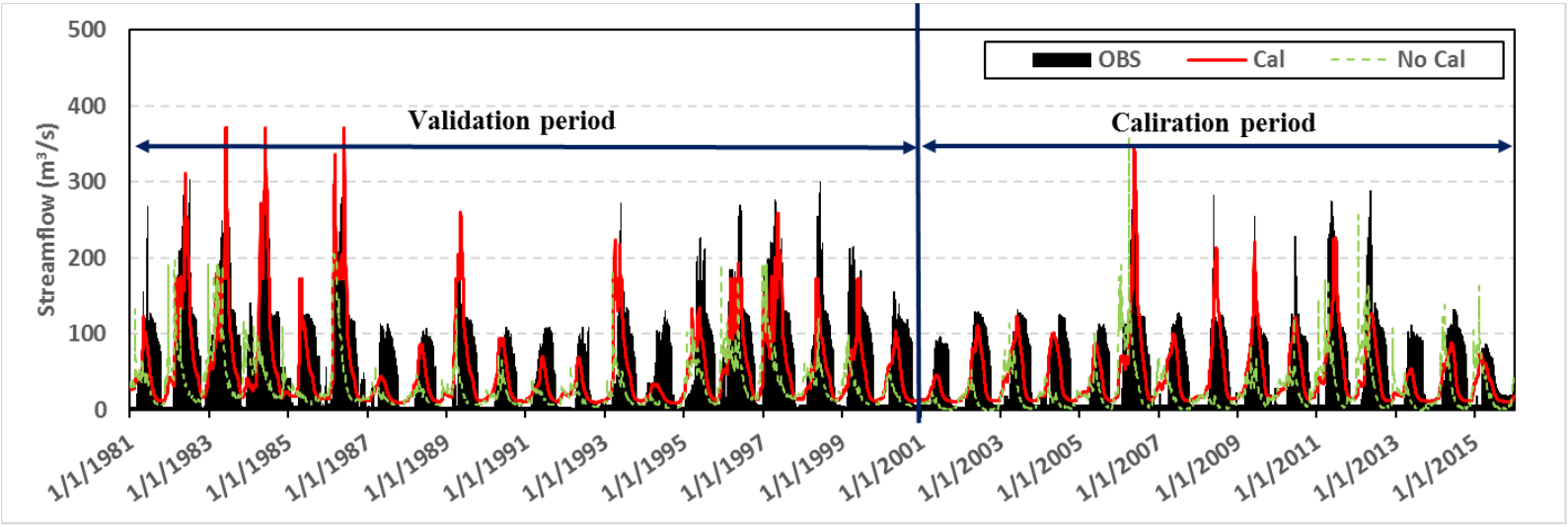


(c) Calibration target point 3

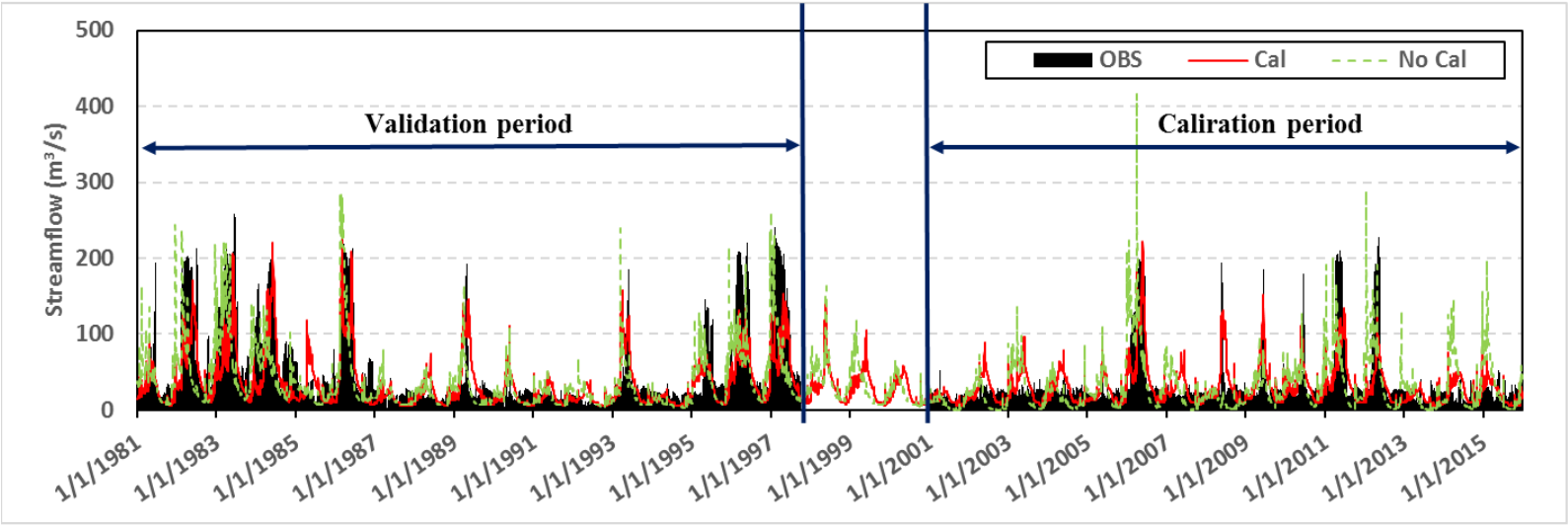


(d) Calibration target point 4

Figure 5.10. Hydrological simulations from HSPF at calibration target points (1-6) in the BRW (Cont.).



(e) Calibration target point 5



(f) Calibration target point 6

Figure 5.10. Hydrological simulations from HSPF at calibration target points (1-6) in the BRW (Cont.).

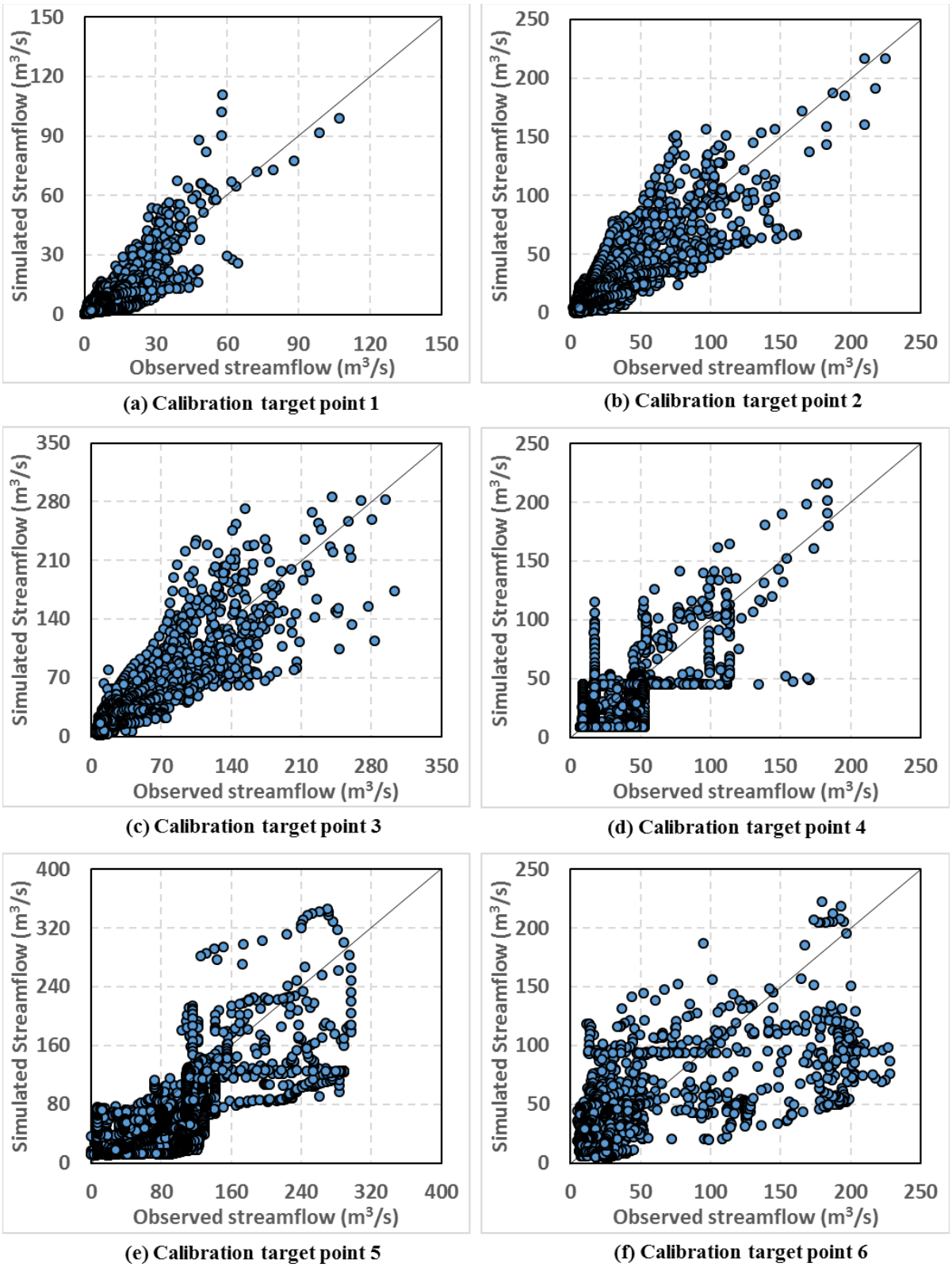


Figure 5.11. Scatter plots of calibration results for calibration target points (1-6) in the BRW.

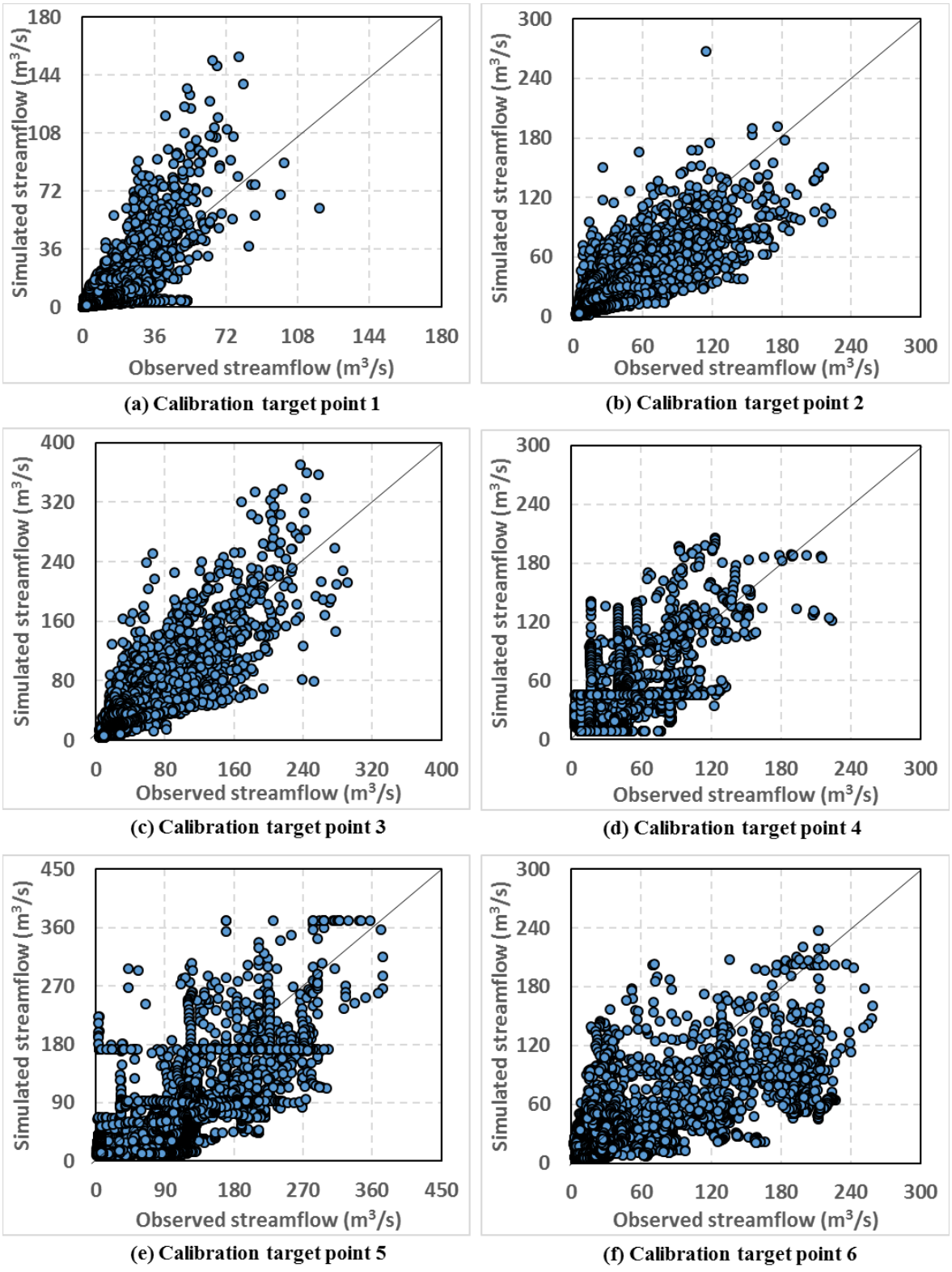


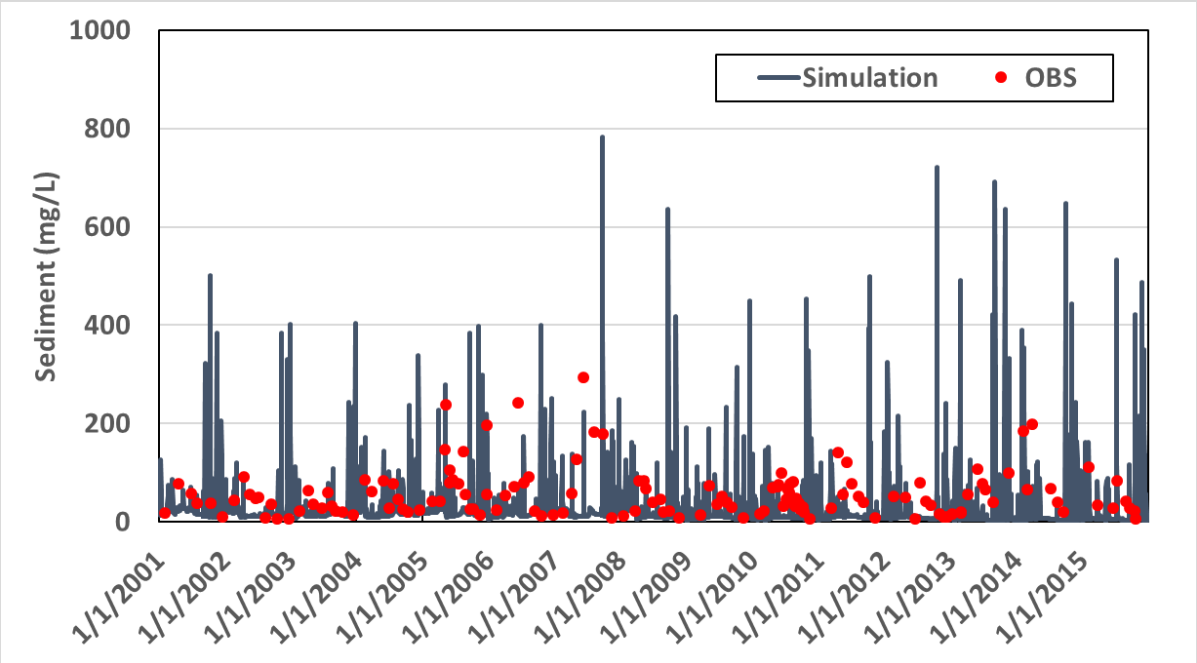
Figure 5.12. Scatter plots of validation results for calibration target points (1-6) in the BRW.

5.4.4. Results of calibrated sediment and water quality using SCO 4

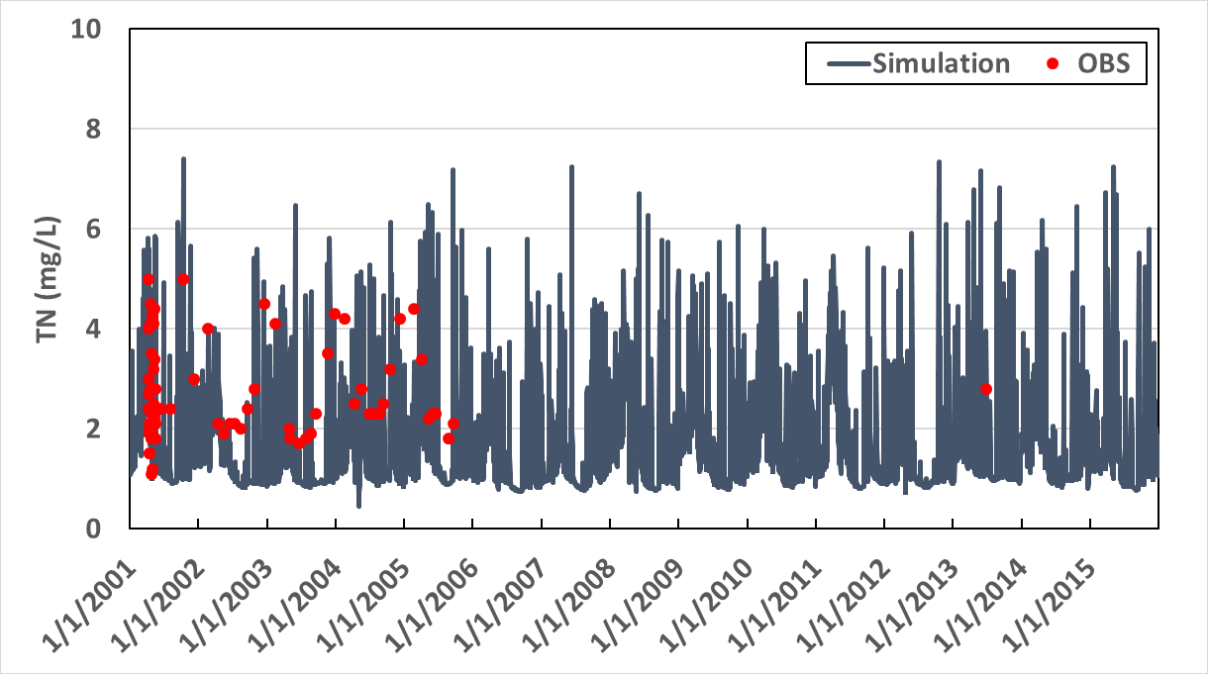
For water quality, no validation was applied due to the insufficient observed datasets. Model parameters of water quality were calibrated depending on available observed daily data. **Table 5.9** and **Figure 2.13** show the model performances for sediment, TN, TP, DO, and water temperature. Based on the model evaluation target by **Table 5.7**, water temperature was very good performance by 6.24%. Sediment, TN, TP, and DO were good performance with a range from 30.05%, 15.10%, 12.19%, and -9.88%, respectively. Therefore, HSPF model simulated the reliable model outputs for water quality (sediment, TN, TP, DO, and water temperature) and well represented hydrologic conditions in the BRW based on model performance statistics.

Table 5.9. Model performance of calibrated water quality.

Evaluation statistic	Sediment	Water Temp	TN	TP	DO
R	0.75	0.96	0.82	0.79	0.94
NS	0.42	0.63	0.22	0.86	0.61
RMSE	445.37	107.79	5.45	0.30	44.98
MAE	23.88	5.22	2.94	0.02	1.01
PBAIS (%)	30.05	6.24	15.10	12.19	-9.88

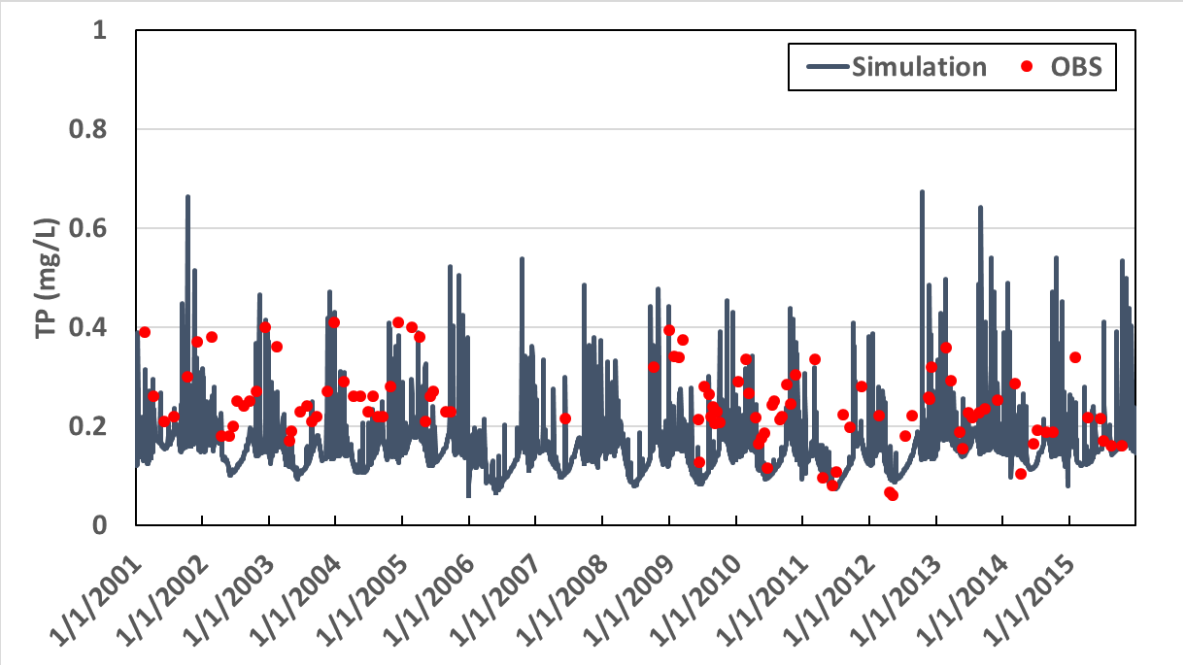


(a) Sediment

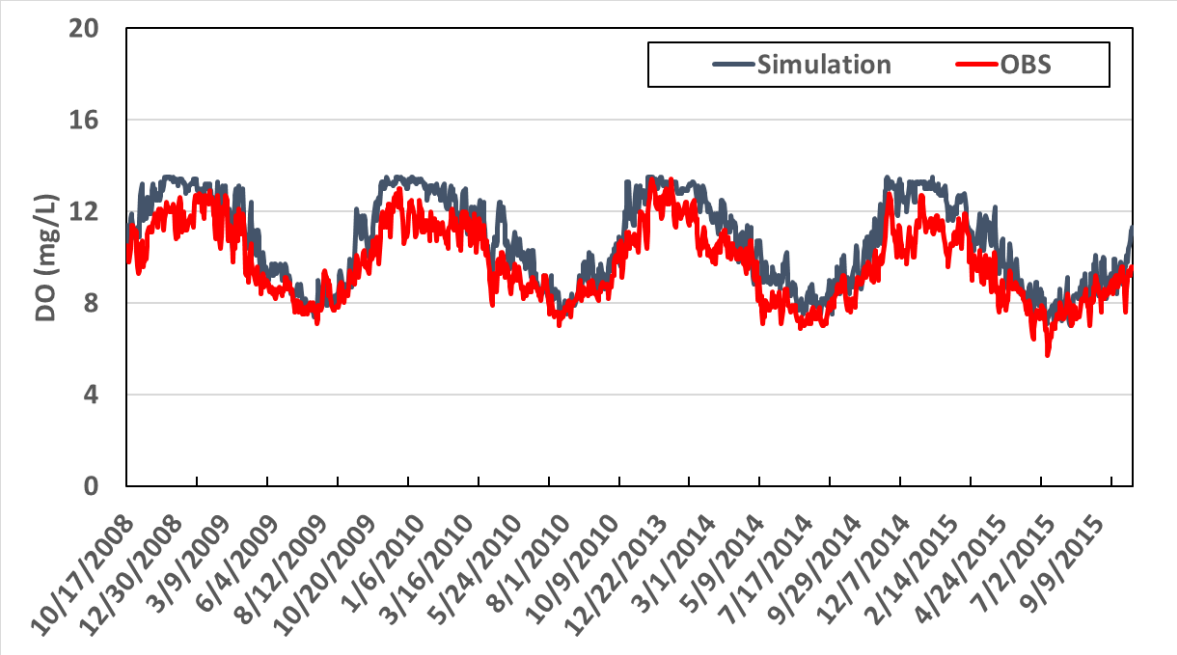


(b) TN

Figure 5.13. Hydrographic of calibrated daily sediment (a), TN (b), TP (c), DO (d), and water temperature (f).

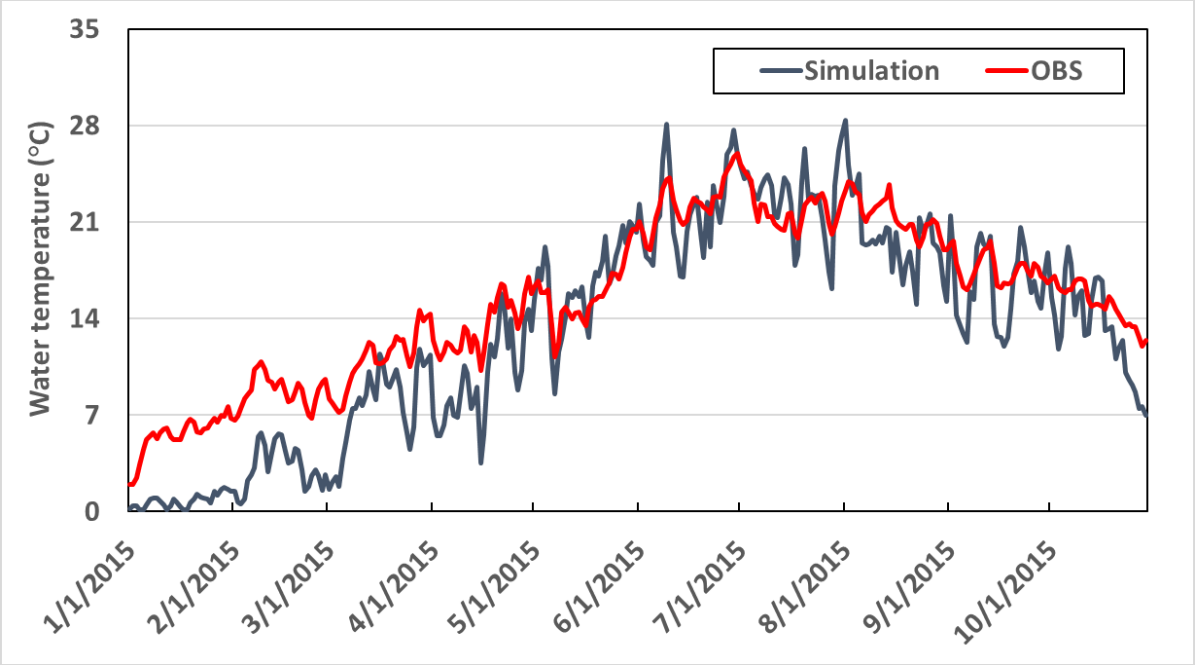


(c) TP



(d) DO

Figure 5.13. Hydrographic of calibrated daily sediment (a), TN (b), TP (c), DO (d), and water temperature (f) (Cont.).



(e) Water temperature

Figure 5.13. Hydrographic of calibrated daily sediment (a), TN (b), TP (c), DO (d), and water temperature (f) (Cont.).

CHAPTER 6. IMPACTS OF CLIMATE AND LAND USE CHANGES ON STREAMFLOW AND WATER QUALITY

6.1. Introduction

Changes in climate and LULC may affect water quality and quantity at the urban-rural interface around the world. Urbanization with land use changes is one of the most significant environmental factors. It can has negative impacts on ecological and hydrologic condition in many waterways due to the increased impervious land segments (e.g. industrial, commercial, and residential land use). The impervious surfaces can accelerate the increase in Nonpoint source (NPS) load, runoff, and sedimentation and the infiltration reduction (Childers et al., 2014; US.EPA, 2000; Wei et al., 2015). Global warming and climate change alter general patterns of air temperature and precipitation. The elevated air temperature in spring and winter results in early snowmelt. It can also affect water quality degradation along with surface water depletion in late spring and summer so that water managers may face the water resources management problems through the year. The variation in future water quality and quantity induced by climate change requires proactive management strategies to mitigate their impacts. Several studies have been conducted to investigate hydrological processes driven by LULC change (Brath et al., 2006; Guo et al., 2008; Lee and Chung, 2007) or climate change (Chen et al., 2011; Prowse et al., 2006; Sellami et al., 2016). Although the combined impacts of climate and LULC change will accelerate the negative hydrologic and ecological effects, few studies have been conducted to investigate how urbanization and climate change can affect water quality and quantity in the urban-rural interface, such as the BRW. Therefore, the evaluation of LULC change coupled with climate variability is a major research topic to

advance water resources planning and management in this chapter.

To understand the effects of LULC change on hydrology, it is imperative to have explicit LULC information (Strayer et al., 2003) in modest spatial resolution. For example, land developments (e.g. barren, housing, and commercial development) will change flow patterns such as timing and peak flows within the watershed. LULC change projections gives useful information for decision makers to visualize future LULC variation, optimize management practices, and improve LULC planning by mitigating the negative consequences of land development (Heistermann et al., 2006). Hydrologic models that incorporate spatial information data can be used to predict the variations of hydrologic dynamics driven by LULC change. Several LULC modeling frameworks have been developed to investigate potential LULC change for the United States (US), but data on LULC scenarios have few constraints. Thus, these data mostly have a small number of land use classification (e.g. urban, forest) and limit our ability to characterize uncertainties associated with future land development scenarios in coarse spatial resolution. Therefore, we use thematically detailed future LULC change products provided by USGS to improve modeling accuracy based on four qualitative and quantitative emission scenarios listed in the Intergovernmental Panel on Climate Change (IPCC) Special Report (Sohl et al., 2014). Note that these LULC products have an appropriate spatial resolution (e.g. 250 m) and various LULC classes.

Climate change is traditionally assessed using climate model simulations, such as those from the Coupled Model Inter-comparison Project Phase 5 (CMIP 5). There are many GCMs available from CMIP5, but several studies have discussed the uncertainty in the selection of climate models which cannot be avoided (Eisner et al., 2012; Jung et al., 2011; Kay et al., 2009; Kingston et al., 2010,). It is necessary to evaluate GCMs that better represent

local or regional climatology because GCMs itself cannot illustrate hydro stats at a particular watershed due to coarse spatial resolution (e.g. 100-500 km). Therefore, statistical or dynamical downscaling methods are typically used to transfer coarse resolution outputs from GCMs to the hydro model-friendly outputs in terms of adequate spatial and temporal scales (e.g, 12 km x 12 km and hourly time step or finer). Xu et al. (2005) proposed three steps to perform the impact of climate change at catchment scales. The first step is to select the different climate projections from GCMs using different emission scenarios. The second step is to apply appropriate downscaling methods to better represent local climatology from the GCMs outcomes. Lastly, the downscaled GCMs is applied to the hydrologic models for water availability in the study area over next few decades. Burger et al. (2012) reported that the downscaling methods show different performances in reproducing climate extremes. Therefore, it may need to find an appropriate downscaling method to better address local water issues on both quality and quantity simulations using GCM's inputs.

In this chapter, therefore, we investigate the impacts of changes in LULC and climate in the BRW. **Figure 6.1** shows a flowchart for the application of model input data associated with urbanization, climate change, and the combined impact of LULC and climate changes. Future climate series and LULC change projections were applied to examine the separate and combined impacts of climate and LULC changes on water quality and quantity using the optimized HSPF model.

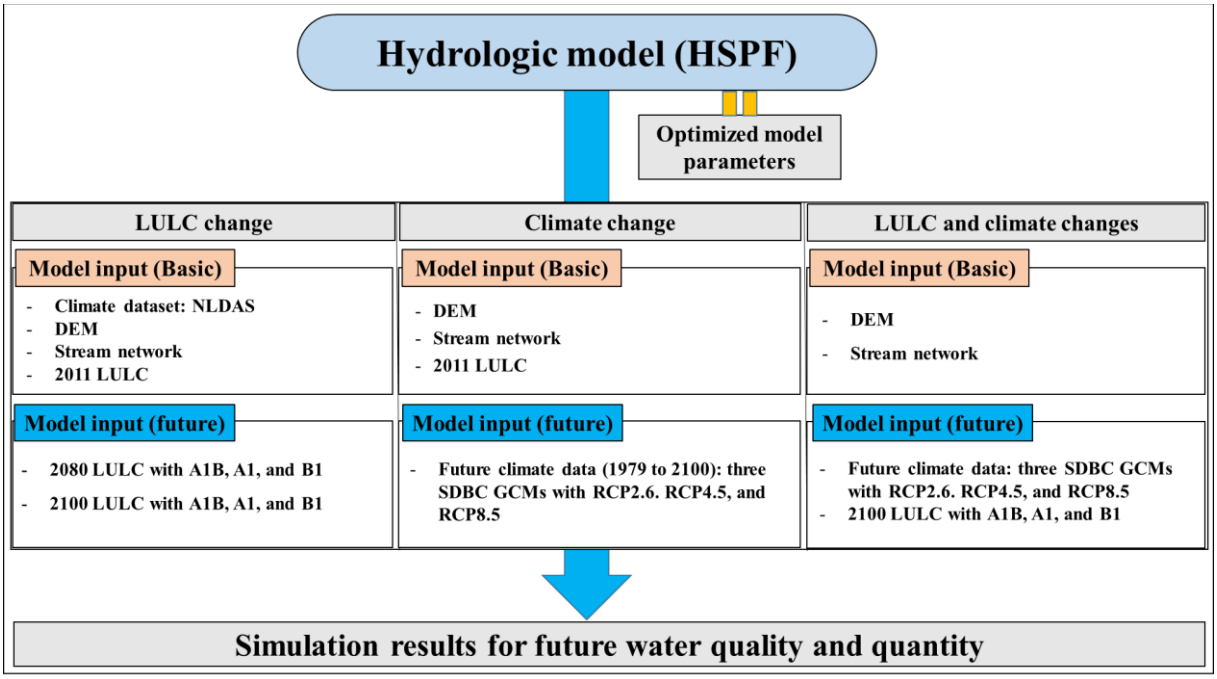


Figure 6.1. The flowchart for application of future LULC, climate change, and the combined impact of climate and LULC changes using HSPF model.

6.2. Materials and methods

6.2.1. Future Land use and land cover (LULC)

Future LULC projections were developed by USGS using the forecast scenarios of land use change (FORE-SCE) model based on IPCC SERS scenarios across the United States for the years 1992-2100 (Sohl et al., 2014). They have 250 meter by 250 meter high spatial resolution and consist of 12 LULC categories as 2011 LULC product. IPCC SRES scenarios have storylines (e.g. A1B, A2, and B1) based on global/regional economic, technological, and environmental cooperation, and economic growth. **Table 6.1** shows the detail information about IPCC SRES scenarios applied future LULC products. **Figure 6.2** indicates a schematic of the LULC project structure with the major data components. LULC modeling was individually simulated for each of the 84 level III ecoregions as mapped based on ecoregion publication (US.EPA, 1995). FORE-SCE model was applied for the spatial LULC modeling

(Sohl et al., 2012a and 2012b). FORE-SEC utilized a modular approach as originally developed by the Conversion of Land Use and its Effects (CLUE) series of LULC modeling to explain the both bottom-up and top-down drivers of change (Verburg et al., 1999; Verburg et al., 2008). It produced spatially explicit LULC using the spatial allocation component. For a non-spatial demand component, it creates future proportions of LULC variations at an aggregated regional level.

In this chapter, 2011 LULC provided by USGS NLCD database was used to simulate current hydrologic condition for water quality and streamflow. The 2080 and 2010 LULC products with IPCC-SRES storylines (A1B, A2, and B1) were used to analyze future hydrologic conditions. Future LULC products were reclassified by seven land use classification such as “Urban”, “Barren / Mining”, “Cropland”, “Forest”, “Grassland”, “Shrubland”, and “Water / Wetland” to apply into HSPF model.

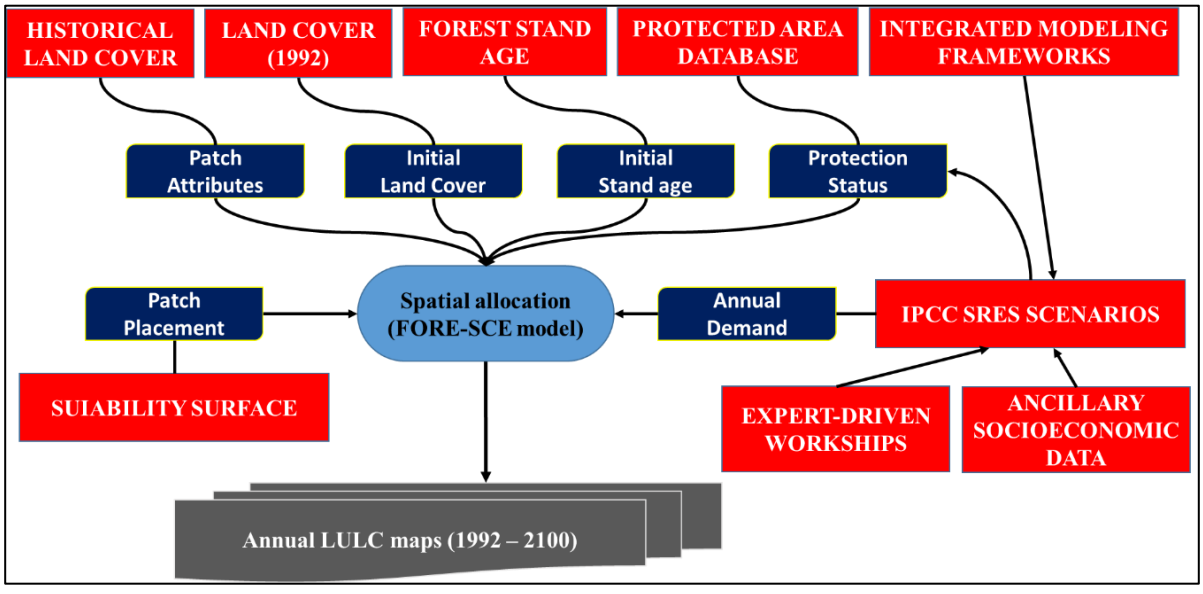


Figure 6.2. Schematic of general future LULC modeling based on IPCC SRES from 1992 to 2100 (Sohl et al., 2014).

Table 6.1. Information of IPCC SRES scenarios (Sohl et al., 2012a).

Scenarios	Description
A1B	- Very rapid economic growth. - Peak global population in mid-century and then declining. - The rapid introduction of new and more efficient technologies. - Substantial reduction in regional differences in per capital income for convergence among regions, capacity building, and increased cultural and social interactions.
A2	- Very heterogeneous world. - Self-reliance and preservation of local identities. - Fertility patterns across regions cover slowly. - Continuously increasing the global population. - Primarily regionally oriented economic development. - More fragmented technological change. - Slower technological change than other storylines.
B1	- A convergent world with the same global population. - Peak global population in mid-century and then declining. - Rapid changes in economic structures and information economy with reduction in material intensity - The introduction of clean and resource-efficient technologies - Improved equity without additional climate initiatives for global solutions (e.g. economic, social, and environmental sustainability)

6.2.2. Global Climate Models (GCMs) from the fifth phase of the Climate Model Intercomparison Project (CMIP5)

GCMs produce the Earth's climate using mathematic equations associated with atmospheric, oceanic, biotic processes, interactions, and feedback. It is the primary tools to provide reasonably reliable global-, hemispheric, and continental-scale climate information. GCMs are used to understand future climate scenarios based on the increased greenhouse gas concentrations. In general, GCMs typically include datasets in coarse spatial resolution (e.g. 100-500 km).

GCM simulations of the 20^{th} and 21^{st} centuries have been conducted with several models through the CMIP5 which was used in the IPCC Fifth Assessment Report (IPCC, 2013). Future scenarios that consider future rates of technological development, social responses, and population growth, and collective mitigation efforts are summarized through representative concentration pathways (RCPs), including high emission scenarios (RCP8.5), midrange mitigation emission scenarios (RCP4.5), and low emission scenarios (RCP2.6). The RCP labels such as RCP8.5 represent the estimation of the radiative forcing in 2100. For example, RCP8.5 shows the increasing of the radiative forcing throughout 21century before it reaches a level about 8.5 W/m^2 at the end of the century.

Figure 6.3 shows the flowchart of creating future climate data for HSPF input using GCMs. Total 27 GCMs were applied to select the best GCMs which better represents historical climate condition at the study area (**Table 6.2**). Seasonal distribution of precipitation and temperature were utilized to analyze the performance of raw GCMs in the BRW from Jan 1981 to Dec 2005. Three GCMs were selected to apply Statistical Downscaling and Bias Correction (SDBC) and temporal downscaling; Two GCMs were

selected from higher ranked models using skill scores and RMSE. One GCM was selected from Rupp et al (2013) study which reported the highest model performance for Pacific Northwest area. After that, SDBC GCMs were imported to create WDM file as model input using method shown in **Figure 5.5**. Lastly, the performance of SDBC GCMs was evaluated comparing with NLDAS-2. More detail information about the GCM selections and downscaling methods indicates as below section.

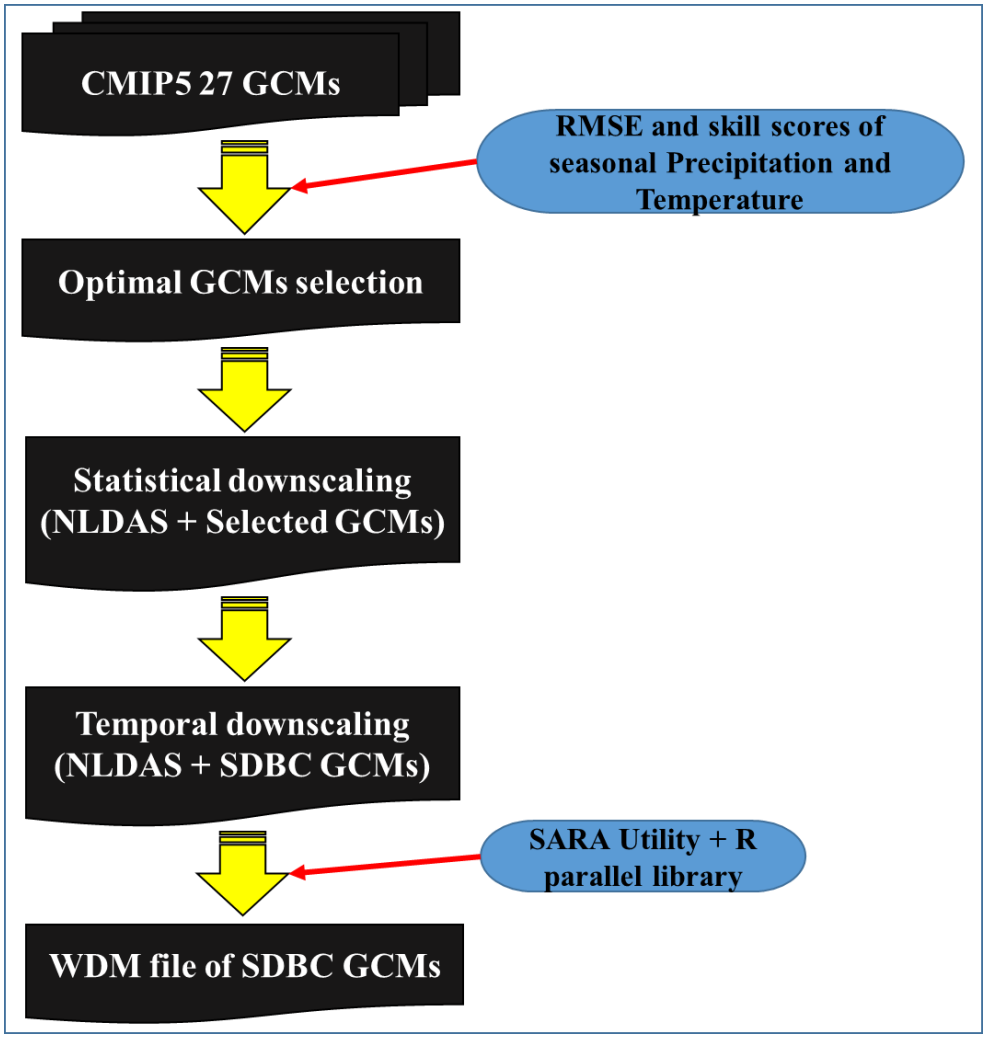


Figure 6.3. The flowchart of creating HSPF model weather data for future climate using optimal GCMs.

Table 6.2. The information of total GCMs.

Numb	GCM name	Institution	Spatial resolution (degree)		Ensemble
			Longitude	Latitude	
1	bcc-csm1-1	Beijing Climate Center, China Meteorological Administration	2.8125	2.7906	1
2	bcc-csm1-1-m		1.1250	1.1215	1
3	canesm2	Canadian Centre for Climate Modeling and Analysis	2.8125	2.7906	4
4	ccsm4	National Center for Atmospheric Research	1.2500	0.9424	1
5	cesm1-bgc	National Science Foundation, Department of Energy, National Center for Atmospheric Research	1.2500	0.9424	1
6	cesm1-cam5		1.2500	0.9424	1
7	cesm1-fastchem		1.2500	0.9424	1
8	cmcc-cm	Centro Euro-Mediterraneo per I Cambiamenti Climatici	1.8750	1.8653	1
9	cnrm-cm5	Centre National de Recherché Meteorologiques / Centre European de Recherché et Formation Avancees en Calcul Scientifique	1.4063	1.4008	5
10	gfdl-cm3	Geophysical Fluid Dynamics Laboratory	2.5000	2.0000	1
11	gfdl-esm2g		2.5000	2.0225	1
12	gfdl-esm2m		2.5000	2.0225	1
13	giss-e2-h-cc	NASA Goddard Institute for Space Studies	2.5000	2.0000	1
14	giss-e2-r		2.5000	2.0000	5
15	giss-e2-r-cc		2.5000	2.0000	1
16	hadgem2-ao	National Institute of Meteorological Research/Korea Meteorological Administration	1.8750	1.2500	1
17	inmcm4	Institute for Numerical Mathematics	2.0000	1.5000	1
18	ipsl-cm5a-lr	Institute Pierre-Simon Laplace	3.7500	1.8974	1
19	ipsl-cm5a-mr		2.5000	1.2676	1
20	ipsl-cm5b-lr		3.7500	1.8974	1
21	miroc4h	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	0.5625	0.5616	1
22	miroc5		1.4063	1.4008	1
23	miroc-esm		2.8125	2.7906	1
24	miroc-esm-chem		2.8125	2.7906	1
25	mpi-esm-mr	Max Planck Institute for Meteorology (MPI-M)	1.8750	1.8653	1
26	mri-cgcm3	Meteorological Research Institute	1.1250	1.1215	1
27	mri-esm1		1.1250	1.1215	1

6.2.2.1. Optimal GCM selections

Several GCMs are available to represent future climate condition, but all GCMs do not necessarily reflect historical climate conditions (1981-2005) for the study area. Therefore, the best GCMs, which agree with the historical climate conditions, were selected to analyze future climate change in the BRW. RMSE and two skill score methods (Heidke Skill Score (HSS), Peirce Skill score (PSS)) were used to compute the accuracy of historical GCMs (retrospective forecast) based on the observed seasonal precipitation and temperature. The detail desecration and equation of RMSE are shown in equation (3.12) of **Chapter 3**. Positive skill score values for HSS and PSS indicate a minimal level of acceptable performance for GCMs results. Contingency tables consist of four part of seasonal precipitation (P) or temperature (T) (less than 20th percentiles, between 20th and 50th percentiles, between 50th and 80th percentiles, higher than 80th percentiles corresponding percentiles of seasonal precipitation or temperature distribution). HSS and PSS were calculated based on a count from contingency table. The HSS and PSS equation are given by:

$$PSS = \frac{\sum_{i=1}^{I} P(y_i, o_i) - \sum_{i=1}^{I} P(y_i)P(o_i)}{1 - \sum_{i=1}^{I} P(y_i)P(o_i)} \tag{6.1}$$

$$HSS = \frac{\sum_{i=1}^{I} P(y_i, o_i) - \sum_{i=1}^{I} P(y_i)P(o_i)}{1 - \sum_{i=1}^{I} P(y_i)P(o_i)} \tag{6.2}$$

where, $P(y_i, o_i)$ is the number of forecasts in category i that had observations in category i; $P(y_i)$ is the number of forecasts in category i; $P(o_i)$ is the number of observation in category i; I is the number of contingency tables.

The calculated skill scores and RMSE were normalized to examine optimal GCMs using as below equitation.

$$Z = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (6.3)$$

where, Z is a normalized value between 0 and 1; X_i is each data point I; X_{min} is the minimal value for all the data points; X_{max} is the maximum value for all the data points.

For hydrologic modeling, precipitation is more important variable than temperature. Therefore, we applied a weight factor to precipitation (0.7) and temperature (0.3) for the final GCM selections. Two GCMs with three different future climate sceneries (RCP2.5, RCP4.5, and RCP8.5) were selected by the higher normalized value calculated from seasonal precipitation and temperature in the BRW. Furthermore, we considered one more GCM from Rupp et al (2014) study. They evaluated CMIP5 20th century climate simulation for the Pacific Northwest US using observed data for the 20th century. The performance and credibility were assessed using correlation and variance of mean seasonal spatial patterns, amplitude of seasonal cycle, diurnal temperature range, annual- to decadal-scale variance, long-term persistence, and regional teleconnections to El Niño Southern Oscillation (ENSO). Cnrm-cm5 model indicated the highest model performance using the simple method. So, total three GCM models (two models from GCM selection method, one model from a reference paper) were selected to investigate the variations of water quality and quantity by climate change in the BRW.

6.2.2.2. Statistical downscaling

Raw GCMs cannot examine fine scale heterogeneity of climate change conditions and the impact assessment and planning of hydrologic conditions at local scales due to the coarse spatial resolution. They include the uncertainty of future climate change at local climate condition. Therefore, it is necessary downscaling process to simulate potential impacts on

crop production, hydrology, and species distribution and to establish a relationship between present day local climate and atmospheric condition at fine scales (e.g. 10 – 50 km). Spatial and temporal aspects of climate projections are performed in the downscaling process. Spatial downscaling is to derive finer resolution from raw GCM at a specific location. Temporal downscaling is to generate fine scale temporal information from coarse temporal GCM output (e.g. from monthly to daily time step). Several methods have been developed to apply downscaling associated with creating of fine spatial resolution and reduction of local climate uncertainties, but there is no official guidance to meet researchers and decision maker's requirements best. Therefore, downscaling methods have been developed to improve local climate uncertainties and perceive the impact of climate change studies using the improved technical and specialized explanations. Dynamical and statistical downscaling are the common methods to combine the information on local conditions with large-scale climate projections. Dynamical downscaling uses a regional climate model (RCM) to generate realistic climate information at a spatial resolution of approximately 20-50 km. Statistical downscaling utilizes the empirical relationship between historical or current large atmospheric and local variables. Statistical downscaling is commonly applied method using a relatively simple bias correction (Johnson and Sharma., 2011; Piani et al., 2010; Wood et al., 2004). Statistical downscaling assumes the statistical relationship between the predictor and predict and remains stable into the future. The advantages of the statistical downscaling method are computationally inexpensive and straightforward enough to adapt for a specific purpose than dynamical downscaling.

6.2.2.2.1. Statistical Downscaling and Bias correction (SDBC) method

The selected three GCMs are not appropriate to apply hydrologic model by themselves since they have coarse spatial resolution (e.g. 100km by 100 km) and climate uncertainties. Indeed, downscaled GCMs are mostly available for precipitation and temperature. Although the downscaled GCM data have the GCM climatology, additional efforts are required to derive climate patterns retrieved from NLDAS at finer scales. Wind speed and humidity data are not usually bias corrected using NLDAS-2.

In this study, for the selected three GCMs, SDBC method developed by Ahmed et al (2013) was applied to create finer spatial resolution (e.g. 12 km by 12 km spatial resolution) and to reduce local climate uncertainties for precipitation, temperature, wind speed, solar radiation, and humidity. SDBC can be appropriately utilized to apply spatial downscaling and bias correction for the amount of GCM data and is computationally efficient. This method is the mixed methods to be used sequentially and a modified method of the bias correction and spatial downscaling (BCSD) proposed by Wood et al (2004) and Maurer and Hidalgo (2007). BCSD has one limitation for the sensitivity of the finally downscaled results to the initial spatial resolution at which bias correction is applied. Ahmed et al (2013) analyzed that the resulting finer scale GCM data using SDBC provided more credible future facing for impact assessment which takes the future climate itself.

For the first step, the spatial downscaling was applied using a factor value [Equation (6.4) and (6.5)]. For example, precipitation and wind speed factors are the ratio of GCM and NLDAS-2 climate variables aggregated to the raw GCM resolution. Temperature, solar radiation, and humidity factors are the difference between GCM and NLDAS-2 climate variables aggregated to the raw GCM resolution.

$$F_v = G_v / N_v \text{ (for precipitation, wind speed)} \quad (6.4)$$

$$F_v = G_v - N_v \text{ (for temperature, solar radiation, humidity)} \quad (6.5)$$

where, F_v is factor value, G_v is the GCM output value, and N_v is the NLDAS value.

The downscaling of the factor values was then used synergraphic mapping system (SYMAP) algorithm developed by Shepard (1968). Combining the downscaled factor values with NLDAS-2 (e.g. multiplication for precipitation and wind speed, and addition for temperature, solar radiation, and humidity) created a downscaled GCM datasets. The downscaled GCM derives finer scale pattern from NLDAS-2. After the spatial downscaling, bias correction was applied using a quantile-based mapping of the probability density functions (synthetic gamma distribution) as the last step to correct uncertainties of GCMs. The bias correction assumes the biases represents the same pattern in both present and future climate simulations. It was based on the comparison between Cumulative Distribution Function (CDF) for NLDAS-2 and GCM data within the same time periods. For future climate condition, it assumed that the difference between the GCM and NLDAS-2 during the reference period also applied to the future period that means the adjustment function is the same time. However, the difference between CDFs for the historical and future period was taken into account. CDF was calculated based on the month-specific probability distribution for monthly GCM and NLDAS-2 data of precipitation, temperature, solar radiation, humidity, and wind speed. Bias correction was applied using Equation (6.6) for historical GCMs and Equation (6.7) developed by Li et al. (2010) for future GCMs. The inverse CDF of the gamma function was used to apply bias correction for GCMs from NLDAS-2. For example, suppose that historical GCM has the 230 mm precipitation in January 2005 and the CDF value is recorded about 99.45% probability in gamma domain. Precipitation of NLDAS-2

corresponding to the same CDF indicates as 210 mm having about 99.45% probability in gamma domain. The precipitation value of GCM is then replaced by 210 mm having the same probability of 99.45% in its own gamma domain for the historical condition. Also, future GCM data has the 130 mm precipitation in January 2090 and the CDF value is recorded about 98.45% probability in gamma domain. Precipitation of NLDAS-2 and historical GCM corresponding to the 98.45% CDF indicates as 180 mm and 140 mm, respectively. The difference between NLDAS-2 and GCM under the current climate (180mm – 140mm) is summed from the future GCM (130 mm) to get the bias corrected value. So, the precipitation value of future GCM is replaced by 170 mm.

$$\bar{x}_{g-c} = F_{o-c}^{-1}\left(F_{g-c}\left(x_{g-c}\right)\right) \tag{6.6}$$

$$\bar{x}_{g-f} = x_{g-f} + F_{o-c}^{-1}\left(F_{g-f}\left(x_{g-f}\right)\right) - F_{g-c}^{-1}\left(F_{g-f}\left(x_{g-f}\right)\right) \tag{6.7}$$

where, F is the CDF of either the observation (o) or GCM (g) for a historical period or current climate (c) or future period (f), F^{-1} is the inverse CDF, x is the climate variation of precipitation and temperature, and $\bar{x}$ is the bias corrected climate variation.

6.2.2.2.2. Temporal downscaling

Temporal downscaling is required to disaggregate SDBC GCMs from monthly to daily data, which is used as model input. Therefore, simple disaggregation method developed by Acharya and Ryu (2014) was applied for three SDBC GCMs. As shown in **Figure 6.4**, the first step in disaggregation process is the selection of target grid (TG), which is SDBC GCM, and source grid (SG), which is NLDAS-2. The SG has both daily and monthly historical climate data, whereas the TG has only monthly climate data. Basically, the daily climate

variables at TG is calculated based on monthly climate variables at TG and daily and monthly climate variables at SG. Once a month and year are for climate variable disaggregation at TG (e.g. 3-month window centering the selected month at TG is chosen from historical NLDAS-2 monthly climate variables at SG). For example, climate variable disaggregation in February at TG requires information for 3-month windows including January, February, and March over historical NLDAS-2 monthly climate variables at SG. Therefore, total monthly precipitation for the selected window at TG is compared with the same window for every year in NLDAS-2 at SG. After that, RMSE is calculated to identify the best match year and months representing monthly total precipitation between TG and SG during the predefined three monthly. Daily precipitation index (DPI) is then calculated as follow. DPI is defined as the ratio of daily precipitation to monthly total precipitation for the selected month at SG. Using DPI at SG and monthly total SDBC GCMs at TG, the disaggregation was simulated for precipitation, temperature, wind speed, solar radiation, and humidity from 1979 to 2100 for SDBC GCMs.

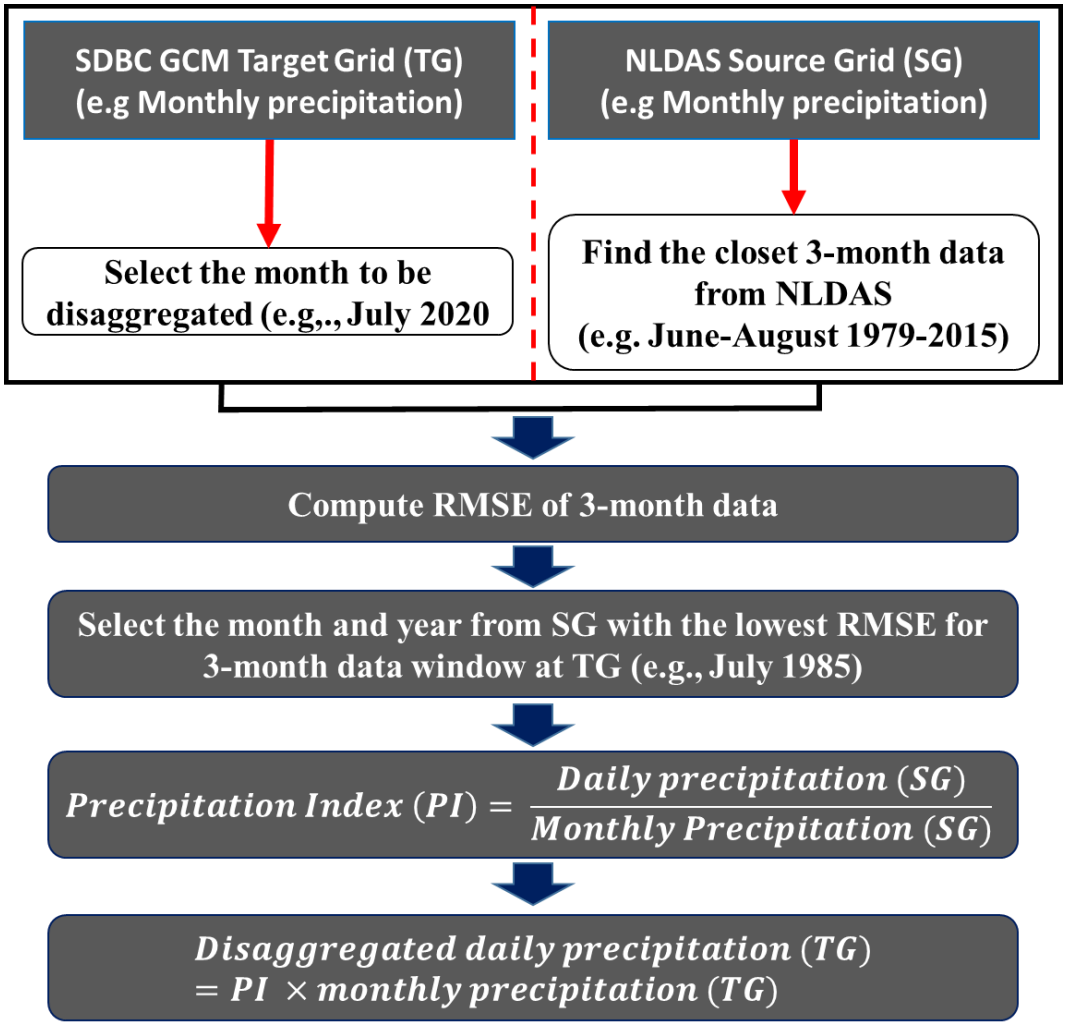


Figure 6.4. Simple temporal disaggregation method.

.

6.2.2.3. Potential Evapotranspiration (PET) of SDBC GCMs

PET is one of the critical elements for hydrological simulation. GCMs do not provide PET data, but we can calculate PET from datasets of SDBC GCMs using Penman-Monteith method (Allen et al, 1998). The Penman-Monteith PET is caluclated using Equation (4.1) in **Chpater 4**. For PET computation, climate forcing data including wind speed, temperature, and the solar radiation data retrieved from SDBC GCMs. Finally, all required datasets were compiled into WDM to facilitate HSPF's input and output processes using **Figure 5.5** method in **Chapter 5**.

6.3. Results

6.3.1. Future LULC variation

Figure 6.5 shows the spatial distribution of the spatially modeled LULC data from 2011 to 2080 and 2100 under three different scenarios (A1B, A2, and B1) in the BRW. **Table 6.3** and **Future 6.6** provide a summary of the projected change for 2080 and 2100 LULC under three scenarios. Urban land increased under three scenarios in 2080 and 2100 LULC relative to 2011 LULC condition. 2100 LULC A2 scenario indicated the greatest increase in urban land (46.79 %) since it assumes high economic growth and very high population growth globally. 2100 LULC A1B and B1 scenarios showed to increase urban land by 13.18 % and 7.51 % due to the similar global population assumption. Also, A1B scenario represented the increase in agricultural land by 3.12 % in 2100 LULC. A2 and B1 scenarios showed the decrease in agricultural land by 7.35% and 1.78% in 2100 LULC due to the conversion of agricultural land to urban land. Barren/mining land was sharply increased by 2,618% (542 km^2) under A1B scenario, 2,260% (470 km^2) under A2 scenario, and 2,180% (455 km^2) under B1 scenario in 2100 LULC comparing with 2011 barren/mining land (20 km^2), respectively. The modeled LULC products indicate the conversion of grassland to barren/mining land by grassland degradation. Forest land was converted to barrel/mining land by deforestation. There was a modest increase in the wetland by 4.56 % under A1B scenario, 3.48 % in under A2 scenario, and 8.64 % under B1 scenario in 2100 LULC.

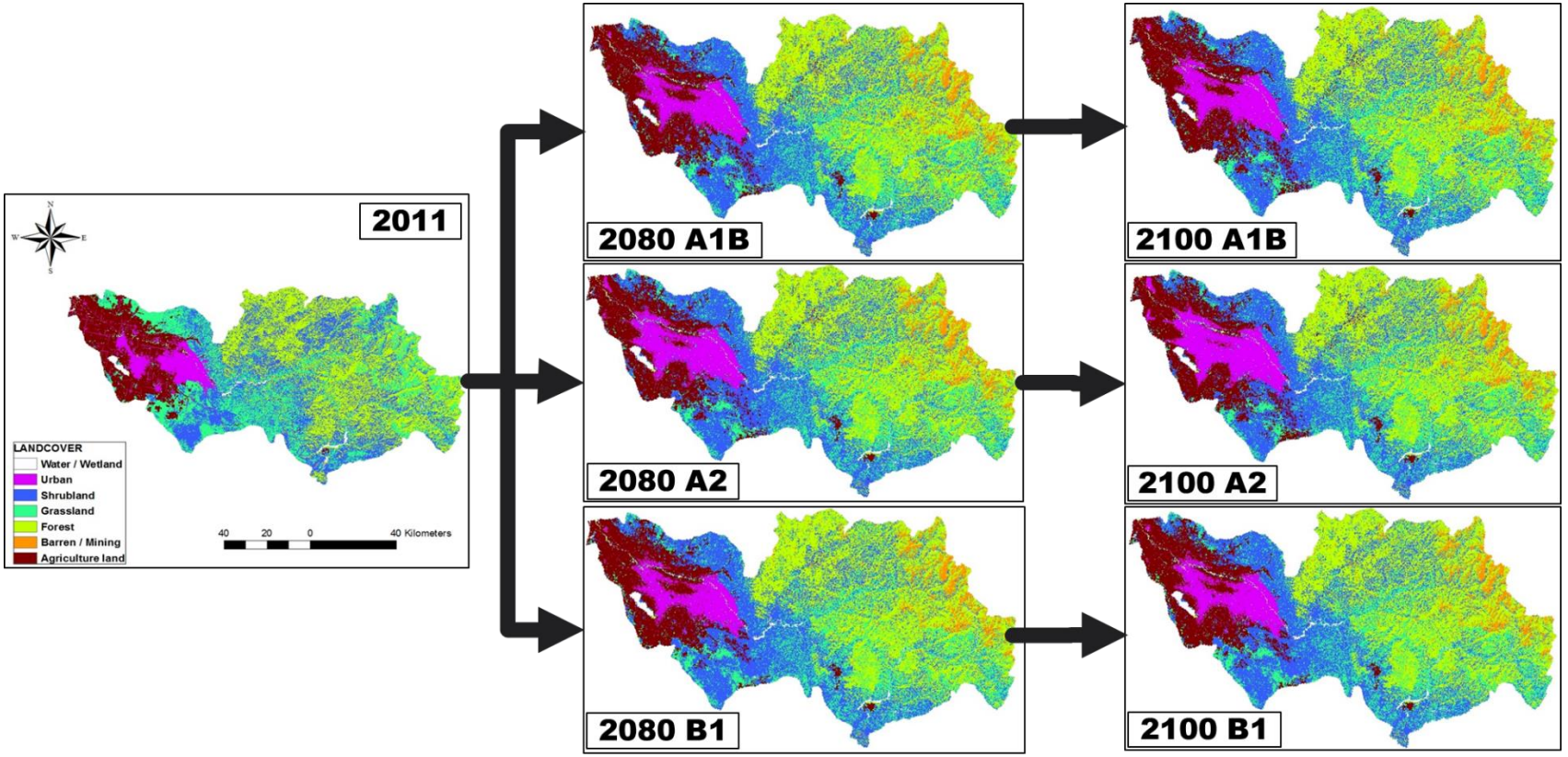
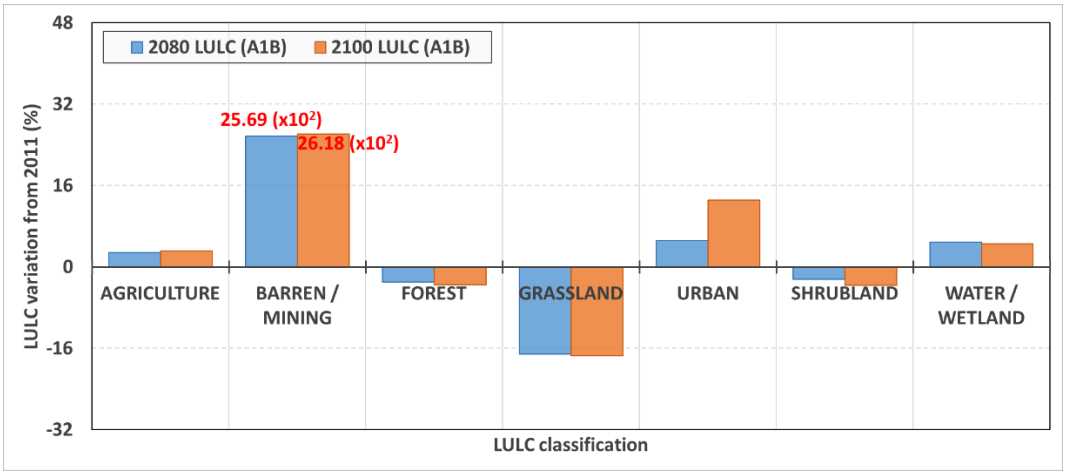


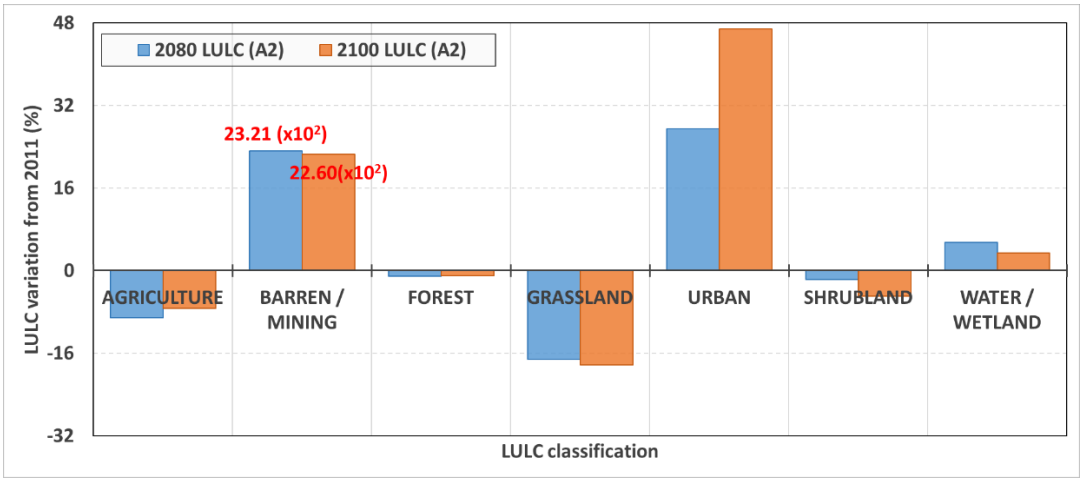
Figure 6.5. Spatial distribution of future LULC changes.

Table 6.3. LULC classification and variation for 2011, 2080, and 2100 LULC.

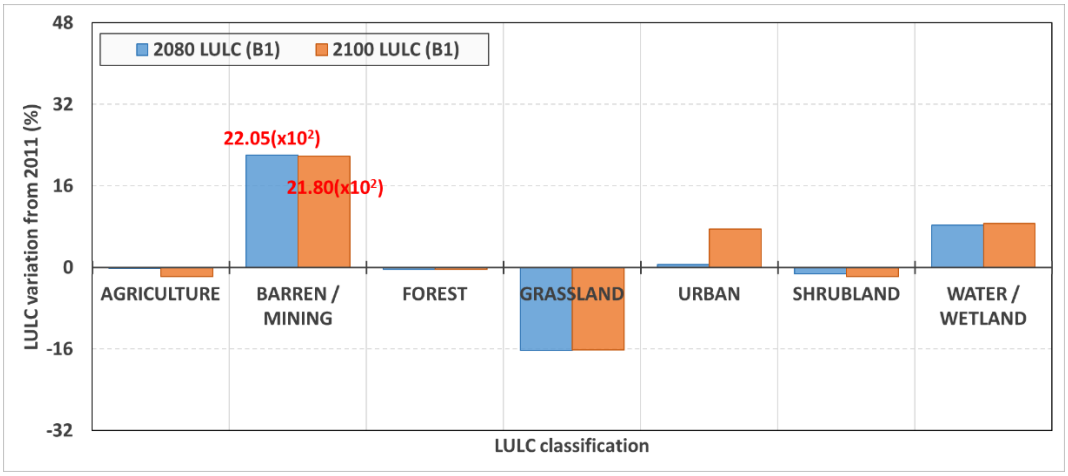
Classification	2011 LULC (km^2)	2080 LULC (km^2)			2100 LULC (km^2)		
		A1B scenario	A2 scenario	B1 scenario	A1B scenario	A2 scenario	B1 scenario
Barren / Mining	20 (0.19 %)	532 (5.09%)	483 (4.62%)	459 (4.40%)	542 (5.19%)	470 (4.51%)	455 (4.35%)
Cropland	1,276 (12.22%)	1,312 (12.57 %)	1,156 (11.10%)	1,274 (12.20%)	1,316 (12.60%)	1,182 (11.32%)	1,253 (12.01%)
Forest	2,992 (28.66%)	2,902 (27.80%)	2,959 (28.34%)	2,981 (28.55%)	2,887 (27.66%)	2,964 (28.40%)	2,980 (28.55%)
Grassland	2,450 (23.47%)	2,030 (19.45%)	2,028 (19.43%)	2,051 (19.65%)	2,023 (19.38%)	2,002 (19.18%)	2,053 (19.66%)
Shrubland	3,023 (28.96%)	2,950 (28.26%)	2,971 (28.46%)	2,984 (28.58%)	2,914 (27.91%)	2,874 (27.53%)	2,967 (28.43%)
Urban	564 (5.40%)	593 (5.68%)	718 (6.88%)	567 (5.43%)	638 (6.11%)	827 (7.92%)	606 (5.80%)
Water / Wetland	115 (1.10%)	121 (1.16%)	121 (1.16%)	125 (1.19%)	120 (1.15%)	119 (1.14%)	125 (1.20%)
Total	10,439 (100%)						



(a) A1B scenario



(b) A2. scenario



(c) B1 scenario

Figure 6.6. 2080 and 2100 LULC variations under A1B, A2, and B1 scenarios relative to 2011 LULC.

6.3.2. Impacts of future LULC changes.

6.3.2.1. Streamflow variation

Monthly maximum, mean, and minimum streamflow were used to evaluate the impacts of LULC changes over the entire simulation period (Jan. 1981 – Dec. 2015). **Table 6.4 – 6 and Figure 6.7 (a, b, and c)** show the simulation results of monthly maximum, mean and minimum streamflow under three scenarios in 2080 and 2100 LULC. 2080 and 2100 LULC changes under three scenarios (A1B, A2, B1) led to an increase in monthly maximum and mean streamflow in average, but they were decreased monthly minimum streamflow. For seasonal variation, monthly mean streamflow was decreased up to 14.22% in summer and increased up to 4.11% in winter under A1B, A2, and B1 scenarios in 2100 LULC. 2080 LULC B1 scenario was slightly increased in mean streamflow than other LULC change scenarios due to the evapotranspiration reduction by the decreased vegetation coverage relative to 2011 LULC condition. 2100 LULC A2 scenario indicated the biggest increases in maximum (41.17%) and mean streamflow (3.68%) relative to 2011 LULC condition because the largely urbanized land contributed to increasing of the impervious surface area in the BRW. These results confirm that urbanization was usually associated with an increase in maximum and mean streamflow due to the impervious land increasing. Also, the increased impervious land and decreased vegetation coverage (grassland to barren/mining land) eventually affected the increase in streamflow caused by the reduction of evapotranspiration and rainfall infiltration.

6.3.2.2 Sediment load variation

Table 6.7 and **Figure 6.7 (d)** show monthly sediment load variations under three

scenarios in 2080 and 2100 LULC. 2080 and 2100 LULC led to notably higher monthly sediment load except August. The main contribution of monthly sediment load reduction was a decrease in monthly mean streamflow in August. Monthly sediment load in winter was generally higher than non-winter loads under three scenarios in 2100 LULC. The average monthly sediment load was increased by 12.09% under A1B scenario, 24.02% under A2 scenario, and 9.05% under B1 scenario in 2080 LULC. From 2100 LULC condition, the average monthly sediment load was increased by 16.80% under A1B scenario, 34.16% under A2 scenario, 13.21% under B1 scenario, respectively. 2100 LULC A2 scenario represented the biggest increase in monthly sediment load due to a large of urbanized land.

6.3.2.3. TN load variation

Table 6.8 and **Figure 6.7 (e)** indicate monthly TN load variations under three scenarios in 2080 and 2100 LULC. 2080 and 2100 LULC yielded the higher monthly TN loads throughout the year relative to 2011 LULC condition. Monthly TN load change followed monthly mean streamflow variation because TN discharge during the surface runoff was strongly related to LULC change. For 2080 LULC condition, the average monthly TN load was increased by 134.62 % under A1B scenario, 134.08% under A2 scenario, and 111.25% under B1 scenario. For 2100 LULC condition, the average monthly TN load was increased by 131.55% under A1B scenario, 134.83% under A2 scenario, 112.10% under B1 scenario, respectively. 2100 LULC A2 scenario was the biggest increase of monthly TN load due to a higher rate of urban and barren/mining land. Especially, barren/mining land was mostly converted from forest or grassland, so monthly TN load was about twice bigger than 2011 LULC condition.

6.3.2.4. TP load variation

Table 6.9 and **Figure 6.7 (f)** indicate monthly TP load variations under three scenarios in 2080 and 2100 LULC. 2080 and 2100 LULC led to the increase in monthly TP load throughout the year relative to 2011 LULC. Spring and winter TP loads were highly increased under three scenarios in 2080 and 2100 LULC because TN load was associated with streamflow variation along the land use change. For 2080 LULC condition, the average monthly TP load was increased by 14.74 % under A1B scenario, 19.35% under A2 scenario, and 10.39% under B1 scenario. For 2100 LULC condition, the average monthly TP load was increased by 14.26% under A1B scenario, 22.92% under A2 scenario, and 11.97% under B1 scenario, respectively. 2100 LULC A2 scenario was the biggest increase in monthly TP load due to the growth in higher urbanization. As a result of this study, it is clear that urbanization affects the increases in monthly streamflow, sediment, TN, and TP loads.

Table 6.4. Changes in monthly maximum streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

Month	2011 LULC (m^3/s)	LULC scenarios	2080 LUCL (m^3/s)	2100 LUCL (m^3/s)	Variation (%)	
					2080 LUCL	2100 LUCL
1	124.88	A1B	133.56	138.32	6.95	10.76
		A2	147.28	159.32	17.94	27.58
		B1	131.32	135.52	5.16	8.52
2	196.00	A1B	208.04	212.52	6.14	8.43
		A2	219.52	229.88	12.00	17.29
		B1	205.80	209.72	5.00	7.00
3	234.92	A1B	251.16	257.04	6.91	9.42
		A2	264.60	276.64	12.63	17.76
		B1	249.20	253.96	6.08	8.10
4	185.92	A1B	209.72	220.36	12.80	18.52
		A2	229.60	250.04	23.49	34.49
		B1	208.32	215.04	12.05	15.66
5	219.80	A1B	224.00	224.56	1.91	2.17
		A2	225.68	227.36	2.68	3.44
		B1	223.16	223.72	1.53	1.78
6	218.40	A1B	241.64	237.72	10.64	8.85
		A2	248.92	246.12	13.97	12.69
		B1	232.12	232.40	6.28	6.41
7	130.48	A1B	145.04	147.28	11.16	12.88
		A2	154.84	161.28	18.67	23.61
		B1	141.68	144.48	8.58	10.73
8	60.48	A1B	71.40	72.80	18.06	20.37
		A2	76.44	80.64	26.39	33.33
		B1	70.00	71.40	15.74	18.06
9	59.92	A1B	66.64	68.04	11.21	13.55
		A2	72.24	86.52	20.56	44.39
		B1	65.24	66.64	8.88	11.21
10	50.12	A1B	62.44	66.08	24.58	31.84
		A2	73.64	85.40	46.93	70.39
		B1	60.76	63.84	21.23	27.37
11	54.60	A1B	81.76	89.60	49.74	64.10
		A2	103.60	121.80	89.74	123.08
		B1	78.96	85.96	44.62	57.44
12	89.88	A1B	117.32	126.84	30.53	41.12
		A2	144.76	167.16	61.06	85.98
		B1	113.68	122.36	26.48	36.14
Mean	135.45	A1B	151.06	155.10	15.89	20.17
		A2	163.43	174.35	28.84	41.17
		B1	148.35	152.09	13.47	17.37

Table 6.5. Changes in monthly mean streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

Month	2011 LULC (m^3/s)	LULC scenarios	2080 LUCL (m^3/s)	2100 LUCL (m^3/s)	Variation (%)	
					2080 LUCL	2100 LUCL
1	23.42	A1B	24.02	24.23	2.56	3.46
		A2	24.62	25.14	5.10	7.34
		B1	23.92	24.12	2.11	2.99
2	29.42	A1B	29.77	30.01	1.19	2.01
		A2	30.22	30.71	2.70	4.37
		B1	29.76	29.92	1.15	1.68
3	38.54	A1B	39.31	39.66	2.00	2.91
		A2	39.73	40.35	3.09	4.69
		B1	39.37	39.54	2.14	2.59
4	50.14	A1B	51.27	51.71	2.25	3.14
		A2	51.45	52.10	2.61	3.91
		B1	51.42	51.52	2.56	2.76
5	67.32	A1B	69.32	68.68	2.97	2.02
		A2	69.50	69.07	3.25	2.61
		B1	68.38	68.41	1.57	1.62
6	63.68	A1B	65.25	62.09	2.46	-2.49
		A2	65.23	62.14	2.43	-2.42
		B1	61.96	61.90	-2.70	-2.80
7	34.63	A1B	35.08	33.46	1.30	-3.39
		A2	35.09	33.48	1.34	-3.31
		B1	33.40	33.40	-3.54	-3.55
8	20.71	A1B	20.84	19.87	0.65	-4.02
		A2	20.83	19.87	0.61	-4.05
		B1	19.85	19.86	-4.13	-4.11
9	14.07	A1B	14.23	13.72	1.14	-2.52
		A2	14.38	13.94	2.19	-0.94
		B1	13.62	13.68	-3.22	-2.81
10	9.98	A1B	10.26	10.08	2.79	0.96
		A2	10.51	10.46	5.36	4.84
		B1	9.92	10.02	-0.55	0.41
11	10.45	A1B	10.99	11.08	5.21	6.08
		A2	11.46	11.79	9.74	12.87
		B1	10.82	10.99	3.62	5.18
12	16.26	A1B	17.16	17.39	5.52	6.94
		A2	17.94	18.57	10.34	14.22
		B1	16.99	17.25	4.50	6.08
Mean	31.55	A1B	32.29	31.83	2.50	1.26
		A2	32.58	32.30	4.06	3.68
		B1	31.62	31.72	0.29	0.84

Table 6.6. Changes in monthly minimum streamflow for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

Month	2011 LULC (m^3/s)	LULC scenarios	2080 LUCL (m^3/s)	2100 LUCL (m^3/s)	Variation (%)	
					2080 LUCL	2100 LUCL
1	6.30	A1B	6.33	6.33	0.44	0.44
		A2	6.38	6.47	1.33	2.67
		B1	6.30	6.33	0.00	0.44
2	9.52	A1B	9.46	9.44	-0.59	-0.88
		A2	9.41	9.38	-1.18	-1.47
		B1	9.46	9.46	-0.59	-0.59
3	10.02	A1B	10.02	10.02	0.00	0.00
		A2	10.02	10.02	0.00	0.00
		B1	10.02	10.02	0.00	0.00
4	11.20	A1B	11.23	11.20	0.25	0.00
		A2	11.23	11.20	0.25	0.00
		B1	11.23	11.20	0.25	0.00
5	13.13	A1B	13.19	13.22	0.43	0.64
		A2	13.19	13.19	0.43	0.43
		B1	13.19	13.22	0.43	0.64
6	9.55	A1B	9.55	9.55	0.00	0.00
		A2	9.55	9.55	0.00	0.00
		B1	9.55	9.55	0.00	0.00
7	9.10	A1B	8.32	8.12	-8.62	-10.77
		A2	8.29	8.18	-8.92	-10.15
		B1	8.04	8.46	-11.69	-7.08
8	5.21	A1B	5.18	5.18	-0.54	-0.54
		A2	5.18	5.18	-0.54	-0.54
		B1	5.18	5.21	-0.54	0.00
9	4.87	A1B	4.84	4.84	-0.57	-0.57
		A2	4.84	4.84	-0.57	-0.57
		B1	4.84	4.84	-0.57	-0.57
10	4.87	A1B	4.84	4.84	-0.57	-0.57
		A2	4.84	4.84	-0.57	-0.57
		B1	4.84	4.84	-0.57	-0.57
11	4.98	A1B	4.87	4.87	-2.25	-2.25
		A2	4.87	4.87	-2.25	-2.25
		B1	4.87	4.90	-2.25	-1.69
12	5.57	A1B	5.46	5.49	-2.01	-1.51
		A2	5.52	5.52	-1.01	-1.01
		B1	5.46	5.52	-2.01	-1.01
Mean	7.86	A1B	7.77	7.76	-1.17	-1.33
		A2	7.78	7.77	-1.09	-1.12
		B1	7.75	7.80	-1.46	-0.82

Table 6.7. Changes in monthly sediment load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

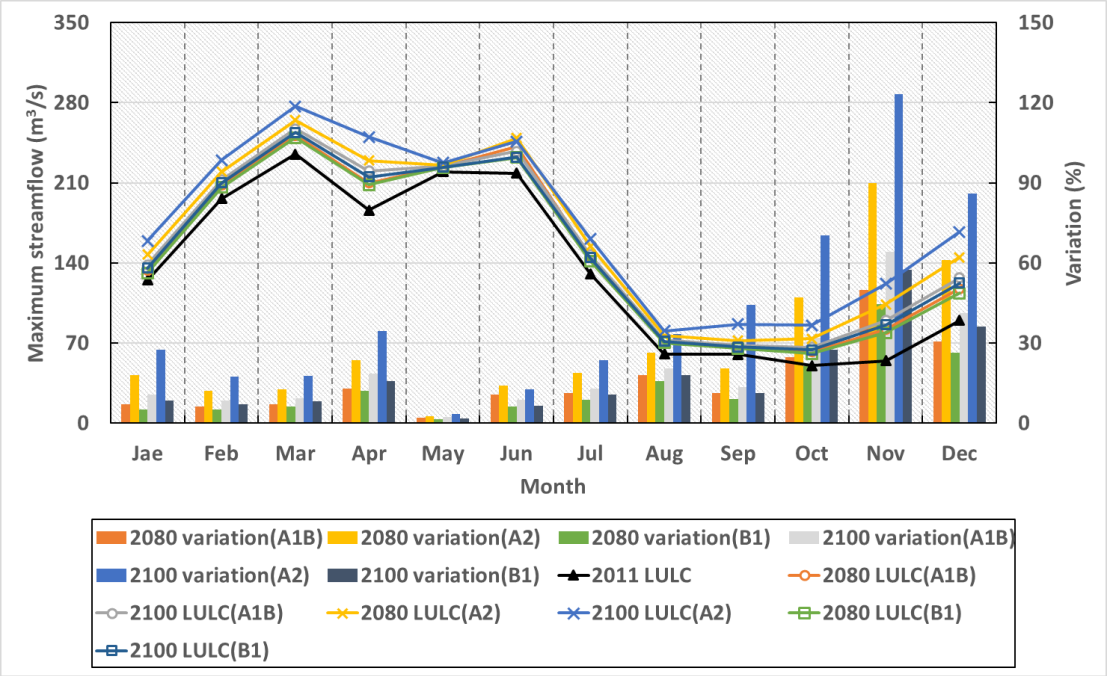
Month	2011 LULC (100×ton/ month)	LULC scenarios	2080 LUCL (100×ton/ month)	2100 LUCL (100×ton/ month)	Variation (%)	
					2080 LULC	2100 LULC
1	20.64	A1B	23.15	24.66	12.15	19.46
		A2	26.94	30.27	30.52	46.67
		B1	22.61	23.86	9.53	15.59
2	21.98	A1B	24.19	25.65	10.03	16.67
		A2	27.57	30.66	25.42	39.47
		B1	23.72	24.85	7.90	13.07
3	26.46	A1B	29.43	30.97	11.20	17.01
		A2	32.38	35.31	22.35	33.44
		B1	28.98	30.06	9.51	13.60
4	27.95	A1B	31.43	32.83	12.43	17.46
		A2	33.17	35.32	18.64	26.34
		B1	31.22	31.92	11.68	14.20
5	38.25	A1B	43.40	44.93	13.45	17.46
		A2	46.04	48.61	20.37	27.08
		B1	42.52	43.56	11.17	13.87
6	30.91	A1B	32.75	32.39	5.97	4.82
		A2	33.56	33.36	8.58	7.95
		B1	31.18	31.53	0.90	2.02
7	17.44	A1B	19.47	19.57	11.70	12.22
		A2	20.38	20.70	16.91	18.70
		B1	18.76	19.10	7.61	9.54
8	10.19	A1B	9.57	9.55	-6.11	-6.28
		A2	9.75	9.69	-4.40	-4.90
		B1	9.27	9.36	-9.07	-8.16
9	15.76	A1B	17.13	17.84	8.65	13.17
		A2	19.15	20.83	21.46	32.16
		B1	16.60	17.28	5.31	9.61
10	15.40	A1B	18.72	19.80	21.51	28.54
		A2	21.57	24.02	40.05	55.91
		B1	18.12	19.11	17.65	24.03
11	19.44	A1B	23.84	25.35	22.66	30.40
		A2	27.66	31.03	42.33	59.65
		B1	23.04	24.34	18.56	25.22
12	20.09	A1B	24.39	26.25	21.42	30.69
		A2	29.34	33.62	46.06	67.39
		B1	23.67	25.29	17.85	25.93
Mean	22.04	A1B	24.79	25.82	12.09	16.80
		A2	27.29	29.45	24.02	34.16
		B1	24.14	25.02	9.05	13.21

Table 6.8. Changes in monthly TN load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

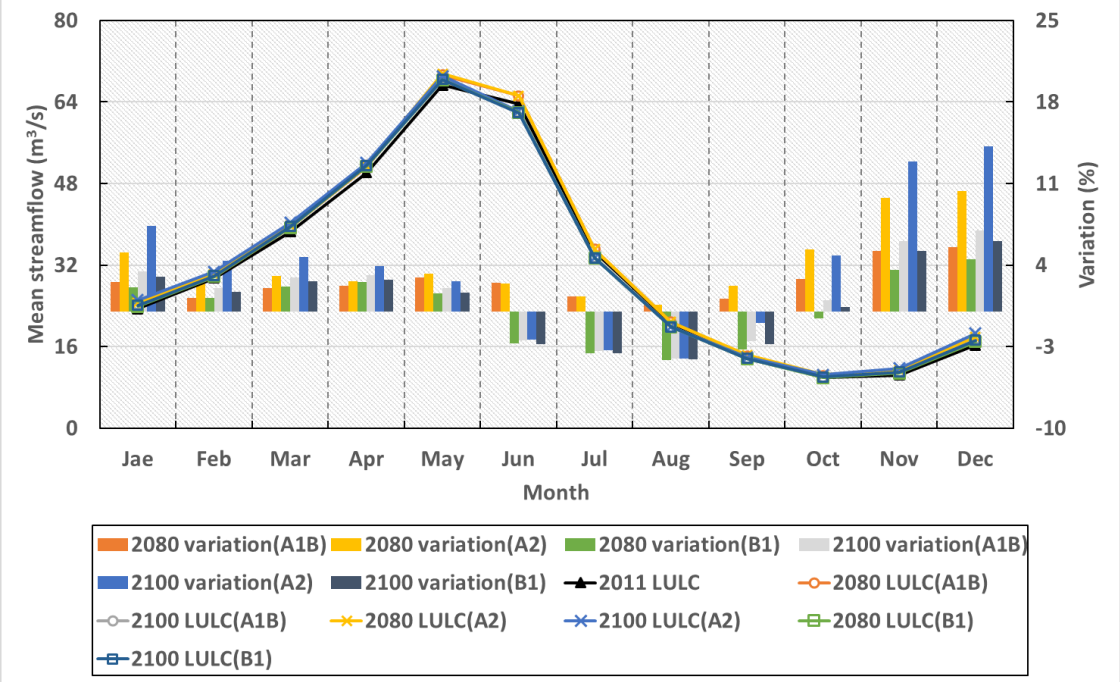
Month	2011 LULC (ton/month)	LULC scenarios	2080 LUCL (ton/month)	2100 LUCL (ton/month)	Variation (%)	
					2080 LULC	2100 LULC
1	105.34	A1B	174.98	182.43	66.11	73.18
		A2	167.10	171.67	81.08	91.40
		B1	190.75	201.62	58.62	62.96
2	119.92	A1B	188.95	192.90	57.56	60.86
		A2	178.17	181.16	64.07	72.76
		B1	196.75	207.18	48.57	51.07
3	136.63	A1B	232.01	233.73	69.81	71.07
		A2	215.12	216.72	69.29	77.32
		B1	231.29	242.26	57.45	58.62
4	132.28	A1B	280.98	282.71	112.41	113.72
		A2	253.74	254.18	104.38	109.64
		B1	270.35	277.31	91.82	92.15
5	157.21	A1B	420.36	416.62	167.39	165.01
		A2	373.85	372.52	156.28	155.76
		B1	402.89	402.08	137.81	136.96
6	129.24	A1B	430.48	410.02	233.17	217.33
		A2	375.03	370.28	218.51	202.19
		B1	411.54	390.46	190.25	186.58
7	67.29	A1B	231.56	220.41	244.13	227.56
		A2	201.91	199.36	228.53	210.28
		B1	221.06	208.78	200.08	196.28
8	38.57	A1B	133.45	126.47	245.97	227.87
		A2	115.97	114.34	228.97	208.96
		B1	126.89	119.17	200.65	196.42
9	39.51	A1B	102.97	99.37	160.61	151.49
		A2	90.84	90.86	157.43	154.33
		B1	101.72	100.49	129.90	129.95
10	42.10	A1B	87.40	86.42	107.59	105.27
		A2	78.92	80.34	113.83	122.35
		B1	90.03	93.61	87.45	90.82
11	62.37	A1B	111.06	114.38	78.06	83.39
		A2	104.11	107.43	93.95	108.11
		B1	120.97	129.80	66.91	72.24
12	82.73	A1B	142.84	150.41	72.65	81.80
		A2	136.86	141.62	92.62	104.91
		B1	159.36	169.52	65.43	71.18
Mean	92.76	A1B	211.42	209.66	134.62	131.55
		A2	190.97	191.71	134.08	134.83
		B1	210.30	211.86	111.25	112.10

Table 6.9. Changes in monthly TP load for 2080 and 2100 LULC under three scenarios relative to the 2011 LULC condition.

Month	2011 LULC (ton/month)	LULC scenarios	2080 LUCL (ton/month)	2100 LUCL (ton/month)	Variation (%)	
					2080 LULC	2100 LULC
1	10.80	A1B	12.55	12.99	16.26	20.31
		A2	13.66	14.40	26.56	33.41
		B1	12.39	12.72	14.78	17.78
2	12.63	A1B	14.07	14.29	11.42	13.14
		A2	14.75	15.51	16.81	22.83
		B1	13.79	14.03	9.21	11.08
3	14.69	A1B	16.31	16.39	11.03	11.56
		A2	16.71	17.55	13.74	19.47
		B1	15.89	16.09	8.20	9.52
4	16.74	A1B	18.61	18.70	11.18	11.67
		A2	18.79	19.46	12.21	16.21
		B1	18.22	18.33	8.82	9.48
5	19.89	A1B	22.84	22.72	14.84	14.24
		A2	23.03	23.48	15.79	18.06
		B1	22.13	22.22	11.25	11.74
6	16.94	A1B	19.86	19.08	17.22	12.64
		A2	19.83	19.37	17.03	14.35
		B1	18.70	18.69	10.39	10.28
7	10.53	A1B	12.02	11.52	14.15	9.40
		A2	12.01	11.74	14.07	11.49
		B1	11.34	11.33	7.64	7.55
8	7.22	A1B	8.02	7.61	11.08	5.43
		A2	7.97	7.74	10.42	7.26
		B1	7.51	7.50	4.10	3.95
9	6.27	A1B	7.04	6.80	12.23	8.49
		A2	7.21	7.33	15.03	16.90
		B1	6.61	6.69	5.41	6.67
10	5.41	A1B	6.24	6.14	15.30	13.49
		A2	6.59	6.91	21.72	27.64
		B1	5.88	6.02	8.62	11.23
11	6.23	A1B	7.53	7.71	20.83	23.83
		A2	8.29	8.92	33.11	43.15
		B1	7.26	7.51	16.52	20.62
12	8.02	A1B	9.73	10.17	21.35	26.87
		A2	10.88	11.57	35.70	44.27
		B1	9.60	9.92	19.75	23.79
Mean	11.28	A1B	12.90	12.84	14.74	14.26
		A2	13.31	13.66	19.35	22.92
		B1	12.44	12.59	10.39	11.97

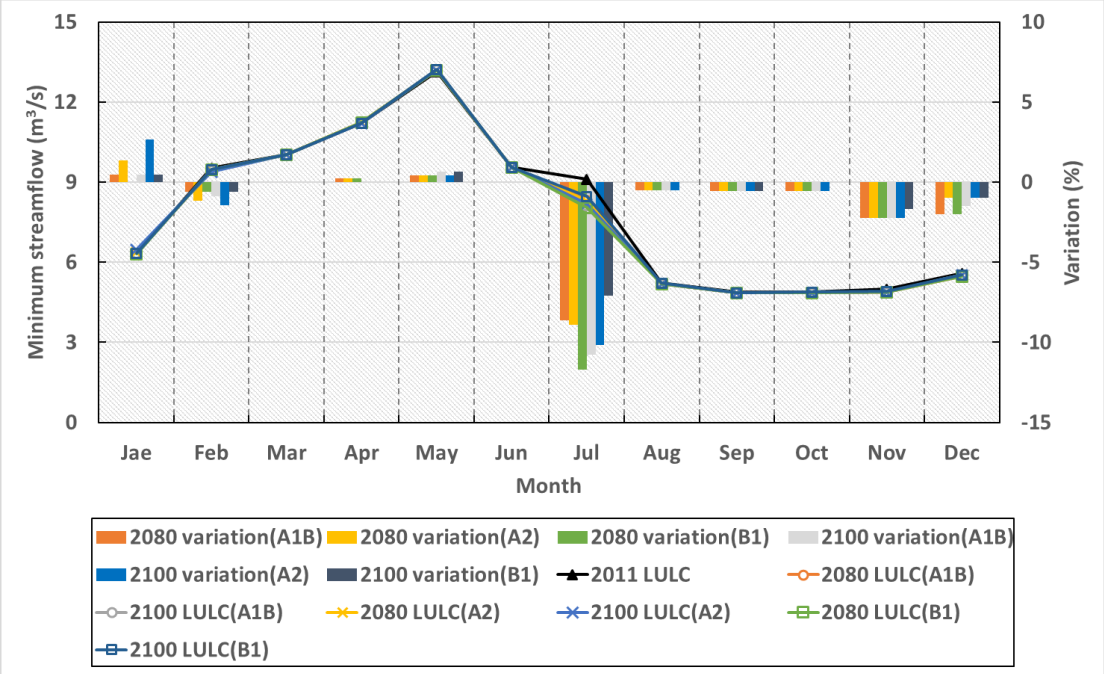


(a) Maximum streamflow

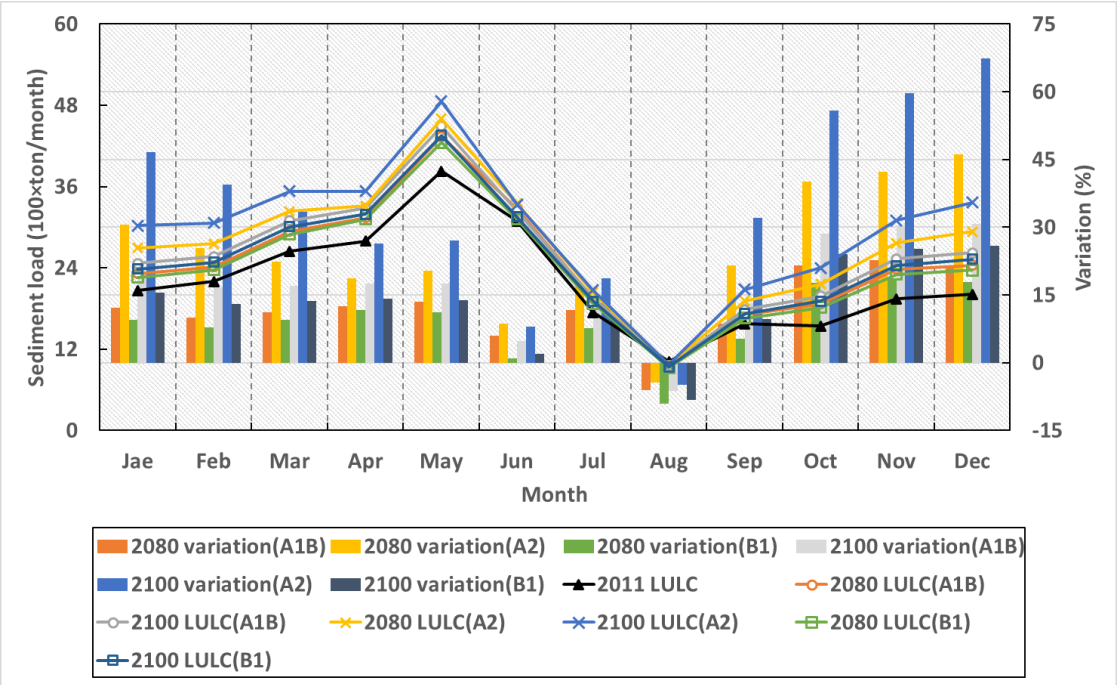


(b) Mean streamflow

Figure 6.7. Changes in monthly maximum streamflow (a), mean streamflow (b), minimum streamflow (c), sediment load (d), TN load (e), and TP load (f) in the 2080 LULC and 2100 LULC under three scenarios three scenarios.

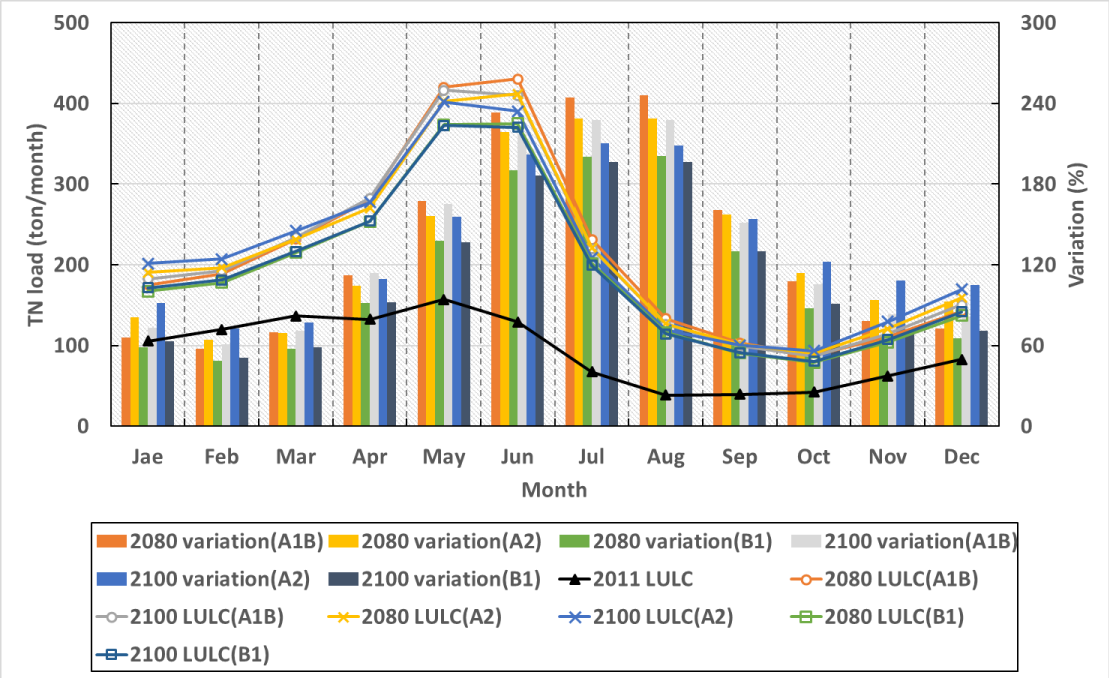


(c) Minimum streamflow

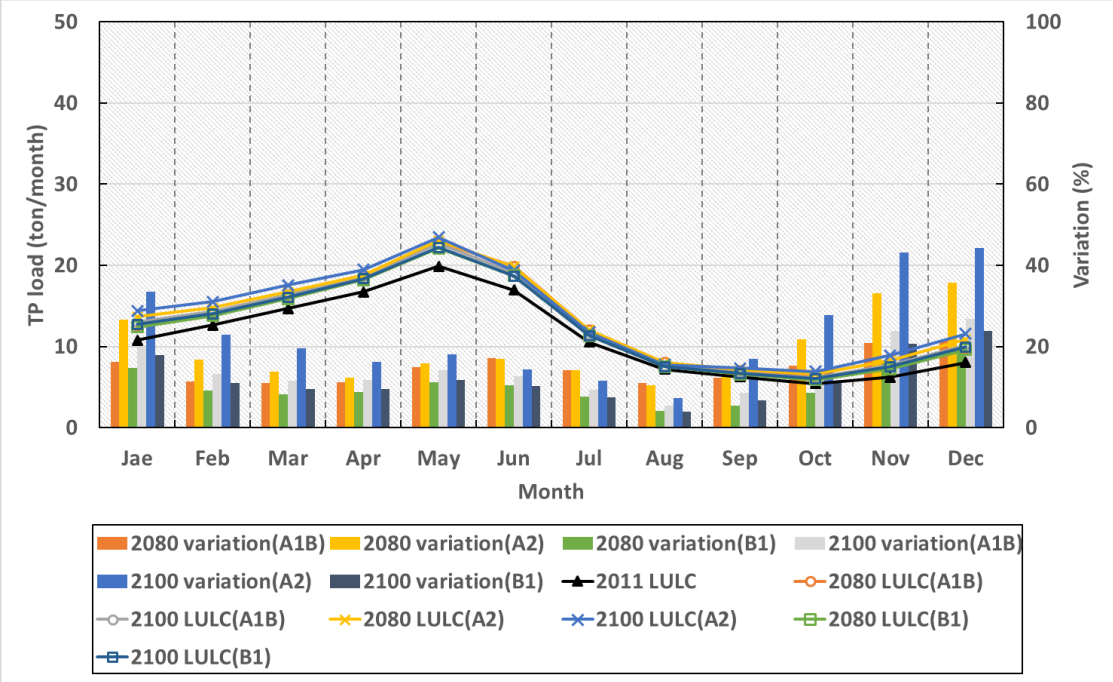


(d) Sediment load

Figure 6.7. Changes in monthly maximum streamflow (a), mean streamflow (b), minimum streamflow (c), sediment load (d), TN load (e), and TP load (f) in the 2080 LULC and 2100 LULC under three scenarios (Cont.).



(e) TN load



(f) TP load

Figure 6.7. Changes in monthly maximum streamflow (a), mean streamflow (b), minimum streamflow (c), sediment load (d), TN load (e), and TP load (f) in the 2080 LULC and 2100 LULC under three scenarios (Cont.).

6.3.3. The selection of optimal GCMs

Total 27 GCMs were applied to select optimal GCMs which better represent the historical climate condition based on seasonal precipitation and temperature in the BRW. **Table 6.10** shows normalized values for skill scores and RMSE and ranked results of GCMs. The highest ranked model was Canesm2 based on weighting results. Bcc-csm1-1m model was the second ranked model. The lowest normalized score showed giss-e2-r model by 0.11. Canesm2 and bcc-csm1-1m models were selected the best GCMs which well represented historical climate condition for the BRW. Cnrm-cm5 model, which was analyzed the highest-ranked model for the Pacific Northwest by Rupp et al. (2014), was also considered for the reliable climate change study. Three GCMs (canesm2, bcc-csm1-1m, and cnrm-cm4) were finally selected to apply SDBC and analyze the impacts of climate change using HSPF model.

Table 6.10. The GCMs ranked according to weighting values between seasonal precipitation and temperature using normalized HSS, FSS, and RMSE scores.

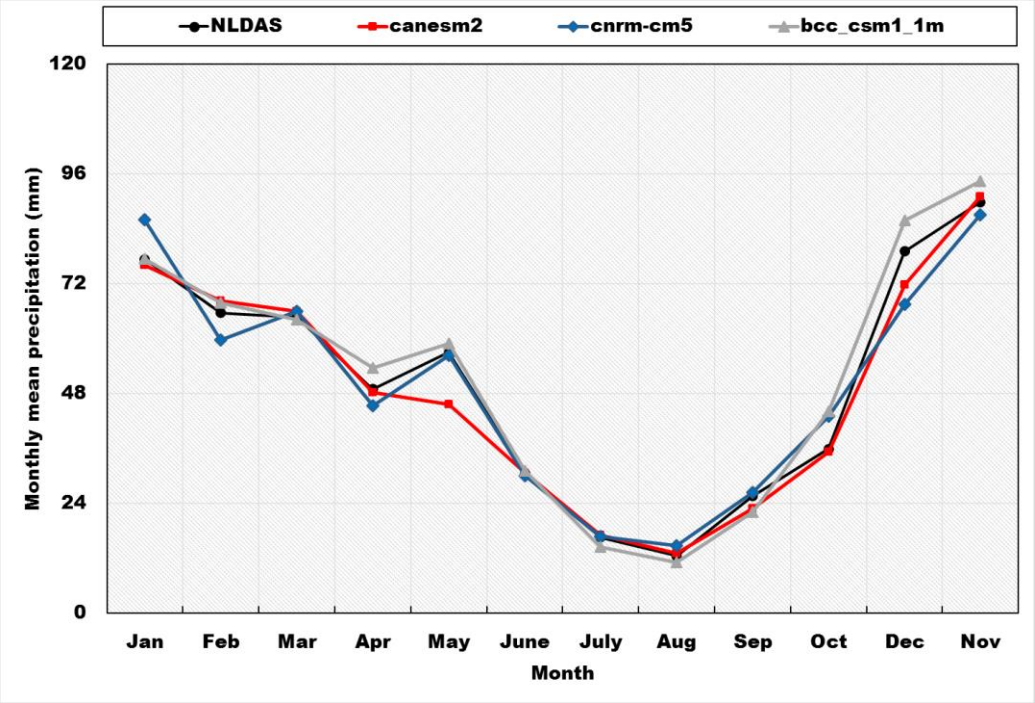
GCM number	Precipitation (P)				Temperature (T)				Weight (P *0.7 + T *0.3)	Ranking
	HSS	FSS	RMSE	Mean	HSS	FSS	RMSE	Mean		
1	0.70	0.70	0.78	0.73	0.15	0.14	0.23	0.17	0.56	11
2	***1.00***	***1.00***	***0.49***	***0.83***	***0.78***	***0.77***	***0.79***	***0.78***	***0.81***	***2***
3	***0.82***	***0.81***	***1.00***	***0.88***	***0.98***	***0.99***	***0.86***	***0.94***	***0.90***	***1***
4	0.41	0.41	0.56	0.46	1.00	1.00	1.00	1.00	0.62	9
5	0.55	0.54	0.54	0.54	0.92	0.92	0.96	0.93	0.66	8
6	0.54	0.53	0.83	0.63	0.91	0.91	0.88	0.90	0.71	4
7	0.56	0.56	0.64	0.59	0.95	0.95	0.91	0.94	0.69	6
8	0.55	0.55	0.67	0.59	0.90	0.90	0.98	0.93	0.69	5
9	0.33	0.32	0.69	0.45	0.05	0.05	0.20	0.10	0.34	17
10	0.40	0.40	0.66	0.49	0.10	0.10	0.12	0.11	0.37	16
11	0.49	0.48	0.87	0.61	0.00	0.00	0.03	0.01	0.43	14
12	0.34	0.33	0.73	0.47	0.19	0.19	0.31	0.23	0.40	15
13	0.17	0.16	0.39	0.24	0.35	0.34	0.33	0.34	0.27	20
14	0.16	0.16	0.34	0.22	0.16	0.15	0.00	0.10	0.19	26
15	0.00	0.00	0.33	0.11	0.15	0.14	0.03	0.11	0.11	27
16	0.44	0.43	0.94	0.61	1.00	1.00	0.63	0.87	0.69	7
17	0.17	0.17	0.17	0.17	0.34	0.33	0.54	0.40	0.24	24
18	0.30	0.30	0.89	0.50	0.35	0.33	0.24	0.31	0.44	13
19	0.66	0.65	0.82	0.71	0.30	0.29	0.52	0.37	0.61	10
20	0.17	0.17	0.47	0.27	0.46	0.44	0.40	0.43	0.32	18
21	0.19	0.18	0.59	0.32	0.90	0.91	0.78	0.86	0.48	12
22	0.68	0.67	0.82	0.72	0.58	0.58	0.97	0.71	0.72	3
23	0.16	0.16	0.01	0.11	0.53	0.51	0.86	0.63	0.27	21
24	0.26	0.26	0.00	0.18	0.47	0.45	0.74	0.55	0.29	19
25	0.18	0.25	0.02	0.15	0.38	0.38	0.80	0.52	0.26	22
26	0.17	0.17	0.13	0.16	0.29	0.28	0.43	0.34	0.21	25
27	0.21	0.21	0.09	0.17	0.35	0.34	0.53	0.41	0.24	23

6.3.4. The performance evaluation of SDBC GCMs

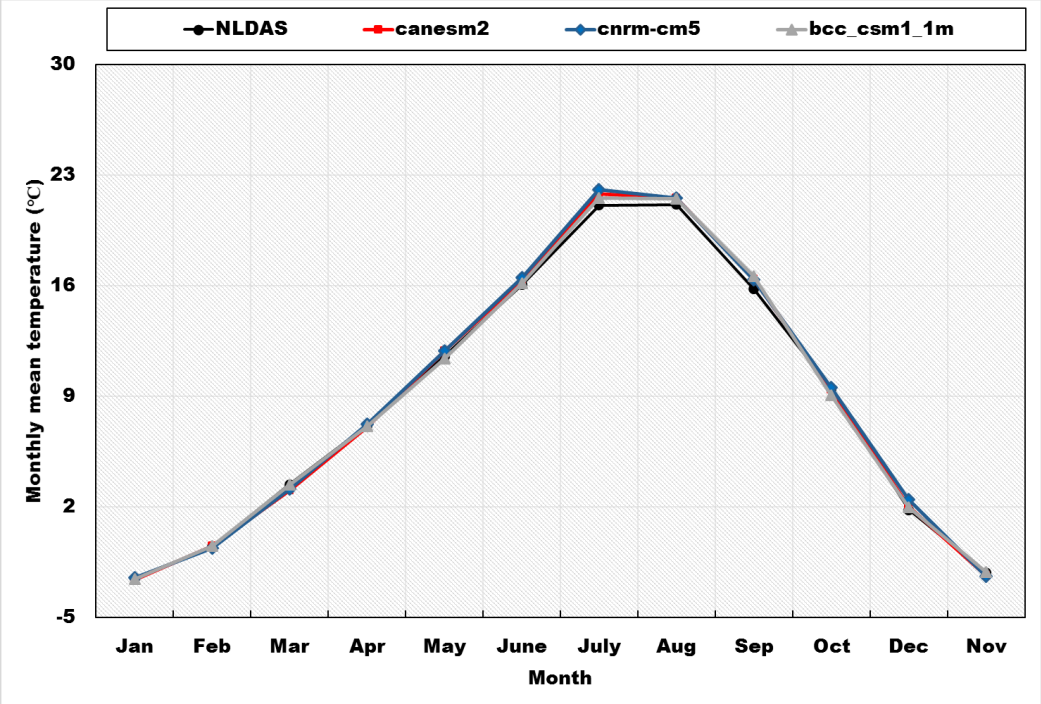
SDBC method was applied to create fine spatial resolution and reduce climate uncertainties for three GCMs (bcc-csm1-1m, canesm2, and cnrm-cm5 models). The performance of SDBC GCMs was evaluated using monthly NLDAS-2 data (1981-2005). **Table 6.11** and **Figure 6.8** show the comparisons results of performance measures for SDBC GCMs. Overall, SDBC GCMs agreed well with the historical monthly precipitation, temperature, solar radiation, specific humidity, and wind speed variations (**Figure 6.8**). Temperature, solar radiation, and specific humidity were calculated higher performance statistics for HSS, FSS, RMSE, MAE, and R^2, respectively. Also, PBIAS was computed below ± 5% for monthly precipitation, temperature, solar radiation, specific humidity, and wind speed of three SDBC GCMs.

Table 6.11. The performance comparisons for monthly SDBC GCMs.

Climate variable	GCMs	HSS	FSS	RMSE	MAE	R^2	PBIAS
Precipitation	bcc_csm1_1m	0.17	0.17	46.53	33.38	0.41	3.47
	canesm2	0.16	0.16	48.07	33.80	0.35	-3.12
	cnrm-cm5	0.16	0.16	48.17	33.33	0.36	-0.65
Temperature	bcc_csm1_1m	0.65	0.65	2.56	2.05	0.96	1.68
	canesm2	0.64	0.64	2.67	2.11	0.95	2.43
	cnrm-cm5	0.66	0.66	2.60	2.07	0.96	4.12
Solar radiation	bcc_csm1_1m	0.80	0.80	12.99	10.10	0.99	-0.09
	canesm2	0.79	0.79	13.43	10.18	0.99	-0.06
	cnrm-cm5	0.81	0.81	13.02	9.97	0.99	-0.21
Specific Humidity	bcc_csm1_1m	0.40	0.40	0.00	0.00	0.83	0.81
	canesm2	0.43	0.43	0.00	0.00	0.84	1.73
	cnrm-cm5	0.41	0.41	0.00	0.00	0.82	1.57
Wind speed	bcc_csm1_1m	0.09	0.09	0.93	0.72	0.19	-0.75
	canesm2	0.08	0.08	0.93	0.73	0.18	-2.73
	cnrm-cm5	0.11	0.11	0.87	0.69	0.30	-2.56

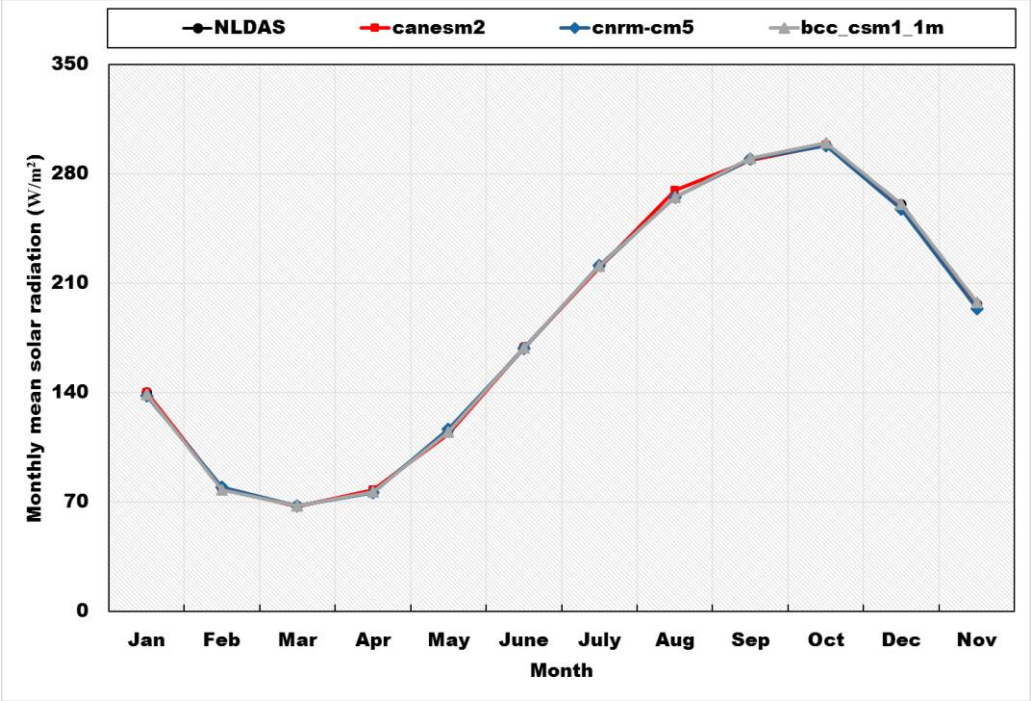


(a) Precipitation

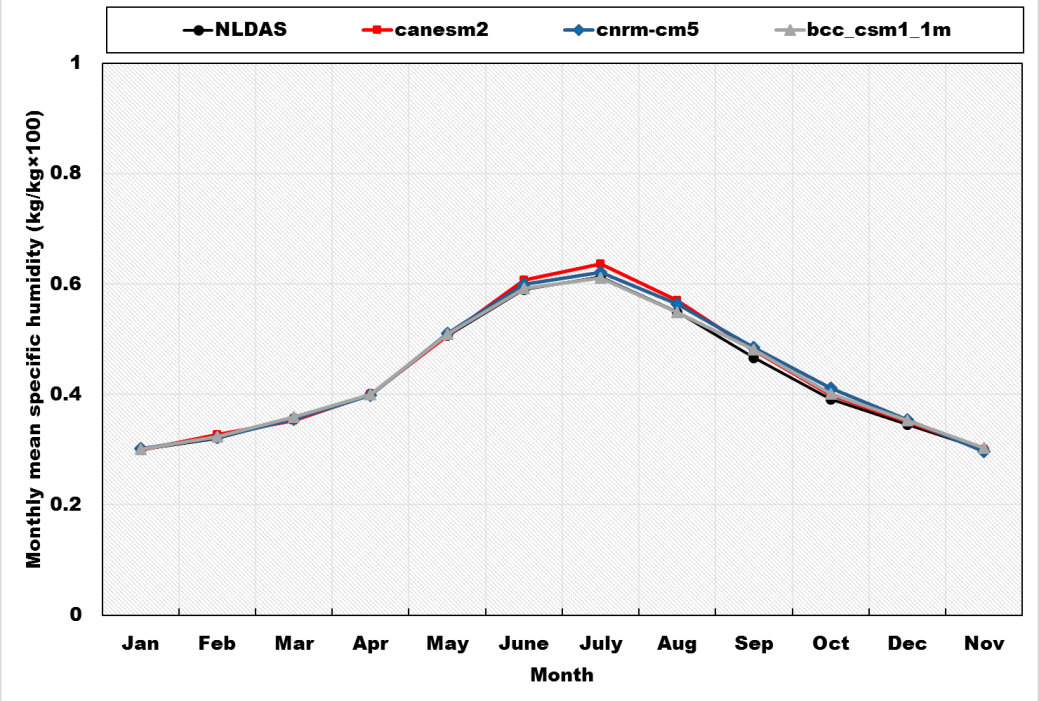


(b) Temperature

Figure 6.8. Comparison of historical 25-year monthly climate data (1981- 2005) for NLDAS-2 and SDBC GCMs.

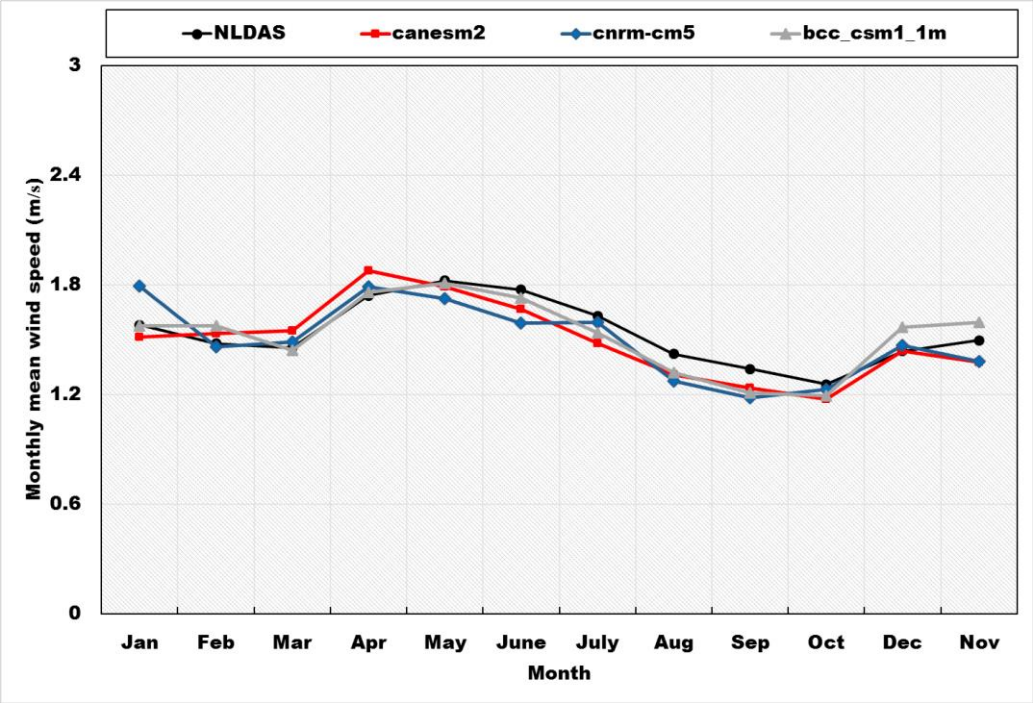


(b) Solar radiation



(d) Humidity

Figure 6.8. Comparison of historical 25-year monthly climate data (1981- 2005) for NLDAS-2 and SDBC GCMs (Cont.).



(e) Wind Speed

Figure 6.8. Comparison of historical 25-year monthly climate data (1981- 2005) for NLDAS-2, MACA, and SDBC GCMs (Cont.).

6.3.5. Future precipitation and temperature generated by three SDBC GCMs

Three SDBC GCMs were used to analyze the changes in future precipitation (%) and temperature (°C) under each emission scenario. **Table 6.12** shows the changes in annual and season precipitation for three SDBC GCMs in the future periods (F1: 2021-2045, F2: 2046-2070, F3: 2071-2095) according to three RCPs comparing with the baseline period. Future precipitation did not indicate a significant trend from three SDBC GCMs under RCP2.6, RCP4.5, and RCP8.5. Precipitation was considerably increased in winter and decreased in summer under RCP8.5 throughout the future periods. The comparisons of precipitation variations before and after SDBC indicated that mean annual precipitation of raw bcc-csm1-1m model was increased up to 31.26% under all RCPs in the future periods. However, SDBC bcc-csm1-1m model showed the decreases in mean annual precipitation by 0.16% under RCP2.6 in the F1 and 5.38% under RCP4.5 in the F2 relative to the baseline period. Mean annual precipitation of raw canesm2 model was increased up to 21.87% under RCP8.5 in the future periods relative to the baseline period, but SDBC canesm2 model indicated the increase in precipitation by 2.22-2.93 % under RCP4.5 and RCP8.5 in the F1 and F2. The raw cnrm-cm5 model was increased mean annual precipitation by 3.58% under RCP4.5 and RCP8.5 in the F3 relative to the baseline period. SDBC cnrm-cm5 model showed the increases in mean annual precipitation under all RCPs in the future periods relative the baseline period. Three SDBC GCMs represented a substantial increase trend of precipitation under RCP8.5 in winter. According to the average variation of three SDBC GCMs, future precipitation was significantly increased under RCP8.5 except for cnrm-cm5 model, but it was slightly increased or decreased under RCP2.6 relative to the baseline period.

Table 6.13 indicates the changes in annual and season temperature for three SDBC

GCMs in the future periods according to three RCPs comparing with baseline period. Canesm2 model showed a large increase in temperature from all season than those of baseline. Bcc-csm1-1m and cnrm-cm5 models indicated a consistent increase in temperature under all RCPs. Summer and winter temperature of SDBC GCMs were highly increased under RCP4.5 and RCP8.5 in the future periods than spring and fall temperature. Mean annual temperature increased from 1.29 ℃ to 4.26 ℃ by bcc-csm1-1m model, 2.63℃ to 6.47℃ by canesm2 model, and 0.85℃ to 5.18℃ by cnrm-cm5 model under all RCPs in the future periods, respectively.

Table 6.12. Future precipitation variations under RCP2.6, RCP4.5, and RPC8.5 climate change in three future periods relative to the baseline period (1981-2005).

SDBC GCMs	Seasonal	Baseline (mm/month)	Variation (%)								
			F1 (2021-2045)			F2 (2046-2070)			F3 (2071-2095)		
			RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5
Bcc-csm1-1m	Winter	77.60	-0.21	7.14	20.89	6.34	12.21	24.75	4.61	-3.01	36.18
	Spring	56.91	3.05	8.42	1.30	-3.43	0.04	-2.04	2.63	-13.54	-2.27
	Summer	19.87	1.33	-20.76	-15.88	-12.26	-24.70	-10.88	-9.87	-15.41	-27.26
	Autumn	46.87	-4.62	-9.47	24.72	12.81	-5.97	-6.27	5.94	4.84	-3.39
	Mean	50.31	-0.16	0.88	12.61	3.24	0.89	6.43	2.93	-5.38	9.83
Canesm2	Winter	77.60	7.62	-1.34	-9.78	-6.89	-2.38	1.65	2.08	0.26	34.31
	Spring	56.91	-2.91	-5.87	-1.18	-7.00	1.84	-0.10	2.01	14.44	17.19
	Summer	19.87	-9.59	18.97	11.20	6.46	17.61	48.79	31.26	47.83	69.16
	Autumn	46.87	2.64	-8.23	2.10	4.59	6.83	12.32	-8.13	-2.55	12.58
	Mean	50.31	1.78	-2.22	-2.51	-2.93	2.93	8.30	2.56	8.31	27.85
Cnrm-cm5	Winter	77.60	-0.62	16.32	15.26	-1.89	21.58	24.03	1.28	27.52	25.71
	Spring	56.91	26.32	3.07	13.92	12.42	15.15	27.02	20.60	19.43	22.29
	Summer	19.87	30.55	7.34	-7.11	13.23	0.47	-18.63	33.40	0.18	-10.39
	Autumn	46.87	23.25	6.69	-7.56	34.53	-2.49	5.11	13.81	5.52	27.44
	Mean	50.31	15.64	9.44	7.36	12.13	12.07	16.26	12.83	17.41	21.58

Table 6.13. Future temperature variations under RCP2.6, RCP4.5, and RPC8.5 climate change in three future periods relative to the baseline period (1979-2005).

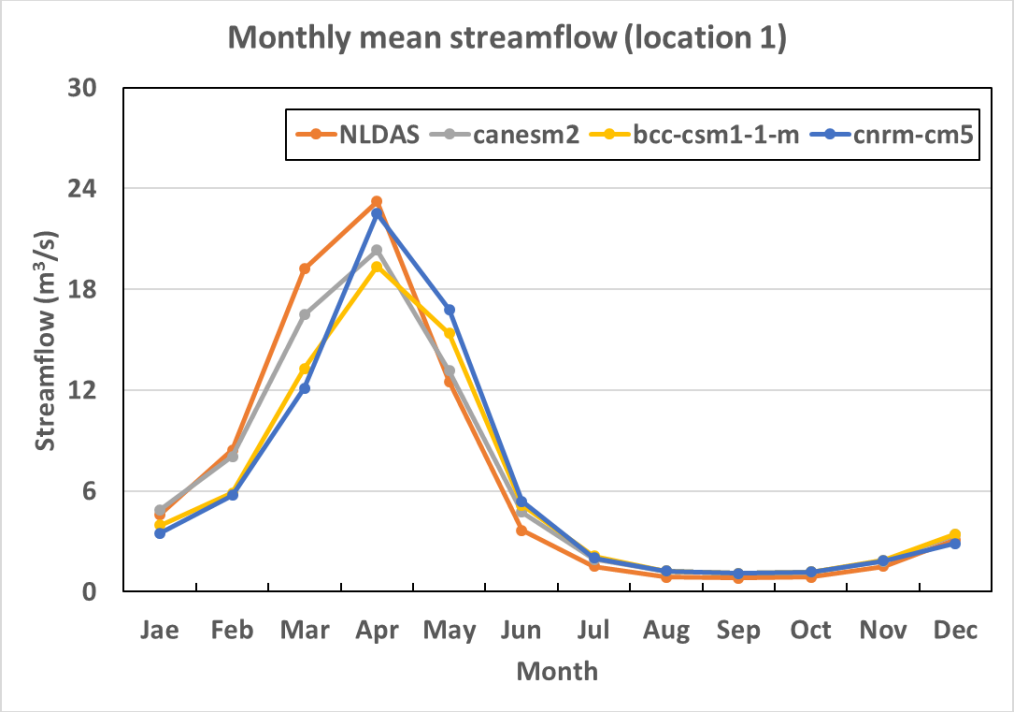
SDBC GCMs	Seasonal	Baseline (°C)	Variation (°C)								
			F1 (2021-2045)			F2 (2046-2070)			F3 (2071-2095)		
			RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5
Bcc-csm1-1m	Winter	-1.73	1.13	0.91	1.90	1.37	1.64	2.98	1.86	2.06	4.01
	Spring	7.36	0.95	0.99	1.24	1.37	1.31	2.09	1.16	2.37	3.65
	Summer	19.43	2.04	2.26	2.53	2.03	2.95	3.99	2.06	3.01	5.69
	Autumn	9.02	1.03	1.38	1.63	1.49	1.88	2.91	1.54	2.64	3.70
	Mean	8.52	1.29	1.39	1.83	1.56	1.95	2.99	1.65	2.52	4.26
Canesm2	Winter	-1.73	2.45	1.87	3.03	2.19	3.15	4.61	2.43	3.45	6.44
	Spring	7.36	2.66	2.46	2.81	2.53	3.62	4.55	3.10	3.37	5.42
	Summer	19.43	3.33	2.97	3.63	3.77	4.32	5.74	2.90	4.10	8.01
	Autumn	9.02	2.07	2.18	2.64	2.31	2.67	4.44	2.28	3.52	6.02
	Mean	8.52	2.63	2.37	3.03	2.70	3.44	4.83	2.68	3.61	6.47
Cnrm-cm5	Winter	-1.73	1.71	1.87	2.27	2.08	3.05	3.72	2.08	3.57	5.27
	Spring	7.36	0.39	1.28	1.45	0.82	2.18	2.58	1.22	2.56	4.32
	Summer	19.43	0.61	1.91	2.31	1.53	2.94	4.07	1.46	3.83	6.26
	Autumn	9.02	0.70	1.26	1.57	0.86	2.02	3.30	1.26	2.78	4.87
	Mean	8.52	0.85	1.58	1.90	1.32	2.55	3.42	1.50	3.19	5.18

6.3.6. Simulated streamflow evaluation for three SDBC GCMs

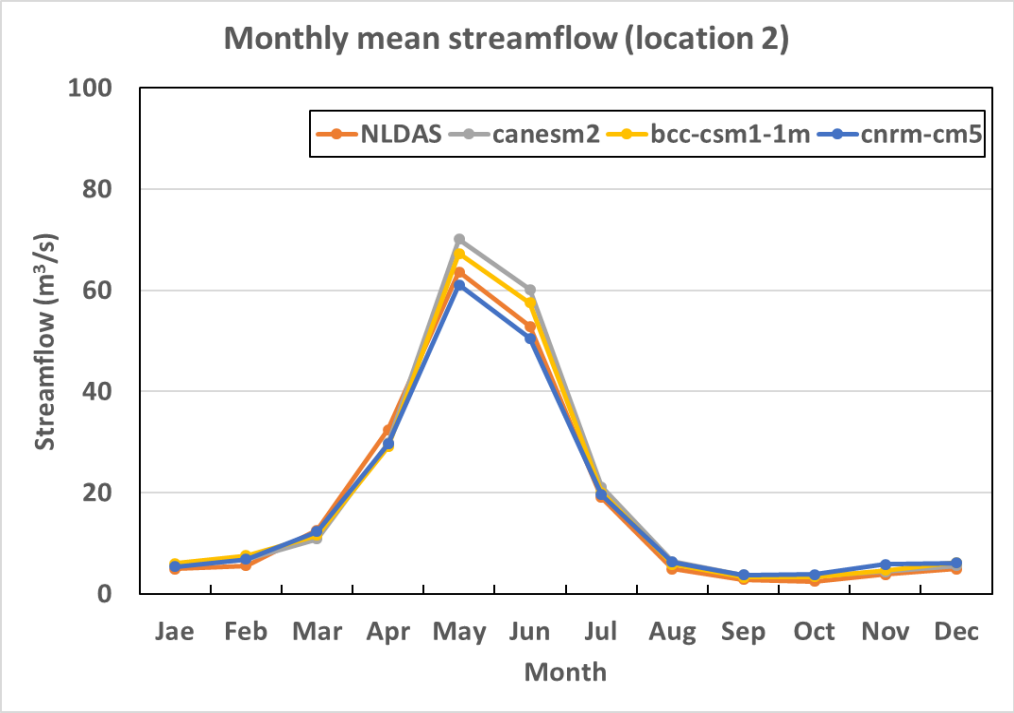
25-year monthly streamflow patterns for three SDBC GCMs and NLDAS-2 were compared during the baseline period (1981-2005) at calibration target points (CTPs). Three SDBC GCMs exhibited a homogenous trend during the historical streamflow simulation for CTP 1, 2, 3, 4, and 5 as represents in **Figure 6.9**. The difference of such simulation for three SDBC GCMs ranged from -7.92% to -2.53% at CTP 1, 0.47% to 8.61% at CTP 2, -2.16% to 2.98% at CTP 3, 3.53% to 8.20% at CTP 4, -0.18% to 4.30% at CTP 5, and -12.04% to -6.22% at CTP 6 (**Table 6.14**). The simulated streamflow difference between NLDAS-2 and SDBC GCMs was not exceeded ±12% for all calibration target points. Therefore, three SDBC GCMs well simulated the historical streamflow condition in the BWR based on the model performance evaluation using the difference results. Acceding to the results of streamflow simulation at the mouth of the watershed, canesm2 model was the most similar streamflow pattern for NLDAS-2 streamflow during the baseline period.

Table 6.14. Comparison of mean streamflow for three SDBC GCMs at all calibration target points.

Calibration target points (CTP)			Mean streamflow (m3/s)				Difference (%)		
No	Lat	Lon	NLDAS	Canesm2	Bcc-scm1-1m	Cnrm-cm5	Canesm2	Bcc-scm1-1m	Cnrm-cm5
1	43.65	-115.99	6.70	6.53	6.17	6.36	-2.53	-7.92	-5.09
2	43.50	-115.31	17.49	18.99	18.52	17.57	8.61	5.88	0.47
3	43.67	-115.73	32.94	33.58	33.94	32.20	1.96	2.98	-2.16
4	43.34	-115.48	23.36	25.27	24.66	24.18	8.20	5.58	3.53
5	43.53	-116.06	50.83	53.02	52.25	50.74	4.30	2.78	-0.18
6	43.78	-116.97	35.09	32.90	32.61	30.86	-6.22	-7.06	-12.04

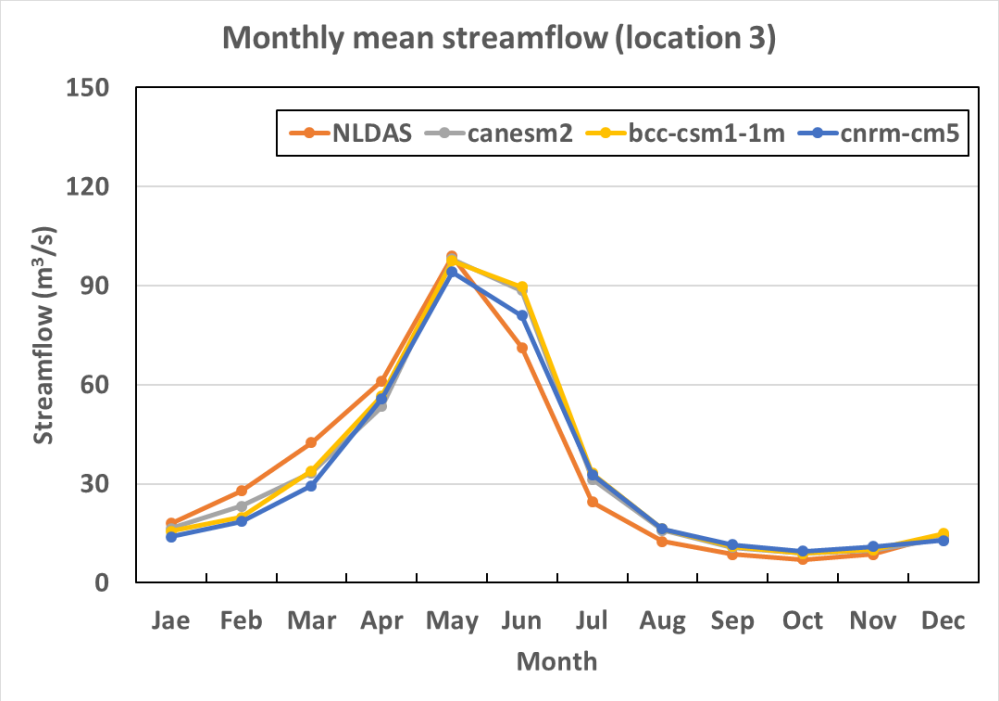


(a) Calibration target point 1

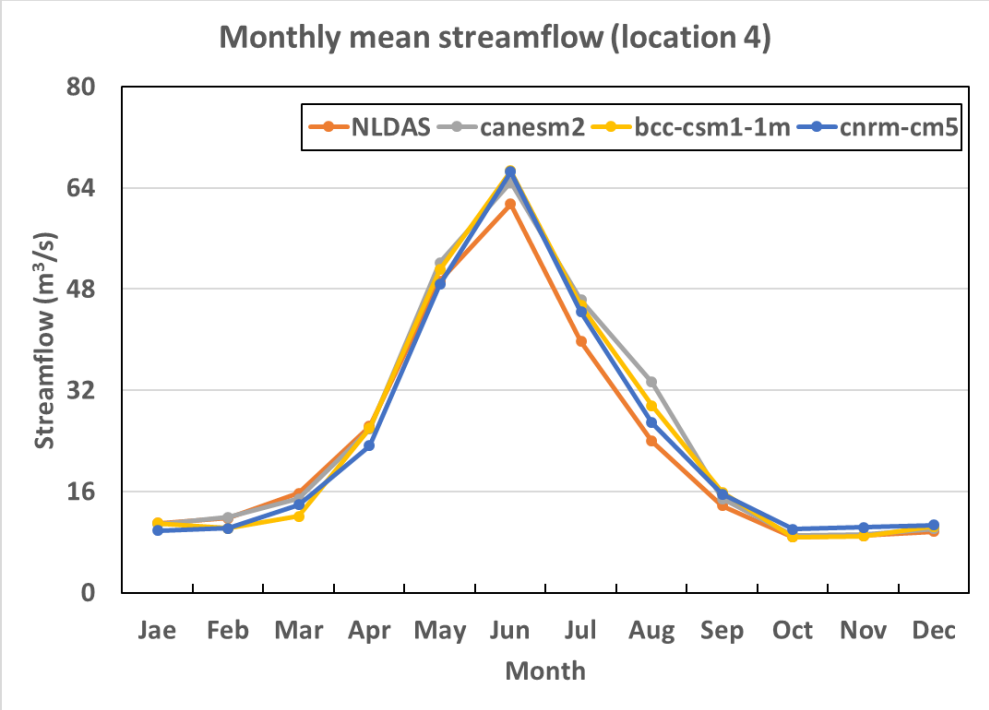


(b) Calibration target point 2

Figure 6.9. Comparison of monthly mean streamflow between NLDAS and SDBC GCMs at all calibration target points for the baseline period (1981-2005).

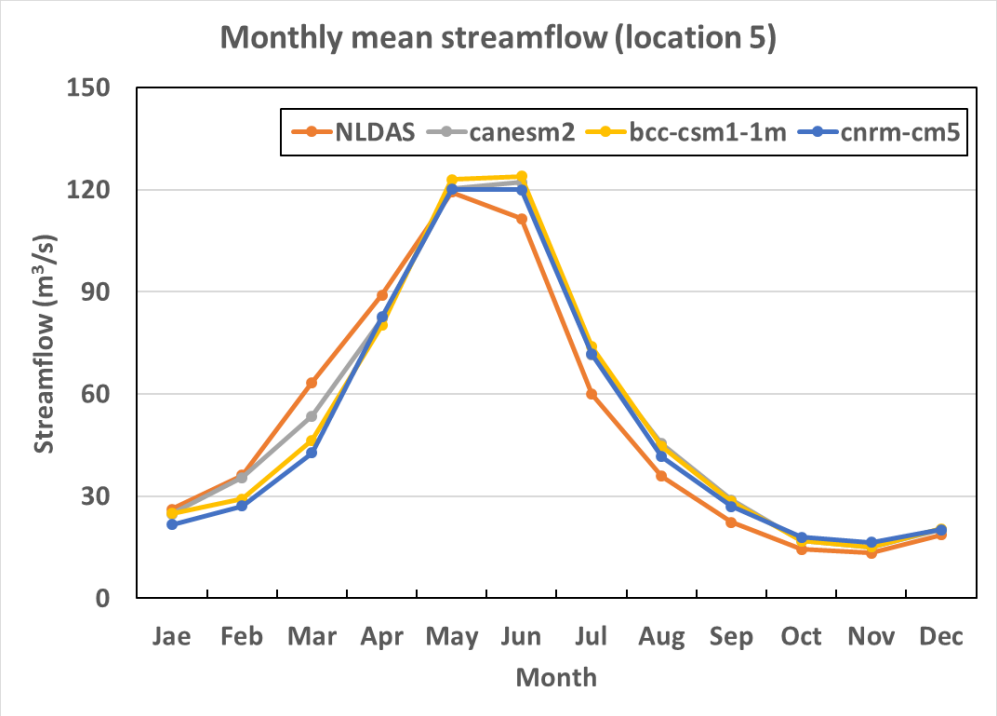


(c) Calibration target point 3

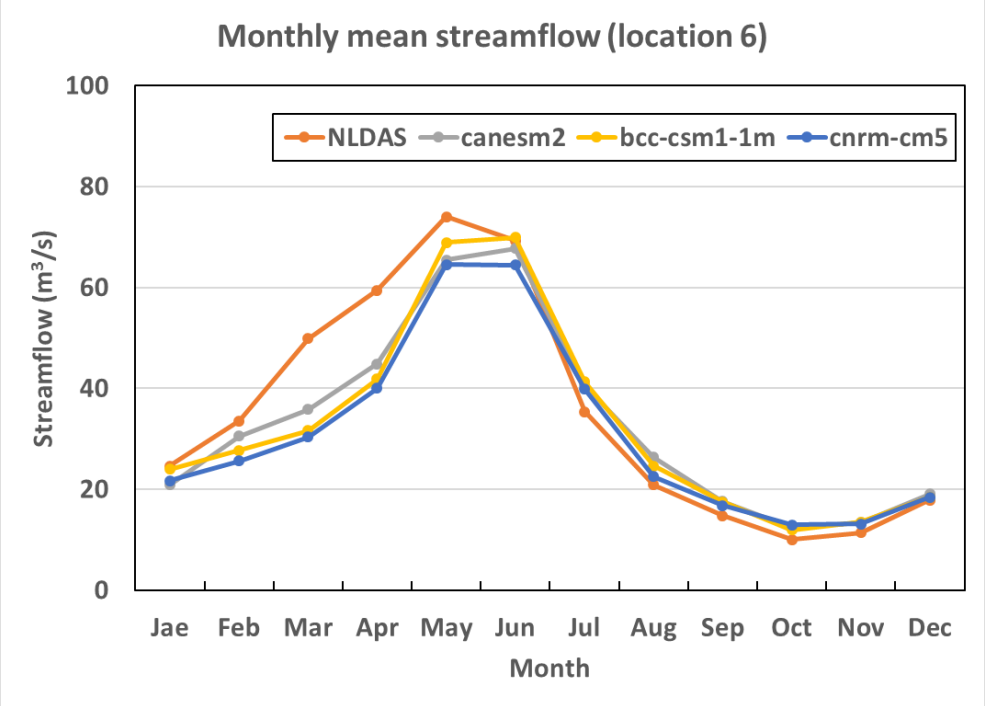


(d) Calibration target point 4

Figure 6.9. Comparison of monthly mean streamflow between NLDAS and SDBC GCMs at all calibration target points for the baseline period (1981-2005) (Cont.).



(e) Calibration target point 5



(f) Calibration target point 6

Figure 6.9. Comparison of monthly mean streamflow between NLDAS and SDBC GCMs at all calibration target points for the baseline period (1981-2005) (Cont.).

6.3.7. Impacts of changes in climate and LULC.

6.3.7.1. Maximum streamflow variation

To examine the potential combined impacts of climate and 2100 LULC changes on watershed streamflow, **Table 6.15** and **Figure 6.10-12** show 25-year monthly maximum streamflow for the baseline (1981-2005) and future periods (F1, F2, and F3). Bcc-csm1-1m model indicated the reduction of monthly maximum streamflow in spring and summer because of the decrease in spring and summer precipitation. Monthly maximum streamflow was mostly increased under all RCPs in the winter over the F1, F2, and F3 (**Figure 6.10**). In the BRW, due to monthly maximum streamflow was spring and summer-dominate and precipitation decreased, monthly maximum streamflow was reduced up to -27.18% under RCP2.6 and -16.44% under RCP8.5. However, RCP4.5 in the F1 increased monthly maximum streamflow by 16.17% relative to the baseline period. The combined impact of climate and 2100 LULC changes showed that monthly maximum streamflow was increased by urbanization coupled with climate change in summer and fall even though it was lower than the baseline period. The distinct increase in monthly maximum streamflow in spring and winter were observed comparing with climate change alone. 2100 LULC A2 under RCP4.5 simulated the greatest increase in the average monthly maximum streamflow by 33.33% in the F1 period relative to the baseline period.

Canesm2 model showed the similar monthly maximum streamflow pattern as bcc-csm1-1m model, but it was stimulated a significant increase in monthly maximum streamflow in winter during the F2 relative to the baseline period (**Figure 6.11**). Although monthly maximum streamflow was decreased under RCP2.6 in the F1, F2, and F3, it was increased by 1.01% in the F2 and 1.99% in the F3 under RCP8.5. RCP8.5 represented the biggest increase

of monthly maximum streamflow by 21.41% in F3 due to the increases in spring and winter precipitation. For the combined impact of climate and 2100 LULC changes, the increases in precipitation and the urban land were highly affected increasing of monthly maximum streamflow comparing with climate change alone. 2100 LULC A2 under RCP8.5 showed the greatest increase in monthly maximum streamflow by 46.66% in the F3 relative to the baseline period.

Cnrm-cm5 model also indicated similar variation pattern of monthly maximum streamflow as two SDBC GCMs (**Figure 6.12**). In future periods, cnrm-cm5 model indicated the increasing pattern of winter and spring precipitation under RCP4.5 and RCP8.5. Monthly maximum streamflow was increased by 19.04% under RCP4.5 and 13.04% under RCP8.5 relative to the baseline period. For the combined impact of climate and 2100 LILC changes, monthly maximum streamflow was increased under all RCPs in the future periods comparing with climate change condition. 2100 LULC A2 under RCP4.5 showed the greatest increase in monthly maximum streamflow by 39.24% in the F3 relative to the baseline period due to a large urban development and winter precipitation increasing.

3.3.7.2. Mean streamflow variation

Table 6.16 and **Figure 6.13 - 15** show 25-year monthly mean streamflow for the baseline and the future periods. Bcc-csm1-1m model indicated the reduction of monthly mean streamflow under all RCPs in winter and early spring over the future periods (**Figure 6.13**). Due to the increases in spring and winter temperature of bcc-csm1-1m model, monthly mean streamflow in early spring and winter was also increased. However, monthly mean streamflow was decreased in summer due to the reduction of summer precipitation and the

impact of early snow melting in the mountain area. Under RCP8.5, monthly mean streamflow was increased by 10.55 % in the F1 by the increase in mean annual precipitation. The combined impact of climate and 2100 LULC changes showed that monthly mean streamflow was increased by urbanization coupled with climate change comparing with climate change alone. 2100 LULC A2 under RCP8.5 indicated the greatest increase in monthly mean streamflow by 15.64% in the F1 relative to the baseline period.

Canesm2 model showed a large increase of monthly mean streamflow in spring, but monthly mean streamflow in summer and fall was decreased under all RCPs in F1, F2, and F3 due to the reduction of summer and fall precipitation and early snow melting in the mountain area relative to baseline period (**Figure 6.14**). Under all RCPs, monthly mean streamflow was decreased in the future periods. RCP8.5 in the F1 was the highest decrease in monthly mean streamflow by 24.08%. For the combined impact of climate and 2100 LULC changes, monthly mean streamflow was all decreased under all RCPs in the future periods except for 2100 LULC A2 under RCP8.5 in the F3 because monthly mean streamflow in summer and winter was highly decreased than the baseline period.

Cnrm-cm5 model indicated monthly mean streamflow was increased in winter and spring and decreased in summer and fall relative to the baseline period (**Figure 6.15**). Monthly mean streamflow was increased by 12.34-4.11% under RCP2.6, 1.91-3.19% under RCP45, and 1.66-2.14% under RCP8.5 in the future periods. The combined impact of climate and 2100 LULC changes indicated that monthly mean streamflow was varied by different LULC and climate change scenarios. 2100 LULC A2 under RCP2.6 showed the greatest increase in monthly mean streamflow by 17.42% in the F1 relative to baseline period because of the large increase in urbanization and spring and summer rainfall.

6.3.7.3. Minimum streamflow variation

Table 6.17 and **Figure 6.16 - 18** show 25-year monthly minimum streamflow for the baseline and the three future periods. Bcc-csm1-1m model indicated the increases in spring and summer minimum streamflow under RCP2.6 and RCP4.5 in the F1 and F2 (**Figure 6.16**). The largest increase of monthly minimum streamflow was about 13.93% under RCP8.5 in the F1. Canesm2 model showed similar monthly minimum streamflow variation as bcc-csm1-1m model. Higher minimum streamflow in spring and summer resulted in the increasing of the monthly minimum streamflow by 10.08% under RCP2.5 in the F2, and 13.93% and 10.30% under RCP4.5 and RCP8.5 in the F3 relative to baseline period, respectively (**Figure 6.17**). Cnrm-cm5 model indicated the increase in monthly minimum streamflow under all RCPs in the future periods relative to the baseline period (**Figure 6.18**). The combined impact of climate and 2100 LULC changes represented the slightly declining of monthly minimum streamflow comparing with climate change alone. However, LULC change rarely affected the variation of monthly minimum streamflow for three SDBC GCMs under all RCPs in the future periods.

6.3.7.4. Sediment load variation

Table 6.18 and **Figure 6.19 - 21** show 25-year monthly sediment loads for the baseline and the future periods. For Bcc-csm1-1m model, monthly sediment load was decreased under RCP2.6 and RCP4.5 in summer and fall and slightly increased under RCP8.5 in winter due to the increasing of winter streamflow (**Figure 6.19**). The average monthly sediment load in the future periods tended to decrease up to 30.73% under RCP2.6, 37.23% under RCP4.5, and 27.58% under RCP8.5. The combined impact of climate and 2100 LULC

change resulted in the large increase in monthly sediment load under all RCPs in winter and spring over the future periods like as monthly mean streamflow variation. 2100 LULC A2 under RCP8.5 in the F1 showed the greatest increase in the average monthly sediment load by 32.23% due to the higher urban area and monthly mean streamflow.

Canesm2 model showed similar results like as monthly sediment load variations of bcc-csm1-1m model (**Figure 6.20**). The average monthly sediment load was decreased up to 34.53% under RCP2.6, 35.40% under RCP4.5, and 35.83% under RCP8.5 in the future periods. The combined impact of climate and 2100 LULC changes resulted in the large increase of sediment loads under all RCPs in winter and spring over the future periods as well. 2100 LULC A2 under RCP85 indicated the biggest increase in the average monthly sediment load by 54.71% relative to the baseline period.

Cnrm-cm5 model indicated similar pattern of monthly sediment load variation as other SDBC GCMs (**Figure 6.21**). The average monthly sediment load was decreased up to 30.32% under RCP2.6, 24.97% under RCP4.5, and 25.13% under RCP8.5 in the future periods. The combined impact of climate and LULC changes resulted in the large increase of monthly sediment loads under all RCPs in winter and spring over the future periods. 2100 LULC A2 under RCP85 represented the biggest increase in the average monthly sediment load by 34.99% in the F3 relative to the baseline period. Urbanization offsets the decrease in sediment load that was associated with climate change scenarios. The sediment load increasing was more affected by urbanization than climate change relative to the baseline period.

6.3.7.5. TP load variation

Table 6.19 and **Figure 6.22 - 24** show 25-year monthly TP loads for the baseline and

the future periods. Bcc-csm1-1m model indicated that TP load was generally decreased in summer and spring and slightly increased in winter such as monthly mean streamflow variation under all RCPs over the future periods (**Figure 6.22**). The average monthly TP load tended to decrease by 1.72% under RCP2.6 in the F1, 9.95% under RCP4.5 in the F3, and 3.02 % under RCP8.5 in the F3, respectively. It was increased by 0.95-2.96% under RCP2.5 in the F2 and F3, 0.50-10.10% under RCP4.5 and RCP8.5 in the F1 relative to the baseline period. The combined impact of climate and LULC changes resulted in the large increase of TP loads under all RCPs in winter and spring over the future periods. Therefore, in four seasons, monthly TP load was all increased under three RCPs and LULC change scenarios. 2100 LULC A2 under RCP8.5 indicated the biggest increase in the average monthly TP load by 33.56% in the F1 due to the increases in urbanization and monthly mean streamflow.

Canesm2 model indicated similar monthly TP load variations like bcc-csm1-1m model (**Figure 6.23**), but the average monthly TP load was decreased under all RCPs in the future periods rather than Bcc-csm1-1m model. The combined impact of climate and LULC changes resulted in the large increase in monthly TP loads under all RCPs in winter and spring over the future periods. 2100 LULC A2 under RCP8.5 indicated the largest increase in the average monthly TP load by 21.16% in the F3 relative to the baseline period.

Overall, cnrm-cm5 model indicated that monthly TP load was increased in the spring and summer and was similar in the fall and winter under all RCPs over the future periods relative to the baseline period (**Figure 6.24**). The average monthly TP load tended to increase by 1.92-9.86% under RCP2.6, 1.50-6.26% under RCP4.5, and 1.33-4.75% under RCP8.5 in the future periods due to the increase in mean streamflow. The combined impact of climate and LULC changes represented the large increase in monthly TP loads under all RCPs in

winter and spring over the future periods. The average monthly TP load was increased under three RCPs and LULC change scenarios in the future periods. 2100 LULC A2 under RCP2.6 indicated the biggest increase in the average monthly TP load by 33.58% in the F1 because mean streamflow and urbanization were higher than other LULC change scenarios under RCP4.5 and RCP8.5. Consequently, monthly TP load profoundly affects by streamflow variation and urbanization.

6.3.7.6. TN load variation

Table 6.20 and **Figure 6.25 - 27** show 25-year monthly TN loads for the baseline and the future periods. Bcc-csm1-1m model indicated monthly TN load was decreased in spring, summer and fall but increased in winter under all RCPs over the future periods (**Figure 6.25**). The average monthly TN load tended to decrease up to 8.37% under RCP2.6, 20.99% under RCP4.5 and 5.01% under RCP8.5 in the future periods except for RCP8.5 in the F1 relative to the baseline period. The combined impact of climate and 2100 LULC changes indicated the significant increases in monthly TN loads under all RCPs in the future periods throughout the year, but the variation pattern of monthly TN loads was as similar the baseline period. 2100 LULC A2 under RCP8.5 indicated the biggest increase in average monthly TN load by 131.79% in the F1 period.

For canesm2 model, monthly TN load was increased in winter but decreased in spring, summer, and fall under all RCPs over the future periods (**Figure 6.26**). The average monthly TN load was decreased up to 18.42% under RCP2.6, 16.58% under RCP4.5, and 20.78% under RCP8.5 in the future periods. However, the combined impact of climate and 2100 LULC changes was simulated the large increasing of TN loads under all RCPs in the future

periods. 2100 LULC A2 under RCP2.6 indicated the largest increase in the average monthly TP load by 93.44% in the F1 relative to the baseline period.

Cnrm-cm5 model showed same increasing or decreasing patterns such as bcc-csm1-1m and canesm2 models under all RCPs in the future periods (**Figure 6.27**). The average monthly TN load in the future periods tended to increase up to 10.80% under RCP2.6, 3.84% under RCP4.5, and 1.34% under RCP8.5. The average monthly TN load was decreased by 4.20% under RCP2.6 and 2.04% under RCP4.5 in the F2 due to the reduction of monthly mean streamflow. The combined impact of climate and 2100 LULC changes showed the large increases in monthly TN loads under all RCPs in spring and winter over the future periods. 2100 LULC A2 under RCP2.5 indicated the highest increase in the average monthly TN load by 132.42% in the F1. Monthly TP load highly affects by monthly streamflow variations and urbanization as well as the increase in barren/mining land.

6.3.7.7. Water temperature variation

Table 6.21 and **Figure 6.28-30** show 25-year monthly water temperature for the baseline and the three future periods. Three SDBC GCMs indicated the increase in monthly water temperature under all RCPs in the future periods. In spring and winter, the increase in monthly water temperature was dominated due to monthly air temperature rise under RCP8.5. Canesm2 model indicated the largest increase in the average monthly water temperature by 4.49 ℃ under RCP85 in the F2 than other SDBC GCMs. The combined impact of climate and 2100 LULC changes was negligible because LULC change was rarely affected monthly water temperature variation.

Table 6.15. Variations of monthly maximum streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (m^3/s)	LULC condition	Emission pathway	Future period (m^3/s)			Variation (%)		
			F1*	F2*	F3*	F1	F2	F3
234.92	**bcc-csm1-1m**							
	2011	RCP2.6	191.64	181.57	171.06	-18.42	-22.71	-27.18
	2100 (a1b)		203.91	182.83	181.58	-13.20	-22.17	-22.71
	2100 (a2)		219.93	189.78	193.82	-6.38	-19.21	-17.49
	2100 (b1)		201.00	182.42	179.65	-14.44	-22.35	-23.53
	2011	RCP4.5	272.90	170.54	145.41	16.17	-27.41	-38.10
	2100 (a1b)		313.21	229.72	163.70	24.11	-15.50	-33.02
	2100 (a2)		291.56	198.50	157.36	33.33	-2.22	-30.32
	2100 (b1)		288.71	193.44	155.10	22.90	-17.66	-33.98
	2011	RCP8.5	215.93	219.24	196.29	-8.08	-6.67	-16.44
	2100 (a1b)		221.19	226.20	213.22	-5.84	-3.71	-9.24
	2100 (a2)		246.38	250.07	235.40	4.88	6.45	0.21
	2100 (b1)		216.66	225.46	209.80	-7.77	-4.03	-10.69
	canesm2							
	2011	RCP2.6	208.13	214.58	171.95	-11.40	-8.66	-26.81
	2100 (a1b)		208.80	217.78	201.63	-11.12	-7.30	-14.17
	2100 (a2)		228.03	219.55	230.62	-2.93	-6.54	-1.83
	2100 (b1)		208.42	216.67	197.13	-11.28	-7.77	-16.09
	2011	RCP4.5	156.06	237.30	239.59	-33.57	1.01	1.99
	2100 (a1b)		168.93	253.98	245.63	-28.09	8.11	4.56
	2100 (a2)		177.10	276.47	250.27	-24.61	17.69	6.54
	2100 (b1)		166.26	250.99	244.46	-29.23	6.84	4.06
	2011	RCP8.5	277.23	192.82	190.34	-17.92	-18.98	21.41
	2100 (a1b)		196.96	203.35	313.19	-16.16	-13.44	33.32
	2100 (a2)		204.26	212.53	344.54	-13.05	-9.53	46.66
	2100 (b1)		196.00	200.75	308.45	-16.57	-14.55	31.30
	cnrm-cm5							
	2011	RCP2.6	247.75	196.22	200.70	5.46	-16.47	-14.57
	2100 (a1b)		261.97	207.96	247.18	11.51	-11.48	5.22
	2100 (a2)		306.16	217.32	301.42	30.32	-7.49	28.31
	2100 (b1)		260.17	205.24	240.50	10.75	-12.63	2.37
	2011	RCP4.5	185.76	217.00	279.66	-20.93	-7.63	19.04
	2100 (a1b)		211.95	242.53	301.11	-9.78	3.24	28.18
	2100 (a2)		237.78	273.67	327.10	1.22	16.49	39.24
	2100 (b1)		207.54	237.76	298.24	-11.65	1.21	26.95
	2011	RCP8.5	161.29	238.32	265.54	-31.34	1.45	13.04
	2100 (a1b)		173.66	245.14	279.64	-26.08	4.35	19.04
	2100 (a2)		196.76	270.49	293.67	-16.24	15.14	25.01
	2100 (b1)		170.94	243.49	277.44	-27.23	3.65	18.10

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.16. Variations of monthly mean streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (m^3/s)	LULC condition	Emission pathway	Future period (m^3/s)			Variation (%)		
			F1*	F2*	F3*	F1	F2	F3
31.77	**bcc-csm1-1m**							
	2011	RCP2.6	28.04	30.92	29.39	-11.74	-2.69	-7.52
	2100 (a1b)		28.82	31.70	30.19	-9.31	-0.24	-5.00
	2100 (a2)		29.41	32.31	30.81	-7.43	1.69	-3.04
	2100 (b1)		28.70	31.58	30.07	-9.67	-0.61	-5.37
	2011	RCP4.5	30.47	29.60	23.98	-4.10	-6.83	-24.54
	2100 (a1b)		31.29	30.36	24.62	-1.51	-4.47	-22.51
	2100 (a2)		31.29	30.36	24.62	0.44	-2.58	-20.77
	2100 (b1)		31.91	30.95	25.17	-1.91	-4.83	-22.82
	2011	RCP8.5	35.13	30.53	29.85	10.55	-3.93	-6.04
	2100 (a1b)		36.02	31.28	30.56	13.37	-1.55	-3.81
	2100 (a2)		36.74	31.95	31.29	15.64	0.57	-1.52
	2100 (b1)		35.88	31.17	30.46	12.94	-1.91	-4.15
	canesm2							
	2011	RCP2.6	27.94	24.67	26.78	-12.07	-22.36	-15.71
	2100 (a1b)		28.61	25.32	27.44	-9.95	-20.30	-13.64
	2100 (a2)		29.20	25.88	28.02	-8.11	-18.56	-11.80
	2100 (b1)		28.51	25.22	27.34	-10.27	-20.62	-13.96
	2011	RCP4.5	25.66	25.94	26.21	-19.23	-18.37	-17.50
	2100 (a1b)		26.31	26.57	26.88	-17.19	-16.38	-15.41
	2100 (a2)		26.83	27.13	27.49	-15.56	-14.61	-13.49
	2100 (b1)		26.21	26.47	26.78	-17.51	-16.69	-15.73
	2011	RCP8.5	24.12	25.08	30.63	-24.08	-21.08	-3.61
	2100 (a1b)		24.72	25.67	31.32	-22.21	-19.20	-1.44
	2100 (a2)		25.24	26.29	32.09	-20.56	-17.26	0.98
	2100 (b1)		24.63	25.58	31.21	-22.50	-19.48	-1.77
	cnrm-cm5							
	2011	RCP2.6	35.69	29.19	33.08	12.34	-8.15	4.11
	2100 (a1b)		36.57	30.03	33.90	15.11	-5.48	6.70
	2100 (a2)		37.31	30.71	34.60	17.42	-3.35	8.91
	2100 (b1)		36.44	29.91	33.78	14.68	-5.88	6.31
	2011	RCP4.5	30.77	32.79	30.68	3.19	-3.43	1.91
	2100 (a1b)		33.63	31.51	33.21	5.84	-0.84	4.51
	2100 (a2)		34.29	32.20	33.94	7.91	1.33	6.82
	2100 (b1)		33.50	31.39	33.08	5.44	-1.22	4.12
	2011	RCP8.5	30.77	32.30	32.45	1.66	2.14	-1.44
	2100 (a1b)		32.01	32.48	31.56	0.74	2.23	-0.67
	2100 (a2)		32.66	33.24	32.34	2.79	4.60	1.79
	2100 (b1)		31.88	32.36	31.45	0.34	1.85	-1.03

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.17. Variations of monthly minimum streamflow by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (m^3/s)	LULC condition	Emission pathway	Future period (m^3/s)			Variation (%)		
			F1	F2	F3	F1	F2	F3
4.87	**bcc-csm1-1-m**							
	2011	RCP2.6	5.54	4.96	4.99	13.71	1.86	2.42
	2100 (a1b)		5.42	4.87	4.92	11.29	-0.06	0.90
	2100 (a2)		5.42	4.87	4.91	11.25	-0.13	0.78
	2100 (b1)		5.44	4.89	4.94	11.74	0.38	1.31
	2011	RCP4.5	5.50	5.53	5.36	12.93	13.48	10.02
	2100 (a1b)		5.38	5.41	5.24	10.40	10.98	7.62
	2100 (a2)		5.38	5.41	5.24	10.40	10.98	7.62
	2100 (b1)		5.38	5.41	5.24	10.35	11.01	7.50
	2011	RCP8.5	5.55	5.26	5.01	13.93	7.90	2.88
	2100 (a1b)		5.43	5.14	4.90	11.39	5.49	0.63
	2100 (a2)		5.42	5.14	4.90	11.28	5.41	0.48
	2100 (b1)		5.45	5.16	4.93	11.83	5.95	1.12
	canesm2							
	2011	RCP2.6	4.99	5.36	4.89	2.51	10.08	0.46
	2100 (a1b)		4.90	5.24	4.81	0.57	7.61	-1.23
	2100 (a2)		4.89	5.24	4.81	0.44	7.57	-1.36
	2100 (b1)		4.92	5.27	4.83	1.01	8.07	-0.78
	2011	RCP4.5	5.35	4.86	5.55	9.74	-0.31	13.93
	2100 (a1b)		5.23	4.77	5.42	7.33	-2.19	11.15
	2100 (a2)		5.23	4.76	5.43	7.28	-2.31	11.41
	2100 (b1)		5.25	4.79	5.44	7.78	-1.73	11.62
	2011	RCP8.5	4.62	4.83	5.37	-5.25	-0.90	10.30
	2100 (a1b)		4.56	4.73	5.27	-6.37	-2.89	8.17
	2100 (a2)		4.56	4.73	5.28	-6.41	-2.90	8.28
	2100 (b1)		4.57	4.75	5.29	-6.11	-2.43	8.62
	cnrm-cm5							
	2011	RCP2.6	5.01	6.59	5.03	2.87	35.32	3.20
	2100 (a1b)		4.92	6.40	4.94	0.91	31.31	1.47
	2100 (a2)		4.91	6.40	4.92	0.80	31.44	1.03
	2100 (b1)		4.94	6.42	4.97	1.36	31.76	1.93
	2011	RCP4.5	5.36	5.97	5.70	10.12	22.55	17.04
	2100 (a1b)		5.25	5.81	5.56	7.70	19.34	14.19
	2100 (a2)		5.25	5.82	5.56	7.67	19.36	14.11
	2100 (b1)		5.27	5.84	5.59	8.15	19.81	14.66
	2011	RCP8.5	5.26	5.44	5.64	8.06	11.73	15.74
	2100 (a1b)		5.18	5.29	5.51	6.27	8.57	13.04
	2100 (a2)		5.18	5.29	5.51	6.26	8.51	13.03
	2100 (b1)		5.20	5.31	5.53	6.73	9.02	13.51

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.18. Variations of monthly sediment load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (100×ton/ month)	LULC condition	Emission pathway	Future period (100×ton/month)			Variation (%)		
			F1*	F2*	F3*	F1	F2	F3
23.04	**bcc-csm1-1-m**							
	2011	RCP2.6	16.18	16.44	15.96	-29.76	-28.63	-30.73
	2100 (a1b)		22.50	22.17	22.47	-2.35	-3.76	-2.49
	2100 (a2)		26.99	26.58	27.12	17.15	15.39	17.71
	2100 (b1)		26.40	25.62	26.17	14.58	11.18	13.60
	2011	RCP4.5	16.43	15.52	14.46	-28.68	-32.66	-37.23
	2100 (a1b)		22.45	20.97	20.48	-2.58	-8.99	-11.11
	2100 (a2)		26.88	25.25	24.76	16.67	9.58	7.49
	2100 (b1)		22.45	20.97	20.48	-2.58	-8.99	-11.11
	2011	RCP8.5	19.02	16.69	16.93	-17.46	-27.58	-26.54
	2100 (a1b)		21.59	21.22	21.48	-6.28	-7.90	-6.75
	2100 (a2)		30.47	27.42	28.30	32.23	19.03	22.85
	2100 (b1)		25.72	22.97	23.98	11.62	-0.29	4.09
	canesm2							
	2011	RCP2.6	15.81	15.08	15.56	-31.36	-34.53	-32.45
	2100 (a1b)		22.04	21.04	20.65	-4.33	-8.69	-10.38
	2100 (a2)		26.40	25.62	26.17	14.58	11.18	13.60
	2100 (b1)		21.15	20.15	20.83	-8.22	-12.52	-9.59
	2011	RCP4.5	14.88	16.24	16.59	-35.40	-29.51	-28.00
	2100 (a1b)		21.12	22.51	23.59	-8.31	-2.29	2.41
	2100 (a2)		25.42	26.80	28.43	10.34	16.30	23.41
	2100 (b1)		20.28	21.55	22.58	-11.96	-6.45	-2.00
	2011	RCP8.5	14.79	16.78	20.94	-35.83	-27.17	-9.12
	2100 (a1b)		21.19	24.10	31.01	-8.02	4.58	34.61
	2100 (a2)		25.47	28.66	35.64	10.56	24.40	54.71
	2100 (b1)		20.34	23.05	29.51	-11.71	0.03	28.08
	cnrm-cm5							
	2011	RCP2.6	18.57	16.05	16.80	-19.41	-30.32	-27.10
	2100 (a1b)		24.87	22.05	22.20	7.92	-4.29	-3.65
	2100 (a2)		29.53	27.08	27.79	28.16	17.52	20.64
	2100 (b1)		23.85	21.10	22.32	3.53	-8.41	-3.13
	2011	RCP4.5	17.48	17.29	17.90	-24.15	-24.97	-22.30
	2100 (a1b)		23.83	23.60	25.36	3.43	2.42	10.09
	2100 (a2)		28.43	28.32	30.19	23.41	22.94	31.06
	2100 (b1)		22.88	22.57	24.19	-0.70	-2.06	5.00
	2011	RCP8.5	17.25	17.86	18.25	-25.13	-22.48	-20.79
	2100 (a1b)		23.47	24.95	26.29	1.85	8.29	14.13
	2100 (a2)		28.13	29.78	31.10	22.08	29.24	34.99
	2100 (b1)		22.52	23.84	25.12	-2.27	3.46	9.04

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.19. Variations of monthly TP load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (ton/month)	LULC condition	Emission pathway	Future period (ton/month)			Variation (%)		
			F1*	F2*	F3*	F1	F2	F3
11.25	**bcc-csm1-1-m**							
	2011	RCP2.6	11.06	11.58	11.36	-1.72	2.96	0.95
	2100 (a1b)		12.55	13.18	12.90	11.56	17.10	14.64
	2100 (a2)		13.41	14.06	13.77	19.20	24.95	22.40
	2100 (b1)		13.10	12.62	12.81	16.44	12.19	13.82
	2011	RCP4.5	11.46	11.34	10.13	1.84	0.77	-9.95
	2100 (a1b)		13.05	12.89	11.43	15.96	14.56	1.54
	2100 (a2)		13.91	13.77	12.22	23.64	22.37	8.60
	2100 (b1)		12.80	12.65	11.23	13.75	12.40	-0.21
	2011	RCP8.5	12.39	11.31	10.91	10.10	0.50	-3.02
	2100 (a1b)		14.11	12.86	12.38	25.38	14.27	10.02
	2100 (a2)		15.03	13.74	13.30	33.56	22.12	18.16
	2100 (b1)		13.84	12.62	12.15	22.97	12.14	8.01
	canesm2							
	2011	RCP2.6	10.78	10.40	10.58	-4.20	-7.59	-5.97
	2100 (a1b)		12.25	11.77	11.99	8.88	4.62	6.52
	2100 (a2)		13.10	12.62	12.81	16.44	12.19	13.82
	2100 (b1)		12.03	11.56	11.77	6.88	2.76	4.60
	2011	RCP4.5	10.43	10.35	10.49	-7.34	-7.97	-6.75
	2100 (a1b)		11.83	11.75	11.98	5.10	4.44	6.44
	2100 (a2)		12.66	12.67	12.68	12.49	12.56	12.69
	2100 (b1)		11.61	11.54	11.71	3.20	2.57	4.05
	2011	RCP8.5	10.06	10.25	11.13	-10.61	-8.87	-1.05
	2100 (a1b)		11.39	11.66	12.70	1.19	3.63	12.88
	2100 (a2)		12.21	12.53	13.63	8.50	11.37	21.16
	2100 (b1)		11.18	11.44	12.15	-0.60	1.66	8.01
	cnrm-cm5							
	2011	RCP2.6	12.36	11.47	11.77	9.86	1.92	4.58
	2100 (a1b)		14.10	13.06	13.32	25.30	16.08	18.39
	2100 (a2)		15.03	13.91	14.15	33.58	23.61	25.79
	2100 (b1)		13.83	12.81	13.13	22.88	13.89	16.71
	2011	RCP4.5	11.96	11.42	11.67	6.26	1.50	3.71
	2100 (a1b)		13.61	13.04	13.31	20.95	15.87	18.31
	2100 (a2)		14.52	13.95	14.21	29.04	24.02	26.31
	2100 (b1)		13.35	12.79	13.06	18.65	13.68	16.09
	2011	RCP8.5	11.79	11.56	11.40	4.75	2.70	1.33
	2100 (a1b)		13.25	13.08	12.90	17.78	16.29	14.66
	2100 (a2)		14.16	14.01	13.84	25.85	24.51	22.98
	2100 (b1)		13.00	12.84	12.67	15.52	14.08	12.56

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.20. Variations of monthly TN load by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

Base [a] (ton/month)	LULC condition	Emission pathway	Future period (ton/month)			Variation (%)		
			F1*	F2*	F3*	F1	F2	F3
94.54	**bcc-csm1-1-m**							
	2011	RCP2.6	86.63	93.53	89.15	-8.37	-1.07	-5.70
	2100 (a1b)		181.89	195.83	188.33	92.39	107.13	99.20
	2100 (a2)		186.91	200.89	193.37	97.69	112.49	104.53
	2100 (b1)		182.88	168.59	173.22	93.44	78.32	83.21
	2011	RCP4.5	93.56	90.97	74.69	-1.04	-3.78	-20.99
	2100 (a1b)		93.56	90.97	74.69	110.68	101.13	62.52
	2100 (a2)		203.31	195.41	159.23	115.05	106.69	68.42
	2100 (b1)		183.50	175.61	142.72	94.09	85.74	50.96
	2011	RCP8.5	104.13	92.31	89.81	10.14	-2.36	-5.01
	2100 (a1b)		214.10	187.56	168.12	126.46	98.39	77.82
	2100 (a2)		219.14	193.29	177.64	131.79	104.45	87.89
	2100 (b1)		197.98	173.78	157.15	109.40	83.81	66.22
	canesm2							
	2011	RCP2.6	84.27	77.13	81.24	-10.87	-18.42	-14.07
	2100 (a1b)		177.53	163.01	168.06	87.78	72.42	77.76
	2100 (a2)		182.88	168.59	173.22	93.44	78.32	83.21
	2100 (b1)		164.52	150.81	155.89	74.01	59.51	64.88
	2011	RCP4.5	78.97	78.87	80.33	-16.47	-16.58	-15.03
	2100 (a1b)		168.61	162.37	170.80	78.34	71.74	80.66
	2100 (a2)		173.73	173.70	172.16	83.75	83.73	82.10
	2100 (b1)		155.91	150.98	152.94	64.91	59.69	61.77
	2011	RCP8.5	74.90	78.83	92.57	-20.78	-16.62	-2.09
	2100 (a1b)		153.37	160.98	166.40	62.22	70.27	76.01
	2100 (a2)		159.26	158.59	174.69	68.45	67.74	84.77
	2100 (b1)		142.48	141.11	157.15	50.71	49.25	66.22
	cnrm-cm5							
	2011	RCP2.6	104.76	90.58	96.67	10.80	-4.20	2.25
	2100 (a1b)		214.09	192.38	191.64	126.45	103.48	102.70
	2100 (a2)		219.74	192.20	218.52	132.42	103.29	131.13
	2100 (b1)		198.42	177.89	184.73	109.87	88.16	95.39
	2011	RCP4.5	98.17	92.62	95.59	3.84	-2.04	1.11
	2100 (a1b)		203.34	191.61	194.77	115.08	102.67	106.01
	2100 (a2)		208.74	198.03	200.88	120.78	109.46	112.47
	2100 (b1)		188.04	178.09	181.23	98.89	88.37	91.69
	2011	RCP8.5	95.81	95.59	95.39	1.34	1.11	0.89
	2100 (a1b)		200.19	191.53	175.51	111.74	102.58	85.64
	2100 (a2)		205.39	198.13	183.81	117.24	109.56	94.41
	2100 (b1)		185.00	178.11	164.25	95.67	88.39	73.73

[a] Base: Baseline period: 1981-2005.

* F1: 2021-2045 period.

* F2: 2046-2070 period.

* F3: 2071-2095 period.

Table 6.21. Variations of monthly mean water temperature by climate change and the combined impact of climate and 2100 LULC changes relative to the baseline period (1981-2005).

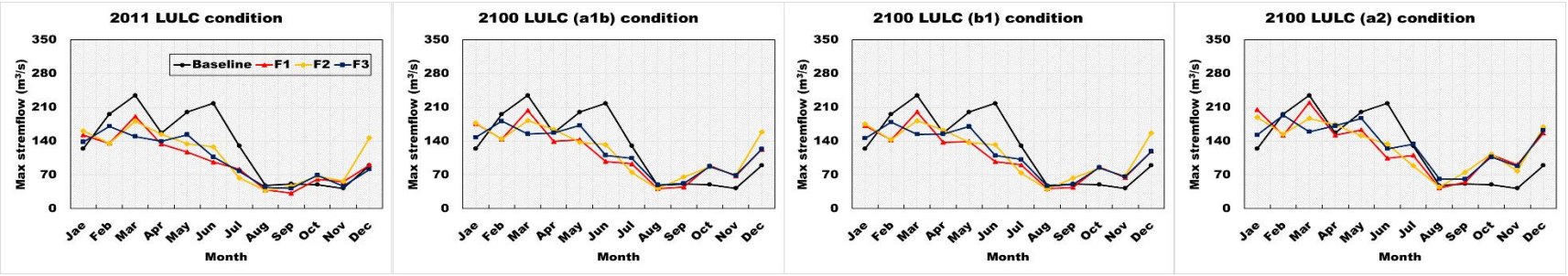
Base [a] (°C)	LULC condition	Emission pathway	Future period (°C)			Variation (°C)		
			F1*	F2*	F3*	F1	F2	F3
9.90	**bcc-csm1-1m**							
	2011	RCP2.6	10.82	10.93	10.97	0.91	1.03	1.06
	2100 (a1b)		9.78	10.95	11.35	0.90	1.02	1.06
	2100 (a2)		10.82	10.94	10.98	0.92	1.04	1.08
	2100 (b1)		9.78	10.96	11.36	0.90	1.02	1.06
	2011	RCP4.5	10.79	11.13	11.62	0.89	1.22	1.71
	2100 (a1b)		10.78	11.12	11.62	0.87	1.22	1.72
	2100 (a2)		10.79	11.14	11.63	0.88	1.23	1.73
	2100 (b1)		10.78	11.12	11.62	0.87	1.22	1.72
	2011	RCP8.5	11.00	11.74	12.66	1.09	1.83	2.76
	2100 (a1b)		10.99	11.74	12.68	1.09	1.84	2.77
	2100 (a2)		11.01	11.75	12.66	1.11	1.85	2.76
	2100 (b1)		10.99	11.74	12.68	1.09	1.84	2.77
	canesm2							
	2011	RCP2.6	11.60	11.83	11.77	1.69	1.93	1.87
	2100 (a1b)		11.60	11.83	11.78	1.70	1.93	1.87
	2100 (a2)		11.61	11.84	11.79	1.71	1.94	1.89
	2100 (b1)		11.60	11.84	11.78	1.70	1.93	1.87
	2011	RCP4.5	11.45	12.24	12.42	1.55	2.33	2.52
	2100 (a1b)		11.45	12.25	12.43	1.55	2.34	2.53
	2100 (a2)		11.47	12.23	12.41	1.56	2.33	2.51
	2100 (b1)		11.45	12.25	12.44	1.55	2.34	2.54
	2011	RCP8.5	11.75	13.10	14.40	1.85	3.19	4.49
	2100 (a1b)		11.76	13.10	14.43	1.85	3.20	4.53
	2100 (a2)		11.77	13.13	14.44	1.87	3.22	4.54
	2100 (b1)		11.76	13.12	12.68	1.85	3.22	2.77
	cnrm-cm5							
	2011	RCP2.6	11.00	11.11	11.24	1.09	1.20	1.34
	2100 (a1b)		10.99	11.11	11.25	1.09	1.20	1.34
	2100 (a2)		11.01	11.14	11.53	1.11	1.24	1.63
	2100 (b1)		10.99	11.11	11.24	1.09	1.20	1.34
	2011	RCP4.5	10.86	11.49	11.86	0.95	1.58	1.96
	2100 (a1b)		10.85	11.49	11.87	0.95	1.59	1.97
	2100 (a2)		10.87	11.51	11.89	0.97	1.61	1.99
	2100 (b1)		10.85	11.49	11.87	0.95	1.59	1.97
	2011	RCP8.5	10.93	11.97	13.15	1.02	2.06	3.24
	2100 (a1b)		10.95	12.00	13.18	1.05	2.09	3.28
	2100 (a2)		10.97	12.02	13.20	1.06	2.11	3.29
	2100 (b1)		10.95	12.00	13.18	1.05	2.09	3.28

[a] Base: Baseline period: 1981-2005.

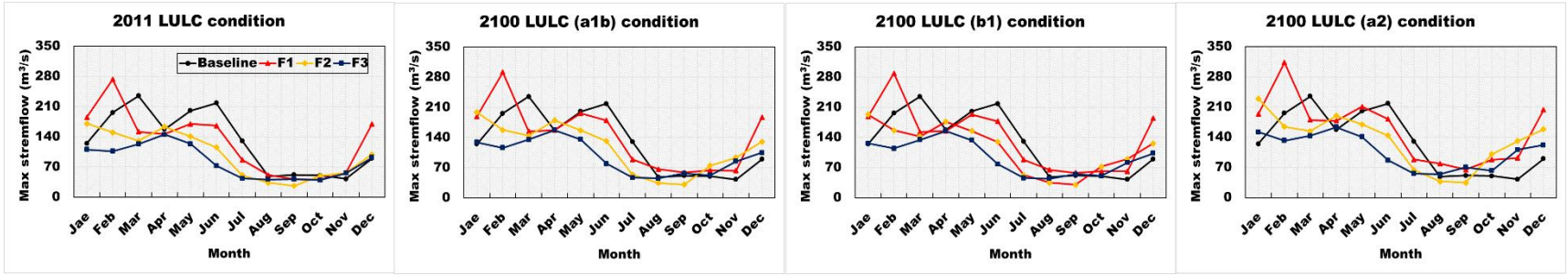
* F1: 2021-2045 period.

* F2: 2046-2070 period.

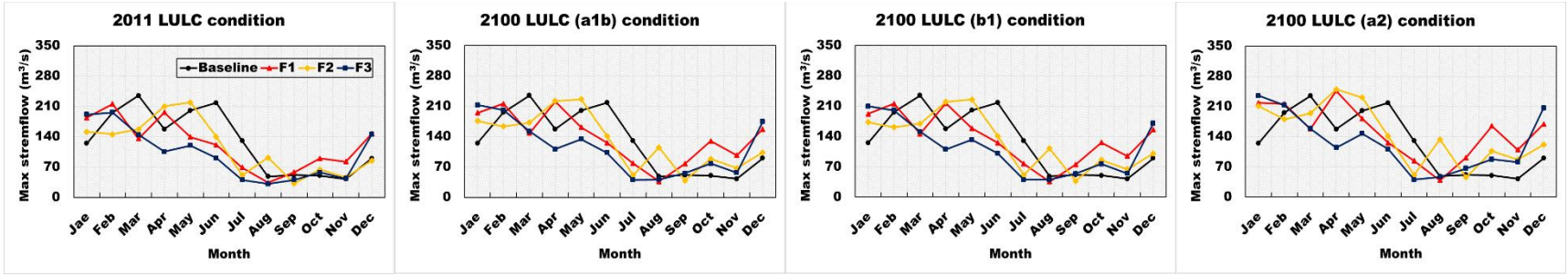
* F3: 2071-2095 period.



(a) RCP2.6

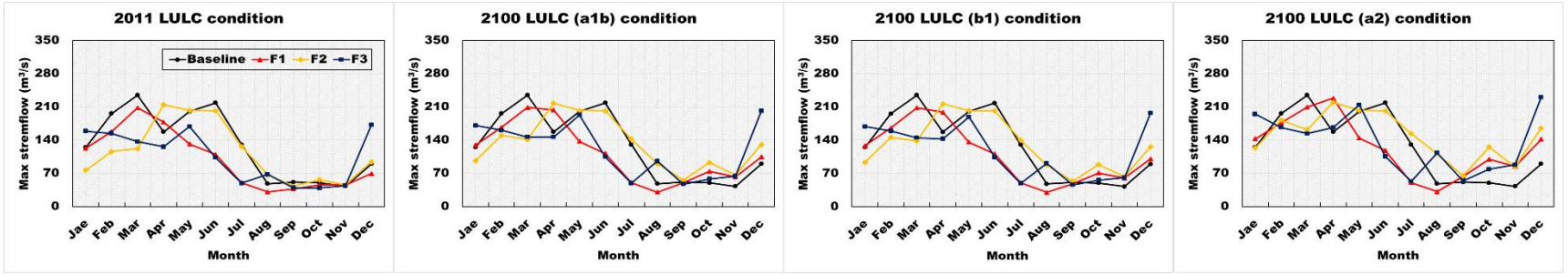


(b) RCP4.5

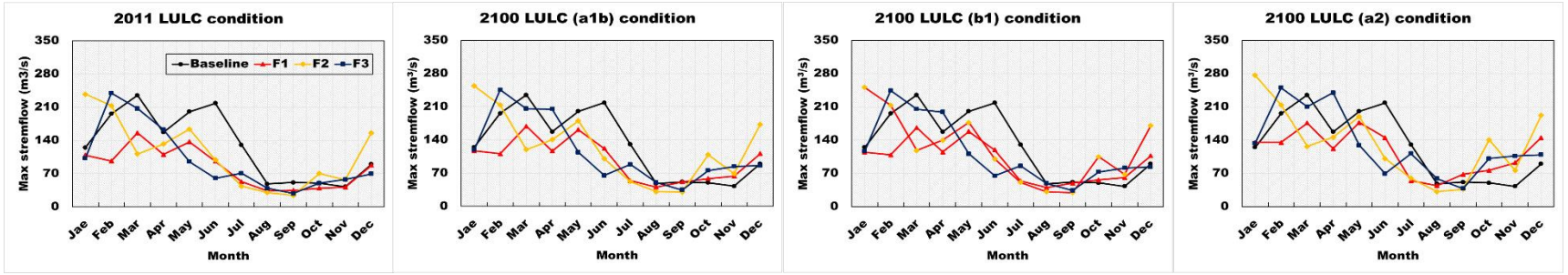


(c) RCP8.5

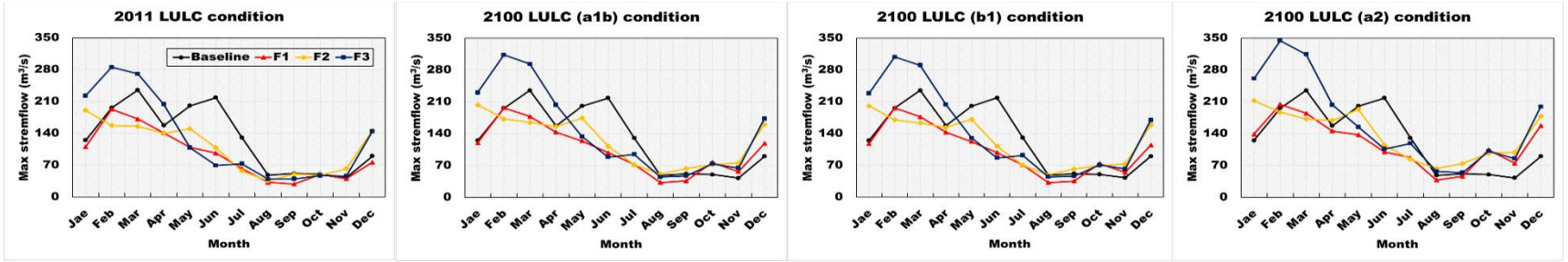
Figure 6.10. Comparison of monthly maximum streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

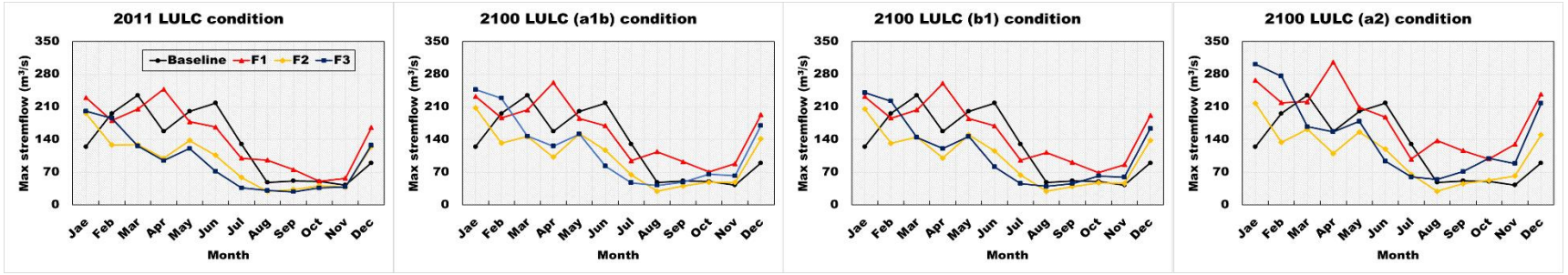


(b) RCP4.5

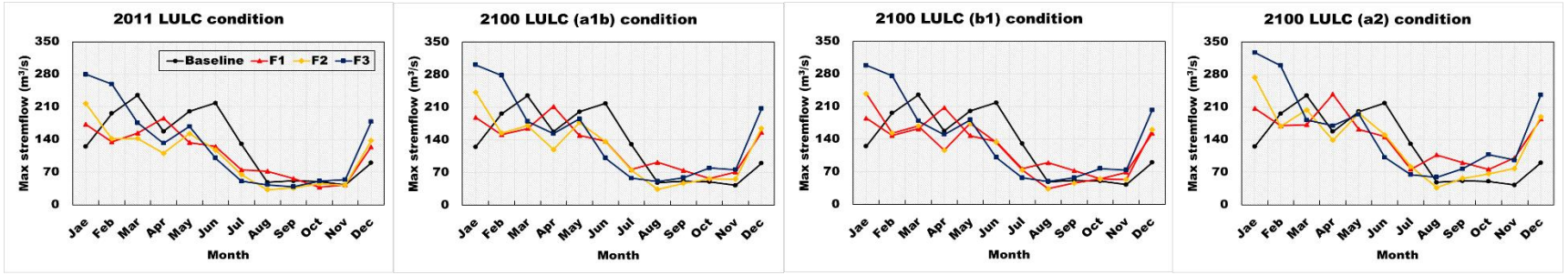


(c) RCP8.5

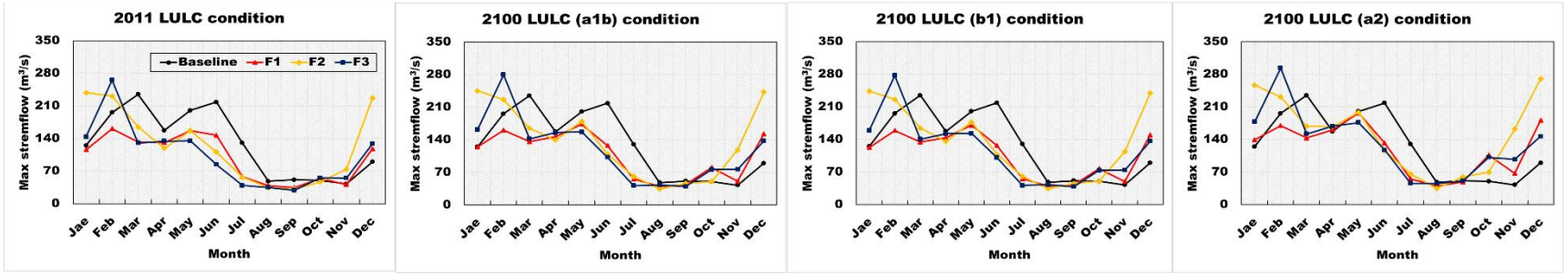
Figure 6.11. Comparison of monthly maximum streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

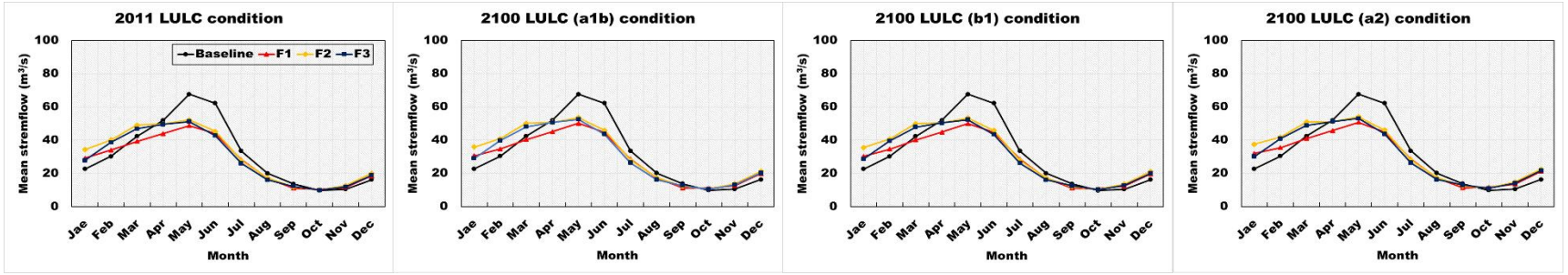


(b) RCP4.5

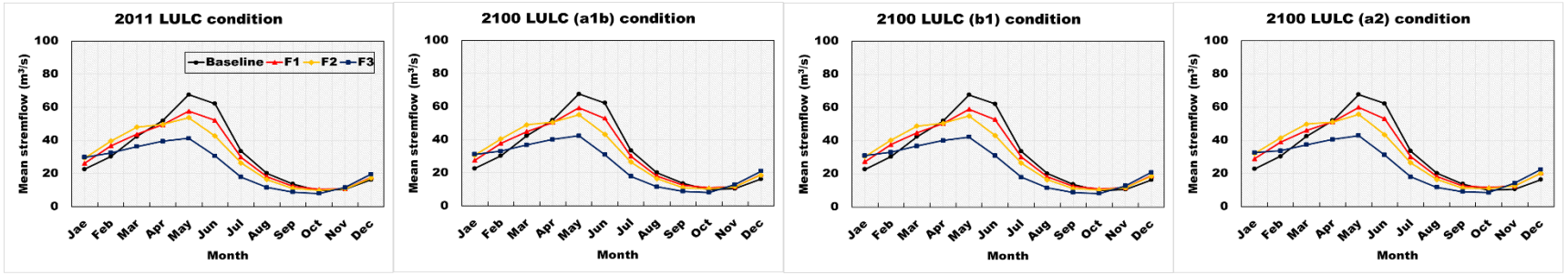


(c) RCP8.5

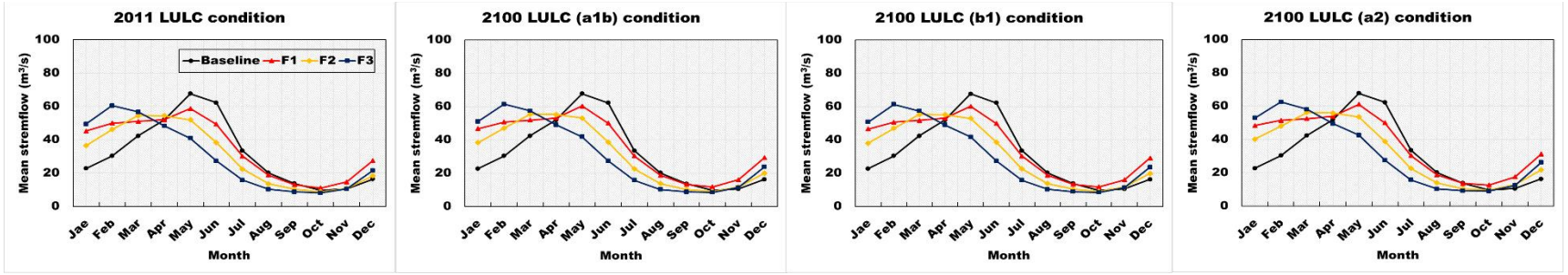
Figure 6.12. Comparison of monthly maximum streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

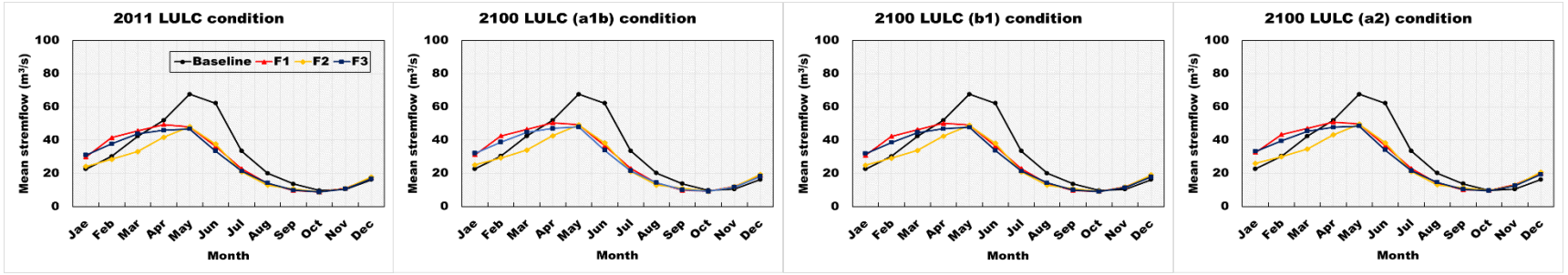


(b) RCP4.5

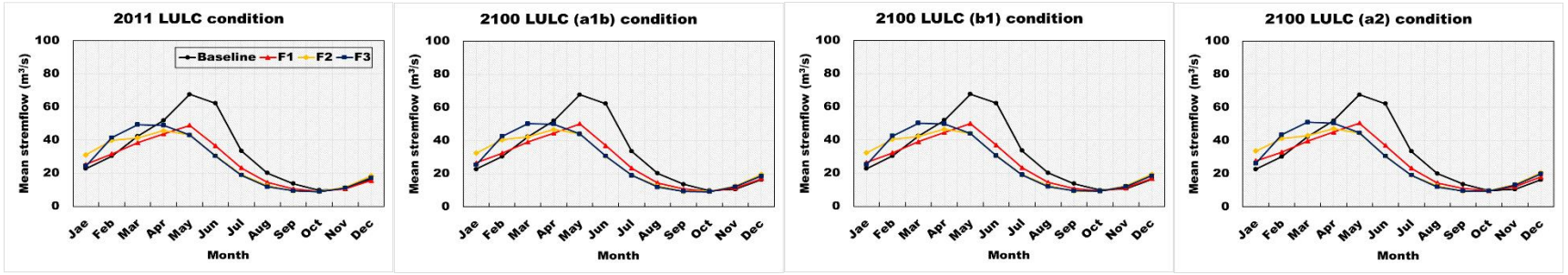


(c) RCP8.5

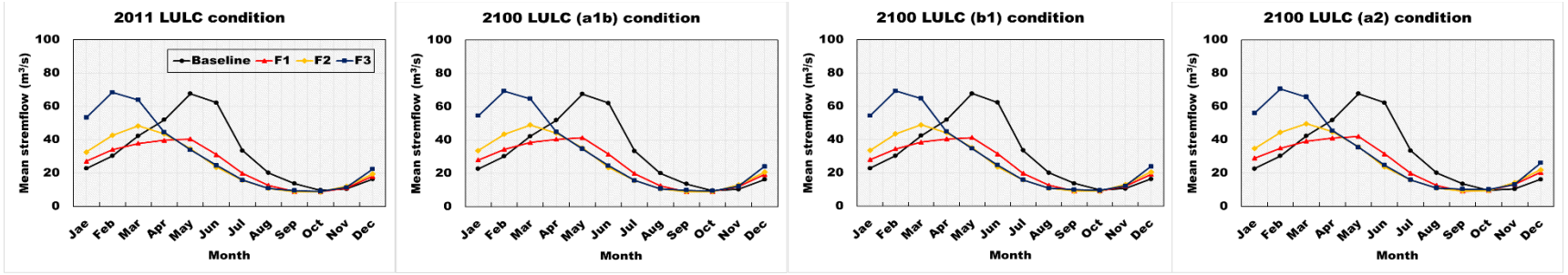
Figure 6.13. Comparison of monthly mean streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

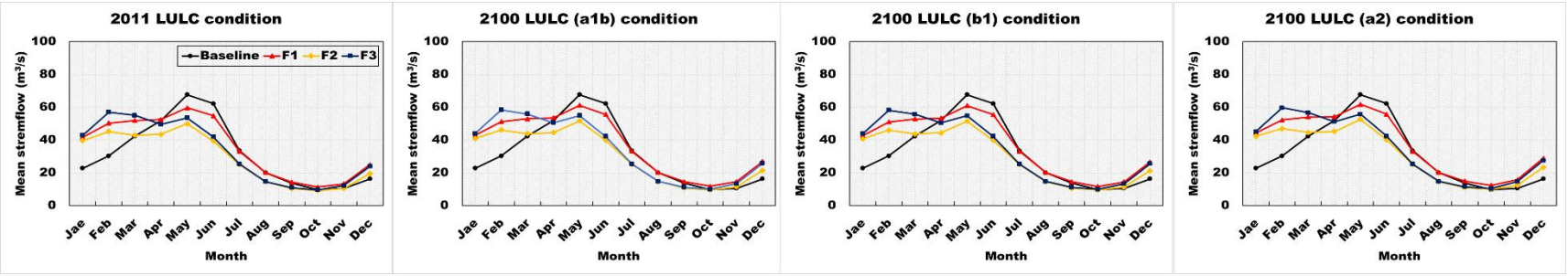


(b) RCP4.5

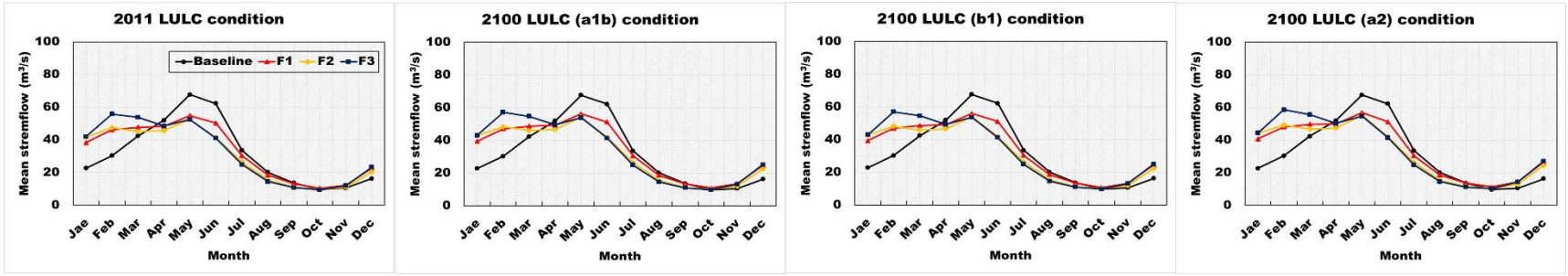


(c) RCP8.5

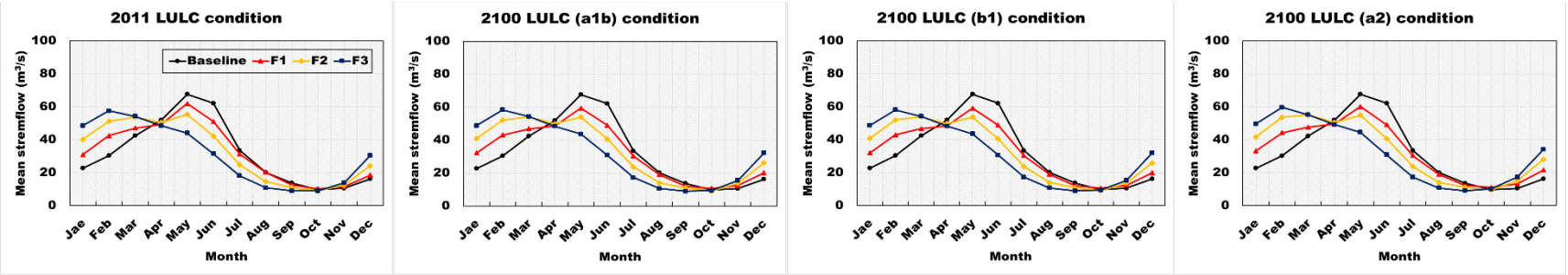
Figure 6.14. Comparison of monthly mean streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

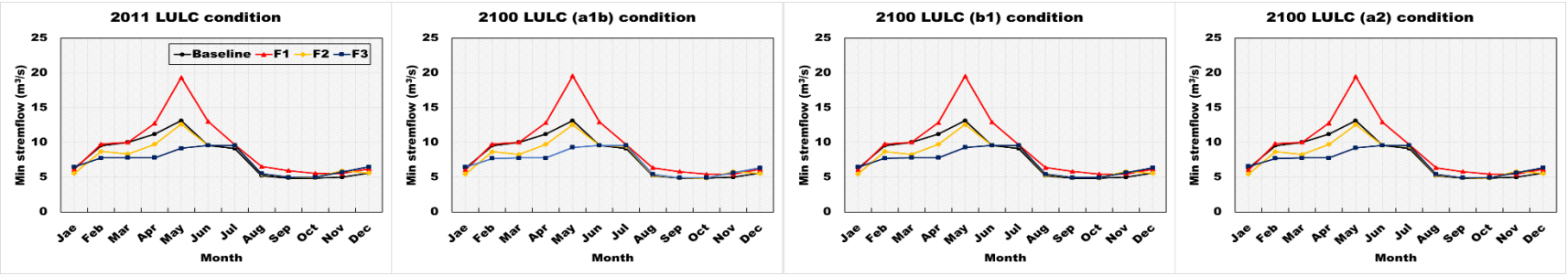


(b) RCP4.5

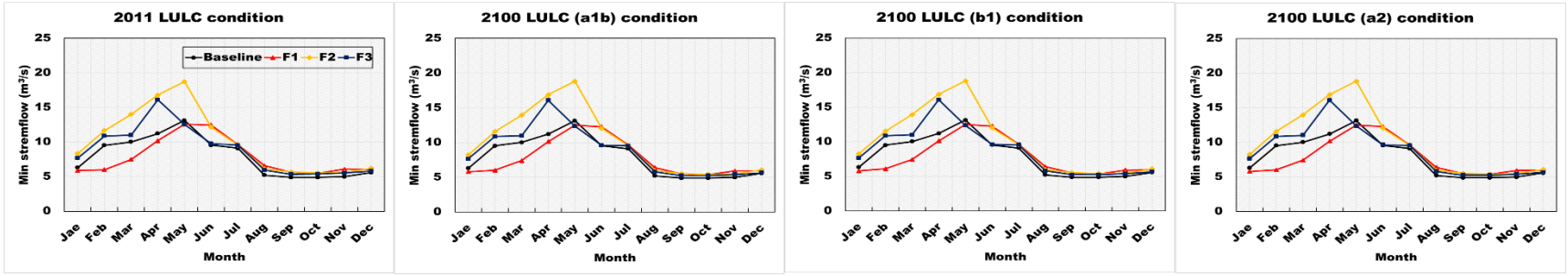


(c) RCP8.5

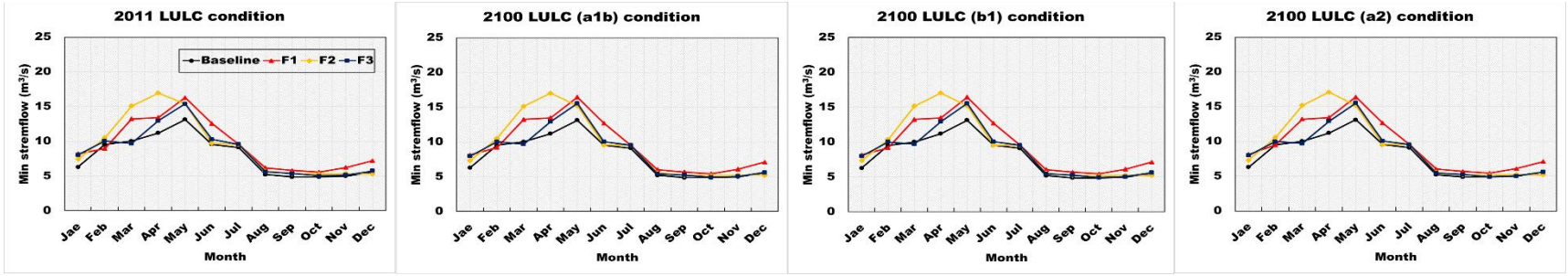
Figure 6.15. Comparison of monthly mean streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

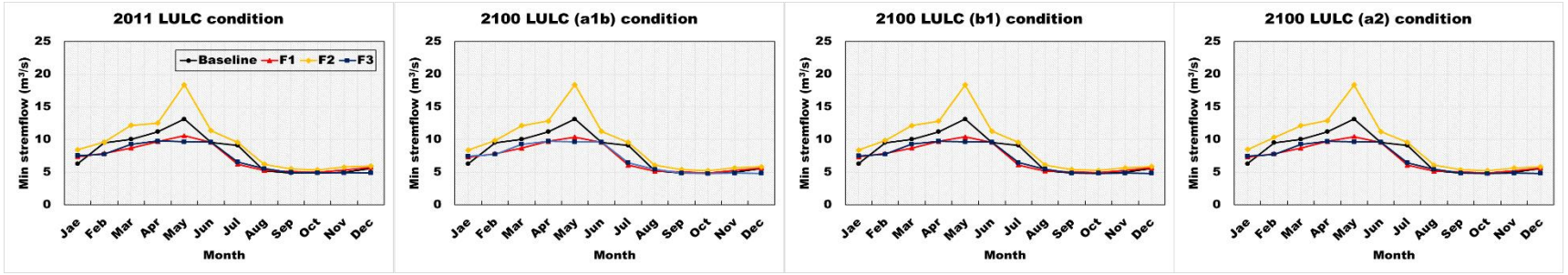


(b) RCP4.5

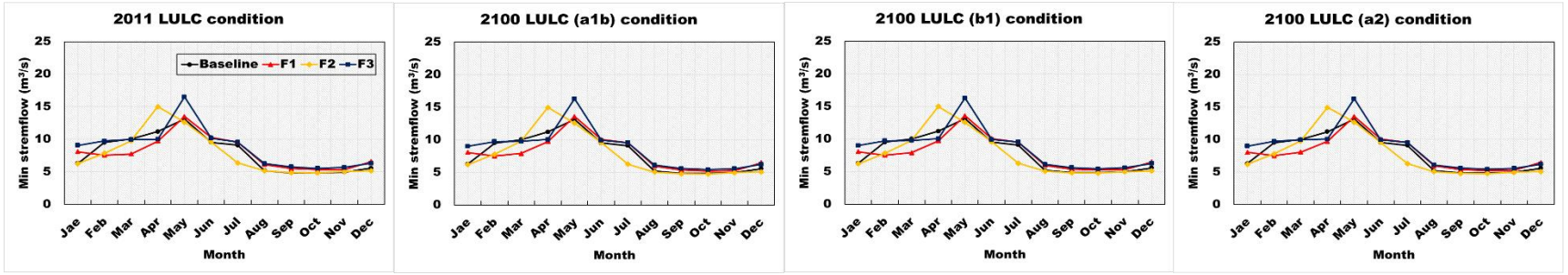


(c) RCP8.5

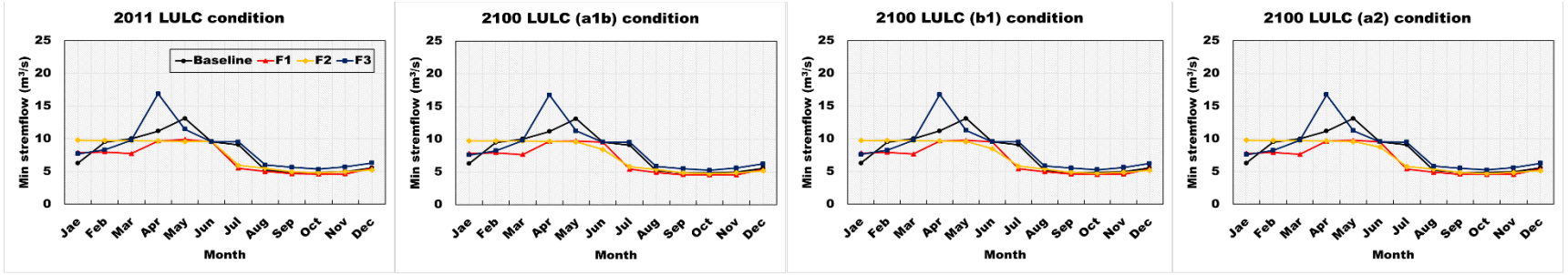
Figure 6.16. Comparison of monthly minimum streamflow for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

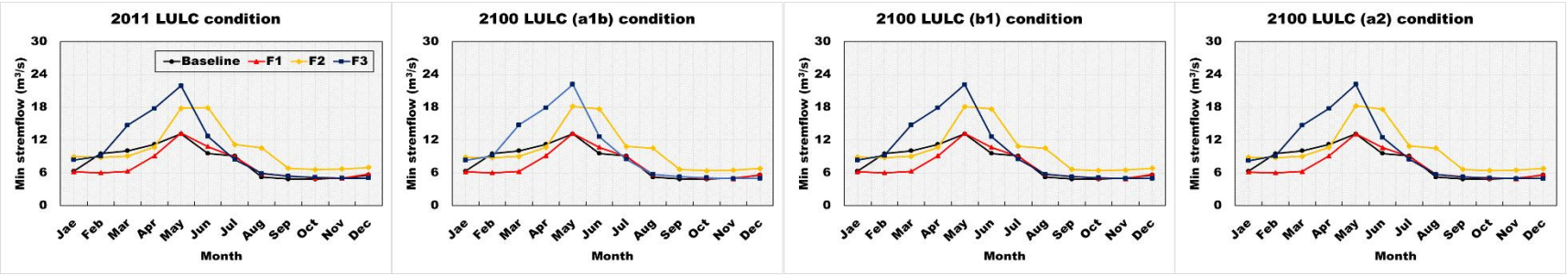


(b) RCP4.5

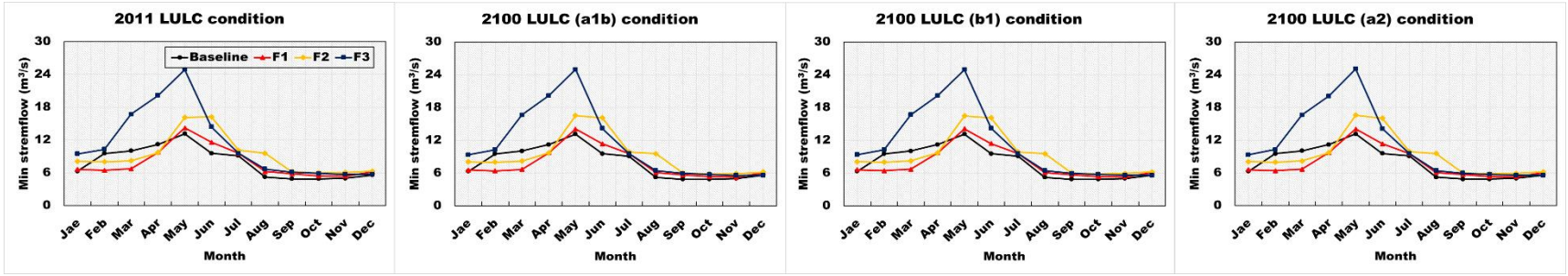


(c) RCP8.5

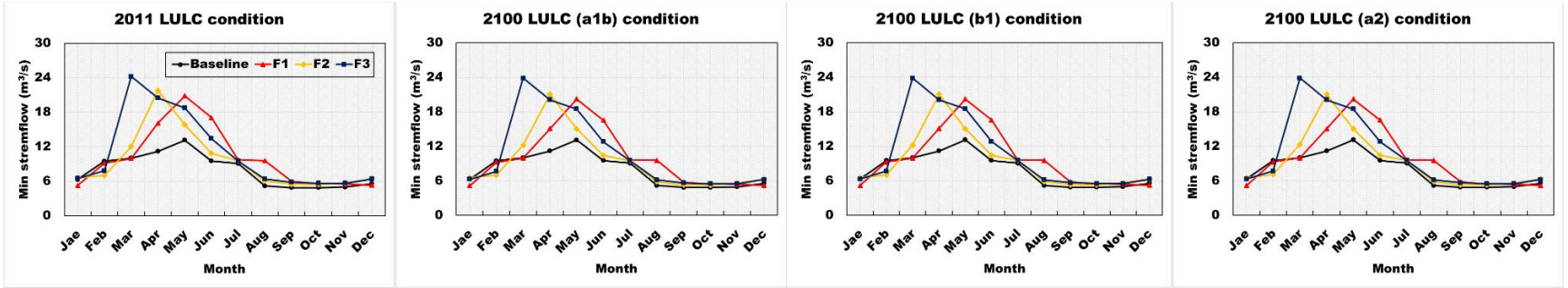
Figure 6.17. Comparison of monthly minimum streamflow for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

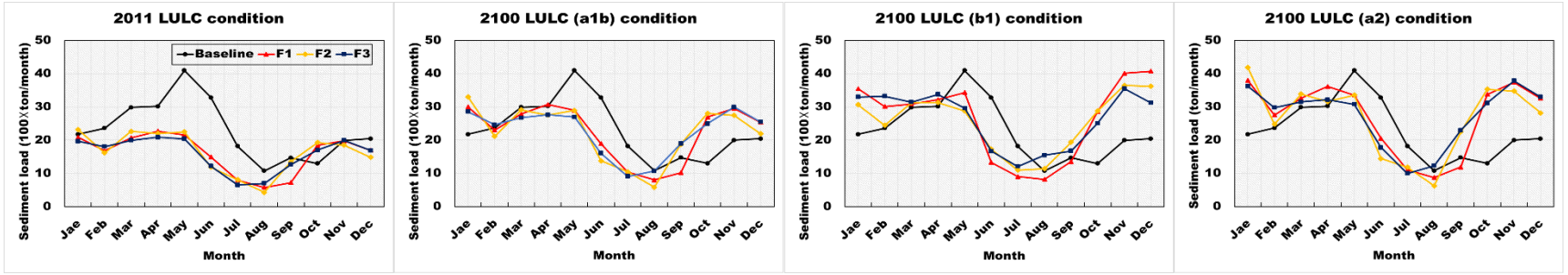


(b) RCP4.5

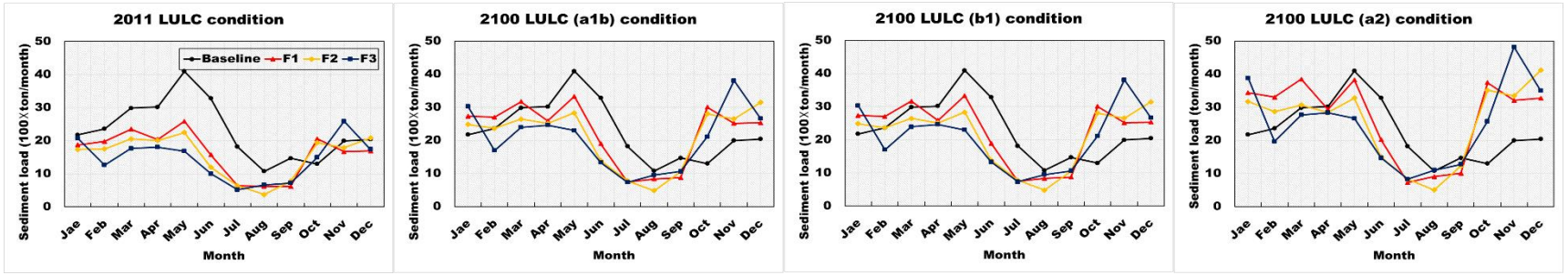


(c) RCP8.5

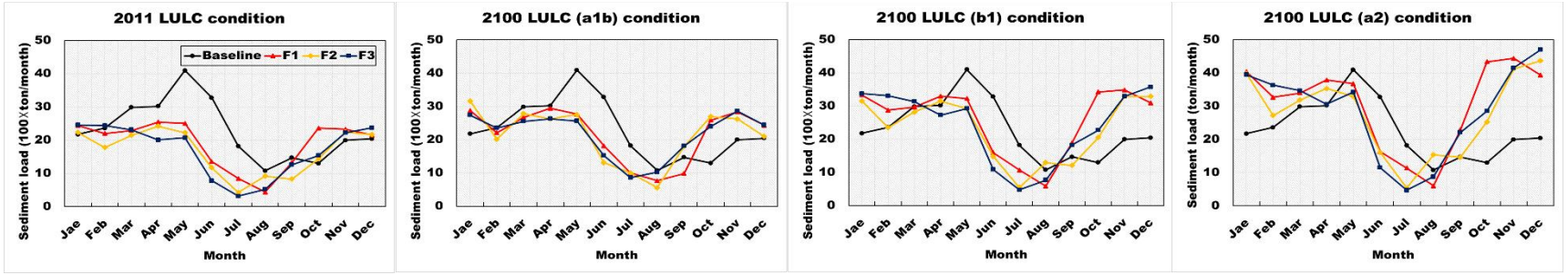
Figure 6.18. Comparison of monthly minimum streamflow for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

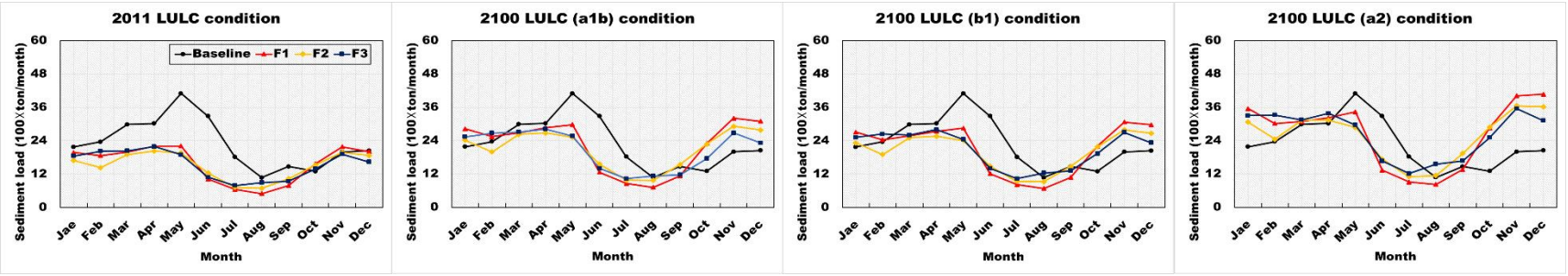


(b) RCP 4.5

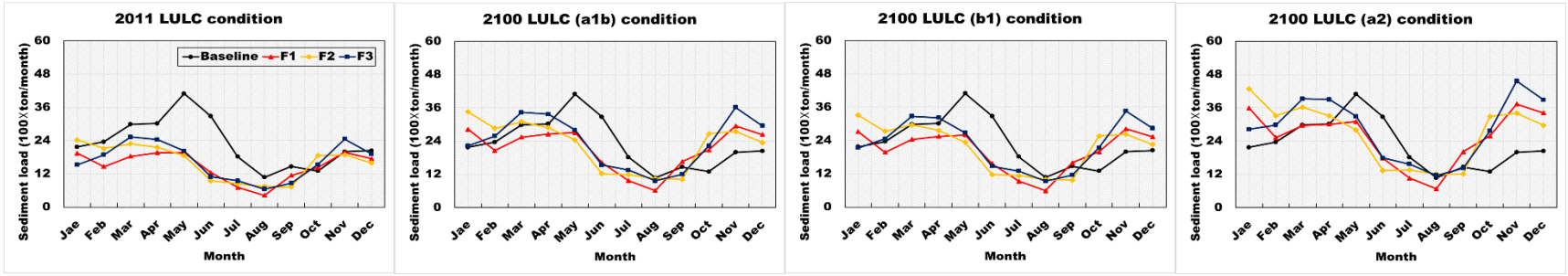


(c) RCP8.5

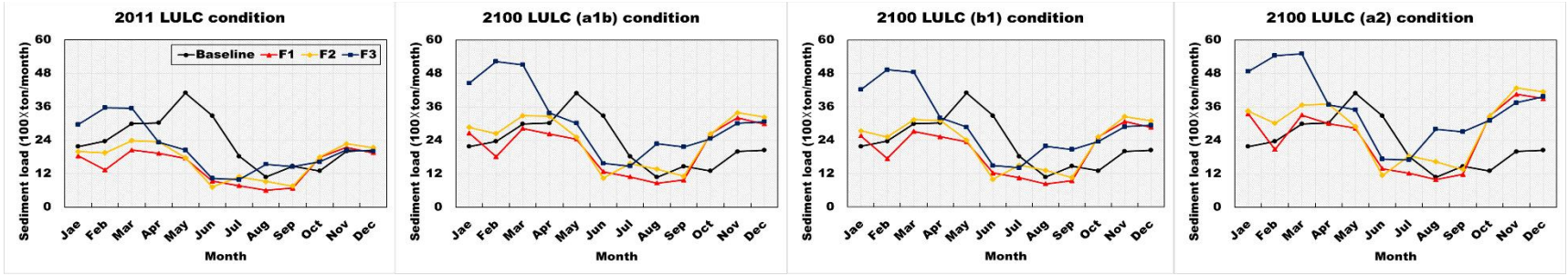
Figure 6.19. Comparison of monthly sediment load for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

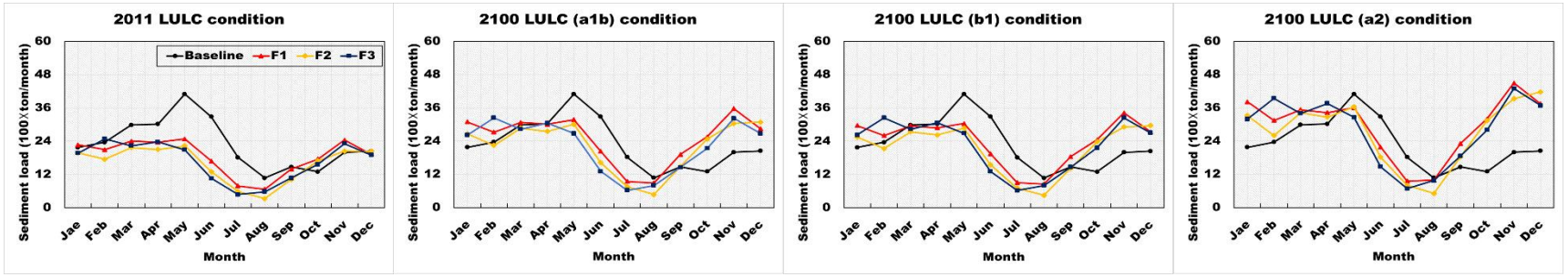


(b) RCP4.5

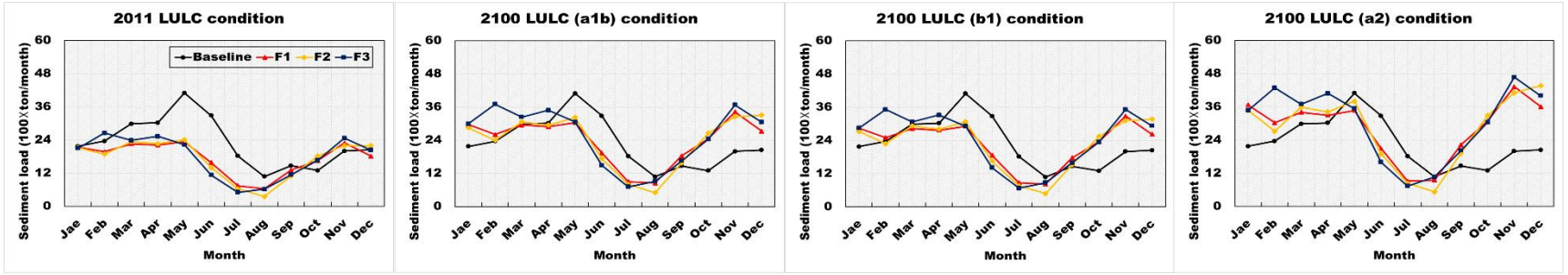


(c) RCP8.5

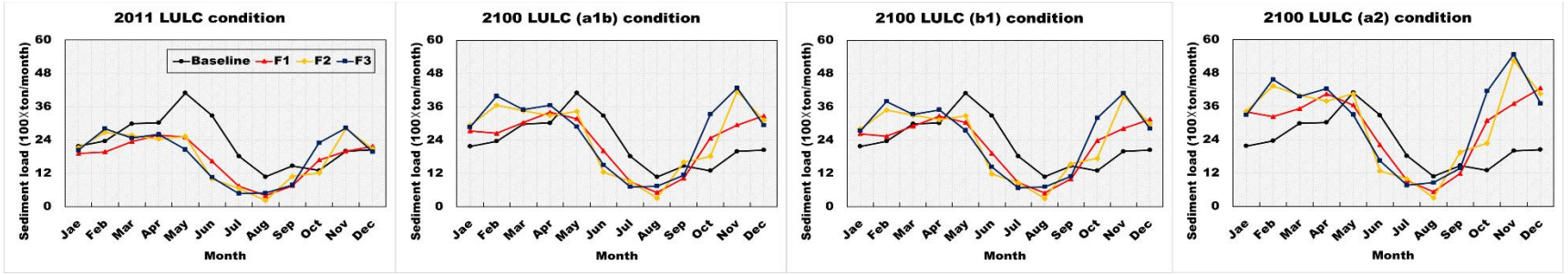
Figure 6.20. Comparison of monthly sediment load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

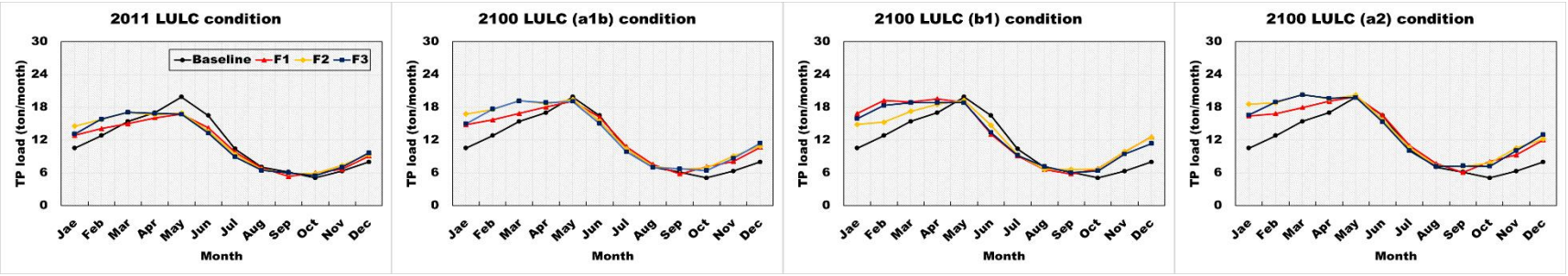


(b) RCP4.5

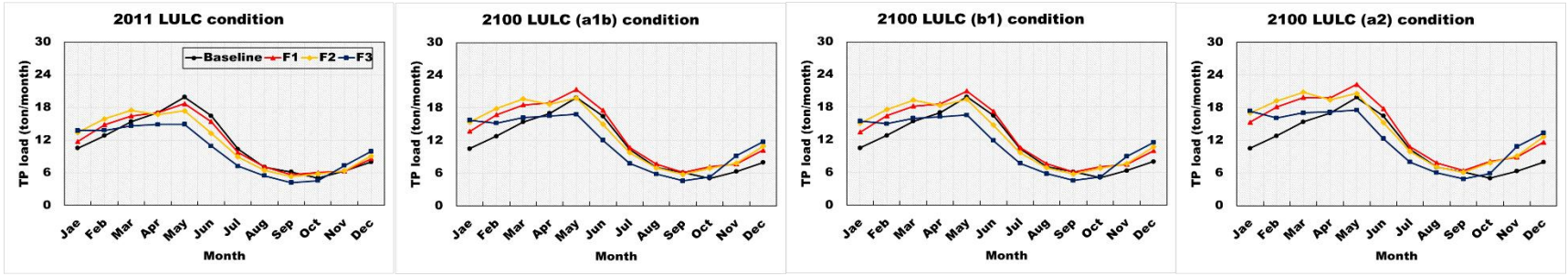


(c) RCP8.5

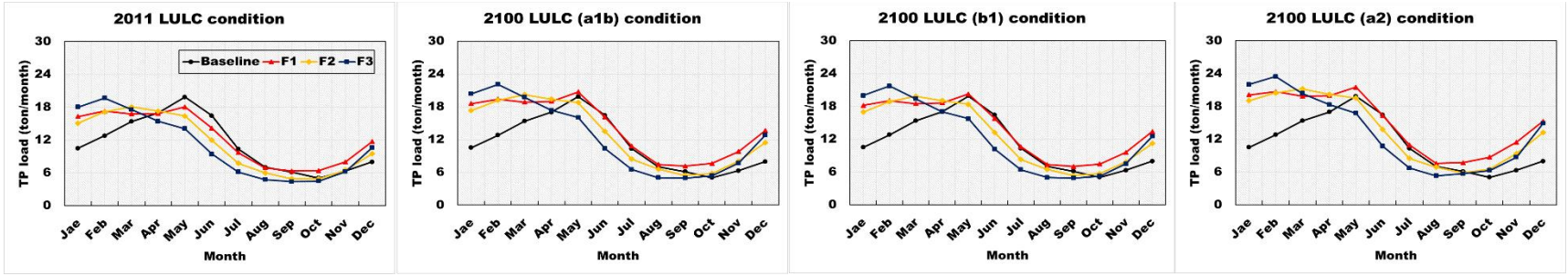
Figure 6.21. Comparison of monthly sediment load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

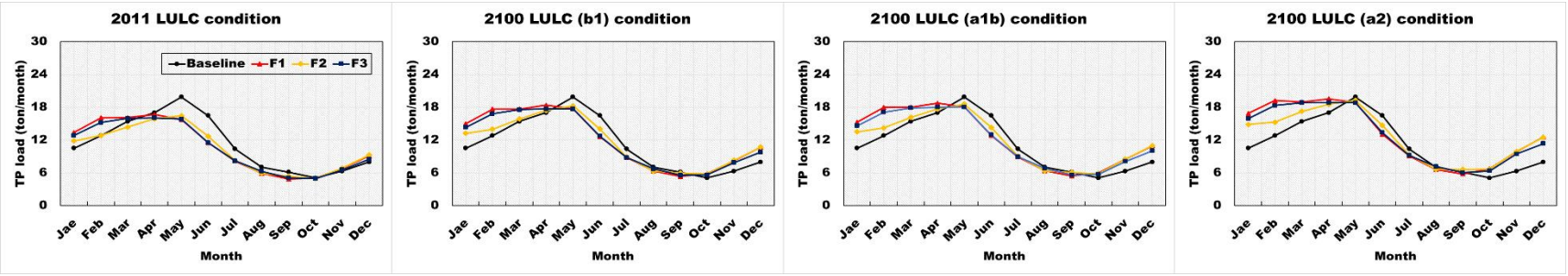


(b) RCP4.5

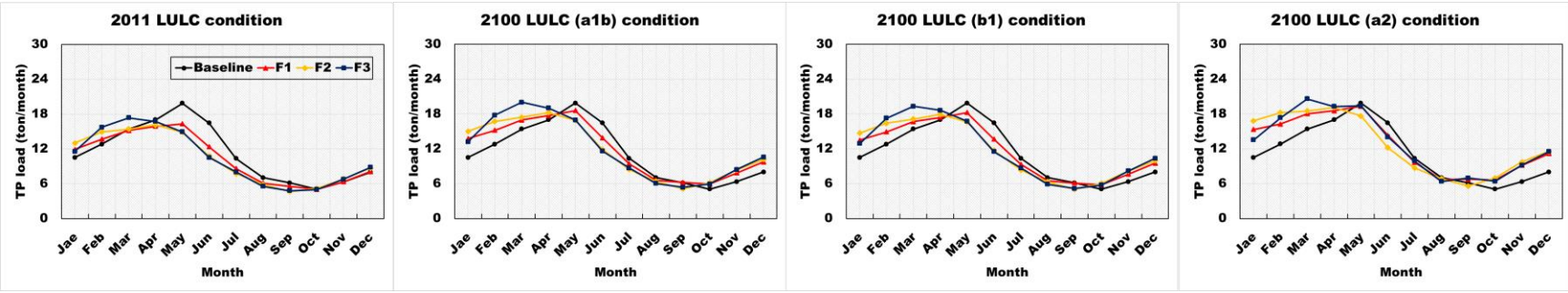


(c) RCP8.5

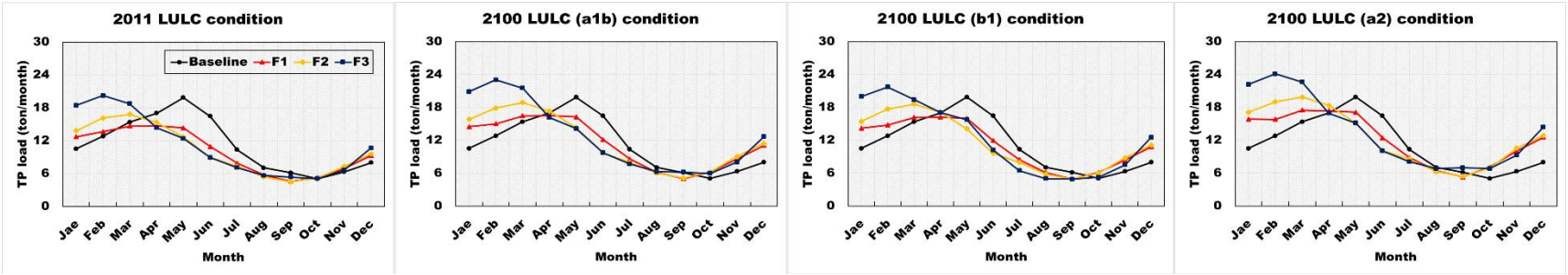
Figure 6.22. Comparison of monthly TP load for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

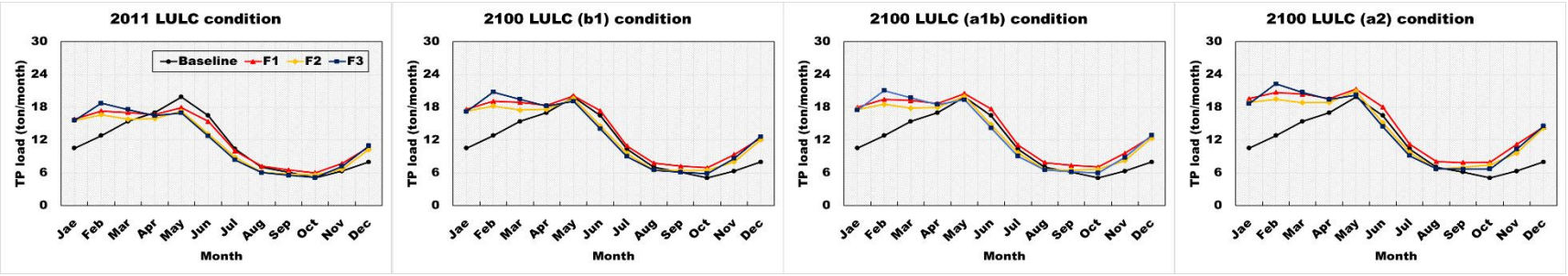


(b) RCP4.5

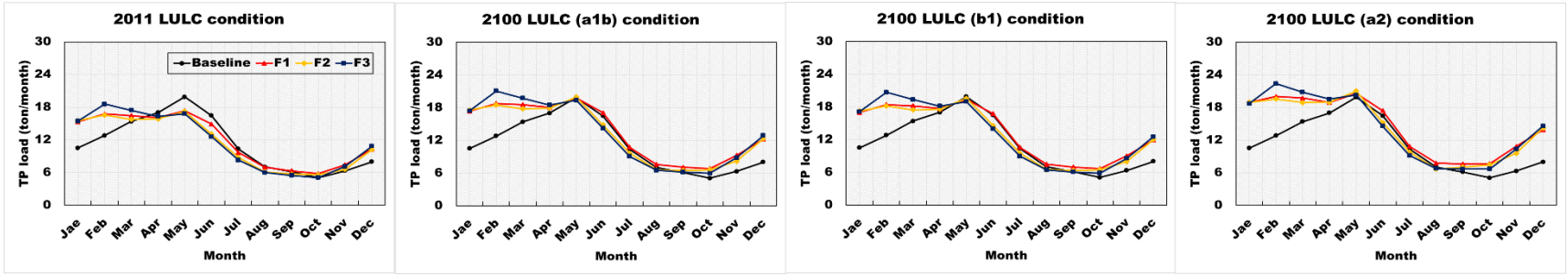


(c) RCP8.5

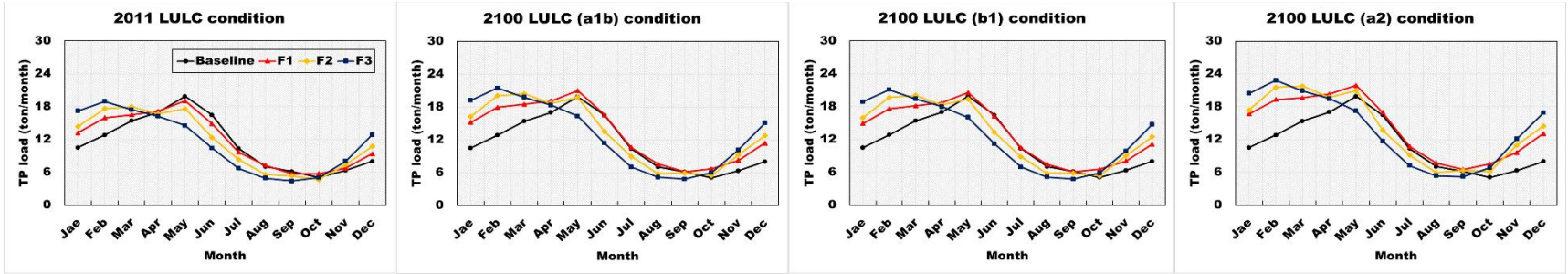
Figure 6.23. Comparison of monthly TP load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

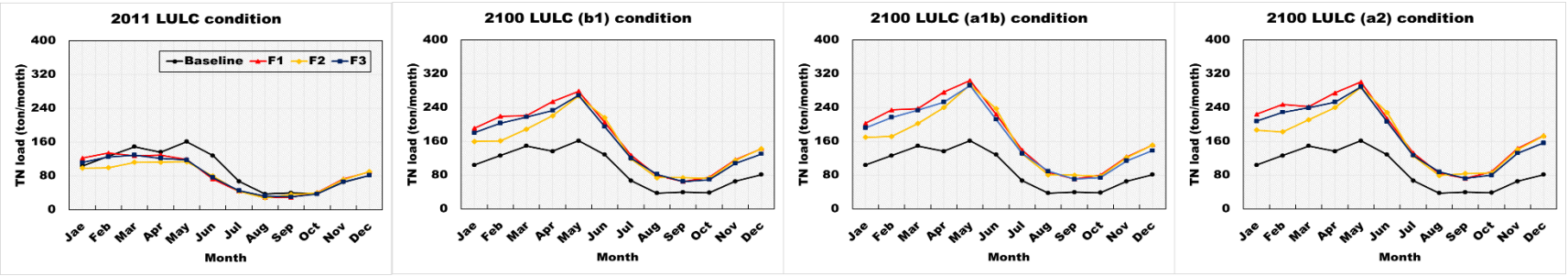


(b) RCP4.5

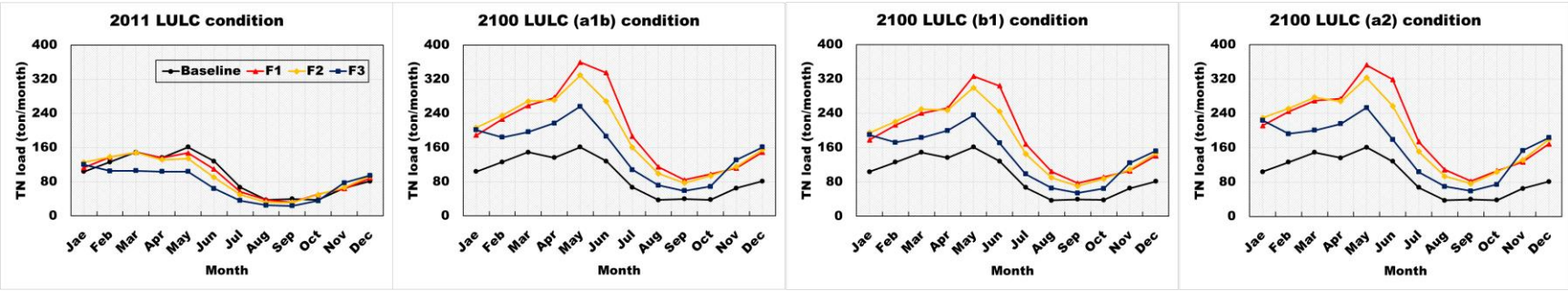


(c) RCP8.5

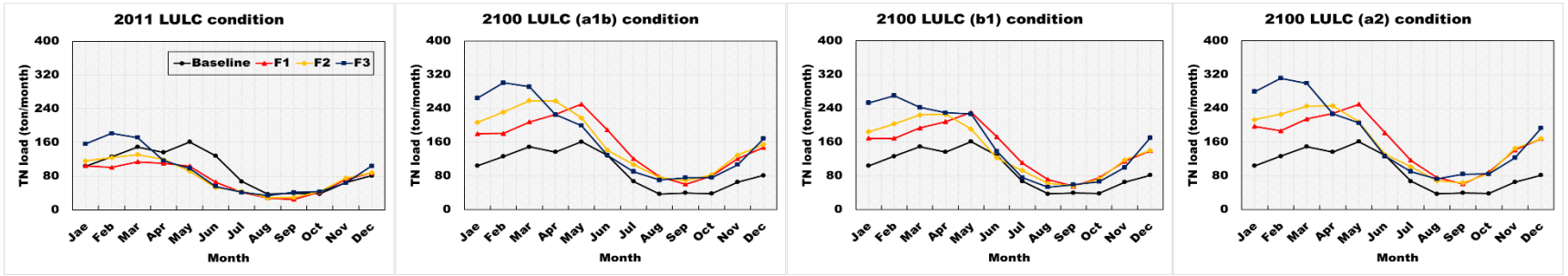
Figure 6.24. Comparison of monthly TP load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

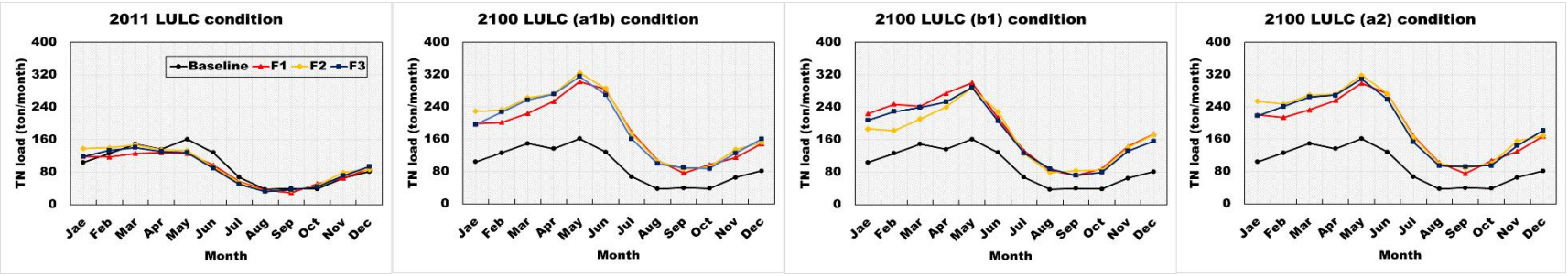


(b) RCP4.5

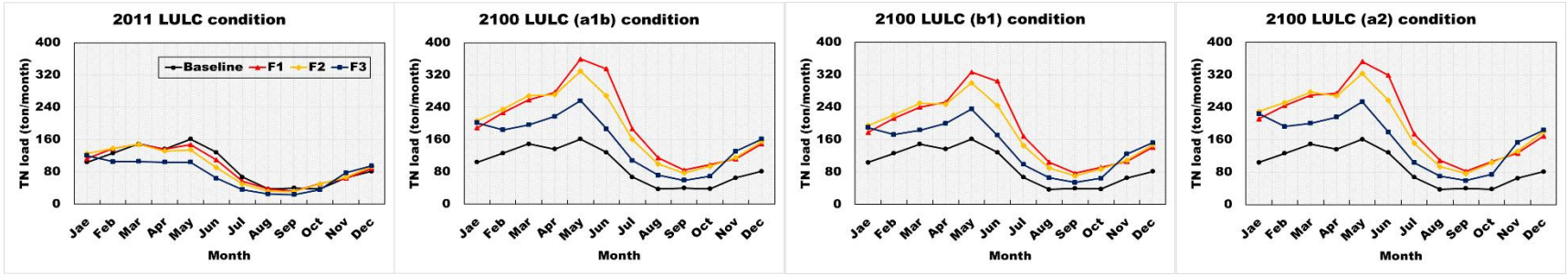


(c) RCP8.5

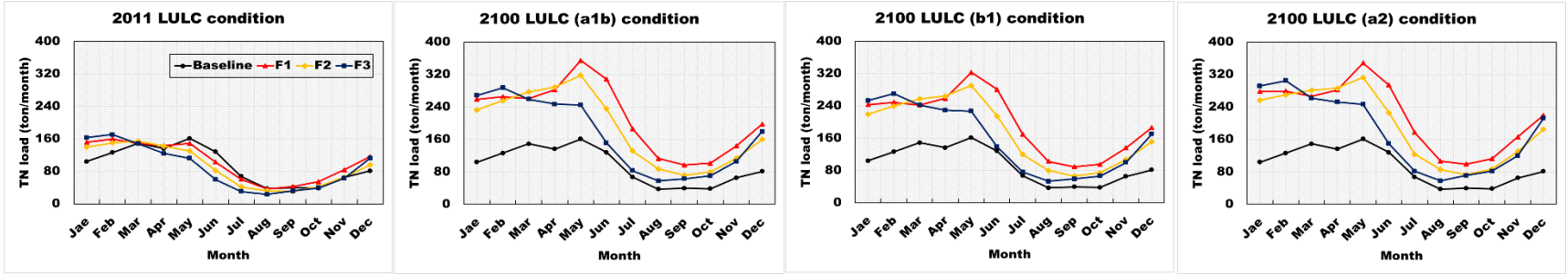
Figure 6.25. Comparison of monthly TN load for bcc-scm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

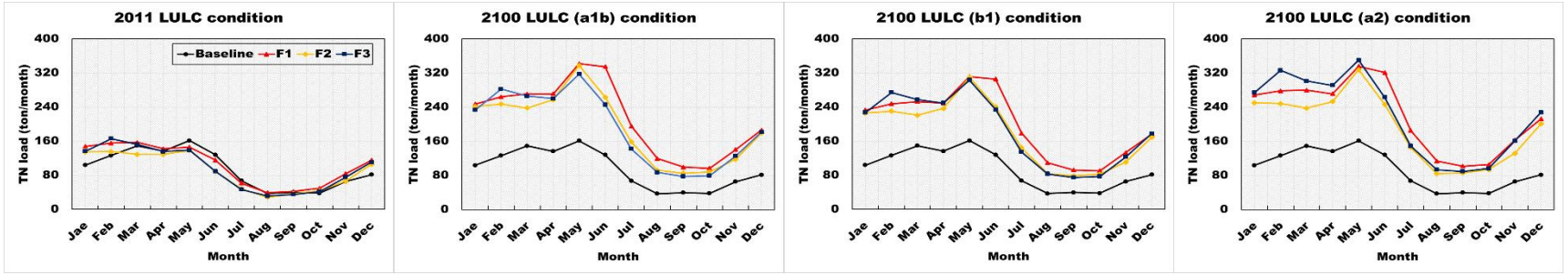


(b) RCP4.5

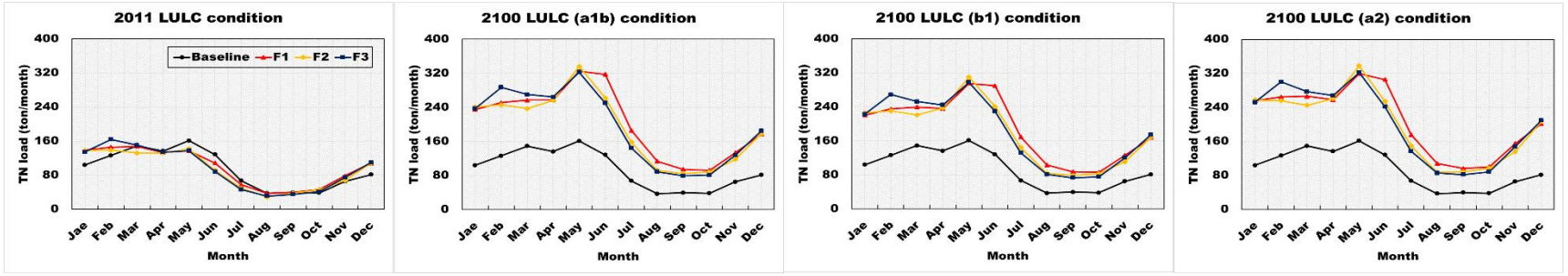


(c) RCP8.5

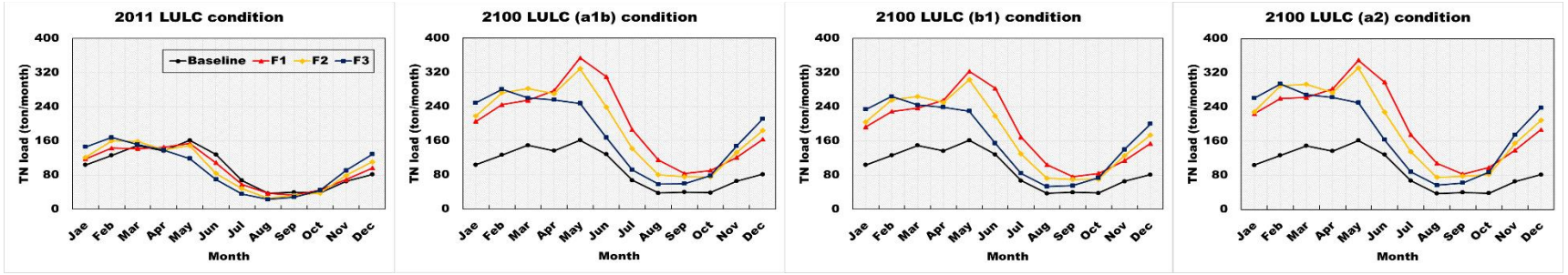
Figure 6.26. Comparison of monthly TN load for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

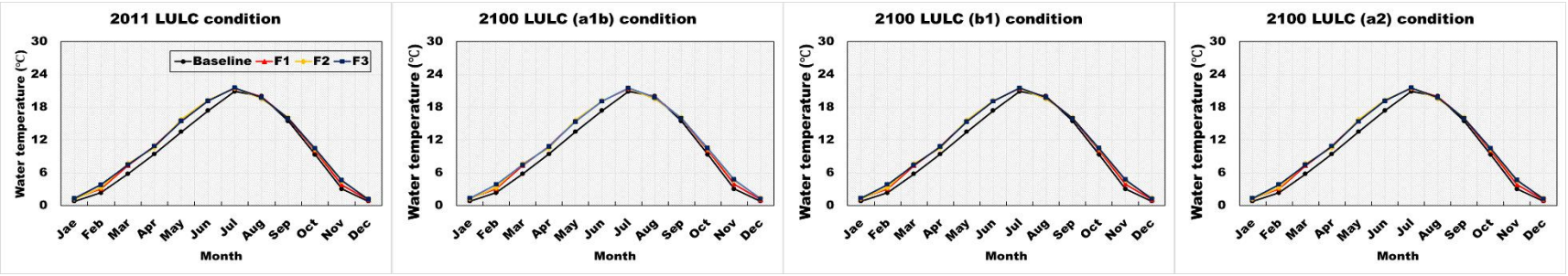


(b) RCP4.5

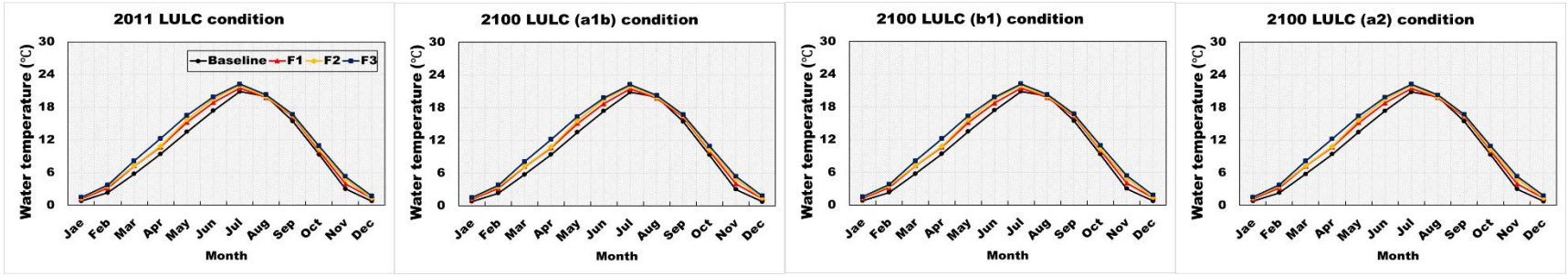


(c) RCP8.5

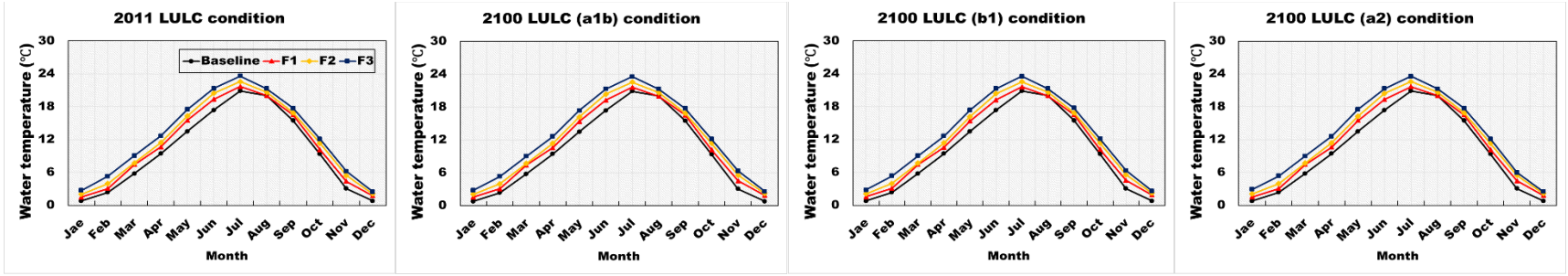
Figure 6.27. Comparison of monthly TN load for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

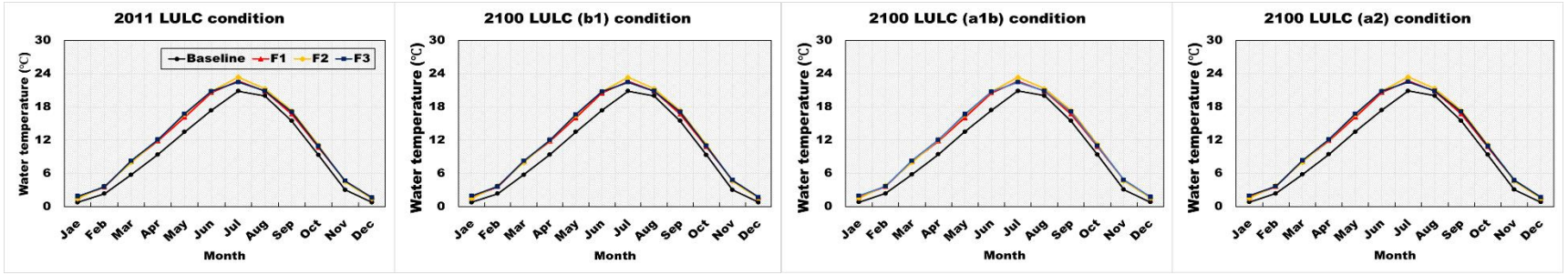


(b) RCP4.5

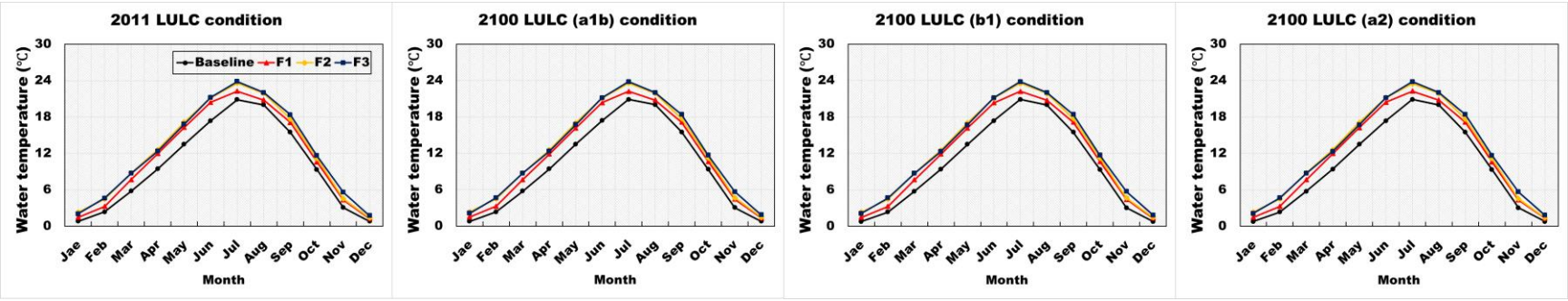


(c) RCP8.5

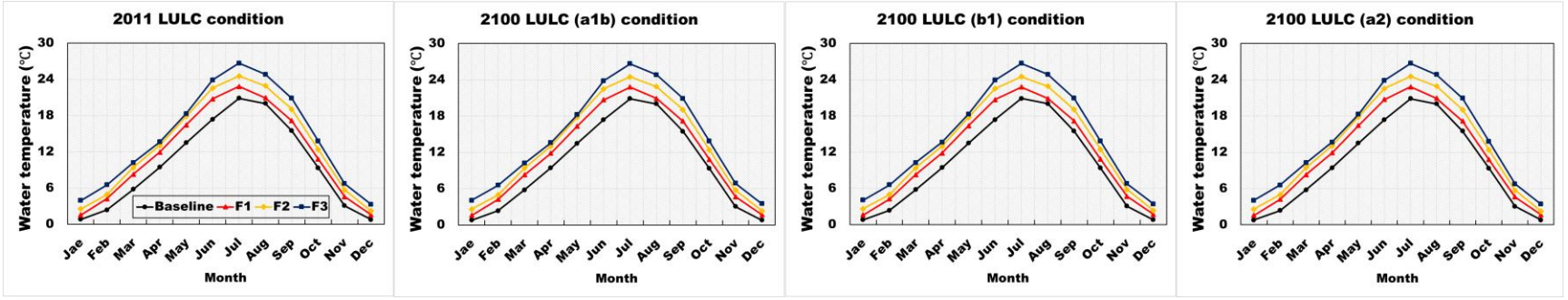
Figure 6.28. Comparison of monthly mean water temperature for bcc-csm1-1m model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6

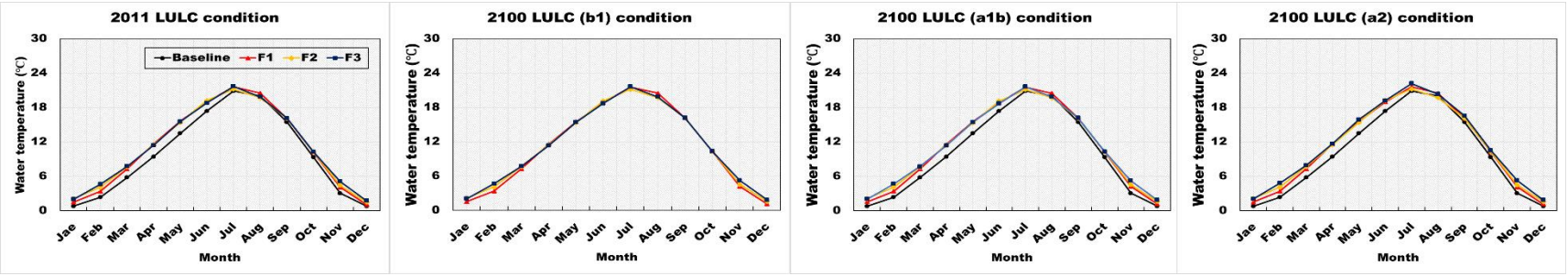


(b) RCP4.5

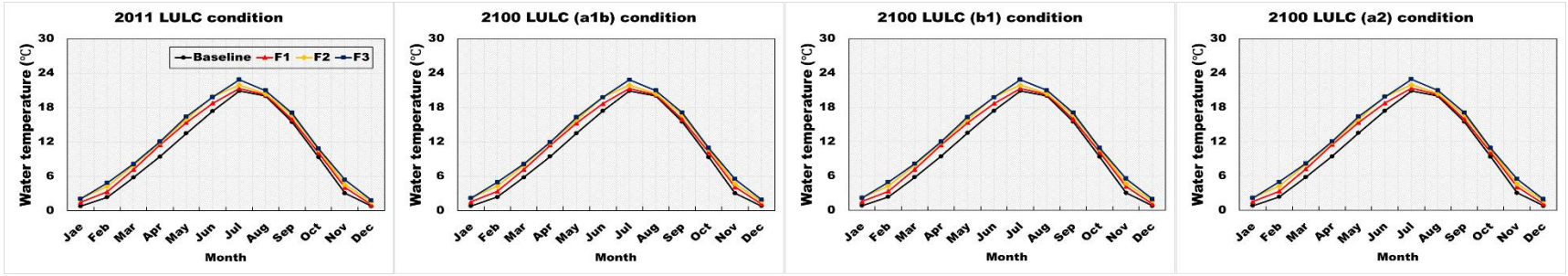


(c) RCP8.5

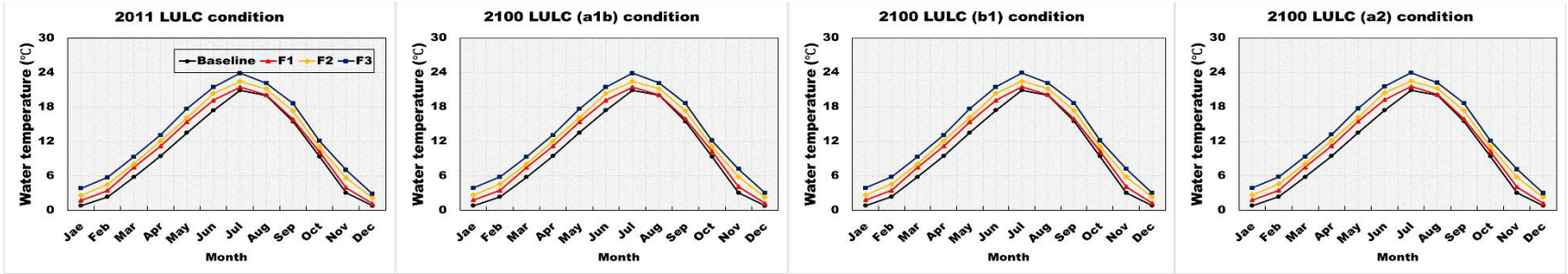
Figure 6.29. Comparison of monthly mean water temperature for canesm2 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).



(a) RCP2.6



(b) RCP4.5



(c) RCP8.5

Figure 6.30. Comparison of monthly mean water temperature for cnrm-cm5 model under three RCPs and 2100 LULC change scenarios from the baseline period (1981-2005) and the future periods (F1, F2, and F3).

CHAPTER 7. DEVELOPMENT OF AN OPTIMIZATION WATERSHED MANAGEMENT TOOL

7.1. Introduction

Water quality degradation is a significant concern to the local communities over the last decades. The problem is compounded by land development, climate change, and intensified agricultural practices. US.EPA (2009) presented 44% out of 3.5 million miles of nation's river and stream were impaired. Urban development, improper application of fertilizer, animal wastes, irrigation water, and forestry are the critical components to result in the eutrophication by excessive nutrient loading into waterbodies (Commission, 1998). The eutrophication decreases property value, disrupts recreation, and creates taste and odor problems (Dodds et al., 2008). Increased demographic and economic pressure continually subject to overexploitation of surface and ground water (IPCC, 2007; IPCC, 2013). Notable management strategy and planning of water resources will be developed and applied by reversing the trend of water quality and quantity degradation in the specific sub-watersheds. Low Impact Development (LID)/Best Management Practice (BMP) implementations on development lands are currently a growing interest so as to improve water quality and quantity affected by climate and LULC changes. LID/BMPs are commonly applied to control water quality and quantity from urban or agricultural land within the specific watersheds. However, LID/BMP effectiveness is related to their types and implementation locations (Giri et al., 2012). The selection of Critical Hotspot (CHS) is a significant task to reduce water pollutant loads by considering the more efficient method for LID/BMPs implementation. Therefore, the identification of CHS is an essential element for water quality improvement

from optimization water resources management. Panagopoulos et al. (2012) developed a decision support system for BMP placement to maximize pollutant reduction while limiting cost. Ghebremichael et al. (2010) applied SWAT model to identify and quantify phosphorus CHSs in the Rock River watershed, which is the primary phosphorus contributor to Lake Champlain.

It is a complicated decision-making process to identify CHSs and determine the management plans for water quality improvement owing to the multiple objective and considerations. The watershed is an appropriate unit to analyze water resources planning and management problems (Mckinney et al., 1999). However, there are several challenges to implementing water resources management plans for mitigating water quality problems in the vast catchments due to the lack of information and resource constraints. It is difficult to aggregate the multiple objectives into a single criterion for performance evaluation in the alternative ranking and selection (Makowski et al., 1996). Multi-criteria decision analysis (MCDA) methods have been developed to provide the decision makers or stakeholders for their unique and personal process. MCDA provides a rich collection of techniques and procedures for structuring decision problems, designing, evaluating, and prioritizing alternatives decisions. MCDA can integrate with geographical data to identify and visualize the CHS for the problematic water pollution area. There is no consensus for optimal MCDA method to address all the criteria. Therefore, the researcher should determine the best selection using their decision-making process. Roy (1981) represented the four main types of decision process using the MCDA method. The first category is the choice problem. It is to select the single optimal option or decrease the group of choices to a subset of equivalent. The second type is the sorting problem. Several alternatives are sorted into ordered and predefined

categories. It is then to regroup the alternatives with similar behaviors or features for clear, organizational, or predictive reasons. The third type is the ranking problem. The alternatives are ordered from best to worst by mean of scores or pairwise comparisons. The final form is the description problem. It is to explain alternatives and consequence. Usually, this step finishes in the first step to understand the characteristics of the decision problem. Several studies have been conducted for the optimization approach using MCDA method. Bryan and Crossman (2008) developed an optimization managements based on regional planning approach to identify geographic priorities for ground natural resource management to assess the most cost-effective strategy. Sadeghi et al. (2009) applied on optimization method to maximize profits from land use while minimizing erosion risk. Cho et al. (2004) coupled Qual2k with a multi-objective evolutionary algorithm (MOEA) to select the adequate treatment type for each wastewater treatment plant to improve water quality at river basin scale.

In **chapter 7**, therefore, the study objectives are to (1) identify CHSs of water pollutants in the watershed before LID/BMPs implementation using three spatial targeting methods, (2) determine the most efficient LID/BMPs at the selected CHSs while minimizing area devoted to LID/BMP implementations, and (3) evaluate water quality improvement and cost-effective for the LID/BMP implementations (**Figure 7.1**). The BRW is selelcted as the study area (**Figure 5.1**).

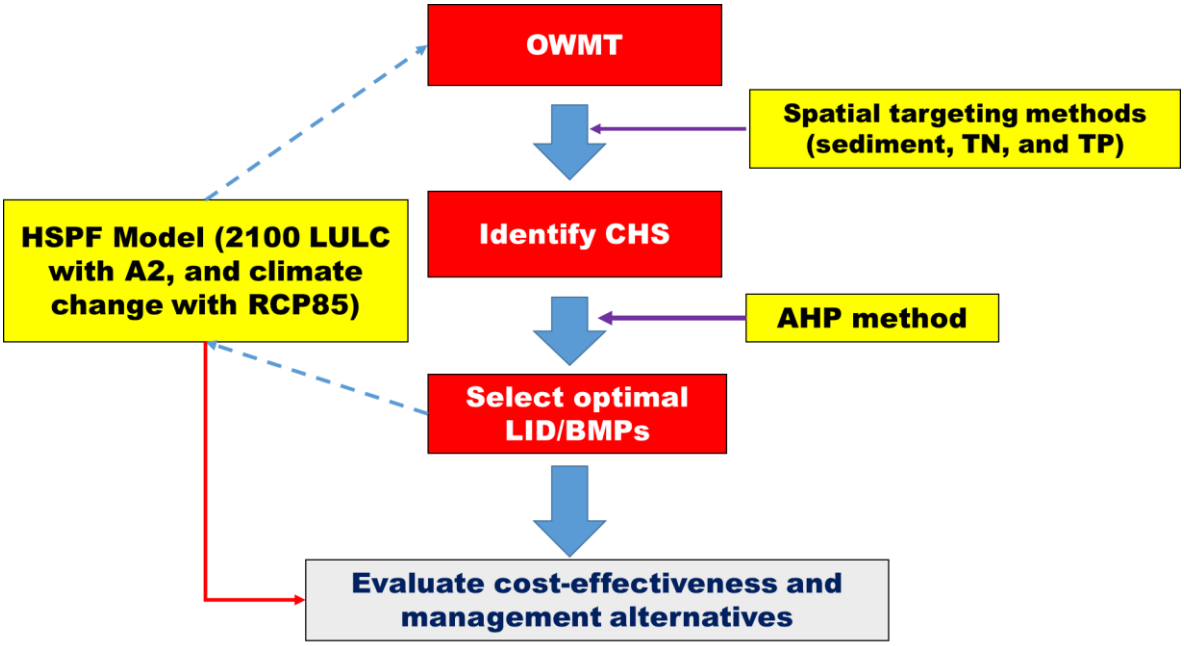


Figure 7.1. Flowchart of LID/BMPs application at CHS using OWMT.

7.2. Model input data

Future climate and LULC data were used bcc-csm1-1 model under RCP8.5 and 2100 LULC A2 scenario to evaluate the watershed management plans by the severe hydrological conditions. Calibrated HSPF model from chapter 2 was used to simulate monthly streamflow, sediment, and water quality from 2071 to 2095.

7.3. Optimization Watershed Management Tool (OWMT)

7.3.1. Spatial targeting methods

Area placement at LID/BMP implementation is highly affected the water quality improvement. Therefore, the CHS identification for pollutant area should be conducted before LID/BMP implementation. Three targeting methods (Concentration Impact Index, Load Impact Index, and Load per sub-basin area Index) were applied to identify CHSs that need the most improvement for the water pollutants (sediment, TN, and TP). Sediment, TN, and TP

outputs estimated from the spatial targeting methods were normalized, and then categorized the high, medium, and low priority areas using quantile method, respectively. For the high priority areas, LID/BMPs were implemented to analyze their effectiveness and water quality improvement.

7.3.1.1. Concentration Impact Index (CII)

CII method is based on the pollutant concentration level in the sub-watershed reaches (Tuppad and Srinivasan, 2008). This method considers pollutants from the sub- watershed as well as the entire upstream conditions. It is effective in addressing localized pollution in tributaries for low and high streamflow condition.

7.3.1.2. Load Impact Index (LII)

LII method is based on the pollutant load in the sub- watershed from the sub-watershed reaches and the entire upstream conditions (Tuppad and Srinivasan, 2008). This method represents useful information of BMPs implementation in high streamflow and contribution of upstream drainage areas.

7.3.1.3. Load per Sub-basin area index (LPSAI)

LPSAI is based on the pollution load from individual sub- watersheds. This method does not include the impact of upstream watersheds (Tuppad and Srinivasan, 2008). This method applies to identify the localized concerns at the sub-watershed scales. Since this approach is highly affected watershed delineation process, it can be considered subjective results.

7.3.2. LID/BMP implementation plans

Five LID/BMPs options, including 1) Bioretention, 2) filter strip, 3) grassed swale, 4) wetland, and 5) detention pond were implemented at CHSs. The reduction rate of sediment, TN, and TP loads was calculated using HSPF model outputs before and after LID/BMP implementations. For the implementation plans, three priority scenarios (equal distribution, economic aspect, and environmental aspect) were applied as shown in **Table 7.1**. For the equal distribution scenario, the weight vector was equally assigned 0.33 for environmental components (sediment, TN, TP), total cost, and area condition, respectively. The economic aspect scenario was given the weight vector to total cost by 0.70, the environmental component by 0.15, and area condition by 0.15. For the environmental aspect scenario, the environmental component was assigned the weight vector by 0.72, and total cost and area condition were assigned the weight vector by 0.14, respectively.

Table 7.1. Weight vectors of criteria properties by application scenarios.

Element	Equal distribution	Economic aspect	Environmental aspect
Sediment	0.11	0.05	0.24
TN	0.11	0.05	0.24
TP	0.11	0.05	0.24
Total cost	0.33	0.70	0.14
Area condition	0.33	0.15	0.14

7.3.3. MCDA development for optimal LID/BMP selection

Five LID/BMPs were implemented to determine the optimal solution for the water quality and quantity improvement at the selected CHSs in the BRW. MCDA method was

performed to evaluate the most effective LID/BMP at each sub-watershed. The data layers were assessed with one another by means of a pairwise comparison matrix with regard to their significance to locating by applying APH technique.

7.3.3.1. Analytical Hierarchy Process (AHP)

AHP uses multiple variables for the decision making. It can be employed to formulate the problem and solve the problems hierarchically (Nagaraju et al., 2012). APH is a widely used method based on the additive weighting (Banai, 1993; Zhu and Dale, 2001). This method has two distinctive ways to apply. The first way can be used to derive the weights associated with alternatives. This approach performs the problems involving numerous alternatives when a pairwise comparison of the alternatives is not possible (Eastman et al., 1993). Another way can be used to aggregate the priority for all level of the hierarchy structure based on the level representing alternatives. AHP method is modeled as a hierarchical structure with the goal, alternative, and criteria by evaluating the alternatives. Criteria can be divided sub-criteria for detailed analysis. Priorities are constituted with the criteria of the hierarchy. The pairwise comparisons of each criterion are applied to assign the importance of criteria using the scale as shown in **Table 7.2**. A comparison matrix summarizes all information by the pairwise comparisons to determine preference or the overall priorities for the hierarchy. The final ranking is determined based on the normalized principal priority vector (Eigenvector) calculated by the comparison matrix. From this matrix, a consistency check will be performed. The computed consistency ratio should be less than 10 %.

Table 7.2. Scale of relative important (adapted from Saaty, 1980)

Intensity of Importance	Description
1	The alternatives being compared contribute equally to the defined object
3	One alternative is favored slightly over the other in terms of achieving the defined objectives.
5	One alternative is favored strongly over the other in terms of achieving the defined objective.
7	One alternative is favored very strongly over the other in terms of achieving the defined objectives.
9	The evidence favoring one alternative over the other is absolute in terms of achieving the defined objectives.
2, 4, 6, 8	Intermediate values available express user-defined comparison.

7.3.3.1.1. Consistency Index (CI)

CI, which is related to the eigenvalue method which developed by Saaty (1980).

$$\mathrm{CI} = \frac{\gamma_{max} - n}{n - 1} \tag{7.1}$$

$$\mathrm{CR} = \frac{CI}{RI} \tag{7.2}$$

where, r_{max} is the maximal eigenvalue, n is the number of rows or columns in the pairwise matrix, CR is the consistency ratio (CR), RI is the random index (**Table 7.3**). If CR is less than 10%, then the matrix is of an acceptable consistency.

Table 7.3. Random indices from Saaty (1980).

n	3	4	5	6	7	8	9	10
RI	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

7.3.3.1.2. Priorities derivation using approximate method

Priorities derivation using approximate method is based on two simple steps. First is the summation of the elements of the row as below question.

$$r_i = \sum_i a_{ij} \tag{7.3}$$

Next step is the normalization of the sums using equation (6.3) in **chapter 6**.

$$\text{pi} = \frac{r_i}{\sum_i r_i} \tag{7.4}$$

Finally, given a decision matrix the final priorities, denoted by A^i_{AHP} of the alternatives in terms of all criteria combined are determined according to the following equation.

$$A^i_{AHP} = \sum_{j=1}^{N} a_{ij} w_j \tag{7.5}$$

7.3.4. Total cost of LID/BMP installation

Total cost for LID/BMP implementations and cost per unit pollutant load reduction (sediment, TN, TP loads) were calculated using **Table 7.4**. Total cost was the sum of construction and maintenance cost for LID/BMP implementations. All costs were used US dollars. Cost per unit pollutant load reduction was utilized to calculate cost per kg pollutant reduction for sediment, TN, and TP on average.

Table 7.4. Construction costs and annual maintenance costs of LID/BMPs (Brown and Schueler, 1997; County, 2014; Dhalla and Zimmer, 2010; NCDENR, 2007; USEPA, 1999)

LID / BMP	Construction Cost ($/ha)	Annual maintenance cost (% of construction cos)
Bioretention	151,200	6
Detention Pond	12,200	4
Wetland	15,500	4
Grassed Swale	9,000	6
Filter strip	3,400	3

7.4. Results

7.4.1. The location prioritization for spatial targeting methods

7.4.1.1. CII method

CII method was based on pollution concentration for sediment, TN, and TP in each sub- watershed reach of the BRW (**Figure 7.2**). The watershed was divided into the high, medium and low priority areas for sediment, TN, and TP concentration. The results were then normalized to determine the final location prioritization. **Table 7.5** shows the selected sub-basins by the high, medium, and low priority areas in the BRW. Total 16 sub- watersheds were selected for CHSs (the high priority areas). CHSs were mostly located in the LBRW. Th main contribution of water pollutants was mainly caused from urban, agricultural, and barren/mining land. For the high priority areas (CHSs), sediment concentration ranged from 17.76 mg/L to 21.97 mg/L. TN concentration ranged from 21.9 mg/L to 176.85 mg/L and TP concentration ranged 0.16 mg/L to 1.01 mg/L. The normalized results ranged from 0.13 to1 at CHSs. The available drainage area of LID/BMP implementation was 112,918.70 ha (**Table 7.6**).

7.4.1.2. LII method

LII method was applied to determine CHSs based on the pollutant loads for each sub-watershed reach **(Figure 7.3**). Total 15 sub- watersheds were selected for CHSs as shown in **Table 7.5**. CHSs were mostly located near the watershed outlet or in the middle of the BRW, which is below the large reservoir (Luck Peak Dam), because pollutant loads from the sub-basins and the entire upstream were accumulated within the large reservoir and near the watershed outlet. For the high priority area, the ranges of sediment load were from 10.240 ton/month to 27,420 ton/month. The ranges of TN load were from 138 ton/month to 310 ton/month. The ranges of TP load were 8.69 ton/month to 14.47 ton/month. The normalized result ranges were from 0.44 to1 at CHSs. The available drainage area of LID/BMP implementation was 73,650.87 ha.

7.4.1.3. LPSAI method

LASAI method was used to determine CHSs using the pollutant loads at each sub-watershed which identified the pollutant from location source. This method classified the high priority area for sub-watersheds that release a relatively large amount of pollutant in a small area due to the inherent area dependence. Total 17 sub-watersheds were selected for CHSs (**Table 7.5)**. For the high priority area, sediment load ranged from 0.25 ton/ha to 2.07 ton/ha. TN load ranged from 24.77 kg/ha to 155.02 kg/ha, and TP load ranged 1.17 kg/ha to 10.06 kg/ha. The normalized result ranged from 0.13 to1 at CHSs. The available drainage area of LID/BMP implementation was 43,740.08 ha.

Table 7.5. The classification ranges for the priority areas and the selected sub-watersheds by spatial targeting methods.

Method	Components	Classification range			Selected sub-watersheds for CHSs
		High	Medium	Low	
CII	Sediment (mg/l)	21.97 >	21.97 – 7.99	7.99 <	14, 21, 28, 42, 44, 49, 50, 52, 71, 72, 73, 74, 75, 76, 77, 78
	TN (mg/l)	2.28 >	2.28 – 1.64	1.64 <	
	TP (mg/l)	0.16 >	0.16 – 0.09	0.09 <	
	Normalized Result	0.13 >	0.13 – 0.07	0.07 <	
LII	Sediment (100×ton/month)	10.24 >	10.24 – 2.21	2.21 <	47, 48, 61, 63, 64, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78
	TN (ton/month)	138.00 >	138.00 – 29.20	29.20 <	
	TP (ton/month)	8.69 >	8.69 – 0.78	0.75 <	
	Normalized Result	0.44 >	0.44 – 0.06	0.06 <	
LPSAI	Sediment (ton/ha)	0.25 >	0.25 – 0.03	0.03 <	44, 45, 47, 55, 59, 60, 61, 62, 64, 67, 68, 71, 72, 73, 76, 77, 78
	TN (kg/ha)	24.77 >	24.77 – 1.92	1.92 <	
	TP (kg/ha)	1.17 >	1.17 – 0.07	0.07 <	
	Normalized Result	0.13 >	0.13 – 0.01	0.01 <	

Table 7.6. LID/BMPs implementation area based on spatial target methods.

Targeting method	Implementation area (ha)
CII	112,918.70
LII	73,650.87
LPSAI	43,740.08

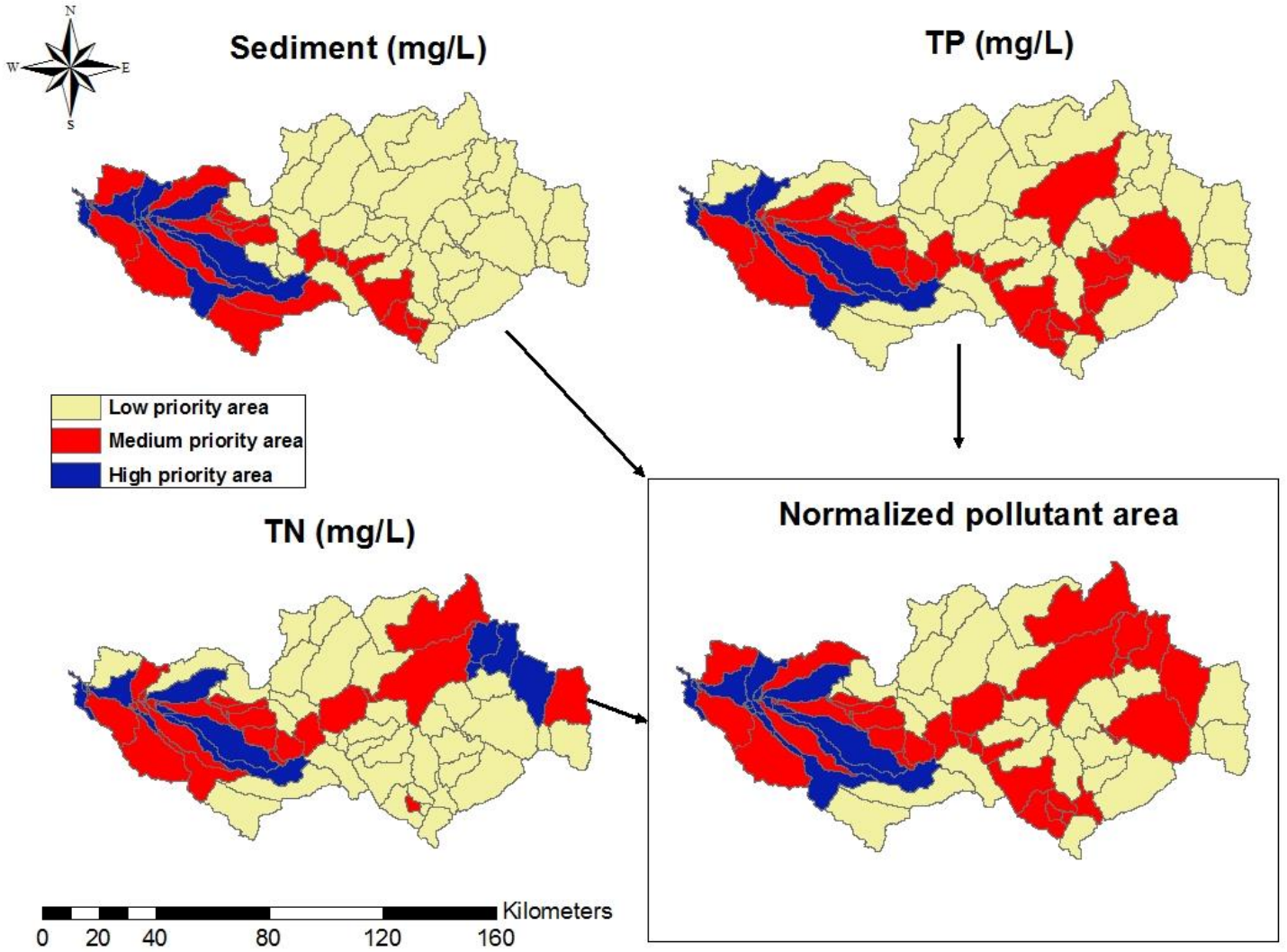


Figure 7.2. Normalized priority areas for CII method using sediment, TN, and TP concentration.

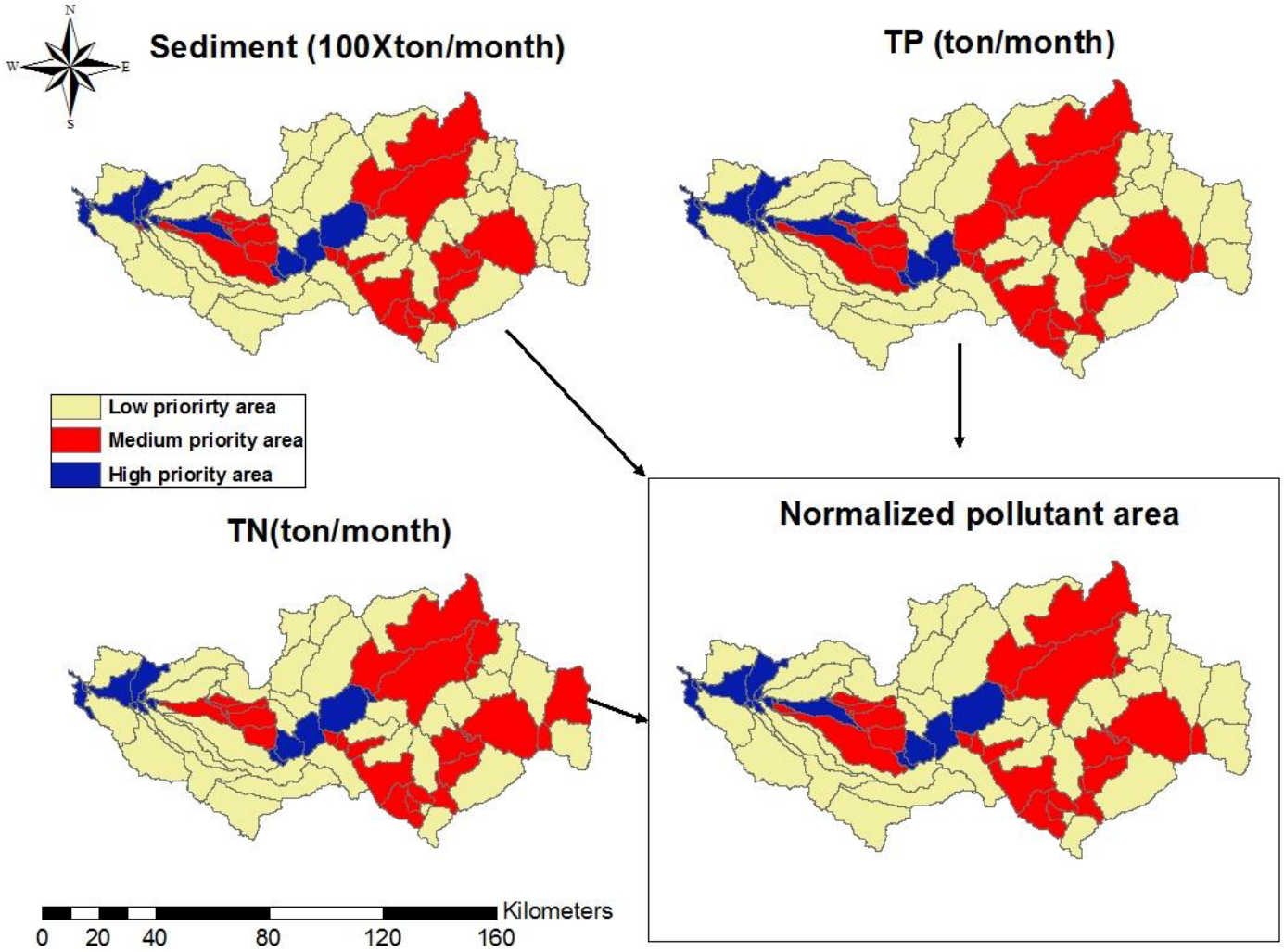


Figure 7.3. Normalized priority areas for LII method using sediment, TN, and TP loads.

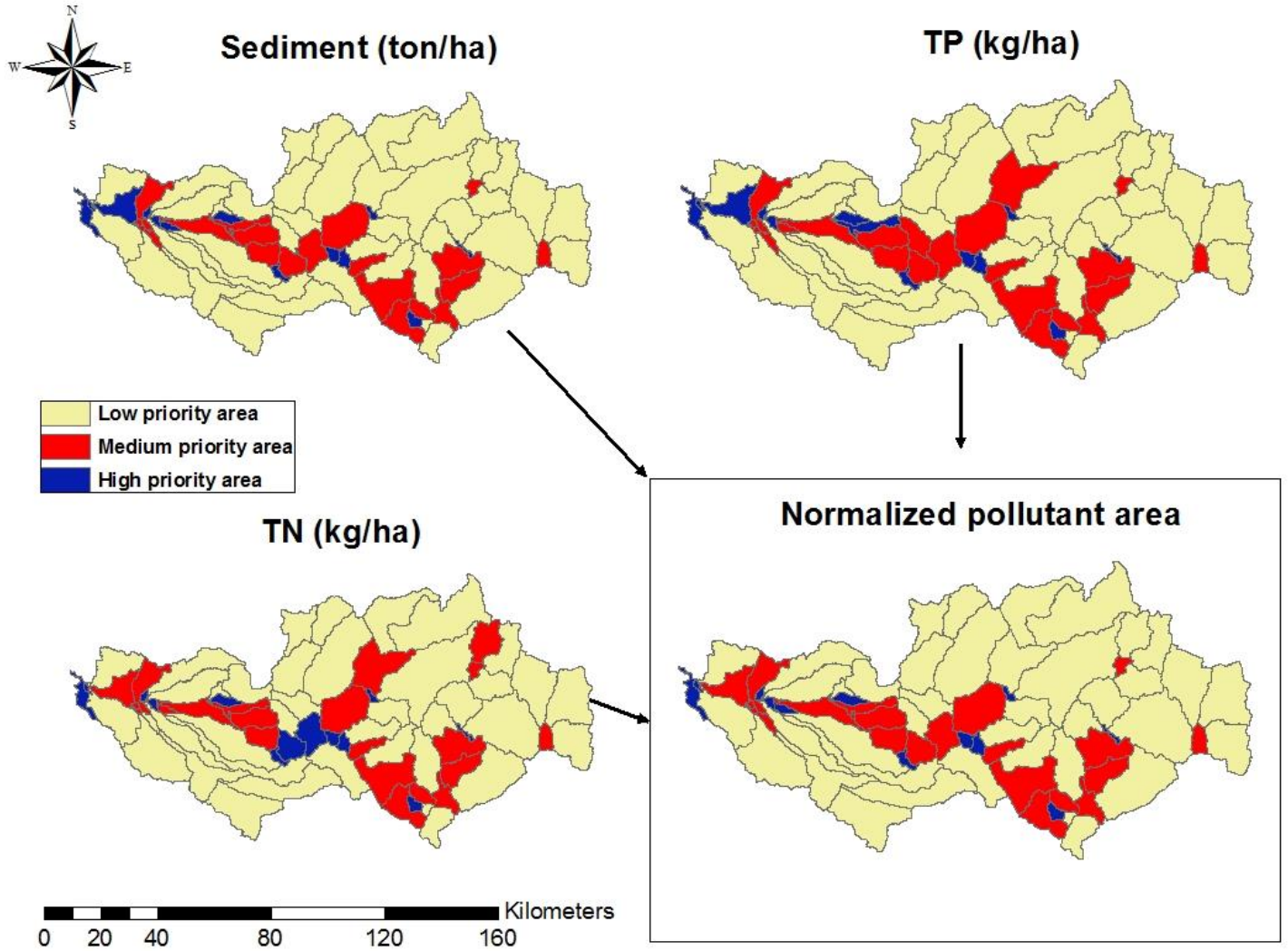


Figure 7.4. Normalized priority areas for LPSAI method using sediment, TN, and TP loads per each sub-watershed.

7.4.2. The best LID/BMP selections at CHSs using AHP method

Optimal LID/BMPs were identified by the implementation plan scenarios using AHP method. The pairwise comparison matrices for CII method were developed for sediment (**Table 7.7**), TP (**Table 7.8**), and TN (**Table 7.9**) on the basis of the reduction rate at each sub-watershed level. Similar tables for LII and LPSAI methods were also developed based on the pollution load reduction at each sub-watershed. The developed pairwise comparison matrices were calculated 0.091% CR by sediment, 0.001% CR by TP, and 0.01% CR by TN. Therefore, these matrixes were an acceptable consistency as they were all less than 10% CR. **Table 7.10** shows the weight vector for sediment, TN, and TP calculated from the pairwise comparison matrices. **Table 7.11** indicates the weight vector for total LID/BMP implementation cost and area condition of sub-watershed 78. The weight vector of area condition was varied by slope, drainage area, and land use at each CHS. The weight vectors for each criterion associated with LID/BMPs were assigned into a single matrix at sub-watershed 78 as shown in **Table 7.12**. **Table 7.13-15** represent the selected individual LID/BMPs at the CHSs using three priority scenarios for three spatial targeting methods. In the equal distribution scenario, grassed swale was installed at 5 sub-watersheds by CII and LPSAU methods and 3 sub-watersheds by LII method. Detention pond was implemented at one sub-watershed by CII and LPSAI methods, and 2 sub-watersheds by LII method. Filter strip was placed at 10 sub-watersheds by CII method and 11 sub-watersheds by LII and LPSAI methods. In the economic aspect scenario, filter strip was implemented for all sub-watersheds by three spatial targeting methods due to the lowest construction cost (**Table 7.4**). In the environmental aspect scenario, detention pond was placed at 5 sub-watersheds by three spatial targeting methods. Bioretention was selected at 11 sub-watersheds by CII method and

12 sub-watersheds by LII method. Filter strip was placed at 13 sub-watersheds by LPSAI method.

Table 7.7. Pairwise comparison matrix created for sub-watershed 78 based on sediment reduction for CII method at sub-watershed level.

LID/BMPs	Bioretention	Detention pond	Filter strip	Grassed swale	Wetland
Bioretention	1.00	1.23	2.01	2.28	1.50
Detention pond	0.81	1.00	1.78	2.06	1.27
Filter strip	0.50	0.56	1.00	1.27	0.66
Grassed swale	0.44	0.49	0.79	1.00	0.56
Wetland	0.67	0.79	1.51	1.78	1.00
λ^{max} = 5.004, CI = 0.001, CR = 0.091 %					

Table 7.8. Pairwise comparison matrix created for sub-watershed 78 based on TP reduction for CII method for CII method at sub-watershed level.

LID/BMPs	Bioretention	Detention pond	Filter strip	Grassed swale	Wetland
Bioretention	1.00	0.91	1.17	0.91	1.17
Detention pond	1.10	1.00	1.27	1.00	1.27
Filter strip	0.86	0.79	1.00	0.79	1.00
Grassed swale	1.10	1.00	1.27	1.00	1.27
Wetland	0.86	0.79	1.00	0.79	1.00
λ^{max} = 5.000, CI = 0.000, CR = 0.001 %					

Table 7.9. Pairwise comparison matrix created for sub-watershed 78 based on TN reduction for CII method at sub-watershed level.

LID/BMPs	Bioretention	Detention pond	Filter strip	Grassed swale	Wetland
Bioretention	1.00	1.46	1.63	1.28	1.46
Detention pond	0.68	1.00	1.16	0.85	1.00
Filter strip	0.62	0.86	1.00	0.75	0.86
Grassed swale	0.78	1.18	1.34	1.00	1.18
Wetland	0.68	1.00	1.16	0.85	1.00
λ^{max} = 5.000, CI = 0.000, CR = 0.01%					

Table 7.10. Weight vector calculation of LID/BMPs for sediment, TN, and TP reduction for CII method at sub-watershed 78.

LID/BMPs	Sediment	TN	TP
Bioretention	0.26	0.20	0.23
Detention pond	0.23	0.21	0.19
Filter strip	0.16	0.19	0.18
Grassed swale	0.14	0.21	0.21
Wetland	0.21	0.19	0.19

Table 7.11. Weight vector calculation of LID/BMPs for total cost and area condition.

LID/BMPs	Cost	Area condition*
Bioretention	0.10	0.125
Detention pond	0.21	0.250
Filter strip	0.35	0.125
Grassed swale	0.25	0.250
Wetland	0.19	0.250

*this weight vector is based on sub-watershed 78. This value is varied by each sub-watershed.

Table 7.12. Decision matrix of LID/BMPs for all criteria developed for sub-watershed 78.

LID/BMPs	Sediment	TN	TP	Total cost	Area condition
Bioretention	0.26	0.20	0.23	0.10	0.13
Detention pond	0.23	0.21	0.19	0.21	0.25
Filter strip	0.16	0.19	0.18	0.35	0.13
Grassed swale	0.14	0.21	0.21	0.25	0.25
Wetland	0.21	0.19	0.19	0.19	0.25

Table 7.13. The selected individual LID/BMPs based on each application scenario at CHSs using CII method.

Sub-watershed	Equal condition	Economic condition	Environmental condition
14	Grassed swale	Filter strip	Detention pond
21	Grassed swale	Filter strip	Detention pond
28	Filter strip	Filter strip	Bioretention
42	Grassed swale	Filter strip	Detention pond
44	Grassed swale	Filter strip	Detention pond
49	Filter strip	Filter strip	Bioretention
50	Filter strip	Filter strip	Bioretention
52	Filter strip	Filter strip	Bioretention
71	Filter strip	Filter strip	Bioretention
72	Detention pond	Filter strip	Detention pond
73	Filter strip	Filter strip	Bioretention
74	Filter strip	Filter strip	Bioretention
75	Filter strip	Filter strip	Bioretention
76	Grassed swale	Filter strip	Bioretention
77	Filter strip	Filter strip	Bioretention
78	Filter strip	Filter strip	Bioretention

Table 7.14. The selected individual LID/BMPs based on each application scenario at CHSs using LII method.

Sub-basin	Equal condition	Economic condition	Environmental condition
47	Filter strip	Filter strip	Bioretention
48	Filter strip	Filter strip	Bioretention
61	Filter strip	Filter strip	Bioretention
62	Filter strip	Filter strip	Bioretention
63	Filter strip	Filter strip	Bioretention
64	Filter strip	Filter strip	Bioretention
68	Detention pond	Filter strip	Detention pond
69	Filter strip	Filter strip	Bioretention
70	Grassed swale	Filter strip	Detention pond
71	Detention pond	Filter strip	Bioretention
72	Grassed swale	Filter strip	Detention pond
73	Filter strip	Filter strip	Bioretention
74	Filter strip	Filter strip	Bioretention
75	Grassed swale	Filter strip	Detention pond
76	Filter strip	Filter strip	Bioretention
77	Grassed swale	Filter strip	Detention pond
78	Filter strip	Filter strip	Bioretention

Table 7.15. The selected individual LID/BMPs based on each application scenario at CHSs using LPSAI method.

Sub-basin	Equal condition	Economic condition	Environmental condition
44	Grassed swale	Filter strip	Detention pond
45	Grassed swale	Filter strip	Detention pond
47	Grassed swale	Filter strip	Detention pond
55	Filter strip	Filter strip	Filter strip
59	Grassed swale	Filter strip	Detention pond
60	Grassed swale	Filter strip	Filter strip
61	Filter strip	Filter strip	Filter strip
62	Filter strip	Filter strip	Filter strip
64	Filter strip	Filter strip	Filter strip
67	Detention pond	Filter strip	Detention pond
68	Filter strip	Filter strip	Filter strip
71	Filter strip	Filter strip	Filter strip
72	Filter strip	Filter strip	Filter strip
73	Filter strip	Filter strip	Filter strip
76	Filter strip	Filter strip	Filter strip
77	Filter strip	Filter strip	Filter strip
78	Filter strip	Filter strip	Filter strip

7.4.3. Evaluation of LID/BMP effectiveness

The pollutant load reduction and efficiency of LID/BMPs implementation vary among the LID/BMPs types, placement, and spatial targeting methods. **Table 7.16** and **Figure 7.5** show the pollutant load reduction per unit area (kg/ha) and total cost per unit pollutant load reduction (1,000 $/kg) associated with the priority scenarios and spatial targeting methods. Overall, filter strip was the most economical LID/BMP for all targeting methods at the sub-watershed levels compared with other LID/BMPs. An intermediate effectiveness of LID/BMPs was detention pond and bioretention due to higher construction cost and pollutant load reduction rate.

The effectiveness of spatial targeting methods using the priority scenarios was analyzed that LPSAI method was the greatest reduction of the average pollutant load by 4.213 kg/ha in economic aspect scenario. CII method was the most cost-effective by 195,728

dollars/kg due to the lower total cost per unit pollutant load reduction for TN and TP in economic aspect scenario. In the environment aspect scenario, LPSAI method showed the greatest reduction of the average pollutant load by 4.007 kg/ha. It was also the most cost-effective by 371,453 dollars/kg owing to a higher pollutant load reduction within the small area of ID/BMP implementation. In the equal distribution scenario, LPSAI method indicated the most efficient for the average pollutant load reduction by 3.762 kg/ha and cost-effective and by 234,736 dollars/kg, respectively.

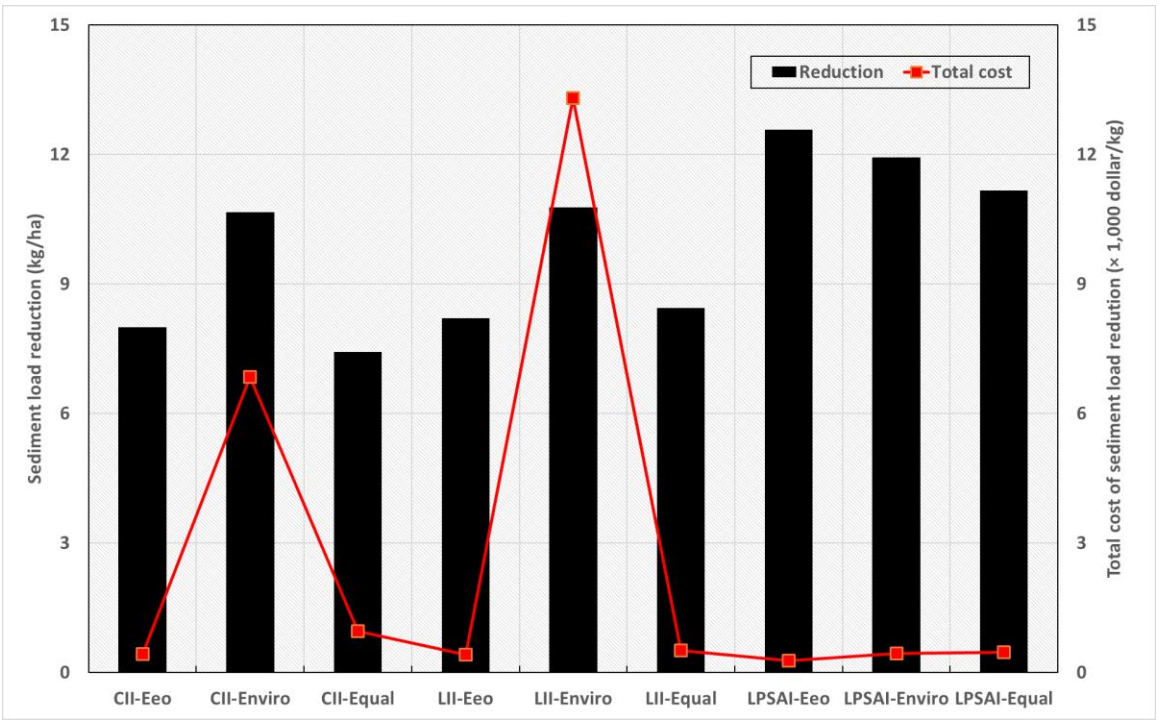
The reduction efficiency of each pollutant load indicated CII method was the most efficiency to reduce TN and TP loads by 0.147 kg/ha and 0.012 kg/ha in the environmental aspect scenario, respectively. LPSAI method was the most efficiency to decrease sediment loads by 12.565 kg/ha in the economic aspect scenario.

The cost-effective for the priority scenarios application indicates that CII method was the most cost-effective by TN and TP load reductions in the economic aspect scenario, while LPSAI method was the most cost-effectiveness for sediment load reduction. In the environment aspect and equal distribution scenarios, LPSAI method was observed the most cost-effective for sediment, TN, and TP load reductions.

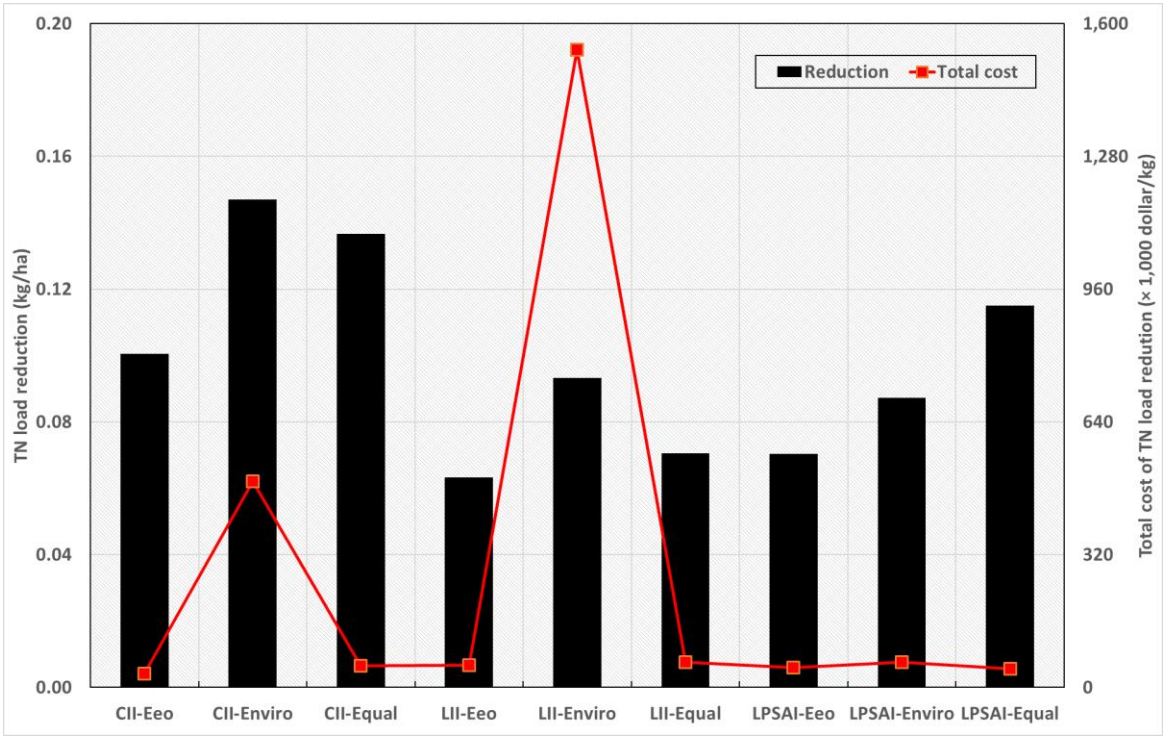
Water resources management plans for water quality improvement should be considered various LID/BMP implementation strategies related to the target pollutant elements, reduction rate, and budget. Therefore, the study results will provide useful insights for the broader view of the water resources management plans to stakeholders and decision makers.

Table 7.16. Total pollutant load reduction per unit area and total cost per unit pollutant load reduction for LID/BMPs implementations.

Priority Scenario	Targeting method	Load reduction per unit area (kg/ha)				Total cost per unit load reduction (1,000 $/kg)			
		Sediment	TN	TP	Average	Sediment	TN	TP	Average
Economic aspect	CII	7.996	0.101	0.006	2.701	0.425	33.818	552.941	195.728
	LII	8.199	0.063	0.004	2.755	0.415	53.681	916.651	323.582
	LPSAI	12.565	0.070	0.004	4.213	0.271	48.374	813.474	287.373
Environment aspect	CII	10.661	0.147	0.012	3.607	6.843	496.152	6,000.311	2,167.769
	LII	10.774	0.093	0.009	3.626	13.310	1537.079	1,521.6591	5,588.993
	LPSAI	11.928	0.087	0.005	4.007	0.442	60.418	1,053.500	371.453
Equal distribution	CII	7.416	0.137	0.010	2.521	0.950	51.570	740.291	264.270
	LII	8.444	0.071	0.004	2.839	0.503	60.160	1027.458	326.707
	LPSAI	11.164	0.115	0.008	3.762	0.464	44.991	658.755	234.736

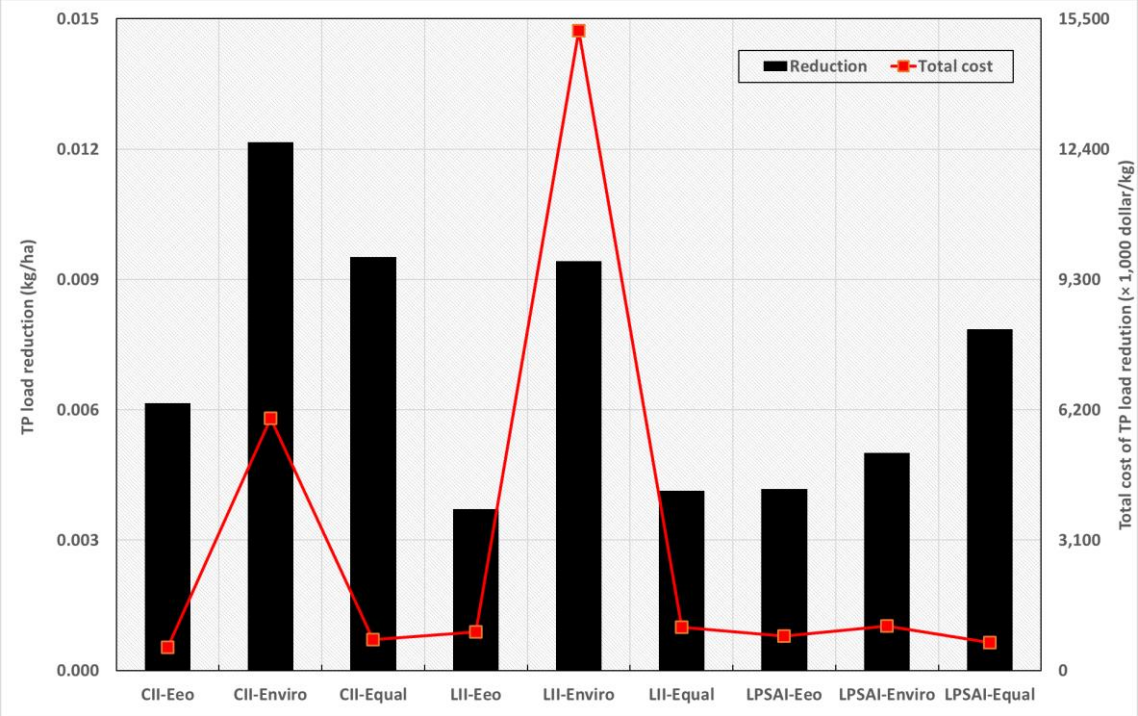


(a) Sediment load

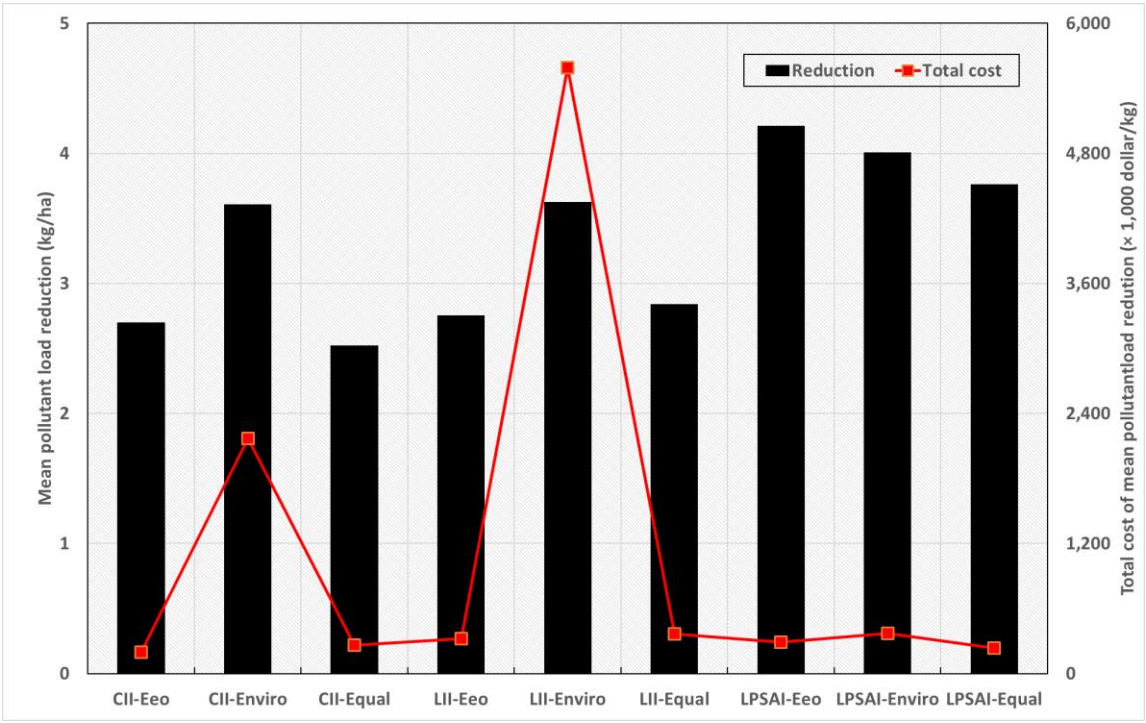


(b) TN load

Figure 7.5. Pollutant load reduction per unit area and total cost per unit pollutant load reduction for different targeting methods with priority consideration scenarios.



(c) TP load



(d) Average pollutant load (sediment, TN, and TP)

Figure 7.5. Pollutant load reduction per unit area and total cost per unit pollutant load reduction for different targeting methods with priority consideration scenarios (Cont.).

CHAPTER 8. SUMMARY AND CONCLUSIONS

8.1. Quantifying a threshold of missing values for gap filling processes in daily precipitation

Chapter 2 was to identify a proper level of missing values (threshold) by enhancing the existing estimation method (GDF) for gap filling in daily precipitation series for hydroclimate research and hydrological fields. The GDF method was used to examine what level of missing values can be plausible for further hydrological processes, such as streamflow simulations. Four different sampling windows, including whole, front, middle, and rear section of datasets at different missing levels (5, 10, 15, 16, 17, and 18 %), are considered to mimic real world setting. And, 1,000 samples are generated using an AR model to conduct statistical tests and analysis. AR (23) model was selected and utilized to generate 1,000 realizations of precipitation datasets at two different locations, including Boise airport, Idaho and Lincoln airport, Nebraska to cross-validate the proposed threshold.

The R^2 values and RDIs were computed and used to determine what missing levels can be considered as a threshold within statistical analysis framework along with hypotheses tests and p-value analysis. Based on evaluation criteria, which are $R^2 > 0.8$ and RDIs>0, it appears that the missing level at 15 % is plausible to construct daily precipitation series during gap filling processes. Although we conclude that 15 % missing level is plausible for the GDF method, it doesn't necessarily that 15 % missing level is an absolute value for other interpolation techniques, such as Mn, IDW, RA, KE, and others. Additional research is needed to verify if the proposed 15 % missing level is also acceptable for such cases.

Another insight from **chapter 2** is related to temporal scale up. Thus, **chapter 2** focused on gap filling for daily precipitation series so that the missing level at 15 % can be conservative for monthly gap filling. So, we may conclude that the advantage gained by temporal time scale at daily will be tangible unless significant systematic and random errors are embedded in precipitation records.

8.2. A heuristic gap filling method for daily precipitation series

Chapter 3 was conducted to evaluate how CA can improve the gap filing processes as a preliminary task for daily precipitation series. A total of 116 stations, which are less than 15% missing level of daily precipitation are used across the state of Idaho. All stations are evaluated first using CA and divided into 12 clusters based on geophysical conditions, including latitude, longitude, and altitude. Since the study area are dominated by maritime climate, it is necessary to minimize errors (e.g., the overestimated and/or underestimated precipitation) affected by the distance between stations nearby. We proposed a new method termed "gamma distribution function with statistical correction coupled with cluster analysis (GSCCA)" to improve the estimation of missing daily precipitation data. The existing methods, including GME, IDW, OK are also examined by employing CA to compare their performance with GSCCA. Using skill scores, performance measures are evaluated to see how individual method can represent the precipitation realization in the mountain west, such as the state of Idaho. The results of **chapter 3** are as follows:

1. GSCCA is a promising method to estimate daily precipitation at TSs, where the data missing level is less than 15% in the sense that all statistic and skill scores are higher than its counterparts of other methods (e.g., GME, IDW, OK).

2. The CA employment for gap filling methods can improve the overall performance as opposed to non-CA implementation. Note that the average R values of GSC, GME, IDW, and OK method increases by 17.19%, 9.84%, 8.93%, and 37.50 %, respectively, when CA is employed as opposed to non-CA implementation. The average RMSE and MAE also advocate the overall performance improvement as shown in **Table 3.4**. Additional condition, such as rain and no rain was also elevated to see how the proposed method can represent such a condition reasonably.
3. Based on skill scores as shown in **Figure 3.5**, the result clearly show GSC well presents rain or no rain condition with the highest HSS, CSI, and PSS values, 0.55, 0.51, and 0.56 on average, respectively.

Chapter 3 delivers few critical observations and findings. First, CA is a good approach to improve the existing gap filling methods. Second, GSC is a better estimation method as opposed to other methods when CA is employed for daily precipitation series. Although the proposed method is a heuristic and feasible solution, additional studies will be required to promote broad applications. Thus, since characteristics of various geophysical and environmental variables, including topography, climate, and other environmental forcing are critical to interact dynamically, more experiments in different climate and landscape regions may be needed for future work.

8.3. Threshold of basin discretization levels for HSPF simulations with NEXRAD inputs

High-quality hydrologic data and automatic calibration methods were used to evaluate how basin discretization affects HSPF simulations with NEXRAD inputs. For the study area,

four drainage areas, including Tahlequah, Siloam Springs, Elk, and Baron Fork, were selected for performance evaluation, model parameterization, and calibration. PEST, an automatic calibration package, was utilized to calibrate interior points of each drainage area, rather than using the observed streamflow at outlet points. RMSE, R, and NS were computed to evaluate the model performance at different catchment scales. The results of **chapter 4** are as follows:

1. Overall performances of HSPF are reasonably good at the Baron Fork, Tahlequah, Siloam Springs drainages. However, the Elk River basin shows a rather poor performance compared with other results; thus, the model performance should be enhanced through additional calibration efforts, such as minor adjustment characterizing local parameter sets (e.g., soil porosity) when necessary.
2. Two groups constrained by R and NS values were identified to evaluate how the performance of the model responds to the derived threshold of basin discretization from this study, which is approximately 23% of the drainage area.
3. The results showed that the model performance at all CPs above the threshold clearly improve as basin size decreases, whereas the performance below this threshold is inconclusive. For example, the R values at CP10, CP11, and CP14, which are 3.62, 1.97, and 1.49%, respectively, of the Tahlequah drainage area, decrease as the R values increase at Siloam Springs, Elk, and Baron Fork River, regardless of catchment sizes within the drainage area below the threshold (23%).
4. A similar result was also observed in NS values that fluctuate (e.g., decrease or increase) for the catchment size of 433 km2 (17.43% of the Tahlequah drainage). This pattern continues at Siloam Springs drainage and Baron Fork basin.

Consequently, the results in **chapter 4** indicate that HSPF performs well when the catchment size is above 23% of the drainage area. Although this study provides useful insights for hydrologists to determine appropriate catchment sizes during basin delineation processes, additional studies may be required for reliable applications. Thus, future experiments focusing on the basin size threshold of HSPF using different basin characteristics and different data sets may be helpful to leverage broad applications of the proposed approach. These more diverse regional applications will support the conclusion provided in **chapter 4**.

8.4. Improving hydrological simulation using HSPF in computer parallelism

Computer parallelism is a useful tool to reduce the amount of computation time and cost. Hydrologic model calibration coupled with computer parallelism gives various opportunities to improve the model performance. In this study, four different calibration scenarios were tested to evaluate computer parallelism. BEOPEST was utilized to parallelize HSPF model calibration in sLCS. Performance measures (e.g., PT, PP, PS, PE) for computer parallelism were used and other hydrological stats (R, NS, RMSE, MAE, and PABIS) were calculated to evaluate the model performance. The optimal calibration scenario was selected for further analysis to improve model performance for interior calibration target points in the BRW. The results of **chapter 5** are as follows:

1. Parallel time (PT) during HSPF calibration was tremendously reduced 76.09%, 88.63%, 89.30%, and 90.41% from SCO 1, SCO 2, SCO 3, and SCO 4, respectively when 14 core were used rather than single core. From all other parallel performance measures (PT, PP, PS, PE), SOC 4 outperforms others.

Therefore, the application of many model parameters is more effective when applying the parallelized calibration technique.

2. SCO 4 was selected the optimal calibration scenario based on NS (0.31), R (0.62), RMSE (30.03), and PBIAS (-10.65%) at the mouth of the watershed. The model performance measures of SCO 4 indicated greater model performance than other calibration scenarios.
3. SCO 4 was applied for further analysis to improve model performance for interior Calibration Target Points (CTPs) in the BRW. The calibrated streamflow results showed that CTP 3, 4, and 6 were very good performance by calibration and CTP 1 and 2 were good performance. The validated streamflow results indicated that CTP 1, 3, and 4 were very good performance and CTP 2 is fair performance. However, CTP 5 was poor performance by calibration and validation.
4. Finally, the calibrated sediment and water quality results showed that water temperature was very good performance. Sediment, TN, TP, and DO were good performance according to model performance evaluation based on PBAIS.

In chapter 5, we can conclude that computer parallelism with many parameters at multiple sub-basins benefits hydrologists to improve hydrological simulation. This study would be a good case by comparing parallel performance at multiple calibration target points using multiple core processes. The method used in this study would be useful for hydrologists to apply computer parallelism to their own applications. In the future, further studies need to be completed to evaluate parallel performance and model improvement using different model parameters from other hydrologic models.

8.5. Impacts of climate and land use changes on streamflow and water quality

From **chapter 6**, we investigated the response of hydrological processes for the separate and combined impacts of future changes in climate and LULC using three SDBC GCMs in the BRW. HSPF model simulated streamflow and water quality under each scenario during the future peirods. Study results indicate as follows:

1. The impacts of LULC change were especially clear in the monthly variations of streamflow, sediment, TN, and TP loads. Urbanization had a much greater effect on an increase in magnitude and seasonal variability of monthly streamflow, sediment, TN, and TP loads. Barren/mining land highly affected the increase of TN load comparing with 2011 LULC condition.
2. SDBC GCMs (bcc-cm1-1m, cancesm2, and cnrm-cm5 models) showed good performance for historical monthly precipitation, temperature, wind speed, humidity, and solar radiation. They well simulated historical monthly streamflow compared with NLDAS simulation results at calibration target point 6.
3. For The hydrologic responses by climate change indicated monthly mean streamflow, sediment, TN, and TP loads were decreased in late spring, summer, and fall while they were increased in early spring and winter relative to 2011 LULC condition due to the future changes in precipitation and temperature under three RCPs. Water temperature was increased along with the increased air temperature under three RCPs over the future periods relative to 2011 LULC.
4. The combined impact of climate and LULC changes indicated similar to the climate change alone, but with increased streamflow, sediment, and TN, and TP load throughout the year. Also, it was predicted to significantly increases in

monthly mean streamflow, sediment, TP, and TN loads in spring and winter comparing with climate change alone. Therefore, the combined impact of future climate and LULC changes represented the problems of water shortage in summer and severe water pollution in spring and winter.

In **chapter 6**, we can conclude that urbanization affects the increases in monthly streamflow, sediment, TN, and TP loads by the increased impervious surface area. For climate change, monthly streamflow, sediment, TN, and TP loads were increased in early spring and winter due to the early snow melting and increase in winter precipitation, while they were decreased in summer owing to the decrease in summer precipitation. The combined impact of climate and LULC changes accelerated the increase in water pollution in the spring and winter. Understanding the changes in water quality and quantity caused by the separate and combined impacts of LULC and climate changes is significant elements for sustainable water resources management. The seasonal variations of streamflow and water quality in the BRW is expected to become more severe in the future. Also, a steadily increase in urbanization of the BRW will be believed to increase the intensity of drought or flooding. Long-term water resources management plans should be flexible and adaptable to diverse perspectives on these impacts. The results of **chapter 6** can contribute to effective management planning for land use development and mitigation of negative effects by LULC and climate changes.

8.6. Development of an optimization watershed management tool

LID/BMPs are one of the primary options in improving surface water quality. Optimal placement of LID/BMPs plays a critical role in an accomplishment of maximum pollutant load reduction while minimizing costs at CHSs. So, **cheater 7** was carried out with the goals

of investigating the CHSs using three targeting methods and selecting optimal LID/BMPs at the selected CHSs using AHP method. The study results are as follows.

1. CHSs were determined for the high priority areas using normalized pollutant results (sediment, TN, and TP) based on three spatial targeting methods. 16 sub-watersheds were selected as CHSs using CII method. They were mostly located in the LBRW. LII method identified 15 sub-watersheds for CHSs. They were located near the watershed outlet or in the middle area of the BRW. LPSAI method identified 17 sub-watersheds for CHSs.
2. AHP method was applied to identify optimal LID/BMPs for the priority scenarios (equal, economic, and environmental aspect) at CHSs from three spatial targeting methods. In the economic aspect scenario, filter strip was placed at all CHSs regardless of spatial targeting methods due to the lowest construction cost. In the environmental aspect scenario, 5 detention ponds and 11 biorentention were implemented by CII method, 5 detention ponds and 12 bioretentions were placed by LII method, and 5 detention ponds and 13 filter strips were selected by LPSAI method.
3. Varying effectiveness results were analyzed when LID/BMP was implemented for the priority scenarios at CHSs based on the three spatial targeting methods. LPSAI method was the most efficiency to reduce average pollutant load by 4.213 kg/ha in economic aspect scenario, 4.007 kg/ha in the environment aspect scenario, and 3.762 kg/ha in the equal distribution scenario. CII method was the most cost-effective by 195,728 dollars/kg in the economic aspect scenario. LPSAI method was most cost-effective by 371,453 dollars/kg in the environmental aspect scenario

and 234,736 dollars/kg in the equal distribution scenario, respectively.

LID/BMPs were a proven to control water quality in the watershed. The effectiveness of LID/BMP implementation were subject to change with the placement locations and consideration objectives like economic or environmental aspects. Spatial targeting methods can be utilized based on the goals stated in LID/BMP implementation plans. Therefore, the results of **chapter 7** would provide a platform in with diverse aspects (equal distribution, economic, and environmental) of LID/BMP implementation. The results of **chapter 7** can strengthen science-based decision making and provide useful perspectives to policy makers in cost-effective water resources management. However, in the future, it will be necessary to improve OWMT by expanding the number of scenarios through conducting a survey of farmers or policy makers in the watershed to apply their preference and criteria for LID/BMPs implementation.

References

Acharya, A., and Ryu, J. H. (2013). Simple method for streamflow disaggregation. *Journal of Hydrologic Engineering, 19*(3), 509-519.

Adhikary, S.K., Muttil, N., and Yilmaz, A.G. (2015). Genetic Programming-Based Ordinary Kriging for Spatial Interpolation of Rainfall. *Jounral of Hydrologic Engineering*, 10.1061/(ASCE)HE.1943-5584.0001300 , 04015062.

Ahmed, K. F., Wang, G., Silander, J., Wilson, A. M., Allen, J. M., Horton, R., and Anyah, R. (2013). Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the US northeast. *Global and Planetary Change, 100*, 320-332.

Ahrens, B. (2006). Distance in spatial interpolation of daily rain gauge data. *Hydrology and Earth System Sciences*. 10:197–208.

Ajami, N., Gupta, H., Wagener, T., and Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, 298(1–4), 112–135.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds), *2nd International Symposium on Information Theory*, 267–281.

Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences*, 24:419–428.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). Crop Evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome, 300*(9), D05109.

Aulinas, M., Nieves, J. C., Cortés, U., and Poch, M. (2011). Supporting decision making in urban wastewater systems using a knowledge-based approach. *Environmental Modelling & Software, 26*(5), 562-572.

ASCE. 1996. Hydrology Handbook, *2nd edition. American Society of Civil Engineers ASCE, New York*.

Banai, R. (1993). Fuzziness in Geographical Information Systems: contributions from the analytic hierarchy process†. *International Journal of Geographical Information Science, 7*(4), 315-329.

Barry, R.G., and Chorley, R. J. (1987). Atmosphere, Weather and Climate, 5*th ed. Routledge: London*.

Bell, V. A., and Moore, R. J. (1998). A grid-based distributed flood forecasting model for use with weather radar data: Part 2. Case studies. *Hydrology and Earth System Sciences*, 2(3), 283–298.

Bicknell, B., Imhoff, J., Kittle Jr, J., Jobes, T., and Donigian Jr, A. (2001). Hydrologic Simulation Program-Fortran (HSPF) User's Manual for Version 12. *US Environmental Protection Agency, National Exposure Research Laboratory, Athens, GA*.

Borah, D., and Bera, M. (2003). Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Transactions of the ASAE, 46*(6), 1553.

Boyle, D. P., Gupta, H. V., and Sorooshian, S. (2001). Toward improved streamflow forecasts: Value of semidistributed modeling. *Water Resources Research*, 37(11), 2749–2759

Brath, A., Montanari, A., and Moretti, G. (2006). Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). *Journal of Hydrology, 324*(1), 141-153.

Breiger, R. L., Boorman, S. A., and Arabie, P. (1975). An algorithm for clustering relation data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12 (3), 328-384.

Brown, R. G. (2004). Engineering a Beowulf-style compute cluster. *Duke University Physics Department*.

Brown, W., and Schueler, T. (1997). The economics of stormwater BMPs in the mid-Atlantic region. prepared for: Chesapeake research consortium. Edgewater, MD. *Center for Watershed Protection. Ellicott City, MD*.

Bryan, B. A., and Crossman, N. D. (2008). Systematic regional planning for multiple objective natural resource management. *Journal of environmental management, 88*(4), 1175-1189.

Bunkers, M. J., Miller J. R., and Degaetano, A. T. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *Journal of Climatology*, 9, 130-146.

Bürger, G., Murdock, T., Werner, A., Sobie, S., and Cannon, A. (2012). Downscaling extremes-an intercomparison of multiple statistical methods for present climate. *Journal of Climate, 25*(12), 4366-4388.

Calder, IR. (1993). In handbook of hydrology. Maidment D (ed.), McGraw-Hill, New York, 18.12–18.13.

Carpenter, T. M., and Georgakakos, K. P. (2004). Continuous streamflow simulation with the HRCDHM distributed hydrologic model. *Journal of Hydrology*, 298(1–4), 61–79

Chang, K.T. (2009). Introduction to geographic information systems. *New Tork: McGraw Hill*.

Chattopadhyay, S., Jhajharia, D., and Chattopadhyay, G. (2011). Univariate modelling of monthly maximum temperature time series over northeast India: neural network versus Yule-Walker equation based approach. *Meteorological Applications*, 18(1):70–82.

Chmielewski, M. S., and Grzymala-Busse, J. W. (1996). Global discretization of continuous attributes as preprocessing for machine learning. *International Journal of Approximate Reasoning*, Vol 15(4), 319-331.

Chen, J., Brissette, F. P., and Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology, 401*(3), 190-202.

Chen, M.S., Han, J., and Yu, P.S. (1996). Data mining: an overview from a database perspective. *IEEE Trans Knowl Data Eng*, 8(6):866–883.

Chiew, F., Kamaladasa, N., Malano, H., and McMahon, T. (1995). Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agricultural Water Management, 28*(1), 9-21.

Childers, D. L., Pickett, S. T., Grove, J. M., Ogden, L., and Whitmer, A. (2014). Advancing urban sustainability theory and action: Challenges and opportunities. *Landscape and Urban Planning, 125*, 320-328.

Cho, J. H., Sung, K. S., and Ha, S. R. (2004). A river water quality management model for optimizing regional wastewater treatment using a genetic algorithm. *Journal of environmental management*, *73*(3), 229-242.

Commission, T. N. R. C. (1998). State of Texas 1998 Clean Water Act Section 303 (d) List and Schedule for Development of Total Maximum Daily Loads. *SFR-58. Austin, TX: TNRCC*.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Duan, Q. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(D22).

County, F. (2014). National Pollutant Discharge Elimination System.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of applied meteorology*, *33*(2), 140-158.

Davis, J.C. (2002). Statistics and data analysis in geology. *New York: John Wiley*.

Dhalla, S., and Zimmer, C. (2010). Low Impact Development Stormwater Management Planning and Design Guide, Version 1.0. *Toronto and Region Conservation for the living city*.

Ding, J., Jiang, Y., Fu, L., Liu, Q., Peng, Q., and Kang, M. (2015). Impacts of land use on surface water quality in a Subtropical River Basin: a case study of the Dongjiang River Basin, Southeastern China. *Water*, *7*(8), 4427-4445.

Dirks, K.N., Hay, J.E., Harris, D. (1998). High-resolution studies of rainfall on Norfolk island part II: interpolation of rainfall data. *Journal of Hydrology*, 208(3–4):187–193.

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., Thornbrugh, D. J. (2008). Eutrophication of US freshwaters: analysis of potential economic damages. *Environmental Science & Technology*, *43*(1), 12-19.

Doherty, J., and Johnston, J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model. *Journal of the American Water Resources Association*, 39(2), 251–265.

Doherty, J., and Skahill, B. E. (2006). An advanced regularization methodology for use in watershed model calibration. *Journal of Hydrology, 327*(3), 564-577.

Donigian, A. S. (1977). Agricultural runoff management (ARM) model version II: refinement and testing. *Environmental Protection Agency*, Office of Research and Development, Environmental Research Laboratory.

Donigian Jr, A. (2000). HSPF training workshop handbook and CD: Lecture.

Donigian Jr, A., and Crawford, N. (1976). *Modeling Pesticides and Nutrients on Agricultural Lands, Office of Research and Development, US Environmental Protection Agency*.

Eastman, J. R., Kyem, P. A., and Toledano, J. (1993). A procedure for multi-objective decision making in GIS under conditions of conflicting objectives. *Paper presented at the Proceedings of European conference on geographical information systems*, EGIS.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, Vol 95, 14863-14868.

Eisner, S., Voss, F. and Kynast, E. (2012) Statistical Bias Correction of Global Climate Projections—Consequences for Large Scale Modeling of Flood Flows. *Advances in Geosciences*, 31, 75-82.

Faloutsos, C., and Oard, D. W. (1998). A Survey of Information Retrieval and Filtering Methods. *Technical Report of the Computer Science Department, University of Maryland*, No. CS-TR-3514.

Fienen, M. N., Muffels, C. T., and Hunt, R. J. (2009). On constraining pilot point calibration with regularization in PEST. *Ground Water*, 47(6), 835–844.

Fulton, R. A., Breidenbach, J. P., Seo, D.-J., and Miller, D. A. (1998). The WSR-88D rainfall algorithm. *Weather Forecast.*, 13(2), 377–395.

Gallagher, M., and Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software, 22*(7), 1000-1020.

Garcia, B.I.L., Sentelhas, P.C., Tapia, L., Sparovek, G. (2006). Filling in missing rainfall data in the Andes region of Venezuela, based on a cluster analysis approach. *Revista Brasileira Agrometeorologia*, 14 (2), 225–233.

Garcia, M., Peters-Lidard, C.D., and Goodrich, D.C. (2008). Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. *Water Resources Research,* 44(W05S13):1–14.

Gardner, J. W. (1991). Detection of vapours and odours from a multisensory array using pattern recognition Part 1. Principal and component and cluster analysis. *Sensors and Actuators B: Chemical,* Vol(4), 109-115.

Ghebremichael, L., Veith, T., and Watzin, M. (2010). Determination of critical source areas for phosphorus loss: Lake Champlain basin, Vermont. *Transactions of the ASABE, 53*(5), 1595-1604.

Giri, S., Nejadhashemi, A. P., and Woznicki, S. A. (2012). Evaluation of targeting methods for implementation of best management practices in the Saginaw River Watershed. *Journal of environmental management, 103*, 24-40.

Glotić, A., Glotić, A., Kitak, P., Pihler, J., and Tičar, I. (2014). Parallel self-adaptive differential evolution algorithm for solving short-term hydro scheduling problem. *IEEE Transactions on Power Systems, 29*(5), 2347-2358.

Göncü, S., and Albek, E. (2010). Modeling climate change effects on streams and reservoirs with HSPF. *Water resources management, 24*(4), 707-726.

Goegebeur, M., and Pauwels, V. R. N. (2007). Improvement of the PEST parameter estimation algorithm through extended Kalman filtering. *Journal of Hydrology,* 337(3–4), 436–451.

Goldberg, A.S. (2000). A course in econometrics. *Harvard University Press,* Cambridge.

Gorgan, D., Bacu, V., Mihon, D., Rodila, D., Abbaspour, K., and Rouholahnejad, E. (2012). Grid based calibration of SWAT hydrological models. *Natural Hazards and Earth System Sciences, 12*(7), 2411-2423.

Goswami, M., and O'Connor, K.M. (2007). Real-time flow forecasting in the absence of quantitative precipitation forecasts: a multi-model approach. *Journal of Hydrology,* 334(1–2):125–140.

Greene, D. R., and Hudlow, M. D. (1982). Hydrometeorologic grid mapping procedures. *AWRA Int. Symp. on Hydrometeorology*, American Water Resources Association, Denver, CO.

Grid Analysis and Display System (GrADS) [Computer software]. Center for Ocean-Land-Atmosphere Studies (COLA), *George Mason University*, Fairfax, VA.

Guo, H., Hu, Q., and Jiang, T. (2008). Annual and seasonal streamflow responses to climate and land-cover changes in the Poyang Lake basin, China. *Journal of Hydrology, 355*(1), 106-122.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering, 4*(2), 135-143.

Hansen, P., and Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming*, 79(1), 191-215.

Hasan, M. M., and Croke, B. F. W. (2013). Filling gaps in daily rainfall data: a statistical approach.*20th International Congress on Modeling and Simulation, Adelaide, Australia, 1-6 December 2013*.

Heistermann, M., Müller, C., and Ronneberger, K. (2006). Land in sight?: Achievements, deficits and potentials of continental to global scale land-use modeling. *Agriculture, Ecosystems & Environment, 114*(2), 141-158.

Hipel, K. W., and McLeod, A. I. (1994). Time series modelling of water resources and environmental systems, *Elsevier, Amsterdam*.

Howarth, R. W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., and Jordan, T. (1996). Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences *Nitrogen cycling in the North Atlantic Ocean and its watersheds* (pp. 75-139): Springer.

Hubbard, K. G., Goddard, S., Sorensen, W. D., Wells, N., and Osugu, T. T. (2005). Performance of Quality Assurance Procedures for an Applied Climate Information System. *Journal of Atmospheric and Oceanic Technology*, 22(1), 105-112.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R., Feddema, J., Fischer, G., and Janetos, A. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic change, 109*(1-2), 117-161.

Hydrocomp, I. (1976). *Hydrocomp Simulation Programming: Operations Manual* (2nd ed.). Palo Alto, California.

Hruschka, E. R., and Ebecken, N. F. (2003). A genetic algorithm for cluster analysis. *Intelligent Data Analysis*, 7(1), 15-25.

Isaake, H.E., and Srivastava, R.M. (1989). "An Intrudction to Applied Geostastistics." *Oxford University Press*, Oxford.

Im, S., Brannan, K. M., and Mostaghimi, S. (2003). Simulating hydrologic and water quality impacts in an urbanizing watershed. *Journal of the American Water Resources Association*, 39(6), 1465–1479.

Ing, C.-K., Sin, C.-y., and Yu, S.-H. (2010). Efficient selection of the order of an AR(infinity): A unified approach without knowing the order of integratedness. *Econometric Society, 10th World Congress, August 17-21, Shanghai, China*.

IPCC. (2007). The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change (pp. 337-383): et al. Cambridge: *Cambridge University Press*.

IPCC. (2013). IPCC, 2013: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.

Jarvie, H. P., Oguchi, T., and Neal, C. (2002). Exploring the linkages between river water chemistry and watershed characteristics using GIS-based catchment and locality analyses. *Regional Environmental Change, 3*(1-3), 36-50.

Jensen, M. E., and Burman, R. D., and Allen, R. G. (1990). Evapotranspiration and Irrigation Water Requirements. *New York, NY: ASCE, 978-0-87262-765-5,* pp330.

Jensen, M. E., and Haise, H. R. (1963). Estimating evapotranspiration from solar radiation. *Journal of Irrigation and Drainage Engineering*, 89(IR4), 15–41.

Journel, A.G., and Huijbregts, C.J. (1978). *Mining Geostatistics Academic, New York*.

Johnson, F., and Sharma, A. (2011). Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Water Resources Research, 47*(4).

Jordi, A., and Wang, D.-P. (2012). sbPOM: A parallel implementation of Princeton Ocean Model. *Environmental Modelling & Software, 38*, 59-61.

Jung, I.W., Chang, H. and Moradkhani, H. (2011) Quantifying Uncertainty in Urban Flooding Analysis Considering Hydro-Climatic Projection and Urban Development Effects. *Hydrology and Earth System Sciences*, 15, 617-633.

Kanevski, M., and Maignan, M. (2004). Analysis and modeling of spatial environmental data. *Italy: EPEL Press*.

Kalyanapu, A. J., Shankar, S., Pardyjak, E. R., Judi, D. R., and Burian, S. J. (2011). Assessment of GPU computational enhancement to a 2D flood model. *Environmental Modelling & Software, 26*(8), 1009-1016.

Kay, A.L., Davies, H.N., Bell, V. and Jones, R.G. (2009) Comparison of Uncertainty Sources for Climate Change Impacts: Flood Frequency in England. *Climatic Change*, 92, 41-63.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W., and Doherty, J. (2007). Comparison of hydrologic calibration of HSPF using automatic and manual methods. *Water Resources Research*, 43(1), W01402.

Kim, J., and Ryu, J. H. (2016). A Heuristic gap filling method for daily preciptiation series. *Water Resources Management, Vol.7*(11), 2275-2294.

Kim, J., and Ryu, J. H. (2015). Quantifying a threshold of missing values for gap filling processes in daily precipitation Serices. *Water Resources Management, Vol.29*(11), 4173-4184.

Kim, J., and Ryu, J. H. (2013). Threshold of basin discretization levels for HSPF simulations with NEXRAD inputs. *Journal of Hydrologic Engineering, 19*(7), 1401-1412.

Kingston, D.G. and Taylor, R.G. (2010) Sources of Uncertainty in Climate Change Impacts on River Discharge and Groundwater in a Headwater Catchment of the Upper Nile Basin, Uganda. *Hydrology and Earth System Sciences*, 14, 1297-1308. http://dx.doi.org/10.5194/hess-14-1297-2010.

Koren, V. I., Finnerty, B. D., Schaake, J. C., Smith, M. B., Seo, D.-J., and Duan, Q. Y. (1999). Scale dependencies of hydrology models to spatial variability of precipitation. *Journal of Hydrology.*, 217(3–4), 285–302.

Kumar, C. N. S., Ramulu, V. S., Reddy, K. S., Kotha, S., and Kumar, C. M. (2012). Spatial Data Mining using Cluster Analysis. *International Journal of Computer Science & Information Technology (IJCSIT)*, 4(4), 71-77.

Lecca, G., Petitdidier, M., Hluchy, L., Ivanovic, M., Kussul, N., Ray, N., and Thieron, V. (2011). Grid computing technology for hydrological applications. *Journal of Hydrology, 403*(1), 186-199.

Lee, K. S., and Chung, E. S. (2007). Hydrological effects of climate change, groundwater withdrawal, and land use in a small Korean watershed. *Hydrological Processes, 21*(22), 3046-3056.

Lerat, J., Perrin, C., Andréassian, V., Loumagne, C., and Ribstein, P. (2012). Towards robust methods to couple lumped rainfall–runoff models and hydraulic models: A sensitivity analysis on the Illinois River. *Journal of Hydrology, 418*, 123-135.

Li, T., Wang, G., Chen, J., and Wang, H. (2011). Dynamic parallelization of hydrological model simulations. *Environmental Modelling & Software, 26*(12), 1736-1746.

Li, H., Sheffield, J., and Wood, E.F. (2010). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate change AR4 models using equidistant quantile matching. *Journal of Geophysical Research*, Vol, 115, D10101, doi:10.1029/2009JD012882.

Li, X., Wei, J., Li, T., Wang, G., and Yeh, W. W.-G. (2014). A parallel dynamic programming algorithm for multi-reservoir system optimization. *Advances in Water Resources, 67*, 1-15.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H. (1982). Hydrology for Engineers. *McGraw-Hill*, New York.

Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.

Llyod, C.D. (2005). Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology*, 308, 128-150.

Lu. G.Y., and Wong, D.W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences*, 34(9), 1044-1055.

Lund, R., and Li, B. (2010). Revisiting Climate Region Definitions via Clustering. *Journal of Climatology*, 22, 1787-1800.

Luo, L., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., and Sheffield, J. (2003). Validation of the North American land data assimilation system (NLDAS) retrospective forcing over the southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *108*(D22).

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal on Applied Mathematics*, 11(2), 431–441.

Maarek, Y. S., Berry, D. M., and Kaiser, G. E. (1991). An Information Retrieval Approach for Automatically Constructing Software Libraries. *IEEE Transactions on Software Engineering*, Vol 17(8), 800-813.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observation. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 281-297.

Makowski, M., Somlyódy, L., and Watkins, D. (1996). MULTIPLE CRITERIA ANALYSIS FOR WATER QUALITY MANAGEMENT IN THE NITRA BASIN1: Wiley Online Library.

Mair, A., and Fares, A. (2011). Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island. *Journal of Hydrologic Engineering*, 16(4), 371-383.

Markovic, R. D. (1965). Probability Functions of Best Fit to Distributors of Annual Precipitation and Runoff. *Hydrology papers Colorado State University Fort Collins Colorado, P 119*.

Marshall, E., and Randhir, T. (2008). Effect of climate change on watershed system: a regional analysis. *Climatic change*, *89*(3-4), 263-280.

Maurer, E., and Hidalgo, H. (2007). Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences Discussions*, *4*(5), 3413-3440.

McCuen, R. H. (1998). Hydrologic Analysis and Design. *Prentice-Hall, NJ*.

McCuen, R. H., Knight, Z., and Cutter, A. G. (2006). Evaluation of the Nash-Sutcliffe efficiency index. *Journal of Hydrology Engineering*, 10.1061/(ASCE) 1084-0699(2006)11:6(597), 597–602.

McIntosh, B. S., Ascough, J., Twery, M., Chew, J., Elmahdi, A., Haase, D., and Jakeman, A. J. (2011). Environmental decision support systems (EDSS) development–challenges and best practices. *Environmental Modelling & Software, 26*(12), 1389-1402.

McKinney, D. C. (1999). *Modeling water resources management at the basin level: Review and future directions* (Vol. 6): IWMI.

Michaud, J., and Sorooshian, S. (1994). Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resources Research*, 30(3), 593–605.

Mitchell, K. E., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, 109(D7), D07S90.

Monteith, J. L. (1965). Evaporation and the environment. *State and Movement of Water in Living Organisms, XIX Symp. Society for Experimental Biology*, Cambridge University Press, Cambridge, U.K.

Nagaraju, D., Nassery, H., and Adinehvandi, R. (2012). Determine suitable sites for artificial recharge using hierarchical analysis process (AHP), remote sensing and geographical information system. *International Journal Earth Sciences and Engineering, 5*(5), 1328-1335.

Nan, Z., Wang, S., Liang, X., Adams, T. E., Teng, W., and Liang, Y. (2010). Analysis of spatial similarities between NEXRAD and NLDAS precipitation data products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*(3), 371-385.

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*(3), 282-290.

NCDENR (North Carolina Department of Environment and Natural Resources). (2007). *Stormwater BMP Costs*.

Onishi, Y., and Wise, S. E. (1982). Mathematical model, SERATRA, for sediment-contaminant transport in river and its application to pesticide transport in Four Mile and Wolf Creeks in Iowa. *BPNL, RICHLAND, WA(USA). 1982.*

Oyi, V. (2013). *Humidity conversion formulas.*

Pan, W., Lin, J., and Le, C. T. (2002). Model-based cluster analysis of microarray gene-expression data. *Genome Biology*, Vol 3(2), 1-8.

Panagopoulos, Y., Makropoulos, C., and Mimikou, M. (2012). Decision support for diffuse pollution management. *Environmental Modelling & Software, 30*, 57-70.

Pedrycz, W. (1990). Fuzzy sets in pattern recognition: Methodology and methods. *Pattern Recognition*, 23(1), 121-146.

Piani, C., Haerter, J., and Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology, 99*(1-2), 187-192.

Pinker, R. T., Tarpley, J. D., Laszlo, I., Mitchell, K. E., Houser, P. R., Wood, E. F., and Cosgrove, B. A. (2003). Surface radiation budgets in support of the GEWEX Continental-Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres, 108*(D22).

Price, R. K. (2011). Urban hydroinformatics: data, models, and decision support for integrated urban water management: *IWA publishing*.

Prowse, T., Beltaos, S., Gardner, J., Gibson, J., Granger, R., Leconte, R., and Toth, B. (2006). Climate change, flow regulation and land-use effects on the hydrology of the Peace-Athabasca-Slave system; Findings from the Northern Rivers Ecosystem Initiative. *Environmental Monitoring and Assessment, 113*(1-3), 167-197.

Raupach, M., and Finnigan, J. (1988). Single layer models of evaporation from plant canopies are incorrect but useful, whereas multilayer models are correct but useless: Discuss. *Australian Journal of Plant Physiology, 15*(6), 705-716.

Reed, S. M., and Maidment, D. R. (1999). Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling. *Journal of Hydrology Engineering*, 10.1061/(ASCE)1084-0699(1999)4:2(174), 174–182.

Refsgaard, J. C., and Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research.*, 32(7), 2189–2202.

Ryu, J. H. (2009). Application of HSPF to the distributed model intercomparison project: Case study. *Journal of Hydrology Engineering*, 10.1061/(ASCE)HE .1943-5584.0000054, 847–857.

Ryu, J.H., Palmer, R.N., Wiley, M.W., Jeong, S. (2009). Mid-range streamflow forecasts based on climate modeling statistical correction and evaluation. *J Am Water Resour Assoc* 45(2):355–368.

Rouholahnejad, E., Abbaspour, K. C., Vejdani, M., Srinivasan, R., Schulin, R., and Lehmann, A. (2012). A parallelization framework for calibration of hydrological models. *Environmental Modelling & Software, 31*, 28-36.

Roy, B. (1981). A multicriteria analysis for trichotomic segmentation problems. *Multiple criteria analysis: Operational methods*, 245-257.

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., and Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research: Atmospheres, 118*(19).

Saaty TL (1980) The analytic hierarchy process. *McGraw- Hill, New York*

Sadeghi, S., Jalili, K., and Nikkami, D. (2009). Land use optimization in watershed scale. *Land Use Policy, 26*(2), 186-193.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). Applied Modeling of Hydrologic Time Series. *Water Resources Publications, LLC*, P.O. Box 2841, Littleton, CO 80161, USA, pp468.

Schreüder, W. A. (2009). *Running BeoPEST.* Paper presented at the Proceedings, *PEST Conference*.

Sellami, H., Benabdallah, S., La Jeunesse, I., and Vanclooster, M. (2016). Quantifying hydrological responses of small Mediterranean catchments under climate change projections. *Science of the Total Environment, 543*, 924-936.

Seo, D.-J., and Breidenbach, J. P. (2002). Real-time correction of spatially nonunifrom bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*, 3(2), 93–111.

Seong, C., Her, Y., and Benham, B. L. (2015). Automatic calibration tool for Hydrologic Simulation Program-FORTRAN using a shuffled complex evolution algorithm. *Water, 7*(2), 503-527.

Sevruk, B. (1996). Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, 42(1-4), 237-246.

Shamir, E., Megdal, S. B., Carrillo, C., Castro, C. L., Chang, H.-I., Chief, K., and Nelson, K. M. (2015). Climate change and water resources management in the Upper Santa Cruz River, Arizona. *Journal of Hydrology, 521*, 18-33.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Paper presented at the Proceedings of the 1968 23rd ACM national conference*.

Simanton, J. R., and H. B. Osborn. (1980). Reciprocal-distance estimate of point rainfall. *Journal of Hydraulic Engineering*, 106, 1242–1246.

Simolo, C., Brunetti, M., Maugeri, M., and Nanni, T. (2010). Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *International Journal of Climatology*, 30(10), 1564-1576.

Sliva, L., and Williams, D. D. (2001). Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water research, 35*(14), 3462-3472.

Smith, M. B., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S. (2004). The distributed model intercomparison project (DMIP): Motivation and experiment design. *Journal of Hydrology*, 298(1–4), 4–26

Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S. M., Cui, Z., and Anderson, E. A. (2012). Results of the DMIP 2 Oklahoma experiments. *Journal of Hydrology, 418*, 17-48.

Smith, R. E., Goodrich, D. C., Woolhiser, D. A., and Unkrich, C. L. (1995). KINEROS: A kinematic runoff and erosion model in computer models of watershed hydrology. *Water Resources Publications*, Littleton, CO, 697–732.

Sohl, T. L., Sayler, K. L., Bouchard, M. A., Reker, R. R., Friesz, A. M., Bennett, S. L., Soulard, C. (2014). Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States. *Ecological Applications, 24*(5), 1015-1036.

Sohl, T. L., Sleeter, B. M., Sayler, K. L., Bouchard, M. A., Reker, R. R., Bennett, S. L., and Zhu, Z. (2012a). Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. *Agriculture, Ecosystems & Environment, 153*, 1-15.

Sohl, T. L., Sleeter, B. M., Zhu, Z., Sayler, K. L., Bennett, S., Bouchard, M., and Liu, S. (2012b). A land-use and land-cover modeling strategy to support a national assessment of carbon stocks and fluxes. *Applied Geography, 34*, 111-124.

Sprinthall, R.C. (2011). Basic Statistical Analysis. 9th Edition. Pearson Education Group.

Strayer, D. L., Beighley, R. E., Thompson, L. C., Brooks, S., Nilsson, C., Pinay, G., and Naiman, R. J. (2003). Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. *Ecosystems, 6*(5), 407-423.

Strengers, B., Leemans, R., Eickhout, B., de Vries, B., and Bouwman, L. (2004). The land-use projections and resulting emissions in the IPCC SRES scenarios as simulated by the IMAGE 2.2 model. *GeoJournal, 61*(4), 381-393.

Teegavarapu, R.S.V. (2007). Use of universal function approximation in variance-dependent surface interpolation method: An application in hydrology. *Journal of Hydrology*, 332(1-2), 16-29.

Teegavarapu, R.S.V, Tufail, M,l and Ormsbee, L (2009), Optimal Functional Forms for Estimation of Missing Precipitation Records, *Journal of Hydrology*, 2009, 374, 106-115.

Teegavarapu, R.S.V., Meskele, T., and Pathak, C. (2011). Geo-spatial grid-based transformation of precipitation estimate using spatial interpolation methods. *Computers and Geoscience*;doi:10.1016/j.cageo.2011.07.004.

Teegavarapu, R.S.V. (2012). Spatial interpolation using nonlinear mathematical programming models for estimation of missing precipitation records. *Hydrological Sciences*. 57(3), pp383-405.

Teegavarapu, R.S.V. (2014a). Statistical corrections of spatially interpolated missing precipitation data estimates. *Hydrological processes*, 28, 3789-3808.

Teegavarapu, R.S.V. (2014b). Missing precipitation data estimating using optimal proximity metric-based imputation, nearest-neighbor classification and cluster-based interpolation methods. *Hydrological Sciences Journal*, 59(11), 2009-2026.

Teegavarapu, R.S.V., and Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312, 191-206.

Teegavarapu, R. S. V., Meskele, T., and Pathak, C. S. (2012). Geo-spatial grid-based transformations of precipitation estimates using spatial interpolation methods. *Computers and Geosciences,* 40(1), 28-39.

Tong, S. T., Sun, Y., Ranatunga, T., He, J., and Yang, Y. J. (2012). Predicting plausible impacts of sets of climate and land use change scenarios on water resources. *Applied Geography, 32*(2), 477-489.

Tuppad, P., and Srinivasan, R. (2008). Bosque River environmental infrastructure improvement plan: Phase II BMP modeling report. College Station, Tex: Texas A&M University, *Texas AgriLife Research*.

Tu, J., and Xia, Z.-G. (2006). Assessing the impact of land use changes on water quality across multiple spatial scales in eastern Massachusetts. *Middle States Geographer, 39*, 34-42.

Turkes, M., and Tatli, H. (2011). Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929 – 2007. *Journal of Climatology*, 31, 2055-2067.

US.EPA. (1995). Level III Ecoregions of the Continental United States. *National Health and Environmental Effects Research Laboratory*. EPA.

US.EPA. (1999). Preliminary data summary of urban storm water best management practices. *United States Environmental Protection Agency, Office of Water.< http://www. epa. gov/waterscience/guide/stormwater/# nsbd*.

US.EPA. (2000). Our built and natural environments: A technical review of the interactions between land use, transportation, and environmental quality.

US.EPA. (2005a). Greenhouse gas mitigation potential in US forestry and agriculture. *Environmental Protection Agency. EPA*.

US.EPA. (2005b). National Management Measures to Control Nonpoint Source Pollution from Urban Areas.

US.EPA. (2009). National Water Quality Inventory: Report to Congress: 2004 Reporting Cycle. *US Environmental Protection Agency, Office of Water, Washington DC*.

Van Delden, H., Seppelt, R., White, R., and Jakeman, A. J. (2011). A methodology for the design and development of integrated models for policy support. *Environmental Modelling & Software, 26*(3), 266-279.

Verburg, P., Veldkamp, A., and Fresco, L. (1999). Simulation of changes in the spatial pattern of land use in China. *Applied Geography, 19*(3), 211-233.

Verburg, P. H., Eickhout, B., and van Meijl, H. (2008). A multi-scale, multi-model approach for analyzing the future dynamics of European land use. *The Annals of Regional Science, 42*(1), 57-77.

Vieux, B.E. (2001). Distributed Hydrologic Modeling using GIS. *Water Science and Technology Library. Kluwer Academic Publishers*.

Wang, D., Smith, M. B., Zhang, Z., Reed, S. M., and Koren, V. I. (2000). Statistical comparision of mean real precipitation estimates from WSR-88d, operational and historical gage networks. *Proc., 15th Conf. on Hydrology*, American Meteorological Society, Boston.

Wang, G., Wu, B., and Li, T. (2007). Digital yellow river model. *Journal of Hydro-environment Research, 1*(1), 1-11.

Wang, K., and Shen, Z. (2012). A GPU-based parallel genetic algorithm for generating daily activity plans. *IEEE Transactions on Intelligent Transportation Systems, 13*(3), 1474-1480.

Watson, D.F., and Philip, G.M. (1985). A refinement of inverse distance weighted interpolation. *Geo-Processing*, 2, 315-327.

Wei, Y.-L., Bao, L.-J., Wu, C.-C., He, Z.-C., and Zeng, E. Y. (2015). Assessing the effects of urbanization on the environment with soil legacy and current-use insecticides: A case study in the Pearl River Delta, China. *Science of the Total Environment, 514*, 409-417.

Westerberg, I., Walther, A., Guerrero, J.L., Coello, Z., Halldin, S., Xu, C.Y., Chen, D., and Lundin, L.C. (2010). Precipitation data in a mountainous catchemnt in Honduras: quality assessment and spatiotemporal characteristics. *Theor Appl Climatol*, Vol 101, 381-396.

Wilson, C. O., and Weng, Q. (2011). Simulating the impacts of future land use and climate changes on surface water quality in the Des Plaines River watershed, Chicago Metropolitan Statistical Area, Illinois. *Science of the Total Environment, 409*(20), 4387-4405.

Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences*, Academic Press.

Wood, E. F. (1995). Heterogeneity and scaling land-atmospheric water and energy fluxes in climate systems. *Space and time scale variability and interdependencies in hydrological processes*, R. A. Feddes, ed., Cambridge University Press, West Nyack, NY, 3–19.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic change*, *62*(1-3), 189-216.

Wu, K., and Xu, Y. J. (2006). EVALUATION OF THE APPLICABILITY OF THE SWAT MODEL FOR COASTAL WATERSHEDS IN SOUTHEASTERN LOUISIANA1. *Journal of the American Water Resources Association*, *42*(5), 1247.

Wu, Y., Li, T., Sun, L., and Chen, J. (2013). Parallelization of a hydrological model using the message passing interface. *Environmental Modelling & Software*, *43*, 124-132.

Xia, Y., Fabian, P., Stohl, A., and Winterhalter, M. (1999). Forest climatology: estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology*. 96, pp 131-144.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., and Livneh, B. (2012a). Continental- scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., and Meng, J. (2012b). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., and Wood, E. F. (2014). Evaluation of multi-model simulated soil moisture in NLDAS-2. *Journal of Hydrology*, *512*, 107-125.

Xie, H., and Lian, Y. (2013). Uncertainty-based evaluation and comparison of SWAT and HSPF applications to the Illinois River Basin. *Journal of Hydrology*, *481*, 119-131.

Xu, C.-y., Widén, E., and Halldin, S. (2005). Modeling hydrological consequences of climate change—progress and challenges. *Advances in Atmospheric Sciences*, *22*(6), 789-797.

Xu, Z., Godrej, A. N., and Grizzard, T. J. (2007). The hydrological calibration and validation of a complexly-linked watershed–reservoir model for the Occoquan watershed, Virginia. *Journal of Hydrology, 345*(3), 167-183.

Yalew, S., van Griensven, A., Ray, N., Kokoszkiewicz, L., and Betrie, G. D. (2013). Distributed computation of large scale SWAT models on the Grid. *Environmental Modelling & Software, 41*, 223-230.

Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R. C., Sadeghi, A., and Arnold, J. G. (2013). Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environmental Modelling & Software, 46*, 208-218.

Zhang, X., Srinivasan, R., Zhao, K., and Liew, M. V. (2009). Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrological Processes, 23*(3), 430-441.

Zhang, Y., Reed, S., and Kitzmiller, D. (2011). Effects of retrospective gauge-based readjustment of multisensor precipitation estimates on hydrologic simulations. *Journal of Hydrometeorology*, 12(3), 429–433.

Zhao, G., Bryan, B. A., King, D., Luo, Z., Wang, E., Bende-Michl, U., and Yu, Q. (2013). Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. *Environmental Modelling & Software, 41*, 231-238.

Zhao, P., Xia, B., Hu, Y., and Yang, Y. (2013). A spatial multi-criteria planning scheme for evaluating riparian buffer restoration priorities. *Ecological engineering, 54*, 155-164.

Zhou, H., and Gao, C. (2011). Assessing the risk of phosphorus loss and identifying critical source areas in the Chaohu Lake watershed, China. *Environmental management, 48*(5), 1033-1043.

Zhu, X., and Dale, A. P. (2001). JavaAHP: a web-based decision analysis tool for natural resource and environmental management. *Environmental Modelling & Software, 16*(3), 251-262.

Appendix A: Installation and configuration of DRBL

1. Before the DRBL downloading, list file is modified by adding the repositories.

   - editing the file from terminal: "*sudo nano /etc/apt/sources.list*"

   - adding the below two lines to the end of the file from the file edition program and save the file:

     "*deb http://free.nchc.org.tw/ubuntu maverick main restricted universe multiverse*",

     "*deb http://free.nchc.org.tw/drbl-core drbl stable*"

2. Download DRBL key

   - "*sudo wget http://drb1.nchc.org.tw/GPG-KEY-DRBL*"

   - "*sudo apt-key add GPG-KEY-DRBL*"

3. Update Advanced Packing Tool (APT) for existing software packages and install DRBL packages.

   - "*sudo apt-get update*"

   - "*sudo apt-get install drbl*"

4. After DRBL installing, we need to set it up to suit the SSC kernel

   - setup the DRBL setting: "*sudo /usr/sbin/drblsrv –i*"

5. From first screen of the DRBL installing (**Figure A.1**), Select "No" for the network installation boot images.

```
**************************************************.
Hint! When a "yes or no" option is available, the default value is uppercase
. E.g. (y/N) the default is "N", so when you press "Enter" without typing "Y
 or N" it will be as if you typed "N" and then "Enter". If you are not sure
which option to choose just press "Enter" key.
**************************************************.
**************************************************.
Installing DRBL for Debian Linux...
**************************************************.
Do you want to install the network installation boot images so that you can
let the client computer install some GNU/Linux distributions (Debian, Ubuntu
, RedHat Linux, Fedora Core, Mandriva, CentOS and OpenSuSE...) via a network
 connection?  !!NOTE!! This will download a lot of files (Typically > 100 MB
) so it might take a few minutes. If the client computer has a hard drive th
at you may install GNU/Linux onto, put a Y here. If you answer "no" here, yo
u can run "drbl-netinstall" to install them later.
[y/N] N
```

Figure A.1. The first step of DRBL installation.

6. From the **Figure A.2**, select “NO” for the displaying option.

```
**************************************************.
This GNU/Linux distribution uses one kernel to support SMP and non-SMP arch.
**************************************************.
Do you want to use the serial console output on the client computer(s)?
If you do NOT know what to pick, say "N" here, otherwise the client computer
(s) may show NOTHING on the screen!
[y/N] N
```

Figure A.2. Displaying option for slave nodes.

7. For the option of upgrading operating system (**Figure A.3**), select “Yes” if you need to upgrade.

Updating data will be automatically downloaded via the internet. Otherwise, select “NO”

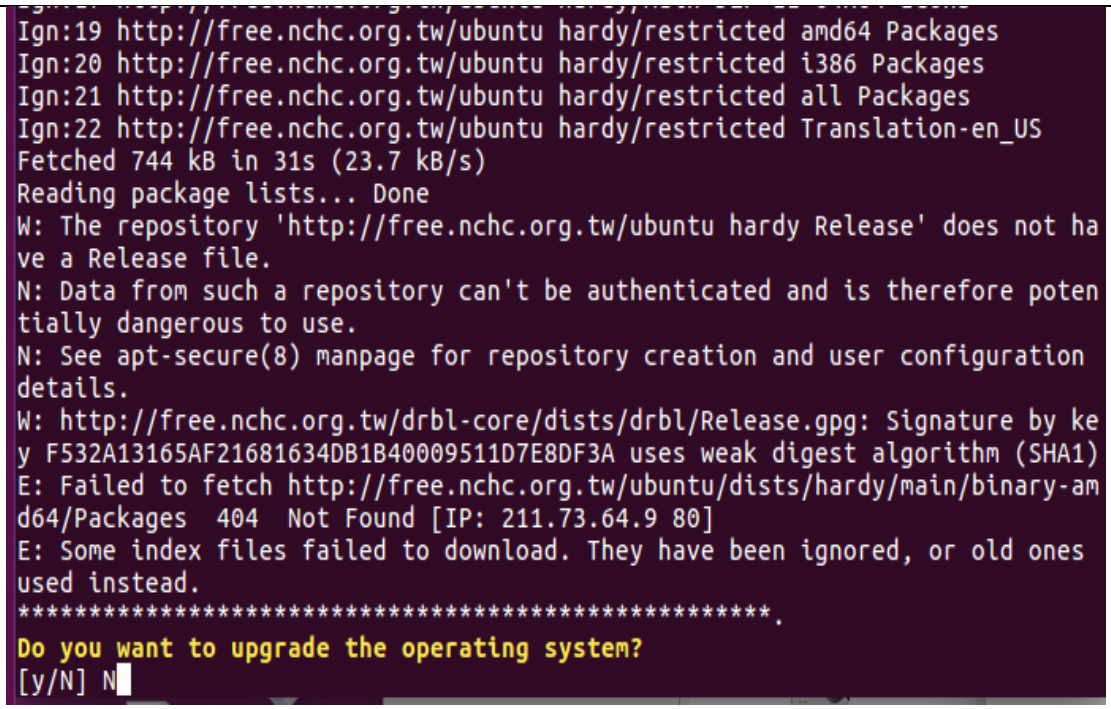


Figure A.3. Upgrading the operating system.

8. Select 1 for the generic kernel from DRBL server from **Figure A.4** (setup completion).

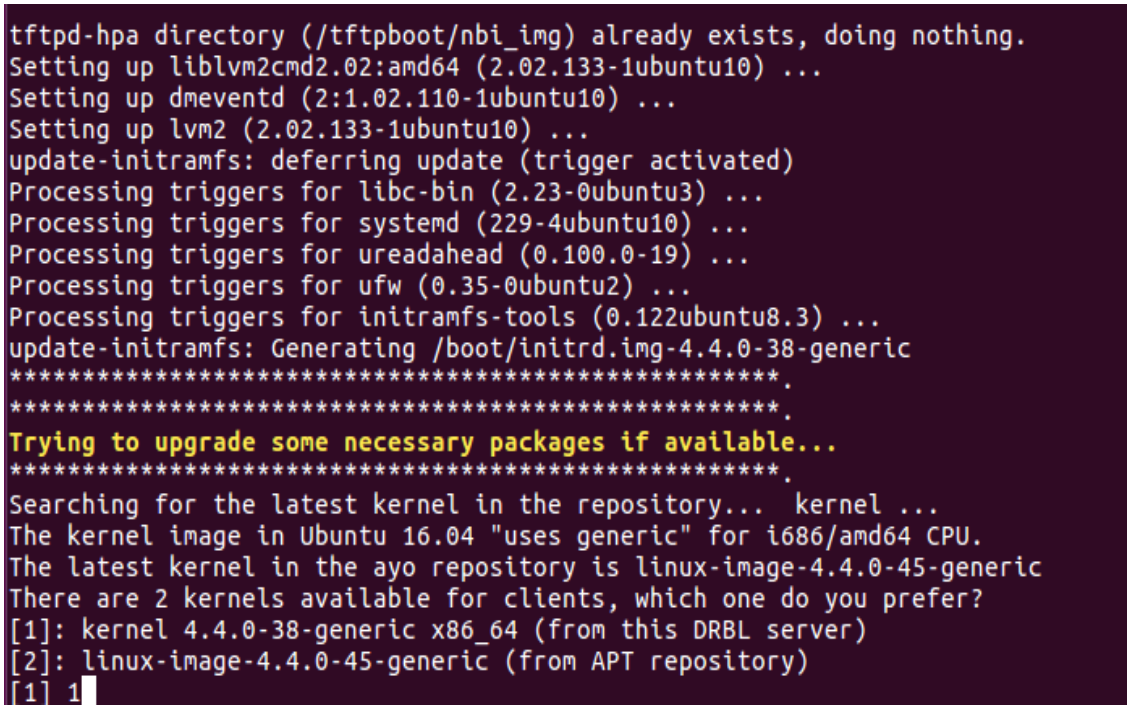


Figure A.4. Two options of the kernel selection.

9. Do configuration of network setting to a manual IP of 192.168.1.1 from Network Connections

- From IPv4 settings in LAN edition icon, change the method to Manual option and then adding the address (192.168.1.1) and subnet (255.255.255.0).

10. Do configuration the master node for the slave node connected.

- "*sudo /usr/sbin /drblpush –i*"

11. Assign DNS, NIS/YP, and prefix domain name (**Figure A.5**).

- DNS is the domain name system. DNS enables the master server to use their assigned name instead of internet protocol network addresses. In this study, we set DNS as jungjin.org.

- NIS/YP is network information service or yellow pages. It indicates the master node name. In this study, we set the master name as "*jungjinlinux*".

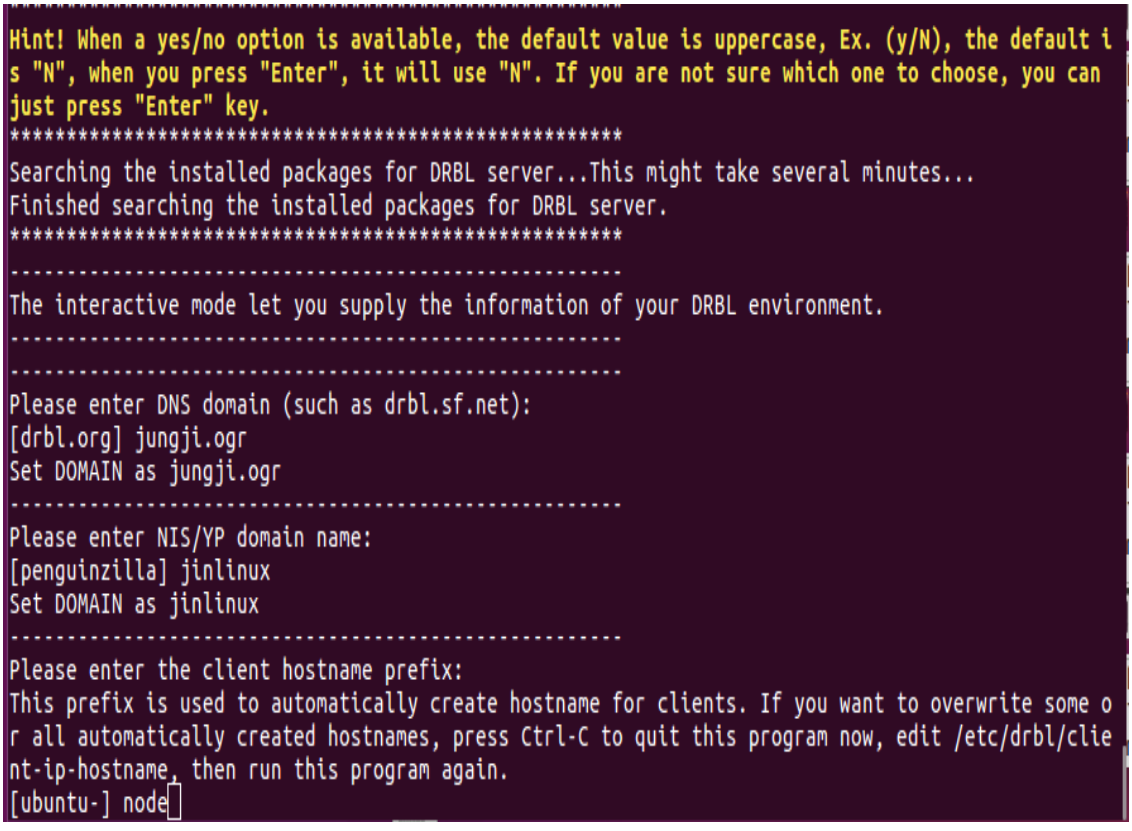


Figure A.5. Assigning the names of DMS domain, NIS, and slave nodes.

12. Select the LAN card which is connected to the slave nodes. In this study, wlan0 was the internet connection and eth0 was the network connection between master and slave nodes. For the collection of the MAC of nodes, entering the "q". And then, give the initial number of the slave node address should be started associated with the number of slave node we have. In this study, we gave the initial value as 1. Slave nodes had IP's from 192.168.1.1. Finally, **Figure A.6** shows the number of slave nodes based on we have. In this study, we had 6 slave nodes and their IPs were from 192.168.1.1 to 19.1.16.1.6.

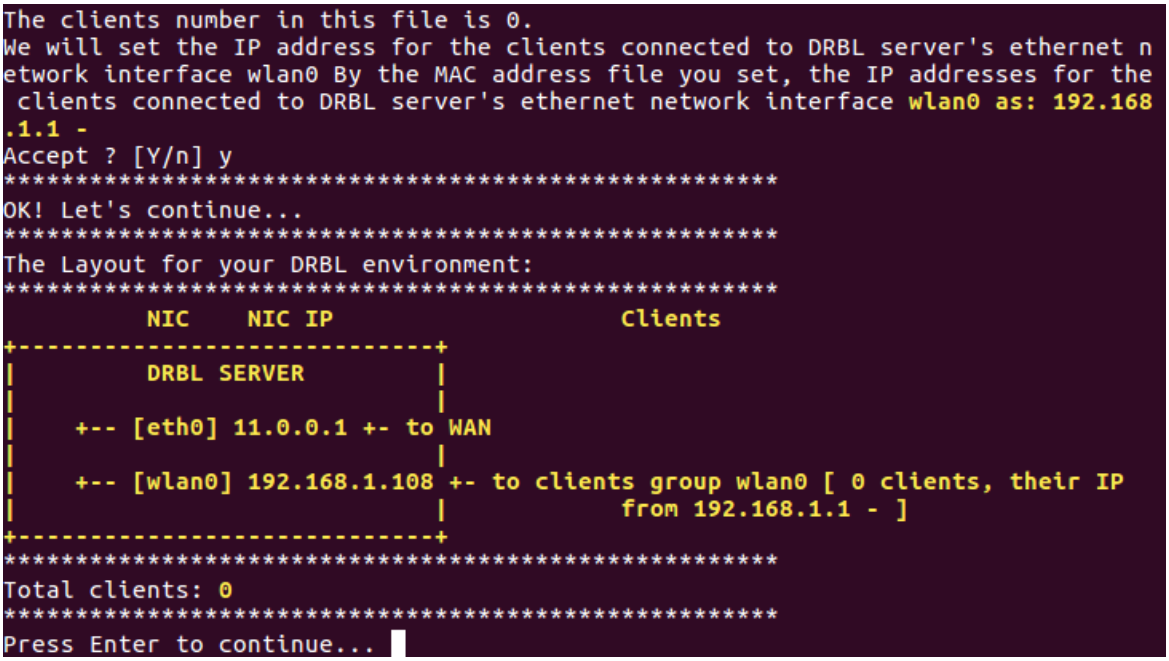


Figure A.6. The results of internet connection (eth0) and network connection (wlan0) between master and slave nodes.

13. Next step, select "1" as DRBL Single System Image", and select "2" for "not use clonezilla" in the following steps because we wanted to build as Beowulf cluster framework (**Figure A.7**).

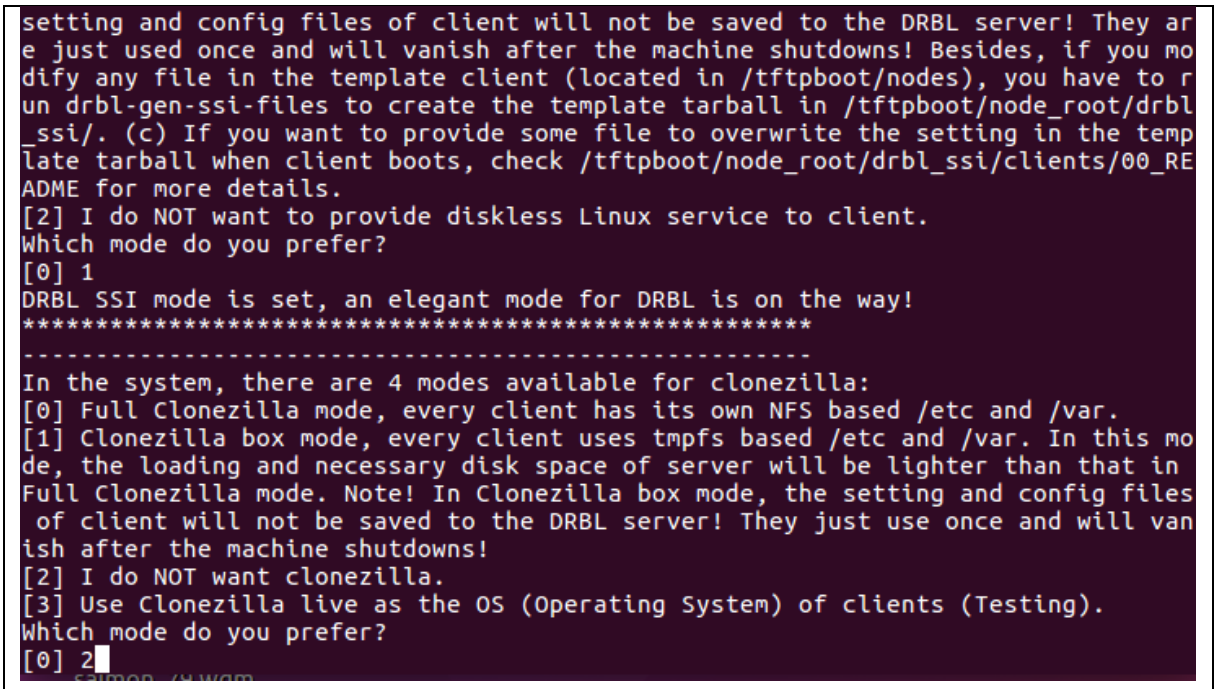


Figure A.7. Selecting the DRBL single system image and not clonezilla mode for the setting up the Beowulf cluster.

14. For the slave nodes mode after boot, select Text mode since the graphical mode is a waste of resources to use. For the boot prompts and passwords of slave nodes, select “NO” because they do not have keyboards (**Figure A.8**).

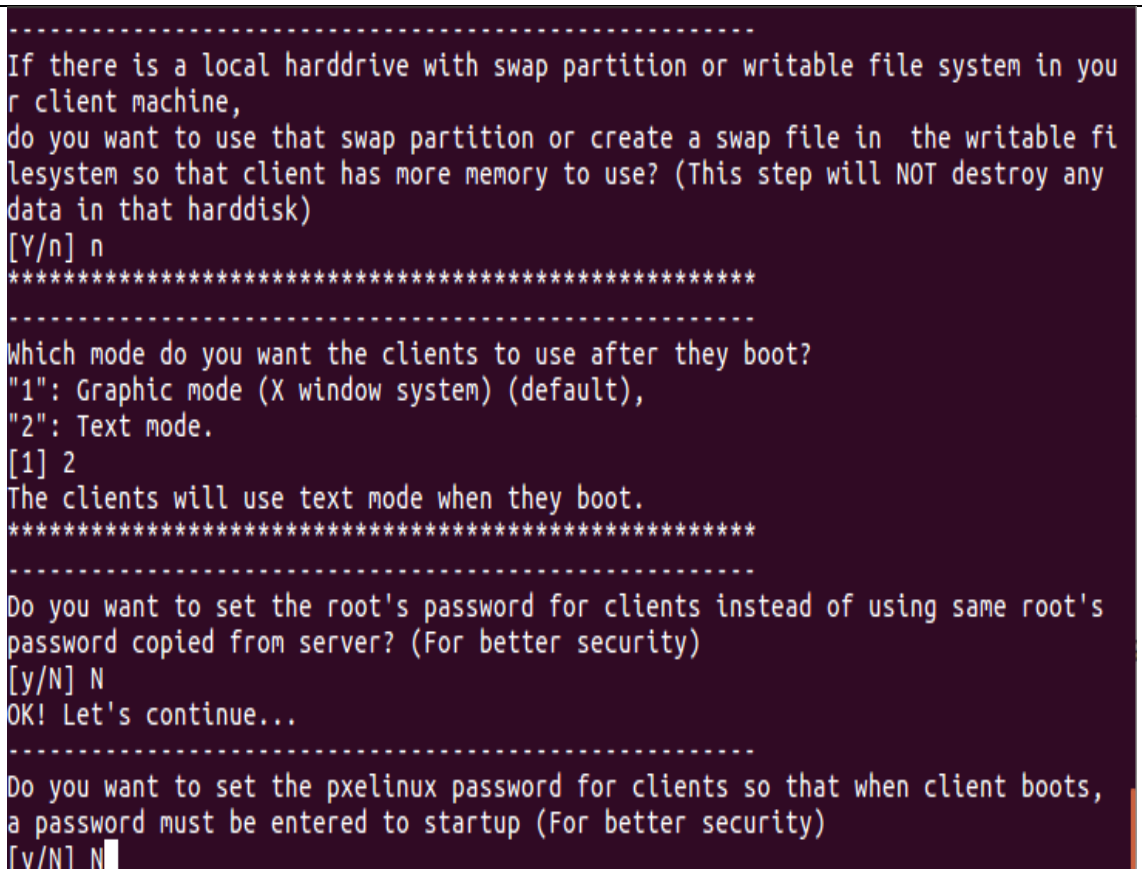


Figure A.8. Boot mode selection and root password setting.

15. The last settings follow the default selection. This is the final setting for DRBL custom to the node environment.

16. Controlling DRBL from DRBL Control Server.

- from the terminal, execute the DRBL program "*/opt/drbl/sbin/dcs*".

- Select "All select all the clients" for setting the model for all slave nodes as **Figure A.9**.

- Select "remote-linux-txt Client remote Linux, textmode, powerful client" for the final step as **Figure A.10**. From this mode, we can switch the slave node setting anytime.

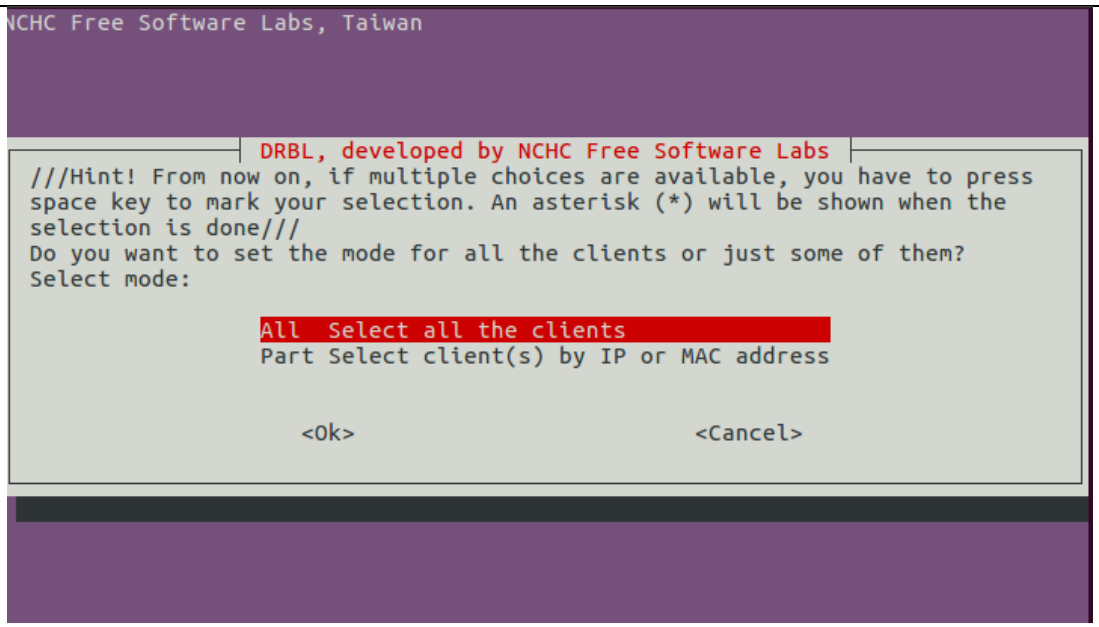


Figure A.9. The mode setting for all slave nodes.

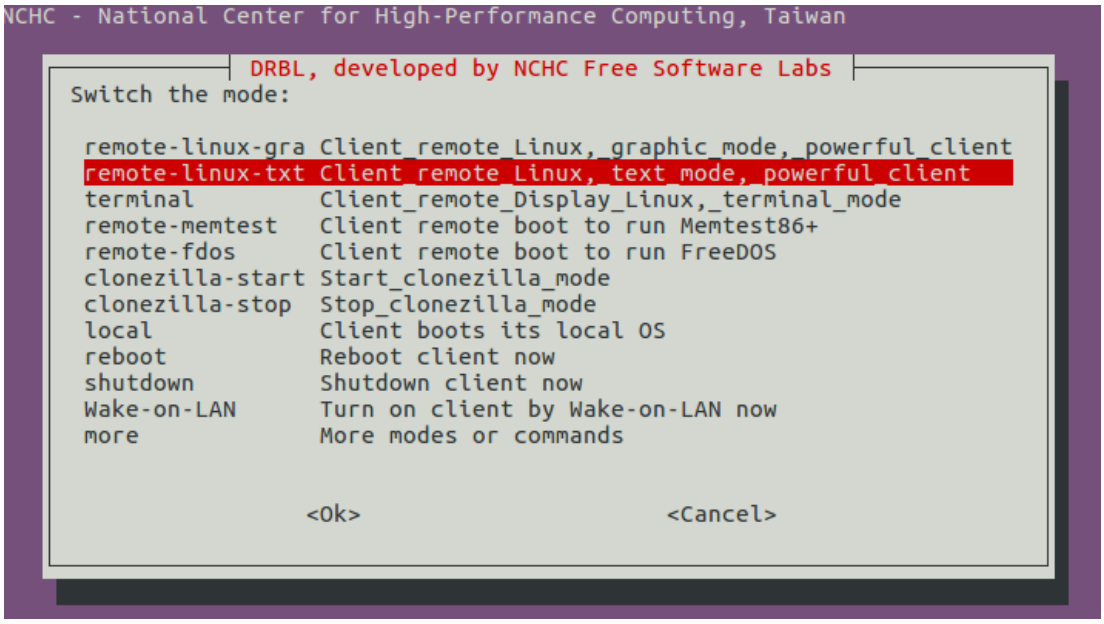


Figure A.10. Final setting for the mode.

17. After the setting, diskless Linux Cluster is completed.