

APPLICATIONS OF INTERPRETABLE
DATA-DRIVEN APPROACHES IN CLIMATE,
HYDROLOGY, AND WATER RESOURCE
MANAGEMENT

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Xiao Peng

December 2021

© 2021 Xiao Peng
ALL RIGHTS RESERVED

APPLICATIONS OF INTERPRETABLE DATA-DRIVEN APPROACHES IN
CLIMATE, HYDROLOGY, AND WATER RESOURCE MANAGEMENT

Xiao Peng, Ph.D.

Cornell University 2021

The recent advances in sensing technology and machine learning have offered new opportunities in hydro-climate research with tremendous data and highly complex, nonlinear models. However, a model tends to lose interpretability when it becomes complex, which can degrade its transferability under out-of-sample conditions. Therefore, the tradeoff between model complexity (accuracy) and interpretability must be considered. In this dissertation, I implement some interpretable data-driven approaches through a typical chain of uncertainty propagation in three major sectors of climate, hydrology, and water resource management. My research serves as an early effort to quantify uncertainty from two major sources, climate forcing and hydrological responses, and to incorporate them into water resource management.

I first examine the feasibility of a regularized regression model in uncovering teleconnections between regional precipitation and large-scale climates from both physical and statistical perspectives. Next, we explore the potential of a Bayesian estimation method as a cheap surveillance tool for monitoring dynamics of rainfall-runoff responses for small watersheds. Finally, we adopt the evolutionary multi-objective direct policy search framework to quantitatively compare how forcing data of different resolutions can alter water system optimization. Based on these projects, we figure that 1) data-driven approaches can guide exploration of physical knowledge and 2) computational costs can be

reduced by using physical constraints.

BIOGRAPHICAL SKETCH

In 2016, Xiao Peng started pursuing a PhD degree in School of Civil and Environmental Engineering at Cornell University under the supervision of Dr. John Albertson. Xiao has a particular interest in analyzing and interpreting data and his research covers investigation of teleconnections between large-scale climate variables, development of a surveillance model for detecting changes in regional hydrological responses, and optimization of inter-connected reservoir systems driven by forcing data of different resolutions. Before coming to Cornell, Xiao received a B.Eng degree from Department of Hydraulic Engineering, Tsinghua University in 2015.

This work is dedicated to my parents.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. John Albertson for his kind and continuous support throughout the PhD program. And I appreciate Dr. Scott Steinschneider and Dr. Patrick Reed for serving in my committee and providing technical help in statistical modeling and system optimization, respectively. Last but not least, I want to thank all my EFMH colleagues and Cornell friends for fostering a friendly community.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1 A General Introduction	1
2 Seasonal Predictability of East Africa Short Rain	9
2.1 Introduction	9
2.2 Data	14
2.3 Methods	16
2.4 Results & Discussions	19
2.4.1 Model skill with shuffled data using spatially-averaged SPI	19
2.4.2 Spatial variability of predictive skill	20
2.4.3 Comparison between climate composites and regression predictor patterns	21
2.4.4 The lead-lag analysis between large-scale climate indices .	30
2.5 Conclusions	32
3 Long-term Forecast of TRHR Wet-season Precipitation	35
3.1 Introduction	35
3.2 Data	40
3.3 Methods	43
3.3.1 The regularized regression	43
3.3.2 Other regression models	45
3.3.3 The correlation analysis	46
3.4 Results & Discussions	47
3.4.1 Comparison of regression models	47
3.4.2 Source of high model skill	52
3.4.3 An alternate multinomial regression model	57
3.5 Conclusions	62
4 Bayesian Surveillance for Rainfall-Runoff Responses	64
4.1 Introduction	64
4.2 Data	67
4.3 Methods	70
4.3.1 Synthetic Data Generator	70
4.3.2 Bayesian Estimation Model	71
4.3.3 Instantaneous Unit Hydrograph Model	74
4.4 Results & Discussions	76

4.4.1	Model Validation	76
4.4.2	Model Application	81
4.5	Conclusions	86
5	Impacts of Information Resolutions on Reservoir Operations	88
5.1	Introduction	88
5.2	Case Study Description	93
5.2.1	Upper Yellow River Multi-Reservoir System	93
5.2.2	Simplified Water Budget Model	95
5.2.3	Synthetic Streamflow Generator	97
5.3	Methods	98
5.3.1	Formulation of Objectives	100
5.3.2	Formulation of DPS policies	102
5.3.3	Many-Objective Optimization	104
5.3.4	Robustness Analysis	105
5.3.5	Time-varying Sensitivity Analysis	106
5.4	Results & Discussions	107
5.4.1	Policy Comparison & Validation	107
5.4.2	Robustness Analysis	113
5.4.3	Time-varying Sensitivity Analysis	117
5.5	Conclusions	120
6	Conclusions	122
A	Supplementary Information for Chapter 2	125
B	Supplementary Information for Chapter 3	127
C	Supplementary Information for Chapter 5	131

LIST OF TABLES

3.1	Correlations between L and prediction skill at varying lead times for different regression models	53
4.1	The transition matrix estimated using historical rainfall events during the selected summer periods for the Fall Creek watershed	70
4.2	Ranges of IUH parameter values [Θ_{min} , Θ_{max}]	76
5.1	Time series of pre-determined water demands $W(t)$ at Lanzhou based on chang2014optimized,bai2015synergistic.	102
5.2	Pearson's correlation coefficients between the historical operations and operations determined by the representative DPS policies (for 324 samples)	113
5.3	Two sets of baseline performance used in computing S scores in the robustness analysis	114

LIST OF FIGURES

1.1	Frequency as measured by interest over time of the term ‘data-driven’ being searched on google worldwide over Jan 2011 to Aug 2021. Interest over time is a rescaled metric of popularity over time. A value of 100 is the peak popularity while a value of 50 indicates the term is half as popular. Data source: Google Trends (https://www.google.com/trends).	2
2.1	The spatial-averaged EASR SPI series from 1981 to 2017	15
2.2	Model skill, shown here as a Spearman correlation between model predictions and observations, as a function of lead time for the testing period using shuffled spatially-averaged EASR SPI. The model predictors are anomalies of: (a) SST, (b) 850 hPa HGT, (c) 500 hPa OMEGA, (d) 850 hPa u wind, (e) 200 hPa u wind, (f) 850 hPa v wind, and (g) 200 hPa v wind. The 25 th – 75 th ranges are plotted in green lines; the median CCs are plotted in red diamonds. The one-tailed $p = 0.1$ significance level is plotted in the black dashed line.	20
2.3	Distribution of regression model skill at each EASR SPI grid cell and for different lead times. The model is fit to SST data between 1981-2007, and correlation coefficients are shown for the testing period (2008-2017). The 25 th – 75 th ranges are plotted in green lines; the median CCs are plotted in red diamonds. The one-tailed $p = 0.1$ significance level is plotted in the black dashed line.	22
2.4	Maps of the best intra-season prediction skill for different lead time ranges. The regression analysis is conducted for periods of: (a) S1 (lead times = 0-3 months), (b) S2 (lead times = 4-6 months), (c) S3 (lead times = 7-9 months), and (d) S4 (lead times = 10-12 months), and the best intra-season skill is defined as the highest CC score for the testing period within the corresponding season. The grid cells with significant prediction skill (one-tailed $p = 0.1$) are marked with red dots.	23
2.5	Coefficient maps for anomalies of (a-d) SST, (e-h) HGT, (i-l) OMEGA during periods of S1 (a, e, i), S2 (b, f, j), S3 (c, g, k), and S4 (d, h, l). Coefficients values are shown in color.	24

2.6	Coefficient maps for anomalies of (a–d) 850 hPa wind and (e–h) 200 hPa wind during periods of (a, e) S1, (b, f) S2, (c, g) S3, and (d, h) S4. Coefficients of the u winds are visualized using zonal arrows, with an eastward arrow denoting a positive coefficient; coefficients of the v winds are visualized using meridional arrows, with a northward arrow denoting a positive coefficient. Amplitudes of the coefficients are represented using the lengths of the arrows. For each atmosphere level (i.e., 850 hPa and 200 hPa), the u and v wind coefficient maps are combined by adding the two vector fields.	25
2.7	Composite maps for the difference between historical extreme wet and dry events for SST (shading), 850 hPa HGT (contour) and 850 hPa winds (arrows) for (a) S1, (b) S2, (c) S3, and (d) S4. Results are only shown if significant when compared to a bootstrapped 90% confidence interval.	27
2.8	Composite maps for difference the between historical extreme wet and dry events for 500 hPa OMEGA (shading), 850 hPa HGT (contour) and 200 hPa winds (arrows) for (a) S1, (b) S2, (c) S3, and (d) S4. Results are only shown if significant when compared to a bootstrapped 90% confidence interval.	28
2.9	Regions used to define large-scale climate indices used in the lead-lag analysis, including the long-lead SST dipole index (SL_p and SL_n), the long-lead height dipole index (HL_p and HL_n), the long-lead meridional wind index (VL), and the short-lead SST dipole index (SS_p and SS_n).	31
2.10	Correlation coefficients (Pearson’s) at varying lead times prior to the short rain season between (a) SL and SS, (b) SL and HL, (c) SL and VL, and (d) VL and SS. The significant correlations are marked with black circles (one-tailed $p = 0.001$). A ‘(-1)’ in the axis label indicates the month from the previous calendar year.	32
3.1	(a) The study region of TRHR (black box) as plotted in an elevation map based on ETOPO-5 [1] and (b) time series of the spatially averaged precipitation from CHIRPS (red) and CMA (blue) after standardization.	42
3.2	A schematic of the proposed the regression model coupling max pooling and elastic net	44
3.3	Comparison of the predictive skill in the testing period from the elastic net regression models with no pooling (dark blue), max pooling (light blue), median pooling (green) and minimum pooling (yellow). The p-value = 0.1 significance level is plotted in the black dashed line.	49
3.4	Predictive skill for the testing set as a function of lead time reported in NSE scores	50

3.5	Prediction skill for the testing set of 2001-2019 measured by correlation coefficients as a function of lead times for elastic net (red), OLS (blue), EOF (green), and CCA (yellow) regression models. The p-value = 0.1 significance level is plotted in black dashed lines	51
3.6	Comparison between predictive skill from the elastic net model (bars) and L_s (a blue diamond line) at varying lead times. The statistical significance levels are not shown here as we are comparing correlation coefficients calculated using samples of different lengths (20 for the predictive skill and 39 for the L_s).	53
3.7	Correlation maps between the TRHR precipitation and the SSTs over the training period of 1981-2000 (top) and the testing period of 2001-2019 (bottom) at the lead time of 14 months	55
3.8	Correlation maps between the TRHR precipitation and the SSTs over the training period of 1981-2000 (top) and the testing period of 2001-2019 (bottom) at the lead time of 22 months	56
3.9	Accuracy of predicting the wet-dry state of TRHR precipitation at varying lead times as reported in S scores. The baseline skill using a null model of totally random guess is plotted in a black dashed line. The bootstrapped p-value = 0.1 significance level is plotted in a blue dashed line (10,000 samples for each lead time).	58
3.10	Maps of regression coefficients from the elastic net models using true-amplitude (top) and binary (bottom) predictand precipitation at the lead time of 14 months	59
3.11	Maps of regression coefficients from the elastic net models using true-amplitude (top) and binary (bottom) predictand precipitation at the lead time of 22 months	60
3.12	Lagged correlations between the TRHR precipitation and the SST indices. The p-value = 0.01 significance level is plotted in the black dashed line.	61
4.1	Model skill measured by relative errors as a function of SNR for (a) λ , (b) k , and (c) θ . Performance of the initial grid searching and the Bayesian estimation model are plotted in blue and red lines, respectively. At each SNR , the synthetic experiment is repeated for 50 times. The 25 th –75 th percentile ranges are plotted as shaded areas and the median model skill are plotted in diamond lines. The reference SNR based on observational data from the Fall Creek watershed is indicated by a vertical black dashed line. And observational data from the Fall Creek watershed have SNR values no smaller than this reference level for all selected summer periods as indicated by the leftward arrow.	78

4.2	Empirical cumulative distribution functions of model skill measured by (a) Pearson’s correlation coefficients (CC) and (b) Nash-Sutcliffe Efficiency (NSE) scores for runoff estimations. Performance of models fitted using MLE and Bayesian parameter estimations are plotted in the red and blue curves, respectively. . . .	79
4.3	Comparison between the Bayesian (x axis) and MLE (y axis) parameter estimations for (a) λ , (b) k , and (c) θ . The 1:1 reference lines are plotted in black dashed lines and the coefficients of determination (R^2) for 65 summer periods are reported in the subplot titles.	80
4.4	A diagram of two types of IUHs: the ‘diffusion’ type (red) and the ‘advection’ type (blue).	82
4.5	Time series of Bayesian estimations of (a) λ , (b) k , and (c) θ . The year of 1943 when Cornell moved its rainfall station is indicated by a vertical red dashed line.	83
4.6	Time series of (a) total summer rainfall, (b) total summer runoff, and (c) <i>IDR</i> . The year of 1943 when Cornell moved its rainfall station is indicated by a vertical red dashed line.	84
4.7	(a) A map of median <i>IDRs</i> for watersheds with ‘good’ NSE scores where sizes of circles are proportional to watershed length scales. (b) A scatter plot between watershed length scales and median <i>IDRs</i> and its fitted line (black).	86
5.1	(a) A map of the upper Yellow River basin [2]. The upper Yellow River multi-reservoir system consists of 12 reservoirs (Index 2-13) with the two largest reservoirs being Longyangxia reservoir (Index 2) and Liujiaxia reservoir (Index 13). (b) A schematic of the simplified reservoir system. Controllable releases are only assumed for LYX (s^y) and LJX (s^j) and all the 10 smaller reservoirs in between are lumped into one reservoir s^l	94
5.2	A diagram of the analytical framework for comparing expectation performance and policy robustness in managing the Upper Yellow River reservoir system with true-amplitude and binary flow data.	99
5.3	Parallel axis plot of expectation objectives from ‘true’ (red) and ‘binary’ (blue) policies for four objectives: total hydropower production (totE), minimum 10-day hydropower production (minE), water deficit (WD), and end-of-horizon water storage (SP). Objective ranges are shown and a direction of increasing preference is indicated by a downward arrow. The shading corresponds to the expectation objective of total hydropower production. A darker value means greater total hydropower production.	107

5.4	Parallel axis plot of expectation objectives from 'true' (red) and 'binary' (blue) policies tested using historical flow data over 2001-2009. Maximum and minimum objective values are shown and a direction of increasing preference is indicated by a downward arrow. Performance of the actual historical operations is plotted in a black line and the shading corresponds to mean square distance between the historical performance and DPS policy performance. The darker value means a DPS policy has more similar performance compared to historical operations. The 'true' and 'binary' policies that have the closest performance compared to the historical operations are defined as the representative policies and are plotted in dark red and blue lines, respectively.	109
5.5	Comparison between times series of LYX (top) and LJX (bottom) reservoir outflow determined using historical operations (black) and representative policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.	111
5.6	Comparison between times series of LYX (top) and LJX (bottom) reservoir water storage determined using historical operations (black) and representative policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.	112
5.7	Robustness reported in S scores under changing initial conditions of water storage for the 'true' (a) and 'binary' (b) representative policies and the difference ('true' - 'binary') (c). The strict baseline performance based on the historical operations is used. The original condition under which DPS policies are trained is marked by a black cross.	114
5.8	Robustness reported in S scores under changing flow distributions for the 'true' (a,d) and 'binary' (b,e) representative policies and the difference ('true' - 'binary') (c,f). Both the strict (top, or a,b,c) and relaxed (bottom, or d,e,f) baseline performance are used. The original condition under which DPS policies are trained is marked by a black cross.	116
5.9	Sensitivities of prescribed LYX (left) and LJX (right) release decisions determined by the 'true' (top) and 'binary' (bottom) policies. The total variances of release decisions are decomposed to contributions of different sources (S^y in red, S^j in pink, Q^{inf} in orange, Q^{lat} in magenta, and interactions in blue).	118
5.10	Variances of (a) LYX and (b) LJX release decisions estimated using simulation samples across SOWs. Variances using the 'true' (red) and 'binary' policies are compared and reduced uncertainties are observed in release decisions when the true-amplitude flow data is used.	119

A.1	Model skill at different lead times for the testing period (2008-2017) using the SSTAs to predict rainfall of (a) Oct-Nov (blue) and Oct-Nov-Dec (red) and (b) Oct (blue), Nov (red), and Dec (yellow); (c) Scatter plot of OND rainfall against ON rainfall.	125
A.2	Maps of the best intra-season prediction skill for different lead time ranges using the GPCP precipitation product. The regression analysis is conducted for lead times ranging from 0-12 months and the best intra-season skill is defined as the highest CC score for the testing period within the corresponding season. The grid cells with significant prediction skill (one-tailed $p = 0.1$) are marked with black circles.	126
B.1	Difference of averaged SSTAs after and before 2000. The anomalies are based on the whole study period of 1979-2019. Assuming SSTAs are independent normal variables, significant differences (p -value = 0.05 one-tailed) are marked with black dots. 38.1% (511 of 1343 grids) show significantly positive trends while only 2.8% (38 of 1343 grids) show significantly negative trends.	127
B.2	Correlation coefficients between the observed and estimated TRHR wet-season precipitation over the testing period of 2001-2019 using varying λ s are shown in a box plot. A geometric sequence of 10 λ s in the range of [17, 53] are tested here for developing the elastic net regression models while λ is set to 28 (the MinMSE λ from a 5-fold CV) in the paper. Major analyses in the paper are focused on the two long-term predictive spikes at lead times of 14 and 22 months, and the results show that these two spikes are not sensitive over a rather broad range of λ s. The p -value = 0.1 significance level is plotted in a black dashed line.	128
B.3	Box plots of testing period model skill as measured by correlation coefficients with random splitting (repeated for 25 times). The p -value = 0.1 significance level is plotted in a horizontal black dashed line. Model performance is degraded with random splitting. A major reason is that the regression model can only 'learn' the statistical associations from the extreme precipitation events and the extreme events are relatively uniformly distributed across the training and testing period in the original splitting case (1981-2000 for training and 2001-2019 for testing). Extreme events being concentrated in either training or testing period could lead to significant degradation of model performance. And this effect is further exaggerated when correlation coefficient is used for measuring model skill.	129

B.4	For a lead time of 22 months, (a) scatter plot of observed precipitation anomalies against the model estimations for the training period (red circles) and the testing period (blue diamonds), and (b) comparison between observed precipitation anomalies (red line) and the rescaled estimations (blue line).	130
C.1	A scatter plot between 10-day pre-LYX inflow and total LYX-LJX lateral flow based on observational data over 2001-2009. The best linear fit is plotted in a black dashed line and the 95% confidence interval (CI) linear fits are plotted in red dashed lines.	131
C.2	Rescaled hypervolume as a function of number of function evaluations (NFE) for 50 random seeds	131
C.3	Comparison between empirical functions for estimating pre-dam water head given water storage for LYX (red) and LJX (blue). For the same amount of increased water storage, pre-dam water head is always raised by a larger margin for LJX.	132
C.4	Comparison between times series of total water storage ($R_{lyx} + R_{ljx}$) determined using historic operations (black) and compromise policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.	132
C.5	A schematic of the two-sided effects of altering flow distributions on policy robustness. Increasing probability of extreme flow events will make both drought and flood events more likely and thus change the original distribution of performance (red) to a flatter distribution (blue). Robustness as measured by the S scores (i.e., success ratios) can be improved when compared against a strict baseline performance (vertical black dashed line on the right) and can be degraded when compared against a relaxed baseline performance (vertical black dashed line on left right).	133

CHAPTER 1

A GENERAL INTRODUCTION

‘Without data you’re just a person with an opinion.’ These words of Dr. W. Edwards Deming, the master of continual improvement of quality, have gained dramatically increasing popularity these days, and reflect a prevalent trend of people attaching more importance to data-based facts across many fields. The statement is particularly relevant in science where people want facts rather than opinions. Concurrently, a significant growing interest has been observed in data-driven approaches (Figure 1.1). Generally speaking, a data-driven approach is determined by or dependent on the collection or analysis of data. Applications of data-driven approaches include but are not limited to prediction [3], causal inference [4], decision making [5], visualization [6], etc. Also, data-driven analysis is often used interchangeably with statistical analysis in research since both analyses attempt to uncover patterns by collecting and interpreting data. More recently, ‘data-driven’ is being linked to some intangible aspects as people start creating a ‘data-driven’ culture in companies [7], schools [8], and even societies [9].

Data-driven approaches have a long history in solving water-related problems, of which only a very brief introduction can be offered here. In this dissertation, their applications are categorized into three major sectors including climate, hydrology, and water resource management. In the climate community, statistical techniques have played an important role since the birth of climatology when very little knowledge of physical modeling was available. Some earliest work can date back to 1800s when scientists tried to study climate in a descriptive way by comparing simple statistics of climate variables (e.g., sample

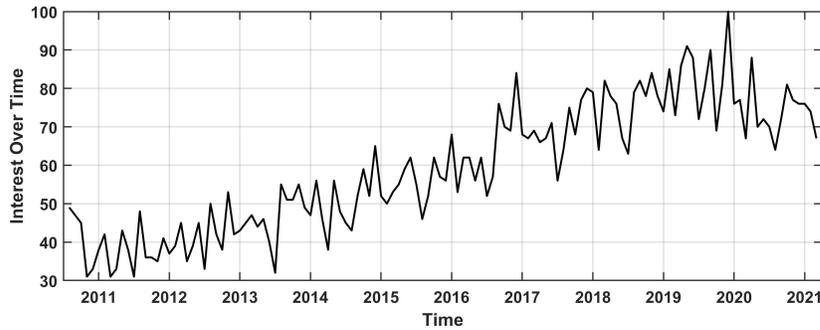


Figure 1.1: Frequency as measured by interest over time of the term 'data-driven' being searched on google worldwide over Jan 2011 to Aug 2021. Interest over time is a rescaled metric of popularity over time. A value of 100 is the peak popularity while a value of 50 indicates the term is half as popular. Data source: Google Trends (<https://www.google.com/trends>).

mean and standard deviation) [10]. Development of climate science was accelerated around 1860s when observations became more reliable [11]. The more readily available data also boosted development of related statistical techniques since it requires more sophisticated methods for information extraction and provides opportunities for statistician to elaborate their innovated approaches [11]. Since then, a diverse of data-driven approaches have been implemented in exploring hidden associations between climate variables and investigating underlying physical processes. Some well-established and widely-used approaches include regression [12], correlation [13], analysis of variance (ANOVA) [14], eigenvalue techniques [15], frequency analysis [16], etc. And for more comprehensive details about statistical classics in climate research, one can refer to Von Storch and Zwiers [10] and Von Storch and Navarra [11].

Application of data-driven approaches has followed a similar path in the field of hydrology. Statistical frequency analysis was first introduced into hydrology in early 1900s by Fuller [17], and was followed by the famous unit

hydrograph model [18]. The early efforts were focused on using linear and empirical relationships between hydrological observations (e.g., precipitation, runoff) due to limited knowledge of physical processes and data scarcity. Later on, thanks to advances of sensing technology, observations became available for variables that were impossible to measure (e.g., surface fluxes) [19], and more refined methods were developed for aid understanding of nonlinear interactions between variables [20, 21]. Even in present days when physically-based modeling is the primary way for causality analysis in hydrology, data-driven approaches provide many essential functions in interpreting model results like uncertainty propagation [22], sensitivity estimation [23], and parameter identification [24].

Research in water resource management has a relatively short history compared to the other two sectors. Multi-unit and multi-purpose water resource systems were studied in a systematic and quantitative way for the first time in 1950s in the Harvard Water Program (HWP) led by a group of engineers and economists [25, 26]. While many studies in climate and hydrology are based on dynamic models governed by physical laws (e.g., general circulation models [27] and distributed hydrology models [28]), such tools are lacking in the community of water resource management. Even though some of the systems to be optimized are based on physical models, they are often treated like black box models and are being replaced by more cost-effective surrogates [29]. And the nature of optimization problems in water resource management makes data-driven techniques like sampling a vital part , especially in managing systems under deep uncertainty [30].

In past few decades, a broad range of water-related problems are becoming

more challenging due to rapidly changing climate and more extreme weather events. The Intergovernmental Panel on Climate Change (IPCC) just released its 6th assessment report (AR6) which suggests that human-induced climate change is affecting many weather and climate extremes across the globe and the scale of recent climate changes is unprecedented [31]. These facts greatly amplify uncertainties in model estimations of climate and hydrology variables [32, 33] and put question marks on empirical models based on an assumption of stationarity [34]. In addition, societal preferences are being reshaped by growing awareness of changing climate [35], which adds extra complexity to problems in water resource management by increasing dimensionality of objectives and decision variables. On one hand, new objectives and constraints are proposed in dimensions of environmental and social impacts [36, 37]; On the other hand, water resource systems are integrated and coordinated with other systems for efficient and sustainable operations [38, 39].

A recent boom in machine learning and sensing techniques has equipped us with new data-driven tools to tackle these challenges. And thanks to advances in computational power, highly complex models (e.g., deep neural networks) are successfully implemented in uncovering relationships between hydro-climate variables [40] or developing accurate emulators [41]. However, machine learning is not a panacea and correlation does not necessarily imply causation. A risk must be considered that models often become harder or even impossible to interpret when they get more complex [42]. To test these complex models against well-established physical knowledge then becomes hard, and lack of interpretability can sometimes be problematic when applying models on out-of-sample conditions [43]. Spurious correlations is an example of potential issues [44, 45] and a more detailed definition of spurious correlation can be

seen in Simon [46]. Therefore, more consideration must be put on the tradeoff between accuracy and interpretability when development data-driven models depending on research purpose. A noteworthy potential solution is to inform data-driven models with physical constraints so as to 1) avoid spurious patterns [47] and 2) to limit function spaces to explore [48].

In this dissertation, data-driven approaches that have good interpretability and are informed by or tested against physical constraints are examined across the sectors of climate, hydrology, and water resource management:

Chapter 2: We investigate the predictability of East African short rains at long (up to 12 month) lead times by relating seasonal rainfall anomalies to climate anomalies associated with the predominant Walker circulation, including sea surface temperatures (SST), geopotential heights, zonal and meridional winds, and vertical velocities. The underlying teleconnections are examined using a regularized regression model that shows two periods of high model skill (0–3-month lead and 7–9-month lead) with similar spatial patterns of predictability. We observe large-scale circulation anomalies consistent with the Walker circulation at short lead times (0–3 months) and dipoles of SST and height anomalies over the Mascarene high region at longer lead times (7–9 months). These two patterns are linked in time by anticyclonic winds in the dipole region associated with a perturbed meridional circulation (4–6-month lead). Overall, these results suggest that there is potential to extend forecast lead times beyond a few months for drought impact mitigation applications.

Chapter 3: Skillful long-lead climate forecast can be based on teleconnections between regional climate and large-scale circulations. Recent innovations in machine learning provide powerful tools in exploring linear/nonlinear teleconnec-

tions between climate variables. However, while it is hard to give physical interpretation of the more complex models, the simple models can be vulnerable to over-fitting, especially when dealing with the highly “non-square” climate data. In Chapter 3, as a compromise of interpretability and complexity, we test a regularized and interpretable regression model coupling pooling and elastic net. Performance of the model is evaluated in estimating the Three-Rivers Headwater Region wet-season precipitation using sea surface temperatures (SSTs) on Pacific Ocean and Indian Ocean at lead times of 0–24 months. The model shows better long-term predictive skill when compared against some commonly used regression methods including the Ordinary Least Squares, Empirical Orthogonal Function, and Canonical Correlation Analysis regressions. A correlation analysis indicates that the high skill relates to persistent regional correlation patterns between the predictand precipitation and predictor SSTs. Furthermore, flexibility of the model is demonstrated using its logistic regression version, which shows good predictive skill around a long lead time of 22 months. Consistent clusters of SSTs are found to contribute to both models. Two SST indices are defined based on the major clusters of predictors and are observed to be significantly correlated with the predictand precipitation at the corresponding lead times. In conclusion, the proposed regression model demonstrates great flexibility and advantages in dealing with collinearity while preserving simplicity and interpretability, and shows potential as a cheap preliminary analysis tool to guide further study using more complex models.

Chapter 4: Understanding dynamics of hydrological responses is essential in producing skillful runoff forecast. This can be quantitatively done by tracking changes in hydrology model parameters that represent physical characteristics. In Chapter 4, we implement a Bayesian estimation method in continuously es-

timating hydrology model parameters given observations of rainfall and runoff for small watersheds. The method is coupled with a conceptual hydrology model using a Gamma distribution-based Instantaneous Unit Hydrograph. The whole analytical framework is tested using synthetic data as well as observational data from the Fall Creek watershed. The results show that the Bayesian method can well track the hidden parameters that change inter-annually. Then the model is applied to examine temporal and spatial variability of the rainfall-runoff responses and we find 1) a systematic shift in the rainfall-runoff response for the Fall Creek watershed around 1943 and 2) a statistically significant relationship between rainfall-runoff responses and watershed sizes for selected NY watersheds. Our results demonstrate potential of the Bayesian estimation method as a rapid surveillance tool in monitoring and tracking changes of hydrological responses for small watersheds.

Chapter 5: Advances in sensing technology and machine learning offer new opportunities in reservoir management under a changing climate by providing highly-accurate hydroclimate observations and forecasts. However, such high-resolution data may not be accessible in less developed areas and water managers may be limited to data of poor resolutions. In Chapter 5, we attempt to address effects of this mismatch on system optimization performance in a quantitative way by managing a multi-reservoir system using forcing data of different resolutions (i.e., perfect versus binary). The analysis is illustrated using the upper Yellow River reservoir system where a series of 12 reservoirs serves to supply water for downstream cities and agricultural districts while providing electric power to the Northwest Grid and load-balancing capacity for other renewable energy (e.g., photovoltaic and wind). Representative policies that approximate actual operations are developed using the Evolutionary Multi-

Objective Direct Policy Search approach and their performance in maximizing hydropower production, load balancing capacity and minimizing water deficit are evaluated. Switching from binary data to perfect data only leads to negligible improvement in expectation objectives but significantly improved robustness under altered conditions of initial water storage and flow distributions. By further decomposing time-varying sensitivities of the policies, we observe that both policies heavily rely on variability in water storage in making release decisions while the policy trained with perfect data can incorporate more information from flow, which makes them less sensitive to climate uncertainty. Overall, our study suggests that benefits of using high-resolution hydroclimate information can be concentrated in improvements of policy robustness and stability rather than expectation performance for large hydropower systems operating on relatively long timescales.

CHAPTER 2

SEASONAL PREDICTABILITY OF EAST AFRICA SHORT RAIN

2.1 Introduction

Food security has been and continues to be a major challenge in Eastern Africa. This region is particularly vulnerable to extreme weather events like droughts and floods, because the regional economy and food production are largely dependent on rainfed agriculture. As an example, the 2011-2012 drought caused a disastrous food crisis that impacted more than 13 million people in the Horn of Africa, killing between 50,000-100,000 people [49]. These impacts perpetuate a cycle of poverty [50] and highlight the critical need for improved climate resiliency in the region. Towards this end, this study seeks to improve seasonal predictions of rainfall in East Africa, with a particular focus on extended lead times (> 6 months), to enable early warning systems and mitigative actions that can help reduce the impacts of major droughts and floods.

Eastern African rainfall seasonality is bimodal, with two distinct rainy periods termed the 'long rains' (Mar-Apr-May, MAM) and the 'short rains' (Oct-Nov-Dec, OND or Oct-Nov, ON). Inter-annual variability of rainfall from the two seasons are relatively independent [51], with higher variability in the short rains but more total rainfall during the long rain season [52]. Further, inter-annual variability in the short rains is more closely related to external forcing [53, 54]. This motivates our focus on improving the predictability of the East African Short Rains (EASR) in the present study.

There is a well-established teleconnection between EASR and forcing from

large-scale circulation over the Indian Ocean basin. Under normal conditions, the short rains are influenced by a Walker circulation cell characterized by surface westerlies over the Indian Ocean, ascending air over Indonesia, upper-level easterlies, and descending air over East Africa [55]. Intensification of this circulation can result in drier conditions in East Africa. This cell, and by extension East African rainfall, can be modulated by underlying sea surface temperature anomalies (SSTAs) in the Indian Ocean. Saji et al. [56] related tropical East African rainfall anomalies to the Indian Ocean dipole (IOD), a gradient in SSTAs with positive events defined by warm waters off the coast of East Africa and cold waters near Indonesia. Black et al. [57] identified a similar relationship but found it to be strongly nonlinear, with rainfall anomalies only associated with extreme and persistent SSTA events that reverse the normal zonal SST gradient for several months. Under these situations, strong easterly winds in the northern-central Indian Ocean weaken the westerly surface flow that normally transports moisture away from East Africa. The anomalous easterlies can even reverse the full Walker cell, enhancing convergence and convection over East Africa and leading to greater rainfall in that region during boreal autumn.

The dynamical pathways were further explored by Ummenhofer et al. [58], who used ensemble simulations to show that the warm western SST pole of the IOD has a larger role than the eastern cold pole in promoting stronger easterly winds from the Indian Ocean and moisture convergence over East Africa. Bahaga et al. [59] also confirmed the dominant role of the western pole of the IOD using AGCM simulations, which demonstrated that the warm western pole initiated a Gill-type response. In Hastenrath et al. [60], the important role of upper-level atmospheric motion was addressed: the summer warming and high stand of upper-tropospheric topography over south Asia enables strong

upper-tropospheric easterlies over the northern-central Indian Ocean in the following boreal autumn, while lack of such a mechanism in boreal spring leads to less predictability of the long rains. A more recent study by Nicholson [61] utilizing a 139-year rainfall record showed that the links between the Walker cell and the East African short rains were significantly weaker during certain historical periods (e.g., 1920-1960) and stressed the time dependence of the links. The non-stationary correlation between Indian Ocean SSTs (e.g., IOD) and EASR was further investigated by Bahaga et al. [62], which suggests that more inter-annual variability of the EASR was explained by the Indian Ocean SSTs in recent years.

Many studies have demonstrated the potential for seasonal prediction of the EASR based on the above teleconnection patterns [52, 54, 63, 64, 65]. Much of the prediction skill appears linked to dynamics internal to the Indian Ocean, and in particular, measures of the Walker cell across the Indian Ocean. While there is some predictability associated with dynamical fields, the east-west SSTA gradient provides superior prediction skill up to three months lead time. Mutai et al. [52] developed a regression model using the Jan-Aug-Sept (JAS) and OND principal components associated with (VARIMAX) rotated empirical orthogonal functions (EOFs) of the global SSTs, which yielded correlation coefficients of [0.56 - 0.78] in the testing period at a lead-time of 1 month. Hastenrath et al. [54] used a stepwise regression model with multiple predictors that are measures of the Walker cell (e.g., surface and upper-level zonal winds, vertical air motion, SST gradient, and pressure difference), and produced cross-validated correlation coefficients within the range of [0.02, 0.74] at a lead-time of 1 month. Nicholson [64] used a similar regression framework and achieved cross-validated r scores of 0.7 at a lead-time of 2 months primarily using atmo-

spheric predictors, and r scores within [0.6, 0.71] at a lead-time of 5 months using mainly SST predictors. Other work [63, 66] also suggests that ENSO plays a limited role in controlling modulations in the Walker cell over the Indian Ocean, and thus has less value for seasonal forecast development compared to direct observations of the Indian basin circulation. Also, Bahaga et al. [62] disentangled the IOD and ENSO influence on the EASR, showing that the ENSO influence is mediated by an in-phase occurrence of IOD.

To the best of our knowledge, the longest lead times of skillful seasonal forecasts have been limited to around 3-5 months. These lead times are related to measures of the Walker cell used in forecast development, which are first observed as SSTAs over the East Indian Ocean that develop in June [67]. These lead times, while useful, may not be long enough to enable decision makers to effectively execute mitigative actions. For example, if water managers knew that there was a high likelihood of drought in the short rains prior to or during the previous long rain season (6-8 month lead), they could prepare by reducing the allocation of water supplies provided by the long rains to mitigate water shortages later in the year. Similarly, long lead times are needed for use in cropping decisions. This motivates the question of whether we can extend seasonal forecast lead times of the East African short rains by focusing on large-scale climate processes that trigger subsequent perturbations in the Walker cell.

We posit that a focus on the Mascarene High and its relationship to inter-annual variability of EASR presents a promising approach for this extension. The Mascarene High (MH) is a semi-permanent subtropical ridge in the southern hemisphere off the coast of Madagascar, and has been shown to have great impacts on variability of the Asian Monsoon [68, 69]. Black et al. [57] proposed

that anomalies in the MH could initiate a perturbed meridional circulation cell, that in turn could influence the Walker circulation and by extension EASR. The link between the MH and EASR was further studied by Manatsa et al. [70], who used correlation and composite analyses for flood and drought events based on observations and reanalysis data to explore the connection and found that EASR variability is strongly linked to the zonal displacement of the MH. When the MH eastern ridge is anomalously displaced to the west (east), the south east trade winds over the Southern Indian Ocean strengthen (weaken), which alters the SST pattern and leads to suppressed (enhanced) convection over East Africa. However, importantly, the seasonally lagged relationship between anomalies in the MH, the Walker circulation, and EASR has not been sufficiently studied to determine if there is potential to extend forecast lead times of EASR. This research gap is a major motivation of the present paper.

In this study, we first explore potential predictors from a predictor pool consisting of multiple climate fields over the Indian Ocean using a regularized regression model. Instead of only using spatially averaged rainfall as the predictand, the regression analysis is repeated for rainfall at every grid cell to examine spatial variability of the predictive skill. Then, regression coefficient maps are compared with Indian Ocean basin composites prior to historical extreme EASR events to better understand the predictive signals being identified by the regression models. Using a cross-validated procedure and additional lead-lag analyses on large-scale climate fields, we assess the robustness of the regression models and the resulting insights. Based on these results, we propose a theory on the potential for extended predictability of EASR.

2.2 Data

Monthly rainfall data between 1981 to 2017 is collected from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) [71] for its fine spatial resolution and relatively long temporal coverage. The gridded rainfall data incorporates satellite imagery ($0.05^\circ \times 0.05^\circ$ resolution) and in-situ station data and is re-interpolated to grids of 0.5° in this study. The CHIRPS rainfall performed very well when compared against in-situ observations, likely because of its direct inclusion of rain gauge data and microwave images during calibration [72]. This study focuses on the equatorial East African region (land area in $[5^\circ\text{S}-5^\circ\text{N}, 33^\circ\text{E}-47^\circ\text{E}]$), similar to the short rains area used in Bahaga et al. [62]. This region has a well-defined rainfall regime with consistent, bimodal seasonality [73]. Average rainfall between Oct and Nov is used to characterize the short rains in this study. Previous studies of the short rain season have either used rainfall over the OND [52, 51, 62] or the ON [54, 53, 63, 61] period to represent the short rains. Our preliminary analysis (Figure A.1) suggested that predictability was slightly higher for the ON than for the OND rainfall, and when regressions were built separately using Oct, Nov, and Dec rainfall, there was little connection between Dec rainfall and large-scale climate fields. Therefore, we selected ON rainfall as our short rain index to better isolate the predictive signal.

The distribution of the EASR is highly skewed, which can overly emphasize a few extreme wet events in statistical analysis of the data. Therefore, the standardized precipitation index (SPI) is used based on a fitted gamma distribution [74]. The EASR series at all grid cells were first averaged and then the SPI was calculated based on the spatial average. Extreme events are defined as the

events with absolute values of SPIs larger than 1 based on the spatial-averaged EASR SPI, which is shown in Figure 2.1.

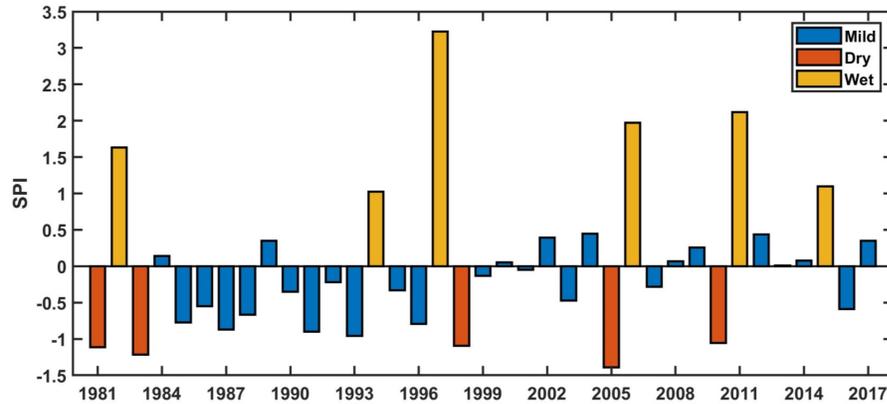


Figure 2.1: The spatial-averaged EASR SPI series from 1981 to 2017

The monthly gridded SSTs (1° latitude by 1° longitude) over 1980-2017 are collected from the Hadley Centre Sea Ice and Sea Surface Temperature data set [75]. Only SSTs in the Indian Ocean basin are used here since previous studies suggest the influence of SSTs on the EASR is 'local' [58, 59]. The Indian Ocean basin domain is based on the ocean boundary definition from the National Oceanic and Atmospheric Administration (NOAA). Several other climate fields ($2.5^\circ \times 2.5^\circ$ resolution) over the region of [65°S - 35°N , 20°E - 150°E] are also taken from the NCEP/NCAR Reanalysis II product over the period from Jan 1980 to Dec 2017, including 850h Pa geopotential height (height or HGT), 850 hPa (low) and 200 hPa (upp) zonal/meridional winds (u/v winds), and 500 hPa vertical motion in the atmosphere (omega). Standardized anomalies are computed for all variables except rainfall by first subtracting the monthly climatology and then dividing by the local standard deviation to reduce the effects of the dominant seasonal cycle and varying amplitudes at different latitudes. This step is done locally at each grid cell.

2.3 Methods

Linear models between the EASR SPI and several gridded climate variables (SST, height, u/v wind, and omega anomalies) at different lead times are developed using elastic net regression [76]. Elastic net is a penalized regression that is a balance between ridge [77] and the Least Absolute Shrinkage and Selection Operator (LASSO) [78] regression and promotes group selection, i.e., the inclusion or exclusion of groups of strongly correlated variables in the model. This approach helps manage collinearity and avoid over-fitting by selecting predictors among a large candidate set but retains groups of potentially important predictors instead of arbitrarily assigning the effect to a single variable (i.e., LASSO regression). A general form is given by

$$(\alpha, \beta) = \operatorname{argmin} \left\{ \frac{1}{2N} \sum_i^N \left(y_i - \alpha - \sum_j \beta_j x_{i-t,j} \right)^2 - \lambda P(\beta) \right\} \quad (2.1)$$

where the penalty term $P(\beta)$ is given by

$$P(\beta) = \frac{1-k}{2} \sum_j \beta_j^2 + k \sum_j |\beta_j| \quad (2.2)$$

Here, y_i is the EASR SPI at time i ; $x_{i-t,j}$ is the value of a climate variable at grid cell j at time $i-t$; (α, β) are the estimated regression coefficients; and λ is a non-negative regularization parameter. The parameter k controls the balance between ridge and LASSO penalties and is set to 0.01 in this study after preliminary testing of model performance with different k values. Models with different values of k (i.e., 1, 0.1, 0.01) yield similar model skill and a smaller k is used for better visualization (models with larger k would assign effects to very

few predictors, leading to very sparse coefficient maps). Based on this parameterization, the model selects several groups of gridded climate variables from the original data fields that are representative of the larger spatial patterns of climate related to the short rains.

The hyperparameter λ is typically chosen using a J-fold cross validation (CV) based on the training data [78, 79]. However, to repeat the CV for every climate variable and every lead time is not feasible given the computation expense. In addition, a constant λ helps ensure the degree of shrinkage is the same across each variable and lead time, making it easier to compare the results. Therefore, for each of the climate variables we fix the value of λ for all lead times. We first select λ using a 10-fold cross validation for 0 lead-time based on all of the data between 1981-2017. This step is repeated separately 100 times for each climate variable to account for variability introduced through the random selection of fold. The values of λ are then collected and the median values are selected and used in the following analysis. We note that sensitivity tests suggest that the regression model has low sensitivity to the value of λ .

With the hyperparameters (i.e., k and λ) determined, the models are fitted and tested 200 times using shuffled data, where in each iteration 27 randomly selected years are used for model training and the remaining 10 years are used for model testing. This cross-validation procedure is used to assess the robustness of regression model results. The spatial-averaged EASR SPI is used as the predictand and the analysis in this step is conducted separately for each individual climate variable using the corresponding λ s. Also, to account for spatial variability of the predictability, the regression is repeated using SPI at every grid by fitting the model using only SST as the predictor. These models are

only fitted over 1981-2009 and tested over 2008-2017. All results are evaluated in terms of the distribution of model skill, reported in the Spearman's correlation coefficients (CC). The statistical significance of the regression coefficients is not straightforward to calculate using a regularized regression, and so is not reported [80].

The elastic net regression will lead to a small number of selected grid cells for a subset of climate predictors that are most related to the EASR at different lead times. The spatial patterns of these selected climate predictors are also compared against climate composites to ensure the models are consistent with those composite patterns. According to our definition of EASR extremes, there are 6 wet extreme events (1982, 1994, 1997, 2006, 2011 and 2015) and 5 dry extreme events (1981, 1983, 1998, 2005, 2010) in the 1981-2017 record. We composite SST, height, u/v wind and omega anomalies 0-12 months prior to these extreme dry and extreme wet events. We examine composites on the difference between wet and dry events and separately for each type of event, to consider asymmetric controls on wet and dry events that are driven by different processes [57, 61]. A bootstrapping significance test is applied in the composite analysis, in which 1000 composite means of 6 randomly selected years (for the wet composite) and 5 randomly selected years (for the dry composite) are collected and their difference are calculated (i.e., 1000 composite differences). Based on these bootstrapped composites, only the results that are statistically significant at the 90% confidence level are shown.

Finally, a lead-lag analysis is conducted to further examine the patterns found in the regression and composite analyses. We define indices of large-scale climate fields for several of our covariates in key regions of the Indian Ocean

basin identified in the previous analyses, and then examine the lead-lag correlations of these indices with each other and with other indices from previous studies used to characterize the Walker circulation. The goal of this analysis is to help develop a better understanding of the causal chain of climate processes that lead to long-lead predictability of the EASR.

2.4 Results & Discussions

2.4.1 Model skill with shuffled data using spatially-averaged SPI

Regression model skill for the testing period using shuffled data is shown in Figure 2.2. For all climate variables, consistent decreasing trends in accuracy are observed for out-of-sample predictions as lead time increases from 0 to 5/6 months, with median CC values above or around $p = 0.1$ significance for concurrent (lag-0) predictions to around zero skill at a lead time of 6 months. We note the level of skill exhibited by the elastic net regression at lead times of 0-6 months is similar to that from the multiple linear regression by Nicholson [64]. Interestingly though, at longer leads (6-10 months) and for some climate predictors, out-of-sample prediction performance increases to levels comparable to lead times of 0-2 months, before dropping again to non-significant levels at lead times of 11-12 months. When SSTAs are used as predictors and at a lead time of 8 months, even the 25th percentile of prediction skill is above the 0.1 significance level. At a lead time of 10 months, the median CCs are around the significance level for models using height, omega and 200 hPa u wind; at a lead time of

12 months, the median CCs are above the significance level for models using height and omega.

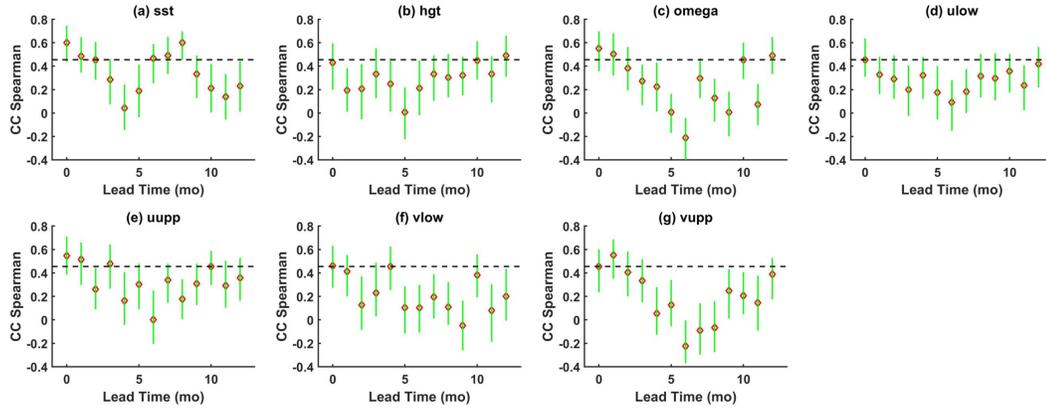


Figure 2.2: Model skill, shown here as a Spearman correlation between model predictions and observations, as a function of lead time for the testing period using shuffled spatially-averaged EASR SPI. The model predictors are anomalies of: (a) SST, (b) 850 hPa HGT, (c) 500 hPa OMEGA, (d) 850 hPa u wind, (e) 200 hPa u wind, (f) 850 hPa v wind, and (g) 200 hPa v wind. The 25th – 75th ranges are plotted in green lines; the median CCs are plotted in red diamonds. The one-tailed $p = 0.1$ significance level is plotted in the black dashed line.

2.4.2 Spatial variability of predictive skill

Spatial variability of the predictive skill is examined by repeating the regression analysis for SPI at each grid cell. Only SST is used as the predictor and the data are not shuffled in this analysis. Model skill as a function of lead time is shown in Figure 2.3. Consistent patterns are observed: at lead times of 0-5 months, model skill drops from above the 0.1 significant level to around zero skill; at a lead time of 8 months, model skill rises above that observed at a 0-lead. To make visualization of the results easier, we divide the lead times into ranges roughly

based on behavior of the prediction skill: 1) S1: lead times of 0-3 months; 2) S2: lead time of 4-6 months; 3) S3: lead times of 7-9 months; and 4) S4: lead times of 10-12 months. Prediction skill maps for our study area during the 4 seasonal lead times are shown in Figure 2.4. During S1 and S3, at most grid cells (83% and 80% for S1 and S3, respectively), significant prediction skill is observed for least one month within the season. However, only 54% and 62% of the grid cells are found to have significant prediction skill during S2 and S4, respectively. More grid cells with significant prediction skill are distributed along the coast for S4 and further inland for S2. Similar spatial patterns are observed for S1 and S3, in which only some grid cells around the central region are found to not have significant prediction skill. The gap region with relatively poor predictive skill overlaps with the eastern part of the East African Highlands. While the poor predictive skill could be physically based, we cannot rule out the possibility that it is resulted from the elevation-associated bias in the rainfall data [72]. We do note that consistent patterns of predictive skill were observed when the analysis was repeated using rainfall from the Global Precipitation Climatology Project (GPCP) (Figure A.2).

2.4.3 Comparison between climate composites and regression predictor patterns

The predictive skill at lead times of 7-9 months is notable because the anomalous Walker circulation has not yet been initiated [67], suggesting there is some other mechanism besides the tropical Walker Cell acting as a source of predictability. To better understand this source and determine how it propagates into shorter

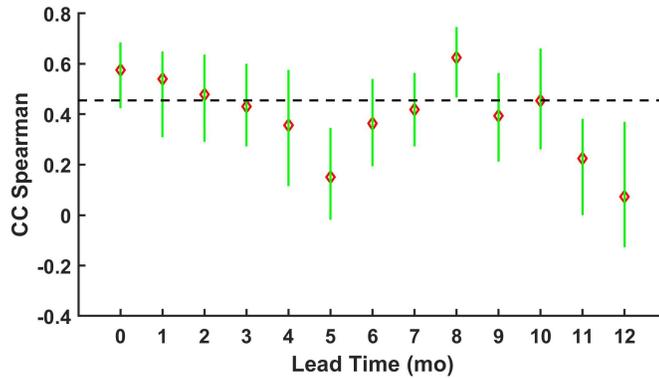


Figure 2.3: Distribution of regression model skill at each EASR SPI grid cell and for different lead times. The model is fit to SST data between 1981-2007, and correlation coefficients are shown for the testing period (2008-2017). The 25th – 75th ranges are plotted in green lines; the median CCs are plotted in red diamonds. The one-tailed $p = 0.1$ significance level is plotted in the black dashed line.

lead times, spatial distributions of regression coefficients for each period are examined in Figure 2.5 and 2.6. The coefficients are fitted using the full series of the spatially-averaged ESAR SPI between 1981-2017 and the regression analysis is done individually for each climate variable. Results from the u wind and v wind models are combined to produce coefficient maps for lower (850 hPa) and upper (200 hPa) winds. The coefficients represent sparse selection of predictor variables under elastic net regression from the large candidate set for the entire domain and highlight those regions that provide the best source of predictability at different lead times. It is worth mentioning here that the regularization procedure is designed to select only a small subset of grid cells to be included in the final model as a way to control for collinearity amongst the full covariate set. Therefore, the set of regression coefficients can be used to identify ‘hotspots’ where a certain climate variable most strongly contributes to the teleconnection, but the coefficients should not be interpreted as representing the true spatial ex-

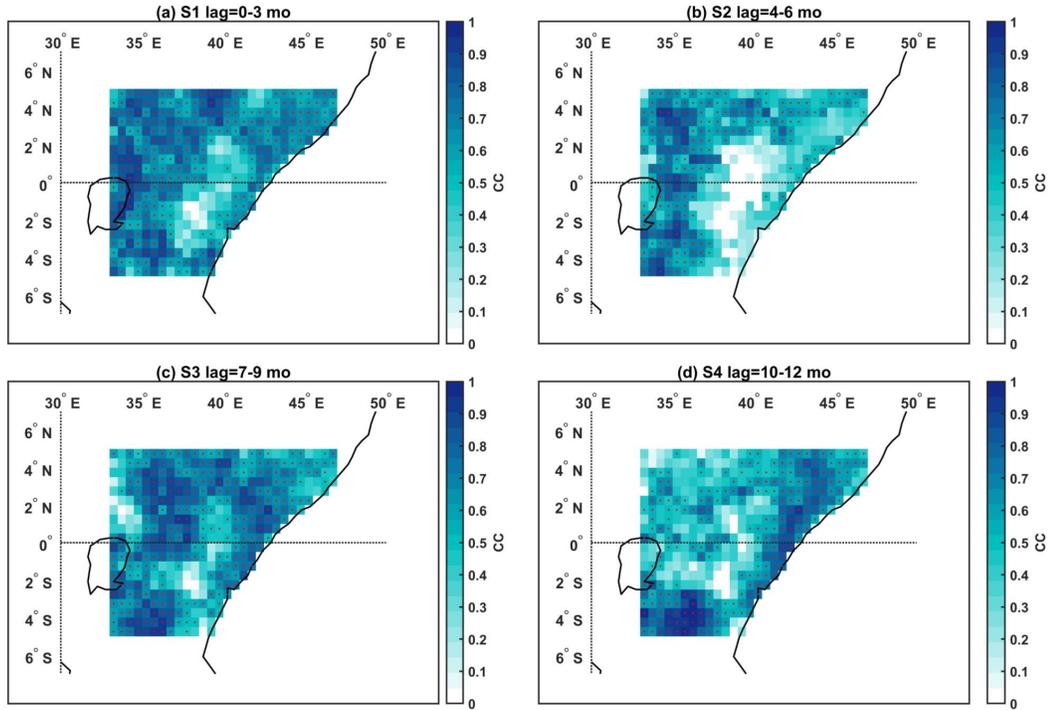


Figure 2.4: Maps of the best intra-season prediction skill for different lead time ranges. The regression analysis is conducted for periods of: (a) S1 (lead times = 0-3 months), (b) S2 (lead times = 4-6 months), (c) S3 (lead times = 7-9 months), and (d) S4 (lead times = 10-12 months), and the best intra-season skill is defined as the highest CC score for the testing period within the corresponding season. The grid cells with significant prediction skill (one-tailed $p = 0.1$) are marked with red dots.

tent of significant relationships between the climate fields and EASR. Hence, the coefficient maps are compared against composite maps (Figure 2.7-2.8) of wet-dry years that include no regularization and therefore present the teleconnection patterns more continuously.

During S4, negative (cold) SST coefficients are observed off the coast of the Horn of Africa while positive (warm) SST coefficients are observed near the Sumatra region, resembling a negative-phase Walker circulation (Figure 2.5d). Other characteristics of this negative phase are also observed: positive (subsidi-

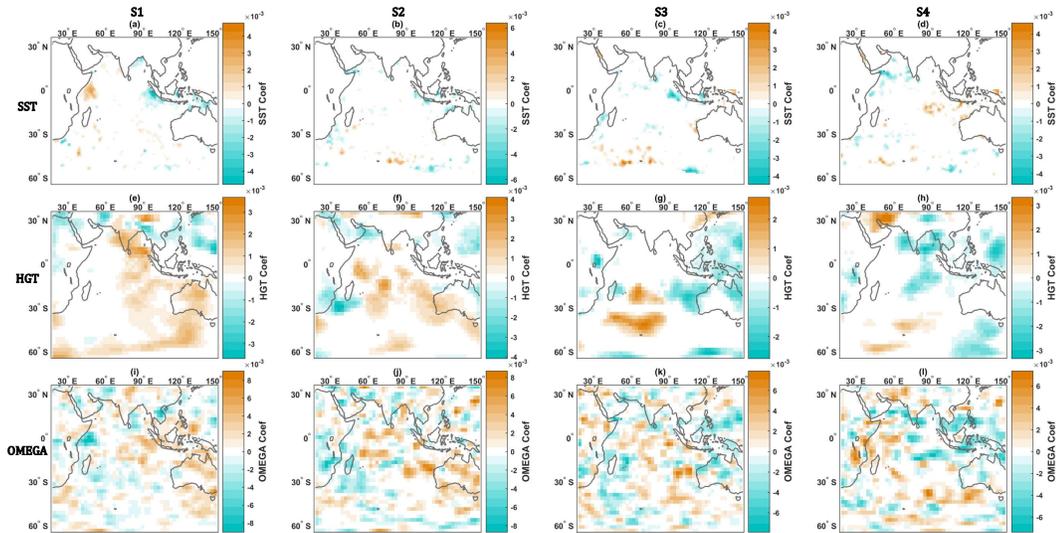


Figure 2.5: Coefficient maps for anomalies of (a-d) SST, (e-h) HGT, (i-l) OMEGA during periods of S1 (a, e, i), S2 (b, f, j), S3 (c, g, k), and S4 (d, h, l). Coefficients values are shown in color.

ing) omega coefficients are observed over East Africa (Figure 2.5l) and westerlies are observed over the tropical Indian Ocean for the 850 hPa winds (Figure 2.6d). All these factors contribute to less convection over East Africa and less transport of water vapor from ocean to land, which result in a drier short rain event. This suggests that dry short rain seasons tend to precede wetter seasons in the following year, which is supported by the negative autocorrelation of the EASR SPI at a 1-year lead-time (Spearman's rank correlation yields -0.47 for 36 samples, one-tailed p-value < 0.005).

S3 has notably high prediction skill observed before the initiation of the perturbed Walker circulation. In the SST coefficient map (Figure 2.5c), clusters of positive (warm) coefficients are observed around the Mascarene High area over the southern Indian Ocean and negative (cold) coefficients are sparsely distributed over the central and northern Indian Ocean. Positive height coefficients are also observed in areas slightly further to the north (Figure 2.5g). During S2,

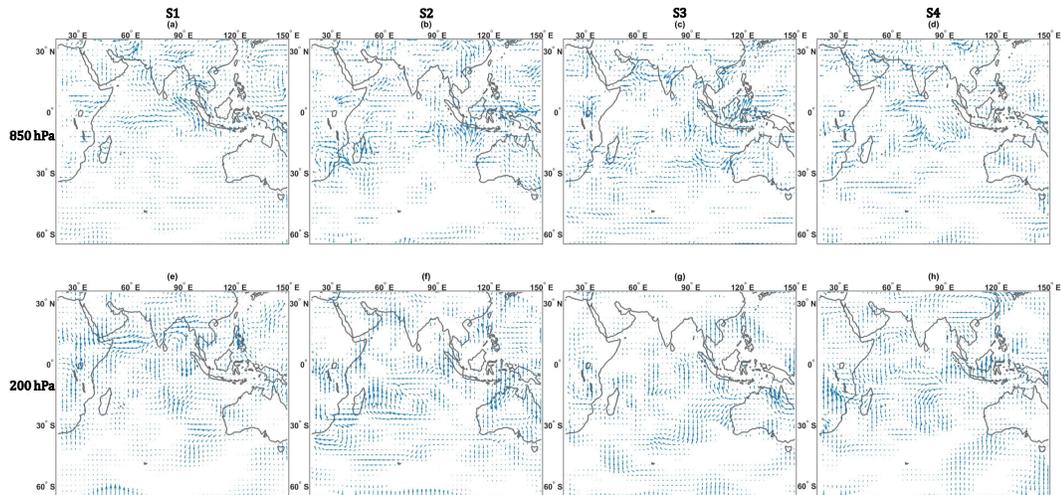


Figure 2.6: Coefficient maps for anomalies of (a – d) 850 hPa wind and (e – h) 200 hPa wind during periods of (a, e) S1, (b, f) S2, (c, g) S3, and (d, h) S4. Coefficients of the u winds are visualized using zonal arrows, with an eastward arrow denoting a positive coefficient; coefficients of the v winds are visualized using meridional arrows, with a northward arrow denoting a positive coefficient. Amplitudes of the coefficients are represented using the lengths of the arrows. For each atmosphere level (i.e., 850 hPa and 200 hPa), the u and v wind coefficient maps are combined by adding the two vector fields.

the model ceases to select SST predictors, as the SSTA and height features associated with the perturbed MH fades out and the modulated Walker cell is not fully developed (Figure 2.5b). Instead, the model emphasizes southerlies around the Eastern Indian Ocean near the Sumatra region for 850 hPa winds (Figure 2.6b). A plausible explanation for the drop in model skill during the transition period of S2 is that the climate variables chosen by the model (i.e., winds) are noisy, both in terms of measurement (i.e., reanalysis model) error and possibly with respect to geographic location, and therefore are more difficult for the regression model use when predicting EASR. Patterns of coefficients during S1 are very consistent and well described by the Walker circulation theory: positive (warm) SST coefficients (Figure 2.5a) and negative (uplifting) omega coefficients (Figure

2.5i) are observed in the western basin and opposite coefficients for the eastern basin; negative coefficients are observed over the tropical Indian Ocean for low-level winds (Figure 2.6a). The Eastern Indian Ocean is dominated by positive height coefficients (Figure 2.5e). Another relevant point is that the southerlies for low-level winds around the Eastern Indian Ocean persists from S2 to S1 (Figure 2.6b-2.6a).

We note that for all seasons, regression coefficients for omega are rather noisy (Figure 2.5i-l) and predictive skill using this variable is erratic (Figure 2.2c). Additional analysis (not shown) suggests these effects are a result of the spatial heterogeneity in the omega field, which can lead to instability in the regression model fit and predictions (as compared to smoother fields like SST and HGT).

Results from the composite analysis for the difference between historical extreme wet and dry events are shown in Figure 2.7 and 2.8. These are consistent with the patterns seen in the individual coefficient maps but aid in highlighting the important features. During the previous short rain season (S4, Figure 2.7d and 2.8d), anomalous tropical westerlies at low levels and easterlies at high levels are observed and associated with a tropical Indian Ocean SSTA gradient, with cold waters in the west and warm anomalies in the eastern Indian Ocean. The Sumatra region is dominated by negative omega anomalies, or enhanced uplifting, which is consistent with the local warm water pool and negative height anomalies at that time. The low-level pressure gradient and uplifting over Sumatra are also associated with anomalous surface westerlies that advect moist air away from East Africa, reducing the short rains.

During S3 (Figure 2.7c and 2.8c), a second dipole of SST anomalies appears

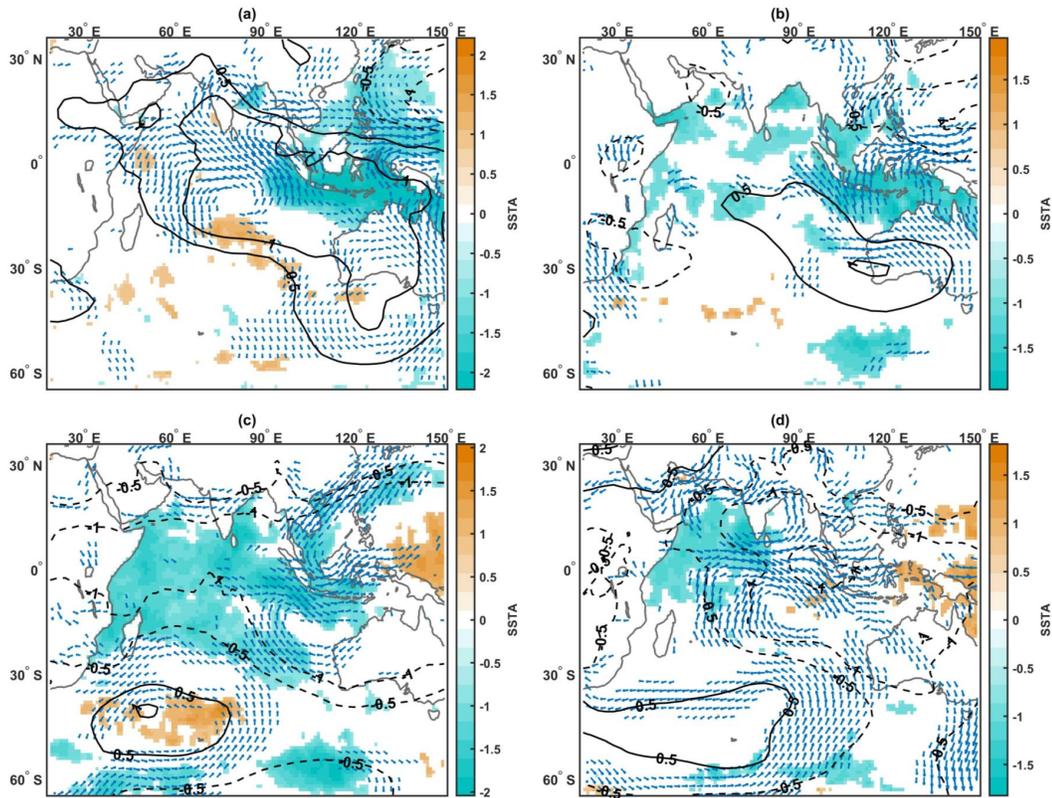


Figure 2.7: Composite maps for the difference between historical extreme wet and dry events for SST (shading), 850 hPa HGT (contour) and 850 hPa winds (arrows) for (a) S1, (b) S2, (c) S3, and (d) S4. Results are only shown if significant when compared to a bootstrapped 90% confidence interval.

over the central and southern Indian Ocean southeast of Madagascar (the Mascarene High area). A similar dipole pattern of height is observed. The warm SST pole is also dominated by anomalously positive height with a low-level anti-cyclonic wind pattern forming around it. During S2 (Figure 2.7b and 2.8b), the dipole of SST dissipates, and cold water continues to accumulate over the tropical Eastern Indian Ocean.

During S1 (Figure 2.7a and 2.8a), the anomalous easterly winds at low levels on the northern extent of the anticyclone reach the equator and strengthen. Over this same timeframe, cold water keeps accumulating over the tropical eastern

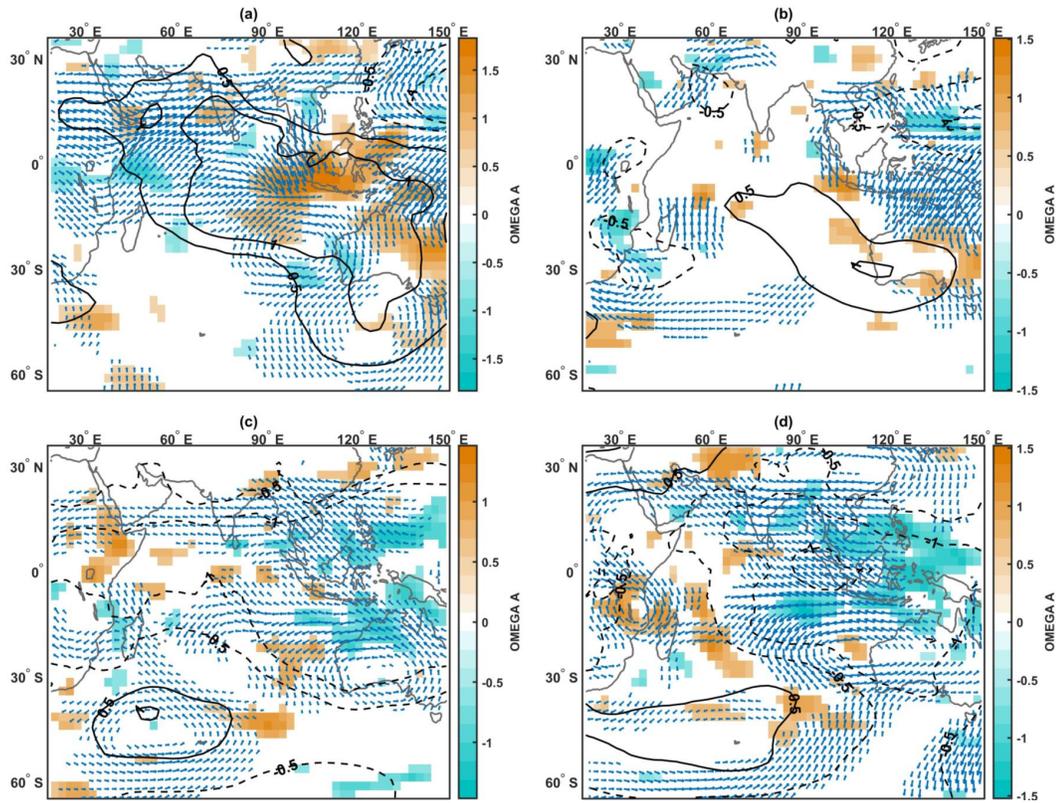


Figure 2.8: Composite maps for difference the between historical extreme wet and dry events for 500 hPa OMEGA (shading), 850 hPa HGT (contour) and 200 hPa winds (arrows) for (a) S1, (b) S2, (c) S3, and (d) S4. Results are only shown if significant when compared to a bootstrapped 90% confidence interval.

Indian Ocean near the Sumatra region, with mild, warm SSTAs in the western basin. Anomalous positive height anomalies dominate the tropical eastern Indian Ocean around Sumatra prior to the short rain season. Positive omega anomalies (subsiding air) are also observed over Sumatra while negative omega anomalies (uplifting air) are observed over the Horn of Africa. At high levels, the tropical region is dominated by anomalous westerlies. These anomalous conditions are all consistent with perturbations to the Walker circulation that have been proposed previously as controlling factors for the short rains and useful measures for seasonal forecast development [52, 54, 63, 64].

Similar composite maps are also examined for wet extreme and dry extreme composites separately. Overall, the average structures of the wettest and driest seasons appear relatively symmetric and represent opposite phases of the ‘Wet-Dry’ composite maps. Some noticeable differences are that in the wet-only events, the dipole of SSTAs is much more significant than that in the dry-only events, when the SSTAs over the Mascarene High area are relatively muted. In general, the long-lead progressions averaged over wet and dry short rain seasons tend to mirror each other.

Based on the analysis above, extreme short rain events appear associated with dipoles of SST and height anomalies and an anticyclone-like wind pattern at long lead-times (S3, 7-9 months in advance). We interpret the dipole of SST and height anomalies that peak during S3 as the manifestation of an intensified Mascarene High. The eastern branch of that circulation advects cold water to the eastern Indian Ocean near Sumatra from high latitudes, while the western branch of the low-level high advects equatorial warm water further south. As the dipole mode dissipates during S2, we draw on the theory proposed by Black et al. [57] to explain how processes maintain or even enhance the cold-water pool over the eastern Indian Ocean. We hypothesize that the local cold-water pool in that region modulates the Hadley circulation, perturbing meridional winds at a larger scale. This perturbed meridional circulation induces anomalous southerlies that keep transporting cold water to the eastern Indian Ocean from high latitudes. This process is also consistent with our observations during S2 and S1 when anomalous southerlies are observed over the eastern tropical Indian Ocean in the ‘Wet-Dry’ composites, along with subsiding air over Australia near 30°S. The enhanced cold pool around Sumatra then exerts a positive feedback to the anomalous Walker circulation, which provides the strongest source

of predictability for the short rains 0-3 months (S1) in advance.

2.4.4 The lead-lag analysis between large-scale climate indices

To further test the proposed causal chain, three indices are defined based on key features of the regression coefficient maps and climate composites from the analyses above (Figure 2.9). These include a long-lead SST dipole index (SL), a long-lead pressure dipole index (HL), and a long-lead meridional wind index (VL). The SL is defined as the difference between spatially averaged SSTs in two regions: ($50^{\circ}S - 35^{\circ}S, 40^{\circ}E - 80^{\circ}E; SL_p$) and ($0^{\circ} - 20^{\circ}N, 50^{\circ}E - 100^{\circ}E; SL_n$). The HL is defined as the difference between spatially averaged 850 hPa geopotential heights in those same two regions. The SL and HL are used to represent dipoles in SSTs and heights that emerge with an anomalous Mascarene High during S3. The VL is defined as the spatially averaged 850 hPa meridional wind in the region ($15^{\circ}S - 5^{\circ}N, 90^{\circ}E - 100^{\circ}E; VL$) and is used to quantify anomalous southerlies that transport cold water from higher latitudes during S2-S1. As a measure of the Walker circulation during S1, a short-lead SST dipole index (SS) based on Hastenrath et al. [54] is defined as the difference between averaged SSTs in ($5^{\circ}S - 10^{\circ}N, 45^{\circ}E - 55^{\circ}E; SS_p$) and ($15^{\circ}S - 5^{\circ}S, 90^{\circ}E - 110^{\circ}E; SS_n$).

Figure 2.10 shows the lead-lag correlations of these different climate indices. Here, correlations in the upper left (lower right) are associated with situations when the index on the vertical-axis leads (lags) the index on the horizontal-axis. Figure 2.10a demonstrates a strong statistical coupling between the Mascarene High-related dipole in SSTs (SL) and heights (HL) at lead times between 6-10 months prior to the short rains. The correlations indicate that anoma-

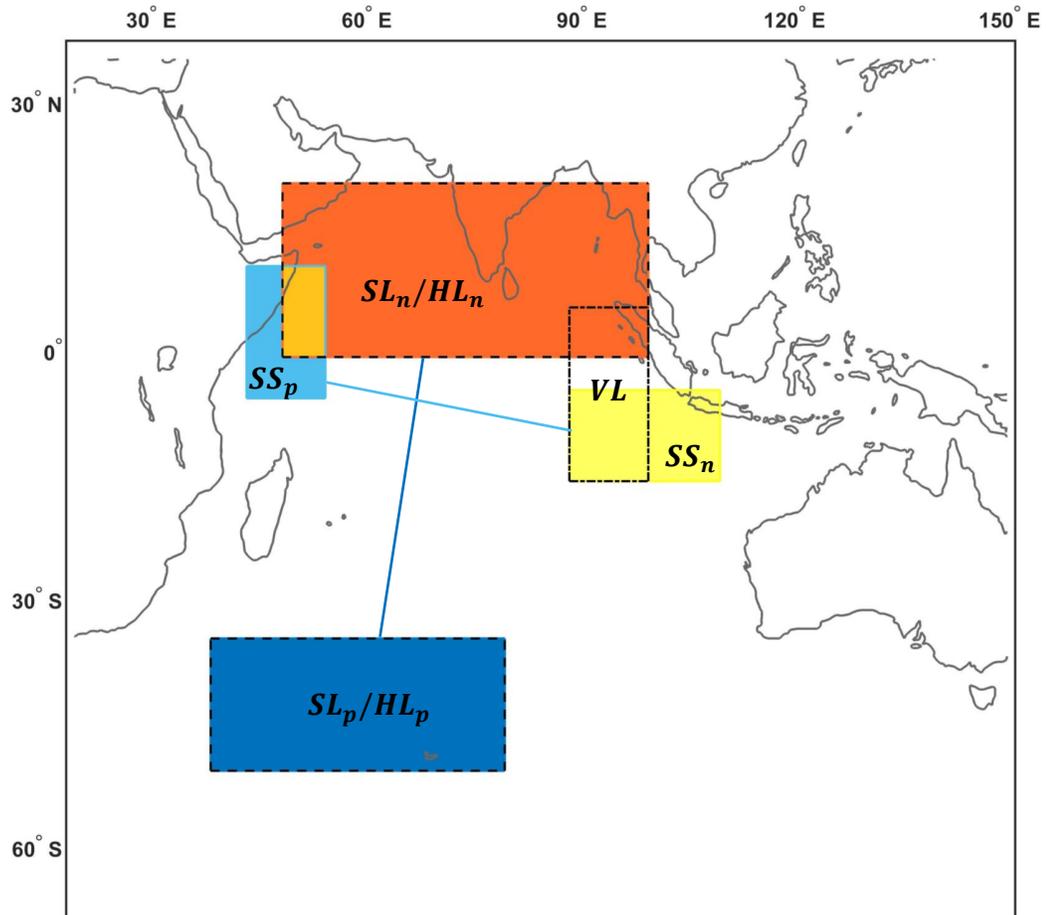


Figure 2.9: Regions used to define large-scale climate indices used in the lead-lag analysis, including the long-lead SST dipole index (SL_p and SL_n), the long-lead height dipole index (HL_p and HL_n), the long-lead meridional wind index (VL), and the short-lead SST dipole index (SS_p and SS_n).

lies in heights precede those in SSTs by between 1-3 months, suggesting that the SST anomalies are forced by atmospheric anomalies. The comparison between SL and VL in Figure 2.10b highlights that the long-lead dipole in SSTs for much of the year preceding the short rains, but especially at a lag of 7-10 months (December-March), is related to anomalous southerly winds in August and May (2 and 5 months prior to the start of the short rains, respectively). Figure 2.10c then shows that those southerly winds exhibit strong concurrent or

slightly lagged relationships with the short-lead SST dipole (SS) that provides a measure of the anomalous Walker cell. This is consistent with the theory that the Walker circulation is enhanced during S1 by persistent southerlies transporting cold water from higher latitudes to the eastern Tropical Indian Ocean. We also see that the SS and VL indices are correlated in August and May (for VL leading SS and vice versa), suggesting a link between the southerly winds and the enhanced Walker cell during the transitional S2 season. Finally, we note that the long-lead SL index is highly correlated to the short-lead SS index, with the correlation reaching 0.58 when the SL index leads the SS index by 8 months (not shown).

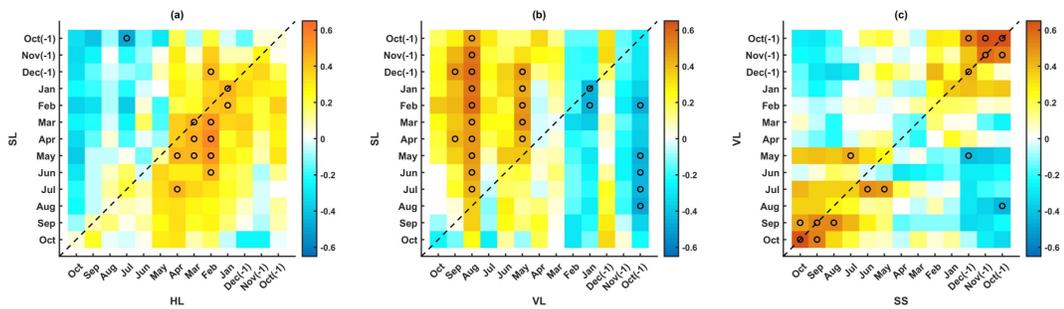


Figure 2.10: Correlation coefficients (Pearson's) at varying lead times prior to the short rain season between (a) SL and SS, (b) SL and HL, (c) SL and VL, and (d) VL and SS. The significant correlations are marked with black circles (one-tailed $p = 0.001$). A '(-1)' in the axis label indicates the month from the previous calendar year.

2.5 Conclusions

In short, we propose a causal chain of physical processes that explains the high predictive skill of EASR anomalies at lead times of 7-9 months (S3) (explained here for wet events):

1) A strengthened Mascarene High first appears and forms dipoles of SST and height anomalies during **lead times of 7-9 months** (Figure 2.5c, 2.5g and 2.7c);

2) The associated dipole mode triggers an anticyclone-like wind anomaly that advects cold water to the Eastern Indian Ocean region from high latitudes during **lead times of 4-6 months** (Figure 2.6b, 2.6c, 2.7b and 2.7c);

3) The cold water over the East Indian Ocean perturbs the local Hadley circulation, which further perturbs the meridional circulation at larger scales, enhancing the cold-water pool at **lead times of 4-6 months** [57];

4) The cold-water pool modulates the Walker circulation by inducing anomalous easterlies, bringing moist air to land, and enhancing uplift and convection over East Africa at **lead times of 0-3 months** (Figure 2.6a, 2.6a and 2.7a).

These patterns are identified by a regularized regression model, in which high predictive skill is observed for lead times of 0-3 months (S1) and lead times of 7-9 months (S3). Maps of regression coefficients show that certain groups of predictors were selected for those two lead times: at lead times of 0-3 months (S1), the model selected cold water over the Sumatra region and anomalous easterlies at 850 hPa, while at lead times of 7-9 months (S3), the model selected the dipole of SSTAs with a warm pole over the Mascarene High area and a cold pole over the northern and central Indian Ocean. Consistent patterns are also observed in the composite analysis, in which large-scale climate fields prior to historical extremes are examined and compared. A set of lead-lag analyses also supported the proposed causal chain. Our findings of high predictability at lead times of 7-9 months suggest that there is significant potential to extend

lead times in forecast models for East Africa, which could prove valuable for water management in the region.

While the results of this empirical study are promising, this study was limited to 37 years of observations. There is evidence that teleconnection between the Indian Ocean and EASR are non-stationary [61, 62]. Therefore, further analysis using general circulation model (GCM) based experiments is needed to confirm the dynamic pathways linking the two stages of high predictability and different climate fields. Such an experiment would require a modeling exercise that activates or suppresses SST and pressure anomalies associated with the Mascarene High, and then determines whether the activation of Mascarene High anomalies in the model are associated with the dynamic pathways in the Indian Ocean basin we identify in this study. This effort is left for future work.

CHAPTER 3

LONG-TERM FORECAST OF TRHR WET-SEASON PRECIPITATION

3.1 Introduction

Skillful long-lead (seasonal to annual) climate forecast is of great importance in managing large water systems. Examples include but are not limited to making water transferring plans for multi-reservoir systems running at annual to inter-annual time scales [81, 82], informing long-term agricultural decision making [83, 84], and developing early warning system for disaster mitigation [85, 86]. While local climate variability always fails to persist through such long lead times, the prediction can be made possible using long-lead teleconnections between regional climate and large-scale circulations. Anomalies of large-scale atmospheric circulations can be anchored by ocean memory due to massive heat capacity of ocean water and be released to perturb other circulations at a much later time [87, 88]. These perturbations can therefore be indicated by sea surface temperature anomalies (SSTAs). There are already well-established SST-based climate indices that have seen good use in long-lead climate forecasts such as the Niño SST indices [89, 90, 91], the Pacific Decadal Oscillation (PDO) [92, 93], the Tropical Northern Atlantic (TNA) and the Tropical Southern Atlantic (TSA) indices [94] and etc.

Approaches commonly used in developing long-term prediction models based on large-scale teleconnections can be roughly categorized into two classes: 1) physically-based simulation and 2) statistical models. While the physically-based simulation is widely used in investigating causality chains of climate processes, it is usually computationally intensive and requires expertise

for parameter calibration [95, 96]. Its statistical counterpart, in the meantime, provides an easy access to examining statistical associations between climate variables which could be further used to develop prediction models. The statistical models are becoming increasingly popular thanks to advances of sensing technology and internet which makes tremendously more data available at exceptionally high temporal and spatial resolutions [97]. Early efforts of statistical modelling are featured by qualitative analysis comparing time series of different climate variables [98, 10]. Most early work relied on insights of expert researchers and were done with data of rather limited sizes. Recent innovations in machine learning have developed powerful tools for examining linear/non-linear associations between climate variables in massive volumes in a more automated way. Just to name a few examples here: Kernelization is used to extend study domain from linear associations to nonlinear associations [99, 100]. Data processing tips like pooling and convolution are used to enhance model robustness by discarding/smoothing noises [101, 102]. Of the many machine learning approaches, neural network has become extremely popular across a wide range of spatial scales (local-global) [103, 104, 105, 40, 106]. Ham et al. [40] even successfully extended lead time of skillful ENSO forecast to one and a half years using a convolution neural network trained on historical simulations, which beat many state-of-the-art dynamical systems in terms of correlation skill for the Niño 3.4 Index. However, even though efforts are being made to improve model interpretability [107, 108, 109], tools for explaining the machine learning models are still insufficient [107], and to find physical interpretation of these models is usually hard or even impossible due to high model complexity. In this study, we looked at a generalized regularized regression method (i.e., elastic net) coupled with pooling as a compromise between model interpretability

and complexity, and examined its performance in predicting regional seasonal precipitation based on large-scale SST anomalies.

Regression has been broadly used in climate research and related fields. Typical applications include 1) change point detection [110, 111], 2) developing forecast models [112, 104, 113] and 3) identification of covariates with high predictive skill [114, 115]. Not only can a regression model identify linear relationships between climate variables at a given temporal basis (e.g., monthly or annual), but also it has good interpretability for guiding further research using more complex, nonlinear statistical methods or physically-based modeling experiments. These two features make regression especially popular in exploring teleconnections between regional climate and large-scale circulations. Hurrell [116] used a multivariate linear regression model to link changes in northern hemisphere temperature to extratropical climate indices. Krishnamurti et al. [112] developed a superensemble method for improving weather and climate forecast skills by using coefficients from multiple regressions. Wakabayashi and Kawamura [114] extracted four major teleconnection patterns in predicting Japan summer climate anomalies by combining the empirical orthogonal function (EOF) and regression. Van Oldenborgh and Burgers [117] developed a synthetic precipitation generator with regression models using the Niño 3.4 Index as the sole regressor to examine decadal variation in global ENSO-precipitation teleconnections. Yang and DelSole [118] used the regression coefficient maps to explore teleconnections between ENSO and fields of different climate fields. More recently, Zhang et al. [119] examined teleconnections between the Arctic sea ice decline and major climate indices using a quantile regression analysis. Only a few examples are listed here for context, as our intent is to test a generalized regression model with regularization in long-term seasonal precipitation

forecast instead of doing a fully review applications of regression in climate research.

However, over-fitting and over-parameterization are important issues for most regression analysis. These issues are particularly pertinent in climate research, since remote sensing data is usually highly 'non-square' (i.e., the total number of time series largely exceeds the length of the time series). Therefore, it typically is necessary to reduce effective dimensionality of the problems. Two commonly used approaches have been: 1) to select only a few dominant features/patterns (e.g., the Principal Component Analysis, PCA [114, 120, 121] or the Canonical Correlation Analysis, CCA [122, 118]); 2) to use only a few well-established climate indices [117, 123, 124, 119]. However, both methods have intrinsic disadvantages: Traditional dimensionality reduction methods like PCA and CCA try to decompose the global covariance structure of predictors (PCA) or between the predictand and predictors (CCA) and can miss important regional patterns while most climate indices are only defined by prior knowledge, which could limit the domain where we want to explore potential teleconnections.

In the past few decades, regularization has become increasingly popular in dealing with multicollinearity in regression. Two regularization approaches commonly used with regression are the L-1 norm (the least absolute shrinkage and selection operator, LASSO [78]) and the L-2 norm (the ridge regression [77]) regularizations. Other popular regularization approaches include the Akaike's Information Criterion (AIC) [125] and the Bayesian Information Criterion (BIC) [126]. Both L-1 and L-2 norm regularizations have shown good performance in alleviating or avoiding over-fitting in regression models in climate research

[115, 127, 128, 129, 121]. Yet, it should be noted that the ridge regression does not directly provoke sparsity of the regression model while the LASSO regression tends to assign non-zero value to only one of many correlated predictors which can make the model difficult to interpret. Lack of interpretability of the lasso model is also pointed out in a recent paper from Stevens et al. [130] where a graph-guided variation is used as an extra regularization to improve robustness of the regression model in predicting Southwestern US winter precipitation. Here, we propose to use the elastic net regression [76] which linearly combines the LASSO and ridge regression regularizations. While the LASSO regularization guarantees sparsity of the model, the ridge regression regularization helps improve visualization of the regression coefficient map and therefore, interpretability of the model [131]. On top of that, a pooling layer is added before developing the regression model. The pooling layer is commonly used in machine learning for reducing spatial dimensions [132, 133, 134]. And this extra pooling layer should help improve model robustness by avoiding a realistic problem that major 'hot' regions defining large-scale circulations are not fixed to certain spatial grids naturally.

The proposed model is tested to predict the Three-Rivers Headwater Region (TRHR) wet-season precipitation using the Pacific Ocean and Indian Ocean SSTs. The TRHR, located in the eastern Tibetan Plateau (TP), is often called China's Water Tower as from it flow the three major rivers of China: the Yellow River, the Yangtze River, and the Lancang (Mekong) River. Consequently, the TRHR plays a critical role in providing invaluable ecological goods and services as well as other resources like energy and food. While great efforts have been devoted to studying teleconnections between precipitation over the broader TP and large-scale climates [135, 136, 137, 138], quantitative studies focusing solely

on the TRHR are rather limited and only look at short lead times [139, 140]. In this study, we extend the forecast lead time up to 24 months and compared performance of elastic net against some widely-used regression methods including OLS multi-linear regression, the EOF regression, and the CCA regression. The precipitation is predicted in true amplitudes and binary states (wetter or drier than normal) to demonstrate flexibility of the model. Here, we seek a model that is computationally tractable for fast decision making support for stakeholders while retaining a relatively direct physical interpretation to aid further investigation of the underlying physical processes.

3.2 Data

We base our analysis on monthly precipitation data from Jan 1981 through Dec 2019 as collected from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). The original gridded precipitation data incorporates satellite data with in-situ station data, with a resolution of 0.05° by 0.05° [71]. In this study, we spatially averaged the TRHR precipitation over a rectangular area masking 89°E to 103°E and 31°N to 37°N as shown in Figure 3.1a. The monthly climatology for the spatially averaged precipitation is monomodal and shows that over 80% of the annual precipitation falls during the 5-month period of May-Sept, which we define as the wet season in this study (also known as the growing season for the TRHR [141]).

The CHIRPS precipitation is double checked against station-based precipitation (1988-2017) collected from the China Meteorological Administration (CMA). 55 stations with missing data ratio lower than 20% within the study

region are selected and a time series of arithmetically averaged precipitation is compared against that from the CHIRPS precipitation. Systematic shifts are observed around 2000 for both precipitations though the shift is less significant for the station-based precipitation. While difference in the shifts can be due to a non-uniform distribution of selected CMA stations, we do not want to diverge into this topic. Instead, the shift is removed by separately standardizing precipitation over 1981-2000 and 2001-2019. The standardization is done by subtracting the mean and dividing with the standard deviation. The standardized CHIRPS precipitation shows good consistency with the station-based precipitation as plotted in Figure 3.1b with a Pearson's correlation coefficient of 0.67 for 39 samples (p -value < 0.01). A binary time series of TRHR precipitation is defined for the multinomial regression model and the two states are defined as: 0 or dry for negative standardized precipitations, and 1 or wet for positive standardized precipitations.

SST is selected as the primary predictor since it can indicate perturbations in large-scale atmospheric circulations 'anchored' in ocean memory [87, 88]. Also, the SST field is less spatially heterogeneous compared to other commonly-used climate variables including geopotential height, vertical velocity of atmosphere (OMEGA) and wind velocities [131], which can help improve robustness of the regression models. Monthly SST data is collected from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) data set with a spatial resolution of 1° by 1° [75] over Jan 1979 - Dec 2019. Only SSTs from the Pacific Ocean and Indian Ocean basins are used to limit our study to regional processes and the basin ranges are based on the definitions from the National Oceanic and Atmospheric Administration (NOAA) via https://www.nodc.noaa.gov/woce/woce_v3/wocedata_1/woce-uot/summary/bound.htm. Similar positive

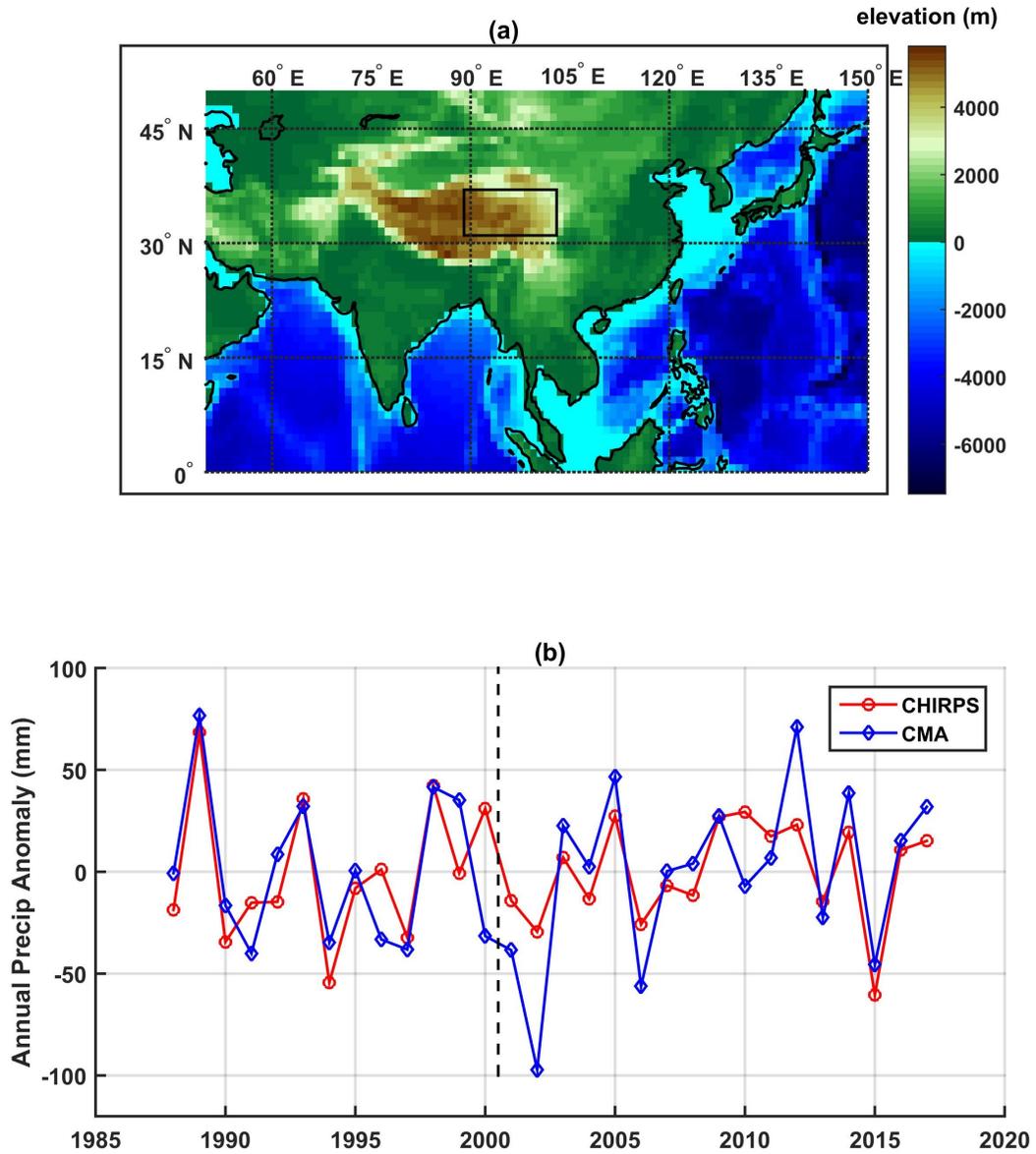


Figure 3.1: (a) The study region of TRHR (black box) as plotted in an elevation map based on ETOPO-5 [1] and (b) time series of the spatially averaged precipitation from CHIRPS (red) and CMA (blue) after standardization.

shifts are observed for most parts of the Pacific Ocean and Indian Ocean as seen in Figure B.1. To be consistent with the standardization of precipitation, the SSTs are also standardized separately for 1979-2000 and 2001-2019 to remove effects of the trends. This step is to ensure that model skill as measured by correlation

coefficients in the later sections will not be biased by the trends. The only difference here is that standardization of the SSTs is done locally at each grid cell and uses monthly climatology means and standard deviations to remove the dominant seasonal cycle.

3.3 Methods

3.3.1 The regularized regression

Here, we propose a two-step generalized regression model with regularization for dealing with collinearity when developing linear prediction models. The model first reduces dimensionality of predictors with a pooling layer, which is a commonly used method for down-sampling input representations. Then a regularized regression model is fitted using the pooled predictors to estimate real-valued or categorical predictand. A schematic is shown in Figure 3.2. In the pooling step, a new grid (in green) is defined by some characteristic values (e.g., maximum or median) of the smaller grids. An extra step of standardization is used to remove the systematic shift in precipitation around 2000 and to rescale precipitation and SSTs (since they have different amplitudes). In this study, we compared model performance using different pooling approaches (i.e., maximum/median/minimum pooling) with a squared window of four grid cells by four grid cells.

The regularized regression is given by Equation 3.1 with predictors x , predictand y , and regression coefficient (β, β_0) (β_0 is the intercept). Flexibility of this formula is achieved by using different deviance functions (Dev) for predictands

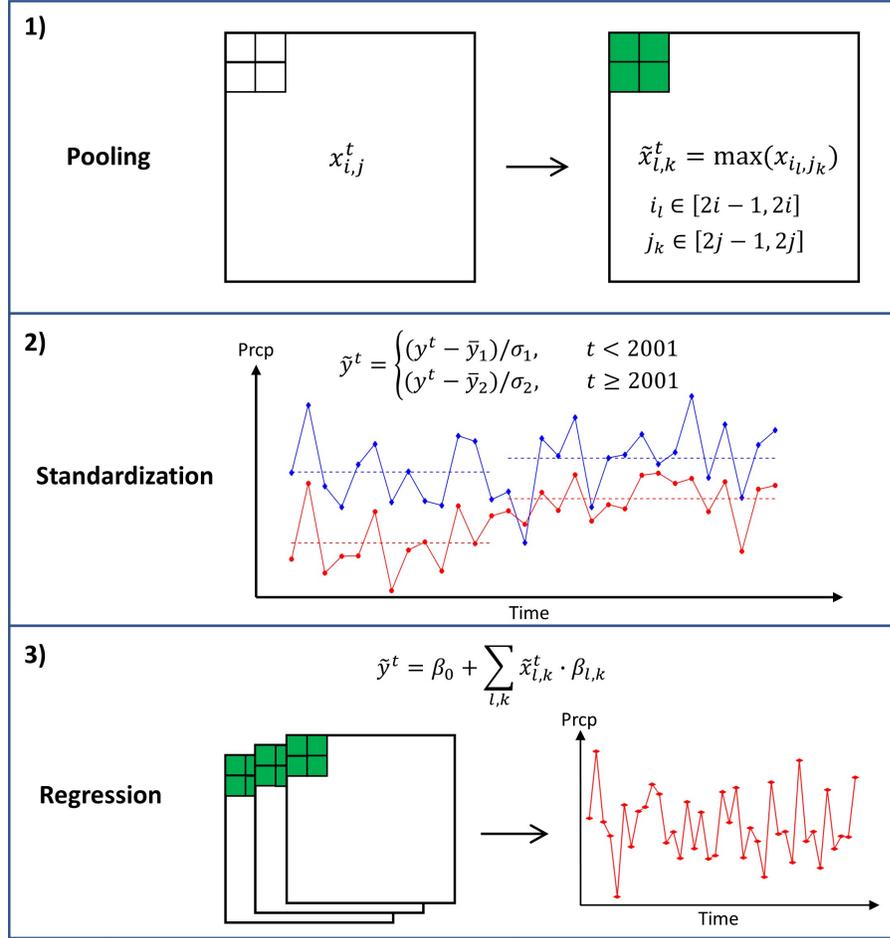


Figure 3.2: A schematic of the proposed the regression model coupling max pooling and elastic net

of different types. For example, the mean squared error (MSE) function is used for estimating real-valued predictands and the log-likelihood function is used for categorical predictands [142]. The elastic net regression is adopted here and the regularization term uses a linear combination of L-1 and L-2 norms of the regression coefficients as shown in Equation 3.2.

$$(\beta, \beta_0) = \min_{\beta, \beta_0} (Dev(x, y; \beta, \beta_0) + \lambda P_\alpha(\beta)) \quad (3.1)$$

$$P_\alpha(\beta) = \frac{1 - \alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \quad (3.2)$$

There are two hyperparameters in the model: α and λ . α balances the regularization between L-1 and L-2 norms of the regression coefficients β and is set to 0.01 for better visualization [131]. λ is usually decided using a k-fold (e.g., 5-fold) cross validation (CV) [78] and the λ value associated with the minimum cross-validated deviance function value (e.g., minimum MSE value for true-amplitude predictands) is used (often referred to as the MinMSE λ). However, this procedure can be computationally burdensome since we have to repeat the CV for all lead times. Therefore, a constant λ is firstly determined using the training set data at 0-month lead and is used for all lead times. A preliminary study demonstrates that significant predictive skill spikes in the testing period are not sensitive over a rather broad range of λ values as shown in Figure B.2. For the true-amplitude predictand, the Pearson's correlation coefficient (CC) and the Nash–Sutcliffe efficiency (NSE) score are used for model performance evaluation. For the binary predictand, an accuracy score S is defined as given by Equation 3.3 where $\mathbf{1}$ is an indicator function and \hat{y} is the predicted probability of y being 1 (i.e., wet).

$$S = \frac{1}{N} \sum_{i=1}^N \left\{ y_i \cdot \mathbf{1}(\hat{y}_i \geq 0.5) + (1 - y_i) \cdot \mathbf{1}(\hat{y}_i < 0.5) \right\} \quad (3.3)$$

3.3.2 Other regression models

Performance of the elastic net is compared against some commonly used regression methods in the two-step scheme including the OLS multilinear regression

[116], the EOF regression [114], and the CCA regression [143]. All regression methods use the same pooled SSTAs as predictors. Even though pooling can alleviate the issue of over-fitting, dimensionality of the pooled predictors is still highly non-square (39 years x 1343 grid cells). The OLS and CCA regression seek for a linear combination of predictors that maximizes its correlation with the predictand and does not regularize the model complexity. The EOF regression first projects the original predictors onto some 'dominant' basis vectors (often referred to as EOFs) by decomposing the covariance matrix of the predictors, and then uses the EOFs as the new predictors. It can implicitly regularize the model complexity by using only a few EOFs that explain most variance of the original predictors. The EOF is implemented such that the original (pooled) SSTAs are projected onto a set of orthonormal time series which constitute the new predictors. It should be noted that it is impossible to develop a prediction model this way since we are using data from testing set to construct the basis vectors. Here, the most dominant 50 EOFs accounting for over 88% variance of the original SST data are used as the new predictors in the EOF regression model.

3.3.3 The correlation analysis

A correlation analysis is designed to measure if any correlation patterns between the predictand precipitation and the predictor SSTAs persist through time. We propose a new correlation metric L to quantitatively measure persistence of any correlation patterns: for a certain lead time, we first compute lagged correlations between the TRHR precipitation anomalies and SSTAs at every ocean grid cell for 1981-2000 and reshape the correlation map into a col-

umn vector denoted by M_1 ; then this step is repeated for 2001-2019 to compute the column vector M_2 ; at last, L is defined by the correlation coefficient between the vectorized correlation maps M_1 and M_2 . L is bounded by a upper limit of 1, which represents an extreme scenario where correlations between TRHR anomalies and SSTAs are perfectly consistent before and after 2001. Under such scenario, a good regression model trained on data during the training period of 1981-2000 should produce significantly high predictive skill during the testing period of 2001-2019. However, L being close to zero does not necessarily mean no predictive potential for regression models. Since L measures persistence of the global correlations between the predictand and predictors, a regression model could select as predictors some regional clusters of SSTAs that have persistent correlations with the predictand precipitation. The metric L is used here to estimate how much degradation of performance is resulted from over-fitting by comparing against the testing period predictive skill from the regression models.

3.4 Results & Discussions

3.4.1 Comparison of regression models

Performance of the regularized regression models in predicting the TRHR precipitation in true amplitudes is examined in this section. The period of 1981-2000 is set as the training period while 2001-2019 is set as the testing period to be consistent with the standardization procedure. Since both pooling and regularization are designed for effective dimensionality reduction and thus to

avoid over-fitting, we first justify using the extra step of pooling by comparing predictive skill of the regularized regression models with and without pooling. A comparison of testing period correlation coefficients is shown for regression models without pooling and with maximum, median, and minimum pooling in Figure 3.3. Two spikes are observed at lead times of 13-14 months and 21-24 months and similar model skill patterns are observed using randomly split data (in Figure B.3). Significant improvement in model skill is shown for lead times of 13-14 months when pooling is used and at the lead time of 13 months, the predictive skill drops to below p -value = 0.1 significance level using non-pooled SSTAs. For lead times of 21-24 months, consistent improvement, though less significant, is observed. The improvement could be due to the fact that while there exist some consistent large-scale circulation patterns, the signals may not be fixed to certain grid cells depending on the spatial resolution and projection coordinate system. Therefore, the model robustness can be improved by including signals from neighboring grid cells with pooling. However, though not examined here, one must be careful with choosing the pooling window size since displacement of some circulation patterns can be important indicators of climate anomalies [144, 70] and this information may not be resolved when the pooling window is too large. It should be noted that λ is re-calibrated for the regression models using non-pooled SSTAs. And the statistical significance for the regularized regression is not straightforward to calculate and thus is not reported [80]. The maximum pooling is used in the following analyses.

Model skill for the testing period data as measured by NSE scores is reported in Figure 3.4. Consistent patterns are observed as two spikes of NSE scores are found around lead times of 14 and 22 months. However, even at those two lead times, the predictive skill is barely satisfactory. By further looking at com-

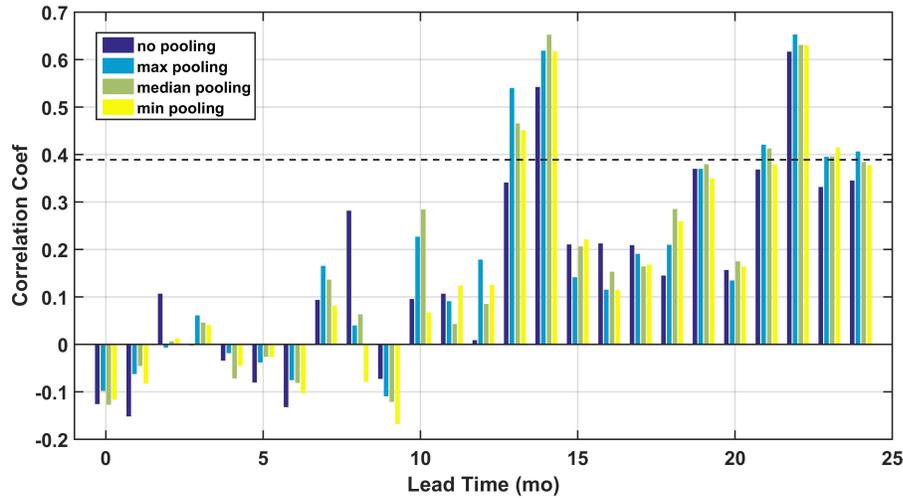


Figure 3.3: Comparison of the predictive skill in the testing period from the elastic net regression models with no pooling (dark blue), max pooling (light blue), median pooling (green) and minimum pooling (yellow). The p-value = 0.1 significance level is plotted in the black dashed line.

parison between the observed and estimated precipitation, we figured that the elastic net model markedly underestimated amplitudes of predictands. However, the extent of shrinking is consistent across the training and testing periods (as shown in Figure B.4a). Thus, in practice, one could ‘learn’ how much the amplitude is shrunk by looking at the training data and can then rescale testing period estimations. The rescaled estimations show significantly improved NSE scores: $NSE = 0.36$ for lead time of 14 months and 0.38 for lead time of 22 months (as shown in Figure B.4b). A plausible explanation is that the elastic net regression sacrifices accuracy in amplitude estimation for model robustness by selecting only a few predictors and shrinking amplitudes of regression coefficients. This effect is more significant with highly non-square data as in our case since greater regularization must be applied. Therefore, amplitude-based measures such as NSE and the root mean square error (RMSE) may not be applicable

for model skill evaluation. The rescaling method discussed above is not recommended since the model is designed to be biased for improved robustness. In following analyses, the Pearson’s correlation coefficient is used as the primary measure of model skill.

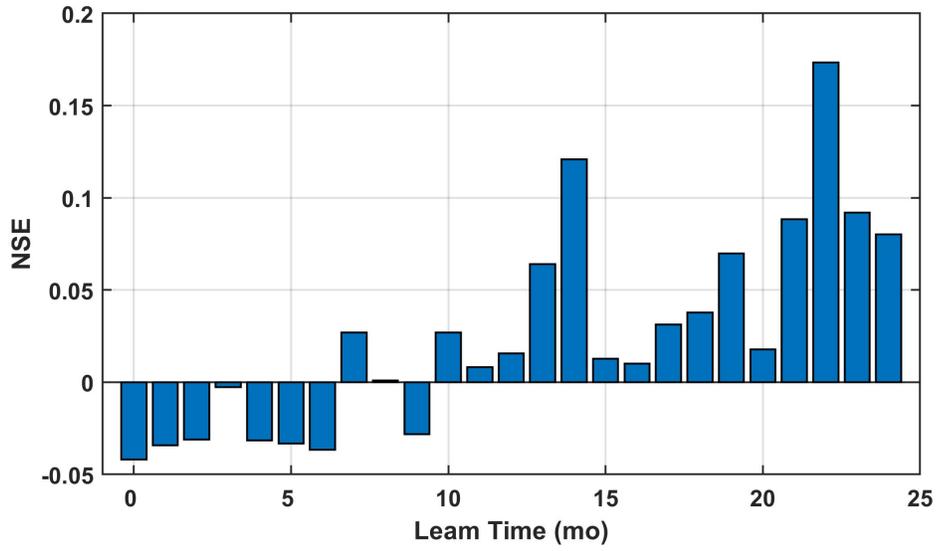


Figure 3.4: Predictive skill for the testing set as a function of lead time reported in NSE scores

We then justify the use of regularization by comparing predictive skill of the regression models with and without regularization. Comparison of the model skill from the elastic net, OLS, EOF and CCA regression models are shown in Figure 3.5. Statistically significant positive predictive skill is only observed for the elastic and CCA regressions models. The elastic net models show two spikes of good predictive skill at lead times of 13-14 months and 21-24 months. Statistically significant positive skill is observed for the CCA regression models only at the lead time of 18 months while that of the elastic net regression almost hits the p -value = 0.1 significance level at the lead time of 19 months. Overall, the elastic net regression shows more potential in addressing linear associations

between the TRHR precipitation anomalies and the SSTAs. While statistically significant positive skill are only found at rather long lead times, this does not necessarily mean that there is no connection between the TRHR precipitation and large-scale climate fields at shorter lead times. We are limiting our analysis to only using SSTs from the Pacific and Indian Oceans which is only one sector of the complex large-scale circulations including a wider range of variables like geopotential heights, humidity, vertical velocity of atmosphere (OMEGA) and horizontal winds etc. Furthermore, we are limiting our analysis in the frame of linear models as we only compare different types of regression models here. Instead of developing accurate forecast models, our intent is to examine how pooling and regularization would improve performance of the linear models at rather low costs. The better performance of elastic net is understandable here since it explicitly regularizes model complexity and provokes sparsity in regression coefficients by using the L-1 norm regularization.

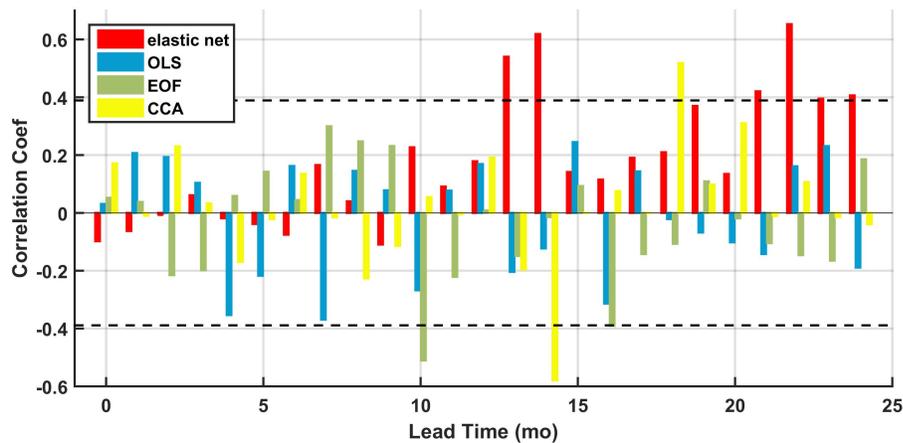


Figure 3.5: Prediction skill for the testing set of 2001-2019 measured by correlation coefficients as a function of lead times for elastic net (red), OLS (blue), EOF (green), and CCA (yellow) regression models. The p-value = 0.1 significance level is plotted in black dashed lines

3.4.2 Source of high model skill

A correlation analysis is conducted to measure at a certain lead time, how well a linear model based on the global correlation between the TRHR precipitation and the SSTs can perform. The potential predictive skill is estimated by the new correlation metric L defined earlier as L measures how the time-shifted global correlation patterns persist from 1981-2000 (the training period) to 2001-2019 (the testing period). A comparison between L and the predictive skill of the elastic net model at varying lead times is shown in Figure 3.6. Spikes in L are observed at lead times of 7, 14, 18, and 22 months. Three of the spikes coincide with good predictive skill from the elastic net model (i.e., lead times of 14, 19, and 22 months) while only one of the spike coincide with good skill from other regression models (i.e., lead time of 18 months for the CCA regression model). To estimate how much potential are realized for each model, correlation coefficients are computed between L and model skill from regression models over lead times of 0-24 months and are reported in Table 3.1. The only statistically significant correlation is found for the elastic net model (0.82 for 25 samples, p -value < 0.01) while rather low correlations are found for other regression models. The results suggest that the OLS, EOF, and CCA regression model do not perform well even when there exist persistent correlations between the predictand precipitation and the predictor SSTs.

Possible explanations are proposed here based on algorithms of the regression methods. For the OLS and CCA regressions, the models could be overfitted to noisy predictor signals for high training skill as both methods decompose the covariance between the predictand precipitation and predictor SSTs to seek a linear combination that either minimizes the MSE (for OLS) or maximizes

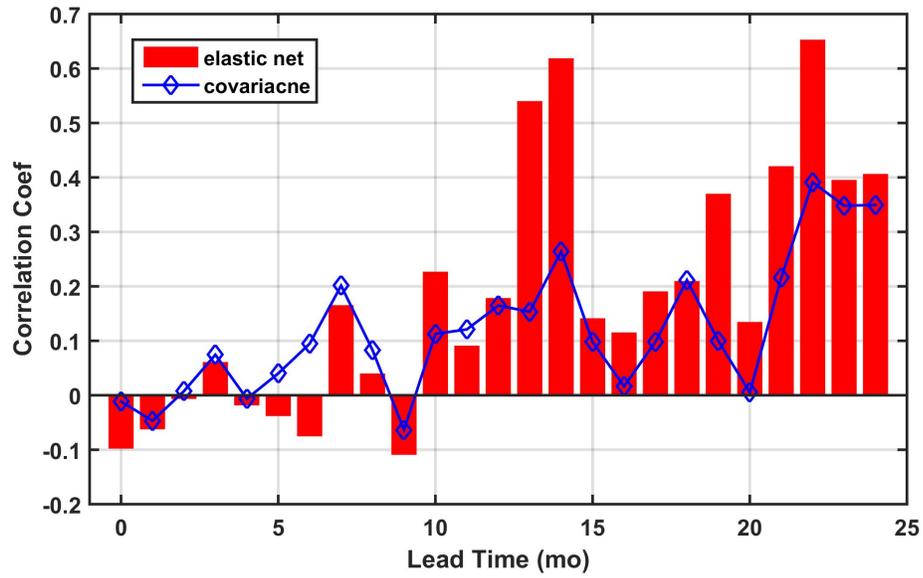


Figure 3.6: Comparison between predictive skill from the elastic net model (bars) and L_s (a blue diamond line) at varying lead times. The statistical significance levels are not shown here as we are comparing correlation coefficients calculated using samples of different lengths (20 for the predictive skill and 39 for the L_s).

Table 3.1: Correlations between L and prediction skill at varying lead times for different regression models

<i>Model</i>	<i>CC</i>
Elastic Net	0.82
OLS	0.01
EOF	-0.07
CCA	-0.11

the correlation (for CCA). Thus, the models can perform poorly on test period samples. As for the EOF regression, the EOF re-constructs the predictors by projecting the global covariance of the SSTs onto some dominant orthonormal basis

vectors (EOFs). There are two limiting factors: 1) the assumption of orthogonality may not be appropriate for some physical variables like SST; 2) any regional persistent correlation patterns between the TRHR precipitation and the SSTs could be lost if they do not make significantly large contribution to the global covariance. Technically, we are not predicting the TRHR precipitation with the EOF regression models since data of the full study period (1981-2019) is used for constructing the new set of predictors (i.e., EOFs).

Interestingly, comparably high predictive skill are observed for the elastic net models at lead times of 14 and 22 months while persistence of the global correlation is much lower at the lead time of 14 months as indicated by L . We specifically looked at the correlation maps between the TRHR precipitation and the SSTs before and after 2000 for the lead times of 14 and 22 months as shown in Figure 3.7 and 3.8, respectively. For the lead time of 14 months: before 2000, the correlation map features a cluster of positive correlations [180E-210E, 45S-15S] to the east of Australia and an extended band of positive correlations over the mid-north Pacific Ocean [120E-210E, 15N-30N]. Scattered and less significant positive correlations are observed over the northern Indian Ocean and to the west of South America; After 2000, the correlation map is dominated by two major clusters of positive correlations to the east of Australia [180E-210E, 45S-15S] and to the west of South America [260E-280E, 60S-15S], and one major cluster of negative correlations over the southwestern Indian Ocean [30E-60E, 60S-30S]. Less significantly positive correlations are observed over the north-western Pacific which overlaps with the extended band before 2000. For the lead time of 22 months, both correlation maps before and after 2000 are dominated by large clusters of positive correlations over the northern Indian Ocean [60E-90E, 15S-15N] and eastern tropical Pacific Ocean [210E-270E, 15S-0]. The major difference

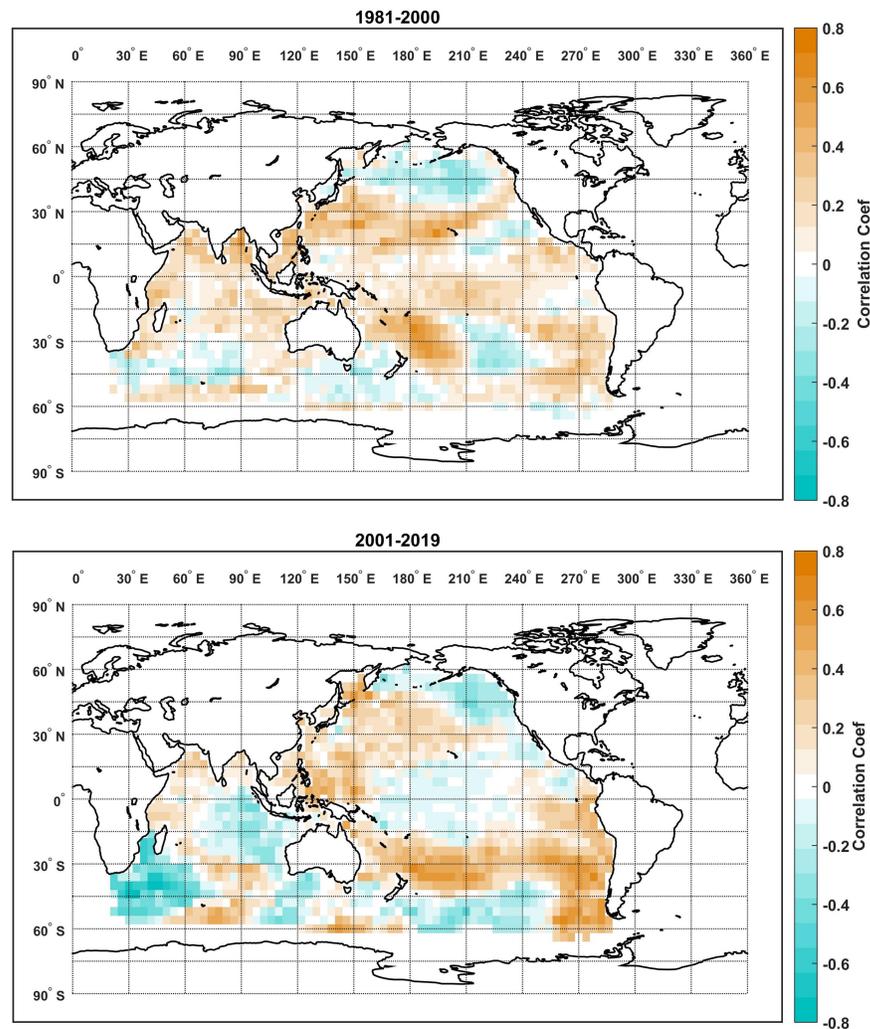


Figure 3.7: Correlation maps between the TRHR precipitation and the SSTs over the training period of 1981-2000 (top) and the testing period of 2001-2019 (bottom) at the lead time of 14 months

is that clusters of positive correlations over the southern-eastern Pacific and the mid-western Pacific [120E-150E, 15N-30N] get enhanced in correlation amplitude and extended in spatial coverage. A comparison between Figure 3.7 and 3.8 suggests a higher level of persistence in the global correlation between the predictand precipitation and the predictor SSTs at the lead time of 22 months, which is consistent with the greater value of L in Figure 3.6. However, a smaller

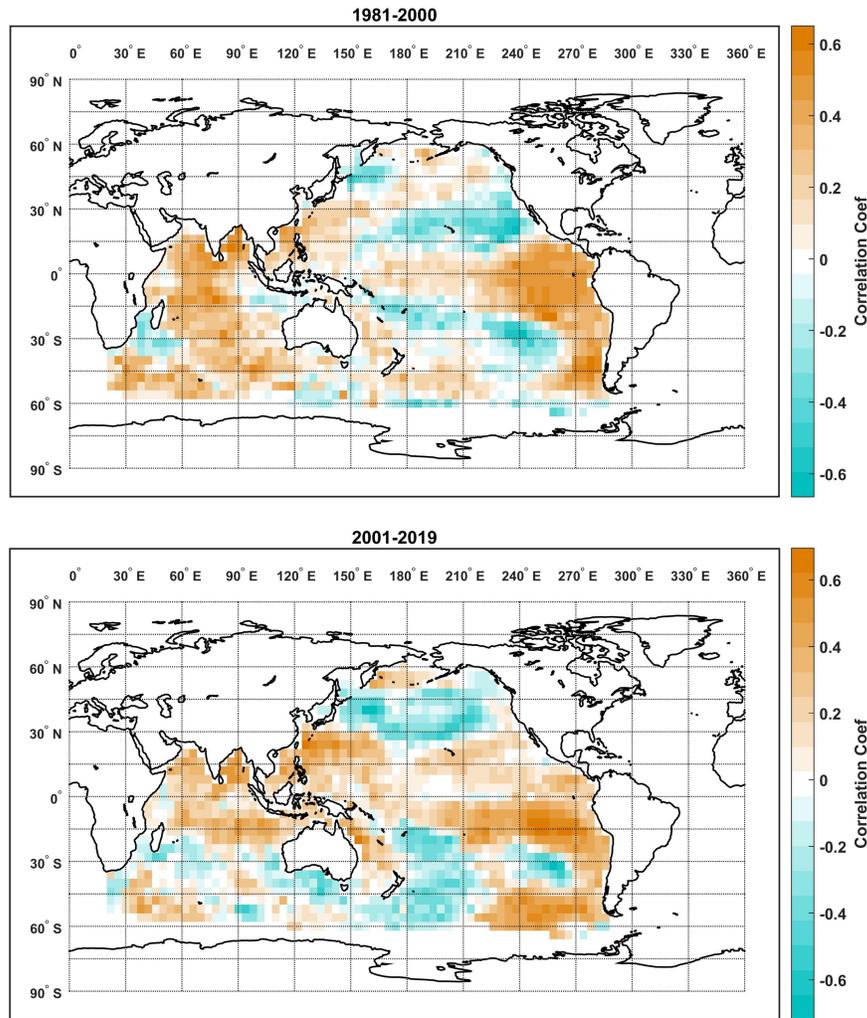


Figure 3.8: Correlation maps between the TRHR precipitation and the SSTs over the training period of 1981-2000 (top) and the testing period of 2001-2019 (bottom) at the lead time of 22 months

value of L does not necessarily mean low predictive potential for a linear model as some regional persistent correlation patterns can be utilized (e.g., the positive correlation cluster to the east of Australia at the lead time of 14 months). The correlation maps are further compared with regression coefficient maps in Section 4.3 as we attempt to interpret the high model skill of the elastic net regression.

3.4.3 An alternate multinomial regression model

In this section, the binary precipitation is predicted using the multinomial regression version (i.e., the elastic net logistic regression) of our model. While new machine learning techniques like the classification and regression tree (CART) can have better model skill in multi-class prediction of climate variables [145, 146], the elastic net logistic regression is tested to demonstrate flexibility and consistency with different types of deviance functions. The multinomial regression may have more use in practical application since amplitudes of the predictand tend to be underestimated because of the regularization [131]. The logistic regression is implemented by simply replacing the MSE function with the log-likelihood function for Dev [142]. Though the logistic regression is a special case of the multinomial regression, the model should be easy to generalize for predictand of more than two categories by separately fitting a regularized Poisson regression for each category, of which the coefficients are used to estimate coefficients of the multinomial regression model [147].

To extend the sample size, the leave-one-out cross validation is used (the testing sample size is thus increased from 19 to 39 here). The leave-one-out CV is not used in the previous analysis for predicting real-valued predictand since the Pearson's correlation coefficient is used for performance evaluation and the leave-one-out CV could result in bias for violating continuity in data. Accuracy reported in the S scores as a function of lead time is shown in Figure 3.9. A consistent spike of high accuracy (over 70% accuracy) is observed at around lead times of 22 month. While a local maximum of accuracy is observed at the lead time of 14 months, it is not statistically significant. A major reason could be that the S score is not an equivalent measure of the correlation coefficient as we

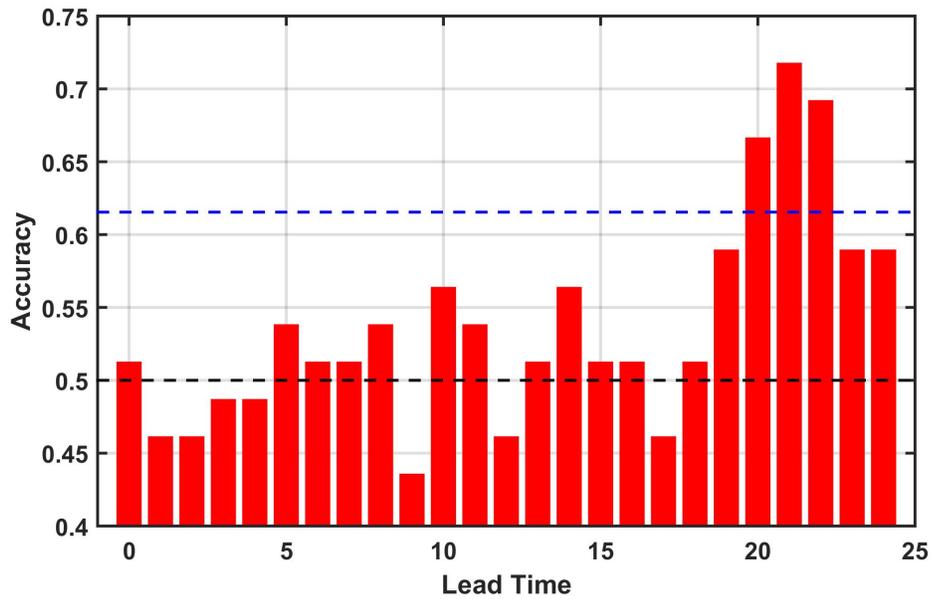


Figure 3.9: Accuracy of predicting the wet-dry state of TRHR precipitation at varying lead times as reported in S scores. The baseline skill using a null model of totally random guess is plotted in a black dashed line. The bootstrapped p-value = 0.1 significance level is plotted in a blue dashed line (10,000 samples for each lead time).

used in Figure 3.3 and 3.5. An example of high correlation but low S is when a model can well predict extreme events but do poorly for less extreme events.

Consistency between the regression models using predictand and deviance function of different types is further examined by comparing maps of the regressions coefficients at lead times of 14 and 22 months as shown in Figure 3.10 and 3.11, respectively. For the lead time of 14 months, both models feature a major cluster of positive coefficients to the east of Australia while more non-zero coefficients are observed for the logistic regression model over the southeastern Pacific Ocean and the Indian Ocean. As for the lead time of 22 months, both maps are dominated by a large cluster of positive coefficients over the eastern tropical Pacific Ocean. While positive coefficients are observed over the northern Indian

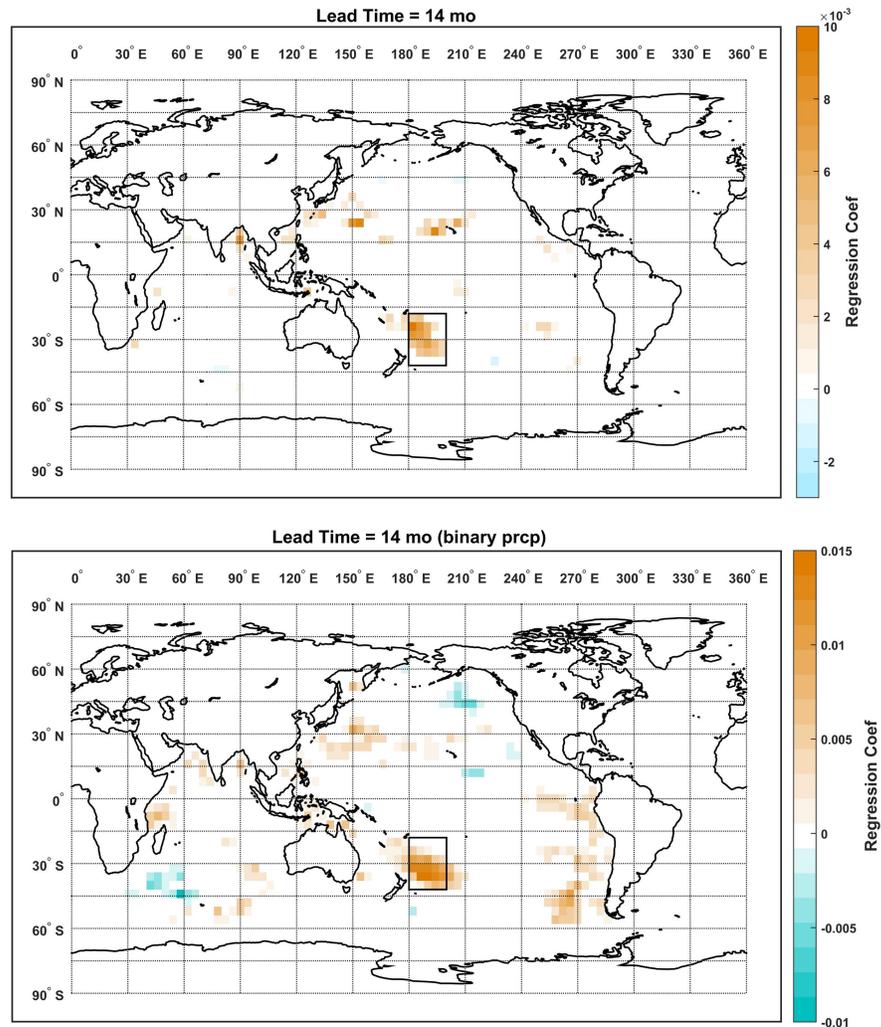


Figure 3.10: Maps of regression coefficients from the elastic net models using true-amplitude (top) and binary (bottom) predictand precipitation at the lead time of 14 months

Ocean for both models, the coefficients are more sparsely distributed when the true-amplitude predictand is used. More non-zero coefficients are found for the logistic regression models at both lead times, which could be due to a less optimal regularization as we only did the cross validation on a relatively sparse sequence of λ s with a lead time of 0 month. But overall, the major clusters of non-zero regression coefficients are consistent across the models. And this is

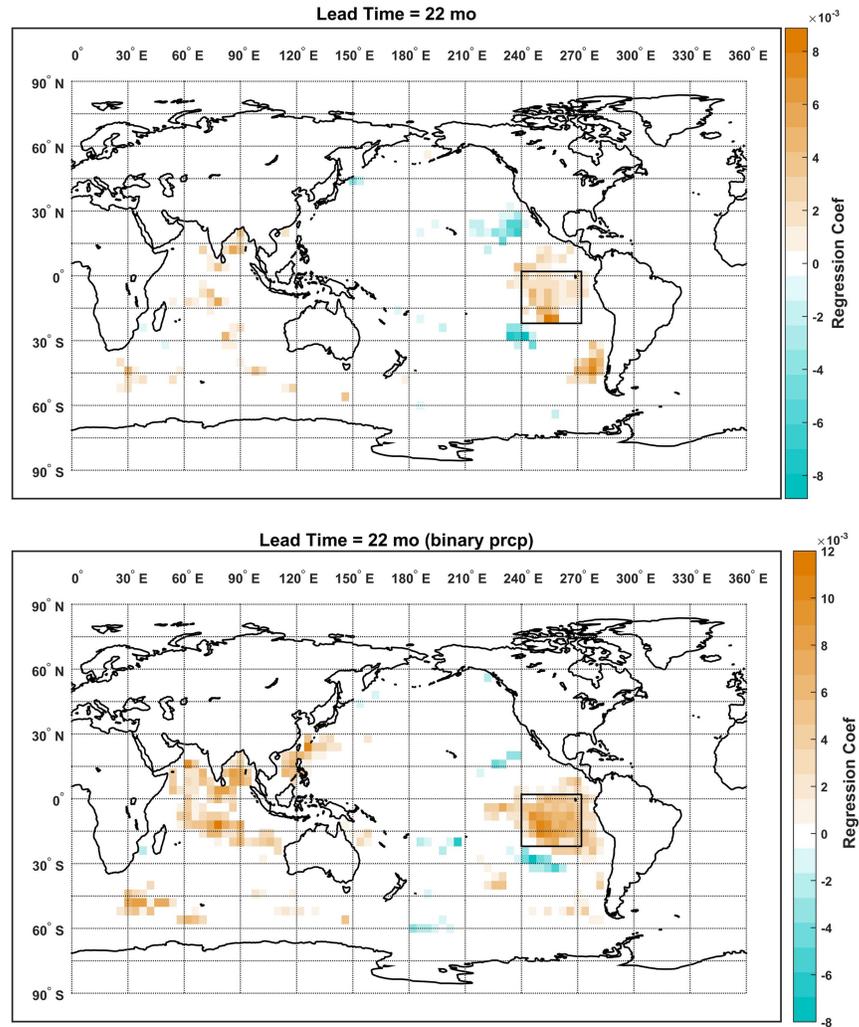


Figure 3.11: Maps of regression coefficients from the elastic net models using true-amplitude (top) and binary (bottom) predictand precipitation at the lead time of 22 months

also confirmed by the results that statistically significant correlations are found between the vectorized coefficient maps of the two models. The correlation coefficients are 0.45 for the lead time of 14 months and 0.44 for the lead time of 22 months (1343 samples), and are 0.35 for the lead time of 14 months (327 samples) and 0.40 for the lead time of 22 months (359 samples) when only non-zero coefficients are considered.

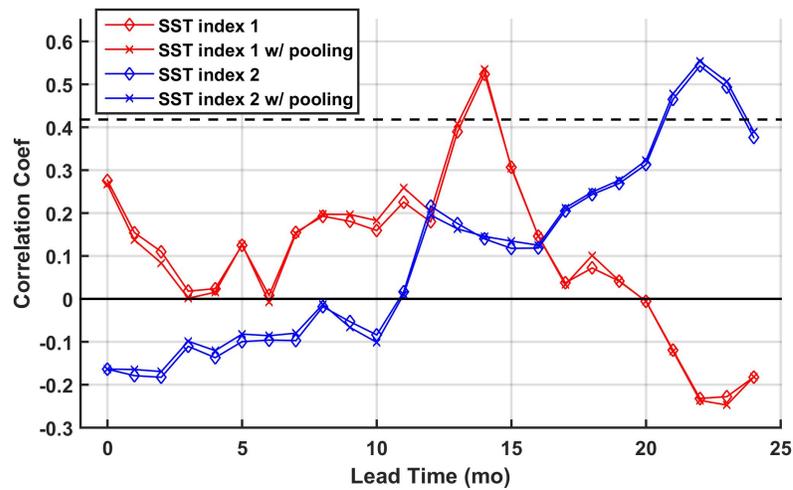


Figure 3.12: Lagged correlations between the TRHR precipitation and the SST indices. The p-value = 0.01 significance level is plotted in the black dashed line.

The comparably high predictive skill at the lead time of 14 months is easy to interpret if we compare the coefficient maps from Figure 3.10 and 3.11 to the correlation maps from Figure 3.7 and 3.8. Though less persistence is observed for the global correlation at the lead time of 14 months, the elastic net model managed to select the regional persistent cluster of positive correlations to the east of Australia. At last, two SST indices are defined based on the consistent patterns across the two regression models: the SST index 1 is defined by mean SST over the domain of [180E-200E, 42S-18S]; the SST index 2 is defined by mean SST over the domain of [240E-272E, 22S-2N]. Lagged correlations between the TRHR precipitation and the SST indices computed using non-pooled and pooled SSTs are plotted in Figure 12. Statistically significant positive correlations are observed at the corresponding lead times and highly consistent results are shown for non-pooled and pooled SSTs. The results suggest that the proposed framework managed to select certain regional SSTs that are persistently correlated with the predictand precipitation. Though correlation does

not necessarily imply causation, the elastic net regression models show good potential here in guiding further research with its highly interpretable and flexible models.

3.5 Conclusions

In this paper, we tested a generalized regression model with regularization coupled with pooling in predicting the TRHR wet-season precipitation at lead times of 0-24 months using the Pacific and Indian Ocean SSTs. The regression is first tested using true-amplitude predictand and is compared against some widely-used regression models including the OLS, CCA, and EOF regressions. Significantly good predictive skill are observed using the elastic net regression models for certain long lead times which are further examined using a correlation analysis. The results demonstrate that the elastic net regression achieves good performance in identifying and using consistent correlation patterns while the other three regression models show relatively poor performance. Low model skill at shorter lead times can be due to the fact that only SST is used as the sole predictor while teleconnection signals can propagate through other climate fields chronologically. A multinomial elastic net regression model is then used to demonstrate flexibility and consistency of the proposed framework. Consistent model skill and regression coefficient maps are observed even when predictand and deviance functions of different types are used. By comparing the lagged correlation maps and regression coefficient maps, we find that the elastic net models manage to select regional persistent correlation patterns as the contributing predictors. The other widely-used regression models are heavily based on global covariances either between the predictand and the predictors or

within the predictors, and thus are vulnerable to over-fitting. At last, two SST indices are defined based on the major clusters of non-zero coefficients from the elastic net models and are found to be significantly correlated with the TRHR precipitation at the corresponding lead times. Overall, the proposed framework demonstrates good interpretability in identifying covariates with high predictive skill and the potential in guiding further investigation using more complex, nonlinear statistical models or physically based modeling experiments.

CHAPTER 4

BAYESIAN SURVEILLANCE FOR RAINFALL-RUNOFF RESPONSES

4.1 Introduction

Skillful runoff forecast is of great importance in informing decision making in water resource management across various sectors including but not limited to hydropower plant operation [148], agriculture resource planning [149], and disaster mitigation [150, 151]. At long timescales (e.g., seasonal to inter-annual), runoff amplitudes are usually largely determined by rainfall amplitudes. But at short timescales (e.g., hourly to weekly), some key variables for defining high-frequency water disasters (e.g., flooding) like timings and rates of peak flow are also determined by the rainfall-runoff responses as controlled by the local hydrological properties [152, 153]. To understand dynamics of rainfall-runoff responses can aid the decision making of water managers by reducing uncertainties in the runoff forecasts/estimations [154, 155].

The past decades have witness significant advances in sensing technology, which are making massive volumes of earth observation data available at exceptionally fine resolutions in space and time [156, 157]. However, these observational data are highly concentrated in some easy-to-measure variables like rainfall and runoff. Though it is hard to directly measure the variable rainfall-runoff response or its hidden variables, such information can be inferred from observations of rainfall and runoff. Past studies already show that some parameters of physically-based hydrology models can reflect physical properties of watersheds like river morphology [158] or land use types [159]. These properties can determine the rainfall-runoff response and thus tracking estimations of

the corresponding parameters can help us learn dynamics of the local responses [160]. Based on this fact, the problem of learning non-stationary rainfall-runoff responses can be refined to estimating time-varying parameters of physics-informed hydrology models given observations of some easy-to-measure variables like rainfall and runoff.

Most methods designed for model calibration are suited for this job and they can be roughly categorized into two classes: 1) the Maximum Likelihood Estimation (MLE) method that searches for the optimal parameter values that maximize the likelihood function for a given set of observations [161, 162]; and 2) the Bayesian method (also often referred to as the Maximum a Posteriori (MAP) method) that estimates the probability distribution function (PDF) of parameters given observations [163, 164]. Either method has its own advantages and represents different perspectives of approaching a problem. MLE is often described as the frequentists' methodology while MAP is often described as the Bayesian's methodology [165]. For situations where information of prior distributions and sources of uncertainty are known, the Bayesian method is the preferable option since it can simultaneously estimate parameters and quantify the associated uncertainties [166]. In addition, for high-dimensional problems, the simulated PDF can converge to the true PDF at a controllable rate using Bayesian methods [167] while the computational cost can increase exponentially as the number of parameters increases when using MLE methods (e.g., grid search methods) [168]. These advantages make the Bayesian method suitable for inferring parameters of hydrology models here since most hydrology models use not multiple but many parameters [169].

Bayesian estimation models has already been widely used in calibrating hy-

drology models. Some early effort can date back to 1970s when Vicens et al. [170] attempted to reduce parameter uncertainties by using informative prior PDFs based on regional information. This work focused on fusing information from prior knowledge and sample observations and resembles a lot of studies in data assimilation and filtering (e.g., the famous Kalman filter by Kalman [171]). Since then, the Bayesian estimation method has gained rapidly increasing popularity in uncertainty quantification for hydrology model parameters. Yeh [24] described the parameter identification procedures in groundwater hydrology as an inverse problem and introduced the Bayesian solution using examples of the composite objective function [172] and the Kalman filter. Kitanidis [173] compared information fusion processes in a Bayesian analysis with different prior and likelihood distributions (e.g., when the normality assumption is not satisfied). Further sophisticated Bayesian methods are applied in tuning hydrology models. Bates and Campbell [163] tested a Markov Chain-Monte Carlo scheme in inferring parameters for a conceptual rainfall-runoff model. Moradkhani et al. [155] implemented the ensemble Kalman filter in iterative hierarchical scheme to simultaneous estimate hyperparameters and state variables given observations of rainfall and runoff. Thanks to the rapid development of computation technology, a lot more simulation-based techniques are developed and tested in learning hydrology model parameters like the particle filter and its derivatives [174, 175, 176]. While all these previous studies have validated the Bayesian estimation method in calibrating hydrology models, most of them are done in a context of data assimilation or uncertainty quantification. Even when Moradkhani et al. [155] managed to estimate time-varying parameters and the associated uncertainties, they highlighted their method as a potential tool to update hydrology simulations and decompose uncertainties from dif-

ferent sources. Here, we want to address the opportunity of using Bayesian estimation methods in learning dynamics of rainfall-runoff responses. And this can be quantitatively done by iteratively feeding the analysis with batches of rainfall and runoff observations collected from different historical periods or different watersheds.

In this paper, we propose to test an analytical framework for monitoring changes in hydrological responses by tracking estimations of hydrology model parameters given rainfall and runoff observations. The framework is based on a Bayesian estimation method and is specifically designed for small watersheds. For large watersheds, a lot of information would be lost by aggregating spatially highly-heterogeneous hydrological properties into a few parameters and this would lead to huge uncertainties in parameter estimations. The model is tested using both synthetic and observational data of rainfall and runoff, and is applied to investigate temporal and spatial variability of rainfall-runoff responses.

4.2 Data

The analytical framework is first validated using synthetic rainfall and runoff data generated from historical records of the Fall Creek watershed. Fall Creek is a fourth-order stream that flows through Ithaca, NY and drains to Cayuga Lake. The watershed of Fall Creek has a relatively small drainage area of 324 km² and consists of mixed forested and agricultural areas [177]. Like most other NY watersheds, the Fall Creek watershed is dominated by a saturation-excess process and runoff is usually only generated when and where soils are

highly saturated [178, 179]. The watershed is chosen for having relatively long records of ground-based precipitation and runoff observations at a daily timescale. Runoff data is collected from the National Water Information System (NWIS) of the United States Geological Survey (USGS) [180]. The runoff station (USGS 04234000 Fall Creek) is located at [42°27'12"N, 76°28'22"W] and has a continuous record from February 1925 to the current year. Rainfall data is collected from the Global Historical Climate Network-Daily (GHCN-D) database that integrates station-based observations from numerous sources in producing daily climate records with quality assurance [181, 182, 183]. The rainfall station (GHCND:USC00304174) is maintained by Cornell [42°26'57"N, 76°26'57"W] and has a continuous record from April 1925 to the current year. And data over 1925-2017 is used in the model validation part.

After the model validation, the framework is also tested on other NY watersheds. We try to examine the spatial variability of rainfall-runoff responses in this part and therefore, collect data for many small watersheds but over a shorter study period. 102 watersheds are selected and their information can be found in the Supplementary Materials. Daily runoff data is collected from the USGS with a temporal coverage of 1979-2006. Considering no rainfall stations can be found near some of the USGS stations, gridded rainfall is used as collected from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation over CONUS provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their website at <https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html>. Quality control is performed on gauge records from multiple sources over 30,000 stations to provide a suite of unified precipitation products in the CPC data project [184, 185, 186]. The CPC gridded rainfall has a spatial resolution of 0.25 ° longi-

tude by 0.25° latitude and for each USGS station, rainfall from the nearest grid cell is used.

Since we want to focus on monitoring changes in local hydrological responses at an inter-annual timescale in this study, the predominant effects of seasonality are removed by limiting each 'batch' to only one boreal summer period (June-July-August). And instead of separating individual storm events before fitting the hydrology model, data over the whole summer period is used which essentially averages the parameter estimations over multiple storm events. Each summer period can be adaptively extended by moving the starting date earlier so that effects of prior storm events will not be missed. The extension is limited to $[0, 10]$ days and stops when runoff at the starting date is smaller than some thresholds (e.g., the 50th percentile value of runoff over the whole summer period). A summer period is dropped if it includes missing values or has significantly high runoff rate at the starting date even after the extension procedure. This step is to guarantee consistency in lengths of summer periods and thus in comparing model performance across different periods. The main reason is because even at the same statistical significance levels, the absolute value of a model skill metric (e.g., the Pearson's correlation coefficient used in this study) can change when the sample size changes.

4.3 Methods

4.3.1 Synthetic Data Generator

A two-step synthetic data generator is developed here to 1) sample rainfall based on historical records and 2) to produce runoff with some given noise levels for a typical summer period (i.e., 92 days). The rainfall events are first assigned using a Markov chain that transitions between the rainy (R) and sunny (S) states. The transition probabilities are estimated based on the historical Fall Creek rainfall events during the selected summer periods and the transition matrix is shown in Table 4.1. The initial state follows a Bernoulli distribution estimated based on the rainfall data at the starting dates of the selected summer periods such that $P(R_0) = 0.49$ and $P(S_0) = 0.51$.

Table 4.1: The transition matrix estimated using historical rainfall events during the selected summer periods for the Fall Creek watershed

Current \ Previous	Sunny	Rainy
Sunny	0.66	0.51
Rainy	0.34	0.49

After generating a time series of rainfall ‘indicators’ that determine the rainy-sunny state on each day, the intensities are simply sampled from historical rainfall intensities with replacements (i.e., bootstrapping). Assume the local rainfall-runoff response is governed by some model denoted by \mathcal{M} such that $\tilde{R} = \mathcal{M}(P; \Theta)$. \tilde{R} is the resulting runoff and Θ is a set of hidden parameters that are determined by physical properties and modulate the hydrological response.

Then the noisy synthetic runoff data (R_t) can be generated by

$$R_t = \mathcal{M}(P_t; \Theta_t) + \varepsilon_t \quad (4.1)$$

A time-invariant additive Gaussian noise is assumed such that $\varepsilon_t \sim N(0, \sigma^2)$. This simplifies the comparison of model skill with different noise levels as well as computations in the Bayesian estimation model (which will be detailed in the section 3.2). We will explore other options of noises in future work but such assumption has already been widely used in previous studies [163, 155, 187]. And the specific form of the hydrology model \mathcal{M} used in this study will be described in section 3.3.

4.3.2 Bayesian Estimation Model

We denote observations of rainfall P_t and runoff R_t at time t by D_t such that $D_t = (P_t, R_t)$. Assuming the local rainfall-runoff response is controlled by some set of hidden parameters Θ that cannot be easily measured, the goal of this study is to infer Θ given some observations of D (i.e., $P(\Theta|D)$). Since it is hard to directly simulate $P(\Theta|D)$, the Bayes' Theorem is adopted as given by

$$P(\Theta|D) = \frac{P(D|\Theta) \cdot P(\Theta)}{P(D)} \quad (4.2)$$

The target PDF (often referred to as the posterior PDF) $P(\Theta|D)$ can then be estimated by calculating the likelihood PDF $P(D|\Theta)$, the prior PDF $P(\Theta)$, and the evidence PDF $P(D)$. The evidence PDF $P(D)$ can be regarded as a rescaling factor to make the resulting posterior PDF sum to one. Using the example of

a continuous random variable of Θ , $P(D)$ can be solved by integrating over all possibilities of Θ as given by

$$P(D) = \int P(D \cap \alpha) d\alpha = \int P(D|\alpha) \cdot P(\alpha) d\alpha \quad (4.3)$$

Analytical solutions for the integral in Equation 3 are only available for some specific combinations of prior and likelihood functions [188, 189]. To numerically estimate the integral is always computationally intensive and even impossible for high dimensional random variables. Instead of directly solving the Equation 2, we propose to simulate the posterior PDF using a MCMC approach [190], which avoid computing the evidence PDF. The classic Metropolis Hastings algorithm [191] is adopted here and consists of 3 major steps for constructing a Markov process of which the stationary distribution (π_∞) converges to the target posterior PDF.

1. Choose a transition kernel (also sometimes referred to as a proposal distribution) $q(y|x)$ such that $q(y|x) > 0$ for all $x, y \in \chi$ (χ is the parameter space).
2. Define an acceptance rate $\alpha_{x,y}$ as given by

$$\alpha_{x,y} = \min\left(1, \frac{\pi_\infty(y)q(x|y)}{\pi_\infty(x)q(y|x)}\right) \quad (4.4)$$

3. Repeatedly simulate the Markov process: at time k , given state $x_k = x$
 - i) simulate next state of the Markov process $y \sim q(y|x)$,
 - ii) generate u from an uniform distribution $u \sim U[0, 1]$,
 - iii) if $u < \alpha_{x,y}$, set $x_{k+1} = y$; otherwise, set $x_{k+1} = x$.

If we substitute the stationary distributions in Equation 3 with the posterior distributions, the acceptance rate is then computed by

$$\alpha_{x,y} = \min\left(1, \frac{P(D|y)P(y)q(x|y)}{P(D|x)P(x)q(y|x)}\right) \quad (4.5)$$

The transition kernel essentially determines how fast the Markov process explores unknown parameter space. A variety of transition kernels have been developed and tested for improving efficiency of the MCMC algorithm [192, 193]. However, since our intent is to test feasibility of the Bayesian estimation model in tracking changes in hydrological responses rather than develop the most efficient algorithm, based on the principle of indifference [194], the commonly used uniform distribution is assumed for the transition kernel $q(y|x)$ such that $y \sim U[x - \Delta x, x + \Delta x]$. Assuming we have no prior knowledge of the watersheds, a uniform distribution is also assumed for the prior distribution $P(\Theta)$ such that $\Theta \sim U[\Theta_{min}, \Theta_{max}]$. The final acceptance rate can then be computed by

$$\alpha_{x,y} = \min\left(1, \frac{P(D|y)}{P(D|x)} \cdot \mathbf{1}_{[\Theta_{min}, \Theta_{max}]}(y)\right) \quad (4.6)$$

where $\mathbf{1}_A(y)$ is an indicator function which returns 1 if $y \in A$ and 0 if not. Compared to the posterior distribution $P(\Theta|D)$, the likelihood distribution $P(D|\Theta)$ is usually easier to estimate when the governing model is already known. However, the simulated Markov process can be trapped by local optima when using the uniform transition kernel with small step sizes [195], and this effect is particularly significant when simulating high dimensional random variables. We propose to initialize the Markov process with a searching procedure over some coarse grids in the range of $[\Theta_{min}, \Theta_{max}]$. The size of a coarse grid is set to $\frac{\Theta_{max} - \Theta_{min}}{10}$

and thus the total number of grids is 10^p for a p -dimensional parameter variable. $P(D|\Theta)$ is estimated using center parameter values of each coarse grid cell and the parameter value that maximizes $P(D|\Theta)$ is used to initiate the Markov process.

4.3.3 Instantaneous Unit Hydrograph Model

The Instantaneous Unit Hydrograph (IUH) model is a widely used conceptual model for runoff estimation and has a long history of successful implementation in hydrology [196, 197, 198, 199]. And it is well-established that IUHs can reflect some physical properties of a watershed like geomorphology characteristics [196, 198, 200]. This fact suggests potential of monitoring changes in local rainfall-runoff responses by tracking changes of IUHs. To study dynamics of IUHs in a quantitative way, here we propose to use a parameterized form as given by

$$h(t; \lambda, k, \theta) = \lambda \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-t/\theta} \quad (4.7)$$

The formula is essentially a rescaled Gamma distribution: λ is a scaling factor that accounts for loss of water in runoff generation processes (e.g., evapotranspiration); and (k, θ) are shaping parameters for the Gamma distribution. The Gamma distribution is used for its flexibility in modeling different types of IUHs and capability of simulating the low-pass filtering effect [201, 202]. Substituting the parameterized IUH into \mathcal{M} in Equation 1, we can approximate the noisy runoff by

$$R(t) = \tilde{R}(t) + \varepsilon_t = \sum_0^T P(t - \tau)h(\tau; \lambda, k, \theta) + \varepsilon_t \quad (4.8)$$

where the resulting runoff \tilde{R} is modeled as the convolution of rainfall and the IUH ($\tilde{R}(t) = P(t) \otimes h(t; \Theta)$). T is set to 14 days in this study. The probability of some given runoff observations conditioned on a set of parameters (λ, k, θ) can then be easily modeled using a Gaussian distribution ($R(t) \sim N(\tilde{R}(t), \sigma^2)$) and the acceptance rate in Equation 6 can be computed by

$$\alpha_{x,y} = \min \left(1, \exp \left(\frac{-1}{2\sigma^2} \sum_{t=1}^N (R(t) - \tilde{R}(t; h(y)))^2 - \frac{-1}{2\sigma^2} \sum_{t=1}^N (R(t) - \tilde{R}(t; h(x)))^2 \right) \cdot \mathbf{1}_{[\Theta_{min}, \Theta_{max}]}(y) \right) \quad (4.9)$$

The hyperparameter σ^2 can be approximated by the mean squared error (MSE) using the MLE parameter estimations [187] as given by

$$\sigma^2 = \frac{\sum_{t=1}^N (R(t) - \tilde{R}^{MLE}(t))^2}{N} \quad (4.10)$$

A physical interpretation of Equation 10 is that if we assume observations are governed by the proposed model, then MSE is minimized using the MLE parameters and should represent the true noise level. In this study, we denote the mean values of parameter samples from MCMC by the Bayesian parameter estimations. We replace the MLE parameter estimations with the Bayesian parameter estimations and Equation 10 is implemented in an iterative way:

1. The PDF of $P(\Theta|D)$ is simulated using MCMC with some arbitrarily assigned constant for σ^2 .

2. \tilde{R}^{MLE} is approximated using the Bayesian parameter estimations and σ^2 is updated.

3. The MCMC is repeated using updated σ^2 and step 2-3 are repeated.

For simplicity, only one iteration is done and a prelim analysis shows that one iteration can already guarantee convergence for the approximated σ^2 ($R^2 = 0.998$ for 65 samples in comparing the approximated σ^2 s after one and two iterations).

4.4 Results & Discussions

4.4.1 Model Validation

Synthetic Experiment

Our analytical framework is first tested using synthetic data of summer rainfall and runoff. For each summer period, the IUH parameters are assumed constant and are drawn from an uniform prior distribution with ranges shown in Table 4.2. These ranges are arbitrarily assigned to cover a broad parameter space and the same prior distribution is also used in the following analysis.

Table 4.2: Ranges of IUH parameter values [$\Theta_{min}, \Theta_{max}$]

Parameters	λ	k	θ
Ranges	[0, 0.6]	[0, 6]	[0, 10]

Model skill is evaluated using relative errors which are defined as ratios of absolute errors of parameters and the corresponding ranges ($err(\Theta) = \Delta\Theta/$

$(\Theta_{max} - \Theta_{min})$). And the model is tested with noises of different levels. To facilitate the comparison of synthetic experiments and case studies using observational data, we use a relative metric of the signal-to-noise ratio (SNR) here rather than the absolute noise amplitudes. And SNR is simply defined as a ratio of the standard deviation of noisy runoff to the standard deviation of noise ($SNR = \sigma(R)/\sigma(\varepsilon)$). Figure 4.1 shows model skill as a function of SNR for three IUH parameters (λ, k, θ). At each SNR , the synthetic experiment is repeated for 50 times. The 25th – 75th percentile ranges are plotted as shaded areas and the median model skill are plotted in diamond lines. A comparison between model skill of the initial grid searching (blue) and the Bayesian estimation model (red) demonstrates that the MCMC can always improve accuracy of parameter estimations when SNR is greater than 1. And when the noise is comparable to even larger than the signal (i.e., $SNR < 1$), performance of both method degrades significantly in terms of median relative errors as well as uncertainty ranges. A reference SNR level is approximated using the ratio of standard deviation of noisy runoff observations to that of runoff residuals. The runoff residual is calculated by fitting the IUH model with Bayesian parameter estimations using the observational data from the Fall Creek watershed. The smallest value of SNR of all selected summer periods is also plotted in Figure 4.1 and the results indicate a good potential performance when applying our Bayesian estimation model on real-world observational data. For any SNR greater than the reference level, the median relative errors should be smaller than 6% for all parameter estimations (or smaller than 3% for (λ, k)).

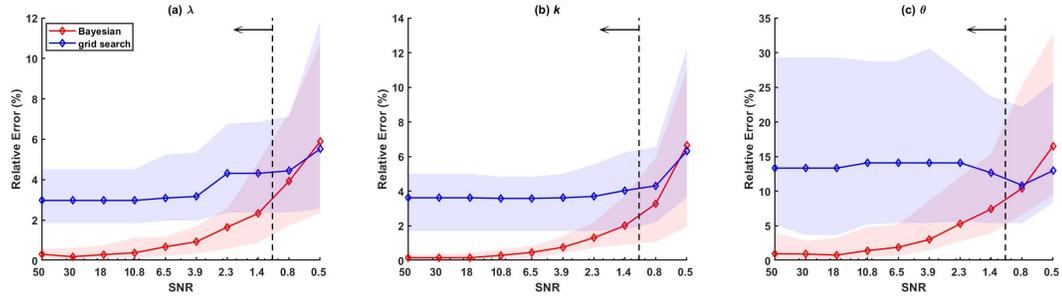


Figure 4.1: Model skill measured by relative errors as a function of SNR for (a) λ , (b) k , and (c) θ . Performance of the initial grid searching and the Bayesian estimation model are plotted in blue and red lines, respectively. At each SNR , the synthetic experiment is repeated for 50 times. The 25th – 75th percentile ranges are plotted as shaded areas and the median model skill are plotted in diamond lines. The reference SNR based on observational data from the Fall Creek watershed is indicated by a vertical black dashed line. And observational data from the Fall Creek watershed have SNR values no smaller than this reference level for all selected summer periods as indicated by the leftward arrow.

Validation Using Observational Data

In addition to using synthetic data, we also examine the model performance using observational data of rainfall and runoff from the Fall Creek watershed. 65 summer periods are selected over the whole period of 1925-2017 and IUH parameters are estimated for each summer period. Since true parameter values are not known here, MLE parameter estimations are used in the comparison. The MLE parameter estimations are computed by randomly searching in the parameter space $[\Theta_{min}, \Theta_{max}]$ to minimize MSE between the observed and estimated runoff and should represent the best possible parameter values given our conceptual IUH model.

We first fit the IUH models using both MLE and Bayesian parameter estimations and their performance in estimating runoff are compared in Figure 4.2.

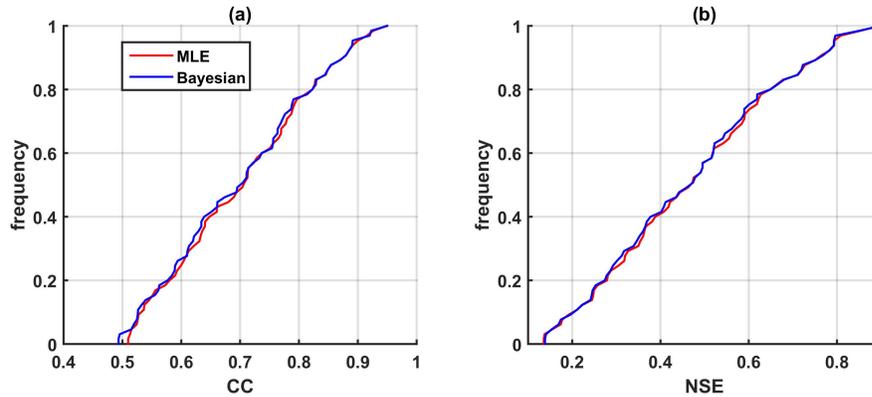


Figure 4.2: Empirical cumulative distribution functions of model skill measured by (a) Pearson's correlation coefficients (CC) and (b) Nash-Sutcliffe Efficiency (NSE) scores for runoff estimations. Performance of models fitted using MLE and Bayesian parameter estimations are plotted in the red and blue curves, respectively.

Very negligible difference are observed in model skills using MLE and Bayesian parameters as measured by both CCs and NSE scores. And comparable good model skill are observed for runoff estimations using both sets of parameters: the median CCs are greater than 0.7 and the median NSE scores are greater than 0.49. The good model skill should also validate the Gamma distribution-based IUH in modeling the rainfall-runoff response in the Fall Creek watershed. No statistical significance levels are reported in the figure since the effective degrees of freedom depend on autocorrelations in the data [203] and vary significantly across different summer periods (considering autocorrelations in runoff/rainfall are largely affected by frequencies and intensities of storm events.)

The Bayesian parameter estimations are then compared against their MLE counterparts as shown in Figure 4.3. Great consistency are observed between the MLE and Bayesian parameter estimations: the R^2 values calculated using the

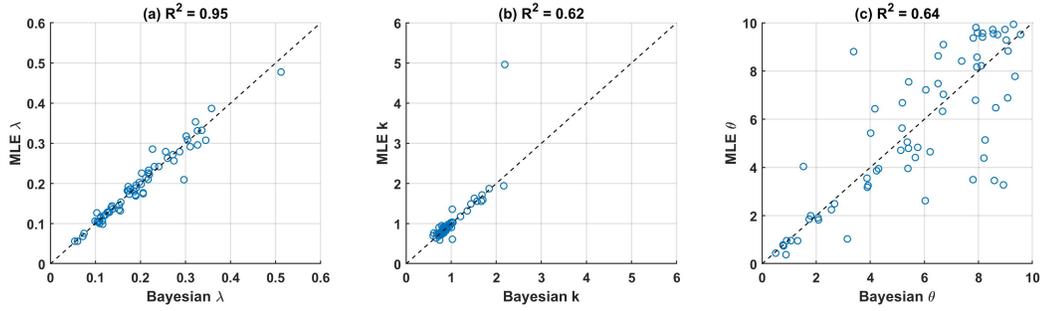


Figure 4.3: Comparison between the Bayesian (x axis) and MLE (y axis) parameter estimations for (a) λ , (b) k , and (c) θ . The 1:1 reference lines are plotted in black dashed lines and the coefficients of determination (R^2) for 65 summer periods are reported in the subplot titles.

1:1 reference line are 0.95, 0.62, and 0.64 for (λ, k, θ) , respectively. Of the three parameters, the highest R^2 value is found for λ . A possible explanation can be that MSE in calculating $\alpha_{x,y}$ (in Equation 9) is more 'sensitive' to λ since it solely determines the integral amplitude of the IUH while the shape parameters (k, θ) only collectively determine the IUH shape. For example, different combinations of (k, θ) can result in very similar IUH shapes and therefore, more deviations between the MLE and Bayesian estimations are expected for (k, θ) .

Also, the Bayesian estimations of λ are found to be consistent with the runoff coefficients approximated using the ratio of total amount of runoff to that of rainfall. The R^2 value calculated using the 1:1 reference line is 0.65 and the correlation coefficient is 0.87 for 65 summer periods (one tailed p-value < 0.001). On one hand, this should validate the Gamma distribution-based IUH in simulating the local rainfall-runoff response since λ is specifically designed to account for the loss of water. On the other hand, the approximated runoff coefficients can be used as the true λ values and thus the results suggest that the Bayesian estimation model can well infer and track the time-varying hidden

parameters.

4.4.2 Model Application

A Historical Change Point in the Fall Creek Watershed

The analytical framework is then used to study dynamics of rainfall-runoff responses in the Fall Creek watershed through 1925-2017. Since it is hard to compare IUHs of different shapes in a quantitative way, we first define a metric of the initial decay rate (*IDR*) as given by

$$IDR = \frac{\partial h / \partial t}{h} \Big|_{t=0.05} = \left(\frac{k-1}{t} - \frac{1}{\theta} \right) \Big|_{t=0.05} \quad (4.11)$$

IDR is essentially the relative initial derivative of IUH and is estimated at $t = 0.05$ [day] to avoid a zero denominator. Using *IDR*, we can easily distinguish two typical IUHs: the IUHs that monotonically decreases over time are defined as the 'diffusion' type while those with wave-like patterns are defined as the 'advection' type. A diagram of the two IUH types is shown in Figure 4.4 and an IUH with greater-than-zero *IDR* falls into the category of 'advection' and 'diffusion' if otherwise. The two types can have direct use in informing flooding mitigation since they determine timings and rates of peak flow differently. Also the two IUH types can reflect different water travel times and therefore indicate different hydrological properties.

Time series of Bayesian parameter estimations are plotted in Figure 4.5. Systematic shifts are observed around 1943 for the shape parameter

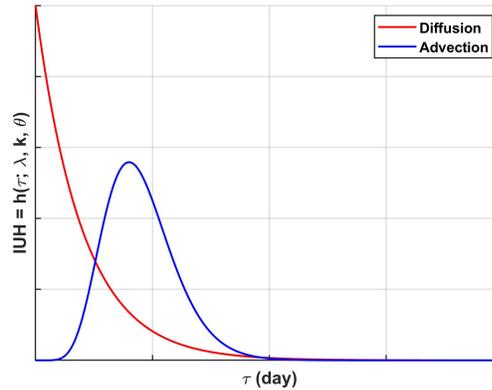


Figure 4.4: A diagram of two types of IUHs: the 'diffusion' type (red) and the 'advection' type (blue).

(k, θ) : values of k drop significantly and values of θ increase by a smaller margin after around 1943. No such change point is found for the rescaling parameter λ . The decreasing k can turn IUHs from the 'advection' type to the 'diffusion' type while the increasing θ has the opposite effect. We find that historical patterns of IUHs are dominated by changes in k when we compare time series of *IDR* with those of summer rainfall and runoff in Figure 4.6. Before 1943, the local IUH was dominated by the 'advection' type and after 1943, the local IUH switched to the 'diffusion' type, which is consistent with the rapid drop of k values. However, such change cannot be learned by looking at time series of summer rainfall and runoff. And this suggests that the change could occur only in the temporal distribution of water. Physically, this change point can indicate that water in the Fall Creek watershed flows faster (i.e., has shorter travel time) after 1943, which can possibly be due to changed in land uses like construction of paved roads. However, we could not find any strong support evidence from chronicles of the Fall Creek watershed or the Ithaca city so far. Interestingly, the change point coincided with relocation of the rainfall station when Cornell moved it from the Roberts Hall (on the ground outside of the building)

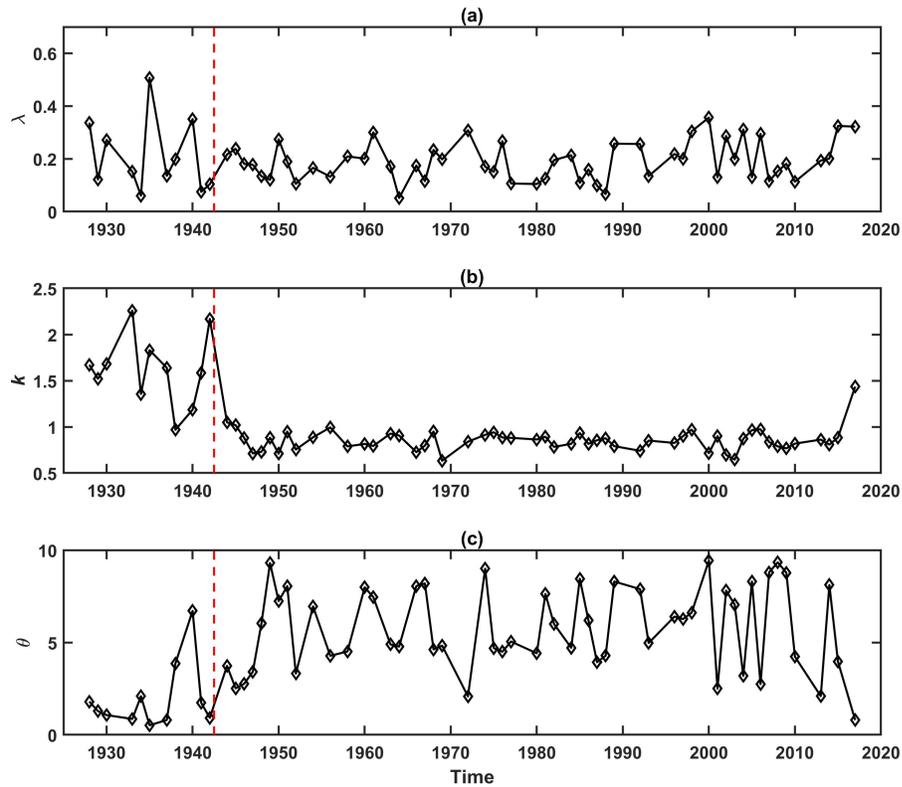


Figure 4.5: Time series of Bayesian estimations of (a) λ , (b) k , and (c) θ . The year of 1943 when Cornell moved its rainfall station is indicated by a vertical red dashed line.

[42°26'55"N, 76°28'45"W] to the College of Agriculture Experimental Farm at the Caldwell Field [42°26'56"N, 76°27'38"W] in June 1943. While the relocation of rainfall station can lead to changes in latency between rainfall and runoff observations, we expect the effect to be minor since 1) the rainfall stations was only move one mile east in 1943 (for reference, a hurricane typically travels at a forward speed of 5-20 miles per hour [204]) and 2) the station was moved again to its current location on the Game Farm Road [42°26'57"N, 76°26'57"W] in June 1969. We did not observe any changes in IUH parameters associated with the second relocation. But still, the Bayesian estimation model manages to identify

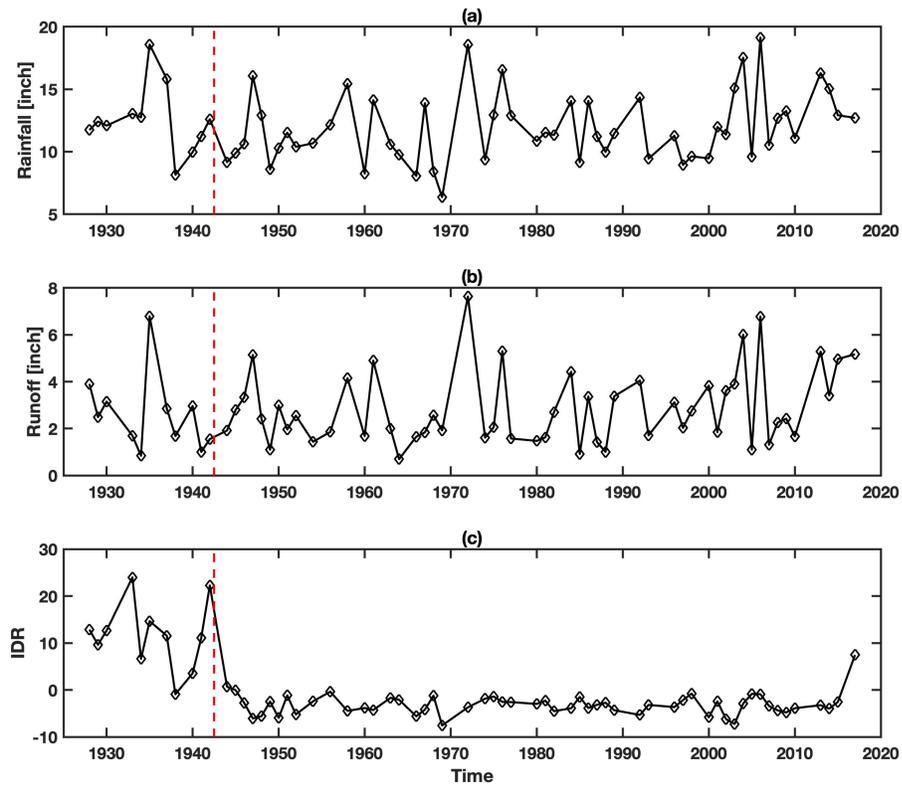


Figure 4.6: Time series of (a) total summer rainfall, (b) total summer runoff, and (c) *IDR*. The year of 1943 when Cornell moved its rainfall station is indicated by a vertical red dashed line.

a systematic shift in the local rainfall-runoff response that cannot be learned by looking at the easy-to-measure variables like rainfall or runoff.

Relationships between IUHs and Watershed Sizes

In the previous section, the Bayesian estimation model is applied to examine temporal variability of rainfall-runoff responses. In this section, we extend its application to examining rainfall-runoff responses of many watersheds by repeating the analysis for 102 NY watersheds. Considering a relatively short

study period is used (1979-2006) here, the hydrological responses are assumed constant and the median *IDRs* for each watershed are compared to investigate the spatial patterns. A summer period is only selected when the NSE score is better than 0.2 for the runoff estimation using the Bayesian parameters. And only watershed with more than 5 selected summer periods are considered. This is because the IUH cannot represent the local rainfall-runoff response when it fails at estimating the runoff.

Figure 4.7a shows a map of median *IDRs* for all selected watersheds (or watershed with 'good' NSE scores). While more positive *IDR* values are found in the northwestern region, this pattern can be due to a non-uniform distribution of watersheds of different sizes. To examine relationships between rainfall-runoff responses and watershed sizes, we first define a watershed length scale simply using the square root of the watershed drainage area. A scatter plot between the median *IDRs* and watershed length scales is shown in Figure 4.7b and a positive relationship is observed. A smaller watershed is more likely to be governed by the 'diffusion' type of IUH ($IDR < 0$) while a larger watershed is more likely to be governed by the 'advection' type of IUH ($IDR > 0$). The positive relationship is found to be statistically significant as examined by a F test (one-tailed p-value < 0.01). The results here appear to be intuitive since water usually needs longer travel time to reach the outlet in a larger watershed. But the underlying physical causalities require further rigorous investigation with runoff generation processes and channel network properties taken into consideration.

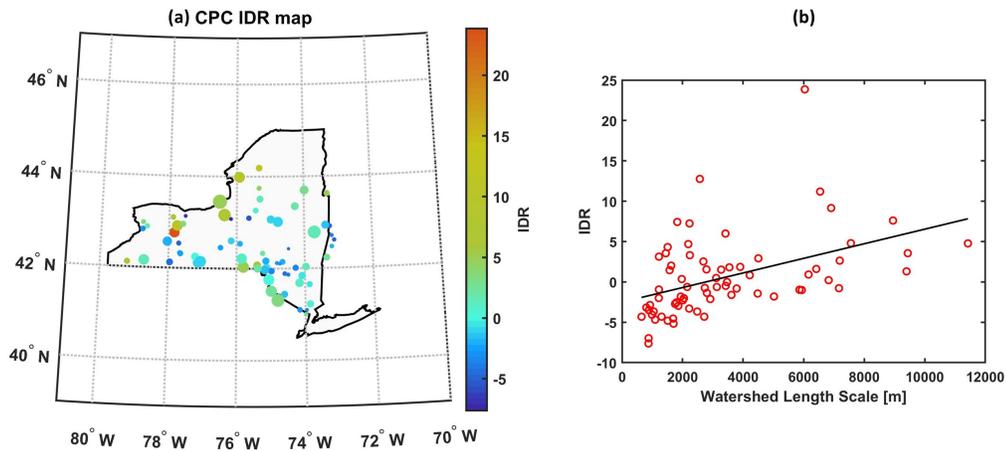


Figure 4.7: (a) A map of median *IDRs* for watersheds with 'good' NSE scores where sizes of circles are proportional to watershed length scales. (b) A scatter plot between watershed length scales and median *IDRs* and its fitted line (black).

4.5 Conclusions

In this study, performance of a Bayesian estimation model (i.e., MCMC) is examined in inferring hydrology model parameters given observations of rainfall and runoff for small watersheds. The IUH conceptual model based on a modified Gamma distribution is adopted and the whole analytical framework is validated using both synthetic and observational data. In the synthetic simulation, great accuracy is observed for Bayesian parameter estimations with relative errors smaller than 6% when the noise is smaller than the reference level. In absence of true parameter values, the Bayesian parameter estimations are compared against the MLE parameter estimations when observational data is used. Comparable model skill are observed for runoff estimations using MLE and Bayesian parameter estimations and good consistency is observed for all parameters between the two sets. Then the Bayesian model is applied to study temporal and spatial variability of rainfall-runoff responses by tracking param-

eter estimations over time or different watersheds. A systematic shift is identified for the Fall Creek watershed when its IUH switched from the 'advection' type to the 'diffusion' type around 1943. This change point cannot be learned by looking at the easy-to-measure variables like summer rainfall and runoff. The underlying cause remains unknown considering the shift can be due to either or both effects of changes in land uses and the relocation of the rainfall station. Furthermore, rainfall-runoff responses of 102 NY watersheds are studied but over a shorter study period. A statistically significant positive relationship is observed between *IDRs* and watershed sizes, which suggests that smaller watersheds are more likely to be governed by IUHs of the 'diffusion' type. Overall, our study demonstrates feasibility of the Bayesian estimation model in monitoring dynamics of hydrological responses and detecting change points for small watersheds by tracking the hidden variables (i.e., model parameters). And the model shows potential as a rapid surveillance tool to aid adapting local strategies of water resource management and disaster mitigation to changes of rainfall-runoff responses by iteratively updating the analytical framework with new batches of observational data.

CHAPTER 5
IMPACTS OF INFORMATION RESOLUTIONS ON RESERVOIR
OPERATIONS

5.1 Introduction

Reservoirs are man-made filters that modulate hydrological variables like streamflow to meet multi-sector demands and its operation plays an essential role in water resource management. Advantages of multi-purpose reservoirs were first addressed by the engineering profession in early 1900s [205] and reservoir operation optimization has since grown to be one of the most studied subject in the water resource community [206]. Most early studies focus on maximizing the economic utility of reservoir like hydropower production [207, 208, 209]. And correspondingly, the traditional optimization approaches were developed for single objective formulations and are limited in scalability to resolving multi- or many-objective problems [210]. In the recent decades, a broader range of issues have been taken into consideration when optimizing reservoir operations [211] and some examples include ecological requirements [212], biodiversity conservation [213], and climate impacts [214]. The changing preferences of water managers are adding importance onto the problem of reservoir operation optimization in many disciplines other than water resource management. And considering an emerging global trend of carbon neutrality, operation optimization for hydroelectric reservoirs is expected to gain more importance in the future.

The problem is intrinsically stochastic since most reservoir systems rely on uncertain hydroclimate forecasts in decision making. These hydroclimate vari-

ables like precipitation and streamflow can have numerous sources of uncertainty ranging from large-scale atmospheric circulations [215] to local changes of land use [216], which make them highly variable in time and space. According to the 2021 Global Energy Review by the International Energy Agency (IEA) [217], hydropower is expected to remain the world's largest source of uncertainty in renewable energy production.

And the problem is now being made more challenging than ever due to greater uncertainty in forcing and increased model complexity. The Intergovernmental Panel on Climate Change (IPCC) just released its sixth assessment report (AR6) in which human-induced climate change is found to be affecting many weather and climate extremes across the globe and the scale of recent climate changes is described as 'unprecedented' [31]. This indicates even further amplified uncertainty in hydrological forcings on which reservoir operation plans are directly based for 1) uncertainty in key climate variables like temperature and precipitation can directly propagate to uncertainty in hydrological variables like runoff and streamflow [218, 219, 220] and 2) regional hydrological properties (e.g., precipitation-runoff response) can be altered by climate changes and weather extremes [221, 222, 223]. In the meantime, reservoirs are being designed and operated for multiple purposes other than solely hydropower production (e.g., flood control, irrigation, and societal and ecological services) and networks of multiple inter-connected reservoirs are being coordinated at the same time for large water systems [224, 225, 226], which adds to the complexity of reservoir operation by increasing dimensions of objectives as well as decision variables.

Advances in sensing technology and machine learning have provided new opportunities in informing decision makings for reservoir operation by narrowing

down uncertainties in key hydroclimate forcings like precipitation [227, 228] and streamflow [229]. However, such innovations usually impose high requirements on computational performance [230] and expertise in hyperparameter selection [231]. These conditions are much harder to satisfy in less developed areas where proliferation of large hydroelectric dams have been seen in the past decades and the challenge of reservoir operation optimization is particularly relevant [232]. In addition, these areas usually have very limited ground stations and must rely on remote sensing data for measuring key hydroclimate forcings. But situations can be made worse by misinterpretation of remote sensing data due to scarcity of station-based data and lack of ground validation [233]. Motivated by this imbalance between the increasing problem complexity and uncertainty and the lack of access to highly accurate observations/forecasts, we conduct an information sensitivity analysis here to inform this question: how can finer-resolution forcing data potentially improve performance of reservoir operation optimization?

Perfect forecasts of hydroclimate forcings are not possible considering the intrinsic uncertainties in atmospheric circulations and hydrological processes. Biased models have been developed that sacrifice model accuracy for improved robustness [76] and multinomial models can be a good alternative option in developing robust hydroclimate forecasts [234, 235]. In this study, we propose to manage a complex water system with the two extreme cases of multinomial forecasts, the perfect or true-amplitude flow data versus binary flow data. But it should be noted that the whole analytical framework is easy to generalize for comparing forcing data of different types of uncertainties.

The upper Yellow River multi-reservoir system is selected for its critical role in

providing many essential services. The system consists of 12 reservoirs with a total installed capacity of 12.6 GWh [2] and is one of the major electricity suppliers for China's Northwest Grid. The Longyangxia hydropower plant of the system is integrated with world's 11th largest photovoltaic plants (the Longyangxia Dam Solar Park) and its hydropower production is directly used in load balancing for solar energy [236]. In addition to supplying water to riparian cities for agricultural, industrial and domestic uses, releases from the system also greatly affect water supply for two major irrigation districts of China, Qingtongxia Irrigation District and Hetao Irrigation District [237]. Part of the upper Yellow River basin is under ecological regulations of the Three-Rivers Headwater National Nature Reserve [238]. Other functions include flood and ice control [239]. These diverse functions make operation optimization for the system a complex and complicated problem with many conflicting objectives. Also, the area is experiencing significant changes in climate (e.g. decreasing wet-season precipitation and rising air temperature) and land use [240, 241, 141] which can amplify uncertainty in hydroclimate forcing. And what makes acquisition of fine-resolution forcing data harder is the fact that there are only sparsely distributed ground stations in this area under management of the China Meteorological Administration [242]. All these factors make the upper Yellow River reservoir system a suitable case for investigating impacts of information resolution in informing operations of large-scale hydropower systems.

Past studies have majorly used aggregated objective formulations in optimizing operations for the upper Yellow River reservoir system which can be roughly divided into two categories: 1) to transform some objectives into constraints to reduce dimension of the objective space [239, 2] and 2) to use a weighted combination of individual objectives (and constraints) as the performance indi-

cator [243, 244]. Most aggregated objective formulations can help alleviate the 'curse of dimensionality' [245] by reducing space of objectives to explore so as to satisfy computational requirements [246]. However, key assumptions made in these formulations like relative importance of competing objectives or selections of objectives to be transformed (into constraints) are largely determined by preferences of the person who formulates the problem. Such prior knowledge can lead to severe decision biases [247] and this risk is even further amplified considering decision makers' preferences are being reshaped by the increasing environmental awareness these days [248].

The Multi-Objective Direct Policy Search (EMODPS) framework developed by Giuliani et al. [249] not only overcomes the issues of aggregated objective formulations, but also enables quantitative comparisons of policies developed using different input data under climate uncertainty. The EMODPS framework optimizes parameters of closed loop nonlinear control policies (Direct Policy Search, or DPS policies) in determining water system operations using a Multi-Objective Evolutionary Algorithm (MOEA). The DPS control strategy approximates parameterized policies conditioned on a set of input variables [250, 251]. Compared to open loop inter-temporal strategies, DPS can develop adaptive operating policies that have good interpretability and can be easily coupled with a MOEA for optimization [249, 252]. While most past studies focus on optimizing water release decisions given some specific inflow conditions (usually the historic inflow) [239, 244, 2], DPS is more flexible and can guide operations under out-of-sample flow conditions since its policies are trained under a set of uncertain states of the world (SOWs) [251]. The Borg-MOEA proposed by Hadka and Reed [253] is used as the multi-objective optimization algorithm in this study for its enhanced effectiveness and efficiency due to its multi-operator adaptiv-

ity [254]. Examples of successful implementations of Borg-MOEA in optimizing water system operations can be seen in the diagnostic assessment by Reed et al. [255].

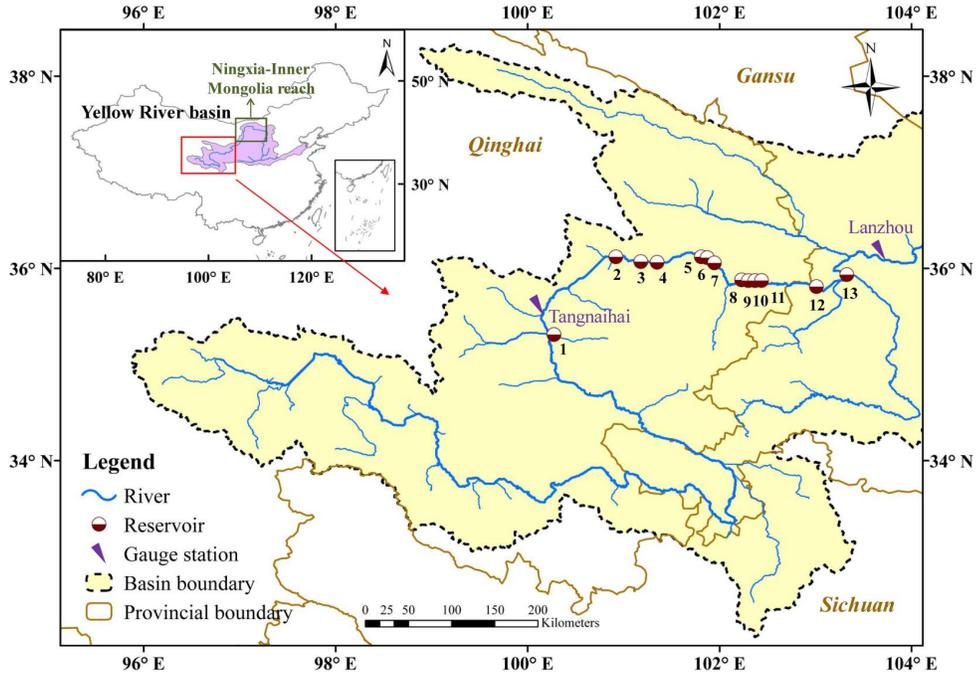
Other than expectation objectives, policy robustness and sensitivities are also examined as important complementary metrics of policy performance. Robustness measures insensitivity of policies across a variety of futures or system characteristics [256] and robustness decision making (RDM) has become increasingly popular in managing water systems [257, 258, 259]. Here, the changing climates are simulated using altered flow distributions and policy robustness are compared when trained with flow data of different resolutions. While the DPS control strategy can generate nonlinear, adaptive operating policies, real-world water managers favor more simple, static rules. To enhance policy implementability, we attempt to further understand how our analytical framework coordinate variability of different hydroclimate variables. To this end, building on Quinn et al. [252], a time-varying sensitivity analysis (TVSA) is performed to understand how the reservoir operations are determined differently when when data availability changes.

5.2 Case Study Description

5.2.1 Upper Yellow River Multi-Reservoir System

The upper Yellow River basin is dominated by a semi-arid monsoon climate [260] and about 90% of annual precipitation fall during a 6-month rainy season of May-October [261]. As response to the water scarcity since 1960s [262], water

a) Upper Yellow River basin map



b) A simplified water budget model

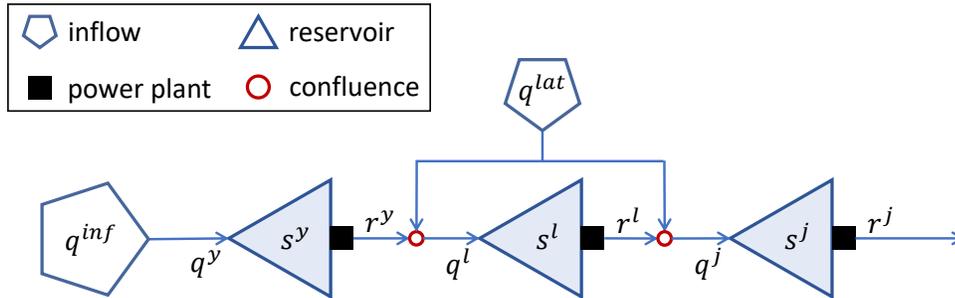


Figure 5.1: (a) A map of the upper Yellow River basin [2]. The upper Yellow River multi-reservoir system consists of 12 reservoirs (Index 2-13) with the two largest reservoirs being Longyangxia reservoir (Index 2) and Liujiaxia reservoir (Index 13). (b) A schematic of the simplified reservoir system. Controllable releases are only assumed for LYX (s^y) and LJX (s^j) and all the 10 smaller reservoirs in between are lumped into one reservoir s^l .

of the whole Yellow River basin has been distributed following the '87' Water Allocation Plan (as proposed in 1987) which divides the typical annual available streamflow (i.e., the 50th percentile streamflow) of 58 billion m³ into two

parts: 21 billion m^3 for sustaining river eco-environments (e.g., sediment flushing) and 37 billion m^3 for domestic use. Operations of the upper Yellow River reservoir system also follow this '87' plan of which further details can be seen in xia2012development. Typically, water managers make annual water release plans under the guidance of Yellow River Conservancy Commission (YRCC) on a monthly or bi-weekly basis using streamflow climatologies, and adaptively adjust the plan when more observational data becomes available. A map of the basin is shown in Figure 5.1a and the system consists of 12 serial reservoirs (Index 2-13). In this study, we focus on optimizing release decisions of two pivotal reservoirs, the Longyangxia (LYX, Index 2) and the Liujiaxia (LJX, Index 13) reservoirs since they are the only reservoirs with large enough capacities to operate on a weekly or longer timescale. Maximum storage capacities are about 20 billion m^3 and 5 billion m^3 for LYX and LJX, respectively while the rest 10 reservoirs between LYX-LJX have a total capacity of less than 3.5 billion m^3 [2]. For reference, total annual natural inflow is estimated at around 21.6 billion m^3 based on flow record over 2001-2009.

5.2.2 Simplified Water Budget Model

Water releases from LYX and LJX are optimized and the rest 10 small reservoirs in between are lumped into one reservoir to simplify water budget model of the system. 10-day historical flow and operation data over 2001-2009 are collected and for consistency, release decisions are being made on a 10-day basis in this study. The lumped reservoir is assumed to have zero storage capacity (i.e., no change of water storage) on the 10-day basis and such simplification is common in past studies [239, 244, 2, 263]. A schematic of the lumped reservoir system is

plotted in Figure 5.1b and the simplified water budget model is given by

$$\frac{ds^i}{dt} = q^i - r^i \quad (5.1)$$

$$q^{i+1} = r^i + q^{lat,i+1} \quad (5.2)$$

where q^i , s^i , r^i are inflow, storage and outflow of the i^{th} reservoir, respectively. The inflow q^{i+1} is sum of outflow from the previous reservoir (r^i) and lateral flow between i^{th} and $i + 1^{th}$ reservoirs ($q^{lat,i+1}$). If we denote water release decisions from the i^{th} reservoir by d^i , some physical constraints applied to the system are

1) Reservoir storages of LYX and LJX are limited in their corresponding designed ranges: $s^i \in [s^{i,min}, s^{i,max}]$.

2) Excessive water must be discarded as abandoned water a^i and is not used in hydropower production: $a^i = \max(0, s^i + q^i - d^i - s^{i,max})$.

3) Outflow r^i from each reservoir cannot exceed its total available water: $r^i = \min(d^i + a^i, s^i + q^i - s^{i,min})$.

Considering the water scarcity that the basin has been experiencing in the past decades and its semi-arid climate, instead of using extra objectives for flooding risks, we limit release decisions from i^{th} reservoir during t^{th} 10-day period in time-varying ranges $[d^{i,min}(t), d^{i,max}(t)]$ based on historical operations from 2001-2009. The normalized release decisions $D^i(t)$ are then used as decision variables in operation optimization and have a range of $[0, 1]$. The normalization is given by

$$D^i(t) = \frac{d^i(t) - d^{i,min}(t)}{d^{i,max}(t) - d^{i,min}(t)} \quad (5.3)$$

Hydropower production for each reservoir is then approximated by product of water head difference and volume of released water as given by Equation 4. For i^{th} reservoir, μ_i is a pre-specified efficiency coefficient, ΔH_i is difference between pre-dam and post-dam water heads, and g is the gravity constant. For the smaller reservoirs between LYX and LJX, ΔH_i s are set to constant water head differences. For LYX and LJX, pre-dam and post-dam water heads are approximated using empirical polynomial functions conditioned on water storage and outflow. Further information of reservoir specifications (e.g., μ_i and ΔH_i) and the empirical water head functions can be seen in si2019revealing.

$$E_i = \mu_i \cdot g \cdot \Delta H_i \cdot \min(d^i, r^i) \quad (5.4)$$

5.2.3 Synthetic Streamflow Generator

Key forcing variables including pre-LYX inflow (q^{inf}) and LYX-LJX lateral flow (q^{lat}) are sampled from a two-step streamflow generator. q^{inf} contributes a major part ($\sim 80\%$) of total natural inflow and its normalized value ($Q^{inf}(t)$) is first generated using a modified fractional Gaussian noise (mFGN) method [264] to maintain temporal autocorrelations. The normalization is done by 1) computing logarithmic values of true inflow data ($q^{inf}(t)$) (to reduce data skewness) and 2) subtracting climatological means ($\mu^{inf}(t)$) and dividing with climatological standard deviations ($\sigma^{inf}(t)$). The reverse normalization is given by Equation 5.

$$q^{inf}(t) = \exp(\mu^{inf}(t) + Q^{inf}(t) \cdot \sigma^{inf}(t)) \quad (5.5)$$

q^{lat} is then generated based on a good linear relationship between q^{inf} and q^{inf} ($R^2 = 0.53$ for 324 samples, p-value < 0.01 as shown in Figure C.1) which guarantees a consistent spatial correlation pattern. We simply assume that q^{lat} follows a Gaussian distribution of which the mean and standard deviation are conditioned on q^{inf} as given by Equation 6. (β_0, β_1) are intercept and slope of the line of best fit and $(\Delta\beta_0^\alpha, \Delta\beta_1^\alpha)$ are their α confidence intervals. The 95% confidence intervals are used to approximate the $4\text{-}\sigma$ uncertainty range of a Gaussian distribution. In addition, assuming precipitation is spatially uniformly distributed, q^{lat} is distributed to each reservoir as proportional to the corresponding drainage area before computing the hydropower production [244].

$$q^{lat}(t) \sim N\left(\beta_0 + \beta_1 \cdot q^{inf}(t), \left(\frac{\Delta\beta_0^{0.95} + \Delta\beta_1^{0.95} \cdot q^{lat}(t)}{4}\right)^2\right) \quad (5.6)$$

5.3 Methods

In this study, we contribute an analytical framework for quantitatively comparing impacts of information resolutions in managing a complex hydropower system. Our methodology builds off the EMODPS framework [249] to develop representative policies that approximate actual operations under different data availability and assess their expectation performance, robustness under changing climates, and time-varying sensitivities. The whole framework is described by a diagram as shown in Figure 5.2. A simplified water budget model is developed as an emulator of the complex hydropower system in section 2.2, which is

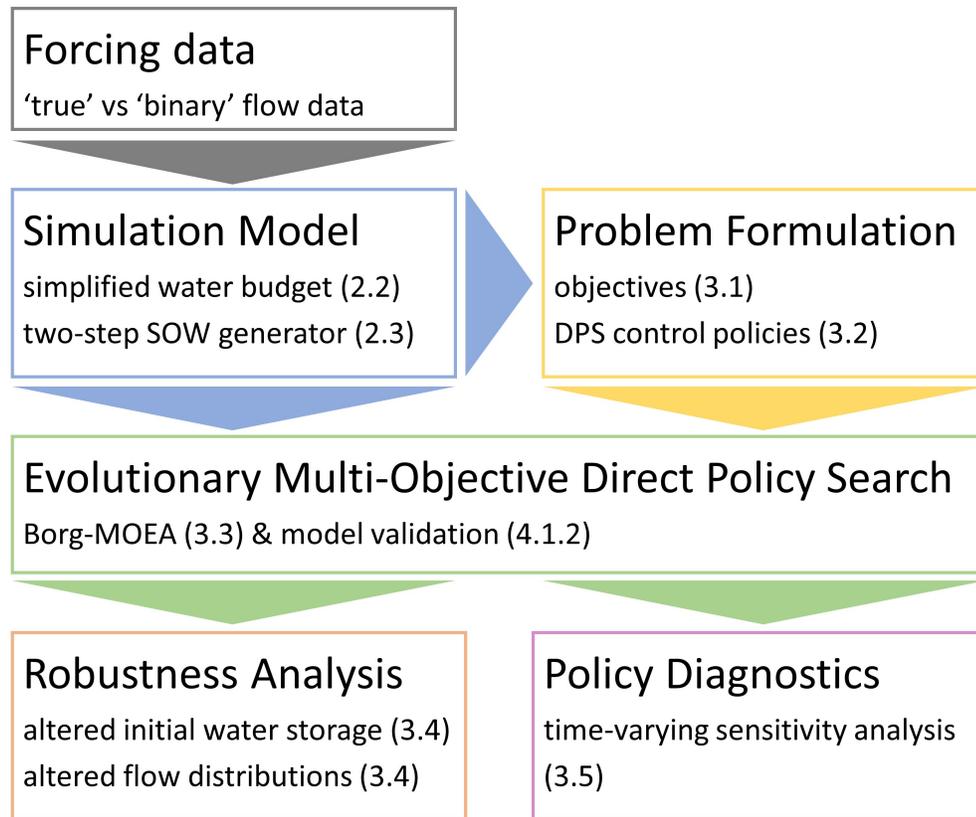


Figure 5.2: A diagram of the analytical framework for comparing expectation performance and policy robustness in managing the Upper Yellow River reservoir system with true-amplitude and binary flow data.

forced by varying states-of-the-worlds (SOWs) sampled using a two-step synthetic streamflow generator described in 2.3. Major utilities are addressed and modeled as objectives in 3.1, and closed-loop DPS policies are formulated conditioned on state variables in 3.2. The water system controlled by DPS policies is then optimized using the Borg-MOEA as introduced in 3.3 to develop representative policies by comparing against historical operations (in 4.1.2). A scenario discovery analysis is then conducted to evaluate robustness of representative policies trained using true and binary flow data in 3.4 following herman2015should. And based on quinn2019controlling, a TVSA is performed in

3.5 to understand impacts of forcing data resolutions in informing reservoir operations.

5.3.1 Formulation of Objectives

As introduced in the previous section, major utilities of the upper Yellow River reservoir system include hydropower production, load balancing for photovoltaic energy, supplying water for downstream cities and irrigation districts, and mitigating ice-jam floods. These utilities are formulated here as three key objectives to maximize total hydropower production and minimum 10-day hydropower production, and to minimize a water deficit index. For each policy, the system is simulated and optimized under forcing of synthetically generated flow data of $T = 36$ 10-day periods (i.e., one year) and this step is repeated for $N = 100$ SOWs. Objectives across the 100 SOWs are averaged for performance evaluation which represents a risk-neutral preference of decision makers [226].

The first objective is to maximize total hydropower production (J^{totE}) over the simulation horizon of one year as given by Equation 7. It should be noted that, although the 10 smaller reservoirs between LYX-LJX are lumped into one reservoir to simplify the water budget model, hydropower production is separately estimated for each individual reservoir considering their different designed water head differences and efficiency coefficients. This is also why q^{lat} is distributed to each reservoir before computing the hydropower production.

$$\text{Maximize } J^{totE} = \frac{1}{N} \sum_{k=1}^N \sum_{t=1}^T \sum_i E_k^i(t) \quad (5.7)$$

The second objective is to maximize minimum 10-day hydropower production (J^{minE}) as given by Equation 8. This objective is designed to measure reliability of load balancing using hydropower for other renewable energy. Technically, the reliability is function of both hydropower output and load balancing demand (e.g., peak-to-valley output difference of renewable energy). Since only limited information of variations of the renewable energy outputs is known here, the objective is re-framed to provide as much buffer capacity for load balancing as possible during any 10-day period. Therefore, the minimum 10-day hydropower production is to be maximized and such formulation is also adopted in si2018evaluating.

$$Maximize J^{minE} = \frac{1}{N} \sum_{k=1}^N \min_t \sum_i E_k^i(t) \quad (5.8)$$

The third objective is to minimize a water deficit index (J^{WD}) based on some pre-defined time-varying water demands (W') as given in Equation 9. The water demands are first specified at Lanzhou separately for wet and dry seasons on a monthly basis. During the wet season (Apr-Oct), minimum streamflow is regulated to meet water demands of riparian cities and downstream irrigation districts according to the '87' Plan and is distributed based on monthly flow climatologies [239]. During the dry season (Nov-Mar), minimum streamflow is regulated to mainly avoid ice jam flooding [265]. Water demands at Lanzhou for the two seasons are combined and shown in Table 5.1. These water demands are then rescaled using a ratio of annual streamflow measured at post-LJX to that measured at Lanzhou. A factor of safety (α) is used to avoid frequently crossing bottom lines of water demands in reservoir operations and α is set to 1.5 in this study.

$$\text{Minimize } J^{WD} = \frac{1}{N} \sum_{k=1}^N \frac{1}{T} \sum_{t=1}^T \frac{\max(0, r_k^j(t) - \alpha \cdot W(t))}{\alpha \cdot W(t)} \quad (5.9)$$

Table 5.1: Time series of pre-determined water demands $W(t)$ at Lanzhou based on chang2014optimized,bai2015synergistic.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Streamflow (m^3/s)	550	500	350	750	1100	900	800	750	750	800	650	600

A fourth objective is proposed to maximize relative total water storage at the end stage (J^{SP} , SP for sustainability penalty) as given by Equation 10. This last objective is designed to avoid unreasonable and unsustainable decisions like over-depletion towards end of the simulation horizon considering we limit our simulation to only one year (i.e., reservoir operations cease after one year) in this study.

$$\text{Maximize } J^{SP} = \frac{1}{N} \sum_{k=1}^N \frac{s_k^y(T) + s_k^j(T)}{s^{y,max} + s^{j,max}} \quad (5.10)$$

5.3.2 Formulation of DPS policies

A closed loop DPS control strategy is adopted here to, instead of optimizing absolute water releases given specific flow conditions, develop adaptive policies that approximate release decisions conditioned on a set of input variables. DPS policies have good interpretability and can inform decision making under out-of-sample conditions used in further analysis of variance [252] and sensitivity [251]. Gaussian Radial Basis Functions (RBFs) have demonstrated good performance in approximating nonlinear policy functions [266, 252] and are used in

defining DPS policies as shown in Equation 11.

$$D^i(t) = \sum_k^A w_k^i \exp\left(-\sum_l^B \frac{(x_l(t) - c_{l,k})^2}{b_{l,k}^2}\right) \quad (5.11)$$

At time t , water release decision $D^i(t)$ for LYX ($i = y$) or LJX ($i = j$) is approximated by a linear combination of A RBFs (with weighting coefficients w_k^i). Each RBF uses information from B input variables (x_l) and the corresponding centers and radii are $c_{l,k}$ and $b_{l,k}$, respectively. $B = 5$ input variables are used as defined in Equation 12 which include normalized water storage from LYX and LJX, a time-of-year index, normalized pre-LYX inflow and normalized LYX-LJX lateral flow. $(Q^{inf}(t), Q^{lat}(t))$ are normal variables used in the synthetic streamflow generator and are shifted by 2 and then rescaled by 1/4 to be roughly limited by a range of $[0, 1]$ ($Q^{lat}(t)$ is the z-score of $q^{lat}(t)$). To compare impacts of information resolutions in informing reservoir operations, we design a world where only binary flow data is available. Flow-related input variables in this world are changed to $(Bin(Q^{inf}), Bin(Q^{lat}))$. $Bin(y)$ is defined to return a binary state of y such that $Bin(y) = 1/4$ if $y < 0.5$ and $Bin(y) = 3/4$ otherwise.

$$x_l(t) = \left(\frac{s^y(t) - s^{y,min}}{s^{y,max} - s^{y,min}}, \frac{s^j(t) - s^{j,min}}{s^{j,max} - s^{j,min}}, \frac{t\%36}{36}, \frac{Q^{inf}(t) + 2}{4}, \frac{Q^{lat}(t) + 2}{4} \right) \quad (5.12)$$

A rule of thumb suggests number of RBFs equal one plus sum of output and input variables ($A = 1 + 2 + 5 = 8$) [267]. We repeated our simulation experiment using varied A s in [5,8] and observed very consistent objective values in a preliminary study. Therefore, A is set to 5 in the following analysis for less model complexity and potentially better model robustness. And a total number of $A(2B+2) = 60$ parameters are to be optimized in this study. Sum of weighting

coefficients is set to 1 to limit $D^i(t)$ in the range of $[0, 1]$ (i.e., $\sum_k w_k^i = 1, \forall w_k^i \geq 0$). As for other constraints, we set $c_{l,k} \in [0, 1]$ and $b_{l,k} \in [0, 10]$.

5.3.3 Many-Objective Optimization

Borg-MOEA is used to search for approximate Pareto-optimal solutions in the parameter space $(w_k^i, c_{l,k}, b_{l,k})$ that minimize the objective vector $(-J^{totE}, -J^{minE}, J^{WD}, -J^{SP})$. A solution is defined Pareto optimal or non-dominant if no further improvement can be made for one objective without degrading performance for at least one other objective [268]. A set of all such solutions and its corresponding set of objective values are often referred to as the Pareto set and Pareto front, respectively. Borg-MOEA has shown significant advantages over other MOEAs in solving many-objective, multi-modal problems [253, 269, 270]. It has good performance in both elite preservation by using adaptive population sizing and diversity preservation by using ϵ -dominance archiving. Also, a auto-adaptive multi-operator searching process adds more tractability to the algorithm while cutting down computational costs. Since Borg-MOEA is a stochastic optimization algorithm that uses random initialization [253], the optimization experiment is repeated for 50 random seeds for consistency control. Maximum number of function evaluations (NFEs) is set to 400,000. Hypervolumes measured at every 100 NFEs show that a maximum of 400,000 NFEs can guarantee convergence across 50 random seeds in our study (as shown in Figure C.2). (Hypervolume measures 'volume' of space dominated by an approximate Pareto solution in the multi-objective space and is a widely-used metric of performance [271].)

5.3.4 Robustness Analysis

Robustness is used as the other important metric of performance complementary to expectation objectives. In this study, we focus on evaluating robustness of policies with respect to two major sources of uncertainty, initial water storage and flow distribution. Explicitly optimizing operations under varying initial conditions of LYX and LJX water storage will greatly increase complexity of the problem and make it hard to compare different policies. Instead, we train DPS policies with fixed initial conditions and measure robustness of these policies under altered conditions of initial water storage. As for flow distributions, altered SOWs are generated for both q^{inf} and q^{lat} by making extreme flow conditions more likely following kasprzyk2013many. A flow condition is defined extreme if its amplitude is smaller than the 25th percentile value or greater than the 75th percentile value.

A scenario discovery approach is adopted based on herman2015should and consists of 3 major steps: 1) train DPS policies by optimizing system operations in the original SOWs; 2) generated a broad range of SOWs with altered conditions; and 3) re-evaluate policies in the altered SOWs. For each source of uncertainty, 100 altered SOWs are generated using a Latin Hypercube Sampling (LHS) method [272]. In each altered SOW, the optimization experiment is repeated for 100 times and a ratio of successful cases (denoted by S or S score in the following sections) is used to measure robustness based some baseline performance. A policy in a SOW is defined as a success only if its objective vector fully dominates the baseline performance (which are further detailed in Section 4.1). S scores are mapped onto corresponding altered conditions to discover under what scenarios, in the space of uncertain parameters, policies tend

to perform and fail.

5.3.5 Time-varying Sensitivity Analysis

Last but not least, a TVSA is performed here based on [252] to understand how information from input variables controls and coordinates release decisions when driven by flow data of different resolutions. The analysis decomposes variances of release decisions $D^i(t)$ using a first-order Taylor-series expansion as given in Equation 13, and compares contributions from individual input variables and their interactions. Since a set of differentiable basis functions (i.e., the Gaussian RBFs) are used in approximating Pareto-optimal policies, the derivative terms $\frac{\partial D^i}{\partial x_m}$ can be analytically solved as given by Equation 14. And the covariance terms $cov(x_m(t), x_l(t))$ can be numerically estimated using simulation samples across different SOWs. To validate this analysis, a preliminary study is done by comparing total variance of $D_i(t)$ estimated using simulation samples against the sum of variances calculated using the TVSA and good linear relationships ($R^2 > 0.9$, p-value < 0.01 for 36 samples) are observed for both policies trained using true-amplitude and binary flow data.

$$Var(D^i(t)|t) \approx \sum_m^B \sum_l^B \left(\frac{\partial D^i}{\partial x_m} \frac{\partial D^i}{\partial x_l} \cdot cov(x_m(t), x_l(t)) \right) \quad (5.13)$$

$$\frac{\partial D^i}{\partial x^m} = \sum_k^A \left\{ -2w_k^i \frac{x_m(t) - c_{m,k}}{b_{m,k}^2} \exp \left[- \sum_l^B \left(\frac{x_l(t) - c_{l,k}}{b_{l,k}} \right)^2 \right] \right\} \quad (5.14)$$

5.4 Results & Discussions

5.4.1 Policy Comparison & Validation

Multi-objective Comparison

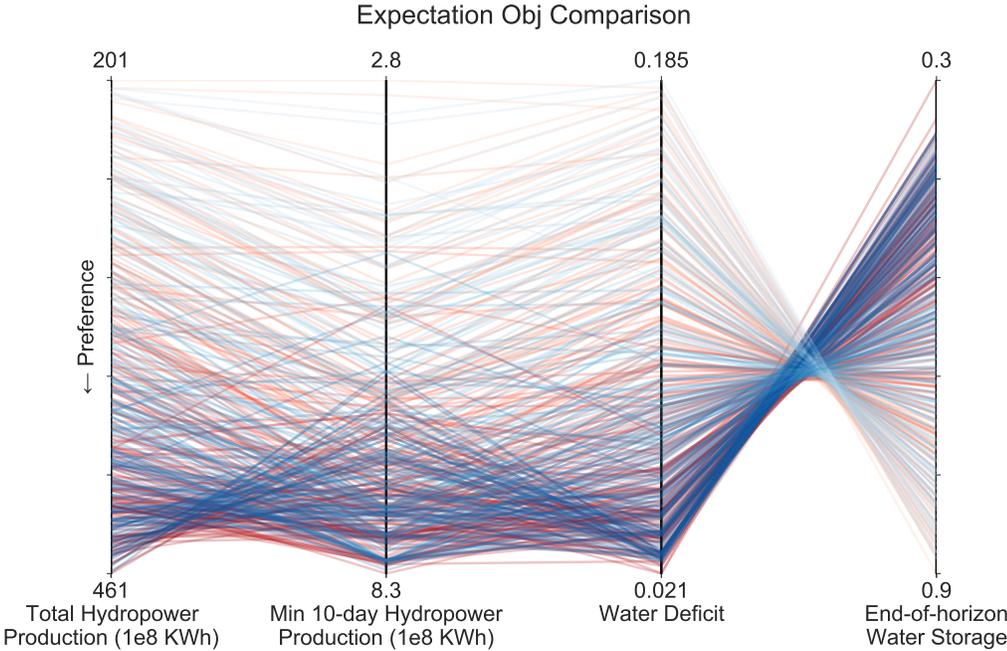


Figure 5.3: Parallel axis plot of expectation objectives from 'true' (red) and 'binary' (blue) policies for four objectives: total hydropower production (totE), minimum 10-day hydropower production (minE), water deficit (WD), and end-of-horizon water storage (SP). Objective ranges are shown and a direction of increasing preference is indicated by a downward arrow. The shading corresponds to the expectation objective of total hydropower production. A darker value means greater total hydropower production.

Comparison of multi-objective performance shows that there is no significant difference in expectation objectives for policies trained using true-amplitude

(denoted by **'true'** policies) and binary (denoted by **'binary'** policies) flow data. By merging Pareto-optimal solutions across all 50 random seeds using ϵ -dominance [253], 97.2% of 'true' policies and 93.6% 'binary' policies are retained. The expectation objective vectors $(J^{totE}, J^{minE}, J^{WD}, J^{SP})$ are rescaled to ranges of [0,1] and are compared in a parallel axis plot in Figure 5.3. Comparable ranges and consistent tradeoffs between conflicting objectives are observed for 'true' and 'binary' policies as plotted in red and blue lines, respectively. End-of-horizon water storage is negatively correlated with the other 3 objectives. Total hydropower production is positively correlated with 10-day minimum hydropower production until it exceeds some threshold (410 KWh, or a rescaled value of around 0.8), and further increasing total hydropower production can lower the minimum 10-day hydropower productions. The tradeoff between $(J^{totE}, J^{minE}, J^{WD})$ and J^{SP} is easy to understand since $(J^{totE}, J^{minE}, J^{WD})$ are mainly determined by total amount of water released from the system while J^{SP} is determined by how much water is saved. As for the negative correlation between J^{totE} and J^{minE} when J^{totE} exceeding some threshold, a possible explanation is that since the pre-dam water head increases monotonically as the water storage increases, to delay releasing water can increase total hydropower production by increasing water head difference in Equation 4 while in the meantime, it can lower 10-day minimum hydropower production during early periods. Overall, no significant difference in expectation performance is observed for DPS policies trained using true-amplitude and binary flow data.

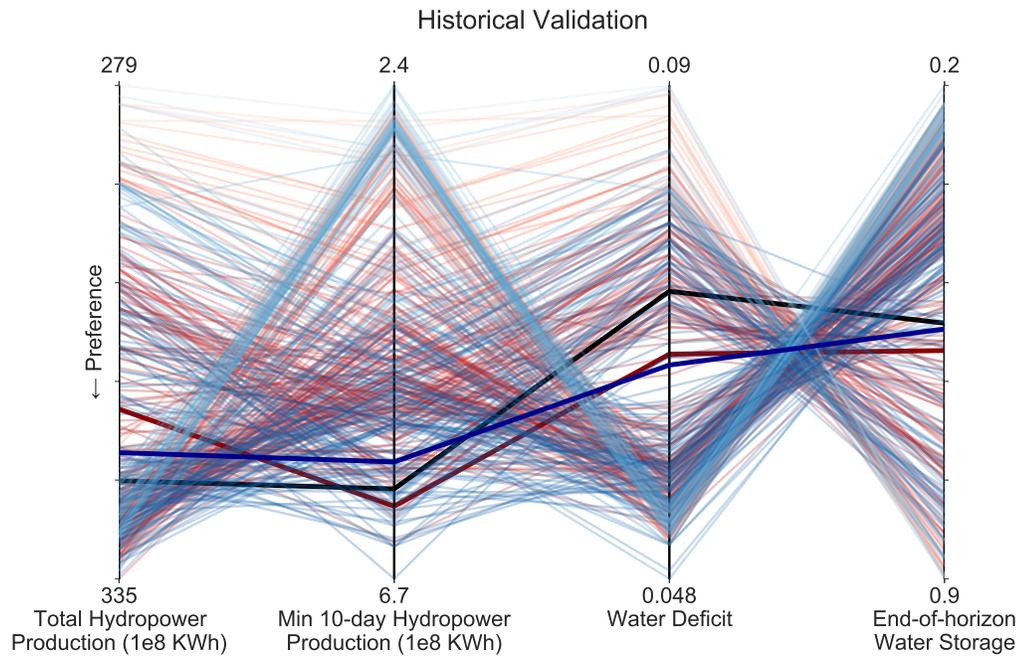


Figure 5.4: Parallel axis plot of expectation objectives from 'true' (red) and 'binary' (blue) policies tested using historical flow data over 2001-2009. Maximum and minimum objective values are shown and a direction of increasing preference is indicated by a downward arrow. Performance of the actual historical operations is plotted in a black line and the shading corresponds to mean square distance between the historical performance and DPS policy performance. The darker value means a DPS policy has more similar performance compared to historical operations. The 'true' and 'binary' policies that have the closest performance compared to the historical operations are defined as the representative policies and are plotted in dark red and blue lines, respectively.

Policy Validation

The DPS policies are then tested using historical flow data and the results indicate that our simplified water budget model can well approximate actual operations. And two representative policies are defined that have the closest expectation objectives compared to the historical operations. This validation exper-

iment is necessary considering many aggressive assumptions are used in simplifying the system. Some major assumptions include: 1) the system is quasi-steady on a 10-day basis, 2) water head differences are constant for smaller reservoirs between LYX-LJX, 3) lateral flow is distributed as proportional to drainage areas, and 4) LYX-LJX lateral flow is linearly dependent on pre-LYX inflow.

Here, the reservoir system is continuously operated as determined by DPS policies and driven by historic flow throughout 2001-2009. The water storage is not reset every year, and impacts of inter-annual flow variations are considered since variations in water storage can carry over across years. This change can potentially degrade policy performance considering that the DPS policies are trained with fixed initial conditions of water storage and no inter-annual variability is modeled in the synthetically generated flow data. Expectation objectives over 2001-2009 using DPS policies are plotted and compared to that using historical operations as shown in Figure 5.4. Two representative policies are defined for cases using true-amplitude and binary flow data such that mean squared difference in expectation objectives is minimized compared to historical performance. Though beat by historical operations in 'totE', both representative policies have better performance in water supply ('WD') and sustainability ('SP'). The 'true' representative policy even outperforms historical operations in 'minE' but it loses in 'totE' by a larger margin compared to the 'binary' representative policy. Even with the potential performance degradation due to being tested under out-of-sample conditions, DPS policies show good performance comparable to that using historical operations.

Operations determined by representative DPS policies are found to be highly consistent with historical operations when we compare trajectories of reservoir

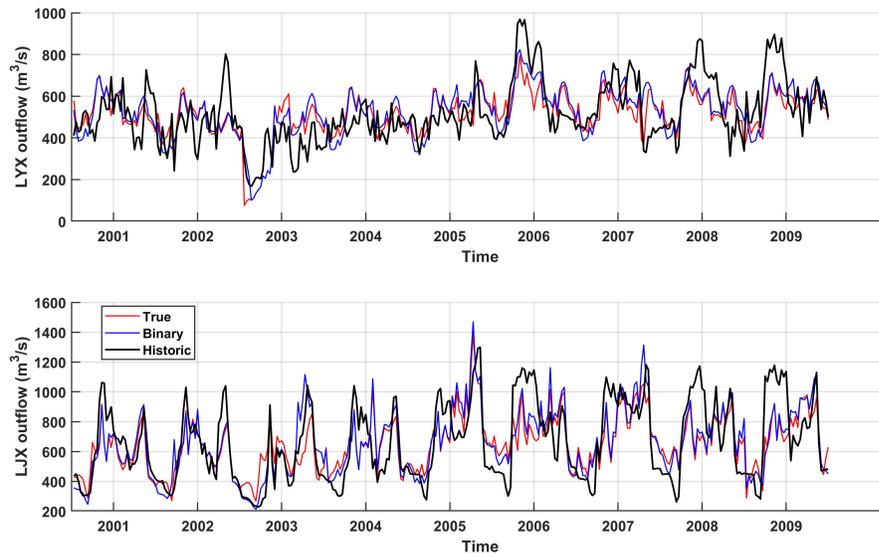


Figure 5.5: Comparison between times series of LYX (top) and LJX (bottom) reservoir outflow determined using historical operations (black) and representative policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.

outflow and water storage in Figure 5.5 and 5.6. Correlations coefficients between operation trajectories are reported in Table 5.2 and all correlations are statistically significant (p -value < 0.001 for 324 samples) except for LJX water storage. The LJX reservoir most of the time is operated at high water storage levels near the maximum capacity while its historical operations are more conservative. A possible explanation is that DPS policies tend to move water from LYX to LJX and save more buffer capacity in LYX so as to increase hydropower production. Different sensitivities of pre-dam water head to water storage are observed (in Figure C.3), and for the same amount of increased water storage, pre-dam water head and thus hydropower production are increased by larger margins for LJX. Also, our simplified system is indifferent between water storage in LYX and LJX (as indicated by Equation 10). Therefore, the trained policies will always operate LJX at high water storage levels to maximize total and 10-

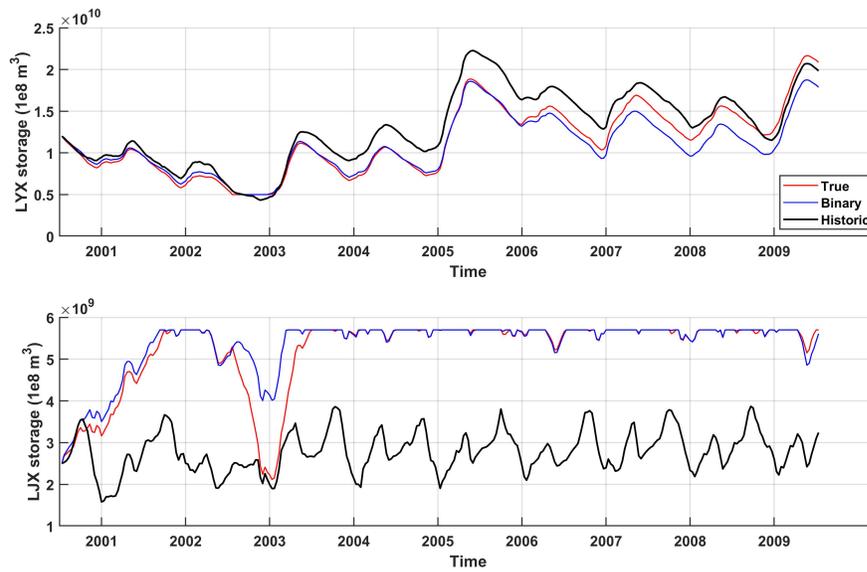


Figure 5.6: Comparison between time series of LYX (top) and LJX (bottom) reservoir water storage determined using historical operations (black) and representative policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.

day minimum hydropower productions while maintaining good performance in minimizing water deficit and maximizing end-of-horizon water storage. This hypothesis is also supported by greater consistency observed between total water storage ($R_{LYX}+R_{LJX}$) determined using representative DPS policies and historical operations (as seen in Figure C.4). The Nash-Sutcliffe Efficiency (NSE) [273] scores are 0.79 and 0.67 for 'true' and 'binary' policies in comparing LYX water storage (to historical operations), and are improved to 0.89 and 0.91 in comparing total water storage. In short, good consistency in both performance and operation trajectories should validate our simplified water system and based on these results, two representative policies that mimic the real-world water managers are defined for further analysis.

Table 5.2: Pearson’s correlation coefficients between the historical operations and operations determined by the representative DPS policies (for 324 samples)

Policy \ Var	LYX outflow	LJX outflow	LYX storage
True	0.59	0.70	0.96
Binary	0.62	0.74	0.97

5.4.2 Robustness Analysis

Before jumping right into the robustness analysis, baseline performance must be defined for discovering feasible and failure scenarios in the space of altered conditions. For an easier physical interpretation, the historical performance over 2001-2009 as shown in Figure 5.4 is adopted as the major set of baseline performance. In each SOW with altered conditions, a policy is defined as successful only when its objective vector fully dominates the baseline objective vector. While increasing initial water storage only monotonically increases available water for the system to manipulate, increasing probability of extreme flow events can have two-sided effects since both drought and flood events become more likely. Under altered flow distributions, a strict baseline performance essentially measures how policies can make more use of excessive water from flood events while a relaxed baseline performance measures how policies avoid degradation of performance when facing more frequent drought events (as shown in Figure C.5). The historical performance sets up a rather strict set of criteria since it requires a ‘successful’ policy to have above average performance in all four objectives. Therefore, a relaxed set of baseline performance is added so that robustness under altered flow distributions from both sides can be evaluated. The two sets of baseline performance are given in Table

5.3.

Table 5.3: Two sets of baseline performance used in computing S scores in the robustness analysis

Baseline \ Obj	totE (1e8 KWh)	minE (1e8 KWh)	WD	SP
Strict	324.4	5.89	0.073	0.58
Relaxed	220.0	4.00	0.110	0.40

Altered Initial Conditions

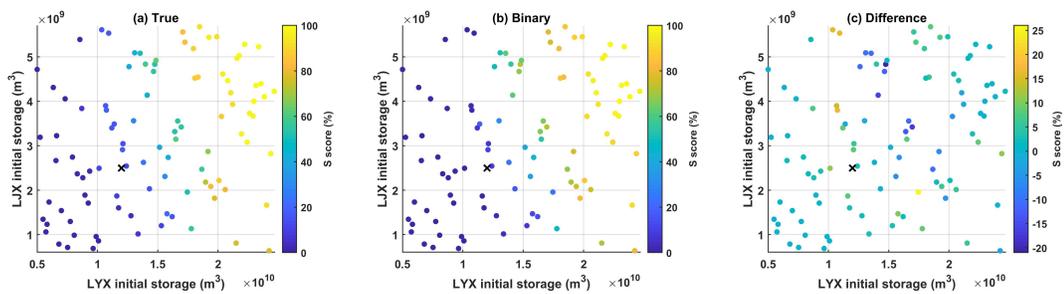


Figure 5.7: Robustness reported in S scores under changing initial conditions of water storage for the 'true' (a) and 'binary' (b) representative policies and the difference ('true' - 'binary') (c). The strict baseline performance based on the historical operations is used. The original condition under which DPS policies are trained is marked by a black cross.

'True' policies are found to be more robust than 'binary' policies under changing initial conditions of water storage. 100 initial conditions are sampled using LHS in ranges of [0, 1] for normalized LYX and LJX water storage, and the two representative policies are tested for 100 SOWs under each altered condition. Comparisons of performance as measured by S scores are plotted in Figure 5.7. 'True' and 'binary' policies show similar patterns that the number of successful cases increases as the total initial water storage increases and the performance

improvement is indifferent to water distribution between LYX and LJX. This is mainly because no objective associated with flooding disasters is explicitly modeled in the system considering basin's semi-arid climate and recent water scarcity. Increasing initial water storage, and thus increasing total available water can directly improve performance in all four objectives (J^{totE} , J^{minE} , J^{WD} , J^{SP}). Difference of robustness shows that the 'true' representative policy outperforms its 'binary' counterpart under 55 altered conditions and only loses under 23 altered conditions.

An interesting pattern observed is that the 'binary' policy has higher S scores than the 'true' policy when total initial water storage is only marginally increased (to about 19-20 billion m³). The improvement is found to be consistent after comparing S scores of both policies under a fixed initial condition of ($R_{LYX} = 16, R_{LJX} = 3.2$) billion m³ with more repeated simulations (i.e., 1000 SOWs). A comparison of flow trajectories suggests that under this specific initial condition, outflow from both LYX and LJX are lifted by larger margins when determined using the 'binary' policy, which should further minimizes water deficit. This explanation is supported by the fact that the 'binary' policy only has markedly better performance in J^{WD} and comparable performance in other objectives when compared to the 'true' policy. In other words, the 'binary' representative policy is more sensitive to changes in water storage since it releases more water given the same amount of increased water storage. However, this effect is less significant when we further increase initial water storage to a stage where WD is no longer the bottleneck objective.

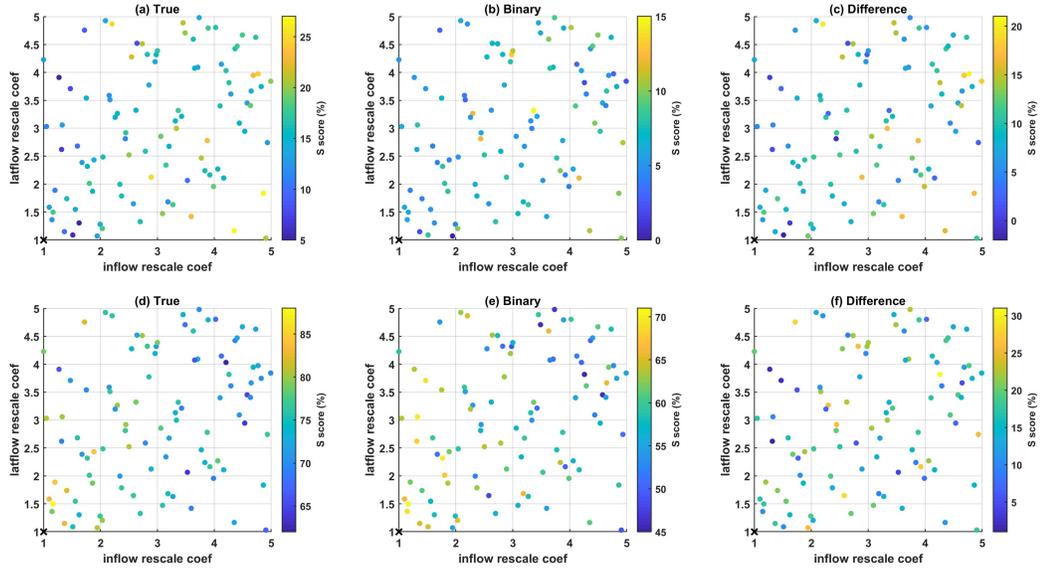


Figure 5.8: Robustness reported in S scores under changing flow distributions for the 'true' (a,d) and 'binary' (b,e) representative policies and the difference ('true' - 'binary') (c,f). Both the strict (top, or a,b,c) and relaxed (bottom, or d,e,f) baseline performance are used. The original condition under which DPS policies are trained is marked by a black cross.

Altered Flow Distributions

Comparisons of robustness under changing flow distributions indicate that the 'true' policy outperforms the 'binary' policy in both making better use of excessive water from flooding events and minimizing performance degradation due to drought events. Altered flow distributions are generated by making extreme flow events more likely by a rescaling coefficient in $[1, 5]$ [257]. Robustness under altered flow distributions for the two representative policies and their difference are plotted in Figure 5.8 based on the two sets of baseline performance. The two-sided effects are observed for both policies as in SOWs with more frequent extreme flow conditions, better and poorer S scores are observed when compared the strict and relaxed baseline performance, respectively. The 'true'

policy has higher S scores than the 'binary' policy under 96 altered flow distributions when the strict baseline performance is used and under all 100 altered flow distributions when the relaxed baseline performance is used. Also, when generating SOWs with altered flow distributions, we adjust the rescaling coefficients rather than specify time series of inflow and lateral flow. Internal sampling variability from the SOW generator still plays an important role in determining flow uncertainty and flow can be non-extreme even when its probability distribution function is stretched. And this should explain why robustness under altered flow distributions appear to be more random compared to results in Figure 5.7.

5.4.3 Time-varying Sensitivity Analysis

Sensitivities of prescribed LYX and LJX release decisions determined using the 'true' and 'binary' policies are plotted in Figure 5.9. Some similarities are observed across the 'true' and 'binary' policies: 1) variations in pre-LYX inflow play the major role in making release decisions during the early periods; 2) contributions from variations in LYX storage grow through time and overtake the pre-LYX inflow to become the predominant factor in the second half of the year. The growing importance of S^y closely relates to the fixed initial conditions of water storage used in simulations. The variances of S^y and S_j are set to zero in the starting period and only gradually increase when streamflow uncertainties propagate to those in water storage. Therefore, little contributions are observed from S^y and S^j during the early periods. It is noteworthy that S^j make minor contributions compared to S^y since the LJX only accounts for 20% of total water storage capacity of the system. These similarities can also help to explain the comparable expectation performance between policies trained using flow

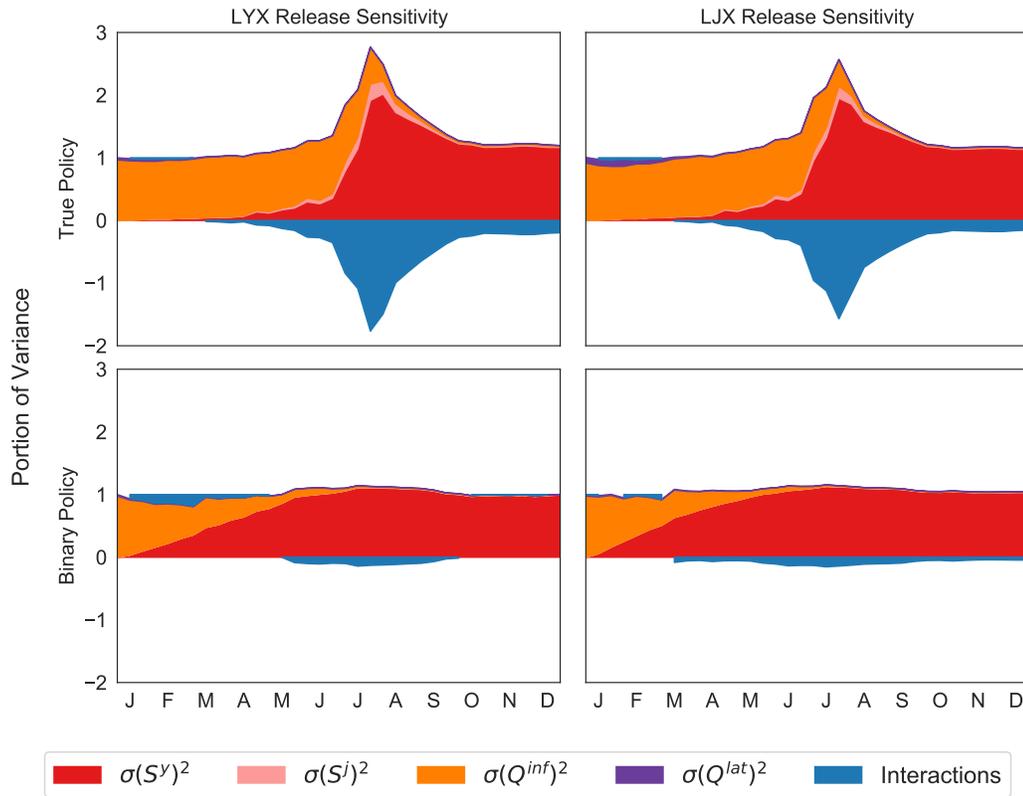


Figure 5.9: Sensitivities of prescribed LYX (left) and LJJ (right) release decisions determined by the 'true' (top) and 'binary' (bottom) policies. The total variances of release decisions are decomposed to contributions of different sources (S^y in red, S^j in pink, Q^{inf} in orange, Q^{lat} in magenta, and interactions in blue).

data of different resolutions. The LYX water storage plays a predominant role in decision making for both representative policies, especially during the last half of the year. Therefore, operations and policy performance are less sensitive to resolutions of flow data since Q^{inf} and Q^{lat} only make limited contributions in informing decision making. A physical interpretation here is that the upper Yellow River reservoir system has a massive capacity of water storage so that it can always store water from natural inflow first and make release decision based on water storage later.

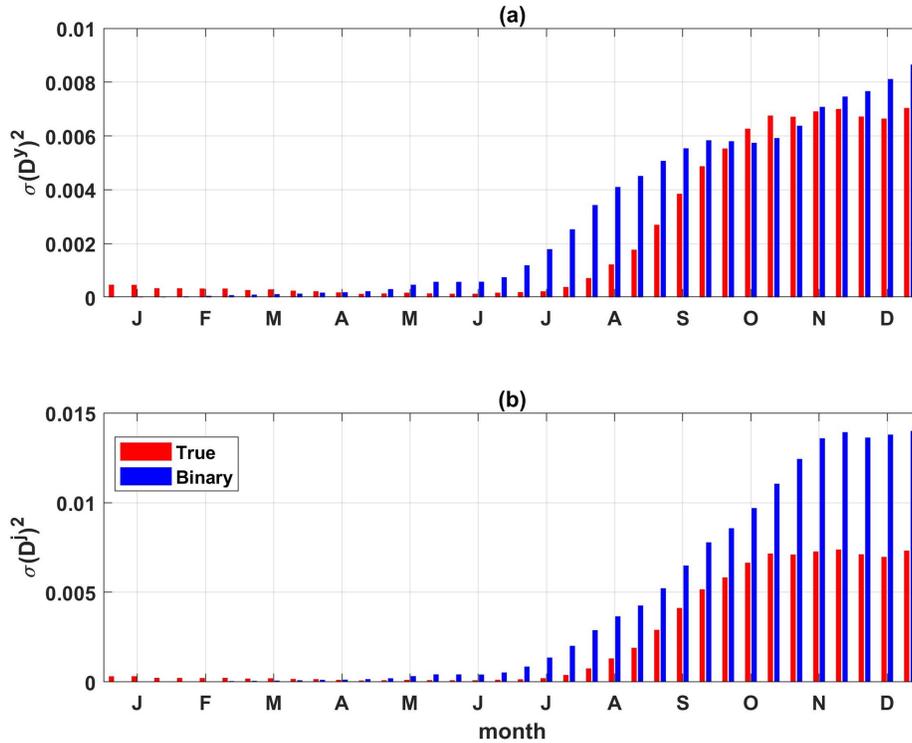


Figure 5.10: Variances of (a) LYX and (b) LJX release decisions estimated using simulation samples across SOWs. Variances using the 'true' (red) and 'binary' policies are compared and reduced uncertainties are observed in release decisions when the true-amplitude flow data is used.

Here, we also want to highlight some differences in sensitivities when flow data of different resolutions are used. Variations in Q^{inf} make more contributions (especially during the first half of the year) and the timing of S^y overtaking Q^{inf} is delayed when the true-amplitude flow data is used. These patterns indicate that more information can indeed be incorporated into making release decisions when flow data of finer resolutions is available. Also, a more significant negative portion of interactions observed for the 'true' policy may suggest that uncertainty in release decisions can be reduced by incorporating more information from flow data. And this hypothesis is supported by smaller variations in release decisions determined by the 'true' policy as shown in Figure 5.10. The re-

sults indicate higher sensitivity of release decisions to climate uncertainty when only binary flow data is used since the same SOW generator is used for training and testing both sets of policies. Though not modeled as a objective in our study, stability or inertia can be an important dimension in evaluating implementability of a policy [251]. And the 'true' policies show superior performance in this dimension for being less sensitive to climate uncertainty.

5.5 Conclusions

Motivated by an emerging imbalance between increasing problem complexity and lack of access to high-resolution data, we conduct a comparison experiment as an attempt to answer this question: how will forcing data of finer resolution improve system optimization performance? Operations for the upper Yellow River reservoir system are optimized under forcing of true-amplitude and binary flow data and their performance are compared. A DPS control strategy is adopted to approximate release decisions conditioned on a set of variables including water storage, flow conditions and a time-of-year index. When compared with historical operations, comparable expectation performance and consistent operations are observed for DPS policies trained using both true-amplitude and binary flow data. Two representative policies are defined to mimic real-world water managers based on historical operations over 2001-2009. Though very negligible difference is observed in expectation objectives, the 'true' policy demonstrates superior robustness under altered initial conditions of water storage and flow distributions as compared to its 'binary' counterpart. At last, a TVSA is used to open black boxes of the two sets of policies by decomposing variances in release decisions to contributions from different in-

put variables. TVSA results suggest both policies are dominated by variations of LYX storage in making release decisions, which should explain the comparable expectation performance. However, having access to high-resolution flow data improves stability of policies by reducing uncertainties in release decisions, which can make them easier to implement in real-world management. Overall, our study suggests that using forcing data of finer resolutions may not necessarily improve expectation performance in reservoir operations (especially for large water systems), but it can develop more robust and stable policies under changing climates. And we suggest decision makers of large water systems consider the tradeoff between costs in data acquisition and potential returns (i.e., marginal improvement in expected performance but significant improvement in robustness) when they seek to buy high-resolution data or to develop new monitoring/forecast systems. Also, our analytical framework can be easily generalized to compare impacts of forcing data of different designed uncertainties in informing water system management.

CHAPTER 6

CONCLUSIONS

Though the recent boom in sensing technology and machine learning has provided new powerful tools in hydro-climate research, a tradeoff between model accuracy and complexity must be considered. A model usually loses its interpretability and becomes hard to validate under out-of-sample conditions when it gets highly complex, and this reflects the prevailing preference of people not trusting black-box models. This dissertation presents applications of some interpretable data-driven approaches through a typical chain of uncertainty propagation in climate-hydrology-water resource management. First, regional precipitation is forecasted based on its teleconnections with large-scale climates. Second, dynamics of local precipitation-runoff responses are diagnosed before translating precipitation into streamflow. Last, water systems are managed as driven by hydrological forcings of different levels of accuracy (i.e., resolutions). On one hand, the data-driven approaches show good potential in guiding exploration of physical processes (Chapter 2 and 3). On the other hand, computational costs can be reduced and model interpretability can be improved when physical constraints are applied (Chapter 4). Furthermore, a quantitative framework is developed for comparing effects of forcing uncertainty (as indicated by data accuracy) on water system management.

In Chapter 2 and 3, elastic net and its generalized version show good model skill in predicting regional precipitation using large-scale climate variables. Chapter 2 examines predictability of EASR based on its teleconnections with multiple predictors including sea surface temperature, geopotential height, vertical atmospheric velocity and horizontal wind speed, and two stages of high pre-

dict skill are observed. By comparing regression coefficients with correlation maps, the dominant role of 'Walker Cell' in affecting EASR is rediscovered for the short-lead stage, and a potential physical causal chain is developed to link long-lead Mascarene High anomalies to short-lead perturbations in the 'Walker Cell'. Chapter 3 limits predictor to sea surface temperature and investigate feasibility of generalized elastic net from a statistical perspective. A comparison against other commonly-used regressions demonstrates effectiveness of regularization and noise smoothing layers (e.g., pooling and convolution) in dealing with collinearity in the highly 'non-square' climate data. And the generalized elastic net manages to select persistent correlation patterns which can guide further physical study with more complex models.

In Chapter 4, a Bayesian estimation method is used to track dynamics of regional hydrological responses. The Instantaneous Unit Hydrograph based on a modified Gamma distribution is adopted to model the local rainfall-runoff responses and its parameters are tracked to represent hydrological properties. The Metropolis Hastings algorithm is implemented to estimate PDFs of the parameters given observations of runoff and runoff. The case study indicates 1) a systematic shift to shorter water travel time around 1943 for the Fall Creek watershed and 2) a statistically significant relationship between rainfall-runoff responses and watershed sizes for selected NY watersheds. These findings cannot be learned by looking at the easy-to-measure variables like rainfall and runoff. The Bayesian estimation method shows potential as a fast surveillance tool to reflect rapid changes in regional hydrological responses as well as to quantify uncertainties in hydrological models.

In Chapter 5, the EMODPS framework is adopted to evaluate effects of forcing

data resolutions (true-amplitude vs binary) on water system management. The complex upper Yellow River reservoir system is optimized in four objectives including hydropower production, load balancing capacity, water supply, and sustainability. While only negligible difference is observed in expectation objectives, policies trained using true-amplitude flow data show superior robustness and stability. By coupling a time-varying sensitivity analysis with the DPS control strategy, we can further understand how the two sets of policies coordinate information from different hydroclimate variables, which helps interpret the difference in policy performance. The results suggest water managers consider the tradeoff between data acquisition costs and potential improvement in policy robustness and stability since improvement in expectation performance can be marginal.

In future work, I intend to integrate the three sectors in a whole framework so that uncertainties from the two major sources, climate forcing and hydrological models, can be quantitatively incorporated into water system management. And I intend to design scenarios of different levels and combinations of uncertainties under which policy performance and sensitivities are compared. Such analysis should help decision makers develop more efficient plans for new sensor networks or forecast model development for uncertainty reduction and system operation optimization.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

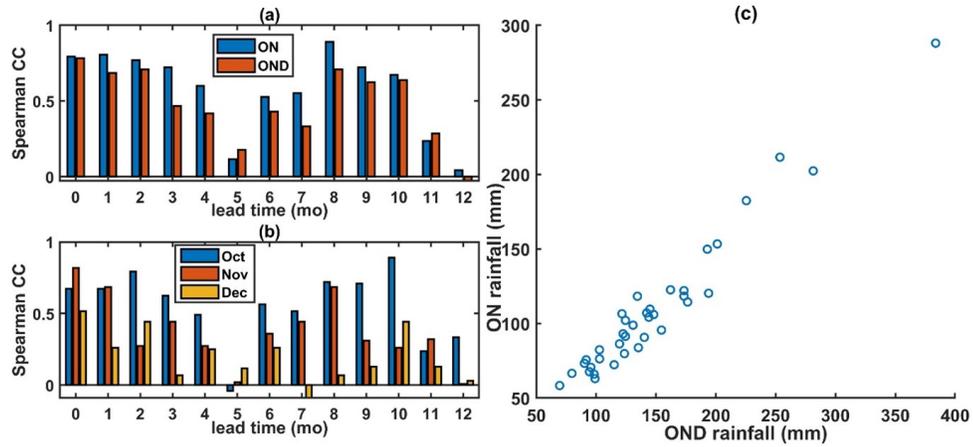


Figure A.1: Model skill at different lead times for the testing period (2008-2017) using the SSTAs to predict rainfall of (a) Oct-Nov (blue) and Oct-Nov-Dec (red) and (b) Oct (blue), Nov (red), and Dec (yellow); (c) Scatter plot of OND rainfall against ON rainfall.

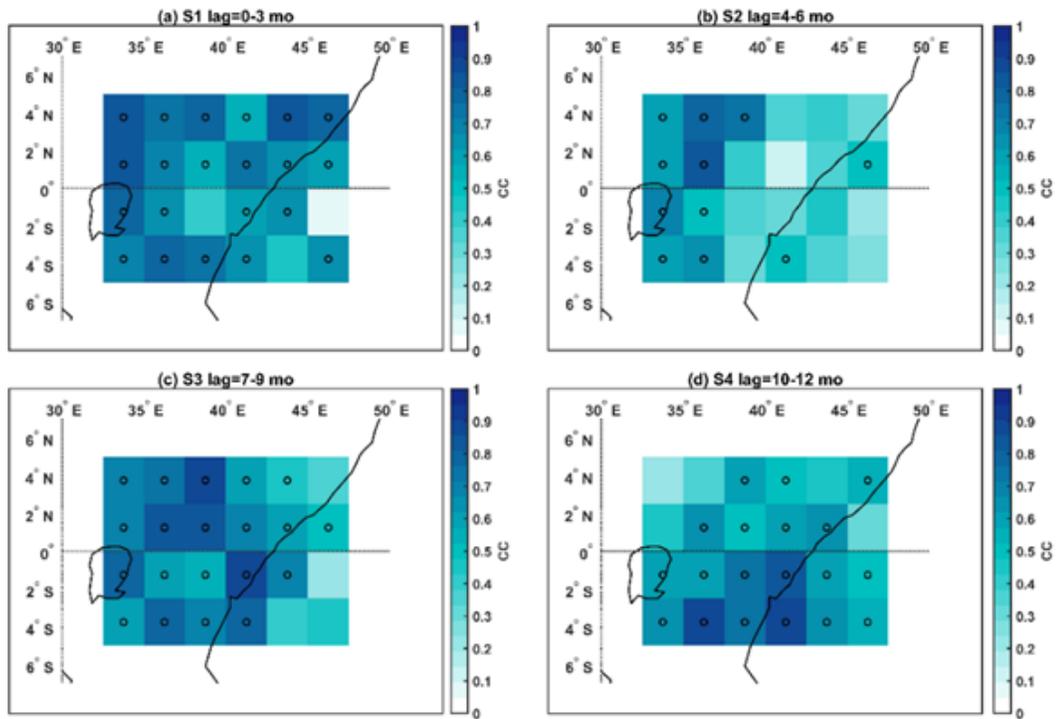


Figure A.2: Maps of the best intra-season prediction skill for different lead time ranges using the GPCP precipitation product. The regression analysis is conducted for lead times ranging from 0-12 months and the best intra-season skill is defined as the highest CC score for the testing period within the corresponding season. The grid cells with significant prediction skill (one-tailed $p = 0.1$) are marked with black circles.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

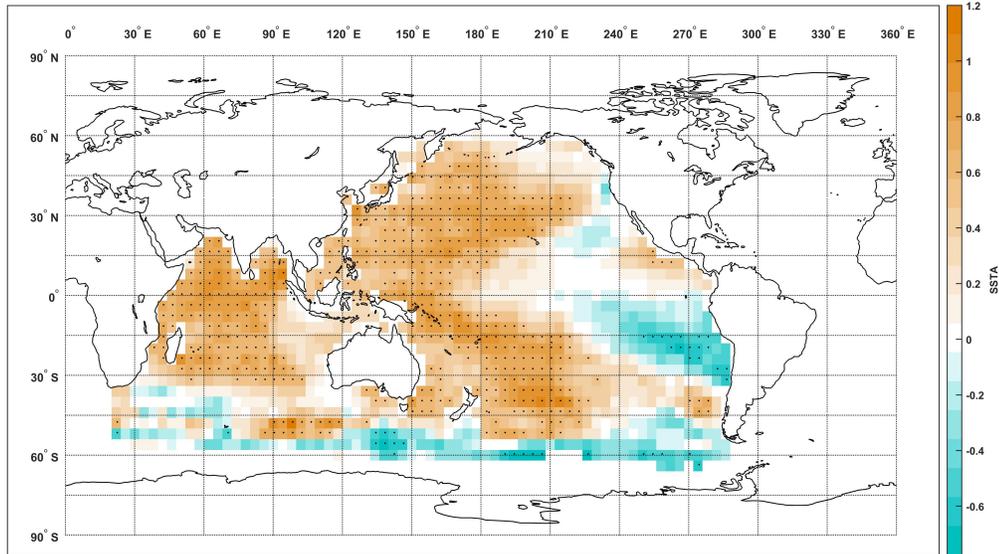


Figure B.1: Difference of averaged SSTAs after and before 2000. The anomalies are based on the whole study period of 1979-2019. Assuming SSTAs are independent normal variables, significant differences (p -value = 0.05 one-tailed) are marked with black dots. 38.1% (511 of 1343 grids) show significantly positive trends while only 2.8% (38 of 1343 grids) show significantly negative trends.

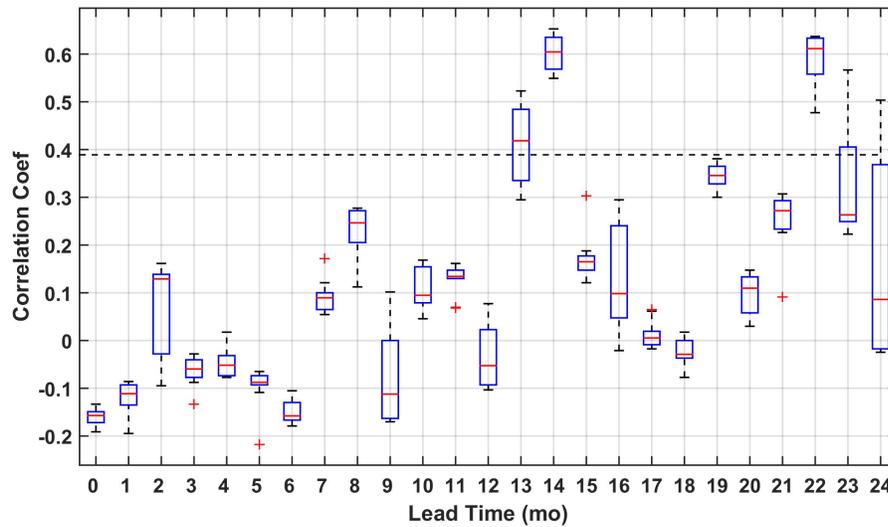


Figure B.2: Correlation coefficients between the observed and estimated TRHR wet-season precipitation over the testing period of 2001-2019 using varying λ s are shown in a box plot. A geometric sequence of 10 λ s in the range of [17, 53] are tested here for developing the elastic net regression models while λ is set to 28 (the MinMSE λ from a 5-fold CV) in the paper. Major analyses in the paper are focused on the two long-term predictive spikes at lead times of 14 and 22 months, and the results show that these two spikes are not sensitive over a rather broad range of λ s. The p -value = 0.1 significance level is plotted in a black dashed line.

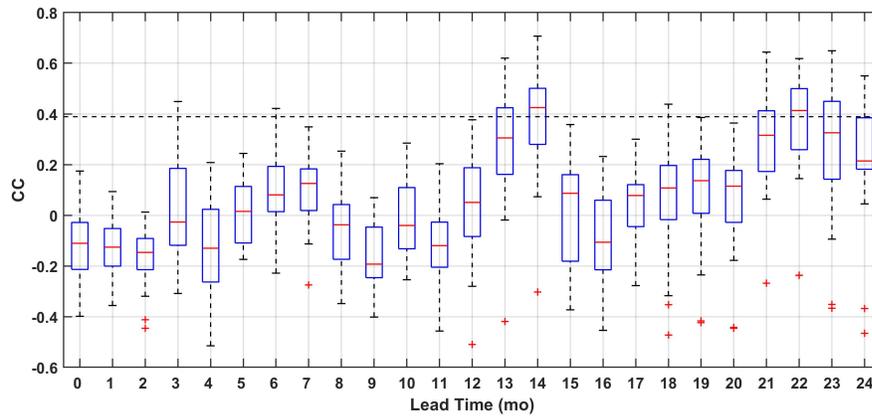


Figure B.3: Box plots of testing period model skill as measured by correlation coefficients with random splitting (repeated for 25 times). The p -value = 0.1 significance level is plotted in a horizontal black dashed line. Model performance is degraded with random splitting. A major reason is that the regression model can only 'learn' the statistical associations from the extreme precipitation events and the extreme events are relatively uniformly distributed across the training and testing period in the original splitting case (1981-2000 for training and 2001-2019 for testing). Extreme events being concentrated in either training or testing period could lead to significant degradation of model performance. And this effect is further exaggerated when correlation coefficient is used for measuring model skill.

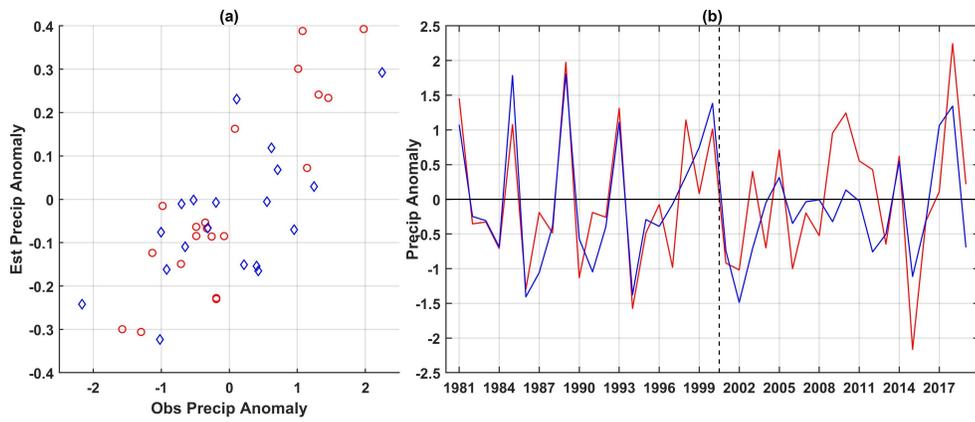


Figure B.4: For a lead time of 22 months, (a) scatter plot of observed precipitation anomalies against the model estimations for the training period (red circles) and the testing period (blue diamonds), and (b) comparison between observed precipitation anomalies (red line) and the rescaled estimations (blue line).

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

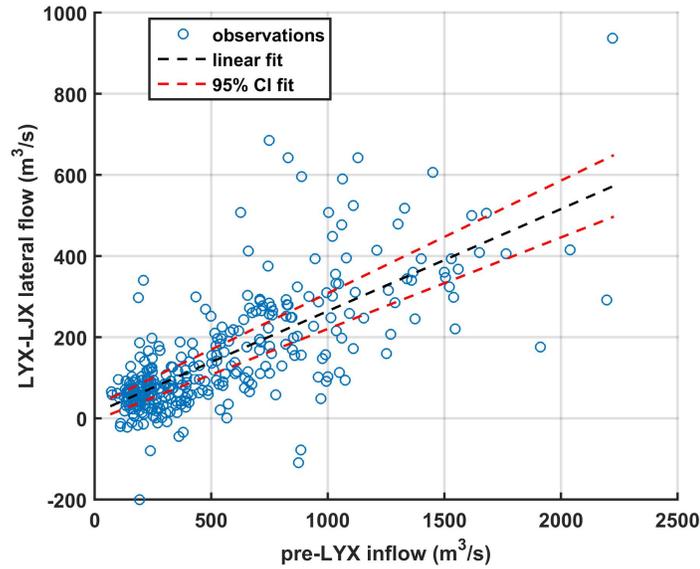


Figure C.1: A scatter plot between 10-day pre-LYX inflow and total LYX-LJX lateral flow based on observational data over 2001-2009. The best linear fit is plotted in a black dashed line and the 95% confidence interval (CI) linear fits are plotted in red dashed lines.

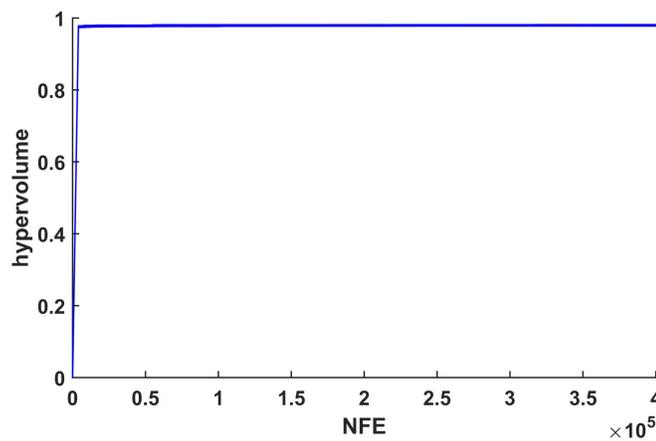


Figure C.2: Rescaled hypervolume as a function of number of function evaluations (NFE) for 50 random seeds

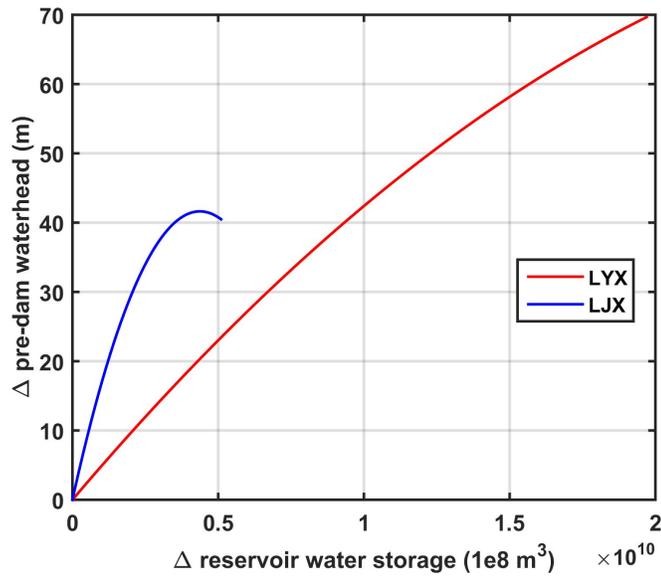


Figure C.3: Comparison between empirical functions for estimating pre-dam water head given water storage for LYX (red) and LJX (blue). For the same amount of increased water storage, pre-dam water head is always raised by a larger margin for LJX.

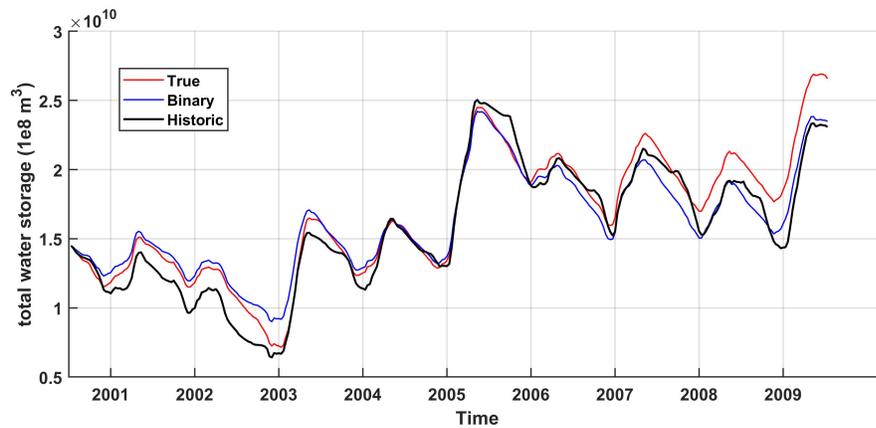


Figure C.4: Comparison between times series of total water storage ($R_{lyx} + R_{ljx}$) determined using historic operations (black) and compromise policies trained with true-amplitude (red) and binary (red) flow data over 2001-2009.

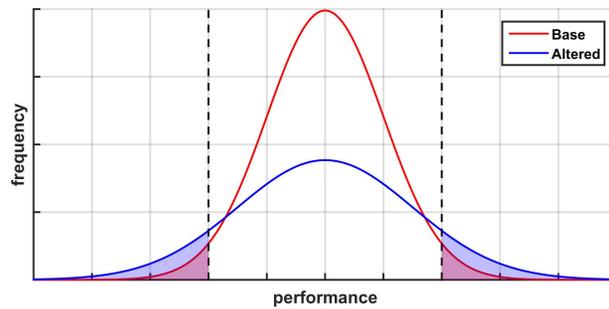


Figure C.5: A schematic of the two-sided effects of altering flow distributions on policy robustness. Increasing probability of extreme flow events will make both drought and flood events more likely and thus change the original distribution of performance (red) to a flatter distribution (blue). Robustness as measured by the S scores (i.e., success ratios) can be improved when compared against a strict baseline performance (vertical black dashed line on the right) and can be degraded when compared against a relaxed baseline performance (vertical black dashed line on left right).

BIBLIOGRAPHY

- [1] National Geophysical Data Center. Etopo-5 bathymetry/topography data, 1988.
- [2] Yuan Si, Xiang Li, Dongqin Yin, Tiejian Li, Ximing Cai, Jiahua Wei, and Guangqian Wang. Revealing the water-energy-food nexus in the upper yellow river basin through multi-objective optimization for reservoir system. *Science of the Total Environment*, 682:1–18, 2019.
- [3] Nicholas P Tatonetti, P Ye Patrick, Roxana Daneshjou, and Russ B Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125):125ra31–125ra31, 2012.
- [4] Justin Grimmer. We are all social scientists now: How big data, machine learning, and causal inference work together. *PS: Political Science & Politics*, 48(1):80–83, 2015.
- [5] Foster Provost and Tom Fawcett. Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1):51–59, 2013.
- [6] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.
- [7] Thomas H Davenport and Randy Bean. Big companies are embracing analytics, but most still don't have a data-driven culture. *Harvard Business Review*, 6:1–4, 2018.
- [8] Penny Noyce, David Perda, and Rob Traver. Creating data-driven schools. *Educational Leadership*, 57(5):52–56, 2000.

- [9] Fan Liang, Vishnupriya Das, Nadiya Kostyuk, and Muzammil M Hus-sain. Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, 10(4):415–453, 2018.
- [10] Hans Von Storch and Francis W Zwiers. *Statistical analysis in climate re-search*. Cambridge university press, 2001.
- [11] Hans Von Storch and Antonio Navarra. *Analysis of climate variability: ap-plications of statistical techniques*. Springer Science & Business Media, 1999.
- [12] JM Gregory, WJ Ingram, MA Palmer, GS Jones, PA Stott, RB Thorpe, JA Lowe, TC Johns, and KD Williams. A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical research letters*, 31(3), 2004.
- [13] K Krishna Kumar, Balaji Rajagopalan, and Mark A Cane. On the weak-ening relationship between the indian monsoon and enso. *Science*, 284 (5423):2156–2159, 1999.
- [14] Stan Yip, Christopher AT Ferro, David B Stephenson, and Ed Hawkins. A simple, coherent framework for partitioning uncertainty in climate pre-dictions. *Journal of Climate*, 24(17):4634–4643, 2011.
- [15] Bryan C Weare, Alfredo R Navato, and Reginald E Newell. Empirical orthogonal analysis of pacific sea surface temperatures. *Journal of Physical Oceanography*, 6(5):671–678, 1976.
- [16] Malte F Stuecker, Axel Timmermann, Fei-Fei Jin, Shayne McGregor, and Hong-Li Ren. A combination mode of the annual cycle and the el niño/southern oscillation. *Nature Geoscience*, 6(7):540–544, 2013.

- [17] Weston E Fuller. Flood flows. *Transactions of the American Society of Civil Engineers*, 77(1):564–617, 1914.
- [18] LeRoy K Sherman. The relation of hydrographs of runoff to size and character of drainage-basins. *Eos, Transactions American Geophysical Union*, 13(1):332–339, 1932.
- [19] Thomas J Schmugge, William P Kustas, Jerry C Ritchie, Thomas J Jackson, and Al Rango. Remote sensing in hydrology. *Advances in water resources*, 25(8-12):1367–1385, 2002.
- [20] Ignacio Rodriguez-Iturbe, Dara Entekhabi, and Rafael L Bras. Nonlinear dynamics of soil moisture at climate scales: 1. stochastic analysis. *Water Resources Research*, 27(8):1899–1906, 1991.
- [21] V Livina, Z Kizner, P Braun, T Molnar, A Bunde, and S Havlin. Temporal scaling comparison of real hydrological data and model runoff records. *Journal of Hydrology*, 336(1-2):186–198, 2007.
- [22] Anna E Sikorska, Alberto Montanari, and Demetris Koutsoyiannis. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. *Journal of Hydrologic Engineering*, 20(1):A4014009, 2015.
- [23] Chi Zhang, Jinggang Chu, and Guangtao Fu. Sobol’s sensitivity analysis for a distributed hydrological model of yichun river basin, china. *Journal of hydrology*, 480:58–68, 2013.
- [24] William W-G Yeh. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water resources research*, 22(2):95–108, 1986.

- [25] Martin Reuss. Is it time to resurrect the harvard water program? *Journal of Water Resources Planning and Management*, 129(5):357–360, 2003.
- [26] Casey M Brown, Jay R Lund, Ximing Cai, Patrick M Reed, Edith A Zagona, Avi Ostfeld, Jim Hall, Gregory W Characklis, Winston Yu, and Levi Brekke. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. *Water resources research*, 51(8):6110–6124, 2015.
- [27] Norman A Phillips. The general circulation of the atmosphere: A numerical experiment. *Quarterly Journal of the Royal Meteorological Society*, 82(352):123–164, 1956.
- [28] Simone Fatichi, Enrique R Vivoni, Fred L Ogden, Valeriy Y Ivanov, Benjamin Mirus, David Gochis, Charles W Downer, Matteo Camporese, Jason H Davison, Brian Ebel, et al. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537:45–60, 2016.
- [29] Saman Razavi, Bryan A Tolson, and Donald H Burn. Review of surrogate modeling in water resources. *Water Resources Research*, 48(7), 2012.
- [30] Jonathan D Herman, Harrison B Zeff, Patrick M Reed, and Gregory W Characklis. Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, 50(10):7692–7713, 2014.
- [31] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and

- B. Zhou. Ipcc, 2021: Summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. 2021.
- [32] M James Salinger. Climate variability and change: past, present and future—an overview. *Increasing climate variability and change*, pages 9–29, 2005.
- [33] Philip K Thornton, Polly J Ericksen, Mario Herrero, and Andrew J Challinor. Climate variability and vulnerability to climate change: a review. *Global change biology*, 20(11):3313–3328, 2014.
- [34] Paul CD Milly, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, and Ronald J Stouffer. Stationarity is dead: Whither water management? *Science*, 319(5863):573–574, 2008.
- [35] Tien Ming Lee, Ezra M Markowitz, Peter D Howe, Chia-Ying Ko, and Anthony A Leiserowitz. Predictors of public climate change awareness and risk perception around the world. *Nature climate change*, 5(11):1014–1020, 2015.
- [36] Yanlai Zhou and Shenglian Guo. Incorporating ecological requirement into multipurpose reservoir operating rule curves for adaptation to climate change. *Journal of Hydrology*, 498:153–164, 2013.
- [37] Ryan R Morrison and Mark C Stone. Evaluating the impacts of environmental flow alternatives on reservoir and recreational operations using system dynamics modeling. *JAWRA Journal of the American Water Resources Association*, 51(1):33–46, 2015.

- [38] Bo Ming, Pan Liu, Shenglian Guo, Xiaoqi Zhang, Maoyuan Feng, and Xi-anxun Wang. Optimizing utility-scale photovoltaic power generation for integration into a hydropower reservoir by incorporating long-and short-term operational decisions. *Applied Energy*, 204:432–445, 2017.
- [39] Yusheng Zhang, Jijian Lian, Chao Ma, Yang Yang, Xiulan Pang, and Lun Wang. Optimal sizing of the grid-connected hybrid system integrating hydropower, photovoltaic, and wind considering cascade reservoir connection and photovoltaic-wind complementarity. *Journal of Cleaner Production*, 274:123100, 2020.
- [40] Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, 2019.
- [41] Jiali Wang, Prasanna Balaprakash, and Rao Kotamarthi. Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model. *Geoscientific Model Development*, 12(10):4261–4274, 2019.
- [42] Yaochu Jin. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement. *IEEE Transactions on Fuzzy Systems*, 8(2):212–221, 2000.
- [43] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- [44] Bruce A McCarl, Xavier Villavicencio, and Ximing Wu. Climate change and future analysis: is stationarity dying? *American Journal of Agricultural Economics*, 90(5):1241–1247, 2008.

- [45] Harvey J Miller and Michael F Goodchild. Data-driven geography. *Geo-Journal*, 80(4):449–461, 2015.
- [46] Herbert A Simon. Spurious correlation: A causal interpretation. *Journal of the American statistical Association*, 49(267):467–479, 1954.
- [47] Nikesh Patel, Jake Nease, Siam Aumi, Christopher Ewaschuk, Jie Luo, and Prashant Mhaskar. Integrating data-driven modeling with first-principles knowledge. *Industrial & Engineering Chemistry Research*, 59(11):5103–5113, 2020.
- [48] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919*, 2020.
- [49] Debbie Hillier and Benedict Dempsey. A dangerous delay: The cost of late response to early warnings in the 2011 drought in the horn of africa. *Oxfam Policy and Practice: Agriculture, Food and Land*, 12(1):1–34, 2012. ISSN 2053-0218.
- [50] Barry J Barnett, Christopher B Barrett, and Jerry R Skees. Poverty traps and index-based risk transfer products. *World Development*, 36(10):1766–1785, 2008.
- [51] Christina Oelfke Clark, Peter J Webster, and Julia E Cole. Interdecadal variability of the relationship between the indian ocean zonal mode and east african coastal rainfall anomalies. *Journal of Climate*, 16(3):548–554, 2003. ISSN 1520-0442.
- [52] C. C. Mutai, M. N. Ward, and A. W. Colman. Towards the prediction of the east africa short rains based on sea-surface temperature-atmosphere

- coupling. *International Journal of Climatology*, 18(9):975–997, 1998. ISSN 0899-8418.
- [53] S. Hastenrath, A. Nicklis, and L. Greischar. Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial indian-ocean. *Journal of Geophysical Research-Oceans*, 98(C11):20219–20235, 1993. ISSN 0148-0227.
- [54] S. Hastenrath, D. Polzin, and P. Camberlin. Exploring the predictability of the ‘short rains’ at the coast of east africa. *International Journal of Climatology*, 24(11):1333–1343, 2004. ISSN 0899-8418.
- [55] S. Hastenrath. Circulation mechanisms of climate anomalies in east africa and the equatorial indian ocean. *Dynamics of Atmospheres and Oceans*, 43(1-2):25–35, 2007. ISSN 0377-0265.
- [56] N. H. Saji, B. N. Goswami, P. N. Vinayachandran, and T. Yamagata. A dipole mode in the tropical indian ocean. *Nature*, 401(6751):360–363, 1999. ISSN 0028-0836.
- [57] Emily Black, Julia Slingo, and Kenneth R Sperber. An observational study of the relationship between excessively strong short rains in coastal east africa and indian ocean sst. *Monthly Weather Review*, 131(1):74–94, 2003.
- [58] C. C. Ummenhofer, A. Sen Gupta, M. H. England, and C. J. C. Reason. Contributions of indian ocean sea surface temperatures to enhanced east african rainfall. *Journal of Climate*, 22(4):993–1013, 2009. ISSN 0894-8755.
- [59] TK Bahaga, G Mengistu Tsidu, F Kucharski, and GT Diro. Potential predictability of the sea-surface temperature forced equatorial east african

- short rains interannual variability in the 20th century. *Quarterly Journal of the Royal Meteorological Society*, 141(686):16–26, 2015.
- [60] S. Hastenrath, D. Polzin, and C. Mutai. Circulation mechanisms of kenya rainfall anomalies. *Journal of Climate*, 24(2):404–412, 2011. ISSN 0894-8755.
- [61] S. E. Nicholson. Long-term variability of the east african ‘short rains’ and its links to large-scale factors. *International Journal of Climatology*, 35(13): 3979–3990, 2015. ISSN 0899-8418.
- [62] Titike K Bahaga, Andreas H Fink, and Peter Knippertz. Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the greater horn of africa during the 20th century. *International Journal of Climatology*, 39(5):2765–2785, 2019.
- [63] Swadhin K Behera, Jing-Jia Luo, Sebastien Masson, Pascale Delecluse, Silvio Gualdi, Antonio Navarra, and Toshio Yamagata. Paramount impact of the indian ocean dipole on the east african short rains: A cgcm study. *Journal of Climate*, 18(21):4514–4530, 2005.
- [64] S. E. Nicholson. The predictability of rainfall over the greater horn of africa. part i: Prediction of seasonal rainfall. *Journal of Hydrometeorology*, 15(3):1011–1027, 2014. ISSN 1525-755x.
- [65] Dean P Walker, Cathryn E Birch, John H Marsham, Adam A Scaife, Richard J Graham, and Zewdu T Segele. Skill of dynamical and ghacof consensus seasonal forecasts of east african rainfall. *Climate Dynamics*, pages 1–25, 2019.
- [66] J. E. Tierney, J. E. Smerdon, K. J. Anchukaitis, and R. Seager. Multidecadal

- variability in east african hydroclimate controlled by the indian ocean. *Nature*, 493(7432):389–392, 2013. ISSN 0028-0836.
- [67] T. Tozuka, S. Endo, and T. Yamagata. Anomalous walker circulations associated with two flavors of the indian ocean dipole. *Geophysical Research Letters*, 43(10):5378–5384, 2016. ISSN 0094-8276.
- [68] A. G. Turner and H. Annamalai. Climate change and the south asian summer monsoon. *Nature Climate Change*, 2(8):587–595, 2012. ISSN 1758-678x.
- [69] F. Xue, H. J. Wang, and J. H. He. Interannual variability of mascarene high and australian high and their influences on east asian summer monsoon. *Journal of the Meteorological Society of Japan*, 82(4):1173–1186, 2004. ISSN 0026-1165.
- [70] D. Manatsa, Y. Morioka, S. K. Behera, C. H. Matarira, and T. Yamagata. Impact of mascarene high variability on the east african ‘short rains’. *Climate Dynamics*, 42(5-6):1259–1274, 2014. ISSN 0930-7575.
- [71] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 2015. ISSN 2052-4463.
- [72] Margaret Kimani, Joost Hoedjes, and Zhongbo Su. An assessment of satellite-derived rainfall products relative to ground observations over east africa. *Remote sensing*, 9(5):430, 2017.
- [73] Wenchang Yang, Richard Seager, Mark A Cane, and Bradfield Lyon. The

- annual cycle of east african precipitation. *Journal of Climate*, 28(6):2385–2404, 2015.
- [74] Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. American Meteorological Society Boston, MA, 1993.
- [75] N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research-Atmospheres*, 108(D14), 2003. ISSN 2169-897x.
- [76] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 67: 301–320, 2005. ISSN 1369-7412.
- [77] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [78] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B-Methodological*, 58(1):267–288, 1996. ISSN 0035-9246.
- [79] L. Breiman and P. Spector. Submodel selection and evaluation in regression - the x-random case. *International Statistical Review*, 60(3):291–319, 1992. ISSN 0306-7734.
- [80] Adel Javanmard and Andrea Montanari. Confidence intervals and hy-

- pothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.
- [81] Theresa M Carpenter and Konstantine P Georgakakos. Assessment of folsom lake response to historical and potential future climate scenarios: 1. forecasting. *Journal of Hydrology*, 249(1-4):148–175, 2001.
- [82] P Block. Tailoring seasonal climate forecasts for hydropower operations. *Hydrology and Earth System Sciences*, 15(4):1355–1368, 2011.
- [83] Maria Carmen Lemos, Timothy J Finan, Roger W Fox, Donald R Nelson, and Joanna Tucker. The use of seasonal climate forecasting in policymaking: lessons from northeast brazil. *Climatic Change*, 55(4):479–507, 2002.
- [84] James W Hansen, Simon J Mason, Liqiang Sun, and Arame Tall. Review of seasonal climate forecasting for agriculture in sub-saharan africa. *Experimental Agriculture*, 47(2):205–240, 2011.
- [85] Donald A Wilhite and Mark D Svoboda. Drought early warning systems in the context of drought preparedness and mitigation. *Early warning systems for drought preparedness and drought management*, pages 1–21, 2000.
- [86] James Verdin, Chris Funk, Gabriel Senay, and Richard Choularton. Climate science and famine early warning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):2155–2168, 2005.
- [87] Shang-Ping Xie, Kaiming Hu, Jan Hafner, Hiroki Tokinaga, Yan Du, Gang Huang, and Takeaki Sampe. Indian ocean capacitor effect on indo-western pacific climate during the summer following el niño. *Journal of Climate*, 22(3):730–747, 2009.

- [88] Shang-Ping Xie, Yu Kosaka, Yan Du, Kaiming Hu, Jasti S Chowdary, and Gang Huang. Indo-western pacific ocean capacitor and coherent climate anomalies in post-enso summer: A review. *Advances in Atmospheric Sciences*, 33(4):411–432, 2016.
- [89] Eugene M Rasmusson and Thomas H Carpenter. Variations in tropical sea surface temperature and surface wind fields associated with the southern oscillation/el niño. *Monthly Weather Review*, 110(5):354–384, 1982.
- [90] Kevin E Trenberth. The definition of el nino. *Bulletin of the American Meteorological Society*, 78(12):2771–2778, 1997.
- [91] Kevin E Trenberth and David P Stepaniak. Indices of el niño evolution. *Journal of climate*, 14(8):1697–1701, 2001.
- [92] Yuan Zhang, John M Wallace, and David S Battisti. Enso-like interdecadal variability: 1900–93. *Journal of climate*, 10(5):1004–1020, 1997.
- [93] Nathan J Mantua, Steven R Hare, Yuan Zhang, John M Wallace, and Robert C Francis. A pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society*, 78(6):1069–1080, 1997.
- [94] David B Enfield, Alberto M Mestas-Nuñez, Dennis A Mayer, and Luis Cid-Serrano. How ubiquitous is the dipole relationship in tropical atlantic sea surface temperatures? *Journal of Geophysical Research: Oceans*, 104(C4):7841–7848, 1999.
- [95] Dimitris Menemenlis, Ichiro Fukumori, and Tong Lee. Using green’s functions to calibrate an ocean general circulation model. *Monthly weather review*, 133(5):1224–1240, 2005.

- [96] Rishi Sahastrabudde and Subimal Ghosh. Does statistical model perform at par with computationally expensive general circulation model for decadal prediction? *Environmental Research Letters*, 16(6):064028, 2021.
- [97] Peng Liu. A survey of remote-sensing big data. *frontiers in Environmental Science*, 3:45, 2015.
- [98] Charles Warren Thornthwaite. An approach toward a rational classification of climate. *Geographical review*, 38(1):55–94, 1948.
- [99] Mumtaz Ali, Ravinesh C Deo, Tek Maraseni, and Nathan J Downs. Improving spi-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *Journal of hydrology*, 576:164–184, 2019.
- [100] Diego Bueso, Maria Piles, and Gustau Camps-Valls. Nonlinear pca for spatio-temporal analysis of earth observation data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8):5752–5763, 2020.
- [101] Naresh Devineni and A Sankarasubramanian. Improving the prediction of winter precipitation and temperature over the continental united states: Role of the enso state in developing multimodel combinations. *Monthly Weather Review*, 138(6):2447–2468, 2010.
- [102] Andrew Schepen, Tongtiegang Zhao, Quan J Wang, and David E Robertson. A bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 australian catchments. *Hydrology and Earth System Sciences*, 22(2): 1615–1628, 2018.

- [103] Lisa Goddard, Simon J Mason, Stephen E Zebiak, Chester F Ropelewski, Reid Basher, and Mark A Cane. Current approaches to seasonal to inter-annual climate predictions. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 21(9):1111–1152, 2001.
- [104] F Mekanik, MA Imteaz, S Gato-Trinidad, and A Elmahdi. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, 503:11–21, 2013.
- [105] YR Fan, W Huang, GH Huang, Z Li, YP Li, XQ Wang, GH Cheng, and L Jin. A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections. *Stochastic environmental research and risk assessment*, 29(6):1557–1569, 2015.
- [106] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.
- [107] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.
- [108] Scott C Worland, Scott Steinschneider, William Asquith, Rodney Knight, and Michael Wiczorek. Prediction and inference of flow duration curves using multioutput neural networks. *Water Resources Research*, 55(8):6850–6868, 2019.
- [109] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. Machine

- learning interpretability: A survey on methods and metrics. *Electronics*, 8 (8):832, 2019.
- [110] Andrew R Solow. Testing for climate change: An application of the two-phase regression model. *Journal of Applied Meteorology and Climatology*, 26 (10):1401–1405, 1987.
- [111] Manfred Mudelsee. Ramp function regression: a tool for quantifying climate transitions. *Computers & Geosciences*, 26(3):293–307, 2000.
- [112] TN Krishnamurti, Chandra M Kishtawal, Timothy E LaRow, David R Bachiochi, Zhan Zhang, C Eric Williford, Sulochana Gadgil, and Sajani Surendran. Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, 285(5433):1548–1550, 1999.
- [113] Viatcheslav V Kharin and Francis W Zwiers. Climate predictions with multimodel ensembles. *Journal of Climate*, 15(7):793–799, 2002.
- [114] Seiji Wakabayashi and Ryuichi Kawamura. Extraction of major teleconnection patterns possibly associated with the anomalous summer climate in japan. *Journal of the Meteorological Society of Japan. Ser. II*, 82(6):1577–1588, 2004.
- [115] Hidetoshi Matsui and Sadanori Konishi. Variable selection for functional regression models via the l1 regularization. *Computational Statistics & Data Analysis*, 55(12):3304–3310, 2011.
- [116] James W Hurrell. Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophysical Research Letters*, 23(6):665–668, 1996.

- [117] Geert Jan Van Oldenborgh and Gerrit Burgers. Searching for decadal variations in enso precipitation teleconnections. *Geophysical Research Letters*, 32(15), 2005.
- [118] Xiaosong Yang and Timothy DelSole. Systematic comparison of enso teleconnection patterns between models and observations. *Journal of climate*, 25(2):425–446, 2012.
- [119] Shuyu Zhang, Thian Yew Gan, and Andrew BG Bush. Variability of arctic sea ice based on quantile regression and the teleconnection with large-scale climate patterns. *Journal of Climate*, 33(10):4009–4025, 2020.
- [120] Justin T Schoof and S C Pryor. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 21(7):773–790, 2001.
- [121] Jingmin Li, Felix Pollinger, and Heiko Paeth. Comparing the lasso predictor-selection and regression method with classical approaches of precipitation bias adjustment in decadal climate predictions. *Monthly Weather Review*, 148(10):4339–4351, 2020.
- [122] Kingtse C Mo. Ensemble canonical correlation prediction of surface temperature over the united states. *Journal of climate*, 16(11):1665–1683, 2003.
- [123] Henning W Rust, Andy Richling, Peter Bissolli, and Uwe Ulbrich. Linking teleconnection patterns to european temperature: a multiple linear regression model. 2015.
- [124] Xuezhi Tan and Dongguo Shao. Precipitation trends and teleconnections

- identified using quantile regressions over xinjiang, china. *International Journal of Climatology*, 37(3):1510–1525, 2017.
- [125] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.
- [126] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- [127] Agus M Soleh, Aji H Wigena, Anik Djuraidah, and Asep Saefuddin. Statistical downscaling to predict monthly rainfall using linear regression with l1 regularization (lasso). *Applied Mathematical Sciences*, 9(108):5361–5369, 2015.
- [128] Timothy DelSole and Arindam Banerjee. Statistical seasonal prediction based on regularized regression. *Journal of Climate*, 30(4):1345–1361, 2017.
- [129] Joonpyo Kim, Hee-Seok Oh, Yaeji Lim, and Hyun-Suk Kang. Seasonal precipitation prediction via data-adaptive principal component regression. *International Journal of Climatology*, 37:75–86, 2017.
- [130] Abby Stevens, Rebecca Willett, Antonios Mamalakis, Efi Foufoula-Georgiou, Alejandro Tejedor, James T Randerson, Padhraic Smyth, and Stephen Wright. Graph-guided regularized regression of pacific ocean climate variables to increase predictive skill of southwestern us winter precipitation. *Journal of Climate*, 34(2):737–754, 2021.
- [131] Xiao Peng, Scott Steinschneider, and John Albertson. Investigating long-range seasonal predictability of east african short rains: influence of the

- mascarene high on the indian ocean walker cell. *Journal of Applied Meteorology and Climatology*, 59(6):1077–1090, 2020.
- [132] Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557*, 2013.
- [133] Dingjun Yu, Hanli Wang, Peiqiu Chen, and Zhihua Wei. Mixed pooling for convolutional neural networks. In *International conference on rough sets and knowledge technology*, pages 364–375. Springer, 2014.
- [134] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2014.
- [135] Douglas I Benn and Lewis A Owen. The role of the indian summer monsoon and the mid-latitude westerlies in himalayan glaciation: review and speculative discussion. *Journal of the Geological Society*, 155(2):353–363, 1998.
- [136] Jeffrey Shaman and Eli Tziperman. The effect of enso on tibetan plateau snow depth: A stationary wave teleconnection mechanism and implications for the south asian monsoons. *Journal of Climate*, 18(12):2067–2079, 2005.
- [137] Lei Feng and Tianjun Zhou. Water vapor transport for summer precipitation over the tibetan plateau: Multidata set analysis. *Journal of Geophysical Research: Atmospheres*, 117(D20), 2012.
- [138] Yiyang Dong, Jiaqi Zhai, Yong Zhao, Haihong Li, Qingming Wang, Shan Jiang, Huanyu Chang, and Zhiyong Ding. Teleconnection patterns of pre-

- precipitation in the three-river headwaters region, china. *Environmental Research Letters*, 15(10):104050, 2020.
- [139] Yu Zhang, Wenyu Huang, and Deyu Zhong. Major moisture pathways and their importance to rainy season precipitation over the sanjiangyuan region of the tibetan plateau. *Journal of Climate*, 32(20):6837–6857, 2019.
- [140] Yang Zhao, Xiangde Xu, Liufeng Liao, Yuhong Wang, Xiaoping Gu, Rui Qin, Yunyun Guo, Zhaoping Kang, Fang Wang, and Min Wang. The severity of drought and precipitation prediction in the eastern fringe of the tibetan plateau. *Theoretical and Applied Climatology*, 137(1-2):141–152, 2019.
- [141] Chen Chen, Tiejian Li, Bellie Sivakumar, Jiaye Li, and Guangqian Wang. Attribution of growing season vegetation activity to climate change and human activities in the three-river headwaters region, china. *Journal of Hydroinformatics*, 22(1):186–204, 2020.
- [142] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [143] Mingdong Sun and Gwangseob Kim. Quantitative monthly precipitation forecasting using cyclostationary empirical orthogonal function and canonical correlation analysis. *Journal of Hydrologic Engineering*, 21(1): 04015045, 2016.
- [144] Shayne McGregor, Axel Timmermann, Malte F Stuecker, Matthew H England, Mark Merrifield, Fei-Fei Jin, and Yoshimitsu Chikamoto. Recent

walker circulation strengthening and pacific cooling amplified by atlantic warming. *Nature Climate Change*, 4(10):888–892, 2014.

- [145] Bahram Choubin, Gholamreza Zehtabian, Ali Azareh, Elham Rafiei-Sardooi, Farzaneh Sajedi-Hosseini, and Özgür Kişi. Precipitation forecasting using classification and regression trees (cart) model: a comparative study of different approaches. *Environmental earth sciences*, 77(8):1–13, 2018.
- [146] Liexing Huang, Junfeng Kang, Mengxue Wan, Lei Fang, Chunyan Zhang, and Zhaoliang Zeng. Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Frontiers in Earth Science*, 9:202, 2021.
- [147] Stuart G Baker. The multinomial-poisson transformation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(4):495–504, 1994.
- [148] Zhiqiang Jiang, Rongbo Li, Anqiang Li, and Changming Ji. Runoff forecast uncertainty considered load adjustment model of cascade hydropower stations and its application. *Energy*, 158:693–708, 2018.
- [149] Reepal Shah, Atul Kumar Sahai, and Vimal Mishra. Short to sub-seasonal hydrologic forecast to manage water and agricultural resources in india. *Hydrology and Earth System Sciences*, 21(2):707–720, 2017.
- [150] Florian Pappenberger, Keith J Beven, NM Hunter, PD Bates, Bingo T Gouweleeuw, J Thielen, and APJ De Roo. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the european flood forecasting system (effs). *Hydrology and Earth System Sciences*, 9(4):381–393, 2005.

- [151] Felix Fundel, Stefanie Jörg-Hess, and Massimiliano Zappa. Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. *Hydrology and Earth System Sciences*, 17(1): 395–407, 2013.
- [152] Juraj M Cunderlik and Taha BMJ Ouarda. Trends in the timing and magnitude of floods in canada. *Journal of hydrology*, 375(3-4):471–480, 2009.
- [153] Hong Xuan Do, Fang Zhao, Seth Westra, Michael Leonard, Lukas Gudmundsson, Julien Eric Stanislas Boulange, Jinfeng Chang, Philippe Ciais, Dieter Gerten, Simon N Gosling, et al. Historical and future changes in global flood magnitude—evidence from a model—observation investigation. *Hydrology and Earth System Sciences*, 24(3):1543–1564, 2020.
- [154] Alberto Montanari and Armando Brath. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 40(1), 2004.
- [155] Hamid Moradkhani, Soroosh Sorooshian, Hoshin V Gupta, and Paul R Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in water resources*, 28(2):135–147, 2005.
- [156] Yan Ma, Haiping Wu, Lizhe Wang, Bormin Huang, Rajiv Ranjan, Albert Zomaya, and Wei Jie. Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51:47–60, 2015.
- [157] Mingmin Chi, Antonio Plaza, Jon Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11):2207–2219, 2016.

- [158] Marwan A Hassan, Roey Egozi, and Gary Parker. Experiments on the effect of hydrograph characteristics on vertical grain sorting in gravel bed rivers. *Water Resources Research*, 42(9), 2006.
- [159] Griet Heuvelmans, Bart Muys, and Jan Feyen. Evaluation of hydrological model parameter transferability for simulating the impact of land use on catchment hydrology. *Physics and Chemistry of the Earth, Parts A/B/C*, 29(11-12):739–747, 2004.
- [160] Vahid Nourani, Ahmad Fakheri Fard, Hoshin V Gupta, David C Goodrich, and Faegheh Niazi. Hydrological model parameterization using ndvi values to account for the effects of land cover change on the rainfall–runoff response. *Hydrology Research*, 48(6):1455–1473, 2017.
- [161] Soroosh Sorooshian, Vijai Kumar Gupta, and James Lloyd Fulton. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility. *Water Resources Research*, 19(1):251–259, 1983.
- [162] Simone Castiglioni, Laura Lombardi, Elena Toth, Attilio Castellarin, and Alberto Montanari. Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. *Advances in Water Resources*, 33(10):1235–1242, 2010.
- [163] Bryson C Bates and Edward P Campbell. A markov chain monte carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water resources research*, 37(4):937–947, 2001.
- [164] Tyler Jon Smith and Lucy Amanda Marshall. Bayesian methods in hydro-

- logic modeling: A study of recent advancements in markov chain monte carlo techniques. *Water Resources Research*, 44(12), 2008.
- [165] Samuel P Nolan, Luca Pezzè, and Augusto Smerzi. Frequentist parameter estimation with supervised learning. *arXiv preprint arXiv:2105.12302*, 2021.
- [166] Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial*. OUP Oxford, 2006.
- [167] Mary Kathryn Cowles and Bradley P Carlin. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.
- [168] Bin Ba, Weijia Cui, Daming Wang, and Jianhui Wang. Maximum likelihood time delay estimation based on monte carlo importance sampling in multipath environment. *International Journal of Antennas and Propagation*, 2017, 2017.
- [169] Gayathri K Devia, B Pa Ganasri, and G Sa Dwarakish. A review on hydrological models. *Aquatic procedia*, 4:1001–1007, 2015.
- [170] Guillermo J Vicens, Ignacio Rodriguez-Iturbe, and John C Schaake Jr. A bayesian framework for the use of regional information in hydrology. *Water Resources Research*, 11(3):405–414, 1975.
- [171] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.
- [172] GR Gavalas, PC Shah, and John H Seinfeld. Reservoir history matching by bayesian estimation. *Society of Petroleum Engineers Journal*, 16(06):337–350, 1976.

- [173] Peter K Kitanidis. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water resources research*, 22(4):499–507, 1986.
- [174] Hamid Moradkhani, Kuo-Lin Hsu, Hoshin Gupta, and Soroosh Sorooshian. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water resources research*, 41(5), 2005.
- [175] Gabriele Manoli, Matteo Rossi, Damiano Pasetto, Rita Deiana, Stefano Ferraris, Giorgio Cassiani, and Mario Putti. An iterative particle filter approach for coupled hydro-geophysical inversion of a controlled infiltration experiment. *Journal of Computational Physics*, 283:37–51, 2015.
- [176] Peyman Abbaszadeh, Hamid Moradkhani, and Hongxiang Yan. Enhancing hydrologic data assimilation by evolutionary particle filter and markov chain monte carlo. *Advances in Water Resources*, 111:192–204, 2018.
- [177] James O Knighton, Arthur DeGaetano, and M Todd Walter. Hydrologic state influence on riverine flood discharge for a small temperate watershed (fall creek, united states): negative feedbacks on the effects of climate change. *Journal of Hydrometeorology*, 18(2):431–449, 2017.
- [178] Zachary M Easton, Pierre Gérard-Marchant, M Todd Walter, A Martin Petrovic, and Tammo S Steenhuis. Hydrologic assessment of an urban variable source watershed in the northeast united states. *Water Resources Research*, 43(3), 2007.
- [179] Helen E Dahlke, Zachary M Easton, Daniel R Fuka, Steve W Lyon, and Tammo S Steenhuis. Modelling variable source area dynamics in a ceap

- watershed. *Ecohydrology: Ecosystems, Land and Water Process Interactions, Ecohydrogeomorphology*, 2(3):337–349, 2009.
- [180] U.S. Geological Survey. Usgs water data for the nation, 1994. URL <https://waterdata.usgs.gov/nwis>.
- [181] Imke Durre, Matthew J Menne, and Russell S Vose. Strategies for evaluating quality assurance procedures. *Journal of Applied Meteorology and Climatology*, 47(6):1785–1791, 2008.
- [182] Imke Durre, Matthew J Menne, Byron E Gleason, Tamara G Houston, and Russell S Vose. Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, 49(8):1615–1633, 2010.
- [183] Matthew J Menne, Imke Durre, Russell S Vose, Byron E Gleason, and Tamara G Houston. An overview of the global historical climatology network-daily database. *Journal of atmospheric and oceanic technology*, 29(7):897–910, 2012.
- [184] Pingping Xie, Mingyue Chen, Song Yang, Akiyo Yatagai, Tadahiro Hayasaka, Yoshihiro Fukushima, and Changming Liu. A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007.
- [185] Mingyue Chen, Wei Shi, Pingping Xie, Viviane BS Silva, Vernon E Kousky, R Wayne Higgins, and John E Janowiak. Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres*, 113(D4), 2008.

- [186] Pingping Xie, M Chen, and W Shi. Cpc unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc*, volume 2, 2010.
- [187] Jerry R Stedinger, Richard M Vogel, Seung Uk Lee, and Rebecca Batchelder. Appraisal of the generalized likelihood uncertainty estimation (glue) method. *Water resources research*, 44(12), 2008.
- [188] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.
- [189] Sean R Eddy. What is a hidden markov model? *Nature biotechnology*, 22(10):1315–1316, 2004.
- [190] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.
- [191] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335, 1995.
- [192] Ziheng Yang and Carlos E Rodríguez. Searching for efficient markov chain monte carlo proposal kernels. *Proceedings of the National Academy of Sciences*, 110(48):19307–19312, 2013.
- [193] Yuttapong Thawornwattana, Daniel Dalquen, and Ziheng Yang. Designing simple and efficient markov chain monte carlo proposal kernels. *Bayesian Analysis*, 13(4):1037–1063, 2018.
- [194] John Maynard Keynes. Chapter iv: The principle of indifference. volume 4, pages 41–64. Macmillan and Company, limited, 1921.
- [195] Hakon Tjelmeland and Bjorn Kare Hegstad. Mode jumping proposals in mcmc. *Scandinavian journal of statistics*, 28(1):205–223, 2001.

- [196] Vijay K Gupta, Ed Waymire, and CT Wang. A representation of an instantaneous unit hydrograph from geomorphology. *Water resources research*, 16(5):855–862, 1980.
- [197] AJ Jakeman, IG Littlewood, and PG Whitehead. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of hydrology*, 117(1-4): 275–300, 1990.
- [198] Kwan Tun Lee and Ben Chie Yen. Geomorphology and kinematic-wave-based hydrograph derivation. *Journal of Hydraulic Engineering*, 123(1):73–80, 1997.
- [199] S Grimaldi, A Petroselli, and F Nardi. A parsimonious geomorphological unit hydrograph for rainfall–runoff modelling in small ungauged basins. *Hydrological Sciences Journal*, 57(1):73–83, 2012.
- [200] Diego da Ros and Marco Borga. Use of digital elevation model data for the derivation of the geomorphological instantaneous unit hydrograph. *Hydrological processes*, 11(1):13–33, 1997.
- [201] James W Kirchner, Xiahong Feng, and Colin Neal. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature*, 403(6769):524–527, 2000.
- [202] PK Bhunya, SK Mishra, and Ronny Berndtsson. Simplified two-parameter gamma distribution for derivation of synthetic unit hydrograph. *Journal of Hydrologic Engineering*, 8(4):226–230, 2003.
- [203] Soroosh Afyouni, Stephen M Smith, and Thomas E Nichols. Effective

- degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, 199:609–625, 2019.
- [204] DB King and OH Shemdin. Radar observation of hurricane wave directions. In *Coastal Engineering 1978*, pages 209–226. 1978.
- [205] Nicholls W Bowden. General problems of design and operation. *Transactions of the American Society of Civil Engineers*, 115(1):803–817, 1950.
- [206] Mohamad I Hejazi, Ximing Cai, and Benjamin L Ruddell. The role of hydrologic information in reservoir operation—learning from historical releases. *Advances in water resources*, 31(12):1636–1650, 2008.
- [207] Warren A Hall and Theodore G Roefs. Hydropower project output optimization. *Journal of the Power Division*, 92(1):67–79, 1966.
- [208] William S Butcher. Stochastic dynamic programming for optimum reservoir operation 1. *JAWRA Journal of the American Water Resources Association*, 7(1):115–123, 1971.
- [209] Ricardo C Harboe, Fereidoun Mobasheri, and William WG Yeh. Optimal policy for reservoir operation. *Journal of the Hydraulics Division*, 96(11):2297–2308, 1970.
- [210] Jazmin Zatarain Salazar, Patrick M Reed, Jonathan D Herman, Matteo Giuliani, and Andrea Castelletti. A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in water resources*, 92:172–185, 2016.
- [211] World Bank. *World development report 2009: Reshaping economic geography*. The World Bank, 2008.

- [212] Hui Wang, Earl D Brill, Ranji S Ranjithan, and A Sankarasubramanian. A framework for incorporating ecological releases in single reservoir operation. *Advances in Water Resources*, 78:9–21, 2015.
- [213] Cui Xu, Zhihao Xu, and Zhifeng Yang. Reservoir operation optimization for balancing hydropower generation and biodiversity conservation in a downstream wetland. *Journal of Cleaner Production*, 245:118885, 2020.
- [214] Hannu Lauri, H de Moel, Philip J Ward, Timo A Räsänen, Marko Keskinen, and Matti Kummu. Future changes in mekong river hydrology: impact of climate change and reservoir operation on discharge. *Hydrology and Earth System Sciences*, 16(12):4603–4619, 2012.
- [215] Theodore G Shepherd. Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7(10):703–708, 2014.
- [216] Bikesh Shrestha, Thomas A Cochrane, Brian S Caruso, and Mauricio E Arias. Land use change uncertainty impacts on streamflow and sediment projections in areas undergoing rapid development: A case study in the mekong basin. *Land Degradation & Development*, 29(3):835–848, 2018.
- [217] International Energy Agency IEA. Global energy review 2021. 2021.
- [218] Balázs M Fekete, Charles J Vörösmarty, John O Roads, and Cort J Willmott. Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2):294–304, 2004.
- [219] Md Abul Ehsan Bhuiyan, Efthymios I Nikolopoulos, Emmanouil N Anagnostou, Jan Polcher, Clément Albergel, Emanuel Dutra, Gabriel Fink, Alberto Martínez-De La Torre, and Simon Munier. Assessment of precipi-

- tation error propagation in multi-model global water resource reanalysis. *Hydrology and Earth System Sciences*, 23(4):1973–1994, 2019.
- [220] Nobuhle P Majozi, Chris M Mannaerts, Abel Ramoelo, Renaud Mathieu, and Wouter Verhoef. Uncertainty and sensitivity analysis of a remote-sensing-based penman–monteith model to meteorological and land surface input variables. *Remote Sensing*, 13(5):882, 2021.
- [221] Virginia H Dale. The relationship between land-use change and climate change. *Ecological applications*, 7(3):753–769, 1997.
- [222] Jing Wang, Kelin Wang, Mingyang Zhang, and Chunhua Zhang. Impacts of climate change and human activities on vegetation cover in hilly southern china. *Ecological engineering*, 81:451–461, 2015.
- [223] Stella M Copeland, Seth M Munson, John B Bradford, and Bradley J Butterfield. Influence of climate, post-treatment weather extremes, and soil factors on vegetation recovery after restoration treatments in the southwestern us. *Applied Vegetation Science*, 22(1):85–95, 2019.
- [224] Deepti Rani and Maria Madalena Moreira. Simulation–optimization modeling: a survey and potential application in reservoir systems operation. *Water resources management*, 24(6):1107–1138, 2010.
- [225] Nay Myo Lin and Martine Rutten. Optimal operation of a network of multi-purpose reservoir: A review. *Procedia Engineering*, 154:1376–1384, 2016.
- [226] Julianne D Quinn, Patrick M Reed, M Giuliani, and Andrea Castelletti. Rival framings: A framework for discovering how problem formulation

- uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*, 53(8):7208–7233, 2017.
- [227] Jian Zhang, Kenneth Howard, Carrie Langston, Steve Vasiloff, Brian Kaney, Ami Arthur, Suzanne Van Cooten, Kevin Kelleher, David Kitzmiller, Feng Ding, et al. National mosaic and multi-sensor qpe (nmq) system: Description, results, and future plans. *Bulletin of the American Meteorological Society*, 92(10):1321–1338, 2011.
- [228] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.
- [229] M Cheng, F Fang, T Kinouchi, IM Navon, and CC Pain. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590:125376, 2020.
- [230] Kim Hazelwood, Sarah Bird, David Brooks, Soumith Chintala, Utku Diril, Dmytro Dzhulgakov, Mohamed Fawzy, Bill Jia, Yangqing Jia, Aditya Kalro, et al. Applied machine learning at facebook: A datacenter infrastructure perspective. In *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 620–629. IEEE, 2018.
- [231] Li Yang and Abdallah Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316, 2020.
- [232] Christiane Zarfl, Alexander E Lumsdon, Jürgen Berlekamp, Laura Ty-

- decks, and Klement Tockner. A global boom in hydropower dam construction. *Aquatic Sciences*, 77(1):161–170, 2015.
- [233] Leopoldo Fergusson, Santiago Saavedra, and Juan F Vargas. The perils of misusing remote sensing data: The case of forest cover. *Documento CEDE*, (15), 2020.
- [234] Balaji Rajagopalan, Upmanu Lall, and Stephen E Zebiak. Categorical climate forecasts through regularization and optimal combination of multiple gcm ensembles. *Monthly Weather Review*, 130(7):1792–1811, 2002.
- [235] Xiao Peng, Tiejian Li, and John D Albertson. Investigating predictability of the trhr seasonal precipitation at long lead times using a generalized regression model with regularization. *Frontiers in Earth Science*, page 693, 2021.
- [236] Yu Gong, Pan Liu, Bo Ming, and Dingfang Li. Identifying the effect of forecast uncertainties on hybrid power system operation: A case study of longyangxia hydro–photovoltaic plant in china. *Renewable Energy*, 178: 1303–1321, 2021.
- [237] Dongyang Ren, Xu Xu, Yuanyuan Hao, and Guanhua Huang. Modeling and assessing field irrigation water use in a canal system of hetao, upper yellow river basin: Application to maize, sunflower and watermelon. *Journal of hydrology*, 532:122–139, 2016.
- [238] Mengdi Fu, Junliang Tian, Yanpeng Zhu, Yu Tian, Zhiping Zhao, and Junsheng Li. Identification of functional zones and methods of target management in sanjiangyuan national park. *Biodiversity Science*, 25(1):71, 2017.

- [239] Tao Bai, Jian-xia Chang, Fi-John Chang, Qiang Huang, Yi-min Wang, and Guang-sheng Chen. Synergistic gains from the multi-objective optimal operation of cascade reservoirs in the upper yellow river basin. *Journal of Hydrology*, 523:758–767, 2015.
- [240] Yurong Hu, Shreedhar Maskey, Stefan Uhlenbrook, and Hongli Zhao. Streamflow trends and climate linkages in the source region of the yellow river, china. *Hydrological Processes*, 25(22):3399–3411, 2011.
- [241] Lan Cuo, Yongxin Zhang, Yanhong Gao, Zhenchun Hao, and Luosang Cairang. The impacts of climate change and land cover/use transition on the hydrology in the upper yellow river basin, china. *Journal of Hydrology*, 502:37–52, 2013.
- [242] Fei Yuan, Bing Wang, Chunxiang Shi, Wei Cui, Chongxu Zhao, Yi Liu, Liliang Ren, Limin Zhang, Yonghua Zhu, Tao Chen, et al. Evaluation of hydrological utility of imerg final run v05 and tmpa 3b42v7 satellite precipitation products in the yellow river source region, china. *Journal of hydrology*, 567:696–711, 2018.
- [243] Yichen Yang, Jianshi Zhao, and Ximing Cai. Decentralized optimization method for water allocation management in the yellow river basin. *Journal of Water Resources Planning and Management*, 138(4):313–325, 2012.
- [244] Yuan Si, Xiang Li, Dongqin Yin, Ronghua Liu, Jiahua Wei, Yuefei Huang, Tiejian Li, Jiahong Liu, Shenglong Gu, and Guangqian Wang. Evaluating and optimizing the operation of the hydropower system in the upper yellow river: A general lingo-based integrated framework. *Plos one*, 13(1): e0191483, 2018.

- [245] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.
- [246] Vira Chankong and Yacov Y Haimes. *Multiobjective decision making: theory and methodology*. Courier Dover Publications, 2008.
- [247] Alizée Lopez-Persem, Philippe Domenech, and Mathias Pessiglione. How prior preferences determine decision-making frames and biases in the human brain. *Elife*, 5:e20317, 2016.
- [248] David L Gadenne, Jessica Kennedy, and Catherine McKeiver. An empirical study of environmental awareness and practices in smes. *Journal of Business Ethics*, 84(1):45–63, 2009.
- [249] Matteo Giuliani, Andrea Castelletti, Francesca Pianosi, Emanuele Mason, and Patrick M Reed. Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. *Journal of Water Resources Planning and Management*, 142(2):04015050, 2016.
- [250] Michael T Rosenstein and Andrew G Barto. Robot weightlifting by direct policy search. In *International Joint Conference on Artificial Intelligence*, volume 17, pages 839–846. Citeseer, 2001.
- [251] Julianne D Quinn, Patrick M Reed, and Klaus Keller. Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. *Environmental Modelling & Software*, 92:125–141, 2017.
- [252] Julianne D Quinn, Patrick M Reed, Matteo Giuliani, and Andrea Castelletti. What is controlling our control rules? opening the black box of mul-

- ti-reservoir operating policies using time-varying sensitivity analysis. *Water Resources Research*, 55(7):5962–5984, 2019.
- [253] David Hadka and Patrick Reed. Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary computation*, 21(2):231–259, 2013.
- [254] David Hadka and Patrick Reed. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evolutionary computation*, 20(3):423–452, 2012.
- [255] Patrick M Reed, David Hadka, Jonathan D Herman, Joseph R Kasprzyk, and Joshua B Kollat. Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in water resources*, 51: 438–456, 2013.
- [256] Jonathan D Herman, Patrick M Reed, Harrison B Zeff, and Gregory W Characklis. How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, 141(10):04015012, 2015.
- [257] Joseph R Kasprzyk, Shanthi Nataraj, Patrick M Reed, and Robert J Lempert. Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, 42:55–71, 2013.
- [258] Riddhi Singh, Patrick M Reed, and Klaus Keller. Many-objective robust decision making for managing an ecosystem with a deeply uncertain threshold response. *Ecology and Society*, 20(3), 2015.

- [259] Robert J Lempert and Sara Turner. Engaging multiple worldviews with quantitative decision support: A robust decision-making demonstration using the lake model. *Risk Analysis*, 41(6):845–865, 2021.
- [260] Akiyo Yatagai and Tetsuzo Yasunari. Interannual variations of summer precipitation in the arid/semi-arid regions in china and mongolia: Their regionality and relation to the asian summer monsoon. *Journal of the Meteorological Society of Japan. Ser. II*, 73(5):909–923, 1995.
- [261] Xiaomang Liu, Tiantian Yang, Koulin Hsu, Changming Liu, and Soroosh Sorooshian. Evaluating the streamflow simulation capability of persiann-cdr daily rainfall products in two river basins on the tibetan plateau. *Hydrology and Earth System Sciences*, 21(1):169–181, 2017.
- [262] Peng Gao, X-M Mu, Fei Wang, and R Li. Changes in streamflow and sediment discharge and the response to human activities in the middle reaches of the yellow river. *Hydrology and Earth System Sciences*, 15(1):1–10, 2011.
- [263] Tao Bai, Jian Wei, Fi-John Chang, Wangwang Yang, and Qiang Huang. Optimize multi-objective transformation rules of water-sediment regulation for cascade reservoirs in the upper yellow river of china. *Journal of Hydrology*, 577:123987, 2019.
- [264] Brian R Kirsch, Gregory W Characklis, and Harrison B Zeff. Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: Practical improvement for generating synthetic streamflows. *Journal of Water Resources Planning and Management*, 139(4):396–406, 2013.
- [265] Jianxia Chang, Xuejiao Meng, ZongZhi Wang, Xuebin Wang, and Qiang

- Huang. Optimized cascade reservoir operation considering ice flood control and power generation. *Journal of Hydrology*, 519:1042–1051, 2014.
- [266] Lucian Busoniu, Damien Ernst, Bart De Schutter, and Robert Babuska. Cross-entropy optimization of control policies with adaptive basis functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(1):196–209, 2010.
- [267] Matteo Giuliani, Daniela Anghileri, Andrea Castelletti, Phuong Nam Vu, and Rodolfo Soncini-Sessa. Large storage operations under climate change: expanding uncertainties and evolving tradeoffs. *Environmental Research Letters*, 11(3):035009, 2016.
- [268] Vilfredo Pareto. *Cours d'économie politique*, volume 1. Librairie Droz, 1964.
- [269] Jazmin Zatarain Salazar, Patrick M Reed, Julianne D Quinn, Matteo Giuliani, and Andrea Castelletti. Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. *Advances in water resources*, 109:196–210, 2017.
- [270] Rohini S Gupta, Andrew L Hamilton, Patrick M Reed, and Gregory W Characklis. Can modern multi-objective evolutionary algorithms discover high-dimensional financial risk portfolio tradeoffs for snow-dominated water-energy systems? *Advances in Water Resources*, 145:103718, 2020.
- [271] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M Fonseca, and Viviane Grunert Da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2):117–132, 2003.

- [272] Michael D McKay, Richard J Beckman, and William J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.
- [273] J Eamonn Nash and Jonh V Sutcliffe. River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.