

Quantifying the Impacts of Initial Condition and Model Uncertainty on
Hydrological Forecasts

by

Caleb Matthew DeChant

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Civil Engineering

Dissertation Committee:
Hamid Moradkhani, Chair
Edward D. Zaron
Stefan Talke
James McNames

Portland State University
2014

UMI Number: 3628148

All rights reserved

INFORMATION TO ALL USERS

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

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



UMI 3628148

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

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



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

Abstract

Forecasts of hydrological information are vital for many of society's functions. Availability of water is a requirement for any civilization, and this necessitates quantitative estimates of water for effective resource management. The research in this dissertation will focus on the forecasting of hydrological quantities, with emphasis on times of anomalously low water availability, commonly referred to as droughts. Of particular focus is the quantification of uncertainty in hydrological forecasts, and the factors that affect that uncertainty. With this focus, Bayesian methods, including ensemble data assimilation and multi-model combinations, are utilized to develop a probabilistic forecasting system. This system is applied to the upper Colorado River Basin for water supply and drought forecast analysis.

This dissertation examines further advancements related to the identification of drought intensity. Due to the reliance of drought forecasting on measures of the magnitude of a drought event, it is imperative that these measures be highly accurate. In order to quantify drought intensity, hydrologists typically use statistical indices, which place observed hydrological deficiencies within the context of historical climate. Although such indices are a convenient framework for understanding the intensity of a drought event, they have obstacles related to non-stationary climate, and non-uniformly distributed input variables. This dissertation discusses these shortcomings, demonstrates some errors that conventional indices may lead to, and then proposes a movement towards physically-based indices to overcome these issues.

A final advancement in this dissertation is an examination of the sensitivity of hydrological forecasts to initial conditions. Although this has been performed in many recent studies, the experiment here takes a more detailed approach. Rather than determining the lead time at which meteorological forcing becomes dominant with respect to initial conditions, this study quantifies the lead time at which the forecast becomes entirely insensitive to initial conditions, and estimating the rate at which the forecast loses sensitivity to initial conditions. A primary goal with this study is to examine the recovery of drought, which is related to the loss of sensitivity to below average initial moisture conditions over time. Through this analysis, it is found that forecasts are sensitive to initial conditions at greater lead times than previously thought, which has repercussions for development of forecast systems.

Acknowledgements

My accomplishments would not be possible without the support of a large number of people, and therefore I must recognize those who aided my academic endeavors. First, I must express my gratitude for the guidance of my advisor, Hamid Moradkhani, who has taught me the necessary skills, and provided me with required resources, to complete this work, and whose continual encouragement has motivated me to complete scientific studies of greater complexity than imagined I was capable. To my dissertation committee members, Dr. Zaron, Dr. Talke and Dr. McNamers, I thank you all for your willingness to both serve on my committee and provide thoughtful questions/comments on my research. Your feedback in the development of this work has provided much needed perspective from experts outside my specific field of study. For all of the members of the Water Resources and Remote Sensing Research group at Portland State University, I appreciate your hard work and positive attitude during all of our collaborations, and for your company during long hours in the lab. In addition to my advisor, committee members and colleagues, I must also thank my family for their support. Thank you to my parents for sustained support, both financially and mentally throughout my academic studies. Finally, I must thank my fiancée Bethany, who has whole heartedly supported my work, despite the difficulty of having a partner working full days, weekends, and even the occasional holiday, and has always encouraged me to strive to achieve my dreams.

Table of Contents

Abstract.....	i
Acknowledgements.....	iii
List of Tables.....	v
List of Figures	vi
List of Abbreviations.....	vii
Chapter 1	
1. Forecasting in Hydrological Applications.....	1
Chapter 2	
2. Framing Forecasts from a Probabilistic Perspective.....	14
Chapter 3	
3. Dissertation Objectives.....	22
Chapter 4	
4. Data and Models.....	31
Chapter 5	
5. Proposed Forecasting Framework.....	40
Chapter 6	
6. Forecast Verification.....	56
Chapter 7	
7. Experimental Setup.....	68
Chapter 8	
8. Results and Discussion.....	79
Chapter 9	
9. Conclusion.....	147
References.....	157

List of Tables

Table 1.....	99
Table 2.....	102
Table 3.....	127
Table 4.....	140

List of Figures

Figure 1.....	35
Figure 2.....	51
Figure 3.....	81
Figure 4.....	83
Figure 5.....	85
Figure 6.....	87
Figure 7.....	91
Figure 8.....	93
Figure 9.....	96
Figure 10.....	100
Figure 11.....	104
Figure 12.....	107
Figure 13.....	111
Figure 14.....	113
Figure 15.....	117
Figure 16.....	120
Figure 17.....	121
Figure 18.....	124
Figure 19.....	125
Figure 20.....	129
Figure 21.....	130
Figure 22.....	132
Figure 23.....	135
Figure 24.....	136
Figure 25.....	139
Figure 26.....	141
Figure 27.....	143
Figure 28.....	145

Figure 29.....146

List of Abbreviations

Advanced Microwave Scanning Radiometer-Earth Observing System.....	AMSR-E
Binomial Cumulative Distribution Function.....	B-CDF
Brier Score.....	BS
Cumulative Distribution Function.....	CDF
Continuous Rank Probability Skill Score.....	CRPS
Critical Success Index.....	CSI
Data Assimilation.....	DA
Ensemble Kalman Filter.....	EnKF
Exceedance Ratio.....	ER
Ensemble Streamflow Prediction.....	ESP
False Alarm Ratio.....	FAR
Hydrologic Ensemble Prediction Experiment.....	HEPEX
Kolmogorov-Smirnov.....	KS
Land Surface Temperature.....	LST
Land Water Storage.....	LWS
Moderate Resolution Imaging Spectroradiometer.....	MODIS
Modified Soil Moisture Index.....	MSMI
National Weather Service.....	NWS
National Weather Service River Forecast Center.....	NWSRFC
Poisson-Binomial Cumulative Distribution Function.....	PB-CDF
Probability Distribution Function.....	PDF
Palmer Drought Severity Index.....	PDSI
Particle Filter.....	PF
Particle Filter with Sequential Bayesian Combination.....	PF-SBC
Probability of Detection.....	POD
Precipitation Regression on Independent Slopes Model.....	PRISM
Quantile-Quantile.....	QQ
Relative Entropy.....	RE

Sacramento Soil Moisture Accounting Model.....	SAC-SMA
Standardized Land Water Index.....	SLWI
Soil Moisture Index.....	SMI
Snow Telemetry.....	SNOTEL
Standardize Streamflow Index.....	SSI
Snow Water Equivalent.....	SWE
Brightness Temperature.....	TB
Upper Colorado River Basin.....	UCRB
Variable Infiltration Capacity.....	VIC

1. Forecasting in Hydrological Applications

1.1 Forecasting in Environmental Systems

Forecasts are a vital aspect of the management of many environmental systems (Beck, 1987; Clark et al., 2001; Levine and D'Antonio, 2003; Shukla, 1998). From weather to biological production, society relies on forecasting from various agencies, consultancies and academic institutions to make decisions about how to manage resources (Bloemhof-Ruwaard et al., 1995; Katz and Murphy, 1997; Stern and Easterling, 1999). While the applications and sources of such forecasts are incredibly diverse, all require the formulation of the forecast problem with a systems approach, commonly based on Reynolds Transport Theorem (Hutter and Jöhnk, 2004). Primarily these systems are derived based on an understanding that certain extensive properties are conserved, typically mass, momentum and energy in most systems of interest. Based on these conservation laws, one can generalize the forecasting problems to have components originating from three categories: Initial conditions, boundary conditions and processes (Araújo and New, 2007). Initial conditions are the states, or storages, of extensive properties within a system at some initial forecast time, processes are the general physical principles that control the dynamics of the system, and boundary conditions are the physical constraints acting on the system states/processes. Assuming that each of these elements can be correctly estimated, accurate forecasting is possible for any system of interest.

1.1.1 Boundary Conditions

Boundary conditions explain the physical constraints placed on the states and processes associated with an environmental system. For these conditions, boundaries determine how the universal processes interact within the specific environmental setting of interest. Boundary conditions may be either static or dynamic. For any fully enclosed system, all boundary conditions will be static. Given this scenario, the state of the system is dynamic in time, but can be determined based solely on the evolution of the physical processes in the system over the designated time. Alternatively, any partially enclosed system will have some dynamic boundary conditions, which reflect the effects of other systems on the system of interest. Since any environmental system cannot be feasibly modeled in its entirety, all practical scenarios will have some combination of static and dynamic boundary conditions. These conditions provide some constraint on the system, but will be derived in very different ways. Static boundary conditions rely on some a priori knowledge of the system, or estimation through some form of calibration. One may theorize that with information about all processes and states of a system, static boundary conditions may be estimated with a sufficiently long time series of observation through inverse modeling (Carrera et al., 2005). As for dynamic parameters, one will be required to have some knowledge of the interacting system, or some observation of that interacting system. Although one could attempt to estimate the dynamic boundary conditions through inverse modeling, this should be avoided in the majority of cases as the problem becomes highly ill-posed (Renard et al., 2010).

1.1.2 Initial Conditions

Initial conditions provide a starting point for trajectories in changes of the system properties. Given a specific state of the system (e.g. storage of mass, momentum, or energy), the system properties will deviate from those conditions in time. Although the system moves from this initial state, making the system increasingly insensitive to initial conditions, many systems will remain sensitive to initial conditions for a significant period (Rabier et al., 1996; Rosati et al., 1997). In any real forecasting scenario, one will need to estimate these initial conditions with sufficient accuracy to develop a reasonable forecast of the system at a later date, or to estimate the total change in the system over time. These initial conditions may be quantified in two ways. First, observations of the system may be available to sufficiently characterize the system states. Second, and primarily for the particular case of known dynamic boundary conditions, one may utilize a model to estimate the initial conditions. The former is generally a preferred scenario, as this only has error in the observation process, as opposed to the observation and simulation errors associated with the latter, yet one can rarely observe the state of an environmental system sufficiently for initialization.

1.1.3 Processes

Processes are the general physical laws that determine the changes of extensive properties throughout a system. Characterization of these processes is a great challenge in any moderately complex system, due to a general inability to completely observe such a system. Although physical reasoning and experimentation may provide avenues to explain the behavior of a system, the interactions between initial and boundary conditions

at feasible spatial and temporal scales for discretized modeling are difficult to unravel. Characterization of these processes is performed through two approaches. First, the processes may be approached from laboratory scale experiments, where the foundational principles of a process may be identified in completely controlled scenario (e.g. Darcy flow in saturated porous media). Although such experimentation provides strong evidence for some system behaviors, the process of scaling these principles up to any useful system is bound to have errors. Alternatively, one may approach the problem from conservation laws, which provide information about the system as a whole, and progress to the smaller scale. In any practical scenario, a forecaster will perform some combination of these methodologies, using physical reasoning when possible, and controlling the system with the top-down approach for applicability to the desired scale.

1.2 Hydrological Forecasting

Hydrological forecasts are of high importance to society, due to the dependency of all aspects of society's functions on the availability of water, and the hazards of water in excess. Humans rely on water for a myriad of issues, not the least being requisite consumption for survival. Agriculture, industry, power generation, transportation, environmental stewardship, wildfire prevention and flood control are all examples of applications that require knowledge of water flows/storages at varying spatiotemporal scales. Such a demand for accurate estimation of hydrological states and fluxes requires the application of the general forecasting approach to the specific hydrological setting. Within a hydrological system, models are generally based on conservation of mass and energy, with conservation of momentum being relatively unimportant at scales of

interest. The simplest models in hydrology are highly conceptualized representations of the water balance (Boyle et al., 2000), but increasingly physically-based models are being developed to account for both the water and energy balance (Gao et al., 2010). Although the primary foci in hydrological forecasting are states and fluxes of water, complete modeling requires accounting for energy within the system, due to the phase changes of water experienced above and below the earth's surface.

Hydrological forecasting systems can generally be described through the state-space framework, as described in equation (1). This framework is entirely consistent with the general environmental forecasting framework previously presented, and assumes that the model follows the first order Markovian criteria, with all necessary information about the system being contained in the previous states.

$$x_t = f(x_{t-1}, u_t, \theta) \quad (1)$$

In equation (1), x_t represents the state vector at time t , which is generated by the forward model operator ($f(\cdot)$). This model represents the hydrologic processes, and therefore requires initial states (x_{t-1}), meteorological forcing data (dynamic boundary conditions) for the current time (u_t), and model parameters (static boundary conditions) (θ) to project forward in time. Often in hydrological forecasting, a subsequent model is required to translate these model states into the prediction or observation space. A typical example is applying a hydrological/hydraulic routing model to translate land surface water fluxes to flow at a watershed outlet. This model is referred to as an observational operator, and is represented in equation (2).

$$y_t = h(x_t, \alpha) \quad (2)$$

In equation (2), y_t represents the forecasted value, based on some extraneous process $h()$, static boundary conditions for that process (α), and the states at the current time step.

1.2.1 Forcing and Parameters as Boundary Conditions

Definition of boundary conditions in hydrological models is complex, owing to the nature of conceptualized models, and the interactions of the land surface with the atmosphere. Within hydrological models, these conditions are typically classified into parameters and forcing. Parameters are the boundary conditions which control the physical processes, which vary spatially, but not in time. These parameters affect the processes in the model, and therefore may relate to either physical or conceptual conditions, with respect to the given location. Examples of physical parameters may be vegetation information, soil properties, slope, aspect and elevation, and examples of conceptual parameters may be representative water storage tank maximums (Burnash et al., 1973), infiltration exponents (Liang et al., 1994), snow melt factor (Anderson, 1973), and storage tank outflow rate (Boyle et al., 2000). In order to characterize these parameters, practitioners use various forms of observations, expert knowledge, and inverse modeling. Forcing data are mass and energy inputs to the region of interest, determined by atmospheric processes. In attempts to avoid modeling the complex interactions of the land surface and atmosphere, hydrological modelers/forecasters typically rely on meteorological observations, for hindcasting experiments, and forecasts generated with by meteorologists, to characterize the dynamic boundary conditions. Overall the definition of boundary conditions in hydrological models becomes quite

complex, and therefore requires a combination of observations, calibration, collaboration with meteorologists, and expert knowledge to define.

1.2.2 Hydrological Initial Conditions

Characterization of initial conditions is typically performed through model simulations, commonly referred to as a “spin-up”, where historical observations of the forcing data are available up to an initial forecast date. The spin-up methodology is popular because nearly all hydrological models have states that are either conceptual or unobservable by currently available methods. These initial conditions primarily include snow water equivalent (SWE), soil moisture content, groundwater levels, and temperatures of the snow/soil. Although these variables have physical values, models often conceptualize them (e.g. theoretical storage tank representing the soil matrix), or require spatial averages that are not readily observed. Through the spin-up methodology, hydrological model simulations are performed up to some initial forecast time. At the initial forecast time, all states considered by the model are saved, and become a starting point for hydrological simulations into the future.

1.2.3 Hydrological Model Processes

Processes at the land surface are poorly understood, and therefore are often of conceptual nature or extrapolated from point-scale/laboratory studies. Owing to the complexities of the land surface structure, land-atmosphere exchanges and environmental interactions, it becomes nearly impossible to accurately observe hydrological phenomena at desired scales for forecasting. In order to overcome these problems, hydrologists either develop a conceptual understanding of the movement of water through a watershed, then

develop an abstract representation of the watershed that is simple enough to simulate (i.e. leaky bucket model), or apply information gained from controlled experiments to the scale of interest (i.e. Darcy flow through soil). Although modeling of the hydrologic cycle requires major assumptions and abstractions, hydrologists rely on the knowledge that mass and energy are conserved to ensure some consistency between model structure and the real world. In addition, the reliance on conservation laws provides hydrological modelers with a basis for model verification (i.e. minimization of bias).

1.3 Water Supply Forecasting

Water supply forecasting broadly describes the quantification of available water for some use over a specified time-scale. Applications of these forecasts can fall into a number of categories including reservoir management for consumptive use, irrigation of cropland, planning of hydropower generation schedules and ensuring availability of environmental water. Within the context these applications, supply forecasts range from monthly to multi-year time-scales (Wood and Werner, 2011), with demand for specific time-scales being determined by stakeholders and water managers (Werner et al., 2013). Within the group of forecast users, a diverse group of farmers, reservoir operators, municipal planners and power supply managers provide competing interests which forecasting agencies attempt to satisfy. Forecasting agencies, including the National Weather Service River Forecast Centers (NWSRFC) and the Natural Resource Conservation Service, and academic groups have developed a range of forecasting techniques to meet user needs (Arumugam et al., 2014). These techniques may take the form of either statistical or dynamic forecasts (Day, 1985). Statistical forecasts simplify

the generalized forecasting system above, and focus directly on correlations between certain environmental information and the quantity of interest. Such forecasting represents the most simplistic of frameworks, and the basis for the original water supply forecasts. Due to a reliance on stationary climate, and a perception that the potential skill of statistical forecasts has an upper limit, forecasters are moving to dynamic frameworks. Through a movement to dynamic, simulation based, forecasts, an understanding of the physics may be leveraged, thus reducing the reliance on climate stationarity, and allowing for continual improvement as research of physical hydrology progresses.

Statistical forecasts of volumetric runoff from the land surface are built on a relationship between one or more environmental variables and streamflow volumes. Some popular sources of information are snow observations (Risley et al., 2005), sea surface temperature (Aziz et al., 2010) and geopotential heights (Grantz et al., 2005). Such relationships are modeled with regression based techniques (Garen, 1992; Moradkhani and Meier, 2010; Pagano et al., 2004) or artificial neural networks (Maier et al., 2010). Within the realm of regression analysis, it has become common to include multiple indicators, which necessitates the removal of variable correlations. Principal component analysis (Garen, 1992) has become quite popular for removing these correlations (e.g. Olden and Poff, 2003; Piechota et al., 1997; Regonda et al., 2006), but more recent developments are available (i.e. Moradkhani and Meier, 2010). Although statistical methods are quite popular for water supply forecasting, these techniques have been questioned by researchers. Due to a lack of physical realism in the statistical models, many researchers have hypothesized that these methods have a potential skill

that is lower than that of dynamic forecast methods, and further argued that these methods are questionable in light of the finding that climate is non-stationary. This dissertation follows the thought that striving for additionally physically-based forecast systems will yield improving results moving forward, and therefore focuses on dynamic forecasting techniques in an attempt to gain information from physical reasoning, and increase the reliability dynamic forecast systems for use under future climatic conditions.

Dynamic forecasting requires the use of a model to simulate future hydrologic processes. The general framework for such a system was laid out earlier in this chapter, and may be observed in both operational and research based forecasting techniques. A prominent example of this is the Ensemble Streamflow Prediction (ESP) method used by the NWSRFCs (Day, 1985). As described earlier, dynamic forecasting of streamflow requires the estimation of initial model states, boundary conditions and future hydrological processes. In regards to the specific case of water supply forecasting, the stationary boundary conditions, referred to as model parameters by hydrologists, are typically treated as part of the process model, and calibrated during some period of time prior to the forecast. After calibration, a spin-up is performed with the model to estimate initial states, and information about future climate, typically precipitation and temperature data, is used to drive the model forward in time. In order to improve forecast accuracy, at least one of these forecast components must be improved. Through improved identification of the initial conditions, boundary conditions and/or hydrologic processes, the accuracy of the final forecast product will become more accurate.

1.4 Drought Forecasting

Drought is a complex hydro-climatic phenomenon that remains poorly understood across various geophysical disciplines. Despite decades of research examining drought, scientists have failed to even reach a consensus on the definition of drought. In addition to being poorly understood, drought is also the costliest natural disaster (Cook et al., 2007), accounting for 41.2% of the economic cost from all natural disasters in the US (Ross and Lott, 2003), and affects the largest population globally of any natural disaster (Wilhite, 2000). Such a combination of misunderstanding and danger has led researchers to focus on understanding drought processes (Mishra and Singh, 2011; Dai, 2011; Zargar et al., 2011), and improving methodologies to mitigate the devastating effects of droughts. Within this line of research, advancing our understanding of droughts is expected to lead to improved quantification and forecasting of drought events.

Quantifying drought is typically performed through index based assessment, where the intensity of a drought is related to the magnitude of the index. The basis for the vast majority of the drought indices used in the scientific literature was forged nearly a half-century ago (Palmer, 1965). A landmark development, the Palmer Drought Severity Index (PDSI) gave researchers the ability to quantify drought in the context of historical information, thus providing an index that may be interpreted with respect to historical climate. Based on a record of past observations, the drought severity is related to historical probability of occurrence. By viewing the state of drought in a historical context, drought is characterized as a deviation from normal conditions, similar to the definition of meteorological drought by Wilhite (2000). This has led to wide ranging

applications of the PDSI for drought quantification, and further advancements of this index (i.e. Surface Water Supply Index). In a simpler fashion, Mckee et al., (1993) developed a Standardized Precipitation Index, under the assumption that droughts are primarily initiated by a lack of precipitation. This method of standardization follows Palmer (1965) to quantify drought conditions in relation to climatology, but demonstrated how this may be applied in a much simpler fashion to focus on a specific aspect of drought. Standardization of hydrological variables has become quite common for drought quantification, leading to the development of many drought indices, each with specific focus. Based on this drought quantification, mitigation measures may be enforced to reduce the impacts of a drought event.

Currently, most drought mitigation systems take a reactionary approach, placing water restrictions when a drought is observed, but developing forecast based mitigation systems will more effectively reduce the overall cost of drought (Jaeger et al., 2013; Pozzi et al., 2013). In addition, moving towards proactive drought management is essential to the future functioning of society in light of recent climate change studies (Dai, 2011). Operational forecasts of drought are produced regularly by the National Oceanic and Atmospheric Administration Climate Prediction Center and the National Drought Monitor, but research forecasts are also available (e.g. <http://hydrology.princeton.edu/forecast/current.php>). In addition, new drought forecasting methodologies are developing, which forecast different types of droughts, and drought characteristics including onset, intensity and duration (Arshad et al., 2012; Kim and Valdés, 2003; Yuan and Wood, 2013). These forecast products are essential in driving

drought mitigation decisions, but advances are still needed to improve management. Similar to water supply forecasting, these advances may be expected to come from improvements in the quantification initial conditions, boundary conditions and hydrological processes that lead to drought conditions.

2 Framing Forecasts from a Probabilistic Perspective

Forecasting of hydrologic variables may be formulated according to the topics covered in Chapter 1, but this framework assumes perfect knowledge of all hydrologically relevant aspects of a region. In any real watershed or region, the spatiotemporal dynamics of hydrological states and processes lead to unavoidable uncertainties. The hydrological cycle cannot be completely observed, nor can it be exactly modeled. Therefore, a forecaster is incapable of perfectly forecasting the phenomena of interest. In order to move forward, a forecaster must acknowledge that all forecasting systems are imperfect, and therefore attempt to quantify the uncertainty in a given forecast.

Forecast uncertainty results from each of the forecasting components. With the knowledge that any forecast can be developed with information about initial states, boundary conditions and processes, and that each of these components will contain some uncertainties, a robust approach for quantifying that forecast uncertainty is through a bottom up framework, examining each forecast component separately. Initial condition uncertainty arises from the inability to accurately observe land surface states. Although a range of advanced observation systems are available, including in-situ and remotely sensing platforms, the spatio-temporal distribution of the predominant hydrological variables is too complex to completely resolve. Alternatively, the forcing, or dynamic boundary conditions, must be gathered from another model to perform simulations into the future. These models may be either atmospheric or statistical in nature, and each type of model will carry significant uncertainties. Model parameters, or static boundary

conditions, may be observed in the case of physically-based model representations, but most impactful parameters in hydrological modeling are unobservable, or conceptual, and therefore require some form of inverse modeling. This inverse modeling will carry uncertainties as the parameters are not perfectly identifiable, due to incomplete information content of the observations, suggesting the presence of parameter error in optimization schemes (Beven and Binley, 1992). Finally, processes are never completely known, and if they were completely known, perfect simulation of those processes would likely be infeasible with currently available computing systems. This is referred to as model structural error, and is likely the most complex uncertainty to address. In order to quantify these uncertainties, methods are taken from probability theory to estimate the certainty of given forecasted outcomes.

2.1 Quantifying uncertainties

2.1.1 Probabilistic Forecasting Methods

Quantifying uncertainty of any phenomena requires the identification of potential estimation errors from the probabilistic perspective. This perspective requires the treatment of some variable(s) of interest as random, being drawn from some probability distribution. At this point, it should be clarified that this dissertation takes the position that hydrological variables are deterministic, and therefore theoretically have potential to be forecasted exactly, but any practical forecaster must treat them as stochastic to represent their incomplete knowledge of the desired information. In this regard, probability theory is relied upon to quantify a forecaster's ignorance, and is not a suggestion that hydrological processes are truly random. In order to frame a forecast from

the probabilistic perspective, a forecaster will typically need to assume parametric Probability Distribution Functions (PDF) to represent the uncertainties of interest (DeChant and Moradkhani, 2014). At all levels of hydrological forecasting, uncertainties may be represented by some PDF. Following the previously described state-space framework, probabilistic modeling of forecasts may be described, as shown in equations (3) and (4).

$$p(x_t) = f(p(x_{t-1}), p(u_t), p(\theta)) + p(\omega_t) \quad (3)$$

$$p(y_t) = h(p(x_t), p(\alpha)) + p(v_t) \quad (4)$$

In the above equations, all variables are represented as probability distributions, as noted with $p()$. Further, each representative model has a time-dependent error term added to it, with ω_t representing the forward model error, and v_t representing the observational model error. These terms represent a time-varying model error, each of which comes from a PDF representing a forecaster's uncertainty about each model's structure. Due to the variety of different uncertainties in this system, it is common to follow the Central Limit Theorem, and therefore assume the overall forecast PDF should approach Gaussianity (Gupta et al., 2009; Reichle et al., 2002; Schoups and Vrugt, 2010). Although this may seem to be a reasonable approximation for the generalized case, the zero boundary condition (no negative mass) for many hydrological variables motivates the use of positive PDFs (Log-Normal or Gamma distributions) (Jayawardena and Lai, 1994). Given the non-linear nature of hydrological models, combining PDFs representing different sources of uncertainty is quite challenging. Generally it is infeasible to perform this analytically, which necessitates the use of Monte Carlo simulations. This involves

simulation of the forecast density to develop an ensemble representation of the forecast PDF. Thus hydrological forecasters have increasingly moved towards ensemble forecasting methods for estimating forecast uncertainty, especially with the recent advances in high-performance computing to overcome the computational burden.

Ensemble prediction methods in the hydrological sciences are prevalent, both in research and operational applications. They are the basis for the operational ESP framework and the focus of the Hydrological Ensemble Prediction Experiment (HEPEX) (Schaake et al., 2007), which is a major research initiative. The extent of operational and research interest highlights the importance of continued study into ensemble forecast methods. In continuing this line of research, this dissertation examines the use of ensemble based forecasting for quantifying hydrologic uncertainty. Following equations (3) and (4), the forecasting problem can be shown in an ensemble framework through the state-space representation.

$$x_{i,t} = f(x_{i,t-1}, u_{i,t}, \theta_i) + \omega_{i,t} \quad (5)$$

$$y_{i,t} = h(x_{i,t}, \alpha_i) + v_{i,t} \quad (6)$$

In equations (5) and (6), each variable is indexed with i , indicating the specific member that value falls within the ensemble. At this point, the problem becomes sampling from the individual probability distributions, and then performing enough simulations to represent the forecast probability distribution. In order to sample each individual error source, one will need an error model for each variable.

Model state error is entirely described within equation (3), with the hydrological model, parameter and forcing errors all contributing to state error. Forcing error requires

some direct error model (equation (7)), which will typically consist of sampling from a representative distribution. Note that in equation (7), the overline indicates an observed/estimated value, which has some unknown error. This distribution will vary depending on the quantity of interest. Similarly, model structural error will typically be sampled from a distribution, which is commonly treated as Gaussian, which may be supported by the argument that the complexities of the model calculations will satisfy the central limit theorem. It is acknowledged that the representativeness the normal distribution is a large assumption, but the examination of alternatives is outside the scope of this dissertation. In equations (8) and (9), η and κ represent multiplication factors to determine the variance of the forward model and observational operator, respectively. These factors treat structural errors as heteroschedastic, based on the assumption that errors likely grow as the forecast deviates from the zero boundary condition, and are therefore less constrained. An alternative to representing structural error is the use of multiple model ensembles (Clemen, 1989; Najafi et al., 2011). Through the simulation of multiple models, it is assumed that the uncertainty in hydrological processes will be implicitly quantified. Finally, parameter error becomes a much more challenging problem, as the parameter errors are intertwined with the data errors. In the hydrological literature, much focus has been placed on parameter error, which is typically estimated with a Bayesian approach (Moradkhani et al., 2012; Thyer et al., 2009).

$$u_{t,i} = Err_{input}(\bar{u}_t) \quad (7)$$

$$\omega_{t,i} \sim N(0, \eta * f(x_{i,t-1}, u_{i,t}, \theta_i)) \quad (8)$$

$$v_{t,i} \sim N(0, \kappa * h(x_{i,t}, \alpha_i)) \quad (9)$$

2.2 Reducing Uncertainties

2.2.1 Reducing State Uncertainty

Reducing state uncertainty within hydrological sciences is necessary for improved forecasting, but this is a challenging task. In order to reduce state uncertainty, one must either improve upon their modeling system, or gain observations of the desired state which is more accurate than the model estimate. Although either of these avenues for improving model states may be reasonable in an extended timeline, as incremental advances in modeling and observing are expected, it is a better strategy to use information from both model simulations and observations to develop a theoretically better product. This methodology is broadly referred to as data assimilation (DA), and has become increasingly popular in hydrological sciences. Within the specific context of this dissertation, a class of DA techniques, referred to as ensemble DA, are highly appropriate. Similar to the use of ensemble methods in representing forecast uncertainty, ensemble DA allows for assimilation with highly non-linear models. By performing DA in an ensemble framework, state uncertainty is being both estimated and reduced simultaneously. Such quantification is highly desirable for the applications examined in this dissertation.

A number of researchers have examined the use of ensemble DA methods for improving land surface state prediction (Andreadis and Lettenmaier, 2006; Clark et al., 2008a; DeChant and Moradkhani, 2011a; De Lannoy et al., 2012; Liu et al., 2012; Margulis et al., 2002; Reichle et al., 2002) and examining the ability of stochastic states to estimate uncertainty reliably (DeChant and Moradkhani, 2012; Leisenring and

Moradkhani, 2011; Leisenring and Moradkhani, 2012; Liu and Gupta, 2007; Moradkhani et al., 2005a, b). The majority of these techniques rely on Bayes Theorem to sequentially reduce the uncertainties in ensemble predictions of state values, based on some observed information. Two primary techniques exist within the hydrological DA literature: the Ensemble Kalman Filter (EnKF) and the Particle Filter (PF). Of these techniques, the EnKF has seen the greatest use in hydrology, and has been shown to effectively reduce errors in the simulation of a range of variables. Although the EnKF is quite popular, this study utilizes the PF, as it has been shown to be a more robust estimator of hydrological uncertainty than the EnKF (DeChant and Moradkhani, 2012), which results from a reduced reliance on Gaussian error structure, and greater applicability to highly non-linear problems. More details on the PF are provided in Chapter 5.

2.2.2 Reducing Model Uncertainty

Reducing model simulation uncertainty has proven challenging since the development of the first simulation models. This point is evidenced by the lengthy discussion of model development philosophy and model error identification in the last 25 years (e.g. Beven, 1989; Butts et al., 2004; Clark et al., 2008b; Jakeman and Hornberger, 1993). The efficacy of different hydrological models and potential ways to improve their individual simulations are regularly discussed, both in the literature and at scientific conferences, yet a single optimal model structure cannot be identified. This has led to the development of a range of modeling systems. With the variety of models currently available in hydrology, and no clear optimal structure, it has become popular to address model structural uncertainty implicitly through multi-model ensembles (Bohn et al.,

2010; Clemen, 1989; Regonda et al., 2006). By having a diverse set of model structures, a forecast is assumed to quantify the uncertainties related to each individual model. Further, one may reduce the error in a multi-model forecast with the use of some observed information. Similar to ensemble DA, Bayes Theorem may be applied to an ensemble of model simulations to reduce the uncertainty of that ensemble. This method relies on the calibration of model weights over a set time period, and then those weights are applied during a forecast period. Due to the ensemble basis of this forecasting system, and the ability to quantify/reduce model structural uncertainty, the use of Bayesian multi-modeling is appropriate for use in this dissertation.

3 Dissertation Objectives

A primary theme of this research is the pursuit of comprehensive accounting of uncertainty in hydrological forecasting. Such a goal is motivated by the assumption that total uncertainty can be reliably estimated, while avoiding unnecessary inflation of that uncertainty, through the proper representation of uncertainty in each forecasting component. Although examples of total error calibration are suggested in the literature (e.g., Montanari and Grossi, 2008), thus simplifying the uncertainty estimation problem, the non-linear relationship between different sources of uncertainty and the forecast value will likely create information loss in a posteriori uncertainty estimation (Renard et al., 2010). A more prudent approach is a comprehensive methodology by individually treating each possible source of uncertainty (DeChant and Moradkhani, 2012; Kuczera et al., 2006; Moradkhani et al., 2012). Specifically, this comprehensive approach requires the treatment of each initial condition, boundary condition, and process with some estimated uncertainty variance. In the specific case of a hydrological model, this will be identified as four sources: Model states (initial conditions), meteorological forcing (dynamic boundary conditions), model parameters (static boundary conditions) and model structure (hydrologic processes). Recent work has shown that accounting for all four of these sources are essential for quantifying uncertainty in forecasts from conceptual hydrological models in the short-term (DeChant and Moradkhani, 2012; Moradkhani et al., 2012), but parameter uncertainty may be less significant at longer time-periods. Since the model is based on a conservation of mass and energy, and specific parameter values tend to focus on rates of energy and water fluxes, it is likely

that model structural uncertainty is dominant, in comparison to parameters, at the seasonal time scales of interest here. By assuming that model structural uncertainty are dominant in comparison to parameter uncertainty, the calibration process is simplified, and the parameter uncertainty is combined with structural uncertainty, which will be referred to as model uncertainty for clarity. Therefore this study will focus on the roles of initial state, forcing and model uncertainty for probabilistic forecasting.

There are an increasing number of operational and research forecasts being cast in a probabilistic manner (Demargne et al., 2013; Madadgar et al., 2012; Yuan et al., 2013), yet none of these forecasts approach the uncertainty estimation problem comprehensively. Most of these forecasts focus predominantly on the uncertainties relative to future weather conditions, thus ignoring initial condition and model uncertainties. Two hypotheses have likely led to the community at large ignoring state and model uncertainty: 1) Forcing uncertainty is assumed to be the largest source of uncertainty; 2) initial states and model processes are the primary source of skill in seasonal forecasting. While the work in this dissertation is in agreement with both of these hypotheses, it is argued that other sources of uncertainty are still significant, and that adding the proper uncertainties to the models/states will not reduce the skill of a forecast. The seasonal forecasting literature generally suggests that state and model uncertainties are significant, based on the consistent finding that most forecasting systems underestimate total uncertainty (Wood and Schaake, 2008, Yuan and Wood, 2012). By focusing entirely on forcing uncertainty, the uncertainties related to initial conditions and the model are removed from the analysis, leading to erroneously small forecast variance.

In order to overcome this issue, seasonal forecasts must move towards treating the initial land surface states and model predictions as probabilistic values, in addition to meteorological forcing. Based on this theme, four objectives can be identified for this study:

1. Demonstrate Reliable Water Supply Forecasting

A first objective in this dissertation is to demonstrate how the proposed methodology can produce reliable probabilistic water supply forecasts at the basin-scale, for seasonal time periods. Volumetric estimates of runoff are necessary for guidance of an array of water management decisions, yet the accuracy of such estimates is often unsatisfactory. To this end, it should be of high priority to ensure that estimates of forecast uncertainty are statistically reliable. Given that probabilistic estimates of volumetric streamflow are reliable, risk within a reservoir system can be more effectively managed, thus reducing the chance of both flood damages and water shortages concurrently.

Research into probabilistic methods for water supply forecasts has developed over the past few decades. A first example is the ESP framework proposed by Twedt et al., (1977) and clarified by Day (1985). ESP works under the assumption that the primary skill in a hydrological forecast is based on land surface conditions, and as such treats initial conditions as deterministic quantities, while leveraging climatological stochastic forcing to account for poor knowledge of future meteorological conditions. The framework itself has prompted a number of studies to attempt to improve seasonal forecasting, including utilizing information from climate indices (Najafi et al., 2012) and

climate modeling products (Mo et al., 2012; Yuan and Wood, 2012). Since the literature suggests that some information about seasonal climate is available through both climate modeling and teleconnections, further studies have examined the assumption that skill is primarily derived from initial conditions (Li et al., 2009; Shukla et al., 2013; Wood and Schaake, 2008; Yossef et al., 2013). With an increasing focus on the relative skill of different aspects of seasonal forecasting, an increasing focus has been placed on determining how best to manage overall uncertainty in the modeling framework. This dissertation proposes an approach to quantifying total forecast uncertainty, and will test the proposed framework with a seasonal forecasting experiment at the basin scale.

2. Demonstrate Reliable Drought Forecasting

A second objective in this dissertation is to demonstrate a methodology to produce reliable probabilistic drought forecasts, which will be achieved by extending the proposed water supply forecasting system to the case of monthly and seasonal drought. This objective is motivated by the desire to develop a proactive drought mitigation system, and the understanding that there is insufficient data to perfectly forecast drought. As a result of this incomplete ability to predict drought, it becomes essential to quantify the certainty which one can place upon a drought forecast. While probabilistic forecasting systems are the norm for many hydrological variables, they are only recently being applied for characterizing drought uncertainty.

Several recent examples of probabilistic drought forecasting are available (Hwang and Carbone, 2009; Madadgar and Moradkhani, 2013; Madadgar and Moradkhani, 2014; Pan et al., 2013 Yuan et al., 2013). While these are generally still in the development

phase, this marks progress in drought forecasting, and a movement away from reactionary drought management. Based on the proposed water supply forecasting framework, a case study for probabilistic drought prediction will be developed to test drought forecast reliability. Within this experiment, it is of high importance to ensure that a forecast is truly reliable. This condition necessitates a critical analysis of reliability metrics, leading this dissertation to propose a new reliability metric with which to analyze probabilistic drought forecast reliability.

3. Critically Examine Traditional Drought Quantification

Forecasting of drought is reliant on available methods for quantifying drought, but the conventional index based assessment is fundamentally flawed. Three specific problems associated with standardized indices are highlighted in this dissertation. First, current droughts indices assume that climate processes are stationary, thus having a constant variability, and that the available historical information covers a sufficiently long time period to characterize that variability. With the community's increasing emphasis on the changing climate (Milly et al., 2008), and the short period of most climate records (Verdon-Kidd and Kiem, 2010), it is unlikely that a statistical index can completely describe the true intensity of a drought, as there is no truly representative dataset of current/future climate. Second, water demand is dynamic. Similar to the non-stationarity of climate, one cannot rely on history to represent the effects of present-day and future drought. Further, spatiotemporal changes in societal water demand (consumption, industrial, irrigation) change much faster than climate processes, adding complexity to drought analysis in human affected environments. Since climate and

demand vary at different timescales, one cannot expect a drought index based on climatology to effectively determine observable impacts. Finally, the standardization of historical data may complicate the examination of spatially distributed drought properties. Hydrological variables are never evenly distributed spatially, leading to an overemphasis on typically dry regions with standardized indices. Since it is impossible to have a reliable drought forecast without a reliable drought index, this dissertation will examine how current drought indices may be unreliable, and suggest a possible starting point for advancing drought indices.

4. Examine the persistence of model initial conditions in forecasts

Hydrological forecasts are known to be strongly affected by all initial conditions in the short term, but the sensitivities to initial conditions at long lead times are not well understood. Over time, a forecast will progressively lose sensitivity to initial conditions, but the lead-time at which initial states are meaningful for a given forecast has yet to be quantified. This is an important point, as hydrological forecasts are in demand at timescales of a season to multiple years (Wood and Werner, 2011), depending on the application of the forecast. Given this scenario, quantifying the sensitivity of future land surface states to initial conditions is highly valuable for hydrologists. Two specific examples are highlighted in this dissertation: 1) examining the recovery time from a drought scenario and 2) understanding the extent to which improved initial conditions will help a given forecast system.

Understanding the relationship between the intensity of drought conditions and the time to recovery is of great importance (Pan et al., 2013). Since a drought is

determined by some deficiency in water, that deficiency will take some amount of time to be alleviated. If the magnitude of that deficiency can be related to the expected recovery time, then mitigation measures may be focused during that expected drought recovery period. Without information about future climate, which is commonly assumed in hydrological forecasting, it is beneficial at the basic level to quantify the recovery time under normal conditions. This provides another view to drought intensity, beyond simply explaining the magnitude of the deficiency. Such information about the typical recovery time may be more functional to the general public and water resources managers alike, and therefore is of interest to society in general. Further, information about sensitivity to initial conditions is useful in considering potential methods to improve a given forecast system as well.

Improving a forecast system involves reducing the uncertainty in at least one of the forecasting components discussed in Chapter 1. Of these three primary components, uncertainty related to initial conditions is the most straightforward to reduce, through the DA methods discussed in section 2.2.1. Although implementation of a DA system has been proven to reduce uncertainty in land surface states, and therefore initial conditions, it does require some resources to develop and maintain. Given this scenario, information about the sensitivity of forecasts in a given basin to initial conditions, at various lead times, would be highly beneficial in the development of forecast systems. If the forecast at the desired lead time is insensitive to the initial conditions, implementation of a data assimilation system may not be warranted. Alternatively, persistent influence of initial

conditions over the forecast at long lead-times would indicate that a DA system is highly beneficial.

Recently, several studies have analyzed the sensitivity of extended forecasts to initial conditions, but have focused on the relative uncertainties in initial conditions and meteorological forcing (Li et al., 2009; Mahanama et al., 2011; Paiva et al., 2012; Shukla and Lettenmaier, 2011; Shukla et al., 2013; Yossef et al., 2013). These studies either compared the ESP methodology (deterministic initial conditions with climatological forcing) with Reverse Ensemble Streamflow Prediction (stochastic initial conditions with deterministic forcing) (Wood and Lettenmaier, 2008), or examined of the ratio of the variability of initial conditions and precipitation during the forecast period. Through these methodologies, it was generally shown that forecasts in snow dominated basins were controlled by initial conditions between three and six month lead times, when the initial forecast date occurred during the accumulation or ablation season, yet only forecasts for very large non-snow dominated basins were controlled by initial conditions beyond a single month. While this analysis provides a compelling argument for the use of data assimilation in short-term forecasts for all basins, and seasonal forecasts during spring and summer for snow dominated basins, such analysis falls short of determining the lead time at which initial conditions provide significant information. For example, forcing may be the dominant source of forecast skill beyond the seasonal time-scale for nearly all basins, but initial conditions may still have a significant impact on forecast uncertainty at longer lead times. Due to this shortcoming, this study attempts to quantify the time at which a forecast becomes entirely insensitive to initial conditions, and examines the rate

at which the information from the initial conditions is lost over time. By quantifying the specific lead time at which a forecast is no longer sensitive to initial conditions, a forecaster can provide clear evidence of the point at which data assimilation will be of no benefit, and provide insight into the recovery time expected from certain drought events. For the remainder of this study, drought recovery will be used to describe the loss of sensitivity to initial conditions, for consistency, and to specify that this dissertation focuses entirely on the effects of below average water storages on seasonal to annual forecasting.

4 Data and Models

4.1 Hydrologic Models

4.1.1 Variable Infiltration Capacity Model

The Variable Infiltration Capacity (VIC) model is a physically-based, distributed model that solves the energy and water balance at the land surface, and spatially discretized units are generally placed on a regular grid (Gao et al., 2010; Liang et al., 1994). In order to perform model calculations, VIC requires soil information, vegetation information, elevation bands, precipitation, maximum and minimum temperature, average wind speed, humidity, and incoming shortwave and longwave radiation for each grid cell. Land surface parameters for VIC simulations were gathered from the Natural Resources Conservation Services STATSGO dataset (soil) and the University of Maryland land cover dataset (vegetation). Elevation bands were defined using the United States Geological Survey's National Elevation Dataset, with information from the Precipitation Regression on Independent Slopes Model (PRISM) yearly precipitation information to aid in the distribution of elevation band precipitation. Simulations were performed over the entire Upper Colorado River Basin (UCRB) at a spatial resolution of 0.25°, which makes 473 model grid cells. Based on the hydrologic fluxes estimated by VIC, excess water is routed to the outlet of the basin with a combination of Nash-Cascade hydrologic routing and Muskingum-Cunge hydraulic routing.

4.1.2 National Weather Service Models

The SNOW-17 (Anderson, 1973) and Sacramento Soil Moisture Accounting (SAC-SMA) (Burnash et al., 1973) models are used by the NWRFCs to provide operational streamflow forecasts for flood and water supply monitoring. These models are coupled, with SNOW-17 handling snow accumulation/ablation calculations and SAC-SMA modeling the soil water storage component. Both SNOW-17 and SAC-SMA have a more conceptual nature to model equations than VIC, leading to an increased reliance on calibration, as opposed to soil and vegetation data. Fortunately, the NWS calibrated parameters for each basin within the UCRB have been made available by the Colorado Basin River Forecast Center. The NWS performs simulations from these models with elevation bands for each sub-basin, leading to 409 discretized units. To run SNOW-17 and SAC-SMA, precipitation, average temperature, and potential evapotranspiration are required. Excess runoff from these models is routed to the outlet with a unit hydrograph for hydrologic routing and Lag/K for hydraulic routing.

4.1.3 Radiative Transfer Models

This study requires a model to estimate the passive microwave radiation from the land surface, which is referred to as a radiative transfer model. For modeling the radiation from the land surface, a combination of the L-Band Microwave Emission Model from the Biosphere (Wigneron et al., 2007), for the case of bare soil, and the Microwave Emission Model for Layered Snowpack (Weismann and Mätzler 1999), when snow cover is present, are used. Radiative transfer calculations from Microwave Emission Model for Layered Snowpack assume horizontally homogeneous layers of snow depth, density,

correlation length, liquid water content and temperature. Simulations from the L-Band Microwave Emission Model from the Biosphere model assumes a homogeneous soil, requiring information about soil temperature, bulk density, sand and clay contents and soil moisture content. Vegetation effects are modeled according to Pullainen et al. (1998), and atmospheric effects are accounted for as described in Durand and Margulis (2006), allowing for comparison of satellite observations. Both the VIC and SNOW-17 models estimate all snow variables necessary, with the exception of grain size, which is modeled according to Jordan (1991), and DeChant and Moradkhani (2011a). Alternatively, only the VIC model estimates physical soil moisture. Due to the conceptual nature of soil water storage in the SAC-SMA model, microwave emission in the 6.9GHz frequency is ignored for assimilation in the NWS models. Outputs from the radiative transfer model are in the form of brightness temperature (TB) for each frequency at each polarization (vertical and horizontal), which is the apparent temperature of earth, as observed by the satellite, assuming the earth is a black body (emissivity is 1).

4.2 Study Area

The proposed study area is the UCRB, defined here as the entire Colorado River Basin upstream of Lee's Ferry (see Figure 1), which is located just downstream of Lake Powell. The UCRB is located in the southwestern US, covering portions of Wyoming, Utah, Colorado, Arizona and New Mexico. The basin drains an area of roughly 280,000km², with forest covering much of the upper elevations and shrub-land covering the valleys. This basin is semi-arid, with the majority of precipitation falling in the higher elevations as snow, and interior lowlands receiving very little precipitation annually.

Although this basin is semi-arid, a large population relies on its runoff. The mean naturalized yearly flow volume at Lee's Ferry is roughly 18 billion cubic meters, providing water to 26 million people, with a minimum designated annual flow from Lake Powell set at 9.3 billion cubic meters. In Figure 1, the gauges of the three major sub-basins (Green River, Colorado Headwaters/Gunnison and San Juan) and at Lee's Ferry are identified. These basins are used for both calibration and large scale validation. In addition to these large sub-basins, 16 smaller sub-basins are used for more detailed analysis, which are not displayed in Figure 1.

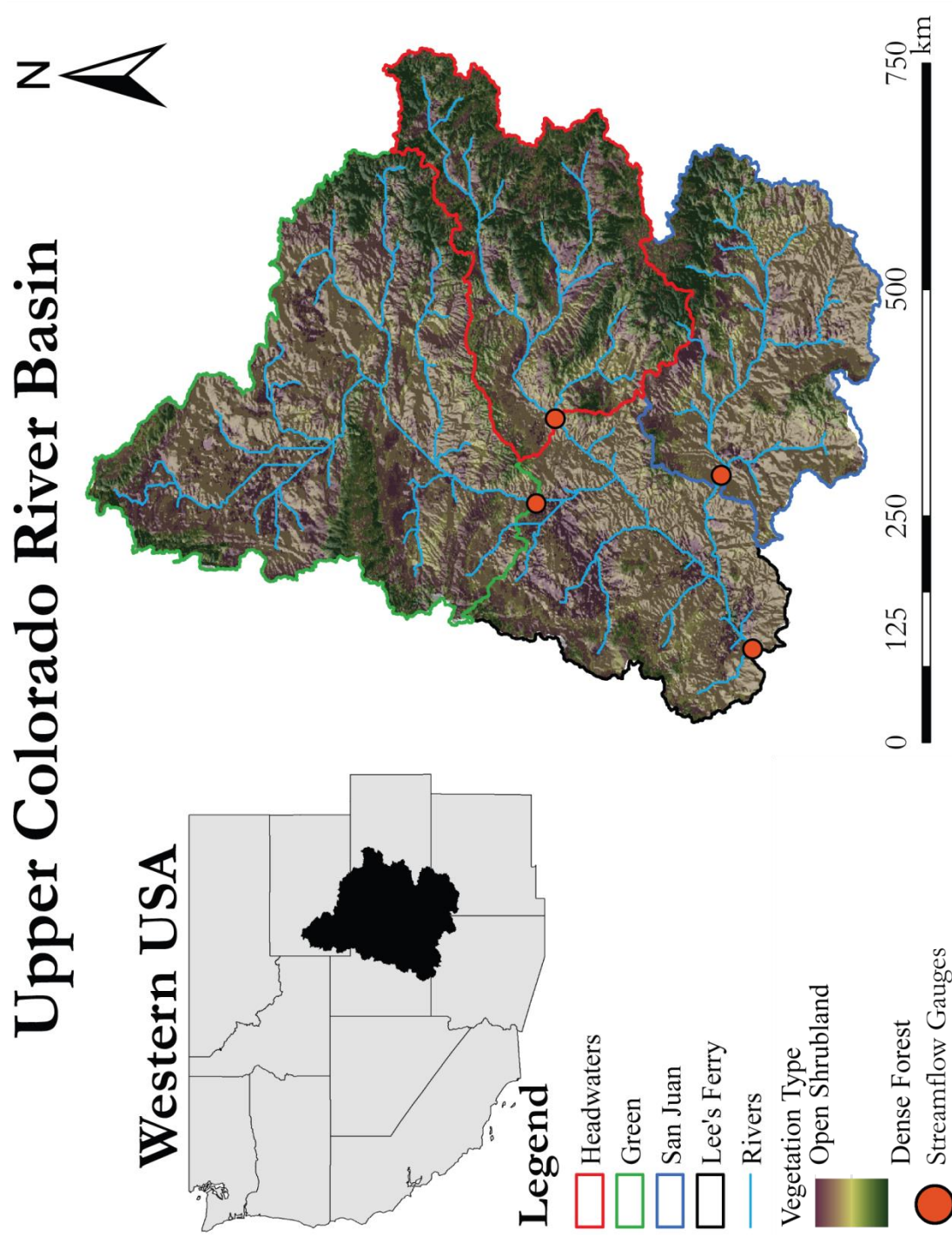


Figure 1. Map of the upper Colorado River Basin, with major sub-basins highlighted, and vegetation type overlying topography.

4.3 Data

4.3.1 Forcing

Forcing datasets required for the VIC model are precipitation, maximum and minimum temperature, wind speed, humidity, and incoming shortwave and longwave radiation. Observed precipitation and temperature data was gathered from the NWS Cooperative Observer Program and Natural Resources Conservation Service Snow Telemetry (SNOTEL) sites, and then spatially distributed over the VIC model grid cells. Distribution of precipitation and temperature was performed with scaling based on PRISM monthly data (Daly et al., 1994), and inverse distance weighting. Wind observations were gathered from the NCEP/NCAR reanalysis dataset (Kalnay et al., 1996) and scaled to the modeling domain similar to the methods in Mauer et al. (2002). Humidity is estimated according to Kimball et al., (1997), shortwave radiation is estimated according to Thornton and Running, (1999), and longwave radiation is estimated according to Bras, (1990), all of which are performed internally in the VIC model. The NWS models require precipitation, temperature and potential evapotranspiration. All forcing for these models was provided by the Colorado Basin River Forecast Center, where precipitation and temperature are estimated from in-situ observations, and potential evapotranspiration is estimated based on pan evaporation rates.

4.3.2 Naturalized Streamflow

The UCRB is a heavily regulated watershed, which makes the use of gauge based observations for forecast verification questionable. A more prudent dataset would remove

the effects of human interactions with the rivers, primarily reservoir operations and water withdrawals. This is referred to as naturalized flow, and requires both in-situ observations and streamflow simulations. Naturalized flow data for the UCRB is provided by the United States Bureau of Reclamation (USDOI, 2005). This dataset contains flow reconstructions throughout the entire Colorado Basin, with 20 reconstruction locations in the upper region. Flows are in the form of monthly volumes. For the sake of this study, 16 of the 20 different flow locations were chosen for verification, based on location within the model routing networks. In addition, this dataset provided the basis for calibration of VIC, which was optimized based on the root mean square error of monthly flow from the four major sub-basins in Figure 1.

4.3.3 Passive Microwave Data

Passive microwave brightness temperature from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) data was used in this study to perform land surface DA. AMSR-E is flown on the Aqua satellite, having a temporal frequency of about 24 hours and a spatial resolution ranging from 12 to 25km. TB was chosen for assimilation in this dissertation as it provides useful information about water stored at the land surface due to its sensitivity to SWE (18.7 and 36.5GHz AMSR-E channels) and soil moisture (6.9 GHz AMSR-E channel). With its sensitivity to both soil moisture and SWE, passive microwave observations provide valuable information about the water storage immediately above and below the land surface. This data is also attractive as it is not obscured by clouds. Passive microwave observations were gathered from the AMSR-E L2A dataset, which was recently shown to have greater information

content, with respect to SWE, than the spatially resampled EASE-Grid products (Li et al., 2012), and then distributed to the centroid of each discretized modeling unit through inverse distance squared weighting.

4.3.4 Land Surface Temperature

Land surface temperature (LST) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) were assimilated in conjunction with TB to improve model estimation of surface temperature, and therefore provide more accurate energy balance estimation and simulations from the radiative transfer models. LST from MODIS is particularly attractive for this dissertation because of the fine resolution observations (1 km by 1 km) and the high frequency (up to 4 observations every 24 hours) of measurements, due to its deployment on both the Aqua and Terra satellites. With high frequency and resolution of the observations, MODIS LST is a powerful source of information about the land surface energy balance. Although the infrared wavelengths observed with this sensor allow for fine spatial resolution, they are obscured by clouds reducing the actual frequency of observations. LST from MODIS is estimated through the Generalized Split-Window Algorithm (Wan and Dozier, 1996), which uses the difference between MODIS bands 31 and 32, along with estimated land surface emissivity, to estimate the temperature of the land surface (Wan and Dozier 1996; Wan et al., 2004). In this study, MODIS LST is aggregated to model resolutions with spatial averaging of observations within each discretized modeling unit.

4.3.5 Data Error Models

Stochastic estimates of precipitation, temperature, wind, and potential evapotranspiration were used to drive the hydrological models. Precipitation uncertainty is assumed to be heteroskedastic and log-normal with a variance of 25% of the magnitude of the nominal value, temperature uncertainty (TB, LST and air temperature) is assumed to be homoskedastic and normal with a standard deviation of 3° C, and both potential evapotranspiration and wind are assumed to have a heteroskedastic normal uncertainty, with variance equal to 25% of the magnitude of the nominal value. The form and magnitude of these uncertainties follow previous studies (DeChant and Moradkhani 2011a,b, 2012; Parrish et al., 2012). In addition, this study also utilizes spatially and temporally correlated perturbations as described by Clark and Slater (2006). For application of this method, a correlation length of 100km and temporal correlation of 5 days was used for all inputs, which was suggested by Clark et al., (2008a)

5 Proposed Forecasting Framework

5.1 Particle Filter

The PF is a methodology to sequentially apply Bayes Theorem for reducing model estimated information at each time when an observation of the system is available (Gordon et al., 1993). Through this sequential updating scheme, the PF provides revised model estimates based on the observation, which is referred to as the posterior. In order to apply the PF, one must start with Bayes Law (equation (10)).

$$p(x | \hat{y}) = \frac{p(\hat{y} | x)p(x)}{p(\hat{y})} \quad (10)$$

In equation (10), $p(x)$ is the prior probability of some model estimated value x , $p(\hat{y})$ is the probability of the observations, $p(\hat{y} | x)$ is the probability of the observation given the model estimate, which is referred to as the likelihood, and $p(x | \hat{y})$ is the revised probability of x given the observation, referred to as the posterior. In addition, the ^ accent will note an observed quantity for the remainder of this dissertation. For sake of this study, the notation x is used to show the model estimated states. Application of equation (10) for the PF requires derivation of a sequential form, which is shown in equation (11).

$$p(x_t | \hat{y}_{1:t}) = \frac{p(\hat{y}_t | x_t)p(x_t | \hat{y}_{1:t-1})}{p(\hat{y}_t | \hat{y}_{1:t-1})} \quad (11)$$

In equation (11), the posterior value is now the probability of the current states, given all observations ($p(x_t | \hat{y}_{1:t})$), the likelihood is the probability of the current observation given the current state ($p(\hat{y}_t | x_t)$), the prior is the probability of the state given all

previous observations ($p(x_t | \hat{y}_{1:t-1})$), and the observation probability becomes the probability of the current observation given all past observations ($p(\hat{y}_t | \hat{y}_{1:t-1})$). Although the likelihood is readily calculated through a representative likelihood function, the prior and observation probabilities are not readily calculated, requiring further attention before evaluation of equation (11) is possible.

Equation (14) shows the final form of sequential Bayes Theorem for use in the filtering problem, which is developed from equations (12) and (13). A first requirement in the application of the PF is that the model is a first order Markov Process. By assuming that the model is Markovian, which is implicit in equation (1), the Chapman-Kolmogorov equation may be applied to estimate the prior probability, as described in equation (12). From this equation, the prior distribution is found to be equivalent to the integral of the product of the transition probability ($p(x_t | x_{t-1})$) and posterior at the previous time. The posterior at the previous time-step will be available at all times greater than 1, and the transition probability is equivalent to the model probability. Further, the observation probability may be estimated by treating the current states as an intermediate variable. Conveniently, equation (13) becomes the integration of the numerator of equation (14), therefore showing that sequential Bayes Law is the normalized product of the likelihood and the integration of the transition probability and posterior at the previous time step. At this point, the likelihood function, transition probability and previous posterior probability are the only quantities necessary to estimate the current posterior.

$$p(x_t | \hat{y}_{1:t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | \hat{y}_{1:t-1}) dx_{t-1} \quad (12)$$

$$p(\hat{y}_t | \hat{y}_{1:t-1}) = \int p(\hat{y}_t | x_t) p(x_t | \hat{y}_{1:t-1}) dx_t \quad (13)$$

$$p(x_t | \hat{y}_{1:t}) = p(x_t | \hat{y}_t, \hat{y}_{1:t-1}) = \frac{p(\hat{y}_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | \hat{y}_{1:t-1}) dx_{t-1}}{\int p(\hat{y}_t | x_t) \left[\int p(x_t | x_{t-1}) p(x_{t-1} | \hat{y}_{1:t-1}) dx_{t-1} \right] dx_t} \quad (14)$$

Equation (14) allows for theoretical application of sequential Bayes Theorem, but further work is required to apply this to a real model. This equation will not have a tractable analytical solution in the applications examined here, and therefore a Monte Carlo experiment is necessitated to simulate the probability distributions. Since this is applied as a Monte Carlo experiment, using an ensemble to represent the state distribution, the posterior probability can be written according to equation (15), where $w_{i,t}$ represents the weight of ensemble member i and time step t , δ is the Dirac Delta function and N is the ensemble size. This equation represents an importance sample at time t . For application of the PF, Sequential Importance Sampling is performed by applying Bayes Theorem at each observation time, which simplifies to the normalized product of the likelihood and prior weights (equation (17)), given that the ensemble of land surface states and predictions are generated from equations (5) and (6). Thus the transition probability is treated through the stochastic model operators. Equation (16) shows the calculation of the likelihood, which is the probability of the residual ($\hat{y}_i - y_{i,t}$) given some expected distribution of the residuals. A Gaussian likelihood is assumed in this dissertation, with variance of R_t , which is estimated as the observation error from section 2.3.5.

$$p(x_t | \hat{y}_t) \approx \sum_{i=1}^N w_{i,t} \delta(x_t - x_{i,t}) \quad (15)$$

$$L(\hat{y}_t | x_{i,t}) = p(\hat{y}_t - y_{i,t} | R_t) \quad (16)$$

$$w_{i,t} = \frac{L(\hat{y}_t | x_{i,t})w_{i,t-1}}{\sum_{i=1}^N L(\hat{y}_t | x_{i,t})w_{i,t-1}} \quad (17)$$

Sequential Importance Sampling provides a weighted ensemble of model states at each time step, which allows for representation of the posterior distribution. Assuming sufficient sample size, properly tuned error estimates, and accurate likelihood function, Sequential Importance Sampling will provide an exact sample from the true posterior distribution, with respect to the modeling framework. Although Sequential Importance Sampling provides the ability to perfectly estimate the posterior, over a sufficient number of model simulations, the required sample size will become too large to remain computationally feasible. In this scenario, the variance of weights will approach zero, as the weight of only a few ensemble members approaches unity, and the rest approach zero. This is referred to as weight degeneration. In order to overcome this problem, a resampling step may be performed, where the ensemble members of high weights are duplicated, and the ensemble members of low weights are discarded, thus ensuring a sample that remains in a meaningful portion of the posterior distribution. This is referred to as Sampling Importance Resampling, and generates a sample of equally weighted ensemble members (Arulampalam et al., 2002). The Sampling Importance Resampling algorithm is used in this dissertation, with resampling being performed at each observation time step. The resampling algorithm used is Multinomial Resampling (Douc and Cappe, 2005).

5.2 Bayesian Multi-Modeling

Bayesian multi-modeling is a class of techniques that provide a weighted sample of models to quantify and reduce model uncertainty. Of these methods, Bayesian Model Averaging is the simplest technique. Bayesian Model Averaging was introduced to the climate forecasting community by Raftery et al. (2005) and later applied to hydrological modeling (Ajami et al., 2007; Duan et al., 2007). Bayesian Model Averaging extends the application of Bayes Law to the case of multiple possible models, where it is assumed that the varying model behavior implicitly represents the uncertainty in those models. In this methodology, an ensemble of models will be averaged according to equation (18). In equation (18), M_k represents the k^{th} model, out of an ensemble of K models, $\hat{y}_{1:t-1}$ is the training data from for all observations up to the previous time step, and $p(M_k | \hat{y}_{1:t-1})$ is the posterior model probability. This equation represents the sum of the product of the posterior model probabilities and the forecast probability distributions ($p(y_{k,t} | M_k)$). In this application, model forecasts are typically deterministic, but probabilistic predictions are required, which necessitates the fitting of a probability distribution to the model. Within hydrological and atmospheric sciences, it is generally assumed that forecast uncertainty is normally distributed, according to equation (19), where $g(y_t | y_{k,t}, \sigma_k^2)$ applies a normally distributed uncertainty to the forecast from model k , with a mean of $y_{k,t}$ and a variance of σ_k^2 . Finally, the probabilistic forecast is estimated according to equation (20), where w_k is the model weight, which is an estimate of the posterior model probability. For evaluation of equation (20), both the model weights and variances are

necessary for model averaging. This creates a two parameter optimization problem, which is typically solved with the Expectation-Maximization algorithm, to maximize the forecast likelihood over the training period. Although Expectation-Maximization is commonly applied, any optimization algorithm may be used, and therefore it may be more prudent to use a global optimization algorithm (Duan et al., 2007).

$$p(y_{1:K,t} | M_{1:K}, \hat{y}_{1:t-1}) = \sum_{k=1}^K p(y_{k,t} | M_k) p(M_k | \hat{y}_{1:t-1}) \quad (18)$$

$$p(y_{k,t} | M_k) = g(y_t | y_{k,t}, \sigma_k^2) \quad (19)$$

$$p(y_{1:K,t} | M_{1:K}, \hat{y}_{1:t-1}) = \sum_{k=1}^k g(y_t | y_{k,t}, \sigma_k^2) * w_k \quad (20)$$

Bayesian multi-modeling may be extended to the case of sequential weight estimation. Similar to the PF, it may be advantageous to update the weights at each available observation, thus having a sequential forecasting system. The sequential form of Bayesian Model Averaging is referred to as Sequential Bayesian Combination, and was proposed in Hsu et al. (2009). In this approach, model weights are calculated based on sequential Bayes Theorem, as shown in equation (21). This leads to dynamic model weights, which are used to create a forecast at each time step, similar to equation (19). In order to evaluate equation (21), some likelihood must be chosen, which is Gaussian in this study, and an a priori variance must be estimated, as shown in equation (22). Estimation of this variance is based on the residuals during the calibration period, as suggested by Hsu et al. (2009).

$$p(M_k | \hat{y}_{1:t}) = \frac{p(\hat{y}_t | M_k) p(M_k | \hat{y}_{1:t-1})}{\int p(\hat{y}_t | M_k) p(M_k | \hat{y}_{1:t-1}) dM} \approx w_{k,t} = \frac{L(\hat{y}_t | M_k) w_{k,t-1}}{\sum_{k=1}^K L(\hat{y}_t | M_k) w_{k,t-1}} \quad (21)$$

$$L(\hat{y}_t | M_k) = p(\hat{y}_t - y_{k,t} | \sigma_k^2) \quad (22)$$

5.3 Combining the Particle Filter and Bayesian Multi-Modeling

A major obstacle for Bayesian multi-modeling is the reliance on calibrated parametric distributions to model forecast uncertainty, as shown in equation (19). Although a parametric distribution allows for a simple approach for uncertainty estimation, it will be limiting based on the knowledge that model residuals typically have complex distributions (i.e. heteroschedastic and non-Gaussian). Since the model residuals are complex, a parametric distribution which properly fits the residuals is unlikely to exist. In order to overcome this problem, Parrish et al., (2012) proposed a method that uses the PF to generate the forecast distribution for model averaging. Since the PF can account for all sources of uncertainty simultaneously, and imposes only a weak assumption of Gaussianity, the PF generates a more appropriate forecast distribution than by simply assuming a distribution is Gaussian. Therefore the predictive distribution from the PF ($p(x_t | \hat{y}_{1:t-1})$) may replace equation (19), which must be expanded to the case of multiple models, as shown in equation (23). In equation (23), the forecast probability is set equal to $p(y_{k,t} | x_t, M_k, \hat{y}_{1:t-1})$, which is conditioned on all previous observations (filtering prior), for the k^{th} model. The forecast probability is expanded to the integration of the product of the prediction from the current states, $p(y_{k,t} | x_t)$ (observational operator), and the posterior at the previous time step from the PF, $p(x_t | M_k, \hat{y}_{1:t-1})$, which

is approximated by the importance sample of the predictions from model k . In order to evaluate the model probability, as is performed in equation (21), a Kernel Smoothing density is used to locate the probability of locations between the ensemble members. For the case of sequential model weight estimation, the model posterior weights are estimated according to equation (24), where $KS(\hat{y}_t | y_{k,t}, M_k)$ is a Kernel Smoothing estimate of the likelihood, conditioned on the probabilistic model prediction. This leads to a final model averaged forecast according to equation (25). Through this methodology, a model averaging system is developed, which simultaneously accounts for initial condition, forcing and model uncertainties, thus providing a useful framework for the forecasts of interest in this dissertation.

$$\begin{aligned} p(y_{k,t} | M_k) &= p(y_{k,t} | x_t, M_k, \hat{y}_{1:t-1}) \\ &= \int p(y_{k,t} | x_t) p(x_t | M_k, \hat{y}_{1:t-1}) dx_t \approx \sum_{i=1}^N \delta(y_t - y_{i,k,t}) w_{i,k,t-1} \end{aligned} \quad (23)$$

$$p(M_k | \hat{y}_{1:t}) \approx w_{k,t} = \frac{KS(\hat{y}_t | y_{k,t}, M_k) w_{k,t-1}}{\sum_{k=1}^K KS(\hat{y}_t | y_{k,t}, M_k) w_{k,t-1}} \quad (24)$$

$$p(y_{1:K,t} | M_{1:K}, \hat{y}_{1:t-1}) \approx \sum_{k=1}^K p(y_{k,t} | x_t, M_k, \hat{y}_{1:t-1}) * w_{k,t} \quad (25)$$

5.4 Ensemble Hydrological Forecasting

Among the most common ensemble hydrological forecasting systems is ESP, which is used by the NWSRFCs to create probabilistic forecasts of streamflow volumes at various lead times. This method leverages stochastic meteorological data to generate a Monte Carlo sample of streamflow predictions to represent the uncertainty of in future

streamflow estimates. As described in equation (26), the ESP technique creates an Monte Carlo sample of streamflow simulations over the forecast period, with each individual simulation referred to as a “trace”, generated from the deterministic states, x_t , resampled historical forcing, $u_{1:Y-1,t}$ (Y represents the current year), and some hydrological model M , from the initial forecast time t , to estimate the probability of volumetric streamflow forecast \tilde{y}_t . Note that $M(x_t, u_{y,t}, \theta, \alpha)$ represents the combined hydrologic model (f) and routing function (h), which forecasts volumetric streamflow over a desired lead time, and $w_{y,t}$ is the weight of each trace from initial forecast time t , which is typically set uniformly to $\frac{1}{Y-1}$. While the common practice is to give each trace equal weight, these weights may be estimated with additional climate information (Najafi et al., 2012), thus developing an importance sample of streamflows.

$$p(\tilde{y}_t | x_t, u_{1:Y-1}, M) = \sum_{y=1}^{Y-1} p(\tilde{y}_t | x_t, u_{y,t}, M) p(u_{y,t}) \approx \sum_{y=1}^{Y-1} \delta(\tilde{y}_t - M(x_t, u_{y,t}, \theta, \alpha)) w_{y,t} \quad (26)$$

Initial states for this method are generated through a deterministic spin-up, performed with observed historical forcing. Starting at this point, the model is forced with resampled historical forcing ($u_{1:Y-1,t}$), beginning at the initial forecast date, for each historical observation year, to estimate meteorological climatology, thus producing a stochastic streamflow forecast. Within this framework, climatology is used to represent the total uncertainty related to the forcing data. By generating a stochastic streamflow forecast through ESP, it is assumed that the initial state estimates are perfect, the resampled historical climate variability represents the future climate uncertainty, which inherently

assumes climate stationarity, and the model is perfect. Given that the assumptions of climate stationarity, accurate model initial state estimates and perfect model structure are not significantly violated, ESP will provide a reliable probabilistic prediction of volumetric streamflow at the desired lead time.

5.5 Framework for Comprehensive Accounting of Hydrological Uncertainty

In any real application of ESP, the assumptions of perfect initial states and model structure will be violated, necessitating a methodology to account for uncertainties in the states and models, which is performed here with a combination of the PF and Sequential Bayesian Combination (referred to here as PF-SBC). Application of PF-SBC to the ESP framework requires two steps, which are outlined in Figure 2. First, DA is performed in each model during the spin-up period to produce stochastic states at each initial forecast date ($x_{1:N,t}$), following the application by DeChant and Moradkhani (2011b). This ensemble of land surface states represents the uncertainty at the initial forecast date, and a resulting distribution of streamflow forecasts from these states is shown in equation (27), where $p(x_{i,t} | \hat{y}_{1,t})$ is the posterior distribution from the PF, and the weights ($w_{i,t}$) are uniform due to the application of Sampling Importance Resampling. Equation (27) describes the single model ESP with DA forecast developed in DeChant and Moradkhani (2011b).

$$\begin{aligned}
 p(\tilde{y}_t | x_{1:N,t}, u_{1:Y-1}, M, \hat{y}_{1,t}) &= \sum_{y=1}^{Y-1} \sum_{i=1}^N p(\tilde{y}_t | x_{i,t}, u_{y,t}, M) p(u_{y,t}) p(x_{i,t} | \hat{y}_{1,t}) \\
 &\approx \sum_{y=1}^{Y-1} \sum_{i=1}^N \delta(\tilde{y}_t - M(x_{i,t}, u_{y,t}, \theta, \alpha)) w_{y,t} w_{i,t}
 \end{aligned} \tag{27}$$

A second step is performing ESP with DA for all models. In this step, each model and initial state ensemble member is propagated forward with the ESP framework, providing a multi-model ensemble forecast, creating $N * (Y - 1) * K$ streamflow traces. At this point, a stochastic multi-model forecast is available, represented by $p(\tilde{y}_t | x_{1:N,t}, u_{1:Y-1}, M_{1:K})$, but this does not account for the relative accuracy of each model, thus creating an overly uncertain forecast. In order to reduce this uncertainty, each model is averaged according to PF-SBC, based on weights estimated from the observations during the spin up period ($\hat{y}_{1:t}$). The proposed methodology estimates the probability distribution of \tilde{y}_t based on resampled historical forcing, stochastic states and multiple weighted models, as represented in equation (28). In the application presented here, both $w_{i,t}$ and $w_{y,t}$ are uniform, and therefore their product is equal to $1/(N*(Y-1))$, making the weight of each trace estimated by equation (29), where $p(M_k | \hat{y}_{1:t})$ is estimated according to equation (24).

$$\begin{aligned}
 & p(\tilde{y}_t | x_{1:N,t}, u_{1:Y-1}, M_{1:K}, \hat{y}_{1:t}) \\
 &= \sum_{y=1}^{Y-1} \sum_{i=1}^N \sum_{k=1}^K p(\tilde{y}_t | x_{i,t}, u_y, M_k) p(u_y) p(x_{i,t} | \hat{y}_{1:t}) p(M_k | \hat{y}_{1:t}) \\
 &\approx \sum_{y=1}^{Y-1} \sum_{i=1}^N \sum_{k=1}^{K-1} \delta(\tilde{y}_t - M_k(x_{i,t}, u_{y,t}, \theta, \alpha)) w_{y,t} w_{i,t} w_{k,t}
 \end{aligned} \tag{28}$$

$$w_{k,i,y,t} = w_{y,t} w_{i,t} w_{k,t} = \frac{1}{N*(Y-1)} p(M_k | \hat{y}_{1:t}) \tag{29}$$

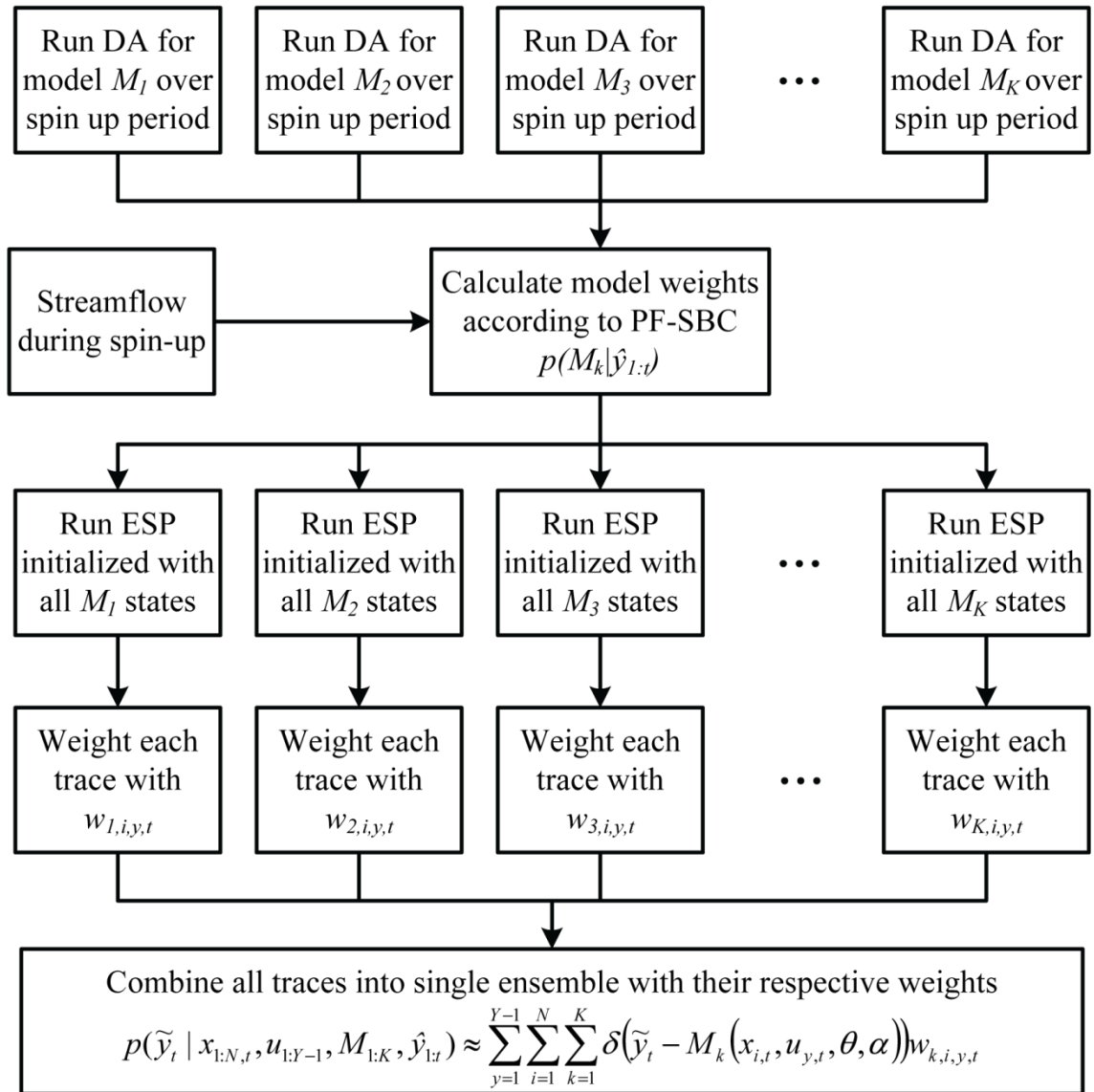


Figure 2. Flowchart describing the application of the PF-SBC algorithm to streamflow forecasting with ESP.

5.6 Drought Quantification

5.6.1 Climatological Indices

Hydrologic drought is quantified in this study with the Standardized Streamflow Index (SSI) (Mckee et al., 1993). SSI is described in equation (30) and (31), where the forecasted flow volume $\tilde{y}_{t,y,i,k}$, for initial start date t , based on forcing year y , ensemble member i and model k , is transformed, based on an assumed distribution $F_{t,m}$, into the space of the observed flow cumulative distribution ($\mu_{t,y,i,k}$). In this dissertation, $F_{t,m}$ is a log-normal distribution fit to all historical flow observations for start date t and the corresponding forecast length m . In Madadgar and Moradkhani (2013), the log-normal distribution was found to be a generally reasonable fit in the Gunnison River Basin, which is a sub-basin of UCRB, motivating its use in this dissertation. After transforming flows into the probability space, they are translated into the normal space with the inverse normal distribution (ϕ^{-1}) to generate the corresponding index value ($SSI_{t,y,i,k}$).

$$\mu_{t,y,i,k} = F_{t,m}(\tilde{y}_{t,y,i,k}) \quad (30)$$

$$SSI_{t,y,i,k} = \phi^{-1}(\mu_{t,y,i,k}) \quad (31)$$

In addition to the SSI, this study will also examine a spatially distributed drought index. This is a standardized index of the total land water storage (LWS), which is defined here as the sum of soil moisture (between the surface and 1 meter below the land surface) and SWE. Throughout this dissertation, the standardized version of the LWS will be referred to as the Standardized Land Water Index (SLWI). Similar to the SSI, the SLWI is estimated based on historical climate during a given month. Following equations

(30) and (31), but substituting the LWS for streamflow, the SLWI is a standardized value assuming a log-normal distribution. With respect to the SLWI, only the VIC model will be used as the SAC-SMA model does not estimate physical soil moisture, and historical climate is based on historical model simulations, as opposed to the SSI which is based on the observed historical record.

5.6.2 Physically-Based Index

In order to develop a physically-based drought index, this dissertation builds on the Soil Moisture Index (SMI) developed in Hunt et al. (2009). The SMI is based on the fraction of available water (F_{AW}), calculated in equation (32), where SM is the soil moisture, SM_{FC} is the field capacity and SM_{WP} is the wilting point. In order to convert the fraction of available water to an index, the value is scaled between $-b$ and b , according to equation (33), which assumes that water stress in crops occurs at a F_{AW} around 0.5 (Baier, 1969), thus providing a physically meaningful drought measure. Though this index was formulated for small scale agricultural purposes, it has the potential to be generalized for a wider drought analysis. In this dissertation, a Modified SMI (MSMI) is presented to account for snow, in addition to soil moisture, thus generalizing the SMI for use in regions where snow is significant. In this index, the soil water content is adjusted with the addition of SWE, according to equation (34), with the assumption that a majority of the SWE will melt and enter the soil matrix. This creates a new F_{AW} value, which may be scaled similarly to the SMI, as shown in equation (35).

$$F_{AW} = \frac{SM - SM_{WP}}{SM_{FC} - SM_{WP}} \quad (32)$$

$$SMI = -b + 2bF_{AW} \quad (33)$$

$$SM_A = \theta + SWE \quad (34)$$

$$MSMI = -b + 2b \frac{SM_A - SM_{WP}}{SM_{FC} - SM_{WP}} \quad (35)$$

5.6.3 Drought Forecast Values

Deterministic and stochastic forecasts of streamflow drought are generated in this dissertation, which requires the designation of drought expectation and probability. From the stochastic forecasts of monthly and seasonal SSI described in section 5.6.1, drought expectation and drought probability may be estimated, assuming a drought threshold of -0.5. This threshold is presented mathematically in equation (36), where $d_{t,y,i,k}$ is a binary value with a value of one indicating drought and a value of zero indicating no drought. For probabilistic forecasting, the drought probability (p_t) is estimated according to equation (37). In order to produce a deterministic forecast, the drought expectation (D_t) is estimated based on the drought probability, where a drought is forecasted if the p_t is greater than 0.5, which is shown in equation (38). Finally, drought observations are described in equation (39), where O_t the drought observation, with one indicating a drought being observed and zero indicates no drought being observed. Based on these values, forecast will be performed as described in Chapter 7.

$$d_{t,y,i,k} = \begin{cases} 1 & \text{if } SSI_{t,y,i,k} \leq -0.5 \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

$$p_t = \sum_{y=1}^{Y-1} \sum_{i=1}^N \sum_{k=1}^K (w_{t,y,i,k} * d_{t,y,i,k}) \quad (37)$$

$$D_t = \begin{cases} 1 & \text{if } p_t > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (38)$$

$$O_t = \begin{cases} 1 & \text{if drought} \\ 0 & \text{otherwise} \end{cases} \quad (39)$$

6 Forecast Verification

6.1 Underlying Theory of Forecast Verification

Forecast verification falls into two categories, deterministic and probabilistic, for two different types of forecasts, continuous and discrete. With respect to the different categories of forecast verification, deterministic measures are more common, as most forecasting frameworks are developed based on deterministic methods, and the framework is more intuitive for a wider audience. The basis for deterministic measures is demonstrating a reduction in errors of the forecast technique. Such a focus on reducing errors is understandable for forecasters and the general public alike, but the utility of deterministic measures may become questionable when noting that they require assumptions about the distribution of errors, which is typically non-gaussian, and may be biased towards overconfident or underconfident forecasts. Therefore this study will focus primarily on probabilistic methods, with the addition of a few deterministic measures for drought forecasting.

Probabilistic measures, for both continuous and discrete forecasts, are necessary for determining the utility of a forecast. Primarily, a user of a probabilistic forecast will be interested in performing some analysis of the risk of an event occurring. Given some level of risk, and the consequence of an event occurring, a practitioner will determine what action, or lack of action, is appropriate to mitigate damages. In order to effectively perform such management of risk, the forecaster will provide information that represents the true probability that an event will occur. Such a forecast is termed reliable, and is defined here as a forecast that satisfies the indistinguishability paradigm (Annan and

Hargreaves, 2010). This paradigm states that a forecast is reliable if the distributions from which the forecast and observation are being drawn are indistinguishable, and therefore the forecasted probabilities may be assumed to follow the same probability distribution as the observation. Given that a forecast is reliable, a second quality is desired: forecast sharpness. Sharpness refers to the certainty of a forecast (tendency towards 0 or 1 for discrete forecasts, and forecast variance approaching 0 for continuous forecasts). Increased sharpness (reduced uncertainty) is noted as a highly desirable attribute of a reliable forecast, indicating to more confident predictions. With these two attributes desired, this dissertation takes the position that an optimal forecast will maximize sharpness, given the condition of reliability (Gneiting et al., 2007, Pal, 2009). Assuming that two reliable forecasts are available for some phenomena, the sharper forecast will be selected. Further, given that a forecast is reliable, it will be chosen over all unreliable forecasts, regardless of sharpness. The following probabilistic measures will be analyzed in this dissertation, and each examines one or both of the desired forecast characteristics.

6.2 Continuous Predictands

6.2.1 Probabilistic Measures

Reliability assessment in forecasts of continuous variables is challenging because the forecaster will not know the true form of the forecast uncertainty. Many forecast verification frameworks will assume that the uncertainty is Gaussian, assuming that the central limit theorem is applicable. Although this strategy is common, it is approximate. Since a forecaster does not know the true form of the uncertainty for each forecast, a better strategy is to rely on the Probability Integral Transform (PIT) theorem. This

theorem states that a random variable, being drawn from any continuous distribution, may be translated into a uniform random variable through the cumulative distribution function (CDF) of the sampling distribution, as shown in equation (40).

$$z = \int_{-\infty}^{\hat{y}} p(y)dy \quad (40)$$

In equation (40), $p(y)$ is the PDF of the forecast variable, \hat{y} is the random variable being drawn from that distribution, and z becomes a uniformly distributed random variable if \hat{y} is drawn from $p(y)$. Based on this theorem, the reliability of the forecast may be examined by testing if a string of observations appear to be random variables drawn from the forecast distributions. Therefore one will test the uniformity of the PIT. Given that the PIT is uniform, the observation and forecast distributions may be deemed indistinguishable, thus validating the hypothesis that the forecast is reliable. Based on the PIT, three probabilistic verification methods are used in this dissertation: The Exceedance Ratio (ER) (Moradkhani et al., 2006; Moradkhani and Meskele, 2009), the reliability (R) metric (DeChant and Moradkhani, 2011b; Renard et al., 2010) and the predictive Quantile-Quantile (QQ) plot (Laio and Tamea, 2007).

The ER can be applied to any predictive quantile range desired, but here it is used to analyze the tails of the distribution (99%, 95% and 90%). Calculation of the ER of a given quantile range is performed according to equations (41), (42), and (43), with equation (41) being the application of the PIT. In these equations, the cumulative probability of each observation, given its respective forecast, is represented by z_t . The uniformity of vector z can then be analyzed at any predictive bound P desired, as shown

in equation (42). Then the ER is estimated as the percentage of observations that fall outside this quantile range, which should be equal to $\frac{1-P}{2}$, which is evident from the uniform distribution. Since the ER is used to examine the tails of the distribution, a metric is also necessary to examine the whole probabilistic forecast. R is used here to examine the entire predictive distribution. This measure describes the average absolute difference between the PIT and the uniform CDF, and is described in equations (41), (44), and (45). Similar to the ER, a vector z is calculated, and then sorted in ascending order, as noted in equation (44). R can then be calculated according to equation (45), where the \hat{z} are compared against the uniform cumulative density to determine the accuracy of the probabilistic forecast. A value of one equals a perfect forecast and a value of zero is the mathematically worst forecast.

$$z_t = \sum_{k=1}^K \sum_{i=1}^N \sum_{y=1}^{Y-1} \chi_{k,i,y,t} \quad \chi_{k,i,y,t} = \begin{cases} w_{k,i,y,t} & \hat{y}_t > \tilde{y}_{k,i,y,t} \\ 0 & \hat{y}_t < \tilde{y}_{k,i,y,t} \end{cases} \quad (41)$$

$$\varepsilon'_t = \begin{cases} 1 & \text{if } z_t > 1 - \frac{1-P}{2} \text{ or } z_t < \frac{1-P}{2} \\ \text{otherwise} & 0 \end{cases} \quad (42)$$

$$ER = 100 \times \sum_{t=1}^T \frac{\varepsilon'_t}{T} \quad (43)$$

$$\hat{z} = \text{sort}(z_{1:t}) \quad (44)$$

$$R = 1 - 2 \sum_{t=1}^T \left| \hat{z}_t - \frac{t}{T} \right| \quad (45)$$

The predictive QQ plot is used as a visual method to diagnose errors in the forecast distribution. This plot compares the \hat{z} vector on the x-axis, and the uniform distribution on the y-axis, to examine the reliability of the forecast distribution. Given

that the plot follows a 1:1 line, this forecast is perfectly reliable ($R=1$). If the predictive QQ line falls above the 1:1 line, the forecast has a high bias (observations have a tendency to fall too low in the forecast distribution) and a predictive QQ line falling below the 1:1 line indicates a low bias. In addition to bias, the over/underconfidence of a forecast distribution can be diagnosed with this plot. If the predictive QQ line falls above the left side of the 1:1 line and crosses the 1:1 line in the middle, the forecast is overconfident (a disproportionately high number of observations being captured by the outer quantiles) and the reverse indicates underconfidence. For a more complete explanation of this plot, see Laio and Tamea (2007).

6.3 Discrete Predictands

6.3.1 Deterministic Measures

Droughts are discrete events with only two possible outcomes, occurrence or non-occurrence, simplifying the performance measures from the more general categorical events to binary. The binary nature of droughts lends itself to three deterministic measures: The Probability of Detection (POD) (equation (46)), the False Alarm Ratio (FAR) (equation (47)) and the Critical Success Index (CSI) (equation (48)) (Wilks, 2006). In these equations, TP (true positives) is the number of correctly forecasted drought occurrences ($O_t = 1$ and $D_t = 1$), FP (false positives) is the number of forecasted drought occurrences in which a drought did not occur ($O_t = 0$ and $D_t = 1$), and FN (false negatives) is the number of drought occurrences that were not forecasted to occur ($O_t = 1$ and $D_t = 0$). POD provides a measure of the likelihood that any given occurrence of

drought will be forecasted. Since this score may be perfect in the event of extreme bias (i.e. all D_t are 1), the FAR is provided to give a measure the probability of a drought prediction being a false positive. Alternatively, FAR will be perfect in extreme bias in the opposite direction (i.e. all D_t are 0). This suggests that an overall metric may be more useful for drought forecast analysis. This is provided with the CSI, where the number of correctly forecasted droughts (TP) divided by the number for forecasts in which a drought is observed, forecasted or both ($TP + FP + FN$). With this mathematical setup, CSI is designed to reward correctly forecasting drought events, and penalize false positives. Therefore the CSI is only maximized when the forecasts perfectly match the observations, providing a generally more useful metric than POD and FAR.

$$POD = \frac{TP}{TP + FN} \quad (46)$$

$$FAR = \frac{FP}{TP + FP} \quad (47)$$

$$CSI = \frac{TP}{TP + FP + FN} \quad (48)$$

6.3.2 Probabilistic Measures

6.3.2.1 Approximate Measures

Among the most common probabilistic verification measures for binary events in hydrometeorology is the Brier Score (BS) (Brier, 1950), shown in equation (49). The BS has several beneficial features that have sustained its continued use over the past half century. First, it is strictly proper (Gneiting and Raftery, 2007) and therefore is optimized if and only if a forecast is perfect with respect to the observation. Propriety also implies

that the forecast is multi-objective, thus examining reliability and sharpness components. This multi-objective nature leads to a second major benefit of the BS, it may be decomposed into its separate components for more detailed analysis (Murphy, 1973). A decomposition of the BS is provided in equation (50), where the reliability, resolution and uncertainty, shown respectively as different terms on the right hand side of equation (50), are displayed as different components of the forecast. This is a common representation of the BS decomposition, but it should be noted that equation (50) assumes sufficient bin size to ignore within-bin variance of the forecast (Stephenson et al., 2008). From this decomposition, the resolution and uncertainty are shown to be properties of the observations, and therefore independent of the forecast, making these components unimportant when comparing forecasts of the same phenomena. In order to calculate the decomposed elements of the BS, it is necessary to group similarly valued probabilistic forecasts into B forecast bins. Each bin has a population of n_b , with an average forecasted probability of \overline{p}_b and an observed frequency of \overline{O}_b , and the final variable to be defined is the overall observation frequency \overline{O} . Of high importance here is the reliability component, which measures the mean square error of the bin frequencies. A rarely discussed point, but important for this discussion, is that the reliability component of this decomposition is a normal approximation to the Binomial Distribution, which leads to a third benefit of the BS. Since each bin represents a sample from the Binomial Distribution, and a normal approximation will asymptotically approach the exact solution with increasing sample size (Feller, 1945), the BS may become a nearly perfect estimator of reliability as sample size increases. A final benefit of the BS is the simplicity of the

measure. With such a simple design for scoring competing probabilistic forecasts, comparison of competing forecasts is readily performed, and will have a high level of accuracy if sample size is sufficient.

$$BS = \frac{1}{T} \sum_{t=1}^T (p_t - O_t)^2 \quad (49)$$

$$BS = \frac{1}{T} \sum_{b=1}^B n_b (\bar{p}_b - \bar{O}_b)^2 - \frac{1}{T} \sum_{b=1}^B n_b (\bar{O} - \bar{O}_b)^2 + \bar{O}(1 - \bar{O}) \quad (50)$$

A second verification method in this study is the reliability diagram (Franz et al., 2003, Wilks, 2006). The reliability diagram employs a similar binning methodology to compare forecasted probabilities with observed frequencies as the BS, but is presented graphically for assessing the accuracy of different forecast bins. Grouped forecast probabilities (\bar{p}_b) are placed on the x-axis and the observed frequencies (\bar{O}_b) are plotted on the y-axis, and compared against the cumulative uniform distribution (one-to-one line). The reliability diagram should be approximately uniform, as the expectation is $\bar{O}_b = \bar{p}_b$ given a reliable forecast distribution. By graphically showing the deviations from the expected observation frequency, bin dependent biases can be observed, helping diagnose forecast errors. Although proximity to the uniform line aids in understanding errors in probabilistic forecasting, it is not entirely consistent with the proper statistical distribution (Binomial), and therefore may not be entirely reliable (Bröcker and Smith, 2007). Through the recognition that each bin is a separate binomial distribution, the deviations shown in the reliability diagram are known to be approximations of the probability that a bin is reliable. Therefore, it is best to view the reliability from the perspective of the Binomial distribution.

The Binomial distribution describes the probability of a number of events occurring (K) over a given number of trials (n_b), with the event having a certain probability ($\overline{p_b}$). The PDF of the binomial distribution is shown in equation (51), where the $\binom{n_b}{K}$ is the binomial coefficient, which calculates the total number of combinations that K events may occur within n_b trials, and each of those combinations has a probability of $\overline{p_b}^K (1 - \overline{p_b})^{(n_b - K)}$, which makes $f(K, n_b, \overline{p_b})$ equal to the probability of the given scenario occurring. In Bröcker and Smith (2007), the Binomial CDF, shown in equation (52), was used to translate the reliability diagram into probability space, for more precise comparisons of different forecasts. This methodology allows for more effective analysis of forecast reliability within the reliability diagram, more accurately determining which forecast is most likely to be reliable.

$$f(K, n_b, \overline{p_b}) = \binom{n_b}{K} \overline{p_b}^K (1 - \overline{p_b})^{(n_b - K)} \quad (51)$$

$$F(K, n_b, \overline{p_b}) = \sum_{k=0}^K \binom{n_b}{k} \overline{p_b}^k (1 - \overline{p_b})^{(n_b - k)} \quad (52)$$

6.3.2.2 Exact Solution

The BS and reliability diagram provide useful tools for comparing probabilistic forecasts, but each suffers from similar drawbacks. Three specific problems with these approaches are identified here. First, each method is approximate. While it was noted previously that the normal approximation to the binomial distribution, and therefore the BS, becomes accurate for large ensemble sizes, in practice there will rarely be enough

observations to make errors negligible. Prior to analysis with the Binomial distribution, this was a similar symptom of the reliability diagram, as it directly analyzes residuals of the forecasted and observed distribution. A second drawback is approaching the problem from the Binomial distribution is limiting. It becomes a balance between having sufficiently small forecast bin variance to reduce errors, and enough observations in each bin to draw a meaningful conclusion. A final drawback is that these metrics are ambiguous. Rather than distinguishing between reliable and unreliable forecasts, the BS and reliability diagram only estimate the probability of reliability. It is more desirable to select all forecasts that may be deemed reliable, and then compare the reliable forecasts solely on their sharpness. In order to overcome these problems, an exact model for reliability is required, which can distinguish between reliable and unreliable forecasts, and does not require binning.

An exact model of probabilistic event forecast verification may be achieved through the generalized form of the Binomial distribution, where probabilities are allowed to vary. This generalized distribution is referred to as the Poisson-Binomial distribution (Hodges and Le Cam, 1960), and is presented in equation (53). In equation (53), S_k is a function identifying all possible combinations of the k droughts occurring throughout the T forecasts (S_k contains $\binom{T}{k}$ combinations), and A signifies all drought occurrences in combination S_k and A^c represents all non-drought occurrences from S_k . From equation (53), the entire string of drought probabilities is used to estimate the sum of the probabilities of each combination of drought events occurring, up to the observed

number of droughts. In the event that all forecasted probabilities are equal, this equation will collapse to the Binomial CDF in equation (52). Equation (53) therefore provides a mathematically exact model of reliability assessment in the probabilistic drought forecast setting. By utilizing the Poisson-Binomial distribution, the probability that a forecast is reliable can be estimated exactly, and without grouping observations. While this method is exact, and avoids the requisite binning process, it does not directly provide a means for distinguishing between reliable and unreliable forecasts.

$$F(k \leq K) = \sum_{k=0}^K \left[\sum_{A \in \mathcal{S}_k} \left(\prod_{t \in A} p_t \prod_{t \in A^c} (1 - p_t) \right) \right] \quad (53)$$

This dissertation follows the method of formal hypothesis testing for distinguishing between reliable and unreliable forecasts, where an attempt is made to reject the null hypothesis that a forecast is reliable. Commonly this takes the form of assuming some significance interval. Given that a cumulative probability falls outside some predefined significance interval (95% in this study), then the hypothesis of reliability will be rejected. More specifically, if $0.025 < F(k \leq K) < 0.975$, the hypothesis of reliability is not rejected. In the event that the hypothesis of reliability is not rejected, the forecast and observation distribution are deemed indistinguishable.

Use of the Poisson-Binomial distribution has one primary drawback. Although the Poisson-Binomial probability density function is computationally tractable, direct estimation of the CDF is computationally infeasible for any useful sample size. In order to overcome this issue, it is possible to use the Discrete Fourier Transform and the Characteristic Function, as demonstrated by Hong (2013), to solve the CDF at any

practically relevant sample size. With the ability to solve the Poisson-Binomial CDF, an exact model of the probabilistic forecast setting is available, and thus the hypothesis of reliability may be properly tested.

6.3.2.3 Sharpness

The probabilistic verification metrics examined to this point have been focused on assessing reliability, but in the event that multiple forecasts are deemed reliable, it becomes necessary to compare the forecasts based on sharpness. Two methods are used in this manuscript for examining the sharpness of forecasts, both of which rely on the understanding that a sharp drought forecast will tend towards probabilities of 0 or 1. First, a histogram of the forecast probabilities examines the distribution of probabilities ($p_{1:T}$), which allows for visual assessment of sharpness as the histogram becomes increasingly U-shaped. Second, the variance of the Poisson-Binomial distribution, as shown in equation (54), is used as a quantitative measure. Since the variance indicates that dispersion in the forecast probability density, it is therefore a measure of the certainty of a forecast. In order to maximize sharpness of a forecast, the variance must be minimized, and therefore the perfect forecast will fall within the significance interval of the Poisson-Binomial distribution, and have a variance of 0.

$$\sigma^2 = \sum_{t=1}^T p_t(1-p_t) \quad (54)$$

7 Experimental Setup

7.1 Water Supply Forecasting

In the water supply forecasting experiment presented in this dissertation, seasonal ensemble forecasts are examined from seven different modeling scenarios. The first two forecast experiments are the standard ESP method, referred to as “open loop”. The first open loop experiment is performed with the NWS models, and the second open loop experiment is performed with the VIC model. Each open loop experiment performs a spin up to an initial forecast date, and then 29 separate forcing years are resampled between 1981 and 2010 to represent climatology (the forecast year is excluded leaving only 29 of the 30 years used). Four more experiments are performed with DA as a spin up. Each model (VIC and NWS) has a spin up with TB only and TB/LST DA. Each DA experiment uses 100 ensemble members, leading to 100 initial condition estimates at each forecast time step, from October 1st, 2002 through September 30th, 2008. With 29 resampled forcing time series’ and 100 initial condition ensemble members, 2900 forecast traces are possible. In the VIC model, it is infeasible to perform all 2900 combinations of initial condition ensemble members and historical forcing years due to computational constraints, and therefore 500 combinations were sampled uniformly from those 2900. Computational demand in the NWS models is significantly lower, and therefore all 2900 combinations are performed. The final experiment is a model averaging experiment, based on PF-SBC, of all 6 prior modeling scenarios. All traces are weighted according to the equations in Chapter 5.5 and Figure 2, and the weighted ensemble is evaluated. Within the PF-SBC algorithm, all streamflow observations are

utilized for model weighting, up to the initial forecast date, which makes the minimum length of training data 3 months, and this is deemed reasonable based on the findings in Parrish et al., (2012). Each forecasting experiment estimates three month volumetric flow from start dates on the 1st and 15th of January through June, in the years 2003 through 2008. This leads to a total of 72 seasonal forecasts from each modeling scenario.

7.2 Probabilistic Verification of Binary Outcomes

A synthetic probabilistic forecasting experiment is performed to demonstrate errors in conventional verification methods. In this example, hypothetical probabilistic forecasts are sampled from a uniform distribution to examine the idealized case (Case 1), as shown in equation (55). In the event that a distribution of probabilistic forecasts deviates from uniform, as will be typical of real forecasts, errors are expected to increase in the BS and reliability diagram. Two alternatives to Case 1 are created to examine the extent of these errors. Case 2 is a skewed distribution, generated from equation (56), and Case 3 is a “U” shaped distribution, with a tendency to forecast towards 0 or 1 (increased sharpness), according to equation (57). J is set to 500 in this experiment.

$$p_{j,1} \sim U(0,1) \quad (55)$$

$$p_{j,2} = p_{j,1}^2 \quad (56)$$

$$p_{j,3} = \begin{cases} p_{j,2} & \text{if } j < \frac{J}{2} \\ 1 - p_{j,2} & \text{otherwise} \end{cases} \quad (57)$$

Observations for each forecast are sampled with a probability equal to the forecasted probability, according to equation (58). This provides randomly distributed

events, which are statistically indistinguishable from the forecasted probabilities, thus ensuring a reliable forecast. Results from these synthetic experiments are presented in section 8.2.

$$O_j = \begin{cases} 1 & \text{if } U(0,1) \leq p_j \\ 0 & \text{otherwise} \end{cases} \quad (58)$$

7.3 Drought Forecasting

Based on the streamflow forecasts generated in the water supply forecasting experiment, drought status is forecasted for up to three month lead times, starting on the 1st and 15th of January through June, for years 2003 through 2008, for 16 sub-basins of the UCRB. This provides a drought forecasting experiment with 1152 individual drought forecasts. From each of these forecasts, the drought status in the 1st, 2nd and 3rd month following the initial forecast date is estimated. From these forecasts, both deterministic and probabilistic forecast verification are performed to examine the utility of data assimilation and model averaging in drought forecasting.

7.4 Assessing the Need for a New Drought Index

Critical examination of drought indices is necessary to determine the drawbacks of standardization, and to suggest ways to move forward. Of the three problems of standardized indices highlighted in chapter 3, this dissertation focuses on the effects that standardization has on spatial drought analysis. Although issues of non-stationarity are equally as important as spatial errors for drought analysis, examination of the effects of spatial errors is much more quantifiable, and therefore is the focus here. In order to examine these spatial errors, this study will examine spatially averaged drought time

series' from different drought indices, and compare them spatially at specific dates of interest. Further, the importance of the physically based index for drought quantification within the context of advanced land surface modeling techniques (i.e. DA) will be discussed.

7.5 Sensitivity of Droughts to Initial Conditions

7.5.1 Simulations

Two separate ensemble LWS estimation experiments are performed in this study: 1) resampling of states from a 30-year historical VIC simulation, from April 1st 1981 through June 30th 2011, referred to as climatology, and 2) forecasting ESP initialized with states estimated from the VIC model with TB/LST data assimilation, referred to as the forecast (DeChant and Moradkhani, 2011b). With respect to climatology, the LWS values are sampled from each year of the historical simulation, providing an ensemble of 30 LWS values at each time step over the forecast lead time, as shown in equation (59), where LWS_t^c represents the climatological distribution of LWS values at time t .

$$LWS_t^c = [LWS_{t,1}^c \quad LWS_{t,2}^c \quad \dots \quad LWS_{t,30}^c] \quad (59)$$

This represents the forecast of greatest possible uncertainty, as it does not utilize information about the initial conditions or meteorological forcing beyond the historical record. Alternatively, the forecast samples initial land surface states from the posterior distribution estimated with data assimilation at the forecast start date, and then performs simulations from those states with meteorological forcing sampled from the same 30 year dataset used to simulate climatology. 500 combinations of the initial conditions and

meteorological forcing time series' were sampled randomly, and a forecast was generated from each of these combinations, thus creating an ensemble forecast of 500 members, as shown in Figure (60).

$$LWS_t^f = [LWS_{t,1}^f \quad LWS_{t,2}^f \quad \dots \quad LWS_{t,500}^f] \quad (60)$$

The forecast has increased information beyond climatology resulting from the land surface state initialization. As the forecast progresses in time, the information added to the forecast from these initial states will reduce over time, leading to the forecast approaching the climatological distribution, and therefore moving away from the initial drought conditions.

The climatology and forecast estimate the LWS over a lead time of 360 days. This provides an extended period which is assumed to be of sufficient lead time for forecasts to approach climatology, thus becoming insensitive to the initial drought conditions. These forecasts are performed from April 1st, to correspond with the date of peak snow water storage, and from July 1st, to correspond roughly to the date of peak soil moisture and minimal snow influence, for each year from 2003 through 2008. By forecasting for a 360 day period, initialized with states on April 1st and July 1st, the time and rate of recovery from drought conditions throughout the UCRB will be quantified for both snow dominated and non-snow dominated seasons. For the analysis here, the LWS is averaged in 10 day increments, which is performed to smooth out the daily noise in the LWS variable, leading to more consistent results. The 10 day increment was chosen based on a comparison of increments ranging from 1 to 30 days, with 10 being a balance of daily noise reduction and retaining sufficient temporal resolution.

7.5.2 Quantifying Drought Recovery Lead Time

The drought recovery study works under the assumption that the basin has fully recovered from a drought at the lead time when the forecast and climatology ensembles become statistically indistinguishable, and therefore may be assumed to be identical. Given that the forecasts and climatology are statistically indistinguishable after a specific time, then one can conclude that the uncertainty in the forecast is entirely a result of the forcing at future times. If the uncertainty is entirely a result of the forcing at any time in the forecast, then the initial drought conditions are completely alleviated, as they no longer affect the hydrologic conditions in the basin. Note that both the climatology and forecasts are based on the same models, and therefore the effects of model error will be negligible. Such an analysis requires a hypothesis testing framework to determine if the forecast ensemble is significantly different from climatology at each 10 day period over the forecast lead time. In this study, a two-sample Kolmogorov-Smirnov (KS) test is used to compare the two distributions (Wolfe and Myles, 1973), with an attempt to reject the null hypothesis that the forecast distribution is equivalent to climatology. Given that the analysis is unable to reject the null hypothesis, it will be assumed that the two distributions are equivalent, and therefore the forecast has entirely recovered from the drought conditions experienced at the initial forecast date.

Hypothesis testing with the KS test relies on the construction of an empirical CDF of two distributions, which is described for the climatological ensemble in equation (61).

$$F_{t,B}^c = \sum_{b=1}^B \sum_{i=1}^N \begin{cases} 1 & LWS_{t,i}^c \leq LWS_{t,b} \\ 0 & otherwise \end{cases} \quad (61)$$

In equation (61), the $F_{t,B}^c$ is the empirical CDF of the climatology at time t , which is estimated over B histogram bins distributed uniformly between the maximum and minimum value from the concatenated LWS_t^c and LWS_t^f arrays, N is the number of ensemble members (30 for climatology), and $LWS_{t,b}^c$ is the maximum LWS value for bin b at time t . Note that this is performed similarly for the forecast ensemble, which will use the notation $F_{t,B}^f$. For the construction of both CDFs, B is set equal to 530 as it is the total number of data points in the concatenated LWS_t^c and LWS_t^f arrays.

After construction of the CDFs, the KS test examines the maximum absolute difference of the forecast and climatology CDFs across all bins, which then must be multiplied by the square root of the ratio of the product and sum of the forecast ($M=500$) and climatology ($N=30$) sample sizes, as shown in equation (62).

$$D_t = \sqrt{\frac{MN}{M+N}} \max(|F_{t,1:B}^f - F_{t,1:B}^c|) \quad (62)$$

From this equation, the two-sample KS statistic (D_t) is estimated, which may be used for testing the null hypothesis. In this application, if D_t is greater than 1.36, then the null hypothesis is rejected (with 95% confidence), and the forecast and climatological ensembles are considered different. At any time step in which the null hypothesis is not rejected, we can assume that the land surface conditions are not significantly different than climatology, thus showing that the basin is no longer experiencing drought conditions and completely insensitive to the initial drought conditions.

7.5.3 Quantifying Drought Recovery Rate

The drought recovery study also seeks to estimate the rate of drought recovery. In order to estimate this rate, information theory provides a useful framework for examining the information provided by the initial drought states. Within information theory, entropy is a basic idea for examining the amount of information a probability distribution contains in respect to some random variable, which is typically estimated with Shannon Entropy (Shannon, 1948) (equation (63)).

$$H(LWS_t^c) = \sum_{j=1}^J p(LWS_{t,j}^c) \log(p(LWS_{t,j}^c)) \quad (63)$$

In equation (63), J is the number of LWS values at which the probabilities are estimated, $p(LWS_{t,j}^c)$ is the probability of the LWS, according to the climatological ensemble, evaluated at time t and value j , $\log(p(LWS_{t,j}^c))$ is the natural logarithm of that probability, and $H(LWS_t^c)$ is the entropy at time t . From this definition, entropy is inversely proportional to information content of a given probability density. Although the probabilities in equation (63) could be estimated from empirical distributions shown in equation (61), the small ensemble sizes will likely lead to biased entropy estimates (Miller, 1955), necessitating interpolation to regions of the LWS space unrepresented by the ensemble members. To ensure minimal bias, both the climatology and forecast ensembles are fit with a Kernel Smoothing Density to estimate the probability between ensemble members, as shown in equation (64) and (65).

$$KS(LWS_{t,j}^c) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{LWS_{t,i}^c - LWS_j}{h}\right) \quad (64)$$

$$p(LWS_{t,j}^c) = \frac{KS(LWS_{t,j}^c)}{\sum_{j=1}^J KS(LWS_{t,j}^c)} \quad (65)$$

In equation (64), N is the ensemble size, LWS_j is the value at which the probability is being estimated, h is the smoothing parameter, K is the kernel, which is chosen here as Gaussian, and $KS(LWS_{t,j}^c)$ is the corresponding Kernel Smoothing Density estimate at time t and value j . This density is then normalized according to equation (65) to estimate the probability at that location and time. The probability is estimated at 1,000 values (J), uniformly distributed between the minimum and maximum values of the concatenated LWS_t^c and LWS_t^f arrays. The smoothing parameter is estimated from equation (66), which is the optimal value for a Gaussian kernel, where σ is the standard deviation of LWS_t^c . This operation may be performed similarly for the forecast ensemble to estimate the information contained in the forecast.

$$h = \left(\frac{4\sigma^5}{3N} \right)^{1/5} \quad (66)$$

Beyond simply quantifying the amount of information in the climatology and forecast ensembles, the goal here is to estimate the amount of information that the forecast contains in relation to climatology. An important note is that these two ensembles quantify the same variable (LWS). Therefore, this study seeks to quantify the relative information loss of the forecast in comparison to climatology, which will therefore estimate the rate at which the forecast loses information extracted from the initial drought states, referred to here as the drought recovery rate. Within this scenario,

the climatology has no information from initial drought status, but contains the same information with respect to meteorological forcing and model structure, and therefore contains the minimum information content that the forecast will achieve. As noted in equation (67), the entropy of the forecast distribution will be less than climatology at all times, except at the time when the forecast and climatology become identical. Note again that entropy is inversely proportional to information content.

$$H(LWS_t^c) \geq H(LWS_t^f) \quad (67)$$

Since the entropy of the climatology will always be greater than or equal to the forecast, and the initial drought state is the only additional information source the forecast contains beyond the climatology, the fraction of the information extracted from the initial drought state by the forecast is equivalent to the ratio of the forecast and climatology entropies. This ratio is referred to as the relative entropy (RE), and is shown in equation (68). The RE (RE_t) is ranges between 0 and 1, with 0 indicating all information in the forecast is derived from the initial drought conditions, and 1 indicating no information in the forecast is derived from the initial drought conditions. Therefore this study will estimate rate of change of the relative entropy, which requires a function to be fit to the estimated RE data points.

$$RE_t = \frac{H(LWS_t^f)}{H(LWS_t^c)} \quad (68)$$

Estimation of drought recovery rate is made difficult by the non-linear nature of the RE metric. Due to this non-linearity, and the upper limit of 1 for the RE value, this study will quantify the drought recovery rate by fitting an exponential function to the time

series' of RE. A simple yet effective function is equation (69), where the least squares fit is deemed sufficient ($\min(\sum (RE(X_t) - f_{RE}(X_t))^2)$).

$$f_{RE}(X_t) = 1 - \frac{1}{X_t^n} \quad (69)$$

This function asymptotically approaches 1 for all positive n values, with increasing values indicating a faster rate of drought recovery. Throughout the analysis of the results, the exponent in equation (69) will be used to quantify drought recovery rate. Equation (69) was chosen over other functions, specifically logarithmic or polynomial functions, because it produced the lowest squared error of the alternatives examined.

8 Results and Discussion

8.1 Water Supply Forecasting

A first assessment of the results from the water supply forecasting experiment is provided with the weights of each model in Figure 3. This figure shows the average model weights from each forecast month, and the overall average weights, as estimated through the PF-SBC algorithm. In general, the NWS models appear to have the highest weights, indicating a higher accuracy in monthly probabilistic streamflow volume estimation, and more reliable estimation of streamflow uncertainty, than the VIC model. However, the VIC model cases were assigned high enough weights to provide a significant contribution in the model averaging framework. In comparing the DA and open loop cases, both DA cases in the NWS models received higher weight than the open loop model, yet the VIC model with DA received an overall lower weight than the open loop case. A general expectation is for the VIC model, which explicitly solves the energy balance to be more effective in estimating inputs to the radiative transfer model and LST, yet the application here does not support this hypothesis. Although the results suggest VIC is less effective for TB and LST DA than the NWS models, this is not a conclusion that should be drawn here due to the differing spatial discretization and calibration schemes. Another important note is the difference in performance of the two DA cases for each model. In the VIC model, the case of combined LST and TB DA outperforms the TB only DA case, but the NWS models perform best without the DA of LST. This is not surprising, as the VIC model explicitly solves for LST, whereas the SNOW-17 model only estimates average pack temperature, thus the VIC model should more accurately

estimate LST. Figure 3 also highlights an important temporal aspect to the performance of these modeling cases. In the DA cases, weights tend to be highest during the snow accumulation season (January, February, and March) and have lower weights into the ablation season (May, June). The reverse is true for the open loop cases. This finding is not surprising, due to the poorer sensitivity of microwave TB to SWE in deeper snowpacks with high liquid water content (DeChant and Moradkhani, 2011a). During the ablation season, the liquid water content in the snowpack remains high, reducing the ability of TB DA to accurately reconstruct SWE. Such temporal variation in DA performance necessitates the use of the PF-SBC methodology, as opposed to batch framework, to allow for dynamic weights, thus providing more accurate weights at each forecast initial condition.

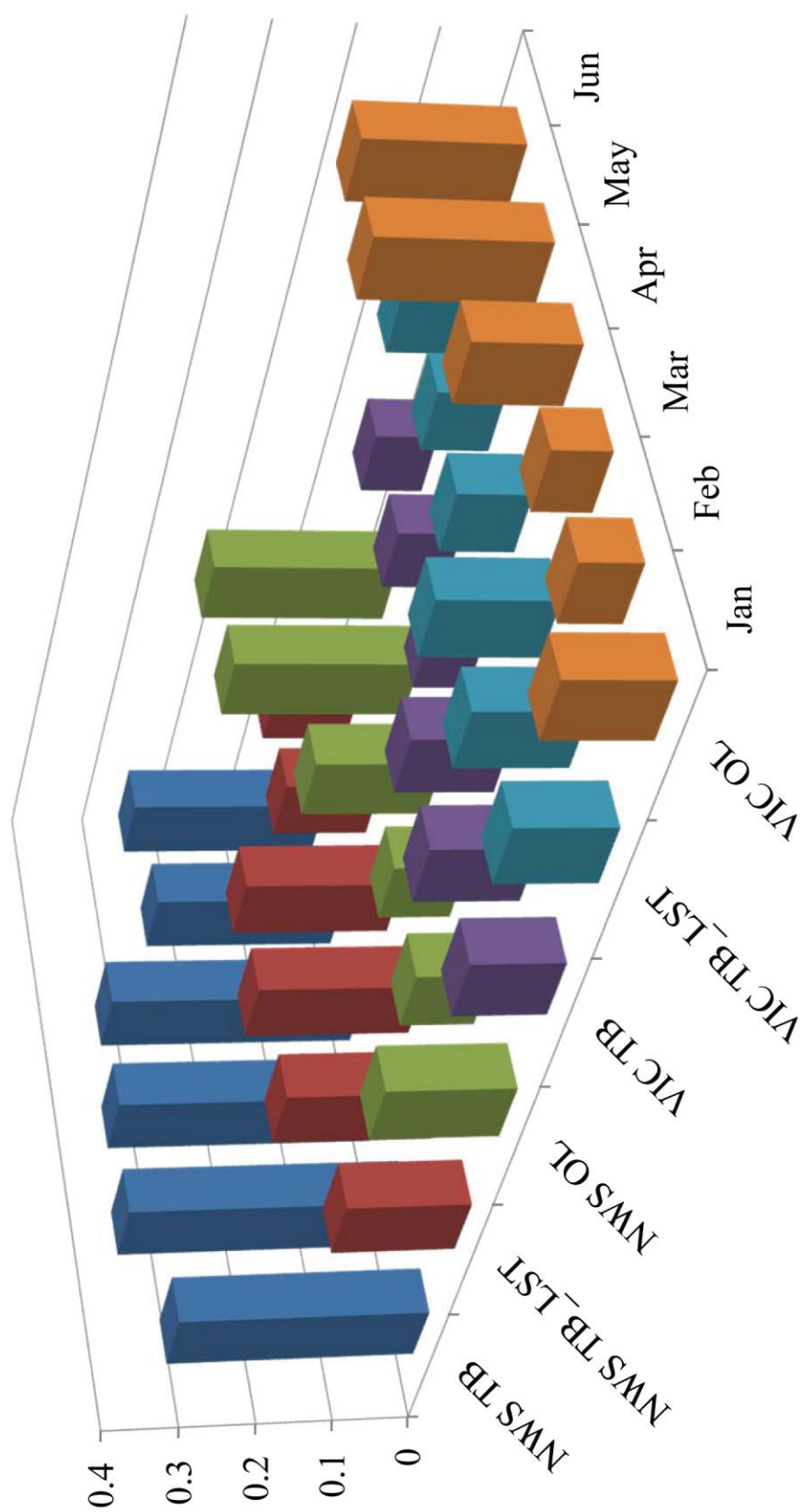


Figure 3. 3-dimensional bar plot of the average weights for each modeling scenario and forecast month, as estimated during the spin-up period.

During the forecast phase, the ability of each modeling case to reliably estimate seasonal runoff volumes is assessed. From a risk management perspective, it is important to examine the accuracy of the tails of the forecast distributions, which may be used as maximum or minimum expected flows. In order to assess the tails of the forecast distributions, the 99%, 95% and 90% predictive bounds from each method are examined, with their respective ERs, in Figure 4. This figure suggests that every modeling scenario is overconfident at every predictive bound (each predictive bound is exceeded at a frequency higher than optimal). At the 99% predictive bound, each DA case reduces the overconfidence, highlighting the importance of initial condition uncertainty, and PF-SBC further reduces the ER to about 3%. This still remains slightly overconfident, yet is a significant improvement over the other modeling cases (the best ESP with DA case had 14.9% ER), thus highlighting the importance of accounting for model errors. At the 95% and 90% ERs, DA continues to improve overconfidence, with the exception of VIC with TB only DA, and PF-SBC again performs best overall. These results suggest that across the entire basin, both ESP with DA and ESP with PF-SBC improve the reliability of the tails of the forecast distributions, but provide no assessment of the performance spatially. In order to examine the performance of the methods across several sub-basins, Figure 5 shows the 99% ER of 16 different sub-basins within the UCRB.

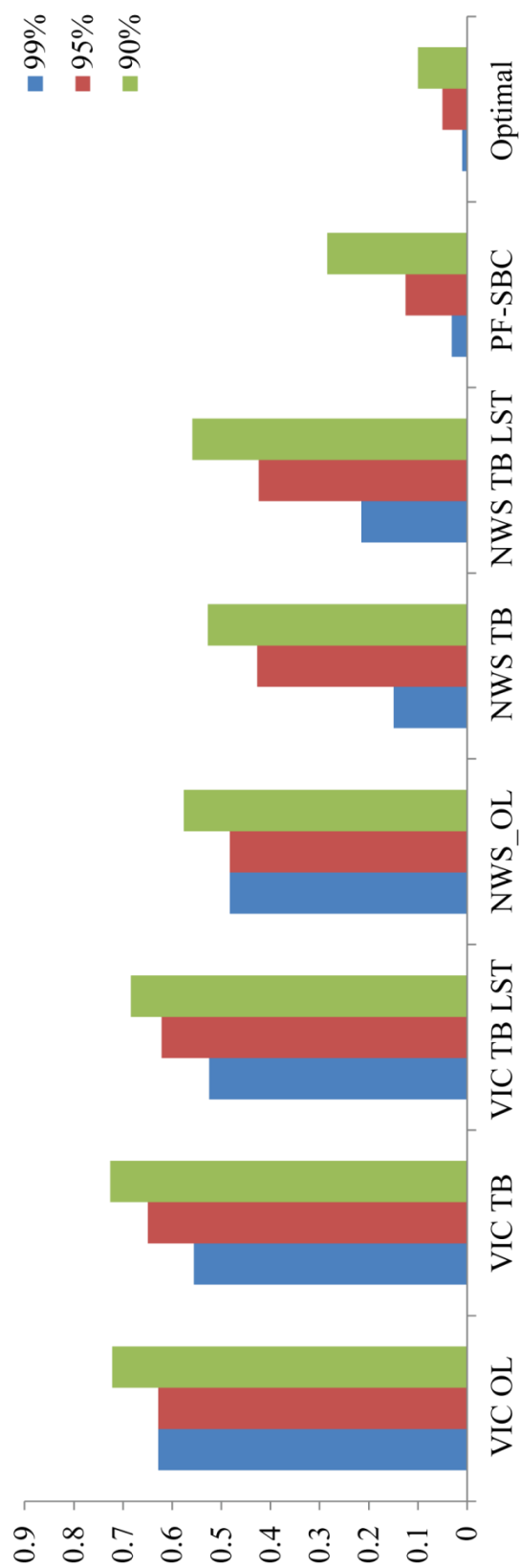


Figure 4. Histogram of the 99%, 95% and 90% exceedance ratios of each modeling scenario, for all seasonal forecasting experiments. The optimal exceedance ratio is on the far right of this figure.

The accuracy of the 99% predictive bounds varies spatially throughout the UCRB. In the VIC open loop model, the San Juan is the only basin that has an ER of lower than 50% (optimal is 1%), which shows that the probabilistic forecasts does not convey the actual magnitude of uncertainty. In the DA cases, a reduction in the 99% ER is observed in every sub-basin outside of the Colorado River headwater region. In this region, DA appears to struggle in improving the accuracy of initial conditions, which is an observation that will be discussed further in relation to later results. The NWS models perform much more consistently throughout the UCRB, with ERs generally around 50%, and greater variability in the Rocky Mountains. In the NWS DA cases, the ER is reduced in all basins, suggesting generally more reliable forecasting of low probability events. DA in these models appears to be more consistently effective than in VIC, which is attributed to the differing spatial discretization, as mentioned in respect to model weights from Figure 3. Finally, PF-SBC further reduces the ER, from which values are consistently below 10%, with the exception of the Dolores River Basin (southwestern sub-basin of the Colorado River headwater region) at around 20%. Overall this shows that DA tends to improve probabilistic prediction, highlighting the importance of initial condition errors, and further improvements are made through PF-SBC, showing the importance of the model error component. Though the results here are encouraging, it is important to also look at the overall reliability of the forecast distributions to ensure that reliability of the entire forecast distribution is improved. This is performed with the *R* metric and is shown spatially in Figure 6.

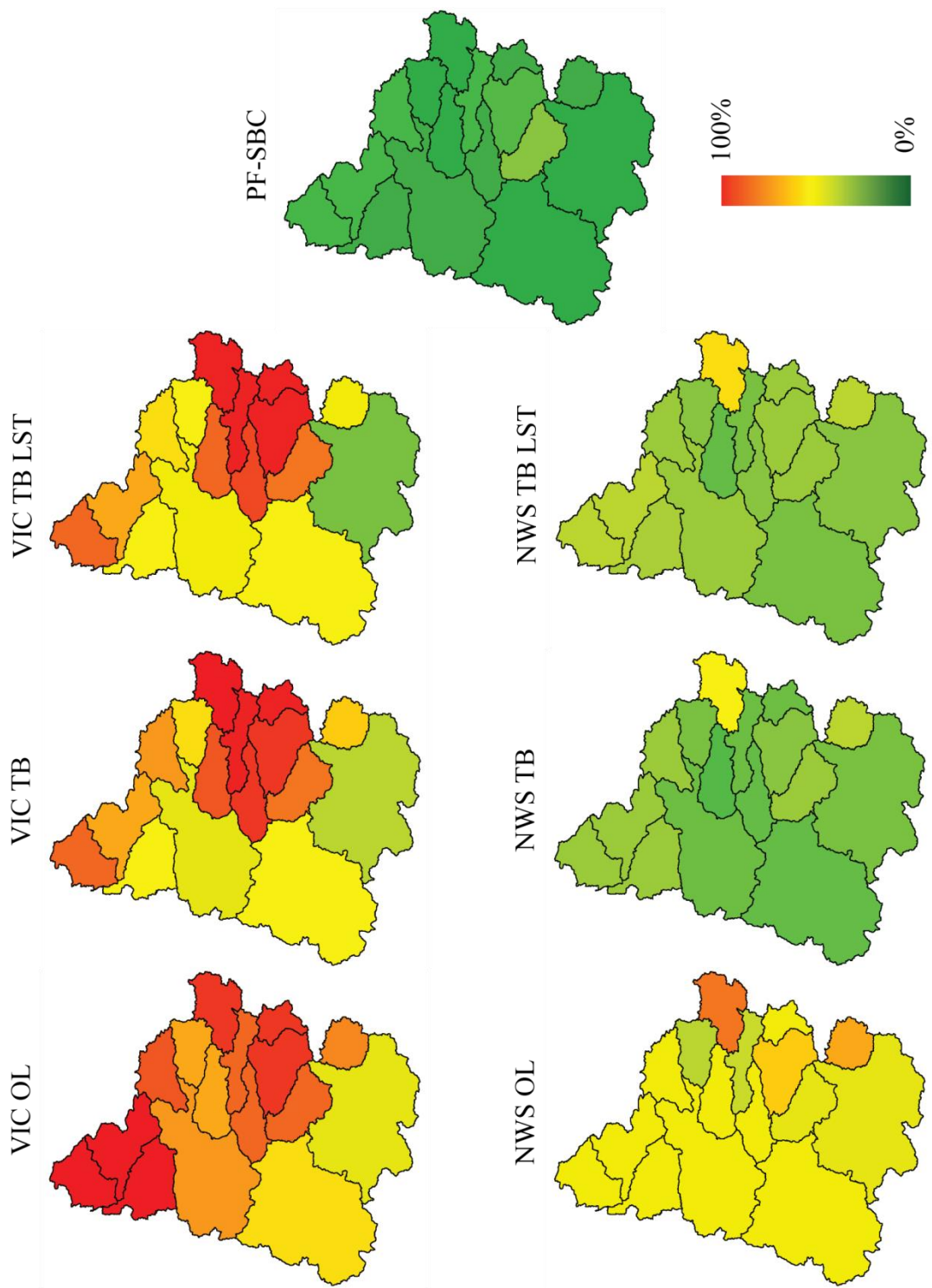


Figure 5. 99% exceedance ratio over 16 sub-basins within the upper Colorado River Basin, for all seasonal forecasts.

In the VIC open loop case, reliability is low in the Green River Basin and in the Colorado Headwaters, with better performance in the San Juan and at Lee's Ferry, which suggests a more accurate reconstruction of initial states in the San Juan than the other sub-basins. In the two DA cases, reliability is improved in both the Green and San Juan basins, but the performance is worse in the headwater basins, similar to the results of the 99% ER. Both DA cases improved the initial condition distribution in the Green and San Juan basins, yet struggled in the Colorado River headwaters. In contrast to the VIC model, the NWS models have less variation in reliability, and no identifiable pattern in performance. Also, the DA cases only show improvements in the San Juan River basin, with slightly worse performance in many other basins. This observation is unexpected in light of the improvements shown in figures 4 and 5. Such an observation suggests that the ESP with DA method is unable to improve the central portions of the predictive distribution. Despite an improvement in the tails for the forecast distributions from the DA cases, the mode of the distribution is not shifted enough to observe general improvements in reliability. Results from the PF-SBC experiment show that the model averaging scenario is superior to all other cases in the majority of basins, though exceptions are present in the upper Green river basin and central Gunnison river basin (south-central portion of the Colorado River headwaters). The superior results from PF-SBC over the entire forecast distribution, in conjunction with unimproved total reliability from DA, suggests that model errors are persistent in the central portion of the forecast distribution. While the initial condition errors were important in estimating the low

probability events, it is essential to manage model error for accurate prediction of the mode of the forecast distribution.

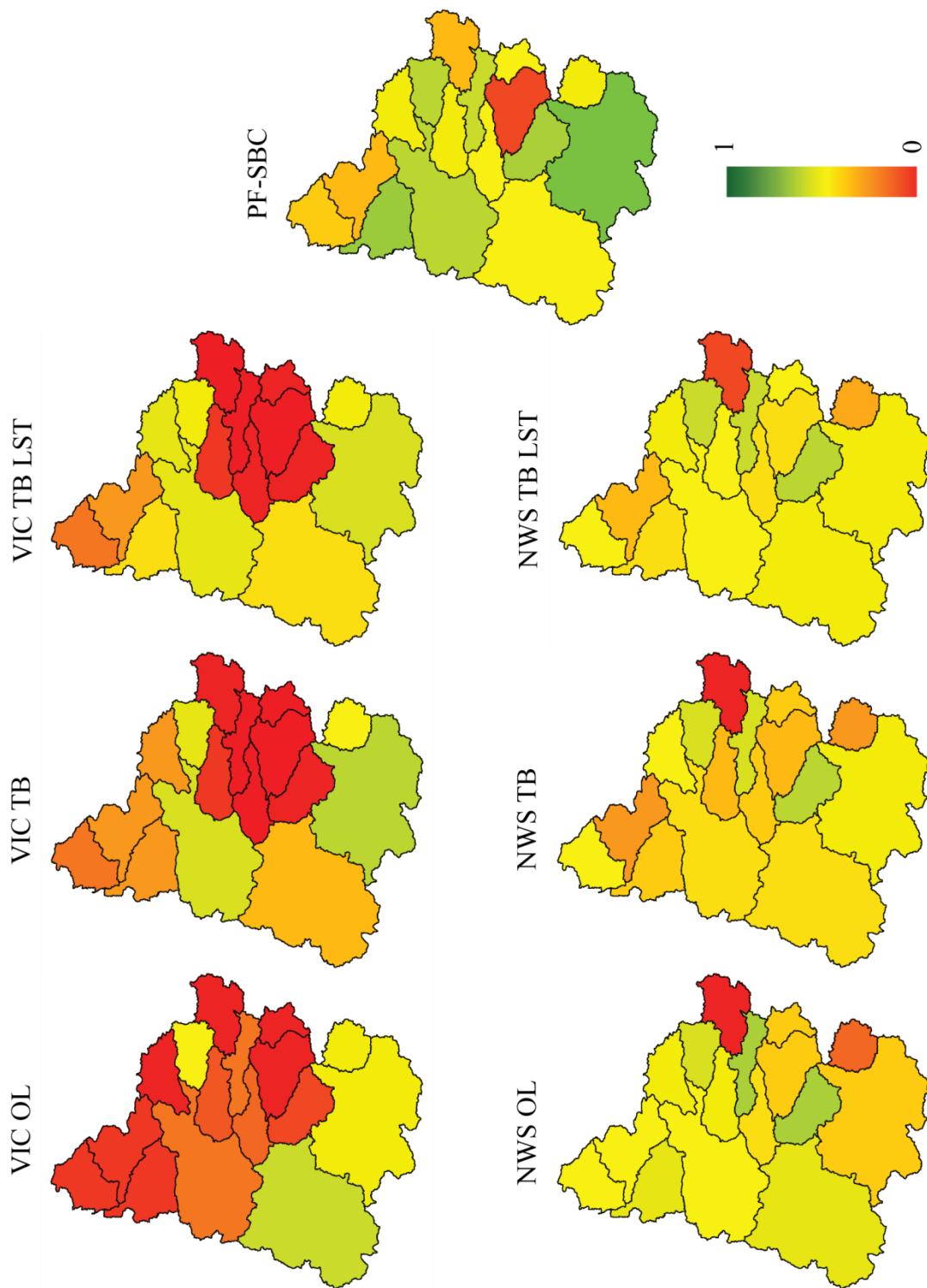


Figure 6. Reliability over 16 sub-basins within the upper Colorado River Basin, for all seasonal forecasts.

The conflicting results regionally, and within different portions of the forecast distributions, requires further examination of the regional forecast distributions. The causes of conflicting information from previous results are diagnosed with the predictive QQ plot from each of the four gauges shown in Figure 1, which are provided in Figure 7. In the upper left plot of Figure 7, it appears that both the NWS models and the VIC model have significant biases, with the NWS models showing a generally high bias (observations primarily falling in the lower portions of the forecast distributions) and the VIC model having a generally low bias (observations primarily falling in the higher portions of the forecast distributions). In the NWS models, DA does not change the bias, though it does move the tails of the distribution to encompass more observations, but the VIC model has increasingly low bias with DA. These differing results highlight the effects of spatial resolution on TB DA in regions of highly complex topography and thick vegetation. From Figure 1, it is apparent that the Colorado River headwaters is the most topographically rough sub-basin, and has the densest forest cover of all regions, causing greater problems for DA in the VIC model. Since the NWS models are distributed based on topography, the basins in this region are much smaller than in the VIC model, and thus the spatial heterogeneity is more effectively modeled. Conversely, in the Green River, the bias in the VIC model was reduced through DA, but again only little change was observed in the NWS bias in the DA case. In the San Juan, which has the lowest density of forest cover and driest climate, the bias in both models is reduced, showing the ability of TB DA to reduce SWE errors in regions of thin vegetation and dry climates. With respect to the PF-SBC case, the forecasts in the headwaters have a high bias, generally

following the NWS model forecasts, but in the Green and San Juan basins, PF-SBC provides a significant reduction in model bias. In basins where both models provide sufficiently accurate forecasts, PF-SBC is capable of effectively leveraging information from multiple model structures to improve forecasts, but the poor forecasts from VIC in the headwaters forces PF-SBC to rely primarily on the NWS modeling cases. Results at Lee's Ferry indicate the strong influence of the Colorado River headwaters in flow at the outlet (on average it provides 45% of the runoff at Lee's Ferry), which is unfortunate because forecasts for the headwaters are the least skillful, and show the least sensitivity to the remotely sensed observations. Overall it appears that the tails of the forecast distributions at Lee's Ferry are most accurate when using the ESP combined with PF-SBC, but this framework is unable to provide much improvement in the mode of the forecast distributions outside of the San Juan and Green River basins.

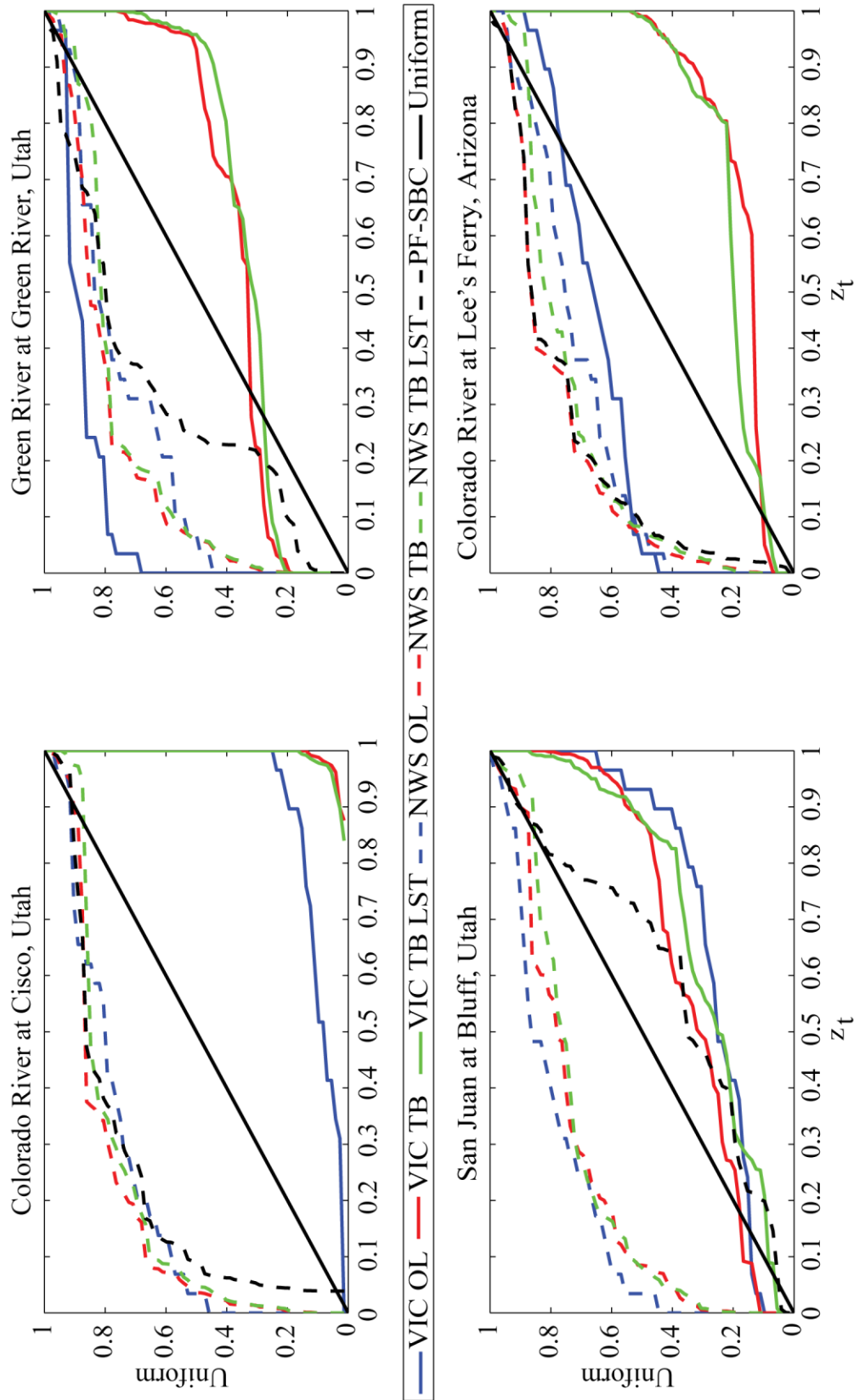


Figure 7. Predictive QQ plot of the seasonal forecasts from the four major sub-basins of the upper Colorado River Basin.

8.2 Probabilistic Verification of Binary Outcomes

The results presented in this section are intended to highlight flaws in the typical reliability assessment of probabilistic hydrological event forecasts. With this purpose, Figure 8 is presented to show how the decomposed BS and reliability diagram assess the reliability of the synthetic forecast cases. In Figure 8, the left plot shows the distribution of reliability values, estimated with the decomposed BS, from 100 replicates of all three synthetic experiments described in Chapter 7.2, each with 10 bins. When comparing these distributions, it is clear that forecast uniformity affects the reliability estimation. Case 1 is shown to have a generally lower reliability score (0 is optimal) than the forecasts that deviate from uniform. Though use of 10 bins is quite common in the literature, the bin size is not small enough to eliminate the effects of forecast uniformity on diagnosing reliability, which may cause misinterpretations of this score. Case 1 may be observed as the best case from this analysis, yet it is the worst case given the knowledge that each is reliable (Case 1 is the least sharp). Alternatively, the reliability diagram (right subplot) shows that each forecast tends towards reliability, but the current diagram is not capable of making the distinction that all forecasts are reliable. Although Case 2 may be noted as questionable, as it has a large deviation within the bin centered at 0.25, all bins are entirely within the statistical variability of this forecasting scenario. In general, Figure 8 suggests that these metrics are useful tools for diagnosing forecast errors, but fall short of distinguishing between reliable and unreliable forecasts.

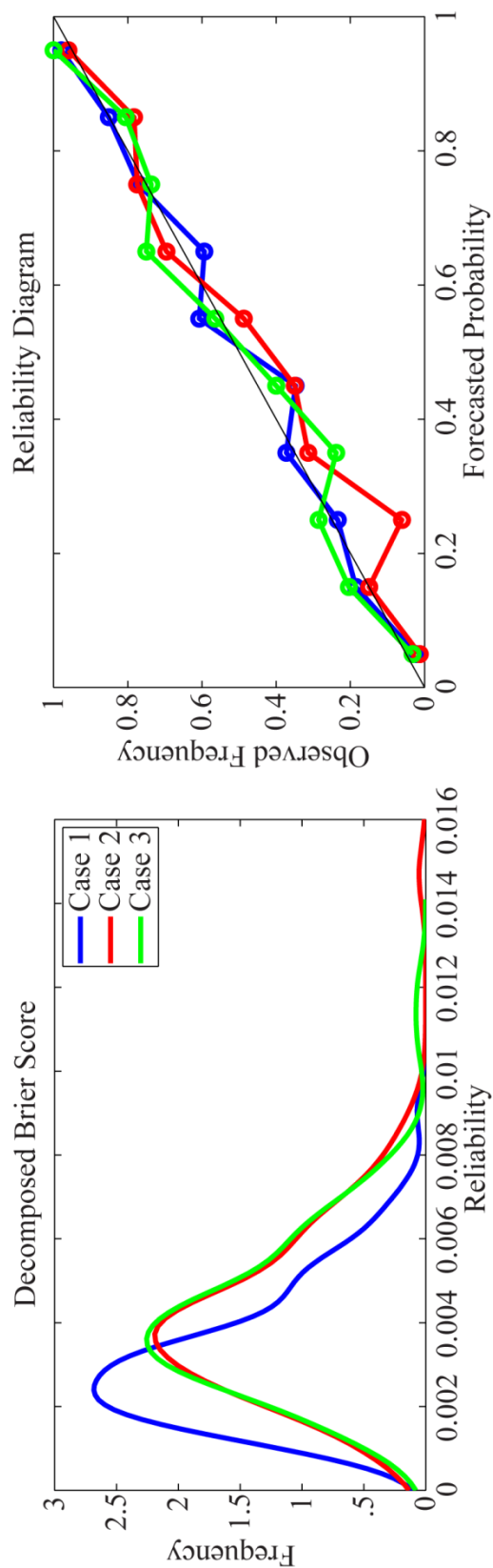


Figure 8. The left plot shows the distribution of reliability values (from 100 different forecast replicates) from the decomposed Brier Score for all three cases. The right plot is the Reliability Diagram for all three cases.

Errors associated with applying the Binomial (B-CDF) and Poisson Binomial (PB-CDF) CDFs for hypothesis testing are quantified through comparison of the significance intervals from each, as provided in Figure 9. In Figure 9, the 95% interval widths for a varying number of forecast bins are presented in the top row. In addition, the middle row shows the histogram of the forecast probabilities for each case, and the corresponding PDF of the approximate and exact solutions, for the case of a single bin, are presented in the bottom row. This figure clearly shows that the B-CDF is wider than the PB-CDF, with that difference being dependent on the uniformity of forecasted probabilities. A wider distribution suggests that the binomial approximation reduces one's ability to reject the hypothesis of reliability, thus increasing the possibility of type II errors. This error is largest in Case 3, which happens to be the sharpest case. Given that all forecasts are reliable, Case 3 would be optimal. In the event that Case 3 is unreliable, it is the most probable to be erroneously deemed reliable, increasing the likelihood of incorrectly selecting Case 3 as the best forecast.

A second observation from Figure 9 is the rapid growth of 95% range with the number of forecast bins. Since the grouping process reduces the sample size at each bin, the 95% significance interval is widened, causing an aggregate effect on the overall determination of reliability. By binning similarly valued forecasts, one vastly reduces the ability to distinguish between reliable and unreliable forecasts, further increasing the chance of Type II errors. This problem is especially concerning in the case of hydrological extremes (i.e. floods, droughts), which are inherently low probability events, making it essential to efficiently use information from every observation. Overall it is

important for forecasts to be verified with as few bins as possible, increasing the effective sample size, maximizing one's ability to reject unreliable forecasts.

A final observation from Figure 9 is that forecast uniformity affects error magnitude. It may be expected that errors in the B-CDF will decrease inversely with the number of forecast bins, as each bin becomes more representative of its members. This is evidenced in Case 1, where the B-CDF approaches the PB-CDF with decreasing bin size. Alternatively, the B-CDF in Case 2 and Case 3 has persistent error even with 10 bins. Rather than approaching the true value, the non-uniform cases display errors nearly independent of bin size. This result suggests that grouping may not significantly reduce the errors associated with the B-CDF. A much simpler and more effective solution is to utilize the PB-CDF in attempts to reject the hypothesis of reliability. In addition to this analysis of probabilistic verification measures, further analysis is provided in section 8.3, with the application of the proposed drought forecasting approach in the UCRB.

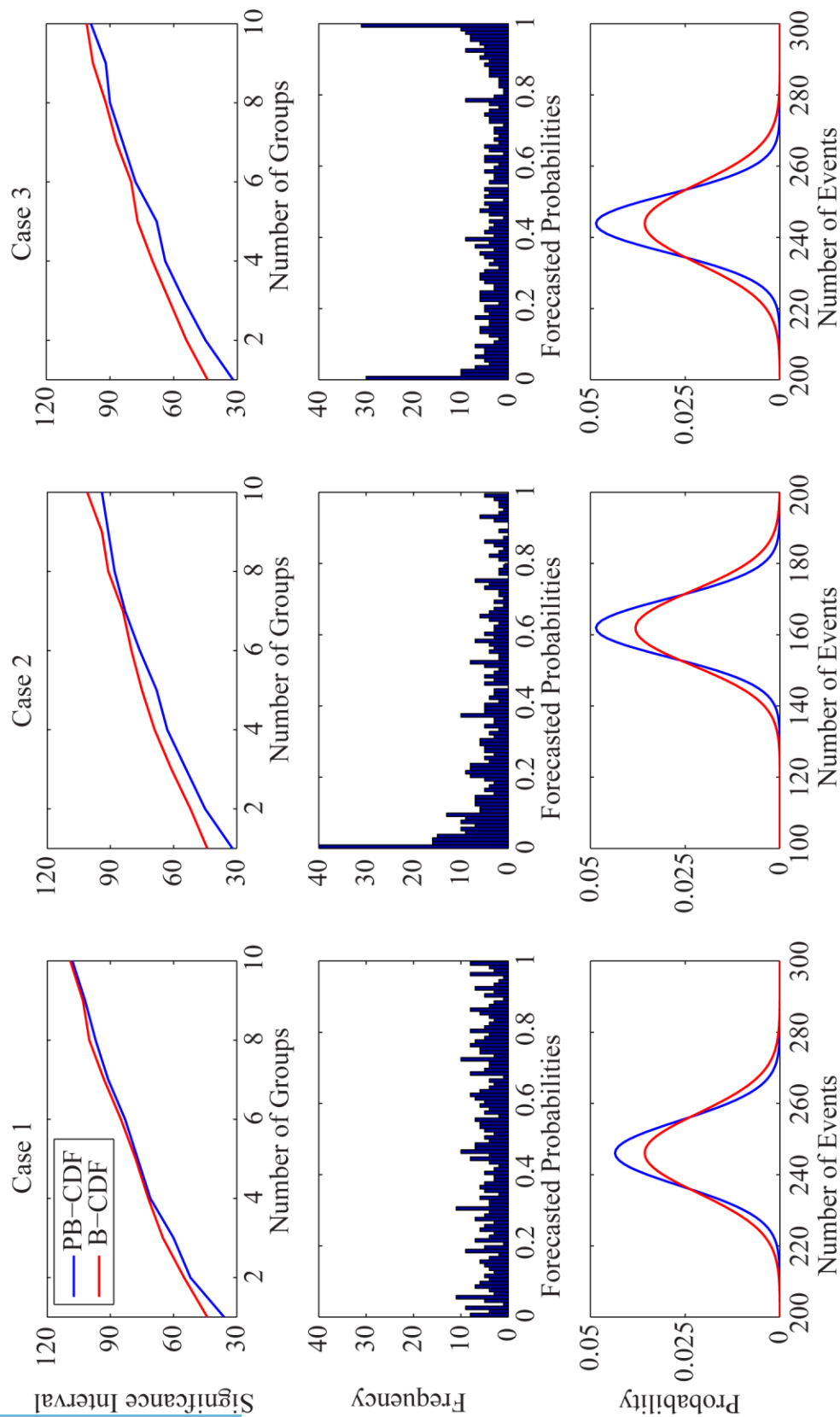


Figure 9. The top row shows the width of the 95% significance interval of the PB-CDF and B-CDF, with respect to the number of bins. The middle row is the histogram of the forecasted probabilities for each case. The bottom row is the probability distribution functions of the Poisson-Binomial and Binomial distributions.

8.3 Drought Forecasting

The examination of the proposed drought forecasting methodology begins with an analysis of the method's ability to deterministically forecast drought. As discussed in section 6.2.1, the POD, FAR and CSI are used in this study to examine the performance of the proposed methodologies for deterministic forecasting. These values are estimated for each forecast method at multiple lead times, as shown in Table 1. From this table, several observations are made. First, the VIC model based forecasts have highest POD and FAR values for every forecast scenario. This suggests that the implementation of the VIC model in this study has a generally low bias. In addition, the NWS models based forecasts have the lowest POD and FAR in all cases, suggesting a generally high bias. Both of these points are supported by results from the water supply forecasting experiment. In terms of overall comparison of forecast performance, the CSI is examined here because it is maximized as POD approaches 1 and the false alarm ratio approaches 0. In addition to Table 1, the CSI of each method and lead time is presented graphically in Figure 10. With respect to the CSI, the DA cases improve upon the open loop simulations in nearly every forecast, with the exception of the 1st month for the NWS models. Further, the combined assimilation of TB and LST leads to an improvement in over TB only assimilation in every model and forecast month. This contrasts slightly with results of the water supply forecasting experiment, as the NWS models were found to perform best with assimilation of TB only. LST assimilation appears to reduce the high bias of the case of TB only assimilation, which is found to be detrimental in water supply forecasting, yet beneficial in drought forecasting. A further finding is that the PF-SBC

case provides the most accurate drought forecasts of all cases, at all lead times. This suggests that model error is a significant factor in monthly to seasonal drought forecasting in the UCRB. By reducing the errors related to both initial conditions, via DA, and the model structure, through multi-modeling, the forecast improvements are achieved in terms of CSI.

Table 1. Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI) of the deterministic drought forecasts from all seven forecasting techniques.

1 st Month	POD	FAR	CSI
NWS_TB_LST	0.3929	0.2771	0.3415
NWS_TB	0.3694	0.2304	0.3326
NWS_OL	0.4800	0.3761	0.3723
VIC_TB_LST	0.8235	0.5853	0.3808
VIC_TB	0.8188	0.5930	0.3734
VIC_OL	0.6000	0.5039	0.3728
PF-SBC	0.5341	0.3401	0.4188
2 nd Month	POD	FAR	CSI
NWS_TB_LST	0.4265	0.2281	0.3787
NWS_TB	0.4160	0.2016	0.3764
NWS_OL	0.4139	0.2566	0.3621
VIC_TB_LST	0.8193	0.5618	0.3996
VIC_TB	0.7878	0.5758	0.3807
VIC_OL	0.5714	0.5152	0.3556
PF-SBC	0.5525	0.3308	0.4340
3 rd Month	POD	FAR	CSI
NWS_TB_LST	0.4649	0.2643	0.3984
NWS_TB	0.4497	0.2452	0.3924
NWS_OL	0.4421	0.2741	0.3789
VIC_TB_LST	0.8121	0.5452	0.4115
VIC_TB	0.7628	0.5563	0.3899
VIC_OL	0.5237	0.5166	0.3358
PF-SBC	0.5655	0.3843	0.4180

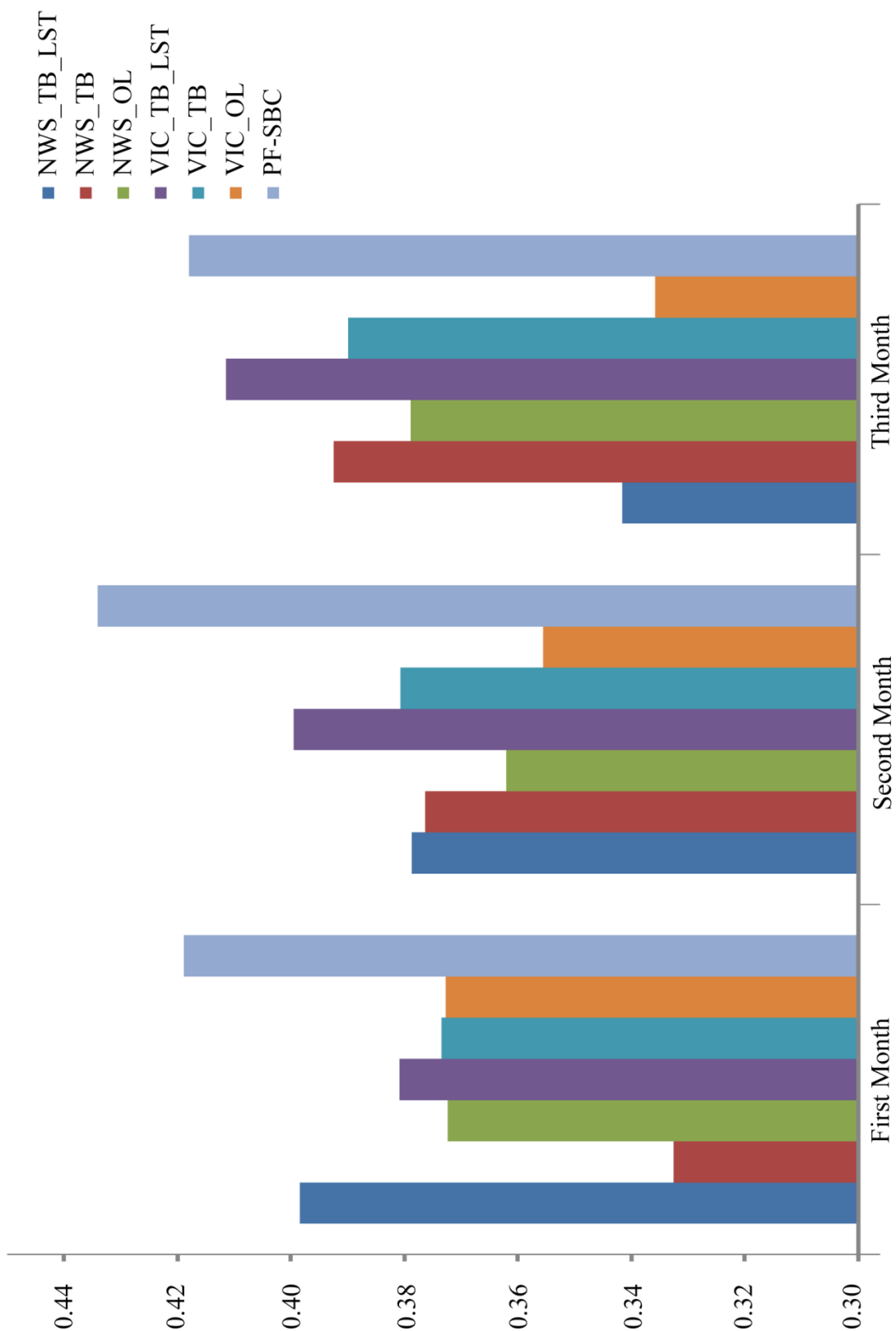


Figure 10. Histogram of the Critical Success Index (CSI) for all forecast months.

Probabilistic forecast verification in this study begins with the results in Table 2, where the reliability, as estimated from the decomposed Brier Score, the number of significant bins from the Reliability Diagram (95% confidence), and the CDF values from the B-CDF and the PB-CDF are presented. The reliability metric and number of significant bins highlight the assessment of drought forecast reliability through a binning approach, and the B-CDF and PB-CDF values show how the probabilistic forecasts can be evaluated as a single group. A few important observations about the accuracy of the forecast methods, and the accuracy of the verification measures, are apparent from Table 2.

Table 2. Reliability (from the decomposed Brier Score), the number of significant bins from the reliability diagram, and the CDF values from the Binomial (B-CDF) and Poisson-Binomial (PB-CDF) distributions.

1 st Month	Binning Methods		CDFs	
	Reliability	Bins	B-CDF	PB-CDF
NWS_TB_LST	0.069	4	1.00	1.00
NWS_TB	0.058	6	1.00	1.00
NWS_OL	0.035	4	0.96	0.99
VIC_TB_LST	0.451	3	0.00	0.00
VIC_TB	0.422	3	0.00	0.00
VIC_OL	0.228	3	0.00	0.00
PF-SBC	0.033	4	0.36	0.29

2 nd Month	Binning Methods		CDFs	
	Reliability	Bins	B-CDF	PB-CDF
NWS_TB_LST	0.057	2	1.00	1.00
NWS_TB	0.060	3	1.00	1.00
NWS_OL	0.046	4	1.00	1.00
VIC_TB_LST	0.349	3	0.00	0.00
VIC_TB	0.350	3	0.00	0.00
VIC_OL	0.281	1	0.00	0.00
PF-SBC	0.035	5	0.61	0.62

3 rd Month	Binning Methods		CDFs	
	Reliability	Bins	B-CDF	PB-CDF
NWS_TB_LST	0.064	1	1.00	1.00
NWS_TB	0.066	3	1.00	1.00
NWS_OL	0.054	4	1.00	1.00
VIC_TB_LST	0.307	2	0.00	0.00
VIC_TB	0.340	3	0.00	0.00
VIC_OL	0.338	1	0.01	0.00
PF-SBC	0.051	2	0.79	0.83

A first observation from Table 2 is that the binning methods often disagree. In the first month, the VIC_TB case has the greatest number of significant bins of all forecast cases and lead times. Conversely, the PF-SBC case produces the lowest reliability (optimal is 0) score of all cases during this forecast month. Further, the PF-SBC has the lowest reliability score of all cases at each lead time, but only produces the greatest number of significant forecast bins during the second forecast month. To explain this incongruity, the reliability diagram for the first month forecasts are presented in Figure 11. In this figure, the VIC_TB case is shown to have six forecast probability bins falling within the 95% significance envelope, yet has outliers at lower forecast probabilities. While a large portion of the reliability diagram may fall within the significant envelope, a few outliers can adversely affect the reliability score, as it is the mean square error from the 1:1 line. With respect to the PF-SBC case, less than half of the forecast bins fall within the significant envelope, yet the PF-SBC produces a reliability diagram that is closer to uniform than the VIC_TB case, leading to a lower reliability score. This observation that both binning metrics typically disagree suggests that at least one metric is suboptimal. Further examination requires the use of all forecasted probabilities simultaneously to maximize the information from the observations.

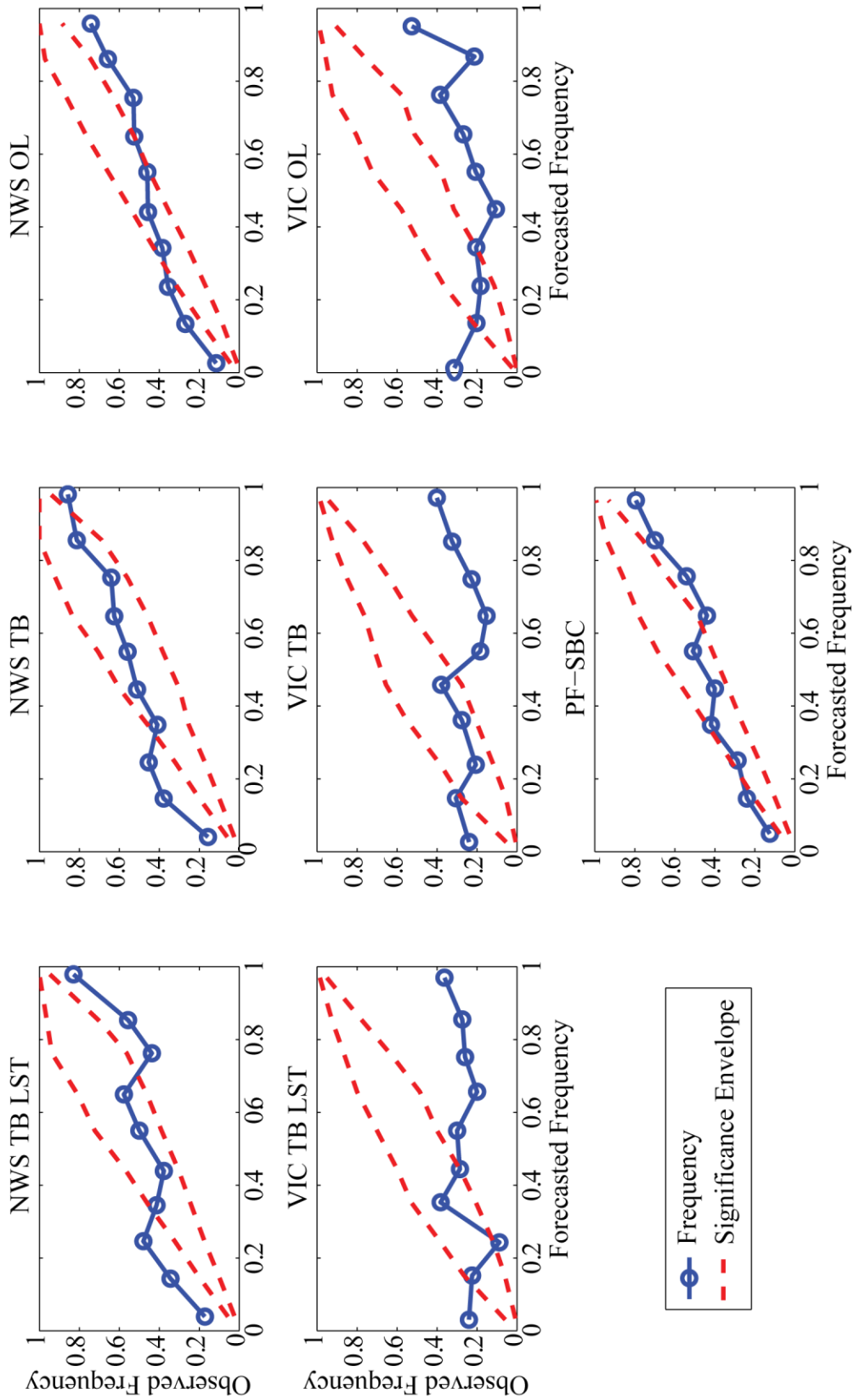


Figure 11. Reliability diagram for the drought forecasts from all methods, during the first forecast month, with the frequency noted by the blue blue circles, with connecting line, and the 95% significance envelope shown with red dashed lines, which was estimated from the Poisson-Binomial Distribution.

The B-CDF and PB-CDF values also differ in their assessment of reliability, but not to the extent of the binning methods. With respect to these scores, a value closer to 0.5 is optimal (the observed number of droughts is closest to the forecasted expected value), but the values should be interpreted from a hypothesis testing viewpoint. As explained in section 6.2.2, if the CDF value is outside of the 2.5% and 97.5% envelope, the hypothesis of reliability is rejected, otherwise the forecast must be assumed to be reliable. From Table 2, the B-CDF suggests that 4 forecast scenarios are reliable, whereas the PB-CDF suggests that only three are reliable. This result is consistent with those presented in section 8.2, where it was shown that use of a Binomial approximation decreases one's ability to reject an unreliable distribution, potentially leading to type 2 errors. The reason for this discrepancy is highlighted in Figure 12, where histograms of the forecasted probabilities are shown for the first month drought forecast. In this month, these forecast probability distributions display a U-shaped pattern, which reduces the accuracy of the Binomial approximation. Note that the Binomial approximation approaches the exact solution as the forecasted probabilities approach the same value (e.g. climatology), but this also minimizes the sharpness of the forecast. Here it is suggested that the PF-SBC is the only case which produces reliable probabilistic predictions, reliably forecasting drought probabilities in all months, based on the analysis of the PB-CDF. In the event that both the PF-SBC and the VIC_OL model were both deemed reliable, as would be the case if the Binomial approximation were relied upon for the first month, the VIC_OL method would be competitive with the PF-SBC method, as both meet the reliability requirements according to the B-CDF. Since both the VIC_OL

and PF-SBC have similar forecast variances (see Figure 13), it is likely that the VIC_OL would be selected for one month forecasting over the PF-SBC method based on the B-CDF value, due to its simpler design. This conclusion is clearly erroneous from the exact solution to the reliability condition. In light of this clear error resulting from the B-CDF value, it is suggested that that the PB-CDF value be used in future studies.

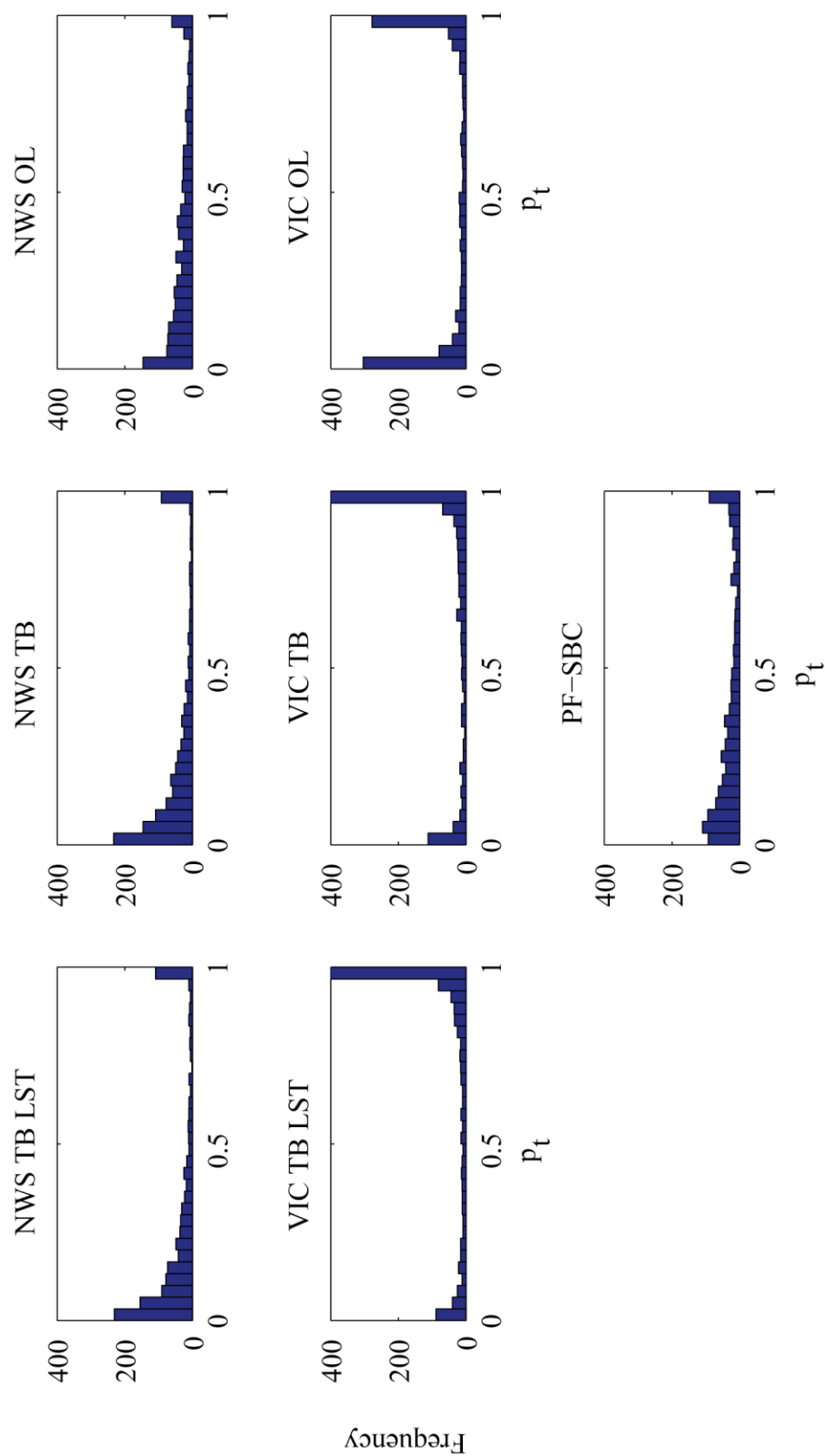


Figure 12. Histogram of the forecasted drought probabilities for each method during the first forecast month.

The prior examination of the forecast cases with the PB-CDF highlights some important issues related to the use of binning metrics. First, a reliable forecast, according to the PB-CDF will not necessarily occupy the significant envelope for all forecasts bins. This is somewhat surprising, but one may realize that the PB-CDF is primarily related to overall bias, whereas the binning approach can assess over/under confidence. Although the PF-SBC method is reliable from the PB-CDF values, it is clearly outside the significant envelope for multiple bins, and therefore may not be reliable at all probabilities. This hints that reliability assessment may be more effectively performed with a combination of single and multiple bin assessments, to determine if the forecast falls in the significant envelope for all bins. In this scenario, one will utilize a single bin scenario to maximize the ability to reject the reliability hypothesis, but also rely on a multiple bin approach to ensure that the model is reliable at each forecast group as well.

Another interesting note is that Table 2 suggests that the reliability score generally agrees with the PB-CDF on which forecast has the highest probability to be reliable. This verifies that the normal approximation assumed in the BS is reasonable, but it is ambiguous in distinguishing between reliable and unreliable forecasts. Two clear examples are observed in the first and third months. In these months, the NWS_OL and PF-SBC cases produce similar reliability, yet clearly differ in their location within the PB-CDF. Further, the NWS_OL case has a lower reliability score during the first forecast month than the PF-SBC has during the third month, yet the PF-SBC is deemed reliable from the PB-CDF in each month, and the NWS_OL case is not. Since there are no clear guidelines for distinguishing between reliable and unreliable forecasts based on the BS,

interpretations of this metric will often lead to over emphasis of sharpness, which is similar to the finding from the synthetic experiment.

Based on the observed location within the PB-CDF from each forecasting case, it appears that the NWS models forecast drought too infrequently, the VIC model forecasts drought too often, and the PF-SBC forecasts transition from over-forecasting to under-forecasting drought frequency with lead time. These results are consistent with the findings from the water supply forecasting experiment, where the NWS models were found to have a high flow bias, and the VIC models was found to have a low flow bias. With respect to the transitioning bias of the PF-SBC case with lead time, it is likely that the high bias is a result of losing sensitivity to the initial conditions over time. Since the model is forced with climatological data, the forecasts will approach climatology with sufficient lead time, which is the point at which sensitivity to initial conditions is negligible. Given that climatology is used as forcing during the forecast period, drought frequency will be under-forecasted with sufficient lead time, as the entire study period is drought prone in comparison to the sampled climatological data. Though the lead time at which the model loses sensitivity to initial conditions is not quantified here, it appears that the effects of approaching climatology may begin to occur in as short as three months. This is further evidenced by a reduction in the sharpness with time as shown in Figure 13. From Figure 13, it is clear that the first month forecast from the PF-SBC is the sharpest forecast of the three months, and is therefore the best forecast given that all three months are forecasted reliably. The sharpness decreases over time, indicating increasing uncertainty as the model approaches climatology. This analysis provides a simplistic

analysis of forecast sensitivity to initial conditions, but this is examined in detail in section 8.5. A final finding from Figure 13 is that the PF-SBC forecast is generally the least sharp, which is due to a more complete accounting of uncertainty. Although it is unfortunate that the best forecast is the most uncertain, this is the most honest accounting of uncertainty, and therefore should be chosen as best of all cases presented.

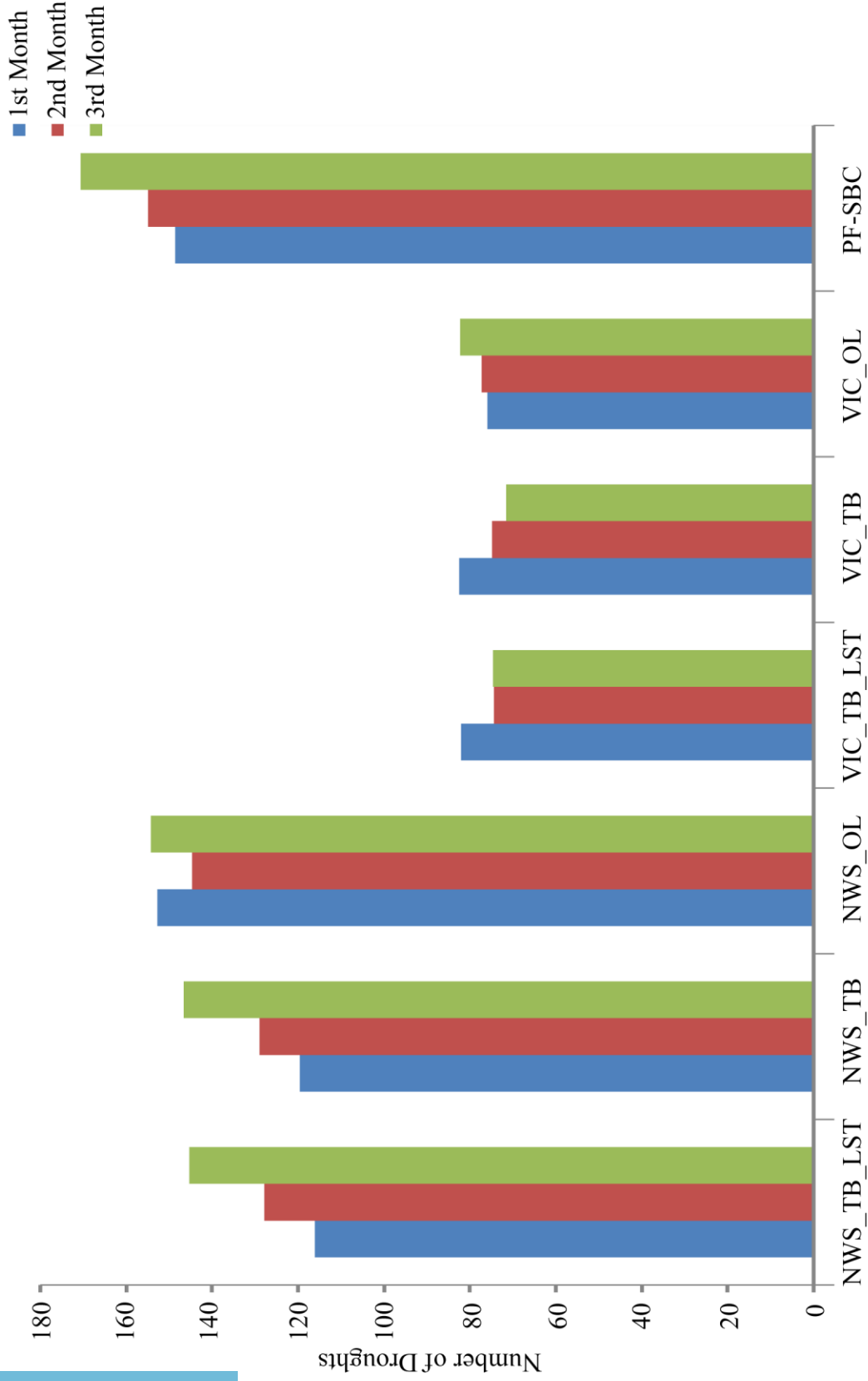


Figure 13. Variance of the the forecasted drought probabilities from each method and during each month.

8.4 Assessing the Need for a New Drought Index

A specific example of the errors associated with spatial analysis of standardized indices is examined with simulations of land surface water states via the VIC model over the UCRB. This example compares the absolute value LWS with the SLWI. Over the entire UCRB, the LWS and SLWI are averaged spatially, and the corresponding time series are plotted in Figure 14.

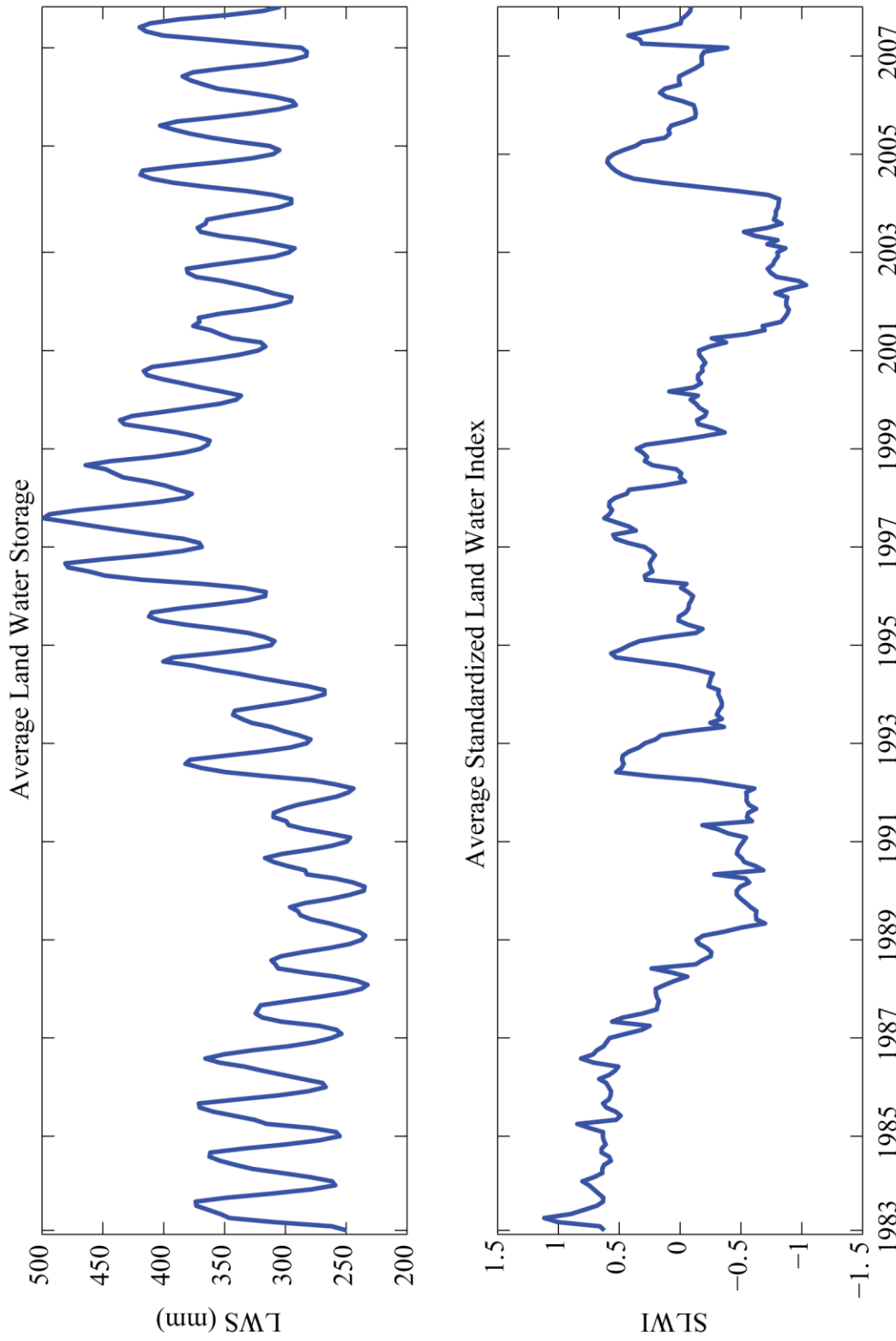


Figure 14. The temporal evolution of the spatially averaged land water storage (LWS) and standardized land water index (SLWI) over the upper Colorado River Basin

The LWS and SLWI in Figure 14 show similar long term patterns, reflecting the relationship between the absolute and standardized values. Although these two time series' show similar long term trends, there are two specific differences that may be observed. First, the LWS shows the annual fluctuations, whereas the standardized index removes these fluctuations by estimating drought intensity with respect to monthly climatology. A second and more important difference is the location of the minimum value, which should correspond to the worst drought over this time period. From this figure, it is apparent that the LWS and SLWI disagree on the most intense drought in this basin. While the LWS indicates that lowest basin water storage during this time period occurred around 1990, the SLWI indicates the worst drought occurred around 2002. The average LWS clearly corresponds to the driest overall time, as it is an absolute measure of basin water storage, but this is improperly identified by the SLWI due to the standardization process. By viewing drought through a historical perspective, some information about overall dryness is lost. Since the LWS and SLWI are based on the same data, the standardization process is identified as causing errors in the analysis of basin-wide drought. This is primarily a result of errors in the standardized methodology in a region of spatially variable water states.

The spatial distribution of water within the UCRB is primarily related to elevation, as shown in Figure 15. In this figure, the relationship between the normalized mean LWS (\bar{S}_i) and elevation percentile (i) of each model grid cell is presented, as shown in equation (64). In equation (64), $S_{i,t}$ is the LWS for cell i at time t , where T is the total number of time steps of data available, which is normalized by dividing the

cumulative mean LWS values ($\sum_{j=0}^i \bar{S}_j$) by the maximum cumulative mean cell LWS value ($\sum_{i=1}^N \bar{S}_i$) in equation (65), where N is the total number of grid cells. Similarly, the cell LWS variance (\hat{S}_i) is estimated according to equation (66), and the cumulative normalized variance is estimated by equation (67). In order for spatial analysis with standardized indices to be optimal, the basin water will need to be uniformly distributed, which is shown with the dashed line in Figure 15. In the UCRB, the actual relationship between average LWS increases non-linearly, as noted by the solid blue line. From this figure, it is clear that the majority of water storage is located in the highest elevation regions. For example, in examining the 70th cell elevation percentile, the corresponding average LWS is roughly 40%, suggesting that 60% of the water storage in the UCRB is in the highest 30% of the basin area. Further, the standardized variability in water storage is presented in the lower subplot of Figure 15. Similar to basin average water storage, the variability in water storage is spatially distributed, with a majority of the basin water variability being controlled by the highest elevation regions. Given this environment, the lower elevation regions are given too much weight through the standardized process, because standardized indices inherently assume uniform distribution of water, thus small deviations from the average in these grid cells have a disproportionately large impact on drought analysis.

$$\bar{S}_i = \frac{1}{T} \sum_{t=1}^T S_{i,t} \quad (64)$$

$$\hat{S}_i = \frac{1}{T} \sum_{t=1}^T (S_{i,t} - \bar{S}_i)^2 \quad (65)$$

$$norm(\bar{S}_i) = \frac{\sum_{j=0}^i \bar{S}_j}{\sum_{i=1}^N \bar{S}_i} \quad (66)$$

$$norm(\hat{S}_i) = \frac{\sum_{j=0}^i \hat{S}_j}{\sum_{i=1}^N \hat{S}_i} \quad (67)$$

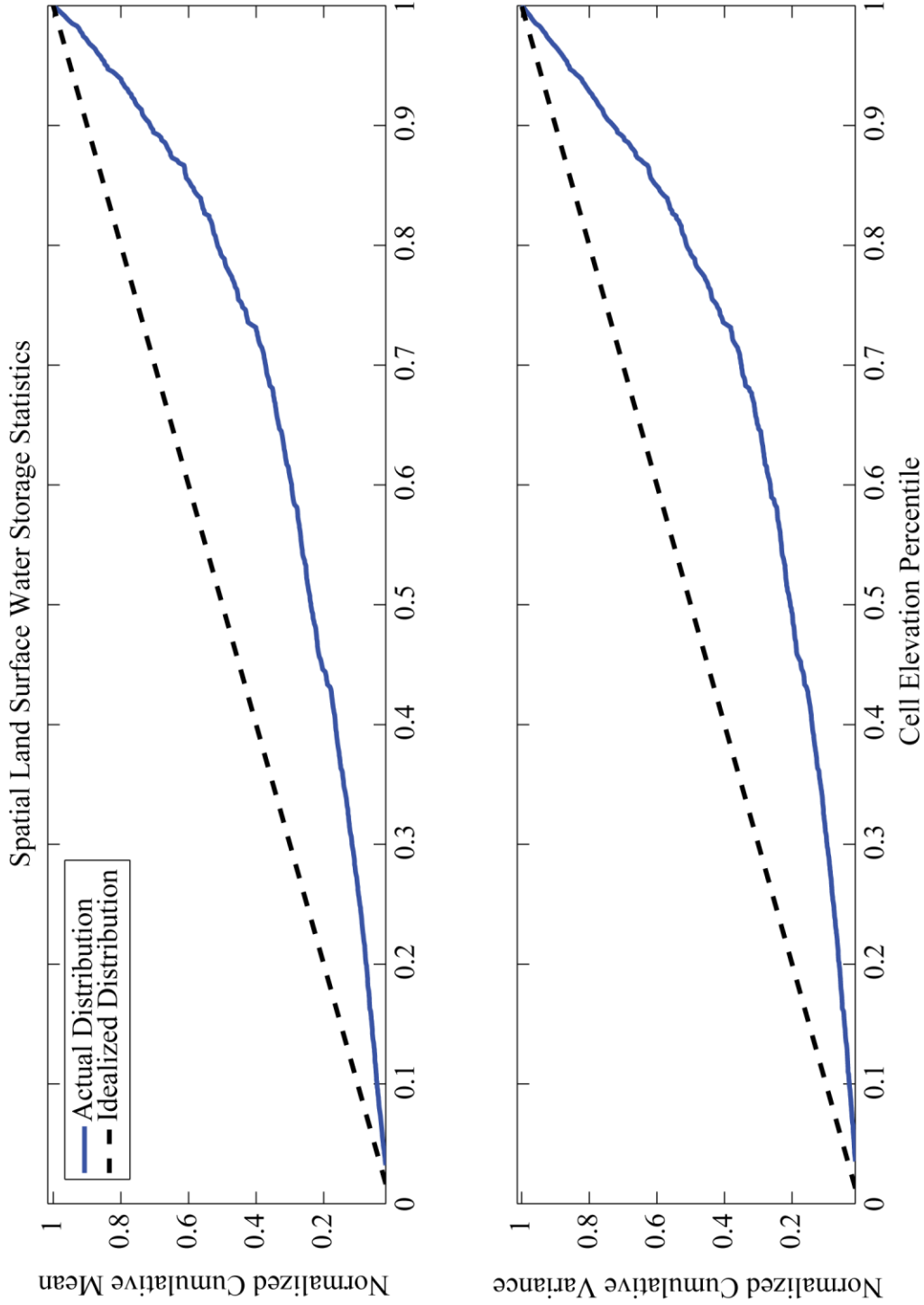


Figure 15. Normalized cumulative mean and variance of water storage in each model grid cell, plotted with respect to the percentile of the elevation of the grid cell.

A comparison of the MSMI with the SLWI is provided in figures 16 and 17. These figures show the spatial distribution of each drought index during May of 1990 (Figure 16) and 2002 (Figure 17). These two months were chosen to correspond to the minimum spring basin water storage (May 1990) and minimum spring SLWI (May 2002). From Figure 16, it is important to note the generally similar values of MSMI throughout the interior portions of the basin, whereas the SLWI indicates a variety of wet and dry states in this region. Such consistency throughout the interior portions in the MSMI suggests that the little variability exists in the LWS throughout this region. From this finding, it is clear that the SLWI is highly sensitivity to small changes in water storage. Since the SLWI is standardized by climatology, and LWS has little variability in dry climates due to generally low precipitation and high evaporation rates, slight shifts in the LWS throughout the interior portion of the UCRB reflect large variability spatially. This observation is important for understanding problems associated with standardized indices.

Further analysis of the errors associated with SLWI is performed by comparing Figure 16 and Figure 17. In contrast to the high spatial variability of drought found with the SLWI in May of 1990, SLWI in May of 2002 more consistently indicates drought throughout the interior portions of the basin, but the MSMI only shows slightly drier status. These low elevation regions receive equal weight from the SLWI when analyzing basin-wide drought, yet the MSMI suggests that these regions have little impact on total basin water storage. Figure 17 also indicates wetter than average conditions in the far northern portion of the basin from both indices. This region receives a much larger

portion of precipitation than the interior portions, and therefore may make up for the drier conditions further south in the basin. When comparing the MSMI between 1990 and 2002, it appears that much of the outer portions of this basin, which are the higher elevation regions, tend to contain more water during 2002 than 1990. Although a general pattern is not apparent from the SLWI in these regions, the MSMI indicates that this regions stored more water during 2002, thus making up for the minor deficits in the lower elevation regions. Although the SLWI suggests a more intense drought overall in 2002, as compared to 1990, this is likely incorrect as the MSMI indicates generally greater LWS throughout the basin in 1990 than 2002. This scenario indicates that standardized indices may be misleading, and that the use of a physically-based index has potential to advance drought monitoring.

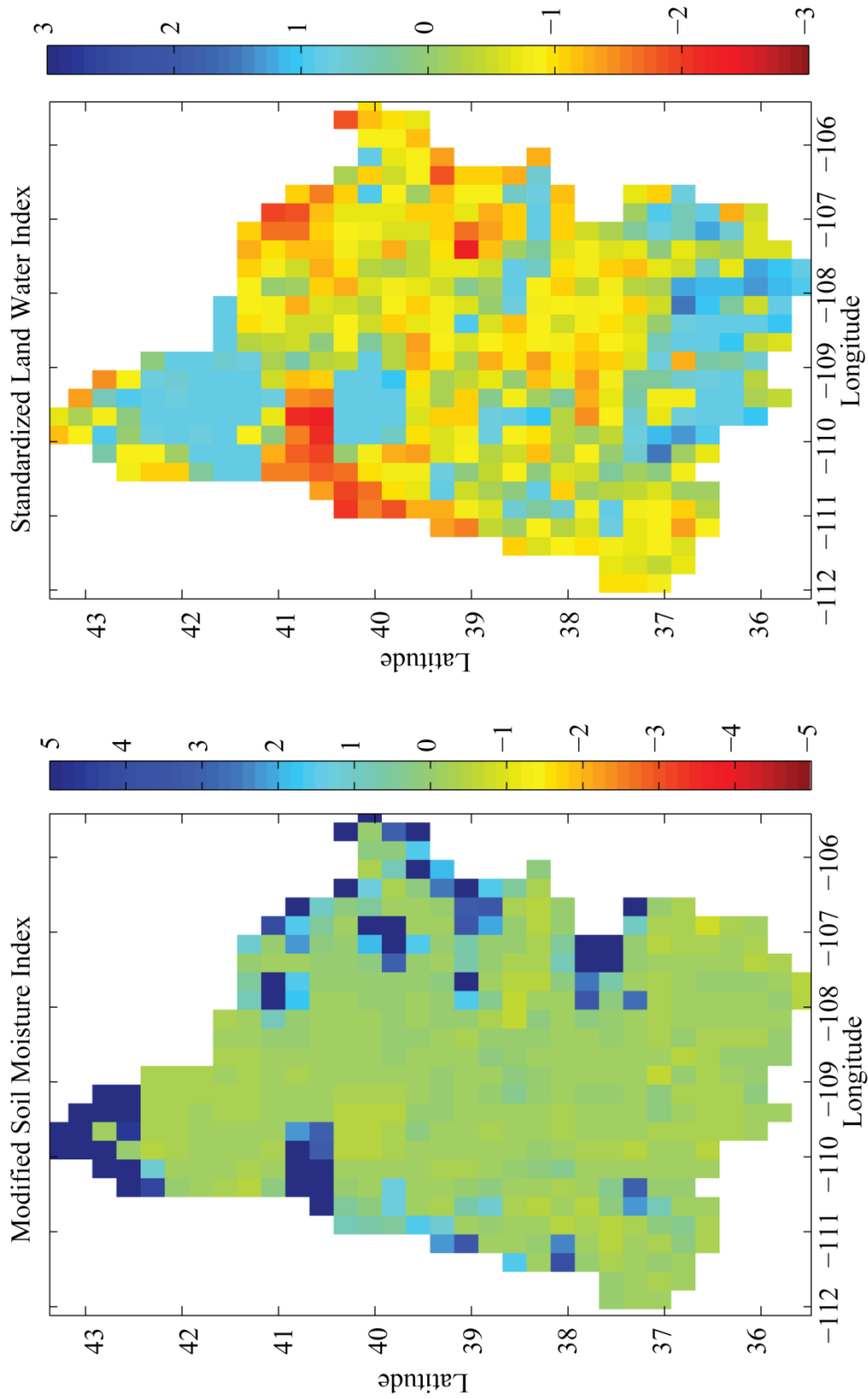


Figure 16. Comparison of the Modified Soil Moisture Index and the Standardized Land Water Index over the upper Colorado River Basin during May of 1990.

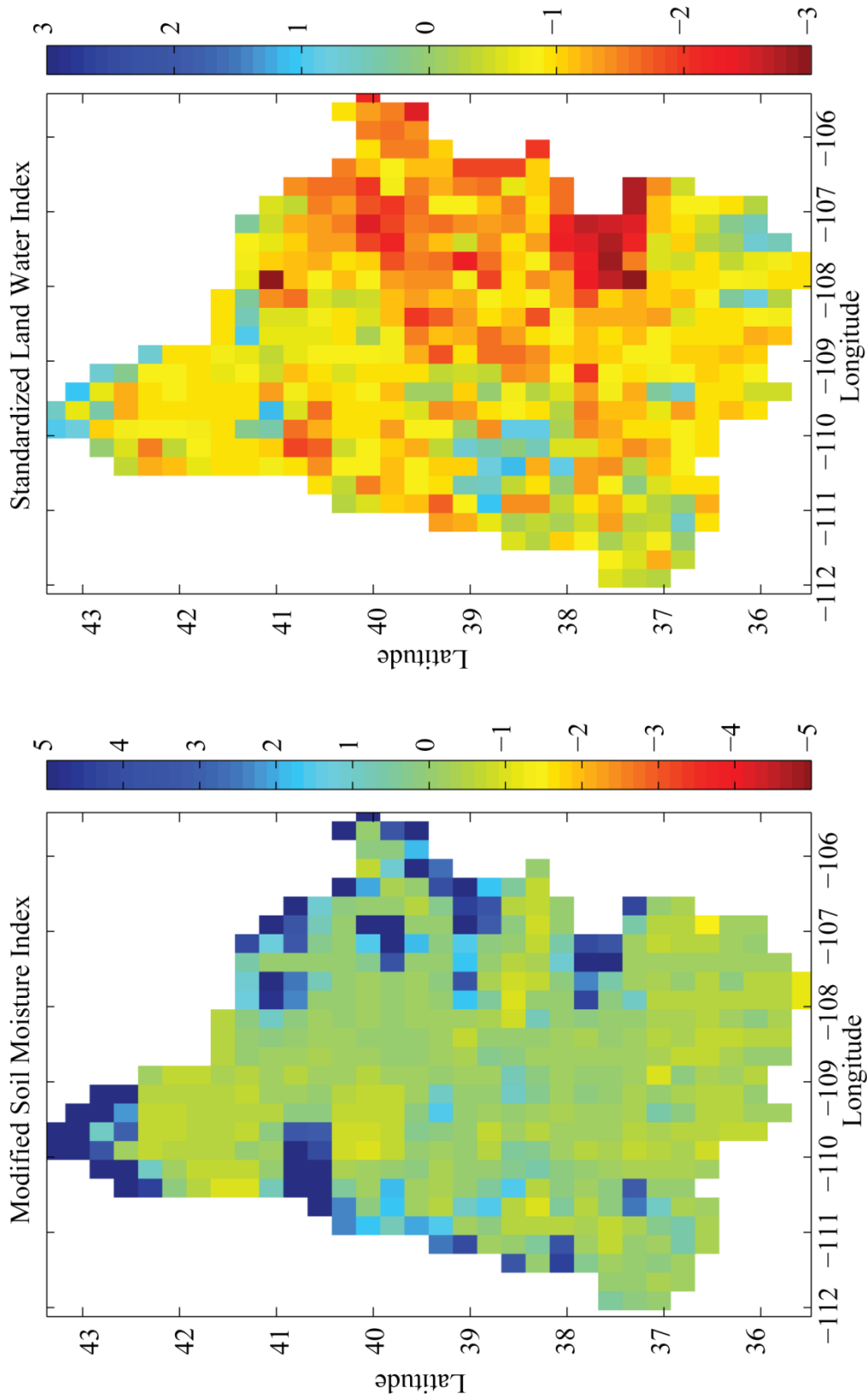


Figure 17. Comparison of the Modified Soil Moisture Index and the Standardized Land Water Index over the upper Colorado River Basin during May of 2002.

Problems with the analysis of standardized indices may also be observed when utilizing advance modeling techniques, such as DA, due to the ability of these techniques to push hydrological states away from a model's own climatology. In order to examine this issue, Figures 18 and 19 are presented to compare the SLWI and MSMI estimated from model simulations and DA. These figures show the spatial distribution of drought within the UCRB during May of 2004. Starting with a comparison of model estimations of SLWI, with and without DA, Figure 18 shows is a stark contrast between the two subplots. In this figure, the SLWI without DA is spatially consistent, having few grid cells showing extreme drought or excess water. Alternatively, when using DA, the SLWI shows significant variability in the central portion of the basin, and many cells at the extremes. This is evidence that the application of standardized indices is problematic when using DA. Since the SLWI is tuned to the model specific climatology, a theoretically more accurate estimation of LWS may actually be incorrect in relation to the historical model distribution. The errors in the model are persistent in the climatology, and are therefore existent when examining the SLWI with the improved states from DA. Given this scenario, it may be advantageous to use a physically-based index, leading to improved drought characterization with increasingly accurate hydrological state estimation.

Figure 19 shows that the MSMI is much more consistent with and without DA than the SLWI. In the no assimilation case, the MSMI shows excess water in the far northern and eastern portions of the basin, with drier regions in the north-central and southern portions. After application of DA, the water storage in the northern and eastern

portions is less wet, and conditions are slightly drier in the interior portions of the UCRB. Unlike the SLWI, the DA case is still a reasonable estimate of the conditions, as it is bound by physical principles and not model climatology. This shows that a physically-based drought index may be more useful when comparing results from model simulations and DA. With a greater applicability to DA estimated states, physically-based indices may become a vital part of drought analysis, as DA systems become increasingly common in operational drought monitoring systems.

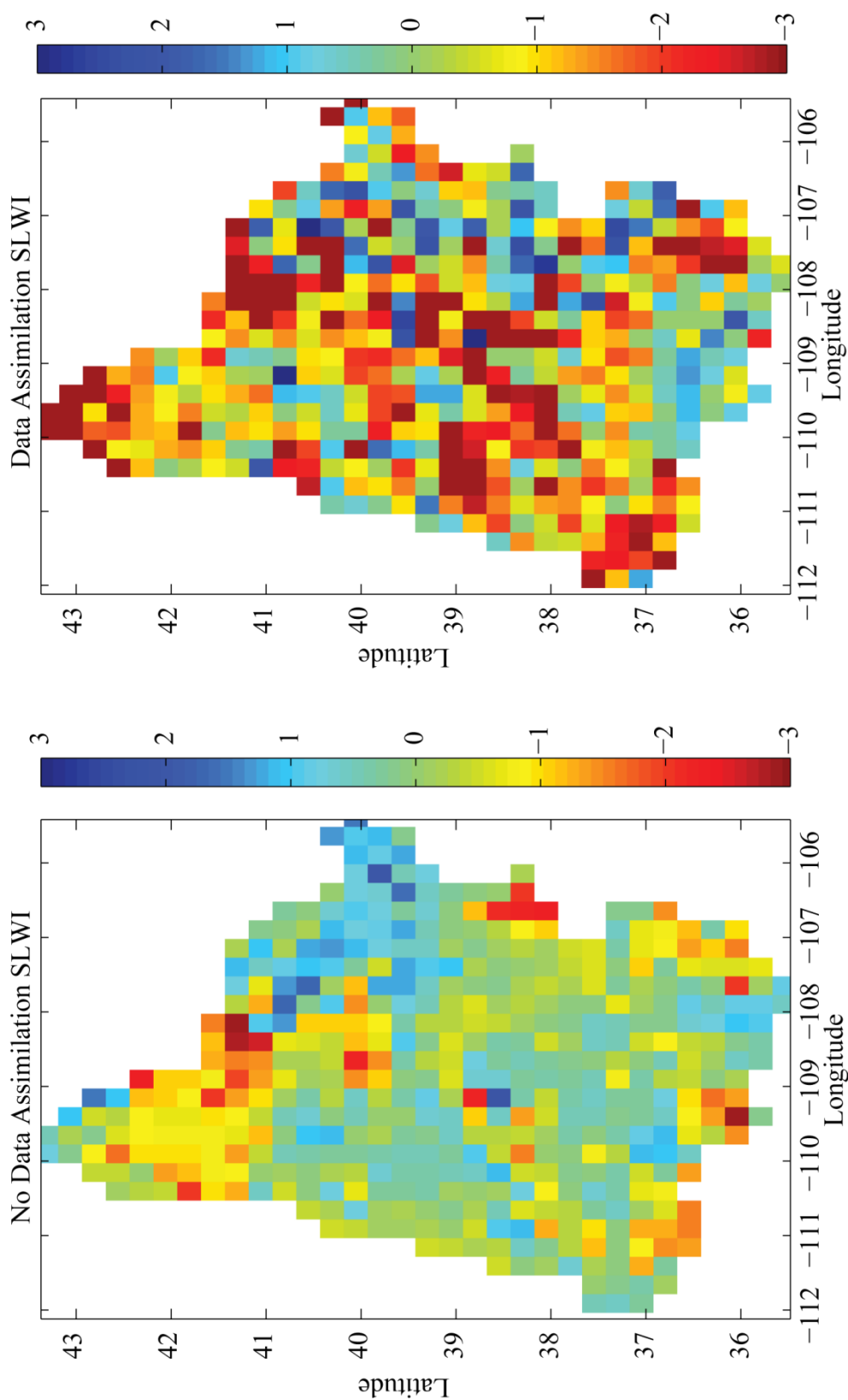


Figure 18. Comparison of the Standardized Land Water Index with and without data assimilation over the upper Colorado River Basin during May of 2004.

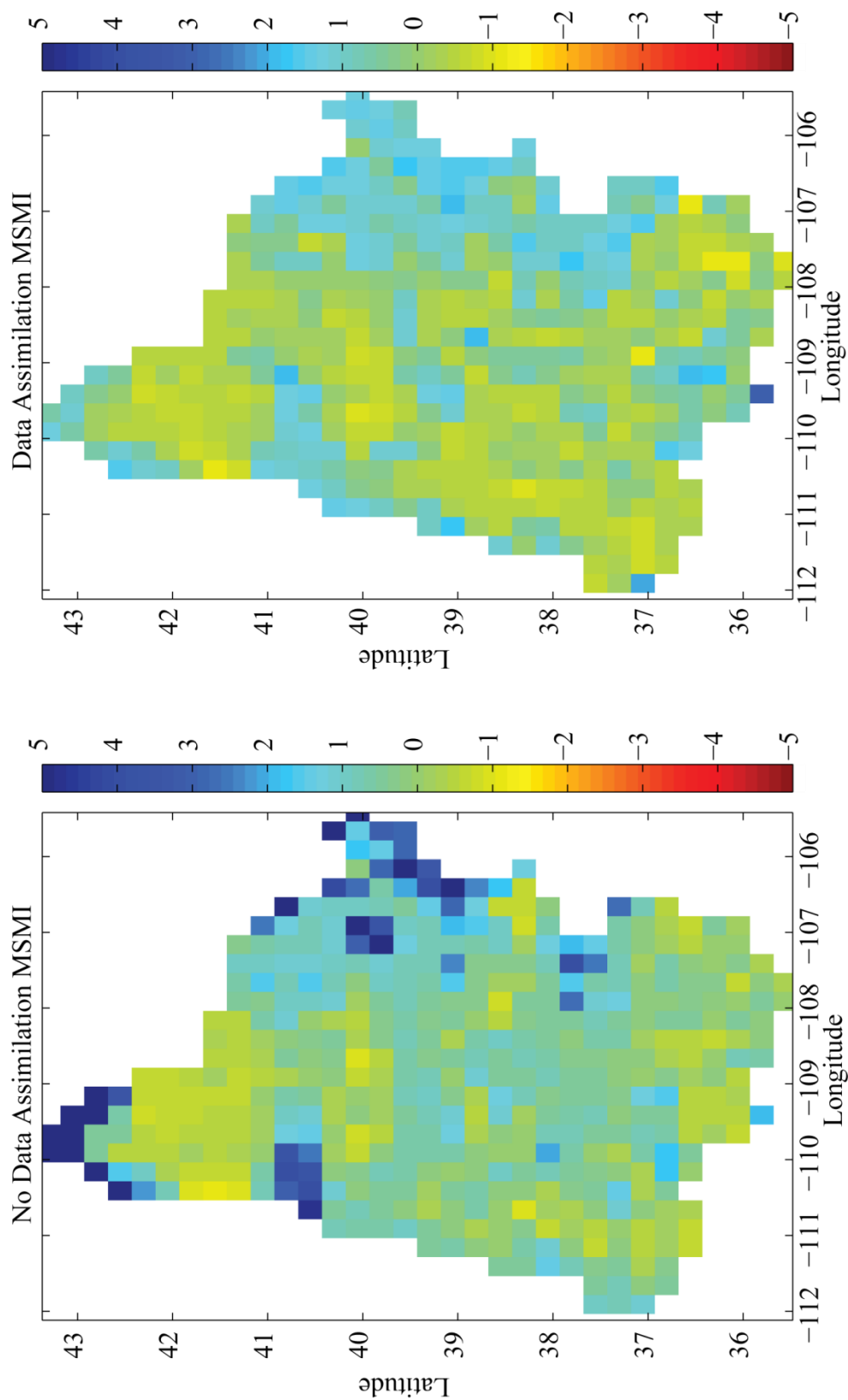


Figure 19. Comparison of the Modified Soil Moisture Index with and without data assimilation over the upper Colorado River Basin during May of 2004.

8.5 Sensitivity of Droughts to Initial Conditions

8.5.1 Forecast Initial Conditions

The time period from October 1st, 2003 through September 30th 2008 was marked by frequent drought conditions, making this a prime study period for analysis of drought forecast recovery. With these persistent drought conditions, the LWS for each forecast date is below average, as shown in Table 3. In this table, the spatially averaged LWS initial condition expected value from the climatological distribution and assimilation of TB and LST into the VIC model is presented. This table clearly shows that LWS from assimilation, at each forecast initialization date, is below the average climatological value, indicating drought conditions. Of these years, 2007 was found to be the worst drought for April over the entire basin, and 2006 was found to be the worst drought for July over the entire basin. Alternatively, the least intense drought year was 2003 for April and 2005 for July. The time for complete recovery from such droughts is of great importance for understanding drought processes and forecasting, and the remainder of this results in this dissertation will seek to quantify the recovery time from these droughts, and the rate at which drought recovers.

Table 3. Basin-wide average land water storage initial condition, in mm, for climatology and each year, from each forecast starting date.

Year	April 1 st	July 1 st
Climatology	190.62	118.00
2003	172.50	85.71
2004	137.35	87.76
2005	155.10	104.32
2006	125.85	73.86
2007	117.32	82.66
2008	132.31	84.60

An initial examination of the forecasts from each start date is provided in Figures 20 and 21. In these figures, the spatially averaged median (dotted line) and 95% predictive bounds (solid lines) are shown for the forecast (red line) and climatology (blue line). This representation shows the progression of reducing forecast bias, and increasing width of the 95% predictive bounds, as climatology is approached. This indicates a slow loss of sensitivity to the forecast initial conditions, which is highly dependent on the drought intensity at the initial forecast date. For example, forecasts starting from each month in 2005 clearly approach climatology, and are indistinguishable from climatology by the end of the 360 day forecast. In addition, the 2005 forecasts approach climatology in a much shorter lead time than all other forecasts, due to the proximity of the initial states to the climatological average. Alternatively, the forecast mean and 95% predictive bounds from April of 2007 do not appear to completely match climatology even at 360 days, which indicates an extended length of memory in the UCRB to initial conditions. Similar observation may be made from the July 2006 forecast, where the forecast required greater lead time to approach climatology than all other years. From a qualitative standpoint, Figures 20 and 21 are very informative of the behavior of forecasts initialized at different drought intensities, but more detailed analysis requires the use of quantitative measures. As described earlier, the measures used here are the KS test and the RE.

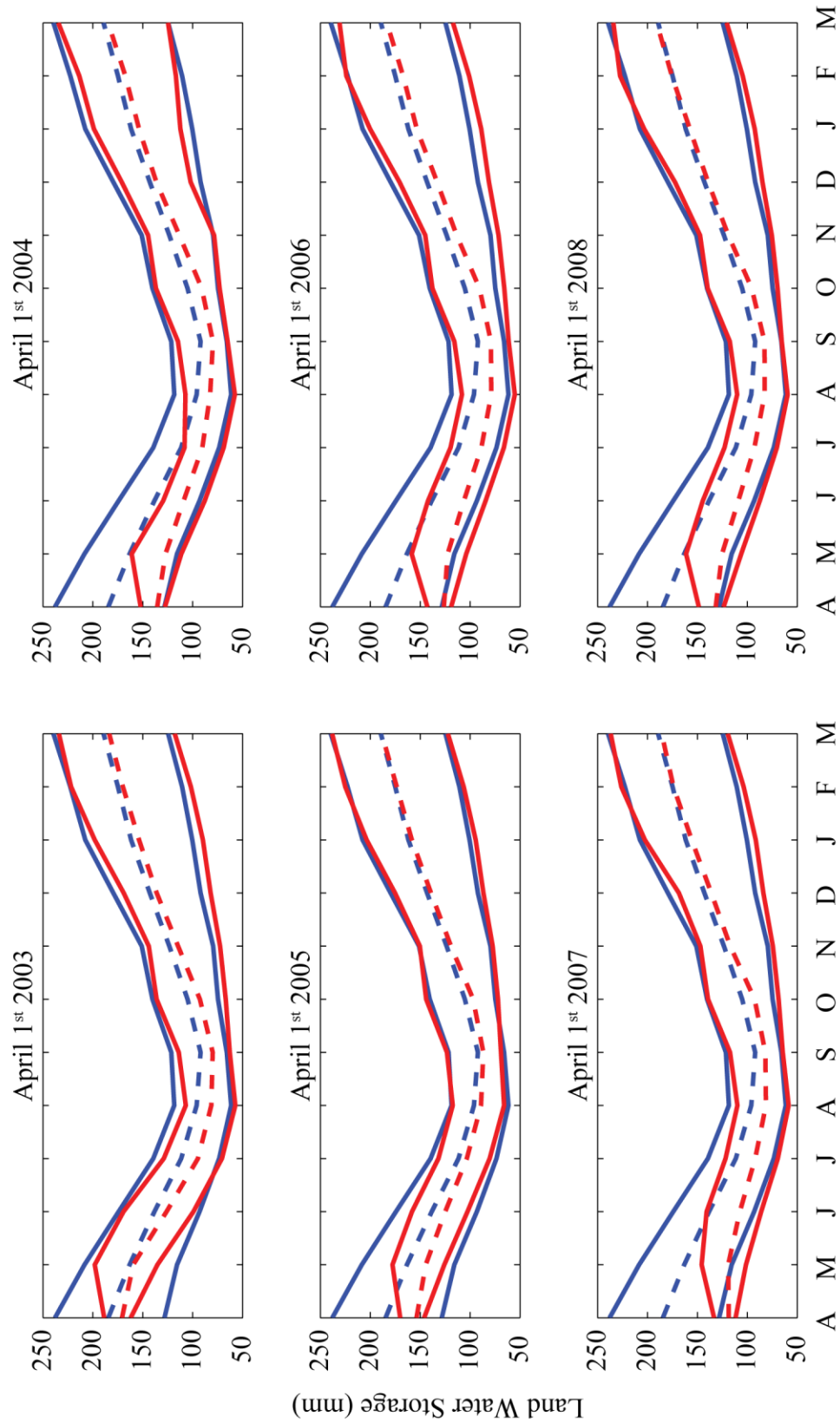


Figure 20. Spatially averaged land water storage (soil moisture + snow water equivalent) for the upper Colorado River Basin, starting on April 1st. The median value is represented with the dotted lines and the 95% predictive intervals are represented with solid lines, with red representing the forecast and blue representing climatology.

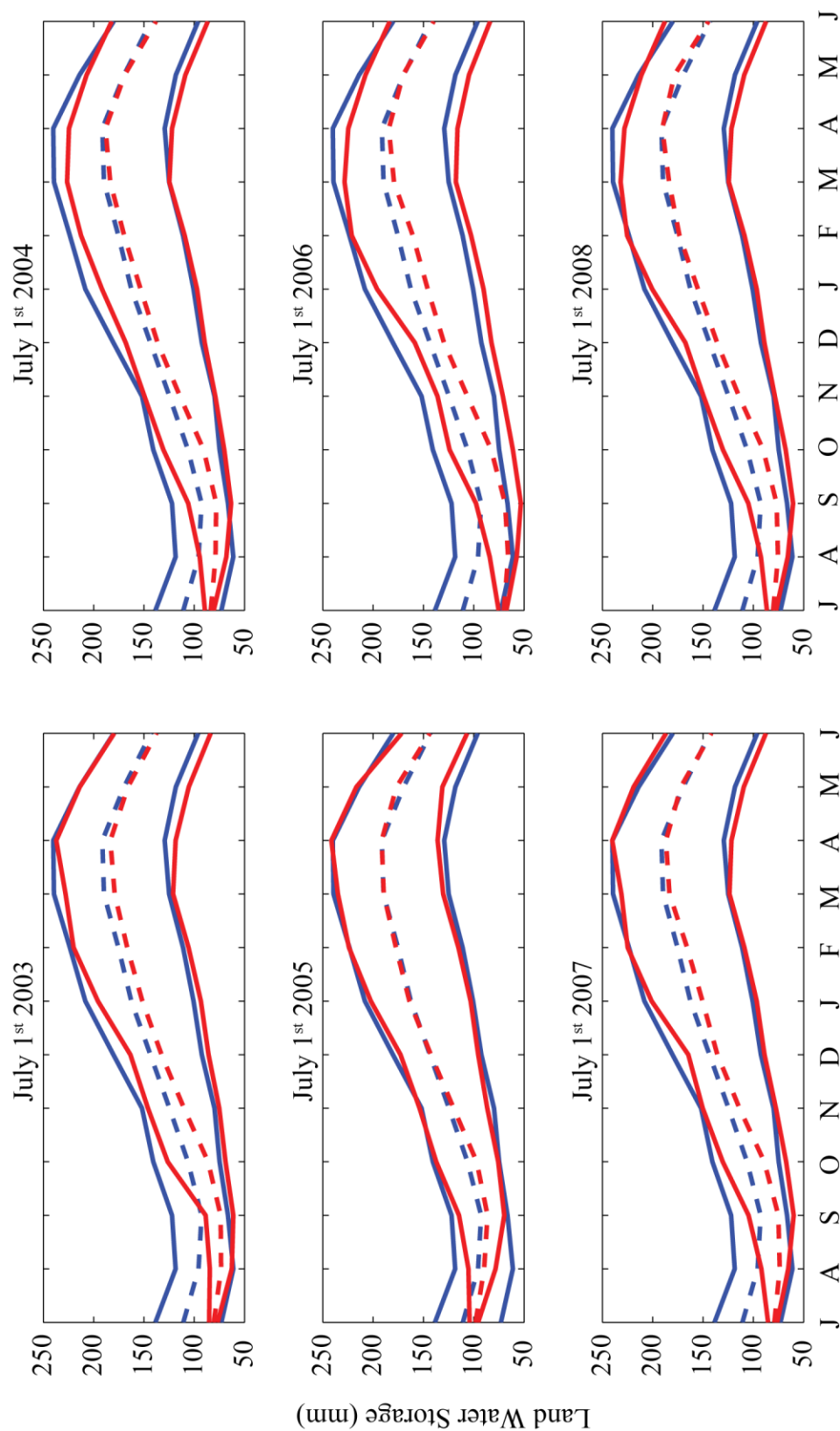


Figure 21. Same as Figure 20, but for all starting dates in July.

8.5.2 Kolmogorov-Smirnov Test

The lead time at which the basin-wide LWS forecast ensemble becomes equivalent to climatological ensemble is quantified with the spatially averaged LWS forecasts, based on the results presented in Figures 20 and 21. Through the KS test, the time at which the spatially averaged forecast and climatology become statistically inseparable is estimated, and is presented in Figure 22. In this figure, the region shaded in black indicates that the distributions are significantly different, and the white region indicates that the forecast and climatology are statistically indistinguishable. Therefore, the time at which the figure transitions from black to white is the drought recovery lead time.

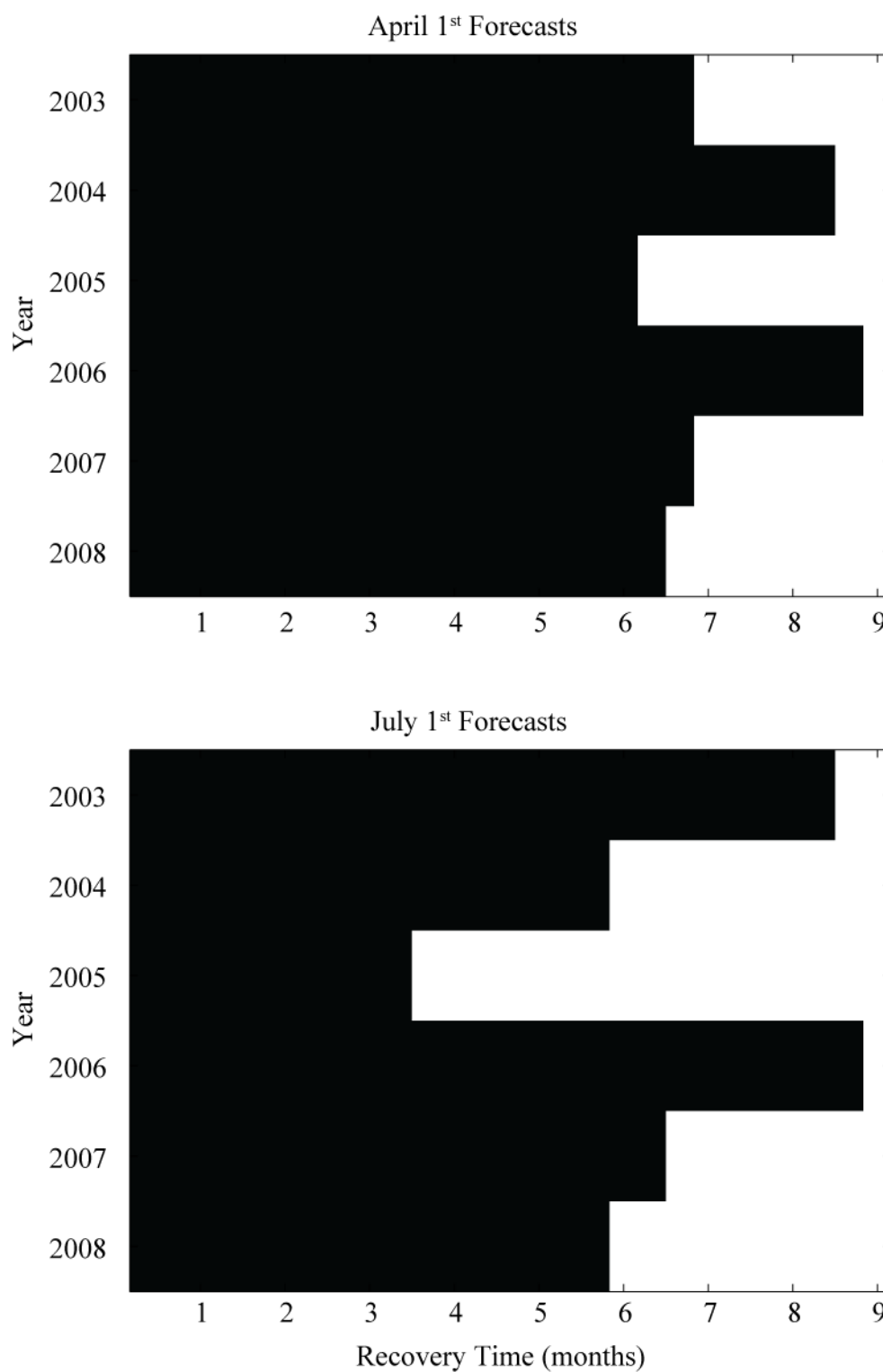


Figure 22. Kolmogorov-Smirnov test for the spatially averaged land water storage, where black regions indicate that the forecast and climatology are significantly different, and white regions indicate the forecast and climatology are indistinguishable.

From Figure 22, it is clear that the magnitude of LWS deficits at the initial forecast date affect the lead time required for recovery. For example, in 2005, forecasts starting in both April and July have their shortest time to recovery, about 6 and 3 months respectively, but display their longest recovery time in 2006, at around 9 months for both. Such a finding is intuitive, as increasingly severe droughts are expected to have increasingly long recovery times. Overall Figure 22 suggests that spatially averaged LWS may take between 6 and 9 (3 and 9) months to recover from the drought conditions observed in April (July). These results suggest different sensitivities to initial conditions than previous studies. From conclusions in previous studies, one would likely assume that drought conditions in April would persist for a maximum lead time of 6 months (Shukla et al., 2013), while drought conditions in July would persist for a maximum lead time of around 3 months (Paiva et al., 2012). Interestingly, the maximum lead time of initial condition influence estimated from previous studies is similar to the minimum drought recovery time determined here. These differing results are explained by the varying perspectives between this study and previous studies. While previous studies examined the relative influence of initial conditions and forcing, this study quantified the lead time at which the initial conditions have no influence. This distinction is important for examining both drought recovery time, and the benefit of improving initial conditions in hydro-meteorological forecasts.

More detailed analysis of drought recovery lead time is performed by viewing spatial patterns of recovery time. Figures 23 and 24 show the spatially distributed drought recovery lead times for the forecasts originating from April and July, respectively. A first

observation from these maps is the extent of the basin which requires nearly the entire 360 day forecast period to recover (shown in red). For forecasts originating in both April and July, large regions in the interior of the UCRB are in deficit conditions throughout the 360 day lead time, with more intense drought years displaying greater recovery time. These interior regions are among the driest portions in the basin, and therefore this scenario may be explained by the difficulty of recovering from a drought when precipitation is sparse even during average climatic conditions. Alternatively, the northern and western portions of the UCRB have among the shortest drought recovery lead times, with some regions reaching climatology within a few months. Due to the increased precipitation in these regions, and the magnitude of typical snow water storage in relation to soil moisture, drought recovery can be quite rapid. Since the LWS in these regions is dominated by snow, it is unlikely that a drought could take more than one year for recovery under normal climatic conditions, due to near complete melt every year, resulting in a clean start after every summer. Alternatively, the LWS in the central portions is dominated by soil moisture, and therefore the memory in this region has the potential to be much longer, as the soil moisture will never reach zero.

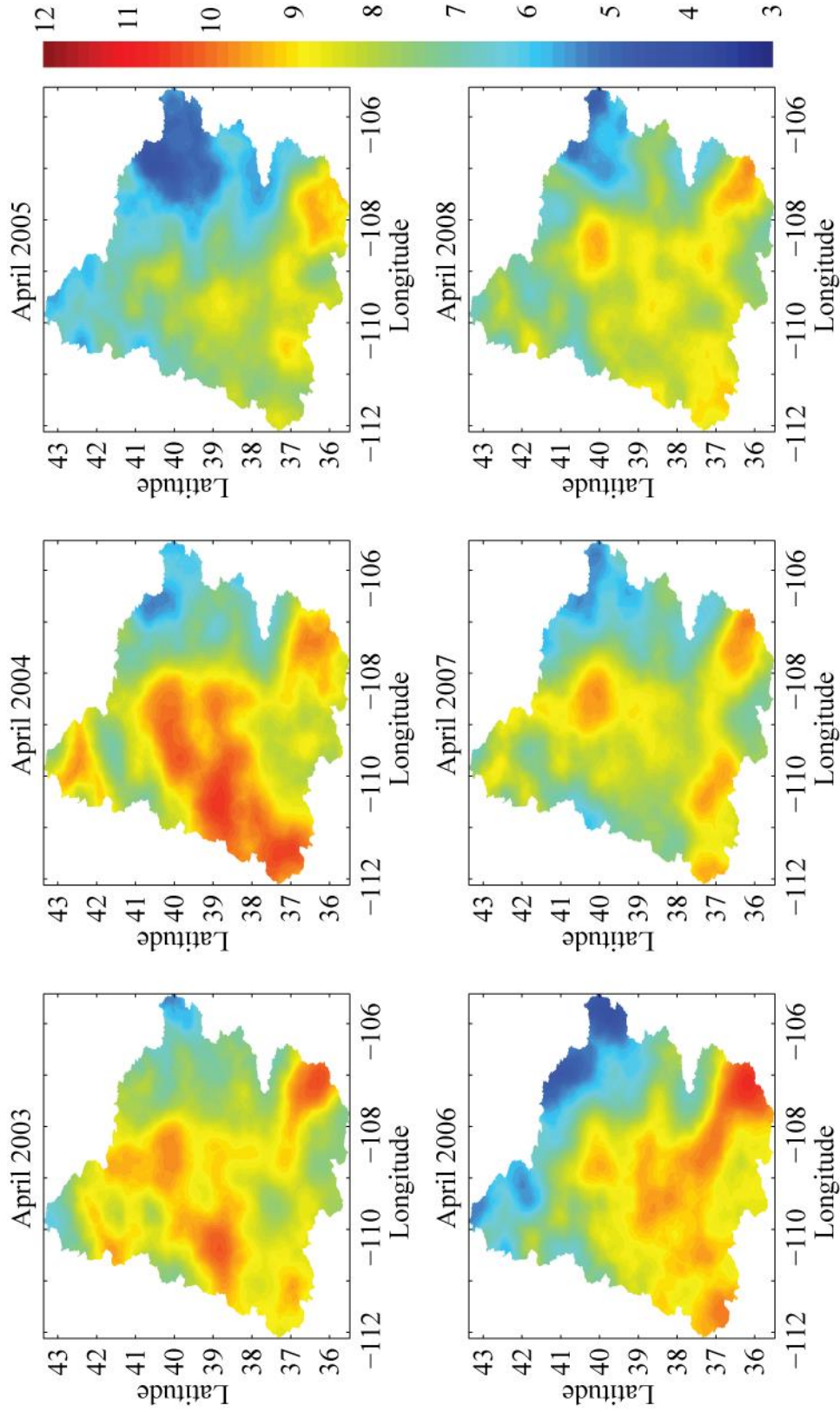


Figure 23. Spatially distributed drought recovery time in the upper Colorado River Basin, starting on April 1st of each year, measured in months.

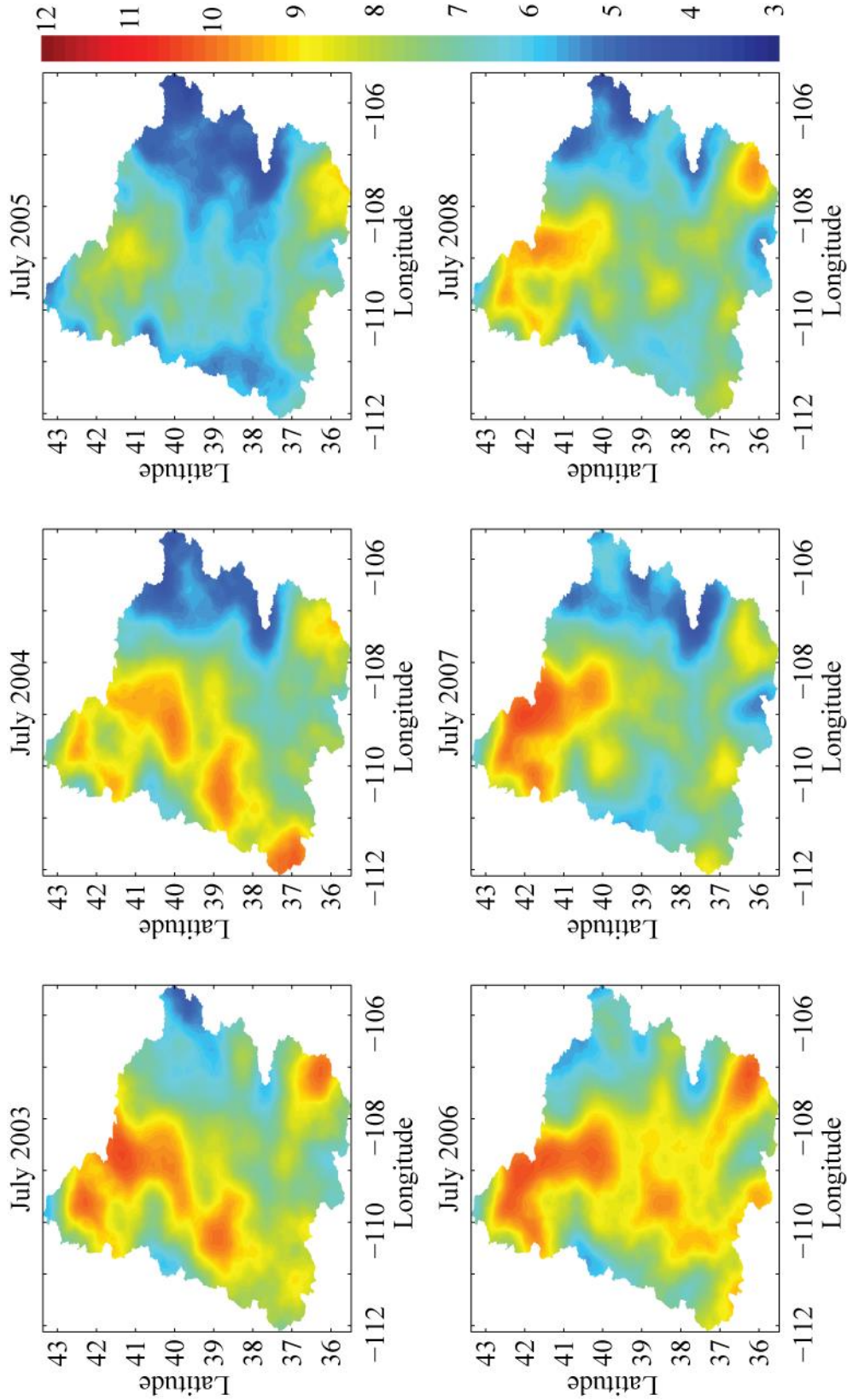


Figure 24. Spatially distributed drought recovery time in the upper Colorado River Basin, starting on July 1st of each year, measured in months.

Further patterns may be observed when comparing April and July forecasts. From Figures 23 and 24, it is clear the July forecasts tend to have shorter drought recovery times in the southern portion of the UCRB, but longer drought recovery times in the northern portion, as compared to the April forecasts. This issue likely has less to do with initial drought status as it does with normal precipitation timing in the central portions of the basin. Although the northern and higher elevation regions receive the vast majority of precipitation during the winter, the southern and interior portions receive a slightly larger portion of precipitation during the summer than in the winter, and therefore some regions have faster recovery times in July than April. An important note here is that this study assumes normal climate conditions over the forecast period, but in the central regions of this basin that assumption is potentially violated. Since much of the summer rains in this basin are due to thunderstorms resulting from moisture emanating from the North American Monsoon (Adams and Andrew, 1997), and moisture over southern portions of the basin and Arizona affecting atmospheric feedbacks (Feng et al., 2013), drought recovery time estimated for the south-central portion of the UCRB in this study may be very conservative. More robust analysis would require a coupled land-atmosphere model, which is outside the scope of this study, as the predominant source of water in the UCRB is the winter and spring westerly storm pattern.

8.5.3 Rate of Information Loss

This manuscript has examined the recovery time from different drought conditions, but to further the analysis, it is important to understand the rate at which drought recovers. Whereas the time for drought recovery is highly dependent on drought

intensity at the initial forecast date, the rate of drought recovery reduces the reliance on drought initial conditions. An example of this point is displayed in Figure 25. In this figure, the RE of the spatially averaged LWS is plotted, with respect to time, and a line is fit to the data points from equation (69). From these results, it is clear that the variance in the rate of drought recovery is relatively constant, as the fitted lines differ only slightly from year to year. In addition, the rate of recovery shows very little variance with respect to start date, as shown in Table 4. All spatially averaged LWS RE functions had an exponent between 0.57 and 0.6, yet spatially distributed calculations ranges from 0.2 to 0.9 (Figure 26). From Figure 26, April forecasts tend to have slightly larger variance in drought recovery rates, but overall the variability seems greater spatially than with respect to initial forecast date. In addition, the distributions of recovery rates from Figure 8 have similar qualities, with an exponent value of around 0.45 having the greatest frequency in each month, and each distribution being positively skewed. This suggests that drought recovery rates from these months have similar spatial patterns, and therefore the specific initial conditions may not have strong influence over the rate of recovery. Further analysis requires a direct comparison of the drought recovery time and rates.

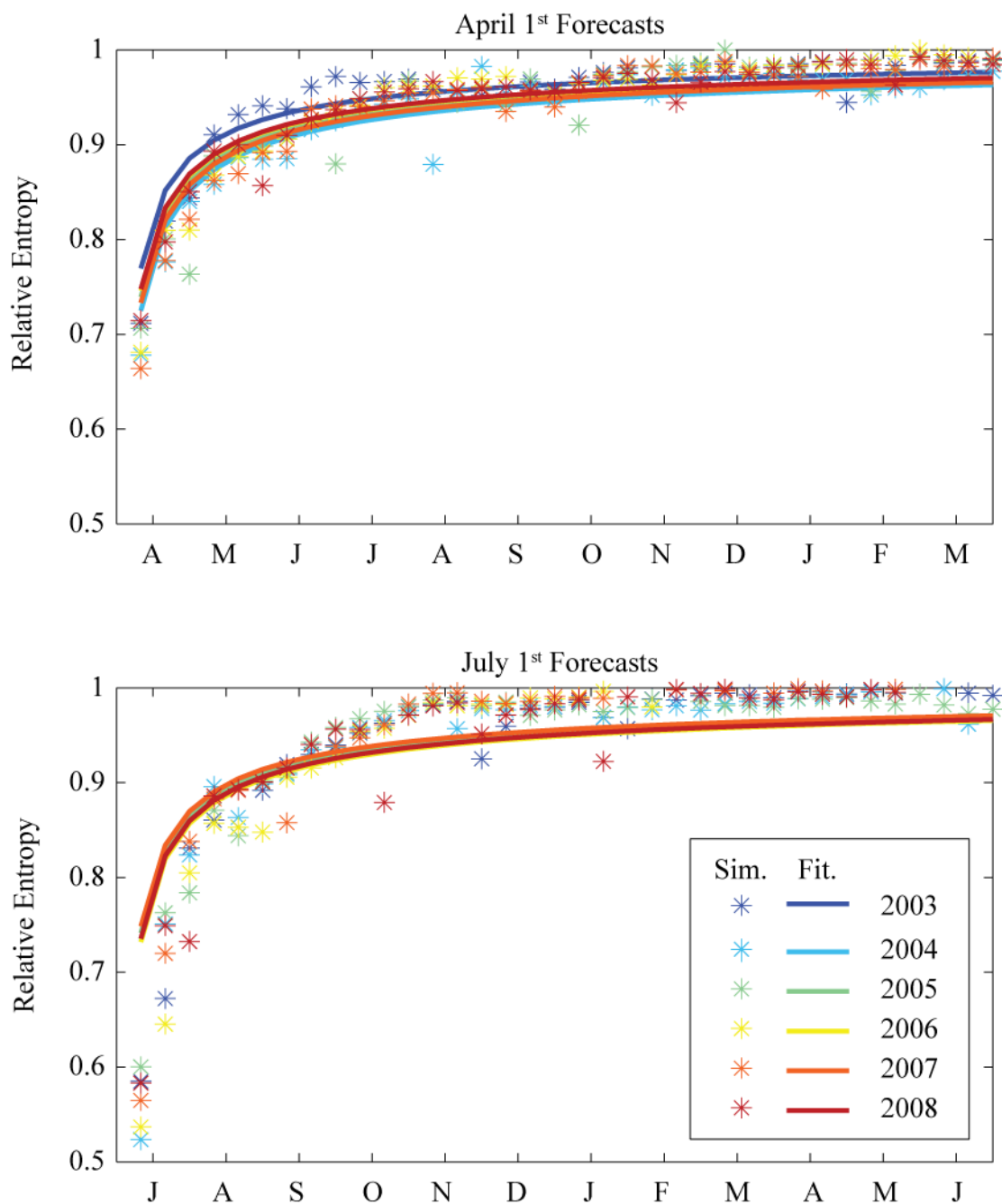


Figure 25. Spatially averaged relative entropies (asterisks), with the fitted (lines) from equation (69), of the land water storage simulations with respect to climatology.

Table 4. Rate of drought recovery from each forecast initialization date, as estimated by the exponent (n) from equation (69).

Year	April 1 st	July 1 st
2003	0.64	0.57
2004	0.56	0.58
2005	0.58	0.59
2006	0.6	0.57
2007	0.57	0.6
2008	0.6	0.58

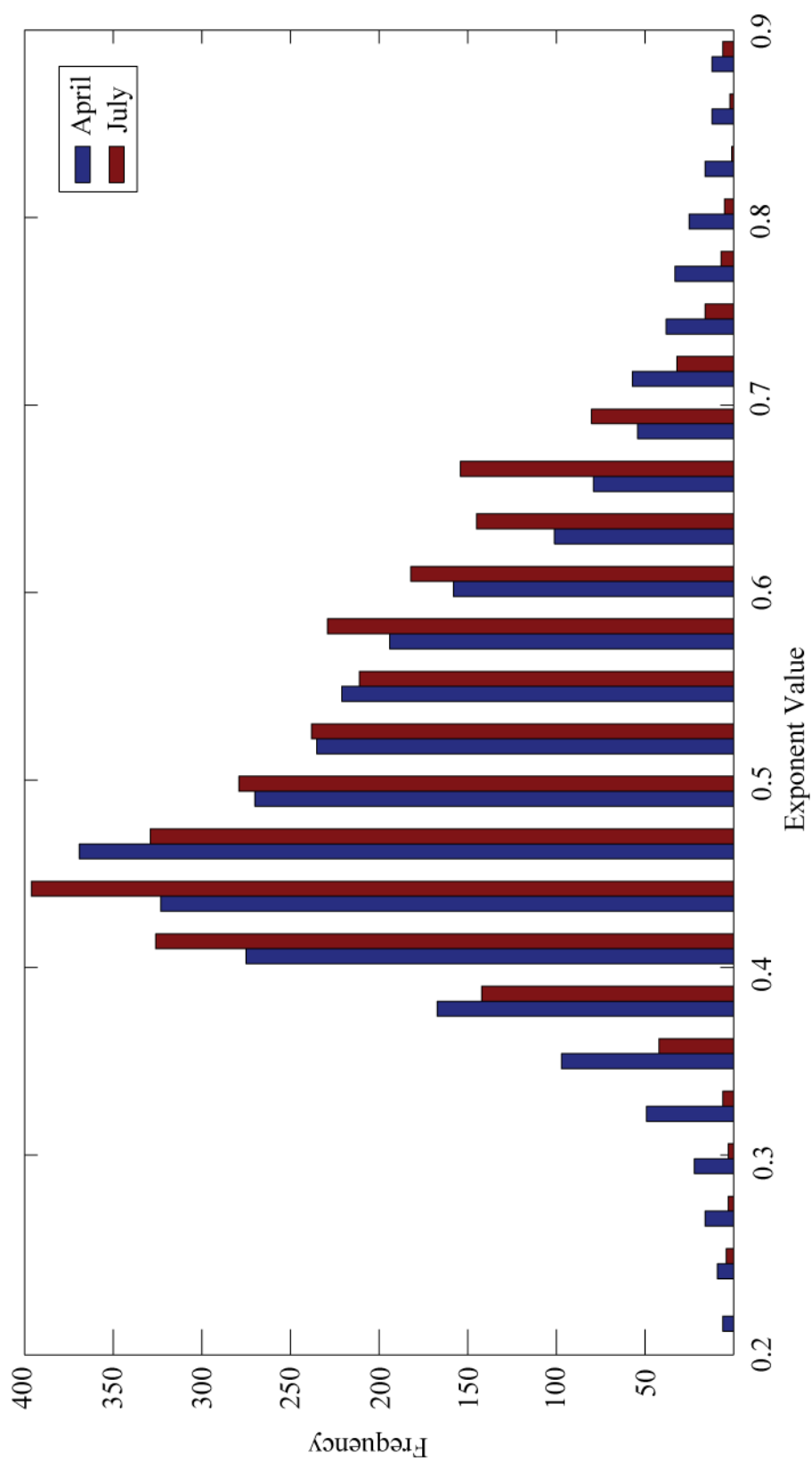


Figure 26. Histogram of the spatially distributed exponent (n) values from equation (69).

Figure 27 presents a comparison of the relationship between drought recovery time and rate for different forecast start dates. In this figure, each combination of drought recovery time (left subplot) and rate (right subplot) between two different start dates (66 unique combinations from the 12 starting dates), for each model grid cell, were plotted. For example, the drought recovery time/rate estimated for the each cell starting in April 1st 2003 is the y component, and the drought recovery time/rate estimated for the each cell starting in April 1st 2004 is the x component, for the first combination in the left subplot. Each subsequent combination of start dates is plotted as well, leading to 31,218 data points from the 66 combinations and 473 model grid cells. By comparing these two subplots, it is clear that the drought recovery rate shows a much more consistent pattern than drought recovery time. This consistent pattern for recovery rate shows a near linear correlation, which suggests that drought recovery rate is largely unaffected by forecast initial drought intensity. Alternatively, the drought recovery time is much more scattered, suggesting that initial drought intensity strongly affect the length of time to drought recovery. Such a finding indicates that the rate of drought recovery is primarily controlled by land surface properties, highlighting the importance of the physical setting of the region affected by drought, and not just the moisture deficit at a given time. Since the drought recovery rate is not significantly changing temporally, as evidenced by Figure 27, yet there is a large range of drought recovery rates experienced in the basin, as shown in Figure 26, it is important to examine the spatial distribution of drought recovery rates to determine drought prone regions.

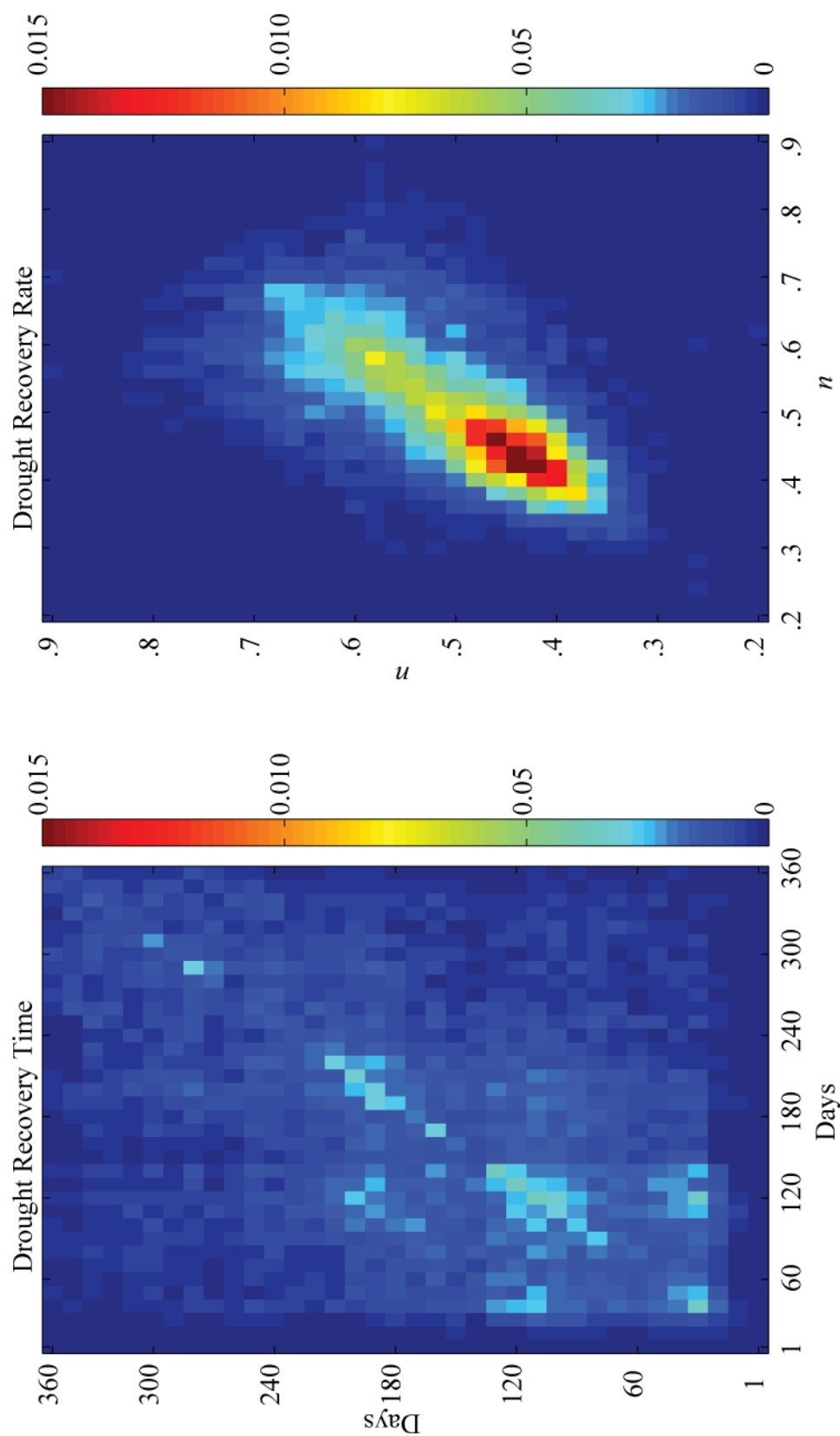


Figure 27. Comparison of the consistency of drought recovery time and rate estimates for each model cell. In these sub-plots, blue indicates rare combinations of time/rate estimates from different start dates, whereas red indicates frequent combinations of time/rate estimates from different start dates. The colorbars to the right of the sub-plots indicate the percentage of estimated time/rate estimate combinations occurring in a given location.

The spatial distribution of drought recovery rates is displayed in Figures 28 and 29. As was expected from Figure 26, these rates are highly variable. In the April forecasts, drought recovery rates tend to be faster in the central to northeastern portions of the basin, and slower in the southern and western portions of the basin. As for July, the fastest drought recovery rates are primarily located in the far eastern portion of the basin, and other portions display a generally slow rate of drought recovery. Starting in both forecast months, the eastern portions of the basin recover from droughts the most rapidly, indicating that this region readily recovers from drought, making this the least drought prone region. Alternatively, the southern and western portions of the basin recover from drought very slowly, indicating that these regions are prone to extended drought conditions. This figure clearly shows that the rate of drought recovery is regional, and therefore is strongly affected by the geographical setting. In this study, the spatial variability in the simulations is caused by differing land surface parameters, indicating that land surface properties (soil types and vegetation cover) strongly affect the ability of a region to recover from drought.

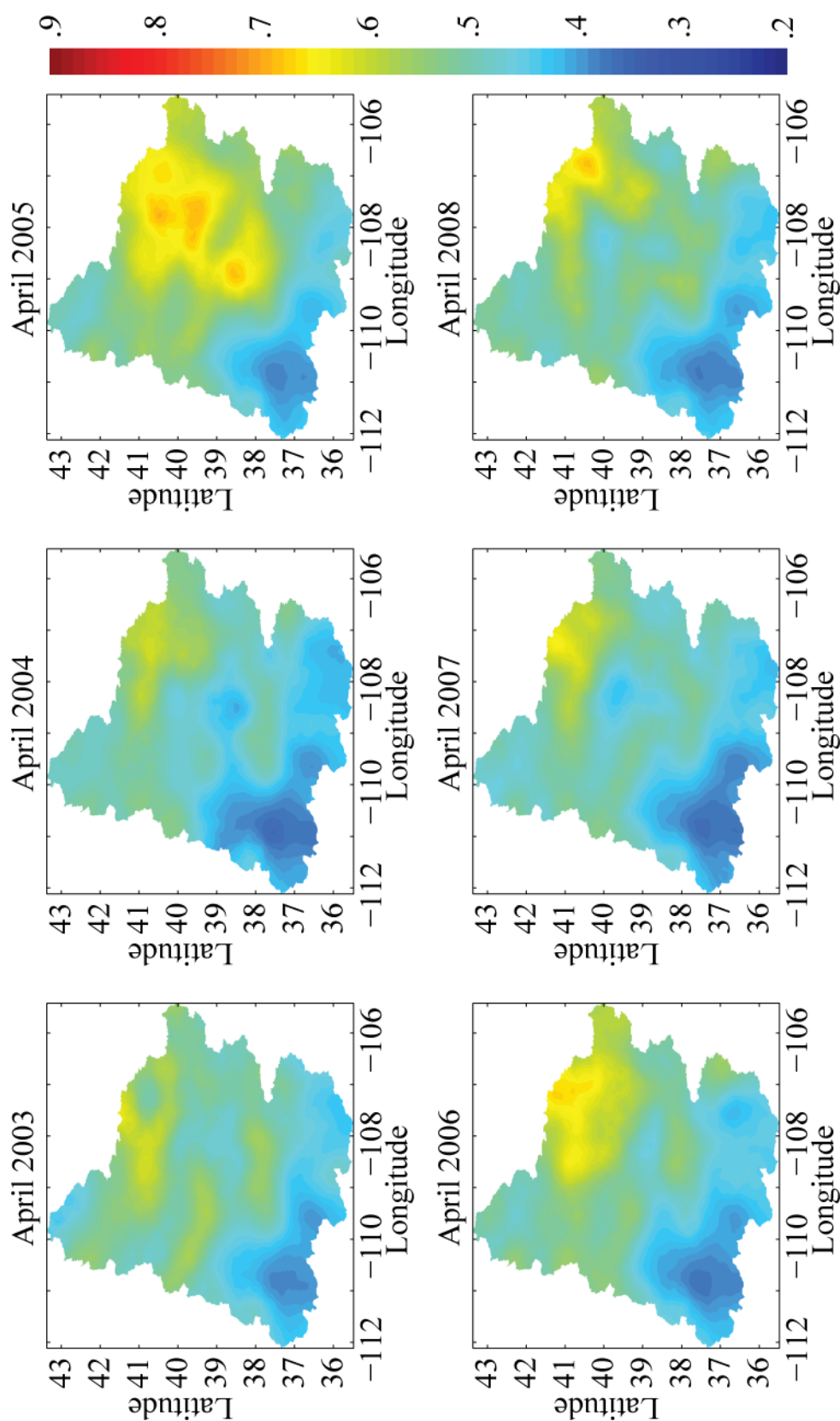


Figure 28. Spatially distributed drought recovery rate in the upper Colorado River Basin, starting on April 1st of each year. Drought recovery rate is estimated as the exponent (n) of equation (69) after being fitted to the relative entropy time series of each grid cell. Larger n values indicate faster recovery rate.

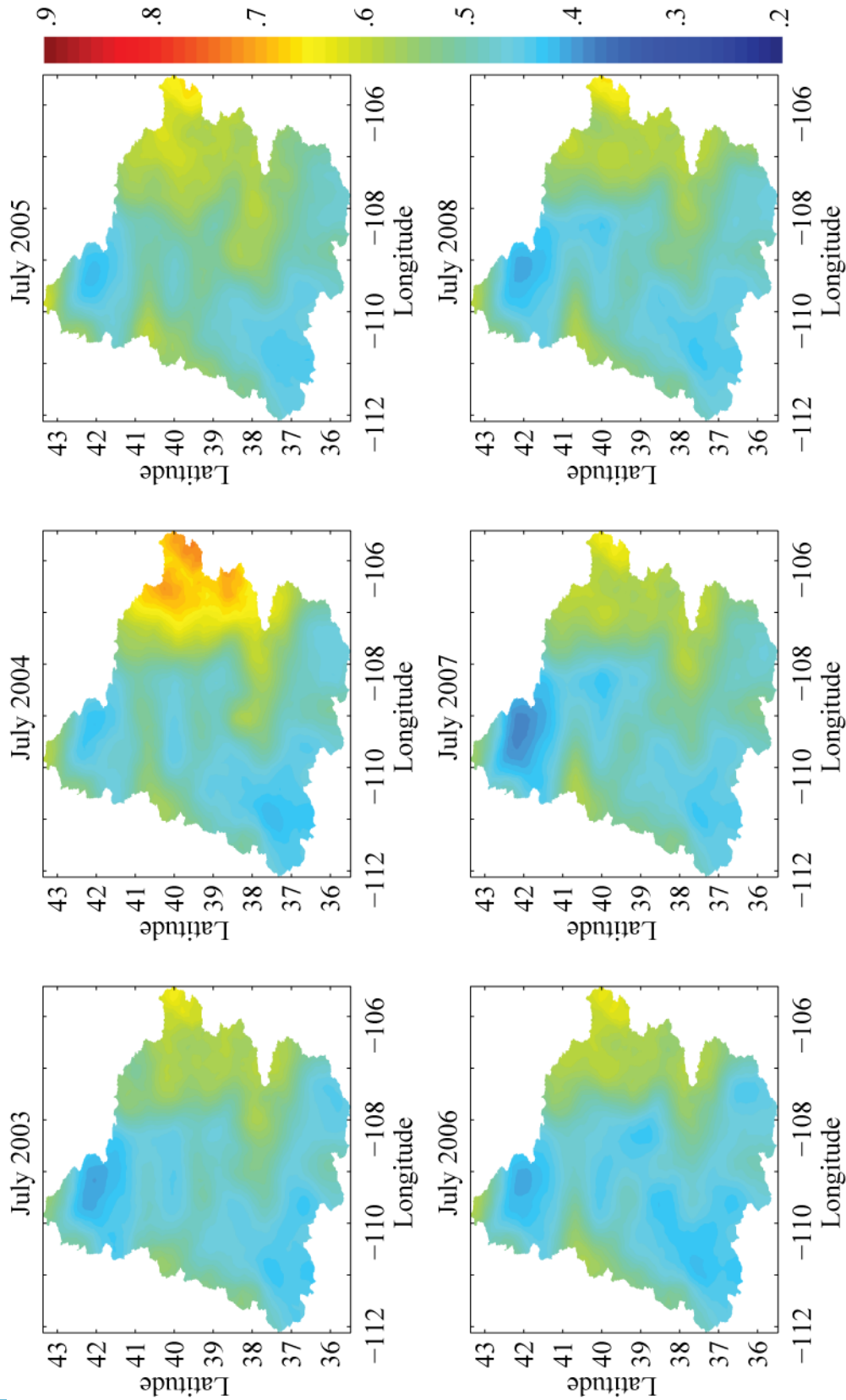


Figure 29. Spatially distributed drought recovery rate in the upper Colorado River Basin, starting on July 1st of each year. Drought recovery rate is estimated as the exponent (n) of equation (69) after being fitted to the relative entropy time series of each grid cell. Larger n values indicate faster recovery rate.

9 Conclusions

9.1 Water Supply Forecasting

This dissertation examined a number of modeling scenarios in attempts to improve the characterization of uncertainty in water supply forecasting. Based on the operational ESP framework, the VIC and NWS models were used to generate probabilistic seasonal streamflow forecasts. Each model had an open loop case, two DA cases (TB only and TB with LST DA), and a model averaging case of all six prior scenarios was performed with PF-SBC. DA was implemented to account for initial condition uncertainty, and PF-SBC is used to account for model uncertainty. Such an experiment is expected to improve the reliability of forecast distributions from ESP, as ESP tends to produce overconfident results.

A first comparison of the modeling scenarios was provided with the relative performance of each during the spin up period. During this time, the NWS models with TB only DA provided the best overall probabilistic prediction, highlighting the effectiveness of the NWS forecasting system, and the sensitivity of passive microwave TB to snowpack states. In addition, the relative accuracy of the models with and without DA over the course of the accumulation and ablation season highlights the temporal characteristics of TB DA performance. In the accumulation season, DA is more effective as the snow tends to be drier, but the open loop models perform better in the ablation season as TB becomes less sensitive to SWE. Relative model weights also suggest that VIC is more efficient in assimilating LST, which was expected as the VIC model

explicitly solves for LST, whereas the SNOW-17 model only estimates average pack temperature.

Forecast reliability was shown to have a strong spatial component in figures 5 and 6. The VIC model produced forecasts with lower skill than the NWS models, with the exception of the San Juan basin, over the UCRB. Both VIC DA cases were capable of improving the forecast reliability in the Green and San Juan basins, suggesting improvements in initial conditions, but produced worse forecasts in the Colorado River headwaters. While the VIC model performed worse with DA in the headwaters, the NWS models improved. This is counterintuitive as the VIC model is expected to be more effective at constructing the correct land surface states, but this result is attributed to the coarser resolution of the VIC model, with respect to the NWS models in this region. In addition, DA in the NWS models provided the least improvement in the headwaters region, in comparison to other regions, which highlights the difficulty of utilizing remotely sensed information to reconstruct land surface states in regions of thick forest cover. This is unfortunate as forest thickness is generally correlated with high precipitation quantity, and thus more important from a water supply perspective. Since the most densely forested region in a basin will tend to be the most important from a water supply perspective, it is imperative that snow observations and DA methods become more efficient in these regions, which is a great challenge for the hydrological research community.

A final conclusion from the water supply forecasting experiment relates to the performance of different portions of the forecast distributions. In general, the

improvements were stronger in the tails of the distributions than in the central portions, with the exception of the PF-SBC case in the Green and San Juan basins. Since DA improved the tails of the distribution in nearly all basins/cases, yet did not consistently improve total reliability, initial condition uncertainty is shown to have stronger control over the reliability of predicting low probability events. Alternatively, PF-SBC improved the overall reliability in nearly all basins, suggesting that model error strongly controls uncertainty in the forecast distribution mode. Overall, this highlights that both initial condition and model error are important factors in seasonal prediction.

The presented methodology shows promise for improving the reliability of seasonal forecasting, by accounting for all sources of forecast uncertainty, but the results from the application here clearly show a need for improvement, as none of the forecasts can be considered truly reliable according to the definition in Chapter 6. It is suggested that these improvements will come from advancements in land surface DA and increasing the number of models to more effectively manage forecast uncertainty. By developing more effective DA systems, the proposed framework will create more accurate and reliable predictions of initial land surface conditions, which were shown to have significant contribution to probabilistic streamflow forecasting. In addition, the inclusion of a greater variety of model structures will more effectively manage model errors, thus leading to more reliable forecast error quantification. A further advancement in quantifying forecast uncertainty may also come from the use of state-parameter estimation within the DA framework (DeChant and Moradkhani, 2012), which is becoming increasingly efficient (Moradkhani et al., 2012), and therefore more applicable

to the spatial extents considered here. Through the application of state-parameter estimation, the estimation of parameter uncertainty may be separated from model structural uncertainty, leading to a more complete accounting of uncertainty.

9.2 Probabilistic Verification of Binary Outcomes

This dissertation used a synthetic experiment to show that conventional methods for assessing the reliability of probabilistic hydrological event forecasts are flawed (i.e. BS and reliability diagram). These methods provide generally useful metrics for comparing probabilistic forecasts, but do not adequately diagnose reliability, nor do they achieve optimal use of observed information. Three important drawbacks to these methods are highlighted: First, available methods for assessment of reliability are approximate, which is evidenced by the dependencies on the uniformity of forecasted probabilities. Second, conventional verification methods cannot distinguish reliable from unreliable forecasts in a statistically accurate way. Last, the requisite grouping process leads to significant loss of information, thus decreasing one's ability to separate reliable and unreliable forecasts. For these reasons, this dissertation suggests the use of a hypothesis test, via the Poisson-Binomial Distribution, in attempts to reject all unreliable forecasts. In the event that the hypothesis of reliability is not rejected for multiple forecasts, the sharpest forecast will be preferred, which is a framework followed in the drought forecasting experiment.

9.3 Drought Forecasting

Results of from the drought forecasting experiment suggest that the combination of DA and multi-modeling (PF-SBC) for seasonal drought forecasting is more reliable than other forecast frameworks presented. Both the deterministic metrics and

probabilistic metrics agree that the PF-SBC case outperforms all others examined here. This result shows the benefits of using DA for land surface state initialization, and the combination of multiple models to constrain model uncertainties. By approaching the problem of drought forecasting through an increasingly comprehensive uncertainty accounting framework, a more accurate and reliable seasonal drought forecasting system was achieved. All of the methods for partial accounting of uncertainty were unable to satisfy the primary reliability condition (PB-CDF), indicating an incomplete description of forecast uncertainty.

Although the deterministic and probabilistic verification measures agree that the PF-SBC method performs the best of all seven forecast methodologies, they provide conflicting evidence about the relative performance of the VIC and NWS based forecasts. In the deterministic measures, the VIC DA cases appear to outperform their NWS counterparts, yet the probabilistic metrics favor DA with the NWS models. Similar to previous studies in comparing deterministic and probabilistic verification (DeChant and Moradkhani, 2012), it is apparent here that a deterministic metric is not necessarily indicative of the probabilistic performance, and therefore both must be examined. In the forecasting scenario presented here, there is a large amount of uncertainty for all forecasting frameworks, and therefore this study suggests that probabilistic metrics should be given higher weight than deterministic measures when comparing model performance. This suggestion is made due to the knowledge that the ultimate goal of forecasting drought is to aid the management of drought risk.

A final conclusion from the drought forecasting experiment is the importance of using an exact solution to the probabilistic verification methodology, to attempt to reject the hypothesis of reliability. Results from this experiment generally agree with the conclusions from the synthetic analysis presented in section 9.2, suggesting that the PB-CDF improves upon conventional reliability metrics, as it is an exact model of the drought verification setting, thus maximizing the information extracted from the observation through a single bin analysis. Although the single bin PB-CDF maximizes one's ability to reject unreliable forecasts, it may miss information related to the over/underconfidence of the forecast, as it directly measures bias. An examination of the reliability diagram suggests that a multiple bin hypothesis test via the PB-CDF may provide additional information for the hypothesis test. With such a framework, this study was able to show that all forecasting systems were unreliable, but that the PF-SBC case is closer to meeting the reliability requirements than all other methods. Unfortunately, the drought forecasting system developed in this dissertation is unable to achieve reliability based on such analysis. This finding suggests that further advancements to the forecasting framework are required, which may come from improved DA, additional models and accounting of parameter uncertainty, as suggested in section 9.1.

9.4 Assessing the Need for a New Drought Index

This dissertation examined the problems associated with standardized indices, and proposes a general framework for advancing drought monitoring. The specific case of poorly representing drought in a basin of extreme spatial water storage distribution was examined. From this analysis, it was shown that standardized indices may provide

misleading analysis of basin-wide drought. In addition, standardized indices were found to have obstacles relating to the use of DA methods, which reduces the viability of standardized indices as the use of DA methods are becoming increasingly common. Beyond the errors related to spatially distributed water storage and DA, it must also be noted that standardized indices will be problematic when analyzing sufficiently dynamic regions, in terms of both hydrological conditions and socio-environmental demand. Based on this information, this dissertation suggests that a movement towards increasingly physically-based drought indices is necessary.

The MSMI was proposed as it is a physically-based drought index. This was shown to have some beneficial attributes, but is still incomplete. Primarily, this index is difficult to interpret. Without any clear guidelines for differentiating between drought and non-drought states, this index still relies on comparison of between multiple dates. While this index may be difficult to interpret, it may be a reasonable starting point for future analysis. If the necessary MSMI required for environmental, agricultural or societal demand can be determined for a region, this index can directly describe potential deficiencies. This may be as simple as accounting for water requirements to support the different types of vegetation cover in a region. In order to move forward, this study proposes the MSMI as a reasonable starting point, and suggests that avenues for providing context to MSMI values must be pursued, with the most immediately apparent avenue being relating MSMI with water demand.

9.5 Sensitivity of Droughts to Initial Conditions

The final experiment of this dissertation examined drought recovery forecasting, under the assumption of normal climate following a drought event. Rather than comparing the relative influence of initial conditions and forcing, as performed in previous studies, the presented results examined the amount of time for the forecast to become insensitive to initial conditions, and the rate at which a forecast loses sensitivity to the initial conditions. Such an examination is a proxy for estimating lead time, and rate, of drought recovery under the assumption that normal climate conditions will occur. Through the quantification of drought recovery time and rate, this experiment simultaneously provides insight into the expected behavior of the UCRB during the recession period of droughts, and the importance of precisely estimating initial conditions in forecasts of up to one year. Not surprisingly, this study found that drought recovery time was related to drought intensity, with increasing intensity requiring greater recovery time. Further, drought recovery time was found to be greater for forecasts originating in April than July, but the difference is smaller than anticipated. In some years, the basin-wide recovery from forecasts originating in April and July were similar, whereas other years displayed a difference of up to three months. In general, one would assume that the April forecasts would have greater influence on drought recovery time, due to the presence of snow water storage, but this is not necessarily the case.

Results from this dissertation clearly contrast with previous work, which is related to the differing perspectives on the importance of drought initial conditions. Due to previous studies focusing on lead time at which forcing becomes the dominant factor, in relation to

initial conditions, and this study seeking to quantify the lead time at which initial states significantly impact the forecast, sensitivity to initial conditions was found to be greater here than in other studies. A maximum of six months of sensitivity was generally found in previous studies, but results here suggest that a forecast may be sensitive to initial conditions beyond one year for select locations, but up to nine months at the basin scale. Further, this dissertation shows that forecasts are sensitive to soil moisture at a minimum of three months, but previous studies generally suggested a maximum of one to two months. Overall these results suggest that initial conditions are important in forecasts of greater lead time than previously thought, and therefore estimation of initial conditions should be considered for forecasts of extended length. This finding suggests that data assimilation systems may have some benefit to forecasts at even a year lead time, although the importance of improved initial conditions will be less than forcing improvements beyond six months, as indicated by previous studies.

Drought recovery times estimated in this study should be treated as conservative. The results presented are based on initial conditions with large uncertainty. Through the data assimilation framework used here, initial conditions remain quite uncertain, and therefore approaches climatology faster than precise estimates would. Assuming that data assimilation science progresses to decreased initial condition uncertainty, the lead time at which a forecast becomes insensitive to initial conditions will increase. This knowledge further motivates the use of data assimilation for initializing forecasts of extended length as it will reduce the uncertainty in the initial condition, thus increasing the information added to the forecast from the initial conditions.

A final conclusion from this dissertation is the difference between drought recovery time and rate of recovery. Since the drought recovery time is highly dependent on initial conditions, this value varies strongly between years and forecast initial date. Alternatively, the drought recovery rate appears to be more related to the specific location than drought intensity. This is evidenced by the small temporal variance of recovery rates for spatially averaged LWS, but large variance for spatially distributed recovery rates (see Figure 26). This conclusion gives some insight into the geographic and climatic settings which drought is most persistent. In the mountainous northeastern region of the UCRB, droughts recover quickly, yet the low lying southern and western regions recover slowly from drought. Drought mitigation within the drier regions of the UCRB is therefore a much greater challenge than in the mountainous regions. Fortunately, the total water storage in this basin is dominated by snow in the mountainous regions, and therefore total basin water storage is expected to recover at a rate faster than that of the dry interior portions.

References

- Adams, D. K., & Comrie, A. C. (1997). The North American Monsoon. *Bulletin of the American Meteorological Society*, 78(10), 2197-2213.
- Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1).
- Anderson, E. (1973). National weather service river forecast system— snow accumulation and ablation model.
- Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29, 872-873-886.
- Annan, J. D., & Hargreaves, J. C. (2010). Reliability of the CMIP3 ensemble. *Geophysical Research Letters*, 37(2), L02703.
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42-47.
- Arshad, S., Morid, S., Mobasheri, M. R., Alikhani, M. A., & Arshad, S. (2012). Monitoring and forecasting drought impact on dryland farming areas. *International Journal of Climatology*.
- Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2), 174-188.

- Arumugam, S., Wood, A., Rajagopalan, B., & Schaake, J. (2014). Improving Forecasts for Water Management. *Eos, Transactions American Geophysical Union*, 95(1), 3-3.
- Aziz, O. A., Tootle, G. A., Gray, S. T., & Piechota, T. C. (2010). Identification of Pacific Ocean sea surface temperature influences of Upper Colorado River Basin snowpack. *Water Resources Research*, 46(7).
- Baier, W. (1969). Concepts of soil moisture availability and their effect on soil moisture estimates from a meteorological budget. *Agricultural Meteorology*, 6(3), 165-178.
- Beck, M. B. (1987). Water quality modeling: a review of the analysis of uncertainty. *Water Resources Research*, 23(8), 1393-1442.
- Beven, K. (1989). Changing ideas in hydrology — The case of physically-based models. *Journal of Hydrology*, 105(1-2), 157-172.
- Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6(3), 279-298.
- Bloemhof-Ruwaard, J. M., Van Beek, P., Hordijk, L., & Van Wassenhove, L. N. (1995). Interactions between operational research and environmental management. *European Journal of Operational Research*, 85(2), 229-243.
- Bohn, T. J., Sonessa, M. Y., & Lettenmaier, D. P. (2010). Seasonal Hydrologic Forecasting: Do Multimodel Ensemble Averages Always Yield Improvements in Forecast Skill? *Journal of Hydrometeorology*, 11(6), 1358-1372.

- Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12), 3663-3674.
- Bras, R. L. (1990). *Hydrology: an Introduction to Hydrologic Science* (pp. 188-241). Workingham: Addison-Wesley.
- Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78(1), 1-3.
- Bröcker, J., & Smith, L. A. (2007). Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting*, 22(3), 651-661.
- Burnash, R. J. C., Ferral, R. L., & McGuire, R. A. (1973). A generalized streamflow simulation system, conceptual modeling for digital computers.
- Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, 298(1-4), 242-266.
- Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., & Slooten, L. J. (2005). Inverse problem in hydrogeology. *Hydrogeology Journal*, 13(1), 206-222.
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., . . . Pizer, W. (2001). Ecological forecasts: an emerging imperative. *Science*, 293(5530), 657-660.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., . . . Uddstrom, M. J. (2008a). Hydrological data assimilation with the ensemble

- Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, 31, 1309.
- Clark, M. P., & Slater, A. G. (2006). Probabilistic Quantitative Precipitation Estimation in Complex Terrain. *Journal of Hydrometeorology*, 7(1), 3-22.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., . . . Hay, L. E. (2008b). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44(12).
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559-583.
- Cook, E. R., Seager, R., Cane, M. A., & Stahle, D. W. (2007). North American drought: reconstructions, causes, and consequences. *Earth-Science Reviews*, 81(1), 93-134.
- Dai, A. (2011). Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change*, 2(1), 45-65.
- Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of applied meteorology*, 33(2), 140-158.
- Day, G. (1985). Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management*, 111(2), 157-170.
- De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., & Pauwels, V. (2012). Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate

- Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*, 48(1).
- Dechant, C., & Moradkhani, H. (2011a). Radiance data assimilation for operational snow and streamflow forecasting. *Advances in Water Resources*, 34(3), 351-364.
- DeChant, C. M., & Moradkhani, H. (2011b). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrol. Earth Syst. Sci.*, 15(11), 3399-3410.
- DeChant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48(4), W04518.
- DeChant, C.M. and H. Moradkhani (2014), Hydrologic Prediction and Uncertainty Quantification, Handbook of Engineering Hydrology, Modeling, Climate Change and Variability, CRC press, Taylor & Francis Group, PP 387-414
- Demargne, J., Wu, L., Regonda, S., Brown, J., Lee, H., He, M., . . . Fresch, M. (2013). The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bulletin of the American Meteorological Society*.
- Douc, R., & Cappe, O. (2005). Comparison of resampling schemes for particle filtering. Paper presented at the Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on.
- Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5), 1371-1386.

- Durand, M., & Margulis, S. A. (2006). Feasibility Test of Multifrequency Radiometric Data Assimilation to Estimate Snow Water Equivalent. *Journal of Hydrometeorology*, 443.
- Feller, W. (1945). On the Normal Approximation to the Binomial Distribution. *The Annals of Mathematical Statistics*, 16(4), 319-329. doi: 10.2307/2236142
- Feng, X., Bosilovich, M., Houser, P., & Chern, J. D. (2013). Impact of land surface conditions on 2004 North American monsoon in GCM experiments. *Journal of Geophysical Research: Atmospheres*.
- Franz, K. J., Hartmann, H. C., Sorooshian, S., & Bales, R. (2003). Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. *Journal of Hydrometeorology*, 4(6), 1105-1118.
- Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T. J., Su, F., . . . Wood, E. F. (2010). Water budget record from Variable Infiltration Capacity (VIC) model. *Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records*.
- Garen, D. C. (1992). Improved techniques in regression-based streamflow volume forecasting. *Journal of Water Resources Planning and Management*, 118(6), 654-670.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 243-268.

- Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), 359-378.
- Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2), 107-113.
- Grantz, K., Rajagopalan, B., Clark, M., & Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, 41(10), W10410.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1), 80-91.
- Hodges, J. L., Jr., & Le Cam, L. (1960). The Poisson Approximation to the Poisson Binomial Distribution. *The Annals of Mathematical Statistics*, 31(3), 737-740.
- Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Computational Statistics & Data Analysis*, 59(0), 41-51.
- Hsu, K. I., Moradkhani, H., & Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, 45(12).
- Hunt, E. D., Hubbard, K. G., Wilhite, D. A., Arkebauer, T. J., & Dutcher, A. L. (2009). The development and evaluation of a soil moisture index. *International Journal of Climatology*, 29(5), 747-759.
- Hutter, K., & Jöhnk, K. (2004). *Continuum methods of physical modeling: continuum mechanics, dimensional analysis, turbulence*. Springer.

- Hwang, Y., & Carbone, G. J. (2009). Ensemble Forecasts of Drought Indices Using a Conditional Residual Resampling Technique. *Journal of Applied Meteorology and Climatology*, 48(7), 1289-1301.
- Jaeger, W.K., A. Platinga, H. Chang, J. McDonnell H. Moradkhani, D. Hulse, R. Hagerty, et al. (2013), Toward a formal definition of water scarcity in natural-human systems, *Water Resources Research*, vol. 49, 1–12, doi:10.1002/wrcr.20249.
- Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8), 2637-2649.
- Jayawardena, A. W., & Lai, F. (1994). Analysis and prediction of chaos in rainfall and stream flow time series. *Journal of Hydrology*, 153(1–4), 23-52.
- Jordan, R. (1991). A one-dimensional temperature model for a snow cover: Technical documentation for SN THERM. 89: COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., . . . Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3), 437-471.
- Katz, R. W., & Murphy, A. H. (1997). *Economic value of weather and climate forecasts*: Cambridge University Press.
- Kim, T.-W., & Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering*, 8(6), 319-328.

- Kimball, J. S., Running, S. W., & Nemani, R. (1997). An improved method for estimating surface humidity from daily minimum temperature. *Agricultural and Forest Meteorology*, 85(1), 87-98.
- Kuczera, G., Kavetski, D., Franks, S., & Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331(1-2), 161-177.
- Laio, F., & Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrol. Earth Syst. Sci.*, 11(4), 1267-1277.
- Leisenring, M., & Moradkhani, H. (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25(2), 253-270.
- Leisenring, M., and H. Moradkhani (2012), Analyzing the Uncertainty of Suspended Sediment Load Prediction Using Sequential Monte Carlo Methods, *Journal of Hydrology*, 468-469, p268-282, 10.1016/j.jhydrol.2012.08.049.
- Levine, J. M., & D'Antonio, C. M. (2003). Forecasting biological invasions with increasing international trade. *Conservation Biology*, 17(1), 322-326.
- Li, D., Durand, M., & Margulis, S. A. (2012). Potential for hydrologic characterization of deep mountain snowpack via passive microwave remote sensing in the Kern River basin, Sierra Nevada, USA. *Remote Sensing of Environment*, 125, 34-48.
- Li, H., Luo, L., Wood, E. F., & Schaake, J. (2009). The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research: Atmospheres*, 114(D4), D04114.

- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7), 14415.
- Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43(7).
- Liu, Y., A.H. Weerts, M. Clark, H.J. Hendricks Franssen, S. Kumar, H. Moradkhani, D.J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec, P. Restrepo, (2012), Toward Advancing Data Assimilation in Operational Hydrologic Forecasting and Water Resources Management: Current Status, Challenges, and Emerging Opportunities, *Hydrology and Earth System Science*, 16, 3863-3887.
- Madadgar, S., Moradkhani, H., & Garen, D. (2012). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*.
- Madadgar, S., & Moradkhani, H. (2013). A Bayesian Framework for Probabilistic Seasonal Drought Forecasting. *Journal of Hydrometeorology*, 14(6), 1685-1705.
- Madadgar, S., and H. Moradkhani (2014), Spatio-temporal Drought Forecasting within Bayesian Networks, *Journal of Hydrology*, 512, 134-146, DOI: 10.1016/j.jhydrol.2014.02.039.
- Mahanama, S., Livneh, B., Koster, R., Lettenmaier, D., & Reichle, R. (2011). Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States. *Journal of Hydrometeorology*, 13(1), 189-203.

- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, 25(8), 891-909.
- Margulis, S. A., McLaughlin, D., Entekhabi, D., & Dunne, S. (2002). Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water resources research*, 38(12), 35-31.
- Maurer, E. P., Wood, A. W., Adams, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *Journal of Climate*, 15, 3237.
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales*, American Meteorological Society Boston, MA, Proceedings of the 8th Conference on Applied Climatology.
- Miller, G. A. (1955). Note on the bias of information estimates. *Information theory in psychology: Problems and methods*, 2, 95-100.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity Is Dead: Whither Water Management? *Science*, 319(5863), 573-574.
- Mishra, A. K., & Singh, V. P. (2011). Drought modeling—A review. *Journal of Hydrology*, 403(1), 157-175.

- Mo, K. C., Shukla, S., Lettenmaier, D. P., & Chen, L. C. (2012). Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophysical Research Letters*, 39(23).
- Montanari, A., & Grossi, G. (2008). Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resources Research*, 44(12), W00B08.
- Moradkhani, H., DeChant, C. M., & Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, 48(12).
- Moradkhani, H., Hsu, K., Hong, Y., & Sorooshian, S. (2006). Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting. *Geophysical Research Letters*, 33(12), L12401.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005a). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2), 135-147.
- Moradkhani, H., Hsu, K.-L., Gupta, H., & Sorooshian, S. (2005b). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5), W05012.
- Moradkhani, H., & Meier, M. (2010). Long-lead water supply forecast using large-scale climate predictors and independent component analysis. *Journal of Hydrologic Engineering*, 15(10), 744-762.

- Moradkhani, H., & Meskele, T. T. (2010). Probabilistic assessment of the satellite rainfall retrieval error translation to hydrologic response. *Satellite Rainfall Applications for Surface Hydrology* (pp. 229-242): Springer.
- Murphy, A. H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, 12(4), 595-600.
- Najafi, M.R., H. Moradkhani, I. Jung (2011), Assessing the Uncertainties of Hydrologic Model Selection in Climate Change Impact Studies, *Hydrologic Processes*, 25(18), 2814-2826.
- Najafi, M. R., Moradkhani, H., & Piechota, T. C. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *Journal of Hydrology*, 442, 105-116.
- Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2), 101-121.
- Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922–2002. *Journal of Hydrometeorology*, 5(5), 896-909.
- Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., & Gonçalves, L. G. G. d. (2012). On the sources of hydrological prediction uncertainty in the Amazon. *Hydrology and Earth System Sciences*, 16(9), 3127-3137.
- Pal, S. (2009). A note on a conjectured sharpness principle for probabilistic forecasting with calibration. *Biometrika*, 96(4), 1019-1023.

- Palmer, W. C. (1965). *Meteorological drought*: US Department of Commerce, Weather Bureau Washington, DC, USA.
- Pan, M., Yuan, X., & Wood, E. F. (2013). A probabilistic framework for assessing drought recovery. *Geophysical Research Letters*, 40(14), 3637-3642.
- Parrish, M. A., Moradkhani, H., & DeChant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*, 48(3), W03519.
- Piechota, T. C., Dracup, J. A., & Fovell, R. G. (1997). Western US streamflow and atmospheric circulation patterns during El Niño-Southern Oscillation. *Journal of Hydrology*, 201(1), 249-271.
- Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., . . . Lawford, R. (2013). Toward Global Drought Early Warning Capability: Expanding International Cooperation for the Development of a Framework for Monitoring and Forecasting. *Bulletin of the American Meteorological Society*, 94(6), 776-785.
- Pullianen, J., Tigerstedt, K., Huining, W., Halliikainen, M., Matzler, C., & Wiesmann, A. (1998). *Retrieval of geophysical parameters with integrated modeling of land surfaces and atmosphere models/inversion algorithms, summary report*. [Helsinki, Finland]: Helsinki University of Technology, Laboratory of Space Technology.
- Rabier, F., Klinker, E., Courtier, P., & Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions. *Quarterly Journal of the Royal Meteorological Society*, 122(529), 121-150.

- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5), 1155-1174.
- Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E. (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, 42(9).
- Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1), 103-114.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5), W05521.
- Risley, J. C., Gannett, M. W., Lea, J. K., & Roehl, E. A. (2005). *An analysis of statistical methods for seasonal flow forecasting in the Upper Klamath River Basin of Oregon and California*: US Department of the Interior, US Geological Survey.
- Rosati, A., Miyakoda, K., & Gudgel, R. (1997). The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Monthly weather review*, 125(5), 754-772.
- Ross, T., & Lott, N. (2003). *A climatology of 1980-2003 extreme weather and climate events*: US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite Data and Information Service, National Climatic Data Center.

- Schaake, J. C., Hamill, T. M., Buizza, R., & Clark, M. (2007). HEPEX: the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society*, 88(10), 1541-1547.
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10).
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379-423. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Shukla, J. (1998). Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, 282(5389), 728-731.
- Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrol. Earth Syst. Sci.*, 15(11), 3529-3538.
- Shukla, S., Sheffield, J., Wood, E. F., & Lettenmaier, D. P. (2013). On the sources of global land surface hydrologic predictability. *Hydrol. Earth Syst. Sci.*, 17(7), 2781-2796.
- Stephenson, D. B., Coelho, C. A. S., & Jolliffe, I. T. (2008). Two Extra Components in the Brier Score Decomposition. *Weather and Forecasting*, 23(4), 752-757.
- Stern, P. C., & Easterling, W. E. (1999). *Making climate forecasts matter*: National Academies Press.

- Thornton, P. E., & Running, S. W. (1999). An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology*, 93(4), 211-228.
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., & Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45(12).
- Twedt, T. M., Schaake Jr, J. C., & Peck, E. L. (1977). *National Weather Service extended streamflow prediction [USA]*.
- U.S. Department of the Interior (2005), Natural flow and salt computation methods, Bur. of Reclam., Salt Lake City, Utah.
- Verdon-Kidd, D. C., & Kiem, A. S. (2010). Quantifying Drought Risk in a Nonstationary Climate. *Journal of Hydrometeorology*, 11(4), 1019-1031.
- Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *Geoscience and Remote Sensing, IEEE Transactions on*, 34(4), 892-905.
- Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. L. (2004). Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing*, 25(1), 261-274.
- Werner, K., Averyt, K., & Owen, G. (2013). River Forecast Application for Water Management: Oil and Water? *Weather, Climate, and Society*, 5(3), 244-253.

- Wiesmann, A., & Mätzler, C. (1999). Microwave emission model of layered snowpacks. *Remote Sensing of Environment*, 70(3), 307-316.
- Wigneron, J. P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M. J., Richaume, P., . . . Calvet, J. C. (2007). L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, 107(4), 639-655.
- Wilhite, D. A. (2000). Drought as a natural hazard: concepts and definitions. *Drought, a global assessment*, 1, 3-18.
- Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences: an introduction* (2nd ed.). Oxford: Academic.
- Wolfe, D. A., & Hollander, M. (1973). *Nonparametric statistical methods*. John Wiley New York.
- Wood, A., & Werner, K. (2011). Development of a seasonal climate and streamflow forecasting testbed for the Colorado River Basin. *Development*, 3, 6.
- Wood, A. W., & Lettenmaier, D. P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, 35(14), L14401.
- Wood, A. W., & Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, 9(1), 132-148.
- Yossef, N. C., Winsemius, H., Weerts, A., van Beek, R., & Bierkens, M. F. P. (2013). Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resources Research*, 49(8), 4687-4699.

- Yuan, X., & Wood, E. F. (2012). Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resources Research*, 48(12), W12519.
- Yuan, X., & Wood, E. F. (2013). Multimodel seasonal forecasting of global drought onset. *Geophysical Research Letters*, 40(18), 4900-4905.
- Yuan, X., Wood, E. F., Roundy, J. K., & Pan, M. (2013). CFSv2-based seasonal hydroclimatic forecasts over conterminous United States. *Journal of Climate*, 26, 4828–4847.
- Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews*, 19(NA), 333-349.