

OPERATIONAL APPLICATIONS OF MACROSCALE HYDROLOGIC MODELS
FOR SEASONAL STREAMFLOW FORECASTS
IN THE WESTERN UNITED STATES

Eric A. Rosenberg

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

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Anne C. Steinemann, Chair

Jessica D. Lundquist

Erkan Istanbuluoglu

Program Authorized to Offer Degree:

Civil and Environmental Engineering

©Copyright 2012

Eric A. Rosenberg

University of Washington

Abstract

OPERATIONAL APPLICATIONS OF MACROSCALE HYDROLOGIC MODELS
FOR SEASONAL STREAMFLOW FORECASTS
IN THE WESTERN UNITED STATES

Eric A. Rosenberg

Chair of the Supervisory Committee:

Professor Anne C. Steinemann

Department of Civil and Environmental Engineering

Despite a number of benefits for seasonal streamflow forecasts, macroscale hydrologic models (MHMs) remain underutilized by the operational community, due partly to misalignments between experimental and operational methodologies. This research addresses this problem through a series of conceptual frameworks that leverage unexploited strengths of MHMs and are demonstrated in the western United States. A hybrid dynamical-statistical approach is developed and tested in the 14 major watersheds of California's Central Valley drainage. The approach employs gridded precipitation and snow water equivalent (SWE) simulated by the Variable Infiltration Capacity (VIC) MHM as predictors in equations generated via the principal components regression methodology of the Natural Resources Conservation Service (NRCS). Results offer improvement over forecasts issued by California's Department of Water Resources, with particular benefits in watersheds poorly sampled by observations.

The approach is then modified to capitalize on its ability to identify locations with strong predictive power for network design applications. The modified approach is applied toward the expansion of the NRCS SNOTEL network in 24 western U.S. basins using

two forecasting scenarios – one assuming the conventional predictors of SWE and precipitation, and one considering soil moisture as an additional predictor. Results indicate that, for most basins, substantial skill improvements are only possible when soil moisture is considered. Furthermore, locations identified as optimal for soil moisture sensors are primarily found at low- to mid-elevations, in contrast to the higher elevations typically occupied by SNOTEL stations.

Finally, the significance of groundwater for seasonal streamflow forecasts is assessed by evaluating its contribution to interannual streamflow anomalies in the 29 Colorado River sub-basins. Terrestrial water storage changes are simulated by a version of VIC modified to include an underlying aquifer. These estimates are evaluated with satellite data and basin-scale water balances derived from observations. Simulated groundwater storage changes are then compared to those derived via baseflow recession analysis. Statistical analyses reveal little relationship between groundwater and streamflow anomalies, suggesting that operational forecasts are likely not degraded by the omission of groundwater conditions for the Colorado River. Viewed collectively, the research provides conceptual and operational contributions toward the adaptation of MHMs to a forecasting environment.

TABLE OF CONTENTS

I. INTRODUCTION	1
II. STATISTICAL APPLICATIONS OF PHYSICALLY BASED HYDROLOGIC MODELS FOR SEASONAL STREAMFLOW FORECASTS	6
2.1 Introduction	6
2.2 Study area	8
2.3 Methods	11
2.3.1 Statistical approach	12
2.3.2 Physically based simulation	15
2.3.3 Hybrid model	16
2.3.4 Control models	18
2.3.5 Surrogate observational data	20
2.4 Results	21
2.4.1 Forecast skill analysis	21
2.4.2 Analysis by water year type	26
2.4.3 Geospatial analysis of predictors	32
2.5 Conclusions	38
III. INFORMING HYDROMETRIC NETWORK DESIGN FOR STATISTICAL SEASONAL STREAMFLOW FORECASTS	41
3.1 Introduction	41
3.2 Study area	43
3.3 Data and methods	47
3.3.1 NRCS methodology	47
3.3.2 Hybrid approach	48
3.3.3 Augmented predictor analysis	49
3.3.4 Network design	50
3.3.5 Evaluation metrics	51
3.3.6 Operational applications	52
3.4 Results	53

3.4.1 Analysis of forecast skill	56
3.4.2 Geospatial analysis	57
3.4.3 Analysis of predictor variables	62
3.4.4 Operational analysis	64
3.5 Discussion and conclusions	65
IV. ON THE CONTRIBUTION OF GROUNDWATER STORAGE TO INTERANNUAL STREAMFLOW ANOMALIES IN THE COLORADO RIVER BASIN	68
4.1 Introduction	68
4.2 Study area	70
4.3 Data and methods	72
4.3.1 Hydrologic models	74
4.3.2 Basin-scale water balance	77
4.3.3 Satellite-derived terrestrial water storage change	78
4.3.4 Baseflow recession analysis	79
4.3.5 Statistical analysis	81
4.4 Results	82
4.4.1 Model performance	82
4.4.2 Satellite-derived storage changes	85
4.4.3 Recession analysis	87
4.4.4 Statistical analysis	90
4.5 Summary and conclusions.....	94
V. CONCLUSIONS.....	96
VI. REFERENCES	100

LIST OF FIGURES

2.1	The 14 watersheds of the Sacramento (blue), San Joaquin (green), and Tulare Lake (red) hydrologic regions, forming the study area for the paper.	9
2.2	Watershed and expanded predictor areas (top); numbers of simulated predictors (middle); and numbers of observed predictors (bottom).	19
2.3	Jackknife standard error as a percentage of mean streamflow for a shrinking target period over water years 1956 to 2005	22
2.4	10th and 90th percentiles of forecast residuals as a percentage of mean streamflow for a constant Apr–Jul target period over water years 1956 to 2005	24
2.5	Jackknife standard error for PCR forecasts with a shrinking target period and a 25-year calibration period (1981–2005) in the American. Incorporating snow sensor data as predictors allowed observation-based PCR forecasts to be generated in the additional months of December, January, and June	26
2.6	April 1 Nash-Sutcliffe efficiency scores for years of (from top to bottom) all, wet, normal, and dry water year types	28
2.7	May 1 Nash-Sutcliffe efficiency scores for years of (from top to bottom) all, wet, normal, and dry water year types. Note that PCR forecasts are for a shrinking target period while DWR forecasts are for a constant Apr-Jul target period	29
2.8	April 1 forecast bias as a percentage of mean target period streamflow for (from top to bottom) PCR forecasts using observed data, surrogate gridpoints, all gridpoints, and DWR’s official forecasts	30
2.9	May 1 forecast bias as a percentage of mean target period streamflow for (from top to bottom) PCR forecasts using observed data, surrogate gridpoints, all gridpoints, and DWR’s official forecasts. Note that PCR forecasts are for a shrinking target period while DWR forecasts are for a constant Apr-Jul target period	31
2.10	(Left) April 1 SWE and (right) October–March precipitation predictor locations for April 1 hybrid forecasts in the Feather. See text for more details	33
2.11	(Left) April 1 SWE and (right) October–March precipitation predictor locations for April 1 hybrid forecasts in the San Joaquin. See text for more details	34
3.1	The 24 basins in the study. Water resource regions are delineated by dashed lines, and federally protected wilderness areas are shown in green	46

3.2	Forecast skill for the augmented predictor analysis. The y-axis represents standard error as a percentage of mean streamflow for a shrinking target period. Lines labeled “NRCS” reflect operational NRCS forecast skill; lines labeled “NRCS predictors” reflect forecasts for which the requirement of month-to-month predictor consistency is relaxed	54
3.3	Forecast skill for the network design analysis. The y-axis is as in Fig. 3.2, with various labeling schemes as indicated. Note that the lines labeled “NRCS predictors” and “All grid cells” (EXP1 and EXP2) match those in Fig. 3.2. For some plots, lines may be obscured by others with similar values	55
3.4	Selected spatial results from the network design analysis. The sizes of the black circles are proportional to the improvements in forecast skill offered by their respective grid cells, and the green and red circles correspond to the green and red lines in Fig. 3.3.....	57
3.5	Spatial statistics for the dominant predictor types corresponding to the forecast improvements in Fig. 3.4. Grid cells for which the specified predictor type was selected for the forecast model are indicated in the leftmost plots by black dots, with the green and red dots corresponding to the green and red circles in Fig. 3.4. Climatology plots for SWE and water year-to-date precipitation (PCP) are shown in units of mm.	59
3.5	(continued) Spatial statistics for the dominant predictor types corresponding to the forecast improvements in Figure 3.4. Climatology plots for SM2 and SM3 are shown in units of volumetric water content.....	60
3.6	Ranges of correlations between observed and gridded/simulated predictor data. The numbers of observed data records are shown at top for their respective predictor types	61
3.7	Results of binary snow cover comparisons between VIC and MODIS for those grid cells that were selected as best overall in the network design analysis. The numbers above each bar represent the number of forecast months for which comparisons were performed	62
3.8	Percentage of seasonal streamflow variance that is explained by each predictor for forecast models labeled at top.....	63
3.9	Skill of various operational forecasts for 6 of the 7 basins in the Upper Colorado water resource region. Lines labeled “NRCS” correspond to those in Fig. 3.2. Shaded blue and brown areas indicate expected skill for NRCS forecasts that incorporate data from an additional observation station under the EXP1 and EXP2 scenarios.....	64

4.1	The Colorado River basin, including the 29 flow locations monitored by USBR, and the principal aquifer systems as given by Miller (1999), Robson and Banta (1995), and Whitehead (1996). Lee’s Ferry (station 20) is indicated in blue.....	71
4.2	Average annual precipitation over water years 1950–2008 (left) and equilibrium water table depth as simulated by SIMGM (right).	76
4.3	The locations of the 72 “reference-quality” watersheds (in blue) used in the baseflow recession analysis. The 29 USBR gauges and sub-basins are shown for reference.....	79
4.4	Nash-Sutcliffe efficiency scores for VIC and VIC-SIMGM in each of the 29 sub-basins.....	83
4.5	Time series of the subsurface storage terms (from top to bottom): third layer soil moisture by volumetric water content, third layer soil moisture in mm, total soil moisture anomaly, aquifer storage anomaly, and total subsurface storage anomaly.....	84
4.6	Comparisons between MODIS-derived and VIC ET (left), results of the basin-scale water balance (middle), and comparisons between simulated and observed streamflow (right) for selected sub-basins.....	86
4.7	Comparisons between VIC-simulated and GRACE-derived changes in storage	87
4.8	Recession plots and associated storage functions (insets) for four reference watersheds. Annual changes in groundwater storage as derived from storage functions are compared with those derived from VIC-SIMGM below their respective recession plots. Watersheds are located in sub-basin 2 (09075700), sub-basin 3 (09107000), sub-basin 9 (09208000), and sub-basin 18 (09352900), with drainage areas provided in parentheses.....	88
4.9	Comparison of monthly change in groundwater storage (y-axis, in mm) as derived from VIC-SIMGM and recession analysis. The number of reference watersheds used for each sub-basin is provided at lower right	89
4.10	Scatter plots for Lee’s Ferry of: (a) annual storage change v. annual streamflow, (b) annual storage change v. Oct. 1 storage anomaly, and (c) annual storage change v. previous water year streamflow. Annual storage changes were calculated between Oct. 1 and Oct. 1 of the following year. Red, yellow, and blue circles denote dry, normal, and wet water years, respectively	90
4.11	Ranges of simulated annual storage change as a percentage of annual streamflow observations for VIC (gray boxes) and VIC-SIMGM (white boxes). The central mark in each box indicates the median, and the edges of the box are the 25th and	

75th percentiles. Whiskers extend to a maximum of ~2.7 standard deviations from the median, and outliers are indicated by dots91

4.12 Scatter plots of forecast skill v. kappa for various storage term combinations, reflecting an Apr.-Jul. target period. The larger dots in each plot are for Lee's Ferry, with the numbers inside indicating the lead time in months93

LIST OF TABLES

2.1	Average annual statistics for the 3 hydrologic regions (bold) and the 14 watersheds in the study. For hydrologic regions, area and precipitation data are from CDWR [2009], and runoff data are from Dziegielewski et al. [1993] and CSWRB [1951]. For watersheds, drainage areas were calculated from huc250k shape files, precipitation was calculated by areal averaging VIC forcing data over water years 1956–2005, and runoff was calculated from CDEC data of unimpaired streamflows at the points indicated in Figure 2.1 (also over water years 1956–2005). The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation. mcm = million cubic m; taf = thousand acre-feet.	10
2.2	Summary of the forecasts compared in this study. Note that “surrogate” SWE predictors are a subset of “simulated” SWE predictors, and “surrogate” P predictors are a subset of “gridded” P predictors.	11
2.3	Predictor index variables used in DWR’s forecast equations for each of the 14 watersheds in the study. SWE = snow water equivalent; SWE _{hi} = high elevation SWE; SWE _{lo} = low elevation SWE; P _{OctMar} = Oct–Mar precipitation for the current water year; P _{AprJun} = Apr–Jun precipitation for the current water year; RO _{OctMar} = Oct–Mar runoff for the current water year; RO _{OctFeb} = Oct–Feb runoff for the current water year; RO _{preAprJul} = Apr–Jul runoff for the previous water year; and RO _{pre2wy} = total runoff for the previous two water years	13
2.4	Predictor statistics for the selected Apr 1 forecast equation in the Feather. Freq. = frequency of predictor occurrence in the top 30 equations; Corr. = correlation between predictor and predictand; CV = coefficient of variation; Mean = mean climatological value in mm; Reg. Coef. = regression coefficient in the selected equation; and Mean Water Contr. = product of the mean and the regression coefficient, which is equivalent to the mean “contribution” of the predictor to the streamflow forecast in mcm. The elements of PC1 and PC2 represent the loadings in the eigenvectors for each of the predictors, and explained variance was determined from the respective eigenvalues.	36
2.5	Predictor statistics for the selected Apr 1 forecast equation in the San Joaquin. Abbreviations are as defined for Table 2.4.	37
3.1	Identification data for the 24 basins in the study	44
3.2	Summary statistics for the 24 basins in the study. Mean values were calculated over the calibration period. The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation.	45

4.1 Average annual statistics for the 29 sub-basins over water years 1950–2008. The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation. The mean annual contribution represents the percentage of the total runoff at Imperial Dam (station 29).....73

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of many people. First and foremost, I would like to thank my advisor Anne Steinemann for her endless encouragement, selflessness, and friendship. I owe much to the kindness and support she has shown me these past few years. I am deeply grateful to my supervisory committee – Jessica Lundquist, Erkan Istanbuluoglu, Maury Roos, and Nate Mantua – for their thoughtful guidance and feedback throughout the course of this effort. I am grateful to Dennis Lettenmaier for challenging my abilities and providing me with incredible opportunities. I also greatly appreciate the friendship and wisdom of Steve Burges, who always had time for a chat.

Andy Wood deserves an immense amount of thanks for stepping in to offer guidance and advice when it was needed most, and for planting the seeds of this research. This research could not have been realized without the generous assistance and suggestions of David Rizzardo at the California Department of Water Resources and David Garen at the Natural Resources Conservation Service. Much of the funding for this work was supported by grants to the University of Washington from the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration, and the United States Geological Survey.

I am grateful to my fellow graduate students for their support, feedback, and friendship over the years. My friends both within and outside the university have been a reliable source of optimism and laughter. Words cannot express my love and appreciation for my parents, Maxine and Louis, who taught me what's important in life, and for my sisters, Leslie and Stacey, who always believed in me. Finally, to my wife Alex, who has patiently traveled on this journey with me and sacrificed so much, I am forever in your debt.

DEDICATION

To Alex, for more reasons than I could ever list; and to Tabitha and Mo, for keeping me company through the late hours of the night.

I. INTRODUCTION

Water managers in the western United States have long faced the challenge of meeting a variety of demands with limited and uncertain supplies. The history of the region is rife with examples of conflicts over this resource, resulting in the complex supply system and legal framework that exists today. One such conflict, the “Lake Tahoe Water Wars” at the turn of the twentieth century, motivated the region’s first snow surveys and the genesis of modern-day seasonal streamflow forecasts. Today, seasonal streamflow forecasts are critical for a broad array of sectors ranging from irrigated agriculture, flood control, and municipal water supply to endangered species protection, power generation, and recreation. The scientific literature abounds with case studies that illustrate their consequences, including both success stories [e.g., *Pagano et al.*, 2001] and failures [e.g., *Glantz*, 1982].

This dissertation is directed at the advancement of seasonal streamflow forecasting methodologies, the improvement of forecast skill, and the extension of forecast lead times. These objectives are motivated by the enablement of more informed decisions for water allocation, management, and planning.

Operationally, seasonal streamflow forecasting in the western U.S. is performed at the federal level by the Natural Resources Conservation Service (NRCS) and the National Weather Service (NWS), and at the state level (in California) by the Department of Water Resources (DWR). Seasonal streamflow prediction models can be divided into two broad categories – those that are data-driven and those that are process-driven. In the present context, data-driven models include the principal components regression (PCR) methodology of NRCS [*Garen*, 1992], the statistical water supply methodology of NWS [*Hartman and Henkel*, 1994], and the multiple regression methodology of DWR, all of which are derivatives of the region’s earliest forecasting models developed by *Church* [1935]. Process-driven models include the ensemble streamflow prediction (ESP) methodology of the NWS River Forecast System (NWSRFS) [*Day*, 1985], which gained

prominence around the time of the California drought of 1976–77 [Twedt *et al.*, 1977, 1978]. The NWSRFS employs the Sacramento Soil Moisture Accounting model [Burnash *et al.*, 1973], a so-called “conceptual simulation model” that is a derivative of the Stanford Watershed Model of Linsley and Crawford [1960]. Both data-driven and process-driven models have their advantages and limitations, which are discussed in greater detail in the chapters that follow.

Macroscale hydrologic models (MHMs) refer to a class of process-driven models that were developed to represent the land surface in general circulation models (GCMs). They are descendants of the “bucket model” of Manabe [1969] and the soil-vegetation-atmosphere transfer schemes (SVATs) introduced in the 1980s to perform this function. Distinguishing features of MHMs include distributed or semi-distributed spatial discretizations and parameterizations based on physical principles, although most include some level of empirical generalization in their structure. In addition to the typical GCM functions of climate change simulation and numerical weather prediction, MHMs are particularly well-suited for very large river basins and have been used in a multitude of “off-line” hydrologic studies at continental and regional scales over the past 20 years [e.g., Nijssen *et al.*, 1997; Mitchell *et al.*, 2004; Zaitchik *et al.*, 2010].

The utility of MHMs for seasonal streamflow forecasts has also been explored in a number of studies, some of which have focused on the western U.S. [e.g., Georgakakos *et al.*, 2005; Wood and Lettenmaier, 2006; Koster *et al.*, 2010]. Many of these studies have examined this issue through ensemble-based methods, which can exploit the ability of MHMs to incorporate weather and climate forecasts [e.g., Hamlet and Lettenmaier, 1999] and assimilate remote sensing data [e.g., McGuire *et al.*, 2006], among other advantages. Yet MHMs contain strengths for statistical prediction approaches as well, including the ability to characterize current hydrologic variables in much greater spatial detail than can be provided by point observations alone, and the leveraging of physical algorithms to simulate hydrometeorological situations not found in the historic training period.

This dissertation develops new methods of exploiting MHMs for seasonal streamflow forecasts, expansion of the hydrometric infrastructure that is the foundation for statistical forecasts, and a more complete understanding of the hydrologic mechanisms of seasonal streamflow predictability. The concepts are developed in the western U.S., with a specific focus on their operational utility. Despite the demonstrated value of MHMs for seasonal streamflow forecasts in previous studies, experimental methods have been slow to infiltrate the operational community, partly because many established operational methodologies remain statistically based. This dissertation is motivated by an overarching interest in the practical integration of MHMs within decision support settings, a vital need that has received inadequate attention in the scientific literature, despite an array of potential benefits for water resources management. A central theme is therefore the adaptation of new approaches to the established methodologies, which is critical to their ultimate implementation [NRC, 2008].

The dissertation is divided into three parts, each of which presents a unique approach to this problem, and together which provide a cohesive analysis of underlying issues. In the first, a hybrid dynamical-statistical methodology is developed that employs gridded observed precipitation and model-simulated snow water equivalent (SWE) data as predictors in statistical models generated via the principal components regression methodology of NRCS. To simulate SWE, we employ the Variable Infiltration Capacity (VIC) macroscale hydrologic model [Liang *et al.*, 1994], which is typical of the land surface models in GCMs and has been used in countless studies worldwide [e.g., among the most cited, Maurer *et al.*, 2002; Barnett *et al.*, 2005; Mote *et al.*, 2005]. The approach is tested for 14 watersheds in California's Central Valley drainage using calibration parameters adapted from DWR's forecasting environment, and results are compared to DWR's operational forecast skill.

The second part of the dissertation was inspired by user feedback for the hybrid approach, which indicated a particularly useful feature to be the ability to identify predictor locations that most greatly contribute to forecast skill. Here the approach is extended to adapt this strength for network design applications. The concept is tested for

the expansion of the NRCS SNOTEL network in 24 western U.S. basins using two forecasting scenarios – one that assumes the currently standard predictors of SWE and water year-to-date precipitation, and one that considers soil moisture as an additional predictor variable. Resulting improvements are spatially and temporally analyzed, attributed to dominant predictor contributions, and evaluated in the context of operational NRCS forecasts, ensemble-based NWS forecasts, and historical as-issued NRCS/NWS coordinated forecasts.

The third part of the dissertation focuses more directly on the hydrologic mechanisms underlying seasonal streamflow predictability. Despite the significance of initial hydrologic conditions for seasonal streamflow prediction in the western U.S. [*Shukla and Lettenmaier, 2011*], the role of groundwater in both statistical and ensemble-based operational forecasts has largely been ignored. This issue is addressed by evaluating the contribution of groundwater storage to interannual streamflow anomalies in the 29 tributary sub-basins of the Colorado River. Monthly and annual changes in total basin storage are simulated by two implementations of VIC – the standard release of the model, and an alternate version that has been modified to include the SIMple Groundwater Model (SIMGM) [*Niu et al., 2007*], which represents an unconfined aquifer underlying the soil column. These estimates are compared to those resulting from basin-scale water balances derived exclusively from observational data (gridded precipitation, naturalized streamflow observations, and satellite-based estimates of evapotranspiration) and changes in terrestrial water storage from the Gravity Recovery and Climate Experiment (GRACE) satellites [*Swenson and Wahr, 2006*]. Changes in simulated groundwater storage are then compared to those derived via baseflow recession analysis for 72 reference-quality watersheds from the GAGES-II database [*Falcone et al., 2010*]. Finally, estimates are statistically analyzed for relationships to interannual streamflow anomalies, and predictive capacities are compared to those of SWE and soil moisture.

The overarching science questions and subordinate questions that this dissertation seeks to address are as follows:

1. Can a hybrid framework, which combines model-simulated initial conditions with the regression-based methods used operationally, improve seasonal forecast skill? Can such an approach provide the ability to generate late-season forecasts, when snow exists at higher elevations but most observing stations are snow-free?
2. How can macroscale hydrologic models be employed to inform network design for statistical seasonal streamflow forecasts? Can soil moisture provide additional predictive skill in a statistical framework beyond that given by conventional predictors?
3. How significant are interseasonal and interannual groundwater anomalies for seasonal streamflow forecasts? How does the coupling of a macroscale hydrologic model to an explicit groundwater model affect soil moisture estimates and their predictive capacity?

These questions are addressed in the following three chapters. Chapter II [Rosenberg *et al.*, 2011] describes the hybrid framework and tests the approach in California's Central Valley drainage to address question 1. Chapter III [Rosenberg *et al.*, 2012a] adapts this framework for network design and evaluates its utility in the western U.S. to address question 2. Chapter IV [Rosenberg *et al.*, 2012b] evaluates the performance of a coupled surface-groundwater model in the Colorado River basin to address question 3. Viewed in its entirety, the research is intended to offer insights into new applications of MHMs for seasonal streamflow forecasts, and to form a foundation for the advancement of MHMs in the operational forecasting context of the western U.S.

II. STATISTICAL APPLICATIONS OF PHYSICALLY BASED HYDROLOGIC MODELS FOR SEASONAL STREAMFLOW FORECASTS

This chapter has been published in its current form in *Water Resources Research* [Rosenberg et al., 2011].

2.1 Introduction

The scarcity of water has defined much of the history in the western United States and continues to be one of its most complex and pressing public issues today. Decisions related to water usage have significant economic consequences, often with far-reaching implications that affect the welfare of the general public [e.g., Glantz, 1982]. It is therefore critical to ensure that this public resource is used most efficiently, which inherently involves accurate forecasting of its future availability.

The western U.S. is also distinctive in that over half of its annual streamflow is derived from snow, which acts as a natural reservoir at higher elevations until it runs off in the spring. Since the 1930s, this relationship has been exploited by the Natural Resources Conservation Service (NRCS) [Helms et al., 2008], which now works with National Weather Service (NWS) River Forecast Centers (RFCs) to issue seasonal water supply forecasts throughout the western U.S. [Franz et al., 2003]. California, which began water supply forecasting in 1929, remains the lone state to conduct its own snow surveys and issue independent forecasts under the direction of its Department of Water Resources (DWR) [Hart and Gehrke, 1990].

Notwithstanding more recent modifications to the mechanics of the approach [Garen, 1992; Pagano et al., 2009], most early forecasting techniques remain largely in active use today. The basic framework is similar for NRCS and DWR, with each relying on multiple regression techniques to relate a collection of predictors (snow water equivalent,

or SWE, accumulated precipitation, antecedent runoff, and in some cases, seasonal climate indices such as those based on the El Niño/Southern Oscillation) to a predictand (seasonal streamflow volume). From January to June, NRCS produces monthly forecasts for various target periods at 732 locations throughout the West, with unofficial “guidance” outlooks released on a daily basis for a subset of locations [*Pagano et al., 2009*]. DWR’s Bulletin 120 contains monthly forecasts of April–July streamflow for roughly 40 locations from February to May, with weekly updates issued from February to mid-June or as conditions warrant (DWR, 2008, personal communication). Predictor data is acquired from various ground-based sources, including storage precipitation gauges and United States Geological Survey (USGS) streamflow data that have been adjusted (“naturalized”) to account for upstream anthropogenic effects such as reservoirs or diversions. For SWE data, NRCS employs both manual snow course and automated snow sensor observations from its Snow Telemetry (SNOTEL) network. DWR depends primarily on snow course data from California’s Cooperative Snow Surveys Program for its first of the month forecasts, using snow sensors to help estimate missing data or correct erroneous data and relying on them exclusively for its weekly updates.

As a complement to statistical methodologies, Ensemble Streamflow Prediction (ESP) employs hydrologic and river routing models to produce forecasts of watershed runoff and streamflow [*Day, 1985; Twedt et al., 1977*]. In recent years, ESP has become a more central component of NWS water supply forecasting activities in the western U.S., as implemented in the NWS River Forecast System (RFS) [*Anderson, 1973; McEnery et al., 2005*], and has also been applied on a limited basis by NRCS. However, with a few exceptions [e.g., *Kim et al., 2001*, who evaluated the performance of ESP in Korea], studies that use “hindcasts” or “re-forecasts” have not demonstrated that ESP can generate significantly more accurate volumetric forecasts than those from existing statistical systems [*Pagano et al., 2009*]. Yet efforts to incorporate ESP into operational decision support systems have advanced despite these and other hurdles [e.g., *Georgakakos et al., 2005*].

Prediction approaches that incorporate physically based hydrologic models contain strengths not present in purely statistical systems. These include the ability to characterize current hydrologic variables in much greater spatial detail than can be provided by point observations alone [*Li et al.*, 2009; *Wood and Lettenmaier*, 2006]. Another is the leveraging of physical algorithms to simulate hydrometeorological situations not found in the historic training period, which is also possible via stochastic weather generators [see, e.g., *Wilks*, 1992]. Distributed, physically based estimates are useful not only for dynamical simulations, but can expand the predictor set for statistical forecasting applications as well. A hybrid approach that combines the initial conditions provided by a physically based hydrologic model with the regression-based methods used operationally has the potential to improve seasonal forecast skill.

In this paper, we explore the utility of a hybrid prediction approach in a case study involving DWR's seasonal forecasting system. The study was motivated by an overarching interest in the practical integration of model-based hydrologic simulation and prediction methods within water resources decision support settings. The approach was motivated by discussions with DWR personnel, who indicated that adaptations specifically tailored to their established statistical methodology were more likely to be implemented than a larger technological change toward purely model-based forecasting. We therefore give particular attention to comparisons of a hybrid approach with DWR's operational water supply forecasts.

2.2 Study area

California's high demand for water is fulfilled by a complex water supply system, including most notably the State Water Project (SWP), operated by DWR, and the Central Valley Project (CVP), operated by the U.S. Bureau of Reclamation. Together, they deliver roughly one-third of the 34 billion cubic m (bcm) (28 million acre-feet [maf]) of water consumed annually statewide, with local projects, groundwater, and Colorado River imports providing the rest [*CDWR*, 2009]. Initial SWP water allocations are generally issued in late November/early December, although these are based mainly

on current reservoir conditions and conservative hydrologic projections (i.e., climatology). Most key decisions regarding water supply usage (e.g., crop selection, groundwater needs) are reserved for January or February when the first snow surveys are conducted, and final allocations are typically issued in May (DWR, 2011, personal communication).

Precipitation in California varies greatly from more than 3550 mm (140 in) in the northwestern part of the state to less than 100 mm (4 in) in the southeastern part [CDWR, 2003]. With a climate dominated by the Pacific storm track, 75 percent of this precipitation falls between November and March, with the majority occurring from December through February [Carle, 2009]. Orographic effects generated by California's massive granite backbone, the Sierra Nevada, cause much of this precipitation to fall as snow on its western slopes. The resulting runoff forms the Central Valley drainage, which acts as a funnel for the state's two longest rivers, the Sacramento and San Joaquin, as they make their way to San Francisco Bay and the Pacific Ocean.

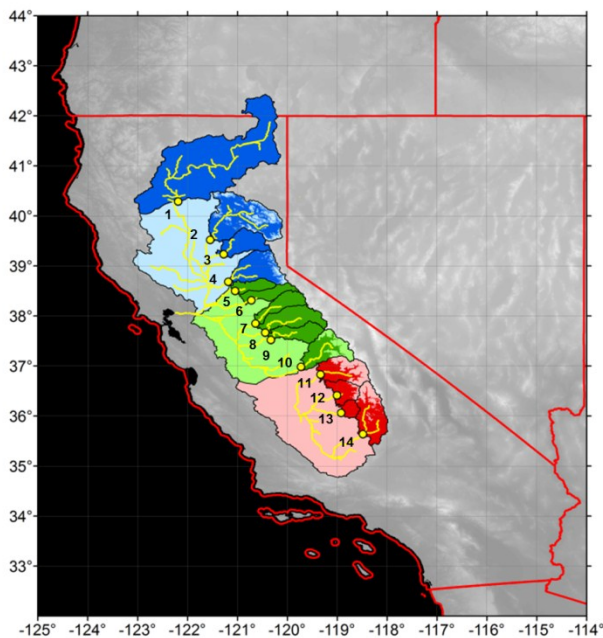


Figure 2.1: The 14 watersheds of the Sacramento (blue), San Joaquin (green), and Tulare Lake (red) hydrologic regions, forming the study area for the paper.

Contained within the Central Valley drainage are three distinct hydrologic regions, which together account for about half of the state's average annual streamflow of 88 bcm (71 maf). The regions are further subdivided into 14 major watersheds whose seasonal streamflows are forecast by Bulletin 120 at the primary locations in Figure 2.1; typical response times for these basins range from about 6 hours (Tule) to 8 days (Upper Sacramento) [USACE, 2001]. Summary statistics for each of the hydrologic regions and watersheds

are provided in Table 2.1. The Sacramento region is the wettest of the three, providing the bulk of the SWP and CVP exports to the agricultural areas and population centers in the drier south. The San Joaquin region is characterized by watersheds with higher

Table 2.1: Average annual statistics for the 3 hydrologic regions and the 14 watersheds in the study. For hydrologic regions, area and precipitation data are from CDWR [2009], and runoff data are from Dziegielewski et al. [1993] and CSWRB [1951]. For watersheds, drainage areas were calculated from USGS data, precipitation was calculated by areal averaging VIC forcing data over water years 1956–2005, and runoff was calculated from CDEC data of unimpaired streamflows at the points indicated in Figure 2.1 (also over water years 1956–2005). The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation. mcm = million cubic m; taf = thousand acre-feet.

	Drainage Area [km ² (mi ²)]	Annual Prec [mm (in)]	Annual Runoff [mcm (taf)]	Apr-Jul Runoff [mcm (taf)]	Annual Runoff Ratio
Sacramento R.	70,600 (27,200)	930 (36.7)	27,600 (22,400)	N/A	0.42
1 Upper Sac.	23,100 (8,900)	880 (34.7)	11,000 (8910)	3080 (2500)	0.54
2 Feather	9340 (3610)	1030 (40.7)	5700 (4620)	2200 (1780)	0.59
3 Yuba	3090 (1190)	1600 (62.8)	2930 (2380)	1240 (1010)	0.59
4 American	4890 (1890)	1270 (50.1)	3350 (2720)	1530 (1240)	0.54
San Joaquin R.	39,400 (15,200)	670 (26.3)	9700 (7900)	N/A	0.37
5 Cosumnes	1390 (540)	1000 (39.3)	480 (390)	160 (130)	0.35
6 Mokelumne	1500 (580)	1260 (49.5)	930 (750)	570 (460)	0.49
7 Stanislaus	2550 (990)	1110 (43.8)	1440 (1170)	870 (710)	0.51
8 Tuolumne	3980 (1540)	1060 (41.8)	2410 (1950)	1500 (1220)	0.57
9 Merced	2750 (1060)	970 (38.2)	1240 (1010)	780 (630)	0.47
10 San Joaquin	4340 (1680)	970 (38.2)	2260 (1830)	1550 (1260)	0.54
Tulare Lake	44,100 (17,000)	390 (15.2)	4100 (3300)	N/A	0.24
11 Kings	4000 (1550)	950 (37.5)	2120 (1720)	1510 (1220)	0.56
12 Kaweah	1450 (560)	910 (35.7)	560 (450)	350 (280)	0.42
13 Tule	1020 (390)	700 (27.4)	180 (150)	80 (60)	0.25
14 Kern	5370 (2070)	560 (21.9)	900 (730)	570 (460)	0.30

elevations that generally reach the peaks of their hydrographs later in the spring. The Tulare Lake region is naturally an endorheic basin, separated from the San Joaquin by a low, broad ridge that is overtopped by the Kings River in the wettest of years [CDWR, 2009; Carle, 2009]. Note that the Cosumnes and Mokelumne, while grouped in the San Joaquin region by Bulletin 120, are hydrologically distinct and more or less independent of the San Joaquin River system.

2.3 Methods

A summary of the forecasts compared in this study is presented in Table 2.2. Both principal components regression (PCR) and Z-Score regression methodologies were adapted from NRCS as detailed below. In contrast to the rest of the West, NRCS does not issue forecasts for watersheds in the Central Valley drainage, although it serves other parts of California such as the Klamath River and Lake Tahoe basins. The NWS, on the other hand, is represented in the region by the California Nevada RFC (CNRFC), which issues its own water supply forecasts using a combination of ESP and its own version of NRCS' PCR models (NWS, 2011, personal communication).

Table 2.2: Summary of the forecasts compared in this study. Note that “surrogate” SWE predictors are a subset of “simulated” SWE predictors, and “surrogate” P predictors are a subset of “gridded” P predictors.

Source of Predictor Data			Regression Methodology	Calibration Period (yrs)
SWE	P	RO		
Observed <i>courses</i>	Observed <i>gauges</i>	Observed	Multiple Linear (DWR)	50
Observed <i>courses</i>	Observed <i>gauges</i>	Observed	PCR, Z-Score	50, 25
Observed <i>courses and sensors</i>	Observed <i>gauges</i>	Observed	PCR, Z-Score	25
Surrogate <i>courses</i>	Surrogate <i>gauges</i>	Observed	PCR, Z-Score	50, 25
Surrogate <i>courses and sensors</i>	Surrogate <i>gauges</i>	Observed	PCR, Z-Score	50, 25
Simulated <i>all</i>	Gridded <i>all</i>	Observed	PCR, Z-Score	50, 25

2.3.1 Statistical approach

The statistical forecasting models of both DWR and NRCS can be represented as

$$Q = f(SWE, P, RO)$$

where the target period streamflow (Q) is a function of three general categories of predictor variables – snow water equivalent (SWE), accumulated precipitation (P), and antecedent runoff (RO). DWR relies on standard multiple regression to develop its forecast equations, which employ various types of these predictor variables as listed in Table 2.3. The two that are common to all watersheds, SWE and accumulated precipitation, are weighted indices of observations at multiple locations (typically 10–20 for SWE and about 10 for precipitation) in and around the watershed boundaries. For the six watersheds with more drastic ranges of topography (Feather, American, San Joaquin, Kings, Kaweah, and Kern), SWE is further divided into high-elevation and low-elevation indices. Runoff, which is measured at the same stations for which the forecasts are issued, is more subjectively used depending upon the specific characteristics of each watershed; forecasts in the Upper Sacramento, for example, employ total runoff over the previous two water years in consideration of the greater water retention properties of the volcanic soils in that region (DWR, 2008, personal communication). The form of the calibrated equation varies by watershed, but generally consists of a polynomial model with predictors that have been transformed (typically via a power function) to account for nonlinearities.

Table 2.3 shows that several of DWR’s predictors describe conditions that may be unknown at the time of a forecast. A forecaster working on February 1 does not know the current water year’s October–March precipitation, for example. To account for this discrepancy, DWR relies on what are termed “future variables,” which (as the name suggests) extrapolate current conditions of predictor variables to future conditions using

Table 2.3: Predictor index variables used in DWR’s forecast equations for each of the 14 watersheds in the study. SWE = snow water equivalent; SWE_{hi} = high elevation SWE; SWE_{lo} = low elevation SWE; P_{OctMar} = Oct–Mar precipitation for the current water year; P_{AprJun} = Apr–Jun precipitation for the current water year; RO_{OctMar} = Oct–Mar runoff for the current water year; RO_{OctFeb} = Oct–Feb runoff for the current water year; $RO_{preAprJul}$ = Apr–Jul runoff for the previous water year; and RO_{pre2wy} = total runoff for the previous two water years.

	SWE	SWE_{hi}	SWE_{lo}	P_{OctMar}	P_{AprJun}	RO_{OctMar}	RO_{OctFeb}	$RO_{preAprJul}$	RO_{pre2wy}
1	x			x	x		x		x
2		x	x	x	x	x			
3	x			x	x			x	
4		x	x	x	x			x	
5	x			x	x				
6	x			x	x				
7	x			x	x	x			
8	x			x	x				
9	x			x	x			x	
10		x	x	x	x	x		x	
11		x	x	x	x			x	
12		x	x	x	x	x			
13	x			x	x	x		x	
14		x	x	x	x	x		x	

their long-term medians. Thus, using the same example, the observed precipitation from October to January is added to the long-term median precipitation for February and March to derive the total October–March precipitation used in the February 1 forecast. Similarly, SWE predictors always describe conditions on April 1 (when peak SWE is typically considered to occur), and forecasts always reflect the April–July target period, even when they are made later in the water year and a portion of the target period’s streamflow has already been observed. This practice allows the use of a single equation for all dates forecasts are issued, achieving greater month-to-month consistency in

predictor variables and a larger sample size for equation calibration. Note that DWR's final forecasts are the result of balancing predictions at several points in each watershed, regional comparisons of trends and relationships, and "forecaster feel" or experience (DWR, 2011, personal communication).

In contrast to DWR, NRCS uses a regression approach based on principal components that dates to the early 1990s. A well-known complication in multivariate regression is collinearity among the predictor variables. DWR's practice of constructing composite indices from data of like data type (e.g., a single SWE value that is a weighted index of multiple SWE observations) partially circumvents this problem, but because the weights are subjectively determined outside of the regression, they are not statistically optimal [Garen, 1992]. PCR is a method of restructuring the predictor variables into uncorrelated principal components, which become the regressors.

The NRCS approach considers only data known at the time of forecast as candidate predictors, which leads to the use of separate equations with varying predictors for each forecast date. Regression coefficients are determined by arranging principal components in order of decreasing eigenvalue (explained variance), developing an equation that sequentially retains only those principal components deemed significant via a t-test, and inverting the transformation so that coefficients are expressed in terms of the original predictor variables. The routine employs an iterative search algorithm that optimizes variable combinations by developing all possible equations resulting from an increasing number of predictors. With each additional variable, the standard error resulting from a jackknife procedure is used to order the equations, and the top 30 equations are identified. When the top 30 equations no longer change from one round to the next, a final equation is selected by striking a compromise between jackknife standard error and month-to-month variable consistency. NRCS' complete regression methodology is fully detailed by Garen [1992] and Garen and Pagano [2007].

Z-Score regression is a heuristic statistical method that takes advantage of predictor collinearity to relax the requirement for serial completeness [Pagano et al., 2009; Garen

and Pagano, 2007]. Individual predictors are first pooled into groups of like data type, and each observation is converted into a Z-Score, or standardized anomaly. For each data type, the coefficient of determination (R^2) between standardized predictors and standardized predictand are used to generate an index wherein each element is a weighted average of all available Z-Scores for that year. The computed index for each data type is then itself standardized, and the process is repeated to create a single composite index reflective of all the predictor data available each year. This composite index is then regressed against the standardized streamflow volumes using a least squares method. While PCR remains the official forecasting methodology for NRCS, Z-Score regression is currently used to provide daily objective guidance to users for a subset of forecast locations [Pagano et al., 2009].

2.3.2 Physically based simulation

Regardless of the regression technique, all of the aforementioned methods employ observed data as predictors. In this study, we contribute simulated SWE predictors using the snow model [Andreadis et al., 2009] contained in the Variable Infiltration Capacity (VIC) macroscale hydrologic model [Liang et al., 1994]. VIC is a semi-distributed grid-based model that is typical of land surface models (LSMs) used in most numerical weather prediction and climate models [Wood and Lettenmaier, 2006]. Like other LSMs, VIC solves the water and energy balance at each time step, but is distinguished by its parameterization of subgrid variability in soil moisture, topography, and vegetation. VIC has been successfully applied in a number of research studies involving major river basins worldwide [Nijssen et al., 1997], and was demonstrated to reproduce SWE, soil moisture, and runoff data that compared favorably with observed data for watersheds of varying size across the conterminous U.S. [Maurer et al., 2002].

Snow accumulation and ablation processes within VIC are simulated using a two-layer energy and mass balance approach. A thin surface layer is used to solve energy exchange with the atmosphere, while the lower or pack layer is used as storage to simulate deeper snowpacks [Andreadis et al., 2009; Cherkauer and Lettenmaier, 2003]. The model

contains an explicit overstory canopy interception scheme that accounts for sublimation, meltwater drip, and mass release of intercepted snow.

For this study, VIC implementations described in *VanRheenen et al.* [2004] and *Tang and Lettenmaier* [2010] were adapted to a spatial resolution of 1/16 degree (roughly 5–7 km at this latitude). Each grid cell was further subdivided into as many as five elevation bands, depending on elevation range. VIC as a whole was implemented at a 24-hour time step, with the embedded snow model using a 1-hour time step. The model was forced by daily precipitation and maximum/minimum temperature data from National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer stations, and daily wind data from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis Project, following the methods of *Maurer et al.* [2002].

2.3.3 Hybrid model

Two objectives guided the development of the hybrid models: first, to produce methods relevant to DWR’s current operational forecasting setting, and second, to allow objective comparisons between the respective skill of simulated and observed predictors. Our regression approach is therefore a compromise between the aforementioned statistical techniques. We employed the same predictor variables in each watershed as those listed in Table 2.3, and we calibrated our equations over the same 50-year period (water years 1956 to 2005) used by DWR. Instead of standard multiple regression, however, we adopted the PCR and Z-Score methodologies used by NRCS. PCR forecasts were generated using the equation with the median jackknifed standard error of the best 30 models resulting from the search algorithm. This choice “handicapped” our results to match more closely those that might be expected operationally, which typically reflect models with slightly less than optimal skill due to the additional selection criterion of maintaining consistent predictors from month-to-month (see section 2.3.1). In actuality, the difference in skill between the median equation and the best equation was negligible.

Unlike DWR's methodology, predictor variables were used only as they were known at the time of the forecast; SWE data were not extrapolated to April 1, and data representing accumulated quantities were used only when in the past. A forecast issued on February 1 included just October–January precipitation in its value for the October–March precipitation predictor and did not use the April–June precipitation predictor at all, for example. We also imposed the criterion that predictors have a correlation of at least 0.3 with the predictand (as is the practice at NRCS) and at least 10 non-zero values over the 50-year period. In addition, rather than using a constant target period regardless of the forecast date, we predicted shrinking target periods reflecting only the months remaining in a season. A forecast on April 1 was thus for the entire April–July season, but a forecast on May 1 was made just for May–July. This choice has two benefits. First, in contrast to DWR's practice of using April 1 SWE as a predictor even in forecasts issued later in the season, our usage of current conditions would result in constant target period forecasts that employ late-season SWE to predict prior streamflow. Second, by focusing our attention on just the water remaining in the snowpack, we are better able to test the late-season performance of a hybrid approach, which can integrate SWE simulated at the highest elevations at a time when snow courses or sensors at lower elevations may already be snow free.

The hybrid model predictors included VIC-simulated SWE data and gridded precipitation forcing data as described in section 2.3.2. Runoff predictors were obtained from observed records of unimpaired streamflow, archived at the California Data Exchange Center (CDEC) (available at <http://cdec.water.ca.gov>), for the locations shown in Figure 2.1. All SWE data occurred on the first of the month, as at DWR. The domain from which the simulated data were selected depended on the locations of DWR's snow and precipitation observing stations, many of which are located outside catchment boundaries (DWR, 2008, personal communication). Watershed boundaries were expanded (“offset”) by either 1/4, 1/2, or 3/4 degree in latitude or longitude in order to encompass all of the observing stations, and simulated data were compiled for all grid points included in the expanded domains. Since observations external to catchment boundaries are often considered proxies for unmonitored points within, some may question the need for

external data when all internal data are simulated. Yet the weather patterns resulting in precipitation and snow yield covariability at these locations as well, and thus the offsets offer an opportunity to assess the relationship of predictive skill to distance from watershed boundaries. Figure 2.2 shows watershed areas, expansion offsets, and expanded areas (top), along with the numbers of simulated SWE and precipitation predictors (middle) for each watershed. The greater numbers of SWE than precipitation predictors result from the multiple elevation bands present within each grid cell, as discussed in section 2.3.2.

It should be noted that our hybrid approach does not account for structural uncertainty in forecasts, a topic that has received considerable attention in the recent hydrologic literature [e.g., *Devineni et al.*, 2008; *Sharma and Chowdhury*, 2011]. Structural uncertainty, which concerns forecast errors that can be attributed to deficiencies in model formulations, is particularly relevant when models are based on a large number of candidate predictors. A simple strategy for addressing this issue could involve a static combination of forecast models. In the present context, such an approach might derive a weighted average of forecasts resulting from each of the 30 models selected by the PCR search algorithm, with weights determined based on model fitness. While static combination has been shown to improve the stability of model predictions, however, unwanted side-effects such as biases in the estimation of high and low flows are likely to result [*Sharma and Chowdhury*, 2011]. Other strategies such as dynamic model combination could better address the issue of structural uncertainty but would increase system complexity and lie outside the scope of the study to investigate.

2.3.4 Control models

Our “control” forecast models were developed using the same approach as the hybrid models (i.e., PCR and Z-Score), but incorporated observed data obtained from CDEC for

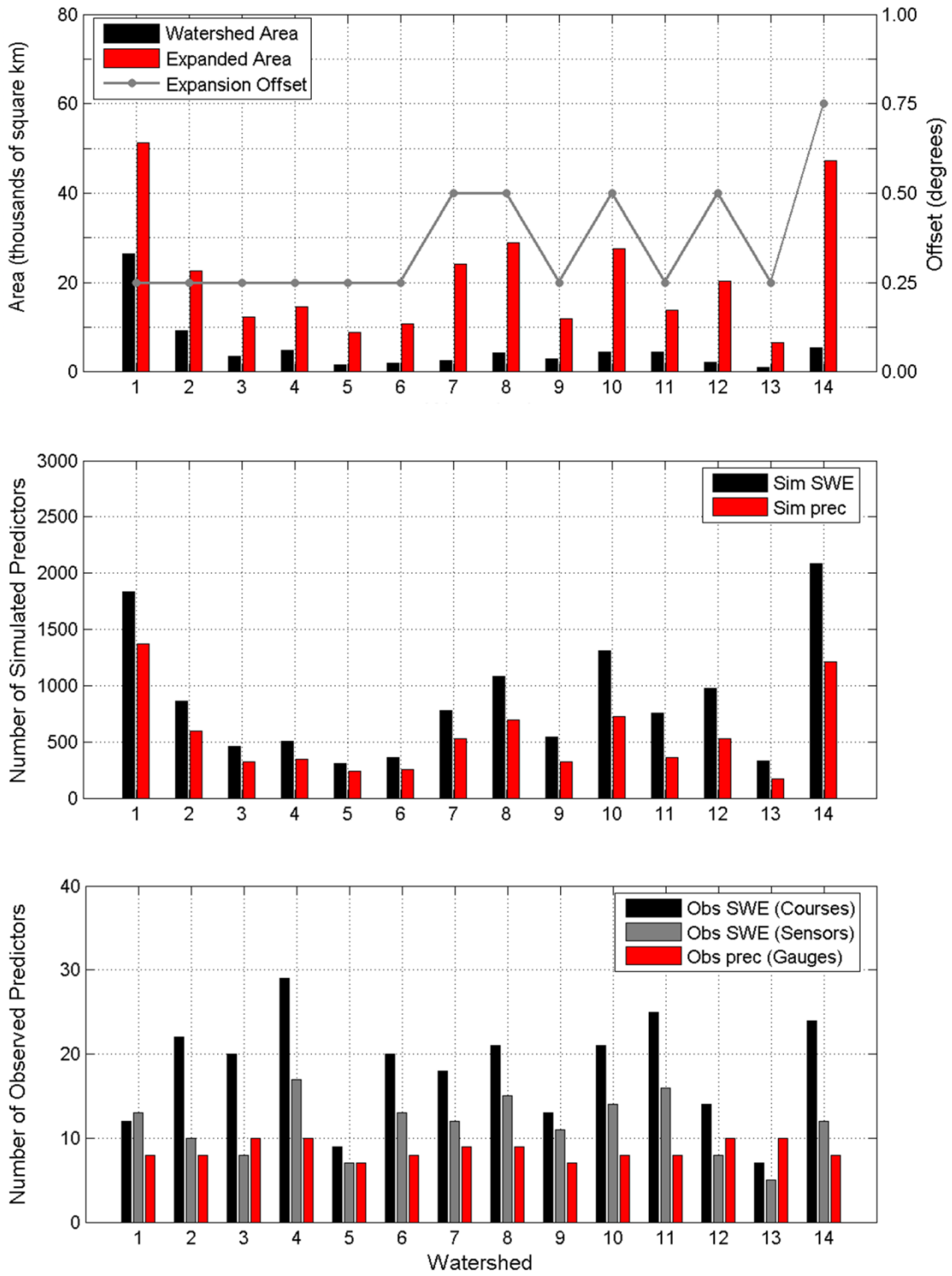


Figure 2.2: Watershed and expanded predictor areas (top); numbers of simulated predictors (middle); and numbers of observed predictors (bottom).

stations used by DWR in their monthly forecasts. In contrast to forecasts based on simulated data, however, forecasts based on observed data are less straightforward. Many observed records are serially incomplete, an issue that can be circumvented using an approach like Z-Score regression, but which complicates methods like PCR. Moreover, DWR's official SWE predictors comprise manual snow course observations but not automated snow sensor observations, which occasionally suffer inaccuracies due to problems like "ice bridging" (DWR, 2008, personal communication). Thus, official observations are generally taken only as needed in the months of February to May, precluding the development of control models at other times of the year.

To address this issue, we incorporated snow sensor data by selecting a collection of snow sensors that mimicked the official snow courses. Most snow sensor records, however, date only as far back as the late 1970s or early 1980s, whereas DWR's official equations are calibrated over a 50-year period from water years 1956 to 2005. To support a consistent comparison between prediction approaches, we devised two types of control models: one that included just snow course data and was calibrated over the full 50-year period used by DWR, and one that included both snow course and snow sensor data and was calibrated over a 25-year period from water years 1981 to 2005. The numbers of snow courses, snow sensors, and rain (precipitation) gauges used for each watershed are indicated in Figure 2.2 (bottom). We estimated missing SWE and precipitation observations (as required by PCR) by regressing stations with missing data against those of like data type within the same watershed, which is the practice preferred by NRCS (NRCS, 2010, personal communication). Stations that had less than 80 percent of a complete record were excluded from the analysis.

2.3.5 Surrogate observational data

Despite the efforts detailed above, SWE records for many watersheds remained incomplete in June and July. We therefore added an analysis based on "surrogate" observational data, which consisted of either gridded precipitation data or simulated SWE data at grid points adjacent to each observing station. Surrogate SWE data were selected

from the model's elevation band most closely corresponding to the elevation of the observing station.

The surrogate approach allows us to quantify the benefit of additional predictor candidates in a statistical model, which is a central question of this study. Comparisons between the predictive skill of simulated and observed data are complicated by differences between the two data sets. The surrogate approach circumvents the uncertainty associated with these differences since the surrogate data are selected from, and thus a subset of, the simulated data. The surrogate data also provide a baseline from which to compare the skill of actual observations, or to estimate their potential skill in months with insufficient data. These comparisons are discussed in section 2.4, which also includes comparisons with forecasts retrospectively generated using DWR's current regression equations, circa 2006 (DWR, 2008, personal communication).

2.4 Results

2.4.1 Forecast skill analysis

We first present our results from the surrogate analysis to establish a baseline of performance. Jackknifed standard error comparisons between forecasts based on the full suite of simulated data and those based on just the selected surrogate data are presented in Figure 2.3 for both Z-Score and PCR methodologies. All results are based on forecasts generated using linear forms of regression models. Similar experiments based on nonlinear equations did not yield more normally distributed residuals or lower standard errors than their linear counterparts.

Several patterns emerge from the results of Figure 2.3. For each watershed, PCR forecasts are comparable to or better than those produced using a Z-Score approach. Increasing the number of candidate predictors consistently results in superior forecast skill under the PCR approach, with models developed from all gridpoints in a domain (hybrid forecasts)

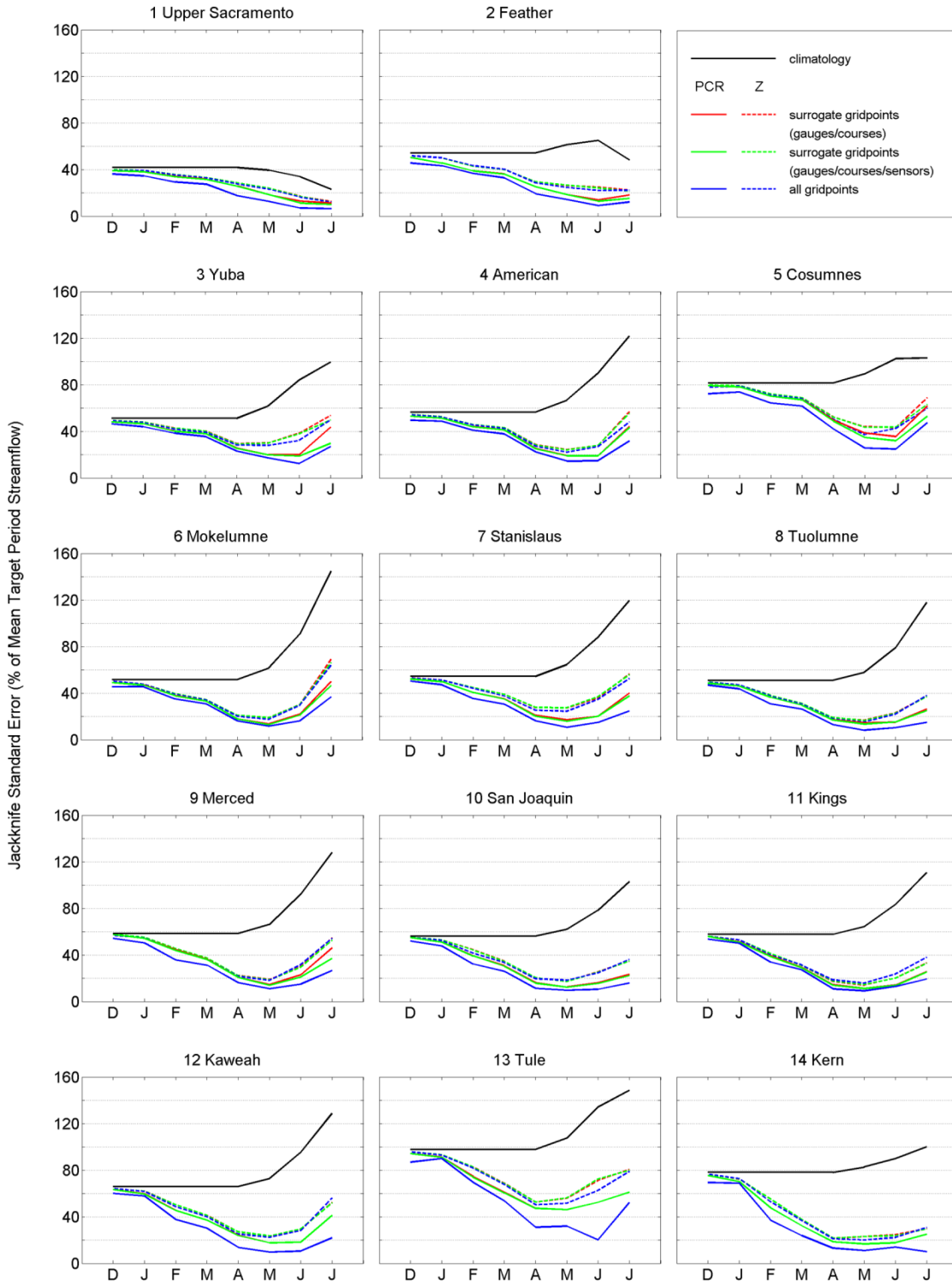


Figure 2.3: Jackknife standard error as a percentage of mean streamflow for a shrinking target period over water years 1956 to 2005.

achieving the lowest standard errors. For almost all watersheds, including the Yuba and Merced in particular, the largest improvements occur late in the snowmelt season in the months of June and July, supporting the notion that a hybrid model can exploit the ability to simulate SWE at high elevations. The greatest overall improvements occur in the Cosumnes and Tule, which are less snowmelt-dominant and have lower elevations and streamflows than other watersheds in the study. These results appear to be mostly a reflection of the coverage offered by existing observation networks, which are less dense than in other catchments, and the particular performance of the hybrid forecasts in wet years (section 2.4.2), which tend to dominate standard error for all years because of the positive skew in the streamflow distributions. Note, however, that even with these improvements, forecast error is still higher in the Cosumnes and Tule than other watersheds due to the higher coefficients of variation (CV) of their Apr–Jul streamflows (0.81 and 0.97, respectively).

Results using a Z-Score approach tell a different story. In most watersheds, forecasts based on the full set of simulated data are comparable to or, at best, marginally better than those based on just the selected surrogate data. In the Kings River watershed, forecasts based on the full set of simulated data are actually worse than those based on just the selected surrogate data. Interestingly, as in the PCR approach, the largest improvements occur in the Cosumnes and Tule River basins. The poor performance of Z-Score was clearly due to a lack of a search routine to first screen out those predictors with negligible predictive skill, thus diluting the predictive skill of the group as a whole. As a test, we performed an additional analysis whereby only those stations selected for the PCR model were used as Z-Score predictors, and found the results to be comparable to those obtained by PCR alone. However, because this screening routine was already embedded in the PCR approach, Z-Score was abandoned for the rest of the analysis.

Figure 2.4 presents data reflecting the 10th and 90th percentiles of forecast residuals as a percentage of mean streamflow. These “funnel plots,” so named because of their shape, compare DWR’s forecasts with PCR forecasts based on three types of predictor data – the

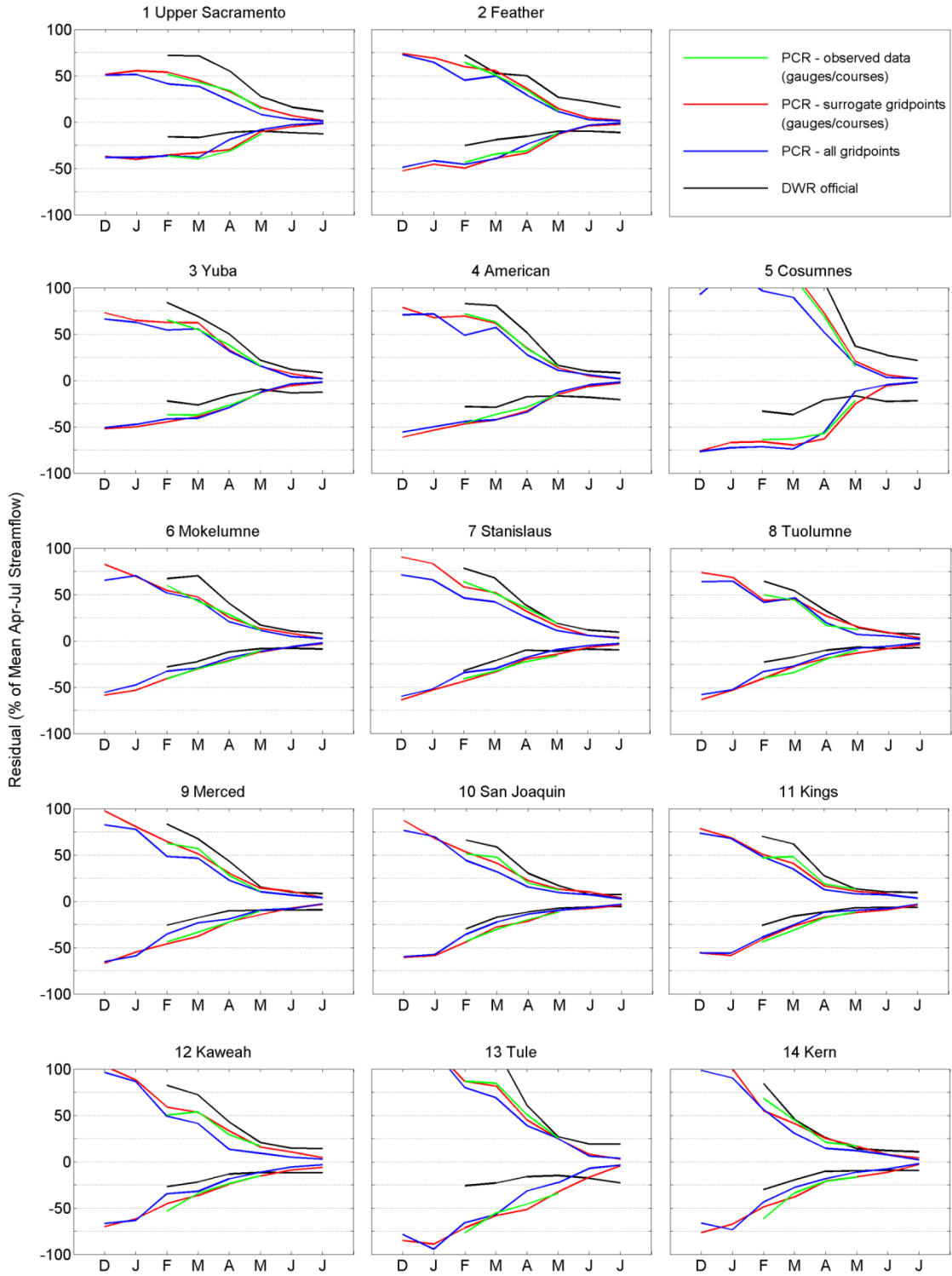


Figure 2.4: 10th and 90th percentiles of forecast residuals as a percentage of mean streamflow for a constant Apr-Jul target period over water years 1956 to 2005.

full set of simulated data for each domain, selected surrogate data, and observed data that have been filled using a Z-Score regression methodology (see section 2.3.4). Because all four forecast types are calibrated over the 50-year 1956–2005 period, forecasts that employ observed or surrogate snow sensor data were excluded. Residuals from DWR’s forecasts are shown beginning in February, which is the first month they are available each water year, and residuals from PCR forecasts based on observed data are shown only for February through May, since those were the only months that permitted the development of complete SWE records. Funnel plots are typically employed by DWR as a measure of model skill (DWR, 2008, personal communication), and as such, reflect the entire Apr–Jul season rather than a shrinking target period as in Figure 2.3. PCR forecasts issued after April 1 were therefore adjusted by adding streamflow observed since April 1 to account for the full period. This unequal advantage should be noted when comparing them to DWR’s official forecasts in these plots.

A striking result from Figure 2.4 is the close correspondence between forecast residuals based on observed data (green lines) and those based on simulated surrogate data at selected gridpoints (red lines). Despite the differences between the two data sets, a surrogate approach appears to produce forecasts that are remarkably similar to those based on their observed counterparts. The implications of this result suggest that forecast skill for months without sufficient observed data can be reasonably reproduced using forecasts based on surrogate (estimated) predictor data. As a check on the interannual variability of modeled SWE, we calculated time series of composite indices that were weighted averages of the surrogate SWE data, and then compared these indices to those based on observed SWE data used by DWR. Correlations between the two were high, ranging from 0.84 to 0.97 for April 1 SWE, depending on the watershed. We also checked whether a relationship existed between the difference in modeled and observed indices and DWR’s historical forecast errors, but found only weak correlations at best.

As in Figure 2.3, the funnel plots indicate that most forecasts based on the full set of simulated data offer at least some improvement over PCR forecasts based on either observed or surrogate data. When averaged over both 10th and 90th percentiles, for

example, improvements in April 1 forecasts range from about 1-2% (Yuba and American) to 12-13% (Kaweah and Tule). In terms of streamflow volume alone, the greatest improvement in April 1 forecasts occurs in the Upper Sacramento, where a difference of ~11% equates to a reduction in forecast error of about 340 million cubic m (mcm) (280 thousand acre-feet [taf]). The apparent late-season superiority of PCR forecasts over DWR’s official forecasts should be tempered by the incongruity between them noted above. However, the earlier season PCR residuals are relatively unbiased (i.e., well-centered around the zero percentile) in contrast to the official forecast residuals, which appear shifted in the positive direction. This asymmetry is most likely due to the nonlinearity of DWR’s equations and will be addressed in greater detail in the next section.

Efforts to incorporate snow sensor data as predictors offered little additional information about the late-season performance of observation-based PCR forecasts. Despite the use of a smaller 25-year calibration period, the sparseness of these data still left most watersheds with incomplete SWE records in the months of June and July. For exceptions such as the American River watershed, results corroborated the above findings that skill metrics for surrogate-based forecasts were reasonable indicators of skill for forecasts based on their observed counterparts (Figure 2.5).

2.4.2 Analysis by water year type

In addition to evaluating forecast performance in all years, we assessed performance in wet, normal, and dry year categories (defined by terciles). These groupings were analyzed via another commonly used skill metric, the Nash-Sutcliffe coefficient of

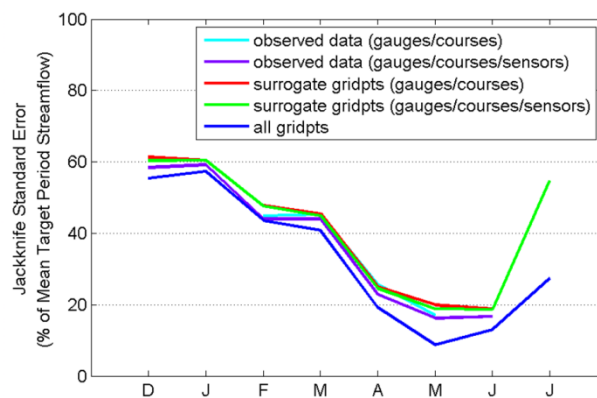


Figure 2.5: Jackknife standard error for PCR forecasts with a shrinking target period and a 25-year calibration period (1981–2005) in the American. Incorporating snow sensor data as predictors allowed observation-based PCR forecasts to be generated in the additional months of December, January, and June.

efficiency (NS) [Nash and Sutcliffe, 1970]. An NS score of 1 is perfect, 0 indicates skill equal to that of climatology, and values less than 0 denote negative skill.

NS scores for each of the 14 watersheds are given in Figure 2.6 for forecasts issued on April 1, which is considered the start of the melt season and thus a benchmark for comparison, and in Figure 2.7 for forecasts issued on May 1, when final allocations for SWP contractors are issued. Figure 2.6 (top) shows scores over all water year types and resembles the results in Figures 2.3 and 2.4. PCR forecasts based on the full set of simulated data result in the best scores, those based on either observed or surrogate data result in slightly lower scores, while DWR's official forecasts generally do not perform as well. The next plot (wet years only) is similar to the top plot in its skill rankings, although the scores are lower for each watershed. NS scores for normal and dry water years, however, are less coherent. For normal years, DWR's forecasts outperform PCR forecasts in the Sacramento watersheds, while PCR forecasts outperform DWR's forecasts in many of the other watersheds; for dry years, the reverse is more often true. In Figure 2.7, which shows results for forecasts issued on May 1, scores are consistently higher. Forecasts in the top two plots show similar patterns to those in Figure 2.6, with considerable wet-year improvements in both figures for the Cosumnes and Tule (watersheds 5 and 13). Conspicuously, however, DWR's forecasts score highest for normal and dry years in Figure 2.7. Note that May 1 PCR forecasts are for streamflow from May to July while DWR's forecasts are for streamflow from April to July, although the effect of this disparity on NS scores is most likely small.

The variation in performance of the methods in Figures 2.6 and 2.7 is largely explained by the variations in mean forecast bias shown in Figures 2.8 and 2.9. Note that the biases in these plots reflect forecasts minus observations, as opposed to the residuals presented in Figure 2.4, which reflect observations minus forecasts in accordance with DWR's methodology. The top three plots in Figure 2.8, which are all based on PCR forecasts, show similar patterns. Biases calculated over all water years are consistently zero, as expected for PCR models. Biases calculated for wet years are minimal, while those for normal and dry years are slightly larger. In contrast, DWR's forecasts exhibit a markedly

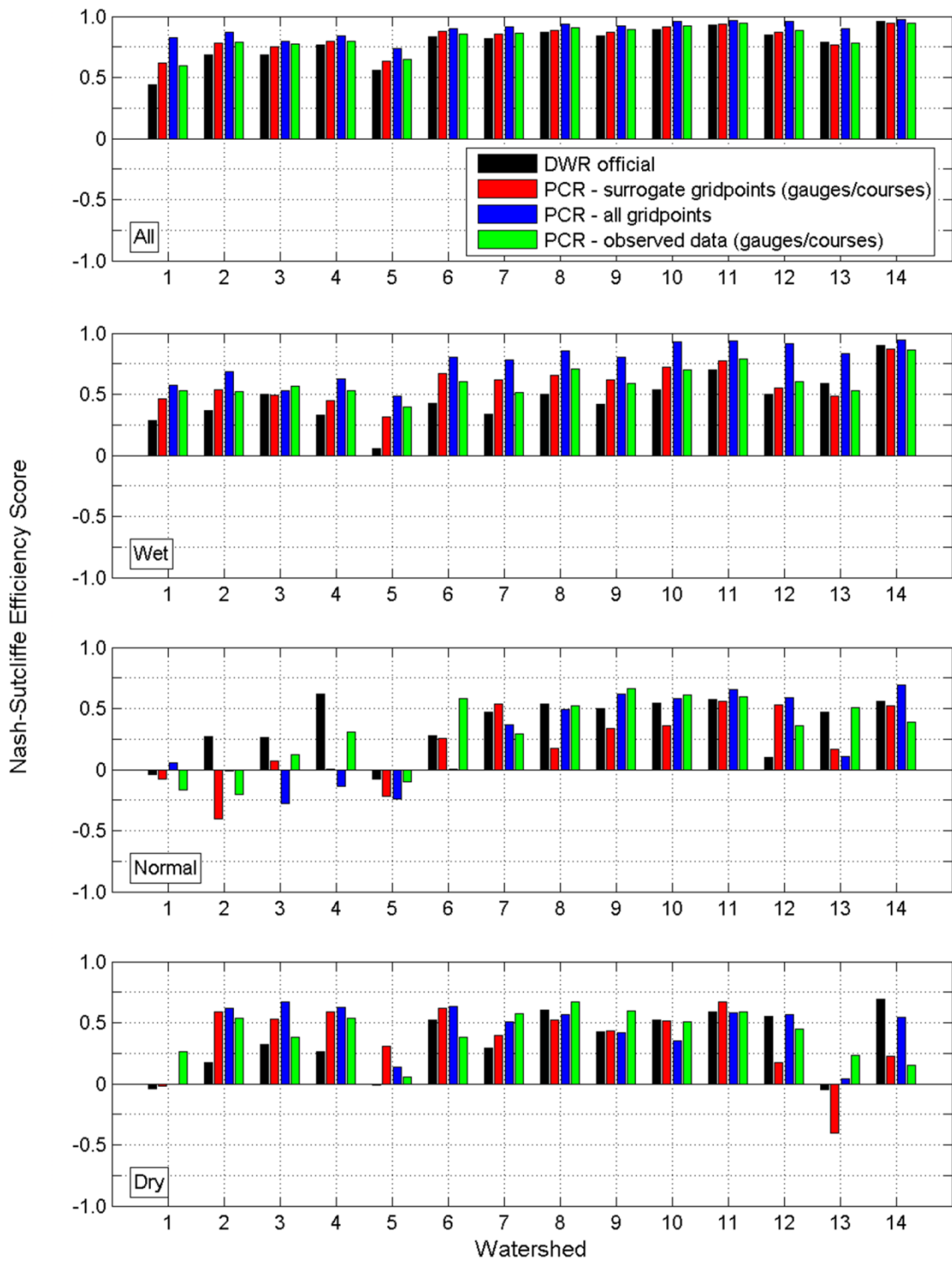


Figure 2.6: April 1 Nash-Sutcliffe efficiency scores for years of (from top to bottom) all, wet, normal, and dry water year types.

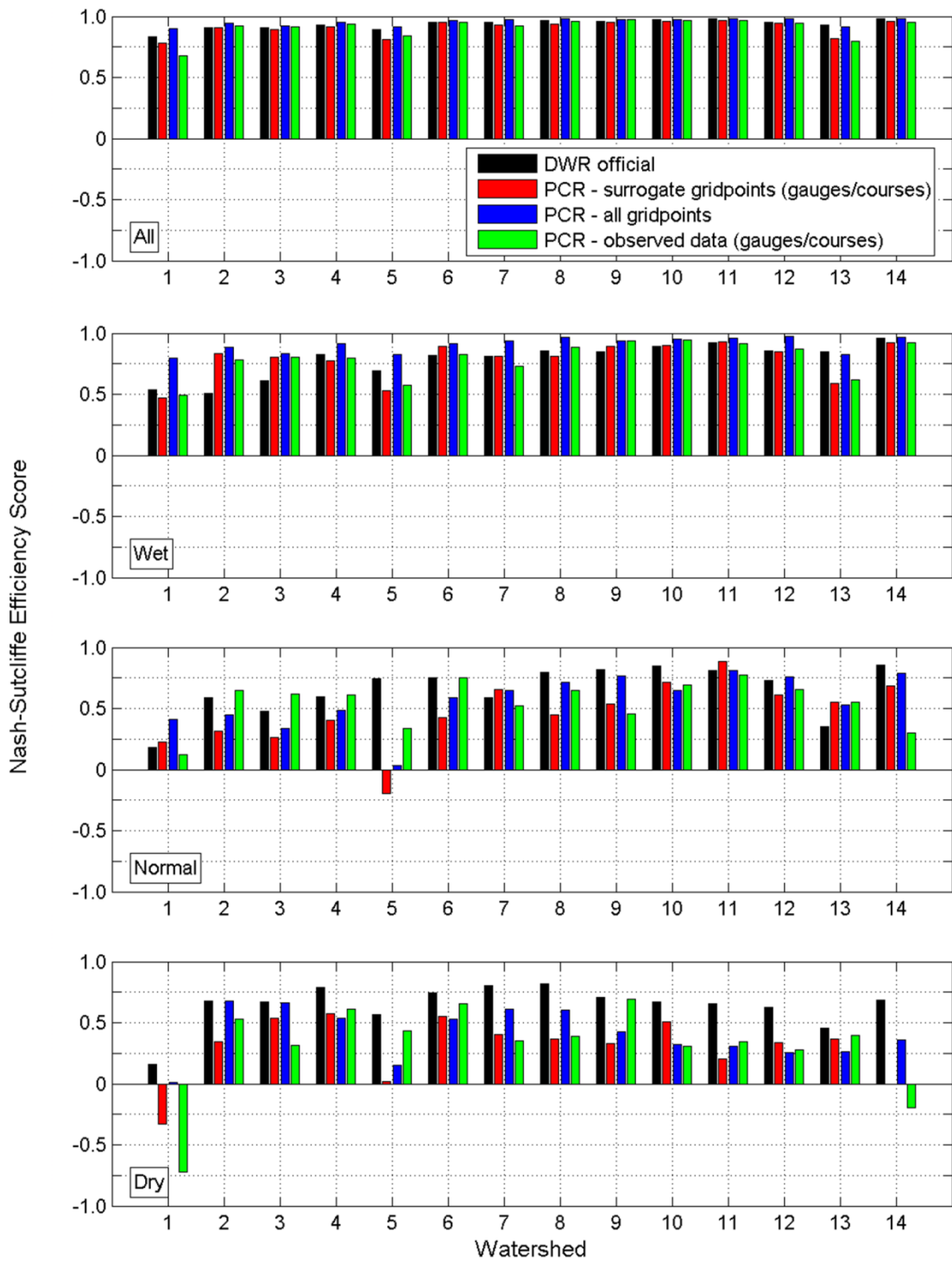


Figure 2.7: May 1 Nash-Sutcliffe efficiency scores for years of (from top to bottom) all, wet, normal, and dry water year types. Note that PCR forecasts are for a shrinking target period while DWR forecasts are for a constant Apr-Jul target period.

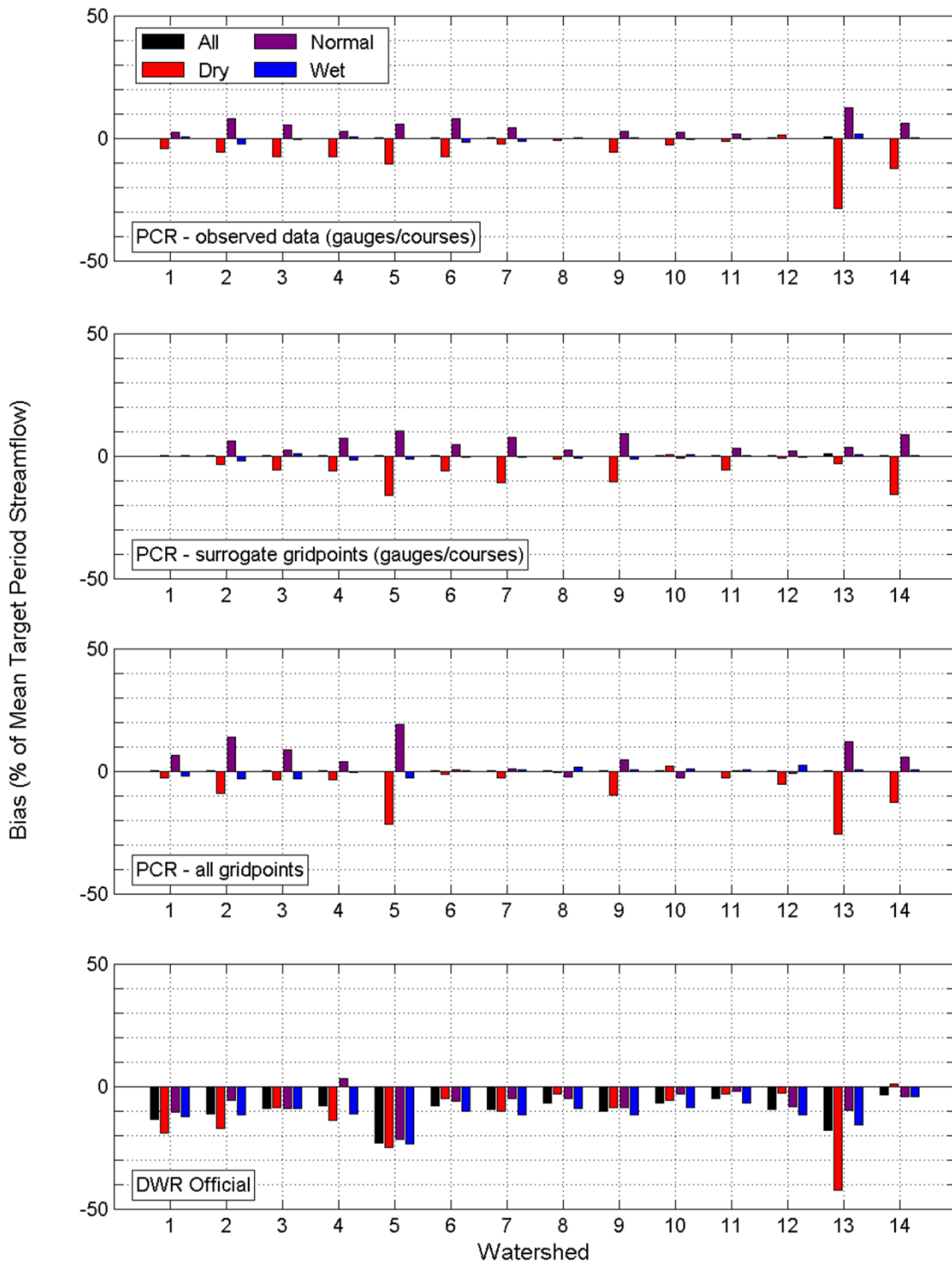


Figure 2.8: April 1 forecast bias as a percentage of mean target period streamflow for (from top to bottom) PCR forecasts using observed data, surrogate gridpoints, all gridpoints, and DWR's official forecasts.

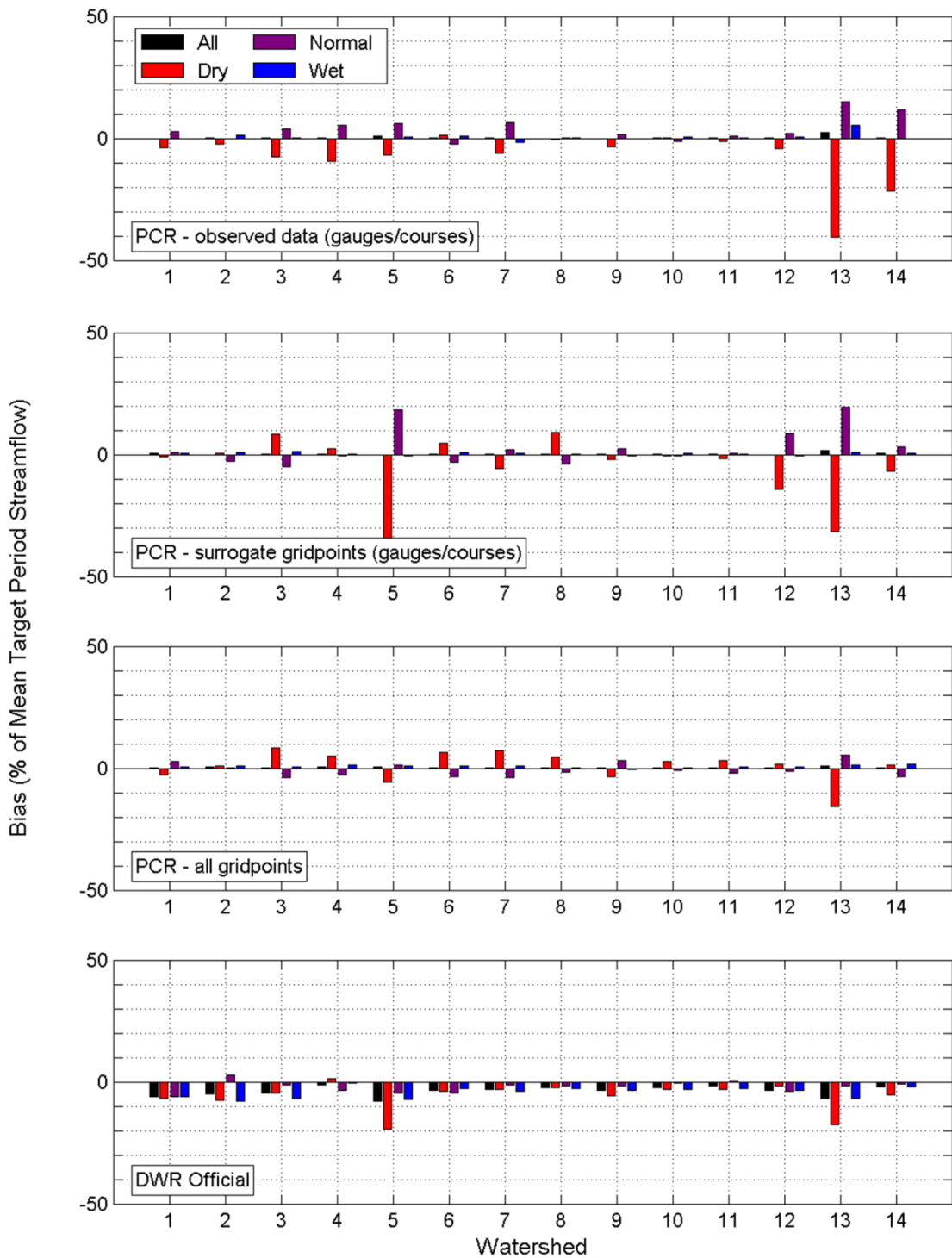


Figure 2.9: May 1 forecast bias as a percentage of mean target period streamflow for (from top to bottom) PCR forecasts using observed data, surrogate gridpoints, all gridpoints, and DWR's official forecasts. Note that PCR forecasts are for a shrinking target period while DWR forecasts are for a constant Apr-Jul target period.

negative bias for each water year type, indicating a tendency to underpredict streamflow. This explains some of the difference in NS scores; where DWR's forecasts are more biased than PCR forecasts (all and wet years), their scores are lower, but for the several watersheds in which their normal and dry year forecasts are less biased, their scores are higher. An example of the latter condition occurs for normal year forecasts in the Sacramento watersheds, due in part to residual patterns that are less linear than in other parts of the domain. Interestingly, DWR's May 1 forecasts (Figure 2.9, bottom) are still negatively biased, but much less so than their April 1 forecasts. PCR forecasts on May 1 are generally also less biased, but for many watersheds DWR's dry year biases are now smaller, thus explaining their better performance in that category.

Despite the superior performance of the hybrid forecasts overall, their limitation in the dry year category warrants comment, especially given California's sensitivity to water scarcity. To its credit, the hybrid model's dry year forecasts perform well in the Sacramento region, which is the source of most of the state's water supply. However, these forecasts are generally less skillful in other watersheds, and thus the model may benefit from a calibration strategy that better recognizes different hydroclimate regimes.

2.4.3 Geospatial analysis of predictors

Applying a search algorithm in combination with PCR also represents a systematic method of determining optimal variable combinations for predictive purposes. In the context of a gridded set of candidate predictor data, the approach offers a means for analyzing predictor locations.

To illustrate the potential utility of the method in this role, we assess predictor locations in the Feather and San Joaquin watersheds (Figures 2.10 and 2.11, respectively). At the upper left of each figure is a topographic map of the watershed's predictor domain, including offsets described earlier. The plots in the lower two rows depict the April 1 SWE (left) and Oct–Mar precipitation (right) predictors that were chosen by the hybrid model for the April 1 forecasts. The black circles in the middle row represent all of the

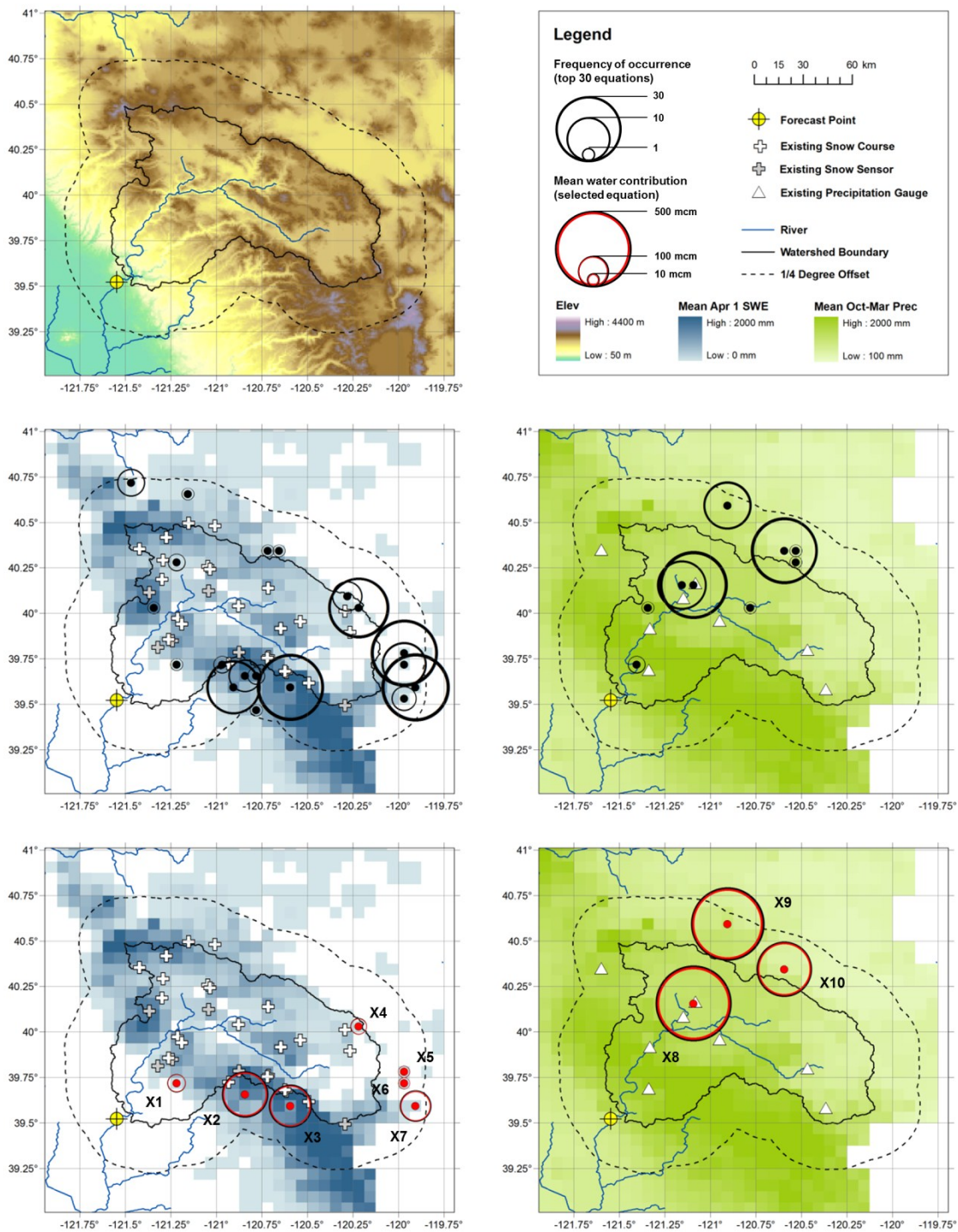


Figure 2.10: (Left) April 1 SWE and (right) October–March precipitation predictor locations for April 1 hybrid forecasts in the Feather. See text for more details.

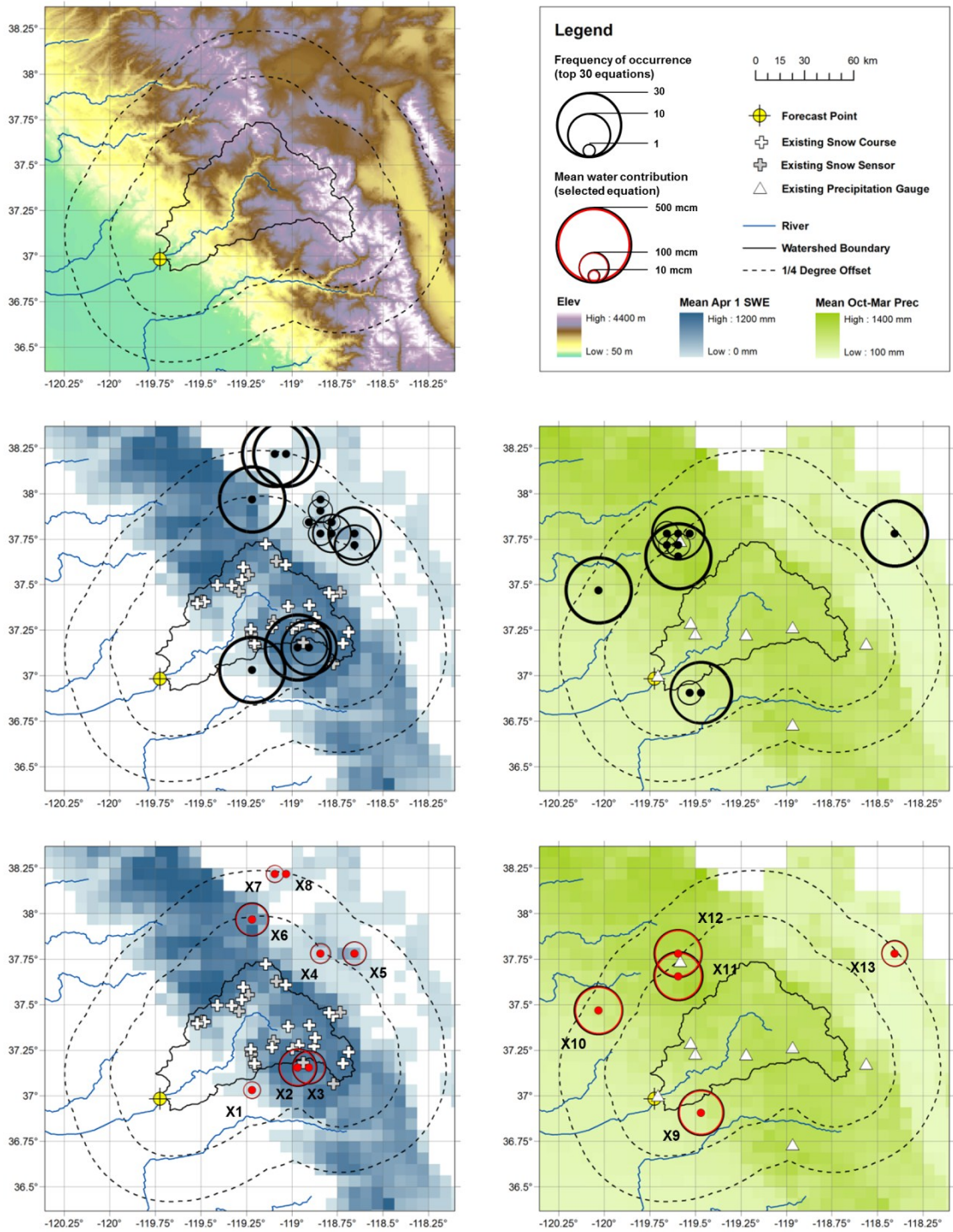


Figure 2.11: (Left) April 1 SWE and (right) October–March precipitation predictor locations for April 1 hybrid forecasts in the San Joaquin. See text for more details.

predictors that appear at least once in any of the top 30 equations, with the size of each circle proportional to the frequency with which the predictor appears in the equations. The red circles in the bottom row represent what we term the “mean water contribution” of each predictor in the final selected equation (i.e., the equation having the median jackknifed standard error of the best 30 models, as described in section 2.3.3). Each predictor’s mean water contribution is the product of its regression coefficient in the selected equation and its mean value over the 50-year period, thus representing the average influence of the predictor on forecasted streamflows.

Underlying the selected predictors are maps of climatology for each predictor type. Grid cells shown in color were used as predictor candidates (i.e., those having a correlation of at least 0.3 with the predictand and at least 10 non-zero values over the 50-year period, as described in section 2.3.3). Some initial patterns are discernable. For both SWE and precipitation, the most frequently selected predictors seem to occur in clusters, indicating that those locations contain important information for streamflow prediction. The relationship of these clusters to climatology is not obvious; some occur in regions of high average values (such as the SWE predictors in the south-central part of the Feather) and others occur in regions of lower average values (such as the SWE predictors in the southeastern corner of the Feather). In terms of mean water contribution, predictors in regions of higher climatology generally have more influence, but those in regions of lower climatology are significant as well. Similar patterns are evident for precipitation predictors and for the San Joaquin watershed.

Additional statistical data for the predictors used in the selected equation are given in Tables 2.4 and 2.5. The correlations and CVs presented are revealing. There is generally a direct relationship between climatological averages and correlation; those predictors with higher means generally have higher correlations as well. On the other hand, climatological average is generally inversely related to CV, with predictors having the lowest means possessing the highest CVs. The two columns at right show the eigenvector loadings for each predictor in the two principal components deemed significant (the Apr 1 equations for both the Feather and San Joaquin just happened to

Table 2.4: Predictor statistics for the selected Apr 1 forecast equation in the Feather. Freq. = frequency of predictor occurrence in the top 30 equations; Corr. = correlation between predictor and predictand; CV = coefficient of variation; Mean = mean climatological value in mm; Reg. Coef. = regression coefficient in the selected equation; and Mean Water Contr. = product of the mean and the regression coefficient, which is equivalent to the mean “contribution” of the predictor to the streamflow forecast in mcm. The elements of PC1 and PC2 represent the loadings in the eigenvectors for each of the predictors, and explained variance was determined from the respective eigenvalues.

ID	Type	Freq.	Corr.	CV	Mean (mm)	Reg. Coef.	Mean Water Contr. (mcm)	PC1	PC2
X1	Apr 1 SWE	1	0.42	2.65	28.22	1.42	40.01	0.22	-0.04
X2	Apr 1 SWE	7	0.81	0.80	428.40	0.59	250.83	0.39	0.14
X3	Apr 1 SWE	29	0.82	0.96	313.50	0.69	217.03	0.39	0.21
X4	Apr 1 SWE	24	0.43	2.80	1.38	22.77	31.35	0.24	-0.46
X5	Apr 1 SWE	29	0.36	3.98	0.83	19.96	16.59	0.21	-0.55
X6	Apr 1 SWE	13	0.36	3.99	0.57	29.14	16.49	0.21	-0.55
X7	Apr 1 SWE	30	0.71	1.45	112.47	1.03	116.15	0.34	0.03
X8	Oct–Mar Prec	30	0.83	0.37	808.40	0.68	552.87	0.39	0.14
X9	Oct–Mar Prec	15	0.83	0.39	668.87	0.77	517.84	0.38	0.19
X10	Oct–Mar Prec	29	0.67	0.59	368.16	0.79	292.32	0.31	0.26
Explained Variance (%)								49	24

contain two principal components each). For both watersheds, the first principal component represents the general spatial distribution of water availability in the basin, i.e., with higher weightings in areas with higher climatological averages. The second principal component reflects variability from that pattern, according higher weights to predictors with lower climatological averages and higher CVs. Note that the regression coefficients (described in section 2.3.1) have been inverted from the principal component transformation to allow expression in terms of the original predictor variables. Thus, with the exception of the y-intercept (which is not shown), the regression coefficient column completely describes the final equation for the April 1 forecast.

Table 2.5: Predictor statistics for the selected Apr 1 forecast equation in the San Joaquin. Abbreviations are as defined for Table 2.4.

ID	Type	Freq.	Corr.	CV	Mean (mm)	Reg. Coef.	Mean Water Contr. (mcm)	PC1	PC2
X1	Apr 1 SWE	30	0.41	3.51	2.86	12.14	34.73	0.10	-0.76
X2	Apr 1 SWE	30	0.96	0.57	681.50	0.25	167.85	0.31	-0.10
X3	Apr 1 SWE	21	0.92	0.65	634.46	0.21	130.89	0.30	-0.04
X4	Apr 1 SWE	4	0.88	1.21	134.67	0.36	48.24	0.30	0.17
X5	Apr 1 SWE	19	0.86	0.86	417.06	0.16	66.94	0.29	0.16
X6	Apr 1 SWE	30	0.88	0.66	465.78	0.28	129.91	0.28	-0.09
X7	Apr 1 SWE	30	0.84	1.33	98.02	0.36	35.46	0.29	0.23
X8	Apr 1 SWE	30	0.66	1.64	59.02	0.11	6.49	0.24	0.41
X9	Oct–Mar Prec	26	0.93	0.40	397.33	0.55	217.34	0.30	-0.05
X10	Oct–Mar Prec	30	0.92	0.40	612.12	0.40	247.36	0.29	-0.17
X11	Oct–Mar Prec	30	0.93	0.39	960.14	0.26	251.46	0.30	-0.16
X12	Oct–Mar Prec	19	0.91	0.40	937.98	0.26	242.94	0.29	-0.17
X13	Oct–Mar Prec	30	0.76	0.63	111.19	0.65	72.04	0.26	0.18
Explained Variance (%)								76	8

The results imply that most of the primary skill is derived from predictors with higher climatological averages, but important information is also contained on the “fringes” of these primary areas. This finding is particularly relevant for SWE, which is “transient” at the lower elevation approaches to the high SWE areas. Most ground-based observations of SWE are located in areas of relatively high and persistent (non-transient) snow accumulation, but these may co-vary more strongly with each other than with less measured transient areas (e.g., those that appear in the second principal components of this analysis). In addition, many of the best predictors are located outside watershed boundaries, a result noted for other watersheds as well. It will not surprise statistical

forecasters that a location need not be “in-basin” to contribute to streamflow predictability. Although this analysis did not formally separate predictor selection, calibration, and validation, and thus some potential for predictor selection bias may exist, jackknifing reduces the risk of this bias by separating validation from predictor selection and calibration.

Also shown on the maps in Figures 2.10 and 2.11 are the locations of existing ground-based observing stations. A comparison of these locations with those of the predictors selected by the hybrid model is instructive. While some predictors, such as the SWE predictors in the south-central part of the Feather, are located in areas already well served by observing stations, many, such as the precipitation predictors in the northeast corner of the Feather, are not. This suggests that distributed model simulations coupled with statistical analysis may provide a useful tool for improving or expanding existing networks, and is an area for future research.

2.5 Conclusions

By combining physically based predictor variables with statistically based prediction methods, we demonstrated a hybrid approach that leverages strengths from both in a real-time, operational forecasting framework. Hybrid forecasts are shown to attain skill comparable to those based on observed data when a selected number of predictor variables are employed, and superior skill when the full set of simulated data are considered. Although this study focuses only on SWE in order to conform with operational practice, various studies have shown a contribution of soil moisture to streamflow predictability as well [*Koster et al.*, 2010; *Maurer et al.*, 2004; *Wood and Lettenmaier*, 2008]; thus along with other simulated fields such as runoff, it is worth examining as a potential input to this framework in future studies.

By simulating SWE at the highest elevations, a hybrid approach also allows the generation of late-season forecasts when most observing stations are snow-free. This feature of the model holds particular value for the catchments of the San Joaquin and

Tulare Lake regions, which contain peaks as high as 4400 m (14,500 ft) and typically experience longer snow persistence. For the San Joaquin, Kings, and Kern specifically, roughly 10% of each watershed lies above the highest snow observations at 3400 m (11,000 ft), indicating sizable areas that are ungauged once the snowline has reached this elevation. Benefits may be most notable for watersheds with relatively small reservoirs that must balance late-season water supply with flood control considerations. The classic example of this scenario is the San Joaquin watershed's Millerton Reservoir, whose capacity of 640 mcm (520 taf) must contend with an average Apr–Jul runoff of 1550 mcm (1260 taf) (Table 2.1). This low storage to runoff ratio prohibits carryover storage, amplifying the shortfall risks associated with late-season over-forecasts that are common in dry years (DWR, 2011, personal communication). In wet years, it is not unusual for Jun–Jul runoff alone to reach 1300 mcm (1060 taf), for which even a respectable 5% under-forecast equates to 65 mcm (53 taf) of unanticipated runoff, or roughly 10% of the reservoir's capacity, at a time when reservoir levels are likely to be high. Similar issues occur in the Merced, Kings, Kaweah, and Kern watersheds, all of which struggle with low storage to runoff ratios and limited downstream capacities to manage snowmelt flooding, the most recent of which occurred in water year 2006 [see, e.g., *Martin*, 2006].

Beyond its forecasting ability, a hybrid model holds potential as a tool for rationalizing predictor locations. While somewhat unique in the context of water supply forecasting, our geospatial analysis is similar to those that have long been common in the atmospheric sciences [see, e.g., *Wallace and Gutzler*, 1981]. One could conceive of additional experiments designed specifically to determine the next best location for an observing station within a ground-based system. Thus, whether hybrid approaches find use in current systems of operational forecasts, statistical analyses of distributed datasets can help us to assess and improve the infrastructure that makes them possible.

Results of the study have been well received by the Hydrology Branch of DWR's Division of Flood Management. Given the well-established use of its current system and the economic and computational expense of a physically based model, however, it is unclear whether DWR would switch to a hybrid approach in the near term (although it is

possible that they could use real-time simulations currently run by CNRFC, which could also be leveraged in case of ground equipment failure). Perhaps most valued, therefore, is the geospatial capacity of the approach. As DWR moves forward in preserving and expanding its data collection network, it has been increasingly asked to justify the cost and environmental impacts of its observing stations (e.g., disturbing a pristine wilderness area with a data collection tower). A hybrid model provides the agency with a tool to rationalize its geographic choices, not to mention the trickledown effect of these locations on data quality and forecast improvements (DWR, 2011, personal communication).

Opportunities for a hybrid approach exist in other parts of the western U.S. as well, particularly in snowmelt-dominant watersheds with relatively sparse observation networks. Possible candidates include those with a considerable percentage of their domain in the National Wilderness Preservation System, for which observation equipment may be restricted [*Landres et al.*, 2003]; examples include the Wind River (Wyoming), Flathead River (Montana), and Gunnison River (Colorado) basins (NRCS, 2011, personal communication). Benefits can also be realized in watersheds with a more transient snowpack, as demonstrated for the Cosumnes and Tule in the present study.

Raw operational forecasts are subject to a great deal of scrutiny and adjustment before they are issued, and so the actual impact of a hybrid system is difficult to predict. Yet the advantages of the approach should not be overlooked. Physically based models forced by mostly low-elevation temperature and precipitation data simulate SWE with biases to be sure, but accurately enough to add value for statistical forecasts. This paper presents one means of exploiting that information resource within an operational framework.

III. INFORMING HYDROMETRIC NETWORK DESIGN FOR STATISTICAL SEASONAL STREAMFLOW FORECASTS

This chapter has been submitted in its current form to the *Journal of Hydrometeorology* [Rosenberg et al., 2012a].

3.1 Introduction

The design of hydrometric networks is a classical problem in hydrometeorology. Central to this issue is that benefits produced by the possession of data should never be less than the data collection costs themselves [Rodríguez-Iturbe and Mejía, 1974]. One can approach the problem by either eliminating redundant stations to reduce costs, or augmenting a network to increase benefits, and countless studies have examined these strategies using a variety of methods for a range of data collection objectives [see, e.g., Mishra and Coulibaly, 2009].

With respect to surface water networks, a number of network design studies have been motivated by the more accurate prediction of streamflow, such as for flood hazard management [e.g., Tsintikidis et al., 2002; Volkmann et al., 2010] or water supply [e.g., Peck and Schaake, 1990]. The methodology used by these studies is necessarily dictated by the prediction model for which the network data are intended. Because many of these prediction models are either physically based or lumped conceptual runoff simulation models, an implicit assumption is that more accurate estimates of mean areal precipitation result in more accurate streamflow predictions, and often interpolation methods such as kriging are applied to historical observations for the network design approach [Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Volkmann et al., 2010]. These methods, however, may be less appropriate for statistical prediction models based on point observations of hydrologic variables.

One such prediction model is that of the Natural Resources Conservation Service (NRCS) of the U.S. Department of Agriculture. Since 1935, a principal responsibility of NRCS has been the publication of water supply forecasts in the American West [*Helms, 2008*]. End users of the forecasts serve a broad array of objectives ranging from irrigated agriculture, flood control, and municipal water supply to endangered species protection, power generation, and recreation [*Pagano et al., 2009*]. Traditionally, NRCS has coordinated with the National Weather Service (NWS) to publish forecasts at the beginning of each month from January through June for several hundred locations throughout the region.

To generate forecasts, NRCS relies on a multivariate principal components regression (PCR) methodology based primarily on point observations of initial hydrologic conditions (IHCs) [*Garen et al., 1992*]. The source of the IHCs is an extensive network of roughly 1200 snow courses and 850 SNOw TELelemetry (SNOTEL) sites scattered throughout the western U.S. (NRCS National Water and Climate Center, personal communication, 2012). Over the past 10–15 years, roughly 40% of SNOTEL stations have been retrofitted with sensors of soil moisture and several other environmental parameters (so-called “enhanced” SNOTEL stations). Data from the NRCS Soil Climate Analysis Network (SCAN) [*Schaefer et al., 2007*] are also available, although most records are still too short for use in statistical forecasts.

New SNOTEL installations have averaged roughly 13 per year since about half of the currently active stations were installed in 1979 and 1980, with the frequency of new installations varying highly from state to state. Equipment costs for a standard installation total \$25,000, a sum that is quickly outpaced by annual operation and maintenance costs (NRCS NWCC, personal communication, 2012). Site selection for new SNOTEL stations is generally an ad hoc process that may be influenced by the offer of funding cooperation from a user group, land availability, road proximity, and a qualitative perception of monitoring network needs. For years, the guiding philosophy was simply to situate new stations at existing snow courses, providing continuous monitoring at locations that otherwise would have monthly or less frequent reports, and

enabling intercomparison between SNOTEL and snow course observations. A Geographic Information System (GIS) tool was recently developed to identify monitoring gaps [Perkins *et al.*, 2010], but does not employ quantitative metrics that evaluate impacts of new stations on streamflow forecast accuracy. Yet budgetary limitations argue that network augmentation to support hydrologic prediction should prioritize drainage basins for which skillful forecasts are most critical and identify sites with the best potential to offer skill improvements.

This paper presents a hydrometric network design approach toward the objective of enhancing statistical prediction models. The specific focus of the paper is the development of a forecast skill-oriented technique for informing NRCS SNOTEL network expansion decisions. We employ a hybrid dynamical-statistical approach that combines the dimension-reducing power of the NRCS PCR methodology with the spatially distributed nature of a physically based, macroscale hydrology model. Principal components analysis is a longstanding network design technique in its own right [e.g., Fiering, 1965; Morin *et al.*, 1979] and, as described previously, physically based hydrologic models have been used extensively in these types of studies as well. The innovations presented herein are in the combination of these techniques for network design and the usage of simulated data as surrogates for point observations in the prediction model.

3.2 Study areas

A collection of study basins was selected to represent the diversity of physiographic, climatic, and existing operational network conditions across the western U.S. Selection criteria stipulated that watersheds be headwater basins, since forecasts for points further downstream are typically based on routed relationships with upstream forecasts rather than point observations of IHCs. NRCS hydrologists were each asked to select 5–10 watersheds of interest from their respective forecast regions, for a total of 24 study basins from six water resource regions (Table 3.1). Reasons for selection tended to vary from basin to basin. Some, such as DTTM8 and HLWM8 in the Missouri water resource

Table 3.1: Identification data for the 24 basins in the study.

NWS ID	NWS Name	USGS ID	Water Resource Region
BITM8	Bitterroot R nr Missoula, MT	12352500	Missouri
BMDC2	Gunnison R at Blue Mesa Res, CO	09124700	Upper Colorado
CHSO3	Sprague R nr Chiloquin, OR	11501000	California
CLLU1	Weber R nr Coalville, UT	10130500	Great Basin
CROW4	Wind R nr Crowheart, WY	06225500	Missouri
DNRC2	Rio Grande R nr Del Norte, CO	08220000	Rio Grande
DTTM8	Teton R nr Dutton, MT	06108000	Missouri
DURU1	Duchesne R nr Randlett, UT	09302000	Upper Colorado
EGLC2	Colorado R nr Dotsero, CO	09070500	Upper Colorado
EGLN5	Cimarron R at Eagle Nest Res, NM	07206000	Arkansas-White-Red
GBRW4	Green R nr Fontenelle Res, WY	09211200	Upper Colorado
HALI1	Big Wood R at Hailey, ID	13139500	Pacific Northwest
HHWM8	S Fork Flathead R at Hungry Horse Dam, MT	12359800	Pacific Northwest
HLWM8	Musselshell R at Harlowton, MT	06120500	Missouri
HRDM8	Little Bighorn R nr Hardin, MT	06294000	Missouri
HREN2	Humboldt R nr Elko, NV	10318500	Great Basin
JLKW4	Snake R at Jackson Lake Dam, WY	13010065	Pacific Northwest
MBLC2	Yampa R nr Maybell, CO	09251000	Upper Colorado
MONO3	N Fork John Day R at Monument, OR	14046000	Pacific Northwest
MPHC2	Dolores R at McPhee Res, CO	09167500	Upper Colorado
NGTC2	N Platte R at Northgate, CO	06620000	Missouri
NVRN5	San Juan R nr Navajo Res, Archuleta, NM	09355500	Upper Colorado
PRTI1	Priest R nr Priest River, ID	12395000	Pacific Northwest
RCEN5	Rio Chama R at El Vado Res Inflow, NM	08285500	Rio Grande

region, were chosen because of highly variable climatologies, rendering forecasting for these basins more difficult. Others are critical forecast points for larger river systems such as the Klamath (CHSO3), Bighorn (CROW4), and North Platte (NGTC2). Those in the Colorado and Rio Grande were selected as a general cross section of the major tributaries across their respective regions. Still others, such as MONO3 and PRTI1 in the

Table 3.2: Summary statistics for the 24 basins in the study. Mean values were calculated over the calibration period. The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation.

NWS ID	Drainage Area (km ²)	Mean Elevation (m)	Forecast Calibration Period	Forecast Target Period	Mean Annual Prec (mm)	Mean Annual Runoff (mm)	Annual Runoff Ratio	Mean Annual Runoff (mcm)	Mean Target Period Runoff (mcm)
BITM8	7290	1780	1981–2010	Apr–Jul	860	280	0.33	2020	1430
BMDC2	8940	3050	1986–2010	Apr–Jul	640	130	0.20	1130	790
CHSO3	4050	1610	1981–2010	Apr–Sep	560	130	0.23	530	260
CLLU1	1130	2430	1981–2010	Apr–Jul	670	170	0.25	200	100
CROW4	4900	2730	1984–2010	Apr–Jul	590	150	0.25	720	470
DNRC2	3420	3220	1986–2010	Apr–Sep	800	220	0.28	770	630
DTTM8	3390	1360	1981–2010	Apr–Jul	440	30	0.07	100	50
DURU1	9820	2390	1986–2010	Apr–Jul	480	70	0.15	680	440
EGLC2	11,380	2900	1981–2010	Apr–Jul	720	210	0.29	2330	1700
EGLN5	430	2870	1981–2010	Mar–Jun	650	40	0.06	20	10
GBRW4	11,090	2470	1981–2010	Apr–Jul	470	120	0.26	1350	910
HALI1	1660	2350	1983–2010	Apr–Jul	740	240	0.32	400	280
HHWM8	3000	1900	1986–2010	Apr–Jul	1260	620	0.49	1850	1460
HLWM8	2910	1750	1984–2010	Apr–Jul	520	30	0.06	100	60
HRDM8	3350	1470	1981–2010	Apr–Jul	460	60	0.13	200	120
HREN2	7200	1930	1981–2010	Apr–Jul	380	30	0.08	230	170
JLKW4	1260	2510	1986–2010	Apr–Jul	1250	630	0.50	790	580
MBLC2	8830	2390	1986–2010	Apr–Jul	680	150	0.22	1310	1080
MONO3	6530	1380	1981–2010	Mar–Jul	610	190	0.31	1250	950
MPHC2	2120	2750	1987–2010	Apr–Jul	610	200	0.33	420	330
NGTC2	3710	2760	1984–2010	Apr–Jul	680	100	0.15	360	270
NVRN5	8440	2520	1987–2010	Apr–Jul	620	150	0.24	1240	840
PRTI1	2340	1150	1984–2010	Apr–Jul	1000	610	0.61	1420	940
RCEN5	2270	2610	1981–2010	Mar–Jul	640	140	0.22	320	280

Pacific Northwest, were chosen because of sparse existing networks but numerous options for new installations.

Table 3.2 presents various physiographic and climatic statistics for each of the basins in the study. Study basin locations are shown in Fig. 3.1, which also includes a map of the National Wilderness Preservation System, for which observation equipment is restricted [e.g., Landres *et al.*, 2003]. As shown, several of the study basins occupy land within this system. Thus, an additional motivation in this study is to determine whether parts of these wilderness areas are important for seasonal streamflow prediction. HHWM8 is a classic example of this scenario, but others (e.g., EGLC2, GBRW4, and NVRN5) contain wilderness areas as well, primarily at higher elevations near basin boundaries.

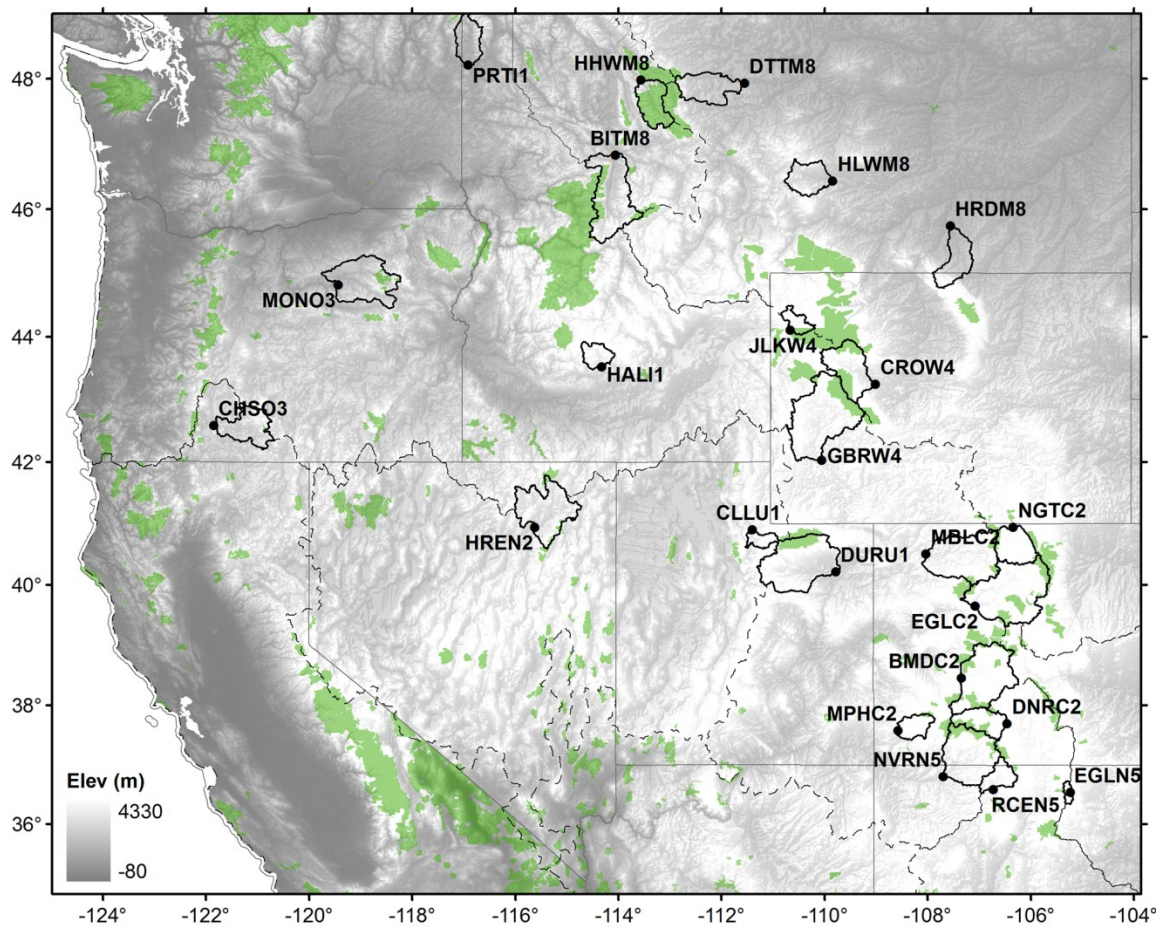


Figure 3.1: The 24 basins in the study. Water resource regions are delineated by dashed lines, and federally protected wilderness areas are shown in green.

3.3 Data and methods

We formulated our approach to specifically address the question, “Given an existing hydrometric network for a particular watershed, where is the next best location to install a SNOTEL station?” The elements of this approach are detailed in the sections that follow.

3.3.1 NRCS methodology

The statistical forecasting approach of NRCS treats each point (i.e., station) observation as an independent predictor in forecasts of seasonal streamflow. Principal components are used to circumvent the well-known problem of predictor collinearity in multivariate regression. Forecast models are developed by restructuring predictor variables into principal components, arranging principal components in order of decreasing eigenvalue (explained variance), developing an equation that sequentially retains only those principal components deemed significant via a t-test, and inverting the transformation so that coefficients are expressed in terms of the original predictor variables. In addition to snow water equivalent (SWE) and water year-to-date precipitation from NRCS snow courses and SNOTEL stations, other predictors such as NWS Cooperative Observer precipitation observations, U.S. Geological Survey streamflow observations, and teleconnection indices are occasionally used.

The NRCS procedure employs an iterative search routine that optimizes variable combinations by developing all possible equations resulting from an increasing number of predictors. With each additional variable, the jackknifed standard error (JSE) is used to order the equations, and the top 30 equations are identified. When the top 30 equations cease to evolve, a final equation is selected by balancing the objectives of a low JSE and month-to-month variable consistency [for details, see *Garen, 1992; Garen and Pagano, 2007*].

For the present study, current operational forecast equations and historical naturalized streamflow observations were obtained directly from NRCS. Table 3.2 lists the forecast

target periods for each of the basins; for those in which forecasts are issued for multiple target periods, the Apr–Jul period was arbitrarily selected for the study. Historical predictor data were downloaded from <http://www.wcc.nrcs.usda.gov/reportGenerator/>. Calibration periods for the forecast models described herein were set to water years 1981 to 2010 unless insufficient predictor or predictand data were available, in which case the calibration period was adjusted as listed in Table 3.2.

3.3.2 *Hybrid approach*

The hybrid framework of *Rosenberg et al.* [2011] exploits the distributed nature of macroscale hydrologic models to expand predictor sets for statistical forecasting applications. As developed for the pilot study of California’s Central Valley drainage, the method uses the Variable Infiltration Capacity (VIC) macroscale hydrology model [*Liang et al.*, 1994] to simulate SWE at a 1/16-degree (~5–7 km) spatial resolution for a given watershed. SWE and the gridded precipitation model forcing data are then used as predictors in the NRCS PCR methodology described in section 3.3.1. *Rosenberg et al.* [2011] found hybrid forecasts to achieve superior skill to observation-based PCR forecasts, with particular benefits late in the snowmelt season, when VIC provides insight into SWE at the highest elevations after snow courses or sensors at lower elevations may already be snow free. As in the NRCS methodology, a noteworthy feature of the hybrid approach is the ability to generate parsimonious models from a large number of predictor candidates, with those in *Rosenberg et al.* [2011] sometimes numbering in the thousands.

In this study, we employed the 1/16-degree VIC forcing and simulated water balance variables of *Livneh et al.* [2012] as predictors for hybrid forecasts. As in *Rosenberg et al.* [2011], predictors with fewer than 10 nonzero entries or correlations with the predictand (i.e., streamflow observations) of less than 0.3 were removed from the candidate pool. In addition, the optimal forecast equation for each forecast month was selected simply as the model with the lowest JSE for that month, and did not incorporate the aforementioned NRCS preference for predictor consistency.

3.3.3 *Augmented predictor analysis*

We examined the predictability afforded by SNOTEL-observed variables to determine those to include in the network design. Variables known to influence the water balance and consequently seasonal streamflow were evaluated. Candidate variables included the traditional predictors of SWE and water year-to-date precipitation, and also two additional variables, soil moisture and air temperature. Soil moisture has long been recognized for its potential to improve seasonal streamflow forecasts [e.g., *Boardman*, 1936; *Clyde*, 1940] and has also been the subject of some recent work on SNOTEL data [e.g., *Lea and Harms*, 2011], but has not been directly incorporated by NRCS in an operational statistical framework. Air temperature influences the cold content of snow, which in turn affects the timing and possibly the efficiency of snowmelt runoff [*Speers et al.*, 1996]. Yet daily temperature information, which is available from most SNOTEL stations, has not, to our knowledge, been used in a statistical runoff prediction context.

The comparison of predictor skill was based entirely on VIC forcings and simulated hydrologic variables. Accumulated water year-to-date precipitation and average water year-to-date minimum/maximum temperature were calculated at monthly intervals from daily forcing data, which were interpolated from NOAA Cooperative Observer station data and scaled (in the case of precipitation) to match climatology data from the parameter-elevation regressions on independent slopes model (PRISM) [for details, see *Maurer et al.*, 2002]. Soil moisture data were taken from VIC simulations for each of three model soil layers (termed SM1, SM2, and SM3) on the first of each month. First-of-the-month SWE data were drawn from simulations for up to five elevation bands for each grid cell. Basin-wide data were compiled for the calibration periods indicated in Table 3.2 and processed through the PCR algorithm described in section 3.3.1. Results corroborated the well-known predictive role of SWE, particularly for those basins at higher elevations. Soil moisture and precipitation, which (like streamflow) can be viewed as a proxy for soil moisture in statistical forecasts [*Speers et al.*, 1996], demonstrated comparable and useful skill in most basins. Temperature, on the other hand, added little skill to forecasts and was excluded from further analysis.

Two experiments were used to examine the potential for improvement beyond current operational forecast skill. In the first (EXP1), gridded precipitation and simulated SWE for all grid cells and snowbands in each basin were added to NRCS predictors in the pool of predictor candidates. In the second (EXP2), basin-wide soil moisture simulations for each of the three soil layers were added to the pool of predictor candidates in EXP1. The results of these experiments estimate the limits of skill that are possible by completely sampling major water balance variables in each basin at the modeled spatial resolution of 1/16 degree.

3.3.4 Network design

The network design approach was formulated as a variation on the experiments described above. Instead of adding modeled data for all grid cells to the pool of predictor candidates at once, we iterated through the grid cells and sequentially added data from just one grid cell at a time. This ensured that improvements in forecast skill were due exclusively to predictor data from that grid cell, whereas in the prior experiments, it would have been impossible to isolate the effects of a single gridded/simulated predictor. Any improvement in JSE over that of the baseline pool of NRCS predictors was then recorded. If multiple elevation bands were present for a given grid cell, this step was repeated for each of the elevation bands, and only the largest improvement was recorded.

We then identified the modeled predictor variable(s) underlying each improvement and generated spatial images of key statistics for each of the five predictor types (SWE, water year-to-date precipitation, SM1, SM2, and SM3). Results were tallied by computing the grid cell offering the best forecast improvement each month and the grid cell offering the best average improvement over all forecast months, which was selected as the best location for a new station. This latter computation was performed in terms of millions of cubic meters, rather than percentage of mean target period streamflow, so as not to give undue weight to forecast improvements later in the snowmelt season. While having the

effect of biasing results towards earlier-season forecasts, this was deemed a more acceptable tradeoff from the perspective of streamflow volume.

3.3.5 Evaluation metrics

A central premise of the study is that gridded/simulated data used in the analysis are suitable proxies for observed conditions. Although some studies have suggested that mean SNOTEL SWE values are generally unrepresentative of mean grid element SWE [e.g., *Molotch and Bales, 2005*], a more meaningful measure for statistical forecasts is the correspondence of variance. Accordingly, we computed correlations between SWE and precipitation observations from existing NRCS stations in each of the 24 basins and gridded/simulated data for the nearest grid cell. Because PRISM data are partially based on NRCS SNOTEL observations, precipitation correlations were expected to be high, but are nonetheless useful. Computations were performed only on data relevant to operational forecasts, i.e., those occurring on the first of the month.

For soil moisture, SNOTEL stations within each of the study basins were scanned for historical observations, and soil moisture records beginning in water year 2006 or earlier were obtained (checks of SCAN stations did not result in any meeting this criteria). This ensured at least five years of data for comparison with simulations, which extended through water year 2010. In total, soil moisture observations were obtained for 59 SNOTEL stations in 17 basins. Each station had records for at least three depths – typically 2”, 8”, and 20” (~50 mm, 200 mm, and 500 mm), and occasionally 4”, 11”, and 40” (~100 mm, 280 mm, and 1000 mm). Soil moisture correlations were computed for the simulated soil layer nearest in depth to each observation (typically SM1 for 2”/4”, SM2 for 8”/11”, and SM2 or SM3 for 20”/40”).

As another means of validating modeled data, we used imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS), specifically the MOD10C2 data set available since February 2000 [*Hall et al., 2006*]. This data set consists of fractional snow cover data at a 0.05-degree spatial resolution for eight-day periods, offering a

significant reduction in the number of cloud-obscured pixels over the daily snow cover products. For each of the locations selected in the network design that were associated with a SWE predictor, we determined the nearest MODIS pixel to the respective VIC grid cell, and extracted data for the eight-day period that fell nearest the first of each month (over water years 2001–2010) for which simulated SWE were included in the prediction model. We then performed binary snow cover comparisons between the MOD10C2 and VIC-simulated data, assigning a value of 1 to any nonzero MOD10C2 fractional snow cover data and any nonzero simulated SWE data.

We also assessed the relative contributions of the predictor variables in each prediction model using standardized partial regression coefficients. Let

$$\hat{y}_i = b_0 + \sum_{k=1}^K b_k x_{i,k}$$

represent a multivariate prediction model, where \hat{y}_i is the prediction in year i , b_0 is the y-intercept, b_k is the partial regression coefficient of the k^{th} predictor, and $x_{i,k}$ is the value of the k^{th} predictor in year i . The standardized partial regression coefficient (or “beta-weight”) of the k^{th} predictor can then be calculated as $\beta_k = b_k (s_{x_{i,k}}/s_y)$, where $s_{x_{i,k}}$ and s_y are the standard deviations of the k^{th} predictor and the predictand, respectively [McCuen, 1985]. We squared the standardized coefficient to obtain the variance in the predictand that is explained by the k^{th} predictor, and divided by the sum of all the squared standardized coefficients to express in percentage form.

3.3.6 *Operational application*

Finally, we evaluated the potential impacts of a new SNOTEL station in the context of operational forecast skill. Recall that an important difference between the NRCS forecasting approach and the one adopted herein involves the NRCS practice of maintaining consistent predictors from month to month. To account for this discrepancy,

we reformulated our optimized forecast models to conform to this approach, i.e., forced these models to include all current NRCS predictors in addition to gridded/simulated data from the grid cell selected in the network design analysis. We employed some subjectivity in these models and only included the new predictor data if an improvement in skill resulted.

As an additional basis for comparison, we generated seasonal streamflow hindcasts for study basins in the Upper Colorado water resource region. Hindcasts were produced using current (as of 2012) operational implementations of the NWS ensemble streamflow prediction (ESP) and statistical water supply (SWS) methodologies for the same periods used in the calibration of our forecast models (Table 3.2). ESP involves forcing the Sacramento Soil Moisture Accounting model with recently observed meteorological data in order to “spin it up” to current conditions, and then with a series of historical meteorological data beginning on the same Julian day to generate an ensemble of projected streamflows [Day, 1985]; the SWS approach is similar to the PCR methodology of NRCS [Hartman and Henkel, 1994]. We also obtained historical as-issued NRCS/NWS coordinated forecasts as available (NRCS NWCC, personal communication, 2012). Because of its utility in comparisons of this type, here we used R^2 between predicted and observed streamflows as our skill metric.

3.4 Results

We present the results of our analyses in the order they were described above. Section 3.4.1 provides skill results for the augmented predictor analysis and the network design exercise, and section 3.4.2 examines the network design results in a geospatial context. Section 3.4.3 presents verification metrics for the gridded/simulated data and evaluates the relative contributions of predictor variables. Section 3.4.4 provides a final assessment of forecast skill from an operational perspective.

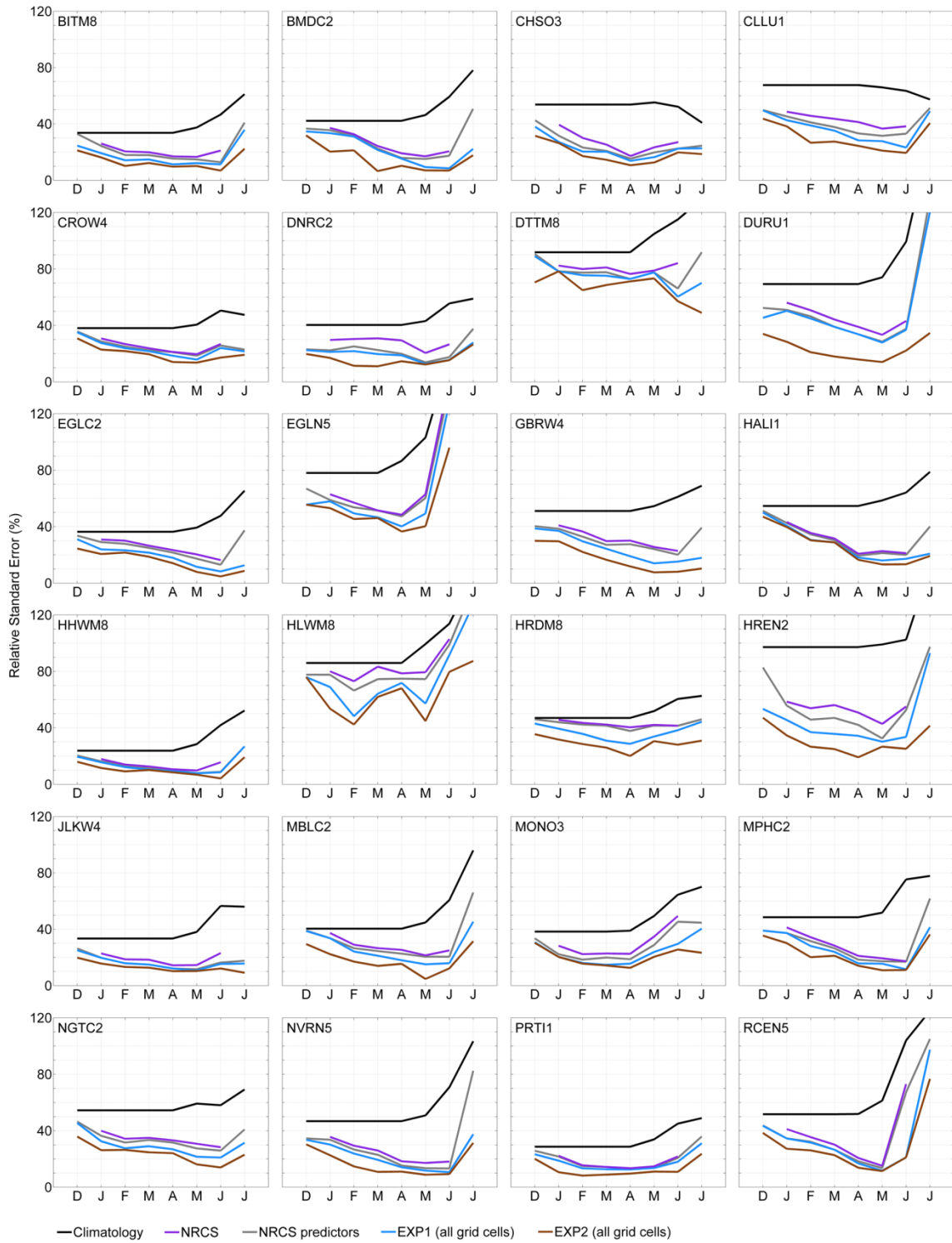


Figure 3.2: Forecast skill for the augmented predictor analysis. The y-axis represents standard error as a percentage of mean streamflow for a shrinking target period. Lines labeled “NRCS” reflect operational NRCS forecast skill; lines labeled “NRCS predictors” reflect forecasts for which the requirement of month-to-month predictor consistency is relaxed.

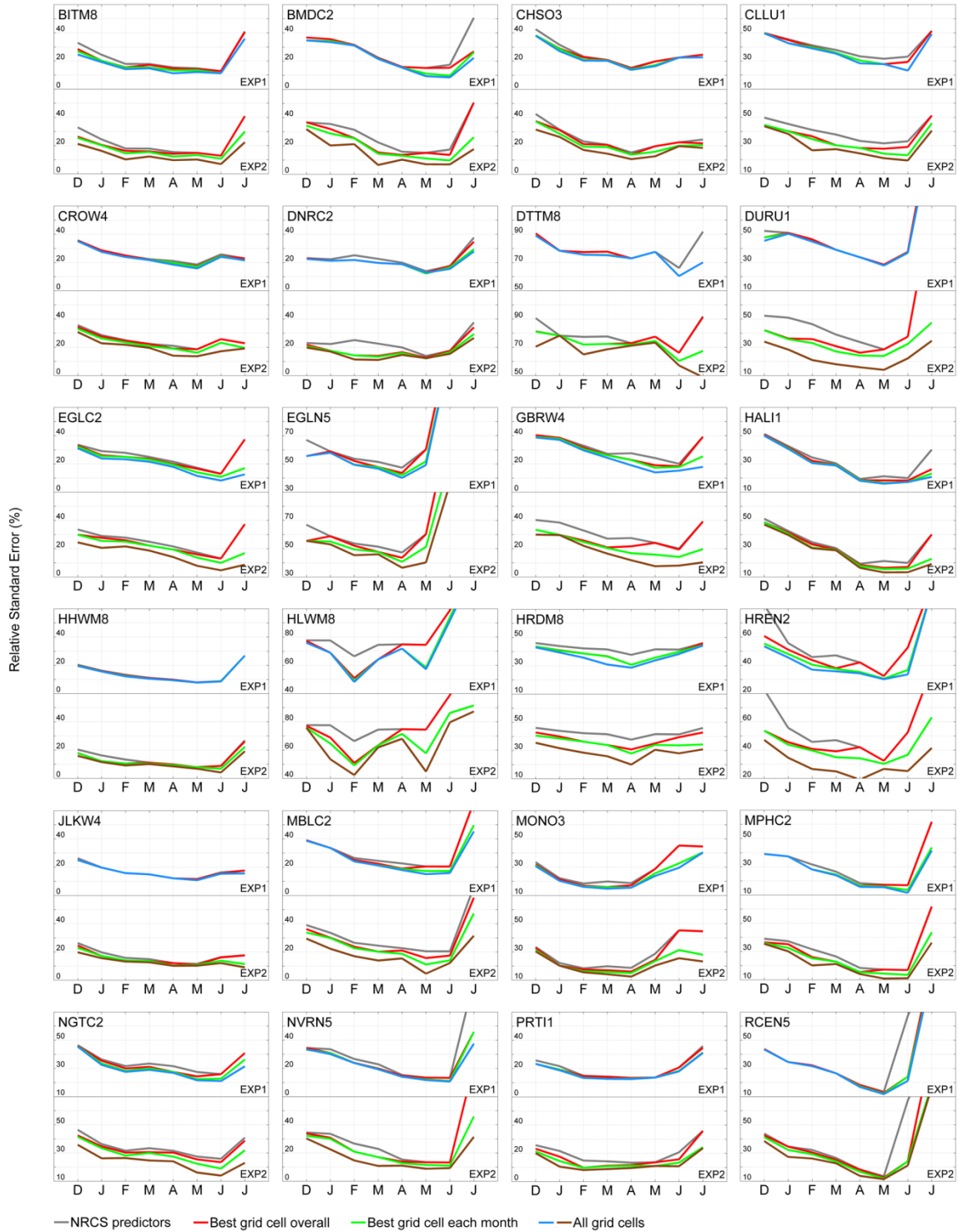


Figure 3.3: Forecast skill for the network design analysis. The y-axis is as in Fig. 3.2, with various labeling schemes as indicated. Note that the lines labeled “NRCS predictors” and “All grid cells” (EXP1 and EXP2) match those in Fig. 3.2. For some plots, lines may be obscured by others with similar values.

3.4.1 Analysis of forecast skill

Fig. 3.2 compares the skill of the different forecasting approaches in the augmented predictor analysis in terms of relative standard error (standard error as a percentage of the mean of the predictand). Here the predictand is streamflow for a shrinking target period, i.e., only the portion of the entire target period remaining for a given forecast month. Differences between NRCS forecast skill with and without the month-to-month consistency requirement are generally small, with the consistency requirement degrading skill between near zero (HALI1, PRTI1) and ~10% (HREN2). In most basins, improvements over the baseline equations (i.e., NRCS predictors without the month-to-month consistency requirement) are fairly nominal for EXP1 (which uses precipitation and SWE as predictors), with the more pronounced improvements occurring in basins that are sparsely sampled (e.g., HLWM8) or later in the snowmelt season as in *Rosenberg et al.* [2011]. Improvements for EXP2 (which also uses soil moisture as predictors), however, are more conspicuous, particularly in basins such as DURU1 and HREN2.

Fig. 3.3 presents results from the network design analysis. For each of the basins, results are given in two panels for EXP1 and EXP2, with the gray lines (the uppermost line in each panel) and the blue and brown lines (the lowermost line in the top and bottom panels, respectively) matching those in Fig. 3.2. Immediately below the gray lines, the red lines represent forecast skill for the best grid cell overall in the network design analysis, while the green lines (immediately below the red lines) represent forecast skill for the best grid cell each month. In general, improvements resulting from the inclusion of predictor data for a single additional grid cell are proportional to those found from the inclusion of predictor data for all grid cells basin-wide, a correspondence that is not surprising. Although improvements are typically possible for at least one grid cell in every forecast month, the grid cell offering the best improvement overall tends to improve forecasts either during the accumulation season or the ablation season, but rarely both. For EXP1, for example, ablation season improvements occur in BMDC2 and DTTM8, while accumulation season improvements occur in HLWM8 and HREN2. For EXP2, most of the best grid cells overall offer improvements exclusively during the

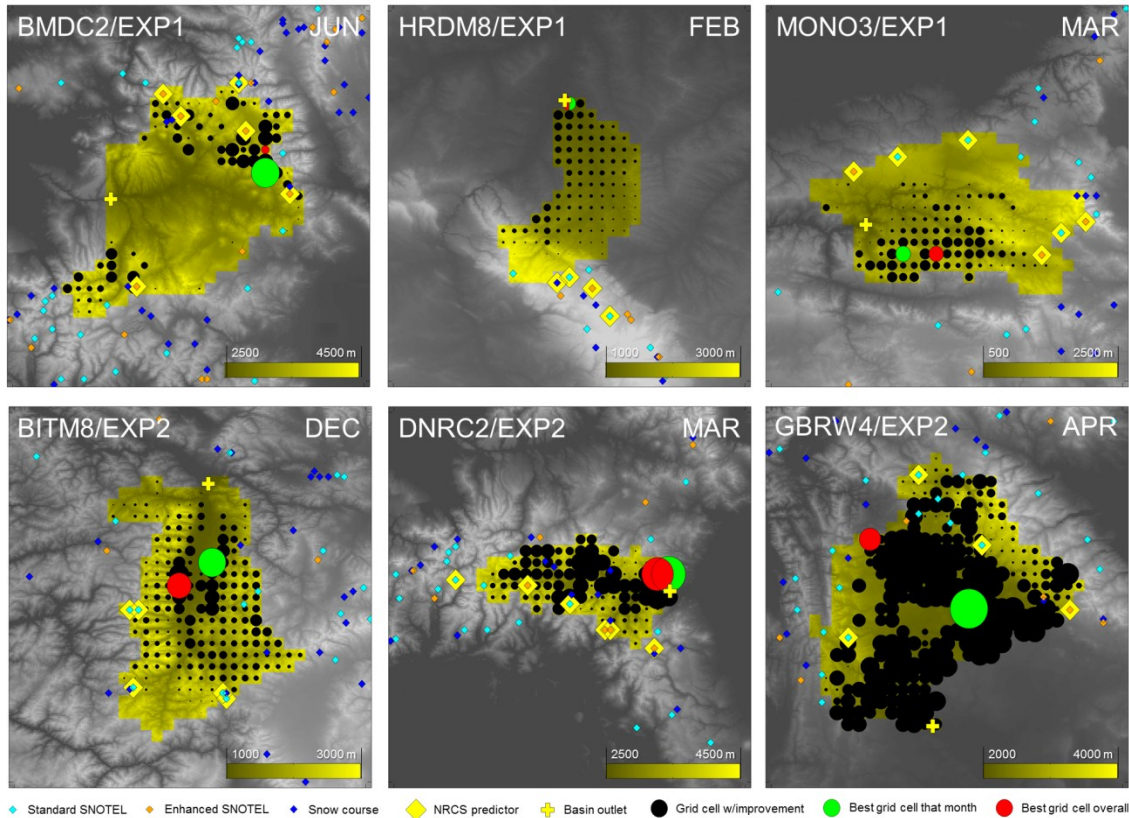


Figure 3.4: Selected spatial results from the network design analysis. The sizes of the black circles are proportional to the improvements in forecast skill offered by their respective grid cells, and the green and red circles correspond to the green and red lines in Fig. 3.3.

accumulation season. This is a consequence of the grid cell locations, which typically differ for the two seasons as described in greater detail below.

3.4.2 Geospatial analysis

Analysis of the spatial patterns associated with the improvements provides insight into the hydrologic mechanisms from which they arise. For a representative set of study basins, Fig. 3.4 shows the locations of grid cells that resulted in forecast improvements for a given month, including the best grid cells that month and over all months. Results for June forecasts using EXP1 predictors at BMDC2 are typical of the late-season improvements that occurred in some of the basins. Comparison of this map with its corresponding plot in Fig. 3.3 indicates that the best grid cell for that month provides an

improvement of 5–10% in relative standard error; the best grid cell overall is situated nearby. Both are located at relatively high elevations, though lower than the basin's peak elevation. Inspection of the top two rows in Fig. 3.5 reveals information about the predictor variables underlying these improvements, the most dominant of which is SWE. The leftmost spatial image depicts average June 1 SWE for those grid cells that were considered predictor candidates (see section 3.3.2). The climatological value for both best overall and month-optimized locations is less than ~200 mm, with the latter location falling on the fringe of the zone with any measurable SWE that time of year. The right two plots in this row reveal that patterns for correlation (with the predictand) tend toward the inverse of those for coefficient of variation (CV, i.e., the ratio of the standard deviation to the mean), a result noted for other basins as well. The selected locations fall in areas of relatively high correlation, but also somewhat higher CV than most of the other grid cells considered as predictor candidates, indicating that these locations add some measure of SWE variability to the predictor pool.

Note that the stations used as predictors in the NRCS forecast models are a subset of all the existing stations in the vicinity of the basin. This subset was selected by NRCS hydrologists to maximize forecast skill, discarding predictor data from other stations because they did not improve forecast performance. The results for BMDC2 also illustrate two limitations of our approach, which is that the locations of existing stations, and issues of scale between the grid and the point element, are not explicitly accounted for in the selection methodology. Forecast improvements in grid cells occupied by existing stations sometimes occurred, particularly in basins with dense existing networks.

A different scenario is presented for February forecasts at HRDM8 in Fig. 3.4. This is an example of a relatively ungauged basin, where NRCS relies on data from nearby stations for predictors in their current forecast models. Improvements tend to increase toward the northwesterly direction, with the largest occurring in the same grid cell as the basin outlet. The corresponding plots in Fig. 3.5 indicate that these improvements are due almost entirely to precipitation, with snow covered area accounting for just a small

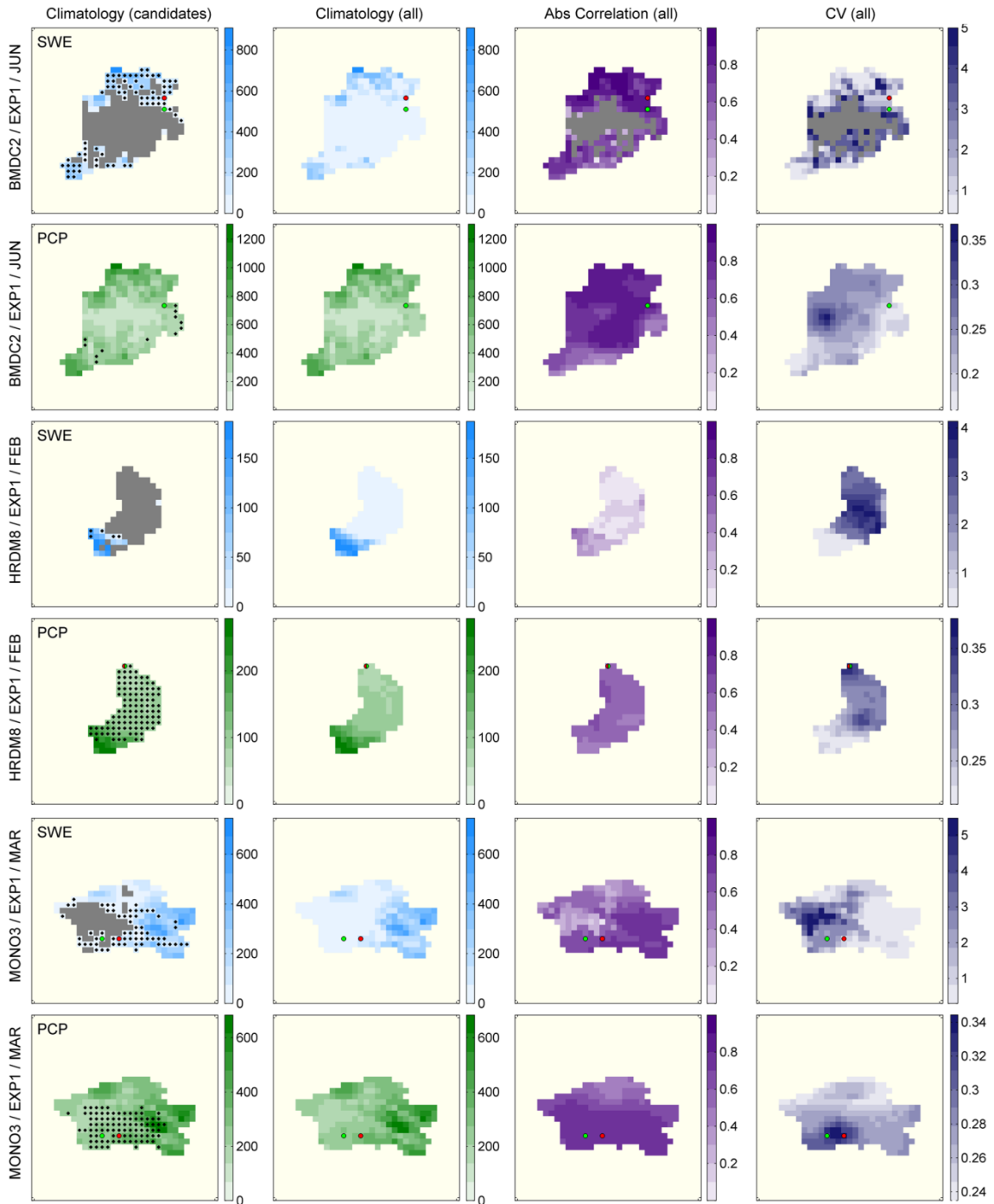


Figure 3.5: Spatial statistics for the dominant predictor types corresponding to the forecast improvements in Fig. 3.4. Grid cells for which the specified predictor type was selected for the forecast model are indicated in the leftmost plots by black dots, with the green and red dots corresponding to the green and red circles in Fig. 3.4. Climatology plots for SWE and water year-to-date precipitation (PCP) are shown in units of mm.

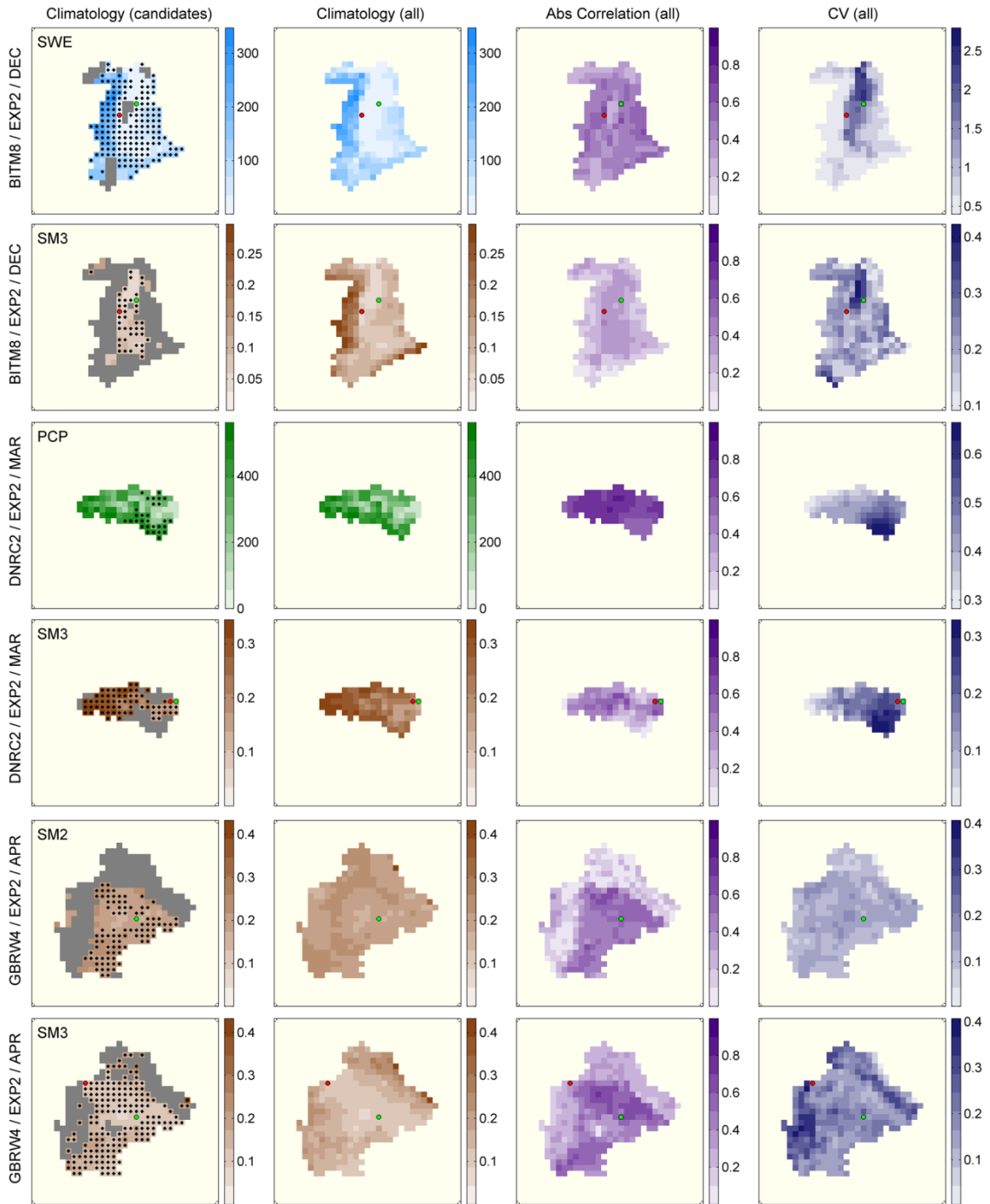


Figure 3.5 (continued): Spatial statistics for the dominant predictor types corresponding to the forecast improvements in Figure 3.4. Climatology plots for SM2 and SM3 are shown in units of volumetric water content.

percentage of the basin at its southern tip. As in the prior example, the grid cell identified as best is located in an area of both high correlation and variability and, additionally, in a region of low intercorrelation with the other predictors in the forecast model. For March forecasts at MONO3, the selected locations also occupy an ungauged area of the basin. Fig. 3.5 indicates that these improvements are due to both SWE and precipitation, suggesting that the ability to exploit the predictive power of multiple variables is yet another reason for selection.

Fig. 3.4 also presents examples of forecast improvements from EXP2, which, as depicted by the sizes of these circles, are larger than those for EXP1. Other commonalities include the forecast months, which are at various points in the accumulation season, and the locations of the grid cells offering improvements, which occur primarily in the basin valleys. The predictor patterns underlying these improvements (Fig. 3.5), however, are slightly different for each basin. For BITM8, the dominant two predictor types are SWE and SM3 (the deeper, larger soil layer). Both best-overall and month-optimized locations are found in areas of high variability and somewhat higher correlations for each predictor type. One striking result is the low

correlation for SM3 at the higher elevations along the basin's periphery, despite these locations being generally wetter than those in the center. This pattern occurs again at DNRC2, where the largest improvements seem to follow the river's course. For GBRW4, SM2 and SM3 combine to offer the largest forecast improvements, also in the lowlands of the basin. Inspection of soil types for areas offering improvements from soil moisture predictors revealed no consistent patterns. For BITM8, DNRC2, and

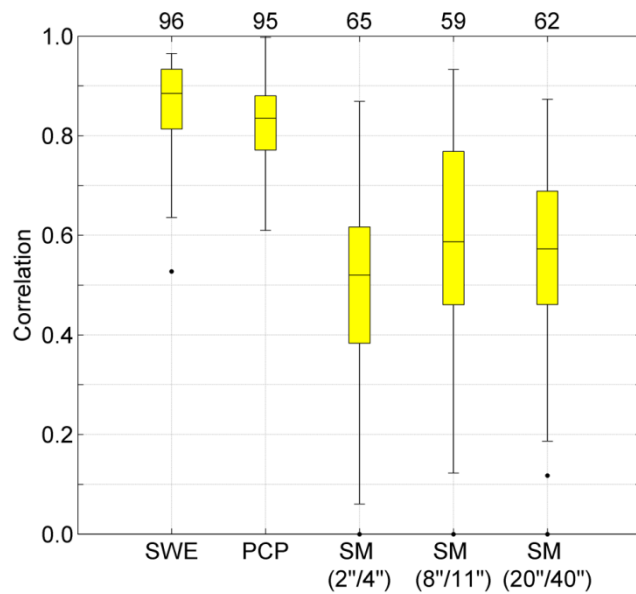


Figure 3.6: Ranges of correlations between observed and gridded/simulated predictor data. The numbers of observed data records are shown at top for their respective predictor types.

GBRW4, for example, predominant soil types include sandy loam, clay loam, and silt loam, respectively, at the depth of the third soil layer [Miller and White, 1998]. Note that, in general, we found SNOTEL installations to be unjustified in wilderness areas; for HHWM8, for example, we found little to no improvement in forecast skill, and in other basins, comparable improvements were found both within and outside wilderness boundaries.

3.4.3 Analysis of predictor variables

A summary of the correlations between station observations and their gridded/simulated counterparts is shown in Fig. 3.6. In general, correlations for SWE and precipitation were strong, while those for soil moisture were more divergent, a result consistent with other studies [see, e.g., Koster et al., 2009]. Since, as noted above, many of the locations selected in the network design analysis were associated with SWE predictors at the fringes of annual snowpack, satellite snow cover comparisons for these sites acquired additional importance. The results (Fig. 3.7) varied from basin to basin, but generally indicated a binary match percentage of about 80%. Thus, the correspondence of observed and gridded/simulated variables generally supported the network design assumptions, particularly for SWE and precipitation, though somewhat less so for soil moisture.

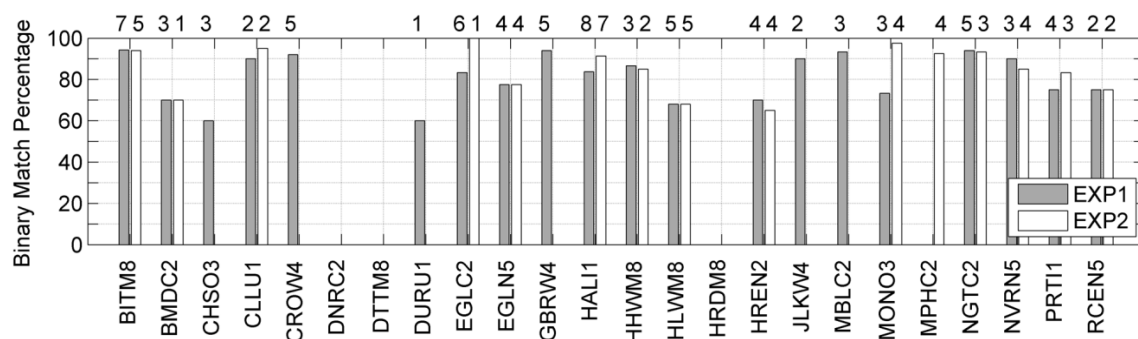


Figure 3.7: Results of binary snow cover comparisons between VIC and MODIS for those grid cells selected as best overall in the network design analysis. The single-digit numbers above each bar represent the number of forecast months for which comparisons were performed.

An examination of predictor contributions for various forecast models and basins reveals some interesting patterns (Fig. 3.8). When only the standard predictors are considered (EXP1), the contribution of precipitation is generally large, while the influence of SWE tends to increase as the year progresses, depending on the basin. When soil moisture is included (EXP2), however, the contribution of precipitation is diminished, while the influence of soil moisture appears stronger in the early part of the water year and weakens as the year progresses. These results are similar to the improvements in forecast skill for EXP1 and EXP2, which tended to be greater during the ablation season and accumulation season, respectively, and consistent with expectations for the western U.S.

[e.g., Wood and Lettenmaier, 2008]. Interestingly, soil moisture predictors are mostly from the bottom two soil layers during the accumulation season, transitioning to the top two layers during the ablation season. Note that, for some of these basins, streamflow observations are used in the NRCS forecast models as described in section 3.3.1. Also note the relatively strong presence of observed predictors for forecast models labeled “All Grid Cells,” even though the vast majority of predictor candidates for these models are gridded or simulated.

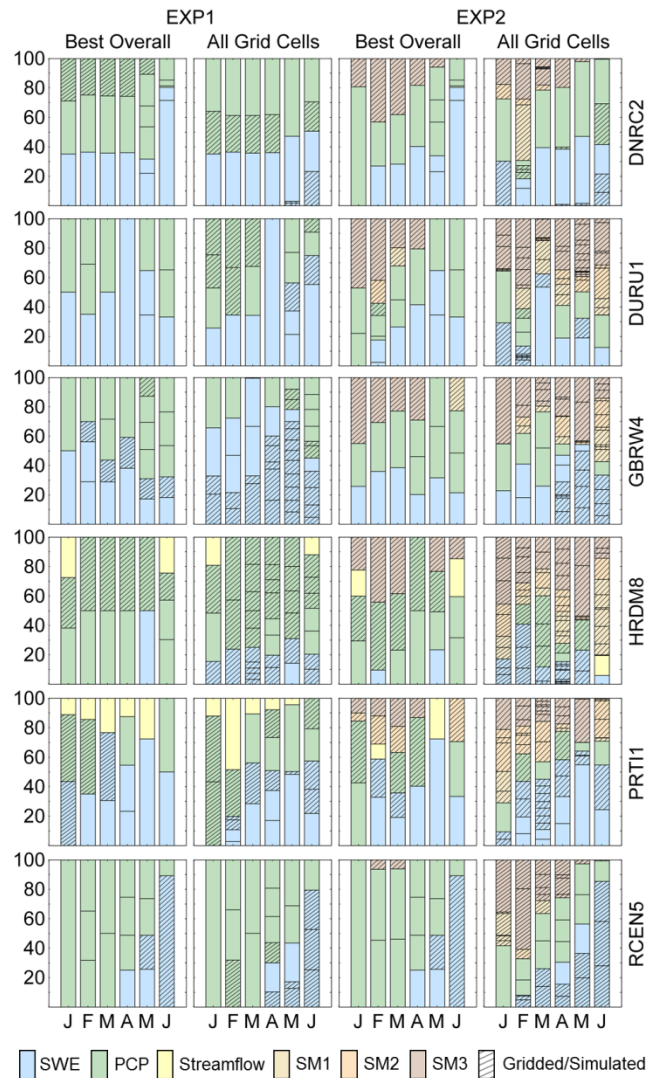


Figure 3.8: Percentage of seasonal streamflow variance that is explained by each predictor for forecast models labeled at top.

3.4.4 Operational analysis

Skill results for the reformulated forecast models are shown for six of the seven Upper Colorado study basins in Fig. 3.9. The shaded areas represent the range of skill that can be expected by adding a predictor to the NRCS forecast models, with the lower bounds determined as described in section 3.3.6 and the upper bounds equivalent to the lines representing the best grid cell overall in Fig. 3.3. Given the subjective nature of the NRCS forecast model selection process, we expect actual skill to fall somewhere in between. Comparisons of these plots with those representing NWS forecast skill reveal a mixture of results. In two of the basins (BMDC2 and NVRN5), current NRCS skill is comparable to or better than that of NWS. In the other four basins, current NWS/ESP forecasts show slightly greater skill than their NRCS counterparts, but the addition of a predictor to the NRCS forecast models is expected to equalize their performance.

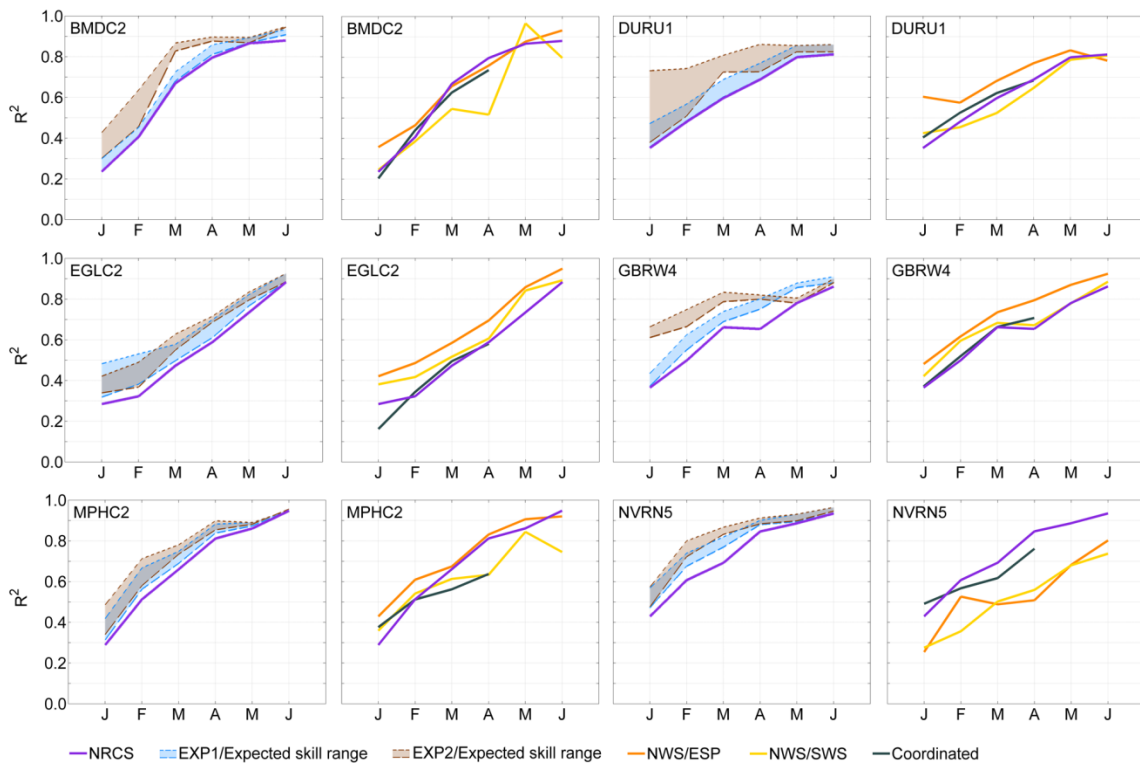


Figure 3.9: Skill of various operational forecasts for 6 of the 7 basins in the Upper Colorado water resource region. Lines labeled “NRCS” correspond to those in Fig. 3.2. Shaded blue and brown areas indicate expected skill for NRCS forecasts that incorporate data from an additional observation station under the EXP1 and EXP2 scenarios.

3.5 Discussion and conclusions

We have demonstrated a skill-oriented, hybrid dynamical-statistical methodology to inform the expansion of hydrometric networks for statistical seasonal streamflow forecasts. While the approach was developed and tested in the western U.S., it is appropriate for any setting in which seasonal streamflow forecast skill is strongly influenced by IHCs. Similarly, the foundation of the approach can be generalized to other water resources applications involving the use of point observations for statistical prediction models, such as those involving groundwater networks or water quality networks.

Evaluation of the method in the western U.S. revealed that locations identified as optimal for SNOTEL placement are primarily concentrated in regions of high correlation with seasonal streamflow, with additional commonalities including high predictor variability, low intercorrelation with existing predictor data, and the ability to exploit the predictive power of multiple water balance variables. When only SWE and precipitation predictors are considered, these tend to occur at the margins of the average snowpack for a given forecast month, though resulting improvements in skill are only notable for basins with sparse existing networks or late in the snow ablation season. One can speculate that the mechanism behind this relationship is related to a second mode of lower-frequency variability that is not captured by predictors with higher climatological averages. Another possibility involves the concept of the snow depletion curve [see, e.g., *Shamir and Georgakakos, 2007*], with the binary snow cover signal at the identified location implicitly providing some indication of snow covered area in the basin and, consequently, the mean areal SWE of the snowpack.

When soil moisture is added as a predictor, improvements in skill are more significant. As above, the mechanism responsible for this result is potentially related to an uncorrelated mode of lower-frequency variability; the power spectra of soil moisture, for example, has been shown to exhibit a strong coherence to that of indicators of climatic extremes like the Palmer Drought Severity Index [*Lakshmi et al., 2004*]. The largest

improvements are found during the accumulation season, when selected locations are typically concentrated in low- to mid-elevations, transitioning to higher elevations as the water year progresses. These patterns are likely also related to snow cover, which shields soil from moisture variations at higher elevations where winter precipitation falls entirely as snow and no winter melt occurs. In contrast, soil moisture at lower elevations, where soil profiles are typically deeper, is more active and better able to reflect the degree of water year-to-date precipitation. As the melting snow exposes the underlying ground surface, the once dormant high elevation soil moisture is again altered in relation to the snowpack volume.

Several recent studies have focused on characterizing the importance of IHCs [e.g., *Wood and Lettenmaier, 2008; Shukla and Lettenmaier, 2011*], and soil moisture in particular [*Koster et al., 2010; Mahanama et al., 2012*], for water supply forecasts. *Mahanama et al. [2012]* found that, outside of spring, the impact of soil moisture initialization on ensemble forecasts dominated over that of snow initialization, and fall soil moisture initialization contributed to skill at particularly long leads. This study shows that such soil moisture predictability can be harnessed in operational statistical water supply forecasts, particularly during the accumulation season via (in most basins) data from low- to mid-elevations. For many basins, even a single soil moisture predictor can enhance forecast skill, and the approach described herein can be used to rank basins in order of those with the most to gain.

The distinction between observed and modeled data represents an important issue in this study. One potential reason for the nominal improvements found for EXP1 is that the gridded precipitation data used in the analysis are partially based on the same SNOTEL observations already in the NRCS statistical models. At the same time, it is uncertain whether the greater forecast improvements found for EXP2 can be expected from in-situ observations. Correlations between simulated and observed soil moisture are generally inconsistent, possibly because of fundamental differences between these two quantities; *Koster et al. [2009]*, for example, have described simulated soil moisture as a “model-specific index of wetness” with no direct observational analog. Nonetheless, the strong

presence of observed predictors in forecast models for which the majority of predictor candidates are simulated or gridded (section 3.4.3) suggests that soil moisture observations should be at least as useful as proxy simulated data for statistical forecasts, although at present, their short record lengths preclude testing this hypothesis. Incidentally, because the methodology described herein employs forcings based on fixed PRISM climatology data, it does not account for spatial distributions of precipitation that may vary in a given year from the long-term mean, which can result in large forecast errors where statistical models are based on point observations [Lundquist *et al.*, 2010]. While installing strategically placed in-situ sensors could alleviate this problem, using the methodology for this purpose would require changes to orographic precipitation gradient assumptions and modifications to the PRISM forcing scheme.

An ancillary finding of this research is that simulated hydrologic data can also be combined with observations to improve operational statistical water supply forecasts, a strategy that may prove more effective than network augmentation in the near term. Indeed, an interesting question is whether advances in computing power and numerical models will render investments in new observations less worthwhile from a forecast skill perspective than those in simulation model- (and ensemble-) based prediction methods. This research demonstrates that, at present, statistical forecasts are comparable in skill to model-based forecasts, and synergies result from their combination. Nonetheless, forecast benefits resulting from SNOTEL installations today are difficult to realize in a statistical framework until enough time has elapsed to develop a statistical climatology. The effects of climate change on hydrologic forecast methods are also a relevant topic, since nonstationarity implies that the statistical properties of future events will no longer resemble those of the past [e.g., Milly *et al.*, 2008; Wood, 2007; Brekke *et al.*, 2010]. We suggest, however, that well-placed observations provide important indications of actual conditions in any climate, and that statistical forecasts will remain useful both for their ability to capture linear predictability at relatively low cost, and as benchmarks against which to evaluate the skill of more intensive prediction approaches.

IV. ON THE CONTRIBUTION OF GROUNDWATER STORAGE TO INTERANNUAL STREAMFLOW ANOMALIES IN THE COLORADO RIVER BASIN

This chapter has been submitted in its current form to *Hydrology and Earth System Sciences* [Rosenberg *et al.*, 2012b].

4.1 Introduction

Among the most important contributors to the skill of a streamflow forecast are the estimation of initial hydrologic conditions (IHCs, i.e., basin water storage at the time of forecast) and prediction of future meteorological anomalies [Mahanama *et al.*, 2012]. Despite some demonstrated skill in seasonal climate forecasting [see, e.g., Stern and Easterling, 1999; Troccoli *et al.*, 2008], most meteorological forecasts for leads longer than about a month are of limited skill [Shukla and Lettenmaier, 2011]. Thus, at seasonal lead times, accurate streamflow forecasts are possible mostly in situations where future runoff is more strongly related to catchment water storage than to meteorological anomalies during the forecast period [Wood and Lettenmaier, 2008; Shukla and Lettenmaier, 2011]. In the American West, this condition is the basis for the statistical seasonal streamflow forecasts issued by the Natural Resources Conservation Service (NRCS) [Garen, 1992] and is an implicit attribute of the dynamically generated ensemble streamflow predictions issued by the National Weather Service (NWS) [Day, 1985].

For many rivers in the western U.S., more than half of the annual streamflow is derived from snowmelt, and snow water storage has historically been the single most significant predictor for statistical streamflow forecasts [Church, 1935]. The opportunity to exploit the relationship between soil moisture and runoff in statistical streamflow forecasts was also recognized in early studies [e.g., Boardman, 1936; Clyde, 1940], although accumulated precipitation is more typically used as a proxy due to a paucity of in-situ soil moisture observations [Speers *et al.*, 1996; Koster *et al.*, 2010]. Nevertheless, recent

modeling studies have demonstrated that early-season soil moisture can have a significant influence on seasonal streamflow, even where annual hydrographs are dominated by spring snowmelt [*Koster et al.*, 2010; *Mahanama et al.*, 2012].

One contribution to basin storage that traditionally has received less attention in streamflow forecasts is groundwater. By sustaining baseflow and laterally redistributing subterranean water, groundwater discharge provides an important link in the hydrologic cycle. With the exception of arid climates where it can be essentially disconnected from the land surface, groundwater also receives surplus during wet periods and offsets deficits during drought [*Fan et al.*, 2007], providing the ability to carry over storage from one year to the next. Although known to be the largest of the storage terms in quantity, however, the magnitude of groundwater's temporal variability relative to that of soil moisture and snow is often poorly understood. How significant are interseasonal and interannual groundwater anomalies for seasonal streamflow forecasts?

The answer to this question has several important implications. One concerns the accuracy of operational seasonal streamflow forecasts, which do not explicitly account for groundwater conditions, although they are sometimes indirectly accounted for by terms such as previous year runoff in statistical forecasts. Another involves the representation of the subsurface in land surface models (LSMs). Notwithstanding their physical basis, LSMs are fundamentally simplifications of natural processes. Until recently, most have lacked a groundwater representation entirely, typically formulating lower boundary conditions either as zero flux or as drainage under gravity [*Maxwell and Miller*, 2005]. Yet such simplifications can significantly bias the estimation of soil water flux and streamflow, and without an explicit representation of the water table, the land surface water budget may not close other than on very long time averages [*Yeh and Eltahir*, 2005]. Consequently, a number of groundwater parameterizations have been proposed [e.g., *Liang et al.*, 2003; *Niu et al.*, 2007], and some research has suggested that inclusion of an explicit aquifer model can reduce the sensitivity of model performance to incorrect parameter choices [*Gulden et al.*, 2007]. Fewer studies, however, have been

devoted to understanding the effect of these modifications on variations in total basin storage, particularly at the interannual scale.

The objective of this paper is to evaluate the contribution of groundwater storage to interannual streamflow anomalies, and hence to seasonal and interannual streamflow predictability, in the Colorado River basin, which is iconic in the American West. It has been described as the most regulated and over-allocated river in the world [NRC, 2007], with some recent research suggesting that current water deliveries are not sustainable [Barnett and Pierce, 2009]. Yet modeling studies of the potential effects of climate change on Colorado River streamflow have been notably incongruent [Hoerling *et al.*, 2009], suggesting a misunderstanding of catchment processes. Thus, while our study is motivated by seasonal streamflow forecasts, an additional interest is in evaluating hydrologic models, which are typically validated only by streamflow, by providing an observation-based assessment of total basin storage anomalies.

We give particular attention to results over the last decade for several reasons. First, conditions have been especially dry in the Colorado River basin during this period [Quinlan, 2010], rendering accurate water supply forecasts particularly important. Second, focusing on the recent past permits better assessment of results from institutional memory. Finally, we are able to supplement hydroclimatic data sets over the last decade with remote sensing observations that were previously unavailable, specifically, estimates of evapotranspiration derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) and total water storage estimates based on the Gravity Recovery and Climate Experiment (GRACE) satellites.

4.2 Study area

The Colorado River flows for 2300 km (1450 mi) through seven U.S. and two Mexican states to its mouth at the Gulf of California (Figure 4.1). Its 630,000 km² (240,000 mi²) drainage area was divided for purposes of the Colorado River Compact of 1922 (and consequently for many water management purposes) into an Upper Basin and a Lower

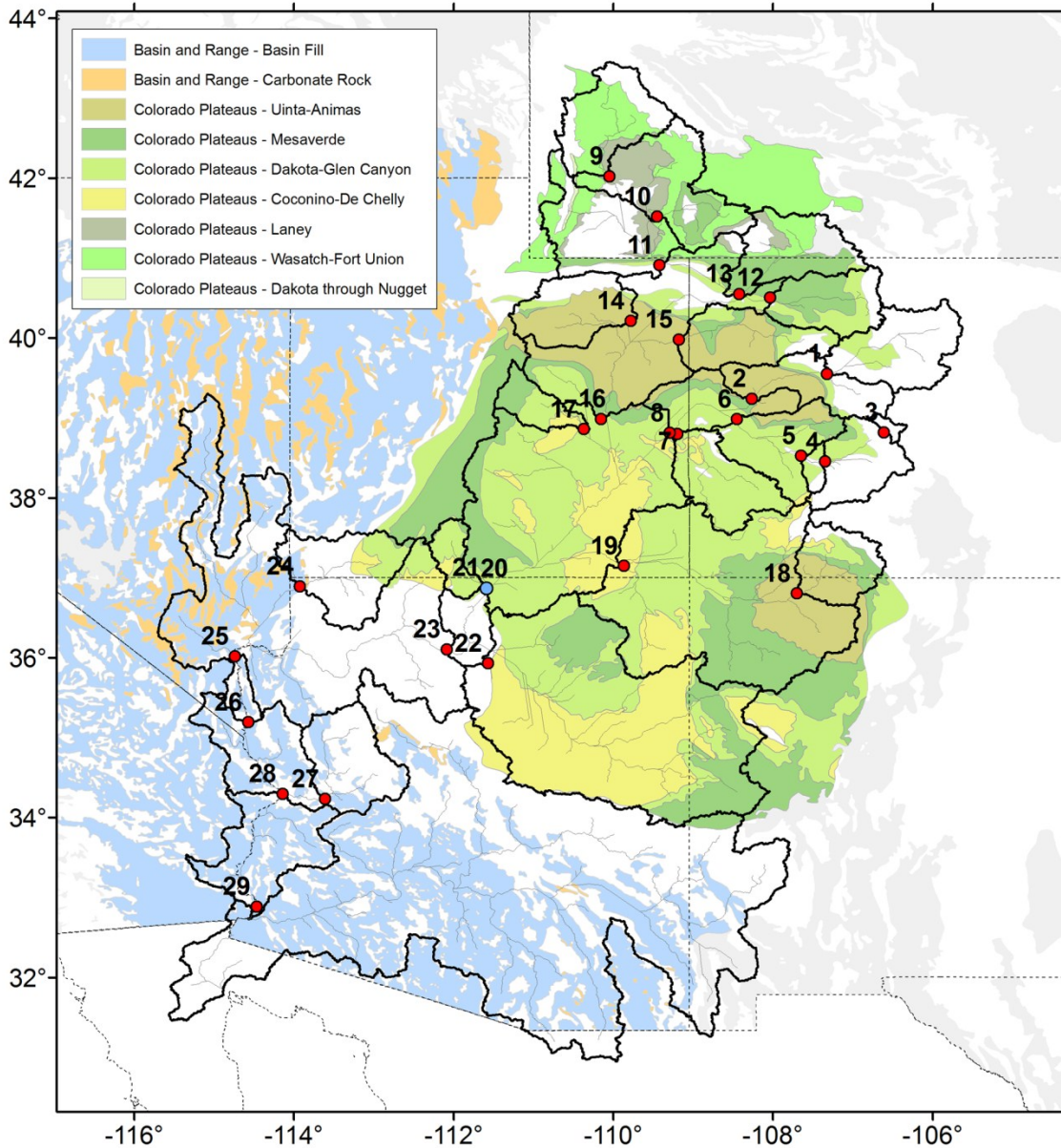


Figure 4.1: The Colorado River basin, including the 29 flow locations monitored by USBR, and the principal aquifer systems as given by Miller (1999), Robson and Banta (1995), and Whitehead (1996). Lee's Ferry (station 20) is indicated in blue.

Basin at Lee's Ferry, Arizona. The hydrograph of the Colorado is dominated by snowmelt, with roughly 70% of its annual streamflow derived from this source. Furthermore, an estimated 85% of its streamflow originates from just 15% of the basin area located in the headwaters of the Southern and Middle Rocky Mountains [Christensen and Lettenmaier, 2007]. The majority of the basin is comprised of desert or

semiarid rangeland, which generally receives less than 250 mm (10 in.) of precipitation per year. Most precipitation in the high elevation streamflow source areas occurs in winter and spring and comes from eastward-tracking Pacific storm systems [Robson and Banta, 1995]. The Colorado has a combined reservoir storage capacity of over 74.0 billion cubic meters (60.0 million acre-feet), or roughly four times the mean annual flow at Lee's Ferry, providing a buffer against a significant temporal variability that includes an historic range in annual streamflow at Lee's Ferry of 6 to 28 bcm (5 to 23 maf) [Christensen and Lettenmaier, 2007; USDOJ, 2000]. 85% of this storage is in Lakes Powell and Mead, operated by the U.S. Bureau of Reclamation (USBR). Table 4.1 lists average annual statistics for the 29 sub-basins in which naturalized streamflow is estimated by USBR.

Three principal aquifer systems, as defined by the U.S. Geological Survey (USGS), are included in the basin (Figure 4.1) [Miller, 1999]. The largest of these is the Colorado Plateaus aquifer system, which contains predominantly sandstone whose porosity is low, such that groundwater moves primarily along joints, fractures, and bedding planes. Surficial aquifers in this system include the Uinta-Animas, Mesaverde, Dakota-Glen Canyon, Coconino-De Chelly, Laney, and Wasatch-Fort Union [Robson and Banta, 1995; Whitehead, 1996]. The remaining two aquifer systems are located within the Lower Basin and include the Basin and Range basin fill aquifers, generally consisting of unconsolidated gravel, sand, silt, and clay, and the Basin and Range carbonate rock aquifers, consisting of limestone and dolomite [Robson and Banta, 1995]. Notably, the Rocky Mountain regions of the basin are not associated with principal aquifer systems.

4.3 Data and methods

Our experimental approach consisted of two stages. First, we estimated monthly and annual changes in total basin storage and its three main elements – snow water equivalent (SWE), soil moisture, and groundwater – using a combination of physically based hydrologic models (section 4.3.1), basin-scale water balances (section 4.3.2), remote sensing data (section 4.3.3), and baseflow recession analyses (section 4.3.4). Second, we

Table 4.1: Average annual statistics for the 29 sub-basins over water years 1950–2008. The annual runoff ratio is defined as the ratio of annual runoff to annual precipitation. The mean annual contribution represents the percentage of the total runoff at Imperial Dam (station 29).

	Drainage Area (km ²)	Annual Precip (mm)	Annual Runoff (mm)	Annual Runoff Ratio	Annual Runoff (bcm)	Mean Annual Contrib
1 Colorado R. at Glenwood Springs, CO	11,805	658	211	0.32	2.50	13.0%
2 Colorado R. near Cameo, CO	20,850	658	202	0.31	4.21	22.0%
3 Taylor R. below Taylor Park Res., CO	658	726	269	0.37	0.18	0.9%
4 Gunnison R. below Blue Mesa Dam, CO	8943	615	139	0.23	1.24	6.5%
5 Gunnison R. at Crystal Res., CO	10,269	618	148	0.24	1.52	7.9%
6 Gunnison R. near Grand Junction, CO	20,534	566	135	0.24	2.77	14.5%
7 Dolores R. near Cisco, UT	11,862	475	79	0.17	0.94	4.9%
8 Colorado R. near Cisco, UT	62,419	558	128	0.23	7.97	41.6%
9 Green R. below Fontenelle Res., WY	11,085	497	144	0.29	1.60	8.4%
10 Green R. below Green River, WY	36,260	364	47	0.13	1.71	8.9%
11 Green R. near Greendale, UT	50,117	367	47	0.13	2.35	12.3%
12 Yampa R. near Maybell, CO	8832	685	166	0.24	1.46	7.6%
13 Little Snake R. near Lily, CO	9661	451	57	0.13	0.56	2.9%
14 Duchesne R. near Randlett, UT	9816	481	95	0.20	0.93	4.9%
15 White R. near Watson, UT	10,412	472	64	0.14	0.67	3.5%
16 Green R. near Green River, UT	116,162	411	55	0.13	6.41	33.5%
17 San Rafael R. near Green River, UT	4217	400	45	0.11	0.19	1.0%
18 San Juan R. near Archuleta, NM	8443	640	144	0.23	1.22	6.4%
19 San Juan R. near Bluff, UT	59,570	362	39	0.11	2.34	12.2%
20 Colorado R. at Lee's Ferry, AZ	289,562	405	61	0.15	17.56	91.6%
21 Paria R. at Lee's Ferry, AZ	3652	296	6	0.02	0.02	0.1%
22 Little Colorado R. near Cameron, AZ	68,529	304	3	0.01	0.19	1.0%
23 Colorado R. near Grand Canyon, AZ	366,744	384	49	0.13	17.98	93.8%
24 Virgin R. at Littlefield, AZ	13,183	370	16	0.04	0.21	1.1%
25 Colorado R. below Hoover Dam, AZ/NV	444,703	368	42	0.11	18.58	97.0%
26 Colorado R. below Davis Dam, AZ/NV	448,847	366	42	0.11	18.88	98.5%
27 Bill Williams R. below Alamo Dam, AZ	11,999	335	9	0.03	0.10	0.5%
28 Colorado R. below Parker Dam, AZ/CA	473,193	360	40	0.11	19.07	99.5%
29 Colorado R. above Imperial Dam, AZ/CA	488,215	353	39	0.11	19.16	100.0%

analyzed the contribution of groundwater storage to interannual streamflow anomalies using the statistical techniques described in section 4.3.5.

4.3.1 Hydrologic models

We estimated total basin storage anomalies using two versions of the Variable Infiltration Capacity (VIC) macroscale hydrology model [Liang *et al.*, 1994]. VIC is a semi-distributed grid-based model that is typical of LSMs used in numerical weather prediction and climate models [Wood and Lettenmaier, 2006], and has been successfully applied in studies of regions across the conterminous US (CONUS) and worldwide [e.g., Sheffield *et al.*, 2009; Mishra *et al.*, 2010; Mahanama *et al.*, 2012]. Like other LSMs, VIC solves the water and energy balance at each time step, but is distinguished by its parameterization of subgrid variability in soil moisture, topography, and vegetation. In the standard release of VIC (4.0.6 in this study, herein referred to as VIC), no distinction is made between saturated and unsaturated zones in the subsurface.

For our second version (herein referred to as VIC-SIMGM), we modified VIC 4.0.6 to incorporate the SIMple Groundwater Model (SIMGM) of Niu *et al.* [2007]. SIMGM is one of several recent models that parameterize groundwater as a lumped, unconfined aquifer beneath a multi-layer soil column [e.g., Gedney and Cox, 2003; Yeh and Eltahir, 2005]. It is included in Community Land Model (CLM) versions 3.5 [Oleson *et al.*, 2008] and 4.0 [Oleson *et al.*, 2010] and Noah-MP [Niu *et al.*, 2011]. In SIMGM, groundwater discharge is parameterized as an exponential function of the water table depth:

$$Q_b = Q_{b\max} e^{-f z_{\nabla}} \quad (1)$$

where $Q_{b\max}$ is the maximum groundwater discharge when the water table depth is zero, z_{∇} is the water table depth, and f is the decay factor. Groundwater recharge (Q_r) is parameterized by Darcy's law and is positive when water enters the aquifer:

$$Q_r = -K_a \frac{-z_T - (\psi_{\text{bot}} - z_{\text{bot}})}{(z_T - z_{\text{bot}})} \quad (2)$$

where K_a is the aquifer hydraulic conductivity, z_{bot} is the depth to the bottom of the soil column, and ψ_{bot} is the matric potential of the bottom soil layer. The time rate of change of aquifer storage (dW_a/dt) is then equal to $Q_r - Q_b$, and the water table depth is computed by scaling aquifer storage by the specific yield (S_y). To incorporate SIMGM in VIC, we added a lumped, unconfined aquifer directly to the base of the lowest (third) soil layer and replaced VIC's baseflow scheme with that of SIMGM (eq. 1). As in CLM, the water table is allowed to move within and between soil layers and the aquifer, in which case eq. 2 is modified following *Niu et al.* [2007]. Hydraulic conductivity between soil layers is computed as a function of soil texture and water content, whereas hydraulic conductivity of the aquifer decays exponentially with depth from the saturated hydraulic conductivity (K_{sat}) of the lowest soil layer. In VIC-SIMGM, the surface runoff parameterization is identical to that of VIC. Because of differences between VIC and CLM, we do not expect the parameter values in eq. 1 to match those of *Niu et al.* [2007].

A limitation of SIMGM is the lack of a direct connection between surface water and groundwater in its parameterization. Because bank storage potentially accounts for a large portion of the interannual hydrologic storage that affects interannual streamflow variations [see, e.g., *Meyboom*, 1961], this likely has implications for our analysis. A second limitation is the lack of a representation of inter-grid cell (lateral) groundwater flow in SIMGM. Nonetheless, with the exception of some very recent work by *Zampieri et al.* [2012], most of the land surface groundwater models that have been proposed have similar drawbacks. Thus, we note these limitations but offer results from VIC-SIMGM for comparative purposes.

Meteorological forcing data were gridded from precipitation and maximum/minimum temperature data from National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer stations and wind data from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis Project. These data were derived for the period 1949 to 2010 at a 1/8-degree

resolution using the methods of *Livneh et al.* [2012], who have extended the data set of *Maurer et al.* [2002]. The first nine months of 1949 were reserved for model spin-up, so that the period of analysis effectively began at the start of water year 1950. Model parameters for VIC were adopted from *Christensen et al.* [2004; subsequently used by *Christensen and Lettenmaier, 2007*]. For VIC-SIMGM, calibration was performed to monthly naturalized streamflow data for water years 1971–1980 by adjusting Q_{bmax} , f , S_y , d_{mid} (the depth of the middle soil layer), d_{bot} (the depth of the bottom soil layer), and b_{inf} (the infiltration shape parameter in VIC). Naturalized streamflow data for the 29 stations in Table 4.1 were obtained for the period 1906 to 2008 from USBR (<http://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html>). For each of the nested sub-basins, calibration was performed in a stepwise fashion, with parameters for the upstream-most sub-basins estimated first. Those parameters were then retained for calibrations to streamflows further downstream.

The VIC-SIMGM implementation required the additional step of first achieving an equilibrium water table depth (WTD) prior to running the simulation. To spin up this

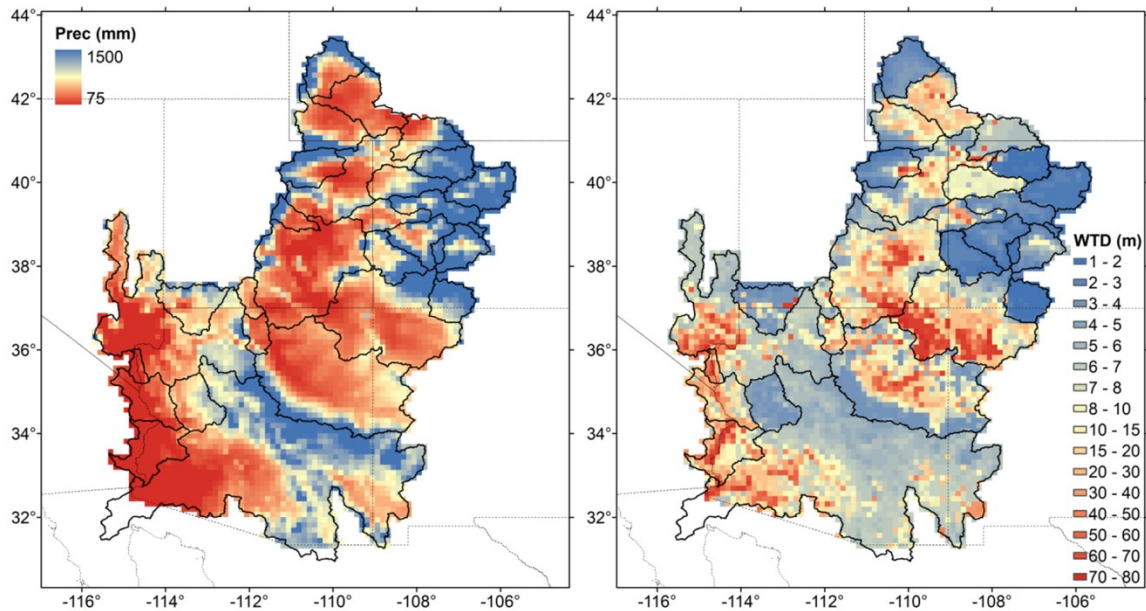


Figure 4.2: Average annual precipitation over water years 1950–2008 (left) and equilibrium water table depth as simulated by SIMGM (right).

term, we ran the model using the observed forcings back-to-back for a period of 2000 years. The process was expedited by removing the exponential decay function for K_{sat} in the aquifer and simultaneously increasing K_{sat} by up to a factor of 10^6 . Results ranged from a depth of 1.3 m to 80 m and generally resemble the pattern of annual precipitation for the basin (Figure 4.2).

4.3.2 Basin-scale water balance

A central question in our study is how well VIC captures the variability in hydrologic storage for the Colorado River basin. This is not unlike a more traditional problem in water resources, which is how well a reservoir's stage-storage relationship captures the variability in its contents. *Langbein* [1960] addressed such a question for Lake Mead by carefully comparing changes in reservoir volumes as determined by a stage-storage relationship from a recent bathymetric survey with changes in volumes as calculated through a water budget of observed inflows and outflows. He found that gains in volume as derived from the water budget exceeded those derived from the stage-storage relationship in wet years, and losses in volume as derived from the water budget exceeded those derived from the stage-storage relationship in dry years, with the magnitudes of the residuals proportional to the changes in volume. He attributed these residual quantities to a "hidden" storage term that was neglected in the stage-storage relationship, namely, bank storage in the sediment of the reservoir, which he estimated at roughly 3 million acre-feet, as compared to the lake's usable capacity of 27 million acre-feet at the time. In a similar study, *Murdock and Calder* [1969] estimated about 6 million acre-feet of bank storage for Lake Powell, which also has a usable capacity of 27 million acre-feet.

In this study, we adopted a similar approach to assess any "hidden" components in the modeled storage quantities by comparing them with changes in storage that are derived exclusively from observational data. Rather than a single reservoir, however, we adopted each of the 29 sub-basins as our control volumes, and computed changes in storage using the hydrologic continuity equation:

$$\Delta S = P - Q - ET \quad (3)$$

We used gridded precipitation data (at a 1/8-degree spatial resolution) and naturalized streamflow taken from the monthly USBR data described in Section 4.3.1. For *ET*, we used a satellite-based product from MODIS [Tang *et al.*, 2009], aggregated from 0.05-degree to a 1/8-degree resolution. The MODIS-based product is one of several recent *ET* datasets derived primarily or entirely from remote sensing data [e.g., Ferguson *et al.*, 2010; Zhang *et al.*, 2010]. These data, arguably, are the nearest alternative to *ET* observations available at this scale.

For the pilot study area of the Klamath River basin, Tang *et al.* [2009] found daily *ET* biases of less than 15% when compared with ground flux tower observations and Landsat-based estimates, with a tendency for the MODIS-based product to underestimate seasonal *ET*. They noted that the algorithm was most effective over areas containing a substantial diversity in vegetation types. In an analysis of a similar product, Ferguson *et al.* [2010] found that satellite-based *ET* was also biased low when compared with VIC-simulated *ET* for the continental US, except for the Colorado and Great Basins, where it was biased high. They suggested that the most likely explanation for the high bias was the lack of a constraint of soil water availability for the remote sensing product.

4.3.3 *Satellite-derived terrestrial water storage change*

As an additional basis for comparison, we analyzed estimates of terrestrial water storage change (TWSC) from the GRACE satellite mission since its 2002 launch. Despite a relatively coarse effective spatial resolution of several hundred kilometers, GRACE data have demonstrated utility for quantifying changes in hydrologic storage in a number of recent studies. Syed *et al.* [2008] found good agreement between GRACE-derived TWSC and Global Land Data Assimilation System simulations at global and continental scales. Strassberg *et al.* [2009] found that GRACE-derived data were highly correlated with in-situ soil moisture and groundwater observations for the High Plains aquifer, and

Grippa *et al.* [2011] showed that GRACE data adequately reproduced the interannual variability of water storage estimated by nine LSMs over West Africa. In a study of nine major US river basins, Gao *et al.* [2010] found that GRACE data tended to underestimate TWSC when compared with VIC simulations, although they noted that errors tended to be of smaller magnitude than those for satellite-based *ET* and precipitation.

The GRACE data sets used here were processed by the Center for Space Research (CSR) at the University of Texas and were filtered to remove spatially correlated errors that result in north-south data “stripes” [Swenson and Wahr, 2006]. Data were provided at a 1-degree spatial resolution for the period 2002–2010 and represent “equivalent water thickness” as computed from observations collected continuously over monthly intervals.

4.3.4 Baseflow recession analysis

We inferred storage changes for subsets of the domain using baseflow recession analysis.

For this we used daily streamflow data from 72 “reference-quality” gages in the GAGES-II database

(http://water.usgs.gov/GIS/metadata/usgs wrd/XML/gagesII_Sept2011.xml;

Figure 4.3). This is an update to a compilation of all USGS stream gages in CONUS that are either currently active or have at least 20 years of complete-year flow records since 1950, and for which watershed boundaries can be reliably delineated [Falcone *et al.*, 2010].

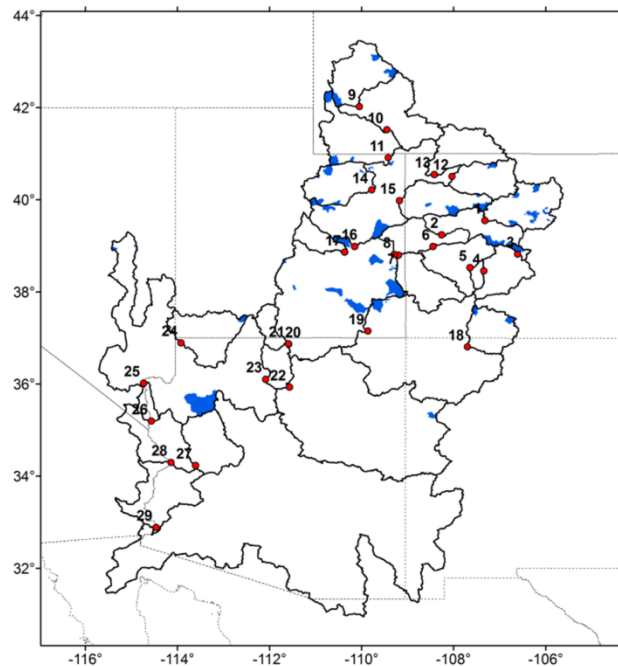


Figure 4.3: The locations of the 72 “reference-quality” watersheds (in blue) used in the baseflow recession analysis. The 29 USBR gauges and sub-basins are shown for reference.

We adopted two separate approaches for our recession analysis. The first,

more conventional, approach utilizes the classical forms of the recession equation as proposed by *Barnes* [1939] and *Maillet* [1905]:

$$Q = Q_0 k_r^t \quad (4)$$

where Q_0 is the streamflow at some arbitrary time $t = 0$ and k_r is the recession constant. For each of the 72 GAGES watersheds, we derived a single recession constant k_r using the semi-automated methodology of RECESS, available for download from <http://water.usgs.gov/software/lists/groundwater> and described by *Rutledge* [1998].

Our second approach is based on the work of *Brutsaert and Nieber* [1977], who described a family of recession curves by the derivative of the nonlinear equivalent of eq. 4:

$$dQ/dt = -aQ^b \quad (5)$$

where a and b are constants. *Vogel and Kroll* [1992] further showed that b can be assumed to be one, in which case only a needs to be estimated from streamflow data. The reciprocal of a has been labeled as the recession timescale τ [*Eng and Milly*, 2007], which is thus equal to $-Q/(dQ/dt)$ and can be used to relate streamflow Q to basin water storage S . For each streamflow record, we identified recession periods using the same criteria as in RECESS, which assumes discharge originates entirely from groundwater at $N = A^{0.2}$ days following each streamflow peak, with A representing the drainage area in square miles [*Linsley et al.*, 1982]. We then followed the methodology of *Kirchner* [2009] and *Krakauer and Temimi* [2011] to compute τ from streamflow observations during these recession periods and develop $S - Q$ curves for each watershed. Values of S were derived for the beginning of each water year (Oct. 1) as those corresponding to the minimum streamflow during the 30-day window of Sep. 16 to Oct. 15, thus assuring at least a high probability that this streamflow was baseflow-dominant, even if it did not fall during a recession period in the strict sense. We then computed annual storage changes for comparison with our other estimates.

Kirchner [2009] further argued that storage-discharge relationships derived from periods of recession could be assumed for other parts of the hydrograph, as evidenced by the ability to invert these relationships to successfully infer $(P - ET)$ from streamflow observations for his two test basins in the United Kingdom. *Krakauer and Temimi* [2011] later used this assumption to estimate monthly storage as the mean of hourly S derived from observed Q , finding that it did not necessarily hold in 61 small watersheds across the U.S. Bearing these results in mind, we tested this approach to also infer monthly storage changes from our daily streamflow data.

4.3.5 Statistical analysis

We performed several first-order statistical analyses on the estimates of storage and storage change. Recognizing that a key issue in our study's main objective relates to the contribution of carryover storage from the previous water year, we proposed the naïve hypothesis that interannual hydrologic storage contributes most to streamflow during years of drought, much like a reservoir is drained to compensate for dry conditions. We tested this hypothesis by evaluating the relationship between change in total water year storage and water year streamflow volume, assuming that, except for losses to evapotranspiration, any negative change in water year storage can be considered a contribution to streamflow from the previous water year. We also examined relationships between water year storage change and Oct. 1 storage anomaly, and water year storage change and previous water year streamflow volume, as a basis for comparison.

We then explored the utility of groundwater estimates for seasonal streamflow forecasts, which are typically issued for the target period April–July in the Colorado River basin by the NWS Colorado Basin River Forecast Center. Here we adapted the dimensionless parameter κ , which was introduced by *Mahanama et al.* [2012] as the ratio of the standard deviation of total basin storage (at a forecast lead of zero) to that of precipitation during the forecast target period. As such it is essentially a comparison between the known and unknown water volumes that determine streamflow, providing an

approximate measure of the predictability that can be derived solely from IHCs. Using ensemble streamflow forecasts for periods ranging from one to six months, *Mahanama et al.* [2012] and *Shukla and Lettenmaier* [2011] found first order relationships between κ and forecast skill for 23 basins and 48 hydrologic sub-regions across CONUS, respectively. We extended the κ concept to measure the predictive capacity of IHCs at leads greater than zero, and also to compare the variability of individual storage terms in addition to total basin storage:

$$\kappa = \sigma_w / \sigma_p \quad (6)$$

where σ_w is the standard deviation of groundwater, soil moisture, SWE, or various combinations thereof on the forecast issue date, and σ_p is the standard deviation of precipitation up to and including the forecast target period. We then developed simple forecasts of April–July streamflow volume via multiple linear regression with the basin-averaged storage terms, and examined the co-variability of the skill (R^2) of these forecasts and our estimates of κ .

4.4 Results

We present our findings in a similar fashion to the methods described above. Sections 4.4.1 to 4.4.3 provide results from the model simulations, satellite data comparisons, and recession analyses in order to establish a most plausible set of storage change estimates. Section 4.4.4 analyzes these estimates to assess the contribution of groundwater storage to interannual streamflow anomalies.

4.4.1 Model performance

Figure 4.4 shows Nash-Sutcliffe scores for both VIC and VIC-SIMGM in each of the 29 sub-basins. The performance of the two models was quite similar. At Lee’s Ferry, scatter plots of simulated vs. observed annual streamflows appeared almost identical,

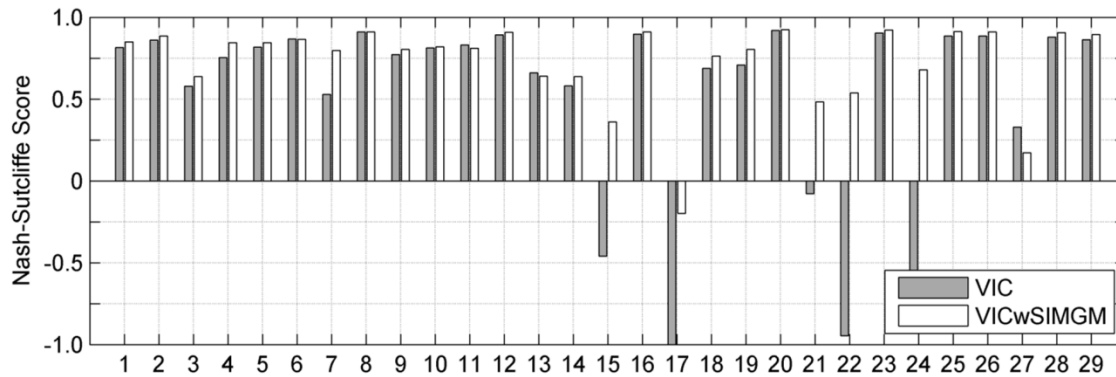


Figure 4.4: Nash-Sutcliffe efficiency scores for VIC and VIC-SIMGM in each of the 29 sub-basins.

showing good agreement between the two terms over the study period. Results for other sub-basins were comparable. For VIC-SIMGM, basin-averaged WTDs exhibited a clear seasonality and ranged from an average of 1.2 m (sub-basin 18) to 27.5 m (sub-basin 19). As a check of these values, we referred to an inventory of USGS wells in CONUS that were opened in surficial, unconfined aquifers hydraulically linked to the land surface [Miguez-Macho *et al.*, 2008]. When we screened for records long enough to evaluate interannual variability, however, only 64 wells in the study domain resulted, including a majority in the Lower Basin where they were less relevant for the purposes of the study. The remaining wells were too sparse to be useful for our analysis.

Details on subsurface storage simulations are given for Lee’s Ferry and its largest headwater basin (Glenwood Springs) in Figure 4.5. As noted by *Niu et al.* [2007] for CLM, VIC-SIMGM resulted in bottom soil layers that were wetter by volumetric water content than in VIC. Differences in wetness were proportional to WTDs and were most evident during the snow accumulation season. On the other hand, bottom layer soil moisture volumes were generally smaller for VIC-SIMGM in sub-basins with shallow WTDs, and total soil moisture anomalies were likewise smaller for VIC-SIMGM. These differences, however, were approximately equal to aquifer storage anomalies, so that total subsurface storage anomalies were roughly the same for both models. The mechanism behind these differences is apparent upon examination of the soil layer depths – for those headwater sub-basins with the shallowest WTDs, calibration resulted in soil columns that

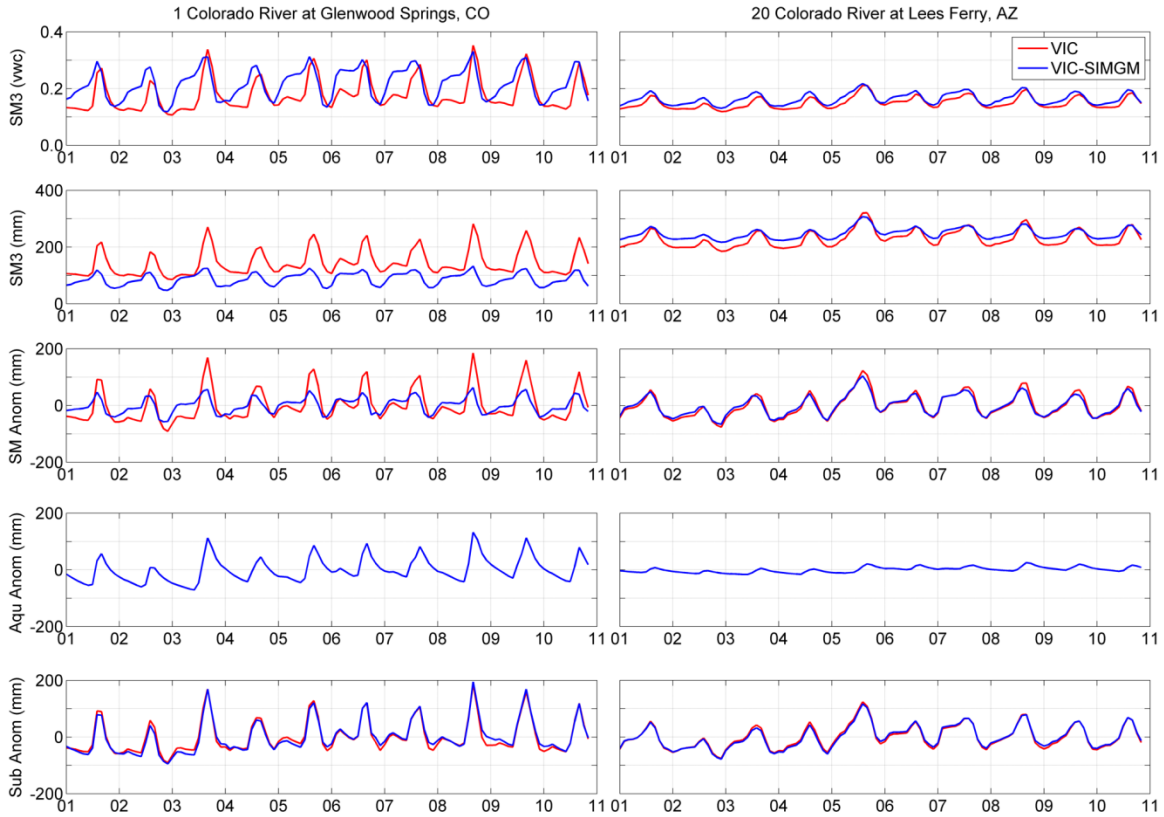


Figure 4.5: Time series of the subsurface storage terms (from top to bottom): third layer soil moisture by volumetric water content, third layer soil moisture in mm, total soil moisture anomaly, aquifer storage anomaly, and total subsurface storage anomaly.

were shallower in VIC-SIMGM than in VIC in order to conserve mass. Differences between the models in total basin storage were likewise negligible.

Comparisons of flux terms for the two models were generally unremarkable, despite different baseflow parameterizations and independent calibrations. In VIC-SIMGM, baseflow constituted a slightly higher percentage of runoff for regions of shallow WTDs, and the annual recession of the simulated hydrograph was noted to typically better match that of the observed hydrograph. *ET* was also somewhat lower in VIC-SIMGM than in VIC, a result that is consistent with other studies of LSM groundwater parameterizations [e.g., Liang *et al.*, 2003].

Modeled storage terms were verified against observations to the extent possible. Measured SWE data were obtained for all active NRCS snow courses and SNOTEL stations within the basin from <http://www.wcc.nrcs.usda.gov/reportGenerator> (116 snow courses and 186 SNOTEL sites total). In each sub-basin, April 1 SWE was computed by averaging the available observations for each water year. VIC-derived estimates were similarly obtained by averaging SWE simulated at the grid cells in which the observations were located. Direct comparisons of these averages generally showed significant differences in magnitude, with observed estimates consistently higher than simulated estimates, a result that is not surprising given the mismatch in scale from point to grid cell. Standardized Z-scores of these estimates, however, compared more favorably.

4.4.2 Satellite-derived storage changes

As a precursor to performing the basin-scale water balance, we compared MODIS-derived estimates of *ET* with VIC estimates of *ET* for each sub-basin. As shown in Figure 4.6 (left), the two estimates generally matched well over the period of MODIS observations. Some slight discrepancies can be seen, such as a tendency for VIC to estimate more *ET* than MODIS for winter months in sub-basin 12 and for summer months in sub-basin 22. The most consistent difference between the two products occurs in water year 2002, for which MODIS estimates substantially more *ET* than VIC. Since 2002 was an extremely dry year (driest or second driest for most sub-basins), this is likely a result of the lack of a constraint of soil water availability for the MODIS product (see Section 4.3.2).

Monthly changes in storage as derived from the basin-scale water balance and VIC are shown in the center column of Figure 4.6, and for comparative purposes, simulated and observed streamflow hydrographs are shown for the same time period at right. The two storage estimates match quite well for all sub-basins, except for water year 2002 due to the previously noted discrepancy in *ET*. Following the work of *Langbein* [1960], annual changes in VIC-simulated storage were plotted against the residuals of the two quantities

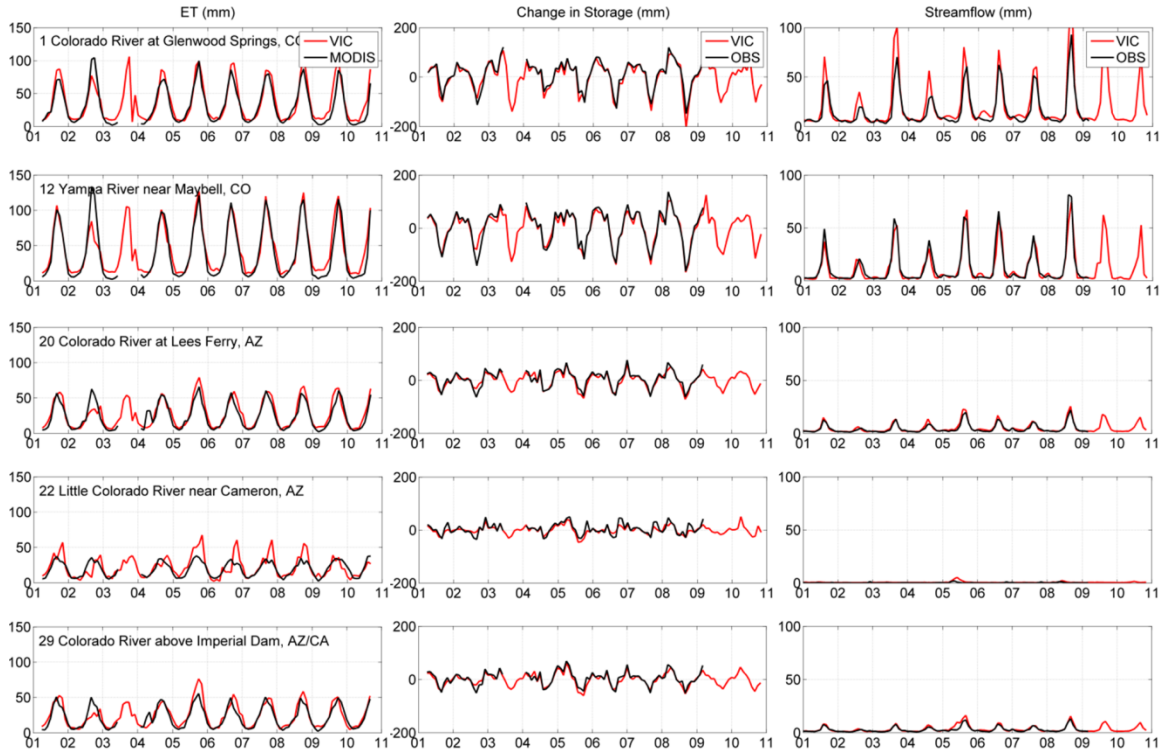


Figure 4.6: Comparisons between MODIS-derived and VIC ET (left), results of the basin-scale water balance (middle), and comparisons between simulated and observed streamflow (right) for selected sub-basins.

(not shown) to infer any element of hydrologic storage that is not captured by VIC (see Section 4.3.2). Despite the good agreement at a monthly time step, however, no consistent relationship was found at an interannual time scale, due in part to the inadequacy of the sample size.

Figure 4.7 compares simulated and GRACE-derived changes in storage for the sub-basins of Lee’s Ferry (20) and Imperial Dam (29). Reasonable agreement between the estimates can be seen, although those derived from GRACE are noticeably noisier than those simulated by VIC. Interestingly, the GRACE data generally indicate more modest changes in storage than VIC, particularly during the summer months of June and July; VIC-SIMGM only slightly reduces this difference.

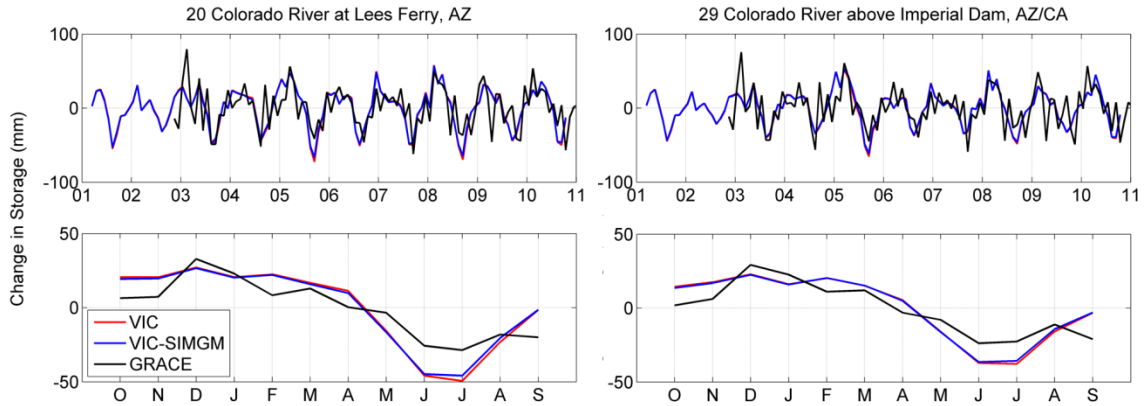


Figure 4.7: Comparisons between VIC-simulated and GRACE-derived changes in storage.

4.4.3 Recession analysis

Baseflow recession constants as derived from RECESS were generally consistent across watersheds. Most results for watersheds in headwater basins fell in the upper part of the typical range at about 0.96–0.99, with only one watershed in sub-basin 9 indicating a recession constant smaller than 0.95. Recession constants for watersheds in the lower part of the Upper Basin were somewhat lower, with a few near 0.85.

Recession plots of $-dQ/dt$ vs. Q are given for four representative watersheds in Figure 4.8. The gray dots in these plots represent all recession observations for each watershed, while the black dots are binned averages following the methodology of *Kirchner* [2009]. Recession timescales (τ) derived from the fitted curves tended to increase with decreasing streamflow but generally were in the 45 ± 15 day range cited by *Brutsaert* [2008]; for the watersheds shown in Figure 4.8, for example, τ ranged from ~ 20 days (gauge #09352900) to ~ 80 days (gauge #09075700) at a Q of 1 mm/day. Storage functions derived from these recession plots are shown in the inset graphs of Figure 4.8 (note that the y-axis on these graphs is labeled with respect to an arbitrary datum), and annual changes in groundwater storage estimated from these functions are shown below their respective recession plots. For comparative purposes, we also show changes in VIC-SIMGM aquifer storage for the grid cell nearest the centroid of each watershed,

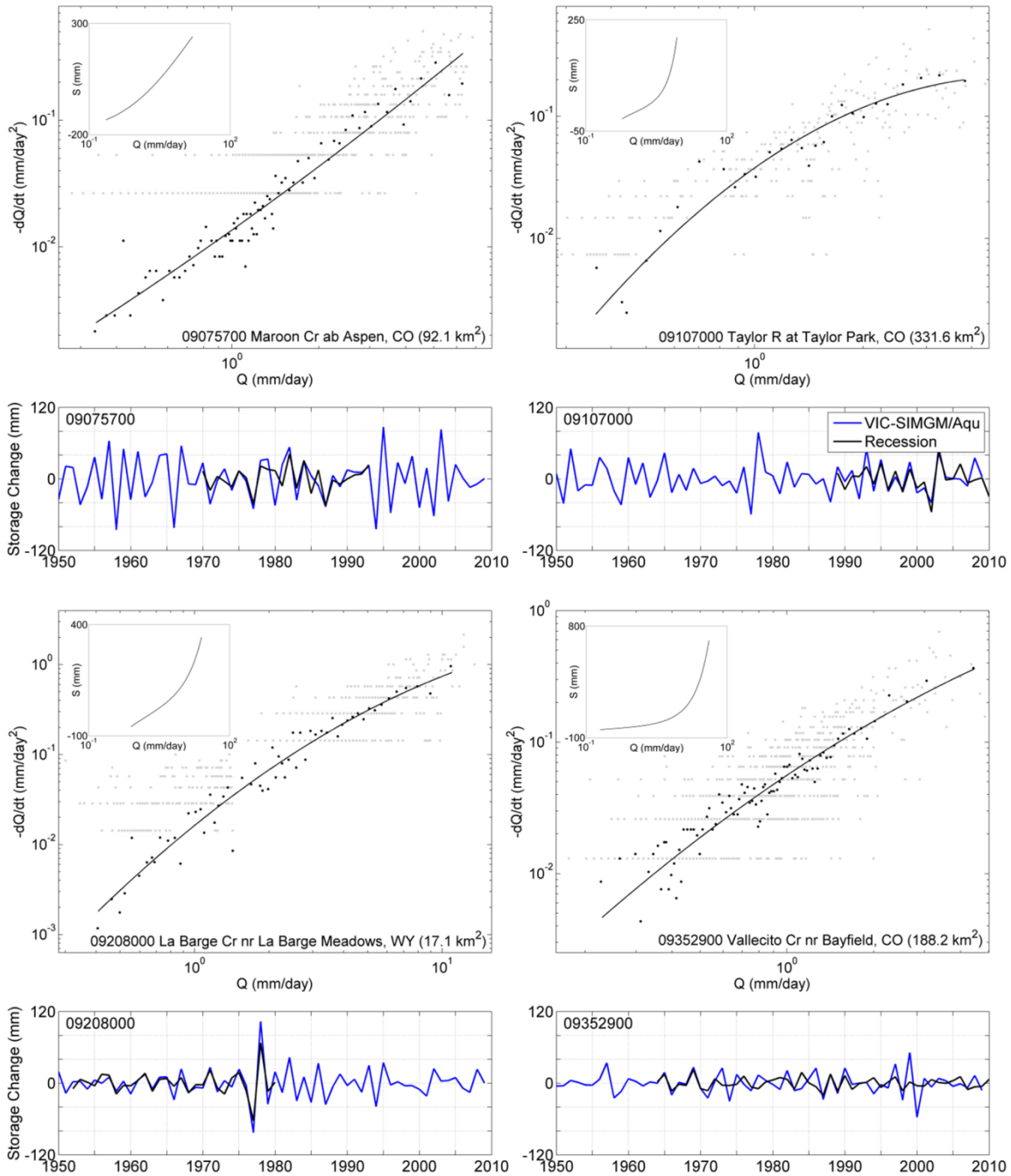


Figure 4.8: Recession plots and associated storage functions (insets) for four reference watersheds. Annual changes in groundwater storage as derived from storage functions are compared with those derived from VIC-SIMGM below their respective recession plots. Watersheds are located in sub-basin 2 (09075700), sub-basin 3 (09107000), sub-basin 9 (09208000), and sub-basin 18 (09352900), with drainage areas provided in parentheses.

which typically covered an area about the same size (and no larger than 2 or 3 grid cells at most). For the watersheds shown here, the ranges of variability in the two estimates are roughly similar, with storage changes for gauge #09107000 and 09208000 matching particularly well. For some of the other watersheds, we noticed a tendency for the simulated groundwater storage to have greater interannual variability than the storage inferred from the recession analysis.

Figure 4.9 presents results from the monthly storage analysis, which assumed that storage-discharge relationships derived from periods of recession could be applied to other parts of the hydrograph as well. To obtain sub-basin-wide estimates of storage change, storage time series were developed for all watersheds with complete records for 2001–2010, and storage averages (weighted by watershed area) were then computed where possible (about half of the sub-basins). Somewhat remarkably, these estimates matched up well with those derived from VIC-SIMGM for sub-

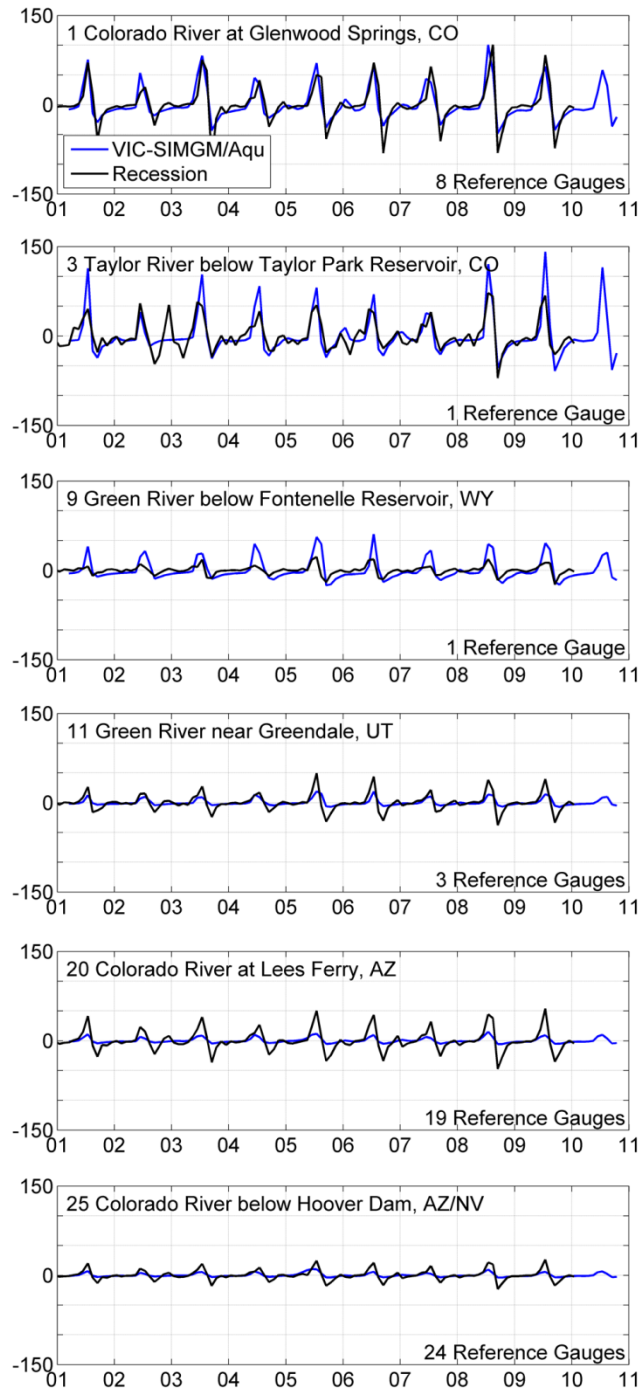


Figure 4.9: Comparison of monthly change in groundwater storage (y-axis, in mm) as derived from VIC-SIMGM and recession analysis. The number of reference watersheds used for each sub-basin is provided at lower right.

basins 1, 3, and (to a lesser extent) 9. The agreement in 3 is perhaps less surprising given that the reference watershed (gauge #09107000, which is also shown in Figure 4.8) accounts for roughly half the area of the sub-basin, but the same cannot be said for sub-basins 1 and 9. For sub-basins further downstream (11, 20, and 25), recession-derived storage changes tend to vary more greatly than those derived from VIC-SIMGM, which can probably be attributed to the headwater bias of the reference watersheds.

4.4.4 Statistical analysis

As tests of the hypothesis that interannual hydrologic storage contributes most to

streamflow during dry years, comparisons among simulated storage estimates and annual streamflow observations are shown for Lee's Ferry in Figure 4.10; for these plots streamflow records were divided into terciles of dry (red circles), normal (yellow circles), and wet (blue circles) water years. As shown in the top two plots (a), correlations between simulated total water year storage change and observed water year streamflow volume are weak at best, with a slight tendency to lose storage in dry years that is modestly more pronounced in VIC-SIMGM.

Correlations between simulated annual storage change and simulated Oct. 1 storage anomalies (b), on the other hand, are somewhat stronger, which is fairly intuitive – what goes

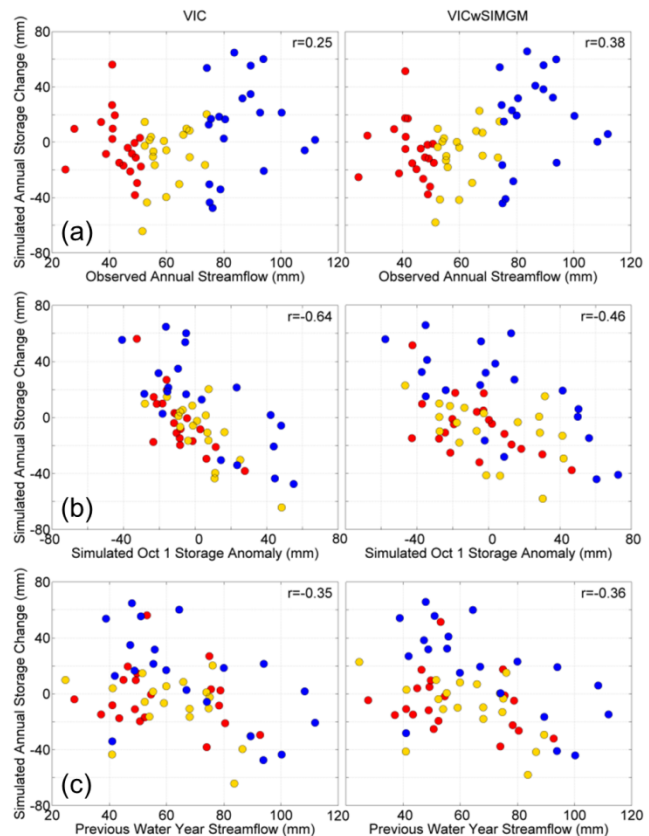


Figure 4.10: Scatter plots for Lee's Ferry of: (a) annual storage change v. annual streamflow, (b) annual storage change v. Oct. 1 storage anomaly, and (c) annual storage change v. previous water year streamflow. Annual storage changes were calculated between Oct. 1 and Oct. 1 of the following year. Red, yellow, and blue circles denote dry, normal, and wet water years, respectively.

up must come down and vice versa. We therefore hypothesized that simulated annual storage changes were related to observed streamflow volumes from the previous water year (c), but again found only weak correlations, due in part to the fact that previous water year streamflow volumes are poor indicators of the previous water year's storage change (and hence Oct. 1 storage anomalies) to begin with, as shown in (a). Results for other sub-basins were comparable, as were comparisons between simulated storage estimates and simulated streamflow volumes. Thus, annual streamflow volumes appear to bear little relation to interannual hydrologic storage, and whether interannual hydrologic storage (and consequently, interannual groundwater storage) contributes to streamflow appears to be more a function of the initial storage conditions, being more or less equally likely in a wet or dry year.

Figure 4.11 provides an alternative perspective on this issue by comparing the range of simulated annual storage change as a percentage of that year's streamflow observation for each of the 29 sub-basins. Expressed in this way, the ranges are smallest for sub-basin 1, the single largest contributor of runoff among the headwater basins, and are minimal for similarly significant sub-basins 2, 8, and 12. On the other hand, ranges are largest for Lower Basin sub-basins that yield much smaller runoff volumes. Thus, interannual hydrologic storage appears of least importance where streamflow matters most, and most

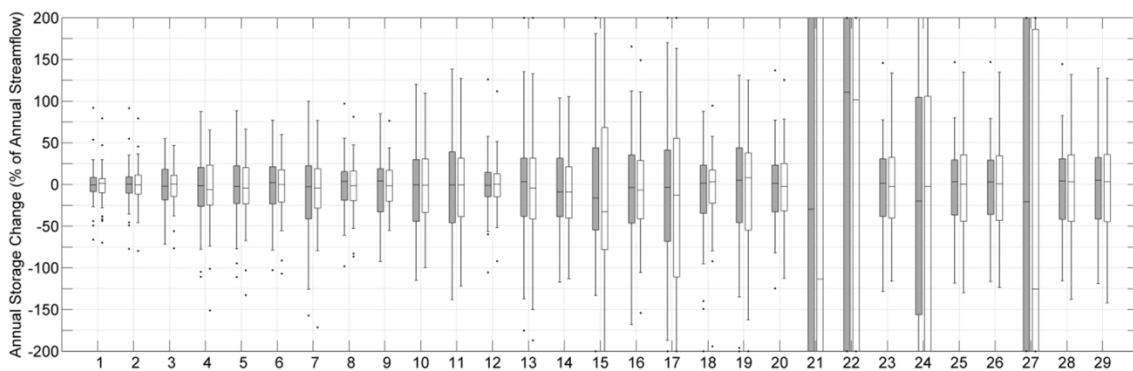


Figure 4.11: Ranges of simulated annual storage change as a percentage of annual streamflow observations for VIC (gray boxes) and VIC-SIMGM (white boxes). The central mark in each box indicates the median, and the edges of the box are the 25th and 75th percentiles. Whiskers extend to a maximum of ~2.7 standard deviations from the median, and outliers are indicated by dots.

where streamflow matters least. As noted in Section 4.4.1, it is difficult to discern any consistent difference between storage changes for VIC and VIC-SIMGM; the range of variability is wider for VIC in sub-basins 7 and 11 and for VIC-SIMGM in sub-basins 15 and 17, but by and large the two are comparable. At Lee's Ferry, both suggest a lower quartile of about -25% and an upper quartile of about +25%.

Results from the κ analysis are shown in Figure 4.12. The top two plots compare skill (here expressed as R^2) for regression forecasts of an April–July target period against κ at lead zero, with each dot representing a different sub-basin, and each color a different simulated storage term or combination of terms. The largest dots in each plot are for Lee's Ferry. For VIC, SWE and soil moisture exhibit comparable estimates of κ , although κ is more variable for SWE across sub-basins, and forecasts based on SWE are noticeably more skillful. As expected, total storage results in values of R^2 and κ that are greater than for either of the individual storage terms. These patterns are largely replicated for VIC-SIMGM, although soil moisture exhibits marginally lower values of κ than for VIC (as discussed in Section 4.4.1), despite seemingly no loss in forecast skill. Aquifer storage has the lowest κ of all, with (somewhat disproportionately) little to no forecast skill. A first order relationship between R^2 and κ is not particularly obvious from these plots, perhaps due to the small sample size.

The middle two plots in Figure 4.12 compare individual simulated storage terms at leads up to six months (i.e., forecasts issued on Oct. 1), with the numbers inside the larger dots denoting the lead time for Lee's Ferry. Similar patterns can again be seen: SWE exhibits the highest skill, VIC-SIMGM soil moisture exhibits lower values of κ than for VIC, and aquifer storage exhibits the lowest values of κ with negligible forecast skill. Though somewhat noisy, a first order relationship between R^2 and κ is here more apparent, with the odd exception of aquifer storage. For the bottom two plots, which show results for total basin storage, this first order relationship is clearly visible, demonstrating that the κ concept can be extended to lead times beyond zero, at least where total basin storage is concerned.

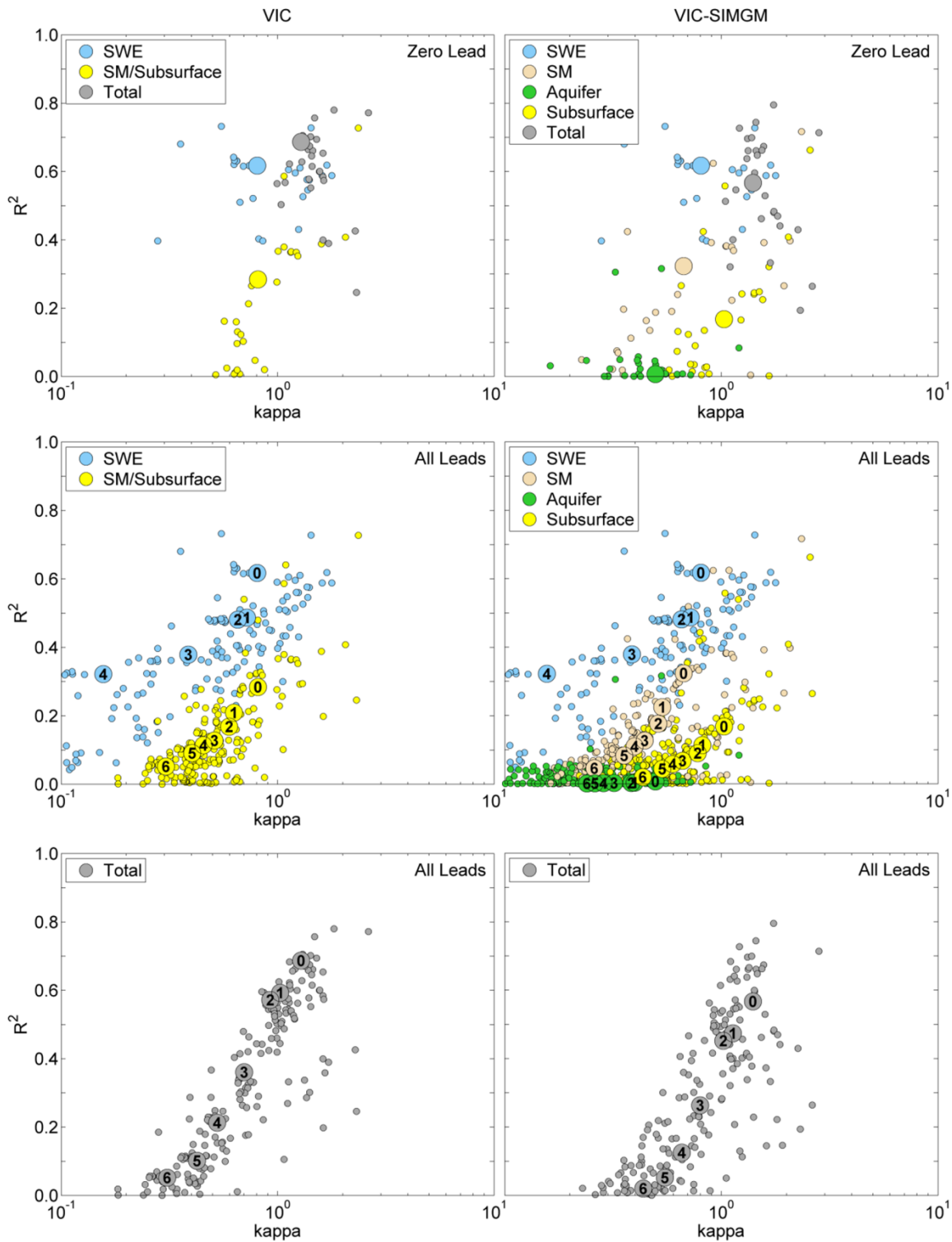


Figure 4.12: Scatter plots of forecast skill v. kappa for various storage term combinations, reflecting an Apr.-Jul. target period. The larger dots in each plot are for Lee's Ferry, with the numbers inside indicating the lead time in months.

The apparently counterintuitive result for aquifer storage merited additional investigation. Correlations between basin-averaged aquifer storage and Apr.-Jul. streamflow were found to be lower than for other storage terms, but the reason was not obvious. Autocorrelation functions were strong and, somewhat logically, higher than for other terms. We tested whether different combinations of grid cells resulted in greater predictive capacity than the collective basin average using a search routine based on principal components regression [Rosenberg *et al.*, 2011]. Skill improved only marginally, however, and negligibly in relation to the improvement found by adopting this approach for other storage terms.

4.5 Summary and conclusions

Our analysis quantitatively assessed the significance of groundwater storage to interannual streamflow anomalies in the Colorado River basin. We compared several estimates of total basin storage change and found that VIC simulations yield similar results regardless of whether a groundwater representation is included. We also found that these estimates compare favorably with those obtained from observation-based basin-scale water balances and GRACE measurements. Further, we found that changes in VIC-SIMGM-simulated groundwater storage were similar to those derived from a baseflow recession analysis.

Assessments of the co-variation between simulated annual storage change and water year streamflow volumes revealed essentially no relationship between these two terms for any of the 29 sub-basins. Similarly, interannual hydrologic storage accounted for only a small percentage of annual streamflow in the headwater sub-basins from which most of the basin's runoff originates. Simulated estimates of groundwater storage exhibited less variability and weaker seasonal streamflow predictive skill than either SWE or soil moisture at every lead time. Thus, we conclude that groundwater storage does not provide a significant contribution to interannual streamflow anomalies in the Colorado River basin.

The implications of these results are noteworthy for both operational and long-term planning purposes. Operationally, they suggest that current statistical and ensemble-based water supply forecasts, neither of which account for groundwater conditions, are likely not detrimentally affected by this omission in the Colorado River basin. With respect to a longer timeframe, the results imply that there is little dependence of one year's discharge on that of the previous year, an issue that has been cited as potentially helping to reconcile modeled projections of climate-change-induced reductions in Colorado River streamflow [Hoerling *et al.*, 2009]. Nonetheless, the contribution of groundwater storage to interannual streamflow anomalies on a global scale is an important and relatively unexplored issue. The methods presented herein can be used to evaluate this issue for other locales.

V. CONCLUSIONS

This dissertation has presented conceptual and operational contributions to the field of water resources engineering. Original concepts exploiting MHMs for operational seasonal streamflow forecasts were developed, demonstrated, and evaluated through applications in the western U.S. Central to this research was the premise that a significant barrier for many prior efforts has been the misalignment between experimental and operational methodologies. Accordingly, the three preceding chapters offered different but complementary approaches to this problem.

Chapter II addressed the first science question, “Can a hybrid framework, which combines model-simulated initial conditions with the regression-based methods used operationally, improve seasonal forecast skill?” The results presented in this chapter suggest that a hybrid framework offers an effective alternative to more conventional forecasting methodologies. Hybrid forecasts based on simulated data acting as surrogates for ground-based observations were found to perform comparably to those based on their observed counterparts. When a larger selection of grid points was considered, hybrid forecasts achieved greater skill, with the greatest benefits in watersheds that are poorly sampled by ground-based observations. For California’s Central Valley drainage, hybrid forecasts demonstrated skill superior to that of DWR’s operational forecasts, with particular improvements for the Cosumnes and Tule.

To the question, “Can such an approach provide the ability to generate late-season forecasts, when snow exists at higher elevations but most observing stations are snow-free?” Chapter II showed that, by providing estimates of SWE along mountain peaks, a hybrid framework can improve forecasts in the latter parts of the ablation season. The implications of this result are perhaps most significant for watersheds with relatively small reservoirs that must balance late-season water supply with flood control considerations, including those in the San Joaquin and Tulare Lake hydrologic regions in the context of the study area.

The second science question was addressed in Chapter III. To the question, “How can macroscale hydrologic models be employed to inform network design for statistical seasonal streamflow forecasts?” this chapter described a skill-oriented approach based on a variation of the hybrid methodology presented in Chapter II. The approach was demonstrated to provide the ability to both rank basins in order of those with the most to gain from a new station, and identify locations within each basin that most appreciably improve forecast skill. Under a forecasting scenario that assumed the currently standard predictors of SWE and water year-to-date precipitation, these locations tended to occur at the margins of the climatological snowpack, providing a second mode of lower-frequency variability that is not captured by predictors with higher climatological averages. As in Chapter II, improvements were most significant for sparsely sampled basins such as the Musselshell and Humboldt.

To the question, “Can soil moisture provide additional predictive skill in a statistical framework beyond that given by conventional predictors?” Chapter III indicated that, for basins with dense existing networks, substantial improvements were only possible when soil moisture was considered, particularly during the accumulation season. Furthermore, locations identified as optimal for soil moisture sensor installation were primarily found in regions of low- to mid-elevation, in contrast to the higher elevations where SNOTEL stations are traditionally situated. These results corroborate prior research while demonstrating that soil moisture data can explicitly improve operational water supply forecasts, and that statistical forecast skill is comparable to ensemble-based skill. In addition, the study demonstrated that simulated hydrologic data can be combined with observations to improve statistical water supply forecasts, a strategy that may prove more effective than new hydrometric installations in the near term.

Chapter IV addressed the final question, “How significant are interseasonal and interannual groundwater anomalies for seasonal streamflow forecasts?” The methods in this chapter form a foundation for the assessment of this problem on a global scale. The corresponding results suggest that groundwater storage bears little relation to interannual

streamflow anomalies, and that groundwater anomalies do not provide a significant contribution to seasonal streamflow predictability, in the Colorado River basin. Operationally, the implication of these results is that current statistical and ensemble-based water supply forecasts, neither of which account for groundwater conditions implicitly or explicitly, are likely not detrimentally affected by this omission in the Colorado River basin. Over a longer time frame, the results imply that there is little dependence of one year's discharge on that of the previous year, an issue that has been cited as potentially helping to reconcile modeled projections of climate-change-induced reductions in Colorado River streamflow.

Finally, to the question, "How does the coupling of a macroscale hydrologic model to an explicit groundwater model affect soil moisture estimates and their predictive capacity?" Chapter IV indicated that a groundwater model substantially increases soil moisture estimates by volumetric water content, particularly during the accumulation season for regions of shallow WTD. Conversely, these same estimates were shown to be smaller than those without a groundwater model in terms of soil moisture volume, an outcome attributed to the propensity of the coupled surface-groundwater model to result in shallower soil layer depths in order to offset the aquifer storage term and conserve mass. Despite these variations, however, statistical analyses revealed essentially no difference in predictive capacity between soil moisture estimated with and without a groundwater representation.

Taken as a whole, the research presented in this dissertation provides an important basis for the adaptation of MHMs to the operational forecasting environment. In the western U.S., where IHCs are so critical to forecast skill, MHMs can help to improve the assessment of catchment storage conditions in an operational context. The three studies in this dissertation explicitly demonstrated this ability for the three main elements of catchment storage: SWE, soil moisture, and groundwater. They also demonstrated how MHMs, which were intended for global and continental applications, can be adapted for analyses at the more focused watershed scales appropriate to operations. The case studies

described provide evidence that MHMs can move beyond the limitations of more conventional approaches and offer insights unavailable by other means.

Future work will seek to transfer the methods developed in this dissertation to other regions. For example, the hybrid methodology of Chapter II may be suitable for environments other than snowmelt-dominant. The network design approach of Chapter III is promising for areas of intensifying water development. The methods of Chapter IV may prove useful in humid settings where the role of groundwater is potentially more significant.

From a conceptual perspective, a noteworthy issue explored in this research is the synergistic skill of data-driven (i.e., statistical) and process-driven (i.e., ensemble-based) prediction models. As computing power and numerical algorithms progress, it will become important to analyze improvements in process-driven models within the context of less intensive prediction frameworks. Furthermore, the implications of climate change for statistical prediction approaches warrant further work. The water resources community is at a crossroads of management paradigms. While continued research on the effects of nonstationarities is essential, there is also a critical need to provide guidance within established models today.

VI. REFERENCES

- Anderson, E. A. (1973), National Weather Service River Forecast System – Snow Accumulation and Ablation Model, *NOAA Tech. Memo*, NWS Hydro-17, 217 pp.
- Andreadis, K. M., P. Storck, and D. P. Lettenmaier (2009), Modeling snow accumulation and ablation processes in forested environments, *Water Resour. Res.*, 45(W05429), 1–13.
- Barnes, B. S. (1939), The structure of discharge-recession curves, *Trans. Am. Geophys. Union*, 20, 721–725.
- Barnett, T. P., J. C. Adam, and D. P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions, *Nat.*, 438, 303–309.
- Barnett, T. P., and D. W. Pierce (2009), Sustainable water deliveries from the Colorado River in a changing climate, *Proc. Natl. Acad. Sci.*, 106(18), 7334–7338.
- Boardman, H. P. (1936), The effect of soil-absorption on snow-survey forecasting of stream-flow, *4th Western Snow Conf.*, Jan 31, Pasadena, CA.
- Brekke, L. D., D. Garen, K. Werner, and D. Laurine (2010), Projecting climate change impacts on seasonal water supply forecasting error, *90th Annual Meeting of the Am. Meteorol. Soc.*, Jan 19, Atlanta, GA.
- Brutsaert, W. (2008), Long-term groundwater storage trends estimated from streamflow records: Climatic perspective, *Water Resour. Res.*, 44, W02409, doi:10.1029/2007WR006518.
- Brutsaert, W., and J. L. Nieber (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, *Water Resour. Res.*, 13(3), 637–643.
- Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), *A generalized streamflow simulation system – Conceptual modeling for digital computers*, Technical Report, Joint Federal and State River Forecast Center, U.S. National Weather Service and California Department of Water Resources, Sacramento, 204 pp.
- CDWR (California Department of Water Resources) (2003), California’s Groundwater: Update 2003, *DWR Bulletin 118*, Available at

<http://www.water.ca.gov/groundwater/bulletin118/bulletin118update2003.cfm>,

Accessed 1 Oct 2009.

CDWR (2009), California Water Plan: Update 2009, *DWR Bulletin 160-09*, Available at

<http://www.waterplan.water.ca.gov/cwpu2009/index.cfm>, Accessed 12 Jan 2010.

CSWRB (California State Water Resources Board) (1951), Water Resources of California, *CSWRB Bulletin 1*.

Carle, D. (2009), *Introduction to Water in California*, California Natural History Guides No. 76, University of California Press, Berkeley, California.

Cherkauer, K. A., and D. P. Lettenmaier (2003), Simulation of spatial variability in snow and frozen soil, *J. Geophys. Res.*, 108, D22, 8858.

Christensen, N., and D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, *Hydrol. Earth Syst. Sci.*, 11, 1417–1434.

Christensen, N. S., A. W. Wood, N. Voisin, D. P. Lettenmaier, and R. N. Palmer (2004), Effects of climate change on the hydrology and water resources of the Colorado River Basin, *Clim. Change*, 62, 337–363.

Church, J. E. (1935), Principles of snow surveying as applied to forecasting stream flow, *J. Agric. Res.*, 51(2), 97–130.

Clyde, G. D. (1940), Soil-moisture studies as an aid in forecasting runoff from snow-cover, *8th Western Snow Conf.*, Apr 19, Portland, OR.

Day, G. N. (1985), Extended streamflow forecasting using NWSRFS, *J. Water. Resour. Plan. Manage.*, 111(2), 157–170.

Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, 45, W09404, 1–22, doi:10.1029/2006WR005855.

Dziegielewski, B., H. P. Garbharran, and J. F. Langowski, Jr. (1993), *Lessons learned from the California drought (1987–1992)*, IWR Rep. 93-NDS-5, Institute for Water Resources, US Army Corps of Engineers, Fort Belvoir, Virginia.

Eng, K., and P. C. D. Milly (2007), Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges, *Water Resour. Res.*, 43, W01201, doi:10.1029/2006WR005293.

- Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador (2010), GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, *Ecology*, 91, 621.
- Fan, Y., G. Miguez-Macho, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, *J. Geophys. Res.*, 112, D10125.
- Ferguson, C. R., J. Sheffield, E. F. Wood, and H. Gao (2010), Quantifying uncertainty in a remote sensing based estimate of evapotranspiration over the continental United States, *Int. J. Remote Sens.*, 31(14), 3821–3865.
- Fiering, M. B. (1965), An optimization scheme for gaging, *Water Resour. Res.*, 1(4), 463–470.
- Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales (2003), Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin, *J. Hydrometeorol.*, 4, 1105–1118.
- Gao, H., Q. Tang, C. R. Ferguson, E. F. Wood, and D. P. Lettenmaier (2010), Estimating the water budget of major US river basins via remote sensing, *Int. J. Remote Sens.*, 31(14), 3955–3978.
- Garen, D. C. (1992), Improved techniques in regression-based streamflow volume forecasting, *J. Water Resour. Plan. Manag.*, 118(6), 654–670.
- Garen, D. C., and T. C. Pagano (2007), Statistical techniques used in the VIPER water supply forecasting software. NRCS-USDA Engineering-Snow Survey and Water Supply Forecasting Technical Note 210-2, 18 pp, Available at http://www.wcc.nrcs.usda.gov/ftpref/downloads/factpub/wsf/technotes/Tech_note_statistical_techniques_in_Viper.pdf, Accessed 12 Jan 2010.
- Gedney, N., and P. M. Cox (2003), The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, *J. Hydrometeorol.*, 4, 1265–1675.
- Georgakakos, K. P., N. E. Graham, T. M. Carpenter, A. P. Georgakakos, and H. Yao (2005), Integrating climate-hydrology forecasts and multi-objective reservoir management for northern California, *Eos Trans. AGU*, 86(12), 122, 127.
- Glantz, M. H. (1982), Consequences and responsibilities in drought forecasting: the case of Yakima, 1977, *Water Resour. Res.*, 18(1), 3–13.

- Grippa, M., L. Kergoat, F. Frappart, Q. Araud, A. Boone, P. de Rosnay, J. M. Lemoine, S. Gascoin, G. Balsamo, C. Ottlé, B. Decharme, S. Saux-Picart, and G. Ramillien (2011), Land water storage variability over West Africa estimated by Gravity Recovery and Climate Experiment (GRACE) and land surface models, *Water Resour. Res.*, 47(W055489), 1–18.
- Gulden, L. E., E. Rosero, Z. L. Yang, M. Rodell, C. S. Jackson, G. Y. Niu, P. J. F. Yeh, and J. Famiglietti (2007), Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil profile?, *Geophys. Res. Lett.*, 34, L09402.
- Hall, D. K., G. A. Riggs, and V. V. Salomonson (2006), *MODIS/Terra Snow Cover 8-Day L3 Global 0.05deg CMG V005*, Oct 2000 to Nov 2010, National Snow and Ice Data Center, Boulder, CO. Digital media, updated daily.
- Hamlet, A. F., and D. P. Lettenmaier (1999), Columbia River streamflow forecasting based on ENSO and PDO signals, *J. Water Resour. Plan. Manag.*, 125(6), 333–341.
- Hart, D., and F. Gehrke (1990), Status of the California Cooperative Snow Surveys Program, *58th Western Snow Conference*, April 17–19, Sacramento, California.
- Hartman, R. K., and A. F. Henkel (1994), Modernization of statistical procedures for water supply forecasting, *62nd Western Snow Conference*, April 18–21, Santa Fe, New Mexico.
- Helms, D., S. E. Phillips, and P. F. Reich (2008), *The history of snow survey and water supply forecasting: Interviews with US Department of Agriculture pioneers*, Natural Resources Conservation Service, US Department of Agriculture, Washington, DC.
- Hoerling, M., D. Lettenmaier, D. Cayan, and B. Udall (2009), Reconciling projections of Colorado River streamflow, *Southwest Hydrol.*, 8, 20–21, 31.
- Kim, Y. O., D. I. Jeong, and H. S. Kim (2001), Improving water supply outlook in Korea with ensemble streamflow prediction, *Water Int.*, 26(4), 563–568.
- Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912.
- Koster, R. D., S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle (2010), Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, *Nat. Geosci.*, 3, 613–616.

- Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma (2009), On the nature of soil moisture in land surface models, *J. Clim.*, 22, 4322–4335.
- Krakauer, N. Y., and M. Temimi (2011), Stream recession curves and storage variability in small watersheds, *Hydrol. Earth Syst. Sci.*, 15, 2377–2389.
- Lakshmi, V., T. Piechota, U. Narayan, and C. Tang (2004), Soil moisture as an indicator of weather extremes, *Geophys. Res. Lett.*, 31, L11401.
- Landres, P., J. Alderson, and D. J. Parsons (2003), The challenge of doing science in wilderness: historical, legal, and policy context, *George Wright Forum*, 20(3), 42–49.
- Langbein, W. B. (1960), Chapter J: Water Budget, in *Comprehensive Survey of Sedimentation in Lake Mead, 1948–49*, U. S. Geol. Survey Professional Paper 295.
- Lea, J., and D. Harms (2011), Developing NRCS SNOTEL and SCAN soil moisture parameters for water supply forecasting applications, *79th Western Snow Conf.*, Apr 18–21, Lake Tahoe/Stateline, NV.
- Li, H., L. Luo, E. F. Wood, and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, 114, D04114.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically-based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, 99, 14415–14428.
- Liang, X., Z. Xie, and M. Huang (2003), A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model, *J. Geophys. Res.*, 108(D16), 8613
- Linsley, R. K., and N. H. Crawford (1960), Computation of a synthetic streamflow record on a digital computer, *Int. Assoc. Sci. Hydrol. Publication*, 51, 526–538.
- Linsley, R. K., M. A. Kohler, and J. L. H. Paulhus (1982), *Hydrology for Engineers, 3rd Ed.*, McGraw-Hill, New York, 508 pp.
- Livneh, B., E. A. Rosenberg, C. Lin, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier (2012), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous U.S.: Update and extensions, *J. Clim.* (submitted).

- Lundquist, J. D., Minder, J. R., Neiman, P. J., and E. Sukovich (2010), Relationships between barrier jet heights, orographic precipitation gradients, and streamflow in the northern Sierra Nevada, *J. Hydrometeorol.*, 11, 1141–1156.
- Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle (2012), Soil moisture, snow, and seasonal streamflow forecasts in the United States, *J. Hydrometeorol.*, 13, 189–203.
- Maillet, E. (1905), Essai d'hydraulique souterraine et fluviale, *Librairie Sci.*, A. Herman, Paris.
- Manabe, S. (1969), Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Weather Rev.*, 97, 739-774.
- Martin, G. (2006), San Joaquin levees: Sodden, leaking and very frail, *San Franc. Chron.*, 14 Apr 2006.
- Maurer, E. P., D. P. Lettenmaier, and N. J. Mantua (2004), Variability and potential sources of predictability of North American runoff, *Water Resour. Res.*, 40, W09306, 1–13.
- Maurer, E. P., A. W. Wood, J. C. Adam, and D. P. Lettenmaier (2002), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *J. Clim.*, 15, 3237–3251.
- Maxwell, R. M., and N. L. Miller (2005), Development of a coupled land surface and groundwater model, *J. Hydrometeorol.*, 6, 233–247.
- McCuen, R. H. (1985), *Statistical Methods for Engineers*, Prentice-Hall, Englewood Cliffs, NJ.
- McEnery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson (2005), NOAA's Advanced Hydrologic Prediction Service: Building pathways for better science in water forecasting, *Bull. Am. Meteorol. Soc.*, 86(3), 375–385.
- McGuire, M., A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2006), Use of satellite data for streamflow and reservoir forecasts in the Snake River basin, *J. Water Resour. Plan. Manag.*, 132(2), 97–110.
- Meyboom, P. (1961), Estimating ground-water recharge from stream hydrographs, *J. Geophys. Res.*, 66, 1203–1214.

- Miguez-Macho, G., H. Li, and Y. Fan (2008), Simulated water table and soil moisture climatology over North America, *Bull. Am. Meteorol. Soc.*, 89(5), 663–672.
- Miller, J. A. (1999), *Ground Water Atlas of the United States: Introduction and National Summary (HA 730-A)*, U. S. Geol. Survey, Reston, Virginia.
- Miller, D. A., and R. A. White (1998), A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, *Earth Interact.*, 2.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity is dead: Whither water management? *Sci.*, 319, 573–574.
- Mishra, A. K., and P. Coulibaly (2009), Developments in hydrometric network design: A review, *Rev. Geophys.*, 47, RG2001.
- Mishra, V., K. A. Cherkauer, and S. Shukla (2010), Assessment of drought due to historic climate variability and projected future climate change in the midwestern United States, *J. Hydrometeorol.*, 11, 46–68.
- Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, 109, D07S90.
- Molotch, N. P., and R. C. Bales (2005), Scaling snow observations from the point to the grid element: Implications for observation network design, *Water Resour. Res.*, 41, W11421, doi:10.1029/2005WR004229.
- Morin, G., J. P. Fortin, W. Sochanska, J. P. Lardeau, and R. Charbonneau (1979), Use of principal component analysis to identify homogenous precipitation stations for optimal interpolation, *Water Resour. Res.*, 15(6), 1841–1850.
- Mote, P. W., A. F. Hamlet, M. Clark, and D. P. Lettenmaier (2005), Declining mountain snowpack in western North America, *Bull. Am. Meteorol. Soc.*, 86(1), 39–49.
- Murdock, J. N., and L. Calder (1969), Bank storage – Lake Powell, Bureau of Reclamation Region 4, Salt Lake City, Utah.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models, Part I: A discussion of principles, *J. Hydrol.*, 10, 282–290.

- NRC (National Research Council) (2007), *Colorado River Basin Water Management: Evaluating and Adjusting to Hydroclimatic Variability*, National Academies Press, Washington, DC.
- NRC (National Research Council) (2008), *Review of CCSP Draft Synthesis and Assessment Product 5.3: Decision-Support Experiments and Evaluations Using Seasonal to Interannual Forecasts and Observational Data*, National Academies Press, Washington, DC, 56 pp.
- Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow simulation for continental-scale river basins, *Water Resour. Res.*, 33, 711–724.
- Niu, G. Y., Z. L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, 112, D07103.
- Niu, G. Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local scale measurements, *J. Geophys. Res.*, 116, D12109, doi:10.1029/2010JD015139.
- Oleson, K. W., et al. (2008), Improvements to the Community Land Model and their impact on the hydrological cycle, *J. Geophys. Res.*, 113, G01021.
- Oleson, K. W., et al. (2010), Technical description of version 4.0 of the Community Land Model (CLM), *NCAR Tech Note NCAR/TN-478 +STR*, Natl. Cent. for Atmos. Res., Boulder, Colorado.
- Pagano, T. C., D. C. Garen, T. R. Perkins, and P. A. Pasteris (2009), Daily updating of operational statistical seasonal water supply forecasts for the western US, *J. Am. Water Resour. Assoc.*, 45(3), 767–778.
- Pagano, T. C., H. C. Hartmann, S. Sorooshian (2001), Using climate forecasts for water management: Arizona and the 1997–1998 El Niño, *J. Am. Water Resour. Assoc.*, 37(5), 1139–1153.
- Pardo-Igúzquiza, E. (1998), Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, *J. Hydrol.*, 210, 206–220.

- Peck, E. L., and J. C. Schaake (1990), Network design for water supply forecasting in the West, *Water Resour. Bull.*, 26(1), 87–99.
- Perkins, T. R., J. K. Marron, and A. G. Goodbody (2010), ArcGIS technique to evaluate the SNOTEL data network, *2nd Joint Federal Interagency Conf.*, Jun 27–Jul 1, Las Vegas, NV.
- Quinlan, P. (2010), Lake Mead’s water level plunges as 11-year drought lingers, *The New York Times*, 13 Aug. 2010.
- Robson, S. G., and E. R. Banta (1995), *Ground Water Atlas of the United States: Arizona, Colorado, New Mexico, Utah (HA 730-C)*, U. S. Geol. Survey, Reston, Virginia.
- Rodríguez-Iturbe, I., and J. M. Mejía (1974), The design of rainfall networks in time and space, *Water Resour. Res.*, 10(4), 713–728.
- Rosenberg, E. A., A. W. Wood, and A. C. Steinemann (2011), Statistical applications of physically based hydrologic models to seasonal streamflow forecasts, *Water Resour. Res.*, 47, W00H14, doi:10.1029/2010WR010101.
- Rosenberg, E. A., A. W. Wood, and A. C. Steinemann (2012a), Informing hydrometric network design for statistical seasonal streamflow forecasts, *J. Hydrometeorol.* (submitted).
- Rosenberg, E. A., E. A. Clark, A. C. Steinemann, and D. P. Lettenmaier (2012b), On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin, *Hydrol. Earth Syst. Sci.* (submitted).
- Rutledge, A. T. (1998), Computer programs for describing the recession of ground-water discharge and for estimating mean ground-water recharge and discharge from streamflow data – Update, *Water-Resources Investigations Report 98-4148*, U. S. Geol. Survey, Reston, Virginia.
- Schaefer, G. L., M. H. Cosh, and T. J. Jackson (2007), The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), *J. Atmos. Ocean Technol.*, 24, 2073–2077.
- Shamir, E., and K. P. Georgakakos (2007), Estimating snow depletion curves for American River basins using distributed snow modeling, *J. Hydrol.*, 334, 162–173.

- Sharma, A., and S. Chowdhury (2011), Coping with model structural uncertainty in medium-term hydro-climatic forecasting, *Hydrol. Res.*, 42(2–3), 113–127, doi: 10.2166/nh.2011.104.
- Sheffield, J., K. M. Andreadis, E. F. Wood, and D. P. Lettenmaier (2009), Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events, *J. Clim.*, 22, 1962–1981.
- Shukla, S., and D. P. Lettenmaier (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, 3529–3538.
- Speers, D. D., D. M. Rockwood, and G. D. Ashton (1996), Chapter 7: Snow and snowmelt, *Hydrology Handbook*, ASCE Manuals and Reports on Engineering Practice No. 28, 2nd ed., Am. Soc. of Civil Eng., New York, 437–476.
- Stern, P. C., and W. E. Easterling (eds.) (1999), *Making Climate Forecasts Matter*, National Research Council Report, National Academies Press, Washington, DC.
- Strassberg, G. B., R. Scanlon, and D. Chambers (2009), Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States, *Water Resour. Res.*, 45, W05410, doi:10.1029/2008WR006892.
- Swenson, S. C., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, 33, L08402, doi:10.1029/2005GL025285.
- Syed, T. H., J. S. Famiglietti, M. Rodell, J. Chen, and C. R. Wilson (2008), Analysis of terrestrial water storage changes from GRACE and GLDAS, *Water Resour. Res.*, 44, W02433, doi:10.1029/2006WR005779.
- Tang, Q., and D. P. Lettenmaier (2010), Use of satellite snow-cover data for streamflow prediction in the Feather River basin, California, *Int. J. Remote Sens.*, 31(14), 3745–3762.
- Tang, Q., S. Peterson, R. H. Cuenca, Y. Hagimoto, and D. P. Lettenmaier (2009), Satellite-based near-real-time estimation of irrigated crop water consumption, *J. Geophys. Res.*, 114, D05114, doi: 10.1029/2008JD010854.
- Troccoli, A., M. Harrison, D. L. T. Anderson, and S. J. Mason (eds.) (2008), *Seasonal Climate: Forecasting and Managing Risk*, NATO Science Series IV: Earth and Environmental Sciences, Springer Academic Publishers, 499 pp.

- Tsintikidis, D., K. P. Georgakakos, J. A. Sperflage, D. E. Smith, and T. M. Carpenter (2002), Precipitation uncertainty and raingauge network design within Folsom Lake watershed, *J. Hydrol. Eng.*, 7(2), 175–184.
- Twedt, T. M., J. C. Schaake, Jr., and E. L. Peck (1977), National Weather Service extended streamflow prediction, *45th Western Snow Conference*, April 18–21, Albuquerque, New Mexico.
- Twedt, T. M., R. J. C. Burnash, and R. L. Ferral (1978), Extended streamflow prediction during the California drought, *46th Western Snow Conference*, April 18–20, Otter Crest, Oregon.
- USACE (US Army Corps of Engineers) (2001), *HEC-HMS for the Sacramento and San Joaquin River basins comprehensive study*, PR-46, Institute for Water Resources, US Army Corps of Engineers.
- USDOI (U. S. Department of Interior) (2000), Colorado River interim surplus criteria, *Final Environmental Impact Statement, Volume I*, Bureau of Reclamation.
- VanRheenen, N. T., A. W. Wood, R. N. Palmer, and D. P. Lettenmaier (2004), Potential implications of PCM climate change scenarios for Sacramento-San Joaquin River basin hydrology and water resources, *Clim. Change*, 62, 257–281.
- Vogel, R. M., and C. N. Kroll (1992), Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, 28, 2451–2458, doi:10.1029/92WR01007.
- Volkman, T. H. M., S. W. Lyon, H. V. Gupta, and P. A. Troch (2010), Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain, *Water Resour. Res.*, 46, W11554.
- Wallace, J. M., and D. S. Gutzler (1981), Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Mon. Wea. Rev.*, 109, 784–812.
- Whitehead, R. L. (1996), *Ground Water Atlas of the United States: Montana, North Dakota, South Dakota, Wyoming (HA 730-I)*, U. S. Geol. Survey, Reston, Virginia.
- Wilks, D. (1992), Adapting stochastic weather generation algorithms for climate change studies, *Clim. Change*, 22, 67–84.

- Wood, A. W. (2007), The effects of climate change on water supply forecasting in the Feather River basin, *4th Annual California Climate Change Conf.*, Sep 10–13, Sacramento, CA.
- Wood, A. W., and D. P. Lettenmaier (2006), A test bed for new seasonal hydrologic forecasting approaches in the western United States, *Bull. Am. Meteorol. Soc.*, 87(12), 1699–1712.
- Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, 35, L14401.
- Yeh, P. J. F., and E. A. B. Eltahir (2005), Representation of water table dynamics in a land surface scheme, Part I: Model development, *J. Clim.*, 18, 1861–1880.
- Zaitchik, B. F., M. Rodell, and F. Olivera (2010), Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme, *Water Resour. Res.*, 46, W06507.
- Zampieri, M., E. Serpetzoglou, E. N. Anagnostou, E. I. Nikolopoulos, and A. Papadopoulos (2012), Improving the representation of river–groundwater interactions in land surface modeling at the regional scale: Observational evidence and parameterization applied in the Community Land Model, *J. Hydrol.*, 420–421, 72–86.
- Zhang, K., J. S. Kimball, R. R. Nemani, and S. W. Running (2010), A continuous satellite derived global record of land surface evapotranspiration from 1983–2006, *Water Resour. Res.*, 46, W09522, doi:10.1029/2009WR008800.

CURRICULUM VITAE

EDUCATION

PhD	University of Washington, Civil and Environmental Engineering, 2012
MS	Columbia University, Civil Engineering and Engineering Mechanics, 2004
BS	Northwestern University, Civil Engineering, 1998

LICENSURES

Registered Professional Engineer, State of New York (License No. 081107)

EMPLOYMENT

Lecturer, University of Washington, Seattle, WA, 2011

- *Engineering Hydraulics and Hydrology (CEE 345)* – Principal instructor for junior-level class of 60 students, offered winter and spring quarters, 2010–2011 academic year. Teaching evaluations ranked among the top 10% of all courses in the department.

Graduate Research Assistant, University of Washington, Seattle, WA, 2006–2010

- *Decision Support through Earth Science Results, Sponsor: NASA* – Designed systems to improve water resources management in the western US through the use of hydrologic modeling, remote sensing data assimilation, and seasonal climate forecasts. Interfaced closely with Natural Resources Conservation Service and California Department of Water Resources.
- *Washington State Climate Change Impacts Assessment, Sponsor: Washington State Congress* – Evaluated potential impacts of climate change on state’s urban stormwater infrastructure. Analyzed historical precipitation records and simulations of future rainfall to assess changes in precipitation extremes. Coordinated efforts among research team members.
- *West-Wide Drought Forecasting System – A Scientific Foundation for NIDIS, Sponsor: USGS* – Developed methods to incorporate USGS streamflow and well data into a drought forecast and nowcast model for the National Integrated Drought Information System. Assessed performance of a coupled surface-groundwater model and quantified contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin.

Senior Principal Engineer, Hazen and Sawyer PC, New York, NY, 2000–2006

- *South Richmond Drainage Project, Staten Island, NY* – Designed 30 stormwater “best management practices” (BMPs) ranging from meandering streams to

- constructed wetlands. Generated contract drawings, prepared erosion and sediment control plans, provided design services during construction, and assisted in project management.
- *Inner Harbor Combined Sewer Overflow Abatement, New York, NY* – Managed project to replace deteriorated sluice gates in 70 regulators located throughout Brooklyn and Manhattan. Developed contract drawings and specifications, prepared cost estimates, assisted in bidding process, and reviewed shop drawings.
 - *Catskill/Delaware Ultraviolet Disinfection Facility, Mount Pleasant, NY* – Performed hydrologic study, prepared conceptual design, and generated preliminary plans of 10-acre on-site wetlands mitigation for 2 BGD facility.

Design Engineer, KPFF Consulting Engineers, Seattle, WA, 1998–2000

- *Port of Longview Berth 8 Wharf, Longview, WA* – Developed designs of grading, dredging, drainage, and utility systems. Provided design services during construction and generated cost estimate for this \$8 million project.
- *Pier T Container Yard, Long Beach, CA* – Conducted hydrologic analysis and prepared storm drainage designs for 200-acre shipping facility. Directed profiling of storm mains and assisted in drafting of contract drawings.
- *Seattle Iron & Metals Relocation, Seattle, WA* – Prepared preliminary and final designs of grading, drainage, and utility systems for scrap metal facility. Coordinated and designed all improvements within public right-of-way.

Construction Inspector, NYSDOT, Binghamton, NY, Summer 1996

- *State Route 7 Bridge Replacement, Harpursville, NY* – Conducted topographical surveys, slump tests, and soil moisture studies for the replacement of a two-lane, 250-foot-long bridge. Calculated quantity estimates for earthwork, paving, and structural concrete. Monitored and evaluated safety and performance of contractor.

REFEREED PUBLICATIONS

Rosenberg EA, Wood AW, Steinemann AC (2012) Informing hydrometric network design for statistical seasonal streamflow forecasts. *Journal of Hydrometeorology* (submitted)

Rosenberg EA, Clark EA, Steinemann AC, Lettenmaier DP (2012) On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin. *Hydrology and Earth System Sciences* (submitted)

Livneh B, Rosenberg EA, Lin C, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP (2012) A long-term hydrologically based dataset of land surface fluxes and states for the conterminous US: Update and extensions. *Journal of Climate* (submitted)

Rosenberg EA, Wood AW, Steinemann AC (2011) Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. *Water Resources Research*, 47(W00H14):1–19, doi:10.1029/2010WR010101

Rosenberg EA, Keys PW, Booth DB, Hartley D, Burkey J, Steinemann AC, Lettenmaier DP (2010) Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State. *Climatic Change*, 102(1–2):319–349, doi:10.1007/s10584-010-9847-0

Whitely Binder LC, Barcelos JK, Booth DB, Darzen M, Elsner MM, Fenske R, Graham TF, Hamlet AF, Hodges-Howell J, Jackson JE, Karr C, Keys PW, Littell JS, Mantua N, Marlow J, McKenzie D, Robinson-Dorn M, Rosenberg EA, Stöckle CO, Vano JA (2010) Preparing for climate change in Washington State. *Climatic Change*, 102(1–2):351–376, doi:10.1007/s10584-010-9850-5

Tang Q, Rosenberg EA, Lettenmaier DP (2009) Use of satellite data to assess the impacts of irrigation withdrawals on Upper Klamath Lake, Oregon. *Hydrology and Earth System Sciences*, 13:617–627.

ADDITIONAL PUBLICATIONS

Westerling AL, Barnett T, Gershunov A, Hamlet AF, Lettenmaier DP, Lu N, Rosenberg E, Steinemann AC (2008) Climate forecasts for improving management of energy and hydropower resources in the western US. California Energy Commission, PIER Energy-Related Environmental Research Program.

Vokral J, Gumb D, Cavallaro AD, Overall N, McCulloch FA, Rosenberg E (2003) Restoring Staten Island's riparian ecology. *Stormwater*, 4(7):12–24.

Vokral J, Gumb D, Cavallaro AD, Mehrotra S, Rosenberg E (2003) Wetlands at work. *Civil Engineering*, 73(2):56–63.

SELECTED CONFERENCE PRESENTATIONS

Rosenberg EA, Clark EA, Steinemann AC, Lettenmaier DP (2011) The utility of soil moisture estimates from a land surface model in seasonal streamflow forecasts. *91st Annual Meeting of the American Meteorological Society*, Jan 23–27, Seattle, WA.

Rosenberg E (2010) Statistical forecasts with simulated predictors: A case study of the Sierra Nevada. *56th Annual Meeting of the California Cooperative Snow Surveys*, Oct 13–15, Bishop, CA.

Rosenberg E, Wood AW, Steinemann AC, Lettenmaier DP (2010) Development of a hybrid dynamical-statistical model for seasonal streamflow forecasts. *Hydrology in the 21st Century: Links to the past, and a vision for the future*, Steve Burges Retirement Symposium, Mar 24–26, 2010, Seattle, WA.

Ho SL, Lo M, Famiglietti J, Swenson S, Anderson K, Syed TH, Rosenberg E, Lettenmaier D (2009) Total water storage change over the San Joaquin and Sacramento River basins: Comparing GRACE and observational data. *2009 AGU Fall Meeting*, Dec 14–18, San Francisco, CA.

Rosenberg E (2009) Historical and future trends in precipitation extremes and their impacts on the stormwater infrastructure of Washington State. *The Water Center's 19th Annual Review of Research*, Feb 18, Seattle, WA.

Rosenberg E, Tang Q, Wood AW, Steinemann AC, Lettenmaier DP (2008) Statistical applications of physical hydrologic models and satellite snow cover observations to seasonal water supply forecasts. *2008 AGU Fall Meeting*, Dec 15–19, San Francisco, CA.

Keys PW, Rosenberg EA, Booth DB, Steinemann AC, Lettenmaier DP (2008) Precipitation extremes and the impacts of climate change on stormwater in Washington State. *2008 AGU Fall Meeting*, Dec 15–19, San Francisco, CA.

Lundquist JD, Lott F, Minder J, Rosenberg E (2008) Storm tracks and spatial precipitation patterns in the Sierra Nevada, California: Implications for hydrologic modeling and water resource planning. *2008 AGU Fall Meeting*, Dec 15–19, San Francisco, CA.

Rosenberg E, Steinemann AC (2008) Utility of seasonal climate outlooks for water supply forecasts in northern California. *The Water Center's 18th Annual Review of Research*, Feb 14, Seattle, WA.

Rosenberg E, Wood AW, Tang Q, Steinemann AC, Imam B, Sorooshian S, Lettenmaier DP (2007) Improving water resources management in the western United States through use of remote sensing data and seasonal climate forecasts. *5th Annual Climate Prediction Applications Science Workshop*, Mar 20–23, Seattle, WA.

Deb-Moorjani D, Vokral J, Gumb D, Cavallaro AD, Mehrotra S, Rosenberg E (2005) The Staten Island Bluebelt: A case study in urban watershed management. *ASCE Watershed Management Conference 2005*, Jul 19–22, Williamsburg, VA.

Gumb D, Brauman R, Mehrotra S, Rosenberg E, Deb D (2004) A study of water quality benefits provided by BMPs in the Staten Island Bluebelt. *6th Annual Conference on Watershed Protection, 2nd Annual New York City Watershed Science and Technical Conference*, Sep 21–22, Fishkill, NY.

REVIEW COMMITTEES

Hossain F (2013) *Climate Vulnerability: Understanding and Addressing Threats to Essential Resources – Water Volume*. Elsevier (in preparation)

Snover AK, Whitely Binder L, Lopez J, Willmott E, Kay J, Howell D, Simmonds J (2007) *Preparing for Climate Change: A Guidebook for Local, Regional, and State Governments*. In association with and published by ICLEI – Local Governments for Sustainability, Oakland, CA.

Journals: *Water Resources Research*

TEACHING

Engineering Hydraulics and Hydrology (CEE 345, University of Washington)

OTHER PROFESSIONAL PRACTICE

World Resources Institute, Washington, DC – Developed gridded, regional scale runoff simulations for Aqueduct, an online water risk assessment tool for globally significant river basins (2012)

HONORS AND AWARDS

UW Civil and Environmental Engineering Luther E. Gregory Scholarship (2011)
AWWA Archie Rice Scholarship (Highest award for Pacific Northwest Section) (2009)
AWRA Rod Sakrison Fellowship (Highest award for Washington State Chapter) (2007)
UW Water Center Andy Studebaker Fellowship (2006)
ASCE Outstanding Student Award for Illinois (1998)
Alpha Lambda Delta Freshman Honor Society (1995)

PROFESSIONAL AFFILIATIONS

Member, American Society of Civil Engineers
Member, American Water Resources Association
Member, American Geophysical Union
Student Member, American Meteorological Society
Student Member, American Water Works Association