ABSTRACT

MAZROOEI, AMIRHOSSEIN. Quantifying and Reducing Uncertainty from Multiple Sources in Forecasting Monthly to Seasonal Land-Surface Attributes. (Under the direction of Dr. Sankar Arumugam.)

Water resources management at monthly to seasonal scales typically require medium-long range (1-month to 3-month ahead) hydrologic forecasts such as streamflow and soil moisture attributes. Over last decades, considerable progress have been made in understanding the low-frequency climate variability and its impact on hydrologic forecasts, particularly for short-range lead times (e.g. daily to weekly). Yet, the quality of forecasts has not been improved substantially for practical applications, hence the necessity of understanding potential improvements and advancements from multiple error sources in the streamflow forecasting over multiple time scales is still being felt. Mainly, the streamflow dynamics in a catchment depend on two major influential components: 1) atmospheric conditions over the watershed and 2) land surface state (e.g. soil moisture, snow cover, etc.) which is also known as Initial Hydrologic Conditions (IHCs). Realistic estimates of initial hydrologic conditions provide critical information of a catchment especially for long-term (e.g. seasonal) hydrologic predictions, although they are hard to be acquired or measured directly. Furthermore, other sources of uncertainty involved in hydrologic predictions such as deficiency of hydrologic models and inaccurate estimates of model parameters also intensify the modeling issues. Inherent chaos of the climate and atmospheric systems is also a barrier in obtaining accurate climate forecasts. Significant advances in weather forecasting were made during last decades after launching weather satellites along with achieving on-going improvements in atmospheric models. However current precipitation and temperature forecasts are still not reliable beyond 10 days
which is a critical time scale for management purposes. The predictability of climate systems especially for seasonal lead times relies on boundary conditions, such as sea surface temperature. Thus a better understanding of climatic variability can lead to more accurate seasonal climate forecasts.

Several strides in hydrology and meteorology have been made in order to improve the estimation of initial hydrologic conditions of a land surface. Unfortunately, due to inaccurate description of the land surface state, the exact IHCs can not be quantified or easily measured from the field studies. This implies that there is an unavoidable uncertainty related to IHCs that will be propagated to hydrologic predictions. However, the potential improvement in state variable estimations can be made through Data Assimilation (DA) techniques which uses the model estimates and its error covariance along with independent observations to update the state variables. Most of the studies on hydrologic data assimilation use remotely sensed soil moisture or snow data, while limited efforts have been made to utilize the observed streamflow records from gauging stations.

In this dissertation, different methods of developing land-surface attributes (i.e. streamflow and soil moisture) for medium-range lead times (e.g. monthly to seasonal) are employed and tested along with the quantification of the uncertainty in the forecasted products through various statistical analyses and verification metrics. and then try to quantify the uncertainty in probabilistic forecasts of streamflow and deterministic forecasts of soil moisture. Furthermore, with the intention of reducing error in land-surface state, data assimilation techniques are applied into different types of hydrologic models and the enhancements in the model products are assessed and presented.
Quantifying and Reducing Uncertainty from Multiple Sources in Forecasting Monthly to Seasonal Land-Surface Attributes

by
Amirhossein Mazrooei

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Civil Engineering

Raleigh, North Carolina
2017

APPROVED BY:

Dr. Kumar Mahinthakumar
Dr. Sujay Kumar

Dr. Erin Hestir
Dr. Tushar Sinha

Dr. Sankar Arumugam
Chair of Advisory Committee
DEDICATION

To my lovely parents, Shahnaz and Habib, and to my inspiring sister and brother in law Parisa and Shervin, due to their endless love and support. I certainly would not be where I am today without your nurture, guidance, and care. I love you.
Amirhossein Mazrooei was born in Esfahan, Iran on August 1989. He received his high school diploma in Mathematics and Physics from Ejei high school, Iranian National Organization for Developing Exceptional Talents (NODET) in 2006. Amirhossein earned his B.S. degree in Civil Engineering at Sharif University of Technology (Tehran, Iran) in 2012. Then, he decided to join North Carolina State University to pursue his Masters of Science and Doctor of Philosophy in Water Resources Engineering under the direction of Prof. Sankar Arumugam. Developing and improving streamflow and soil moisture forecasts was the main focus of his research.
ACKNOWLEDGEMENTS

First, I would like to express my sincerest gratitude to my advisor, Dr. Sankar Arumugam for his continuous support and guidance and invaluable kindness during my graduate study at NC State. It was a privilege to have learned from you. I would also like to appreciate my advisory committee members, Drs. Kumar Mahinthakumar, Erin Hestir, Sujay Kumar, Tushar Sinha, Raju Vatsavai and Andy Wood for their scientific and technical assists and inspirations throughout my study.

Further, I would like to thank my friends and colleagues at the Climate, Hydrology and Water Resources (CHWR) group Dr. Jason Patskoski, Dr. Seung Beom Seo, Dr. Rajarshi Bhowmik, Dr. Dominic Libera, Dr. Sudarshana Mukhopadhyay, Bandar Almutari, and Harminder Singh whose contribution helped me a lot to complete my research. Also, I really feel grateful to work with state climate office of North Carolina Prof. Ryan Boyles and Rebecca Ward.

And lastly, I would like to extend my special thanks to my family and friends Dr. Amir Rezvani, Dr. Ali Kasiri, Shahla Rezvani, Zahra Farid, Maryam Mohaghegh, Naser Soltani, Rahil Fazel, Morteza Jafarabadi, Mehran Jafari, Alireza Afiyat, Payam Tabrizian, Nasim Naseri, Negin Naseri, Tahmineh Dehghan, Monica Zeinalizadeh, Samaneh Pourmojib, Pejman Dehghani, Aliasghar Marjani, Hamid Kazem, Nooshafarin Mohammadzadeh, Mahtab Fakhimi, Arya Farhood, Ladan Gahramani, Nahid Mehraban, Habib Rahimi, Milad Hallaji, Amirhossein Norouzi, Behrooz Keshavarzi, Amir Koolivand, Daniel Davis, Alireza Abbasion, Shahin Safavizadeh, Arash Bozorgi, Alireza Mashadi, Sahand Saberi, Amir Ghanbari, Mehrdad Nasiri, Kourosh Mohammadi, and many others who helped me throughout all the years at NC State to create memorable experiences. And to whom I’ve forgotten, I am truly sorry. Thank you too!
# TABLE OF CONTENTS

LIST OF TABLES ................................................................. vii

LIST OF FIGURES ............................................................... ix

INTRODUCTION ........................................................................ 1

Chapter 1 Probabilistic Downscaling ........................................ 4
  1.1 Introduction .................................................................. 5
  1.2 data ........................................................................... 9
    1.2.1 Study Area ......................................................... 9
    1.2.2 Streamflow Data ................................................. 10
    1.2.3 Climate Forecasts ............................................... 11
  1.3 Methodology .................................................................. 12
    1.3.1 Candidate Models ............................................... 12
    1.3.2 Model Calibration and Validation ......................... 15
    1.3.3 Forecasting Skill Metrics ..................................... 16
  1.4 Results ......................................................................... 18
  1.5 Discussion ..................................................................... 30

Chapter 2 SM Forecasting ......................................................... 34
  2.1 Introduction .................................................................. 35
  2.2 data and methodology ................................................ 37
    2.2.1 SMAP .................................................................. 37
    2.2.2 ECHAM4.5 Precipitation Forecasts ......................... 38
    2.2.3 NOAH3.2 LSM ................................................... 38
    2.2.4 Meteorological Forcings of LSM ......................... 39
    2.2.5 Experimental Setup ............................................. 39
  2.3 Results ......................................................................... 41
  2.4 Discussion ..................................................................... 45

Chapter 3 EnKF Data Assimilation ............................................. 47
  3.1 Introduction .................................................................. 48
  3.2 data ........................................................................... 52
    3.2.1 Streamflow Data ............................................... 53
    3.2.2 Precipitation and Temperature Data .................... 53
    3.2.3 Potential Evapotranspiration Data ....................... 54
    3.2.4 Precipitation Forecasts ...................................... 54
  3.3 Methodology .................................................................. 55
    3.3.1 "abcd" Model ....................................................... 55
3.3.2 Ensemble Kalman Filter (EnKF) ........................................... 61

3.4 Results .................................................................................. 65
  3.4.1 Assessment of ”abcd” Model Performance without EnKF ........... 65
  3.4.2 Role of EnKF in Improving ‘abcd’ Model ............................... 66
  3.4.3 Scenarios of State Variables in EnKF .................................. 68
  3.4.4 EnKF Usefulness in Operational Streamflow Forecasting ........... 71

3.5 Discussion .............................................................................. 72

Chapter 4 Variational DA in VIC .................................................. 75
  4.1 Introduction ........................................................................... 76
    4.1.1 Previous Work ............................................................... 78
  4.2 Data ....................................................................................... 81
    4.2.1 Tar River Basin ............................................................. 81
    4.2.2 Streamflow Observations .............................................. 81
    4.2.3 Observed Meteorological Forcings ................................. 82
    4.2.4 ECHAM4.5 Precipitation Forecasts ............................... 83
  4.3 Methodology .......................................................................... 84
    4.3.1 Hydrologic Model .......................................................... 84
    4.3.2 Variational Data Assimilation ....................................... 85
    4.3.3 Implementation of VAR in VIC model ......................... 88
    4.3.4 Streamflow Forecasting Approach ............................... 89
    4.3.5 Evaluation Metrics .......................................................... 91
  4.4 Results ................................................................................... 93
    4.4.1 VIC Model Performance .................................................. 93
    4.4.2 Role of VAR-DA in Streamflow Simulation .................. 94
    4.4.3 Role of VAR-DA in Streamflow Forecasting ................ 97
    4.4.4 Role of VAR-DA Over Different Time Scales ............... 100
  4.5 Discussion ............................................................................ 102

CONCLUSION AND FUTURE WORK ........................................... 107

BIBLIOGRAPHY ........................................................................... 110

APPENDICES .............................................................................. 139
  Appendix A  Common Data Assimilation Algorithms .................... 140
  Appendix B  Supplemental Results ............................................ 143
LIST OF TABLES

Table 1.1  USGS gauge sites characteristics .......................................................... 10
Table 1.2  Improved forecast attributes (positive/bold values) in terms of $\Delta BS = BS^{PCR} - BS^{MLR}$, $\Delta REL = REL^{PCR} - REL^{MLR}$, and $\Delta RES = RES^{MLR} - RES^{PCR}$, by utilizing MLR forecasts over PCR forecasts for BN and AN months ........................................ 22
Table 3.1  Description of the parameters of the 'abcd' model ............................. 56
Table 4.1  VIC model performance summery over the validation period ........ 94
Table B.1  Performance summary of monthly deterministic streamflow forecasting using ECHAM4.5 precipitation forecasts during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths $AW$, "m" denotes month(s) and "d" stands for day(s). "+" sign indicates the improved verification metric and "−" sign denotes degradation in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest. 150
Table B.2  Performance summary of monthly deterministic streamflow forecasting using ensemble mean of 42 ensemble members generated under the ESP scheme during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths $AW$, "m" denotes month(s) and "d" stands for day(s). "+" sign indicates the improved verification metric and "−" sign denotes degradation in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest. 151
Table B.3  Performance summary of monthly probabilistic streamflow forecasting under the ESP scheme for Below Normal (BN) months and Above Normal (AN) months during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths $AW$, "m" denotes month(s) and "d" stands for day(s). Upward arrow indicates an improvement in the score and downward arrow shows the degradation of the metric in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Map of the US with the Sunbelt specified as the red shaded area as well as the location of the selected river basins and their considered USGS gauging stations.</td>
</tr>
<tr>
<td>1.2</td>
<td>Comparison of MLR and PCR models performances on cross-validated forecasts in a seasonal time scale.</td>
</tr>
<tr>
<td>1.3</td>
<td>Comparison of MLR and PCR models performances on split sample forecasts in a seasonal time scale.</td>
</tr>
<tr>
<td>1.4</td>
<td>Comparison of skill in individual streamflow forecasts between the models MLR1 and MLR2.</td>
</tr>
<tr>
<td>1.5</td>
<td>Cloud plot comparison of skill in streamflow forecasting for Below Normal (BN), Normal (N), and Above Normal (AN) months between the MLR model and PCR model under both cross-validation and split sample validation. The median lines of the cloud points are projected on each axes for each category.</td>
</tr>
<tr>
<td>1.6</td>
<td>Difference between MLR and PCR performances ($\Delta RPSS = RPSS_{MLR} - RPSS_{PCR}$) with respect to the skewness of monthly streamflows for all the selected basins during Winter season. Positive $\Delta RPSS$ means MLR performs better than PCR. Significant positive relationship were found after fitting linear regression.</td>
</tr>
<tr>
<td>1.7</td>
<td>Decomposed contribution of probabilistic information from Precipitation forecasts (P) and antecedent streamflow observations (Q) as input variables in determining the skill of MLR streamflow forecasts under cross validation approach.</td>
</tr>
<tr>
<td>1.8</td>
<td>The performance range of MLR model in categorical streamflow forecasting as a function of ensemble size of precipitation forecasts during four seasons. The solid line, the darker area, and the lighter area in each color represents the median, IQR, and 90% confidence interval of the RPSSs respectively derived from 50 iterations at each ensemble size.</td>
</tr>
<tr>
<td>2.1</td>
<td>RRMSE of the bias corrected 1-3 months ahead soil moisture forecasts based on the SMAP soil moisture observations.</td>
</tr>
<tr>
<td>2.2</td>
<td>Correlation coefficient between 1-3 months ahead soil moisture forecasts with the SMAP soil moisture observations. Grid cells with insignificant correlations (based on 18 months of data) are grayed out.</td>
</tr>
<tr>
<td>2.3</td>
<td>Scatter plot of soil moisture residuals for two sample grid cells with good and bad forecasting skills. The residuals are centered around the mean of SMAP observations.</td>
</tr>
</tbody>
</table>
Figure 4.1 Location of the Tar River basin and USGS gauging station #02089500 located at Tarboro North Carolina

Figure 4.2 Seasonality of Tar River Basin

Figure 4.3 Schematic of the variational data assimilation approach. Black squares represent the model states (not in the streamflow dimension), triangles are the streamflow observations, $UF$ and $AW$ denote DA Update Frequency and Assimilation Window parameters respectively. The hydrologic model is then initialized using the state conditions achieved either from Open Loop (OL) simulation or Data Assimilated (DA) simulations.

Figure 4.4 Timeseries of monthly streamflow observations along with streamflow simulations from Open Loop (OL) scheme and Data Assimilation (DA) experiment using $UF = 15\text{days}$ and $AW = 10\text{days}$.

Figure 4.5 $\Delta KGE (KGE_{DA} - KGE_{OL})$ Improvement in Streamflow Simulations after applying VAR data assimilation for different $AW$ and $UF$ lengths using all flows (gray-scaled plots), only low flow analysis (i.e. observed streamflow less than $10^{th}$ percentile $Q < Q_{p=0.1}$) (orange plots) and only high flow analysis (i.e. observed streamflow more than $90^{th}$ percentile $Q > Q_{p=0.9}$) (blue plots) during the period 1991-2010. The first row shows the statistics quantified using daily flows and the second row plots are generated using average monthly flows. The skill of open-loop simulations for each of the analyses are shown in the plot titles in terms of KGE.

Figure 4.6 Timeseries of monthly forecasted streamflow acquired from both deterministic and probabilistic forecasting approaches. Prior to monthly forecasting VIC model is initialized either based on Open Loop (OL) states (first row) or through Data Assimilated (DA) initial conditions (second row) using $AW = 15\text{days}$.

Figure 4.7 The magnitude of the model skill improvements due to data assimilation over daily time steps of a monthly prediction. Each color represents the DA application with a unique assimilation window length $AW$.

Figure B.1 Evaluation of Spearman’s rank correlation between downscaled monthly precipitation forecasts and observed precipitation at $1/8^\circ$ grid cells over Tar River basin, NC.

Figure B.2 Timeseries of optimal scale factor $K$ for various DA experiments with different $UF$ and $AW$ parameters. Y axis of the plots are limited between 0 and 2, with the middle thick sign showing the value 1 for the $K$ multiplier ($K = 1$ denotes no change in the background model state).
Figure B.3 Mean and Variance of $K$ multiplier computed during 20 years of simulations for various DA experiments with different $UF$ and $AW$ parameters.
INTRODUCTION

Water resources management at monthly-to-seasonal time scales typically requires 1-month to 3-month ahead hydrologic forecasts such as streamflow and soil moisture attributes. Over last decades, considerable progress have been made in understanding the low-frequency climate variability and its impact on developing seasonal hydrologic forecasts. Yet, the quality of forecasts has not been improved substantially for practical applications, hence the necessity of understanding potential strategies for improving and advancing uncertainty reduction methods in the streamflow forecasting from multiple error sources [Welles et al., 2007]. Mainly, the streamflow dynamics in a catchment depend on two major influential components: 1) atmospheric conditions over the watershed and 2) land surface states (e.g. soil moisture, snow cover, etc.) which is also known as Initial Hydrologic Conditions (IHCs). Seasonal climate forecasts are routinely issued by climate prediction centers. Given the atmospheric forcings, realistic estimates of initial hydrologic conditions is critically required especially for seasonal hydrologic predictions, although they are hard to be acquired or measured directly. Furthermore, other sources of uncertainty involved in hydrologic predictions such as model and parameter uncertainties also degrade the skill in streamflow predictions. In this regard, substantial effort have been made in order to reduce the prediction errors due to parameter uncertainty as well as hydrologic model uncertainty [Beven and Binley, 1992; Vrugt et al., 2003; Wilby, 2005; Li et al., 2015].

Inherent chaos of the climate and atmospheric systems is also a barrier in obtaining accurate climate forecasts. Significant advances in weather forecasting were made during last decades after launching weather satellites along with achieving on-going improvements in atmospheric models (e.g. GCMs) [Anderson and Veltishchev, 1973; Bauer et al., 2015].
However, current precipitation and temperature forecasts are still not reliable at seasonal time scales, which is a critical time scale for management purposes. The predictability of climate systems especially for seasonal lead times relies on boundary conditions, particularly sea surface temperature (SST) [Lorenz, 1993]. Thus a better understanding of climatic variability (e.g. El Nino-Southern Oscillation-ENSO) can lead to more accurate seasonal climate forecasts. Reducing uncertainty in climate models and forcings through multimodel combination of climate forecasts will also reduce uncertainty in streamflow forecasts [Devineni et al., 2008; Devineni and Sankarasubramanian, 2010b; Li et al., 2016]. In addition, better estimation of soil moisture state enhance the prediction skill in numerical climate models [Fennessy and Shukla, 1999].

Several strides in hydrology and meteorology have been made in order to improve the estimation of initial hydrologic conditions of a land surface [Pagano et al., 2004; Liu et al., 2012]. Unfortunately, due to inadequate observations of the land surface state, the exact IHCs can not be quantified [Li et al., 2009] or easily measured from the field studies. This implies that there is an unavoidable uncertainty related to IHCs that will be propagated in hydrologic predictions. However, the potential improvement in state variable estimations can be made through data assimilation techniques which uses the model estimates and its error covariance along with independent observations to update the state variables. Most of the studies on hydrologic data assimilation use remotely sensed soil moisture or snow data, while limited efforts have been made to utilize the observed streamflow records from gauging stations.
Outline of the Dissertation

In this dissertation, different methods of developing land-surface attributes (i.e. streamflow and soil moisture) for medium-range lead times (e.g. monthly to seasonal) are employed and tested along with the quantification of the uncertainty in the forecasted products through various statistical analyses and verification metrics. and then try to quantify the uncertainty in probabilistic forecasts of streamflow and deterministic forecasts of soil moisture. Furthermore, with the intention of reducing error in land-surface state, data assimilation techniques are applied into different types of hydrologic models and the enhancements in the model products are assessed and presented.

The dissertation is organized as follows: we first explore a new approach in utilizing the climate forecasts and its application for categorical streamflow forecasting. Next, seasonal soil moisture forecasts are developed and their skill is assessed by comparing with the remotely sensed observations from the Soil Moisture Active Passive (SMAP) satellite. Next we focus on the application of ensemble data assimilation techniques in a lumped watershed model in order to update model state variables based on streamflow observations and improvements introduced to the monthly streamflow forecasts are identified. And in the last chapter, the application of variational data assimilation method into a land surface model is tested for operational forecasting purposes. This research tries to include both stochastic and physical modeling approaches in hydrologic forecasting. Towards this, we also focus over the contiguous US, with the purpose of better understanding the models’ behaviors across various hydroclimatic regimes.
Utilizing Probabilistic Downscaling Methods to Develop Streamflow Forecasts from Climate Forecasts

Abstract

Statistical information from ensemble of climate forecasts can be utilized in improving the streamflow predictions by using different downscaling methods. This study investigates the use of Multinomial Logistic Regression (MLR) in downscaling large-scale ensemble climate forecasts into basin-scale probabilistic streamflow forecasts of categorical events over major river basins across the US Sunbelt. The performance of MLR is then compared with the categorical forecasts estimated from the traditional approach Principal Component Regression (PCR). Results from both cross-validation and split-sampling reveal that in general the probabilistic categorical forecasts from MLR model have more accuracy and
exhibit higher Rank Probability Skill Score (RPSS) compared to the PCR probabilistic forecasts. MLR forecasts are also more skillful than PCR forecasts during the winter season and for basins that exhibit high interannual variability in streamflows. The role of ensemble size of precipitation forecasts in developing MLR-based streamflow forecasts was also investigated. Given the simplicity involved in MLR, it offers an alternate reliable approach in developing categorical streamflow forecasts.

1.1 Introduction

Reliable monthly to seasonal streamflow forecasts can significantly improve the management of water resources systems and its subsequent plans [Hamlet et al., 2002]. Towards this, hydrologists provide deterministic-style forecasts that estimate the volume of streamflow for a month or a season ahead. Alternatively, water managers are interested in categorical and probabilistic streamflow forecasts which represent the probability of occurrence of predefined events such as below-normal or above-normal streamflows [Krzysztofowicz, 2001; Ahmadisharaf et al., 2016a]. Streamflow forecasts are typically developed using either dynamical modeling or statistical modeling. In the dynamical approach, downscaled climate forecasts from General Circulation Models (GCM) such as precipitation and temperature forecasts are forced into physical hydrologic models (e.g. Lumped or distributed rainfall-runoff models) in order to develop streamflow predictions and other terrestrial hydrologic fluxes [Vrugt et al., 2006; Lohmann et al., 2004; Sinha and Sankarasubramanian, 2012]. For the purpose of developing ensemble streamflow forecasts, the hydrologic models can be fed with ensemble of forcings and/or ensemble of perturbed initial hydrologic conditions (ICs). Another approach is to pool forecasts from multiple
models to develop multimodel ensemble which reduces the model uncertainty \cite{Ajami et al., 2006; Li and Sankarasubramanian, 2012; Devineni et al., 2008}.

Statistical model forecasting is based on developing a statistical relationship between relevant climatic predictors and/or some available observations such as initial soil moisture and streamflow conditions prior to the forecasting. Commonly, statistical models such as Principal Component Regression (PCR) assumes normality of predictands and linearity between predictors and predictands \cite{Hsu et al., 1995; Garen, 1992; Sankarasubramanian et al., 2008}. However it is well-known that nonlinear relationship exists between runoff and precipitation \cite{Jakeman et al., 1993; Sankarasubramanian and Vogel, 2003}. For a long time, considerable progress has been made on incorporating the influence of Sea Surface Temperature (SST) anomalies such as El Nino Southern Oscillation (ENSO), which partially holds information in tropics and subtropics towards seasonal precipitation and streamflow forecasting \cite{Ropelewski and Halpert, 1986; Tottle et al., 2005; Hamlet and Lettenmaier, 1999}. In addition to SST anomalies, statistical models can benefit from the natural persistence of streamflow in order to improve their forecasting skill through using information of past streamflow conditions. Previous studies have employed various types of statistical modeling techniques including parametric, semi-parametric and re-sampling methods (e.g. K-Nearest Neighbors (K-NN), Model Output Statistics (MOS), etc.) in order to develop probabilistic streamflow forecasts over a specific watershed or across a region \cite{Grantz et al., 2005; Clark and Hay, 2004}.

In general, statistical modeling approach is simpler and computationally less costly compared to the dynamical modeling method. In dynamical modeling, hydrologic models (e.g. Land Surface Models) require climatic forcings with finer spatio-temporal resolution.
than that of climate forecasts issued from GCMs. Addressing this resolution mismatch is a big challenge in itself and requires the application of spatial downscaling and/or temporal disaggregation approaches on GCM outputs before forcing them into hydrologic models [Wood et al., 2002; Wood and Lettenmaier, 2006; Yuan et al., 2011]. In addition, both downscaling and disaggregation procedures introduce errors into the climatic forcings and subsequently hydrologic products; thus might significantly influence the reliability of streamflow forecasts, mostly depending on their location and forecasting season. [Mazrooei et al., 2015; Sinha et al., 2014; Seo et al., 2016].

Climate forecasts, that are available in large spatial scales, are typically issued in the form of ensembles, which quantify the uncertainty due to the initial conditions [Goddard et al., 2003; Doblas-Reyes et al., 2005]. Various studies have focused on reducing uncertainty in climate forecasts by combining multiple models [Barnston et al., 2003; Weigel et al., 2008] and developing different strategies for providing ensembles that quantify uncertainty in atmospheric conditions [Kumar et al., 2001; Li et al., 2008] and hydrologic states [Shukla and Lettenmaier, 2011]. However, for statistical downscaling, most studies used only the ensemble mean of climate forecasts for developing streamflow forecasts which ignores the probabilistic information in the data [Sinha and Sankarasubramanian, 2013]. Few studies that pointed out the significance of these probabilistic information use ensemble spread or a subset of skillful ensemble members in their modeling process in order to gain more accuracy in the forecasted products [Wilks and Hamill, 2007; Regonda et al., 2006].

The main intent of this study is to evaluate the value of ensemble climate forecasts in developing categorical streamflow forecasts based on statistical downscaling. Here, we propose the use of multinomial logistic regression (MLR) for downscaling probabilistic
information inside the large scale ensemble climate forecasts into basin-scale and utilize it for probabilistic streamflow forecasting. Unlike binary logistic regression that deals with only two categories of events, MLR is able to develop categorical probabilities for multiple outcomes and predefined events. The other advantage of MLR is that the model input can be both probabilistic data such as probabilistic information derived from climate ensembles as well as deterministic data such as initial land surface conditions of a catchment. In this study, we have employed precipitation forecasts from ECHAM4.5 GCM along with past streamflow observations to build MLR model for developing 1-month ahead categorical streamflow forecasts over six river basins that span across various hydroclimatic regimes located in the US Sunbelt. The performance of MLR model was then compared with the commonly used PCR model in order to identify the added value of using probabilistic information in the ensemble climate forecasts for improving streamflow forecasts. Results from our experiments address the following science questions associated with categorical streamflow forecasting:

1. How is the MLR model performance compared to PCR model in developing probabilistic categorical streamflow forecasts during different seasons?

2. What is the information added to the categorical streamflow forecasts developed using the MLR model?

3. Is there any relationship between the skill of MLR model and river basins’ characteristics and their regimes?

4. What is the role of climate forecast ensemble size in determining the skill of probabilistic streamflow forecasts?
1.2 DATA

CHAPTER 1. PROBABILISTIC DOWNSCALING

1.2 Study Area and Data

1.2.1 Study Area

In this study, we consider six river basins that fall under arid, semi-arid, or humid hydroclimatic regimes across the US Sunbelt (Fig. 1.1). US Sunbelt is defined as the region south of the 37°N latitude that has short and mild winters and extended summers.
1.2. DATA

Table 1.1. USGS gauge sites characteristics

<table>
<thead>
<tr>
<th>Basin Name</th>
<th>State</th>
<th>USGS Identification</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Drainage Area [km(^2)]</th>
<th>Mean Streamflow [cfs]</th>
<th>Aridity Index</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep</td>
<td>NC</td>
<td>#02102000 35°37'37&quot;</td>
<td>79°06'58&quot;</td>
<td>3,714 (1,434)</td>
<td>1,466</td>
<td>0.89 Humid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACF</td>
<td>FL,GA,AL</td>
<td>#02535800 30°42'03&quot;</td>
<td>84°51'33&quot;</td>
<td>44,548 (17,200)</td>
<td>22,287</td>
<td>0.89 Humid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guadalupe</td>
<td>TX</td>
<td>#08167500 29°51'37&quot;</td>
<td>98°23'00&quot;</td>
<td>3,405 (1,315)</td>
<td>461</td>
<td>0.50 Dry sub-humid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rio Grande</td>
<td>NM</td>
<td>#08276500 36°19'12&quot;</td>
<td>105°45'16&quot;</td>
<td>25,200 (9,730)</td>
<td>737</td>
<td>0.41 Semi-arid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Pedro</td>
<td>AZ</td>
<td>#09471000 31°37'33&quot;</td>
<td>110°10'26&quot;</td>
<td>3,195 (1,234)</td>
<td>41</td>
<td>0.26 Arid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verde</td>
<td>AZ</td>
<td>#09508500 34°04'23&quot;</td>
<td>111°42'56&quot;</td>
<td>15,170 (5,858)</td>
<td>613</td>
<td>0.32 Semi-arid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1 presents detailed information about the selected river basins and their streamflow gauge stations.

1.2.2 Streamflow Data

Streamflow data is obtained from the U.S. Geological Survey’s (USGS) Hydro-climatic Data Network (HCDN) [Slack et al., 1993] which consists of streamflow gauging stations with minimal anthropogenic impacts from upstream reservoir operations, land use changes, and groundwater pumping. This dataset encompasses daily, monthly, and annual mean discharge values for 1,659 sites across the United States. In this study, we used monthly streamflow records of the selected river basins during the period of 46 years from 1957 to 2002 (HCDN gauge numbers are shown in Fig. 1.1).
1.2.3 Climate Forecasts

Precipitation forecasts from ECHAM4.5 GCM model was obtained from International Research Institute of Climate and Society (IRI) data library [Li and Goddard, 2005]. ECHAM model is developed by Max Planck Institute and is currently used in real time climate forecasting [Roeckner et al., 1992]. We considered ECHAM4.5 since it has a long period of retrospective monthly precipitation forecasts and studies have shown that ECHAM4.5 precipitation forecasts can provide a reliable skill in streamflow forecasting [Mazrooei et al., 2015; Sinha et al., 2014]. In this study we used both ECHAM4.5 monthly simulations and monthly forecasts both available at $2.8\times2.8$ resolution. In the simulation scheme, the observed Sea Surface Temperatures (SST) were used to feed the ECHAM4.5 GCM and the simulated ensemble of monthly precipitation contains 85 members. On the other hand, in the forecast scheme, the climate model is forced with the updated SST forecasts developed using the constructed analogue SST (CA-SST) [Van den Dool, 1994] to develop the climate forecasts which has 24 ensemble members and available from 1957 till present. We decided to include the ECHAM4.5 simulations because it has sufficient number of ensemble members to evaluate if increased ensemble size provides additional information for developing probabilistic categorical streamflow forecasts. Here, we obtained 1-month ahead precipitation data from both ECHAM4.5 simulations and forecasts for the period of 46 years (1957-2002).
1.3 Methodology

For each river basin, the grid point of ECHAM4.5 precipitation forecasts having the highest correlation with the observed streamflow was selected among all the grid points that overlay or neighbor the basin boundary. Then the precipitation forecasts from the selected grid point and the observed monthly streamflow prior to the forecasting time step were considered as the two predictors of our modeling framework, along with the observed streamflow in that forecasting month considered as the predictand variable. Two models, Multinomial Logistic Regression (MLR) and traditional Principal Component Regression (PCR) were considered to develop streamflow forecasts. The models were evaluated based on leave-5-out cross validation and split sample validation to develop probabilistic streamflow tercile forecasts. This section provides more details about both models and validation techniques.

1.3.1 Candidate Models

Principal Components Regression (PCR)

Climate Predictability Tool (CPT) from IRI [Mason and Tippett, 2016] was employed in order to develop the PCR model. For a given month, the predictor variables for the PCR model include the ensemble mean of the precipitation forecasts ($\mu_{prec}$) and the previous month observed streamflow ($Q_{t-1}$) and PCR model estimates the conditional mean and variance of the streamflow during the forecasting time step. Using the above two statistics and by assuming the predictand follows a normal distribution, we quantify the probabilistic forecasts of below normal, normal, and above normal streamflow events.
1.3. METHODOLOGY

(\(P(Q_{BN}), P(Q_N), P(Q_{AN})\)) based on the climatological thresholds. PCR is one of the widely used Model Statistic Output (MOS) approaches in both hydrological research fields and operational forecasting frameworks [Antolik, 2000; Garen, 1992; Hamill et al., 2006; Pagano, 2008; Li et al., 2014; Arumugam et al., 2015]. However, the disadvantage of this method is that it only considers the ensemble mean and ignores the probabilistic information in the ensemble of climate forecasts.

Multinomial Logistic Regression (MLR)

Multinomial logistic regression is an extension of binary logistic regression, which enables the model to develop probabilistic predictions for multiple categories and outcomes. MLR is also capable of accepting inputs from mixed data types such as probabilities and deterministic variables. In order to feed MLR with the probabilistic information from the ensemble of climate forecasts, we quantified the probability of below normal \(P(Q_{BN})\), normal \(P(Q_N)\), and above normal \(P(Q_{AN})\) precipitation forecasts. For this purpose, we first pooled together all the forecasted precipitation ensembles over the study period (1957-2002) for a given month and then the 33rd and 67th model climatological percentiles were computed. Based on these climatological statistics of the ECHAM4.5 GCM, we estimated the probabilities of categorical precipitation forecasts in two different ways:

(i) Bin Counting: For each time step, the number of ensemble members that fall between the GCM model climatological 33rd and 67th percentiles were counted and then divided by the total number of ensembles (either 24 or 85 depending on the source of precipitation forecasts) in order to develop precipitation forecast tercile probabilities.

We assigned the term **MLR1** to the MLR model that uses this type of precipitation
1.3. METHODOLOGY

(ii) Distribution Fitting: For each time step, a log-normal distribution was fit to the precipitation ensemble and the CDF of the model climatological 33rd and 67th quantiles were computed. MLR model fed with this type of predictor is called MLR2 in the following.

The predictor matrix $X$ (Eqn. 1.1) for MLR model consists of probabilities of categorical precipitation events as well as observed streamflow prior to the forecasting time step. We didn’t include $P(P_{AN})$ in the predictors since it can be derived from the other two probabilities thus it doesn’t add any additional information into the modeling framework. MLR model requires categorized outcomes as the predictand, thus for a given month we converted the streamflow observations to nominal outcomes based on the 33rd and 67th quantiles of the historical records. The MLR model then estimates the coefficient matrix $B$ (Eqn. 1.2) for the regression. In MLR model, number of rows in the coefficient matrix should be equal to number of predictors plus one and number of columns is equal to number of outcome categories minus one so in our study the matrix $B$ would have 4 rows and 2 columns. Finally based on the equations 1.3 and 1.4 and the rule that probabilities sum up to 1, the probabilities of the BN, N, and AN categories of streamflow forecasts were computed.

$$X_t = \begin{bmatrix} P(P_{BN})_t & P(P_{N})_t & Q_{t-1} \end{bmatrix}$$  \hspace{1cm} (1.1)
1.3. METHODOLOGY

CHAPTER 1. PROBABILISTIC DOWNSCALING

\[
B = \begin{bmatrix}
B_{01} & B_{02} \\
B_{11} & B_{12} \\
B_{21} & B_{22} \\
B_{31} & B_{32}
\end{bmatrix}_{4 \times 2} \tag{1.2}
\]

\[
\ln \left( \frac{\mathbb{P}(Q_{BN})}{\mathbb{P}(Q_{AN})} \right) = B_{01} + B_{11} \mathbb{P}(P_{BN}) + B_{21} \mathbb{P}(P_{N}) + B_{31} Q_{t-1} \tag{1.3}
\]

\[
\ln \left( \frac{\mathbb{P}(Q_{N})}{\mathbb{P}(Q_{AN})} \right) = B_{02} + B_{12} \mathbb{P}(P_{BN}) + B_{22} \mathbb{P}(P_{N}) + B_{32} Q_{t-1} \tag{1.4}
\]

1.3.2 Model Calibration and Validation

MLR and PCR models were conducted and evaluated based on two different validation techniques which are explained here.

1.3.2.1 Cross Validation

Cross validation technique is a procedure that assesses the performance of a model by calibrating the model with a subset of the data and validating it for the left out data [Craven and Wahba, 1978]. In this study, for a given month, leave-5-out cross validation is performed by removing a 5-year window of data centered at the forecasting time step, calibrating the model based on the remaining 41 years of data, and then evaluating the model performance on the forecasting year.
1.3. METHODOLOGY

1.3.2.2 Split Sample Technique

In the split sample technique, unlike the cross-validation only a single subset of data is used to train the models. Here, The first 26 years of data (from 1957 to 1982) is used as the calibration period and the remaining 20 years (from 1983 to 2002) as the validation period.

Each of the explained validation techniques has its own advantages and drawbacks. Cross validation can stabilize the structure of the model since it uses different subsets of training data. On the other hand, in the split sample test the model has lesser years for the calibration and it is shown that it has lesser skill too, comparing to cross-validated forecasts [Goutte, 1997; Moradkhani et al., 2004]. Split sample test is commonly used and beneficial when there is a short set of dataset so it is not rational to leave out the available data. In addition, the split sample validation technique provides a more rigorous way to evaluate the skill of the model towards operational forecasting as it only uses past data to develop the forecasts [Klemes, 1986].

1.3.3 Forecasting Skill Metrics

Candidate models’ performance is evaluated over their validation period using Rank Probability Skill Score (RPSS) which is a widely used statistic in probabilistic forecasting verification [Epstein, 1969; Wilks, 2006]. RPSS indicates the improvement of the model skill relative to the climatology in forecasting the category in which the streamflow observation lies. First, Rank Probability Score (RPS, Eqn. 1.5) is computed which expresses the closeness of the forecasted CDF of the categorical probabilities to the observed CDF for
1.3. METHODOLOGY

CHAPTER 1. PROBABILISTIC DOWNSCALING

each time step and then RPSS is computed by comparing the RPS of forecast with the RPS of climatology (Eqn. 1.7),

$$RPS_t = \frac{1}{K} \sum_{k=1}^{K} [P_t^k - O_t^k]^2$$  \hspace{0.5cm} (1.5)

$$RPS_{t \text{climatology}} = \frac{1}{K} \sum_{k=1}^{K} [\bar{O}^k - O_t^k]^2$$  \hspace{0.5cm} (1.6)

$$RPSS_t = 1 - \frac{RPS_t}{RPS_{t \text{climatology}}}$$  \hspace{0.5cm} (1.7)

where in the above equations, $P_t^k$, $O_t^k$, and $\bar{O}^k_t$ are the forecasted, observed, and climatological cumulative probabilities respectively at the time step $t$ for the $k^{th}$ category.

In addition to RPSS, in order to diagnose different attributes of the forecasting skill for a specific category (e.g. BN or AN), we employed commonly used Brier Score (BS) [Brier, 1950; Weigel et al., 2007], and its empirical decomposition into reliability (REL), resolution (RES), and uncertainty (UNC) were quantified for model intercomparison purposes based on the following equations:

$$BS = \frac{1}{N} \sum_{t=1}^{N} (p_t - y_t)^2$$  \hspace{0.5cm} (1.8)

$$BS = REL - RES + UNC$$  \hspace{0.5cm} (1.9)

$$REL = \sum_{d=1}^{D} \frac{n_d}{N} \left( \frac{o_d}{n_d} - d \right)^2$$  \hspace{0.5cm} (1.10)

$$RES = \sum_{d=1}^{D} \frac{n_d}{N} \left( \frac{o_d}{n_d} - \bar{d} \right)^2$$  \hspace{0.5cm} (1.11)
1.4. RESULTS

\[ UNC = \bar{o}(1 - \bar{o}) \]  

where \( N \) is total number of forecast probabilities, \( y_t \) is binary observation (\( y_t = 1 \) if event happened and \( y_t = 0 \) otherwise), and the issued forecasts \( p_t \) were assumed to have \( D \) distinct values (i.e. \( p_t \in \{ P_1, ..., P_D \} \)) for all timesteps \( t \). \( n_d \) denotes the number of times when the \( d^{th} \) forecast was issued, and \( o_d \) refers to the total number of events that have been observed when the \( d^{th} \) forecast was issued. \( \bar{o} \) is the climatological event frequency which is equal to 0.33 (i.e. BN or AN months) in this case.

1.4 Results

Based on the two MLR approaches that are described in section 1.3, we selected the individual model with higher skill for discussing the results. Fig. 1.2 compares the cross-validated RPSS between MLR and PCR for all six basins during four seasons. RPSS metric ranges from negative infinity to 1 where for better visualization purposes (i.e. better visualize the variability in the plots) we limited the RPSS axes from -1 to 1 which illustrates almost 95% of the data points on average. All the positive RPSS medians in the box plots in Fig. 1.2 suggest that the MLR and PCR forecasts perform better than the climatology (i.e. RPSS median greater than zero). In addition, the MLR model is more successful than the PCR model in forecasting streamflows in almost all the cases as it indicates higher median RPSS and interquartile range (IQR) (Fig. 1.2). According to the box plots, Rio Grande river basin exhibits the best forecasting performance since more than 75% of RPSS values from both models are above zero for all seasons. In general, the MLR model delivers a wider distribution of RPSSs compared to the PCR model.
indicating a higher uncertainty in streamflow forecasts. MLR performs better than PCR in the Guadalupe river basin during the winter season with the first quartile of MLR being approximately above the third quartile of PCR highlighting that in 75% of occasions MLR would be more skillful.

Fig. 1.3 shows the same analysis of RPSS values developed under the split-sample validation technique. We again see that both MLR and PCR forecasts are more skillful than the climatology. Comparing the RPSS median between the two forecasting models, MLR model still performs better than the PCR model in all basins over all seasons with just one exception during spring season in the Deep river basin. From Fig. 1.2 and Fig. 1.3 we notice that the skill of cross-validated forecasts vary less in comparison to the split-sample validation. The reason is that the cross-validation approach considers more data for model training resulting in less variability in model skill. We also notice under both validation techniques that MLR model’s distributions of RPSSs are more skewed towards higher values indicating that MLR is able to issue better streamflow predictions over the considered period suggesting higher reliability of MLR.

Furthermore, another analysis was conducted on the cross-validated forecasts during BN and AN months using BS and its decomposed components. Table 1.2 shows the difference in forecast verification metrics comparing MLR and PCR forecasts over each river basin. Positive (Bold) numbers indicate the improvements achieved by utilizing MLR forecasts over PCR forecasts. REL and RES values were computed using 20 distinct forecast probabilities (i.e. $D = 20$) during 46 years of monthly forecasts (i.e. $N = 552$). Note that UNC component does not differ between MLR and PCR forecasts thus it is not included in this table. This assessment reveals that MLR forecasts consistently
Figure 1.2. Comparison of MLR and PCR models performances on cross-validated forecasts in a seasonal time scale.
Figure 1.3. Comparison of MLR and PCR models performances on split sample forecasts in a seasonal time scale.
1.4. RESULTS

Table 1.2. Improved forecast attributes (positive/bold values) in terms of $\Delta BS = BS^{PCR} - BS^{MLR}$, $\Delta REL = REL^{PCR} - REL^{MLR}$, and $\Delta RES = RES^{MLR} - RES^{PCR}$, by utilizing MLR forecasts over PCR forecasts for BN and AN months

<table>
<thead>
<tr>
<th>Basin</th>
<th>BN $\Delta BS$ [$\times 10^{-2}$]</th>
<th>BN $\Delta REL$ [$\times 10^{-2}$]</th>
<th>BN $\Delta RES$ [$\times 10^{-2}$]</th>
<th>AN $\Delta BS$ [$\times 10^{-2}$]</th>
<th>AN $\Delta REL$ [$\times 10^{-2}$]</th>
<th>AN $\Delta RES$ [$\times 10^{-2}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep</td>
<td>2.35</td>
<td>-0.43</td>
<td>2.86</td>
<td>0.22</td>
<td>-0.52</td>
<td>0.84</td>
</tr>
<tr>
<td>ACF</td>
<td>1.82</td>
<td>1.40</td>
<td>0.64</td>
<td>-0.26</td>
<td>-0.35</td>
<td>0.26</td>
</tr>
<tr>
<td>Guadalupe</td>
<td>6.36</td>
<td>2.13</td>
<td>4.43</td>
<td>3.51</td>
<td>0.89</td>
<td>2.71</td>
</tr>
<tr>
<td>Rio Grande</td>
<td>2.97</td>
<td>1.13</td>
<td>1.87</td>
<td>0.37</td>
<td>0.03</td>
<td>0.48</td>
</tr>
<tr>
<td>San Pedro</td>
<td>2.72</td>
<td>-0.05</td>
<td>2.86</td>
<td>2.26</td>
<td>0.73</td>
<td>1.92</td>
</tr>
<tr>
<td>Verde</td>
<td>3.13</td>
<td>-0.36</td>
<td>3.55</td>
<td>2.23</td>
<td>1.20</td>
<td>1.60</td>
</tr>
</tbody>
</table>

have lower brier score for BN months and most likely for AN months over all the basins. This is particularly because of the greater resolution in MLR forecasts, thus containing more information compared to PCR forecasts (an ‘uninformed’ forecaster has $RES = 0$). However, $\Delta REL$ values imply that sometimes MLR has lower reliability (a reliable forecaster has $REL = 0$) compared to PCR, but this degradation is dominated by the improved resolution and eventually results in better (lower) brier score.

We also compared the performance of two MLR models that were fed with two types of climatic probabilistic information. One uses bin counting (MLR1) to quantify the precipitation probabilities and the other one uses fitting a log-normal distribution (MLR2) in order to obtain the probabilistic information from the precipitation ensembles (details in section 1.3). Fig. 1.4 shows the individual RPSSs for all the monthly time steps under two validation techniques. Also the months are categorized based on below normal (BN), normal (N), and above normal (AN) months which are obtained by comparing observed streamflow with the climatological terciles. Based on Fig. 1.4 we see that the
points are scattered almost equally on the two sides of the diagonal line indicating no significant difference in forecasting skill whether using bin counting or fitting a log-normal distribution to the precipitation ensemble.

Fig. 1.5 demonstrates the comparison between the skill of MLR and PCR models in categorical streamflow forecasting for each individual monthly time step during the validation period. In order to simplify the comparison across the models, the medians of the RPSS values for each category were shown on each axes as colorized lines. In this figure, since the majority of points and consequently the intersection of median lines are located below the diagonal line, it indicates that the RPSS of MLR is higher than PCR. By looking at the point clouds in Fig. 1.5a we see that the skill of PCR model varies less (less deviation in RPSS values) in terms of forecasting normal (N) flows compared to other two categories. MLR model performs more precisely in arid regions like Guadalupe and Rio Grand river basins, since most of the points are scattered on the right side of the boxes. Based on the medians of the RPSSs in Fig. 1.5a, both MLR and PCR models have a better skill in forecasting AN and BN months, compared to N months. It is harder to conclude this by looking at Fig. 1.5b since there are fewer data points in the split sample validation as we have limited sample size under the different flow categories. In addition, the better performance of MLR is distinct in the arid basins.

Further, the seasonal relationship between the mean monthly $\Delta$RPSS ($\Delta$RPSS = $RPSS_{MLR} - RPSS_{PCR}$) and the skewness of monthly flows, $\gamma$ based on Eqn. 1.13:

$$\gamma_m = \frac{E(Q_m - \bar{Q}_m)^3}{\sigma_{Q_m}^3}$$  \hspace{1cm} (1.13)

where $\sigma$ is the standard deviation of flow values for a given month $m$ computed over
1.4. RESULTS

Figure 1.4. Comparison of skill in individual streamflow forecasts between the models MLR1 and MLR2
1.4. RESULTS

CHAPTER 1. PROBABILISTIC DOWNSCALING

Figure 1.5. Cloud plot comparison of skill in streamflow forecasting for Below Normal (BN), Normal (N), and Above Normal (AN) months between the MLR model and PCR model under both cross-validation and split sample validation. The median lines of the cloud points are projected on each axes for each category.
1.4. RESULTS

46 years of observations, were computed. This assessment was conducted by fitting a linear regression to the monthly data collected from all the basins during each season. Based on this seasonal analysis, we infer that a significant positive relationship between $\Delta RPSS$ and $\gamma$ exists only in the winter season during which the estimated regression slope is greater than zero with the p-values less than 0.05 under both cross validation and split sample validation (Fig. 1.6). It is well known that climate models show more skill in forecasting during the winter season across the Southern US due to the teleconnections associated with ENSO conditions [Devineni and Sankarasubramanian, 2010b; Oh and Sankarasubramanian, 2012]. In addition, the selected basins also exhibit increased inter-annual variability in winter flows resulting in higher variations in skewness.

Similar analysis was also conducted based on monthly average $\Delta RPSS$ for each basin. Results showed that Deep and Rio Grande river basins exhibit a significant positive relationship between monthly $\Delta RPSS$ and $\gamma_m$. This suggests that MLR model performs better than PCR model in forecasting mode when skewness of monthly flows are higher, which typically occurs with higher inter-annual variability in streamflows. The improved performance of basins with high skewness also arises from the ability of MLR model to accommodate skewed nature of streamflow forecasts, whereas the PCR model assumes log-normal distribution, which forces the skewness of the streamflow forecast to be fixed ($\gamma$ is zero in the log-transformed plane but $\gamma = 3CV + CV^3$ in the original plane, where $CV$ is the coefficient of variation of flows).
Figure 1.6. Difference between MLR and PCR performances \( \Delta \text{RPSS} = \text{RPSS}_{MLR} - \text{RPSS}_{PCR} \) with respect to the skewness of monthly streamflows for all the selected basins during Winter season. Positive \( \Delta \text{RPSS} \) means MLR performs better than PCR. Significant positive relationship were found after fitting linear regression.

### Relative Contribution of Model Inputs

As mentioned earlier, the MLR model is forced with probabilistic information from precipitation forecasts along with deterministic streamflow observation from prior months. To further understand the role of each input component in determining the skill of probabilistic streamflow forecasting, MLR modeling was performed under two scenarios - the first one being probabilistic information from precipitation forecasts \( MLR(P) \) alone as a predictor, and the second being the antecedent streamflow observations \( MLR(Q_{t-1}) \) alone as a predictor - for developing streamflow forecasts. The forecasted products from these schemes were then compared to the original MLR modeling scheme which uses both input variables \( MLR(P, Q_{t-1}) \) and were evaluated based on the differences in average rank probability scores (RPS). Fig. 1.7 illustrates the median of RPS differences.
1.4. RESULTS  

### CHAPTER 1. PROBABILISTIC DOWNSCALING

![Graph](image)

Figure 1.7. Decomposed contribution of probabilistic information from Precipitation forecasts (P) and antecedent streamflow observations (Q) as input variables in determining the skill of MLR streamflow forecasts under cross validation approach.

**between each of the scenarios and the original modeling scheme, which basically denotes the role of the input component that was excluded from each scenario.** For instance, the comparison between $MLR(Q_{t-1})$ and $MLR(P, Q_{t-1})$ schemes quantifies the contribution of precipitation forecast (i.e. $[P(P_{BN})_t, P(P_N)_t]$, see Eqn.1.1) in enhancing the skill of streamflow forecasting. Based on this, we see that contribution of streamflow input dominates in comparison to the contribution of precipitation probabilities. This is mostly due to the hydrological persistence of the basins compared to the probabilistic precipitation forecast available from the climate model.
1.4. RESULTS

CHAPTER 1. PROBABILISTIC DOWNSCALING

Ensemble Size Analysis

One of the objectives of our study is to analyze the role of ensemble size of precipitation forecasts in determining the performance of streamflow forecasting via MLR model. In this regard, we decided to use ECHAM4.5 precipitation simulations with 85 ensemble members instead of actual ECHAM4.5 forecasts which has only 24 ensemble members. We performed the cross-validated MLR model 50 times for each ensemble size ranges from 10 to 85 with increments of 5. For a given ensemble size, in each one of the 50 iterations that MLR model was fitted, we randomly selected ensemble members out of the 85 members to form the specified ensemble size which was subsequently used to obtain the probabilities of precipitation forecasts. This helped us to evaluate 50 MLR models under a given ensemble size. However, when the ensemble size increases the number of unique subsets of ensemble members would become more limited which results in less variation in model’s skill, which is also clear in Fig. 1.8. This figure illustrates the results of this analysis for the selected basins during four seasons that are represented in different colors with the shaded colors as the spread of the RPSS’s. This analysis reveals that the skill of the MLR model is almost independent of the precipitation ensemble size since there is not a significant unique trend in the plots. It is important to note that the MLR skill in this figure should not necessarily be same as Fig. 1.2 since we used precipitation forecasts from ECHAM4.5 simulations which has 85 members, whereas all the results presented in figures 1.2-1.6 were developed using retrospective ECHAM4.5 forecasts which has 24 ensemble members.
1.5 **Discussion and Concluding Remarks**

Categorical streamflow forecasts which provides information on the probability of occurrence of below-normal and above-normal flows are useful in contingency planning and in allocating resources for the shift in streamflow potential. Furthermore, categorical forecasting also provides the change in the seasonal streamflows from climatology; hence they are easy to communicate. In this regard, we applied Multinomial Logistic Regression (MLR) as a potential approach to develop probabilistic categorical streamflow forecasts by using...
the probabilistic information in climate forecasts and previous month streamflow. Thus, coarse-scale ensemble-based precipitation forecasts from ECHAM4.5 GCM along with past observations of streamflow from HCDN dataset were used as the predictors of MLR model in order to issue 1-month ahead streamflow forecasts consisted of Below Normal (BN), Normal (N), and Above Normal (AN) streamflow occurrences over six river basins across the US Sunbelt. We compared the performance of MLR model with the traditional approach, Principal Component Regression (PCR), which is commonly used to obtain the categorical precipitation and streamflow forecasts. Our findings demonstrated that MLR and PCR models both have a higher forecasting skill than climatology for almost all the seasons by having the median of RPSS values greater than zero. Also, the analysis of both models under cross validation and split sample validation techniques revealed that the MLR model has a higher skill in producing categorical streamflow forecasts comparing to traditional regression alternatives such as the PCR model. The reason is that the MLR model is capable of utilizing the probabilistic information in the climate forecast ensemble while the PCR model is built based on the mean of the ensembles, thereby not considering the ensemble spread in issuing categorical forecasts. Further, MLR structure is based on the multinomial distribution, which can naturally accommodate the skewness exhibited in the conditional distribution of flows. Hence, MLR model performs more accurately in arid basins and during months with high skewness in flows over humid basins. Grantz et al. [2005] has shown that for snow-dominated river basins (i.e. Rio Grande and Verde river basins), incorporating the information from large-scale climatic forecasts with the winter snowpack initial conditions as models predictors potentially ends up with a higher streamflow forecasting skill. Thus, for basins under snow-melt regime, one could
potentially consider snow water equivalent (SWE) as a predictor instead of streamflow records particularly for predictions during the melting seasons. However, for non snowmelt months, considering antecedent streamflow would be a good strategy to develop monthly streamflow forecasts. For basins under rainfall-runoff regime, consideration of additional predictors such as remotely-sensed soil moisture products (e.g. SMAP) and groundwater levels (e.g. from USGS Climate-Groundwater Response Network) could also be given to enhance the forecasting skill. Even though this study shows the contribution of precipitation forecasts in improving the skill is limited, one could also consider tercile forecasts from multi-model ensemble which in general improves the reliability of climatic probabilistic forecasts [Devineni and Sankarasubramanian, 2010a; Singh and Sankarasubramanian, 2014]. We also infer that the skill of the streamflow forecasts in general improves for large basins in comparison to the smaller basins. For instance, Deep River basin has relatively the lowest skill compared to the rest of the basins. Thus, basins dominated with significant groundwater storage (e.g., ACF) having strong persistence in streamflows are expected to have its’ skill contributed mostly by previous month streamflow conditions.

In this study, two different methods were employed and evaluated to estimate the probabilistic information inside the climate forecasts. The MLR model was forced with tercile precipitation forecasts estimated either by (1) counting the ensemble members lies in each category or by (2) fitting a log-normal distribution to the forecasted ensembles. Results revealed that there is no significant change in the skill of the MLR model between these two information extraction approaches. In addition, the role of the ensemble size of precipitation forecasts was evaluated in estimating the categorical streamflow forecasts using MLR model. Our analyses showed that probabilistic information collected from 10-25
members are enough to estimate streamflow forecasts while further increase in ensemble size did not result in any significant and consistent improvements in the skill of categorical streamflow forecasting. Thus, the proposed MLR approach offers an alternate approach to issue categorical streamflow forecasts that are typically needed in communicating the change in monthly/seasonal streamflow potential.

acknowledgments

We would like to thank the anonymous reviewers whose suggestions and comments helped us in improving the manuscript. In addition, we thank our funding sources: National Science Foundation grants CBET-0954405, CBET-1204368, and CCF-1442909.
Comparison of NOAH3.2 Monthly-to-Seasonal Soil Moisture Forecasts with SMAP Satellite Observations over the Southeast US

Abstract

Providing accurate current and future soil moisture (SM) conditions are critical for initialization of weather models, agricultural planning, and hydrological studies. This study develops monthly to seasonal top layer SM forecasts by implementing NOAH3.2 Land Surface Model using ECHAM4.5 precipitation forecasts over the US Southeast and then evaluates the SM forecasting skill by comparing with remotely sensed SM observations collected by SMAP satellite. Our results indicate potential in developing real-time SM
forecasts with the retrospective 18-month comparison showing a statistically significant correlations of 0.62, 0.57, and 0.59 over 1-3 month lead times respectively. Comparison of the SM forecast with the 2007-2008 monitored drought indexes also indicate potential in issuing SM forecasts to support agricultural planning and operations.

2.1 Introduction

Seasonal climate forecasts are very beneficial for short-term planning and management of water and agriculture systems [Li et al., 2014; Indeje et al., 2006; Sinha et al., 2014]. Most evaluation of climate forecasts has traditionally focused only on the skill in predicting precipitation and streamflow [Hoerling et al., 2009; Devineni et al., 2008]. Limited efforts have focused on the utility of climate forecasts for agriculture systems by evaluating the skill in predicting seasonal to interannual variability in crop yield [Indeje et al., 2006; Hansen et al., 2006]. For instance, rain-fed agriculture heavily depends on actual soil moisture (SM) conditions and the stress that crops are likely to face during the growing phase. Further, improved SM forecasts also provide useful information in capturing the land-surface feedback from the atmosphere [Koster and Suarez, 2001; Eltahir, 1998]. Quantifying and predicting the land surface conditions have also been shown to improve monthly to seasonal forecasts using regional climate models [Berger and Entekhabi, 2001]. Studies have shown the importance of initial land-surface conditions in improving seasonal streamflow forecasting [Wood et al., 2002; Mishra et al., 2012]. However, systematic validation of land-surface attributes such as SM, which is critical for streamflow forecasting and for initializing RCMs, has not been carried out. The main challenge in validating SM forecasts is the limited availability of observed SM over a long period across a large
2.1. INTRODUCTION

SM observations using microwave remote sensing began in the late 1970s with the Scanning Multichannel Microwave Radiometer \cite{owe1992, guha2004} and continued with the Special Sensor Microwave Imager \cite{choudhury1990, lakshmi1997a, lakshmi1997b}. In the past decade with the launch of AMSR (Advanced Microwave Scanning Radiometer) \cite{njoku2003} there was a decade long dataset (2002-2011) for SM from space. This effort continued with the European Space Agency SMOS (Soil Moisture and Ocean Salinity Mission) \cite{kerr2012} and currently we have SMAP (Soil Moisture Active Passive Mission) \cite{entekhabi2010, chan2016, colliander2017, ahmadalipour2017c}. The SMMR and AMSR were C band missions with a penetration depth of a few cm (less than 5), SSM/I - lowest frequency of 19 GHz (penetration depth around 1-2cm) and low spatial resolution (SSM/I - 50km, AMSR - 25km, SMMR - 150km, SMOS - 25km). SMAP is L band (penetration depth 0-5cm) and a high spatial resolution (9km) makes it the best mission for observing SM. However, we face a) challenges in getting observed SM over large domain, and b) limitations of previous other SM products - as mentioned above. Recent remotely-sensed SM products from Soil Moisture Active Passive (SMAP) provide a great opportunity in evaluating the skill of monthly-to-seasonal (M2S) SM forecasts. The main intent of this study is to develop M2S SM forecasts developed from a Land Surface Model (LSM) forced with climate forecasts and to evaluate its skill by comparing it with SM observations from SMAP products over the Southeast US (SEUS). We also evaluate the SM forecasts in predicting 2007-2008 historical drought conditions over the region by comparing with the drought conditions issued by the US drought monitor. The next section briefly describes the data and forecast methodology, followed by the skill evaluation. Finally, we conclude
2.2 Hydroclimatic Data and Forecast Methodology

In this study, we used SMAP observations to validate the ability of NOAH3.2 LSM in forecasting SM at M2S time scales over the SEUS using the NASA’s Land Information System (LIS version 6.2) framework [Kumar et al., 2006]. Our forecasting scheme uses precipitation forecasts from ECHAM4.5 Atmospheric General Circulation Model (AGCM) along with climatology of meteorological variables (except precipitation) included in the phase 2 of the North American Land Data Assimilation System (NLDAS-2) dataset to implement the LSM.

2.2.1 SMAP

The SMAP Mission was launched on January 31, 2015 and has L-band radiometer designed to measure near surface (0-5 cm) SM and land surface freeze/thaw conditions with complete global coverage in 2-3 days. In this paper we utilized the Level-3 SMAP radiometer global daily SM products available at a 0.36° spatial resolution. This data is available for the time period April 2015 to present where we used monthly averaged SM products for a 20 months period from April 2015 to November 2016. SMAP instrument and mission description are provided in greater details in [Entekhabi et al., 2010] and smap.jpl.nasa.gov.
2.2.2 ECHAM4.5 Precipitation Forecasts

Monthly updated precipitation forecasts from ECHAM4.5 AGCM were obtained from the International Research Institute for Climate and Society (IRI) Climate Data Library [Li and Goddard, 2005]. The ECHAM4.5 climate forecasts are issued at a resolution of 2.8° from January 1957 with forecasts available up to 7-month lead time for every month consisting of 24 ensemble members. Constructed analogue Sea Surface Temperature (SST) forecasts have been used to develop the ECHAM4.5 AGCM climate forecasts. The spatio-temporal resolution of the climate forecasts is much coarser than the accepted resolution of the LSM forcing variables, thus statistical downscaling and disaggregation methods were employed in order to address this mismatch. We used the ensemble mean of monthly precipitation forecasts to spatially downscale from 2.8° to 1/8° by developing Principal Component Regression (PCR) with the observed precipitation available from Maurer et al. [2002]. Following that, the downscaled monthly precipitation forecasts were temporally disaggregated to daily time scale using a Kernel nearest neighbor (K-NN) approach in order to develop 1-3 month ahead daily precipitation forecasts. Further details of spatial downscaling and temporal disaggregation methods can be found in Mazrooei et al. [2015].

2.2.3 NOAH3.2 LSM

NOAH LSM has been developed from 1993 through multi-institutional cooperation [Chen et al., 1996; Niu et al., 2011] and has been widely used in operational weather and climate predictions. NOAH is a 1-D column hydrological model which can be executed in either
coupled or uncoupled mode. NOAH3.2 is the version used in our study that requires near-surface atmospheric forcing as inputs and computes surface energy and water balance variables. NOAH3.2 is capable of developing output variables such as streamflow, soil moisture and soil temperature for various vertical layers, snowpack depth, snowpack water equivalent, canopy water content, and other energy flux and water flux terms [Ek et al., 2003]. The NOAH3.2 within the LIS6.2 was primarily used for developing the SM forecast.

2.2.4 Meteorological Forcings of LSM

The meteorological forcing (other than precipitation) for the NOAH LSM were acquired from the phase 2 of the North American Land Data Assimilation System (NLDAS-2) dataset [Mitchell et al., 2004]. NLDAS2 data is available at 1/8° spatial resolution and an hourly temporal scale from 1979 till present. We used the mean of these hourly forcings over a period of 31 years (1979-2010) in order to provide the meteorological forcing for the NOAH model. The initial land-surface conditions of the NOAH model were updated using the last hourly values prior to the forecasting time step.

2.2.5 Experimental Setup

Land-surface Initial Conditions (ICs) are one of the key components of seasonal hydrologic forecasting where the predictability of the surface fluxes associated with the starting conditions of the LSMs (Wood et al., 2016 - JHM). In order to prepare ICs for the forecasting framework, observed meteorological forcings during the month prior to the forecasting month from NLDAS-2 dataset were ingested into NOAH3.2 model in a simulation mode within the NASA’s LIS. The simulated ICs at the end of the month
were then used as ICs for the forecasting period. Under the forecasting mode, spatially
downscaled and temporally disaggregated precipitation forecasts along with the hourly
climatology of the NLDAS-2 forcing (excluding precipitation) and simulated end-of-the-
month ICs were used to run NOAH3.2 LSM on a monthly basis to develop 3-month ahead
monthly forecasts of SM. Mazrooei [2014b] used the above setup to develop streamflow
forecasts for four target basins over the US Sunbelt. The resulted LSM products are at
0.25° spatial resolution and the domain of the modeling is set to Southeast US (defined
as grid cells south of 37°N latitude and east of 96°W longitude).

Mean monthly SM forecasts over top 10cm soil layer was obtained by averaging daily
values of forecasted SM from NOAH3.2 LSM for each month. Since the spatial resolution of
SMAP data (0.36°) is slightly larger than 0.25°, we employed a simple upscaling technique
using a weighted average based on overlying area fractions in order to create SM forecasts
at 0.36° resolution. In order to quantify the forecasting skill for a given grid cell, a simple
uniform bias correction was applied to the SM forecasts based on the difference between
the mean of SMAP observations and the mean of forecasted time series. The skill of the
bias-corrected SM forecasts is then evaluated based on correlation coefficient and Relative
Root Mean Squared Error (RRMSE) by comparing with the respective 18-month values
of SMAP SM observations for three different forecast times. RRMSE for a given lead time
is simply the root mean square error (RMSE) between the monthly SM forecasts and the
corresponding SMAP observations divided by the 18-month average SM from SMAP.
Results

Figure 2.1 (Figure 2.2) shows the RRMSE (correlation coefficient) between the bias-corrected SM forecasts and SMAP observations for 1-3 month lead time. Since we have only 18 data points, grids with insignificant correlations at 95% confidence interval $(\pm 1.96/\sqrt{n}$, where $n = 18$) are plotted in gray scale. As expected, the forecasting error increases from 1-month to 3-month lead time, but the spatial patterns in RRMSE remains the same with increase in lead time. The spatially averaged RRMSE over southeast were equal to 0.040, 0.042 and 0.041 for 1-month, 2-month and 3-month lead times respectively. Among all the 1055 grid cells covering our study domain, only 23% of the grid cells, mostly located in the southeast side of Appalachian mountain, showed consistent increase in RRMSE. This minimal change in RRMSE can be associated with the strong memory (persistence) of SM during seasonal lead times [Koster et al., 2010].

Based on correlation coefficients in Figure 2.2, when the lead time increases from 1 month to 3 months, more grid cells with insignificant correlations appear specifically over the region south side of Appalachian trains. On the other hand, areas with undeveloped or weakly-developed wetland soils [Effland, 2008] such as Mississippi, Alabama, and east side of Texas state indicate increased correlation coefficients in longer forecasting lead times. The average correlation coefficient is 0.62, 0.57 and 0.59 for 1 month, 2 month and 3 month lead times respectively with all the correlations being statistically significant for
2.3. RESULTS

CHAPTER 2. SM FORECASTING

Figure 2.1. RRMSE of the bias corrected 1-3 months ahead soil moisture forecasts based on the SMAP soil moisture observations.

Figure 2.2. Correlation coefficient between 1-3 months ahead soil moisture forecasts with the SMAP soil moisture observations. Grid cells with insignificant correlations (based on 18 months of data) are grayed out.
2.3. RESULTS

CHAPTER 2. SM FORECASTING

Figure 2.3. Scatter plot of soil moisture residuals for two sample grid cells with good and bad forecasting skills. The residuals are centered around the mean of SMAP observations.
2.3. RESULTS

CHAPTER 2. SM FORECASTING

the 18 months of SMAP observations.

To understand how the forecasts capture the variability in SMAP observations, two grid cells with high and low skill in forecasting were selected and anomalies around the mean of SMAP observations are presented (Figure 2.3). The first row shows a scatter plot between the anomalies of the forecasts and SMAP observations for a grid cell with relatively low RRMSE (0.019 on average) and strong correlation (0.726 on average) located in the east part of Georgia. The second row shows similar information for a grid cell located in SC with poor forecasting skill (high RRMSE and low correlation). Under good forecast, the model exhibits limited variability in forecasting below-normal anomalies (i.e., drought conditions) particularly for 2-month and 3-month lead times. For bad forecast, the model in general lacks ability to predict both anomalous conditions.

We also examined the NOHA3.2 SM forecasting skill for the period that the SMAP satellite was not launched. We compared the forecasted SM products with the US Drought Monitor (USDM) data from National Drought Mitigation Center [Svoboda et al., 2002] for the 2007-2008 historical drought over the Southeast [Seager et al., 2009]. For this purpose, we developed 1-month ahead SM forecasts over the southeast where the NOAH3.2 model was initialized on Oct 1st 2007 using NLDAS-2 all meteorological forcings and then the model was forced with 1-month ahead ECHAM4.5 precipitation forecasts and other forcings with corresponding monthly climatology. Then, for a given grid cell, the percentile of the SM forecast was computed from the monthly SM climatological distribution, which is constructed based on the 20-year October SM simulations of NOAH3.2 model (1991-2010). Figure 2.4a demonstrates the forecasted percentile of SM being less than 0.5 which indicates the below-normal (i.e. drought) SM conditions compared to the
2.4 Discussion

The main focus of this study is to develop M2S SM forecasts using NOAH3.2 LSM and ECHAM4.5 precipitation forecasts and also evaluate their skill by comparing with the 18-month remotely-sensed SM observations from SMAP satellite. Both the skill metrics, correlation coefficient and RRMSE, showed that the forecasted SM captured the climatology. Comparing Figure 2.4a with the USDM drought monitoring conditions (from droughtmonitor.unl.edu) issued on Oct 16th 2007 (Figure 2.4b), we see the 1-month ahead SM forecast captures the spatial pattern of drought severity indicating the hard-hit regions with lower percentiles of SM forecast. Thus, initializing the LSM with simulated NLDAS2 conditions and forcing the NOAH3.2 with ECHAM4.5 forecasted precipitation provides useful information for developing real-time M2S SM forecasts.

Figure 2.4. Comparison between the (a) NOAH3.2 retrospective 1 month ahead soil moisture forecasted percentiles with (b) the actual drought conditions from U.S. Drought Monitor (USDM) during historical drought of Oct 2007.
variability in SM from SMAP. SM forecast also show slight reduction in skill as the lead time increases. Comparison with the reported USDM conditions also indicate the SM forecasts capture the spatial variability in drought conditions suggesting potential for developing real-time M2S SM forecasts to inform potential dry/wet conditions to support agricultural planning. To disseminate these forecasts with agencies, the hydroclimatology group at NC State has developed a real-time streamflow and soil moisture forecasting portal (http://climate.ncsu.edu/water/map) that automatically updates the percentiles of SM forecast by comparing it with the climatological distribution of SM obtained from the simulated NOAH3.2 model. Since the skill of the SM forecasts decay little with increase in lead-time, we infer the most of the skill comes from the IC of the LSM. Mazrooei and Sankarasubramanian [2017] show most of the skill in developing tercile streamflow forecasts primarily come from IC. One could also potentially reduce the uncertainty in SM forecasts by considering multiple GCMs and LSMS and perform optimal combination to improve the skill of forecasts [Arumugam et al., 2015; Seo et al., 2016]. Multimodel precipitation forecasts tend to improve the reliability of climate forecasts which could potentially improve the skill in predicting below-normal SM conditions. Further, as more SMAP observations become available, we could evaluate the error in SM prediction and use that information for assimilating LSM ICs for subsequent forecasting. We will also explore 9km soil moisture for comparisons and for initial conditions. However we in the process of downscaling the SMAP 9km soil moisture to 1km by using the MODIS (Moderate Resolution Imaging Spectroradiometer) vegetation and surface temperature. This will offer a higher spatial resolution for model forecasting and skill comparisons.
Chapter 3

Utilizing EnKF Data Assimilation in Improving Monthly Streamflow Forecasts over Contiguous US

Abstract

Among different sources of uncertainty in hydrologic modeling (i.e. model structure, parameter estimation, input data, etc.) consecutive error reduction from model initial conditions can prevent a model from drifting away from reality and consequently improving the model estimates. Most approaches that evaluated correction of initial conditions through data assimilation have predominantly used distributed models focusing primarily on improving model simulations as opposed to evaluating the prediction in a forecasting context. This paper investigates the utility of Ensemble Kalman Filter (EnKF) data assimilation in which available observed streamflow data is exploited to update state
variables of "abcd" conceptual water balance model in order to forecast streamflow for 857 river basins across the conterminous United States (CONUS). Our results demonstrate that skill in predicting seasonal streamflow improves after EnKF application for most basins across the U.S. Sunbelt, whereas EnKF reduces the accuracy in streamflow prediction over snow dominated basins which is primarily due to the bias in snowmelt model. Also, EnKF improves low flow predictions, particularly during the summer season, resulting in 40% reduction in R-RMSE on average. Furthermore, the increased model skill due to EnKF was noticed under a streamflow forecasting, due to the enhancement of initial conditions even though the skill is predominantly dependent on the precipitation forecasts.

### 3.1 Introduction

Water balance models conceptually address the physical processes included in the transformation of the incoming precipitation - rainfall and snowmelt - into outflows such as evapotranspiration, groundwater recharge and runoff. These models are applied at daily to monthly time scales with different degrees of complexity representing the physical processes. Monthly water balance models were first introduced in the 1940s [Thornthwaite, 1948] and later modified in 1950s [Thornthwaite and Mather, 1955, 1957]. Such models are widely used in different research fields such as agricultural studies, climate impacts on streamflow, groundwater assessment, and watershed planning [Denmead and Shaw, 1962; Alley, 1984; Vandewiele and Elias, 1995; Keshavarzi and Bakshi, 2012; Xu and Halldin, 1997; Vörösmarty et al., 2000; Sankarasubramanian and Vogel, 2003].

It is common to consider different types of uncertainty (e.g. imprecise inputs, inexact initial conditions, inaccurate model parameters, and deficiency of model structure) in
hydrological modeling to quantify their impacts on model outputs [Kitanidis and Bras, 1980; Mazrooei et al., 2015; Sinha et al., 2014]. For instance, considerable effort has been made on improving the estimation of model parameters [Beven and Binley, 1992; Vrugt et al., 2003; Wilby, 2005]. Due to high nonlinearity and complexity of the watershed dynamics, it is a tough challenge to perfectly calibrate the models. Even if models are well-calibrated using manual/automatic calibration, the deficiency in the model structure will eventually cause mismatch between output and historical records [Ajami et al., 2004; Moradkhani et al., 2005; Li and Sankarasubramanian, 2012]. Studies have explicitly focused on reducing the model uncertainty by optimally combining multiple model predictions [Hagedorn et al., 2005; Weigel et al., 2008; Singh and Sankarasubramanian, 2014; Yan and Moradkhani, 2016]. The multimodel combination approach is effective since it helps to overcome the spurious confidence of individual models and also reduces the propagated errors by optimal combination [Li et al., 2016; Kaiser Khan et al., 2014].

Several studies have focused on reducing uncertainties from input data and initial model states [Li et al., 2016; Moradkhani and Sorooshian, 2009]. Inaccuracy in estimated initial states causes the model trajectory to drift away from the observations resulting in error in model predictions, particularly in finer temporal resolutions. This divergence can be alleviated by assimilating the recent observations of soil moisture and streamflow in order to improve model state variables [Rüdiger et al., 2006; Vrugt et al., 2006; Komma et al., 2008; Li et al., 2016]. Successful Data Assimilation (DA) techniques were first introduced by meteorologists and applied to improve operational weather forecasting over the last couple of decades [Rabier, 2005; Reichle, 2008]. DA has also been used in oceanography to improve ocean dynamic predictions [Bennett, 1992; Keshavarzi and Kim,
2016; Pham et al., 1998; Bertino et al., 2003]. However, application of DA in hydrological studies has started lately due to lack of suitable large-domain observations of initial states such as soil moisture as well as due to lack of proper quantification of the uncertainty both in observations and hydrologic models [Liu et al., 2012]. Nonetheless, the increasing availability of remotely-sensed observations of soil moisture and groundwater along with increased interest in real-time streamflow forecasting has made DA as an important step in streamflow prediction [Board et al., 2007].

The theory behind data assimilation relies on quantifying errors in the model predictions and observations at the end of a timestep, and then updating the initial states based on the magnitude of error in predicting the precious timestep outputs. DA methods are divided into two types: variational and sequential methods. Variational data assimilation is more commonly used in numerical weather prediction whereas sequential data assimilation is used frequently in hydrological models since it works for distributed data sets [McLaughlin, 2002; Seo et al., 2003; Bertino et al., 2003]. Several sequential data assimilation algorithms have been introduced so far which are explained briefly in Appendix A.

Data assimilation in hydrological studies have focused on both lumped and distributed models with the purpose of improving model simulations [Moradkhani and Sorooshian, 2009]. In snow dominated catchments where the dynamics of stored snowpack plays an important role in streamflow predictions, assimilation of snow water equivalent (SWE) or snow cover extent (SCE) data has resulted in better performance of hydrological models [Sun et al., 2004; Slater and Clark, 2006; Lee et al., 2005; Liston and Hiemstra, 2008; Dechant and Moradkhani, 2011]. Similarly, assimilation of near-surface soil moisture data
3.1. INTRODUCTION

CHAPTER 3. ENKF DATA ASSIMILATION

improves the estimation of land surface fluxes as well as error reduction in streamflow predictions [Pauwels et al., 2001; De Lannoy et al., 2007; Reichle, 2008; Montzka et al., 2011; Kumar et al., 2009]. Several studies reported the enhancement in operational short-term hydrological forecasting with the aid of assimilation of streamflow records in distributed models [Vrugt et al., 2006; Rüdiger et al., 2006; Clark et al., 2008]. A fundamental difference between assimilating variables such as soil moisture and snow observations with assimilating streamflow observations is the fact that in-situ records of streamflow are available over sufficient number of gauge stations for long time periods. On the other hand, ground measurements of snow and soil moisture content are hard and expensive to acquire, especially over a large spatial domain, and most of the DA studies tend to use remotely sensed data of these variables which covers a large domain, although it has short length of observations and contains large measurement errors. In addition, assimilation of the streamflow observation incorporates the information from the entire basin response and also captures the spatial variability within the watershed in generating the runoff. Hence, observed streamflow data could be used to update the state variables over previous time steps based on errors in streamflow predictions [Pauwels and De Lannoy, 2006].

The goal of this paper is to assess the utility of data assimilation in improving streamflow simulations/forecasts using monthly water balance models over the contiguous US (CONUS). We first present the hydroclimatic data set along with the details of the lumped watershed model and the approach of assimilating streamflow records through EnKF. Then the performance of the hydrologic model is assessed with and without EnKF application with respect to the improvements achieved in streamflow simulations and.
3.2. DATA

CHAPTER 3. ENKF DATA ASSIMILATION

Figure 3.1. Locations of the 857 selected HCDN stations across the 18 HUC level 2 geographical regions. HCDN basins are classified snow-melt and rainfall-runoff regimes based on the type of precipitation in influencing the runoff.

Following that, the utility of EnKF is also explored in developing monthly streamflow forecasts. Finally, the findings are summarized along with remarks for application of EnKF in hydrological forecasting.

3.2 Hydroclimatic Data and Climate Forecasts

The temporal span of this study is a 39 water-year period from 1971 to 2010. The first 19 years of the data (1971-1990) is used to calibrate the models (see section 3.3.1) and the rest of the 20 years (1991-2010) for validation. Monthly time series of variables such as potential evapotranspiration (PET), minimum and maximum temperature ($T_{\text{min}}, T_{\text{max}}$), streamflow observations ($Q$), and forecasted and observed precipitation ($P$) were acquired.
3.2. DATA

from multiple sources that are explained next.

3.2.1 Streamflow Data

The streamflow observation data is obtained from the Hydro-climatic Data Network (HCDN) compiled by Slack et al. [1993]. HCDN stations are stream gauges whose records are not influenced by anthropogenic activities such as upstream storage and groundwater pumping [Sankarasubramanian and Vogel, 2003; Vogel and Sankarasubramanian, 2005]. This dataset includes average daily, monthly, and annual records of streamflow over almost 1700 unique stations across the entire country for the past century. Since the start date and end date of the records are not similar for all the stations and have data gaps in the length of their records, we limited our analyses to 857 stations that has a continuous monthly time series of streamflow observations for the period of 1971 to 2010 (Fig. 3.1).

3.2.2 Precipitation and Temperature Data

Monthly time series of precipitation (P) as well as minimum temperature (T_{min}) and maximum temperature (T_{max}) over the period of 1971 to 2010 are obtained by taking the spatial average over each HCDN basin using the Maurer et al. [2002] dataset. This meteorological observations are available at 1/8° spatial resolution on a daily and monthly temporal scale for the CONUS derived by interpolating between observations from weather gauges across the country to develop a gridded observed meteorological dataset. However, interpolated data is likely to contain significant errors under specific circumstances such as complex topographies (e.g. mountain areas), low gauge density in a region, and high seasonality of meteorological variables [Mazrooei, 2014a]
3.2.3 Potential Evapotranspiration Data

Daily timeseries of potential evapotranspiration (PET) over the period of 1971-2010 are obtained from Penman-Monteith method from the Variable Infiltration Capacity (VIC) land surface model setup using observed meteorological data. VIC model is a distributed model that has Penman-Monteith method to estimate PET at 1/8° spatial resolution \cite{Liang et al., 1994, 1996a}. In this study, the 1/8° daily PET estimations were spatially averaged over the HCDN basin boundaries and aggregated to monthly time scale to develop monthly PET.

3.2.4 Precipitation Forecasts

Monthly updated precipitation forecasts consisting of 24 ensemble members at 2.8° spatial resolution and monthly timestep from ECHAM4.5 General Circulation Model (GCM) are obtained from International Research Institute of Climate and Society (IRI) data library \cite{Li and Goddard, 2005}. These retrospective precipitation forecasts are then downscaled to 1/8° spatial resolution using Principal Component Regression (PCR) statistical approach. For a given month, for each 1/8° grid cell, the ensemble mean of monthly precipitation forecasts from four nearest 2.8° neighboring grids are selected as predictors and the corresponding observed monthly precipitation of the finer grid cell is assigned as predictand. Then, an adaptive PCR model is developed using the 25 years of data prior to the forecasting year as training period. For example, to obtain spatially downscaled forecast for a given 1/8° grid for January 1991, antecedent 25 years of January data (i.e., from 1966 to 1990) served as the training period. Since the coarse-scale gridded
precipitation forecasts have high spatial correlation with each other, PCR is selected
to build the statistical model between predictors and predictand. PCR is capable of
transforming the predictors in such a way that minimizes the inter-correlation between
predictors thus each of them has a unique valuable information [Stock and Watson, 2002;
Sankarasubramanian et al., 2008; Li et al., 2014].

3.3 Methodology

3.3.1 ”abcd” Model

The ”abcd” model is a nonlinear, physically-based and lumped water balance model which
accepts monthly precipitation (P) and potential evapotranspiration (PET) as inputs,
obtains soil moisture storage (S) and groundwater storage (G) as state variables, and
producing streamflow (Q) as the output. This model was first introduced by Thomas
[1981] and Thomas et al. [1983] as a decent model in performing regional water resources
assessment. This model was later compared with various monthly water balance models
Vandewiele et al. [1992]. The ”abcd” model has four parameters a, b, c, and d, each has
a specific physical interpretation which is described in table 3.1. We also enhanced the
’abcd’ model with a temperature index method to account for snow storage and melting.

During each monthly simulation, two internal variables termed as ”available wa-
ter” (W_t) and ”evapotranspiration opportunity” (Y_t) are defined and estimated using
precipitation, evapotranspiration, and soil moisture based on following equations:
Table 3.1. Description of the parameters of the 'abcd' model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interpretation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>The propensity of runoff to occur before soil is fully saturated</td>
<td>$0 \leq a \leq 1$</td>
</tr>
<tr>
<td></td>
<td>The upper bound on the accumulation of the evapotranspiration and soil moisture storage. It depends on the ability of the catchment to hold water within the upper soil horizon.</td>
<td>$0 \leq b$</td>
</tr>
<tr>
<td>b</td>
<td>Fraction of streamflow which arises from groundwater discharge (Baseflow Index)</td>
<td>$0 \leq c \leq 1$</td>
</tr>
<tr>
<td>d</td>
<td>The average groundwater residence time</td>
<td>$0 \leq d \leq 1$</td>
</tr>
<tr>
<td>e</td>
<td>Melt Factor (only for snow dominated basins) represents the snowmelt release in a month when the mean temperature is greater than freezing temperature</td>
<td>$0 \leq e$</td>
</tr>
</tbody>
</table>

\[
W_t = P_t + S_{t-1} \tag{3.1}
\]
\[
Y_t = E_t + S_t \tag{3.2}
\]

where \(P_t\) and \(E_t\) represent the precipitation and actual evapotranspiration at timestep \(t\). \(S_{t-1}\) and \(S_t\) denotes the soil moisture storage at the beginning and the end of period \(t\) respectively. \(Y_t\) is postulated as a nonlinear function of \(W_t\) (Eq. 3.3).

\[
Y_t(W_t) = \frac{W_t + b}{2a} - \sqrt{\left(\frac{W_t + b}{2a}\right)^2 - \frac{W_t b}{a}} \tag{3.3}
\]
Evapotranspiration opportunity $Y_t$ is further partitioned into actual evapotranspiration $E_t$ and fraction of soil moisture storage $S_t$ by relating the rate of soil moisture loss to potential evapotranspiration (Eq. 3.4).

$$E_t = Y_t(1 - exp(-PE_t/b)) \quad (3.4)$$

Water available for runoff ($W_t - Y_t$) is then separated into upper zone contribution to runoff (direct runoff) $(1 - c)(W_t - Y_t)$ and recharge to groundwater $c.(W_t - Y_t)$ by the parameter $c$. Groundwater recharge is added to the lower soil zone state variable ($G_{t-1}$) in order to compute the groundwater storage $G_t$ at the end of period $t$. Finally, groundwater discharge to the stream channel is modeled according to the linear relationship $d.G_t$, where $d$ is the fourth parameter. Based on continuity, $G_t$ is updated using Eq. 3.5) and the total streamflow is computed based on Eq. 3.6:

$$G_t = G_{t-1} + c.(W_t - Y_t) - d.G_t \quad (3.5)$$
$$Q_t = (1 - c).(W_t - Y_t) + d.G_t \quad (3.6)$$

“abcd” Model with Snow Component

Most basins in higher latitudes are dominated by snow accumulation and snowmelt processes. Hence, an extension of “abcd” model coupled with a snowmelt model is utilized which accounts for the snow storage and melting process using a temperature index method. Towards this, precipitation is modeled as effective precipitation $P_t^e$ and snowfall based on a comparison between the monthly mean temperature $T_t$ and a base near-freezing
temperature $T_b$. When precipitation falls as snow, $P_t^e$ is zero and a simple snow accounting model computes the snow accumulation $A_t$ as the sum product of the previous accumulated snow storage and snowfall. In warmer months (i.e. $T_t > T_b$) effective precipitation is defined as the sum of snowmelt $M_t$ and rainfall. In addition, snow accumulation is computed by subtracting snowmelt from the accumulated snow storage (Eq. 3.7 and Eq. 3.8).

\[
\begin{align*}
\text{If } T_t \leq T_b & \rightarrow \begin{cases} 
A_t = A_{t-1} + P_t \\
P_t^e = 0
\end{cases} \\
\text{If } T_t > T_b & \rightarrow \begin{cases} 
M_t = \min\{A_{t-1}, e(T_t - T_b)\} \\
A_t = A_{t-1} - M_t \\
P_t^e = P_t + M_t
\end{cases}
\end{align*}
\]

where the parameter $e$ is a new model parameter termed as the melt factor and the snowmelt is estimated using a temperature index method. The minimum operator assures that the modeled snowmelt in a given month cannot exceed the snow storage. Effectively, the 'abcd' model with snowmelt stores the precipitation in the colder months and lags its influence in producing runoff during melting months. The initial snow accumulation $A_0$ in the abcd-snow model was assumed to be zero since the model simulation starts at October and generally basins experience the lowest snow storage around the end of September. In addition, initial groundwater recharge $G_0$ and initial soil moisture storage $S_0$ were assumed to be equal to monthly climatology of streamflow and precipitation respectively.

The abcd-snow model was tested in three scenarios using either minimum, maximum,
3.3. METHODOLOGY

or average monthly temperature as $T_t$ model input. Comparison between the skill of the simulations under these scenarios (not shown) revealed that feeding abcd-snow model with monthly minimum temperature $T_{min}$ results in lower errors between simulated streamflows and the observations.

Rainfall-Runoff and Snowmelt Driven Basins

We assigned the ”snow” or ”no snow” tags to each HCDN basin based on the assumption whether snow ablation plays an important role in basin dynamics or not. Our approach for this classification is based on the grouping suggested by Martinez and Gupta [2010] where they used predominant snow cover data from Sturm et al. [1995] along with modification of the snow versus no-snow contour through visual inspection. Here for simplicity, all the HCDN basins inside the HUC regions 3, 6, 8, 11, 12, 13, and 15 were defined as no snow basins since they are located in the Sunbelt (region south of the 37°N latitude) which has extended summers and short and mild winters. Furthermore, basins in HUC 17 and 18 located in West coast were also identified as no-snow basins and the basins in high latitude regions were listed as snow basins. Based on this, we ended up with 340 no-snow basins and 517 snowmelt dominated basins (Fig. 3.1) where 'abcd' model and 'abcd' model with snowmelt model are used respectively for streamflow simulations.

'abcd' Model Calibration and Validation

The performance of the 'abcd' model is evaluated with 39 years of monthly data for calibration and validation purposes. Correspondingly, the dataset was divided into two parts, i.e. the calibration period containing 19 years of data (water year 1972-1990) and
3.3. METHODOLOGY

CHAPTER 3. ENKF DATA ASSIMILATION

the validation period of 20 years (water year 1991-2010). The data in the calibration period were used to find the optimum values of the model parameters through ‘trust-region-reflective’ optimization algorithm in MATLAB that minimizes the Mean Square Errors (MSE) between simulated flows and the observations.

The ‘abcd’ Model performance during the validation period was then evaluated using Relative Root Mean Square Error (R-RMSE) for each month (Eq. 3.9). Root Mean Square Error (RMSE) is a common used metric to indicate the accuracy in simulated flows [Hyndman and Koehler, 2006].

\[
R - RMSE_m = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Q_{sim}^{m,t} - Q_{obs}^{m,t})^2}
\]

(3.9)

where \(m\) and \(t\) denote a specific month and year respectively, \(Q_{sim}^{m,t}\) and \(Q_{obs}^{m,t}\) refer to the corresponding modeled and observed flows respectively, and \(n\) is the total number of years in the evaluation period (20 years). We considered R-RMSE since it is a normalized version of the RMSE based on the climatology of the streamflow, thus it allows us to compare the model performance across different months and also across the basins.

In addition to model precision, for the purpose of measuring the strength of the model in predicting the variability of the observed flows, Pearson’s correlation coefficient and Nash-Sutcliffe Efficiency (NSE) [Nash and Sutcliffe, 1970] are also considered. In our analysis, Pearson’s correlation is checked for statistical significance at 95% confidence level (i.e. \(1.96/\sqrt{n-3}\) where \(n\) denotes the length of vector of data points used in calculating the correlation [Steel et al., 1960]).
### 3.3.2 Ensemble Kalman Filter (EnKF)

Sequential data assimilation (i.e. filtering) ingests the up-to-date observed information into the model state variables in order to improve model initial conditions prior to next step forecasting. Kalman Filter (KF) is a sequential filter method introduced by Kalman [1960] which is appropriate for models with linear dynamics, while for non-linear models, extended Kalman Filter (EKF) can be used [Jazwinski, 2007]. However, for large scale systems, both methods are computationally demanding and EKF may produce instabilities or divergence if nonlinearities are strong [Miller et al., 1994; Reichle et al., 2002]. To overcome these limitations, the Ensemble Kalman Filter (EnKF) was first introduced by Evensen [1994] and later modified by Burgers et al. [1998]. EnKF propagates the updated ensemble states from previous timestep through a nonlinear model and computes the forecast state error covariance matrix afterwards. Consequently, when observations are provided, the state ensemble can be updated based on a statistical relationship between state covariance and observational error covariance [Moradkhani et al., 2005].

According to the general framework of EnKF [Evensen, 2003], nonlinear function $f$ estimates prior ensemble state $x_i^-$ at time $t$ based on updated (posterior) state from previous timestep $x_i^{+1}$, model parameters set $\theta$, and ensemble of forcing data perturbations $u_i^t$ (Eq. 3.10 and Eq. 3.11). The function $H$ relates the current model state ensemble $x_i^-$ to the observation ensemble $y_i^t$ (Eq. 3.12) generated by adding the noise $\epsilon_i^t$ to the actual observation $y_t$ (Eq. 3.13):
3.3. METHODOLOGY  

CHAPTER 3. ENKF DATA ASSIMILATION

\[ x_{i}^{i-} = f(x_{i-1}^{i+}, \theta, u_{i}^{i}) \] (3.10)

\[ u_{i}^{i} = u_{i} + \eta_{i}^{i}, \quad \eta_{i}^{i} \sim N(0, \sum_{x}^{u}) \] (3.11)

\[ \hat{y}_{i}^{i} = H(x_{i}^{i-}, \theta) \] (3.12)

\[ y_{i}^{i} = y_{i} + \epsilon_{i}^{i}, \quad \epsilon_{i}^{i} \sim N(0, \sum_{y}^{y}) \] (3.13)

where \(i\) denotes the \(i^{th}\) ensemble member \((i = 1, ..., N)\), \(\eta_{i}^{i}\) is the added noise to forcing data with covariance \(\sum_{x}^{u}\), and \(\epsilon_{i}^{i}\) is the noise added to the actual observation with predefined covariance \(\sum_{y}^{y}\). For each timestep \(t\), when the observation becomes available, the posterior model state ensemble can be computed based on a linear correction of the prior state ensemble:

\[ x_{i}^{i+} = x_{i}^{i-} + K_{t}(y_{i}^{i} - \hat{y}_{i}^{i}) \] (3.14)

where \(K_{t}\) refers as the Kalman gain matrix that corrects the model prediction depending on the accuracy of the observation and is quantified as:

\[ K_{t} = \sum_{x}^{x-} H^{T}(H\sum_{x}^{x-} H^{T} + \sum_{y}^{y})^{-1} \] (3.15)

where \(\sum_{x}^{x-}\) represents the prior state error covariance and can be estimated if the true state variables are known \((\sum_{x}^{x-} = E[(x_{i}^{i} - x_{i}^{true})(x_{i}^{i} - x_{i}^{true})^{T}])\). Eq. 3.15 shows the standard form of Kalman gain where the computation of \(\sum_{x}^{x-}\) is necessary. In ensemble based approach (EnKF), Kalman gain can be quantified as Eq. 3.16 which is a modified
version of the standard kalman gain (Eq. 3.15) [Evensen, 2003] and the estimation of
\( \sum^x_t \) is not explicitly needed, thereby making it easier to implement into systems with
high nonlinear dynamics with relatively low computational complexity:

\[
K_t = \sum^x_t \hat{y} - (\sum^\hat{y} + \sum^y_t)^{-1}
\]

(3.16)

where \( \sum^x_t \hat{y} \) is the forecast cross covariance between the prior state variables \( x^t \)
and the measurement prediction \( \hat{y} \), and \( \sum^\hat{y} \) is the forecasts error covariance matrix of the
measurement prediction \( \hat{y} \).

**EnKF Application on "abcd" Model**

By its design, EnKF can easily be employed in various hydrologic models. Here, we
denote the "abcd" model as a nonlinear system \( H(\theta, X^t_{i+1}, U^t_i) \) which accepts the model
parameter set \( \theta \in [a, b, c, d] \) (Table 3.1), initial model states ensemble \( X^t_{i-1} \in [S^t_{i-1}, G^t_{i-1}] \)
consisted of two state variables soil moisture storage and groundwater storage, and
ensemble of perturbed forcing data \( U^t_i \in [P^t_i, PE^t_i] \) containing precipitation and potential
evapotranspiration. The \( H \) model then estimates the prior state variables and ensemble
of streamflow \( \hat{Q}^t_i \) at time \( t \) using equations 3.1-3.6 in the 'abcd' model formulation:

\[
[\hat{Q}^t_i, X^t_i] = H(\theta, X^t_{i+1}, U^t_i)
\]

(3.17)

For the snow dominated basins, the parameter \( e \) and perturbed monthly minimum
temperature data \( T_{min} \) are also added to the inputs of \( H \). When the streamflow observation
\( Q_t \) is available, updated ensembles of soil moisture storage and groundwater storage at
time \( t \) can be computed in order to be used in the next time step model integration. EnKF implementation in the "abcd" model is listed in three steps (i.e. initialization, forecast, and update steps) where the last two steps are the natural steps in EnKF approach \[\text{Burgers et al.}, 1998; \text{Reichle et al.}, 2002\]. Here, we apply those steps in the context of 'abcd' model for streamflow forecasting described as following:

i . **Initialization:** Model parameters are estimated using "abcd" model without EnKF by minimizing the MSE during calibration period (water years 1972-1990) as described in section 3.3.1. The initial ensemble of the two state variables \( S_{0,1 \times N} \) and \( G_{0,1 \times N} \) are generated using normal distribution with mean and variance equal to the mean and 0.5% variance of the monthly climatological precipitation and streamflow records of September respectively and stored into the state matrix \( X_{0,2 \times N} \). Similarly, ensemble of perturbed forcing data is created around the observed value at time \( t \) with variance equal to 0.5% variance of the forcings during that month over the calibration period.

ii . **Forecast step:** For each ensemble member \( i \), estimation of streamflow and prior model state \( \hat{\mathbf{Q}}_t, \mathbf{X}_t \) are developed using Eq. 3.17 which is a generalized form of the equations expressed in section 3.3.1.

iii . **Update step:** An ensemble of streamflow observation \( \mathbf{Q}_t \) is developed by perturbing the observation at time \( t \) with a Gaussian noise having variance equal to 0.5% variance of that month streamflow. The covariance matrices \( \sum_t \hat{\mathbf{Q}} \) and \( \sum_t \mathbf{Q} \) are then calculated using the estimated and observed streamflow ensembles respectively. In addition, we compute the cross covariance matrix between the prior state and estimated flow \( \sum_t \mathbf{X} \hat{\mathbf{Q}} \) using Eq. 3.18. Next, Kalman gain is quantified by Eq. 3.19 which is a reshaped form
of Eq. 3.16 and finally model states are updated through Eq. 3.20.

\[
\sum_t X^\hat{Q}_t = \left( X_i - E[X_i] \right) \left( \hat{Q}_t - E[\hat{Q}_t] \right)^T \quad (3.18)
\]

\[
K_t = \sum_t \hat{Q}_t \left( \sum_t \hat{Q} + \sum_t^Q \right)^{-1} \quad (3.19)
\]

\[
X^{i+}_t = X^{i-}_t + K_t (Q^i_t - \hat{Q}_t) \quad (3.20)
\]

3.4 Results

3.4.1 Assessment of ”abcd” Model Performance without EnKF

For each basin, monthly streamflow simulations during water years 1991-2010 from ”abcd” model without EnKF were evaluated in terms of R-RMSE (Eq. 3.9) and Pearson’s correlation coefficient over four seasons (Fig. 3.2). Results show that simulated flows during winter and fall seasons have the lowest errors compared to other seasons, while ’abcd’ model performs poorly during summer season. The ’abcd’ model has the lowest errors over the Southeast and some parts of the West coast, and the best performance is displayed in central and southern regions of Appalachian highlands. Correlation maps in Fig. 3.2 indicate significant correlation in predicting observed flows over approximately 91% of the selected HCDN stations across the CONUS. This shows the ability of ’abcd’ model in simulating the variability of streamflows during all the seasons particularly over Southeast, Mideast, Northeast, Midwest, and West Coast basins. The skill in streamflow prediction using ’abcd’ model is in line with the results of Martinez and Gupta [2010] and
3.4. RESULTS

Figure 3.2. Performance of 'abcd' model expressed as R-RMSE (first row) and Pearson’s Correlation Coefficient (second row) in simulating observed flows over the CONUS. Grayed-out basins indicate high errors in simulated flows ($R - RMSE > 1$) or insignificant correlation coefficient at 95% confidence level.

Hay and McCabe [2002] studies, in which the performance of a conceptual water balance model is evaluated over the CONUS.

3.4.2 Role of EnKF in Improving 'abcd' Model

Based on the EnKF methodology described earlier, state variables including soil moisture storage and groundwater storage were updated by assimilating monthly streamflow observations at each model iteration. R-RMSE of pure "abcd" simulations as well as simulations with EnKF (abcd-EnKF) for a specific month were compared and their difference is presented in Fig.3.3. The R-RMSE of the abcd-EnKF is computed based on the ensemble mean of the simulations with ensemble size $N = 100$. From Fig.3.3, EnKF is able to reduce the errors in simulated flows during all 12 months over most of the
3.4. RESULTS

CHAPTER 3. ENKF DATA ASSIMILATION

Figure 3.3. Differences in R-RMSE in simulating observed flows with EnKF and without EnKF through 'abcd' model. Positive (negative) values indicate the improvements (degradations) introduced to the simulated flows due to EnKF.

Figure 3.4. Differences in R-RMSE in simulating observed low (high) flows due to EnKF in 'abcd' model shown in upper (lower) row. Low (high) flows correspond to months where the flows are lesser (greater) than the 10th (90th) percentile of climatology. Positive (negative) values in the figure indicate decreased (increased) error in simulations due to application of EnKF.
catchments in Southeast, South, and Southwest. Such error reduction is experienced over 78% of the basins located in the U.S. Sunbelt. Furthermore, the strongest improvement from EnKF application is felt from September to February months for most basins in the central HUC regions (i.e. 10,11,12,13, and 14) resulting in an average 40% decrease in monthly R-RMSE. However, EnKF performance is poor during November to May over Northeast and Midwest regions where most of the catchments are snow dominated. Our analysis show that abcd-EnKF with snow model underestimates the flows from November to March and overestimates in the spring season (figure not shown).

Further analyses were conducted on the skill of the model in simulating low and high flows (streamflows less than 10\textsuperscript{th} percentile and greater than 90\textsuperscript{th} percentile respectively according to monthly climatology of the observations). From Fig. 3.4 we infer that abcd-EnKF is more advantageous in low flow predictions compared to high flows, where at least 80% of stations displayed 7% reduction in R-RMSE on average in low flow predictions after utilizing EnKF. This emphasizes the importance of data assimilation in drought management. In the case of high flows, we see EnKF increases the errors over most basins in the Northern half of US. However, few basins in HUCs 14 and 16 showed a significant improvement during fall and winter seasons. Over US Sunbelt, basins located in Florida, Georgia, South Carolina and Southern California can benefit from EnKF in reducing the errors in predicting high flows.

3.4.3 Scenarios of State Variables in EnKF

By definition, ”abcd” model has two state variables (i.e. soil moisture storage and groundwater storage) while in EnKF we can decide to select both of them simultaneously
3.4. RESULTS

CHAPTER 3. ENKF DATA ASSIMILATION

Figure 3.5. NSE in simulating observed flows through application of EnKF on state variables (a) Soil moisture storage ($S_t$) and groundwater storage ($G_t$), (b) only groundwater ($G_t$), and (c) only soil moisture ($S_t$). Stations with $NSE < 0$ are not shown.

or just one of them to be updated when the model is integrating forward in time. Based on this, three scenarios of simulations with different sets of updated state variables (i.e. both $G_t$ and $S_t$, only $G_t$, and only $S_t$) were performed and analyzed. Fig 3.5 illustrates the model skill in streamflow prediction through the defined EnKF scenarios based on NSE coefficient computed over the validation period. Only the stations with NSE value greater than zero are printed here which represents the superiority of the model performance over climatology skill. In addition, total number of the plotted stations is presented next to each map along with the median of their NSE values in parenthesis. As we can see in this figure, the spatial distribution of NSE values are almost same in the three scenarios, meaning that different sets of state variables does not have any role in determining the EnKF performance, thus updating both state variables of ‘abcd’ model is considered subsequently.
3.4. RESULTS

CHAPTER 3. ENKF DATA ASSIMILATION

Figure 3.6. Seasonal R-RMSE (first row) and correlation (second row) in forecasting observed monthly flows using 'abcd' model (without EnKF) during four seasons. Stations with R-RMSE values greater than 1 or insignificant correlation coefficients (corresponding to 95% confidence level) are shown as gray.

Figure 3.7. Difference in R-RMSE ($\Delta$R-RMSE = R-RMSE$_{abcd}$ – R-RMSE$_{abcd-EnKF}$) and correlation coefficient ($\Delta$Correlation = CORR$_{abcd-EnKF}$ – CORR$_{abcd}$) within the forecasted streamflows from "abcd" model applied with and without EnKF. Positive (negative) values indicate the improved (degraded) forecasting skill in terms of both R-RMSE and correlation coefficients.
3.4.4 EnKF Usefulness in Operational Streamflow Forecasting

To develop 1-month ahead streamflow forecasts, spatially downscaled and temporally disaggregated precipitation forecasts from ECHAM4.5 GCM were forced into the 'abcd' model during the validation period. In this fashion, optimal model parameters are estimated during the calibration period using observed forcings. As we infer from Fig.3.6, most of the stations exhibit poor forecasting skill in terms of R-RMSE as well as correlation. Comparison between skills of 1-month ahead forecasts and simulation (Fig.3.2) shows that the main source of error in the forecasted streamflow is due to imprecise precipitation forcings. This forcing error propagates into model state variables when the model integrates forward in time, which could be corrected by DA. Following this, 1-month ahead precipitation forecasts were applied to ’abcd’ model with EnKF and its performance was compared to pure ”abcd” model forecasts to quantify the role of EnKF application in streamflow forecasting (Fig.3.7).

Results illustrate that EnKF consistently improves forecasting over majority of stations in Southeast, South, Southwest and West coast regions, however it increases the forecasting error over Northwest region. During fall and winter seasons, most basins located in Western U.S. can benefit from EnKF since the forecast performance in terms of R-RMSE and correlation coefficient has been improved. EnKF also improves the forecasting skill over Midwest region during summer and fall seasons especially in capturing the variability in streamflow. Fig.3.7 only presents the changes in the verification metrics and in order to determine the basins with significant skill in forecasting, it should be compared with Fig.3.6. Based on this comparison, we found out that only few stations have R-RMSE lesser than one, while a significant number of stations exhibit significant correlation in
forecasting streamflow. This suggests that abcd-EnKF is more successful in capturing the variability comparing to bias reduction in forecasting monthly streamflows. EnKF also improves the forecasted streamflows over Southeast stations during fall and winter and over West coast region during spring and summer seasons due to correction of initial conditions.

3.5 Summary and Discussion

This paper tries to highlight the application of Ensemble Kalman Filter (EnKF) in hydrologic data assimilation using "abcd" conceptual water balance model. Monthly streamflow observations from the HCDN dataset was used to update the two defined state variables in the "abcd" model, soil moisture and ground water. Totally, 857 river basins across the conterminous U.S. (CONUS) were classified as either "snow" or "no snow" catchments. Observed meteorological forcings during the water years 1971-2010 were used to run the model with/without EnKF implementation over each basin, in which the first 19 years of data were used for estimation of model parameters and the rest 20 years for validation of the simulated flows. Furthermore, 1-month ahead forecasted precipitation data from ECHAM4.5 GCM were spatially downscaled and temporally disaggregated and forced into the 'abcd' model in order to assess the model performance. The comparison between the skill of 1-month ahead forecasts from 'abcd' model and abcd-EnKF model was conducted to assess the role of EnKF in an actual streamflow forecasting scheme. Our findings suggest that simulations from "abcd" model have an acceptable performance over Southeast, Mideast, and West coast particularly during fall and winter seasons. Further, 91% of the stations exhibit significant correlation coefficient between simulations and
observations during all seasons. After implementing EnKF, error reduction is noticeable in streamflow simulations as well as forecasts over the stations located in Southeast, South, Southwest, and West coast regions. Nevertheless, EnKF extremely increases the errors in snow dominated basins located in Northwest and Midwest during fall and winter seasons, which is partially due to underestimation and overestimation of flows during snow accumulation and snow melt periods respectively. Further analyses on the low/high flow simulations revealed that EnKF is more effective in error reduction of low flows compared to high flows and this is in harmony with the conclusion of Aubert et al. [2003]. Three sets of state variables (i.e. both $G_t$ and $S_t$, only $G_t$, and only $S_t$) were considered for updating in EnKF and our analyses did not find any significant differences in improving the model performance. Overall, EnKF enhances the model performance of monthly simulations in at least 30% of the stations while this number increases to its maximum of 75% during Aug-Oct-Nov months.

Several recommendations can be considered for future work as a subsequence of this study. Our EnKF methodology uses a fixed value of input error variance and observational error while these parameters can vary either during each time step. Hence, based on the basin’s location and characteristics further assessment can be done on tuning these parameters within DA [Moradkhani et al., 2005]. Also, it seems that the ”snow” vs. ”no-snow” basin classification method that we took in this study influences the skill of the models particularly when we deal with models with incorporated EnKF (e.g. Fig.3.3 January). The difference in the results between two regions (e.g. boundary between HUCs 3 and 2 in Mideast) demonstrate that the hypothesis of the ”abcd” model with snow component or classification of snow regions requires considerable attention. In this regard,
employing a more complex method in estimating snow accumulation/melt can reduce the model errors, while the increase in the computational cost of the model iterations is expected. According to Aubert et al. [2003] coupled assimilation of both streamflow and soil moisture observations into conceptual models results in better performance in terms of flow forecasting. While the lack of a comprehensive dataset of soil moisture records is still a barrier, this analysis show that observed streamflow could be used in DA applications and for correction of model states towards better streamflow forecasting.
Chapter 4

Variational Assimilation of
Gauge-Measured Streamflow
Records in Monthly Streamflow
Simulation and Forecasting

Abstract

Uncertainties associated with the initial conditions of soil moisture have been recognized as one of the main sources of errors in hydrologic modeling, specifically in a rainfall-runoff regime. Apart from the recent advances in Data Assimilation (DA) for improving hydrologic predictions, this study explores variational assimilation (VAR) of gauge-measured daily streamflow data for improving initial state of soil moisture content for Variable Infiltration Capacity (VIC) model. This research is conducted for the Tar River basin in NC during a
20-year period (1991-2010). Furthermore, the roles of two critical parameters of VAR-DA, the optimal frequency of data assimilation application and optimal length of assimilation window are assessed in improving streamflow predictions. Results suggest that updating VIC model’s initial conditions every 20 days using a 7-day assimilation window can result in the most improved predictions evaluated by Kling-Gupta Efficiency (KGE) and Nash-Sutcliffe Efficiency (NSE). In addition, the potential gain from VAR-DA framework is quantified and discussed under two streamflow forecasting schemes - ECHAM4.5 1-month ahead precipitation forecasts for deterministic streamflow forecasting and Ensemble Streamflow Prediction (ESP) approach for probabilistic forecasting - for developing 1-month ahead forecasts. This study also evaluates the persistence of the DA positive effect in the monthly predictions by quantifying the enhanced accuracy in daily flows. Analyses show that the effectiveness of the updated initial conditions from DA lasts only for a shorter time scale - a week to ten days - in monthly streamflow predictions.

4.1 Introduction

Accurate quantification of uncertainty in weather and hydrologic forecasts helps water resources managers towards better decision making. In recent years, substantial effort has been devoted to address the importance of the hydrologic uncertainty quantification especially in the forecasting process [Schaake et al., 2006; Pappenberger and Beven, 2006; Brown, 2010; Ahmadalipour et al., 2017a]. Several major sources of uncertainty have been identified in hydrologic predictions, that include uncertainty in model structure and model parameters, initial hydrologic conditions, and hydrometeorologic forcings [Ajami et al., 2007; Salamon and Fegen, 2010]. Thus, it is essential to effectively quantify and reduce
these uncertainties for developing accurate and reliable hydrologic forecast products to support risk-based decision making [Weerts et al., 2010; Pappenberger et al., 2008, 2011; Charney et al., 1969; Sankarasubramanian et al., 2009; Li et al., 2014; Arumugam et al., 2015].

Data Assimilation (DA) is considered as an effective technique that is able to reduce the errors in model states and parameters and consequently improve the model predictability. The basic theory behind DA is to optimally combine the information from model predictions and past available observations to update the model initial conditions. DA have been widely used in meteorology, oceanography, and operational weather services over the last couple of decades, specifically after the launch of first weather satellites. The role of DA has been well proven in weather forecasting, yet its application in hydrologic studies is in its infancy [Liu et al., 2012]. Meanwhile, considerable advances in theoretical development of hydrologic data assimilation techniques have been proposed from simple direct insertion methods to complex sequential and smoothing filtering methods [Kumar et al., 2009; DeChant and Moradkhani, 2012; Wang and Cai, 2008]. In land data assimilation, ensemble Kalman filter (EnKF) is one of the most frequent and used methods in improving the model states, however its application to distributed hydrologic models requires sophisticated computational resources. On the other hand, hydrologic studies that have employed Variational (VAR) data assimilation technique [Seo et al., 2003; Liu and Gupta, 2007; Seo et al., 2009; Lai et al., 2014] suggested proper simplifications due to its technical complexities. VAR is a commonly used method in global atmospheric assimilation schemes and has much unexploited potential in operational hydrology. In spite of the great research into hydrologic DA, only few studies have been focused on DA
4.1. INTRODUCTION

formulation and application in real-time operational forecasting for short-range lead times and attempted to quantify the performance gain due to DA [Seo et al., 2003; Vrugt et al., 2006].

The abundance of hydrologic observations collected during last decades (in-situ or remote sensing) has motivated the need to use them for improving hydrologic predictions through data assimilation techniques.

Particularly in recent years, the potential has been increased due to availability of remotely sensed data of soil moisture and snow cover area/extent from satellite observations [Pauwels et al., 2001; Andreadis and Lettenmaier, 2006; Clark et al., 2008; Reichle et al., 2008]. Remotely sensed observations provide a focal point for the DA community since it is the only observed information available for many parts of the world and can be collected during 6-hourly periods, thus it has a great potential in atmospheric modeling and weather predictions. However, historical in-situ observations such as gauge-measured streamflow records are available for a much longer period of time (i.e. almost a decade) and contain very less measurement errors compared to satellite observations. Hence, assimilating streamflow observations for correcting model initial states provides a great opportunity to improve streamflow predictions.

4.1.1 Previous Work

Seo et al. [2003] employed variational assimilation (VAR) in order to assimilate hydrologic observations such as streamflow and hydrometeorological observations such as precipitation for improving operational hydrological forecasting. A coupled hydrologic model consisting of Sacramento (SAC) model and unit hydrograph model (UH) were used to issue streamflow
forecasts. Their study reveals that VAR approach significantly improved the accuracy of the streamflow forecasts for three basins at short lead times (40 hours). They also showed that VAR technique is more appropriate compared to other DA methods like Kalman filter in operational forecasting problems since it has less computational demand. Rüdiger et al. [2006] conducted a synthetic study on three catchments in Australia in order to assess the role of assimilating observed streamflow for soil moisture retrieval. This study uses the Catchment Land Surface Model (CLSM) along with variational assimilation for an initial state optimization by minimizing deviation from observed data. It was found that the assimilation of streamflow has a significant improvement in the retrieval of soil moisture profile, but showed limitations in retrieving the surface soil moisture state. Application of simultaneous optimization and data assimilation (SODA) approach uses EnKF to assimilate streamflow observations along with a stochastic optimization approach in order to find the optimal model parameters for operational streamflow forecasting [Vrugt et al., 2006]. They implemented the SODA method into the Sacramento watershed model and their results demonstrate a 30% improvement in the model forecast performance. They also showed that a higher skill is exhibited with low flows. Clark et al. [2008] explored the application of the ensemble Kalman filter with streamflow observations in a distributed hydrologic model TopNET. They found that EnKF approach is not an appropriate choice for the situations when there is a nonlinear relationship between state variables and observations. In another experiment, they transformed streamflow to log space before computation of the error covariance matrix and improved the EnKF performance. In addition, they discussed the importance of time lags between model states and streamflow that can be taken into account by a bias-aware method named as retrospective ensemble
Kalman filter (REnKF) according to Pauwels and De Lannoy [2006].

For catchments located in a rainfall-runoff regime, the hydrologic predictability is mostly dependent on the true estimation of the initial soil moisture conditions [Mahanama et al., 2012]. Furthermore in such basins, rainfall is the major contributor to the streamflow, in another word it is the key source of uncertainty [Beven and Kirkby, 1979]. Thus the skill of streamflow forecasting for such basins can be enhanced through advances in climatic forecasting and improving the model state conditions. Monthly precipitation forecasts from global climate models contain inevitable amount of uncertainty that is latter propagated into streamflow forecasts. Although, comparing to hydrologic forecasts from Ensemble Streamflow Prediction (ESP) approach [Day, 1985; Wood et al., 2016; Sinha et al., 2014], there is still more information in streamflow forecasts derived from climate forecasts. Majority of past hydrologic DA studies have focused either on a conceptual hydrologic model or a lumped distributed model which solely considers the terrestrial hydrologic dynamics. Another motivation in this study is to explore the usability of data assimilation into Land Surface Models (LSM) in which more complex modeling components such as interactions between land and atmosphere, vegetation dynamics, land use and land change, soil temperature, and etc. are taken into account with a finer modeling time steps. [Cox et al., 2000; Feddema et al., 2005; Bonan and Levis, 2006; Zeng, 2010]. Also, most of the hydrologic DA studies explored the role of data assimilation in short-range forecasting lead times from hourly to weekly. Here, streamflow observations at basin outlet are assimilated into a LSM in order to assess the improvements in the streamflow forecasting for longer forecasting lead times (e.g. monthly).

This manuscript is organized as follows: Section 4.2 provides details about the used
4.2. DATA

4.2 Study Area and Data

4.2.1 Tar River Basin

The Tar River Basin (Fig.4.1) located in North Carolina state is selected for this study. The selected study area is part of Tar-Pamlico river basin which consists of two sub-basins, Tar River Headwaters (HUC03020101) and Fishing Watershed (HUC03020102). The total drainage area of this domain is 2,183 square miles and the basin has experienced several flood events due to various hurricanes such as Fran (1996), Floyd (1999), and Isabel (2003).

4.2.2 Streamflow Observations

Daily Observed streamflow data for the period 1949-2010 is obtained from the US Geological Survey (USGS) at Tar River at Tarboro (USGS#02083500). This site is classified as one of the virgin basins in US and is included in the Hydro-Climatic Data Network (HCDN) [Slack et al., 1993] database since it receives minimal influence from anthropogenic factors such as upstream reservoir and groundwater pumping [Sinha and Sankarasubramanian, 2012; Vogel and Sankarasubramanian, 2005]. Fig. 4.2 shows the seasonality of the observed flows for the Tar River basin. Typically, the basin experiences
4.2. DATA

4.2.1 Data Set

For hydrologic modeling purposes, observed meteorological data is obtained from Maurer et al. [2002] dataset over the Tar River basin (Fig. 4.1). This data is derived by interpolating the weather gauge observations across the country and it is available at 1/8° (~150 km² cell area) spatial resolution and a daily temporal scale from January 1949 till December 2010. The meteorological forcing time series used as the hydrologic model input variables are listed as daily total precipitation [mm/day], maximum and minimum daily temperatures, and precipitation intensity [mm/hr].

4.2.2 Hydrological Model

The hydrological model used in this study is VIC (Variable Infiltration Capacity), which is a process-based model that simulates the hydrological cycle at the catchment scale. VIC is widely used for regional and global hydrological studies due to its ability to simulate a wide range of hydrological processes including infiltration, soil moisture, and surface runoff.

4.2.3 Observed Meteorological Forcings

For hydrologic modeling purposes, observed meteorological data is obtained from Maurer et al. [2002] dataset over the Tar River basin (Fig. 4.1). This data is derived by interpolating the weather gauge observations across the country and it is available at 1/8° (~150 km² cell area) spatial resolution and a daily temporal scale from January 1949 till December 2010. The meteorological forcing time series used as the hydrologic model input variables are listed as daily total precipitation [mm/day], maximum and minimum daily temperatures, and precipitation intensity [mm/hr].
4.2. DATA

Figure 4.2. Seasonality of Tar River Basin

[°C], and average 10-meter wind speed [m/s]. These four variables are the minimum set of variables required by the hydrologic model in this study. In order to estimate terrestrial fluxes, other forcing variables are also required by the hydrologic model, which are described in section 4.3.1.

4.2.4 ECHAM4.5 Precipitation Forecasts

In order to develop hydrologic forecasts, climatic forecasts from the ECHAM4.5 general circulation model (GCM) are obtained from the International Research Institute of Climate and Society (IRI) data library [Li and Goddard, 2005]. The ECHAM4.5 climatic forecasts are available at monthly time scale from 1957 till present at a coarse spatial resolution of 2.8° for up to 7-month lead time. Constructed analogue Sea Surface Temperature (SST) forecasts were utilized to develop the ECHAM4.5 climate forecasts in 24 ensemble members.
Nevertheless, the selected hydrologic model in this study (section 4.3.1) accepts forcing with a finer spatio-temporal resolution than the raw product of ECHAM4.5 GCM, hence we downscaled ECHAM4.5 forecasts to 1/8° resolution and disaggregated the monthly forecasts to daily time steps. For spatial downscaling, a Principal Component Regression (PCR) model is developed between the ensemble mean of the 2.8° monthly precipitation forecasts and the monthly observed precipitation at 1/8° resolution from Maurer et al. [2002]. PCR model is trained using 1957-1990. Next, a temporal disaggregation technique based on a K-NN algorithm in time dimension [Prairie et al., 2007] were employed in order to resample daily precipitation from the downscaled monthly forecasts. Further explanation of the spatial downscaling and temporal disaggregation schemes as well as the skill of precipitation forecasts over the Tar River basin can be found in Appendix B.

4.3 Methodology

4.3.1 Hydrologic Model

The model employed in this study is consisted of two components: 1) Variable Infiltration Capacity (VIC) [Wood et al., 1992; Liang et al., 1994, 1996b] model is executed which is a semi-distributed physically-based hydrologic model, capable of simulating runoff and other terrestrial variables for a set of grid cells independently and 2) a routing model [Lohmann et al., 1996, 1998] is then performed to transport surface runoff and baseflow from each grid cell to the river system and eventually estimate the streamflow at the basin outlet. Here, VIC model is performed in 3-hour time steps, considering 3 soil layers over 40 grid cells of Tar River basin (Fig. 4.1). The direct runoff is quantified as the excess
water from saturation and infiltration at the top two soil layers and the baseflow is derived from the bottom soil layer using a generalized version of the Arno model [Franchini and Pacciani, 1991].

Here, the VIC model is manually calibrated for Tar River basin using 40 years of data (1951-1990) by minimizing Root Mean Square Error (RMSE) between simulated and observed flows [Sinha et al., 2014]. The parameters considered in the calibration process are listed as maximum soil moisture content, infiltration shape parameter, evapotranspiration parameter, and baseflow parameter [Li et al., 2015]. VIC model is typically implemented with daily meteorological forcings expressed in section 4.2.3, and it evenly divides the daily totals into sub-daily values to run at the model timestep. Other than those variables, VIC internally estimates forcings such as air pressure, relative humidity, incoming radiations, etc. through various complex algorithms [Kimball et al., 1997; Thornton and Running, 1999; Bras, 1990].

Subsequently after VIC simulations, runoff and baseflow fluxes are routed to the edge of each grid cell based on Lohmann et al. [1996] routing model using couple of input files containing information of each grid cell’s flow direction, flow velocity, unit hydrograph, etc. derived from delineating the watershed.

### 4.3.2 Variational Data Assimilation

4D (1D in time, 3D in space) variational data assimilation (VAR) is a widely used approach which was introduced and being employed in atmospheric and oceanography for decades, while it is still at infancy in hydrologic studies [Ide et al., 1997; Liu et al., 2012]. Basically, VAR methods are smoothers that seeks the optimal initial conditions so
that the model prediction best fits the observation within a user-specified assimilation 
window. In addition, depending on the hydrologic problem, the VAR technique can be 
utilized as one-dimensional (1D-VAR), three-dimensional (3D-VAR), or four-dimensional 
(4D-VAR) \cite{Liu and Gupta, 2007} where time always comes in as the first dimension and 
the rest defines the dimensions in space. A standard variational DA problem tries to 
minimize the cost function $J$ (Eqn. 4.1) which is a weighted sum of squared distances from 
the background state ($J_b$ cost function of background) as well as the difference between 
simulated flows and the observed flows ($J_o$ cost function of observations) distributed over 
a time interval. In another word, $J_b$ component is the resistance of the background state 
to change and $J_o$ component penalizes the discrepancy between the observations and the 
model simulations:

$$
J(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2}(y - H[x])^T R^{-1} (y - H[x])
$$

where $x$ is the model state, $x_b$ denotes the background state, $y$ is observation, $H[x]$ is 
the operator that maps the model state $x$ to the observation field, $B$ is the background 
error covariance matrix, and $R$ is the observational error covariance matrix. $B$ contains 
information on the importance of the model background state for different locations and 
it is practically difficult to measure and treat in VAR problems. There are few known 
methods such as ensemble method or analysis of forecast differences in order to estimate 
$B$ but most of the times since the "true" state is unknown it brings up a major challenge 
in the problem \cite{Bannister, 2008}.

Furthermore, minimizing $J$ cost function and finding the optimal solution of variational
4.3. METHODOLOGY

CHAPTER 4. VARIATIONAL DA IN VIC

Figure 4.3. Schematic of the variational data assimilation approach. Black squares represent the model states (not in the streamflow dimension), triangles are the streamflow observations, UF and AW denotes DA Update Frequency and Assimilation Window parameters respectively. The hydrologic model is then initialized using the state conditions achieved either from Open Loop (OL) simulation or Data Assimilated (DA) simulations.

Techniques can be computationally expensive when the number of unknown states and/or parameters increases, and sometimes for nonlinear and high dimensional hydrologic applications, it is very difficult and often impossible to render a comprehensive solution. Accordingly, for VAR applications, approximations and simplifications are taken into account such as linearizing the state and/or observation equations. Still it is challenging to solve VAR explicitly, even after simplifications, thus most of the studies use random search, gradient search, or numerical techniques such as a reverse adjoint model in order to compute the gradient of the cost function of the controlled state vector.
4.3.3 Implementation of VAR in VIC model

The ultimate goal of VAR-VIC incorporation is to update model initial state conditions at the forecasting timestep, using all available observations (i.e. both observed forcings and streamflow observations). In this study, based on the available computational resources we considered a simplified version of the general VAR approach in order to implement the variational data assimilation process [Le Dimet and Talagrand, 1986; Liu and Gupta, 2007] where the ultimate objective is to minimize the cost function of observations solely \( J_o \) within an assimilation window. Here, soil moisture is considered as the state variable to be updated in VIC state file for each DA cycle. The overall number of soil moisture elements containing in the VIC state file is 804, which is the product of 268 sub-grid vegetation/land covers (over 40 grid cells) and 3 soil layers. Furthermore, in order to decrease the number of decision variables in the optimization problem, DA is simplified into a 1D problem by considering a constant multiplier \( k \) to adjust the soil moisture elements in the model background state. Since we apply one constant multiplier to all the grid cells, so the relative variability of the model state would be preserved over the basin area. Hence minimizing the change in the background state can be neglected and the \( J \) cost function can be revised as:

\[
J(x_k) = J_o = \sum_{T_0}^{T_{-AW}} (y_t - H_t[x_k])^T R_t^{-1} (y_t - H_t[x_k])
\]

where in the above expression, \( x_k \) refers to the analysis state derived from linearly scaling the background state (i.e. \( x_k = k \times x_b | k \in [0, 2] \)), \( T_0 \) is the time of forecast, \( T_{-AW} \) is the beginning of the assimilation window, and \( H_t[x_k] \) is the simulated flow at time \( t \).
4.3. METHODOLOGY

when VIC is initialized with $x_k$ (Fig. 4.3). $R_t$ is the daily observational error computed based on 0.05% of variance of observed daily flows computed using 62 years (1949-2010) of records. For a given data assimilation time frame, model background state $x_b$ at $T_{-AW}$ can be derived by implementing VIC model with observed meteorological forcing along with un-updated initial conditions $X^{-}$ from past (i.e. "open loop" simulations or "control run"). Next, based on a brute force search method, VIC model is executed with different $x_k$ state conditions to create an ensemble of streamflow predictions. Then the cost function $J$ is computed based on available streamflow observations, resulting in the determination of optimal solution $x_k^*$. Finally, the optimal state conditions are used to initialize the model at the beginning of assimilation window and drop the updated state conditions $X^+$ at the forecasting timestep $t_0$.

The two main parameters in the explained DA framework are 1) the length of assimilation window $AW$ and 2) the state update frequency $UF$ (also known as DA cycles). In this study, a retrospective application of the explained DA were performed throughout the entire study period (1991-2010) in different scenarios depending on U.F. and A.W. lengths. For each scenario, $UF$ and $AW$ are selected as 7days, 10days, 15days, 20days, 30days, or 60days. Note that the coarser resolution $UF$ has, the less DA applications are involved in the model simulation. The results of enhancements in streamflow prediction for each scenario are presented in section 4.4.

4.3.4 Streamflow Forecasting Approach

Generally, developing skillful monthly-to-seasonal streamflow forecasts depends on two major contributors: 1) accurate estimation of initial moisture conditions of the basin and 2)
the skill of monthly-to-seasonal climatic forecasts. In this study, with the purpose of issuing
deterministic streamflow forecasts, VIC model is forced with spatially and temporally
downscaled precipitation forecasts from ECHAM4.5 GCM (more details in section 4.2.4)
along with daily climatology of temperature and wind speed forcings. The ECHAM4.5
forecasts used in this study are monthly updated 1-month ahead forecasts, thus for
consistency reasons, UF parameter in the DA application in streamflow forecasting (section
4.4.3) is selected as 30days. To compare the performance of 1-month ahead forecasts with
another candidate forecast, we considered a probabilistic streamflow forecasting approach,
known as Ensemble Streamflow Prediction (ESP) [Day, 1985; Wood and Lettenmaier,
2008], that uses climatological forcings. Under the ESP scheme, VIC model is forced with
an ensemble of meteorological forcings from the past historical records (climatology). We
used 42 years of data (1949-1990) prior to the study time frame (1991-2010) in order
to generate an ensemble streamflow forecast with 42 ensemble members. For instance,
monthly streamflow forecast for February 1993 is obtained by feeding VIC model with
observed February meteorological forcings collected from 1949 till 1990. ESP is a traditional
forecasting approach in National Weather Service (NWS), operationally used by European
Center for Medium-Range Weather Forecasts (ECMWF), since it allows to estimate the
uncertainty in a forecast under climatological forcings. Under both forecasting approaches
in this study, VIC is initialized either with un-updated initial conditions from open loop
simulation or with the updated conditions obtained from VAR-DA applications with
different $UF$ and $AW$ parameter sets.
4.3.5 Evaluation Metrics

The performance of VAR-DA and open loop simulations/forecasts are evaluated using different measures such as: Nash-Sutcliffe Efficiency (NSE) score [Nash and Sutcliffe, 1970], Kling-Gupta Efficiency (KGE) score [Gupta et al., 2009], Relative Root Mean Square Error (R-RMSE), Spearman’s Rank Correlation Coefficient, and % Bias (Table 4.1). For probabilistic ESP forecasts, we used Brier Score and its decomposed components [Brier, 1950; Weigel et al., 2007]. NSE is derived by taking average of the squared differences between the predicted and observed flows normalized by the variance of the observed flows (Eqn. 4.3). NSE ranges between 1 (i.e. perfect fit) and $-\infty$, whereas $NSE < 0$ means that the model is no better than the mean of observations as a predictor. Since NSE is quantified using squared differences, model performance is overestimated during peak flows and is underestimated during low flows. Hence, NSE is not suggested to use it as a verification metric for low-flow and high-flow predictions. Thus for extreme flow analyses, KGE metric is preferred which simultaneously accounts for correlation coefficient, mean bias, and relative variability in the predictions and observations (Eqn. 4.4). KGE also ranges between 1 to $-\infty$ with 1 denoting a perfect prediction. In addition, to better diagnose the model performance, Relative Root Mean Square Error (R-RMSE) (Eqn. 4.5), Spearman’s Rank Correlation, and % Bias metrics are also presented.

\[
NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{t}^{VIC} - Q_{t}^{obs})^2}{\sum_{t=1}^{n}(Q_{t}^{obs} - \bar{Q}^{obs})^2}
\]  

\[
KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma^{VIC}}{\sigma^{obs}} - 1\right)^2 + \left(\frac{\bar{Q}^{VIC}}{\bar{Q}^{obs}} - 1\right)^2}
\]
4.3. METHODOLOGY

CHAPTER 4. VARIATIONAL DA IN VIC

\[
R - \text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Q_{t}^{VIC} - Q_{t}^{obs})^2} \quad (4.5)
\]

These verification metrics are also used in order to compute the overall performance of the ESP scheme based on the mean ensemble. However, the added value of conducting the ESP scheme is that it allows to quantify the forecasting skill for specific flow categories such as Below-Normal (BN) or Above-Normal (AN) flows. In this regard, we employed Brier Score (BS) [Brier, 1950] (Eqn. 4.6) as the verification metric and its decomposed components such as Reliability (Eqn. 4.7) and Resolution (Eqn. 4.8) which provide a better insight into different aspects of forecast quality. Basically, the reliability metric represents the bias in the probabilistic forecasts and the resolution metric indicates how much climatic uncertainty is explained in the forecast (i.e. can be seen as the amount of information in the forecast). The first step to measure BS is to compute probabilistic forecast \( p_t \) for a predefined category from ESP runs. For example for BN events, the probabilistic forecast is quantified as the number of ensemble members with the forecasted streamflow quantity below the 33\(^{rd}\) percentile of streamflow climatology (i.e. \( Q < Q_{33}^{obs} \)) of the basin. Then BS metrics can be measured based on following equations:

\[
BS = \frac{1}{N} \sum_{t=1}^{N} (p_t - y_t)^2 \quad (4.6)
\]

\[
REL = \sum_{d=1}^{D} \frac{n_d}{N} \left( \frac{O_d}{n_d} - P_d \right)^2 \quad (4.7)
\]

\[
RES = \sum_{d=1}^{D} \frac{n_d}{N} \left( \frac{O_d}{n_d} - \bar{\delta} \right)^2 \quad (4.8)
\]
where $N$ is the total number of forecasts, $y_t$ is the observation transformed to a binary scale (i.e. $y_t$ is 1 if the event happened and 0 otherwise). $D$ denotes the number of distinct forecast probabilities issued (i.e. $p_t \in \{P_1, ..., P_D\}$), $n_d$ is the number of forecasts lies in the $d^{th}$ category, $o_d$ is the total number of observed events when the $d^{th}$ forecast was issued, and $\overline{o}$ is the climatological event frequency.

4.4 Results

4.4.1 VIC Model Performance

Daily flow predictions for the Tar River basin were obtained by running the VIC and routing models in an open loop scheme during the period 1988-2010. The first three years of simulations (i.e. 1988-2010) were ignored as the model spin-up period and the rest twenty years are evaluated at a monthly time step based on the observed flows at the basin outlet. Table 4.1 summarizes the performance of the VIC model using mean monthly flows for a given month along with the overall skill during 1991-2010. The overall NSE during validation period is quantified as 0.72, the monthly NSE and R-RMSE values suggest that the model is well calibrated for spring and summer seasons while it performs poorly during November-January and August months. Although all the monthly correlation coefficients are statistically significant (i.e. greater than $1.96/\sqrt{n - 3}$, where $n = 20$ years for a given month) which suggest a strong positive relation between the modeled and observed flows, but R-RMSE and %BIAS values indicate that the error in predictions are particularly high during Summer and Fall seasons. Further, it is noticeable that there is always a positive bias in the simulations inferring that the model tends to overestimate...
Table 4.1. VIC model performance summary over the validation period

<table>
<thead>
<tr>
<th>Month</th>
<th>NSE</th>
<th>KGE</th>
<th>Rank</th>
<th>Corr</th>
<th>R-RMSE</th>
<th>%BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>-0.12</td>
<td>0.40</td>
<td>0.84</td>
<td>0.57</td>
<td>41.4</td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td>0.47</td>
<td>0.69</td>
<td>0.88</td>
<td>0.37</td>
<td>29.2</td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td>0.74</td>
<td>0.77</td>
<td>0.96</td>
<td>0.28</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td>0.87</td>
<td>0.81</td>
<td>0.97</td>
<td>0.22</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>0.87</td>
<td>0.85</td>
<td>0.90</td>
<td>0.31</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>0.86</td>
<td>0.85</td>
<td>0.89</td>
<td>0.34</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>Jul</td>
<td>0.42</td>
<td>0.42</td>
<td>0.84</td>
<td>0.75</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td>0.15</td>
<td>0.26</td>
<td>0.86</td>
<td>0.90</td>
<td>62.7</td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td>0.91</td>
<td>0.49</td>
<td>0.86</td>
<td>0.67</td>
<td>50.5</td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td>0.50</td>
<td>0.23</td>
<td>0.72</td>
<td>1.06</td>
<td>71.5</td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td>0.35</td>
<td>0.38</td>
<td>0.83</td>
<td>0.82</td>
<td>54.9</td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td>-0.20</td>
<td>0.34</td>
<td>0.82</td>
<td>0.74</td>
<td>54.3</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.72</td>
<td>0.62</td>
<td>0.91</td>
<td>0.52</td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>

the flows.

### 4.4.2 Role of VAR-DA in Streamflow Simulation

As described in section 4.3.3, there are two main parameters in our VAR-DA framework that influence the accuracy of streamflow simulations and forecasts. In order to assess the sensitivity of the DA framework to these parameters, a set of experiments are performed, where in each scenario VIC model state conditions are updated in different frequencies (Update Frequency, UF) by assimilating a range of past available streamflow observations (Assimilation Window AW). Then the skill of data assimilated streamflow simulations are quantified based on the observed flows and are compared to the skill of the open loop simulation (i.e. simulation with un-updated initial conditions). Figure 4.4 shows a sample time series from a VAR-DA experiment with $UF = 15\text{days}$ and $AW = 10\text{days}$ along with
4.4. RESULTS

Figure 4.4. Timeseries of monthly streamflow observations along with streamflow simulations from Open Loop (OL) scheme and Data Assimilation (DA) experiment using $UF = 15\text{days}$ and $AW = 10\text{days}$.

the open loop simulation and the streamflow observations. Here, the term "simulation" (sometimes known as "perfect forecast") indicates that the hydrologic model is fed with all observed meteorological forcing data.

Figure 4.5 displays the difference between the KGE score from DA simulations with different settings and the OL simulation. The statistics shown in this figure are computed either using daily (first row) or monthly (second row) flows. Further, the analyses of DA impact on low flows (flows less than 10$^{th}$ percentile of the climatology) and high flows (flows greater than 90$^{th}$ percentile of the climatology) are included. In all experiments, VAR-DA predictions are enhanced compared to the open loop simulation specifically for shorter assimilation windows and update frequencies between 10 days and 30 days. On average, KGE of the predictions during the entire 20 years of study time frame is increased by 0.3 after implementing DA framework, which results in a KGE score greater
4.4. RESULTS

Figure 4.5. $\Delta KGE (KGE_{DA} - KGE_{OL})$ Improvement in Streamflow Simulations after applying VAR data assimilation for different AW and UF lengths using all flows (gray-scaled plots), only low flow analysis (i.e. observed streamflow less than 10th percentile $Q < Q_{p=0.1}$) (orange plots) and only high flow analysis (i.e. observed streamflow more than 90th percentile $Q > Q_{p=0.9}$) (blue plots) during the period 1991-2010. The first row shows the statistics quantified using daily flows and the second row plots are generated using average monthly flows. The skill of open-loop simulations for each of the analyses are shown in the plot titles in terms of KGE.

than 0.8 for both daily and monthly flow analyses. The model skill in predicting low flows are particularly lower than normal, possibly due to the bias in model calibration. However, the strongest improvement from DA is found in low flow predictions in terms of KGE. NSE metric is not used to determine the effect of VAR-DA in low/high flow analyses, since it is sensitive to extreme flow conditions as explained in section 4.3.5.
4.4. RESULTS

On the other hand, minimal improvements are found when using a long assimilation window. For example, in the case of $AW = 60\text{days}$ the initial conditions of the simulations are updated by incorporating past 2 months of daily streamflow observations which is beyond the residence time of water in the soil moisture zone. A longer $AW$ also results in a smoother changes in the scaling factor $K$ since it behaves as a moving average window with long bandwidth, where there is an overlay between the past and future iterations so the minimum variation in the predictions is expected. A detailed statistical analysis of the $K$ multiplier is reported in Appendix B. In addition, updating model state conditions in shorter cycles (i.e. higher frequencies) does not necessarily result in the highest improvement, although it is expected because of the greater number of DA implementations. It is inferred from figure 4.5 that updating initial conditions every 15-20 days results in better streamflow predictions due to correction of initial conditions.

4.4.3 Role of VAR-DA in Streamflow Forecasting

For a given month, streamflow forecast is developed either using 1-month ahead precipitation forecast from ECHAM4.5 GCM or employing the ESP approach that uses precipitation climatology (more details in section 4.3.4). For each of the forecasting approaches, VIC model is initialized either with un-updated state conditions from the open loop simulation or with the updated state conditions from the DA simulation schemes prior to forecast issue time. Figure 4.6 shows the timeseries of monthly streamflow forecasts developed under these forecasting schemes. The effect of VAR-DA application in streamflow forecasting is evident in this figure where basically DA tries to keep the model on the observation track. For instance during Summer and Fall 1993, the forecasts using the
4.4. RESULTS

CHAPTER 4. VARIATIONAL DA IN VIC

Figure 4.6. Timeseries of monthly forecasted streamflow acquired from both deterministic and probabilistic forecasting approaches. Prior to monthly forecasting VIC model is initialized either based on Open Loop (OL) states (first row) or through Data Assimilated (DA) initial conditions (second row) using $AW = 15$ days.

Open loop initial states (top figure) start to drift apart from the observations, while using monthly updated initial conditions (bottom figure) minimizes this divergence. Further, under the ESP approach coupled with DA, it can be seen that the forecasted ensemble mean is shifted towards the observations and the forecasted ensemble spread is narrowed indicating the uncertainty reduction in probabilistic forecasting. In general, DA is able to improve the skill in monthly streamflow forecasting by $+0.08$ in terms of NSE and $+0.02$ in terms of KGE. These quantities are the averaged improvements computed from DA applications with different $AW$ lengths. A detailed analysis of the results as well as the
skill of open loop simulation is shown in Appendix B (Table B.1). In addition table B.2 provides the statistical analysis of mean flows (ensemble mean, EM) as one of the output information from the ESP forecast. Generally, the ensemble mean specifies the element of forecast that can be predicted with confidence (i.e. most likely outcome), however it has a poor performance in capturing the risk of extreme streamflow events. Tables B.1 and B.2 suggest that DA is always effective in enhancing streamflow forecasts. Also, a comparison between these tables reveals that the streamflow forecasts using ECHAM4.5 precipitation forecasts perform better than the ensemble mean of ESP runs, thus utilizing climate forecasts is more informative than using climatological forcings. Table B.3 provides a detailed analysis of skill enhancement in probabilistic forecasting for Below Normal (BN, below 33rd percentile climatology of respective monthly flows) and Above Normal (AN, above 67th percentile climatology of respective monthly flows) streamflow events. Data assimilation is able to reduce the brier score approximately by 0.05 in BN and 0.02 in AN months where the brier score of open loop forecasts for both categories is about 0.18. This improved brier score is due to a slight improvement in resolution, but mostly related to the enhancement of the reliability score which can be interpreted as the bias reduction in the probabilistic forecasts. In contrast to simulation scheme, under forecasting experiment it is found that selecting a long assimilation window such as 15-20 days results in the best improvement in the skill of streamflow forecasts, based on all the verification metrics (Tables B.1 and B.2). However in the case of probabilistic forecasting, assimilating a shorter window (e.g. 7-15 days) is more successful. Monthly statistical analysis (results are not shown) suggest that DA improves forecasting skill mostly during Fall and Winter seasons compared to spring and summer seasons. This is not completely surprising since
the skill of streamflow forecasting over Southeast is particularly governed by soil moisture conditions rather than model forcings during wet seasons [Mahanama et al., 2012; Sinha and Sankarasubramanian, 2013].

The magnitude of improvements measured in the VAR-DA simulation schemes (figure 4.5) is much larger than the improvements in forecasting scheme. Under the simulation scheme, since VIC model is fed with the observed meteorological forcings, the role of forcing error is minimized and thus the skill deference between OL and DA experiments is totally inherited from the updated initial conditions. However in the forecasting scheme, due to the fact that the imprecision of forcings dominates the forecasting skill, thus the DA improvements are found smaller and almost similar across different AW lengths.

4.4.4 Role of VAR-DA Over Different Time Scales

As mentioned before, VIC model products are generated in daily time steps. In order to examine the effect of data assimilation at daily time steps (when the model is initialized and executed at the beginning of each month) the $KGE$ statistic is computed based on the daily flows under different lead times. For example, to examine the role of DA for a 7-day ahead simulation/forecasting, the flow predictions from the 1st day of a month till the 7th day are pooled together over all the months during validation period 1991-2010 and $KGE$ is computed by comparing with the respective 7-day observations. Similarly, for a monthly prediction (31-day ahead) all the data points in the timeseries are included in score computation and analyses.

Figure 4.7 shows the difference between $KGE$ scores from OL and DA experiments under different lead times where the last data point in each curve shows the average
improvement from monthly predictions. Note that there is no temporal overlay in the forecasting steps while there is an overlay in the analyzed timeseries for each lead time. Figure 4.7 also illustrates the contribution of updated model initial conditions over model time steps during a monthly simulation and forecasting. It is found that the role of DA is positive in short-range lead times, after 7-8 days from the forecast issue time. The skill of streamflow simulations stay same with no significant improvements beyond 7-8 days, while the skill of streamflow forecasting in both OL and DA experiments start to decline which is primarily due to the lack of skill in precipitation forecasts. In fact, the consistency of curves under the simulation schemes convey that the skill of OL and DA predictions remain almost same after 7-8 days and the skill under simulation is higher due to observed forcings. On the other hand, under the forecasting schemes, the negative slope after 7-8 days reflects the domination of forcings errors, which starts canceling out the positive effect of DA, however $\Delta KGE$ still remains positive even for long analysis windows, indicating the positive effect of DA.

The overall lag in DA impact to reach its peak in figure 4.7 can be interpreted as the lag in the updated soil moisture to show its effect at the basin outlet location. It is expected to have earlier peak times for smaller basins. Also this analysis suggests that for a short-term (medium-term) forecasting, it is better to use shorter (longer) assimilation windows. This analysis is conducted only over the simulation schemes with $UF = 1\text{month}$ since in the other scenarios with shorter $UF$ lengths model state updating happens at irregular days of months during the 20-year period of our study, thus the DA improvement plot with respect to different daily lead times in a month could not be interpreted.
4.5. DISCUSSION

Most studies in the hydrometeorologic literature focus on the Data Assimilation (DA) potential using observed forcings, with limited/no studies in understanding the role of DA impact in land surface models in the context of real-time streamflow forecasting. Further, most DA efforts have been focused on demonstrating the potential of assimilating remotely sensed data, with few studies utilizing ground observations such as streamflow records for improving land-surface model states. The goal of the work presented here is mainly to describe the application of the variational (VAR) data assimilation in which gauge-measured streamflow observations are utilized to update states in VIC distributed hydrologic model. Consequently, the ability of VAR data assimilation in improving VIC model soil moisture initial state is explored by quantifying the improvements in streamflow simulation and forecasting over a 20-year period (1991-2010) for the Tar river basin in NC.

Figure 4.7. The magnitude of the model skill improvements due to data assimilation over daily time steps of a monthly prediction. Each color represents the DA application with a unique assimilation window length $AW$. 
Two key parameters in the DA framework are defined for analysis: Update Frequency \((UF)\) and Assimilation Window \((AW)\), with the former determining the period between each DA application and the latter expresses the length of past daily observations to incorporate in the DA framework. The role of DA in improving the model simulation (i.e. using observed forcings to implement the model) and the sensitivity of DA to \(UF\) and \(AW\) are then assessed by conducting 36 experiments based on different combinations of VAR-DA parameters selected from a discrete set of intervals ranges from 7 days to 2 months.

Comparison of streamflow predictions between open loop (i.e. control run) and DA coupled simulations suggest that the DA application is always successful in enhancing the streamflow modeling. The maximum potential of DA is achieved when a shorter assimilation window is selected with \(UF\) around 20 days for the selected river basin. A shorter assimilation window is more effective since it mainly considers most recent basin observations and excludes the information beyond the basin’s memory characteristics. However, the optimal \(UF\) length is expected to be dependent on basin characteristics - soil type, hydraulic conductivity, and drainage area - so it is expected to be different for each river basin. Furthermore, the role of DA in operational monthly forecasting is explored by employing two forecasting approaches. Deterministic monthly streamflow forecasts are developed by feeding the VIC model with 1-month ahead spatially downscaled and temporally disaggregated precipitation forecasts from ECHAM4.5 GCM. Probabilistic streamflow forecasts (ESP) are developed using the climatological forcings, based on the historical precipitation observations from 1949 to 1990. Investigating the performance of DA under the deterministic forecasting scheme indicate that DA improves the skill in
monthly forecasting by increasing NSE by +0.08 and KGE by +0.02 on average. It is also noticed that there is limited improvement in the skill among different assimilation window lengths since the skill of climatic forcing controls the skill of streamflow forecasts. Considering that fact, we also quantified the persistence of the impact of data assimilated initial conditions in improving daily streamflow predictions. It is found that the effectiveness of updated initial conditions lasts for 7-8 days ahead from the forecast issue time.

Rendering a comprehensive VAR optimization problem for a distributed hydrologic model is very difficult and computationally intensive. Hence, based on our available computational resources, the complexity of DA problem in this study is simplified to a 1-D problem by considering only one decision variable which is basically defined as a multiplier factor that adjusts the model background soil moisture contents in different grid points and different soil layers over the selected basin during each DA implementation. Even with a simplified version of the problem, solving the iterative optimization problem is not easy and most of the studies often use an numerical approximation algorithm or adjoint model technique to find an optimal solution. This is mainly due to the high nonlinearity of the hydrologic systems such as the complicated interactions between the basin initial conditions and the dynamics of the basin outlet. Furthermore, accurate estimation of the model background error covariance is challenging and usually it is obtained through an ensemble approach by executing model with a set of perturbed forcings. In our study, since the analysis state is derived by uniform scaling of the background state, the spatial covariance of soil moisture state is preserved. In this fashion, the cost function estimate in VAR algorithm can be simplified by considering the observational cost function and disregarding the objective function related to minimizing the background cost function.
since it prevents the background state to change.

It is shown that data assimilation can be significantly beneficial in improving the hydrologic predictions, yet its appropriate application requires a certain amount of understanding on its behavior to gain maximum potential. Data assimilation is sometimes criticized for violating the water mass balance assumption in the hydrologic model, however its potential benefit in improving the model products cannot be ignored. Data assimilation applications can be used not only for better state estimation but also to enhance parameter estimation in a combined approach. It is more rational to quantify the effect of DA in the absence of model bias, which is another advantage of data assimilation application. With the intention of applying bias correction as well as quantifying the sole role of DA in hydrologic forecasting, a recursive bias estimation should be coupled into the DA framework at each iteration resulting in a two-stage estimation algorithm (known as "separate bias estimation" in DA literature)[Friedland, 1969; Dee and Da Silva, 1998] and hence more computational cost. In addition, a sequential data assimilation has the potential to capture the changes in hydrologic flow patterns, hence it acts similar to a dynamic bias correction.

Majority of hydrologic DA studies are focused on short-term hydrologic predictions ranges from hourly to maximum few days, but here we tried to understand the DA impact in monthly-to-seasonal forecasting which can be used to support water management. Further, in real-time operational forecasting, the technical complexity of DA problem due to nonlinearity of hydrologic process and high dimensionality of the model initial conditions should be considered properly. Our study shows that DA is overall beneficial in forecasting low and high flows using the proposed simplified version of VAR, but DA also
degraded the model predictions in certain months. Thus, it is also important to assess and quantify the negative effects of DA and subsequent shortcomings in practice especially for specific cases such as extreme events. Finally, significant advances in hydrologic DA should be well communicated among researchers and forecasting centers in order to reach a pathway for transition of the hydrologic DA research into operational forecasting applications.

**Acknowledgement**

This work was supported by National Science Foundation grants CBET-0954405, CBET-1204368, and CCF-1442909. The authors would also like to thank the University Corporation for Atmospheric Research (UCAR) Advanced Studies Program (ASP) as well as National Center for Atmospheric Research (NCAR) for providing high-performance computing support for this project [Yellowstone, 2012].
CONCLUSION AND SCOPE FOR FUTURE WORK

Various strategies for uncertainty reduction in streamflow forecasts were investigated in this dissertation. In the first chapter, the usefulness of climatic probabilistic forecasts in forecasting monthly streamflow conditions were assessed for selected river basins across the US Sunbelt. It is then shown that Multinomial Logistic Regression (MLR) can be considered as a promising alternative for the traditional Principal Component Regression (PCR) model in order to issue probabilistic forecasts for categorical events such as below normal and above normal streamflow conditions. Aside from the MLR and PCR comparison, other methods such as quantile mapping, K-NN approach, predictors for MLR, and role of climate ensemble size were also tested. Analyses showed that the observed streamflow data in fact adds more information to the streamflow forecasts in comparison to the coarse-scale climatic information from GCM particularly during drier conditions. However, ensemble precipitation forecasts have the potential to improve the sub-seasonal flow predictions. This work could be extended for other river basins with different characteristics and under specific hydroclimatic conditions. Also, soil moisture or snow cover remotely sensed data could be considered as external predictors which basically contribute in the accuracy of initial conditions.

The second chapter provides a quality assessment of the seasonal soil moisture forecasts based on the soil moisture observations collected by the recently launched SMAP satellite. The forecasts are developed by feeding the NOAH3.2 Land Surface Model (LSM) with 1-3 month ahead precipitation forecasts from ECHAM4.5 GCM along with NLDAS2 meteorological forcings. It is found that the forecasted SM captures the spatial variability
of SM observations from SMAP. The limitation of SMAP data is that it is available only for last few years. For future analyses, it is suggested to develop seasonal soil moisture hindcasts from 1979 until now, since NLDAS2 data is also available for the same period, thus the SM products can be compared with the European Space Agency’s (ESA) Climate Change Initiative (CCI) soil moisture dataset available for the same period. In addition, the uncertainty in the satellite observations should be considered in such comparisons.

In the third chapter, ‘abcd’ lumped monthly watershed model is used in simulating and forecasting streamflows for almost 900 virgin river basins across the CONUS. It is found that conceptual models are capable of simulating flows that explain the interannual variability of the observed flows. In addition, Ensemble Kalman Filter (EnKF) data assimilation (DA) is applied in order to reduce the errors in the model state variables and the potential gain due to DA is analyzed. However, for basins located in the snow-dominated areas EnKF did not improve the model skill. This degradation is potentially due to the limitations of snow ablation model (based on a temperature index method) included in the ‘abcd’ model which seems to have bias in estimating snow ablation due to sublimation. In this regard, this work could be extended by employing a more complex method of snow accumulation/melt estimation. In addition, assimilation of snow observations over these basins is suggested in order to better reduce the errors in initial conditions.

And finally the last problem explores the usability of variational (VAR) data assimilation in VIC land surface model. The motivation of using VAR data assimilation over EnKF, is to incorporate a window of recent observations into the assimilation scheme, while EnKF method only considers the most recent information. In addition, implementing
an ensemble DA method into a distributed LSM would increase the complexity of the problem to a great extent, and addressing such problems requires great computational resources. Furthermore, last chapter is focused on operational hydrologic forecasting by utilizing two different forecasting approaches. Results indicate that VAR data assimilation is able to significantly enhance the performance of streamflow predictions. Moreover, it is identified that for a study case over Tar River basin in NC, the effectiveness of updated initial conditions can not be seen beyond 7-8 days in forecasting daily flows. Data assimilation is not only used to improve the model states, but it also has the potential to gain a better system identification and parameter estimation as well. In addition, the application of data assimilation techniques in streamflow forecasting can be a potential future research, and it can be achieved by investigating a combination of different methods of modeling and data assimilation techniques based on increasing number of observations available in estimating land-surface states. These provide exciting opportunities in reducing uncertainty in streamflow prediction through various strategies.


Ajami, N. K., H. Gupta, T. Wagener, and S. Sorooshian, Calibration of a semi-distributed


Andreadis, K. M., and D. P. Lettenmaier, Assimilating remotely sensed snow observations into a macroscale hydrology model, Advances in water resources, 29(6), 872–886, 2006.


Arumugam, S., R. Boyles, A. Mazrooei, and H. Singh, Experimental reservoir storage forecasts utilizing climate-information based streamflow forecasts, 2015.


Li, S., and L. Goddard, Retrospective forecasts with echam4. 5 agcm iri technical report. 05–02, 2005.

Li, S., L. Goddard, and D. G. DeWitt, Predictive skill of agcm seasonal climate forecasts subject to different sst prediction methodologies, *Journal of Climate*, 21(10), 2169–2186, 2008.


Mashhadi, A., M. E. Shafiee, and E. Z. Berglund, Agent-based modeling to simulate the

Mason, S. J., and M. K. Tippett, Climate predictability tool version 15.3. 9, 2016.


Mazrooei, A., Decomposition of sources of errors in seasonal streamflow forecasting over the us sunbelt using multiple land surface models., 2014a.


Pappenberger, F., and K. J. Beven, Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water resources research, 42*(5), 2006.


Regonda, S. K., B. Rajagopalan, and M. Clark, A new method to produce categorical streamflow forecasts, *Water resources research*, 42(9), 2006.


Roeckner, E., et al., Simulation of the present-day climate with the {ECHAM-3} model: impact of model physics and resolution, 1992.


Sinha, T., and A. Sankarasubramanian, Role of climate forecasts and initial land-surface conditions in developing operational streamflow and soil moisture forecasts in a rainfall-


APPENDICES
Common Data Assimilation Algorithms

Some of the common used data assimilation techniques are briefly explained in this section according to Kalnay [2003].

Direct Insertion

Direct insertion is one of the oldest and simplest approaches to data assimilation. In this method, the forecasted model states are directly replaced with the observations. This approach assumes that the model has no useful information and the observations are perfectly true, which both disregards important information provided by the model and preserves observation errors [Houser et al., 1998].

Statistical Correction

This method is a derivative of the direct insertion method which tries to adjust the mean and variance of the model states with regard to the observations. This approach assumes
that the model structure is correct but contains a bias. First the model simulations are scaled by the ratio of the observation standard deviation to simulated standard deviation. Then the bias in scaled estimations are corrected by the difference between the means of the simulations and observations [Houser et al., 1998].

**Optimal Interpolation**

The Optimal Interpolation approach (OI), sometimes called as statistical interpolation, approximates the optimal gain matrix K by using an approximated (fixed) state covariance matrix for all the time steps, and is often given by prescribed variances and correlation functions calculated only based on the distance.

**3D Variational**

This approach directly solves the iterative optimization problem in order to minimize error between the forecast model state and the observations over space for each time step [Parrish and Derber, 1992]. 3D-var has the same assumption for the state covariance matrix as in the optimal interpolation is typically used. The solution of 3D-var gives an analysis which is very similar to direct insertion method.

**Kalman Filter**

All of the Kalman filter (KF) data assimilation methods calculate the gain matrix K by directly forecasting the states covariance matrix. In the traditional KF approach the K matrix is identified by applying the standard error propagation on the linearized state model. The difference between Kalman filter and extended Kalman filter is that the
forecast model is linearized using a Taylor’s series expansion while the same forecast and update equations are used in both methods. A further approach to estimate the state covariance matrix is the ensemble Kalman filter (EnKF) where the covariances are calculated from an ensemble of state forecasts using a Monte Carlo approach rather than a single discrete forecast of covariances [Evensen, 2003].
Supplemental Results: Evaluation of Forecasted Precipitation and Streamflow Over Tar River Basin

Precipitation forecasts from ECHAM4.5 atmospheric General Circulation Model (GCM) were collected from the International Research Institute of Climate and Society (IRI) data library [Li and Goddard, 2005] in order to feed VIC hydrologic model and develop deterministic streamflow forecasts. Climate forecasts from GCMs are typically available at monthly temporal resolution and at a large spatial scale (e.g. $2.8^\circ \times 2.8^\circ$ is the resolution of ECHAM4.5 model) while land surface models (LSMs) are mainly designed for a finer spatial scale (e.g. $1/8^\circ \times 1/8^\circ$ is the resolution of VIC model) and to iterate at shorter time steps (e.g. daily or sub-daily). Over the last decade, a majority of scientific efforts were devoted to spatio-temporal downscaling problems either using statistical models [Wood et al., 2004; Maurer and Hidalgo, 2008] or dynamical local models [Leung et al.,
2004; Wood et al., 2002; Wilby et al., 2000] in order to address this mismatch between the resolution of climate models and land surface models. In addition, Mazrooei [2014b] conducted a systematic evaluation and quantification of uncertainty (error) propagated to streamflow forecasting by spatial downscaling and temporal disaggregation approaches. This appendix will focus on the explanation of the statistical approaches used to downscale climate model outputs in this dissertation.

**Spatial Downscaling**

Monthly updated ECHAM4.5 precipitation forecasts available at 2.8° spatial scale, and daily observed precipitation data at 1/8° spatial scale were obtained over the period from 1961 to 2010. In order to spatially downscale the ECHAM4.5 GCM monthly precipitation forecasts from 2.8° to 1/8° spatial scale, a Principal Component Regression (PCR) approach is adapted. Given a specific month, for each 1/8° grid cell, four nearest 2.8° grids were selected as predictors while the observed monthly precipitation at the 1/8° grid cell was assigned as predictand. Since the predictors are correlated, we developed a PCR model to estimate the downscaled monthly precipitation by considering previous 25 years of data as training period prior to the forecasting year. For example, to obtain spatially downscaled monthly forecast for a given 1/8° cell for January 1991, previous 25 years of January observations (i.e., from 1966 to 1990) were served as training period of PCR.
Temporal Disaggregation

In order to create a daily time series of precipitation forecast over a month, for the given monthly downsampled precipitation forecasts from ECHAM4.5 GCM, the temporal disaggregation method explained in Prairie et al. [2007] is adapted. This method uses Kernel nearest neighbor approach (K-NN) to identify "K" numbers of months that have similar historical monthly conditions to the downsampled monthly forecasts based on leave-one out cross-validation. Based on the identified "K" neighbors of monthly precipitation, the disaggregation approach basically resamples the daily precipitation using Lall and Sharma kernel [Lall and Sharma, 1996], which gives the highest (lowest) weight to the historical monthly precipitation that is the closest (farther) to the forecasted monthly precipitation in terms of euclidean distance. Thus, daily precipitation forecasts are generated for each month from 1991 to 2010. It is important to note that the disaggregated precipitation at each 1/8° grid point is obtained independently to its neighbor grid points. This may result in reduced spatial correlation in the ensembles of disaggregated precipitation. Besides, this could potentially add error to the forecasted precipitation particularly during the winter season when there is a high spatial correlation across the observed daily precipitation. This issue could be improved by adding spatial probability functions in the disaggregation procedure in order to preserve the spatial correlation of daily precipitation time series between neighboring grid points.
Skill of ECHAM4.5 Precipitation Forecasts

Figure B.1 illustrates the skill of the spatially downscaled precipitation forecasts at monthly temporal scale over the period 1991-2010. Note that the significance level of correlation coefficient using 20 data points is 0.47 at 95% confidence level based on Steel et al. [1960]. The figure suggests that the precipitation forecasts from ECHAM4.5 GCM are not able to capture the variability in rainfall significantly. In addition, negative correlation coefficients quantified during summer season infer that the precipitation forecasts are not promising during this season. Thus, the imprecision in precipitation forcing is expected to be propagated later in streamflow forecasting. In this regard, improving forcing data and data assimilation applications hold a great potential for hydrologic research community.

Soil Moisture Scaling Factor $K$

Figures B.2 and B.3 show the time series and statistics of the optimal scale factor $K$ applied the soil moisture elements in VIC initial state file through data assimilation algorithm. It is found that mean of $K$ values during 20 years of simulations tends to stay below the one value indicating that DA generally tries to reduce the soil moisture contents in the model. This is mostly due to the bias in the model which tends to overestimates the flows. It is obvious that by selecting a longer length for the $UF$ parameter, the number of DA implementations during 20-year period of simulations is less, so it results in less variation of the $K$ value. Also selecting a longer $AW$ parameter results same since DA incorporates a longer timeseries of past observations so the change in $K$ between two
Figure B.1. Evaluation of Spearman’s rank correlation between downscaled monthly precipitation forecasts and observed precipitation at 1/8° grid cells over Tar River basin, NC.
DA cycles is smoother. On the other hand, when a short $UF$ and $AW$ is selected there are more oscillations in the optimal $K$ factor over time, which results in higher variance statistic.

**Skill of Streamflow Forecasting**

Following tables B.1, B.2, and B.3 provide a summary of the streamflow forecast enhancements due to data assimilation application. The tables first show the deterministic or probabilistic forecasting skills under the Open Loop (OL) scheme following with illustrating the changes found in different verification metrics after data assimilation applied. The "+" sign in the tables indicate that the verification metric is improved, for instance in terms of %BIAS it denotes model bias reduction.
Figure B.2. Timeseries of optimal scale factor $K$ for various DA experiments with different $UF$ and $AW$ parameters. Y axis of the plots are limited between 0 and 2, with the middle thick sign showing the value 1 for the $K$ multiplier ($K = 1$ denotes no change in the background model state).
Table B.1. Performance summary of monthly deterministic streamflow forecasting using ECHAM4.5 precipitation forecasts during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths AW, ”m” denotes month(s) and ”d” stands for day(s). ”+” sign indicates the improved verification metric and ”−” sign denotes degradation in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest.
Table B.2. Performance summary of monthly deterministic streamflow forecasting using ensemble mean of 42 ensemble members generated under the ESP scheme during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths AW, ”m” denotes month(s) and ”d” stands for day(s). ”+” sign indicates the improved verification metric and ”−” sign denotes degradation in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest.

<table>
<thead>
<tr>
<th>Streamflow Forecasting using Ensemble Mean (EM) from ESP method</th>
<th>OL</th>
<th>DA [AW]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2m</td>
<td>-1m</td>
</tr>
<tr>
<td>NSE</td>
<td>0.10</td>
<td>+0.09</td>
</tr>
<tr>
<td>KGE</td>
<td>0.27</td>
<td>+0.05</td>
</tr>
<tr>
<td>RANK CORR</td>
<td>0.73</td>
<td>+0.01</td>
</tr>
<tr>
<td>RRMSE</td>
<td>1.11</td>
<td>+0.06</td>
</tr>
<tr>
<td>%BIAS</td>
<td>38.48</td>
<td>+13.90</td>
</tr>
</tbody>
</table>
Table B.3. Performance summary of monthly probabilistic streamflow forecasting under the ESP scheme for Below Normal (BN) months and Above Normal (AN) months during the entire study period 1991-2010. OL column denotes the skill of forecasts initialized with un-updated state conditions from open loop simulation and DA columns represent the forecasts initialized with updated states through data assimilation with different assimilation window lengths $AW$, ”m” denotes month(s) and ”d” stands for day(s). Upward arrow indicates an improvement in the score and downward arrow shows the degradation of the metric in comparison with the OL column. Each row is colorized regardless of the other rows based on the magnitude of the improvements, where green is the greatest value and red is the lowest.

<table>
<thead>
<tr>
<th>Below Normal Flows (Q&lt;Q^{33rd} percentile)</th>
<th>DA [AW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OL</td>
<td>-2m</td>
</tr>
<tr>
<td>BS 0.186</td>
<td>0.02</td>
</tr>
<tr>
<td>REL 0.059</td>
<td>0.02</td>
</tr>
<tr>
<td>RES 0.094</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Above Normal Flows (Q&gt;Q^{67th} percentile)</th>
<th>DA [AW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OL</td>
<td>-2m</td>
</tr>
<tr>
<td>BS 0.185</td>
<td>0.01</td>
</tr>
<tr>
<td>REL 0.045</td>
<td>0.02</td>
</tr>
<tr>
<td>RES 0.065</td>
<td>0.00</td>
</tr>
</tbody>
</table>