Journal of Hydrology 519 (2014) 2967–2977

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination

Caleb M. DeChant, Hamid Moradkhani *

Department of Civil and Environmental Engineering, Portland State University, United States

ARTICLE INFO

Article history:
Available online 2 June 2014

Keywords:
Streamflow forecasting
Data assimilation
Multi-modeling
Ensemble streamflow prediction

SUMMARY

Uncertainties are an unfortunate yet inevitable part of any forecasting system. Within the context of seasonal hydrologic predictions, these uncertainties can be attributed to three causes: imperfect characterization of initial conditions, an incomplete knowledge of future climate and errors within computational models. This study proposes a method to account for all threes sources of uncertainty, providing a framework to reduce uncertainty and accurately convey persistent predictive uncertainty. In currently available forecast products, only a partial accounting of uncertainty is performed, with the focus primarily on meteorological forcing. For example, the Ensemble Streamflow Prediction (ESP) technique uses meteorological climatology to estimate total uncertainty, thus ignoring initial condition and modeling uncertainty. In order to manage all three sources of uncertainty, this study combines ESP with ensemble data assimilation, to quantify initial condition uncertainty, and Sequential Bayesian Combination, to quantify model errors. This gives a more complete description of seasonal hydrologic forecasting uncertainty. Results from this experiment suggest that the proposed method increases the reliability of probabilistic forecasts, particularly with respect to the tails of the predictive distribution.

1. Introduction

Uncertainty is pervasive throughout hydrologic forecasting. A general lack of information, and skillful modeling frameworks, leads to forecast products that do not have sufficient ability to be relied upon in an entirely deterministic manner. In the specific case of seasonal streamflow, volumetric estimates of runoff are necessary for guidance of an array of water management decisions, yet the accuracy of such estimates is often unsatisfactory (Moradkhani and Meier, 2010). To this end, it should be of high priority to ensure that estimates of forecast uncertainty are statistically reliable. Given that probabilistic estimates of volumetric streamflow are reliable, risk within a reservoir system can be more effectively managed, thus reducing the chance of both flood damages and water shortages concurrently.

Research into probabilistic methods for seasonal forecasts has developed over the past few decades. A first example is the Ensemble Streamflow Prediction (ESP) framework proposed by Twedt et al. (1977) and clarified by Day (1985). ESP works under the assumption that the primary skill in a hydrologic forecast is based on land surface conditions, and as such treats initial conditions as deterministic quantities, while leveraging climatological stochastic forcing to account for poor knowledge of future meteorological conditions. The framework itself has prompted a number of studies to improve seasonal forecasting, including utilizing information from climate indices (Najafi et al., 2012) and climate modeling products (Mo et al., 2013; Yuan and Wood, 2012). Since the literature suggests that some information about seasonal climate is available through both climate modeling and teleconnections, further studies have examined the assumption that skill is primarily derived from initial conditions (Li et al., 2009; Shukla et al., 2013; Wood and Schaake, 2008; Yossef et al., 2013). With an increasing focus on the relative skill of different aspects of seasonal forecasting, an increasing focus has been placed on determining how best to manage overall uncertainty in the modeling framework.

Interest in probabilistic forecasting is increasing within the hydro-meteorological research and operational communities (Brown et al., 2010; Demargne et al., 2013; Madadgar et al., 2012; Yuan et al., 2013), yet forecast systems rarely approach the uncertainty estimation problem holistically. Primarily, in the seasonal streamflow forecasting realm, these uncertainties arise from meteorological forcing of the model, initial land surface conditions, and model uncertainty. The forecasting community

* Corresponding author. Tel.: +1 503 725 2436.
E-mail address: hamidm@pdx.edu (H. Moradkhani).

http://dx.doi.org/10.1016/j.jhydrol.2014.05.045

has understandably focused predominantly on the uncertainties relative to future weather conditions, as these uncertainties will become dominant at most forecast lead times (Shukla et al., 2013; Yossef et al., 2013). More specifically, the forecast steadily loses sensitivity to the initial conditions over time, but the forecast will remain sensitive to forcing at all lead times. Though such methods lead to probabilistic flow estimates, these methods generally underestimate the forecasting uncertainty (Wood and Schaake, 2008; Yuan and Wood, 2012). The objective of this study is to develop and test a technique to improve the quantification of uncertainty in probabilistic forecasting of seasonal volumetric streamflow, thus reducing the problem of overconfidence. In order to overcome this overconfidence, this study proposes a movement towards treating the initial land surface states and models as probabilistic values, in addition to meteorological forcing.

Probabilistic estimation of land surface states has proven to be a challenge throughout the land surface modeling community, but great strides are being made in the field of ensemble data assimilation (DA) (Moradkhani, 2008). A number of researchers have been looking into the use of DA methods for improving land surface state prediction (Andreadis and Lettenmaier, 2006; Clark et al., 2008; De Lannoy et al., 2012; Margulis et al., 2002; Reichle et al., 2002) and examining the ability of stochastic states to estimate uncertainty reliably (DeChant and Moradkhani, 2011a; Leisenring and Moradkhani, 2010; Liu and Gupta, 2007; Moradkhani et al., 2005a,b). The extensive literature on ensemble DA within hydrologic models motivated the use of ensemble DA techniques for probabilistic initial state estimation (ESP-DA) as presented in DeChant and Moradkhani (2011b). Results from this study suggested that accounting for initial condition uncertainty in ESP improves the reliability of seasonal streamflow forecasting, but that results remain overconfident. An important issue that potentially causes this persistent overconfidence is the assumption that model uncertainty is insignificant. Thus model error must also be examined in a seasonal forecasting framework.

An increasingly popular method to account for model uncertainty is through multi-model ensembles (Bohn et al., 2010; Regonda et al., 2006). By having a diverse set of models, a forecast implicitly accounts for the errors related to each individual model. Multi-modeling via Bayesian Model Averaging (BMA) is becoming an increasingly popular technique throughout hydrologic forecasting (Ajami et al., 2007; Duan et al., 2007; Raftery et al., 2005), which has also been extended to estimate the posterior model probability sequentially in time, which is referred to as Sequential Bayesian Combination (SBC) (Hsu et al., 2009). Recently, both BMA and SBC have been shown to be completely compatible with ensemble DA (Parrish et al., 2012), leading the current study to propose the use of model averaging within the previously developed ESP-DA framework to simultaneously account for initial condition and model uncertainty. Through these advancements, it is possible to move hydrologic forecasting towards a more complete accounting of uncertainty (Liu et al., 2012). Thus the hypothesis of this study is that ESP with DA and SBC will lead to more reliable probabilistic forecasts of seasonal streamflow, in comparison to traditional ESP, and the previously examined ESP-DA methodology.

2. Methods

2.1. Study area

The study examines streamflow forecasting throughout the Upper Colorado River Basin (UCRB), defined here as the entire Colorado River Basin upstream of Lee's Ferry (see Fig. 1), which is located just downstream of Lake Powell. The UCRB is located in the southwestern US, covering portions of Wyoming, Utah, Colorado, Arizona and New Mexico. The basin drains an area of roughly 280,000 km^2, with forest covering much of the upper elevations and shrub land covering the valleys. The mean naturalized yearly flow volume at Lee's Ferry is roughly 18 billion cubic meters, providing water to 26 million people with a minimum designated annual flow from Lake Powell set at 9.3 billion cubic meters. In Fig. 1, the gauges of the three major sub-basins (Green River, Colorado Headwaters/Gunnison and San Juan) and at Lee's Ferry are identified. These four gauges are used to examine overall forecast reliability, whereas spatial aspects of forecast accuracy are analyzed over 16 smaller sub-basins.

3. Models and data

3.1. Hydrologic models

3.1.1. Variable infiltration capacity model

The VIC model is a physically-based, distributed model that solves the energy and water balance at the land surface, and spatially discretized units are generally placed on a regular grid (Gao et al., 2010; Liang et al., 1994). In order to perform model calculations, VIC requires soil information, vegetation information, elevation bands, precipitation, maximum and minimum temperature, average wind speed, humidity, and incoming shortwave and longwave radiation for each grid cell. Land surface parameters for VIC simulations were gathered from the Natural Resources Conservation Services STATSGO dataset (soil) and the University of Maryland land cover dataset (vegetation). Elevation bands were defined using the USGS National Elevation Dataset, with information from the Precipitation Regression on Independent Slopes Model (PRISM) yearly precipitation information to aid in the distribution of elevation band precipitation. Readers are referred to Section 2.5 for information about the forcing data. Simulations were performed over the entire UCRB at a spatial resolution of 0.25°, which makes 473 model grid cells. Based on the hydrologic fluxes estimated by VIC, excess water is routed to the outlet of the basin with a combination of Nash-Cascade hydrologic routing and Muskingum–Cunge hydraulic routing.

3.1.2. National Weather Service River Forecast Center Models

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

3.2. Observations

3.2.1. Passive microwave radiance

Passive Microwave (PM) brightness temperature (T_b) from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) data was used in this study to perform land surface DA. T_b was chosen for this study as it provides useful information