**Title:** NMME Precipitation and Temperature Forecasts for the Continental United States and Europe: Diagnostic Evaluation and Development of Multi Model Applications
**PI:** Gabriele Villarini; **Co-PI:** A. Allen Bradley
**Report Year:** Final Report **Grant #:** NA15OAR4310073

## Results and Accomplishments

As listed in the proposed work, the main goals of this research are to:

1) Examine the potential value and applicability of five GCMs that are part of the NMME project in forecasting monthly and seasonal precipitation and temperature over the continental United States and Europe. Particular emphasis will be placed on the evaluation of the quality of the forecasts for extreme conditions (e.g. low rainfall and/or high temperature; heavy rainfall) potentially leading to flood and drought.

2) Develop a multi-model averaging procedure to increase the forecast skill of these models. The technique assigns weights that define the likelihood of historical outcomes given the forecasts, and can facilitate applications in ensemble prediction of water resources and streamflow based on resampling of historical weather observations.

Since the beginning of the project, we have accomplished these goals, even exceeding what proposed in the first place. Here is a summary of our accomplishments:

### Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA.

We have examined the forecasting skill of eight Global Climate Models (GCMs) from the North-American Multi-Model Ensemble (NMME) project (CCSM3, CCSM4, CanCM3, CanCM4, GFDL2.1, FLORb01, GEOS5, and CFSv2) over seven major regions of the continental United States (based on the 2009 National Climate Assessment Report and modified by dividing the Great Plains Region into North and South). Monthly total precipitation and monthly reference mean temperature at 2 meters were obtained for all available lead times and ensemble members over the continental United States. To verify model skill, we used the Parameter-elevation Regression on Independent Slopes Model (PRISM) climate mapping system as reference dataset for the continental United States. PRISM's temporal and spatial resolutions are monthly and approximately 4 km.

The skill of the monthly forecasts is quantified using the mean square error skill score. This score is decomposed to assess the accuracy of the forecast in the absence of biases (potential skill) and in the presence of conditional (slope reliability) and unconditional (standardized mean error) biases.

We have summarized the forecasting skill of each model according to the initialization month of the forecast and lead time. Results indicate that the most skillful predictions occur at the shortest lead times and decline rapidly thereafter (Figure 1). Spatially, potential skill varies little, while actual model skill scores exhibit strong spatial and seasonal patterns primarily due to the unconditional biases in the models. The conditional biases vary little by model, lead time, month, or region. Overall, we have found that the skill of the ensemble mean is equal to or greater than that of any of the individual models.

We have also tested the models' ability to predict extended periods of extreme climate conducive to eight 'billion-dollar' historical flood and drought events (Figure 2). At the seasonal scale, the drought events are better forecasted than the flood events, and are predicted equally well in terms of high temperature and low precipitation. Overall, our findings provide a systematic diagnosis of the strengths and weaknesses of the eight models over a wide range of temporal and spatial scales.

These analyses are related to the first main goal. Also, it is worth highlighting that we have analyzed eight GCMs, rather than the five proposed.

## Development of a statistical/dynamical framework for seasonal streamflow forecasting in an agricultural watershed

We have developed a statistical-dynamical prediction framework providing probabilistic seasonal streamflow forecasts from low to high flows for the Raccoon River at Van Meter, a 8900-km$^2$ catchment located in central-western Iowa. Statistical model fits for each discharge quantile (from seasonal minimum to maximum; predictands) are based on observed basin-averaged total seasonal precipitation and annual row crop (corn and soybean) production acreage (predictors). Using the most recently-updated relationship between predictand and predictors, we produce forecasts from one to ten months ahead of the given season based on annual row crop acreage from the previous year (persistence forecast) and the monthly precipitation forecasts provided by dynamical predictions from eight GCMs from the NMME. Additionally, observed precipitation from the month preceding each season is used to characterize antecedent soil moisture conditions.

The skill of our forecast discharge is assessed both in deterministic and probabilistic terms for all lead times, flow quantiles, and forecast seasons. Overall, the system produces relatively skillful streamflow forecasts with no strong dependence on lead-time and good prediction of high flows (see Figure 3 for an example).

The seasonal flow forecast accuracy is notably improved by weighting the contribution of individual GCMs in a superensemble, and by the inclusion of antecedent precipitation to characterize initial conditions.

By examining different merging techniques and applying the NMME forecasts for streamflow predictions, this study addresses some of the issues listed in the second main goal of the work.

## On the Predictability of Precipitation across the United States

The PI has investigated the spatial distribution of predictability of daily precipitation across the United States. The emphasis is on determining the rate of increase in predictability with spatio-temporal averaging, by defining three predictability statistics (maximum predictability, predictive error and predictive instability) based on the nonlinear finite time Lyapunov exponent. Predictability increases monotonically with temporal averaging, while spatial averaging has minimal influence, pointing to the possible spatially invariant nature of precipitation dynamics. Modeling the precipitation dynamics at relatively coarser scales of 1°×1° and higher temporal scales of 5-10 days could markedly improve the predictability statistics.

Even though not part of the main proposed goals, the work was performed to gain further insights into the predictability of precipitation in terms of chaos theory.

## Bayesian weighting of North American Multi-Model Ensemble (NMME) outputs for improved flood and drought prediction across Europe

Temperature and precipitation forecasts are increasingly being used in weighted multi-model ensemble schemes to produce climate predictions with heightened predictive skill and reduced uncertainty. To enable such multi-model approaches, the NMME project collects and archives retrospective and real-time forecasts from participating centers at monthly to seasonal lead times (0.5 to 11.5 months). Here, we use temperature and precipitation forecasts from eight NMME General Circulation Models (CCSM3, CCSM4, CanCM3, CanCM4, CFSv2, GEOS5, GFDL2.1, and FLORb01) across four broad climatic European regions: Subarctic-Polar, Mediterranean, Temperate and Humid-Continental. We use the ENSEMBLES daily gridded observational dataset (E-OBS) as reference data to evaluate the skill of the NMME in forecasting monthly and seasonal precipitation and temperature at all lead times and for all four regions.

Three different multi-model weighting procedures are developed and compared using: 1) the simple mean of the individual model ensembles (giving equal weights to each model member); 2) the Bayesian weighted mean of the individual model ensembles; and 3) the Bayesian weighted mean of all of the model members. We build multi-model ensembles using a simple new Bayesian model averaging procedure that leverages the strength of each model at different lead times for different months. Using the relationship between hindcast model forecasts and observations from the verification, the procedure assigns weights to historical observations given the multi-model forecasts; the weights represent the likelihood of each historical outcome given the multi-model forecasts. The weighted samples of observations not only define an optimal bias-corrected multi-model ensemble forecast, they can also be used to selectively weight historical forcings as an atmospheric ensemble pre-processor method for hydrologic forecasting.

The monthly and seasonal skill of the different multi-model weighting procedures is assessed relative to the reference data by decomposing the mean square error skill score in the absence of biases (potential skill), and in the presence of conditional (slope reliability) and unconditional (standardized mean error) biases. Last, we assess the skill of these models in forecasting some of the costliest droughts and floods that have occurred across Europe in recent decades.

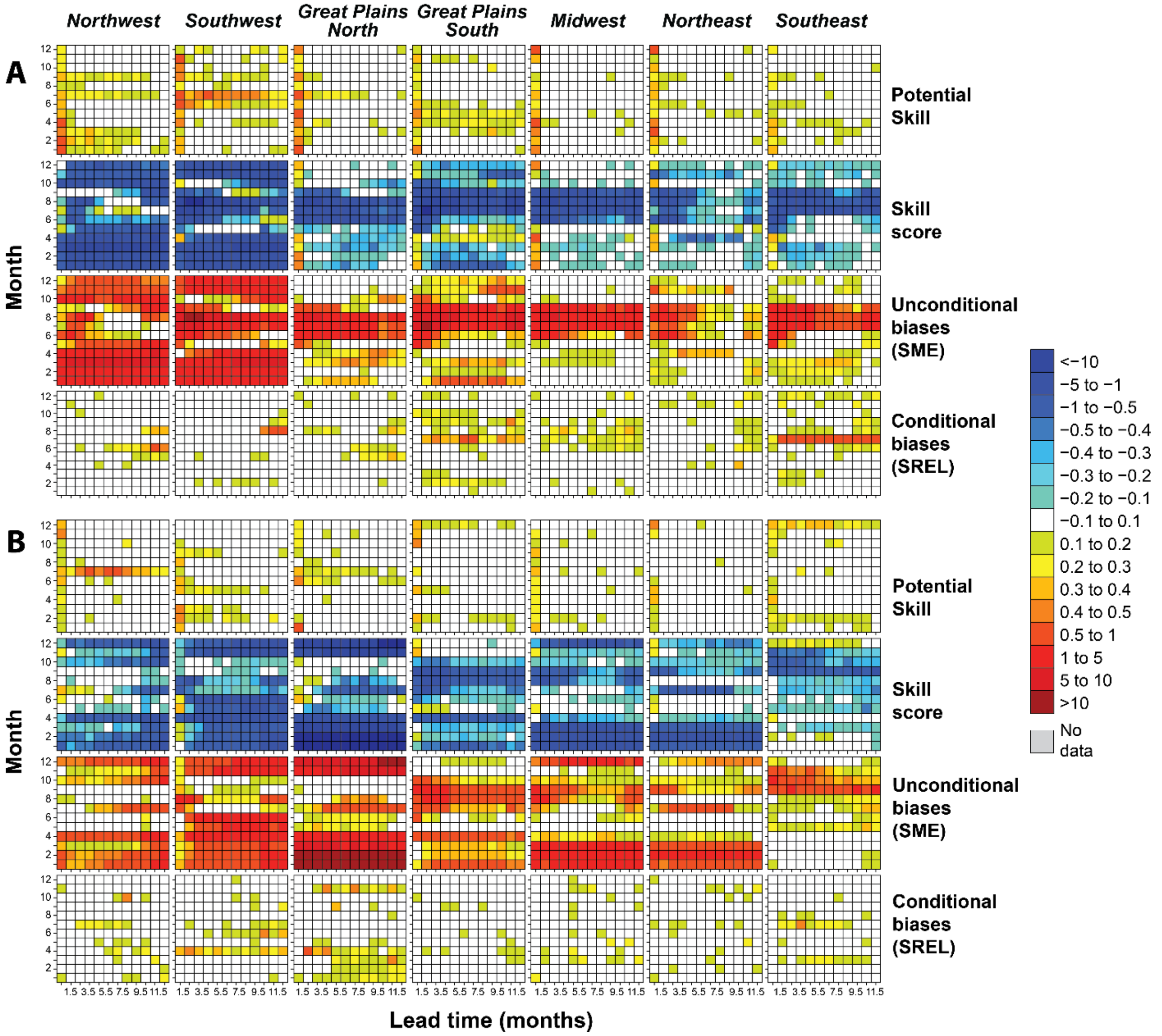

**Figure 1. Color maps indicating average skill of the eight-model ensemble mean for (A) Temperature and (B) Precipitation.** For each individual color map (1 box), x-axis indicates the lead time of the climate forecast, ranging from 0.5 to 11.5 months; y-axis indicates the month that is forecasted, ranging from 1 (January) to 12 (December). Labels at the top of the figure indicate each of the 7 regions considered in this work (Northwest, Southwest, Great Plains North, Great Plains South, Midwest, Northeast, and Southeast). Right side of the figure indicates the computed components of the ensemble's skill: Potential skill, Skill score, Unconditional biases (SME), and Conditional biases (SREL). The color scale on the right side of the figure is used for all components of the skill score, and ranges from less than -10 (blue shades) to more than 10 (red shades). From Slater et al. (2016a)

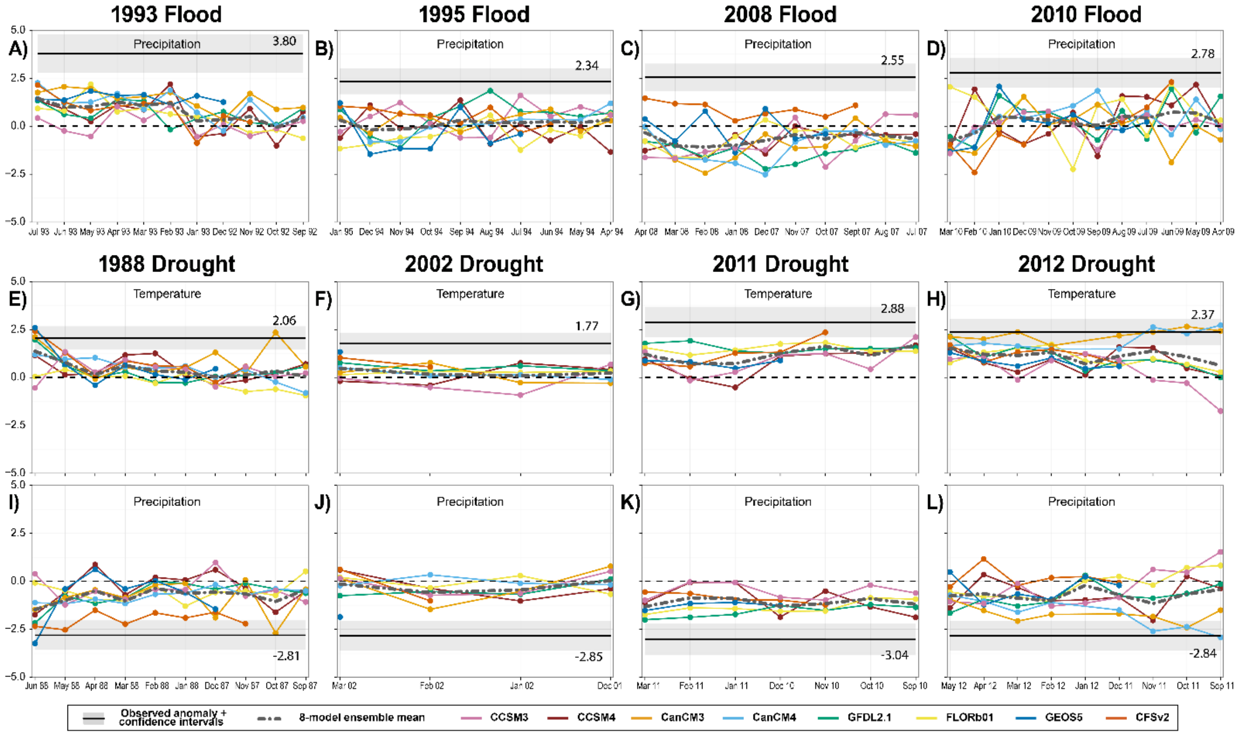

**Figure 2**. **Skill of the eight NMME models in predicting four flood and four drought events, in comparison with the observed climatology**. Flood and drought events (A-L) are identified in the title above the panels. Thick horizontal black line indicates the PRISM observed climatological anomaly, with 95% confidence intervals indicated as shaded grey rectangles in the background. NMME anomalies are indicated as colored lines. Long/short-dashed black line indicates the eight-model ensemble mean. Panels F and J: note that GEOS5 only exhibits one lead time and CFSv2 two, because the event lasted for nine months and these models only issue nine- and ten-month lead times, respectively. Panels G and K: note that the two Canadian models have data gaps in 2011, so are not included in the evaluation of the 2011 March-August drought. From Slater et al. (2016a)

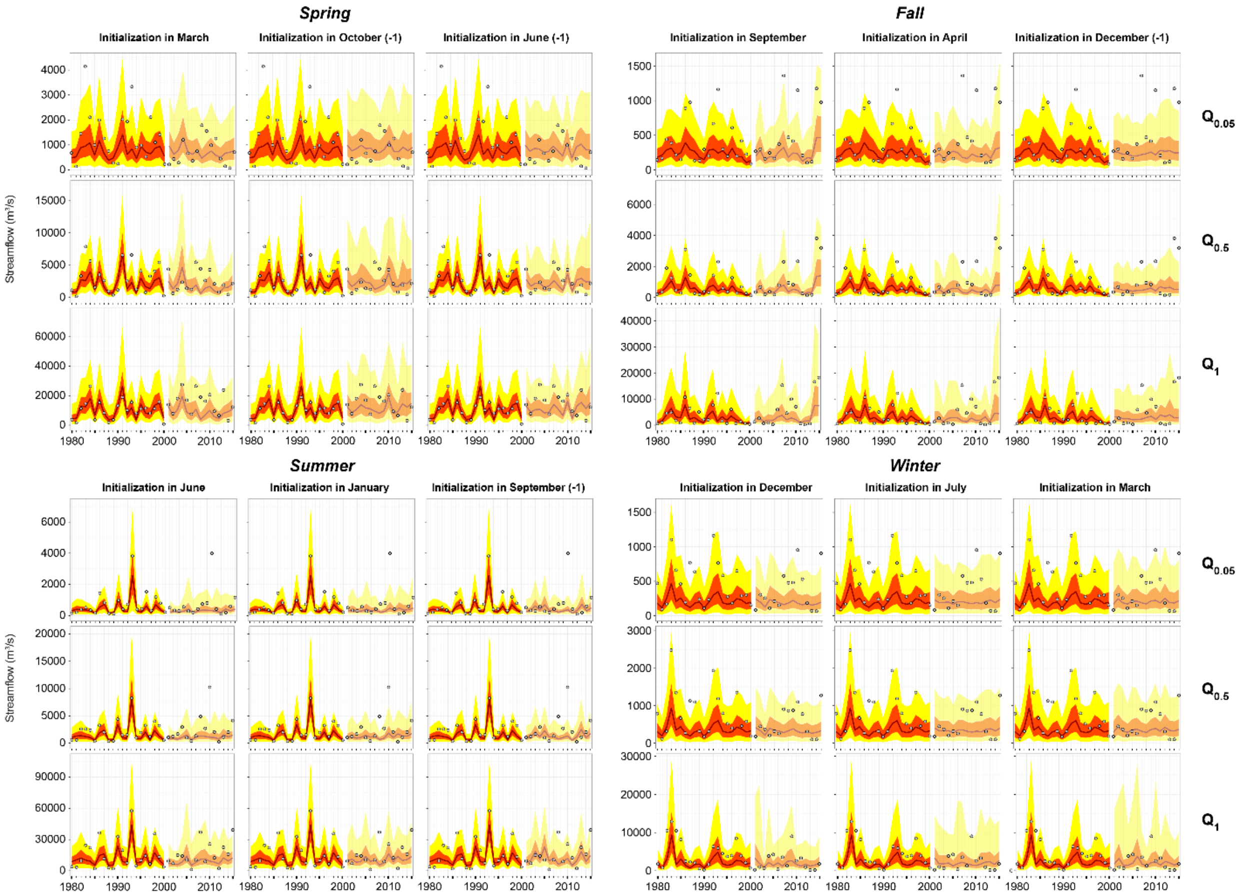

**Figure 3: Time series showing the superensemble fit (1980-2000) and forecast (2001-2015) against the observed values**. For every season, and three different initialization times (0.5, 5.5 and 9.5 months ahead of the season), five quantiles of the predicted discharge distribution are shown ($Q_{0.05}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$ and $Q_{0.95}$) for three quantiles (rows): low flow ($Q_{0.05}$), median flow ($Q_{0.50}$), and maximum seasonal flow ($Q_1$). The dark red line represents the median ($Q_{0.50}$) of the predicted distribution, the orange region the area between $Q_{0.25}$ and $Q_{0.75}$, and the yellow region the area between $Q_{0.05}$ and $Q_{0.95}$. The gray circles indicate the observed values. Predicted values are only shown for 1980-2000, even though the models have been fit to the entire period 1927-2015, in order to highlight the detail in the models. From Slater et al. (2016b).

## Highlights of Accomplishments

- We have examined the forecasting skill of eight Global Climate Models (GCMs) from the North-American Multi-Model Ensemble (NMME) project (CCSM3, CCSM4, CanCM3, CanCM4, GFDL2.1, FLORb01, GEOS5, and CFSv2) over seven major regions of the continental United States and four European regions.
- Results indicate that the most skillful predictions occur at the shortest lead times and decline rapidly thereafter. Spatially, potential skill varies little, while actual model skill scores exhibit strong spatial and seasonal patterns primarily due to the unconditional biases in the models. The conditional biases vary little by model, lead time, month, or region. Overall, we find that the skill of the ensemble mean is equal to or greater than that of any of the individual models. A much improved performance is obtained by developing multi-model ensemble averages based on Bayesian weighting.
- We have tested the models' ability to predict extended periods of extreme climate conducive to historical flood and drought events in the United States and Europe. At the seasonal scale, the drought events are better forecasted than the flood events, and are predicted equally well in terms of high temperature and low precipitation.
- We have used the precipitation forecasts from eight GCMs from the NMME to produce seasonal streamflow forecasts (from low to high flows) for an agricultural watershed in Iowa. Overall, the system produces relatively skillful streamflow forecasts with no strong dependence on lead-time and good prediction of high flows. The seasonal flow forecast accuracy is notably improved by weighting the contribution of individual GCMs in a superensemble, and by the inclusion of antecedent precipitation to characterize initial conditions.

## Publications from the Project

Slater, L.J., G. Villarini, and A.A. Bradley, 2016a: Evaluation of the skill of North-American Multi-Model Ensemble (NMME) global climate models in predicting average and extreme precipitation and temperature over the continental USA. DOI:10.1007/s00382-016-3286-1, Climate Dynamics (in press).

Slater, L.J., G. Villarini, A.A. Bradley, and G.A. Vecchi, 2016b: A statistical/dynamical framework for seasonal streamflow forecasting in an agricultural watershed. Climate Dynamics (submitted).

Slater, L.J., G. Villarini, A.A. Bradley, 2016c: Improved climate predictions across Europe using Bayesian weighting of North American Multi-Model Ensemble (NMME) forecasts (in preparation).

## PI Contact Information

Gabriele Villarini
The University of Iowa
306 C. Maxwell Stanley Hydraulics Laboratory
Iowa City, Iowa 52242
Office: (319) 384-0596
Email: gabriele-villarini@uiowa.edu