HEPEX Testbed Progress Report #1

Basins in the Southeast U.S.

Lifeng Luo (*Princeton University, Princeton, NJ*)
lluo@princeton.edu

Eric F. Wood (*Princeton University, Princeton, NJ*)
efwood@princeton.edu

Project Purpose

The purpose of the HEPEX testbed in the Southeast US is to address the following HEPEX science questions:

1) How do we generate skillful and reliable meteorological forcing for seasonal hydrologic forecasting?
2) How do we generate the hydrologic ensembles that reflect the total uncertainties?
3) How can climate information, such as climate model forecasts or teleconnections, be used reliably in seasonal hydrologic forecasting?
4) How do we validate hydrologic ensembles for extreme events?

These are essential challenges that the community faces in seasonal hydrologic predictions. To help answering these questions, the following research objectives have been established:

1) Understand the potential predictability of the climate system (more specifically precipitation and temperature) at seasonal time scales.
2) Identify regions where potential predictability exists in dynamic climate models
3) Develop methodology to process atmospheric information from dynamic climate models to meet the needs in hydrologic applications, including information merging, bias correction, spatial and temporal downscaling, etc.
4) Identify the ensemble size, initial conditions and models that are necessary to provide unbiased and reliable probabilistic forecast
5) Explore the possibility of multi-climate model and multi-hydrologic model approach in seasonal hydrologic prediction.
6) Share data and other software tools with the community so that we can move towards a community based hydrologic prediction system.

Accomplishments

1) Background

Studies in the last two decades or so have demonstrated the possibility and feasibility of seasonal climate predictions with dynamical climate models. At present, seasonal climate predictions are been made routinely at several operational weather and climate prediction centers and research institutes, including the National Centers for Environmental Prediction (NCEP), the European Centre for Mid-range Weather prediction (ECMWF) and International Research Institute on climate change (IRI) at Columbia University. The predictions with coupled climate models have shown skills over the tropics. In the mid-latitudes, their skill is improving, and some of them have shown skills that are equal to statistical models (Saha *et al.* 2005). There are possibilities that these seasonal climate forecasts can contribute to the development of seasonal hydrological prediction capabilities. One of the questions that we want to address with the southeast testbed is how we can better use of the information provided by seasonal climate forecast in seasonal hydrologic predictions.

Two major issues related to using seasonal climate forecast from dynamic climate models is the mismatch in spatial and temporal scales, and the uncertainties associated with this information. Driving hydrological models requires minimum amount of information about surface meteorology at daily level in most cases, while information from seasonal climate forecast is normally available at monthly timescale and over large regions. Therefore, significant efforts are needed to do the temporal downscaling of climate model forecast in a reasonable manner. The uncertainties associated with climate model forecast are determined by the potential predictability of the climate system at seasonal timescale and the actual predictive skill of the models. The southeast U.S. has been shown to have relative better predictability in several studies.

2) Methodology development

During the past two years, our major accomplishment has been the development of the Bayesian merging method for combining seasonal climate forecast from multiple climate models with observed climatology to produce our best estimate of the future conditions. The Bayesian merging method uses the observed climatology as the prior, and builds the likelihood function of each model forecast with a linear regression between the model hindcasts and historical observations. The basic idea is that models with better skill should receive higher weights in the averaging (merging). If all models are lack of skill, the observed climatology should receive more weight. The method is documented by a paper recently submitted to *JGR-Atmosphere* and is currently under review.

3) Supporting Data Set Development

To facilitate the participation of other research groups in this testbed, the data sets being used to develop the seasonal hydrologic forecast system is available to the community via request, and will be made available via ftp in the near future. These data sets include: (1)

Observed atmospheric forcing. (2) NLDAS realtime forcing, (3) NCEP Climate Forecast System (CFS) hindcast and forecast (4) European Union DEMETER project seasonal climate forecast (5) Streamflow data from USGS.

a. Observed atmospheric forcing

Maurer *et al.* (2002) presented a long term atmospheric forcing dataset based on station observations for the conterminous United States and portions of Canada and Mexico. The dataset spans the period 1949–2000, and is at a 3-h time step with a spatial resolution of 1/8 degree. The temperature and precipitation in the data set is a gridded product derived directly from station observations, which is different from reanalysis that is normally used as observations in many studies. The dataset is aggregated at daily and monthly time scales. The monthly values are further aggregated to climate model grid. This data set has been widely used by the hydrometeorology community in various studies (references). This date set can serve three purposes in seasonal hydrologic prediction studies. The aggregated monthly climate model scale observation is used to evaluate the climate model forecast skill. The daily 1/8 degree observation can be used as the base for spatial and temporal downscaling during the forecast period. The land surface states from an offline simulation forced by this daily forcing dataset serve as a proxy of observations in that large scale continuous observations of soil moisture do not exist. There are other long term observations of precipitation and temperature available at coarse resolution that can be used for the first purpose, but this dataset keeps the consistency between observations at coarse and find scales.

b. NLDAS realtime forcing

The North American Land Data Assimilation System (NLDAS) is a multi-institutional project with the goal of providing better land surface initial conditions to operational weather and climate forecasts. One important contribution of NLDAS to the community is the creation and archive of a near realtime atmospheric forcing at 1/8 degree resolution for the conterminous US with good quality (Mitchell *et al* 2004, Cosgrove *et al*. 2003 and Luo *et al*. 2003). This hourly forcing is a combination of Eta Data Assimilation System and observations including rain gages and radar observations of precipitation and satellite retrievals of the surface shortwave radiation. The realtime NLDAS forcing can be used as an realtime extension of the Maurer data set.

c. NCEP Climate Forecast System and its hindcast

The NCEP Climate prediction System (CFS) is a fully coupled ocean-land-atmosphere dynamical seasonal prediction system, and it is the current operational system used in NCEP for seasonal prediction. It consists an interactive ocean component that is the GFDL Modular Ocean Model version 3 (MOM3)(Pacanowski and Griffies 1998) and an atmospheric component that is the atmospheric model in the NCEP global forecast system. The atmospheric component is the same as the one used for operational weather forecasting without additional tuning for climate applications except for having a coarser resolution (T62L64). The atmospheric and oceanic components are coupled and exchange daily averaged quantities, such as heat and momentum fluxes once a day, with no flux correction. Full interaction between these two components is confined to 66°S to 50°N, due to the difference in latitudinal domain.

In realtime, the CFS is run twice a day from 0Z and 12Z initial conditions from the atmosphere and ocean of 7 days ago (Saha *et al*., 2005). The atmospheric initial condition is obtained from NCEP Reanalysis-2 (Kanamitsu *et al*. 2000) and the oceanic initial condition is obtained from NCEP Global Ocean Data Assimilation System (GODAS). The forecast period covers the first partial month and 9 full months into the future. With such a setup, there are about 60 members available for each nine month forecast period. Besides the daily forecasts, a set of fully coupled retrospective predictions spanning the period of 1981-2004, with 15 members per month out to nine months into the future, have been produced with the CFS (Saha *et al*., 2005). These retrospective forecasts are important for bias correction and calibration of the operational seasonal prediction. More details about CFS and its forecast products can be found in Saha *et al*, (2005) and NCEP web sites.

We have processes the CFS hindcast data and also have the capability to process the forecast in realtime. This dataset can be used to spatial and temporal downscaling studies.

d. European Union DEMETER seasonal forecast

DEMETER is the acronym of the EU-funded project entitled "Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction". The objective of the project is to develop a well-validated European coupled multi-model ensemble forecast system for reliable seasonal to interannual prediction (Palmer *et al*., 2005). Seven atmospheric general circulation models from operational and research institutes in Europe participated in the project. These atmospheric GCMs were coupled with one ocean GCM to produce a multimodel seasonal forecast 6 months into the future with 9 ensemble members from each model. A 20-year hindcast dataset was created with the DEMETER system and is available from the European Centre for Mid-range Weather Forecast (ECMWF). The hindcasts start from the first day of February, May, August and November initial conditions. More information about DEMETER can be found in Palmer *et al*. (2005) and ECMWF web site.

One big advantage of the Southeast testbed is that we provide a platform to develop methodology for using seasonal climate forecasts from multiple climate models. The hindcast from DEMETER is a big contributor to the multimodel platform.

e. Streamflow data from U.S. Geological Survey

Daily streamflow data for the last 60 years are obtained for hundreds of stream gages from the U.S. Geological Survey. This dataset is used in the hydrological model calibration, and forecast evaluation. Additionally, this long term dataset can also be used in adjustment of streamflow prediction based on the cumulative distribution functions of observed monthly streamflow and modeled streamflow from the offline simulation. This is to take care of possible biases in streamflow prediction when calibration is not sufficient.

4) Example Results

As most of our research activity was over the Ohio River basin, here we present some example results of the Ohio River basin. The datasets are also available over the Southeast US, and the same research activities can be easily extend back to the Southeast.

Over the Ohio River basin, we have produced a complete set of hindcast for the 20-year period (1981-2000) with initial date on February, May, August and November. Each forecast set consists of three type of forecast: a climatologic forecast (ESP) and CFS-based forecast (CFS) and multimodel-based forecast (CFS+DEMETER). The atmospheric forcing (precipitation and air temperature) and model output (soil moisture, streamflow, snow, etc) are all available. Figure 1 illustrate the ensemble mean forecast of total column soil moisture for a 6-month period starting May 1990. During this period, most of the Ohio river basin has higher than normal soil moisture while the lower part of the basin developed drying episode then disappeared by the end of the forecast period.

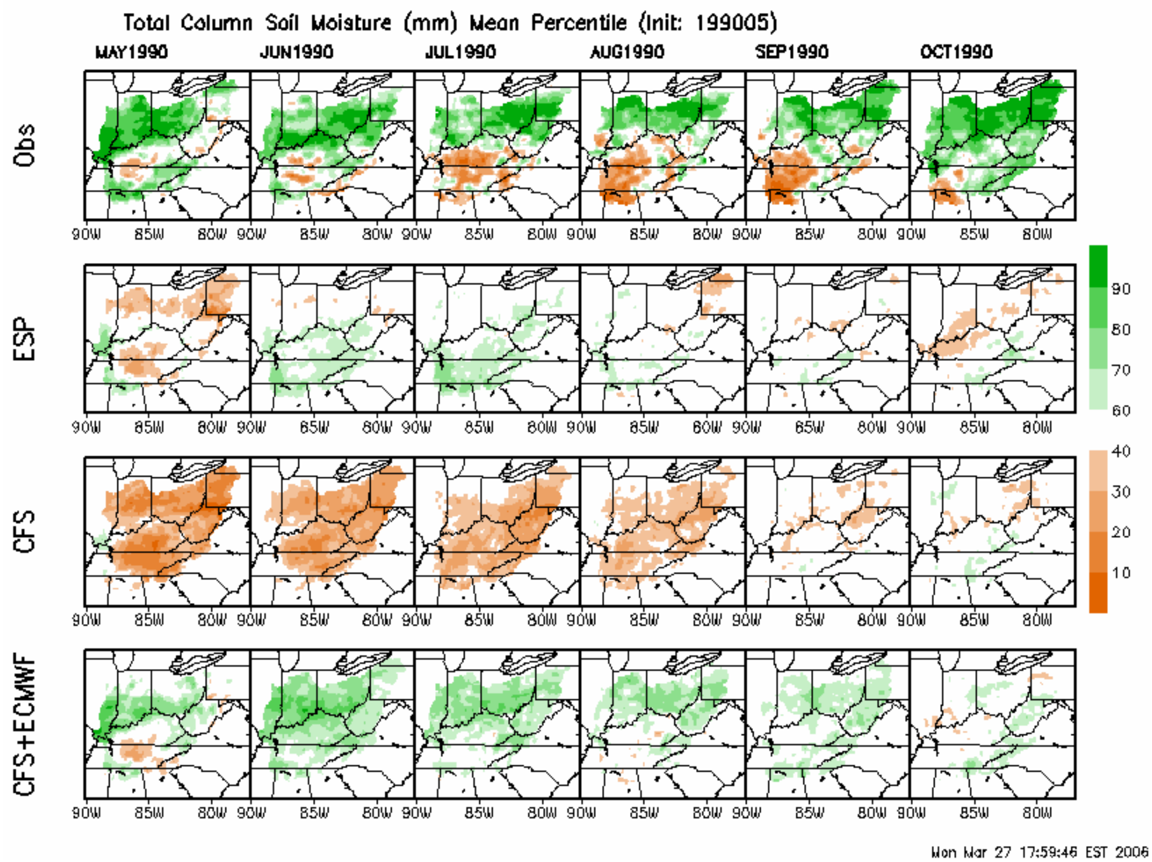

Figure 1: Soil moisture predictions over the Ohio River basin from three types of forecast. What shown is the percentile of the ensemble mean forecasts.

Among the three types of forecast, multimodel-base forecast shows the most skill over the region. These three types of forecast all start with the same initial condition, but because the precipitation forecasts are so different, they produced very different soil moisture sequences. An example of streamflow prediction is presented in Figure 2. It is of our interests to perform the similar type of study over the southeast US, and also to include other hydrologic models, and to test other type of merging and downscaling techniques.

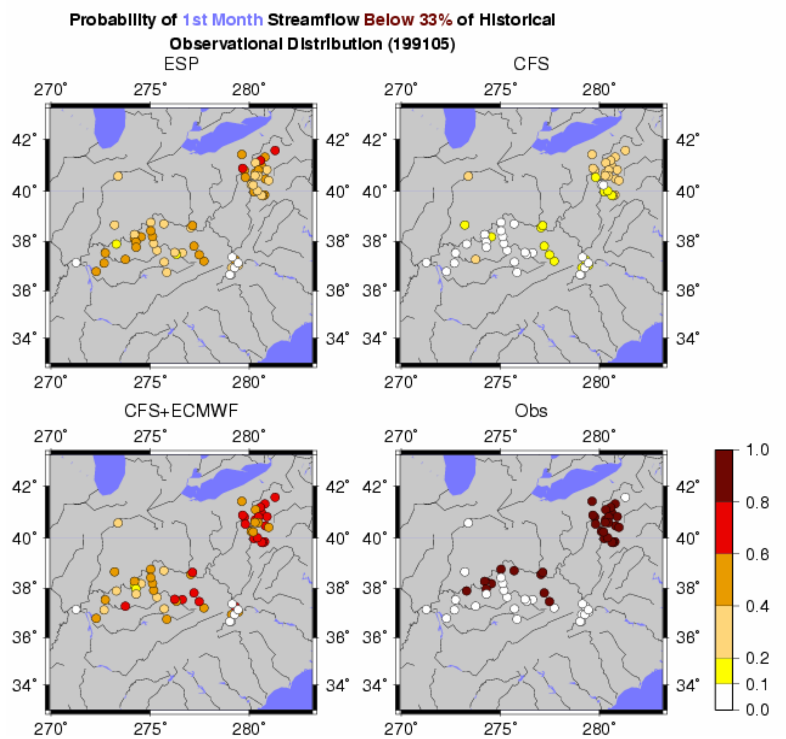

Figure 2: Probabilistic streamflow prediction products from the seasonal hydrologic forecast system over the Ohio River basin. At each gage station, the probability of an event, defined as monthly streamflow below 33% of the historical observational distribution, is estimated from the ensemble members. The probabilistic forecast is compared with actual observations expressed as 0 or 1.

Future Plans

Our future research activities in the next year will include the following:

1) Continued development and improvement of the Bayesian merging method and its application in atmospheric pre-processor
   The Bayesian merging method is currently only applied to precipitation on monthly timescale. We plan to apply it to air temperature and possibly other forcing variables. When creating atmospheric forcing to drive land surface models, we only apply the posterior distribution on a montly basis. It is necessary to derive distributions for other aggregated period, i.e, first two months, etc. This is expected to improve the forecast skill since precipitation over two months period is more predictable than monthly precipitation for a given lead time. We will also test the idea of adding spatial aggregated information in the merging and forcing generation.

2) Testing of alternative merging and downscaling algorithms
   As one of the HEPEX testbed for downscaling, we would like to collaborate with other research groups to test our other downscaling algorithms, then compare them to find the strength and weakness of each algorithm.

3) Collaboration with other testbed to move towards shorter timescale forecast (2-15 days) and towards CHPS

   Discussion with Dr. John Schaake in the last few weeks has generated tremendous interests in extend our seasonal forecast to a shorter leadtime using NCEP Global Forecast System (GFS) forecast and the ensemble pre-processor developed by Dr. John Schaake. This will help us move towards a seamless suite of forecast system at different lead times.

4) Supporting data set development.
   Supporting data sets will continuously be developed and shared with the community.

References and selected publications and presentations

Luo, L. and E. F. Wood 2006: Assessing the idealized predictability of precipitation and temperature in the NCEP Climate Forecast System, *Geophys. Res. Lett.*, **33**, L04708, doi:10.1029/2005GL025292.

Lifeng Luo, Eric F. Wood and Ming Pan 2006: Bayesian Merging of Multiple Climate Model Forecasts for Seasonal Hydrologic Predictions. JGR-Atmosphere, in review.

Luo, L. and E. F. Wood 2006: Seasonal hydrological prediction with VIC model over the Ohio River basin. In preparation for submission to *Journal of Hydrometeorology.*