Experimental forecasts of streamflow

Martyn P. Clark

Presentation at the
Office of Hydrologic Development
21st April, 2004

Integrated set of forecast inputs (days—seasons)…

(different models for different forecast lead times)

Integrated set of forecast inputs (days—seasons)…

(different models for different forecast lead times)

<u>Short-term:</u> Bias-corrected output from a regional model

- 72-hour forecasts from the regional reanalysis

<u>Medium-Range</u>: Downscaled output from a global forecast model

- 14-day forecasts from the CDC frozen version of NCEP's MRF model

<u>Seasonal time scales:</u> Dis-aggregated probabilistic forecasts

- weather generator conditioned on climate indices
- weather generator conditioned on probabilistic forecasts

Integrated set of forecast inputs (days—seasons)…

(different models for different forecast lead times)

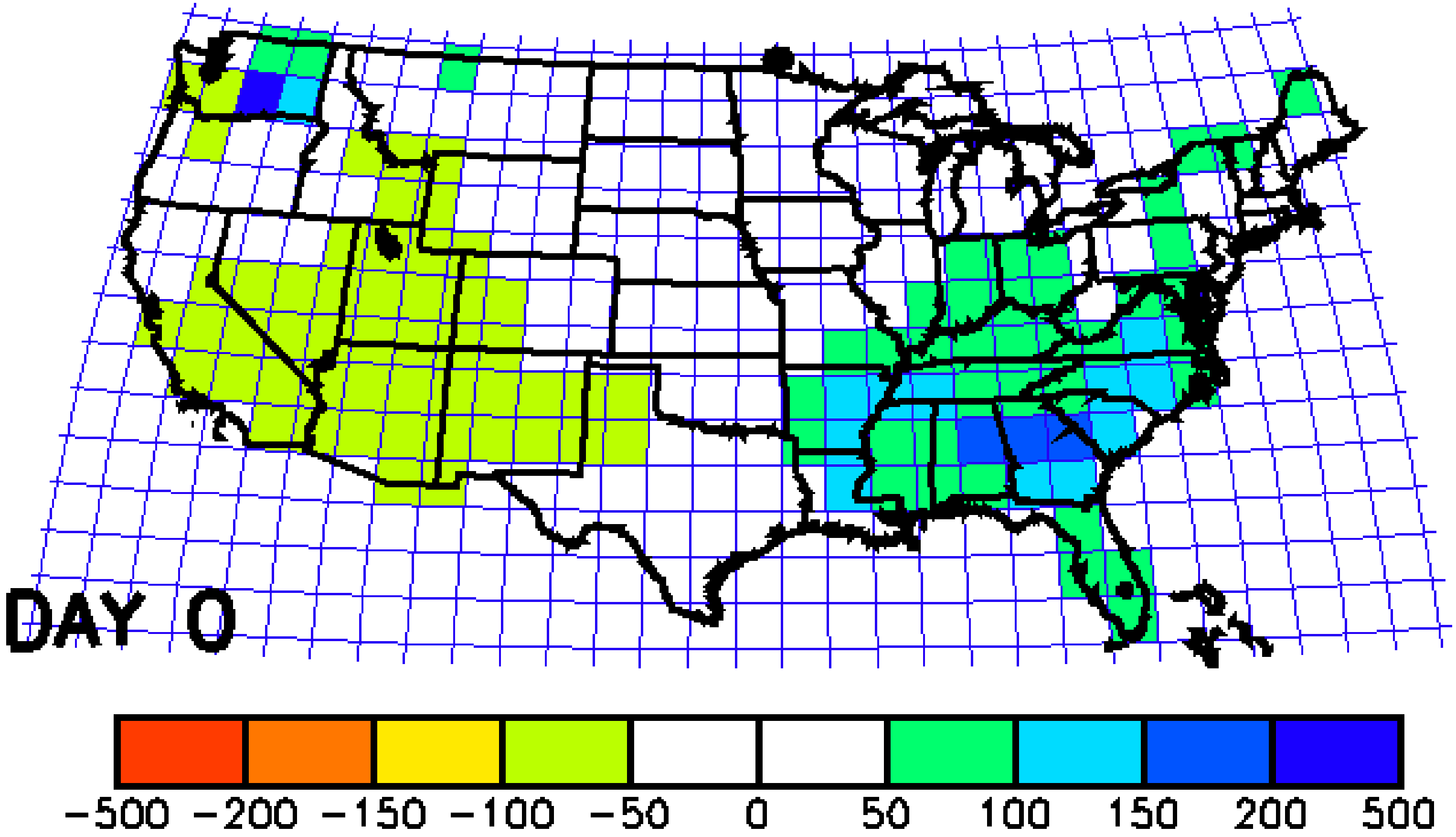
Short-term: Bias-corrected output from a regional model
- 72-hour forecasts from the regional reanalysis

Medium-Range: Downscaled output from a global forecast model
- 14-day forecasts from the CDC frozen version of NCEP's MRF model

Seasonal time scales: Dis-aggregated probabilistic forecasts
- weather generator conditioned on climate indices
- weather generator conditioned on probabilistic forecasts

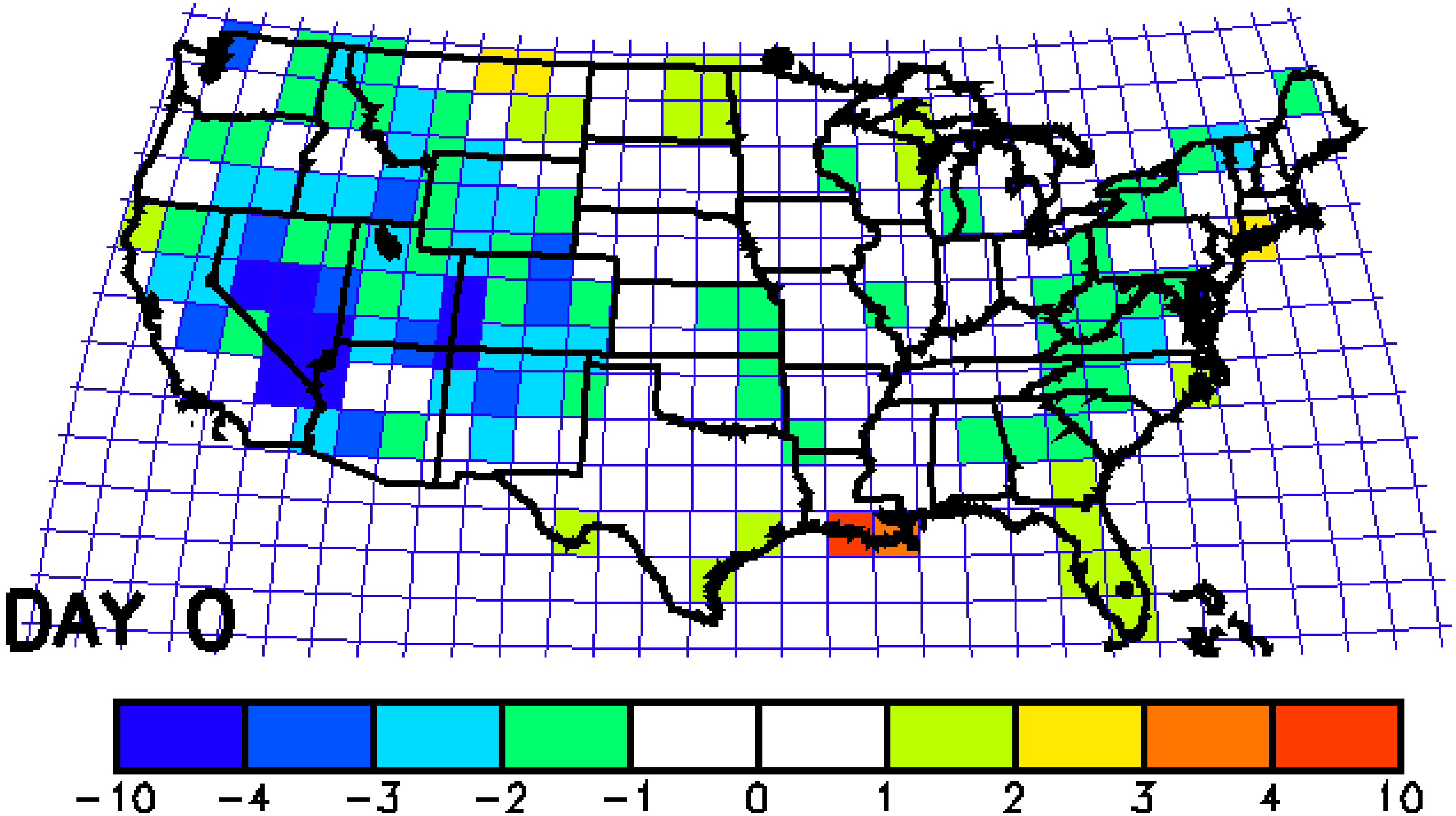
Requirements:
- an ensemble *daily* sequences of weather
- preserve inter-site correlations, temporal persistence, and correlations between variables
- minimize abrupt changes when a new model is introduced

PRECIPITATION BIASES



Precipitation biases are in excess of 100% of the mean

TEMPERATURE BIASES



Temperature biases are in excess of 3^{o}C

The CDC Re-forecast experiment

- Jeff Whittaker and Tom Hamill at the NOAA-CIRES Climate Diagnostics Center have used the 1998 NCEP MRF to generate medium-range forecasts for the period 1979 to the present.

- CDC are continuing to run the 1998 NCEP MRF in real time.

- The NWP hindcast (1979-2001) is used to develop regression models between MRF output and precipitation and temperature at individual stations, and apply the regression coefficients to the CDC experimental forecasts in real-time.

- The resultant local-scale precipitation and temperature forecasts are used as input to the CBRFC hydrologic modeling system to provide real-time forecasts of streamflow.

Downscaling approach

- For hydrologic applications we need to:
  - Obtain reliable local-scale forecasts of precipitation and temperature
  - Preserve the spatial variability and temporal persistence in the predicted temperature and precipitation fields
  - Preserve consistency between variables

- Multiple linear Regression with forward selection

$$Y = a_0 + a_1X1 + a_2X2 + a_3X3 \ldots + a_nXn + e$$

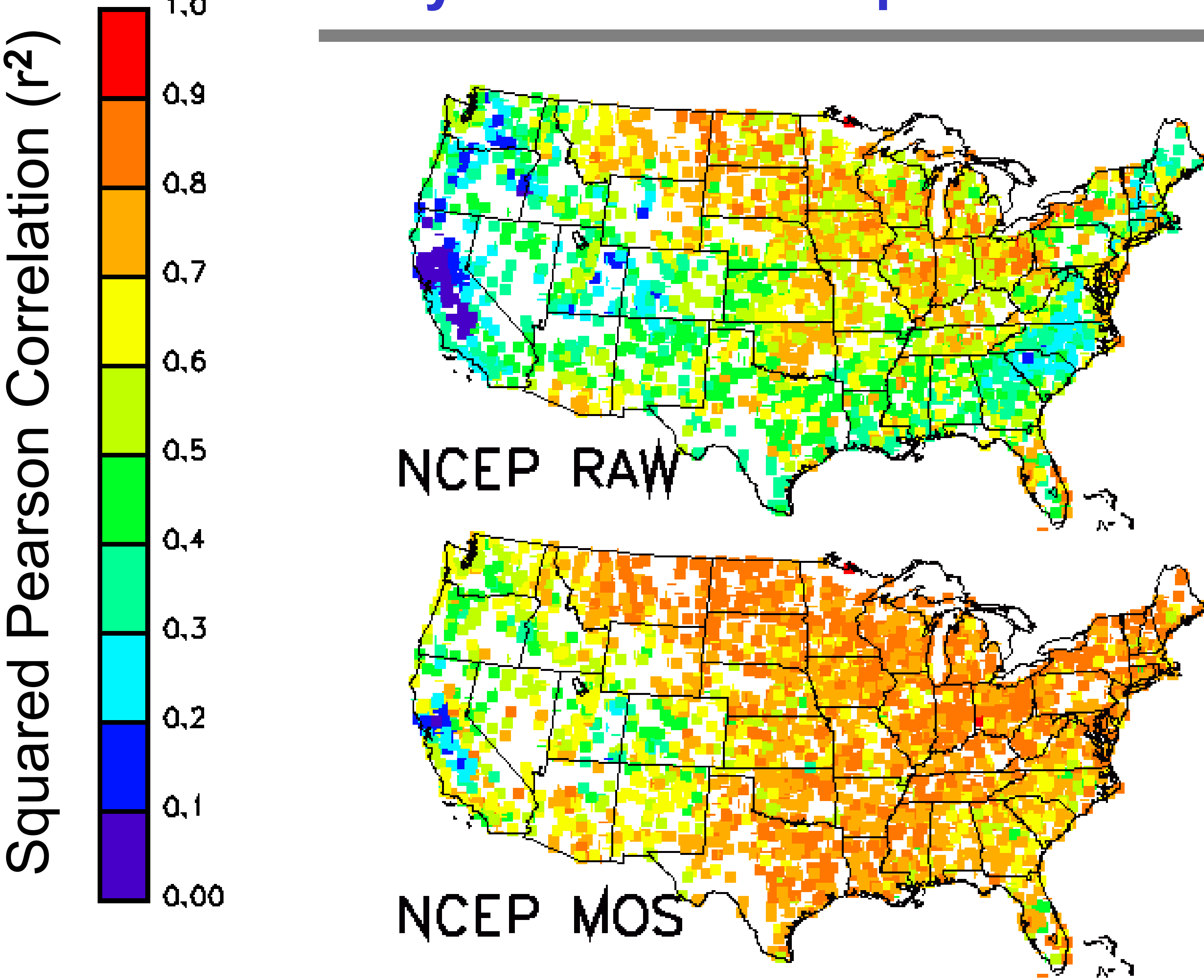
- A separate equation is developed for each station, each forecast lead time, and each month.

- Use cross-validation procedures for variable selection – typically less than 8 variables are selected for a given equation

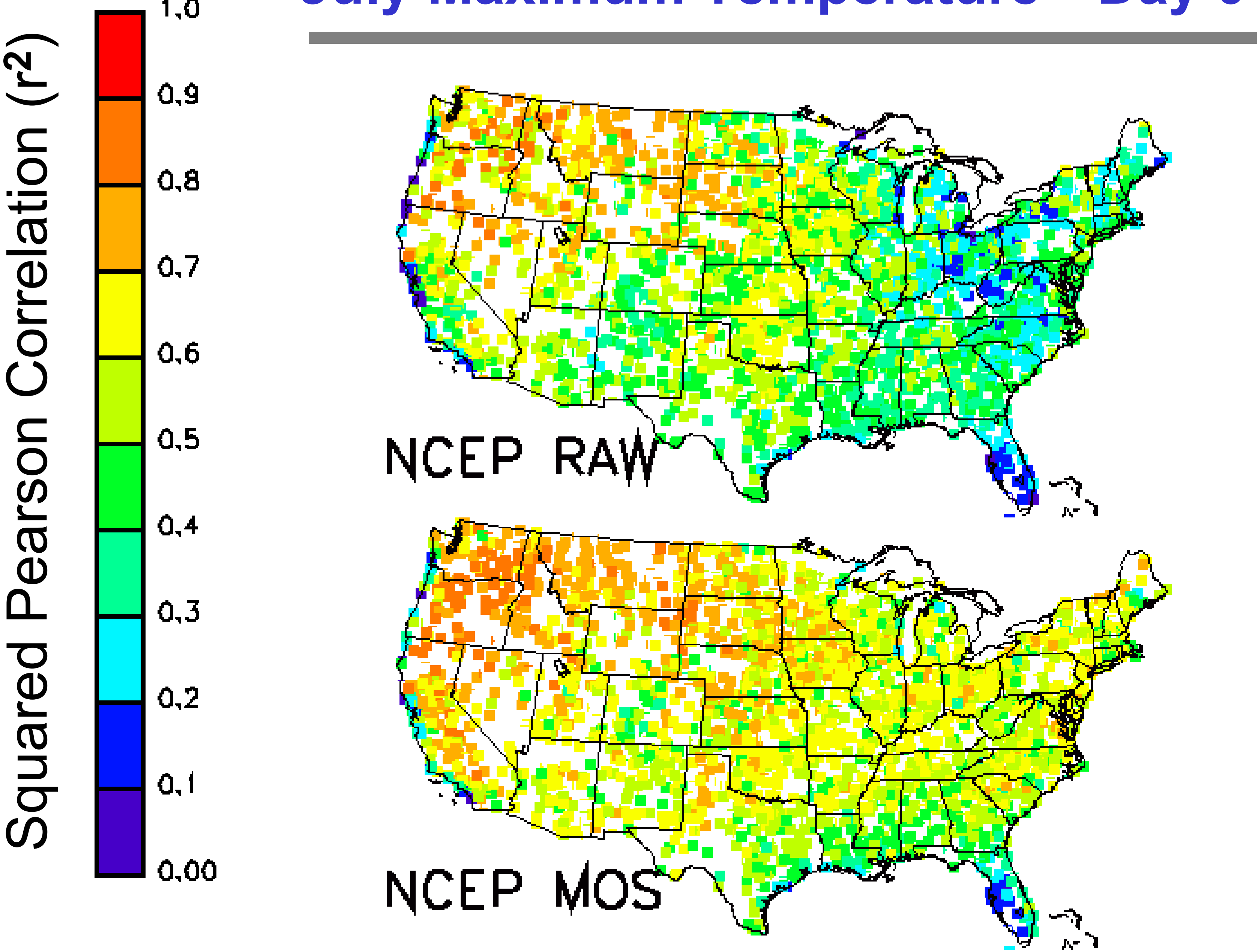
- Stochastic modeling of the residuals in the regression equation to provide ensemble time series

- Shuffling of the ensemble output to preserve the observed spatial variability, temporal persistence, and consistency between variables.

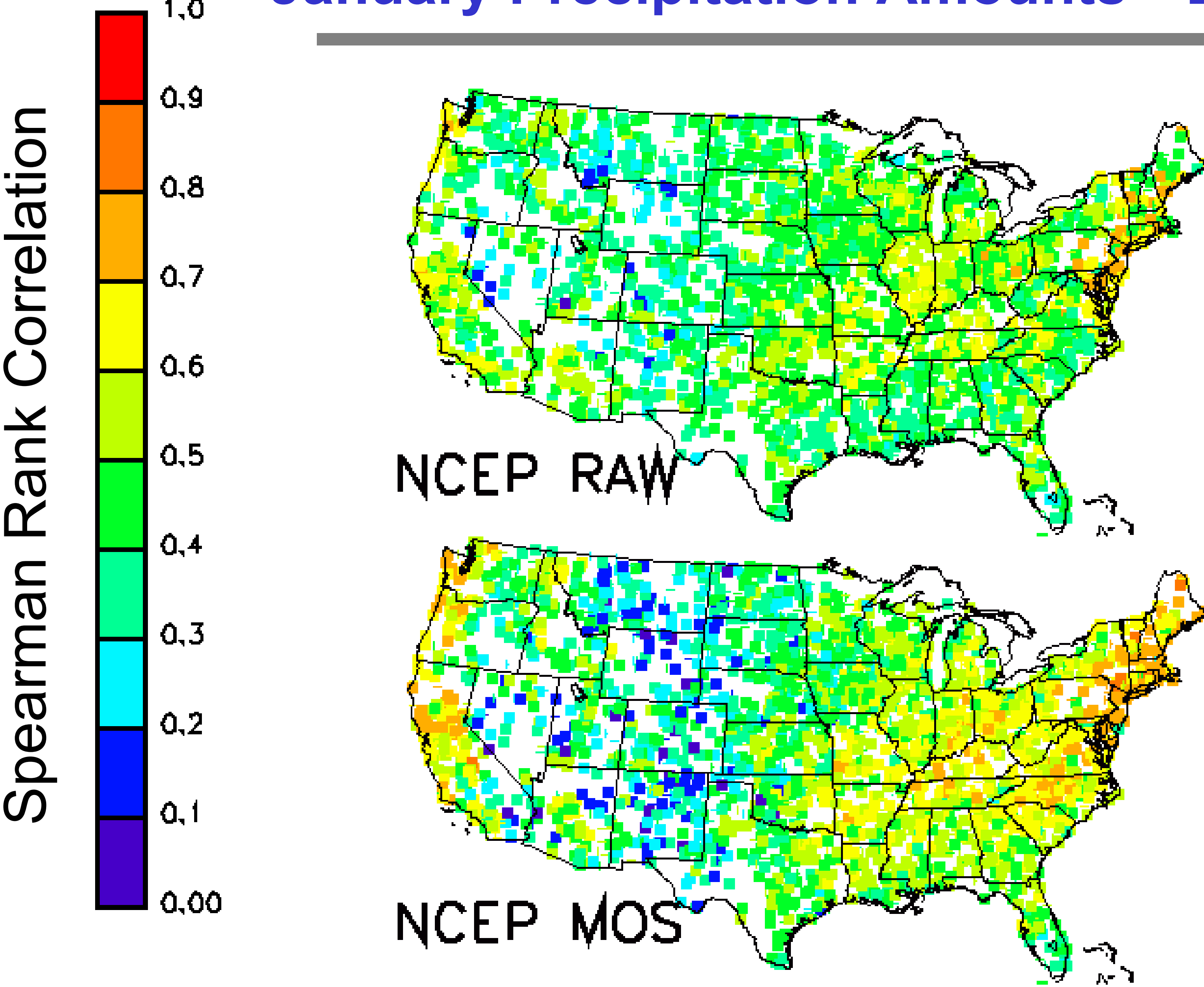
January Maximum Temperature—Day 0



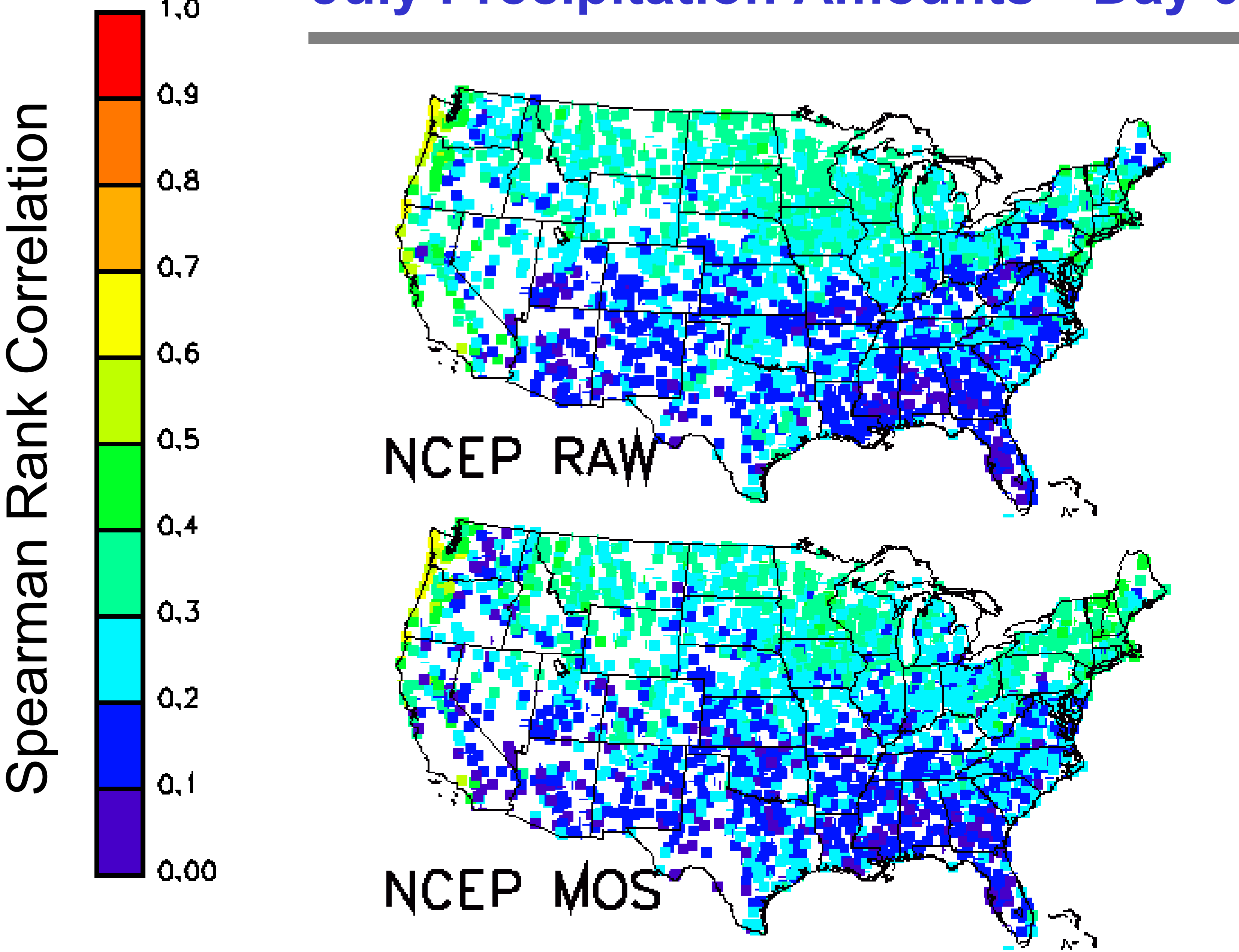
July Maximum Temperature—Day 0



January Precipitation Amounts—Day 0



July Precipitation Amounts—Day 0



Hydrologic Model

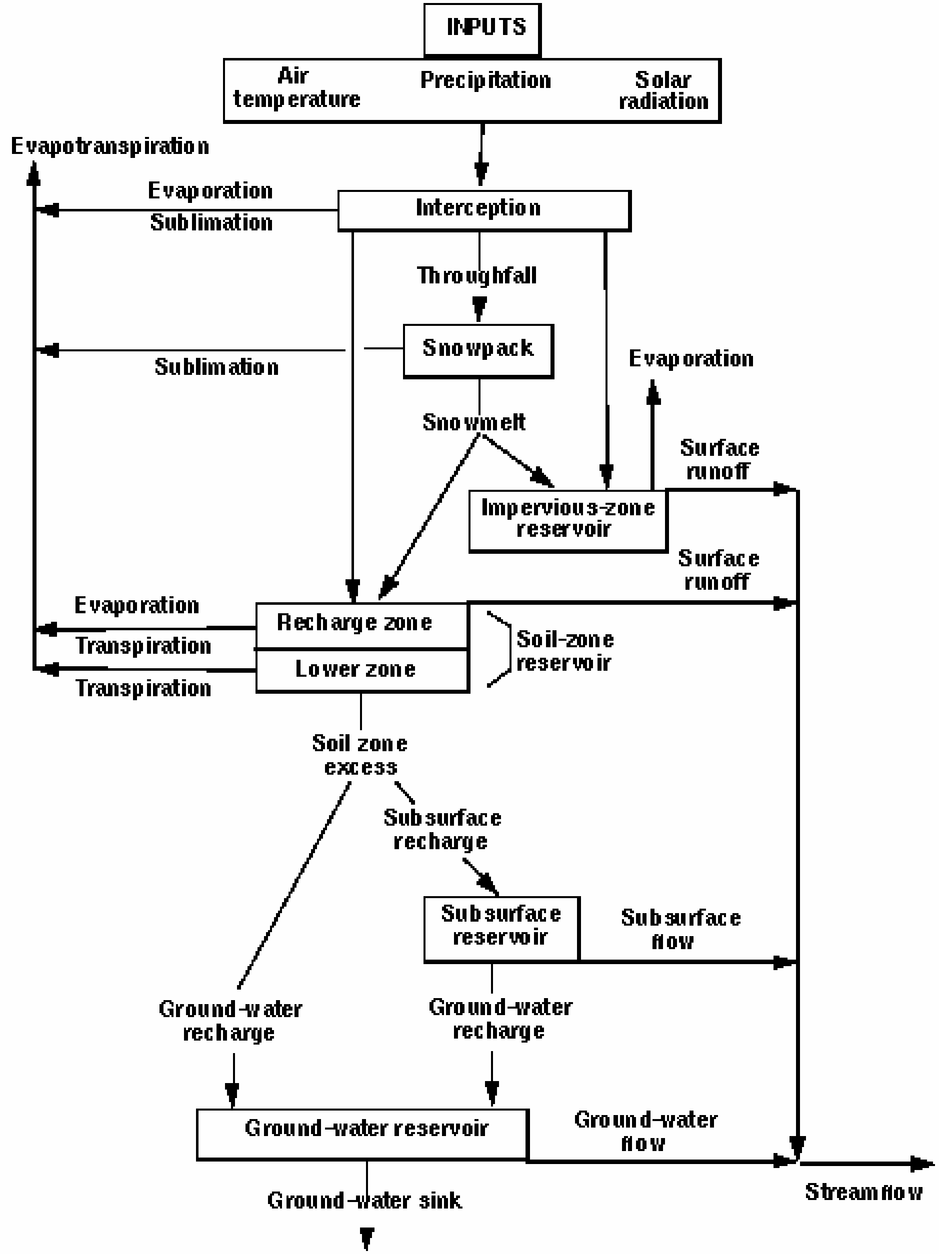
Precipitation Runoff Modeling System (PRMS)

[distributed –parameter, physically-based watershed model]

Implemented in:

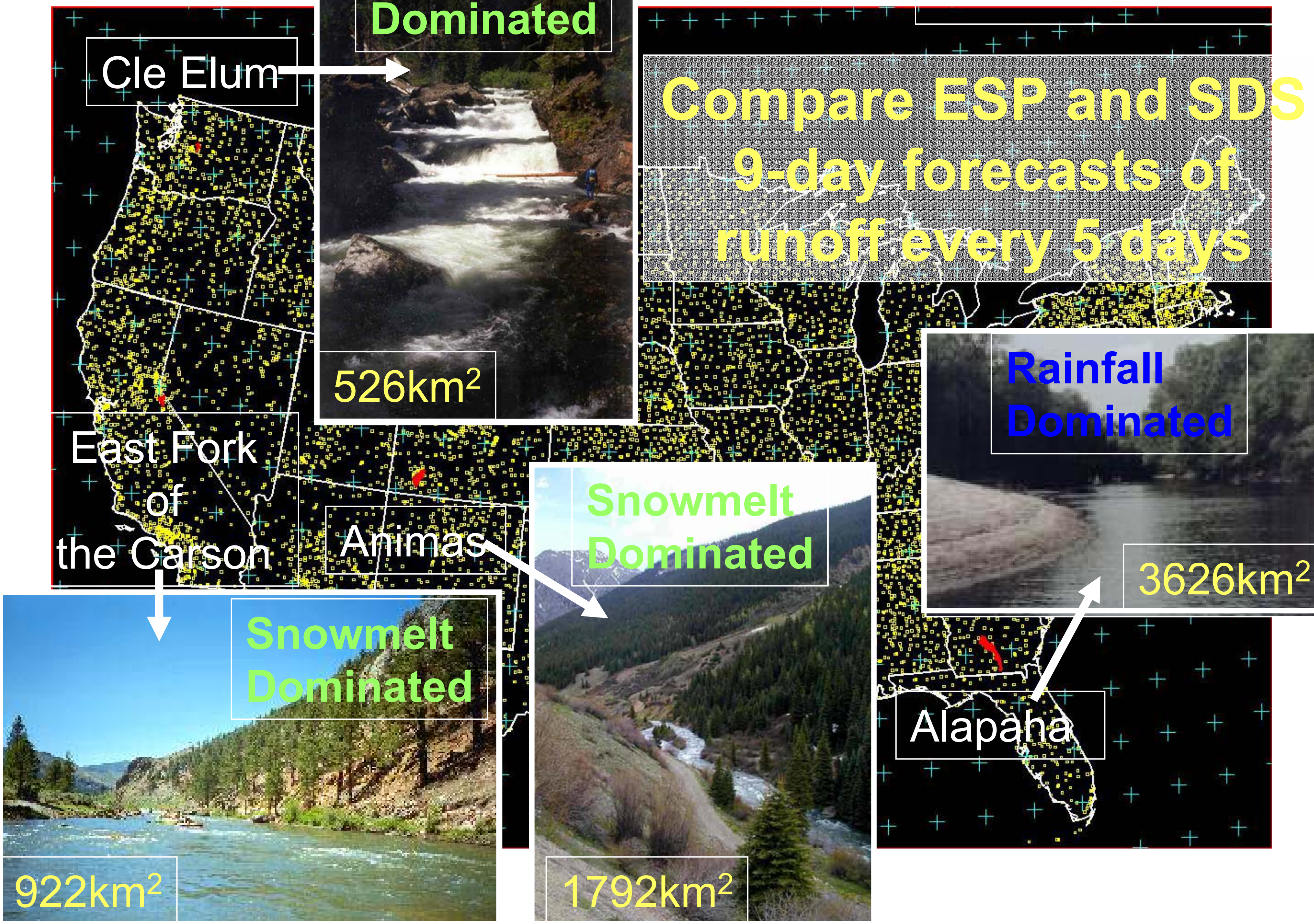
The Modular Modeling System (MMS)

[A set of modeling tools to enable a user to selectively couple the most appropriate algorithms]

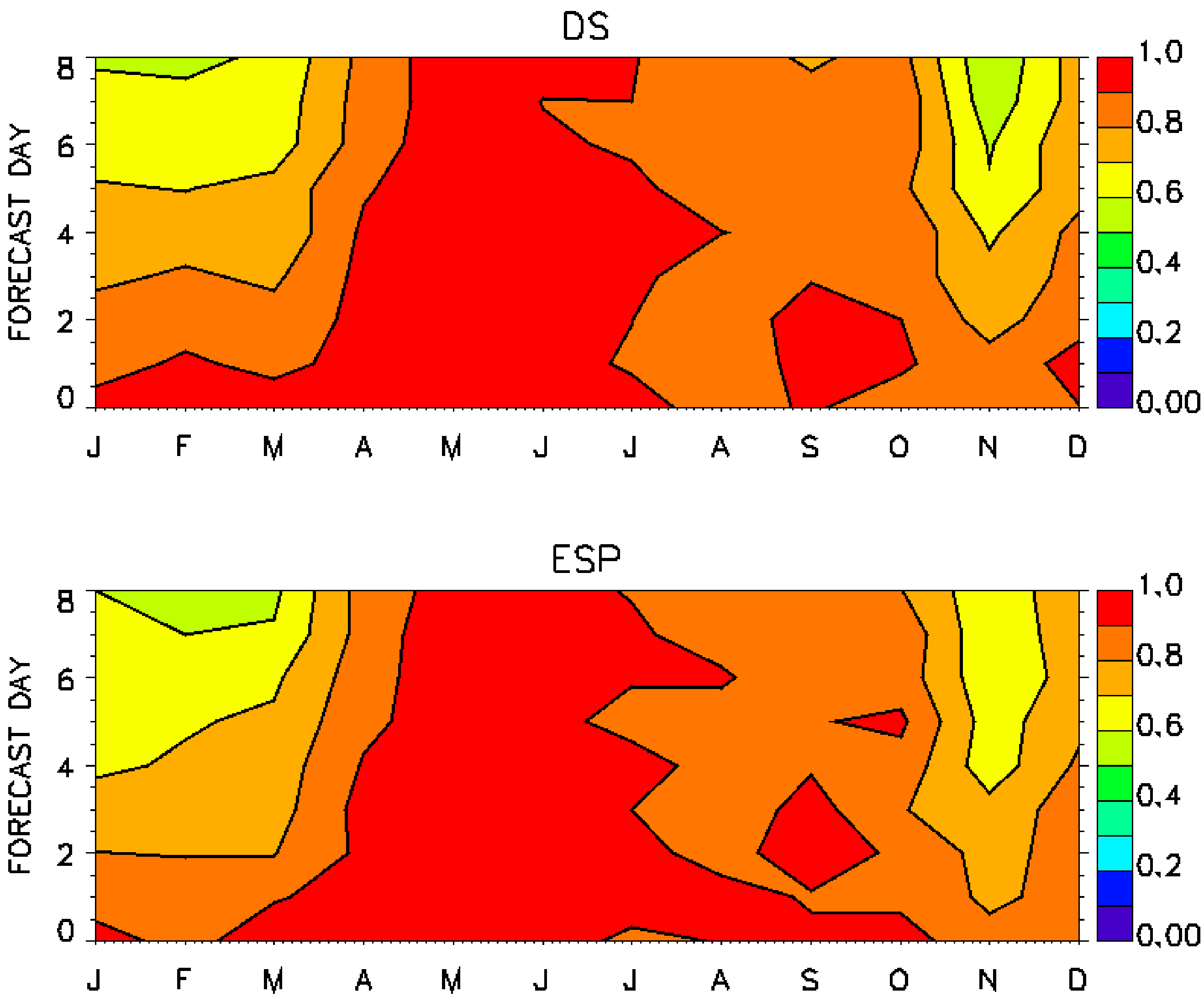


BASINS

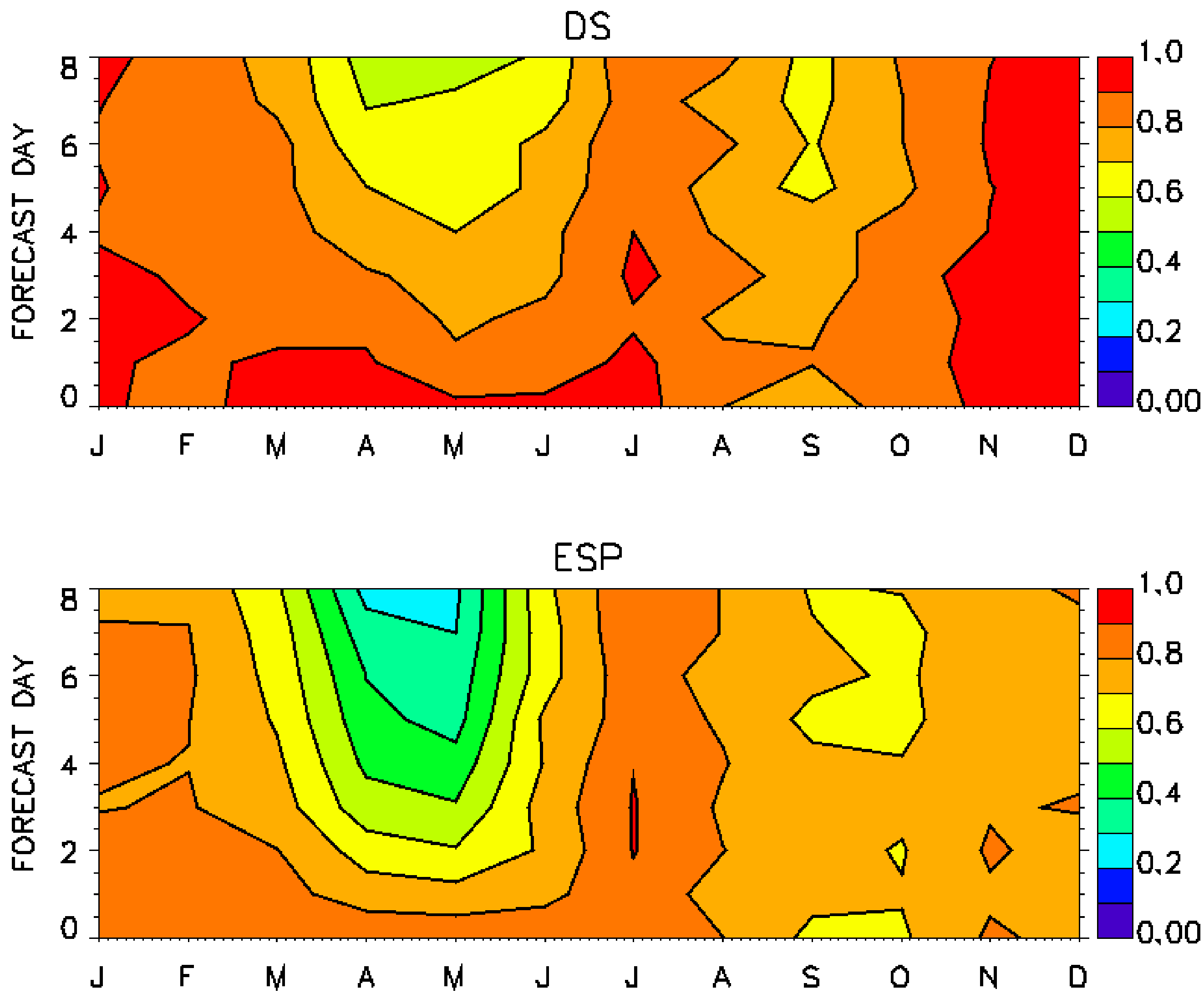
Compare ESP and SDS 9-day forecasts of runoff every 5 days



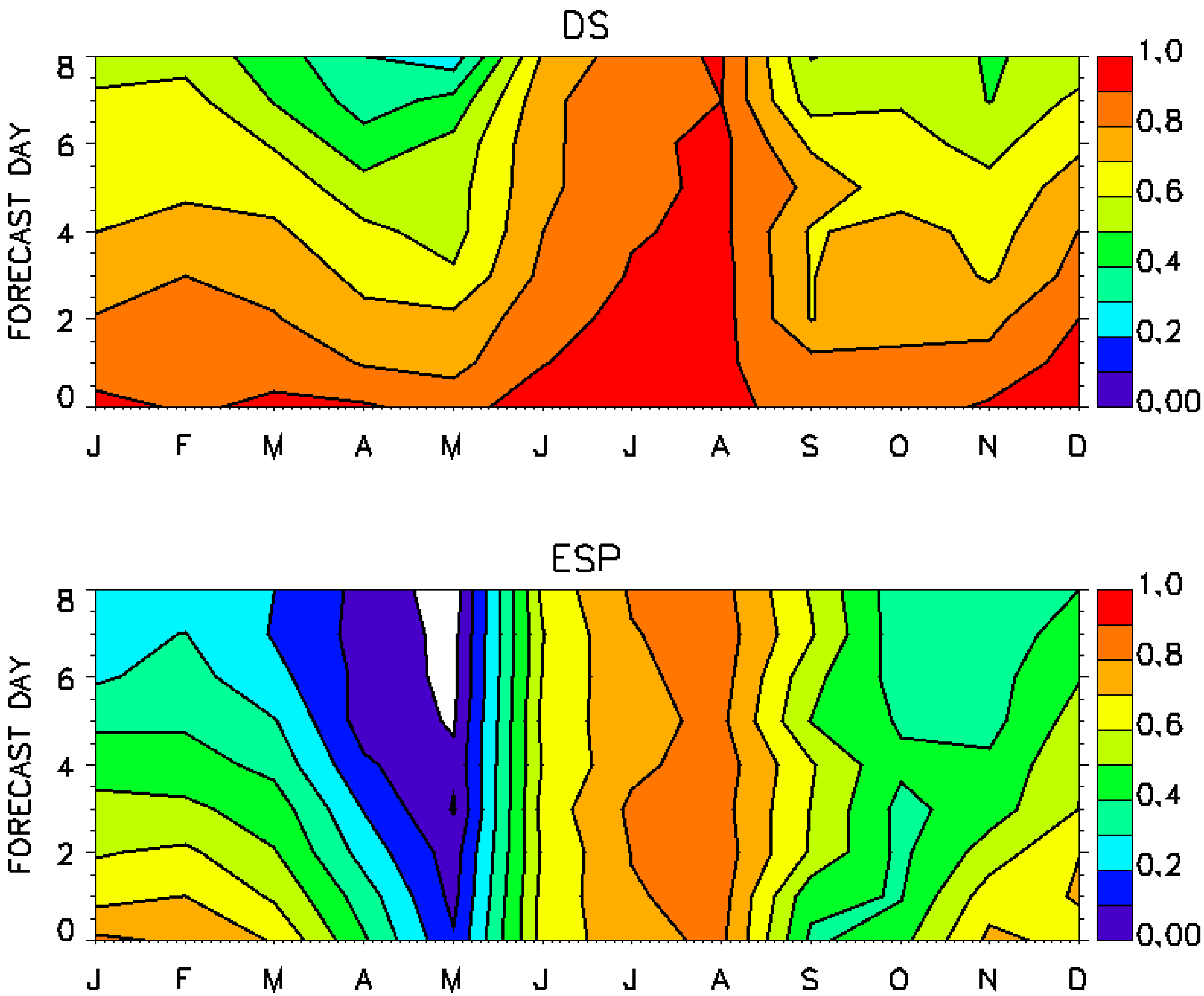
Alapaha River Basin (Southern Georgia)



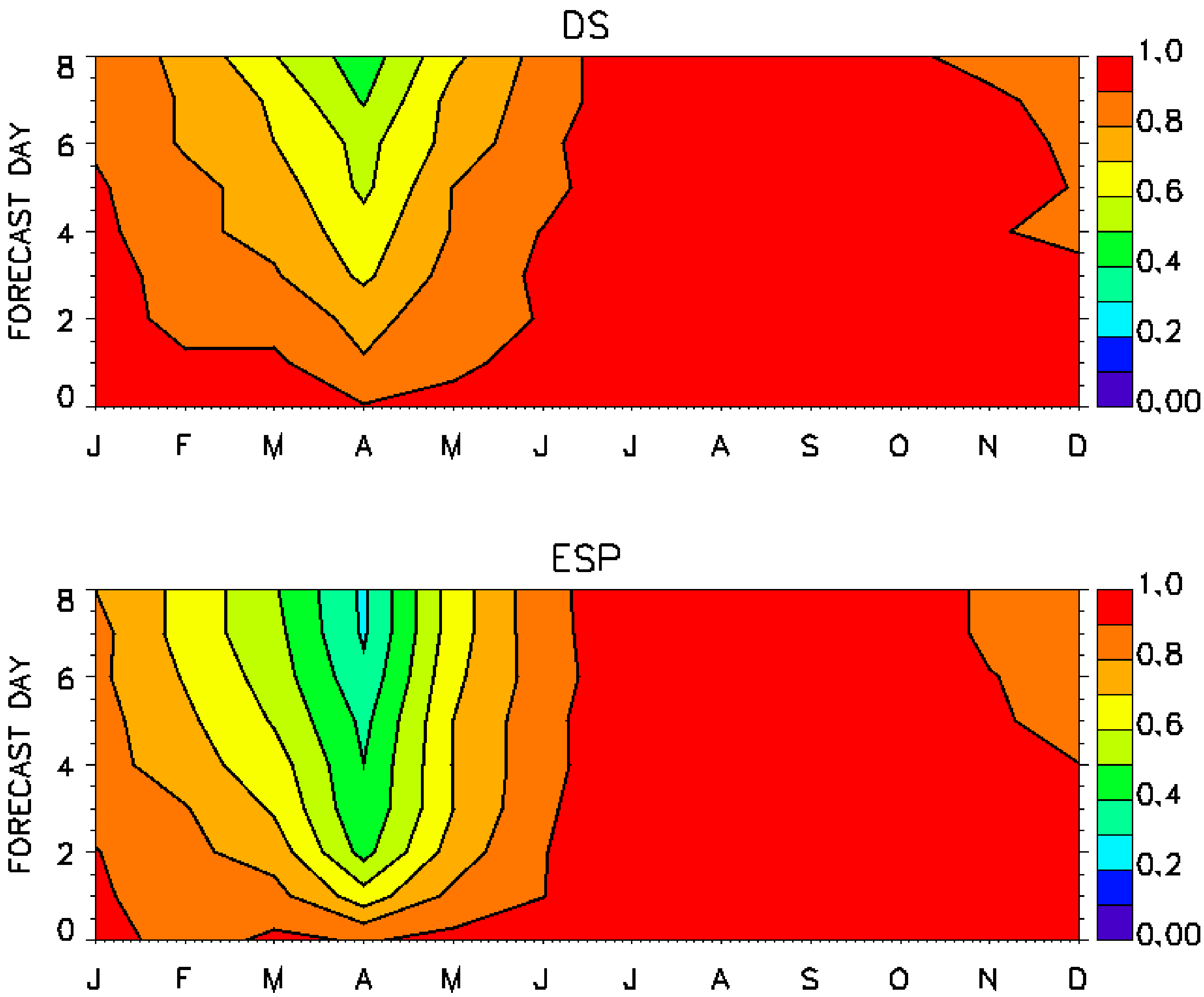
Animas River Basin (Southwest Colorado)



Cle Elum River Basin (Central Washington)



Carson River Basin (CA/NV Border)



Seasonal predictions… the weather generator model

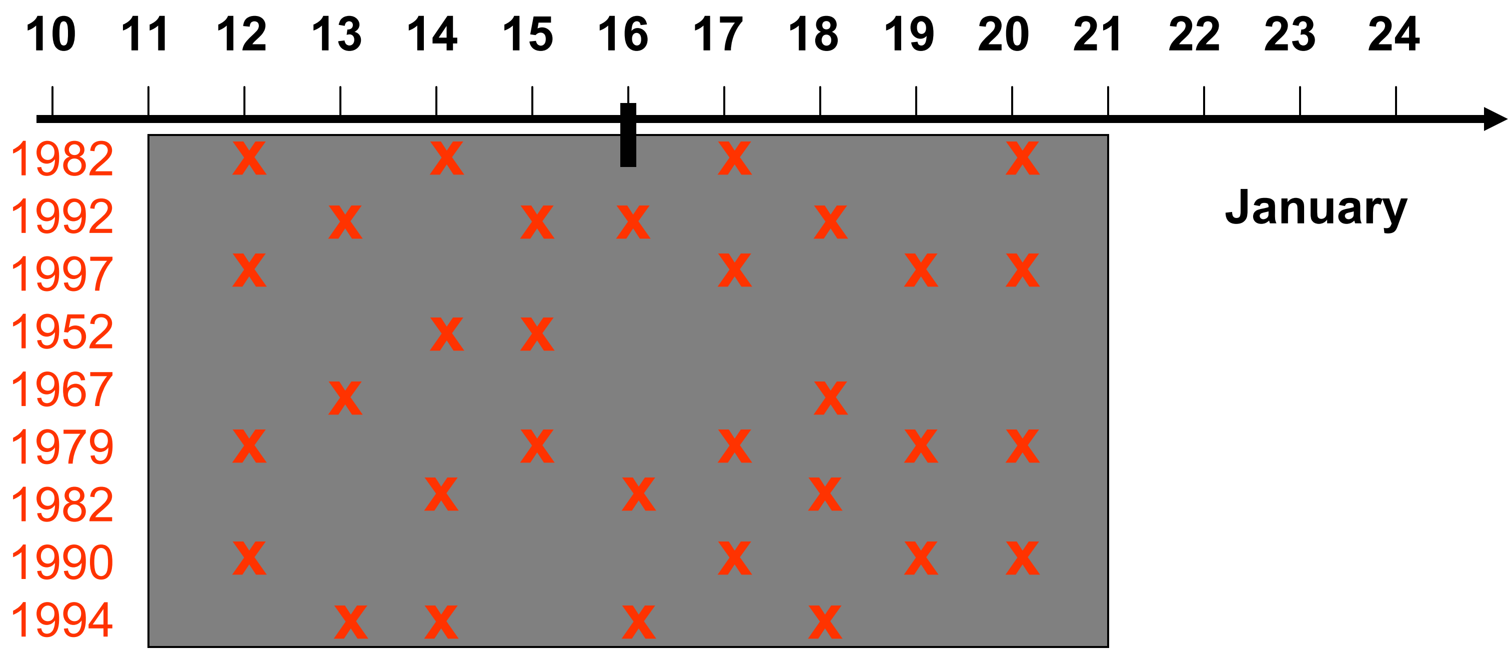
(1) Identify a subset of years from the historical record, such that the CDF from the selected years matches the CDF from the probabilistic forecast
(2) Re-sample data from the subset of years nens times
(3) Re-order the ensembles to preserve observed inter-site correlations, observed temporal persistence, and observed correlations between variables

The weather generator model… (seasonal predictions)

(1) **Identify a subset of years from the historical record, such that the CDF from the selected years matches the CDF from the probabilistic forecast**

(2) Re-sample data from the subset of years nens times

(3) Re-order the ensembles to preserve observed inter-site correlations, observed temporal persistence, and observed correlations between variables Re-sample data from the historical record nens times



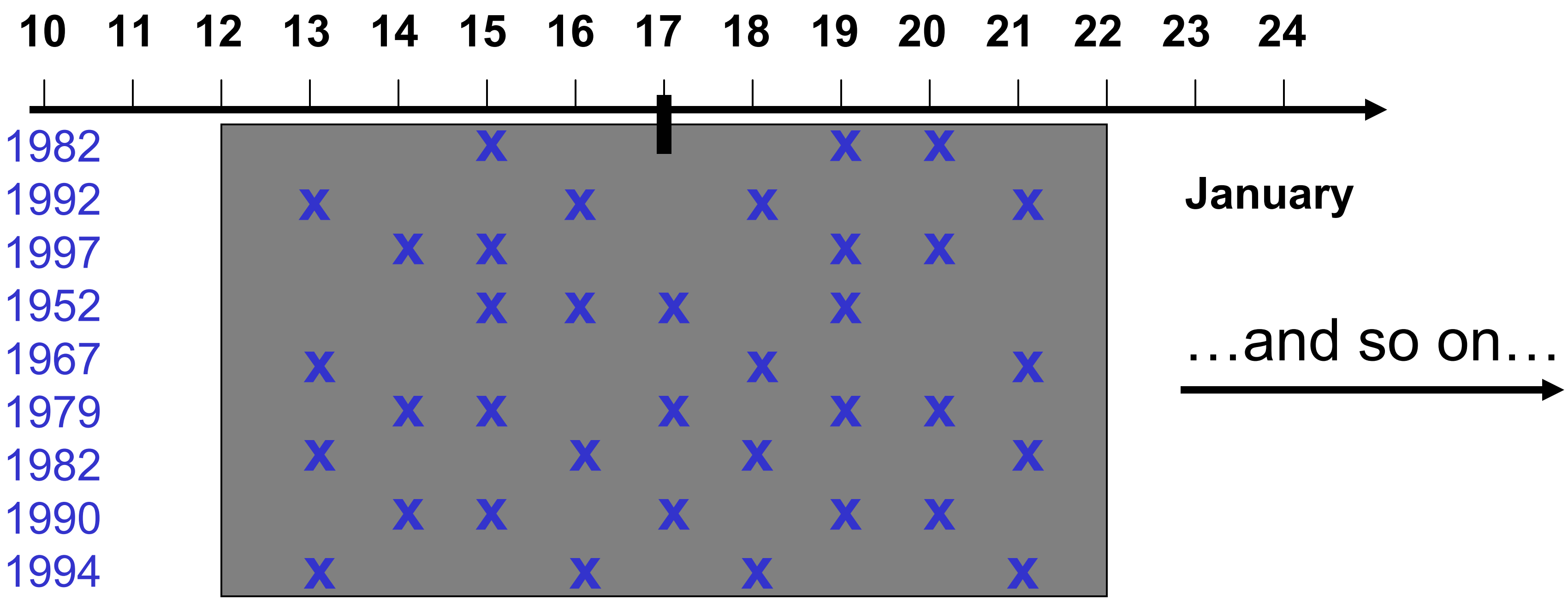
For 16th January, select an ensemble of data from a biased set of years

The weather generator model… (seasonal predictions)

(1) **Identify a subset of years from the historical record, such that the CDF from the selected years matches the CDF from the probabilistic forecast**

(2) Re-sample data from the subset of years nens times

(3) Re-order the ensembles to preserve observed inter-site correlations, observed temporal persistence, and observed correlations between variables Re-sample data from the historical record nens times



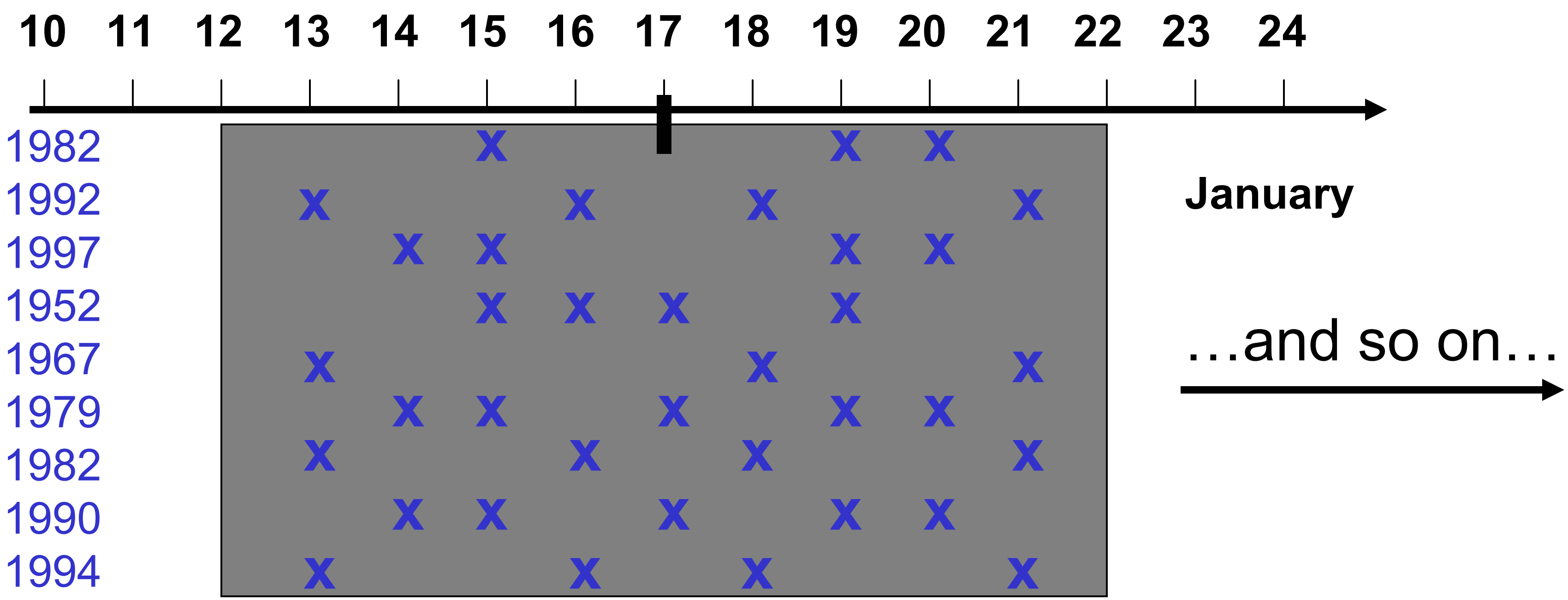
For 17th January, select an ensemble of data from a biased set of years

The weather generator model… (seasonal predictions)

(1) **Identify a subset of years from the historical record, such that the CDF from the selected years matches the CDF from the probabilistic forecast**

(2) Re-sample data from the subset of years nens times

(3) Re-order the ensembles to preserve observed inter-site correlations, observed temporal persistence, and observed correlations between variables Re-sample data from the historical record nens times

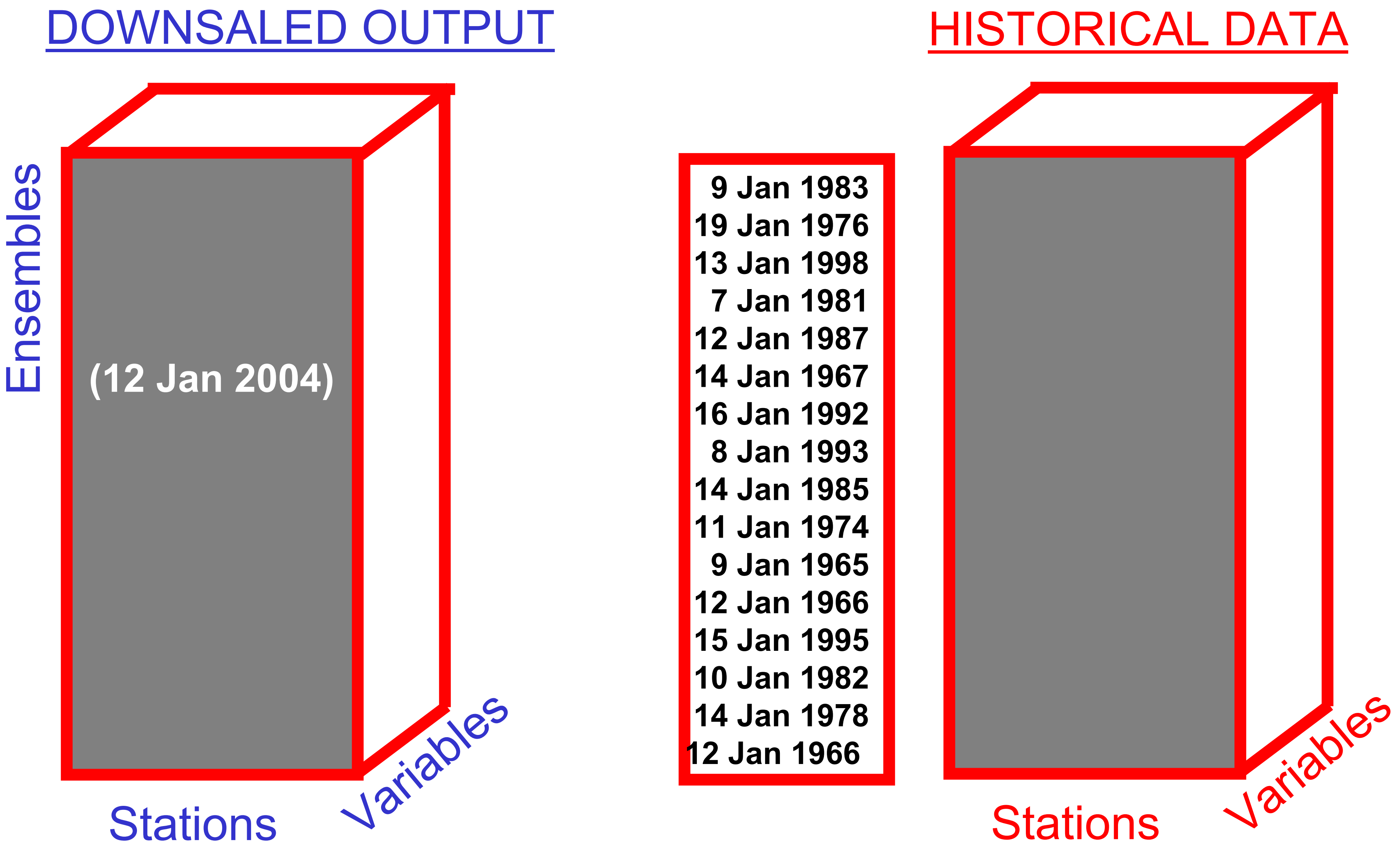


<u>For 17th January</u>, select an ensemble of data from a biased set of years

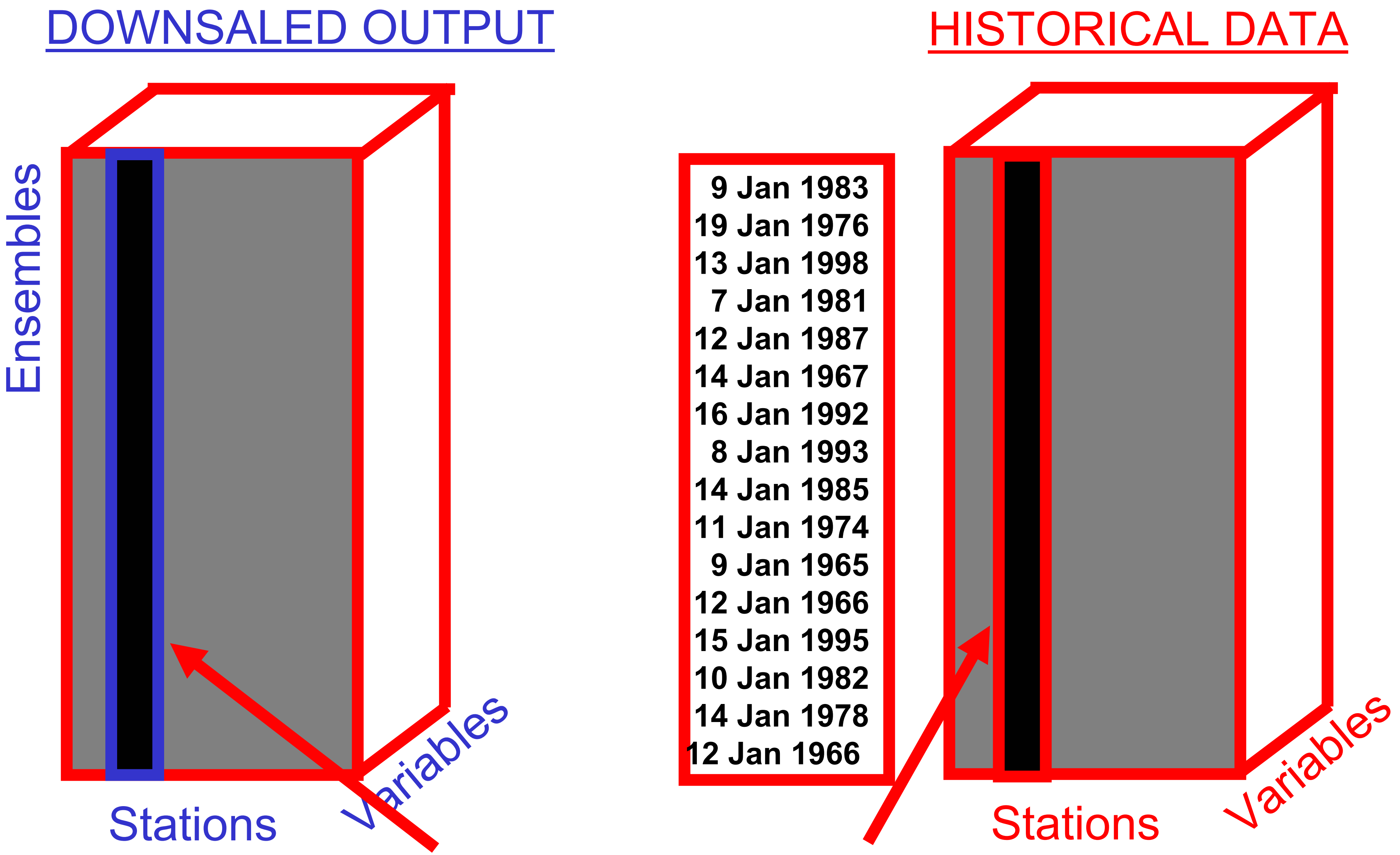
Schaake Shuffle

A method for reconstructing space-time variability in forecasted precipitation and temperature fields

The Schaake Shuffle



The Schaake Shuffle



SELECT VECTORS OF ENSEMBLES FROM THESE MATRICES

The Schaake Shuffle

ENSEMBLE OUTPUT (RANKED)

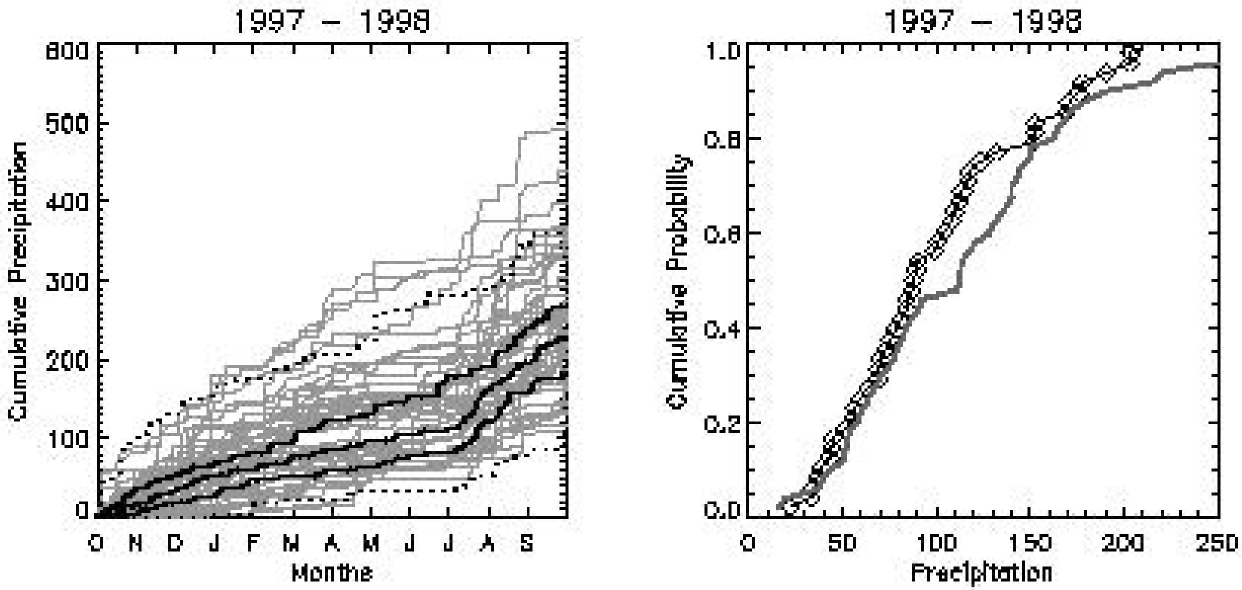
X =	23.4, 25.7, 26.3, 27.2, …, N	**(ACTUAL VALUES)**
r =	1, 2, 3, 4, …, N	**(RANKING)**

HISTORICAL DATA (RANKED)

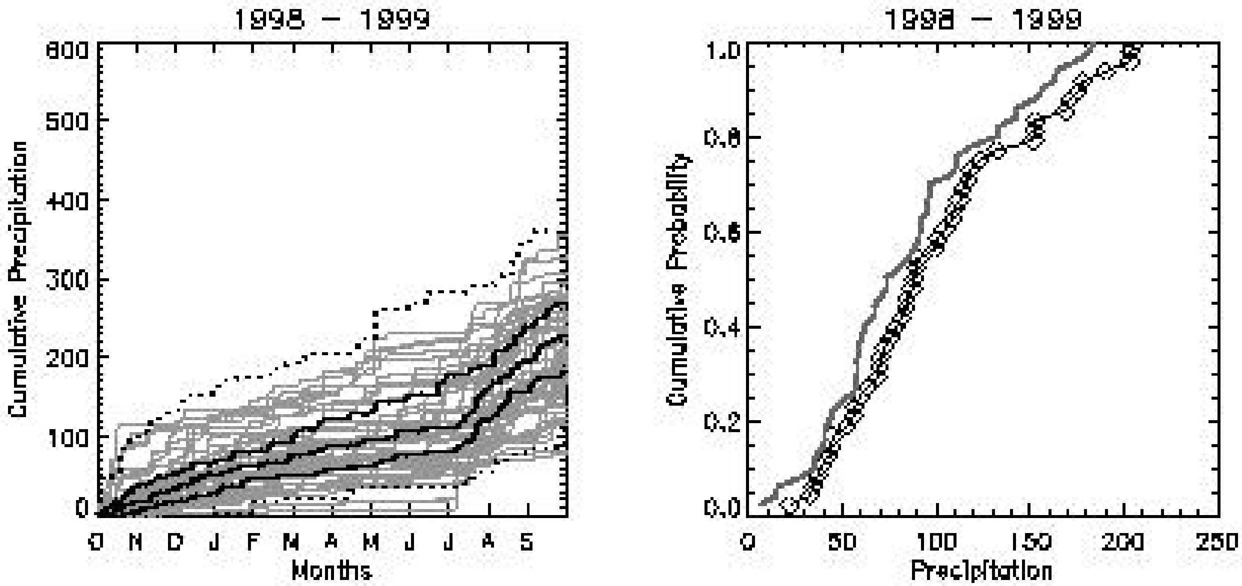
Y =	19.6, 21.2, 23.4, 24.1, …, N	**(ACTUAL VALUES)**
q =	97, 24, 38, 65, …, N	**(ENSEMBLE MEMBER)**

$x^{ss}(q) = x(r)$, $r=1,\ldots,N$ **(e.g., ens 97 is taken as the lowest value)**

Conditioning on CPC forecasts

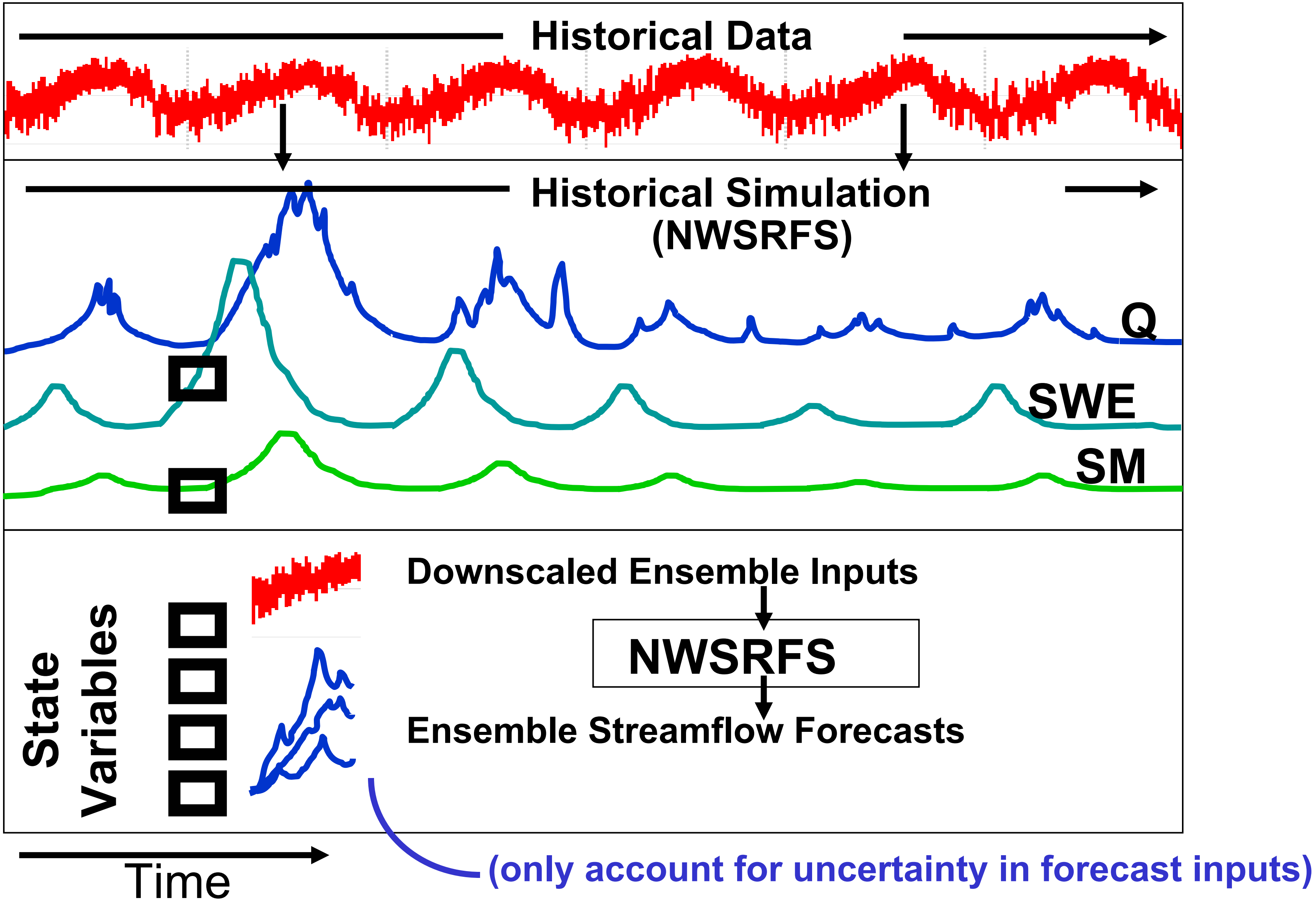


El Nino



La Nina

Model-based streamflow forecasting method…

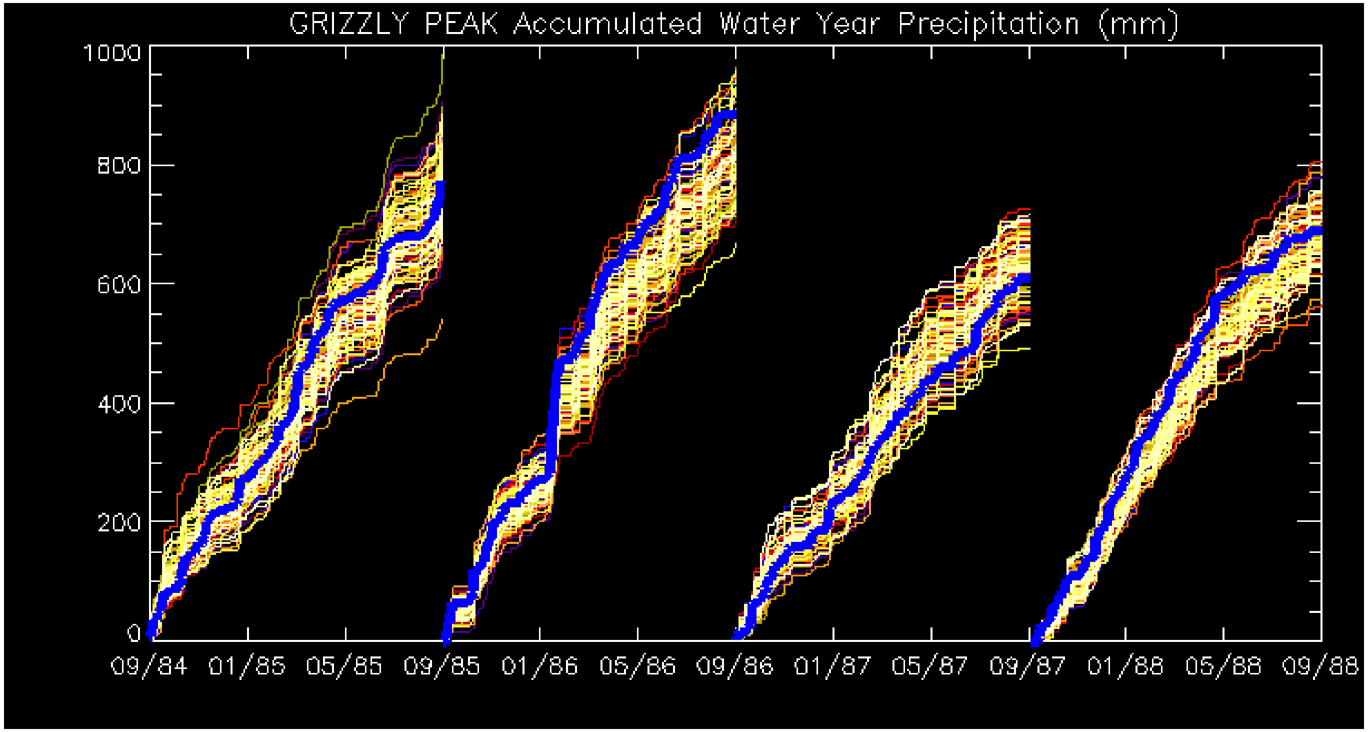


Uncertainty in basin initial conditions…

(1) Stochastic input forcings

- regression techniques used to estimate spatial fields of model forcings (precipitation, temperature)
- topographic characteristics (lat, lon, elev) used as predictiors; a different regression equation is developed for each day
- residuals in the regression equations are modeled stochastically to produce ensemble time series

Stochastic Forcings

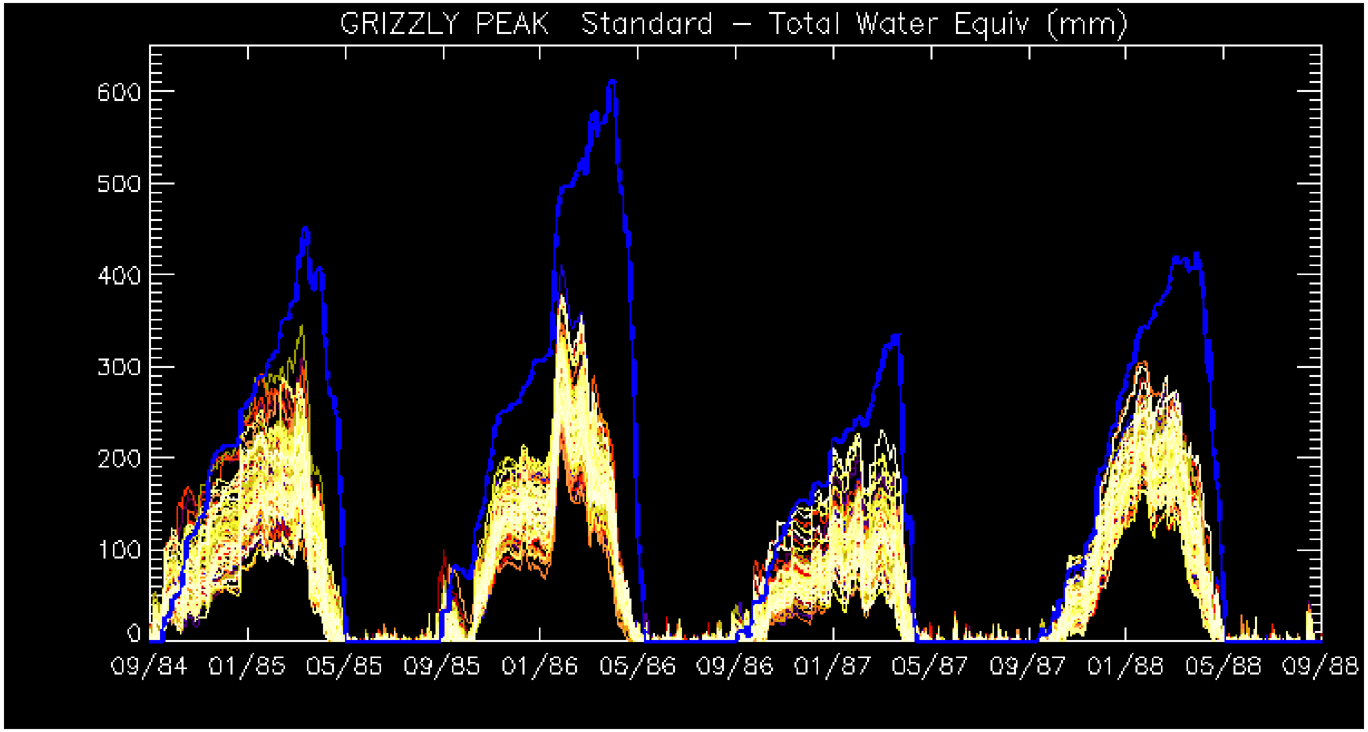


Uncertainty in basin initial conditions…

(1) Stochastic input forcings

- regression techniques used to estimate spatial fields of model forcings (precipitation, temperature)
- topographic characteristics (lat, lon, elev) used as predictiors; a different regression equation is developed for each day
- residuals in the regression equations are modeled stochastically to produce ensemble time series

Snow-17 Simulations

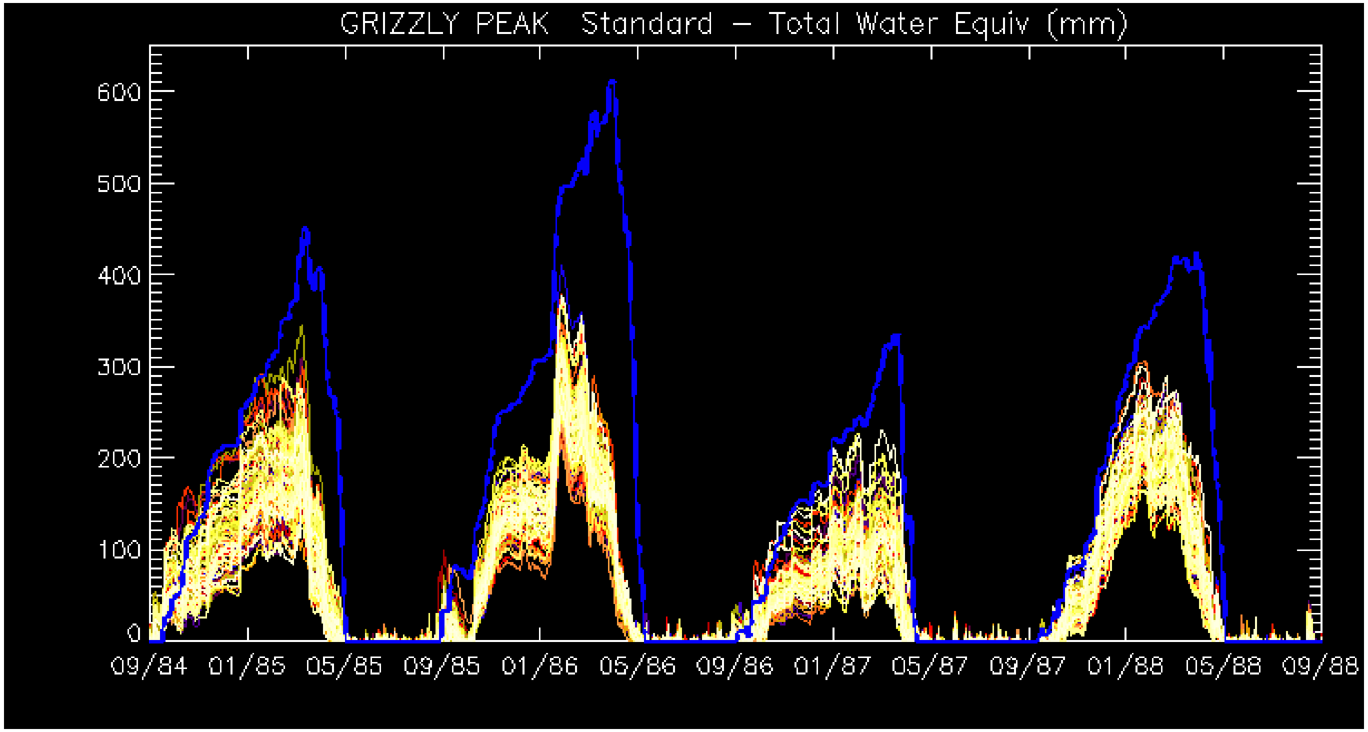


State Updating…

(1) Screened ensembles

- restrict attention to ensemble members that are closest to (the model equivalent of observations) at the start of the forecast period

Snow-17 Simulations



State Updating…

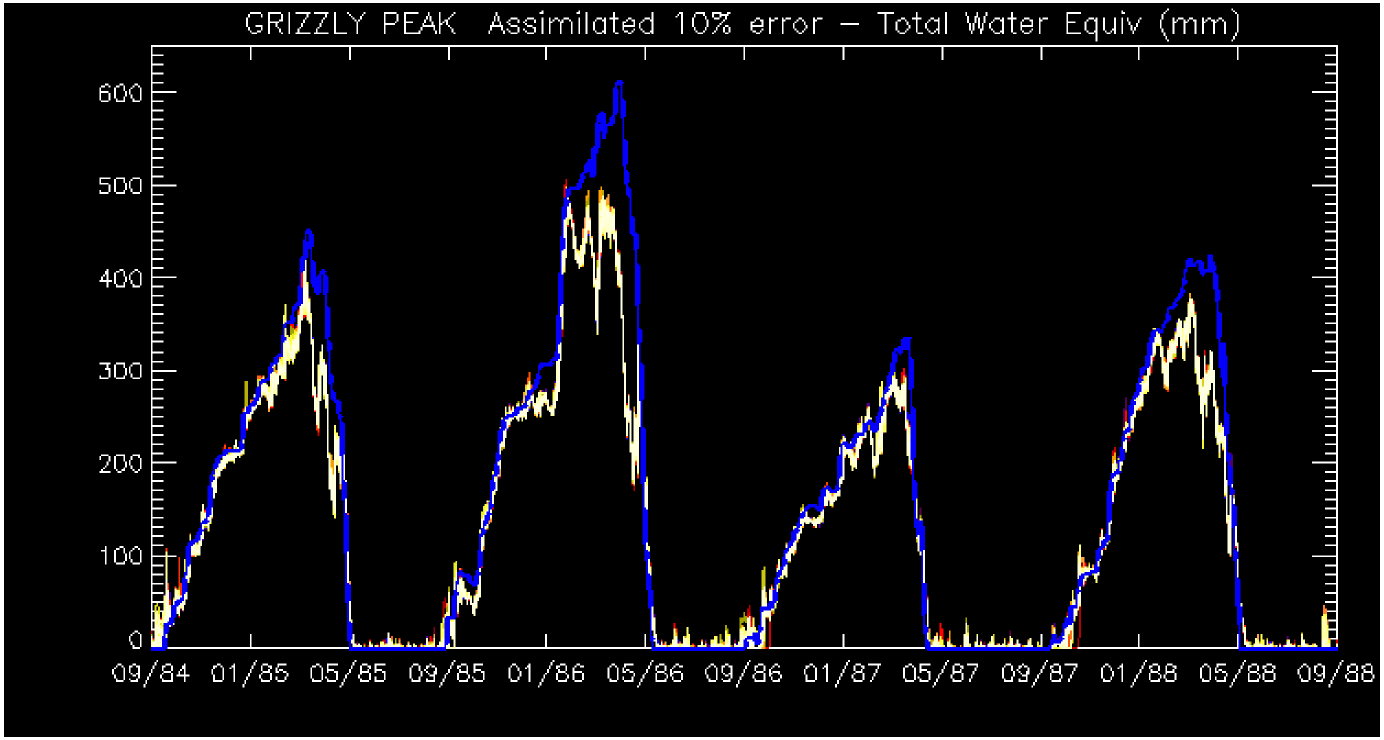
(1) Screened ensembles

- restrict attention to ensemble members that are closest to (the model equivalent of observations) at the start of the forecast period

(2) State updating

- Use of data assimilation methods (e.g., the ensemble Kalman filter) to update model estimates of snow water equivalent

Updating using EnKF



Model issues…

(1) Perturbed parameters

- development of methods to estimate parameter uncertainty, and use perturbed parameters to estimate uncertainty in basin initial conditions and model simulations of streamflow

(2) Model Structure / Complexity – (the Regional Reanalysis Conundrum)

- desire to match the complexity of the model to available data
- often do not have forcing data to use physically-based methods to simulate the land-surface energy balance
- Regional Reanalysis to the rescue—but model likely contains biases
- do not have data to evaluate model biases

- **research is needed to determine the model complexity that can be supported in light of the availability and quality of forcing data**

(3) Diagnosis of model errors

- evaluate model errors to understand which processes dominate in different river basins and which methods can be used effectively to improve streamflow forecasts.