Multi-Site Streamflow Forecast Framework

Application to the Upper Colorado River Basin

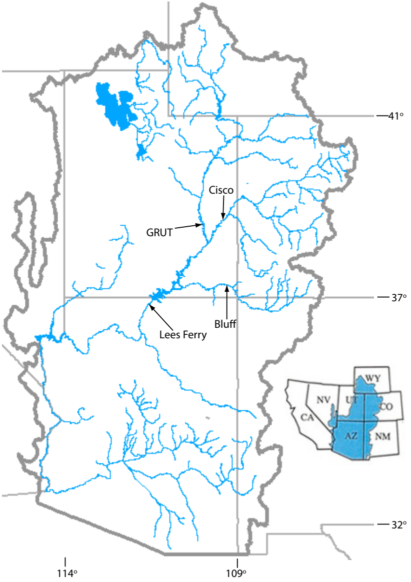
Cameron Bracken
Humboldt State University

Research Experience For Undergraduates
University of Colorado Boulder
Advisor: Balaji Rajagopalan

August 9, 2007

Outline

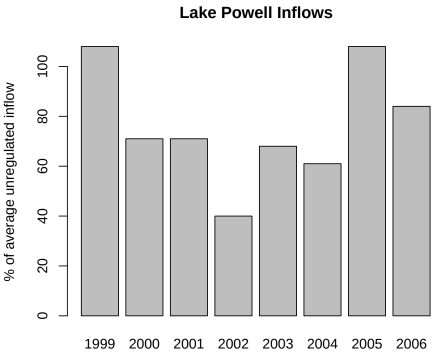
Study Area



- Parts of seven states with an area of 303,450 mi^2.
- Elevations ranging from 200 to 14,200 ft.
- Nearly 80% of the streamflow in the UCRB is due to snowmelt.
- 19 important gages in the Upper Colorado River Basin.
  - Lees Ferry and three main tributaries

Motivation

- The recent dry period (2000-2004) in the Upper Colorado River Basin (UCRB).
- Lowest levels in Lake Powell since 1980.
- Water managers need accurate forecasts at long lead times for several spatial locations

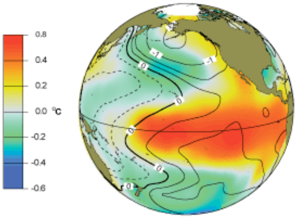


Current Modeling Approach

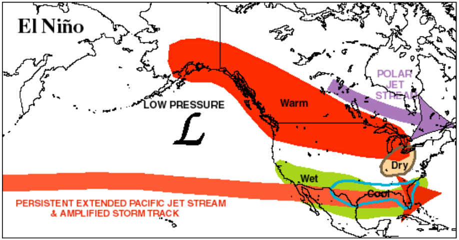
- The Colorado Basin River Forecast Center (CBRFC) makes streamflow forecasts at Fees Ferry.
  - The current model is primarily driven by snowpack data so forecasts cannot be made before January 1.
  - Number of predictions is limited by the number of years in the historical record.
- The CBRFC forecast is used in the Bureau of Reclamation's "24 month study".
  - No uncertainty is captured.
  - Does not use large-scale climate information.

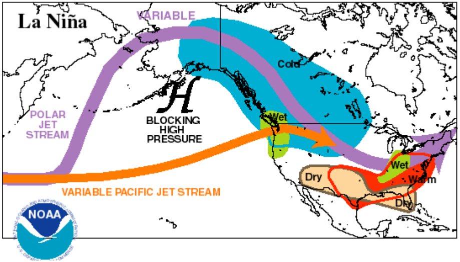
Large-Scale Climate Influence on Basin-Scale Hydrology.

How can skillful predictions be made in the earlier in the winter/spring season when snowpack data is unavailable or incomplete?
Large-scale climate variables can be used as predictors of peak season streamflow [Grantz et al. 2005] [Regonda et al. 2006].



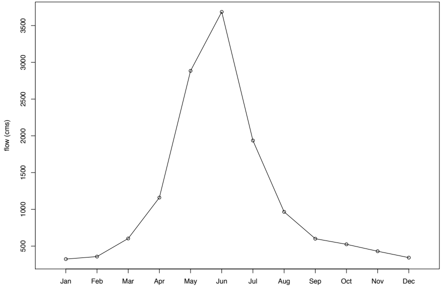
Warm Phase ENSO

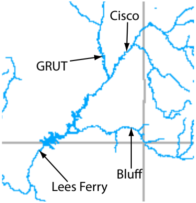




Multi-Site Considerations

- A river system is an interconnected network.
- Forecasts of multiple sites must be made simultaneously to preserve spatial dependencies.
- Solve this problem by creating an index or imaginary gage I.



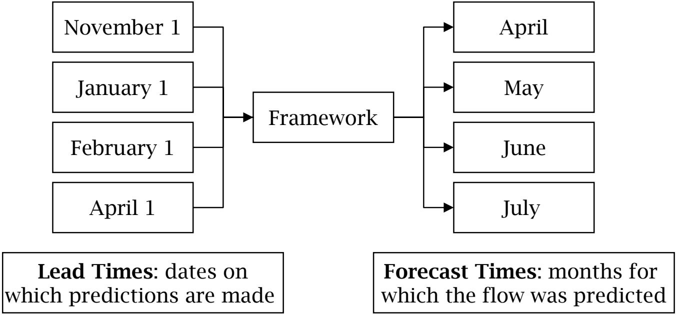


I gage=Lees Ferry + GRUT + Bluff + Cisco

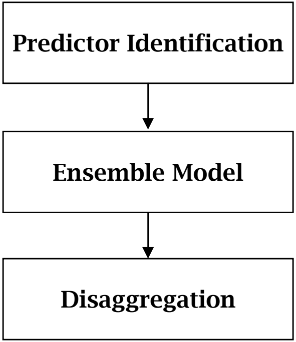
Overall Goal

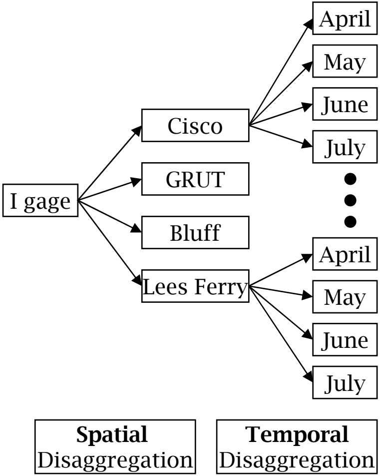
The overall goal of the study is to:

1. Simultaneously forecast many spatial locations while preserving spatial dependencies.



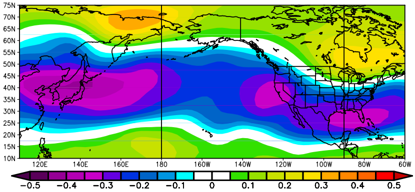
Framework


Predictor Identification
Ensemble Model
Disaggregation

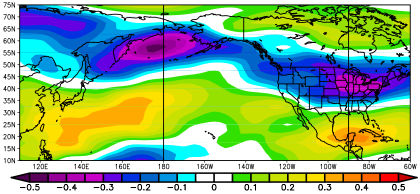

I gage
Cisco
GRUT
Bluff
Lees Ferry
April
May
June
July
April
May
June
July
Spatial Disaggregation
Temporal Disaggregation

Predictor Identification (April 1 Forecast)

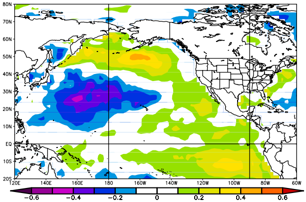
Spring season streamflow correlated with March climate data.



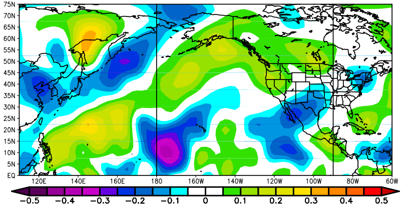
(a) 500 mb Geopotential Height



(b) 500 mb Zonal Winds



(c) Sea Surface Temperature

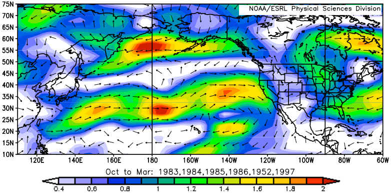


(d) 500 mb Meridonal Winds

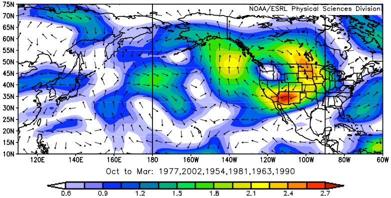
Other Predictors

- Snow Water Equivalent (SWE) is used as an indicator of snowpack. SWE data is on available for the February 1 and April 1 predictions.
- The Palmer Drought Severity Index (PDSI) is used as a surrogate for soil moisture content.

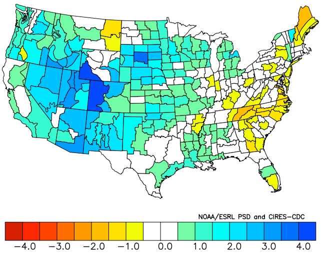
Composite Maps



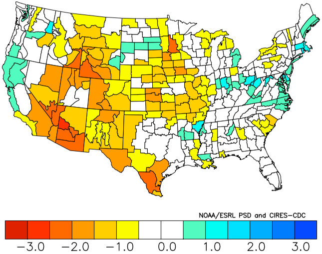
(a) Vector Winds (wet years)



(b) Vector Winds (dry years)



(c) PDSI (wet years)



(d) PDSI (dry years)

Multi-Model Forecasting

The model for each year t has the form

$$y_t = f(\mathbf{x}_t) + \varepsilon$$

where $\varepsilon \sim N(0, \sigma_t)$ and the matrix $\mathbf{x}_t$ is a vector of predictors. If the function f is global and linear then the model is traditional linear regression.
In this study we approximate f using a locally weighted polynomial (LWP) model. There are many advantages to LWP models which are well documented [Regonda et. al. 2006].

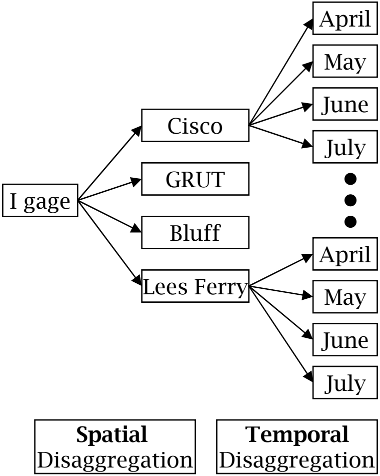
Multi-Model Ensemble Selection

Using an objective criteria, based on minimizing the Generalized Cross Validation statistic, models were determined for each lead time.

Lead Time	Predictors	PDSI	GPH	ZNW	MDW	SST	SWE
apr1	2	1	0	0	0	0	1
feb1	2	0	0	0	0	1	1
feb1	2	1	0	0	0	0	1
feb1	3	0	0	0	1	1	1
jan1	3	0	1	1	0	0	0
jan1	2	0	1	1	0	0	0
nov1	2	1	1	0	0	0	0
nov1	1	1	0	0	0	0	0

Table: Multimodels for each lead time. "1" indicates the presence of a predictor and "0" indicates the absence of a predictor.

The Disaggregation Problem



The disaggregation problem can be thought of as simulation from the conditional probability density function $f(\mathbf{x}|z)$ where $\mathbf{x}$ is a vector of flows and $\sum x_i = z$ [Prairie et al. 2007].

Results:Drop One Cross-Validation

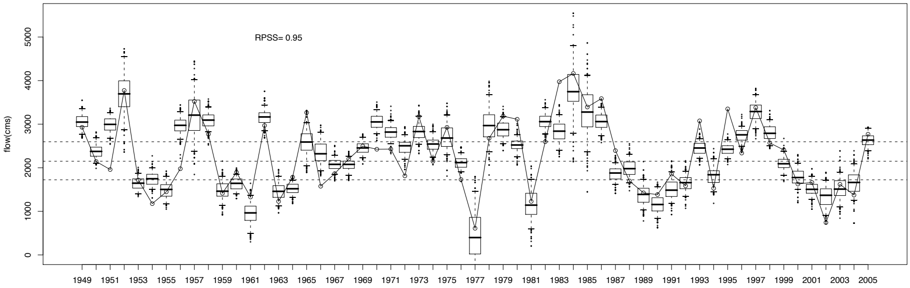


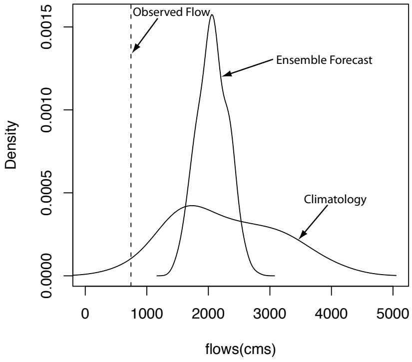
Figure: Drop One Cross-Validation forecast for the I gage on April 1.

Validation mode	apr1	feb1	jan1	nov1
Leave-one	0.95	0.85	0.65	0.28
Retroactive	0.80	0.85	0.29	0.10

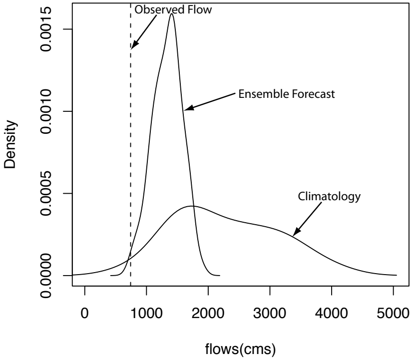
Table: I gage forecast skills

2002

2002 was the second lowest flow year in record. Though the model does not capture the magnitude of the flow, the predictions shift in relation to climatology from the November to April predictions.



(a) November 1



(b) April 1

Retroactive Forecast

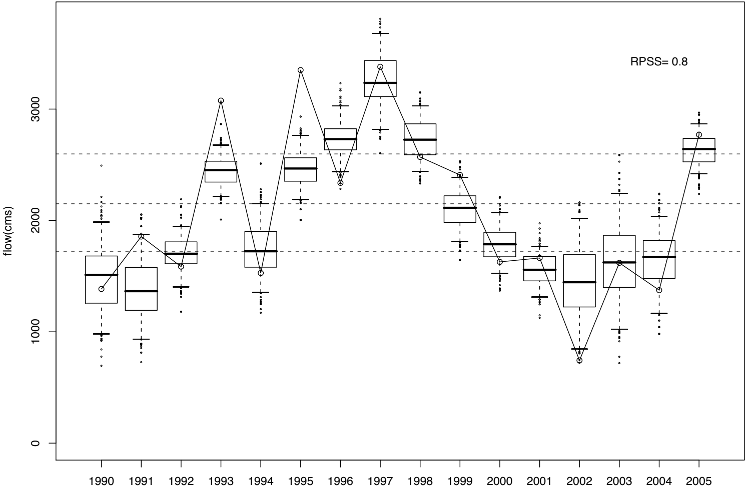


Figure: Retroactive forecast for the I gage on April 1.

Spatial Disaggregation Skills

Validation mode	Site	apr1 RPSS	feb1 RPSS	jan1 RPSS	nov1 RPSS
Retroactive	Cisco	0.97	0.73	0.62	0.78
Retroactive	GRUT	0.54	0.46	0.17	0.58
Retroactive	Bluff	0.37	0.40	0.15	0.15
Retroactive	Lees Ferry	0.87	0.73	0.62	0.60

Table: Disaggregated I gage forecast skill.

Spatial Disaggregation at Lees Ferry

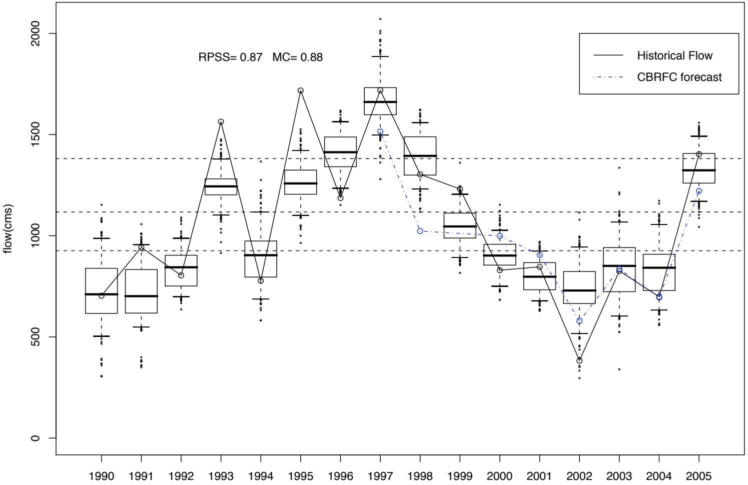


Figure: Retroactive forecast for Lees Ferry on April 1.

Spatial Disaggregation at Lees Ferry

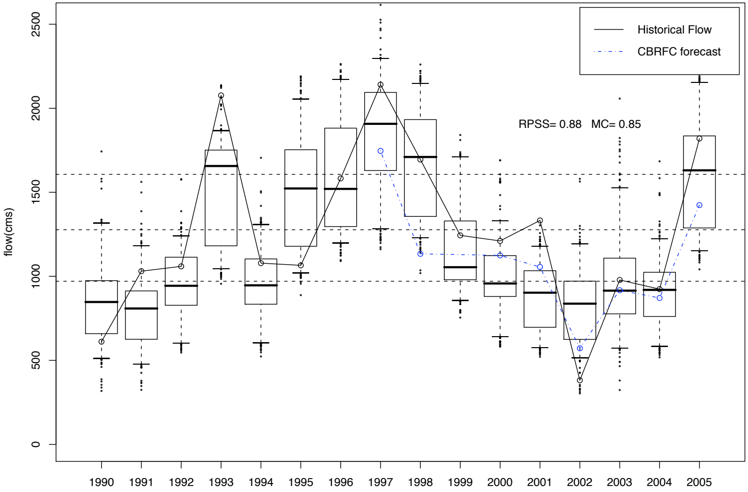


Figure: April 1 Retroactive forecast for May flows at Lees Ferry.

Conclusion and Further Reserach

- This multi-site framework is a simple and parsimonious method for streamflow prediction.
- Predictors need only be developed for one index gage.
- Disaggregation preserves spatial dependencies of sites on a river network.
- More than four sites could be incorporated easily by making a new Index gage.
- Forecasts with positive skill can be made as early as November 1.
- Forecasts were comparable to CBRFC forecasts thought not better all the time.
- It may be possible to combine both models in a Bayesian framework.
- To truly understand their usefulness, these long lead forecasts would need to be used to drive a decision model.

References

Katrina Grantz, Balaji Rajagopalan, Martyn Clark, and Edith Zagona. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, 41, 2005.

James Prairie, Balaji Rajagopalan, Upmanu Lall, and Terrance Fulp. A stochastic nonparametric technique for space-time disaggregation of streamflows. *Water Resources Research*, 2007.

Satish Kumar Regonda, Balaji Rajagopalan, Martyn Clark, and Edith Zagona. A multimodel ensemble forecast framework: Application to spring seasonal flows in the gunnison river basin. *Water Resources Research*, 42, 2006.

Disaggregation Algorithm I

1. The first step is to generate the orthonormal rotation matrix, $\mathbf{R}(d)$ using the Gram-Schmidt algorithm. Note that the $d \times d$ matrix $\mathbf{R}$ is only a function of the the dimension d and has the property $\mathbf{R}^T = \mathbf{R}^{-1}$.
2. Obtain the the matrix , $\mathbf{X}$, of historical flows and rotate into the matrix $\mathbf{Y}$ by
$$\mathbf{Y} = \mathbf{RX}$$
where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t)$, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t)$ and t is the number of years in the simulation. For example if disaggregating to four spatial location, $d = 4$ and $\mathbf{X}$ will contain flows for each site down the rows and each year across the columns. $\mathbf{Y}$ is now $(\mathbf{U}, \mathbf{z}d^{-0.5})^T$. Where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_t)$.
3. One simulated value, z_{sim}, from the MME to be disaggregated is selected.

Disaggregation Algorithm II

4. One $\mathbf{u}_i$ corresponding to year i is randomly selected from among the closest K years to z_{sim} based on the weight function

$$W(k) = \frac{1}{k \sum_{i=1}^{K} \frac{1}{i}} \quad \text{where } k = 1, 2, \ldots K.$$

5. The new vector $\mathbf{y}^*$ is constructed as $\mathbf{y}^* = (\mathbf{u}_i, z_{sim} d^{-0.5})$.
6. Transform back to the original space and generate the vector of disaggregated flows (which preserve additivity) by

$$\mathbf{x}^* = \mathbf{R}^T \mathbf{y}^*$$