Sources of Skill and Error in Long-Range Streamflow Forecasts for the Columbia River Basin:

A Comparison of the Role of Hydrologic State Variables and Winter Climate Forecasts

JISAO CSES Climate Impacts Group
and the Department of Civil and Environmental Engineering
University of Washington
April, 2003

Alan F. Hamlet
Dennis P. Lettenmaier
Pacific Decadal Oscillation

A history of the PDO

El Niño Southern Oscillation

A history of ENSO
In 5 out of 7 test years, accurate categorical ENSO forecasts (warm, neutral, cool) have been available in June preceding the water year. By October simple persistence gives an accurate forecast.

1998 ✔
1999 ✔
2000 ✔
2001 X
2002 ✔
2003 ✔
2004 X
Variable Infiltration Capacity (VIC) Model

Vicariance Infiltration Capacity (VIC)
Macroscopic Hydrologic Model

Cell Energy and Moisture Fluxes

Grid Cell Vegetation Coverage

Variable Infiltration Curve

$i = i_i [1 - (1 - A)^{10}]$

Point Infiltration Capacity, $i$

Fraction of Area

Baseflow Curve

$D_i = D_i^{m} = W_i W_2^{0.5} W_2$

Layer 2 Soil Moisture, $W_2$

Columbia River 1/8° degree resolution
Routing Flow Network
Schematic for Forecasting Experiments Using Resampled Observed Data

Select Temperature and Precipitation Data from Historic Record Associated with Forecast Climate Category

Run Initialized Hydrologic Model

Ensemble Streamflow Forecast

Climate Forecast

- ENSO
- PDO
Red line is long-term simulated climatological mean
On October 1
Initial soil moisture accounts for about 16% of the range of flows in the subsequent summer. For normal and high flow years the timing is not significantly altered.
In dry years, streamflow timing is also affected.
By April 1 basin snowpack is sufficient to produce an accurate summer streamflow forecast using simple regression techniques.
Definitions:

**Forecast Value:**
Economic (or intangible) value of the forecast in the context of some management decision process or other systematic response by a particular forecast user.

Metrics of this type are designed to show the value to a particular forecast user.

**Forecast Skill:**
Forecast performance as measured by a quantitative skill metric relative to some standard (e.g. climatology).

Skill metrics are designed to objectively compare different forecasts.
Red = observed  
Blue = ensemble mean

October 1 PDO/ENSO Forecast  

January 1 ESP forecast

1952
**A Simple Skill Metric:**

Take the central tendency of the forecast (ensemble mean) and compare to the long term mean of observations (climatology) using the MSE of the forecast relative to the observed value being forecast.

\[
\text{Skill}_1 = 1 - \frac{\text{MSE (ensemble mean)}}{\text{MSE (climatology)}}
\]

or when evaluating over the historic period:

\[
\text{Skill}_1 = 1 - \frac{\text{MSE (ensemble mean)}}{[\text{variance of observations}]}
\]

Where MSE is the mean squared error relative to the observed value being forecast.

Note that a skill of 1.0 is a perfect forecast, a skill of 0.0 is equivalent to climatology, and a negative value means that climatology is a better forecast.
What’s wrong with the simple metric in the previous example?

Metrics that use only the central tendency of each forecast pdf will fail to distinguish between red, green, and aqua forecasts, but will identify the purple forecast as inferior. **Example metric:** MSE of ensemble mean compared to MSE of long term mean of observations (variance of obs.)
A More Sophisticated Skill Metric:

In this case the MSE of the forecast relative to the observation is calculated for each of the ensemble members, and the average of these squared errors over all ensemble members is calculated. Note that in this case both the forecast and the climatology are treated as an ensembles as opposed to a single deterministic trace.

Skill_2 = 1 - \[ \frac{\sum (\text{forecast} - \text{obs})^2}{N} / \frac{\sum (\text{climatology} - \text{obs})^2}{M} \]

where N is the number of forecast ensemble members and M is the number of climatological observations.

The difference between this skill metric and the simple one is that this metric rewards accuracy, but also punishes spread. So a forecast with a tighter distribution and the same central tendency as the climatology will achieve a higher skill than climatology using this metric.
More sophisticated metrics that reward accuracy but punish spread will rank the forecast skill from highest to lowest as aqua, green, red, purple.

**Example metric:** average RMSE of ALL ensemble members compared to average RMSE of ALL climatological observations.
A Comparison of Forecast Skill between:

1) January 1 ESP forecasts (no climate forecast)

2) ENSO-Based Forecasts for October 1, November 1, and December 1 (hydrologic initial conditions plus climate forecast)
Oct
Skill = 0.115
1952

Nov
Skill = 0.146
1952

Dec
Skill = 0.323
1952

Jan ESP
Skill = 0.074
1952
Comparison of Skill For Warm ENSO Years

Skill Score

-1.00  -0.80  -0.60  -0.40  -0.20  0.00  0.20  0.40  0.60  0.80  1.00

OCT ENSO
NOV ENSO
DEC ENSO
JAN ESP

Skill Score

Annual Variation of Oct. ENSO, Nov. ENSO, and Jan. ENSO

- OCT ENSO
- NOV ENSO
- JAN ENSO

Comparison of Skill For Cool ENSO Years

Skill Score


Legend:
- OCT ENSO
- NOV ENSO
- DEC ENSO
- JAN ESP
Comparison of Skill For ENSO Neutral Years

Skill Score

OCT ENSO
NOV ENSO
DEC ENSO
JAN ESP

Winter Climate Forecasts Dominate

Hydrologic State Variables Dominate

June
December
March
Conclusions

PDO/ENSO forecasts are worth more than $100 million per year in the Columbia basin in terms of hydropower revenue alone (value of forecast), but an objective comparison with a Jan 1 ESP forecasts is useful in understanding skill and error characteristics in the context of water management.

Forecast skill metrics that reward accuracy and punish spread are more appropriate than simple metrics based on the central tendency of the forecasts when the forecasts have different variability than the climatology and there are relatively small shifts in the central tendency of the forecast from year to year.

Fall and early winter ENSO-based long-lead forecasts for the Columbia basin (based on resampling methods) typically have a lower skill than January 1 ESP forecasts based on persistence of hydrologic state (soil moisture and snow), but frequently have higher skill than climatology. Prior to December 1 the ENSO based forecasts are considerably less robust than the Jan 1 ESP forecasts. For ENSO neutral years the skill metrics are questionable due to the distribution of flows within the ensemble.

These results suggest that long-range streamflow forecasts based on long-range climate forecasts should be interpreted differently than existing statistical forecasts based on hydrologic state variables in order to account for the different error characteristics of the two forecast systems.