Ensemble Forecast Verification

Kristie J. Franz

Introduction

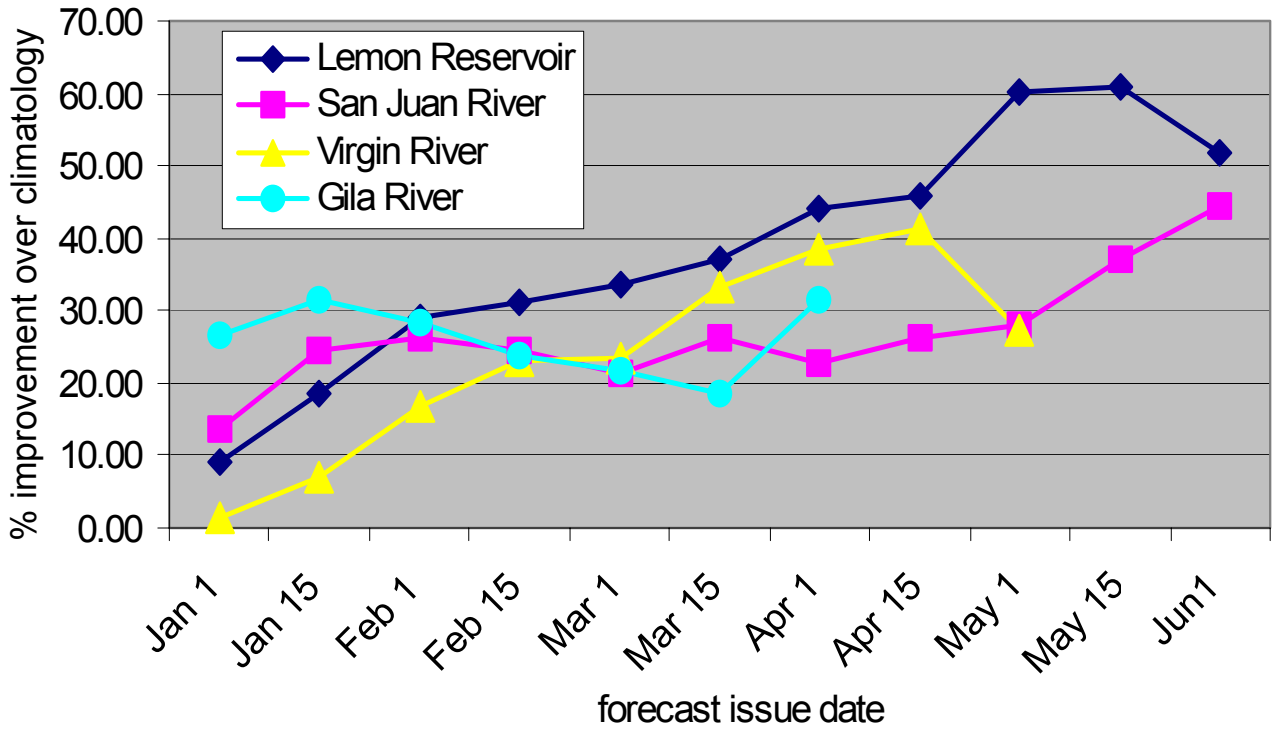
- **ESP water supply hindcasts were generated for 14 forecast points in Colorado River Basin.**
- **Several verification methods were tested.**
  (Franz et al., 2003)
- **The 3 distribution-based methods studied were found to give useful and detailed information about the forecasts performance.**

Verification Statistics Tested

1) Ranked Probability Score (RPS) and Skill Score (RPSS)

(Epstein, 1969; Wilks, 1995)

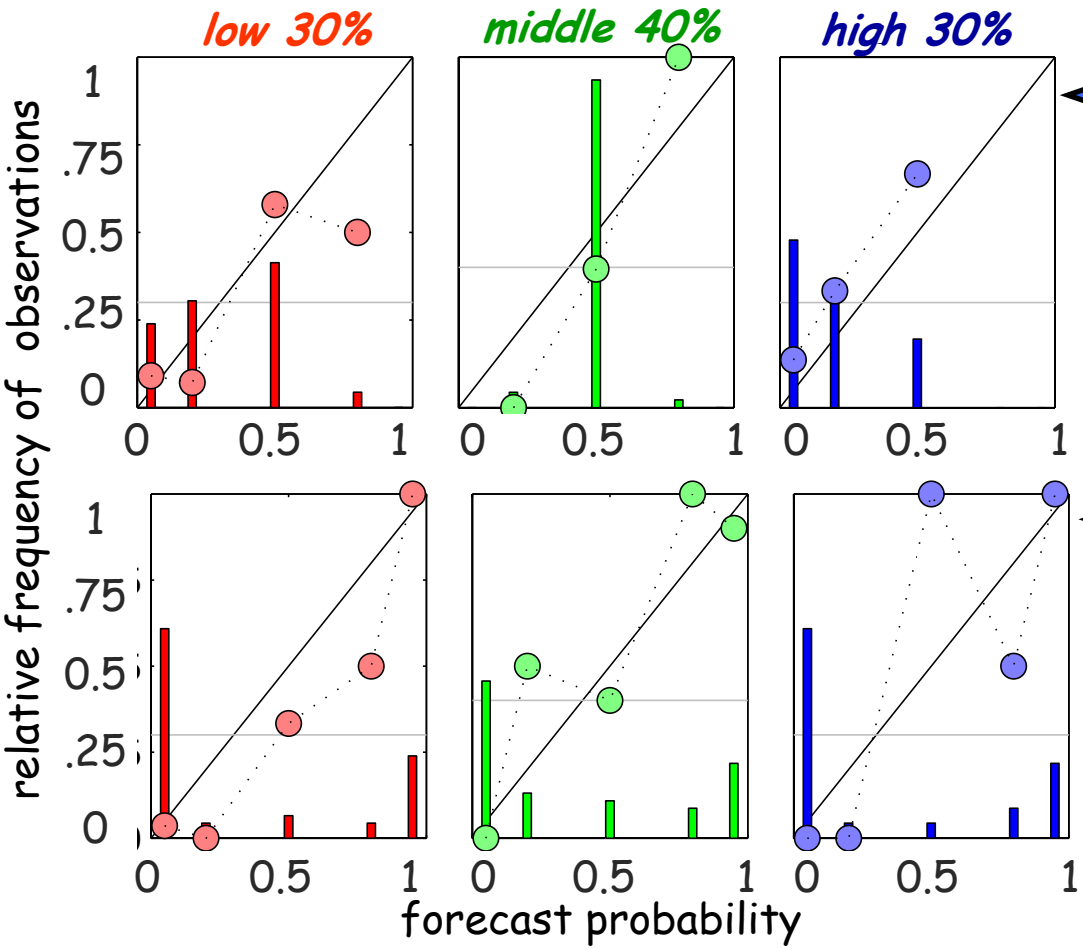
- single summary score of forecast accuracy
- considers the magnitude & distance between the observation & forecast probability
- RPSS gives relative skill of forecast compared to another (climatology is used below)



Verification Statistics Studied

2) Reliability (p(O|F)) (Murphy & Winkler, 1987,1992; Wilks, 1995)

- Determines whether flows occurred at the frequency at which they were forecast.



Jan 1 hindcasts for the Animas River are fairly reliable but have low confidence

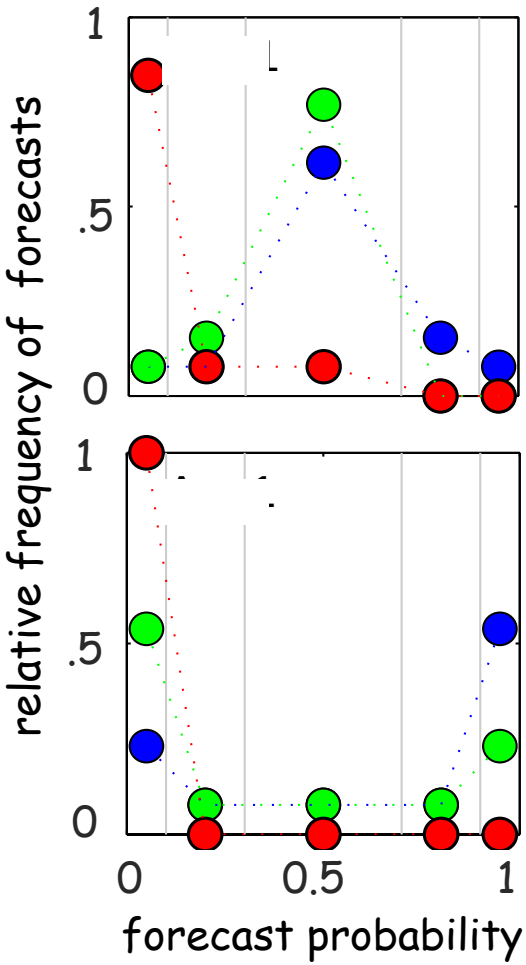
June 1 hindcasts show good reliability where forecast confidence is strong, but poor where sample sizes are low.

Verification Statistics Studied

3) Discrimination (p(F|O)) (Murphy & Winkler, 1987,1992; Wilks, 1995)

- Determines whether the forecasts predicted the flow that was observed.

Low (30%)
Mid (40%)
High (30%)

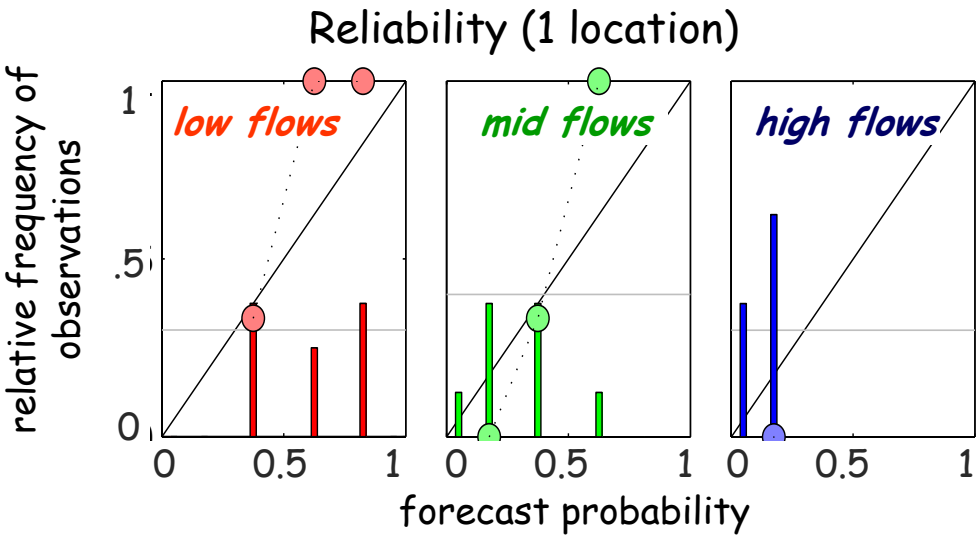


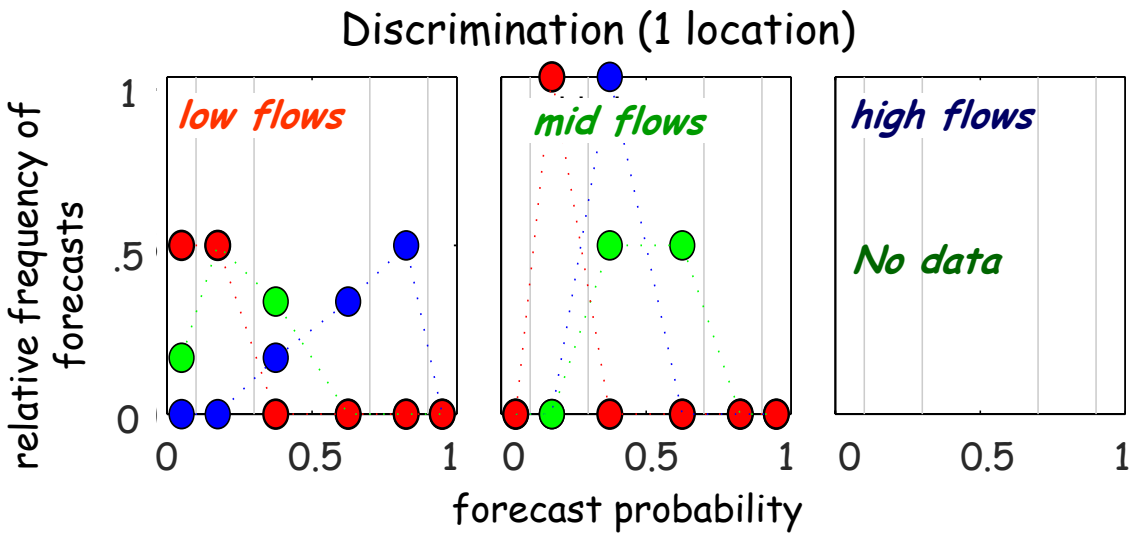
When high flows occurred for Gila River…

Jan 1 hindcasts indicated that low flows were least likely to occur, but did not discriminate between high and middle flows.

April 1 hindcasts showed improved discrimination but were not perfect (middle flows were occasionally forecasted with high probability even though they did not occur).

Testing methods for operational use





- All ESP forecasts and historical stage observations that were archived by the Ohio RFC were tested.
- Limited record of observed and forecast data is problematic. Poor forecast performance could be due to:
  - ✓ inaccurate forecasts
  - ✓ improper flow category identification due to poor observed record
  - ✓ inadequate forecast sample size

Low (25%)
Mid (50%)
High (25%)

What we have learned…

- Distribution-based verification methods are appropriate for ESP.
- RPS & RPSS are considered very useful from the forecaster's perspective (provide a single number, easy to understand and calculate).
- Discrimination & reliability are more complicated and may be more difficult implement (require large sample sizes and more involved interpretation).
- Obstacles to operational implementation:
  - ✓ Inadequate observational and forecast archives.
  - ✓ Understanding of the applicability of hindcasts for predicting operational forecast performance.
  - ✓ User education.
  - ✓ Methods may not be appropriate for short-term forecasts.
  - ✓ Interpreting statistics for run-time forecast modifications.

References

Franz, K.J., Hartmann, H., Sorooshian, S., Bales, R., 2003: Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado river basin. Journal of Hydrometeorology, 4 (6), 1105-1118.

Murphy, A.H. and Winkler, R.L., 1987: A general framework for forecast verification. Monthly Weather Review, 115, 1330-1338.

Murphy, A. H., and Winkler, R. L., 1992: Diagnostic verification of probability forecasts. International Journal of Forecasting, 7, 435-455.

Wilks, D.S., 1995: Forecast verification. Statistical Methods in the Atmospheric Sciences, Academic Press, 467 p.