Best Practices for Quantifying Improvements in Predictions

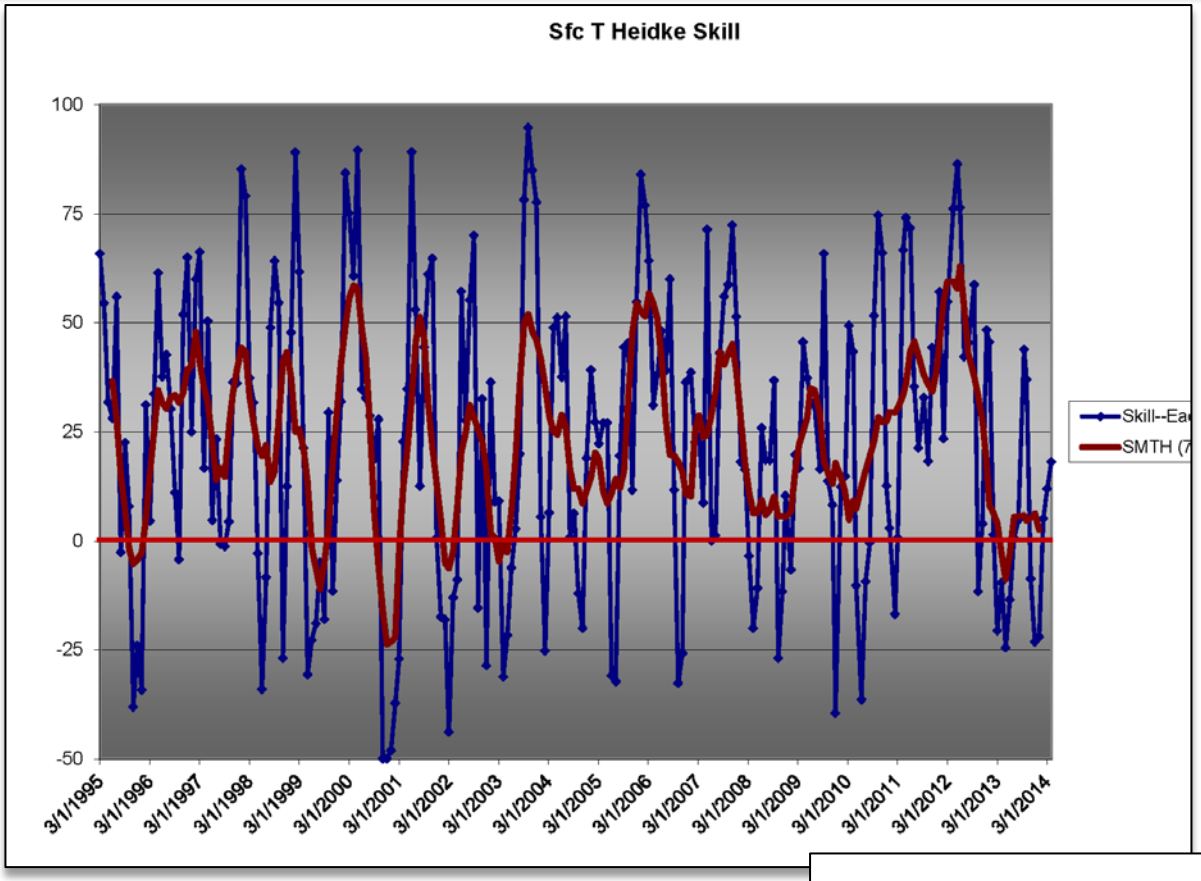
Arun Kumar

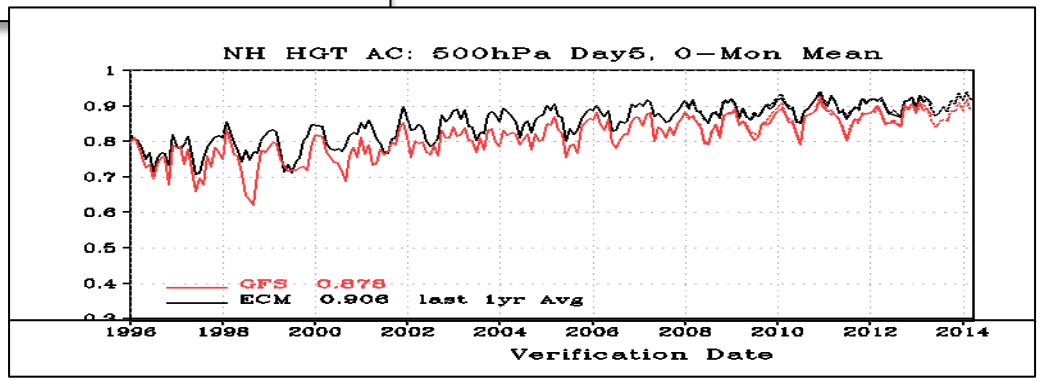
Climate Prediction Center

arun.kumar@noaa.gov

- Reason to quantify improvements in predictions due to our efforts in
  - Improving models and assimilation systems
  - Improvements in observational platforms
  - Gains in computing
  - Improvements in forecasting infrastructure (longer hindcasts; bias correction procedures; increase in ensemble size and resolution)

Skill of CPC's Operational Surface Temperature Forecast





Why Best Practices to Quantify Prediction Skill?

- Riddled with anecdotal information
- Quantify ROI of our efforts
- Compare different forecast systems on an equal footing
- Shed some information on predictability estimates

Various Aspects of Verification Process

- Verification datasets
- Different verification measures
- Forecast system related aspects (skill depends on ensemble size; length of verification period etc.)
- Context of Relevancy (in terms of what to verify)

Verification Datasets

- Multiplicity of verification data sets
- Verification data sets can make a difference in prediction skill
- Verifying against its own analysis generally enhances skill

Different Verification Methods

- An endless list of verification measures
- Deterministic, categorical, probabilistic; However, various verification measures are related
- Verification period
- Cross validation…Dangers of looking ahead in the assessment of prediction skill

nature geoscience

Retracted

LETTERS

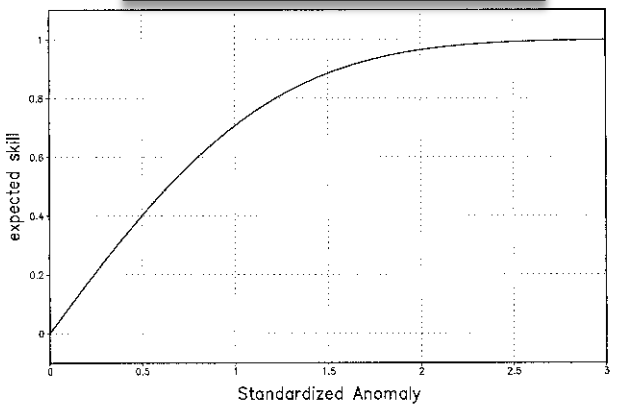
PUBLISHED ONLINE: 23 MARCH 2014 | DOI: 10.1038/NGEO2118

One-year-out cross-validated hindcasts instead require cross-validation of the MCA pattern-generation in the year out, thus performing an MCA on the remaining years. Following this approach, the cross-validated skill in hindcasting the winter NAO index using September SIC over the whole Arctic is 0.08, indicating that there is no predictive skill from Arctic sea-ice variability. The cross-validated NAO skill

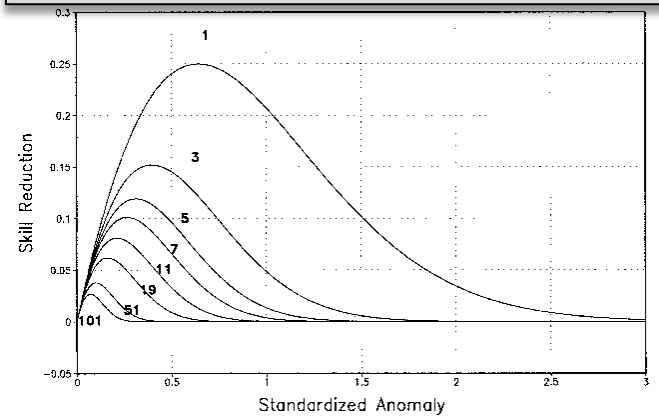
Model Related Aspects

- Ensemble size
- Bias correction procedures prior to verification
  - Look ahead
  - Lead time dependent bias correction

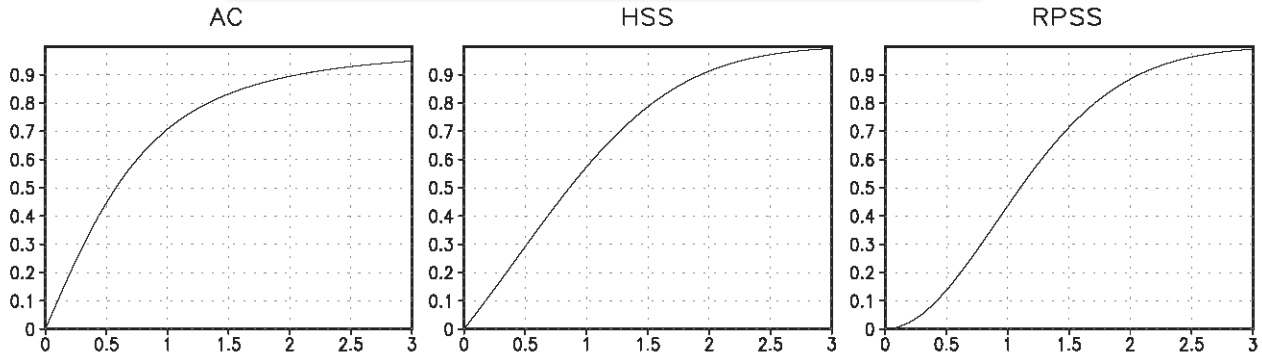
AC and SN Ratio



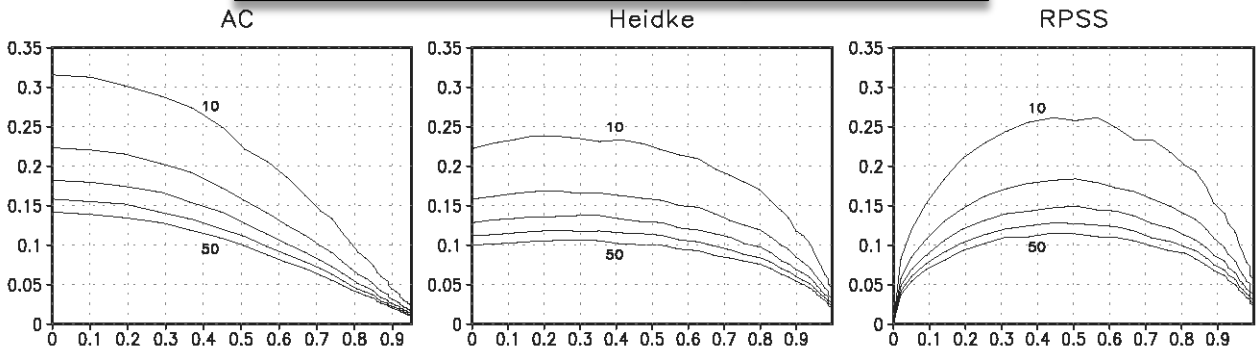
Gain in Skill with Ensemble Size



Skill Measures and SN Ratio



Error Bars and Sample Size

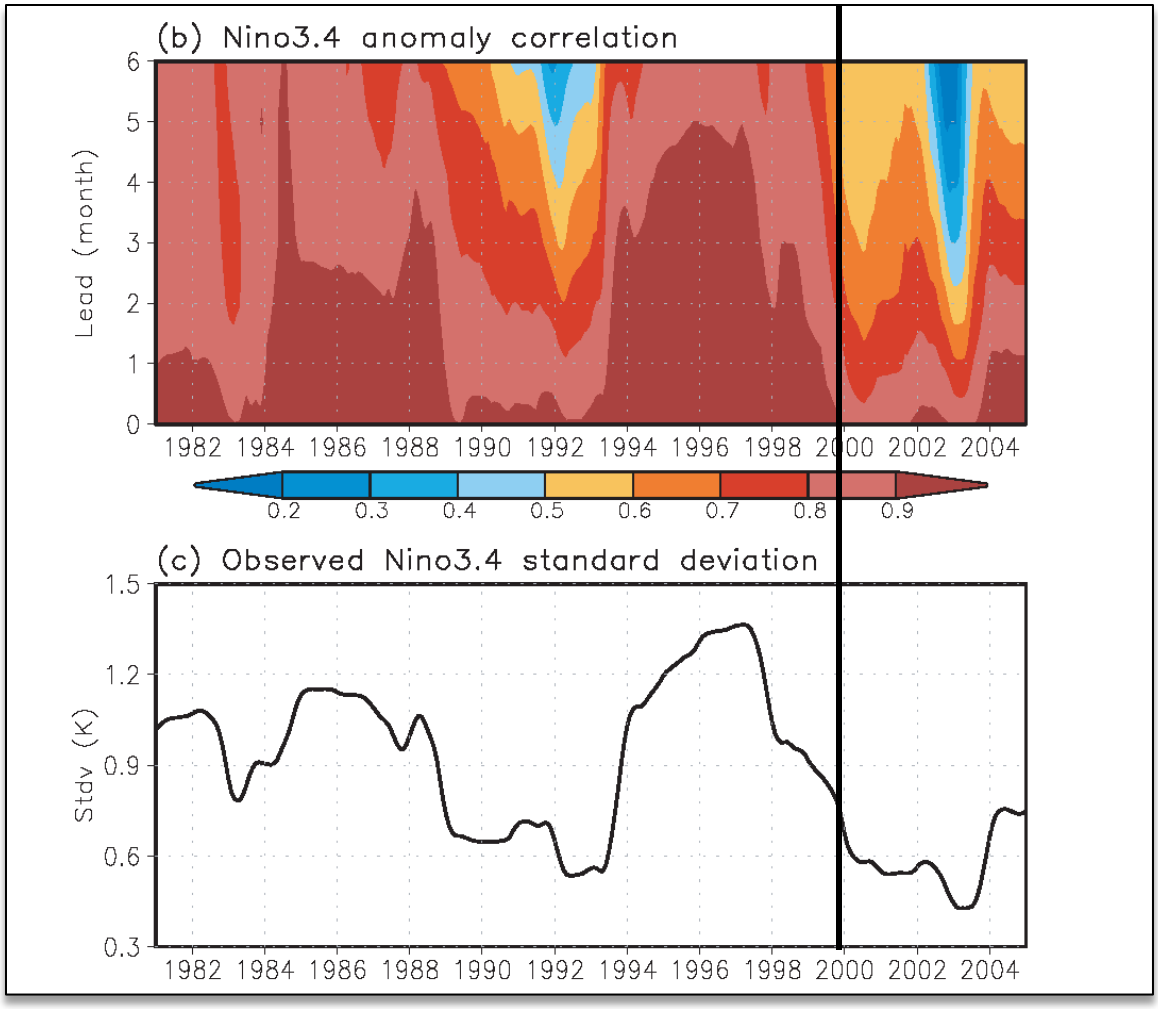


Recommendations

- Use a fix hindcast period over which to verify
- Recommend independent verification data sets
- Clearly state the ensemble size
- Quantify error bars (e.g., Monte Carlo approach)

Caveats to be Aware About

- LF variability (pessimistic or optimistic assessment)
- Due to changes in observational systems average prediction skill based on a prior period may not be representative of real-time skill


(b) Nino3.4 anomaly correlation
Lead (month)
0
1
2
3
4
5
6
1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
(c) Observed Nino3.4 standard deviation
Stdv (K)
0.3
0.6
0.9
1.2
1.5
1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004