Toward Short-Range Ensemble Prediction of Mesoscale Forecast Skill

Eric P. Grimit and Clifford F. Mass
University of Washington
Big Questions

- How much confidence does one have in today’s short-range ensemble forecast? Can the skill be predicted?

- How does one define skill, anyway?

- Is there a single metric of the ensemble forecast distribution that is useful for this purpose?
The Traditional Spread/Skill Relationship

Spread/Error Correlation Theory
(Houtekamer 1993; Whitaker and Loughe 1998)

\[ \rho^2(\sigma, |E|) = \frac{2}{\pi} \frac{1 - \exp(-\beta^2)}{1 - \frac{2}{\pi} \exp(-\beta^2)} \ ; \beta = \text{std}(\ln \sigma) \]

- Spread/error correlation depends on the time variation of spread
- For constant spread ($\beta = 0$) $\rho = 0$
- For large spread variability ($\beta \to \infty$) $\rho \to \sqrt{2/\pi} < 0.8$
- Spread is a more useful predictor of error when it is extreme (large or small relative to its median)
- Assumes that E is the ensemble mean error
**Simple Idealized Spread/Skill Model**

1. Draw today’s “forecast uncertainty” from a log-normal distribution (Houtekamer 1993 model).
   \[ \ln(\sigma) \sim N(\ln(\sigma_m), \beta) \]

2. Create synthetic ensemble forecasts by drawing M values from the “true” distribution (perfect ensemble).
   \[ F_i \sim N(Z_m, \sigma) \; ; \; i = 1, 2, \ldots, M \]

3. Draw the verifying observation from the same “true” distribution.
   \[ V \sim N(Z_m, \sigma) \]

4. Calculate ensemble spread and skill using varying metrics.
Simple Idealized Spread/Skill Model

- Assumed Gaussian statistics
- Varied:
  1) spread variability
  2) finite ensemble size
  3) spread and skill metrics

- Stochastically simulated ensemble forecasts at a single grid point with 10,000 realizations (cases)
- Perfect model (verification is drawn from the same distribution)

STD = Standard Deviation

MAE = ensemble Mean Absolute Error

Correlation increases with spread variability and ensemble size
What Measure of Skill?

- STD is a better predictor of the average ensemble member error than of the ensemble mean error.

  \[
  \text{MAE} = | \overline{E} | \\
  \text{AEM} = | E |
  \]

- Different measures of variation in the ensemble may be required for other measures of skill.
  For example: categorization into climatological bins.
Empirical Spread/Skill Correlations

- Early MM5 Ensemble Results (2000)
- Expanded MM5 Ensemble Results
  Analysis-Centroid Mirroring Ensemble (ACME)
  Cool season 2001-2002
  Cool season 2002-2003
- Temporal Ensemble Results
Initial Results

- Spread/error correlations (0.4-0.6) suggested that mesoscale errors could be predicted a priori.
- Following Houtekamer (1993) and Whitaker and Loughe (1998), spread/error correlations increased (0.6-0.8) when only cases with high and low spread were considered.

(see Grimit and Mass 2002)

Ensemble Size = 5 members
   (AVN, CMC, ETA, NGM, NOGAPS)

Verification Period: Jan - Jun 2000
   (102 cases)

Verification Strategy: Interpolate Model to Observations

Variable: 10-m Wind Direction

Spread = Standard Deviation
Error = Mean Absolute Error  (domain averaged)
ACME\textsuperscript{core} Results

- Spread/error correlations using MAE are somewhat lower than initial results.
- As in the simple model, STD is a better predictor of the AEM error, than MAE.
- Spread/error correlations improve when only cases with high and low spread are considered, as before.
- Not as much separation between extreme and non-extreme spread after F24.
- Average errors by spread category still have good separation.

**Ensemble Size = 7 members**

(AVN, CMC, ETA, GASP, NOGAPS, TCWB, UKMO)

**Verification Period:** Dec 2001 – Mar 2002

(71 cases)

**Verification Strategy:** Interpolate Model to Observations

**Variable:** 10-m Wind Direction

**Spread = Standard Deviation**

**Error = Average Ensemble Member error**
**ACMEcore Results**

- Spread/error correlations slightly lower than before, except much lower from F06-F15.
- More noticeable diurnal cycle.
- Results using MAE match up very well with H93 theory. “Cleaner” results, overall.
- Now excellent separation between extreme and non-extreme spread cases.
- Average errors by spread category still have good separation, although somewhat less.

**Ensemble Size = 8 members**  
(AVN, CMC, ETA, GASP, JMA, NOGAPS, TCWB, UKMO)

**Verification Period:** Nov 2002 – Feb 2003  
(105 cases)

**Verification Strategy:** Interpolate Model to Observations

**Variable:** 10-m Wind Direction

**Spread = Standard Deviation**  
**Error = Average Ensemble Member error**
Correlation and Ensemble Size

ACME 12-KM SPREAD-ERROR CORRELATION — ALL SURFACE / WIND DIR

Correlation Coefficient

Lead Time (hrs)
**Temporal (Lagged) Ensemble**

[Hoffman and Kalnay 1983; Reed et al. 1998; Roebber 1990]

- Roebber (1990) found weak correlations (r < 0.5) between lagged forecast spread and lagged forecast mean absolute error for SLP
- Did not look for correlation between lagged forecast spread and current forecast skill

\[ T - 2.0 \]

\[
\begin{array}{c}
\text{CENT-MM5} \\
\text{CENT-MM5} \\
\text{CENT-MM5} \\
\text{CENT-MM5}
\end{array}
\]

\[ \downarrow \]

\[
\begin{array}{c}
\text{F48} \\
\text{F36} \\
\text{F24} \\
\text{F12}
\end{array}
\]

\[ T - 1.0 \]

Using lagged ACME Centroid Forecasts

- Yields mesoscale spread
- Less sensitive to a single synoptic-scale model’s time variability
- Best forecast estimate of “truth” (without any prior post-processing)

\[ M = 4 \]

[equal weighting of lagged forecasts]
Temporal Spread/Error Correlations

- Lagged CENT-MM5 ensemble spread has moderate to strong correlation ($r = 0.7-0.8$) with the lagged CENT-MM5 ensemble mean skill (not shown).
- Weaker correlation with current mean skill, but is still a useful secondary predictor.

Relatively weak correlation with current ensemble mean skill.
Summary and Extensions

Prediction of forecast skill depends largely on the definition of skill itself.

- User-dependent needs

For near-surface wind direction, ACME$^{\text{core}}$ variance (STD) is a good predictor of both average ensemble member error (AEM) and ensemble mean error.

The variance of a temporal ensemble (using lagged centroid forecasts from ACME) is a useful second predictor of errors.

Current and future work:

- Grid-based verification with RUC20
- Use a simple bias removal (as in Eckel talk)
- Analyze ACME skill prediction of other weather parameters
- Analyze ACME skill prediction using binned quantities
“No forecast is complete without a forecast of forecast skill!”
-- H. Tennekes, 1987

QUESTIONS?