Redefining the Ensemble Spread-Skill Relationship from a Probabilistic Perspective

Eric P. Grimit and Clifford F. Mass

Department of Atmospheric Sciences
University of Washington, Seattle, WA USA

1. The Ensemble Spread-Skill Relationship

- Based on the premise that ensemble spread should be related to forecast error.
- Often characterized by the correlation between the standard deviation of the ensemble forecasts and the absolute error of the ensemble mean forecast – the spread-error correlation.
- Actual ensemble spread-error correlations have been disappointing, often less than 0.5, especially for short-range ensemble forecasts (SREF) of near-surface weather variables.
- Larger spread-error correlations (0.6-0.7) have been observed with the spread of spatial averaging, forecast bias correction, and considering only extreme spread cases (Grimit and Mass 2012; Stensrud and Yussouf 2003).
- However, correlations of spatial averages can be misleading (ecological correlation) and the practical use of spatial mean error forecasts is suspect.
- Simple, statistical arguments (Houtekamer 1993) show that ensemble spread-error correlation is a function of the day-to-day ensemble spread variability (σ) and thus has an upper limit near 0.8.

A couple studies show that categorical measures of forecast spread (e.g., statistical entropy or mode population) are better than continuous measures (e.g., variance) at discriminating forecast success from failures (Toth et al. 2001; Zehrman 2001).

- A probabilistic approach that uses the full forecast error distribution is warranted.
- Probabilistic error forecasts can be used to evaluate the skill of ensemble forecasts.

4. UW Short-Range Ensemble Forecast Data

- The U.S. Mesoscale Ensemble (UME),
- A multimodel-based short-range ensemble forecast (SREF) system for 24-hour forecasts of high-impact weather events.
- Forecasting using MM5 at 12- and 36-km horizontal resolutions and all models at 3-hourly intervals.
- A parallel, eight-member ensemble that includes physics diversity (UME+).

- Evaluation Period:
  - 361 cases (27 OCT 2003 – 31 MAR 2004)
  - 271 cool season (Oct.-Mar.) cases
  - 90 warm season (Apr.–Sept.) cases
  - All forecasts initialized at 0000 UTC
  - All observations initialized at 1200 UTC

- Parameters of Forecast:
  - Assumed 925 hPa
  - 72-hour period
  - 10-m Wind Speed (WIND 10) (U, V Components)
  - 10-m Wind Direction (WIND D) (30-Min Avg)

- Verification Data:
  - NCEP Radar Update Cycle 3 (NRC3)
  - 12 km EnSRF (Ensemble-based Reanalysis, NCEP)
  - 12 km EnSRF (Ensemble-based Reanalysis, NCEP)

- Verification of Model:
  - 361 cases, US Mesoscale Ensemble (UME)
  - 271 cases, cool season
  - 90 cases, warm season

2. A Probabilistic Perspective

- The ensemble spread should be related to forecast error.
- Using a statistical model of the ensemble (Grimit and Mass 2012) to generate large numbers of idealized ensemble forecasts, it is shown that ensemble spread (σ) is related to forecast error (ε).
- While small spread guarantees a small error, large spread guarantees a large error.
- Spread-based conditional error climatologies (CECs) are proposed to form idealized ensemble forecasts.

Assumed:
- The predictive skill of spread-based conditional error climatologies can be related to statistical consistency.

3. Idealized Forecast Error Prediction

- A statistical model of spread-skill relationship from a probabilistic perspective is warranted.
- Modified Statistical Model of Spread-Skill (MSM) and SREP (Stensrud and Yussouf 2003)
- The predictive skill of spread-based conditional error climatologies can be related to statistical consistency.

Worked Example of Spread-Skill Relationship:
- Draw today’s “forecast uncertainty” from a log-normal distribution conditioned by ensemble spread category, form the probabilistic forecast error predictions.

5. Real Forecast Error Prediction

- What is the best performance we can expect given a perfect probabilistic error forecast?
- The spread-skill relationship can be redefined as probabilistic error forecasts.
- A couple studies show that categorical measures of forecast spread (e.g., statistical entropy or mode population) are better than continuous measures (e.g., variance) at discriminating forecast success from failures (Toth et al. 2001; Zehrman 2001).
- The predictive skill of spread-based conditional error climatologies can be related to statistical consistency.

6. Summary and Conclusions

- Does the spread of mesoscale short-range ensemble forecasts contain reliable information from which the expected errors in near-surface weather forecasts can be estimated accurately?
- Several recent studies have used real-time diagnostics of model performance, e.g., verification of high impact events with SREF forecasts.
- Using a log-normal model, it is shown that ensemble spread variability can be related to statistical consistency.
- A probabilistic approach that uses the full forecast error distribution is warranted.
- Probabilistic error forecasts can be used to evaluate the skill of ensemble forecasts.

- The predictive skill of spread-based conditional error climatologies (CECs) is proposed as a robust measure of the spread-skill relationship.
- Poor performance of spread-based CEC skill is nearly equal to the spread-skill relationship by reducing the temporal spread variability (not shown).
- The spread-skill relationship can be redefined as probabilistic error forecasts.