Errors in Error Variance Prediction and Ensemble Post-Processing

Elizabeth Satterfield[1], Craig Bishop[2]

[1]National Research Council, Washington, DC, USA; [2]Naval Research Laboratory, Monterey CA, USA

Ensemble Forecasts

- Accurate predictions of forecast error variance are vital to ensemble weather forecasting.
- The variance of the ensemble provides a prediction of the error variance of the ensemble mean or, possibly, a high resolution forecast.

How Should Ensemble Forecasts Be Post-Processed?

- Adjust a raw ensemble variance prediction to a new value that is more consistent with historical data?
- Must one attempt to describe the distribution of forecast error variance given an imperfect ensemble prediction?

A Simple Model of Innovation Variance Prediction

- The "true state" is defined by a random draw from a climatological Gaussian distribution.

$$x^t \sim N\left(0, \sigma_C^2\right)$$

- The error of the deterministic forecast is defined by a random draw from a Gaussian distribution whose variance is a random draw from a climatological inverse Gamma distribution of error variances.

$$\overline{x^f} = x^t + \varepsilon^f, \text{ where } \varepsilon^f \sim N(0, \sigma_i^2)$$

$$\sigma_i^2 \sim \rho_{\text{prior}}(\sigma^2) = \Gamma^{-1}(\alpha_{prior}, \mathrm{B}_{prior})$$

An "Imperfect" Ensemble Prediction

- We assume an ensemble variance s^2 is drawn from a Gamma distribution $s^2 \sim \Gamma(k,\theta)$ with mean a(ω-R).

$$s^2 = a(\sigma_i^2) * \eta + s_{min}^2$$

Sensitivity

Stochastic

Define a Posterior Distribution of Error Variances Given An Ensemble Variance

Distribution of s^2 given ω

Prior climatological distribution

$$\rho_{\text{post}}(\sigma \mid s^2) = \frac{L(s^2 \mid \sigma^2)\rho_{\text{prior}}(\sigma^2)}{\int_0^\infty L(s^2 \mid \sigma^2)\rho_{\text{prior}}(\sigma^2)\, d\omega}$$

Adjust ensemble variances to be consistent with the mean or the mode of the posterior distribution?

Naïve ensemble prediction

Climatological mean

$$\overline{\sigma^2 \mid s^2} = w_e \sigma_n^2 + (1 - w_e)\overline{\sigma_{\text{prior}}}$$

- This result shows optimal error variance is a combination of a static climatological prediction and a flow dependent prediction: A theoretical justification for Hybrid DA.
- Bishop and Satterfield (2012) show how parameters defining the prior and posterior distributions can be deduced from a time series of innovations paired with ensemble variances.

Ensemble Post-Processing

When all distributions are Gaussian, Bayes' Theorem gives,

$$\left(\overline{\underline{\mathbf{x}}_P^f}\right) = \left[\frac{\frac{1}{\sigma_S^2}}{\frac{1}{\sigma_i^2 \mid s^2} + \frac{1}{\sigma_C^2}}\right]\left(\overline{x^f}\right) + \left[\frac{\frac{1}{\sigma_C^2}}{\frac{1}{\sigma_i^2 \mid s^2} + \frac{1}{\sigma_C^2}}\right]\left(\overline{x^t}\right)_C$$

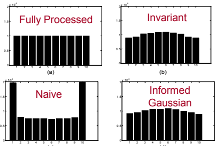
Error variance is not precisely known, but we know that given a forecast and an inaccurate ensemble variance there is an inverse-gamma distribution of possible true innovation variances

$$\frac{1}{\sigma_P^2} = \frac{1}{\sigma_i^2 \mid s^2} + \frac{1}{\sigma_C^2}$$

$$\left(x_i^f\right)_P = \left(\overline{\underline{\mathbf{x}}_P^f}\right) + x'_i, \;\; x'_i \sim N(0, \sigma_P)$$

What is the value of knowledge of the variance of the variance

- **Method 1: Invariant** $\sigma_i^2 \mid s^2 = \overline{\sigma^2}_{prior}$
  - The variance fixed to be the climatological average of forecast error variance. The ensemble prediction of variance is ignored.
- **Method 2: Naïve** $\sigma_i^2 \mid s^2 = \sigma_n^2$
  - The naïve ensemble prediction of the variance is used. All climatological information is ignored.
- **Method 3: Informed Gaussian** $\sigma_i^2 \mid s^2 = \overline{(\sigma \mid s^2)}_{post}$
  - The mean of the posterior distribution is used. The variable nature of forecast error variance is ignored.
- **Method 4: Fully Processed** $\sigma_{ij}^2 \mid s^2 \sim \Gamma^{-1}(\alpha_{post}, \mathrm{B}_{post})$
  - Each ensemble member (j) draws a random variance from the posterior.



Rank frequency histogram obtained from a M=1000 member ensemble by defining 10 bins based on percentile values. The experiments used the parameter values ω'^2=10, *k*=0.5, and *a*=0.5.

Weather Roulette

(Hagedorn and Smith, 2008)

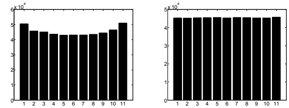
- Two players, A and B, each open a weather roulette casino
- Each start with a $1 bet and reinvest their winnings every round (fully proper variant)
- A and B use their forecast to:
  - set the odds of their own casino ~1/pA(v) , 1/pB(v)
  - distribute their money in the other casino ~ pA(v), pB(v)
- They define 100 equally likely climatological bins
- Value of the ensemble is interpreted using an *effective daily interest rate*

	M=2	M=4	M=6
(a) Invariant	2.0130	3.1625	3.9063
(b) Naïve	28.6550	6.9860	2.8573
(c) Informed-Gaussian	0.6055	0.2269	0.1206

The effective daily interest rate obtained by a gambler using the FP ensemble in casinos whose odds were set by (a) invariant, (b) naïve ,and (c) informed Gaussian ensembles. Columns show the effective number of ensemble members used to generate the variance prediction. Smaller values of *M* correspond to less accurate predictions of error variance. Values shown represent the mean taken over seven independent trials.

Results from the Lorenz '96 Model

- We use 10 variable Lorenz '96 model and Ensemble Transform Kalman Filter (ETKF)
- The ETKF ensemble is re-sampled to create a suboptimal M member ensemble to form the error covariance matrix and produce a suboptimal analysis.
- Only the full ETKF analysis and analysis ensemble are cycled



M=10 member random sample ensemble   M=10 member post-processed ensemble

Conclusions

- Knowledge of varying error variances is beneficial in ensemble post-processing
- Rank Frequency Histograms and the weather roulette *effective daily interest rate* show improvement when varying error variances are considered
- Our ensemble post-processing scheme has proved effective with both synthetic data and data from a long Lorenz '96 model run.

Acknowledgments: CHB gratefully acknowledges support from the U.S. Office of Naval Research Grant# 9846-0-1-5 . This research was performed while ES held a National Research Council Research Associateship Award