Hidden Error Variance Theory: Deriving Optimal Combinations of Static and Flow Dependent Variances

Elizabeth Satterfield[1] , Craig Bishop[2] and David Kuhl[3]
[1]NRC Monterey, CA, [2]NRL Monterey, CA
[3] NRL Washington, DC

5th EnKF Workshop, May 21-24, 2012

Rensselaerville, New York

"A conundrum of predictability research is that while the prediction of flow dependent error distributions is one of its main foci, chaos hides flow dependent forecast error distributions from empirical observation."

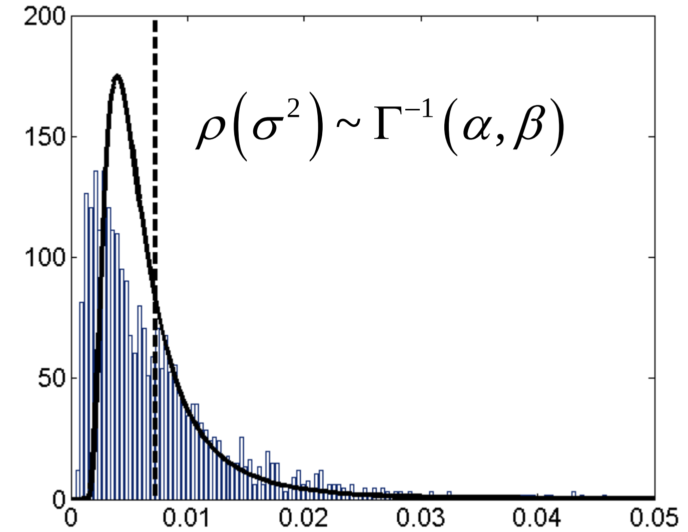
Bishop and Satterfield (2012a,b, MWR, in review), Satterfield and Bishop (2012a,b, to be submitted)

Lorenz Model Experiments

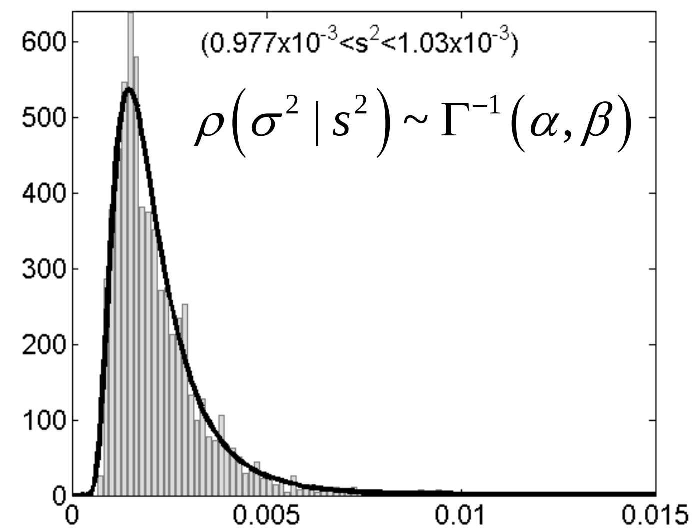
- We use 10 variable Lorenz '96 model and Ensemble Transform Kalman Filter (ETKF).

- We create 25,000 independent time series of analyses and forecasts, each having the same true state and differing only in random draws of observation error, model error, and initial condition error.

- Error Variances are computed for each spatio-temporal point by averaging the squared error across each independent forecast.

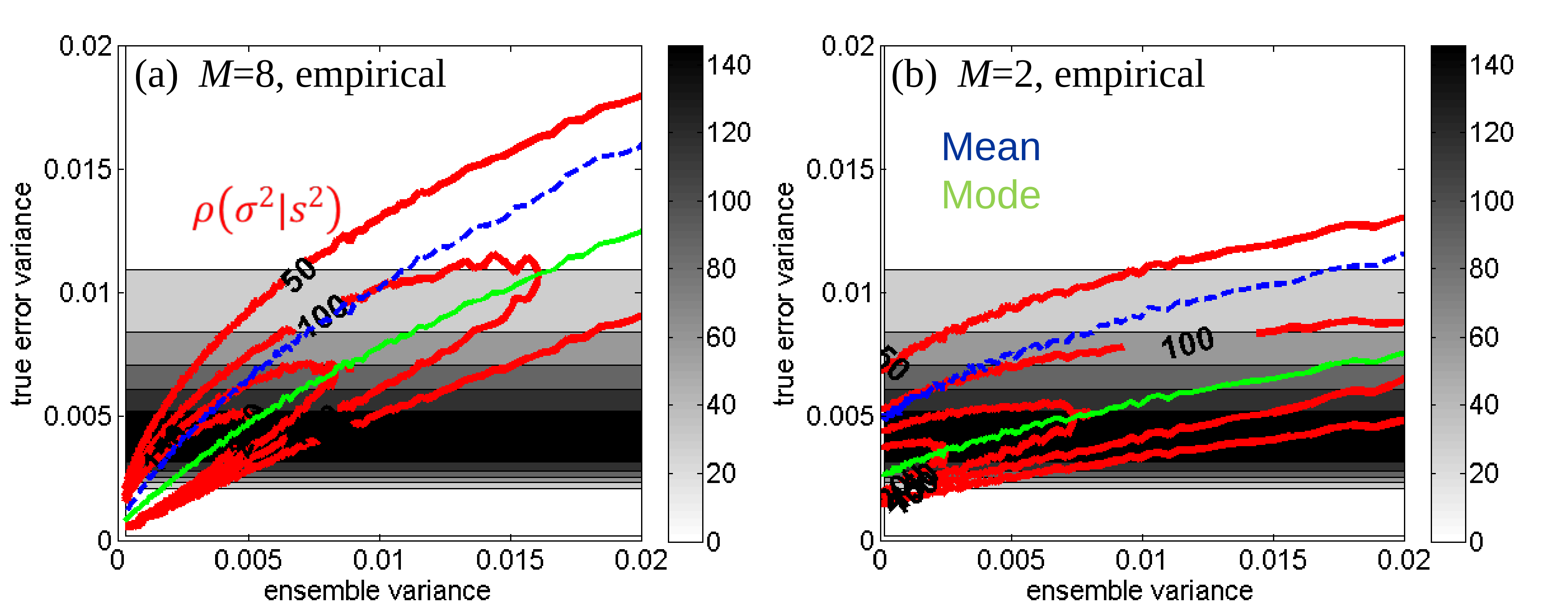


Climatological pdf of true error variances



true error variance given small range of s^2

Empirical estimation of pdf of true error variance given ensemble variance from 25000 trials



Red lines depict empirical estimate of pdf of true error variance (ordinate axis) given fixed values of ensemble variance (abscissa axis). Thin green and blue lines give the mode and mean of the empirical estimates of the mode and mean of these estimates. Panels (a) and (b) show the empirical estimates for random sample ensembles of sizes *M=2* and *M=8*, respectively. The grey shading gives an inverse-gamma pdf fit to the climatological pdf of true error variances.

A Simple Model of Innovation Variance Prediction

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

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

Error Variance

$$\sigma_i^2 \sim \rho_{prior}(\sigma^2) = \Gamma^{-1}(\alpha, \beta)$$

- We assume an imperfect ensemble prediction is drawn from a Gamma distribution with mean $a\sigma_i^2 + s^2_{min}$.

sensitivity

stochastic

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

$$\overline{s^2} = a(\sigma_i^2) + s^2_{min}$$

$$s^2 \sim \Gamma(k, \theta), \quad \theta = (a\sigma_i^2 + s^2_{min}) / k$$

- Bayes' Theorem is used to define a posterior distribution of error variances given an imperfect ensemble prediction

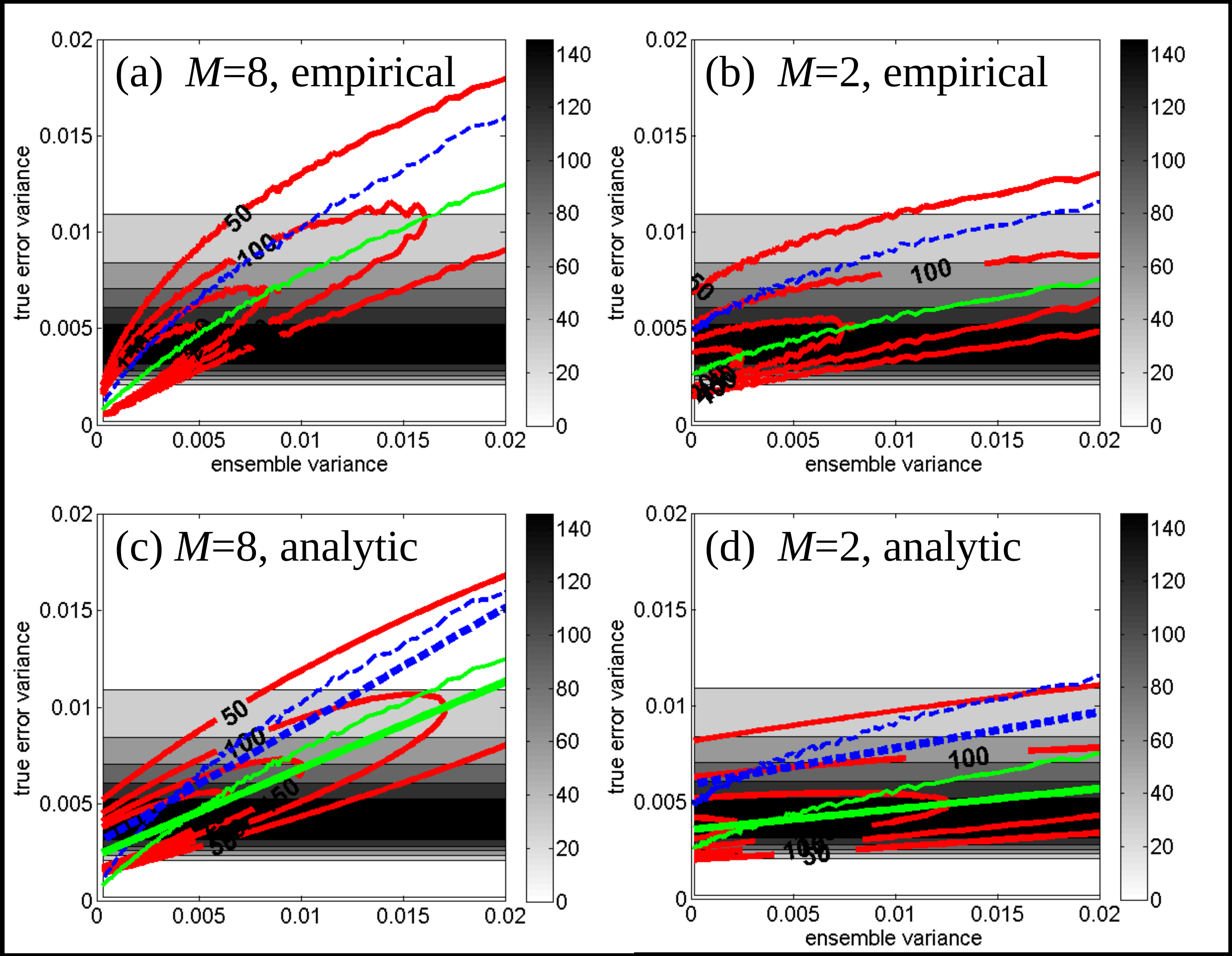
Likelihood distribution of s^2 given a particular σ

Climatological Prior Distribution

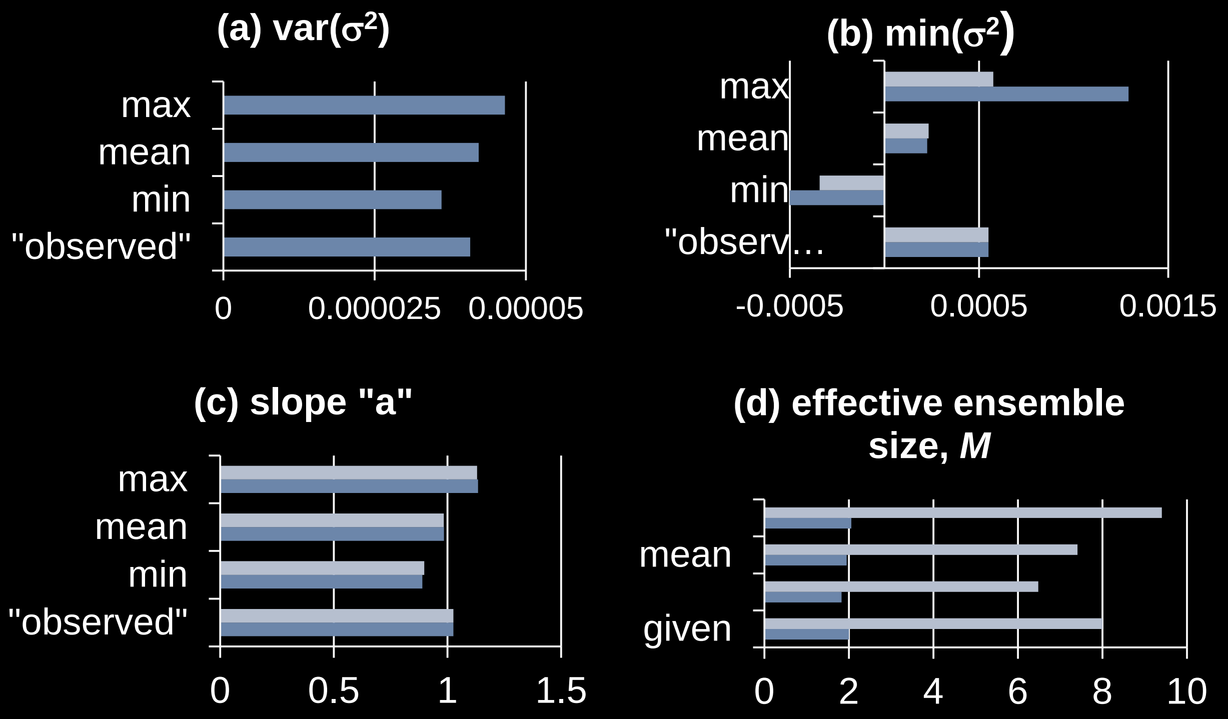
$$\rho_{\text{post}}(\sigma^2 \mid s^2) = \frac{L(s^2 \mid \sigma^2)\rho_{\text{prior}}(\sigma^2)}{\int_0^\infty L(s^2 \mid \sigma^2)\rho_{\text{prior}}(\sigma^2) d\sigma}$$

Analytic model of $\rho\left(\sigma^2 \mid s^2\right)$ gives good match to empirically estimated $\rho\left(\sigma^2 \mid s^2\right)$

$\rho(\sigma^2|s^2)$
Mean
Mode
$\rho(\sigma^2)$



New theory successfully recovers hidden parameters from a single run



- Retrieved hidden parameters, var(σ^2), σ^2, *a* and *M* are shown in plots (a), (b), (c) and (d), respectively
- “Observed” values are obtained from the 175 DA cycles of the 25,000 “replicate systems”
- Minimum, mean and maximum of the values retrieved from 21 sets of 200,000 DA cycles.
- Light blue bars: M=2, Dark blue bars: M=8
- The “given” ensemble sizes in (d) are the random sample ensemble sizes used to degrade the quality of the ETKF ensemble variance.

Combining Static and Flow Dependent Variances

σ^2 : error variance

s^2 : ensemble variance

ω : innovation variance

R : observation error variance

k : parameter defining distribution of ensemble variances

α : parameter defining distribution of error variances

Flow dependent de-biased ensemble prediction

$$\sigma_n^2 = \frac{s^2 - s_{min}^2}{a}$$

static prior mean

$$\overline{\sigma^2 \mid s^2} = \frac{k\sigma_n^2}{k+\alpha-1} + \frac{(\alpha-1)\overline{\sigma_{prior}^2}}{k+\alpha-1} = w_1\sigma_n^2 + w_2\overline{\sigma_{prior}^2}$$

$$= w_1\left[\frac{s^2 - s_{\min}^2}{a} + \left(\sigma_{min}^2\right)\right] + w_2\left(\overline{\sigma_{prior}}^2\right)$$

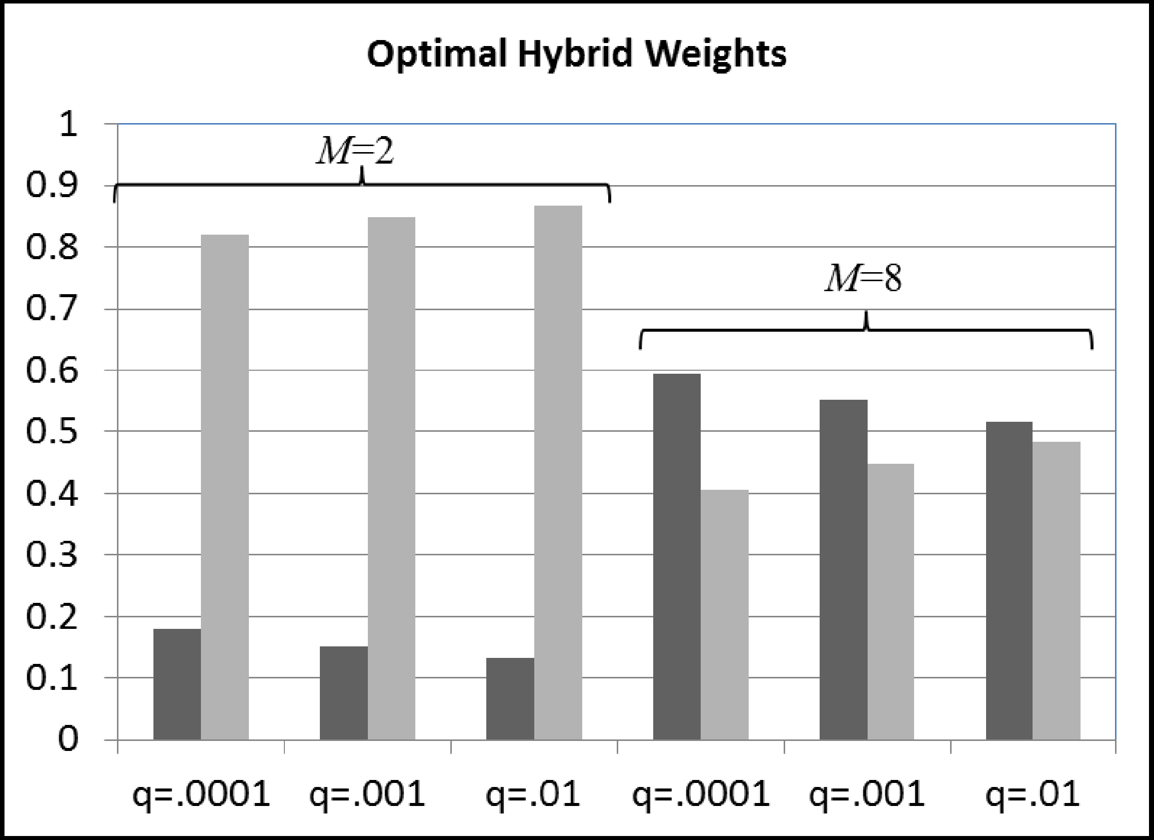
$$\mathbf{P}_{Hybrid}^f = w_1\left\{\frac{\mathbf{P}_{\text{Ensemble}}^f}{a} + \left[\left(\sigma_{min}^2\right) - \frac{s_{\min}^2}{a}\right]\tilde{\mathbf{Q}}_{\text{climatology}}^{\min}\right\} + w_2\mathbf{P}_{\text{climatology}}^f$$

Flow-dependent ensemble prediction

Correlation matrix of unavoidable errors

Static climatological term

Variation of optimal weights with model error and ensemble size, M



Variation of weights for mean of posterior distribution of true error variances with model error *q* and given effective ensemble size *M*. Black bars give the weights for the de-biased *flow-dependent* ensemble variance while grey bars give the corresponding weights for the *static* mean of the climatological error variances.

Perturbed Observations Update

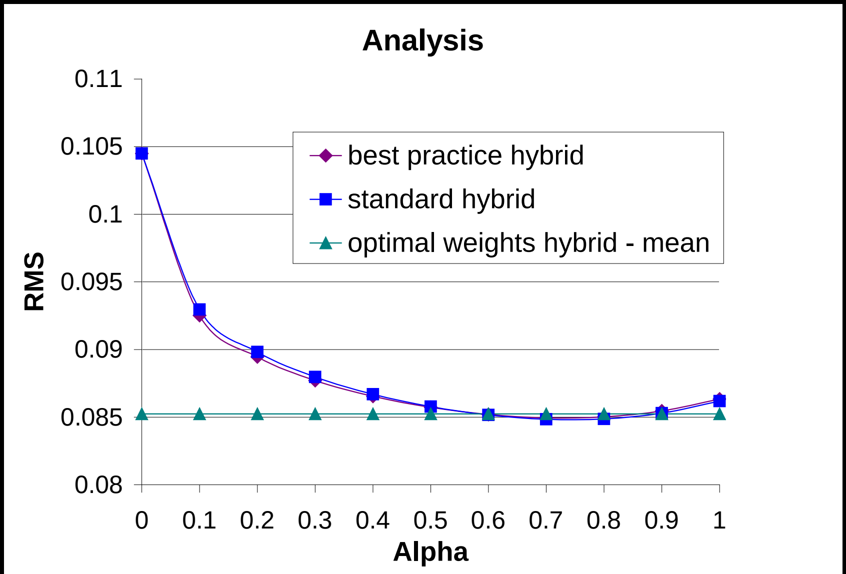
- A suboptimal M=32 member ensemble is generated using a perturbed observations update.
- A climatological error covariance matrix ($P^f_{climatology}$) is formed by collecting forecast errors for 100,000 time steps (using an 100% ensemble based error covariance matrix)

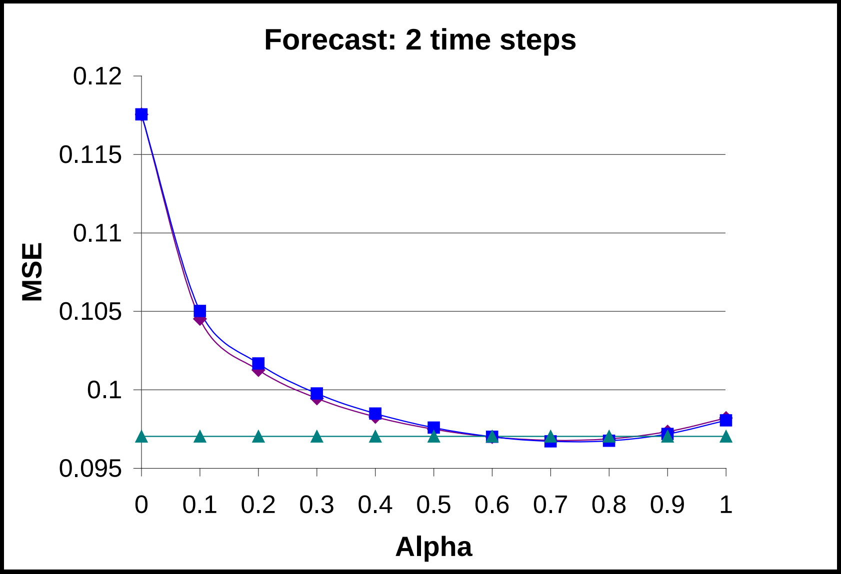
- P^f_{hybrid} is computed at each time step and used in the ETKF DA scheme to obtain an analysis, which is cycled.

- We compute the "best practice" hybrid and the "standard" hybrid for all alpha values for comparison.

  "Best Practice" hybrid: The ensemble based P^f is corrected by a factor of

$$B = \frac{\overline{\sigma^2}}{\overline{s^2}} = \frac{\overline{\varpi} - R}{\overline{s^2}}$$

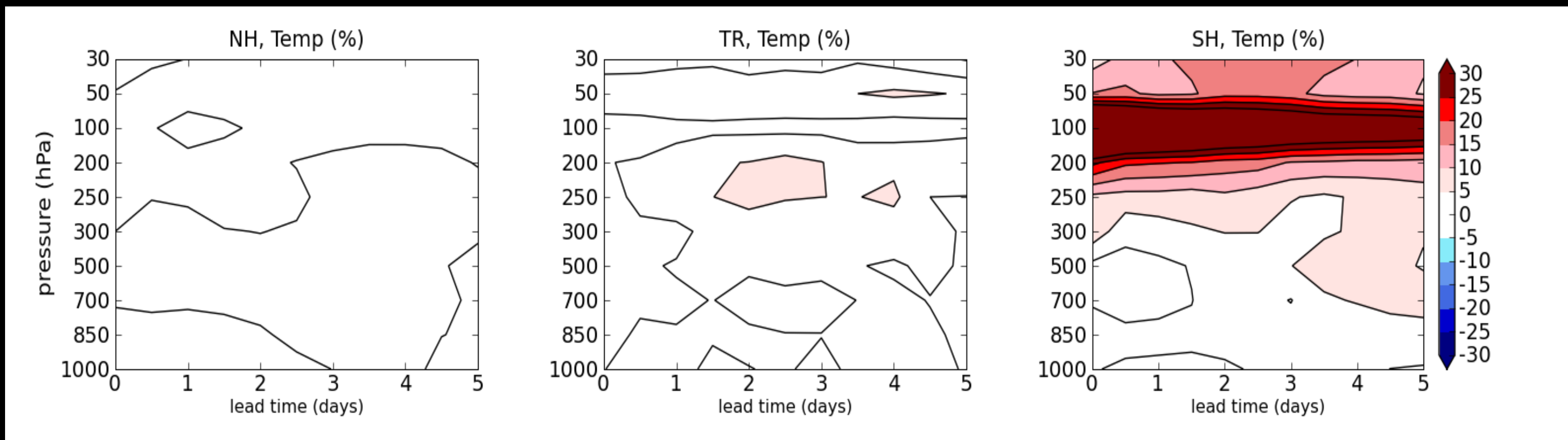




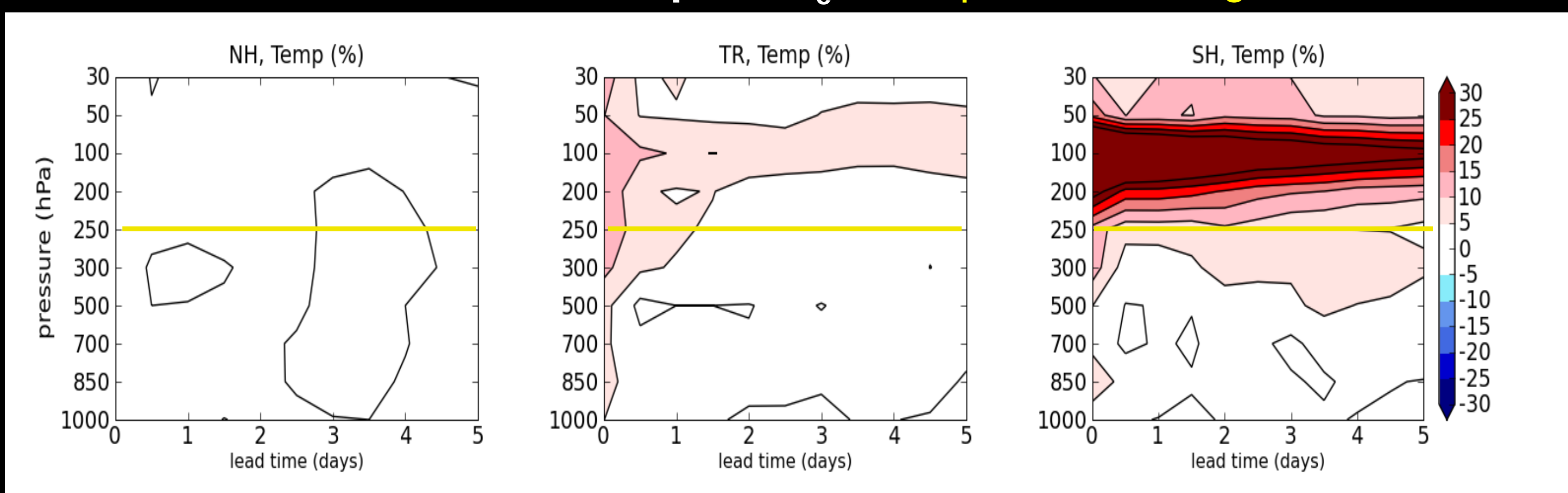
NAVDAS-AR-Hybrid Results

figures provided by David Kuhl

Alpha=0.5



Alpha=W_e computed for 6 regions



Conclusions

1. A simple theory of the relationship between ensemble variances and error variances has been developed.
2. This theory provides a new method for estimating from an archive of pairs of innovations and variances (s^2,ω)
   a. the climatological pdf of error variances,
   b. the pdf of ensemble variances given a true error variance, and
   c. The Posterior Mean of the pdf of error Variances (PMV)
3. Our approximations of an inverse-gamma PDF for the prior climatological distribution of innovation variances and a gamma PDF for the likelihood distribution of ensemble variances given an innovation variance are reasonably accurate for the Lorenz '96 system.
4. Equation for PMV provides a theoretical justification for Hybrid DA systems, which linearly combine static and flow-dependent covariances.
5. Recovery and application of optimal weights of flow dependent and climatological variances has been demonstrated in the Lorenz '96 system
6. The theory developed was just for variances. A complete theoretical basis for the hybrid will require a corresponding treatment of *covariances*.