Variances and Correlations in Hybrid 4DVAR and the use of Climatological Ensembles

David Kuhl[1], Tom Rosmond[2], Craig Bishop[3], Elizabeth Satterfield[4]

[1]NRC/Naval Research Laboratory, Washington DC

[2]Science Application International Corp., Forks, WA

[3]Naval Research Laboratory, Monterey, CA

[4]NRC/Naval Research Laboratory, Monterey, CA

The 9th Workshop on Adjoint Model Applications In Dynamic Meteorology

Cefalu, Sicily, Italy 10-14 October 2011

Session: Hybrid Techniques

Thursday October 13, 2011 11:25AM

Motivation

- Many groups (CMC, NCEP and Met Office) have found that hybrid assimilation results in improved analyses and forecasts
- Part I: We tried our version of hybrid assimilation with the observation space 4D-Var Navy system and found similar results
- Part II: Contribution of variances versus correlations to the improvement
- Part III: Ensemble based climatological mean error covariance

Part I: NAVDAS-AR Hybrid System

$$\underline{\mathbf{P}}_{NAVDAS-AR} = \begin{bmatrix} \mathbf{P}_0^b & \mathbf{P}_0^b\mathbf{M}^T \\ \mathbf{M}\mathbf{P}_0^b & \mathbf{M}\mathbf{P}_0^b\mathbf{M}^T + \mathbf{Q} \end{bmatrix}$$

- $\underline{\mathbf{P}}_{NAVDAS-AR}$ is the error covariance matrix for NAVDAS-AR specified at all time steps of the DA window
- We replace the conventional $\mathbf{P}_0^b$ of NAVDAS-AR with a hybrid $\mathbf{P}_0^b$
- The hybrid $\mathbf{P}_0^b$ is a combination of the conventional and ensemble covariances:

$$\mathbf{P}^b_{0\_Hybrid} = (1-\alpha)\mathbf{P}^b_{0\_CONV} + \alpha\mathbf{P}^b_{0\_ENS}$$

NAVDAS-AR Conventional $\mathbf{P}^b_{0\_CONV}$

$$\mathbf{P}^b_{0\_CONV} = \mathbf{D}^{1/2}_{CONV}\mathbf{C}_{CONV}\mathbf{D}^{1/2}_{CONV}$$

- Variances ($\mathbf{D}_{CONV}$):
  - Geo-pot. height and temperature are in exact hydrostatic balance
  - Geo-pot. height and winds are approximately geostropically balanced in the extratropics and independent in tropics
- Correlations ($\mathbf{C}_{CONV}$):
  - Isotropic correlation model based on balanced and unbalanced correlations separable in the vertical and horizontal (see Chapter 4 Daley and Barker 2000)
- **Strengths**:
  - High rank
  - Preserves some aspects of geophysical balances
- **Weaknesses**:
  - Not flow dependent
  - Horizontal length scale independent of height may not apply in both troposphere and stratosphere
  - Balance assumptions are incorrect in boundary layer and stratosphere

Flow Dependent Ensemble $\mathbf{P}^b_{0\_ENS}$

$$\mathbf{P}^b_{0\_ENS} = \left[\frac{1}{K-1} \sum_{i=1}^{K} \left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^T \right] \odot \mathbf{C}$$

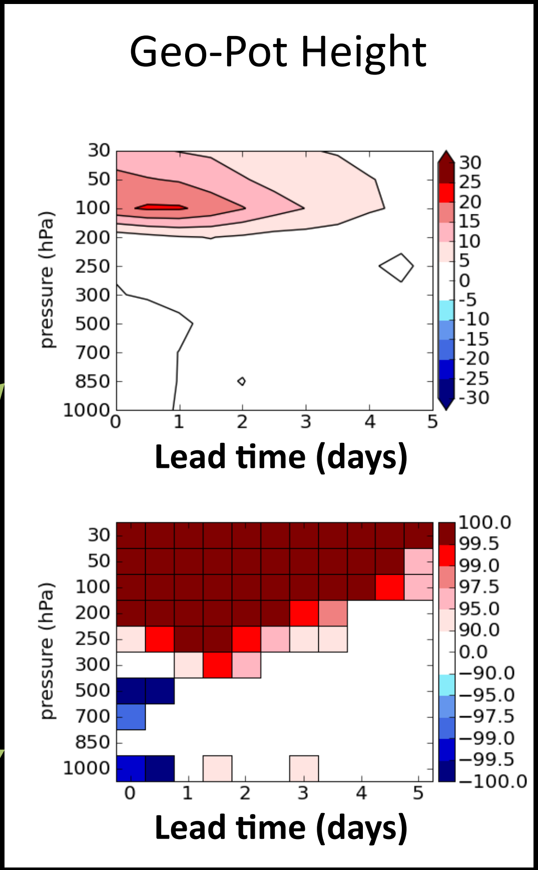
- Where:
  - $\mathbf{x}_i - \bar{\mathbf{x}}$ is the ensemble perturbation
  - K is the number of ensemble members
  - $\mathbf{C}$ is localization matrix
- Ensemble is created using the 5-banded Ensemble Transform (ET)
  - Mean: 3-hour forecast of 4D-var analyses at high resolution
  - Covariances (balance of):
    - Operational 3D-Var variances
    - 3-hour forecast of ensemble members at low resolution
- **Strengths**:
  - Flow dependent errors of the day
  - Multivariate balances implied by the localized ensemble correlations
- **Weaknesses**:
  - Localization damages geophysical balances
  - Cycled ensembles (ET, ETKF, EnKF, etc) often result in variances that are too small in some regions and too large in others. Getting this correct is a work in progress.

Experimental Setup

- Cycling analysis from Nov. 20, 2008 to Dec. 31, 2008
- Discard first 8 days of analysis for ensemble spin-up
- Model resolution: T119L42 outer, T47L42 inner
- Ensemble resolution (same as inner): T47L42
- 32 Ensemble Members
- Assimilating only conventional observations (no radiances)
- Verification:
  - 5-day forecasts from each analysis
  - Verification of forecasts with radiosondes

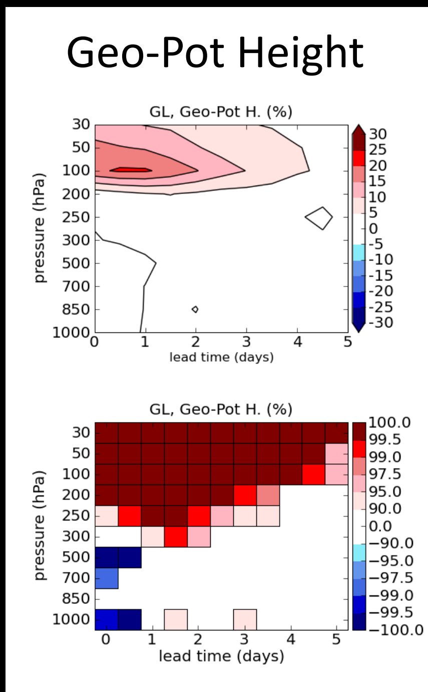
Conventional vs. Hybrid $\alpha = 0.5$

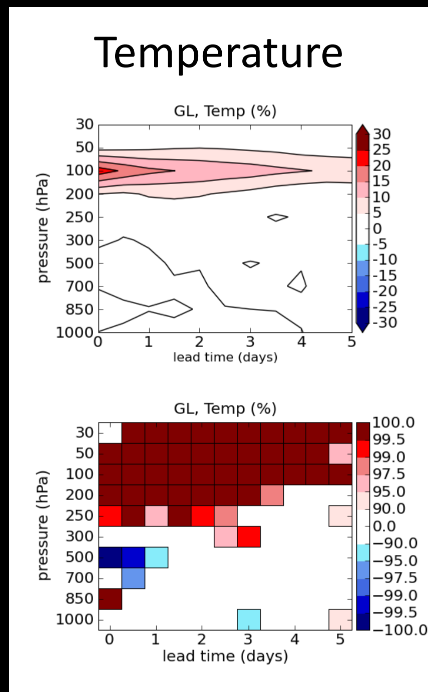
- Experiment comparison:
  - Blue is win for Conventional $\mathbf{P}^b_{0\_CONV}$
  - Red is win for Hybrid $\mathbf{P}^b_{0\_Hybrid}$
- Percentage reduction/increase of rms error relative to conventional
  - RMS error is computed relative to radiosondes at different forecast lead times 0-5 days
  - Forecasts were launched every 12 hours from Nov. 28, 2008 to Dec. 31, 2008
- Statistical significance of RMS errors difference

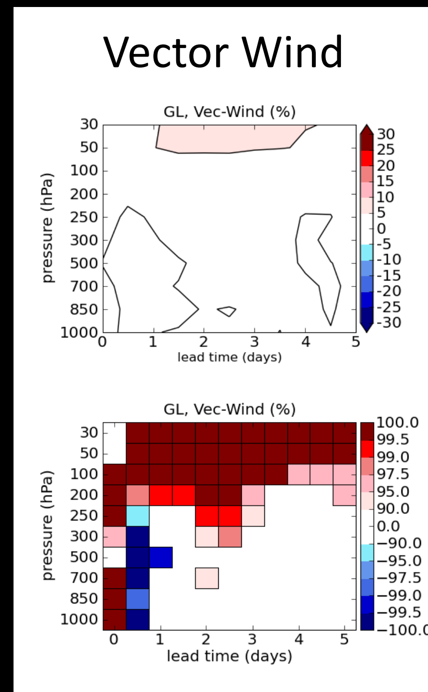


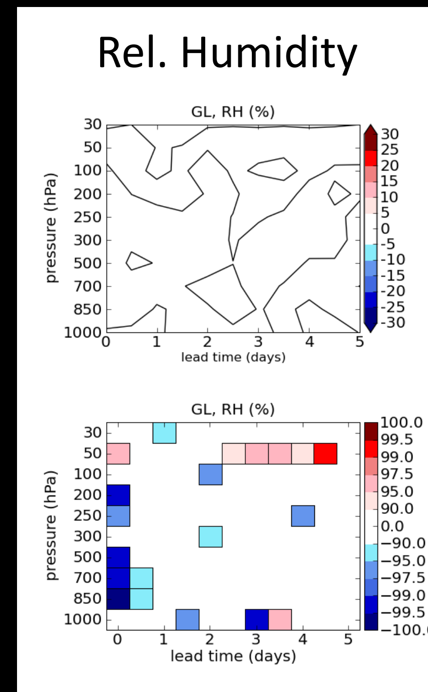
Conventional vs. Hybrid $\alpha = 0.5$

$$\mathbf{P}^{b}_{0\_Hybrid} = 0.5\mathbf{P}^{b}_{0\_CONV} + 0.5\mathbf{P}^{b}_{0\_ENS}$$









Red is a win for Hybrid (alpha=0.5), Blue is a win for Conventional

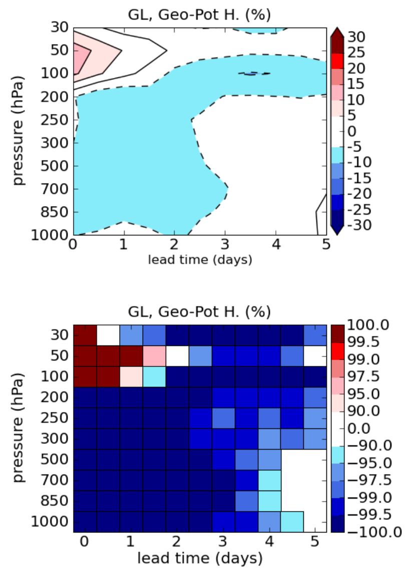
Similar results as others: hybrid assimilation produces better forecasts

Our improvements to the conventional method are found in stratosphere

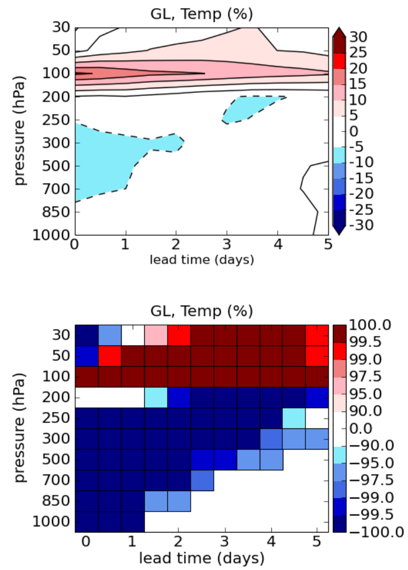
Conventional vs. Hybrid $\alpha = 1.0$

$$\mathbf{P}^{b}_{0\_Hybrid} = \mathbf{P}^{b}_{0\_ENS}$$

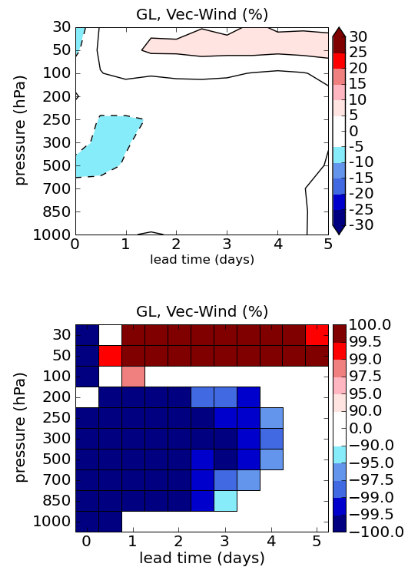
Geo-Pot Height



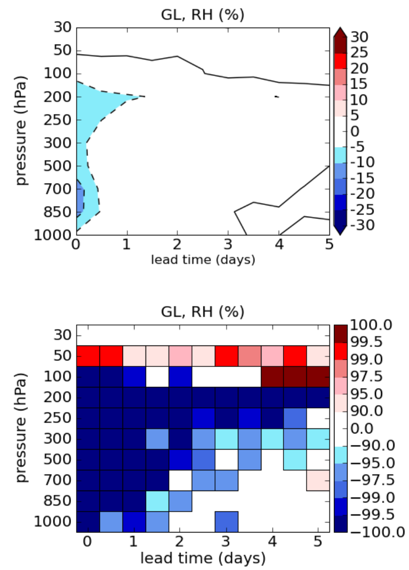
Temperature



Vector Wind



Rel. Humidity



Red is a win for Ensemble (alpha=1.0), Blue is a win for Conventional

Similar results as others: Ensemble alone is mixed result

We clearly see ensemble contributes positive impact to the stratosphere

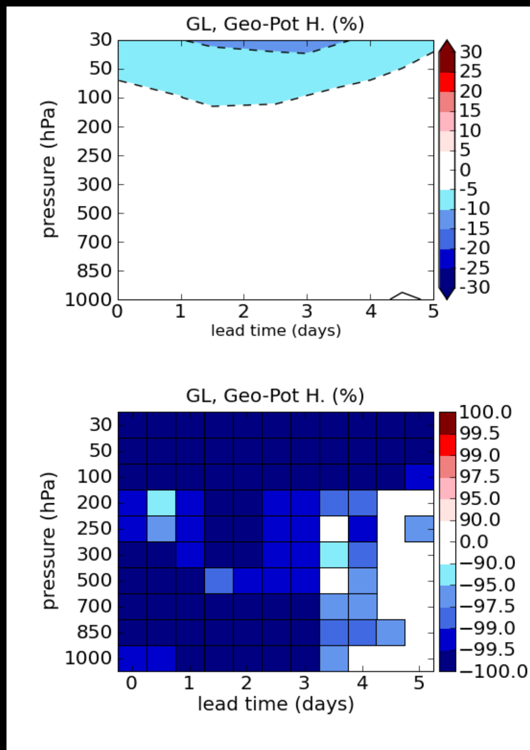
Part II: What part of $\mathbf{P}_{0\_ENS}^{b}$ is contributing to the positive impacts?

1. Variances? (Is the improvement due to the ensemble contribution to the variance estimate?).
   - Test using $\mathbf{P}_0^b = \mathbf{D}_{ENS}^{1/2}\mathbf{C}_{CONV}\mathbf{D}_{ENS}^{1/2}$
2. Correlations? (ensemble length scales, or multi-variate correlations superior?).
   - Test using $\mathbf{P}_0^b = \mathbf{D}_{CONV}^{1/2}\mathbf{C}_{ENS}\mathbf{D}_{CONV}^{1/2}$

Variance/Correlation Impact

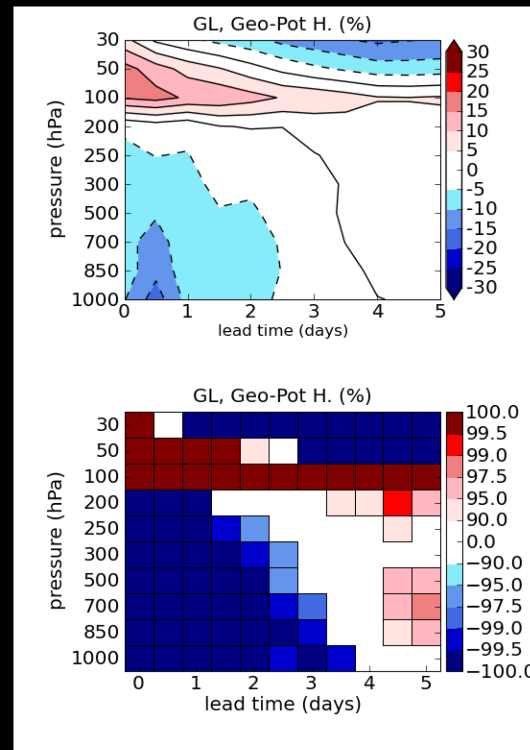
Ens. Var./Conv. Corr.

$$\mathbf{P}_0^b = \mathbf{D}_{ENS}^{1/2}\mathbf{C}_{CONV}\mathbf{D}_{ENS}^{1/2}$$



Ens. Corr./Conv. Var.

$$\mathbf{P}_0^b = \mathbf{D}_{CONV}^{1/2}\mathbf{C}_{ENS}\mathbf{D}_{CONV}^{1/2}$$



Red is a win for Experiment, Blue is a win for Conventional

We clearly see the ensemble correlations is where the positive stratospheric impact is coming from.

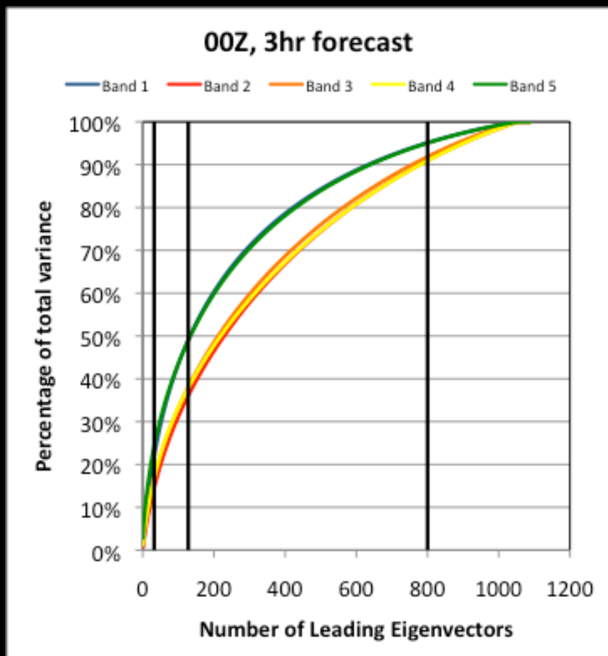
Part III: Climatological Ensemble

- Archives of ensemble perturbations at 0,6,12, and 18 UTC were created from our 40 day ensemble run. The covariance of these perturbations provide estimate of the climatological error covariance.
- Motivated in part by Bishop and Satterfield's theory for the distribution of error variances given an inaccurate ensemble variance which shows that optimal error variance prediction is a (Hybrid) linear combination of a climatological error variance ($\sim \mathbf{P}^b_{0\_CONV}$) and ensemble variance ($\sim \mathbf{P}^b_{0\_ENS}$).
- **Strengths**:
  - Multivariate balances implied by the localized averaged ensemble correlations
  - No need for online forecasts
- **Weaknesses**:
  - Flow dependent errors of the day
  - Localization damages balance

Our Climate Ensemble

- Collect 34 days (Nov. 28th to Dec. 31st) of 32 member flow dependent ensembles
- Collect into 4 diurnal groups (00Z, 06Z, 12Z and 18Z) of 1,088-members
- Produce smaller ensemble sets:
  – Use Singular Value Decomposition to calculate the eigenvectors of the members and arrange them from leading to trailing eigenvector.
  – Collect either 32, 128 or 800 leading eigenvector ensemble members

Our Climate Ensemble

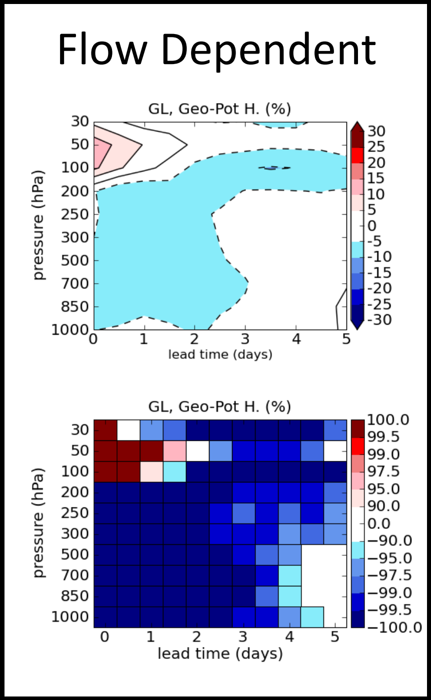


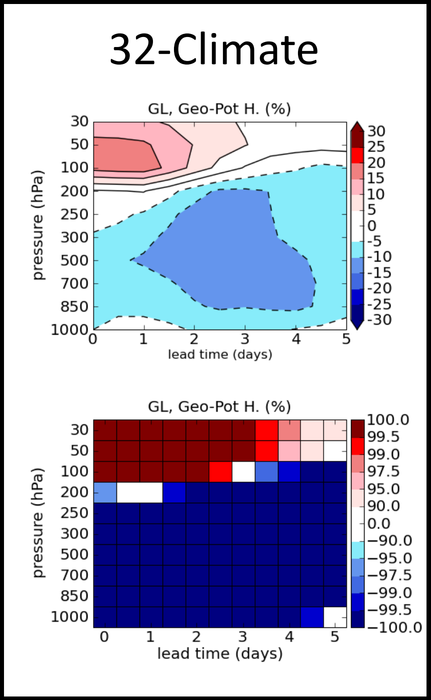
Percentage of Total Variance			
	Number of Leading Eigenvectors		
	32	**128**	**800**
Band 1	22%	49%	95%
Band 2	14%	36%	91%
Band 3	17%	38%	92%
Band 4	17%	38%	91%
Band 5	24%	49%	95%

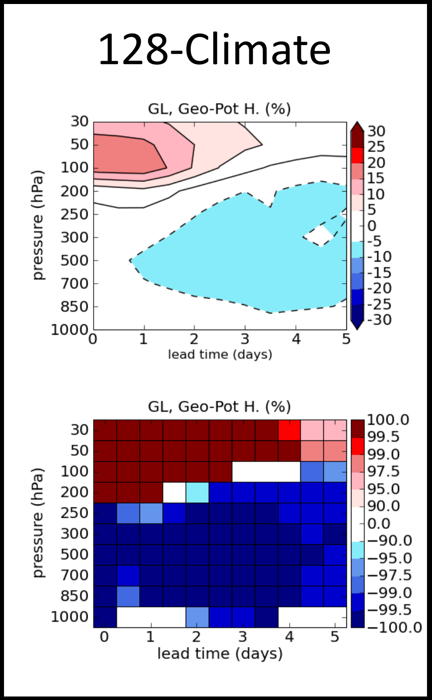
- The sets of leading eigenvectors (32, 128 or 800 members) are normalized to have the same variance as the total initial 1,088 member ensemble
- The localization (relative to 32-member flow dependent ensemble): 800-Climate is slightly larger, 128 and 32-Climate is the same

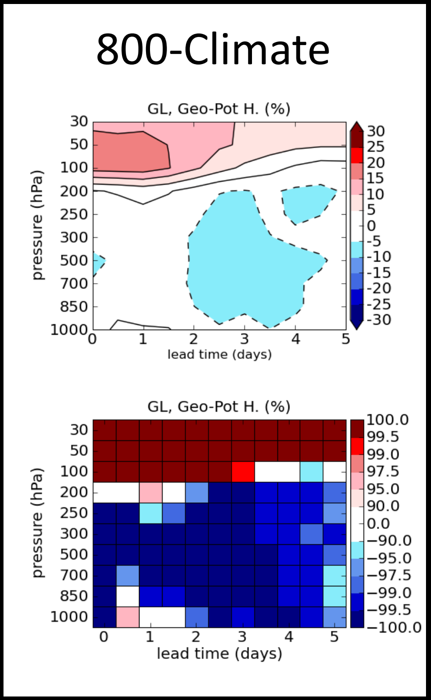
Flow Dependent compared to Climate

$$\mathbf{P}^b_{0\_Hybrid} = \mathbf{P}^b_{0\_ENS} \qquad \mathbf{P}^b_{0\_Hybrid} = \mathbf{P}^b_{0\_CLIM}$$









Red is a win for Experiment, Blue is a win for Conventional

Impact of clmatological ensemble is similar to flow dependent ensemble

Static Hybrid Assimilation

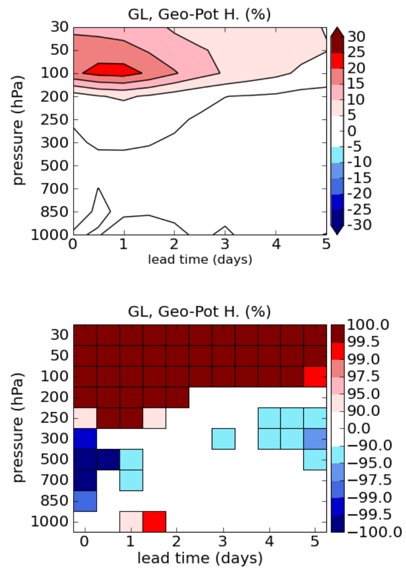
- Here, we linearly combine the conventional static error covariance matrix with the static climatological error covariance matrix.

$$\mathbf{P}^{b}_{0\_Static} = (1-\beta)\mathbf{P}^{b}_{0\_CONV} + \beta\mathbf{P}^{b}_{0\_CLIM}$$

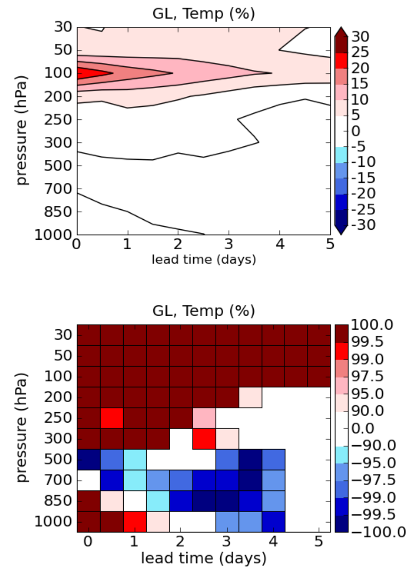
Conventional vs. Static Hybrid $\beta = 0.5$

$$\mathbf{P}^b_{0\_Static} = 0.5\mathbf{P}^b_{0\_CONV} + 0.5\mathbf{P}^b_{0\_CLIM}$$

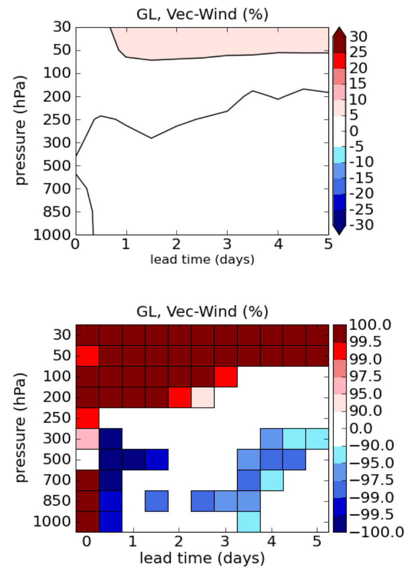
Geo-Pot Height



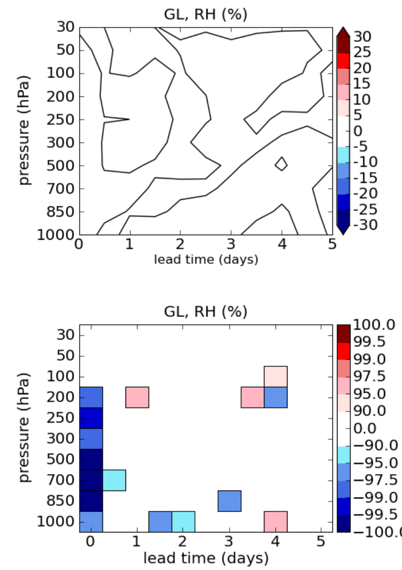
Temperature



Vector Wind



Rel. Humidity



Red is a win for Hybrid (alpha=0.0, beta=0.5), Blue is a win for Conventional
The Hybrid 128-Member climate performs as well as the flow dependent hybrid. And it takes less online computational time.

Conclusions

- Part I: Our New Hybrid Assimilation System
  - Hybrid ensemble system improved forecasts
  - Ensemble on its own improved stratosphere but degraded troposphere
- Part II: Ensemble correlations and variances
  - Experiments switching variances and correlations suggest that ensemble correlations are the source of the improvements in stratosphere
- Part III: Climatological ensemble
  - The climatological ensemble can be used to improve the static background error covariance
- Experiments at operational resolutions and with a full set of operational observations are underway

Climate Hybrid Assimilation

- With climate hybrid assimilation we combine the of the conventional and climate to form a better static $\mathbf{P}^b_{0\_Static}$ and then combine with the ensemble to capture any flow dependent structures:

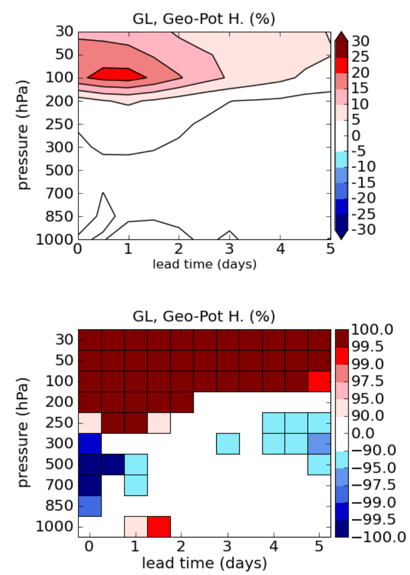
$$\mathbf{P}^b_{0\_Static} = (1-\beta)\mathbf{P}^b_{0\_CONV} + \beta\mathbf{P}^b_{0\_CLIM}$$

$$\mathbf{P}^b_{0\_Hybrid} = (1-\alpha)\mathbf{P}^b_{0\_Static} + \alpha\mathbf{P}^b_{0\_ENS}$$

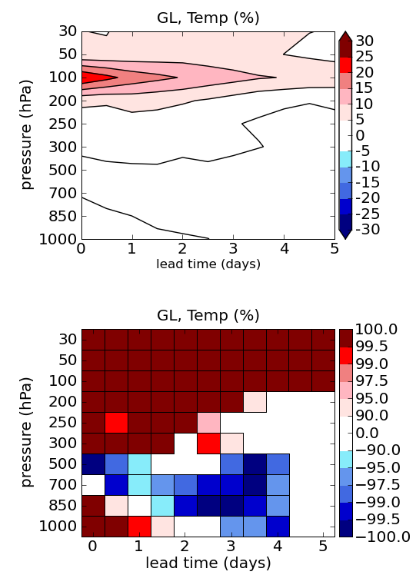
Only Static: $\alpha = 0.0 \quad \beta = 0.5$

$$\mathbf{P}^b_{0\_Hybrid} = 0.5\mathbf{P}^b_{0\_CONV} + 0.5\mathbf{P}^b_{0\_CLIM}$$

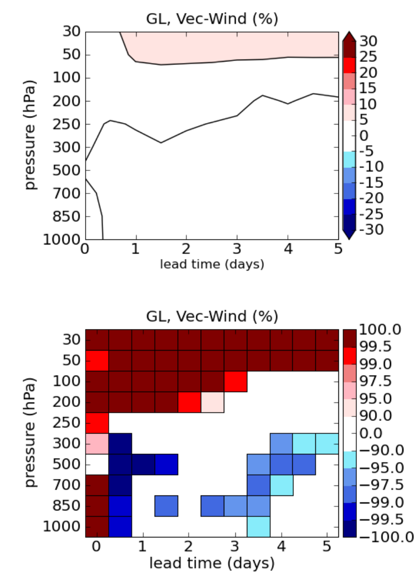
Geo-Pot Height



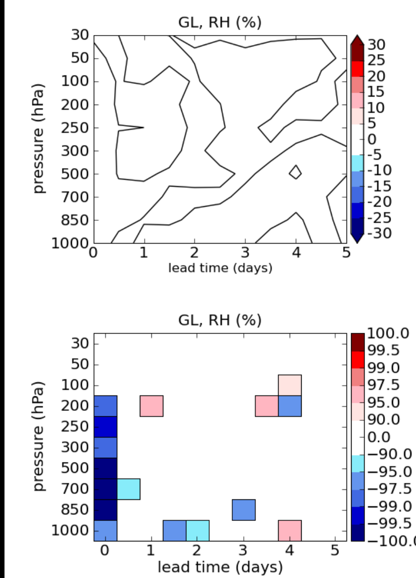
Temperature



Vector Wind



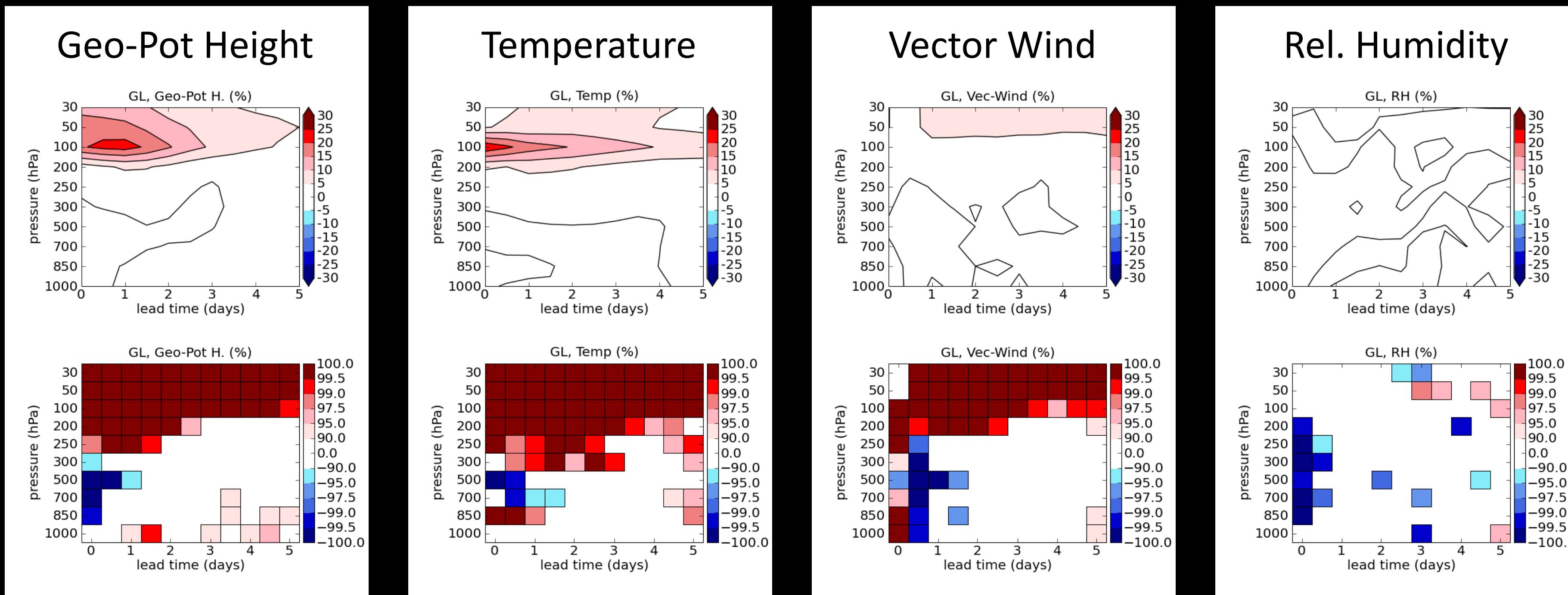
Rel. Humidity



Red is a win for Hybrid (alpha=0.0, beta=0.5), Blue is a win for Conventional
The Hybrid 128-Member climate performs basically as well as the flow dependent hybrid. And it takes less online computational time.

Full Mixture: $\alpha = 0.5 \quad \beta = 0.5$

$$\mathbf{P}^b_{0\_Hybrid} = 0.5\left[0.5\mathbf{P}^b_{0\_CONV} + 0.5\mathbf{P}^b_{0\_CLIM}\right] + 0.5\mathbf{P}^b_{0\_ENS}$$



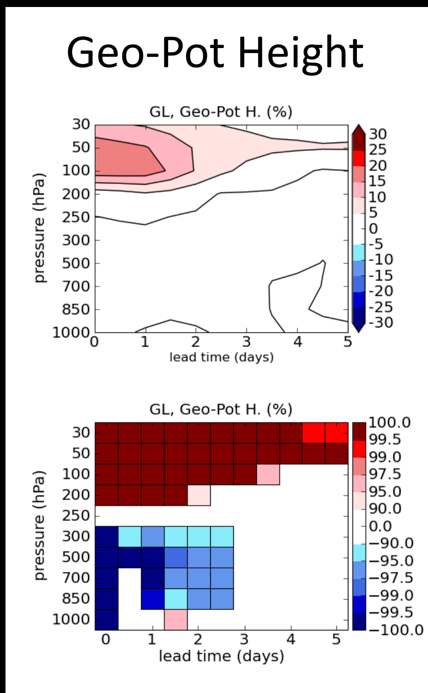
Red is a win for Hybrid (alpha=0.5, beta=0.5), Blue is a win for Conventional

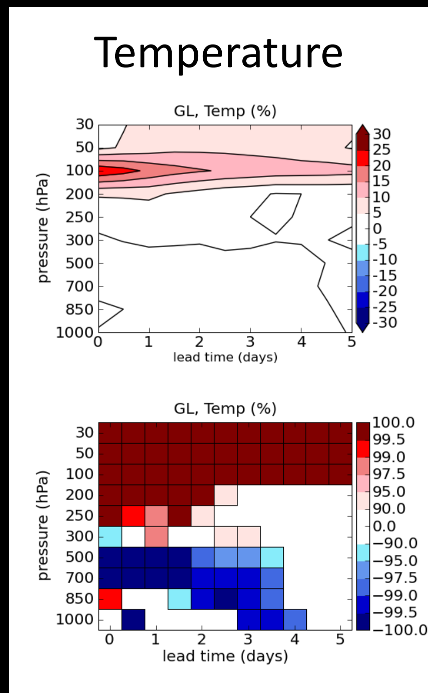
We are using the best static we've come up with, this is probably

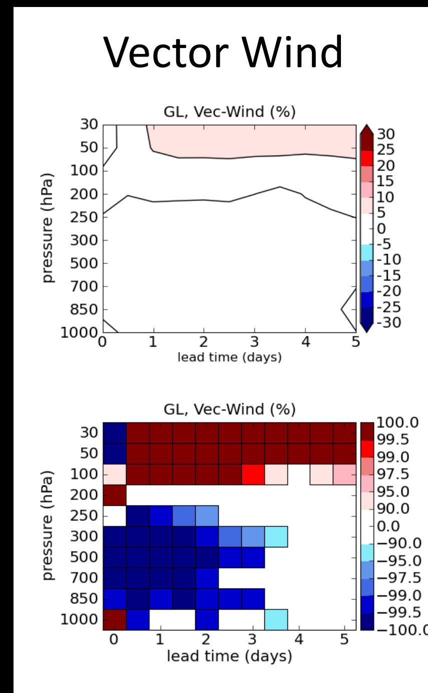
The best results we've seen.

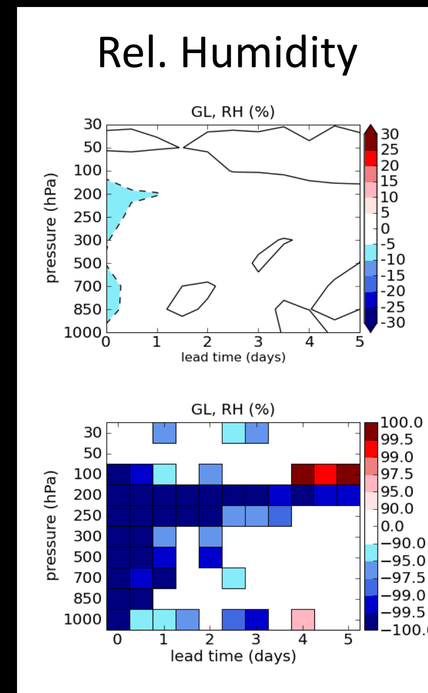
Only Ensemble: $\alpha = 0.5 \quad \beta = 1.0$

$$\mathbf{P}^b_{0\_Hybrid} = 0.5\mathbf{P}^b_{0\_CLIM} + 0.5\mathbf{P}^b_{0\_ENS}$$









Red is a win for Hybrid (alpha=0.5, beta=1.0), Blue is a win for Conventional

We are only using information from the ensemble, this shows that a PbN hybrid with no TLM/adjoint may be possible