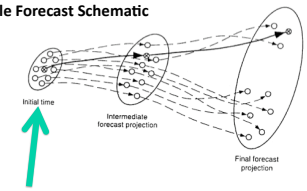
Ensemble Kalman Filter Development at ESRL

Jeffrey S. Whitaker, Thomas M. Hamill, Gilbert P. Compo, Philip Pegion (PSD)

in collaboration with GSD, NRL, NCAR, Universities and NCEP/EMC

What is an Ensemble Kalman Filter (EnKF)?

Ensemble Forecast Schematic



Where does this part come from?

- in operational NWP using variational data assimilation, perturbations are added to analysis.
- perturbations may not sample PDF of analysis error.
- the EnKF is a data assimilation technique that automatically defines an ensemble of initial states that is a random sample of (estimated) analysis error.

How does it work?

Given a single ob y^o with expected error variance R, an ensemble of model forecasts $\mathbf{x}^b$ (model priors), and an ensemble of predicted obs $\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$ (observation priors) :

Step 1: Update observation priors.

(1a) $\underline{\mathbf{y}}^a = (1-K)\underline{\mathbf{y}}^b + Ky^o$ *update for ob prior means*

(1b) $\mathbf{y}^{a\prime} = (1-K)^{1/2}\mathbf{y}^{b\prime}$ *rescaling of ob prior perturbations*

Where $K = \mathrm{var}(\mathbf{y}^b)/(\mathrm{var}(\mathbf{y}^b)+R)$, underline denotes means, prime denotes perturbations, superscript *b* denotes prior, *a* denotes analysis.

Linear interpolation between observations and observation prior mean with weight K (0<=1<=K), and a rescaling of ob prior ensemble so that $\mathrm{var}(\mathbf{y}^a)=(1-K)\mathrm{var}(\mathbf{y}^b)$.

when $\mathrm{var}(\mathbf{y}^b) \ll R$, all weight given to prior.
when $\mathrm{var}(\mathbf{y}^b) \gg R$, all weight given to observation.

Step 2: Update model priors.

Let $\Delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b$ be the increment to the model priors, $\Delta\mathbf{y} = \mathbf{y}^a - \mathbf{y}^b$ be the increment to the observation priors.

(2) $\Delta\mathbf{x} = \mathbf{G}\Delta\mathbf{y}$ *computation of increments to model prior*

where $\mathbf{G} = \mathrm{Cov}(\mathbf{x}^b, \mathbf{y}^{bT})/\mathrm{Cov}(\mathbf{y}^b, \mathbf{y}^{bT})$

Linear regression of model priors on observation priors

Only change model prior when $\mathbf{x}^b$ and $\mathbf{y}^b$ are correlated within the ensemble.

What's it good for?

- Since $\mathbf{P}^b$ is computed from an ensemble, it is flow dependent (it knows about the flow situation).
- Analysis accuracy improved relative to 3DVar (particularly for poorly observed variables) – comparable to 4DVar.
- Initial ensemble represents the uncertainty in the "flow of the day" – leading to better estimate of uncertainty in the forecast.

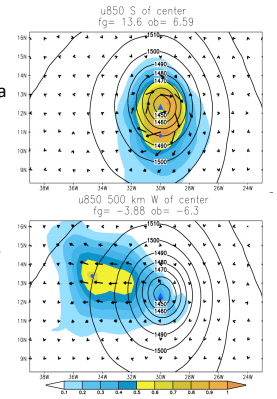
Benefits of Flow-Dependent Background Errors

Example 1: Tropical cyclone

Hurricane Fred 00Z 9 Sep

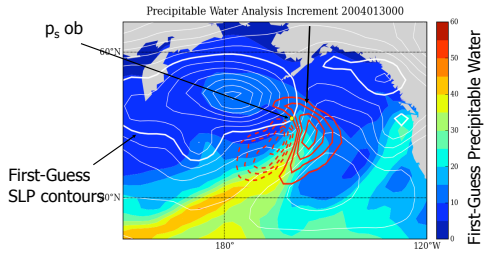
Single ob increments for 850 hPa zonal wind ob 1 m/s different than background.

Analysis "knows" where the hurricane is.



Solid contours: *850 hPa background geopotential height*
Colors: *wind speed increment*
Arrows: *vector wind increment*
Blue triangle: *hurricane center*
Blue circle: *location of ob*

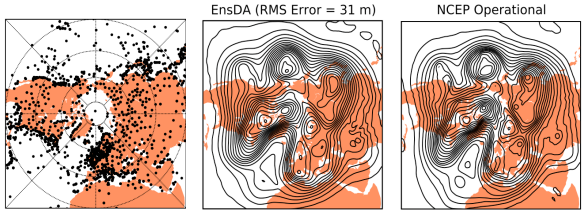
Example 2: Mid-Lat cyclone



Surface pressure observation can improve analysis of integrated water vapor (flow-dependent cross-variable covariance in $\mathbf{P}^b$).

Applications

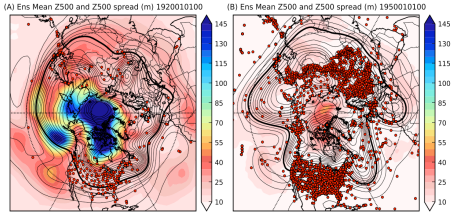
Historical Reanalysis



Surface pressure obs circa 1930 O(100's)

EnKF analysis at 500 hPa using only those p_s obs

NCEP operational analysis using O(1000000's) obs



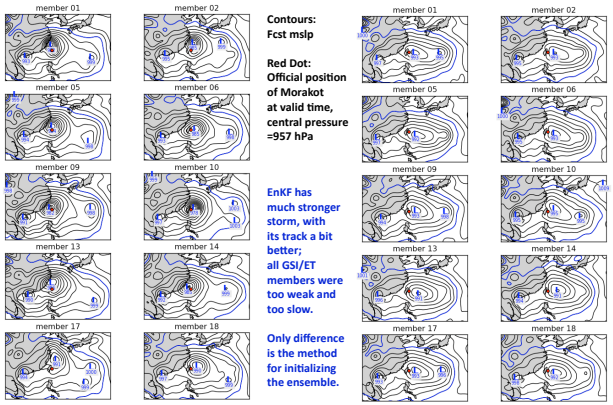
Example contrasting 1920 and 1950 – color shading is an estimate of analysis uncertainty.

Typhoon Morakot

Ens. Size: 20 members (10 shown); INI: 2009080500; Lead Time: 54 hours; Valid: 2009080706, at landfall

EnKF ensemble, T382 resolution, 20 members
Central pressure 978-991 hPa

GSI/ET ensemble, T382 resolution, initialized with GSI analysis plus operational perturbations
Central pressure: 991-995 hPa



Contours: Fcst mslp

Red Dot: Official position of Morakot at valid time, central pressure =957 hPa

EnKF has much stronger storm, with its track a bit better; all GSI/ET members were too weak and too slow.

Only difference is the method for initializing the ensemble.