EnKF with the **CNMCA** Regional forecasting system: ***recent developments***

Massimo Bonavita[1], Lucio Torrisi, Francesca Marcucci
CNMCA National Meteorological Center,Italy

[1] Current affiliation: ECMWF, Reading (UK)

OUTLINE

- The Ensemble Kalman Filter approach : pro and cons
- Ensemble Data Assimilation at CNMCA :
  Configuration of the system
- Progress report on model error parameterization experiments
- Lessons learned and outstanding issues

Ensemble Kalman Filter

Uses an ensemble of N system states to parametrize the distribution

$$P^b = \frac{1}{N-1} X^b X^{b^T} \qquad X^b = x^b - \overline{x}^b$$

follows the time evolution of the mean and covariance (Gaussian assumption) by propagating the ensemble of states

LETKF FORMULATION (Hunt et al,2007)

$$\tilde{H}_n = H(x_n^b) - \overline{H}(x^b)$$

$$\tilde{P}^a = \left[(\tilde{H}^T R^{-1} \tilde{H} + (N-1) I \right]^{-1}$$

$$K = X^b \tilde{P}^a \tilde{H}^T R^{-1}$$

$$X^a = X^b W^a$$

$$W^a = \left[(N-1) \tilde{P}^a \right]^{1/2} \quad (\text{Square root filter})$$

$$\overline{x}^a = \overline{x}^b + K(y^o - H(\overline{x}^b))$$

$$x^a = \overline{x}^a + X^a$$

The analysis ensemble mean is the linear combination of forecast ensemble states which best fits the observational dataset

Ensemble Kalman Filter

PLUS

- Algorithm simplicity;
- Avoid the need of tangent linear approximation;
- Low order and explicit representation of B matrix (error of the day);

LETKF

- Analysis is performed locally (local analyses for each grid column, obs selection) → intrinsically parallel;
- Avoids serial processing of observations (allows taking into account correlated observation errors inside local patches);
- Inverse matrix computed in the low order ensemble space at every grid point
- Natural 4D extension (4D-LETKF, Hunt et al,2004)

Ensemble Kalman Filter

MINUS

SOLUTIONS

- misspecification of observation error matrix **R** ;
- errors in observation operator (i.e. representativeness);

→ **Common to 3D-Var,4D-VAR**

- sampling errors (limited ensemble size) → **Localization**
- forecast model deficiencies → **Inflation factors**
- sampling errors due to nonlinearities (especially at very high resolution where highly-nonlinear processes are to be represented/parameterized - microphysics, turbulence, surface fluxes); → **OUTER LOOP**
- non-gaussianity of forecast and observation errors

EnKF : errors handling

total forecast errors = internal errors + external error

$$\mathbf{P}_{\mathrm{i}}^{b} = \mathbf{M}_{x_{i-1}^{a}} \mathbf{P}_{\mathrm{i}}^{\mathrm{a}} \mathbf{M}_{x_{i-1}^{a}}^{\mathrm{T}} + \boxed{\mathbf{Q}}$$

Errors in initial state and their dynamical growth

Model deficiencies (model error)

EnKF estimates the background error covariances from an ensemble of forecasts which allows them in theory to include information on the flow-dependent error of day (both temporally and spatially variant) → better than 3D-Var

BUT ensemble spread only represents growth of initial condition errors

overconfidence on background → system decoupled from truth

SOLUTIONS

MULTIPLICATIVE INFLATION

$$Q = 0 \qquad P^b \rightarrow (1+\Delta)P^b$$

Implicitly assumes that model errors have the same error structure as the internal errors so that their error covariance **Q** can be represented by dynamically evolved error covariance $\mathbf{P}^{\mathbf{b}}$

ADDITIVE INFLATION

$$Q = Q(0, q)$$

Add random perturbations with a certain covariance structure and zero mean
Select perturbations consistent in structure with model errors
→explore unstable directions that lie outside analysis subspace

EnKF at CNMCA: proof of concept

CNMCA Implementation and upgrades (with respect to last COSMO GM)

- **30 member** ensemble (based on HRM) at 0.25° (~ **28Km**) grid spacing (EURO-HRM domain), **40 hybrid p-sigma vertical levels** (top at 10 hPa)
- Initial ensemble from different EURO-HRM forecasts valid in ± 48h around start time
- boundary from ECMWF IFS for all members (not perturbed)
- **6-hourly** assimilation cycle run for 15 days
- $(T, u, v, \mathbf{q_v}, P_s)$ set of control variables
- Operational 3DVar cycle run in parallel at same spatial resolution
- Observations: RAOB (Tuv), SYNOP(SP), SHIP(SP), BUOY(SP), **AIREP, AMDAR, ACAR, AMV, MODIS, WPROF**
- **800 Km** circular local patches (obs weight smoothly decay $\propto r^{-1}$)

EnKF at CNMCA

Model Parameterization experiments

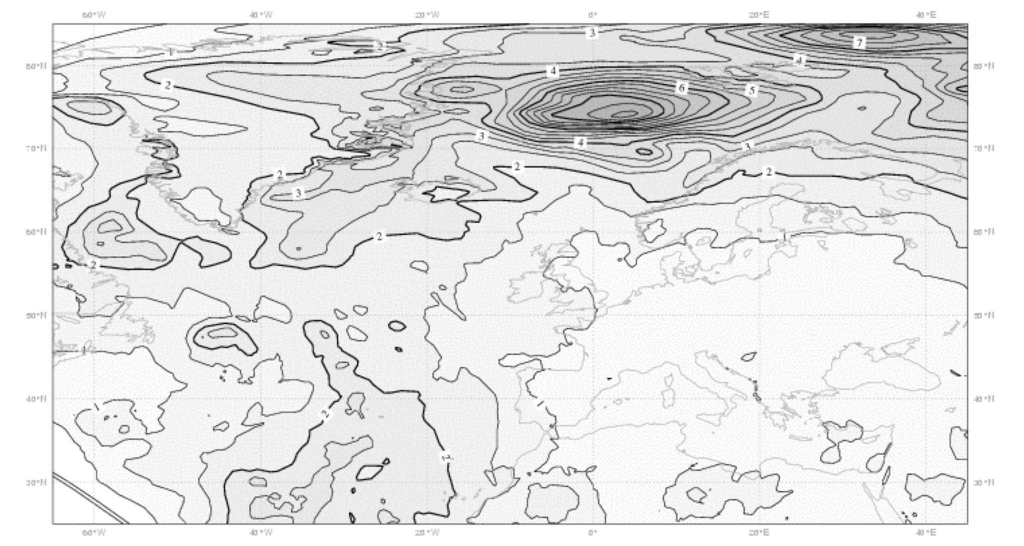
ML1: MULTIPLICATIVE INFLATION FACTOR TIME and MODEL LEVEL varying

$$(1+\Delta) = \frac{d^T_{o-b} d_{o-b} - Tr(R)}{Tr(HP^bH)}$$

\+ temporal smoothing algorithm

simple scalar Kalman filter approach (Li et al,2007)

R slow varying → estimated on long assimilation period instead of simultaneously

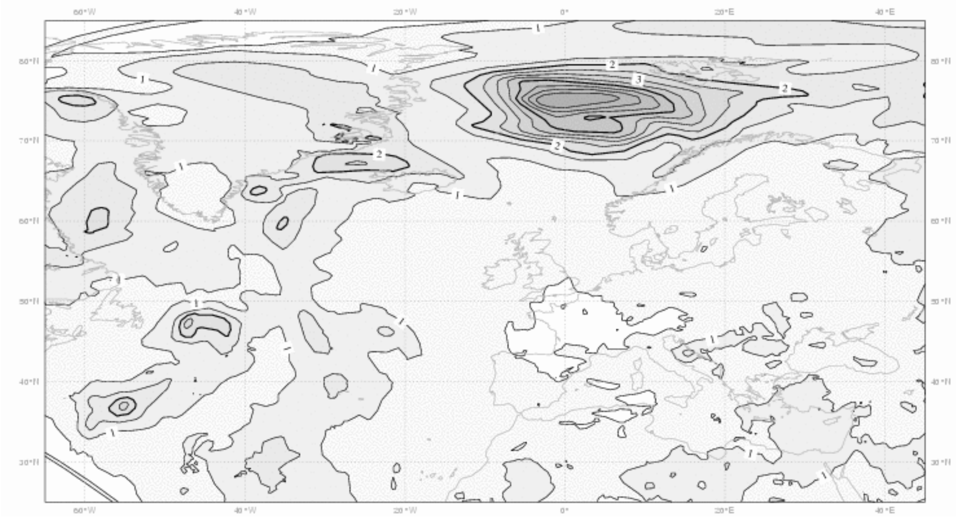


Background ensemble FCST standard deviation (ps)

Unrealistic forecast error variance in observation poor regions and too small variance in observation rich regions

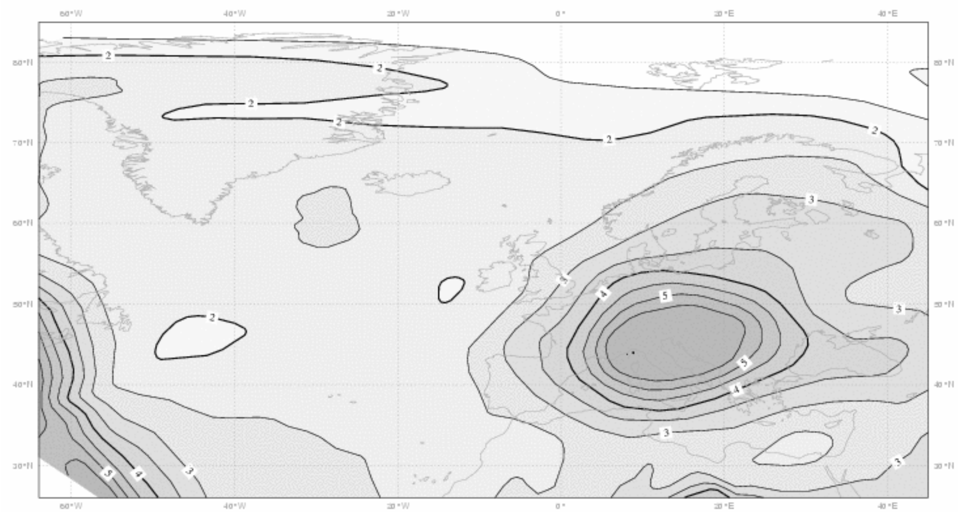
EnKF at CNMCA: Model Parameterization experiments

ML2: 3D ADAPTIVE INFLATION FACTOR



Background ensemble forecast standard deviation of surface pressure

- more realistic representation of forecast uncertainty



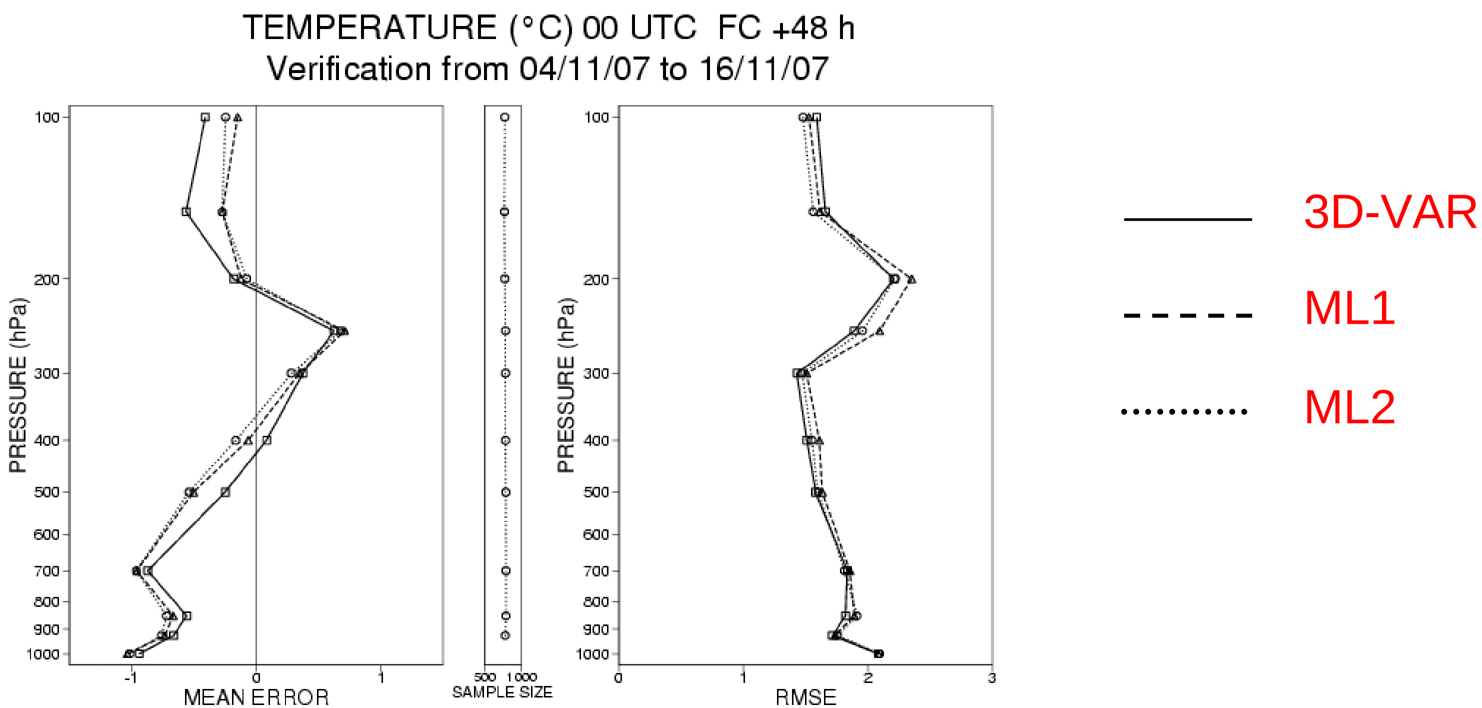
Spatial structure of inflation factor @ 500 hPa

- reflects the underlying observations' distribution
- higher values in the better observed regions confirm the idea that in these regions the forecast error is dominated by the poorly known model error
- need compensation for the lack of spread at the borders of the integration domain

→ boundary perturbation

EnKF at CNMCA: Model Parameterization experiments

Temperature forecast (+48h) verification against RAOB



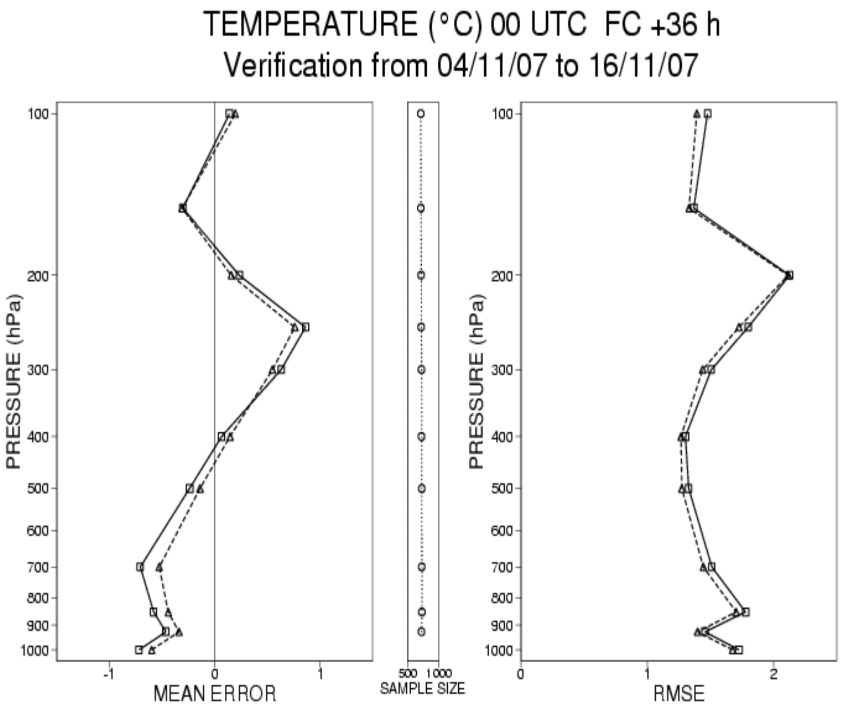
EnKF at CNMCA: Model Parameterization experiments

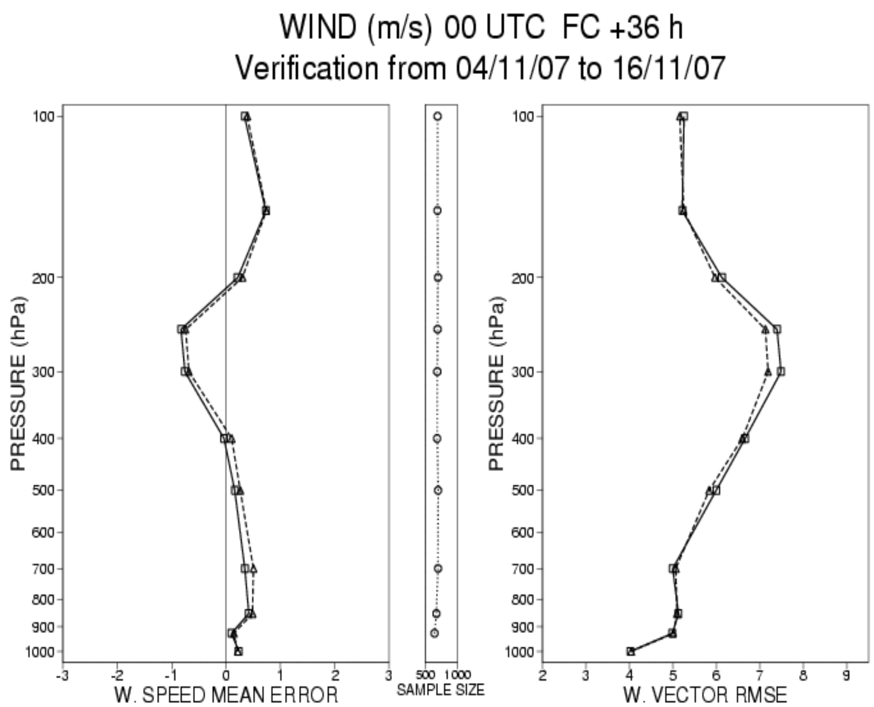
ML3: ADDITIVE INFLATION FACTOR

Additive perturbation derived from randomly selected, scaled 24-hour forecast differences

$$x_i^{\,a} = x^{a}{}_{e(i)} + r q_i$$

$$\overline{q}_i = 0$$

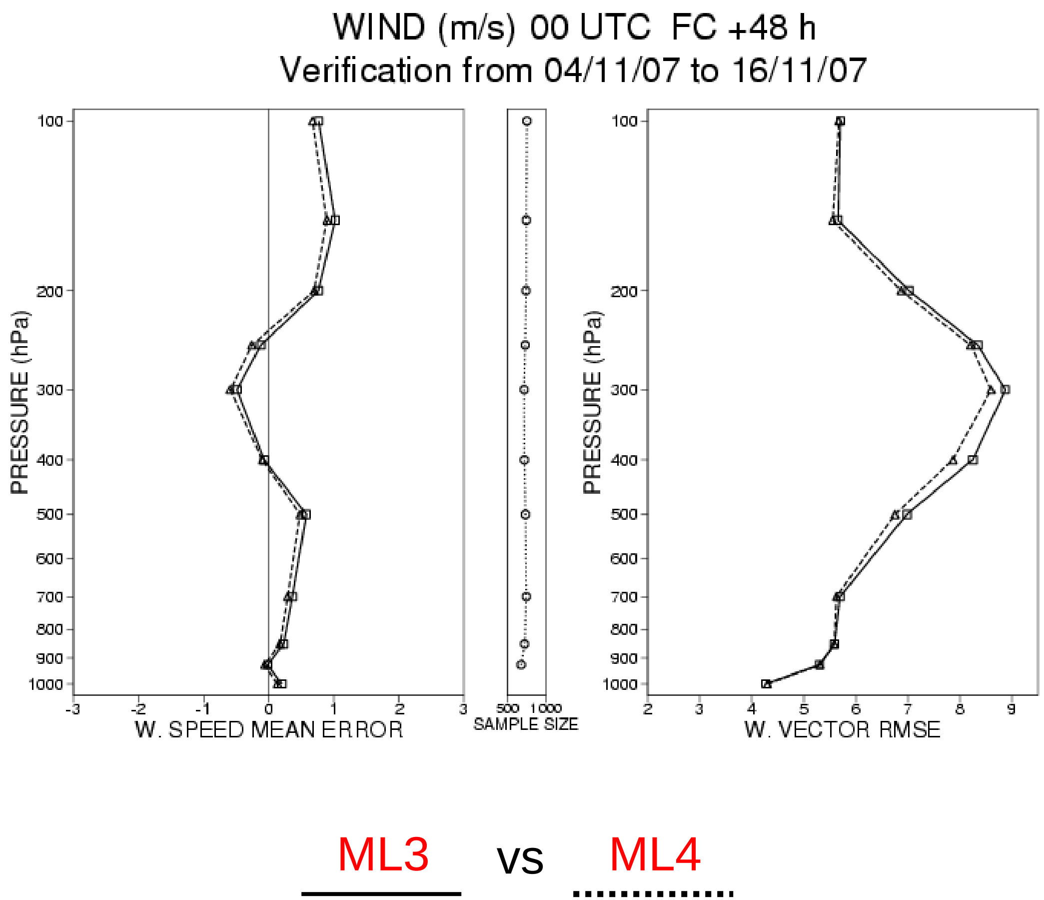




ML2 vs ML3

EnKF at CNMCA: Model Parameterization experiments

ML4: 3D ADAPTIVE MULTIPLICATIVE + ADDITIVE INFLATION FACTOR

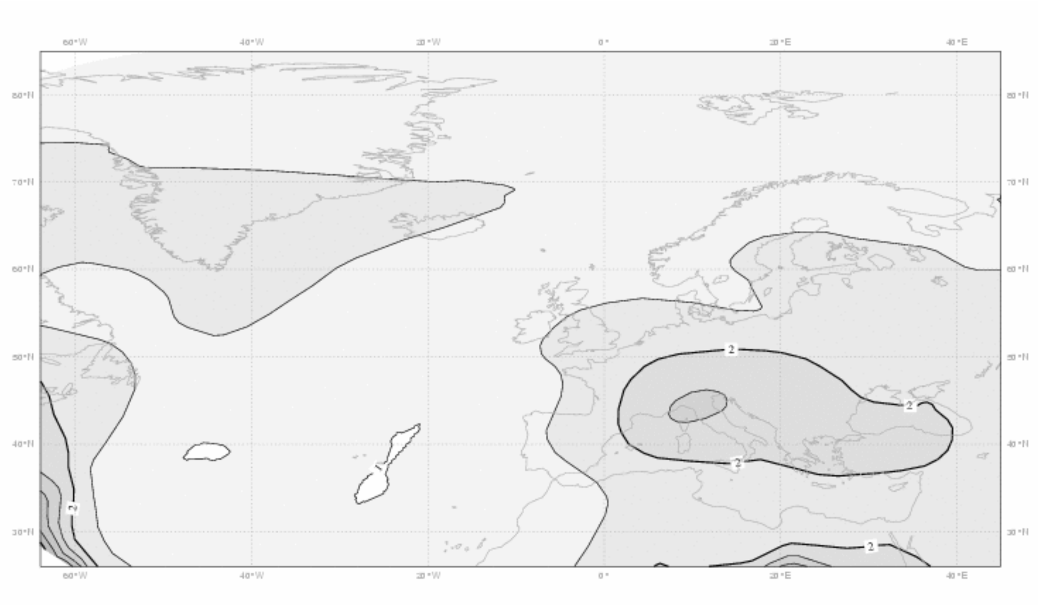


Wind profile forecast (+48h) verification against RAOB

- Light improvement of wind forecast scores
- Temperature scores unchanged

EnKF at CNMCA: Model Parameterization experiments

ML4: 3D ADAPTIVE MULTIPLICATIVE + ADDITIVE INFLATION FACTOR



Spatial structure of inflation factor @ 500 hPa

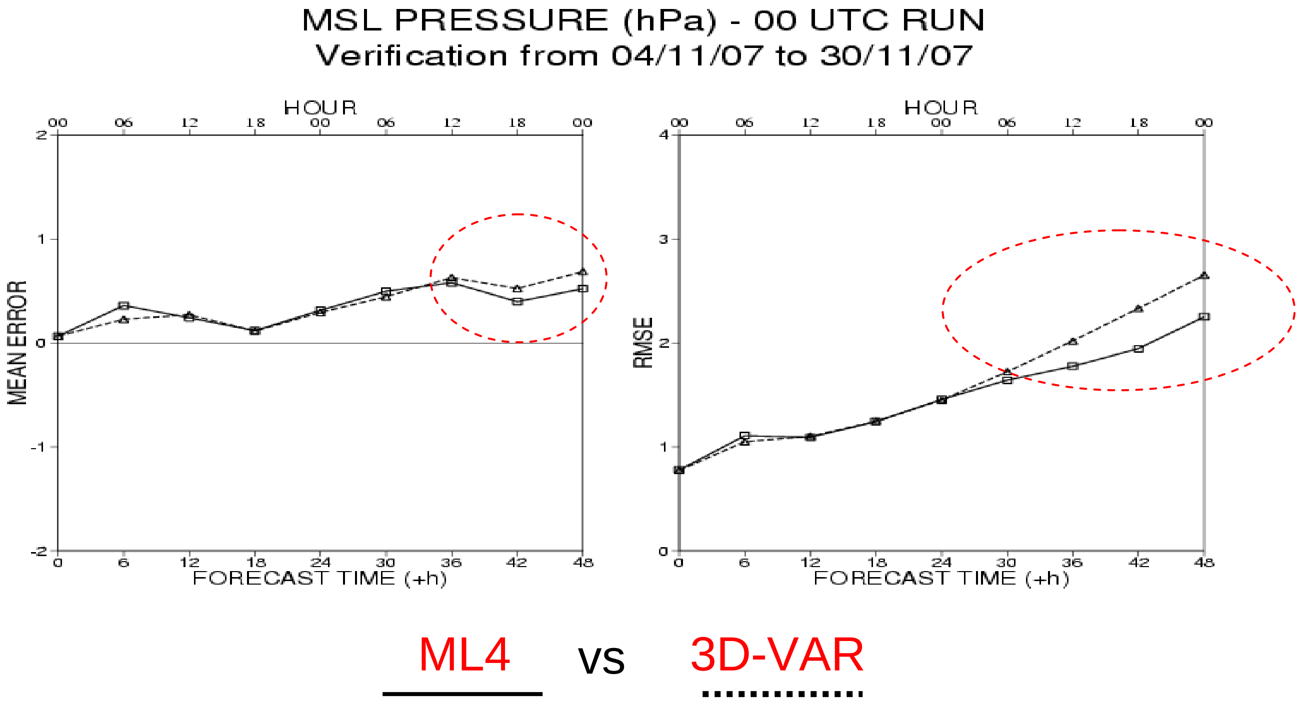
- multiplicative factor has become a 2nd order effects which corrects for locally (space/time) insufficient forecast spread

EnKF at CNMCA

EXTENDED FORECAST EXPERIMENT: LETKF VS 3D-VAR

- **ML4** configuration
- 6-hourly assimilation cycle run for **30 days** (1-30 Nov 2007)

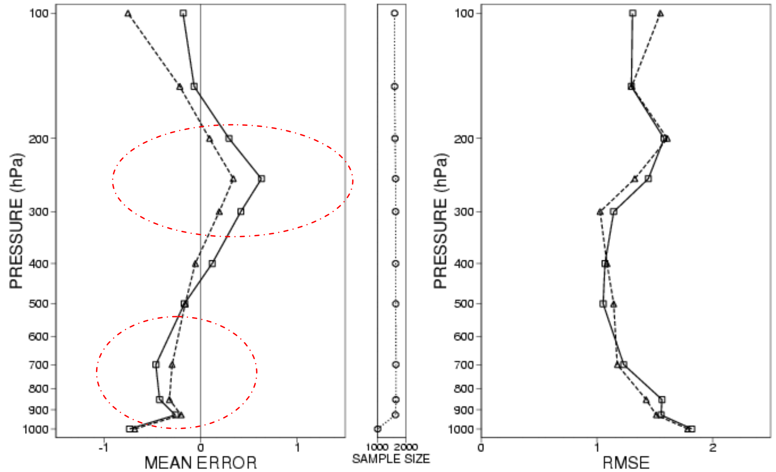
Bias and root means square error



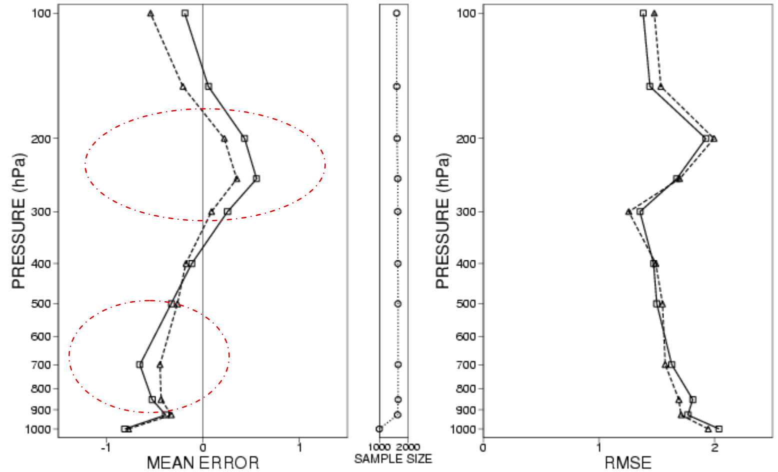
Data assimilation at CNMCA : LETKF vs 3D-VAR

ML4 vs 3D-VAR

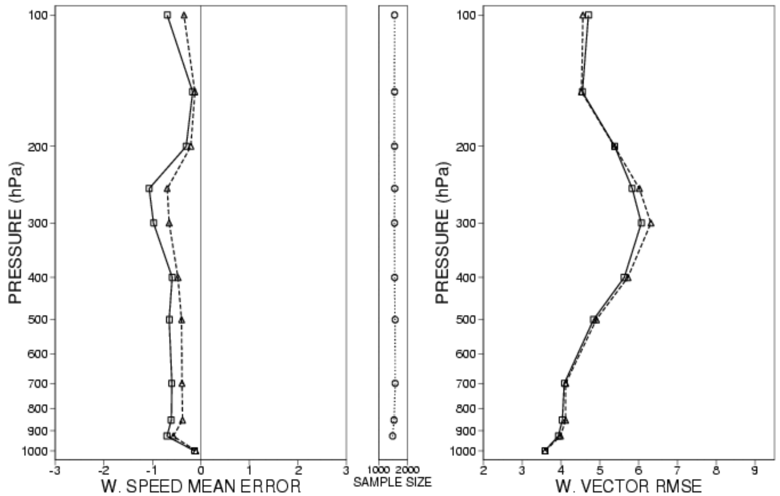
TEMPERATURE (°C) 00 UTC FC +24 h
Verification from 04/11/07 to 30/11/07



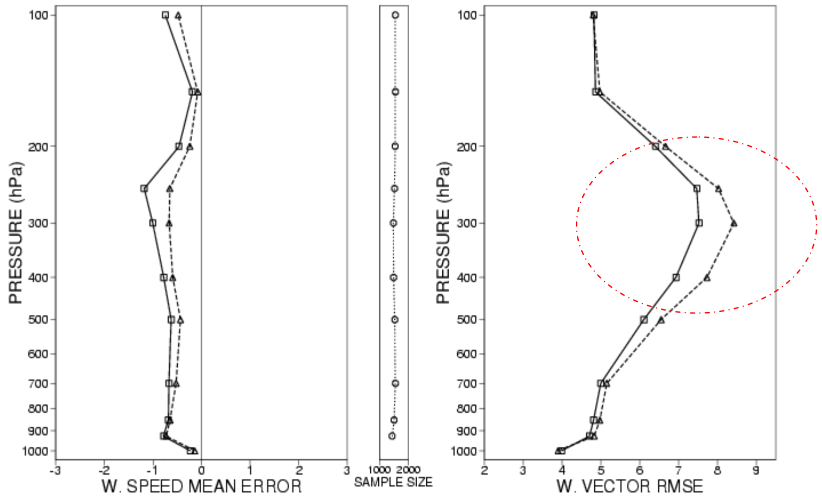
TEMPERATURE (°C) 00 UTC FC +48 h
Verification from 04/11/07 to 30/11/07



WIND (m/s) 00 UTC FC +24 h
Verification from 04/11/07 to 30/11/07

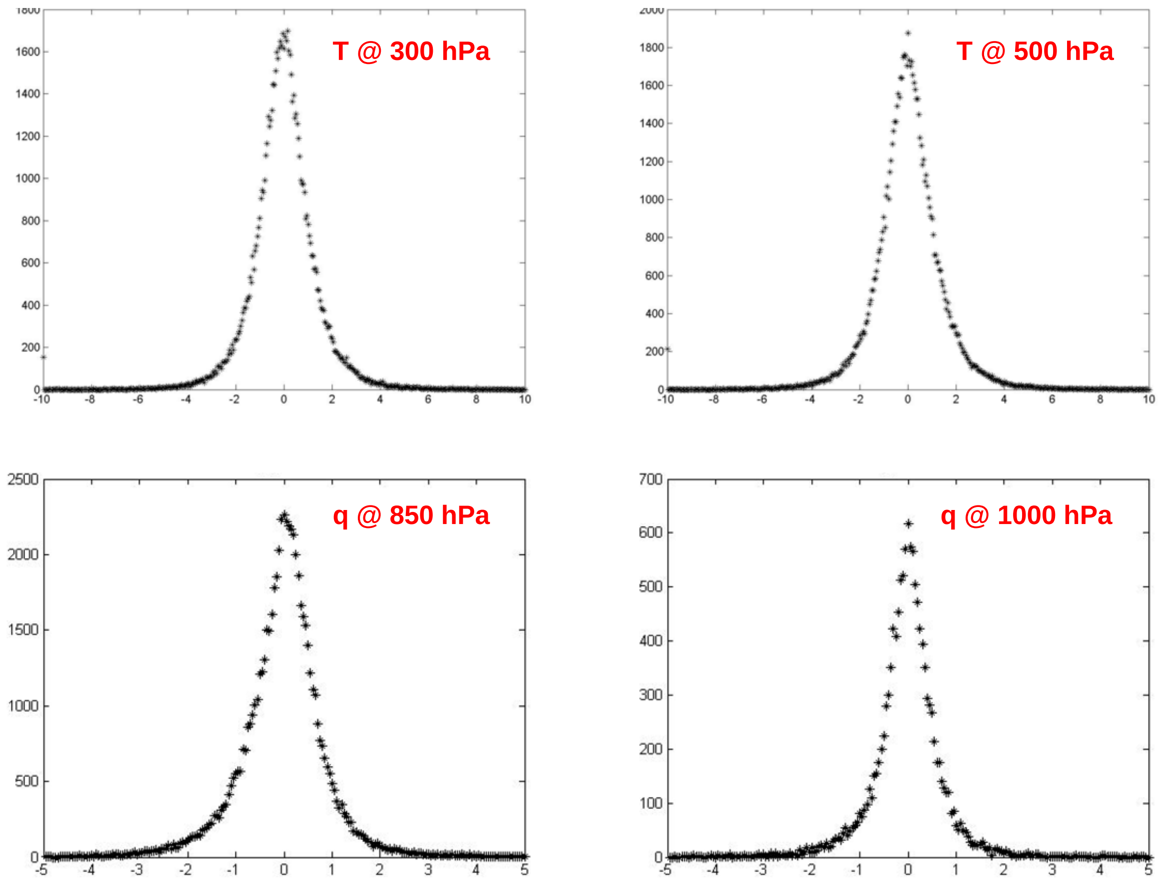


WIND (m/s) 00 UTC FC +48 h
Verification from 04/11/07 to 30/11/07



EnKF at CNMCA : ENSEMBLE DISTRIBUTION

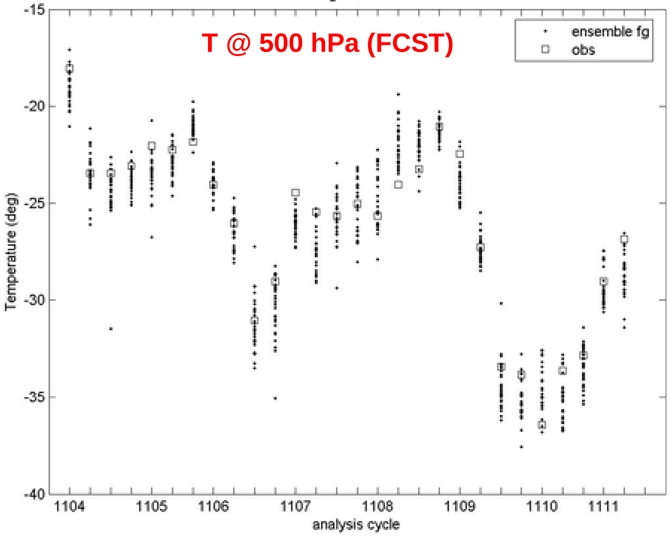
- Deterministic ensemble square root filters show a tendency to collapse onto a single state
- Need verification of gaussianity assumption in the forecast and analysis ensemble

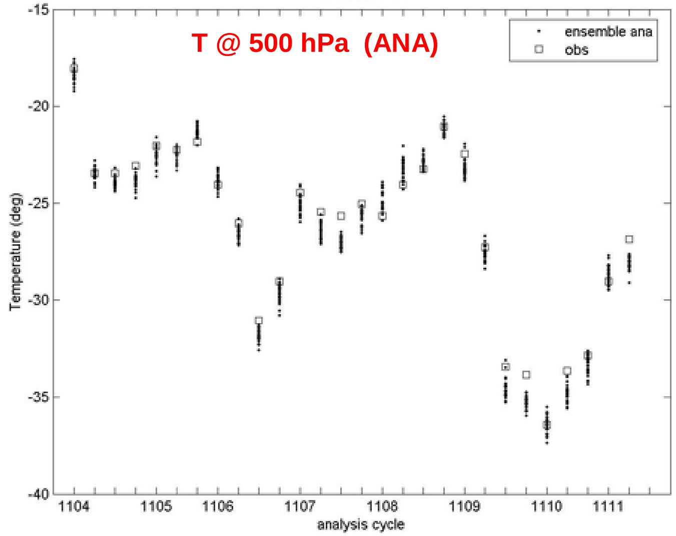


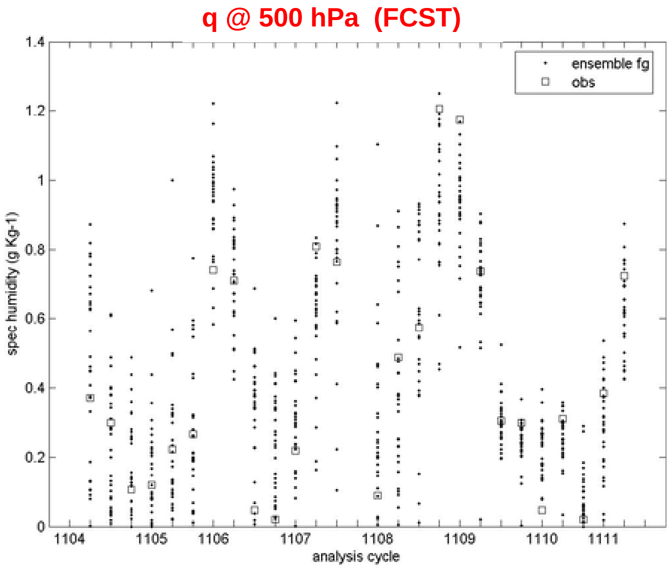
Forecast t+6h ensemble distribution around the mean at all RAOB locations

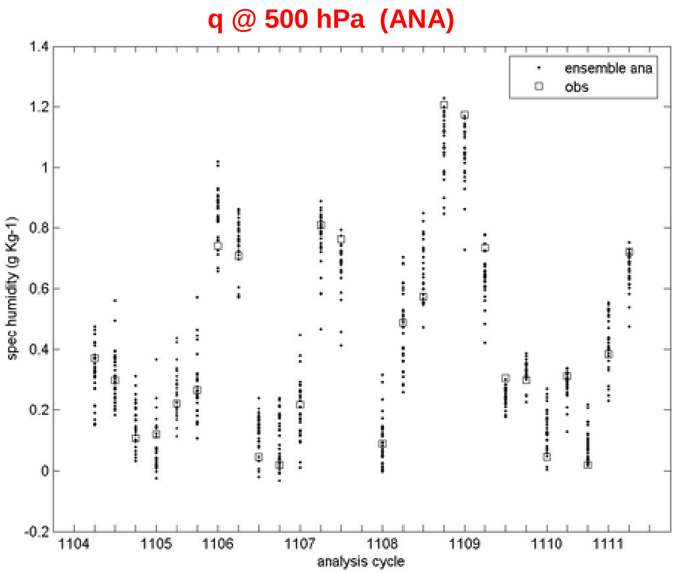
EnKF at CNMCA : ENSEMBLE DISTRIBUTION

TIME SERIES of FORECAST and ANALYSIS ensemble distribution @ RAOB 11520





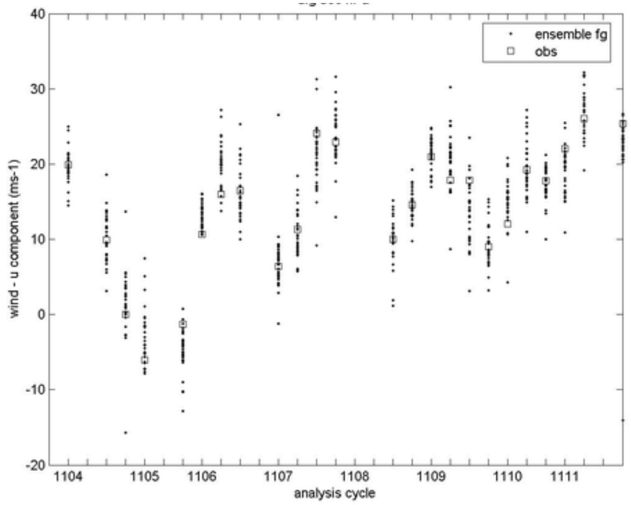




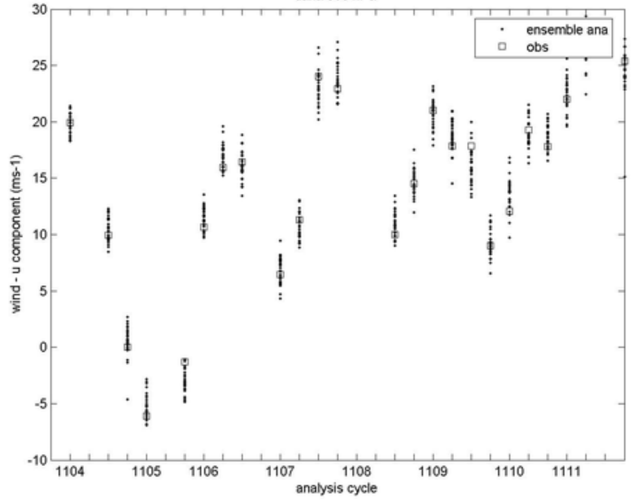
EnKF at CNMCA : ENSEMBLE DISTRIBUTION

TIME SERIES of FORECAST and ANALYSIS ensemble distribution @ RAOB 11520

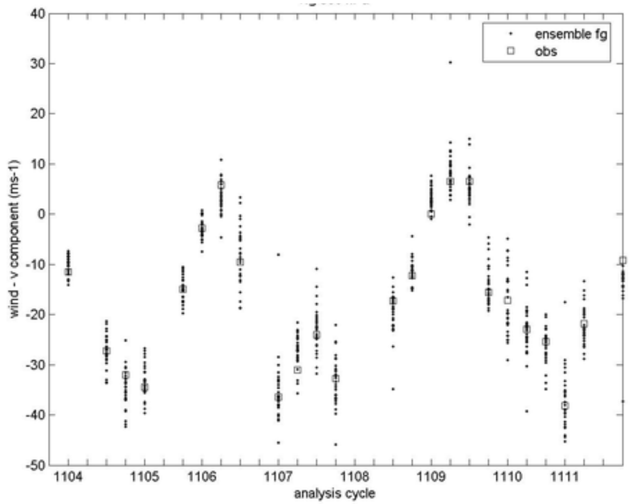
u @ 500 hPa (FCST)



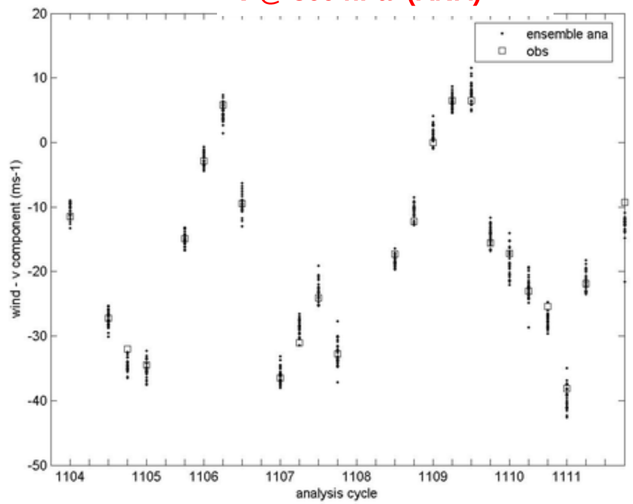
u @ 500 hPa (ANA)



v @ 500 hPa (FCST)

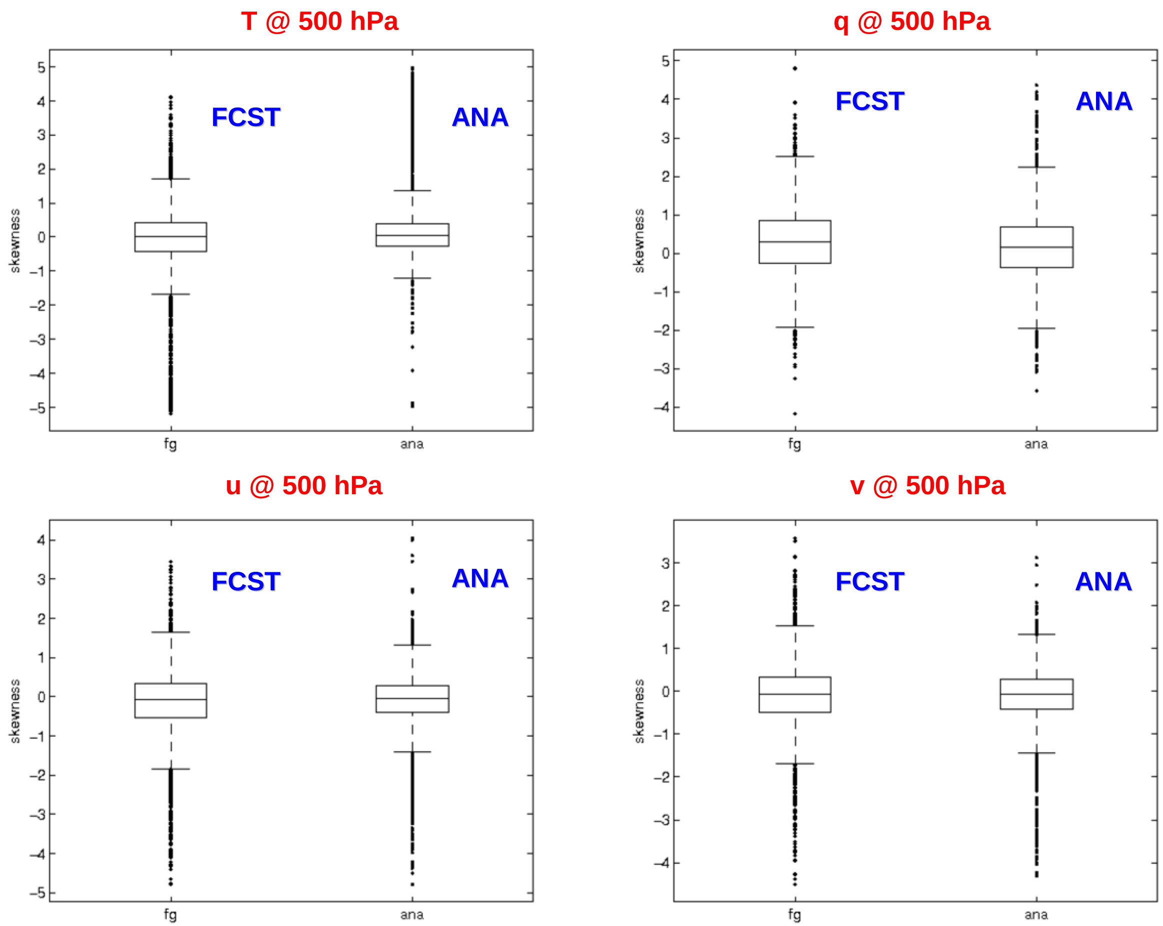


v @ 500 hPa (ANA)



EnKF at CNMCA : ENSEMBLE DISTRIBUTION

SKEWNESS DISTRIBUTION FOR ENSEMBLE FORECAST **PDF**



RESULTS DISCUSSION

- EnKF has proved to be relatively easy to implement, stable, with good computational scalability
- LETKF based forecasts generally outperform 3DVar in terms of RMSE metric
- LETKF confirms more sensitivity to model systematic errors
- A combination of additive and (adaptively) multiplicative covariance inflation techniques seems adequate to combat filter divergence symptoms and provide a reliable first-guess ensemble
- Both ensemble forecasts and analysis distributions look close to Gaussian

FUTURE IMPROVEMENTS

- Treatment of nonlinearities based on OUTER LOOP iterations
- Treatment of MODEL BIAS
- Assimilation of Radiance observations
- Perturbation of boundary conditions