Model Error in the ECMWF Ensemble of Data Assimilations

Massimo Bonavita

ECMWF

Acknowledgments: Lars Isaksen, Martin Leutbecher

ECMWF

Outline

- Estimation of analysis and background error statistics with the Ensemble of Data Assimilations (EDA)

- Parameterizations of model error in the ECMWF EDA

- Impact of model error parameterizations on the EDA sample statistics

- Conclusions and Plans

ECMWF

The EDA method

For a linear system the data assimilation update is:

$$\mathbf{x}_a^k = \mathbf{x}_b^k + \mathbf{K}_k\left(\mathbf{y}^k - \mathbf{H}_k\mathbf{x}_b^k\right)$$

$$\mathbf{x}_b^{k+1} = \mathbf{M}_k\mathbf{x}_a^k$$

Under the assumptions of statistically independent background ($\mathbf{P}^b$), observation ($\mathbf{R}$) and model errors ($\mathbf{Q}$), the evolution of the system error covariances is given by:

$$\mathbf{P}_k^a = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^b\left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$

$$\mathbf{P}_{k+1}^b = \mathbf{M}_k\mathbf{P}_k^a\mathbf{M}_k^T + \mathbf{Q}_k$$

ECMWF

The EDA method

Consider now the evolution of the same system where we perturb the observations and the forecast state (t_{k+1}) with random noise drawn from the respective error covariances:

$$\widetilde{\mathbf{x}}_a^k = \widetilde{\mathbf{x}}_b^k + \mathbf{K}_k\left(\mathbf{y}^k + \boldsymbol{\eta}_k - \mathbf{H}_k\mathbf{x}_b^k\right)$$

$$\mathbf{x}_b^{k+1} = \mathbf{M}_k\mathbf{x}_a^k + \boldsymbol{\zeta}_k$$

where $\boldsymbol{\eta}_k \sim \mathrm{N}(\mathbf{0},\mathbf{R})$, $\boldsymbol{\zeta}_k \sim \mathrm{N}(\mathbf{0},\mathbf{Q})$.
If we define the differences between the perturbed and unperturbed state $\boldsymbol{\varepsilon}_a \equiv \widetilde{\mathbf{x}}_a - \mathbf{x}_a$ and $\boldsymbol{\varepsilon}_b \equiv \widetilde{\mathbf{x}}_b - \mathbf{x}_b$, their evolution and the evolution of their sample statistics are governed by the same equations (Fisher et al., 2005):

$$\boldsymbol{\varepsilon}_a^k = \boldsymbol{\varepsilon}_b^k + \mathbf{K}_k\left(\boldsymbol{\eta}_k - \mathbf{H}_k\boldsymbol{\varepsilon}_b^k\right)$$

$$\boldsymbol{\varepsilon}_b^{k+1} = \mathbf{M}_k\boldsymbol{\varepsilon}_a^k + \boldsymbol{\zeta}_k$$
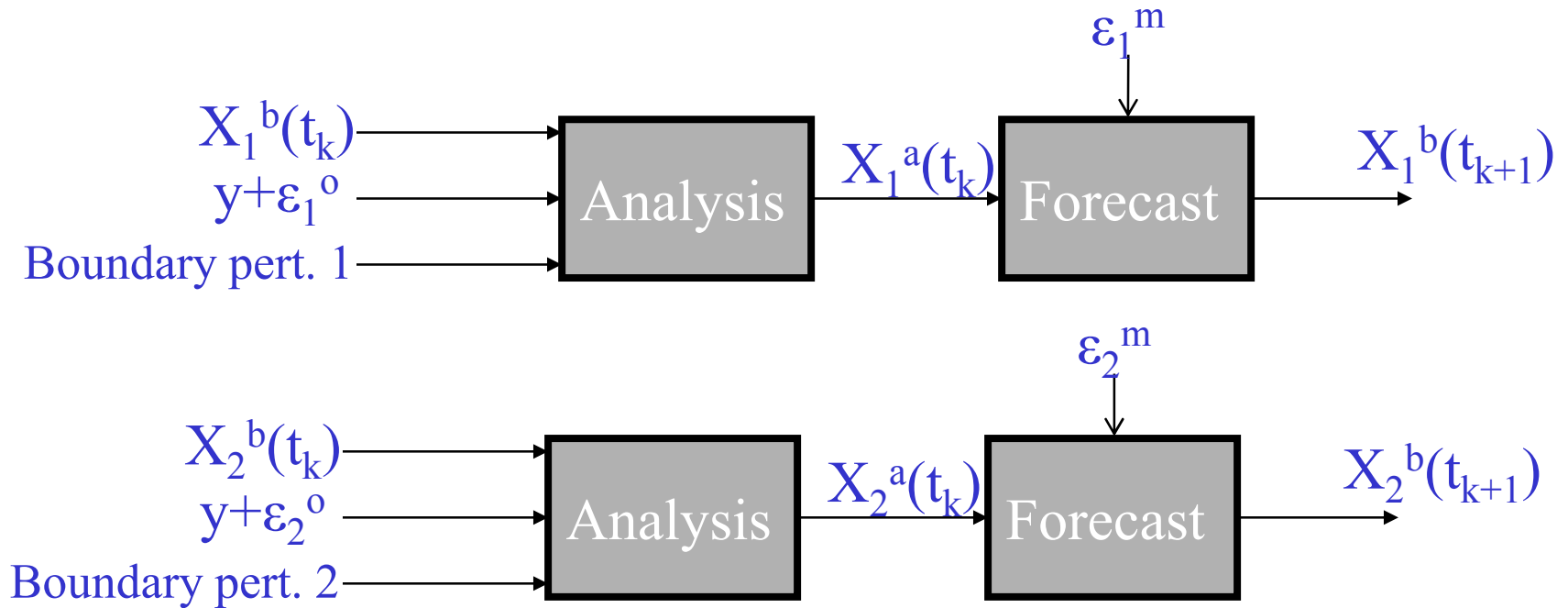
ECMWF

The EDA method

$$\left\langle \boldsymbol{\varepsilon}_k^a \left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \left\langle \boldsymbol{\varepsilon}_k^b \left(\boldsymbol{\varepsilon}_k^b\right)^T \right\rangle \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T$$

$$\left\langle \boldsymbol{\varepsilon}_{k+1}^b \left(\boldsymbol{\varepsilon}_{k+1}^b\right)^T \right\rangle = \mathbf{M}_k \left\langle \boldsymbol{\varepsilon}_k^a \left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle \mathbf{M}_k^T + \mathbf{Q}_k$$

This implies that:

1. We can use an ensemble of perturbed assimilation cycles to simulate the errors of our reference assimilation cycle;

2. The ensemble of perturbed DAs should be as similar as possible to the reference DA (i.e., same or similar **K** matrix)

3. The applied perturbations $\boldsymbol{\eta}_k$, $\boldsymbol{\zeta}_k$ must have the required error covariances (**R**, **Q**);

ECMWF

The EDA method

$$\varepsilon_1{}^m$$

$$X_1{}^b(t_k)$$
$$y+\varepsilon_1{}^o$$
Boundary pert. 1

Analysis

$$X_1{}^a(t_k)$$

Forecast

$$X_1{}^b(t_{k+1})$$

$$\varepsilon_2{}^m$$

$$X_2{}^b(t_k)$$
$$y+\varepsilon_2{}^o$$
Boundary pert. 2

Analysis

$$X_2{}^a(t_k)$$

Forecast

$$X_2{}^b(t_{k+1})$$

....

ECMWF

The EDA method

- **10** ensemble members using 4D-Var assimilations

- **T399** outer loop, T95/T159 inner loops. (Reference DA: T1279 outer loop, T159/T255/T255 inner loops)

- Perturbation applied to all relevant sources of uncertainty:

   1. **Observations** randomly perturbed according to their prescribed **R** matrix;

   2. **SST** perturbed with climatological error structures

   3. **Model error** represented by stochastic methods

ECMWF

Model error representation

- In the EDA we do not represent model error through a covariance matrix (Kalman filter – Weak constraint 4D-Var) but with physically plausible Monte Carlo realizations

- Two model error parameterizations are available in the IFS:
  1. Stochastically Perturbed Parameterization Tendencies(**SPPT**)
  2. Stochastic Kinetic Energy Backscatter (**SKEB**)

- They are both used in the operational Ensemble Prediction System (EPS)

- Only the SPPT parameterization is currently used in operational EDA

- Goal: Converge toward a unified representation

ECMWF

Model error representation

Stochastically Perturbed Parameterization Tendencies (SPPT)

- **Physics tendencies** P perturbed by ΔP = rP, with r a random pattern

- Gaussian distribution, truncated at 2 (instead of uniform distr.)

- Same pattern r for T; q; u; v

- Random pattern r uses AR-1 processes in spectral space and is smooth in space and time

- Three components with different correlation scales: 6 h, 3 d, 30 d and 500 km, 1000 km, 2000 km

- Improved version of the original SPPT scheme (stochastic physics, Buizza, Miller & Palmer (1999))

from: M. Leutbecher

Model error representation

Stochastic Kinetic Energy Backscatter (SKEB)

- Rationale: A fraction of the dissipated energy is backscattered upscale and acts as **streamfunction forcing** for the resolved-scale flow (Shutts and Palmer 2004, Shutts 2005, Berner et al. 2009)

$$\text{Streamfunction forcing} = [bD]^{1/2} \ F(x; t);$$

  where b,D,F denote the backscatter ratio, the (smoothed) total dissipation rate and the 3-dim evolving pattern

- Total dissipation rate: sum of
  1. "Numerical" dissipation: loss of KE by numerical diffusion + interpolation in semi-Lagrangian advection;
  2. Dissipation from orographic gravity wave drag parameterization;
  3. An estimate of the deep convective KE production
- Boundary layer dissipation is omitted
- see Tech Memo 598, Palmer et al. (2009) for details

from: M. Leutbecher

ECMWF

Impact of model error parameterizations

- The impact of different model error parameterizations can be evaluated in the context of an EDA

- This is arguably a different and more stringent test than is possible with an EPS:

  1. Effects accumulate over assimilation cycles;

  2. Background errors for use in the deterministic analysis require the estimation of **multivariate** pdfs;

  3. Background errors span a larger portion of phase space than forecast errors at longer lead times

ECMWF

Impact of model error parameterizations

- To test these ideas three EDA cycles have been run with different model error parameterizations:

    1. **fgk7**: no model error parameterizations;

    2. **fi8s**: SPPT parameterization;

    3. **fgk6**: SPPT + SKEB parameterizations

- All other aspect of the EDA setup are equal

- Results shown are time averages over a 20 day period (i.e., 40 assimilation cycles; 20100405 – 20100425)

ECMWF

Impact of model error parameterizations

EDA Temperature spread – no model error (fgk7)

Mean Vertical
EDA T spread
Cross Section
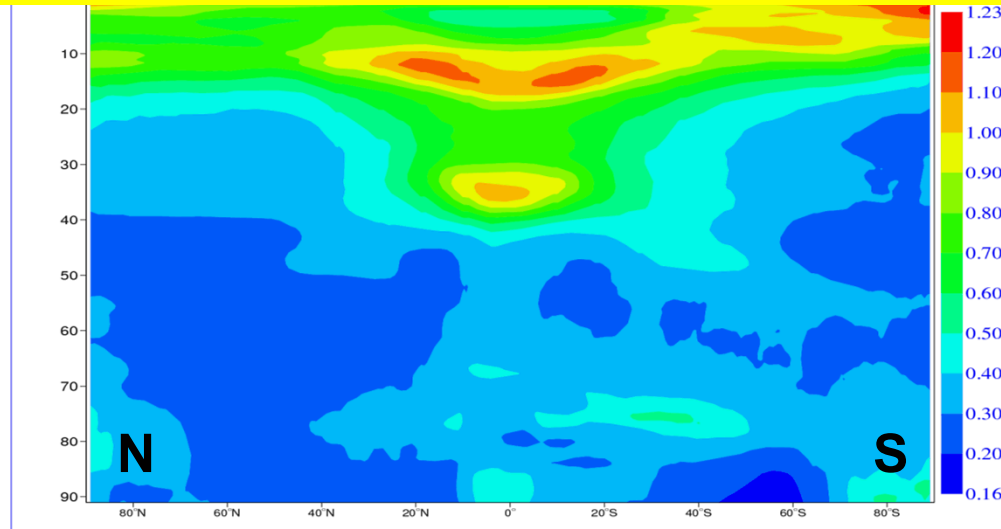
Impact of SPPT on T spread (fi8s-fgk7)

Addit. Impact of SKEB on T spread (fgk6-fi8s)

Impact of model error parameterizations

EDA Vorticity spread – no model error (fgk7)

Mean Vertical
EDA Vo spread
Cross Section

Average of vort 20100405 600 step 3 Expver FGK7 (180.0W-180.0E)

N   S
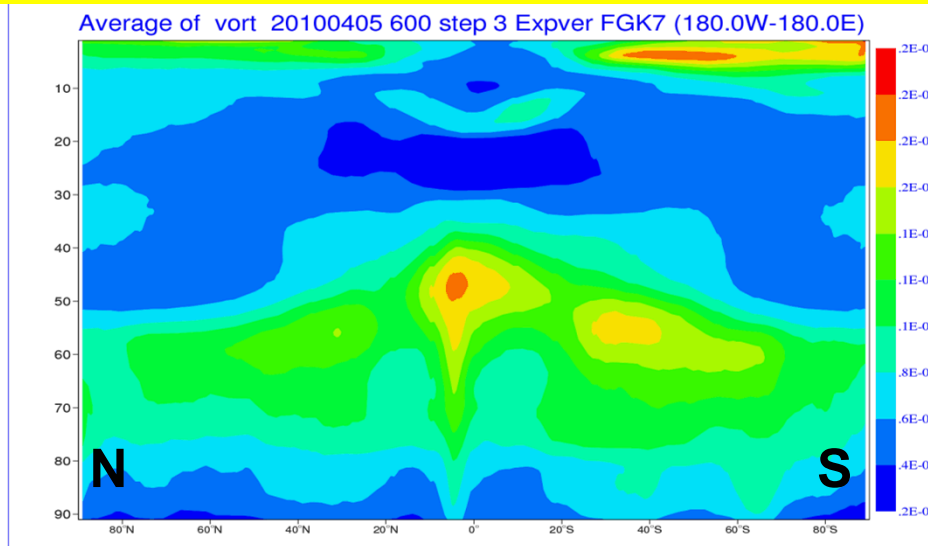
Impact of SPPT on VO spread (fi8s-fgk7)

Average of vort 20100405 600 step 3 Expver fi8s (180.0W-180.0E)

Addit. Impact of SKEB on VO spread (fgk6-fi8s)

Average of vort 20100405 600 step 3 Expver fgk6 (180.0W-180.0E)

Impact of model error parameterizations
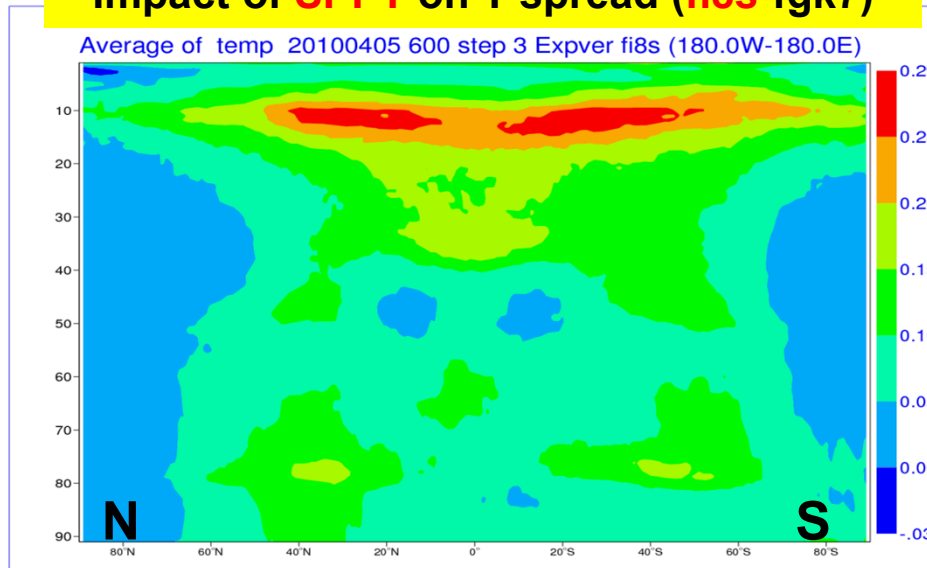
Impact of SPPT
EDA T spread

Impact of **SPPT** on T spread (**fi8s**-fgk7)

SPPT on T spread: model lev 48 (200 hPa)

SPPT on T spread: model lev 78 (850 hPa)

Impact of model error parameterizations

Impact of SKEB
EDA T spread
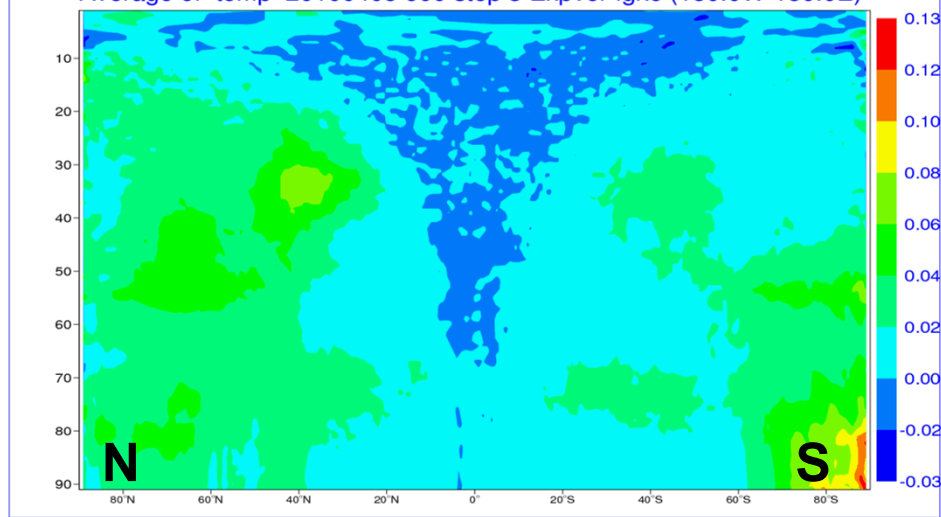
SKEB on T spread: model lev 38 (100 hPa)

**SPPT on T spread: model lev 78 (850 hPa)**
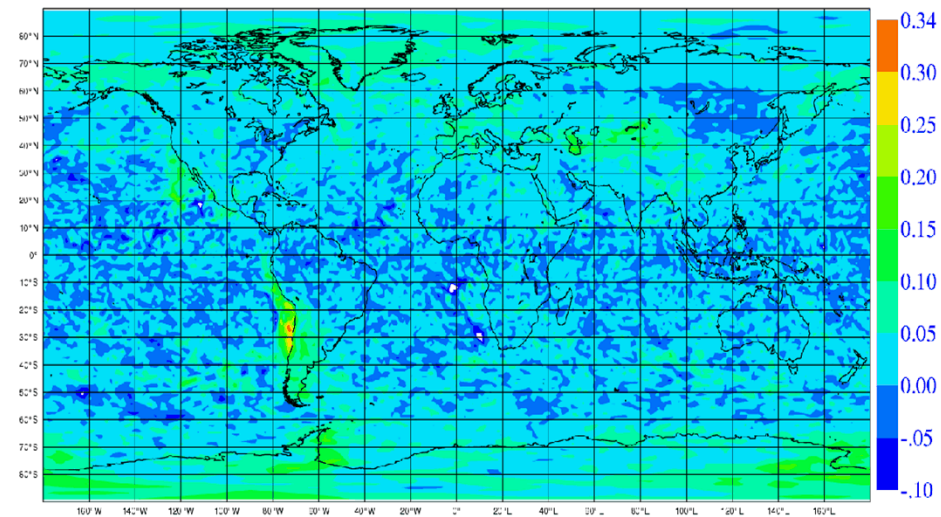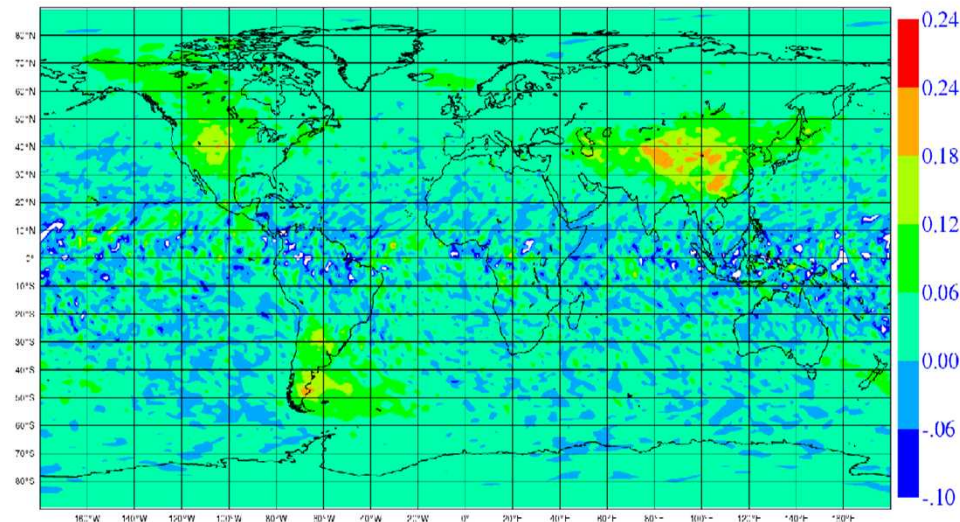
Impact of model error parameterizations

Impact of SPPT
EDA Vo spread

Impact of SPPT on VO spread (fi8s-fgk7)
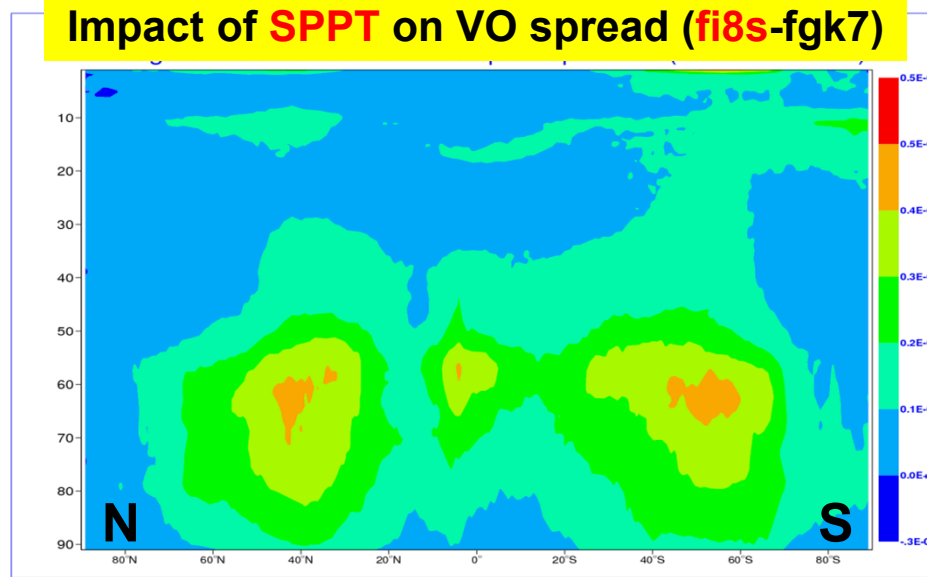
SPPT on VO spread: model lev 64 (500 hPa)
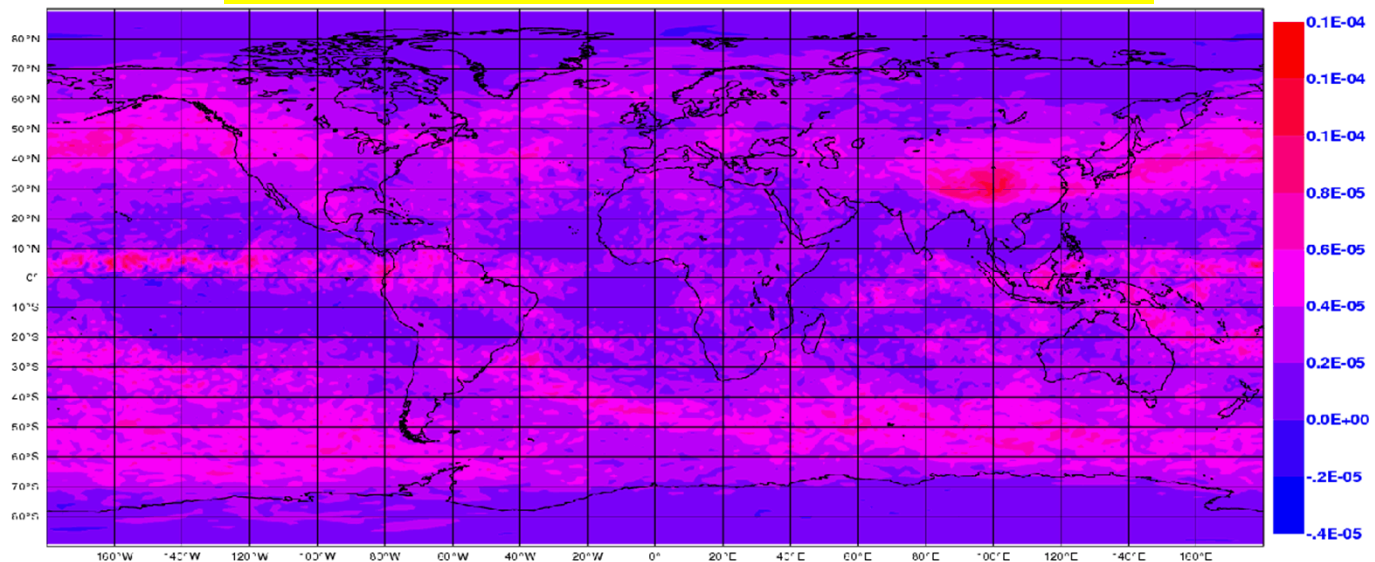
Mod

Impact of model error parameterizations

Impact of SKEB
EDA Vo spread
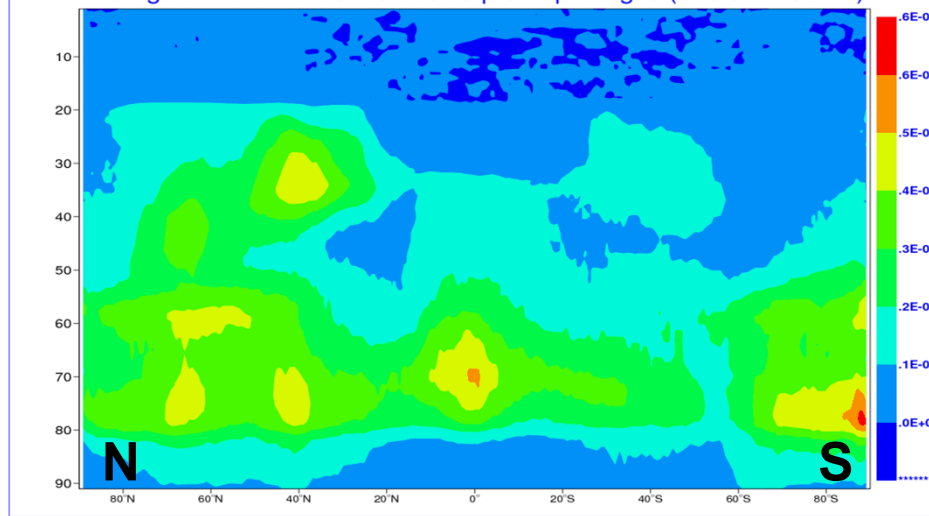
SKEB on VO spread: model lev 30 (50 hPa)

SKEB on VO spread: model lev 78 (850 hPa)
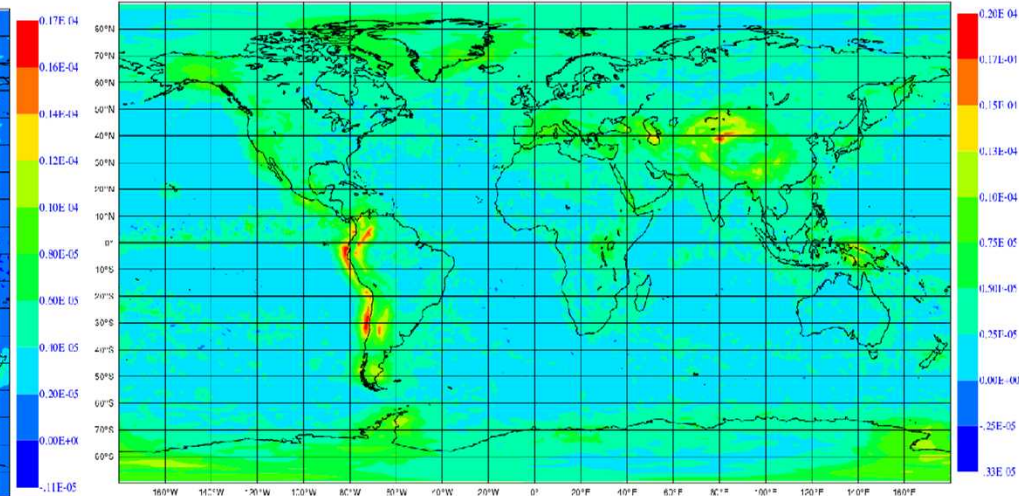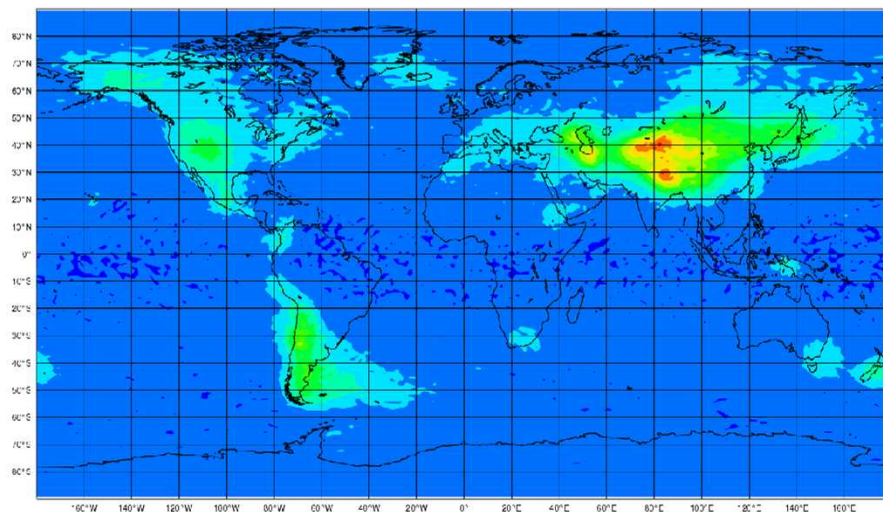
Impact of model error parameterizations

BG Errors triggered by orographic wave activity are seen in analysis increments maps
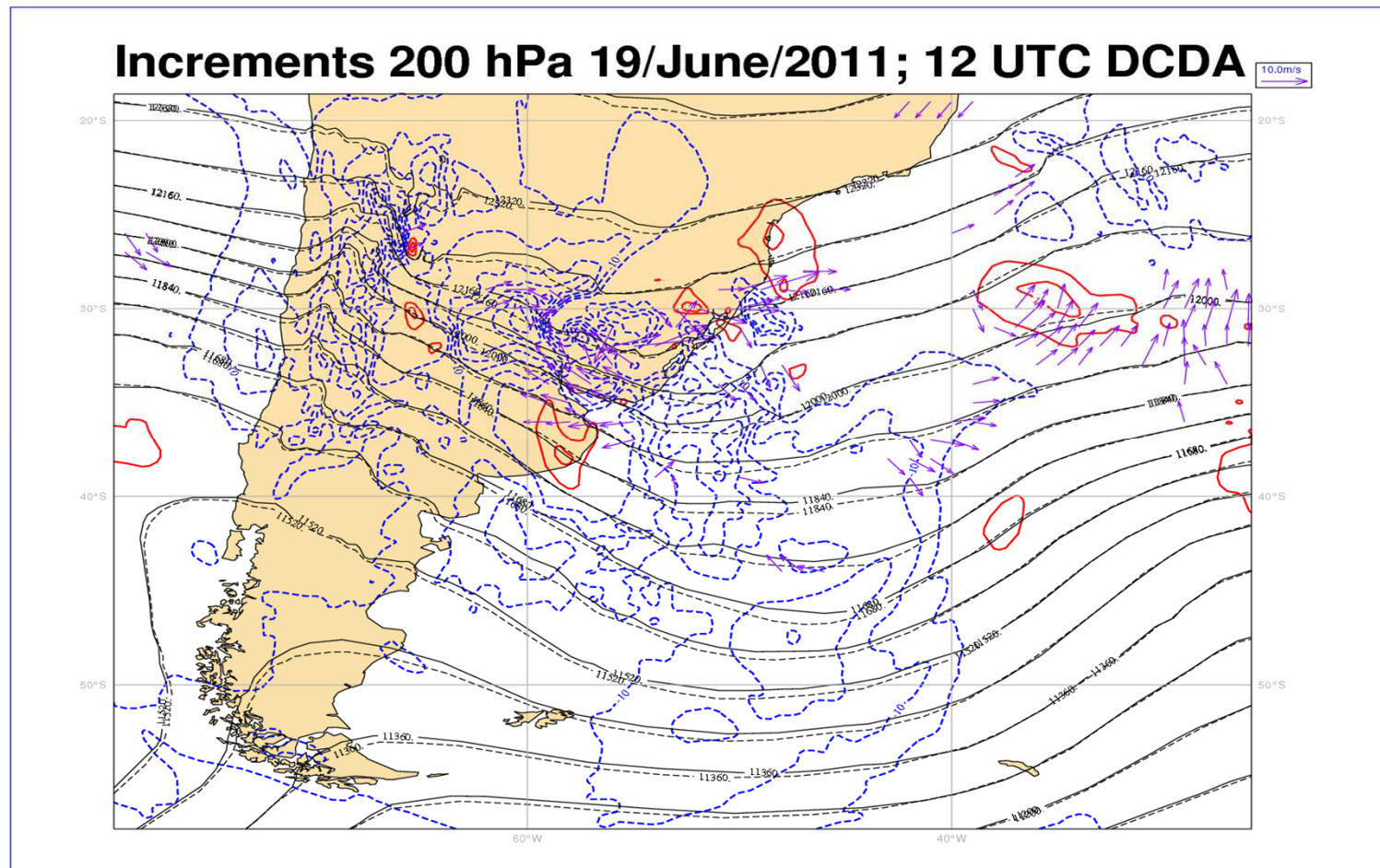
Increments 200 hPa 19/June/2011; 12 UTC DCDA

Impact of model error parameterizations

Does the use of model error param. make the EDA spread a "better" predictor of background error StDev?
(perceived background errors = control bg – operational ana)

BG Error Temperature lev 78 (850 hPa)
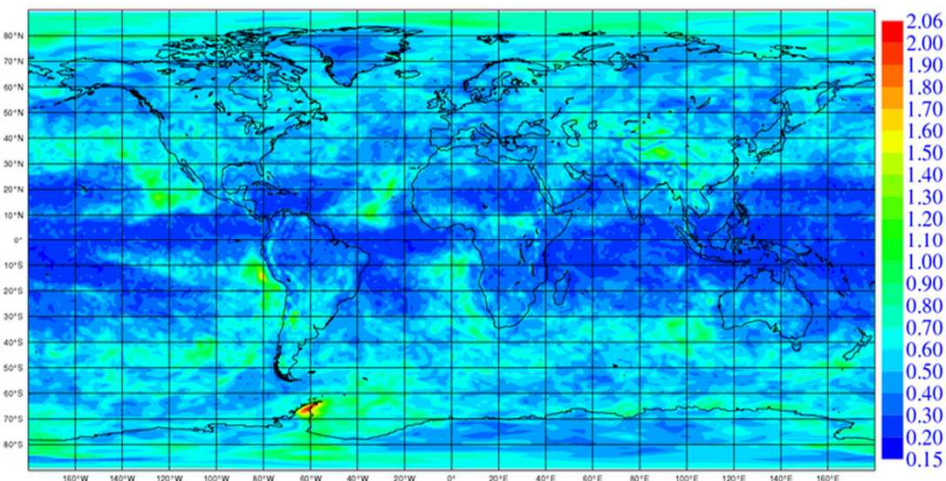
Impact of SPPT + SKEB on EDA T spread lev 78 (850 hPa)

ECMWF
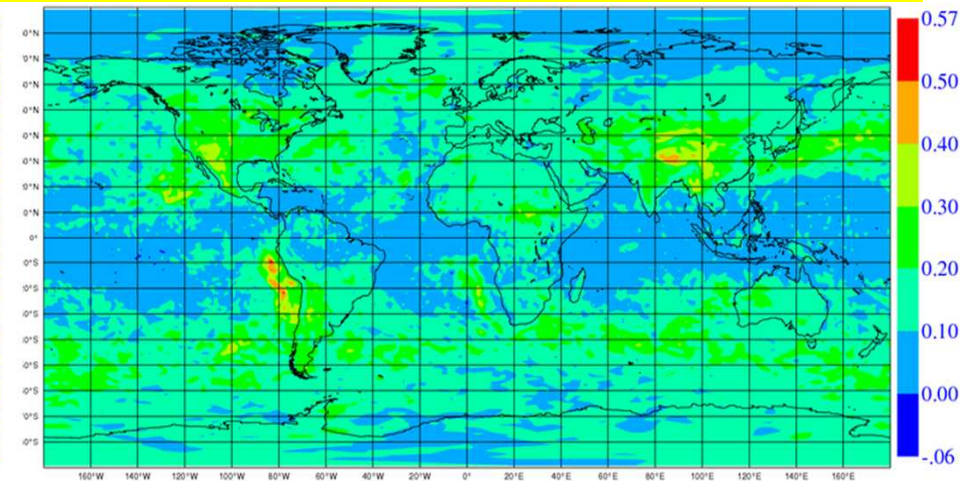
Impact of model error parameterizations
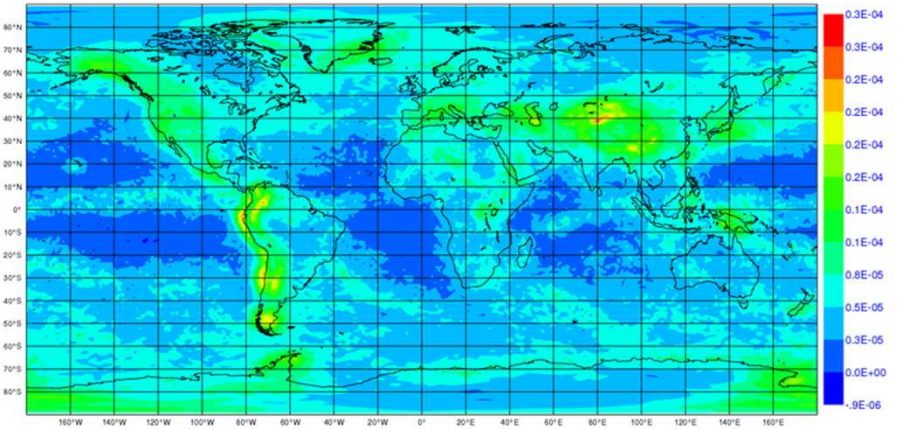

BG Error Vorticity lev 64 (500 hPa)
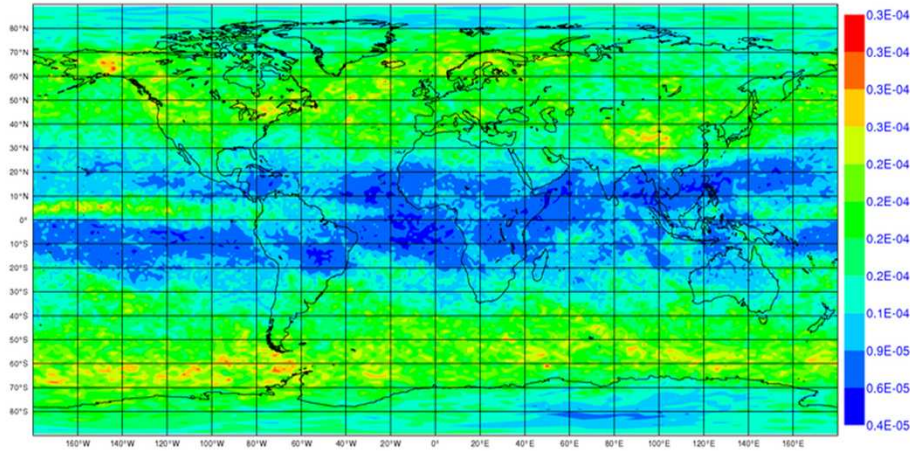
Impact of SPPT + SKEB on EDA Vo spread lev 64
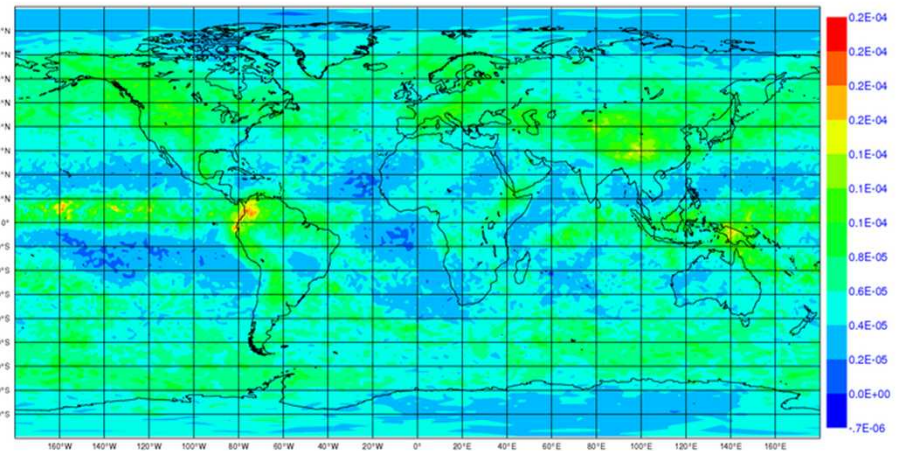
BG Error Vorticity lev 78 (850 hPa)
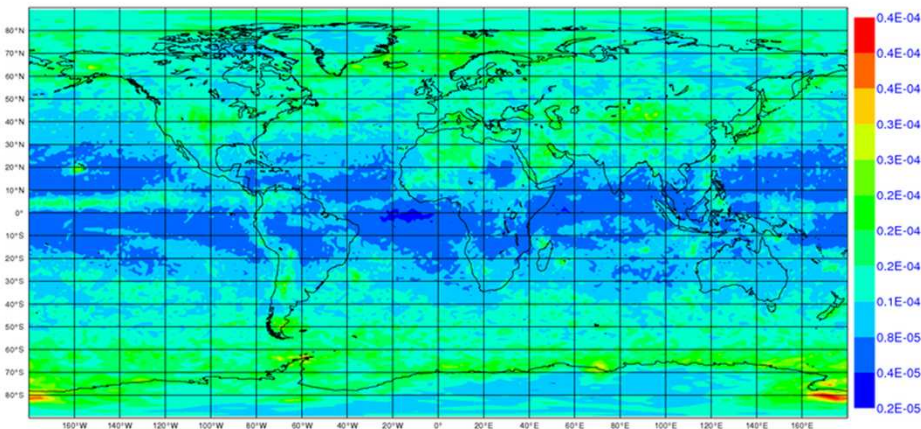
Impact of SPPT + SKEB on EDA Vo spread lev 78
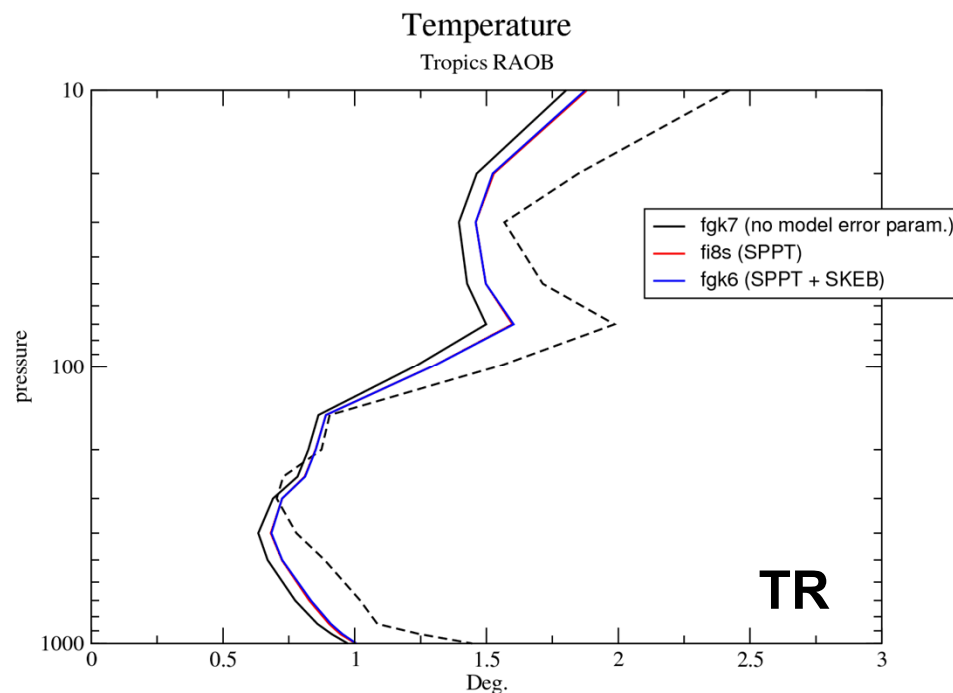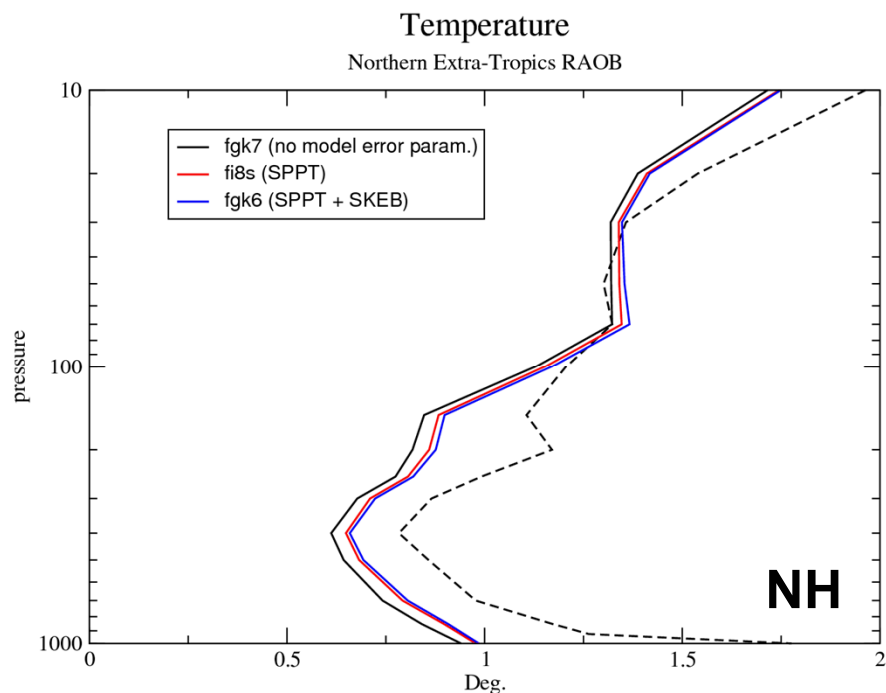
Impact of model error parameterizations

Does the use of model error param. make the EDA spread a "better" predictor of background errors StDev?

For a reliable (E)DA:

$$\left\langle \mathbf{d}_i \mathbf{d}_i^{\mathbf{T}} \right\rangle = \mathbf{HB}_i \mathbf{H}^T + \mathbf{R}$$

$$diag\left(\left\langle \mathbf{d}_i \mathbf{d}_i^{\mathbf{T}} \right\rangle\right) = diag\left(\mathbf{HB}_i \mathbf{H}^T\right) + diag(\mathbf{R}) \qquad (1)$$

Innovation Variance = Expected Innovation Variance

ECMWF

Impact of model error parameterizations

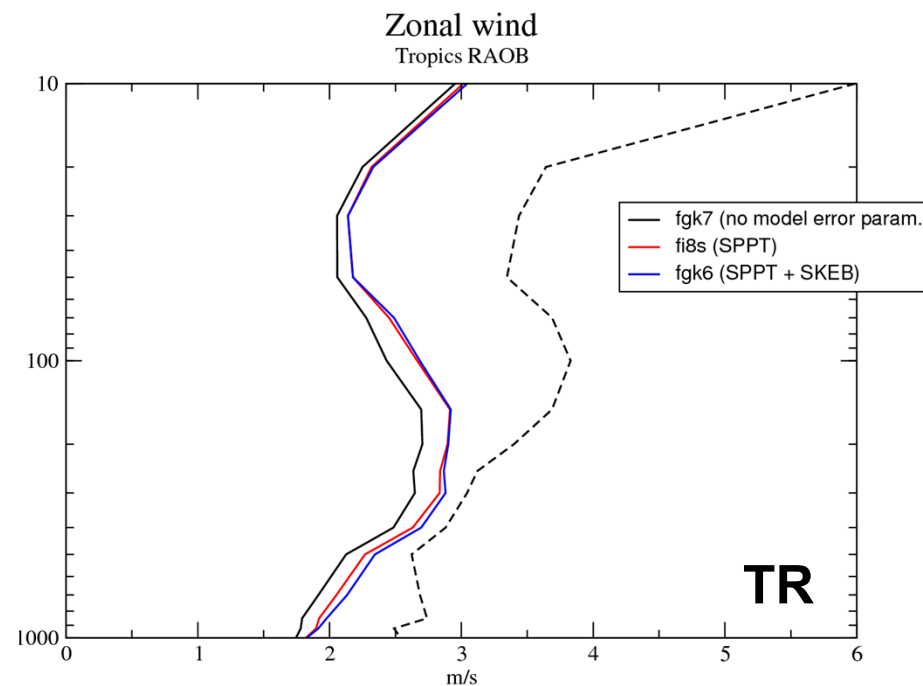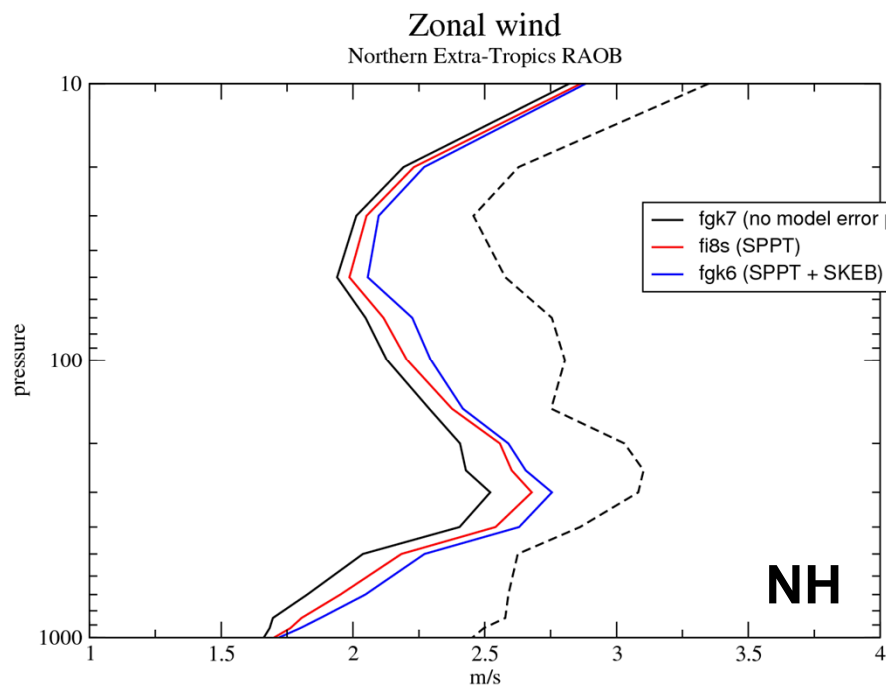
Radiosonde temperature obs.

Innovation StDev (dashed line)

Expected Innovation StDev (continuous line)

Impact of model error parameterizations

Radiosonde zonal wind obs.

Innovation StDev (dashed line)

Expected Innovation StDev (continuous line)

Impact of model error parameterizations

**Does the use of model error param. make the EDA spread a "better" predictor of background errors StDev?**
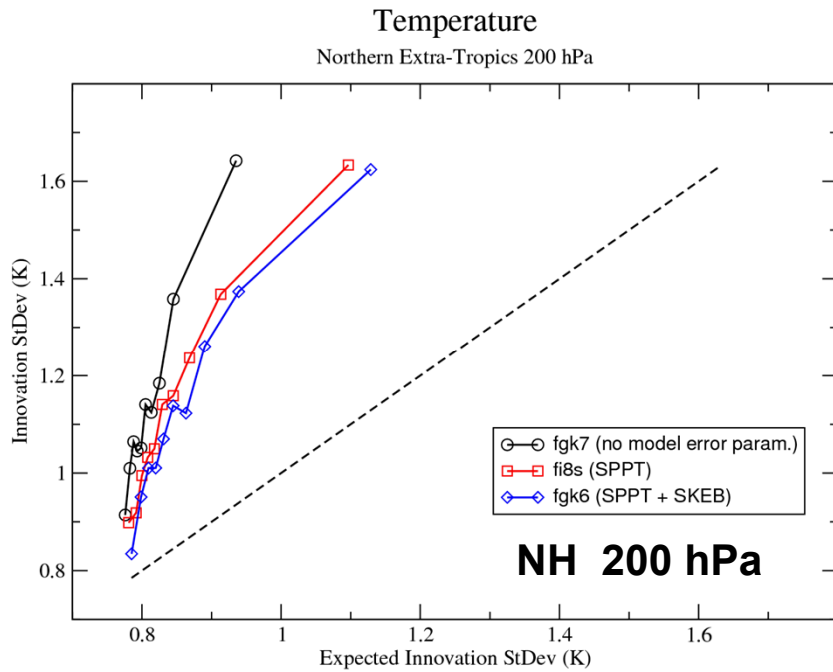
EDA spread should be a good predictor of the magnitude of background errors in a **statistical sense**: larger (smaller) EDA spread should correspond to larger (smaller) background errors, on average

We evaluate this by binning

$$ diag\left(\left\langle \mathbf{d}_i \mathbf{d}_i^{\mathrm{T}} \right\rangle\right) = diag\left(\mathbf{H}\mathbf{B}_i\mathbf{H}^T\right) + diag(\mathbf{R}) $$
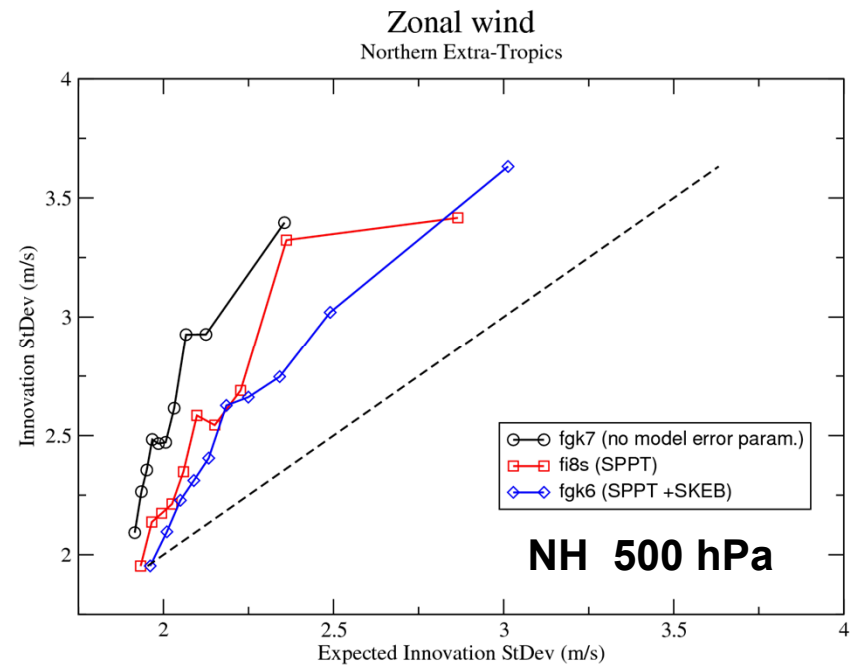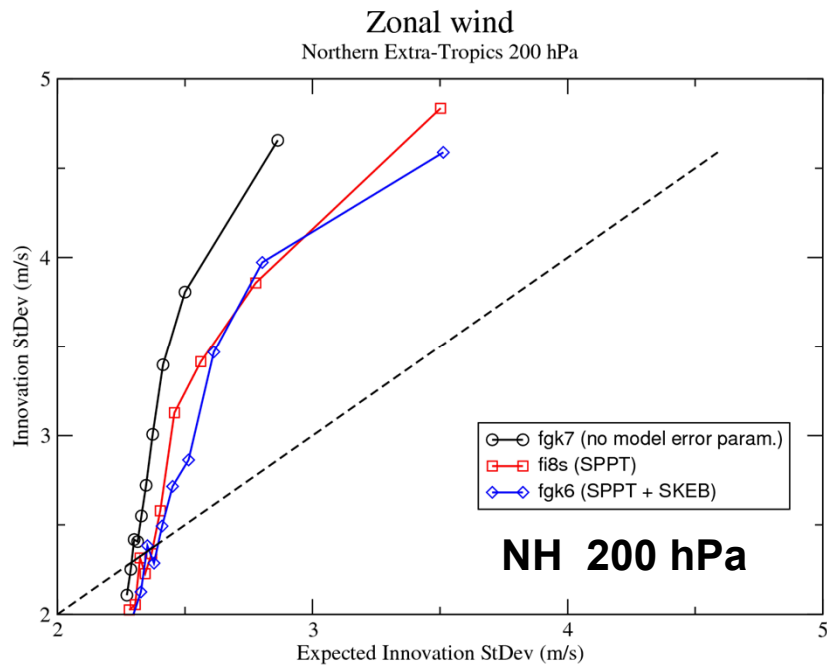
according to the magnitude of the expected innovation variance (Wang & Bishop, 2003)

ECMWF

Impact of model error parameterizations
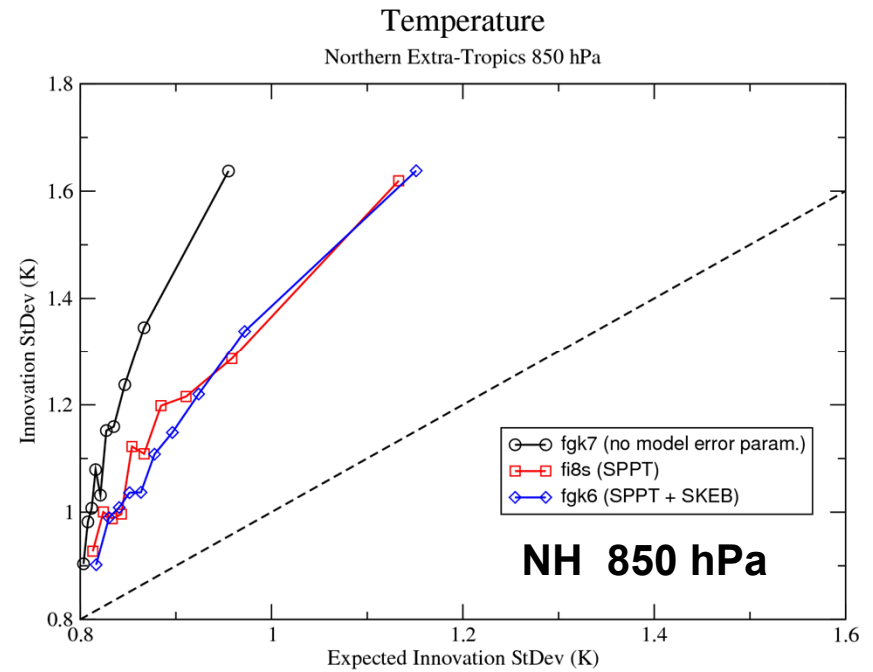
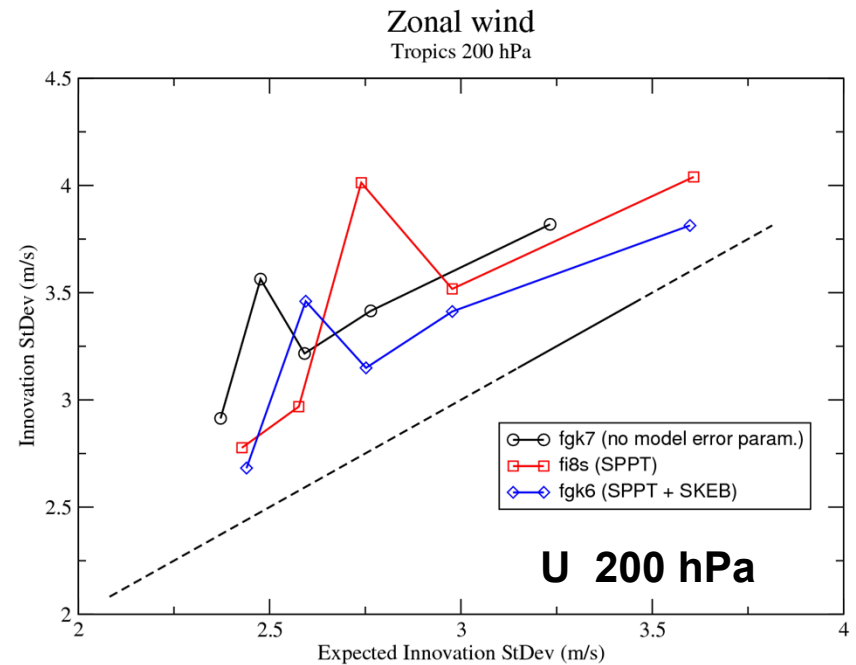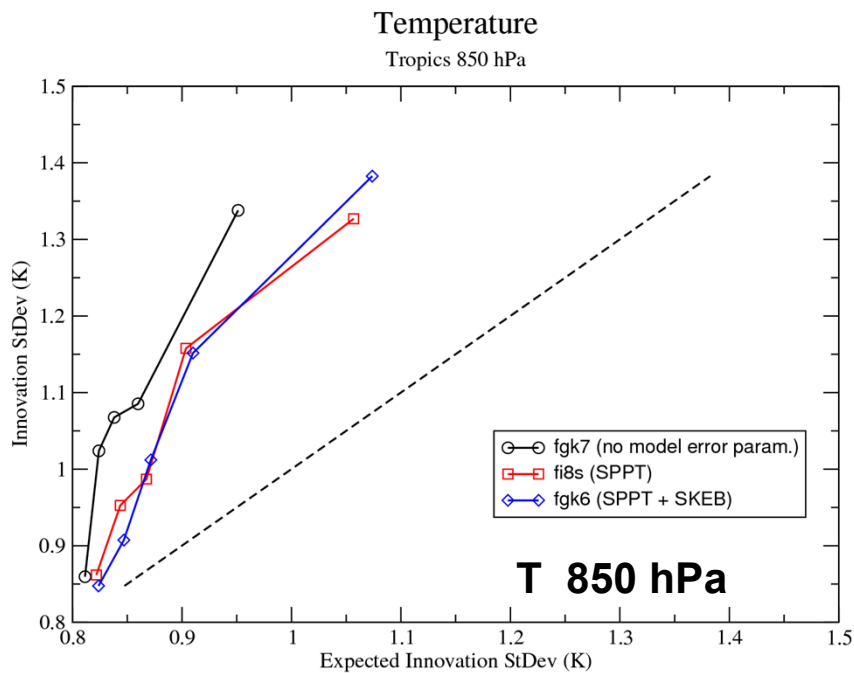

Radiosonde temperature obs.
Northern Extra-Tropics

Impact of model error parameterizations

Radiosonde zonal wind obs.

Northern Extra-Tropics

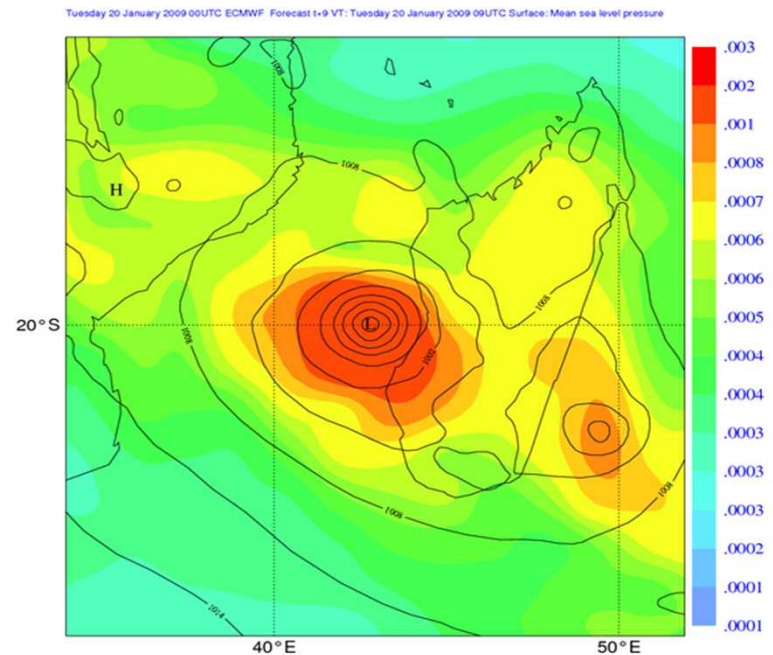# Impact of model error parameterizations

Radiosonde obs., Tropics
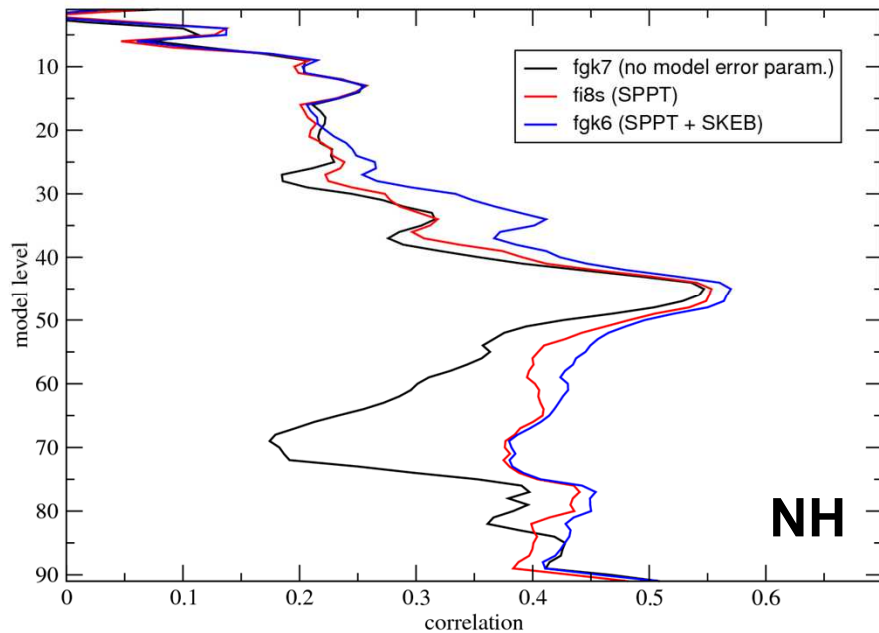
Impact of model error parameterizations

EDA spread should be able to predict the magnitude of background errors in a "**deterministic" sense**: areas where EDA spread is large (small) w.r. to its climatological mean should correspond to areas of larger (smaller) background uncertainty.

The effect of model error param. can be quantified by looking at the time averaged spatial correlation between EDA spread and perceived background errors
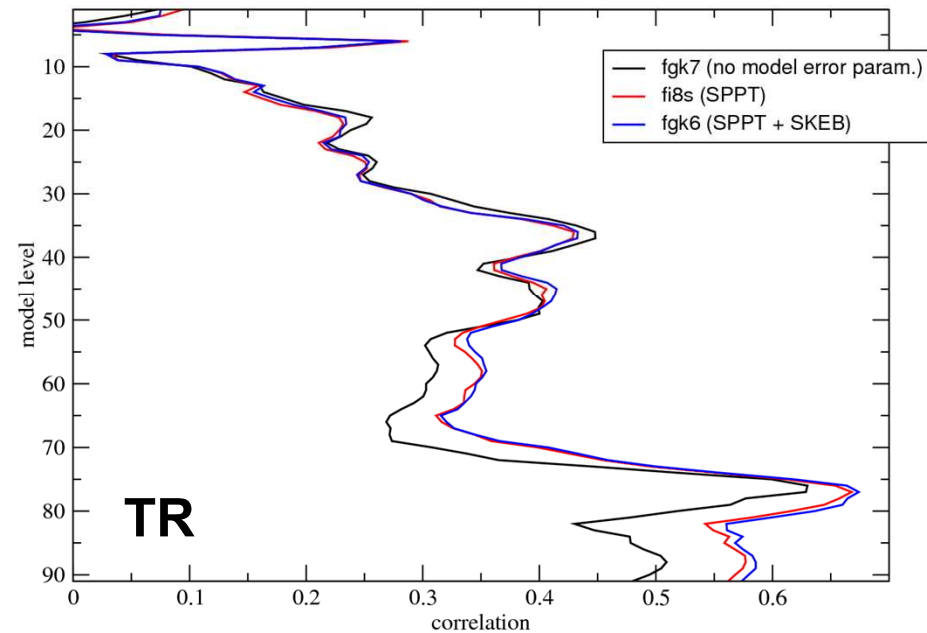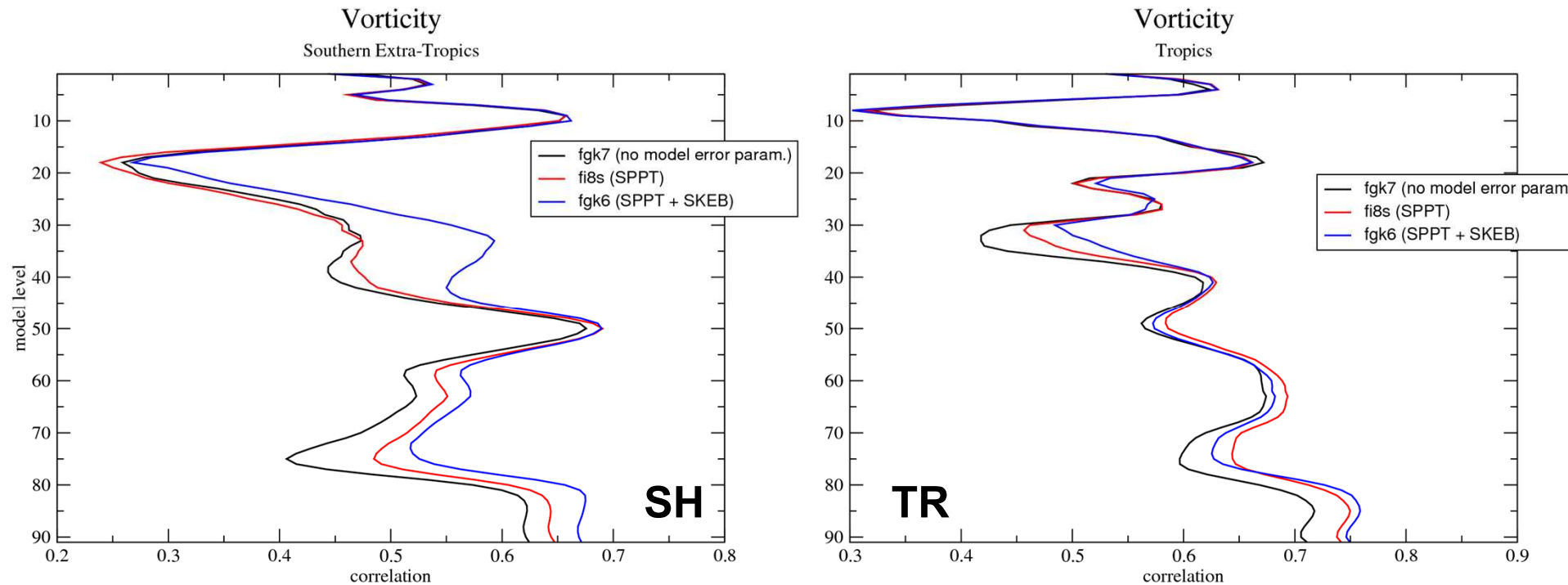
Impact of model error parameterizations

Time averaged spatial correlation between **temperature** EDA spread and perceived background errors

Impact of model error parameterizations

Time averaged spatial correlation between **vorticity** EDA spread and perceived background errors

Impact of model error parameterizations

EDA perturbations are also used to diagnose the **covariance** structures of background errors.

This is currently done offline to estimate a climatological **B** in 4D-Var, and it is envisaged to be done online to compute a flow-dependent **B** when a sufficiently large EDA will be available.

We would like the estimate how much the EDA perturbations project onto the background error patterns.

An approx. measure is the 'Perturbation versus Error Correlation Analysis' (**PECA**, Wei and Toth, 2003)

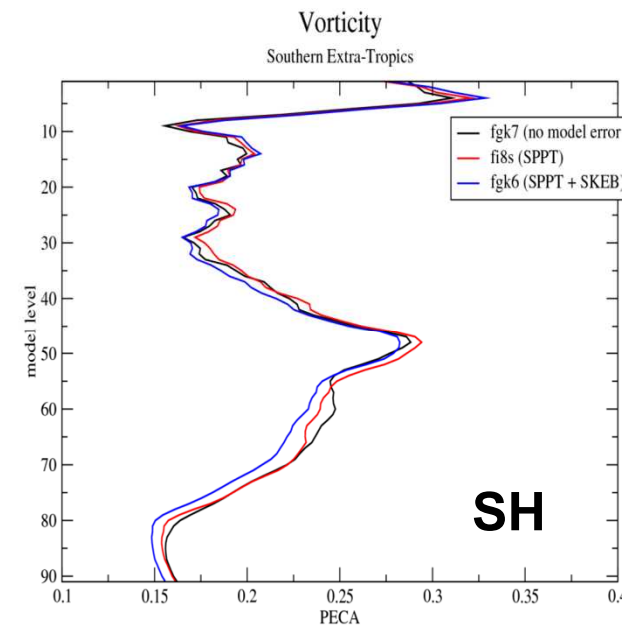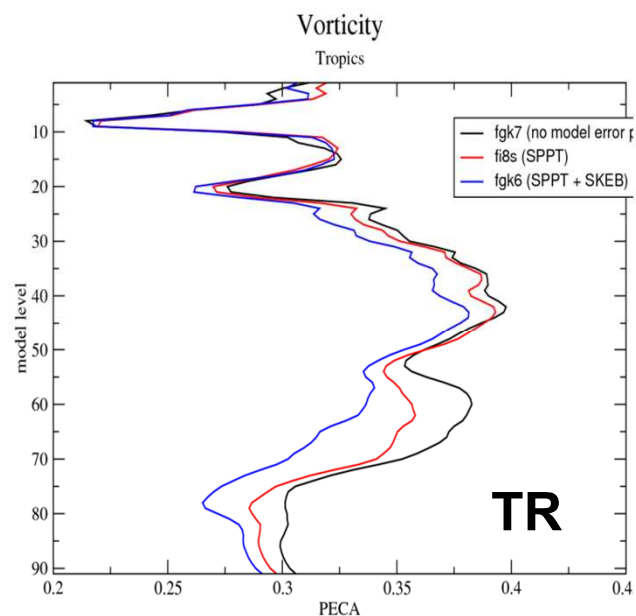**ECMWF**

Impact of model error parameterizations

$$PECA \equiv \frac{1}{N_{ens}} \sum_{j=1}^{Nens} cor\left(\varepsilon_{\mathbf{b}}, pert_j\right)$$

Where $\varepsilon_{\mathbf{b}}$ = control bg - verifying analysis,
$pert_j$ = member$_j$ bg – control bg

What is the impact of model error parameterizations on this diagnostic?

ECMWF

Impact of model error parameterizations

$$PECA \equiv \frac{1}{N_{ens}} \sum_{j=1}^{Nens} cor\left(\varepsilon_{\mathbf{b}}, pert_j\right)$$
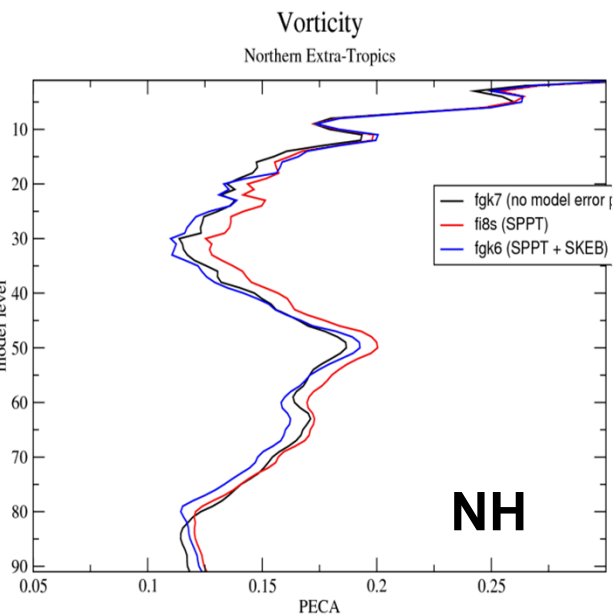
Should we be concerned?

Conclusions and Plans

- The EDA is arguably the best tool available to us to estimate analysis and background errors

- The EDA depends on a 'correct' specification of the sources of uncertainties: observation errors, model errors, boundary condition errors. This makes it a stringent test for errors parameterizations

- Model errors are represented in the EDA using physically plausible Monte Carlo realizations also used in the ECMWF EPS: SPPT and SKEB

- The use of model error parameterizations improves the statistical reliability of the EDA short range forecasts.

ECMWF

Conclusions and Plans

- EDA still underdispersive wrt radiosonde innovations (and own analysis): problems in **R** or model error tuned for EPS?

- The use of model error parameterizations improves the spatial correlation of the EDA spread with the 'perceived' background errors standard deviation

- The use of the SKEB param. in addition to the SPPT gives additional benefit, especially for the representation of wind field errors

- These positive diagnostics indications will be verified in an assimilation experiment using errors from an EDA with same model error parameterizations as in EPS
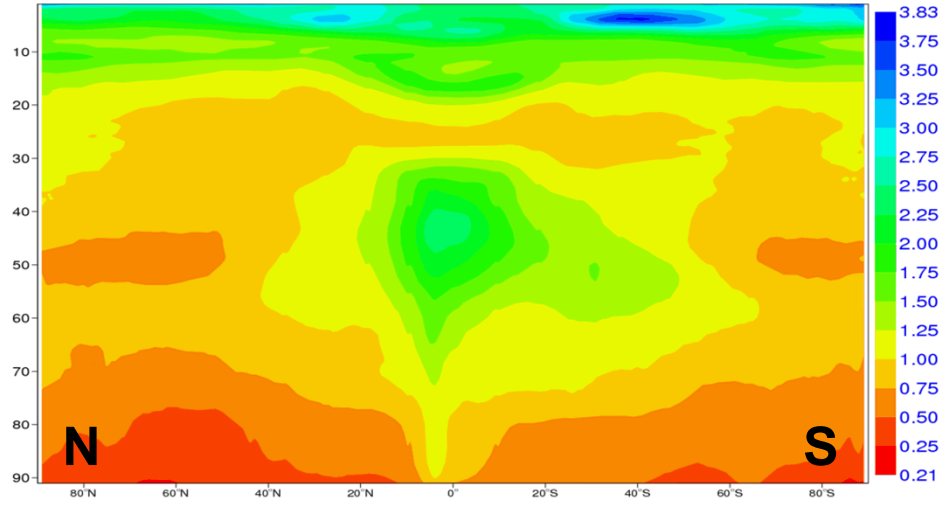
ECMWF

Conclusions and Plans

- The impact of current model error param. (especially SKEB) on the EDA ability to simulate bg error covariances needs further investigation: assimilation experiments using climatological **B** statistics derived from EDA with different model errors are required

ECMWF

Impact of model error parameterizations

Mean Vertical
EDA U spread
Cross Section

EDA U wind spread – no model error (fgk7)

Impact of SPPT on U spread (fi8s-fgk7)

Addit. Impact of SKEB on U spread (fgk6-fi8s)