Relevance of Climatological Background Error Statistics for meso-scale Data Assimilation

Jelena Bojarova & Nils Gustafsson

= HARMONIE

The linear regression model for background error statistics

The model for background error statistics in the HARMONIE data assimilation was formulated by Berre (2000)

$$X = \mathbf{B}^{1/2}\xi \qquad \mathbf{B}^{-1/2} = VD^{-1}F$$

where

***B**$^{-1/2}$* is the inverse of square-root of the background error covariance,

F is horizontal 2-dimensional Fourier transform from physical grid-point space to spectral space,

***D**$^{-1}$* is a de-correlation operator,

V is a vertical transform utilizing the eigenvectors of vertical covariance matrices.

It is assumed that the spectral components for different wave-lengths are statistically uncorrelated what significantly simplifies the formulation of background error covariance in spectral space. The price to pay is the homogeneous background error statistics in physical space. The horizontal isotropy is assumed in addition what allows to simplify formulation of background error covariance even more and to represent horizontal covariances via 1D covariance spectra for control variable.

The balance operator D is derived in spectral space through step-wise multivariate statistical regression technique for each wave number component separately

$$\begin{aligned} \zeta &= \zeta \\ \eta &= MH\zeta + \eta_u \\ (T, P_s) &= NH\zeta + P\eta_u + (T, P_s)_u \\ q &= QH\zeta + R\eta_u + S(T, P_s)_u + q_u \end{aligned}$$

The regression is performed on an ensemble of short range forecasts differences.

Two different ensemble sets for generation of background error statistics for HARMONIE AROME 2.5

DS – downscaling of coarse resolution operational ECMWF EDA (T399, 91 vertical level, 12 window 4DVAR)

EDA – high-resolution HARMONIE AROME 2.5 EDA (6h 3DVAR, conventional observations, perturbed observations for ensemble members except for control) + ECMWF EDA on lateral boundaries

Grid-point structure functions for longitudinal wind speed component in east-west direction

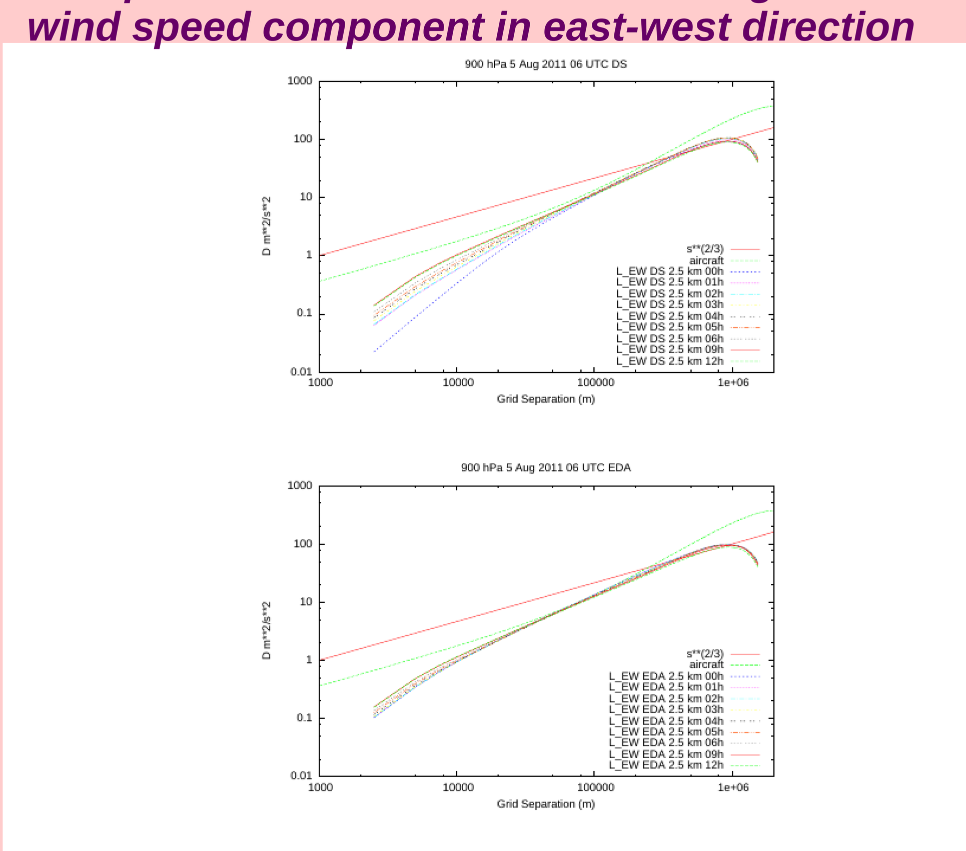


Figure 1: Structure functions for the longitudinal component of the wind in the east-weast direction. Horizontal domain average at model level 47 (≈ 900 hPa) for 5 August 2011 06 UTC. Separate curves for forecast lengths +0h, +1h, +2h, +3h, +4h, +5h, +6h, +9h and +12h. The theoretical slope of the structure function $s^{2/3}$ as well as the structure functions derived from aircraft observations are also included. Downscaling (top) and EDA (bottom).

Horizontal spectral density of unbalanced humidity error

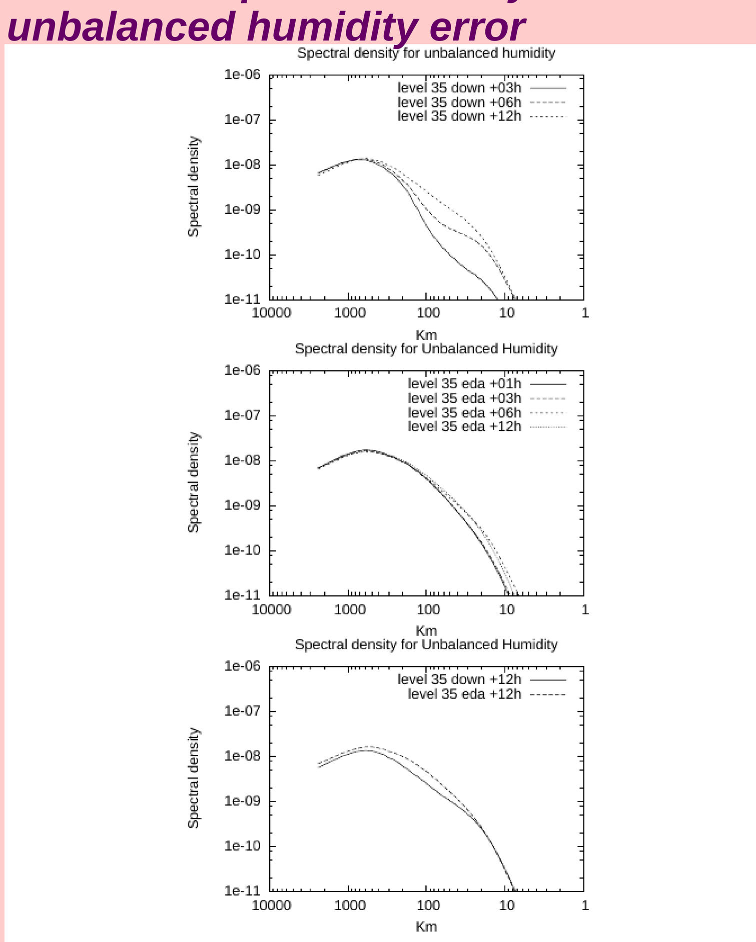


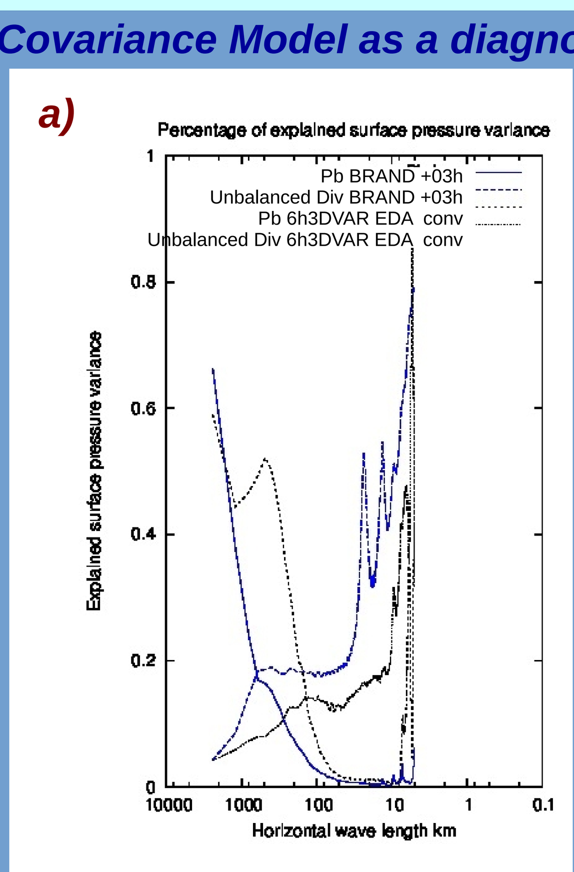
Figure 2: Spectral densities of +1h, +3h, +6h and +12h unbalanced humidity background error as estimated by the ensemble downscaling (top) and the Ensemble Data Assimilation (EDA, middle) techniques. Model level 35. Spectral densities of +12h unbalanced humidity background error as estimated by the downscaling and EDA techniques (bottom).

DS based structure functions are associated with a strong spinup during first 6-12 hours of integration. The adaptation to a high-resolution orography is likely to be one of the main processes behind the spinup. The use of low-resolution ensemble for the purposes of high-resolution data assimilation will always be associated with such interpolation processes. One can see that for humidity fields for example certain spin-up processes are on-going even after 12 h of integration with HARMONIE AROME

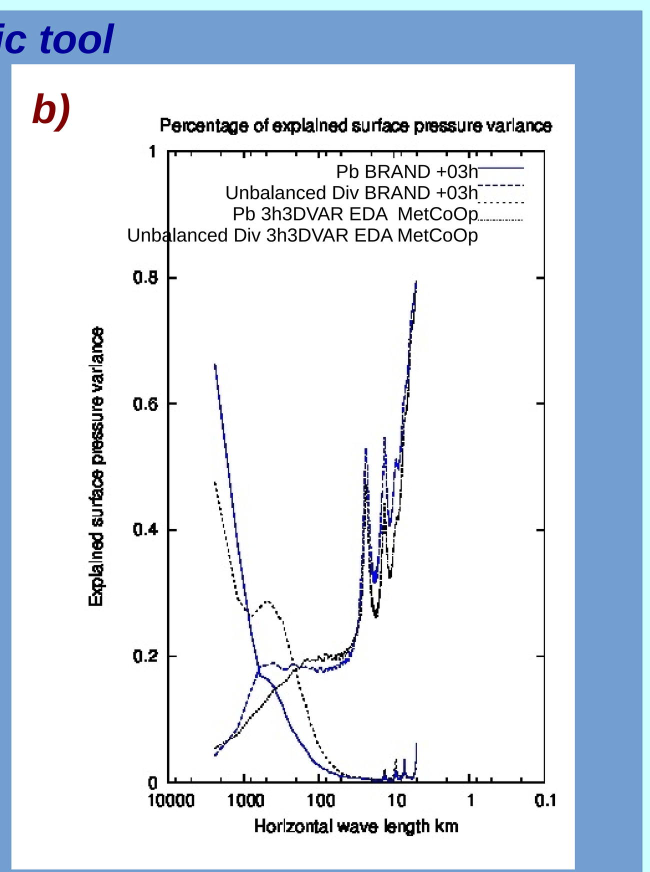
Climatological Background Error Covariance Model as a diagnostic tool

In the plots to right we show a percentage of the explained surface pressure variance computed from 3 different ensemble sets . The plot **a)** curves correspond to statistics obtained from **BRAND** ensemble and from HARMONIE AROME 2.5 EDA conv (**6h3DVAR**; conventional observing network only). The plot **b)** curves correspond to statistics obtained again from **BRAND** ensemble and from HARMONIE AROME 2.5 EDA MetCoOP (**3h3DVAR**, high resolution observing network including radar data, GNSS, ATOVS). Structure functions are computed from +03h forecast differences for three cases. First of all, we can notice that the gravity wave signature (a strong correlation between mass-field and divergence at scales < 25km) is much stronger for **3h3DVAR** than **6h3DVAR**. This indicates that an initialisation procedure has to be applied even for a high-resolution data assimilation. Although the model integration manages to reduce the level of noise, the existence of gravity waves diminishes efficiency of the data assimilation procedure with waves carrying away a part of the observation increment.

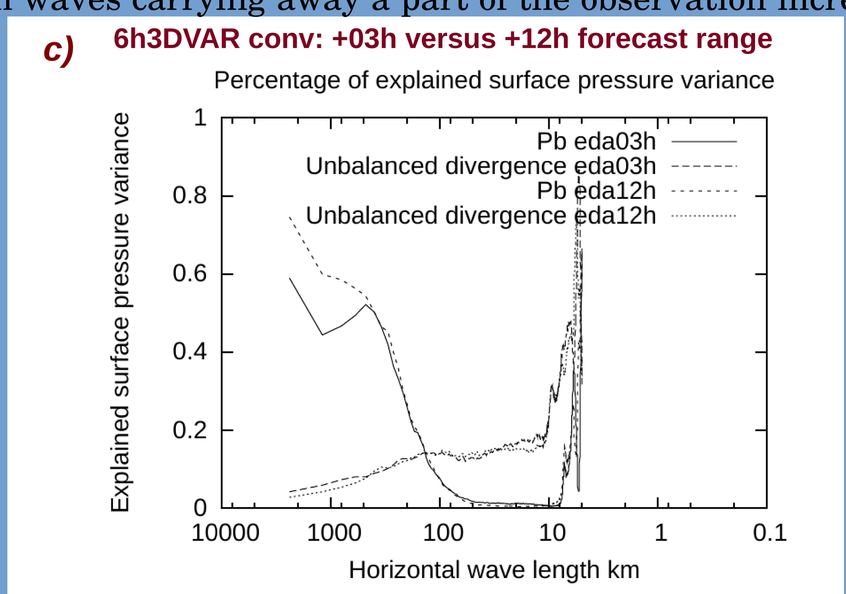
a)



b)

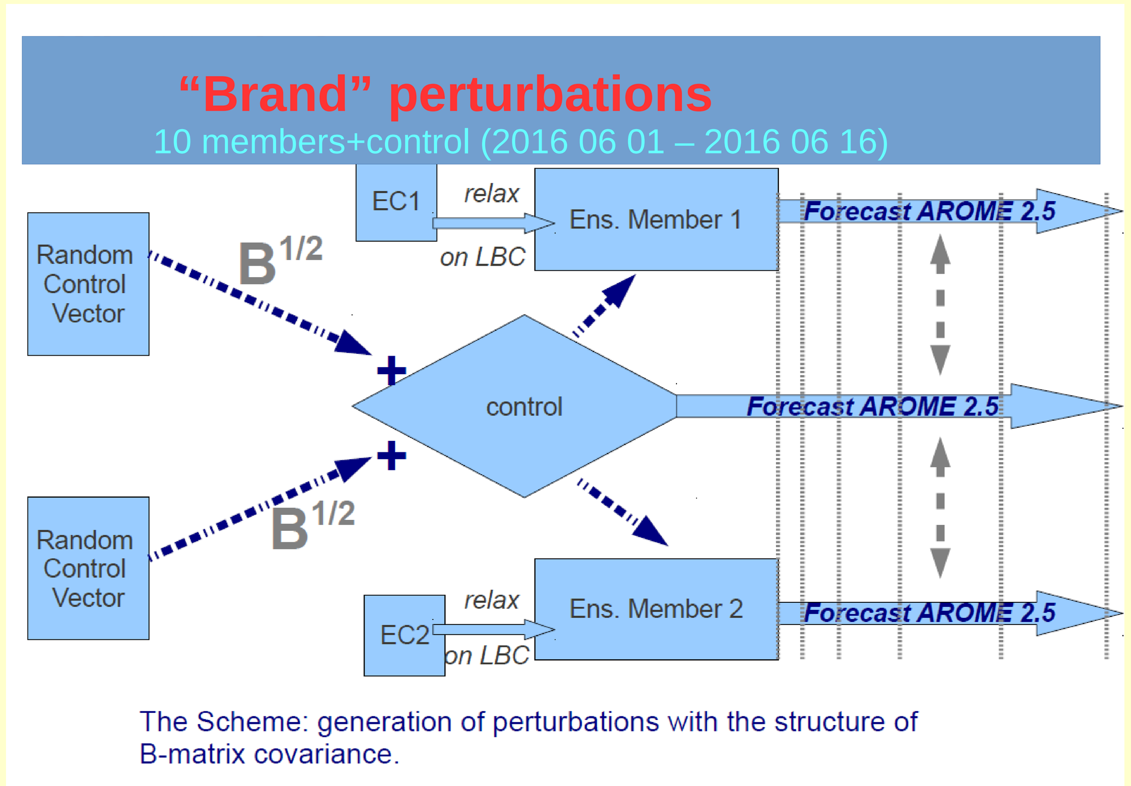


c)



Secondly we can notice a pick in the "Pb" curve at scales 200-300 km when statistics are computed from EDA ensembles, a high narrow pick in case of **6h3DVAR EDA conv** and a smoother lower pick in case of **3h3DVAR EDA MetCoOp**. Such a pick does not appear in statistics computed from the **BRAND** ensemble. We claim that such a "pick" is a result of the observation perturbation methodology for ensemble generation. Properties of a short range ensemble reflects the way in which an ensemble was generated. In **6h3DVAR EDA conv** the random unstructured perturbations were induced at the scale corresponding to the scales of conventional observing system. The background error covariance model is used to transform this unstructured noise into structured perturbations. The model integration will transfer perturbations injected at particular scales to the entire model space. However as we see from plot **c)** to the left , most of the injected energy will be transferred to larger scales.

Alternative ensemble generation approach



BRAND – ensemble perturbations are generated in control vector space and projected to the physical space applying square-root of B-matrix covariance. The perturbations are added to the short range control forecast and relaxed on the LBC's towards ECMWF EDA.

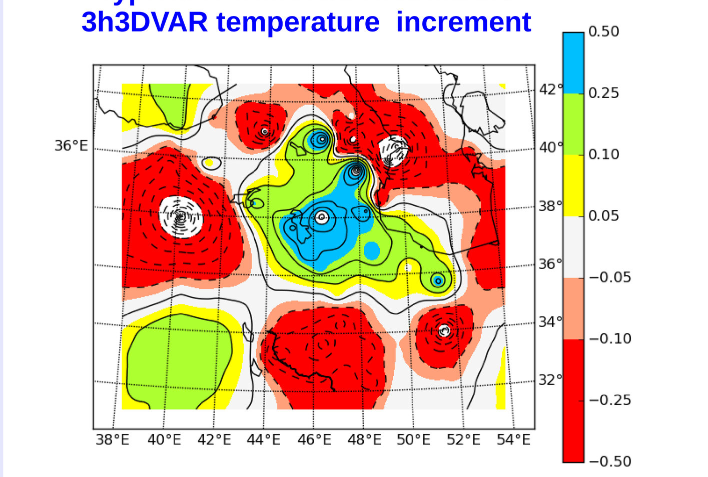
BRAND perturbations perturb entire model space in a structured way; they impose structures as described by the background error covariance model, the same model that is used to form the analysis increments from observations.

EDA perturbations inject random unstructured spread on the scales of observing system. EDA relies on DA technique and the model integration to spread the injected energy in the structured way to entire model space.

Filtered analysis increment

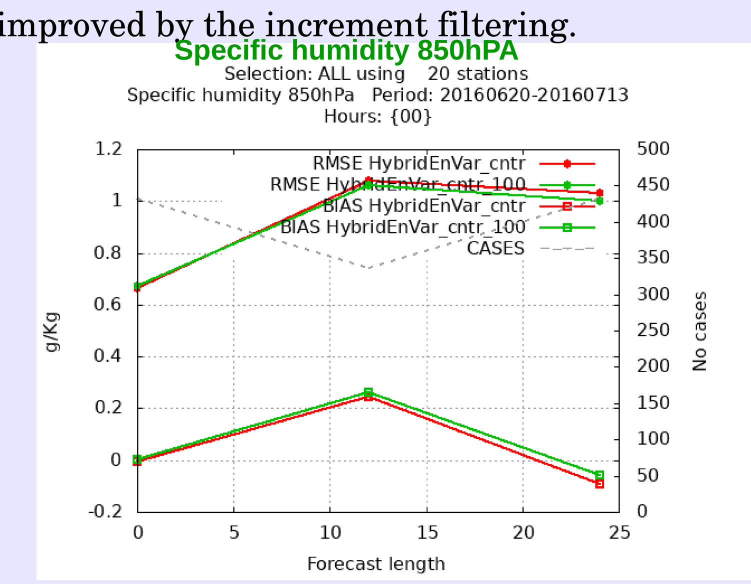
A typical HARMONIE AROME 2.5 3h3DVAR analysis temperature increment is shown in plot to the right. One can clearly see large scale structures with superimposed small scale noise. This agrees well with the diagnostics above based of the statistical forecast error structures. This small scale noise is due to gravity waves present in the ensemble from which the structure functions were derived. We claim that the gravity waves do not reflect the genuine nature of meso-scale flow but rather appear as artefacts of ensemble generation methodology lacking the initialisation step. Below we show the comparison of the verification scores between a 1 month reference run (standard 3h3DVAR; red) and the "filtered analysis increment" experiment (**3h3DVAR 100**; green) where the only 100 longest waves (>25km) were updated during assimilation. Most of the verification scores are improved by the increment filtering.

Typical HARMONIE AROME 2.5 3h3DVAR temperature increment

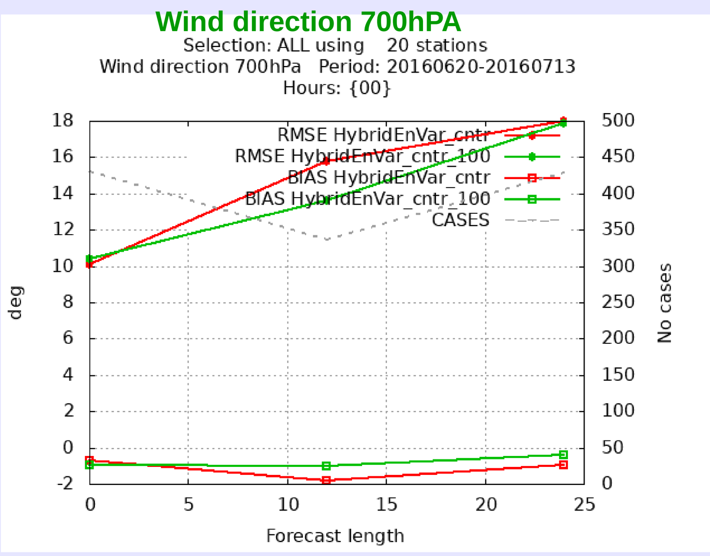


Verification scores: 2016 06 20 – 2016 07 13

Specific humidity 850hPA



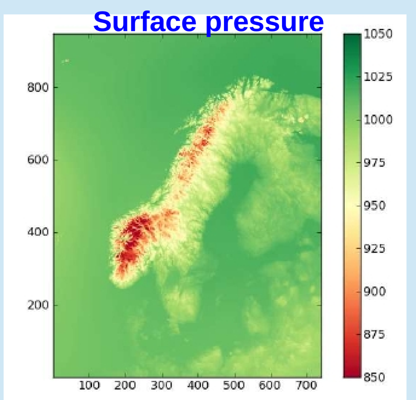
Wind direction 700hPA



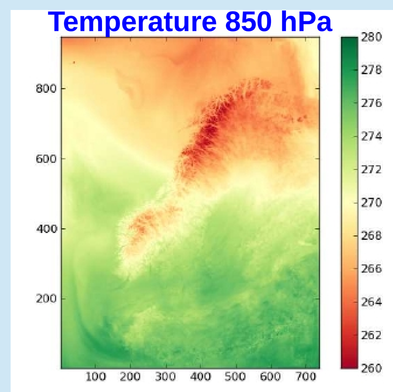
On the basic assumptions behind climatological background error covariance model

The model for the HARMONIE background error statistics is based on the three assumptions of stationarity, homogeneity and isotropy with respect to horizontal covariances. A very large size ensemble is needed to obtain a stable model for background error statistics when it is derived statistically from the ensemble of model differences. Horizontal averaging (through assumption of the horizontal homogeneity) is a very efficient remedy to obtain stable and smooth background error statistics. The assumption of horizontal isotropy adds even more smoothing. Although such a model may be attractive from a computational efficiency point of view, one may ask how representative stationary, isotropic and homogeneous horizontal correlations are for the forecasts with a strong case-to-case variability. Our experience shows that variational data assimilation at convective scale profits a lot even if a relatively simple ensemble is used to describe consistently error-of-the-day (for example via a Hybrid formulation) . This is probably due to the fact that convective motion happens in the low part of atmosphere which is strongly affected by orographic conditions and even a relatively small ensemble size is enough to capture inhomogeneity and anisotropy associated with orography.

Surface pressure

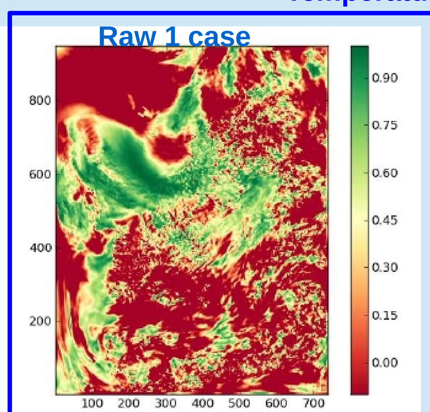


Temperature 850 hPa

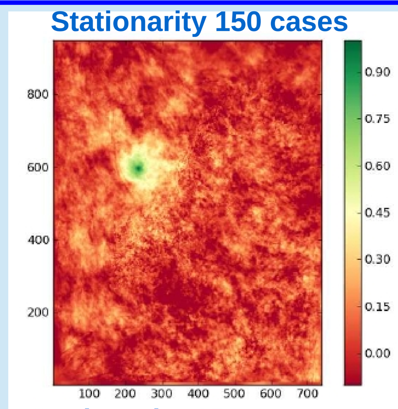


Temperature autocorrelation.

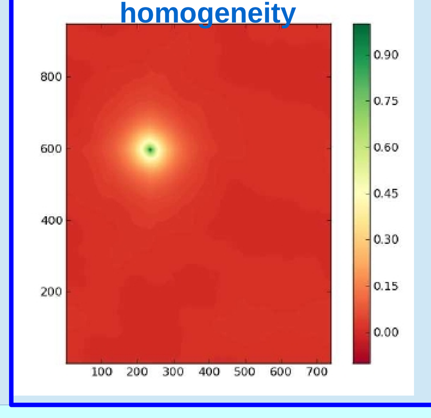
Raw 1 case



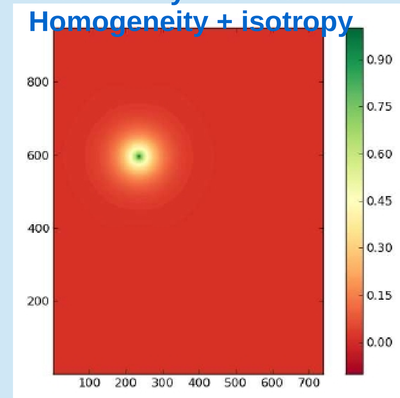
Stationarity 150 cases



Stationarity 150 cases + homogeneity



Stationarity 150 cases + Homogeneity + isotropy



In the plots to the right the balance operator **D** derived from **BRAND** and **3h3DVAR EDA MetCoOp** are presented. MetCoOp observing network is heavily dominated by mass observations. The geostrophic balance imposed by the background error covariance model is clearly seen in the structure functions derived from 3h3DVAR EDA MetCoOp data set.