Nonlinearity and non-Gaussianity in 4D-Var (and beyond)

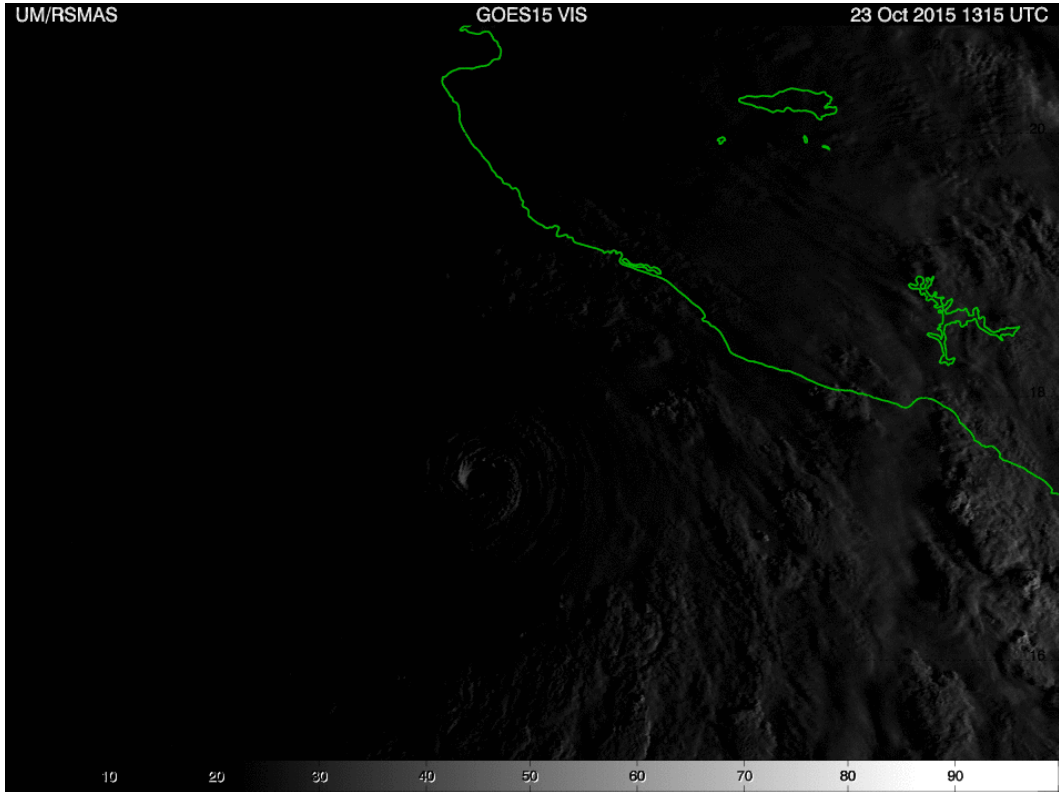
Massimo Bonavita and many colleagues

ECMWF

Special acknowledgements: Peter Lean, Elias Holm, Philippe Lopez

Hurricane Patricia (20-24 Oct. 2015)

“Second-most intense tropical cyclone on record worldwide, with a minimum estimated atmospheric pressure of 872 hPa” (NOAA NHC, Trop. Cyclone Rep. EP202015)

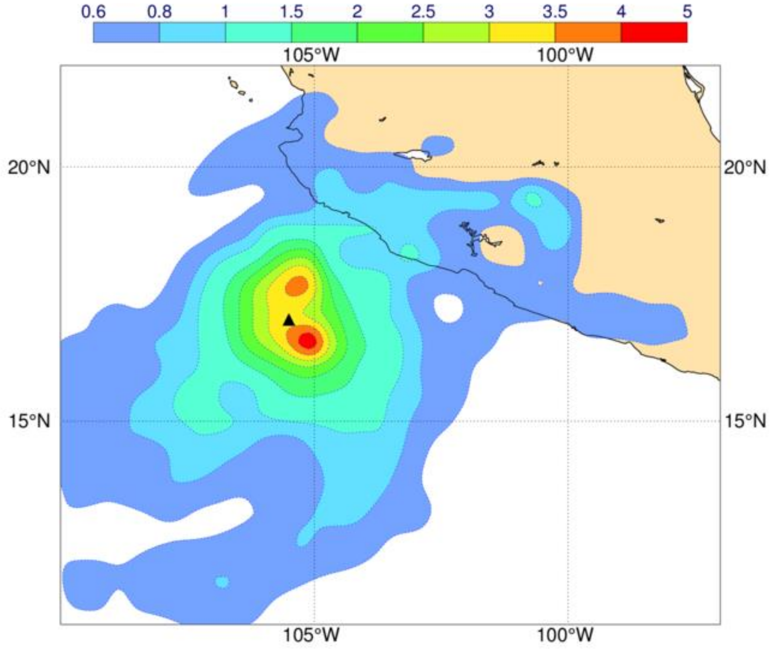


NOAA GOES-15 VIS
Credits: University of Miami's Rosenstiel School of Marine and Atmos. Science

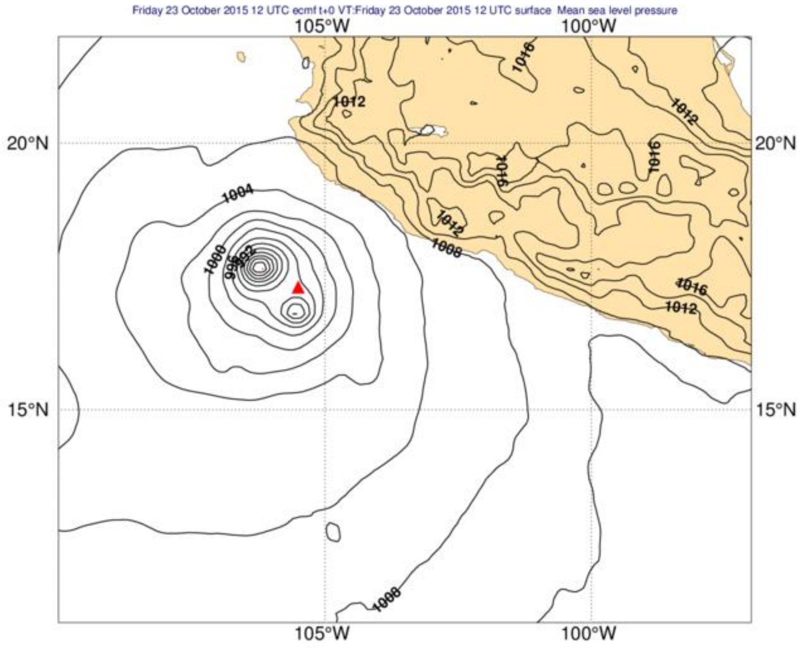
Hurricane Patricia (20-24 Oct. 2015)

"Second-most intense tropical cyclone on record worldwide, with a minimum estimated atmospheric pressure of 872 hPa" (NOAA NHC, Trop. Cyclone Rep. EP202015)

0069 IFS **EDA B** errors 23-10-2015 12UTC



0069 IFS Analysis 23-10-2015 12UTC



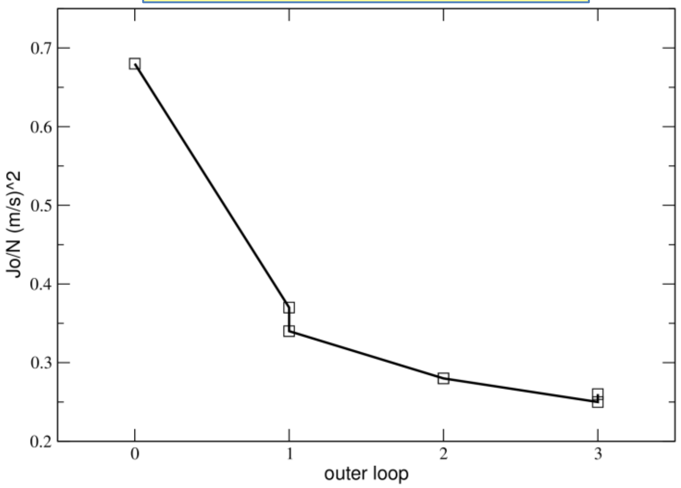
ECMWF

Hurricane Irma (Sept. 2017)

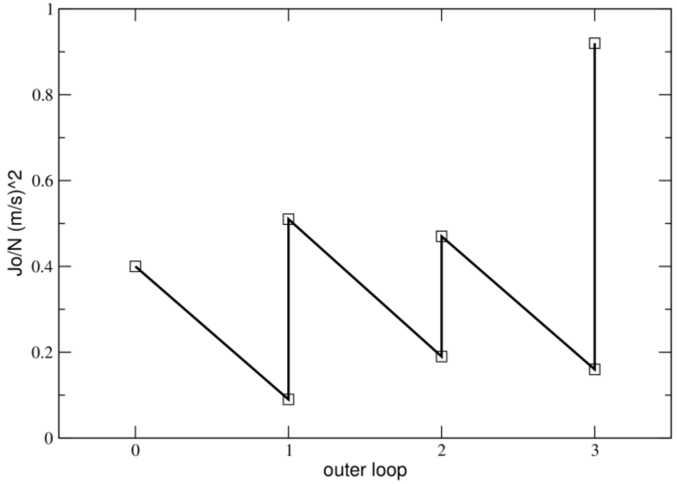
VIIRS image from NOAA Suomi NPP satellite, 5/09/2017 17.06UTC

Evolution of the **Jo** cost function during the IFS 4D-Var minimization, 04-09-2017 12UTC

Radiosonde winds



Near-core Dropsonde winds



Outline

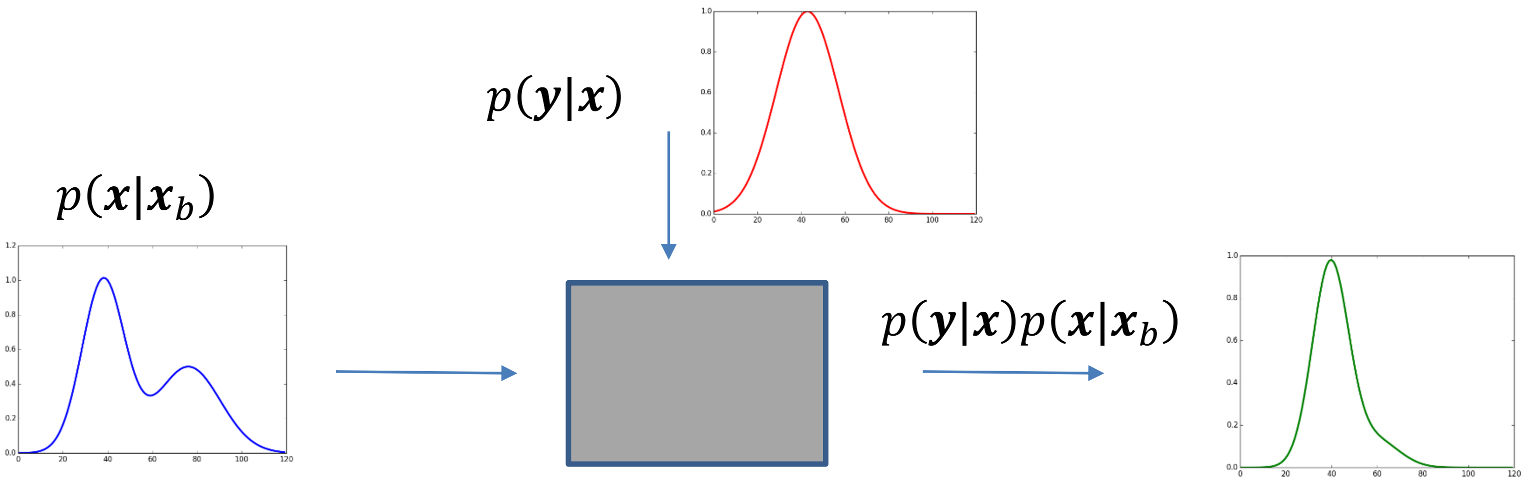
- Nonlinear and non-Gaussian effects: background
- Dealing with nonlinearity
- Dealing with non-Gaussianity
- The way forward

Nonlinearity and non-Gaussianity

- Nonlinear and non-Gaussian effects are **inextricably linked topics**: model and observation operator nonlinearities inevitably produce non-Gaussian posteriors etc.
- Subject of very many studies in simplified models (Miller et al, 1994; Pires et al, 1996; Evensen, 1997; Verlaan and Heemink 2001, Bocquet et al, 2010,…)
- Sudden spike of interest at ECMWF in the mid 2000s, e.g. Andersson et al., 2005; Trémolet, 2005; Radnóti et al., 2005
- Worth revisiting in light of:
  1. Much higher resolution of current models and data assimilation;
  2. Vastly increased use of non-linear observations (e.g., all-sky radiances)

The Bayes perspective

- We can think of the analysis process as updating our prior knowledge about the state, represented by the background forecast pdf, with new observations, represented by their pdf:



$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{x}_b)$$

- $p(\boldsymbol{x}|\boldsymbol{x}_b)$ = prior pdf (encapsulate our knowledge about the state before new observations)
- $p(\boldsymbol{y}|\boldsymbol{x})$ = observations likelihood (pdf of the observations conditioned on the state)
- $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})$= posterior pdf (updated pdf of the state after the analysis)

The Gaussian approximation

- Not making assumptions on the shape of the prior and the likelihood pdf makes the Bayesian problem difficult (i.e., analytically and/or computationally intractable)
- Usual choice is to assume a Gaussian distribution for the both the observations' likelihood and the prior pdf
- Why Gaussian?
    1. Mathematically tractable problem;
    2. Full distribution characteristics defined by only its first two moments (mean and covariance);
    3. Supported by the Central Limit Theorem;
    4. Maximum entropy distribution for given variance (=> taking a Gaussian pdf we are making the least amount of assumptions on the underlying population)
    5. Assuming linear model and observation operators the **posterior** (analysis) distribution $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})$ can also be expressed as a **Gaussian** distribution (conjugate distributions)

Mean-finding methods

- Once we know the form of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})$ we have a choice:

  1) Either we can solve for the **mean** and the **covariance** of the posterior Gaussian distribution:

$$\boldsymbol{x}_a = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})\boldsymbol{dx}$$

$$\mathbf{P}_a = \int (\boldsymbol{x}-\boldsymbol{x}_a)(\boldsymbol{x}-\boldsymbol{x}_a)^T\, p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})\boldsymbol{dx}$$

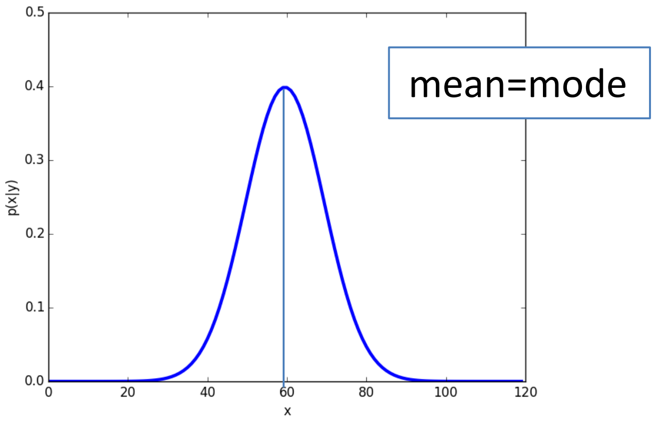
  Methods based on this approach include Optimum Interpolation (O.I.), Kalman Filter, Ensemble Kalman Filter and Smoother (EnKF/S), 3/4D Ensemble Var (EnsVar). This is the minimum variance solution or the best linear unbiased estimate (BLUE).

Mode-finding methods

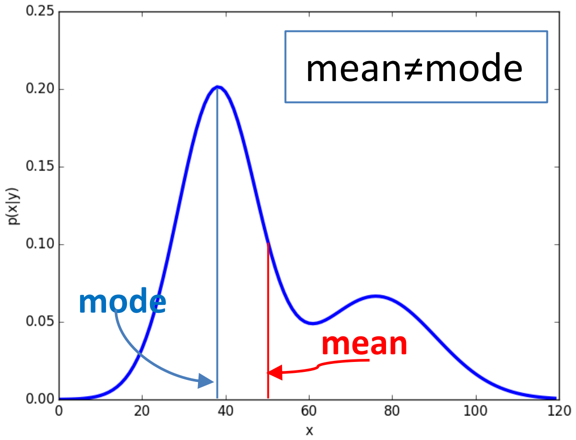
2) Alternatively we might want to estimate the **mode** of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})$, i.e. find the analysis $\boldsymbol{x}_a$ as the state that corresponds to the maximum of the posterior distribution:

$$\boldsymbol{x}_a = \arg\max_{\boldsymbol{x}}\big(p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})\big)$$

This way of attacking the problem leads to the variational approach (3D-Var, 4D-Var). The solution found in this way is called the maximum a-posteriori probability state (MAP).



In the linear, Gaussian world solutions 1) and 2) coincide



In the real world they do not!

Nonlinearity and non-Gaussianity in 4D-Var

- The 4D-Var cost function is simply the neg log of the posterior pdf:

$$J(\boldsymbol{x}) \propto -\log\left(p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{x_b})\right) \propto -\log\left(p(\boldsymbol{y}|\boldsymbol{x})\right) - \log\left(p(\boldsymbol{x}|\boldsymbol{x}_b)\right)$$

- Nonlinear effects arise when the relationship between observations and model state is nonlinear (this will also make $p(\boldsymbol{y}|\boldsymbol{x})$ non-Gaussian)
- The prior error distributions of $\boldsymbol{y}|\boldsymbol{x}$ and $\boldsymbol{x}|\boldsymbol{x}_b$ can be non-Gaussian to start with
- Both effects have the potential to introduce multiple minima in the cost function and make the minimisation problematic

Nonlinear effects

"The road to wisdom? — Well, it's plain and simple to express:
Err
and err
and err again
but less
and less
and less."

Piet Hein
(Danish mathematician, 1905-1996)

Nonlinear effects

- Nonlinear 4D-Var:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^{\mathrm{T}}\mathbf{P}_b^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{y}_k - G_k(\mathbf{x}_0)\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{y}_k - G_k(\mathbf{x}_0)\right) \qquad (1)$$

where $G_k = H_k \circ M_{t_0 \to t_k}$ is a generalised observation operator that includes **model** propagation

- Solving (1) directly is not feasible:

  1. Direct computation of minimum of $J(\mathbf{x}_0)$ is impossible for any realistic model;

  2. Nonlinear G_k can lead to nonconvex cost functions

Nonlinear effects

- Incremental 4D-Var (Courtier et al, 1994) approximates nonlinear cost function as a sequence of minimizations of quadratic cost functions defined in terms of perturbations $\delta\mathbf{x}_0$ around a sequence of "progressively more accurate" trajectories $\mathbf{x}^g$:

$$J(\delta\mathbf{x}_0) = \frac{1}{2}\left(\delta\mathbf{x}_0 + \mathbf{x}_0^g - \mathbf{x}_b\right)^{\mathrm{T}}\mathbf{P}_b^{-1}\left(\delta\mathbf{x}_0 + \mathbf{x}_0^g - \mathbf{x}_b\right) + \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{d}_k - \mathbf{G}_k(\delta\mathbf{x}_0)\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{d}_k - \mathbf{G}_k(\delta\mathbf{x}_0)\right)$$

  where $\mathbf{G}_k$=$\mathbf{H}_k\mathbf{M}_{t_0\to t_k} = \mathbf{H}_k\mathbf{M}_{t_{k-1}\to t_k}\mathbf{M}_{t_{k-2}\to t_{k-1}} \ldots \mathbf{M}_{t_0\to t_1}$ is the linearised version of G_k around the latest model trajectory and $\mathbf{d}_k = \mathbf{y}_k - G_k\left(\mathbf{x}_0^g\right)$ are the corresponding observation departures

- The incremental formulation has many advantages:

  1. Reduced computational cost through the use of lower resolution/complexity linearised models
  2. Quadratic cost function guarantees convergence and uniqueness
  3. Quadratic cost function allows use of efficient gradient-based minimisers

Nonlinear effects

- Going from nonlinear to incremental formulation requires the tangent linear (TL) approx.:

$$\mathbf{y}_k - G_k(\mathbf{x}_0) = \mathbf{y}_k - G_k\left(\mathbf{x}_0^g + \delta\mathbf{x}_0\right) = \mathbf{y}_k - G_k\left(\mathbf{x}_0^g\right) - \mathbf{G}_k(\delta\mathbf{x}_0) - \frac{1}{2}(\delta\mathbf{x}_0)^{\mathrm{T}}\left(\frac{\partial\mathbf{G}_k}{\partial\mathbf{x}}\right)_{\mathbf{x}^g}(\delta\mathbf{x}_0) - O(\|\delta\mathbf{x}_0\|^3) \approx$$

$$\mathbf{y}_k - G_k\left(\mathbf{x}_0^g\right) - \mathbf{G}_k(\delta\mathbf{x}_0)$$

- The TL approximation implies either or both:

  1. Small increments (when scaled w.r.to observation errors);

  2. Small sensitivity of $\mathbf{G}_k$=$\mathbf{H}_k\mathbf{M}_{t_0\rightarrow t_k}$to linearization trajectory $\mathbf{x}^g$=> approx. linear behaviour of $\mathbf{H}$ and $\mathbf{M}$

- So, how nonlinear are M and H?

Model nonlinearities

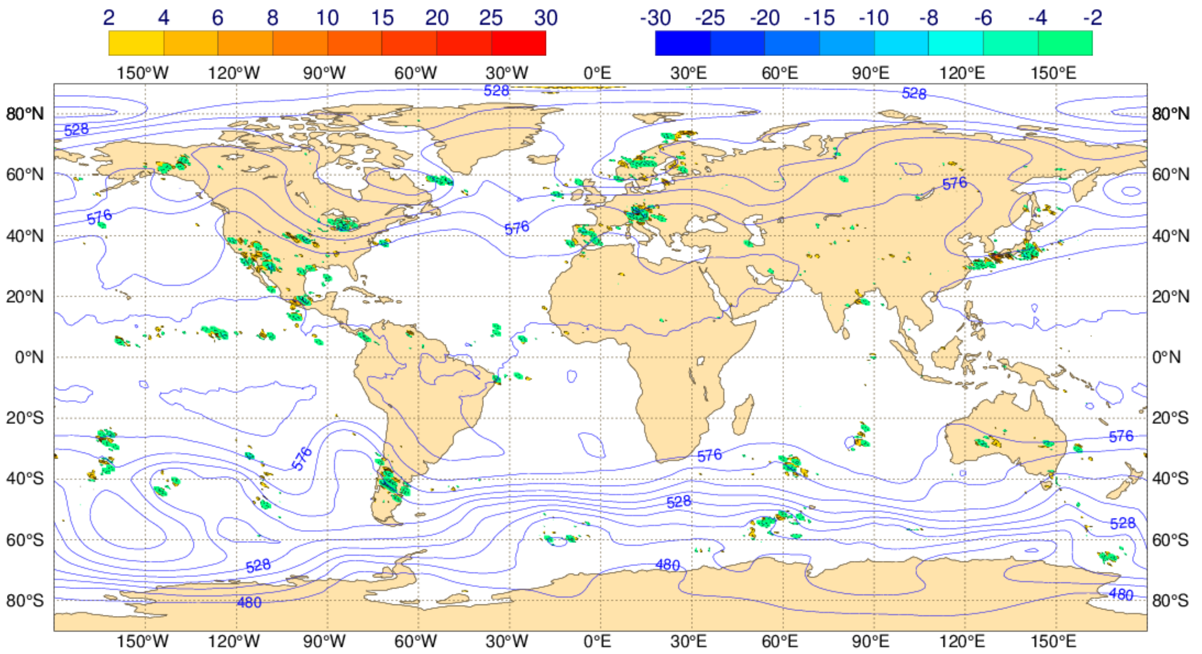
- Model non-linearity: how much initial increments evolved by the linearised model and the nonlinear model differ?

$$stdev\left(M(\mathbf{x}^t + \delta\mathbf{x}_0) - \left(M(\mathbf{x}^t) + \mathbf{M}(\delta\mathbf{x}_0)\right)\right)$$

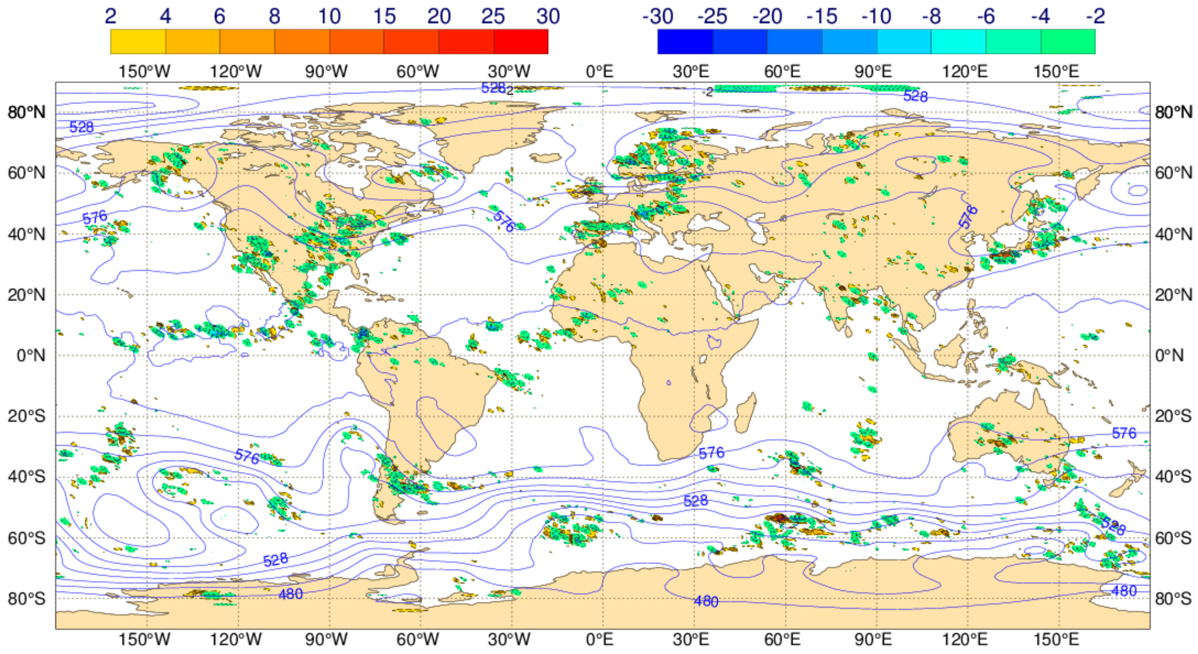
Vorticity 500 hPa (units: $10^{-5}s^{-1}$)

4D-Var, TCo639 outer loops, TL191/191 inner loops

T+3h



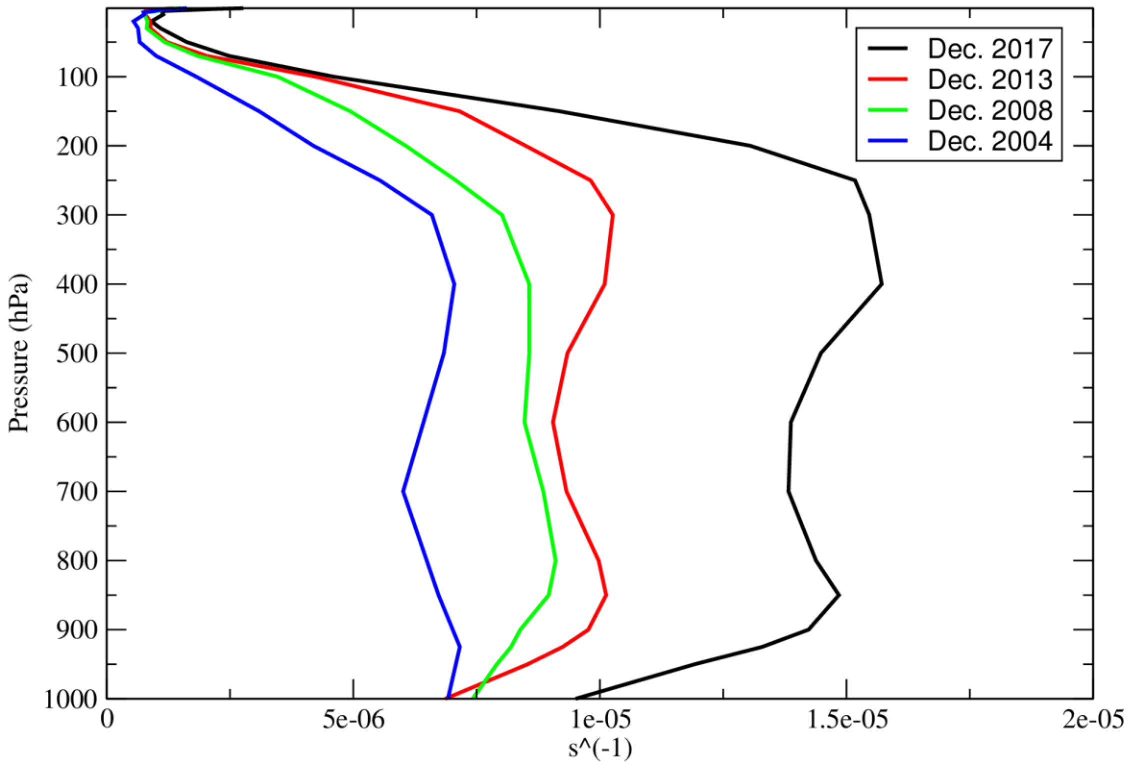
T+9h



Model nonlinearities

$$StDev\left(M(\mathbf{x}^t + \delta\mathbf{x}_0)\ - \left(M(\mathbf{x}^t) + \mathbf{M}(\delta\mathbf{x}_0)\right)\right)$$

Vorticity



- Larger nonlinearities than in the past, due to:

1. Increased resolution (40 km -> 9 km)
2. Increase mismatch of outer-inner loop resol. (from 3 -> 5)
3. Less diffusive model

Observation nonlinearities

- The other source of nonlinearity is from observation operators

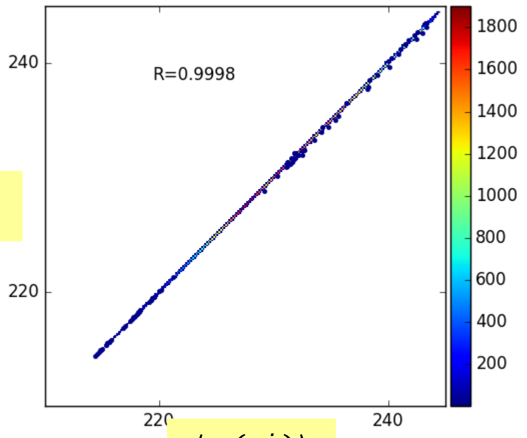
- For linear observation operators in an ensemble DA:

$$H\left(\mathbf{x}_b^{ctrl}\right) = H\left(M\left(\left\langle \mathbf{x}_a^i \right\rangle\right)\right) \approx \left\langle H\left(\mathbf{x}_b^i\right)\right\rangle \qquad i = 1, \dots, N_{ens}$$

Observation nonlinearities

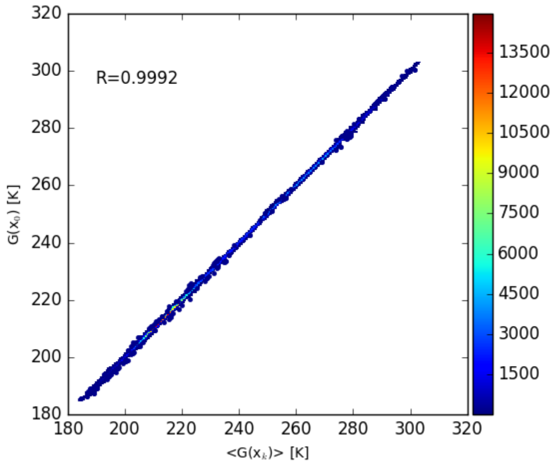
AMSU-A ch. 6
Integrated tropospheric temperature

$H\left(\mathbf{x}_b^{ctrl}\right)$

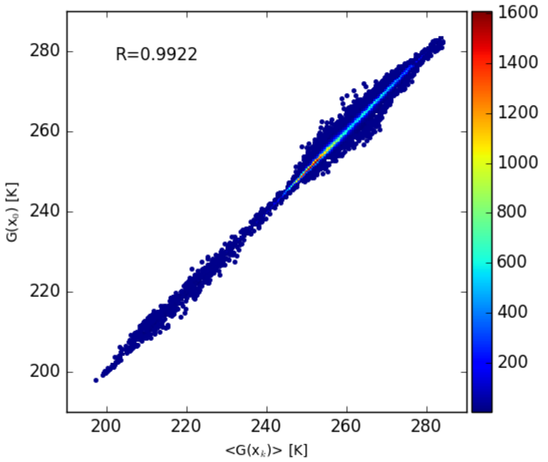


$\langle H\left(\mathbf{x}_b^{i}\right)\rangle$

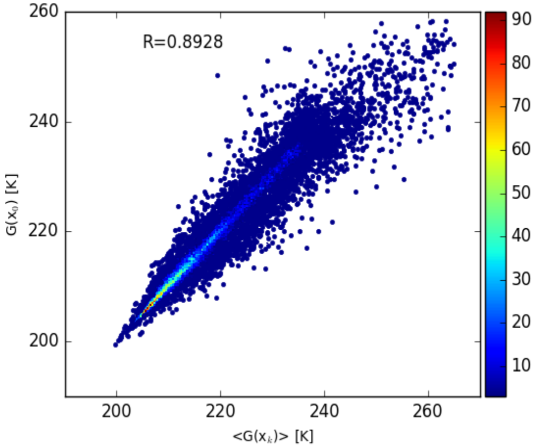
Radiosonde temperature
Point temperature measurement



ATMS ch. 20
Tropospheric humidity



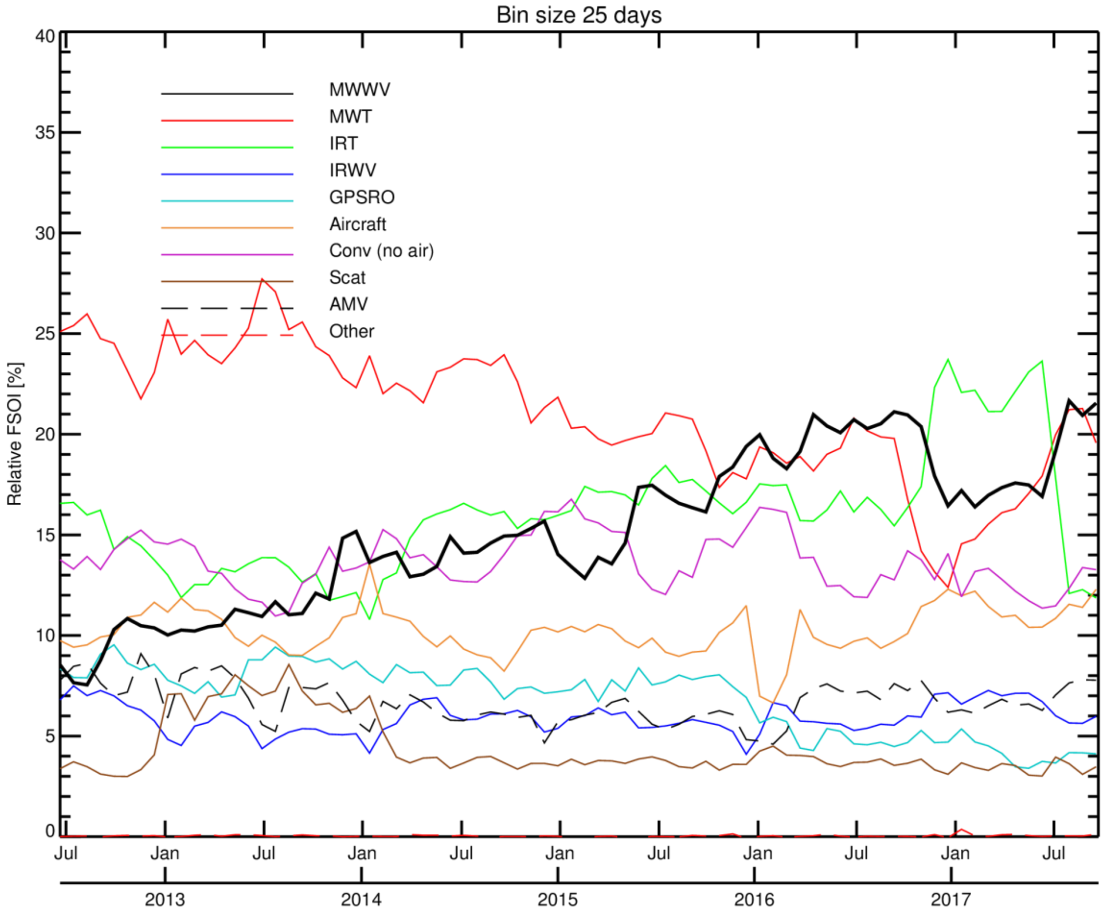
AMSR-2 ch. 6
Cloud liquid water



Observation nonlinearities

Nonlinear $\mathbf{H}$ effects in 4D-Var become increasingly important because of **ever increasing influence of nonlinear observations on the analysis**

Forecast sensitivity (FSO) of major observing systems in ECMWF DA



from Alan Geer

Nonlinear effects

- Nonlinear effects, from both the model and the observations are important;
- The current trend shows they will become even more important in the future
- How do we deal with them?

Nonlinear effects

- The validity of the TL approximation implies either or both:

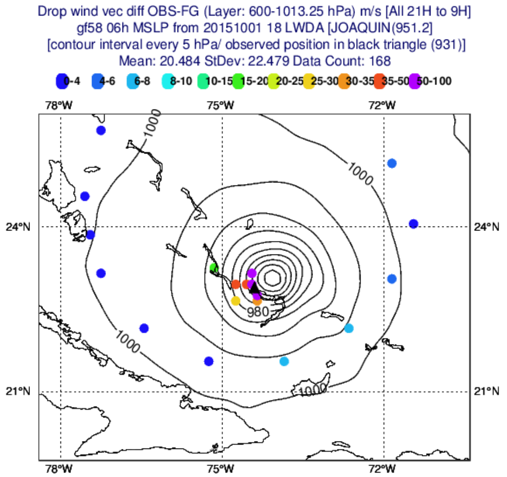
  1. Small increments around the linearisation point (when scaled w.r.to observation errors);

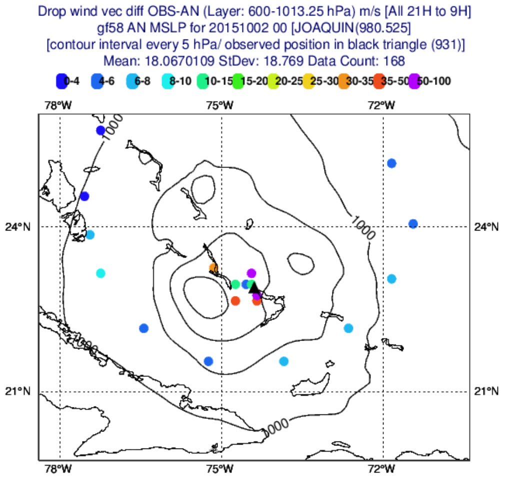
  $$\left\|\mathbf{R}^{-1/2}\delta\mathbf{x}_0\right\| \ll 1$$

  2. Small sensitivity of $\mathbf{G}_k$=$\mathbf{H}_k\mathbf{M}_{t_0\rightarrow t_k}$ to linearization trajectory => ~~approx. linear behaviour of $\mathbf{H}$ and $\mathbf{M}$~~

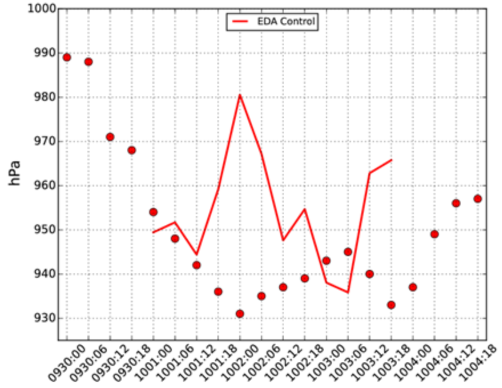
Nonlinear effects: large increments

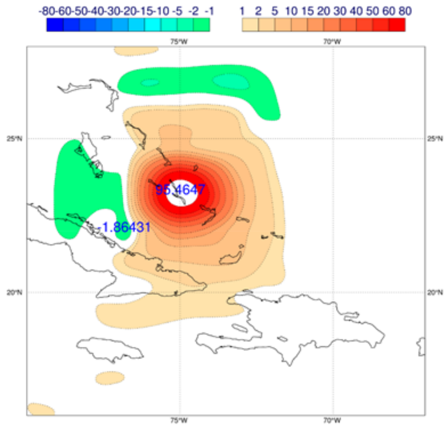
- When does the incremental approach break down?





Tropical Cyclone Joaquin, 2015-10-02 00UTC
Near TC core O-B wind departures: 30 - 80 m/s
Assumed Observation error StDev: 1.8 - 2.2 m/s



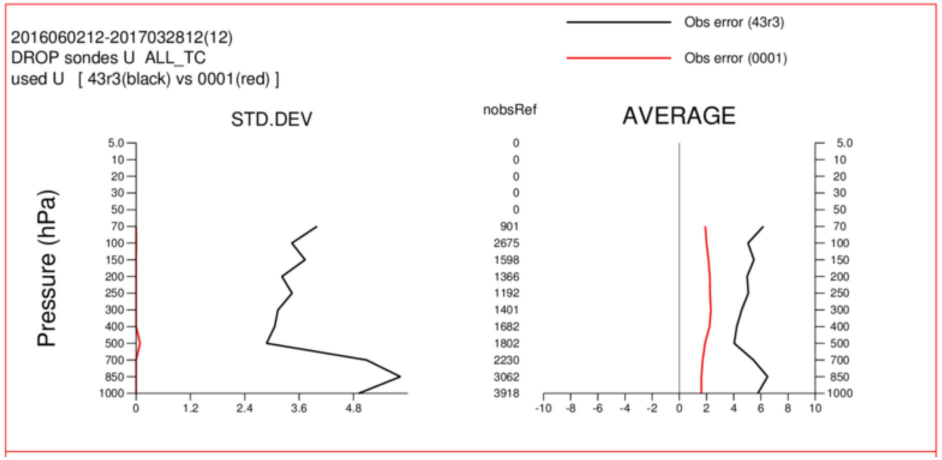


Bonavita et al., 2017: On the initialization of Tropical Cyclones. ECMWF Tech. Memo. n. 810

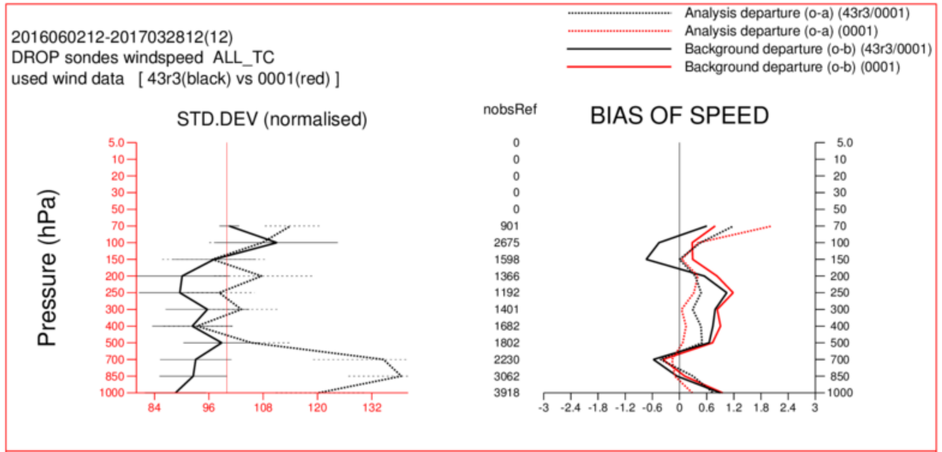
Nonlinear effects: large increments

- Remedy: reduce increments by increase of prescribed Observation Errors (taking representativeness error into account)

$$\left\langle \left(y - G\left(\mathbf{x}_0^b\right)\right)^2 \right\rangle = \sigma_b^2 + \sigma_o^2 = \sigma_b^2 + \sigma_{o,I}^2 + \sigma_{o,R}^2 + \sigma_{o,H}^2$$



Observation error model for dropsondes winds

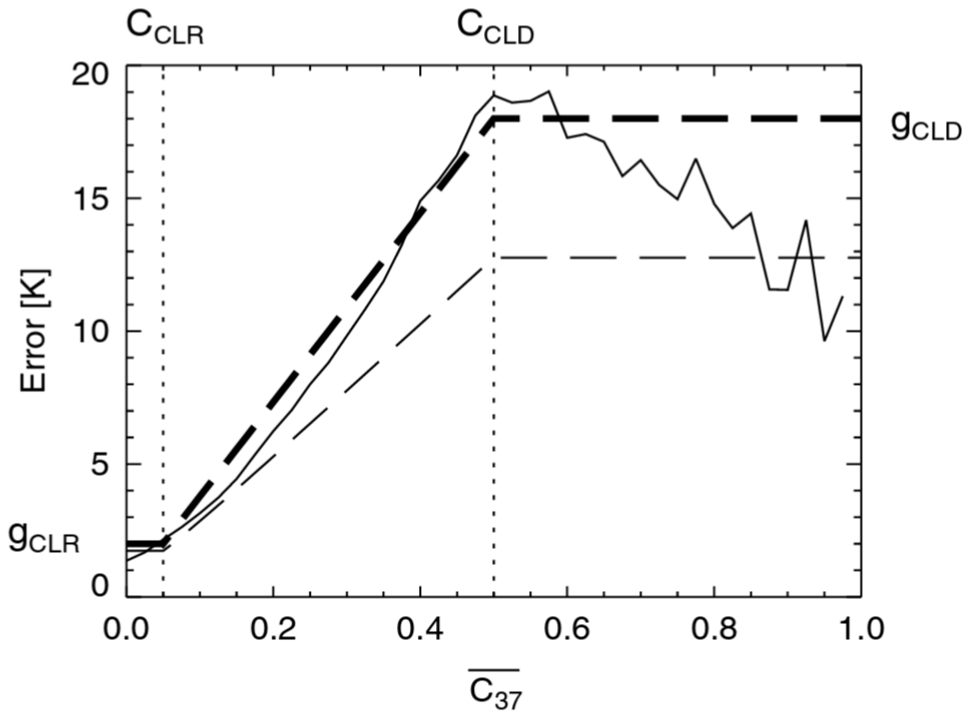


O-B departures' statistics for dropsondes wind speed

Bonavita et al., 2017. ECMWF Tech. Memo. n. 810

Nonlinear effects: large increments

- Remedy: reduce increments by increase of prescribed Observation Errors (taking representativeness error into account)



Observation error model for AMSRE ch.19v all-sky radiances as a function of “symmetric” (forecast and observed) cloud amount

from Geer and Bauer, 2011

Nonlinear effects: MDA

- The idea of modulating observation errors when observation departures are large can be generalised

- The observations likelihood can be formally written as (Neal, 1996):

$$p(y|x) = p(y|x)^{\Sigma_{i=1}^{N} 1/\alpha_i} = \Pi_{i=1}^{N} p(y|x)^{1/\alpha_i}, \qquad with\ \alpha_i > 0, \qquad \Sigma_{i=1}^{N} 1/\alpha_i = 1$$

- The Bayes posterior then becomes:

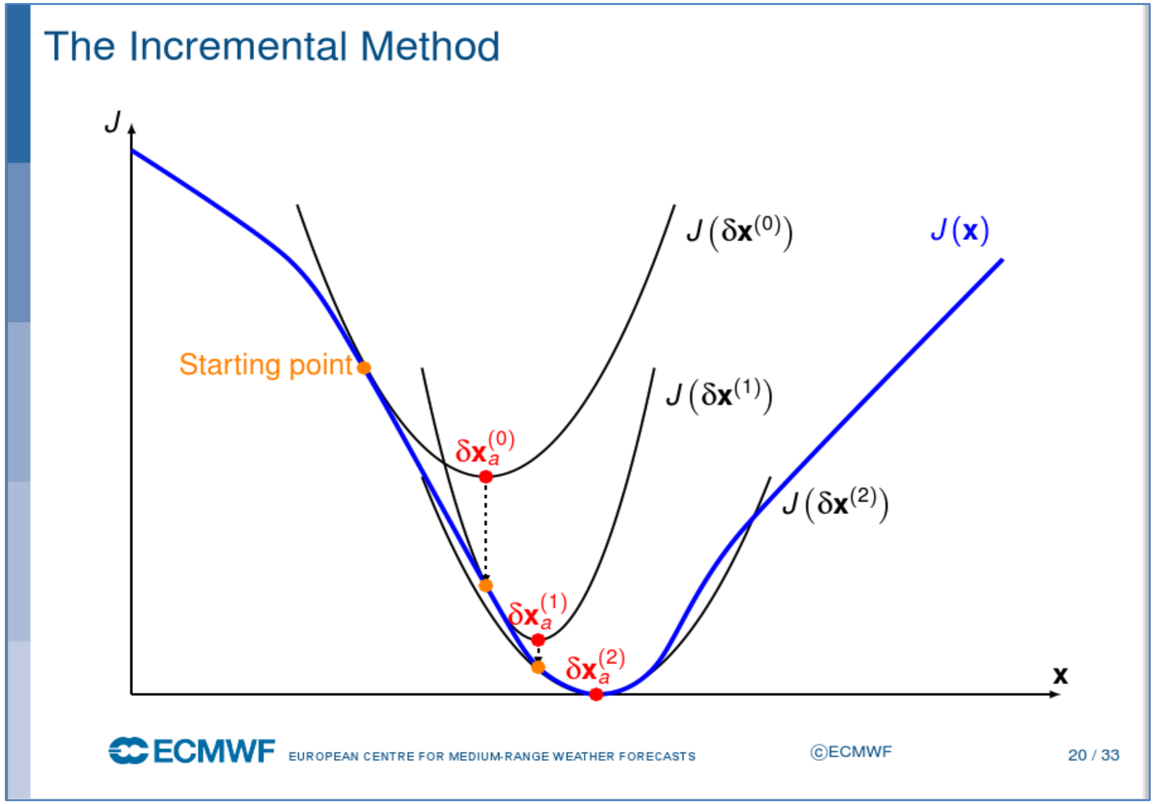
$$p(x|y, x_b) \propto p(x_b|x)p(y|x) = p(x_b|x)\Pi_{i=1}^{N} p(y|x)^{1/\alpha_i}$$

- This expression can be written as a recursion starting from the background and progressively updating the guess state. This is called ES-MDA (Ensemble Smoother with Multiple Data Assimilations, Emerick and Reynolds, 2012, 2013; Evensen, 2018)

- Maximising this recursion is equivalent to minimising a series of successive cost functions with the obs error covariances modulated by α_i

Nonlinear effects: the incremental approach

- Incremental 4D-Var deals with nonlinearities by a succession of quadratic optimization problems around progressively more accurate first guess trajectories (approximate Gauss-Newton method, Gratton et al 2007):

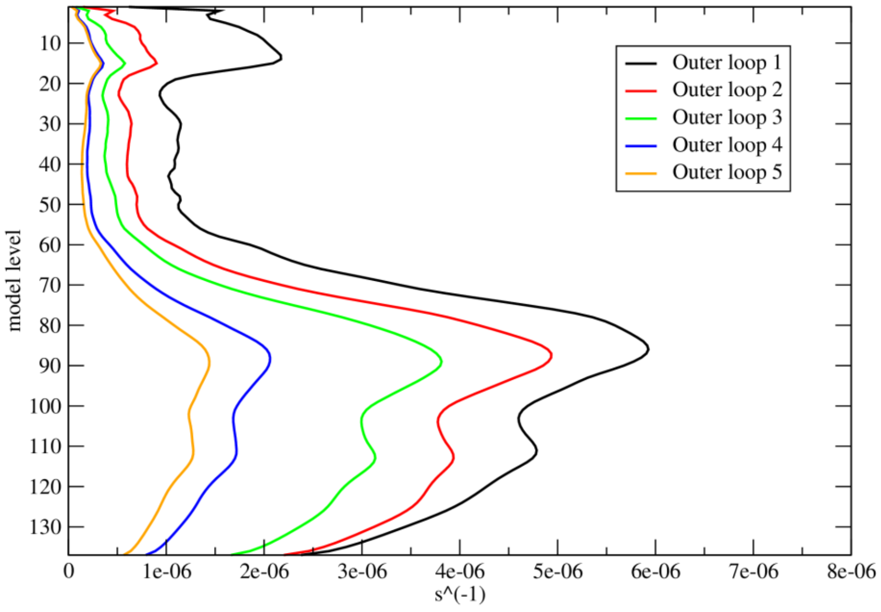


S. Massart

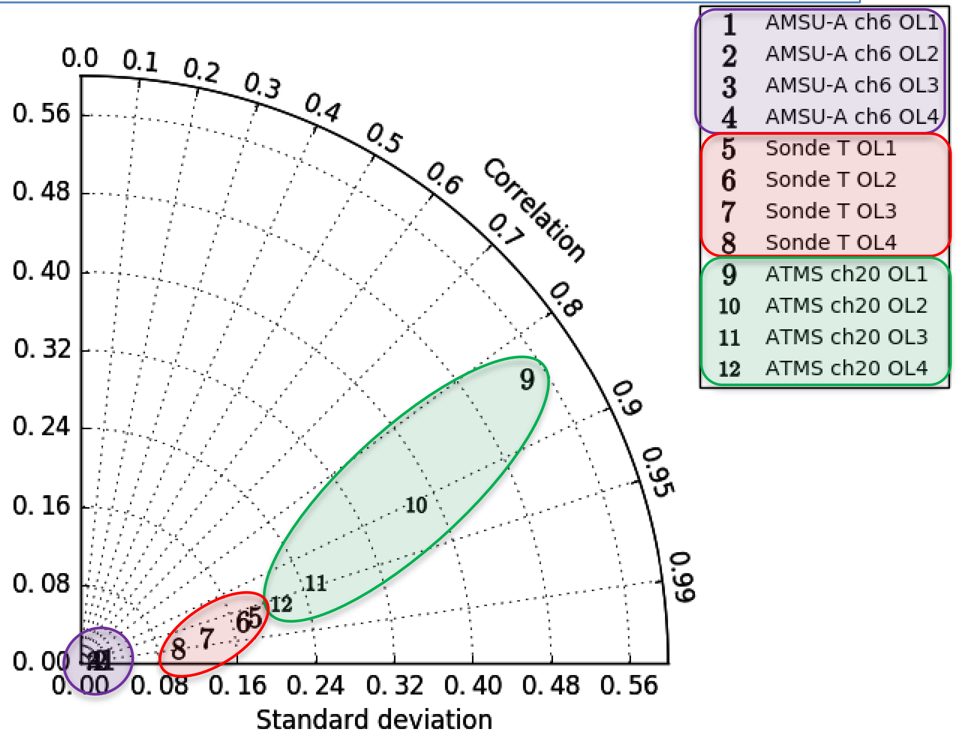
Nonlinear effects: the incremental approach

- Incremental 4D-Var deals with nonlinearities by a succession of quadratic optimizations around progressively more accurate first guess trajectories => progressively smaller increments => more accurate local linearisation!

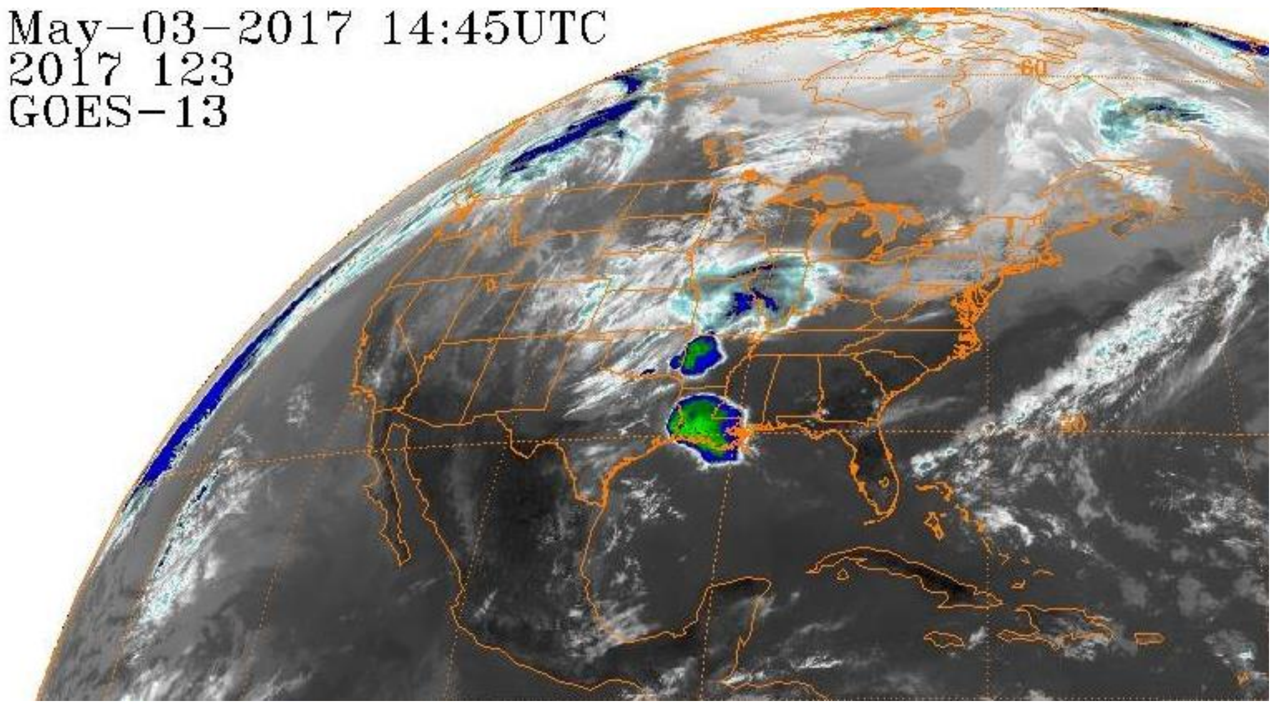
StDev of **vorticity analysis incr.** in successive minimizations



Taylor diagram for differences in obs-dep for nonlinear and linearised trajectories

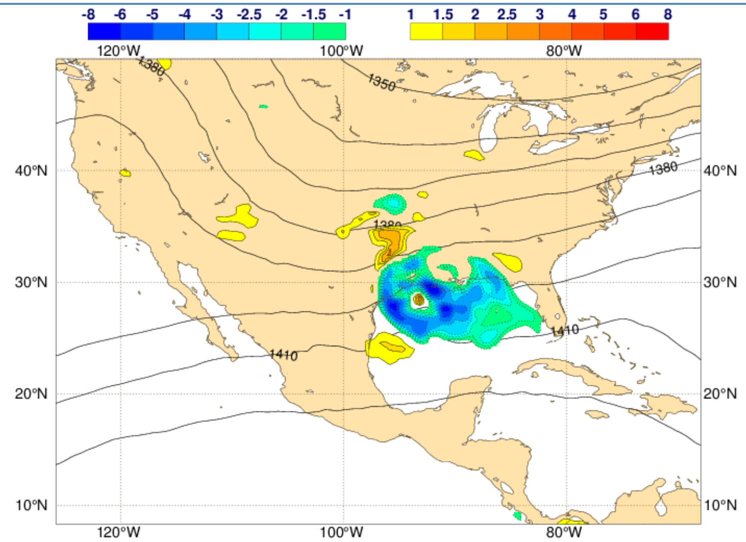


Nonlinear effects: the incremental approach

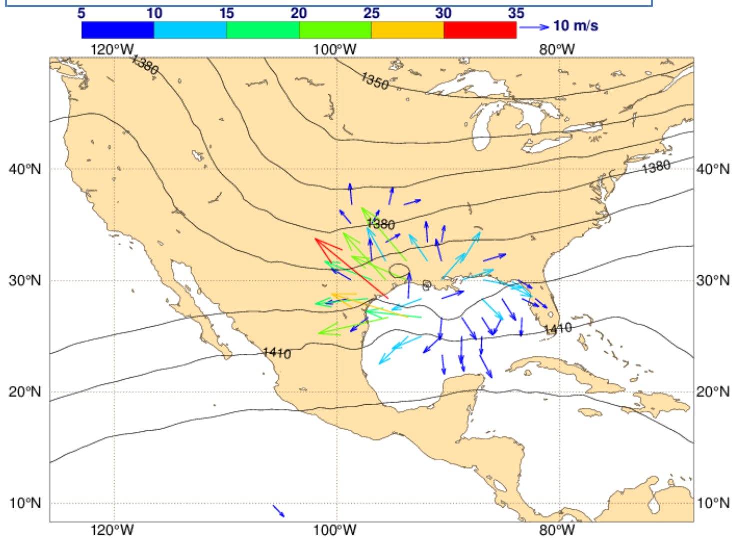


Source: www.ncdc.noaa.gov/gibbs/

150 hPa temp. analysis increments, 2017/05/03 12UTC
ECMWF operational analysis

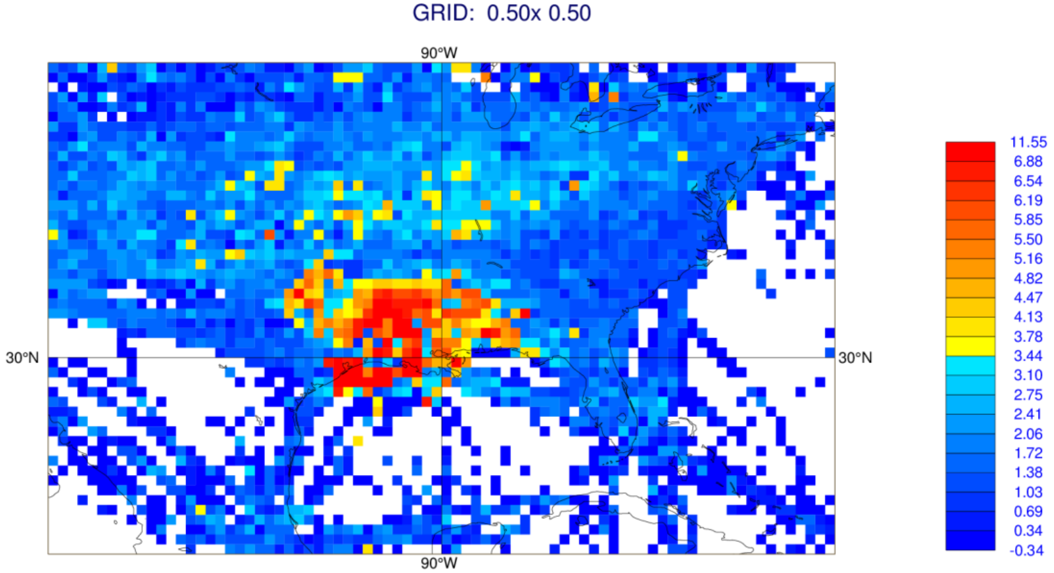


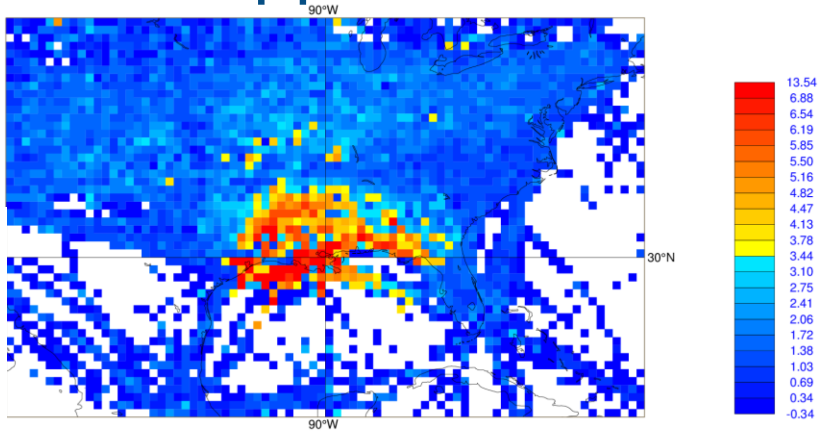
150 hPa wind analysis increments



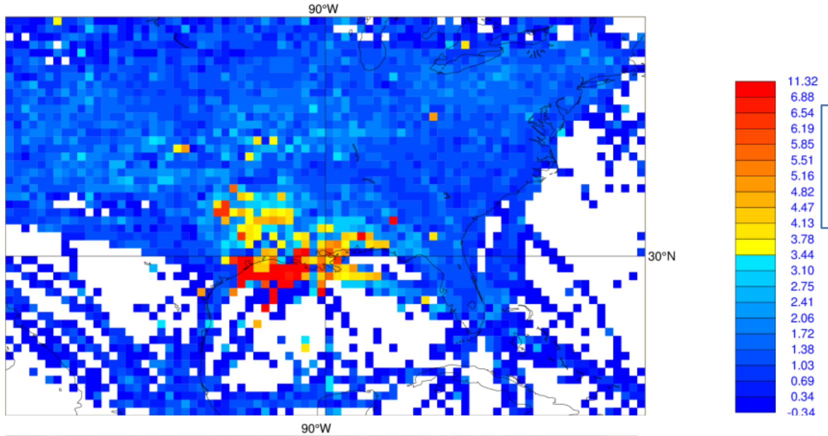
Nonlinear effects: the incremental approach

StDev of wind O-B increments, 100-400 hPa
Area Avg: **2.12** m/s

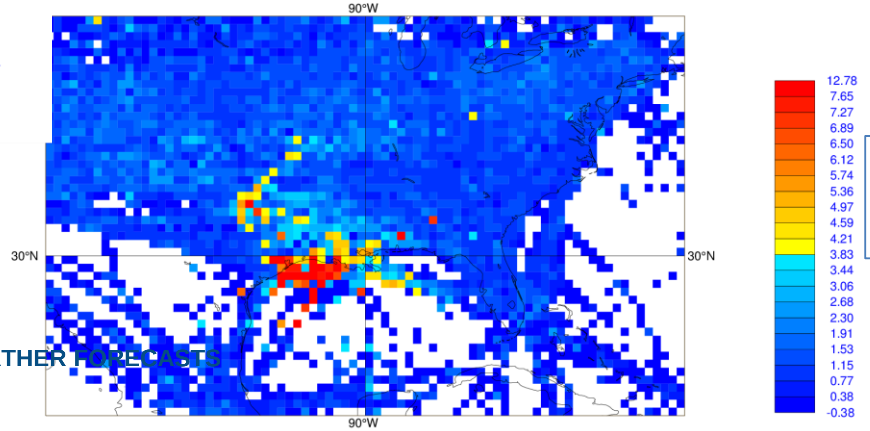




StDev of O-A inc: 1Outer L.
Area Avg: **1.91** m/s



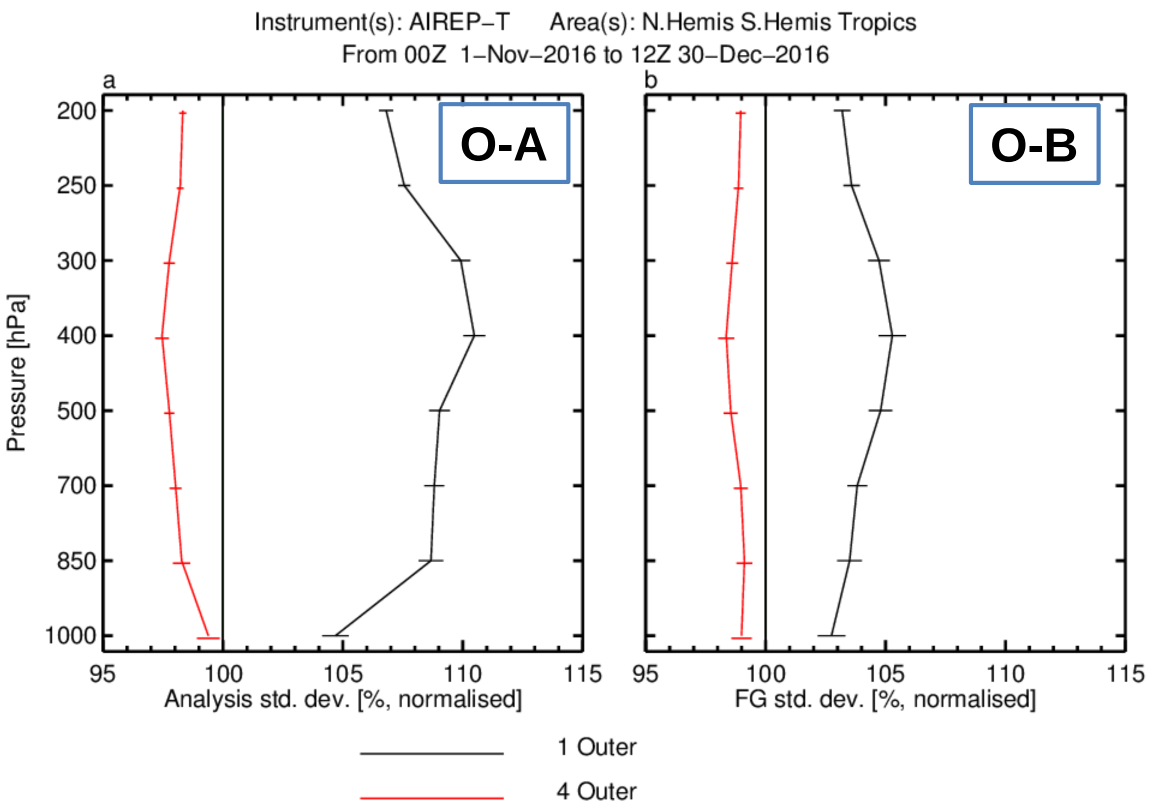
StDev of O-A inc.: 3Outer L.
Area Avg: **1.57** m/s

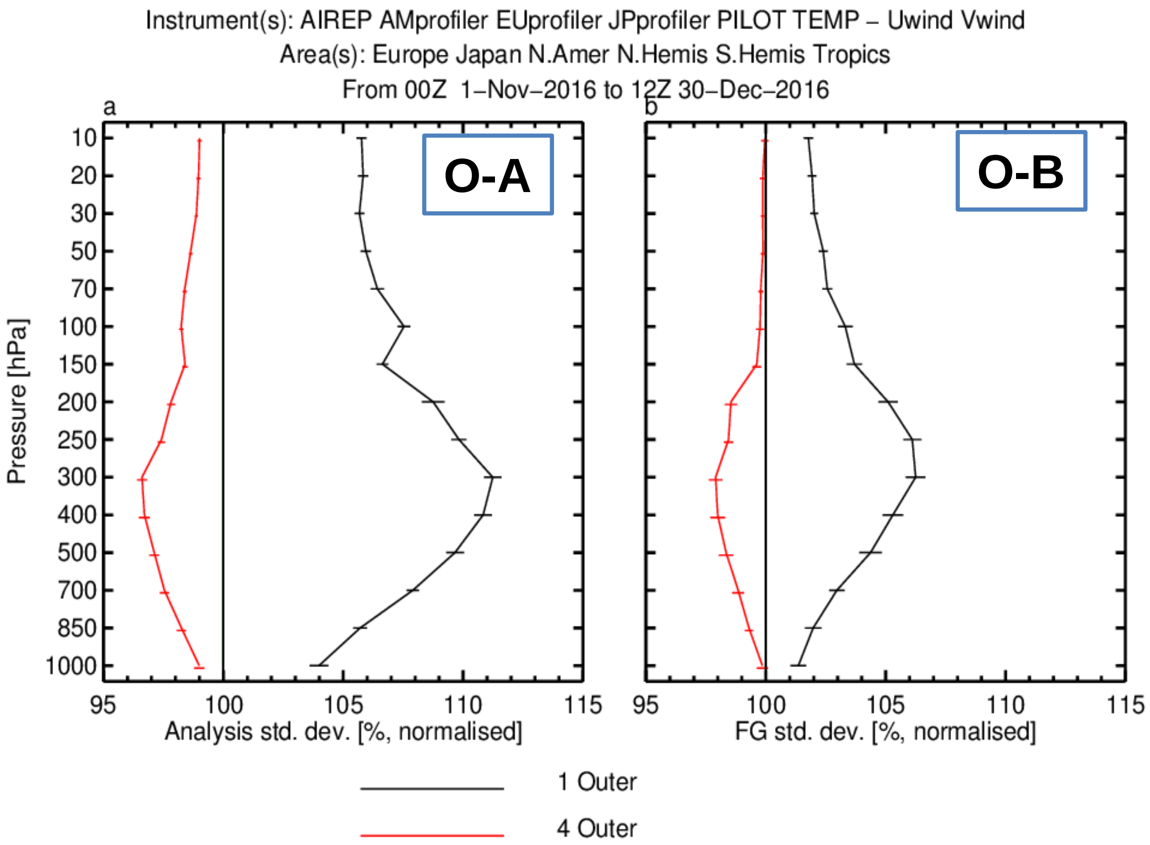


StDev of O-A inc.: 5Outer L.
Area Avg: **1.49** m/s

Nonlinear effects: the incremental approach

- How important is the capacity to run outer loops for global analysis and forecast skill?
- Relative difference of observation departures of 1 OL and 4 OL wrt 3OL control





Nonlinear effects: the incremental approach

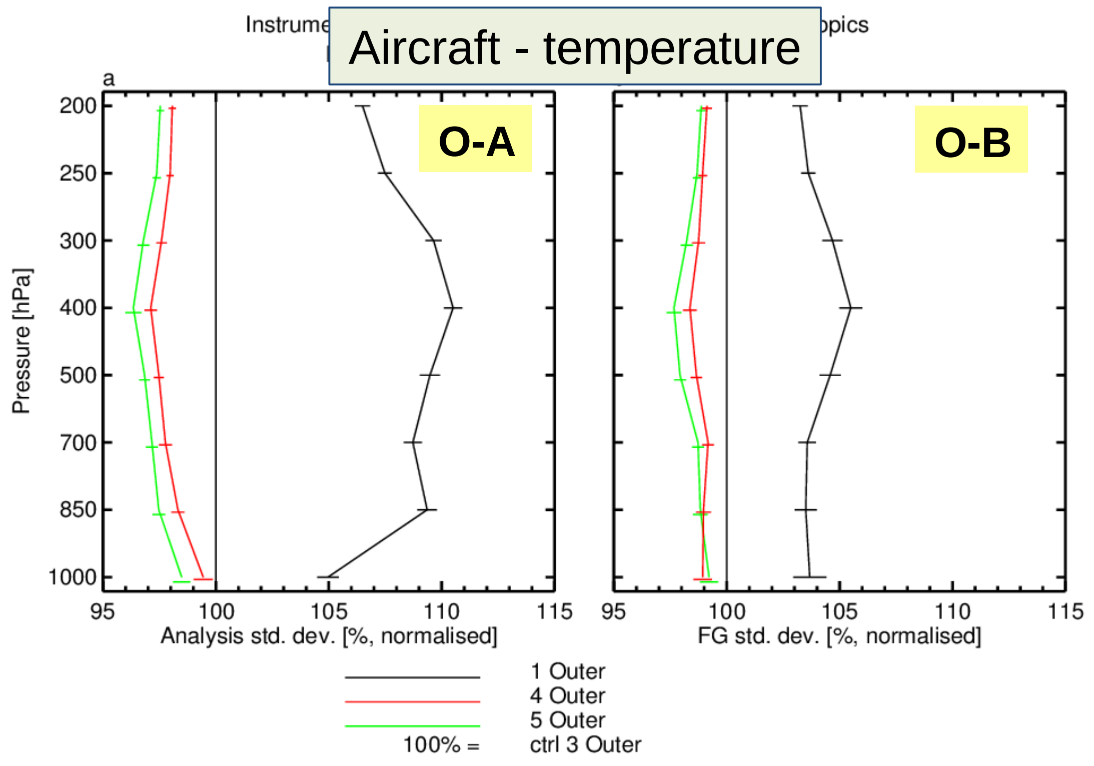
- Running incremental updates is key to control nonlinearity in 4D-Var
- This is true in EnKF/EnKS/EnsVar world as well: IEnKF/IEnKS/IES, etc (e.g. Chen and Oliver, 2012; Sakov et al, 2012; Emerick and Reynolds, 2013; Bocquet and Sakov, 2014)
- In EnKF/EnKS/EnsVar the analytic gradients are replaced by their ensemble approximations:

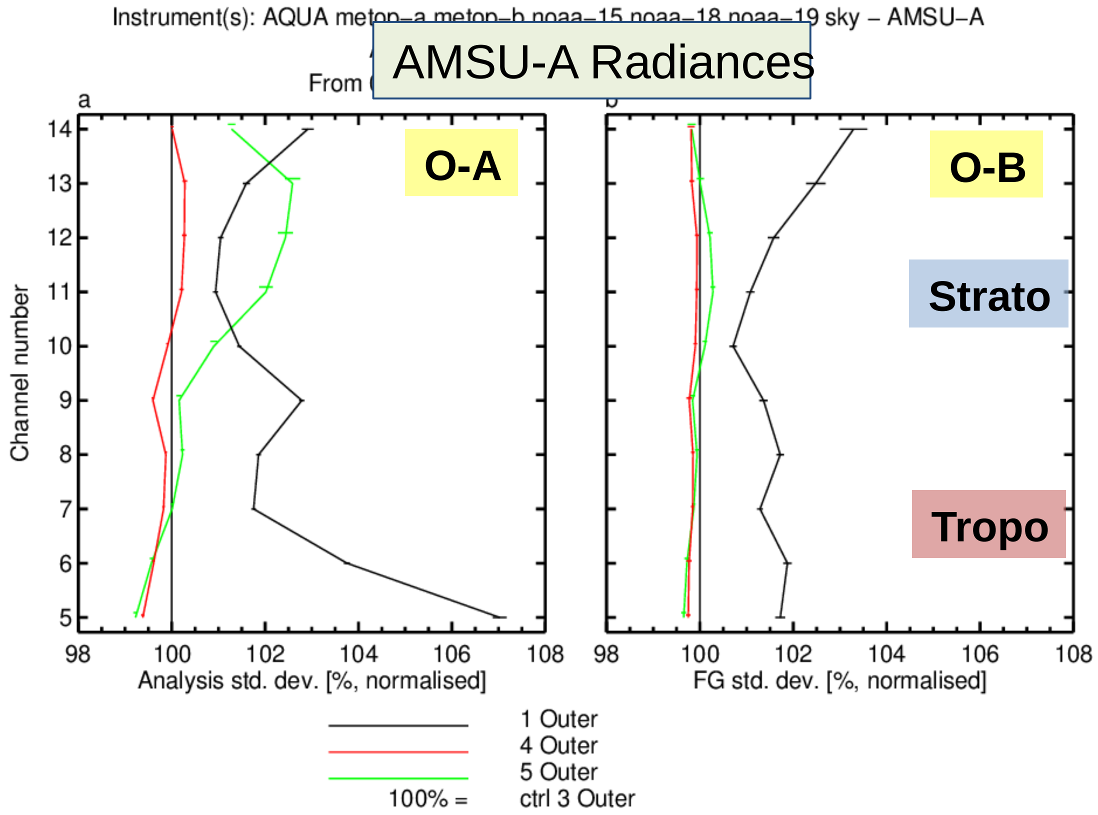
$$\mathbf{B}(\mathbf{HM})^T \rightarrow \mathbf{B}_{xy}^{ens}$$
$$(\mathbf{HM})\mathbf{B}(\mathbf{HM})^T \rightarrow \mathbf{B}_{yy}^{ens}$$

- This implies that the iterated versions of the EnKF/EnKS/EnsVar need to re-run the ensemble at each iteration to compute the updated sensitivities
- Another consequence is these ensemble re-runs cannot be computed before the observations are available

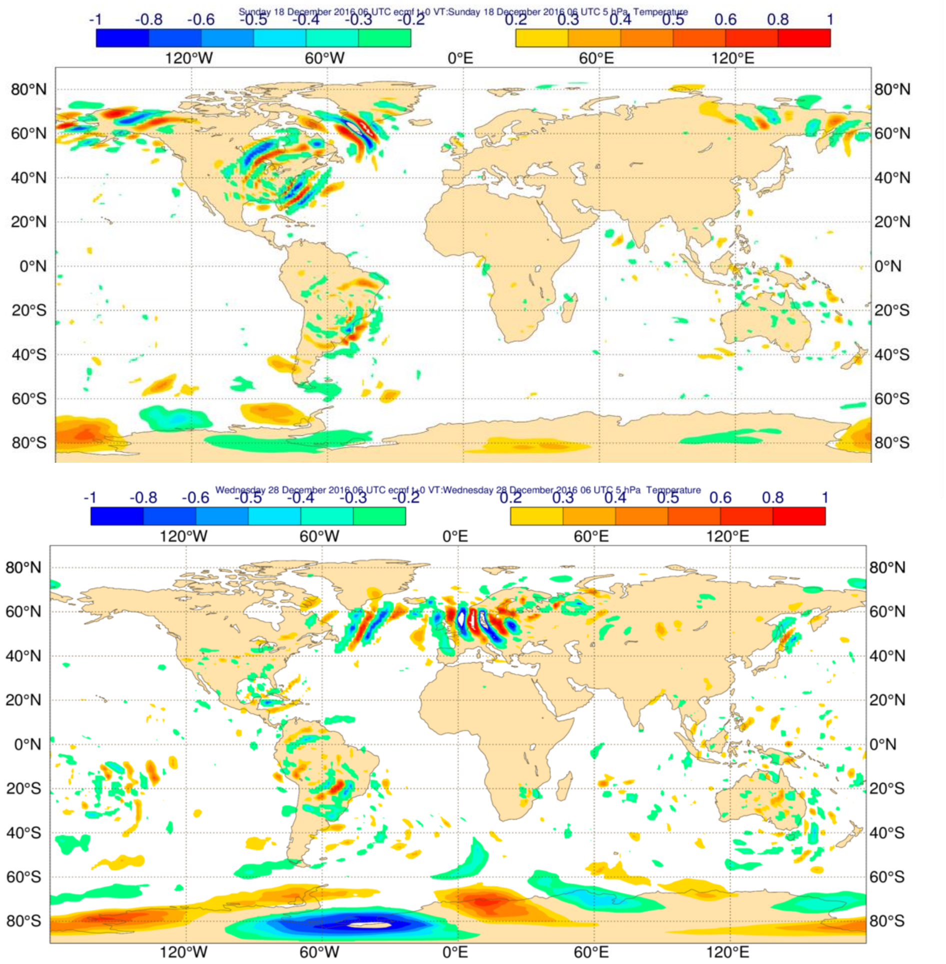
Nonlinear effects: the incremental approach

- Running incremental analysis updates is key to control nonlinearity in 4D-Var
- Can we see the limits of this approach yet?

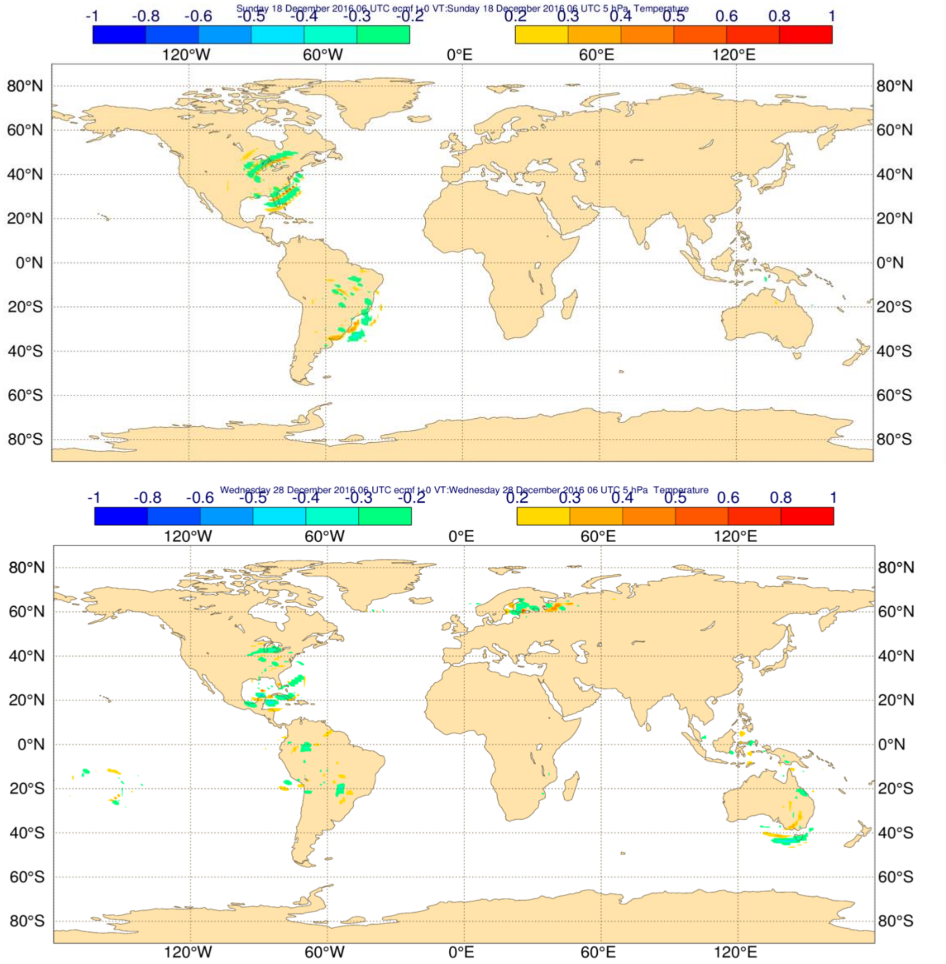




Difference between analysis increments computed by **nonlinear and linearised** models 9 hours in the assimilation window (**temperature ~5 hPa**)



Nonlinear Model tstep=450s; Linearised Model tstep=900s

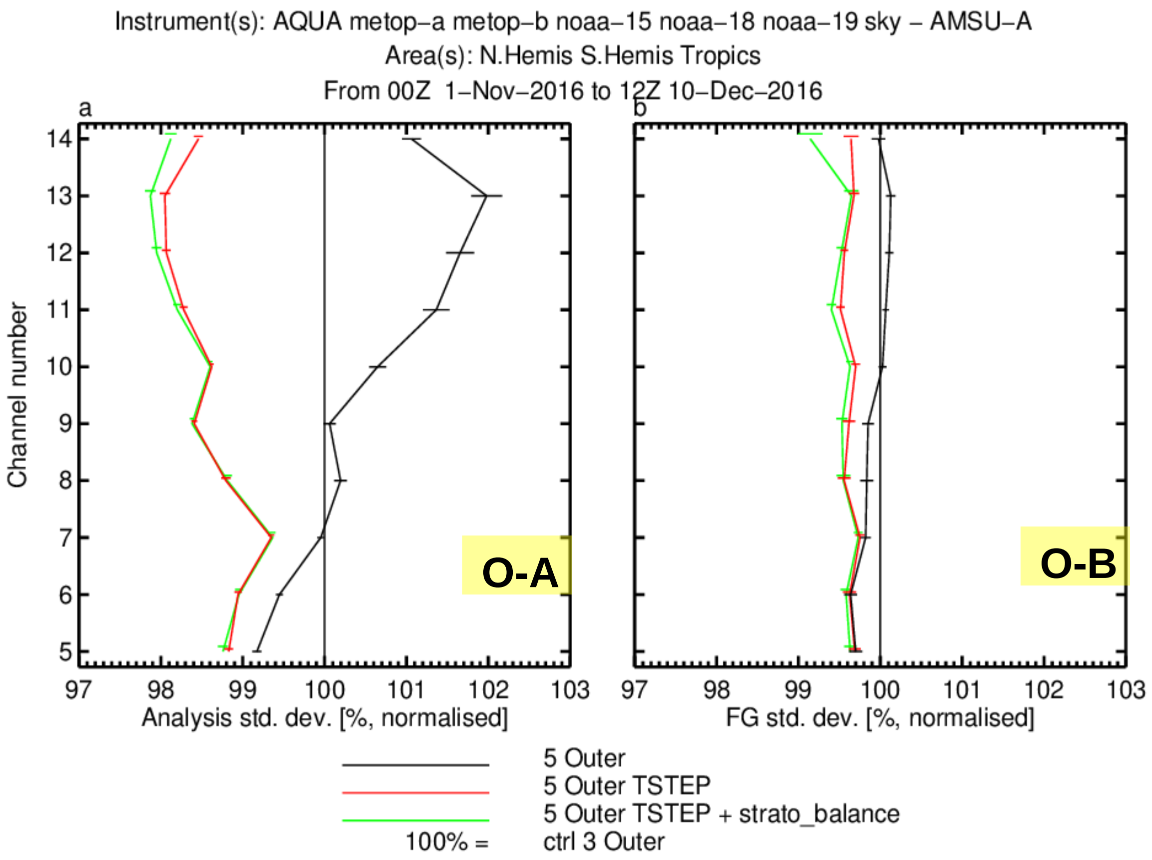


Nonlinear Model tstep=450s; Linearised Model tstep=450s

Nonlinear effects: the incremental approach

- **Matching timesteps** in outer/inner loops

AMSUA Radiances



Nonlinear effects: the incremental approach

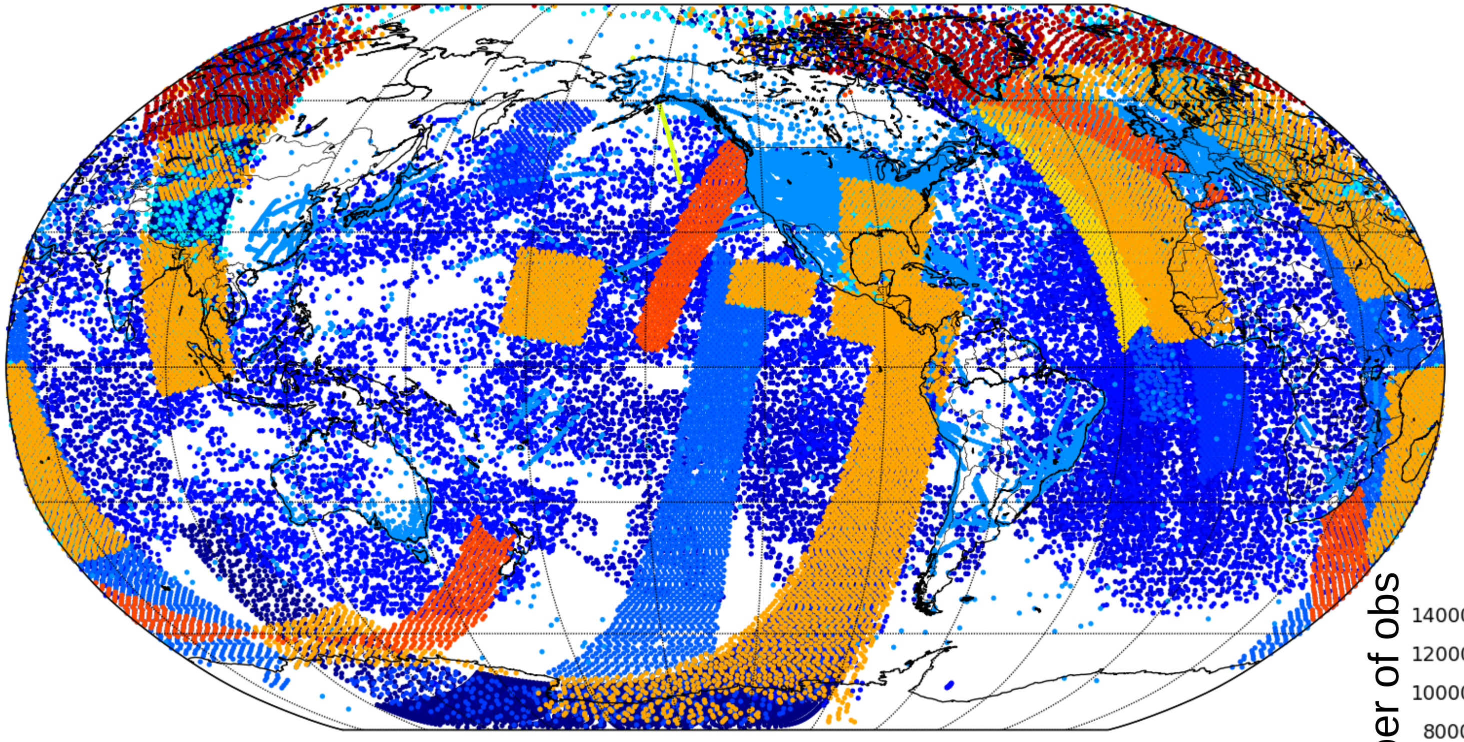
- Incremental approach is very effective in tackling nonlinearity problems arising either from the model and/or from the observations
- Its effectiveness is significantly enhanced by:
    1. Increasing the number of outer loop re-linearisations
    2. Matching outer/inner loop timesteps
    3. Increased inner loops resolution
    4. Tighter convergence criteria of the minimisation
- All of the above steps require more time for doing the analysis…
- …which means starting the analysis earlier…
- …which means using less observations

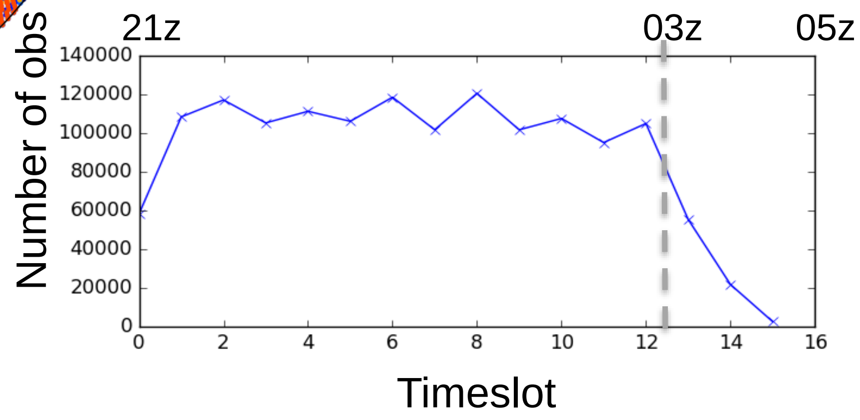
An impossible conundrum?

Continuous data assimilation (Lean et al, 2018)

- Key point: Start running data assimilation **before** all of the observations have arrived:
  1. Most of the assimilation is removed from the time critical path
  2. Configurations which were previously unaffordable can now be considered

Additional observations assimilated in Continuous DA configuration

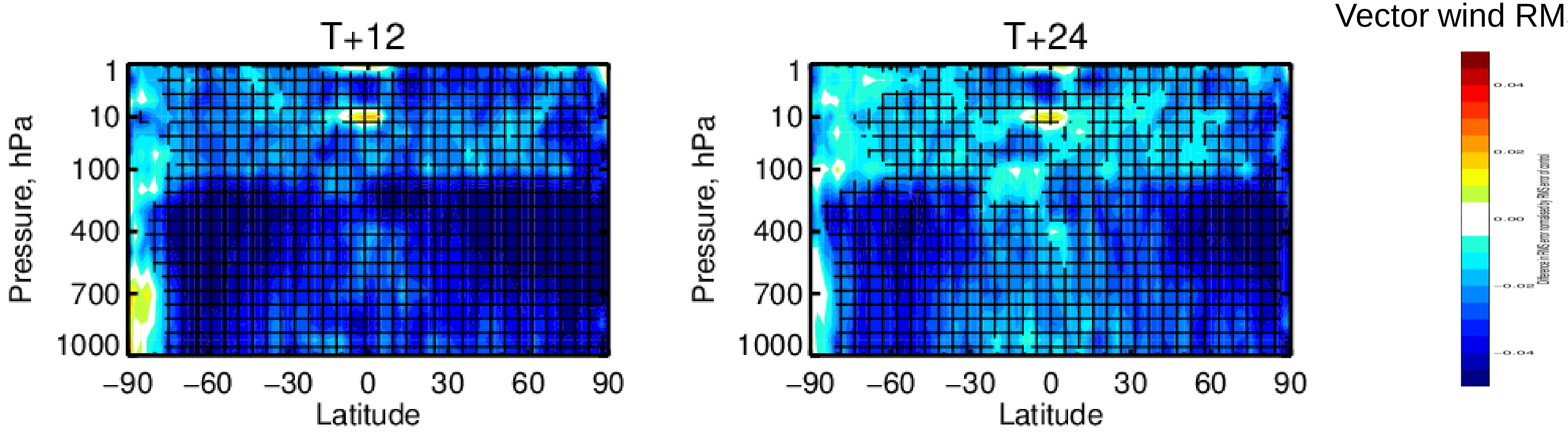




from Peter Lean

Continuous DA: Forecast improvements

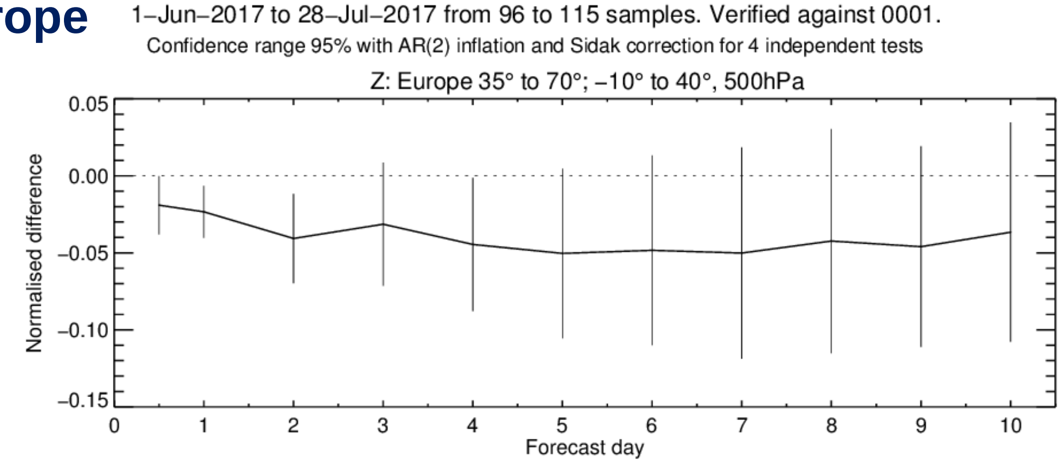
Vector wind RMS error



(bad)

(good)

Europe



Outline

- Non-Gaussianity

Non-Gaussianity

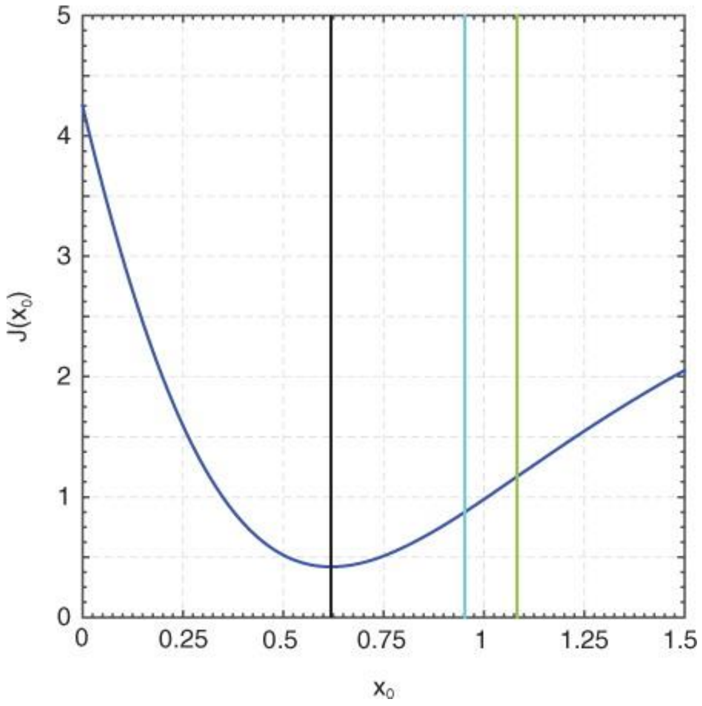
- The statistical interpretation of 4D-Var (and all other DA algorithms used in global NWP!) relies on a Gaussian assumption about the sources of information and their evolution:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^{\mathrm{T}}\mathbf{P}_b^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{y}_k - G_k(\mathbf{x}_0)\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{y}_k - G_k(\mathbf{x}_0)\right)$$

- But what if these random variables $(\mathbf{x}_0 - \mathbf{x}_b)$ and $\left(\mathbf{y}_k - G_k(\mathbf{x}_0)\right)$ are non-Gaussian?

Non-Gaussianity

- From a theoretical perspective, the 4D-Var and the BLUE solution will be different:



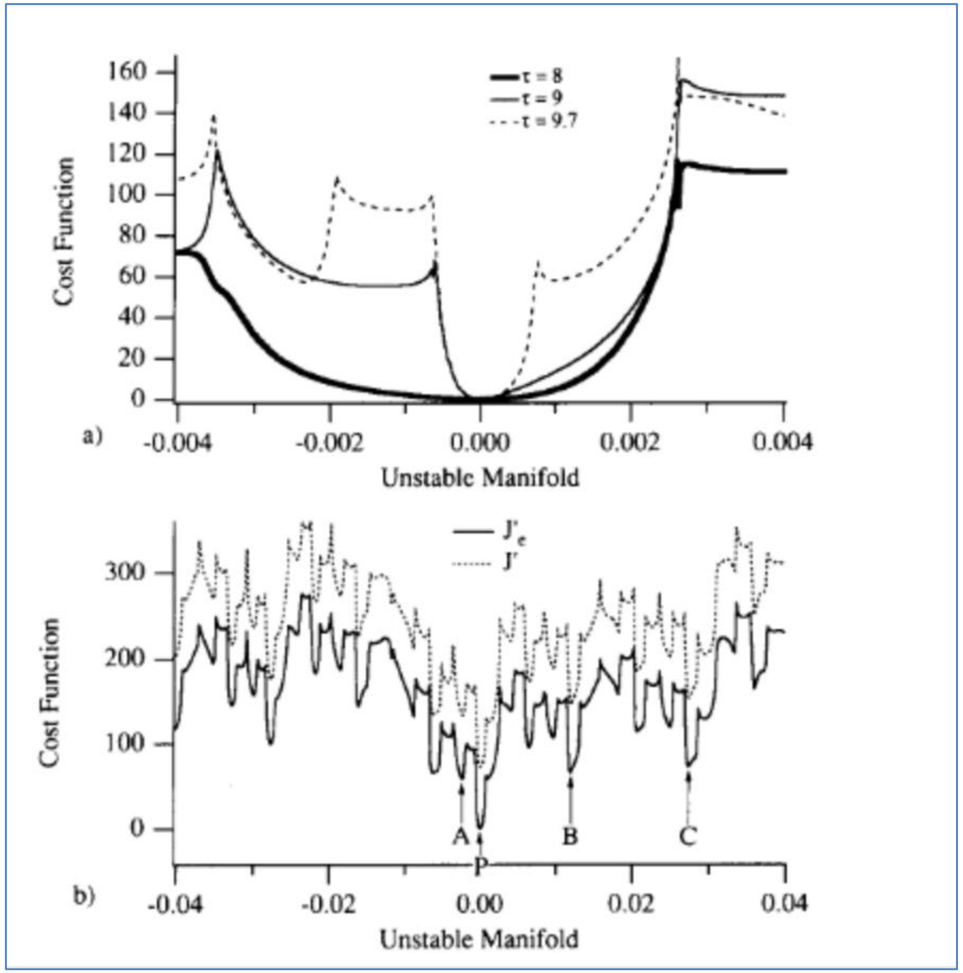
Posterior mode (increm. 4D-Var)
BLUE (Ens. Kalman smoother)
True posterior mean

Hodyss, Bishop and Morzfeld
Tellus 2016

- As importantly, the inverse of the Hessian of the cost function around the mode is a poor estimator of the posterior covariance

Non-Gaussianity

From a more practical point of view, will 4D-Var converge? What will it converge to?



Pires, Vautard and Talagrand
Tellus 1996

Non-Gaussianity: Observations

- Let us start by looking at the pdf of $\left(\mathbf{y} - G(\mathbf{x}_0)\right)$: Which errors contribute to it?

A standard derivation (e.g., Hoffman, 2018) leads to:

$$\mathbf{y} - G(\mathbf{x}_0) = \mathbf{e}^I + \mathbf{e}^H + \mathbf{e}^R - \mathbf{G}(\mathbf{x}_0 - \mathbf{x}_0^t) \ ,$$

where $\mathbf{e}^I$ is the instrument error, $\mathbf{e}^H$ are the errors of the observation operator (including its linearization) and $\mathbf{e}^R$ are the errors due to the scale mismatch between observations and model (representativeness error).

At the start of the minimization $\mathbf{x}_0 \rightarrow \mathbf{x}_0^b$:

$$\mathbf{y}^o - G\left(\mathbf{x}_0^b\right) = \mathbf{e}^I + \mathbf{e}^H + \mathbf{e}^R - \mathbf{G}\mathbf{e}^b,$$

Where $\mathbf{e}^b$ is the background error.

Except for instrument errors, all other error sources in the background departures can potentially introduce non-Gaussian errors (e.g., state-dependence of $\mathbf{e}^H$ and $\mathbf{e}^R$, non-Gaussianity of $\mathbf{e}^b$, state-dependence of $\mathbf{G}$, …).

Non-Gaussianity: Observations

- To quantify the distance of the observed pdf from the expected Gaussian distribution we use a measure common in information theory, the **Kullback–Leibler divergence** (D_{KL})

- KLD measures the distance of the sample distribution $\mathbf{Q}$ to the prior distribution $\mathbf{P}$ by:

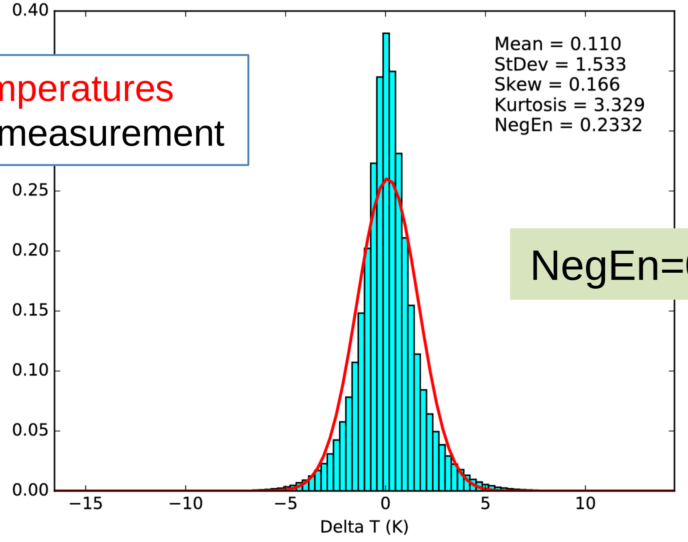
$$D_{KL}(Q||P) = \sum_i Q(i)\log(Q(i)/P(i))$$

- When the prior distribution is a Gaussian, D_{KL} is called the **negentropy** of the sample distribution

- The negentropy is always positive and is equal to **zero** iff the sample Q is **Gaussian** almost everywhere

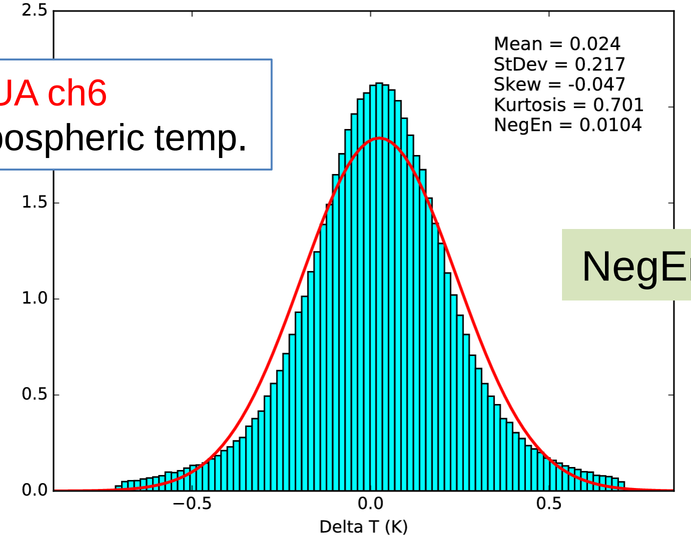
Non-Gaussianity: Observations

Radiosonde temperatures
Point temperature measurement



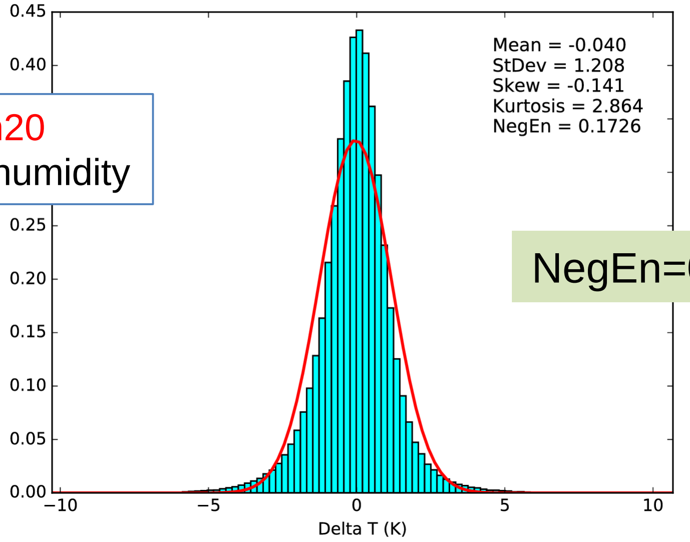
NegEn=0.2332

AMSUA ch6
Integrated tropospheric temp.



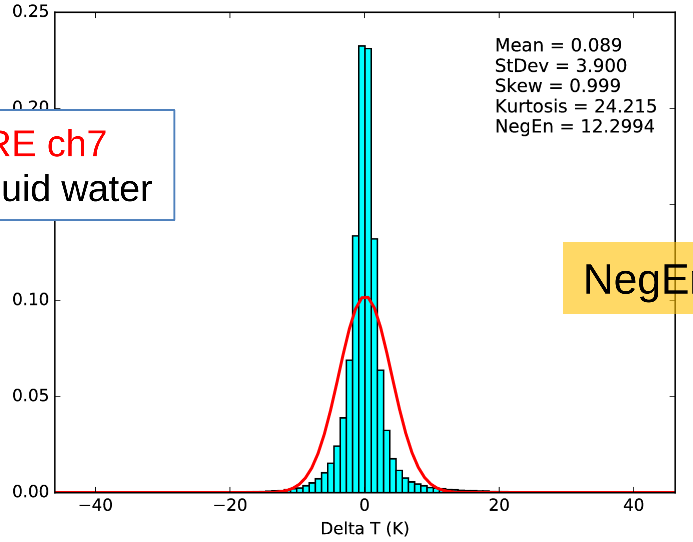
NegEn=0.0104

ATMS ch20
Tropospheric humidity



NegEn=0.1726

AMSRE ch7
Cloud liquid water



NegEn=12.2994

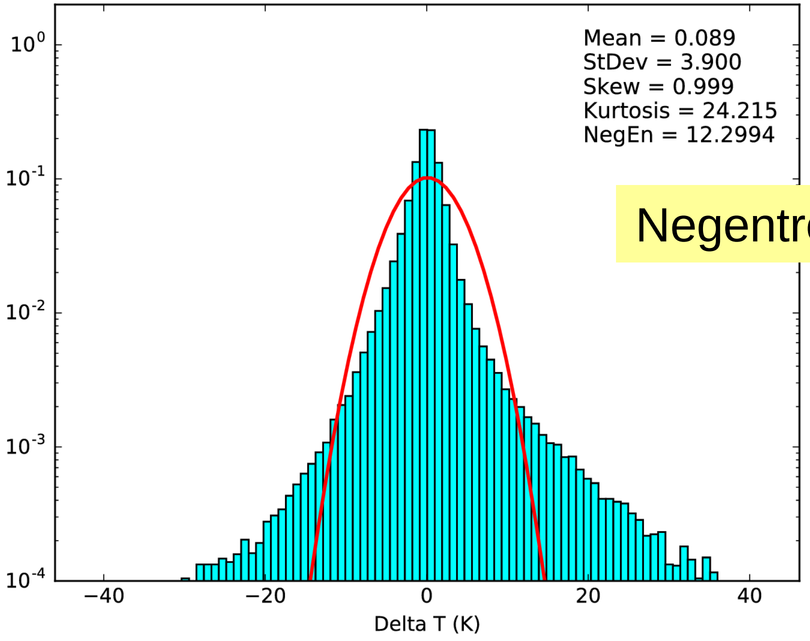
Non-Gaussianity: Observations

- Dealing with non-Gaussianity: **Gaussian Anamorphosis**
  - Transform variable of interest into new variable with (more) Gaussian statistics, and perform analysis in new space
  - This transformation can be applied to the observed quantities, the control variable or both (Amezcua and van Leeuwen, 2014)
  - In the variational minimization we actually need the normalised departures: $\mathbf{R}^{-1/2}\left(\mathbf{y} - G\left(\mathbf{x}_0\right)\right)$
  - Thus, a way to achieve Gaussian anamorphosis is to identify an "observation" error model that makes the normalised departures more nearly Gaussian (Geer and Bauer, 2011)
  - Note that also the **Huber norm** (Tavolato and Isaksen, 2015) can be viewed as a form of Gaussian Anamorphosis (Bonavita et al., 2017)

Non-Gaussianity: Observations

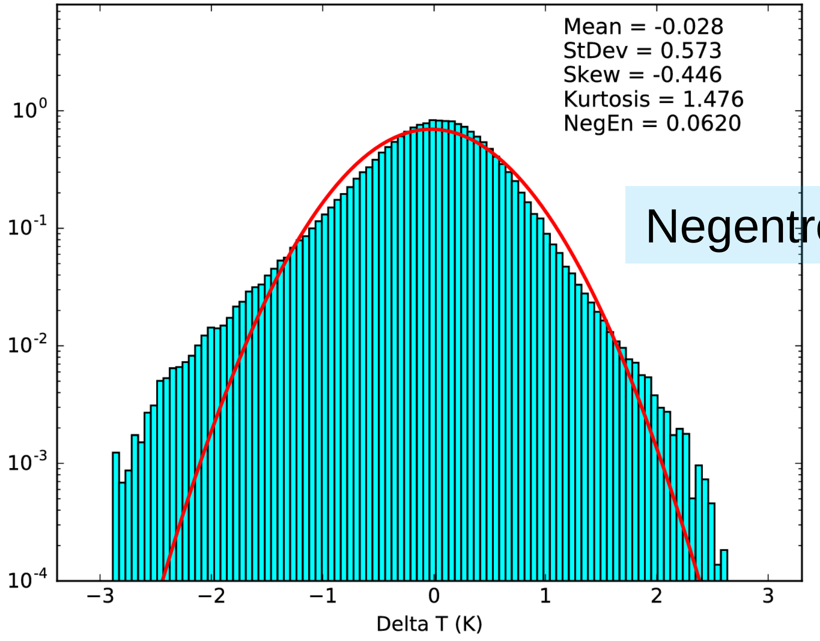
- Dealing with non-Gaussianity: **Gaussian Anamorphosis**

AMSR-2 ch. 7: $\mathbf{y} - G(\mathbf{x}_0)$



Negentropy=12.299

AMSR-2 ch. 7: $\mathbf{R}^{-1/2}\left(\mathbf{y} - G(\mathbf{x}_0)\right)$

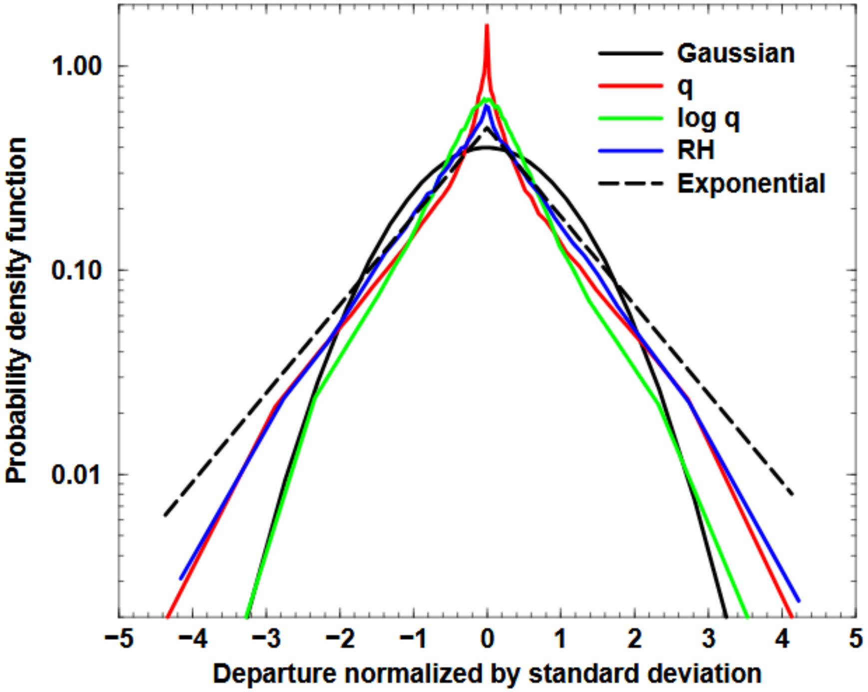


Negentropy=0.062

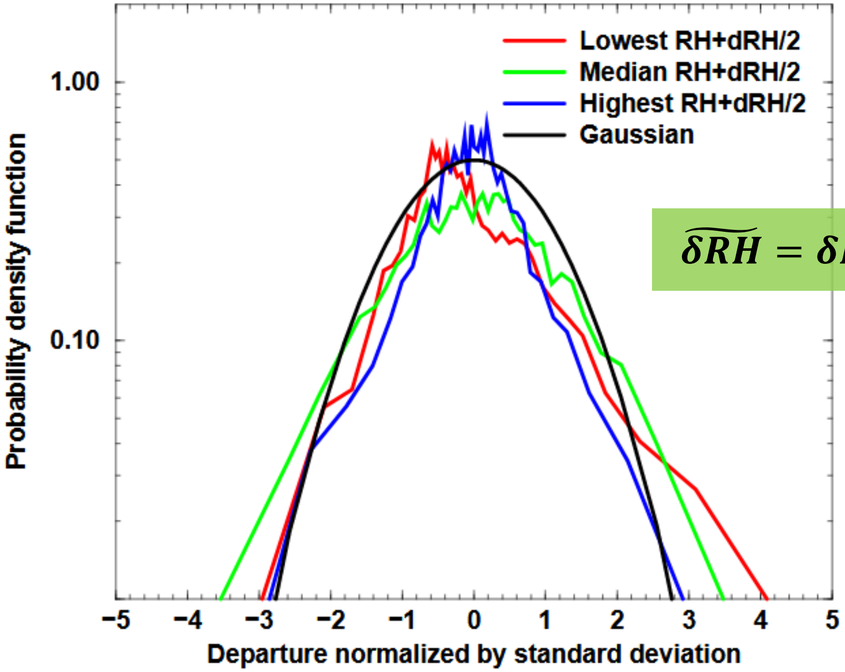
Non-Gaussianity

- The same approach of Gaussian anamorphosis can be used for the control variable $(\mathbf{x}_0 - \mathbf{x}_b)$
- A typical example is the humidity variable, which is physically bounded and presents large spatial/temporal variability

Distribution of EDA background fcst differences for different humidity variables at 850hPa.
From Hólm et al, 2003



Distribution of EDA background fcst differences for the **Hólm humidity CV** at 850hPa.
From Hólm et al, 2003



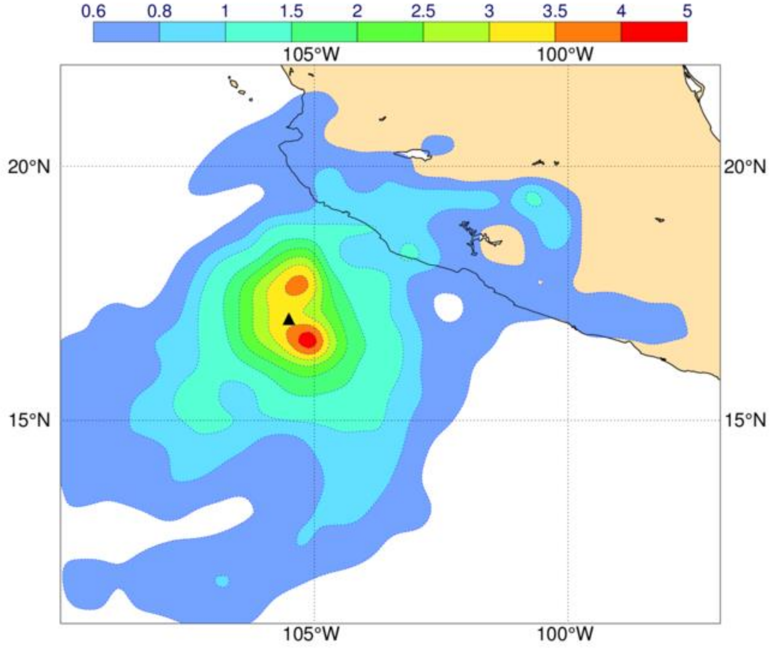
$$\widetilde{\delta RH} = \delta RH / \sigma\left(RH^b + {}^1\!/_2\, \delta RH\right)$$

OR

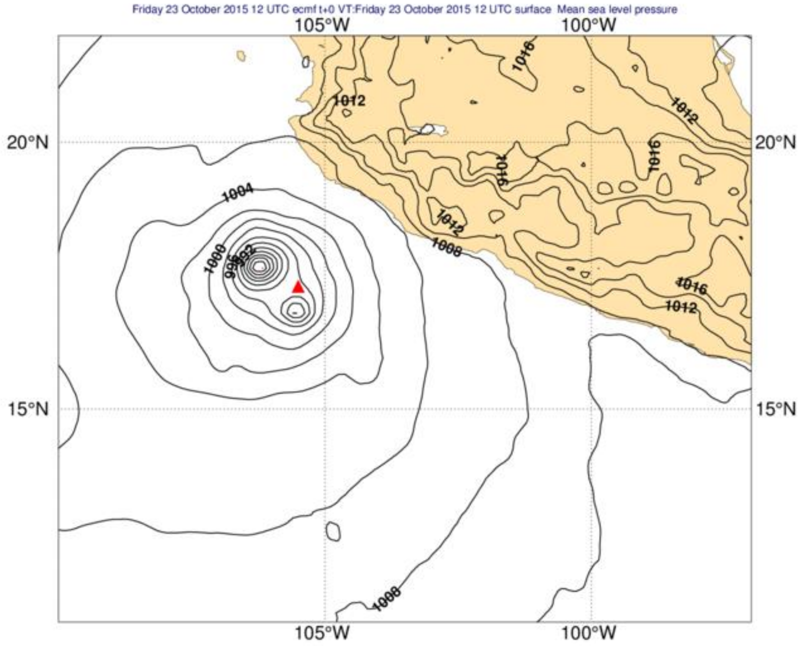
Non-Gaussianity

- Non-Gaussianity in humidity/cloud/precipitation variables is “expected”
- We have seen some of the regularisation techniques used to deal with it
- With increasing model resolution, and thus increasing nonlinearities, we can also expect to see macroscopic effects of non-Gaussianity in the BG forecasts of dynamical fields too!

0069 IFS **B** MSLP errors 23-10-2015 12UTC



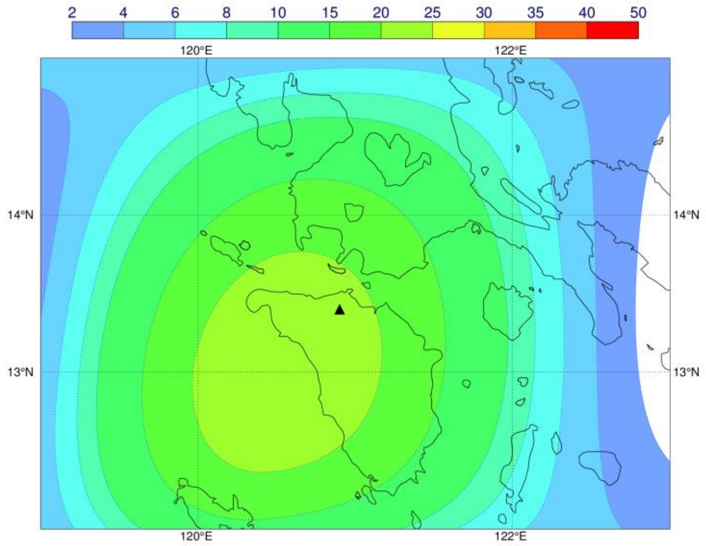
0069 IFS MSLP Analysis 23-10-2015 12UTC



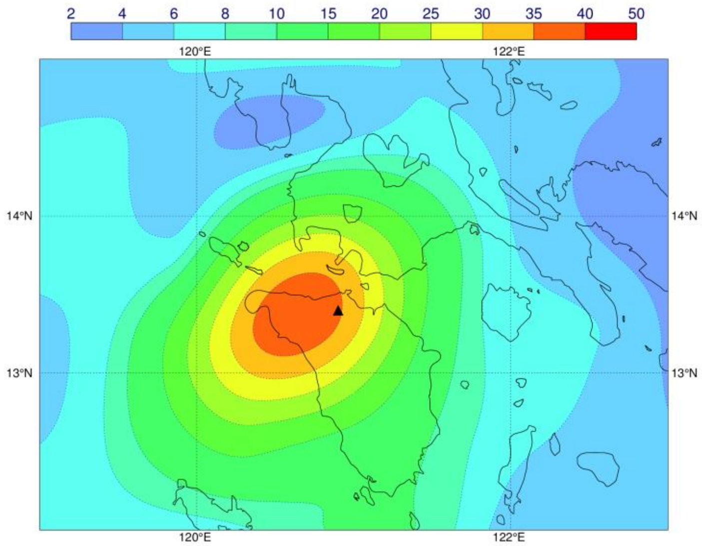
Non-Gaussianity

- Sometimes the non-Gaussianity in the forecasts is pathological, i.e. it is a symptom of other problems in the assimilation (bad convergence, initialisation issues, etc.)
- But other times the non-Gaussianity in the BG forecasts is structural, i.e. it reflects true uncertainty in the analysis/short-range evolution
- In general, this uncertainty is highly **scale-dependent**: the closer (i.e., higher resolution) we look the more non-Gaussian features will appear!

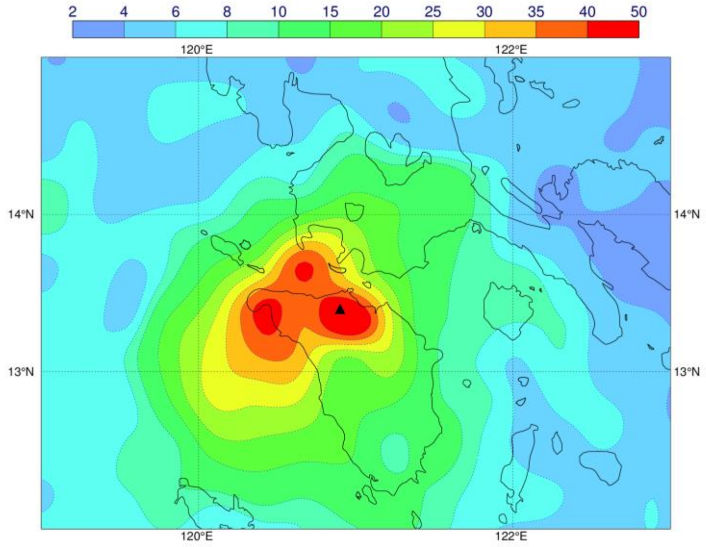
Typhoon Melor, EDA Vorticity errors 850hPa, **T159**



Typhoon Melor, EDA Vorticity errors 850hPa, **T319**



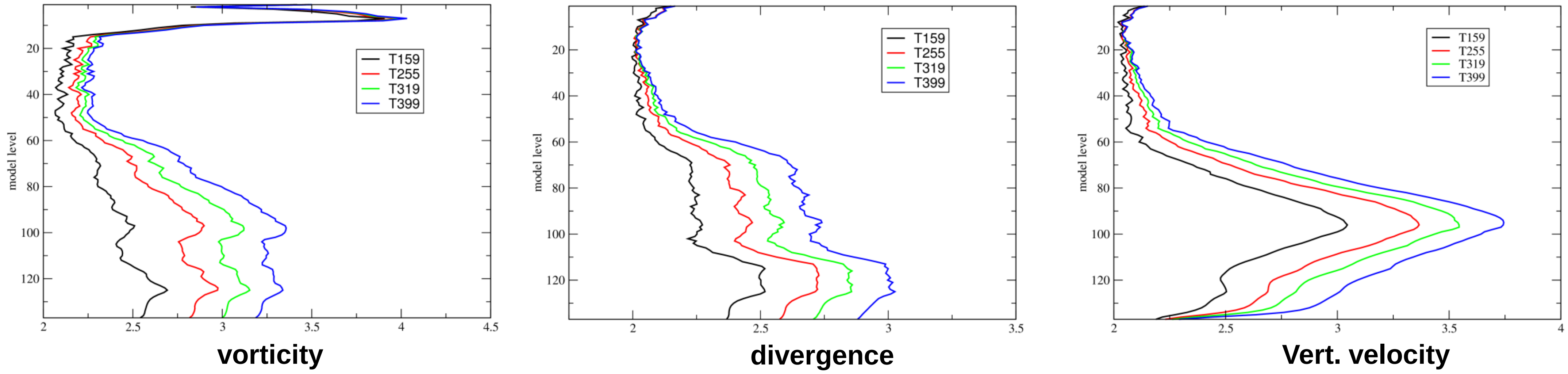
Typhoon Melor, EDA Vorticity errors 850hPa, **T639**



Non-Gaussianity

- We can quantify this scale-dependency of non-Gaussianity, using e.g. a standard test of normality
- In the following we use the D'Agostino K^2 metric (D'Agostino et al, 1990) to detect deviations from normality due to skeweness and kurtosis
- For distributions close to Gaussian $K^2 \sim \chi^2(k=2) \rightarrow mean, stdev \approx 2$

EDA bg distribution K^2 as a function of spectral trunc.



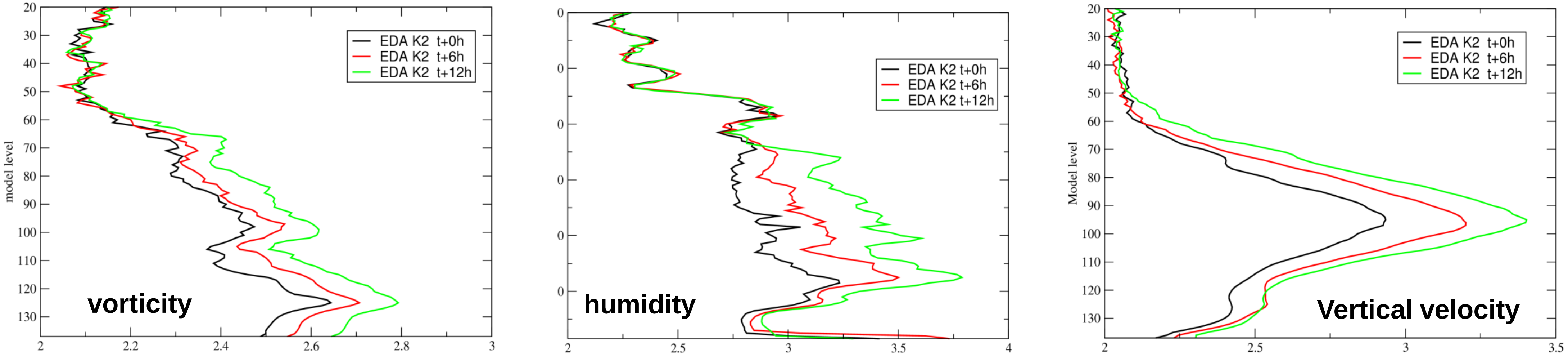
Non-Gaussianity

- Using error estimates at higher resolution increases non-Gaussianity of the assimilation problem → higher condition numbers in minimisation, increased risk of poor convergence → spatial filtering of error covariances helps (not only for filtering sampling noise!)
- Also suggests that progressive increase of minimisation resolution in multi-incremental 4D-Var is a good idea!
- Choice of more well-behaved dynamical control variables could also be helpful as for the humidity cv (e.g., Legrand et al, 2016)

Non-Gaussianity

- Non-Gaussianity of control variable is also very sensitive to length of nonlinear model forecast:

K^2 of EDA forecasts as a function of forecast range



Non-Gaussianity

- The rapid growth of non-Gaussian errors in the assimilation window a points to the opportunity of **more frequent analysis updates and shorter assimilation windows**
- Going to shorter assimilation windows is possible, but it has its drawbacks…

1. Initial balance: it currently takes 6 to 9 hours for the initial adjustment process of the IFS (Bonavita et al, 2017)
2. It takes 6 to 12 hours to synchronise atmosphere-ocean updates in the outer loop coupling framework (Laloyaux et al, 2018)
3. 12h window 4D-Var has marginally better scores than 6h window 4D-Var
4. More frequent analysis updates impose stronger constraints on observations' timeliness (though continuous DA provides a solution)

Non-Gaussianity: Global Optimization of non-convex functions

- Non-Gaussian errors lead to non-convex functions and the hard problem of their minimisation

- One type of approach to non-convex global minimisation is systematic.
  - Methods are guaranteed to converge with a predictable amount of work, e.g. **simulated annealing**, **grid box, genetic sampling** methods.
  - Drawback: amount of work makes them intractable for large-dimensional problems

Non-Gaussianity: Global Optimization of non-convex functions

- There is however a class of **perturbative** methods that have been found to work well in a number of real world, large scale global minimisation problems (Wu, 1997; Mohabi et al, 2015): Homotopy/Continuation methods
  - Embed the target cost function $J(\mathbf{x})$ in a family of cost functions $\{J_0(\mathbf{x}), J_1(\mathbf{x}), .., J_n(\mathbf{x})\}$ such that $J_n(\mathbf{x}) = J(\mathbf{x})$ and $J_{i-1}(\mathbf{x})$ is more "well-behaved" than $J_i(\mathbf{x})$ (i.e., it is convex over a larger subset of the CV space)
  - The algorithm proceeds by solving a sequence of progressively harder optimization problems starting from the solution of the previous minimization and using standard convex minim. tools

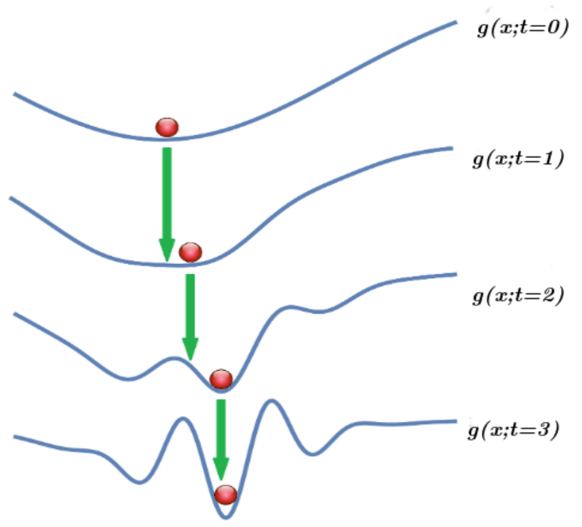


Figure 1: Plots show g versus x for each fixed time t.

Algorithm 1 Algorithm for Optimization by Continuation Method

1: Input: $f : \mathcal{X} \to \mathbb{R}$, Sequence $t_0 > t_1 > \cdots > t_n = 0$.

2: x_0 = global minimizer of $g(x; t_0)$.
3: **for** $k = 1$ **to** n **do**
4: x_k = Local minimizer of $g(x; t_k)$, initialized at x_{k-1}.
5: **end for**
6: Output: x_n

From Mohabi and Fisher, 2015

Non-Gaussianity: Global Optimization of non-convex functions

- What would a Homotopy/Continuation algorithm look like in the context of our sequential, incremental 4D-Var based DA?

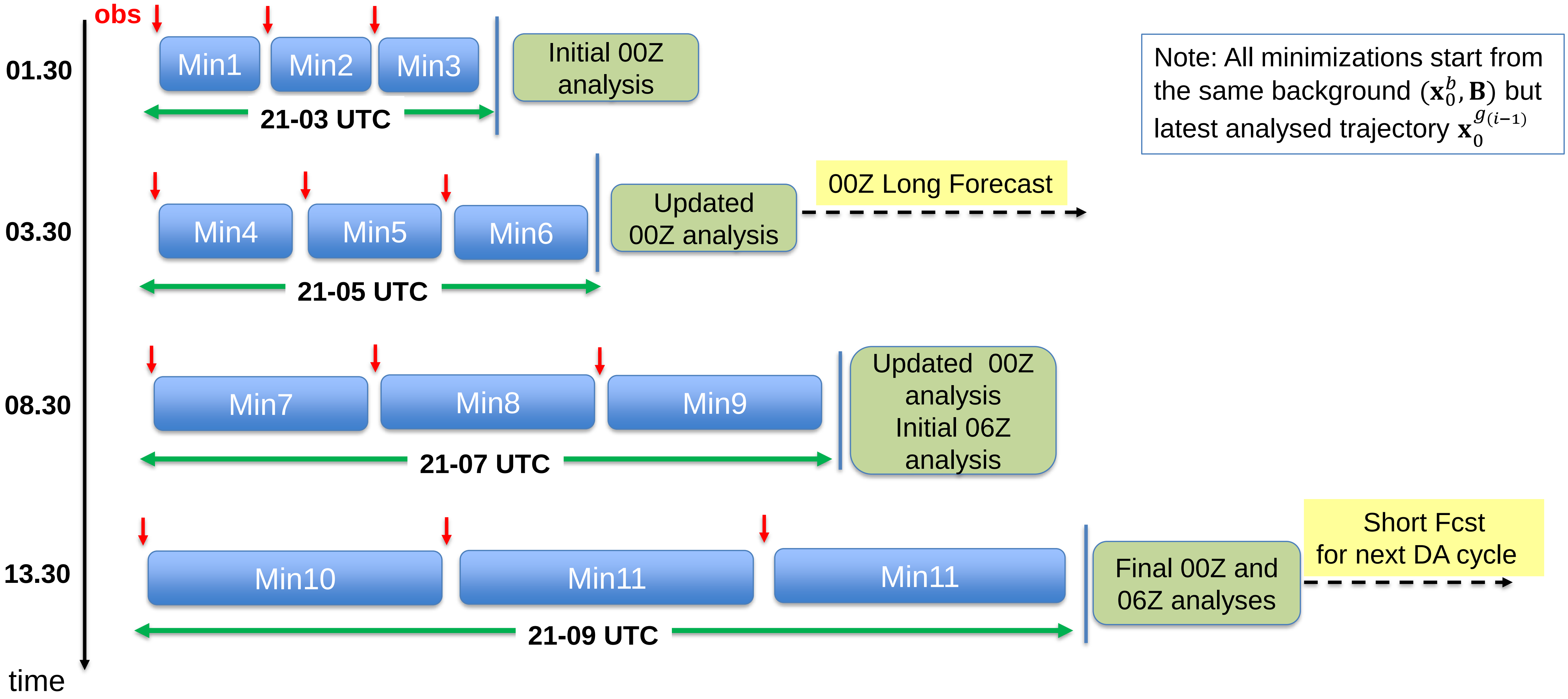
- Let us recall the main sensitivities of DA to nonlinearity/non-Gaussianity:

  1. Resolution of the minimization;

  2. Length of the assimilation window

- Based on these ideas, a 4D-Var Homotopy/Continuation algorithm would look like a natural extension of the continuous DA concept, i.e.:

Continuous Long Window DA



Continuous Long Window DA

- The homotopy defined by the continuous LWDA embeds the target 12-hour cost function in a series of increasingly nonlinear cost functions over a progressively longer assimilation window ($\{J_6(\mathbf{x}), J_8(\mathbf{x}), J_{10}(\mathbf{x}), J_{12}(\mathbf{x})\}$

- Each set of analysis updates goes from low to high resolution minimisations

- Continuous LWDA extends in the time domain the multi-incremental 4D-Var approach

- Continuous LWDA has significant additional advantages for operational NWP:

    1. A single DA cycle (No need for a separate “Early Delivery” suite)
    2. A more time-uniform exploitation of available computing resources
    3. Improved timeliness of analysis products
    4. Vastly improved resilience

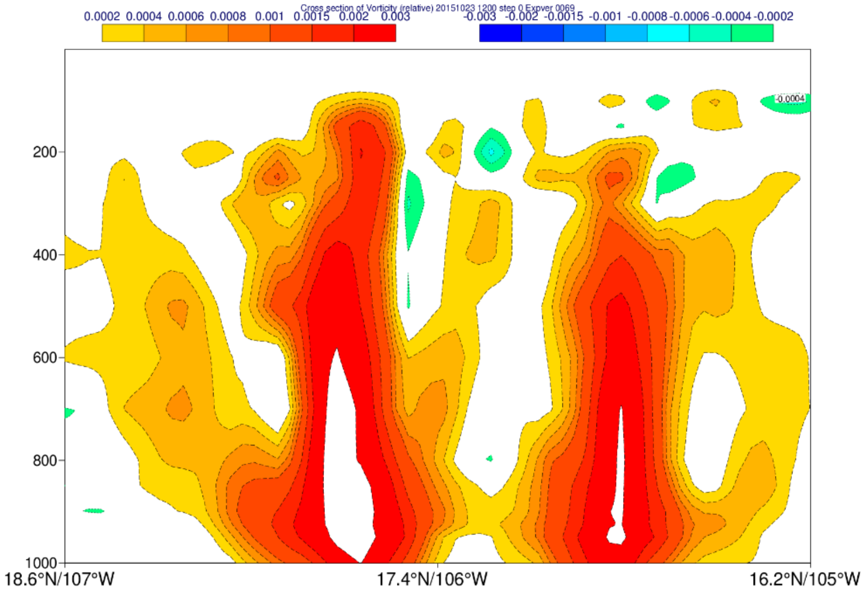
- It is not a completely new idea in NWP. Variations on this idea were already discussed by Pires et al, 1996 and Jarvinen et al, 1996!!

Final remarks

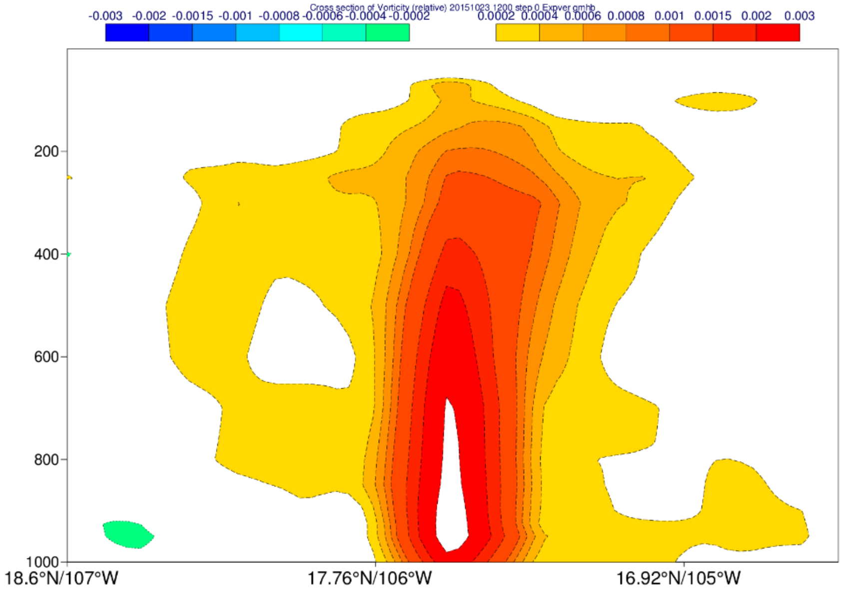
- Nonlinearity and non-Gaussianity are a central theme for data assimilation methods in global NWP, and even more in DA for Local Area Models
- They are bound to become even more important in the future as models increase in resolution and fidelity and the majority of new observations are increasingly nonlinear (we want to assimilate lightning obs!!)
- Fully nonlinear, non-Gaussian DA methods are computationally intractable for global NWP (but wait for the next lecture!)
- In the data assimilation toolbox there are a number of methods that can be applied to deal with these problems based on two general ideas: regularization of the problem and perturbative convergence to solution
- Continuous DA and Continuous Long Window DA provide a promising framework to control nonlinearity and non-Gaussianity in an effective and <u>efficient</u> manner

Many thanks for your attention!

Hurricane Patricia v1.0



Hurricane Patricia v2.0



References

Amezcua J. and P. J. Van Leeuwen, 2014: Gaussian anamorphosis in the analysis step of the EnKF: a joint state-variable/observation approach, Tellus A: Dynamic Meteorology and Oceanography, 66:1

Andersson, E., M. Fisher, E. Hólm, L. Isaksen, G. Radnóti, Y. Trémolet, 2005: Will the 4D-Var approach be defeated by nonlinearity? ECMWF Tech. Memo. N. 479

Bonavita, M., M. Dahoui, P. Lopez, F. Prates, E. Hólm, G. De Chiara, A. Geer, L. Isaksen and B. Ingleby, 2017: On the Initialisation of Tropical Cyclones. ECMWF Tech. Memo. n. 810.

Bonavita, M., P. Lean and E. Hólm, 2018: Nonlinear effects in 4D-Var. Nonlin. Proc. Geo., https://doi.org/10.5194/npg-2018-20, 2018.

Bocquet, M., C.A. Pires, and L. Wu, 2010: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. Mon. Wea. Rev., 138, 2997–3023

Courtier, P., Thépaut, J.-N. and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol. Soc., 120, 1367-1387, doi: 10.1002/qj.49712051912, 1994

D’Agostino, R. B., A. Belanger and R B. D’Agostino, Jr, 1990: A suggestion for using powerful and informative tests of normality. The American Statistician. 44 (4): 316–321

Emerick, A.A. and A.C. Reynolds, 2013: Ensemble Smoother with multiple data assimilation. Comput. Geosci. 55, 3-15

Evensen, G., 1997: Advanced Data Assimilation for Strongly Nonlinear Dynamics. Mon. Wea. Rev., 125, 1342–1354

Evensen, G., Analysis of iterative ensemble smoothers for solving inverse problems, Computat. Geosci., 2018a.

Geer, A. J. and Bauer, P.: Observation errors in all-sky data assimilation. Q.J.R. Meteorol. Soc., 137: 2024–2037.

Hodyss, D., C. H. Bishop and M. Morzfeld, 2016: To what extent is your data assimilation scheme designed to find the posterior mean, the posterior mode or something else?, Tellus A: Dynamic Meteorology and Oceanography, 68:1

Hoffman, R.N., 2018: The Effect of Thinning and Superobservations in a Simple One-Dimensional Data Analysis with Mischaracterized Error. Mon. Wea. Rev., 146, 1181–1195, https://doi.org/10.1175/MWR-D-17-0363.1

Hólm, E.V., Andersson, E., Beljaars, A.C.M., Lopez, P., Mahfouf, J.-F., Simmons, A., Thépaut, J.-J., 2002: Assimilation and 5 Modelling of the Hydrological Cycle: ECMWF's Status and Plans. ECMWF Tech. Memo. 383, 2002

Huber, P.J, 1981: Robust Statistics. John Wiley & Sons.

Järvinen, H. , Thépaut, J. and Courtier, P., 1996: Quasi-continuous variational data assimilation. Q.J.R. Meteorol. Soc., 122: 515-534.

Legrand, R., Michel, Y., and Montmerle, T., 2016: Diagnosing non-Gaussianity of forecast and analysis errors in a convective-scale model, Nonlin. Processes Geophys., 23, 1-12

Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. J. Atmos. 30 Sci., 51, 1037–1056

Mobahi, H. and J. W. Fisher III, 2015: A Theoretical Analysis of Optimization by Gaussian Continuation, Proc. of 29th Conference on Artificial Intelligence (AAAI)

Neal, R.M., 1996: Sampling from multimodal distributions using tempered transitions. Stat. Comput. 6(4), 353-366.

Pires, C., Vautard, R. and Talagrand, O., 1996: On extending the limits of variational assimilation in nonlinear chaotic systems, Tellus. A, 48, 96–121.

Radnòti, G., Y. Trémolet, E. Andersson, L. Isaksen, E. Hólm, M. Janiskova: Diagnostics of linear and incremental approximations in 4D-Var revisited for higher resolution analysis. ECMWF Tech Memo 479, 2005.

Trémolet, Y., 2004: Diagnostics of linear and incremental approximations in 4D-Var. Q.J.R. Meteorol. Soc., 130: 2233–2251.

Verlaan, M. and A.W. Heemink, 2001: Nonlinearity in Data Assimilation Applications: A Practical Method for Analysis. Mon. Wea. Rev., 129, 1578–1589

Wu, Z. (1997). Global continuation for distance geometry problems. SIAM Journal of Optimization, 7, 814–836. 323

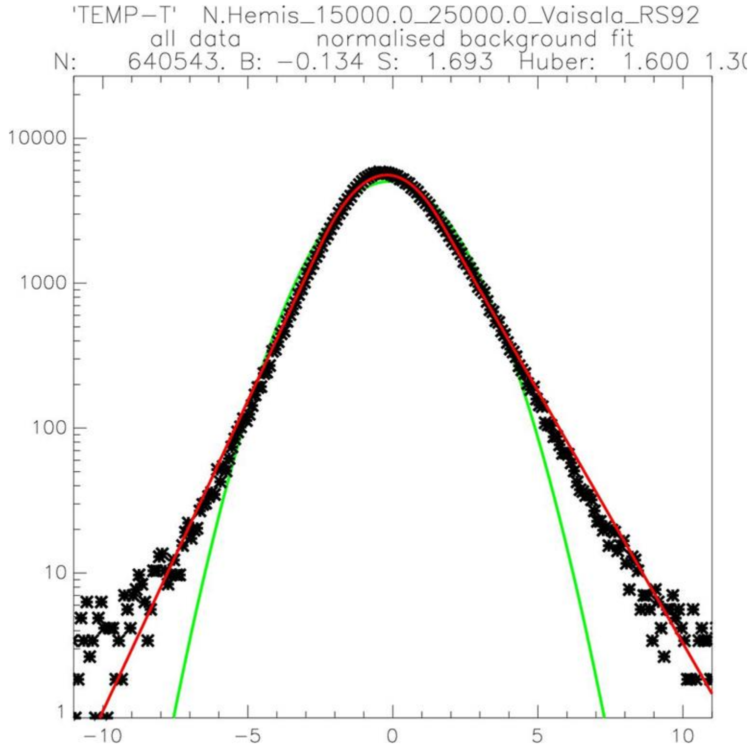
Additional Slides

Non-Gaussianity: Observations

- Dealing with non-Gaussianity: **Robust estimation** (Huber, 1981)
  - The presence of heavy tails in the O-B statistics indicates that outliers are significantly more probable than implied by a Gaussian error distribution
  - Empirically, it is found that for many observations a Huber-type metric provides a better fit to observed departures:

$$p(y|x) = \begin{cases} \dfrac{1}{\sigma_o\sqrt{2\pi}} \exp\left(\dfrac{a^2}{2} - |a\delta|\right) & \text{if } a < \delta \\ \dfrac{1}{\sigma_o\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\delta^2\right] & a \le \delta \le b \\ \dfrac{1}{\sigma_o\sqrt{2\pi}} \exp\left(\dfrac{b^2}{2} - |b\delta|\right) & \text{if } \delta > b \end{cases}$$

where $\delta = \mathbf{y} - G(\mathbf{x}_0)$



Best Gaussian fit
Best Huber norm fit

Tavolato and Isaksen, 2015

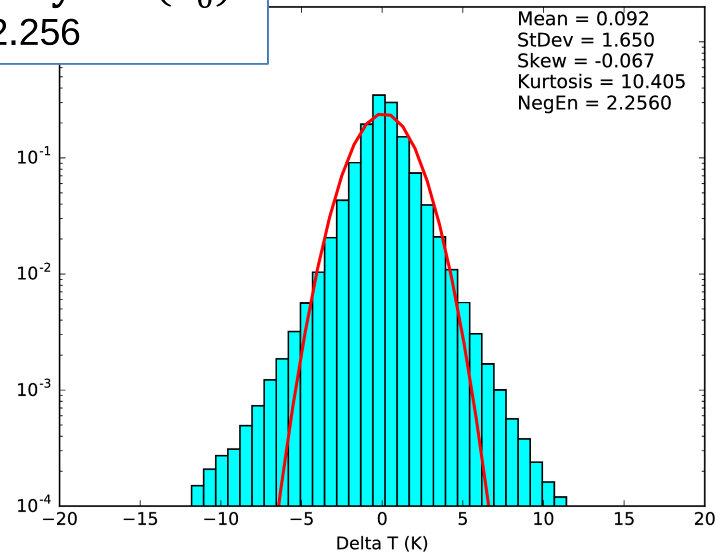
Non-Gaussianity: Observations

- The cost-function modified by the Huber norm is still convex, thus it does not cause problems to a gradient-based minimization
- In practice, the Huber-based quality control works by adjusting the expected observation error variance proportionally to the guess departures during the re-linearization (Bonavita et al, 2017):

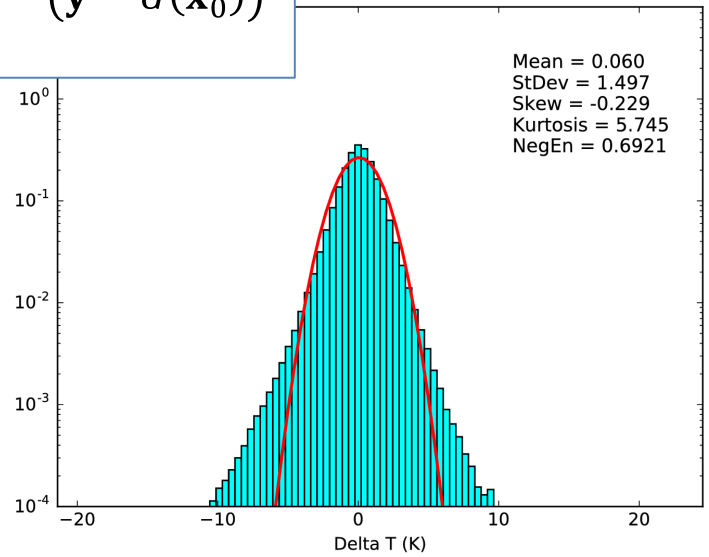
$$\sigma_o^2 \rightarrow \sigma_o^2\left(\frac{|y-G(x_0)|}{c}\right),$$

- In this way, it implicitly achieves a form of Gaussian anamorphosis:

Radiosonde T: $\mathbf{y} - G(\mathbf{x}_0)$
Negentropy=2.256



Radiosonde T: $\mathbf{R}^{1/2}(\mathbf{y} - G(\mathbf{x}_0))$
Negentropy=0.692



Non-Gaussianity: Control vector

- The change of variable of the Hólm transform $\widetilde{\delta RH} = \delta RH/\sigma\left(RH^b + {}^1\!/_2\,\delta RH\right)$ has made the J_B cost function nonlinear!

$$J_B\big(f(\delta RH)\big) = \frac{1}{2}\big(f(\delta RH)\big)^{\mathrm{T}}\mathbf{P}_b^{-1}\big(f(\delta RH)\big), \qquad f(\delta RH) = \delta RH/\sigma\left(RH^b + {}^1\!/_2\,\delta RH\right)$$

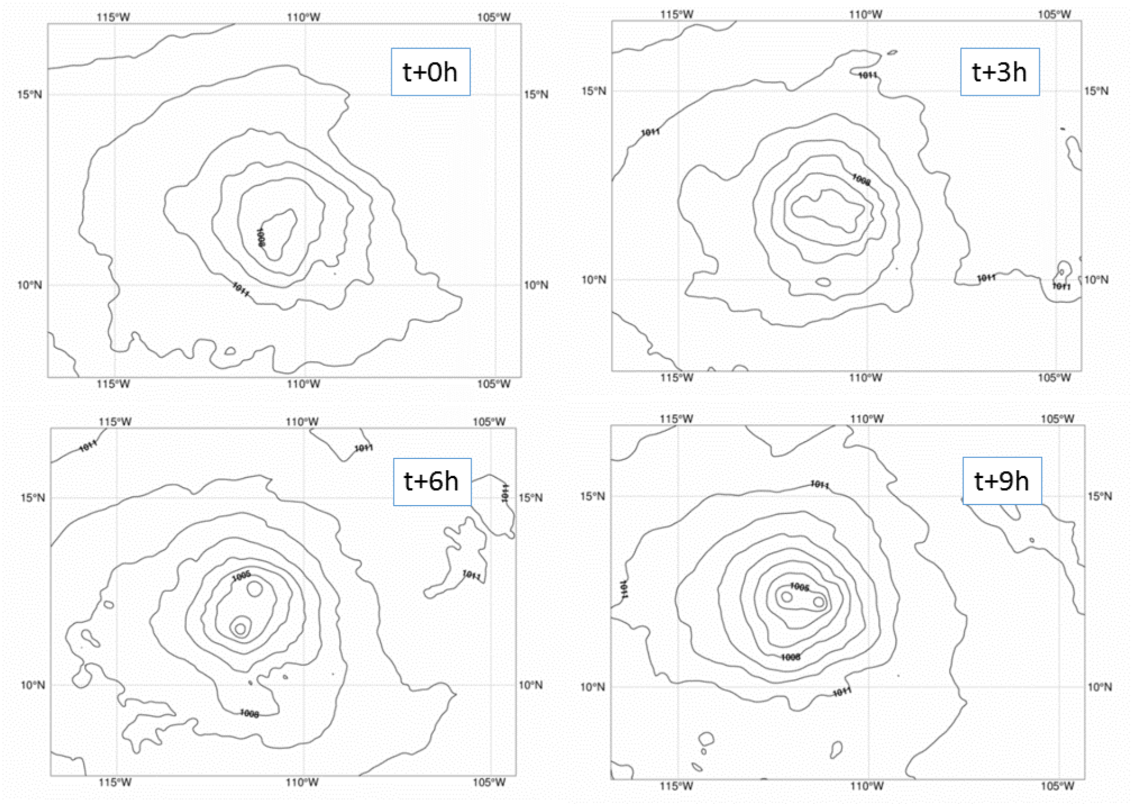
- The outer-inner loop mechanism comes to the rescue (again!):

  1. Inner loop: Minimise as a function of $\widetilde{\delta RH} = \delta RH/\sigma\left(RH^b\right)$
  2. Outer loop: solve for δRH the nonlinear equation $\widetilde{\delta RH} = \delta RH/\sigma\left(RH^b + {}^1\!/_2\,\delta RH\right)$, add the increment and compute updated guess trajectory

Non-Gaussianity

- Sometimes the non-Gaussianity in the forecasts is pathological, i.e. it is a symptom of other problems in the assimilation

    1. Initialisation problems in some of the EDA members (and the HRES too):

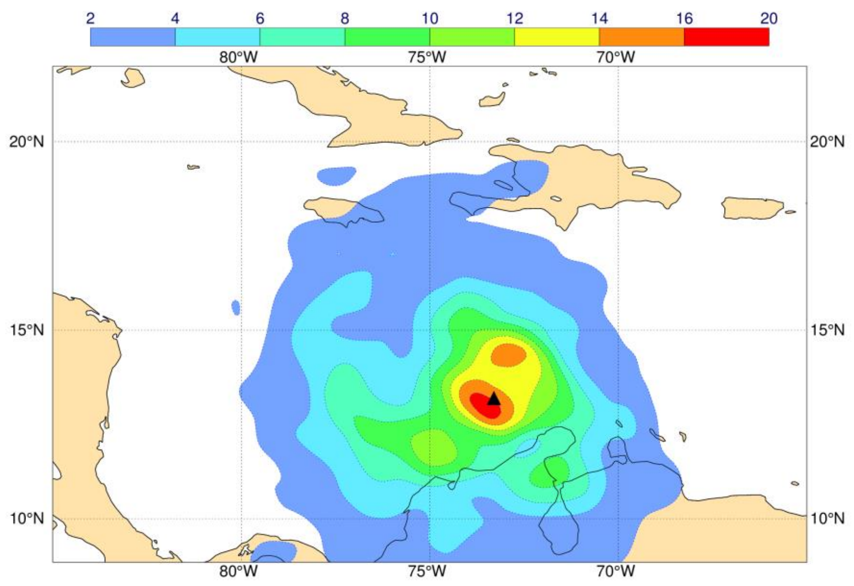


Tropical Storm Eugene: Evolution of the MSLP background forecast for the operational IFS started on 2017-07-07 at 18UTC

Non-Gaussianity

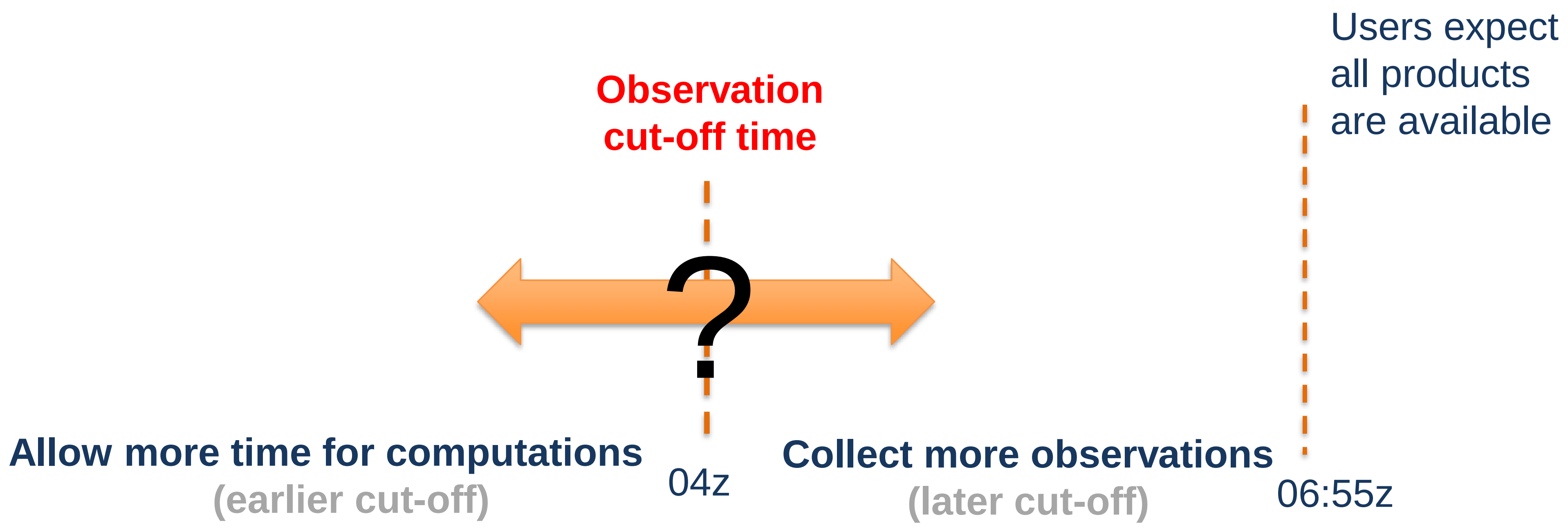
- Sometimes the non-Gaussianity in the forecasts is pathological, i.e. it is a symptom of other problems in the assimilation

    1. Minimisation has not converged in all EDA members



Tropical Cyclone Matthew, 2016-10-01 21UTC
EDA MSLP background errors
Lack of convergence due to dropsonde O-B in the 20-80 m/s range!

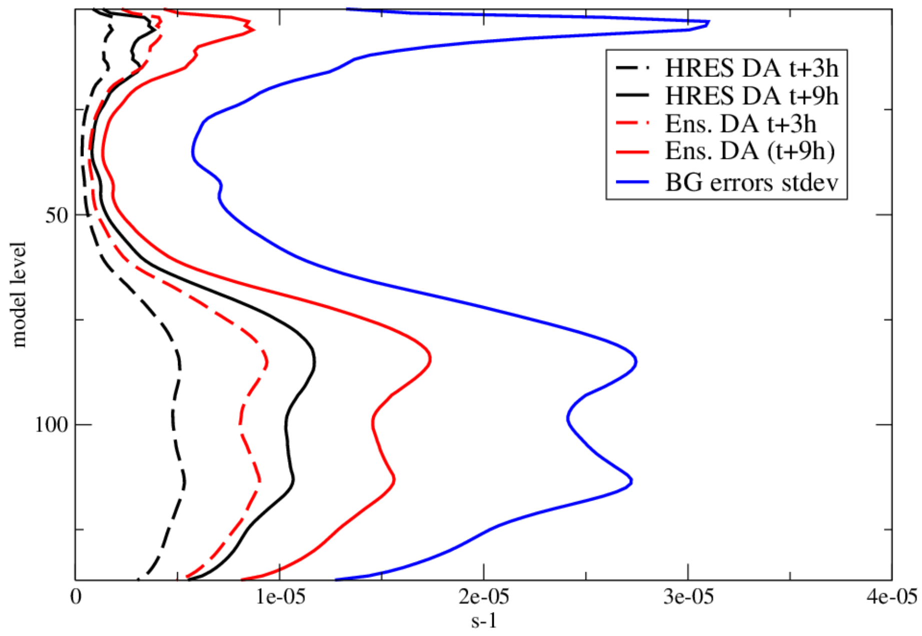
Current system: Trade-off



Model nonlinearities

$$StDev\left(M(\mathbf{x}^t + \delta\mathbf{x}_0) \; - \left(M(\mathbf{x}^t) + \mathbf{M}(\delta\mathbf{x}_0)\right)\right)$$

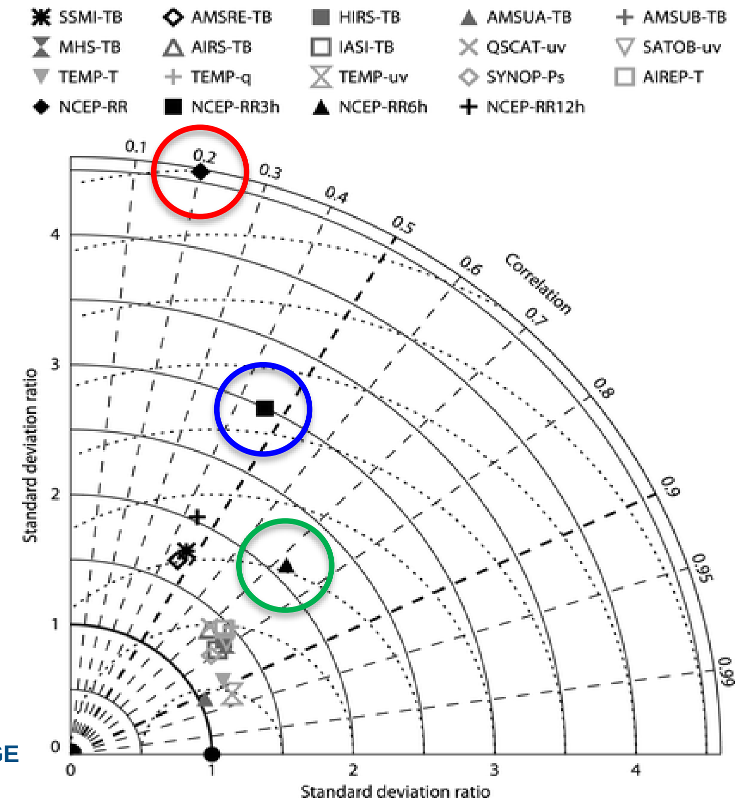
Vorticity



- Predominant in the Troposphere
- 10 to 50% of size of background errors
- Rapid increase of nonlinearity with length of assimilation window

Non-Gaussianity: Observations

- Another example is the explicit change of variable applied in the assimilation of radar and gauge **precipitation** at ECMWF (Lopez, 2011).
- This change of variable has two stages: 1) From 1-hourly to 6-hourly accumulated precipitation (**RR_{1h}** -> **RR_{6h}**) => improves linearity of DA problem



NCEP-RR1h
NCEP-RR3h
NCEP-RR6h

Taylor diagram of linearised vs nonlinear guess departures after 1st minimization for various observing systems. (Lopez, 2011)

Non-Gaussianity: Observations

- This change of variable has two stages, 1) From hourly to 6-hourly accumulated precipitation (**RR_{1h} -> RR_{6h}**): this improves the linearity of the DA problem
- The second stage changes from $RR_{6h} \rightarrow \ln(RR_{6h}+1)$: this improves the Gaussianity of the problem