Diagnosing Non-Gaussianity of Forecast and Analysis errors in a Convective Scale Model.

Météo-France, CNRM/GMAP

Raphaël Legrand, **Yann Michel** and **Thibaut Montmerle**

03/06/15
10^{th} Adjoint Workshop
Roanoke, West Virginia

Outlines

The Gaussian hypothesis

Bayesian formulation of the analysis process yields

$$\boxed{\underbrace{\mathcal{P}_a(\mathbf{x}|\mathbf{y})}_{\text{Analysis error PDF}} \propto \underbrace{\mathcal{P}_o(\mathbf{y}|\mathbf{x})}_{\text{Obs. error PDF}} \times \underbrace{\mathcal{P}_b(\mathbf{x})}_{\text{Background error PDF}}}$$

Background and observation errors are usually modeled with Gaussian distributions as: $\mathcal{P}_o(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, and $\mathcal{P}_b(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$.

Nonlinear dynamics yield non-Gaussian PDF of error (Bocquet et. al. 2010)

Aim of the study:

Diagnosing deviation from Gaussianity in forecast and analysis errors.

Methodology:

Run normality tests to diagnose Non-Gaussianity (NG) from distributions of perturbations sampled from an ensemble of assimilation.

Diagnostic of Non-Gaussianity (NG)

Deviation from Gaussianity is measured using K^2-statistics of the D'Agostino test (D'Agostino, 1970).

$$\boxed{K^2 = \underbrace{\left(\widehat{skewness}\right)^2}_{asymmetry} + \underbrace{\left(\widehat{kurtosis}\right)^2}_{peakedness}}$$

$\widehat{skewness} \sim \mathcal{N}(0,1)$, transformation of the 3^{rd} central moment.

$\widehat{kurtosis} \sim \mathcal{N}(0,1)$, transformation of the 4^{th} central moment.

♣ $K^2 \sim \chi^2(2) \rightarrow$for hypothesis testing of H_0:"the distribution is Gaussian",
H_0 is rejected at 95% confidence level, when $K^2 > 5.991$.

Diagnostic

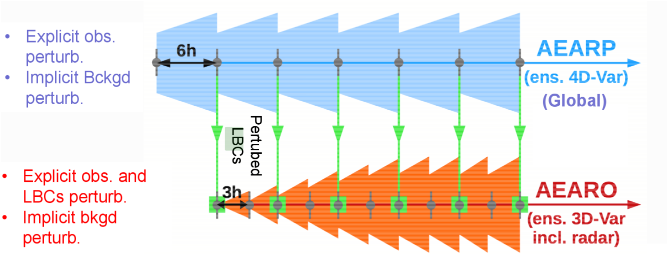
- discrimination according to the PDF's shape: asymmetry, peakedness
- cheap and parallelizable univariate test.
- this test could be use for sample sizes >30.

METEO FRANCE

Ensemble Data Assimilation (EDA)

Background error PDF is sampled using a Monte-Carlo approach with N perturbations $\delta\mathbf{x}_i$ of an ensemble data assimilation:

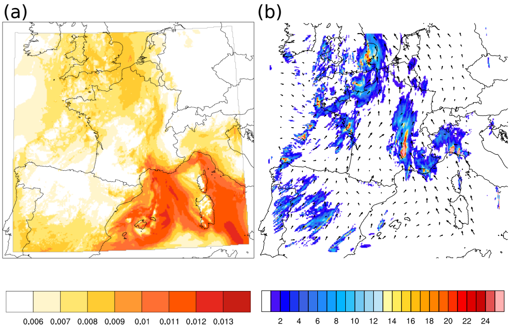
$$\delta\mathbf{x}_i = \mathbf{x}_i - \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \qquad \text{, for i=1..N}$$



Fisher 2003 ; Kucukkaraca and Fisher (2006); Berre et al 2006

<u>Dataset:</u> a 90-members ensemble (described in Ménétrier et al. (2014)) of the convective scale model AROME-France.

AROME simulation of the 04/11/11



3h-forecast of (a) specific humidity (q, kg/kg) at ≈ 920hPa and (b) surface precipitation (mm/h) for 1 member valid at 03UTC, the 04/11/11

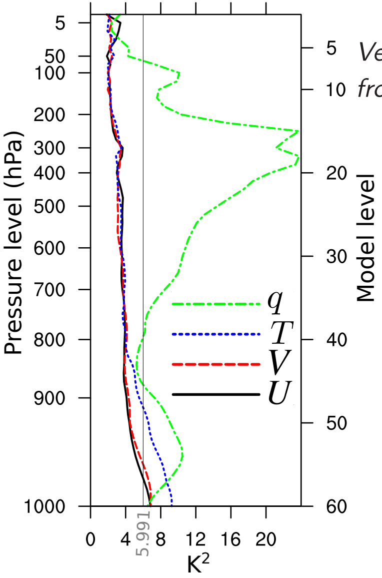
Meteorological situation of the 4th of November 2011:

- strong southerly convergent flow triggering deep convection (HYMEX research program, Ducrocq et al. 2014)
- cold active front, North-West of France

Outlines

METEO FRANCE

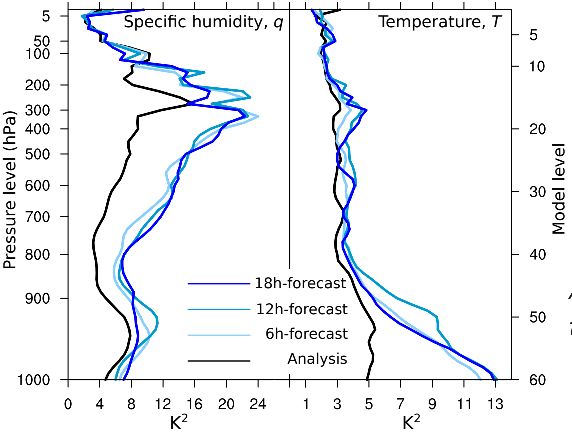
Overview of NG in background errors



Vertical profiles of averaged K^2 for 4 model var. from a 90-members of AROME 3h-forecasts

- largest NG for q, especially in boundary layer and the high troposphere.
- U, V, and T close to Gaussianity above 850hPa
- NG for U, V, and T in the boundary layer

Time evolution



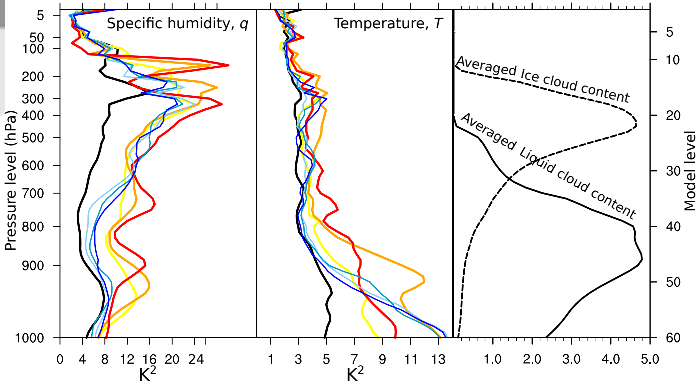
Averaged K^2 profiles: from the analysis to 18h-forecast.

- main increase of NG during the 6 first hours
- for q, large evolution in all free troposphere.
- For T, evolution in boundary layer.

Time evolution and Cloud processes

Cloud mask:

Cloudy mask: points where the vertically integ. cld content $> 0.5 g.kg^{-1}$ in the majority of the ensemble members

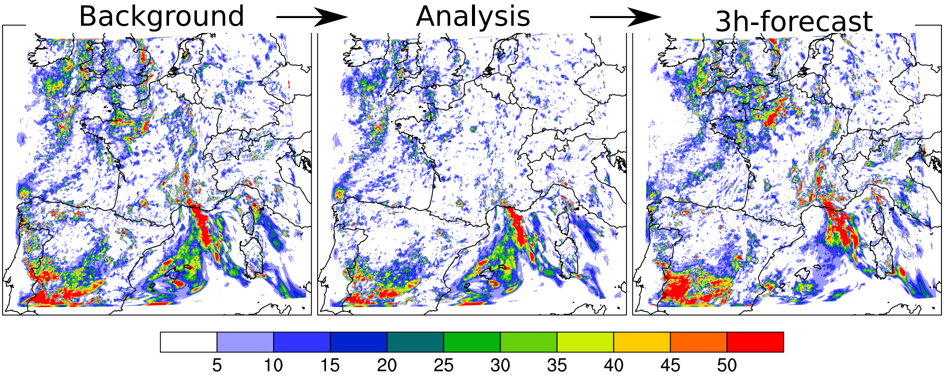


K^2 profiles averaged over "cloudy" points or "clear sky" points

- for q, NG in "cloudy" areas (displacement errors and diabatitic processes?)
- for T, NG in boundary layer (turbulent and radiative processes?)

METEO FRANCE

Impact of data assimilation on NG



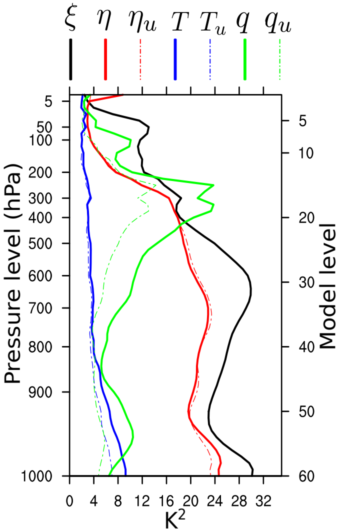
Maps of K^2 for q at level 52 ($\approx 920hPa$) during a cycle of assimilation.

- similarities between horizontal NG structures and meteorological features
- large decrease of NG during analysis step over well-observed areas
- recovery of NG after 3h of model integration

METEO FRANCE

NG in control space of the assimilation

Averaged profile of K^2 in 3h-forecasts for 4 control variables:



Multivariate formulation (Berre 2000)

$$\underbrace{\begin{pmatrix} \xi \\ \eta_u \\ T_u \\ q_u \end{pmatrix}}_{control} = \mathbf{B}_u^{-\frac{1}{2}} \mathbf{K}^{-1} \underbrace{\begin{pmatrix} \xi \\ \eta \\ T \\ q \end{pmatrix}}_{model}$$

$\mathbf{K}^{-1}$: inv. of balance operator

$\mathbf{B}_u^{-\frac{1}{2}}$: spatial transform

- ξ and η_u have strong NG over whole troposphere
- T_u and q_u are closer to Gaussianity than their balanced counterparts T and q.

Conclusion

Aim of the study:

Diagnosing deviation from Gaussianity in forecast and analysis errors for the convective scale model AROME in an Ensemble Data Assimilation framework.

- use of D'Agostino test (K^2) based on PDF's shape
- background error PDF sampled with a 90-members EDA

Main results

Forecast errors:

- q has the largest NG. For T, U, and V, NG only in boundary layer.
- main increase of NG during the 6 first hours
- cloud processes and surface processes are expected to enlarge NG.

Analysis errors:

- 3D-Var assimilation reduce NG in well-observed areas
- mass control variables ξ, and $\eta_u \rightarrow$ largest NG within control variables.
- T_u and q_u are more Gaussian than T and q.

Perspectives

Questions and Future work

- our findings may have implication for the choice of the control variables: choice of more Gaussian alternative dynamical variables.
- since displacement errors yield NG (Lawson and Hansen, 2005), diagnostics of NG may be used to evaluate improvements brought by the correction of displacement errors (Ravela, 2007).

Publication

Legrand, Michel and Montmerle: Diagnosing Non-Gaussianity of Forecast and Analysis errors in a Convective Scale Model $\Rightarrow$ submitted to NPG

METEO FRANCE

End

Legrand, Michel, and Montmerle

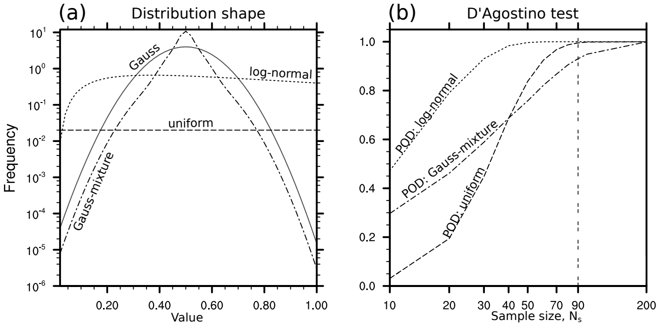
References

- Berre, 2000: Estimation of Synoptic and Mesoscale Forecast Error Covariances in a Limited-Area Model, M. W. R., 128, 644–667
- D'Agostino, 1970: Transformation to Normality of the Null Distribution of G1, Biometrika,679–681
- Ducrocq et al., 2014: "HyMeX-SOP1, the field campaign dedicated to heavy precipitation and flash flooding in the northwestern Mediterranean", B. A. M. S., 138, 7, 1083–1100, http://journals.ametsoc.org/doi/full/10.1175/BAMS-D-12-00244.1
- Bocquet, Pires, and Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation, M. W. R., 138, 8, 2997–3023
- Lawson and Hansen, 2005: Alignment error models and ensemble-based data assimilation, M. W. R., 133, 1687-1709
- Ménétrier, Montmerle, Berre, and Michel, 2014: Estimation and diagnosis of heterogeneous flow-dependent background-error covariances at the convective scale using either large or small ensembles, Q. J. R. M. S., 140, 683, 2050–2061
- Ravela, 2007: Data assimilation by field alignment, Physica D: Nonlinear Phenomena, 230, 127-145

Evaluation of D'Agostino test

Evaluation of D'Agostino test

Probability Of Detection (POD) is the probability that a test accurately rejects the tested hypothesis H_0 (e.g. "the PDF is Gaussian").



When testing different shapes of non-Gaussian distribution (a), values of POD with different sample sizes (b).

METEO FRANCE