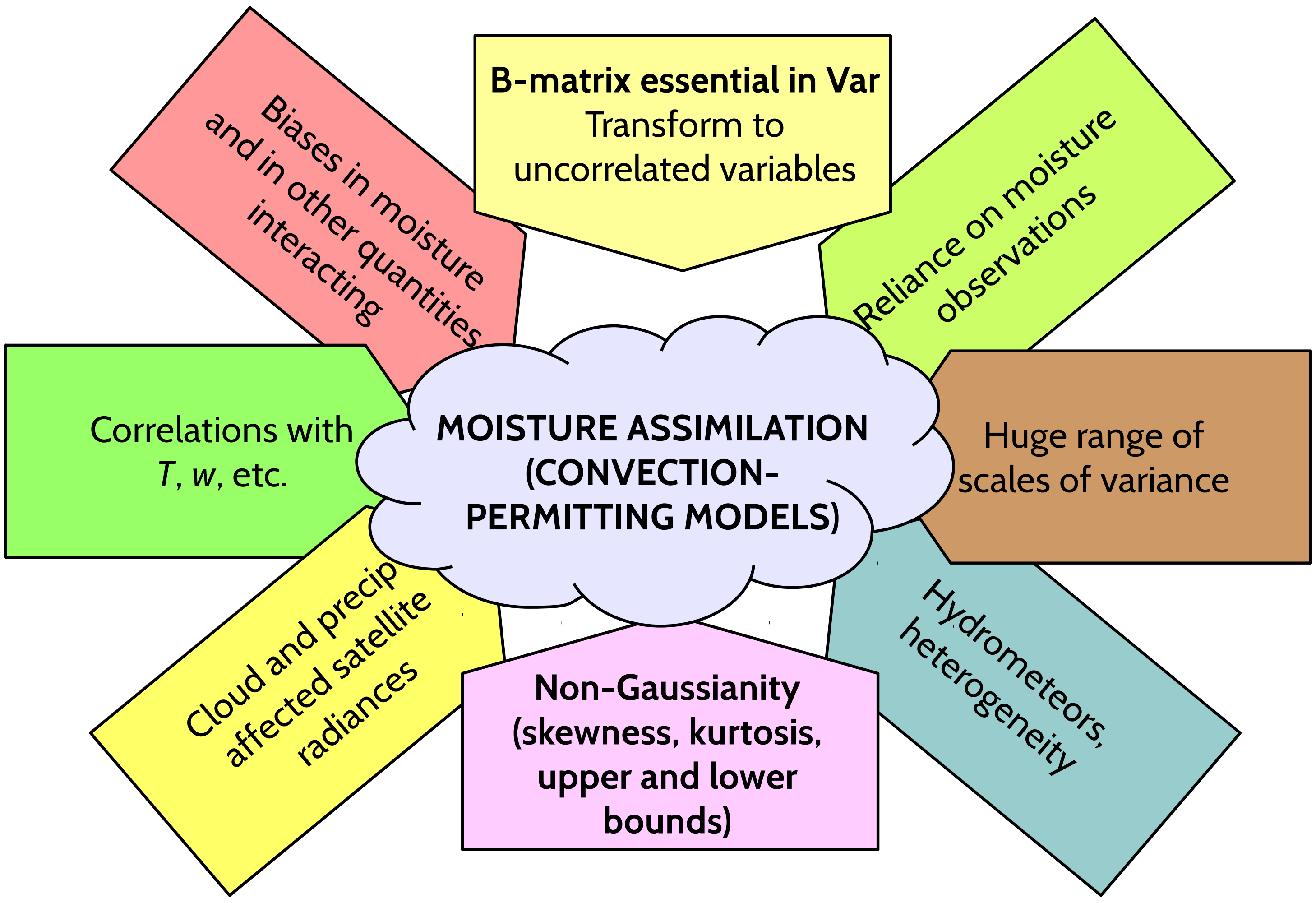
Non-Gaussianity of Moisture Errors in Data Assimilation



Ross Bannister*, David Livings*, Marek Wlasak†,
Oscar Martinez-Alvarado*
* University of Reading, UK
† Met Office, Exeter, UK

Project FRANC

Abstract

Assimilation of humidity information (e.g. specific humidity, q) has raised some interesting and challenging issues (see above). One such issue is its non-Gaussianity, which is an inevitable property given the bounded nature of specific humidity ($0 \leq q \lesssim q_{\rm sat}$). There have been various attempts in the last decade to account for this non-Gaussianity in the background errors of q in operational data assimilation systems. Here we compare the merits of two of them, namely (i) Gaussian anamorphosis, and (ii) the 'symmetrising' transform of Hólm et al. (2002) - used by Gustaffson et al. (2011) and later adapted by Ingleby et al. (2013). Emphasis is placed on assimilation in convective-scale models.

Keywords: Background error covariances, Non-Gaussianity, Gaussian anamorphosis, Hólm transform

How to incorporate non-Gaussian stats in DA

1. The particle filter

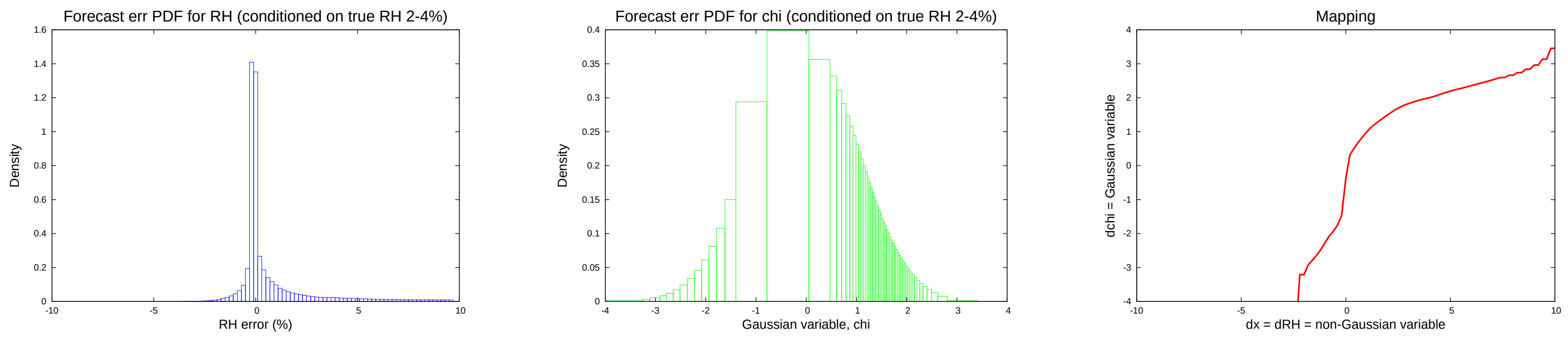
- Very general. Unsuitable for operational use (currently . . .).

2. Transform between 'non-Gaussian' and 'Gaussian' perturbations

$$\delta\chi = T(\delta x)$$

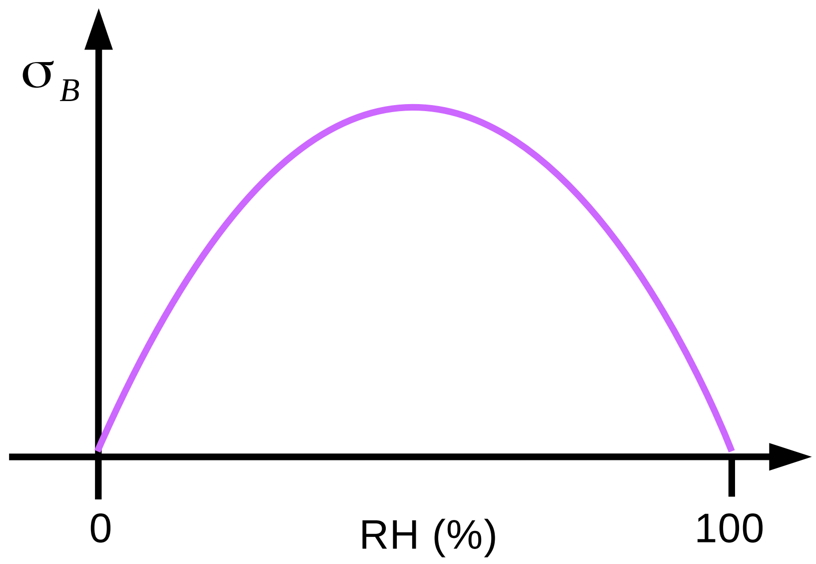
$\delta\chi$: has Gaussian bg errors ↑ ; δx : (model space) has non-Gaussian bg errs ↑

- Example: log-normal transformation.
- . . . or more generally: 'Gaussian anamorphosis'.



Method à la Hólm

- In DA, $\sigma_{\rm B}$ (the background error standard deviation) modulates how much we are allowed to modify $x_{\rm B}$.
- $\sigma_{\rm B}$ can be a function of RH.



- In Hólm's method the standard deviation, $\sigma_{\rm B}$, is a function of the RH value averaged between the background and analysis values:

$$\sigma_{\rm B}\left([x_{\rm B}+x_{\rm A}]/2\right).$$

Given

$$x_{\rm A} = x_{\rm B} + \delta x,$$

this leads to the implicit non-linear Hólm transform:

$$\delta\chi = \frac{\delta x}{\sigma_{\rm B}\left(x_{\rm B}+\delta x/2\right)}.$$

- The allowed increments reduce closer to the boundaries.

Basis of Hólm's transform

- If x_1 and x_2 are independent and drawn from the same distribution then $\delta x = x_1 - x_2$ has a symmetric distribution when conditioned on $\overline{x} = (x_1 + x_2)/2$.

Proof:

$$p(x_1, x_2) = p(\delta x, \overline{x}) \begin{vmatrix} \frac{\partial \delta x}{\partial x_1} & \frac{\partial \overline{x}}{\partial x_1} \\ \frac{\partial \delta x}{\partial x_2} & \frac{\partial \overline{x}}{\partial x_2} \end{vmatrix} = p(\delta x, \overline{x})$$

$$\therefore p(-\delta x, \overline{x}) = p(x_2, x_1) = p(x_1, x_2) = p(\delta x, \overline{x})$$

$$p(\delta x|\overline{x}) = \frac{p(\delta x, \overline{x})}{p(\overline{x})} \qquad \therefore p(\delta x|\overline{x}) = p(-\delta x|\overline{x})$$

- Consider the law of total probability:

$$p(\delta x) = \int_{-\infty}^{\infty} d\phi \, p(\delta x|\phi) p(\phi),$$

where $p(\delta x|\phi)$ is the probability conditioned on an arbitrary variable ϕ.

  - If $p(\delta x|\phi)$ is a Gaussian with the same mean and variance $\forall \phi$, then $p(\delta x)$ is also Gaussian.
  - If we could find a conditioning, $p(\delta x|\phi)$, such that this is true, then we may construct a Gaussian (or at least symmetric) variable.

- Strategy:

  - Find $p(\delta x|\phi)$ which is close to Gaussian.
  - Construct a new variable which has mean 0 and variance 1:

$$\delta \chi = \frac{\delta x - \overline{\delta x(\phi)}}{\sigma(\delta x|\phi)}.$$

  - Choose $\phi = \overline{x}$ then the conditional distribution has zero mean, $\overline{\delta x(\overline{x})} = 0$, and Hólm's transform follows.

Notes on Hólm's transform

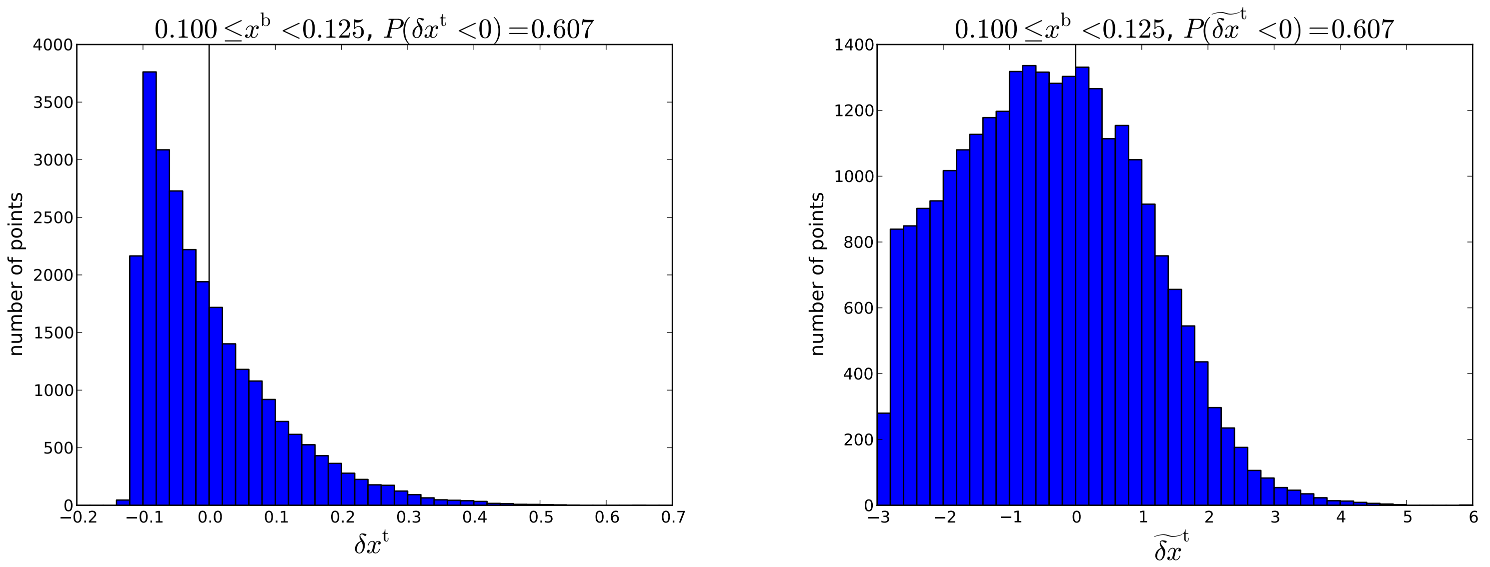
No transform of the form $\delta\chi = a\,\delta x$ where $a > 0$ can turn an arbitrary asymmetric distribution into a symmetric distribution.

Proof:

- A symmetric distribution has its median at zero: $p(\delta\chi < 0) = 1/2$.
- Suppose $p(\delta x)$ is asymmetric in such a way that $p(\delta x < 0) \neq 1/2$.
- The transform preserves the sign of δx ...
- ... so $p(\delta\chi < 0) = p(\delta x < 0) \neq 1/2$.
- The distribution of $\delta\chi$ cannot therefore be symmetric.

Test with some synthetic data

$p(\delta x)$ $p(\delta\chi)$



We do however expect Hólm's transform to have some useful properties for data assimilation.

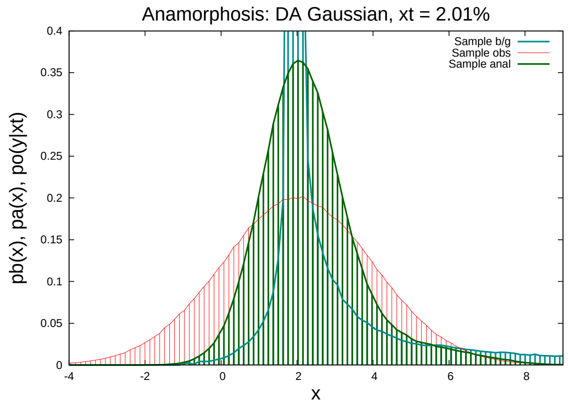
Monte-Carlo data assimilation experiments

- Background error PDFs are computed from 35 pairs of RH forecasts from the MetO UKV model.
  - NMC method.
  - '$T+6$' minus '$T+3$' forecast error proxy.
  - Dry (RH 2%), medium (RH 61%) and moist (RH 99%) scenarios.
- Run an ensemble of scalar data assimilations (10^6 samples).
  - Obs sampled from a Gaussian with $\sigma_{\mathrm{O}} = 2\%$ (allowed to go 'out of bounds').
  - Bgs sampled from the relevant non-Gaussian (allowed to go 'out of bounds').
- Assimilation performed with anamorphosis:
  - Controls: Gaussian DA with σ_{B} found from non-Gaussian distribution.
  - Tests: Non-Gaussian DA.
- Assimilation performed with Hólm:
  - Controls: Gaussian DA with σ_{B} found from non-Gaussian distribution.
  - Tests: Non-Gaussian DA with Hólm normalisation.

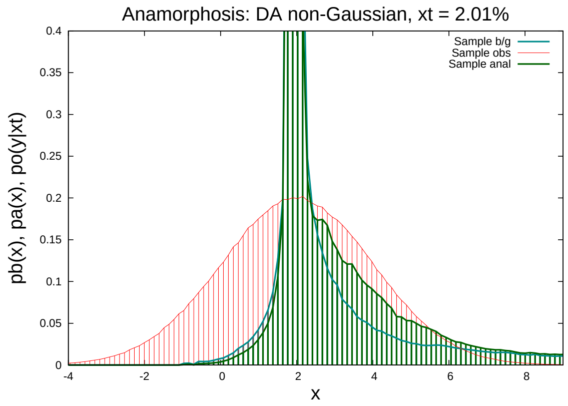
Anamorphosis results

	Gaussian DA	non-Gaussian DA

Dry

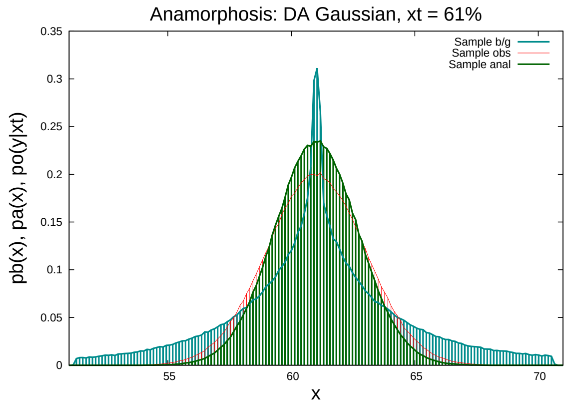


$$p_A(x < 0) = 0.019$$
$$\text{skew}(p_A) = 1.004$$

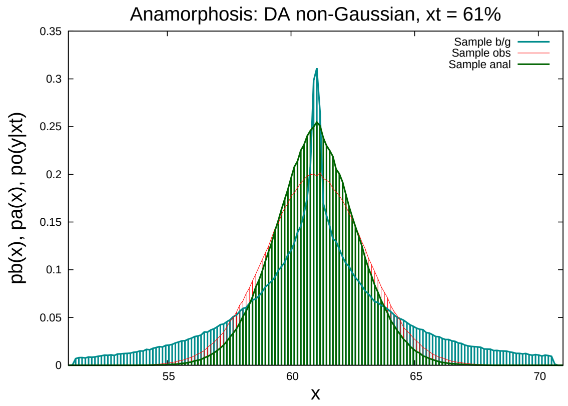


$$p_A(x < 0) = 0.002$$
$$\text{skew}(p_A) = 2.044$$

Medium

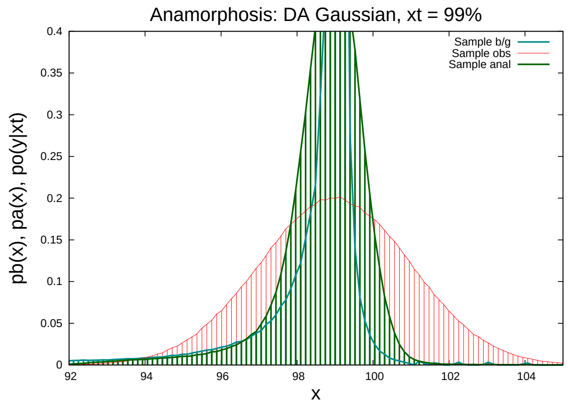


$$\text{skew}(p_A) = 0.003$$

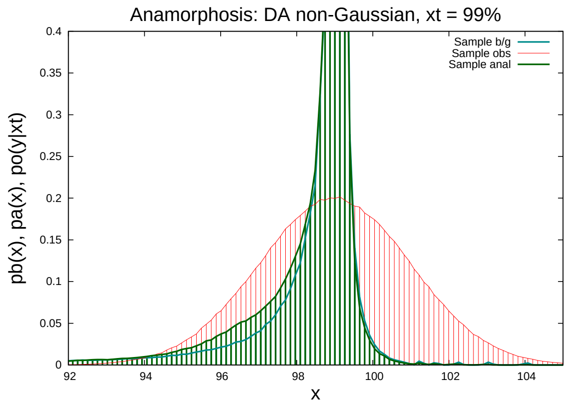


$$\text{skew}(p_A) = 0.008$$

Moist



$$p_A(x > 100) = 0.056$$
$$\text{skew}(p_A) = -1.739$$

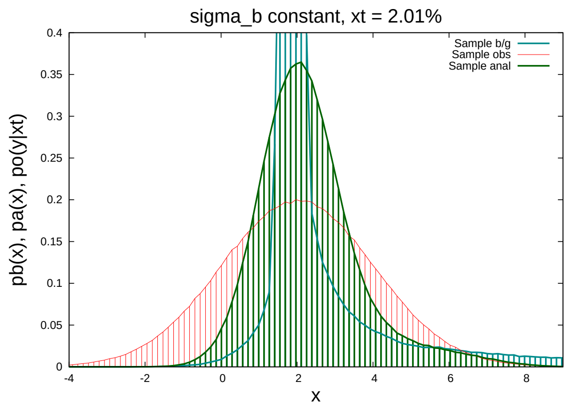


$$p_A(x > 100) = 0.008$$
$$\text{skew}(p_A) = -2.684$$

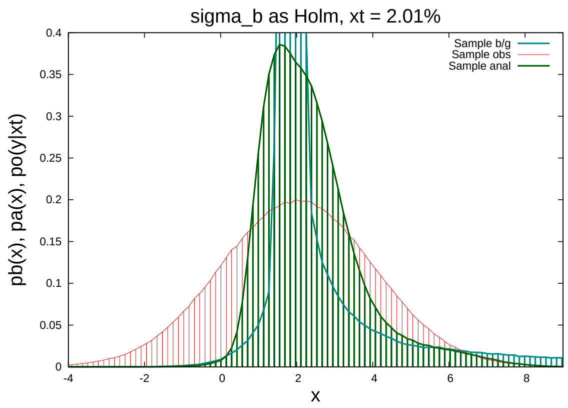
Hólm results

	Constant σ_{B}	σ_{B} according to Hólm

Dry

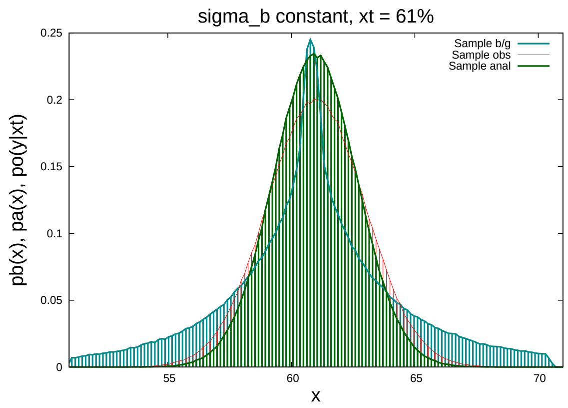


$$p_{\mathrm{A}}(x < 0) = 0.022$$
$$\mathrm{skew}(p_{\mathrm{A}}) = 1.030$$

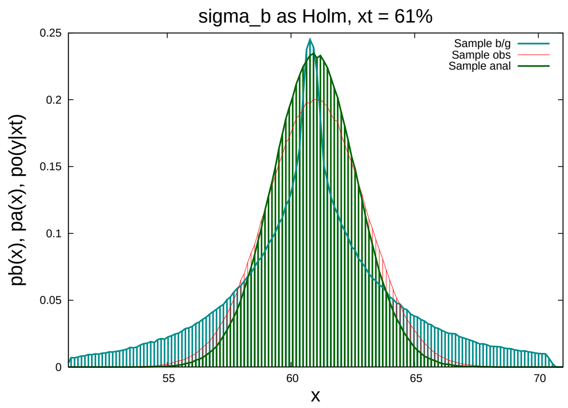


$$p_{\mathrm{A}}(x < 0) = 0.004$$
$$\mathrm{skew}(p_{\mathrm{A}}) = 1.306$$

Medium

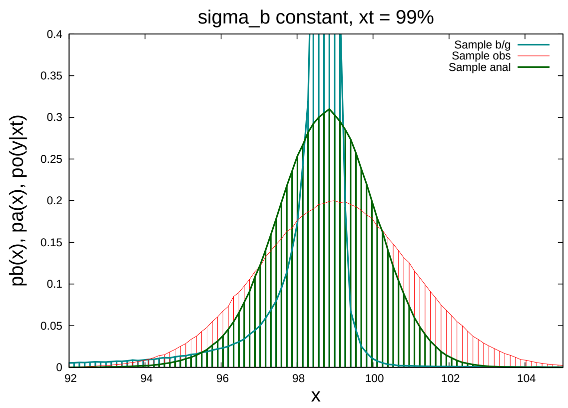


$$\mathrm{skew}(p_{\mathrm{A}}) = 0.006$$

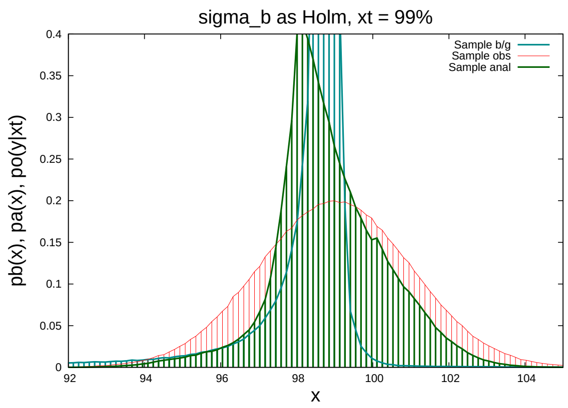


$$\mathrm{skew}(p_{\mathrm{A}}) = 0.006$$

Moist



$$p_{\mathrm{A}}(x > 100) = 0.156$$
$$\mathrm{skew}(p_{\mathrm{A}}) = -0.183$$



$$p_{\mathrm{A}}(x > 100) = 0.192$$
$$\mathrm{skew}(p_{\mathrm{A}}) = 0.096$$

Conclusions

- Non-Gaussianity of errors should be considered in many circumstances . . .
  - Avoids 'out-of-bounds' in DA.
- . . . but most operational DA schemes rely on Gaussianity.
- Non-Gaussianity can be accounted for using many methods:
  - Particle filters.
  - Transform methods:
    * Gaussian anamorphosis.
    * (Special example log-normal).
  - Non-constant normalisation:
    * As Hólm.
- We consider non-Gaussian background errors, but observations can also be non-Gaussian.
- Gaussian anamorphosis uses the correct non-Gaussian distribution.
- Hólm attempts to control analysis increments based on a variable σ_{B} that is a function of the average of the background and the analysis.
- Gaussian anamorphosis is more successful than Hólm in our experiments.

References

- Hólm E., Andersson E., Beljaars A., Lopez P., Mahfouf J.-F., Simmons A., Thépaut J.-N., Assimilation and modelling of the hydrological cycle: ECMWF's status and plans, ECMWF Tech. Memo. 383 (2002).
- Gustafsson N., Thorsteinsson S., Stengel M., Hólm E. Use of a nonlinear pseudo-relative humidity variable in a multivariate formulation of moisture analysis, Q.J.R.Meteor.Soc. 137, 1004-1018 (2011).
- Ingleby N.B., Lorenc A.C., Ngan K., Rawlins F., Jackson D.R., Improved variational analyses using a nonlinear humidity control variable, Q.J.R.Meteor.Soc. 139, 1875-1887 (2013).

Please also see poster by Hristo Chipilski (poster No. 9 on Wednesday)