National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

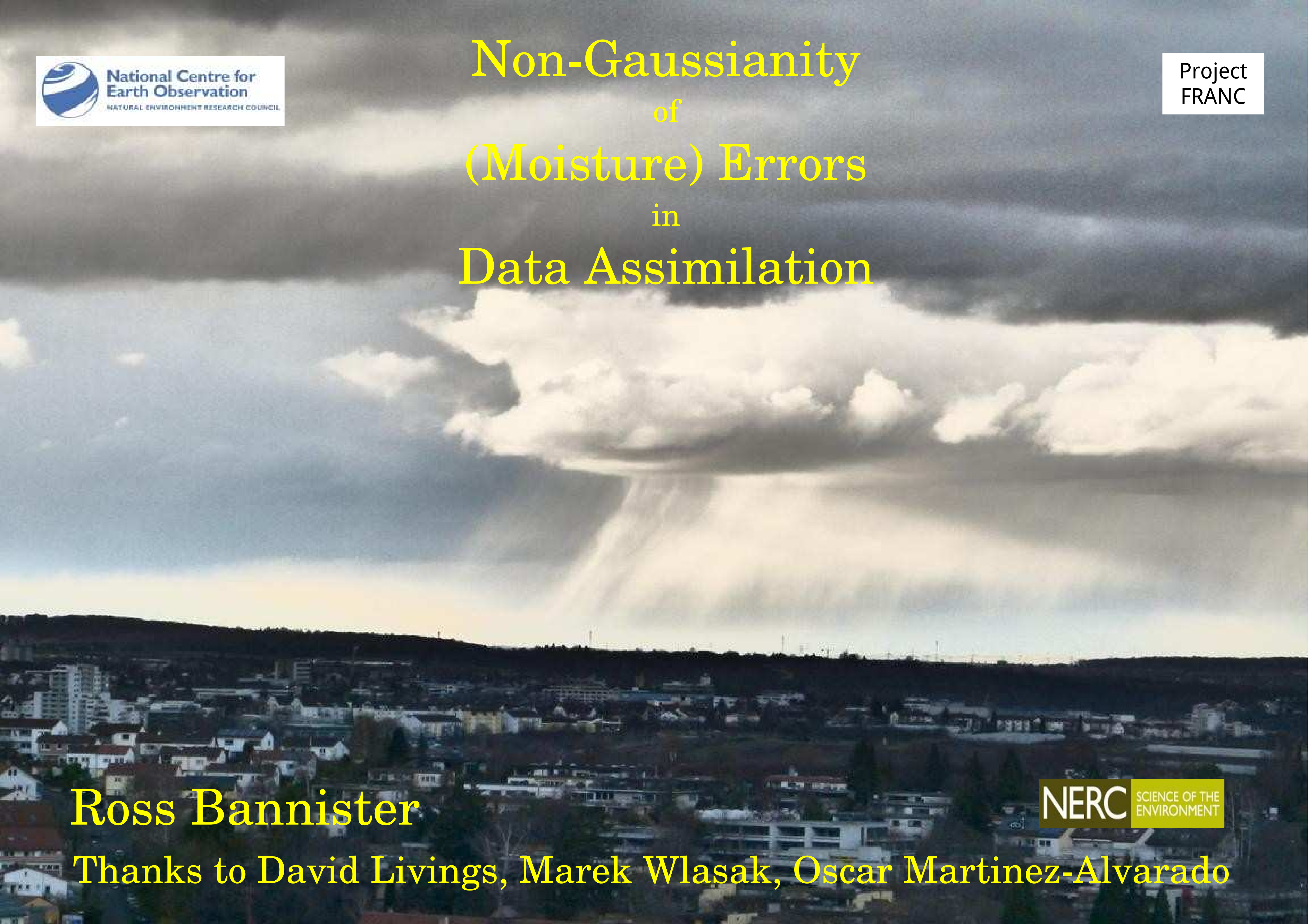
Project FRANC

Non-Gaussianity of (Moisture) Errors in Data Assimilation

Ross Bannister

Thanks to David Livings, Marek Wlasak, Oscar Martinez-Alvarado

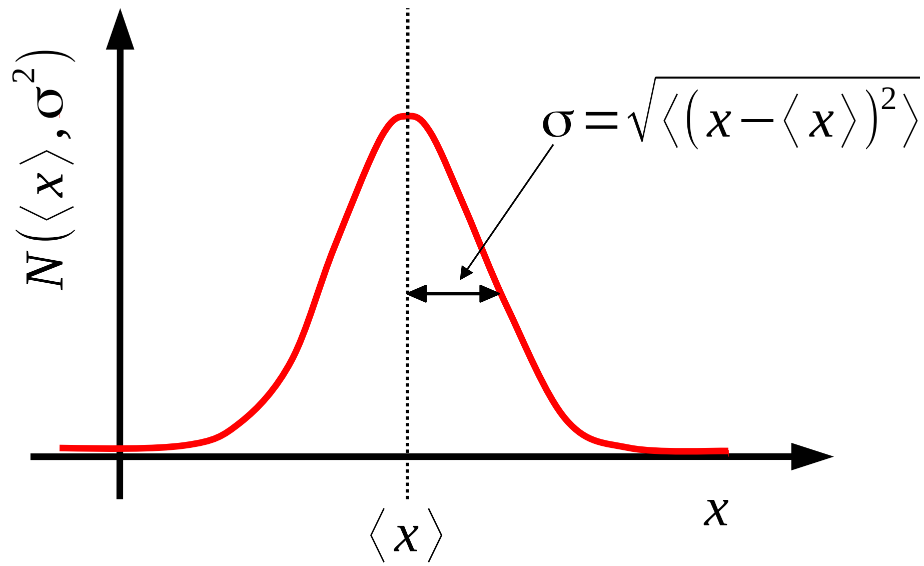
NERC SCIENCE OF THE ENVIRONMENT



Non-Gaussianity of (Moisture) Errors in DA

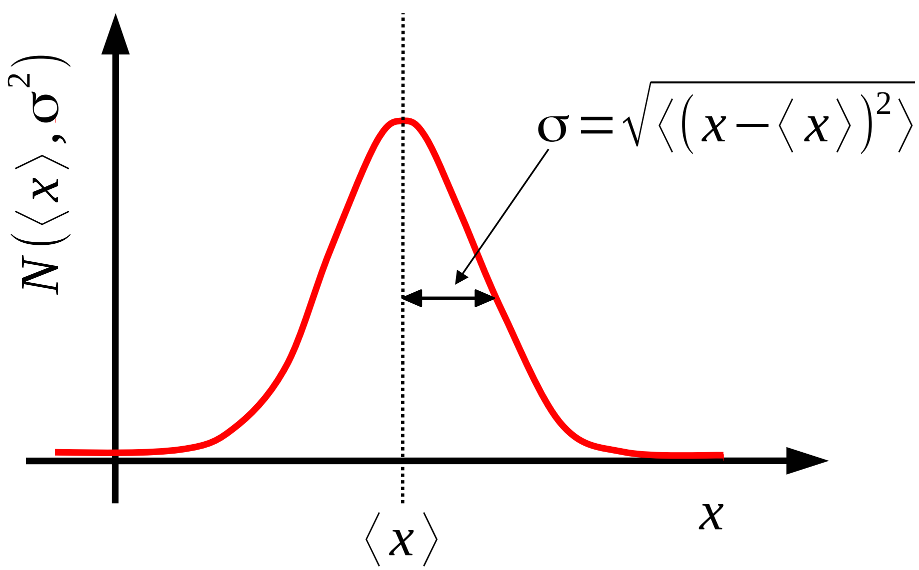
- PDF $\sim$ possible values of a quantity realisable.

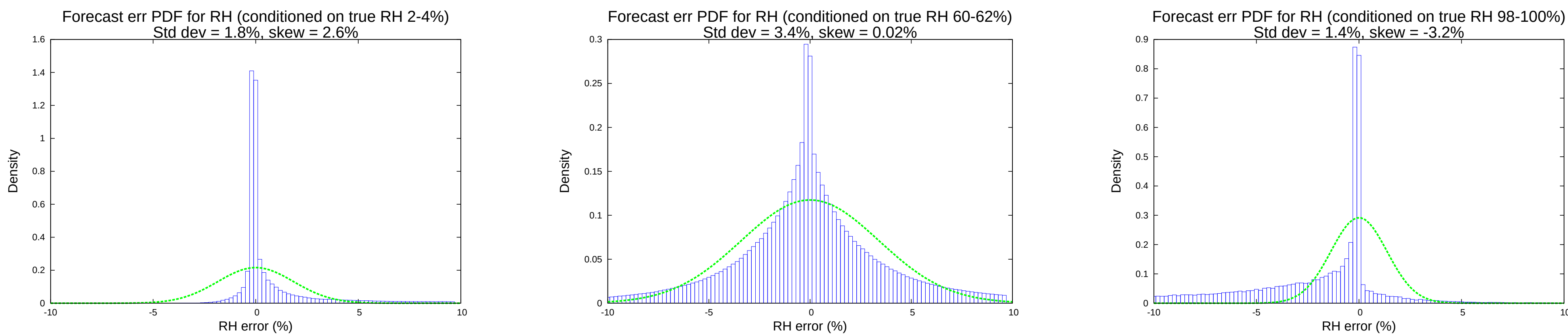
- Normal (Gaussian) distribution used widely.
  - Mathematically convenient.
  - Compact representation.
  - Many quantities are Gaussian.
  - Many are not!
  - Most data assimilation systems assume Gaussianity.



Non-Gaussianity of (Moisture) Errors in DA

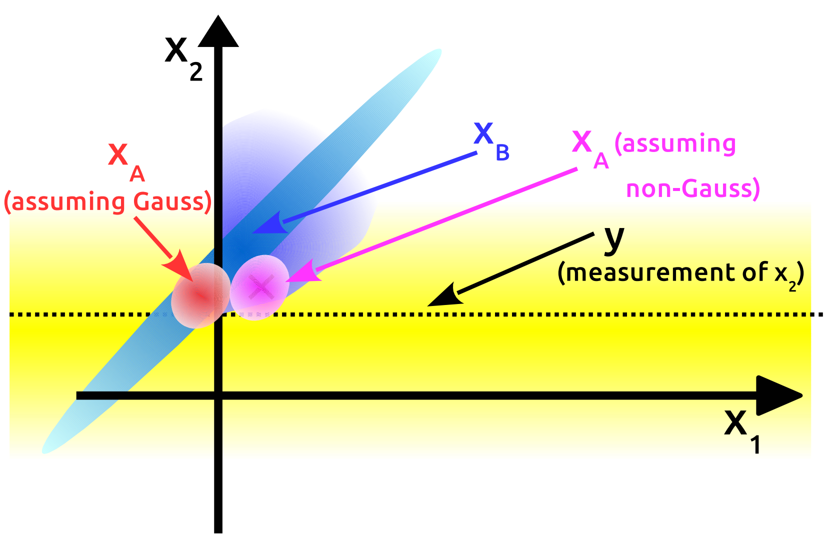
- PDF $\sim$ possible values of a quantity realisable.
- Normal (Gaussian) distribution used widely.
  - Mathematically convenient.
  - Compact representation.
  - Many quantities are Gaussian.
  - Many are not!
  - Most data assimilation systems assume Gaussianity.



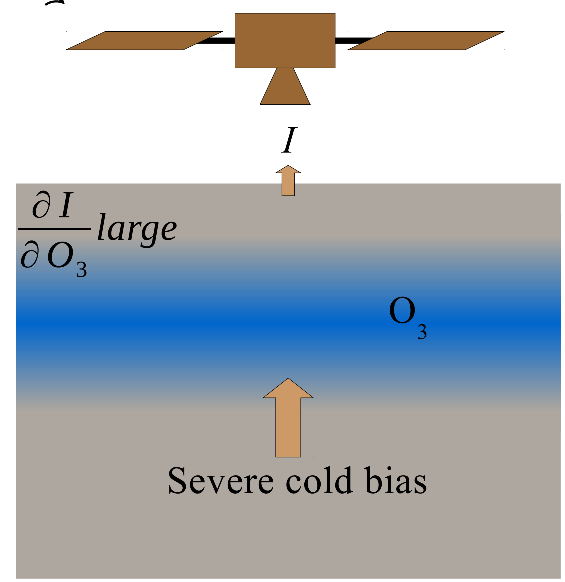


Potential shortcomings assuming Gaussian stats in DA (for the bg errs)

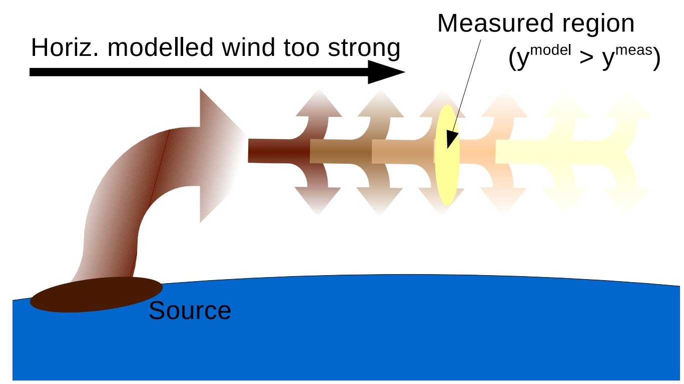
- Suboptimal analyses.
- Multi-modalities will be ignored.
- Can lead to unphysical values (e.g. neg concentrations). How can this happen?
  - For direct obs, the analysis is a linear combination of the b/g and the obs.
  - Can still get neg concentrations in unobserved quantities via correlations. ↘



- Conceivable when state is to be inferred. E.g. radiance assimilation ↘. . .



- . . . surface flux estimation ↙.



How to incorporate non-Gaussian stats in DA

1. The particle filter

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

2. Transform (T) between 'non-Gaussian' and 'Gaussian' perts

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

$$\begin{array}{rl} \uparrow & \quad\uparrow \\ \delta\chi: \text{ errors that have} & \delta x: \text{ (model space) errors that have} \\ \text{Gaussian bg errors} & \text{non-Gaussian bg errs} \end{array}$$

- Example: log-normal transformation.

How to incorporate non-Gaussian stats in DA

1. The particle filter

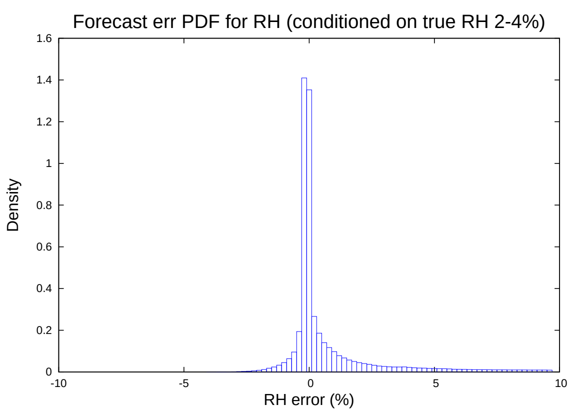
- Very general. Unsuitable for operational use (currently ...).

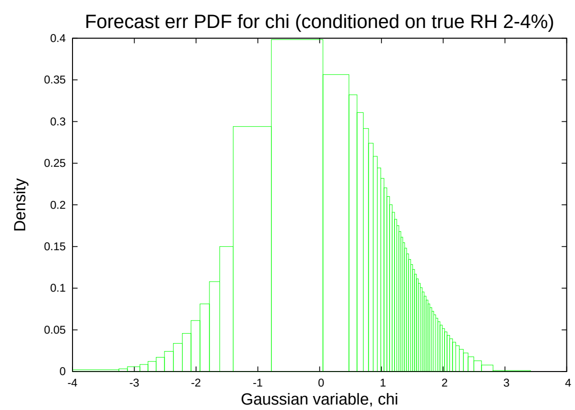
2. Transform (T) between 'non-Gaussian' and 'Gaussian' perts

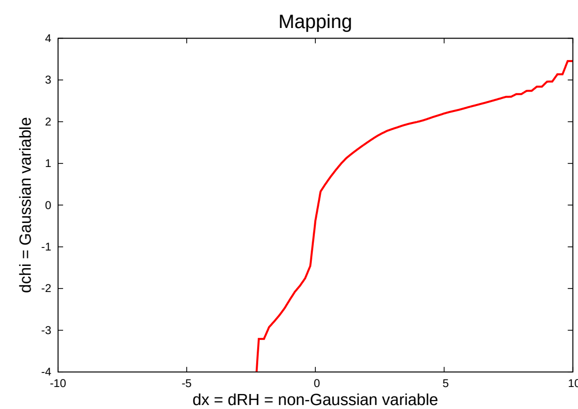
$$\delta\chi = T(\delta x)$$

$$\begin{array}{ll} \uparrow & \uparrow \\ \delta\chi: \text{ errors that have} & \delta x: \text{ (model space) errors that have} \\ \text{Gaussian bg errors} & \text{non-Gaussian bg errs} \end{array}$$

- Example: log-normal transformation.
- More generally: 'Gaussian anamorphosis'.



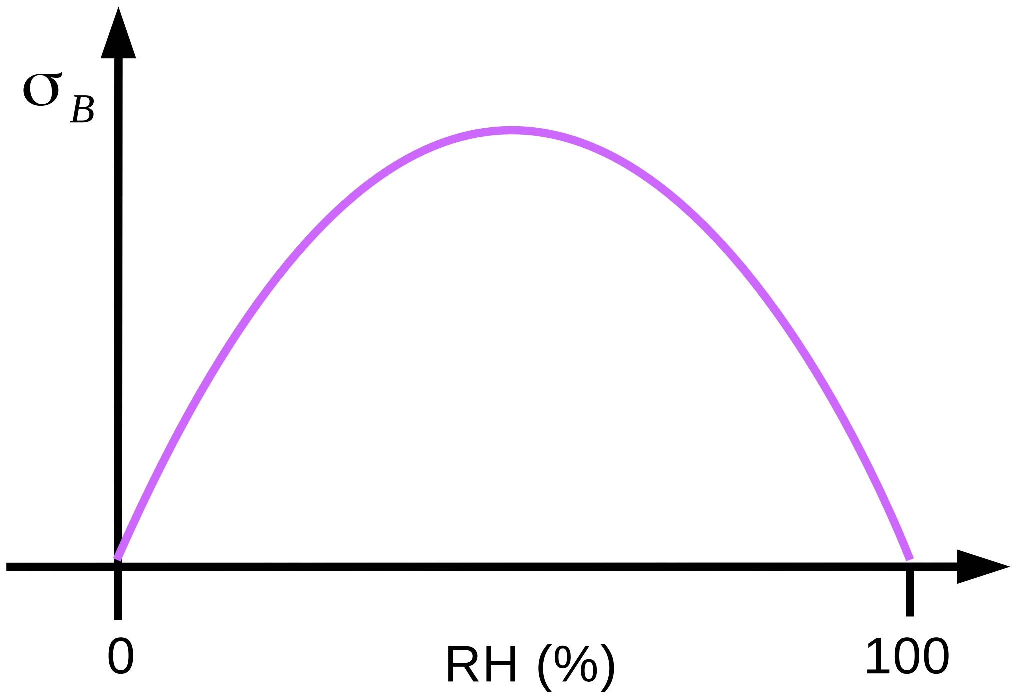




How to incorporate non-Gaussian stats in DA (cont.)

3. 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 the Hólm method the standard deviation is conditioned on the RH value averaged between the background and analysis values:

$$
\begin{aligned}
\delta x &= \sigma_{\rm B}([x_{\rm B} + x_{\rm A}]/2)\, \delta\chi, \\
\text{where } x_{\rm A} &= x_{\rm B} + \delta x,
\end{aligned}
$$

leading to the implicit non-linear Hólm transform:

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

- The allowed increments reduce closer to the boundaries.

Monte-Carlo experiments

- Background error PDFs are computed from 35 pairs of 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.

Monte-Carlo experiments

- Background error PDFs are computed from 35 pairs of 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 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').

Monte-Carlo experiments

- Background error PDFs are computed from 35 pairs of 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 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.

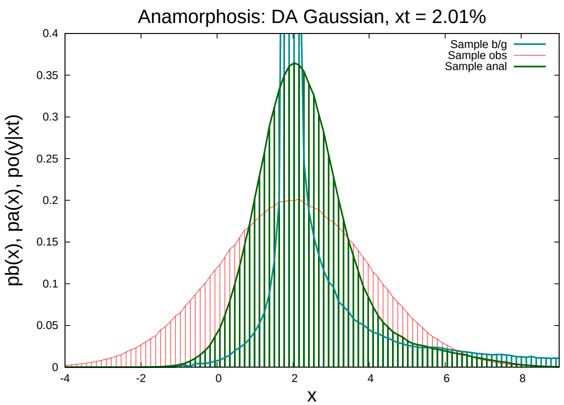
Monte-Carlo experiments

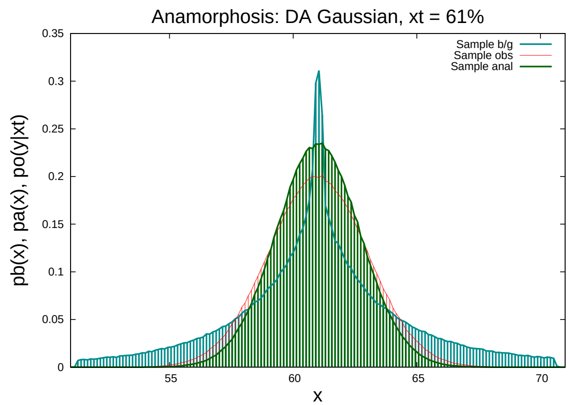
- Background error PDFs are computed from 35 pairs of forecasts from the MetO UKV model.
  - NMC method.
  - $T-6$ minus $T-3$ forecast error proxy.
  - Dry (RH 2%), medium (RH 61%) and moist (RH 99%) scenarios.
- Run an ensemble of data assimilations (10^6 samples).
  - Obs sampled from a Gaussian with $\sigma_{\mathrm{O}} = 2\%$ (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: 'Gaussian' DA with Hólm conditioning.

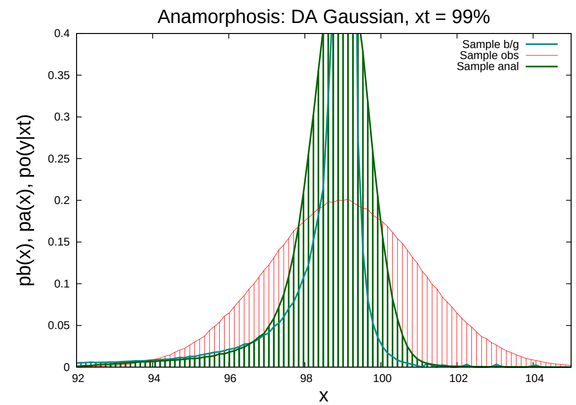
Anamorphosis results

Dry	Medium	Moist

Runs using Gaussian data assimilation





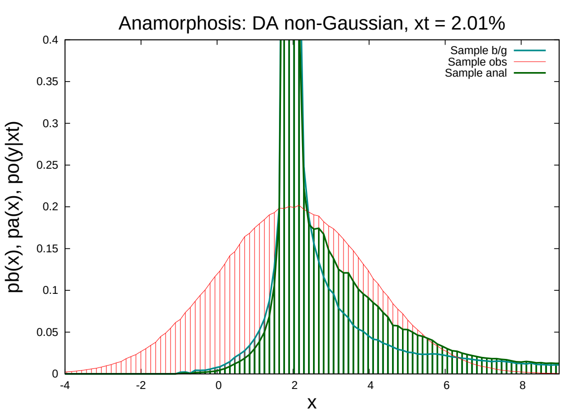


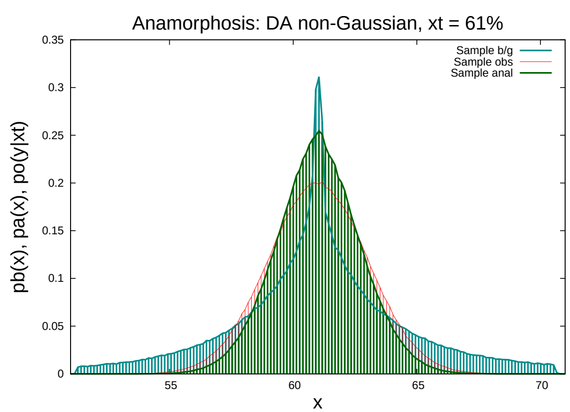
Dry:
$$p_{\mathrm{A}}(x < 0) = 0.019$$
$$\mathrm{skew}(p_{\mathrm{A}}) = 1.004$$

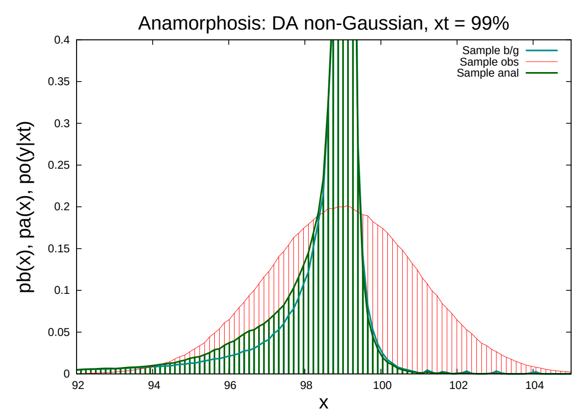
Medium:
$$\mathrm{skew}(p_{\mathrm{A}}) = 0.003$$

Moist:
$$p_{\mathrm{A}}(x > 100) = 0.056$$
$$\mathrm{skew}(p_{\mathrm{A}}) = -1.739$$

Runs using non-Gaussian data assimilation







Dry:
$$p_{\mathrm{A}}(x < 0) = 0.002$$
$$\mathrm{skew}(p_{\mathrm{A}}) = 2.044$$

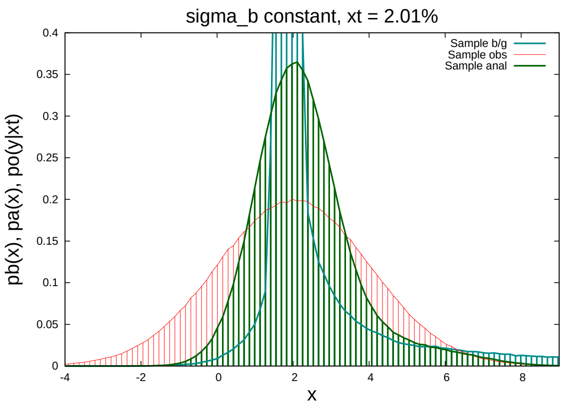
Medium:
$$\mathrm{skew}(p_{\mathrm{A}}) = 0.008$$

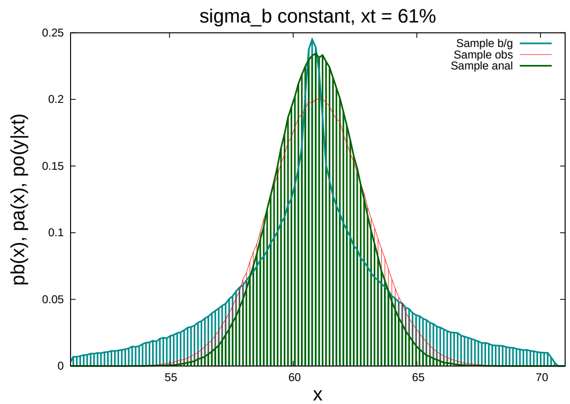
Moist:
$$p_{\mathrm{A}}(x > 100) = 0.008$$
$$\mathrm{skew}(p_{\mathrm{A}}) = -2.684$$

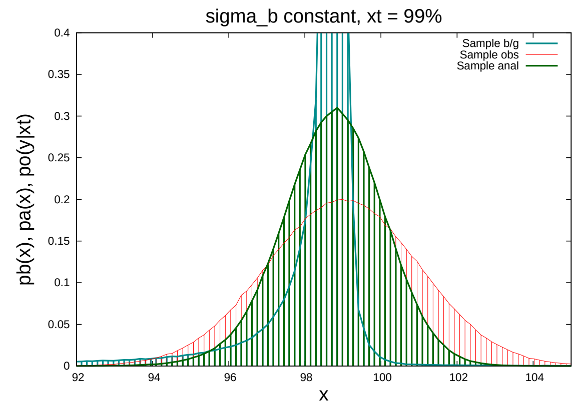
Hólm results

Dry	Medium	Moist

Runs using data assimilation with constant σ_{B}





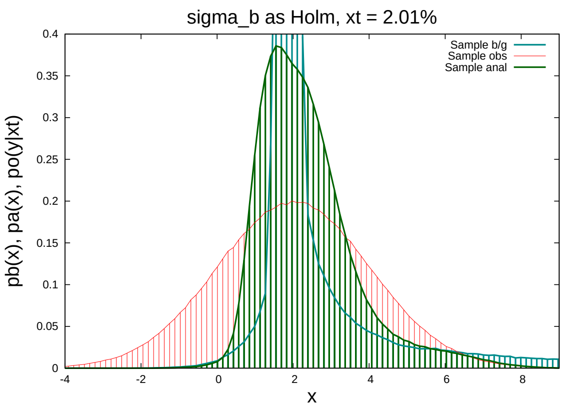


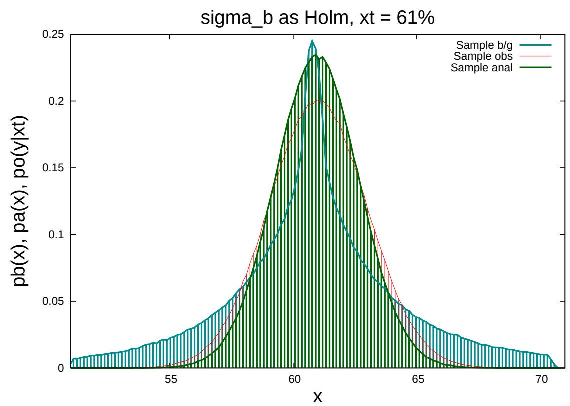
Dry: $p_{\mathrm{A}}(x < 0) = 0.022$, $\mathrm{skew}(p_{\mathrm{A}}) = 1.030$

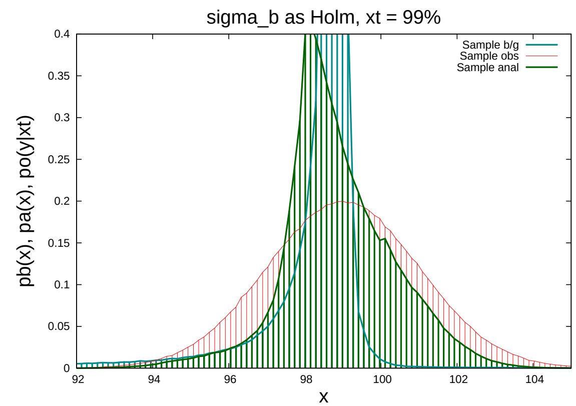
Medium: $\mathrm{skew}(p_{\mathrm{A}}) = 0.006$

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

Runs using data assimilation with σ_{B} according to Hólm







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

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

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

Conclusions

- Non-Gaussianity of errors should be considered in many real-world 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-linear conditioning:
    * 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} conditioned on the average of the background and the analysis.
- Gaussian anamorphosis is more successful than Hólm in our experiments (proportion of points out-of-bounds and skewness measures).