Handling non-Gaussian and nonlinear issues with variational methods

Elías Valur Hólm

ECMWF

May 11, 2010

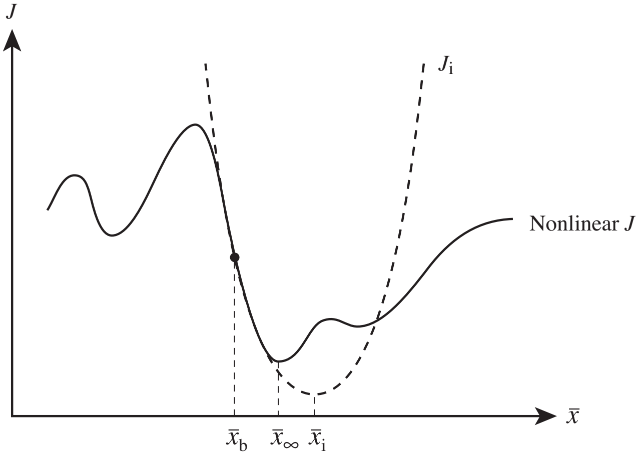
Outline

1. Sources of nonlinearity
2. Model operator with tangent-linear model and physics
3. Observation operators
4. Observation error modelling
5. Background error modelling
6. Gaussianization of background/observation pdfs

Sources of nonlinearity in variational data assimilation

- The full nonlinear problem is solved as a sequence of linear problems (outer loops), each of which is made up of several iteration of the inner loops using tangent linear models $\mathbf{M}$ and observation operators $\mathbf{H}$ at reduced resolution.
- In each iteration the costfunction is linearized around a previous estimate $\mathbf{x}_i$, which gives a quadratic costfunction J_i.
- For weakly nonlinear systems this procedure works well and is widely used.
- The convergence depends on:
  - the accuracy of the first estimate of $\mathbf{x}$ (the background $\mathbf{x}_b$)
  - and the nonlinearity of J.
- There are Monte Carlo assimilation algorithms which deal directly with the nonlinearity. However, often this is not trivial for large dimensional problems (e. g. large number of particles/members needed for capturing nonlinearities with particle filters/ensemble Kalman filters)

The inner/outer loop solution of the nonlinear problem



How to handle model nonlinearities?

- The error in the tangent linear model perturbation at reduced resolution is estimated by comparison with differences of nonlinear integrations,

$$\mathbf{M}\delta\mathbf{x}_0 - (\mathcal{M}(\mathbf{x} + \delta\mathbf{x}_0) - \mathcal{M}(\mathbf{x}))$$

- Factors affecting this difference are:
  - Number of outer loops - each relinearization increases the chance of capturing nonlinearities (presently 3 at ECMWF).
  - Resolution of the inner loops - ideally same as outer loops, but needs to be reduced (at ECMWF factor 3 balances accuracy/efficiency).
  - Length of the assimilation window - longer windows decrease the accuracy of the TL approximation (presently 12-24 h at ECMWF).
  - Model dynamics and physics - chaotic transitions and on/off processes decrease linearity (at ECMWF we gradually enhance linearized physics).
  - Size of the analysis increments - smaller increments with better observations/analysis.

Effect of model variable choice on nonlinearities

- Some model variables are more linear than others, e. g.:
  - Condensation/evaporation rapidly converts between humidity and condensated water, but the sum is more constant in time, $(q, q_l, q_i) \longrightarrow q_t = q + q_l + q_i$.
  - Stratospheric chemistry rapidly converts between NO and NO_2 during the night, but the sum is more constant, $(NO, NO_2) \longrightarrow NO_x = NO + NO_2$.
- We can keep the original model variables in the nonlinear model $\mathcal{M}$ and only use the combined variables in the TL model $\mathbf{M}$ to make the linear approximation better.
- We can also only use the combined variables in the control variable with the aim to get more Gaussian statistics. Or use a different combination for the control variable, e. g. statistics for $\delta q_u = \delta q - \alpha \delta T$ taking into account correlations and balances!
- Lesson is that the correct answer for any model depends on your own research to find the variable giving best linearity and accuracy.

How to handle observation operator nonlinearities?

- The error in the tangent linear observation operator can be estimated by comparing with the difference between two nonlinear observation operator calculations,

$$\mathbf{H}\delta\mathbf{x}_i - (\mathcal{H}(\mathbf{x}_i + \delta\mathbf{x}_i) - \mathcal{H}(\mathbf{x}_i))$$

- Factors affecting this difference are:
  - The linearity of the observation operator.
  - Resolution of the inner loops.
  - The length of the assimilation window.
  - The model physics.
  - The size of the analysis increments - smaller increments will reduce the agitation of nonlinearities above.

The linearity of the observation operator

- Many operators are linear or nearly linear, like those just requiring interpolation of model variables, e. g. airplane temperature.
- Some operators include physical on/off processes, like cloud and rain observations, which require a careful formulation of linear physics: microwave radiances and radar.
- Infrared radiances are sensitive to the presence of clouds and sea ice, abrupt changes to radiance depending on whether it is clear/cloudy with/without ice.

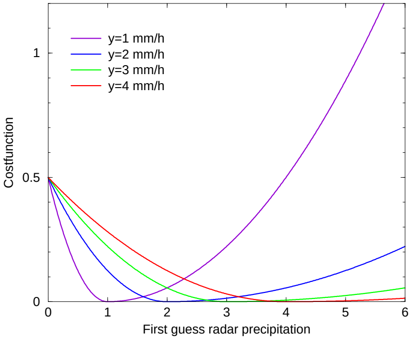
Sources of nonlinear observation errors

- Observation errors are often accurately modelled by Gaussian distributions, e. g. radiosonde temperature errors.
- Non-Gaussian observation error distributions have different sources:
  - The observations include bad observations affected by systematic errors, causing skewed or even multimodal distributions - this is looked at in the lecture on quality control, e. g. Huber norm for robust estimation.
  - The observations contain different populations with different error characteristics, e. g. cloudy and non-cloudy infrared radiances, again possibly causing skewed/multimodal distributions.
  - The observations are not direct, but retrievals, with biases and ambiguities of the retrievals causing skewed/multimodal distributions, e. g. satellite winds from cloud motion vectors (height assignment not unique) and scatterometer winds, which have a multimodal mapping from the measured backscatter to winds.
  - The observations are of bounded quantities, causing skewed distributions, e. g. rain observations, humidity and some chemical tracer observations.

How to model non-Gaussian observation errors?

- For multi-modal distributions, the first step is to identify the characteristics of the different sub-distributions and model the observation errors separately for each class of observations, e. g. cloudy vs. non-cloudy IR radiances.
- More generally, the variance and bias of the observations may vary continuously as a function of some stratifying variable, e. g. the amplitude of the observation itself. An example is rain observations.
- The cause of skewed distributions also needs to be identified. If caused by a superposition of different populations, the same techniques apply as for the multimodal case.
- Skewed distributions caused by boundedness of the observations are modelled with asymmetric observation error distributions, e. g. for rain observations there is larger chance for the error to be on the high than the low side when the observed value is low.

Radar precipitation observations



- Example of asymmetric costfunction for precipitation (Koizumi, Ishikawa and Tsuyuki, JMA):

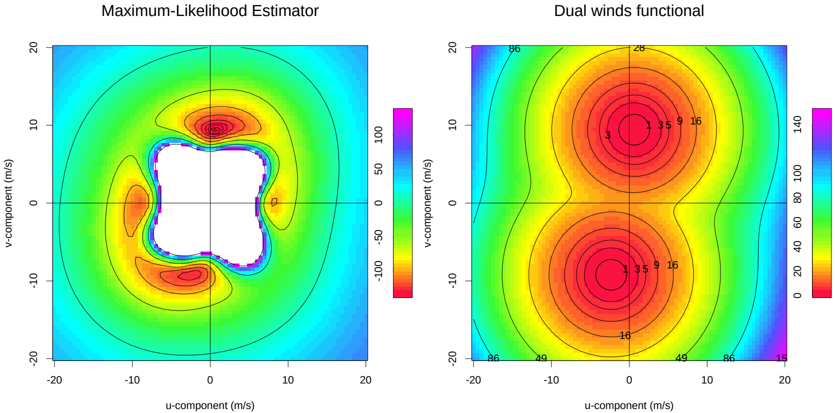
$$J_{rain} = \frac{1}{2r^2}(y - (\mathcal{H}_{rain}(\mathbf{x}_b))^2$$

where $r = r_c$ if $\mathcal{H}_{rain}(\mathbf{x}_b) \leq y$, $r = 3r_c$ if $\mathcal{H}_{rain}(\mathbf{x}_b) > y$, with $r_c = max(y, 1mm/h)$.

How to handle non-Gaussian observation errors in the assimilation?

- Possibilities depend on the minimization algorithm used.
- For the conjugate gradient method, which requires a continuous Hessian and thus strictly quadratic costfunction, we must re-linearize J_o at each outer loop level.
- For the quasi-Newton method, like M1QN3 from INRIA, which can handle weak nonlinearities, J_o is either re-linearized w. r. t. the observation error at outer loop level as above, or the nonlinear observation error is used in the inner loops. This works especially well if the observation error model maintains a convex costfunction, which does not create any additional minima.
- Useful probability distributions giving convex costfunction include:
  - Gaussians with different variances joined at zero: Good for skewed distributions.
  - Huber distribution: A Gaussian in the middle is joined to exponential distributions to the left and to the right. Good for skewed and fat tail distributions with more frequent outliers than a Gaussian.

Scatterometer wind observation costfunction



- The costfunction for wind is multimodal (left), but can be simplified by combining the most probable solutions (right). (Courtesy Mark Leitner)

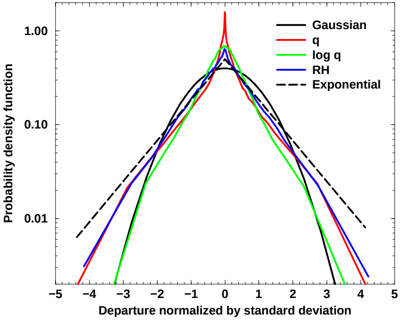
How to handle non-Gaussian background errors?

- The sources of nonlinearities coming from the model integration producing a background state were discussed above.
- If the background errors become non-Gaussian, we need to treat these as perturbations on top of the Gaussian framework, if we are to stay with efficient algorithms which assume quadratic costfunctions. Generally works well in combination with the techniques discussed above to handle nonlinearities at outer loop level (re-linearization etc).
- The main problems arise when:
  - Variables change very abruptly, e. g. due to on/off physical processes: cloud variables.
  - Sharp gradients in horizontal/vertical cause anisotropies: boundary layer top, fronts.
  - Variables contain scales spanning several orders of magnitude: upper tropospheric humidity spans over 4 orders of magnitude.
  - Variables are physically bounded within limits: humidity, clouds, tracers.

Anisotropic and flow-dependent horizontal and vertical correlations

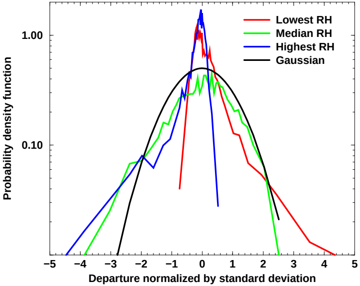
- Anisotropies can be introduced in several ways:
  - Diffusion operator, digital filter and gridpoint formulations of the background error all allow 3D anisotropy. Need specification of coefficients!
  - Coordinate transforms in both horizontal and vertical: geostrophic coordinate transform, boundary layer top and tropopause following vertical coordinates.
- The specification of coefficients needed to produce anisotropy can be made flow dependent (based on last outer loop trajectory).
- In general a statistical regression of error correlations as function of variables/gradients is needed, preferably guided by physical insights/principles.
- A natural extension is to include information from all ensemble members in an ensemble assimilation system.

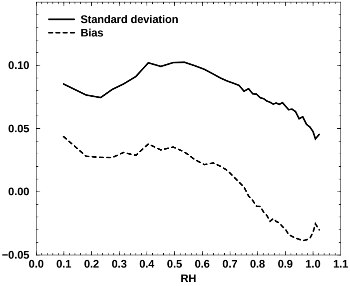
Symptoms of inhomogeneous background error variances



- Global humidity forecast difference histograms at $850hPa$ model level show peak at zero, where all the small errors gather - typical of an inhomogeneous situation.
- Using different variables does not help much.

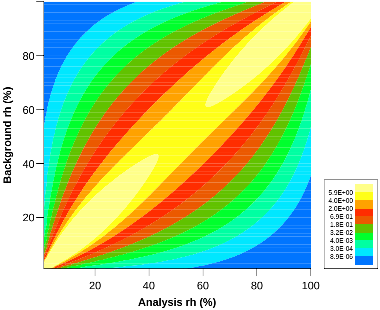
Symptoms of bounded background error variances





- Relative humidity histograms as above for different bins of background relative humidity show asymmetry at boundaries, accompanied by biases.

The joint pdf $P^b(rh^1, rh^2)$

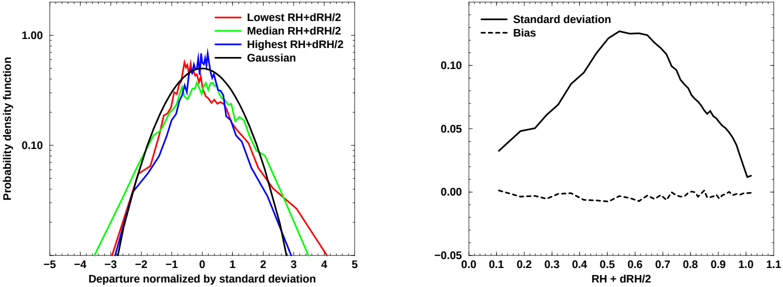


- The joint pdf (from forecast differences) is asymmetric w. r. t. rh^1, but symmetric w. r. t.

$$\frac{1}{2}(rh^1 + rh^2) = rh^1 + \frac{1}{2}\delta rh$$

- We have labelled the two forecasts analysis and background, but should be fc1/fc2.

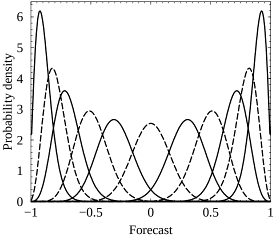
Normalized and symmetrized humidity statistics

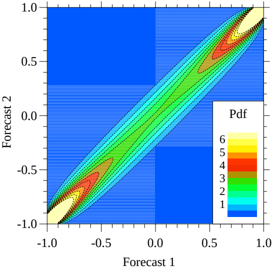


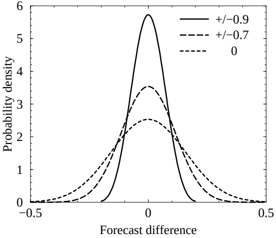
- The symmetric pdf $P(\delta rh | rh^b + \frac{1}{2}\delta rh)$ can be modelled by a normal distribution
- The variance changes with $rh^b + \frac{1}{2}\delta rh$, and the bias is zero. A control variable with approximately a normal distribution $\mathcal{N}(0,1)$ is obtained by nonlinear normalization

$$\widetilde{\delta rh} = \frac{\delta rh}{\sigma(rh^b + \frac{1}{2}\delta rh)}$$

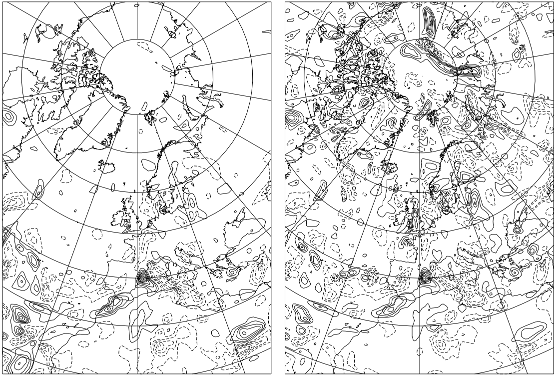
Example: Background error for a harmonic oscillator amplitude







Effect of normalization on homogeneity



- Forecast differences (ca. 850 hPa) for specific humidity δq (left) and normalized relative humidity $\widetilde{\delta rh}$ (right).
- The normalized relative humidity gives a more homogeneous field for the following normalization in spectral or wavelet space.

Implementing nonlinear humidity background errors

- The background error costfunction J_b is now nonlinear

$$J_b = f^T(\delta rh)B^{-1}f(\delta rh)$$

$$f(\delta rh) = \frac{\delta rh}{\sigma(rh^b + \frac{1}{2}\delta rh)}$$

- Implementation requires linear inner loops with nonlinearities in outer loops
  - For outer loop i, inner loops use $\widetilde{\delta rh^i} = \delta rh^i / \sigma(\frac{1}{2}(rh^b + rh^{i-1}))$.
  - In outer loops $\widetilde{\delta rh}$ is given. Solve δrh from the nonlinear equation

$$\frac{\delta rh}{\sigma(rh^b + \frac{1}{2}\delta rh)} - \widetilde{\delta rh} = 0$$

Direct ‘gaussianization’ of background/observation pdfs

- Nonlinear error in both background and observation costfunctions can be transformed analytically to obtain a more Gaussian costfunction, e. g. logarithmic transform of the control variable. An analytic transform like this may not fit the statistics.
- More generally, a numerical transform based on forecast difference statistics can be derived, based on the pdf $P(\delta \mathbf{x})$, or even better the conditional pdf $P(\delta \mathbf{x}|\Phi)$, where Φ is a suitable stratifying variable which separates large and small variances for fields containing different scales (like humidity or atmospheric tracers).
- For a given Φ we just need to find a transformation $f(\delta \mathbf{x}, \Phi)$ of the $\delta \mathbf{x}$ axis so that the probability that $\eta \leq \delta \mathbf{x}$ equals the probability that $\xi \leq f(\delta \mathbf{x}, \Phi)$ for a normal gaussian distribution

The 'gaussianization' transform using cumulative pdfs

- The cumulative pdf Π is related to the gaussian cumulative pdf Π_G by

$$\Pi(\delta\mathbf{x}|\Phi) = \int_{-\infty}^{\delta\mathbf{x}} P(\eta|\Phi)d\eta = \int_{-\infty}^{f(\delta\mathbf{x},\Phi)} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} d\xi = \Pi_G(f(\delta\mathbf{x},\Phi))$$

- Inverting the gaussian cumulative distribution then gives

$$f(\delta\mathbf{x},\Phi) = \Pi_G^{-1}(\Pi(\delta\mathbf{x}|\Phi))$$

- This will always work, but we may end up with a bias term which does not fit into the standard analysis framework - the costfunction needs to be modified to account for the bias.
- However, we could also apply this transform as a finishing step after the symmetrizing transform and in that way avoid the bias term.