Model Error in Variational Data Assimilation

Yannick TRÉMOLET

September 2003

---

1. Introduction to Model Error in VDA.
2. Weak constraint 4D-Var: Theory.
3. Weak constraint 4D-Var: Practice.
4. Choice of control variable and covariance matrix.
5. Future directions.

Model Error in Data Assimilation

- Sequential methods, Kalman Filter (Dee, DAO),
- Ensemble KF (Houtemaker, Evensen),
- 4D-PSAS (Oceanography),
- Representers method (Bennett),
- Weak constraint 4D-Var.

Does Model Error Affect 4D-Var ?

The forecast is run with an imperfect non-linear T511 model.

The analysis is run with a linear T159 model.

Even if the high resolution model was perfect 4D-Var would see an imperfect model.

Model error is present in the IFS assimilation system.

Does it degrade the quality of the analysis ?

Statistical Linear Estimation

Theory tells us that, if all 4D-Var hypotheses are verified, x_a is the BLUE of the true initial condition and:

$$E((J_o)_{min}/p) = 1.$$

If model error is present (D. Dee, 1995):

$$E((J_o)_{min}/p) = 1 + s_n^2.$$

Assimilation window	p	$(J_o)_{min}$	$(J_o)_{min}/p$
6h (4 cycles)	1510574	676277.2	0.45
12h (2 cycles)	1474381	768568.0	0.52
24h (1 cycle)	1483961	934998.2	0.63

$(J_o)_{min}$ at the end of the minimisation (IFS CY23R4).

What is model error ?

Assuming the true state of the atmosphere x^t is known:

Observation error: $y_i = \mathcal{H}(x_i^t) + \epsilon_i^o$.

Analysis error: $x_i^a = x_i^t + \epsilon_i^a$.

Forecast error: $x_i^f = x_i^t + \epsilon_i^f$ where $x_i^f = \mathcal{M}(x_{i-1}^a)$.

Model error: $x_i^t = \mathcal{M}(x_{i-1}^t) + \eta_i$.

None of these quantities can be computed.

Variational Data Assimilation

Theoretical knowledge of the system:

- Equations governing the physical state of the system:

$$\mathcal{G}(x) = 0,$$

- Equations relating the state of the system to observations:

$$\mathcal{H}(x) = y.$$

Taking into account the uncertainties:

$$\begin{aligned} \mathcal{G}(x) &= \varepsilon_g, \\ y - \mathcal{H}(x) &= \varepsilon_h, \end{aligned}$$

with error covariance matrices C_g and R.

Variational Data Assimilation

Combining the two sources of information, the *a posteriori* probability distribution for x given the observations y is (from Bayes theorem):

$$P(x|y) = \alpha \exp\left(-\frac{1}{2}[y - \mathcal{H}(x)]^T R^{-1} [y - \mathcal{H}(x)] - \frac{1}{2}\mathcal{G}(x)^T C_g^{-1} \mathcal{G}(x)\right)$$

The problem of finding the maximum of the probability distribution can be replaced by the problem of finding the minimum of:

$$\begin{aligned} J(x) &= -\ln(P_a(x)) \\ &= \frac{1}{2}[y - \mathcal{H}(x)]^T R^{-1} [y - \mathcal{H}(x)] + \mathcal{G}(x)^T C_g^{-1} \mathcal{G}(x). \end{aligned}$$

Variational Data Assimilation

In meteorology, a prior estimate of the state of the system (background x_b) is known with error ε_b and background error covariance matrix B.

The cost function becomes:

$$\begin{aligned} J(x) &= \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[y - \mathcal{H}(x)]^T R^{-1}[y - \mathcal{H}(x)] \\ &+ \frac{1}{2}\mathcal{F}(x)^T C_f^{-1}\mathcal{F}(x) \end{aligned}$$

where $\mathcal{F}$ represents the remaining theoretical knowledge after background information has been accounted for.

No hypothesis has been made regarding x yet !

Same formulation as Sasaki, 1970.

3D Variational Data Assimilation

- x is the 3D state of the atmosphere at analysis time,
- $\mathcal{F}$ includes balance constraints,
- $\mathcal{H}$ is a (sophisticated) 3D operator.

4D Variational Data Assimilation

- x is the 4D state of the atmosphere during the assimilation window,
- $\mathcal{F}$ includes equations governing the evolution of the atmosphere (model $\mathcal{M}$) and other constraints (DFI...),
- $\mathcal{H}$ is a (sophisticated) 4D operator, accounting for the time dimension:
  - serially correlated observations,
  - observations used at correct time.

Strong constraint 4D-Var

- Model is perfect,
- x is a function only of the initial condition x_0 (the size of the control variable is reduced):

$$x_i = \mathcal{M}_i(x_{i-1}) = \mathcal{M}_{0,i}(x_0),$$

- constraints associated to the model disappear from $\mathcal{F}$ (no model error covariance matrix),
- $\mathcal{H}$ is the same 4D operator,
- operational implementation.

Weak constraint 4D-Var

- The model is not perfect,
- x is a 4D control vector,
- Model error verifies:

$$\eta_i = x_i - \mathcal{M}_i(x_{i-1}) = \mathcal{F}_i(x_i),$$

- Model error covariance matrix Q has to be defined,
- Choices of control vector $x = (x_i)_{i=0,\ldots,n}$ and $\chi = (x_0, (\eta_i)_{i=1,\ldots,n})$ are equivalent.

Weak constraint 4D-Var

Linear evolution of perturbations:

$$\delta x_i = M_i \delta x_{i-1} + \delta \eta_i = M_{0,i} \delta x_0 + \sum_{j=1}^{i} M_{j,i} \delta \eta_j$$

Variation of the total cost function:

$$\begin{aligned} \delta J(x_0, \eta) \quad = \quad & \delta x_0^T B^{-1} (x_0 - x_b) + \delta x_0^T \sum_{i=1}^{n} M_{i,0}^T H_i^T R_i^{-1} d_i \\ & + \sum_{j=1}^{n} \delta \eta_j^T \sum_{i=j}^{n} M_{i,j}^T H_i^T R_i^{-1} d_i + \sum_{j=1}^{n} \sum_{k=1}^{n} \delta \eta_j^T Q_{j,k}^{-1} \eta_k \end{aligned}$$

- $\frac{\partial J}{\partial \eta_j}$ is obtained by accumulating contributions from the adjoint at steps $i = j, \ldots, n$.
- The gradient of the cost function is still obtained by one backward integration of the adjoint.

Size of the problem

Current 4D-Var operational resolution:

- Horizontal: T159,
- Vertical: 60 levels,
- Time-step: 1800s for 12 hours.

The size of the control variable is:

- Perfect model: $N = 7.7 \times 10^6$,
- Weak constraint: $(n + 1) \times N = 1.9 \times 10^8$.

The model error covariance matrix would have 1.9×10^{16} elements and occupy 131,331 Tb of memory.

Sources of information

- 3×10^6 observations are available each day to estimate 1.9×10^{16} elements of Q.
- At today's rate of observation it would take 6 billion years to gather as many observations as there are parameters in Q.
- Assuming there is no redundancy in observed quantities and model error can be separated from other sources of error...
- There is not enough information to solve the problem:

Approximations are required !!!

Model Error and Model Bias

- 4D-Var is *bias blind*: it assumes errors are unbiased or that biases have been removed.
- Biases for each of the errors considered:

$$\beta_b = < x^t - x_b > \ , \ \varepsilon_b = (x^t - x_b) - \beta_b \ \text{ and } \ B = < \varepsilon_b^T \varepsilon_b >,$$

$$\beta_o = < y^t - \mathcal{H}(x^t) > \ , \ \varepsilon_o = (y^t - \mathcal{H}(x^t)) - \beta_o \ \text{ and } \ R = < \varepsilon_o^T \varepsilon_o >,$$

$$\beta_f = < \mathcal{F}(x^t) > \ , \ \varepsilon_f = \mathcal{F}(x^t) - \beta_f \ \text{ and } \ Q = < \varepsilon_f^T \varepsilon_f > .$$

- Unbiased variables are noted with $\widetilde{(.)}$: $\tilde{x_b} = x_b - \beta_b$.

Model Error and Model Bias

- The variational data assimilation cost function should be:

$$\begin{aligned} J(x) &= \frac{1}{2}(x-\tilde{x_b})^T B^{-1}(x-\tilde{x_b}) + \frac{1}{2}[\tilde{y}-\tilde{\mathcal{H}}(x)]^T R^{-1}[\tilde{y}-\tilde{\mathcal{H}}(x)] \\ &+ \frac{1}{2}\tilde{\mathcal{F}}(x)^T Q^{-1}\tilde{\mathcal{F}}(x). \end{aligned}$$

- 4D-Var is in fact minimising:

$$\begin{aligned} J(x) &= \frac{1}{2}(x-\tilde{x_b}+\beta_b)^T B^{-1}(x-\tilde{x_b}+\beta_b) \\ &+ \frac{1}{2}[\tilde{y}-\tilde{\mathcal{H}}(x)+\beta_o]^T R^{-1}[\tilde{y}-\tilde{\mathcal{H}}(x)+\beta_o] \\ &+ \frac{1}{2}[\tilde{\mathcal{F}}(x)+\beta_f]^T Q^{-1}[\tilde{\mathcal{F}}(x)+\beta_f] \end{aligned}$$

- The analysis is biased.

Bias Example

- Scalar z to be estimated from a and b (with $\sigma_a = \sigma_b = 1.0$),
- Cost function: $J(z) = (z-a)^2 + (z-b)^2$,
- Minimum is reached for $z = (a+b)/2$.
- If a and b are biased, the actual cost function should be:

$$J(z) = (z-\tilde{a})^2 + (z-\tilde{b})^2$$

- Minimum reach for $z = (\tilde{a}+\tilde{b})/2$.
- Biases do affect the final estimate.
- $J_{min} = (a-b)^2/2$ or $J_{min} = (\tilde{a}-\tilde{b})^2/2$.

Model Error and Model Bias

- Biases are the mathematical expectation of the errors, or an *ensemble average* of the errors, not a time average.
- The time averaged error is sometimes (abusively) called *forecast bias* or *model bias*.
- Time averaged error is usually non-zero. It can be unbiased in expectation terms (*ie* in the realisations dimension).

Weak constraint 4D-Var: Practice

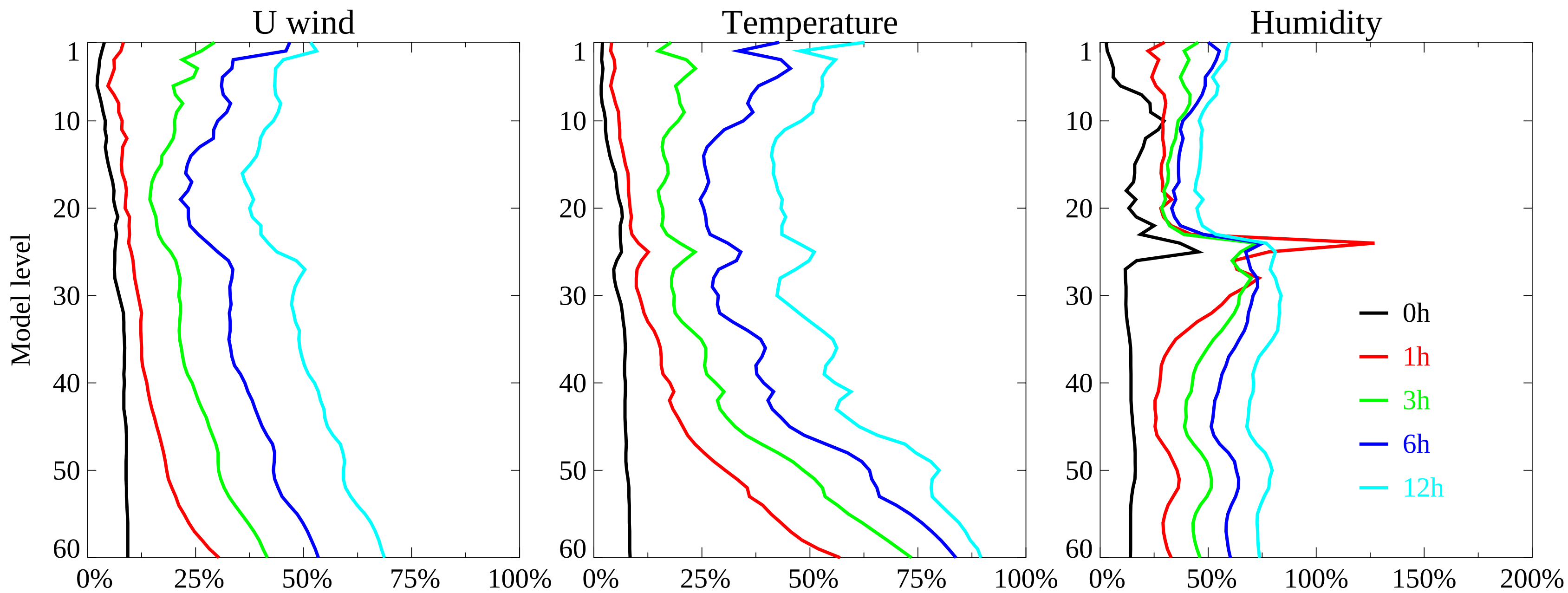
What can really be done ?

1. Choice of model error control variable,
2. Model error statistics,
3. Using model error.

Characteristics of model error

- Some components are constant (orography),
- Some components are almost periodic (diurnal cycle),
- Some components are flow dependent (physical processes),
- Model error is correlated in time (in addition to growth),
- Discretisation and numerical errors may be more random,
- In incremental 4D-Var context, model error is the sum of:
  - Error between the atmosphere and HR NL model,
  - Error between inner and outer loop: HR NL model vs. LR TL model with limited physics.

Inner Loop Approximations



Evolution of the relative error in the T159 tangent linear model with respect to the T511 forecast model over the assimilation window.

Representing Model Error

1. Constant forcing $\eta_i = \eta$,
2. Markov chain: $\eta_i = \frac{\mu}{\mu+(1-\mu^2)^{1/2}} \eta_{i-1} + \frac{(1-\mu^2)^{1/2}}{\mu+(1-\mu^2)^{1/2}} r_k$ (Zupanski),
3. Fourier series expansion (Diurnal cycle),
4. Spin-up/down term (vanishing term).

Choice of control variable

The 4D control variable $\{x_i,\ \ i = 0, \ldots, n\}$ can be replaced by:

x_0	x_0	x_0	x_0
η_i	$\eta_i = \eta$	$\eta_i = 0$	$\eta_i = 0$
$x_i = \mathcal{M}(x_{i-1}) + \eta_i$	$x_i = \mathcal{M}(x_{i-1}) + \eta$	$x_i = \mathcal{M}(x_{i-1})$	$x_i = x_0$
$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
Weak constraint 4D-Var	Practical Implementation	4D-Var	3D-Var

Model Error Covariance Matrix

- Define statistics for model error from the model's implementation (numerics and physics),
- Based on B matrix (Zupanski),
- Statistics on innovation and residual (Daley): Method based on KF and not (easily) applicable to 4D-Var,
- Ensemble of slow modes (Phillips, Cohn and Parrish),
- Online estimation $Q(\alpha)$ (Dee),
- Statistics on A and B, using $B = MAM^T + Q$,
- Comparison between LR linear runs and HR nonlinear runs.

$$Q = \alpha B$$

Computational cost

- Cost per iteration:
  - Double the size of the control vector,
  - Double cost of linear algebra (at most, cost in $\beta + n\tau$),
  - Add forcing in linear model and adjoint (negligible additional cost),
  - Overall, cost of linear algebra is negligible.
- Number of iterations:
  - Fixed number of iterations in first minimisation,
  - Depends on conditioning in following minimisations,
  - Efficient preconditioning: Lanczos algorithm (M. Fisher).

Computational cost (measured)

Elapsed time	Strong contraint	Weak constraint
Dot product	3.3 ms	4.9 ms
SIM4D	89.2 sec	91.9 sec
Memory	703 Mb	769 Mb

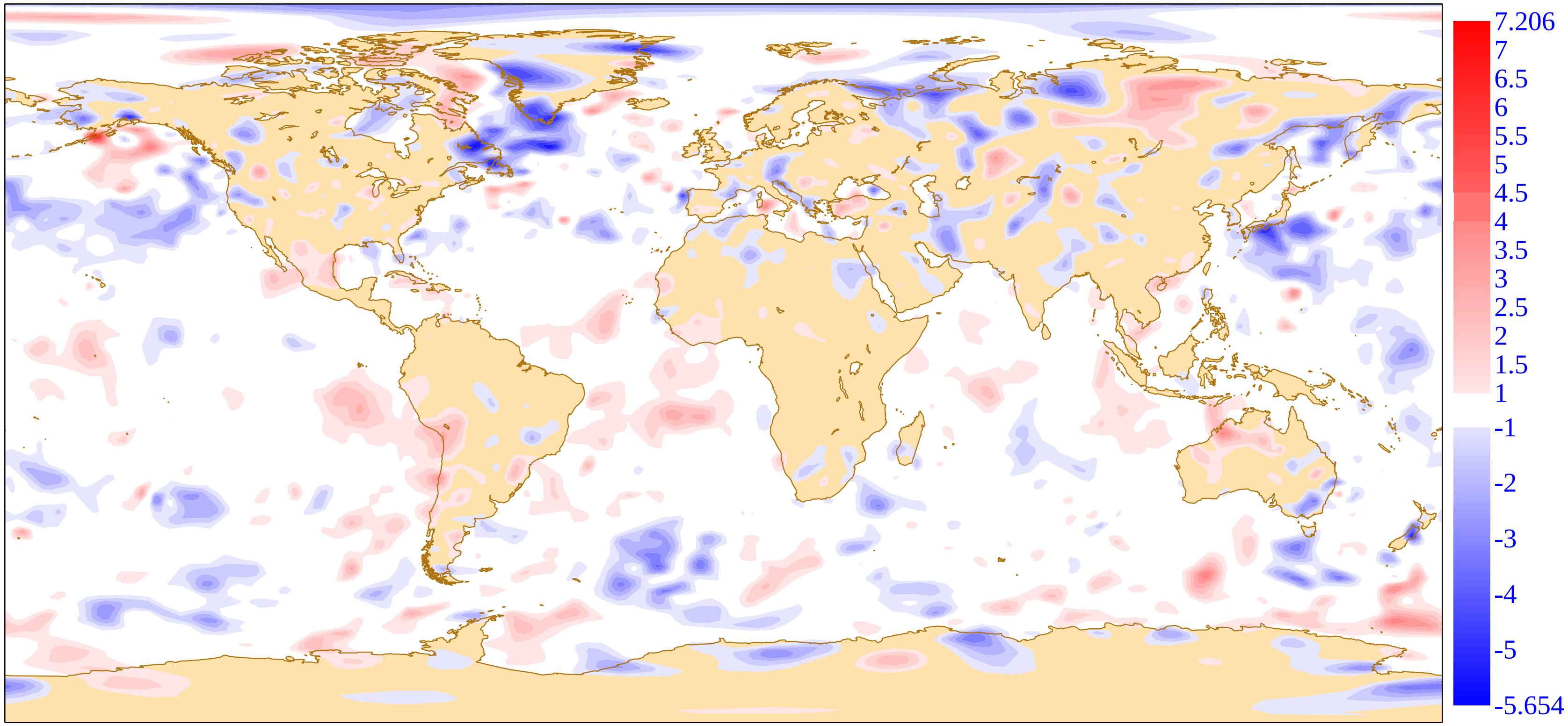
Measured elapsed time and memory for T159 minimisation with the IFS on IBM SP, 128 CPUs

	Strong contraint	Weak constraint
First minimisation	6575.9	10996.7
Second minimisation	2007.1	1302.0

Measured condition number in IFS (CY26R3)

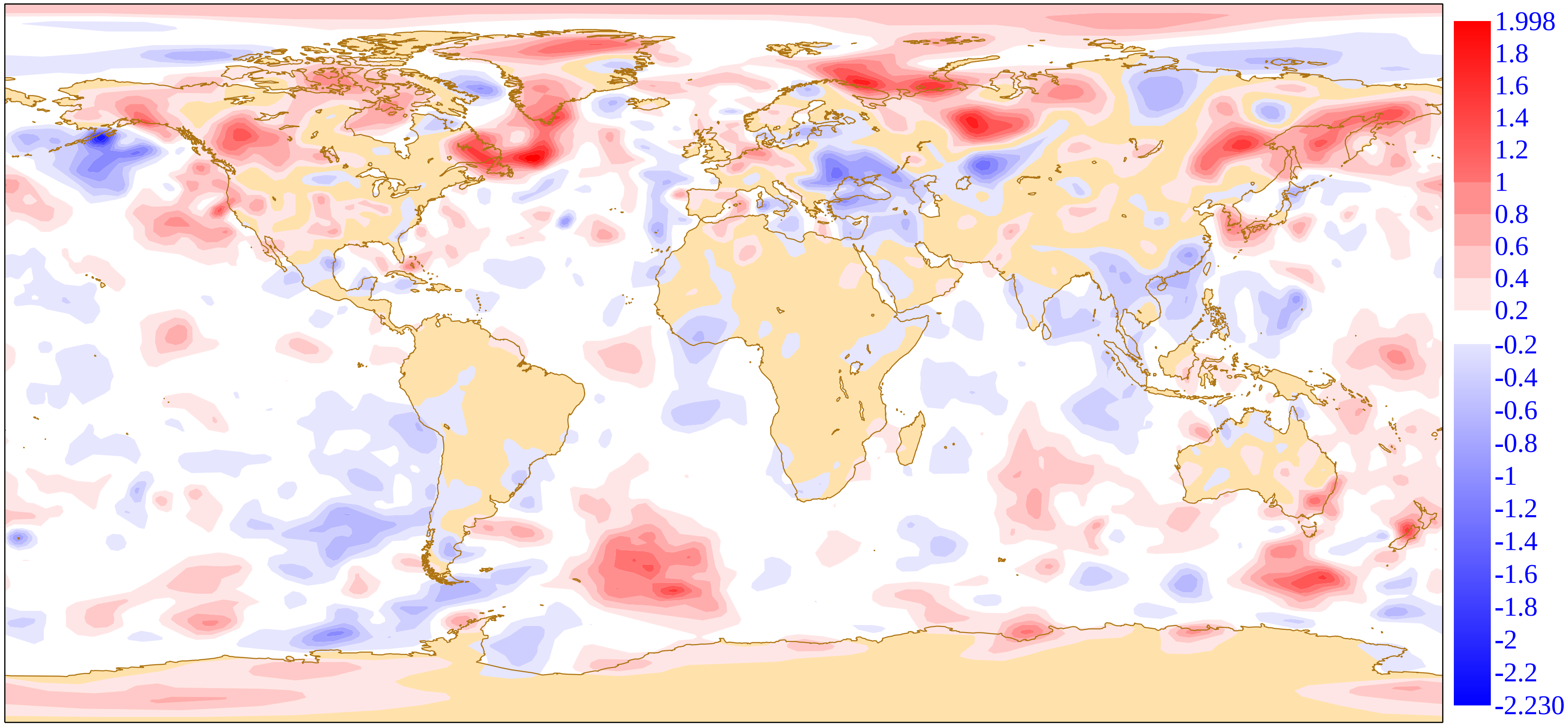
Preliminary results

Model Error (Constant Forcing), $Q = 10^{-1}B$, Surface Pressure



Preliminary results

Impact on Initial Condition, $Q = 10^{-1}B$, Surface Pressure



Model Error Evolution

- Knowing x_0 and η, x_i can be computed:

$$\begin{aligned} x_i &= \mathcal{M}_i(x_{i-1}) + \eta_i \\ &= \mathcal{M}_i\left(\mathcal{M}_{i-1}\left(\ldots\left(\mathcal{M}_1(x_0) + \eta_1\right)\ldots\right) + \eta_{i-1}\right) + \eta_i \end{aligned}$$

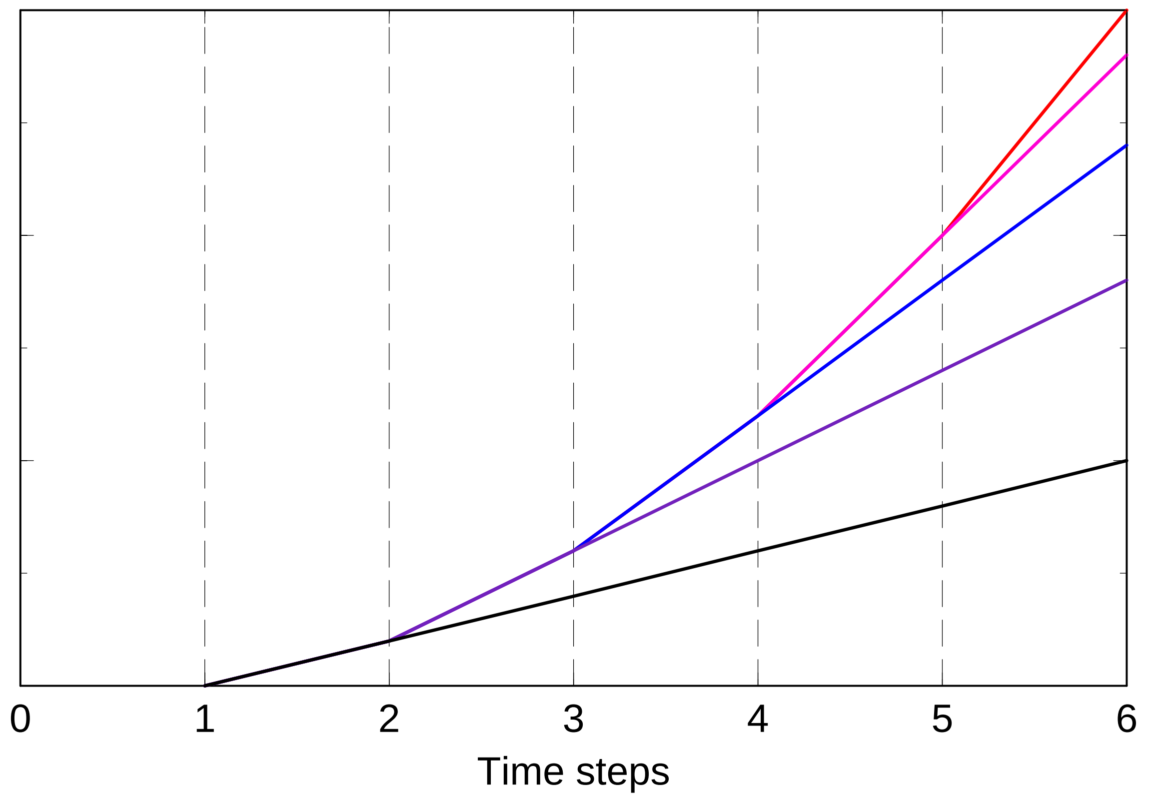
- If model errors are small enough, $\mathcal{M}_i$ can be linearised:

$$x_i = x_i^m + \sum_{j=1}^{i} M_i \ldots M_{j+1} \eta_j$$

where $x_i^m = \mathcal{M}_i(\ldots(\mathcal{M}_1(x_0))\ldots)$ is the perfect model forecast.

- Early components dominate the resulting impact of model error.
- Constant model error is influenced mostly by early errors.

Impact of Forcing



- Accumulated impact of early forcing in forecast is larger,
- It dominates in the cost function.

Incremental 4D-Var

- Incremental formulation gives:

$$x_i = \mathcal{M}_i \left(\mathcal{M}_{i-1} \left(\ldots \left(\mathcal{M}_1 (x_b + \delta x_0) + \eta_1 \right) \ldots \right) + \eta_{i-1} \right) + \eta_i$$

- Linearisation leads to:

$$x_i = x_i^b + M_i \ldots M_1 \delta x_0 + \sum_{j=1}^{i} M_i \ldots M_{j+1} \eta_j$$

- δx_0 can be identified with η_0.
- Because Q and B are proportional, δx_0 and η are constrained in the same directions, with relative amplitudes controlled by α.
- They both predominantly retrieve the same information.

Another choice of control variable

x_0	x_0	x_0
η_i or β_i	$\eta_i = \eta$	$\beta_i = \beta$
$x_i = \mathcal{M}(x_{i-1}) + \eta_i$ or $x_i = \mathcal{M}_{0,i}(x_0) + \beta_i$	$x_i = \mathcal{M}(x_{i-1}) + \eta$	$x_i = \mathcal{M}_{0,i}(x_0) + \beta$
$\Downarrow$	$\Downarrow$	$\Downarrow$
Weak constraint 4D-Var	Constant Forcing	Bias

- The choices of control vector $x = (x_i)_{i=0,\dots,n}$, $\chi = (x_0, (\eta_i)_{i=1,\dots,n})$ or $\chi' = (x_0, (\beta_i)_{i=1,\dots,n})$ are equivalent.
- The approximations $\eta_i = \eta$ and $\beta_i = \beta$ are not.

Control of Model Bias

- Controlling model bias means:

$$x_i - x_i^m = \beta$$

- All the time components of model error have equal influence on the cost function.
- It is a good representation of the time averaged error. This is the *bias* for some authors.
- Coded in the IFS *with extra features*.

Propagation of Error Covariance Matrix

We have $x_{i+1} = \mathcal{M}(x_i^a)$ and $x_{i+1}^t = \mathcal{M}(x_i^t) + \eta_i$.
Taking the difference leads to:

$$\epsilon_{i+1}^f = \mathcal{M}(x_i^a) - \mathcal{M}(x_i^a - \epsilon_i^a) - \eta_i.$$

A first order approximation gives $\epsilon_{i+1}^f = M\epsilon_i^a - \eta_i$.

If analysis error and model error are uncorrelated (?):

$$E(\epsilon_{i+1}^f(\epsilon_{i+1}^f)^T) = E(M\epsilon_i^a(M\epsilon_i^a)^T) + E(\eta_i\eta_i^T),$$

which is:

$$P^f = MP^aM^T + Q.$$

Phillips (86) hypothesis: P^f is dominated by Q if analysis has frequent access to good data (supported by HL86-LH86).

Model Error Covariance Matrix

- Cohn and Parrish (91) based on Phillips (86),
- Model error consists of uncorrelated slow modes:

$$Q = V_s S V_s^*$$

  and S is diagonal.
- Energy spectrum is known: the elements of S depend on a few (one) parameters.
- Forecast error covariance matches results from Hollingsworth and Lönnberg (86) and LH86.

Estimation of error covariance parameters

Can be done online (Dee).

Q is parameterised by α.

Expectation of innovation vector $v = y - Hx$ is:

$E(vv^T) = E(\epsilon^o(\epsilon^o)^T) + E(H\epsilon^f(H\epsilon^f)^T) - E(\epsilon^o(H\epsilon^f)^T) - E(H\epsilon^f(\epsilon^o)^T)$.

If correlation between observation and forecast error can be neglected:

$$E(vv^T) = HP^fH^T + R.$$

Remember that $P^f = MP^aM^T + Q$,

$$E(vv^T) = H(MP^aM^T + Q(\alpha))H^T + R = S(\alpha).$$

Estimation of error covariance parameters

$$E(vv^T) = H(MP^aM^T + Q(\alpha))H^T + R = S(\alpha)$$

Maximum likelihood value for α is obtained for:

$$\alpha^{ML} = \arg\min_{\alpha}(\log\det S(\alpha) + v^T S^{-1}(\alpha) v)$$

$$\text{and} \quad \frac{\partial f}{\partial \alpha_i} = \text{Trace}\left[(S^{-1} - S^{-1}vv^T S^{-1})\frac{\partial S}{\partial \alpha_i}\right].$$

Lanczos method (Fisher).

Simple expression of $S(\alpha)$?

Problem can be simplified: if $S = \alpha S_0$ then $\alpha^{ML} = \frac{1}{p} v^T S_0^{-1} v$.

Using model error

- Model error is valuable information which can be used in the forecast integration,
- Add a forcing term in the forecast model: Empirical tests have shown good results (Saha, Thiébaux and Morone),
- Model bias can be added at post-processing stage.
- Model error might help identify model deficiencies,
- Model error term can be used in sensitivity computations.

Future developements in the IFS

Start by weak constraint 4D-Var with:

- Control variable: constant forcing or bias,
- Covariance matrix: αB.

Future developments:

- Cycling, archiving, verification.
- Q based on slow modes.
- Online estimation of parameters of Q.
- Model error variable in time.
- Cross correlations: model error is state dependent and should be correlated with background.