# Allowing for model error in strong constraint 4DVAR

Kat Howes
Supervised by Dr Amos Lawless and Dr Alison Fowler
Sponsored by NERC

University of Reading

2nd June 2015

# Contents

- Model errors
- Development of a combined model error and observation error covariance matrix for use in 4DVar
- Estimation of the combined matrix with diagnostics
- Results

# Model error

**Models** are **best representations** of **true** dynamical **systems**

$$\mathbf{x}_i = \widetilde{\mathbf{M}}_{\{i-1\}\rightarrow i}\mathbf{x}_{i-1} \qquad i = 1, 2, ...$$

- Inaccurate parameter specifications
- Inaccurate parametrisations of sub-grid physical processes
- Inaccurate specification of boundary conditions
- Numerical schemes only approximate solutions
- Poor model resolution

# Random model error

$$
\begin{aligned}
\mathbf{x}_i^t &= \mathbf{M}_{\{i-1\}\rightarrow i}\mathbf{x}_{i-1}^t \\
&= \widetilde{\mathbf{M}}_{\{i-1\}\rightarrow i}{\mathbf{x}^t}_{i-1} + \boldsymbol{\eta}_i \qquad\qquad i = 1, 2, ...,
\end{aligned}
$$

where the model error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$.

# 4Dvar Four dimensional variational data assimilation

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0)^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0),$$

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{y}_N \end{pmatrix} \quad \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0\to 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{0\to N} \end{pmatrix} \text{ and } \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{R}_1 & 0 & \cdots & 0 \\ \vdots & 0 & 0 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{R}_N \end{pmatrix}.$$

- $\epsilon_b = \mathbf{x}^b - \mathbf{x}^t{}_0$ with $\epsilon_b \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$,
- $\epsilon_{ob} = \hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}^t{}_0$ with $\epsilon_{ob} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}})$.

# 4Dvar with erroneous model

$$\mathcal{J}(\mathbf{x}_0) \quad = \quad \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\hat{\mathbf{y}} - \widetilde{\mathbf{H}}\mathbf{x}_0)^T{\mathbf{R}^*}^{-1}(\hat{\mathbf{y}} - \widetilde{\mathbf{H}}\mathbf{x}_0),$$

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{y}_N \end{pmatrix} \qquad \widetilde{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\widetilde{\mathbf{M}}_{0\to 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N\widetilde{\mathbf{M}}_{0\to N} \end{pmatrix}$$

$$\boxed{\epsilon^*_{ob} = \hat{\mathbf{y}} - \widetilde{\mathbf{H}}{\mathbf{x}^t}_0 \text{ with } \epsilon^*_{ob} \sim \mathcal{N}(?, ?)}$$

# Combined model error and observation error

$$\begin{aligned} \epsilon_{obi} &= \mathbf{y}_i - \mathbf{H}_i \mathbf{M}_{0\to i} \mathbf{x}^t{}_0, & \epsilon_{obi} &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i) & (1) \\ \epsilon^*_{obi} &= \mathbf{y}_i - \mathbf{H}_i \widetilde{\mathbf{M}}_{0\to i} \mathbf{x}^t{}_0, & \epsilon^*_{obi} &\sim \mathcal{N}(?, ?) & (2) \end{aligned}$$

Substracting (1) from (2) and rearranging,

$$\begin{aligned} \epsilon^*_{obi} &= \epsilon_{obi} + \mathbf{H}_i(\mathbf{M}_{0\to i}\mathbf{x}^t{}_0 - \widetilde{\mathbf{M}}_{0\to i}\mathbf{x}^t{}_0), \\ &= \epsilon_{obi} + \mathbf{H}_i \sum_{j=1}^{i} \widetilde{\mathbf{M}}_{j\to i}\eta_j, \\ < \epsilon^*_{obi} > &= \mathbf{0}. \end{aligned}$$

# Combined model error and observation error covariance

Let,

$$\mathbf{R}^*_{(i,k)} = < \epsilon^*_{ob\,i}(\epsilon^*_{ob\,k})^T > .$$

Then,

$$\mathbf{R}^*_{(i,k)} = \begin{cases} \mathbf{R}_0 & \text{for i=k=0} \\ \mathbf{R}_i + \mathbf{H}_i \left[ \sum_{j=1}^{min(i,k)} \widetilde{\mathbf{M}}_{j\to i} \mathbf{Q}_j \widetilde{\mathbf{M}}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{for i=k} \\ \mathbf{H}_i \left[ \sum_{j=1}^{min(i,k)} \widetilde{\mathbf{M}}_{j\to i} \mathbf{Q}_j \widetilde{\mathbf{M}}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{otherwise.} \end{cases} \quad (3)$$

# Combined model error and observation error covariance matrix

$$\mathbf{R}^* = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{R}_1 + \mathbf{Q}^*_{(1,1)} & \mathbf{Q}^*_{(1,2)} & \cdots & \mathbf{Q}^*_{(1,N)} \\ \vdots & \mathbf{Q}^*_{(2,1)} & \mathbf{R}_2 + \mathbf{Q}^*_{(2,2)} & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ 0 & \mathbf{Q}^*_{(N,1)} & \cdots & \cdots & \mathbf{R}_N + \mathbf{Q}^*_{(N,N)} \end{pmatrix}.$$

- increase in block diagonal terms due to model error,
- off diagonal block model error covariance terms (which represent time correlations),
- model error covariance terms increase over time.

# Increased model uncertainty over time

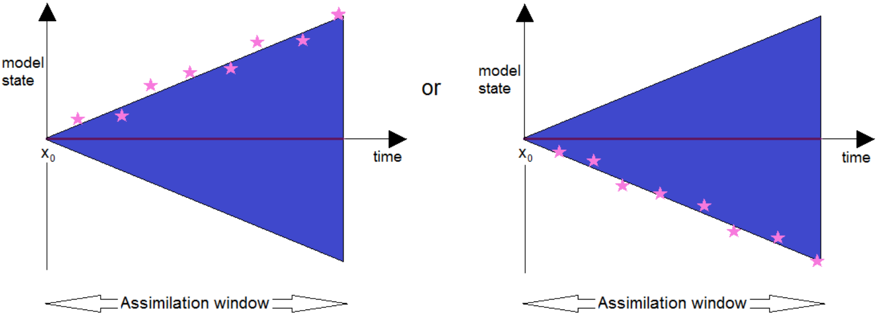


Figure: Worse case scenario observations are normally distributed around the true model which evolves further from the best known model over time.

# Increase in analysis accuracy: Scalar case

- Erroneous model $x_i = \widetilde{\beta} x_{i-1}$,
- true model state $x^t{}_i = \beta x^t{}_{i-1} = \widetilde{\beta} x^t{}_{i-1} + \eta_i$
- direct observations at time $t_1$ with operator $h_1 = 1$,
- ${\sigma_{ob}}^{*2} = {\sigma_{ob}}^2 + {\sigma_q}^2$

Difference in the analysis error variance,

$$ {\sigma_A}^2 - {\sigma_{A^*}}^2 = \frac{{\sigma_q}^4 {\sigma_b}^4 \widetilde{\beta}^2}{(\widetilde{\beta}^2 {\sigma_b}^2 + {\sigma_{ob}}^2 + {\sigma_q}^2)(\widetilde{\beta}^2 {\sigma_b}^2 + {\sigma_{ob}}^2)^2} \geq 0. $$

# Increase in analysis accuracy: Scalar case

Increase in analysis accuracy more significant,

- increase in: model error, observation accuracy,
- decrease in: background accuracy

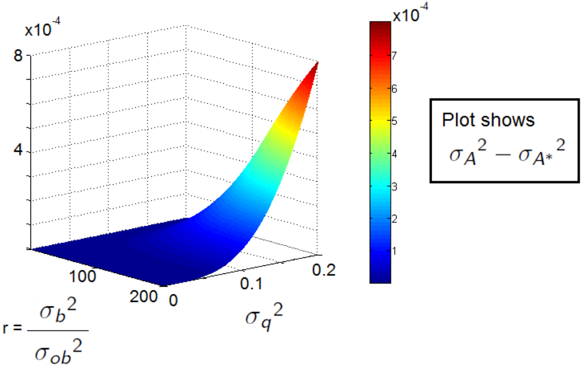


Figure: Analysis accuracy for scalar case $\widetilde{\beta} = 1$, $\sigma_{ob}{}^2 = 10^{-3}$

# How do we specify the model error statistics?

$$
\mathbf{R}^*{}_{(i,k)} = \begin{cases} \mathbf{R}_0 & \text{for i=k=0} \\ \mathbf{R}_i + \mathbf{H}_i \left[ \sum_{j=1}^{min(i,k)} \widetilde{\mathbf{M}}_{j\to i} \mathbf{Q}_j \widetilde{\mathbf{M}}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{for i=k} \\ \mathbf{H}_i \left[ \sum_{j=1}^{min(i,k)} \widetilde{\mathbf{M}}_{j\to i} \mathbf{Q}_j \widetilde{\mathbf{M}}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{otherwise,} \end{cases} \tag{4}
$$

How can we specify $\mathbf{Q}_j$?

# Consistency diagnostics

- Desroziers et al. developed diagnostics for use as quality checks for **B** and **R** with no model evolution [1].
- One of these diagnostics has been previously formulated to include model evolution with an erroneous model [2]. Let the innovation vector,

$$(\mathbf{d}^o{}_b)_1 \quad = \quad \mathbf{y}_1 - \mathbf{H}_1\widetilde{\mathbf{M}}_{0\to 1}\mathbf{x}^b.$$

Taking the statistical expectations of innovations ,

$$E[(\mathbf{d}^o{}_b)_1(\mathbf{d}^o{}_b)_1{}^T] \quad = \quad \mathbf{R}_1 + \mathbf{H}_1\widetilde{\mathbf{M}}_{0\to 1}\mathbf{B}\widetilde{\mathbf{M}}_{0\to 1}{}^T\mathbf{H}_1{}^T + \mathbf{H}_1\mathbf{Q}_1\mathbf{H}_1{}^T.$$

[1] G. Desroziers, L. Berre, B. Chapnik, P. Poli: Diagnosis of observation, background and analysis-error statistics in observation space Quarterly *Journal of the Royal Meteorological Society*, vol. 131, 2005, pp. 3385-3396.

[2] D. P. Dee: On-line estimation of error covariance parameters for atmospheric data assimilation *Monthly weather review*, vol.123, 1995, pp. 1128-1145.

# Estimation of combined error covariance matrix

We have developed a method to estimate $\mathbf{R}^*{}_{(i,k)}$. Let,

- $(\mathbf{d}^o{}_b)_i = \mathbf{y}_i - \mathbf{H}_i \widetilde{\mathbf{M}}_{0\to i}\mathbf{x}^b$,
- $(\mathbf{d}^o{}_b)_k = \mathbf{y}_k - \mathbf{H}_k \widetilde{\mathbf{M}}_{0\to k}\mathbf{x}^b$.

Then,

$$\boxed{\mathbf{R}^*{}_{(i,k)} = E[(\mathbf{d}^o{}_b)_i (\mathbf{d}^o{}_b)_k{}^T] - \mathbf{H}_i \widetilde{\mathbf{M}}_{0\to i} \mathbf{B} \widetilde{\mathbf{M}}_{0\to k}{}^T \mathbf{H}_k{}^T. \qquad (5)}$$

Note $\mathbf{H}_i \widetilde{\mathbf{M}}_{0\to i} \mathbf{B} \widetilde{\mathbf{M}}_{0\to k}{}^T \mathbf{H}_k{}^T$ can be estimated for a very large system using the randomisation method [3].

[3] E. Andersson: Modelling the temporal evolution of innovation statistics *Proceedings of Workshop on recent developments in data assimilation for atmosphere and ocean, ECMWF, Reading*, 2003, pp. 153-164.

# Idealized coupled nonlinear model

Couples the Lorenz 63 system and 2 linear equations (Molteni et al. [4]),

$$
\begin{aligned}
\dot{x} &= -\sigma x + \sigma y + \alpha v, \\
\dot{y} &= -xz + rx - y + \alpha w, \\
\dot{z} &= xy - bz, \\
\dot{w} &= -\Omega v - k(w - w^*) - \alpha y, \\
\dot{v} &= \Omega(w - w^*) - kv - \alpha x,
\end{aligned}
\tag{6}
$$

where $\sigma = 10$, $r = 30$, $b = \frac{8}{3}$, $k = 0.1$, $\Omega = \frac{\pi}{10}$ and $w^* = 2$.

- Runge-Kutta 2nd order method with fixed time step $\Delta t = 0.01$ used to approximate solution of coupled ODE's $\mathbf{x}_i = \widetilde{\mathcal{M}}_{\{i-1\}\to i}(\mathbf{x}_{i-1})$.
- Best known representation of the system that contains model error.

---

[4] F. Molteni, L. Ferranti, T.N. Palmer, P. Viterbo: A dynamical interpretation of the global response to equatorial Pacific SST anomalies *Journal of climate*, **vol.6**, 1993, pp. 777-795.

# True idealized coupled nonlinear model

Parameter perturbation method: Stochastic forcing simulation [5], but with Gaussian error distributions and random error at each time-step.

The true parameter values $\sigma^t$, $k^t$ and $\alpha^t$ change at every time-step,

- ${\sigma_i}^t = \gamma_\sigma \sigma$, where $\gamma_\sigma \sim \mathcal{N}(\mathbf{1}, \frac{1}{12}^2)$,
- ${k_i}^t = \gamma_k k$, where $\gamma_k \sim \mathcal{N}(\mathbf{1}, \frac{1}{6}^2)$,
- ${\alpha_i}^t = \gamma_\alpha \alpha$, where $\gamma_\alpha \sim \mathcal{N}(\mathbf{1}, \frac{1}{12}^2)$,

The difference between the true and erroneous model at each time can be considered as additive model error $\boldsymbol{\eta}_i$ of the form,

$$\mathbf{x}_i^t = \mathcal{M}_{\{i-1\}\rightarrow i}(\mathbf{x}_{i-1}^t) = \widetilde{\mathcal{M}}_{\{i-1\}\rightarrow i}({\mathbf{x}^t}_{i-1}) + \boldsymbol{\eta}_i \qquad i = 1, 2, ...500.$$

[5] R. Buizza, M. Miller, T.N. Palmer: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system *Quarterly Journal of the Royal Meteorological Society*, vol.125, 1999, pp. 2887-2908.

# Numerical experiments: design

- Assimilation window length 500 time-steps of length $\Delta t = 0.01$, with all variables observed every 10 time-steps directly. Let $\mathbf{B} = \mathbf{R}_i = 10^{-4}\mathbf{I}$.
- Perturb the true model states using $\mathbf{B}$ and $\mathbf{R}_i$ respectively to produce background model state $\mathbf{x}^b$ and observations $\mathbf{y}_i$.
- Select background vector $\mathbf{x}^b$ and perturb using $\mathbf{B}$ to obtain a sample of 20 background values (note these are all at initial time $t_0$).
- For each observation time $t_i$: select observation vector $\mathbf{y}_i$ and perturb using $\mathbf{R}_i$ to obtain a sample of 20 observations.
- Use these samples to estimate $(\mathbf{d}^o{}_b)_i = \mathbf{y}_i - \mathbf{H}_i\widetilde{\mathbf{M}}_{0\to i}\mathbf{x}^b$ at each observation time $t_i$.
- Take the expectations of the innovation products $E[(\mathbf{d}^o{}_b)_i(\mathbf{d}^o{}_b)_i{}^T]$ at each observation time $t_i$.
- Calculate $\mathbf{R}^*{}_{(i,i)} = E[(\mathbf{d}^o{}_b)_i(\mathbf{d}^o{}_b)_i{}^T] - \mathbf{H}_i\widetilde{\mathbf{M}}_{0\to i}\mathbf{B}\widetilde{\mathbf{M}}_{0\to i}{}^T\mathbf{H}_i{}^T$.

Compare 4DVar analysis accuracy using $\mathbf{R}^*$ as opposed to $\hat{\mathbf{R}}$.

# Numerical experiments: results

- Method 1: use $\hat{\mathbf{R}}$ in 4DVar
- Method 2: use $\mathbf{R}^*$ in 4DVar

| Variable | Truth | Analysis Method 1 | Error % Method 1 | Analysis Method 2 | Error % Method 2 |
|---|---|---|---|---|---|
| $x$ | -3.4866 | -3.1111 | 10.77 | -3.4829 | 0.11 |
| $y$ | -5.7699 | -5.2994 | 8.15 | -5.7843 | 0.25 |
| $z$ | 18.341 | 18.6500 | 1.68 | 18.3464 | 0.03 |
| $w$ | -10.7175 | -10.8140 | 0.90 | -10.7181 | 0.01 |
| $v$ | -7.1902 | -7.9787 | 10.97 | -7.1928 | 0.04 |

Table: Analysis from Method 1 and Method 2.

- Method 1: 81 iterations, Method 2: 2 iterations.
- Stopping criteria $\frac{\|\mathcal{J}({\mathbf{x}_0}^k)\|}{\|\mathcal{J}({\mathbf{x}_0}^1)\|} < 10^{-3}$.

# Numerical experiments: results

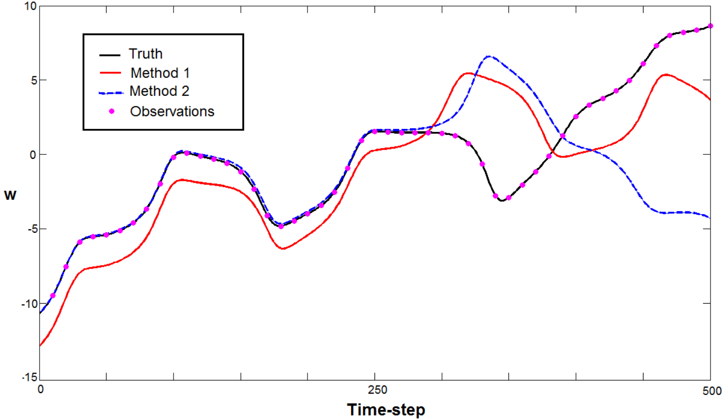


Figure: Comparing analysis trajectory for $w$ over the assimilation window from Method 1: 4DVAR with $\hat{\mathbf{R}}$ and Method 2: 4DVAR with $\mathbf{R}^*$. (Large parameter errors with s.d's of 1 for $\sigma$, $k$ and $\alpha$).

# Numerical experiments: summary

When the model used in 4DVar is erroneous, using $\mathbf{R}^*$ as opposed to $\hat{\mathbf{R}}$ increases the analysis accuracy at the initial time.

Experimental results have shown the increase is most significant when,

- the model error is large,
- the observations are very accurate,
- the background is very inaccurate,
- assimilation window length is increased,
- increased observation frequency.

# Summary

- Derived an expression for the true covariance of the error in the observational cost function term in strong constraint 4DVar in the presence of model error.
- This matrix contains both model error and observation error statistics.
- Developed a method to estimate this combined matrix using diagnostics.
- We have shown using the combined model error and observation error covariance matrix, as opposed to only the observation error covariance matrix, increases the analysis accuracy.
- Application of the method suited to reanalysis, where the objective is to best estimate the analysis at the initial time and start of an assimilation window (not beneficially applicable to long term forecasts).

Thank you for listening

Any questions?