Parameterisation Estimation using Data Assimilation

Matthew Lang
University of Reading

Supervisors: Peter Jan van Leeuwen
Phil Browne

What’s the problem?

- In NWP, parameterisations are used to compensate for errors in the model.
  - Eg. Sub-grid turbulence, radiation
- Errors can be due to:
  - A lack of scientific understanding
  - A lack of computing power
- Large uncertainties in models come from parameterisations
- Improving parameterisations is currently done in an ad-hoc fashion, eg. via parameterisation tuning

How do we help?

- A new method of parameterisation estimation using data assimilation
- State estimation is used to reduce the errors in the model trajectory by combining observations with previously known information.
- We use the differences between a data assimilation trajectory and an analysis forecast
- This gives us information about structures in the errors of the prior model

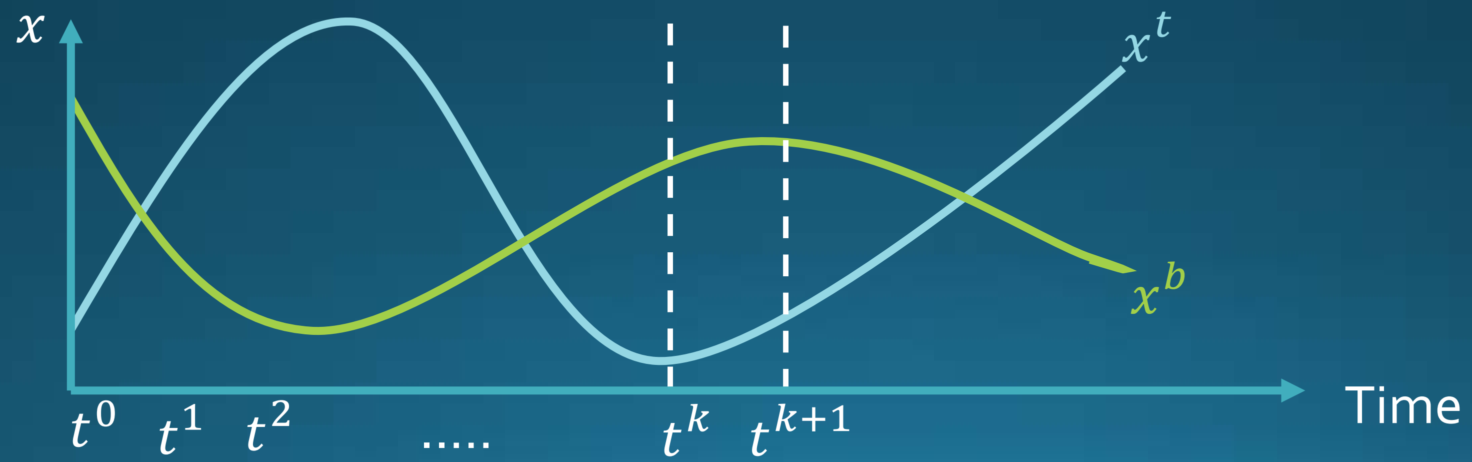
Our method

- We have a true state, x^t , that we wish to approximate, given by:

$$\mathrm{x}_{\mathrm{k}+1}^{t} = \mathrm{f}\left(x_k^t, G^t\left(\alpha^t\right)\right) + \xi_k^t$$

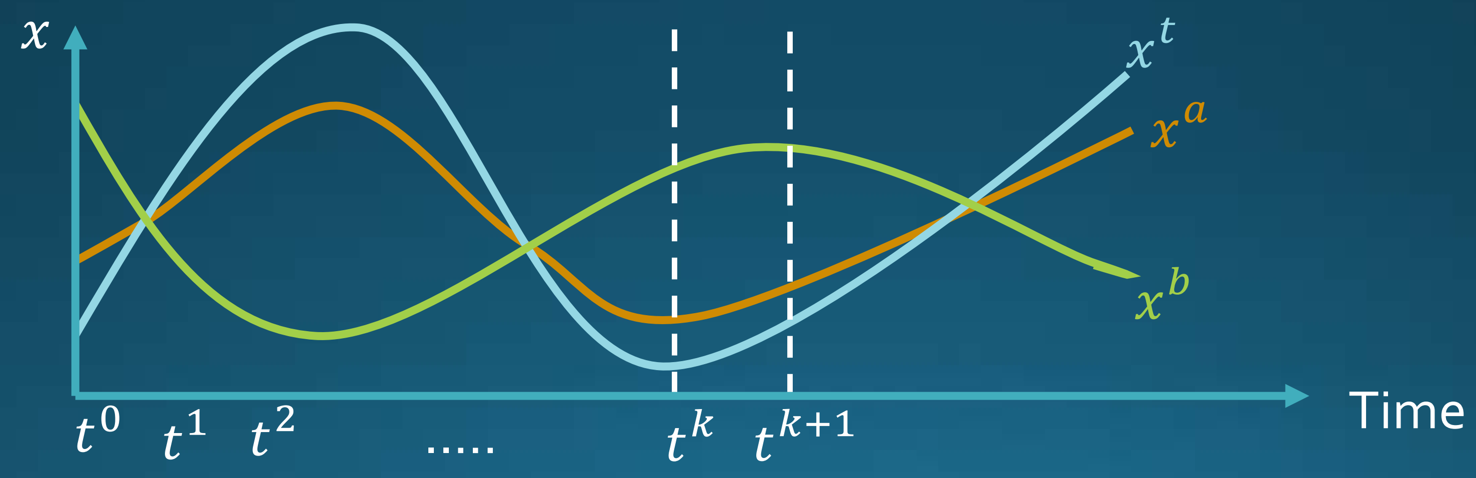
- Our prior estimate of the state, x^b , with the form:

$$\mathrm{x}_{k+1}^{b} = \mathrm{f}\left(\mathrm{x}_k^b, \mathrm{G}^b\left(\alpha^b\right)\right) + \xi_k^b$$



Our method

- When we run our data assimilation scheme, we will get a trajectory of points that are a better representation of the truth

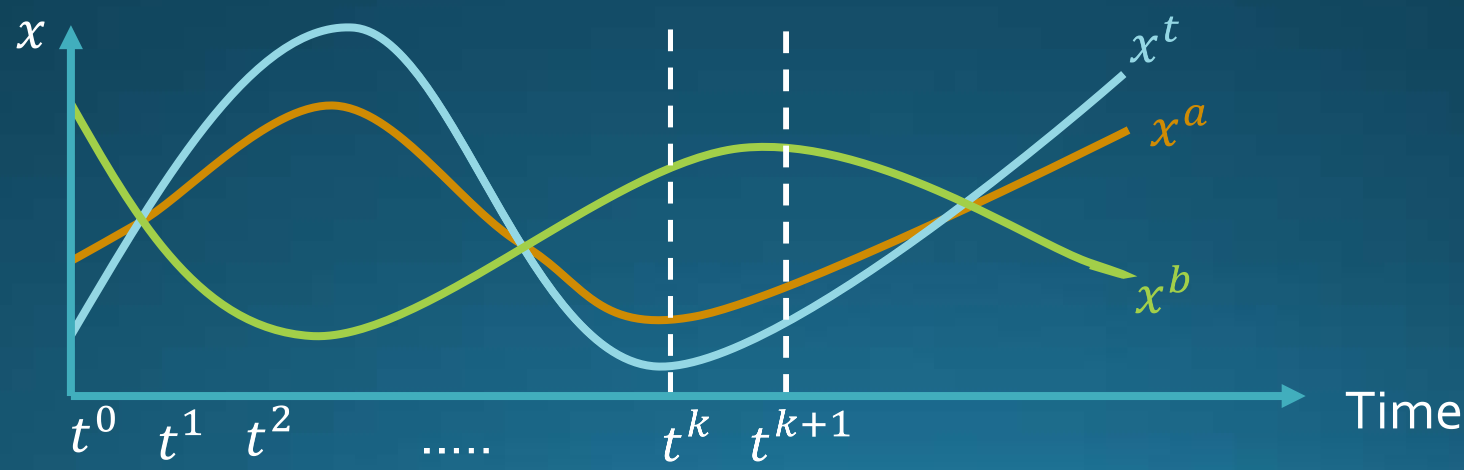


Our method

- Define a new, unknown, parameterisation $G^a(\alpha^a)$ such that the evolution of the data assimilation analysis can be written as:

$$x_{k+1}^a = f(x_k^a, G^a(\alpha^a)) + \xi_k^a$$

- As x^a is more accurate than x^b
  - $G^a(\alpha^a)$ is expected to be more accurate than $G^b(\alpha^b)$



Our method

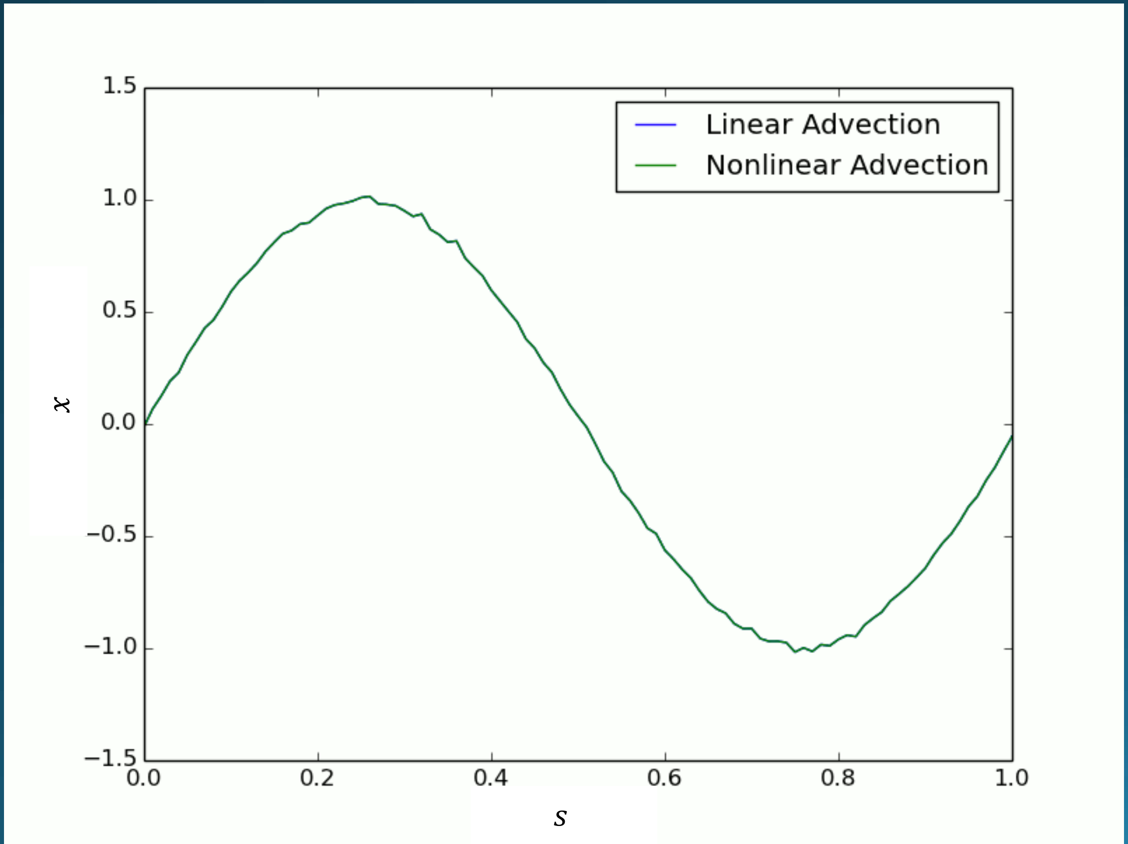
- Zooming into a single time step of the DA trajectory

Our method

- We create a new variable $\tilde{x}_{k+1}$.
  - Is the analysis state evolved one time step using the prior parameterisation $\mathrm{G}^{b}(\alpha^{b})$
  - Represents the analysis forecast
- Compute the differences between x^{a} and $\tilde{x}$ over the whole domain at all timesteps

Motivating example

- This method is applied to estimated the functional difference between a linear advection scheme and a nonlinear advection scheme over [0,1] with periodic boundary conditions



True model

$$\frac{\partial x}{\partial t} + (x+3)\frac{\partial x}{\partial s} = \xi_q$$

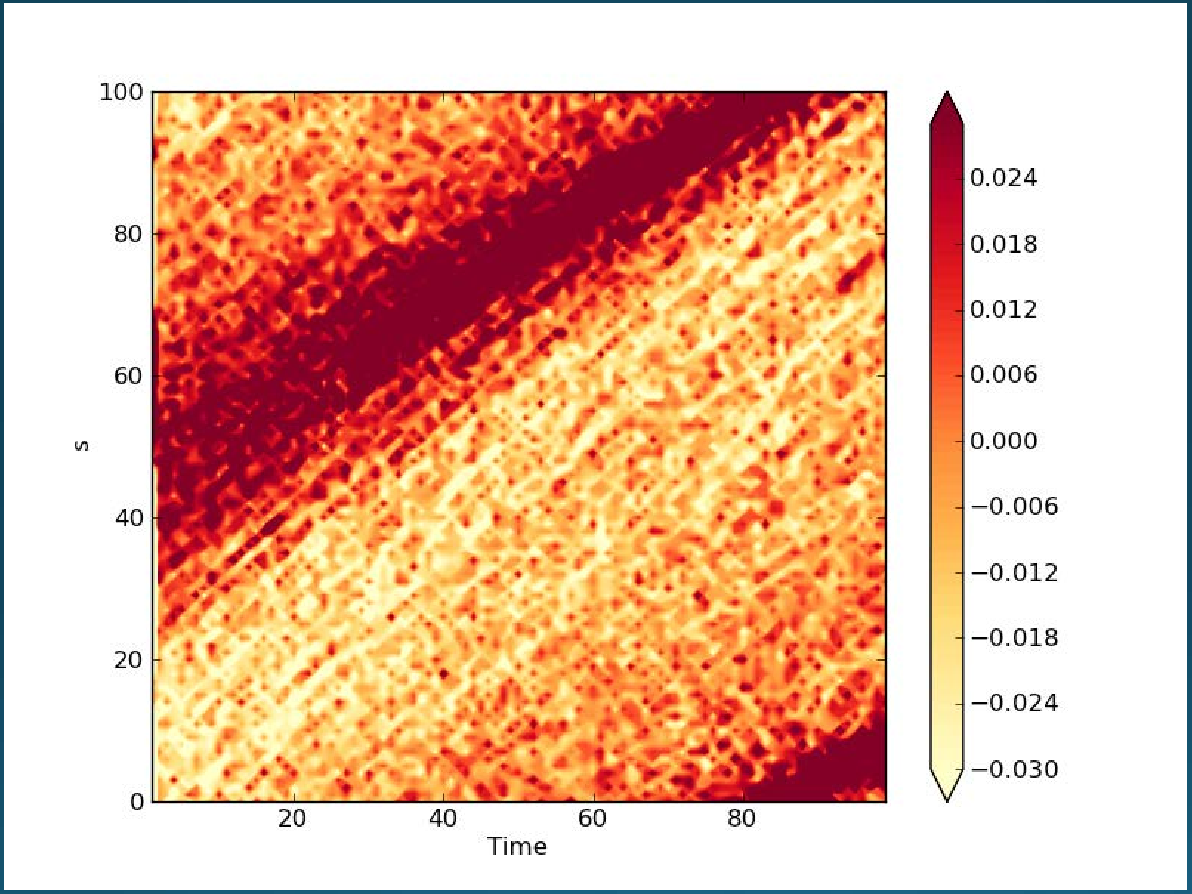
Prior model

$$\frac{\partial x}{\partial t} + \frac{\partial x}{\partial s} = \xi_q$$

Experiment specifications

- For these experiments, we will be using the Ensemble Kalman Smoother (EnKS) with the following specifications:
  - 100 time steps
  - 100 grid points
  - 40 ensemble members
  - Observations are at all times and grid points
  - Model error standard deviation=0.01
  - Observation error standard deviation=0.005
  - Localisation radius = 0.05

Generating $x^a - \tilde{x}$



- The mean (over the ensemble) of $x^a - \tilde{x}$ plotted for each gridpoint and timestep generated by the EnKS using the prior model

Extracting the model error structure

- To extract the model error present, it is required that the structure in $x^a - \tilde{x}$ be determined

- To do this, a test function, $g(x)$, is defined with the form:

$$g(x) = \alpha_0 f_0(x) + \cdots + \alpha_n f_n(x)$$

- To define the test function, expertise regarding the model is required to choose relevant terms that may be the cause of model error

Defining the test function

- It is expected that the functional form of the model error is given by:

$$x\frac{\partial x}{\partial s} + 2\frac{\partial x}{\partial s}$$

- Given prior knowledge of the advection equation, the test function is defined as the following:

$$g(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 \frac{\partial x}{\partial s} + \alpha_4 x \frac{\partial x}{\partial s} + \alpha_5 x^2 \frac{\partial x}{\partial s} + \alpha_6 \frac{\partial^2 x}{\partial s^2} + \alpha_7 x \frac{\partial^2 x}{\partial s^2} + \alpha_8 x^2 \frac{\partial^2 x}{\partial s^2}$$

- This function contains the terms we expect, plus extra terms that indicate our uncertainty about what the true model error is

Analysing test function

- Once an adequate test function is determined, the test function is split into separate sub-functions:

$$g_0(x) = \alpha_0 f_0(x)$$
$$g_1(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x)$$
$$\vdots$$
$$g_n(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x) + \cdots + \alpha_n f_n(x)$$

- Nonlinear regression is used to estimate optimal coefficients for each of the $g_i(x)$ ($i = 0, 1, \ldots, n$) that best fits the data

The Bayesian Information Criterion

- To assess the quality of the terms, the Bayesian Information Criterion (BIC) values are computed for each $g_i(x)$, where

$$BIC = \mathrm{k}\ \log N - 2\log \mathcal{L}$$

$k =$ Number of terms, N=Number of fitting points in the regression, $\mathcal{L}$ is maximised likelihood function

- The BIC represents the trade-off between how well the model fits the data given and the complexity of the model

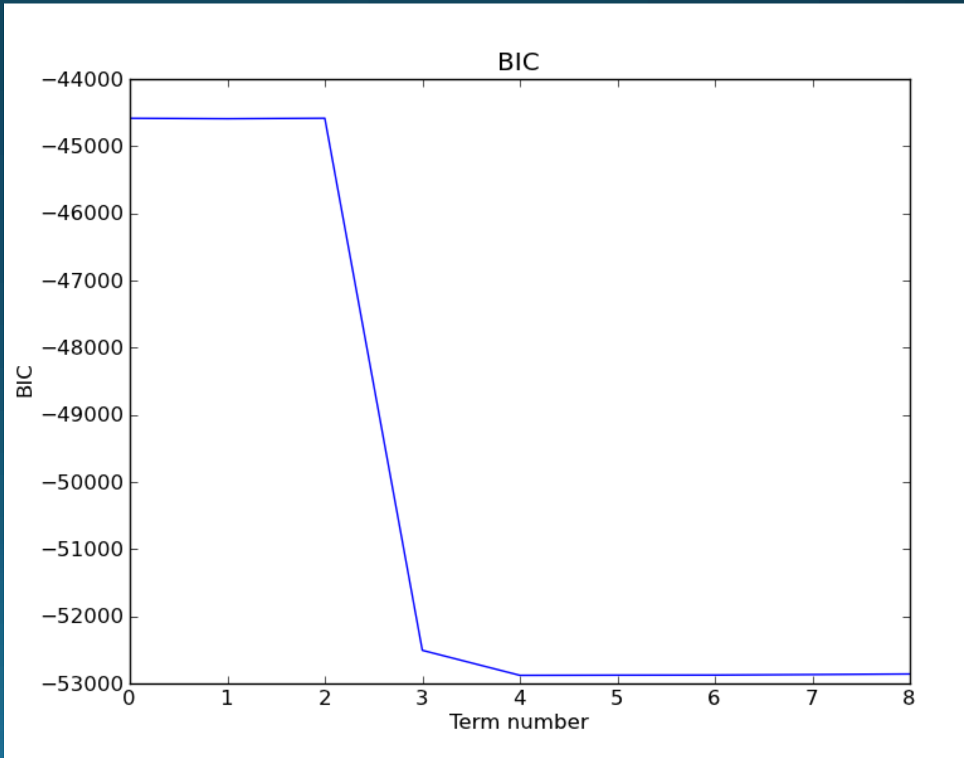
- Smaller values of BIC indicate 'better' models

- The greatest decreases in BIC correspond to the terms with the 'most' information about the structure of the model error

BIC of Functional Estimates obtained

$$g(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 \frac{\partial x}{\partial s} + \alpha_4 x \frac{\partial x}{\partial s} + \alpha_5 x^2 \frac{\partial x}{\partial s} + \alpha_6 \frac{\partial^2 x}{\partial s^2} + \alpha_7 x \frac{\partial^2 x}{\partial s^2} + \alpha_8 x^2 \frac{\partial^2 x}{\partial s^2}$$

- Most information is added by terms 3 and 4 which represent $\frac{\partial x}{\partial s}$ and $x\frac{\partial x}{\partial s}$ respectively
- Test function is reordered based on greatest decreases in BIC and nonlinear regression is performed again

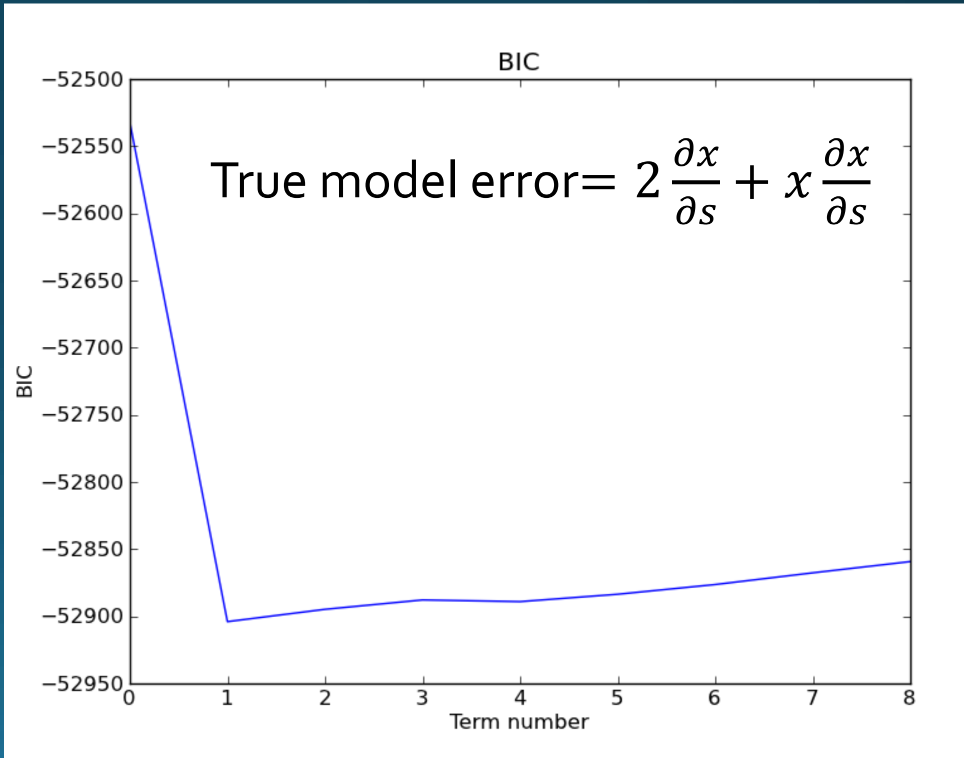


Reordered Functional Estimates obtained

$$h(x) = \alpha_0 \frac{\partial x}{\partial s} + \alpha_1 x \frac{\partial x}{\partial s} + \alpha_2 x + \alpha_3 + \alpha_4 \frac{\partial^2 x}{\partial s^2} + \alpha_5 x^2 \frac{\partial x}{\partial s} + \alpha_6 x \frac{\partial^2 x}{\partial s^2} + \alpha_7 x^2 + \alpha_8 x^2 \frac{\partial^2 x}{\partial s^2}$$

- Minimum occurs after the addition of Term 1 (corresponding to $x \frac{\partial x}{\partial s}$), hence from regression analysis optimal test function is given by:

$$h_1(x) = 2.026 \frac{\partial x}{\partial s} + 0.902 x \frac{\partial x}{\partial s}$$



Remarks and Conclusions

- Parameterisation Estimation Method is shown to work well for the advection model
- Picks out the optimal functional form of the model error
- It can also be used for parameter estimation
- The method is very dependent upon how well the DA scheme used performs
- Current work is looking into sensitivity of method to data assimilation specifications (eg. observation densities)

Any questions?

Results: Functional estimates of $x^a - \tilde{x}$

$g_0(x) = -0.043$

$g_1(x) = -0.045 + 0.754x$

$g_2(x) = -0.284 + 0.740x + 0.499x^2$

$g_3(x) = -0.268 + 0.040x + 0.470x^2 + 2.002\frac{\partial x}{\partial s}$

$g_4(x) = -0.280 + 0.028x + 0.178x^2 + 2.026\frac{\partial x}{\partial s} + 0.905x\frac{\partial x}{\partial s}$

$g_5(x) = -0.282 + 0.065x + 0.183x^2 + 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s} - 0.220x^2\frac{\partial x}{\partial s}$

$g_6(x) = -0.291 + 0.141x + 0.211x^2 + 2.084\frac{\partial x}{\partial s} + 0.881x\frac{\partial x}{\partial s} - 0.172x^2\frac{\partial x}{\partial s} + 0.001\frac{\partial^2 x}{\partial s^2}$

$g_7(x) = -0.288 + 0.135x + 0.267x^2 + 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s} - 0.182x^2\frac{\partial x}{\partial s} + 0.001\frac{\partial^2 x}{\partial s^2} + 0.001x\frac{\partial^2 x}{\partial s^2}$

$g_8(x) = -0.290 + 0.117x + 0.273x^2 + 2.087\frac{\partial x}{\partial s} + 0.907x\frac{\partial x}{\partial s} - 0.216x^2\frac{\partial x}{\partial s} + 0.002\frac{\partial^2 x}{\partial s^2} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.001x^2\frac{\partial^2 x}{\partial s^2}$

Results: Functional estimates of $x^a - \tilde{x}$

$$g_0(x) = -0.043$$

$$g_1(x) = -0.045 + 0.754x$$

$$g_2(x) = -0.284 + 0.740x + 0.499x^2$$

$$g_3(x) = -0.268 + 0.040x + 0.470x^2 + 2.002\frac{\partial x}{\partial s}$$

$$g_4(x) = -0.280 + 0.028x + 0.178x^2 \boldsymbol{+ 2.026\frac{\partial x}{\partial s} + 0.905x\frac{\partial x}{\partial s}}$$

$$g_5(x) = -0.282 + 0.065x + 0.183x^2 \boldsymbol{+ 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s}} - 0.220x^2\frac{\partial x}{\partial s}$$

$$g_6(x) = -0.291 + 0.141x + 0.211x^2 \boldsymbol{+ 2.084\frac{\partial x}{\partial s} + 0.881x\frac{\partial x}{\partial s}} - 0.172x^2\frac{\partial x}{\partial s} + 0.001\frac{\partial^2 x}{\partial s^2}$$

$$g_7(x) = -0.288 + 0.135x + 0.267x^2 \boldsymbol{+ 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s}} - 0.182x^2\frac{\partial x}{\partial s} + 0.001\frac{\partial^2 x}{\partial s^2} + 0.001x\frac{\partial^2 x}{\partial s^2}$$

$$g_8(x) = -0.290 + 0.117x + 0.273x^2 \boldsymbol{+ 2.087\frac{\partial x}{\partial s} + 0.907x\frac{\partial x}{\partial s}} - 0.216x^2\frac{\partial x}{\partial s} + 0.002\frac{\partial^2 x}{\partial s^2} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.001x^2\frac{\partial^2 x}{\partial s^2}$$

Results:Reordered functional estimates

$$h_0(x) = 2.002\frac{\partial x}{\partial s}$$

$$h_1(x) = 2.026\frac{\partial x}{\partial s} + 0.902x\frac{\partial x}{\partial s}$$

$$h_2(x) = 2.026\frac{\partial x}{\partial s} + 0.902x\frac{\partial x}{\partial s} + 0.0319x$$

$$h_3(x) = 2.026\frac{\partial x}{\partial s} + 0.906x\frac{\partial x}{\partial s} + 0.033x - 0.195$$

$$h_4(x) = 2.046\frac{\partial x}{\partial s} + 0.875x\frac{\partial x}{\partial s} + 0.130x - 0.190 + 0.001\frac{\partial^2 x}{\partial s^2}$$

$$h_5(x) = 2.084\frac{\partial x}{\partial s} + 0.883x\frac{\partial x}{\partial s} + 0.147x - 0.190 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.172x^2\frac{\partial x}{\partial s}$$

$$h_6(x) = 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s} + 0.142x - 0.162 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.181x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2}$$

$$h_7(x) = 2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s} + 0.135x - 0.288 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.182x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.267x^2$$

$$h_8(x) = 2.087\frac{\partial x}{\partial s} + 0.908x\frac{\partial x}{\partial s} + 0.117x - 0.290 + 0.002\frac{\partial^2 x}{\partial s^2} - 0.216x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.273x^2 - 0.001x^2\frac{\partial^2 x}{\partial s^2}$$

Results:Reordered functional estimates

$$h_0(x) = 2.002\frac{\partial x}{\partial s}$$

$$h_1(x) = \boldsymbol{2.026\frac{\partial x}{\partial s} + 0.902x\frac{\partial x}{\partial s}}$$

$$h_2(x) = \boldsymbol{2.026\frac{\partial x}{\partial s} + 0.902x\frac{\partial x}{\partial s}} + 0.0319x$$

$$h_3(x) = \boldsymbol{2.026\frac{\partial x}{\partial s} + 0.906x\frac{\partial x}{\partial s}} + 0.033x - 0.195$$

$$h_4(x) = \boldsymbol{2.046\frac{\partial x}{\partial s} + 0.875x\frac{\partial x}{\partial s}} + 0.130x - 0.190 + 0.001\frac{\partial^2 x}{\partial s^2}$$

$$h_5(x) = \boldsymbol{2.084\frac{\partial x}{\partial s} + 0.883x\frac{\partial x}{\partial s}} + 0.147x - 0.190 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.172x^2\frac{\partial x}{\partial s}$$

$$h_6(x) = \boldsymbol{2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s}} + 0.142x - 0.162 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.181x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2}$$

$$h_7(x) = \boldsymbol{2.079\frac{\partial x}{\partial s} + 0.910x\frac{\partial x}{\partial s}} + 0.135x - 0.288 + 0.001\frac{\partial^2 x}{\partial s^2} - 0.182x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.267x^2$$

$$h_8(x) = \boldsymbol{2.087\frac{\partial x}{\partial s} + 0.908x\frac{\partial x}{\partial s}} + 0.117x - 0.290 + 0.002\frac{\partial^2 x}{\partial s^2} - 0.216x^2\frac{\partial x}{\partial s} + 0.001x\frac{\partial^2 x}{\partial s^2} + 0.273x^2 - 0.001x^2\frac{\partial^2 x}{\partial s^2}$$