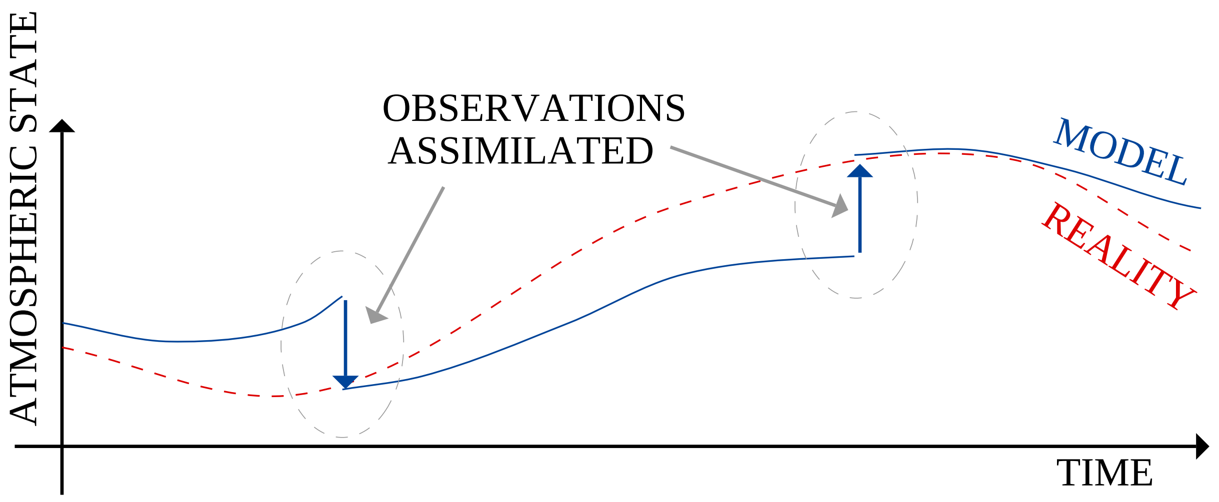
4. DATA ASSIMILATION FUNDAMENTALS

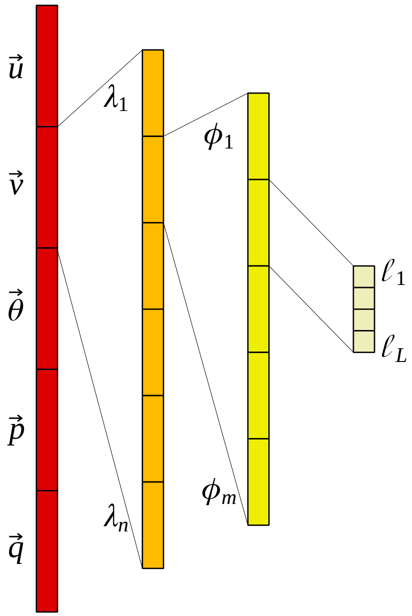
... [the atmosphere] "is a chaotic system in which errors introduced into the system can grow with time ... As a consequence, data assimilation is a struggle between chaotic destruction of knowledge and its restoration by new observations."

Leith (1993)



The vector notation for fields and data and the need for an a-priori

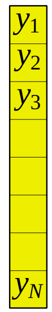
The 'state vector', $\vec{x}$



$\vec{u}$ zonal wind field
$\vec{v}$ meridional wind field
$\vec{\theta}$ potential temperature
$\vec{p}$ pressure
$\vec{q}$ specific humidity
λ longitude
ϕ latitude
ℓ vertical level

NWP models: $> \; 10^7$ elements $(5 \times n \times m \times L)$

The 'observation vector', $\vec{y}$



NWP models: $\sim 10^6$ elements

The 'forward model', $\vec{h}$

$$\vec{y} = \vec{h}[\vec{x}] + \vec{\varepsilon}$$

- No. of obs. << No. of (unknown) elements in $\vec{x}$.
- This is an under-constrained (and inexact) inverse problem.
- Need to fill-in the missing information with prior knowledge.

An 'a-priori' state (a.k.a. 'first guess', 'background', 'forecast') is needed to make the assimilation problem well posed.

Vectors and matrices

Vector/matrix notation is a powerful and compact way of dealing with large volumes of data.

•A matrix operator acts on an input vector to give an output vector, e.g.

$$\vec{x}^{(2)} = \mathbf{A}\vec{x}^{(1)} \qquad \begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \ldots \\ x_N^{(2)} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1N} \\ A_{21} & A_{22} & \ldots & A_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ A_{N1} & A_{N2} & \ldots & A_{NN} \end{pmatrix} \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \ldots \\ x_N^{(1)} \end{pmatrix}$$

•Matrix products do not commute in general, e.g.

$$\vec{x}^{(3)} = \mathbf{ABC}\vec{x}^{(1)} \neq \mathbf{CBA}\vec{x}^{(1)}$$

•Some matrices can be inverted (must be 'square' and non-singluar), e.g.

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \ldots \\ x_N^{(1)} \end{pmatrix} = \begin{pmatrix} (A^{-1})_{11} & (A^{-1})_{12} & \ldots & (A^{-1})_{1N} \\ (A^{-1})_{21} & (A^{-1})_{22} & \ldots & (A^{-1})_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ (A^{-1})_{N1} & (A^{-1})_{N2} & \ldots & (A^{-1})_{NN} \end{pmatrix} \begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \ldots \\ x_N^{(2)} \end{pmatrix}$$

(matrices are complicated to invert, ie $(A^{-1})_{ij} \neq A_{ij}^{-1}$.)

•The matrix transpose make rows into columns and columns into rows (also for vectors), e.g.

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1N} \\ A_{21} & A_{22} & \ldots & A_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ A_{N1} & A_{N2} & \ldots & A_{NN} \end{pmatrix}, \mathbf{A}^T = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{N1} \\ A_{12} & A_{22} & \ldots & A_{N2} \\ \ldots & \ldots & \ldots & \ldots \\ A_{1N} & A_{2N} & \ldots & A_{NN} \end{pmatrix}$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_N \end{pmatrix}, \qquad \vec{x}^T = (x_1,\ x_2,\ \ldots\ x_N)$$

•The inner product ('scalar' or 'dot' product), e.g.

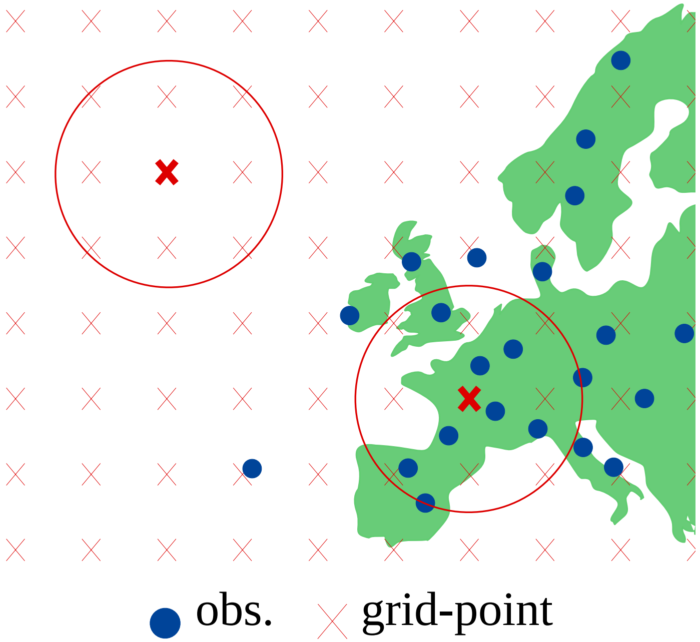
$$\vec{x}^{(2)T}\vec{x}^{(1)} = x_1^{(2)}x_1^{(1)} + x_2^{(2)}x_2^{(1)} + \ldots + x_N^{(2)}x_N^{(1)} = \text{scalar}$$

•The outer product (a matrix), e.g.

$$\vec{x}^{(2)}\vec{x}^{(1)T} = \begin{pmatrix} x_1^{(2)}x_1^{(1)} & x_1^{(2)}x_2^{(1)} & \ldots & x_1^{(2)}x_N^{(1)} \\ x_2^{(2)}x_1^{(1)} & x_2^{(2)}x_2^{(1)} & \ldots & x_2^{(2)}x_N^{(1)} \\ \ldots & \ldots & \ldots & \ldots \\ x_N^{(2)}x_1^{(1)} & x_N^{(2)}x_2^{(1)} & \ldots & x_N^{(2)}x_N^{(1)} \end{pmatrix}$$

Early data assimilation ("objective analysis")
The method of "successive corrections"

Bergthorsson & Doos (1955), Cressman (1959)



- Analysis is a linear combination of nearby observations, and an a-priori.

- Analysis → obs. (obs-rich regions).
- Analysis → a-priori (obs-poor regions).

✔Simple scheme to develop.
✔Computationally cheap.
✔Use a-priori in absence of observations.

✘Poor account of error statistics of obs and a-priori.
✘Direct observations only.
✘Anomalous spreading of obs. information.
✘No multivariate relations (e.g. geostrophy).

'Optimal' Interpolation (OI)

Introduced in the 1970s - a more powerful formulation of data assimilation

$$\vec{x}_A = \vec{x}_B + \mathbf{K}(\vec{y} - \vec{h}[\vec{x}_B])$$

- **K** is a rectangular matrix operator (the 'gain matrix').
  - **K** depends upon the error covariance matrices **B** and **R**, and linearization **H**.
  - $\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}$. The Best Linear Unbiased Estimator (BLUE).
- **R** : observation error covariance matrix.
  - Describes the error statistics of the observations (see later).
- **B** : background error covariance matrix.
  - Describes the error statistics of the a-priori state (see later).
- **H** : linearized observation operator.
  - $\vec{h}[\vec{x}_B + \delta\vec{x}] \approx \vec{h}[\vec{x}_B] + \mathbf{H}\delta\vec{x}$.

✔Account taken of a-priori and obs. error statistics.
✔Allows assimilation of some indirect obs.
✔Use a-priori in absence of obs.
✔Works as an inverse model.

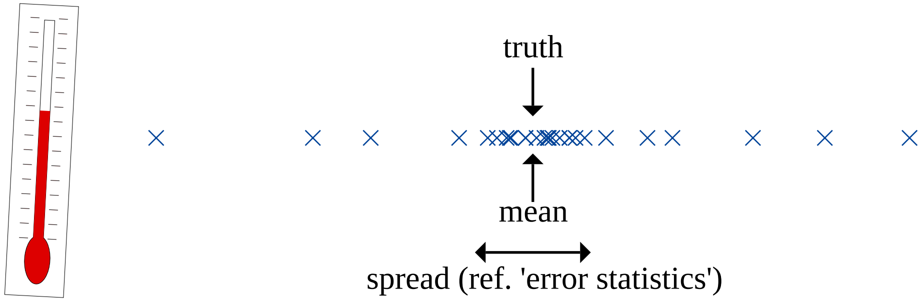
✘Too expensive for single global solution.
✘Difficult to know **B**.
✘No consistency with the equations of motion.

Types of errors

1. Random errors

Data	Arising from
Obs.	Noise
Forecast	Stochastic processes in model, init. conds.
Assim.	Input data

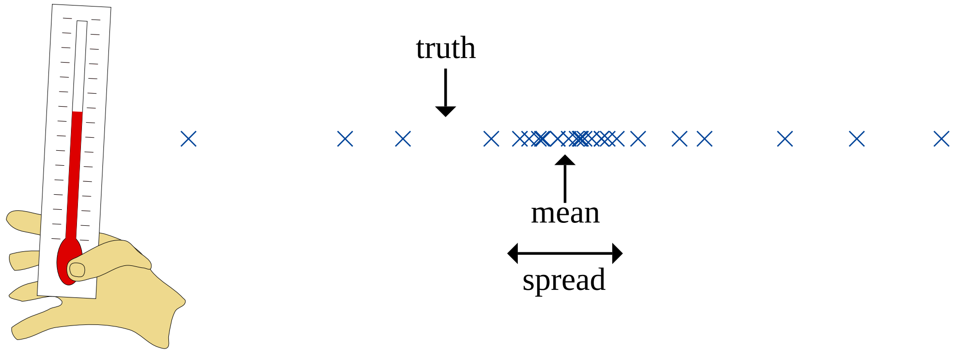
E.g. repeated measurement of temperature:



2. Systematic errors

Data	Arising from
Obs.	E.g. reading errors
Forecast	Model formulation, init. conds.
Assim.	Input data, formulation

E.g. As before but with biased thermometer:

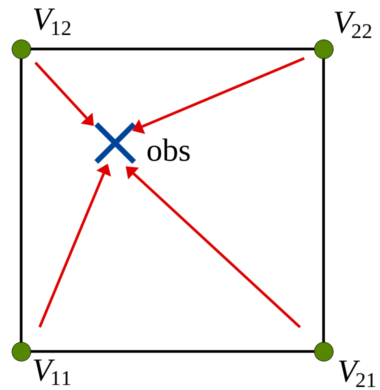


Biases should be corrected where possible.

3. Representativeness error

Data	Arising from
Assim.	Unresolved variability

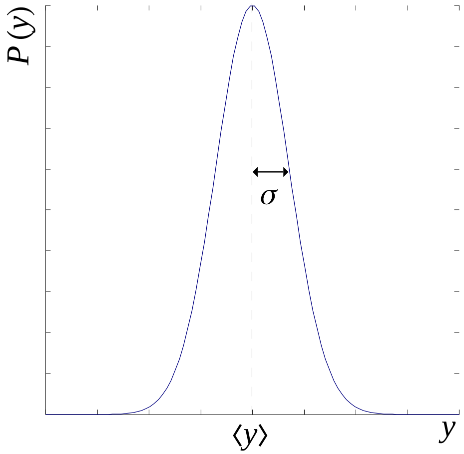
E.g. Interpolation of model grid values to location of observation (forward operator $\vec{h}[\vec{x}]$):



Probability distribution functions

- Error statistics are described by a probability density function (PDF).
- PDFs of random and representativeness errors are often expressed together.
- They are often approximated by the normal (Gaussian) distribution.

A scalar (ie a single piece of information), y



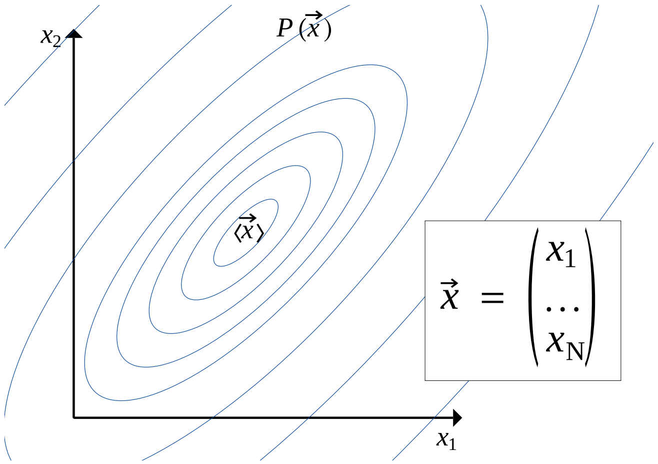
$$P(y) \propto \exp - \frac{(y - \langle y \rangle)^2}{2\sigma^2}$$

Mean : $\langle y \rangle$

Variance : σ^2

(Std. dev. : σ)

A vector (multiple pieces of information), $\vec{x}$



$$\vec{x} = \begin{pmatrix} x_1 \\ \dots \\ x_N \end{pmatrix}$$

$$P(\vec{x}) \propto \exp\left(-\frac{1}{2}(\vec{x} - \langle \vec{x} \rangle)\,\mathbf{B}^{-1}(\vec{x} - \langle \vec{x} \rangle)\right)$$

Mean : $\langle \vec{x} \rangle$

$$\text{Covariance : } \mathbf{B} = \begin{pmatrix} \langle \delta x_1^2 \rangle & \langle \delta x_1 \delta x_2 \rangle & \dots & \langle \delta x_1 \delta x_N \rangle \\ \langle \delta x_2 \delta x_1 \rangle & \langle \delta x_2^2 \rangle & \dots & \langle \delta x_2 \delta x_N \rangle \\ \dots & \dots & \dots & \dots \\ \langle \delta x_N \delta x_1 \rangle & \langle \delta x_N \delta x_2 \rangle & \dots & \langle \delta x_N^2 \rangle \end{pmatrix}$$

$$= \langle \delta\vec{x}\delta\vec{x}^T \rangle \quad \text{where} \quad \delta\vec{x} = \vec{x} - \langle \vec{x} \rangle$$

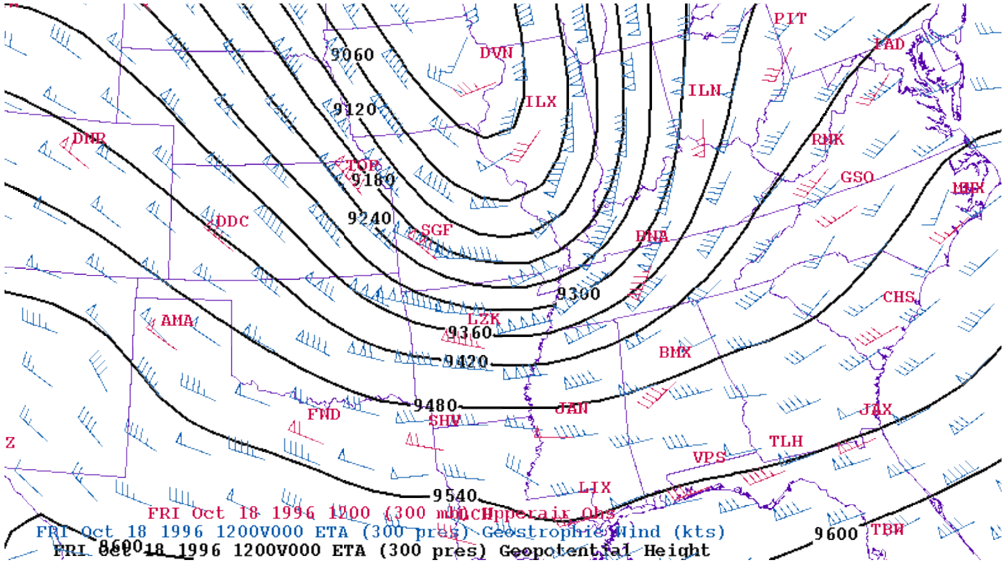
Example of covariances: forecast as the a-priori

- $\vec{x}_B$ is a forecast and so the equations of motion will influence strongly the covariance patterns.

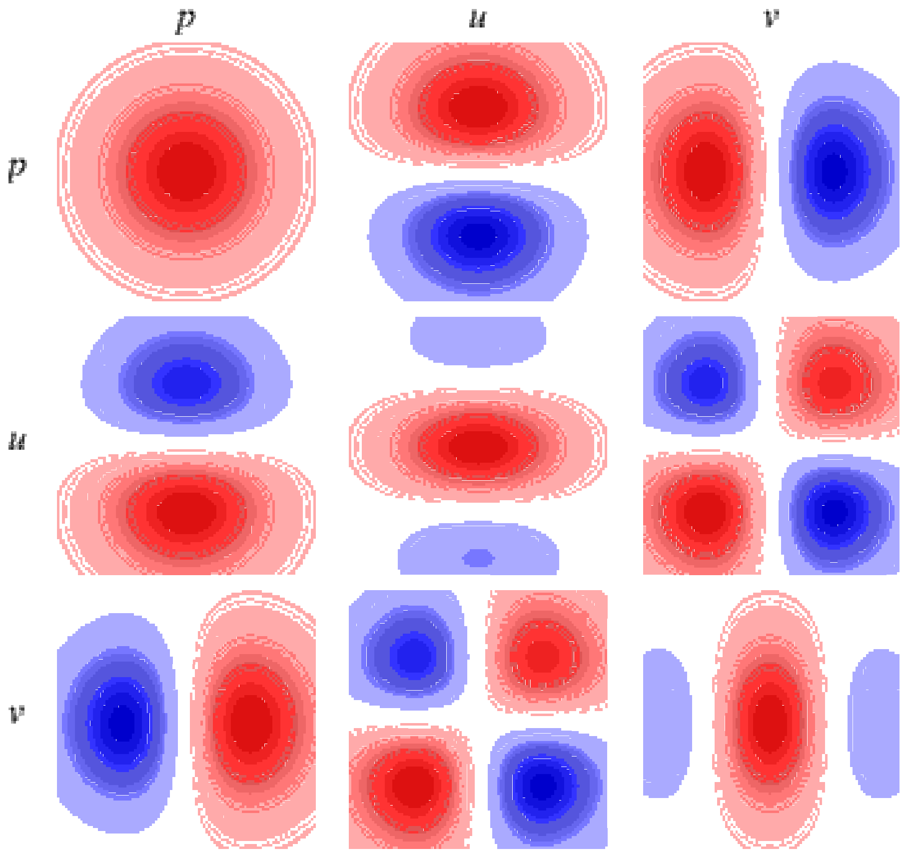
<u>Example</u>: Geostrophic error covariances

Geostrophic balance:

$$v = -\frac{1}{\rho f}\frac{\partial p}{\partial x} \qquad u = \frac{1}{\rho f}\frac{\partial p}{\partial y}$$



Courtesy, Univ. of Washington



Pressure-pressure covariances assumption:

$$\langle \delta p_i \delta p_j \rangle = \sigma^2 \exp -\frac{r_{ij}^2}{2L^2} \qquad \sqrt{2}L \sim 750 \text{ km}$$

Variational data assimilation

The 'method of least squares' - simple version

$$J(\vec{x}) = (\vec{x} - \vec{x}_B)^2 + (\vec{y} - \vec{h}[\vec{x}])^2$$

J : cost function (a scalar)

$\vec{x}_B$: a-priori (background) state

$\vec{y}$: observations

$\vec{h}[\vec{x}]$: observation operator (forward model)

$\vec{x}$: variable $\quad \vec{x}_A = \vec{x}|_{\min J} =$ "analysis"

Carl Fredrich Gauss 1777-1855

The 'method of least squares' - considering error statistics

$$J(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1} (\vec{x} - \vec{x}_B) + \frac{1}{2}(\vec{y} - \vec{h}[\vec{x}])^T \mathbf{R}^{-1} (\vec{y} - \vec{h}[\vec{x}])$$

$\mathbf{B}$: background error covariance matrix

$\mathbf{R}$: observation error covariance matrix

- This is the form used in operational weather forecasting, deriving satellite retrievals, etc.
- Non-Euclidean L_2 norm.
- Assumes perfect forward model, unbiased data.
- This is consistent with a Gaussian model of error statistics (next slide).
- 'Var.' is efficient enough to solve the global problem.

The Bayesian view of data assimilation

Bayes' Theorem

Rev. Thomas Bayes 1702-1761

$$\left.\begin{aligned} P(\vec{y},\vec{x}) &= P(\vec{x}\,|\,\vec{y})P(\vec{y}) \\ P(\vec{x},\vec{y}) &= P(\vec{y}\,|\,\vec{x})P(\vec{x}) \end{aligned}\right\} \quad P(\vec{x}\,|\,\vec{y}) = \frac{P(\vec{y}\,|\,\vec{x})P(\vec{x})}{P(\vec{y})}$$

$$\propto P(\vec{x})P(\vec{y}\,|\,\vec{x})$$

$$P(\vec{x}\,|\,\vec{y}) \propto \exp\left(-\frac{1}{2}(\vec{x}-\vec{x}_B)^T\mathbf{B}^{-1}(\vec{x}-\vec{x}_B)\right)\exp\left(-\frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T\mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})\right)$$

$$\propto \exp-\left(\frac{1}{2}(\vec{x}-\vec{x}_B)^T\mathbf{B}^{-1}(\vec{x}-\vec{x}_B)+\frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T\mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})\right)$$

Maximum likelihood $\Rightarrow$ Minimum penalty, J

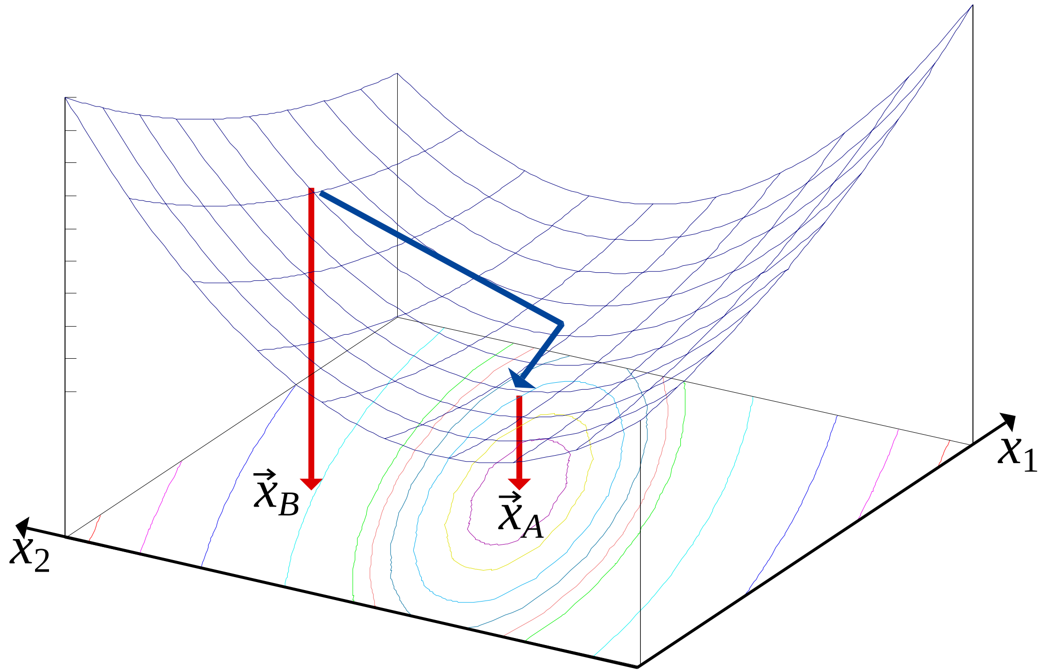
$$J[\vec{x}] = \frac{1}{2}(\vec{x}-\vec{x}_B)^T\mathbf{B}^{-1}(\vec{x}-\vec{x}_B)+\frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T\mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})$$

$$\vec{x}_A = \vec{x}\big|_{\min J} = \text{"analysis"}$$

Minimising the cost function

The problem reduces to a (badly conditioned) optimisation problem in 10^7-dimensional phase space.

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \frac{1}{2}(\vec{h}[\vec{x}] - \vec{y})^T \mathbf{R}^{-1}(\vec{h}[\vec{x}] - \vec{y})$$



- Descent algorithms minimize J iteratively.
- They need the local gradient, $\nabla_{\vec{x}} J$ of the cost function at each iteration.
- The adjoint method is used to compute the adjoint.
- The curvature^{-1} (a.k.a. inverse Hessian, $(\nabla^2_{\vec{x}} J)^{-1}$) at $\vec{x}_A$ indicates the error statistics of the analysis.
    - A very badly conditioned problem.

Algebraic minimization of the cost function

Under simplified conditions the cost function can be minimized algebraically.

Assume that the linearization of the forward model is reasonable

$$\vec{h}[\vec{x}] \approx \vec{h}[\vec{x}_B] + \mathbf{H}(\vec{x} - \vec{x}_B)$$

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \frac{1}{2}(\mathbf{H}(\vec{x} - \vec{x}_B) - (\vec{y} - \vec{h}[\vec{x}_B]))^T \mathbf{R}^{-1}(\mathbf{H}(\vec{x} - \vec{x}_B) - (\vec{y} - \vec{h}[\vec{x}_B]))$$

1. Calculate the gradient vector

$$\nabla_{\vec{x}} J = \begin{pmatrix} \partial J / \partial x_1 \\ \partial J / \partial x_2 \\ \\ \partial J / \partial x_N \end{pmatrix} = \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \mathbf{H}^T \mathbf{R}^{-1}(\vec{h}[\vec{x}] - \vec{y})$$

2. The special $\vec{x}$ that has zero gradient minimizes J (this cost function is quadratic and convex)

$$\nabla_{\vec{x}} J\big|_{x_A} = 0$$

$$\vec{x}_A = \vec{x}_B + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

$$= \vec{x}_B + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

This is the OI formula with the BLUE!

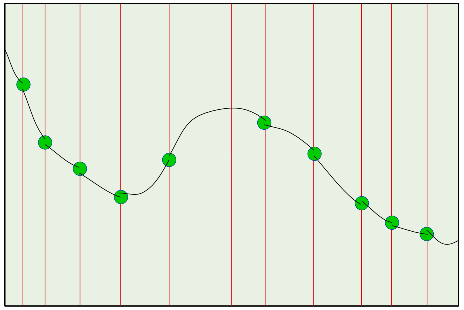
Types of data assimilation

Sequential data assimilation methods

- Data assimilation performed at each batch of observations.
- Model forecast made between batches (for background).
- E.g. OI, KF*, EnKF*, etc.

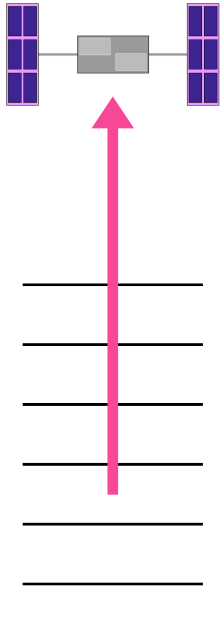
✓Explicit formula used for analysis.

✗Very expensive.



1d-Var

- Data assimilation performed for vertical profile only, where satellite makes observations.
- Used as a 'pre-main-assimilation' step to produce vertical profiles of model quantities (retrievals) from satellite radiances.



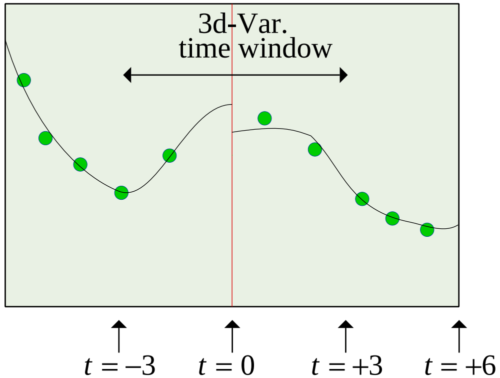
* Kalman Filter (KF) and Ensemble Kalman Filter (EnKF).

3d-Var

- Data assimilation performed every 6 hours.
- 6 hour model forecast between analysis times (for background).
- Adequate for re-analysis.

✓Relatively cheap.

✗Observations within ±3 hours are not at analysis time†.
✗No dynamical constraint used.

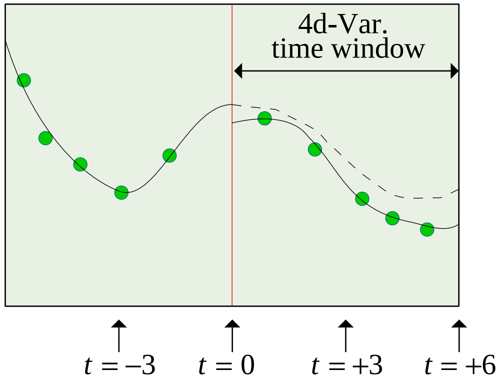


4d-Var

- Data assimilation performed every 6 (12) hours.
- 6 (12) hour model forecast between analysis times (for background).
- Used (e.g.) by ECMWF and Met Office for operational weather forecasting.

✓Model used as a dynamical constraint.
✓Observations are compared to the model trajectory at the correct time.

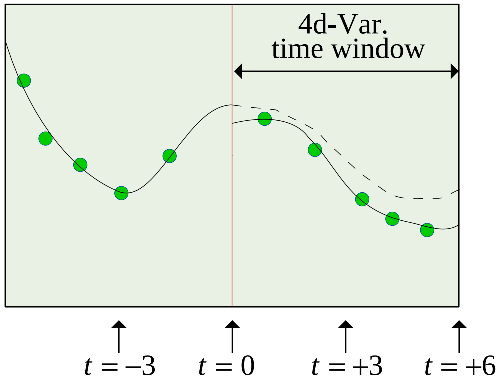
✗Perfect model assumption‡.
✗Expensive, but not unfeasible.

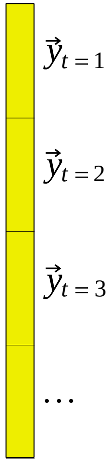


† 3dFGAT '3d First Guess at Time' is half-way between 3d and 4d-Var.

‡ 'Strong constraint' - it is possible to use 4d-Var with a model under the 'weak-constraint' formulation.

The 4d-Var cost function





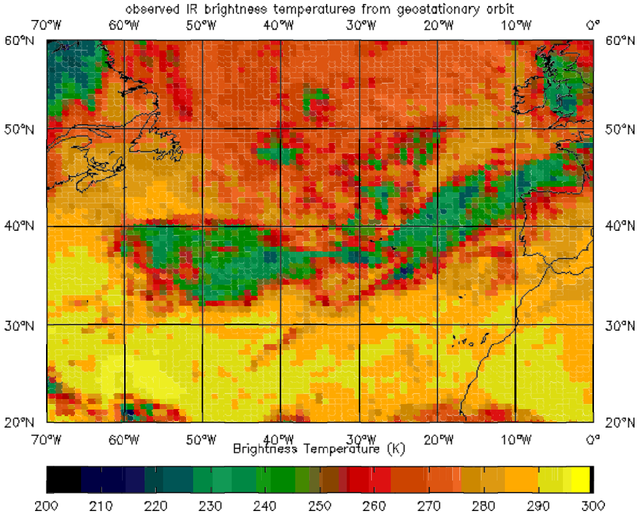
$$J[\vec{x}_0] = \frac{1}{2}(\vec{x}_0 - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x}_0 - \vec{x}_B) + \frac{1}{2}\sum_t (\vec{h}_t[\vec{x}_t] - \vec{y}_t)^T \mathbf{R}_t^{-1}(\vec{h}_t[\vec{x}_t] - \vec{y}_t)$$

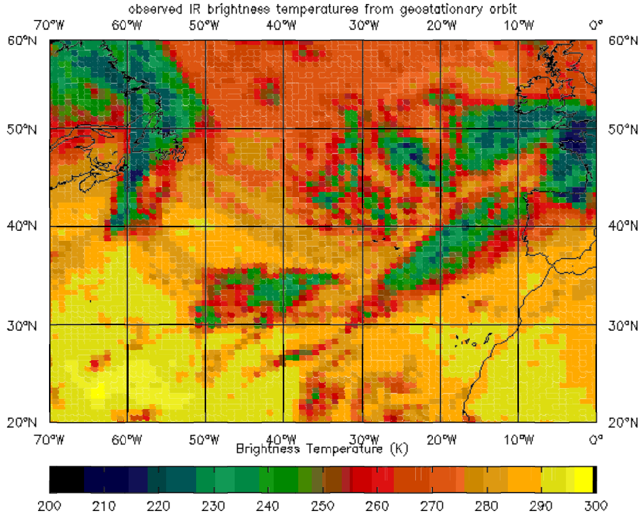
- The observation vector comprises subvectors, $\vec{y}_t$ for time t.
- The observation operator $\vec{h}_t$ acts on model state $\vec{x}_t$.
- Vary $\vec{x}_0$ in the minimization - the state at the start of the 4d-Var. cycle.
- Future states in the cycle are computed with the forecast model, $\vec{x}_t = \vec{M}_{t \leftarrow 0}\vec{x}_0$.
- Forward model is the composite operator $\vec{h}_t M_{t \leftarrow 0}$.
- Important issues:
  - Tangent linear model (and its adjoint) needed and can be difficult to find.
  - Forecast model can be highly non-linear (e.g. sensitive dependence in model's convection scheme - on/off 'switches').

Assimilation of sequences of satellite images in 4d-Var

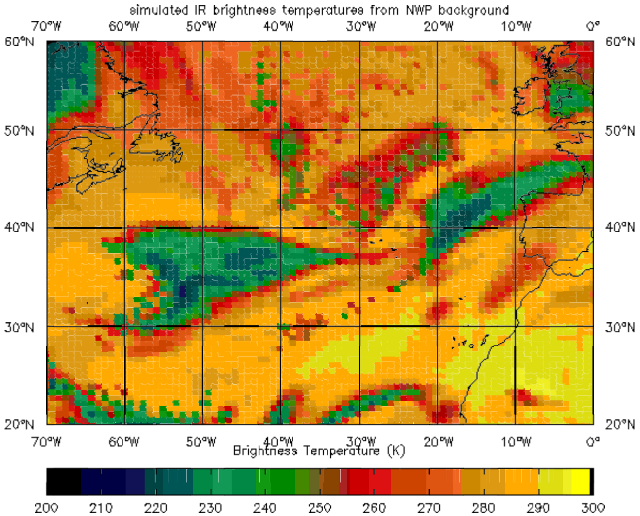
(Courtesy Samantha Pullen, Met Office)

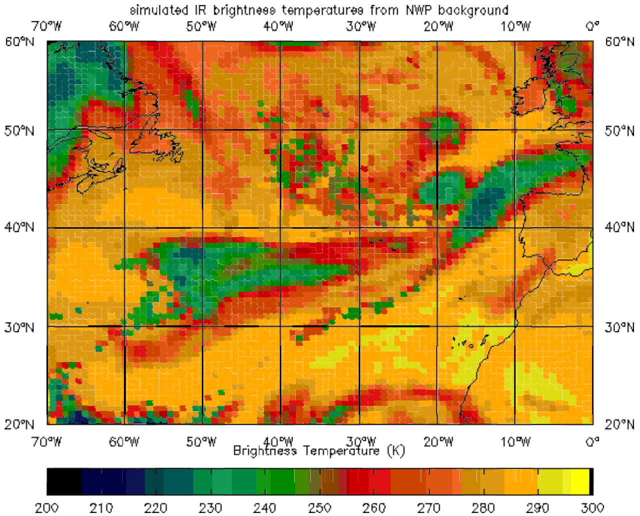
Sequence of observed brightness temperatures

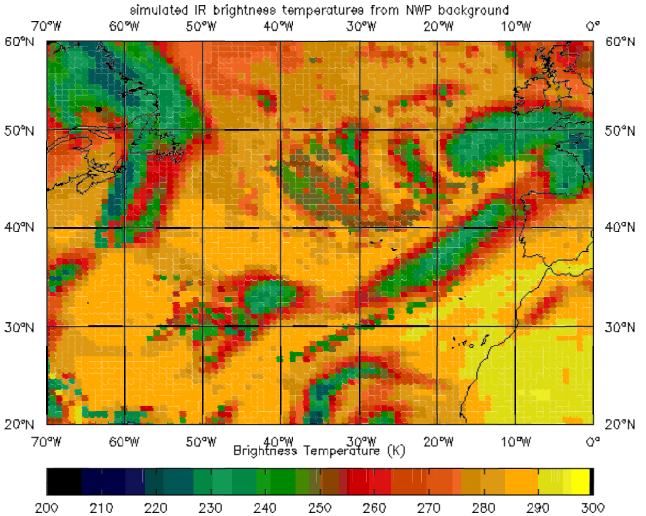




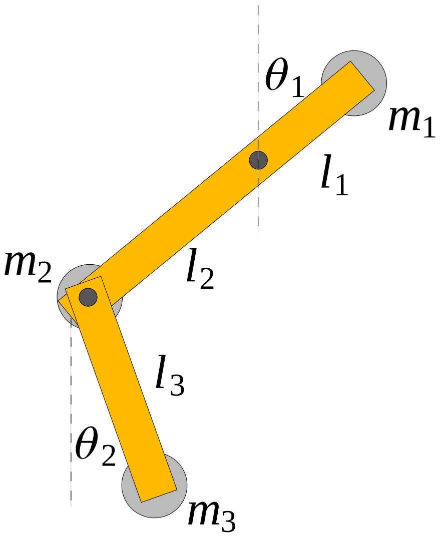
Sequence of simulated brightness temperatures







'4d-Var.' demonstration with a double pendulum



$$\vec{x} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix}$$

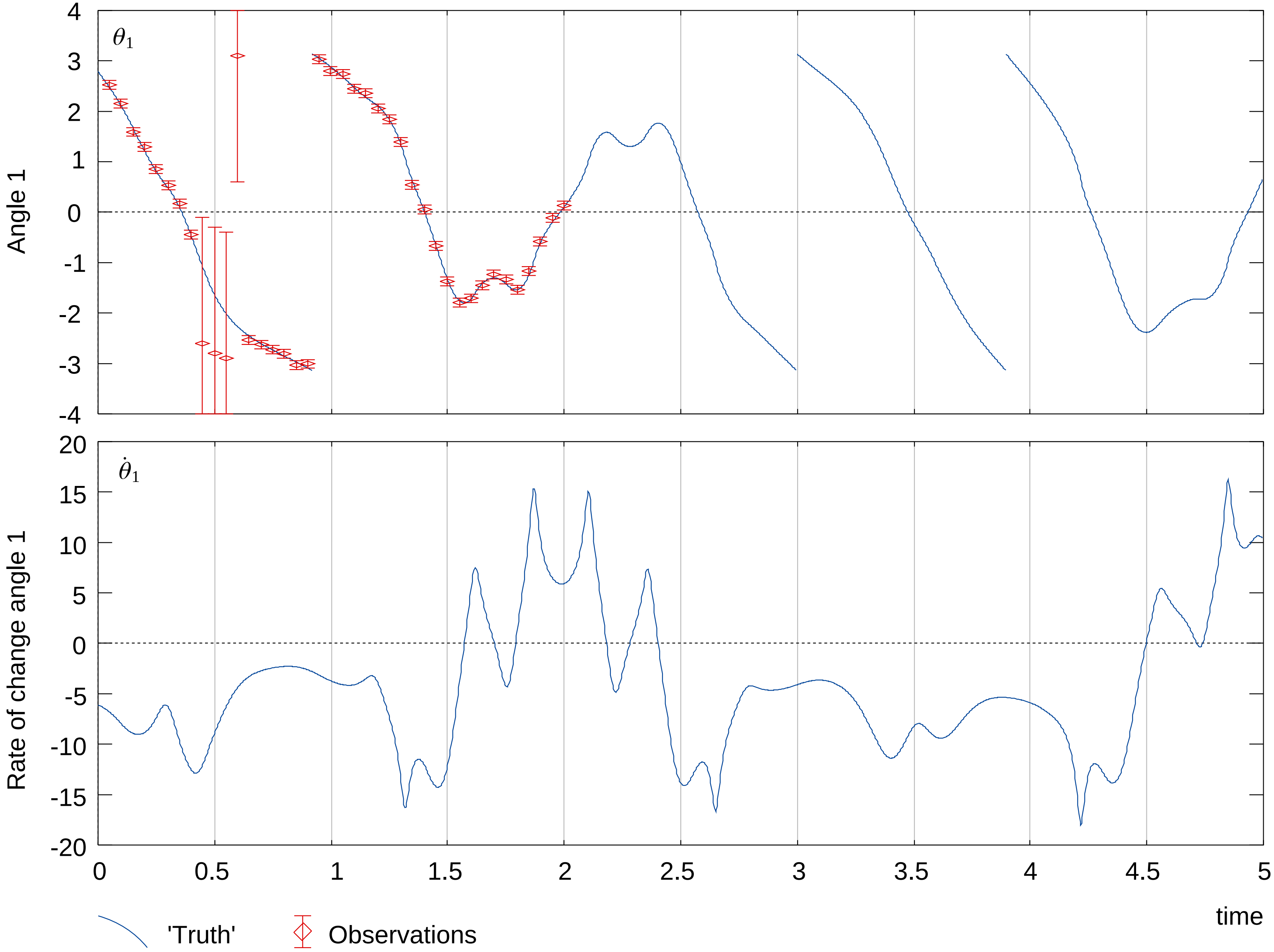
$$L = T - V \qquad \frac{\mathrm{d}}{\mathrm{d}\,t}\left(\frac{\partial L}{\partial \dot{\theta}_i}\right) = \frac{\partial L}{\partial \theta_i}$$

$$V = g m_1 l_1 \cos\theta_1 - g m_2 l_2 \cos\theta_1 - g m_3 \left(l_2 \cos\theta_1 + l_3 \cos\theta_2\right)$$

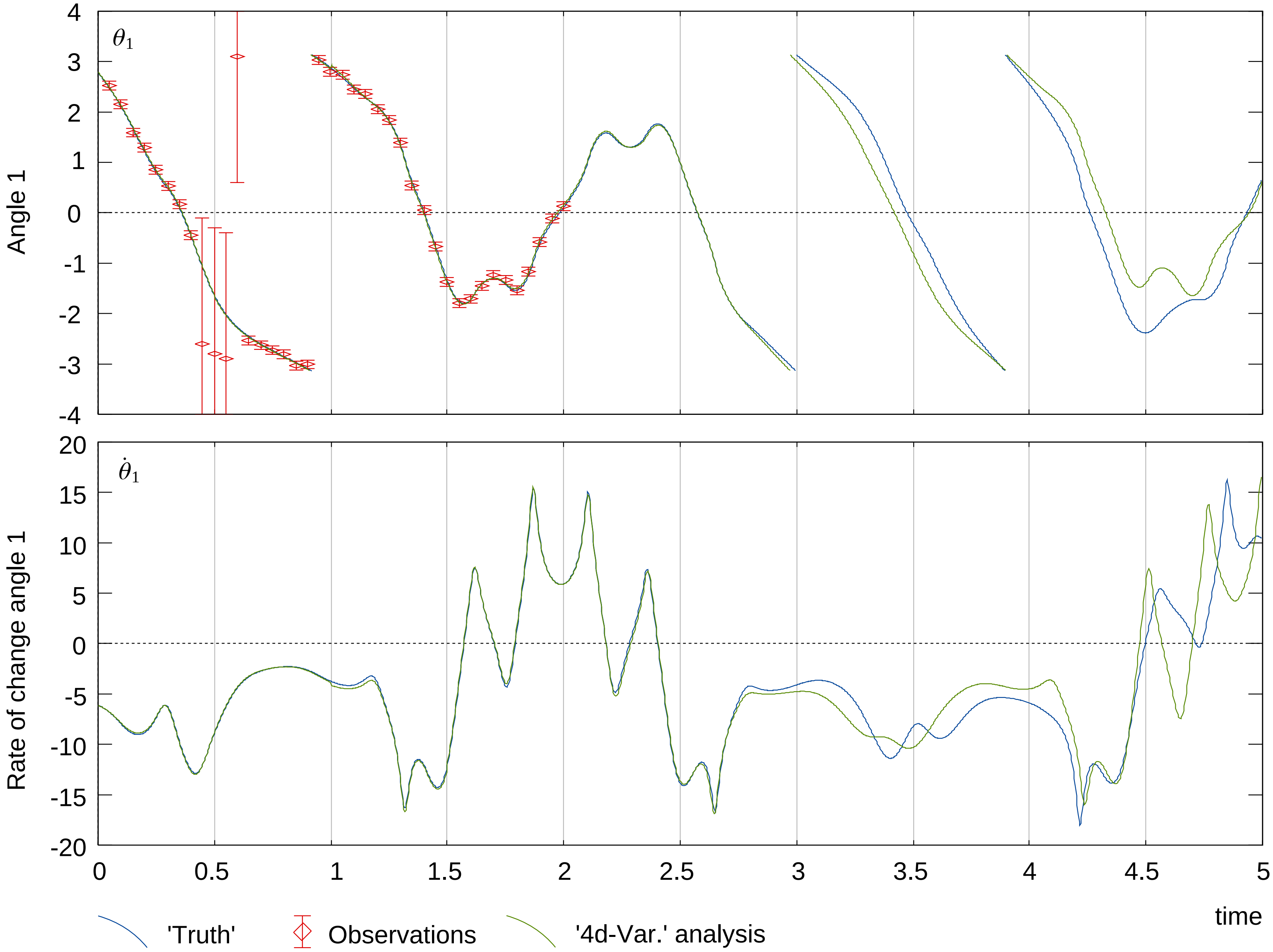
$$T = \tfrac{1}{2} m_1 (\dot{x}_1^2 + \dot{y}_1^2) + \tfrac{1}{2} m_2 (\dot{x}_2^2 + \dot{y}_2^2) + \tfrac{1}{2} m_3 (\dot{x}_3^2 + \dot{y}_3^2)$$

- Demonstrate '4d-Var' with an OSSE - 'Observation System Simulation Experiment'.
- Also known as a 'twin experiment'.
  - Choose a set of initial conditions and run the model (truth).
  - Add random noise to generate pseudo-observations.
  - Forget the truth and try to recover it by assimilating the observations.
  - Use observations of θ_1 and θ_2 only (no observations of $\dot{\theta}_1$ and $\dot{\theta}_2$).

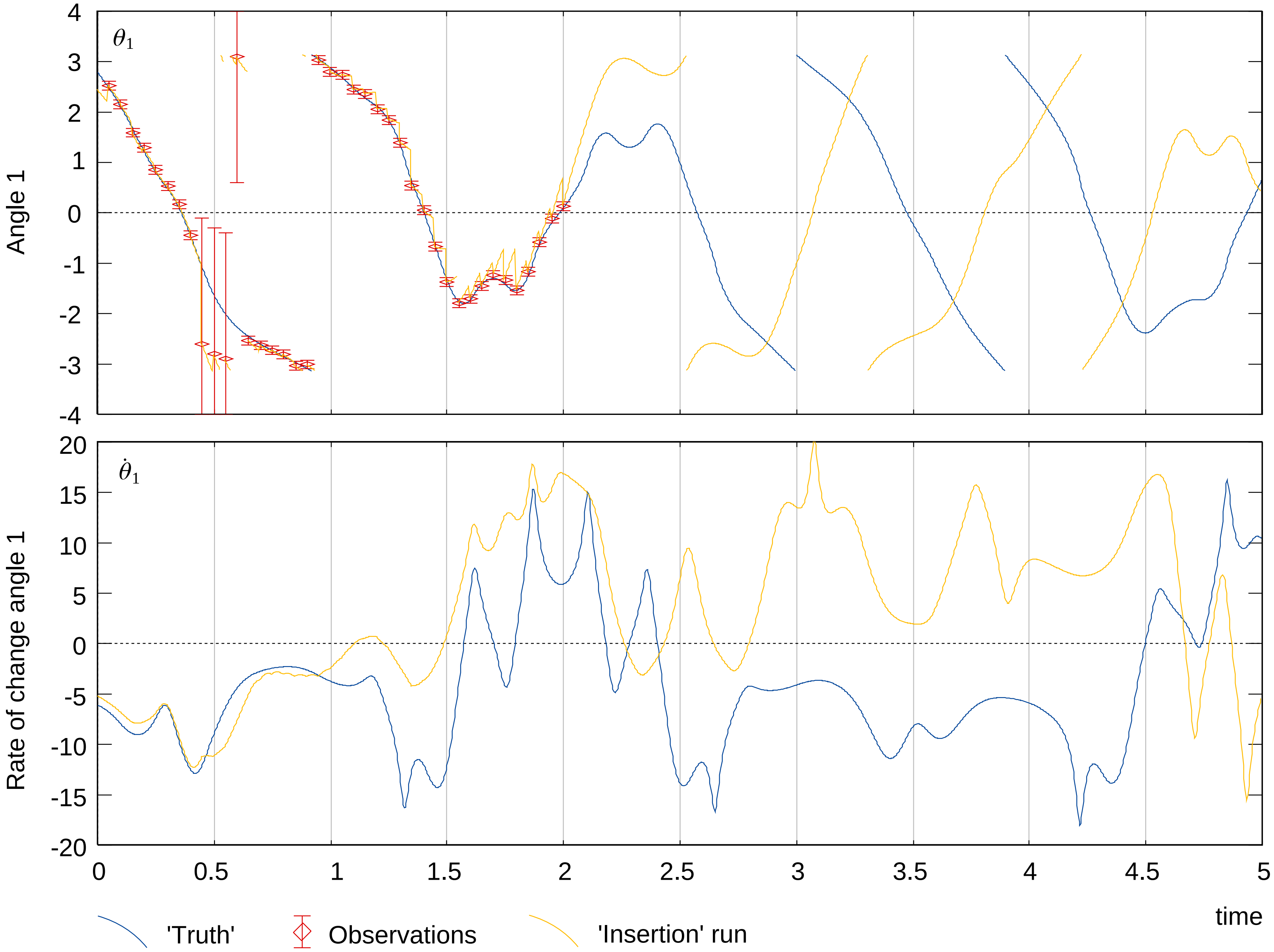
OSSE demonstration (double pendulum) - truth run



OSSE demonstration (double pendulum) - '4d-Var.' run



OSSE demonstration (double pendulum) - 'obs insertion' run



Issues with data assimilation

- Data assimilation is a computer intensive process.
  - For one cycle, 4d-Var. can use up to 100 times more computer power than the forecast.
- The **B**-matrix (forecast error covariance matrix in Var.) is difficult to deal with.
  - Assimilation process is very sensitive to **B**.
  - Least well-known part of data assimilation.
  - In operational data assimilation, **B** is a $10^7 \times 10^7$ matrix.
  - Need to model the **B**-matrix - use technique of 'control variable transforms'.
  - In reality **B** is flow dependent. Practically, **B** is quasi-static.
- Data assimilation replies on optimality. Issues of suboptimality arise if:
  - Actual error distributions are non-Gaussian,
  - **B** or **R** are inappropriate.
  - Forward models are inaccurate or are non-linear.
  - Data have biases.
  - Cost function has not converged adequately (in Var.).
- Assimilation can introduce undesirable imbalances.
- Quantities not constrained by observations can be poor (e.g. diagnosed quantities):
  - Precipitation.
  - Vertical velocity, etc.