Data Assimilation

Alan O'Neill
University of Reading, UK

The Kalman Filter

Kalman Filter
(*expensive*)

Use model equations to propagate B forward in time.

B $\longrightarrow$ B(t)

Analysis step as in OI

Evolution of Covariance Matrices

$$\mathbf{x}_b^{\;n+1} = M(\mathbf{x}_a^{\;n}) \;= M(\mathbf{x}^n) + \mathbf{M}\varepsilon_a^n$$

where M is the non-linear model, $\mathbf{M}$ is the tangent linear model, and the epsilons are vectors

$$\mathbf{x}^{\;n+1} = M(\mathbf{x}^n) - \varepsilon_m$$

Subtract: $\varepsilon_b^{\;n+1} = \mathbf{x}_b^{n+1} - \mathbf{x}^n = \mathbf{M}\varepsilon_a^{\;n} + \varepsilon_m$

The forecast error covariance is: $\mathbf{B}^{n+1}$

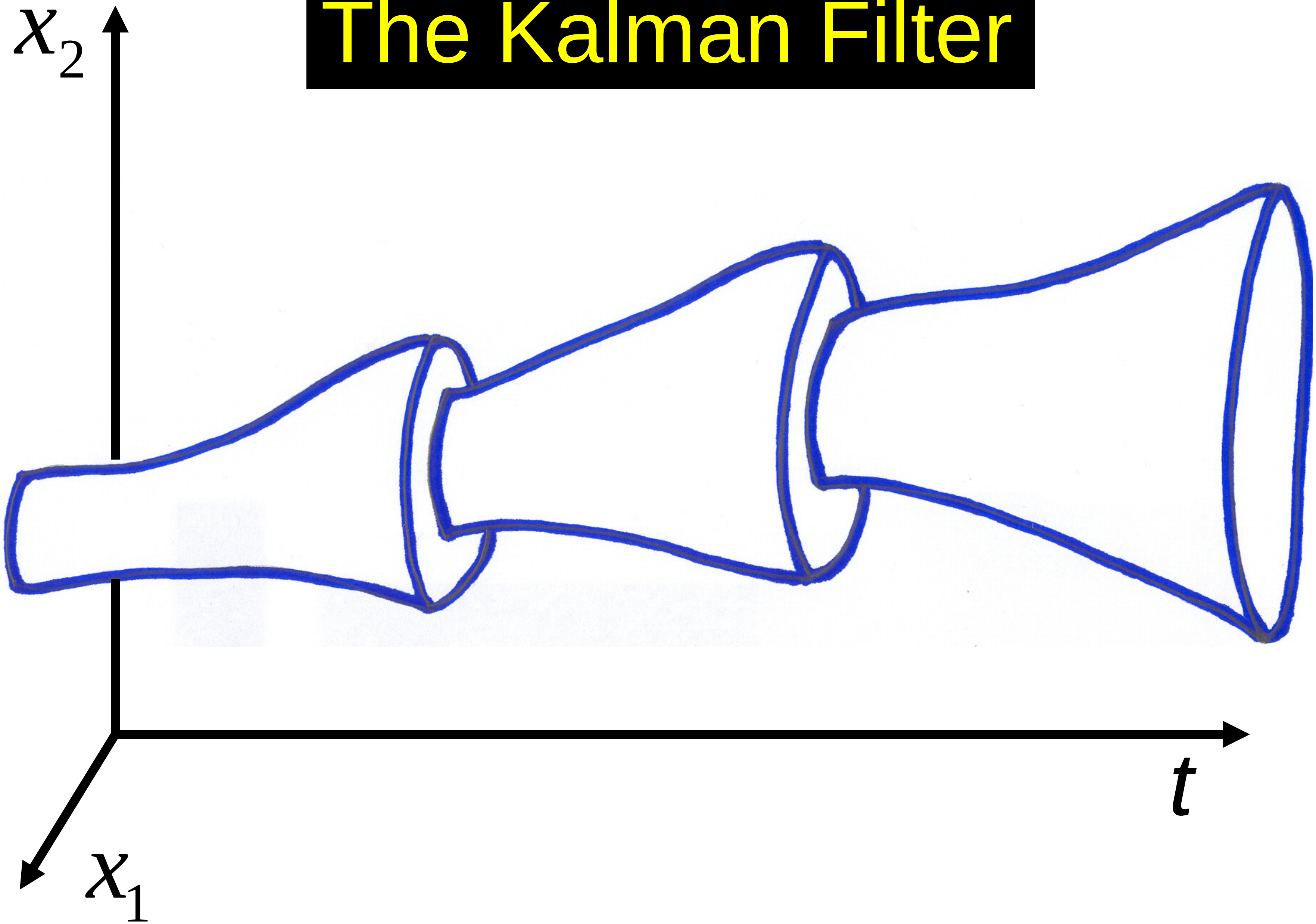
$$= < (\varepsilon_b^{\;n+1})(\varepsilon_b^{\;n+1})^{\mathrm{T}} >$$

$$= \mathbf{M}(t_n)\mathbf{P}_a\mathbf{M}^{\mathrm{T}}(t_n) + \mathbf{Q}(t_n) \;\text{ where }\; \mathbf{Q} = < \varepsilon_m\varepsilon_m^{\;\mathrm{T}} >$$

where $\mathbf{P}_a = < (\varepsilon_a^{\;n})(\varepsilon_a^{\;n})^{\mathrm{T}} >$

$$\mathbf{P}_a^{\;-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$$

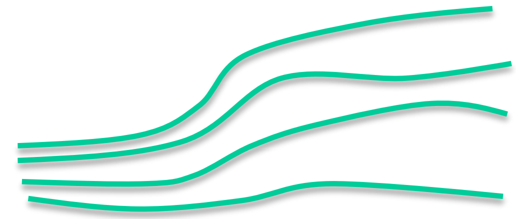
The Kalman Filter



Extended Kalman Filter

- Allows for the model to be *non-linear* and imperfect and for the observation operator to nonliear.
- Reduces to the standard KF when linearity holds (and looks like it algorithmically).
- The EKF linearises locally in time about the nonlinearly evolving state estimate.
- Very expensive to implement because of the very large dimension of the state space (~ 10^6 – 10^7 for NWP models).

Ensemble Kalman Filter

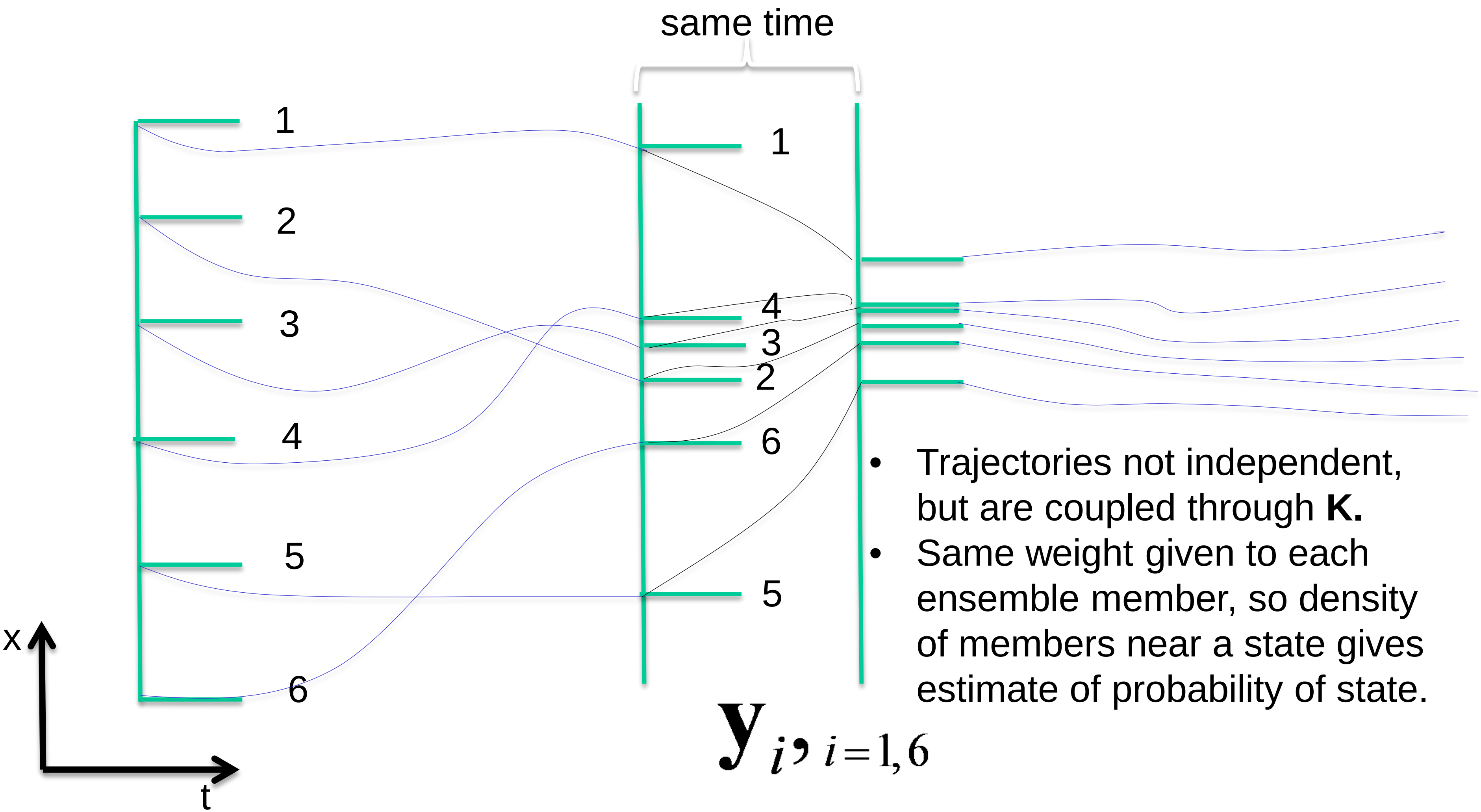


- Carry forecast error covariance matrix forward in time by using ensembles of forecasts:

$$\mathbf{B}^{n+1} \approx \frac{1}{K-1} \sum_{k \neq 1}^{K} (\mathbf{x}_k^{n+1} - < \mathbf{x}^{n+1} >)(\mathbf{x}_k^{n+1} - < \mathbf{x}^{n+1} >)^{\mathrm{T}}$$

- Only ~ 10 + forecasts needed.
- Does not require computation of tangent linear model and its adjoint.
- Does not require linearization of evolution of forecast errors.
- Fits in neatly into ensemble forecasting.

The Ensemble Kalman Filter



- Trajectories not independent, but are coupled through **K.**
- Same weight given to each ensemble member, so density of members near a state gives estimate of probability of state.

$\mathbf{y}_i, \; i=1,6$

observation vector perturbed slightly for each ensemble member

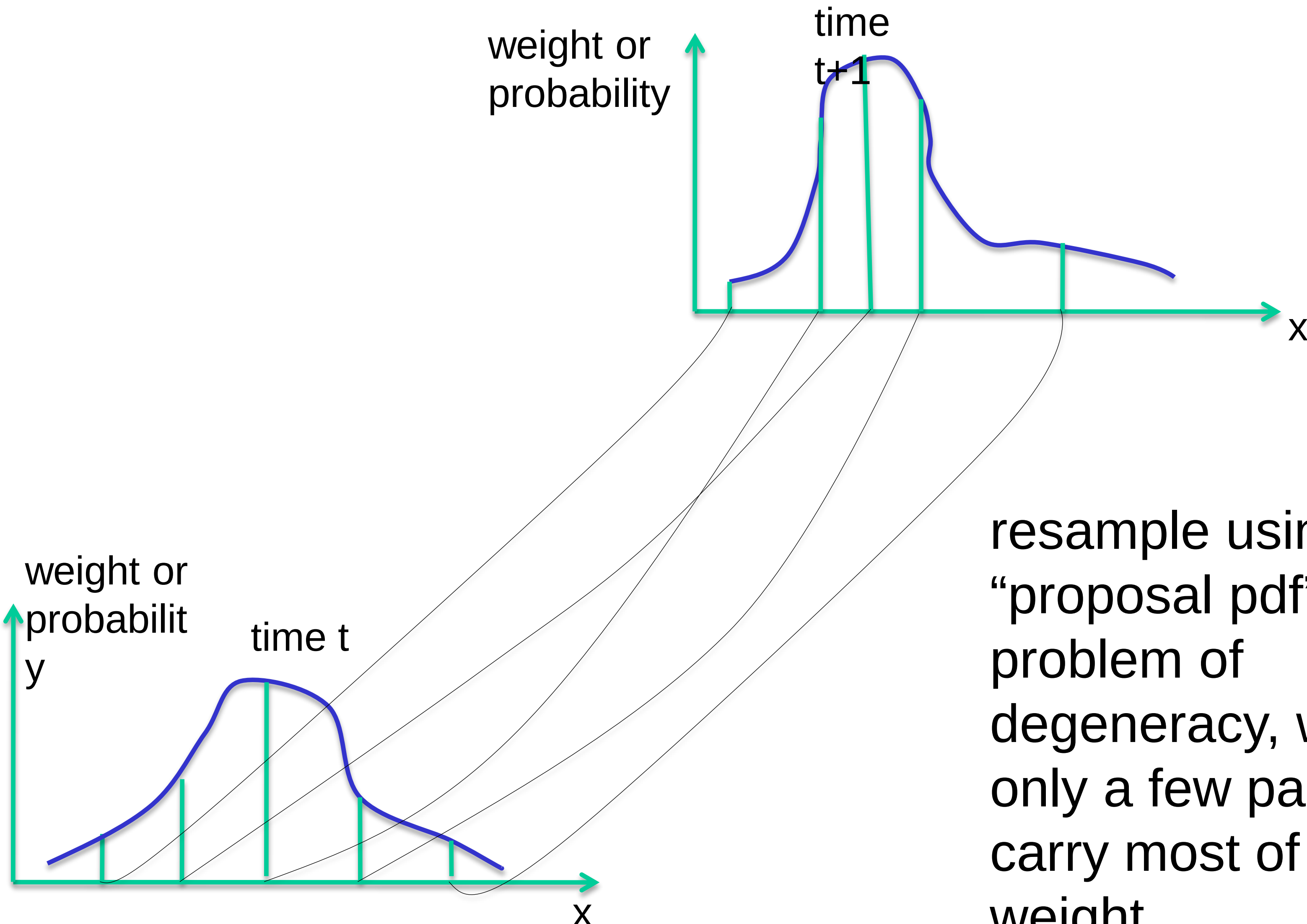
Particle Filter: moving to nonlinear, non-Gaussian methods

- Idea in particle filtering is to try to represent the probability density function of the state of a model by a number of random draws, called “particles.”
- Use Bayes’s theorem to update the weight (or probability) assigned to a given particle using observations.
- Has some similarity to Ensemble Kalman Filter, but there are significant differences.

Particle Filter in outline: dealing with nonlinear systems and allowing for non-Gaussian statistics

- Run an ensemble of trajectories with a model.
- Give each trajectory a weight to represent the probability that the actual state corresponds to the trajectory at a given time.
- Use observations to update the weights by using Bayes's theorem.
- In simple implementation, position of trajectory not altered by observations, just its weight (cf. Kalman filter where position of trajectory is altered by observations, but weights remain equal for all trajectories).

Particle Filter



resample using a "proposal pdf" to avoid problem of degeneracy, where only a few particles carry most of the weight

Particle Filter

$$p_b^t(x) \propto \sum_i w_i^t \delta(x - x_i^t)$$

Evolve the x_i^t to time $t+1$ with the model to get $p_b^{t+1}(x)$ keeping the w_i the same (add in some noise to represent model error).

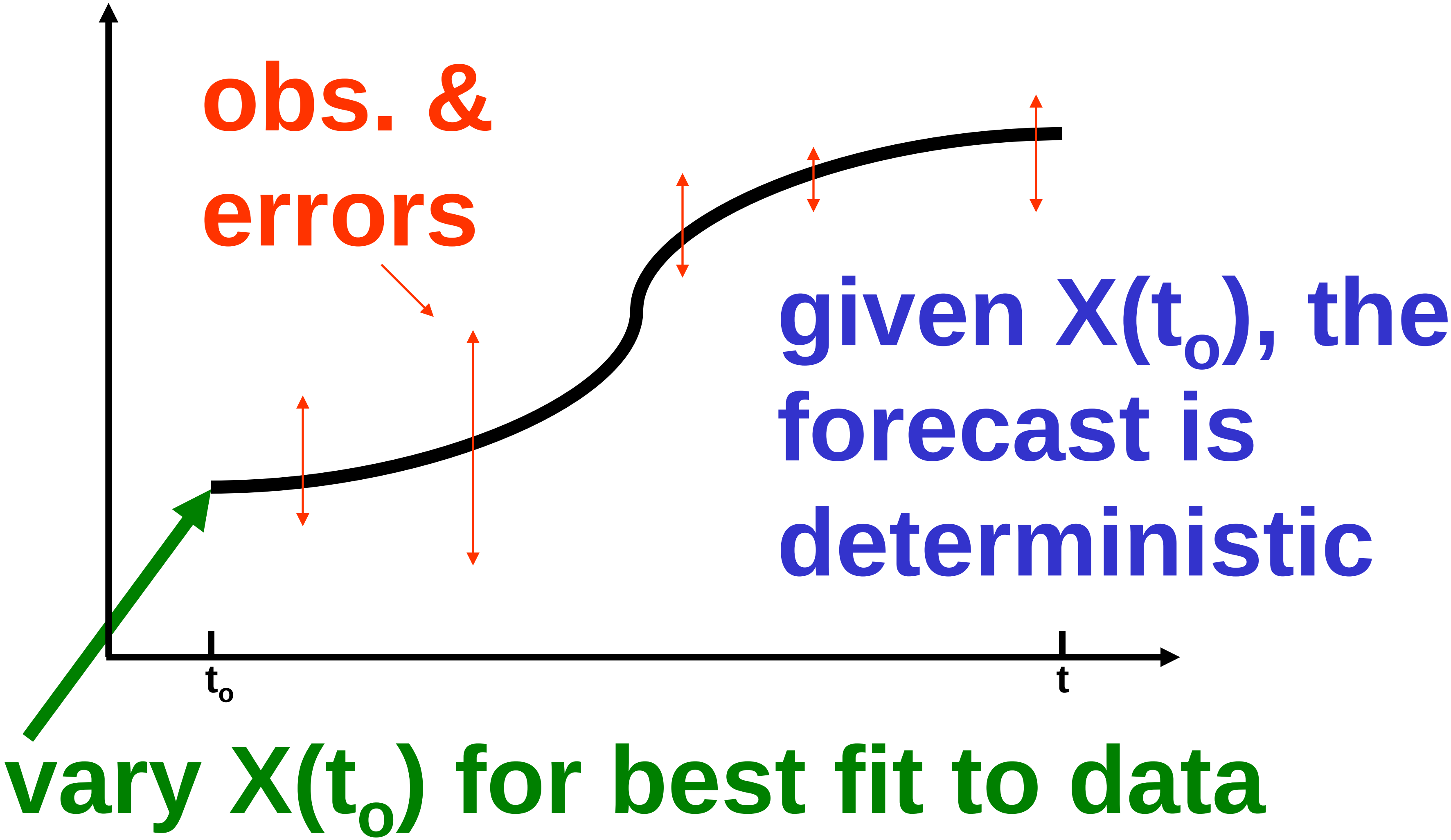
Use Bayes's theorem & observtions at time $t+1$ to update the weights

$$p_a^{t+1}(x \mid y) = \sum_i w_i^t p(y \mid x_i^{t+1}) \delta(x - x_i^{t+1})$$

Expected value of x (or f(x)) determined by BOTH density of trajectories near x and by their weights (cf. Kalman filter where only density matters since weights the same).

4d-Variational Assimilation

4D Variational Data Assimilation



4d-Variational Assimilation

$$J(\mathbf{x}(t_0)) = \frac{1}{2}\sum_{i=0}^{N}[\mathbf{y}_i - h(\mathbf{x}_i)]^{\mathrm{T}}\mathbf{R}_i^{-1}[\mathbf{y}_i - h(\mathbf{x}_i)]$$

$$+\frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}_b(t_0)]^{\mathrm{T}}\mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}_b(t_0)]$$

where $\mathbf{x}(t_i) = M_{0\to i}(\mathbf{x}(t_0))$ i.e. the model is treated as a strong constraint

Minimize the cost function by finding the gradient ("Jacobian") with respect to the control variables in $\mathbf{x}(t_0)$ $\partial J / \mathbf{x}(t_0)$

4d-VAR comments

- The 2nd term on the RHS of the cost function measures the distance to the background $\mathbf{x}_b(t_0)$ at the beginning of the interval.
- The term helps join up the sequence of optimal trajectories found by minimizing the cost function for the observations.
- The “analysis” is then the optimal trajectory in state space. Forecasts can be run from any point on the trajectory, e.g. from the middle.

Some Matrix Algebra

$$J = J(\mathbf{x}(\mathbf{x}_0))$$

Then $\dfrac{\partial J}{\partial \mathbf{x}_0} = \left(\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} \dfrac{\partial J}{\partial \mathbf{x}}$

adjoint of the model

$$M : \mathbf{x}_0 \mapsto \mathbf{x}$$

Let J have the following form: $J = \dfrac{1}{2}\mathbf{z}^{\mathrm{T}}(\mathbf{x})\mathbf{A}\mathbf{z}(\mathbf{x})$

Then it can be shown that $\dfrac{\partial J}{\partial \mathbf{x}} = \left(\dfrac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^{\mathrm{T}} \mathbf{A}\mathbf{z}$

Combining these results: $\dfrac{\partial J}{\partial \mathbf{x}_0} = \left(\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} \left(\dfrac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^{\mathrm{T}} \mathbf{A}\mathbf{z}$

4d-VAR for Single Observation

$$J(\mathbf{x}(\mathbf{x}_0)) = \frac{1}{2}[\mathbf{y} - h(\mathbf{x}(\mathbf{x}_0))]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - h(\mathbf{x}(\mathbf{x}_0))]$$

obs. term only

By using results on slide "Some Matrix Algebra":

$$\frac{\partial J}{\partial \mathbf{x}_0} = -\mathbf{L}_{0\to t}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - h(\mathbf{x}(\mathbf{x}_0))] \equiv -\mathbf{L}_{0\to t}^{\mathrm{T}}\mathbf{d}$$

where $\mathbf{L}_{0\to t}^{\mathrm{T}} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} = \frac{\partial M_{0\to t}^{\mathrm{T}}(\mathbf{x}_0)}{\partial \mathbf{x}_0}$,

adjoint of tangent linear model

$$\mathbf{L}_{0\to t} = \mathbf{L}_{t_{n-1}\to t} \ldots \mathbf{L}_{t_1\to t_2}\mathbf{L}_{0\to t_1}$$

$$\therefore \mathbf{L}_{0\to t}^{\mathrm{T}} = \mathbf{L}_{0\to t_1}^{\mathrm{T}}\mathbf{L}_{t_1\to t_2}^{\mathrm{T}} \ldots \mathbf{L}_{t_{n-1}\to t}^{\mathrm{T}}$$

$\Rightarrow$ backward integration in time of TLM

4d-VAR Procedure

- Choose $\mathbf{x}_0 , \mathbf{x}_0^b$ for example.
- Integrate full (non-linear) model forward in time and calculate $\mathbf{d}$ for each observation.
- Map $\mathbf{d}$ back to *t*=0 by backward integration of TLM, and sum for all observations to give the gradient of the cost function.
- Move down the gradient to obtain a better initial state (new trajectory “hits” observations more closely)
- Repeat until some STOP criterion is met.

note: not the most efficient algorithm

Comments

- 4d-VAR can also be formulated by the method of Lagrange multipliers to treat the model equations as a constraint. The adjoint equations that arise in this approach are the same equations we have derived by using the chain rule of partial differential equations.
- If model is perfect and B_0 is correct, 4d-VAR at final time gives same result as extended Kalman filter (but the covariance of the analysis is not available in 4d-VAR).
- 4d-VAR analysis therefore optimal over its time window, but less expensive than Kalman filter.

Incremental Form of 4d-VAR

- The 4d-VAR algorithm presented earlier is expensive to implement. It requires repeated forward integrations with the non-linear (forecast) model and backward integrations with the TLM.
- When the initial background (first-guess) state and resulting trajectory are accurate, an incremental method can be made much cheaper to run on a computer.

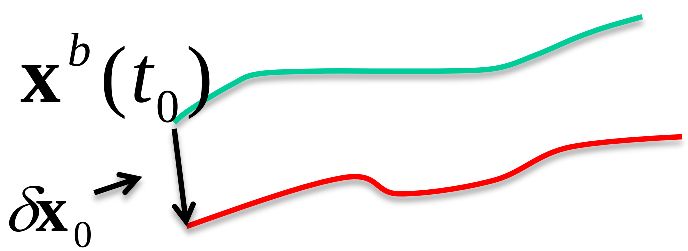
Incremental Form of 4d-VAR

The incremental form of the cost function is defined by

$$J(\delta\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0)^{\mathrm{T}}\mathbf{B}_0^{-1}(\delta\mathbf{x}_0)$$

where $\delta\mathbf{x}_0 = \mathbf{x}(t_0) - \mathbf{x}^b(t_0)$

$$+\frac{1}{2}\sum_{i=0}^{N}[\mathbf{y}_i - H(\mathbf{x}^f(t_i)) - \mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0]^{\mathrm{T}}\mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}^f(t_i)) - \mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0]$$



Taylor series expansion about first-guess trajectory starting from

$\mathbf{x}^f(t_i)$

$\mathbf{x}^b(t_0)$

Minimization can be done in lower dimensional space

4D Variational Data Assimilation

- Advantages
  - consistent with the governing eqs.
  - implicit links between variables
- Disadvantages
  - very expensive
  - model is strong constraint

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Summary of basic principles

- DA is concerned with estimating the state of a system given:
  - observations (direct [e.g. in-situ] and indirect [e.g. remotely sensed]),
  - forecast models (to provide a-priori data, given too-few obs),
  - observation operators (to connect model state with obs).
- All data have uncertainties, which must be quantified.
  - DA estimates are sensitive to uncertainty characteristics, which are often poorly known.
  - Many observations and model have systematic as well as random errors.
  - Should take into account all sources of error in the system.
- DA theory is suited mostly to errors that are Gaussian distributed.
  - Most errors are non-Gaussian and non-linearity is synonymous with non-Gaussianity.
- DA problems are computationally expensive and require intensive development effort.

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.

Leading methods of solving the DA problem

Method	Description	Pros	Cons
A. Data insertion	Set grid points to observation values	1. Easy to do	1. No respect of uncertainty 2. What about observation voids? 3. Can't deal with indirect observations
B. Variational data assimilation	Minimize a cost function Many flavours: 3D, 4D, weak/strong constraint	1. Respect of data uncertainty 2. Direct and indirect observations **3.** $\mathbf{P}_f$ gives smooth and balanced fields 4. Efficient 5. Can deal with (weakly) non-linear **h**	**1.** $\mathbf{P}_f$ is difficult to know, often static and suboptimal 2. High development costs **3. h**: need tangent linear, **H** and adjoint, $\mathbf{H}^T$ 4. Gaussian pdf
C. Kalman filtering	Evaluate KF equations	1. As B.1, B.2, B.3 **2.** $\mathbf{P}_f$ adapts with the state	1. As B.3, B.4 2. Difficult to use with non-linear **h** 3. Prohibitively expensive for large *n*
D. Ensemble Kalman filtering	Approximate KF equations with ensemble of *N* model runs Many flavours	1. As B.1,B.2, B.4, B.5, C.2 **2. h**: do not need **H** and $\mathbf{H}^T$ 3. Have measure of analysis spread	1. As B.4 2. Serious sampling issues when $N << n$ 3. Need ensemble inflation and localization schemes to overcome D.2
E. Hybrid	Cross between C/D	1. As B.1, B.2, B.3, B.4, B.5, C.2	1. As D.2
F. Particle filter	Assign weights to ensemble members to represent any pdf	1. As. B.1, B.2 2. Can deal with non-linear **h** 3. Can deal with non-Gaussian pdf 4. Have measure of analysis spread	1. As D.2 2. Inefficient – members often become redundant 3. Need special techniques to overcome F.2

Some Useful References

- Atmospheric Data Analysis by R. Daley, Cambridge University Press.
- Atmospheric Modelling, Data Assimilation and Predictability by E. Kalnay, C.U.P.
- The Ocean Inverse Problem by C. Wunsch, C.U.P.
- Inverse Problem Theory by A. Tarantola, Elsevier.
- Inverse Problems in Atmospheric Constituent Transport by I.G. Enting, C.U.P.
- Dynamic Data Assimilation, Lewis et al. C.U.P
- Data Assimilation: the ensemble KF, G. Evensen, Springer
- Quantitative Remote Sensing of Land Surfaces, S Liang, Wiley
- ECMWF lecture notes: www.ecmwf.int

END