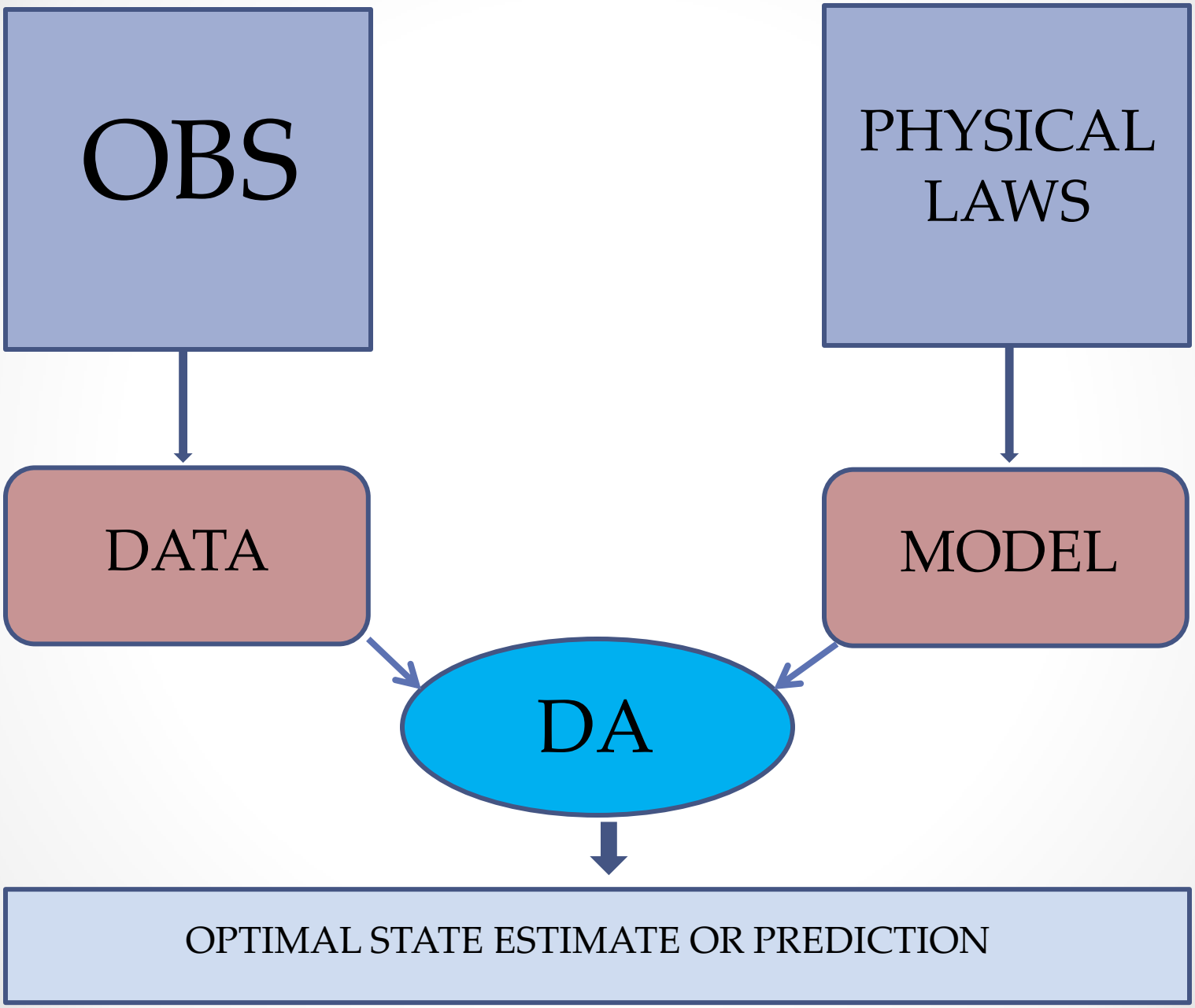
Is Data Assimilation Relevant to Climate Research?

Christopher Jones
Warwick Mathematics Institute,
University of Warwick, and
Department of Mathematics,
University of North Carolina at Chapel Hill

Newton Institute, Cambridge: August, 2010


OBS
PHYSICAL LAWS
DATA
MODEL
DA
OPTIMAL STATE ESTIMATE OR PREDICTION

Model:

$x \in \mathbb{R}^N$ -state vector comprising all relevant dynamical variables

(e.g. flow velocity, temperature, salinity, etc. at each grid point)

$dx^f = M(x^f, t)dt$ -prognostic model

$dx^t = M(x^t, t)dt + \sigma(t)dt$ -actual evolution

$E[\sigma(t)\sigma^T(t')] = \delta(t - t')Q(t)$ -covariance of the model residual

Observations:

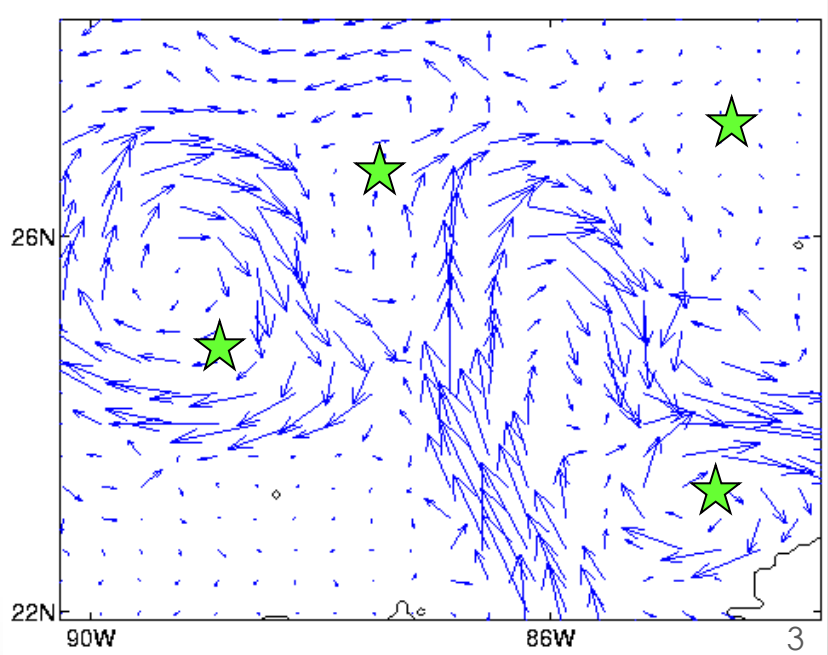
$\eta_i^o = H_i[x_i^t] + \boldsymbol{\varepsilon}_i$

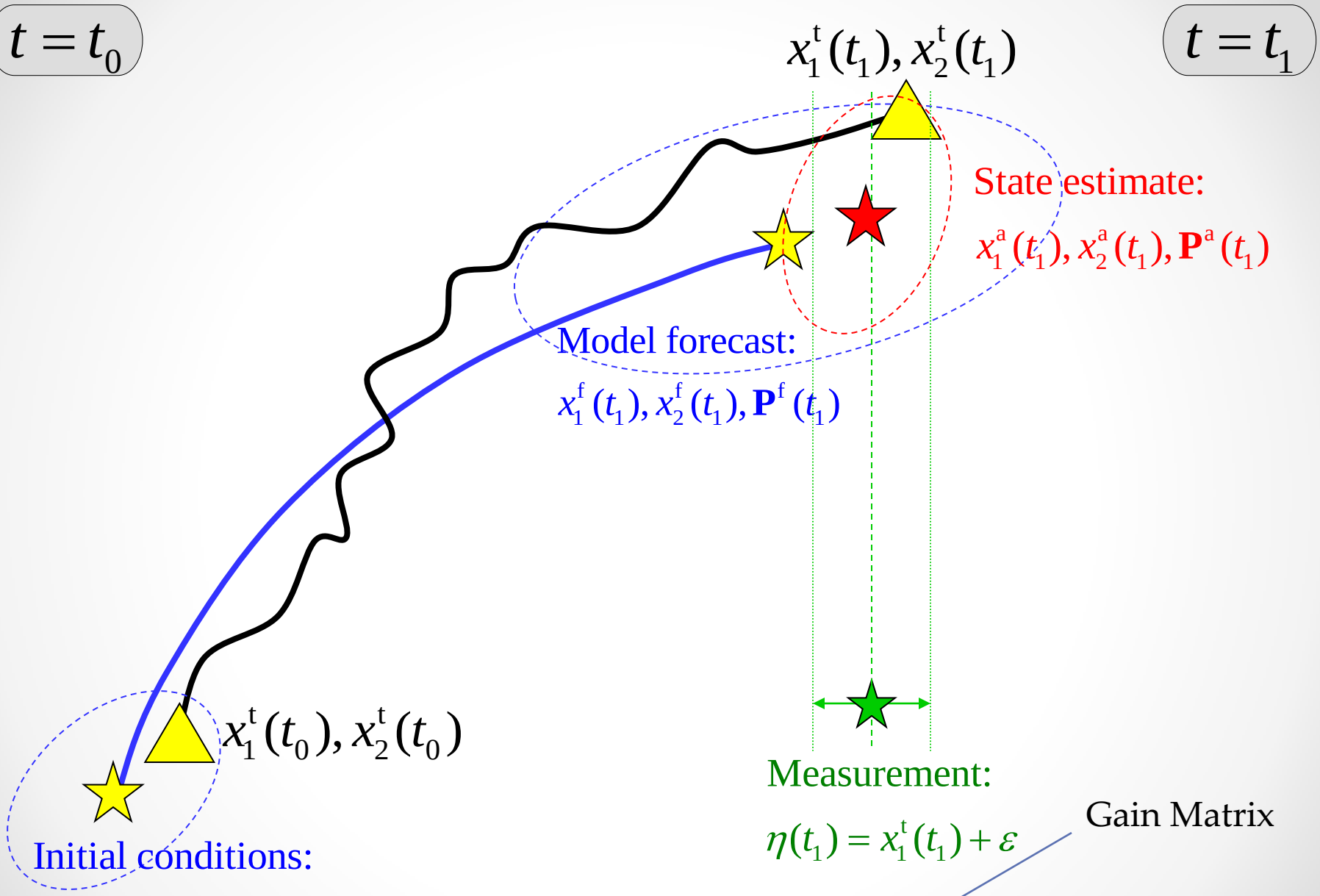
H_i -observation operator

$\boldsymbol{\varepsilon}_i$ -observation error

$\eta_i^o \in \mathbf{R}^L$, typically $L \ll N$

$E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_m^T] = \delta_{im}\mathbf{R}_i$ -Obs error covariance

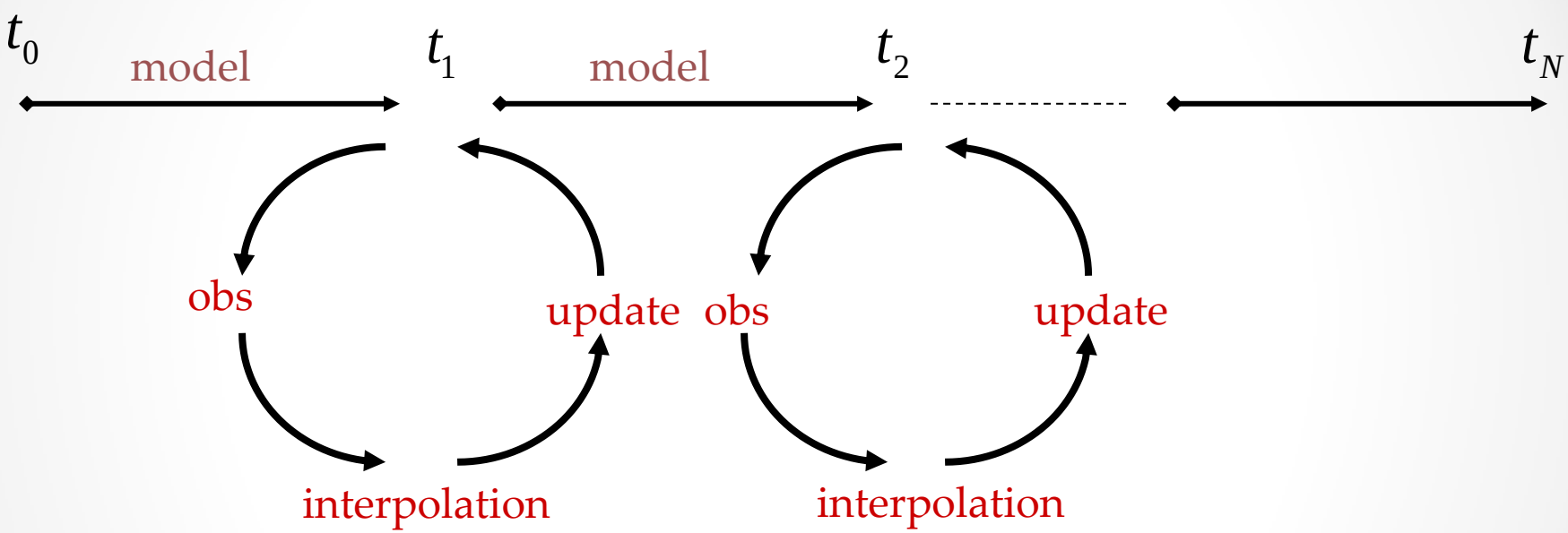




$$x^a = x^f + K\left(\eta - H(x^f)\right)$$

Data Assimilation in Predictive Mode

Model + observations $\longrightarrow$ prediction



Interpolation at $t = t_i$:

$$x_i^{\,a} = x_i^{\,f} + K_i\left(\eta_i - H(x_i^{\,f})\right)$$

Sequential DA/Forward Problem/Filtering

$$x^a = x^f + K\left(\eta - H(x^f)\right)$$

Where the matrix K is determined by the requirement that x^a minimise the cost function:

$$J(x) = \left\langle x - x^f, \left(P^f\right)^{-1}\left(x - x^f\right)\right\rangle + \left\langle \eta - H\left(x^f\right), R^{-1}\left(\eta - H\left(x^f\right)\right)\right\rangle$$

$$\Downarrow$$

$$K = P^f H^T \left(R + HP^f H^T\right)^{-1}$$

H (linearised) observation operator

R observation error covariance matrix

P^f determined by DA scheme being used!

$$x^a = x^f + K\left(\eta - H(x^f)\right)$$

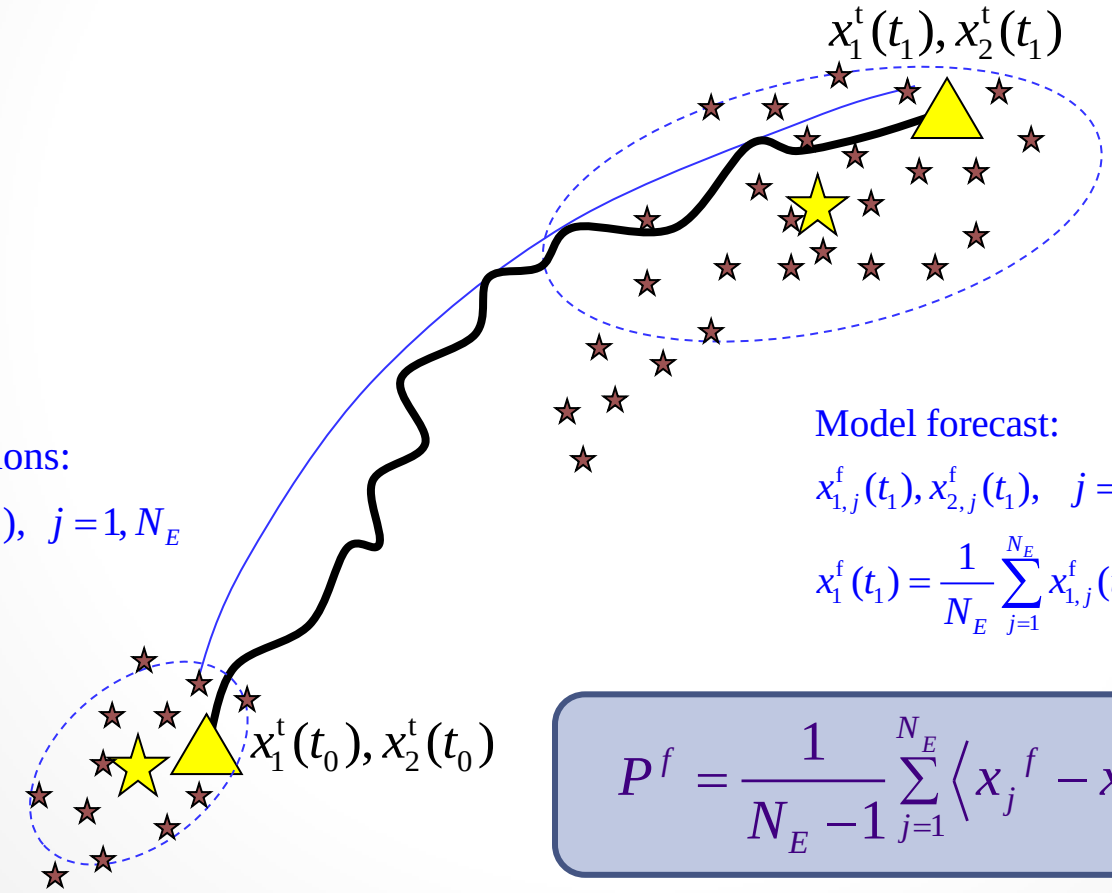
$$K = P^f H^T \left(R + HP^f H^T\right)^{-1}$$

Possibilities for calculating P^f

1. P^f is constant background error covariance matrix (Optimal Interpolation=OI)
2. If model is linear, P^f is evolved under model (Kalman Filter=KF)
3. If model is nonlinear, P^f is evolved under tangent linear model (Extended Kalman Filter=EKF)
4. P^f is built out of ensembles evolved under full nonlinear model (Ensemble Kalman Filter=EnKF)

Ensemble Kalman Filter (EnKF)

Error covariance is predicted via solution of full nonlinear system for a Monte-Carlo ensemble of states



Initial conditions:

$$x_{1,j}^{\mathrm{f}}(t_0), x_{2,j}^{\mathrm{f}}(t_0), \quad j = 1, N_E$$

Model forecast:

$$x_{1,j}^{\mathrm{f}}(t_1), x_{2,j}^{\mathrm{f}}(t_1), \quad j = 1, N_E$$

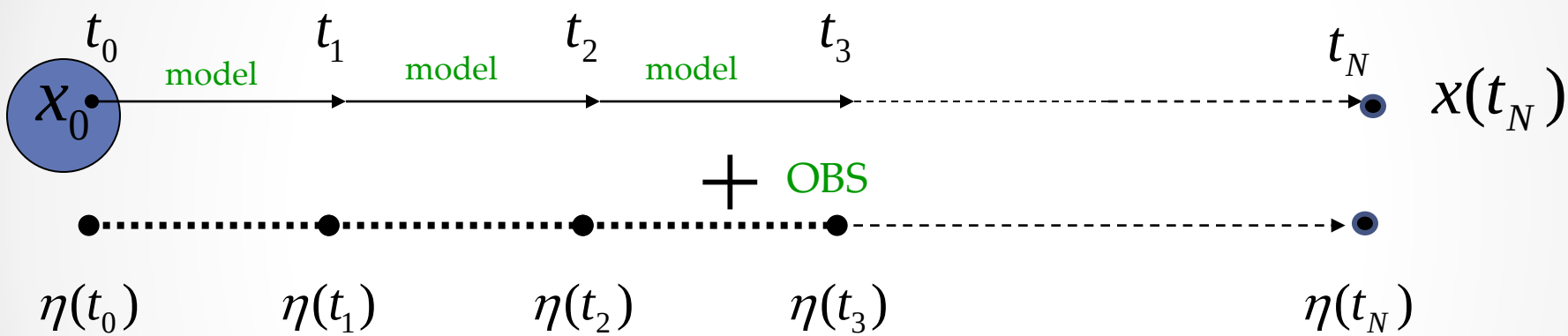
$$x_1^{\mathrm{f}}(t_1) = \frac{1}{N_E}\sum_{j=1}^{N_E} x_{1,j}^{\mathrm{f}}(t_1), \quad x_2^{\mathrm{f}}(t_1) = \frac{1}{N_E}\sum_{j=1}^{N_E} x_{2,j}^{\mathrm{f}}(t_1)$$

$$P^f = \frac{1}{N_E - 1}\sum_{j=1}^{N_E}\left\langle x_j^{\,f} - x^f, x_j^{\,f} - x^f \right\rangle$$

DA in State Estimation Mode

Model run + observations $\longrightarrow$ state estimate

$t = t_1, \ldots, t_N$



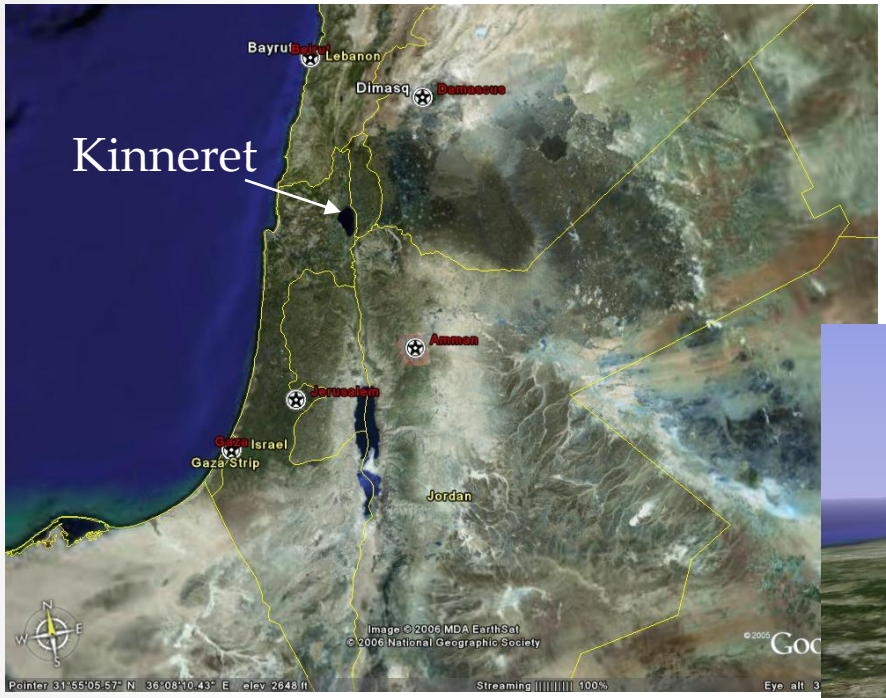
4DVAR: Minimise the cost function:

$$J(x) = \left\langle x_0 - x_0^*, B^{-1}\left(x_0 - x_0^*\right)\right\rangle + \sum_1^N \left\langle \eta_j - H(x(t_j)), R_j^{-1}\left(\eta_j - H(x(t_j))\right)\right\rangle$$

$x_0^* =$ estimate $\qquad B =$ background error covariance matrix

Variational DA/Inverse Problem/Smoothing

Fishkill in Lake Kinneret



Vernieres et al. (2006)

Conjecture: due to "lifting" of lower layer of oxygen-free water

- Occasional "fishkill"
- Feeding of 5,000??

Model:

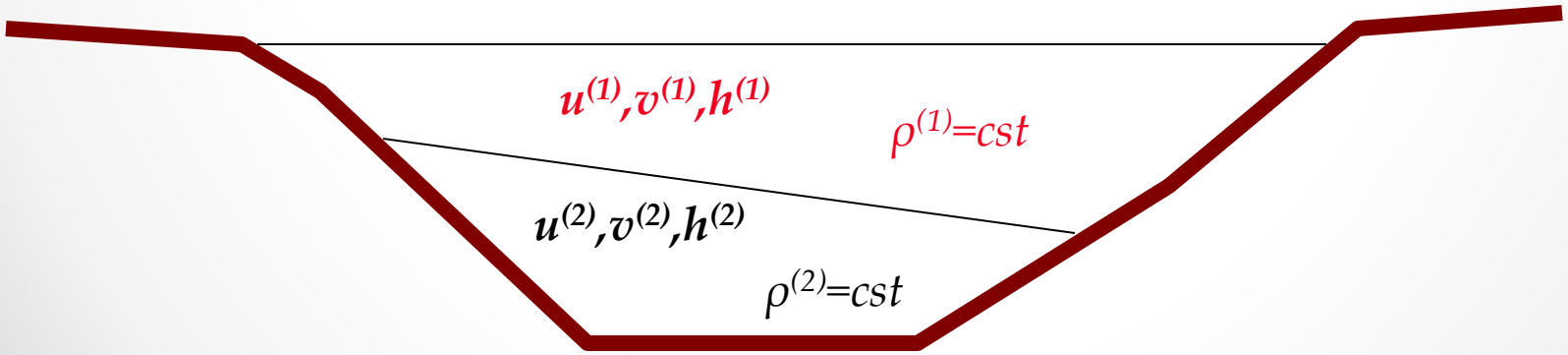
- Stably stratified during summer
- Strong westerly sea breeze

$$\left.\begin{aligned}\frac{du^{(1)}}{dt} - fv^{(1)} + g\frac{\partial}{\partial x}\left(h^{(1)} + h^{(2)} + D\right) - A_h\nabla^2 u^{(1)} - F_u = 0\\ \frac{dv^{(1)}}{dt} + fu^{(1)} + g\frac{\partial}{\partial y}\left(h^{(1)} + h^{(2)} + D\right) - A_h\nabla^2 v^{(1)} - F_u = 0\end{aligned}\right\}$$ Top layer momentum

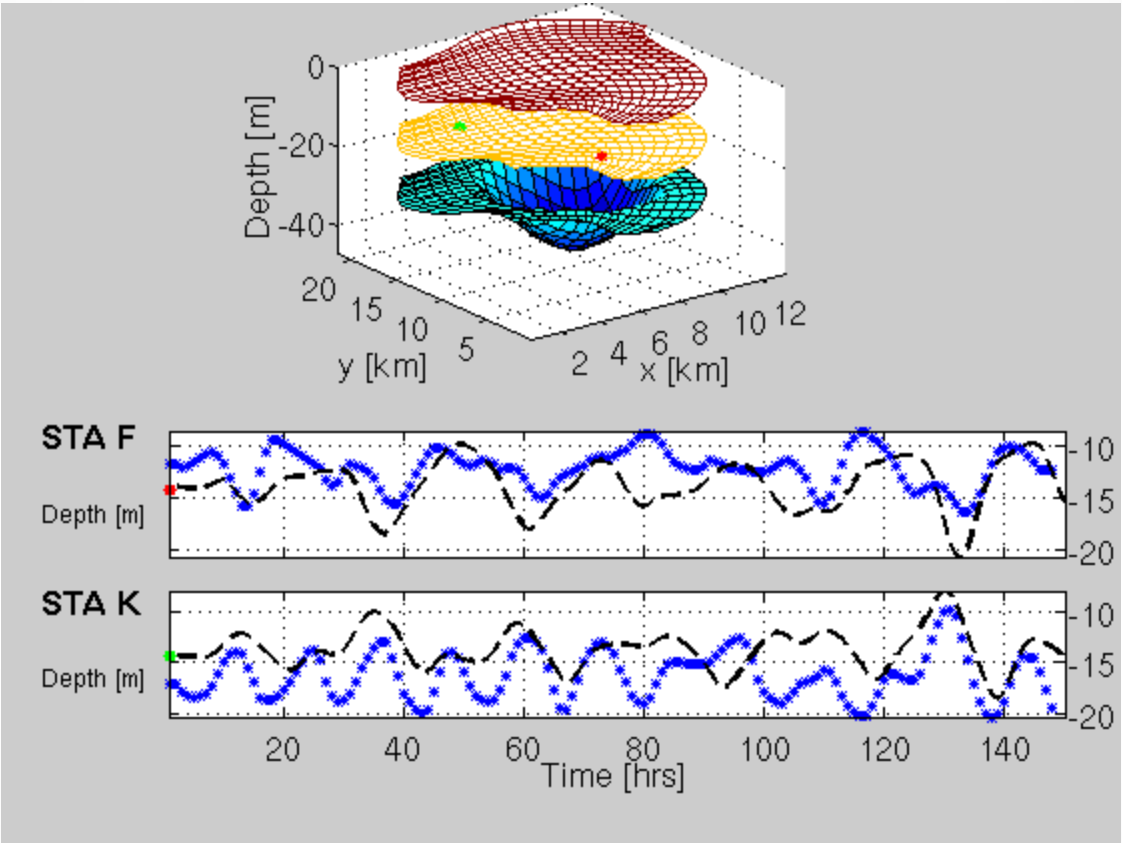
$$\frac{dh^{(1)}}{dt} = 0$$

$$\left.\begin{aligned}\frac{du^{(2)}}{dt} - fv^{(1)} + g\frac{\partial}{\partial x}\left(h^{(1)} + h^{(2)} + D\right) + g'\frac{\partial}{\partial x}\left(h^{(2)} + D\right) - A_h\nabla^2 u^{(1)} = 0\\ \frac{dv^{(2)}}{dt} + fu^{(1)} + g\frac{\partial}{\partial y}\left(h^{(1)} + h^{(2)} + D\right) + g'\frac{\partial}{\partial y}\left(h^{(2)} + D\right) - A_h\nabla^2 v^{(1)} = 0\end{aligned}\right\}$$ Bottom layer momentum

$$\frac{dh^{(2)}}{dt} = 0$$

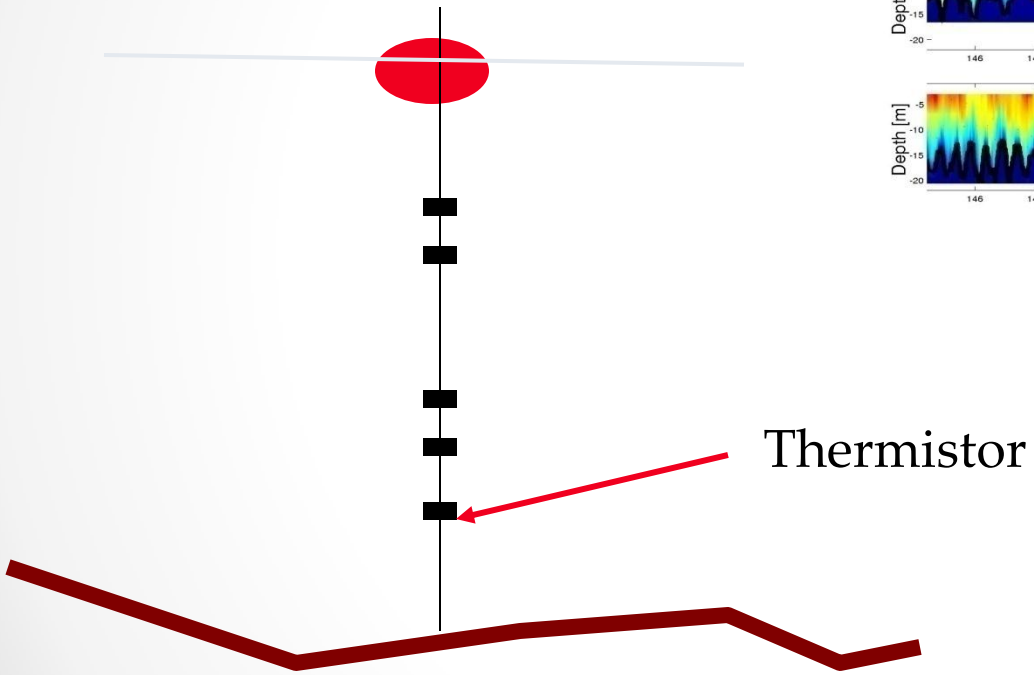


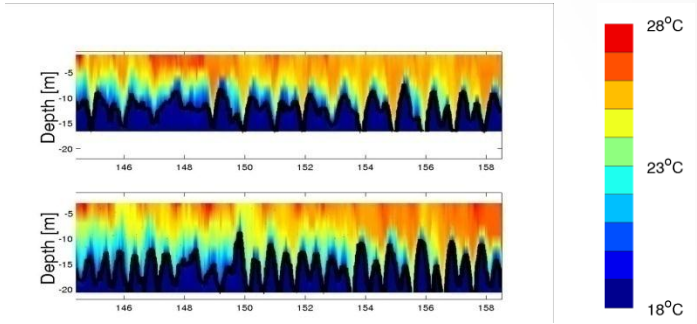
Model run


Depth [m]
0
-20
-40
20
15
10
5
y [km]
2
4
6
8
10
12
x [km]
STA F
Depth [m]
-10
-15
-20
STA K
Depth [m]
-10
-15
-20
20
40
60
80
100
120
140
Time [hrs]

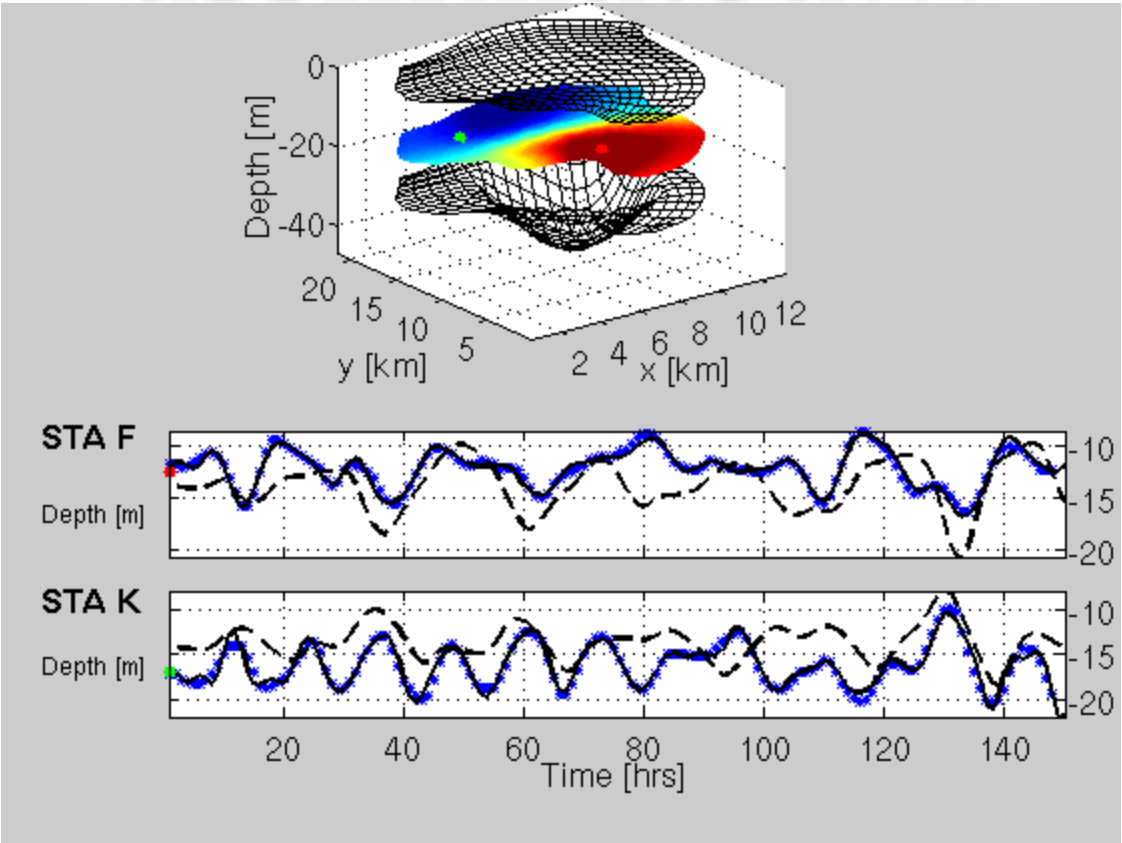
Data

- Thermistor chains



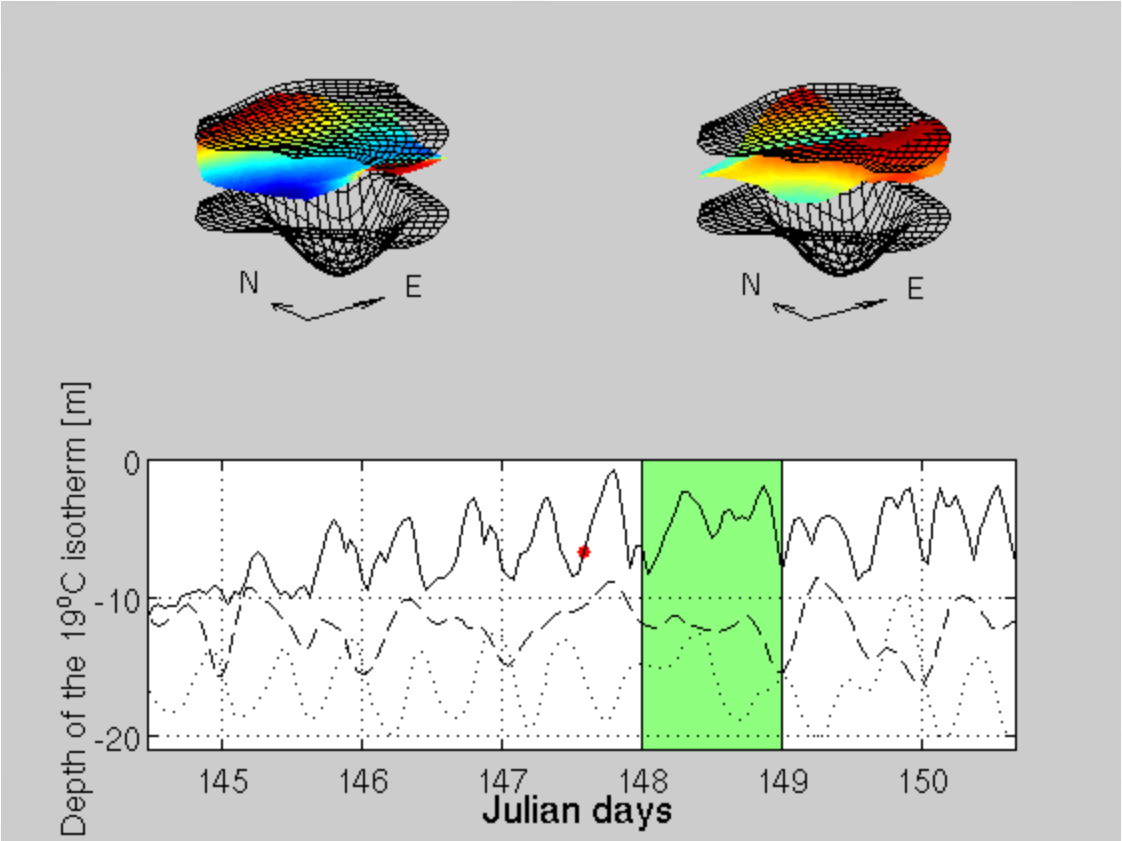


With thermistor data assimilated…



A comparison

Model alone



DA applied

Neither model nor data on their own show "fishkill,"
But, together, they do!

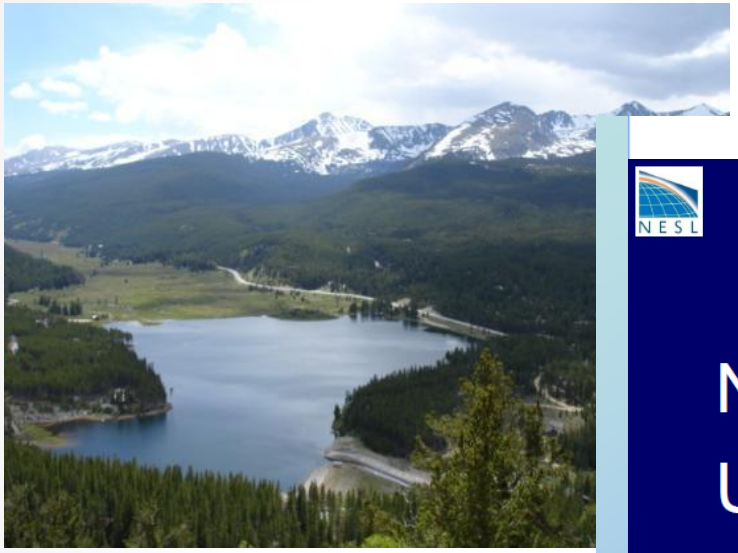
Climate: is DA relevant?

Weather prediction: DA is strikingly successful

Climate prediction: neither approach is, in an obvious way, useful or even appropriate

Conventional wisdom: climate prediction is NOT an initial value problem, but a forcing and boundary value problem

CESM Annual Meeting Breckenridge June, 2010



Antarctic sea ice cover

Challenges

New system complexity

Ultra-high resolution

Data assimilation

Needs of user community

Where are we heading?

Beachhead for DA in Climate

Paleoclimate: state estimation DA is becoming recognised as an appropriate and useful tool

Ocean: IPCC AR5 focus on decadal predictions- need to get ocean initial condition right! DA can be very useful in this endeavour.

Key issues in both areas: *long* timescales and *sparse* data

Workshop: Data Assimilation and Climate Research, June 2010

Dalton Minimum 1790-1820

(Van der Schrier and Barkmeijer, Climate Dynamics 2005)

- Cold in Western Europe
- Decreased solar activity
- Volcanoes: 1809 and 1815

Bjerknes Hypothesis:

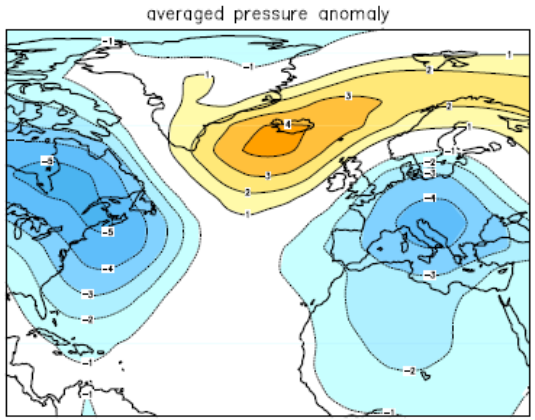
Anomalous ocean-atmosphere interaction

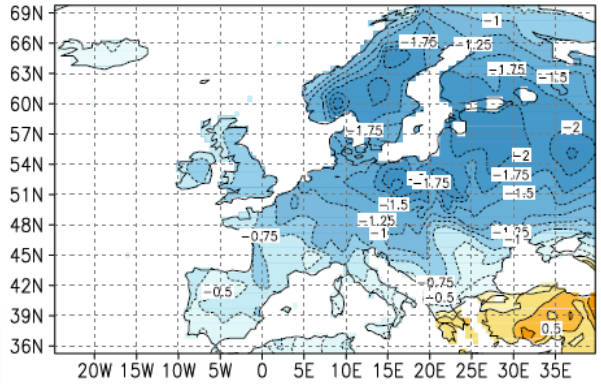
↓

Southerly flow of cold polar waters into NE Atlantic Ocean

↓

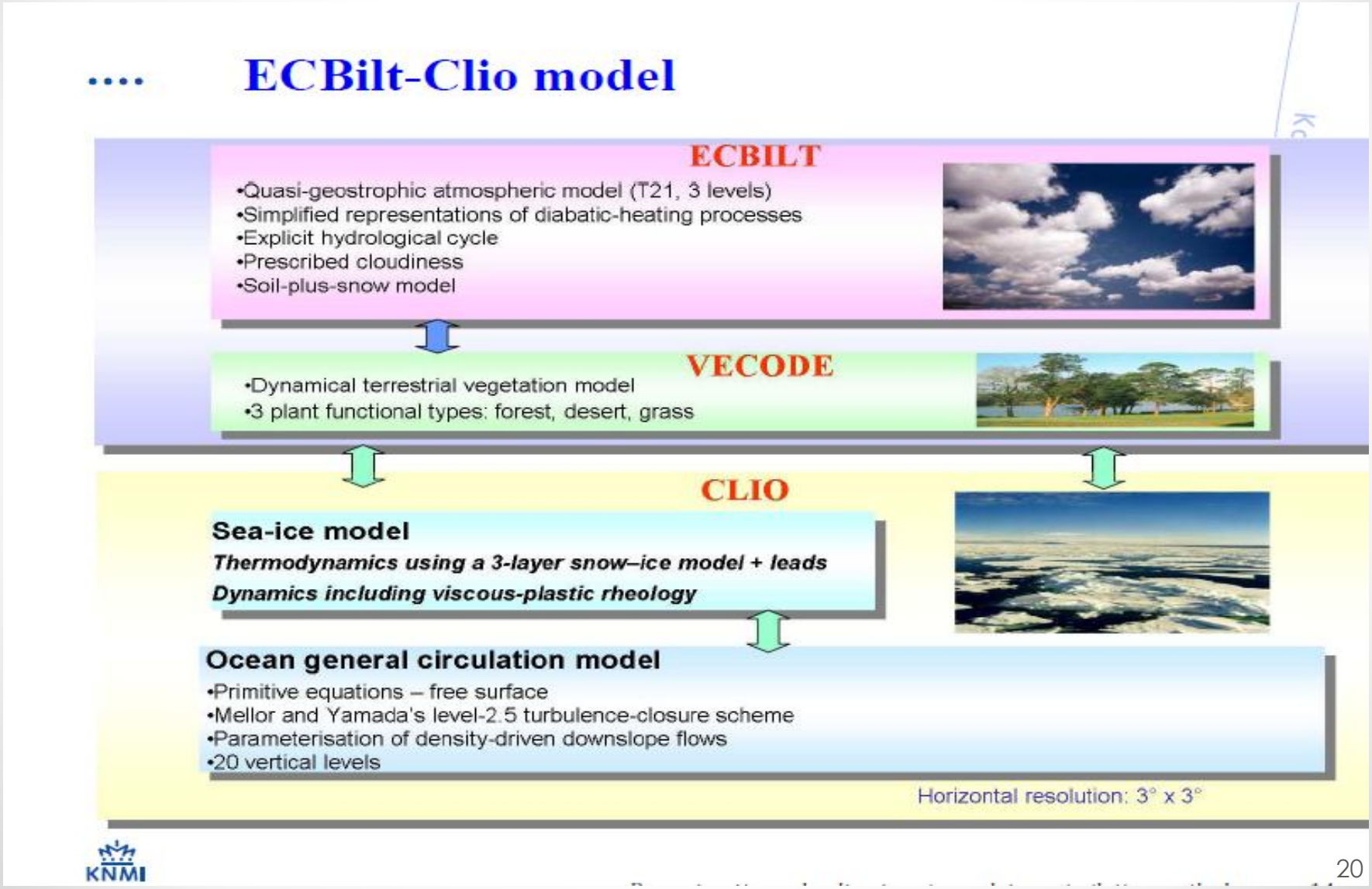
Western Europe is cold even into Summer months





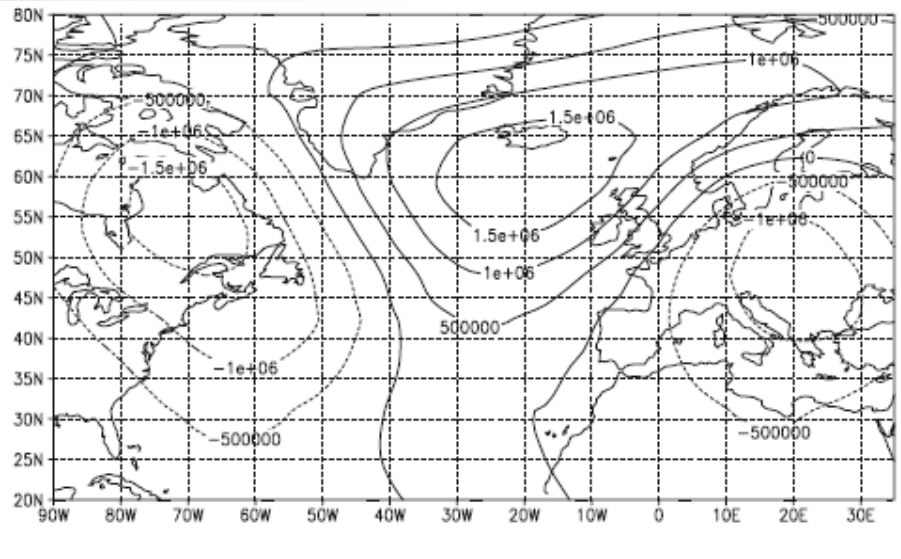
Luterbacher et al. (2004)

Van der Schrier and Barkmeijer introduce a DA technique to assess relationship between anomalous SLP and cold period

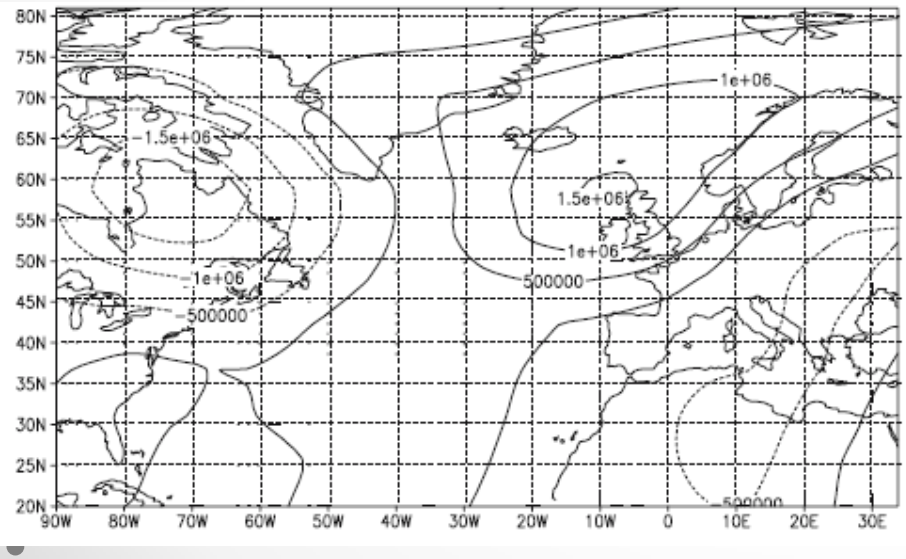


Atmospheric circulation:

QG streamfunction ($m^2 s^{-1}$)



Target-used in DA



Difference (DJF) between DA and control runs

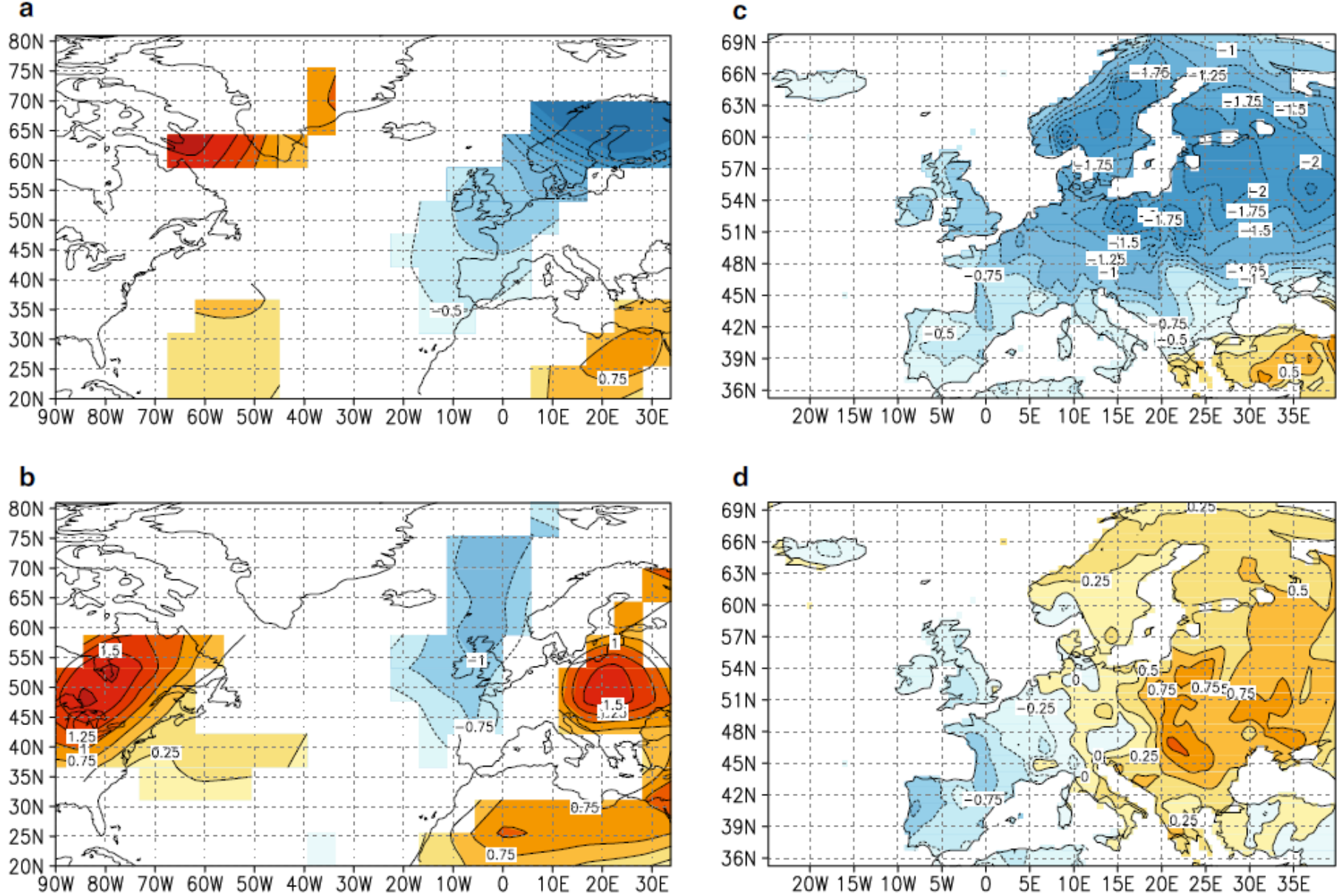
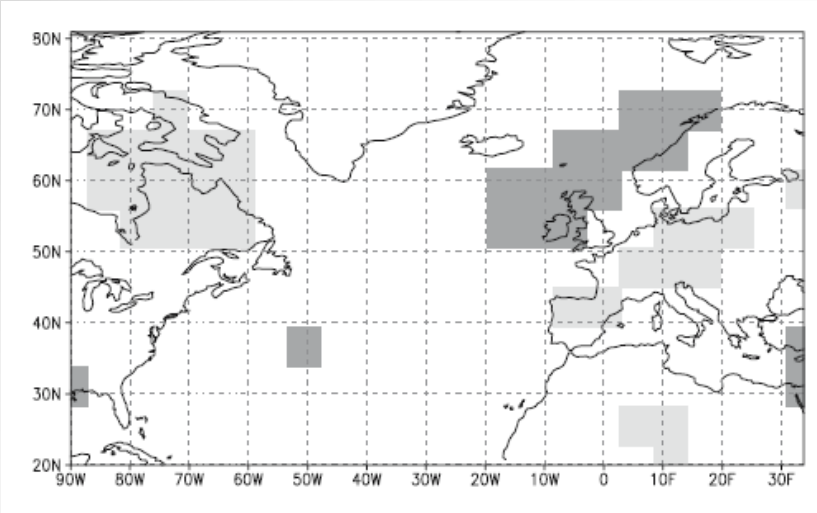
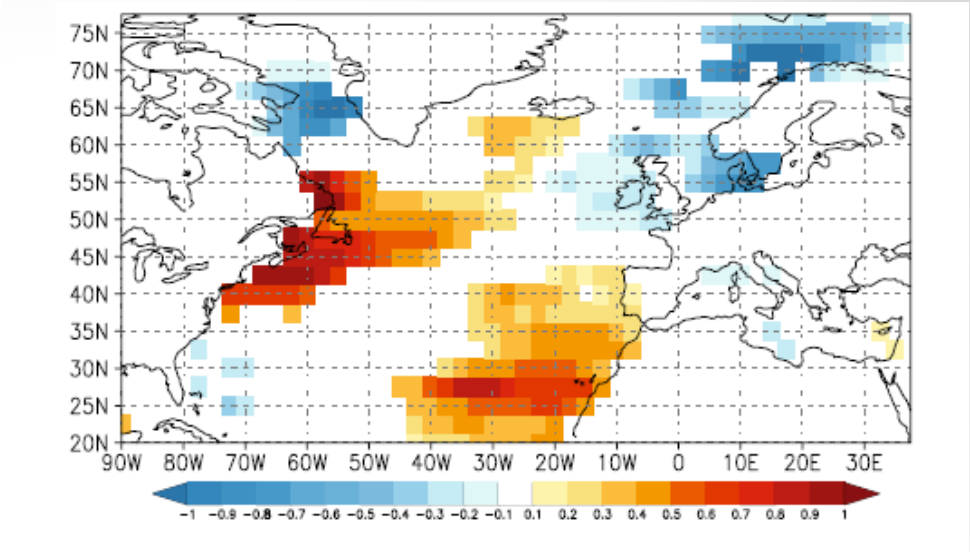


Fig. 13 Difference in 2 meter temperature, averaged over 25 year, for the winter (DJF) season (fig. a) and the summer (JJA) season (fig. b) between the data-assimilated and the control run. In both seasons, north-western Europe is under the data-assimilated run colder than in the control run. All temperature changes shown are statistically significant at the 95% level. Figures c and d show the difference in reconstructed surface air temperature for the period 1790-1820 AD with respect to 1971-2000, for the winter (DJF) season (fig. c) and the summer (JJA) season (fig. d). Reconstructed temperatures are from Luterbacher et al. (2004)



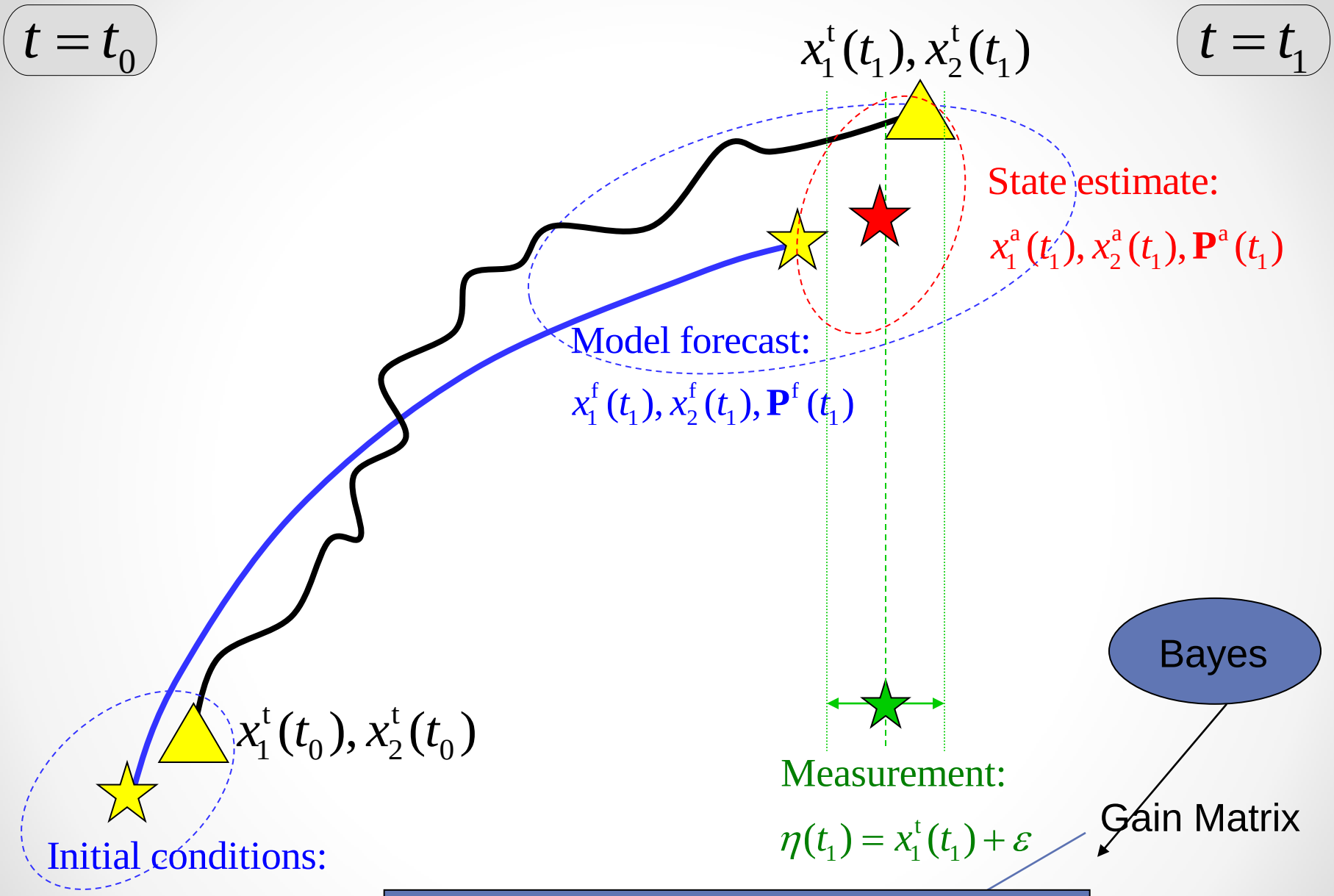
Anomalous wind-stress curl (DJF)



SST difference between DA and control (DJF)

Conclusions:

1. Anomalous ocean/atmosphere interaction is directly related to Western European cooling,
2. Independent of possible primary forcing due to solar or volcanic activity
3. MOC weakening is present but not significant



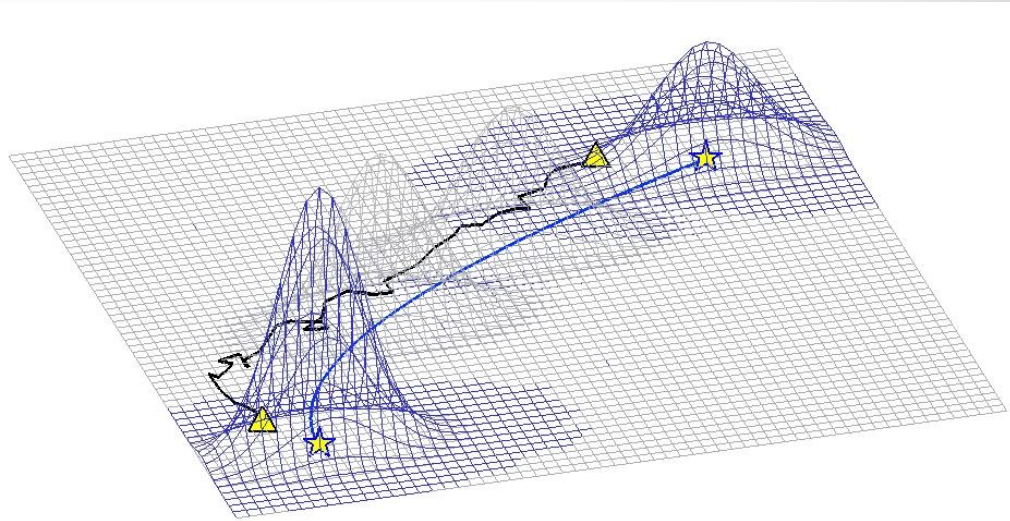
$$P^{\mathrm{posterior}}(x|\eta) = P^{\mathrm{obs}}(\eta|x) P^{\mathrm{prior}}(x)$$

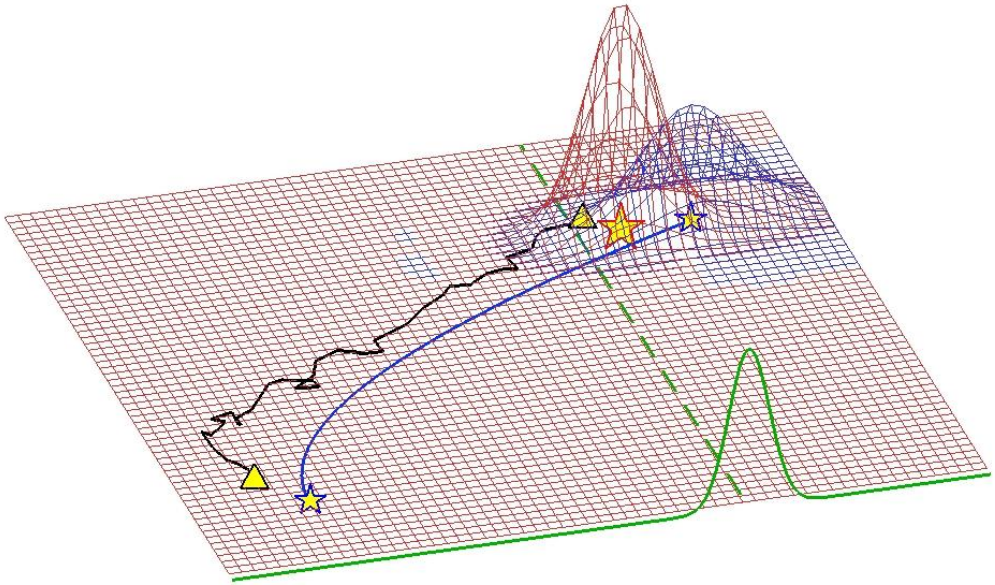
$$x^a = x^f + K(\eta - (Hx^f))$$

Forecast step:

$$p(\mathbf{x}, t_0) \rightarrow p(\mathbf{x}, t_1)$$

$$\frac{\partial p}{\partial t} + \frac{\partial (M_i p)}{\partial x_i} = \frac{1}{2} \frac{\partial^2 (Q_{ij} p)}{\partial x_i \partial x_j}$$





Bayes step (update/analysis):

$$p(\mathbf{x}, t_1) \rightarrow p(\mathbf{x}, t_1 \mid \mathbf{y}^{\mathrm{o}})$$

$$p(\mathbf{x}, t_1 \mid \mathbf{y}^{\mathrm{o}}) = \frac{p(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}) p(\mathbf{x}, t_1)}{\int p(\mathbf{y}^{\mathrm{o}} \mid \mathbf{z}) p(\mathbf{z}, t_1) d\mathbf{z}}$$

But: computationally prohibitive, state $\approx 10^6$

Posterior and Cost Function

Predictive:

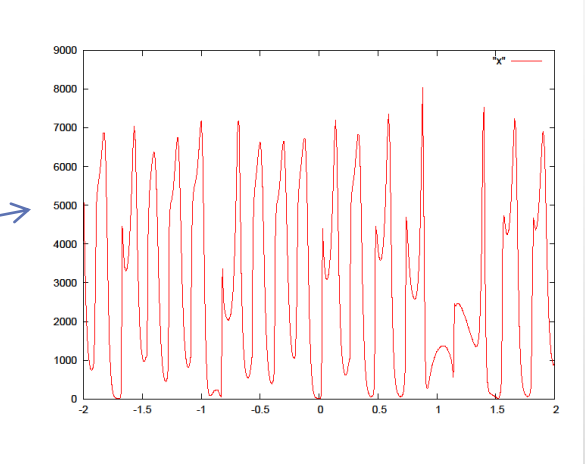
$$P(x\,|\,\eta) \propto \exp\left(-J(x)\right) = \exp\left(-\left\langle x - x^f, \left(P^f\right)^{-1}\left(x - x^f\right)\right\rangle - \left\langle \eta - H\left(x^f\right), R^{-1}\left(\eta - H\left(x^f\right)\right)\right\rangle\right)$$

State estimation:

$$P(x\,|\,\eta) \propto \exp\left(-J(x)\right) = \exp\left(-\left\langle x_0 - x_0^*, B^{-1}\left(x_0 - x_0^*\right)\right\rangle - \sum_1^N \left\langle \eta_j - H(x(t_j)), R_j^{-1}\left(\eta_j - H(x(t_j))\right)\right\rangle\right)$$

The minimisation gives the posterior mode

log posterior of the Lorenz '63 system (courtesy of Jochen Voss)



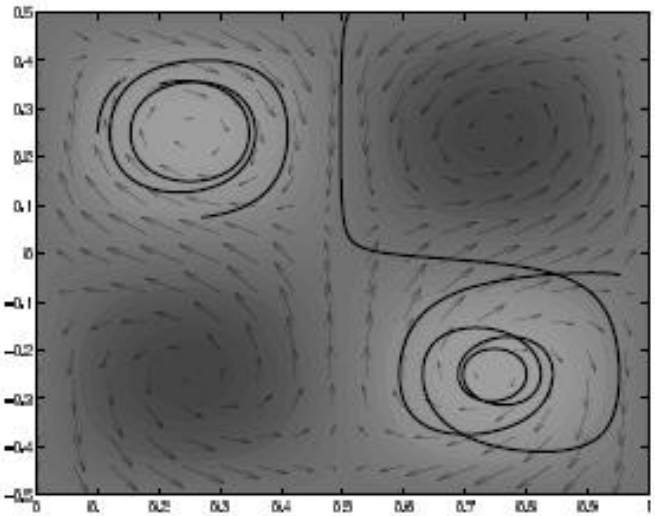
See: Data Assimilation: Mathematical and Statistical Perspectives, Apte, J, Stuart and Voss, IJNMF 2008

Perturbed Cellular Flow Field

$$\frac{\partial u}{\partial t} = v - \frac{\partial h}{\partial x},$$
$$\frac{\partial v}{\partial t} = -u - \frac{\partial h}{\partial y},$$
$$\frac{\partial h}{\partial t} = -\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y},$$

$$u(x,y,t) = -2\pi l \sin(2\pi kx)\cos(2\pi ly)u_0 + \cos(2\pi my)u_1(t),$$
$$v(x,y,t) = 2\pi k \cos(2\pi kx)\sin(2\pi ly)u_0 + \cos(2\pi my)v_1(t),$$
$$h(x,y,t) = \sin(2\pi kx)\sin(2\pi ly)u_0 + \sin(2\pi my)h_1(t),$$

$$\dot{u_0} = 0,$$
$$\dot{u_1} = v_1,$$
$$\dot{v_1} = -u_1 - 2\pi m h_1,$$
$$\dot{h_1} = 2\pi m v_1,$$

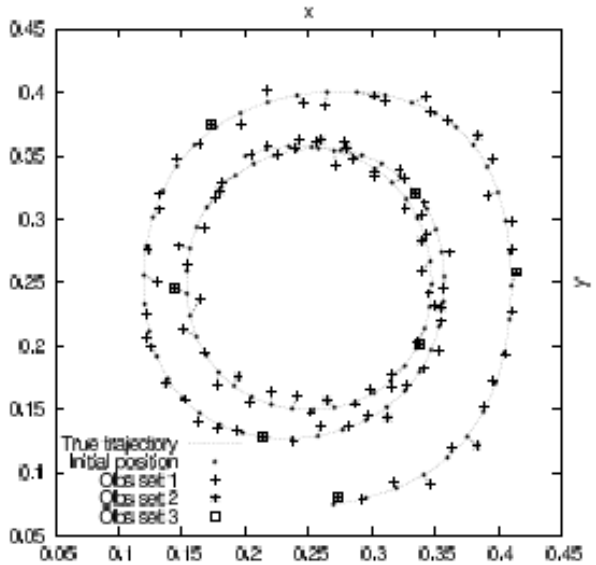


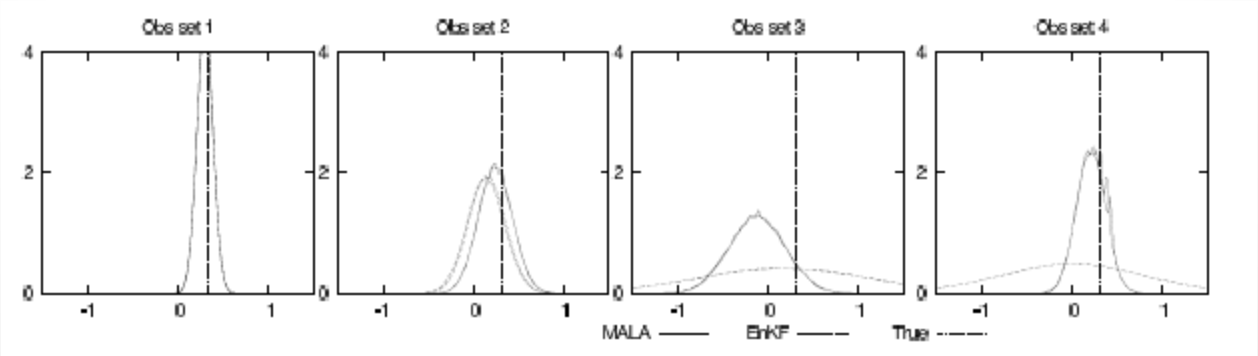
$$\dot{x} = u(x,y,t)$$
$$\dot{y} = v(x,y,t)$$

Apte, Stuart and J., Tellus A 2008

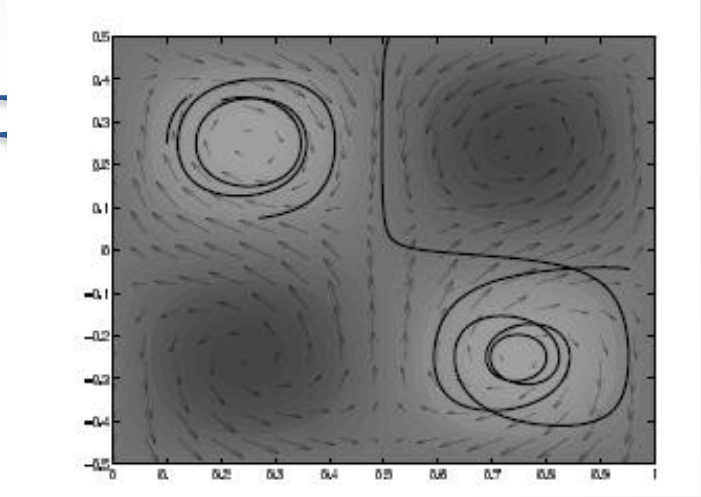
Assimilating from trajectory staying in one cell

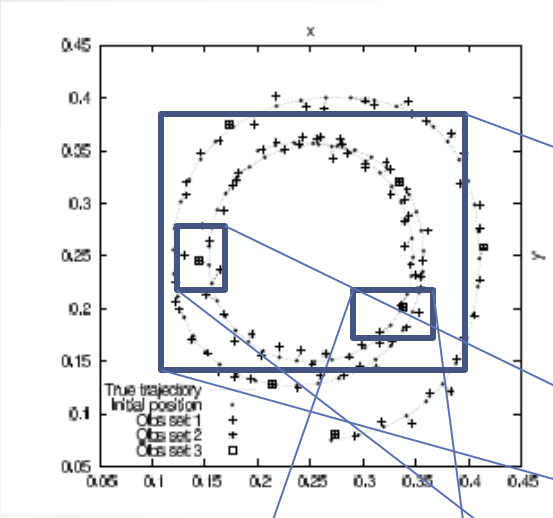
Compare:
- EnKF
- Metropolis-Hastings

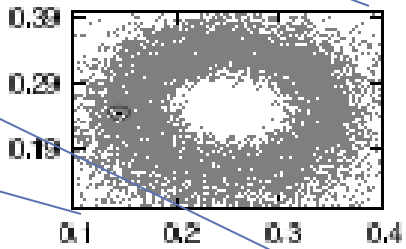




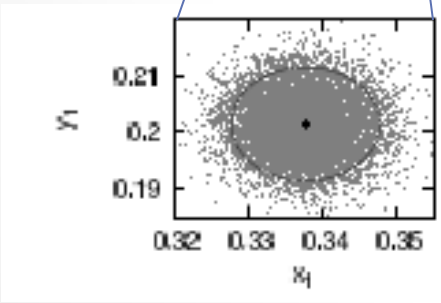
Problem with EnKF



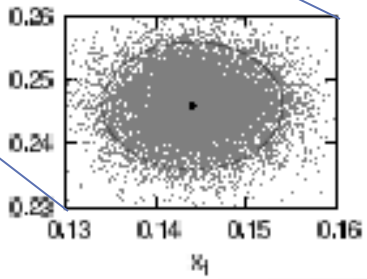




Before second observation



After first observation and assimilation



After SECOND observation and assimilation

Climate Models

$$\frac{du}{dt} + fv = -\frac{1}{\rho}\frac{\partial P}{\partial x} + F^x,$$

$$\frac{dv}{dt} - fu = -\frac{1}{\rho}\frac{\partial P}{\partial y} + F^y,$$

$$0 = -g - \frac{1}{\rho}\frac{\partial P}{\partial z} + F^z,$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho U = 0,$$

$$\frac{\partial T}{\partial t} + \nabla \cdot TU = F^Q,$$

$$\frac{\partial S}{\partial t} + \nabla \cdot SU = F^S.$$

$$\rho = \rho(P, T, S, \ldots),$$

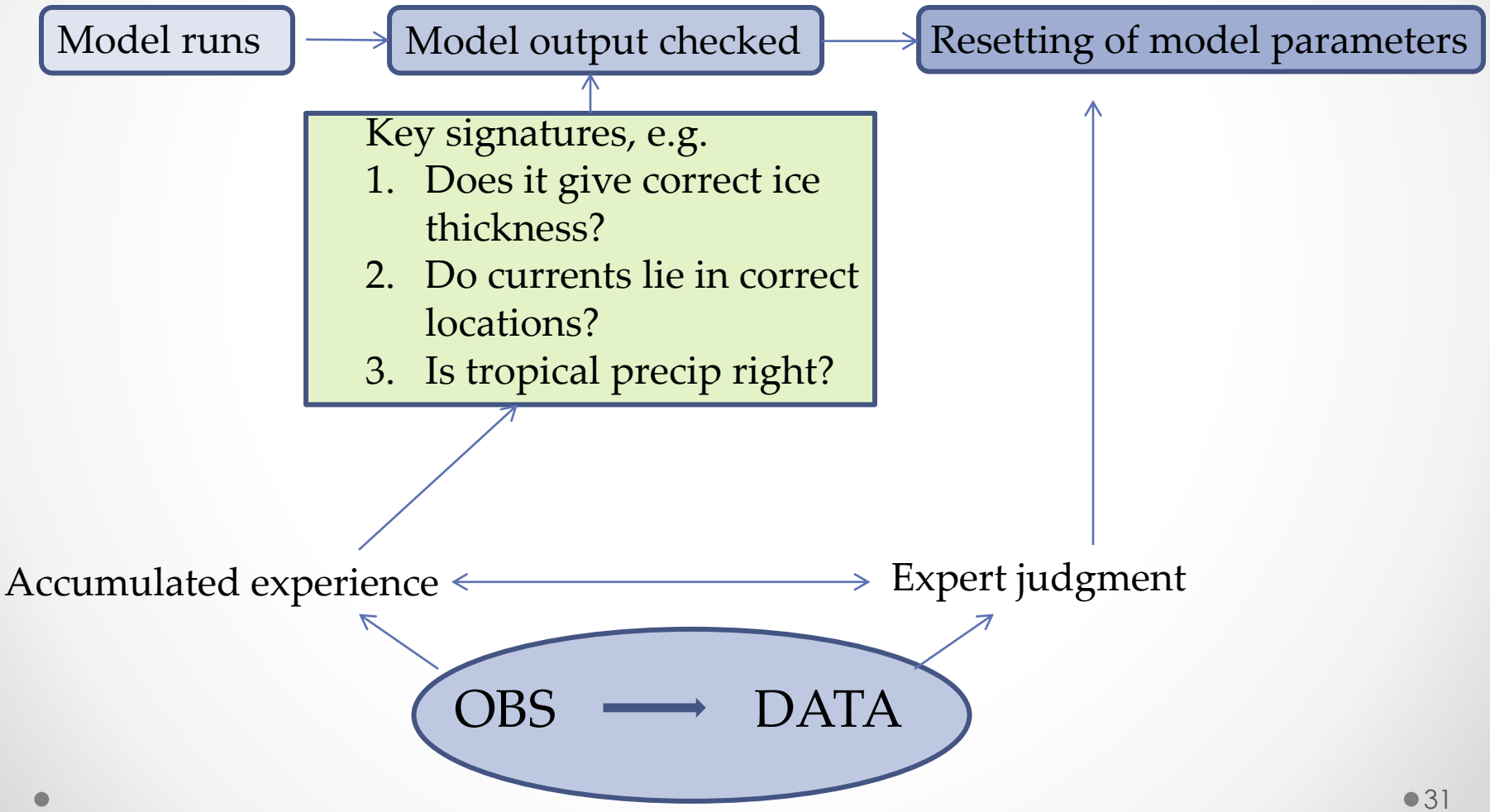
Primitive eqns of GFD solved on grid

Dynamical core of model

MUCH of climate science is devoted to setting the forcing terms. This includes:

1. Parametrising sub-gridscale processes
2. Parametrising missing effects
3. Information flow to other parts of model
4. Setting ambient conditions (radiative forcing)
5. Tuning the model

Tuning the Model



Preconceptions

1. Uncertainty is an intrinsic property of "The" Model, or a collection of models, and it can be found by just looking at them hard enough

Uncertainty of the model can have its origin in many places: data used in tuning, expert judgment etc. Perhaps we need to unpack the model?

2. Data assimilation's primary use is to get the "best" estimate

Data assimilation tells us how to put information together from models and data how the cumulative error is formed. The error may be arbitrarily complex and the issue is how to sample the PDF

- DA could be used in tuning climate models
- Its use would be similar to the way the paleoclimate work by van der Schrier and Barkmeijer
- If done from a Bayesian perspective, the procedure would have implications for uncertainty due to model tuning.

Challenges

- DA is nowhere near ready to be applicable to climate
- Need to develop DA to deal with
  - High dimensions
  - Nonlinearity
  - Sparse data
  - Multiple timescales (and space)
- Shift perspective from getting "best estimate" of truth to finding PDF that represents all available information