Data Assimilation for Earth System Models

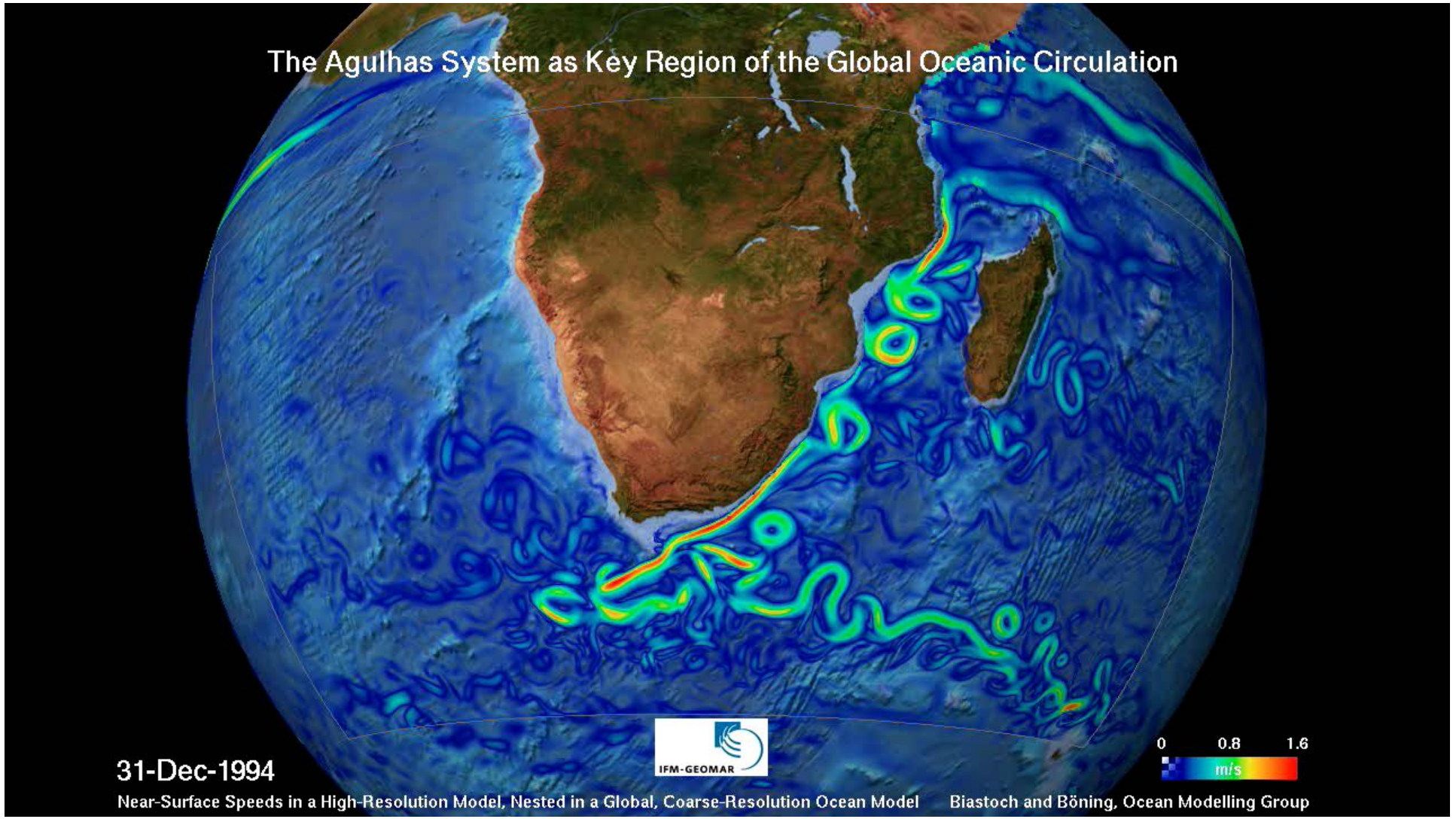
Peter Jan van Leeuwen

Department of Atmospheric Science, Colorado State University

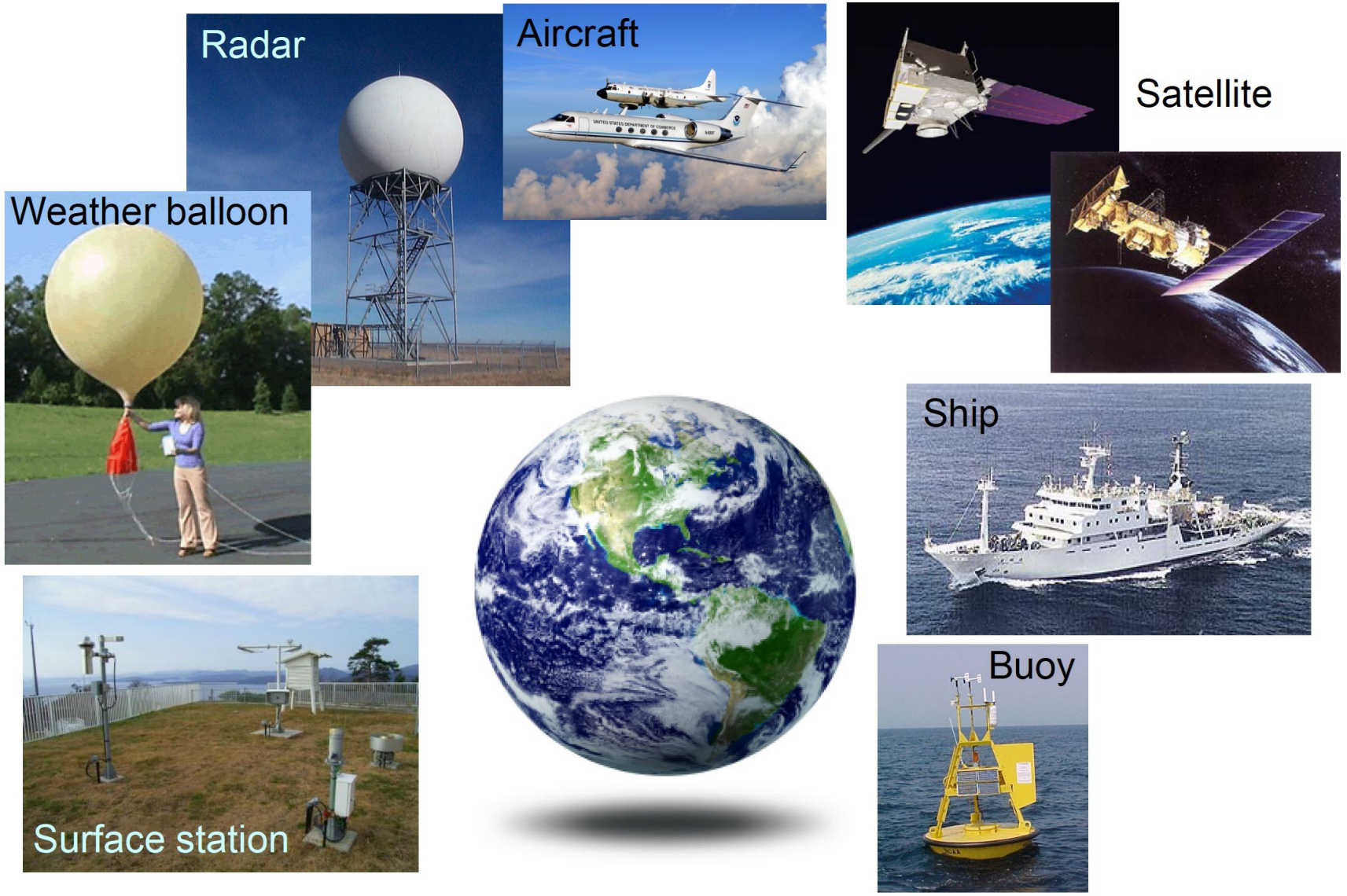
Why DA for Earth System Models?

- Improve S2S and climate predictions,
- Construct re-analyses of four-dimensional climate system to increase scientific understanding (think ERA5, the most used data set in the atmospheric sciences)
- Earth system model improvement:
  - estimate Earth system model parameters
  - estimate Earth system model parameterizations/missing physics (chemistry, biology,…)
- Intelligent monitoring: where should we measure what?

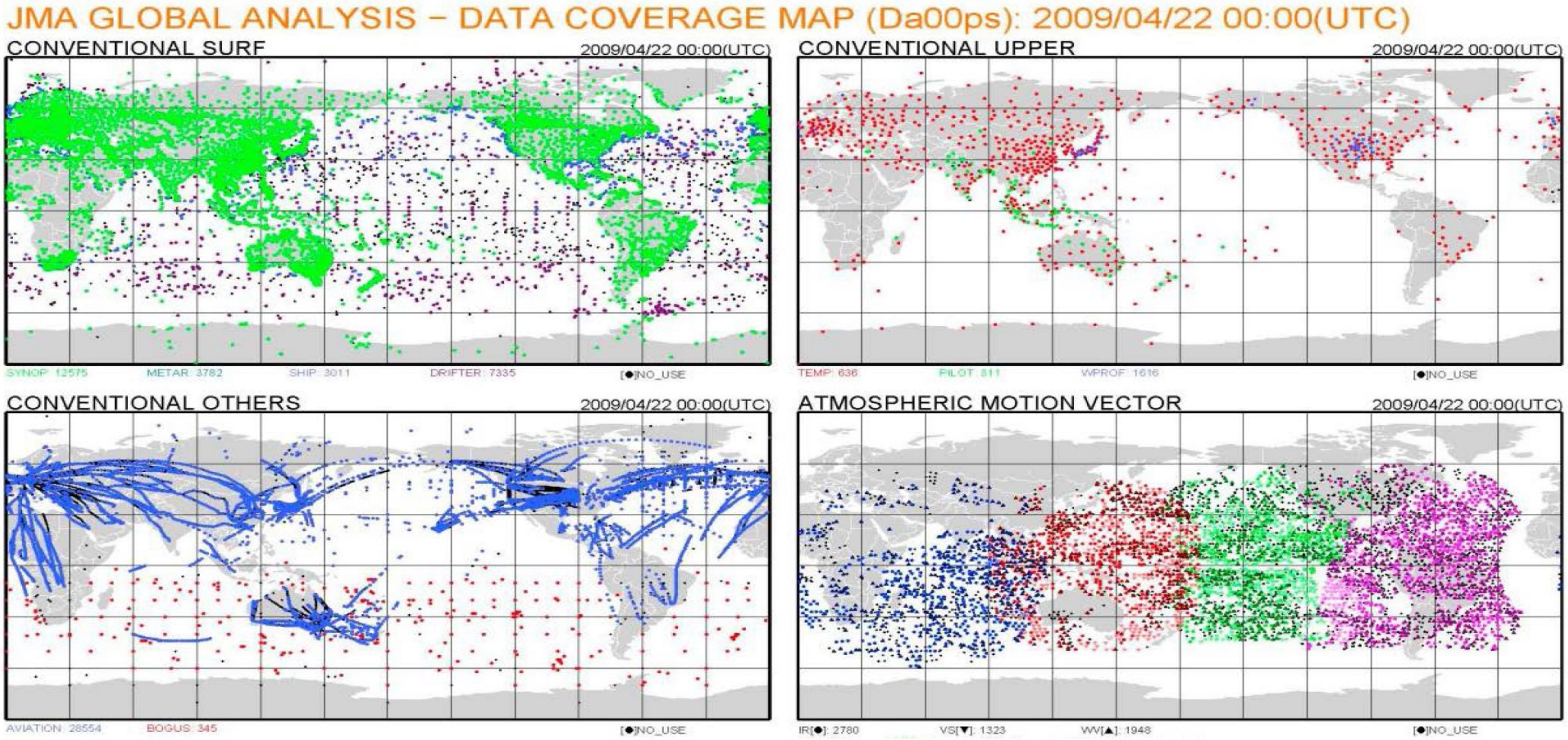
Example of an ocean model: chaotic dynamics



The Global observing system for weather prediction

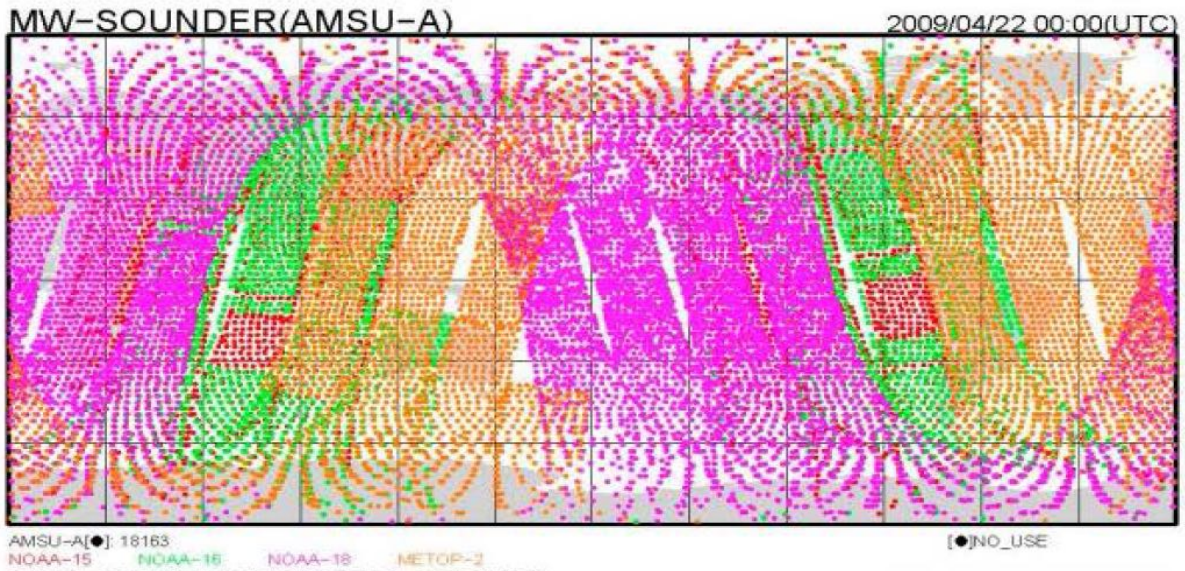


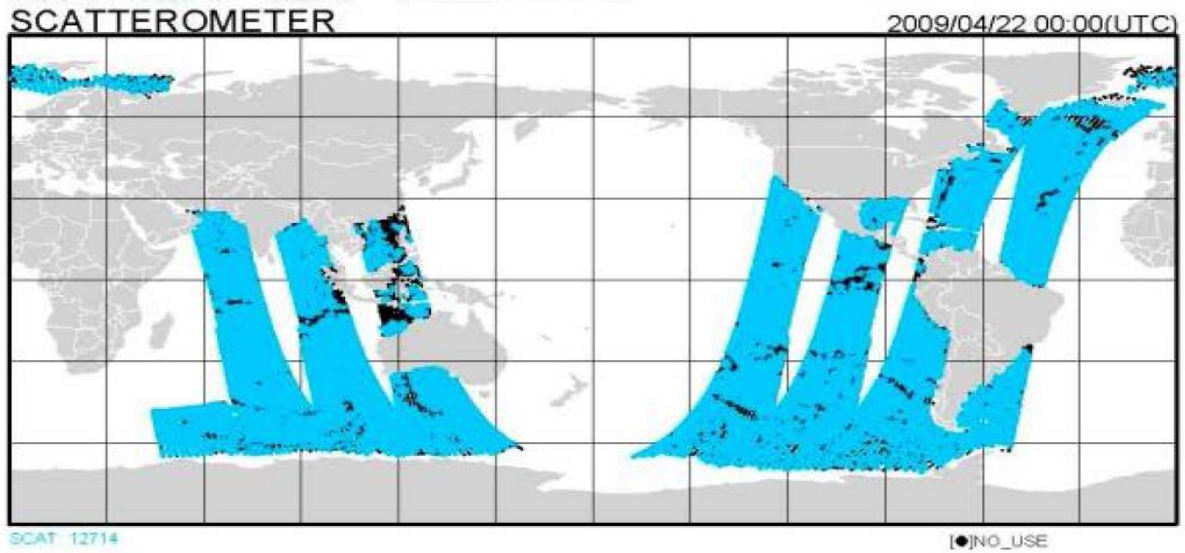
Observation coverage

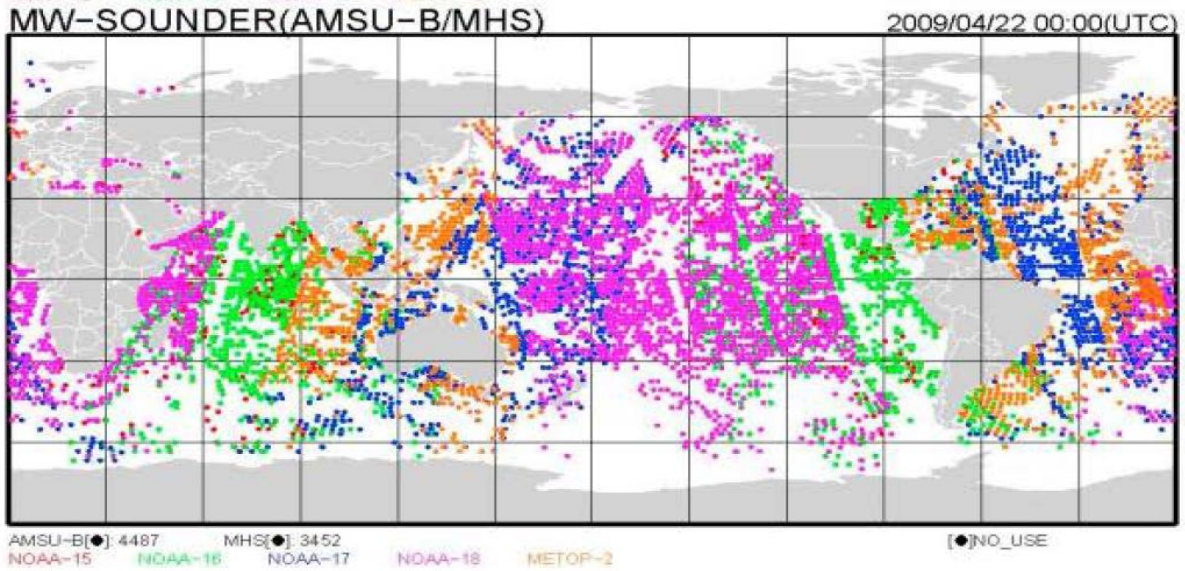


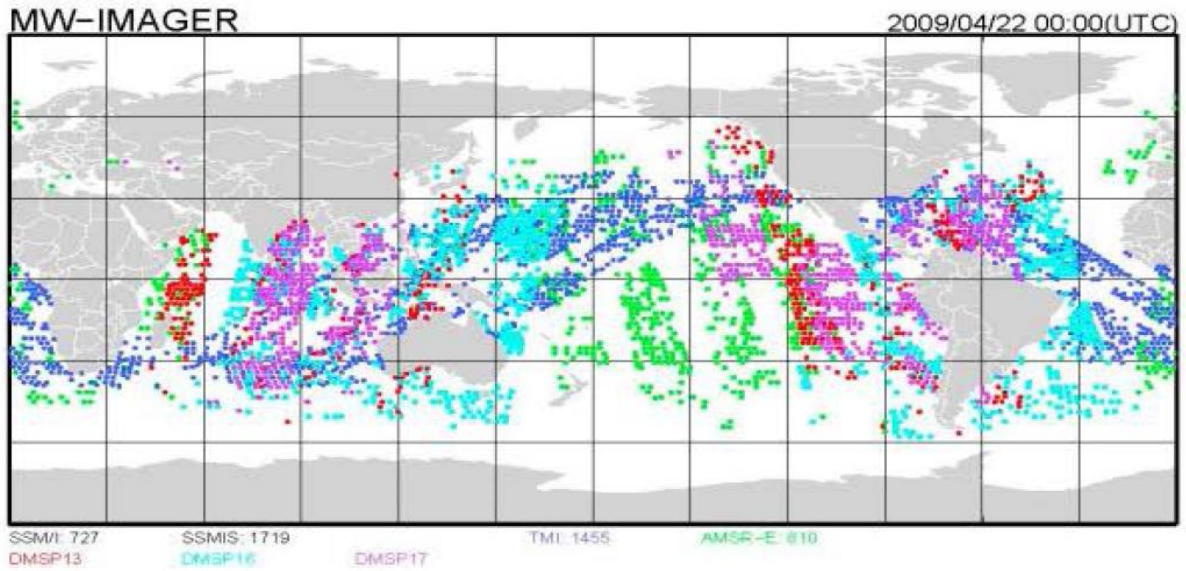
World's effort! (no border in the atmosphere)

Observation coverage satellites

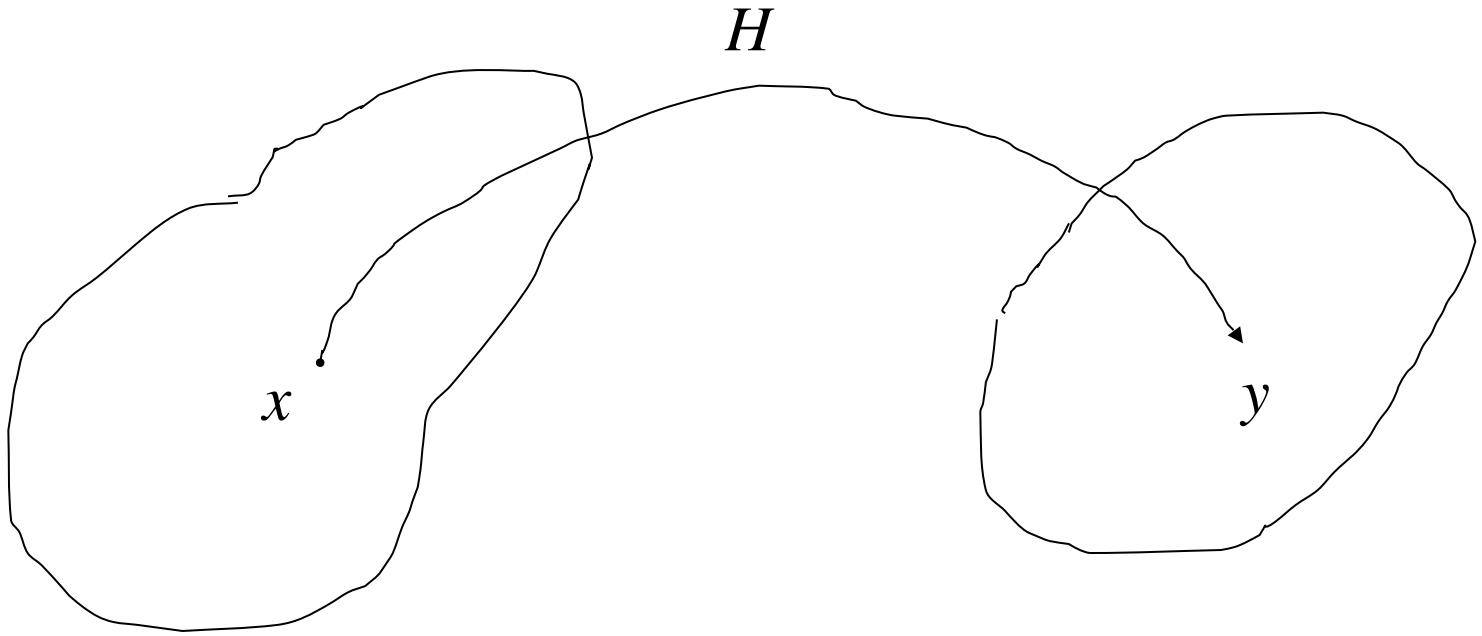








How to connect observations to model variables?



$$y^n = H(x^n) + \epsilon^n$$

The function *H* is called the *observation operator.*

Data assimilation: the basics

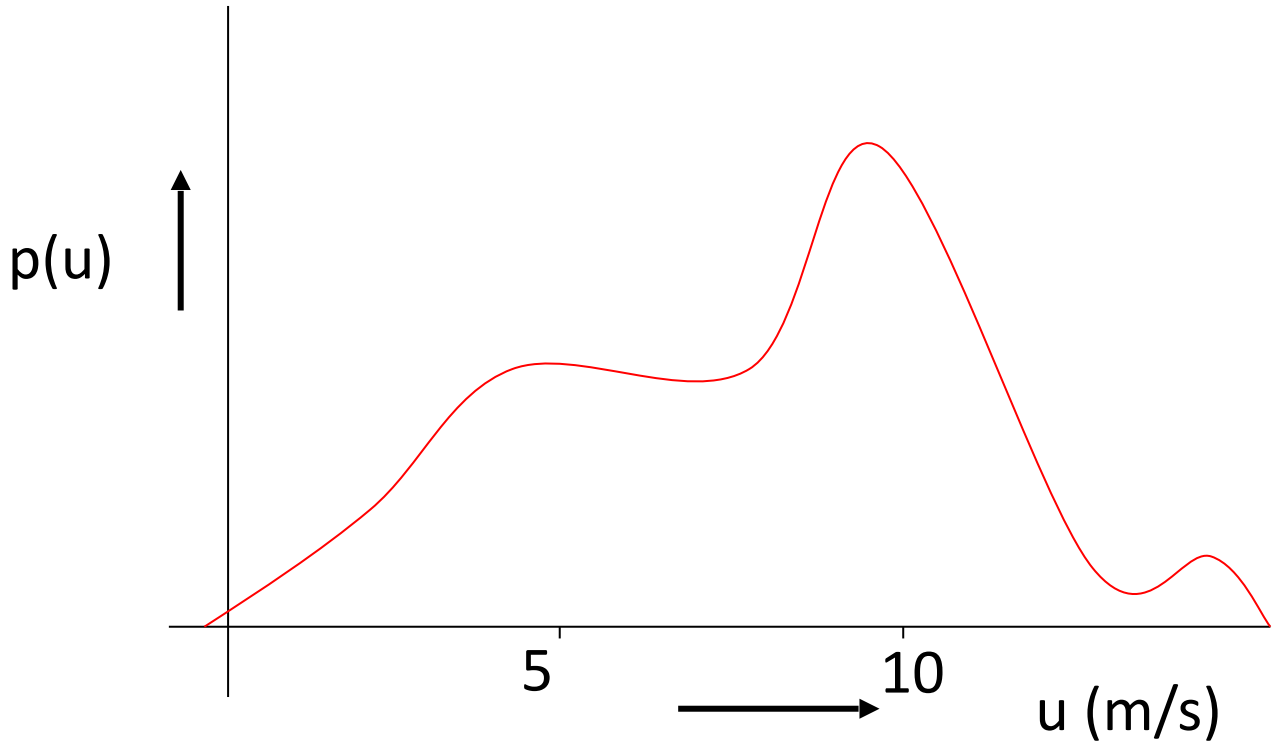
Data assimilation is the science of combining many sources of information:

from 1) a numerical model and 2) many observations.

These sources have their uncertainty, and this uncertainty is crucial when we combine the sources of information

The most general way to incorporate uncertainties is via probability density functions.

Mathematical description: probability density functions (pdf)



The variable *x* can be ‘anything’: the state of the atmosphere, a concentration field, sources and sinks, a trajectory of a concentration field, model parameters, or combinations of these. Or it can be a whole Earth system model.

Intermezzo: conditional pdf

Conditional pdf: $$p(x, y) = p(x|y)p(y)$$

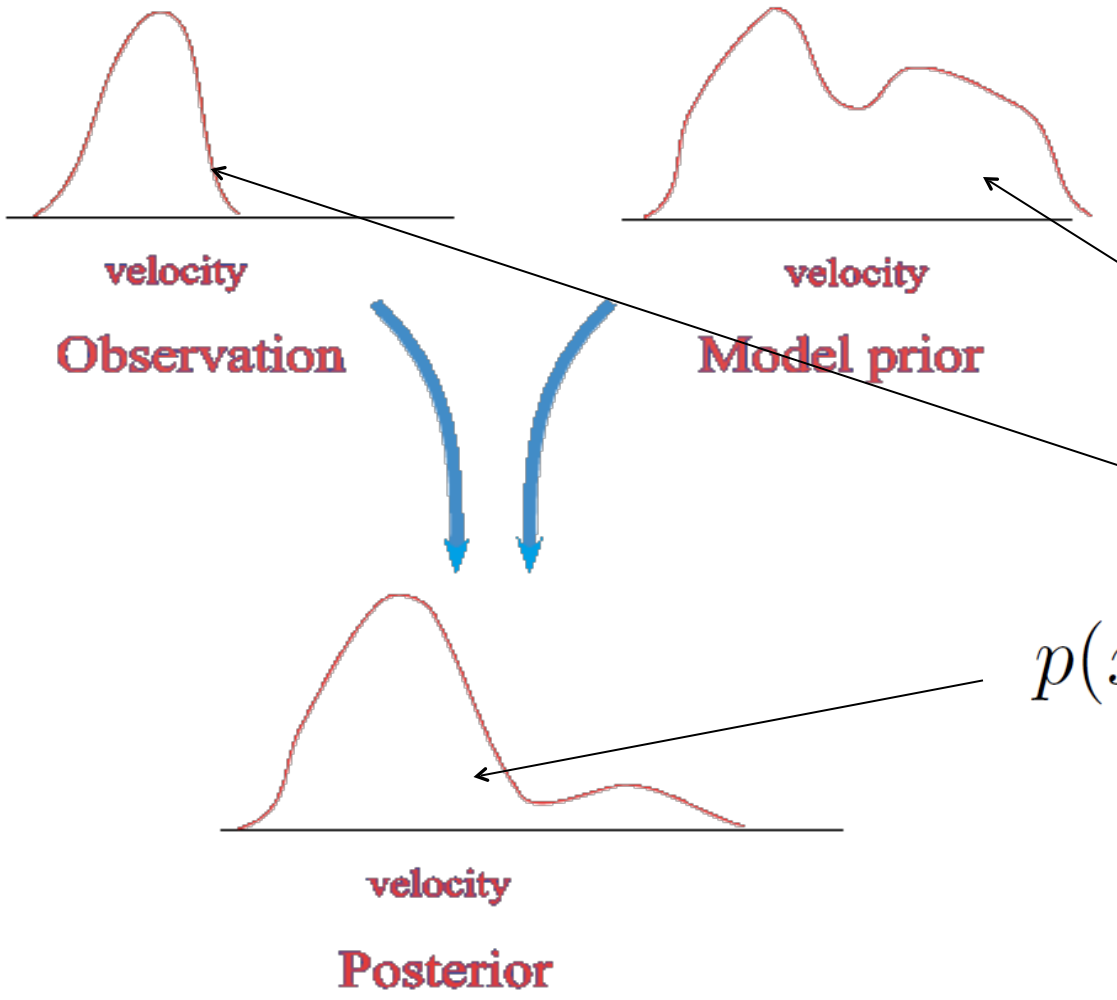
'The probability of X=x and Y=y is equal to prob of X=x given Y=y, times the probability that that event Y=y occurs.'

Similarly: $$p(x, y) = p(y|x)p(x)$$

Combine: $$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

This is Bayes Theorem, the basis of data assimilation, machine learning, any estimation problem!

Data assimilation: general formulation



Bayes theorem:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

The solution is a pdf!

No inversion!

The likelihood

The likelihood is given by:

$$p(y|x)$$

The observations are generated as

$$y = H(x_{true}) + \epsilon_{true}$$

with $p(\epsilon)$ known, e.g.

$$p(\epsilon) = N(0, R)$$

How do we find an expression for the likelihood from this? Note that *x* is the active variable, while the observations *y* are given when we do data assimilation. Hence, the likelihood is NOT the pdf of the observations!

What is the Likelihood ?

$p(y|x)$ is the probability (density) that we find observations y, given that the model state is x. We can write

$$y = H(x) + \epsilon \quad \text{or} \quad \epsilon = y - H(x)$$

where x and y are known. Hence, this is equal to the probability (density) that this measurement error epsilon ϵ appears, which is equal to $p(\epsilon)$. We thus find, for Gaussian observation errors:

$$p(y|x) = p(\epsilon) = A \exp\left(-\frac{1}{2}\epsilon^T R^{-1}\epsilon\right)$$

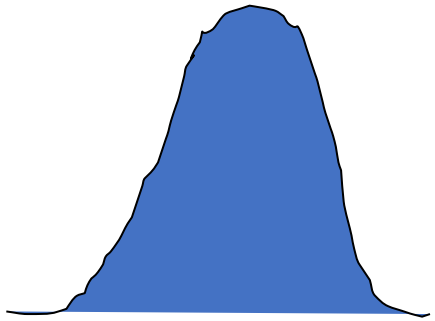
$$= A \exp\left[-\frac{1}{2}\left(y - H(x)\right)^T R^{-1}\left(y - H(x)\right)\right]$$

Big Data

- How big is the nonlinear data-assimilation problem?
- Assume we need 10 frequency bins for each variable to build the joint pdf of all variables.
- Let's assume we have a modest model with a million variables.
- Then we need to store $10^{1,000,000}$ numbers.
- The total number of atoms in the universe is estimated to be about 10^{80}.
- So, the data-assimilation problem is larger than the universe…

- And data assimilation becomes the **art** of finding the best **approximate** method for the problem at hand.

The Gaussian assumption



$$p(T) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\left(T-\overline{T}\right)^2}{2\sigma^2}\right)$$

Prior pdf: multivariate Gaussian:

$$p(x) \propto \exp\left[-\frac{1}{2}(x-x_b)^T B^{-1}(x-x_b)\right]$$

Likelihood: multivariate Gaussian

$$p(y|x) \propto \exp\left[-\frac{1}{2}(y-H(x))^T R^{-1}(y-H(x))\right]$$

(Ensemble) Kalman Filter/Smoother I

Use Gaussianity in Bayes Theorem:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Multiplication, assuming observation operator *H* is linear:

$$p(x|y) \propto \exp\left[-\frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) - \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx)\right]$$

Complete the squares to find again a Gaussian (only for linear H !):

$$p(x|y) \propto \exp\left[-\frac{1}{2}(x - x_a)^T P^{-1}(x - x_a)\right]$$

(Ensemble) Kalman Filter/Smoother II

Two possibilities to find the expressions for the mean and covariance:

1) Completing the squares
2) Assume solution is linear combination of model and observations.

Both lead to the Kalman filter equations, which is just the least squares solution (best linear unbiased estimator, BLUE):

Posterior mean: $$x_a = x_b + K(y - Hx_b)$$

Posterior covariance: $$P = (I - KH)B$$

with Kalman gain: $$K = BH^T(HBH^T + R)^{-1}$$

(Ensemble) Kalman Filter/Smoother II

Two possibilities to find the expressions for the mean and covariance:

1) Completing the squares
2) Assume solution is linear combination of model and observations.

Both lead to the Kalman filter equations, which are just the least squares solutions (best linear unbiased estimator, BLUE):

$$x_a = x_b + BH^T(HBH^T + R)^{-1}(y - Hx_b)$$

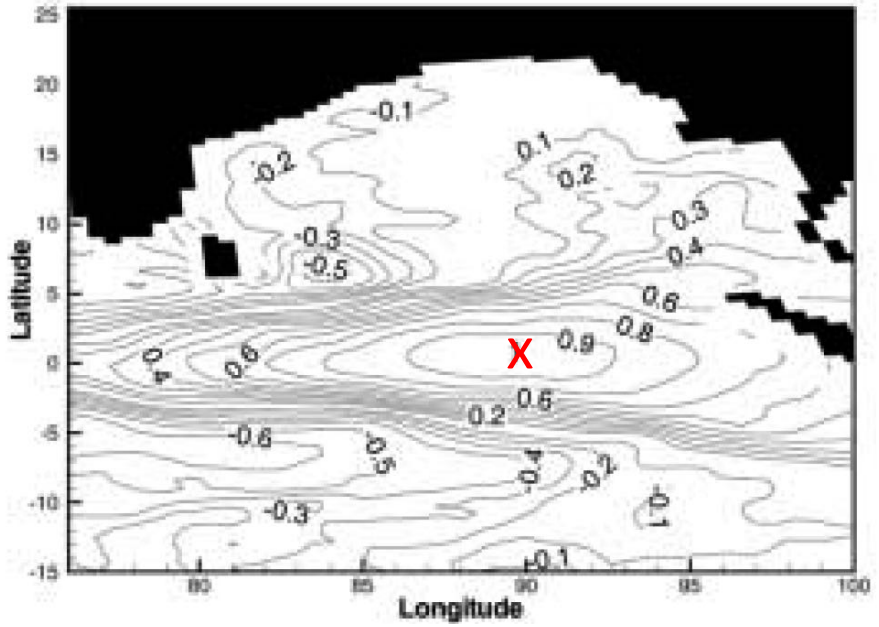
influence region — weighting — innovation

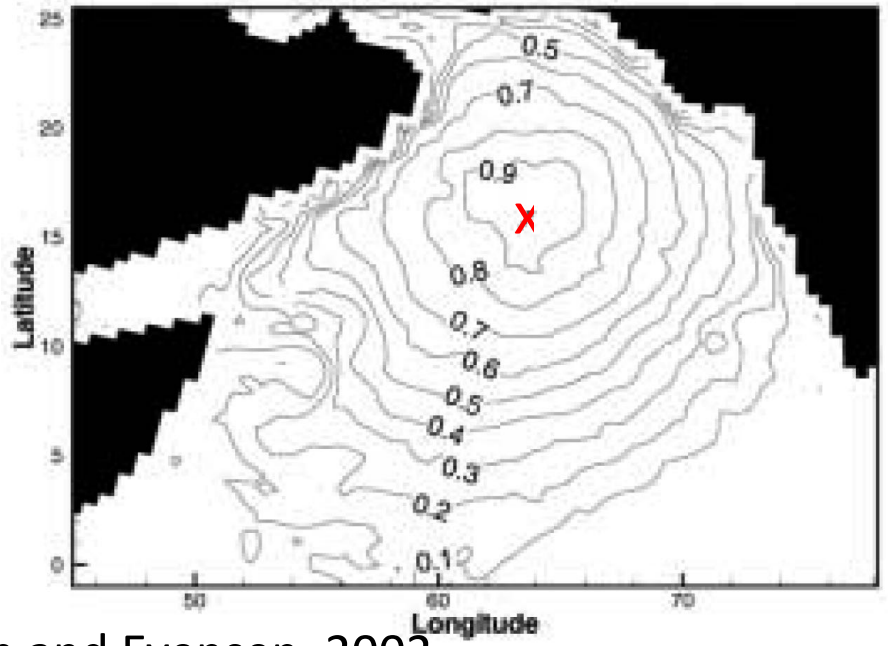
$$K = BH^T(HBH^T + R)^{-1}$$

The model error covariance:

Tells us how model variables co-vary.

Examples of spatial correlation of SSH at X and SST in the Indian Ocean





In the Kalman filter this comes in via the BH^T term:

$$x_a = x_b + BH^T(HBH^T + R)^{-1}(y - Hx_b)$$

Note, can also do this over time!

Haugen and Evensen, 2002

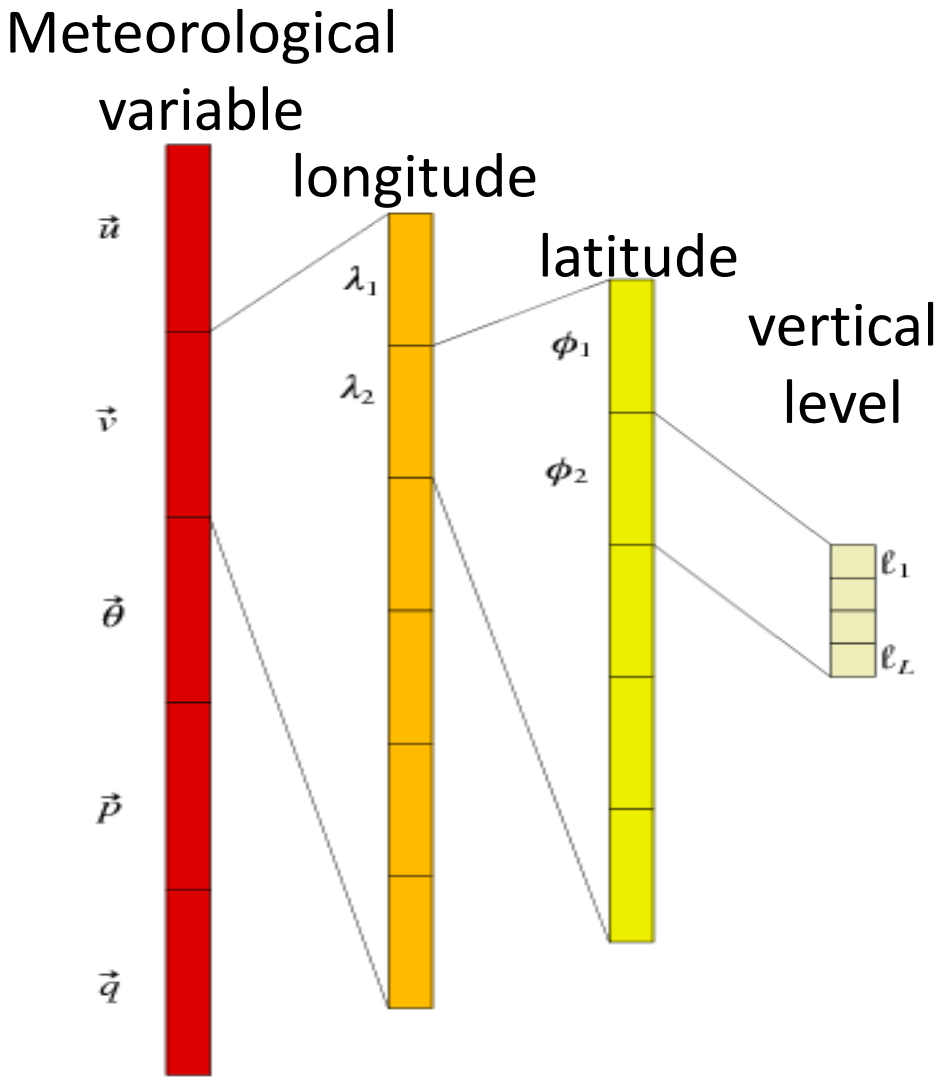
Kalman Filter in high-dimensions…

We need the mean and the covariance for the Kalman Filter.

For numerical weather prediction the state dimension is 10^{11}:

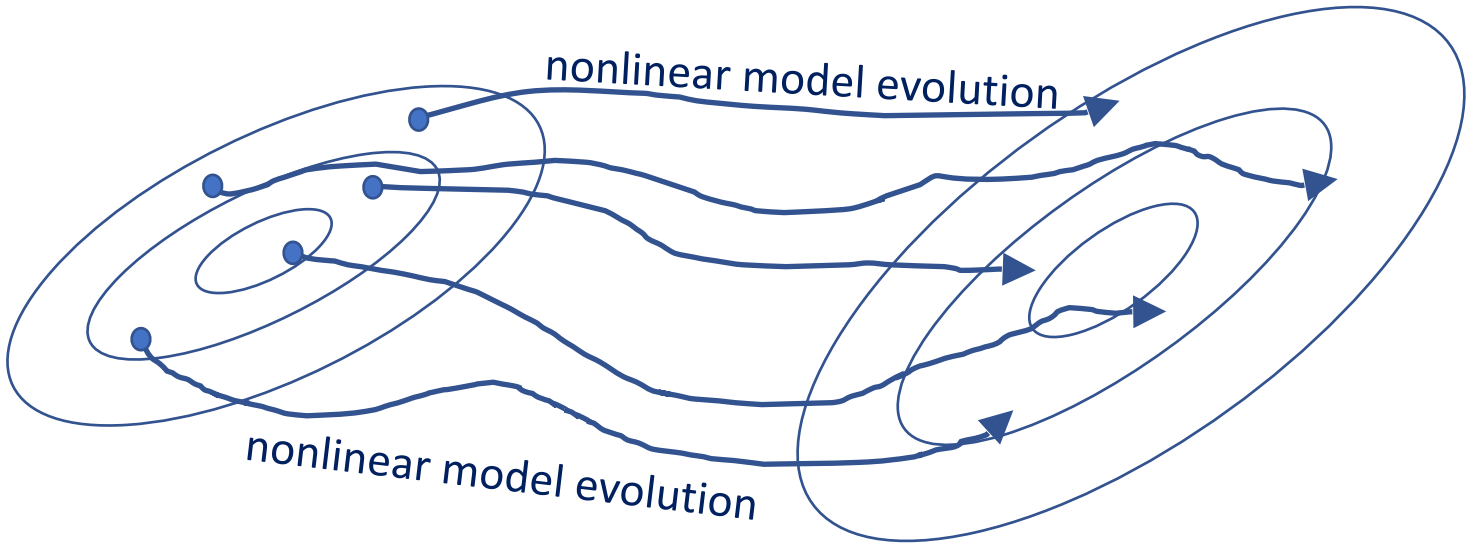
That means a covariance matrix with 10^{22} elements.

No computer today can store that….



Ensemble Kalman filters and smoothers

Use ensemble to store and propagate the mean and covariance matrix.



Ensemble Kalman Filter: $BH^T = X_{t2}\,(HX_{t2})^T$

Ensemble Kalman Smoother: $BH^T = X_{t1}\,(HX_{t2})^T$

Many variants: SEnKF, SEnKS, ETKT, EAKF,…

Can allow for weakly nonlinear *H* via iterative variants.

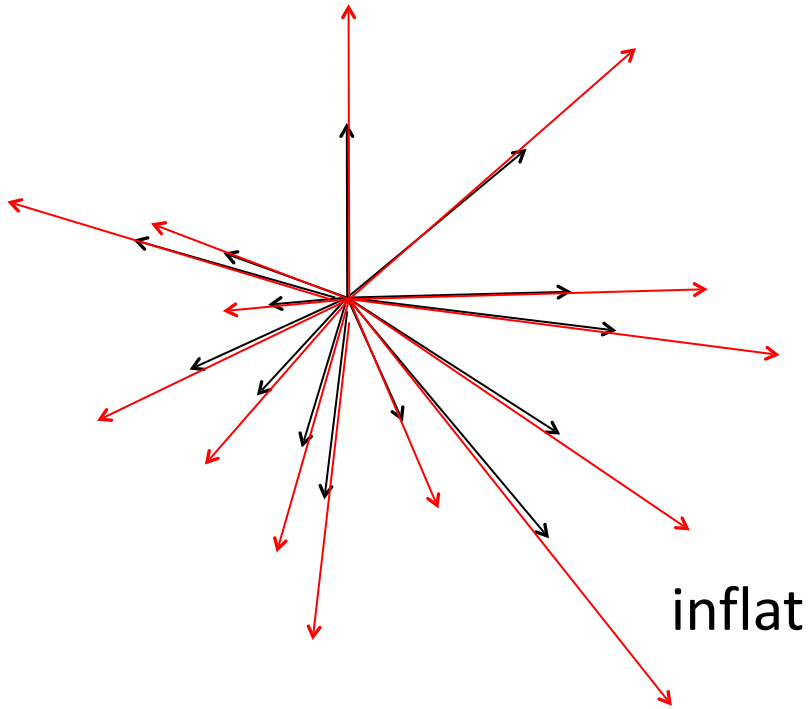
Issues Ensemble Kalman filters/smoothers

Two effects of finite sample size:

- Underestimation of sample covariance.
- Spurious long-range correlations.

Fixes:

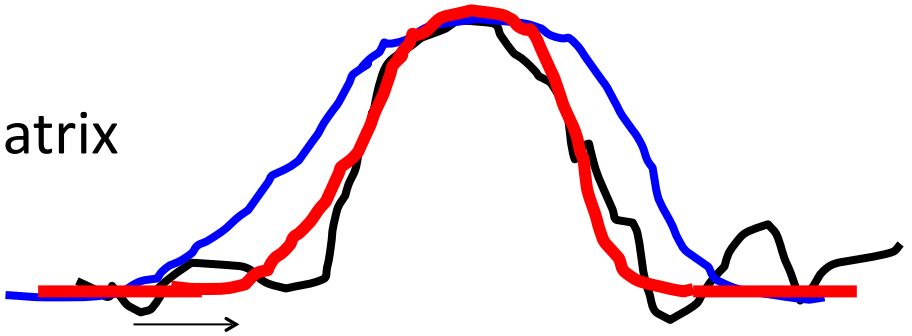
- Covariance inflation
- Covariance localization



inflation

Localization: multiply with smooth correlation matrix

$$B_{loc} = C \circ B$$



Variational methods

A variational method looks for the most probable state, which is the maximum of this posterior pdf also called the mode.

Instead of solving for the maximum, one solves for the minimum of a so-called costfunction.

The posterior pdf can be rewritten as $p(x|y) \propto \exp\left[-\frac{1}{2}J\right]$

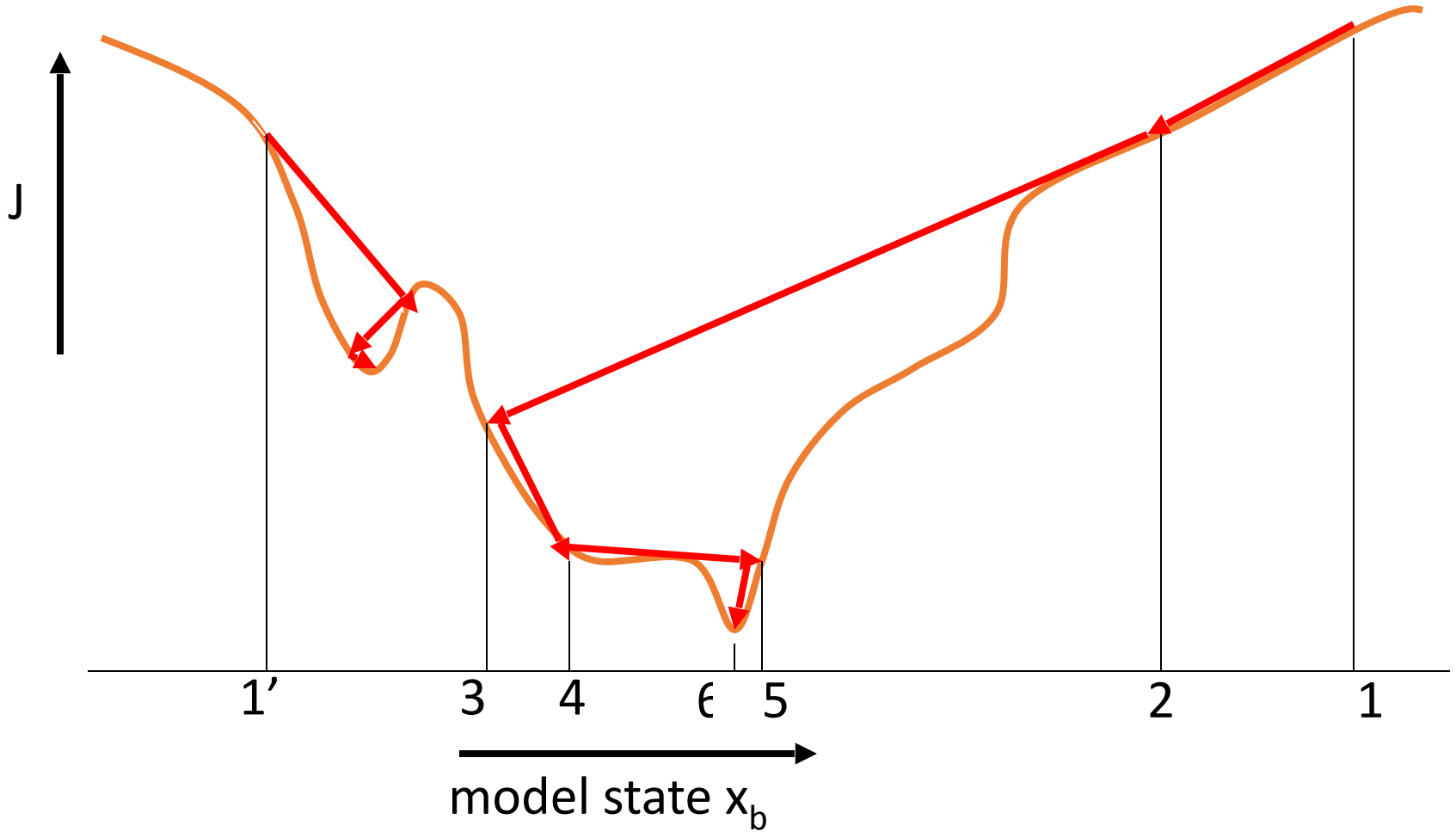
in which *J* is costfunction or penalty function

$$J = (x - x_b)^T B^{-1}(x - x_b) + (y - H(x))^T R^{-1}(y - H(x))$$

in which *B* has to be determined from climatological physics.

Find *min J* from variational derivative: $\frac{\delta J}{\delta x} = 0$, leading to 3DVar

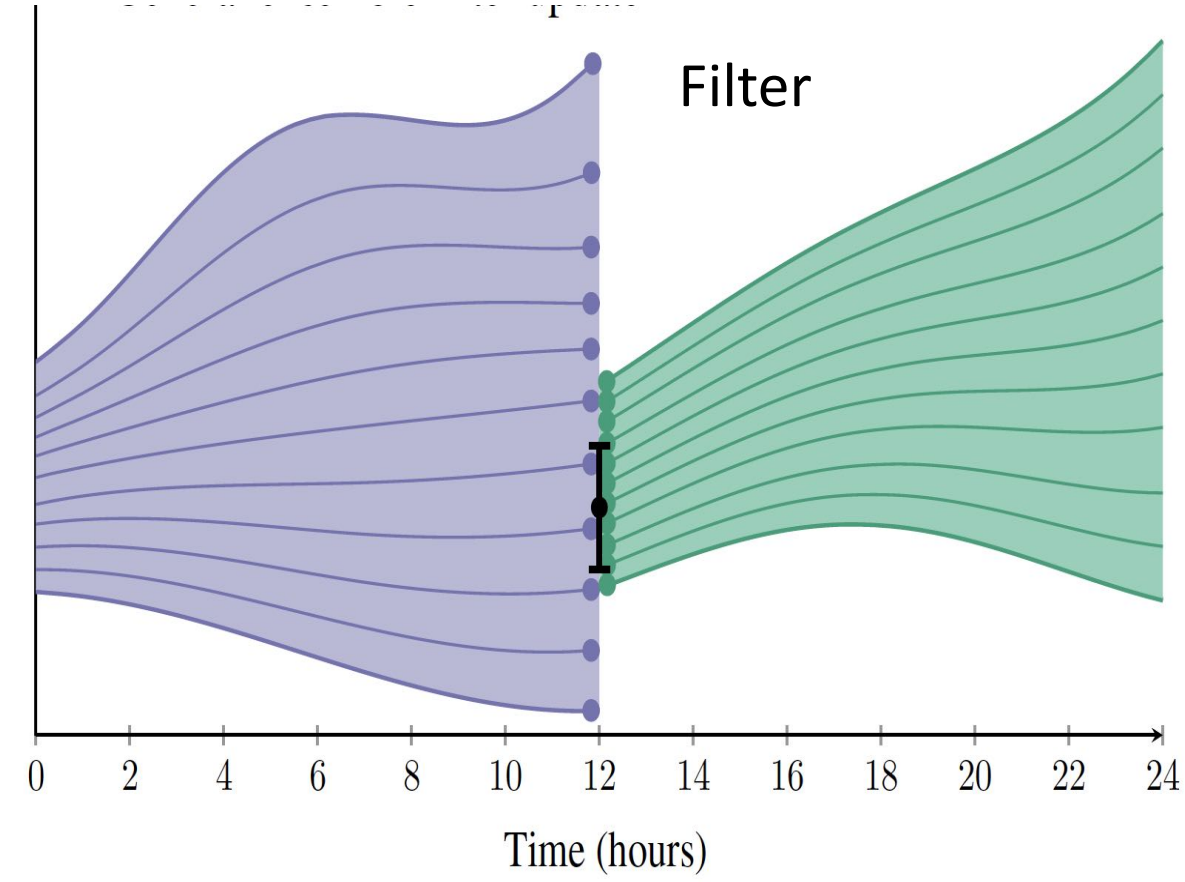
Gradient descent methods: e.g. Gauss-Newton iterations

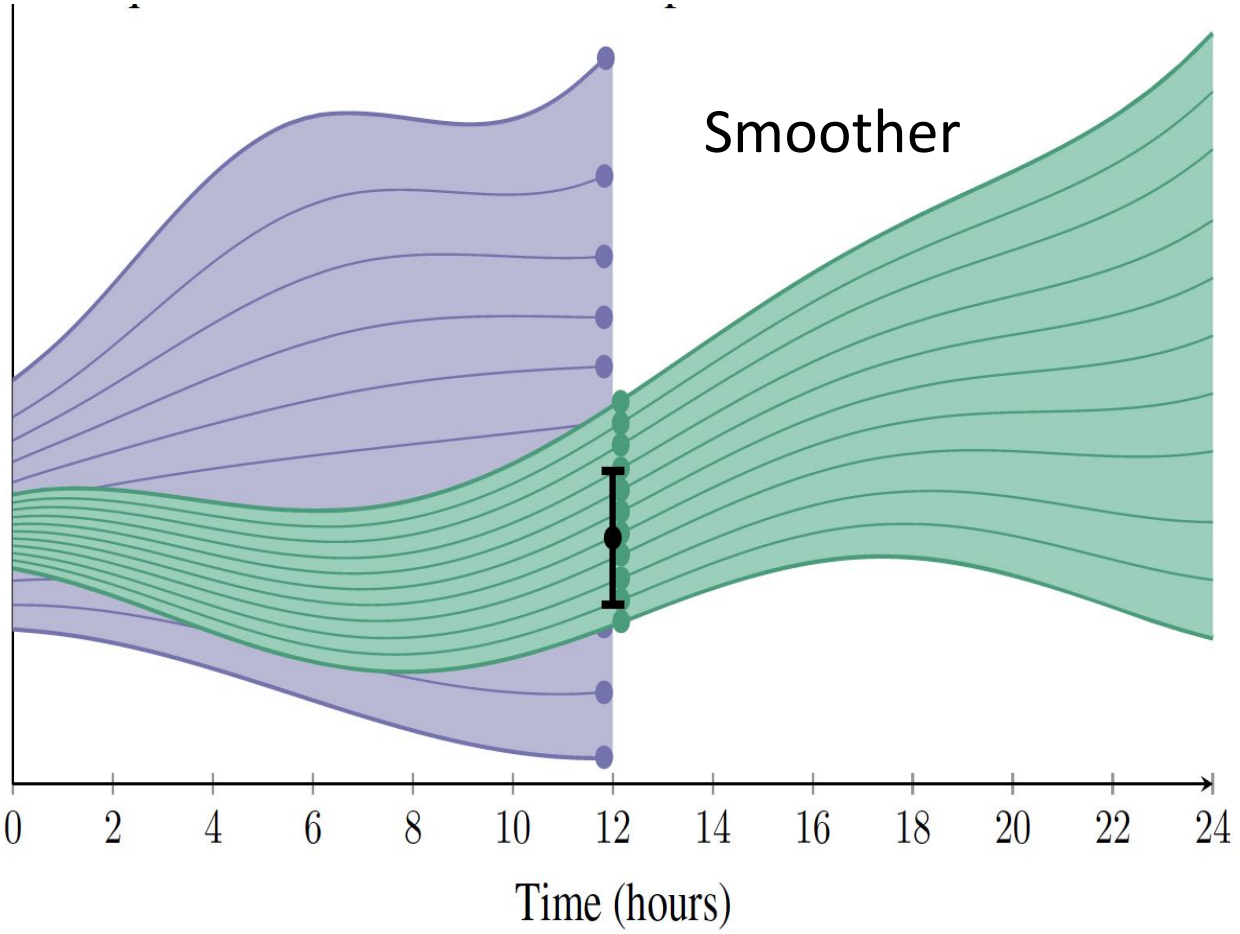


Sequential data assimilation

Two ingredients: $x(t+1) = M(x(t)) + error.$ $y(t) = H(x(t)) + error$

Two modes of data assimilation:





4DVar: include the time dimension

The total costfunction that we must minimize now becomes:

$$J = (x - x_b)^T B^{-1}(x - x_b) + \sum_{t_{obs}=1}^{M} (y_i - H_i(x))^T R^{-1}(y_i - H_i(x))$$

in which the observation operator H_i contains the forward model:

$$H_i(x) = H(M_{0\rightarrow i}(x))$$

This nonlinear costfunction is again minimized iteratively.

4Dvar has been the workhorse for weather forecasting for the last 20 years.
(Machine learning can be seen as special kind of 4DVar, with many simplifications.)

Popular data-assimilation methods

(Iterative) Ensemble Kalman Filters

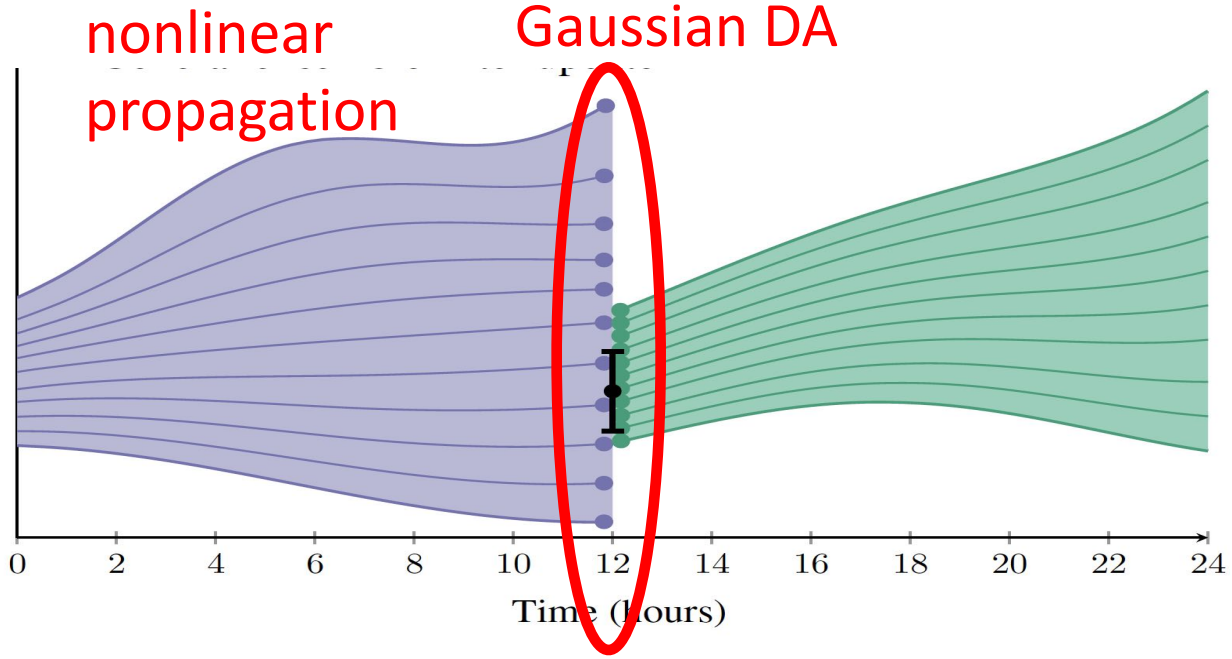
- Ensemble size typically too small, need localization and inflation

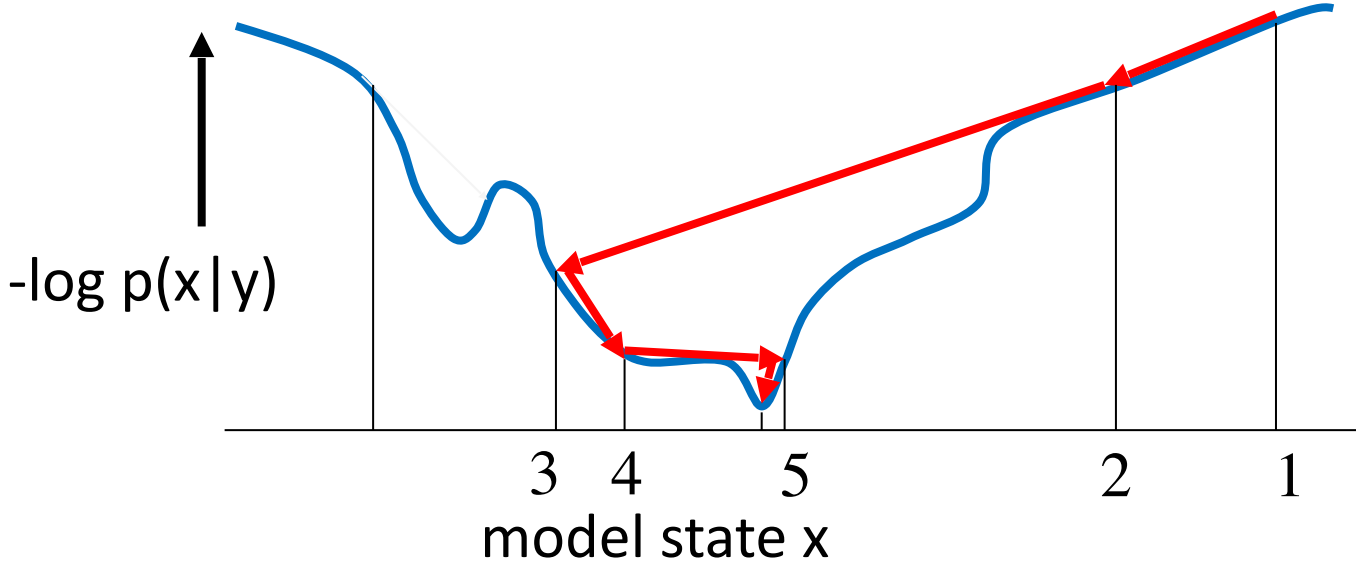
Variational methods find mode of posterior pdf, Gauss-Newton iteration

- Gaussian prior with **fixed** prior covariance, *H* can be weakly nonlinear, obs errors Gaussian
- Hard to find uncertainty estimate

'Hybrid' methods e.g. ECMWF uses ensemble of variational members

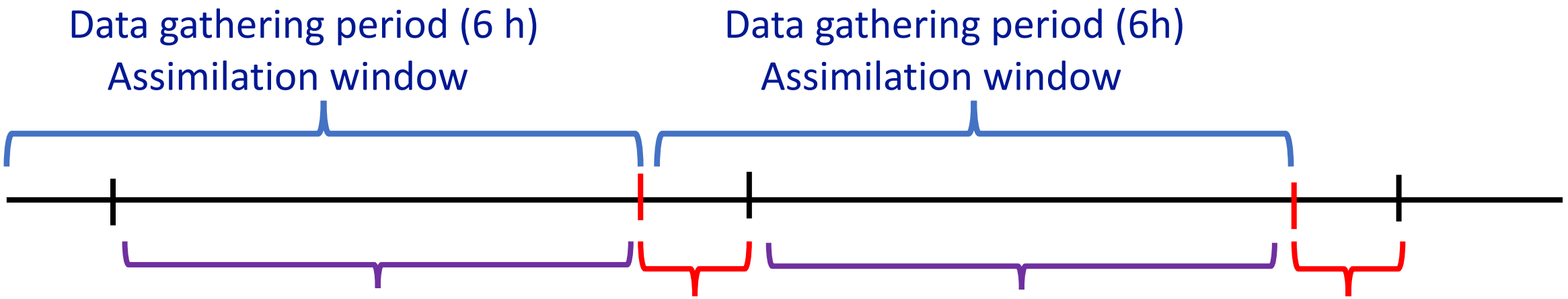
Nonlinear data-assimilation methods





Weather Prediction: 4Dvar

Atmospheric state vector of dimension 10^{11}.
Observation vector 10^{8} (and this is only 5% of all the observations).

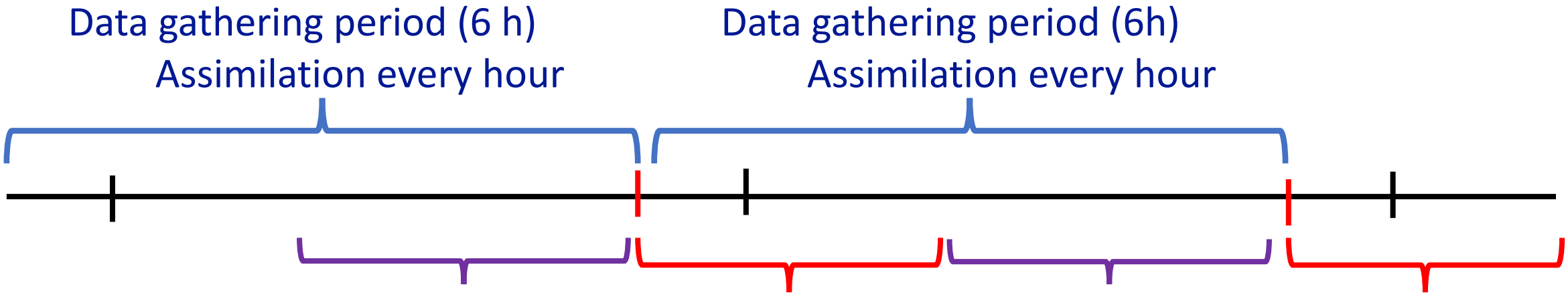


Computation: Model forecasts Data assimilation Model forecasts Data assimilation

4DVar (Gauss-Newton, first and second level preconditioning, prior covariance matrix, correlated observation errors, representation errors, …)

Weather Prediction: EnKF

Atmospheric state vector of dimension 10^{11}.
Observation vector 10^{6} (?) .



Computation: Model forecasts Data assimilation Model forecasts Data assimilation

EnKF (Inflation and localization for prior covariance matrix, uncorrelated observation errors, representation errors, …)

Ocean-Atmosphere (coupled) data assimilation

Approaches:	Issues:
1) EnKF separate	1) Obs of system 1 cannot influence other system directly
2) 3DVar separate	2) Obs of system 1 cannot influence other system directly Prior model covariance?
3) EnKF whole system	3) Localization radius?
4) 4Dvar whole system	4) Prior model covariance? Time-scale separation?

Ocean-Atmosphere 4Dvar data assimilation

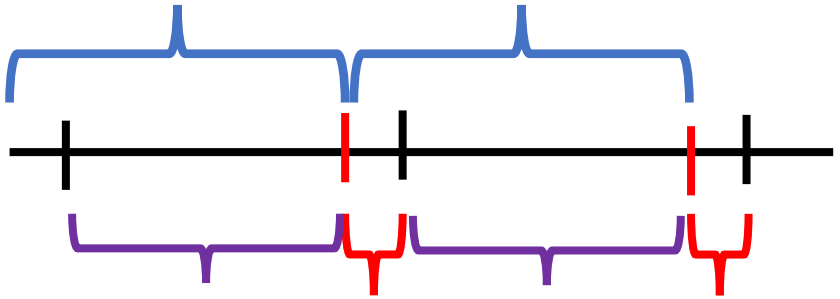
Solution strategies:

1. Separate 4Dvar in Atmosphere and Ocean
2. Strongly coupled over 12 hour
3. Strongly coupled over a few days with ’smoothed’ atmosphere
4. Weakly coupled: Exchange fields after each inner loop.

Not solved, active area of research.

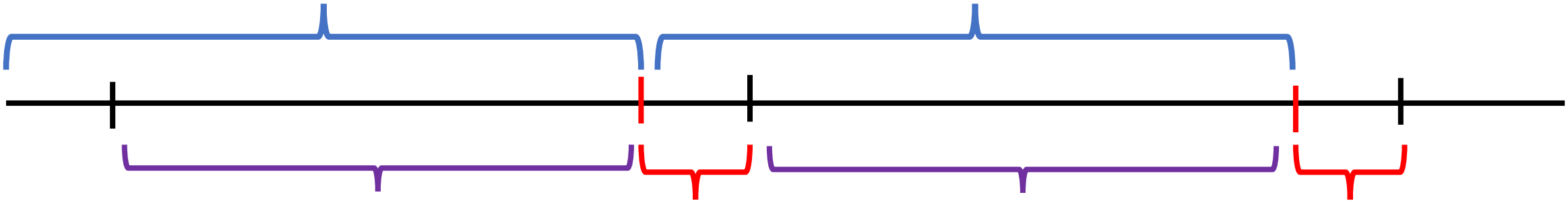
Ocean-Atmosphere (coupled) 4DVar

Biggest problem is different operational time scales



Atmosphere:

Data-assimilation window 6 to 12 hours



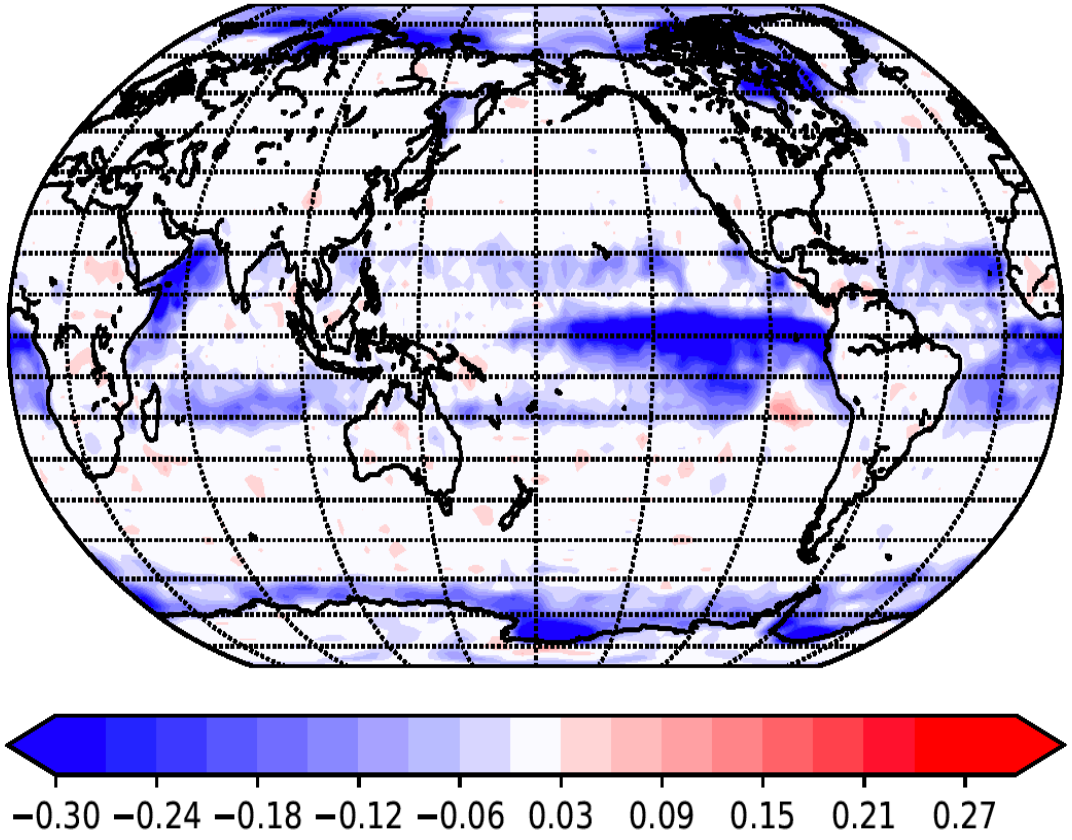
Ocean:

Data assimilation window 3 days

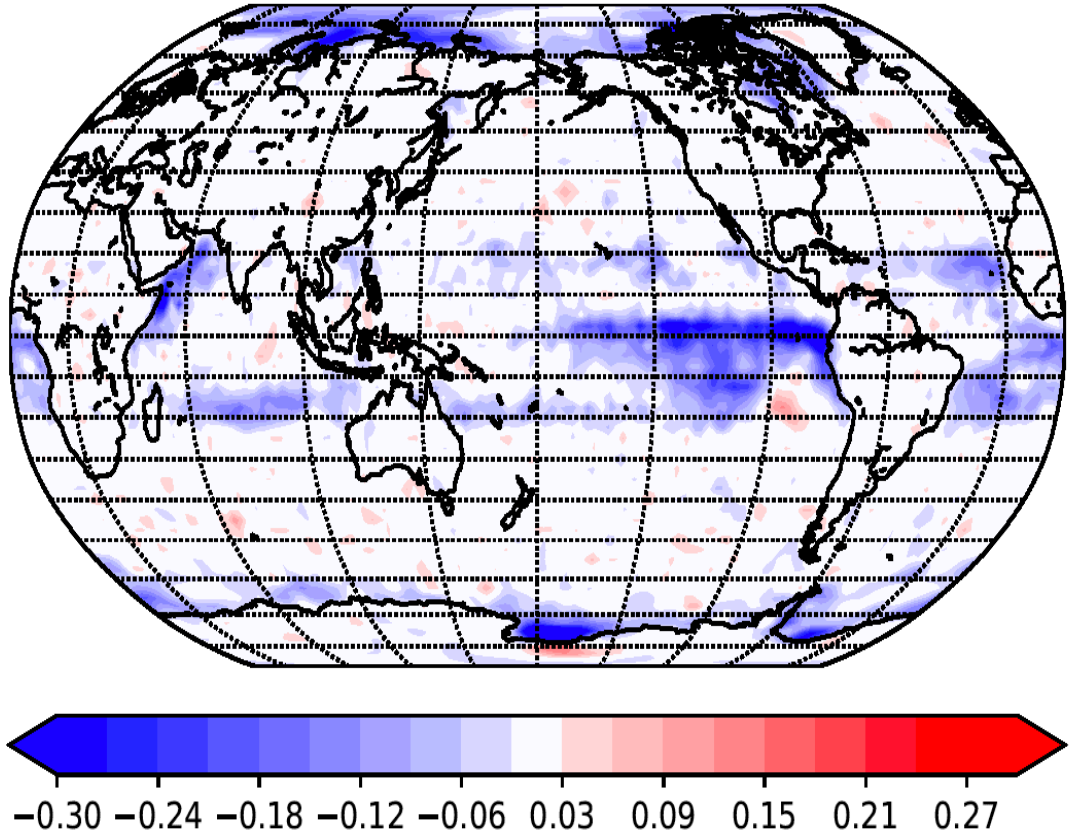
Ocean data come in late, e.g. ARGO buoys once every 10 days.

ECMWF weakly coupled Earth system DA

Normalised difference in rms error of T at 1000hPa T+12hrs

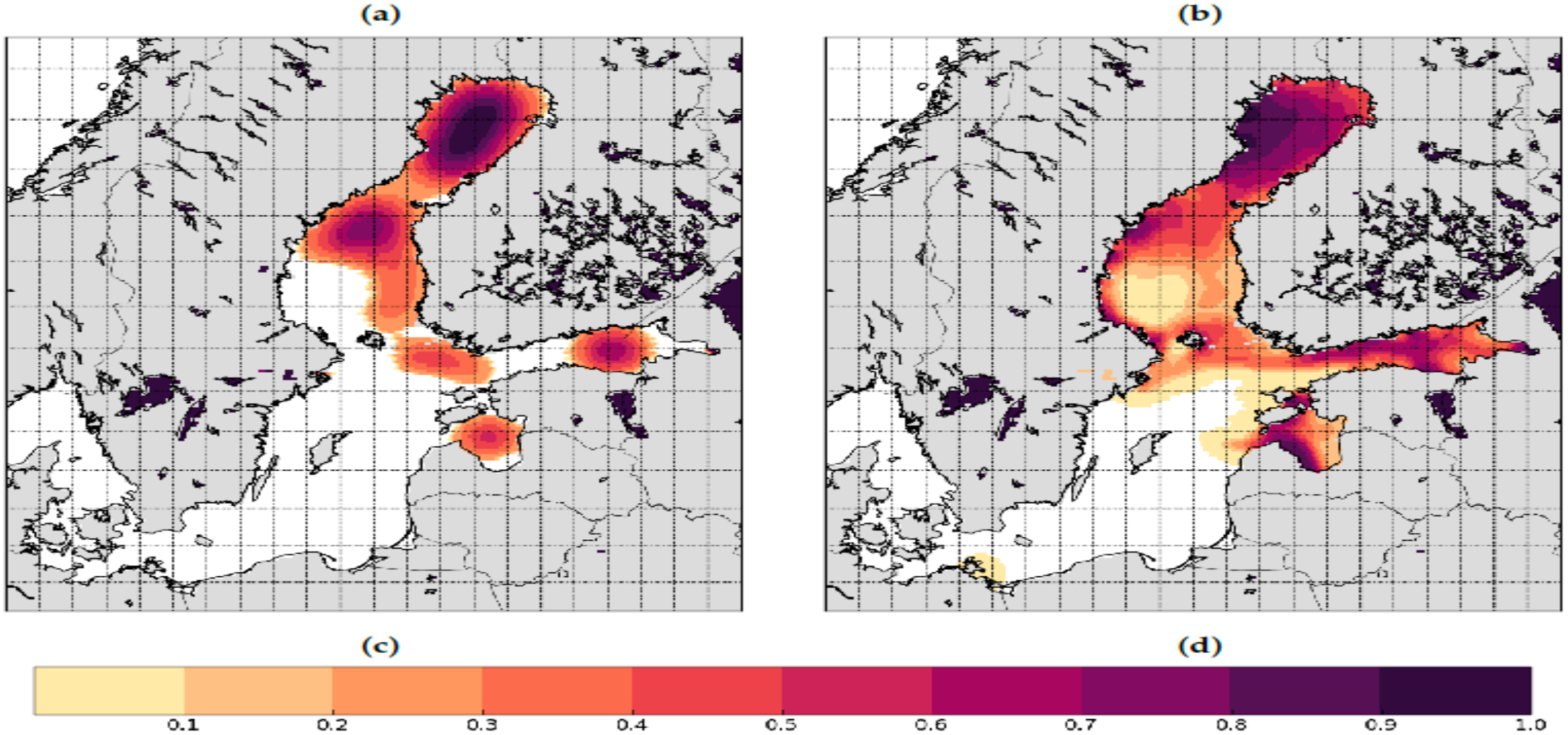


Normalised difference in rms error of T at 1000hPa T+48hrs



ECMWF weakly coupled Earth system DA

sea-ice concentration assimilation a) uncoupled, b) weakly coupled



Ocean-Atmosphere (coupled) data assimilation

1. Cross components effects are generally stronger in the direction from the slow to the fast scale, so that observations of the slow scale may benefit the fast, but,
2. Intra-component effects are much stronger in the fast scale. The fast scale must be controlled by frequently enough observations to prevent error growth and affect the slow scale.
3. The coupling changes the Lyapunov spectrum, and it seems important to also control neutral and weakly stable modes.

Coupled data assimilation in Ensemble Filters

Kuramoto-Shivashinsky coupled model:

Domain size of 32 for Atmos and
256 for Ocean on a 1024 nodes grid.

$\epsilon > 1$: ocean evolves on a slower scale than the atmosphere.

Uncoupled (left) and Coupled (right); $\epsilon =$

Coupled data assimilation

Uncoupled (left) and Coupled (right); ϵ

Uncoupled (left) and Coupled (right); ϵ =

Correlations, using 1000 ensemble members

Uncoupled

Coupled (right); ϵ =

Correlations, using 1000 ensemble members

Uncoupled

Coupled (right); ϵ =

The best localization scale for small ensemble size?

Vastly different scales…, and how to do cross-system covariances?

Adaptive localization

Local analysis updates a gridpoint using only “nearby” observations.

- Distance-based truncation of “spatially” remote obervations.
- Correlation-based truncation of “weakly correlated” observations.

We define a correlation distance as

Truncating when the correlation distance

Hence: spatial distance becomes irrelevant, only correlation value matters.

Adaptive localization

Root-mean-square errors deviation

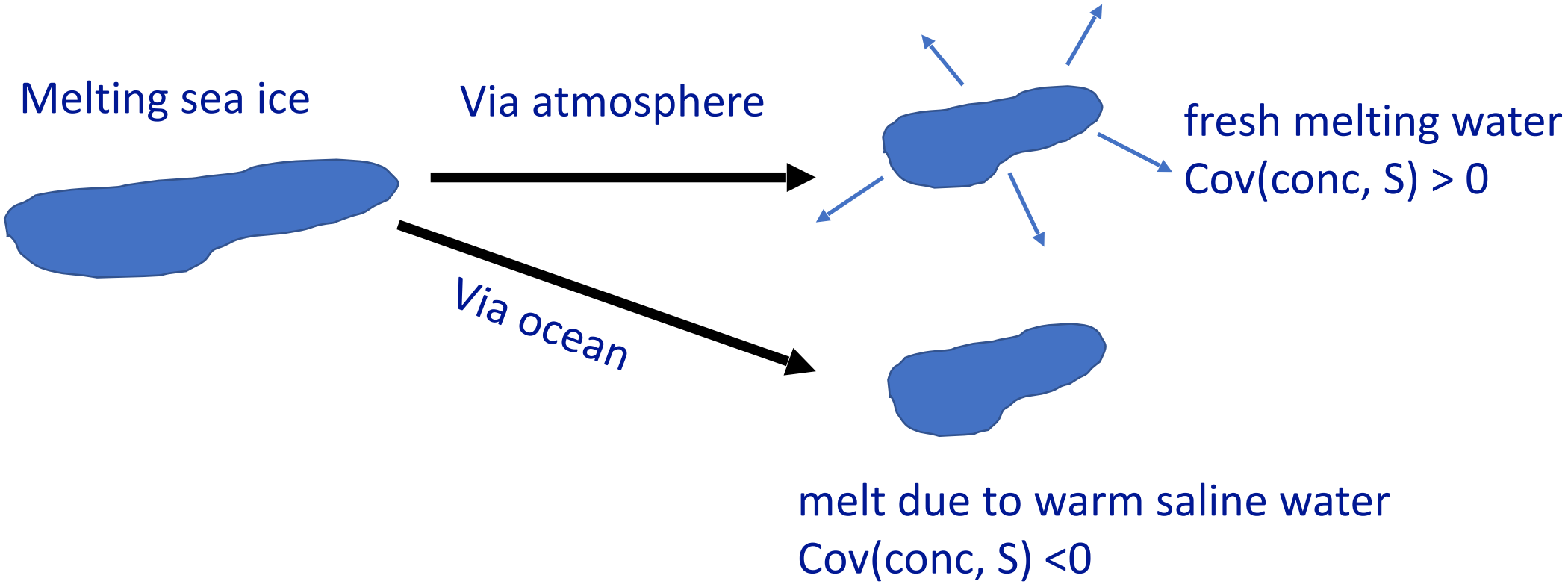
Ensemble standard

Adaptive localization seems the way to go !

Sea-ice –ocean and –atmosphere data assimilation

Sea-ice DA is still in its infancy, EnKF most advanced method, but problem is highly nonlinear:

1) sea-ice boundary (covariances can ‘flip’)



Sea-ice –ocean and –atmosphere data assimilation

2) sea-ice concentration in [0,1], so not Gaussian

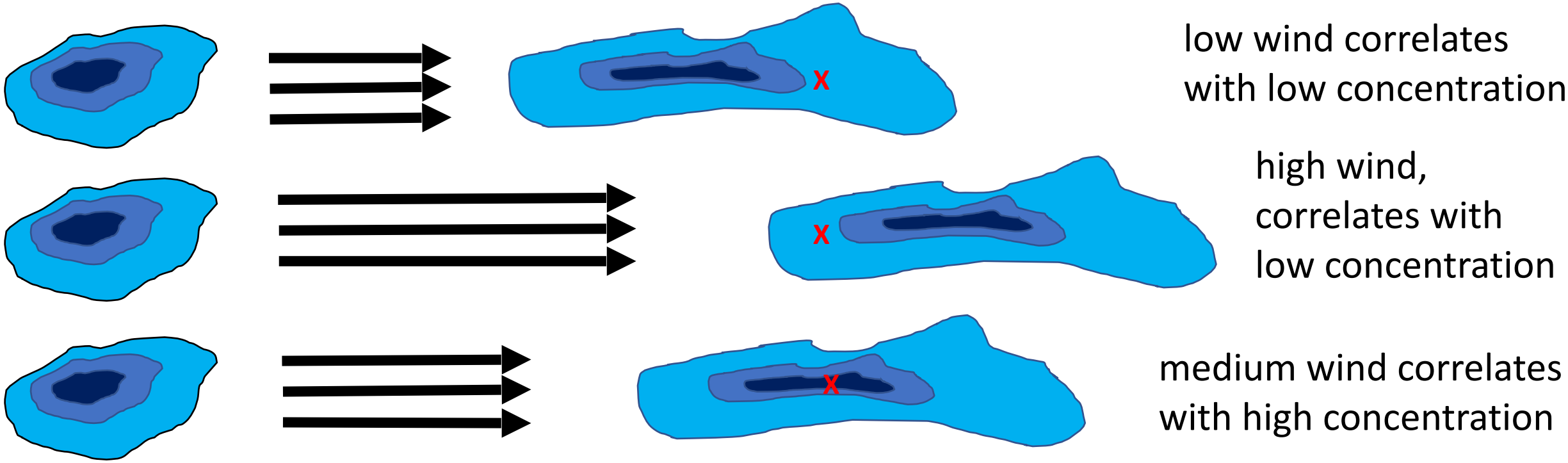
3) Sea-ice thickness > 0, so not Gaussian

4) Equation of state highly nonlinear (ridging)

5) observational difficulties (melt ponds, snow on ice, sea-ice thickness categories)

Atmospheric chemistry - meteorology

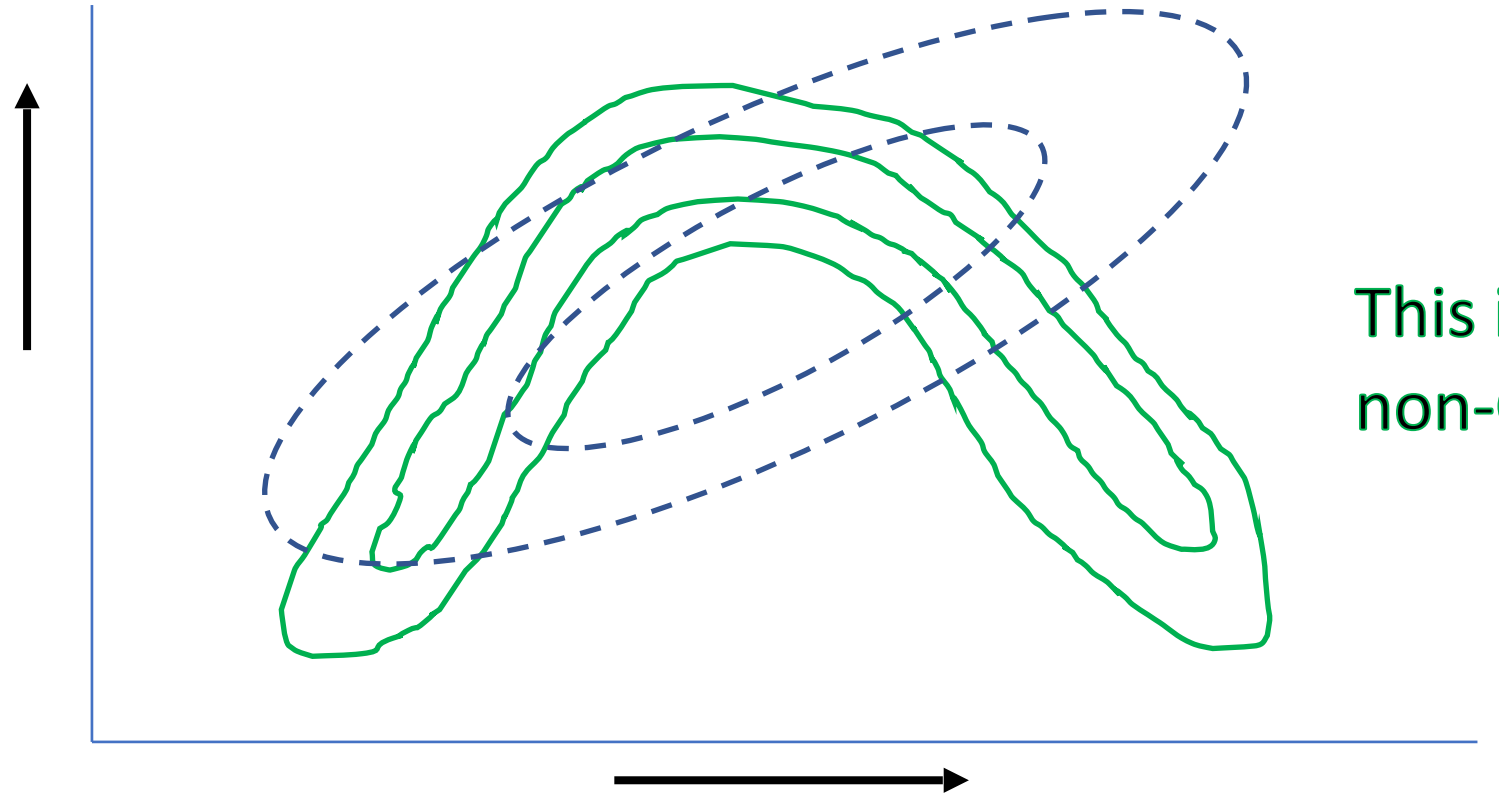
- Meteorology and atmospheric chemistry can be highly coupled.
- Even when influence of atmospheric chemistry on meteorology is neglected, the relations between e.g. concentration fields and winds allow for updates in meteorology.
- But coupling often highly nonlinear



Pdf of concentration at time t and wind strength

$$H_{0\to t}(x) = H_t(M_{0\to t}(x))$$

= concentration at time t



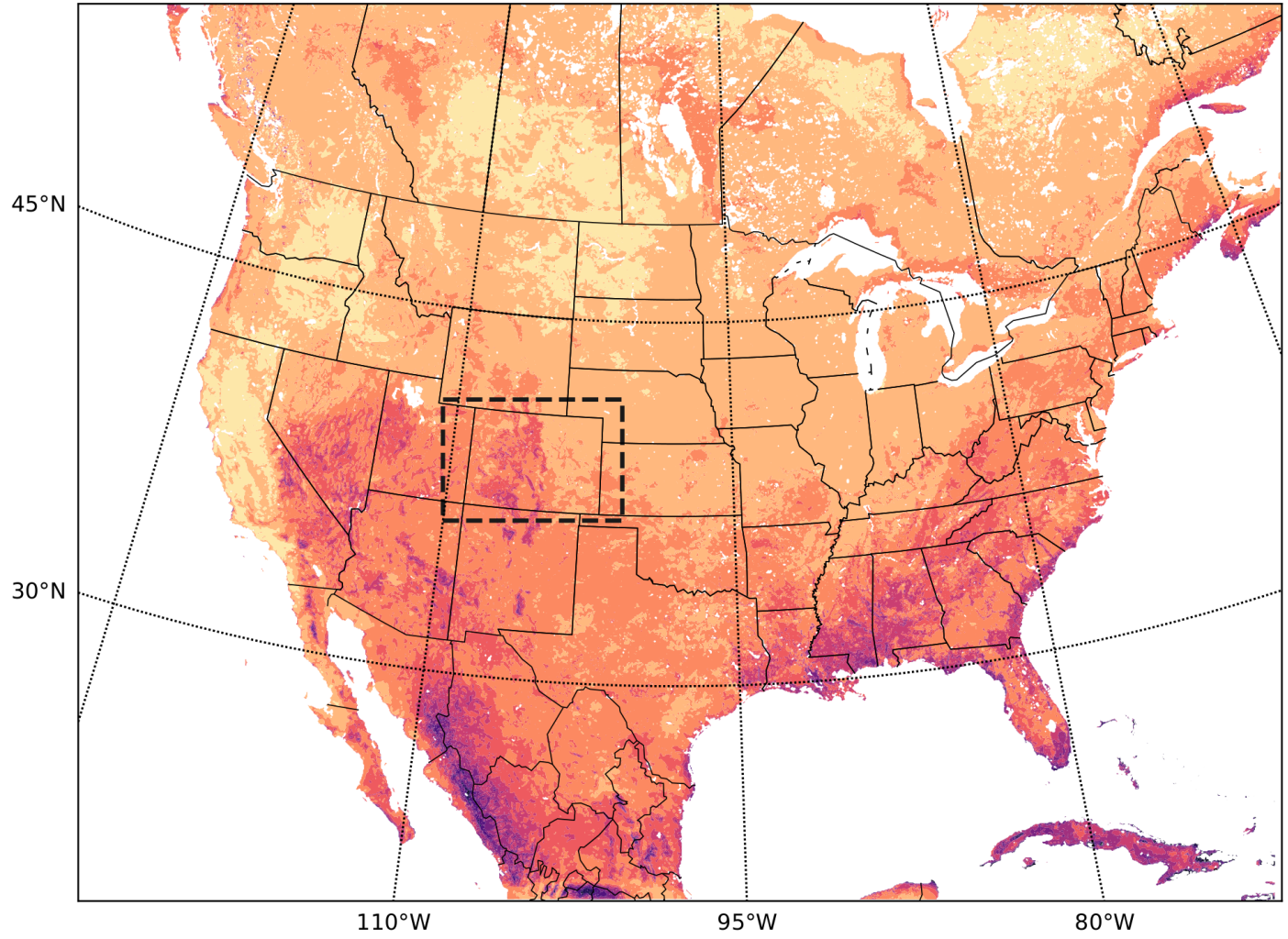
This is highly non-Gaussian !

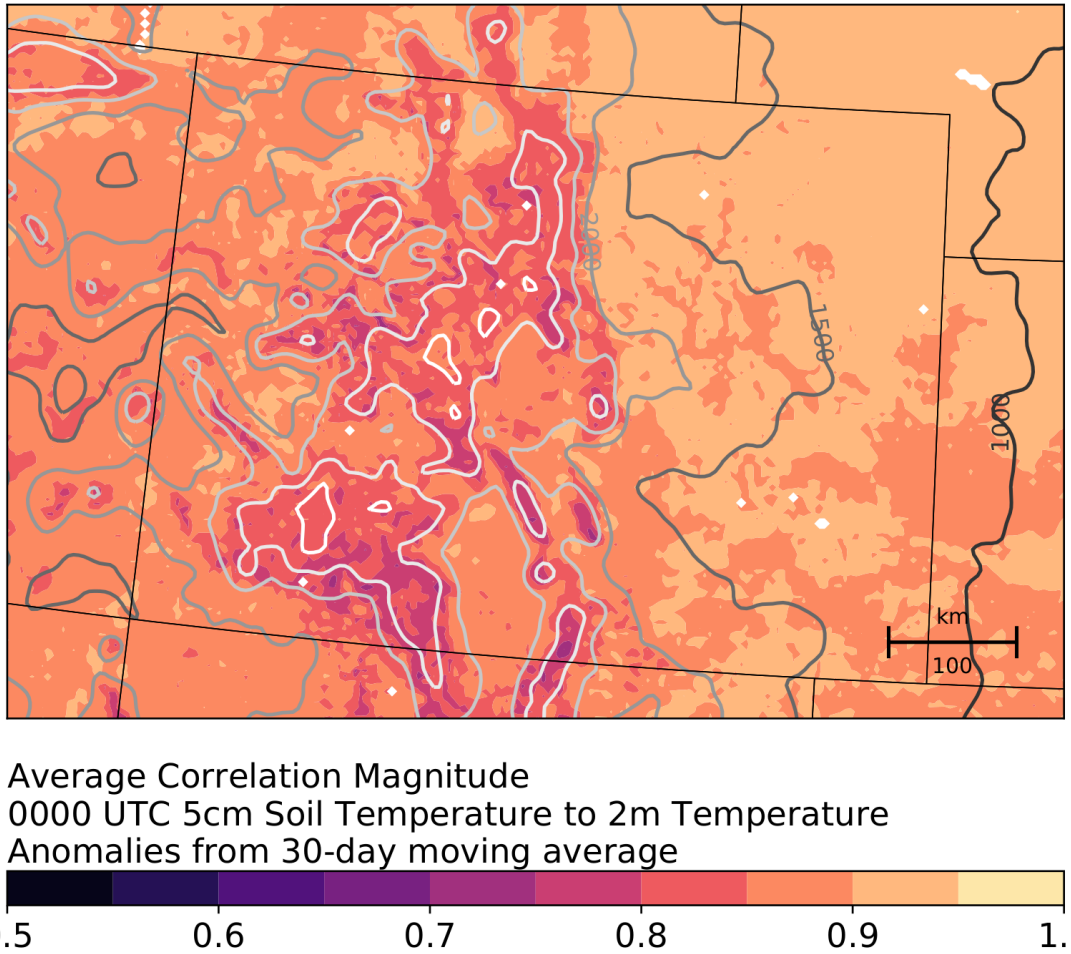
$M_{0\to T}(x)$ = Wind strength = transport error

(see e.g. Anderson 2020, MWR)

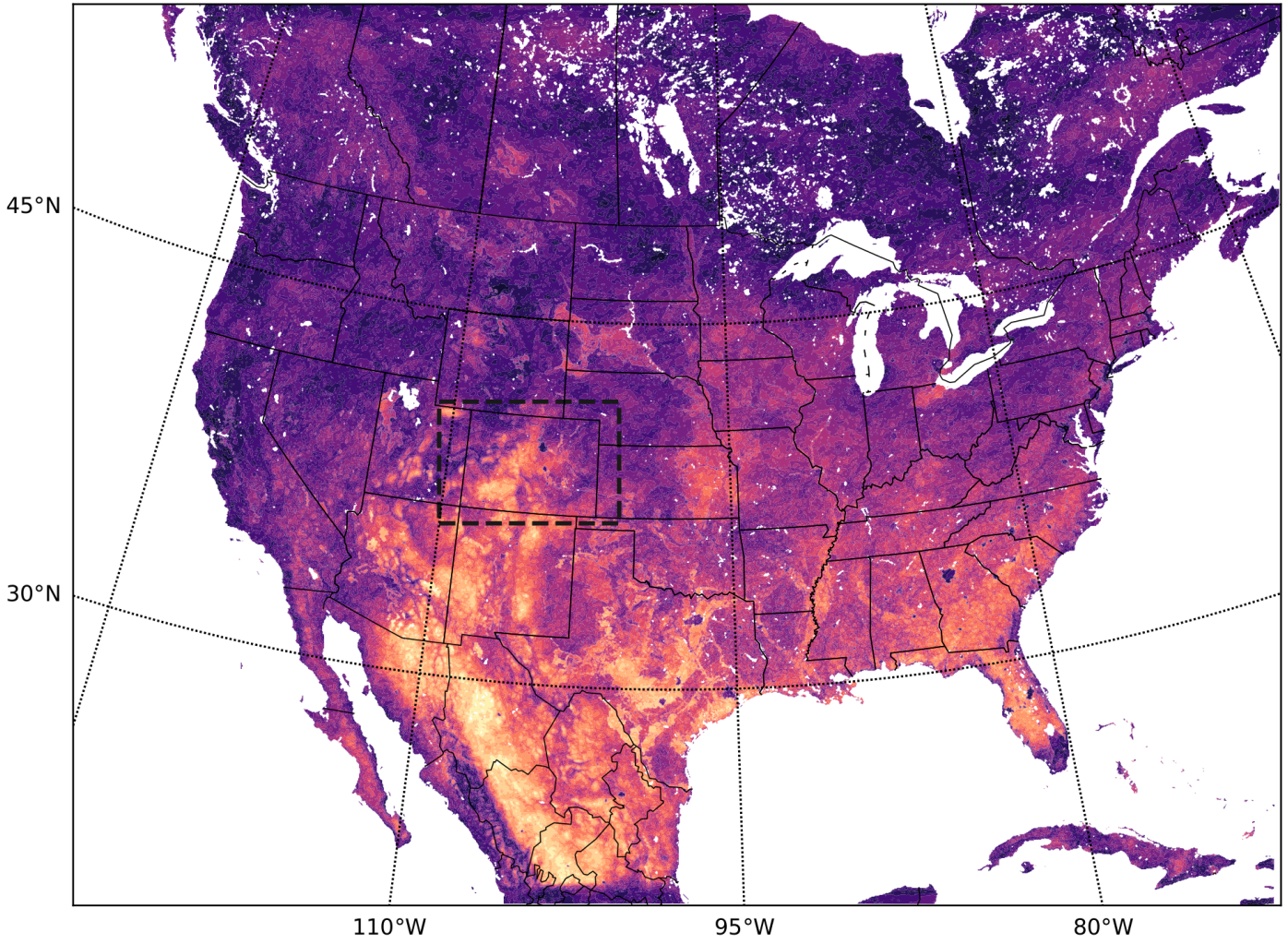
Land-atmosphere DA

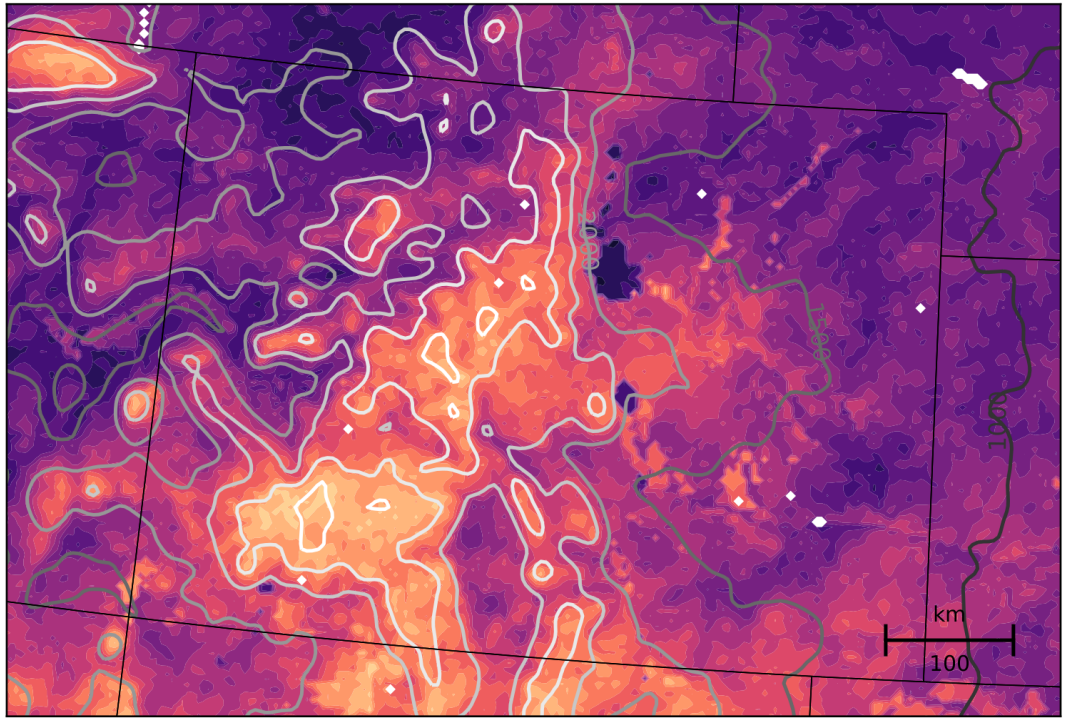
Highly irregular covariances





Land-atmosphere DA





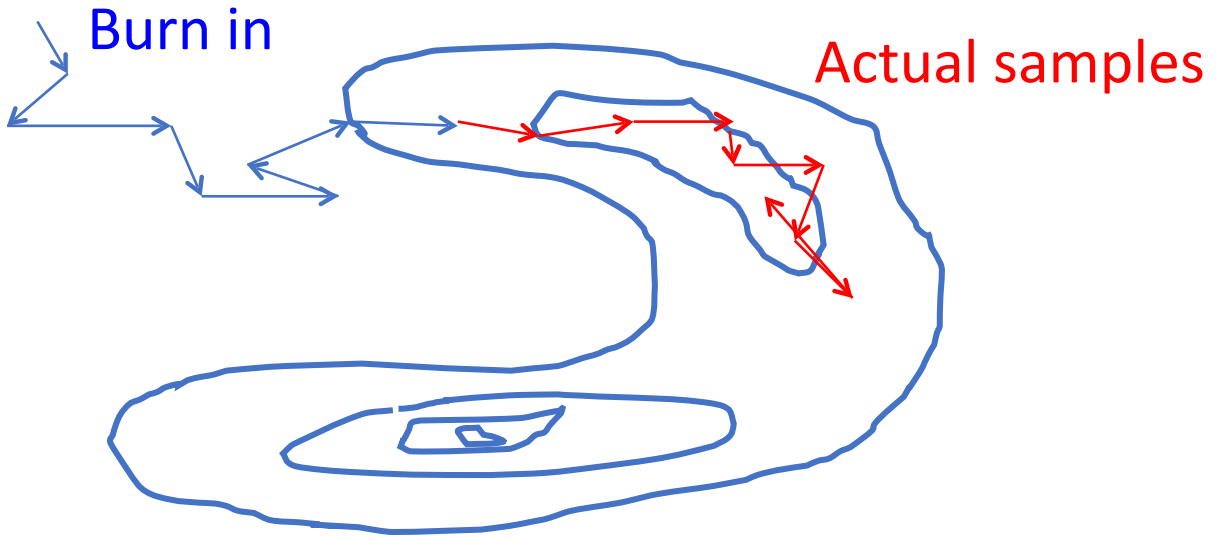
Average Correlation Magnitude
0000 UTC 5cm Soil Moisture to 2m Specific Humidity
Anomalies from 30-day moving average

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7

Nonlinear data-assimilation methods I

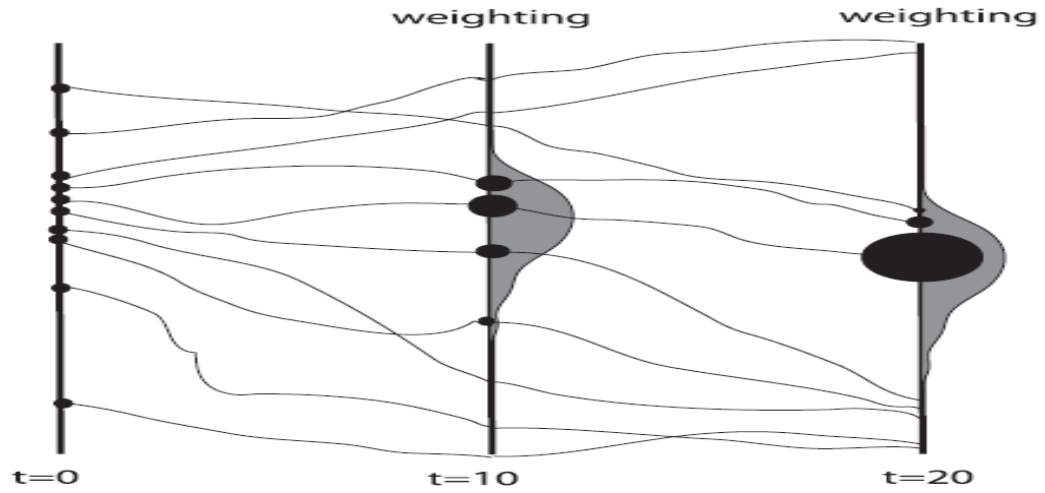
Markov-Chain Monte-Carlo methods
(e.g. Metropolis-Hastings, Langevin sampling, Hamiltonian Monte-Carlo)

- These schemes are sequential in generating ensemble members
- Only for small (<10) dimensions



Particle Filters/Smoothers

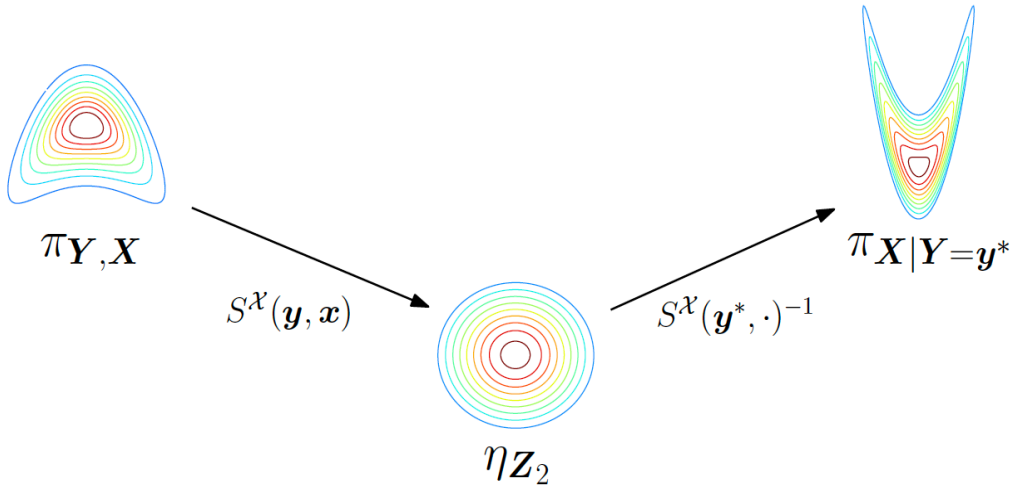
- Generate samples in parallel, use Importance sampling. High-dimensional variants biased.
- Highly efficient schemes use tempering (iterative likelihood refinement)



Nonlinear data-assimilation methods II

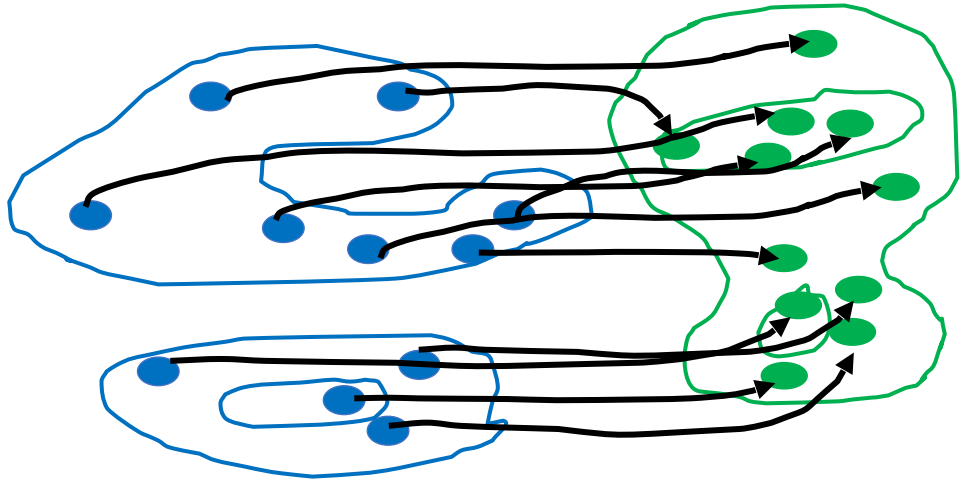
One-step Optimal transportation filters

- Find transport map between posterior and reference (Gaussian) pdf, and between prior and reference pdf
- Use e.g. triangular map



Particle Flow filters/smoothers

- Flow in pseudo time
- Stochastic versions can be made unbiased
- used in high-dimensions!

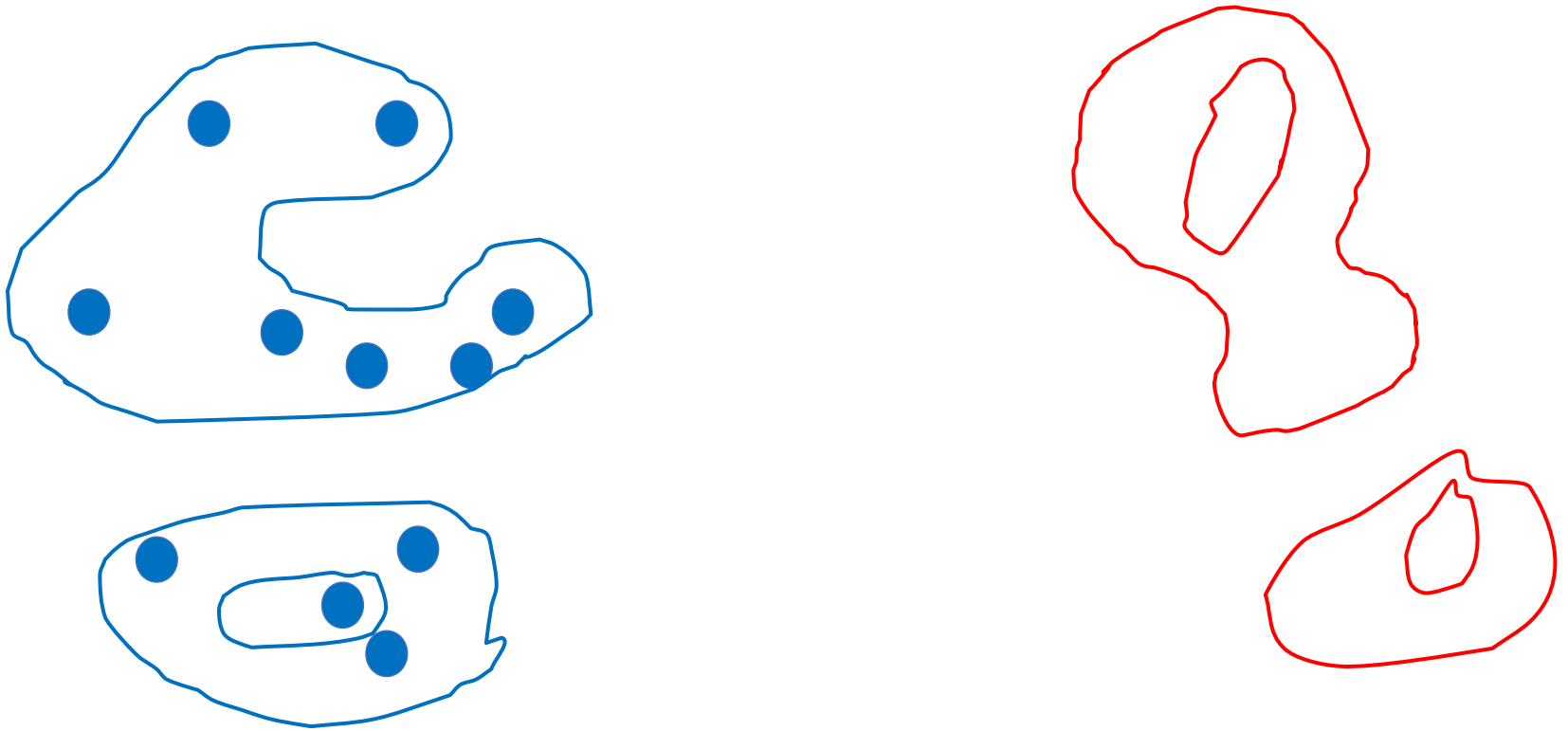


Before observations
'Prior'

After observations
'Posterior'

Particle Flows: propagation of pdf from prior to posterior

Particle flow in *pseudo time*

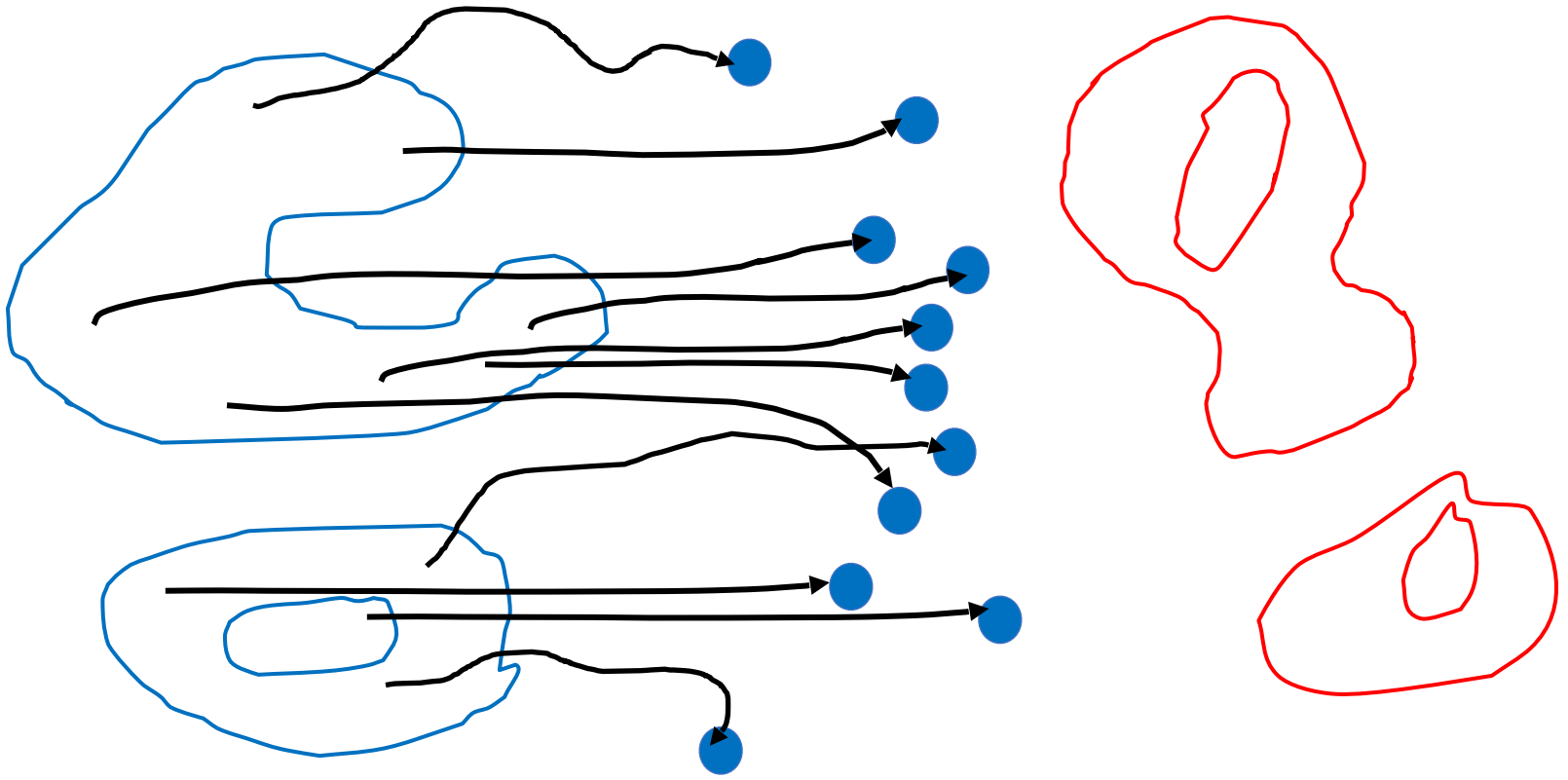


Prior pdf

Posterior pdf

The prior and posterior can be for a model state (filter) or a model trajectory (smoother) or a set of parameters, or a combination of these.

Particle flow in *pseudo time*

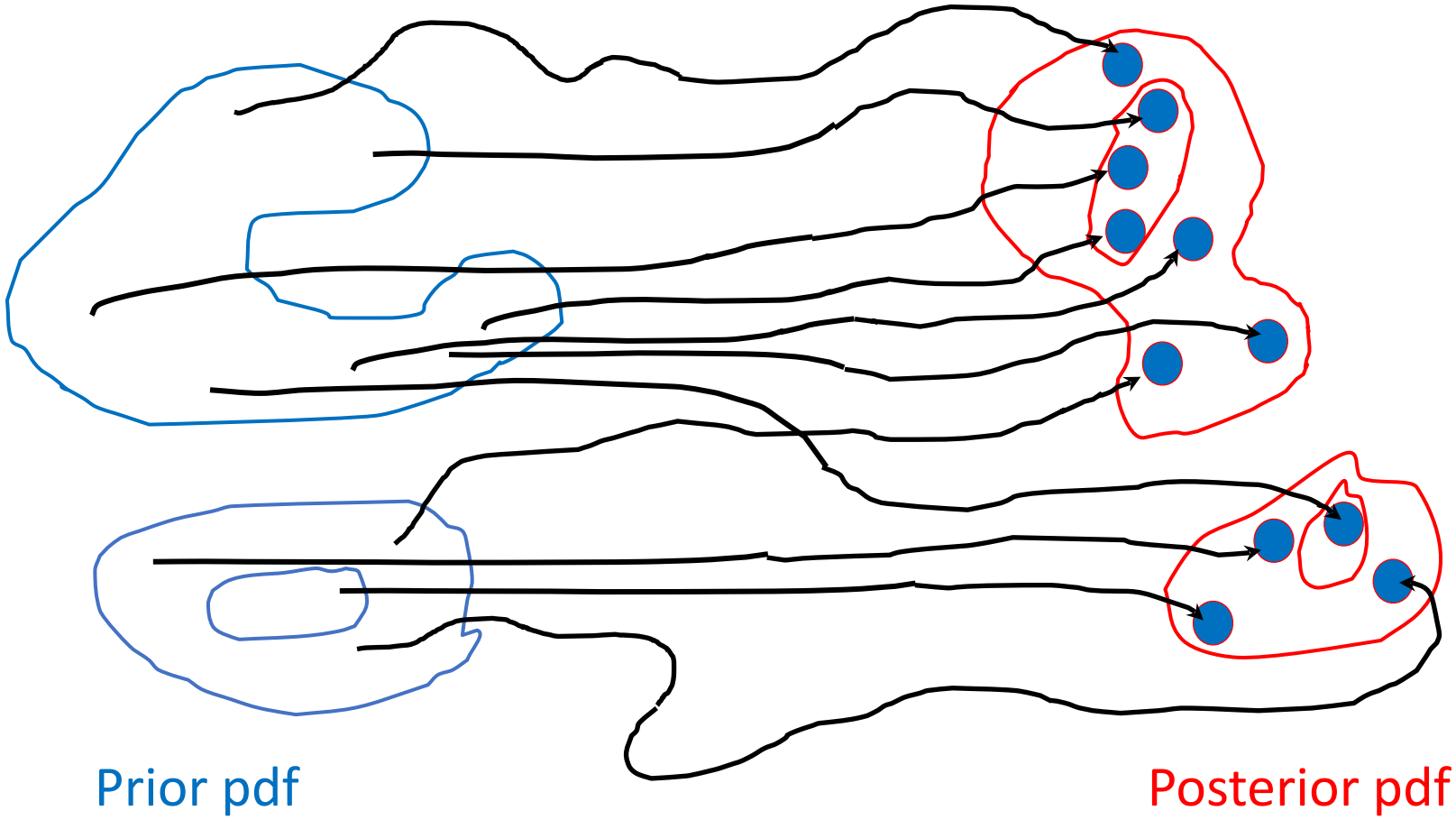


Prior pdf

Posterior pdf

Particle Flow converged on posterior pdf

Not degenerate by construction



Hu, C-C, and P.J. van Leeuwen (2021) A particle flow filter for fully nonlinear high-dimensional data assimilation., Q.J. Royal Meteorol. Soc., doi: 10.1002/qj.4028

Particle Flow Filter on a high dimensional atmospheric model

5-layer primitive equation model of the atmosphere with variables U, V, T, SLP

28.200 gridpoints

Assimilation of sea-level pressure (SLP)

1)

2)

Ensemble size 25 particles

Compare EnKF and PFF

Linear observation operator with Gaussian observation errors

Domain averaged values

Nonlinear observation operator with Gaussian observation errors

Domain averaged values

Conclusions

- Data assimilation theory is simple, but developing practical schemes is complicated.
- DA for Earth system models in its infancy
- Main effort in coupled ocean-atmosphere, which seems most complicated due to size of problem and different time scales
- DA problem is be coming highly nonlinear
- Atmospheric chemistry will be a killer…
- All agencies are in high need of DA expertise
- ML is unlikely to take over soon.

Shameless plug…

- Free open access book
- doi: 10.1007/978-3-030-96709-3