Jet Propulsion Laboratory
California Institute of Technology

State Estimation

Part 3: Practical Matters

Ichiro Fukumori

Jet Propulsion Laboratory, Caltech

ECCO Summer School 2019

Main Points from Lectures 1 & 2

1) State estimation (data assimilation) is fundamentally an inverse problem,

observations

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \mathbf{H} & & & & & \cdots \\
\cdots & & \mathbf{H} & & & & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\
\cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots
\end{pmatrix}
\approx
\begin{pmatrix}
\vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots
\end{pmatrix}
$$

model equations

2) Inverse problems using data are invariably ill-posed mathematically for which inverse methods provide particular solutions,

3) Estimates by minimum variance and least-squares estimates are identical when assumptions are the same,

4) Kalman filter and RTS Smoother are minimum variance estimators that solve the state estimation problem sequentially in time and by separate constraint,

5) Adjoint method is a least-squares estimator that solves the state estimation problem iteratively by descent optimization, using the model’s adjoint to compute the gradient of the least-squares’ misfit.

Outline

1. Basic Machinery (Monday)

   The mathematical problem (inverse problem), Linear inverse methods, Singular value decomposition (SVD), Rank deficiency, Gauss-Markov theorem, Minimum variance estimate, Least-squares,

2. Methods of state estimation (Tuesday)

   Kalman filter, Rauch-Tung-Striebel smoother, Adjoint method,

3. Practical Matters (this lecture)

   Error estimation, representation error, covariance, approximate Kalman filters, other data assimilation methods (Optimal Interpolation, 3DVAR).

What is "data error"?

$$\mathbf{R} \equiv \left\langle \delta\left(\hat{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}\right)\delta\left(\hat{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}\right)^T \right\rangle$$

Objective function in adjoint method

$$\begin{aligned} J \equiv \sum_{t=1}^{M} &\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \\ &+\left[\hat{\mathbf{x}}(0,+)-\hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+)-\hat{\mathbf{x}}(0)\right] \\ &+\sum_{t=0}^{M-1}\left[\hat{\mathbf{u}}(t,+)-\hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+)-\hat{\mathbf{u}}(t)\right] \\ &-2\sum_{t=1}^{M}\boldsymbol{\mu}(t)^T\left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+)-\mathbf{G}\hat{\mathbf{u}}(t-1,+)\right] \end{aligned}$$

Kalman filter

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

What is “data error”?

“Data error” in state estimation is not simply “noise in data” but includes ***model representation error***.

$$\mathbf{R} \equiv \left\langle \delta\left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)\delta\left(\hat{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}\right)^T \right\rangle$$

true value of the model state

The true value of the model may not necessarily correspond to a noise-free observation, because of missing physics from the model; e.g.,

1. Subgridscale variability,
2. Baroclinic variability for a barotropic model,
3. Tides in a non-tidally forced model,
4. Atmospheric pressure-driven variability in a model without pressure forcing.

How can data error be set?

An estimate of data error can be obtained by comparing observations with a model simulation.

signal

data error

$$\mathbf{y} = \mathbf{s} + \mathbf{n}$$

$$\mathbf{Hx}_{sim} \equiv \mathbf{m} = \mathbf{s} + \mathbf{p}_{sim}$$

state error of simulation

Assuming signal and errors are mutually uncorrelated, i.e.,

$$\langle \mathbf{sn}^T \rangle = \mathbf{0} \quad \langle \mathbf{sp}_{sim}{}^T \rangle = \mathbf{0} \quad \langle \mathbf{np}_{sim}{}^T \rangle = \mathbf{0}$$

covariance between data and model simulation is,

$$\langle \mathbf{yy}^T \rangle = \langle \mathbf{ss}^T \rangle + \langle \mathbf{nn}^T \rangle$$

$$\langle \mathbf{mm}^T \rangle = \langle \mathbf{ss}^T \rangle + \langle \mathbf{p}_{sim}\mathbf{p}_{sim}{}^T \rangle$$

$$\langle \mathbf{ym}^T \rangle = \langle \mathbf{ss}^T \rangle$$

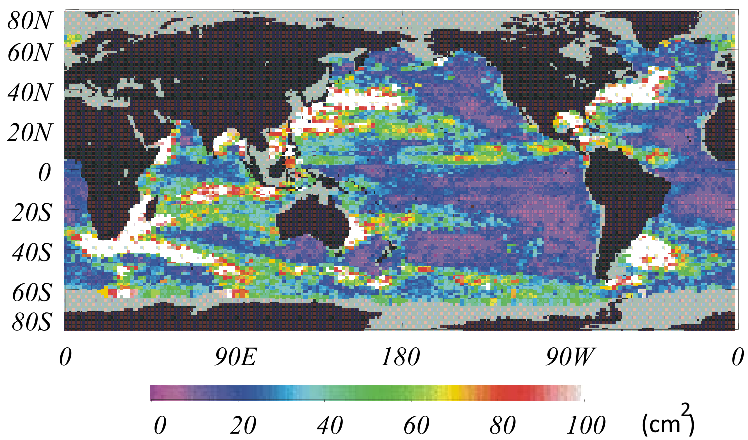
Which can be solved for $\langle \mathbf{nn}^T \rangle \equiv \mathbf{R}$ and $\langle \mathbf{p}_{sim}\mathbf{p}_{sim}{}^T \rangle \equiv \mathbf{HP}_{sim}\mathbf{H}^T$

[*Fu et al., 1993*]

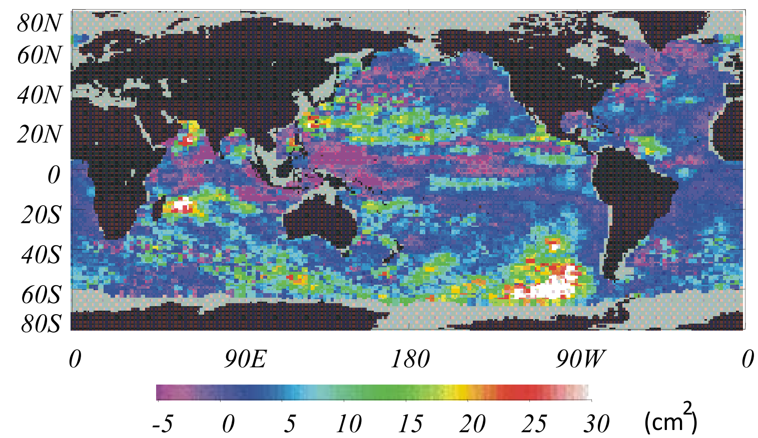
Example of an error estimate

Data and model simulation error estimates (variance) of sea level, comparing altimetry data (TOPEX/Poseidon) and a global MITgcm with 1-deg resolution.

$$\mathbf{R} \equiv \langle \mathbf{n}\mathbf{n}^T \rangle = \langle \mathbf{y}\mathbf{y}^T \rangle - \langle \mathbf{y}\mathbf{m}^T \rangle$$



$$\mathbf{H}\mathbf{P}_{sim}\mathbf{H}^T = \langle \mathbf{p}\mathbf{p}^T \rangle = \langle \mathbf{m}\mathbf{m}^T \rangle - \langle \mathbf{y}\mathbf{m}^T \rangle$$

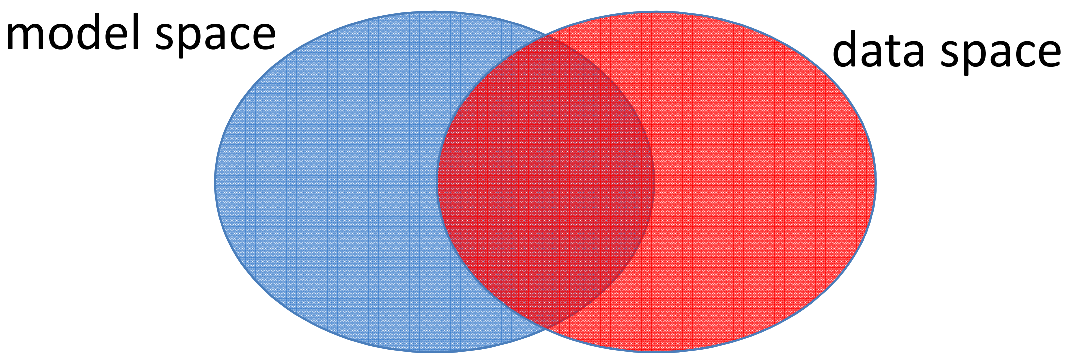


[*Fukumori et al., 1999*]

Significance of State Estimation

Given what "data error" is,

State estimation constrains what is consistent with both model and observations, not entirely what observations measure or the combination of the two.



observations

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \mathbf{H} & & & & & \cdots \\ \cdots & & \mathbf{H} & & & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \cdots \\ \cdots & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

model equations

State estimation corrects model's common part with data.

What is model error?

$$\mathbf{Q} \equiv \left\langle \delta\hat{\mathbf{u}}\delta\hat{\mathbf{u}}^T \right\rangle \qquad \text{where} \quad \mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$

Objective function in adjoint method

$$\begin{aligned} \mathrm{J} \equiv \sum_{t=1}^{M} & \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right]^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,+)\right] \\ & + \left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right]^T \mathbf{P}(0)^{-1}\left[\hat{\mathbf{x}}(0,+) - \hat{\mathbf{x}}(0)\right] \\ & + \sum_{t=0}^{M-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right]^T \mathbf{Q}(t)^{-1}\left[\hat{\mathbf{u}}(t,+) - \hat{\mathbf{u}}(t)\right] \\ & - 2\sum_{t=1}^{M}\boldsymbol{\mu}(t)^T\left[\hat{\mathbf{x}}(t,+) - \mathbf{A}\hat{\mathbf{x}}(t-1,+) - \mathbf{G}\hat{\mathbf{u}}(t-1,+)\right] \end{aligned}$$

Kalman filter

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

RTS smoother

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

What we treat as model error dictates what is being estimated.

What is being estimated?

Control (includes model forcing & errors in the model equation)

State estimate $\hat{\mathbf{x}}(t) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$

In general $\bar{\mathbf{x}}(t) = \mathbf{A}\bar{\mathbf{x}}(t-1) + \mathbf{B}\bar{\mathbf{w}}(t-1) + \boldsymbol{\Gamma}\bar{\mathbf{q}}(t-1)$

where $\mathbf{G}\mathbf{u} \equiv \begin{pmatrix} \mathbf{B} & \\ & \boldsymbol{\Gamma} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{q} \end{pmatrix}$

$$\delta\mathbf{u} \equiv \begin{pmatrix} \delta\mathbf{w} \\ \delta\mathbf{q} \end{pmatrix} \qquad \mathbf{Q} = \left\langle \delta\mathbf{u}\delta\mathbf{u}^T \right\rangle = \begin{pmatrix} \left\langle \delta\mathbf{w}\delta\mathbf{w}^T \right\rangle & \left\langle \delta\mathbf{w}\delta\mathbf{q}^T \right\rangle \\ \left\langle \delta\mathbf{q}\delta\mathbf{w}^T \right\rangle & \left\langle \delta\mathbf{q}\delta\mathbf{q}^T \right\rangle \end{pmatrix}$$

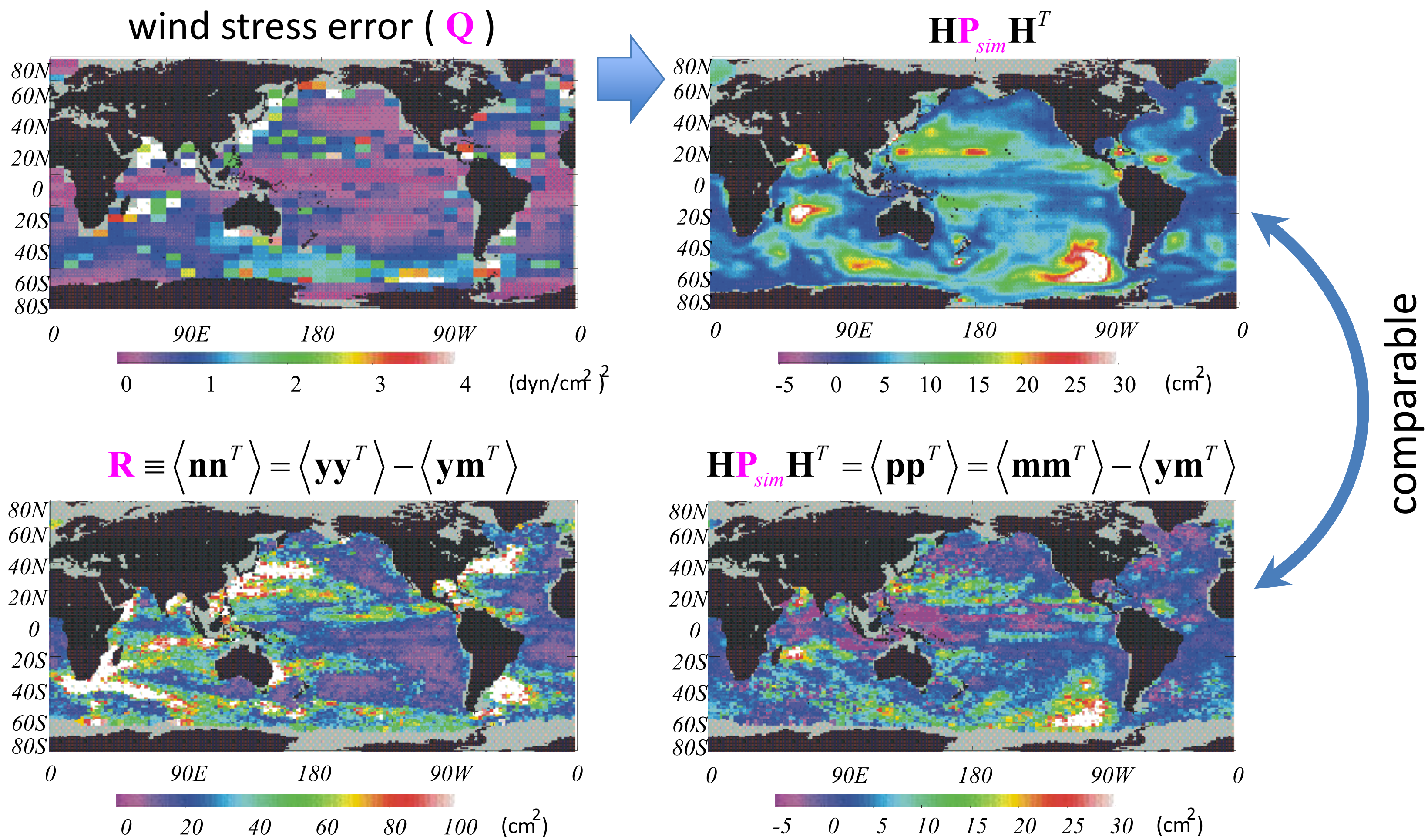
In most ECCO estimates, we set

$$\mathbf{Q} = \left\langle \delta\mathbf{u}\delta\mathbf{u}^T \right\rangle = \begin{pmatrix} \left\langle \delta\mathbf{w}\delta\mathbf{w}^T \right\rangle & 0 \\ 0 & 0 \end{pmatrix}$$

the veracity of which is assessed a posteriori.

Example of model error estimate

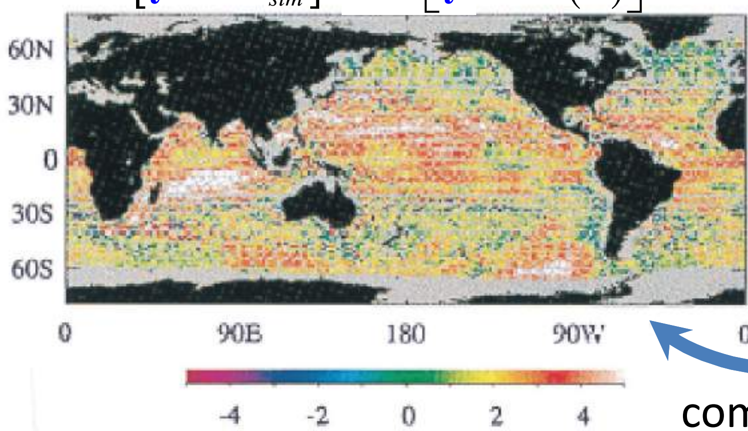
Wind error chosen such that model simulation error estimate is comparable to that from data-simulation comparison.



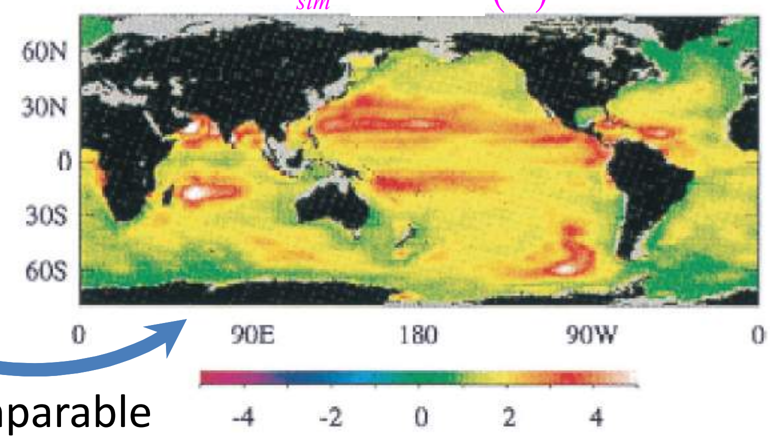
Example of model error estimate

Consistency between formal error estimates
and model-data differences

$\mathrm{var}\left[\hat{\mathbf{y}} - \mathbf{H}\hat{\mathbf{x}}_{sim}\right] - \mathrm{var}\left[\hat{\mathbf{y}} - \mathbf{H}\hat{\mathbf{x}}(-)\right]$

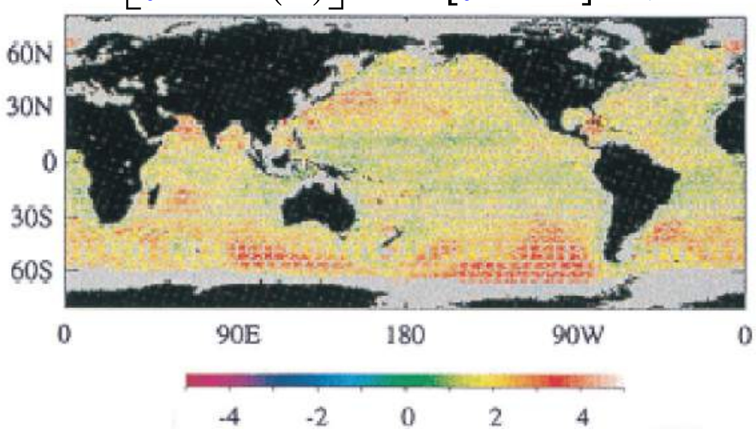


$\mathbf{H}\mathbf{P}_{sim}\mathbf{H}^T - \mathbf{H}\mathbf{P}(-)\mathbf{H}^T$

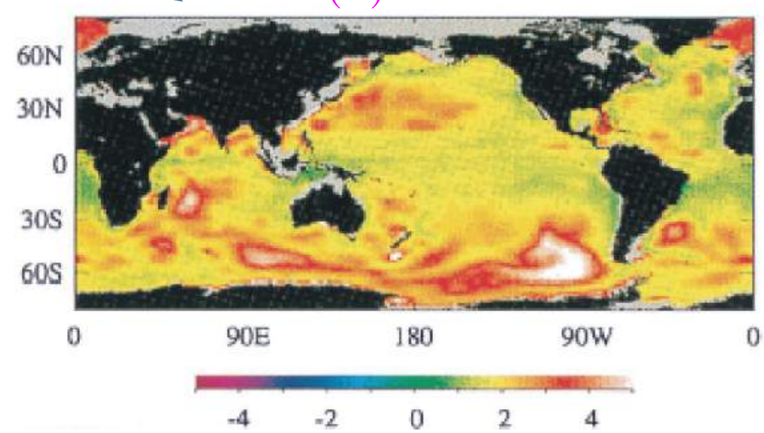


comparable

$\mathrm{var}\left[\hat{\mathbf{y}} - \mathbf{H}\hat{\mathbf{x}}(-)\right] - \mathrm{var}\left[\hat{\mathbf{y}} - \mathbf{H}\hat{\mathbf{x}}\right]$



$\mathbf{H}\mathbf{P}(-)\mathbf{H}^T - \mathbf{H}\mathbf{P}\mathbf{H}^T$

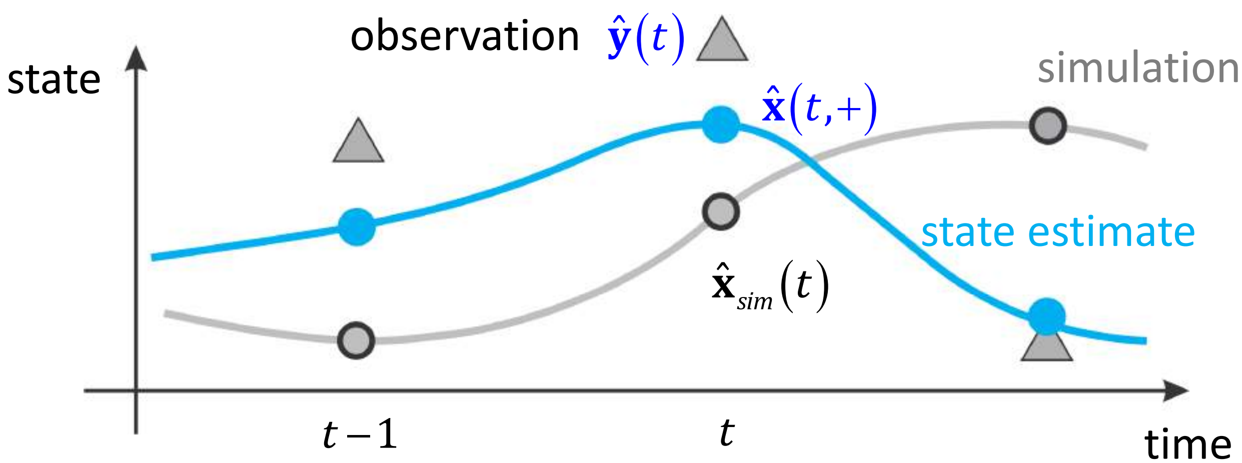


Significance of State Estimation

A state estimate provides

a) a ***more accurate*** description of *what is being estimated* than a simulation does,

b) a ***more complete*** description of *what is being estimated* than observations do.

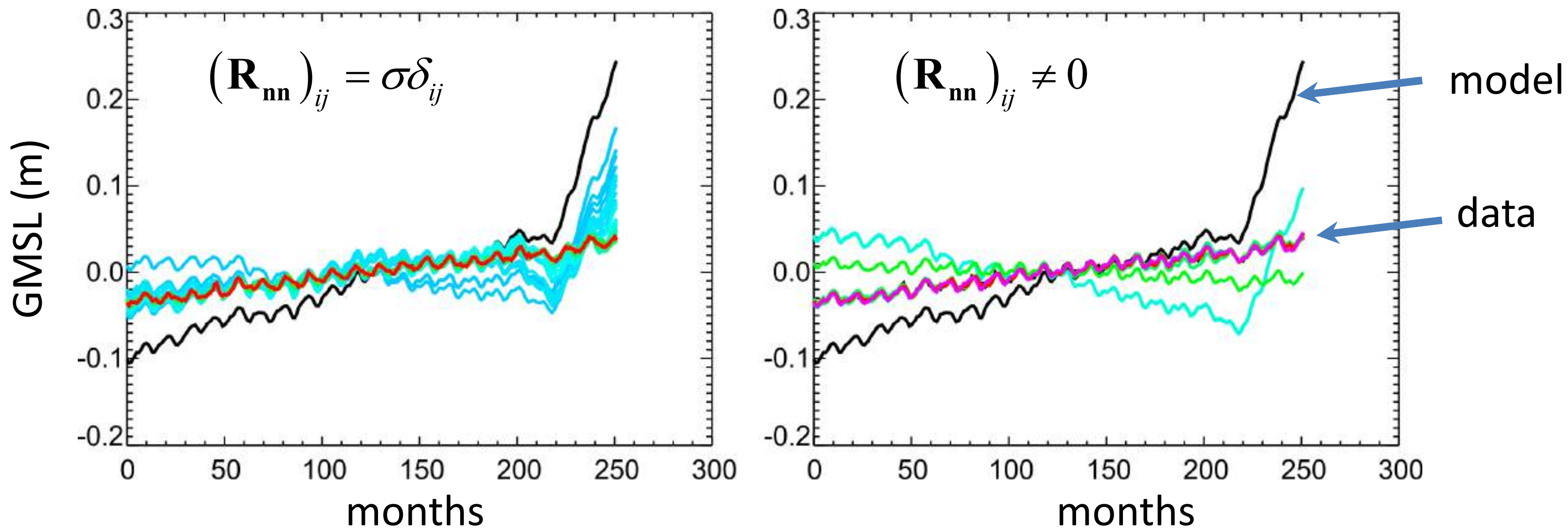


Significance of Correlation

$$J = (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T \mathbf{R}_{nn}^{-1} (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}) + \mathbf{x}^T \mathbf{R}_{xx}^{-1} \mathbf{x}$$

The weights (prior error covariance) define the estimation problem and using different weights amount to solving different problems. In particular, ignoring correlation (cross-covariance) can hamper optimization;

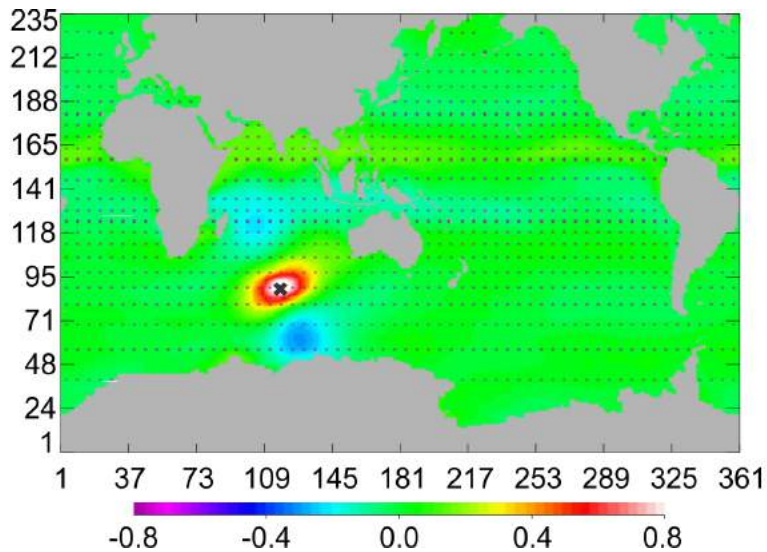
e.g., fitting global mean sea level (GMSL) rise. $\frac{d\eta}{dt} = EmP$



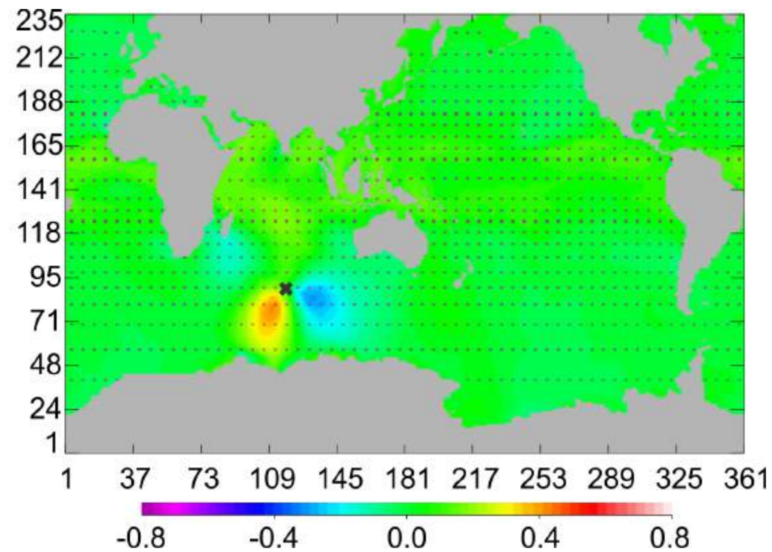
Significance of Correlation

Atmospheric controls are often not only correlated spatially but also between different variables.

$\langle \tau_x \tau_x \rangle$



$\langle \tau_x \tau_y \rangle$



Summary of data and control errors

1. Data error in state estimation is not only inaccuracies in the measurements but also includes models' representation error,

2. Data error and simulation error (consequence of process noise) can be estimated from comparing data and model simulation,

3. Model's control error (process noise) must be chosen judiciously and examined for consistency. It also defines what is being estimated by the model-data synthesis,

4. Covariance (correlation) among data error and among controls are as fundamental to estimation as the error variance themselves.

Evaluating Covariances

Evaluating covariance (correlation) can be difficult owing to its large dimension, which motivates approximations by taking advantage of certain inherent low degrees of freedom (d.o.f.), thus reducing computational requirements; e.g.,

- Large-scale variations often dominate covariance (e.g., atmospheric synoptic systems),
- Small-scale variations often have limited spatial extent.

reduced state

$$\mathbf{u} \approx \mathbf{B}'\mathbf{u}_r \quad \text{where} \quad \dim(\mathbf{u}_r) \ll \dim(\mathbf{u})$$

$$\langle \mathbf{u}\mathbf{u}^T \rangle \approx \langle \mathbf{B}'\mathbf{u}_r (\mathbf{B}'\mathbf{u}_r)^T \rangle \approx \langle \mathbf{B}'\mathbf{u}_r\mathbf{u}_r^T\mathbf{B}'^T \rangle \approx \mathbf{B}'\langle \mathbf{u}_r\mathbf{u}_r^T \rangle \mathbf{B}'^T$$

Estimate $\langle \mathbf{u}\mathbf{u}^T \rangle$ by evaluating the smaller $\langle \mathbf{u}_r\mathbf{u}_r^T \rangle$.

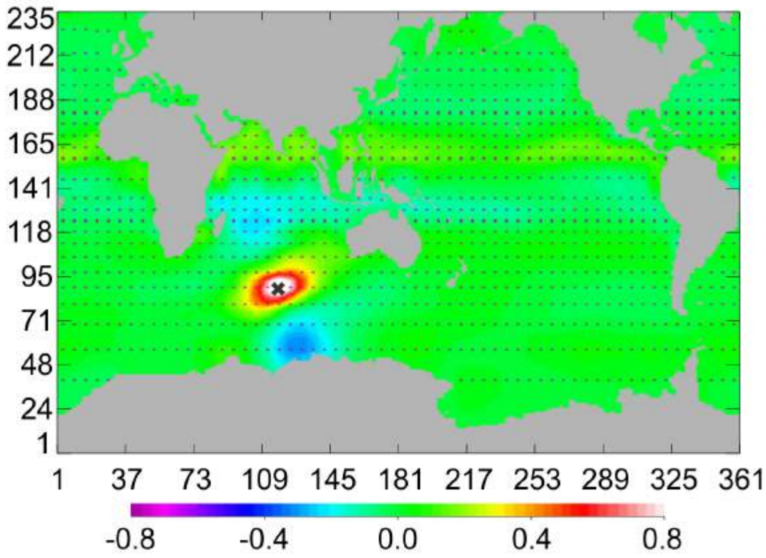
Quantifying correlation with reduced d.o.f.

Large-scale approximation of atmospheric control by objective mapping from a coarse grid.

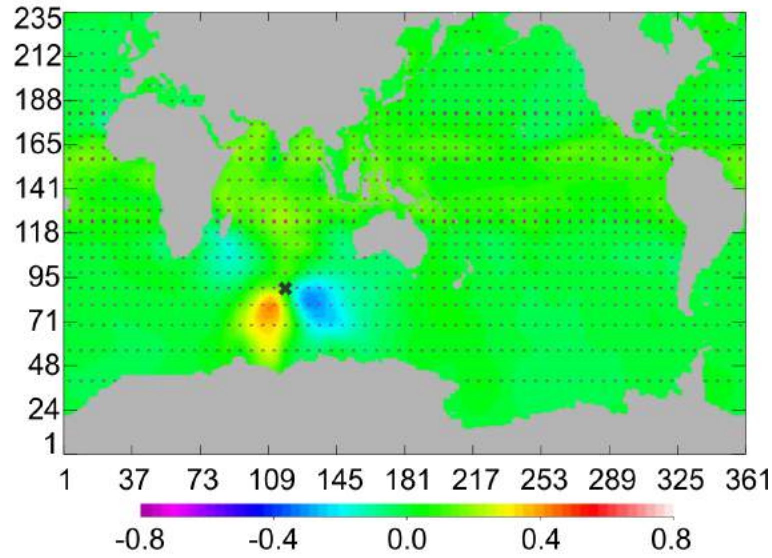
$$\mathbf{u} \approx \mathbf{B}'\mathbf{u}_r$$

100,000-point fine grid ($\mathbf{u}$); 1000-point coarse grid ($\mathbf{u}_r$)

$\langle \tau_x \tau_x \rangle$



$\langle \tau_x \tau_y \rangle$



$$\langle \mathbf{u}\mathbf{u}^T \rangle \approx \langle \mathbf{B}'\mathbf{u}_r (\mathbf{B}'\mathbf{u}_r)^T \rangle \approx \langle \mathbf{B}'\mathbf{u}_r \mathbf{u}_r^T \mathbf{B}'^T \rangle \approx \mathbf{B}' \langle \mathbf{u}_r \mathbf{u}_r^T \rangle \mathbf{B}'^T$$

200,000×200,000 2,000×2,000

Computational requirements of Kalman Filter

Forecast

$$\hat{\mathbf{x}}(t,-) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{G}\hat{\mathbf{u}}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$$

Correction

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t,-)$$

This error covariance's time-integration step is the most computationally demanding step of the Kalman filter algorithm, because the dimension of $\mathbf{P}(t)$ is the square of $\hat{\mathbf{x}}(t)$; e.g.,

- Dimension of $\hat{\mathbf{x}}(t)$ in ECCO version 4 is 10 million (40 MB),
- Storage of a single $\mathbf{P}(t)$ is equivalent to 10 million $\hat{\mathbf{x}}(t)$ (400 TB),
- Integration of $\mathbf{P}(t)$ requires 20 million times of $\hat{\mathbf{x}}(t)$; Integrating $\hat{\mathbf{x}}(t)$ for 26-years requires 0.5 days, Integrating $\mathbf{P}(t)$ for 26-years requires 27,000 years.

Reduced-State Kalman Filter & Smoother

Estimate only dominant and/or particular elements of the error covariance matrix, thus reducing both storage and computational time of Kalman filtering.

range space

null space

$$\mathbf{x} = \mathbf{B}\mathbf{x}_r + \mathbf{N}\mathbf{x}_n \qquad \dim(\mathbf{x}_r) \ll \dim(\mathbf{x}) \qquad \langle \mathbf{x}\mathbf{x}^T \rangle \approx \mathbf{B}\langle \mathbf{x}_r\mathbf{x}_r^{\,T} \rangle \mathbf{B}^T$$

$$\mathbf{u} = \mathbf{B}'\mathbf{u}_r + \mathbf{N}'\mathbf{u}_n \qquad \dim(\mathbf{u}_r) \ll \dim(\mathbf{u}) \qquad \langle \mathbf{u}\mathbf{u}^T \rangle \approx \mathbf{B}'\langle \mathbf{u}_r\mathbf{u}_r^{\,T} \rangle \mathbf{B}'^T$$

Model equation;

$$\mathbf{x}(t,-) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$

$$\mathbf{B}\mathbf{x}_r(t,-) + \mathbf{N}\mathbf{x}_n(t,-) = \mathbf{A}\left[\mathbf{B}\mathbf{x}_r(t-1) + \mathbf{N}\mathbf{x}_n(t-1)\right] + \mathbf{G}\left[\mathbf{B}'\mathbf{u}_r(t-1) + \mathbf{N}\mathbf{u}_n(t-1)\right]$$

pseudo inverse

$$\begin{aligned}\mathbf{x}_r(t,-) &= \mathbf{B}^+\mathbf{A}\left[\mathbf{B}\mathbf{x}_r(t-1) + \mathbf{N}\mathbf{x}_n(t-1)\right] + \mathbf{B}^+\mathbf{G}\left[\mathbf{B}'\mathbf{u}_r(t-1) + \mathbf{N}\mathbf{u}_n(t-1)\right] \\ &= \mathbf{B}^+\mathbf{A}\mathbf{B}\mathbf{x}_r(t-1) + \mathbf{B}^+\mathbf{G}\mathbf{B}'\mathbf{u}_r(t-1) + \left[\mathbf{B}^+\mathbf{A}\mathbf{N}\mathbf{x}_n(t-1) + \mathbf{B}^+\mathbf{G}\mathbf{N}'\mathbf{u}_n(t-1)\right]\end{aligned}$$

If range & null space are physically independent such that

$$\mathbf{B}^+\mathbf{A}\mathbf{N}\mathbf{a} \approx \mathbf{B}^+\mathbf{N}\mathbf{b} = \mathbf{0} \qquad \mathbf{B}^+\mathbf{G}\mathbf{N}'\mathbf{c} \approx \mathbf{B}^+\mathbf{N}\mathbf{d} = \mathbf{0}$$

then $\mathbf{x}_r(t,-) = \mathbf{B}^+\mathbf{A}\mathbf{B}\mathbf{x}_r(t-1) + \mathbf{B}^+\mathbf{G}\mathbf{B}'\mathbf{u}_r(t-1)$ provides a “smaller” system for error evaluation.

Reduced-State Kalman Filter & Smoother

The reduced-state model operators (matrices) can be derived explicitly and then used in the Kalman filter algorithm to evaluate the state error covariance matrix.

$$\mathbf{x}_r(t,-) = \mathbf{B}^+\mathbf{A}\mathbf{B}\mathbf{x}_r(t-1) + \mathbf{B}^+\mathbf{G}\mathbf{B}'\mathbf{u}_r(t-1)$$

$$= \mathbf{A}_r\mathbf{x}_r(t-1) + \mathbf{G}_r\mathbf{u}_r(t-1)$$

pseudo inverse

where $\mathbf{A}_r \equiv \mathbf{B}^+\mathbf{A}\mathbf{B}$ $\quad \mathbf{G}_r \equiv \mathbf{B}^+\mathbf{G}\mathbf{B}'$

Given model $\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{G}\mathbf{u}(t) = F\left(\mathbf{x}(t),\mathbf{u}(t)\right)$

a perturbation satisfies $F\left(\mathbf{x}(t)+\boldsymbol{\delta},\mathbf{u}(t)\right) \approx F\left(\mathbf{x}(t),\mathbf{u}(t)\right) + \mathbf{A}\boldsymbol{\delta}$

$$\mathbf{B}^+\mathbf{A}\boldsymbol{\delta} \approx \mathbf{B}^+\left[F\left(\mathbf{x}(t)+\boldsymbol{\delta},\mathbf{u}(t)\right) - F\left(\mathbf{x}(t),\mathbf{u}(t)\right)\right]$$

Setting $\boldsymbol{\delta} \approx \mathbf{B}\mathbf{e}_j$ where $\mathbf{e}_j$ is the *j*-th column of identity matrix spanning $\mathbf{x}_r$

(column *j* of $\mathbf{A}_r$) = $\mathbf{A}_r\mathbf{e}_j = \mathbf{B}^+\left[F\left(\mathbf{x}+\mathbf{B}\mathbf{e}_j,\mathbf{u}\right) - F\left(\mathbf{x},\mathbf{u}\right)\right]$

Similarly

(column *j* of $\mathbf{G}_r$) = $\mathbf{G}_r\mathbf{e}'_j = \mathbf{B}^+\left[F\left(\mathbf{x},\mathbf{u}+\mathbf{B}'\mathbf{e}'_j\right) - F\left(\mathbf{x},\mathbf{u}\right)\right]$

Reduced-State Kalman Filter & Smoother

Approximate Kalman filter and RTS smoother with reduced-state error covariance matrix.

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\hat{\mathbf{x}}(t-1,+) = \hat{\mathbf{x}}(t-1) + \mathbf{L}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\hat{\mathbf{u}}(t-1,+) = \hat{\mathbf{u}}(t-1) + \mathbf{M}(t)\left[\hat{\mathbf{x}}(t,+) - \hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t)\mathbf{H}(t)^T\mathbf{R}(t)^{-1}$$

$$\mathbf{L}(t) \equiv \mathbf{P}(t-1)\mathbf{A}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{M}(t) \equiv \mathbf{Q}(t-1)\mathbf{G}^T\mathbf{P}(t,-)^{-1}$$

$$\mathbf{P} \approx \mathbf{B}\mathbf{P}_r\mathbf{B}^T$$

$$\mathbf{K}(t) \approx \mathbf{B}\mathbf{P}_r(t)\mathbf{B}^T\mathbf{H}(t)^T\mathbf{R}(t)^{-1}$$

$$\mathbf{L}(t) \approx \mathbf{B}\mathbf{P}_r(t-1)\mathbf{A}_r{}^T\mathbf{P}_r(t,-)^{-1}\mathbf{B}^+$$

$$\mathbf{M}(t) \approx \mathbf{B}'\mathbf{Q}_r(t-1)\mathbf{G}_r{}^T\mathbf{P}_r(t,-)^{-1}\mathbf{B}^+$$

Alternate form of Kalman filter

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

$$= \mathbf{P}(t)\mathbf{H}(t)^T\mathbf{R}(t)^{-1}$$

[Fukumori & Malanotte-Rizzoli, 1995]

Reduced-State Kalman Filter & Smoother

Some things to note.

1) Typically, do matrix-vector operations than matrix-matrix operations,

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}_r(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}_r(t) \equiv \mathbf{B}\mathbf{P}_r(t)\mathbf{B}^T\mathbf{H}(t)^T\mathbf{R}(t)^{-1}$$

2) An algorithmic adjoint can be used in place of left multiplying a matrix transpose,

3) Pseudo inverses do not necessarily need to be computed explicitly, e.g., objective mapping $\mathbf{B}$ 100,000×1,000

$$\mathbf{B}^+ = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T$$ 1,000×1,000

4) Multiple reduced-state filters/smothers can be combined (Partitioned filter & smoother; Fukumori, 2002)

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \sum_i \mathbf{K}_i(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

Other Approximate Kalman Filters

1) Singular Evolutive Extended Kalman (SEEK) filter [Pham et al., 1998]

Let the reduced state basis function evolve in time. $\mathbf{x}(t) \approx \mathbf{B}(t)\mathbf{x}_r(t)$

$$\mathbf{P}(t) \approx \mathbf{B}(t)\mathbf{P}_r(t)\mathbf{B}(t)^T \quad \text{where} \quad \mathbf{B}(t) = \mathbf{A}\mathbf{B}(t-1)$$

Then, $\mathbf{P}(t,-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T$

becomes

$$\begin{aligned}
\mathbf{B}(t)\mathbf{P}_r(t,-)\mathbf{B}(t)^T &\approx \mathbf{A}\mathbf{B}(t-1)\mathbf{P}_r(t-1)\mathbf{B}(t-1)^T\mathbf{A}^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T \\
&\approx \mathbf{B}(t)\mathbf{P}_r(t-1)\mathbf{B}(t)^T + \mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T \\
&\approx \mathbf{B}(t)\left[\mathbf{P}_r(t-1) + \mathbf{B}^+(t)\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T\mathbf{B}^+(t)^T\right]\mathbf{B}(t)^T
\end{aligned}$$

Thus $\mathbf{P}_r(t,-) \approx \mathbf{P}_r(t-1) + \mathbf{B}^+(t)\mathbf{G}\mathbf{Q}(t-1)\mathbf{G}^T\mathbf{B}^+(t)^T$

Other Approximate Kalman Filters

2) Ensemble Kalman filter (EnKF) [Evensen, 1994]

Instead of time-integrating the covariance matrix itself, integrate a small collection of states (ensemble) and use its sample covariance instead.

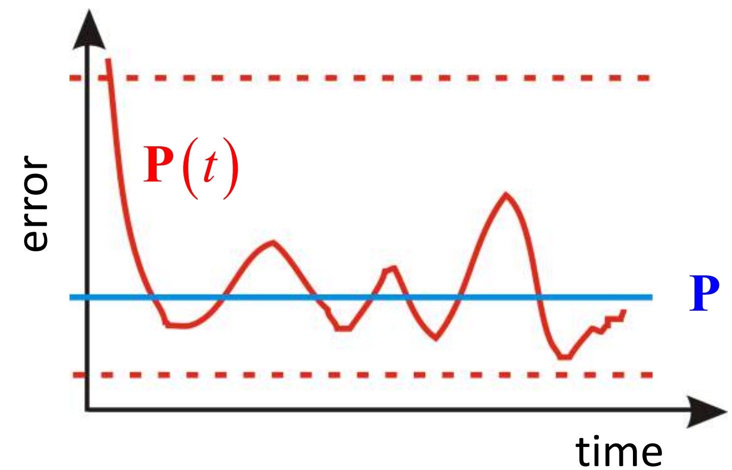
$$\mathbf{P} \approx \frac{1}{L}\mathbf{X}\mathbf{X}^T \quad \text{where} \quad \mathbf{X} \equiv \begin{pmatrix} \mathbf{x}_1 - \bar{\mathbf{x}} & \mathbf{x}_2 - \bar{\mathbf{x}} & \cdots & \mathbf{x}_L - \bar{\mathbf{x}} \end{pmatrix} \qquad \bar{\mathbf{x}} \equiv \frac{1}{L}\sum_{i=1}^{L}\mathbf{x}_i$$

$$\mathbf{x}_i(t) = \mathbf{A}\mathbf{x}_i(t-1) + \mathbf{G}\mathbf{u}_i(t-1) \qquad \text{for each member } i$$

Other Approximate Kalman Filters

3) Steady-state Kalman filter and RTS smoother

Estimation errors often approach an asymptotic limit which could be used in place of the time-variable error, thus significantly reducing the computational requirements of Kalman filtering and RTS smoothing.



$$\mathbf{P}(t,-) \approx \mathbf{P}(-)$$

$$\mathbf{P}(t) \approx \mathbf{P}$$

$$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}(t)^T \mathbf{R}(t)^{-1} \quad \Rightarrow \quad \mathbf{K}(t) \approx \mathbf{P}\mathbf{H}(t)^T \mathbf{R}(t)^{-1}$$

$$\mathbf{L}(t) = \mathbf{P}(t-1)\mathbf{A}^T \mathbf{P}(t,-)^{-1} \quad \Rightarrow \quad \mathbf{L} \approx \mathbf{P}\mathbf{A}^T \mathbf{P}(-)^{-1}$$

$$\mathbf{M}(t) = \mathbf{Q}(t-1)\mathbf{G}^T \mathbf{P}(t,-)^{-1} \quad \Rightarrow \quad \mathbf{M} \approx \mathbf{Q}\mathbf{G}^T \mathbf{P}(-)^{-1}$$

[Fukumori et al., 1993]

Adjoint Error Estimate

The adjoint method has an advantage of not requiring explicit state error estimates, but the lack thereof is also a shortcoming.

The Hessian of the least-squares problem's objective function provides the formal error estimates' inverse for the adjoint method, but the Hessian is much too large to be evaluated explicitly, necessitating approximations for error evaluation.

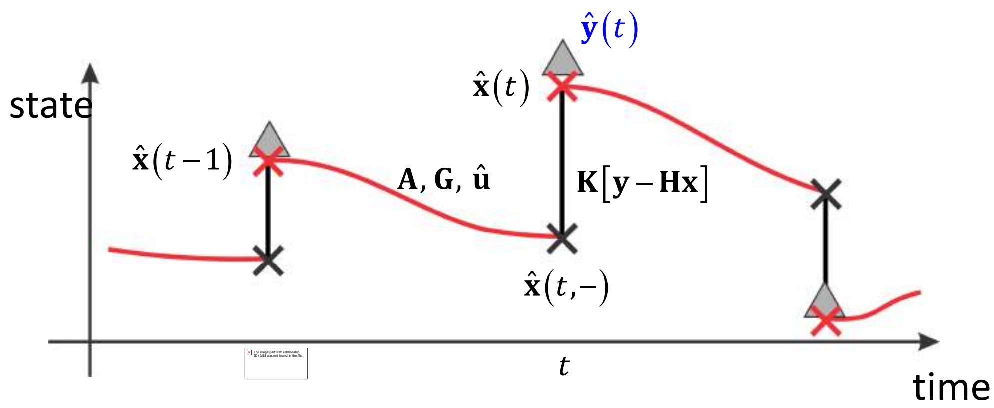
1. AD tools can generate code to evaluate Hessian-vector products, which can be used to compute the dominant modes of the error,
2. The tangent linear model can be used to map the control error to particular state errors of interest.

[Moore et al., 2011; Kalmikov & Heimbach, 2014]

Other Common Data Assimilation Methods

Simplified methods of data assimilation are still in common use today owing to their simplicity compared to formal state estimation methods.

1) Nudging
2) Optimal Interpolation (OI)
3) Three-dimensional Variational method (3DVAR)



Other Common Data Assimilation Methods

1) Nudging

$$\hat{x}_i(t) = \hat{x}_i(t,-) + \gamma\left[\hat{y}_i(t) - \hat{x}_i(t,-)\right]$$

where $\hat{y}_i(t)$ is data processed to be coincident in space and time with the model variable x_i and γ is a prescribed coefficient (nudging coefficient) describing relaxation of model variable x_i towards $\hat{y}_i(t)$

Nudging differs from Kalman filter in weighting coefficient and how data is used.

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K}(t) \equiv \mathbf{P}(t,-)\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}(t,-)\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

Other Common Data Assimilation Methods

2) Optimal Interpolation (OI)

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t,-) + \mathbf{K}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t,-)\right]$$

$$\mathbf{K} \equiv \mathbf{P}\mathbf{H}(t)^T\left[\mathbf{H}(t)\mathbf{P}\mathbf{H}(t)^T + \mathbf{R}(t)\right]^{-1}$$

In OI, covariance matrix $\mathbf{P}$ is prescribed whereas in Kalman filtering $\mathbf{P}$ is computed taking both model physics and past observations into account.

Other Common Data Assimilation Methods

3) Three-dimensional variational method (3DVAR)

$$\mathrm{J} \equiv \left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t)\right]^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t)\right] + \left[\hat{\mathbf{x}}(t)-\hat{\mathbf{x}}(t,-)\right]^T \mathbf{P}^{-1}\left[\hat{\mathbf{x}}(t)-\hat{\mathbf{x}}(t,-)\right]$$

3DVAR combines coincident data with model in a least-squares sense, minimizing the misfit above iteratively by descent optimization with prescribed weights.

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}(t)} \equiv -\mathbf{H}(t)^T \mathbf{R}(t)^{-1}\left[\hat{\mathbf{y}}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t)\right] + \mathbf{P}^{-1}\left[\hat{\mathbf{x}}(t)-\hat{\mathbf{x}}(t,-)\right]$$

Given the equivalence of least-squares and minimum variance estimate, 3DVAR is equivalent to OI.

Summary of Covariance & Other Data Assimilation

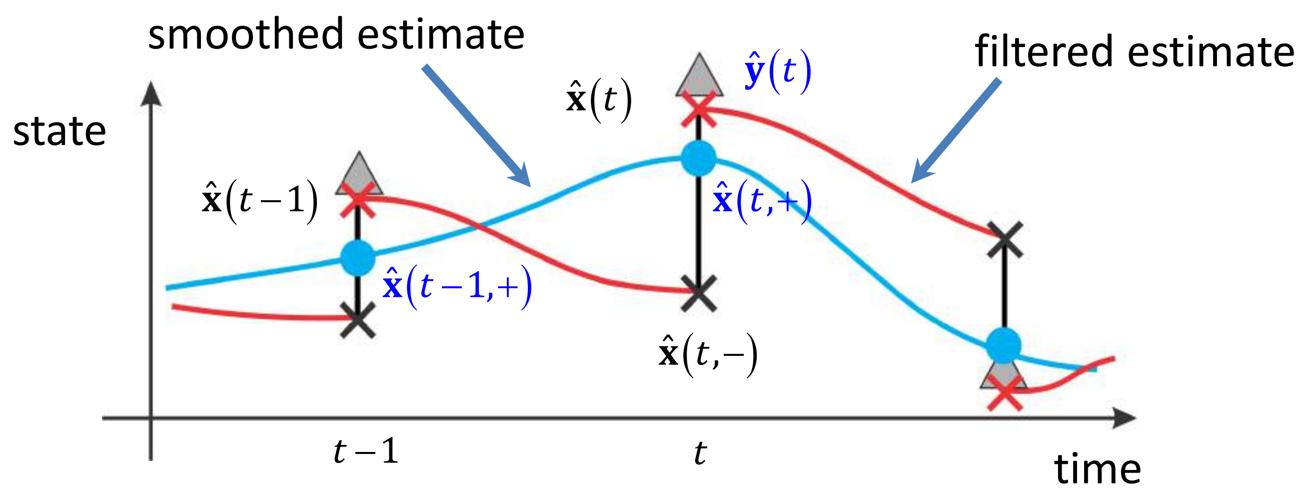
1. State estimations' error covariance can be derived by approximation in space, time and variables (e.g., state reduction, Hessian products), providing fundamental measures of the estimate,

2. Resulting errors also permit application of Kalman filtering and smoothing to practical problems,

3. Common methods of data assimilation (e.g., nudging, OI, 3DVAR) are filtering algorithms with prescribed state errors instead of computed ones from first principles.

Common data assimilation lack smoothing.

Concluding Remarks



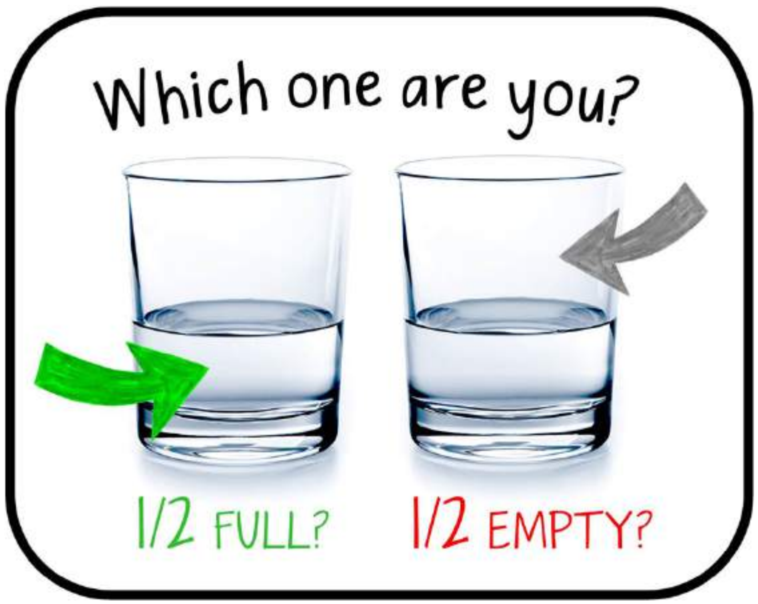
1) Smoothed estimates distinguish state estimation (and ECCO) from common data assimilation,

2) The unique virtue of smoothed estimates is its description of the state's temporal evolution (blue curve), not the discrete estimates (blue circles) per se.

The model and its adjoint permit interrogation of this evolution; e.g., budgets, adjoint decompositions.

Concluding Remarks

3) While the fidelity of state estimation will continue to improve (e.g., more data, better models, better estimation), the results where it already has skill can be used to learn something new about the ocean, and is ripe for innovation (e.g., budget analysis, adjoint decomposition).



References

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99,** 10143-10162, doi:10.1029/94jc00572.

Fu, L. L., I. Fukumori, and R. N. Miller, 1993: Fitting Dynamic-Models to the Geosat Sea-Level Observations in the Tropical Pacific-Ocean .2. A Linear, Wind-Driven Model. J Phys Oceanogr, 23, 2162-2181, doi:10.1175/1520-0485(1993)023<2162:Fdmttg>2.0.Co;2

Fukumori, I., J. Benveniste, C. Wunsch, and D. B. Haidvogel, 1993: Assimilation of Sea-Surface Topography into an Ocean Circulation Model Using a Steady-State Smoother. *J Phys Oceanogr*, **23,** 1831-1855, doi:10.1175/1520-0485(1993)023<1831:Aossti>2.0.Co;2.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An Approximate Kalman Filter for Ocean Data Assimilation - an Example with an Idealized Gulf-Stream Model. *J Geophys Res-Oceans*, **100,** 6777-6793, doi:10.1029/94jc03084.

Fukumori, I., R. Raghunath, L. L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results? *J Geophys Res-Oceans*, **104,** 25647-25665, doi:10.1029/1999jc900193.

Fukumori, I., 2002: A partitioned Kalman filter and smoother. *Mon Weather Rev*, **130,** 1370-1383, doi:10.1175/1520-0493(2002)130<1370:Apkfas>2.0.Co;2.

Kalmikov, A. G., and P. Heimbach, 2014: A Hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM Journal on Scientific Computing*, **36,** S267-S295, doi:10.1137/130925311.

Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, and J. Zavala-Garay, 2011: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems Part I - System overview and formulation. *Prog Oceanogr*, **91,** 34-49, doi:10.1016/j.pocean.2011.05.004.

Tuan Pham, D., J. Verron, and M. Christine Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16,** 323-340, doi:10.1016/S0924-7963(97)00109-7.

Jet Propulsion Laboratory
California Institute of Technology