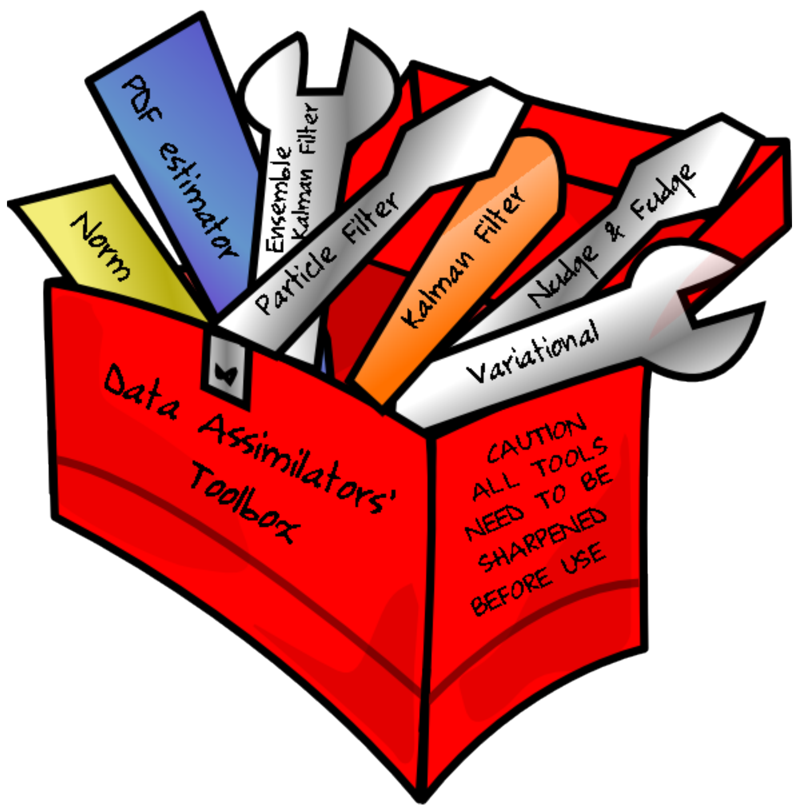
State estimation from imperfect knowledge - a data assimilation toolbox

Probability, Stochastic modelling and Financial mathematics seminar, 2nd Oct 2014, Leeds.

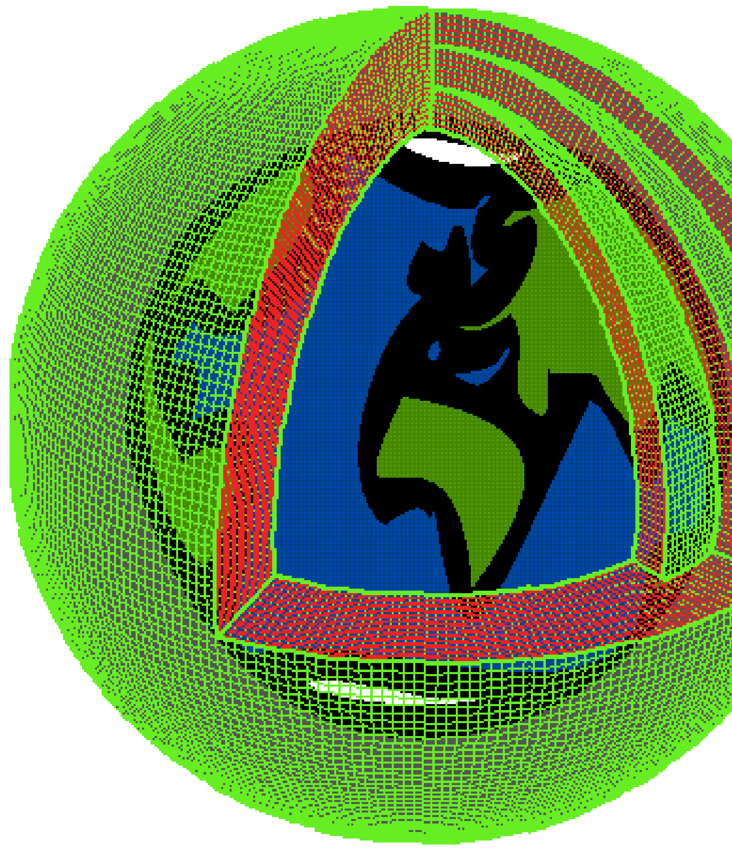
Ross Bannister
National Centre for Earth Observation, University of Reading, UK



The data assimilation problem (with some jargon)

- Data assimilation:
  - State estimation, data fusion, history matching, retrieving, inverse modelling.
- Estimate possible 'truths' of system.
  - State vector, $\mathbf{x} \in \mathbb{R}^n$.
- Observations, $\mathbf{y} \in \mathbb{R}^p$:
  - Some representation of the truth.
  - Direct and indirect observations.
- Models, $\mathcal{M}, \mathcal{H}$:
  - Linking $\mathbf{x}$ to $\mathbf{y}$.
- Prior information, $\mathbf{x}_\mathrm{B}$ (background), $\mathbf{x}_\mathrm{f}$ (forecast).
- Posterior, $\mathbf{x}_\mathrm{A}$ (analysis).

- Uncertainty:
  - Errrors are everywhere - in state, in observations, in model, in representation ...
  - PDF, $p(\mathbf{x})$.
  - Ensemble, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$.
- Constraints (strong and weak):
  - Dynamical model.
  - Physical balance.
  - Smoothness.
- Applications:
  - NWP - Numerical Weather Prediction.
  - Atmospheric/ocean physics.
  - Astronautics/aeronautics.
  - Astrophysics.
  - Seismology.
  - &c &c.

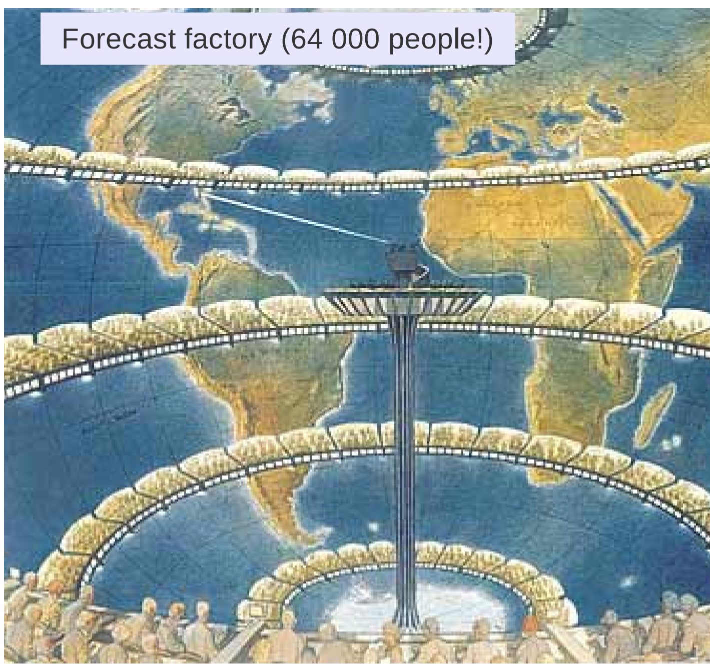


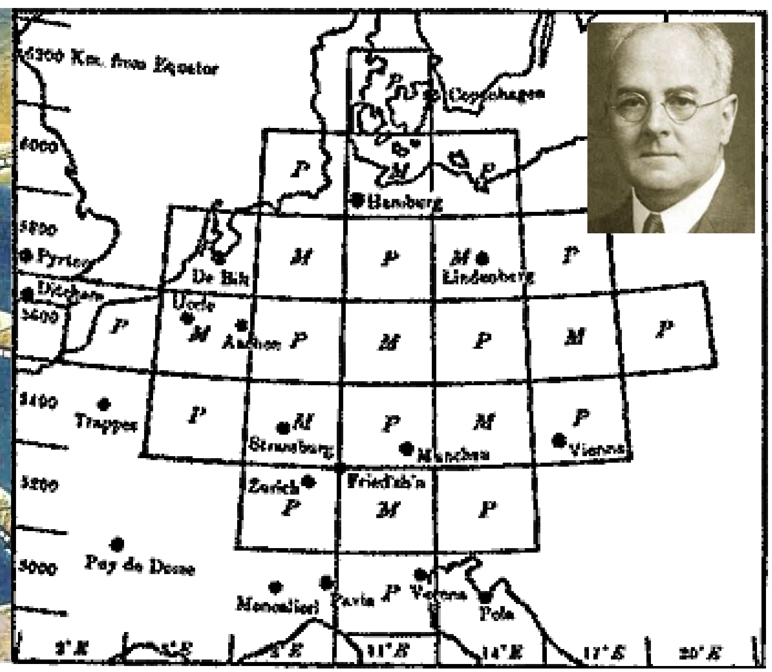
Typical values for NWP:
$n \sim 10^9$, $p \sim 10^8$.

The NWP problem

Data assimilation is used in weather forecasting to estimate the initial conditions of a large numerical model of the atmosphere.

- L.F. Richardson (1922) attempted a hind-cast (by hand!) for 20th May 1910.
- Primitive eq-based forecast model (eqs of motion used for large-scale flows): resolution $\Delta\lambda = 3°$, $\Delta\phi = 1.8°$, 5 vertical levels.
- 'Data assimilation' was done for mass variables (T, p) separately from wind (u, v) (i.e. univariate) - interpolate obs subjectively.
- A disastrous forecast: $\Delta P/\Delta t \approx 145$ hPa $/6$ hours (note surface pressure is ~ 1000 hPa).
- Catastrophic growth rate not due to an inadequate model, but due to imbalance between mass and wind at $t = 0$.





Meteorological balances

Initial conditions of meteorological models need to be appropriately balanced.

$$
\begin{aligned}
\frac{D\delta'}{Dt} &= \mathcal{M}' + \mathcal{W}' + \text{other terms}, \\
\frac{Dw'}{Dt} &= \mathcal{P}' + \mathcal{T}' + \text{other terms}, \\
\text{horiz. wind div.: } \delta' &= \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}, \\
\text{'mass' term: } \mathcal{M}' &= c_p\theta_0\nabla_z\Pi' + c_p\theta'\nabla_z\Pi_0, \\
\text{'wind' term: } \mathcal{W}' &= -f\mathbf{k}\cdot(\nabla\times\mathbf{u}') + \mathbf{k}\cdot(\mathbf{u}'\times\nabla f), \\
\text{'pressure' term: } \mathcal{P}' &= \theta_0\frac{\partial\Pi'}{\partial z}, \\
\text{'temperature' term: } \mathcal{T}' &= \theta'\frac{\partial\Pi_0}{\partial z}, \\
\Pi_0, \Pi' &: \text{Ref. and pert. pressure} \\
\theta_0, \theta' &: \text{Ref. and pert. temperature}
\end{aligned}
$$

$$
\begin{aligned}
f &= 2\Omega\sin(y/a), \\
\Omega &= 7.29\times 10^{-5}\text{rads}^{-1}, \\
a &= 6.371\times 10^{6}\text{m}, \\
g &= 9.806\text{ms}^{-1}, \\
\mathbf{k} &= \text{vertical unit vector.}
\end{aligned}
$$

Meteorological balances

Initial conditions of meteorological models need to be appropriately balanced.

$$
\begin{aligned}
\cancel{\frac{D\delta'}{Dt}} &= \mathcal{M}' + \mathcal{W}' + \cancel{\text{other terms}}, \\
\cancel{\frac{Dw'}{Dt}} &= \mathcal{P}' + \mathcal{T}' + \cancel{\text{other terms}}, \\
\text{horiz. wind div.: } \delta' &= \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}, \\
\text{'mass' term: } \mathcal{M}' &= c_p \theta_0 \nabla_z \Pi' + c_p \theta' \nabla_z \Pi_0, \\
\text{'wind' term: } \mathcal{W}' &= -f \mathbf{k} \cdot (\nabla \times \mathbf{u}') + \mathbf{k} \cdot (\mathbf{u}' \times \nabla f), \\
\text{'pressure' term: } \mathcal{P}' &= \theta_0 \frac{\partial \Pi'}{\partial z}, \\
\text{'temperature' term: } \mathcal{T}' &= \theta' \frac{\partial \Pi_0}{\partial z}, \\
\Pi_0, \Pi' &: \text{ Ref. and pert. pressure} \\
\theta_0, \theta' &: \text{ Ref. and pert. temperature}
\end{aligned}
$$

$$
\begin{aligned}
f &= 2\Omega \sin(y/a), \\
\Omega &= 7.29 \times 10^{-5} \text{rads}^{-1}, \\
a &= 6.371 \times 10^{6} \text{m}, \\
g &= 9.806 \text{ms}^{-1}, \\
\mathbf{k} &= \text{vertical unit vector.}
\end{aligned}
$$

Geostrophic and hydrostatic balances apply when Ro and W/U are small (extra-tropical large-scale flow).

$$\text{Ro} = \frac{U}{fL} = \mathcal{O}(10^{-1}), \ \frac{W}{U} = \mathcal{O}(10^{-2}).$$

- Perfect geostrophic balance

$$\mathcal{M}' + \mathcal{W}' = 0.$$

- Perfect hydrostatic balance

$$\mathcal{P}' + \mathcal{T}' = 0.$$

Meteorological balances

Initial conditions of meteorological models need to be appropriately balanced.

$$
\begin{aligned}
\cancel{\frac{D\delta'}{Dt}} &= \mathcal{M}' + \mathcal{W}' + \cancel{\text{other terms}},\\
\cancel{\frac{Dw'}{Dt}} &= \mathcal{P}' + \mathcal{T}' + \cancel{\text{other terms}},\\
\text{horiz. wind div.: } \delta' &= \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y},\\
\text{'mass' term: } \mathcal{M}' &= c_p \theta_0 \nabla_z \Pi' + c_p \theta' \nabla_z \Pi_0,\\
\text{'wind' term: } \mathcal{W}' &= -f\mathbf{k}\cdot(\nabla \times \mathbf{u}') + \mathbf{k}\cdot(\mathbf{u}' \times \nabla f),\\
\text{'pressure' term: } \mathcal{P}' &= \theta_0 \frac{\partial \Pi'}{\partial z},\\
\text{'temperature' term: } \mathcal{T}' &= \theta' \frac{\partial \Pi_0}{\partial z},\\
\Pi_0, \Pi' &: \text{ Ref. and pert. pressure}\\
\theta_0, \theta' &: \text{ Ref. and pert. temperature}
\end{aligned}
$$

$$
\begin{aligned}
f &= 2\Omega \sin(y/a),\\
\Omega &= 7.29 \times 10^{-5} \text{rads}^{-1},\\
a &= 6.371 \times 10^{6} \text{m},\\
g &= 9.806 \text{ms}^{-1},\\
\mathbf{k} &= \text{vertical unit vector.}
\end{aligned}
$$

Geostrophic and hydrostatic balances apply when Ro and W/U are small (extra-tropical large-scale flow).

$$
\text{Ro} = \frac{U}{fL} = \mathcal{O}(10^{-1}), \quad \frac{W}{U} = \mathcal{O}(10^{-2}).
$$

- Perfect geostrophic balance

$$\mathcal{M}' + \mathcal{W}' = 0.$$

- Perfect hydrostatic balance

$$\mathcal{P}' + \mathcal{T}' = 0.$$

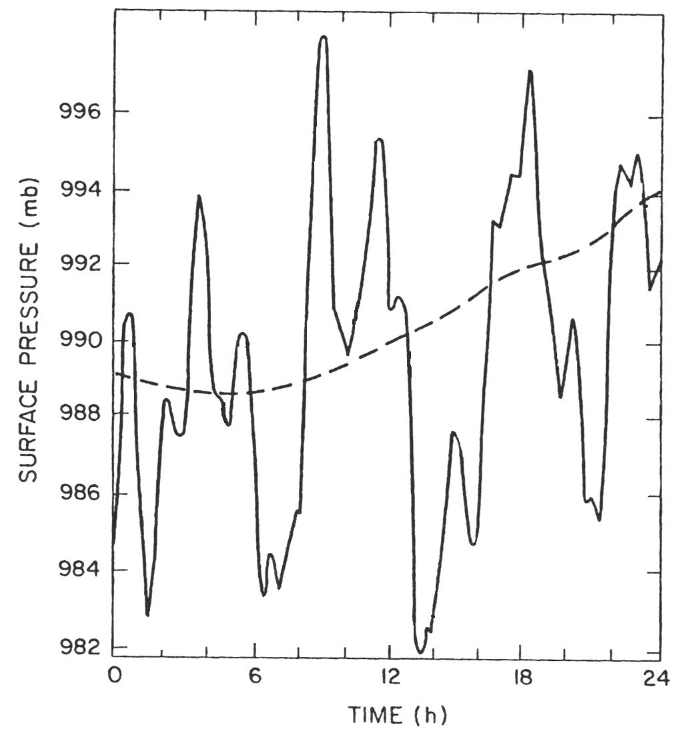
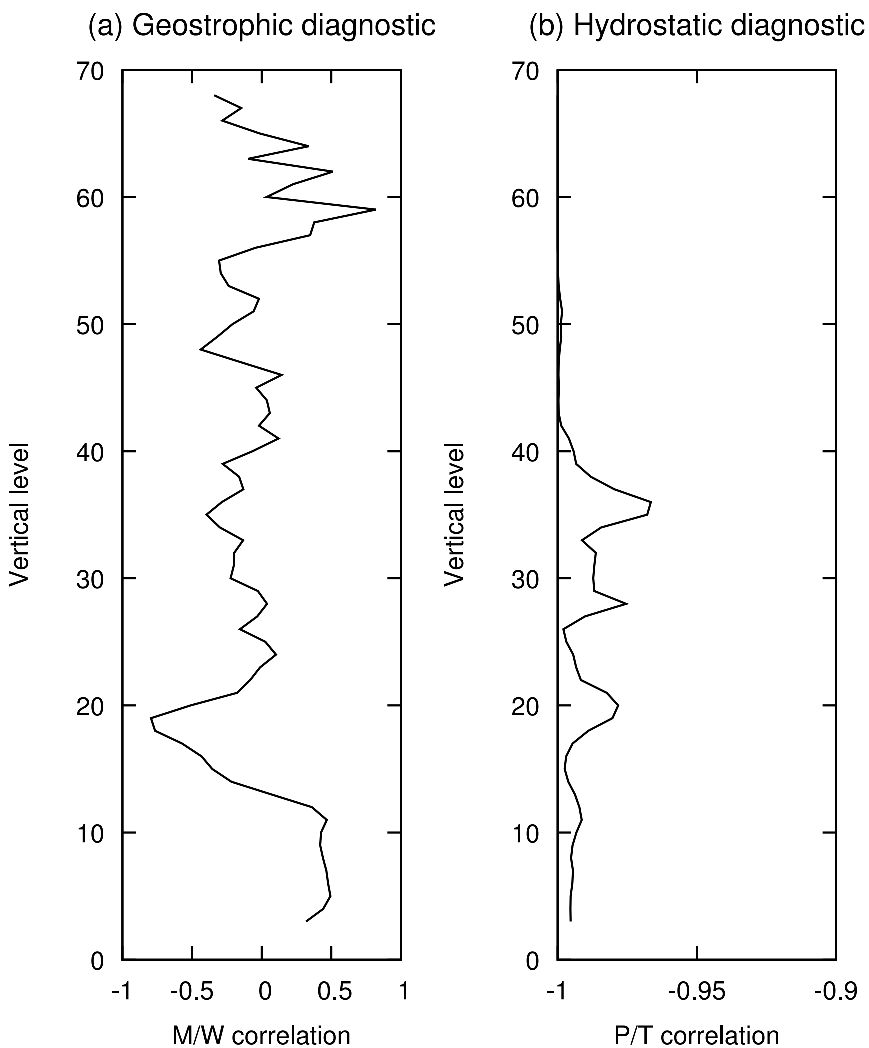
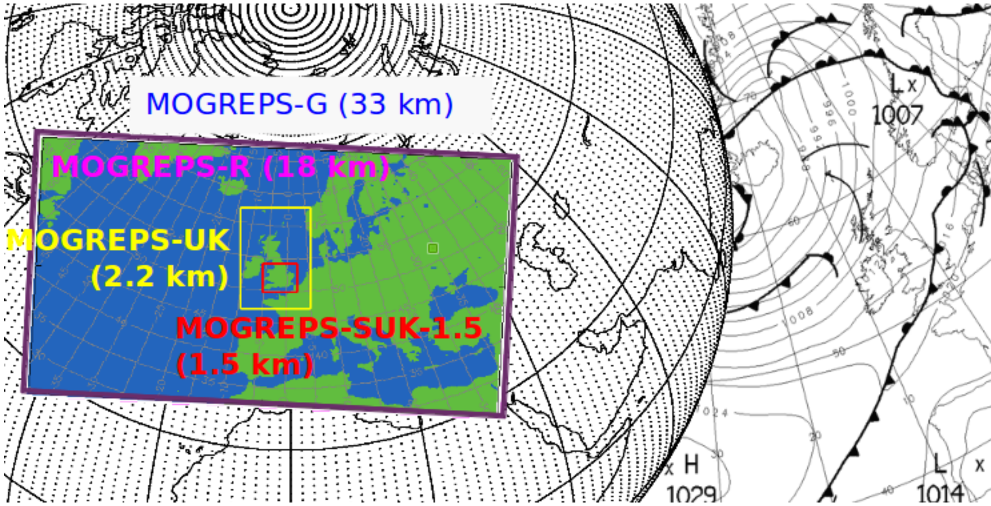


Figure 6.1 Surface pressure as a function of time during the integration of a primitive equations model. Uninitialized (solid), initialized (dashed). (After Williamson and Temperton, *Mon. Wea. Rev.* **109**: 745, 1981. The American Meteorological Society.)

Example geostrophic and hydrostatic correlations



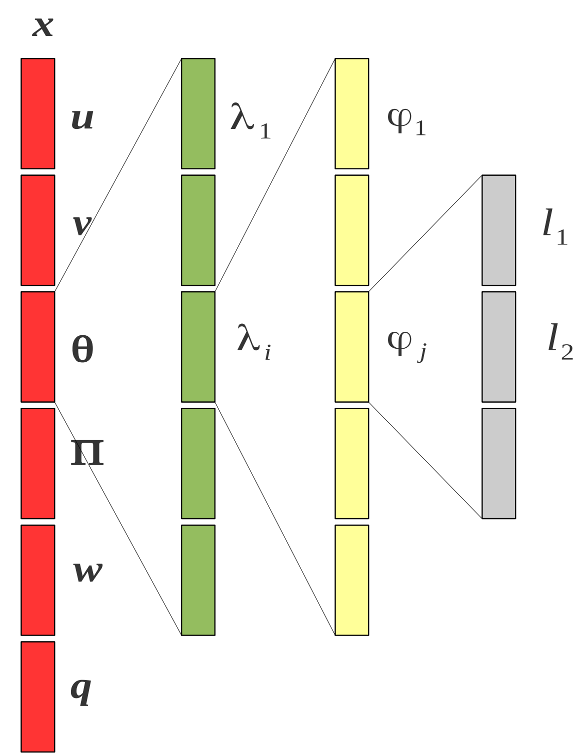


- MOGREPS = Met Office Global and Regional Ensemble Prediction System.
- ETKF = Ensemble Transform Kalman Filter.
- 24 members.
- NDP 1.5 km grid.
- NDP = Nowcasting Demonstration Project.
- Case study: 20/09/11, 15:00 UT.

Correlation $-1 \Rightarrow$perfectly balanced.

Problems to be faced in data assimilation

- General theoretical problems.
  - Representation and quantification of uncertainty.
  - Sampling from a PDF.
  - Non-Gaussian statistics and non-linear models.
  - Understanding the observations.
  - Unknown unknowns.
- General technical problems.
  - Large volume of information.
  - Time efficiency.
  - Parallelizability.
- Plus specific geophysical problems.
  - Constraining an appropriate 'closeness' of balance in the analysis.
    - ∗ In large-scale systems balance is understood ($>$ 100s km - e.g. cyclones, anticyclones).
    - ∗ In convective-scale systems balance is not well understood (sub-km - 10s km - e.g. thunderstorms).
    - ∗ Multi-scale.
  - Moisture and clouds.
  - Phase errors.



A numerical analyst's approach: least squares

- No prior information
  - Have:
    - p observations in $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$, $\mathbf{y}_\mathrm{o} = \mathbf{y}_\mathrm{t} + \boldsymbol{\epsilon}_y$.
    - n unknowns in $\mathbf{x} \in \mathbb{R}^n$.
    - Obs operator / forward model $\mathbf{y}_\mathrm{m} = \mathbf{H}\mathbf{x}$.
  - $\mathcal{J}_1[\mathbf{x}] = \frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x})$.
  - Normal equations:

$$\nabla_x \mathcal{J}_1 = -\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{A}) = 0.$$

    - $\mathbf{x}_\mathrm{A} = \left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}_\mathrm{o}$.
    - OK if $\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}$ is full rank, ill-posed otherwise (always ill-posed if $p < n$ - the case for NWP).

A numerical analyst's approach: least squares

- No prior information
  - Have:
    - p observations in $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$, $\mathbf{y}_\mathrm{o} = \mathbf{y}_\mathrm{t} + \boldsymbol{\epsilon}_y$.
    - n unknowns in $\mathbf{x} \in \mathbb{R}^n$.
    - Obs operator / forward model $\mathbf{y}_\mathrm{m} = \mathbf{H}\mathbf{x}$.
  - $\mathcal{J}_1[\mathbf{x}] = \frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x})$.
  - Normal equations:

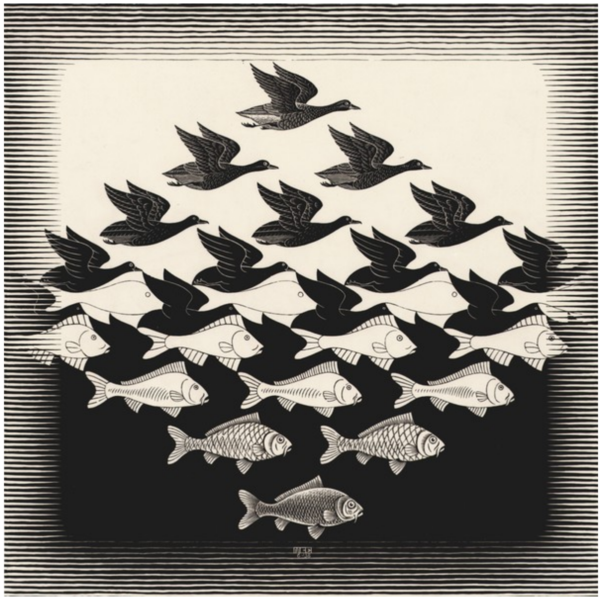
$$\nabla_x \mathcal{J}_1 = -\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{A}) = 0.$$

    - $\mathbf{x}_\mathrm{A} = \left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}_\mathrm{o}$.
    - OK if $\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}$ is full rank, ill-posed otherwise (always ill-posed if $p < n$ - the case for NWP).

- Prior information - regularizing the problem
  - Have:
    - p observations in $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$, $\mathbf{y}_\mathrm{o} = \mathbf{y}_\mathrm{t} + \boldsymbol{\epsilon}_y$.
    - n unknowns in $\mathbf{x} \in \mathbb{R}^n$.
    - A-priori $\mathbf{x}_\mathrm{B} \in \mathbb{R}^n$, $\mathbf{x}_\mathrm{B} = \mathbf{x}_\mathrm{t} + \boldsymbol{\epsilon}_\mathrm{B}$.

$$\begin{aligned}
\mathcal{J}_2[\mathbf{x}] &= \frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_\mathrm{B})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{B} \\
\nabla_x \mathcal{J}_2 &= -\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{A}) + \mathbf{P}_\mathrm{f}^{-1}(\mathbf{x}_\mathrm{A} - \mathbf{x}_\mathrm{B}) = 0,
\end{aligned}$$

$$\begin{aligned}
\mathbf{x}_\mathrm{A} &= \mathbf{x}_\mathrm{B} + (\mathbf{P}_\mathrm{f}^{-1} + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{B}), \\
&= \mathbf{x}_\mathrm{B} + \mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T})^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{B}).
\end{aligned}$$



Bayesian approach

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$
$$\propto P(B|A) \times P(A).$$

Let A be the event $\mathbf{x} \in \mathbb{R}^n$ and B be the event $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$:

$$\underbrace{P(\mathbf{x}|\mathbf{y}_\mathrm{o})}_{\text{posterior}} \propto \underbrace{P(\mathbf{y}_\mathrm{o}|\mathbf{x})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x})}_{\text{prior}}.$$

Bayesian approach

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$
$$\propto P(B|A) \times P(A).$$

Let A be the event $\mathbf{x} \in \mathbb{R}^n$ and B be the event $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$:

$$\underbrace{P(\mathbf{x}|\mathbf{y}_\mathrm{o})}_{\text{posterior}} \propto \underbrace{P(\mathbf{y}_\mathrm{o}|\mathbf{x})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x})}_{\text{prior}}.$$

In the n-dimensional state space:

- Suppose that we represent the PDF on a discreet grid with 10 values per element.
  - Number of pieces of information to represent the PDF is 10^n.
  - Grossly impossible, even with this modest resolution! "The curse of dimensionality."
- Particle filters (with a proposal density) are a possible tool - development needed.

Approximation: assume Gaussian statistics

Form of n-dimensional Gaussian for $\boldsymbol{\epsilon} = (\epsilon_1, \dots \epsilon_n)^{\mathrm{T}}$ with mean $\boldsymbol{\mu} = (\mu_1, \dots \mu_n)^{\mathrm{T}} \in \mathbb{R}^n$ and covariance $\mathbf{S} \in \mathbb{R}^{n \times n}$:

$$\begin{aligned} \underbrace{P(\mathbf{x}|\mathbf{y}_{\mathrm{o}})}_{\text{posterior}} &\propto \underbrace{P(\mathbf{y}_{\mathrm{o}}|\mathbf{x})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x})}_{\text{prior}}, \\ \boldsymbol{\epsilon} &\sim N(\boldsymbol{\mu}, \mathbf{S}), \\ P(\boldsymbol{\epsilon}) &= \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{S})}} = \exp -\frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\mu}). \end{aligned}$$

- Likelihood: $\boldsymbol{\epsilon} \to \mathbf{y}_{\mathrm{o}}$, $\boldsymbol{\mu} \to \mathbf{H}(\mathbf{x})$, $\mathbf{S} \to \mathbf{R} \in \mathbb{R}^{p \times p}$.
- Prior: $\boldsymbol{\epsilon} \to \mathbf{x}$, $\boldsymbol{\mu} \to \mathbf{x}_{\mathrm{B}}$, $\mathbf{S} \to \mathbf{P}_{\mathrm{f}} \in \mathbb{R}^{n \times n}$.

Approximation: assume Gaussian statistics

Form of n-dimensional Gaussian for $\boldsymbol{\epsilon} = (\epsilon_1, \ldots \epsilon_n)^{\mathrm{T}}$ with mean $\boldsymbol{\mu} = (\mu_1, \ldots \mu_n)^{\mathrm{T}} \in \mathbb{R}^n$ and covariance $\mathbf{S} \in \mathbb{R}^{n \times n}$:

$$
\begin{aligned}
\underbrace{P(\mathbf{x}|\mathbf{y}_{\mathrm{o}})}_{\text{posterior}} &\propto \underbrace{P(\mathbf{y}_{\mathrm{o}}|\mathbf{x})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x})}_{\text{prior}}, \\
\boldsymbol{\epsilon} &\sim N(\boldsymbol{\mu}, \mathbf{S}), \\
P(\boldsymbol{\epsilon}) &= \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{S})}} = \exp -\frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\mu}).
\end{aligned}
$$

- Likelihood: $\boldsymbol{\epsilon} \to \mathbf{y}_{\mathrm{o}}$, $\boldsymbol{\mu} \to \mathbf{H}(\mathbf{x})$, $\mathbf{S} \to \mathbf{R} \in \mathbb{R}^{p \times p}$.
- Prior: $\boldsymbol{\epsilon} \to \mathbf{x}$, $\boldsymbol{\mu} \to \mathbf{x}_{\mathrm{B}}$, $\mathbf{S} \to \mathbf{P}_{\mathrm{f}} \in \mathbb{R}^{n \times n}$.

$$
\begin{aligned}
\overbrace{P(\mathbf{x}|\mathbf{y}_{\mathrm{o}})}^{\text{posterior}} &\propto \overbrace{\exp -\frac{1}{2}(\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x}))}^{\text{likelihood}} \times \overbrace{\exp -\frac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{B}})^{\mathrm{T}} \mathbf{P}_{\mathrm{f}}^{-1} (\mathbf{x} - \mathbf{x}_{\mathrm{B}})}^{\text{prior}}, \\
&\propto \exp -\frac{1}{2}\left[(\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_{\mathrm{B}})^{\mathrm{T}} \mathbf{P}_{\mathrm{f}}^{-1} (\mathbf{x} - \mathbf{x}_{\mathrm{B}})\right], \\
&\propto \exp -\mathcal{J}_2[\mathbf{x}], \\
\text{cost function: } \mathcal{J}_2[\mathbf{x}] &= \frac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{B}})^{\mathrm{T}} \mathbf{P}_{\mathrm{f}}^{-1} (\mathbf{x} - \mathbf{x}_{\mathrm{B}}) + \frac{1}{2}(\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}_{\mathrm{o}} - \mathbf{H}(\mathbf{x})),
\end{aligned}
$$

In the n-dimensional state space:

- $\mathbf{P}_{\mathrm{f}}$-matrix needs $\sim n^2/2$ pieces of information.
- Still prohibitive for large n.

$$
\begin{aligned}
\mathbf{R} &= \left\langle \boldsymbol{\epsilon}_y \boldsymbol{\epsilon}_y^{\mathrm{T}} \right\rangle, & \mathbf{P}_{\mathrm{f}} &= \left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle, \\
\mathbf{y}_{\mathrm{o}} &= \mathbf{y} + \boldsymbol{\epsilon}_y, & \mathbf{x}_{\mathrm{B}} &= \mathbf{x} + \boldsymbol{\epsilon}_{\mathrm{B}}.
\end{aligned}
$$

Importance of $\mathrm{P_f}$ in data assimilation

Importance of $\mathrm{P_f}$ in data assimilation

Very important

The forward model - generalization and linearization

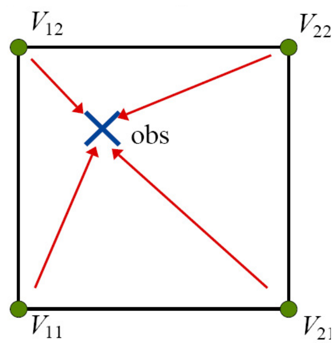
$$
\begin{aligned}
\mathbf{Hx} &\to \tilde{\mathcal{H}}(\mathbf{x}), \\
&\to \begin{pmatrix} \mathcal{H}_0(\mathbf{x}) \\ \mathcal{H}_1(\mathcal{M}_{1\leftarrow 0}(\mathbf{x})) \\ \vdots \\ \mathcal{H}_T(\mathcal{M}_{T\leftarrow 0}(\mathbf{x})) \end{pmatrix}, \\
&\approx \begin{pmatrix} \mathcal{H}_0(\mathbf{x}_\mathrm{B}) \\ \mathcal{H}_1(\mathcal{M}_{1\leftarrow 0}(\mathbf{x}_\mathrm{B})) \\ \vdots \\ \mathcal{H}_T(\mathcal{M}_{T\leftarrow 0}(\mathbf{x}_\mathrm{B})) \end{pmatrix} + \begin{pmatrix} \mathbf{H}_0\delta\mathbf{x} \\ \mathbf{H}_1\mathbf{M}_{1\leftarrow 0}\delta\mathbf{x} \\ \vdots \\ \mathbf{H}_T\mathbf{M}_{T\leftarrow 0}\delta\mathbf{x} \end{pmatrix}, \\
\mathbf{x} &= \mathbf{x}_\mathrm{B} + \delta\mathbf{x}.
\end{aligned}
$$

The forward model - generalization and linearization

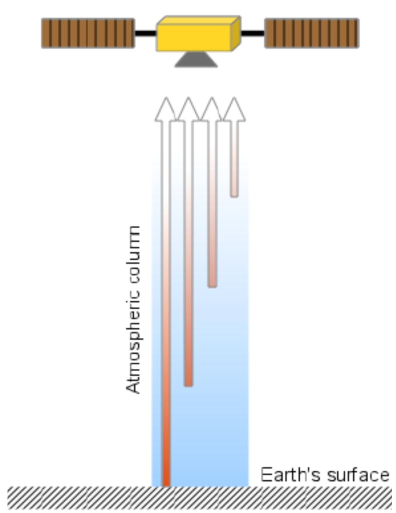
$$
\begin{aligned}
\mathbf{H}\mathbf{x} &\rightarrow \tilde{\mathcal{H}}(\mathbf{x}), \\
&\rightarrow \begin{pmatrix} \mathcal{H}_0(\mathbf{x}) \\ \mathcal{H}_1(\mathcal{M}_{1\leftarrow 0}(\mathbf{x})) \\ \vdots \\ \mathcal{H}_T(\mathcal{M}_{T\leftarrow 0}(\mathbf{x})) \end{pmatrix}, \\
&\approx \begin{pmatrix} \mathcal{H}_0(\mathbf{x}_\mathrm{B}) \\ \mathcal{H}_1(\mathcal{M}_{1\leftarrow 0}(\mathbf{x}_\mathrm{B})) \\ \vdots \\ \mathcal{H}_T(\mathcal{M}_{T\leftarrow 0}(\mathbf{x}_\mathrm{B})) \end{pmatrix} + \begin{pmatrix} \mathbf{H}_0\delta\mathbf{x} \\ \mathbf{H}_1\mathbf{M}_{1\leftarrow 0}\delta\mathbf{x} \\ \vdots \\ \mathbf{H}_T\mathbf{M}_{T\leftarrow 0}\delta\mathbf{x} \end{pmatrix}, \\
\mathbf{x} &= \mathbf{x}_\mathrm{B} + \delta\mathbf{x}.
\end{aligned}
$$

Examples of forward models used in NWP

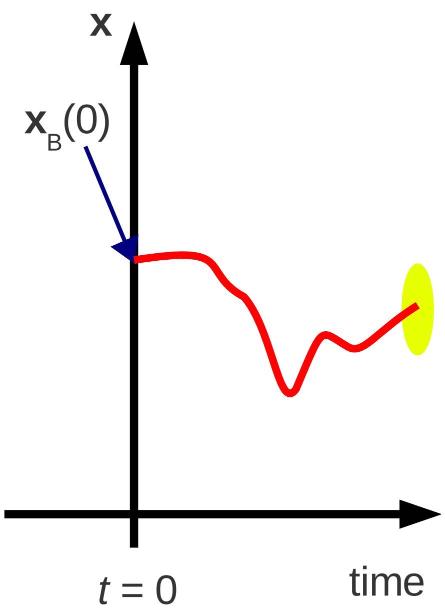
- Interpolation.



- Radiative transfer.

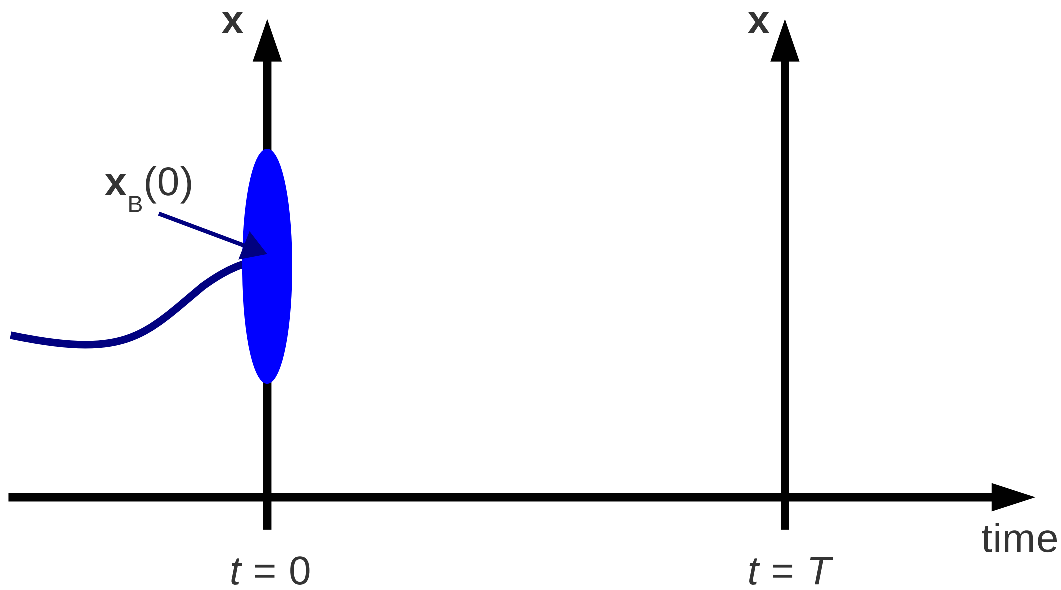


- Future observation.



Approximate solutions - Kalman Filter - a sequential DA method

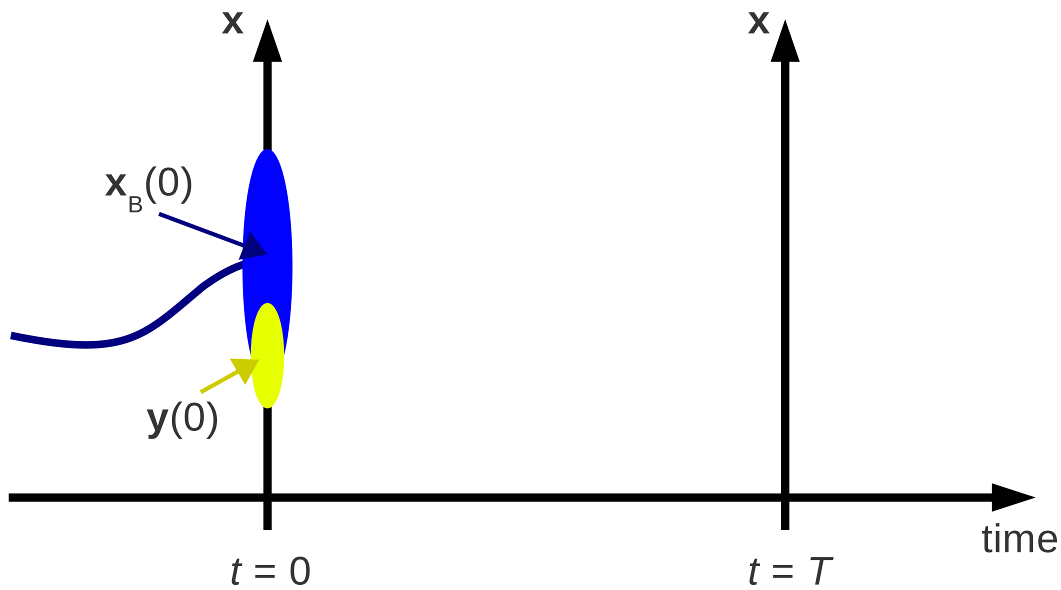
☑ n small (explicit $\mathbf{P}_\mathrm{f}$) ☑ p small ☑ linear $\tilde{\mathcal{H}}$ ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats



1. Background at $t = 0$: $\mathbf{x}_\mathrm{B}(0)$, $\mathbf{P}_\mathrm{f}(0) = \left\langle \boldsymbol{\epsilon}_\mathrm{B} \boldsymbol{\epsilon}_\mathrm{B}^\mathrm{T} \right\rangle$.

Approximate solutions - Kalman Filter - a sequential DA method

☑ n small (explicit $\mathbf{P}_{\mathrm{f}}$) ☑ p small ☑ linear $\tilde{\mathcal{H}}$ ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats

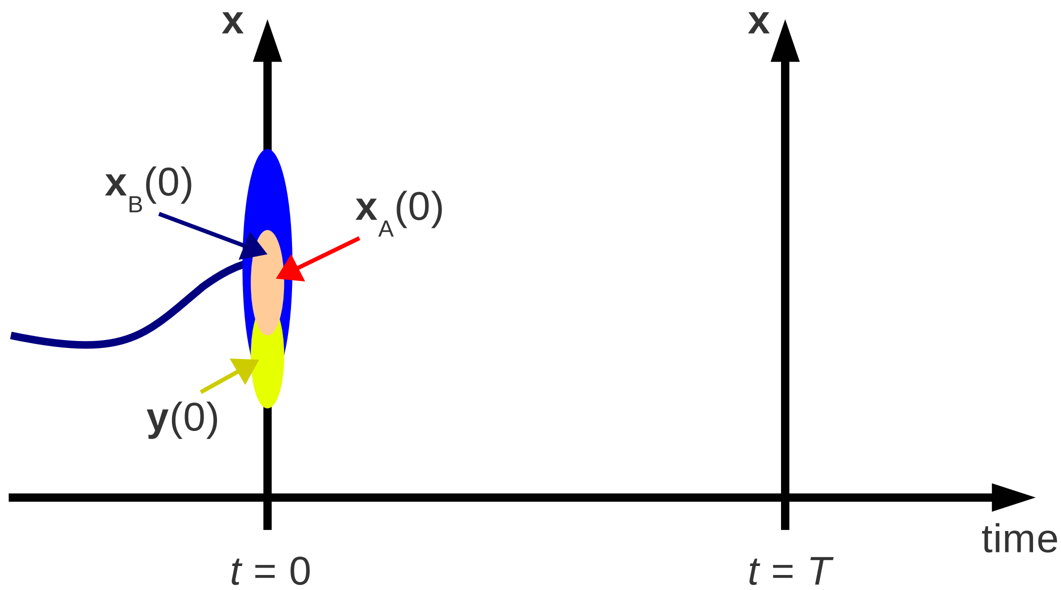


1. Background at $t = 0$: $\mathbf{x}_{\mathrm{B}}(0)$, $\mathbf{P}_{\mathrm{f}}(0) = \left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle$.

2. Introduce observations at $t = 0$: $\mathbf{y}_{\mathrm{o}}(0)$, $\mathbf{R}(0) = \left\langle \boldsymbol{\epsilon}_{y} \boldsymbol{\epsilon}_{y}^{\mathrm{T}} \right\rangle$.

Approximate solutions - Kalman Filter - a sequential DA method

☑ n small (explicit $\mathbf{P}_\mathrm{f}$) ☑ p small ☑ linear $\tilde{\mathcal{H}}$ ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats



1. Background at $t = 0$: $\mathbf{x}_\mathrm{B}(0)$, $\mathbf{P}_\mathrm{f}(0) = \left\langle \boldsymbol{\epsilon}_\mathrm{B} \boldsymbol{\epsilon}_\mathrm{B}^\mathrm{T} \right\rangle$.

2. Introduce observations at $t = 0$: $\mathbf{y}_\mathrm{o}(0)$, $\mathbf{R}(0) = \left\langle \boldsymbol{\epsilon}_y \boldsymbol{\epsilon}_y^\mathrm{T} \right\rangle$.

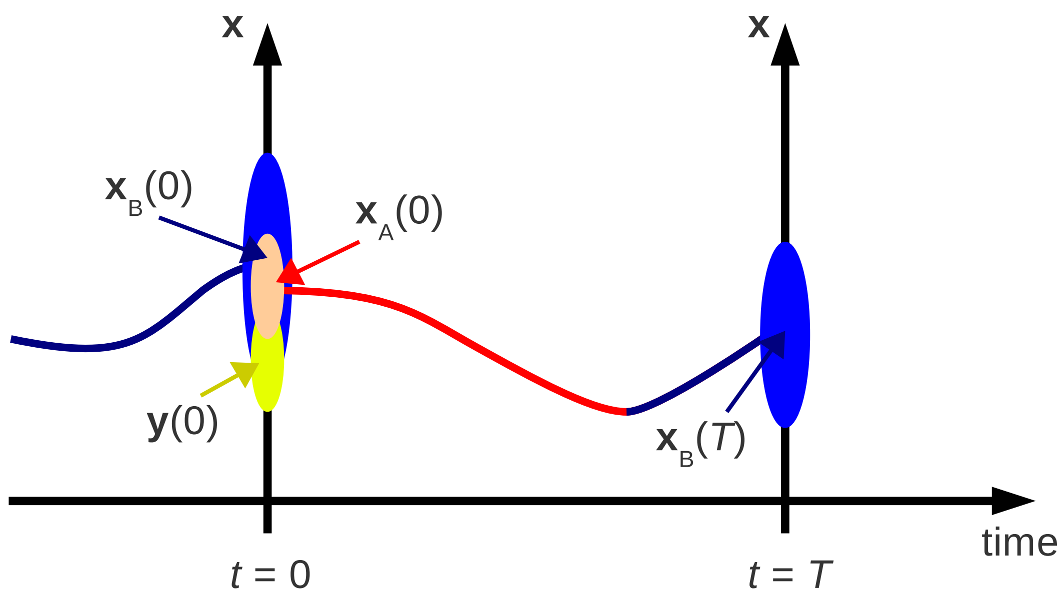
3. Analysis at $t = 0$:

   (a) $\mathbf{x}_\mathrm{A}(0) = \mathbf{x}_\mathrm{B} + \mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T})^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{B})$,

   (b) $\mathbf{P}_\mathrm{A}(0) = \left[\mathbf{I} - \mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T})^{-1}\mathbf{H}\right]\mathbf{P}_\mathrm{f} = \left\langle \boldsymbol{\epsilon}_\mathrm{A} \boldsymbol{\epsilon}_\mathrm{A}^\mathrm{T} \right\rangle$.

Approximate solutions - Kalman Filter - a sequential DA method

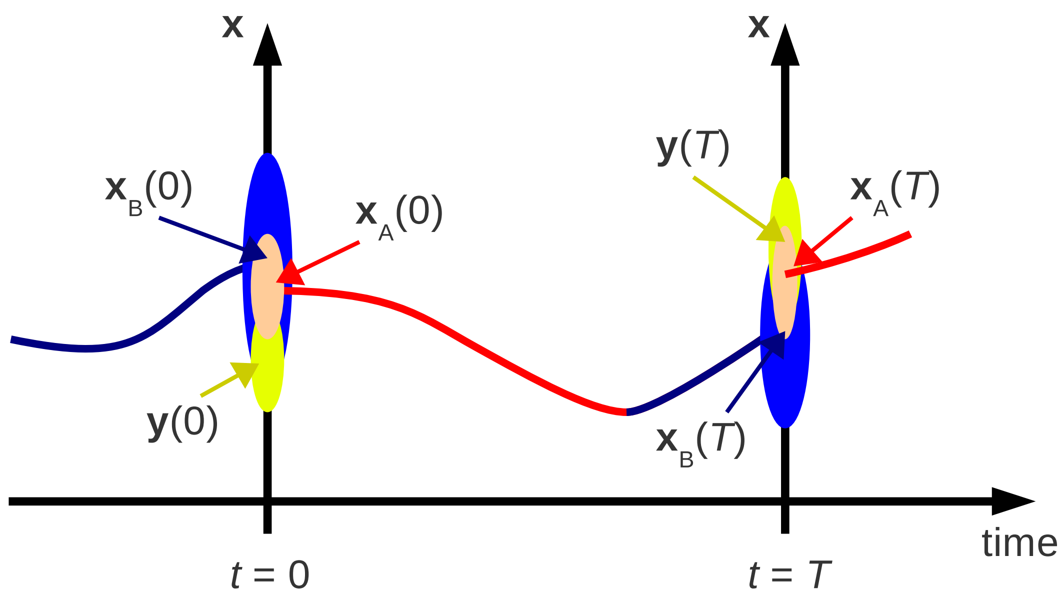
☑ n small (explicit $\mathbf{P}_\mathrm{f}$) ☑ p small ☑ linear $\tilde{\mathcal{H}}$ ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats



1. Background at $t = 0$: $\mathbf{x}_\mathrm{B}(0)$, $\mathbf{P}_\mathrm{f}(0) = \left\langle \boldsymbol{\epsilon}_\mathrm{B}\boldsymbol{\epsilon}_\mathrm{B}^\mathrm{T} \right\rangle$.
2. Introduce observations at $t = 0$: $\mathbf{y}_\mathrm{o}(0)$, $\mathbf{R}(0) = \left\langle \boldsymbol{\epsilon}_y\boldsymbol{\epsilon}_y^\mathrm{T} \right\rangle$.
3. Analysis at $t = 0$:
   (a) $\mathbf{x}_\mathrm{A}(0) = \mathbf{x}_\mathrm{B} + \mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T})^{-1}(\mathbf{y}_\mathrm{o} - \mathbf{H}\mathbf{x}_\mathrm{B})$,
   (b) $\mathbf{P}_\mathrm{A}(0) = \left[\mathbf{I} - \mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T})^{-1}\mathbf{H}\right]\mathbf{P}_\mathrm{f} = \left\langle \boldsymbol{\epsilon}_\mathrm{A}\boldsymbol{\epsilon}_\mathrm{A}^\mathrm{T} \right\rangle$.
4. Forecast to $t = T$:
   (a) $\mathbf{x}_\mathrm{B}(T) = \mathbf{M}\mathbf{x}_\mathrm{A}(0)$,
   (b) $\mathbf{P}_\mathrm{f}(T) = \mathbf{M}\mathbf{P}_\mathrm{A}(0)\mathbf{M}^\mathrm{T} + \mathbf{Q} = \left\langle \boldsymbol{\epsilon}_\mathrm{B}\boldsymbol{\epsilon}_\mathrm{B}^\mathrm{T} \right\rangle, \quad \mathbf{Q} = \left\langle \boldsymbol{\epsilon}_\mathrm{M}\boldsymbol{\epsilon}_\mathrm{M}^\mathrm{T} \right\rangle$.

Approximate solutions - Kalman Filter - a sequential DA method

☑ n small (explicit $\mathbf{P}_{\mathrm{f}}$) ☑ p small ☑ linear $\tilde{\mathcal{H}}$ ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats

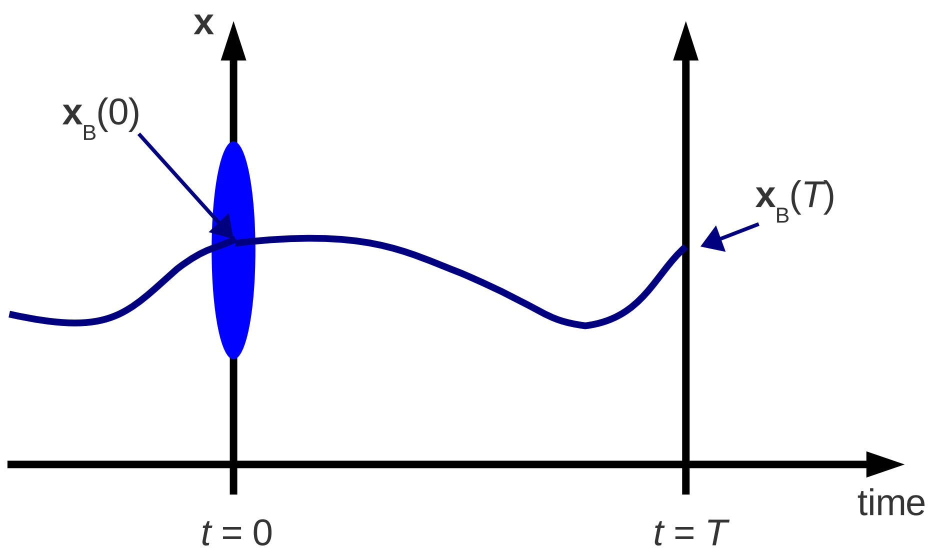


1. Background at $t = 0$: $\mathbf{x}_{\mathrm{B}}(0)$, $\mathbf{P}_{\mathrm{f}}(0) = \left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle$.
2. Introduce observations at $t = 0$: $\mathbf{y}_{\mathrm{o}}(0)$, $\mathbf{R}(0) = \left\langle \boldsymbol{\epsilon}_y \boldsymbol{\epsilon}_y^{\mathrm{T}} \right\rangle$.
3. Analysis at $t = 0$:
   (a) $\mathbf{x}_{\mathrm{A}}(0) = \mathbf{x}_{\mathrm{B}} + \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}})^{-1}(\mathbf{y}_{\mathrm{o}} - \mathbf{H}\mathbf{x}_{\mathrm{B}})$,
   (b) $\mathbf{P}_{\mathrm{A}}(0) = \left[\mathbf{I} - \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{H}\right]\mathbf{P}_{\mathrm{f}} = \left\langle \boldsymbol{\epsilon}_{\mathrm{A}} \boldsymbol{\epsilon}_{\mathrm{A}}^{\mathrm{T}} \right\rangle$.
4. Forecast to $t = T$:
   (a) $\mathbf{x}_{\mathrm{B}}(T) = \mathbf{M}\mathbf{x}_{\mathrm{A}}(0)$,
   (b) $\mathbf{P}_{\mathrm{f}}(T) = \mathbf{M}\mathbf{P}_{\mathrm{A}}(0)\mathbf{M}^{\mathrm{T}} + \mathbf{Q} = \left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle, \quad \mathbf{Q} = \left\langle \boldsymbol{\epsilon}_{\mathrm{M}} \boldsymbol{\epsilon}_{\mathrm{M}}^{\mathrm{T}} \right\rangle$.

Approximate solutions - variational DA - a smoother

☑ n large (implicit/static $\mathbf{P}_\mathrm{f}$) ☑ p large ☑ (non-linear $\tilde{\mathcal{H}}$) ☑ Gaussian stats ☑ (unbiased data) ☑ known model/obs err stats

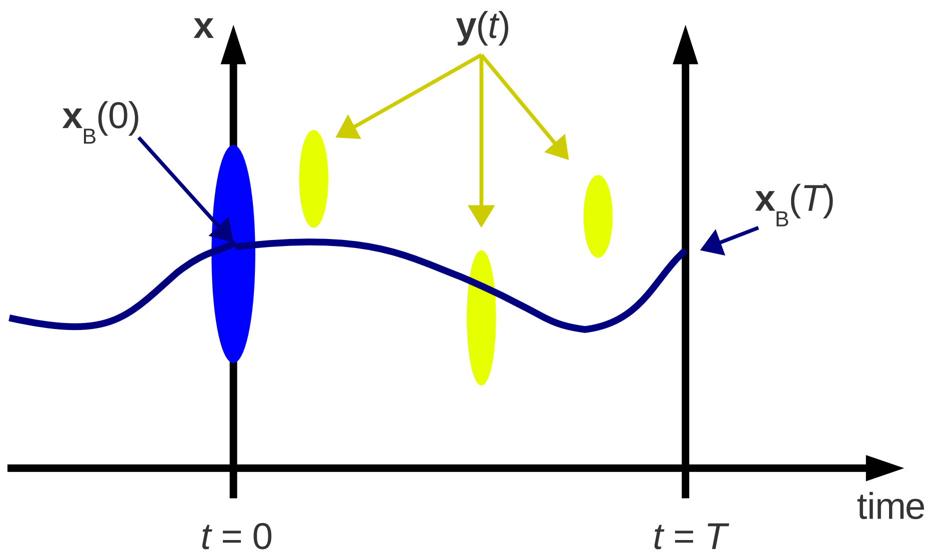
E.g. Incremental, strong constraint 4D-Var $\mathbf{P}_\mathrm{f} \rightarrow \mathbf{B}$



Approximate solutions - variational DA - a smoother

☑ n large (implicit/static $\mathbf{P}_\mathrm{f}$) ☑ p large ☑ (non-linear $\tilde{\mathcal{H}}$) ☑ Gaussian stats ☑ (unbiased data) ☑ known model/obs err stats

E.g. Incremental, strong constraint 4D-Var $\mathbf{P}_\mathrm{f} \rightarrow \mathbf{B}$

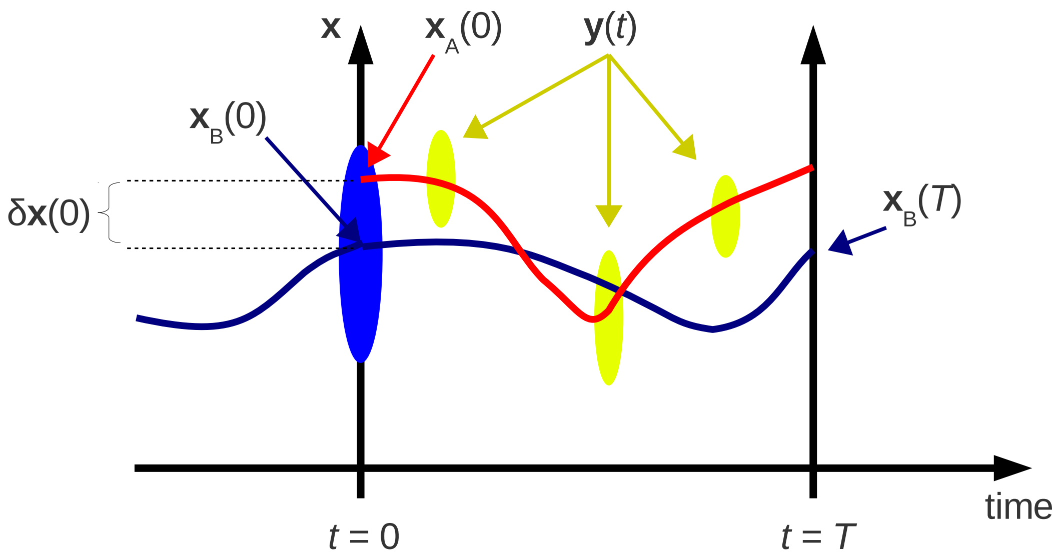


$$
\begin{aligned}
\mathcal{J}[\delta\mathbf{x}(0)] &= \frac{1}{2}\left[\delta\mathbf{x}(0)\right]^\mathrm{T}\mathbf{B}^{-1}\left[\delta\mathbf{x}(0)\right] + \frac{1}{2}\sum_{t=0}^{T}\left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right]^\mathrm{T}\mathbf{R}_t^{-1} \times \\
&\qquad \left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right], \\
\nabla_{\delta x}\mathcal{J} &= \mathbf{B}^{-1}\delta\mathbf{x}(0) - \sum_{t=0}^{T}\mathbf{M}_{t\leftarrow 0}^\mathrm{T}\mathbf{H}_t^\mathrm{T}\mathbf{R}_t^{-1}\left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right], \\
&\qquad\qquad \uparrow \quad \uparrow \qquad\qquad\qquad\qquad\qquad\qquad \uparrow \quad \uparrow \\
&\qquad\quad \text{adjoint operators} \qquad\qquad\qquad\qquad \text{tangent linear operators} \\
\mathbf{x}(0) &= \mathbf{x}_\mathrm{B}(0) + \delta\mathbf{x}(0).
\end{aligned}
$$

Approximate solutions - variational DA - a smoother

☑ n large (implicit/static $\mathbf{P}_\mathrm{f}$) ☑ p large ☑ (non-linear $\tilde{\mathcal{H}}$) ☑ Gaussian stats ☑ (unbiased data) ☑ known model/obs err stats

E.g. Incremental, strong constraint 4D-Var $\mathbf{P}_\mathrm{f} \rightarrow \mathbf{B}$



$$
\begin{aligned}
\mathcal{J}[\delta\mathbf{x}(0)] &= \frac{1}{2}\left[\delta\mathbf{x}(0)\right]^\mathrm{T}\mathbf{B}^{-1}\left[\delta\mathbf{x}(0)\right] + \frac{1}{2}\sum_{t=0}^{T}\left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right]^\mathrm{T}\mathbf{R}_t^{-1} \times \\
&\qquad \left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right], \\
\nabla_{\delta x}\mathcal{J} &= \mathbf{B}^{-1}\delta\mathbf{x}(0) - \sum_{t=0}^{T}\mathbf{M}_{t\leftarrow 0}^\mathrm{T}\mathbf{H}_t^\mathrm{T}\mathbf{R}_t^{-1}\left[\mathbf{y}_\mathrm{o}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_\mathrm{B})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right], \\
&\qquad\qquad \uparrow \quad \uparrow \qquad\qquad\qquad\qquad\qquad\qquad \uparrow \quad \uparrow \\
&\qquad \text{adjoint operators} \qquad\qquad\qquad\qquad \text{tangent linear operators} \\
\mathbf{x}(0) &= \mathbf{x}_\mathrm{B}(0) + \delta\mathbf{x}(0).
\end{aligned}
$$

Making variational DA work

- Key to success of 4D-Var in NWP is the $\mathbf{B}$-matrix.
- This is modelled, e.g., via (linear) change of variables - a control variable transform:
  - $\delta\mathbf{x}(0) = \mathbf{U}\delta\mathbf{v}$.
  - Background errors in the $\delta\mathbf{v}$-representation are assumed to be mutually uncorrelated:

$$
\begin{aligned}
\left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle &\approx \mathbf{B}. \\
\left\langle \left[\mathbf{U}^{-1}\boldsymbol{\epsilon}_{\mathrm{B}}\right] \left[\mathbf{U}^{-1}\boldsymbol{\epsilon}_{\mathrm{B}}\right]^{\mathrm{T}} \right\rangle &\approx \mathbf{I}., \\
\mathbf{U}\mathbf{U}^{\mathrm{T}} &\approx \mathbf{B}.
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{J}[\delta\mathbf{x}(0)] &= \frac{1}{2}\delta\mathbf{v}^{\mathrm{T}}\delta\mathbf{v} + \frac{1}{2}\sum_{t=0}^{T}\left[\mathbf{y}_{\mathrm{o}}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_{\mathrm{B}})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\mathbf{v}\right]^{\mathrm{T}}\mathbf{R}_t^{-1} \times \\
&\qquad \left[\mathbf{y}_{\mathrm{o}}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_{\mathrm{B}})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\mathbf{v}\right], \\
\nabla_{\delta v}\mathcal{J} &= \delta\mathbf{v} - \mathbf{U}^{\mathrm{T}}\sum_{t=0}^{T}\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\left[\mathbf{y}_{\mathrm{o}}(t) - \mathcal{H}_t(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_{\mathrm{B}})) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}(0)\right].
\end{aligned}
$$

Approximate solutions - Ensemble Kalman Filter (EnKF)

☑ n large (sample $\mathbf{P}_{\mathrm{f}}$) ☑ p large ☑ (non-linear $\tilde{\mathcal{H}}$) ☑ Gaussian stats ☑ unbiased data ☑ known model/obs err stats

- The KF update equation:

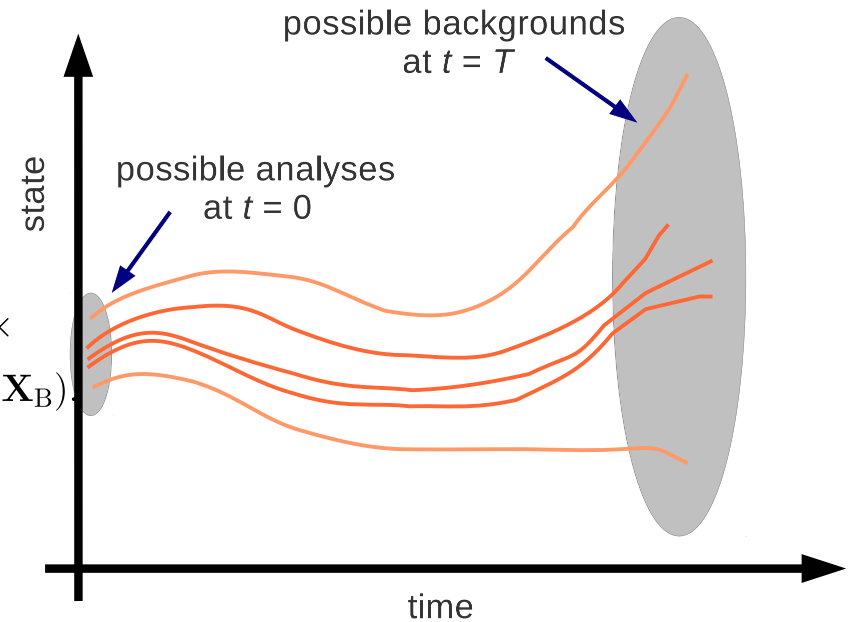
$$\mathbf{x}_{\mathrm{A}} = \mathbf{x}_{\mathrm{B}} + \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}})^{-1}(\mathbf{y}_{\mathrm{o}} - \mathbf{H}\mathbf{x}_{\mathrm{B}}).$$

- Introduce an ensemble of N possible backgrounds (the ensemble), and N perturbed sets of observations:

$$\mathbf{X}_{\mathrm{B}} = \left(\mathbf{x}_{\mathrm{B}}^{(1)} \ \cdots \ \mathbf{x}_{\mathrm{B}}^{(N)} \right), \qquad \mathbf{X}'_{\mathrm{B}} = \left(\underbrace{\mathbf{x}_{\mathrm{B}}^{(1)} - \langle \mathbf{x} \rangle}_{\mathbf{x}_{\mathrm{f}}^{\prime(1)}} \ \cdots \ \underbrace{\mathbf{x}_{\mathrm{B}}^{(N)} - \langle \mathbf{x} \rangle}_{\mathbf{x}_{\mathrm{f}}^{\prime(N)}} \right), \qquad \mathbf{Y}_{\mathrm{o}} = \left(\mathbf{y}_{\mathrm{o}}^{(1)} \ \cdots \ \mathbf{y}_{\mathrm{o}}^{(N)} \right).$$

- The ensemble can be used to approximate $\mathbf{P}_{\mathrm{f}}$:

$$\mathbf{P}_{\mathrm{f}} \approx \mathbf{P}_{\mathrm{f}}^{(N)} = \frac{1}{N-1} \sum_{l=1}^{N} \mathbf{x}_{\mathrm{f}}^{\prime(l)} \mathbf{x}_{\mathrm{f}}^{\prime(l)\mathrm{T}} = \frac{1}{N-1} \mathbf{X}'_{\mathrm{B}} \mathbf{X}'^{\mathrm{T}}_{\mathrm{B}}.$$



- The basic EnKF - ens. members are columns:

$$\mathbf{X}_{\mathrm{A}} = \mathbf{X}_{\mathrm{B}} + \mathbf{X}'_{\mathrm{B}} \left(\mathbf{H}\mathbf{X}'_{\mathrm{B}} \right)^{\mathrm{T}} \left((N-1)\mathbf{R} + \left(\mathbf{H}\mathbf{X}'_{\mathrm{B}} \right) \left(\mathbf{H}\mathbf{X}'_{\mathrm{B}} \right)^{\mathrm{T}} \right)^{-1} \times (\mathbf{Y}_{\mathrm{o}} - \mathbf{H}\mathbf{X}_{\mathrm{B}}).$$

- Cycling:

$$\mathbf{X}_{\mathrm{B}}(T) = \mathcal{M}\left(\mathbf{X}_{\mathrm{A}}(0) \right) + \underline{\boldsymbol{\eta}}.$$

The basic EnKF suffers with sampling error for $N \ll n$.

1. The forecast error covariance matrix is rank deficient

The rank of $\mathbf{P}_\mathrm{f}^{(N)}$ is an indication of the size of the state space spanned by the forecast error ensemble.

$$\mathrm{rank}\left(\mathbf{P}_\mathrm{f}^{(N)}\right) \leq N-1.$$

The analysis increments are restricted to be a linear combination of the forecast error ensemble perturbations in an $N-1$-dimensional space.

2. The forecast ensemble spread will be subject to sampling error

- If the spread is too large then the analysis ens. will over-fit the obs. - too little trust in the fc. ens.
- If the spread is too small then the analysis ens. will under-fit the obs. - too much trust in the fc. ens.
  - Once in this regime, it is difficult to escape as the ens. will (effectively) ignore the obs..
  - This is called "filter divergence" (because we diverge from reality).

Filter divergence means that each ensemble member will (effectively) be free running.

3. The correlations will be subject to sampling error

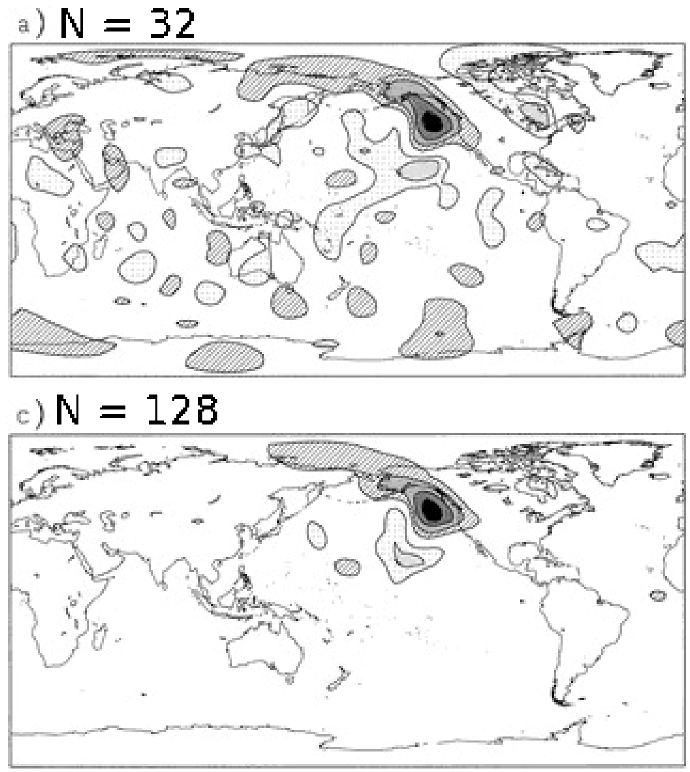
- The error in the sample correlation between errors at locations i and j has expectation:

$$[\mathcal{E}(\delta\mathbf{C}_\mathrm{f}^{(N)})]_{ij} \sim \frac{1}{\sqrt{N}}\left(1-([\mathbf{C}_\mathrm{f}]_{ij})^2\right), \qquad [\mathbf{C}_\mathrm{f}^{(N)}]_{ij} = \frac{[\mathbf{P}_\mathrm{f}^{(N)}]_{ij}}{\sigma_i\sigma_j}$$

  (errors are expected to be large when N small and/or correlations are close to zero).

- Pairs of distant points would be expected to have correlations close to zero.

Sampling error means that we can't trust distant correlations. Left untreated this noise will destroy the benefits of DA (analysis increments will be influenced by distant observations).

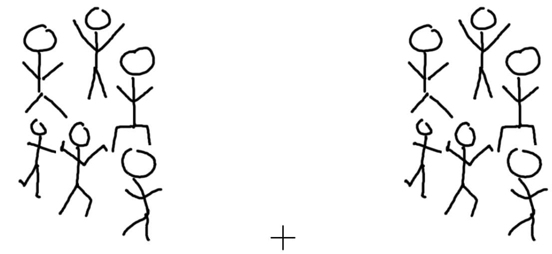


From Houtekamer & Mitchell, 1998

Making progress

What can be done to reduce/mitigate this problem $N \ll n$?

- **Use more ensemble members**.



  - This is expensive.
  - How many is 'enough'?

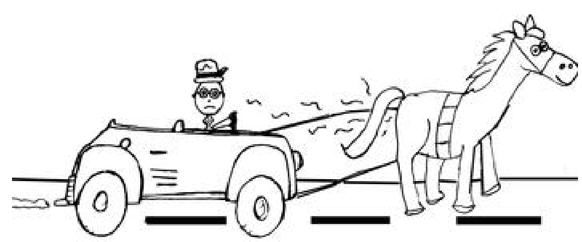
- **Ensemble inflation**.

  - Artificially increase the size of each $\mathbf{x}_{\mathrm{f}}^{\prime(i)}$.
  - How do we know what the ensemble spread should be?

- **Localization**.

  - Eliminate far-field correlations.
  - How should this be done?
  - Does this have any other consequences?

- **Combine ensemble with variational approaches**.

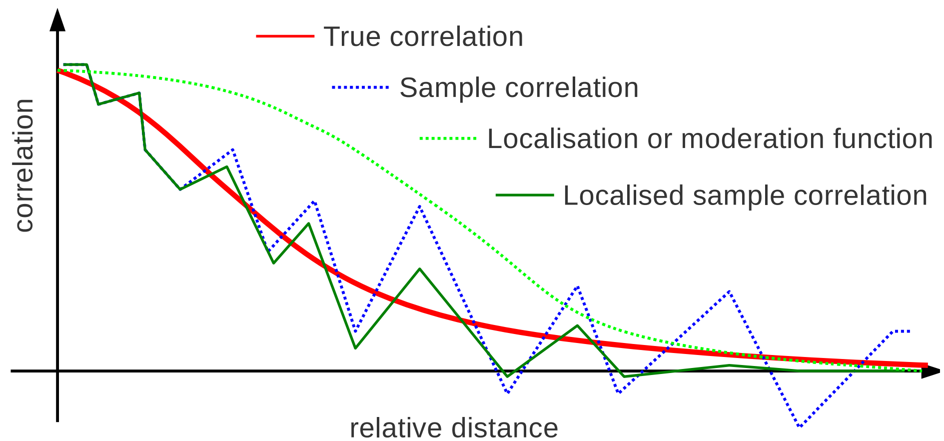


  - Adopt a hybrid method.
  - How to do this?

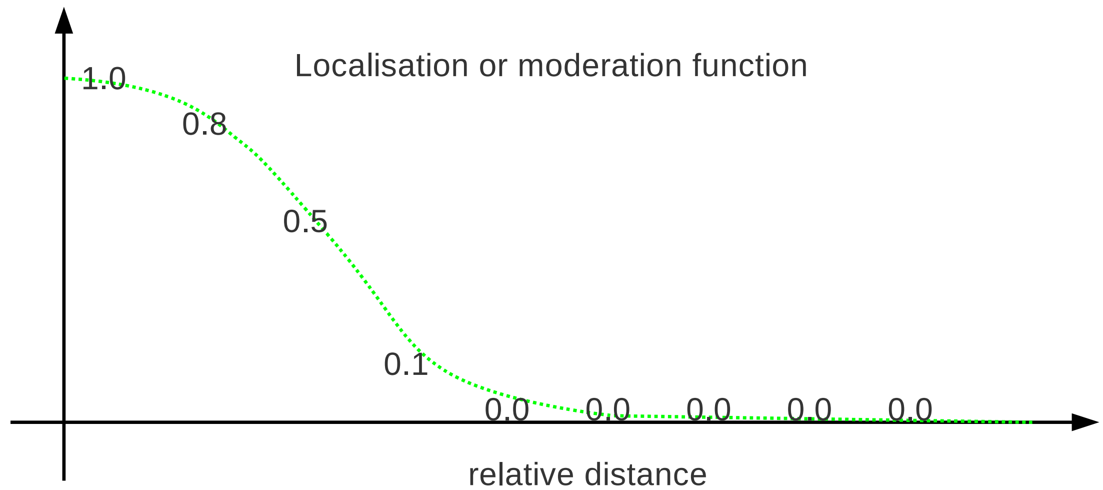
Localization

Many ways of doing localization:

- $\mathbf{R}$-localization.
- $\mathbf{P}_{\mathrm{f}}$-localization:
  - Modify $\mathbf{P}_{\mathrm{f}}^{(N)}$ with a localization/moderation function that decreases with separation.
  - What length-scale? How to do multivariate aspects?
  - Has side effects (e.g. affects length-scales, **affects balance**).



$\mathbf{P}_\mathrm{f}$-localization (Schur/Hadamard product, univariate)



$$
\begin{aligned}
\mathbf{P}_\mathrm{f}^\mathrm{Loc} &= \mathbf{P}_\mathrm{f} \circ \mathbf{\Omega}, \\
&= \begin{pmatrix}
P_{\mathrm{f}11}^{(N)} & P_{\mathrm{f}12}^{(N)} & \cdots & \cdots & P_{\mathrm{f}15}^{(N)} & \cdots & \cdots & P_{\mathrm{f}18}^{(N)} & P_{\mathrm{f}19}^{(N)} \\
P_{\mathrm{f}21}^{(N)} & P_{\mathrm{f}22}^{(N)} & \cdots & \cdots & P_{\mathrm{f}25}^{(N)} & \cdots & \cdots & P_{\mathrm{f}28}^{(N)} & P_{\mathrm{f}29}^{(N)} \\
\vdots & \vdots & \ddots & \cdots & P_{\mathrm{f}35}^{(N)} & \cdots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & P_{\mathrm{f}45}^{(N)} & \vdots & \vdots & \vdots & \vdots \\
P_{\mathrm{f}51}^{(N)} & P_{\mathrm{f}52}^{(N)} & P_{\mathrm{f}53}^{(N)} & P_{\mathrm{f}54}^{(N)} & P_{\mathrm{f}55}^{(N)} & P_{\mathrm{f}56}^{(N)} & P_{\mathrm{f}57}^{(N)} & P_{\mathrm{f}58}^{(N)} & P_{\mathrm{f}59}^{(N)} \\
\vdots & \vdots & \vdots & \vdots & P_{\mathrm{f}65}^{(N)} & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \cdots & \cdots & P_{\mathrm{f}75}^{(N)} & \cdots & \ddots & \vdots & \vdots \\
P_{\mathrm{f}81}^{(N)} & P_{\mathrm{f}82}^{(N)} & \cdots & \cdots & P_{\mathrm{f}85}^{(N)} & \cdots & \cdots & P_{\mathrm{f}88}^{(N)} & P_{\mathrm{f}89}^{(N)} \\
P_{\mathrm{f}91}^{(N)} & P_{\mathrm{f}92}^{(N)} & \cdots & \cdots & P_{\mathrm{f}95}^{(N)} & \cdots & \cdots & P_{\mathrm{f}98}^{(N)} & P_{\mathrm{f}99}^{(N)}
\end{pmatrix}
\circ
\begin{pmatrix}
1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.5 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 \\
0.1 & 0.5 & 0.5 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 \\
0.0 & 0.1 & 0.1 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.5 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 \\
0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0 & 0.8 & 0.5 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0 & 0.8 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0
\end{pmatrix}, \\
P_{\mathrm{f}ij}^\mathrm{Loc} &= P_{\mathrm{f}ij}\Omega_{ij}.
\end{aligned}
$$

Can be extended to multivariate localization. But … we rarely have access to explicit $\mathbf{P}_\mathrm{f}$ or $\mathbf{\Omega}$ matrices $(n \times n)$.

Localization without explicit $\mathbf{P}_\mathrm{f}$ and $\boldsymbol{\Omega}$ matrices

$$\text{Sample forecast error cov. matrix (from } N \text{ dynamical ens. members)} : \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{l=1}^{N} \mathbf{x}_\mathrm{f}'^{(l)} \mathbf{x}_\mathrm{f}'^{(l)\mathrm{T}} = \frac{1}{N-1}\mathbf{X}_\mathrm{B}'\mathbf{X}_\mathrm{B}'^{\mathrm{T}},$$

$$\text{Sample localization matrix (from } K \text{ correlation ens. members)} : \boldsymbol{\Omega}^{(K)} = \frac{1}{K-1}\sum_{k=1}^{K} \boldsymbol{\omega}^{(k)} \boldsymbol{\omega}^{(k)\mathrm{T}} = \frac{1}{K-1}\mathbf{K}\mathbf{K}^{\mathrm{T}},$$

$$\text{One matrix element: } [\mathbf{P}_\mathrm{f}^{(N)}]_{ij} = \frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}'^{(l)} x_{\mathrm{f}j}'^{(l)},$$

$$\text{One matrix element: } [\boldsymbol{\Omega}^{(K)}]_{ij} = \frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}.$$

Localization without explicit $\mathbf{P}_\mathrm{f}$ and $\boldsymbol{\Omega}$ matrices

$$\text{Sample forecast error cov. matrix (from } N \text{ dynamical ens. members) : } \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{l=1}^{N} \mathbf{x}_\mathrm{f}'^{(l)} \mathbf{x}_\mathrm{f}'^{(l)\mathrm{T}} = \frac{1}{N-1}\mathbf{X}_\mathrm{B}'\mathbf{X}_\mathrm{B}'^{\mathrm{T}},$$

$$\text{Sample localization matrix (from } K \text{ correlation ens. members) : } \boldsymbol{\Omega}^{(K)} = \frac{1}{K-1}\sum_{k=1}^{K} \boldsymbol{\omega}^{(k)} \boldsymbol{\omega}^{(k)\mathrm{T}} = \frac{1}{K-1}\mathbf{K}\mathbf{K}^{\mathrm{T}},$$

$$\text{One matrix element: } [\mathbf{P}_\mathrm{f}^{(N)}]_{ij} = \frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}'^{(l)} x_{\mathrm{f}j}'^{(l)},$$

$$\text{One matrix element: } [\boldsymbol{\Omega}^{(K)}]_{ij} = \frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}.$$

$$\begin{aligned}
\text{Localized matrix (Schur product): } [\mathbf{P}_\mathrm{f}^{\mathrm{Loc}}]_{ij} &= [\mathbf{P}_\mathrm{f}^{(N)}]_{ij}[\boldsymbol{\Omega}^{(K)}]_{ij}, \\
&= \left[\frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}'^{(l)} x_{\mathrm{f}j}'^{(l)}\right]\left[\frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}\right], \\
&= \frac{1}{N-1}\frac{1}{K-1}\sum_{l=1}^{N}\sum_{k=1}^{K} \underbrace{x_{\mathrm{f}i}'^{(l)} \omega_i^{(k)}}_{\substack{\text{element } i \\ \text{of } \tilde{\mathbf{x}}'^{(m)}}} \quad \underbrace{x_{\mathrm{f}j}'^{(l)} \omega_j^{(k)}}_{\substack{\text{element } j \\ \text{of } \tilde{\mathbf{x}}'^{(m)}}},
\end{aligned}$$

Localization without explicit $\mathbf{P}_\mathrm{f}$ and $\boldsymbol{\Omega}$ matrices

$$\text{Sample forecast error cov. matrix (from } N \text{ dynamical ens. members)} : \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{l=1}^{N} \mathbf{x}_\mathrm{f}'^{(l)} \mathbf{x}_\mathrm{f}'^{(l)\mathrm{T}} = \frac{1}{N-1}\mathbf{X}_\mathrm{B}' \mathbf{X}_\mathrm{B}'^\mathrm{T},$$

$$\text{Sample localization matrix (from } K \text{ correlation ens. members)} : \boldsymbol{\Omega}^{(K)} = \frac{1}{K-1}\sum_{k=1}^{K} \boldsymbol{\omega}^{(k)} \boldsymbol{\omega}^{(k)\mathrm{T}} = \frac{1}{K-1}\mathbf{K}\mathbf{K}^\mathrm{T},$$

$$\text{One matrix element: } [\mathbf{P}_\mathrm{f}^{(N)}]_{ij} = \frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}'^{(l)} x_{\mathrm{f}j}'^{(l)},$$

$$\text{One matrix element: } [\boldsymbol{\Omega}^{(K)}]_{ij} = \frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}.$$

$$\begin{aligned}
\text{Localized matrix (Schur product): } [\mathbf{P}_\mathrm{f}^{\mathrm{Loc}}]_{ij} &= [\mathbf{P}_\mathrm{f}^{(N)}]_{ij} [\boldsymbol{\Omega}^{(K)}]_{ij}, \\
&= \left[\frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}'^{(l)} x_{\mathrm{f}j}'^{(l)}\right] \left[\frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}\right], \\
&= \frac{1}{N-1}\frac{1}{K-1}\sum_{l=1}^{N}\sum_{k=1}^{K} \underbrace{x_{\mathrm{f}i}'^{(l)} \omega_i^{(k)}}_{\substack{\text{element } i \\ \text{of } \tilde{\mathbf{x}}'^{(m)}}} \quad \underbrace{x_{\mathrm{f}j}'^{(l)} \omega_j^{(k)}}_{\substack{\text{element } j \\ \text{of } \tilde{\mathbf{x}}'^{(m)}}},
\end{aligned}$$

$$\text{Effective ens. } (M = NK \text{ localized ens. members}) = \frac{1}{N-1}\frac{1}{K-1}\sum_{m=1}^{M} \tilde{x}_i'^{(m)} \tilde{x}_j'^{(m)},$$

$$\tilde{\mathbf{x}}'^{(m)} = \mathbf{x}_\mathrm{f}'^{(l)} \circ \boldsymbol{\omega}^{(k)} = \begin{pmatrix} x_{\mathrm{f}1}'^{(l)} \\ \vdots \\ x_{\mathrm{f}n}'^{(l)} \end{pmatrix} \circ \begin{pmatrix} \omega_1^{(k)} \\ \vdots \\ \omega_1^{(k)} \end{pmatrix} = \begin{pmatrix} x_{\mathrm{f}1}'^{(l)} \omega_1^{(k)} \\ \vdots \\ x_{\mathrm{f}n}'^{(l)} \omega_1^{(n)} \end{pmatrix}, \qquad \underbrace{\tilde{\mathbf{X}}}_{n\times M} = \underbrace{\mathbf{X}_\mathrm{B}'}_{n\times N} \triangle \underbrace{\mathbf{K}}_{n\times K},$$

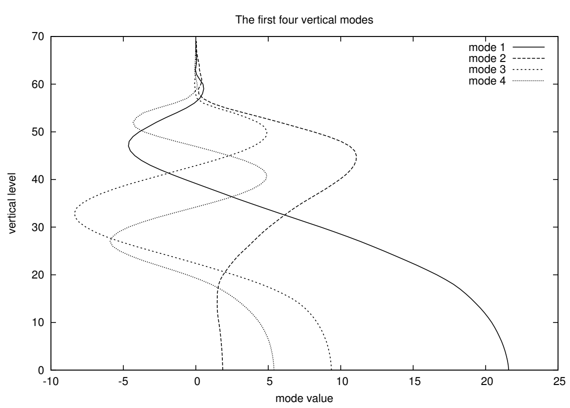
Spectral localization scheme (static localization)

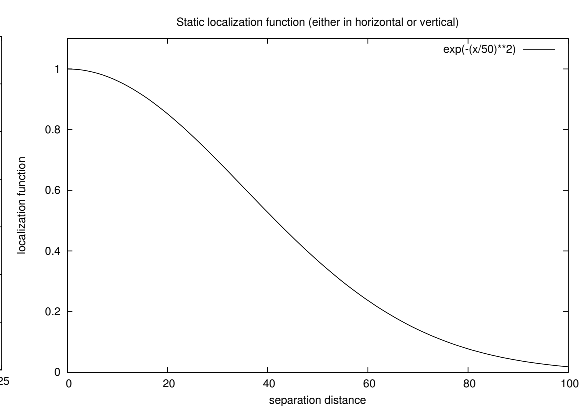
Propose a static and homogeneous model:

$$
\begin{aligned}
\mathbf{K} \dashrightarrow \mathbf{K}_{\text{spec}} &= \overline{\begin{pmatrix} \mathbf{F}_u\boldsymbol{\Lambda}_u^{1/2} \\ \mathbf{F}_v\boldsymbol{\Lambda}_v^{1/2} \\ \mathbf{F}_\theta\boldsymbol{\Lambda}_\theta^{1/2} \\ \mathbf{F}_\Pi\boldsymbol{\Lambda}_\Pi^{1/2} \end{pmatrix}} \in \mathbb{R}^{n\times K}, \\
[\mathbf{F}_p]_{\mathbf{rk}} &= \cos(k_x r_x + \delta_s^x)\cos(k_y r_y + \delta_s^y)\nu(r_z, k_z), \\
[\boldsymbol{\Lambda}_p]_{\mathbf{kk}'} &= \begin{cases} \lambda_p\left(\sqrt{k_x^2+k_y^2}, k_z\right) & \mathbf{k}=\mathbf{k}' \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

If $\mathbf{F}_p$ and $\boldsymbol{\Lambda}_p$ are independent of p:

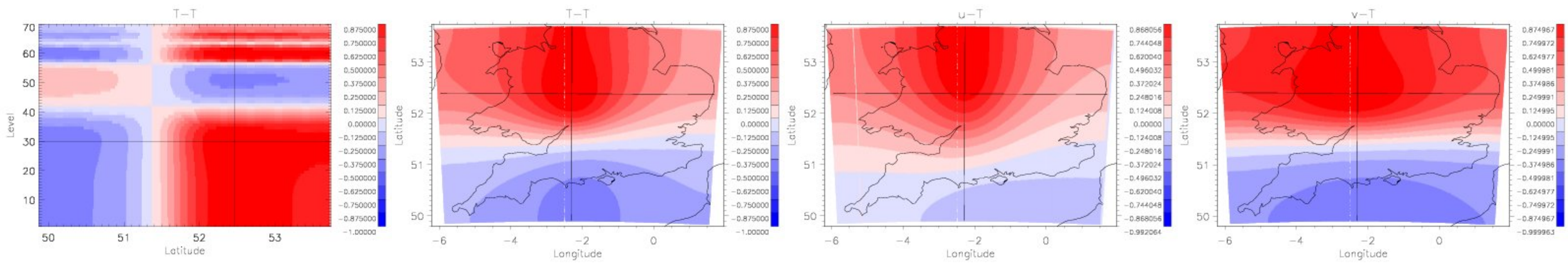
$$
\mathbf{K}_{\text{spec}}\mathbf{K}_{\text{spec}}^{\mathrm{T}} \sim \begin{pmatrix} \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} \\ \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} \\ \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} \\ \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} & \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^{\mathrm{T}} \end{pmatrix}.
$$



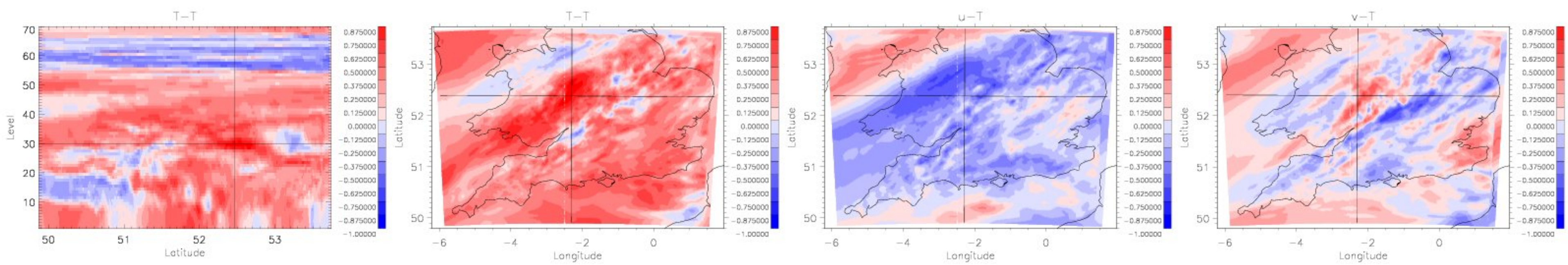


Localization results - spectral

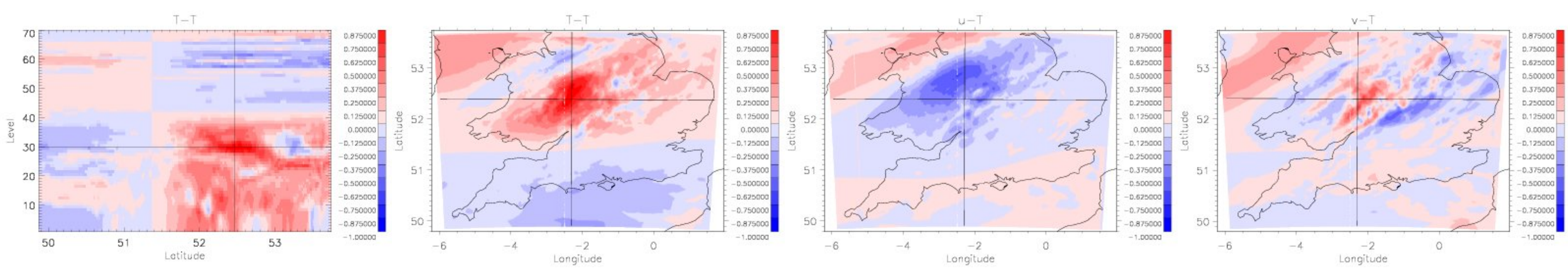
Localization functions, $\mathbf{K}_{\mathrm{spec}}\mathbf{K}_{\mathrm{spec}}^{\mathrm{T}}$**, 58 horizontal wave-numbers, 5 vertical modes,** $K = 290$



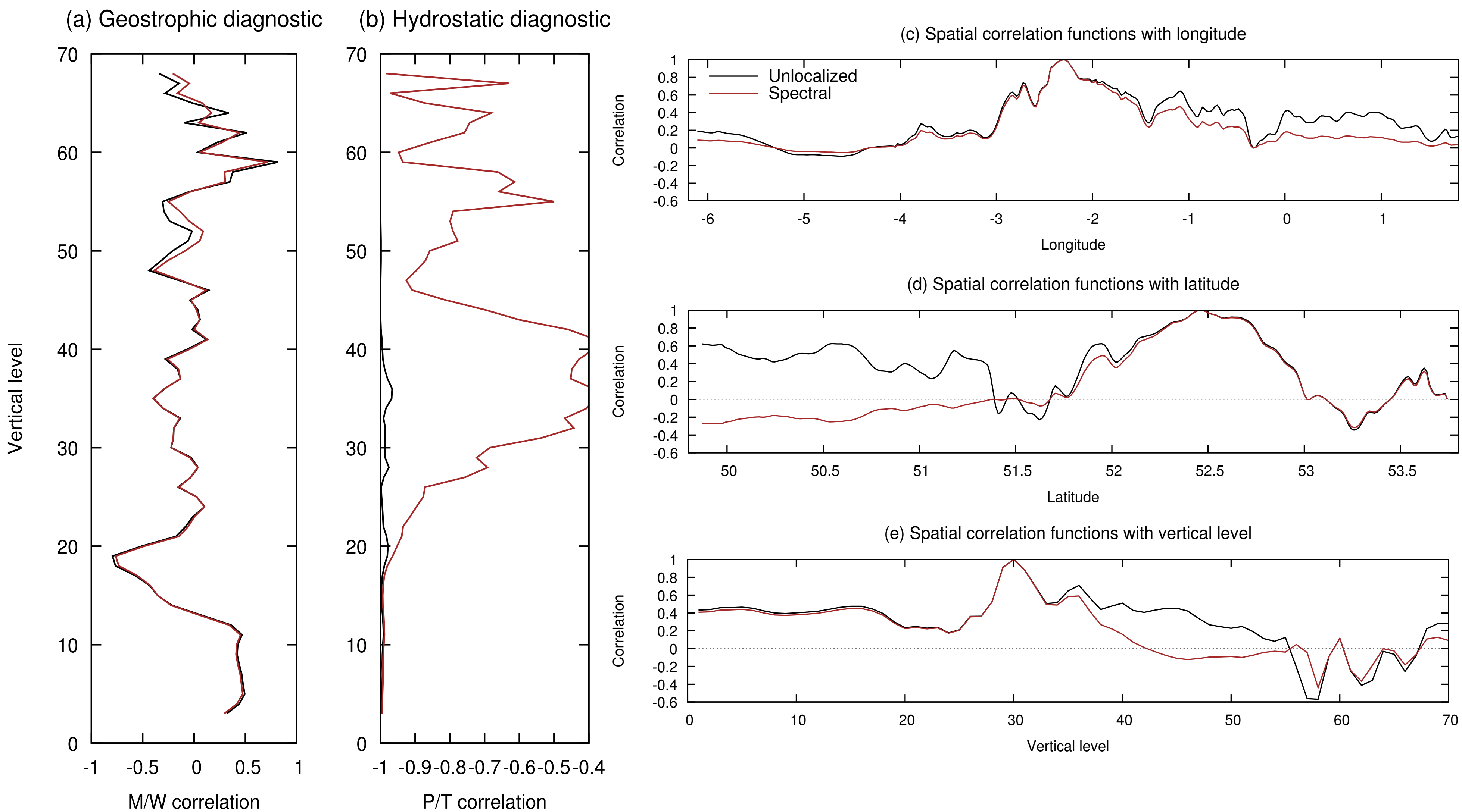
Dynamical correlations (raw), $\mathbf{X}_{\mathrm{B}}'\mathbf{X}_{\mathrm{B}}'^{\mathrm{T}}$



Localized functions, $(\mathbf{X}_{\mathrm{B}}'\mathbf{X}_{\mathrm{B}}'^{\mathrm{T}}) \circ (\mathbf{K}_{\mathrm{spec}}\mathbf{K}_{\mathrm{spec}}^{\mathrm{T}})$



Localization and balance results - spectral, $N = 24,\ K = 290$



SENCORP localization scheme - Smoothed ENsemble COrrelations Raised to a Power) - adaptive[1]

$$\mathbf{\Omega} \dashrightarrow \mathbf{\Omega}_{\text{SENCORP}} = \mathbf{C}^{\circ Q}$$

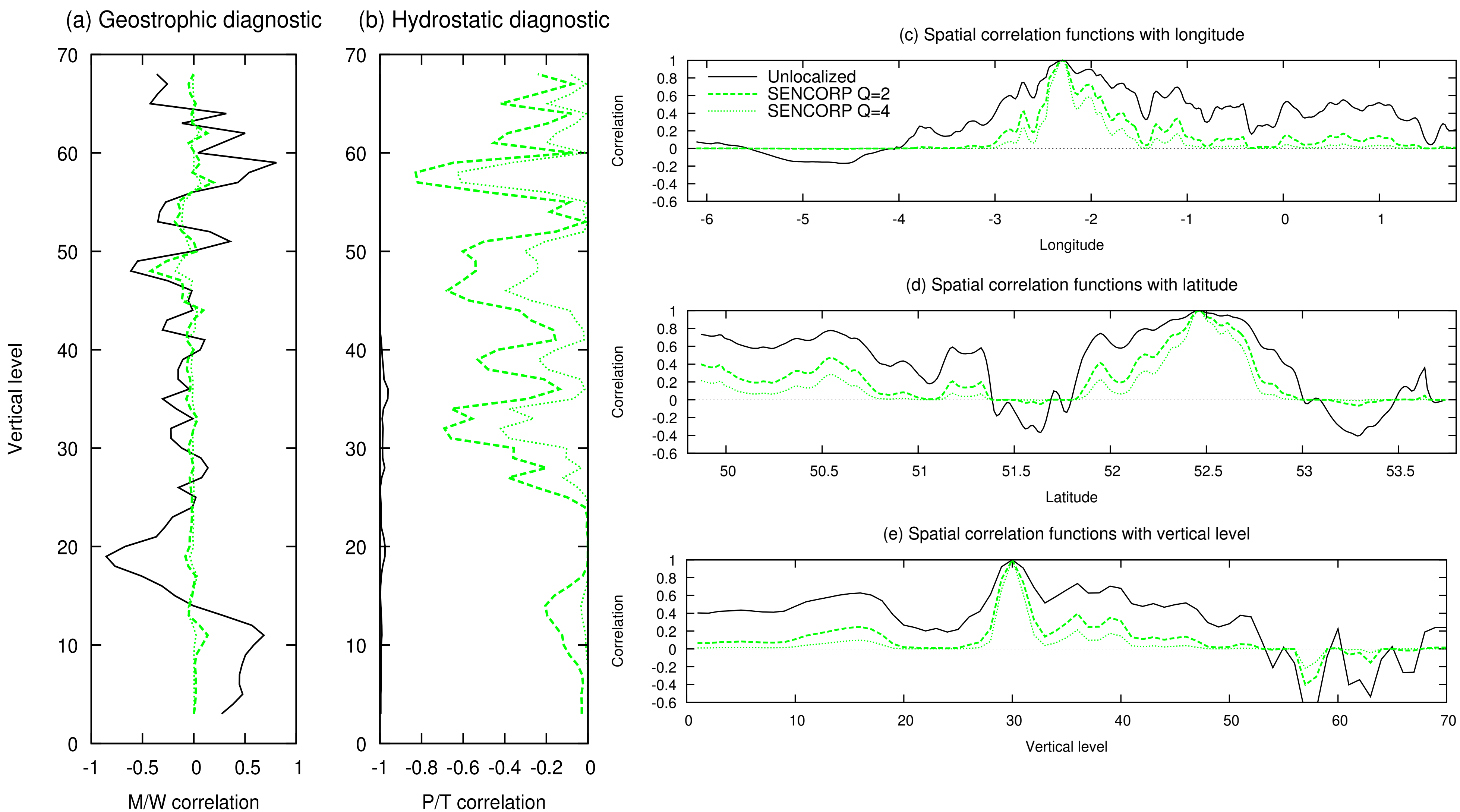
1. From the ensemble members, $\mathbf{X}'_{\text{B}}$, create smoothed members, $\mathbf{W}'_{\text{B}}$.
2. Normalize (sum of squares of each row of $\mathbf{W}'_{\text{B}}$ is 1, call this $\overline{\mathbf{W}}'_{\text{B}}$).
3. Calculate correlation matrix
$$\mathbf{C} = \frac{1}{N-1}\overline{\mathbf{W}}'_{\text{B}}\overline{\mathbf{W}}'^{\text{T}}_{\text{B}}.$$
4. $\mathbf{C}^{\circ Q}$ is the Schur power of $\mathbf{C}$ with itself Q times (Q is even).

$$\mathbf{\Omega}_{\text{SENCORP}} = \mathbf{K}_{\text{SENCORP}}\mathbf{K}^{\text{T}}_{\text{SENCORP}},$$

$$\mathbf{K} \dashrightarrow \mathbf{K}_{\text{SENCORP}} = \sqrt{\frac{K-1}{(N-1)^Q}}\overline{\mathbf{W}}'_{\text{B}} \vartriangle \overline{\mathbf{W}}'_{\text{B}} \vartriangle \overline{\mathbf{W}}'_{\text{B}} \vartriangle \cdots, \qquad \mathbf{K} \in \mathbb{R}^{n\times K}, \qquad K = N^Q.$$

[1] Bishop C.H. and Hodyss D., Flow adaptive moderation of spurious ensemble correlations and its used in ensemble-based data assimilation, Quart. J. Roy. Met. Soc. 133, 2029-2044 (2007), DOI:10.1002/qj.169.

Localization and balance results - SENCORP, $N = 16$, $(Q = 2,\ K = 256)$, $(Q = 4,\ K = 65536)$



ECO-RAP localization scheme - Ensemble COrrelations Raised to a Power) - adaptive[2]

$$\begin{aligned}
\mathbf{K} \dashrightarrow \mathbf{K}_{\mathrm{ECORAP}} &= \overline{\mathbf{C}^{\circ Q} \begin{pmatrix} \mathbf{F}_u \boldsymbol{\Lambda}_u^{1/2} \\ \mathbf{F}_v \boldsymbol{\Lambda}_v^{1/2} \\ \mathbf{F}_\theta \boldsymbol{\Lambda}_\theta^{1/2} \\ \mathbf{F}_\Pi \boldsymbol{\Lambda}_\Pi^{1/2} \end{pmatrix}} \in \mathbb{R}^{n \times K}, \\
[\mathbf{F}_p]_{\mathbf{rk}} &= \cos(k_x r_x + \delta_s^x) \cos(k_y r_y + \delta_s^y) \nu(r_z, k_z), \\
[\boldsymbol{\Lambda}_p]_{\mathbf{kk}'} &= \begin{cases} \lambda_p \left(\sqrt{k_x^2 + k_y^2}, k_z \right) & \mathbf{k} = \mathbf{k}' \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Lots of computations required

$$[\mathbf{K}^{\mathrm{ECORAP}}]_{(\mathbf{r}s)\mathbf{k}} = c_{(\mathbf{r}s)} \sum_{s'} \sum_{r'_x=1}^{n_x} \sum_{r'_y=1}^{n_y} \sum_{r'_z=1}^{n_z} (\mathbf{C}^{\circ Q})_{(\mathbf{r}s)(\mathbf{r}'s')} \mathbf{F}_{(\mathbf{r}'s')\mathbf{k}} (\boldsymbol{\Lambda}_{s'}^{1/2})_{\mathbf{kk}}.$$

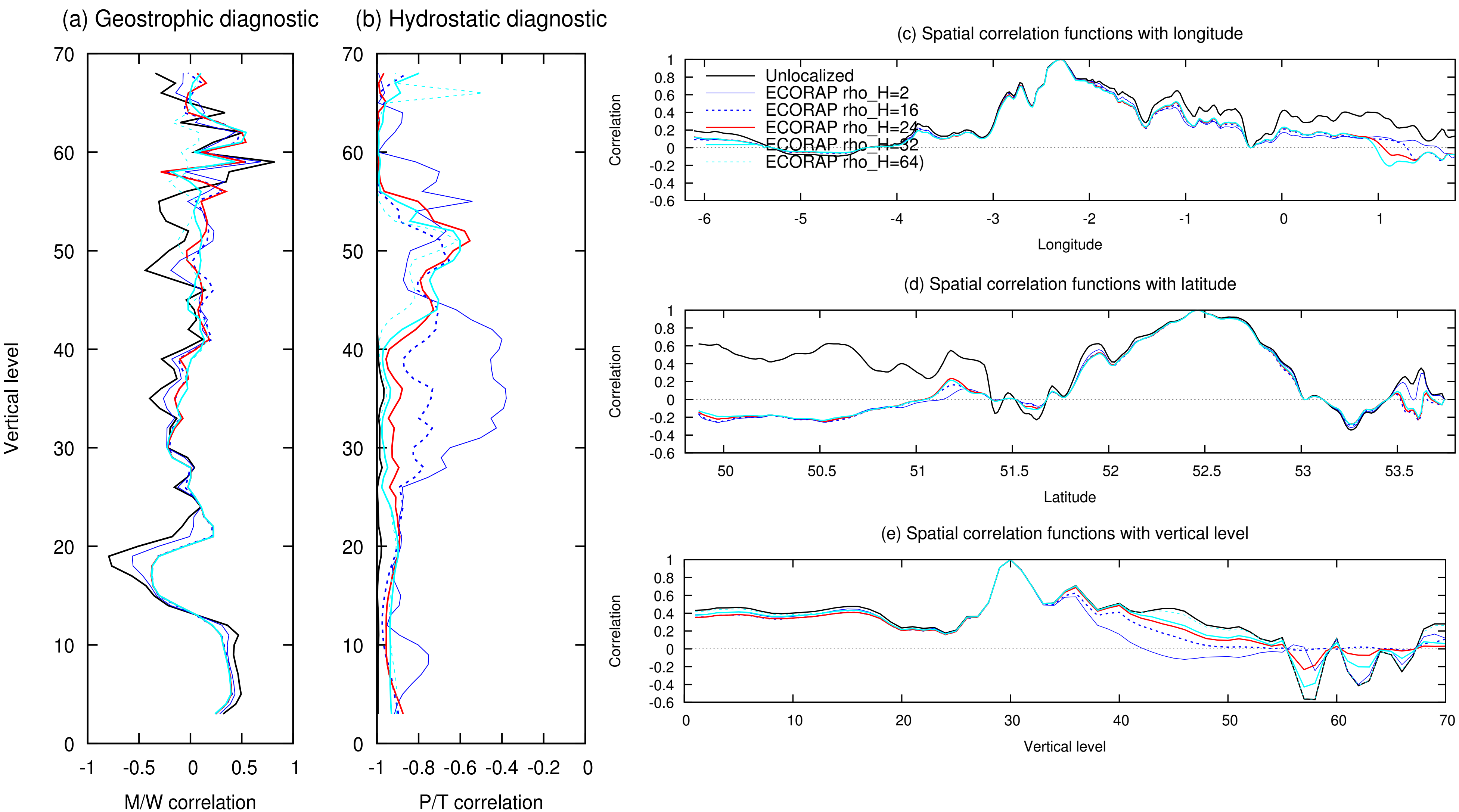
For efficiency (the original motivation...)

$$[\mathbf{K}^{\mathrm{ECORAP}}]_{(\mathbf{r}s)\mathbf{k}} \approx c_{(\mathbf{r}s)} \sum_{s'} \sum_{r'_x=r_x-\rho_{\mathrm{H}}}^{r_x+\rho_{\mathrm{H}}} \sum_{r'_y=r_y-\rho_{\mathrm{H}}}^{r_y+\rho_{\mathrm{H}}} \sum_{r'_z=r_z-\rho_{\mathrm{V}}}^{r_z+\rho_{\mathrm{V}}} (\mathbf{C}^{\circ Q})_{(\mathbf{r}s)(\mathbf{r}'s')} \mathbf{F}_{(\mathbf{r}'s')\mathbf{k}} (\boldsymbol{\Lambda}_{s'}^{1/2})_{\mathbf{kk}},$$

$$\rho_{\mathrm{H}}, \rho_{\mathrm{V}} \text{ influence radii of } \mathbf{C}^{\circ Q} \text{ in horizontal and vertical.}$$

[2] Bishop C.H. and Hodyss D., Ensemble covariances adaptively localized with ECO-RAP, Part 1: Tests on simple error models, Tellus A 61, 84-96 (2009).
Bishop C.H. and Hodyss D., Ensemble covariances adaptively localized with ECO-RAP, Part 2: A strategy for the atmosphere, Tellus A 61, 97-111 (2009).

Localization and balance results - ECO-RAP, $N = 24$, $Q = 2$, $K = 290$, $\rho_{\mathrm{H}} = 0$, $\rho_V = 2, 16, 24, 32, 64$



Summary & Conclusions

- Data assimilation attempts to combine imperfect data from models, from observations distributed in time and space, exploiting any relevant physical constraints, to produce a more accurate and comprehensive picture of the system as it evolves in time.

- All methods are approximate applications of Bayes' Theorem:

  - First moment of posteriori PDF:
    * Variational methods.
  - First and second moments of posteriori PDF:
    * Kalman filter.
    * Ensemble Kalman Filters.
  - Approximate whole PDF:
    * Particle filter.

- Unlikely for a method to work 'off the shelf':

  - E.g. in numerical weather prediction need to understand balances in atmosphere to model $\mathbf{B}$.
  - Can't deal with explicit matrices (n large).

- **Ensemble Kalman Filter**:

  - Suffers sampling error for $N < n$.
  - Use localization to remove long-range correlations in $\mathbf{P}_{\mathrm{f}}^{(N)}$.
  - Localization can destroy valuable information about balance.

- Three localization schemes studied:

  - Spectral (static).
  - SENCORP.
  - ECO-RAP.
  - ECO-RAP seems best compromise between ability to localize and preservation of balance.