The combined parameter and state estimation problem

Geir Evensen and Laurent Bertino

Hydro Research Centre, Bergen, Norway,

Nansen Environmental and Remote Sensing Center, Bergen, Norway

- Combined parameter and state estimation

  “Find the pdf of the parameters and the model solution conditional on a set of measurements.”

- Bayesian formulation.
- Ensemble solutions.

Bayesian formulation

Model equations and measurements:

$$
\begin{aligned}
\frac{\partial \boldsymbol{\psi}}{\partial t} &= \boldsymbol{g}(\boldsymbol{\psi}, \boldsymbol{\alpha}) + \boldsymbol{q}, \\
\boldsymbol{\psi}|_{t_0} &= \boldsymbol{\Psi}_0 + \boldsymbol{a}, \\
\boldsymbol{\alpha} &= \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}', \\
\boldsymbol{\mathcal{M}}(\boldsymbol{\psi}, \boldsymbol{\alpha}) &= \boldsymbol{d} + \boldsymbol{\epsilon}.
\end{aligned}
$$

Bayes theorem becomes:

$$
f(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto f(\boldsymbol{\psi} | \boldsymbol{\alpha}, \boldsymbol{\psi}_0) f(\boldsymbol{\psi}_0) f(\boldsymbol{\alpha}) \, f(\boldsymbol{d} | \boldsymbol{\psi}, \boldsymbol{\alpha}).
$$

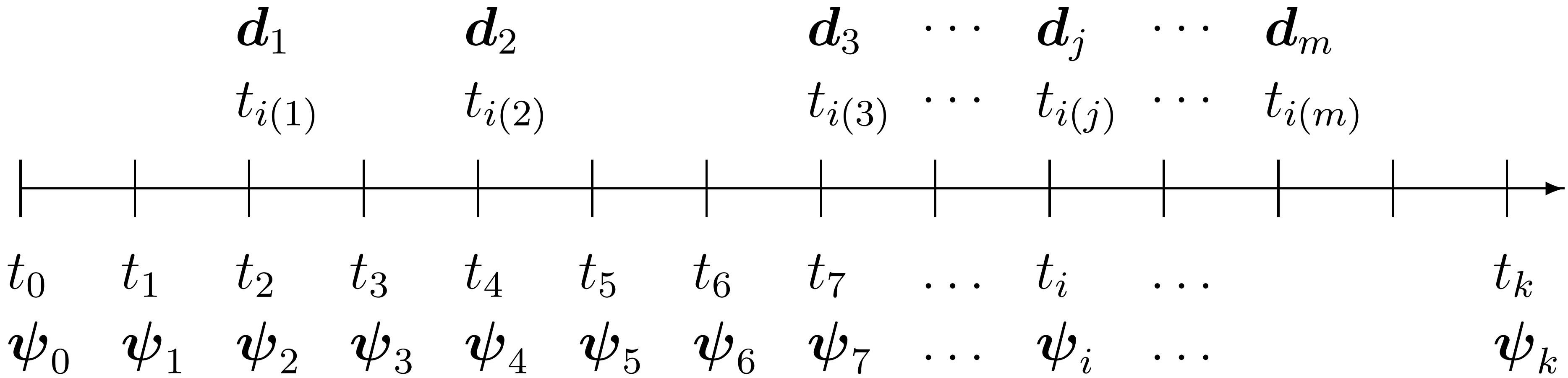
Gaussian priors

$$f(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto \exp\left\{-\frac{1}{2}\mathcal{J}[\boldsymbol{\psi}, \boldsymbol{\alpha}]\right\},$$

$$\begin{aligned}\mathcal{J}[\boldsymbol{\psi}, \boldsymbol{\alpha}] = & \left(\frac{\partial\boldsymbol{\psi}}{\partial t} - \boldsymbol{g}(\boldsymbol{\psi}, \boldsymbol{\alpha})\right)^{\mathrm{T}} \bullet \boldsymbol{W}_{qq} \bullet \left(\frac{\partial\boldsymbol{\psi}}{\partial t} - \boldsymbol{g}(\boldsymbol{\psi}, \boldsymbol{\alpha})\right) \\ & + (\boldsymbol{\psi}_0 - \boldsymbol{\Psi}_0)^{\mathrm{T}} \circ \boldsymbol{W}_{aa} \circ (\boldsymbol{\psi}_0 - \boldsymbol{\Psi}_0) \\ & + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}} \circ \boldsymbol{W}_{\alpha\alpha} \circ (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\ & + (\boldsymbol{d} - \boldsymbol{\mathcal{M}}[\boldsymbol{\psi}, \boldsymbol{\alpha}])^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d} - \boldsymbol{\mathcal{M}}[\boldsymbol{\psi}, \boldsymbol{\alpha}]).\end{aligned}$$

- MLH solution, hard to solve and to compute error estimates.

Discretisation in time



Assume

- Model is first order Markov process.

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0) \propto$$

$$f(\boldsymbol{\alpha}) f(\boldsymbol{\psi}_0) \prod_{i=1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}).$$

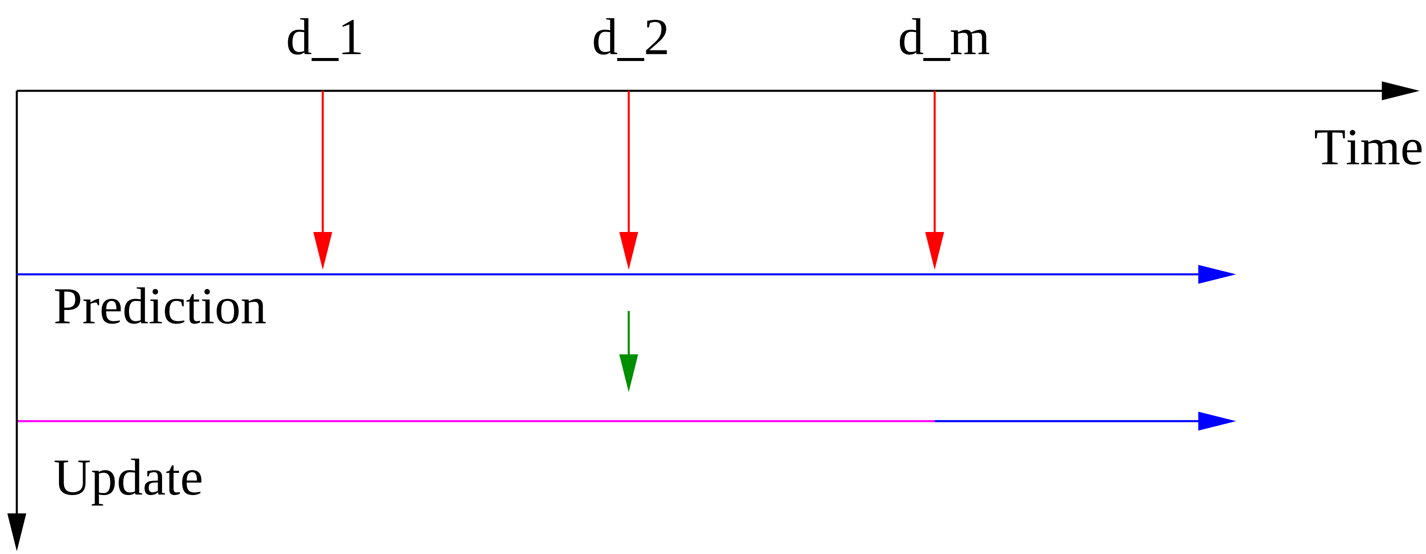
- Measurement errors are independent in time

$$f(\boldsymbol{d} | \boldsymbol{\psi}, \boldsymbol{\alpha}) = \prod_{j=1}^{m} f(\boldsymbol{d}_j | \boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}).$$

Bayes for discrete state

- Bayes theorem then becomes

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto f(\boldsymbol{\alpha}) f(\boldsymbol{\psi}_0) \prod_{i=1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) \prod_{j=1}^{m} f(\boldsymbol{d}_j | \boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}),$$



Rewrite as:

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto f(\boldsymbol{\alpha}) f(\boldsymbol{\psi}_0)$$

$$\prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha})$$

$$\prod_{i=i(1)+1}^{i(2)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_2 | \boldsymbol{\psi}_{i(2)}, \boldsymbol{\alpha}) \cdots$$

$$\prod_{i=i(m-1)+1}^{i(m)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_m | \boldsymbol{\psi}_{i(m)}, \boldsymbol{\alpha})$$

$$\prod_{i=i(m)+1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha})$$

Sequential processing of measurements

- First update

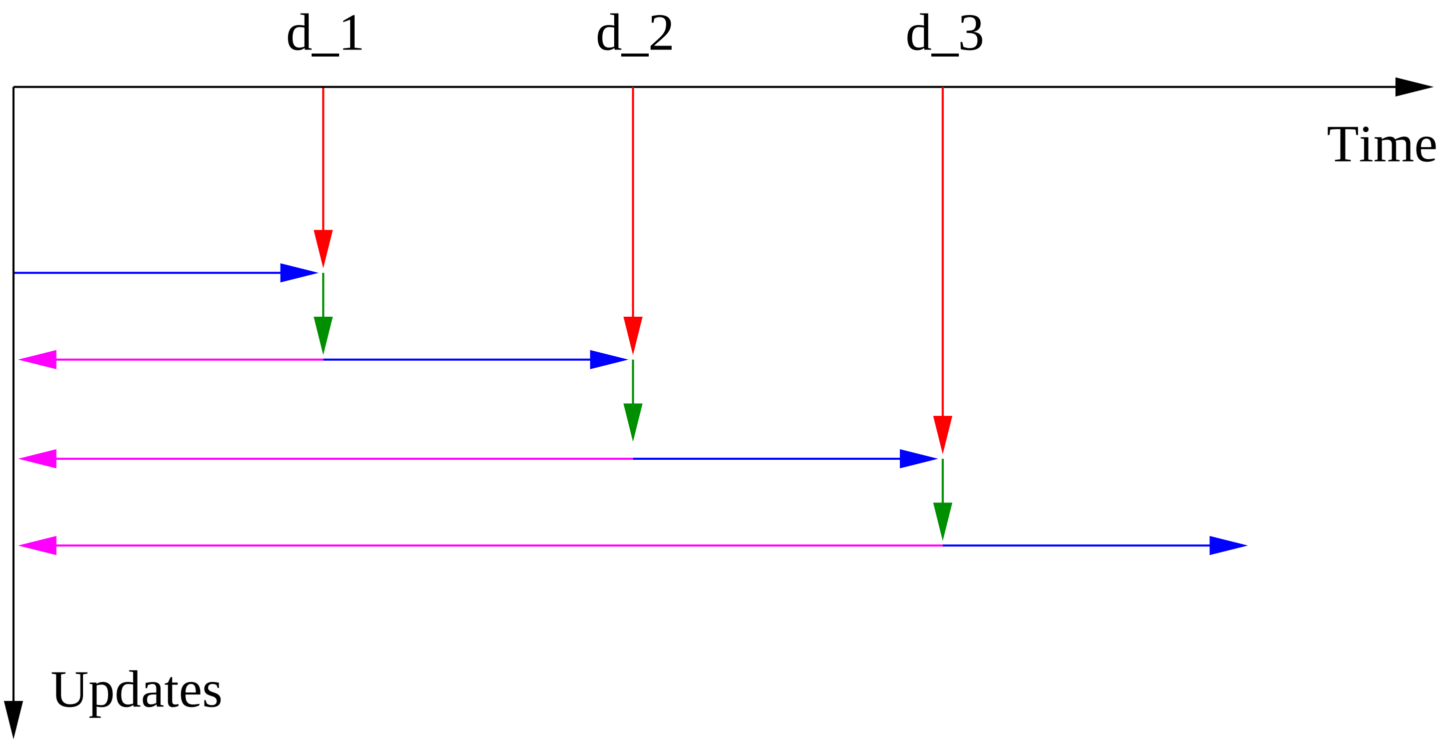
$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}_1) \propto f(\boldsymbol{\alpha}) f(\boldsymbol{\psi}_0) \prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha})$$

- Second update

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i(2)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}_1, \boldsymbol{d}_2) \propto f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}_1) \prod_{i=i(1)+1}^{i(2)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_2 | \boldsymbol{\psi}_{i(2)}, \boldsymbol{\alpha})$$

Summary

- Independent measurements processed sequentially in time.
- Sequence of inverse problems.
- Solution of one sub-problem is prior for next.
- Hard to solve using traditional variational methods.
- Well suited for ensemble methods.



Variance minimizing analysis

Given

- First guess, ψ^{f}, and prediction error covariance, $\boldsymbol{C}_{\psi\psi}^{\mathrm{f}}$.
- Measurements, $\boldsymbol{d}$, and error covariance, $\boldsymbol{C}_{\epsilon\epsilon}$.

The variance minimizing analysis is minimum of

$$\begin{aligned}\mathcal{J}[\boldsymbol{\psi}^{\mathrm{a}}] = &\left(\boldsymbol{\psi}^{\mathrm{a}} - \boldsymbol{\psi}^{\mathrm{f}}\right)^{\mathrm{T}} \bullet \boldsymbol{C}_{\psi\psi}^{-1} \bullet \left(\boldsymbol{\psi}^{\mathrm{a}} - \boldsymbol{\psi}^{\mathrm{f}}\right) \\ &+ \left(\boldsymbol{d} - \boldsymbol{\mathcal{M}}\psi^{\mathrm{a}}\right)^{\mathrm{T}} \boldsymbol{C}_{\epsilon\epsilon}^{-1} \left(\boldsymbol{d} - \boldsymbol{\mathcal{M}}\psi^{\mathrm{a}}\right).\end{aligned}$$

Analysis equations

Update of estimate

$$\boldsymbol{\psi}^{\mathrm{a}} = \boldsymbol{\psi}^{\mathrm{f}} + \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}}\left(\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\left(\boldsymbol{d} - \boldsymbol{\mathcal{M}}\boldsymbol{\psi}^{\mathrm{f}}\right)$$

Error covariance for update

$$\boldsymbol{C}^{\mathrm{a}}_{\psi\psi} = \boldsymbol{C}^{\mathrm{f}}_{\psi\psi} - \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}}\left(\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}$$

Ensemble methods

ES: Ensemble Smoother

EnKS: Ensemble Kalman Smoother

EnKF: Ensemble Kalman Filter

- Ensemble representation for pdfs.
- Ensemble prediction for time evolution of pdfs.
- Linear ensemble analysis scheme:
  - "Variance minimizing".
  - Assumes Gaussian pdf for model prediction.
  - No resampling!

The error covariance matrix

Define ensemble covariances around the ensemble mean

$$\boldsymbol{C}_{\psi\psi} = \overline{\left(\boldsymbol{\psi} - \overline{\boldsymbol{\psi}}\right)\left(\boldsymbol{\psi} - \overline{\boldsymbol{\psi}}\right)^{\mathrm{T}}}$$

- The ensemble mean, $\overline{\psi}$, is the best-guess.
- The error variance is defined by the ensemble spread.
- The smoothness of members defines the covariance.

$\Downarrow$

Representation by a finite ensemble of model states.

Ensemble representation

- Define ensemble matrix

$$\boldsymbol{A}(\boldsymbol{x},t_i) = \begin{pmatrix} \boldsymbol{\psi}^1(\boldsymbol{x},t_i) & \boldsymbol{\psi}^2(\boldsymbol{x},t_i) & \dots & \boldsymbol{\psi}^N(\boldsymbol{x},t_i) \\ \boldsymbol{\alpha}^1(\boldsymbol{x},t_i) & \boldsymbol{\alpha}^2(\boldsymbol{x},t_i) & \dots & \boldsymbol{\alpha}^N(\boldsymbol{x},t_i) \end{pmatrix}.$$

- Parameters augmented to model state.
- The ensemble covariance becomes

$$\boldsymbol{C}^{\mathrm{e}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2,t_i) = \frac{\boldsymbol{A}'(\boldsymbol{x}_1,t_i)\boldsymbol{A}_i'^{\mathrm{T}}(\boldsymbol{x}_2,t_i)}{N-1}.$$

Dynamical evolution of error statistics

- Members evolve according to stochastic model dynamics

$$d\boldsymbol{\psi} = \boldsymbol{g}(\boldsymbol{\psi}, \boldsymbol{\alpha})dt + \boldsymbol{h}(\boldsymbol{\psi})d\boldsymbol{q}$$
$$d\boldsymbol{\alpha} = 0.$$

- The pdf evolves according to Kolmogorov's equation

$$\frac{\partial f}{\partial t} + \sum_i \frac{\partial (g_i f)}{\partial \psi_i} = \frac{1}{2} \sum_{i,j} \frac{\partial^2 f \left(\boldsymbol{h}\boldsymbol{Q}\boldsymbol{h}^T\right)_{ij}}{\partial \psi_i \partial \psi_j}.$$

- Fundamental equation for evolution of error statistics.
- May be solved using Monte Carlo methods.

Ensemble analysis equations

- Define randomized measurements $\boldsymbol{D} = \boldsymbol{d} + \boldsymbol{E}$.
- Variance minimizing ensemble update becomes

$$
\begin{aligned}
\boldsymbol{A}^{\mathrm{a}} &= \boldsymbol{A} \\
&\quad + \boldsymbol{C}^{\mathrm{e}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}}\left(\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{e}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{\mathcal{M}}\boldsymbol{A}\right) \\
&= \boldsymbol{A} \\
&\quad + \boldsymbol{A}'(\boldsymbol{\mathcal{M}}\boldsymbol{A}')^{\mathrm{T}}\left(\boldsymbol{\mathcal{M}}\boldsymbol{A}'(\boldsymbol{\mathcal{M}}\boldsymbol{A}')^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{\mathcal{M}}\boldsymbol{A}\right) \\
&\;\vdots \\
&= \boldsymbol{A}\boldsymbol{X}
\end{aligned}
$$

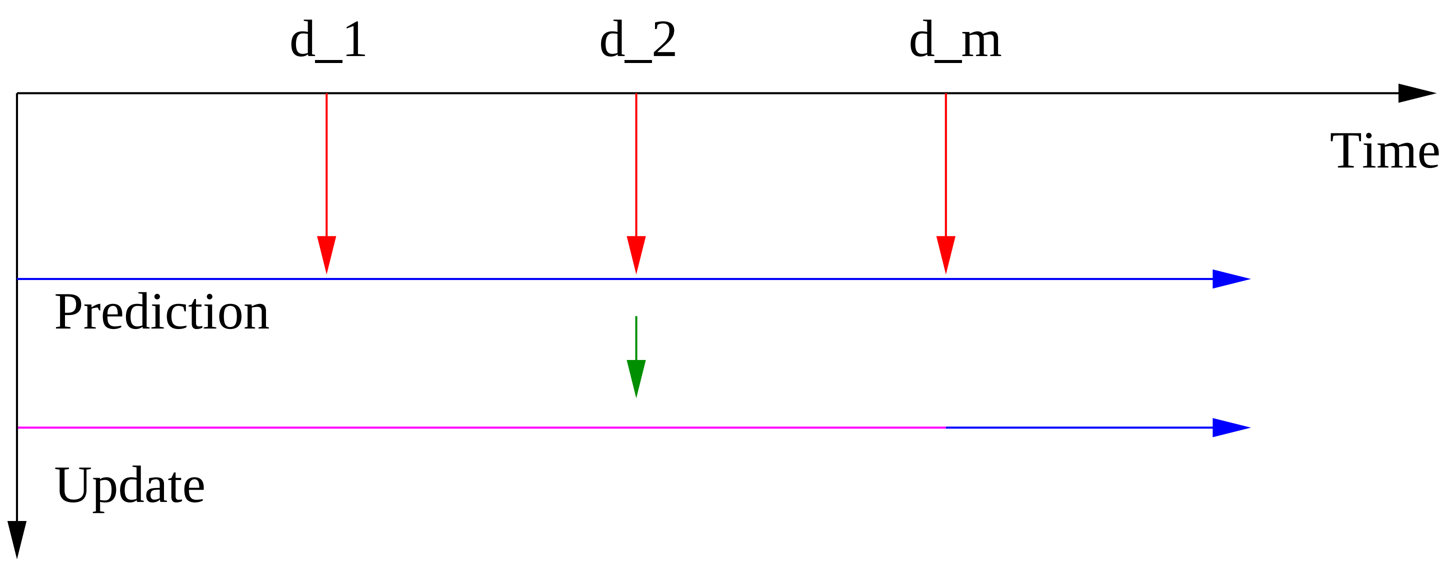
$\Rightarrow$ Updated ensemble with correct mean and covariance!

ES: The Ensemble Smoother

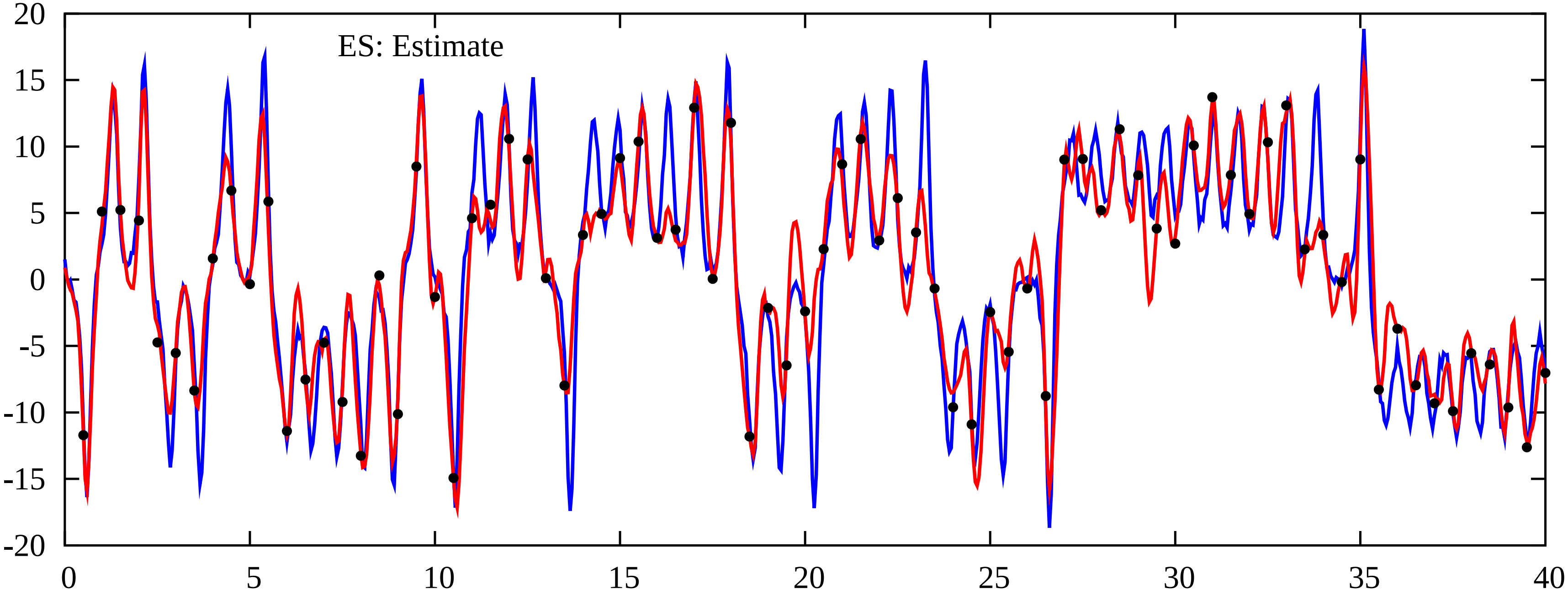
Computes the Bayesian update using linear ensemble equation

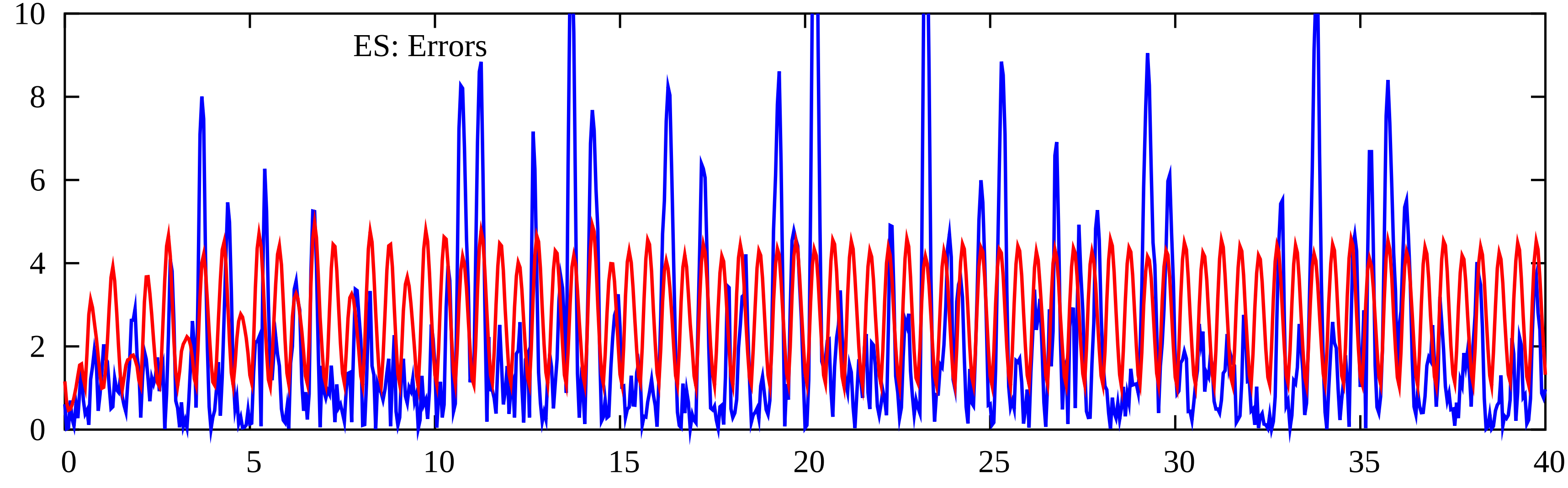
$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}) =$$

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\alpha}, \boldsymbol{\psi}_0) \prod_{j=1}^{m} f(\boldsymbol{d}_j | \boldsymbol{\psi}_{i(j)}, \boldsymbol{\alpha}),$$



ES: Example with Lorenz equations





ES: summary

Gauss–Markov interpolation in space and time.

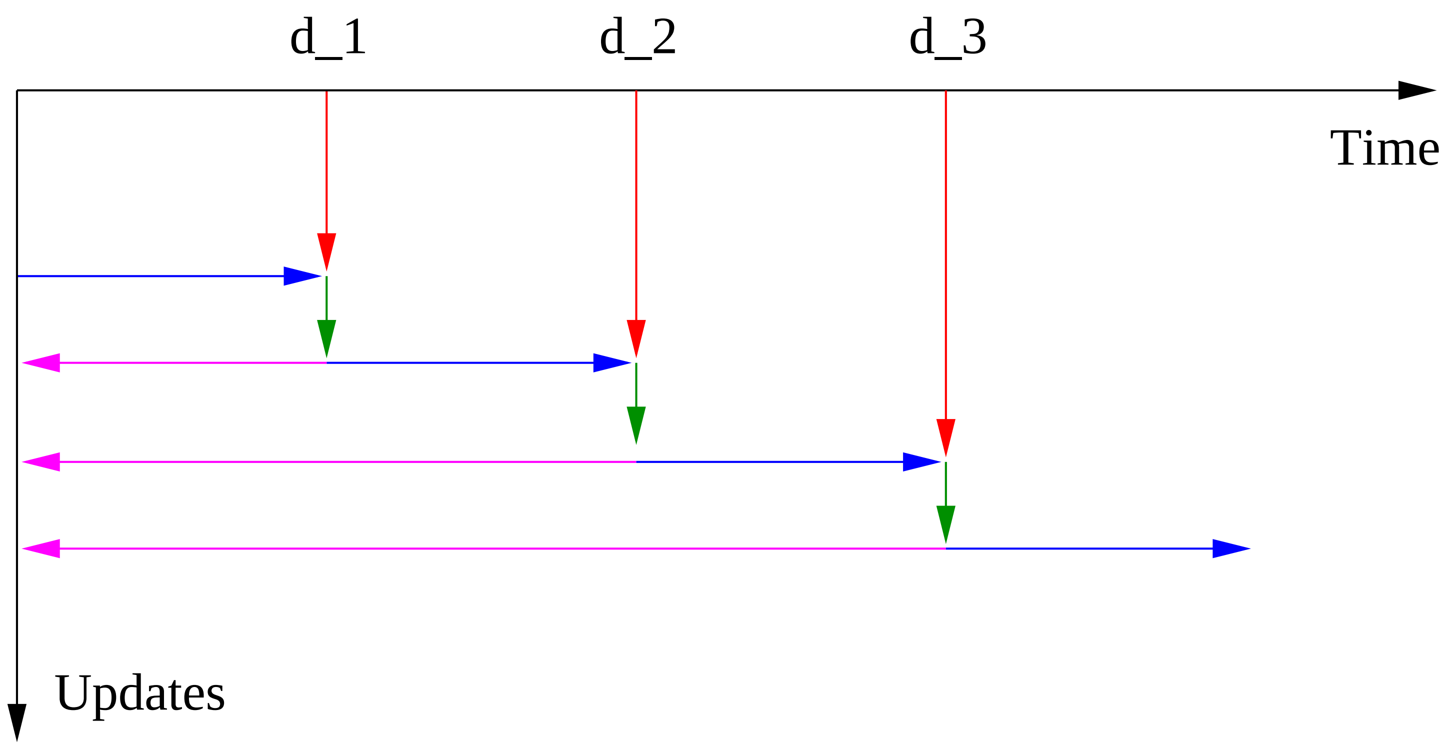
- Creates an ensemble for the model evolution.
- Assumes Gaussian pdf for model evolution.
- Computes variance minimizing ensemble analysis.
- Exact solution for linear problems.

EnKS: The ensemble Kalman smoother

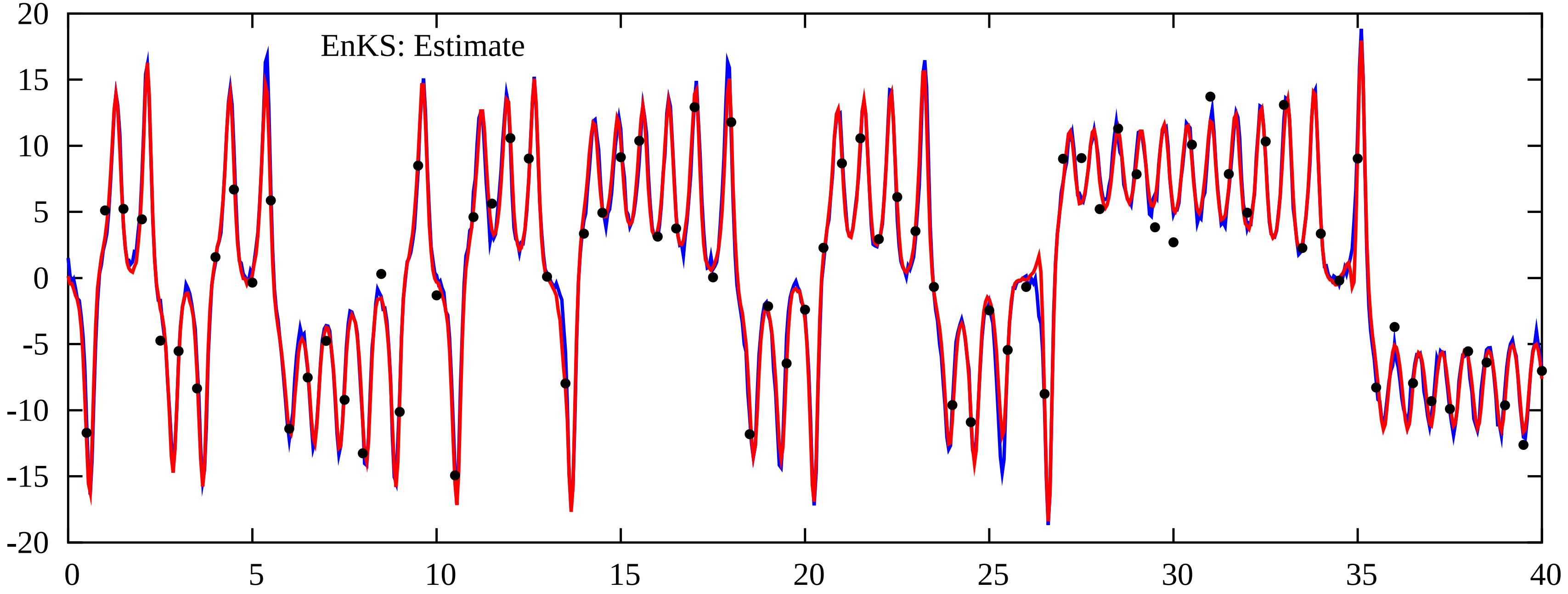
- Bayes with sequential processing of data: First update is

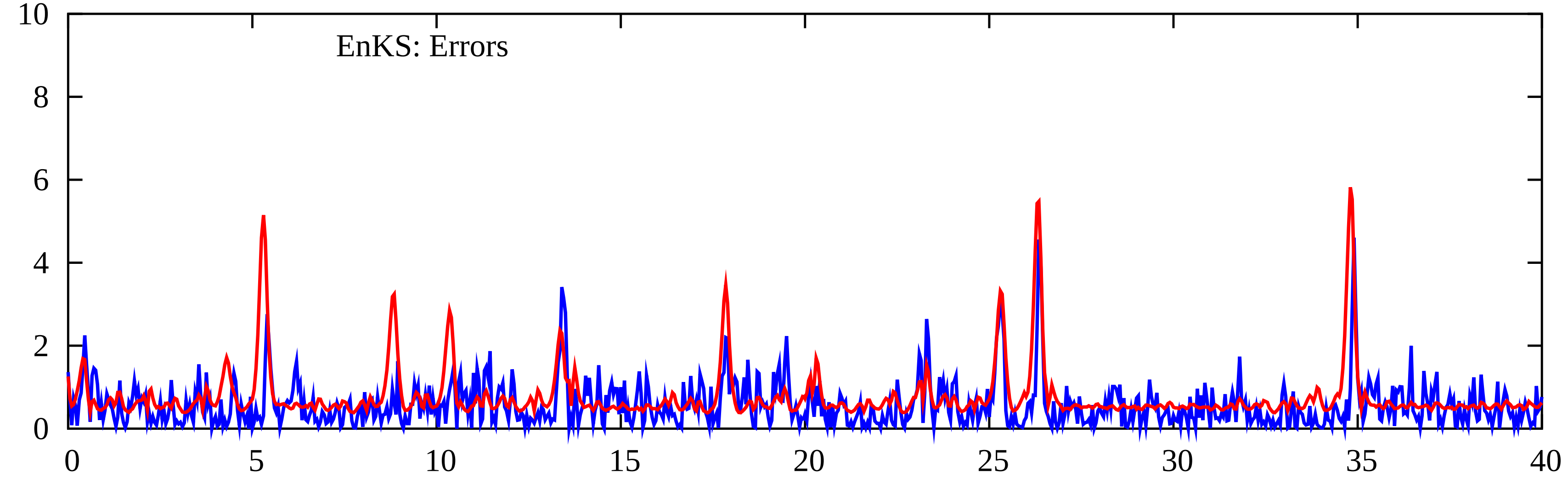
$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}, \boldsymbol{\psi}_0 | \boldsymbol{d}_1) \propto$$

$$f(\boldsymbol{\alpha}) f(\boldsymbol{\psi}_0) \prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}, \boldsymbol{\alpha}) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)}, \boldsymbol{\alpha}),$$



EnKS solution



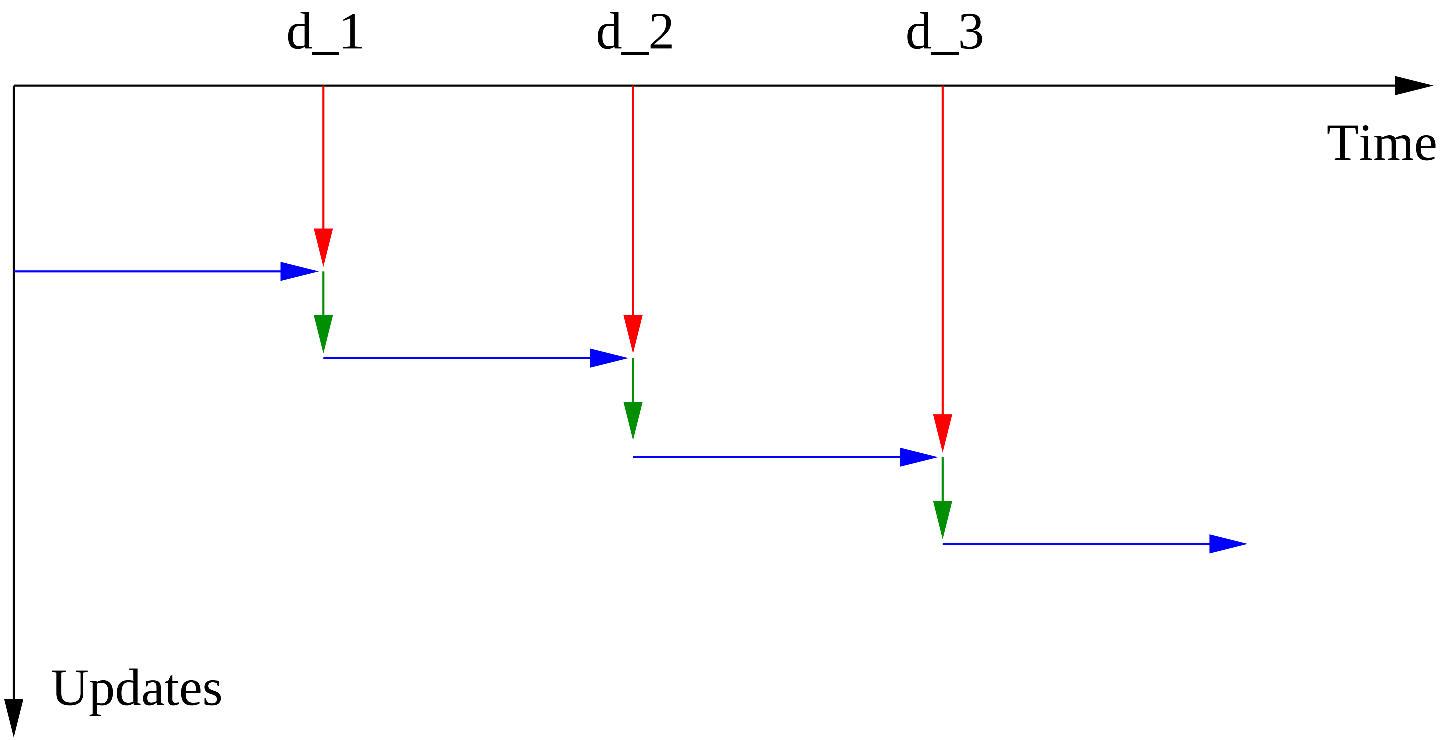


EnKS summary

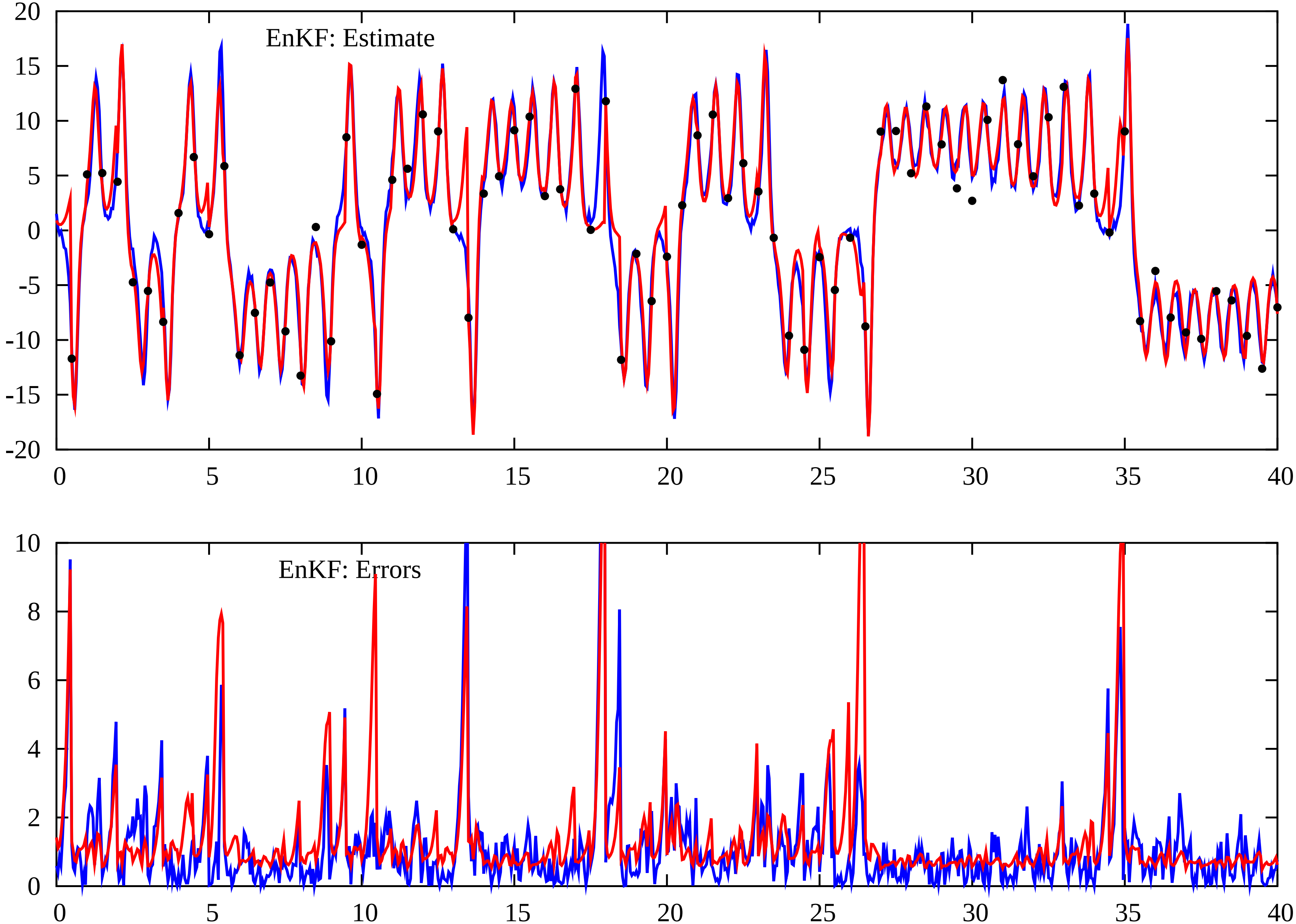
- ES and EnKS give identical results for linear models.
- EnKS is superior to the ES with nonlinear models.
  - Sequential processing of measurements introduces “Gaussianity”.
  - Ensemble is kept close to the true state.

EnKF: Ensemble Kalman Filter

- Special case of EnKS.
- Only update state at data times
- EnKF forms a prior for the EnKS.
- A prediction from EnKF and EnKS will be identical.



EnKF solution



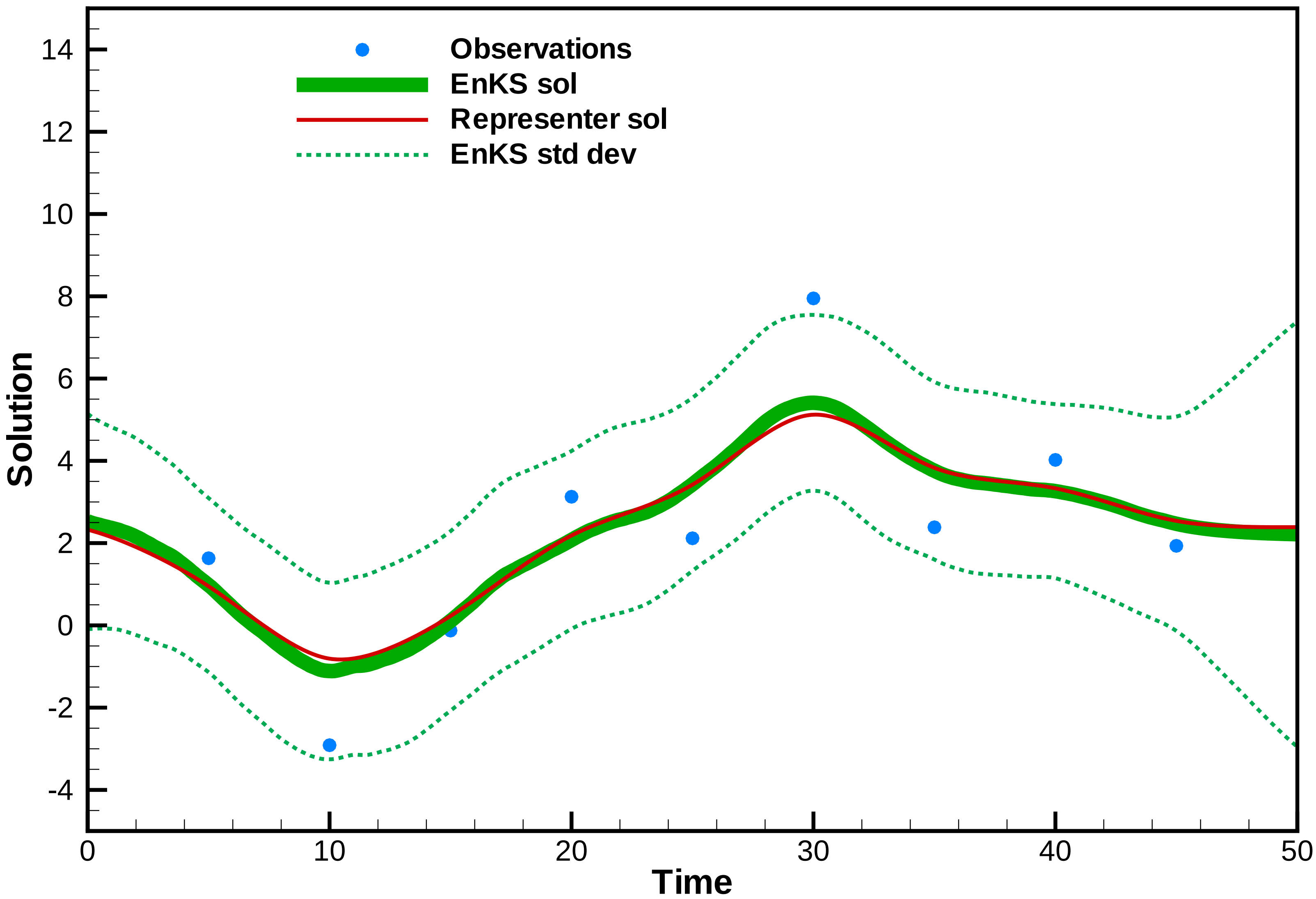
Example

- Scalar model

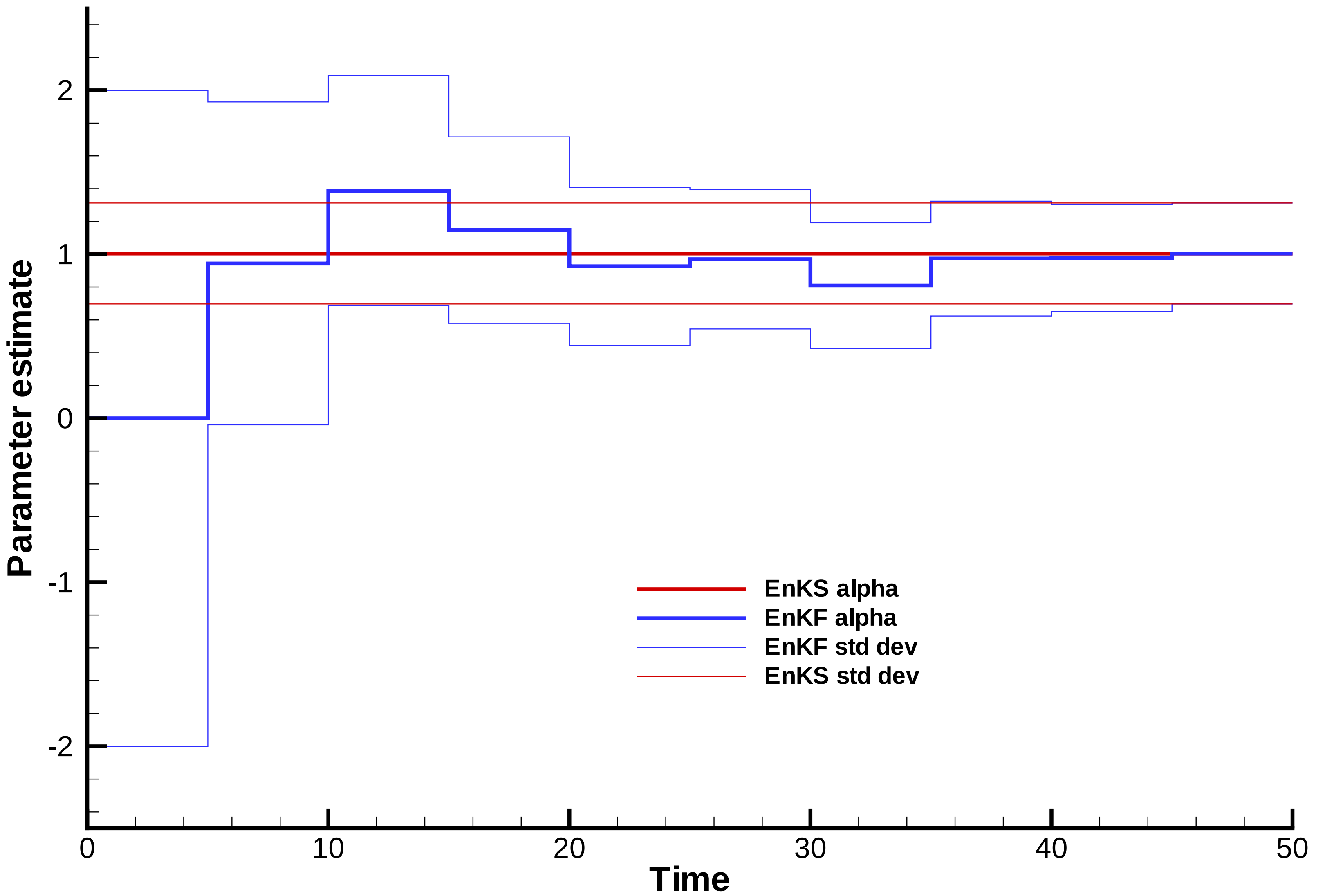
$$\frac{\partial \psi}{\partial t} = 1 - \alpha + q,$$
$$\psi(t=0) = 3 + a,$$
$$\alpha = 0 + \alpha',$$
$$\mathcal{M}(\psi) = \boldsymbol{d} + \boldsymbol{\epsilon}.$$

- True parameter value is $\alpha = 1$.
- Model operator is linear and independent of ψ.
- Solved using EnKF, EnKS and Representer methods.
- Exponential time correlation for model errors.

State and parameter estimation



Estimate of parameter



Summary

www.nersc.no/~geir/EnKF

- Joint state and parameter estimation problem!
- Bayesian formulation:
  - Variational methods (multiple minima; MLH).
    - Hard to compute error statistics!
    - Do not allow for sequential processing of measurements!
  - Ensemble methods (Gaussianity assumption; mean).
    - Provide estimate with error statistics.
    - Allow for sequential processing of measuremets.
      - Introduces Gaussianity!
      - Advantage with nonlinear dynamics.