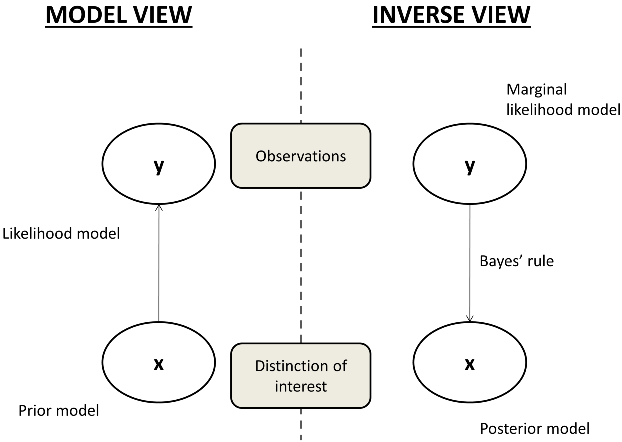
Ensemble Kalman Filter

Jo Eidsvik

Mathematical Sciences, NTNU, Norway

2019

Bayesian illustration



Bayesian approach

- Variables of interest $\boldsymbol{x} = (x_1, \ldots, x_n)$. Prior $p(\boldsymbol{x})$.
- Data $\boldsymbol{y} = (y_1, \ldots, y_N)$. Likelihood $p(\boldsymbol{y}|\boldsymbol{x})$.

Common setting:

- Likelihood: $p(\boldsymbol{y}|\boldsymbol{x})$, for instance $p(\boldsymbol{y}|\boldsymbol{x}) = \text{Normal}(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{\Sigma}_e)$
- Prior: $p(\boldsymbol{x})$, for instance $p(\boldsymbol{x}) = \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Posterior: $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$

Sequential data assimilation

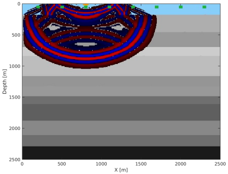
- Structure data or variables in increasing order ('time').
- Update variables recursively moving through the order ('time').

Why do this?

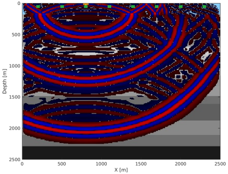
- Structure natural for realistic modeling of variables over order ('time').
- Data arrive sequentially, and we want on-line decisions.
- Split and conquer for handling large problems.

Waveform inversion

- The objective is to predict (background) velocity distribution from raw seismic recordings.
- Use sequential updating (over shot data and receiver time).



Wavefield at time 400 msec.

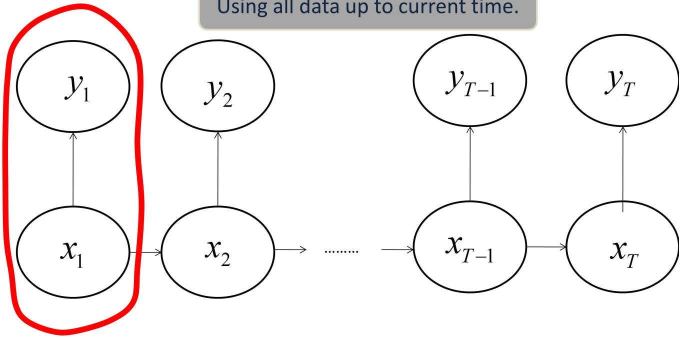


Wavefield at time 750 msec.

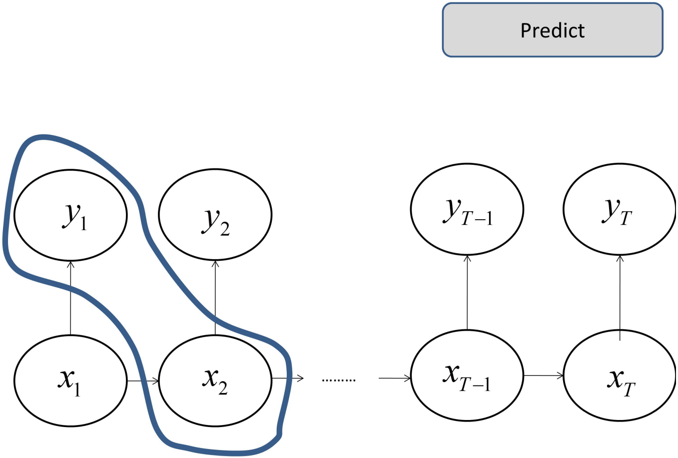
Filter

Filtering

Using all data up to current time.



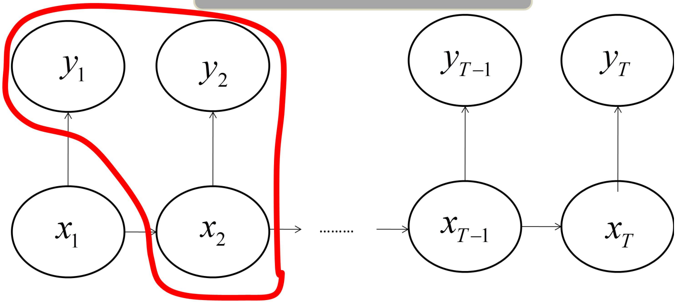
Predict



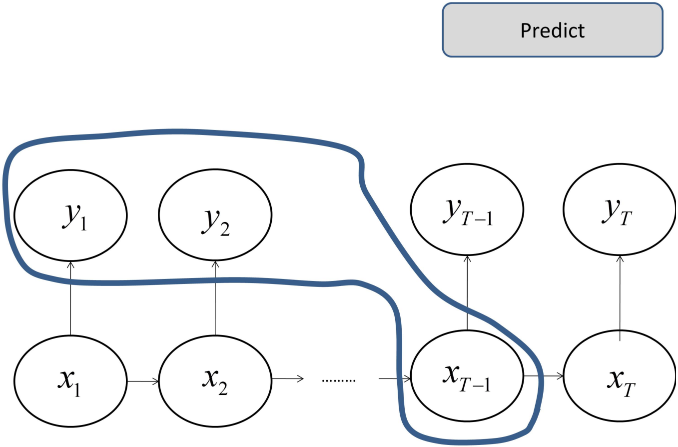
Filter

Filtering

Using all data up to current time.



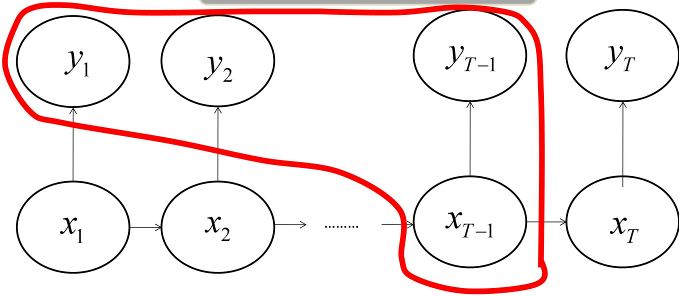
Predict



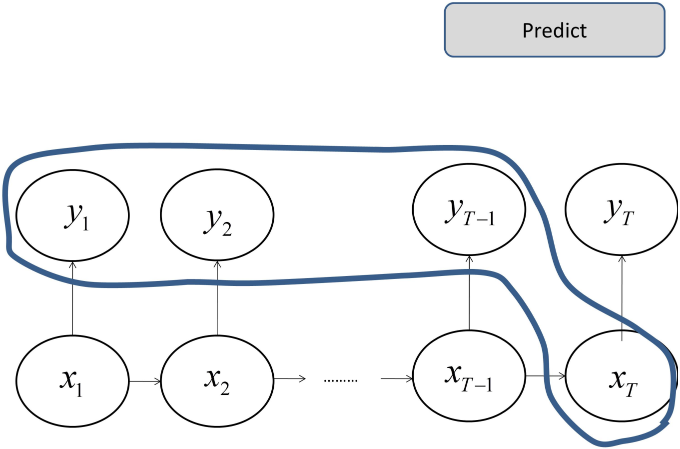
Filter

Filtering

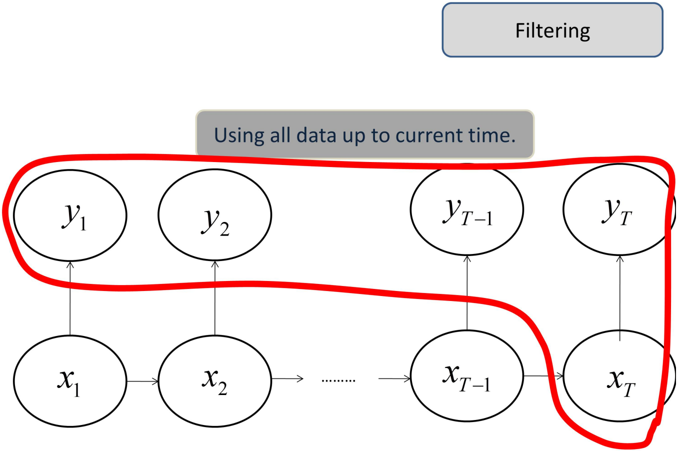
Using all data up to current time.



Predict



Predict


Filtering
Using all data up to current time.
y_1
y_2
y_{T-1}
y_T
x_1
x_2
....
x_{T-1}
x_T

Prediction, filtering and smoothing

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_{t-1}) = \int p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1}$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_t) = \frac{p(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t)}{p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1)}$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_T), \qquad p(\boldsymbol{x}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T)$$

Exact closed-form solutions:

- Markov models (Forward-backward algorithm)
- Gaussian linear models (Kalman filter, smoother)

Not so easy for other models...

Linear Gaussian model assumptions

Conditional independence in process (state) model:

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{y}_1, ...\boldsymbol{y}_{t-1}) = N(\boldsymbol{F}_t \boldsymbol{x}_{t-1}, \boldsymbol{Q}_t)$$

Simplest setting (static model): $\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$

Conditional independence in measurement model:

$$p(\boldsymbol{y}_t | \boldsymbol{y}_1, ...\boldsymbol{y}_{t-1}, \boldsymbol{x}_t) = N(\boldsymbol{G}_t \boldsymbol{x}_t, \boldsymbol{R}_t)$$

Kalman filter

Elegant form for building the Gaussian distribution for prediction and filtering/analysis/assimilation:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_{t-1}) = \int p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1}$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_{t-1}) = N(\boldsymbol{\mu}_{t|t-1},\boldsymbol{\Sigma}_{t|t-1})$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1,....\boldsymbol{y}_t) \propto p(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t) = N(\boldsymbol{\mu}_{t|t},\boldsymbol{\Sigma}_{t|t})$$

Kalman filter : Prediction step

With linear expectation and Gaussian additive noise, the models remain Gaussian. Need mean and covariance.
$\boldsymbol{\delta}_t \sim N(0, \boldsymbol{R})$

$$\boldsymbol{\mu}_{t|t-1} = E(\boldsymbol{F}_t \boldsymbol{x}_{t-1} + \boldsymbol{\delta}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{F}_t \boldsymbol{\mu}_{t-1|t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathrm{Var}(\boldsymbol{F}_t \boldsymbol{x}_{t-1} + \boldsymbol{\delta}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{F}_t \boldsymbol{\Sigma}_{t-1|t-1} \boldsymbol{F}_t^T + \boldsymbol{Q}_t$$

Kalman filter update : Joint Gaussian

$$p(\boldsymbol{x}_t|\boldsymbol{y}_1, ...\boldsymbol{y}_t) \propto p(\boldsymbol{x}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t) = N(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$$

The updating with data $\boldsymbol{y}_t$ relies on this result for joint Gaussian variables:

$$\begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{A,B} \\ \boldsymbol{\Sigma}_{B,A} & \boldsymbol{\Sigma}_B \end{pmatrix}\right]$$

$$[\boldsymbol{x}_A|\boldsymbol{x}_B] \sim N(\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{B,A})$$

Kalman filter update : Joint Gaussian

$p(\boldsymbol{x}_t, \boldsymbol{y}_t | \boldsymbol{y}_1, ...\boldsymbol{y}_{t-1})$ is joint Gaussian

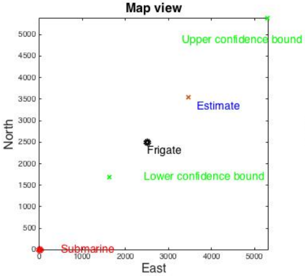
$$\begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_{t|t-1} \\ \boldsymbol{G}_t \boldsymbol{\mu}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{G}_t^T \\ \boldsymbol{G}_t \boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{G}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{G}_t^T + \boldsymbol{R}_t \end{pmatrix}\right]$$

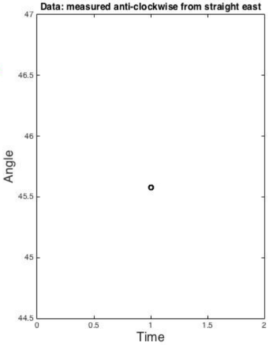
$$[\boldsymbol{x}_t | \boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1] \sim N(\boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{G}_t \boldsymbol{\mu}_{t|t-1}), \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{G}_t \boldsymbol{\Sigma}_{t|t-1})$$

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{G}_t^T [\boldsymbol{G}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{G}_t^T + \boldsymbol{R}_t]^{-1}$$

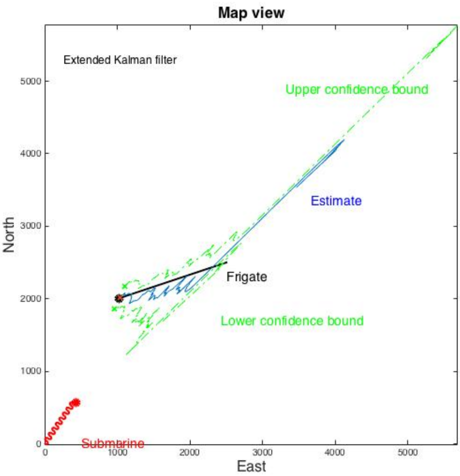
Traditional example

Gained popularity in space missions. Satellite control. Tracking planes and ships.





Traditional example



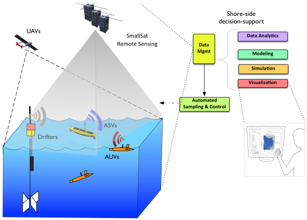
Spatio-temporal Kalman filter

Autoregressive space-time process:

$$\boldsymbol{x}_t = \phi \boldsymbol{x}_{t-1} + N(0, \boldsymbol{Q})$$

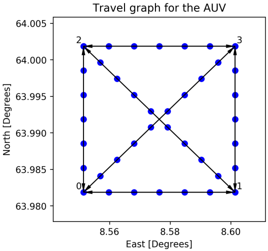
Measure at some or all locations:

$$\boldsymbol{y}_t = \boldsymbol{G}_t \boldsymbol{x}_t + N(0, \boldsymbol{R})$$



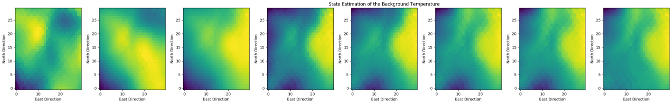
Spatio-temporal Kalman filter

AUV sampling of temperature data



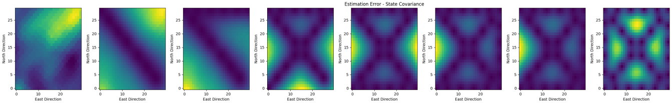
Spatio-temporal Kalman filter

Predictions from dynamic exploration of ocean temperatures.



Spatio-temporal Kalman filter

Prediction variance from dynamic exploration of ocean temperatures.



Approximate algorithms

The Kalman filter is exact only under idealized conditions (linear, Gaussian assumptions).
Approximations must be used in more complex settings.
- common to use linearized, Gaussian approaches, ensemble-based methods or Markov chain Monte Carlo.

Approximate algorithms

Approximations ; pros and cons.

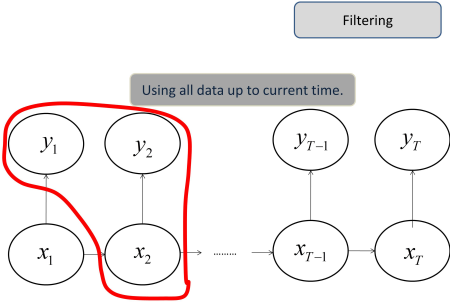
Criterion	EKF	UKF	EnKF	PF
Closed-form updating	V	V	V	
MC based			V	V
Non-linear	w	V	V	V
Scales with dim.	V		V	
Reliable UQ		V	w	V

- focus on EnKF (ensemble-based) methods.

Ensemble Kalman filter (EnKF)

- Monte Carlo based data assimilation
- Forecast using (non-linear) physical relations
- Assimilation based on **linear** update

(Evensen, 2009).



(Other approaches exist (exercise).)

Ensemble-based Kalman approximation

- Ensemble size B.
- Realizations $\boldsymbol{x}^b$, $b = 1, \ldots, B$ from $p(\boldsymbol{x}_1) = \text{Normal}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$.
- Recursively assimilate data over 'time' $t = 1, \ldots, T$.
- Kalman update of bth ensemble member at step t

$$\boldsymbol{x}_t^b = \boldsymbol{x}_t^b + \hat{\boldsymbol{K}}_t(\boldsymbol{y}_t - \boldsymbol{y}_t^b)$$

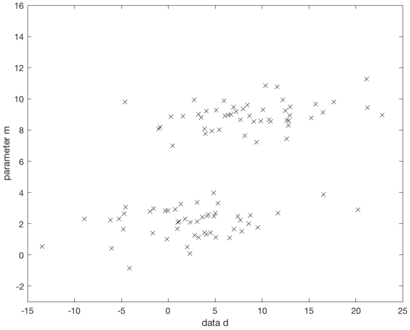
- $\boldsymbol{y}_t^b = \boldsymbol{g}(\boldsymbol{x}_t^b) + N(0, \boldsymbol{R}_t)$.
- Kalman weight matrix $\hat{\boldsymbol{K}}_t = \hat{\boldsymbol{\Sigma}}_{xy,t}\left(\hat{\boldsymbol{\Sigma}}_{yy,t} + \boldsymbol{R}_t\right)^{-1}$ determined empirically from forecast ensembles $(\boldsymbol{x}_t^b, \boldsymbol{y}_t^b)$, $b = 1, \ldots, B$.

EnKF properties

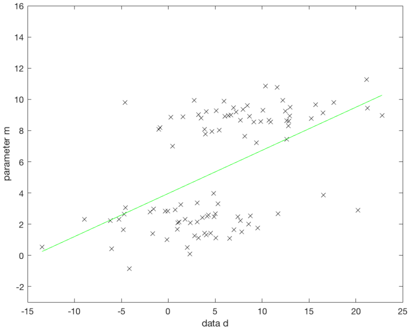
- Ensemble based approximation to filtering distribution.
- Exact for Gaussian linear systems when the ensemble size goes to infinity.
- Unclear how it performs for nonlinear systems (no mathematical results), but in practice ensemble-based KF versions are performing really well.

Univariate example - forecast samples

$$x^b \sim p(x), \qquad y^b = x^b + N(0, 5^2)$$

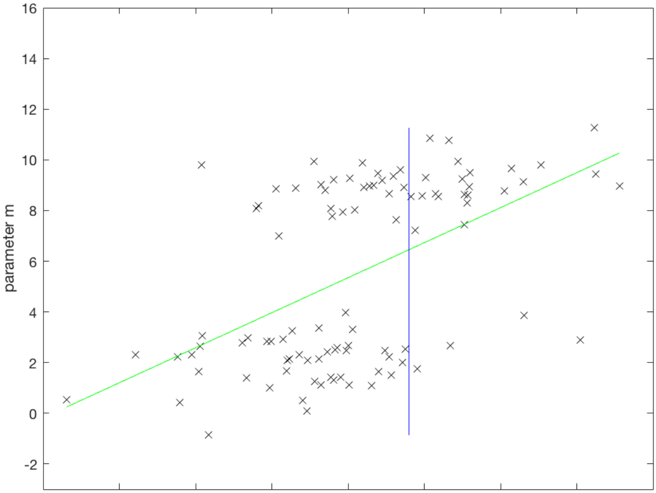


Univariate example - regression fit

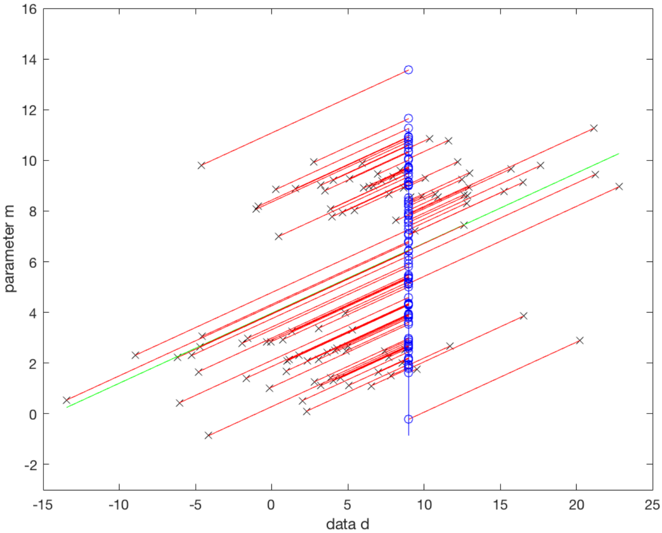


Univariate example - observation

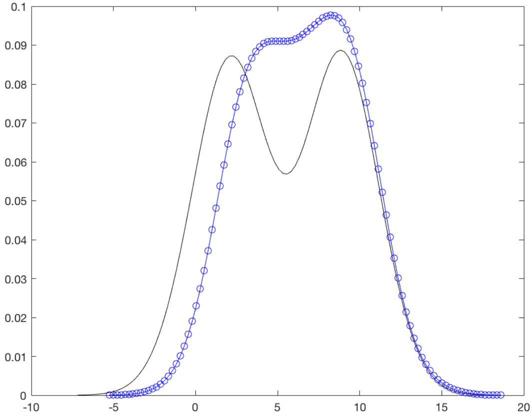
$y = 9.$



Univariate example - analysis or update step



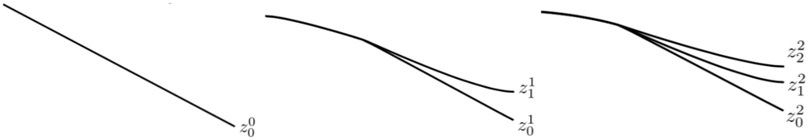
Univariate example - prior and posterior



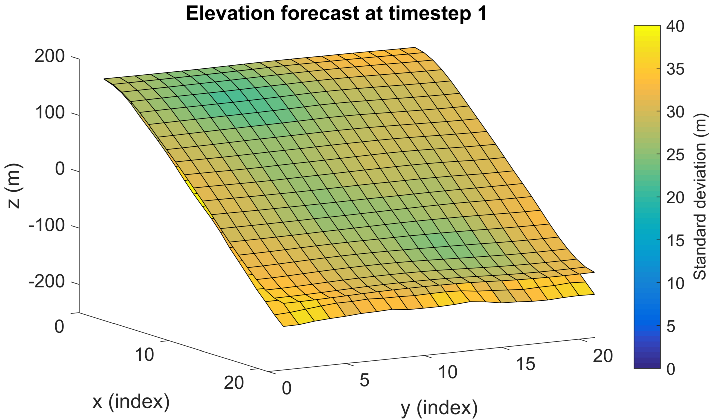
Geologic process models

Differential equation for sedimentation, corrected with data.

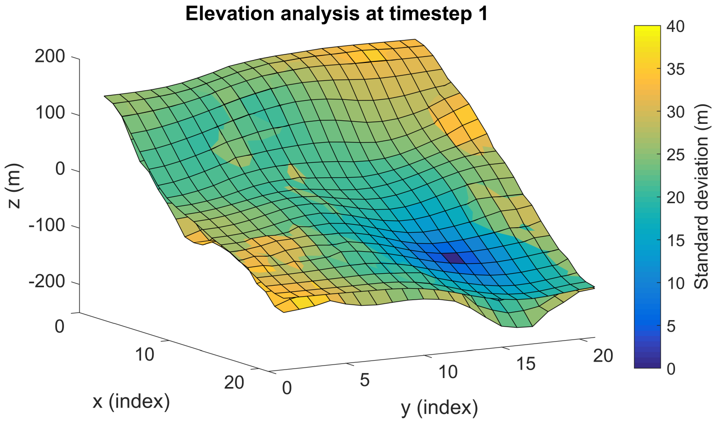
- Start with initial surface z_0^0 at time t_0
- Surface will "diffuse" to yield new top surface z_1^1 at time t_1
- Elevation of surface j at time t_k is z_j^k for $j = 1, \ldots, k$



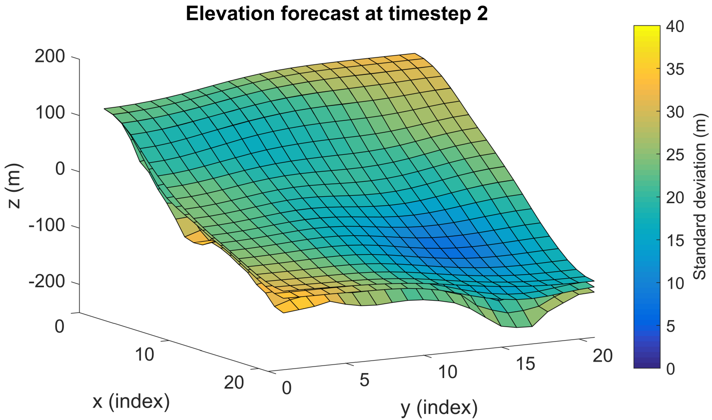
Time evolution of ensemble



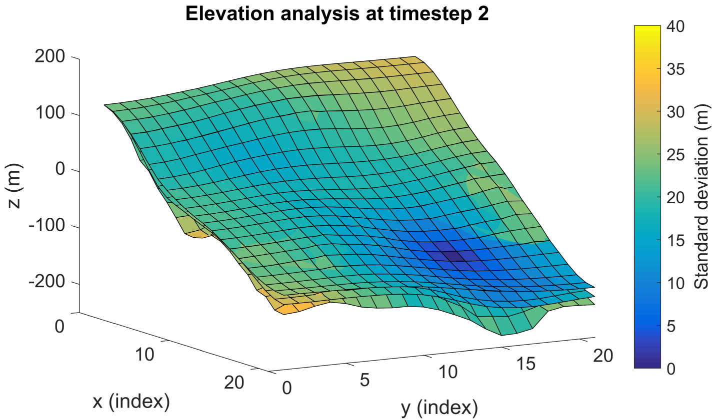
Time evolution of ensemble



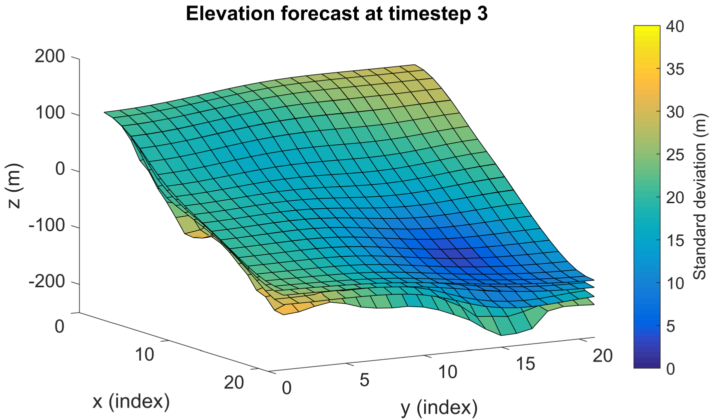
Time evolution of ensemble



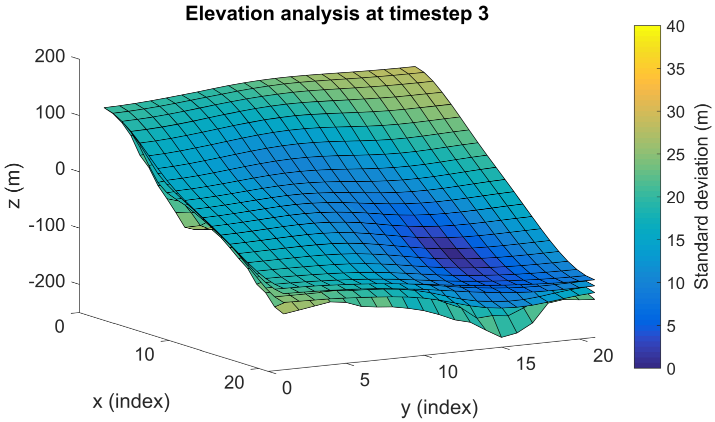
Time evolution of ensemble



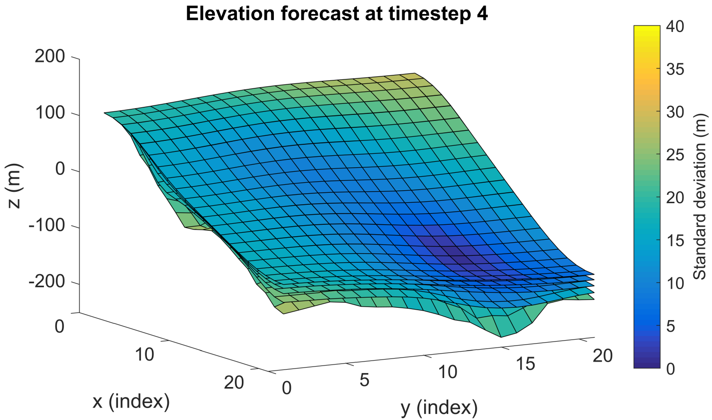
Time evolution of ensemble



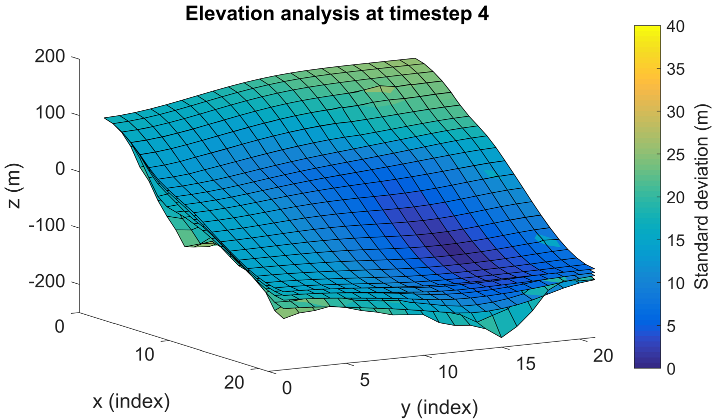
Time evolution of ensemble



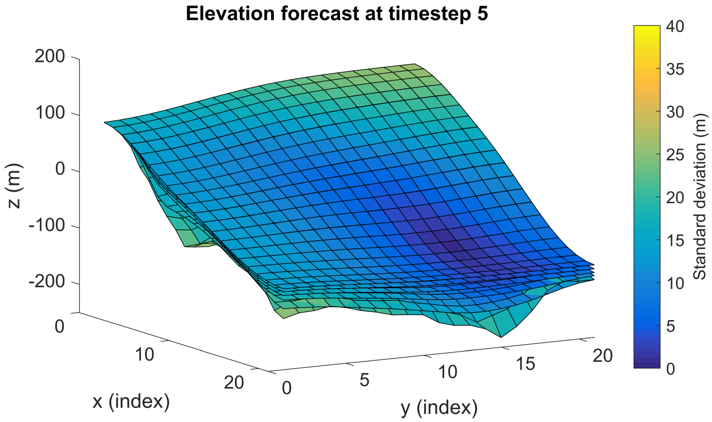
Time evolution of ensemble



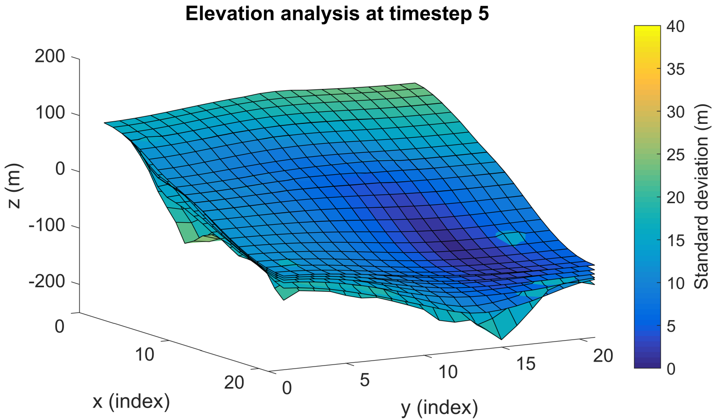
Time evolution of ensemble



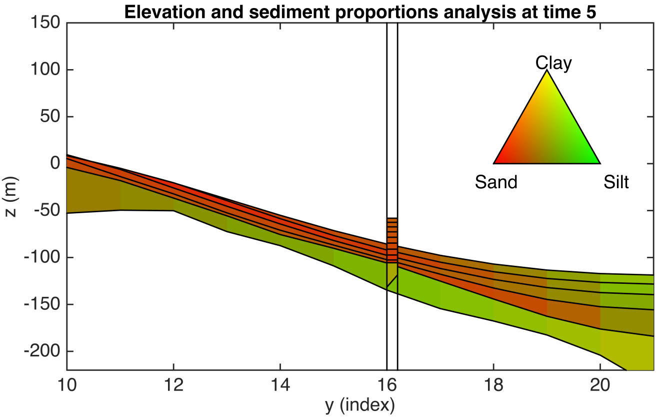
Time evolution of ensemble



Time evolution of ensemble

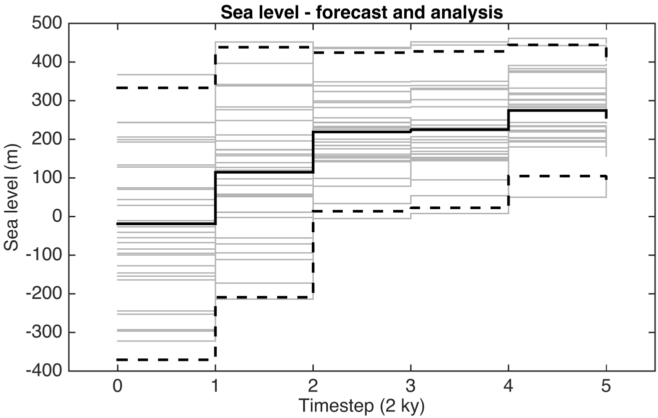


Time evolution of ensemble



Time evolution of ensemble

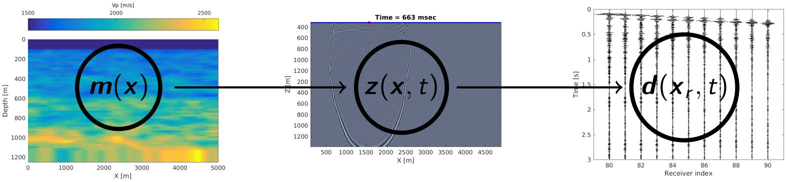
Sea level parameter - constant: $\theta(t) = \theta_0, \; t_{\text{start}} \leq t \leq t_{\text{end}}$



Seismic inversion

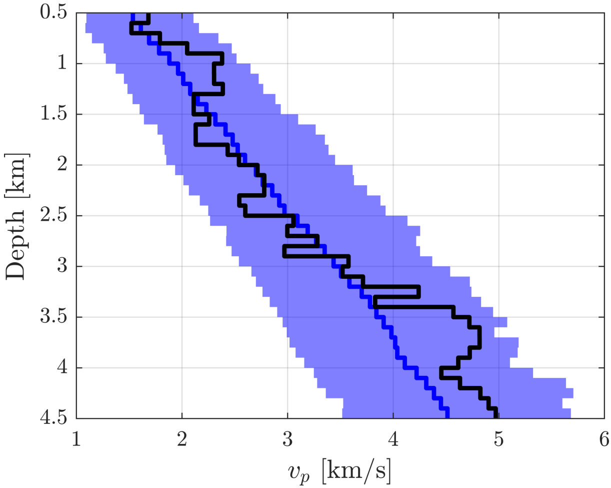
Elastic parameters

Measurements

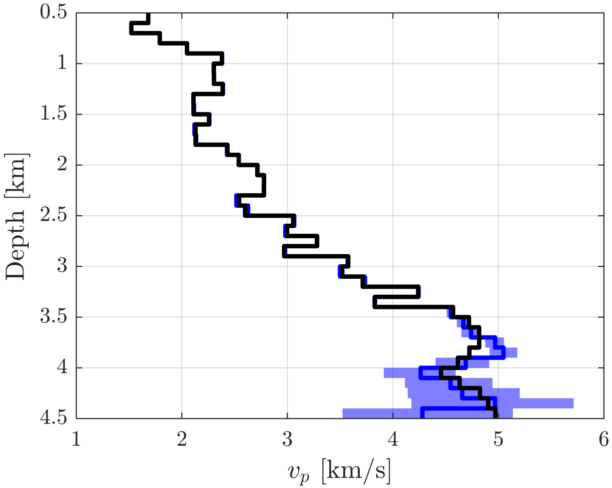


Propagating Wavefield

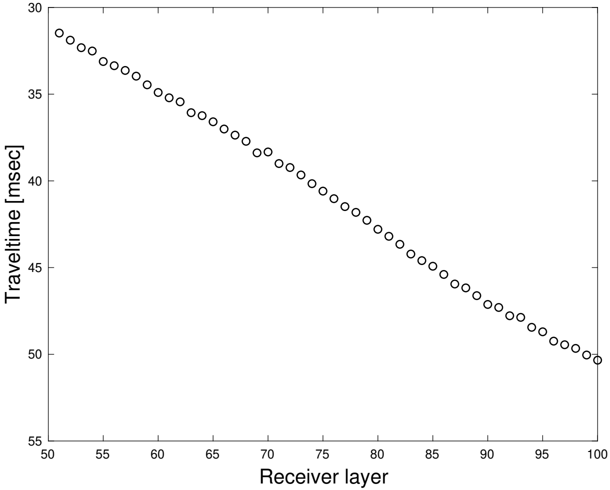
Seismic - prior model



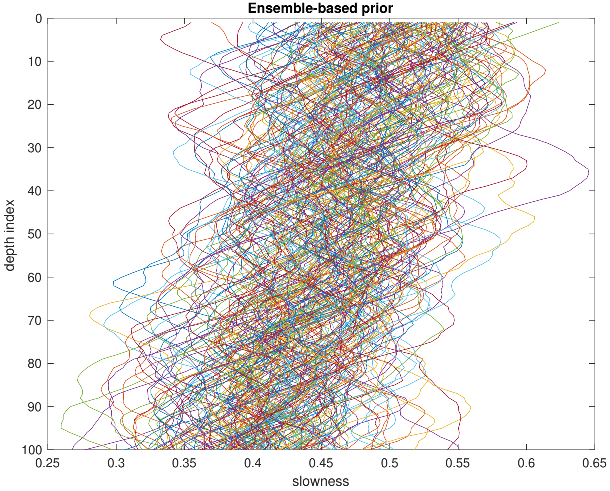
Seismic - posterior model



Exercise : Data



Exercise : Prior ensemble



Exercise : Posterior ensemble