Technical University of Denmark

Online deterministic models: Updating with EnKF

Morten Borup

DTU Environment & Krüger A/S, Veolia VWS Denmark

Morten Grum

Krüger A/S, Veolia VWS Denmark

Peter Steen Mikkelsen

DTU Environment

DTU Environment

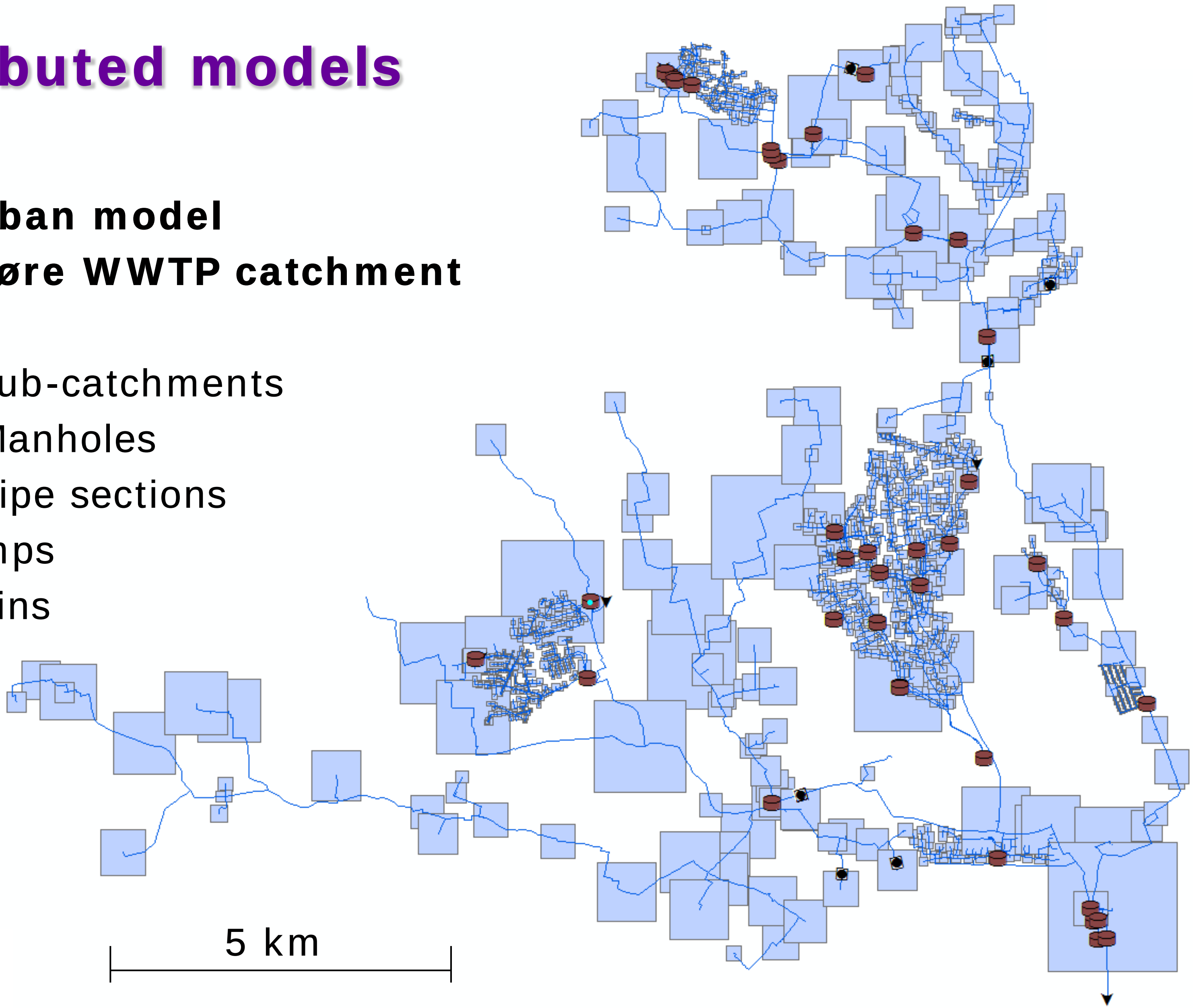
Department of Environmental Engineering

Intelligence Inside

Distributed models

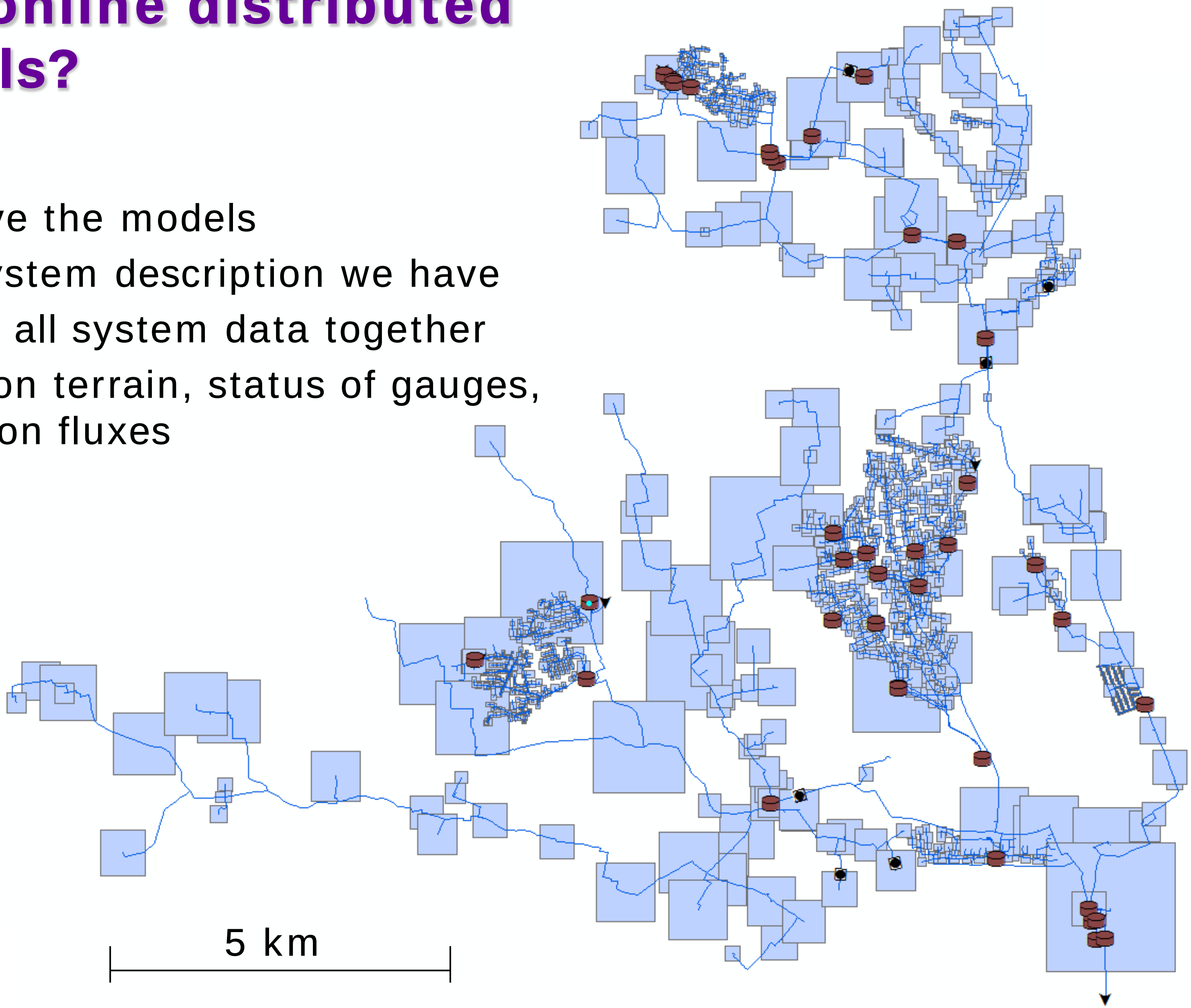
Mike Urban model
of Avedøre WWTP catchment

- 1707 sub-catchments
- 6601 Manholes
- 7749 Pipe sections
- 40 Pumps
- 40 Basins



Why online distributed models?

- We have the models
- Best system description we have
- Can tie all system data together
- Water on terrain, status of gauges, pollution fluxes



Existing DA-metods for MIKE URBAN

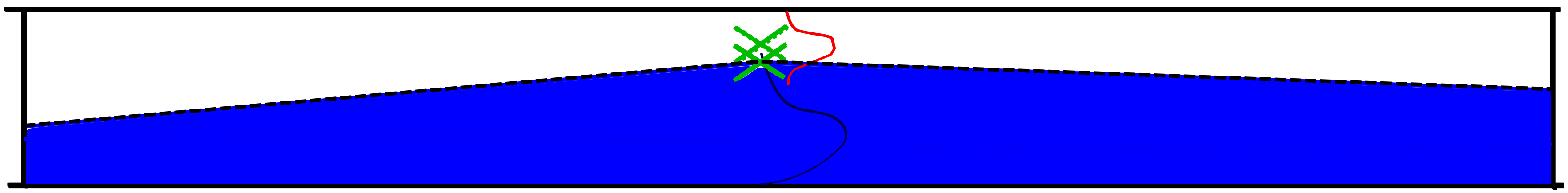
- Deterministic local updating: **MOUSE UPDATE**
  Water levels at single points are corrected

- Deterministic upstream updating: **RDII updating**
  Upstream hydrological lumped conceptual models are updated

- No Hydrodynamic global updating method exists

Optimal updating of distributed models

- Minimizing the error from gauge and model uncertainty
- Correction is distributed from calculated/expected spatial correlation



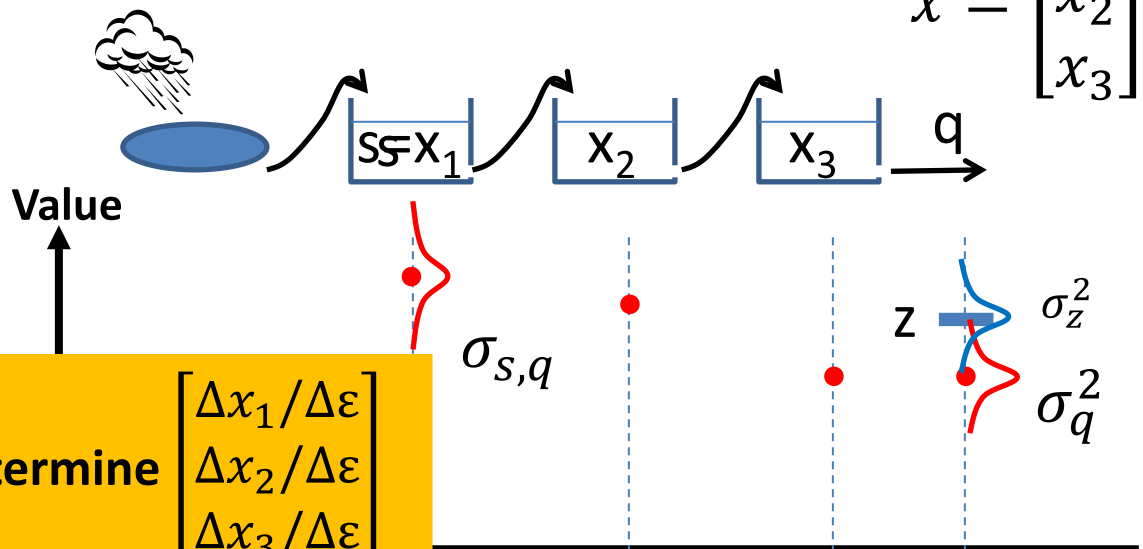
- N states -> NxN crosscorrellations

(States in MIKE URBAN models: 10^4–10^6

-> 10^8–10^{12} crosscorrelations)

Kalman Filter

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



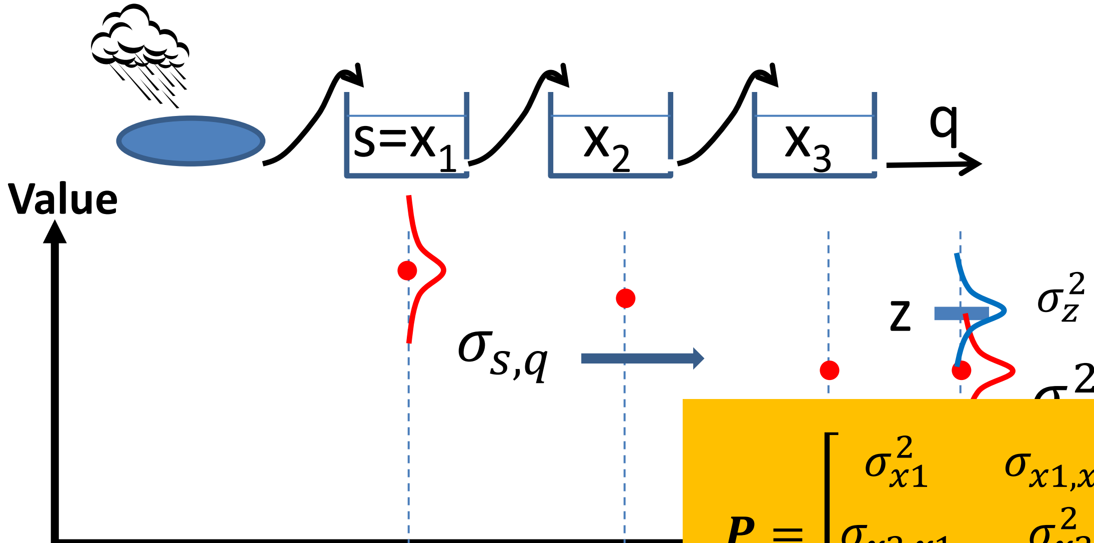
Kalman gain **K** **determine** $\begin{bmatrix} \Delta x_1/\Delta\varepsilon \\ \Delta x_2/\Delta\varepsilon \\ \Delta x_3/\Delta\varepsilon \end{bmatrix}$

Best estimate of s knowing z: $s = \hat{s} + \frac{\sigma_{s,q}}{\sigma_q^2}(\mathrm{z} - \mathrm{q})$

When z is uncertain:

When s is part of vector **x**: $\boldsymbol{x} = \hat{\boldsymbol{x}} + K(\mathrm{z} - \mathrm{q})$

Kalman Filter



$$P = \begin{bmatrix} \sigma_{x1}^2 & \sigma_{x1,x2} & \sigma_{x1,x3} \\ \sigma_{x2,x1} & \sigma_{x2}^2 & \sigma_{x2,x3} \\ \sigma_{x3,x1} & \sigma_{x3,x2} & \sigma_{x3}^2 \end{bmatrix}$$

Error covariance is the cornerstone of KF

Estimated by KF using full error covariance matrix **P**

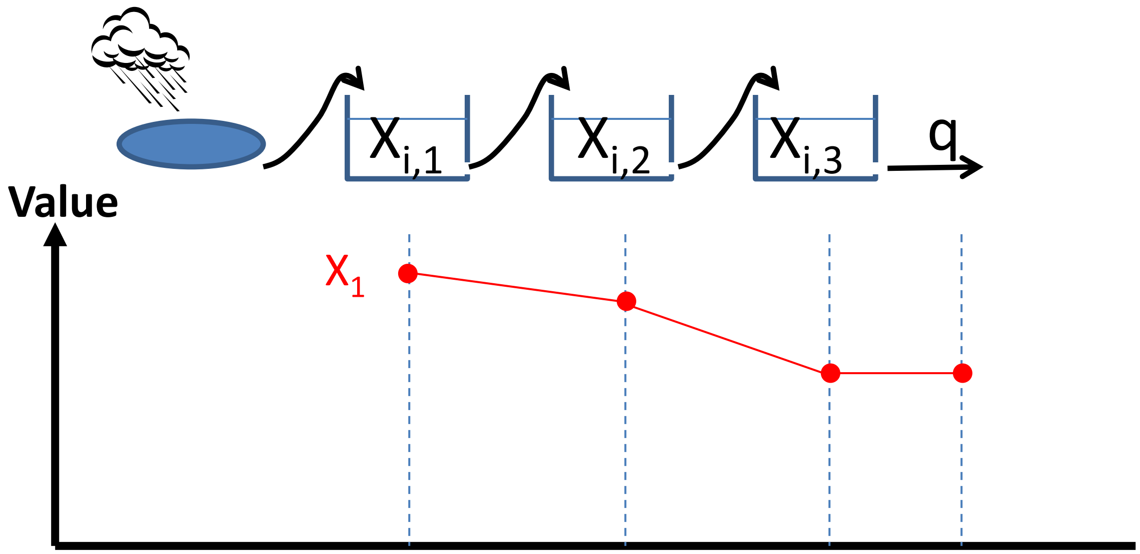
This requires:

1. **Normal distributed** Errors (including any input uncertainty)

2. **Linearity**

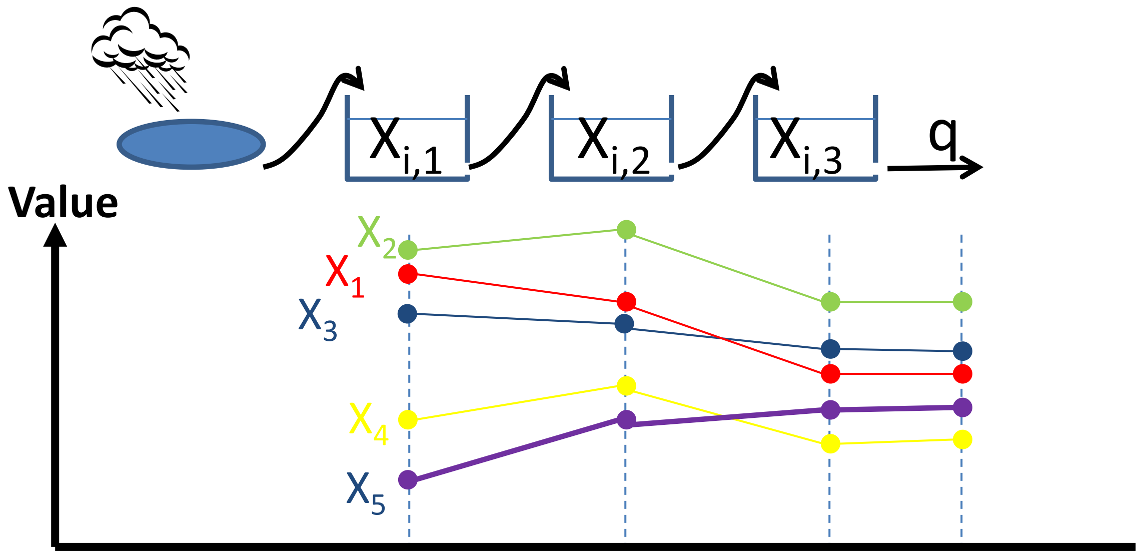
3. **Limited** number of **states** (size of **P** = n*n)

Ensemble **K**alman **F**ilter



Ensembles of models used to represent state uncertainty

Ensemble **K**alman **F**ilter



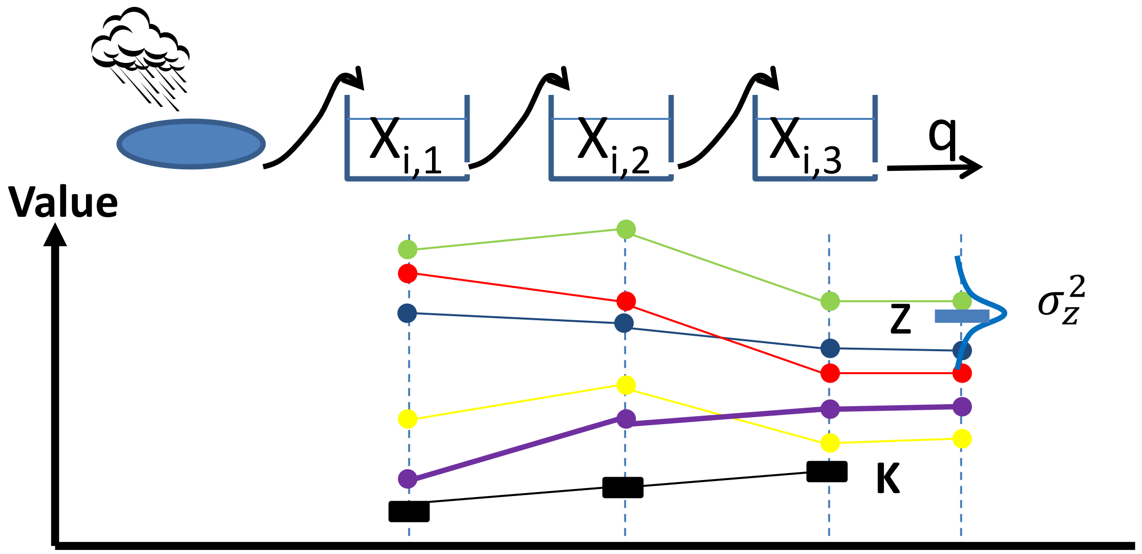
Ensembles of models used to represent state uncertainty

Covariance can be calculated directly from ensemble:

$$\sigma_{1,3} = \frac{1}{N-1}\sum_{i}^{N}(X_{i,1} - \overline{X_{,1}})(X_{i,3} - \overline{X_{,3}})$$

No need for full covariance matrix for finding **K**

Ensemble Kalman Filter

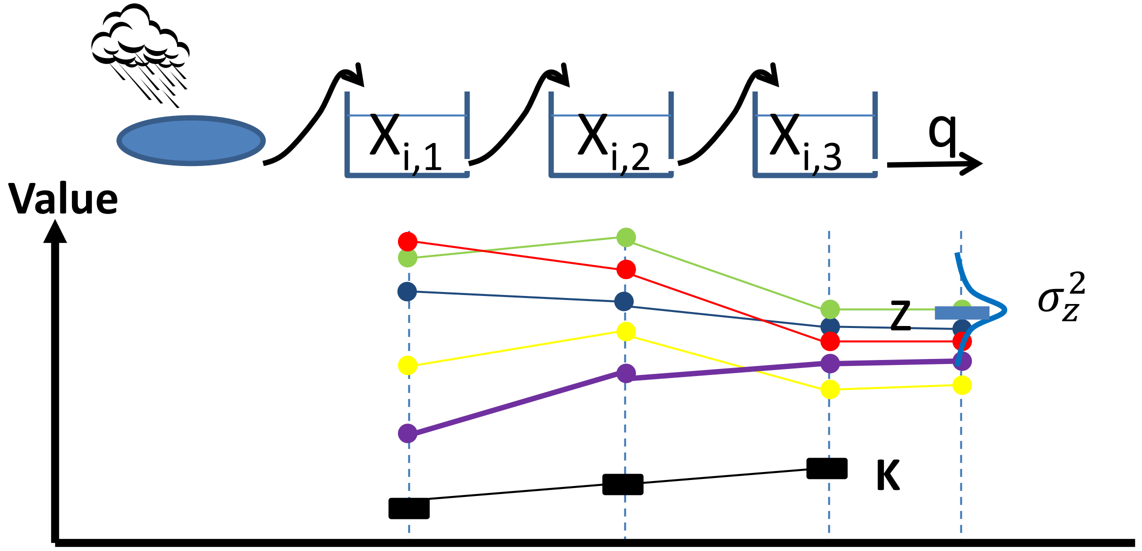


EnKF step-by-step

1.Calculate **K** using the relevant covariances from ensemble

2.Correct each ensemble member using **K**(observed-modelled)

Ensemble **K**alman **F**ilter



EnKF step-by-step

1.Calculate **K** using the relevant covariances from ensemble

2.Correct each ensemble member using **K**(observed-modelled)

3.Run model until next measurements arrives

EnKF vs. KF

- EnKF

  Computational efficient for LARGE models

  Can handle any kind of noise

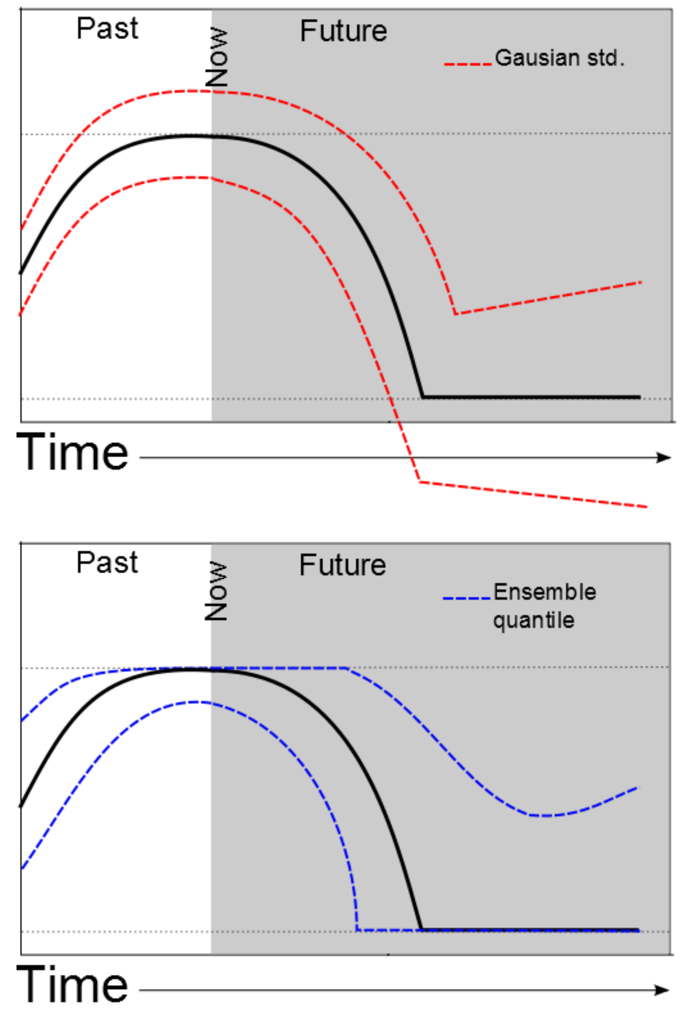
  Handles non-linearity

- KF

  If gauss-linearity then much more efficient for not too big models

EnKF vs. KF

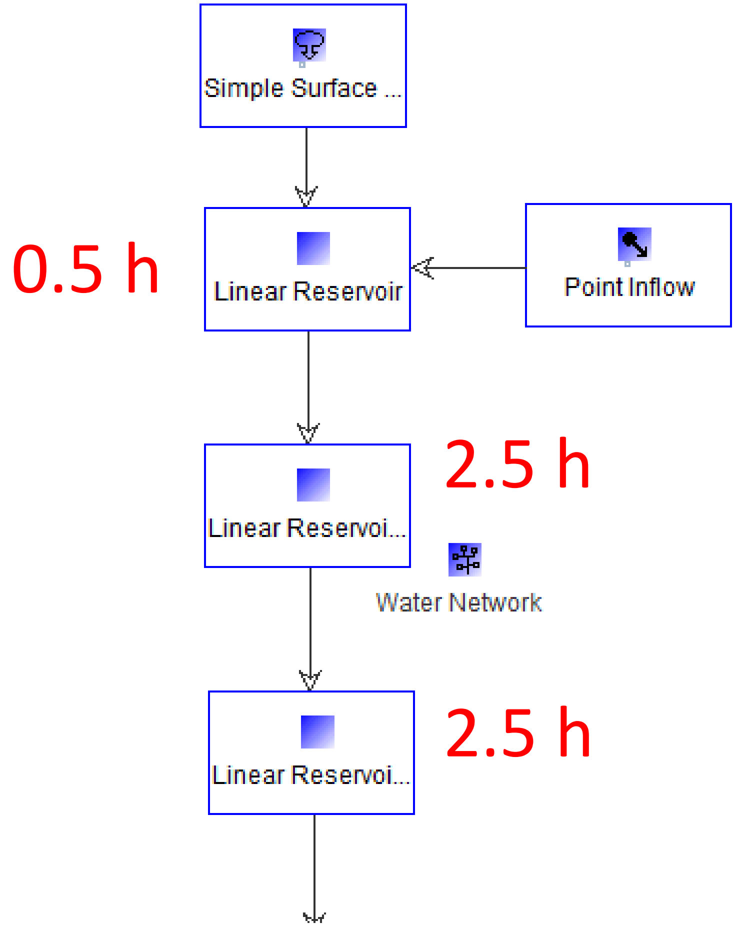
- Uncertainty descriptions



Testing EnKF

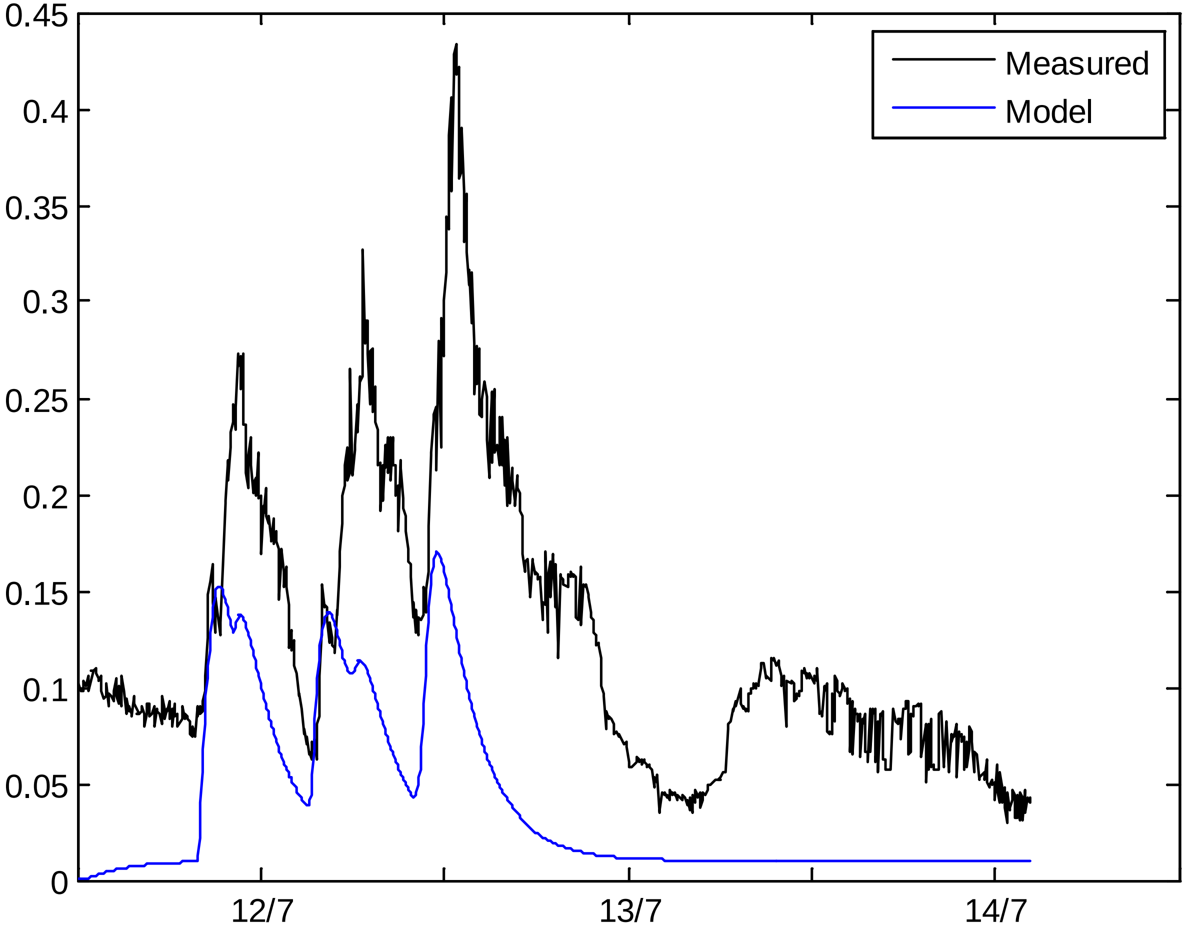
- 2010 flow data from Ballerup

- Simple 3 linear reservoir WA model
- Only noise on first reservoir (state proportional)
- Std. measured flow = 10%



Ballerup 2010

Measured and modelled flow when no update



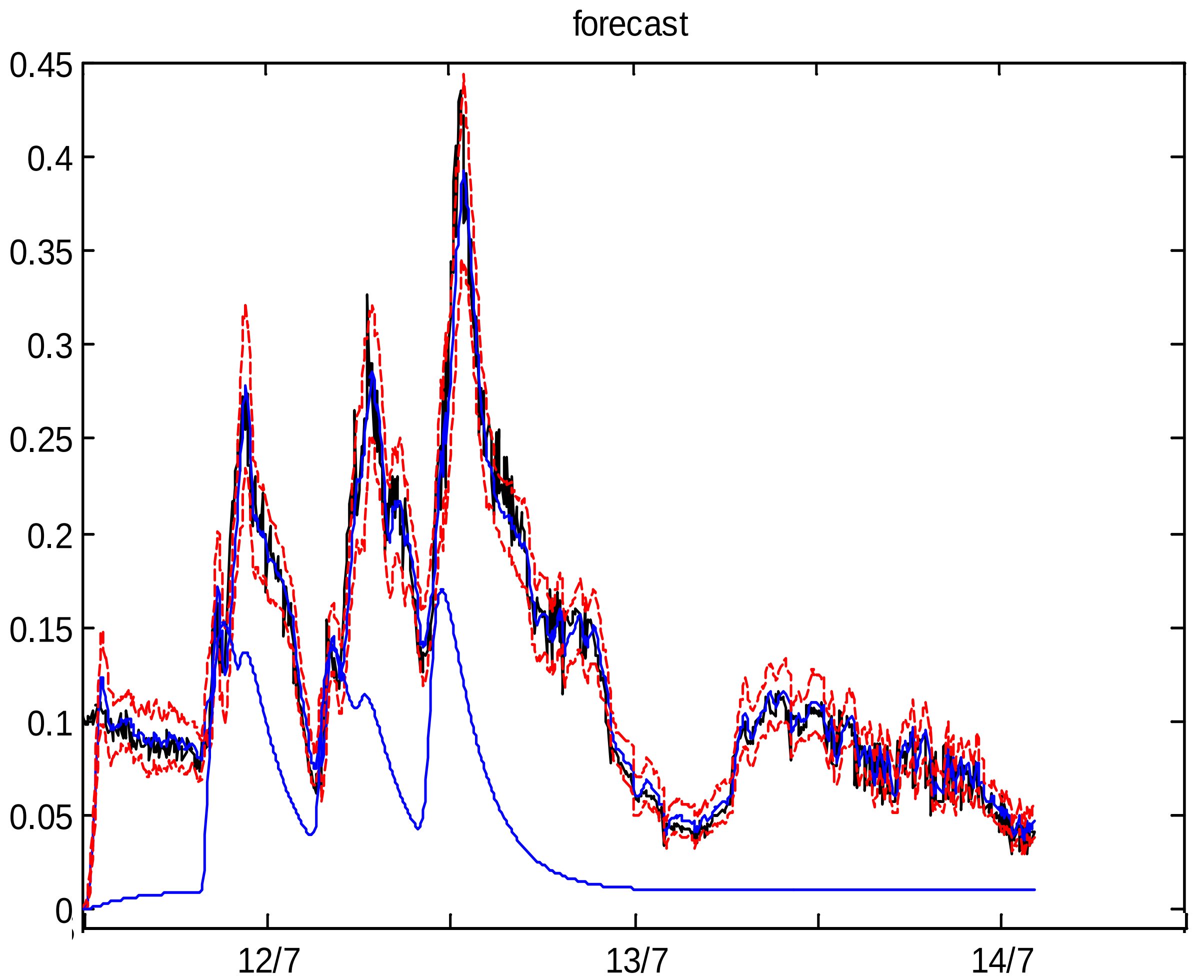
EnKF (N=50)

Measured and modelled flow when no update

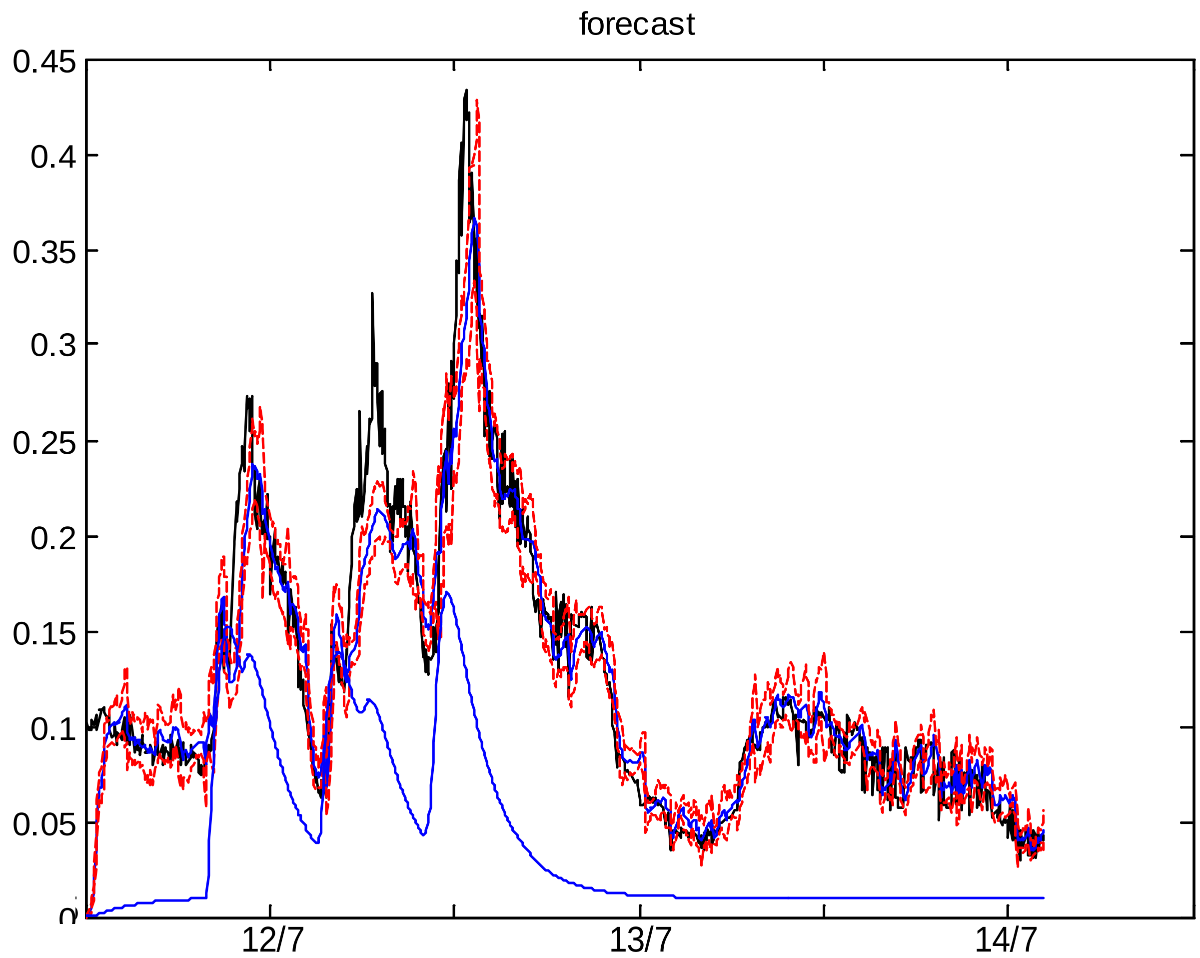
+

One step prediction from using EnKF including 95% prediction interval

Number of ensemble members N = 50



EnKF (N=5)



Measured and modelled flow when no update

+

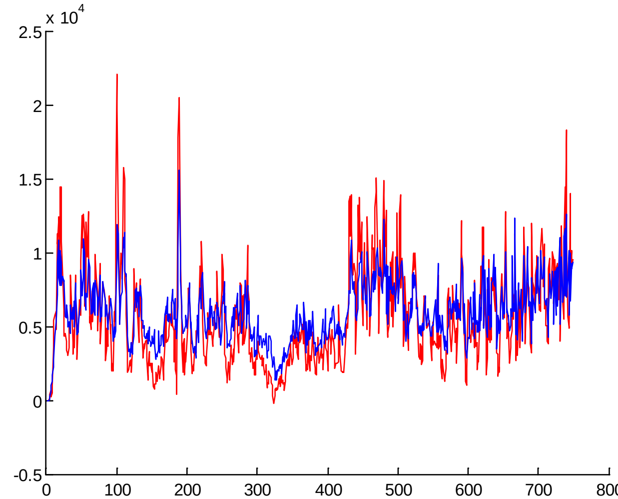
One step prediction from using EnKF including 95% prediction interval

Number of ensemble members N = 5

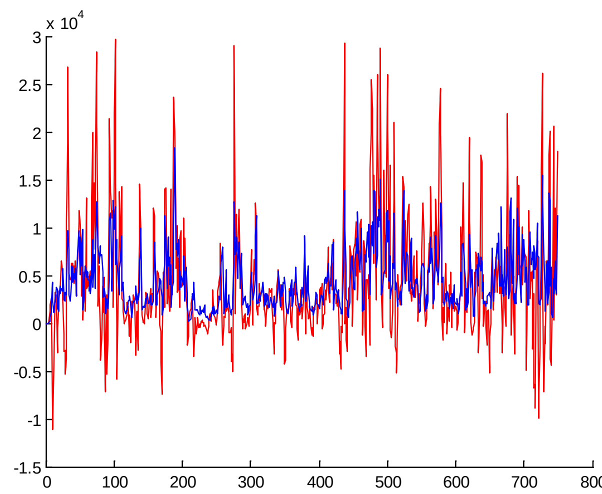
Kalman gain K for different N

K for first (red) and second (blue) reservoirs.

Kalman gain K much more stable when N is high.



N=50



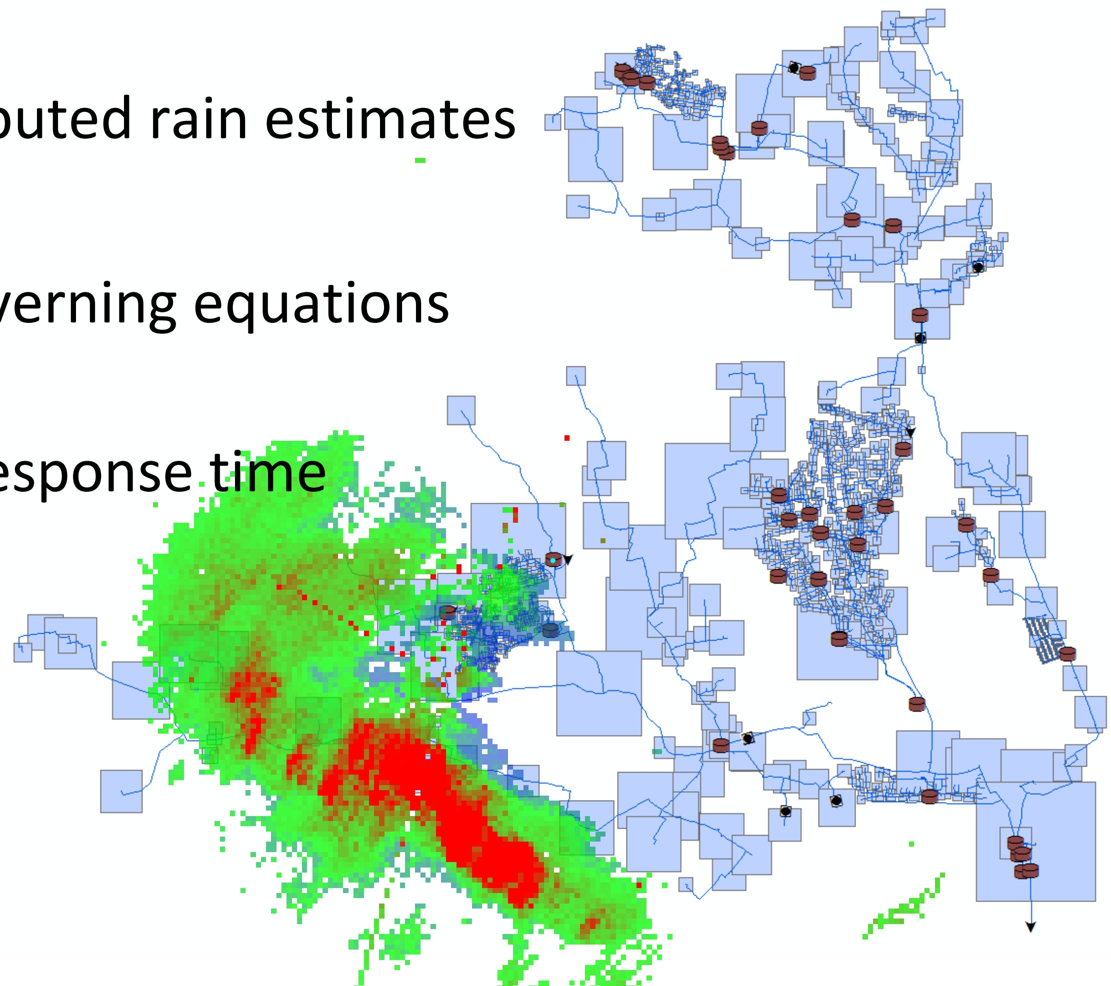
N=5

EnKF and MU

1. Distributed models tend to be big
2. And non-linear
3. Ensembles of distributed rain estimates

Challenges

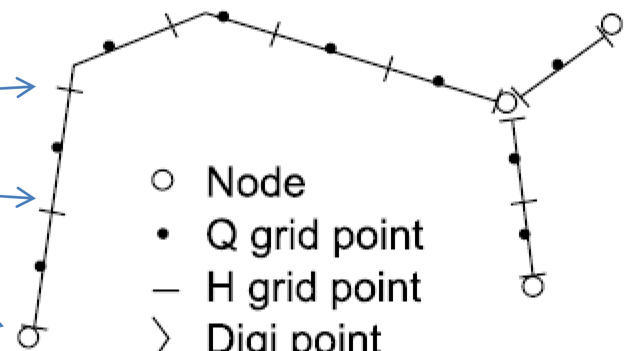
1. Manipulation of governing equations (St. Venant eq.)
2. Slow hydrological response time



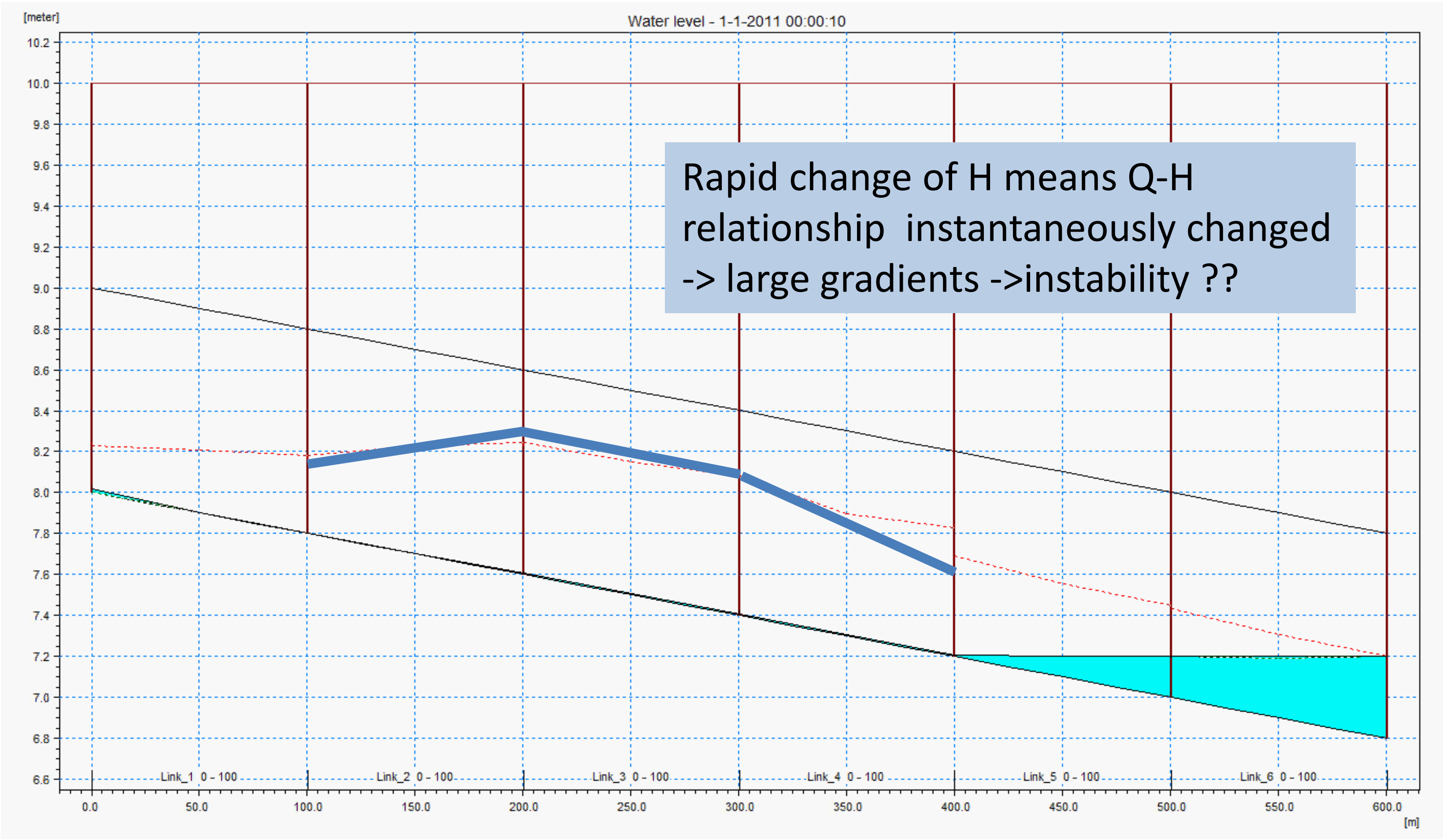
Updating MIKE URBAN

- New computational engine (Mike1D)
- Interface to states (h and Q)

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ .. \end{bmatrix}$$



Example of water level only update in MIKE URBAN



Summary

Ensemble based data assimilation methods are suitable for urban runoff

The states in the new MIKE URBAN hydrodynamic engine can be manipulated without instability

HD data assimilation for MIKE URBAN is on the way

Questions