# Groundwater flow modeling: uncertain boundary conditions and their impact on the forecast

Emanuel Huber[1], Jef Caers[1], Peter Huggenberger[2], Thomas Vienken[3], Manuel Kreck[3], Peter Dietrich[3,4]

[1]Department of Geological Sciences, Stanford University
[2]Applied and Environmental Geology, University of Basel
[3]UFZ-Helmholtz Centre for Environmental Research, Leipzig
[4]Eberhard Karls University of Tübingen

geoENV 2016

# Context

## Subsurface flow

An interplay between

- subsurface heterogeneity
- boundary conditions (inflow/outflow, pressure)

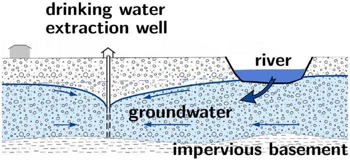
# Context

## Subsurface flow

An interplay between

- subsurface heterogeneity
- boundary conditions (inflow/outflow, pressure)

## Bacterial contamination of groundwater




High discharge events in the river

- $\nearrow$ bacteria concentration in the river
- $\nearrow$ infiltration of river water into the aquifer
- $\nearrow$ possible contamination of groundwater water

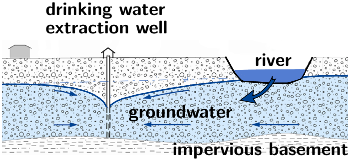
# Context

## Subsurface flow

An interplay between

- subsurface heterogeneity
- boundary conditions (inflow/outflow, pressure)

## Bacterial contamination of groundwater




High discharge events in the river

- $\nearrow$ bacteria concentration in the river
- $\nearrow$ infiltration of river water into the aquifer
- $\nearrow$ possible contamination of groundwater water

☞ Shall the drinking water extraction well be turned off?

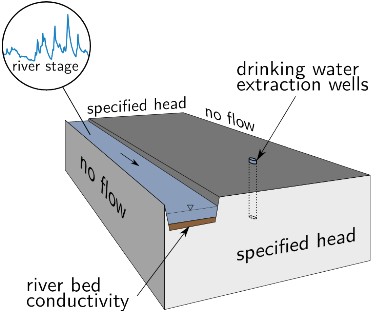
# Objectives

- Uncertainty model for boundary conditions
- Uncertainty quantification of forecast

# Objectives

- Uncertainty model for boundary conditions
- Uncertainty quantification of forecast

## Synthetic case study




Drinking water extraction in one well

River water–groundwater interaction
$\rightarrow$ infiltration $q = C_{\mathrm{riv}}\,(h_{\mathrm{riv}} - h_{\mathrm{gw}})$

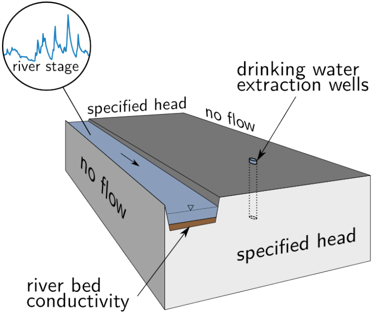
Specified head boundary conditions

Mechanical bacteria filtration in aquifer
$\rightarrow$ exponential decay of $\lambda \cdot s$

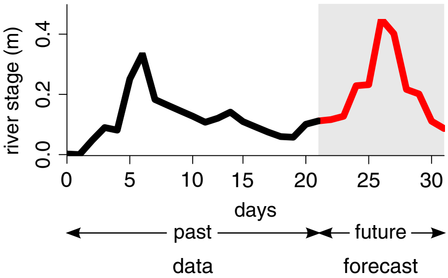
# Objectives

- Uncertainty model for boundary conditions
- Uncertainty quantification of forecast

## Synthetic case study



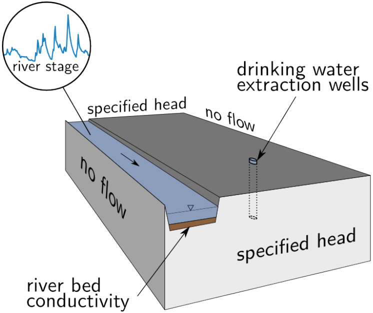

**Prediction** = bacteria concentration for the next 10 days

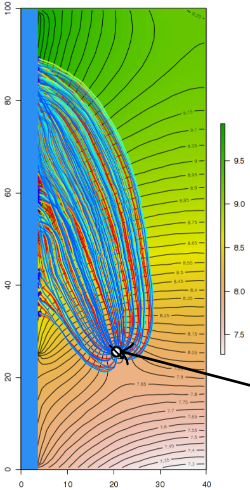


# Objectives

- Uncertainty model for boundary conditions
- Uncertainty quantification of forecast

## Synthetic case study




streamlines
= bacteria pathways

known bacteria
concentration in river

extraction well

# Uncertain boundary conditions

## River water–groundwater interaction

- river bed conductivity $C_{\text{riv}} \rightarrow$ spatially constant, uniform prior.
- river heads ($h_{\text{riv}}$) $\rightarrow$ measured (assumed to be known everywhere)

# Uncertain boundary conditions

## River water–groundwater interaction

- river bed conductivity $C_{\mathrm{riv}} \rightarrow$ spatially constant, uniform prior.
- river heads ($h_{\mathrm{riv}}$) $\rightarrow$ measured (assumed to be known everywhere)

## Specified heads

Interpolation: space-time Gaussian process conditioned to measured groundwater heads

- Matérn (space) and Gaussian (time) covariance functions
- (unknown) linear mean function
- no-flow boundary conditions (specified derivatives)

# Uncertain boundary conditions

## River water–groundwater interaction

- river bed conductivity $C_{\text{riv}} \rightarrow$ spatially constant, uniform prior.
- river heads ($h_{\text{riv}}$) $\rightarrow$ measured (assumed to be known everywhere)

## Specified heads

Interpolation: space-time Gaussian process conditioned to measured groundwater heads

- Matérn (space) and Gaussian (time) covariance functions
- (unknown) linear mean function
- no-flow boundary conditions (specified derivatives)

But river and groundwater heads
are unknown for the next 10 days!

# Uncertain future boundary conditions

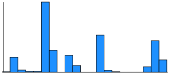
## Convolution ($\star$) model

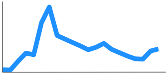
Strong relationship between precipitation and river heads

precipitations

river heads

PAST

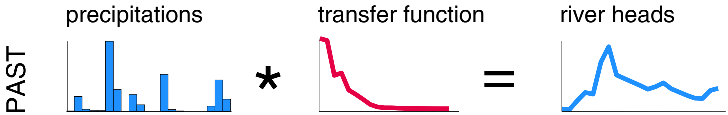




# Uncertain future boundary conditions

## Convolution ($\star$) model

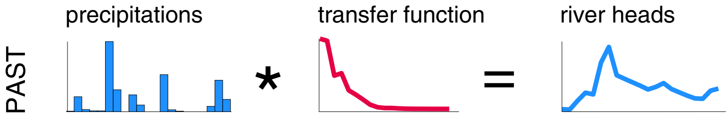
Strong relationship between precipitation and river heads



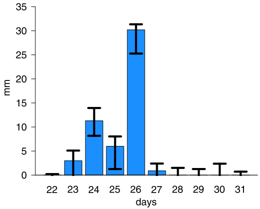
# Uncertain future boundary conditions

## Convolution ($\star$) model

Strong relationship between precipitation and river heads




Uncertain weather forecast $\rightarrow$ random sampling



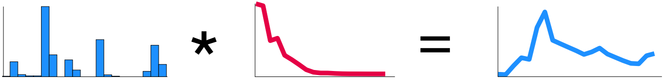
# Uncertain future boundary conditions

## Convolution ($\star$) model

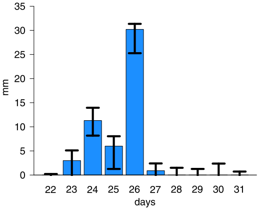
Strong relationship between precipitation and river heads

precipitations transfer function river heads

PAST

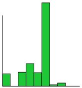


Uncertain weather forecast $\rightarrow$ random sampling




FUTURE

precipitations



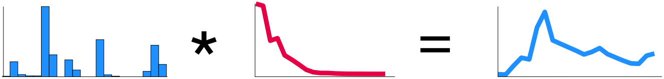
# Uncertain future boundary conditions

## Convolution (⋆) model

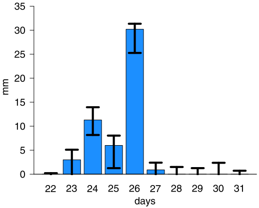
Strong relationship between precipitation and river heads

precipitations transfer function river heads

PAST

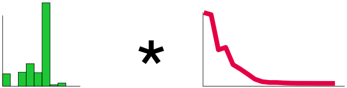



Uncertain weather forecast → random sampling




FUTURE

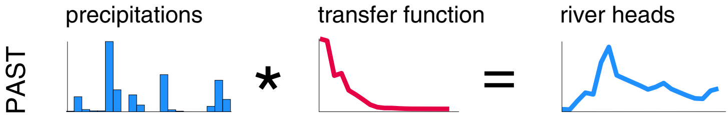
precipitations transfer function



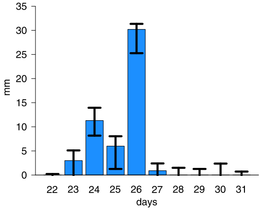
# Uncertain future boundary conditions

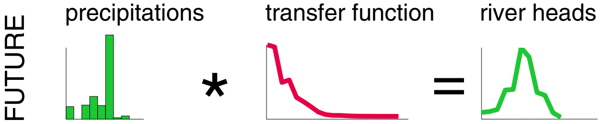
## Convolution ($\star$) model

Strong relationship between precipitation and river heads




Uncertain weather forecast $\rightarrow$ random sampling

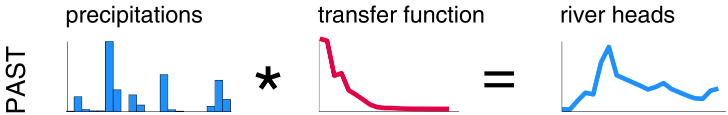


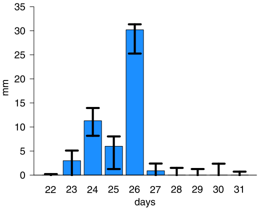
# Uncertain future boundary conditions

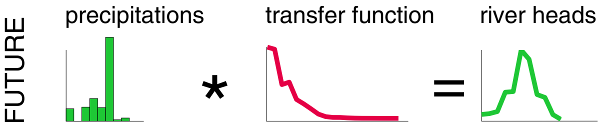
## Convolution ($\star$) model

Strong relationship between precipitation and river heads




Uncertain weather forecast $\rightarrow$ random sampling




☞ Same approach for the groundwater heads

# Spatial uncertainty

## Subsurface heterogeneity uncertainty

Hydraulic conductivity
$\rightarrow$ Gaussian random field (Matérn covariance function)

Porosity, specific storage, specific yield
$\rightarrow$ spatially constant, uniform prior

# Impact of uncertain boundary conditions

## Distance-based general sensitivity analysis

1. Sample 1000 realisations and compute the model response
2. Cluster the realisations based on the model response
3. For each parameter: compare the parameter distribution in each cluster with the global parameter distribution
4. If the distance between both distributions is significant, then the parameter is influential (on the model response)

# Impact of uncertain boundary conditions

## Distance-based general sensitivity analysis

1. Sample 1000 realisations and compute the model response
2. Cluster the realisations based on the model response
3. For each parameter: compare the parameter distribution in each cluster with the global parameter distribution
4. If the distance between both distributions is significant, then the parameter is influential (on the model response)

# Impact of uncertain boundary conditions

## Distance-based general sensitivity analysis

1. Sample 1000 realisations and compute the model response
2. Cluster the realisations based on the model response
3. For each parameter: compare the parameter distribution in each cluster with the global parameter distribution
4. If the distance between both distributions is significant, then the parameter is influential (on the model response)

# Impact of uncertain boundary conditions

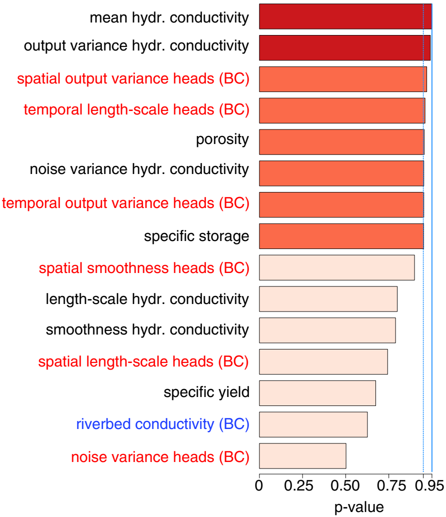
## Distance-based general sensitivity analysis

1. Sample 1000 realisations and compute the model response
2. Cluster the realisations based on the model response
3. For each parameter: compare the parameter distribution in each cluster with the global parameter distribution
4. If the distance between both distributions is significant, then the parameter is influential (on the model response)

# Impact of uncertain boundary conditions

## Distance-based general sensitivity analysis

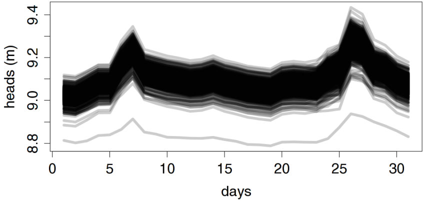
1. Sample 1000 realisations and compute the model response
2. Cluster the realisations based on the model response
3. For each parameter: compare the parameter distribution in each cluster with the global parameter distribution
4. If the distance between both distributions is significant, then the parameter is influential (on the model response)

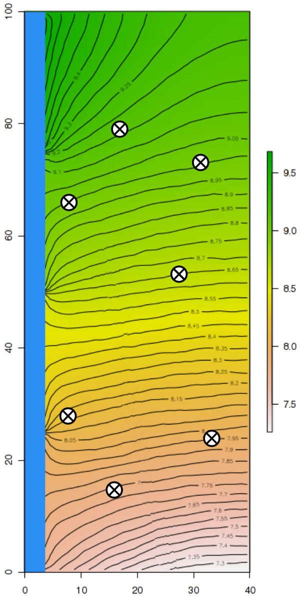


# Statistical prediction

## Data (1000 realisations)

groundwater heads (day 1 - day 31)
7 wells $\times$ 31 time steps

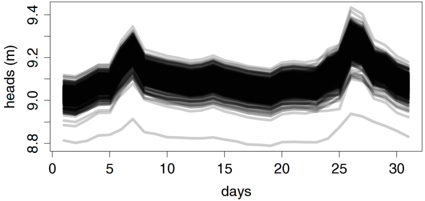


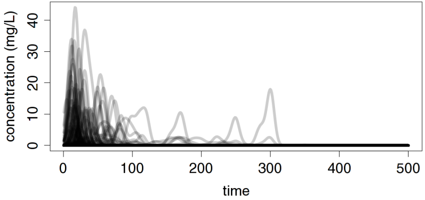
# Statistical prediction

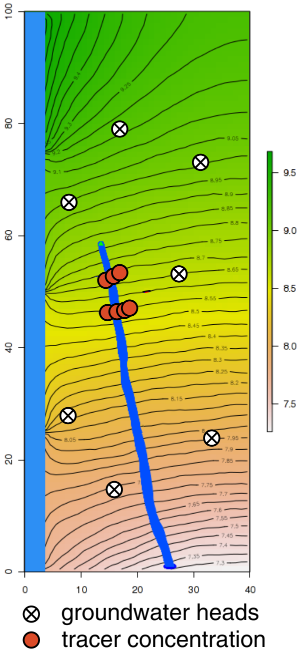
## Data (1000 realisations)

groundwater heads (day 1 - day 31)
7 wells $\times$ 31 time steps




tracer concentration (day 1 - day 5)
7 wells $\times$ 500 time steps

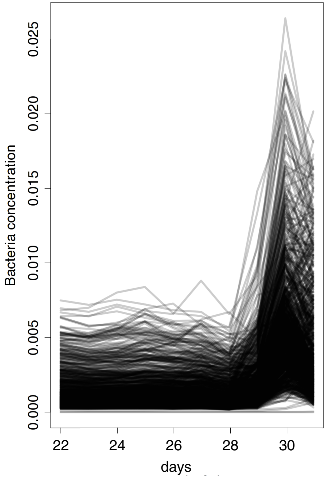


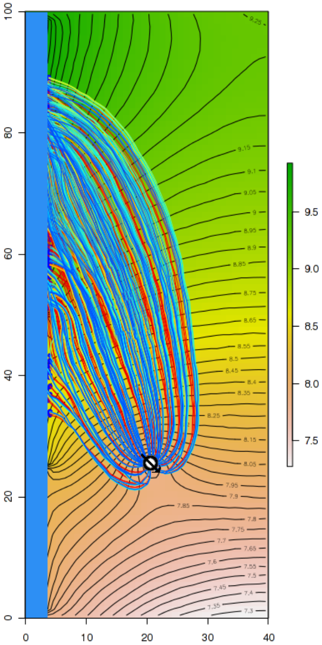



# Statistical prediction

## Forecast (1000 realisations)

bacteria concentration (day 1 - day 31)
1 wells × 10 time steps



# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

Time-series $d(t)$ approximated by a linear combination of B-splines $\Phi_j(t)$
$d(t) \approx \sum_j c_j \Phi_j(t)$

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

functional version of PCA
summations change into integrations

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

$\Downarrow$

merge scores

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

$\Downarrow$

merge scores

$\Downarrow$

PCA

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

$\Downarrow$

merge scores

$\Downarrow$

PCA

PCA

$\Uparrow$

**prediction**

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

$\Downarrow$

merge scores

$\Downarrow$

PCA

$\Downarrow$

$\boxed{\text{CCA}}$

$\Uparrow$

PCA

$\Uparrow$

**prediction**

find two bases **A** and **B** in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized.

$$\mathbf{U} = \mathbf{X}\mathbf{A}$$
$$\mathbf{V} = \mathbf{Y}\mathbf{B}$$

with
$(\mathbf{U}, \mathbf{V})$ = canonical variables
$\mathbf{X}$ = reduced data variable
$\mathbf{Y}$ = reduced forecast variable

# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

⇓

basis function expansion

⇓

functional PCA

⇓

merge scores

⇓

PCA

⇓

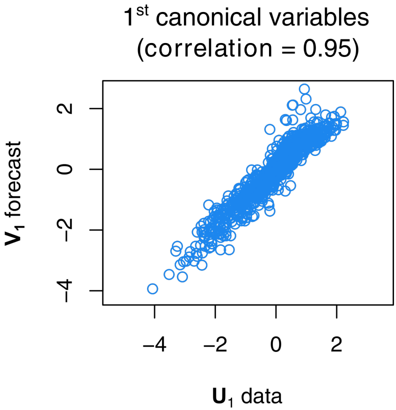
CCA

⇑

PCA

⇑

**prediction**



# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

⇓

basis function expansion

⇓

functional PCA

⇓

merge scores

⇓

PCA

⇓

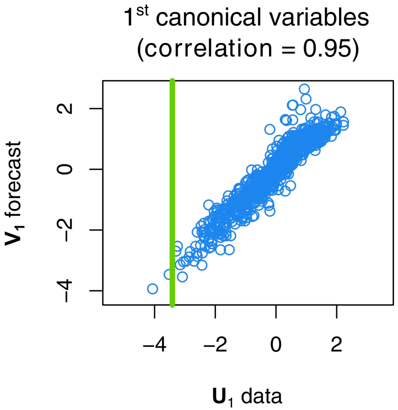
$\boxed{\text{CCA}}$

⇑

PCA

⇑

**prediction**



# Statistical prediction – Method

## Dimension reduction + CCA + linear regression + backtransform

**data**

$\Downarrow$

basis function expansion

$\Downarrow$

functional PCA

$\Downarrow$

merge scores

$\Downarrow$

PCA

$\Downarrow$

$\boxed{\text{CCA}}$

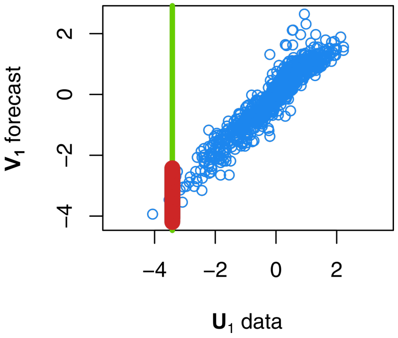
$\Uparrow$

PCA

$\Uparrow$

**prediction**

1$^{st}$ canonical variables
(correlation = 0.95)

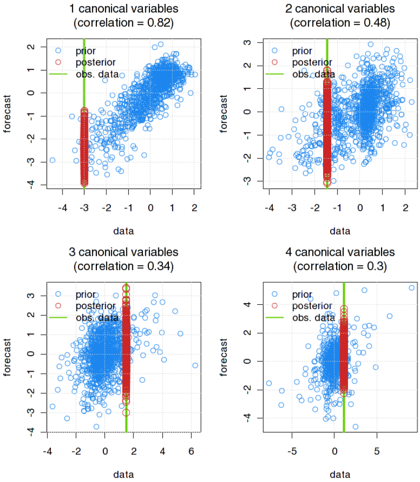


# Statistical prediction – Results

## Canonical correlation space

### tracer data (99.9% of variance)

data (1000 $\times$ 64), forecast (1000 $\times$ 7)

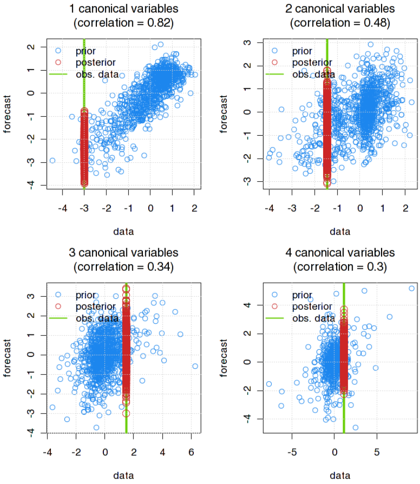


# Statistical prediction – Results

## Canonical correlation space

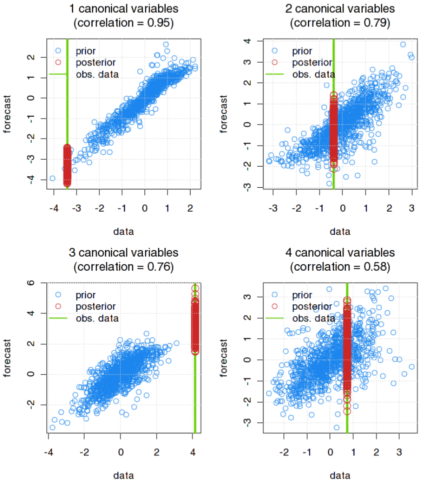
### tracer data (99.9% of variance)

data (1000 $\times$ 64), forecast (1000 $\times$ 7)




### head data (99.9% of variance)

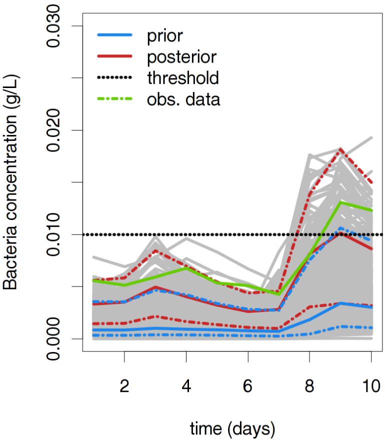
data (1000 $\times$ 16), forecast (1000 $\times$ 7)



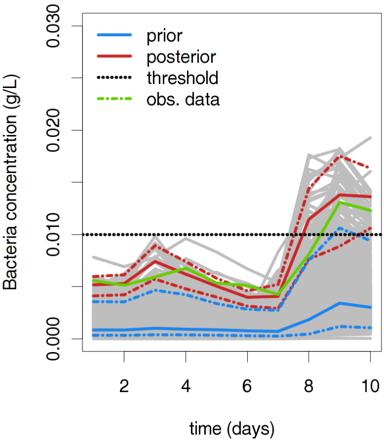
# Statistical prediction – Results

## Prediction

tracer data




head data

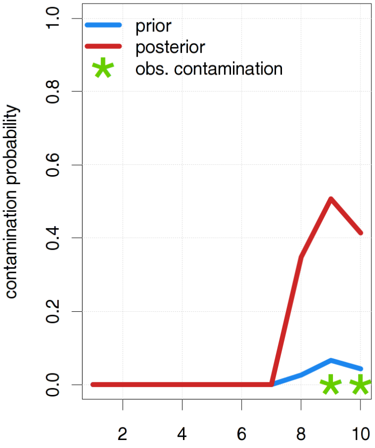



More accurate results with head data (larger log predictive density)

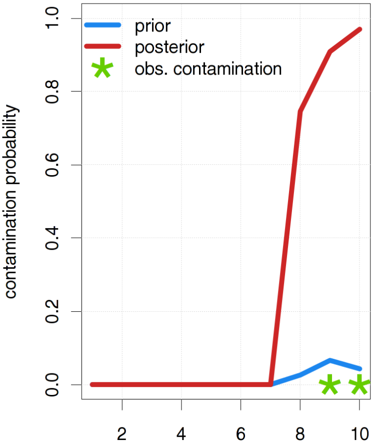
# Statistical prediction – Results

## Decision

tracer data




head data




☞ Turn off the drinking water extraction well!

# Conclusion

- model of uncertain boundary conditions
- distance-based general sensitivity analysis
  $\rightarrow$ importance of boundary conditions (specified heads)
- statistical prediction
  $\rightarrow$ circumvent classical inversion
  $\rightarrow$ relevance of data for prediction

# Conclusion

- model of uncertain boundary conditions
- distance-based general sensitivity analysis
  $\rightarrow$ importance of boundary conditions (specified heads)
- statistical prediction
  $\rightarrow$ circumvent classical inversion
  $\rightarrow$ relevance of data for prediction

## Further research

- model for riverbed conductivity (spatial and temporal)
- statistical prediction for designing monitoring network
- use statistical prediction for resampling