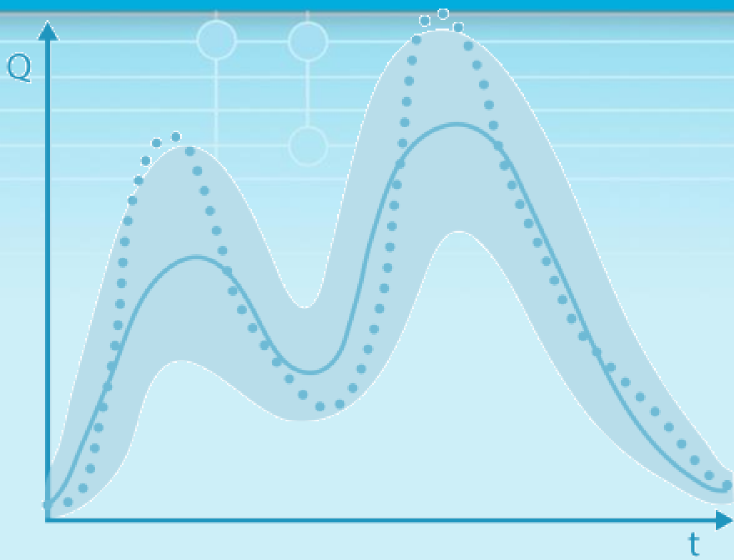
How can statistics help us to get reliable predictions despite model bias?

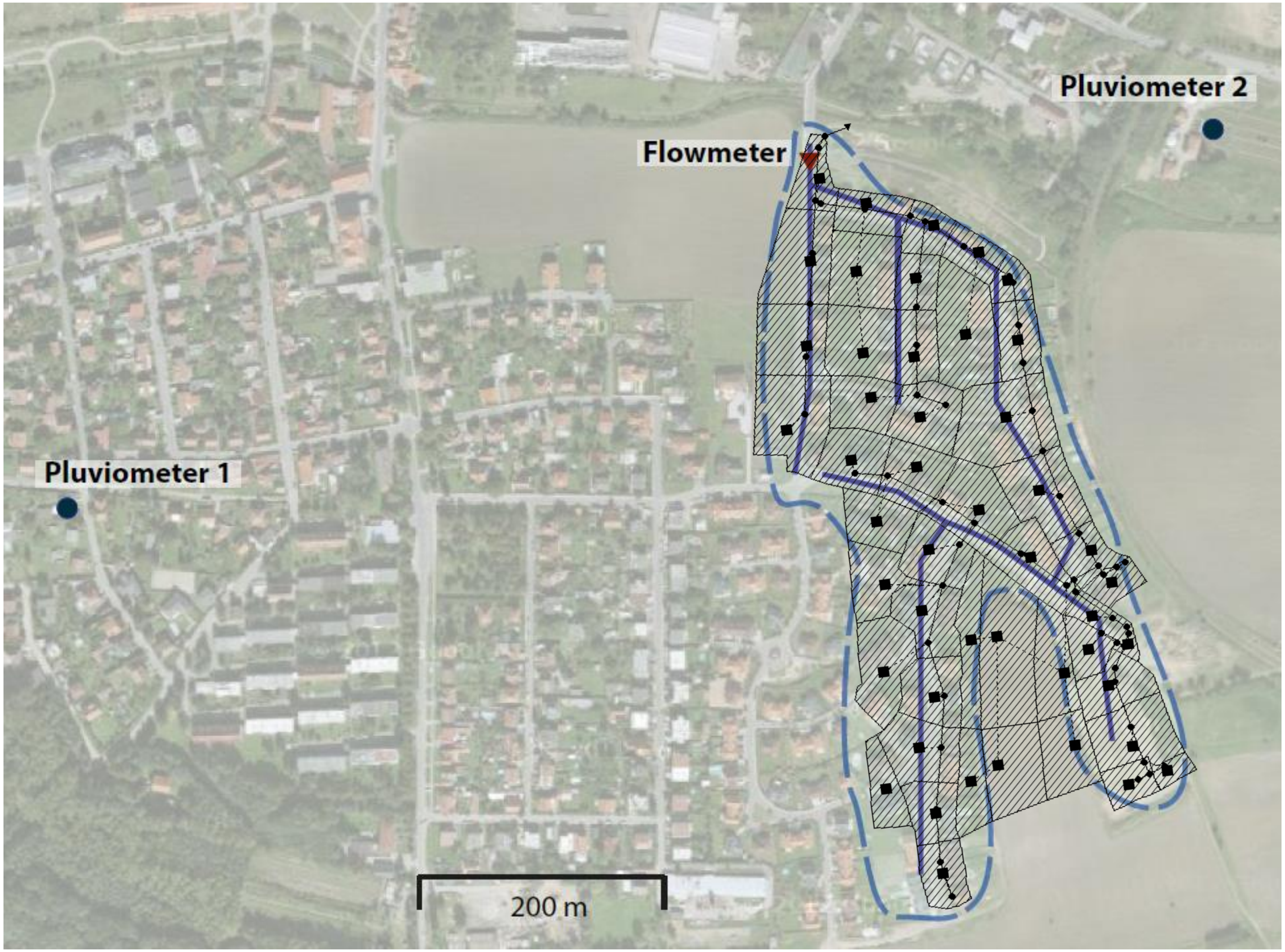


Dario Del Giudice
J. Rieckermann
A. Scheidegger
P. Reichert
C. Albert

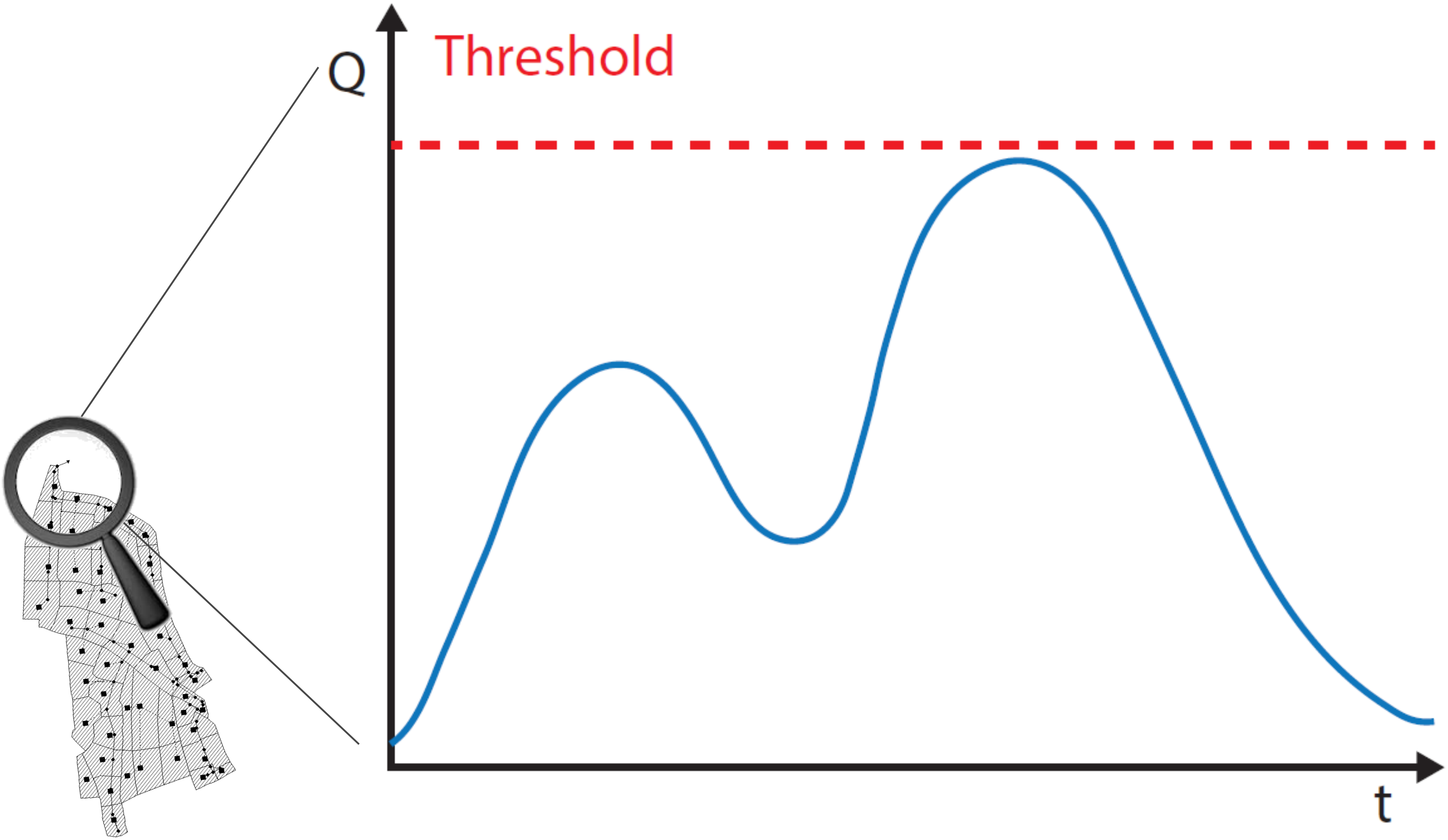
ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

eawag aquatic research

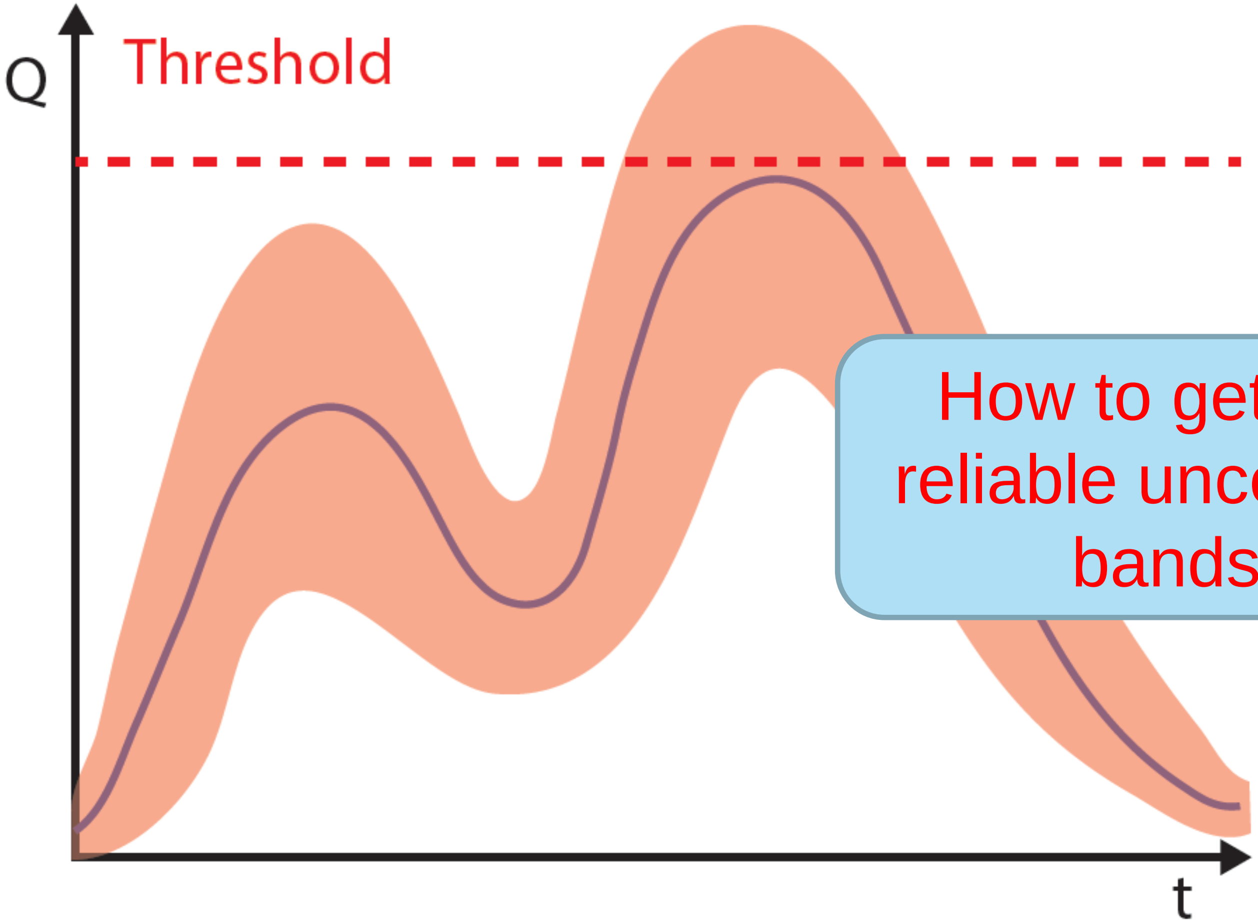
Our typical situation



We want to predict the output

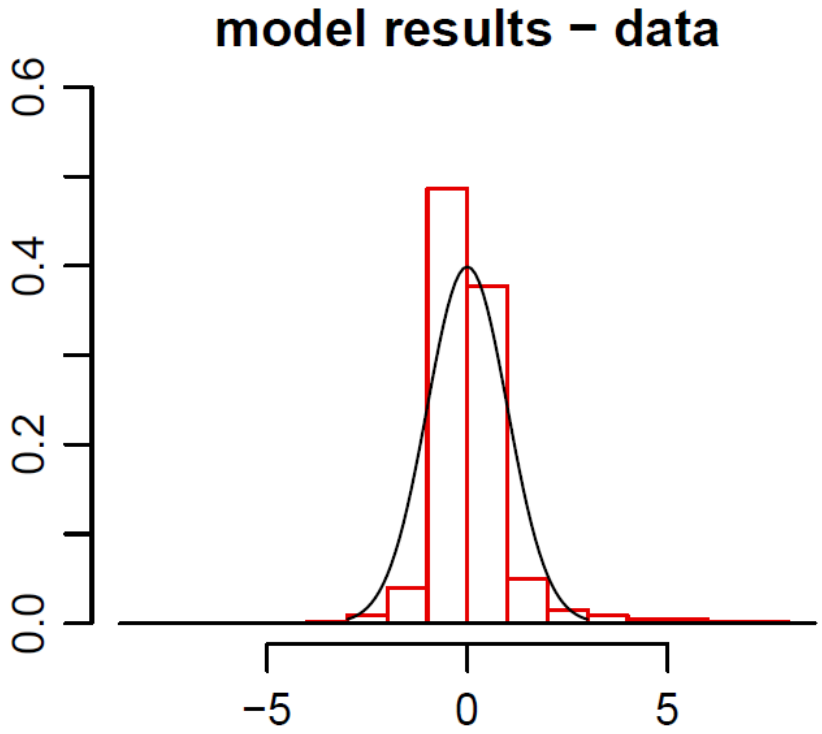


A line is not enough…



How to get such reliable uncertainty bands?

Statistics can come in handy for realistic predictions



Likelihood function

$$f(\boldsymbol{y}_o \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{x}) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det(\boldsymbol{\Sigma}(\boldsymbol{\psi}, \boldsymbol{x}))}}$$

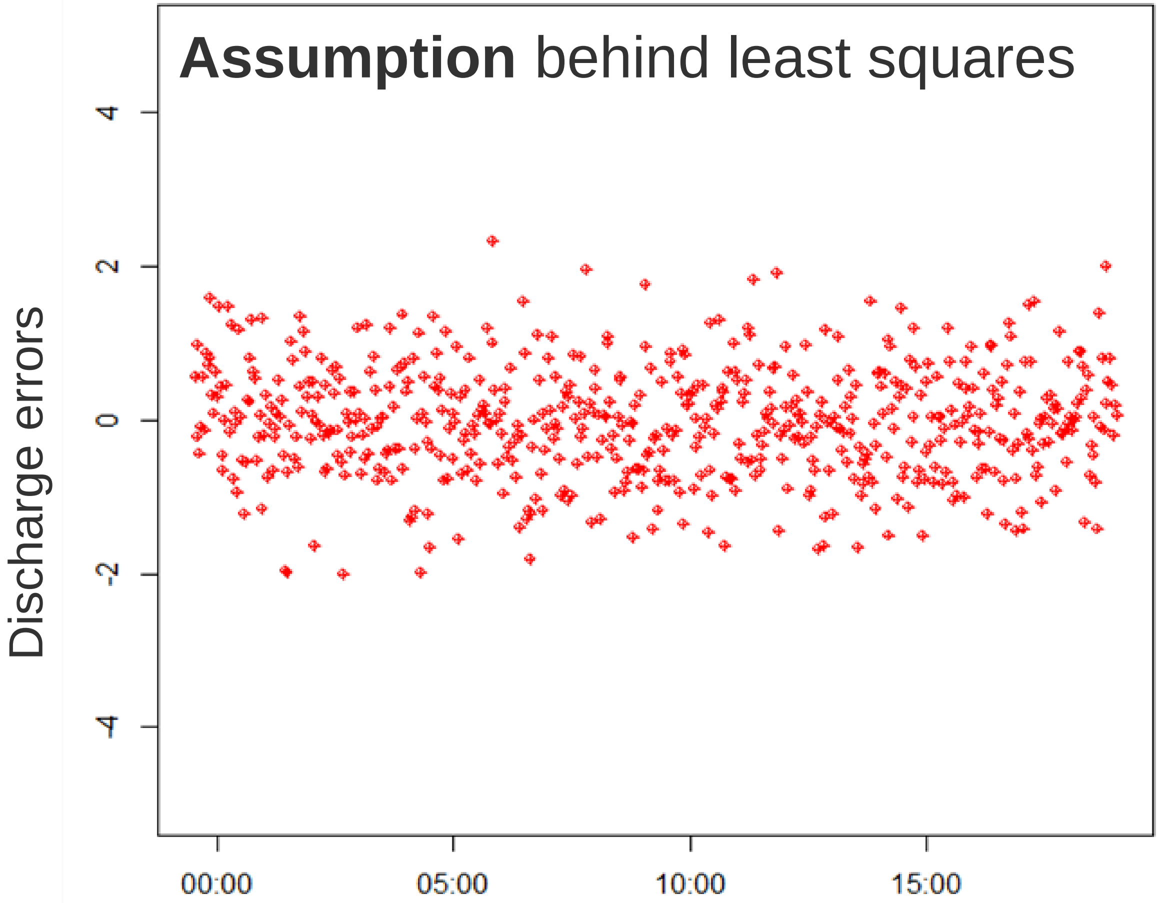
$$\cdot \exp\left(-\frac{1}{2}\left[\tilde{\boldsymbol{y}}_o - \tilde{\boldsymbol{y}}_{\mathrm{M}}(\boldsymbol{\theta}, \boldsymbol{x})\right]^{\mathrm{T}} \boldsymbol{\Sigma}(\boldsymbol{\psi}, \boldsymbol{x})^{-1} \left[\tilde{\boldsymbol{y}}_o - \tilde{\boldsymbol{y}}_{\mathrm{M}}(\boldsymbol{\theta}, \boldsymbol{x})\right]\right) \prod_{i=1}^{n} \frac{\mathrm{d}g}{\mathrm{d}y}(y_{o,i}, \boldsymbol{\psi})$$

Statistically-based doesn’t always mean reliable

To work, uncertainty analysis has to be based on a **realistic likelihood** function!

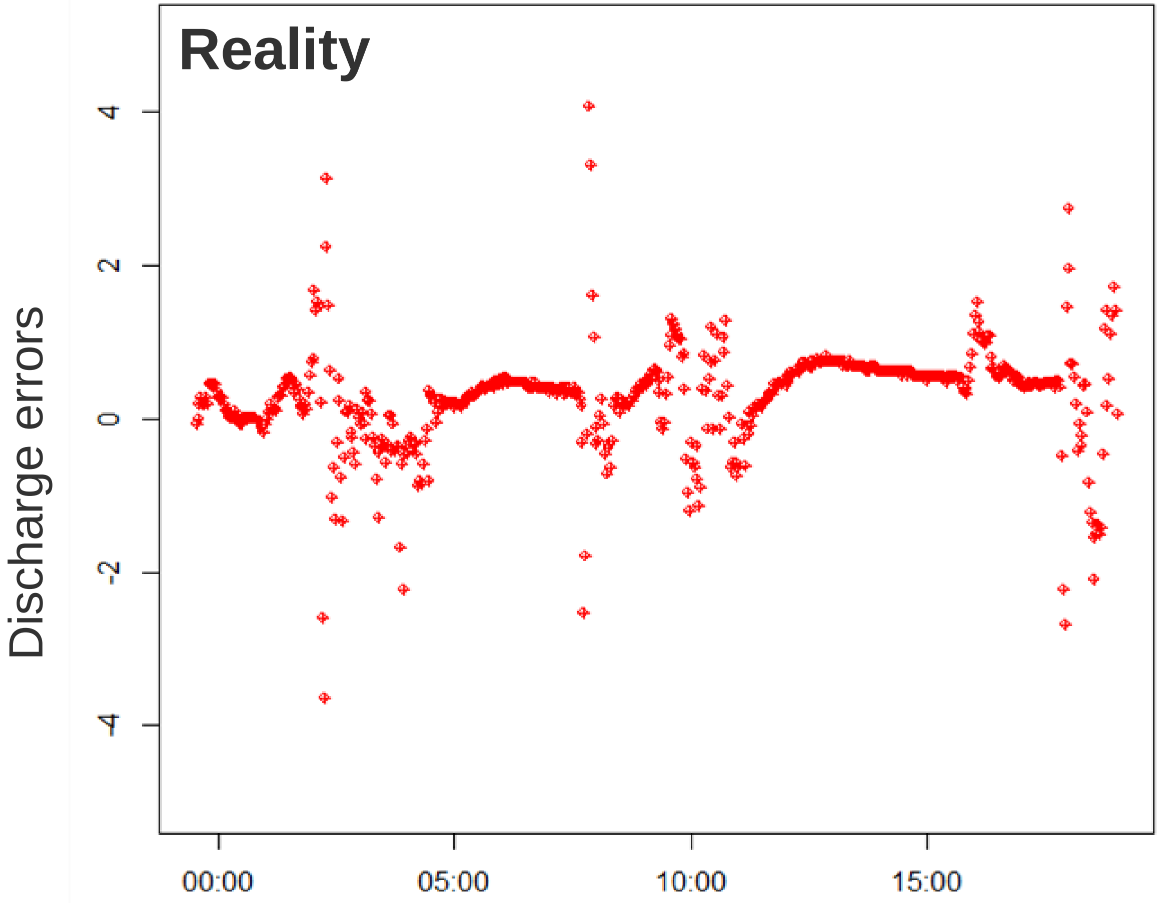
Traditional uncertainty assessment is overly idealized

model results - data = random uncorrelated errors

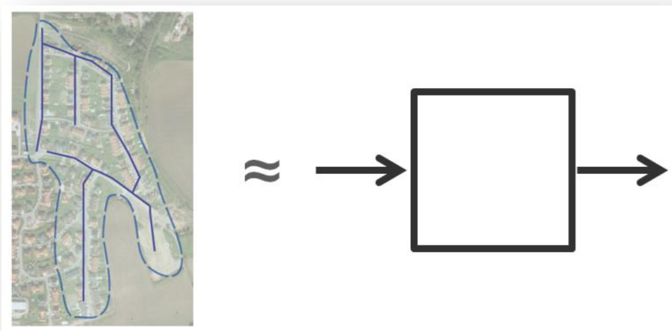


Reliable uncertainty assessment should account for reality

model results - data = non-stationary systematic errors

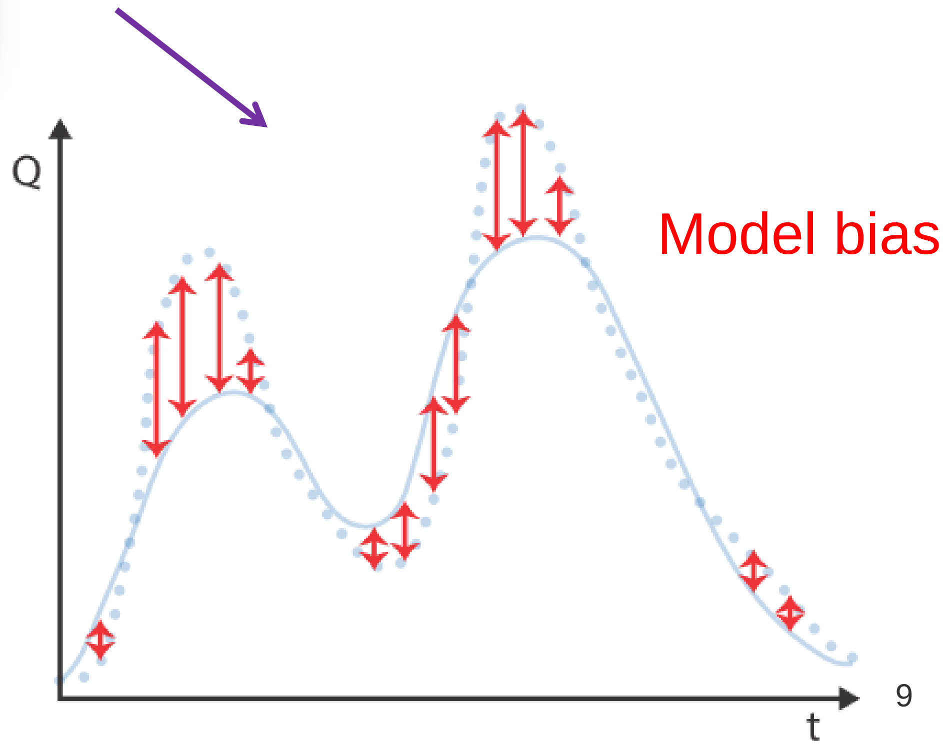


Why do we have these systematic deviations?



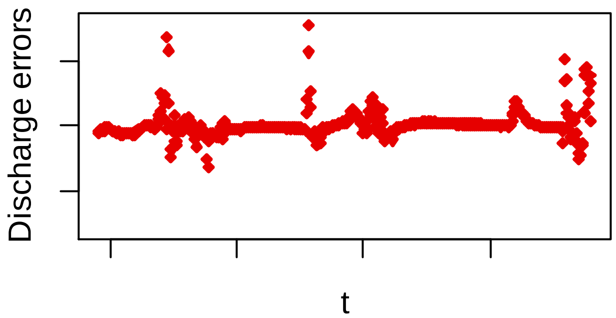
Structural deficits

Input errors

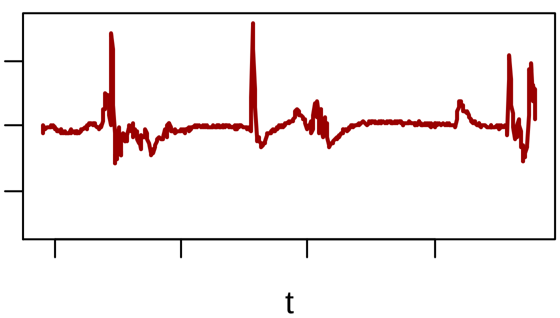


Real errors are better approximated by a bias description

We model the systematic output errors as a rainfall dependent Gauss-Markov process



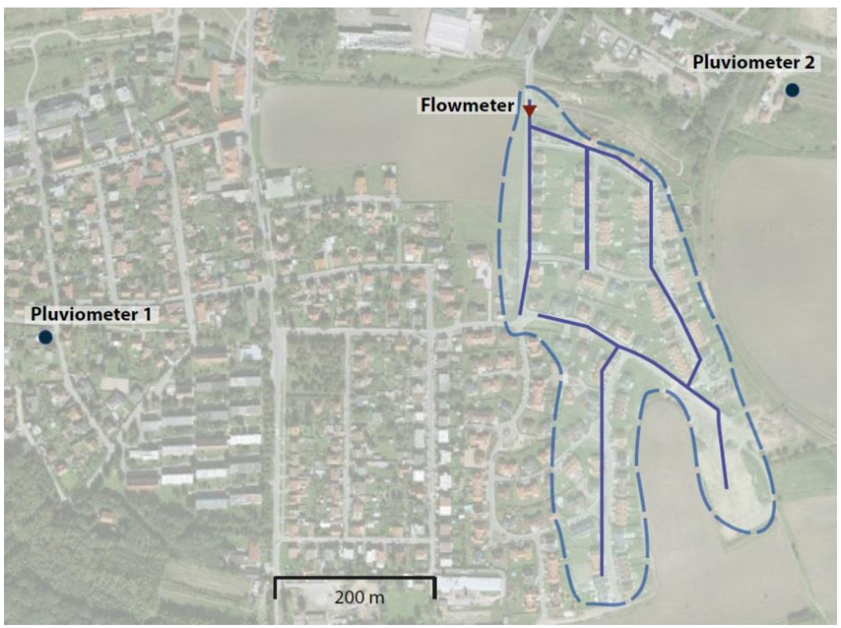
≈

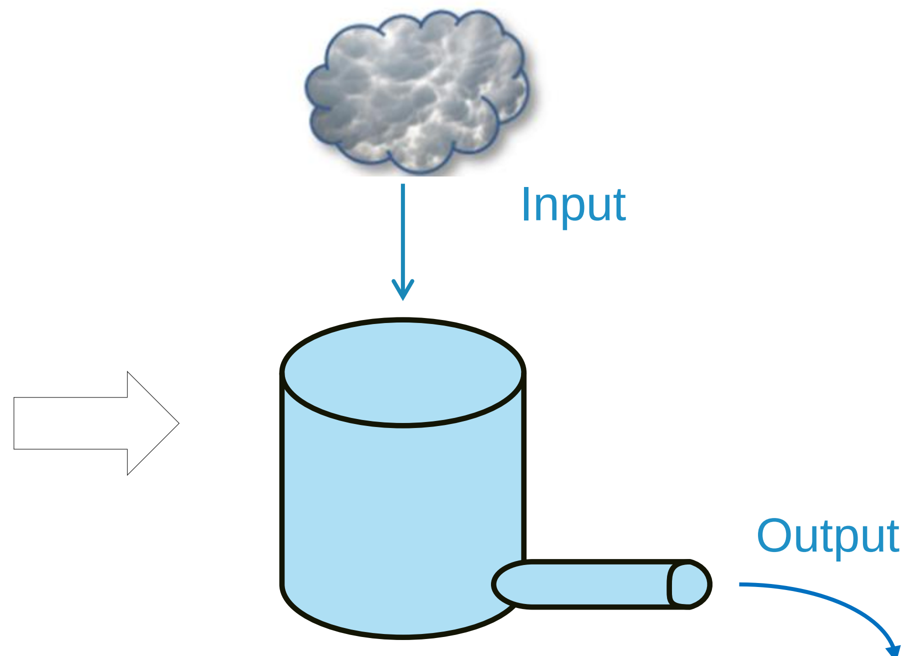


$$\mathrm{d}B_{\mathrm{M}}(t) = -\frac{B_{\mathrm{M}}(t)}{\tau}\mathrm{d}t + \sqrt{\frac{2}{\tau}\left(\sigma_{B_{\mathrm{ct}}}^2 + (\kappa x(t-d))^2\right)}\mathrm{d}W(t)$$

How does it work?

Example: flow simulations over 2 days



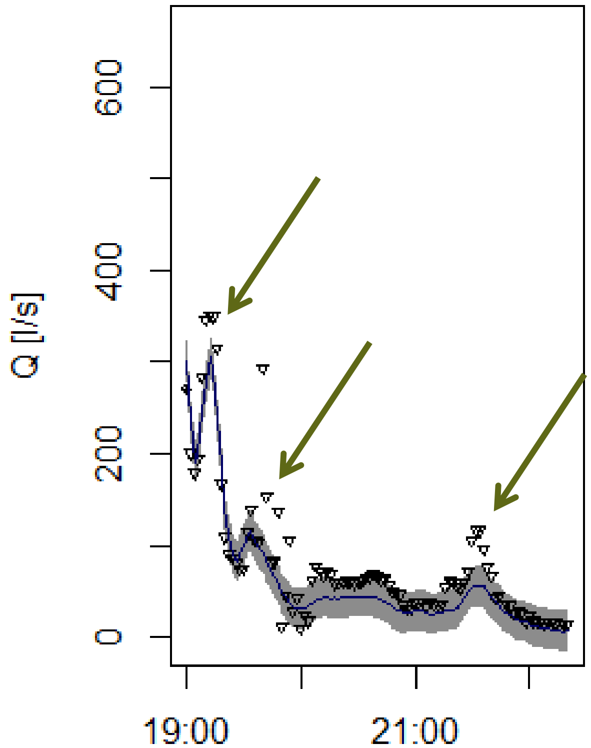


Stormwater catchment, CZ

Hydrologic model

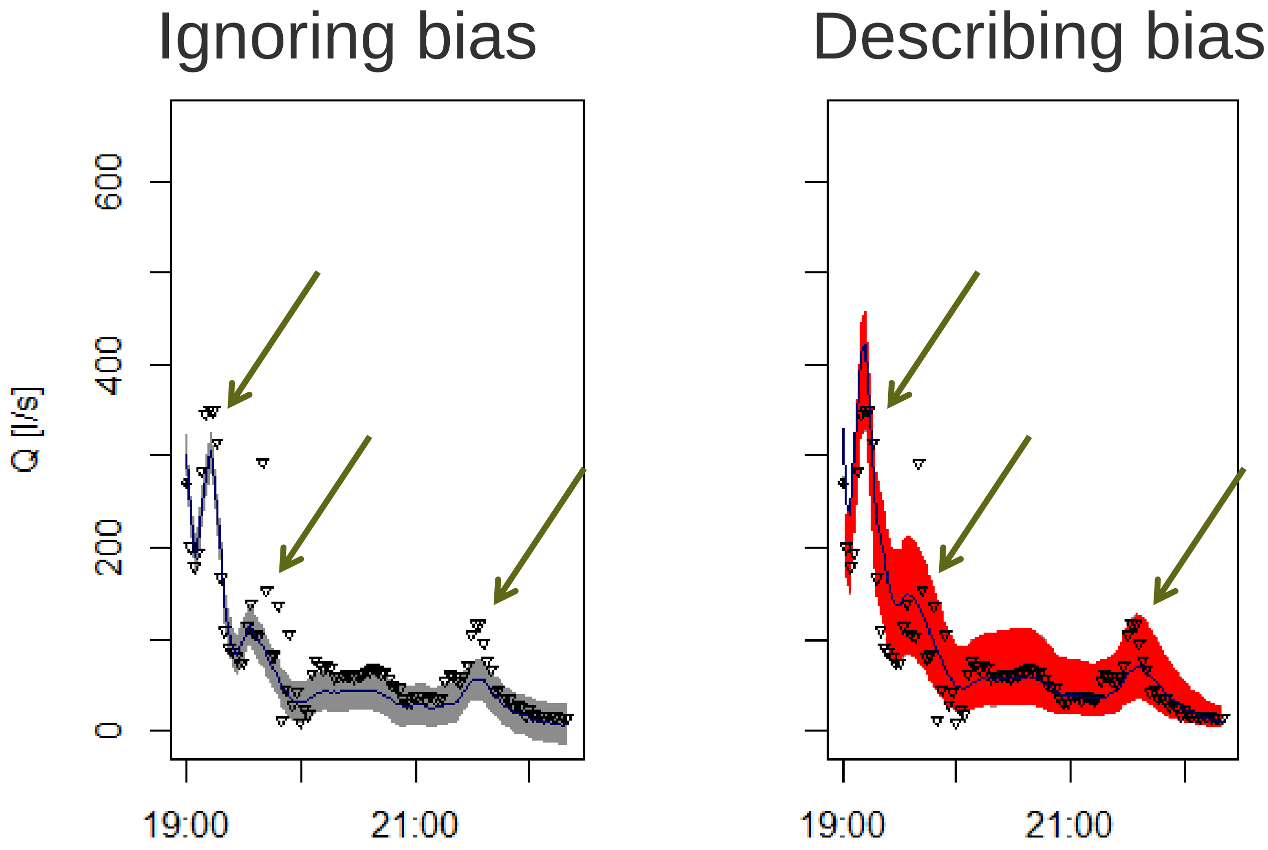
Results: Hindcasting

Ignoring bias



Del Giudice et al., Improving uncertainty estimation (2013) - HESS

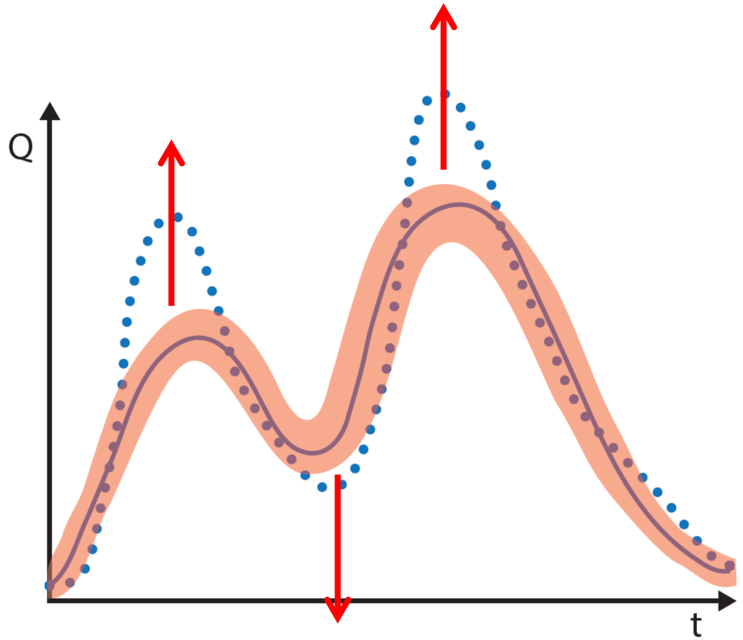
Results: Hindcasting



Del Giudice et al., Improving uncertainty estimation (2013) - HESS

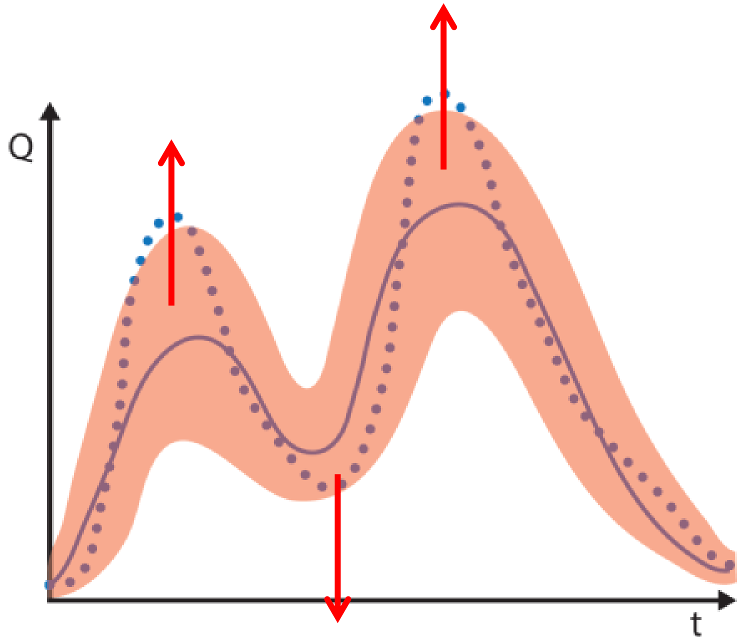
Discussions (I)

Rainfall & structural errors matter: with the bias description predictions are more reliable and robust



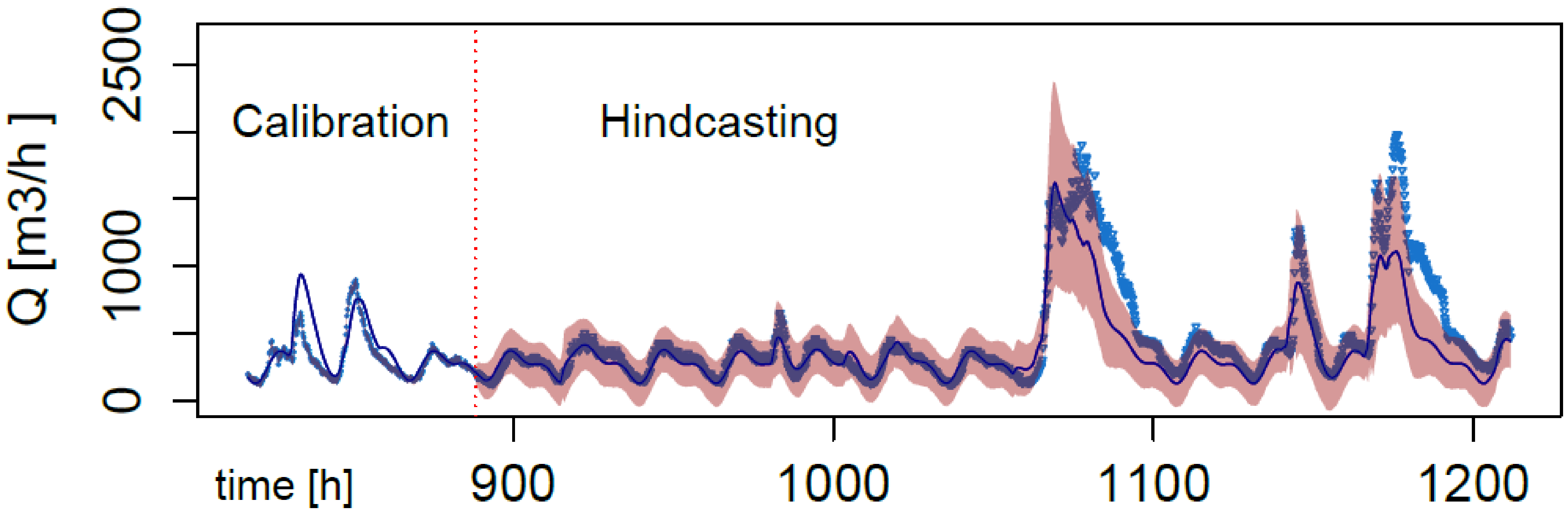
Discussions (I)

Rainfall & structural errors matter: with the bias description predictions are more reliable and robust



Discussions (II)

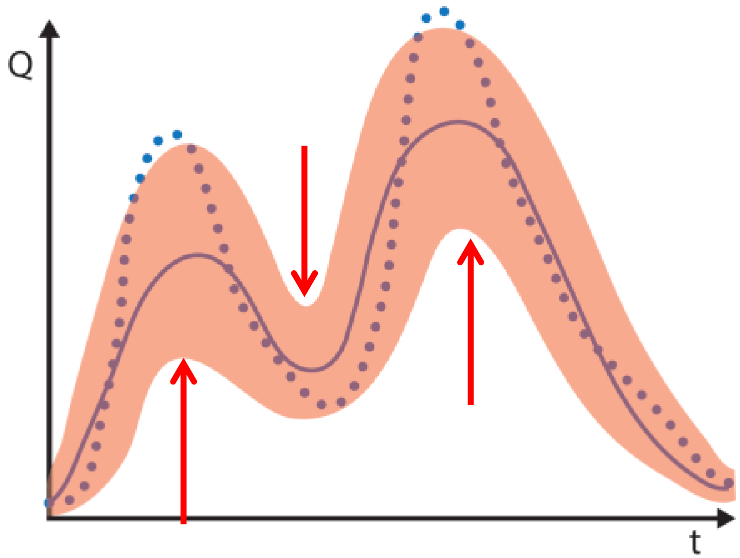
Current implementation is optimal for long-term predictions with parsimonious models



Del Giudice, Mikkelsen et al., Comparing two stochastic techniques (in prep.) - WRR

Discussions (III)

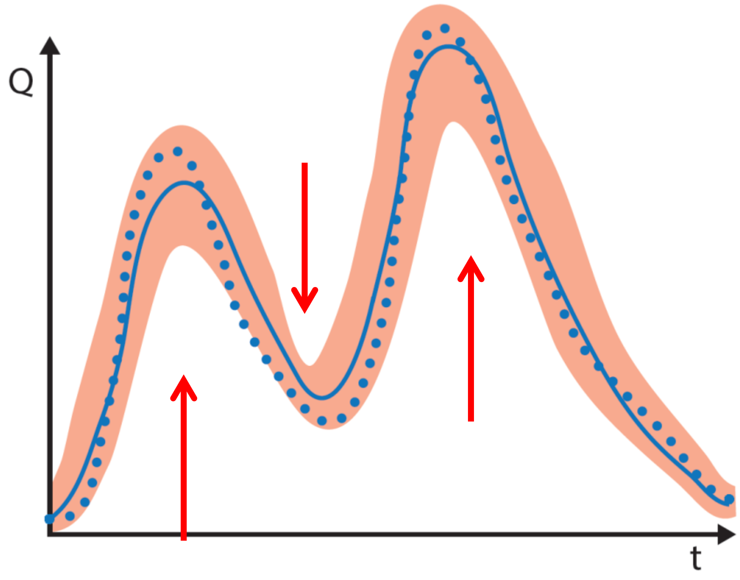
Bias description only improves the assessment of the uncertainties but doesn't reduce them



Del Giudice et al., Model bias and complexity (2014) – EM&S

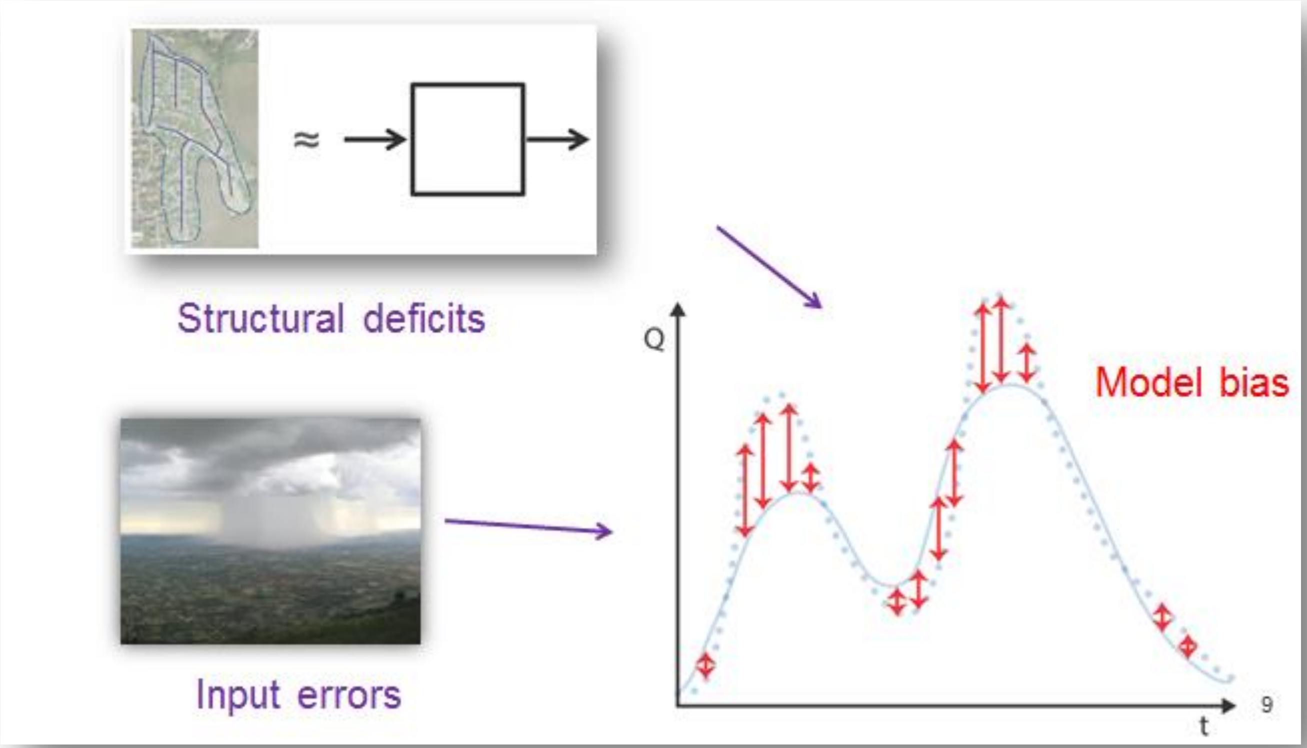
Discussions (III)

Bias description only improves the assessment of the uncertainties but doesn't reduce them



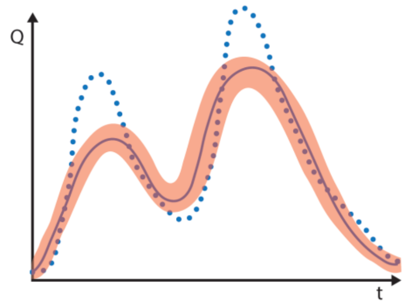
Del Giudice et al., Model bias and complexity (2014) – EM&S

Current challenge: quantify and reduce input uncertainty



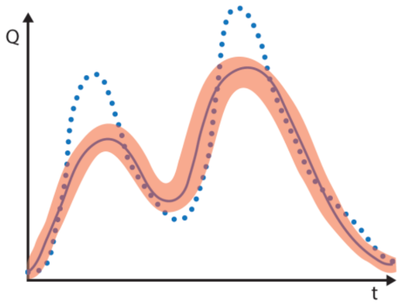
Conclusions

i. Rainfall/structural errors play a crucial role in urban drainage predictions

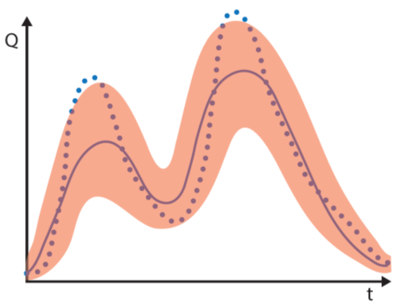


Conclusions

i. Rainfall/structural errors play a crucial role in urban drainage predictions

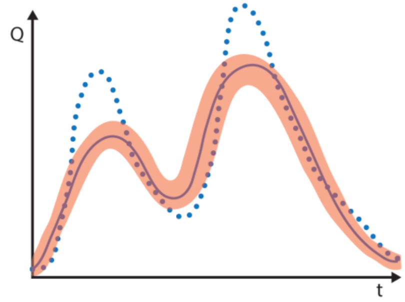


ii. A reliable way to quantify these uncertainties is the bias description

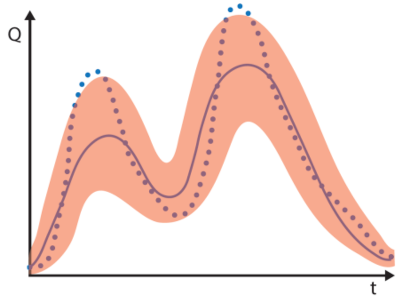


Conclusions

i. Rainfall/structural errors play a crucial role in urban drainage predictions



ii. A reliable way to quantify these uncertainties is the bias description



iii. A more advanced input description will minimize rainfall uncertainties