Uncertainty Estimation of hydrological models by using Gaussian-Process-Regression

Christian Fischer
Kos, 19. October 2013

Introduction

Hydrological models are subject to several uncertainties

- **Sources of uncertainty are for example:**

Data uncertainty
- measurement errors
- regionalization of data

Structural uncertainty
- the model is always only a simplified representation of real world processes
- not all processes can be taken into account

Parameter uncertainty
- different sets of model parameters are equally likely

- Uncertainty of the model determines whether a model is suitable for a specific application

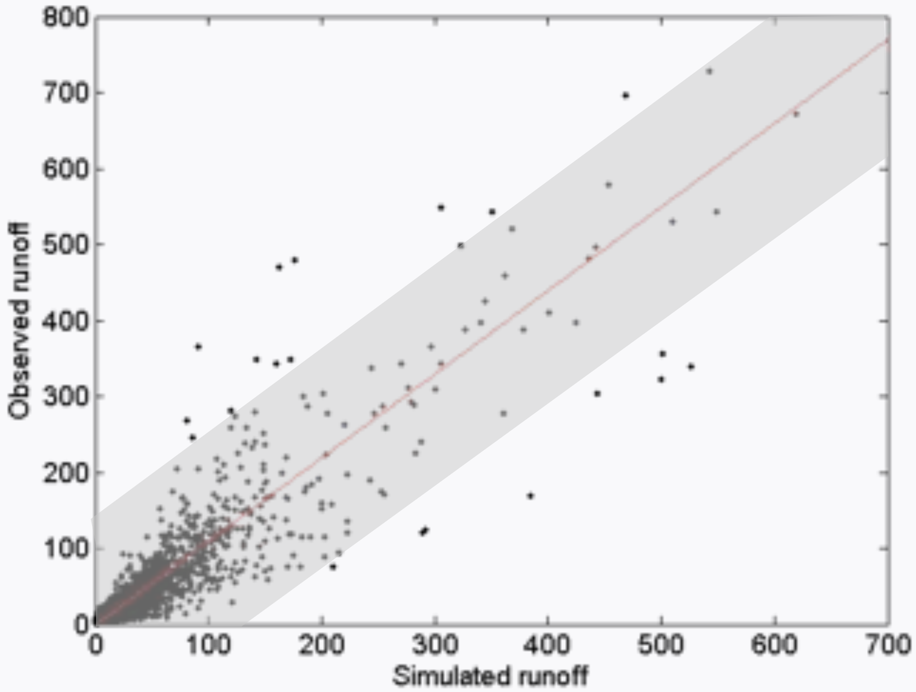
Problem definition

- **Problem:** Estimation of confidence intervals for the simulation of runoff where long measured runoff time-series are available
- **Statistical properties of the model error:**
  - heteroscedastic
  - non-gaussian
  - non-stationary
  - auto-correlated (i.e. not independent)

Properties of the model error

- **Approach:** Predict the observed runoff from the simulated runoff by using methods for function regression

=> Estimation of the prediction error



- Examples of possible methods
  - Linear regression
  - Meta-Gaussian Model (Montanari & Brath, 2004)
  - **Gaussian-Process-Regression (GPR)**

Gaussian-Process-Regression

- Training data are $\boldsymbol{D} = \{\boldsymbol{x}^{(i)}, \boldsymbol{o}^{(i)} \mid \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{n}\}$
- Each input is the simulated runoff ***x***.
- Each observation is a real-valued scalar, which follow a noisy function

o = f (x) + noise.

- Collect inputs in 1 × n matrix, ***X***, and targets in vector, y:

D = {X, y}

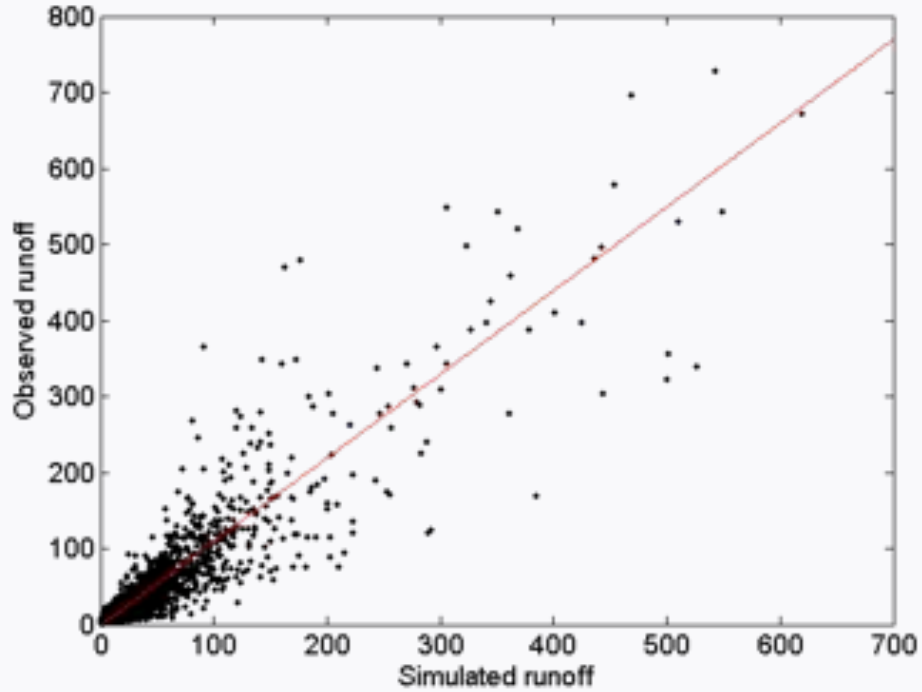
- Wish to infer $\boldsymbol{y}^*$ for unseen input $\boldsymbol{X}^*$, using ***P(y*|X*,D)***
- **Gaussian Process Regression (GPR)** gives $P(\boldsymbol{y}^*|\boldsymbol{X}^*,\boldsymbol{D}) = N(\mu, \sigma^2)$

$$\mu = (K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X}) + s^2 I)^{-1}\boldsymbol{y}$$

$$\sigma^2 = K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})[K(\boldsymbol{X}, \boldsymbol{X})s^2 I]^{-1}K(\boldsymbol{X}, \boldsymbol{X}^*)$$

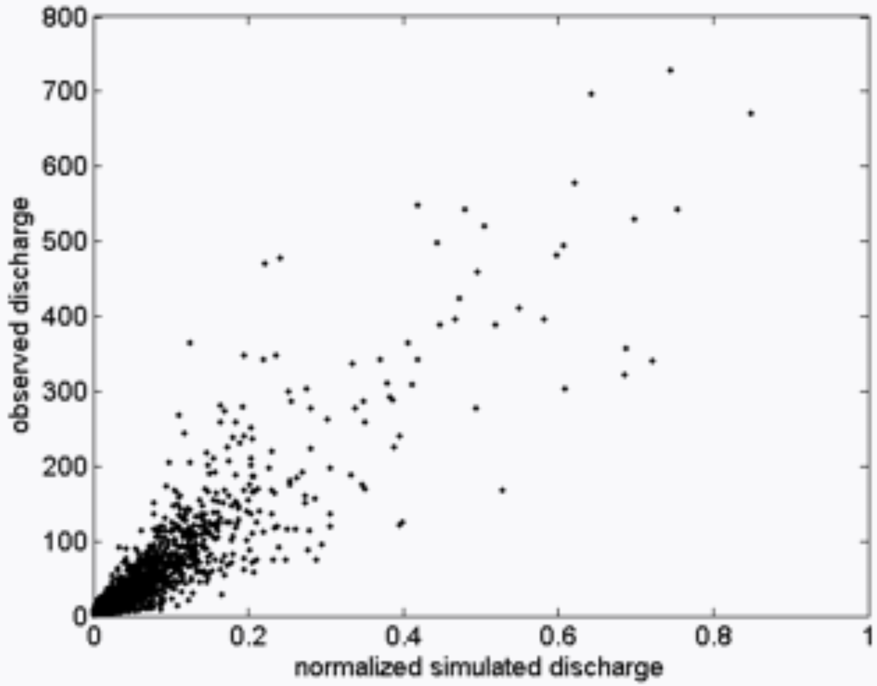
(Rasmussen & Williams, 2006)

Preprocessing steps



Preprocessing steps

Step 1: Linear rescale of simulated discharge

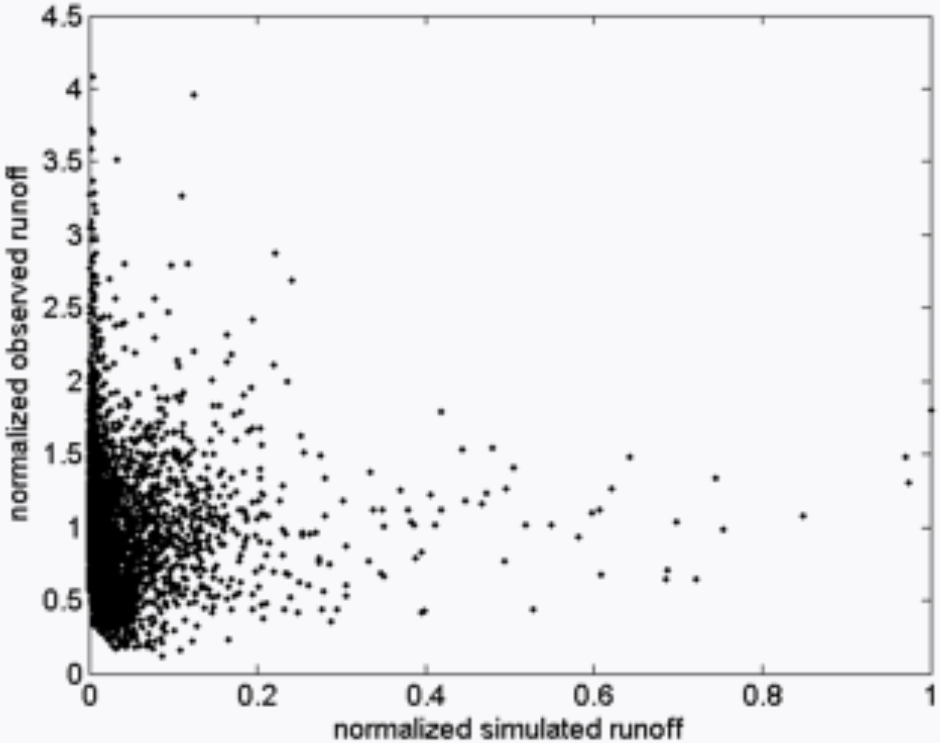


Preprocessing steps

Step 1: Linear rescale of simulated discharge

Step 2: Normalization of the observation

$$o'^{(i)} = \frac{o^{(i)}}{x^{(i)}+\beta}$$



Preprocessing steps

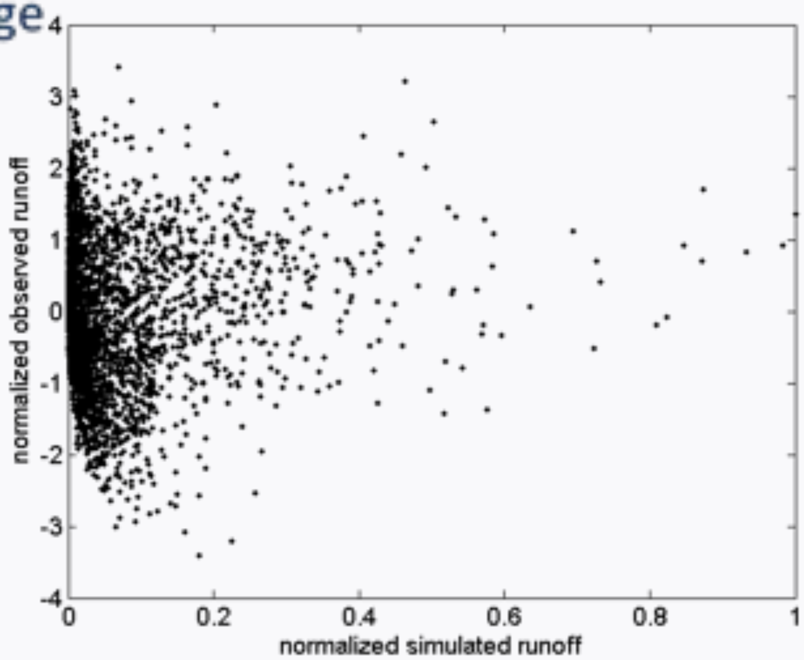
Step 1: Linear rescale of simulated discharge

Step 2: Normalization of the observation

$$o'^{(i)} = \frac{o^{(i)}}{x^{(i)} + \beta}$$

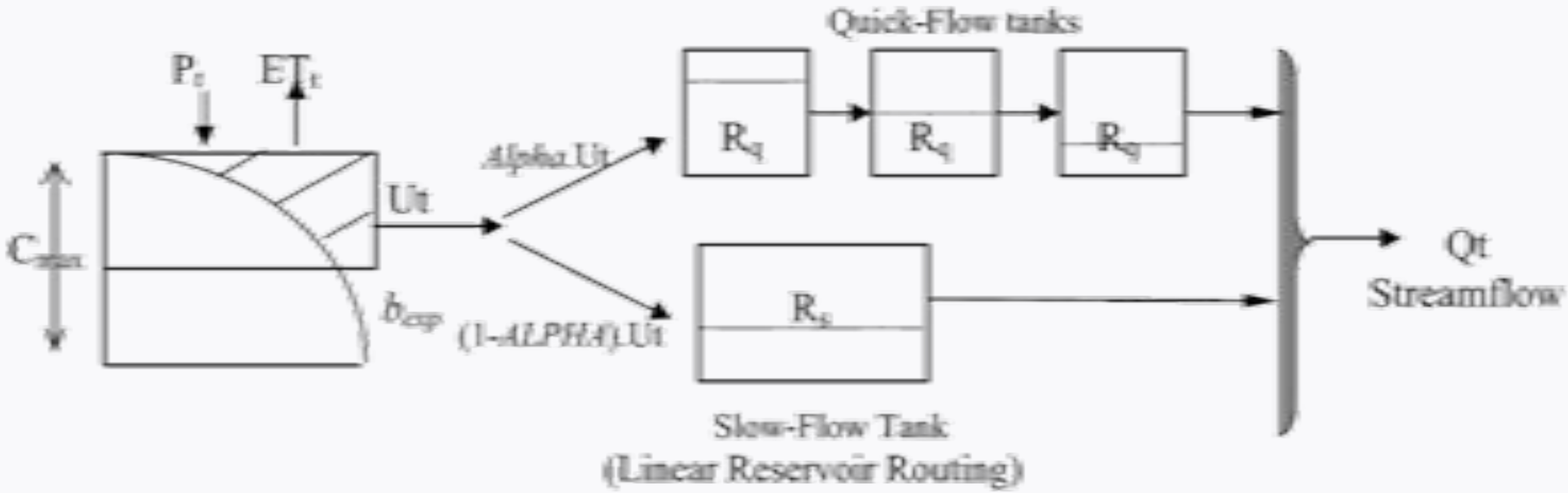
Step 3: Transformation by a Normal Quantile Transformation (NQT)

$$o''^{(i)} = NQT\left(o'^{(i)}\right)$$



Example

Hydrological model HYMOD



(Boyle, 2001)

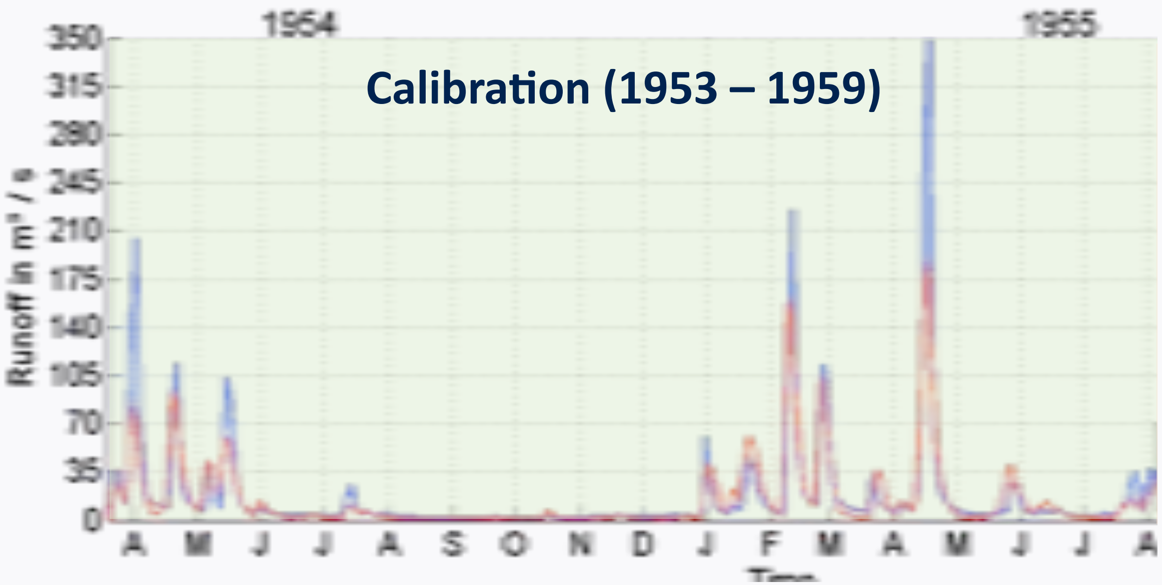
Example

- **Leaf River catchment**

- Used in studies by Vrugt et al. (2003-2013), Ajami et al. (2007), Misirili et al. (2003)

Example

Hydrological model HYMOD



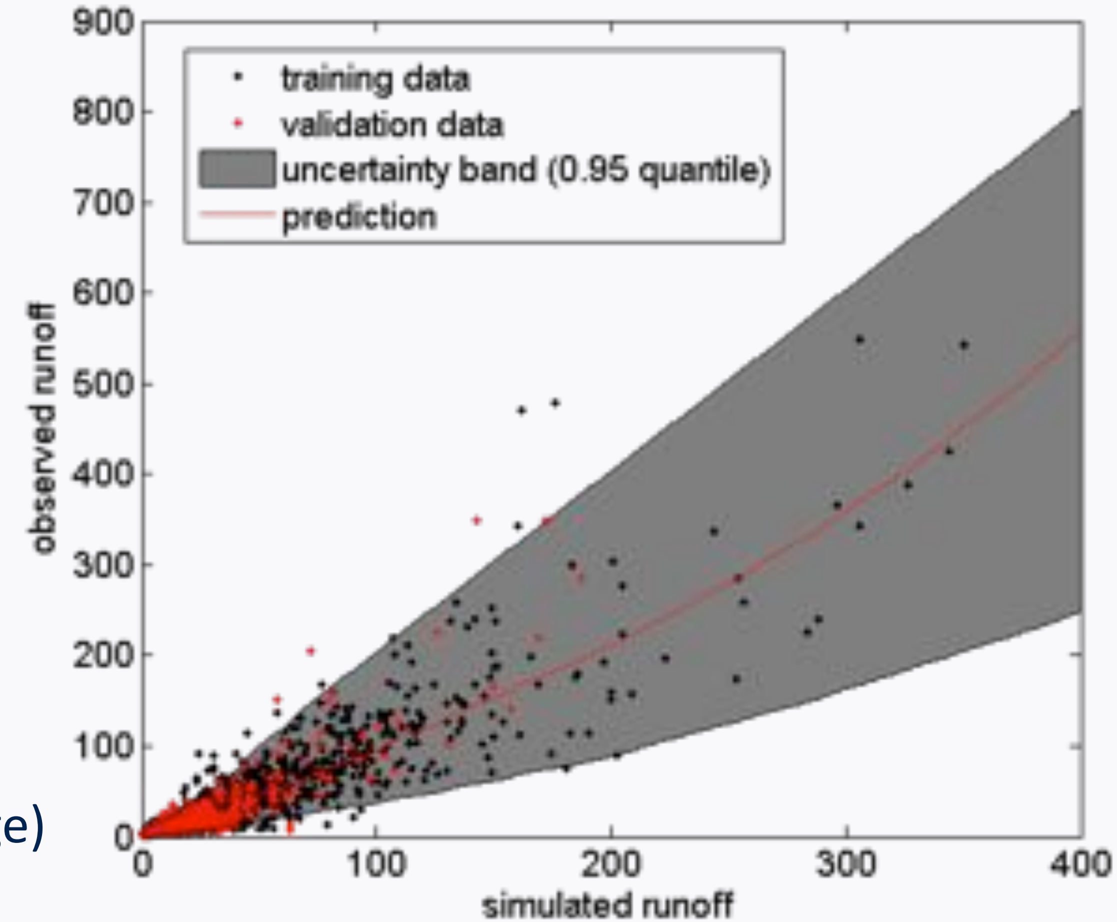
Calibration (1953 – 1959)

Validation (1960 – 1962)

Preliminary result

Gaussian Process Regression

- prediction increases linearly for simulated runoff up to 200m³/s and more than linear after that
- 91% of all observations are within the uncertainty range
- average error is 68 m³/s (124.23% of average discharge)



Enhancements

- Gaussian Process Regression is a multivariate method

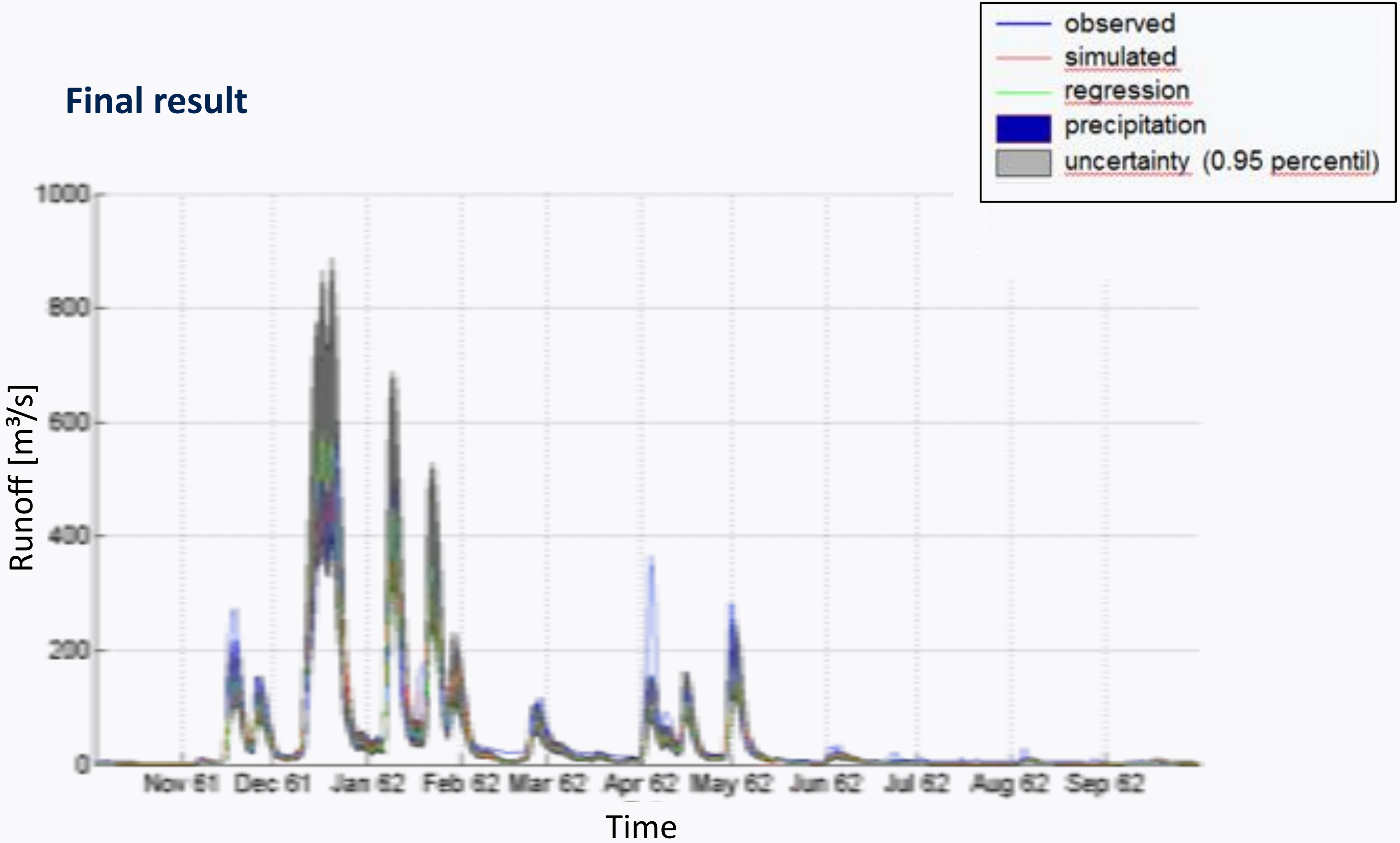
=> use several model states for prediction of observed runoff

- Possible candidates:
  - simulated discharge (Q)
  - rainfall input (R)
  - change of discharge in last 7 days (Q7)
  - total precipitation in last 7 days (R7)

model states	**observations within uncertainty band**	**average range of uncertainty band**
Q	91%	68 m³/s
Q,R	92%	64 m³/s
Q,Q7	90%	66 m³/s
Q, R7	91%	65 m³/s
Q,R, Q7	91%	65 m³/s
Q, R, R7	92%	64 m³/s
Q, Q7, R7	91%	65 m³/s
Q, R, Q7,R7	91%	66 m³/s

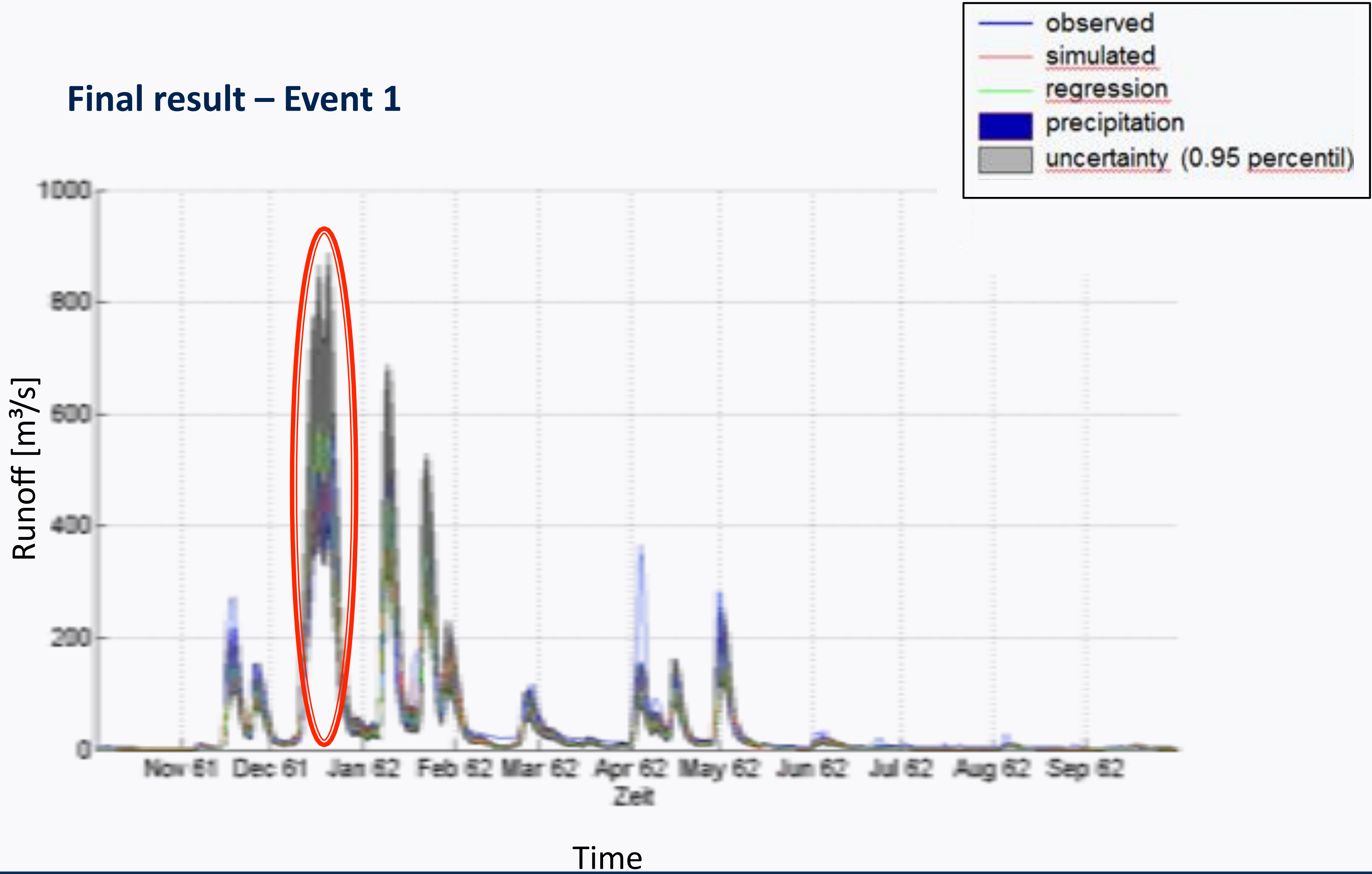
Final result

Final result



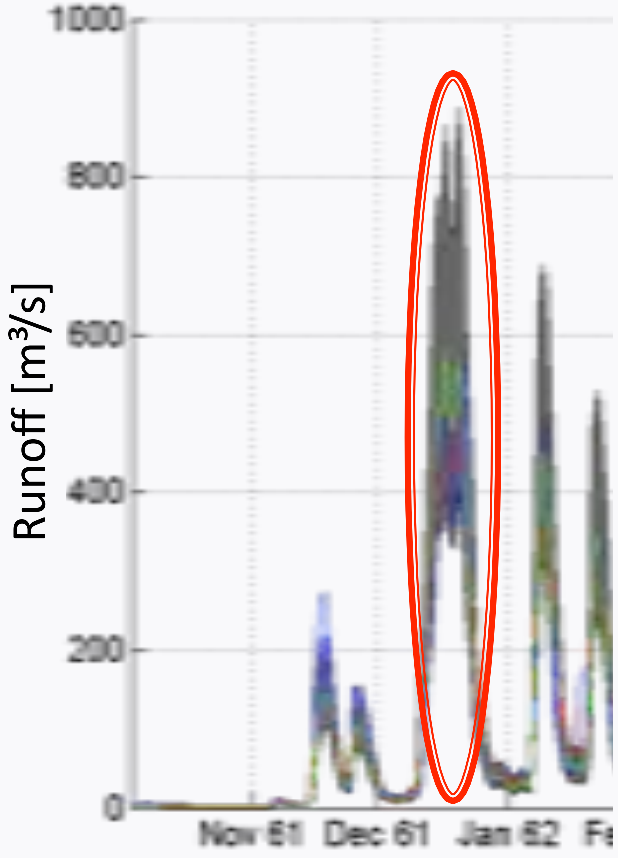
Final result

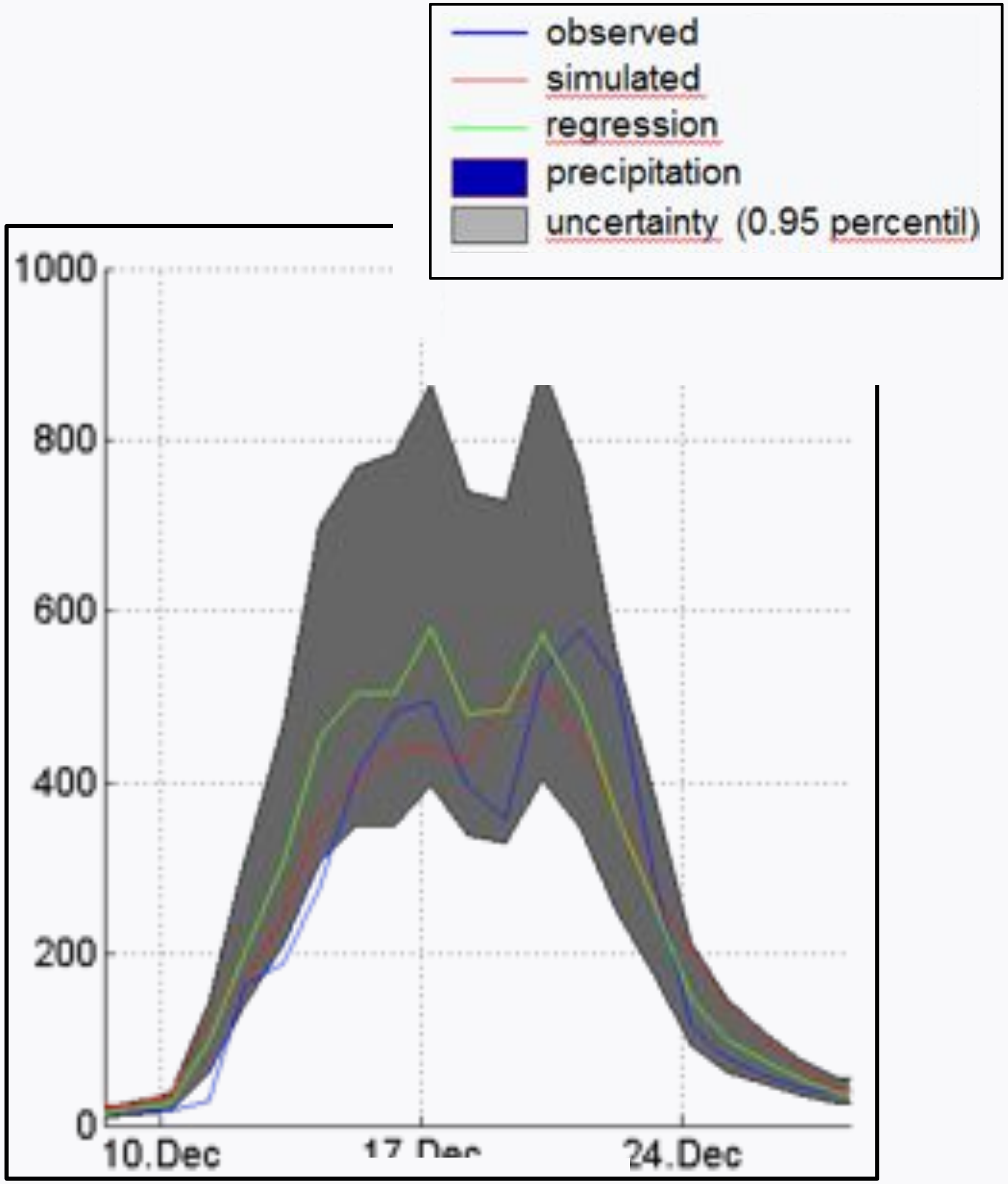
Final result – Event 1



Final result

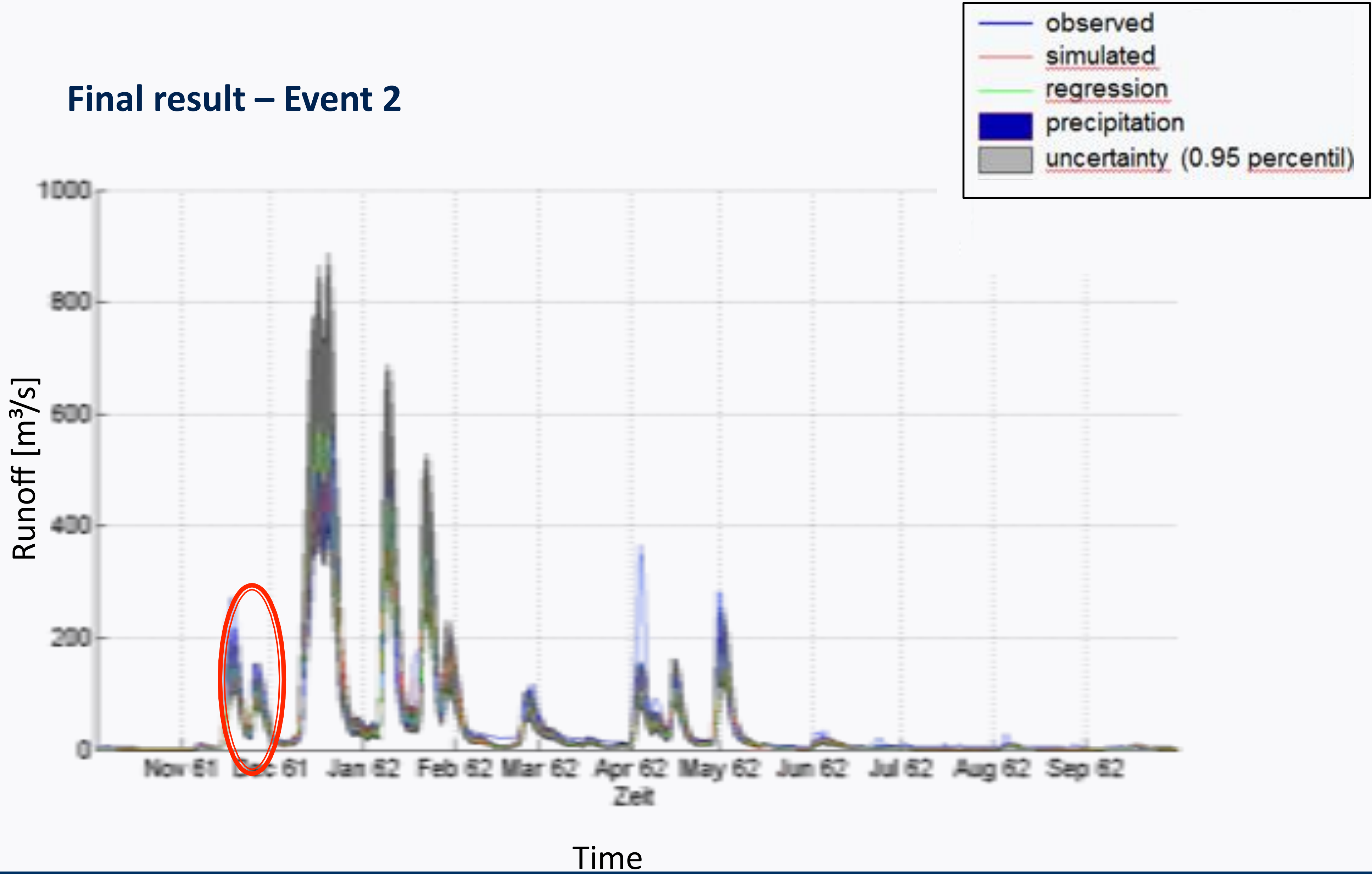
Final result – Event 1





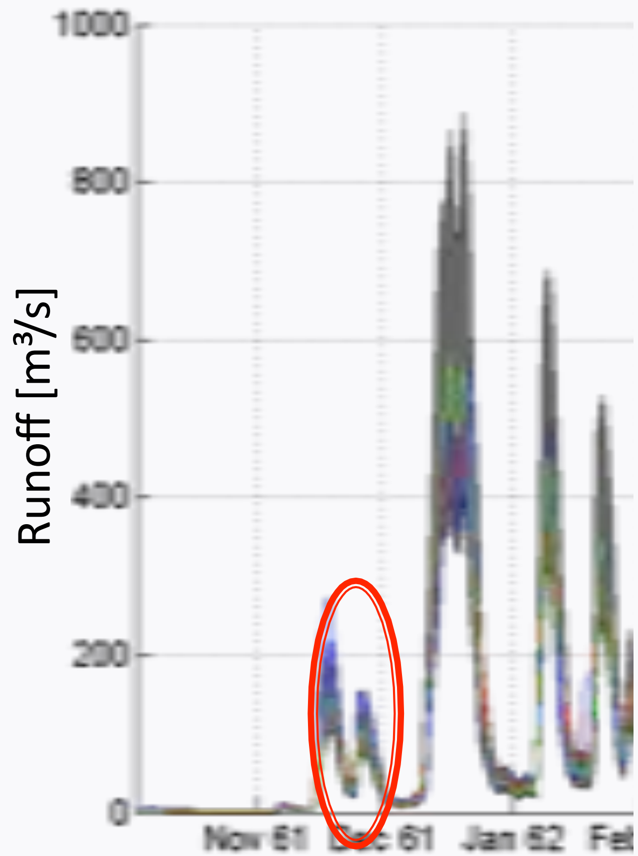
Final result

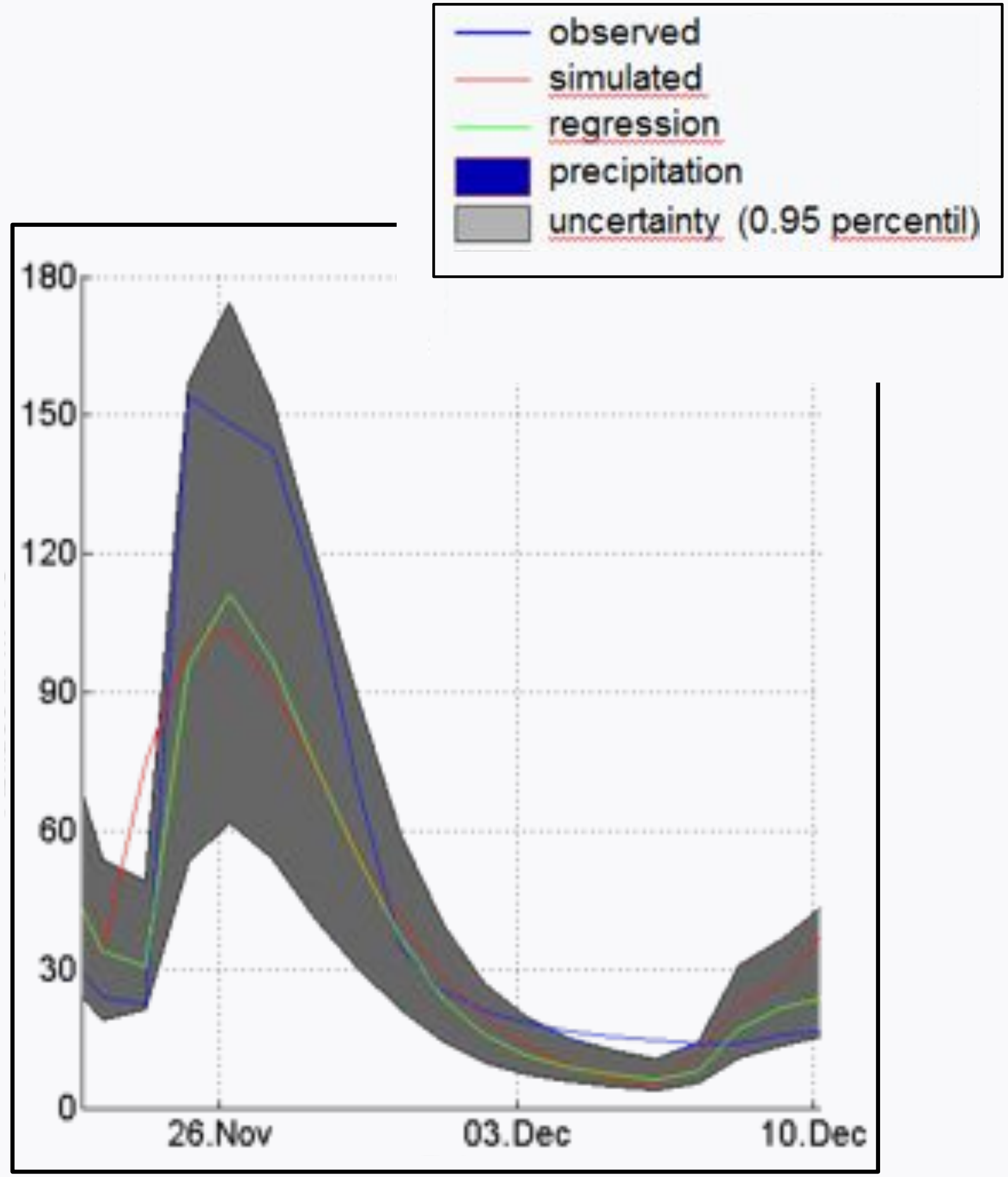
Final result – Event 2



Final result

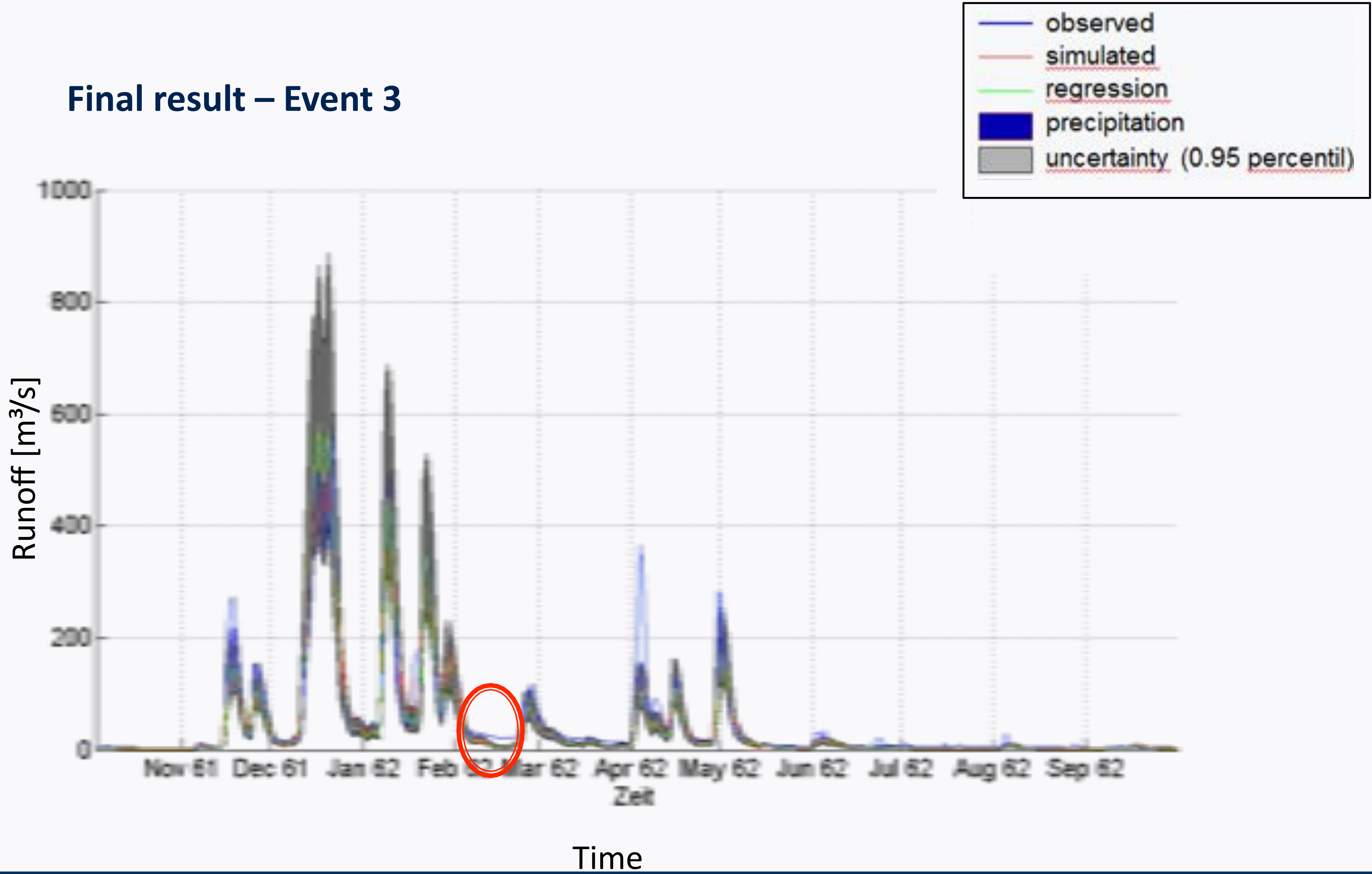
Final result – Event 2





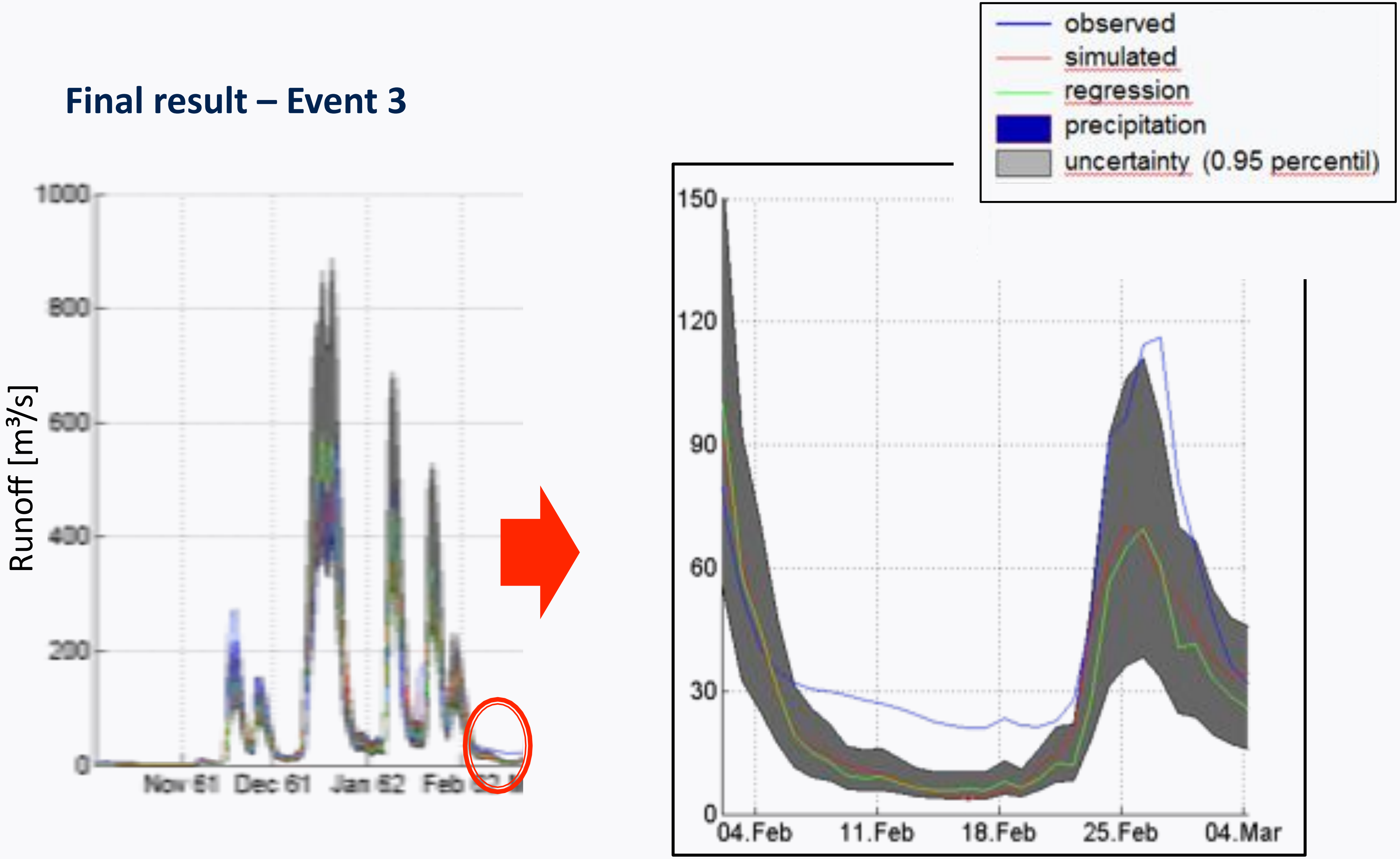
Final result

Final result – Event 3



Final result

Final result – Event 3



Summary

Limitations:

- long measured time series are required
- representation of non-stationary in practice very difficult
- extrapolation of the regression depends on the parameters of the Gaussian Process is modeled

Advantages:

- statistically based prediction of model error
- quality of resulting uncertainty bounds is not bad
- computational effort is relatively low

Conclusion

Gaussian Process Regression is an alternative approach to predict the error of hydrological models

Thank you for your attention!

Uncertainty Estimation of hydrological models by using Gaussian-Process-Regression

Christian Fischer
Kos, 19. October 2013