EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Bias correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment

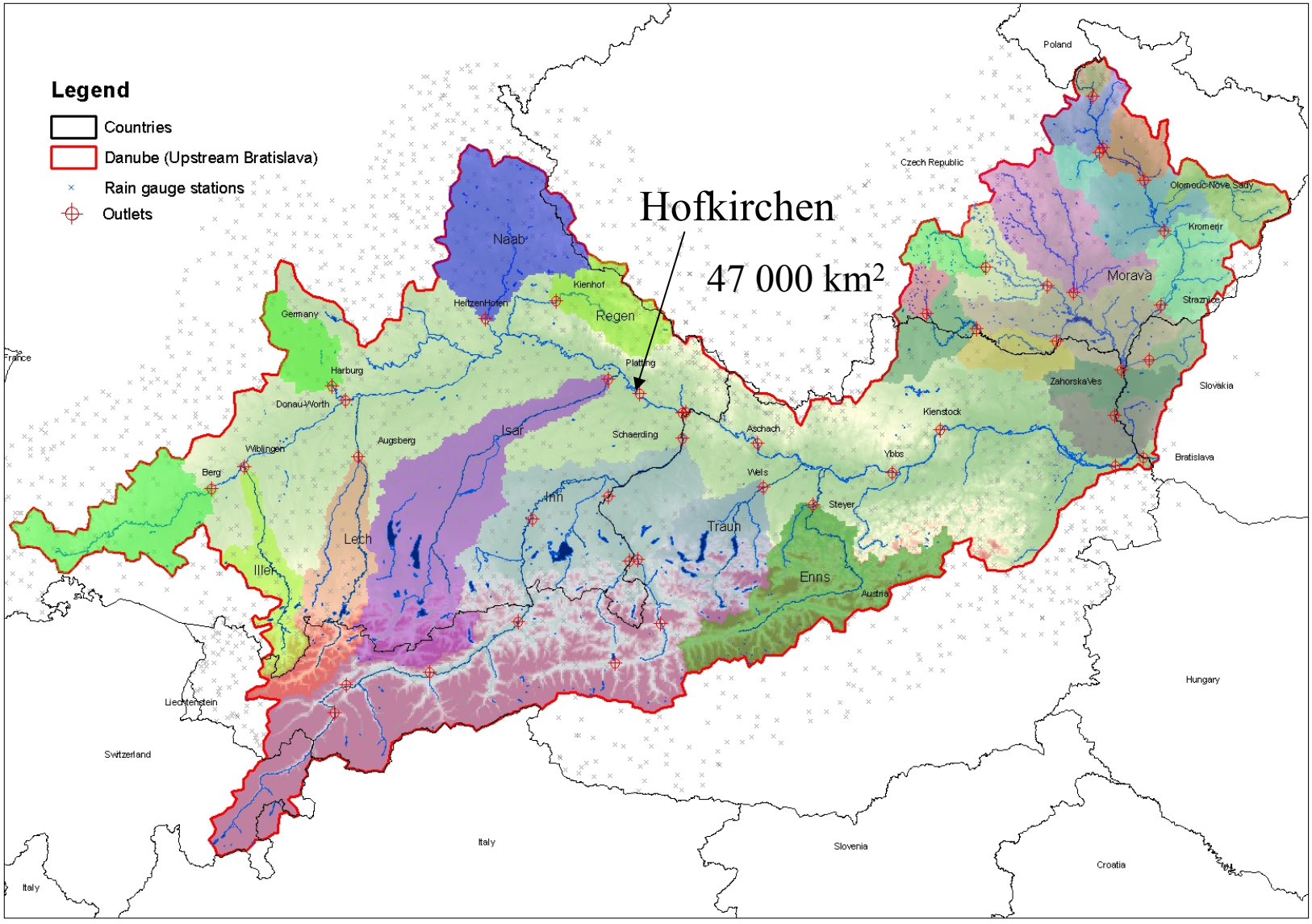
K. Bogner

Outline

- PREVIEW project
  - Catchment
  - Available Forecast data
- Bias Correction Methods
  - Wavelet Transformations + State Space model
  - Support Vector Machines
- Bayesian Model Averaging

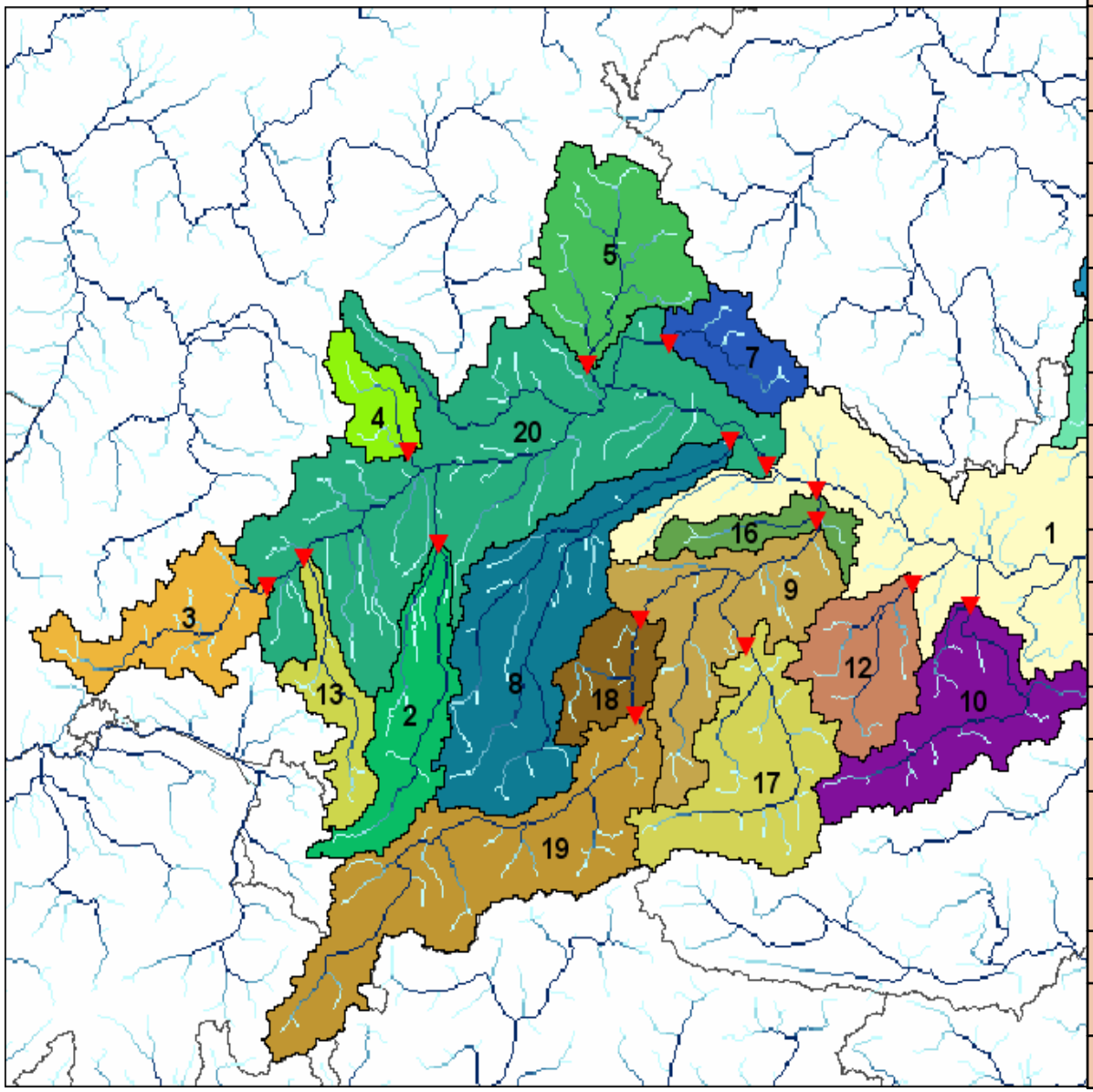
JRC - PREVIEW Objectives

- Evaluation of the added value of medium-range flood forecasting to allow for reliable extended warning times as compared to short-term forecasting.
- Demonstration of the usefulness of probabilistic forecasts based on ensembles - providing complementary information for water authorities.


Legend
Countries
Danube (Upstream Bratislava)
Rain gauge stations
Outlets
Hofkirchen
47 000 km²
Naab
Regen
Isar
Inn
Lech
Iller
Traun
Enns
Morava
Germany
France
Switzerland
Liechtenstein
Italy
Austria
Czech Republic
Slovakia
Hungary
Slovenia
Croatia
Poland
Kienhof
HeitzenHofen
Harburg
Donau-Worth
Wiblingen
Berg
Augsberg
Plattling
Schaerding
Aschach
Wels
Steyer
Ybbs
Kienstock
Zahorska Ves
Bratislava
Olomouc-Nove Sady
Kromeriz
Straznice

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

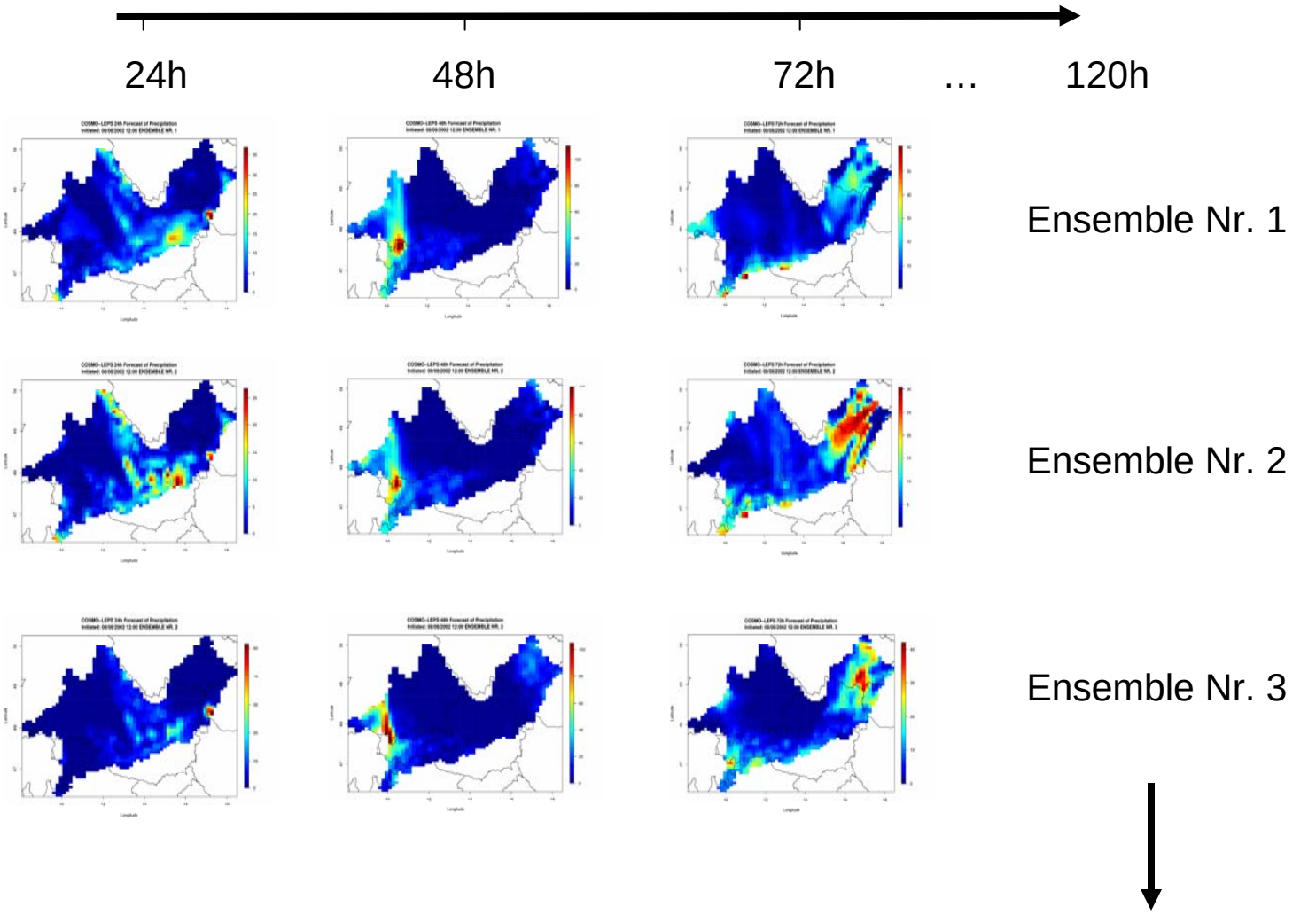
Joint Research Centre



ID	STATION NAME	AREA [km^2]
1	Bratislava	131978
2	Augsburg	3992
3	Berg	3891
4	Harburg	1630
5	Heitzenhofen	5450
6	Ivancice	2625
7	Kienhof	2218
8	Platting	9096
9	Schaerding	24346
10	Steyer	5998
11	Straznice	9418
12	Wels	3537
13	Wiblingen	2247
14	ZahorskaVes	25481
15	Zidlochovce	3928
16	Passau-Ingling	25977
17	Laufen	6174
18	Wasserburg	12004
19	Oberaudorf	9840
20	Hofkirchen	47534

Available weather forecast data

Provider	Product	#	Forecast Period	Spatial Resolution	Forecast Horizon
ECMWF	EPS	51	10/01/2001 - 30/09/2002	~80km	10 days
	VAREPS	51	20/07/2002 - 20/08/2002	~40km	14 days
	Deterministic	1	10/01/2001 - 30/09/2002	~40km	15 days
DWD	GME	1	10/01/2001 - 30/09/2002	~40km	7 days
	COSMO-LME	1	20/07/2002 - 20/08/2002	~7km	3 days
ARPA-SIM	COSMO-LEPS	11	20/07/2002 - 20/08/2002	~10km	5.5 days
IMK	High-resolution	1	03/08/2002 - 13/08/2002	~1km	1.25 days



Example of COSMO-LEPS precipitation forecast

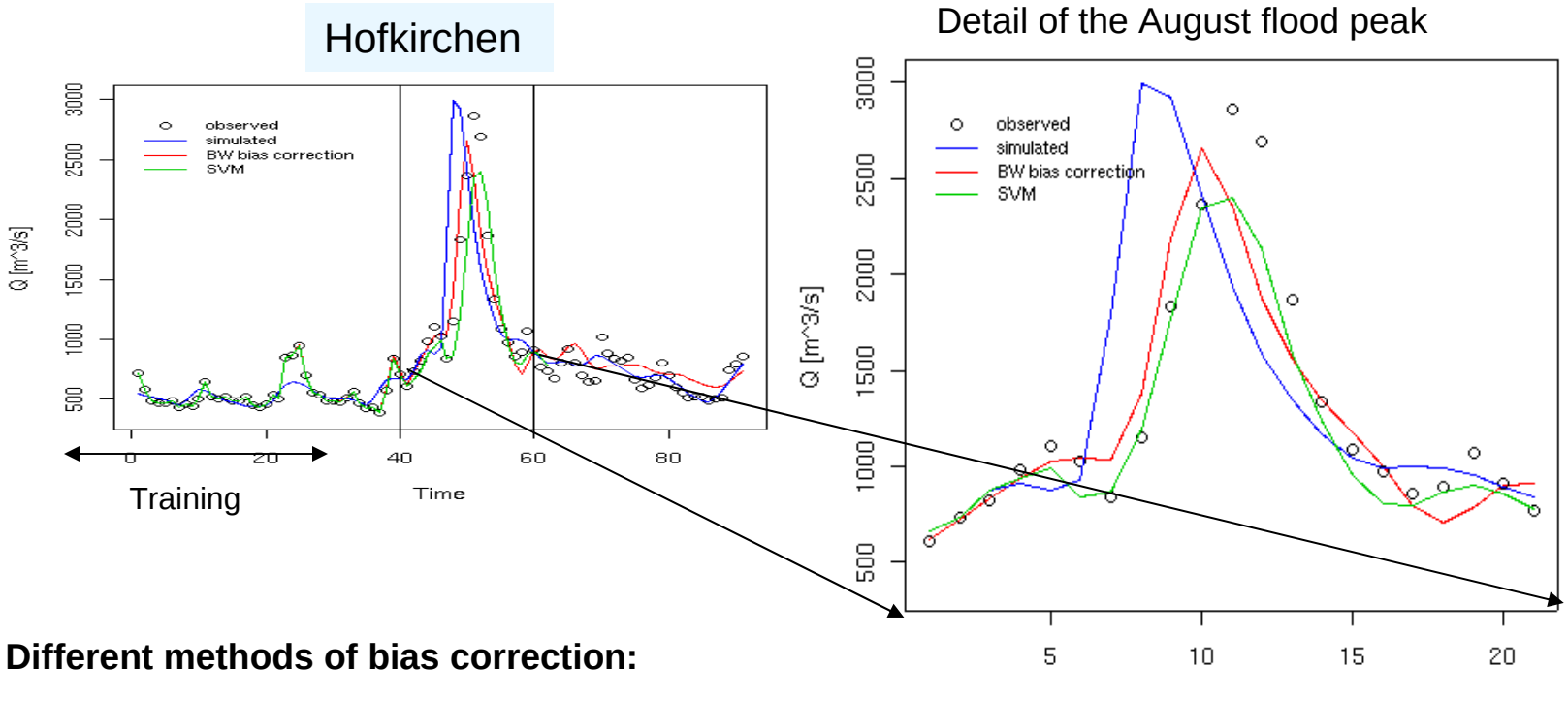
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Bias correction

Results for the August 2002 flood

Objective: Improving the ensemble forecast quality (timing of the peak, peak volume, ..)

Method: Adjusting the ensemble traces using a transformation derived with simulated and observed flows from the Year 2002



Different methods of bias correction:

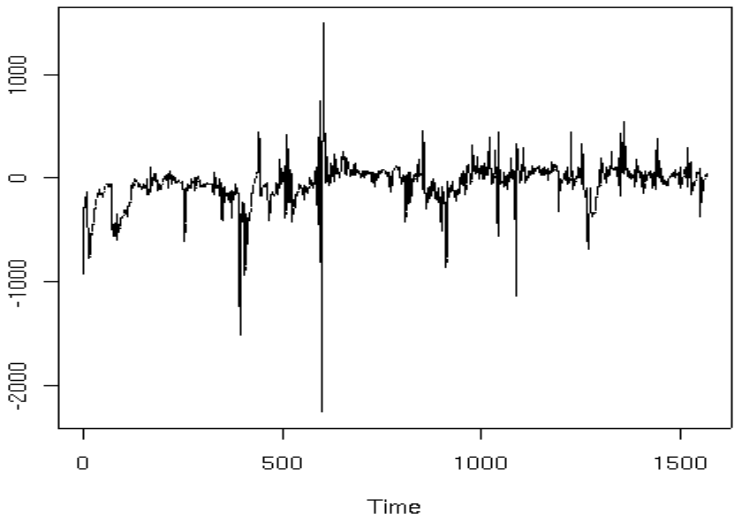
1. Wavelet transformation (breaking the data into a set of scaled and translated versions of a wavelet function) + Bayesian Time Series Analysis (dynamic regression for each scale)

2. SVM: Support Vector Machine (kernel based neural networks)

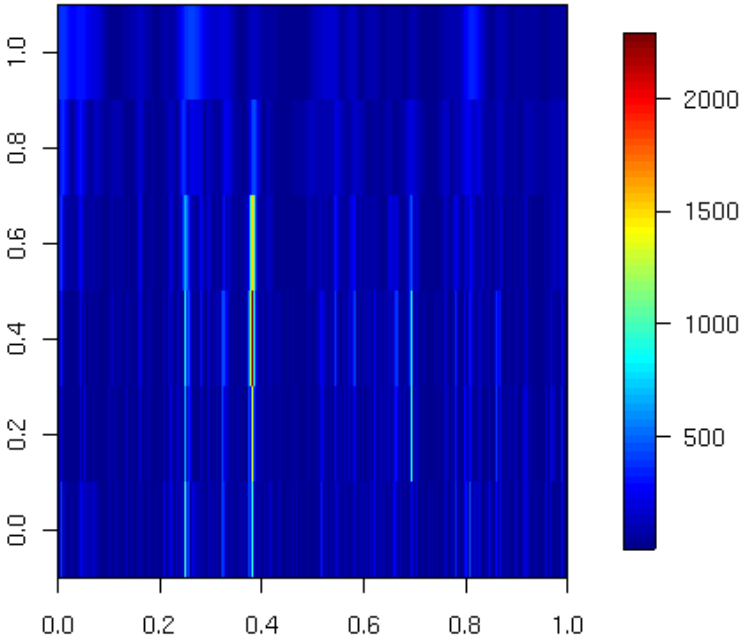
Joint Research Centre

Wavelet based Error Analysis

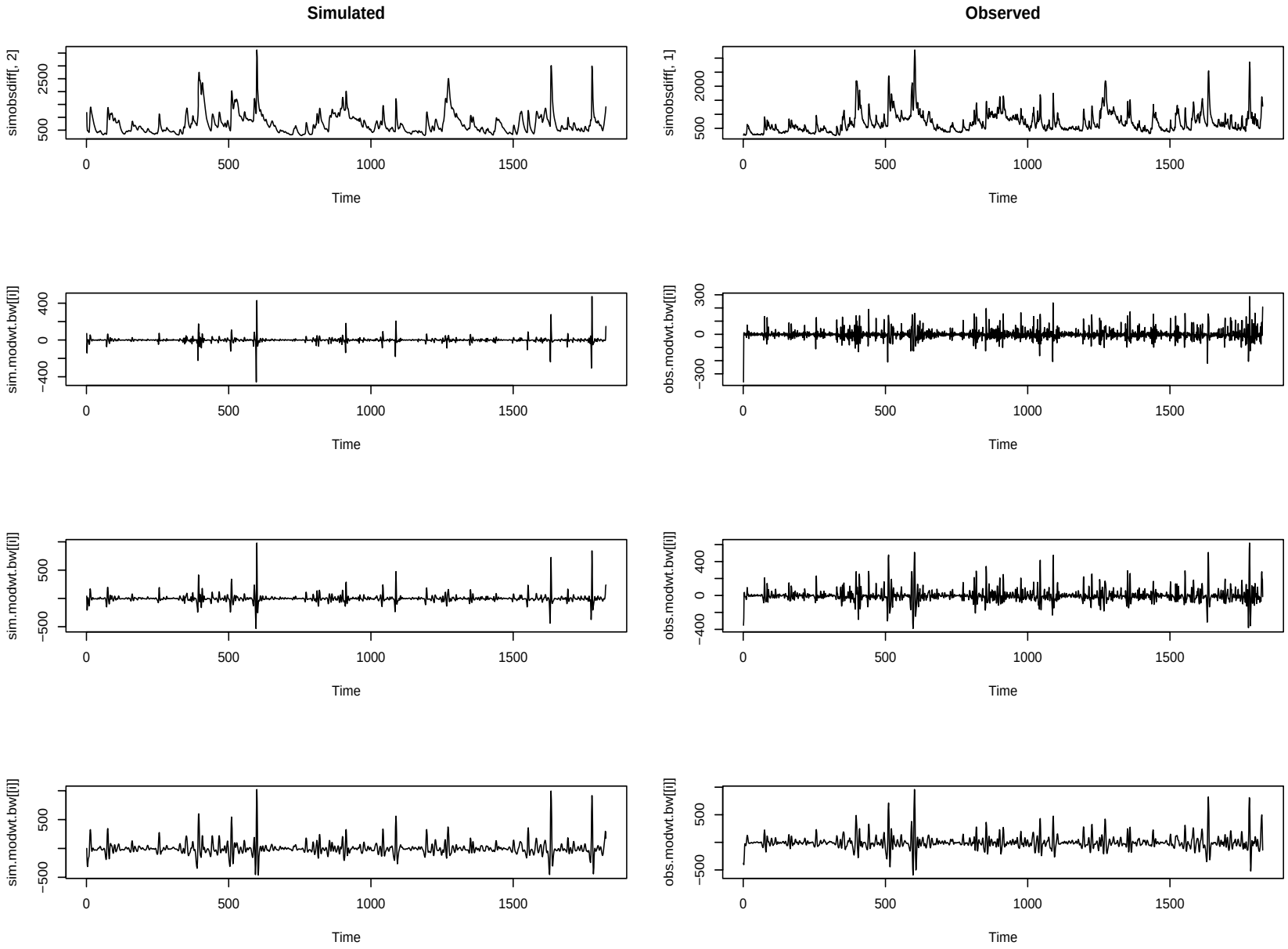
Magnitude and scale of model error show time dependences (e.g. snowmelt season). This requires localization in time and consideration of the time-scale over which error is manifest. This could be done by Wavelet transformation.

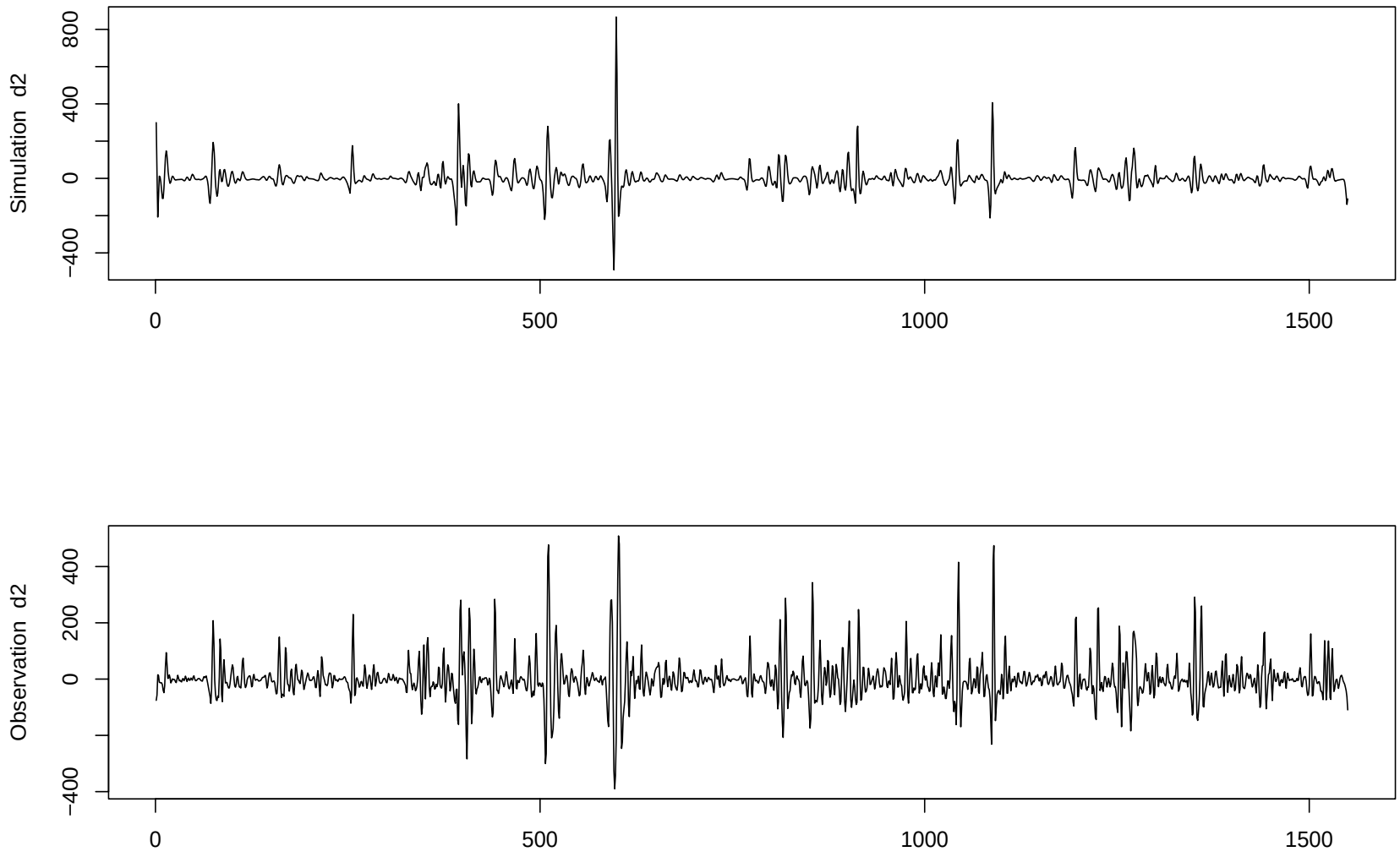


Error: Observed - Simulated

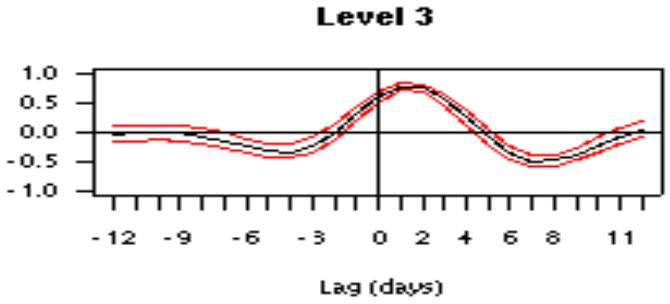


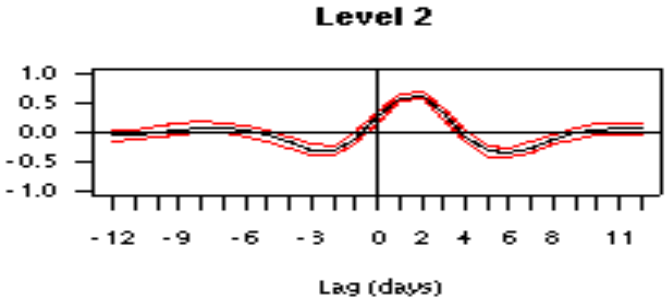
Wavelet transformation of the Error (Time vs. Scale)

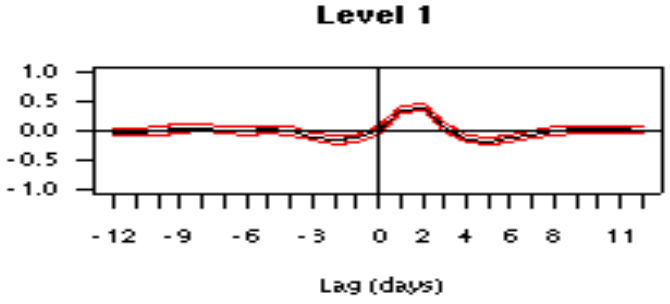

Simulated
Observed
simobsdifff[, 2]
simobsdifff[, 1]
sim.modwt.bwf[[i]]
obs.modwt.bwf[[i]]
Time
0
500
1000
1500


Simulation d2
800
400
0
−400
0
500
1000
1500
Observation d2
400
200
0
−400
0
500
1000
1500

Wavelet Correlation







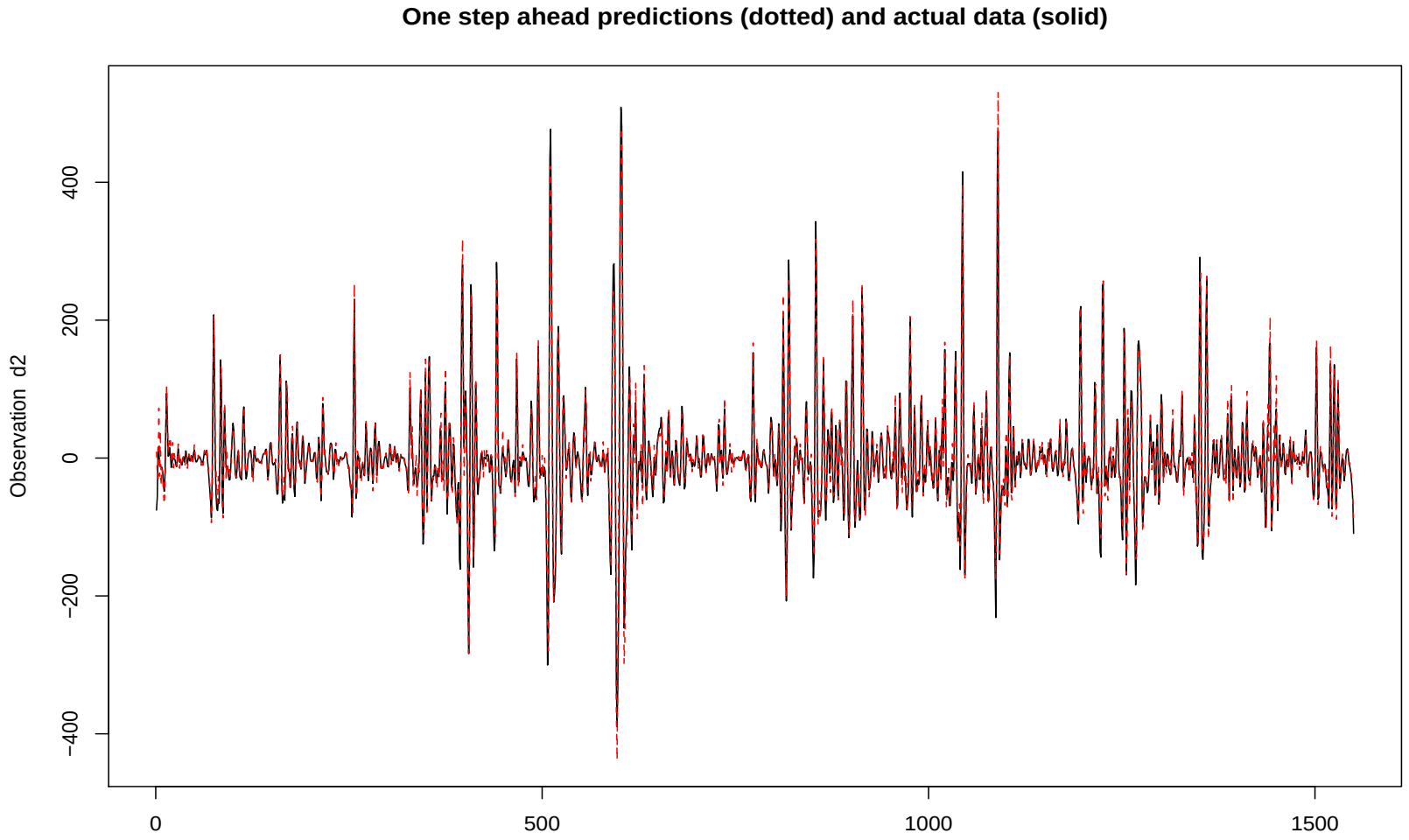
State Space Model

- A linear time-invariant state space representation in innovations form is given by:

$$z_t = Fz_t{-}1 + Gu_t + Ke_t{-}1$$
$$y_t = Hz_t + e_t$$

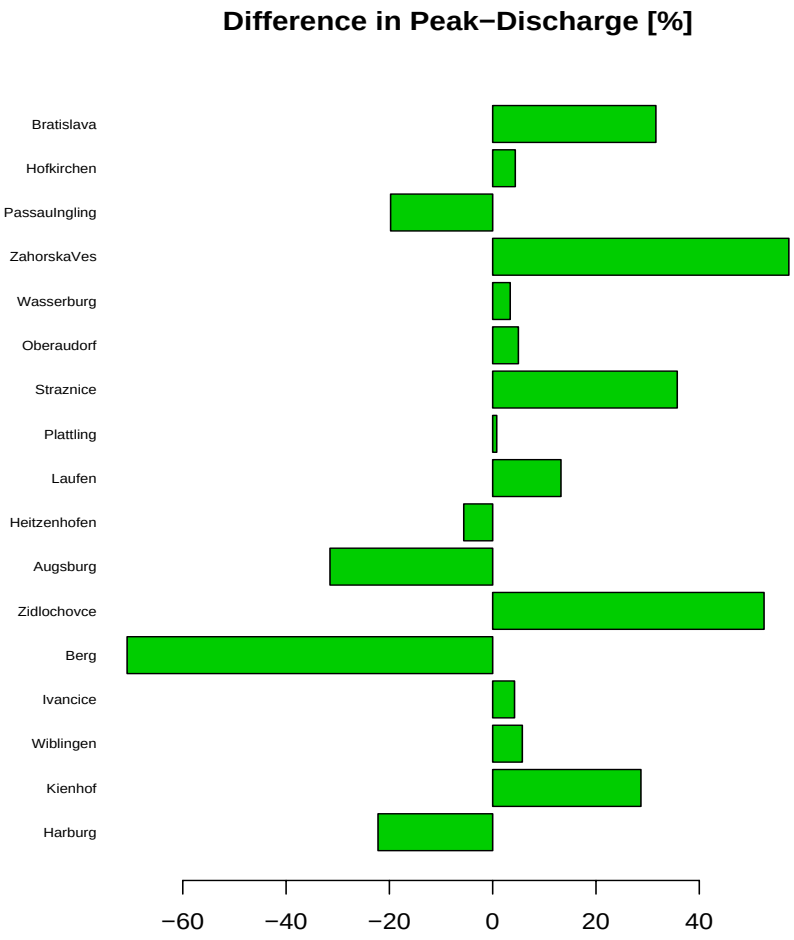
- where z_t is the unobserved underlying n dimensional state vector, F is the
- state transition matrix, G, the input matrix, H, the output matrix, and K, the Kalman gain.

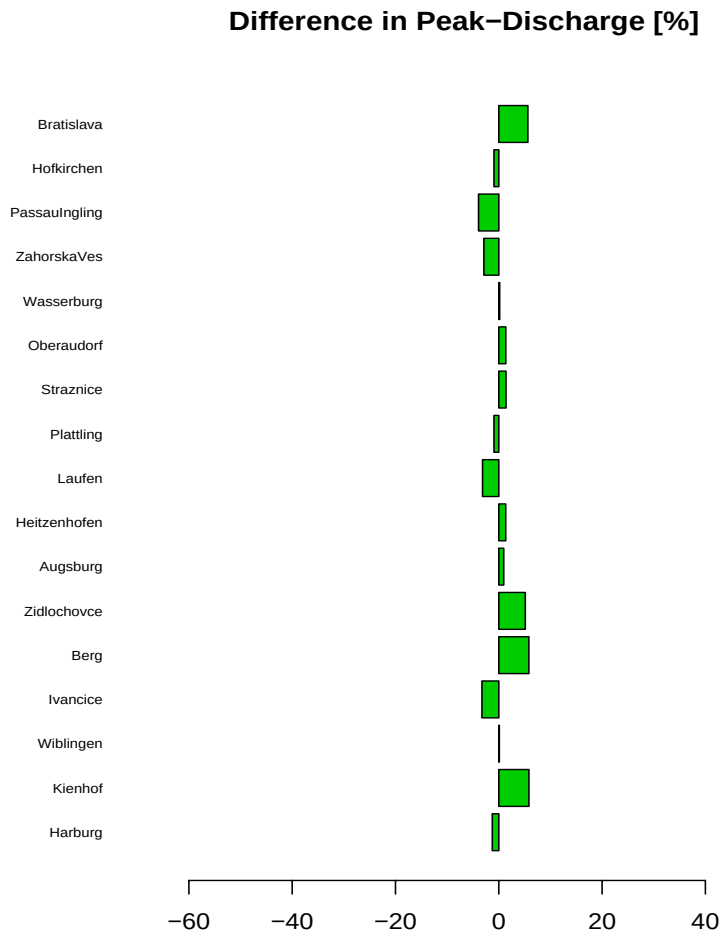
Example of one step ahead prediction for level 2



Difference in peak-discharge [%]

without bias correction and the with bias corrected forecasts (1 day ahead)

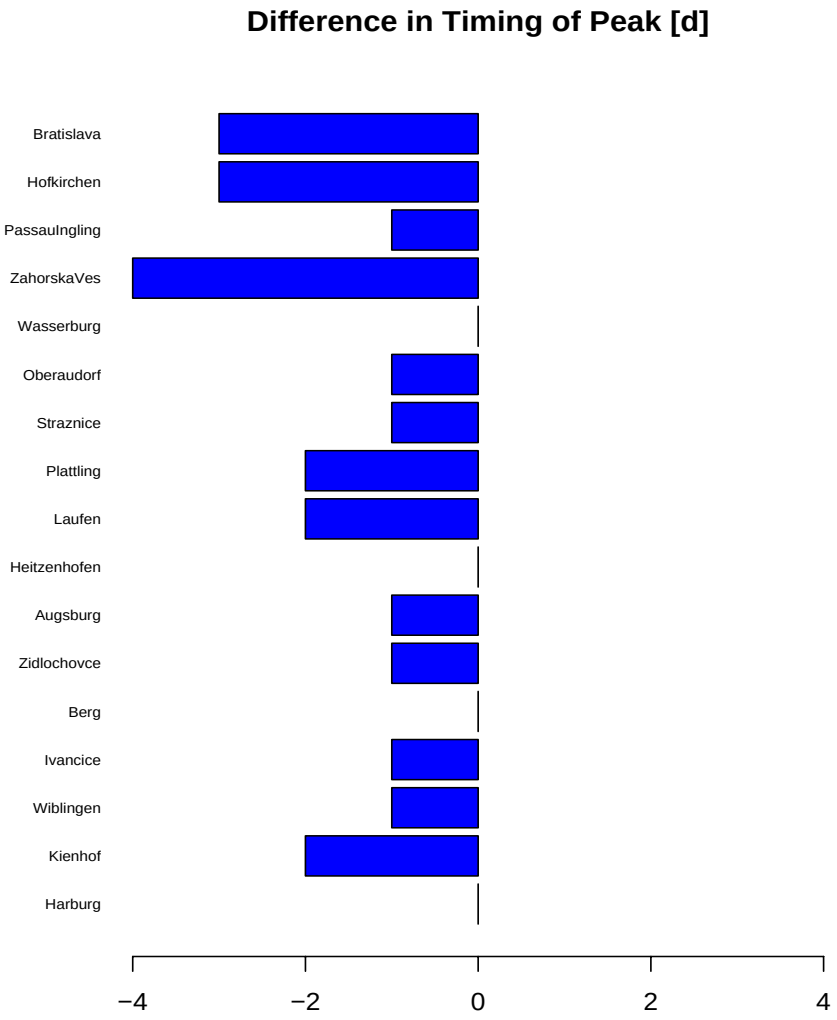


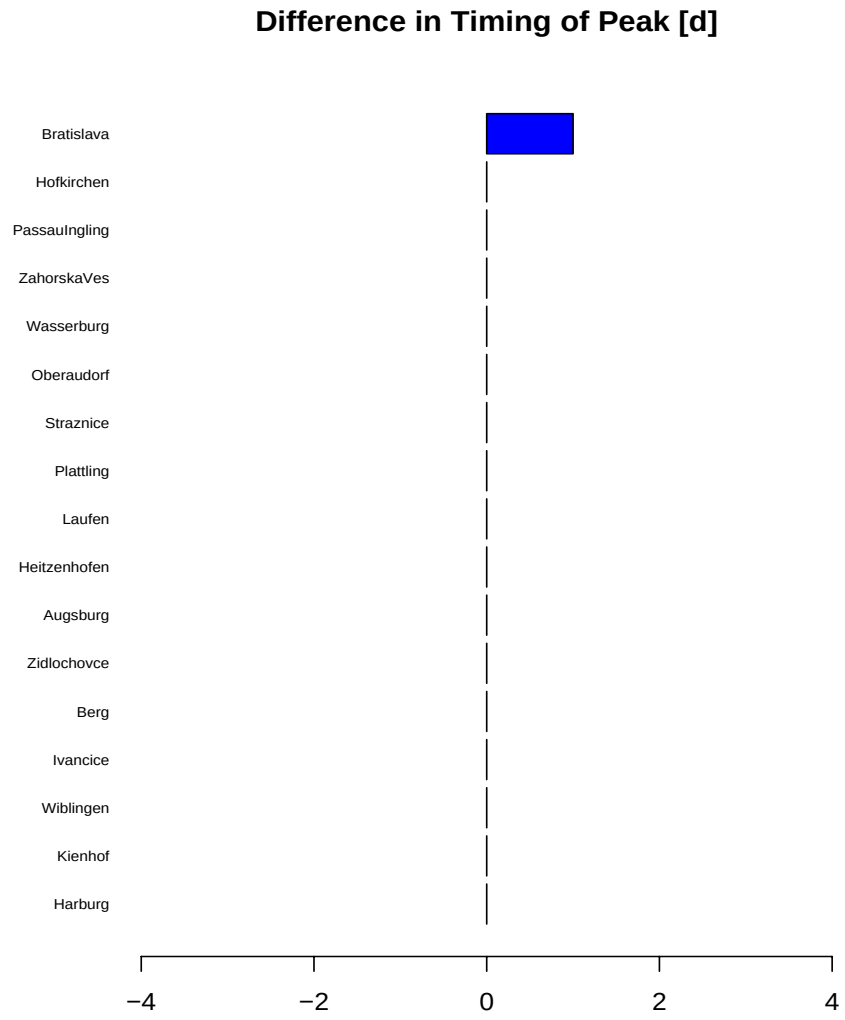


EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

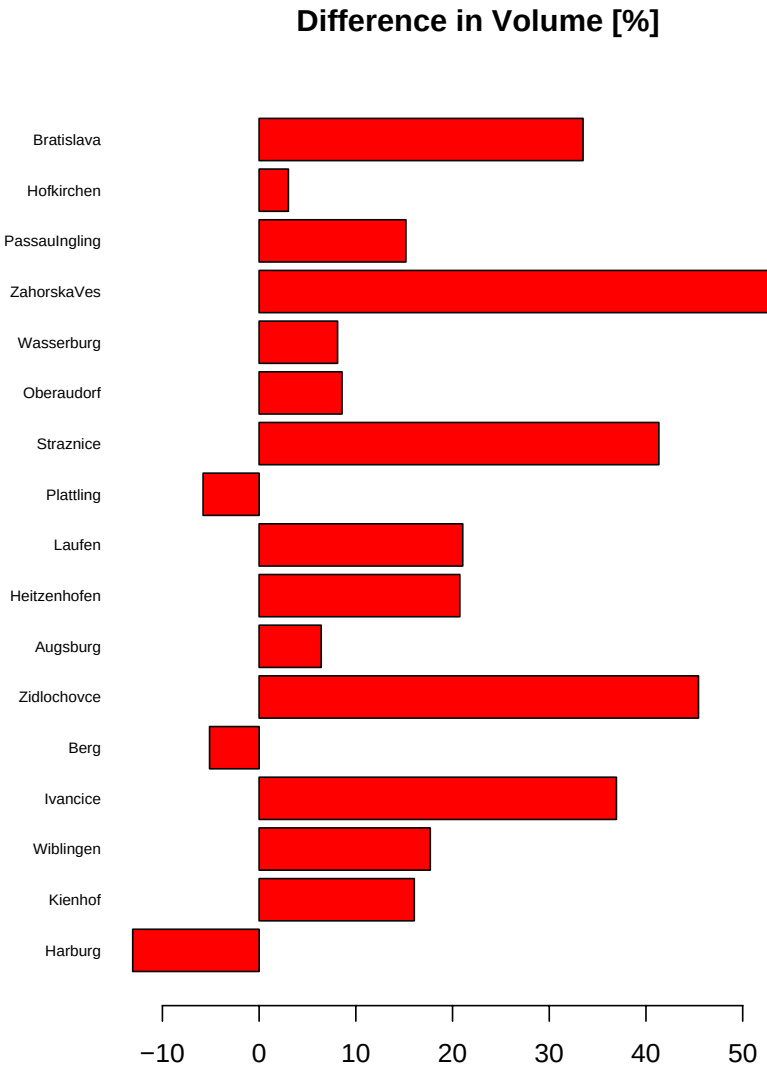
Difference in the timing of the peak [d]

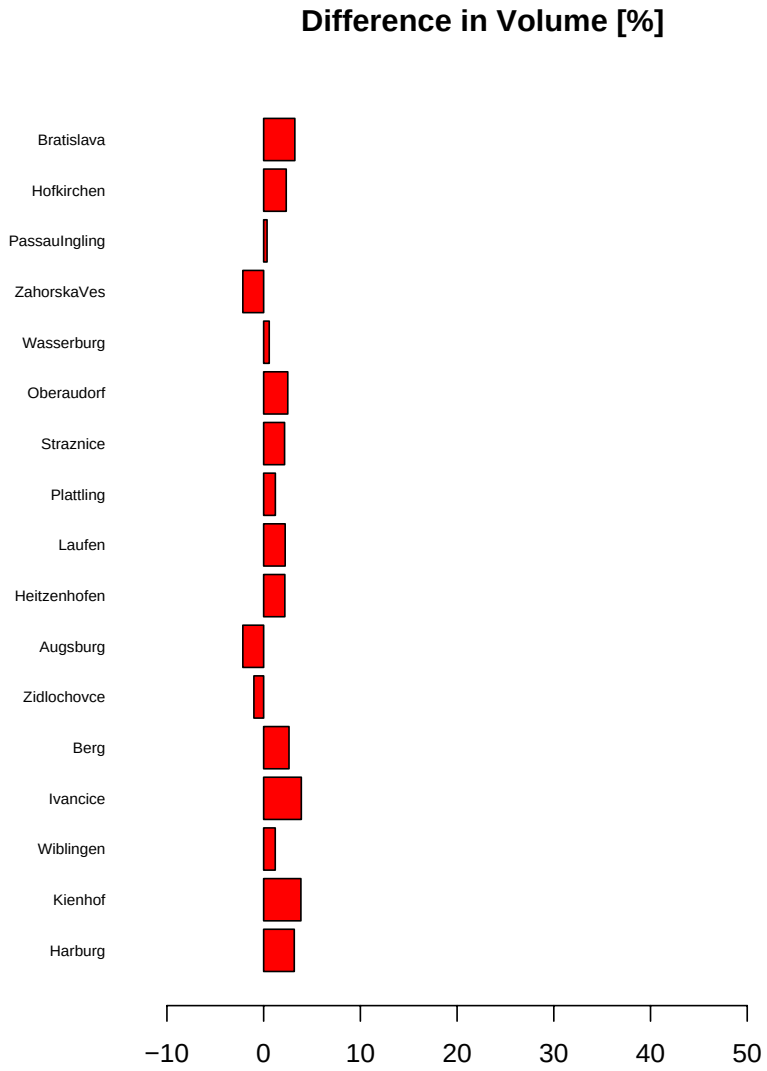
Joint Research Centre

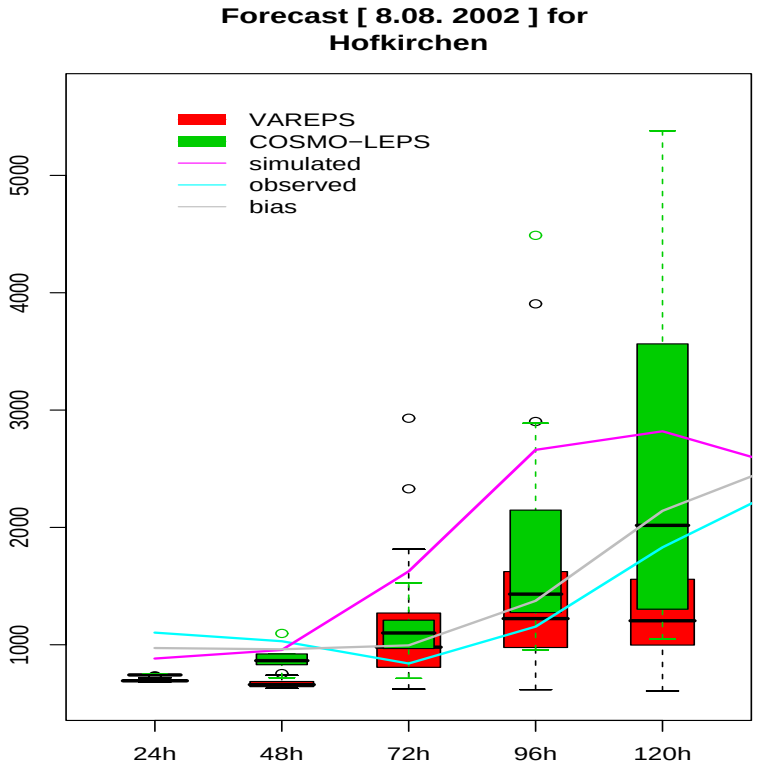


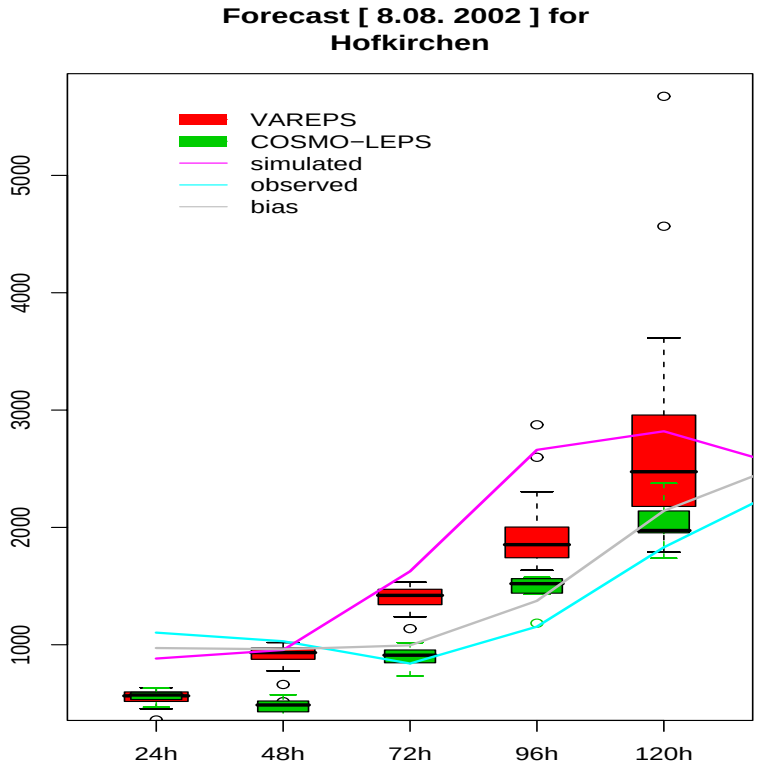


Difference in volume [%]









Example of 5 days forecast for the August 2002 flood event with and without bias corrected ensemble traces (Wavelet DLM)

Kernel based machine learning methods

- Input
  - Data represented by features
  - Features are measures describing a data point
- Output
  - Quantitative (regression) or categorical (classification) values to predict
- Training examples
  - Pairs of input and output
- Target function
  - relationship between input and output
  - unknown
  - approximated by a machine learning algorithm

Support Vector Regression

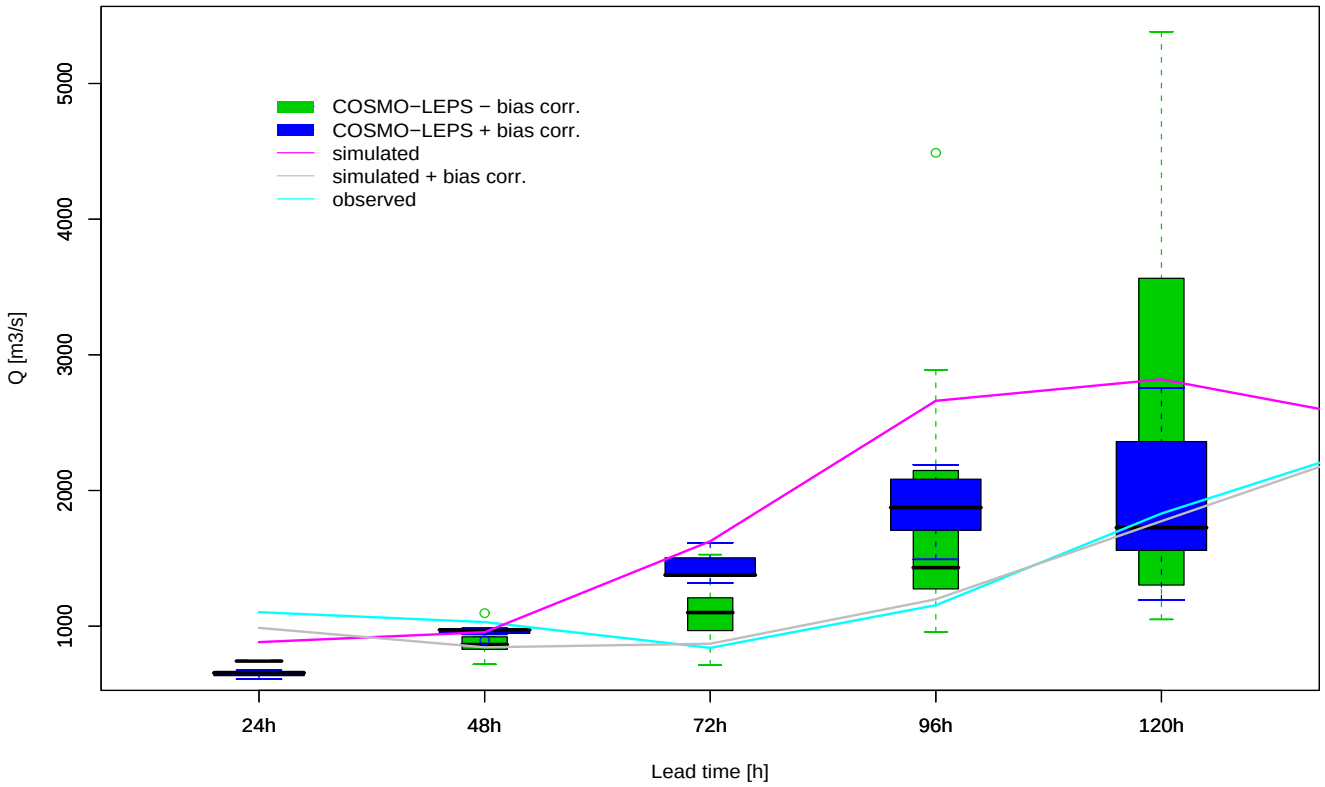
Support Vector Machine (SVM)

- - Statistical learning using support vectors in feature space
- - Presented by V. Vapnik et. al. at AT&T Bell Lab in 1995
- - **Advantages**
  - Greater generalization ability and global minima due to
  - structural risk minimization (SRM)
  - Robust in dealing with corrupted data
- - **Applications**
  - Pattern recognition, document classification, etc.

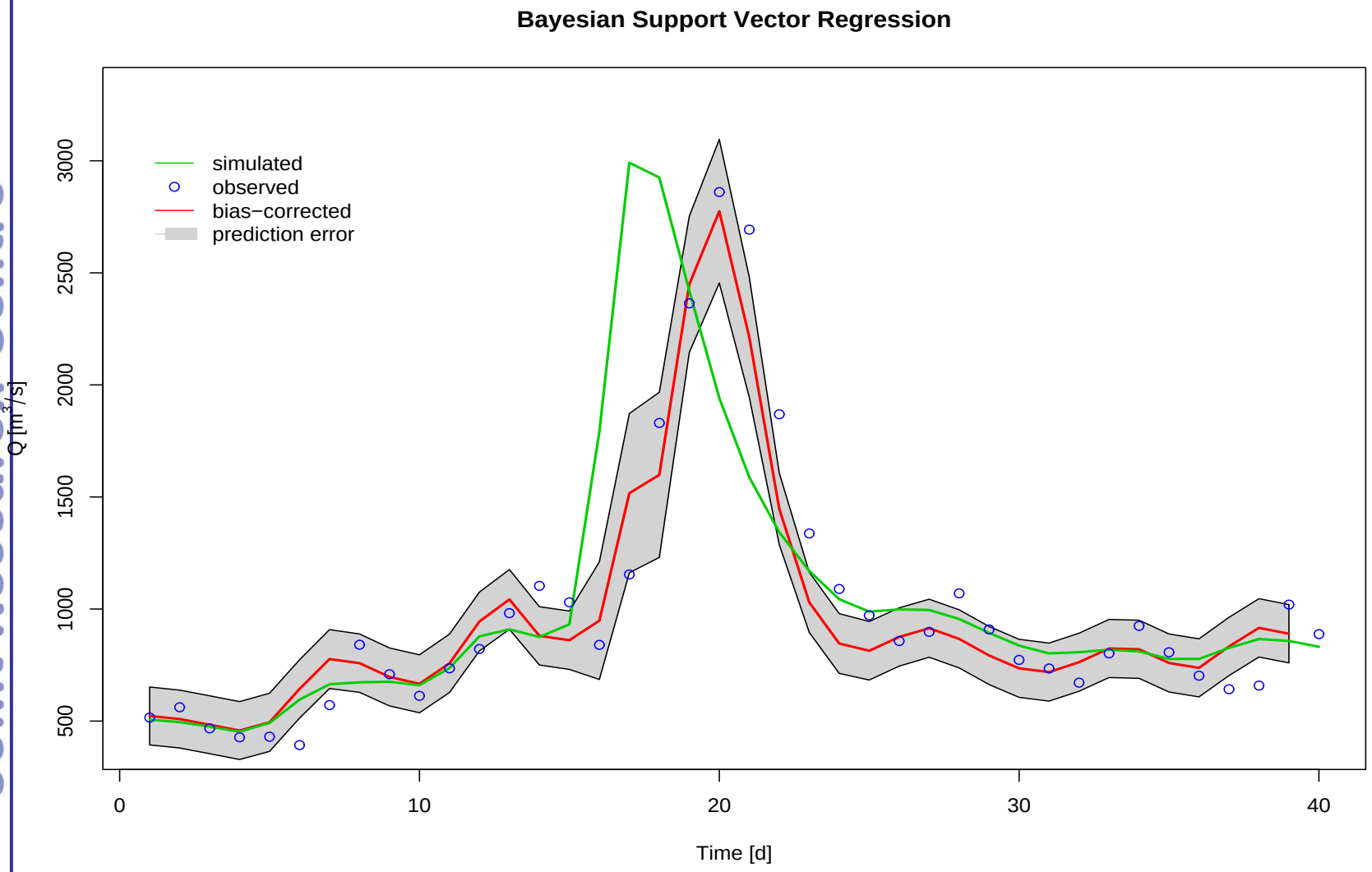
Support Vector Regression (SVR)

- - Time-series forecasting with SVM
- - **Applications**
  - Estimation of power consumption
  - Reconstruction of chaotic systems
  - Forecasting of hydrological time series (see for example:

    Pao-Shan Yu, Shien-Tsung Chen and I-Fan Chang, Support vector regression for real-time flood stage forecasting, Journal of Hydrology, Volume 328, Issues 3-4, 2006, Pages 704-716)



Example 1: SVM Regression COSMO-LEPS forecasts of August 2002



Example 2: Bayesian support vector regression

Bayesian Model Averaging

- Calibrate system PDF (variance) by training and weighting individual Member PDF
- Train member PDF against observations for past days
- Ensemble mean forecast = $\Sigma w_k f_k$

  where

  f_k = result of k^{th} model

  w_k = weight of k^{th} model, related to model's correlation with observations during training

BMA

- Estimate observations (y) via linear combination of model forecasts (f_k):

$$p(y|f_1,\ldots f_K) = \Sigma w_k g(y|f_k)$$

where

$p(y|f_1,\ldots f_K)$ = probability of a given observation based on the set of model forecasts

w_k = model weight, related to correlation w/ obs

$g(y|f_k)$ = probability of a given observation based on forecast from model k alone