ESTIMATION OF THE PARAMETERS FOR HYDROLOGICAL MODELLING

Luka Stravs, Andrej Vidmar, Mitja Brilly

Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia

Abstract: The HBV program and several forecast programs were developed by the machine learning system for the Sava River basin. The model with daily data was elaborated with system analysis and some parameters were estimated by measurements and hydrological analyses. Other parameters were estimated by calibration. The calibration was very successful with R^2 higher than 0.95. The verification yielded better results than were the results gained with calibration. The forecasts made by the machine learning system were better than the HBV model for mean flows but proved unsuccessful for flood simulation. Successful daily forecasts with ALADIN-SI data (precipitation forecasted one and two days in advance) were developed by the HBV model for two flood events in August 2002. The presented investigation was developed as a part of the EU project European flood forecast system – EFFS.
Key words: hydrological modeling, flood wave propagation, Doppler velocity meter, machine learning, HBV model

SCHÄTZUNG DER PARAMETER FÜR DIE HYDROLOGISCHEN MODELLIERUNGEN

Zusammenfassung: Das HBV-Programm und einige Vorhersageprogramme wurden mit dem System des maschinellen Lernens für das Einzugsgebiet des Flusses Sava vorbereitet. Man hat ein Modell mit täglichen Daten mithilfe der Systemanalyse ausgearbeitet und einige Parameter waren mit Messungen und hydrologischen Analysen geschätzt. Andere Parameter waren mit Kalibrieren geschätzt. Das Kalibrieren war sehr erfolgreich mit R^2 höher als 0.95. Die Resultate der Verifikation waren besser als die Resultate des Kalibrierens. Für den durchschnittlichen Abfluss waren die Vorhersagen des Systems des maschinellen Lernens besser als das HBV-Modell, sie waren aber weniger erfolgreich in Hochwassersimulationen. Mit dem HBV-Modell waren erfolgreiche tägliche Vorhersagen mithilfe der ALADIN-SI-Daten hergestellt für zwei Hochwasserereignisse im August 2002. Die Erforschung wurde im Rahmen des EU-Projekts „Das europäische Hochwasservorhersagesystem" (European Flood Forecasting System – EFFS) ausgeführt.
Schlüsselworte: hydrologische Modellierung, Abflussbeschleunigung, Doppleranemometer, maschinelles Lernen, HBV-Modell

1. Introduction

The Sava River is a tributary of the Danube River. The head part of the Sava River basin is located in Slovenia (Figure 1). The IHMS-HBV model was chosen as a hydrological model to be calibrated, verified and used for flow forecasting on the part of the Sava River basin area contributing to the discharge at water station Hrastnik. Urban areas in some parts of the modeled catchment are located in narrow valleys where the settlements as well as the traffic infrastructure are endangered by the possibility of flooding.

River basins are highly complex systems that typically demonstrate the need for estimation of a great number of parameters when simulating hydrological processes. The parameters of river basin models are determined by appropriate measurements and in the process of calibration (Duan et al., 2003). Our aim was to not only calibrate the chosen model by the trial-and-error method, but to estimate some of the initial values of the calibration parameters of the IHMS-HBV model for the selected part of the Sava River basin by actual measurements and their analysis.

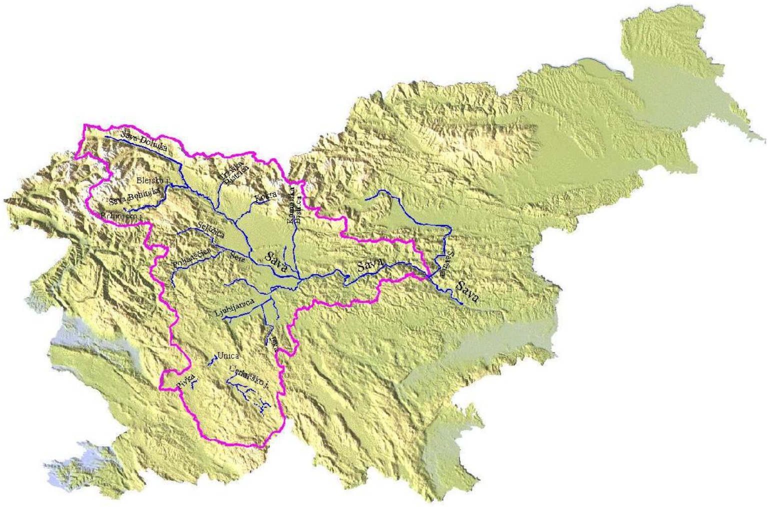

Figure 1. Geographical location of the modeled district on map of Slovenia

2. Calibration of the IHMS-HBV model

Selected part of the Sava River basin (Figure 1) was divided into 5 subbasins with area ranging from 118.04 km^2 to 1883.78 km^2 (Figure 2). Tables 2-4 present basic data about subbasins and meteorological and hydrological stations used for calibration and verification of the model.

Recorded discharge data from water station Medno was used as inflow to the subbasin SA01, discharge data from water stations Moste and Vir was used as outflow from subbasins LJ and KB01 and discharge data from water station Hrastnik was used as total outflow from the district (Figure 2).

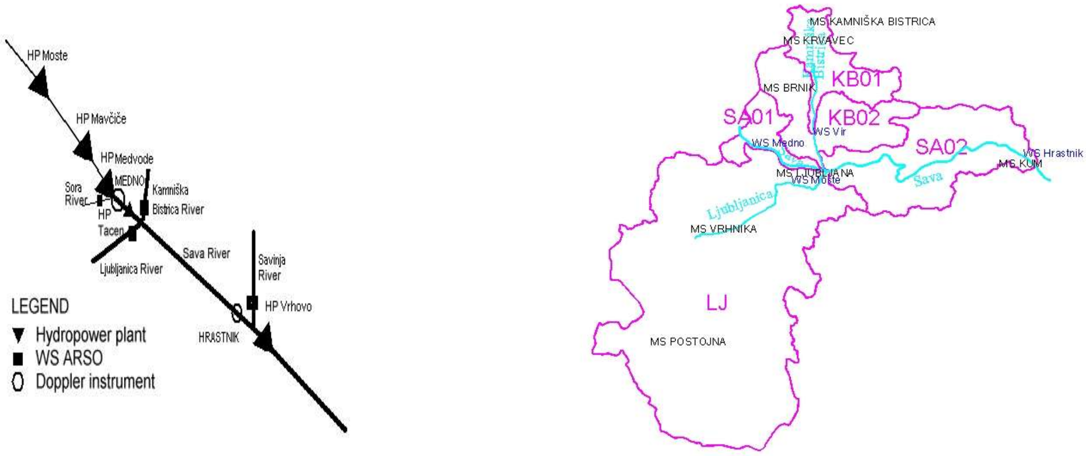

Figure 2. The scheme of the Sava River and modeled district (divided into subbasins) and distribution of water (blue) and meteorological (black) stations

Discharge, precipitation, temperature and evapotranspiration data for period from 31.10.1993 to 31.12.1997 (4 years and 2 months) was used for calibration of the model. Verification of the calibrated model was performed on data from 01.01.1998 to 30.12.1999 and on data for the period of the Central European floods in July-August 2002.

As an initial approximations of some of the IHMS-HBV calibration parameters results from the analysis of the recession periods of the daily discharge hydrographs and results from the empirically estimated operational and flood wave velocity diagrams were used. Other HBV parameters were determined manually by visual inspection and analysis of R^2 and accumulated difference from the calibration results.

Table 1. List of subbasins with their land cover type distribution

		Total area		Forest area		Field area		Water area	
Subbasin	**HBV name**	**[km^2]**	**[%]**	**[km^2]**	**[%]**	**[km^2]**	**[%]**	**[km^2]**	**[%]**
Kamniška Bistrica (WS Vir)	KB 01	207.80	100.00	121.15	58.30	86.24	41.50	0.42	0.20
Kamniška Bistrica (confluence)	KB 02	330.86	100.00	192.89	58.30	137.31	41.50	0.66	0.20
Ljubljanica (confluence)	LJ	1883.78	100.00	1248.63	66.28	610.98	32.43	24.17	1.28
Ljubljanska Sava (confluence)	SA 01	118.04	100.00	68.82	58.30	48.99	41.50	0.24	0.20
Litijska Sava	SA 02	523.49	100.00	360.25	68.82	158.35	30.25	4.89	0.93

Table 2. List of meteorological stations

Meteorological station	Altitude (m a.s.l.)	Type of data used in HBV model
LJUBLJANA	299	Precipitation, temperature, evapotranspiration
BRNIK	362	Temperature
KRVAVEC	1740	Temperature
KUM	1218	Precipitation
POSTOJNA	533	Temperature, evapotranspiration
VRHNIKA	293	Precipitation
KAMNIŠKA BISTRICA	610	Precipitation

Table 3. List of hydrological (water) stations

Station code	Water station	River	Altitude (m a.s.l.)	Area [km^2]
3530	MEDNO	Sava	301.473	2191.4
5080	MOSTE	Ljubljanica	280.798	1762.5
4430	VIR	Kamniška Bistrica	301.203	207.8
3725	HRASTNIK	Sava	195.077	5176.8

2.1 Hydrograph recession analysis

Simple and fast semi-automated analysis of hydrograph recession periods based on matching strip method using only daily discharge data for the 1991-1999 period was performed for nine main Slovenian Upper and Middle Sava River tributaries (Sava Dolinka, Radovna, Sava Bohinjka, Trziska Bistrica, Kokra, Sora, Kamniska Bistrica, Ljubljanica, Savinja rivers). Master recession curves for each tributary were formed from the selected hydrograph recession data (Figures 3-5). For each master recession curve the values of parameters of the exponential function that cause the function to best fit the master recession curve were determined. Obtained results of the analysis were used as a first approximation of the IHMS-HBV K4 parameter for each the Sava River tributaries in the modeled district (Ljubljanica and Kamniska Bistrica rivers) in the calibration of the model.

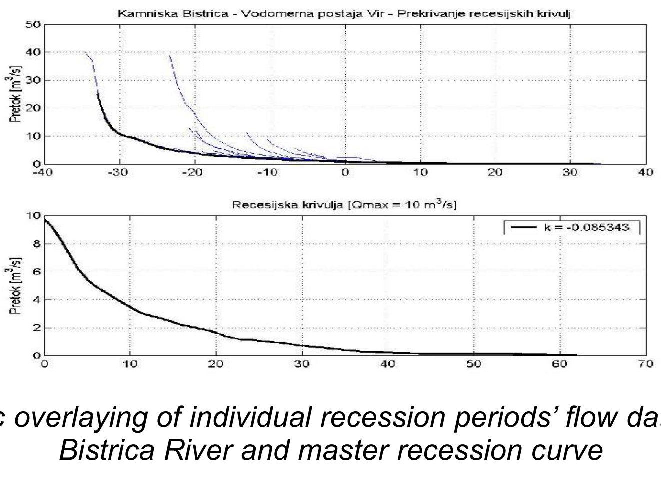

Figure 3. Automatic overlaying of individual recession periods' flow data for the Kamniska Bistrica River and master recession curve

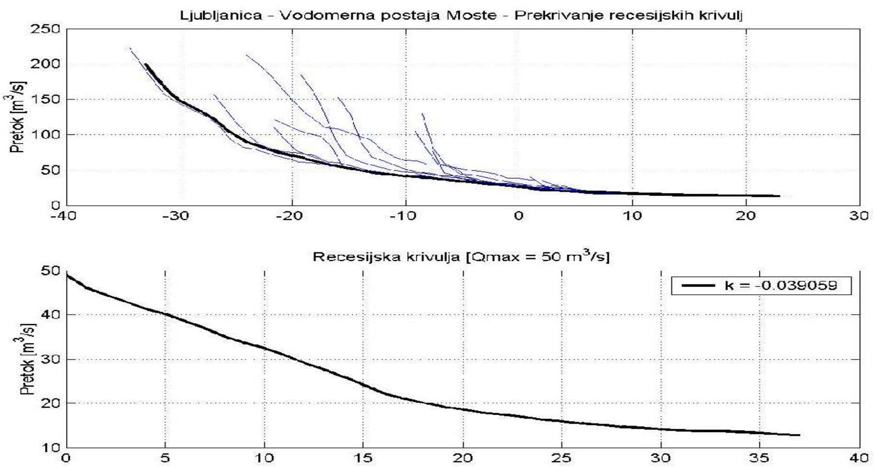

Figure 4. Automatic overlaying of individual recession periods' flow data for the Ljubljanica River and master recession curve

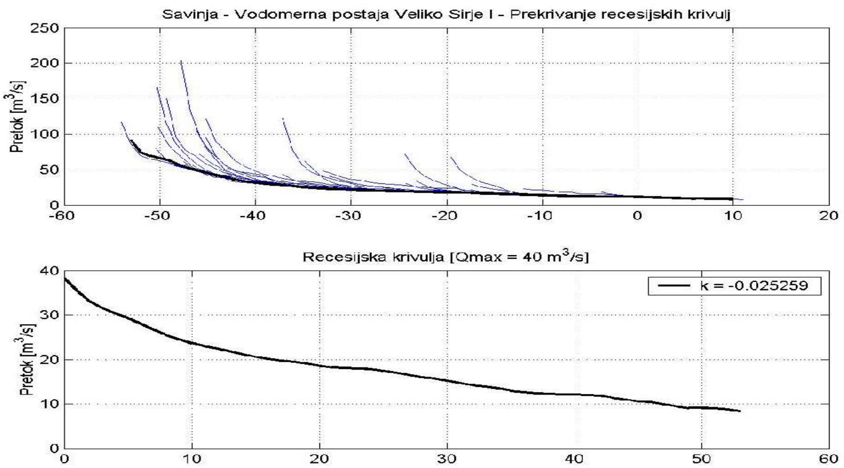

Figure 5. Automatic overlaying of individual recession periods' flow data for the Savinja River and master recession curve

Some of the obtained models for flow forecasting during rainless periods of the year for n-days ahead:

- Ljubljanica River: $Q_{t+n} = Q_t * e^{-0.039059 * n}$
- Kamniška Bistrica River: $Q_{t+n} = Q_t * e^{-0.085343 * n}$
- Savinja River: $Q_{t+n} = Q_t * e^{-0.025259 * n}$

2.2 Channel routing velocity estimation

Channel routing of flood waves is an important process in hydrological modelling aiming to manage the in-stream flows exercising a strong influence on flood peak movements. The channel routing velocity is a function of the in-stream discharge and its respective increase or decrease. Estimation of the channel routing velocity is made by observation of flood peak discharges, however the peak discharge is also influenced by the composition of the river basin network, and topography and rainfall distribution. Time lag of flood wave peaks mostly depends on the channel routing velocity, and to some extent it is allocated to other model parameters. Overestimating or underestimating of the channel routing velocity can result in a significant impact on the estimation of many other parameters in the calibration procedure.

The experimental section of the Sava River was between the Medno water station, 3.5 kilometres downstream of the Medvode hydro power plant and the Hrastnik water station situated upstream of the major tributary, namely the Savinja River (Figure 2). Downstream of the Medno water station is a small Tacen hydro power plant situated on natural runs. Water stations are at a distance of 66.5 kilometres, with the absence of any noteworthy structures or inundated areas in the immediate area of one hundred meters from the stream. Width of the stream bottom of the cross-section at the Medno water station is 50 meters and at the Hrastnik water station it is 54 meters. The elevation difference between the water stations is 105 meters. Two Doppler 1D flowmeters were mounted on rods of the water stage gauges at water stations Medno and Hrastnik. The instruments were installed twenty centimetres above the bottom and oriented downstream, water depth, velocity and temperature data were logged every five minutes. The measurements were carried out in the time period from July to November 2002.

Hydrographs of seven of the recorded flood waves (Table 4) at both water stations were separated on the rising and falling limb and the time of the operational waves' propagation between water stations Medno and Hrastnik was calculated by cross-correlating the recorded discharges at both water stations. Time lag at the maximum value of the cross-correlation function was estimated to be the time of propagation of operational and flood waves used in the analysis (Figure 6).

Table 4. Recorded flood waves' characteristics

	Flood wave duration				**Flood wave peak**			
	Start		**End**		**WS Medno**		**WS Hrastnik**	
Flood wave #	**Date**	**Time**	**Date**	**Time**	**Water stage [m]**	**Discharge [m³/s]**	**Water stage [m]**	**Discharge [m³/s]**
01	04/07/2002	18:35	08/07/2002	7:15	1.99	182	2.90	172
02	14/07/2002	22:20	18/07/2002	16:10	2.51	322	3.42	268
03	24/07/2002	0:35	27/07/2002	16:20	1.47	81	2.84	162
04	10/10/2002	17:00	17/10/2002	16:00	1.91	164	3.76	338
05	17/10/2002	22:40	22/10/2002	6:15	2.96	468	4.78	550
06	22/10/2002	6:20	31/10/2002	9:00	2.76	400	5.23	650
07	03/11/2002	19:00	10/11/2002	18:15	2.13	215	3.46	276

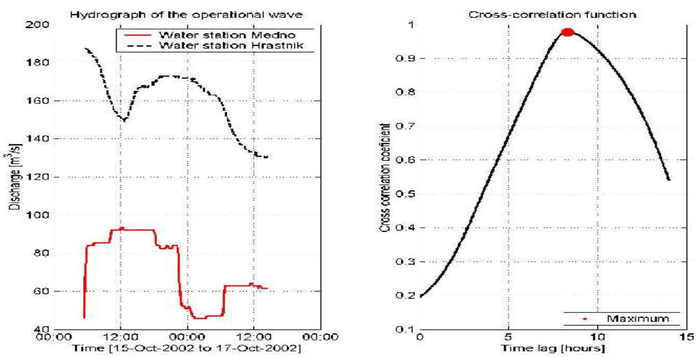

Figure 6. Estimation of the time of the propagation of the operational wave on the falling limb of the flood wave hydrograph by cross-correlation of the recorded discharge data at water stations Medno and Hrastnik

Empirically determined relation of discharge rate at both water stations and time of operational and flood waves' propagation between water stations Medno and Hrastnik

(Figures 7, 8) clearly show non-negligible differences in the channel routing velocity on the rising and falling limb of the flood waves. Suprisingly, time of propagation of the operational waves on the falling limb of the flood waves was shorter (routing velocity was higher).

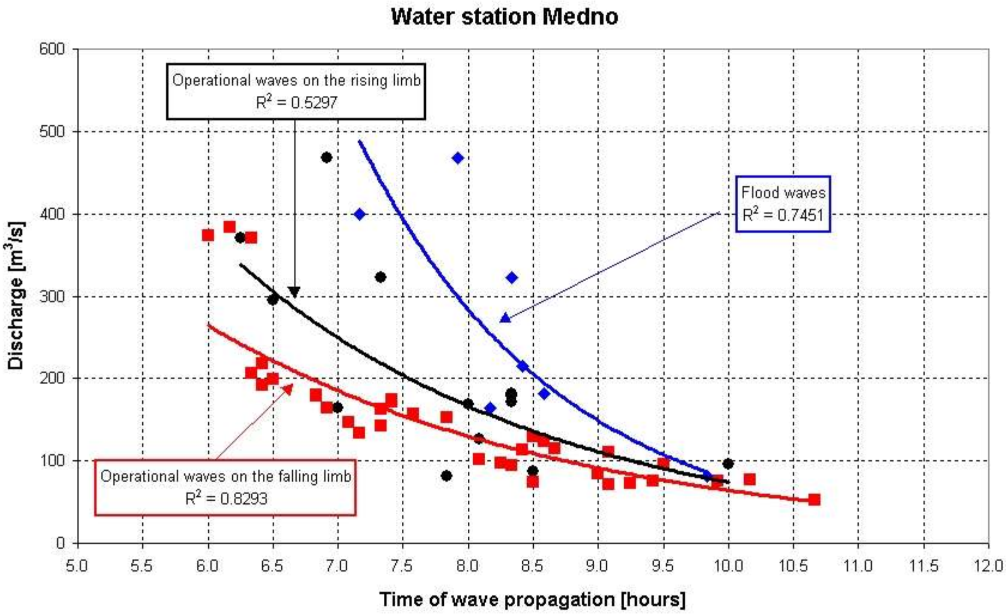

Figure 7. Empirically determined relation of discharge rate at the Medno water station and time of operational and flood waves' propagation between water stations Medno and Hrastnik

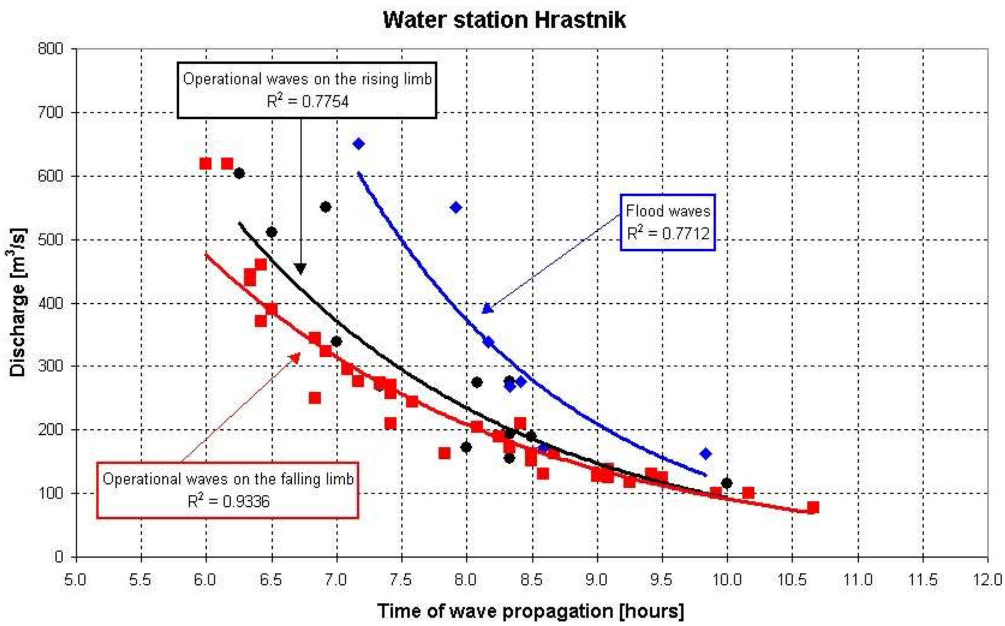

Figure 8. Empirically determined relation of discharge rate at the Hrastnik water station and time of operational and flood waves' propagation between water stations Medno and Hrastnik

3. Model calibration and verification results

Calibration and verification results are shown in Table 5 and Figures 9-11. Verification results for year 1998 only are much better in comparison to those from years 1998 and 1999 together, the improvement of verification results for year 1998 only is especially high for Ljubljanica River subbasin. For the period from June to December 1999 calibrated IHMS-HBV model constantly overestimated low flows, which was not the case in 1998 (at least not so constantly and for such a long time period).

Table 5: Model calibration and verification results

	R^2		
	Calibration (1994-1997)	**Verification (only year 1998)**	**Verification (years 1998 – 1999)**
KB01 (subbasin)	0.77	0.81	0.77
LJ (subbasin)	0.76	0.90	0.78
SA02 (district)	0.87	0.92	0.91

Obtained results for both calibration and verification are quite satisfactory, introduction of additional calibration data and additional subdivision of Ljubljanica River subbasin would probably gain even higher values for HBV efficiency criteron R^2.

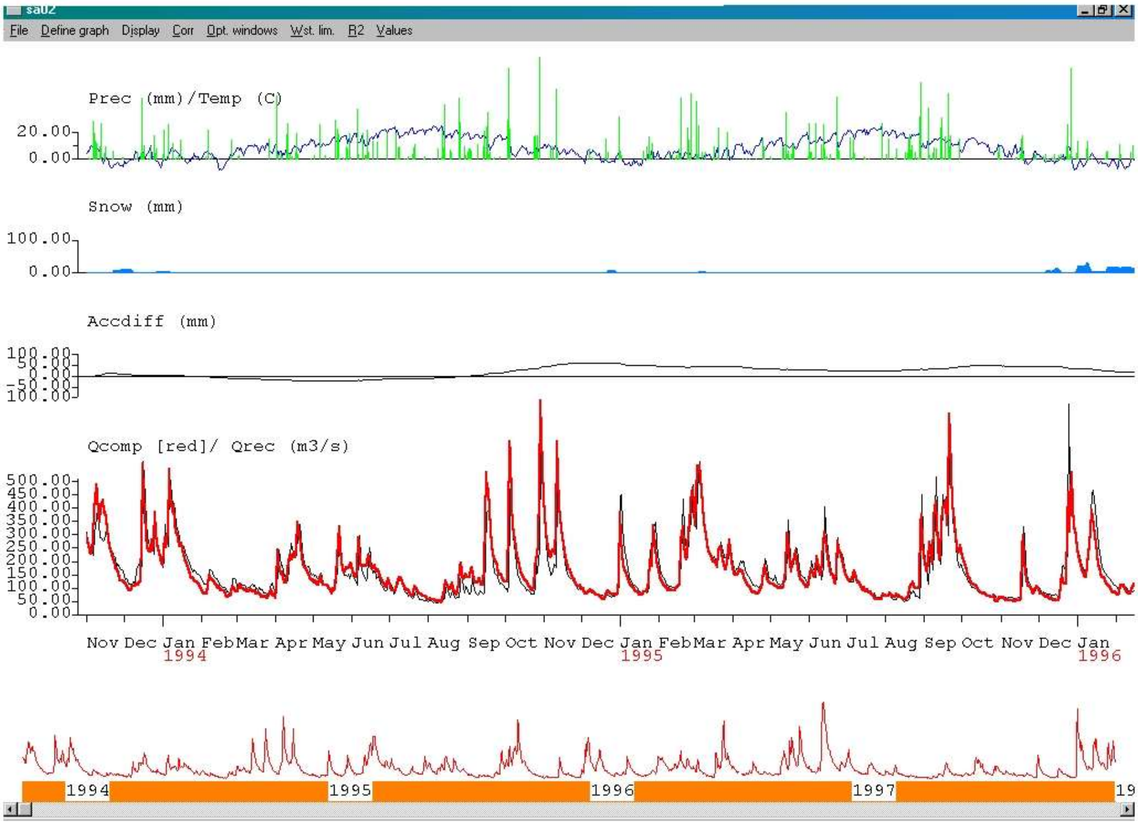

Figure 9. Model calibration results for SA02 (1994-1995)

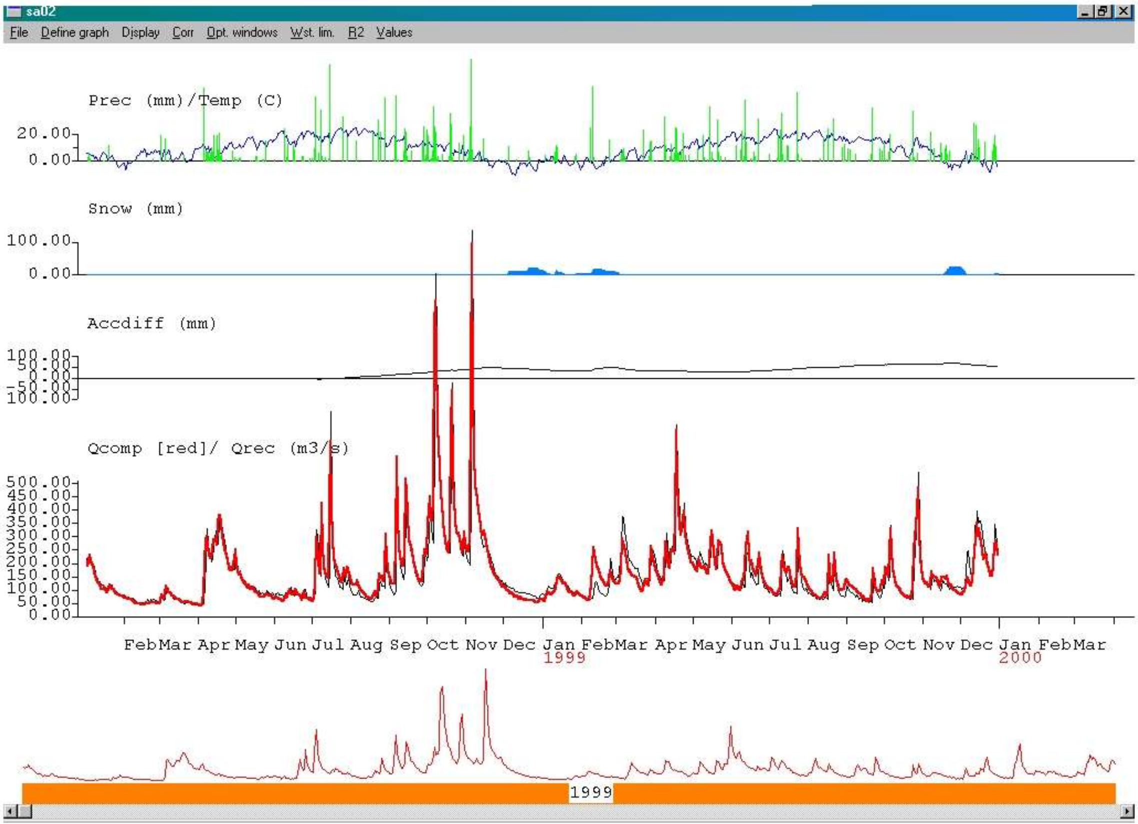

Figure 10. Model verification results for SA02 (1998-1999)

Figure 11. Model verification results for SA02 (July-August 2002 – time of Central European floods)

4. Comparison of the performance of the HBV model and forecasting models obtained by linear regression and M5 machine learning method

By using the data from the same time intervals (calibration period 1994 –1997, verification period 1998-1999 and July-August 2002) and the same hydrological and meteorological stations simple 1- and 2-day ahead forecasting models for daily discharge at water station Hrastnik were built. First type of forecasting models was built by linear regression, the other type by machine learning method M5 (Kompare et al, 1997) as implemented in a system called WEKA, developed at the University of Waikato in New Zealand.

The machine learning method M5 (Breiman et al., 1984 and Karalič 1992) has been successfully used in hydrology for runoff modelling (Kompare et al., 1997). The M5 method constructs models with the presupposition that linear relations with the same parameters are not valid for all ranges of data. The model automatically sequences data into subspaces, determined by different ranges of input data (attributes), and determines best-fit linear equations for each of these subspaces. The input for modelling is the predetermined logical structure as well as chosen model attributes and the necessary test data.

Input data for the 1- and 2-day ahead forecasting models was the last two days' discharge data from all hydrological stations (Table 3) except for Hrastnik and the last two days' precipitation and evapotranspiration data from all meteorological stations (Table 2).

Obtained linear regression forecasting models:

Obtained M5 forecasting models:

```
LM3M1 SavaMedno_
Q_1 <= 67.7 : [1]
|   Ljubljanica_Q_1 <= 42.1 : LM1
|   Ljubljanica_Q_1 >  42.1 : LM2
SavaMedno_Q_1 >  67.7  :_M5 - forecasting model for 1 day ahead:
```

Forecasting models at the leaves:

```
LM1:   SavaHrastnik_Q = 13.3 - 0.00451SavaMedno_Q_2 + 0.941SavaMedno_Q_1 -
       0.00662Ljubljanica_Q_2 + 1.15Ljubljanica_Q_1 + 1.18KamniskaB_Q_1 -
       0.00681KamniskaB_P_2 + 0.598KamniskaB_P_1 + 0.0286Ljubljana_P_1 -
       0.0078Vrhnika_P_2 + 0.785Vrhnika_P_1 - 1.68Postojna_E_2 - 0.0958Ljubljana_E_2
       + 0.0554Ljubljana_E_1

LM2:   SavaHrastnik_Q = 18.9 - 0.00451SavaMedno_Q_2 + 1.03SavaMedno_Q_1 -
       0.26Ljubljanica_Q_2 + 1.2Ljubljanica_Q_1 - 2.81KamniskaB_Q_2 +
       4.39KamniskaB_Q_1 - 0.00681KamniskaB_P_2 + 2.5KamniskaB_P_1 - 1.19Kum_P_2 +
       0.0477Ljubljana_P_1 - 0.0078Vrhnika_P_2 + 0.0372Vrhnika_P_1 -
       11.6Postojna_E_2 + 8.7Postojna_E_1 - 0.167Ljubljana_E_2 + 0.0554Ljubljana_E_1

LM3:   SavaHrastnik_Q = 28.4 - 0.262SavaMedno_Q_2 + 0.999SavaMedno_Q_1 +
       0.913Ljubljanica_Q_1 + 1.72KamniskaB_Q_1 - 0.743KamniskaB_P_2 +
       1.98KamniskaB_P_1 + 0.0249Ljubljana_P_1 - 0.0109Vrhnika_P_2 + 1.56Vrhnika_P_1
       - 0.0812Postojna_E_2 - 0.078Ljubljana_E_2 + 0.0771Ljubljana_E_1
```

Notation:

_Q_ means discharge data from water station, _P_ means precipitation and _E_ means evapotranspiration data from meteorological station;

_1_ means data from one day ago, _2_ means data from two days ago and _3_ means data from three days ago;

Example: SavaMedno_Q_2 means discharge data from two days ago

Comparison of the performance of the built forecasting models and calibrated IHMS-HBV model was done by the means of testing the HBV efficiency criterion R^2, correlation coefficient r (r^2),

---

[1] This particular model reads as:

if discharge at water station Medno (Sava River) one day ago is lower than or equal to 67.7
then
  if discharge at water station Moste (Ljubljanica River) one day ago is lower than or equal to 42.1
    then use model LM1 for prediction
    else use model LM2 for prediction
  end
else use model LM3 for prediction
end

mean absolute error and mean relative absolute error. 1-day ahead M5 forecasting model's performance ranked the best in all validation categories (Table 6). Only by visual comparison of the computed and measured hydrographs (Figure 12,13) it becomes obvious that M5 and linear regression forecasting model seriously lack in prediction of the time of the flood wave peaks. Conceptual models like IHMS-HBV perform much better in prediction of the time of the flood wave peaks which is very

Table 6. Comparison of the modeling results for calibration (1994-1997) and verification data (1998 - 1999) at the Hrastnik water station

	R^2 (Nash Sutcliffe)		r (correlation coeficient)		r^2		Mean absolute error [m^3/s]		Mean relative absolute error [%]	
Model	Calib.	Verif.	Calib.	Verif.	Calib.	Verif.	Calib.	Verif.	Calib.	Verif.
IHMS – HBV	0.87	**0.92**	0.94	0.96	0.88	0.91	26.57	22.88	18.23	16.34
Linear Regression - Forecast 1 day ahead	0.94	0.91	0.97	0.96	0.94	0.91	15.77	19.97	10.32	12.33
Linear Regression - Forecast 2 days ahead	0.69	0.61	0.83	0.78	0.69	0.61	31.04	33.94	19.24	20.01
Machine learning method M5 - Forecast 1 day ahead	**0.95**	**0.92**	0.97	0.96	**0.95**	**0.93**	**13.47**	**16.99**	**7.92**	**9.09**
Machine learning method M5 - Forecast 2 days ahead	0.70	0.62	0.84	0.79	0.70	0.63	29.71	31.99	16.57	16.24

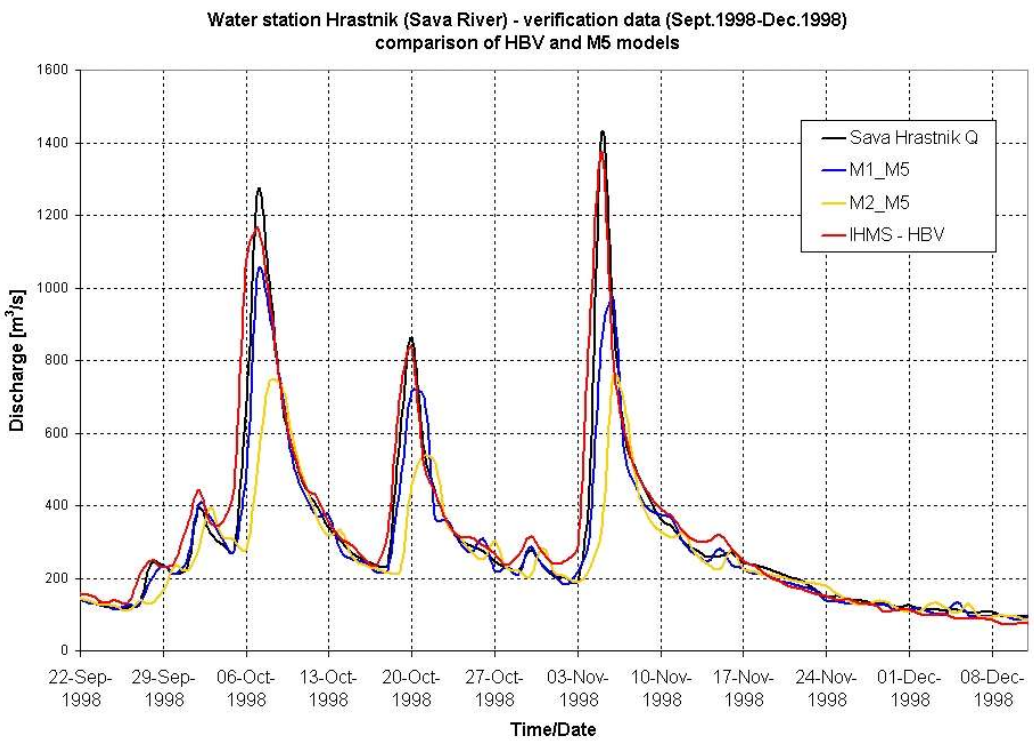

Figure 12. Comparison of the performance of the IHMS-HBV model and M5 (machine learning method) 1- and 2-day ahead forecasting models (verification data Sept. - Dec. 1998)

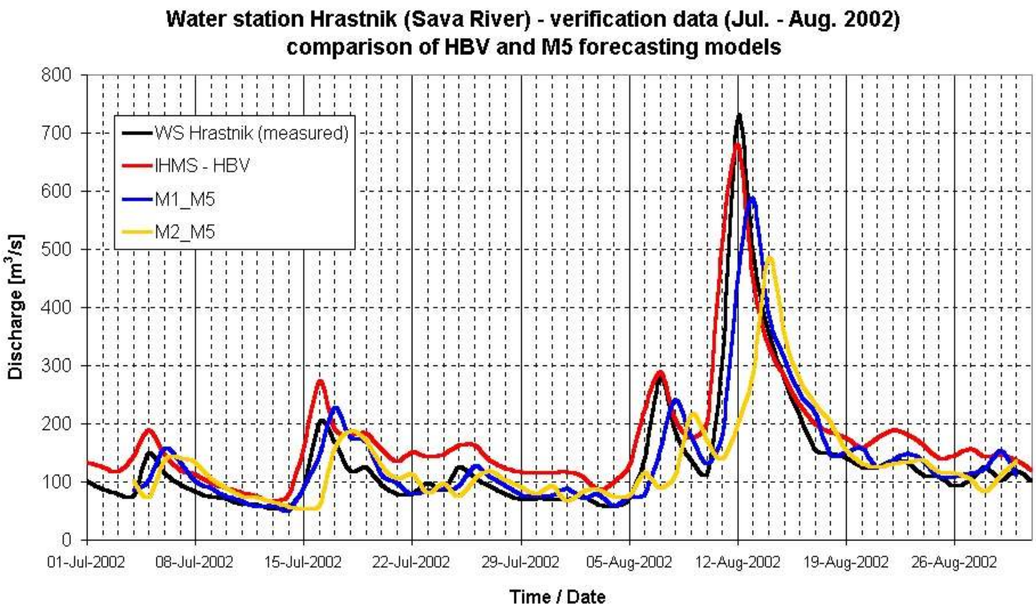

Figure 13. Comparison of the performance of the IHMS-HBV model and M5 (machine learning method) 1- and 2-day ahead forecasting models (verification data Jul. - Aug. 2002)

5. Conclusions

Classic conceptual hydrological IHMS-HBV model for the selected part of the Sava River basin was calibrated not only by the-trial-and-error method, but also on the basis of results of measurements of hydrological processes (flow routing) and baseflow recession analysis in the selected river basin, done before the start of the calibration procedure. Results of additional measurements of hydrological processes were used as first approximations of the calibration parameters. Knowledge about actual hydrological processes in the modeled river basin can help us prevent situations where really good modeling results are achieved but the calibrated parameters of partial processes in the river basin do not represent the actual river basin interactions, which can be easily established by appropriate measurements. By performing hydrological measurements with Doppler 1D flowmeter at water stations Medno and Hrastnik additional knowledge about different water flow velocities on the rising and falling limbs of the flood waves in the selected part of the Sava River basin was aquired, neglecting the influence of which can result in underestimation of peak discharges during highly dynamic floods and less accurate estimation of the time of the flood peaks.

Using machine learning methods for flow forecasting (hindcasting) presents quite an useful alternative to classic hydrological modeling, especially from the view of reducing man-power spent in the process of calibration of the hydrological model. Analysis of only statistical perfomance indicators, when comparing conceptual and machine learning built hydrological models, can mislead us in concluding that machine learning models can easily outperform classical hydrological models. Only by visual comparison of the computed and measured hydrographs it becomes obvious that machine learning forecasting models seriously lack in prediction of the time of the flood wave peaks. Conceptual models like IHMS-HBV perform much better in prediction of the time and value of the flood wave peaks which is very (if not the most) important in flow forecasting.

6. References

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984): *Classification and regresion Trees.* Wadworth, Belmont.

Chow V.T., Maidment D.R., Mays L.W., (1988): *Applied hydrology*. McGraw Hill, Civil Engineering Series.

Duan Q., Gupta H.V., Sorooshian S., Rousseau A.N., Turcotte R., (2003): *Calibration of watershed models*. Water Science and Application 6, AGU, Washington.

Karalič, A. (1992): RETIS – *A knowlage acquisition system. User's manual. Software version 2.07/nf*, Jožef Stefan Institute, Ljubljana, Slovenia.

Kompare, B., I., Steinman, F., Cerar, U., and Džeroski, S.,(1994): *Prediction of rainfal runoff from catchment by data analysis with machine learning tools within artificial intelligence tools*. Acta Hydrotechnica, 15/17: 79-94, Ljubljana, Slovenia.