

UDK/UDC: 004.85:556.342(282)(497.4)

Prejeto/Received: 11.10.2024

Izvirni znanstveni članek – *Original scientific paper*

Sprejeto/Accepted: 05.03.2025

DOI: [10.15292/acta.hydro.2024.09](https://doi.org/10.15292/acta.hydro.2024.09)

Objavljeno na spletu/Published online: 13.03.2025

IDRIJCA AND SOČA/ISONZO RIVER DISCHARGES ESTIMATION FOR MODELLING MERCURY POLLUTION

OCENA PRETOKOV IDRIJCE IN SOČE ZA MODELIRANJE ONESNAŽENOSTI Z ŽIVIM SREBROM

Mateja Škerjanec^{1,*}, Nataša Atanasova¹, Dušan Žagar¹, Gorazd Novak¹

¹ Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova cesta 2, 1000 Ljubljana

Abstract

River discharges play an important role in understanding mercury fate in contaminated catchments. While hydrological and hydraulic models are commonly used to calculate discharges, their complexity and computational costs often pose challenges. This study evaluates the fit of the statistical curve and one of the machine learning methods, namely model trees, and explores their performance in predicting downstream river discharges based on upstream discharge measurements. The model trees method performs better, particularly with high discharges, which transport the vast majority of mercury downstream. The resulting relationships can be used as an input to various models assessing the impact of mercury pollution from the former mine in Idrija and the climate change on mercury transport in the river systems of the Idrija and Soča/Isonzo rivers. The application of such models will improve our understanding of mercury cycling in the contaminated catchment and in the Gulf of Trieste's coastal environment.

Keywords: discharge, mercury, Idrija, Soča river, curve fitting, model trees.

Izveček

Pretoki rek igrajo pomembno vlogo pri razumevanju usode živega srebra v onesnaženih porečjih. Za določitev pretokov se običajno uporabljajo hidrološki in hidravlični modeli, žal pa njihova kompleksnost in stroški, povezani z njihovo postavitvijo, pogosto predstavljajo izziv. Ta študija primerja statistično metodo prilaganja krivulje (angl. curve fitting) in eno izmed metod strojnega učenja, tj. modelna drevesa, ter raziskuje njihovo učinkovitost pri napovedovanju dolvodnih pretokov rek na podlagi gorvodnih meritev pretokov. Modelna drevesa dajejo boljše rezultate, predvsem pri visokih pretokih, ko je tudi transport živega srebra največji. Izračunana razmerja pretokov lahko uporabimo kot vhodne podatke v različnih vrstah modelov za oceno vplivov onesnaženja z živim srebrom iz nekdanjega rudnika v Idriji ter podnebnih sprememb na transport živega srebra po rečnem sistemu Idrije in Soče, katerih uporaba bo pripomogla k boljšemu razumevanju kroženja živega srebra v onesnaženem porečju in v obalnem okolju Tržaškega zaliva.

Ključne besede: pretok, živo srebro, Idrija, reka Soča, prilaganje krivulje, modelna drevesa.

* Stik / Correspondence: mateja.skerjanec@fgg.uni-lj.si

© Škerjanec M. et al.; This is an open-access article distributed under the terms of the [Creative Commons Attribution – NonCommercial – ShareAlike 4.0 Licence](https://creativecommons.org/licenses/by-nc/4.0/).

© Škerjanec M. et al.; Vsebina tega članka se sme uporabljati v skladu s pogoji [licence Creative Commons Priznanje avtorstva – Nekomercialno – Deljenje pod enakimi pogoji 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

The Idrijca and Soča river catchments have been subject to mercury (Hg) pollution from the former Hg mine in Idrija (Slovenia). Hg concentrations are particularly increased in the alluvial sediments along both rivers (Gosar and Žibret, 2011), as the smelting remains were deposited close to the Idrijca riverbank and washed away by the runoff and river floods. Even though the mine has been closed for more than three decades, about 1,500 kg of Hg is still transported to the Gulf of Trieste by the rivers Idrijca and Soča annually (Širca et al., 1999a, 1999b). Previous studies (Žagar et al., 2006; Gosar and Žibret, 2011) confirmed that about 10,000 Mg (i.e. 10^7 kg) of Hg is still available for transport. Particularly during flood events, the Hg bound to soil and sediments is washed off in suspended matter from the flood plains and riverbed, and transported along the river system. The quantity of Hg washed to the Gulf of Trieste by discharges of lower probability (i.e. with a return period of more than 100 years) can exceed 10 Mg (i.e. 10^4 kg) per event (Žagar et al., 2006).

Zhu et al. (2018) discussed the availability of adequate data as a critical factor in selecting the appropriate model to assess Hg transport in river systems. They also listed several models used in various freshwater and marine compartments and the case studies where these models were applied. Although multi-box models representing the computational domain by consecutive zero-dimensional boxes at the level of catchment and small coastal environment are available, e.g., WASP (Ambrose and Wool, 2017), MeRiMod (Carroll et al., 2000; 2001; Žagar et al., 2006, Carroll and Warwick, 2016) and INCA-Hg models (Whitehead et al., 1998a,b; Wade et al., 2002), and have been applied in similar studies, we intend to develop and use a new multi-box model for the Idrijca–Soča river system and the Gulf of Trieste. The disadvantage of all existing multi-box models is that they require numerous parameters that are not always available, as the models were developed within the framework of a specific study, with the measurements adapted to the applied model (Zhu et al., 2018, and the references therein).

More than 90% of particulate-bound pollutants in the Idrijca and Soča rivers are transported during flood events, as the concentration of suspended solids increases exponentially with the discharge (Žagar et al., 2006). Therefore, the prediction of river discharge is crucial for understanding the Hg loading and its fate in the Gulf of Trieste. As a result of climate change, the Mediterranean region is expected to experience higher rainfall intensities and increased runoff carrying accumulated pollutants to the sea (Lun et al., 2021; Bertola et al., 2023). Despite high uncertainties, predicting Hg transport under changing climate conditions requires adequate knowledge of discharge dynamics along the river system.

There are several viable approaches to this task. Commonly, hydrological and hydraulic models are used, as they are the most transparent in terms of explaining the whole process. However, they also require several times more data and a significantly longer, more complex modelling process. Carroll et al. (2000, 2001) and Žagar et al. (2006) applied full 1-D hydraulic models, namely MeRiMod and HEC-RAS (<https://www.hec.usace.army.mil/software/hec-ras/>) to assess discharge dynamics along the river system, and the approach required detailed topography and bathymetry of the entire river systems and flood plains, consumption curves at the boundaries between the compartments, and Manning's coefficients. These parameters are variable in time and must be obtained after each extreme event.

In multi-box models, similarly to any other transport/transformation model, all processes depend on the discharge dynamics within and in between the boxes or grid cells. However, in studies where the discharge is further applied in simplified models or computations of relatively low accuracy, it is difficult to justify the application of complex and computationally expensive modelling procedures. In such cases, we can use alternative data-driven approaches, such as statistical methods or machine learning (ML) algorithms, learning from the available measured data. Although these models have good predictive power, they often lack interpretability and transferability to other domains (Maity et al., 2024; Tripathy and Mishra, 2024).

In this study, we applied two data-driven approaches, namely the curve fitting method and an ML model trees approach, to determine relations between the discharges measured along the selected river system.

ML is a branch of artificial intelligence that deals with acquiring new knowledge and finding rules in data by utilizing experience-based learning. The chosen algorithm first builds a model based on the relations between input and output variables. Next, the model is used to predict output variables (numerical or descriptive) in the event of new or changed input variable values (Jordan and Mitchell, 2015). The advantages of ML methods are the ability to independently uncover trends or patterns in large data sets, a good fit to the training data, and quick adaptation of the models to changed circumstances, whereby it is possible to continuously update them with new data if they become available (Zhong et al., 2021). In environmental sciences, ML methods have been successfully applied for assessing ecological risks and the quality of water and soil, for the optimization of various environmental technologies, the identification of surface and underground water pollution sources, and habitat modelling (Zhong et al., 2021).

Although both curve fitting and model trees are data-driven methods with all the benefits and drawbacks of such an approach, the model trees are considered semi-transparent and an upgrade of the plain regression methods (Quinlan, 1992). They allow the user to apply recursive data-partitioning techniques to automatically construct a model for predicting the values of numerical variables. Due to the data partitioning techniques, they have a better capacity to fit dynamic data. Model trees can efficiently handle large datasets with many attributes and missing data (Behnood et al., 2017). We hypothesize that model trees will outperform the curve-fitting method in discharge prediction, especially when informed by more dynamic data that can be observed during high-flow periods.

The present study aims to apply both methods and compare their results. The obtained relationships will be used in modelling the impact of climate changes to the transport of sediment and sediment-

bound Hg in the river system. This will improve our knowledge on the quantity of Hg in the Gulf of Trieste and contribute to general knowledge on Hg cycling in contaminated catchment and coastal environments. Additionally, the proposed approach to accessing the relationships between the measured discharges could be used in other environmental domains, e.g. to model river water quality or sediment transport along the watercourse.

2. Materials and methods

2.1 Study area and input data

The study area included three gauging stations along the Soča River, namely the stations Log Čezsoški, Kobarid, and Solkan, and two along its tributary Idrijca, i.e., stations Podroteja and Hotešk (Figure 1). Gauging station Kobarid was taken into consideration only during preliminary tests. The present study focused on the other four stations.

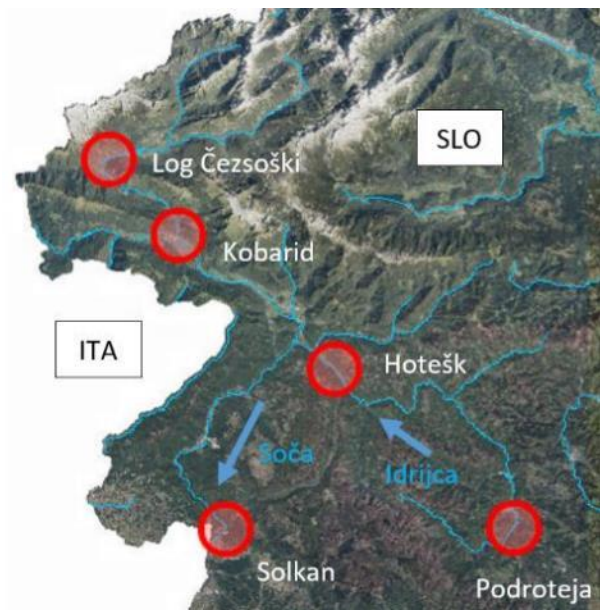


Figure 1: Locations of considered gauging stations (arrows show direction of the flow).

Slika 1: Lokacije obravnavanih merilnih postaj (puščice kažejo smer toka).

The input data were hourly river discharges measured for the period from 1 January 2020 to 31 December 2023, amounting to a time series of more than 35,000 values, acquired from the archives of Slovenian Environment Agency (ARSO). Measured

data from 2020 to 2022 were used for calibration (i.e. to set up or train both the curve fitting and ML models), while the discharges from 2023 were used for the validation.

Using hourly data, we were able to test the effect of different time lags of discharges (up to 24 hours), to account for the influence of discharges measured at the upstream stations at time $t-1$ [h] to $t-24$ [h] on the downstream discharges measured at time t .

2.2 Curve fitting

To investigate the correlation of discharges for consecutive water gauging stations, the Mathworks Matlab R2024a Curve Fitter tool was employed. This tool allows for fitting of curves and surfaces to data, conducting regression analysis using the library of linear and nonlinear models. The library

provides optimized solver parameters and initial conditions to improve the quality of fits (Mathworks Matlab, 2024).

2.3 Model trees

Model trees are hierarchical structures composed of three types of nodes (a root, internal nodes, and leaves) connected by branches (Figure 2). The root is the starting node at the top of the tree. Together with the internal nodes, it contains tests on the input attributes – in our case, discharges measured at upstream gauging stations, considering time lags. The leaves (terminal nodes) contain linear models for calculating a target/class variable – in our case, discharge measured at a particular downstream gauging station in time t . Model trees are interpreted in terms of IF–THEN rules.

Data set

Q_{1t}	Q_{1t-1}	Q_{1t-2}	...	Q_{2t}
3.1	?	?	...	80.2
3.1	3.1	?	...	50.1
3.0	3.1	3.1	...	50.9
3.1	3.0	3.1	...	45.2
3.0	3.1	3.0	...	20.8
3.0	3.0	3.1	...	19.4
...

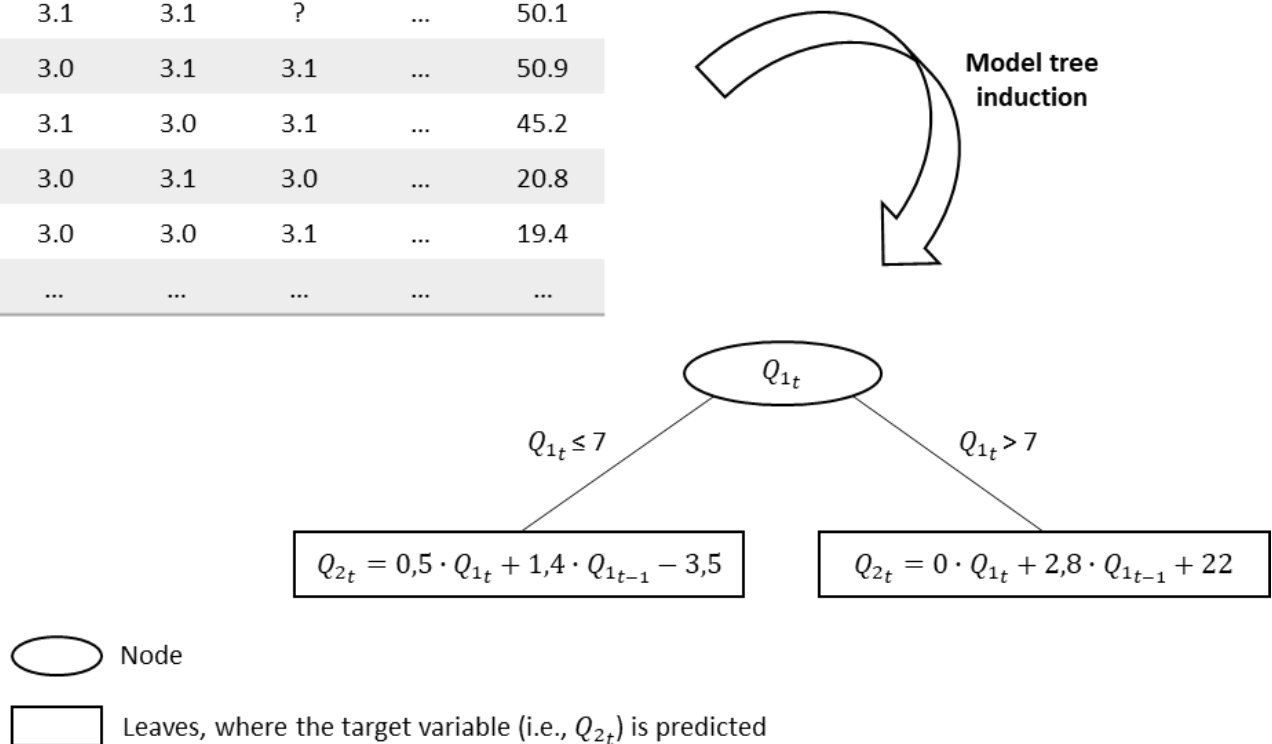


Figure 2: Induction of a model tree from a given data set, as used in the present study.

Slika 2: Indukcija modelnega drevesa iz danega niza podatkov, uporabljena v tej študiji.

In this study, model trees were built using the M5P algorithm (Quinlan, 1992; Wang and Witten, 1997) incorporated into the ML package WEKA (Witten et al., 2011). The M5P algorithm combines linear regression and decision-tree-supervised ML. It constructs a tree by recursively partitioning the training dataset based on the attribute values. The splitting process aims to find the attribute that provides the best split based on selected criteria, e.g., information gain or variance reduction.

To cope with model tree complexity and avoid overfitting, pruning was applied, which improves the transparency of the induced trees by reducing their size (Bratko, 1989). To this end, we used the “minimum number of instances per leaf” criterion, meaning every leaf should contain a minimum number of examples – otherwise, no branching is allowed.

Model trees learn from using a training data set – in our case discharges measured in the period 2020–2022. The quality of the constructed model, as gauged by the accuracy of its predictions, is expressed as a correlation coefficient. To test/validate the model, we used the generated model trees and the testing data set – discharges, measured in 2023.

One option for improving model performance is to reduce the dimensionality of the data, namely the number of attributes. To this end, we applied the automatic attribute selection feature incorporated into WEKA, namely the Correlation-based feature subset evaluation (CfsSubsetEval; Hall, 1999) used to evaluate the correlation between the attributes and the class variable. The CfsSubsetEval

automatically selects the subset of attributes that have a high correlation with the class variable while having low intercorrelation among themselves.

We used the model tree approach to find relations between:

1. discharges at the Hotešk gauging station and the delayed discharges at the upstream Podroteja gauging station, and
2. discharges at the Solkan gauging station and the delayed discharges at the gauging stations Hotešk and Log Čezsoški.

First, we used all of the attributes to build model trees. Next, we performed tree pruning. Finally, we applied the algorithm CfsSubsetEval to automatically select the most informative attributes and possibly improve model performance.

3. Results and discussion

3.1 Application of Curve Fitter tool

3.1.1 Calibration

First, curve fitting was performed for hourly discharges measured at stations along Idrijca, specifically the upstream Podroteja and downstream Hotešk. Correlating the two sets of input data resulted in a best-fit curve. Since the Idrijca takes several hours to travel from one end of the mentioned section to the other, various time lags were taken into consideration. However, for all cases the best-fit curves were linear functions $f(x) = Ax + B$, where $f(x)$ was the discharge at Hotešk and x was discharge at Podroteja, while A and B were coefficients. These results are summarized in Table 1.

Table 1: Curve fitting for discharges measured at stations Podroteja and Hotešk.

Preglednica 1: Prileganje krivulj za pretoke, izmerjene na postajah Podroteja in Hotešk.

time lag	coeff. A	coeff. B	A at 95% confidence	B at 95% confidence	R ²	RMSE (m ³ /s)
1 h	1.92	3.63	1.91 to 1.92	3.49 to 3.78	0.91	10.53
2 h	1.93	3.51	1.92 to 1.94	3.38 to 3.64	0.93	9.60
3 h	1.93	3.48	1.93 to 1.94	3.36 to 3.61	0.93	9.41
4 h	1.92	3.60	1.91 to 1.93	3.46 to 3.73	0.92	10.24
5 h	1.89	3.84	1.88 to 1.90	3.68 to 4.00	0.89	11.85

Values of R^2 and RMSE in the Table 1 indicate that the best correlation was achieved with the time lag of 3 hours. The graphical representation of this correlation is given in Figure 3.

Figure 3 clearly shows that correlation decreases with increasing discharges. Since we are interested mostly in higher discharges (as they trigger the movement of sediment and consequently of Hg), this statistical approach is not optimal.

Next, based on the results showing that the time lag between the Podroteja and Hotešk stations is 3 hours, and assuming the similar distances of the river reaches in consideration, it was estimated that the same time lag exists between stations on the Soča, i.e. Log Čezsoški (upstream location) and Kobarid (downstream location, but still upstream of the confluence). Using the same line of reasoning, it can be estimated that the time lag between the confluence and the station at Solkan is similarly 3 hours, amounting to 6 hours of difference between the discharges at the upstream stations of Podroteja

and Log Čezsoški, and the target downstream station of Solkan.

Again, the Curve Fitter tool was used, but this time to find the best-fit surface for the stations Podroteja, Log Čezsoški and Solkan, separated by a time difference of 6 hours. A time lag of 7 hours was also tested for the comparison. In both cases, the best fit was a surface $f(x,y) = C + Dx + Ey$, where $f(x,y)$ was the discharge at Solkan, x was the discharge at Podroteja, y was the discharge at Log Čezsoški, while C , D , and E were coefficients. These results are summarized in Table 2.

Values of R^2 and RMSE in the Table 2 indicate that better correlation is achieved with the time lag of 6 hours. As expected, the correlation for the surface (considering 3 stations) is not as good as for the curve (considering only 2 stations). Also, it decreases with the increasing time lag. The graphical representation of this surface is given in Figure 4.

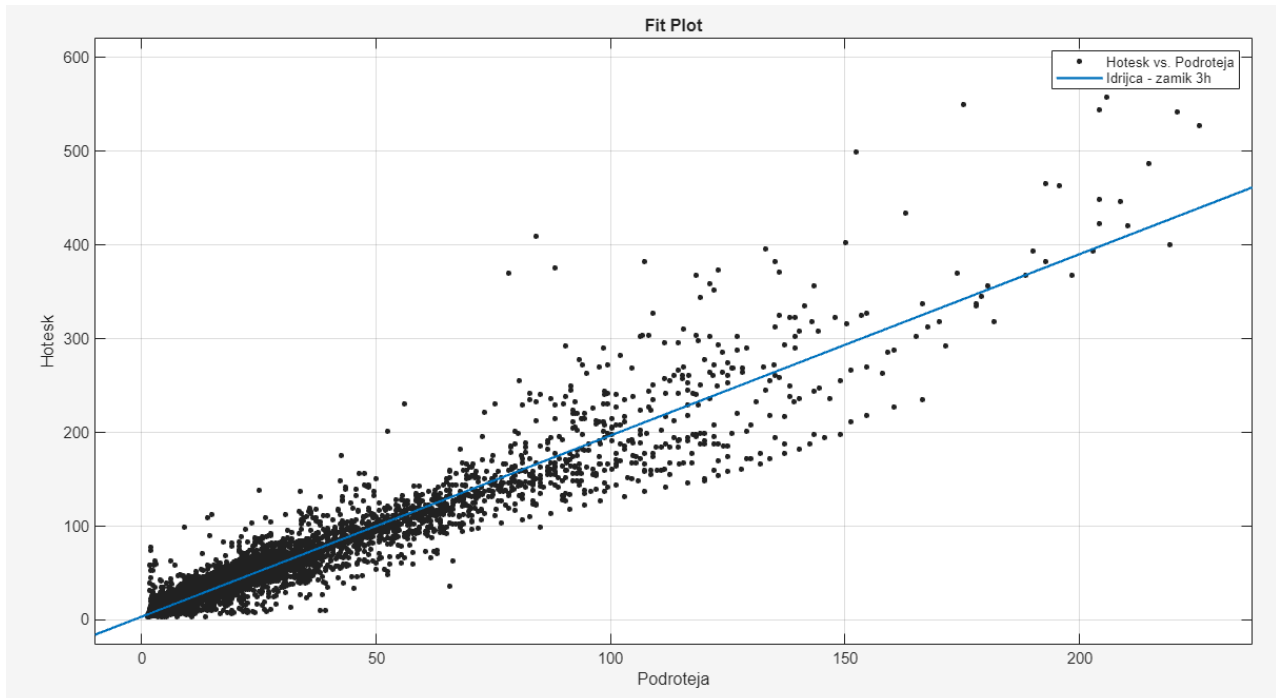


Figure 3: Best-fit curve for discharges at the gauging stations Podroteja (x-axis, in m^3/s) and Hotešk (y-axis, in m^3/s), with the time lag equaling 3 hours.

Slika 3: Najboljša krivulja za pretoke na postajah Podroteja (x-os, v m^3/s) in Hotešk (y-os, v m^3/s), časovni zamik je 3 ure.

Table 2: Best-fit surface for discharges at stations Podroteja, Log Čezsoški, and Solkan.

Preglednica 2: Najbolje prilegajoča se ploskev za pretoke na postajah Podroteja, Log Čezsoški in Solkan.

time lag	C	D	E	C at 95% confidence	D at 95% confidence	E at 95% confidence	R ²	RMSE (m ³ /s)
6 h	3.89	3.64	2.55	3.11 to 4.67	3.60 to 3.68	2.52 to 2.58	0.83	46.32
7 h	4.92	3.61	2.52	4.10 to 5.74	3.57 to 3.65	2.48 to 2.65	0.81	48.77

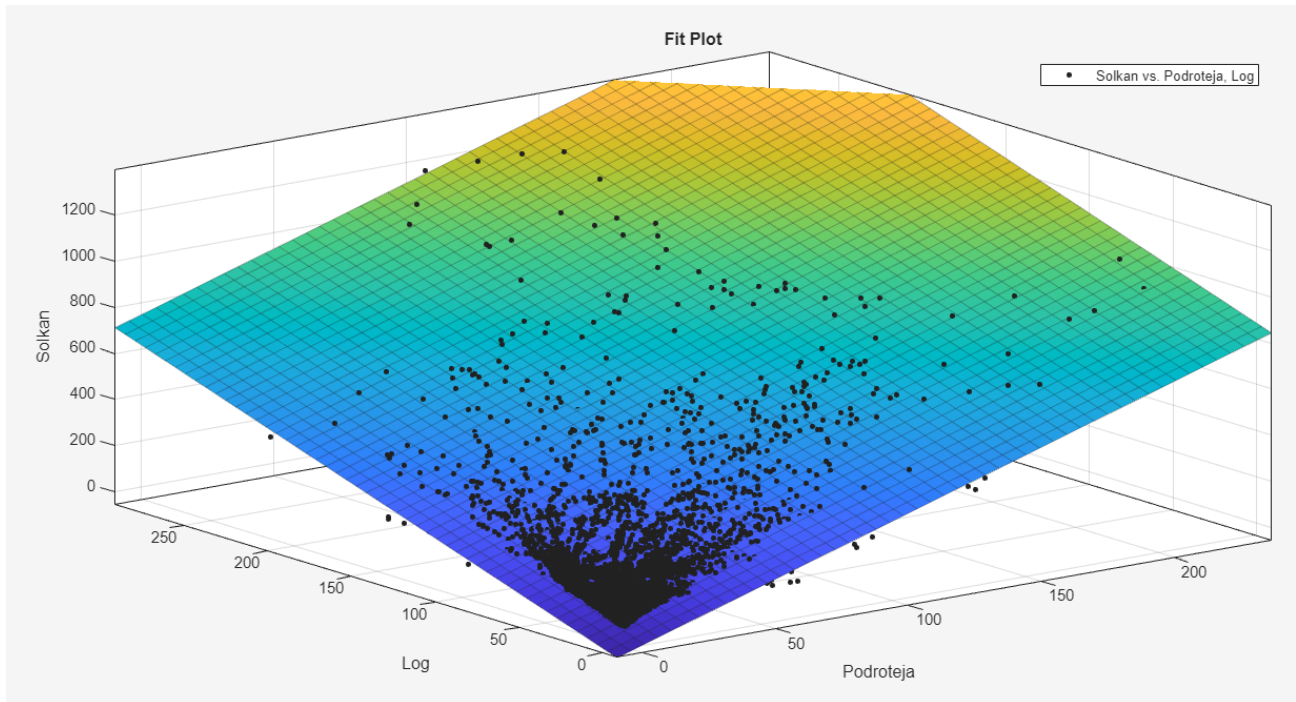


Figure 4: Best-fit surface for discharges measured at Idrija – Podroteja (x -axis, in m³/s), Soča – Log Čezsoški (y -axis, in m³/s), and Soča – Solkan (z -axis, in m³/s); time lag equals 6 hours.

Slika 4: Najbolje prilegajoča se ploskev za pretoke, izmerjene na postajah Idrija – Podroteja (x -os, v m³/s), Soča – Log Čezsoški (y -os, v m³/s) in Soča – Solkan (z -os, v m³/s); časovni zamik je 6 ur.

We get the time series shown in Figure 5 by calculating the discharges at Solkan from the discharges at Podroteja and Log Čezsoški, assuming the 6 h time lag, using the equation listed in Table 2, which can be written out as:

$$Q_{Solkan} = 3.89 + 3.64 \cdot Q_{Podroteja(t-6h)} + 2.55 \cdot Q_{Log(t-6h)} \quad (1)$$

The results show that calculated discharges (particularly peak discharges) are lower than the measured ones.

3.1.2 Validation

Using the same approach as for the calibration (equation (1)) but for a different set of hourly discharges, namely those of 2023, the validation gave the results presented in Figure 6.

Figure 6 shows that the calculated discharges are lower than the measured ones, while in terms of time, the predictions made with Curve Fitter occur close to measured peaks. However, a more detailed analysis showed significant discrepancies.

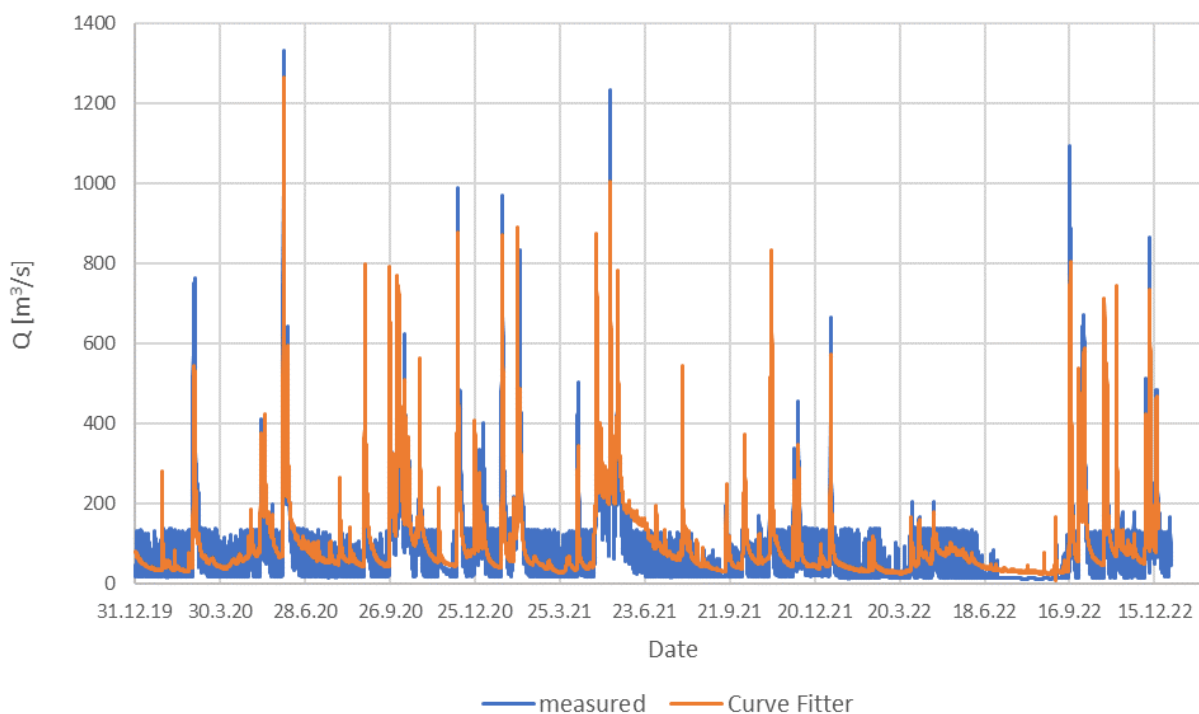


Figure 5: Application of the Curve Fitter tool – results of the calibration for the Solkan gauging station.

Slika 5: Uporaba orodja za prileganje krivulj – rezultati umerjanja za vodomerno postajo Solkan.

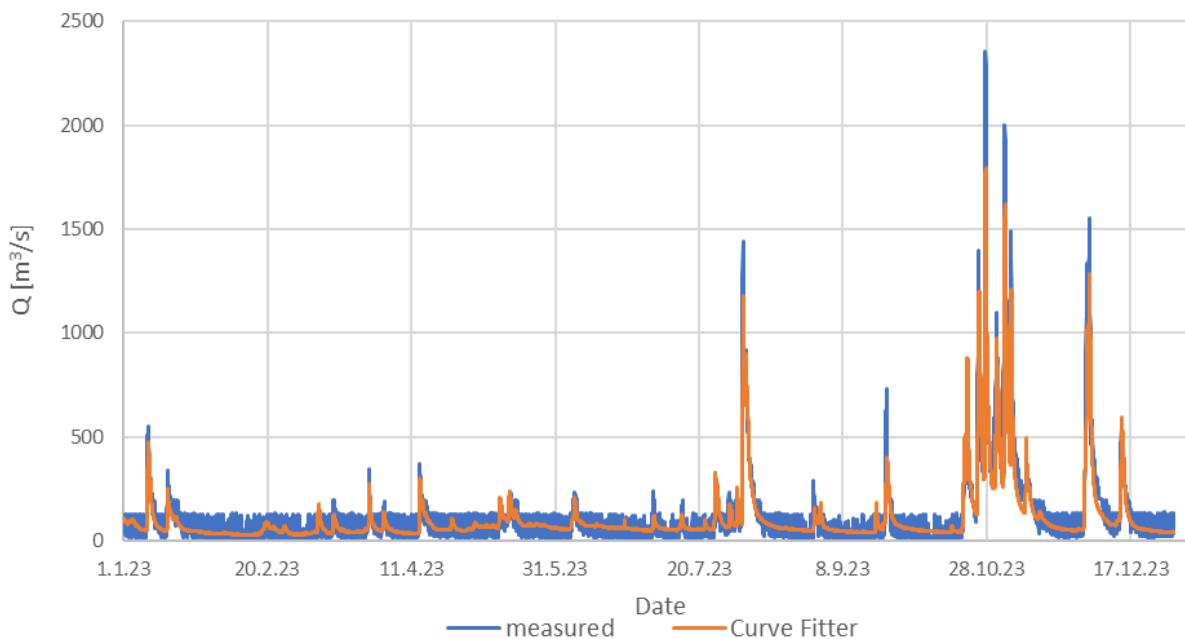


Figure 6: Application of the Curve Fitter tool – results of the validation for the Solkan gauging station.

Slika 6: Uporaba orodja za prileganje krivulj – rezultati validacije za vodomerno postajo Solkan.

Looking at high discharge events at Hotešk ($Q_{\text{Hotešk}} > 200 \text{ m}^3/\text{s}$), which are the most relevant

for the transport of Hg, gave results presented in Table 3.

Based on Table 3, the following can be stated:

1. Measured time lags between Hotešk and Solkan range from -1 h to +5 h, which is on average less than what was assumed during the calibration stage.
2. Predicted (i.e. calculated) times of peak discharges at Solkan are on average more than 3 hours late, when compared against measured hourly discharges.

3. Predicted (i.e. calculated) values of peak discharges at Solkan are in most cases underestimated; the difference between calculated and measured discharges varies from -24% to +8%, giving an average difference of 15% underestimation.

This indicates a need for a better prediction tool, for example the one described in the following sections.

3.2 Application of model trees

3.2.1 Calibration

When forecasting the discharges at the Hotešk station, we tested three variants, using: 1) all the attributes, 2) tree pruning, and 3) the reduced number of attributes according to the CfsSubsetEval analysis. All variants gave similar results – only one leaf containing a single equation (equation (2)), with a correlation coefficient of 0.97.

$$\begin{aligned} \text{Idrijca_Hotesk} = & \quad (2) \\ & 0.48 \cdot \text{Idrijca_Podroteja} \\ & + 1.04 \cdot \text{Idrijca_Podroteja-3} \\ & + 0.43 \cdot \text{Idrijca_Podroteja-4} \\ & + 3.31 \end{aligned}$$

On the other hand, when predicting discharges at the Solkan station (based on the discharges measured at stations Log Čezsoški and Podroteja), three variants are presented. In the first one, all attributes (50) were used. Again, only one leaf was created using a single equation (equation (3)). The correlation coefficient was 0.92.

In the second one, the reduced number of attributes (5) were used. These attributes were selected based on the results of the CfsSubsetEval analysis. The resulting model tree is presented in Figure 7. The two linear models contained in the leaves are given in equations (4) and (5). Here, the correlation coefficient was 0.93.

$$\begin{aligned} \text{Soca_Solkan} = & \quad (3) \\ & - 0.54 * \text{Soca_Log} \\ & + 1.08 * \text{Soca_Log-1} \\ & + 0.82 * \text{Soca_Log-3} \\ & + 1.47 * \text{Soca_Log-5} \\ & - 0.13 * \text{Soca_Log-7} \\ & - 0.18 * \text{Soca_Log-23} \end{aligned}$$

$$\begin{aligned} & + 1.21 * \text{Idrijca_Podroteja} \\ & + 0.45 * \text{Idrijca_Podroteja-3} \\ & - 0.15 * \text{Idrijca_Podroteja-4} \\ & + 0.78 * \text{Idrijca_Podroteja-5} \\ & + 1.15 * \text{Idrijca_Podroteja-6} \\ & - 0.12 * \text{Idrijca_Podroteja-20} \\ & + 0.66 * \text{Idrijca_Podroteja-24} \\ & + 1.73 \end{aligned}$$

$$\text{LM 1:} \quad (4)$$

$$\begin{aligned} \text{Soca_Solkan} = & \\ & 0.51 * \text{Soca_Log-2} \\ & + 1.38 * \text{Soca_Log-4} \\ & + 0.00 * \text{Idrijca_Podroteja} \\ & + 0.00 * \text{Idrijca_Podroteja-4} \\ & + 8.68 * \text{Idrijca_Podroteja-6} \\ & - 3.55 \end{aligned}$$

$$\text{LM 2:} \quad (5)$$

$$\begin{aligned} \text{Soca_Solkan} = & \\ & 0.00 * \text{Soca_Log-2} \\ & + 2.78 * \text{Soca_Log-4} \\ & + 0.00 * \text{Idrijca_Podroteja} \\ & + 3.18 * \text{Idrijca_Podroteja-4} \\ & + 0.01 * \text{Idrijca_Podroteja-6} \\ & + 22.31 \end{aligned}$$

In the third variant, the same reduced number of attributes (5) was used, together with additional tree pruning. This resulted in only one leaf containing a single equation (equation (6)), with a correlation coefficient of 0.92.

$$\text{Soca_Solkan} = \quad (6)$$

$$\begin{aligned} & 0.47 * \text{Soca_Log-2} \\ & + 2.11 * \text{Soca_Log-4} \\ & + 1.26 * \text{Idrijca_Podroteja} \\ & + 0.44 * \text{Idrijca_Podroteja-4} \\ & + 2.06 * \text{Idrijca_Podroteja-6} \\ & + 2.18 \end{aligned}$$

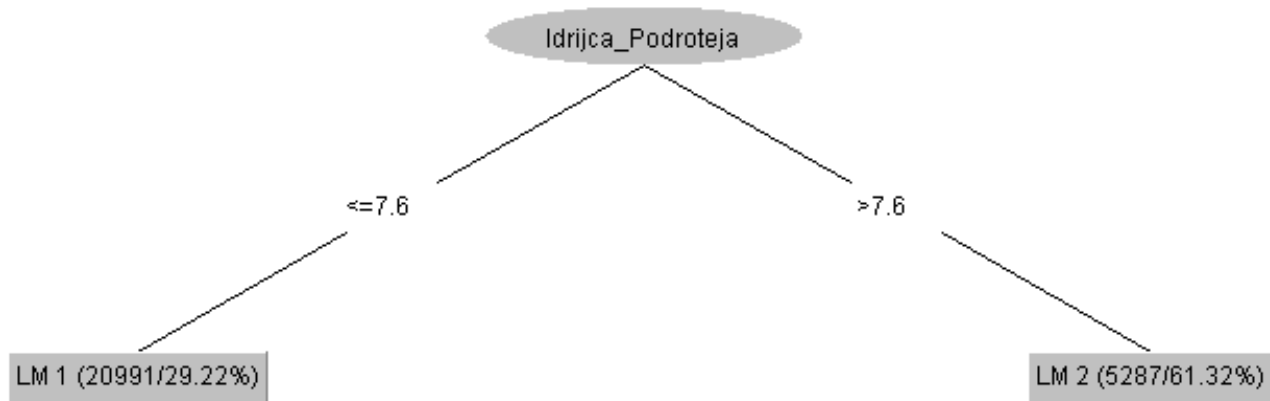


Figure 7: Model tree for predicting discharges at the Solkan gauging station (considering the reduced number of attributes).

Slika 7: Modelno drevo za napovedovanje pretokov na vodomerni postaji Solkan (upoštevajoč manjše število atributov).

Table 3: Validation results of curve fitter method for cases with high discharges of Hotešk.

Preglednica 3: Rezultati validacije metode prilaganja krivulj za primere z visokimi pretoki Hoteška.

date	peaks at Hotešk > 200 m ³ /s		corresponding peaks at Solkan				difference ΔQ [%]
	hour	Q [m ³ /s]	hour	Q [m ³ /s]	hour	Q [m ³ /s]	
04.08.2023	08:00	588.3	07:00	1439.0	08:00	1181.0	-18
27.10.2023	11:00	686.6	14:00	2352.7	15:00	1792.6	-24
05.11.2023	09:00	504.3	11:00	1488.5	13:00	1208.4	-19
01.12.2023	14:00	432.1	19:00	1333.3	2.12. 03:00	1043.4	-22
13.12.2023	18:00	288.0	18:00	552.4	22:00	596.3	+8

Next, we had to decide which linear models to choose for modelling the influence of climate change on the transport of sediment and sediment-bound Hg. For modelling the discharges at the Hotešk gauging station, we could only induce one linear model (equation (2)). From equation (2), it can be seen that the lag times of discharges measured at the upstream Podroteja station up to four hours are relevant for forecasting discharges at the Hotešk station. The lag time of three hours seems to have the greatest impact (due to the highest coefficient value being multiplied by the discharge measured at Podroteja station at time t-3). Comparison of the measured hourly discharges and the results of the selected linear model (namely equation (2)) are presented in Figure 8.

To model the discharges at the Solkan gauging station, we induced three models with almost

identical predictive performance, i.e. in all three cases the correlation coefficient was approx. 0.92. Thus, from a practical perspective and to avoid conditional (IF-THEN) relations, we decided to use equation (6). Here, the lag time of four hours considering Log Čezsoški and six hours considering Podroteja seem to impact discharges at the Solkan station. If we focus on the relationship between the Podroteja and Solkan stations, the lag time of six hours seems to have the greatest impact (due to the highest coefficient value being multiplied by the discharge measured at Podroteja station at time t-6). Comparison of the measured hourly discharges and the results of the selected linear model (equation (6)) are presented in Figure 9. The results show that the calculated peak discharges are mostly underestimated.

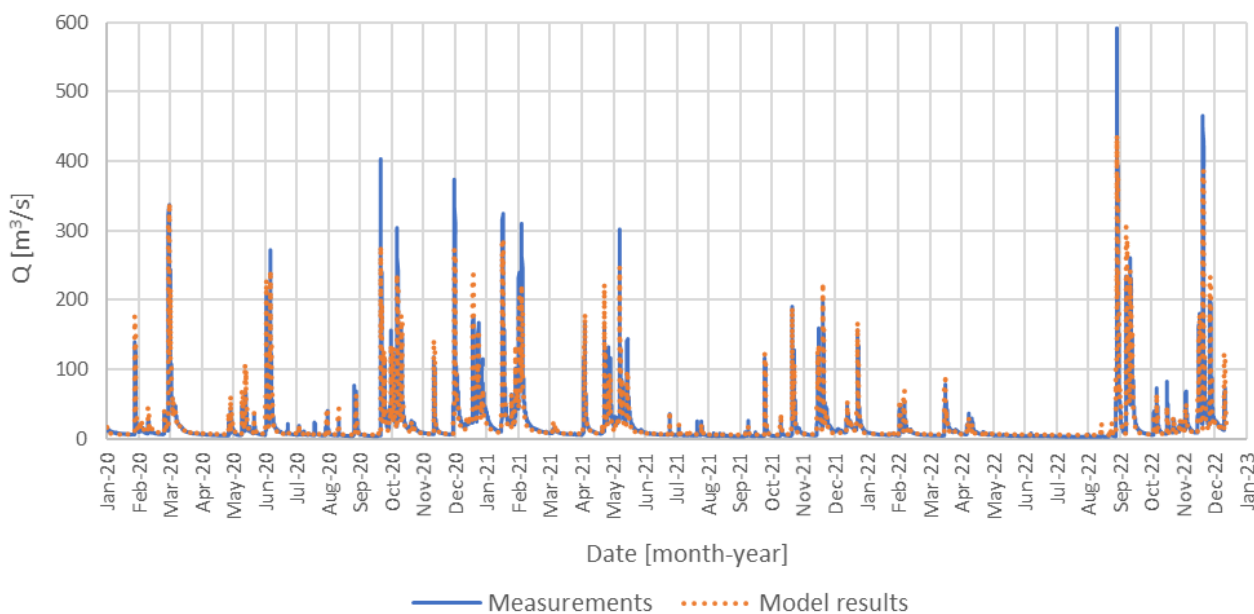


Figure 8: Comparison of the measured hourly discharges and the results of the selected linear model (equation (2)) for the Hotešk gauging station – for the calibration period 2020–2022.

Slika 8: Primerjava izmerjenih urnih pretokov in rezultatov izbranega linearnega modela (enačba (2)) za vodomerno postajo Hotešk – za obdobje umerjanja 2020–2022.

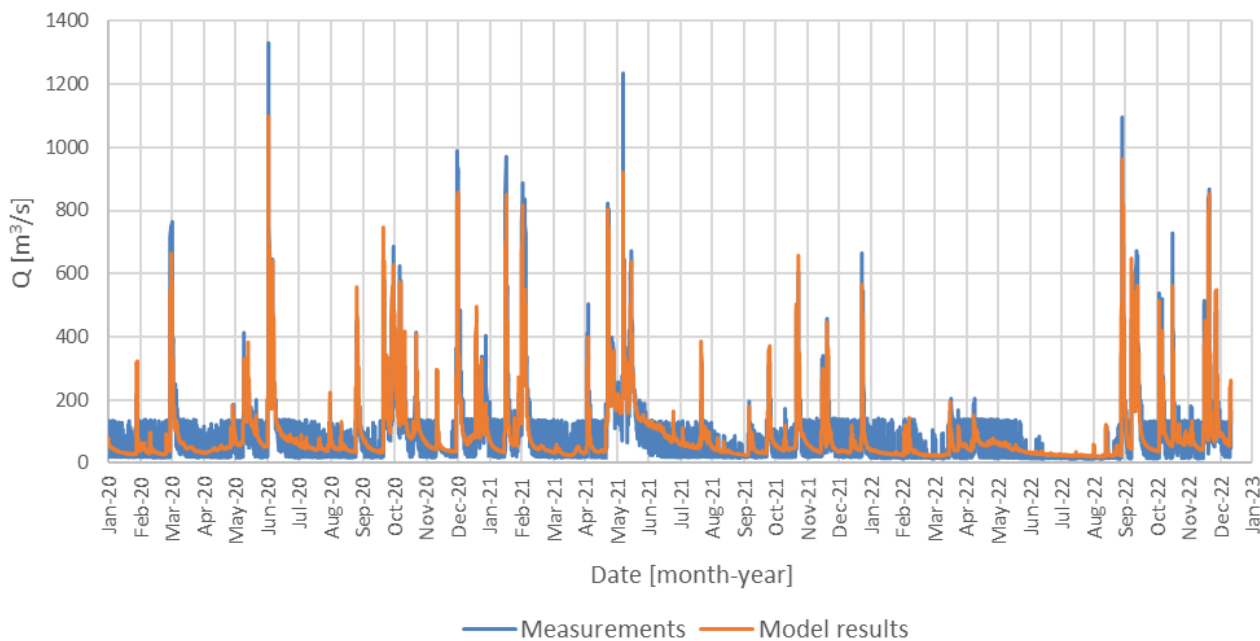


Figure 9: Comparison of the measured hourly discharges and the results of the selected linear model (equation (6)) for the Solkan gauging station – for the calibration period 2020–2022.

Slika 9: Primerjava izmerjenih urnih pretokov in rezultatov izbranega linearnega modela (enačba (6)) za vodomerno postajo Solkan – za umeritveno obdobje 2020–2022.

3.2.2 Validation

To validate equation (2) and assess its usability for predicting discharges at the Hotešk station, we used a different (test) data set, namely discharges, measured at the Podroteja station in 2023. Resulting correlation coefficient was almost the same as within the calibration phase (0.95). The validation results are presented in Figure 10.

The same procedure was repeated for the validation of equation (6) for predicting discharges at the Solkan station. Again, we used a test data set, namely discharges, measured at the Log Čezsoški and Podroteja station in 2023. Resulting correlation coefficient was slightly higher than in the calibration phase (0.96). The validation results for the Solkan station are presented in Figure 11.

Figure 10 and 11 show that calculated peak discharges are again lower than the measured ones, while in terms of time, the predictions of the model tree approach are in accordance with the measured

peaks. Thus, validation confirmed the trend we noticed during the calibration.

Again, we performed further analysis for the top five high-discharge events at the Hotešk station within the validation period ($Q_{\text{Hotešk}} > 200 \text{ m}^3/\text{s}$), as these are the most relevant for the transport of Hg. More details are presented in Table 4.

Based on Table 4, the following can be stated:

1. The predicted values of peak discharges at Solkan are in most cases underestimated; the difference between the calculated and measured discharges varies from -18% to +4%, giving an average difference of -9%, i.e. 9% underestimation.
2. On average, the predicted times of peak discharges at Solkan are about an hour late; however, they vary up to four hours.
3. The measured time lags between Hotešk and Solkan vary between zero and five hours.

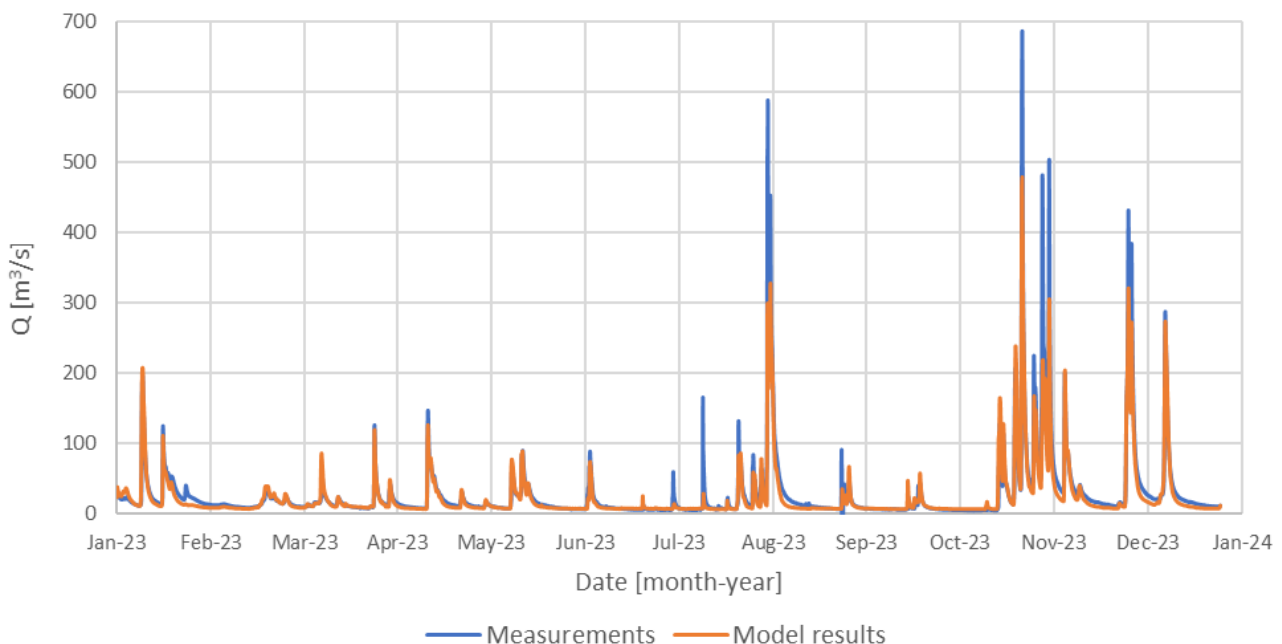


Figure 10: Comparison of the measured hourly discharges and the results of the selected linear model (equation (2)) for the Hotešk gauging station – for the validation year 2023.

Slika 10: Primerjava izmerjenih urnih pretokov in rezultatov izbranega linearnega modela (enačba (2)) za vodomerno postajo Hotešk – za validacijsko leto 2023.

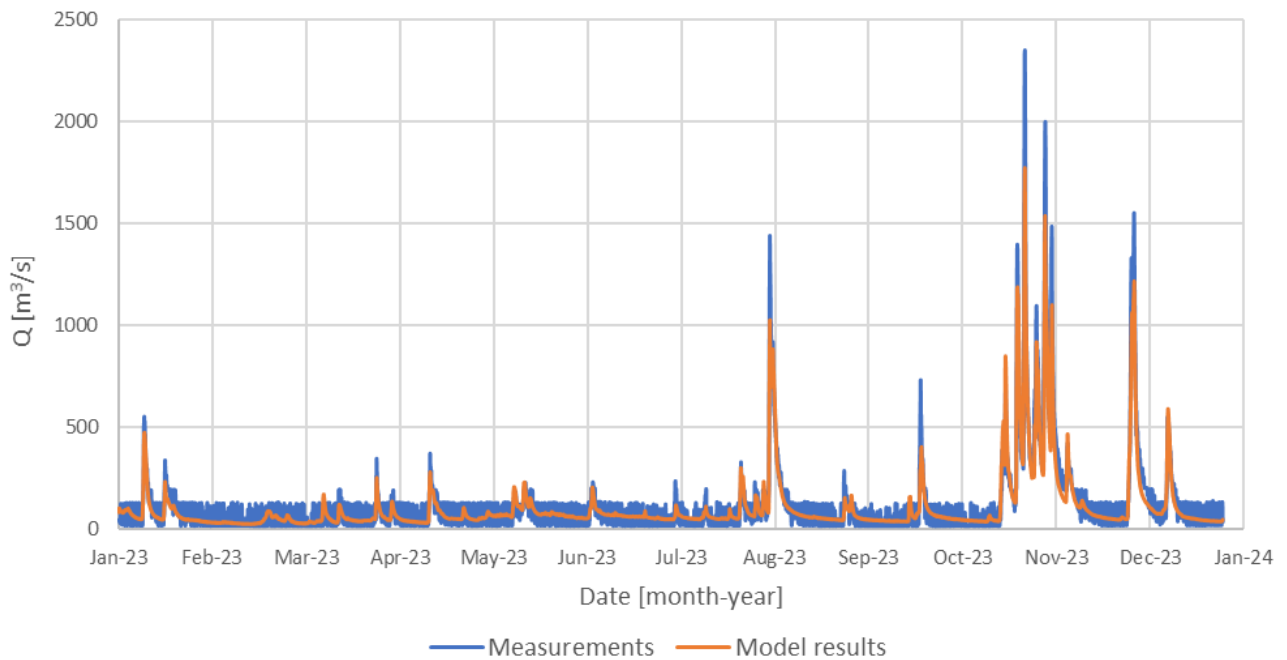


Figure 11: Comparison of the measured hourly discharges and the results of the selected linear model (enačba (6)) for the Solkan gauging station – for the validation year 2023.

Slika 11: Primerjava izmerjenih urnih pretokov in rezultatov izbranega linearnega modela (enačba (6)) za vodomerno postajo Solkan – za validacijsko leto 2023.

Table 4: Validation results of model trees for cases with high discharges of Hotešk.

Preglednica 4: Rezultati validacije modelnih dreves za primere z visokimi pretoki Hoteška.

date	peaks of Hotešk > 200 m ³ /s		corresponding peaks at Solkan				difference ΔQ [%]
	hour	Q [m ³ /s]	measured hour	measured Q [m ³ /s]	calculated hour	calculated Q [m ³ /s]	
4.8.2023	8:00	588.3	7:00	1439.0	9:00	1434.0	0
27.10.2023	11:00	686.6	14:00	2352.7	13:00	1990.0	-15
5.11.2023	9:00	504.3	11:00	1488.5	11:00	1280.6	-14
1.12.2023	14:00	432.1	19:00	1333.3	23:00	1095.2	-18
13.12.2023	18:00	288.0	18:00	552.4	20:00	572.0	+4

3.3 Comparison of both methods

Finally, the results of both methods were compared to the discharges measured in 2023 (i.e. within the validation period) at the Solkan gauging station. Figure 12 shows the comparison for the entire validation period, while in Figures 13–17 we present the results for the particular high-flow events that were also considered in Tables 3 and 4.

Figures 13–17, as well as Tables 3 and 4, show that in most cases the prediction of both peak discharge and the time lag is better with the model tree approach. The underestimation of both methods is

more pronounced at single-peak flood waves, while with two-peak flood waves discrepancies occur in both the discharge and in the time lag.

The correctly calculated time lag is as important for predicting the discharge dynamics as for the correctly calculated peak discharges. In the equations (2) – (6) obtained from the model trees, we used discharges with different lags at the upstream gauges and determined the lag with the maximum effect on the downstream gauges. By applying the same equations calibrated to existing data and introducing different input discharges, we can also predict the temporal dynamics of flood

waves. Considering the distances between the compartments of the multi-box model, the time lag

between the occurrence of the flood wave in each box can be calculated.

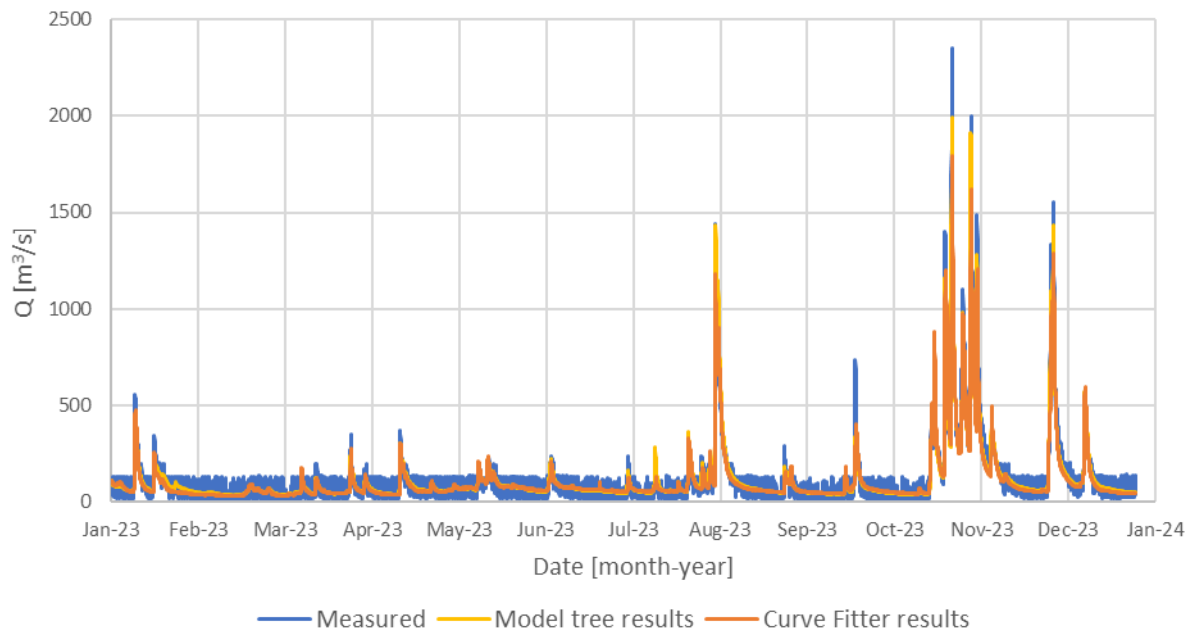


Figure 12: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods for the entire validation period.

Slika 12: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod za celotno validacijsko obdobje.

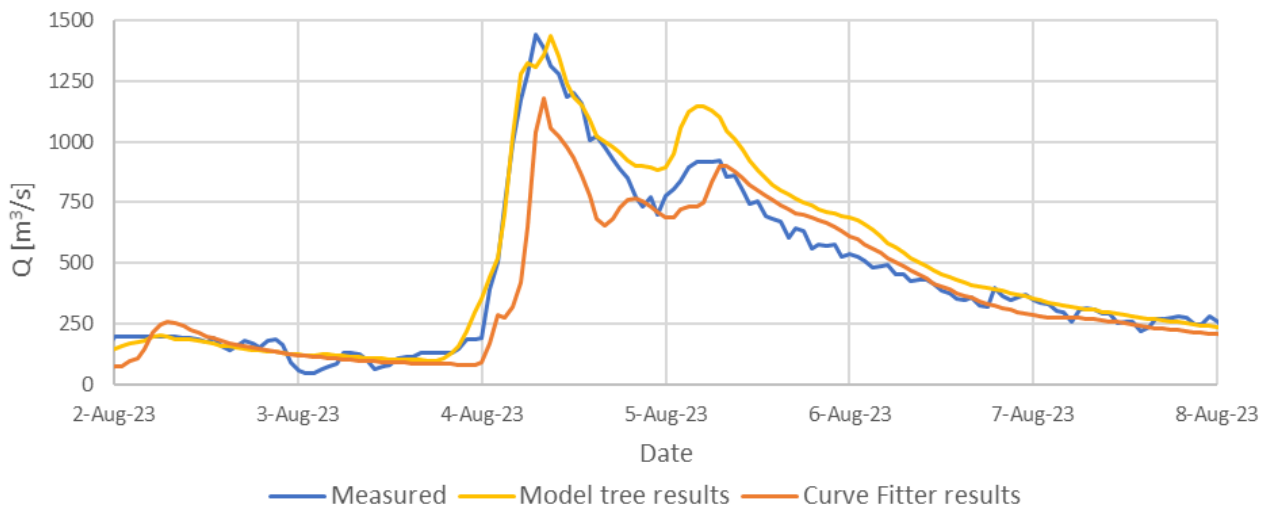


Figure 13: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods – for 4 August 2023.

Slika 13: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod – za 4. avgust 2023.

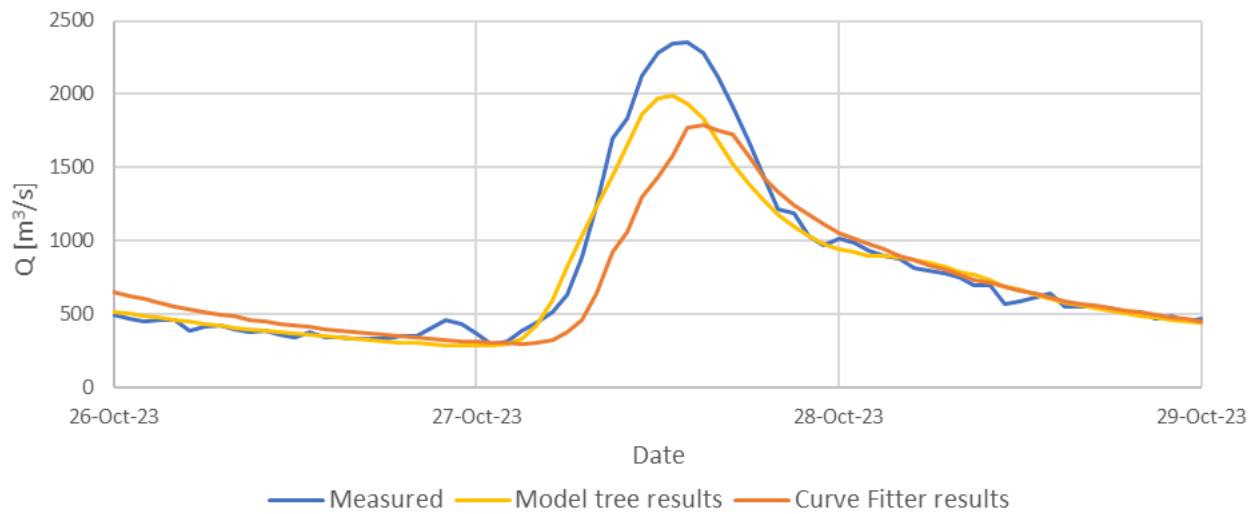


Figure 14: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods – for the 27 October 2023 event.

Slika 14: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod – za dogodek 27. oktobra 2023.

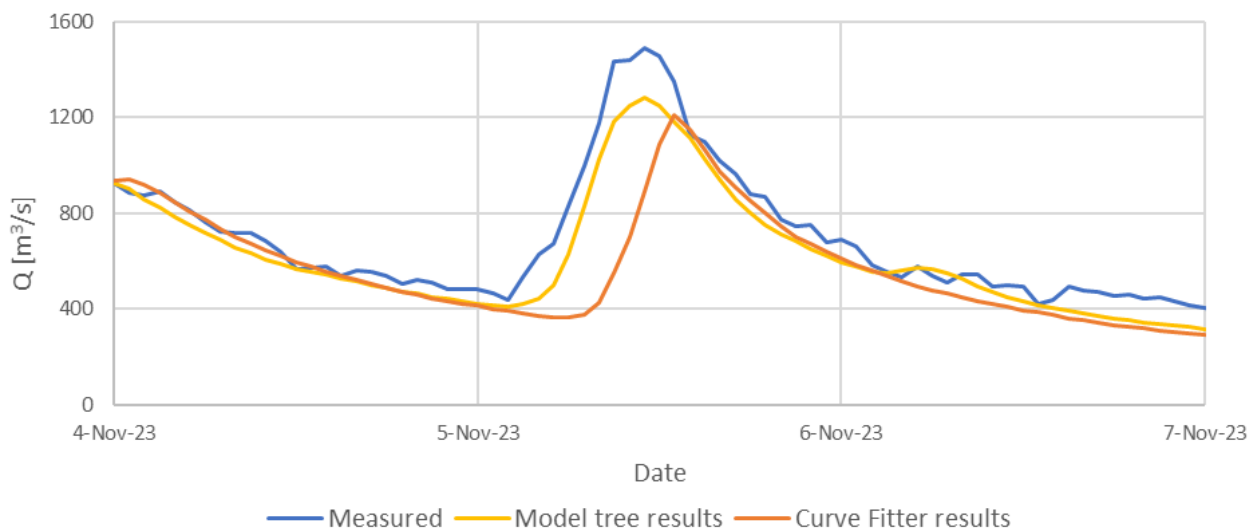


Figure 15: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods – for the 5 November 2023 event.

Slika 15: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod – za dogodek 5. novembra 2023.

We are aware that the flow parameters at successive gauging stations depend on a number of variables that were not considered by the simplifications used. Each meteorological event is unique in terms of its evolution, the spatial and temporal distribution of precipitation, its overall intensity, and the direction and speed at which the precipitation system moves.

In addition, runoff depends on several other parameters (local rainfall intensity, season, snow cover, soil moisture, etc.) which are also unique to each high runoff event. It would be difficult to expect any of the methods used to produce a reliable equation that would cover all the possible cases. However, the predictive capability of both methods

can be considered satisfactory for simulating different discharge scenarios in multi-box models, as the calculation of particulate Hg transport involves further modelling simplifications. The relationship between water discharge and sediment transport is not straightforward and sediment-bound Hg concentrations can vary by a factor of three or

more depending on the origin of the sediment. Underestimated high discharges at successive gauges by 10–20% are therefore acceptable and can be adjusted manually in the multi-box modelling process. The same level of accuracy is expected for the time lag prediction between the gauges and between the compartments of the multi-box model.

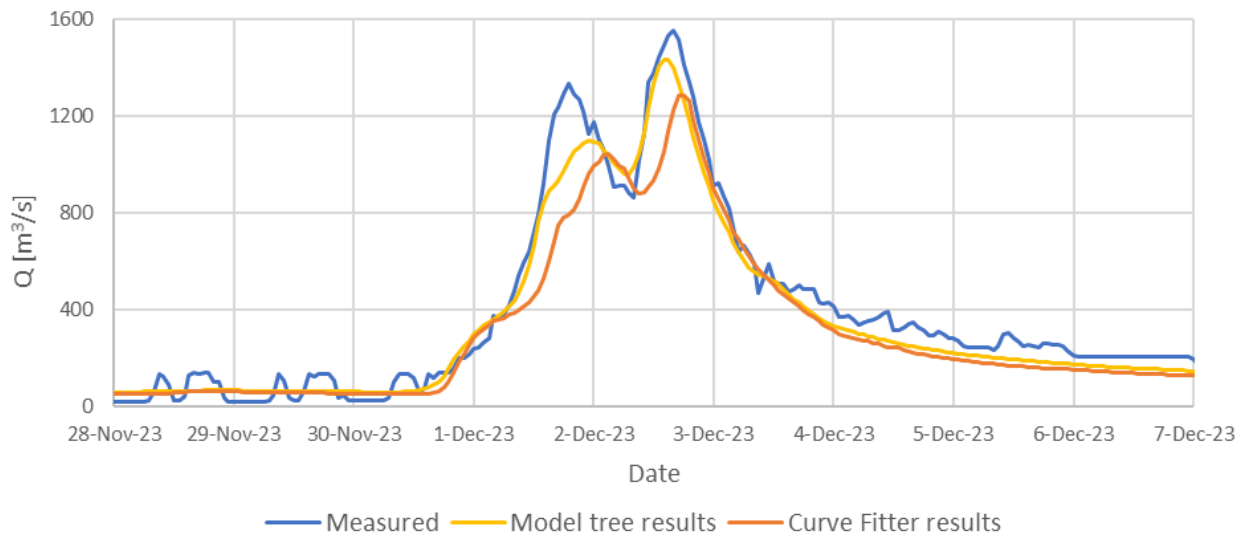


Figure 16: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods – for the 1 December 2023 event.

Slika 16: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod – za dogodek 1. decembra 2023.

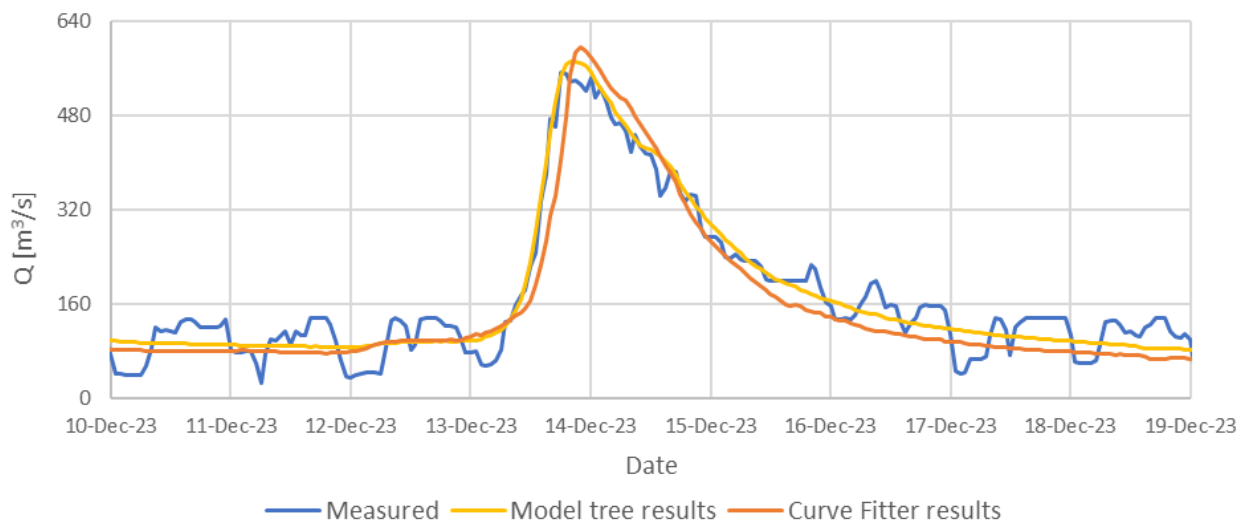


Figure 17: Comparison of the measured hourly discharges at the Solkan gauging station and the results of both methods – for the 13 December 2023 event.

Slika 17: Primerjava izmerjenih urnih pretokov na vodomerni postaji Solkan in rezultatov obeh metod – za dogodek 13. decembra 2023.

Further uncertainties in our calculations and estimates are introduced by including climate changes into the model, namely in terms of the expected increases in precipitation, runoff, soil erosion, river discharge, and consequently increased pollutant transport. Considering these uncertainties, the tools described are sufficient for estimating the temporal and spatial dynamics of discharges along the river system (and in each box of the Hg model) cost-effectively, based on either estimated or calculated hydrographs at the selected (upstream) gauging station.

4. Conclusions

Starting from the same input data, i.e. hourly discharges measured recently at several gauging stations located along the Idrijca and Soča River, we used Matlab's curve fitting tool Curve Fitter and the model tree approach to predict relationship between discharges at successive gauging stations.

Based on the calibration and validation phases (which took into consideration data for three years and one year, respectively) we can conclude that both methods underestimate the peak discharges of the flood waves. The model tree results fit the measured hydrographs better, especially when focusing on specific high-flow events, which confirmed our hypothesis. Looking at the Curve Fitter results, the peaks appear to be more offset than the model tree results. Based on the comparisons presented, we can conclude that, of the approaches tested, the model trees perform better in terms of predicting discharges and the time lag between gauges and will therefore be used in our further work. The improved knowledge about the relationships between discharges along the Idrijca and Soča river systems will allow simulations of Hg transport with multi-box models without using computationally expensive hydrological and/or hydraulic models. The uncertainties in the further application of discharge dynamics due to the complex relationship between discharge and suspended matter transport and in the introduction of climate change effects make the use of more complex hydrological and hydraulic models at least questionable, if not unjustified. We consider the reliability of the established relationships to be

sufficient for simulating sediment and sediment-bound pollutant transport, thus reducing the cost and time of further investigations.

Using the generated data-driven models, we can, with limited reliability, simulate discharges at two gauging stations, namely the Hotešk and Solkan. Based on the identified relationships between the discharges measured along the Idrijca and Soča Rivers, we can determine the time lags (i.e. the flood wave delays) between consecutive compartments along the analyzed system. This will help us set up a Hg multi-box model, which is expected to be a useful tool for better understanding the transport and transformation processes of particulate pollutants and, in particular, Hg dynamics in the river system, inter alia considering the impact of climate change. Nevertheless, modelling the effects of climate change is expected to introduce additional uncertainties likely to exceed the discrepancy between the simulated and measured results presented.

Acknowledgements

The authors would like to thank Mira Kobold from the Slovenian Environmental Agency for providing the data on measured discharges. This research was supported by the Slovenian Research Agency (project No. J1-3033 and research core funding No. P2-0180).

References

- Ambrose, R. B., Wool, T. A. (2017). WASP8 Stream Transport – Model Theory and User's Guide. US EPA, Washington DC, 67 pp. <https://www.epa.gov/sites/default/files/2018-05/documents/stream-transport-user-guide.pdf>.
- Behnood, A., Behnood, V., Gharehveran, M.M., Alyamac, K.E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm, *Constr. Build. Mater.* **142**, 199–207. <https://doi.org/10.1016/j.conbuildmat.2017.03.061>.
- Bertola M, Blöschl G, Bohac M. et al. (2023). Megafloods in Europe can be anticipated from observations in hydrologically similar catchments. *Nat. Geosci.* **16**, 982–988. <https://doi.org/10.1038/s41561-023-01300-5>.

- Bratko, I. (1989). “Machine learning” in Gilhooly, K.J., Ed., *Human and Machine Problem Solving*. Plenum Press, New York and London, 265-287. <https://doi.org/10.1007/978-1-4684-8015-3>.
- Carroll, R.W.H., Warwick, J.J., Heim, K. J., Bonzongo, J. C., Miller, J. R., Lyons W. B. (2000). Simulation of mercury transport and fate in the Carson River, Nevada. *Ecol Model* **125** (2-3), pp. 255-278. [https://doi.org/10.1016/S0304-3800\(99\)00186-6](https://doi.org/10.1016/S0304-3800(99)00186-6).
- Carroll, R.W.H., Warwick, J.J. (2001). Uncertainty analysis of the Carson River mercury transport model. *Ecol Model* **137**, pp. 211-224. [https://doi.org/10.1016/S0304-3800\(00\)00438-5](https://doi.org/10.1016/S0304-3800(00)00438-5).
- Carroll, R.W.H., Warwick, J.J. (2016). Modeling the Highly Dynamic Loading of Mercury Species in the Carson River and Lahontan Reservoir System, Nevada. *JAWRA*, pp. 1207-1222. <https://doi.org/10.1111/1752-1688.12448>.
- Gosar, M., Žibret, G. (2011). Mercury contents in the vertical profiles through alluvial sediments as a reflection of mining in Idrija (Slovenia), *Journal of Geochemical Exploration* **110** (2), 81-91. <https://doi.org/10.1016/j.gexplo.2011.03.008>.
- Hall, M.A. (1999). Correlation Based Feature Subset Selection for Machine Learning. PhD Thesis, University of Waikato, Hamilton, New Zealand, 198 pp.
- HEC-RAS model: <https://www.hec.usace.army.mil/software/hec-ras/>.
- Jordan, M.I., Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects, *Science* **349**, 255-260. <https://doi.org/10.1126/science.aaa8415>.
- Lun D, Viglione A, Bertola M, Komma J, Parajka J, Valent P, Blöschl G (2021). Characteristics and process controls of statistical flood moments in Europe – a data-based analysis, *Hydrol. Earth Syst. Sci.* **25**, 5535–5560. <https://doi.org/10.5194/hess-25-5535-2021>.
- Maity, R., Srivastava, A., Sarkar, S., Khan, M.I. (2024). Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning. *Applied Computing and Geosciences* **24**, 100206. <https://doi.org/10.1016/j.acags.2024.100206>.
- Mathworks Matlab 2024. PDF Documentation for Curve Fitting Toolbox, Curve Fitting Toolbox™ User's Guide.
- Quinlan, J.R. (1992). Learning with Continuous Classes. Proceedings of 5th Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore, 343-348.
- Širca, A., Rajar, R., Harris, R.C., Horvat, M. (1999a). Mercury Transport and Fate in the Gulf of Trieste (Northern Adriatic) – a Two-Dimensional Modelling Approach, *Environmental modelling & software* **14**, 645–655. [https://doi.org/10.1016/S1364-8152\(99\)00006-7](https://doi.org/10.1016/S1364-8152(99)00006-7).
- Širca, A., Horvat, M., Rajar, R., Covelli, S., Žagar, D., Faganeli, J. (1999b). Estimation of mercury mass balance in the Gulf of Trieste, *Acta Adriat.* **40**, 75–85.
- Tripathy, K.P., Mishra, A.K. (2024). Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions. *J. Hydrol.* **628**, 130458. <https://doi.org/10.1016/j.jhydrol.2023.130458>.
- Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., Butterfield, D., Rankinen, K., Lepisto, A. (2002). A nitrogen model for European catchments: INCA, new model structure & equations. *Hydrol. Earth System Sci.* **6**, 559–582. <https://doi.org/10.5194/hess-6-559-2002>.
- Wang, Y., Witten, I.H. (1997). Induction of model trees for predicting continuous classes. Poster papers of the 9th European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague.
- Whitehead, P. G., Wilson, E. J., Butterfield, D. (1998a). A semi-distributed integrated nitrogen model for multiple source in catchments INCA.: Part I – model structure and process equations. *Sci. Total Environ.* **210/211**, 547–558.
- Whitehead, P. G., Wilson, E. J., Butterfield, D., Seed, K. (1998b) A semi-distributed integrated flow and nitrogen model for multiple source assessment in catchments, INCA. Part II, Application to large River Basins in South Wales and Eastern England. *Sci. Total Environ.* **210/211**, 559–583.
- Witten, I.H., Frank, E., Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Burlington, MA, USA. <https://doi.org/10.1016/B978-0-12-374856-0.00018-3>.
- Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.J., Zhang, H. (2021). Machine Learning: New Ideas and Tools in Environmental Science and Engineering, *Environ. Sci. Technol.* **55**, 12741-12754. <https://doi.org/10.1021/acs.est.1c01339>.
- Zhu, S., Zhang, Z., Žagar, D. (2018). Mercury transport and fate models in aquatic systems: A review and synthesis. *Sci. Total Environ.* **639**, 538-549. <https://doi.org/10.1016/j.scitotenv.2018.04.397>.
- Žagar, D., Knap, A., Warwick, J.J., Rajar, R., Horvat, M., Cetina, M. (2006). Modelling of mercury transport and

transformation processes in the Idrijca and Soca river system. *Sci. Total Environ.* **368** (1), 149–163.
<https://doi.org/10.1016/j.scitotenv.2005.09.068>.