Irrigation Reservoir Modeling in Catchments without Measurements of Stream Flow

Milan Cisty, Veronika Soldanova, Barbora Povazanova
Department of Land and Water Resources Management
Slovak University of Technology
Bratislava, Slovakia
e-mail: milan.cisty@stuba.sk

Abstract— **This paper considers the determination of time series of river flows in catchments without direct monitoring of this variable. This paper proposes a method for the acquisition of monthly data, which is useful for various purposes. Different parameters of various water management structures can be determined based on information from such data series, such as irrigation reservoir volumes or water demand for irrigation. While identifying unknown stream flows required for such a calculation, authors suppose that historical climatic data for the given area and flows in nearby river catchments are available. This article includes a description of the method of selecting river catchments such that their measured flows can be used in the calculation of an unknown flow of a different stream. This study compares hydrological modeling, linear regression with regularization, and machine learning methods (support vector machines, random forest). Statistical indicators evaluate the calculated flows with the result that the most suitable approach is the support vector machines method using a linear kernel and LASSO regularisation.**

Keywords-stream flow; ungaged catchment; hydrologic modeling; LASSO; machine learning.

I. Introduction

Accurate modeling of flows at watersheds provides the information required to make optimal decisions and inform the optimal design of water management structures. For some purposes (e.g., flood protection), daily or hourly time series of flows are required, but this paper considers the acquisition of monthly data, which is sufficient for various purposes. Different parameters of various water management structures can be determined on the basis of information from such data series, such as irrigation reservoir volumes or water demand for irrigation. This paper considers a typical application requiring such monthly flow series, namely design of an irrigation reservoir.

Irrigation reservoirs are used to retain water during periods of surplus and to control its subsequent use for irrigation in drought periods. In moderately dry regions, such reservoirs are often used to address water management problems in small river catchments, or in marginal parts of larger river catchments. In such places, the irrigation area does not normally exceed a few hundred hectares. These smaller reservoirs are mostly formed by a front dam with a height of 5–12 m, with a volume not exceeding one million m^3 and a surface area of 2–70 ha. In addition to the irrigation function, they also ensure a minimum flow in the stream under the dam, protection against floods, and the creation of conditions for aquaculture. Smaller streams supply such reservoirs at the margins of river catchments. Such smaller streams often do not have systematic measurements of their flow, and therefore determination of this quantity is the subject of this paper.

In designing a reservoir, it is important to evaluate its function and assess its water management, e.g., its ability to provide the required amount of water for irrigation. The assessment of small water reservoirs represents a set of tasks dealing with the evaluation of the reservoir from the point of view of the quantitative balance of water. For irrigation reservoirs, this is mostly the seasonal, annual, rarely multi-year management of outflow and storage. The water management considerations include the management of the supply volume for securing of required functions (irrigation, recreation, fish breeding), managing the protection function of the reservoir volume (flood protection), and determining the requirements for the outflow, abstraction of water, and other parameters.

The input data used in the water management calculation of reservoirs include the water inflow into the reservoir, demand for water abstraction from the reservoir, data on compulsory outflow of water below the reservoir, and the evaporation and other losses of water from the reservoir. Such data are required retrospectively every month for at least 30 years. Such a long period should include sufficient occurrence of dry and wet years needed for the objective calculations. As stated, data on inflow are often unavailable, as there are rarely long-term measurements of the catchments with a small area.

The contribution of this article is a comparison of the various methods for determining river flows for the calculations needed when designing a small water reservoir. The authors are not aware of the previous works directly aimed at this topic, although, part of this task - flows determination in watersheds without measurements alone, was studied by various authors. A survey of regional methods used in Slovakia, from where the case study presented later in this paper is located, was reported by [1]. A good introduction to the topic is given by [2]. The determination of unmeasured flows can be conducted using hydrological models [3], regression methods [4], or artificial neural networks [5].

In this paper, several methods of undertaking the given task are compared. The objective is to acquire the monthly flows required for the balance calculation verifying the supply function of the irrigation reservoir. In Section II, the acquisition and preparation of the data are described. The methods applied in this study are briefly explained in Section III. In Section IV, the settings of the experimental computations are described, and the results are evaluated and discussed. Finally, Section V summarizes the main findings of this study.

II. Case Study and Data Description

This paper reports a case study of the Parna stream, which is a small mountain stream in the Small Carpathians in Western Slovakia. Its catchment area is 45.59 km^2. To determine the average daily flow in this stream, known flow data from similar nearby catchments are used (Figure 1).

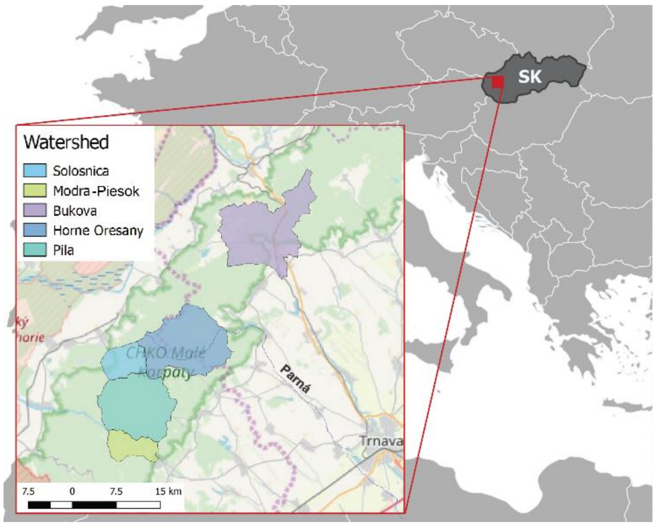

Figure 1. Situation of selected water catchments. © OpenStreetMap contributors (partly)

Data from water metering stations at Bukova (catchment of Trnavka), Modra (catchment of Vistucky stream), Pila (catchment of Gidra), and Solosnica (Solosnicky stream catchment) were used. The daily flow data for the river catchments were obtained from the Slovak Hydrometeorological Institute in Bratislava, Slovakia. Climatic data from the European Climate Assessment & Dataset (ECA&D) were also used in this study. ECA&D is composed of 69 participating organizations from 63 countries. The network of basic metering stations from which the data are derived covers the European and Mediterranean regions. It records 12 climatic elements. The main product of this initiative (E-OBS), which was used in this work, is a daily gridded observational dataset for precipitation, temperature, and sea level pressure for Europe. The climatic data in ECA&D are provided as a spatial time series in the netCDF format for the period 1950–2018, cover a spatial scope of 25°N–75°N and 40°E–75°E, and have a spatial resolution of 0.25°×0.25°. Data from 1 January 1980 to 31 August 2017 were used. Time series of the daily values of potential evapotranspiration in individual water catchments were calculated using the climatic data. The potential evapotranspiration was calculated using a formula proposed by Oudin in [6]. The advantage of Oudin's formula is that it only requires the minimum and maximum temperatures as inputs. As stated in this paper, this simplification of the inputs does not significantly affect the precision of the flow modeling in hydrological models.

III. Methods

The main objective of this paper is to reassess the water management function of small water reservoirs built in the past. Since they were built, the climate conditions have changed, and the demand for irrigation water is likely to have increased. The operating volume of reservoirs is often reduced from its original design value by years of sedimentation. As stated, the limiting factor in these calculations is the fact that the inflow into the reservoirs is, in many cases, not measured, i.e., it cannot be applied in the water balance calculation of the reservoir. The cause of this situation is that small-size streams often lack measurements. In this paper, regression methods and hydrological modeling are compared for determination of such unknown historical flows. To evaluate the performance of these methods, the case study for Horne Oresany reservoir on Parna stream was performed, e.g., in water catchment where the flows are known. The results of the computation methods were evaluated by suitable statistical indicators and by a comparison of the results of the reservoir water balance using either measured or simulated flows. A brief characterisation of these methods is given below.

The most common regression method is Multiple Linear Regression (MLR). MLR analysis is generally used to find the relevant coefficients in the following equation using the least-squares method. The basic equation is:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \ldots + \varepsilon, \qquad (1)$$

where Y is the dependent variable, X_i are explanatory variables, β_0 is the intercept (constant term), β_i is the slope coefficient for each explanatory variable, and ε is the model error term.

A major condition for linear regression is that the explanatory variables X_i must be relatively uncorrelated. However, some correlation is likely to occur in the task addressed in this study. More suitable algorithms than basic linear regression were therefore used. Least Absolute Shrinkage and Selection Operator (LASSO) applied in this paper redefine linear regression to prevent the effect of multicollinearity and help ensure a more stable model by penalizing and subsequently reducing the number of MLR coefficients [7].

A Support Vector Machine (SVM) is a supervised machine learning method that can be used to calculate regression tasks. Its characteristic feature is the kernel trick—a nonlinear mapping that transforms the original training data of a non-linear problem (which is the case in our scenario) into a higher-dimensional form [8]. Another important concept of the SVM methodology is its ability to ignore small errors. As a consequence, the SVM model has good generalization abilities.

Random Forests (RF) [9] are formed by a set of trees, which can either be classification or regression trees, depending on the problem being addressed. An RF prediction is an average of many trees (weak learners) grown on a bootstrap sample of the training data. The user chooses the

 ISBN: 978-1-61208-737-5

number of trees in the forest (ensemble). Each tree is trained using a different bootstrap sample, which causes that different trees are obtained. For the regression task, the values predicted by each tree are averaged to get the final random forest prediction.

Next method used for the calculation of unknown flows is the TUW hydrological model [10]. This model runs on a daily time step and consists of a snow routine, a soil moisture routine, and a flow routing routine. The snow routine simulates snow accumulation and melting using a degree-day concept. The soil moisture routine simulates runoff generation and changes in the soil moisture state of a catchment. Upper and lower soil reservoirs represent runoff routing. A genetic algorithm was used to calibrate the 15 parameters of this conceptual model.

The following formula generally defines the relationship between the inflow of water to the reservoir and outflow of water from the reservoir:

$$\Delta V = (I - O)\, \Delta_t, \tag{2}$$

where ΔV expresses the change in volume of water in the reservoir over time, I is the inflow to the reservoir, O denotes the outputs (outflow, water extraction, and losses), and Δ_t is the time step for the evaluation of the balance.

For this study, to implement the water balance in the reservoir, a computer program was set up in R [11] to model the accumulation of water in the reservoir. The basic objective is to reassess the feasibility of the requested water extraction from the reservoir. The model operates on a monthly time step. Different time step (e.g., two weeks) is also possible, but not necessary for this task, as experiences from building and managing irrigation reservoirs confirms (e.g., flood protection function of a reservoir, where a much smaller time step is required is an entirely different task than that, being solved herein). It is also necessary to note that the use of the previous period, when designing a new reservoir already defines a level of accuracy which will not be significantly improved by a more detailed time step. The hydrological data used in the model are also specified in monthly values.

IV. Results and Discussion

To determine the unknown flows required for the water balance model of the irrigation reservoir, the following steps were evaluated:

1. Analysis of the river catchment above the reservoir, in which the absence of flow measurements is assumed (Velke Oresany catchment, Parna stream), and selection of suitable river basins for their calculation using the hydrological analogy method.
2. Analogous calculation of unknown flows using hydrological modeling.
3. Analogous calculation of unknown flows using statistical and machine learning methods.
4. Calculation of water management balance of the reservoir with measured flows (used only in the context of testing) and with simulated flows acquired by various methods.
5. Evaluation and comparison of results.

A. Selection of Suitable River Catchments

The outflow regime of the river catchment depends on its climate conditions, on topographic features, geological conditions, types of prevailing soils in the river basin, land use of the area, etc. The river basins for the analogous calculation should be similar to the river catchment where the flow is to be determined. Moreover, it is advantageous to select analogous river catchments as close as possible to the catchment of interest, as some climate, topological, geological, and other properties change relatively smoothly, so nearby river catchments could have many similar features, and consequently can have a similar genesis of the outflow (Figure 1).

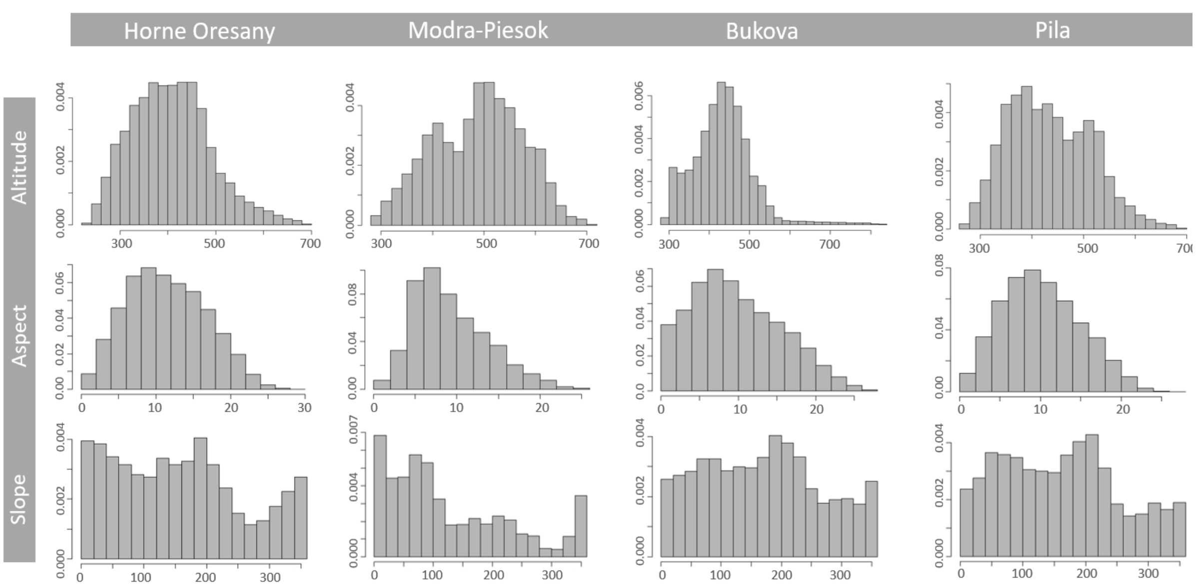

Figure 2 Histograms of altitudes [m a.s.l.], aspect [° north] and slope [° horizontal] of river catchments.

Several analyses were elaborated for this purpose; Selected parts of this analysis are shown in Figure 2 and described in Table I.

Figure 2 shows the analyses of the topography (altitudes, aspect, slope), as the nature of landscape cover, weather, and the amount and type of rainfall change with these properties. Such phenomena will impact water outflow from the river catchment (as will, for example, the winter flow regime). The river catchments used in this study are located at altitudes of approximately 300–700 m above sea level. This histogram representation of altitudes allows a visual assessment of the similarity of river catchments. The most similar catchment to the Horne Oresany catchment, where we want to determine the flows is the Pila basin in all the features shown (altitude, slope, aspect).

The representative part of the analyses is evaluated in Table I. This table and GIS analyses show that, for the assessed river catchments and from the point of view of catchment features influencing the outflow regime, the Parna river catchment is most similar to the Gidra catchment (Pila gauging station) and the Vistucky stream catchment (Modra-Piesok station). Similarity can be seen in numerical evaluation of basic topographic characteristics, percentage of soil types and land use. Different soil types vary in terms of the ratio of infiltration and the outflow of water during periods of rain, in the ability to retain water in the soil, and in other properties that influence the outflow of water from the river catchment. Similarly, land use (for example, a forest versus arable land) also has a significant impact on river basin drainage properties. These two catchments will be therefore preferred in the following calculations. Also other analyses were accomplished, but cannot be described in more detail here for reasons of brevity.

B. Hydrological Modeling

The unknown flows were calculated by the TUW conceptual rainfall-runoff model, with a genetic algorithm used for calibration. The calculation was performed in daily step, and the daily flows were subsequently converted into monthly flows. The calibration was implemented based on flow and climate data from the Pila catchment which, based on previous analysis, was assessed as being the most similar to the river catchment in which the unknown flows were to be calculated. The optimum parameter values of the TUW model were acquired by the genetic algorithm using the flow and climate data from the Pila catchment. These parameters were subsequently applied in the modeling of the river catchment with the unknown flows (Velke Oresany–Parna) using the local climate data. The genetic algorithm population was set to 500, the number of parameters to be determined was 15, and the maximum number of generations was 20. The objective function sought to minimise the Mean Absolute Scaled Error (MASE), a statistical variable suggested by Hyndman [12]. This statistic is preferable to the Nash–Sutcliffe Efficiency [13] in this case, as the final objective is to calculate monthly flows. This is because the MASE does not take into consideration the power of flows, and thus does not emphasize the calculation of large daily values, which are not herein priorities because of the transformation (averaging) of calculated daily values into monthly values. The calculated daily and monthly inflows to the Horne Oresany reservoir are compared with the measured values in Table II.

C. Regression Calculation

To calculate the flows at the Horne Oresany river reservoir on the Parna stream, regression methods were also applied. These computations used the assumption that flow measurements had recently started in the Parna river catchment as a requirement for the determination of longer series of historical flows. The measurements were assumed to have started at the beginning of 2016 (whole period is 1980–2017). The regression relations were then derived based on this period and applied for the whole historical period of interest.

TABLE I. COMPARISON OF RIVER CATCHMENT FEATURES

Watershed	Area (ha)	Surface Characteristics (median)			Soil Types (%)			Land Use (%)				
		Elevation (m a.s.l.)	Slope (°)	Aspect (°)	Loamy	Loamy-sand	Sandy-loam	Arable land	Broad-leaved forest	Mixed forest	Transitional woodland/shrub	Urban fabric
Bukova	4296.1	332.4	9.0	170	82.5	0	17.5	23.9	60.7	6.9	2.1	4.9
Modra-Piesok	937.7	495.1	8.0	92.9	0	54.3	45.7	0	91.5	0	8.5	0
Horne Oresany	3733.1	403.1	11.1	151	38	0	62	0	90.9	0.8	8.2	0
Pila	3289.9	426.7	9.8	161	12.8	16.3	70.9	0	92.4	0	7.1	0.5
Solosnica	1046.5	420.8	16.3	195	100	0	0	0	94.4	0	5.6	0

Methods described in the previous subsections were used for the regression calculation. A grid search combined with a repeated cross-validation methodology was used to find the parameters of these models. In this approach, a set of model parameters from a predetermined grid is sent to the evaluating algorithm. A set of parameters was sent to the repeated cross-validation mechanism, which is used for the evaluation of the parameter combinations [14]. The calculation was performed in R languague [11].

The unknown flows in the Parna stream were calculated using flows from four “analogous” river catchments and from the average daily temperature, rainfall, and potential evapotranspiration in the Parna river catchment. As the flow from a catchment is influenced not only by the current values of climate variables, but also by their values from previous days, climate data from seven days before the date of the prediction were also included. As the longer history of hydro-climatic developments in the catchment must also be described in the input data, three variables summarizing the previous precipitation (cumRAIN7, cumRAIN14, cumRAIN21) and variables summarizing the previous evapotranspiration (cumPET7, cumPET14, cumPET21) were constructed. (The numbers in these variable names denote how many days they are summarising.) In this way, a training set with 35 explanatory variables was created. This set covers the period of anticipated short-term measurements on the Parna stream for 608 days (only data up to August were available for 2017). The test file includes the same variables, but the data relate to the whole period 1980–2017, i.e., it contains 13,738 lines (one per day).

TABLE II. EVALUATION OF MODELS

Statistic	TUW	MLR	LASSO	SVM	RF
RMSE	0.41	0.14	0.14	0.14	0.15
NSE	0.65	0.78	0.78	0.78	0.73
r	0.81	0.89	0.89	0.89	0.88
R2	0.66	0.79	0.79	0.79	0.77
VE	0.65	0.72	0.72	0.72	0.70

RMSE – Root Mean Sqare Error, NSE - Nash-Sutcliffe Efficiency, r - Correlation Coefficient, R2 – Coefficient of Determination, VE - Volumetric Efficiency

The results of the regression calculation are presented in Table II, which indicates that the regression provides relatively balanced results using different statistical and machine learning methods. The best methods are SVM, the regularised linear method using the LASSO-type regularisation, and, surprisingly, the simple MLR. However, the latter method cannot be recommended, as linear regression requires the explanatory variables to be relatively uncorrelated with each other. This multicollinearity principle is violated for this task, and the results using MLR are therefore expected to be unstable; for other streams, they may be less precise than in this case. Relatively poor results are obtained by the random forest technique (RF), which is a popular ensemble tree method. The authors believe that the problem with the random forest method is that it is based on regression trees which have no extrapolation ability. Described regression task used a relatively short training period, and there is a high possibility that there were no extremes during this period. Thus, the random forest was not able to learn such events. Although this does not influence so much, e.g., the simple linear regression, for RF method it is quite important.

D. *Water Balance of the Reservoir*

The water balance of the reservoir was computed using the measured and calculated inflows to the reservoir to verify whether the precision of the monthly flows computed by various analogy methods is sufficient for this purpose. To calculate the balance, a function was programmed in R to take various input data. The main data are the required time series of monthly irrigation amounts and water inflow to the reservoir from Parna stream. Additionally, these data include the requisite altitudes of the minimum and maximum reservoir levels, the minimum and maximum water volume in the reservoir, the irrigated area, evaporation data from the water surface, and the ecological flow that must be respected below the reservoir. The water body of the reservoir is defined by a curve of water surface areas and water volumes depending on the water level in the reservoir. This curve (or table) can be determined from a topographic survey and is also part of the input data.

The balance of the reservoir was firstly calculated using flows calculated by analogy and then using the measured data. Two methods were used to compare the acquired results. The first was the establishment of the irrigation security, which is the ratio of supplied and required irrigation as a percentage:

$$\text{IrrSecured} = (\text{ZSupplied}/\text{ZRequired})*100, \qquad (3)$$

The correlation among the computed annual amount of irrigation water has been also evaluated on the basis of the computations with measured and calculated flows. The resulting data are summarised in Table III, which demonstrates the suitability of the applied methods. The security of irrigation (availability of water in the reservoir) is also expressed in Figure 3, where a relatively high agreement can be seen as regards the security of irrigation when using the calculated and measured flows.

TABLE III. IRRIGATION SECURITY EVALUATION

Method	Yearly correlation with computed by measured flows	Irrigation security in %
Measured flows	1.00	84.96
TUW	0.90	90.23
MLR	0.90	86.79
LASSO	0.92	87.77
SVM	0.90	87.04
RF	0.89	89.88

 ISBN: 978-1-61208-737-5

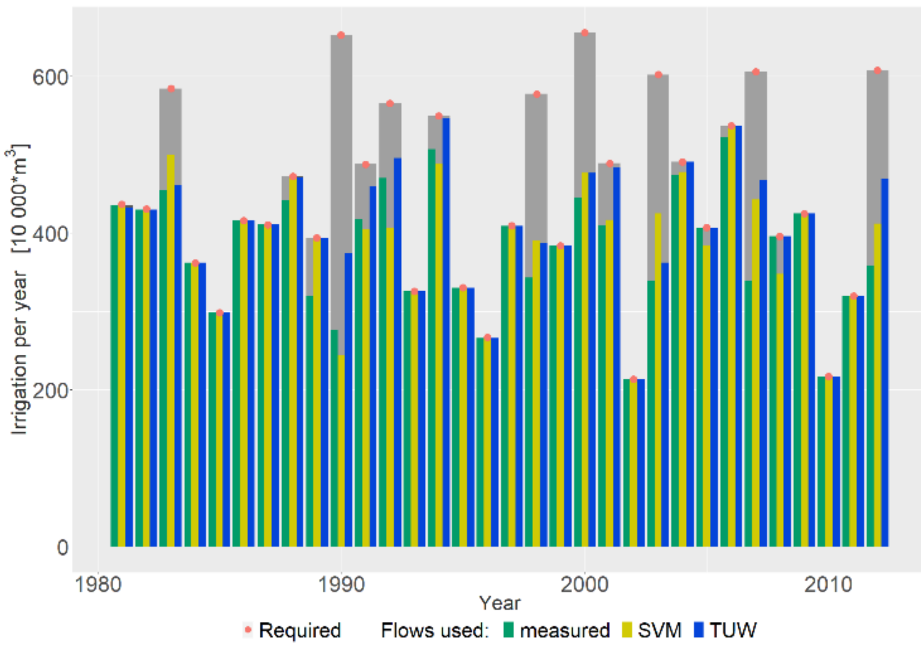

Figure 3. Graphical comparison of security of irrigation.

V. CONCLUSION

The objective of this study was to compare various methods of calculating unknown flows in the context of engineering computations, such as evaluation of water balance of an irrigation reservoir. The methods compared include regression and hydrological modeling methods. If at least short-term measurements are available for the relevant river catchment, a most suitable method is the flow calculation method using regression with LASSO regularisation, as this eliminates the problem of multicollinearity in the input data. Another suitable method is the machine learning SVM method, which offers good generalisation ability. This is a major advantage for computations on small river catchments (such as that used in this study), where relatively significant uncertainties can be expected as regards data and modeling. If no flow data are available, a hydrological model must be used. In this paper, the use of the TUW model was applied. Computation of flows by TUW model was slightly less precise than by using regression methods, but the subsequent verification of calculated flows using TUW model in the context of an irrigation reservoir balance, also demonstrated its usability in the context of engineering calculations.

ACKNOWLEDGEMENT

This work was supported by the Slovak Research and Development Agency under Contract No. APVV-15-0489 and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, Grant No. 1/0662/19.

REFERENCES

[1] S. Kohnova, J. Szolgay, L. Solin, and K. Hlavcova, “Regional methods for prediction in ungauged basins,” Ostrava: Key Publishing, p. 113, 2006, ISBN: 80-87071-02-6

[2] M. Hrachowitz et al., “A decade of predictions in ungauged basins (PUB)—a review,” Hydrological sciences journal, vol. 58(6), pp. 1198-1255, 2013.

[3] A. Viglione et al., “Comparative assessment of predictions in ungauged basins - Part 3: Runo signatures in Austria,” Hydrology and Earth System Sciences, vol. 17, pp. 2263-2279, 2013, doi:10.5194/hess-17-2263-2013, 2017

[4] M. S. Gibbs, H. R. Maier, and G. C. Dandy, “A generic framework for regression regionalization in ungauged catchments,” Environmental modeling & software, vol. 27, pp. 1-14, 2012.

[5] V. Kuzmin et al., “Method of prediction of the stream flows in poorly gauged and ungauged basins,” Journal of Ecological Engineering, vol. 20(1), pp. 180-187, 2019.

[6] L. Oudin, C. Michel, and F. Anctil, “Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?,” Journal of Hydrology, vol. 303/ Nos. 1-4, pp. 275-289, 2005.

[7] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” Journal of Statistical Software, vol. 33(1), pp. 1-22, 2010.

[8] V. Vapnik, “The Support vector method,” International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, pp. 261-271, 1997.

[9] L. Breiman, “Random forests,” Machine Learning, vol. 45(1), pp. 5-32, 2001.

[10] A. Viglione and J. Parajka, “TUWmodel: Lumped hydrological model for education purposes,” R package

version 1.0-1., 2018, Available from: https://CRAN.R-project.org/package=TUWmodel [retrieved July, 2019]

[11] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2019, Available from: https://www.R-project.org [retrieved July, 2019]

[12] R.J. Hyndman, and G. Athanasopoulos, "Forecasting: principles and practice," OTexts, 2018.

[13] F. Lin, X. Chen, and H. Yao, "Evaluating the use of Nash-Sutcliffe efficiency coefficient in goodness-of-fit measures for daily runoff simulation with SWAT," Journal of Hydrologic Engineering, vol. 22(11), p.05017023, 2017.

[14] C. Bergmeir, R.J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," Computational Statistics and Data Analysis, 120, pp.70-83, 2018.