Predicting the Peak Water Level of Flood Waves by Using Statistical Methods on River Rába

PhD student **Gábor KERÉK**[1]

[1] Deputy Head of Department, North Transdanubian Water Directorate, Győr, Hungary,
kerek.gabor@eduvizig.hu

Abstract: *This year, as a result of a joint Hungarian-Austrian development, an early flood warning system of the Rába river basin was completed, based on hydrodynamic model support. The warning system calculates the water flows and levels with a 6-day time advantage in the defined water gauge sections of the catchment. The system also includes a 2-dimensional flood model subsystem, which models the evolution of valley floods in case of a flood wave event.*

As a part of this project, an application was also implemented, which allows statistical-based parameter estimation onto flood peak levels. The basic mathematical method of this app is the multivariable linearized regression. Its parameters are characteristics water levels or water discharges that measured in the same water gauge sections as in the hydrodynamic model.

In this study, I describe the hydrological foundations of the database serving this application and then I evaluate the first results.

Keywords: *Hydrological statistics, Rába, Linearized regression, Flood forecasting, Numerical analysis*

1. River Rába and its catchment, Flood forecasting issues on Rába

Rába is one of the most significant tributaries of the Danube in Hungary. Its origin is in Styria province of Austria, in the Fischbach Alps, at an altitude of around 1200 m above sea level, from two branches. It crosses the Austrian-Hungarian border at Szentgotthárd, flows across Kisalföld region of Hungary, and reaches its receiving river Mosoni-Duna in the city of Győr. The total length of the river is 283 km, and its section in Hungary is 211.5 km. The catchment area is 10270 km^2. [1] One third of its catchment area is in Austria and two thirds in Hungary. The catchment area of river Rába is shown on the next figure.

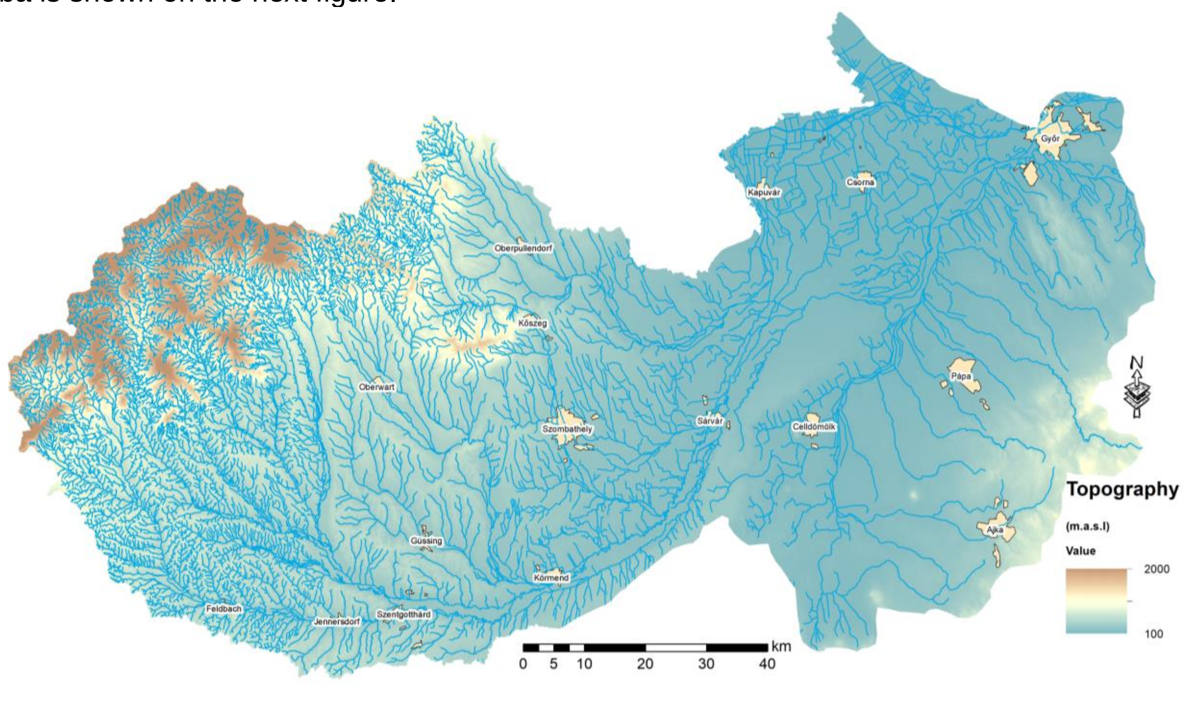

Fig. 1. Rába river basin

The current and the expected hydrological regime of our rivers is the basic of several water management tasks, and it has an essential role in the work of flood defense organizations and in disaster emergency planning. The newly developed early flood warning system on the river basin of Rába can partially fulfil this role, the hydrodynamic model system operating on meteorological boundary conditions estimates the expected water flow and water level in some cross-sections of significant rivers of the river basin for a 6 days ahead. However, in case of flood events, additional forecasts are also made in order to support the necessary measures in technical, logistical and economic sense in flood protection. This method basically not based on the river basin's meteorological forecasts, but on a statistical analysis of long-term water level time series measured in water gauge profiles. The method is based on the principle of linearized regression and it's also suitable for the continuous prediction of time series, however, with an appropriate pre-processing of the time series, the peak water levels of the flood waves can be determined with sufficient accuracy. The processed time series are derived from the long-term measured water level time series of the water system. Their location is shown in Figure 2.

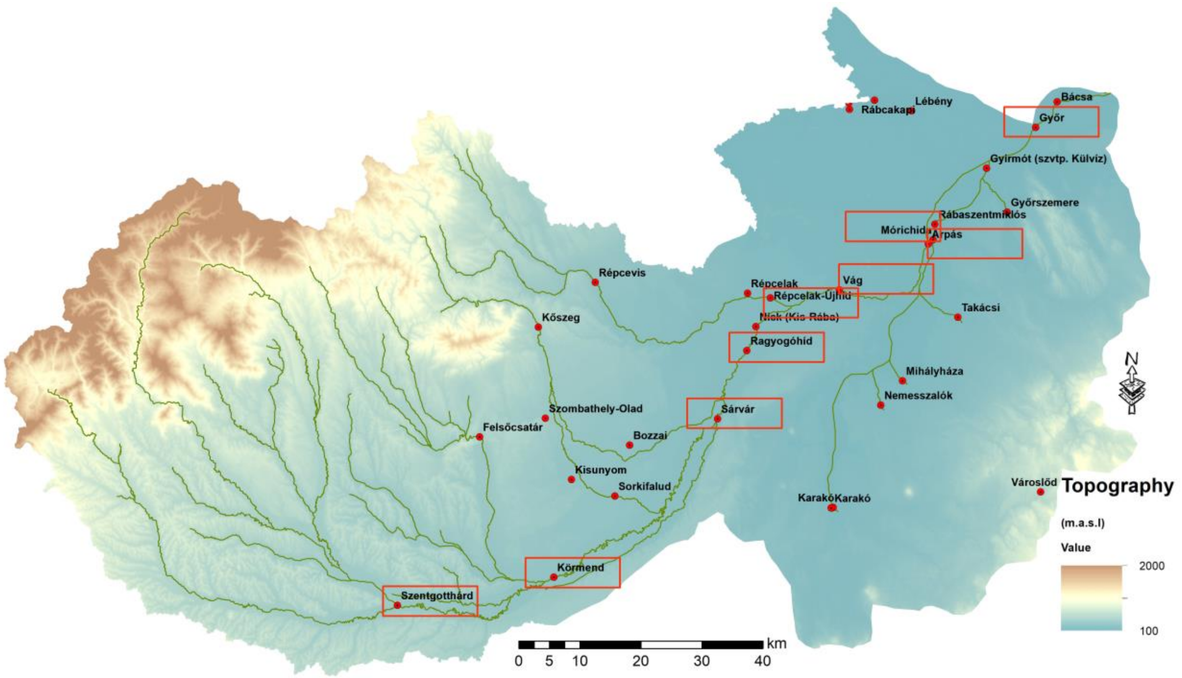

Fig. 2. Water gauges on the river basin of Rába

First, let's briefly review the mathematical foundations of regression analysis.

2. Mathematical basics of linearized regression

Due to its flexibility, this method is applicable for comparing time-invariant datasets, for parameter-estimation, if there is a physical relationship between the individual probability variables. The procedure can be used in cases where the relationship between two probability variables is not clearly defined. In this case, the values of the independent variable belong to a statistical set of the dependent variable, so that the distribution of the dependent variable together with the distribution of the independent variable changes in a specific way. In this case, there is a correlation between the variables. In a correlation relationship, the relationship can be given between the expected values of the variables; thus, the correlation occupies an intermediate space between the well-defined function relationship and the complete independence of the variables (stochastic relationship). The closeness of the relationship is expressed by the correlation coefficient (r).

$$0<r<|1| \quad (1)$$

where the value of the correlation factor 0 refers to the complete independence of the variables and 1 refers to the exact function relationship.

The following tasks may be possible during the regression analysis:

- estimating the parameters (constants) of a function describing the relationship between two variables
- examination of the linearity hypothesis (statistical test of correlation coefficient)
- examining the assumptions about the parameters of the fitted function
- calculation of the confidence interval
- estimating the parameters of the function and the margins of error of physical quantities [2]

In regression analysis, the basic task is to fit a line or curve to a point cloud expressed by the relationship of the variables so that the difference between the points and the theoretical function is as small as possible. Obviously, no curve can be drawn that touches all points; thus, the task is to choose a relation based on theoretical considerations in the knowledge of the examined system, which enforces the physical regularities of the variables as best as possible. E.g. with the usage of power functions any point cloud can be described with an appropriate precision, but it's doubtful that the theoretical function has physical meaning, or the precision can only mathematically proved. The aim is to define an equation in which by substituting the different values of the independent variable (x), we can estimate the values of the dependent variable (y) as accurate as possible. To define this equation the least-squares-based regression analysis is used. [2]

Principle of least squares: the function describing the relationship between the variables and its parameters are defined so that the sum of the squares of the differences between the measured dependent variable values and the values calculated by substituting the same independent variable from the relationship is minimal. In the simplest case the equation of a line is fitted to the point cloud. [2]

The equation of a line is:

$$\boldsymbol{y} = \boldsymbol{a} + \boldsymbol{b} * \boldsymbol{x} \quad (2)$$

Based on the principle of least squares, the most probable values of the parameters are: the minima of the sum of the squares of the deviations of the dependent variable calculated from the function and the measured variable

$$Q = \sum_{j=1}^{n} (y_j - f(x_j))^2 = \sum_{j=1}^{n} (y_j - a - bx_j)^2 = \min. \quad (3)$$

In the equation we consider the parameters of the function (a, b) as variable, and the measured value pairs x_j and y_j are fixed.

The condition of minima:

$$\frac{\partial Q}{\partial b} = 0; \qquad \frac{\partial Q}{\partial b} = \sum_{j=1}^{n} -2(y_j - a - bx_j)x_j = 0$$
$$\frac{\partial Q}{\partial a} = 0; \qquad \frac{\partial Q}{\partial a} = \sum_{j=1}^{n} -2(y_j - a - bx_j) = 0 \tag{4}$$

, these equations ordered and simplified are the following

$$\sum_{j=1}^{n} x_j y_j = b\sum_{j=1}^{n} x_j^2 + a\sum_{j=1}^{n} x_j$$
$$\sum_{j=1}^{n} y_j = an + b\sum_{j=1}^{n} x_j \tag{5}$$

In this form the 'a' and 'b' parameters can be determined directly:

$$b = \frac{n\sum_{j=1}^{n} x_j y_j - \sum_{j=1}^{n} x_j \sum_{j=1}^{n} y_j}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2} = \frac{\overline{x * y} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2}$$
$$a = \frac{\sum_{j=1}^{n} y_j - b\sum_{j=1}^{n} x_j}{n} = \overline{y} - b * \overline{x} \tag{6}$$

The difference between the value pairs xj and yj and the y values calculated from the calculated equation of the line defined by the parameters a and b is called the residue. Based on this, the residual standard deviation can be calculated with the following formula

$$S^{*2}_{reziduális} = \frac{\sum_{j=1}^{n}\left[(y_j - a - bx_j)^2\right]}{n-2} = \frac{\sum_{j=1}^{n}(\Delta y_j)^2}{n-2} \tag{7}$$

, where $\sum_{j=1}^{n}(\Delta y_j)^2$ is the sum of the squares of the deviation. This quantity is the fit standard deviation and it describes the regression fit. The residual standard deviation defines a confidence interval around the regression line in both directions, which can be used to examine the quality of the fit. By using the equation of line, the standard deviation of the 'b' parameter (the slope of line) can also be determined:

$$S_b = \frac{S^{*}_{reziduális}}{S_x\sqrt{n}} = \frac{S^{*}_{reziduális}}{\sqrt{\sum^{n}(x_j - x)^2}}$$
$$S_x = \sqrt{\frac{\sum_{j=1}^{n}(x_j - x)^2}{n}} \tag{8}$$

ISSN 1453 – 7303 **"HIDRAULICA" (No. 4/2020)**
Magazine *of Hydraulics, Pneumatics, Tribology, Ecology, Sensorics, Mechatronics*

The confidence interval is the following: $b \pm t_Q * S_b$ (9)

, where

t_Q is the probability value of the distribution 't' with probability 'Q', so it can be stated with probability 1-Q, that the true value of parameter 'b' is within this interval.

3. Building the database of regression parameters

The aim of the regression analysis is to predict the peak water levels of the flood waves in Rába by processing past hydrological data. In this first stage of the evaluation, I defined a regression-based forecasting model for the main water gauges of the Rába. These prediction points are marked with red squares in Figure 2.

This goal can be achieved by building and applying a well-constructed database that estimates the expected value of a selected target parameter (as dependent variable) based on a statistical sample of the behavioral characteristics of several flood waves in the past. To this end, hydrological data time series of 147 flood waves has been organized into an SQLite database from a total of 13 hydrographic stations between 1965 and 2019 [3].

The calculation parameters of the model are the peak water levels of each flood wave at the indicated water gauge profiles, as shown in Fig. 3. The physical properties of flood waves can be modeled by adding additional parameters. When a flood occurs, the magnitude of the peak water levels crucially depends on the load of the riverbed. When a flood wave arrives onto a riverbed section, and the riverbed is relatively empty, a significant part of the flood volume will be lost for filling the bed, and the longitudinal peak water discharges decreasing from section by section. This effect can be modeled by a parameter, we called 'simultaneous water levels'. This means, that additional values are stored in the database, the actual water levels in the time of an upper station's flood peak.

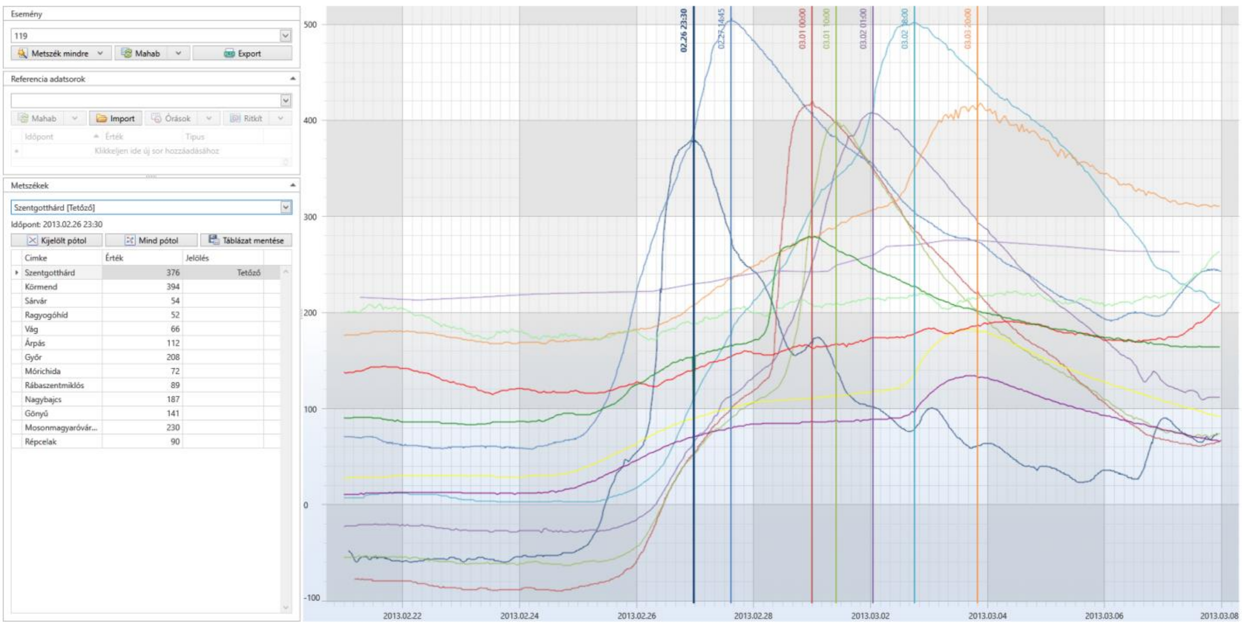

Fig. 3. Evaluation of the flood waves [4]

A part of the database numerically represented in the following figure:

Flood wave			Simultaneous water levels (cm)											Peak water levels (cm)										
ID	Time of Szentgotthárd peak	Szentgotthárd	Körmend	Sárvár	Ragyogóhíd	Vág	Árpás	Győr	Mórichida	Rábaszentmiklós	Nagybajcs	Gönyű	Répcelak	Körmend	Sárvár	Ragyogóhíd	Vág	Árpás	Győr	Mórichida	Rábaszentmiklós	Nagybajcs	Gönyű	Répcelak
1	1982.01.02 15:12	215	345	78	42	87	23	250	90	111	298	235	104	388	165	118	190	213	442	123	130	506	480	166
2	1982.05.24 23:51	182	110	-48	-63	-9	-18	318	-2	-37	416	350	70	250	68	39	90	110	385	22	-25	450	414	72
3	1982.08.09 05:02	84	218	72	41	87	72	302	72	41	309	253	202	264	97	58	134	162	319	100	84	332	284	208
4	1982.10.07 18:42	353	417	166	115	169	210	291	-13	-11	169	120	201	473	362	294	362	420	360	8	52	198	158	288
5	1982.10.07 18:42	353	417	166	115	169	210	291	-13	-11	169	120	201	473	362	294	362	420	392	30	73	198	158	288
6	1982.10.14 22:43	277	340	113	91	179	237	336	11	27	170	140	185	440	362	294	362	420	392	30	73	234	188	274
7	1982.11.14 21:00	131	181	-55	-39	7	12	157	-30	-34	103	54	100	303	106	78	150	183	260	17	14	154	99	171
8	1982.12.19 06:00	231	243	-10	-6	36	50	209	5	-4	203	157	156	392	159	118	205	234	294	80	82	266	206	250
9	1982.12.19 06:00	231	243	-10	-6	36	50	209	5	-4	203	157	156	388	173	130	212	281	366	175	181	286	252	250
10	1982.12.19 06:00	231	243	-10	-6	36	50	209	5	-4	203	157	156	388	173	130	212	281	366	175	181	286	252	250
11	1984.03.02 19:50	75	180	55	-27	20	19	168	-9	-15	128	74	90	250	67	63	127	142	273	6	-6	142	90	115
12	1984.04.04 12:25	226	252	-12	-20	23	5	183	-30	-43	237	178	103	317	100	74	142	170	260	-14	-29	259	202	143
13	1984.05.30 03:00	113	90	-48	-45	7	4	218	-8	-27	287	227	105	251	51	43	99	101	260	-1	-21	322	257	105
14	1985.03.23 05:35	296	367	101	72	147	179	260	63	61	179	119	152	438	295	238	292	381	374	87	74	301	250	211
15	1985.05.08 01:00	266	289	-40	-52	22	15	209	-12	-22	277	201	115	380	136	108	176	229	290	7	-2	310	236	169
16	1985.05.17 01:00	279	96	-42	-47	24	33	245	-18	-34	338	262	63	314	88	65	130	157	306	60	61	373	306	174
17	1985.08.07 16:00	253	87	-94	-90	-34	-30	242	-45	-60	350	268	54	263	59	39	97	126	600	14	59	659	631	126
18	1985.12.30 06:24	73	95	-84	-85	-20	-19	156	53	48	159	85	74	251	56	41	87	123	234	113	114	222	146	124

Fig. 4. Part of the database in numerical form

In the present study, I set up a regression model with multivariable linearized regression to estimate the peak water levels at Körmend, Sárvár and Árpás based on the Szentgotthárd peak water level. To perform this task, the parameters shown in the following table were used for each flood wave:

Table 1: regression parameters used for the model

Independent variable	**Physical meaning of parameter**
$H_{Kö\ T}$	Peak water level at Körmend (cm)
$H_{Sztg\ T}$	Peak water level at Szentgotthárd (cm)
$H_{Sá\ T}$	Peak water level at Sárvár (cm)
$H_{Ár\ T}$	Peak water level at Árpás (cm)
$H_{Kö\ T0Sztg}$	Simultaneous water level at Körmend at the time of Szentgotthárd peak (cm)
$H_{Sá\ T0Sztg}$	Simultaneous water level at Sárvár at the time of Szentgotthárd peak (cm)
$H_{Ár\ T0Sztg}$	Simultaneous water level at Árpás at the time of Szentgotthárd peak (cm)

4. Multivariable linearized regression

In the case of multivariate linear regression, the target dataset is not a function of one, but of several independent datasets. In this case, the regression result is obtained as the result of a first-order polynom; thus, the estimate is obtained as the sum of the independent variables. The aim is to determine the coefficients of the polynom, its principle is the least squares method. [2] The polynomial regression in this case takes the following form:

$$y = a_1 * x_1 + a_2 * x_2 + ... + a_n * x_n \tag{10}$$

To minimize the difference between the defined formula and the measured target time series, partial derivatives of the estimation function must be generated:

$$\sum_{i=1}^{N} (Y_i - F(x_i))^2 \tag{11}$$

, the zero positions of this partial derivatives need to be determined.

The regression model for the above-mentioned water gauges were performed with this method. The calculated function for Körmend is shown on Fig.5. As shown in the figure, a linearized function describes the relationship of the variables, which is a polygon corresponding to the equation (), and accordingly all independent variables that can be involved to the calculation has its first and second powers in the equation. Higher degree polynomials can also be fitted by the application, which describe the relationship of the parameters with a tighter fit, but this can only be verified mathematically. Knowing the physical relationship of the variables, it is advisable to approach the relationship with a polynomial with the lowest possible degree. [2]
In this first part of this study 3 main water gauge sections were involved to this multivariable regression, to estimate the peak water level of the floods. In the next chapter I describe the regression model for the estimation of their water levels.

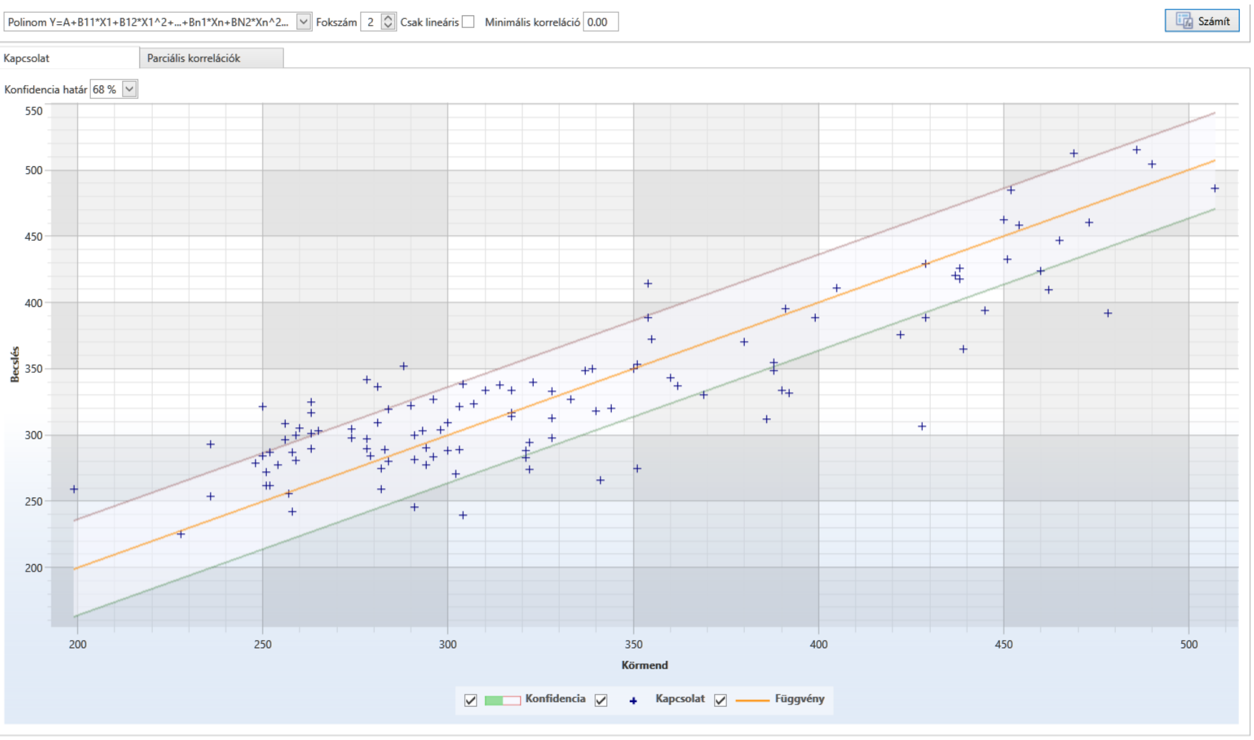

Fig. 5. Multivariable linearized regression for the predicted water level at Körmend [4]

5. Regression model for the main water gauges on Rába from Szentgotthárd peak

The polynomial regression model is shown in the following formulas. The model can be applied from the time of the peak of the Rába in Szentgotthárd. Using formula (12), the estimated peak water level in Körmend can be determined.

$$H_{Kö\,T} = 211{,}3 + 0{,}08 * H_{Kö\,T0Sztg} + 0{,}0007 * H^2_{Kö\,T0Sztg} + 0{,}11 * H_{Sztg\,T} + 0{,}00055 * H^2_{Sztg\,T} \quad (12)$$

The two lower stations' peak water level can be estimated by the following equations (13, 14). In this step, the prediction for Sárvár can be estimated on the basis of the measured Szentgotthárd and the estimated Körmend peaking from the first step. Of course, due to the usage of an estimated parameter, the result has a greater uncertainty than just the uncertainty from the regression.

$$H_{Sá\,T} = 395{,}8 - 2{,}81 * H_{Kö\,T} + 0{,}005 * H^2_{Kö\,T} + 0{,}62 * H_{Sztg\,T} + 0{,}002 * H^2_{Sztg\,T} - 0{,}00004 * H^2_{Sá\,T0Sztg} + 0{,}4 * H_{Sá\,T0Sztg} \quad (13)$$

$$H_{Ár\,T} = 54{,}1 + 0{,}11 * H_{Kö\,T} + 0{,}0005 * H^2_{Kö\,T} - 0{,}025 * H_{Sztg\,T} - 0{,}0002 * H^2_{Sztg\,T} + 0{,}35 * H_{Sá\,T} + 0{,}0006 * H^2_{Sá\,T} + 0{,}32 * H_{Ár\,T0\,Sztg} - 0{,}0004 * H^2_{Ár\,T0Sztg} \quad (14)$$

This phenomenon also valid onto the estimation for Árpás, at the lower section of the Rába (formula no. 14)

6. Conclusions

The quality of the linearized regression can be described by 2 characteristic values: the standard deviation, and the differences between the measured and the calculated values.

The standard deviation of the predicted water levels for each station is between 17… 30 cm, the average of the differences between the measured and estimated water levels is between 15… 25 cm.

The presented method can be used to predict the peak water level of discrete flood waves. The dynamics of flood waves can also be estimated by adding additional parameters (e.g., evaluating daily water level changes).

It is absolutely necessary to integrate the hydrological data measured on the tributaries of the Rába (Pinka, Gyöngyös, Strém, Perint, Lapincs), as well as to include the Austrian section of the catchment.

Further investigation is required to prepare the regression database for continuous water regime forecasting in addition to discrete flood wave forecasting.

Summarized, the linearized regression procedure, applied for both the upper and lower section of the river basin, is suitable to support flood protection with a sufficient time advantage.

The regression application was developed as a subsystem of Rába's early flood forecasting system.

The forecasting system uses hydrodynamic models to make a forecast for the already mentioned water gauges with a time advantage of 6 days.

The results of the regression procedure presented in this study may provide a good control for evaluating the results of the forecasting system.

Rába's flood forecasting system and its results are available on the website https://rf4c.vizugy.hu.

References

[1] West-transdanubian Water Directorate / Nyugat-dunántúli Vízügyi Igazgatóság. "IDENTIFICATION OF SIGNIFICANT WATER MANAGEMENT ISSUES / PROBLEMS IN THE RÁBA RIVER BASIN MANAGEMENT PLANNING UNIT” / „JELENTŐS VÍZGAZDÁLKODÁSI KÉRDÉSEK/PROBLÉMÁK AZONOSÍTÁSA A RÁBA VÍZGYŰJTŐ-GAZDÁLKODÁSI TERVEZÉSI ALEGYSÉGEN.” [Online]. Available: http://www.nyuduvizig.hu/upload/1.1.4.Raba_kesz-bovitett.PDF. Accesed December 10, 2020.

[2] Hüse, Zsolt. "Computer program for the hydrological application of nonlinear methods of univariate and multivariate regression analysis.” / „Egy-és többváltozós regresszióanalízis nemlineáris módszerei hidrológiai alkalmazásának számítógépi programja.” ATIKÖVIZIG, Szeged, 2001.

[3] National Directorate General for Water Management / Országos Vízügyi Főigazgatóság. "Hungarian Hydrological Database (Unified online hydrological data storage and data processing system).” / „Magyar Hidrológiai Adatbázis (Egységes online hidrológiai adattároló és adatfeldolgozó rendszer).” Budapest, 2015 (available: North-transdanubian Water Directorate, 28-32. Arpad St. Gyor, Hungary).

[4] H-FIT Informatikai Bt. Regress4. Regression application / Regress4. Regresszió alkalmazás, 2019. (available: North-transdanubian Water Directorate, 28-32. Arpad St. Gyor, Hungary).