ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 64 58 Number 2, 2016

http://dx.doi.org/10.11118/actaun201664020497

THE USE OF REGRESSION TRIPLET IN HYDRAULIC MODELLING ON THE EXAMPLE OF DETERMINATION OF OVERFALL COEFFICIENT

Petr Pelikán[1], Jana Marková[1]

[1] Department of Landscape Management, Faculty of Forestry and Wood Technology, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Abstract

PELIKÁN PETR, MARKOVÁ JANA. 2016. The use of Regression Triplet in Hydraulic Modelling on the Example of Determination of Overfall Coefficient. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 64(2): 497–502.

The paper is focused on a hydraulic problem of water overfall on hydrotechnic structures, especially outlets and spillways of water reservoirs. The main parameter of such structures is its discharge capacity depending on overfall coefficient, dimensions of spillway, gravitational constant and height of overflowing water jet. The aim of investigation was the mathematical derivation of formula for calculation of overfall coefficient for sharp-crested spillway from observed data. The problem was solved with the aid of statistical method of nonlinear regression analysis, Gauss-Newton algorithm (nonlinear least squares). The objective of investigation was achieved by the design of new equation providing high confidential results.

Keywords: flow rate, nonlinear regression, regression diagnostics, small water reservoir, spillway

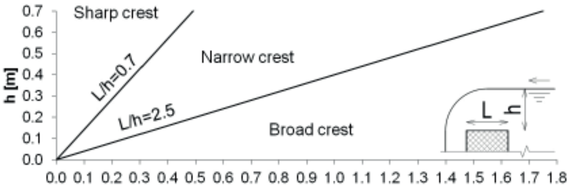
INTRODUCTION

The overfall occurs most often when obstruction of river channel (Kratochvíl, 1991). Water climbs over an obstacle, and after reaching its peak, the spillway edge, it starts to overflow, spill out over the top edge of the wall (e.g. barrage, weir, sluice gate).

The fundamental calculations are derived from water outlet through a gap. The inputs for discharge capacity computation involve the overfall coefficient, dimensions of spillway, gravitational constant, height of overflowing water jet, perhaps even velocity head. Most of these inputs are definite but the determination of overfall coefficient should be ambiguous. The ratio of the height of water on the weir and spillway width is changing in dependence on altering water level behind the damming construction. The most responsible task is represented by determination of overfall type. We are able to distinguish three types of overfall which can occur: broad-crested, narrow-crested and sharp-crested conditions. The depth of the water cascading over the spillway edge – height of overflowing jet (h) and width of the weir (L) are parameters that affect the hydraulic conditions on the spillway and its calculation (Fig. 1).

Further critical parameters such as the shape of the edge of the spillway, length of spillway, its position relative to the axis of flow, lower water level etc (Patočka, 1965). The flow rate determines the energy overflow height and overfall coefficient (Agroskin, Dmitrijev, Pikalov, 1956; Boor and Klopček, 1961). Tab. I summarizes the values of overfall coefficient for broad crested spillways (Jandora and Šulc, 2006).

In the basic equation for the flow rate, the overfall coefficient (μ) is introduced, depending on the type and the height of overflowing water current (h). The overfall coefficient $m = \mu \cdot 2/3$ is usually applied in practical calculations (Krešl, 2001). The series of the empirical formulas were developed for its calculation, e.g. Bazin, Poncelet (Jandora, Stara,



1: *Determination of overfall type*

I: *Overfall coefficient m_b for broad-crested spillway*

Shape of spillway	m_b [–]
Overfall without losses (abstract case)	0.385
Rounded inflow (good conditions)	0.360
Rounded inflow	0.350
Skewed inflow	0.330
Sharp-edged inflow	0.320
Sharp-edged inflow (poor conditions)	0.300

Starý, 2002; Uhmannová, 1999). Resulting values of the coefficient according to the formulas by different authors mostly deviate by 2–3%. The general form of formula for flow rate on rectangular shaped spillway without assuming the velocity head is given by equation (1).

$$Q = mb\sqrt{2g}h^{1.5}, \quad (1)$$

b [m]............length of spillway edge shortened by side contraction,
g [m·s^{-2}] ......gravitational constant,
h [m]............height of overflowing jet,
m [–] ....overfall coefficient,
Q [m^3·s^{-1}] ...flow rate.

The following article deals with the determination of the overfall coefficient (m), especially for the case of water overfall in outlet structure. The outlet, constructed as sluices with the damming function ensured by double wall of timber sluice planks with sealing, is widely used structure for small water reservoirs.

In practice, the determination of overfall coefficient is simplified and is assigned to on the basis of tabular values for the type of overfall. In this relatively frequent calculation of determining can get into the situation that defining the type of overfall and coherent overfall coefficient is ambiguous. The ratio of the height of water on the weir and spillway width is changing during the reservoir drawdown in dependence on overfall jet. Thus the type of spillway changes from sharp-crested to narrow crested (Synková and Zlatuška, 2003) – equation (2).

$$m_n = 0.42\left(0.7 + 0.185\frac{h}{L}\right), \quad (2)$$

m_n [–]...........overfall coefficient for narrow-crested spillway,
L [m] ....width of weir.

MATERIALS AND METHODS

The aim of investigation was the mathematical derivation of formula for calculation of overfall coefficient for sharp-crested spillway. The input data for analysis were represented by empirical tabulated values of overfall coefficient for sharp-crested conditions (Vrána, 1993; Vrána and Beran, 1998), occurring on the outlet structures constructed as sluices with the damming function ensured by double wall of timber sluice planks with sealing (widely used for small water reservoirs in practice). Tab. II summarizes values of overfall coefficient (m_s) for sharp-crested spillway in dependence on height of overflowing water (h).

The problem was solved with the aid of statistical method of nonlinear regression analysis, Gauss-Newton algorithm (nonlinear least squares). The principle of regression lies in the derivation of mathematical equation which describes the relation between dependent and independent variables. The method of least squares is a standard approach to the approximate solution of overdetermined systems, i.e. sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation. When the model function is not linear in the parameters, the sum of squares must be minimized by an iterative procedure (Meloun and Militký, 2004). The regression analysis follows the order listed below:

a) Formulation of the regression model.
b) Determination of model parameters (the best parameter estimates).
c) Determination of statistical significance (whether found model contributes to the specification of dependent variable estimation).
d) Interpretation of results gained by model due to the primary task.

The analysis was performed in algorithmic statistical software QCExpert 3.3 which was primarily developed for purposes of quality control (Kupka, 2012). The program enables interactive data analysis with the aid of modern statistical techniques which facilitate interpretation of results. The modul of nonlinear regression serves for creation and analysis of explicit nonlinear regression models in a general form.

The principle of parameter estimation consists in finding of its optimal values with asymptotic estimates of standard deviation and confidence intervals for a given significance level alpha (α). The estimation target is a function of the independent

II: *Overfall coefficient m_s for sharp-crested spillway*

h [m]	**0.05**	**0.06**	**0.08**	**0.10**	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**	**0.22**
m_s [–]	0.459	0.450	0.439	0.432	0.428	0.424	0.422	0.420	0.419	0.417
h [m]	**0.24**	**0.26**	**0.28**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**	**0.60**	**0.70**
m_s [–]	0.416	0.415	0.415	0.414	0.413	0.412	0.411	0.410	0.410	0.409

variables called the regression function with confidence bands.

The regression diagnostics is solved with the aid of technique called regression triplet (Meloun and Militký, 2002). The analytical method for model construction involves three consequent steps – discussion on data, discussion on model and discussion on method of estimation.

Discussion on Data

The introductory analysis of input data is usually worked out for evaluation its quality for designed model. The process includes the data cleanup and identifying outliers and extremes.

Discussion on Model

The process is focused on evaluation of model quality for given data. We try to guess, mathematically derive and verify the best model in regression analysis. Statistical characteristics of regression represent proper tools for model evaluation.

The multiple correlation coefficient (R) is a measure of the strength of the association between the independent (explanatory) variables and the one dependent (prediction) variable. The determination coefficient (R^2) denotes how much of variance is explained by model. The higher the values are, the better the model.

The mean error of prediction (MEP) represents the prediction error of i-th value of dependent value calculated via regression without i-th point.

$$MEP = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i^2}{(1-H_{ii})^2}, \tag{3}$$

e_i^2 ....square of i-th residua in model,
H_{ii}.................*i*-th diagonal component of H-projection matrix.

The Akaike information criteria (AIC) is a measure of the relative quality of a statistical model for a given set of data. AIC deals with the trade-off between the goodness of fit of the model and the complexity of the model. It is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. AIC does not provide a test of a model in the sense of testing a null hypothesis, i.e. AIC can tell nothing about the quality of the model in an absolute sense. AIC results from residual sum of squares and number of variables.

$$AIC = n \cdot \ln\left(\frac{RSC}{n}\right) + 2m, \tag{4}$$

m ....number of parameters,
RSC ....residual sum of squares.

Both MEP and AIC is very strong indicator of model quality. The lower the values are, the better the model.

Discussion on Estimation Method

The procedures involved in this part of analysis check the fulfillment of all presumptions demanded by the least squares method. The verification of mentioned assumptions is accomplished through the tests of residua. The tests applied in the analysis are represented by the test of heteroscedasticity, test of normality and test of autocorrelation.

Cook-Weisberg test of heteroscedasticity tests the permanency of variance of errors. If the test criteria $S_f < \chi^2_{1-\alpha}(1) = 3.84$, heteroscedasticity does not exist in the data (i.e. any weights do not have to be used).

$$S_f = \frac{\left[\sum_{i=1}^{n}(y_i - y_p)^2 e_i^2\right]^2}{2\sigma^4 \sum_{i=1}^{n}(y_i - y_p)^2}, \tag{5}$$

$$y_p = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{6}$$

e_i ....i-th residua in model,
y_i..................variable in i-th point,
σ^2.................variance,
$\chi^2_{1-\alpha}(1)$ ....chi-squared distribution with one degree of freedom.

The normality of random errors was solved via numerical calculation. Jarque-Bera test proves the normality of distribution of errors with the aid of residua distribution. If the test criteria $L(e) > \chi^2_{1-\alpha}(2) = 5.99$, normality is rejected.

$$L(e) = n\left[\frac{\hat{u}_3^2}{6\hat{u}_2^3} + \frac{\frac{\hat{u}_4}{\hat{u}_2^2} - 3}{24}\right] + n\left[\frac{3\hat{u}_1^2}{2\hat{u}_2} - \frac{\hat{u}_3\hat{u}_1}{\hat{u}_2^2}\right], \tag{7}$$

$$\hat{u}_j = \frac{\sum_{i=1}^{n}\hat{e}_i^j}{n}, \tag{8}$$

e_i ....i-th residua in model,
u_j..................j-th selection moment of residua.

Wald test of autocorrelation tests the presence of autocorrelation on the strength of calculated residua. The most common case is represented by first order autocorrelation $\varepsilon_i = \rho_1\varepsilon_{i-1} + u_i$, where $u_i \sim N(0, \sigma^2)$. For $\rho_1 = 1$ is the case of cumulative errors (ε), for $\rho_1 \leq 1$ is the case of first order autocorrelation (Meloun, 2007). The tested hypothesis is H_0: $\rho_1 = 0$ vs. H_0: $\rho_1 \neq 0$. If the test criteria $W_a < \chi^2_{1-\alpha}(1) = 3.84$, the autocorrelation is not approved in the data.

$$W_a = \frac{n\rho_1^2}{1-\rho_1^2}. \tag{9}$$

Further the graphical methods were used for identification of potential influential points. The Atkinson distance is the modification of Cook distance. Another technique of indication of

III: *Estimation of parameters*

Parameter	Value	Standard error	Hypothesis H_0	Lower CL	Upper CL
p1	–0.01944	0.00027	rejected	–0.02003	–0.01886
p2	0.35305	0.00078	rejected	0.35137	0.35472
p3	0.38541	0.00039	rejected	0.38457	0.38625
p4	0.05245	0.00072	rejected	0.05090	0.05401
p5	0.04849	0.00061	rejected	0.04718	0.04981
p6	0.05270	0.00064	rejected	0.05133	0.05408

influential points is based on diagonal elements of H-projection matrix $H = X(X^TX)^{-1}X^T$ which represent the rate of influence of particular points on regression (X is the matrix of the first partial derivatives of model according to individual parameters in particular values of independent variable). The points situated above the horizontal line are considered as strongly influential in both plots and it is necessary to pay attention to them.

RESULTS AND DISCUSSION

The data are represented by two variables and 20 cases. Independent variable (x) is the height of overflowing water current (h) over spillway. The dependent variable (y) is the value of overfall coefficient (m_s) during sharp-crested conditions (Tab. II). The analysed dataset is founded on empirical measurements and all values were involved for purposes of the first guess of model. The model was designed with parameters p_1, p_2, p_3, p_4, p_5, p_6 – equation (10).

$$y = p_1 \ln(p_2 x) + p_3 x^{p_4} + p_5 \frac{p_6}{x}. \qquad (10)$$

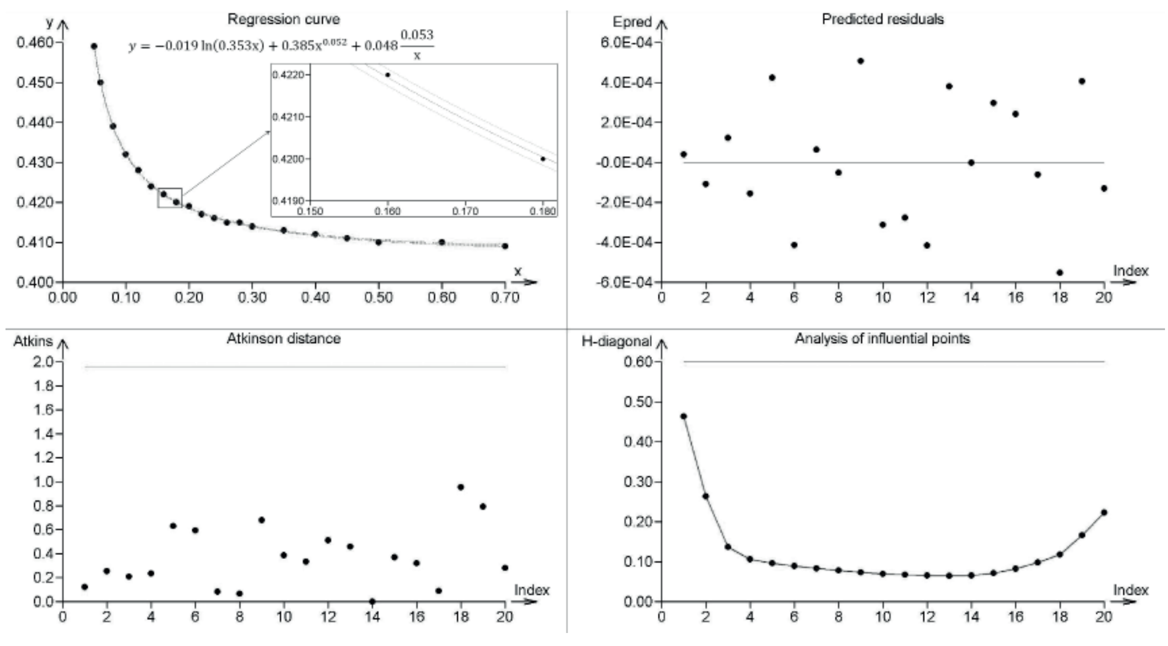
The estimates of particular parameters were realized by means of Gauss-Newton algorithm for solution of nonlinear least squares problems on the significance level α = 0.05 (Tab. III).

The results show that all parameters are statistically significant. The values of all parameters are more than three times higher than their standard errors (Sillén rule). The confidence limits (CL) do not contain 0, so the hypothesis H_0: $p_i = 0$ was rejected in all cases.

The quality of found estimates was rated with the aid of statistical characteristics of regression (Tab. IV). The value of multiple correlation coefficient R shows the designed model is statistically significant. The high value of coefficient of determination R^2

IV: *Statistical characteristics of regression*

Criteria	Value
R	0.99979
R^2	0.99959
MEP	8.98E-08
AIC	–316.493
s(e)	0.00032
RSC	1.47E-06



2: *Regression model and its quality (QCExpert)*

V: *Regression diagnostics – tests of residua*

Cook-Weisberg test of heteroscedasticity	
Value of S_f criteria	0.60007
$\chi^2(1-\alpha, 1)$	3.84146
Result	homoscedasticity
Jarque-Bera test of normality	
Value of L(e) criteria	0.79816
$\chi^2(1-\alpha, 2)$	5.99146
Result	normal distribution
Wald test of autocorrelation	
Value of W_a criteria	1.29402
$\chi^2(1-\alpha, 1)$	3.84146
Result	insignificant autocorrelation

means that 99.96% of points correspond with model. The values of the mean error of prediction MEP and Akaike information criteria AIC are very low. They could be used for a decision of quality among other potential models. Estimation of standard error of residua s(e) and residual sum of squares are represented by very low values as well.

However the described characteristics confirm good quality of regression model, the regression diagnostics were carried out. The diagnostics contain some procedures for interactive data analysis in the method of regression triplet. Fig. 2 depicts graphical results of regression analysis.

The plot of regression facilitates the visual control of how well the data define the best-fit curve. Working-Hotelling 1 – α confidence bands represent the range in which the values of dependent variable are situated with probability of 95%. This includes the uncertainty in the true position of the curve (enclosed by the confidence bands). The range is very narrow indicating very good fit and relevancy of designed model. The plot also contains the formulation of regression function.

The visual analysis of residua was performed via plot of predicted residua. The points are equally distributed about horizontal line crossing zero and the shape of cloud indicates trouble-free model.

The quality of data was reviewed through the plot of Atkinson distance and analysis of influential points. Both plots show that any influential (golden) points are not present.

Numerical tests of residua for verification of the presumptions of least squares method were accomplished as the final part of regression diagnostics (Tab. V). The final values of test criteria are lower that incident quantiles of chi-squared distribution. Hence, the homoscedasticity of residua was proved via Cook-Weisberg test, Jarque-Bera test approved normal distribution of residua and Wald test demonstrated the insignificant rate of autocorrelation.

The statistical analysis based on regression triplet showed that the designed relationship can be considered as suitable model for given dataset. After the substitution of variables (with rounded parameter estimates), the resulting equation for calculation of overfall coefficient in sharp-crested conditions (m_s) in dependence on height of overflowing water current (h) has the form as follows:

$$m_s = -0.019\ln(0.353h) + 0.385h^{0.052} + 0.048\frac{0.053}{h}. \quad (11)$$

The objective of analysis was achieved by the design of new formula which determines the overfall coefficient in any value of water height with accuracy of ±0.0002 (value specified from residua statistics). The gains of the results subsist in facilitation of consequent calculations of flow rate through spillways without the need of inaccurate linear interpolation of overfall coefficient from Tab. II.

CONCLUSION

The problem of determination of overfall coefficient is closely related to the construction of spillway. The ratio of the height of water on the weir and spillway width is changing in dependence on altering water level. We know three types of overfall: broad-crested, narrow-crested and sharp-crested conditions. The investigation dealt with the case of outlet hydrotechnic structure widely used on small water reservoirs for which only empirical values of overfall coefficient for sharp-crested conditions exist. The objective of investigation was the mathematical derivation of a formula fit into measured data.

The problem was solved with the aid of statistical method of nonlinear regression analysis, Gauss-Newton algorithm (nonlinear least squares) in algorithmic statistical software QCExpert 3.3. The methodology of analysis was based on regression triplet using the regression diagnostics which helps to rate the quality of regression model. The analytical method of regression triplet in model construction involves three consequent steps – discussion on data, discussion on model and discussion on method of estimation. Regression diagnostics helped to rate the quality of designed regression model with the aid of residua analysis. Both numerical and graphical model characteristics and tests were applied.

The statistical analysis based on regression triplet showed that the designed relationship can be considered as suitable model for given dataset. The objective of investigation was achieved by the derivation of new formula by substitution of variables which determines the overfall coefficient for sharp-crested conditions in any value of water height with high accuracy. The gains of the results subsist in facilitation of consequent calculations of flow rate through spillways without the need of inaccurate linear interpolation of overfall coefficient.

Acknowledgement

The paper was worked out as a part of research project "Active anti-abrasion structures", reg. no. LDF_VT_2015011, funded by IGA FFWT MENDELU Brno.

REFERENCES

AGROSKIN, I. I., DMITRIJEV, G. T., PIKALOV, F. I. 1956. *Hydraulics II* [in Czech: *Hydraulika II*]. Praha: Státní nakladatelství technické literatury.

BOOR, B., KLOPČEK, A. 1961. *Fundamentals of hydraulics in practice II* [in Czech: *Základy hydrauliky II*]. Bratislava: SVŠT v Bratislave, Slovenske vydavatelstvo technickej literatury.

JANDORA, J., STARA, V., STARÝ, M. 2002. *Hydraulics and hydrology* [in Czech: *Hydraulika a hydrologie*]. Brno: Akademické nakladatelství CERM.

JANDORA, J., ŠULC, J. 2006. *Hydraulics* [in Czech: *Hydraulika*]. Brno: Brno University of Technology.

JANDORA, J., UHMANNOVÁ, H. 1999. *Fundamentals of hydraulics and hydrology – practice* [in Czech: *Základy hydrauliky a hydrologie – cvičení*]. Brno: Akademické nakladatelství CERM.

KRATOCHVÍL, J. 1991. *Hydraulics* [in Czech: *Hydraulika*]. Brno: Brno University of Technology.

KREŠL, J. 2001. *Hydraulics* [in Czech: *Hydraulika*]. 1. ed. Brno: Mendel University in Brno.

KUPKA, K. 2012. *QC.Expert: Interactive statistical data analysis* [in Czech: *QC.Expert: Interaktivní analýza statistických dat*]. Pardubice: TriloByte Statistical Software.

MELOUN, M. 2007. Computer-assisted statistical data analysis – 5. Build-up of regression model and its testing using diagnostic of regression triplet [in Czech: Statistické zpracování vodohospodářských dat – 5. Výstavba regresního modelu diagnostikami regresního tripletu]. *Water management*, 57: 224–230. Available at: http://meloun.upce.cz/en/publishing/original-papers.

MELOUN, M., MILITKÝ, J. 2002. *Compendium of Statistical Analysis of Experimental Data* [in Czech: *Kompendium statistické analýzy experimentálních dat*]. 1. ed. Praha: Academia, The Academy of Sciences of the Czech Republic.

MELOUN, M., MILITKÝ, J. 2004. *Statistical Analysis of Experimental Data* [in Czech: *Statistická analýza experimentálních dat*]. Praha: Academia, The Academy of Sciences of the Czech Republic.

PATOČKA, C. 1965. *Hydraulics* [in Czech: *Hydraulika*]. Praha: Státní nakladatelství technické literatury.

SYNKOVÁ, J., ZLATUŠKA, K. 2003. *Small Water Reservoirs – practice* [in Czech: *Malé vodní nádrže – cvičení*]. 1. ed. Brno: Mendel University in Brno.

VRÁNA, K. 1993. *Fishponds and Special Purpose Reservoirs – practice* [in Czech: *Rybníky a účelové nádrže – cvičení*]. Praha: Czech Technical University in Prague.

VRÁNA, K., BERAN, J. 1998. *Fishponds and Special Purpose Reservoirs* [in Czech: *Rybníky a účelové nádrže*]. Praha: Czech Technical University in Prague.

Contact information

Petr Pelikán: pelikanp@seznam.cz
Jana Marková: jana.markova@mendelu.cz