Suciu, G., Bălănescu, M., Bălăceanu, C.M., Ușurelu, T., Suciu, V., Vasilescu, A., Pasat, A., Istrate, C., Anwar, M. (2018), Predictive model for danube water level using statistical data analysis pp. 78-85. In Gastescu, P., Bretcan, P. (edit, 2018), *Water resources and wetlands*, 4th International Conference Water resources and wetlands, 5-9 September 2018, Tulcea (Romania), p.312

Available online at http://www.limnology.ro/wrw2018/proceedings.html

4th International Conference Water resources and wetlands, 5-9 September 2018, Tulcea (Romania)

A PREDICTIVE MODEL FOR DANUBE WATER LEVEL USING STATISTICAL DATA ANALYSIS

George SUCIU, Mihaela BĂLĂNESCU, Cristina Mihaela BĂLĂCEANU, Teodora UȘURELU, Victor SUCIU, Andrei VASILESCU, Adrian PASAT, Cristiana ISTRATE, Muneeb ANWAR

BEIA Consult International SRL, 16-22 Peroni St., Bucharest, 041386, Romania, + 40 21 332 3005, Email: george@beia.ro, mihaela.balanescu@beia.ro, cristina.balaceanu@beia.ro, teodora.usurelu@beia.ro, victor.suciu@beia.ro, andrei.vasilescu@beia.ro, adrian.pasat@beia.ro, cristiana.istrate@beia.ro, ma@beia.ro,

Abstract

Water have impact on many sectors and is also a resource used in competing domains among agriculture, energy, conservation and human settlements. Given the estimated impact of the climate change, an increase of the vulnerabilities to water-related hazards (i.e.floods) is expected. In the EU around 216 000 people are estimated to be exposed to river flooding and the flood damage could amounting to € 5.3 billion each year. In this context a predictive model for water level of the main european rivers are important to be realized. Beia Consult International has installed and is further developing an automatic system able to continuously monitor the level and water temperature along the Danube and some of its tributary rivers. Till now, between the 21 monitoring points where the telemetry equipments was installed are: Turnu Severin, Gruia, Calafat, Bechet, Corabia, Oltenița, Chiciu, Izvoarele and Unirea. Each remote monitoring installation consists of a Remote Telemetry Unit (RTU), a water level and temperature sensor which is connected to the RTU through an atmospheric pressure relief box and the solar panel that powers the RTU and the sensor. Measurements are displayed as a table and graph. Users can access the platform anywhere there is an internet connection. The implemented system measures the water level and temperature every 3 minutes and calculates the average of 5 measurements every 15 minutes. A predictive model for water level was developed based on measurements for period 2015-2017 from 3 monitoring station (Gruia, Oltenița and Izvoarele). For each measurement point a descriptive statistical analysis was performed to identify the main characteristics of the data series. After completing the missing data (by interpolation method) these was analysed using time series analysis method for developing the model. The model was tested using a validation dataset to ensure the accuracy and efficiency.

Keywords: water level, sensors measurement, telemetry, predictive model

1. INTRODUCTION

Over the years, concrete action has been taken to increase capacity to react, particularly regarding floods or extreme weather phenomena. In 2005, a national flood risk management strategy was set up which included responsibilities and methods of flood prevention and intervention. Over time, climate change has led to abnormal phenomena: winters have become warmer and shorter, which has led to declining seasonal snow and extreme summer temperatures have led to a drop-in water resources and a rising demand for water. Thus, predictive model for water level of the main European rivers are important to be realized (Burnete et al, 2017). The flood prediction usually consists of water level rise forecasting and flood territory mapping. The flood models are categorized in three types of models. The first type - physical model - consist in representation of a scaled copy of a real physical system. The second one is a mathematical model based on mathematical logic and equations, while the last one is based on machine-learnings techniques and are data-driven (K. A. a. C. Morley, 2002). The ANFIS model was used for the simulation and forecasting of floods in the Sieve basin in Italy (Gautam et al, 2001). A fuzzy logic approach was used for clustering the data in the hydrological basin of "Padule di Fucecchio" basin in Italy. This method assures a good prediction of extreme and rare events (Lucheta et al, 2003). The evaluation and the workability of a nonlinear system for prediction of rainfall have performed (Kishtawal et al, 2003). The decision forest regression makes forecasts by using a sequence of base models and combining it. From many decision trees that act with a pure data full decision tree stronger than each tree but by consolidating them is make a better overall achievement (Criminisi et al, 2013). The determination tree machine learning algorithm was used for prediction of flood areas in Kelantan, Malaysia (Tehrany et al, 2013). The problem of proliferation of flood cases with the low appearance of these extreme events can be determined with several approaches. For some circumstances it is possible to construct a counterfeit flood-causing cyclone model, that permits to exercise and test the machine learning-based solution with the broad range of genuine objects, and for cases with an intricate mechanism of flood condition, some similar events with lower scale can be detected in open data (Nikitin et al, 2016. Observed that the neural networks (NN) models advanced in the study were ready to forecast the water levels of Lake Naivasha for four back-to-back months occurring after a given month and given data for six consecutive month's antecedents to that month (Chuku, 2016). Thus, NN renders an efficient and suitable method for forecasting water levels in the lake. This can serve in the water-use formulation and scheduling for domestic, community and agricultural uses. Timely forecasting can also help in flood monitoring, recognition, and restraint in areas prone to floods (Samuel et al, 2017; Feilat et al, 2017). The most important prediction factors for the urban pluvial flooding identified in Noymanee work are water level measured at stations close to the predicted one and historic water level in that station (Noymanee et al, 2017).

2. METHODS

Beia Consult International has installed a monitoring system (Vasilescu et al, 2016) for Danube waters and its affluent rivers using the equipment provided by Adcon Telemetry (Adcon, 2018). Until now, 21 monitoring stations have been installed, nine of them are located on the Danube. The nine locations (figure 1) along the Danube where telemetry equipment was installed are: Turnu Severin, Gruia, Calafat, Bechet, Corabia, Oltenița, Chiciu, Izvoarele and Unirea (Vasilescu et al, 2016). The water monitoring system consists of the A753 remote monitoring unit and the SDI-12 temperature sensor, the OTT-RLS radar level sensor (OTT, 2018), the A850 data storage unit and the addVANTAGE Pro platform, where all the measured data can be accessed and downloaded (Vasilescu et al, 2016). Communication between system components and users is done over the Internet, so that everything is easily reconfigurable and transferable.

Figure 1. Locations of the remote monitoring stations

The initially data sets extracted from platform were represented by date (day, hour, minute) and a water level values registered at every 15 minutes for each measurement point included in the analysis (Gruia, Oltenița and Izvoarele) for the period 2015-2017. The extracted data were preliminary analyzed from the qualitative point of view with the aim to remove the and obtain proper datasets. These datasets were preliminary analyzed using statistical descriptive methods (i.e. mean value, variation, standard deviation, quartile calculation), histograms, boxplot representation and correlation coefficients. The dataset used for the model development were represented by joined homogenized datasets for the years 2015 and 2016, while the dataset used for model testing are the one from 2017. In section 4 are presented the results of this analysis and the predictive model developed. The predictive model obtained using a multiple linear regression method were tested for validation and the results are presented in the section 5.

4. RESULTS AND DISSCUSIONS
4.1 Model development

The initial datasets are represented by 9 datasets that includes the values for water level in the points of analysis (Gruia, Oltenita and Izvoarele) for a 3 years period (2015-2017). Each dataset has 35042 registration who was subject of further processing. Thus, the negative value (corresponding to the situation when water level are under the corrected standard values of the National Administration of Romanian Waters) was removed, the hourly average values were calculated, and the datasets were homogenized in order to have values for all the measuring points in the same date and time. All 9 final dataset has 5342 values each. In the figure 2 are presented the yearly variation of water level in all the 3 points included in the analysis and the multiannual variation to have a graphical representation of the data analyzed.

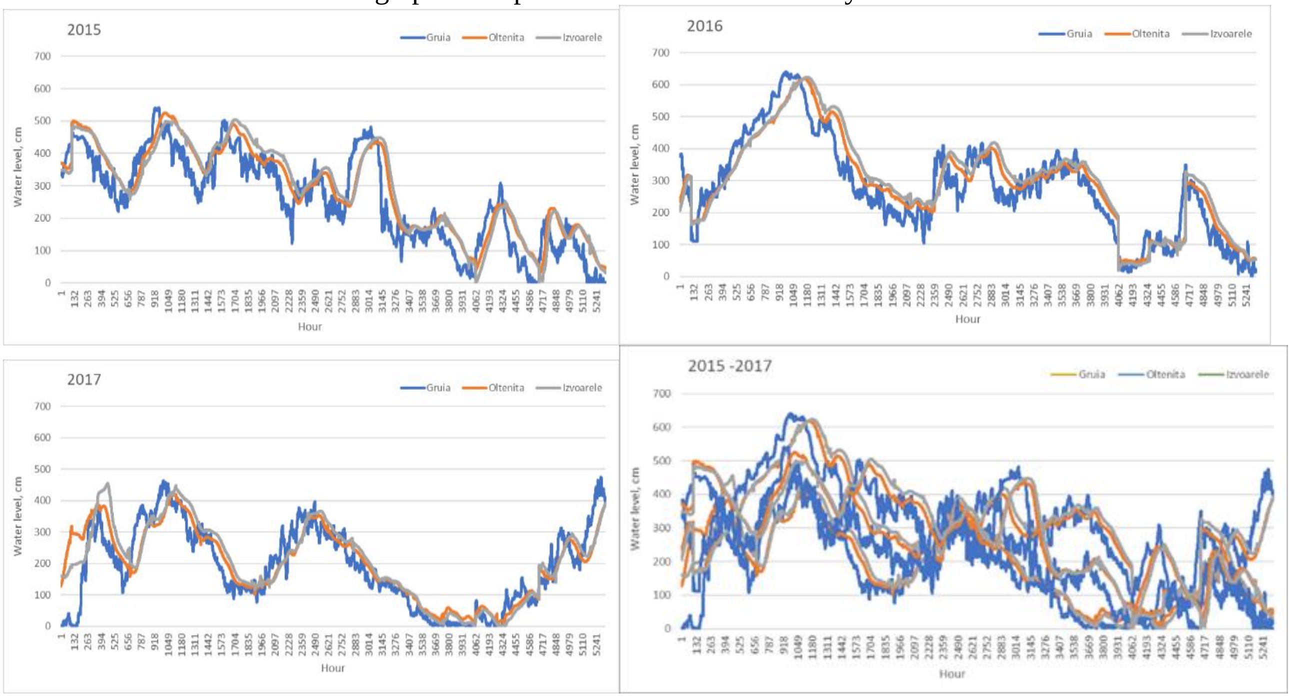

Figure 2. Danube water level variation in the selected measurement points for the period 2015-2017

Table 1 – Statistical parameter for the datasets used in the model development

Statistical parameter	2015			2016			2017		
	Gruia	Oltenita	Izvoarele	Gruia	Oltenita	Izvoarele	Gruia	Oltenita	Izvoarele
Mean	259.11	290.80	293.56	277.64	291.96	304.19	188.32	201.13	205.35
Standard Error	1.94	1.79	1.86	2.04	1.91	1.97	1.73	1.51	1.64
Median	265	299	311	269	292	309	184	206	204
Standard Deviation	142.10	131.31	136.29	149.16	139.74	144.22	126.60	110.48	120.41
Sample Variance	20193.87	17243.33	18576.07	22250.22	19528.15	20799.71	16029.26	12207.12	14500.44
Kurtosis	-1.10	-1.14	-1.13	-0.27	-0.28	-0.34	-1.05	-1.06	-0.93
Skewness	-0.10	-0.10	-0.24	0.35	0.21	0.07	0.14	-0.003	0.013
Range	540	502	505	640	574	604	475	410	456
Minimum	1	23	0	1	44	20	0	9	0
Maximum	541	525	505	641	618	624	475	419	456
Confidence Level for mean value (95.0%)	3.81	3.52	3.65	4.00	3.74	3.86	3.39	2.96	3.23

For the development of the predictive model for Danube water level a preliminary statistical analysis of the datasets was performed. The results, presented in the table 1, show very close value of the statistical parameter between the measurement points (Gruia, Oltenita, Izvoarele) in each year. Also, the values registered in Gruia show a higher variability compared with the other points in every year. In figure 3 are presented histograms for the water level values for each point of measurement and every year included in the analysis. The boxplot representation of the data (figure 4) show a similar variation and the very high correlation coefficients (figure 5) is a strong indication that a multilinear model will be a good fitting for the process.

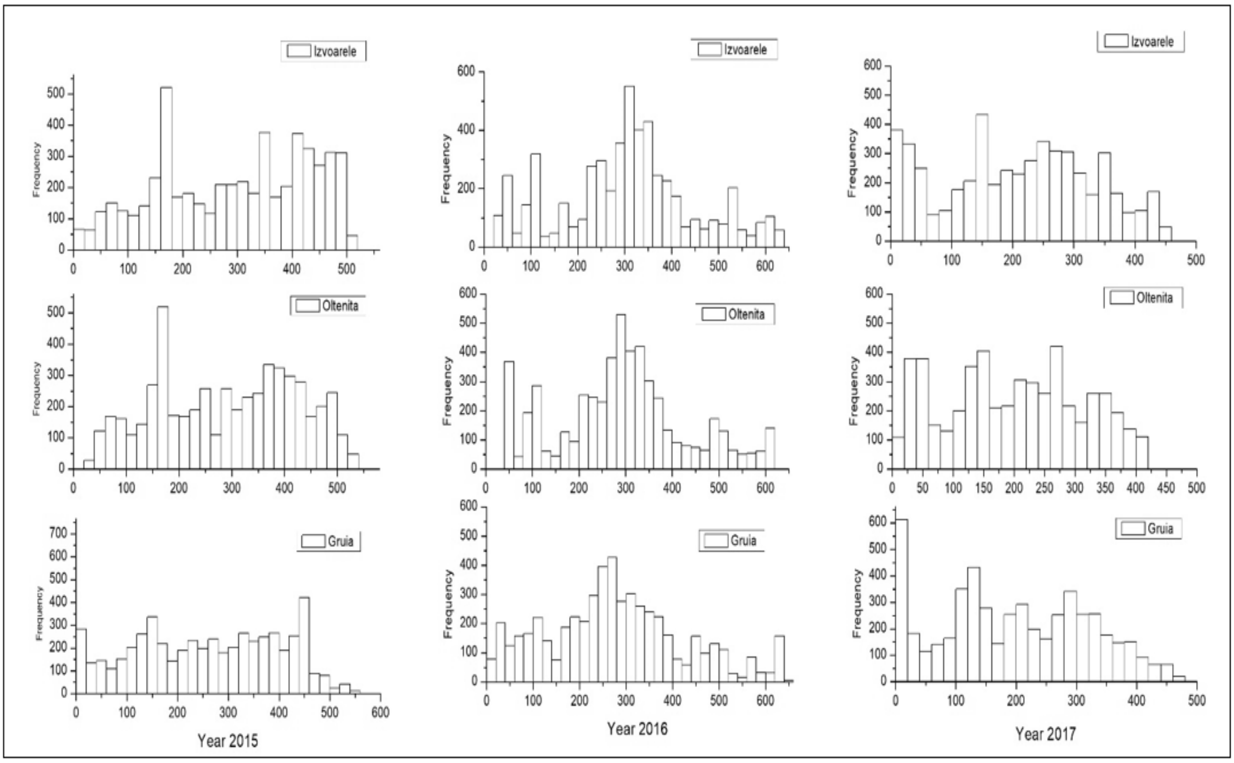

Figure 3. Histograms of values for Danube level in the selected points for the period 2015-2017

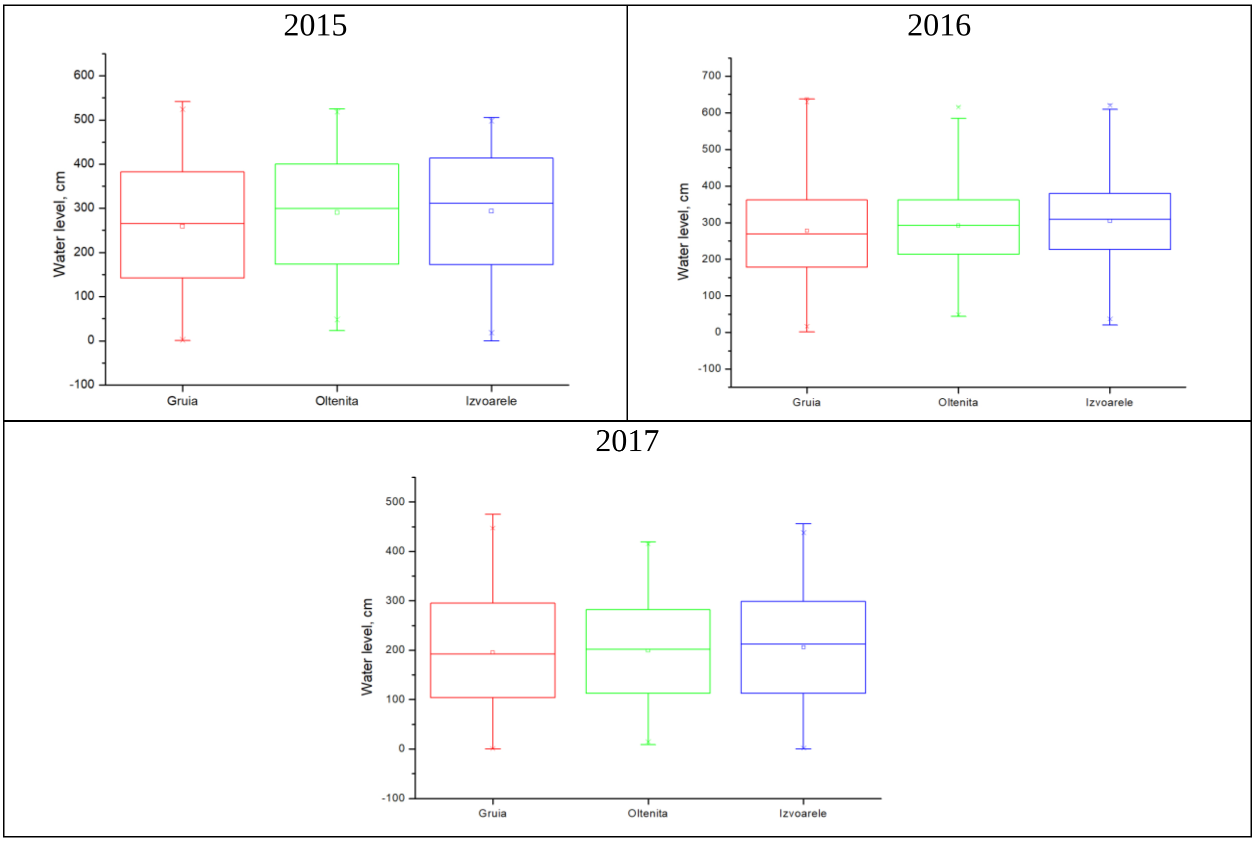

Figure 4. Boxplots of values for Danube water level in the selected points for the period 2015-2017

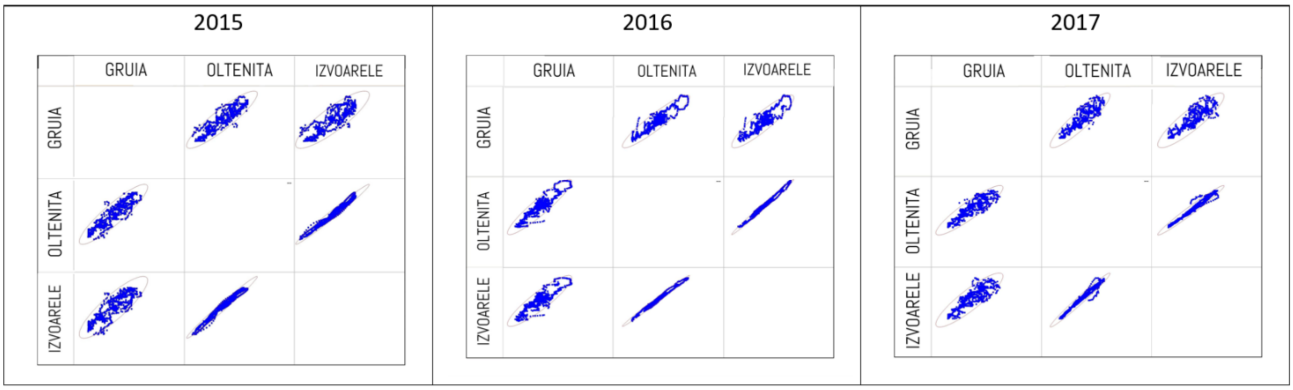

Figure 5. Correlation matrix for datasets used

For the development of the predictive model was used the combined datasets for the years 2015 and 2016 for the all 3 measurement points. Figure 6 shows the statistical characteristics of the 2015-2016 dataset.

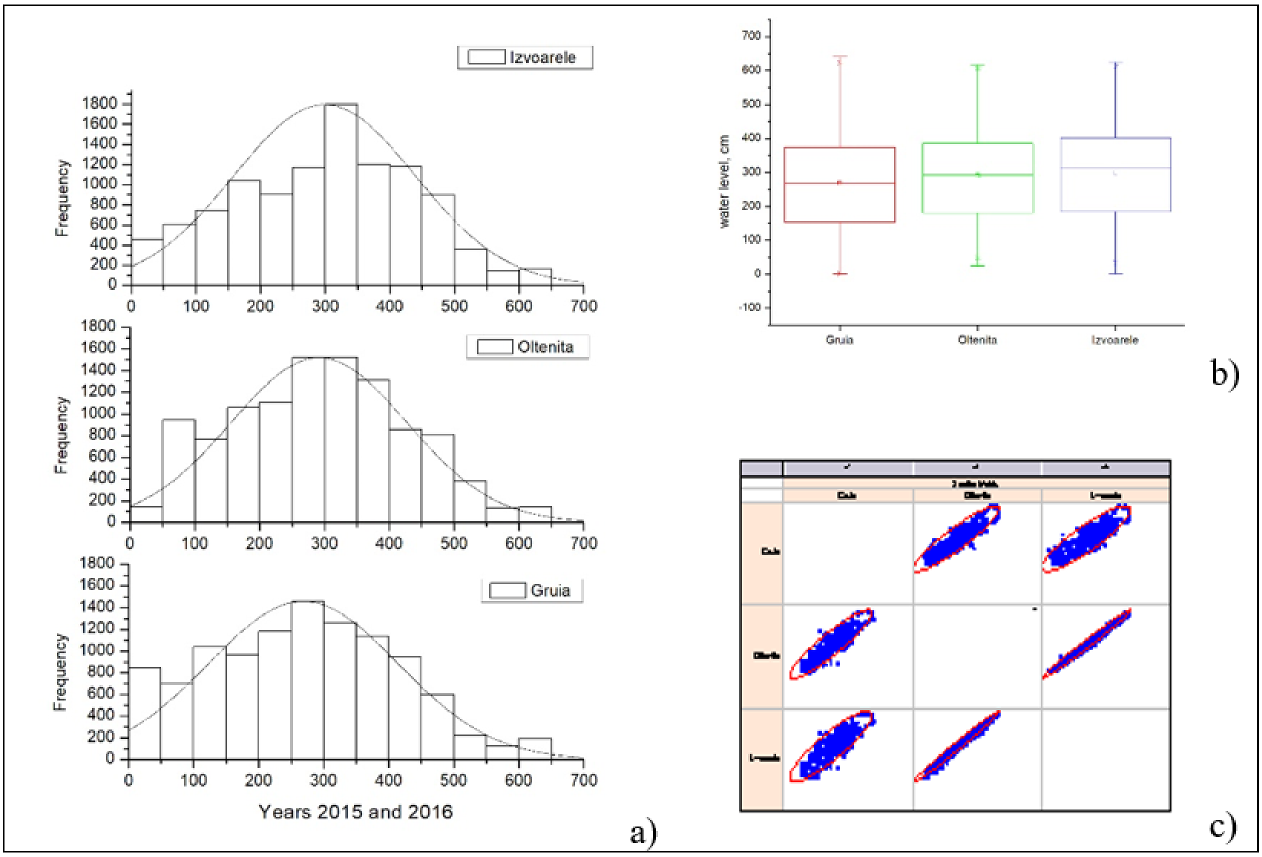

Figure 6. Statistical characteristics of the data set used for the predictive model development

As in the preliminary analysis of each location dataset the correlation between values registered in Oltenița and Izvoarele are very strong (0.99) while between Gruia and Izvoarele are slightly lower (0.88). The equation obtained for the predictive model for estimation of water level in Izvoarele based on multilinear regression method is:

$$z = a \cdot x + b \cdot y + c \tag{1}$$

Where:
x – water level at Oltenița measurement point, cm
y – water level at Gruia measurement point, cm
z – water level at Izvoarele measurement point, cm

The adjusted R-square coefficient is 0.98474 which indicate that the regression prediction approximates the real data points very well. In the Table 2 are presented the values and the standard error obtained for the equation coefficients, while in the Table 33 are show the ANOVA results.

Table 2. Values and standard error for the equation coefficients

Parameter	Value	Standard Error
a	1.1544	0.00311
b	-0.12993	0.00289
c	-2.62354	0.40186

Table 3. ANOVA results for the predictive model

	Degree of Freedom	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	2.07E+08	1.04E+08	344406.6	0
Error	10672	3.21E+06	300.76	-	-
Total	10674	2.10E+08	-	-	-

4.2. Model validation

The model developed was tested for validation on the 2017 dataset. The input in the model were considered the hourly water level measured in Gruia and Oltenița and output are the predicted value for water level in Izvoarele. The predicted values were compared with the measured values in the 2017. As the results show important percentage deviation when in Izvoarele are registered low water level (< 50 cm) and in the beginning of the year (when data measured are affected by the extreme low temperature that generate frozen surface, ice stones), the model was tested on an adjusted dataset (4123 values). This dataset is obtained after removing the values corresponding to the situation presented above and represent 80% of the initial dataset.

The results show that the predictive model provide very good estimation of the registered values. In Figure 7 are presented graphical representation of the measured and predicted values (figure 7.a), the boxplot (figure 7.b), the relative percent error (figure 7.c) and his distribution (figure 7.d).

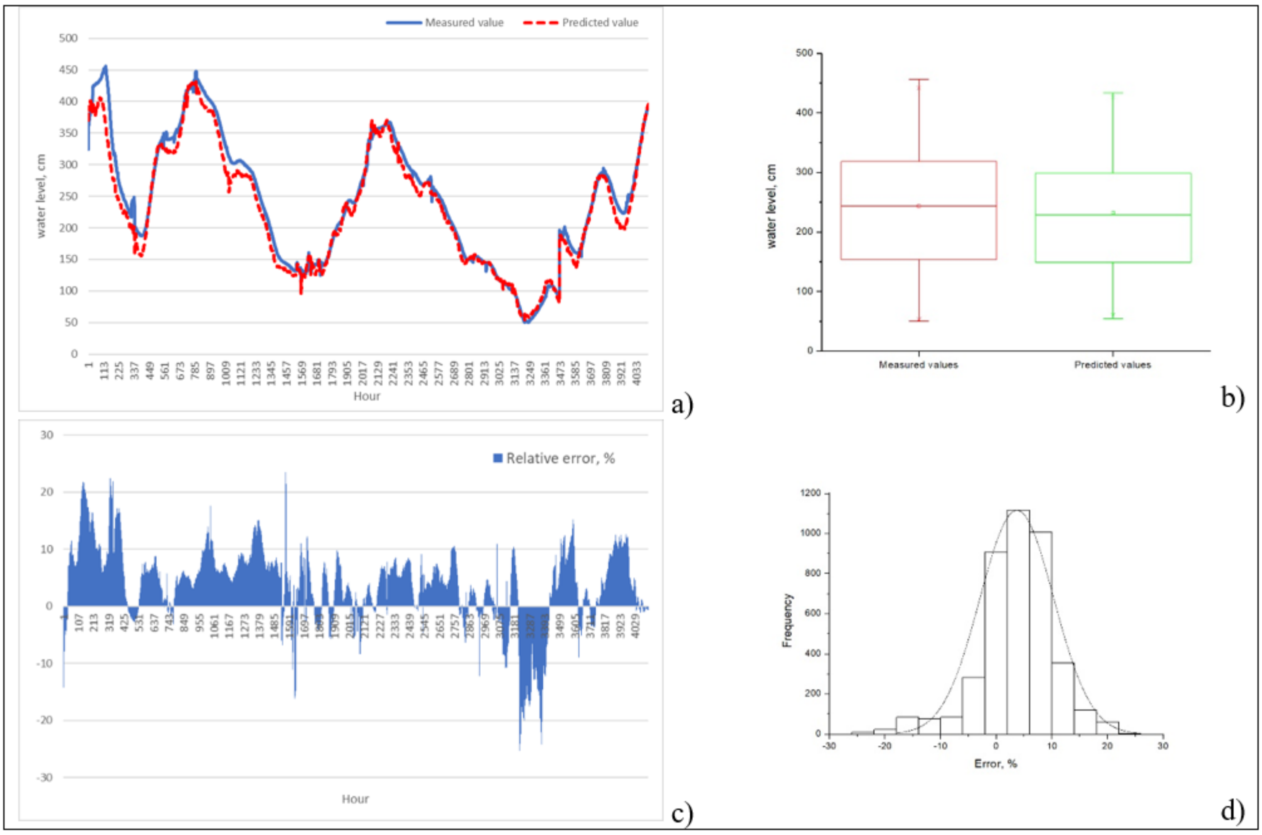

Figure 7. Results of model testing on the adjusted 2017 dataset

The statistical values of the measured and predicted values are very similar (Table 4).

Table 4. Statistical parameters of the measured and predicted water level values

Parameter	Mean value	Standard Deviation	Lower 95% CI of Mean	Upper 95% CI of Mean	Minimum value	1st Quartile (Q1)	3rd Quartile (Q3)	Maximum value	Interquartile Range (Q3 - Q1)	P90	P95
Measured values	243.47	100.07	240.42	246.53	50	154	318	456	164	381	421
Predicted values	232.29	94.19	229.42	235.17	54.53	148.74	298.57	433.12	149.82	368.10	393.31

While the predictive model slightly overestimates (less than 10% of the measured value) the water level for the 79% of the data, in 46% of the data measured the relative percent error are between -5% and +5%. The model underestimates the measured water level with more than 20% in 0.4% of the total measured data and happens when the water level is under 100 cm.

5. CONCLUSIONS

The work presented in the paper show that a multilinear model is a good choice for prediction of Danube water level in the restricted intervals of values. The predictive model developed offer very accurate prediction for the water level in the Izvoarele measurement point based on measurement from Gruia and Oltenita in over 80% of time.

This study represents the first stage in the attempt to realize a predictive model for the Danube waters that covers all the specific situations (including flooding). Further work needs to be performed to develop models for situations when low water level is registered and by including more measurement points (including from some affluent rivers) in analysis. For these cases, other types of models would need to be developed (including by using non-linear mathematical models or machine learning techniques).

ACKNOWLEDGMENT

This work has been supported in part by Executive Unit for Financing Higher Education, Research, Development and Innovation Romania through projects CitiSim (Eureka program, PN-III-P3-3.5-EUK-2016-0043) and Wins@HI (Eureka program, Celtic-Plus cluster, C2015/2-2, PN-III-P3-3,5-EUK-2017-02-0038) and Minister of Research and Innovation Romania through project SmartAgro (contract no. 8592 / 2018). Also, was funded in part by European Union's Horizon 2020 research and innovation program under grant agreements No. 777996 (SealedGRID project) and No. 787002 (SAFECARE project).

REFERENCES

Burnete, C., Tomescu, C., Popescu, A., Samoilă, Dobrin, M., Rusu, V., Popescu, G., Falup, O., Bunda, I., Bălăceanu, C., M., Bălănescu, M., Dascălu, E. (2017). Romania's 7th National Communication under the United Nations Framework Convention on climate change, UNFCCC, Bonn, Germany.

Gautam, K. D., Holz, K. P., (2001). Rainfall-runoff modelling using adaptive neuro-fuzzy systems. Journal of Hydroinformatics, 3(1), pp. 3-10, ISSN 1464-7141, DOI:/10.2166/hydro.2001.0002.

Luchetta, A., Manetti, S., (2003). A real time hydrological forecasting system using a fuzzy clustering approach. Computers & Geosciences, Vol. 29 N° 9, pp. 1111-1117, DOI:10.1016/S0098-3004(03)00137-7.

Kishtawal, M. C., Basu, S., Patadia, F., Thapliyal, K. P (2003). Forecasting summer rainfall over India using Genetic Algorithm. Geophysical Research Letters, Vol. 30, N° 23, 2203.

Criminisi, A., Shotton, J. (2013). Decision Forests for Computer Vision and Medical Image Analysis. Springer Publishing Company, Incorporated. pp.95-107, ISSN 2191-6586, DOI:10.1007/978-1-4471-4929-3.

Tehrany, M. S., Pradhan, B., Jebur, M. N., (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. Journal of Hydrology, Volume 504, pp. 69-79.

Nikitin, N., Spirin, D., Visheratin, A., Kalyuzhnaya, A. (2016). Statistics-based models of flood-causing cyclones for the Baltic Sea region. Procedia Computer Science, pp. Vol. 101, pp. 272–28, DOI:10.1016/j.procs.2017.11.187.

Chuku, C., Oduor, J., Simpasa, A., (2016). Intelligent forecasting of economic growth for African economies: Artificial neural networks versus time series and structural econometric models.

Samuel, I., A., Emmanuel, A., Odigwe, I., A., Felly-Njoku, F., C., (2017). A comparative study of regression analysis and artificial neural network methods for medium-term load forecasting. Indian Journal of Science and Technology. Vol. 10, ISSN 0974-6846.

Feilat, E., A., Talal, Al-Sha'abi, Momani, M., A., (2017). Long-term load forecasting using neural network approach for Jordan's power system. pp. 43-50.

Adcon Telemetry GmbH, Vienna-Klosterneuburg, Austria, website: www.adcon.com.

BEIA Consult International SRL, Telemetry & Telemonitoring, www.beia-telemetrie.ro.

OTT Hydromet GmbH, Kempten, Germany, www.ott.com.

Vasilescu, A., Suciu, G. and Suciu, V. (2016). Monitoring the Danube with equipment from Adcon Telemetry, an OTT Hydromet Business Unit distributed by BEIA Consult International. In Gastescu P., Bretcan P. (edit) Conference proceedings of 3rd International Conference on "Water resources and wetlands", pp. 301-308.

Morley K. A. a. C., (2002). The origin and evolution of complex transfer zones (graben shifts) in conjugate fault systems around the Funan Field, Pattani Basin, Gulf of Thailand," Journal of Structural Geology, pp. 435-449

Noymanee, J., Nikitin, N.O., Kalyuzhnaya, A.V., (2017). Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin, Procedia Computer Science, Volume 119, 2017, Pages 288-297, DOI: 10.1016/j.procs.2017.11.187