

Data-driven discharge analysis: a case study for the Wernersbach catchment, Germany

Eklavya Popat 

Technische Universität Dresden, Institute of Hydrology and Meteorology, Dresden, Germany
Goethe University, Institute of Physical Geography, Frankfurt am Main, Germany, e-mail: Popat@em.uni-frankfurt.de

Alexey Kuleshov, Rico Kronenberg, Christian Bernhofer

Technische Universität Dresden, Institute of Hydrology and Meteorology, Dresden, Germany

DOI: 10.26491/mhwm/112284

54

ABSTRACT. This study focuses on precipitation-discharge data-driven models, with regression analysis between the weighted maximum rainfall and maximum discharge of flood events. It is also the first of its kind investigation for the Wernersbach catchment, which incorporates data-driven models in order to evaluate the suitability of the model in simulating the discharge from the catchment and provide good insights for future studies. The input parameters are hydrological and climate data collected from 2001 to 2009, including precipitation, rainfall-runoff and soil moisture. The statistical regression and artificial neural network models used are based on a data-driven multiple linear regression technique, and the same input parameters are applied for validation and calibration. The artificial neural network model has one hidden layer with a sigmoidal activation function and uses a linear activation function in the output layer. The artificial neural network is observed to model 0.7% and 0.5% of values, with and without extreme values respectively. With less than 1% error, the artificial neural network is observed to predict extreme events better compared to the conventional statistical regression model and is also better suited to the tasks of rainfall-runoff and flood forecasting. It is presumed that in the future this study's conclusions would form the basis for more complex and detailed studies for the same catchment area.

KEYWORDS: Artificial Neural Networks, data-driven modelling, event-based coefficient of rainfall-runoff, precipitation, multi-correlation analysis, soil moisture content.

SUBMITTED: 3 July 2018 | **REVISED:** 21 March 2019 | **ACCEPTED:** 11 September 2019

1. INTRODUCTION

The transformation of rain into a runoff is a complex process that is difficult to fully understand (Hsu et al. 1995; Humphrey et al. 2016). The challenges faced are mainly linked to the non-stationary features of the phenomenon (e.g. trends, seasonality, and jumps) and a highly non-linear relationship between discharge and its driving variables (Cannas et al. 2006; Nourani et al. 2011). Hence, a reliable model for the simulation and prediction of the rainfall-runoff process is in demand, providing important information for integrated water resource management and planning.

With recent developments in computational intelligence, there has been a rapid expansion in the capabilities of empirical modelling, and in particular, in data-driven modelling (DDM). DDM is a fundamental analysis technique that uses the data characteristics of a system and requires less input as parameters compared to other models (Solomatine et al. 2008). As an empirical model, it applies mathematical equations in the analysis of concurrent input and output time series, for example, linear and multi-linear regressions (Clarke 1994).

The most popular computational techniques include artificial neural network (ANN) models, fuzzy rule-based systems (FRBSs), genetic algorithms (GAs), and approaches to model integration. ANN have been successfully applied in modelling rainfall-runoff processes e.g. Hsu et al. (1995), Minns and Hall (1996), Dawson and Wilby (1998), Dibike et al. (1999), Abrahart and See (2000), Abrahart et al. (2008). More specifically, Solomatine and Avila Torres (1996) replicated the behaviour of river basin hydrodynamic/hydrological models and proposed the optimal control of a reservoir, while Bhattacharya and Solomatine (2003) and Sudheer and Jain (2003) modelled stage-discharge relationships.

FRBSs have been successfully applied for drought assessment (Pesti et al. 1996), the prediction of precipitation events (Abebe et al. 2000b), the analysis of groundwater model uncertainty (Abebe et al. 2000a), the control of water levels in polder areas (Lobrecht, Solomatine 1999) and the modelling rainfall-discharge dynamics (Vernieuwe et al. 2005). Moreover, GAs have been used to optimise DDM techniques, such as neural networks (Yao, Liu 1997). Khu et al. (2001) applied genetic programming to real-time runoff forecasting for a catchment in France, while Giustolisi and Savic (2006) used evolutionary regression for groundwater and river temperature modelling.

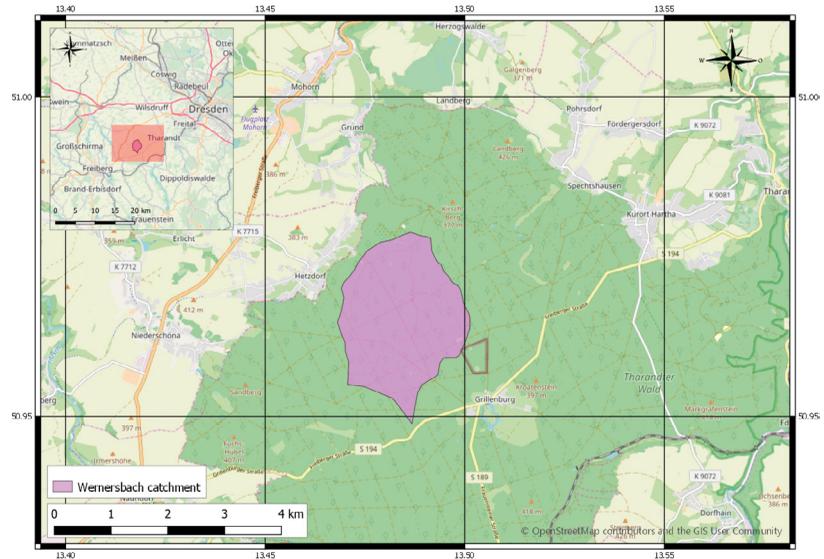


Fig. 1. Wernersbach catchment

The objective of our study is to perform a first of its kind investigation for the Wernersbach catchment, based on an approach incorporating two data-driven models, namely a statistical regression model and an artificial neural network model, in order to validate and add to better understanding on the suitability of the models for the catchment's discharge simulations. It also aims to provide good insights for future studies that would incooperate it's conclusion and develop on its limitations for future research studies.

2. STUDY AREA AND METHODOLOGY

The catchment area of the Wernersbach is located approximately 25 km southwest of Dresden, in the north-western region of the landscape conservation area of Tharandt Forest and the eastern part of the Ore Mountains (latitude 50°58'N, longitude 13°28'E). The area covers 4.6 km² and with an altitude above sea level within the range of 323-424 m.

The location of the catchment is depicted at the top left-hand side of in Figure 1, with the catchment location marked in red on the political boundary of Germany. In addition to this, Figure 1 presents the boundary of the Wernersbach catchment in purple and the river and surrounding areas in dark green.

The local topography and the dominance of forestry as the land usage type are observed to majorly contribute to the deviations from the average climatic conditions. The area mainly comprises low slopes of less than 3° (Bernhofer 2002). These small-scale relief forms have a great impact, as concave relief forms tend

to lead to lower temperatures (cold-air formation), and convex relief forms generally exhibit higher temperatures in the winter. These differences in relief also influence incident radiation, wind and precipitation (Bernhofer 2002).

The climate is essentially characterized by maritime influences whereby the greatest amount of precipitation is observed during the summer months.

The mean annual temperature is 7.5°C, and the mean annual precipitation is approximately 847 mm (Goldberg, Bernhofer 2007). Groundwater occurs only in rhyolite rocks. Therefore, large fluctuations in groundwater levels can occur within a short time period (Gerold et al. 1998).

In the upper surface of the paleorhyolite (porphyry), weathering products exhibit strong cohesive properties. During the early cretaceous, small low-binder sandstones with embedded clay horizons were present in the area. These sandstones are effective in water management. At the layer boundaries of the sandstone and the porphyry layer sources, water discharges are also recorded. Moreover, further marine sandstone deposits have been preserved from the transitional time (i.e. the early cretaceous to the late cretaceous), with weathering products that are prone to frost formation.

The Wernersbach area is used exclusively for forestry purposes. Spruce is widely predominant and is used to strike other conifers and deciduous trees. In order to improve growth conditions for the forest stands at locations affected by water pollution, a dense net-

Table 1. Water balance components of the catchment Wernersbach (Goldberg, Bernhofer 2007)

| Water Balance Components | Measure | Unit(s) |
|----------------------------|---------------------------|-----------|
| Precipitation | 847 | mm |
| Runoff | 240 | mm |
| Evapotranspiration | 607 | mm |
| Minimum runoff (mean) | 2.5 | litre/sec |
| Mean runoff | 35 | litre/sec |
| Maximum runoff (mean) | 1,228 | litre/sec |
| Maximum runoff (estimated) | 80,00-10,000 (13.08.2002) | litre/sec |

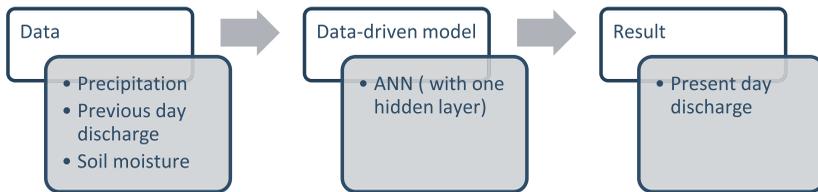


Fig. 2. Modelling process

work of artificial drainage trenches has been constructed (Bernhofer 2002).

The natural conditions of the Wernersbach catchment make it a suitable study area for the quantification and measurement of the water balance in space and time. The water balance components of the catchment are presented in Table 1.

This study uses climate and hydrological data from the Wernersbach catchment collected from 2001 to 2009. All data was initially stored on-site in data loggers and has a temporal resolution of one day. Field observation measurements used for the runoff models include the depth to groundwater table, precipitation, soil moisture at 30 and 60 cm and discharge.

2.1. ARTIFICIAL NEURAL NETWORK

ANN technology is a computational approach inspired by studies of the brain and nervous systems (Luk et al. 2001). The ANN technique has a number of interconnected processing elements that generally operate in parallel with regular configurations. These neural networks are capable of modelling both linear and non-linear systems (Riad et al. 2004).

There are two types of artificial neural networks: (i) the feed-forward neural network; (ii) recurrent/feed-back networks. The multi-layer field network (MLF) is a typical feed-forward network where adjusted weight coefficients are calculated and outputs are highly accurate (Svozil et al. 1997). A very simple model based on historical data, namely, drawing a line to best separate between critical and non-critical conditions, is used for the MLF. The possible applications of this model include flash flood susceptibil-

ity analysis, evaluating the multi-correlation function between rainfall and runoff, soil moisture and the precipitation of a hydrological catchment area.

An MLF neural network consists of neurons that are ordered into layers. The first layer is denoted as the input layer, the last layer is the output layer, and the layers in between are hidden layers (Svozil et al. 1997). The model used in this study, presented in Figure 2, has one hidden layer with a sigmoidal activation function, and a linear activation function in the output layer. The network is prepared using a simple error backpropagation algorithm.

2.2. REGRESSION ANALYSIS

The regression analysis in its simplest form uses two variables, one as the dependent variable and the other as the independent variable, thus making it possible to study the cause and effect of the relationship. A linear regression model that involves more than one independent variable is known as a multiple linear regression (MLR) model. Multiple regression analysis is used to establish the statistical relationship between one dependent variable and one or more independent variables X_1, X_2, \dots, X_p and is of the form (1) (Jaya Rami Reddy 2013) the multiple regression equation. The term linear is used because equation (1) is a linear function of the unknown parameters b_0, b_1, \dots, b_p :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p \quad (1)$$

In non-linear regression analysis, the dependent variables are modelled as a non-linear function of the model parameters and one or more

independent variables (Bilgili 2010). The multiple non-linear regression model (MNL) is a simple and efficient method in producing more accurate maximum daily discharge predictions compared to the ANN, the adaptive neuro-fuzzy inference system and the MLR (Rezaeianzadeh et al. 2014). The multiple non-linear regression equation generally is of the form (Jaya Rami Reddy 2013):

$$Y = b_0 X_1^{b_1} X_2^{b_2} \dots X_p^{b_p} \quad (2)$$

where b_0, b_1, \dots, b_p are the parameters of the non-linear relationship. Multiple non-linear regression problems can be linearized using simple logarithmic transformation by taking the logarithms of both sides of equation (2) (Jaya Rami Reddy 2013):

$$\ln Y = \ln b_0 + b_1 \ln X_1 + b_2 \ln X_2 + \dots + b_p \ln X_p \quad (3)$$

Moreover, the regression of $\ln Y$ on $\ln X_1, \ln X_2, \dots, \ln X_p$ is utilized for estimating b_0, b_1, \dots, b_p .

2.3. MODEL VERIFICATION

A statistical evaluation method can provide an indication of the best model, while the graphical and hydrological interpretation of the presented datasets and models can evaluate this simplistic indication. Based on their soundness and robustness, traditional log-log rating curves have been observed to be superior, regardless of their poor goodness-of-fit statistics (Abrahart et al. 2011).

2.4. AREAL PRECIPITATION

The Thiessen polygon method introduced by Thiessen (1911), is used here to determine the average amount of precipitation over the study area. It is a graphical technique that calculates areas relating to specifically placed rain gauges, deriving an areal value plus a reference, resulting in polygons within polygons (see Fig. 3).

3. GENERAL STUDY APPROACH

We assume an ideal prognosis using the observed precipitation and soil moisture data. This ideal prognosis of the meteorological input allows us to neglect prognosis uncertainties, which are mostly connected to modelled precipitation from dynamic weather forecast models. Therefore, the resulting model performances are mainly defined by the approaches themselves, rather than by accuracy of the data.

The basic architecture of the approach used in this study is presented in Figure 2. The three

white boxes demonstrate the three basic steps, i.e. the data, the data-driven model and the result. The three grey boxes show the sub-routines. The first step consists of inputs that can influence the output, i.e. precipitation, previous day discharge, and soil moisture. The second step consists of the models used to analyze the inputs. The last or output step consists of the predicted values or results, i.e. present day discharge.

The validation strategy was used to create statistically robust results from 100 randomly chosen subsets, in order to quantify the impact of extremes in the prediction. Performance indices were used to compare the performance of both models, namely the coefficient of determination (4), the root mean square error (*RMSE*) (5) and the mean absolute percentage error (*MAPE*) (6).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

In the above equations, y_i represents the observed water discharge, \hat{y}_i represents the forecasted water discharge, \bar{y} represents the average observed water discharge and average predicted water discharge respectively, and n represents the number of observations in both calibration and validation stages.

The R -squared (R^2) value was used to measure how close the data is to the fitted regression line. It provides information on the strength of the linear relationships between the observed and predicted values. Moreover, the *RMSE* (5) represents the prediction of the errors in the model.

All analysis was performed using the R programming package. The Neuralnet R package was used for the ANN model. For the MLR model, an in-house script was written. The ANN model exhibited higher computing times due to the training of the neural network. One catchment at a time was used for both models due to the minimal amount of computer resources required for the MLR (statistical) and the ANN (data-driven). The model ensembles were then used to identify the best estimate for a flood event.

4. RESULTS AND DISCUSSION

The areal significance to the point rainfall values assigned by the Thiessen polygon method for five rain gauges are depicted in Figure 3. The respective sizes of the areas are reported in Table 2.

A relationship was derived between precipitation, soil moisture and discharge, considering days when precipitation was greater than 0.1 mm, and using daily soil moisture data and discharge values for previous days. During 2001-2009, 1,804 days were observed to have a precipitation value of more than 0.1 mm per day. In order to reduce random errors from using linear regression only once, the 1,804 events with precipitation values, rainfall-runoff, and soil moisture were randomly divided into two halves 100 times.

According to the results of the calibration and validation data sets, the median R^2 value for both data sets was observed as 0.887. This indicates a strong correlation between the considered variables. In addition, both data sets exhibit the same distribution shape (Figs. 4 and 5) for the determined R^2 values.

The logarithmic linear dependence can be expressed as follows:

$$\ln Q_{(0)} = 0.228 \ln Q_{(-1)} + 0.002 Sw_{(-1)} + 0.9 \ln P_{(0)} - 0.215 \quad (7)$$

where Q represents discharge, Sw represents soil moisture, P represents precipitation, the indices 0 and -1 denote the current and previous day respectively.

The resultant model is characterized by the values of the selected performance indices presented in Table 3.

In Figure 6, Q_{mod} represents the modelled specific discharge, while Q_{obs} represents the observed specific discharge. Both values are converted to mm, in order to have the same units as precipitation and soil moisture. The conversion to discharge involves multiplying the specific discharge by the size of the Wernersbach catchment. The black line in Figure 6 represents the trend line without considering the extreme values, while the green line represents the trend line when extreme values are considered.

The trend lines demonstrate a consistency in the observed maximum values, particularly for the validation data set. Both data sets suggest that the model tends to slightly overestimate high flood events by 1.01% and 14% for the validation and the calibration data sets, respectively.

The range of predicted errors between the simulated value and the measured values

for the statistical regression model is shown in Figure 8.

It should be noted that 1,798 rain events from the total 1,804 lie within the range of 0-13 mm, with the maximum (1,593 events) within in the range of 0-1 mm.

A good correlation between the simulated and measured specific discharge is observed for the range of 0-13 mm. In contrast, greater uncertainty is associated with discharge above 13 mm, due to the small number of precipitation events within 13-61 mm (only 6 events).

High uncertainty indicates a high demand for data and a small and responsive characterization of the catchment area. Under such conditions, it is difficult to predict any extremes.

The discharge value in the Wernersbach catchment depends primarily on precipitation. A close relationship between all three parameters is exhibited. A significant input from the soil moisture parameter can be observed during events with high discharge values. Using the median absolute deviation (MAD) method (Leys et al. 2013), 198 precipitation events were determined as “outliers” or bankful discharge. These events suggest a rise in discharge and flood, or high flood, events. Here, we denote them as extreme events.

Following its successful launch, the ANN model was validated. The MLR and ANN models’ results are assessed with a statistical representation of the histograms and Q-Q plots. The data in both models runs are with and without extremes (the 198 events described earlier were excluded). Based on this, different scenarios are analyzed.

A perfect bell-shaped curve representing a normal distribution is observed where the allocated dataset does not include extreme values. The non-linear characteristic of the rainfall-runoff process is normalized using a logarithmic transformation. This transformation is applied in order to provide an improved fit of the R^2 values to the normal distribution (Fig. 9).

The *MAPE* value is observed as 153% for the MLR model and 140% for the ANN model. The *MAPE* errors demonstrate that the ANN model has a smaller variance in prediction errors than the MLR model compared to the original dataset scale.

We find that both models overestimate the simulated values. The dataset validation with extreme values for the MLR model is equal to 14%, while the ANN model demonstrates a lower error value of 0.7% for the dataset validation with extreme values and 0.5% without extremes values.

The above result implies that the ANN model better generalizes the variability of high

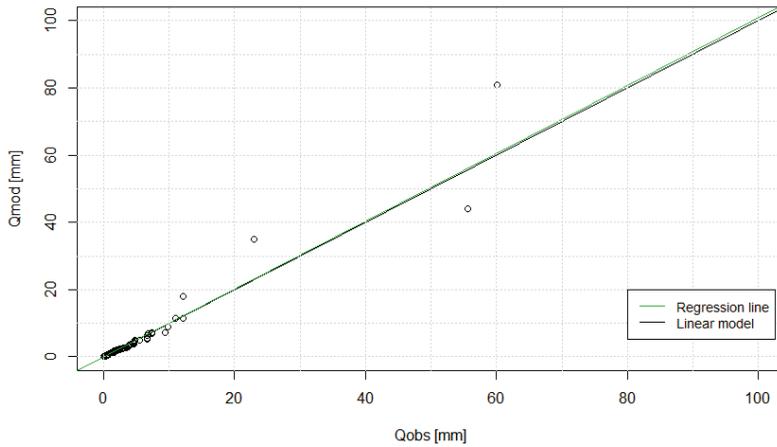


Fig. 7. Dependencies between the simulated and the measured values for the validation data set of the statistical regression model

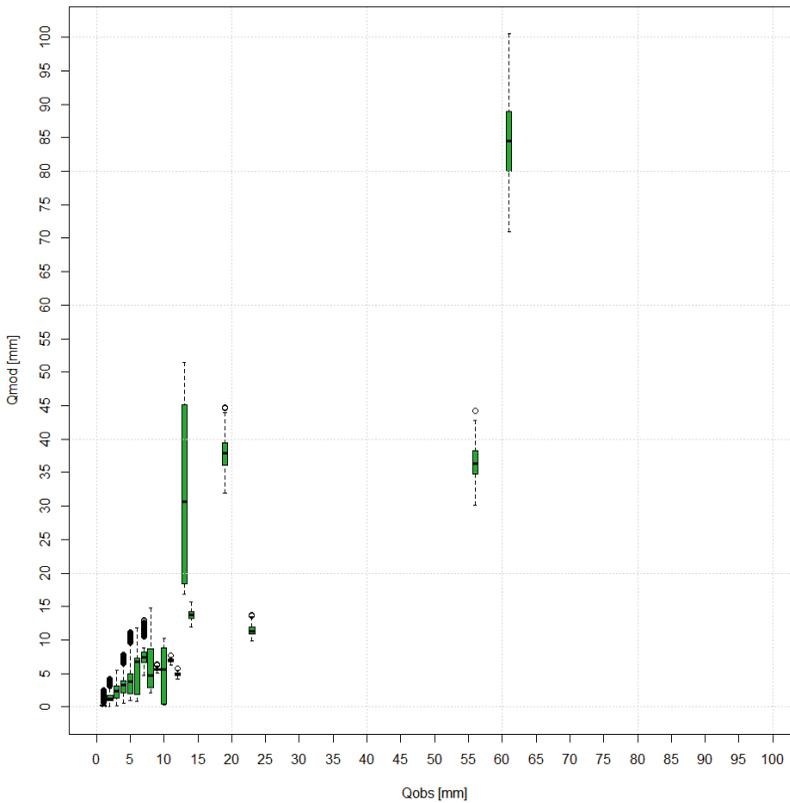


Fig. 8. Box plot showing the range of errors between the simulated and measured values for the statistical regression model

floods in the observation period compared to the MLR regression model. The results from the former can be considered more satisfactory over the whole period, with a large annual variability of extreme events. The estimates are sufficient to provide on-time warnings with minimal errors.

The snowmelt factor was not considered as influential in this study. However, including climate change as a driving factor

and considering studies for an earlier initiation of the snowmelt season (Schneider, Schönbein 2005; Blöschl et al. 2017), increases the duration of the snowmelt season. Thus, it is important to include snow/ice –melt and temperature in the models in order to provide near real-time forecasts. This will be duly addressed in our future and on-going studies.

Soil moisture is often found to have a triggering influence on floods and is included

in the models, depending on the availability of the measured input. In the typical data-poor conditions that characterize flash flood forecasting and warnings, surrogate indexes that implicitly consider the soil moisture initial state, are often extremely useful.

5. CONCLUSION

In this study, we investigate an approach incorporating data-driven modelling for the Wernersbach catchment with three observed parameters: (i) soil moisture, (ii) precipitation of the current day and (iii) precipitation from the previous day. An alternative model is run to compare and validate the results. Both models are on loop fast enough to include several statistical analysis methods, with an efficiency in forecasting results for time stringent cases. Based on the catchment characteristics, the model based on regression analysis is determined as a more effective statistical method for the estimation of flood discharge compared to other methods.

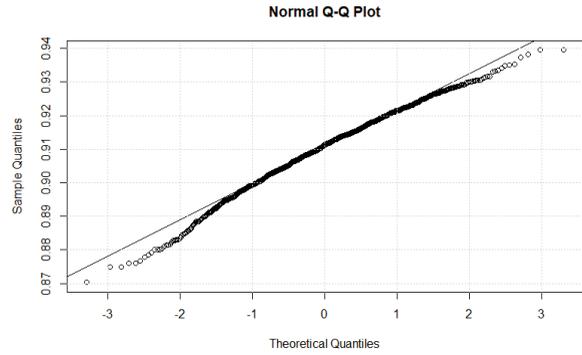
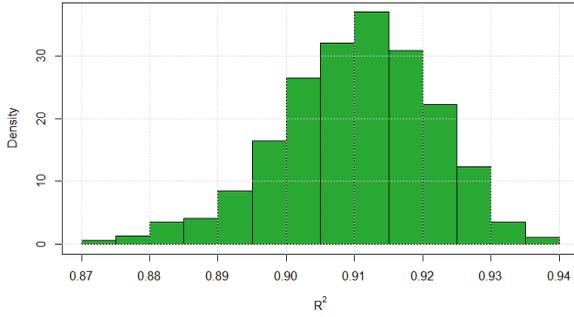
A more detailed alternative approach would involve the use of hydraulic formulae (e.g. Manning equation) or one- or two-dimensional hydraulic models (e.g. the Saint-Venant equations) to convert historical flood levels into historical discharges (Benito et al. 2004). Moreover, flash flood guidance, which tags the rainfall accumulation needed to produce a flood of a given magnitude according to current soil moisture conditions, has proven useful for ungauged basins (Borga et al. 2011).

For the Wernersbach catchment area, our results show that data-driven methods are a feasible alternative to the flash flood guidance approach. the ANN model is generally preferable, even for extreme events, compared to the MLR approach.

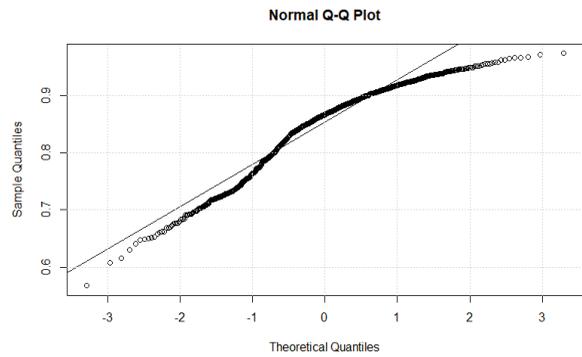
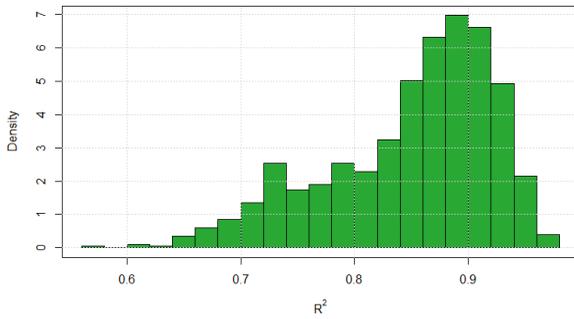
However, the considered temporal scale (2001-2009) do not include significant amount of past major hydrological extreme events and exclusion of uncertainties by climate change limits our study in various aspects. Despite these limitations, we are confident ANN model is better suited to the tasks of rainfall-runoff and flood forecasting over traditional MLR approach.

Nevertheless, as per our current knowledge, this study is the only one of its kind done over the Wernersbach catchment and is presumed to build foundation for further application of ANN models in research focusing on the influence of spatial and temporal rainfall patterns on the estimation of rainfall thresholds, as well as the prediction of soil moisture.

a) ANN normal



b) ANN model run over logarithm



c) MLR normal

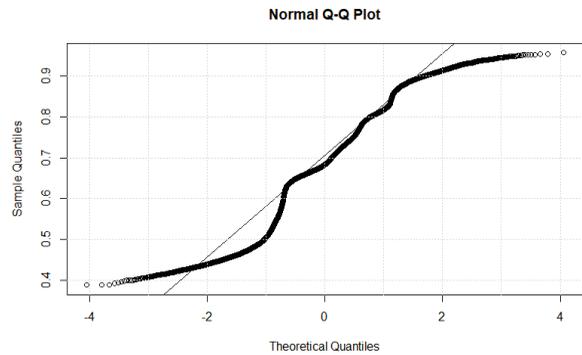
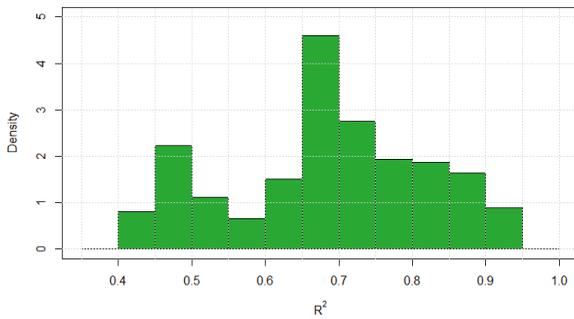
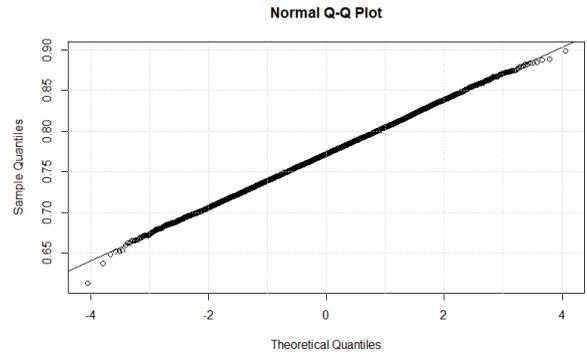
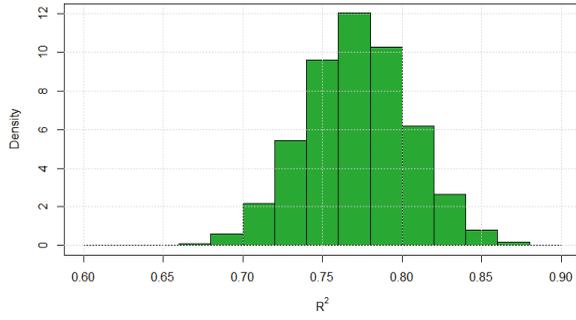
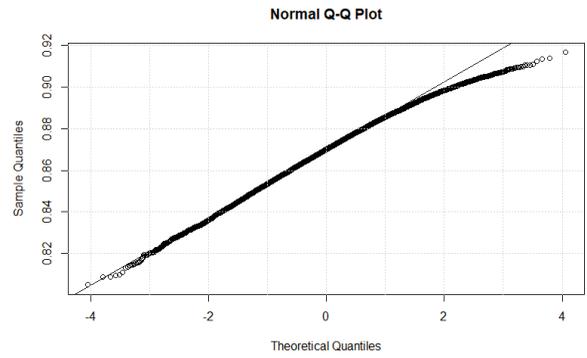
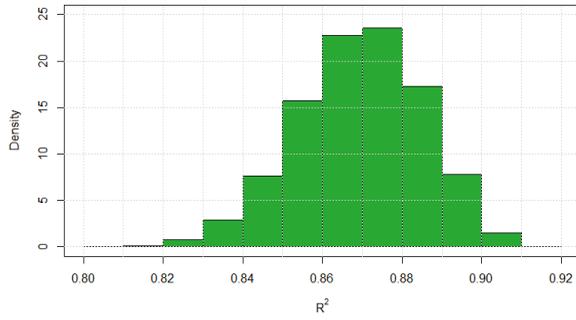


Fig. 9a-c. Distribution of R^2 and Q-Q plots: a) Including extreme data for the ANN model run; b) including extreme data for the ANN model run; c) including extreme data for the MLR model run

d) MLR normal without extremes



e) MLR model run over logarithm



f) MLR model run over logarithm without extremes

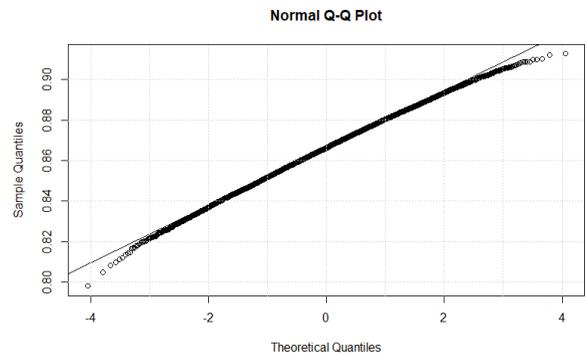
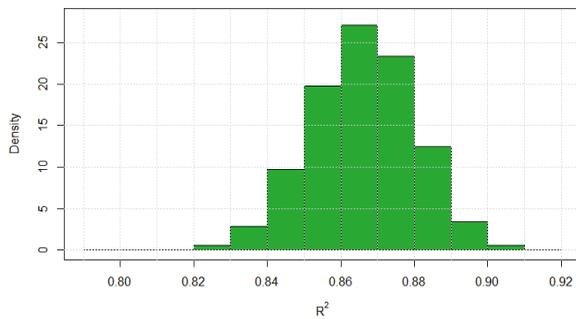


Fig. 9d-f. Distribution of R2 and Q-Q plots: d) without extreme data for the MLR model run; e) including extreme data for the MLR analysis on the dataset run over logarithm; f) without extreme data the MLR analysis on the dataset run over logarithm

REFERENCES

- Abebe A.J., Guinot V., Solomatine D.P., 2000a, Fuzzy alpha-cut vs. Monte Carlo techniques in assessing uncertainty in model parameters, [in:] Proceedings of the 4th International Conference on Hydroinformatics, Iowa City, USA, July 2000, available online <https://pdfs.semanticscholar.org/35ae/0ec4107ac2950e7b-da62042b1056d732e7dc.pdf> (data access 17.09.2019)
- Abebe A.J., Solomatine D.P., Venneker R.G.W., 2000b, Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events, *Hydrological Sciences Journal*, 45 (3), 425-436, DOI: 10.1080/02626660009492339
- Abrahart R.J., Mount N.J., Ghani N.A., Clifford N.J., Dawson C.W., 2011, DAMP: A protocol for contextualizing goodness-of-fit statistics in sediment-discharge data-driven modelling, *Journal of Hydrology*, 409 (3), 596-611, DOI: 10.1016/j.jhydrol.2011.08.054
- Abrahart R.J., See L.M., 2000, Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, *Hydrological Processes*, 14 (11-12), 2157-2172, DOI: 10.1002/1099-1085(20000815/30)14:11/12<2157::AID-HYP57>3.0.CO;2-S
- Abrahart R.J., See L.M., Solomatine D.P., 2008, Practical hydroinformatics: Computational intelligence and technological developments in water applications, Springer-Verlag Berlin Heidelberg, 505 pp., DOI: 10.1007/978-3-540-79881-1
- Benito G., Lang M., Barriendos M., Llasat M.C., Francés F., Ouarda T., Thorndycraft V.R., Enzel Y., Bardossy A., Coeur D., Bobée B., 2004, Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. Review of scientific methods, *Natural Hazards*, 31 (3), 623-643, DOI: 10.1023/B:NHAZ.0000024895.48463.eb
- Bernhofer C., 2002, Guidebook on excursions and internships in the Tarandt forest: materials for hydrological and meteorological field practice, (in German), Technische Universität Dresden, 292 pp.
- Bhattacharya B., Solomatine D.P., 2003, Neural networks and M5 model trees in modeling water level-discharge relationship for an Indian river, [in:] Proceedings of the 11th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 2003, 407-412, available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.3898&rep=rep1&type=pdf> (data access 17.09.2019)
- Bilgili M., 2010, Prediction of soil temperature using regression and artificial neural network models, *Meteorology and Atmospheric Physics*, 110 (1-2), 59-70, DOI: 10.1007/s00703-010-0104-x
- Blöschl G., Hall J., Parajka J., Perdigão R.A.P., Merz B., Arheimer B., Aronica G.T., Bilibashi A., Bonacci O., Borga M., Čanjevac I., Castellarin A., Chirico G.B., Claps P., Fiala K., Frolova N., Gorbachova L., Gül A., Hannaford J., Harrigan S., Kireeva M., Kiss A., Kjeldsen T.R., Kohnová S., Koskela J.J., Ledvinka O., Macdonald N., Mavrova-Guirguinova M., Mediero L., Merz R., Molnar P., Montanari A., Murphy C., Osuch M., Ovcharuk V., Radevski I., Rogger M., Salinas J.L., Sauquet E., Šraj M., Szolgyaj J., Viglione A., Volpi E., Wilson D., Zaimi K., Živković N., 2017, Changing climate shifts timing of European floods, *Science*, 357 (6351), 588-590, DOI: 10.1126/science.aan2506
- Borga M., Anagnostou E.N., Blöschl G., Creutin J.-D., 2011, Flash flood forecasting, warning and risk management: the HYDRATE project, *Environmental Science & Policy*, 14 (7), 834-844, DOI: 10.1016/j.envsci.2011.05.017
- Cannas B., Fanni A., See L.M., Sias G., 2006, Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning, *Physics and Chemistry of the Earth. Parts A/B/C*, 31 (18), 1164-1171, DOI: 10.1016/j.pce.2006.03.020
- Clarke R.T., 1994, *Statistical modelling in hydrology*, 1 edition, Wiley & Sons, Chichester, 426 pp.
- Dawson C.W., Wilby R., 1998, An artificial neural network approach to rainfall-runoff modelling, *Hydrological Sciences Journal*, 43 (1), 47-66, DOI: 10.1080/02626669809492102
- Dibike Y.B., Solomatine D.P., Abbott M.B., 1999, On the encapsulation of numerical-hydraulic models in artificial neural network, *Journal of Hydraulic Research*, 37 (2), 147-161, DOI: 10.1080/00221689909498303
- Gerold G., Cyffka B., Suttmöller J., Krüger J.-P., Herbst M., 1998, Regionalisierung der Abfließbildung über die Aggregation homogener Flächen unter Verwendung des Geographischen Informationssystems ARC/INFO und der Digitalen Reliefanalyse (Programm SARA), [in:] *Hydrologie und Regionalisierung: Ergebnisse eines Schwerpunktprogramms (1992-1998)*, Deutsche Forschungsgemeinschaft (ed.), Wiley-VCH Verlag GmbH, 191-204, DOI: 10.1002/9783527609260.ch9
- Giustolisi O., Savic D.A., 2006, A symbolic data-driven technique based on evolutionary polynomial regression, *Journal of Hydroinformatics*, 8 (3), 207-222, DOI: 10.2166/hydro.2006.020b
- Goldberg V., Bernhofer C., 2007, Water balance of forested catchments, [in:] *Global change: enough water for all?*, L.J. Lozán (ed.), Wissenschaftliche Auswertungen, 72-76
- Hsu K., Gupta H.V., Sorooshian S., 1995, Artificial Neural Network modeling of the rainfall-runoff process, *Water Resources Research*, 31 (10), 2517-2530, DOI: 10.1029/95WR01955
- Humphrey G.B., Gibbs M.S., Dandy G.C., Mair H.R., 2016, A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network, *Journal of Hydrology*, 540, 623-640, DOI: 10.1016/j.jhydrol.2016.06.026
- Jaya Rami Reddy P., 2013, *Stochastic hydrology*, Laxmi Publications Ltd, New Delhi, 259 pp.
- Khu S.T., Liong S.-Y., Babovic V., Madsen H., Muttill N., 2001, Genetic programming and its application in real-time runoff forecasting, *Journal of the American Water Resources Association*, 37 (2), 439-451, DOI: 10.1111/j.1752-1688.2001.tb00980.x
- Leys C., Ley C., Klein O., Bernard P., Licata L., 2013, Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, *Journal of Experimental Social Psychology*, 49 (4), 764-766, DOI: 10.1016/j.jesp.2013.03.013
- Lobbrecht A.H., Solomatine D.M., 1999, Control of water levels in polder areas using neural networks and fuzzy adaptive systems, [in:] *Water industry systems: modelling and optimization applications*, vol. 1, D. Savic, G. Walters (eds.), Research Studies Press, Ltd, Baldock, UK, 509-518
- Luk K.C., Ball J.E., Sharma A., 2001, An application of artificial neural networks for rainfall forecasting, *Mathematical and Computer Modelling*, 33 (6-7), 683-693, DOI: 10.1016/S0895-7177(00)00272-7
- Minns A.W., Hall M.J., 1996, Artificial neural networks as rainfall-runoff models, *Hydrological Sciences Journal*, 41 (3), 399-417, DOI: 10.1080/02626669609491511
- Nourani V., Kisi Ö., Komasi M., 2011, Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process, *Journal of Hydrology*, 402 (1-2), 41-59, DOI: 10.1016/j.jhydrol.2011.03.002
- Pesti G., Shrestha B.P., Duckstein L., Bogárdi I., 1996, A fuzzy rule-based approach to drought assessment, *Water Resources Research*, 32 (6), 1741-1747, DOI: 10.1029/96WR00271

- Rezaeianzadeh M., Tabari H., Arabi Yazdi A., Isik S., Kalin L., 2014, Flood flow forecasting using ANN, ANFIS and regression models, *Neural Computing and Applications*, 25 (1), 25-37, DOI: 10.1007/s00521-013-1443-6
- Riad S., Mania J., Bouchaou L., Najjar Y., 2004, Rainfall-runoff model using an artificial neural network approach, *Mathematical and Computer Modelling*, 40 (7-8), 839-846, DOI: 10.1016/j.mcm.2004.10.012
- Schneider C., Schönbein J., 2005, Climatological analysis of snow reliability and snow-making of winter sports resorts in German low mountain ranges, (in German), *Institute of Nature Sports and Ecology, German Sport University Cologne*, 37 pp.
- Schulla J., 1997, *Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen*, Doctoral Thesis, Swiss Federal Institute of Technology in Zurich, 154 pp., DOI:10.3929/ethz-a-001763261
- Solomatine D.P., Avila Torres L.A., 1996, Neural network approximation of a hydrodynamic model in optimizing reservoir operation, [in:] *Proceedings of the International Conference on Hydroinformatics, Zurich, 1996*, 201-206, available at <https://www.un-ihe.org/sites/default/files/solomatinetorresann-approx-reserv-ophtic-1996.pdf> (data access 18.09.2019)
- Solomatine D.P., See L.M., Abrahart R.J., 2008, Data-driven modelling: concepts, approaches and experiences, [in:] *Practical hydroinformatics*, R.J. Abrahart, L.M. See, D.P. Solomatine (eds.), *Water Science and Technology Library*, 68, Springer, Berlin, Heidelberg, 17-30, DOI: 10.1007/978-3-540-79881-1_2
- Sudheer K.P., Jain S.K., 2003, Radial basis function neural network for modeling rating curves, *Journal of Hydrologic Engineering*, 8 (3), 161-164, DOI: 10.1061/(ASCE)1084-0699(2003)8:3(161)
- Svozil D., Kvasnicka V., Pospichal J., 1997, Introduction to multi-layer feed-forward neural networks, *Chemometrics and Intelligent Laboratory Systems*, 39 (1), 43-62, DOI: 10.1016/S0169-7439(97)00061-0
- Thiessen A.H., 1911, Precipitation averages for large areas, *Monthly Weather Review*, 39 (7), 1082-1084, DOI: 10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2
- Vernieuwe H., Georgieva O., De Baets B., Pauwels V., Verhoest N., De Troch F., 2005, Comparison of data-driven Takagi-Sugeno models of rainfall-discharge dynamics, *Journal of Hydrology*, 302 (1-4), 173-186, DOI: 10.1016/j.jhydrol.2004.07.001
- Yao X., Liu Y., 1997, A new evolutionary system for evolving artificial neural networks, *IEEE Transactions on Neural Networks*, 8 (3), 694-713, DOI: 10.1109/72.572107

