

On the instabilities of estimated distributions of the POT (Peak Over Threshold) low flow characteristics

Wojciech Jakubowski

Wrocław University of Environmental and Life Sciences, The Faculty of Environmental Engineering and Geodesy, Department of Mathematics, Centrum Naukowo-Dydaktyczne, Grunwaldzka Street 55, 50-357 Wrocław, Poland, e-mail: wojciech.jakubowski@up.wroc.pl

Abstract: In a certain stream gauge profile, consider the low flow flows determined with the POT (Peak Over Threshold) method. Each of the flows can be described by three characteristics – deficit, duration and minimal flow. Values of the three-dimensional random variable depend on the choice of truncation level Q_g (threshold flow) – POT method parameter. It is typically assumed that the threshold level is included within the range from $Q_{95\%}$ to $Q_{60\%}$ (Tallaksen, van Lanen 2004). However, in computational practice the Q_g value is determined at the level of either $Q_{90\%}$ to $Q_{70\%}$. This choice is made mainly from the hydrological (not statistical) point of view.

In this paper the influence of the threshold flow on the form of estimated distributions of each of the above three characteristics is considered. The following distributions are chosen:

- GEV (generalised extreme value distribution) – while examining the distribution of extremes;
- log-normal – in the non-extreme case.

In each of the examined stream gauge profiles the following algorithm was used:

1. from the curve of duration sums, two flow values $Q_{90\%}$ and $Q_{55\%}$ are chosen
2. for each flow from the range ($Q_{90\%}$, $Q_{55\%}$), using the Zelenhasić method (1987), a three-dimensional sequence is determined of observed deficits, durations and minimal flow values;
3. for each of the one-dimensional sequences, the parameters of the above distributions are estimated.

The variability of the estimated quantiles and their intervals of confidence were shown with the example of three gauge profiles – Kuripapango (New Zealand), Bogusław (Prosna) and Bystrzyca Kłodzka (Nysa Kłodzka).

Keywords: low flows, probability distributions, threshold level

Submitted 25 May 2014, **revised** 12 July 2014, **accepted** 19 October 2015

1. Introduction

The hydrological term „low flow” is a conventional concept which describes the situation of low water flow in rivers (Ozga-Zielińska, Brzeziński 1997). Literature provides examples of the evolution of both the definition of low flow and the criteria for its determination. From as early as the 1990s and the papers by Ozga-Zielińska (1990), Ozga-Zielińska, Brzeziński (1997), Tallaksen, van Lanen (2004) and Bonacci (1993), a simple and clear definition exists: »Low flow is the time when the water flows are equal or lower than the assumed threshold level Q_g «. This way of defining low flow is called the threshold level method (POT – peak over threshold). This method was applied for the first time by Yevjevich (1967) and it's been used by many authors during the statistic estimation of the distribution of low flow characteristics in a specific gauge station and during its geographical regionalization (see for e.g. Bower et al. 2004). The review of literature papers illustrating the applications of the threshold level

in hydrological practice may be found in the monograph by Tallaksen, van Lanen (2004) and Smakhtin (2001). The standard method of Yevjevich does not cover seasonal changes. Different methods of determining low flow, also with regard to seasonal periodicity, are described in the paper by Hisdal, Tallaksen (2000) and in the WMO instruction (Gustard, Demuth 2009).

The researcher encounters two kinds of difficulties when applying the method of the threshold level Q_g , with the time step of the measurement as 24 hours:

- certain further observed low flows may be mutually dependent, they have arisen during the same hydro-meteorological phenomena;
- in a sequence of observations, a large amount of small low flow is present.

In such cases, Tallaksen (2000) comes forward with a method of aggregating dependent low flow and removing small low flow, insignificant during the analysis of extreme phenomena, from the set of observations. Various criteria are applied in this case. A criterion frequently

used is the two-phase algorithm, proposed by Zelenhasić, Salvai (1987). During the first phase, the dependent low flows are joined, whereas during the second one the low flows, whose duration or deficit volume do not exceed the assumed restricting values, are removed from the observations. The selection of the method parameters depends on the watershed size, its nature and the target of the conducted statistical estimation of the distribution of the low flow characteristics (Tallaksen, van Lanen 2004).

Low flow and the associated phenomena of soil drought and hydrological drought concern many branches of human activity. For this reason the criteria of threshold flow selection are different. Ozga-Zielińska, Brzeziński (1997) claim that *»These criteria are either hydrological or economic. The kind of the assumed criterion depends on the target of the paper, its content should be justified and the obtained results properly interpreted«*. In practice, while selecting Q_g , different criteria are applied. Normally, it is assumed that Q_g takes the values: $Q_{90\%}$, $Q_{70\%}$ (Hisdal 2002; Tallaksen, van Lanen 2004), $Q_{95\%}$, $Q_{70\%}$ (WMO instruction by Gustard, Demuth 2009) or is equal to one of the specific flows Highest Annual Minimum Flow (HMF) (Ozga-Zielińska, Brzeziński 1997) or Average Annual Minimum Flow (AMF) (Kaznowska 2011) or even the first quartile $Q_{75\%}$ (Tokarczyk, Jakubowski 2006).

After selecting the proper threshold level and setting the values of the restricting parameters from the time series of everyday observations, the low flow characteristics are extracted. Normally this is three values, namely – deficit volume, duration and minimum flow. Consequently, the estimation of their probability distribution is executed. First of all, the one dimension distributions are considered, each characteristic separately. It is assumed that for extreme events during the deficit and duration estimations, either the general Pareto distribution (Hisdal 2002; Jakubowski 2011) or the general extreme value distribution GEV (Tallaksen, van Lanen 2004; Gustard, Demuth 2009) are used. The GEV distribution, in its minimum version, is used for the estimation of the minimal outflow. For non-extreme events the lognormal distribution is applied, and the estimation is carried out for low flow deficits and durations only.

The selection of the threshold level Q_g normally takes place with the unchecked assumption that the estimated distributions are stable, which means that even small changes of the threshold level value Q_g cause small changes in the estimated quantile values. Jakubowski (2011) has shown that it is not a rule in the case of the threshold level change range ($Q_{90\%}$, $Q_{55\%}$). While testing the two dimensional general Pareto distribution of the two maximum low flow characteristics deficit and duration, he showed that both the low

flow deficit and its duration for certain threshold levels demonstrate instability. Such instability has an enormous influence on the proper estimating of the probabilities of the low flows extremes – unrecognized results in the overestimating of the low flow duration and deficit.

Instability of one dimensional estimated distribution was also tested by Hisdal (2002). Using the general Pareto distribution, she demonstrated that it exists for low flow deficits with threshold levels lower than $Q_{95\%}$.

In this paper a similar instability is examined, but for the probabilistic description of three extreme characteristics of low flow – deficit, duration and minimal flow. For researching the extremes, a generalised extreme value (GEV) distribution was used with a cumulative distribution function:

$$F(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \text{ when } 1 + \xi\left(\frac{x - \mu}{\sigma}\right) > 0 \quad (1)$$

This distribution form was applied for the deficit and low flow duration. For the minimal flow the GEV distribution was applied in the following adjusted form:

$$F(x) = 1 - \exp\left\{-\left[1 - \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \text{ when } 1 - \xi\left(\frac{x - \mu}{\sigma}\right) > 0 \quad (2)$$

In the non-extreme case a log-normal distribution with the following distribution (only for the deficit and low flow duration) was applied:

$$F(x) = \int_0^x \frac{1}{\sigma\sqrt{2\pi\sigma t}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right) dt \quad (3)$$

2. Computation algorithm

In each of the examined water gauge profiles, the following algorithm was applied:

Step 1. From the flow duration curve two threshold flow values $Q_{90\%}$, and $Q_{55\%}$, are chosen. The criteria for this choice are as follows:

- for threshold flows lower than $Q_{90\%}$, the number of low flows determined from a 40-year sequence of daily flows is too small to correctly estimate distribution parameters (Hisdal 2002);
- flow higher than the $Q_{55\%}$ threshold can hardly be considered low, on the other hand the HMF flow frequently exceeds the $Q_{60\%}$ level, so the choice of $Q_{55\%}$ as the highest defining low flow can be considered optimal.

Step 2. For each observed flow Q_g in the ($Q_{90\%}$, $Q_{55\%}$) range, using the Zelenhasić, Salvai (1987) method, a three-

dimensional sequence of observed deficits, durations and minimal flow values is determined. The method imposes the following restrictions on the determined sequence of observed low flows:

- the minimal interval between subsequent low flows equals m_t days;
- the minimal low flow duration is m_c days;
- the minimal flow deficit – according to which low flows with deficits below $\alpha_0 D_{max}$ are removed from the data sequence. Typically the parameter $\alpha_0 = 0.005$, and D_{max} means the highest noted deficit.

The first limitation eliminates possible dependencies between subsequent low flows, the others exclude insignificant observations. Their function is based on the following rules:

- in the first step, two subsequent low flows which fulfil the first condition are merged into one, their flow deficit and durations are added and for the minimal flow the lowest flow from the combined low flow is chosen;
- in the second step, single low flows that fulfil the other two criteria are removed from the data set.

In the calculations below it was assumed, after Zelenhasić, Salvai (1987), that $m_t = 3$ days, $m_c = 5$ days, while parameter α_0 , similarly to in Jakubowski's paper (2011), varies from 0.03 to 0.08. Such a high parameter α_0 causes the parameter m_c to lose significance – all short low flows, lasting no more than a few days, are eliminated from the data sequence.

Step 3. For each of the one-dimensional sequences, the parameters of the distributions (1) – (3) are estimated. Estimates were conducted using the maximum likelihood method. The goodness of fit was tested with the χ^2 test. At most of the threshold levels there was no reason to reject the hypothesis of goodness-of-fit tests.

Step 4. For the Z characteristic, the estimated distribution and the threshold flow Q_g quantiles $z_{p,g}$ are determined, defined by the probability $Pr(Z \leq z_{p,g}) = 1 - p$. Next, the variance $Var(\hat{z}_{p,g})$ is calculated. This is used to determine the confidence interval of the $z_{p,g}$ quantile. Conforming to the delta method (Coles 2001), the variance $Var(\hat{z}_{p,g})$ is approximated by $\nabla z_{p,g}^T V \nabla z_{p,g}$, where V is the covariance matrix of the distribution's estimated parameters, whereas $\nabla z_{p,g}$ is the derivative vector of the quantile $z_{p,g}$ relative to the distribution's estimated parameters. In the case of the GEV distribution, the matrix V is three-dimensional, while:

$$\nabla z_{p,g} = \left[\frac{\partial z_{p,g}}{\partial \mu}, \frac{\partial z_{p,g}}{\partial \sigma}, \frac{\partial z_{p,g}}{\partial \xi} \right]$$

whereas in the case of the log-normal distribution, V is two-dimensional, and:

$$\nabla z_{p,g} = \left[\frac{\partial z_{p,g}}{\partial \mu}, \frac{\partial z_{p,g}}{\partial \sigma} \right]$$

Step 5. Determination of the ranges of the threshold Q_g , in which the estimated probability distributions are unstable. Let us consider two threshold flows – Q_{g1} , Q_{g2} – such that $|Q_{g1} - Q_{g2}|$ is close to 0. One should then expect that $z_{p,g2}$ should belong to 50% of the interval of confidence for the $z_{p,g1}$ quantile, and symmetrically $z_{p,g1}$ should belong to 50% of the interval of confidence for the $z_{p,g2}$ quantile. If these conditions are met, the estimated distribution of the examined characteristic is stable; if they are not met, the estimated distribution will be unstable.

All numerical computations were carried out using the extended version of Nizowka2003 software (Jakubowski, Radczuk 2004).

3. Influence of the threshold flow on low flow characteristics

Calculations were conducted for flows in three gauging stations selected in order to reflect to some degree the orographical variability and area management of the drainage basin:

- Kuripapango – the drainage basin of the river Ngaruroro, New Zealand. The upper part of the basin, in Kuripango, which closes it with a stream gauge, is an example of near complete forestation of a mountainous drainage area (Tallaksen, van Lanen 2004). This drainage area has never been used agriculturally, and the beech forest covering it is part of the Kaweka Forest Park, and thus possesses the properties of a natural drainage basin – flows from years 1965-2000;
- Bogusław – the drainage basin of the Prosna is an example of a Polish lowland drainage basin under agricultural use – flows from years 1965-2005;
- Bystrzyca Kłodzka – the drainage basin of the Nysa Kłodzka is located in a foothill area and is of a forest-agricultural character – flows from years 1965-2005.

The Fig. (1) – (5) present the influence of threshold flow on the estimated quantiles of the high probability of non-exceeding. The horizontal axis marks the threshold flow Q_g , while the vertical axes denote either the flow deficit of the low flow, its duration, or minimal flow. The obtained results should be considered separately, depending on whether they are extreme low flows, or normal (non-extreme) ones.

3.1. Non-extreme low flows

For their description, a log-normal distribution (3) was used. Figures 1-2 present the estimated quantiles of the log-normal distribution for the duration and low flow deficit on two stream gauges, Kuripapango and Bogusław, and the observed maximum values of low flow deficit or duration, respectively.

Estimated 95% and 90% quantiles were included in 50% confidence intervals – these were determined for each examined truncation level Q_g . For all three curves, both estimated quantiles and the maximum observed characteristic value are non-decreasing functions. We seek a sudden increase in the characteristic value. In most examined stream gauges this increase is stable and subsequent values

of estimated quantiles do not exceed the interval of confidence resulting from the quantile preceding it. In only a few cases is the jump so large that it exceeds these boundaries. An example of such behaviour is the low flows at the station in Bogusław. In the merging of large low flows, at the 70% truncation level a rapid increase of both quantiles occurs – in Fig. 2 it is depicted by a red rectangle.

3.2. Extreme low flows

The course of estimated quantiles of the GEV distribution looks different. In Fig. 3-5 red rectangles mark the ranges of truncation level demonstrating visible instability. This also appears in the Kuripapango station, located in a natural drainage basin. This means that this type of insta-

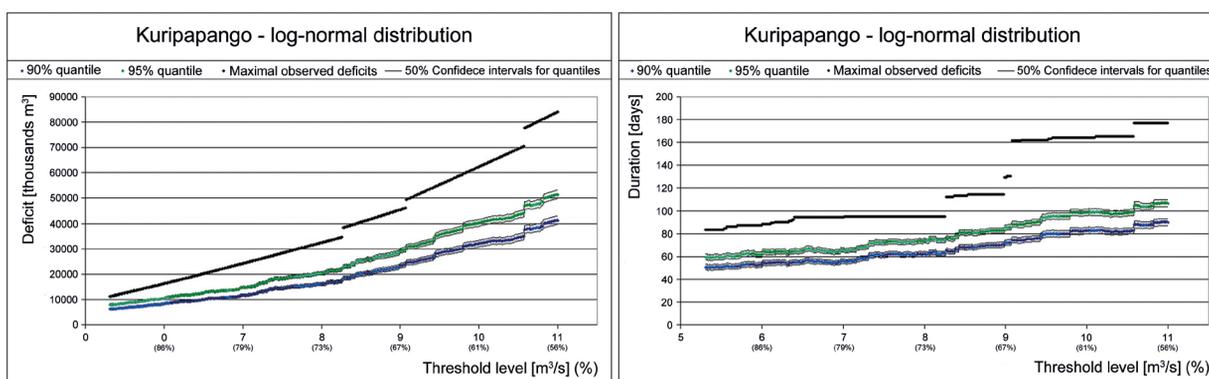


Fig. 1. Kuripapango stream gauge – dependence of estimated quantile of log-normal distribution on the truncation level defining low flow

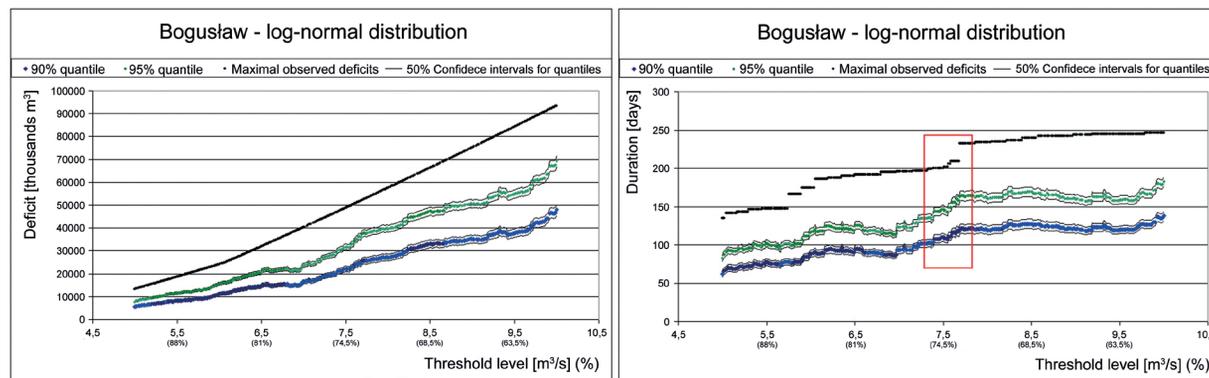


Fig. 2. Bogusław stream gauge – dependence of estimated quantile of log-normal distribution on the truncation level defining low flow

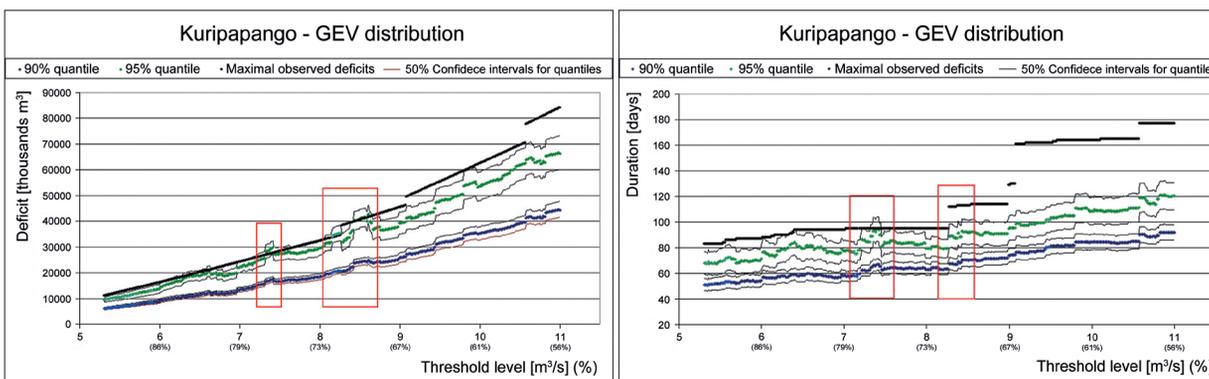


Fig. 3. Kuripapango stream gauge – dependence of the estimated GEV distribution quantile on the truncation level defining low flow

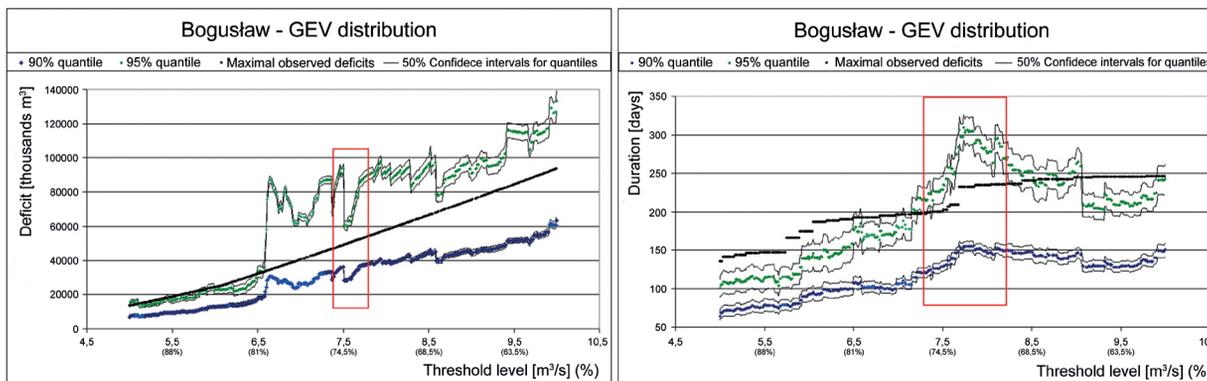


Fig. 4. Bogusław stream gauge – dependence of the estimated GEV distribution quantile on the truncation level defining low flow

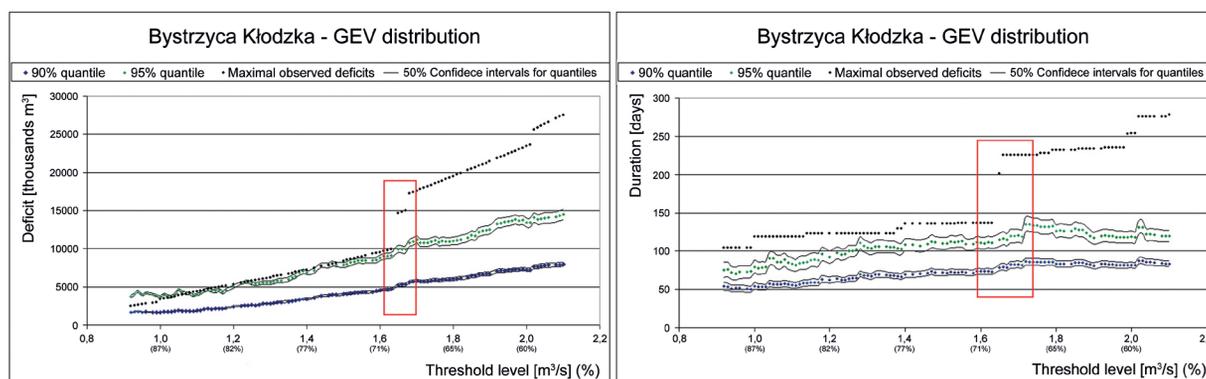


Fig. 5. Bystrzyca Kłodzka stream gauge – dependence of the estimated GEV distribution quantile on the truncation level defining low flow

bility is likely not related to the changes in a drainage basin resulting from its type of use. It is more likely a characteristic of the drainage basin. In drainage basins under agricultural use (Bogusław stream gauge in the Proсна basin) the variability of both duration and deficit estimators is very high. And, as in the case of the previous drainage basin, one does not see a significant cause for the jump in variability of the estimated quantile.

The course of the estimated quantiles in the last of the examined stream gauges – in Bystrzyca Kłodzka (Nysa Kłodzka drainage basin) – does not show such sudden changes. Only at a flow of 1,73 m³/s is a faster increase of the 95% quantile visible.

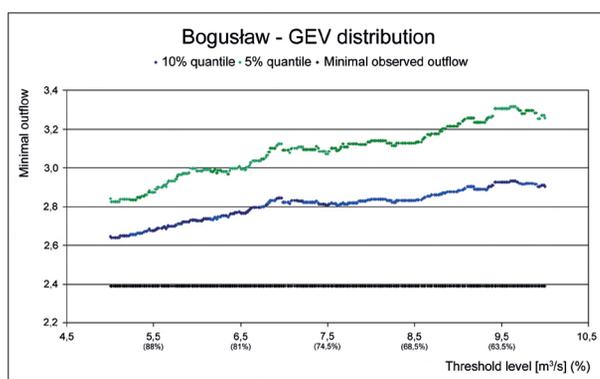


Fig. 6. Bogusław stream gauge – dependence of the estimated GEV distribution quantile on the truncation level defining low flow

3.3. Minimal low flow

For the estimation of the minimal low flow, a GEV distribution was used in the minimum version (2). In the majority of the examined stream gauge sections, a variation was marked in the estimated 5% and 10% quantiles, however it was not as pronounced as in the case of the other two characteristics. The graph below – quantiles of the minimal flow distribution in Bogusław – confirms this. For threshold flows above 7 m³/s the graphs of low flow duration and deficit quantiles show a significant variability. This does not carry over at all to the variability of the estimated quantiles of minimal flow in the low flow period.

4. Conclusions

The presented examples should demonstrate the emerging problem of estimating the tails of low flow duration and deficit distributions. Also, in many other researched sections there appear threshold flows Q_g which demonstrate significant instability. Such instability is related to the precision of estimating the tails of probability distributions of low flow duration and deficit – in other words, estimating the risks of deep, prolonged hydrological droughts. The only possibility of recognising instabili-

ty is the estimation of both distributions for several close threshold flows and comparison of the obtained results.

Bibliography

- Bonacci O., 1993, Hydrological identification of drought, *Hydrological Processes*, 7 (3), 249-262, DOI: 10.1002/hyp.3360070303
- Bower D., Hannah D.M., McGregor G.R., 2004, Techniques for assessing the climatic sensitivity of river flow regimes, *Hydrological Processes*, 18 (13), 2515-2543; DOI: 10.1002/hyp.1479
- Coles S., 2001, *An introduction to statistical modeling of extreme values*, Springer – Verlag London Limited, 209 pp.
- Gustard A., Demuth S. (eds.), 2009, *Manual on Low-flow: Estimation and Prediction*, Operational Hydrology Report No. 50, WMO-No. 1029, Koblenz
- Hisdal H., Tallaksen L.M., 2000, Drought event definition, Technical Report to the ARIDE Project 6, Department of Geophysics, University of Oslo, available at <http://www.hydrology.uni-freiburg.de/forsch/aride/navigation/publications/pdfs/aride-techrep6.pdf> (data access 26.10.2015)
- Hisdal H., 2002, Regional aspects of droughts, Faculty of Mathematics and Natural Sciences, University of Oslo, available at http://www.geo.uio.no/edc/downloads/phd_thesis_hisdal_2002.pdf (data access 26.10.2015)
- Hisdal H., Tallaksen L.M., Frigessi A., 2002, Handling non-extreme events in extreme value modelling of streamflow droughts, [in:] *FRIEND 2002 – Regional Hydrology: Bridging the Gap between Research and Practice*, Proceedings of the Fourth International FRIEND Conference held at Cape Town, South Africa, IAHS Publ. 274, 281-288, available at http://hydrologie.org/redbooks/a274/iahs_274_281.pdf (data access 26.10.2015)
- Jakubowski W., Rączuk L., 2004, Nizowka 2003 software, [in:] *Hydrological Drought*, 48, Processes and estimation methods for streamflow and groundwater, L.M. Tallaksen, H.A.J. van Lanen (eds.), Amsterdam
- Jakubowski W., 2011, Probability distributions in the estimation of hydrological drought (in Polish), *Monografie Uniwersytetu Przyrodniczego we Wrocławiu*, CXVI, 178 pp.
- Kaznowska E., 2011, Analysis of low flow characteristics and drought frequency in agricultural attachments, [in:] *Prediction and Reduction of Diffuse Pollution, Solid Emission and Extreme Flows from Rural Areas – case study of small agricultural catchment*, Wydawnictwo SGGW, 27-46, available at http://ziw.sggw.pl/zaklad/monogra/rozdzial_2.pdf (data access 26.10.2015)
- Ozga-Zielińska M., 1990, Droughts and floods – their definition and modeling (in Polish) *Przegląd Geofizyczny*, XXXV (1-2), 33-44
- Ozga-Zielińska M., Brzeziński J., 1997, *Applied Hydrology* (in Polish), PWN, Warszawa, 323 pp.
- Smakhtin V.U., 2001, Low flow hydrology: a review, *Journal of Hydrology*, 240 (3-4), 147-186, DOI: 10.1016/S0022-1694(00)00340-1
- Tallaksen L.M., 2000, Streamflow drought frequency analysis, [in:] *Drought and drought mitigation in Europe*, V.J. Vogt, F. Somma (eds.), *Advances in Natural and Technological Hazards Research*, 14, 103-117, DOI: 10.1007/978-94-015-9472-1_8
- Tallaksen L.M., van Lanen H.A.J. (eds.), 2004, *Hydrological Drought*, 48, Processes and estimation methods for streamflow and groundwater, Elsevier Science B.V., Amsterdam, 579 pp.
- Tokarczyk T., Jakubowski W., 2006, Temporal and spatial changeability of drought in mountain catchments of the Nysa Klodzka basin, [in:] *Climate Variability and Change – Hydrological Impacts*, Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, IAHS Publ. 308, 139-144, available at <http://iahs.info/uploads/dms/13650.29-139-144-91-308-Taka.pdf> (data access 30.10.2015)
- Yevjevich V., 1967, An objective approach to definition and investigations of continental hydrologic droughts, *Hydrology Papers*, 23, Colorado State University, Fort Collins, available at https://dspace.library.colostate.edu/bitstream/handle/10217/61303/HydrologyPapers_n23.pdf?sequence=1 (data access 30.10.2015)
- Zelenhasić E., Salvai A., 1987, A method of streamflow drought analysis, *Water Resources Research*, 23 (1), 156-168; DOI: 10.1029/WR023i001p00156