

Around and about an application of the GAMLSS package to non-stationary flood frequency analysis

S. E. Debele¹ · E. Bogdanowicz¹ · W. G. Strupczewski¹

Received: 22 March 2017 / Accepted: 12 August 2017 / Published online: 24 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract The non-stationarity of hydrologic processes due to climate change or human activities is challenging for the researchers and practitioners. However, the practical requirements for taking into account non-stationarity as a support in decision-making procedures exceed the up-to-date development of the theory and the of software. Currently, the most popular and freely available software package that allows for non-stationary statistical analysis is the GAMLSS (generalized additive models for location, scale and shape) package. GAMLSS has been used in a variety of fields. There are also several papers recommending GAMLSS in hydrological problems; however, there are still important issues which have not previously been discussed concerning mainly GAMLSS applicability not only for research and academic purposes, but also in a design practice. In this paper, we present a summary of our experiences in the implementation of GAMLSS to non-stationary flood frequency analysis, highlighting its advantages and pointing out weaknesses with regard to methodological and practical topics.

Keywords Non-stationary flood frequency analysis · GAMLSS · Monte Carlo Simulations

Introduction

All evaluations of the risks of extreme high river flows require methods to statistically estimate the upper quantiles on the basis of the observed data. Such methods are widely used in the development of regulations concerning the design of hydraulic structures, embankments, bridges and culverts, flood mitigation strategies and civil protection policies. Furthermore, it is widely recognized that estimation of changes in the frequency of floods related to climatic changes and analyzing trends in the high flow extremes are crucial for the adaptation and mitigation measures undertaken to cope with the negative impacts of climate change. For this purpose new methods taking into account non-stationarity of flood events are developed with a view to replace long-standing and well-established characteristics and principles of engineering design and to move from an equilibrium- or stationary-paradigm to one of a constant evolution that recognizes the dynamic nature of physical and socio-economic processes. Non-stationary statistical analysis can be performed on the observed annual/seasonal maxima series and on projections obtained from coupled hydrological and climate models. However, for methodical, computational and practical reasons this analysis is difficult and uncertain, which casts a shadow over both practical and theoretical efforts. This paper is a sequel to the Debele et al. (2017) paper where an application of the GAMLSS tools to the FFA was compared with two more traditional approaches. The results of the research presented in the latter indicate that the GAMLSS performs better than the other two methods applied from the point of view of its flexibility and a superior treatment of non-stationarity. The aim of this paper is to put the FFA analysis into the context of real-life applications and to

✉ S. E. Debele
s.debele@igf.edu.pl

E. Bogdanowicz
ewabgd@o2.pl

W. G. Strupczewski
wgs@igf.edu.pl

¹ Department of Hydrology and Hydrodynamics, Institute of Geophysics, Polish Academy of Sciences, Ksiecica Janusza 64, 01-452 Warsaw, Poland

discuss the pros and cons of different FFA approaches from the end-user perspective.

We describe briefly the history of flood frequency analysis (FFA) in “[Some historical aspects of flood frequency analysis](#)”. Some methodological problems and practical aspects of non-stationary flood frequency analysis (NFFA) are discussed in “[Non-stationary process of flood formation and NFFA](#)”. The GAMLSS package is described in “[Introduction to the GAMLSS package](#)” and its application to NFFA is presented in “[Hydrological Applications and Assessment of the GAMLSS Software](#)”. The relevant conclusions are presented in “[Conclusions](#)”.

Some historical aspects of flood frequency analysis

The flood frequency analysis arose from practical problem of how to ensure the safety of structures prone to high waters, and for how long. For ages constructors have coped with this problem; however, the methods they used have changed from trial-and-error methods to more and more sophisticated ones accompanying the growing potential of hydrological observations, mathematics and computing capabilities. Many artifacts from ancient times demonstrate the skills of their creators, both in terms of design and construction. The earliest indications of Dutch dike building date from the late Iron Age. In Roman times, dikes and dams were created on the territories of present France, Spain, Portugal, Syria and North African countries. One of the oldest functioning bridges in the world is the Pons Fabricius in Rome, Italy, which was built in 62 BC; the year after Cicero was a consul, to replace an earlier wooden bridge destroyed by fire. The oldest, the Caravan Bridge over the river Meles in Izmir, Turkey is dated c. 850 BC. Intact from antiquity, the bridges have been in continuous use ever since.

In many countries the beginning of the twentieth century was a period of introducing the statistical analysis of the maximum annual flows probability distribution as a basis for determining design characteristic in the form of the upper quantile corresponding to the specified return period—the expected life time of a structure. The return period is adopted according to the structure class but most commonly the 100-year return period is chosen and the corresponding quantile called a 100-year water (or flood) is thought to be a fair balance between protecting society and overly stringent regulation. Former procedures based on water level measurement defined usually the design characteristic as the highest observed water level plus 1 m. The design characteristics assessment from the probability distribution of the annual maximum flows entailed a number of well-known problems. As the true distribution remains unknown, the selection of an appropriate

probability distribution model and the estimation method of its parameters were, and still are, the most important. The limited length of hydrological series makes a discussion of the proper choice of statistical model unjustified; although some goodness-of-fit criteria can point to the best model but their discriminate power remains low. Moreover, when the series length increases (even only by 1 year) the best distribution model can be different from that previously chosen. As the candidate distributions can have different tail heaviness, the consequences to design quantiles are obvious. To avoid this situation and to enable control over the results, in many countries the type of annual maxima distribution, and the estimation method as well, were fixed and the parameters and the quantiles were updated, e.g., every 5 years. Essentially, the chosen distribution form stemmed from tradition and multiple tests of the regional conformity. The estimation methods evolved from simple graphical fitting based on plotting position and a comparison of the empirical and theoretical quantiles to the method of moments, linear moments and mathematically advanced optimization methods (e.g., maximum likelihood method). The inventory and analysis of the FFA models used in different countries at the end of the twentieth century were presented by Cunnane (1989). Many have reached the status of more or less obligatory guidelines and rules [e.g., in the USA—Bulletin No 15 (1969) with further extensions and updates, in the UK—Flood Studies Report (1975) succeeded by Flood Estimation Handbook and Flood Estimation Guidelines, in Poland—Regulations (1969, 2007), and many others].

The research on methods for flood frequency estimation under the various climatic and geographic conditions found in Europe, carried out by the COST FloodFreq project over 20 years later (Castellarin et al. 2012) revealed that nowadays the distribution types in use did not change significantly from Cunnane’s report. They are: Gamma, Pearson 3, Log Pearson 3, Gumbel, Lognormal, Generalized Logistic and Weibull 3 with the newly introduced Generalized Extreme Value distribution (GEV) on the top of the list, Generalized Pareto and Two-Component Extreme Value (TCEV) distributions. Currently, the seasonal approach is more often used in the case of different seasonal flood generating processes within a year (e.g., Guidelines for flood frequency analysis..., 2005; Strupczewski et al. 2009, 2011; Vormoor et al. 2015).

With the explosive development of computing technology the guidelines for the FFA started to be accompanied by software packages often prepared by hydrologists themselves. Later, statisticians and specialists in numerical methods and programming languages professionally developed the software and offered it on the market. In the early 1980s the HOMS (Hydrological Operational Multipurpose System) project was implemented by the World

Meteorological Organisation (WMO) to enable free exchange and transfer of technology in hydrology and water resources. This technology is usually in the form of descriptions of hydrological instruments, technical manuals or computer programs provided by the Hydrological Services of member countries of the WMO. The principle of the project is that the technology transferred is not only ready for use but also works reliably.

The “FFA statistics” used and developed by hydrologists as a separate branch of the statistics based on their own paradigms attracted the interest of good, professional statisticians who enriched its methodological range. Particularly noteworthy is the paper of Langbein (1949).

The R-project, an open source programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing, plays an important role of a platform of statistical software repository and exchange (R Core Team 2017, <https://www.r-project.org/>).

The GAMLSS (Rigby and Stasinopoulos 2005) software is implemented in a series of packages in R language available from CRAN—the R library (<https://www.r-project.org/>). The platform is of general statistical use and it is not dedicated to hydrological problems with their specific requirements and constraints and, in particular, it is not directly applicable to FFA. The methodology of FFA laboriously developed for years is currently facing the challenge of adopting GAMLSS solutions.

Non-stationary process of flood formation and NFFA

To carry out reliable statistical inference we need to make significant assumptions about relations between the observed and unobserved data. Among these, the major assumption is simplicity, consistency and uniformity of Nature which can be understood as a premise of stationarity of the processes generating hydrological extremes, at least in the time scale of human life. Nowadays, the symptoms of climate change are an increasingly prominent topic in scientific and public discourse. Many scientific bodies claim that instrumental data analysis provides evidence that climate is changing and that the process is accelerating. “Rivers are a product of the climate”, Vojekov wrote in his work “Climates of the globe and Russia in particular” (1884). Climate change must therefore result in changes of water resources and flood generation processes as a consequence. If so, the natural question to be answered is the persistence of these changes and how meaningful they can be.

Carbon dioxide, the main suspect in climate change, displays an exceptional persistence that renders its

warming nearly irreversible for more than 1000 years. The warming due to non-CO₂ greenhouse gases, although not irreversible, persists notably longer than the anthropogenic changes in the greenhouse gas concentrations themselves (Solomon et al. 2010). Are these changes already visible in the magnitudes and frequencies of flooding? And further, can the observed trends be extrapolated and to what extent?

Koutsoyiannis (2006) and Koutsoyiannis and Montanari (2007) give a clear answer:

- The fluctuations of the mean are a normal feature of stationary processes, which, together with the limited length of the series of observation may lead to the impression that the underlying process is non-stationary.
- A deterministic function representing trend is a function that can be produced only by deduction, independently of the data (a priori); e.g., by a model that could predict them.
- On the contrary, according to common practice, the “trends” and “shifts” in the means are inferred by induction based on the data (a posteriori).
- Therefore such fitted lines are not deterministic and do not represent non-stationarity.

We do not have reliable stochastic or deterministic models of the process which could simulate the climate drivers and their hydrological impacts with the desired accuracy.

Coupled hydrological and climate models running for various emission scenarios provide us with the projections of seasonal/annual maxima for the near (to 2050) and far (up to 2100) future.

But the problem arises: how can we trust the projections made for the time horizon of 50 years and more, given that the weather forecasts for a period longer than 3 days are not credible? Furthermore, the most uncertain factor in prediction by global climate models is the spatial–temporal distribution of rainfall which is usually the driving cause of floods and drought phenomena. So, a stationary model is sometimes preferable to a non-stationary one when the evolution in time of hydrological processes cannot be predicted reliably (Bayazit 2015; Serinaldi and Kilsby 2015; Lins and Cohn 2011).

Numerous sources of uncertainty influencing the projections of the temporal evolution of future changes (e.g., Serinaldi and Kilsby 2015; Koutsoyiannis 2013), however, do not intimidate us from analyses, hoping that they do give the best assessment of the future extreme events we are able to achieve nowadays. Enthusiasts of the idea of climate change and the possibility of its prediction announced to the world that stationarity is dead (e.g., Milly et al. 2008), and, although difficult, the non-stationary approach to water management has become mandatory.

However, it seems that the question is not whether to use these methods, but how to use them in design and planning practices considering their immense uncertainty and who will benefit (or maybe loose) in a result. Many projects on “managing flood risk to keep pace with climate change” declare tackling flood risk now for future generations.

It should be noted that a number of projects on the adaptation to climate change carried out under the banner of the benefit for future generations improved the much neglected so far domains, for which funding would be difficult to get under the less publicized slogan. Whether the stationarity is dead or not “playing this game with Nature”, we can only monitor the dynamics of our time series and improve climate and hydrological models. Designers can use past events and trends only as an indication of the severity of effects likely to occur in the future, acknowledging the fact that it is very easy to confuse statistical significance with practical or substantive importance. They should focus on the size of the expected effects instead of their statistical significance. The same statistically significant trends for large and small river can be practically important for the small but not important for the large river, where changes due to the trend are only a small fraction of the observed flow and lie completely within the range of accuracy of their assessments.

Taking into account the accuracy of flood flow data is another important problem ignored by the scientists. Development of the FFA and NFFA methods is focused on their mathematical formulation where there is no place for a typically hydrological reasoning. It is shown in the large number of papers published in hydrological journals which are strictly mathematical without reference to actual hydrological situation even in the form of case studies (e.g., Rasmussen 2001; Askhar and Mahdi 2003; Jawitz 2004, and many others).

Introduction to the GAMLSS package

Generalized additive models for location, scale and shape is a modern distribution-based approach to parametric and semi-parametric regression models, where the parameters of the assumed distribution for the response variable can be modeled as linear and/or non-linear and/or non-parametric smoothing functions of the explanatory variables (Rigby and Stasinopoulos 2005). GAMLSS was proposed by Rigby and Stasinopoulos (2005), Stasinopoulos and Rigby (2007) and Stasinopoulos et al. (2008) as a way of overcoming some of the limitations associated with generalized linear models (GLM) and generalized additive models (GAM) (Nelder and Wedderburn 1972; Hastie and Tibshirani 1992, respectively). In GAMLSS the exponential family distribution assumption for the response variable is

relaxed and replaced by a general distribution family called the GAMLSS family.

A GAMLSS model assumes that independent observations y_i of a random variable Y for $i = 1, 2, 3, \dots, n$ have probability distribution function $f_Y(y_i|\theta^i)$ with $\theta^i = (\theta_1^i, \dots, \theta_p^i)$ vector of p parameters accounting for location, scale and shape of the random variable distribution. The number of parameters p is limited to four, and guarantees enough flexibility of distributions in most applications. In GAMLSS explanatory variables are introduced into the model through the ‘predictors’ η_k related to parameters by monotonic link functions $g_k(\cdot)$ —identity, log, inverse function and others. The GAMLSS package involves several important sub-models, relating the distribution parameters to explanatory variables through an additive model (Eq. 1) or semi-parametric additive model in linear (Eq. 2) and non-linear form (Eq. 3).

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^m Z_{jk}\gamma_{jk} \quad (1)$$

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^m h_{jk}(x_{jk}) \quad (2)$$

$$g_k(\theta_k) = \eta_k = h_k(X_k\beta_k) + \sum_{j=1}^m h_{jk}(x_{jk}), \quad (3)$$

where θ_k and η_k are vectors of length n , X_k is a known design matrix (fixed effects design matrix) of order $n \times m$ (a matrix of explanatory variables, i.e., covariates), β_k is a parameter vector of length m and $\sum_{j=1}^m Z_{jk}\gamma_{jk}$ represent random effects term, $h_{jk}(\cdot)$ is an unknown function of explanatory variables x_{jk} (smoothing term) and $h_k(\cdot)$ is a known non-linear function. The dependence shown in Eq. 2 can be linear or smooth through smoothing terms, but this case will not be considered here. If $m = 0$ (additive terms removed), then models described by Eqs. 1–3 are reduced to the fully parametric form. For Eq. 2, for example, it reads:

$$g_k(\theta_k) = \eta_k = X_k\beta_k. \quad (4)$$

The meaning of θ_k depends on parameterization of the distribution. In the FFA mean/standard deviation parameterization θ_k is equal to the mean value and the standard deviation of the distribution assumed. Then the model in Eq. 4 simply reduces to VGLM (vector generalized linear models) case where $\theta_1 = E(Y)$ and $\theta_2 = \sqrt{\text{Var}(Y)}$ and both $g_k(\theta_k) = \eta_k$ are linear functions of the explanatory variables.

The list of distributions included in GAMLSS is impressive and covers over 90 types, and the open design of the system allows attaching others.

The estimation methods in GAMLSS are based on the maximum likelihood principle. Fully parametric models are estimated by the maximum likelihood method. For models with random effects and smoothing (Eqs. 2–4) the penalized likelihood is used. Two basic algorithms are used for maximizing the likelihood. The **CG algorithm** is a generalization of the Cole and Green (1992) algorithm and it uses the first and (expected or approximated) second and cross-derivatives of the log-likelihood function with respect to the distribution parameters. The limitation of CG is that it requires the existence of expected values of the cross-derivatives of the log-likelihood function with respect to location, scale and shape parameters. However, for many probability density functions $\{f_Y(y|\theta)\}$ the parameters are informationally orthogonal; since the expected values of the cross-derivatives of the log-likelihood function are zero. The CG algorithm performs better for distributions with potentially highly correlated parameters. In the case of zero cross-derivatives of likelihood function, the simpler **RS algorithm**, which is a generalization of the algorithm by Rigby and Stasinopoulos (1996a, b), is more suitable. The RS algorithm does not use the cross-derivatives and does not require accurate starting values for the parameters to ensure convergence; and it is faster for larger data sets.

The GAMLSS package gives the opportunity for fast-fitting of different models to the data set and the model selection can be carried out by checking the significance of the fitting improvement, e.g., between stationary and non-stationary model by means of deviance statistics; however, other methods are possible.

The GAMLSS framework of statistical modeling is implemented in a series of packages in R. The packages can be downloaded from the R library, CRAN (<https://www.r-project.org/>). There are more packages for stationary and non-stationary analysis in R; e.g., extRemes (Gilleland and Katz 2016), FAdist (Aucoin 2015), PearsonDS (Becker and Klößner 2017).

Hydrological applications and assessment of the GAMLSS software

As stated above, the list of distributions included in GAMLSS contains over 90 types. Truncated, censored, log and logit transformed and finite mixture versions of these distributions can be also used. However, a hydrologist can be confused since he/she cannot find there the names of the distributions commonly used in FFA, or they are two-parameter distributions while the three-parameter, lower-bounded distributions are mostly in operational use. Furthermore, the names and acronyms of the GAMLSS family distributions and their parameterization, as well, are

difficult for beginners and less experienced users of the package. Sometimes three-parameter FFA distributions are hidden in the generalized parameterization of GAMLSS. For instance, the Reverse Generalized Extreme Family distribution RGE is a re-parameterization of the three-parameter Weibull distribution. It is due to the fact that some parameterizations are computationally preferable to others in the sense that maximization of the likelihood function is easier.

Most GAMLSS-based research in hydrology addresses the application of two-parameter distributions from the GAMLSS family. Some GAMLSS applications in hydrological/climatological time-series are presented by Villarini et al. (2009a, b, 2010a, b, 2012), Machado et al. (2015), López and Francés (2013), Osorio and Galiano (2012), Hudson et al. (2008). GAMLSS have been used to model seasonal rainfall and temperature in Rome by Villarini et al. (2010a); they showed that the GAMLSS models could represent the magnitude and spread in the seasonal time series with parameters being a smooth function of time or teleconnection indices. GAMLSS models have been used for flood frequency analysis in Villarini et al. (2009a, b) and López and Francés (2013). The study by Machado et al. (2015) also applied GAMLSS to model flood data using historical information. Some studies applied GAMLSS to gridded datasets; e.g., Osorio and Galiano (2012) and Galiano et al. (2015) applied GAMLSS modeling framework to develop a methodology to account for non-stationarity existing in climate and hydrological processes and assessing the non-stationary spatial patterns of extreme droughts using gridded data of observed rainfall from regional climate models (RCMs). The summary of important applications of GAMLSS in NFFA is given in Table 1.

A very few hydrological studies have applied the three-parameter GAMLSS distributions, e.g., López and Francés (2013) and Zhang et al. (2015). In both studies, the Generalized Gamma (GG) distribution was applied to develop a framework for flood frequency analysis for annual maximum daily flows and maximum daily precipitation. It should be noted that although GG is a three-parameter distribution, its parameters are scale and two shape parameters. The distribution is defined for values $y > 0$, and does not have a lower bound being a location parameter. So, from the point of view of FFA applications it can be treated as a two-parameter distribution, more flexible due to the two shape parameters. As shown by Strupczewski et al. (2008) two-shape-parameter distributions in the stationary case are likely to have greater flexibility for fitting the right tail of empirical distribution than their lower-bounded parameter counterparts. The accuracy of the skewness coefficient estimation from short series (e.g., Wallis et al. 1974) is low. Therefore, it is commonly

Table 1 Summary of important applications of GAMLSS in non-stationary FFA

References	GAMLSS distribution	Number of distribution parameters
Villarini et al. (2009a, b)	Gumbel (GU); Weibull (WEI); Gamma (GA) and Lognormal (LOGNO)	2
Villarini et al. (2010a, b)	GU; WEI; GA; Logistic (LO) and LOGNO	2
Villarini et al. (2012)	Poisson (PO)	1
López and Francés (2013)	GU; LOGNO; WEI; GA	2
	Generalized Gamma (GG)	3
Machado et al. (2015)	LOGNO	2
Zhang et al. (2015)	GG	3

approved in the NFFA that the shape parameter remains constant. Allowing the skewness to vary in time seems unrealistic in terms of reliable estimation.

A review of the extensive available literature (Table 1) shows that the package and distributions selected from GAMLSS family can be used in the NFFA. It should be stressed that the applications described above are research-based. The reliability of the results obtained using two-parameter models should be evaluated in simulation experiments. None of the above noted papers addressed the practicality and efficiency of GAMLSS algorithm using Monte Carlo simulations.

In this topical issue of the *Acta Geophysica* we present a comparison of three non-stationary approaches to FFA (Debele et al., this issue) with time as a covariate one of them is the GAMLSS distributions family estimated by GAMLSS software. Using the possibility of introducing new distributions to the GAMLSS family we have added Pearson three distribution. The fitting works only with a CG algorithm, with and without covariates. However, our attempts to introduce other three-parameter distributions, namely Lognormal type 3 and GEV, were unsuccessful and the estimation procedures failed. The lower-bounded distributions are generally difficult to fit by the maximum likelihood method and the linear moments method (LMM) is recommended (e.g., Hosking and Wallis 1997; Strupczewski et al. 2001a, b; Markiewicz et al. 2010). In a comparison study we examined the performance of GAMLSS algorithm in the case of false distribution assumption ($T = \text{Lognormal } 3$ with imposed linear trends in mean and standard deviation, $H = \text{RGE}$). The relative bias (RB) and relative root mean square error (RRMSE) were computed for time-dependent moments and 99% quantiles. The RB and the RRMSE were calculated for 10,000 simulations with a sample size of $n = 50, 100$ and 200 generated by considered parent distribution. Compared to the weighted least squares (WLS), described in Strupczewski and Kaczmarek (2001), Kochanek et al. (2013) and Strupczewski et al. (2015), the algorithm GAMLSS showed better efficiency in the estimation of the trend in

the standard deviation and worse in the mean value. This finding also applies to the real flood data when sometimes the algorithm GAMLSS identified the trend in the mean as the opposite to that observed. The accuracy measures RB and RRMSE of the shape parameter were estimated best by GAMLSS. This result, together with the high efficiency in the evaluation of the trend in the standard deviation, has been translated into good estimates of the 99% quantile for the short and long series.

While applying GAMLSS to the NFFA a hydrologist can encounter some other problems. Perhaps the most important is the problem of the confidence intervals (CI) of time-dependent quantiles. The GAMLSS package allows us to evaluate CI by the delta method, likelihood or deviance profiles (Rigby et al. 2014), but in more complicated cases, which are typical in FFA, e.g., seasonal or multi-model approaches, their application is not obvious.

Conclusions

The non-stationarity of flood peak flows due to climate change creates numerous problems of both the theoretical and applied kind. In addition to describing the observed time series the hydrologist's goal is also to predict (extrapolate) future flows for the length of a planning horizon. The uncertainty of predictions derived from theoretical methods governs the results as well as the assumed distribution type. The inductive (data-based) methods of trend detection cannot provide proof of the non-stationarity of flood generation processes, nor for its persistence in the future. However, from a practical point of view, we are interested only in the long-term trends covering the expected life time of design structures (100 years or more for major structures), despite the fact that both accuracy of prediction and demands are decreasing with time.

Climate change and its hydrological impacts became a front page topic in the mass media and scientific journals and political issues creating a sense of danger at the individual, local and global scales. Only a positive trend

detected in the flood data can sustain this feeling, so only positive trends are “politically correct”. Hunters of trends in the time series of high water should be disappointed when a downward trend is detected, or if a positive trend identified as statistically significant turns out to be physically insignificant and therefore irrelevant to hydrological design.

Application of non-stationary methods in the practice of the FFA is conditioned by the availability of software that allows the assessment of the impact of changes in the flood generating processes on design characteristics, i.e. the upper quantiles. In this paper the possibilities of the most popular R software—the GAMLSS package are described with a view to its being applied in flood frequency analysis.

The GAMLSS package is the universal, flexible and complex statistical tool for different fields of application. However, its potential for stationary and non-stationary flood frequency analysis is limited by the distribution types included and the estimation method used. The GAMLSS family of distributions does not include three-parameter distributions with location parameter as the lower bound, the most frequently used in FFA. Only one distribution RGE, re-parametrized Weibull 3 parameter distribution, belongs to the GAMLSS family. Although the software is open to the inclusion of new distribution types, the estimation algorithms can fail. We were able to add only Pearson type 3 distribution to the GAMLSS family. The attempts with other distributions were unsuccessful. We presume that it is due to the maximum likelihood estimation method which is not the most suitable for this kind of distributions.

It is difficult to apply the GAMLSS package. The user should understand distributions and their properties and should make decisions regarding the distribution of the response variable, the choice of explanatory variables and the link functions, the amount of smoothing and random effects. The application of GAMLSS results in estimates of time-varying quantiles which are strongly distribution dependent, so the selection of the suitable distribution is of the first importance.

Looking at the many hydraulic structures built at a time when risk assessment methods were not as complex as they are today and hydrological data were sparse and irregular, we admire the knowledge of water and the environment, engineering skills and common sense of their creators. Perhaps the loss of this sense is the price we pay for technological development.

Acknowledgements This research has been done in the framework of Polish Norwegian Research Programme project CHIHE (PolNor/196243/80/2013), carried out by the Institute of Geophysics, Polish Academy of Sciences. We would like to gratefully acknowledge Robert A. Rigby and Dimitrios Mikis Stasinopoulos for their help in GAMLSS modeling framework.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Askhar F, Mahdi S (2003) Comparison of two fitting methods for the log-logistic distribution. *Water Resour Res* 39(8):7–8
- Aucoin F (2015) FAdist: distributions that are sometimes used in hydrology. R package version 2.2. <https://CRAN.R-project.org/package=FAdist>. Accessed 20 Mar 2016
- Bayazit M (2015) Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. *Environ Process* 2(3):527–542. doi:10.1007/s40710-015-0081-7
- Becker M, Klößner S (2017) PearsonDS: Pearson distribution system. R package version 1.0. <https://CRAN.R-project.org/package=PearsonDS>. Accessed 20 Feb 2017
- Bulletin No. 15 (1969) A uniform technique for determining flood flow frequencies, Hydrology Committee of Water Resources Council
- Castellarin A, Kohnova S, Gaal L, Fleig A, Salinas JL, Toumazis A, Kjeldsen TR, Macdonald N (2012) Review of applied-statistical methods for flood-frequency analysis in Europe, NERC/centre for ecology & hydrology, Wallingford. (ESSEM COST Action ES0901)
- Cole TJ, Green PJ (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 11:1305–1319. doi:10.1002/sim.4780111005
- Cunnane C (1989) Statistical distributions for flood frequency analysis, operational hydrol. Rep. No. 33 WMO-No. 718. World Meteorological Organization, Geneva
- Debele SE, Bogdanowicz E, Strupczewski WG (2017) A comparison of three approaches to non-stationary flood frequency analysis, *Acta Geoph.*, this issue, submitted for publication
- Flood Studies Report (1975) 5 Volumes + maps. Natural Environment Research Council, London
- Galiano SGG, Gimenez PO, Osorio JDG (2015) Assessing nonstationary spatial patterns of extreme droughts from long-term high-resolution observational dataset on a semiarid basin (Spain). *Water* 7(10):5458–5473. doi:10.3390/w7105458
- Gilleland E, Katz RW (2016) extRemes 2.0: an extreme value analysis package in R. *J Stat Softw* 72(8):1–39. doi:10.18637/jss.v072.i08
- Guidelines for flood frequency analysis long measurement series of river discharge (2005) WMO/HOMS Component I81.3.01. <http://www.wmo.int/pages/prog/hwrrp/homs/Components/English/i81301.htm>. Accessed Apr 2017
- Hastie TJ, Tibshirani RJ (1992) Generalized additive models, monographs on statistics and applied probability 43. Chapman & Hall/CRS, Boca Raton
- Hosking JRM, Wallis JR (1997) Regional frequency analysis, an approach based on L-moments. Cambridge University Press, New York
- Hudson IL, Rea A, Dalrymple ML, Eilers PHC (2008) Climate impacts on sudden infant death syndrome: a GAMLSS approach. In: Proceedings of the 23rd international workshop on statistical modelling. pp 277–280
- Jawitz JW (2004) Moments of truncated continuous univariate distribution. *Adv Water Resour* 27:269–281
- Kochanek K, Strupczewski WG, Bogdanowicz E, Feluch W, Markiewicz I (2013) Application of a hybrid approach in nonstationary flood frequency analysis—a Polish perspective.

- Nat Hazards Earth Syst Sci Discuss 1(5):6001–6024. doi:10.5194/nhessd-1-6001-2013
- Koutsoyiannis D (2006) Nonstationarity versus scaling in hydrology. *J Hydrol* 324:239–254
- Koutsoyiannis D (2013) Hydrology and change. *Hydrol Sci J* 58(6):1177–1197. doi:10.1080/02626667.2013.804626
- Koutsoyiannis D, Montanari A (2007) Statistical analysis of hydroclimatic time series: uncertainty and insights. *Water Resour Res* 43(5):W05429. doi:10.1029/2006WR005592
- Langbein WB (1949) Annual floods and the partial-duration series. *Trans Am Geophys Union* 30(6):879–881
- Lins HF, Cohn TA (2011) Stationarity: wanted dead or alive? *J Am Water Resour Assoc* 47(3):475–480. doi:10.1111/j.1752-1688.2011.00542.x
- López J, Francés F (2013) Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrol Earth Syst Sci* 17:3189–3203. doi:10.5194/hess-17-3189-2013
- Machado MJ, Botero BA, López J, Francés FA, Díez-Herrero BG (2015) Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrol Earth Syst Sci* 19:2561–2576. doi:10.5194/hess-19-2561-2015
- Markiewicz I, Strupczewski WG, Kochanek K (2010) On accuracy of upper quantiles estimation. *Hydrol Earth Syst Sci* 14:2167–2175. doi:10.5194/hess-14-2167-2010
- Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: whither water management? *Science* 319:573–574
- Nelder JA, Wedderburn RWM (1972) Generalized linear model. *J R Stat Soc Series A (General)* 135(3):370–384
- Osorio JDG, Galiano SGG (2012) Non-stationary analysis of dry spells in monsoon season of Senegal River Basin using data from Regional Climate Models (RCMs). *J Hydrol* 450–451:82–92. doi:10.1016/j.jhydrol.2012.05.029
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN: 3-900051-07-0
- Rasmussen PE (2001) Generalized probability weighted moments: application to the generalized Pareto distribution. *Water Resour Res* 17(6):1745–1751
- Regulation (2007) Ordinance of the Minister of the Environment of 20 April 2007 on the technical conditions to be met by hydrotechnical structures and their location. *Journal of Laws No. 86 of 2007, item 57 (in Polish)*
- Regulations for computation of the greatest annual discharges for given probability of occurrence to design engineering structures and technical equipment for water management in the field of hydraulic engineering (1969) Central Office of Water Management, Warsaw (in Polish)
- Rigby RA, Stasinopoulos DM (1996a) A semi-parametric additive model for variance heterogeneity. *Statist Comput* 6:57–65
- Rigby RA, Stasinopoulos DM (1996b) Mean and dispersion additive models. In: Hardle W, Schimek MG (eds) *Statistical theory and computational aspects of smoothing*. Physica-Verlag, Heidelberg, pp 215–230
- Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Appl Stat* 54:507–554. doi:10.1111/j.1467-9876.2005.00510
- Rigby RA, Stasinopoulos DM, Heller G, Voudouris V (2014) The distribution toolbox of GAMLSS. <http://www.gamlss.org/wp-content/uploads/2014/10/distributions.pdf>. Accessed 14 Apr 2016
- Serinaldi F, Kilsby CG (2015) Stationarity is undead: uncertainty dominates the distribution of extremes. *Adv Water Resour* 77:17–36. doi:10.1016/j.advwatres.2014.12.013
- Solomon S, Daniela JS, Todd JS, Murphy DM, Plattner G-K, Knutti R, Friedlingstein P (2010) Persistence of climate changes due to a range of greenhouse gases. *PNAS* 107(43):18354–18359. doi:10.1073/pnas.1006282107
- Stasinopoulos DM, Rigby RA (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw* 23(7):1–46
- Stasinopoulos DM, Rigby RA, Akantziliotou C (2008) Introductions on how to use the package in R, 2nd edn. <http://www.gamlss.org/wp-content/uploads/2013/01/gamlss-manual.pdf>. Accessed 16 Mar 2016
- Strupczewski WG, Kaczmarek Z (2001) Non-stationary approach to at-site flood frequency modelling. Part II. Weighted least squares estimation. *J Hydrol* 248(1–4):143–151. doi:10.1016/S0022-1694(01)00398-5
- Strupczewski WG, Singh VP, Feluch W (2001a) Non-stationary approach to at-site flood frequency modelling. Part I. Maximum likelihood estimation. *J Hydrol* 248(1–4):123–142. doi:10.1016/S0022-1694(01)00397-3
- Strupczewski WG, Singh VP, Mitosek HT (2001b) Non-stationary approach to at-site flood frequency modelling. Part III. Flood analysis of Polish rivers. *J Hydrol* 248(1–4):152–167. doi:10.1016/S0022-1694(01)00399-7
- Strupczewski WG, Markiewicz I, Kochanek K, Singh VP (2008) Short walk into two-shape parameter flood frequency distributions. In: VP Singh (ed) *Hydrology and hydraulics*. Water Resources Publications, Littleton, pp 669–716
- Strupczewski WG, Kochanek K, Feluch W, Bogdanowicz E, Singh VP (2009) On seasonal approach to nonstationary flood frequency analysis. *Phys Chem Earth* 34(10):612–618
- Strupczewski WG, Kochanek K, Bogdanowicz E, Markiewicz I (2011) On seasonal approach to flood frequency modelling, Part I: flood frequency analysis of Polish rivers. *Hydrol Process* 26:705–716. doi:10.1002/hyp.8179
- Strupczewski WG, Kochanek K, Bogdanowicz E, Markiewicz I, Feluch W (2015) Comparison of two nonstationary flood frequency analysis methods within the context of the variable regime in the representative polish rivers. *Acta Geoph* 64(1):206–236. doi:10.1515/acgeo-2015-0070
- Villarini G, Serinaldi F, Smith JA, Krajewski WF (2009a) On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resour Res* 45:1–17
- Villarini G, Smith JA, Serinaldi F, Bales J, Bates PD, Krajewski WF (2009b) Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Adv Water Resour* 32:1255–1266. doi:10.1029/2008WR007645
- Villarini G, Vecchi GA, Smith JA (2010a) Modeling the dependence of tropical storm counts in the North Atlantic basin on climate indices. *Mon Weather Rev* 138:2681–2705
- Villarini G, Smith JA, Napolitano F (2010b) Nonstationary modelling of a long record of rainfall and temperature over Rome. *Adv Water Resour* 33:1256–1267
- Villarini G, Smith JA, Serinaldi F, Ntelekos AA, Schwarz U (2012) Analyses of extreme flooding in Austria over the period 1951–2006. *Int J Climatol* 32:1178–1192. doi:10.1002/joc.2331
- Vojekov AD (1884) *Climates of the globe and Russia in particular*, St Petersburg
- Vormoor K, Lawrence D, Heistermann M, Bronstert A (2015) Climate change impacts on the seasonality and generation processes of floods—projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. *Hydrol Earth Syst Sci* 19:913–931. doi:10.5194/hess-19-913-2015
- Wallis JR, Matalas NC, Slack JR (1974) Just a moment! *Water Resour Res* 10(2):211–219. doi:10.1029/WR010i002p00211
- Zhang D, Yan D, Wang YC, Lu F, Liu S (2015) GAMLSS-based nonstationary modeling of extreme precipitation in Beijing–Tianjin–Hebei region of China. *Nat Hazards* 77(2):1037–1053. doi:10.1007/s11069-015-1638-5