Inundation risk for embanked rivers.

W. G. Strupczewski[1], K. Kochanek[1], E. Bogdanowicz[2] and I. Markiewicz[1]

[1] *Institute of Geophysics, Polish Academy of Sciences. Ksiecia Janusza 64, 01-452 Warsaw, Poland (wgs@igf.edu.pl, kochanek@igf.edu.pl, iwonamar@igf.edu.pl)*

[2] *Institute of Meteorology and Water Management. Podlesna 61; 01-673 Warsaw, Poland (ewa.bogdanowicz@imgw.pl)*

ABSTRACT

The Flood Frequency Analysis (FFA) concentrates on probability distribution of peak flows of flood hydrographs. However, examination of floods that haunted and devastated the large parts of Poland lead us to revision of the views on the assessment of flood risk of Polish rivers. It turned out that flooding is caused not only by overflow of the levees' crest but mostly due to the prolonged exposure to high water on levees structure causing dangerous leaks and breaches that threaten their total destruction. This is because, the levees are weakened by long-lasting water pressure and as a matter of fact their damage usually occurs after the culmination has passed the affected location. The probability of inundation is the total of probabilities of exceeding embankment crest by flood peak and the probability of washout of levees Therefore, in addition to the maximum flow one should consider also the duration of high waters in a river channel.

In the paper the new two-component model of flood dynamics: 'Duration of high waters–Discharge Threshold–Probability of non-exceedance' (DqF), with the methodology of its parameters estimation was proposed as a completion to the classical FFA methods. Such model can estimate the duration of stages (flows) of an assumed magnitude with a given probability of exceedance. The model combined with the technical evaluation of probability of levees breach due to the *d*-days duration of flow above alarm stage gives the annual probability of inundation caused by the embankment breaking.

The results of theoretical investigation were illustrated by a practical example of the model implementation to the series of daily flow of the Vistula River at Szczucin. Regardless promising results, the method of risk assessment due to prolonged exposure of levees to high water is still in its infancy despite its great cognitive potential and practical importance. Therefore, we would like to point out the need for and usefulness of the DqF model as complementary to the analysis of the flood peak flows, as in classical FFA. The presented two-component model combined with the routine flood frequency model constitutes a new direction in FFA for embanked rivers

Keywords: inundation risk, embanked rivers, modelling of high waters duration, annual flow peaks, levee leakage

1 INTRODUCTION

The most popular way of flood protection in Poland is the embankment of the rivers. In consequence of this passive way of protection, floods in Poland occur mostly due to the levee breach or to flow over the crest of dikes. Sense of security in floodplains of embanked rivers results from the belief that levees protect against the flood magnitude for which they were designed. So it creates the illusion that if the actual forecasted flood peak does not exceed the safety levels related to levee's designed value one can assume that the risk of water overtopping the dike crest is negligible and so is the risk of flooding in the protected area. The records of floods in Poland show that this is not true; more often the floods are the result of the prolonged exposure to high water on levees. The levees are weakened by water and their disruption occurs when it seems that the danger is over, so after passing culmination. This is particularly dangerous because when the staff responsible for flood protection and local residents breathe sigh of relief the worst is yet to come.

Therefore, apart from the magnitude of the peak flows another important factor should be taken into consideration, the duration of high water levels, in fact, a parameter of the wave's shape. Long-lasting high stages may weaken the levees' structure (soaking) and cause dangerous leaks, blurs and breaks that threaten their destruction. That is why the classical Flood Frequency Analysis (FFA) concerning only the frequency of the annual maximum (AM) flows is not suitable in this case and ought to be supplemented by the analysis of the duration of flows over the given threshold (Bogdanowicz *et al.*, 2011, also Eagleson, 1972; Sivapalan *et al*., 1990; Gioia *et al*., 2008; Iacobellis *et al*., 2011). The joint risk of inundation making allowance for the

two main sources of vulnerability to flood hazard for areas protected by embankments, over-crest flow and levees failure, has been proposed and defined.

In Poland, as in many other countries for each hydrological station two benchmark water levels, called the warning stage and the alarm stage, have been specified. Although warning and alarm stages are assigned to the places where water levels are observed, to the hydrological stations, their determination procedures as well as other inundation risk characteristics take into account, *inter alia*, the elevation of the embankment system for the whole river reach. So, the results of below analysis refer to the river reaches represented by data observed at hydrological stations. The frequency of annual maximum uninterrupted duration, D (in days), of flows over the flood alarm stage (Fig. 1) can be used to assess the risk of flooding due to waning of the levees' strength. The aim of this study is to introduce formal aspects of the Duration–flow–Frequency (DqF) modelling in stationary and non-stationary conditions, to use it to assess the inundation risk due to the levees breach and to combine it with the AM flow model to get the cumulative probability of inundation. In the presented statistical model, the duration is considered as a random variable while the alarm flow discharge is the fixed value. The approach presented here for non-stationary conditions can to some extent resemble the Peak Over Threshold (POT) with covariates techniques developed by Davison and Smith (1990). Looking for similarities to other approaches used in hydrology one can find that the likelihood function of DqF model is for stationary conditions similar to the likelihood function of the censored sample introduced to FFA by Kaczmarek (1977).

The paper is built as follows: in the second section the concept of the inundation risk for embanked river is defined. Then a short review of literature on statistical modelling of flood shape hydrographs with emphasis on one-dimensional models is presented (Section 3). In the next section the Duration–Flow discharge–Frequency (DqF) model is introduced and estimations of its parameter for stationary and non-stationary case are described and discussed. Taking into account the embankment resistance, the annual probability of inundation caused by levees breaching is introduced. To illustrate the proposed way of inundation risk assessment the case study for the Szczucin gauging station at the Vistula River (Southern Poland) is presented (Section 5). The probability of inundation due to levees breaching is compared with the conventional probability of peak flow exceeding the levee crest and the cumulative probability of inundation are computed. The section 6 concludes the paper.

Fig. 1 Definition of the threshold flow discharge and duration in DqF model:
a) the flood wave of d_t duration entirely in the year t;
b) the flood wave starts in the year t and continues in $t + 1$.

2 FLOOD-RISK

Floods occur as a result of water spilling over the crest of embankment ($Q > Q_B$) or more often as a result of prolong existence of high water in the embanked river channel, so when the peak flow discharge exceeds the alarm flow (Q_A) but is lower than the overtopping flow (Q_B is the discharge that overtops levee crests) ($Q_{AQ} < Q_{\max} < Q_B$). One can also distinguish many other causes of floods, such as back water and ice-jams, etc., but they do not stem from the embankment failures and will not be considered in this study.

The annual probability of inundation for embanked river reach is expressed as the total of probability of the two exclusive events (*Fl* stands for 'flood') (see Fig. 2):

$$P(Fl) = P_1(Fl) + P_2(Fl) \tag{1}$$

where the first term comes from the conventional FFA

$$P_1(Fl) = p(Q_{\max} > Q_B) \tag{2}$$

The second term of Eq.(1) defines the probability of inundation caused by levees breaching which depends on both the flood persistency and levees resistance to high water stages which in turns depends on their design and technical condition. Therefore, the $P_2(Fl)$ is expressed as the integral of the product of the value of the hazard index $h(Fl|d)$ which is defined as the probability density of levee breaching caused by the d-days duration of flow over the flow level Q_A and of the pdf of the the d – duration, so $f(d)$ for annual peak flows in the interval $Q_A < Q_{\max}(t) < Q_B$.

$$P_2(Fl) = p\left(Fl \middle| \left(Q_A < Q_{\max} \le Q_B\right)\right) = \int_{0^+}^{\infty} h\left(Fl \middle| d\right) \cdot f(d) \cdot \mathrm{d}d \tag{3}$$

where

$f(d)$ – pdf of the duration d of flows above the alarm stage;

$h(Fl|d)$ – the hazard index being the probability of levee breaching caused by a high water of the duration d.

The value of the hazard index $h(Fl|d)$ tends to 0 for d going to 0 and to 1 for d going to infinity (e.g. Fig. 6). The hazard index $h(Fl|d)$ is determined administratively for the river reach by the Regional Water Management Board based on the technical assessment of flood embankments.

Fig. 2 Two reasons of inundation – an illustration.

Note that collating the annual maximum high flow duration data for analysis one puts $d_t = 0$ (index t marks the t-th year in a series in which the particular event $d = 0$ occurred, $t = 1, 2,\ldots, T$ and T is the length of the series in years) the both for $Q_{max} \leq Q_A$ and $Q_{max} > Q_B$, so 1 inundation yearly is considered and that caused by spilling over crest has the priority over one caused by prolonged high stages. Furthermore note that the weaker is the relationship between annual maximal values of peak flow and duration of flows above the alarm flow (Q_A) the more justified is the separate analysis of the both random variables. The DqF approach is the extension of the conventional FFA performed on a single annual peak flow series. But, even though it does not have to be the same flood that gives the annual flood peak and last longest in the year, the annual peak flows are usually assumed to be temporary independent what has been verified by several investigators and so is assumed here for the annual maximal durations. Due to the poor measurement material – short samples – it would be eve hard to analyse autocorrelations in the series of durations.

The ratio of probabilities P_2 to P_1 and their total is helpful to determine the actions to reduce the risk of flooding, namely the strengthening or heighten the levees (or building parallel levees).

3 THE STATISTICAL MODELLING OF FLOOD HYDROGRAPHS SHAPE

Due to complexity of stochastic nature of river flow process one has to accept a rational ignorance while dealing with flood risk management. In response to practical needs several simple conceptual structures are being developed for statistical modelling of flood hydrographs. The methods of constructing design flood hydrographs are most popular for modelling flood hydrographs. Their reviews is available in e.g. Serinaldi and Grimaldi (2010), Strupczewski (1964, 1966) and Strupczewski *et al*. (2013). The design hydrograph $Q(t)$ with the defined return period of its peak serves both in flood–risk mapping procedures and for designing a reservoir storage capacity and other hydraulic structures sensitive for flood hydrograph magnitude and shape.

The common feature of most of the approaches to flood hydrographs analysis is an avoidance of using a joint probability distribution of parameters describing the shape of the hydrographs while limiting multi-dimensional analysis to conditional expectations further reduced to a regression. The most commonly used variables are flood peak and flood volume.

Extension of the standard FFA for statistical analysis of peak part of flood hydrographs is the one-dimensional model Flow-duration Frequency (QdF) initiated by NERC (1975) and Askhar (1980). In the nineties, Sherwood (1994), Balocki and Burgess (1994). Galea and Prudhomme (1997) laid out the foundations of the present form of the QdF method. Based on the assumption of the convergence of different flood distributions for small return periods Javelle *et al*. (1999), Javelle (2001) introduced a converging approach to the QdF modeling. Here the annual mean maximum peak flood volume (or equivalently the mean excess discharge – $\overline{Q}_d$) corresponding to the given duration (d) is taken (Fig. 3a) as the random variable. Therefore consequently the maximum d-days annual outflow volume $V_d = d \cdot \overline{Q}_d$ is the random variable as well. In fact, the above idea of flood peaks analysis is modelled on the analyses of the Intensity-duration-Frequency (IdF) commonly used for stochastic modelling of high intensity rainfalls and of the QdF analysis of low flows.

To cater for the conventional FFA, the flow discharge (Q_A) corresponding to the alarm stage (H_A) is used here, so the upper limb of the rating curve is regarded as time invariant. The frequency of annual maximum uninterrupted duration of flows, D (in hours, days, etc.), over the flood alarm stage (H_A) (or equivalently over the alarm flow (Q_A)) but excluding floods pouring over the embankment crest (which corresponds to flows exceeding the overtopping flow Q_B) serves to assess the inundation risk of flood spilling out of river channel caused by scouring the levees (Fig. 2). Therefore, the $d_t = 0$ in the [d] time-series means that the threshold discharge, Q_A, has not been exceeded during the t-th year of the series ($Q_{max}(t) < Q_A$)

or that the peak flow has exceeded the overtopping flow ($Q_{max}(t) > Q_B$) where $Q_{max}(t)$ denotes the annual maximum discharge occurred in the t-th year of the sample series. In other words, there is no risk of the dike's damaging due to the prolonged exposure to the high water because the flood wave was either too small to reach the weaken construction of the levee or, the contrary, the flood is such big and sudden that the water immediately overtops the levee's crest. Note that if more than one flood appears in a year, the D and the annual peak flow (Q_{max}) can correspond to different floods (Fig. 1).

Using multi-duration approach, by fitting the appropriate statistical distribution to the extracted samples for various durations, from the relations QdF for various d one can roughly construct the scaled Flood-duration-Frequency curve (QdF). To avoid inconsistency of the estimates of quantile $Q(d, F)$ for various d, the same distribution function is applied for all duration (Javelle, *et al.*, 1999, Castellarin et al., 2004; Iacobellis 2008; Botter et al. 2008) and the quantiles are reduced by the appropriate function $\varphi(d, \boldsymbol{\nu})$ which is decreasing function of d:

$$Q(d,F) = \varphi(d,\boldsymbol{\nu}) \cdot Q(0,F) \text{ for } d = 0,1,2,.. \; ; \; \varphi(0) = 1 \tag{4}$$

where the $\boldsymbol{\nu}$ denotes the vector of parameters which are estimated from the data.

It means that differences in the distributions of various d values result from the differences in the mean value only. Note that $Q(0, F)$ corresponds to the distribution of annual instantaneous peak discharges. The parameters of the function $\varphi(d, \boldsymbol{\nu})$ and $Q(0, F)$ [Eq.(5)] are estimated separately.

Finding that flood persistence is a factor of flood hazard for embanked rivers, Bogdanowicz, *et al.* (2008) modified the above model redefining Q as the annual maximum flow discharge (Q_d) which is continuously exceeded during the period d, wherein the d variable is still treated as a deterministic value (Fig. 3b). The applied way of determining the scaled distribution function does not differ much from the method described by Javelle *et al.*(1999). In parallel, the use of ML method in the presence of the d as the covariate (Strupczewski *et al.* 2001abc, Katz *et al.*, 2002, Stasinopoulos and Rigby, 2007, Stasinopoulos *et al.*, 2008, 2012) is demonstrated for Weibull distribution with the lower bound parameter and the constant shape parameter. Here all parameters are estimated jointly.

However to address the 1-D statistical analysis of the peak part of flood hydrographs directly to the problem of softening and breaching of river embankment, the duration (d) of high stages should be taken as a random variable rather than the mean excess discharge $\overline{Q}_d$ (Javelle, 2001) (Fig. 3a) or the the annual maximum flow discharge (Q_d) (Fig. 3b) (Bogdanowicz *et al.*, 2008). Note that the duration of flood (d) can be more accurate assessed than the peak flow discharge of large floods.

Fig. 3 Definition of the random variables in the QdF models:
a) the mean maximum d-days flow,
b) the annual maximum flow discharge (Qd) continuously exceeded during the period d.

4 FORMAL ASPECTS OF THE DURATION–FLOW–FREQUENCY MODELLING

To address the flood risks arising from softening and washing out the river embankments, Bogdanowicz *et al.* (2011) proposed to take as the subject of analysis the frequency of annual maximum uninterrupted duration, D (in days), of flows over the flood alarm stage (Q_A), the duration (D) is considered as a random variable while the alarm flow discharge (Q_A) is the fixed value (Fig. 1).

The time-series of annual maximum uninterrupted duration, D (in days), of flows over the flood alarm flow Q_A, $\mathbf{d} = (d_1, d_2, \ldots, d_t, \ldots, d_T)$, is the subject of statistical modelling in stationary and non-stationary conditions. The $d_t = 0$, denotes that the Q_A has not been exceeded during the t-th year ($Q_{max}(t) < Q_A$) or that the peak flow has exceeded the overtopping flow ($Q_{max}(t) \geq Q_B$), which means that the priority of overtopping over breaching is given and we rule out the possibility of two inundation floods of the two different origins within one year. Note that the condition $Q_{max}(t) \geq Q_B$ is equivalent to the unconditional inundation, from Eq.(2) $P_1(Fl|Q_{max}(t) \geq Q_B) = 1$, while $Q_B > Q(t) \geq Q_A$ points only possible inundation [see Eq.(3)].

Frequency analyses of hydrological sample with zero discrete values have received relatively little attention. Still there are several approaches for analysis of censored data, including probability plot, regression, weighted-moment estimators, maximum likelihood estimators, and conditional probability analyses (Gilliom and Helsel, 1986; Hass and Scheff, 1990; Harlow, 1989; Helsel, 1990). A consistent approach to the frequency analysis of such data requires using discontinuous probability distribution functions. Jennings and Benson (1969), Interagency Advisory Committee on Water Data (1982), Woo and

Wu (1989), Wang and Singh (1995) among others developed empirical three-parameter models for frequency analysis of hydrologic data containing zero values.

When the available data represent mean daily discharge, the d values are in fact the integer numbers (the exposition can last 1, 2, 3, etc. days) but to maintain the continuity of time we treat them as real numbers and consider d as if it corresponded to the duration range (d – 0.5 day, d + 0.5 day). In particular, for d = 0 (beginning of the time axis) the interval corresponds to the range (0, d + 0.5 day). If a flood starts before the end of a year and is continuing to the next year, the d value is derived for the entire flood wave (from its beginning in one year to its end in the next year) but attributed to the year t when the flood culmination occurred. To get an insight into flood persistence properties, the several threshold stages (Q_T) are considered but not only the alarm stage Q_A.

4.1 Stationary conditions

As far as the probability theory is concerned, the occurrence of zero events can be expressed by placing a non-zero probability mass on a zero value: $P(D = 0) \neq 0$, where D is the random variable, and P is the probability mass (e.g. Strupczewski *et al.*, 2002, 2003, Weglarczyk *et al.*, 2005). Therefore, the parent distribution functions of such hydrologic series would be discontinuous (with discontinuity at 0) and, using the theorem of total probability, their forms can be written as:

$$f(d) = \beta\,\delta(d) + (1-\beta)\,f^{o}(d;\mathbf{g})\cdot 1(d) \tag{5}$$

where β denotes the probability of the zero event, $\beta = P(D = 0)$, $f^{o}(d; \mathbf{g})$ is the conditional probability density function (CPDF), $f^{o}(d; \mathbf{g}) \equiv f(d|D > 0)$, which is continuous in the range $(0, +\infty)$ with a lower bound of 0, and $\mathbf{g}$ is the vector of parameters (containing β or not), $\delta(d)$ is the Dirac's delta function and $1(d)$ is the unit step function. Assuming the infinite upper bound for D seems acceptable and facilitates modelling. Due to discretisied duration d intervals, the probability of exceeding the Q_A flow during one day only equals to $P(d) = \int_{d-1/2}^{d+1/2} f(d)\cdot \mathrm{d}\,d$.

Hydrological samples with zero values are most frequently of exponential-like shape. Weglarczyk *et al.* (2005) modeled CPDF $f^{o}(d; \mathbf{g})$ of (5) by two-parameter distributions, namely by Generalized Pareto, Weibull and Gamma, estimating parameters by the maximum likelihood (ML) and the moments (MOM) methods.

4.1.1 Estimation of the weight parameter β

i) **From the pdf of the duration d [Eq.(5)] and the records $\mathbf{d} = (d_1, d_2, \ldots, d_t, \ldots, d_T)$ for given alarm flow Q_A**

From Eq.(5) one can write the likelihood function as:

$$L = \beta^{n_1}\cdot(1-\beta)^{n_2}\prod_{j=1}^{n_2} f^{o}(d_j;\mathbf{g}) \tag{6}$$

where n_1 and n_2 denote the number of zeros and non-zeros values, respectively.
If $\beta \notin \mathbf{g}$, from ML-equations:

$$\frac{\partial \ln L}{\partial \beta} = \frac{n_1}{\beta} - \frac{n_2}{(1-\beta)} = 0 \tag{7}$$

one can easily find that the ML-estimate of β is

$$\hat{\beta} = \frac{n_1}{n_1 + n_2} \tag{8}$$

so β and $\mathbf{g}$ are estimated by MLM independently**.**

ii) **From CDF of annual maximum floods obtained from FFA**

The better estimate of the β parameter in the sense of definition [Eq.(9)], not its standard error, can be obtained from the CDF of annual peaks providing the selected for Annual Maxima (AM) model fits well upper tail data. Note that the D = 0, denotes that the Q_A has not been exceeded during the t-th year ($Q_{\max}(t) < Q_A$) or that the peak flow has exceeded the overtopping flow ($Q_{\max}(t) > Q_B$) where $Q_{\max}$ denotes the annual maximum discharge, therefore, probability of zero value of D

$$\hat{P}(D=0)=\hat{P}(Q_{\max}<Q_A)+\hat{P}(Q_{\max}>Q_B)=\hat{\beta} \tag{9}$$

should be estimated from CDF of annual peak flows got from FFA rather than from the (0, 1) time series of the **d** record. Having derived from FFA the CDF of the annual peaks $\hat{G}(Q_{\max})\equiv\phi\left(Q_{max},\hat{\mathbf{h}}\right)$ where $\hat{\mathbf{h}}$ is the vector of parameter estimates, one can get the estimate of β as

$$\hat{\beta}=\hat{G}(Q_{max}=Q_A)+\left(1-\hat{G}(Q_{\max}=Q_B)\right). \tag{9a}$$

Note that if more than one flood appears in a year it may happen that the d_t and the annual peak flow $Q_{max(}(t)$ correspond to different floods.

Floods in excess of Q_B are unique in Polish rivers, but if they were they should be in the FFA treated as of unknown magnitude over the thereshold Q_B, thus one deals with first order right censored sample.

4.1.2 Estimation of parameters of the continuous part of Eq.(5)

ML estimate of the parameters (**g**) of the continuous part of PDF [Eq.(6)]: the conditional probability density function (CPDF). $f^{\circ}(d;\mathbf{g})\equiv f(d|D>0)$ of $f^{\circ}(d;\mathbf{g})$, can be obtained by solving the ML system of equations:

$$\frac{\partial \ln L}{\partial \mathbf{g}}=\frac{\partial}{\partial \mathbf{g}}\sum_{j=1}^{n_2}\ln f^{\circ}(d_j;\mathbf{g})=0 \text{ for } \beta\notin\mathbf{g} \tag{10}$$

Since the *d*-samples deprived of the zero values are most frequently of exponential-like shape, the distribution functions in Table 1 are recommended as candidates for $f^{\circ}(d_j;\mathbf{g})$ model.

Table 1. Distribution functions recommended as $f^{\circ}(d_j;\mathbf{g})$ model.

Distribution name	Probability density function	Parameters	Equation nr:
Exponential (Ex)	$f^{\circ}(d;\alpha)=\frac{1}{\alpha}\exp(-d/\alpha)$	α – scale	(11)
Weibull (We) distribution	$f^{\circ}(d;\alpha,b)=\frac{b}{\alpha}\left(\frac{d}{\alpha}\right)^{b-1}\exp(-d/\alpha)$	α – scale $b>0$ – shape	(12)
Generalized Pareto (Pa)	$f^{\circ}(d;\alpha,k)=\frac{1}{\alpha}\left(1-\frac{k}{\alpha}d\right)^{1/k-1}$	$\alpha>0$ – scale $k<0$ – shape	(13)
Generalized Exponential (GE)	$f^{\circ}(d;\alpha,\gamma)=\frac{\gamma}{\alpha}\exp(-d/\alpha)\left[1-\exp(-d/\alpha)\right]^{\gamma-1}$	$\alpha>0$ – scale $\gamma>0$ – shape	(14)
Gamma (Ga)	$f^{\circ}(d;\lambda,\alpha)=\frac{1}{\alpha^{\lambda}\Gamma(\lambda)}d^{\lambda-1}e^{-(d/\alpha)}$	$\alpha>0$ – scale $\lambda>0$ – shape	(15)

Note that Exponential distribution is a special case of all other mentioned above distributions, Eqs.(12)-(15).

The detailed information on the models mentioned above with the methods of ML estimation, one can easily find in hydrological and statistical literature, e.g. in Rao and Hamed (2000) and for *GE* in Gupta and Kundu (2000).

4.2 Non-stationary case

The non-stationary Flood Frequency Analysis has been a subject of numerous publications. Davison, and Smith (1990) dealt with time-related POT approach, Strupczewski and Mitosek (1991) (later completed and published e.g. in Strupczewski and Feluch, 1997abc, 1998ab and Strupczewski et al, 2001abc) dealt with maximum estimation of flood distribution functions within the presence of time as the covariate. Similar approach is presented e.g. in Katz et al (2002) and Stasinopoulos and Rigby (2007), Stasinopoulos *et al*., (2008, 2012). The basic assumption in the classical Flood Frequency Analysis and the Duration-Flood-Frequency modelling is that neither the adopted distribution function nor its parameters change in time. However, the longer the hydrological series, the harder to maintain the assumption of stationarity in the face of a changing environment and climate. (Milly, *et al*., 2008). The non-stationarity of hydrological data ought to be taken into account in FFA for theoretical and empirical reasons, but practical aspects of its introduction into design and planning procedures are not so obvious and simple and pose significant ongoing challenges to the hydrological research and water management policy. One could easily accept the increasing trend in design upper quantiles, but decreasing detected trends may distort decision-making in the engineering

design, evaluation of flood risk and in other flood-related issues. Especially when statistical inference is based on peak flow series of average length currently covering barely 60, 70 elements or on climate change scenarios and their hydrological response that we presume, we are able to predict in a realistic manner. Herein the formal aspects of at site non-stationary Duration-Flow-Frequency modelling are presented while regional Flow-Duration-Frequency modeling being introduced by Cunderlik and Ouarda (2006).

Assuming that only the values of parameters of the continuous part of the PDF may vary with time, but its form remains unchanged, the PDF f can be written as:

$$f\left(d|t\right)=\beta\left(t\right)\delta\left(d\right)+\left[1-\beta\left(t\right)\right]f^{\circ}\left[d;\mathbf{g}\left(t\right)\right]\cdot 1\left(d\right) \tag{16}$$

Assuming the forms of trends and denoting the vectors of their parameters, respectively, as $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ we have got:

$$f\left(d|t\right)=\beta\left(t;\boldsymbol{\theta}\right)\delta\left(d\right)+\left[1-\beta\left(t;\boldsymbol{\theta}\right)\right]f^{\circ}\left[d;t,\boldsymbol{\xi}\right]\cdot 1\left(d\right);\boldsymbol{\theta}\notin\boldsymbol{\xi}. \tag{17}$$

For compact notation let us define the dichotomous variable Y_t given by:

$$Y_t=\begin{cases}1\text{ for }D=0\\0\text{ for }D>0\end{cases} \tag{18}$$

For the time series $\mathbf{d} = (d_1, d_2,\ldots, d_t,\ldots, d_T)$ of the maximal annual duration of river flows exceeding the given threshold, the likelihood function can be expressed as:

$$L=\prod_{t=1}^{T}\beta\left(t;\boldsymbol{\theta}\right)^{y_t}\cdot\prod_{t=1}^{T}\left(1-\beta\left(t;\boldsymbol{\theta}\right)\right)^{1-y_t}\cdot\prod_{t=1}^{T}f^{\circ}\left(d_t;t,\boldsymbol{\xi}\right)^{1-y_t} \tag{19}$$

and the Log-likelihood function

$$\ln L=\sum_{t=1}^{T}y_t\cdot\ln\left(\beta\left(t;\boldsymbol{\theta}\right)\right)+\sum_{t=1}^{T}\left(1-y_t\right)\cdot\ln\left(1-\beta\left(t;\boldsymbol{\theta}\right)\right)+\sum_{t=1}^{T}\left(1-y_t\right)\cdot\ln\left(f^{\circ}\left(d_t;t,\boldsymbol{\xi}\right)\right) \tag{20}$$

As one can see from Eq.(20), the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, as they are independent, can be estimated separately.

4.2.1 Estimation of parameters of the continuous part of Eq.(16) [$f^{\circ}(d; t, \boldsymbol{\xi})$]

The ML estimate of the parameters $\boldsymbol{\xi}$ of CPDF ($f^{\circ}(d; t, \boldsymbol{\xi})$) are obtained by solving the system of equations:

$$\frac{\partial\ln L}{\partial\boldsymbol{\xi}}=\frac{\partial}{\partial\boldsymbol{\xi}}\sum_{t=1}^{T}\left(1-y_t\right)\cdot\ln\left(f^{\circ}\left(d_t;t,\boldsymbol{\xi}\right)\right)=0 \tag{21}$$

while the candidate functions f° are given by Eqs.(11)-(15) however with time dependent parameters in this case (Strupczewski *et al.*, 2001, Strupczewski & Kaczmarek, 2001). The estimates can also be found by direct search for the maximum of Log-likelihood function [the last component of Eq.(20)] with respect to trend parameter vector $\boldsymbol{\xi}$.

The consequence of making allowance for time dependent parameters of. $f^{\circ}(d; \mathbf{g})$ is an increase of the number of parameters to be estimated. Given the small number of non-zero elements in the time series $\mathbf{d} = (d_1, d_2,\ldots, d_t,\ldots, d_T)$, the number of parameters which can be effectively estimated is small. Therefore, we decided to adopt the values of these parameters as independent of time. Then the only non-stationarity lies in the weighting parameter $\beta(t; \boldsymbol{\theta})$ which plays the role of the time-dependent function 'switching' on and off the event of dikes' prolonged exposure to high waters. Note here that the duration d is a parameter that describes the shape of the flood hydrograph, so we assume that the persistence of flood of magnitude $Q_A < Q_{\max} < Q_B$ is not subject to time variability.

4.2.2 Two ways of estimation the time dependent weight parameter $\beta(t; \boldsymbol{\theta})$

The estimation of parameters $\boldsymbol{\theta}$ of the discrete part – weighting parameter $\beta(t; \boldsymbol{\theta})$, in the joint distribution Eq.(17) can be performed in two ways: by regression analysis and on the base of non-stationary distribution of annual maxima with time dependent parameters.

Regression analysis

The variable Y_t represents binary outcomes and has a binomial distribution with parameter:

$$\beta\left(t;\boldsymbol{\theta}\right)=P\left(Y_t=1\right)=P\left(D=0\right) \tag{22}$$

However the trend in β can not be found by means of frequently assumed linear regression. The reasons of being that

- in general linear trend may take the values of probability $\beta(t, \boldsymbol{\theta})$ outside the range from 0 to 1,
- the error term is not homoscedastic, nor it is normally distributed as in normal regression.

In order to avoid values outside the range from 0 to 1 a monotonic transformation of the interval (0,1) is performed to the range $(-\infty, +\infty)$.There are many transformations with this property, but the most popular are two: probit and logit transformations. Both give similar results but logit transform is more convenient for calculations. Probit transformation consists in converting the probability to corresponding quantiles of the standard normal distribution. Logit transformation is given by:

$$\text{logit} = \ln[\beta/(\beta - 1)] \tag{22a}$$

And the trend is modelled as:

$$\text{logit} = a + bt \tag{22b}$$

Inverse transformation leads to the logistic (LO) function β of time t with parameter vector $\boldsymbol{\theta} = [a, b]$.

$$\beta(t;a,b) = \frac{1}{1+e^{-(a+bt)}} \tag{23}$$

Logistic regression is used in many disciplines, medicine, social science, econometrics, in engineering, especially for predicting the probability of failure of a system or product.

The logistic regression coefficients a and b are usually determined using maximum likelihood estimation by iterative process until the improvement of the solution is minute and the procedure is said to have converged. Sometimes, when the considered flow threshold is high and thus number of 'ones' greatly exceeds number of zero values of y_t, the convergence cannot be reached. The failure to converge may indicate that the trend coefficients are not significant or other methods of inference about the trend in β should be applied.

Several measures enable to evaluate the goodness of fitted trend model. Deviance, pseudo-R^2 and odds ratios confidence intervals are the most frequently used. There are two measures of deviance corresponding to the likelihood ratio. One, called model deviance, to compare fitted model to saturated model (a theoretical model with perfect fit) and second, null deviance, which represents the difference between null model [a model with only intercept, so representing the stationary case, β given by Eq.(8)] and saturated model. Model deviance is given by equation:

$$D_{\text{model}} = -2\ln\frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}} \tag{24}$$

and similarly, null deviance:

$$D_{\text{null}} = -2\ln\frac{\text{likelihood of the null model}}{\text{likelihood of the saturated model}} \tag{25}$$

Note that in logistic regression the likelihood of the saturated model ($y_t = \beta(t; \boldsymbol{\theta})$) is equal 1.

The deviance has an approximate chi-square distribution with 1 degree of freedom for each predictor, so 1 in our case. Smaller values of deviance indicates better fit what corresponds to non-significant chi-square values.

Pseudo $-R^2$ is calculated on the base of deviances:

$$\text{Pseudo} - R^2 = \frac{D_{\text{null}} - D_{\text{model}}}{D_{\text{null}}} \tag{26}$$

and interpreted almost like a coefficient of determination in linear regression.

The method via annual maxima distribution with time-varying parameters

An alternative way of analyzing a trend in β is to use the non-stationary CDF of annual peaks with time dependent parameters. From NFFA (Strupczewski et al., 2001) one gets $G = \phi(Q, \mathbf{h}, t)$ where $\mathbf{h}$ – the vector of PDF parameters of the annual flood peaks distribution. Then per analogy to Eq.(9a) one can write:

$$\hat{\beta}(t) = \hat{P}\left[D(t)=0\right] = \hat{P}\left[Q_{\max}(t) \le Q_A\right] + \left\{\hat{P}\left[Q_{\max}(t) > Q_B\right]\right\} = \hat{G}\left(Q_A|t\right) + \left[1 - \hat{G}\left(Q_B|t\right)\right] \tag{27}$$

providing the selected distribution and trend model of its parameters fits well upper tail of data. It would be advisable to compare the results of both methods. Compatibility of the results could serve as the overall test of correctness of the assumptions made.

4.2.3 Probability of inundation during the period (t_1, t_2)

Dealing with hydrologic design, due to non-stationarity, the notion of return period is no longer valid and the probability of inundation should refer to the whole period of life of a hydraulic structure,-not to a single year as has been agreed in the stationary case.

When the parameters of DqF distribution are time dependent, consequently the annual probability of leves breach [Eq.(3)] becomes time dependent: $P_2(Fl, t)$. The probability that at least once in the period (t_1, t_2) the inundation caused by levees breach occurs is expressed as:

$$P_2\left(Fl,(t_1,t_2)\right)=1-\prod_{t=t_1}^{t_2}\left[1-P_2\left(Fl,t\right)\right] \tag{28}$$

Similarly, if the distribution of annual maximum peaks is time dependent, $G = \phi(Q, \mathbf{h}, t)$, the exceedance probability of overflow of the levees' crest, so the probability that [see Eq.(2)], $P(Q_{\max} \geq Q_B, t) = 1 - G(Q_B, t) = P_1(Fl, t)$ is time dependent. Then the probability that the inundation caused by overtopping the embankment crest occurs at least once in the period (t_1, t_2) and can be expressed as

$$P_1\left(Fl,(t_1,t_2)\right)=p\left(Q>Q_B,(t_1,t_2)\right)=1-\prod_{t=t_1}^{t_2}\left[1-P_1\left(Fl,t\right)\right] \tag{29}$$

The total probability of inundation in the period (t_1, t_2) equals to:

$$P\left(Fl,(t_1,t_2)\right)=P_1\left(Fl,(t_1,t_2)\right)+P_2\left(Fl,(t_1,t_2)\right) \tag{30}$$

5 EXAMPLE – SZCZUCIN AT VISTULA RIVER (SOUTHERN POLAND)

To illustrate how the proposed approach works in practice the Szczucin gauge (southern Poland) at the Vistula River has been selected as an example. Recent flooding in the upper Vistula bared the weakness of the system of flood protection, especially unsatisfactory condition of the embankments in the region of Szczucin. One, but not only, major reason for the current state of flood protection infrastructure is a complex history of these lands. When Western European countries formed an effective flood protection schemes Polish south-eastern lands were periphery of three empires, two of which were among the most undeveloped countries of the continent. After regaining independence, social and economic problems associated with merging the various districts of the reborn Poland influenced the poor development of an efficient protection system. For these reasons, embankments built by the World War II do not meet current requirements which were lately even put to higher level. The Polish People's Republic period did not bring any important changes. Although, the embankments have been periodically increased and strengthened, the high cost of post-war reconstruction and industrialization carried out under conditions of socialist economy, did not allow to catch up with Western standards. Lately, the material excavated on the flood land, very often at the immediate vicinity of the embankments, was used for the re-construction. As a consequence, the top layer of inactivated meadow was damaged, what facilitated the filtration of water from the horizontal residual layer under the layer of permeable sealer coat. There are present plans to modernise the dikes and first works have been carried out. The investor claims that the modernisation will reduce the flooding risk by 80%. To assess the risk before and after modernisation (provided that the statement of the investor is right) the following analysis was performed.

The daily flows record covering the period 1951-2006 (n = 56 years) was used in this study. At first the daily records have been controlled and tested with regard to the sharp discontinuities and jumps in data – no particular irregularities have been detected (Fig. 4).

The overtopping flow Q_B was assessed from the rating curve as 10,500 m^3/s which roughly corresponds to two-hundred-years return period of annual peak flow ($Q_{0.5\%}$), the base design value for the Ist class embankments. In fact, there are no annual peak flows exceeding this value in the record. Therefore the Q_B value does not affect the composition of the vector of observation values [$\mathbf{d_t}$]. The alarm threshold for the Szczucin station Q_A = 1690 m^3/s (which means flow of ca. 2-year return period, stage 660 cm), however, for completion a few other thresholds will be analysed, too, namely Q_{Tr} = 700, 1000, 1300 and 2000 m^3/s. The hazard index $h(Fl|d)$ for Q_A = 1690 m^3/s [Eq.(3)] was assessed as:

$$h\left(Fl\middle|d\right)=\begin{cases}0.05\cdot d & \text{for } d\le 20 \text{ days}\\ 1 & \text{for } d>20 \text{ days}\end{cases}. \tag{32}$$

so the embankments cannot withstand the pressure of high waters of more than 20 days.

Fig. 4 Hydrograph of the daily flows at the Szczucin gauging station. Horizontal dashed lines reflect the Q_{Tr} values used in this study.

5.1 Stationary case

The weak correlation between the durations when $d(t) > 0$ (see Fig. 5) and the respective annual maxima $Q_{\max}(t)$ indicates the variety of shapes of flood hydrographs and, as a consequence, d cannot be represented (or replaced rather) in FFA by $Q_{\max}$. It implies the analysis of both d and $Q_{\max}$ by (perhaps) two different types of models. As a model for the parameters of the $f°$ function Generalised Exponential (GE) distribution has been chosen (e.g. Gupta & Kundu, 2000). Among the distributions presented in Eqs.(11)-(15) the GE distribution Eq.(14) performs relatively well in terms of the AIC value and shows stability of numerical ML solutions in estimation of $f°(d;\mathbf{g})$ parameters, regardless the Q_{Tr} threshold applied for the calculations. The list of the GE estimated parameters of the two-component DqF model and β values for different Q_{Tr} including Q_A is presented in Table 2.

Fig. 5 The durations (in days) of the discharge above Q_A = 1690 m^3/s for Szczucin gauging station (1951–2006). The annual maximal durations are in black.

The annual maxima are believed to be adequately described by the heavy-tailed distributions (e.g. Strupczewski, *et al*, 2011), so to cater for the Flood Frequency Analysis (FFA) for extreme values (annual maxima) the β values [Eq.(8)] and $P_1(Fl)$ [Eq,(2)] by means of $Q_{\max}$ series were calculated with the three-parameter Generalised Extreme Value distribution:

$$G\left(q;\alpha,\gamma,\varepsilon\right)=\exp\left\{-\left[1-\frac{\gamma}{\alpha}\left(q-\varepsilon\right)\right]^{1/\gamma}\right\}=G_{GEV}^{\gamma}\left(x\right) \tag{33}$$

From the AM sample covering the period 1951–2006 (n=56 years) we got the ML estimates of GEV parameters equal (for calculations we used our original soft-packages *FloodDurations*, *NonstationaryMLM* and *SDEP* which we can eagerly share with others):

location $\equiv \hat{\varepsilon}$ = 1260.02 m^3/s, *scale* $\equiv \hat{\alpha}$ = 671.39 m^3/s and *shape* $\equiv \hat{\gamma}$ = –0.33.

For completion note that the value of log-likelihood function ln$L$ = –463.231 and thus AIC = 932.461.

Substituting for q into Eq.(33) the chosen Q_{Tr} and Q_B values and then putting the corresponding probabilities to Eq.(9a), one gets the estimates of the weighting parameters display in Table 2.

Table 2. The parameters of the two-component DqF model for Szczucin data.

	First component (by two methods)			**Second component, $f°$is the two-parameter Generalised Exponential**		
Q_{Tr}	n_2	$\beta = n_1/n$ by Eq.(8)	$\beta = \beta(Q_{Tr})$ by Eq.(9)	*scale*	*shape*	lnML/n_2
700	51	0.089	0.076	2.8799	0.2938	–2.63
1000	40	0.286	0.226	4.0392	0.5228	–2.10
1300	32	0.429	0.395	4.8616	0.7464	–1.77
1690*	*23*	*0.589*	*0.577*	*3.4238*	*0.8357*	*–1.62*
2000	17	0.696	0.683	3.7411	0.9126	–1.54

* $Q_{Tr} = Q_A$

One can notice from the Table 2 that β got by means of Eq.(8) and Eq.(9) are quite similar particularly for higher values of Q_{Tr} and for all cases the confidence interval for proportion β includes the value estimated from AM distribution [Eq.(9)].

5.1.1 Assessment of probability of levee breach along Szczucin reach.

Since the event of levee breach is conditioned by the peak flow being in the range of $[Q_A, Q_B]$, Eq.(3) can be written as (see also Fig. 6)

$$P_2(Fl) = (1-\beta)\int_{0^+}^{\infty} h(Fl|d)\cdot f^{\circ}(d)\cdot \mathrm{d}d \tag{34}$$

The pdf of GE [Eq.(15)] for $Q_{Tr} = Q_A = 1690$ m^3/s (Table 2) takes the form

$$\left(1-\hat{\beta}\right)\cdot f^{\circ}(d;\alpha = 3.4238, \gamma = 0.8357) = 0.423\frac{0.2441\cdot \exp\left(-d/3.4238\right)}{\left[1-\exp\left(-d/3.4138\right)\right]^{0.1643}} \tag{35}$$

while the ML estimate of β equals (Table 2) 0.577. Substituting them and the hazard index function defined by Eq.(32) into Eq.(34) and integrating one gets the annual probability of levee breaching $P_2(Fl) = 0.064$. Note that at the same time, and when the same GEV distribution is used [see the Eq.(33) and its parameters below the equation], the probability of flood caused by exceeding embankment crest by annual peak flow: $P_1(Fl) = P(Q_{\max} > Q_B = 10{,}500$ m^3/s$) = 1 - G(Q_B)$ is equal to 0.005, so it is almost insignificant (more than ten times smaller than P_2), hence, the overall probability of flood along Szczucin reach $P = P_1 + P_2 = 0.069$.

Variety of shapes of flood hydrographs one can evaluate by a measure of correlation strength between $Q_{\max}(t)$ and $d(t)$. Due to shape similarity of flood peak parts, a strong dependence between the peak flows (Q_{max}) and the duration above the alarm flow (d) can take place. If it is a case, the probability $P_2(Fl)$ can be assessed on the base of Q_{max} distribution $g(Q_{max})$. Assuming that $d = \psi(Q_{\max})$ one can expressed in Eq.(34) the d variable by the $Q_{\max}$ getting

$$P_2(Fl) = p\left(Fl\left|\left(Q_A < Q_{\max} \le Q_B\right)\right.\right) = \int_{Q_A}^{Q_B} h\left(Fl\left|\psi\left(Q_{\max}\right)\right.\right)\cdot g\left(Q_{\max}\right)\cdot \mathrm{d}Q_{\max} \tag{36}$$

where per analogy to Eq.(32) $h(Fl|\psi(Q_{\max}))$ equals 0 and 1 for Q_A and Q_B, respectively. The Pearson's correlation coefficient $r(Q_{\max}, d)$ for Szczucin equals to 0.83.

Fig. 6 Components of the integral Eq.(34).

Of course, when estimating the risk of a levee breach except the time of high water residence, more technical parameters of levees should be analysed, such as the construction of the levee, the material used for its building, its age, susceptibility to softening, the regime of the river, wind-induced waving and so on. All in all, those who decided to build their houses in the river's proximity behind the levees, sooner or later do experience a catastrophe.

5.2 Non-stationary case

Analysis of long series of hydrological observations on Polish rivers lead us to the conclusion that two random variables whose probability distributions have been considered as components of DqF analysis show different behaviour versus time. The continuous variable – duration of water level above certain stage – in general, shows no trend. It describes the shape of the flood waves which has been stated to be rather stable and, if any trend there exists, it does not pose any effect on the final results of the DqF calculations. On the other hand, a visual assessment of records for Szczucin and other hydrological stations show that the frequency of occurrence of extreme flows (P_1) and flows above (so well below) a given threshold (Q_{Tr}) may reveal some trend. Therefore in this study we focused only on the search of trends in the probability P_1 and in the weighting factor β that plays the role of the time-dependent function 'switching' on and off the event of dikes' prolonged exposure to high waters. These trends have been estimated from the annual peak flow series and by direct analysis of $[d_t]$ vector represented by the sequence of 0 and 1 as given by Eq.(18). In both cases the maximum likelihood method (MLM) has been used for calculation, while the logistic function (23) serves to model the (0,1) duration series.

The estimation $\beta(t)$ for the threshold corresponding to the alarm stage ($Q_A = 1690$ m^3/s) in the form of the logistic function (26) revealed the decreasing trend ($b < 0$), whereas $a = 0.405$, so the $\beta(t)$ takes the form:

$$\beta^{LO}(t) = \left[1+\exp\left(0.002\cdot t - 0.405\right)\right]^{-1} \tag{37a}$$

and the parameters of stationary f° function for selected Q_{Tr} values can be found in Table 2.

The above equation [Eq.(37a)] says that the odds (the ratio of probabilities of events against nonevents: $\beta(t)/(1-\beta(t))$ decreases in average by 0.2% from year to year, that gives the change of β from ca. 0.60 in 1951 to about 0.58 in 2006. However this trend is not statistically significant. The model deviance D_{model} being equal to 75.8286 and the null deviance $D_{\text{null}} = 75.8372$ give the difference with p-value of 0.9264 from chi-square distribution. The value of pseudo – $R^2 = 0.046$ is close to 0. It is likely that this result points on almost stable risk of inundation caused by dike breaches for summer floods that prevail in the reach of the Vistula river represented by Szczucin hydrological station, where changes in the river bed and on the floodplains have not influenced considerably the transportation of high waters. Winter floods can reveal stronger trends due to greater variability of melting condition and observed temperature rise, so, as consequence, the volume of runoff. Small catchments seem to be more susceptible for trends in β. These statements ought to be verified on the larger hydrological data set.

If instead of the logistic (LO) we take the non-stationary Generalised Extreme Value (GEV) distribution function (see stationary case above), assume linear trends in mean value and standard deviation (but not in the parameters of location, scale and shape) and calculate the $\beta(t)$ by means of Non-stationary Flood Frequency Analysis (e.g. Strupczewski, *et al*, 2001, 2009) we obtain:

$$\beta^{GEV}(t) = \exp\left\{-\frac{\left[t\cdot\left(33.067\cdot t - 22453.3\right) + 3.812\cdot 10^6\right]^{1.54}}{\left[1.43\cdot t + 1.339\cdot\sqrt{t\cdot\left(33.067\cdot t - 22453.3\right) + 3.812\cdot 10^6} - 275.269\right]^{3.08}}\right\} \tag{37b}$$

The comparison of the values of the non-stationary log-likelihood function and AIC, $\ln L = -463.078$ and AIC = 936.157, respectively with the stationary results reveals that the supplement by two extra parameters to the model (those responsible for the linear trend in mean and standard deviation) worsen the estimation results. It means that for a given series size ($N = 56$) the detected trends are in fact weak, and perhaps addition of a few new measurements in series can dramatically change their value or even sign. The weakness of the trends in moments are confirmed by the weakness of β time-variability.

The time variability of β functions got by the two approaches are shown in the Fig. 7.

Fig. 7 Non-stationary $\beta(t)$ by two approaches

The above equations [Eqs.(37a) and (37b)] and the diagram (Fig. 7) point at the difference in trend sign of β between the results received by the two approaches (LO and GEV). However, there are similarities, too. The results for both cases say that the value of β is practically time independent (statistically insignificant) within time period 1951 to 2006 and thus maintain the relatively constant balance between the first and the second terms of the DqF probability density function [Eq.(17)]. In consequence, the durations of water stay above Q_A described by the f° function are actually as frequent nowadays as they were in past. On the other hand, the probability P_1 and P_2 (and thus P) are now the functions of t. If we take the GEV-based $\beta(t)$ as an example (as more reliable than LO-based $\beta(t)$) and $t = 1$ (year 1951) one obtains $P_2 = 0.066$. Further, with the non-stationary GEV (by the same parameters as for $\beta(t)$): $P_1 = 0.007$, so in consequence $P = 0.073$. For $t = 56$ (year 2006): $P_1 = 0.004$, $P_2 = 0.064$, so $P = 0.068$, thus the probability of flood in Szczucin dropped by 7% over the half of the century – a judgement whether it is much or not we leave for the reader and decision makers. Please also note that regardless the point in time the ratio P_1/P_2 is similar to the stationary conditions.

However, the probability for the certain point in time may not carry information sufficient for flood protection authority. Therefore, it is interesting to know what is the probability of inundation over the certain period, e.g. 20 years of the exploitation of the dikes in Szczucin. For the GEV non-stationary model (with the parameters mentioned above) and last 20 years of the time series (1986-2006) the probability of overtopping over the levee crest is equal to $P_1 = 0.048$, whereas the dike's breach probability is more than 10 times larger: $P_2 = 0.516$. Overall risk of inundation $P = 0.563$, it is almost 10 times larger than for a single year. The reader also notes easily that again the ratio P_1/P_2 is alike the ratios for the point-in-time non-stationary case as well as for the stationary case.

One has to bear in mind, however, that the linear trend in parameters (in case of the LO) and first two moments (as it was in GEV) is just the simplest of the countless trend patterns that may be employed for the time-dependent models and application of other ways (e.g. parabolic, polynomial, exponential, etc.) usually leads to the overparametrisation and noteworthy complication of numerical calculations. It is so, because

maximum likelihood estimates for time-dependant models require multi-parameter optimisation of relatively 'flat' log-likelihood functions with use of relatively short data-series.

6 CONCLUSIONS

In the paper the new two-component model of flood waves, 'duration of flooding-discharge-probability of non-exceedance' (DqF), with the methodology of its parameters estimation was proposed as a completion to the classical FFA methods. Such model can estimate the duration (d) of stages (and flows) exceeding the assumed magnitude with a certain probability which is of key importance when the river's dikes are prone to the prolonged impact of high waters. The embankments may be weaken by the water, soak and eventually break – this is the most frequent cause of floods in Poland. However, in this study the two main causes of inundation of embanked rivers, namely over-crest flow and wash out of the levees, were combined to assess the total risk of inundation. The proposed DqF modelling approach was generalised to the non-stationary conditions. Therefore, in addition to the maximum flow one should consider also the duration of high waters above the alarm flow Q_A in a river channel. The model combined with the technical evaluation of probability of levees breach expressed by the hazard index gives the annual probability of inundation caused by the embankment failure. The probability of inundation is the total of probabilities of exceeding embankment crest by flood peak and the probability of washout of levees.

The DqF modelling is the consequence of QdF approach developed by Javelle *et al.* (1999, 2000, 2002) and Bogdanowicz *et al.*(2008) but in the first model the gravity is put on the probability of the certain duration above alarming stage/discharge (Q_A) rather than on magnitude of flood itself ($Q_{\max}$) like in the latter case (Fig. 3).

The DqF model in the form of Eq.(5) consists of two terms: $\beta \cdot \delta(d)$ deals with the zero event, $D = 0$, whereas the latter term $(1 - \beta) \cdot f^{o}(d;\mathbf{g}) \cdot 1(d)$ stands for the events when the duration $D > 0$. In general both β and f^{o} in non-stationary case may depend on time. The maximum likelihood method (MLM) was proposed for estimation of β and $\mathbf{g}$ parameters. In the non-stationary case it is convenient to describe the $\beta(t; \boldsymbol{\theta})$ by means of the logistic function (23). However, β and $\beta(t; \boldsymbol{\theta})$ can be also estimated by means of annual peak flows series, $Q_{\max}$, using the routine flood frequency techniques (FF) with distribution functions commonly used in FFA (e.g. GEV) for stationary and non-stationary case, respectively. Note that estimating the weighting factor β and $\beta(t; \boldsymbol{\theta})$ from the duration d time series the information (0,1) for excess the threshold level Q_{Tr} is used exclusively, while basing on the annual peak flow time-series $Q_{\max}$ the information from whole range of recorded flood magnitude is used to assess the trend in the alarm flow Q_A. For $f^{o}(d_j; \mathbf{g})$ model (both stationary and non-stationary) the exponential-like shaped distribution functions are recommended, such as: Exponential, Weibull, Pareto, Generalised Exponential, Gamma and similar.

The calculations for the Szczucin at the Vistula River case study made for several threshold values (Q_{Tr}) including the alarm flow (Q_A) have showed the similar results for the weighting factor β estimated by ML method from the duration time-series and from annual peaks time-series (Table 2). The peak flows that could overtop the embankments have not been detected in the Szczucin's record (1951-2005). According to the hazard function (32) the possibility of levees breaching increases almost tenfold the probability of inundation.

Variability in the Szczucin time series of the d-duration (understood as a time-dependence of the $\mathbf{g}$ parameters of $f^{o}(d_i; \mathbf{g})$) has not been subject of modelling because of the insufficient data and the conviction based on the visual judgment of the (d(t) vs. t) diagram that the trend would be negligibly small. The only trend considered is the trend in the weighting factor β. The significant difference in trend estimates of $\beta(t)$ got by ML method from the direct analysis of $[\mathbf{d}_t]$ vector represented by the sequence of 0 and 1 [Eq.(18)] assuming the logistic function (LO) of time and from GEV distributed annual peak flow series is striking. The results for both cases differ in sign (Fig.6) and moreover they point that the value of β is practically time independent within time period 1951 to 2006. Nevertheless, as long as the change of river regime in time is visible (regardless its origin), one should consider non-stationary modelling accepting (sadly) the fact that the tools available are in their infancy.

The DqF model proved to be the important completion to the traditional FFA concentrating on maximal seasonal or annual discharges. The DqF approach is especially useful in polish specific conditions where the flood protection infrastructure is dated and often does not survive confrontation with prolonged pressure of high waters.

Reliable data and information about floods are indispensable for better understanding the interactions between rivers and flood protection system: embankments, reservoirs and polders. Improvement of statistical

models is essential for engineering design in general and in particular for implementation of flood risk mitigation procedures. Not only has the DqF modelling shown that actual flood risk is greater than the risk assessed by means of classical FFA but also provides quantitative measures which can be used in flood protection systems planning, exploitation and conservation. This measures in form of dependence of inundation risk on river flow (or water level) should be established for other hydrological stations on Polish rivers and their dimensionless versions compared. The geographic information systems technique (GIS) could be used to indicate locations prone to inundation, Also the GIS can be a helpful tool to visualisation and testing trends in the structure of river network and to the regional analysis. These results can constitute the theoretical background to a number of practical decisions in water management issues.

7 ACKNOWLEDGMENTS

The authors would like to thank Prof. Janusz Żelaziński for inspiring discussions and exchange of ideas, without which this work would not have its present form and meaning. Sincere thanks for a sense of realism and keeping both feet on the ground what in face of many sources of uncertainty, so typical for hydrology, helped us to work out practical solution to the problem.

This research project was partly financed by the grant of the Polish National Science Centre titled '*Modern statistical models for analysis of flood frequency and features of flood waves*', decision nr DEC-2012/05/B/ST10/00482, the Grant Iuventus Plus IP 2010 024570 '*Analysis of the efficiency of estimation methods in flood frequency modelling*' and made as the Polish contribution to COST Action ES0901 '*European Procedures for Flood Frequency Estimation (FloodFreq)*'.

8 REFERENCES

Askhar, F., 1980, *Partial duration series models for flood analysis.* PhD Thesis. Ecole Polytechnique de Montréal. Montréal, Canada.

Balocki, J.B. and Burges, S.J., 1994, Relationships between n-day flood volumes for infrequent large floods. *Journal of Water Resources Planning and Management* **120** (6), 794-818.

Bogdanowicz, E., Strupczewski, W.G. and Kochanek, K., 2008, Application of Discharge-duration-Frequency model for description of peak part of flood hydrograph. (in polish) *Przegląd Geofizyczny*. LIII. 3-4, 263-288.

Bogdanowicz, E., Strupczewski, W.G. and Kochanek, K., 2011, Persistence as a factor of flood hazard for embanked rivers. EGU Leonardo Conference "Floods in 3D", Bratislava 23-25 Nov. *Abstract in Proceedings.*

Botter, G., S. Zanardo, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo, 2008, Ecohydrological model of flow duration curves and annual minima, *Water Resour. Res.*, **44**, W08418, doi:10.1029/2008WR006814.

Castellarin, A., R. M. Vogel, and A. Brath, 2004, A stochastic index flow model of flow duration curves, *Water Resour. Res.*, 40, W03104, doi:10.1029/2003WR002524.

Cunderlik, J.M. and Ouarda, T.B.M.J., 2006, Regional flood-duration-frequency modelling in changing environment. *J. of Hydrology* 318, 276-291.

Davison, A.C. and Smith, R.L., 1990, Models for exceedances over high thresholds. *Journal of the Roy. Statist. Soc.. Series B*, **52**, pp. 393-442.

Eagleson, P.S., 1972, Dynamics of flood frequency, *Water Resour. Res.*, 8(4), 878-898

Galéa, G. and Prudhomme, C., 1997, Notations de bases et concepts utiles pour la comprehension de la modélisation synthétique des régimes de crue des basins versants au sens des modeles QdF. *Rev. Sci. Eu.* 1, 83-101.

Gilliom, R.J. and Helsel, D.R., 1986, Estimation of distributed parameters for censored trace level water quality data-1. Estimation techniques. *Water Resourc. Res.* **22**: 1201-1206.

Gioia, A., Iacobellis, V., Manfreda, S., and Fiorentino, M., 2008, Runoff thresholds in derived flood frequency distributions. *Hydrol. Earth Syst. Sci.*, **12**, 1295–1307, doi:10.5194/nhess-12-1295-2008.

Gupta, R. D. and Kundu, D., ,2000, Generalized Exponential Distribution: Different Method of Estimations, *J. Statist. Comput. Simul.*, 2000, Vol. 00, pp. 1 – 22.

Haas, C.N. and Scheff, P.A., 1990, Estimation of averages in truncated samples. *Environ. Sci. Technol.* **24(6):** 912-919.

Harlow, D.G., 1989, Effect of proof-testing on the Weibull distribution. *J. Material Sci.* **24**: 1467-1473.

Helsel, D.R., 1990, Less than obvious: statistical treatment of data below detection limit. *Environ. Sci. Technol.* **24(14)**:1767-1774.

Iacobellis, V., 2008, Probabilistic model for the estimation of T-year flow duration curves, *Water Resources Research*, ISSN: 0043-1397, **Vol. 44**, doi: 10.1029/2006WR005400

Iacobellis, V., Gioia, A., Manfreda, S., Fiorentino, M., 2011, Flood quantiles estimation based on theoretically derived distributions: regional analysis in Southern Italy. *Nat. Hazards Earth Syst. Sci.*, **11**, 673–695, doi:10.5194/nhess-11-673-2011.

Interagency Advisory Committee on Water Data. 1982, *Guidelines for determining flood flow frequency*. Bulletin 17B, U.S. Department of the Interior, Geological Survey, Office of Water Data, Reston, Va.

Javelle, P., 2001, *Charactérization du régime des crues: le modèle débit-durée-fréquence convergent. Approache locale et régionale*, PhD thesis, Camagref-Lyon. Institut National Polytechnique de Grenoble, 268p.

Javelle, P., Galéa, G. and Grésillon, J.M., 2000, L'approche debit-durée-fréquénce: historique et avancées. *Revue des Sciences de la terre et des planétes* 329, 39-44.

Javelle, P., Grésillon, J.M. and Galéa, G., 1999, Discharge-duration-frequency curves modelling for floods and scale invariance. Comptes Rendus de l'Academie des Sciences, *Sciences de la terre et des planets* 329, 39-44.

Javelle, P., Ouarda, T.B.M.J., Lang, M., Bobée, B., Galéa, G. and Grésillon, J.M., 2002, Development of regional flood-duration-frequency curves based on the index-flood method. *J. of Hydrology* 258, 249-259.

Jennings, M.E., Benson M.A., 1969, Frequency curve for annual flood series with zero events or incomplete data. *Water Resour. Res.* **51**: 276-280.

Kaczmarek Z., 1977, *Statistical Methods in Hydrology and Meteorology*. Sec. 4.5, 205-217.Published for the Geological Survey, U.S. Department of the Interior and the National Science Foundation, Washington, D.C (translation of Polish book, 1970), 320 pp

Katz, R.W., Parlange, M.B. and Naveau, P., 2002, Statistics of extremes in hydrology. *Adv. Water Resour.* **25**, 1287-1304.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J. ,2007, Stationarity Is Dead: Whither Water Management? *Science*, February 2008: Vol.319 no. 5863 pp. 573-574; DOI: 10.1126/science.1151915

NERC, 1975, *Flood Studies Report. Estimation of flood volumes over different durations*. V.1, Ch. 5, pp. 243-264, V.2, Ch. 3. Nat. Environ. Res. Council, London, Vols. 1-5, 1100-pp.

Rao, A.R. and Hamed, K.H., 2000, *Flood frequency analysis*. CRC Press.

Serinaldi, F. and Grinaldi, S., (2010). Synthetic Design Hydrographs Based on Distribution Functions with Finite Support. *J. of Hydrol. Eng. ASCE*, V.16, 5, 435-446.

Sherwood, J.M., 1994, Estimation of volume-duration-frequency relations of ungauged small urban streams in Ohio. *Water Resources Bulletin* 30 (2), 261-269.

Sivapalan, M., E. F. Wood, and K. J. Beven, 1990, On hydrologic similarity, 3, A dimensionless flood frequency model using a generalized geomorphologic unit hydrograph and partial area runoff generation, *Water Resour. Res.*, **26**(1), 43-58

Stasinopoulos, D.M. and Rigby, R.A., 2007, Generalized additive models for location scale and shape (GAMLSS) in *R. Journal of Statistical Software*, 2007, 23(7), 1-46.

Stasinopoulos, M., Rigby, B. and Akantziliotou, C., 2008, Introductions on How to Use the GAMLSS Package in R (second edition), January 11, 2008.

Stasinopoulos, M., Rigby, B and Akantziliotou C., 2012, Instructions on how to use the GAMLSS package in R. Second Edition, February 25, 2012. (See Ch.6. Model selection. 6.2. Selecting Explanatory variables using addterm, dropterm and step AIC

Strupczewski, W., 1964, Flood hydrograph equation. (In Polish: Równanie fali powodziowej) *Bulletin du Service Hydrologique et Meteorologique*, Wyd. Komunikacji i Lacznosci, Fascicule 57, 2/1964, 35-58.

Strupczewski, W., 1966, *Statistical analysis of shapes of flood hydrographs* (in Polish: Statystyczna analiza kształtów fal wezbraniowych.). Doctoral Dissertation. Warsaw Technical University, pp. 181 plus Tables 25 and Figures 39

Strupczewski, W.G., Mitosek, H.T., 1991, How to deal with nonstationary time series in the hydrologic projects. *Mitteilungsblatt des Hydrographishen Dienstes in Osterreich*, Nr.65/66,36-40. Presented at IAHS Symposium, Vienna.

Strupczewski, W.G., Feluch, W., ,1997a System of identification of an optimum flood frequency model with time dependent parameters (IDT). In: *Integrated Approach to Environmental Data Management Systems*. Editor Harmanciouglu et al., Kluwer Acad. Publ. 291-300.

Strupczewski, W.G., Feluch, W., 1997b, Floods study of the Polish rivers by the IDT soft tool. *Annales Geophysicae, Part II, Hydrology, Oceans, Atmosphere & Nonlinear Geophysics*, Supplement II to Volume 15, C-310.

Strupczewski, W.G., Feluch W., 1997c, System of identification of an optimum flood frequency model with time dependent parameters (IDT). In: N.B. Harmancioglu et al. (ed) *Integrated Approach to Environmental Data management Systems*, Kluwer Academic Publishers, 291-300. Presented at NATO-ARW "Integrated Approach to Environmental Data Management Systems" Izmir, September.

Strupczewski, W.G., Feluch, W., 1998a, Flood frequency analysis under non-stationarity, *Geographia Polonica*, 71, 19-34.

Strupczewski, W.G., Feluch, W., 1998b, Investigation of trend in annual peak flow series. Part. I. Maximum likelihood estimation. Proceedings of the Second International Conference on Climate and Water V.1, 241-250.

Strupczewski, W.G., Singh, V.P., Feluch, W., 2001a, Non-stationary approach to at-site flood-frequency modelling. Part I. Maximum likelihood estimation. *J. Hydrol*., 248, 123-142.

Strupczewski, W.G., Kaczmarek Z., 2001b, Non-stationary approach to at-site flood-frequency modelling. Part II. Weighted least squares estimation. *J. Hydrol*. ,248, 143-151

Strupczewski, W.G., Singh, V.P., Mitosek, H.T., 2001c, Non-stationary approach to at-site flood-frequency modelling. Part III. Flood analysis of Polish rivers. *J. Hydrol*., 248, 152-167

Strupczewski, W.G., Singh, V.P. and Weglarczyk, S., 2002, Physically based model of discontinuous distribution for hydrological samples with zero values. Proceedings of the Int. Conf. On WaRMAR, Kuwait. Surface Water Hydrology, Singh, Al-Rashed & Sherif (eds). A.A. Balkema Publishers, Swets & Zeitlinger, Lisse, 523-537.

Strupczewski, W.G., Weglarczyk, S. and Singh, V.P., 2003. Impulse response of the kinematic diffussion model as a probability distribution of hydrologic samples with zero values. *J.Hydrol.* 270, 328-351.

Strupczewski, W.G., Kochanek, K., Feluch, W., Bogdanowicz, E. and Singh, V.P., 2009, On seasonal approach to non-stationary flood frequency analysis. *Physics and Chemistry of the Earth* 34 pp 612–618

Strupczewski, W.G., Kochanek, K., Markiewicz, I., Bogdanowicz, E., Weglarczyk, S. and Singh V.P., 2011, On the tails of distributions of annual peak flow. *Hydrology Research* Volume 42, Issue 2-3, 2011, pp. 171-192, DOI: 10.2166/nh.2011.062

Strupczewski, W., Bogdanowicz, E., and Kochanek, K., 2013, Discussion of "Synthetic Design Hydrographs Based on Distribution Functions with Finite Support" by Francesco Serinaldi and Salvatore Grimaldi." *J. Hydrol. Eng*., 18(1), 121–126. doi: 10.1061/(ASCE)HE.1943-5584.0000538

Wang, S.X. and Singh, V.P., 1995, Frequency estimation for hydrological samples with zero values. *J. of Water Resour. Planning and Management*, ASCE, **121**: 98-108.

Woo, M.K. and Wu, K., 1989, Fitting annual floods with zero flows. *Can. Water Resour. J.*, **14**, 10-16.

Weglarczyk, S., Strupczewski, W.G. and Singh, V.P., 2005, Three-parameter discontinuous distributions for hydrological samples with zero values, *Hydrologic Processes*, 19, 2899-2914, DOI: 10-1002/hyp.5787.

Figures

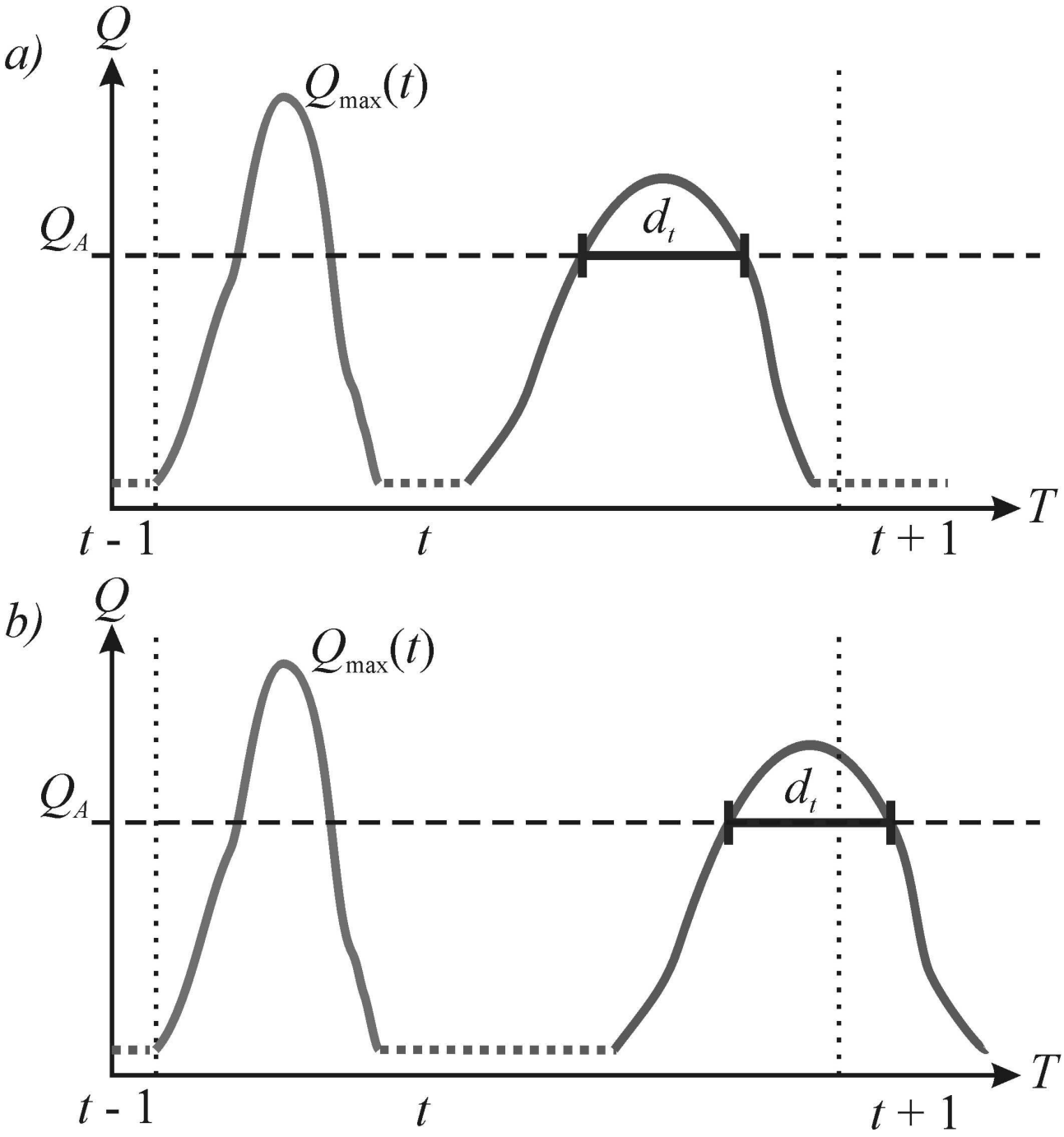

Fig. 1 Definition of the threshold flow discharge and duration in DqF model:
a) the flood wave of *dt* duration entirely in the year *t*;
b) the flood wave starts in the year *t* and continues in *t* + 1.

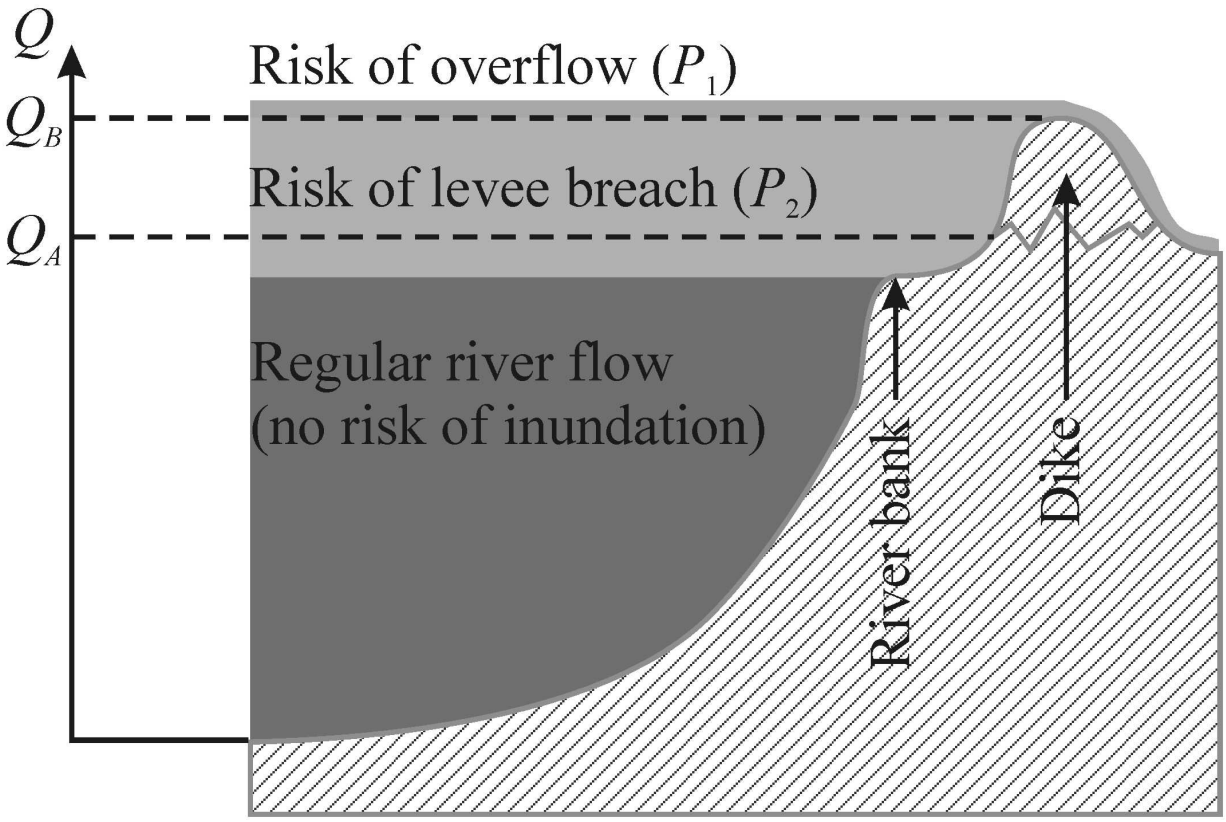

Fig. 2 Two reasons of inundation – an illustration.

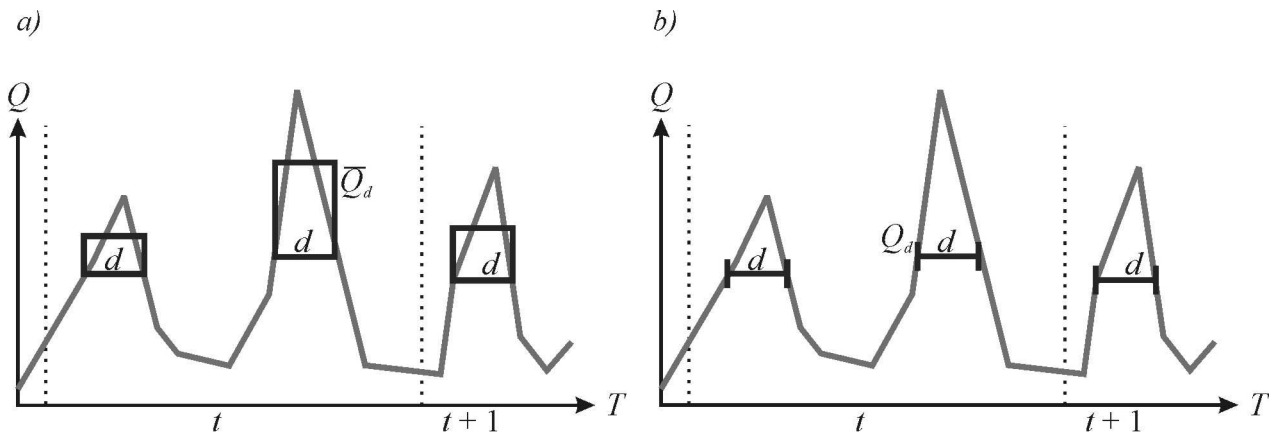

Fig. 3 Definition of the random variables in the QdF models:
a) the mean maximum *d*-days flow,
b) the annual maximum flow discharge (Qd) continuously exceeded during the period *d*.

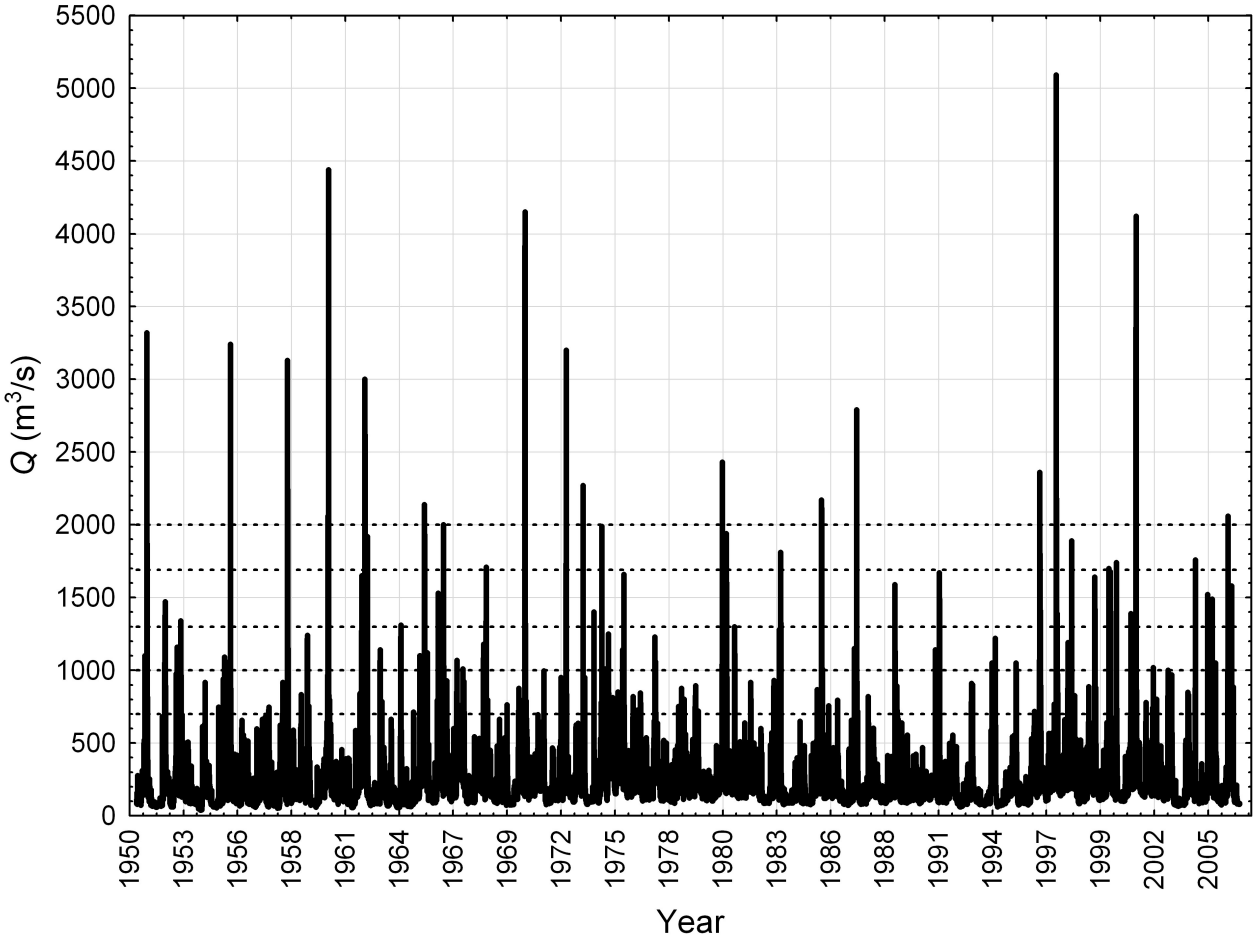

Fig. 4 Hydrograph of the daily flows at the Szczucin gauging station. Horizontal dashed lines reflect the Q_{Tr} values used in this study.

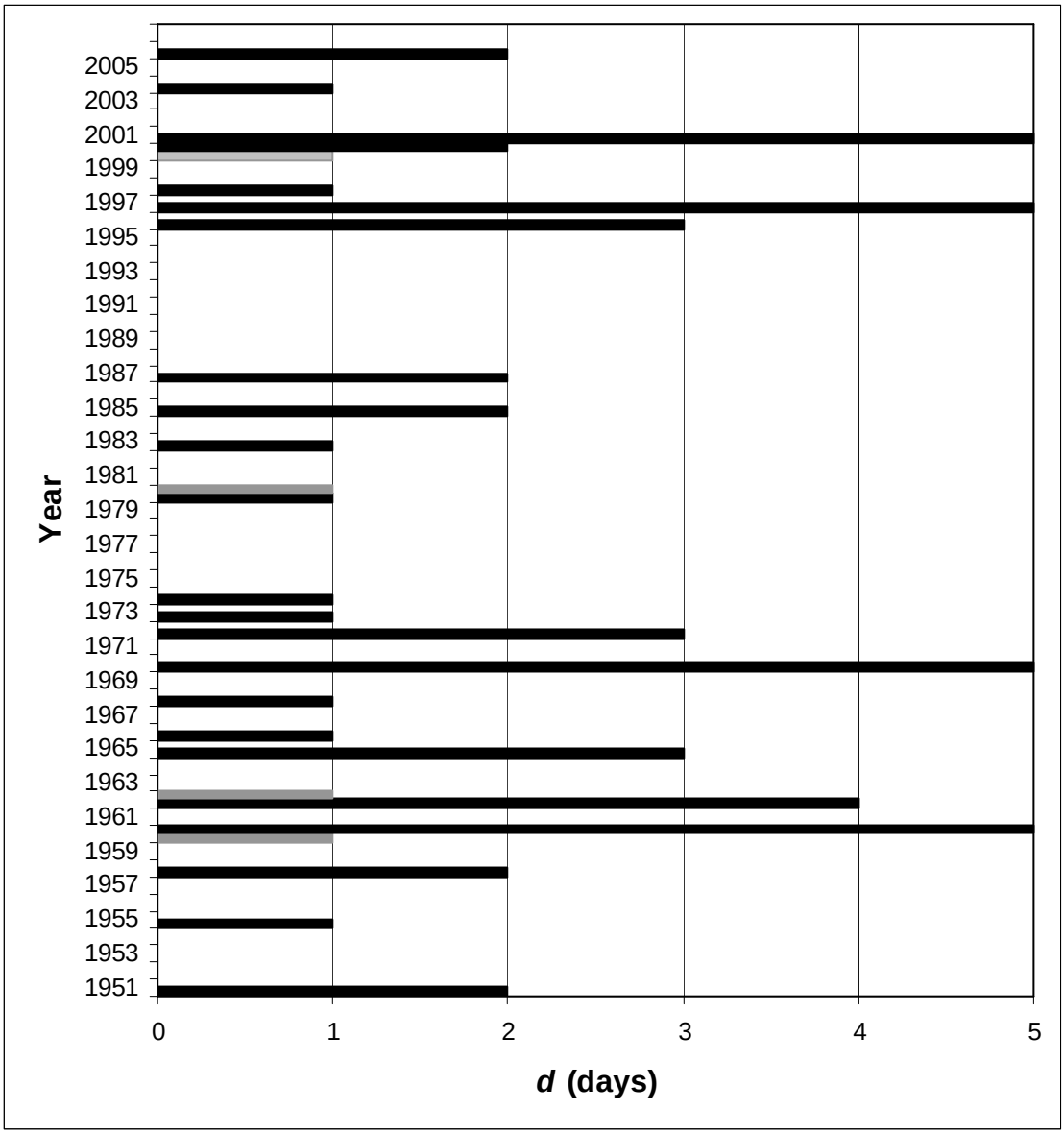

Fig. 5. The durations (in days) of the discharge above Q_A = 1690 m^3/s for Szczucin gauging station (1951–2006). The annual maximal durations are in black.

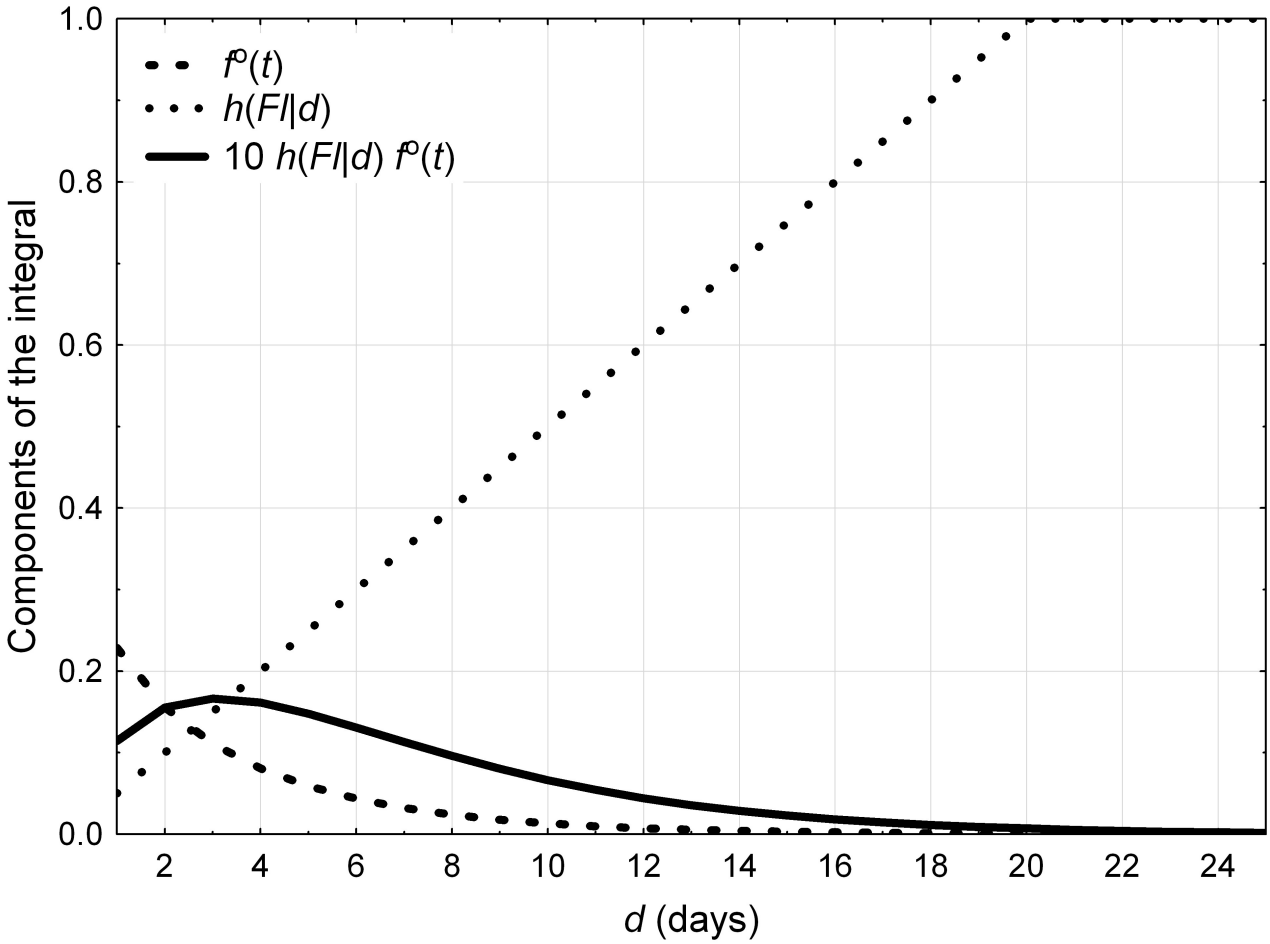

Fig. 6 Components of the integral Eq.(34).

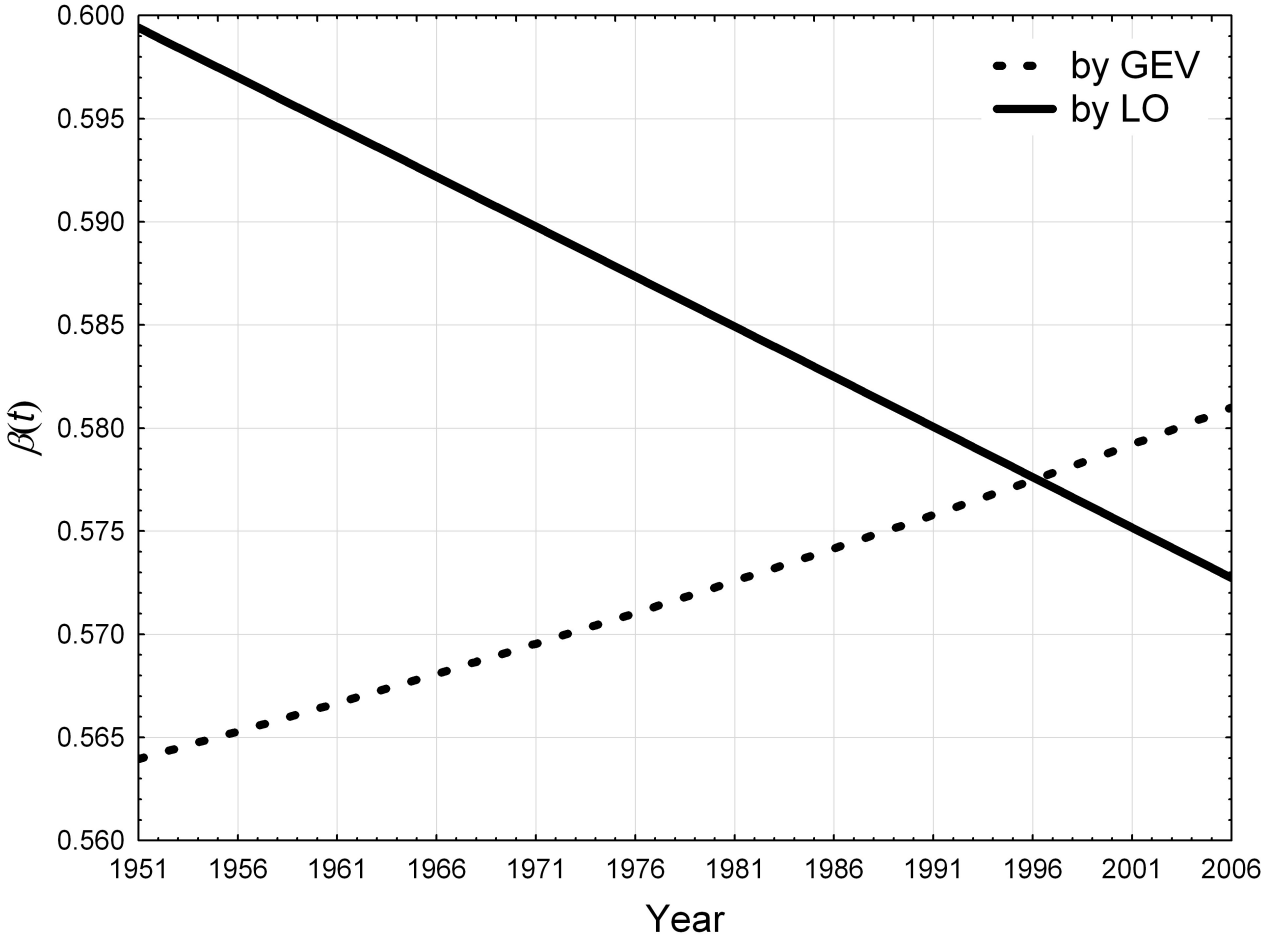

Fig. 7 Non-stationary $\beta(t)$ by two approaches

Tables

Table 1. Distribution functions recommended as $f^{o}(d_i; \mathbf{g})$ model.

Distribution name	Probability density function	Parameters	Equation nr:
Exponential (Ex)	$f^{o}(d;\alpha)=\frac{1}{\alpha}\exp(-d/\alpha)$	α – scale	(11)
Weibull (We) distribution	$f^{o}(d;\alpha,b)=\frac{b}{\alpha}\left(\frac{d}{\alpha}\right)^{b-1}\exp(-d/\alpha)$	α – scale $b > 0$ – shape	(12)
Generalized Pareto (Pa)	$f^{o}(d;\alpha,k)=\frac{1}{\alpha}\left(1-\frac{k}{\alpha}d\right)^{1/k-1}$	$\alpha > 0$ – scale $k < 0$ – shape	(13)
Generalized Exponential (GE)	$f^{o}(d;\alpha,\gamma)=\frac{\gamma}{\alpha}\exp(-d/\alpha)\left[1-\exp(-d/\alpha)\right]^{\gamma-1}$	$\alpha > 0$ – scale $\gamma > 0$ – shape	(14)
Gamma (Ga)	$f^{o}(d;\lambda,\alpha)=\frac{1}{\alpha^{\lambda}\Gamma(\lambda)}d^{\lambda-1}e^{-(d/\alpha)}$	$\alpha > 0$ – scale $\lambda > 0$ – shape	(15)

Table 2. The parameters of the two-component DqF model for Szczucin data.

	First component (by two methods)			**Second component, f^{o} is the two-parameter Generalised Exponential**		
Q_{Tr}	n_2	$\beta = n_1/n$ by Eq.(8)	$\beta = \beta(Q_{Tr})$ by Eq.(9)	*scale*	*shape*	$\ln ML/n_2$
700	51	0.089	0.076	2.8799	0.2938	–2.63
1000	40	0.286	0.226	4.0392	0.5228	–2.10
1300	32	0.429	0.395	4.8616	0.7464	–1.77
1690*	*23*	*0.589*	*0.577*	*3.4238*	*0.8357*	*–1.62*
2000	17	0.696	0.683	3.7411	0.9126	–1.54

* $Q_{Tr} = Q_A$