

FLOOD FREQUENCY ANALYSIS OF LITHUANIAN RIVERS

Diana Sarauskiene¹, Jurate Kriauciuniene²*Laboratory of Hydrology, Lithuanian Energy Institute, Breslaujos 3, LT-44403 Kaunas, Lithuania.**E-mail: ^{1,2}eko@mail.lei.lt*

Abstract. Numerous hydrological data derived from water measurement stations enables us to perform direct statistical analysis of river runoff characteristics. Recently Lithuanian hydrologists pay great attention to analysis of spring flood, as a very special hydrological regime phase. Many studies concerning the main spring flood parameters, their spatial and temporal variability, timing trend analysis, climate change impact on spring flood events are already performed. In this study a probabilistic behavior of floods was studied. Analysis of different flood frequency distributions using long-term data of 32 water measurement stations was performed in order to estimate the best fit distributions for Lithuanian conditions. For evaluation of climate change impact, the assessment of flood frequency distributions for different periods was accomplished.

Keywords: spring flood, flood frequency, probability distributions, climate change

Introduction

Recently significant attention is paid to analysis of spring flood, as a very special hydrological regime phase. Since the 1990s extensive and repeated floods across Europe have been bringing a lot of damage and loss (Handmer 2001, Barredo 2009). That shows failure of modern specialists' efforts to control this natural element, despite the huge money that is spent for flood prevention measures. An ability to predict extreme flood magnitude and behavior is very important for management of flood control, prevention of present and planned hydrotechnical structures, hydropower stations and dams, bridges and culverts. To make reliable estimates of expected extreme flood events, flood frequency estimation techniques are used: series of observed events are analyzed to assume a probabilistic behavior, which is then extrapolated to provide estimates of the likely magnitude of future extreme.

Historical flood flow data for probabilistic modeling are widely used all over the world. The United Kingdom has a set of procedures for flood frequency estimation adopted by most UK regulatory authorities as standard procedures to be used for flood frequency analysis (FFA) and design flood estimation in the UK (the initial statistical procedures were published in 1975: Natural Environment Research Council). The UK Flood Studies Report inspired the nationwide research project VA.PI ("VALutazione PIene", Flood Estimation) in Italy. It was a special program promoted to assess flood discharge values for different return periods in all major river basins in Italy. Researchers from France participated in the

French National research project ExtraFlo that aimed to perform a thorough comparison between FFA approaches currently used in France, based on an extensive dataset of long series of rainfall and runoff (Lang *et al.* 2010). Professionals from the mentioned countries as well as researchers from other European countries have joined the COST Action FloodFreq (European procedures for flood frequency estimation) in order to share their experience and improve their knowledge concerning FFA.

Many flood frequency studies are performed by scientists involved in other nationally funded research projects or by individually working research groups. Great attention to at-site FFA and regional FFA is given in other European countries (Czech Republic – e.g. Blazkova and Beaven 2002, Austria - Bloschl and Merz 2008, Spain - Mediero and Jiménez 2007, etc.) and other continents - Australia (Vogel *et al.* 1993, Peel *et al.* 2001), the United States (Vogel and Wilson 1996, Stedinger and Griffis 2008), African countries (Kjeldsen *et al.* 2002, Abida and Ellouze 2008), etc.

Main spring flood parameters (flood duration, runoff volume and maximum discharges), their spatial and temporal variability, timing trend analysis, climate change impact on spring flood events have already been studied by Lithuanian hydrologists (Gailiūšis *et al.* 2001, Ascila *et al.* 2002, Simaitytė 2007, Meilutytė-Barauskiene and Kovalenkoviene 2007, Dumbrasukas *et al.* 2008, Meilutytė-Barauskienė *et al.* 2010 etc.). Numerous hydrological data derived from water measurement stations enables us to perform direct statistical analysis of river runoff characteristics.

Hydrologists from Lithuanian Energy Institute calculated parameters of probability curve, maximum discharge values of different probability floods for the most of Lithuanian rivers (Gailiusis *et al.* 2001). Only few studies have been made concerning flood frequency distributions in Lithuanian rivers. We do not have any national procedures for flood frequency estimation as well.

Probabilities and return periods of extreme flood peaks of the Nemunas at Kaunas were studied (Simaitytė 2007). A long time data analysis for the Nemunas River was performed to better understand flood processes in the river delta; several probability distributions were compared and selected the one that gave the best results to the extreme values of discharges of the Nemunas (Ascila *et al.* 2002).

In this study an attempt is made to analyze spring floods in Lithuanian rivers. That is one of the first studies that focus on probability distribution fitting to the actual spring flood data in the rivers. The aim was to find the best fit probability distributions for the spring flood data in the rivers from the different hydrological regions, as well as to define the best probabilistic models for the data of different periods in order to find out the climate change impact on the spring flood data distribution.

Data

Daily discharges from 32 water measurement stations (WMS) were used for flood frequency analysis. The water measurement stations are distributed in three hydrological regions of Lithuania (Fig 1): 1 – Western

Lithuania, 2 – Central Lithuania, 3 – Southeastern Lithuania (Gailiusis *et al.* 2001). Additionally the hydrological data from 5 WMS which are on the biggest rivers of Lithuania (the Nemunas and the Neris) were investigated (Table 1). Duration of the discharge data observations is different and ranges from 38 to 88 years. The data of discharge of the Nemunas River in Smalininkai WMS exists from 1812. These data series are among the longest time series in Europe and are very valuable.

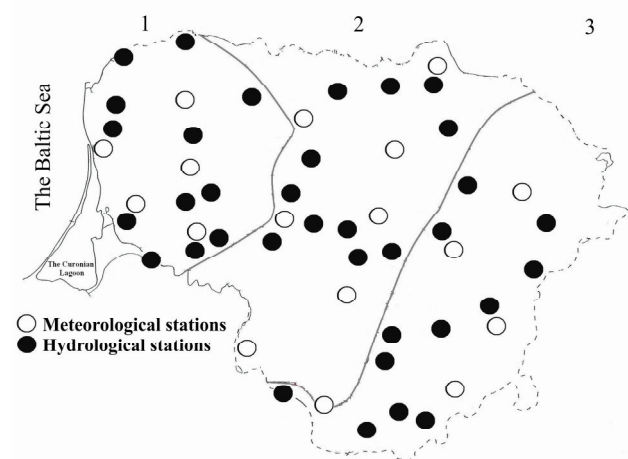


Fig 1. Distribution of hydrological stations in three hydrological regions (1 –Western, 2 – Central, 3 – Southeastern Lithuania)

Table 1. Water measurement stations, periods of observations and results of probability fitting test for flood discharges (EV1- Gumbel, GEV - generalized extreme value, LP3 - log-Pearson type 3, LN - lognormal and LN3 - three-parameter lognormal distribution)

No.	River	WMS	Period of observations	Probability distributions				
				Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Western Lithuania								
1	Venta	Leckava	1949–2008	LN	LP3	LN3	GEV	EV1
2	Bartuva	Skuodas	1957–2008	LN3	LP3	GEV	EV1	LN
3	Venta	Papilė	1949–2008	LN3	LP3	GEV	LN	EV1
4	Minija	Kartena	1925–2008	LN	LP3	LN3	GEV	EV1
5	Rešketa	Gudeliai	1946–1999	LN	LN3	LP3	GEV	EV1
6	Veiviržas	Mikužiai	1954–1999	LN3	GEV	LP3	EV1	LN
7	Akmena	Paakmenis	1955–2008	LP3	GEV	LN3	EV1	LN
8	Jūra	Pajūris	1946–1999	LN	LP3	LN3	GEV	EV1
9	Šyša	Jonaičiai	1961–1999	EV1	GEV	LN3	LP3	LN
10	Jūra	Tauragė	1926–2008	GEV	LN3	LP3	EV1	LN
11	Šešuvis	Skirgailiai	1939–2008	LN	GEV	LN3	LP3	EV1
Central Lithuania								
12	Agluona	Dirvonakiai	1961–1999	LN3	LP3	GEV	LN	EV1
13	Mūša	Ustukiai	1958–2008	LN3	LP3	GEV	LN	EV1
14	Lėvuo	Kupiškis	1955–1999	LP3	GEV	LN3	LN	EV1
15	Dubysa	Lyduvėnai	1933–2008	LN3	LN	LP3	GEV	EV1
16	Alsa	Paalsys	1957–1999	GEV	LP3	EV1	LN3	LN
17	Šušvė	Josvainiai	1940–2008	LP3	EV1	LN3	GEV	LN
18	Dubysa	Padubysys	1930–1999	LP3	GEV	LN3	EV1	LN
19	Nevezis	Dasiūnai	1961–2005	LP3	GEV	LN3	LN	EV1

Southeastern Lithuania								
20	Šventoji	Anykščiai	1928–2008	LN3	LP3	GEV	LN	EV1
21	Žeimena	Kaltanėnai	1961–2001	LP3	GEV	LN3	LN	EV1
22	Šventoji	Ukmergė	1925–2008	LN	LN3	LP3	GEV	EV1
23	Žeimena	Pabradė	1954–2008	GEV	LP3	LN3	LN	EV1
24	Verknė	Verbyliškės	1952–2008	GEV	LN	LP3	LN3	EV1
25	Šešupė	Kalvarija	1954–2004	EV1	GEV	LP3	LN	LN3
26	Ūla	Zervynos	1960–2008	LP3	LN	GEV	LN3	EV1
27	Merkys	Puvočiai	1945–2008	LN3	LP3	GEV	LN	EV1
Biggest Lithuanian rivers								
28	Neris	Jonava	1920–2008	LN3	LP3	GEV	LN	EV1
29	Nemunas	Smalininkai	1812–2008	EV1	GEV	LP3	LN	LN3
30	Neris	Vilnius	1923–2008	LP3	LN3	GEV	LN	EV1
31	Nemunas	Nemajūnai	1920–2008	LP3	GEV	LN3	LN	EV1
32	Nemunas	Druskininkai	1945–2008	LP3	GEV	LN3	LN	EV1

In order to evaluate climate change impact on spring flood events flood frequency analysis was made for different periods. Flood frequency distributions were estimated for the period of 1961-1990 (selected as a reference period), the period after the reference period, as well as for the whole observation data series (durations depending on data availability).

Methods

There are numerous various probability distributions (PD) used in hydrological modeling. Many different studies were carried out in order to find out which PD represents flood frequency data best in the selected regions. In this study a comparison of 5 commonly used probability distributions was performed. Gumbel (EV1), generalized extreme value (GEV), log-Pearson type 3 (LP3), lognormal (LN) and three-parameter lognormal (LN3) were selected for the analysis statistical characteristics of observed flood data of Lithuanian rivers.

Generalized extreme value and *Gumbel* distributions are extreme value distributions widely used for modeling of extreme or rare events such as extreme floods and snowfalls, high wind speeds, extreme temperatures. The *Gumbel* (EV1) distribution, also known as the *Extreme Value Type I* distribution, is unbounded (defined on the entire real axis). The *Generalized Extreme Value* (GEV) distribution is a flexible three-parameter model that combines the Gumbel, Fréchet, and Weibull *maximum* extreme value distributions.

The *log-Pearson Type III* (LP3) is mostly used for annual maximum floods. In the USA, it is a standard distribution for flood frequency analysis.

The *three-parameter lognormal* (LN3) and *lognormal* (LN) distributions are popular in the studies of the frequency analysis of flood as well.

Probability density functions for all studied probability distributions can be found on MathWave Technologies internet site (<http://www.mathwave.com>).

The Kolmogorov-Smirnov test was selected for determination of the best distribution. This test is used to determine if a sample comes from a hypothesized

continuous distribution. It is based on the empirical cumulative distribution function (CDF). The empirical CDF is denoted by (<http://www.mathwave.com>)

$$F_n(x) = \frac{1}{n} \cdot [\text{Number of observations} \leq x] \quad (1)$$

The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical F(x) and the empirical Fn(x) cumulative distributions functions. It is defined as

$$D = \max_{1 \leq i \leq n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \quad (2)$$

The hypothesis regarding the distributional form is rejected at the chosen significance level (*alpha*) if the test statistic, D, is greater than the critical value obtained from a table.

Product of MathWave Technologies – EasyFit 5.4 with the built-in distribution viewer StatAssist was used for probabilistic modeling of flood data of the selected rivers.

Results

Five probability distributions were used as flood frequency models for probabilistic analysis of spring floods in the selected Lithuanian rivers. The aim of the goodness of fit test was to measure the "distance" between the data and the tested distribution, and compare that distance to some threshold value. If the distance (the test statistic) was less than the threshold value (the critical value), the fit was considered good. The critical values depend on the sample size and the significance level chosen ($\alpha = 0.05$). The distribution with the lowest statistic value is the best fitting model.

Table 1 illustrates how the annual maximum discharge data of spring floods of the whole observation period fitted to the selected distributions. The primary analysis of fitting results showed that log-Pearson type III distribution provided the best approximation to the distributions of river flood data in the Southeastern and

Central regions, whereas the best fitting PD for the spring flood data in Western Lithuania was three-parameter lognormal distribution.

In order to determine the dominant probability distributions (PD) for all Lithuanian rivers, the defined PDs were ranked according to the Kolmogorov-Smirnov test (Table 2). The best fitting PD received rank 1, and the worst – 5. Comparison of the summed up PD ranks revealed that log-Pearson type III distribution provided the best approximation to the distributions for the whole observation periods (indicated in Table 1) of the studied rivers. Data sets generated by Gumbel distribution were the least similar to the actual flood data. Unfortunately, reliable conclusions cannot be made, because the lengths of observation data series were not equal (Table 1).

Table 2. Summary of ranks of distributions for the tested WMS data (the whole observation period) according to the Kolmogorov-Smirnov test

PD	Number of WMS that received rank:					Total of ranks
	1	2	3	4	5	
LP3	10	12	8	2	0	66
LN3	9	4	14	3	2	81
LN	6	3	0	15	8	112
GEV	4	12	9	7	0	83
EVI	3	1	1	5	22	138

In order to make the selected data more comparable, spring data of the 3 regions was grouped to different periods: 1961-1990 (used as the reference period), 1961-2008, and 1991-2008 (Table 3). The summing-up of the ranks defined that for 1961-1990 data series LP3 best fitted for water measurement stations of the Western region and GEV – for the rest of water measurement stations. The Western region is close to the Baltic Sea,

Table 3. Total number of PD ranks for the selected periods in the hydrological regions of Lithuania

Periods	Western Lithuania					Central Lithuania				
	LP3	LN3	GEV	LN	EVI	LP3	LN3	GEV	LN	EVI
1961-1990	22	30	26	45	42	19	19	15	36	31
1961-2008	24	27	25	46	43	18	20	17	34	31
1991-2008	19	17	16	27	26	6	12	6	8	13
Periods	Southeastern Lithuania					Nemunas and Neris				
	LP3	LN3	GEV	LN	EVI	LP3	LN3	GEV	LN	EVI
1961-1990	21	26	9	36	28	11	17	10	19	18
1961-2008	12	21	18	33	31	12	12	10	16	25
1991-2008	16	11	10	27	26	12	9	9	25	20

Detected different probability distribution models of flood discharge series in all three hydrological regions for the last period (1991-2008) mean that recently character of flood probability distribution changes. That is a consequence of significant changes of the climatic parameters during the last decades.

The analysis of regional series of temperature, precipitation and river discharge data in the three hydrological regions (Fig 1) revealed that the recent

the rivers in this region distinguish for significant spring runoffs, as well as maximum discharges, resulting in the rain floods and often exceeding discharges of spring floods; the main source of these river feeding is precipitation, that makes 40-70% of annual runoff (Gailiūšis *et al.* 2001). The mentioned features of the marine type climate may cause a different frequency distribution pattern in this region. For the period of 1991-2008 GEV was the probability distribution that received the highest ranks in all regions; and for the data of 1961-2008 LP3 revealed the best fitting in Western and Southeastern rivers, whereas LP3 and GEV had the same ranks in the rivers of Central Lithuania. The same maximum discharge data distribution in the whole country in 1991-2008 may be the result of recent significant changes of hydrometeorological conditions. According to the studies of meteorological data changes in Lithuania (Kriaučiūnienė *et al.* 2008), since 1990 in all regions similar increase of air temperature was observed, whereas the changes of precipitation had high variability in space and time. During the last sixteen year period (1991-2006) increasing of precipitation in Southeastern Lithuania was 34 mm, decreasing in Western and Central regions - 41 and 18 mm comparing with the reference period (1961-1990) (Kriaučiūnienė *et al.* 2008). It could be suggested that during the period of 1991-2008 the conditions of spring flood formation became more similar in the different hydrological regions of Lithuania. LN and EVI probability distributions seemed to be the least suitable for the modeling of the analyzed spring flood data.

Goodness of fit tests of the Nemunas and the Neris flood discharge data revealed that for the all described periods GEV probability distribution showed the highest similarity to the observed data.

observation period of 1991-2008 in comparison with the reference period can be characterized by increased annual (0.9 °C) and seasonal temperatures, by considerably (10-29%) increased precipitation in winter season and as a result – by higher river runoff in winter. Such climatic changes influenced a distribution of the whole river runoff during a year, as well as maximum flood discharges. Changed patterns of the climate parameters influenced changes in spring flood data distribution. Due

to warmer winters with greater amount of precipitation, spring floods tended to begin earlier, lasted longer and in the result – their peaks were less. In the most cases the detected best fit probability distributions for the analyzed individual rivers also changed.

Estimation of probability distribution is necessary in order to forecast the change in frequency of flood events in the future. The PD analysis of flood data showed the best match of generalized extreme value distribution for the empirical spring flood data in all hydrological regions (as well as in the biggest rivers) in 1991-2008. Therefore it can be concluded that recently, in the period of significant climatic changes, GEV probability distribution is the most suitable distribution for flood frequency analysis and for prediction of flood maximum discharges in Lithuanian rivers.

Comparison of the Nemunas flood data showed that in the last period (1991-2008) maximum discharges were about 30% less than in the reference period at Smalininkai, 32% at Druskininkai and 29% at Nemajunai WMS. Big differences of statistical parameters are found in series of maximum discharges in the Nemunas at Smalininkai for the studied periods. Significant variations of the mean, minimal and maximum annual flood discharges are detected in different periods. Flood discharges became significantly smaller in 1991-2008. Figure 2 shows changes of statistical parameters of spring flood data.

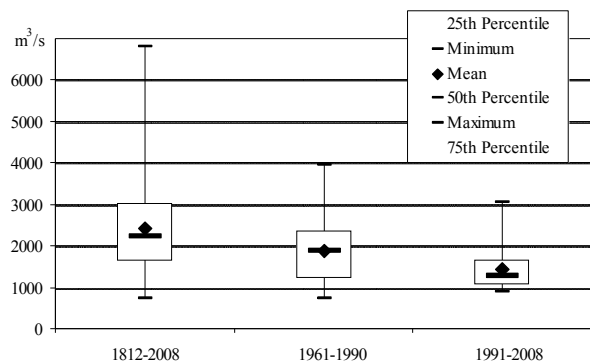


Fig 2. Box plot of statistical parameters of spring floods in the Nemunas at Smalininkai of the three studied periods

Changes of statistical parameters of spring flood data reflect in changes of probabilistic behavior of simulated flood discharge data. Spring flood data in the Nemunas at Smalininkai were best characterized by different extreme value distributions: the whole observation period (1812-2008) by EV1 and both reference (1961-1990) and the recent period (1991-2008) by GEV probability distribution. Figure 3 displays the theoretical probability density functions of the fitted distributions for the different observation periods. Probability distribution function for the data of the last period diverges the most.

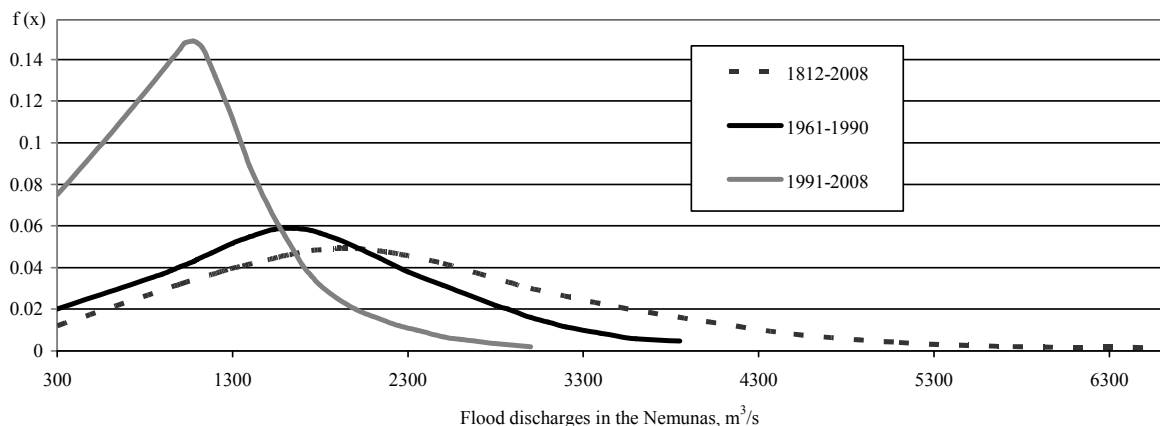


Fig 3. The theoretical probability density functions of the fitted distributions for the flood data of the Nemunas in the different observation periods

Conclusions

Analysis of different flood frequency distributions using long-term data of Lithuanian river runoff was performed in order to estimate the best fit distributions for Lithuanian conditions. The analysis revealed that spring flood data of individual rivers, as well as in the different regions, tended to have different best fit probability distributions for the studied time periods.

The summing-up of the ranks revealed that in 1961-1990 (reference period) log-Pearson type 3 probability

distribution best fitted actual data of spring flood in the Western region and GEV – for the Central and Southeastern regions; in the period of 1991-2008 GEV was the probability distribution that received the highest ranks in all regions.

Recently, in the period of significant climatic changes (1991-2008), GEV probability distribution is the most suitable distribution for flood frequency analysis and could be used to predict of flood maximum discharges in Lithuanian rivers. This probability

distribution could also be used for simulating of flood frequency distribution in ungauged rivers.

LN and EVI probability distributions seemed to be the least suitable (out of the five studied) for the modeling of the analyzed spring flood data.

Acknowledgments

The research described in this paper was supported by COST action ES0901 European procedures for flood frequency estimation.

References

- Abida, H.; Ellouze, M. 2008. Hydrology and Earth System Sciences Probability distribution of flood flows in Tunisia. *Hydrology and Earth System Sciences*, 12: 703–714.
- Ascila, R.; Dumbrasukas, A.; Punys, P. 2002. Hydrological aspects of flooding of the Nemunas river delta. *Transactions of the Lithuanian University of Agriculture and Lithuanian Institute for Water Management*, 18(40): 3–12.
- Barredo, J. I. 2009. Normalized flood losses in Europe: 1970–2006. *Natural Hazards and Earth System Sciences*, 9: 97–104.
- Blazkova, S.; Beven, K. 2002. Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty). *Water Resources Research*, 38, 1139, doi:10.1029/2001WR000500.
- Bloschl, G.; Merz, R. 2008. Estimating flood risk in Austria. *Slovak Journal of Civil Engineering*, 3: 13–20.
- Dumbrasukas, A.; Bagdžiūnaitė-Litvinaitienė, L.; Vyčienė, G. 2008. Trend detection in hydrological series of main Lithuanian rivers. *The 7th International conference "Environmental engineering": selected papers*, May 22–23, 2008 Vilnius, Lithuania. 2: 508–514.
- Gailiūšis B.; Jablonskis J.; Kovalenkoviėnė M. 2001. Lietuvos upės. Hidrografija ir nuotėkis [The Lithuanian Rivers. Hydrography and Runoff]. Kaunas: LEI.792 p. ISBN 9986-492-64-5.
- Handmer, J. 2001. Improving flood warnings in Europe: a research and policy agenda. *Environmental Hazards*, 3: 19–28.
- Kjeldsen T.R.; Smithers J.C.; Schulze R.E. 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index flood method. *Journal of Hydrology*, 255(1-4): 194–211.
- Kriauciuniene J.; Meilutyte-Barauskienė D.; Kovalenkoviėnė M. 2008. Regional Series of temperature, precipitation and runoff for Lithuania. XXV Nordic Hydrological Conference. Reykjavik, Iceland August 11–13, 2008. NHP Report 50 (2): 638–645.
- Lang, M.; Renard, B.; Kochanek, K.; Sauquet, E.; Garavaglia, F.; Paquet, E.; Soubeyrou, J.; Jourdain, S.; Veyssiere, J.; Borch, F.; Neppel, L.; Najib, K.; Arnaud, P.; Aubert, Y.; Auffray, A. Data-Based Comparison of Frequency Analysis Approaches: Methodological Framework and Application to Rainfall / Runoff Data in France American Geophysical Union, Fall Meeting 2010, abstract #NH11A-1117.
- Mathwave: Data Analysis and Simulation [online]. [Viewed on January 3, 2011]. Available on the Internet: <<http://www.mathwave.com>>Mediero, L. and Jiménez, A. 2007. Regional analysis for frequency estimation of annual flood peaks in ungauged basins of Spain. *Geophysical Research Abstracts*, 9, 03251.
- Meilutyte-Barauskienė, D.; Kovalenkoviėnė, M. 2007. Change of spring flood parameters in Lithuanian rivers. *Energetika*. 2: 26–33.
- Meilutyte-Barauskienė, D.; Kriauciūnienė, J.; Kovalenkoviėnė M. 2010. Impact of climate change on runoff of the Lithuanian rivers. *Modern climate change models, statistical methods and hydrological modeling*. Saarbrücken, Germany: LAP LAMBERT Academic Publishing, 55 p. ISBN 978-38383-5833-8.
- Natural Environment Research Council. 1975. *Flood Studies Report*, 5 Volumes + maps, London, UK.
- Peel, M.C.; Wang, Q.J.; Vogel, R.M.; McMahon, T.A. 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*, 46(1): 147–155.
- Simaitytė, J. 2007. Risk-based decision making for flood discharge control. *Doctoral Dissertation*, Kaunas.
- Stedinger, J.R.; Griffis, V.W. 2008. Flood Frequency Analysis in the United States: Time to Update. *Journal of Hydrologic Engineering*, 13(4): 199–204.
- Vogel, R.M.; McMahon, T.A.; Chiew, F.H.S. 1993. Floodflow frequency model selection in Australia. *Journal of Hydrology*, 146: 421–449.
- Vogel, R. M.; Wilson, I. 1996. Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *Journal of Hydraulic Engineering*, 1: 69–76.