AIOC'16

All In One Conferences

Proceedings Book

2nd International Conference on Advances in Statistics

Uncertainties in flood event estimation

Jadran Berbić[a], Neven Kuspilić[b],Eva Ocvirk[b]

[a]*PhD candidate at The Faculty of Civil Engineering, University of Zagreb*
[b]*Kačićeva 26, The Faculty of Civil Engineering, University of Zagreb*

Abstract

Flood event probability is of great interest, both in estimating flood risk and in decision making about design of structures for flood protection. Basic uncertainties classification is on natural and epistemic uncertainties. Concretely, hydrological, hydraulic, statistical and geotechnical uncertainties play role in real value of flood event probability, each of these having two basic concepts of uncertainties involved in it. Climate change and global warming could play significant role in future floods, bringing more uncertainty in statistical description and hydrological processes of water protection systems. Review, classification and description of those uncertainties are given in the paper.

Keywords: flood event; probability; uncertainties

1. Introduction

In decision making about hydraulic structures, that is, flood protection structures, for given solutions of the problem, usually the range of different scenarios and their risk is analyzed. It is quite difficult to say what is really going to happen, and decision makers can only make decisions under uncertainty, knowing only the probabilities and the range of reliability for each possible state of flood protection system. Some uncertainties can be quantified, some cannot. Rational use of the term uncertainty includes the quantification of those.[1] When it is about flood protection structures, the absence of absolutely safe system is the valuable thing to have in mind. So, there always exists dealing with the event causing possible damages, and rational deliberation about it includes defining its probability p within the given range of uncertainty, that is, $p \pm \Delta p$. [1]

The uncertainties included in water resources management can be distinguished in data uncertainties, model uncertainties and technological uncertainties. Data uncertainty is usually said to be the main uncertainty driver and the greatest part of the uncertainty influencing the flood probability estimation. [2] Considering the relatively young hydrological history of, depending of the area, 50-100 years in estimating the events with several times greater return periods, it seems quite convenient to corroborate previous sentence. [3,4]. Not that just hydrological history is relatively young and not filled with longer period of measurements, but various imperfections in measurings, errors done by humans, irregularly calibrated or poorly maintained equipment, inadequate sampling etc. also contributes to the amount of uncertainties. Though models are just simplified representation of complex systems and always have some aspect of uncertainty involved whether through describing the system or through calibration, validation or wrong interpretation of system's nature. [1]

Merz and Thieken (2005) separate natural (inherent, intrinsic) and epistemic uncertainty in the flood frequency analysis. Epistemic uncertainty arises from lack of understanding, measuring and describing the system, its phenomenons and bonds between inputs and outputs.[5] The main flood initiator, without which floods in fact can not occur, are rainfalls and snow melting. As these phenomenons have numerous cause-consequence connections with the whole sequence of different processes in atmosphere, the use of stochastics becomes unavoidable the probabilistic analysis of floods. It is hard, not to say impossible, to engage all of the factors influencing the formation of a flood event. Previously mentioned cause-consequence connections also include the part belonging to the natural uncertainty. Additionally, possible unexpected damages on levees, dams, lack of design and monitoring, increase the epistemic uncertainty making the modeling of probability and reliability of flood structures more challenging. The epistemic uncertainty can be reduced.[5] Natural uncertainty arises from variability of natural phenomenons. Basically, they are induced by climatic, atmospheric, hydrological variations, but also by unconsistencies in levee and foundation ground properties. Principally, the natural uncertainty can not be reduced.[5] In the way of controlling the material properties and choosing the certain places for material excavation, the aspect of unconsitence of building material can be partially bypassed.

Ranzi et al. (2012) include climate change as category of uncertainty beside hydrological, hydraulic, geotechnical and climate change [6]. In the paper the classification of uncertainties which is used is division into statistical, geotechnical, hydraulic and hydrological uncertainty, as they seem to be obvious categories influencing the estimation of high water event probability, but up with having in mind climate variations could influence every of those. The objective of the paper is to notice, describe and classify those uncertainties and their components, and give recommendations about their treatment. The importance of given matter reflects in the risk assessment, highly determined by probability of set of scenarios, and designing the flood protection structures, where uncertainties, that is hydrological events with its reliability, should be indicated.

2. Statistical uncertainty

2.1. Introduction

Although there is the space for discussion about suitable approach for naming and describing this kind of uncertainty, the need and nature of designing flood structures and thus making the statistical analysis, requests usage of the methods and data which are on disposition. Statistical uncertainty can be classified in parameter and distribution uncertainty, the first one arising from unsuficient amount of data and inappropriate method of parameter estimation, and the second one arising from the choice of distribution type for fitting the data. [7] As the usual type of statistical analysis in water resources management is univariate statistics because of its simplicity and practicability, this chapter is written under that assumption. Statistical methods in hydrology, as usual, can be applied if the following conditions are satisfied: sequence is made from random variables, variables are mutually independent, sequence is homogenous, stationary and long enough. [8,9] In flood frequency analysis there are two different approaches that are used for estimating the probability of flood event, that is, return period – annual maximum approach (AMA) and threshold exceedance approach (TEA). In AMA hydrological data is taken in the measuring period of at least 30 years and for every year peak flows are taken into statistical analysis. Therefore, the return period $T(Q)$, in years, of the certain value of flow Q is calculated as follows:

$$T(Q) = \frac{1}{P(q \geq Q)} \tag{1}$$

where $P(q \geq Q)$ is the probability of exceedance of the flow Q.

In TEA, flows with values above the certain threshold are taken in the analysis. It is usually applied in the circumstances in the cases when available data is taken form the measuring period less than 30 years. The return period is calculated as follows:

$$T(Q)=\frac{N}{M}\frac{1}{P(q\geq Q)^*} \tag{2}$$

$$P(q\geq Q)^*=\frac{N}{M}P(q\geq Q) \tag{3}$$

where $P(q \geq Q)$ * is the probability of exceedance of the flow Q calculated using the total number of sequence members, N is the number of unit time intervals and M is the number of sequence members.

2.2. Using data for statistical analysis – discussion

Using the AMA approach theoretically could lead to choosing the flows close enough that some kind of natural dependence between those could be involved, but this is the possibility that rarely happens. Having in mind the year is quite rigid time border, the possibility of appearance of two timely closed high water waves connected with the same atmospheric event, one at the end of the year, one at the beginning of the next year, exists. In TEA, due to the usage of threshold, there is reasonable chance to have two or three mutually dependent events and while selecting the threshold, one high enough which excludes dependence should be chosen. Beside taken assumption of numerous causes influencing the flow in rivers, high water events are usually the consequence of the prevailing of rainfall event or snow melting in the certain period in the relation to other causes. So in both methods, as the high water events are mainly driven by one cause, there is a question of homogeneity involved in analysis. Although events could have similar and greatest peaks in analysis, the nature of high water event including its volume and duration is of interest. Still, flood structures dimensions and the intensity of the flood, is driven by those two parameters and the greatest peak does not necessarily has the greatest volume (in "very" small duration). Thus, in the manner of avoiding the previously said, statistical analysis can be made in events observing way, calculating the probability and return period for every certain high water event important with its significance.

It is implicitly understood that all events (elements) in space $\mathcal{V}$ must be mutually exclusive, equally distributed in space and deplete the space as much as possible. The way the second and third property are fulfilled is made through using the hydrological measurements on disposition. Moreover, without timely longer observations these properties can hardly be improved. As Hrelja (2007) stated that is supposed to have in mind the relatively young history of systematically taken hydrological measurements.[3] In that manner, even if the one certain type of hydrological event is taken, the real space Ω of naturally existing elements representing the same type of event, including those of smaller and those of higher magnitudes, is hardly known, and thus depleted. Analyzing the high water events and predicting those with great return periods (of 100, 1 000, 10 000 years) requires approximation of event distribution using some of the famous distributions (Gauss, Weibull, Pearson, Gamma, Galton and so on) and extrapolation beyond really taken measurements. Mentioned can take the expanded space of possible events $\mathcal{A}$ broader in the relation to the space of measured values $\mathcal{V}$, and even outside the borders of naturally possible elements, as it is shown on the figure 1. Using AMA and TEA, after the specific event with peak flow is chosen for analysis, another chosen event does not necessarily excludes the first one. As it can be seen on figure 2. where similar high water events are marked in the same way. There are displayed average daily flows for three years in a row. In AMA, events from November (1982), March (1983) and September (1984) are in the pot for analysis because in those year they have got the highest (hourly averaged as it is marked with line on figure) peak flow. This does not exclude the similar event in march/april (1982) from happening. That is, using the AMA approach, as the March (1983) event is included in analysis, similar event from previous year is not included. Also, the situation where November (1982) event had happened does not exclude that March/April (1982) event in the same year from happening, but it was not included in analysis as the similar event in march (1983) is included. In TEA, depending of the chosen threshold, this also could happen.

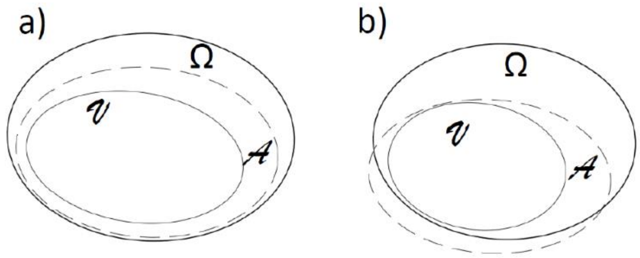

Fig. 1. Data extrapolation (a) extrapolation is kept in the space of all possible elements; (b) extrapolation goes beyond the space of all possible elements

Also, using AMA in this situation does not put the interesting event on the end of every year, which tends to have large magnitude, especially looking the volume of water (the area under the flow) in December (1982).Although the mentioned event is not the greatest of all events, it does not mean that sometimes in next years this event can not exceed all the other events in the sense of volume, no matter as high water event is not observed in measurements.

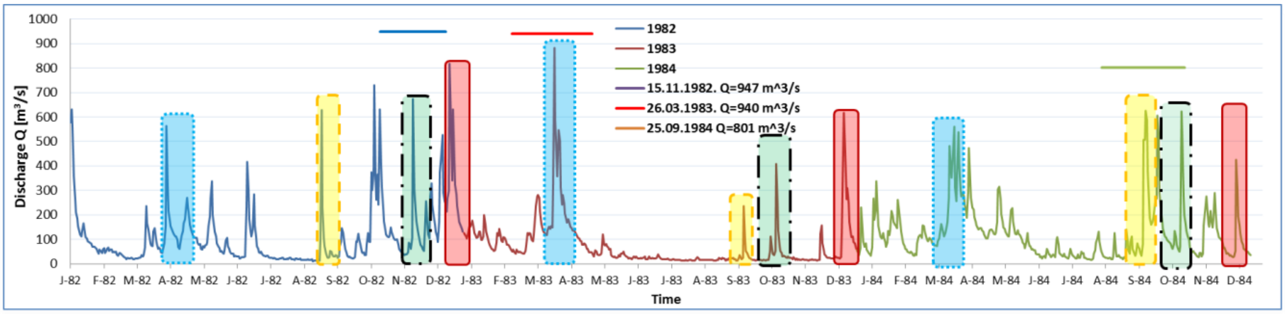

Fig. 2. Hydrograph of Brodarci station for years 1982-1984

As it can be seen, there are unresolved issues regarding the usage of the flows as basic information for statistical analysis of high water events. Still, it should be considered also in the manner of simplicity and practicability, as these methods really are. The epistemic uncertainty arises from defining the appropriate statistical approach to treat high water events. Moreover, analysis should include all events considered as high water events. [10] As volume and duration of the event, beside the flow, are also of the great importance, they have to be included in analysis. So the way these could be avoided is in observing the similar events in some specified time increments which will conduct to situation where the probability and return period for every different event could be estimated. Then, by using the Bayesian approach, probability that any of the events will happen could be estimated. Also, the approach with bivariate statistics, probability depending on volume and duration or flow, can be considered. On the other side, no matter all of this is fulfilled and the approach is perfect, from short period of measurements it is impossible to get the full information about any event. Natural uncertainty arises from the fact that statistical distribution of various hydrological events is susceptible to changes, due to reasons like climate variations, but also due to possible changes in river and basin environment. Thus, the sequence homogeneity and stationarity could be influenced through this type of uncertainty. Possible climate changes and global warming could affect this type of uncertainty. One of the definitions of climate changes says that they influence the statistical distribution of hydrological events.[4] The open question is will it affect the nature of high water events and floods. Flood damages tends to be higher, but the main reason lies in fact that material resources in the vicinity of flood protection structures tends to value more.[3] Due to the last IPCC synthesis report, increasing of the extreme rainfalls and river flows influence intensity and occurrence of floods in regional level, not in global level (IPCC).[11] So, with the temperature increasing the way it goes like today, it is quite possible to have changes in probability of high water events.

3. Geotechnical uncertainty

Flood event connected with geotechnical uncertainty has dual nature. As the first, flood can occur due to undersizing of the structure (levee or earthen dam) and thus the mechanisms which could potentiate the failure are overtopping, erosion by waves, mechanisms interconnected with seepage like piping, hydraulic failure and liquefaction. It is obvious these mechanisms also depend of the nature of high water events. Overtopping intensity depend of the duration of some high enough water level, as also seepage cannot occur without long enough duration needed for water to get through levee. Referring to the chapter 2, these also confirms the importance of the information about duration of the high water event, not just the peak flow. As the second, flood can occur due to the insufficiently robust structure which can be caused by various damages on the levee, including the occurrence of sliding surface, subsidence, damages caused by previous events etc. Geotechnical uncertainty cannot be strictly separated from hydraulic uncertainty because, that is, uncertainty arises from occurrence of hydraulic phenomenons, but also depends on material properties. Thus, failure mechanisms include:

- Sliding surface on the upstream slope or on the the downstream slope
- Overtopping
- Piping through levee or ground
- Hydraulic failure of the outside slope or ground
- Ground and levee material liquefaction
- Erosion by waves
- Ground and/or levee subsidence
- Earthquake
- Different type of damages on levee caused by animals and/or humans, tree damages etc...

As it is obvious, except of material properties, occurrence of these mechanisms depends of the high water nature – intensity of water level increase, decrease and duration. These mechanisms are principally acting in the combination and levee failure due to breaching is complex for full physical and mathematical treatment. Mathematical models usually, as the output result, give the time failure and output hydrograph. [12] This also implies that any estimation of levee failure probabilities is, in the least, challenging.

Occurrence of sliding surface could potentiate the levee failure and thus the occurrence of flood event, if not at the moment of some present, possibly in some of the future high water events. This stands because the most critical situation for slides to occur (also surface slough) is when water level decreases suddenly. Then slopes, saturated with water and additionally loaded with flow from saturated area, have the greatest magnitude of load and the lowest resistance. For calculation of the sliding occurrence probability *P(s)*, both on upstream and downstream slope, the concept of reliability from construction engineering can be used as it is in [13]:

$$P(s) = 1 - \Phi(\beta) \tag{4}$$

where $\Phi(\beta)$ is is the cumulative probability function for the safety index β calculated using the equation:

$$\beta = \frac{ln(\frac{C_{50}}{D_{50}})}{\sqrt{\sigma^2(lnc) + \sigma^2(lnD)}} \tag{5}$$

C_{50} and D_{50} are the median values of the capacity and demand. Capacity is given as the density distribution function of material resistance and demand as the density distribution function of loads acting on levee slope. Label σ stands for the standard deviation of capacity and demand given in lognormal distribution. Full treatment of these mechanisms includes considering the nature of water level decrease, which also emphisize the importance of considering the volume and duration of high water events. So the approximate probability of failure and flooding involves using (4), (5) in combination with (6), having in mind that, due to water level decrease, it is not necessary this triggers the flooding.

Overtopping occurs if water level exceeds the level of levee crown. As the circumstances where, at the same time, there are no any geometrical imperfections of levee and the material is appropriately compacted, it is quite possible that overtopping will potentiate the levee failure. Principally, the levee crown elevation is estimated as the sum of the level of the high water event with certain probability (that is return period) and some value called freeboard. Ignoring the freeboard, the overtopping probability *P(O)* (sometimes also called risk) of the high water with certain return period *T* during the structure life time *LT* is [10]:

$$P(O) = 1 - \left(1 - \frac{1}{T}\right)^{LT} \tag{6}$$

In the cases when during the high water, due to the seepage through levee or ground, water flow removes material particles creating thin channels, the process is called piping. Further increasing of particles removal will enlarge those channels initiating the breaching of the levee. As mathematical models of internal erosions are still developing, probability quantification of levee failure due to piping is done by subjective judgement, so called subjective probabilities. The probability of breaching by (through) piping $P(B_{tp})$ can be estimated using the University of New South Wales method or modified form like in [14]:

$$P(B_{tp}) = \prod_i w_i P_{ref} \tag{7}$$

where w_i are the weights, i characteristics affecting the performance, P_{ref} probability of breaching by piping of a reference levee with fixed characteristics. Characteristics include anything which could contribute to piping like animal burrowing, seepage, subsidence, compactness, existing culverts etc. depending also of the material used for building levees [14]. The same way probabilities of different damages can be estimated, which is already included in the mentioned method.

Levees are usually designed and built in the way the compactness is as high as liquefaction should not occur during the high water event. Ground made of loose material, like sands, could potentiate fluidization, and flow of water coupled with material due to the pressure of water on upstream side could cause the levee failure and very likely the flood event. Although the material fluidization can occur without earthquake, at least as cold flow (creep), still, liquefaction is usually analyzed coupled with trigger like earthquake. As the concrete formulation of levee failure probabilities are not found, soil liquefaction probability could be estimated due to [15,16].

Earthquakes are usually given with return period and its magnitude, and levees are sized in the manner they have got the certain safety factor, that is safety index as it is in (5). The earthquake probability of the certain return period can be calculated using (6), combining with (4), (5) and calculating the probability of the high water level of certain period resulting in the probability of concomitant earthquake with levee failure and flooding.

All of the accounted uncertainties are sometimes included in considering the probability of levee failure, without analyzing every each of those. The probability of failure is then estimated depending of the levee condition, which is roughly poor, medium or good and usually described using the levee fragility curves. The probability of failure is then given as the function of water level as it can be found in [17].

Epistemic uncertainty arises from the possibility to understand and describe all of these complex mechanisms, which usually come in combinations. Natural uncertainty arises from the range of differences in material properties used for building levees and from the nature of high water events. Realization of mentioned mechanisms does not necessarily mean the occurrence of flood event, but as they can always potentiate it, it is worth to notify them, estimate their magnitude and include them in the risk scenarios. Since the mechanisms could come individually or in combination, the effort could be made in estimating the tree event [17], resulting in the probability of all possible events.

4. Hydraulic and hydrological uncertainty

As hydraulic uncertainties cannot be separated by geotechnical, and hydrological uncertainties are also connected with statistical treatment of hydrological events, there also exist some aspects which can be precieved as purely hydraulic and hydrological. Water level and flow depend on the nature of incoming high water waves, thus consumption curve (water level dependency of flow at certain hydrological station) tend to have varying nature, creating a loop around mean value. As velocity and flow of incoming increasing water wave increase results in greater flow for the same level, outcoming decreasing wave causes less flows for the same level [8]. Thus, this part arises from the quality of hydrological measurements and, equipment and their usage, appropriate interpretation of the data, appropriate regression analysis, which are hydrological, moreover epistemic uncertainty. As natural part, hydrological processes are interconnected with climate and atmospheric processes. Climate variations and changes could make the impact on the occurrence and intensity of future floods, which is already mentioned in the second chapter. Another epistemic uncertainty is thus modeling uncertainty and the question of equality of the water level return period and flow return period [18]. That is, after the assumed levee's route and profile, it is necessary to simulate situation using the hydrograph of high water event and consumption curve of some upstream profile. The resulting water levels depends of bed roughness and geometry, as two main drivers of return period enaquality. Thus it is necessary to make calibration and validation, usually with lower return period flows, respectively with measurements on disposition, so these difficulties are hardly avoidable. Epistemic uncertainty arises from the possibility to fully understand and describe the flow in river beds, reservoirs, seepage etc. This also relates with the description of the flow coupled with geotechnical failures. Natural uncertainty arises from the range of differencies in material properties, that is, natural material used for building levees and material of the ground. Variations in materials also mean variations in hydraulic properties. Hydrological processes are interconnected with climate and atmospheric processes. Climate variations and changes could make the impact on the occurrence and intensity of future floods, which is already mentioned in the second chapter.

5. Conclusion

Statistical, geotechnical, hydrological and hydraulic uncertainties were indicated and described in the paper, stating the components and possibilities of estimating them. Further investigation should define the statistical treatment of hydrological events in the methodological way, including the possibilities of using multivariate statistics and joint probabilities, Bayesian approach, event trees in order to recognize events with important information of volume and duration. Those information could also be valid for estimation of risk from geotechnical failure, which cannot be separated from hydraulic uncertainty. Hydrological and hydraulic uncertainty should be treated in the way of defining the reliability of the equality of flow and water level return period.

References

[1] Singh VP, Jain SK, Tyagi A (2007): Risk and Reliability Analysis, American Society of Civil Engineers, USA
[2] Bogardi JJ, Kundzewicz ZW (2004): Risk, Reliability, Uncertainty, and Robustness of Water Resources Systems, UNESCO, 2004
[3] Hrelja, H (2007): Engineering hydrology, The Faculty of Civil Engineering, University of Sarajevo
[4] Simonović SP (2012): Floods in a changing climate, International Hydrology Series, Cambridge
[5] Merz B, Thieken AH (2005): Separating natural and epistemic uncertainty in flood frequency analysis, Journal of Hydrology 309: 114-132
[6] Ranzi R, Bacchi B, Barontini S, Ferri M, Mazzoleni, M (2012): Levee breaches statistics and „geotechnical" uncertainty in flood risk mapping, EGU Leonardo Conference „Hydrology and society", Torino
[7] Vrijling JK, van Gelder PHAJM (2000): Policy implications of Uncertainty Integration in Design, 8th Int. Symp. On Stochastic Hydraulics, 633-646

[8] Žugaj R (2000): Hydrology, The Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb
[9] Pauše Ž (2003): Probability – information, stochastical processes, Školska knjiga, Zagreb
[10] Kuspilić N, Gilja G, Ocvirk E (2015): Calculation of flood event exceedence and ist impact on flood risk, 6th Croatian Water Conference with International Participation, Opatija, 583-592
[11] IPCC (Interggovernmental panel on climate change) (2015): Climate change 2014 – Synthesis Report, WMO, Geneva, Switzerland
[12] Zhu YH, Visser P, Vrijling J (2006): A model for breach erosion in clay-dykes
[13] Hamedifar H, Bea RG, Pestana-Nascimento JM, Roe EM (2014): Role of Probabilistic Methods in Sustainable Geotecnhical Slope Stability Analysis, Precedia Earth and Planetary Science 9, 132-142.
[14] Redaelli M (2011): Estimating the probability of piping-induced breaching of flood embankments, International Symposium on Geotechnical Safety and Risk, Munchen, Germany
[15] EERC (Earthquake Engineering Research Center) (2003): Recent Advances in Soil Liquefaction Engineering: A Unified and Consistent Framework, University of California, Berkley
[16] Moss RES, Seed RB, Kayen RE, Stewart JP, Der Kiureghian A, Cetin KO (2006): CPT-Based Probabilistic and Deterministic Assesment of In Situ Seismic Soil Liquefaction Potential, Journal of Geotechnical and Geoenvironmental Engineering 132 (8), 1032-1052
[17] Hui R (2014): Optimal levee design and flood systems, dissertation University of California, USA
[18] Guganesharajah K, Lyons DJ, Parsons SB, Lloyd BJ (2006): Influence of Uncertainties in the Estimation Procedure of Floodwater Level, Journal of Hydraulic Engineering 132 (10), 1052-1060