

## EVALUATION OF LOAD VALUES USING THE GUMBEL MODEL

S. WOLIŃSKI<sup>1</sup>, T. PYTLOWANY<sup>2</sup>

The paper deals with application of the Gumbel model to evaluation of the environmental loads. According to recommendations of Eurocodes, the conventional method of determining return period and characteristic values of loads utilizes the theory of extremes and implicitly assumes that the cumulative distribution function of the annual or other basic period extremes is the Gumbel distribution. However, the extreme value theory shows that the distribution of extremes asymptotically approaches the Gumbel distribution when the number of independent observations in each observation period from which the maximum is abstracted increases to infinity. Results of calculations based on simulation show that in practice the rate of convergence is very slow and significantly depends on the type of parent results distribution, values of coefficient of variation, and number of observation periods. In this connection, a straightforward purely empirical method based on fitting a curve to the observed extremes is suggested.

*Key words:* snow load, extreme loads, Gumbel model, Monte Carlo simulation, return period.

## 1. INTRODUCTION

The construction of a building must be designed in such a way as to meet the requirements of reliability and, while minimizing the costs, as to be able to take over, in a designed period of use, all the effects and loads that may occur during its realization and use [1,2]. The aforementioned requirement for reliability and costs reduction of realization of the project is closely related to several factors, among others: assessment of the size and likelihood of loads, as well as evaluation of the anticipated physical, social and environmental costs of damaging or excluding the building from use. As for the environmental loads, they may result from the phenomena or events of a natural character. Climatic loads being a result of natural phenomena deserve attention in particular. These phenomena include geophysical phenomena and processes that take place in the earth's crust, in atmosphere, and in hydrosphere. The most typical values of these loads are generally the main representative values caused by these events. In the past they were determined on the basis of observation, simple measurements and historical

---

<sup>1</sup> Assoc. Prof. S. Woliński, Department of Building Structures, Faculty of Civil and Environmental Engineering, Rzeszów University of Technology, Rzeszów, Poland, e-mail: szwolkkb@prz.edu.pl

<sup>2</sup> MA. Department of Civil Engineering, Faculty of Civil Engineering, Institute of Technology, Krosno State College, Krosno Poland, e-mail: tompyt@pwsz.krosno.pl

data. Modern form of their description are usually probabilistic models that take the form of random variables whose parameters and probability distributions are identified on the basis of statistical analysis of results of systematic measurements. These models are descriptive, i.e. they do not relate to the nature of the events that determine their occurrence and parameters. Currently, the most commonly used models are those based on random discrete trials, multiplicative interaction models, and extreme values. Key parameters determining representative value of environmental loads are return period  $T$  (in which the main representative load maybe exceeded on average once) and annual (or this related to other unit period) probability  $p$  of exceeding this value.

The main scientific objects of the article are: a presentation of a review of models of probabilistic live interactions, primarily climatic ones, an analysis of assumptions using the Gumbel model for the greatest value, as well as the results of calculations based on simulation related to the ground snow load and the assessment of the possibilities and consequences of use.

## 2. PROBABILISTIC MODELS OF CLIMATIC LOADS

In semi-probabilistic and probabilistic methods of construction design, basic variables are random values. With regard to the live loads, it is accepted that they are continuous random variables with normal distribution  $N$ , log-normal distribution  $LN$  or the Gumbel distribution  $G$ . Their basic representative values  $F_{rep}$  are the main representative values of  $F_k$ , commonly called the quantiles of order  $p$  of corresponding random variables  $F$ , at a fixed return period  $T$  [1,3]. In this paper, the return period  $T$  equals the number of years that elapse on average between successive exceeds of values  $S_t$  by random process  $S(t)$ . In the case of the main representative value of snow load  $S_k$ , different intensities of snowfall during successive winter periods must be taken into account. If  $S_k$  is determined on the basis of specified return period  $T$ , the maximum annual load may exceed the value of  $S_k$  on average on every  $T$  of measurement period (years).

It is commonly accepted [4, 5] that the probability of exceeding the value corresponding to  $T$ -year return period is  $p = 1/T$ . At the same time, it is worth noting that the probability of exceeding of defined in such a way main representative value is  $p$  and the probability  $P$  of exceeding the main representative values  $x$ -times in reference period  $T$  can be calculated by Bernoulli's equation.

$$(1) \quad P(T, x, p) = \binom{T}{x} p^x (1-p)^{T-x}$$

The following table shows the independent Bernoulli trial referred to the assumed parameters  $p, T$ .

An important class of probabilistic models are those that describe the properties of the considered phenomenon generated by a group of causes difficult to isolate

**Table 1**

Sequence of independent Bernoulli trials.

$T, p$	Probability $P$ that variable $F$ in the period $T$ does not exceed the value of $F_k$	Probability $P$ of exceeding the value $F_k$ in period $T$ one time
T=20 years, p=0,05	0,358	0,642
T=50 years, p=0,02	0,364	0,636
T=100 years, p=0,01	0,366	0,634
where: $F_k$ - the main representative value of the load		

which determine the resultant of uncertainty of the phenomenon. The model, being the resultant variable distribution, can often be determined without knowing the distribution of causes, only from the way they affect the resultant variable. Three basic cases are mostly considered, when the causes are: additive, multiplicative, and when extreme values of causes are important [4,5].

a) Models of summation, based on the central limit theorem, which, in a popular simplification, states that with the increasing number of random variables being the components of the sum, the distribution of their sum approximates to normal distribution. In case of meteorological loads, the sum models are not justified.

b) Multiplicative models are based on the assertion that with the increasing number of random variables being the factors of the product, the distribution of their product approaches to the log-normal distribution. Multiplicative mechanism is applicable to many phenomena and processes, including aggregate crushing, accumulation of fatigue damages, flows in rivers, precipitation and meteorological loads, including ground snow load [6,7].

c) Models of extremes are used when the survival or destruction of a building depends on its ability to meet the minimum or maximum requirements. Application of extreme value distributions is justified in case of live loads, including meteorological and seismic loads, intensity of flows of liquids and gases, other loads and the strength of building materials. The main finding of this theory is that, regardless the type of parent distribution of random variables, uncorrelated extreme values have distributions that are asymptotically convergent to extreme value distributions of three types: type I infinite and exponentially right-or left-falling, type II restricted from below by zero and exponentially falling right, type III unilaterally limited by zero and exponentially falling [4].

### 3. THE GUMBEL MODEL FOR THE MAX VALUE

Since the sixties of the twentieth century, the distributions of maximum values have been commonly used to estimate the parameters of distributions of meteorological loads on building structures. The distribution of the first type, which is the Gumbel double exponential distribution, is most commonly used. Cumulative distribution function of the distribution of variable  $Y$ , the greatest one of many independent random variables

with the exponential character for sufficiently great values of the argument has the following form:

$$(2) \quad F_Y(y) = \exp\{-\exp[-\alpha(y-u)]\}, -\infty \leq y \leq +\infty$$

where:  $u, \alpha$  – distribution parameters: modal value and the measurement of dispersion.

Between the parameters of Gumbel distribution and moments of this distribution, the following interactions occur:

$$(3) \quad m_Y \approx u + \frac{0,577}{\alpha}, \sigma_Y \approx \frac{1,282}{\alpha}, \gamma_Y = 1,1396$$

where:  $m_Y, \sigma_Y, \gamma_Y$  – expected value, standard deviation and skewness.

Projected main representative value of the load as a function of the return period  $T$  can be estimated using the dependence:

$$(4) \quad F_k \equiv Y_k = -\ln\{-\ln[F_Y(y)]\} = \alpha(F_k - u)$$

Values of random variable cumulative distribution function  $F_Y(y)$  for the total number  $n$  of seasonal maxima (e.g. in case of annual snow load) is calculated based on order statistics of the range, arranged in ascending order:

$$(5) \quad F_Y(y_i) = \frac{i}{n+1}$$

For the parameters  $\alpha$  and  $u$  calculated according to (5) for the assumed reference period  $T$ , the main representative value of an action  $F_k$  can be determined:

$$(6) \quad F[F_k = Y_k(T)] = 1 - \frac{1}{T}$$

In the extreme value theory it is not valid to assume that the observed maximum values should fit well into the Gumbel distribution, because it is shown that the distribution of maxima tends asymptotically to the theoretical Gumbel distribution with the increase in the number of sample  $n$ , i.e., the number of observation periods, from which the greatest values have been selected. As it can be noticed, the distribution parameters  $\alpha$  and  $u$  depend on  $n$ . In practice, it is assumed that  $n$  is great enough to use empirical cumulative distribution function in the proper model of random variable  $Y$ . For instance, according to the recommendations of PN-EN 1991-1-3 [8], the ground snow load analysis should include at least 20 measurements of maximum annual values. Functional interaction between the modal value  $u_n$  from the  $n$ -element sample can be approximated, for example, using such dependency [5]:

$$(7) \quad u_n = \frac{1}{\lambda} \ln n, \text{ lub } u_n = \frac{1}{\lambda} \sqrt{\ln n}$$

where:  $\lambda$  parameter determined on the basis of the results of the measurements.

The second fundamental limitation on the applicability of extreme value distributions, usually neglected in engineering analysis, results from the assumption that the extreme value theory is important when the periodic (e.g. annual) extreme values have been selected from the infinite number of independent observations [4]. This assumption is not satisfied even approximately when considering the problem of prediction of some meteorological impacts, such as snow load. Influence of the number of  $m$  sets of observations and their distributions on the main representative value of an action can be analyzed by simulation methods.

#### 4. ASSESSMENT OF THE IMPACT OF NUMBER AND TYPE OF PARENT DISTRIBUTION TO ESTIMATE THE EXTREME VALUES

##### 4.1. PROCEDURE FOR SIMULATION CALCULATIONS

To assess the impact of the number of independent unit periods  $n$ , from which periodic maxima  $y_{i,max}$ , of numbers of measurement results in unit periods  $m$ , and type of parent distribution for the anticipated maximum value  $y_{max}$  of the random variable  $Y$  in the reference period  $T$ , have been selected. The author developed an algorithm and an original program for the simulation based calculations. The procedure for simulation based calculations of extreme values of random variables can be presented as follows:

- a) Generate  $m$  pseudo-random numbers of the assumed probability distribution (e.g.,  $N$  – normal,  $LN$  – log-normal and  $G$  – Gumbel) for specific moments of the distribution of random variable  $Y$ ;
- b) Choose the greatest and the least of  $m$  pseudo-random  $y_{max} = \max(y_1, y_2, \dots, y_m)$  and  $y_{min} = \min(y_1, y_2, \dots, y_m)$ ;
- c) Repeat the drawing  $n$  times as well as the choice of the the greatest and the least value according to points (a) and (b) and write down two sequences of numbers  $y_{j,max}$  and  $y_{j,min}$ ,  $j = 1, 2, \dots, n$ ;
- d) Develop a frequency histogram of empirical distribution and calculate the distribution moments of maximum and minimum values;
- e) Calculate the parameters of the assumed extreme probability distribution with the method of least squares, of moments or of maximum likelihood;
- f) For the selected reference period  $T$  or for the probability of exceeding the main representative value (quantile) of the random variable  $Y$  during the unit (e.g. annual), calculate the main representative value  $Y_k$ .

A pseudo-random number generator from the Matlab package version 6.5 was used for the calculations [9]. The original program was written to draw pseudo-random numbers at a certain distribution, in this case:  $N$ ,  $LN$ , and  $G$ .

## 4.2. RESULTS AND ANALYSIS OF SIMULATION BASED CALCULATIONS

Algorithm of simulation based calculations presented in point 4.1 was used to estimate the impact of the type of distribution, a number of results  $m$  in a period from which there were selected: the greatest value  $y_{j,max}$  and numbers  $j = 1, 2, \dots, n$  of observation periods on the greatest value  $y_{max}$  of variable  $Y$ . The distributions of the random variable  $Y$  of the type  $N$ ,  $LN$ , and  $G$  with the same parameters: the expected value of  $1.0$  and a coefficient of variation:  $0.1$ ,  $0.2$ ,  $0.5$  and  $0.7$  were taken for the calculation. The values of these parameters were calculated by the method of moments on the corresponding values of distribution parameters  $LN$  and  $G$  [10]. For a number of results within a period and the number of unit periods for which the calculations were made, it was adopted:  $m = 10, 20, 50, 100, 1000, 10000$  and  $n = 10, 100, 1000, 10000$ . On the Fig. 1, 2 and 3 there are the results of the simulation based calculations as a dependency graph  $y_{max}(m)$  for selected types of distribution of random variable and value  $n$ .

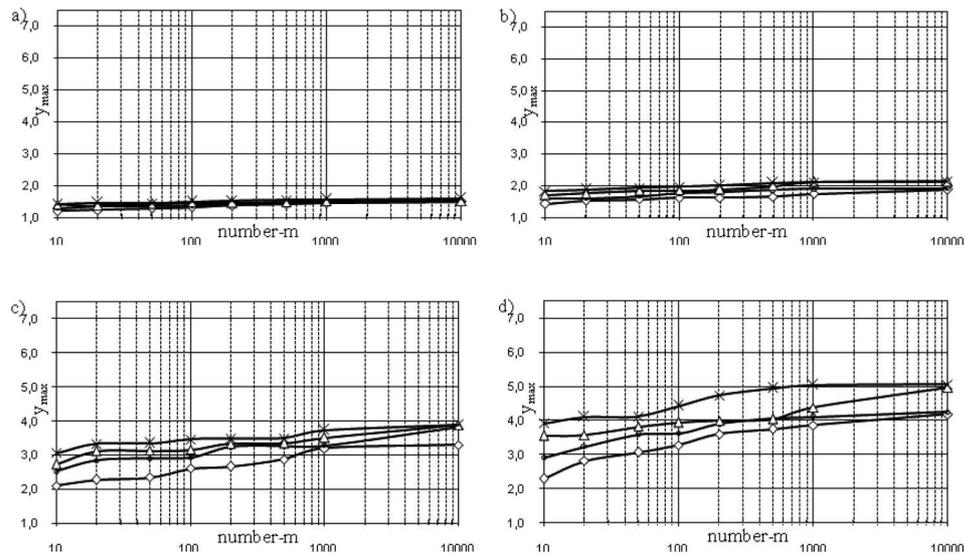


Fig. 1. Dependency graphs  $y_{max}(m)$  for  $n = 10[-\diamond-]$ ,  $10^2 [-\square-]$ ,  $10^3 [-\triangle-]$  and  $10^4[*]$ , distribution  $N$  on expected value  $\hat{y} = 1,0$  and coefficient of variation: (a)  $v=0,10$ ; (b)  $v=0,20$ ; (c)  $v=0,50$ ; (d)  $v=0,70$

Based on the analysis of the results of simulation based calculations, it was found that all the factors included in the calculations had a significant impact on the value  $y_{max}$ . The following are the major conclusions and observations drawn from the results of calculations:

- Regardless of the type of parent distribution, numbers  $m$  of the results in the unit observation periods and maximum values  $n$  (i.e. number of unit periods included), a decisive influence on the maximum values  $y_{max}$ , and on the tendency of their

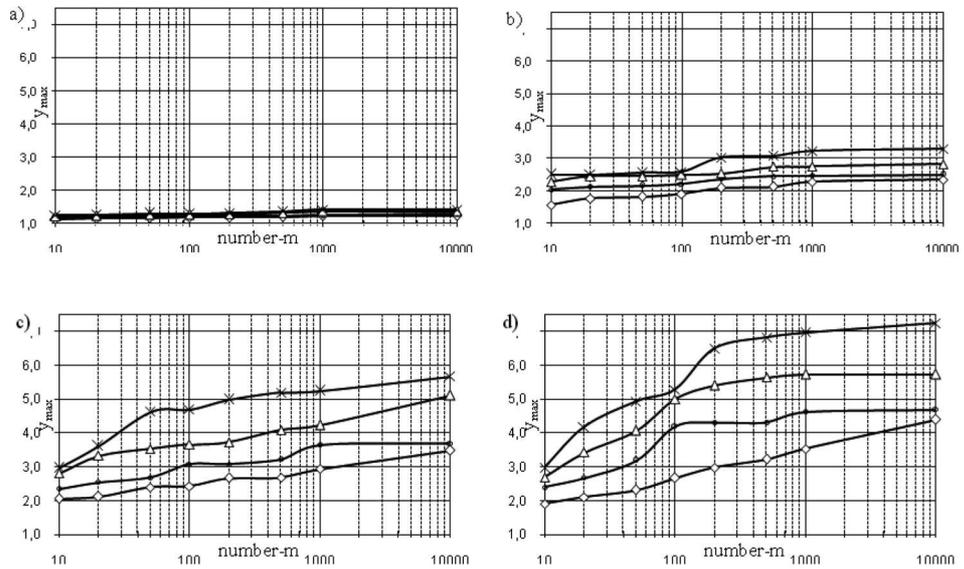


Fig. 2. Dependency graphs  $y_{max}(m)$  for  $n = 10[-\diamond-]$ ,  $10^2 [-\square-]$ ,  $10^3 [-\triangle-]$  and  $10^4[*]$ , of distribution LN value of median and logarithmic coefficient of variation: (a)  $\check{y}=0,995$ ;  $\nu=0,0997$ , (b)  $\check{y}=0,981$ ;  $\nu=0,198$ ; (c)  $\check{y}=0,894$ ;  $\nu=0,472$ ; (d)  $\check{y}=0,819$ ;  $\nu=0,631$

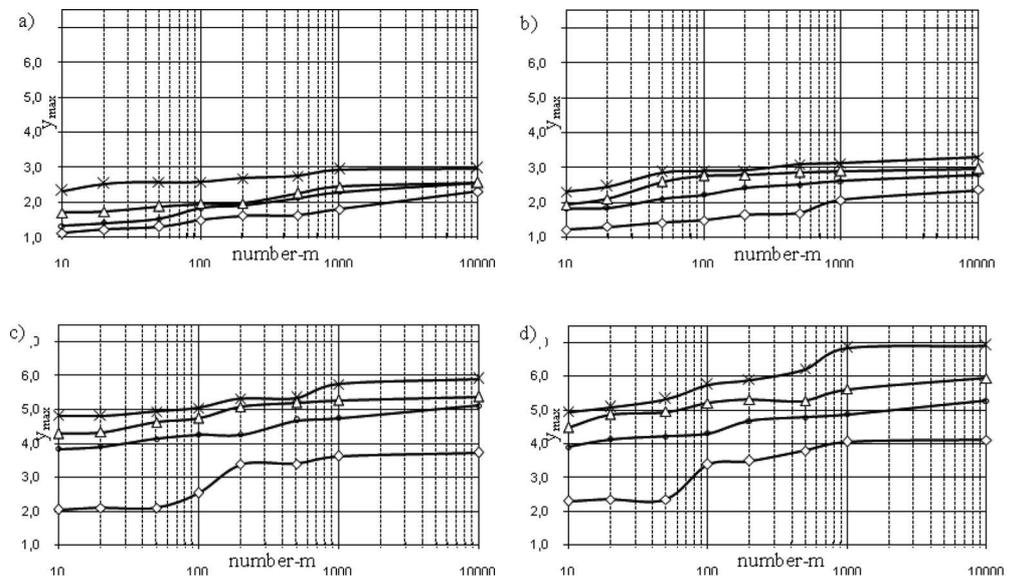


Fig. 3. Dependency graphs  $y_{max}(m)$  for  $n = 10[-\diamond-]$ ,  $10^2 [-\square-]$ ,  $10^3 [-\triangle-]$  and  $10^4[*]$ , of distribution G of modal value and scatter coefficient: (a)  $u=0,955$ ;  $\alpha=0,078$ , (b)  $u=0,91$ ;  $\alpha=0,156$ ; (c)  $u=0,775$ ;  $\alpha=0,39$ ; (d)  $u=0,332$ ;  $\alpha=0,546$

- changes, is a dispersion of results in unit periods, and the coefficient of variation  $\nu$  is its measure.
- Values  $y_{max}$  for determined  $m, n$  and  $\nu$  depend on the type of parent distribution. For normal distribution  $N$  of coefficient of variation  $\nu \leq 0,20$  there is a weak dependence, and for  $\nu \geq 0,50$  the dependence is increasingly strong. A similar, but clearer and stronger dependence occurs in case of the log-normal  $LN$  distribution. Definitely the strongest dependence occurs in case of Gumbel distribution  $G$ .
  - Regardless of the type of parent distribution, the influence on numbers of measurements within unit period  $m$  on the values of  $y_{max}$  is very clear, in particular for the values of  $\nu > 0,20$  and small values  $m$ . The relative stabilization of the tendency of values  $y_{max}$  to grow occurs for  $m > 50 \div 200$ .
  - There was a significant effect on numbers of unit periods  $n$  included in the calculation of the value  $y_{max}$ . Depending on the value of dispersion measurement and on the type of distribution, the growth rate of values  $y_{max}$  decreases significantly for  $n > 100$ , but a clear stabilization of  $Y_{ma}$  is observed for values of  $n = 10^3 \div 10^4$  and  $m > 50 \div 100$ . For the distributions of a type  $LN$  and  $G$  with coefficients of variation  $\nu \geq 0,50$  there are obvious disorders of changes of  $y_{max}$  with the increase of value  $n$ .

## 5. EXAMPLE

An illustration of the impact of the number  $m$  of independent measurement results during the unit period, and the type of parent distribution on the evaluation of the main representative values, is shown on the example of a snow load. Based on the results of measurements of ground snow load from the years 1950-2000 [11] for two points: in Krakow Balice (zone II) and Wroclaw (zone I), the expected values and load standard deviations were estimated with the method of moments. For the assumed parent distributions  $N$ ,  $LN$  and  $G$ , according to the procedure presented in point 4, the sets of quasirandom numbers  $n = T = 50$  of assumed distributions, and number  $m = 10, 10^2, \dots, 5 \times 10^6$  were generated. Then, the maximum values were selected and on the basis of their distribution the main representative values  $S_k$  for the period  $T = 50$  years were calculated.

Figure 4 presents dependency graphs of the main representative values  $S_k$  of ground snow load, calculated as quantiles of order  $p = 1/T = 0,02$  of the number  $m$  of measurement results, for selected measurement points and the assumed types of parent distributions

## 6. CONCLUSIONS

The standard method of calculating the main representative values of the impacts based on the Gumbel model can lead to the results containing significant errors difficult to estimate. It results from the acceptance of a strong and hard to verify assumption

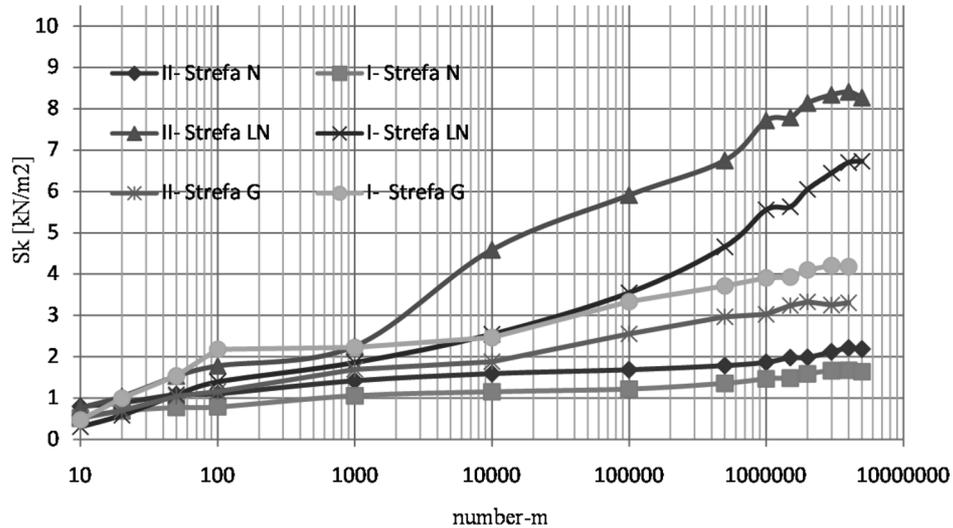


Fig. 4. Dependency graphs of  $S_k$  on the number of measurements and type of parent distribution.

concerning the asymptotic nature of similarity of maximum values distribution to the Gumbel distribution, and sufficiently large, in theory converging to infinity, number of sets of results from which the greatest values have been selected.

The results of simulation based calculations presented in this article show a very strong dependence of assessments obtained with the assumption of Gumbel model on several factors: dispersion of measurement results in time intervals from which maximum values were selected; the number of measurement results in these periods; the type of parent distribution of the results; the number of set of maximum values. Therefore, purely empirical approach based on analysis of second-order regression seems to be a better solution.

## REFERENCES

1. PN-EN 1990 *Eurocode: Basis of structural design*, PKN, Warsaw 2004.
2. ISO 2394 – 1998 (E). *General principles on reliability for structures*.
3. PN-EN 1991 *Eurocode 1: Actions on structures*, PKN, Warsaw 2005.
4. E.J. GUMBEL, *Statistics of extremes*. Columbia Univ. Press, New York, 1958.
5. M. FISZ, *Probability* and PWN, Warsaw, edition II, 1958.
6. J.R. BENJAMIN, C.A. CORNELL, *Probability, Statistics and Decision for Civil Engineers*, WNT, Warsaw 1977.
7. B. ELLINGWOOD, R. REDFIELD, *Ground snow loads for structural design*. Journal of Structural Engineering, ASCE, **109**, 4, 1983, 950-964.

8. PN-EN 1991-1-3 *Eurocode 1: Actions on structures. Part 1-3: General actions– Snow loads*, PKN, Warsaw 2005.
9. B. MROZEK, Z. MROZEK, *Matlab and Simulink*. Helion, Gliwice 2004.
10. J. MURZEWSKI, *Reliability of structures*, Arkady, Warsaw, 1989.
11. IMiGW *Statistic yearbooks of precipitation 1950-2000*.

*Remarks on the paper should be  
sent to the Editorial Office  
not later than September 30, 2012*

*Received January 10, 2012  
revised version  
June 10, 2012*