Application of the distribution of the mth extreme value in land use planning

H. A. van Alderwegen

Institute for Land and Water Management Research (ICW), Wageningen, Netherlands

Summary

In this paper the statistical theory of mth extreme values is elaborated by calculation of the theoretical values of the expected mean and standard deviation of the reduced variate $\bar{y}_{n,m}$ of the mth extremes as functions of the number of observations n, i.e. $n = 1(1)20$, 50 and 100 respectively. This calculation allows to estimate the probability function of the mth extreme value of a stochastic variate based on a limited number of observations.

For six different types of recreation and traffic facilities, calculation of the probability functions of the use intensity on the first to tenth most crowded day in a year is discussed as based on observations during some years of the daily use intensity of these facilities. With the aid of these functions a justified choice of the normative day for the design capacity of similar projects can be made.

Introduction

In the fields of hydraulic engineering and hydrology, application of the theory of extreme values is common practice. For instance, as a rule assessment of the normative height of dikes is based on a frequency analysis of the heights of floods, and determination of the capacity of sewage systems in urban areas is based on a frequency analysis of extreme precipitation values. In such cases an exceedance of the capacity of the project will result in calamities (loss of life, substantial economic losses). Consequently hydraulic engineers and hydrologists primarily will be interested in the return period of the first extreme value of the observed phenomena (flood, precipitation).

In contrast, in the field of land use planning exceedance of the carrying capacity of a man-made facility as a highway or a recreation site rarely leads to a calamity. Exceeding the capacity of a highway for instance (only) will result in traffic congestion and consequently in loss of time for road-users; exceeding the

capacity of a recreation site means less enjoyment for recreational visitors. In accordance with economic principles the design capacity of such land use projects will be based on predicting the use intensity.

We suggest to equate the design capacity of a facility with the expected number of visitors on the mth (most) crowded day in a year with a probability of P %; m and P should be chosen by policymakers depending on the type of facility and should be based on:
— the extent of exceedance as obtained by frequency analysis of the use intensity;
— the physical and social impact of exceedance of the capacity;
— the marginal costs of construction and maintenance of the facilities designed.

In order to know the probability of m exceedances of a certain use intensity per period, the probability function of the mth extreme value of the number of users per day is to be known. Data about the use intensity of such objects almost exclusively are available over only a few years. Calibration of the probability function of the mth extreme value of the number of users, requires the calculated values of the expected mean deviation of the population of mth extremes ($\bar{y}_{n,m}$) and the expected standard deviation of this same population ($\sigma_{n,m}$), where n is the registration period in years.

The procedure with respect to the calculation of these model parameters is discussed in van Alderwegen (1980). The procedure proposed to assess a criterion in land use planning for the design capacity of a man-made facility is given in the present paper.

Distribution of mth extreme value

The distribution of the mth extreme of n observations taken from an initial distribution of the exponential type now will be considered. We speak of mth values when m is counted from the start of a series given in increasing order x_1, $x_2, \ldots, x_{n-1}, x_n$, and of mth extremes when m is small compared to n. Gumbel (1960) has given the asymptotical distribution of the mth extreme, viz:

$$\Phi_m(x_m) = \frac{m^m \alpha_m}{(m-1)!} \exp(-m y_m - m e^{-y_m}) \tag{1}$$

$$y_m = \alpha_m(x_m - u_m) \tag{2}$$

where:
x_m = the mth extreme of the observations of x
$\Phi_m(x_m)$ = distribution of the mth extreme
y_m = reduced mth extreme variate
α_m = scale parameter; reciprocal of the measure of dispersion of the distribution
u_m = mode of the distribution.

The asymptotic probability $\Phi_m(x_m)$ of the mth extreme value of the observations is given by:

$$\Phi_m(x_m) = \exp(-me^{-y_m}) \sum_{v=o}^{m-1} \frac{m^v e^{-vy_m}}{v!} \;.$$ (3)

Fitting of the observations to the distribution Φ_m requires an independent estimate of the distribution parameters α_m and u_m. According to Gumbel (1954) and Stol (1978):

$$1/\alpha_m = s_m/\sigma_{n,m}$$ (4)

$$u_m = \bar{x}_m - \bar{y}_{n,m}/\alpha_m$$ (5)

where:
$\bar{y}_{n,m}$ = expected mean deviation of the population of mth extremes
$\sigma_{n,m}$ = expected standard deviation of the population of mth extremes
$\bar{x}_m$ = mean of the observations of mth extremes
s_m = standard deviation of the observations of mth extremes

The theoretical values $\bar{y}_{n,m}$ and $\sigma_{n,m}$ are the mean and the standard deviation of n values of the reduced variate taken at equidistant probability intervals, according to the formula for the plotting positions:

$$\Phi(y_i) = p(y \leqslant y_i) = i/(n + 1), \quad i = 1(1)n.$$ (6)

When the number of observations of the mth extreme value is finite the theoretical values of $\bar{y}_{n,m}$ and $\sigma_{n,m}$ have not been published as far as we know. This lack of theoretical expected values obstructs the calculation of the parameters α_m and u_m after Eqs. 4 and 5 for the mth extreme values from a limited number of observations of x_m. Therefore the table which contains the theoretical values $\bar{y}_{n,m}$ and $\sigma_{n,m}$ have been calculated for $m = 2(1)10$ and $n = 1(1)20, 50$ and 100. The solutions of $\bar{y}_{n,m}$ are obtained by an iterative process because for $m \geqslant 2$ the inverse function of Eq. 3 are implicit. The result of the calculation is given in Table 1 in which for the number of observations $n = 1(1)20, 50, 100$ the expected mean $\bar{y}_{n,m}$ and standard deviation $\sigma_{n,m}$ is given for $m = 1(1)10$.

Applications in land use planning

To assess a criterion for the designing capacity of, for instance, angling facilities, the following procedure can be employed.

For a representative facility the number of visitors on the ten most crowded days per year are selected over the period 1972 up to 1979. To diminish the influence of autonomous growth in the participation in sport fishery in this period the number of visitors per day (V_{day}) is divided by the total number of visitors

Table 1. Expected mean $\bar{y}_{n,m}$ and expected standard deviation $\sigma_{n,m}$ of the mth extremes as functions of the number of observations n.

	m = 1		m = 2		m = 3		m = 4		m = 5	
	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$
n = 1	.3665	.0000	.1753	.0000	.1150	.0000	.0855	.0000	.0681	.0000
n = 2	.4043	.4984	.1925	.3276	.1261	.2610	.0937	.2233	.0745	.1983
n = 3	.4286	.6435	.2035	.4210	.1332	.3349	.0989	.2862	.0787	.2540
n = 4	.4458	.7315	.2113	.4768	.1382	.3788	.1026	.3236	.0816	.2871
n = 5	.4588	.7928	.2172	.5153	.1419	.4090	.1053	.3493	.0837	.3098
n = 6	.4690	.8388	.2218	.5440	.1449	.4314	.1075	.3682	.0855	.3265
n = 7	.4774	.8749	.2255	.5664	.1473	.4489	.1093	.3830	.0869	.3395
n = 8	.4843	.9043	.2287	.5845	.1493	.4629	.1107	.3949	.0880	.3500
n = 9	.4902	.9288	.2313	.5995	.1510	.4746	.1120	.4047	.0890	.3586
n = 10	.4952	.9496	.2336	.6121	.1524	.4844	.1131	.4130	.0898	.3659
n = 11	.4996	.9676	.2356	.6230	.1537	.4928	.1140	.4201	.0906	.3722
n = 12	.5035	.9833	.2373	.6325	.1548	.5001	.1148	.4262	.0912	.3776
n = 13	.5070	.9971	.2389	.6408	.1558	.5066	.1155	.4317	.0918	.3824
n = 14	.5100	1.0095	.2402	.6482	.1567	.5123	.1162	.4364	.0923	.3866
n = 15	.5128	1.0206	.2415	.6549	.1575	.5174	.1168	.4407	.0928	.3903
n = 16	.5154	1.0306	.2426	.6609	.1582	.5220	.1173	.4446	.0932	.3937
n = 17	.5177	1.0397	.2437	.6663	.1589	.5261	.1178	.4481	.0936	.3968
n = 18	.5198	1.0481	.2446	.6713	.1595	.5299	.1182	.4513	.0939	.3996
n = 19	.5217	1.0557	.2455	.6758	.1600	.5334	.1186	.4542	.0942	.4021
n = 20	.5236	1.0628	.2463	.6800	.1605	.5366	.1190	.4569	.0945	.4045
n = 50	.5485	1.1607	.2575	.7368	.1677	.5797	.1242	.4928	.0986	.4359
n = 100	.5600	1.2065	.2626	.7625	.1709	.5990	.1266	.5088	.1005	.4498

per year (V_{year}) and then multiplied by 100. Furthermore, the possibility of application in other similar facilities is increased. This relative number of visitors is considered to be a statistical variate with a probability distribution of the exponential type. To calculate the parameters α_m and u_m according to Eqs. 4 and 5 and the probabilities of $\Phi(\underline{x}_m \leqslant x)$ for series of values of x, a computer programme has been written.

The observed and calculated cumulative distributions of the relative number of visitors to the fishing waters considered, is plotted in Fig. 1. The fit of the curves seems to be reasonable.

When as a starting point in the planning it is taken that there has to be sufficient space to allow everyone who wishes to participate in angling to do so on, for instance, the fifth crowded day in a modal ($\Phi_5 = 0.50$) season, we can apply Fig. 1 (see →). From this figure can be derived: on the chosen day this kind of facility will be visited by 2.1 % of the total yearly number of visitors and the capacity of the facility will be designed accordingly. In this case, we accept that the

$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$	$\bar{y}_{n,m}$	$\sigma_{n,m}$
m = 6		m = 7		m = 8		m = 9		m = 10	
.0565	.0000	.0483	.0000	.0422	.0000	.0375	.0000	.0337	.0000
.0619	.1801	.0529	.1662	.0462	.1551	.0410	.1459	.0369	.1382
.0653	.2307	.0558	.2128	.0487	.1985	.0432	.1867	.0389	.1769
.0677	.2607	.0578	.2404	.0505	.2242	.0448	.2109	.0403	.1997
.0695	.2812	.0594	.2593	.0518	.2418	.0460	.2274	.0413	.2154
.0709	.2963	.0606	.2732	.0529	.2548	.0469	.2396	.0422	.2269
.0721	.3081	.0616	.2841	.0538	.2649	.0477	.2491	.0429	.2359
.0730	.3176	.0624	.2928	.0545	.2730	.0483	.2567	.0434	.2431
.0738	.3254	.0631	.2999	.0551	.2797	.0488	.2630	.0439	.2490
.0745	.3320	.0637	.3060	.0556	.2853	.0493	.2683	.0443	.2540
.0751	.3376	.0642	.3112	.0560	.2901	.0497	.2728	.0447	.2583
.0757	.3425	.0646	.3157	.0564	.2943	.0501	.2767	.0450	.2620
.0761	.3468	.0650	.3196	.0568	.2979	.0504	.2801	.0453	.2652
.0766	.3506	.0654	.3231	.0571	.3012	.0506	.2832	.0455	.2681
.0769	.3540	.0657	.3262	.0574	.3041	.0509	.2859	.0457	.2706
.0773	.3571	.0660	.3290	.0576	.3067	.0511	.2884	.0459	.2730
.0776	.3598	.0663	.3316	.0579	.3090	.0513	.2906	.0461	.2750
.0779	.3623	.0665	.3339	.0581	.3112	.0515	.2926	.0463	.2769
.0782	.3646	.0668	.3360	.0583	.3131	.0517	.2944	.0464	.2787
.0784	.3668	.0670	.3379	.0585	.3150	.0519	.2961	.0466	.2803
.0818	.3950	.0699	.3638	.0610	.3390	.0541	.3186	.0486	.3015
.0833	.4075	.0712	.3752	.0621	.3495	.0551	.3285	.0495	.3108

capacity of the fishing waters is insufficient during one day per year with a probability of 94 % and, for instance, three days per year with a probability of 78 % ($\leftarrow\leftarrow$).

The return period of the occurrence of a certain number of exceedances can also be derived from the figure where the return period (T) is given. The return period T is defined by

$$T(x_m) = \frac{1}{1 - \Phi_m(x_m)} .$$ (7)

So once in the nine years the capacity based on $\Phi_5 = 0.50$ will be exceeded seven times a year ($\rightarrow\rightarrow\rightarrow$).

With regard to the maintenance of man-made facilities and the conservation of natural resources, project managers are particularly interested, however, in the extent of the exceedance of the capacity. From Fig. 1 it can be derived that
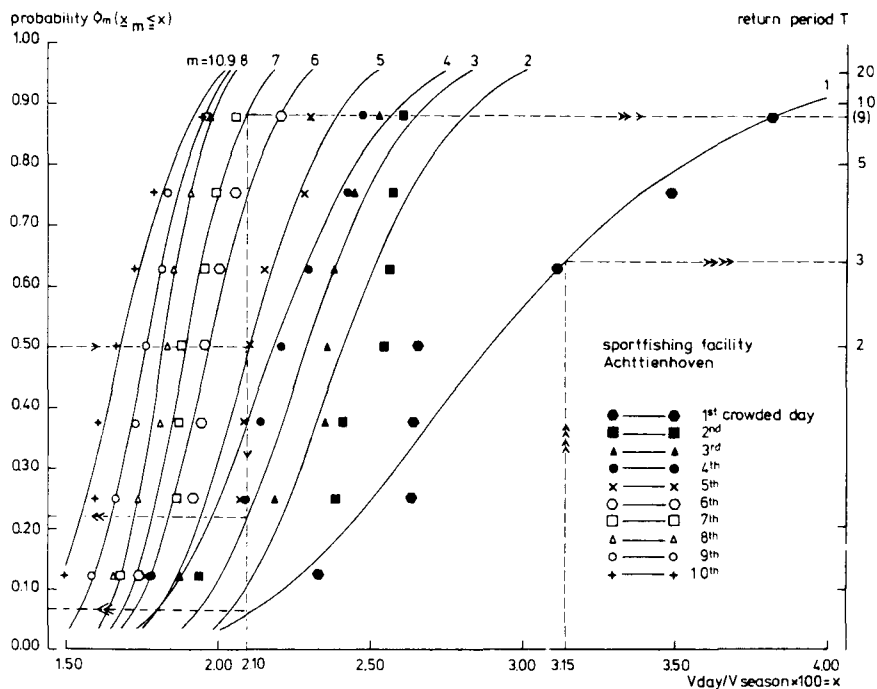
Fig. 1. Estimated and observed probabilities for the first to tenth crowded day in a year related to the number of visitors per year to the facility for anglers near Achttienhoven (data according to the Directorate of Fisheries, The Hague).

the choice of the fifth day with a probability of 50 % leads to the acceptance of an intensity of use of one and a half times the capacity (2.1 × 1.5 % = 3.15 % of the yearly number of visits) every three years (→→→→).

A similar procedure is followed for the number of passages along a point of a main road near Lemmer. The statistical variate now is obtained by dividing the number of passages per day (deg) by the annual daily mean (jeg) as registered from 1973 up to 1979. The results are shown in Fig. 2. For the type of main roads like the one near Lemmer the consequence of two possible chosing capacities are given in Table 2 in terms of the probability of combinations of extent and number of exceedance of the chosen capacity. So, when one accepts that on the 10th most crowded day once in two years the chosen capacity is exceeded (10th day, $\Phi = 0.50$), it will happen with a probability of 9 % that the chosen capacity will be exceeded once in a year by more than 50 %!

The results of the statistical analysis of extreme values of the number of visitors (users) of different kinds of facilities are shown in Fig. 3 as curves that indicate the expected values of the number of visitors on the ten most crowded days
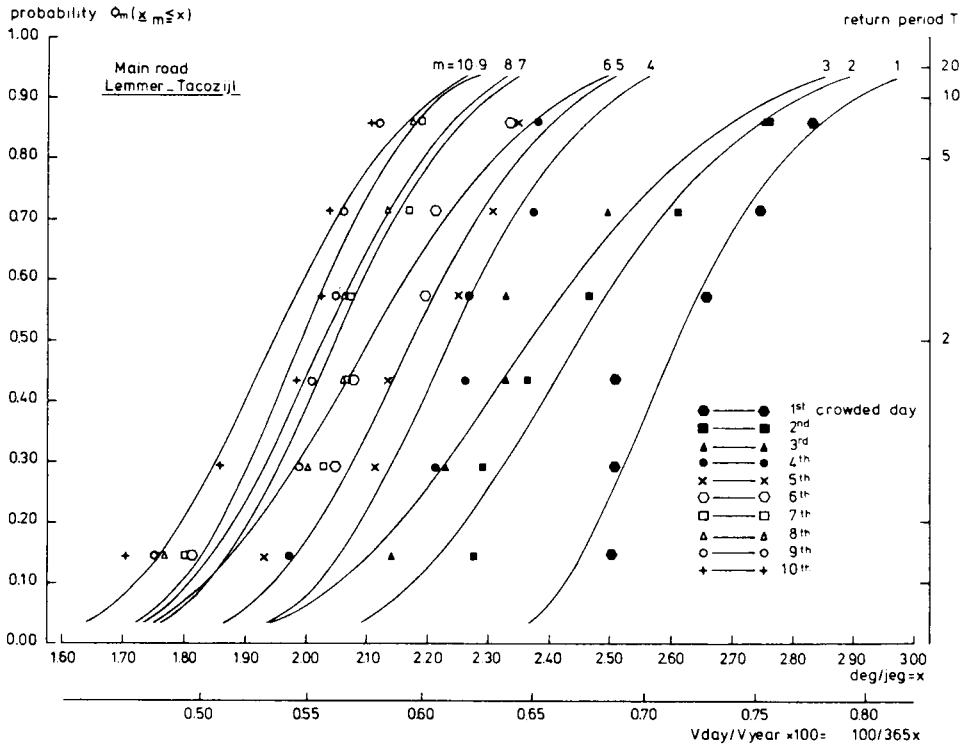
42

Fig. 2. Estimated and observed probabilities for the first to tenth crowded day in a year related to the annual daily mean of passages and the annual number of passages along a main road near Lemmer.

Table 2. Probability (%) of combinations of extent and number of exceedances of capacity on a main road near Lemmer for two chosen capacities.

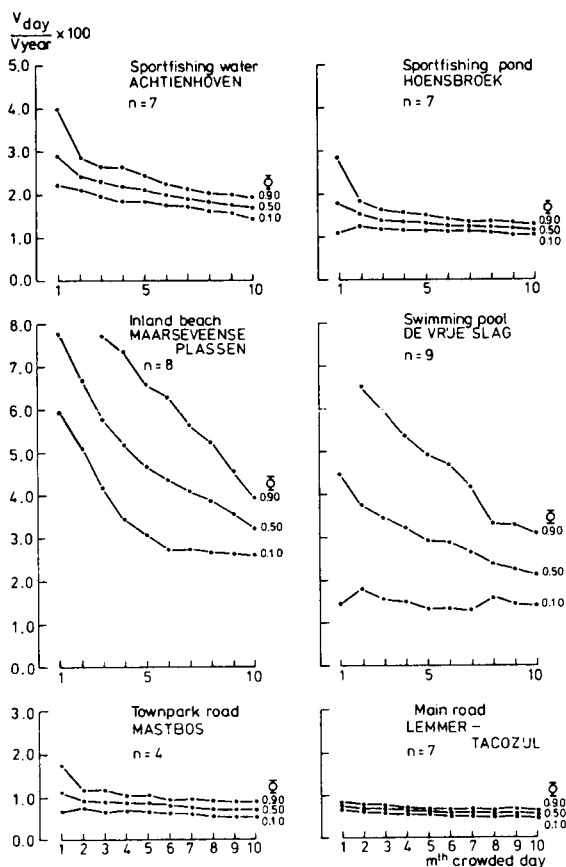| Chosen capacity	Number of exceedances per year	Extent of exceedances (% of capacity)		
		$\geqslant 150$	$\geqslant 125$	$\geqslant 100$
5th day, $\Phi = 0.50$	1	1	30	100
	3	0	13	77
	5	0	1	50
	7	0	0	26
	10	0	0	14
10th day, $\Phi = 0.50$	1	9	88	100
	3	5	40	96
	5	0	11	90
	7	0	3	72
	10	0	1	50

Fig. 3. Curves of the daily number of visitors, expressed as a percentage of the annual number of visitors, expected on the ten most crowded days in a year ($\Phi = 0.10$, 0.50 and 0.90) for six types of facilities; n = number of years during which data were collected.

in a year. The curves concern a sportfishing water in a rural area (Achttienhoven), a sportfishing pond in an urbanized district (Hoensbroek), an inland beach (Maarseveense Plassen), a local swimming pool (De Vrije Slag), a road along a townpark (Mastbos) and a main road in the countryside (Lemmer). In every graph the daily number of visitors is expressed as a percentage of the annual number of visitors.

For inland beaches it appears that 49 % of the annual number of visitors in a modal year ($\Phi = 0.50$) occur on the ten most crowded days, in contrast to the intensity of use of roads where this figure is only 6 to 8 %. Differences in level and course of the six curves clearly demonstrate the necessity of the availability of such curves for such facilities when having to make a choice of what extreme value of the number of visitors (which most crowded day) is to be considered normative for the design capacity. The theory discussed for the mth extreme value therefore is a valuable instrument for the assessment of capacities in land use planning.

44

References

Alderwegen, H. A. van, 1980. Application of the mth extreme value distribution in land use planning. Institute for Land and Water Management Research, Wageningen, Nota 1221. 16 pp.

Gumbel, E. J., 1954. Statistical theory of extreme values and some practical applications. US Department of Commerce, National Bureau of Standards. Applied Mathematics Series No 33.

Gumbel, E. J., 1960. Statistics of extremes. Columbia University Press, New York, USA.

Stol, Ph. Th., 1980. Analysis of extreme values. Proceedings of the Roving Seminar on the Use of Computers in Hydrology and Water Resources Planning, ESCAP, Bangkok, Chapter 7: 58-71. Water Research Series No 52. United Nations, New York, 1980.