Civil Engineering Research Journal
ISSN: 2575-8950

Research Article
Volume 11 Issue 1 - November 2020
DOI: 10.19080/CERJ.2020.11.555802

Civil Eng Res J

Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges

Arie Ben-Zvi*

Hydrologic consultant, Jerusalem, Israel

Submission: October 08, 2020; **Published:** November 03, 2020

***Corresponding author:** Arie Ben-Zvi, Hydrologic consultant, Jerusalem, Israel

Abstract

Temporal distribution, of high runoff discharges, is currently predicted through the exponential distribution. Yet, statistical properties of intervals, in a number of case studies, are found different from those of the exponential distribution. To account for such cases, a more-parameter distribution should be applied. Owing to the inter-dependence between statistical parameters of inter-arrival intervals associated with different threshold discharges, a non-parametric procedure is applied for relating these parameters to the discharge. This procedure is adjusted to yield semi-logarithmic relationships between threshold discharge and the mean and the standard deviation of the intervals between its exceedances. In the present case, these relationships are found representative for high thresholds. The predicted mean intervals, between the highest observed discharges are found closer than the customarily predicted recurrence intervals, to the number of years on record. This validates the suitability of the proposed relationship for determining flood risk and high design discharge.

Keywords: Inter-arrival intervals, High threshold discharges, Semi-logarithmic relationship, Non-parametric procedure, Runoff events, Misrepresentation of the exponential distribution

Introduction

Recurrence interval of peak discharges serves as a measure of flood risk and for determination of design discharges. For a gauged site, it is commonly computed through a three-step procedure: (1) drawing, from the records, a sample of peak discharges, (2) fitting a selected probability distribution to this sample and (3) by use of this distribution, transfer a pre-selected recurrence interval into exceedance probability of a sought discharge (or vice versa: transfer the exceedance probability of a given discharge into the recurrence interval of this discharge). The sampling method and the distribution selected vary within the hydrologic community. But, the third step is common. Its formula is:

$$T(Q) = N / \left(np(Q)\right) \tag{1}$$

where $T(Q)$ is recurrence interval of the discharge Q, N is the number of years on record, n is sample size and $p(Q)$ is exceedance probability of Q. Note: for annual maxima samples, in which the numerical values of series size and of years on record are mostly equal to each other, the equation might be written as:

$$T(Q) = 1 / p(Q) \tag{2}$$

which is dimensionally incorrect and errs for series containing dry years. Step 3, of the procedure, shrinks the recorded information, on temporal distribution of peak discharges, into their mean inter-arrival interval, N/n. All other information, contained in the sample, is abandoned. Consequently, this is the sole statistical parameter applied for computing flood risks and design discharges, regardless of the number of parameters of the distribution selected in step 2. The exponential distribution, commonly selected for computing inter-arrivals of high runoff events, appears suitable for this purpose. It concerns unbounded and independent positive values, its parameter is simply related to the mean inter-arrival interval, the coefficient of variation is equal to 1 and the skewness coefficient is equal to 2. Another advantage of the exponential distribution is its consistency with the Poisson distribution for counting the number of events within time units (usually years). Both of them are one-parameter distributions, and their parameters are related to each other.

Despite the long-term and general acceptance of the above procedure, a number of reservations, regarding its representation of real world, have been published. Lye & Lin [1] analyzed the serial correlation structure of annual flood series from 90 Canadian rivers. They found “that significant long-term serial correlation as measured by the Hurst K statistic is present in a large number of rivers, and that when a peak flow series shows short-term independence, there is still a fairly high probability of long-term dependence. This long-term dependence cannot be disregarded as in traditional flood frequency analysis; it should be taken into account as it may significantly increase the risk associated with future peak flows.” Serinaldi & Kilsby [2] investigated 473 daily stream flow time series across the major river basins in Europe. One of their results is: “the standardized return intervals between over-threshold values exhibit a sub-exponential Weibull-like distribution, implying a higher frequency of return intervals longer than expected under independence, and expected return intervals depending on the previous intervals; this results in a tendency to observe short (long) inter-arrival times after short (long) inter-arrival times.” Todhunter [3] presented a case, of a 120-year long record of a river in North-Western USA, through which he questioned assumptions behind the estimation of the regulatory flood. In this case, the coefficient of variation of annual maximum discharges exhibits a high degree of inter-annual variability, with flood peaks occurring in clusters of values higher or lower than the median peak. That series exhibits statistically significant serial correlation and it is not a true set of random and independent events, required for the common predictions of recurrence intervals.

Ben-Zvi [4] showed a case where the negative binomial distribution is superior to the Poisson distribution in counting runoff events. The negative binomial of events’ counting is consistent with the gamma distribution of events’ inter-arrival intervals. Ben-Zvi [5,6] showed a case where the gamma distribution is superior to the exponential distribution in representing inter-arrival intervals of runoff events. Anderson [7] proposed using the Erlang distribution for modeling inter-arrival times between peak flows. This distribution is related to the gamma distribution. Stoynov et al. [8] proposed a switch-time distribution for modeling inter-arrival times of floods. This distribution is also related to the gamma distribution. Stoynov et al. [9] showed its fit to long-term records of Chinese rivers.

Eichner et al. [10], Koscielny-Bunde et al. [11], Eichner et al. [12], and Lennartz et al. [13] showed the effect of long-term correlation on persistence in a number of phenomena, including floods, and the power-law description of their inter-arrivals. Benson et al. [14] concluded that if only the tail properties of the exceedances and inter-arrivals can be estimated, this class of extreme value densities can be used to obtain recurrence intervals for extreme events with power-law inter-arrivals. Turcotte & Green [15] and Malamud & Turcotte [16] derived, through regressions, power-law relationships between threshold discharges and the mean inter-arrival intervals of their exceedances. Ben-Zvi & Azmon [17] applied a non-parametric procedure, for relating the mean and the standard deviation, of inter-arrival intervals of peak runoff discharges, to threshold discharges. Their results lead to power-law relationships. Ben-Zvi [18] proposed applying this procedure for predicting flood risk. The present article revises that proposal.

Materials and Methods

Table 1: Watersheds and records.

No.	Stream @ station	Area (km^2)	Mean precipitation (mm)	Observation period	Years
1	Hazor @ Ayyelet HaShahar	34	650	1944/5 – 2018/9	71
2	Keziv @ Gesher HaZiv	130	800	1944/5 – 2018/9	72
3	Hillazon @ Yas’ur	163	640	1943/4 – 2018/9	72
4	Hadera @ Gan Shmuel	579	660	1949/0 – 2018/9	70
5	Alexander @ Elyashiv	488	630	1948/9 – 2018/9	67
6	Lakhish @ Yavne – Ashqelon Rd.	996	540	1943/4 – 2018/9	68
7	Shiqma @ Beror Hayil	376	450	1951/2 – 2018/9	68
8	Besor @ Re’im	2586	275	1964/5 – 2018/9	54
9	Zin @ Masos	674	70	1955/6 – 2018/9	59

Notes:

1. The years follow the Israeli meteorological determination, beginning on August 1.
2. The difference between the number of years and the observation period is due to missing data for certain years.
3. A peak discharge, of about 25 m^3/s, was recorded at the site of Hazor station, prior to the observation period.
4. A peak discharge, of about 450 m^3/s, was recorded at the Lakhish station, after the considered observation period.
5. The second highest peak discharge at Hadera station (occurred in January 1960) is estimated, by the author, through comparison of its flooding extent with those of the first and the third highest peak discharges. Peak discharges, of the two other events, occurred in that year, are estimated through comparison with corresponding discharges at Alexander station. They were very low.

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

Records of nine hydrometric stations, of the Israel Hydrological Service, are analyzed here. Their watersheds are exposed to Mediterranean and arid climates, where precipitation falls during the cool winter season and the summers are hot and dry. The selected streams carry direct surface runoff only. Their watersheds are delineated in Figure 1. Main properties of the watersheds and the records are presented in Table 1. Magnitude and occurrence time of peaks of runoff events comprise the data considered here. By definition, successive events are discerned from one another by cessation of flow for at least 24 hours. Events, composing the entire flow during a year, are included in the computation. Years of missing data are skipped. Accordingly, the computed interval, between the last event before a skipped year and the first event after that year, is one year shorter than the actual interval between these events. This involves loss of the information that very high discharges did not occur during a skipped year. An exception to this procedure, where a very high discharge did occur in a potentially skipped year, is described in the last note following Table 1.

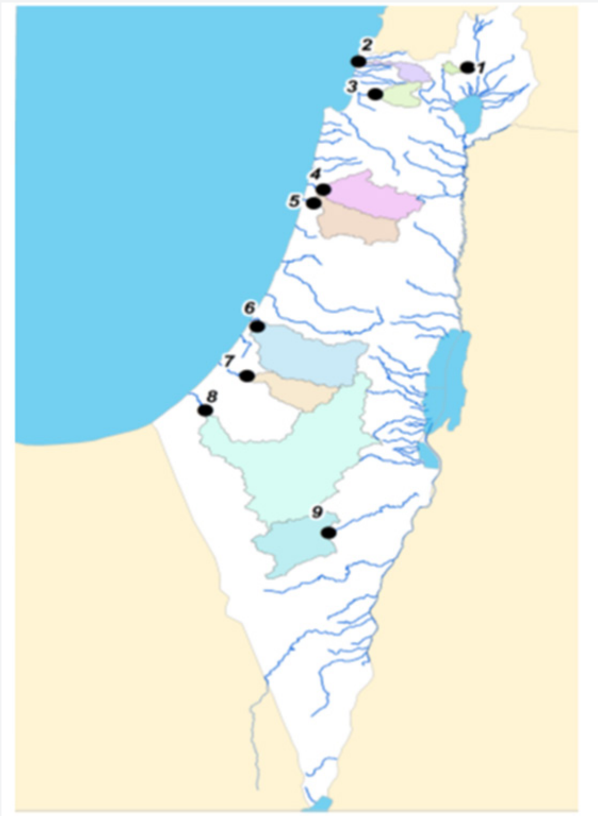

Figure 1: Map of the considered watersheds and station locations.

The present article deals with statistical parameters of the intervals between arrivals of peaks of events exceeding given thresholds. The sum of actual intervals leaves out the two external sub-intervals: from the beginning of the observation period till the first exceedance of the threshold; and from the last exceedance of the threshold till the end of the observation period. To regain this loss, these sub-intervals are added to each other, making an interval that joins the actual ones. By that, the average interval is kept equal to the number of years on record divided by the number of exceedances. The intervals, associated with a given threshold, are nested within those associated with lower thresholds. This causes a certain dependence between statistical parameters of the intervals associated with different thresholds. Therefore, the relationship between parameters and thresholds should be developed through a procedure tolerant to this dependence. Such a procedure is the non-parametric scheme, applied by Ben-Zvi & Azmon [17]. There, it leads to power-law relationships between those variables. The present article applies another variant of that scheme, which leads to an apparently better representation of the actual situation. The selected scheme considers the slopes between realizations of the variables. One variable is threshold discharge and the other is the logarithm of either the mean or the standard deviation of the inter-arrival intervals of peak discharges exceeding (or equal to) the threshold. A slope is the difference between any two realizations of the latter variable divided by the difference between the corresponding

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

realizations of the former variable. The median of all pair-wised slopes, positioned at the point of median values of both variables, makes the sought relationship (e.g. Brauner [19]; Newson [20]; Huth and Pokorna [21]).

Analysis

Figure 2 presents the coefficients, of variation and of skewness, against the mean of the inter-arrival intervals between exceedances of recorded discharges at the different stations. The mean interval ranges from 0.5 year to the number of years on record. The figure reveals a very few cases in which the coefficient of variation is about 1 while the skewness coefficient is about 2. None of these cases is found where the mean interval is at least 5 years. This indicates that arrivals of high events, at the studied stations, do not follow the exponential distribution.

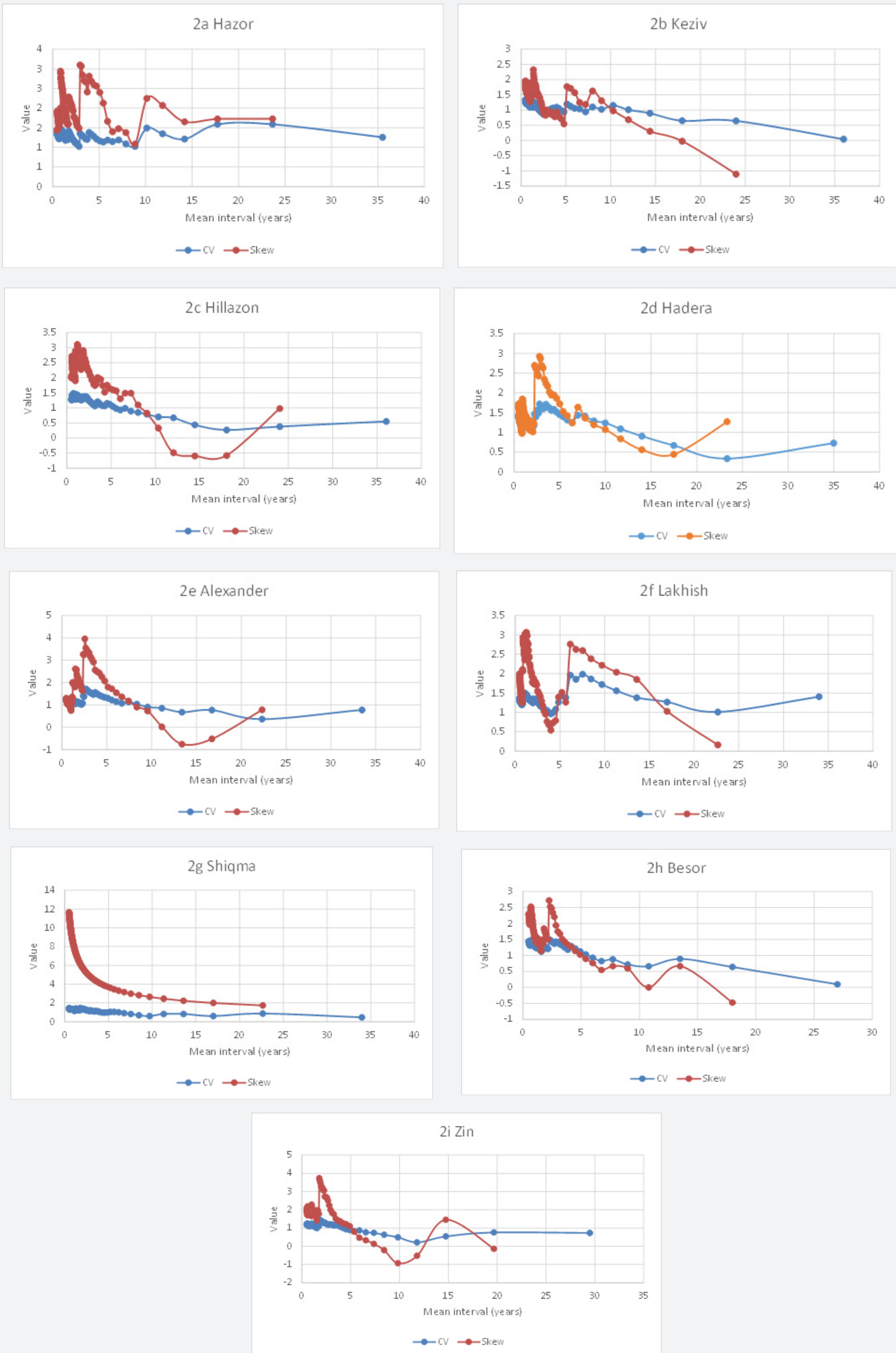

Figure 2: Statistical parameters of the inter-arrival intervals.

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

In search for a representative relationship, the mean intervals are plotted, in Figure 3, against their corresponding threshold discharges. For each station, the data points follow a curve of complex formulation. Use of a linear scale for the discharges and a logarithmic scale for the intervals, as is done in Figure 3, places the upper data points along an about straight line. It means a semi-logarithmic relationship for the upper tails of the variable values. This relationship commences, for the different stations, at a mean interval of about 2.5 to 8.5 years. This observation leads to development of the semi-logarithmic scheme described above. Mean intervals, predicted through the proposed scheme, for the high discharges at the different stations, are presented in Figure 4 along with the recorded ones. For each station, the minimum threshold is selected where the ratio, of predicted to recorded mean interval, is closest to 1. Details of this fit are presented in Table 2.

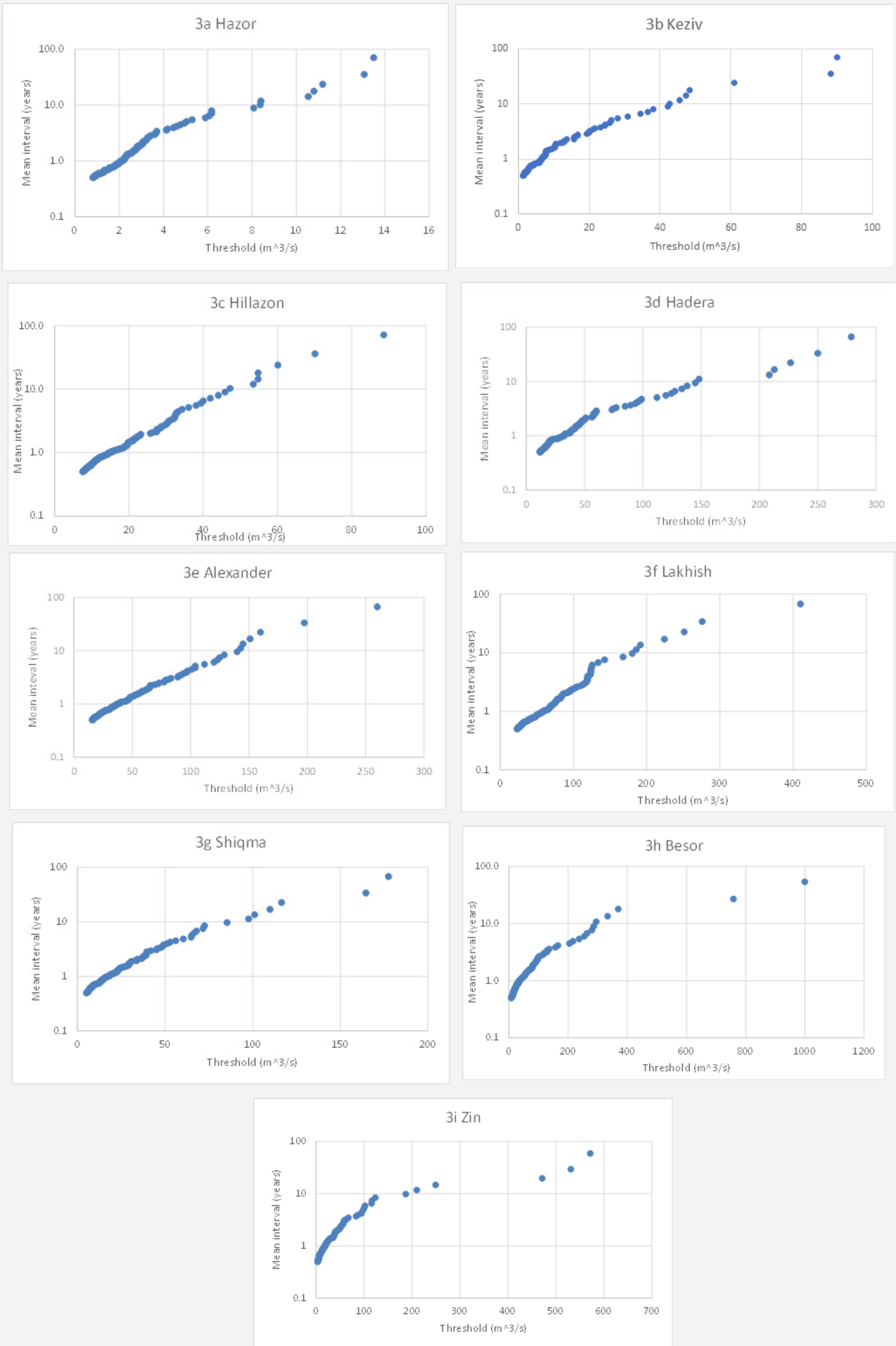

Figure 3: Increase of mean interval with threshold discharge.

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

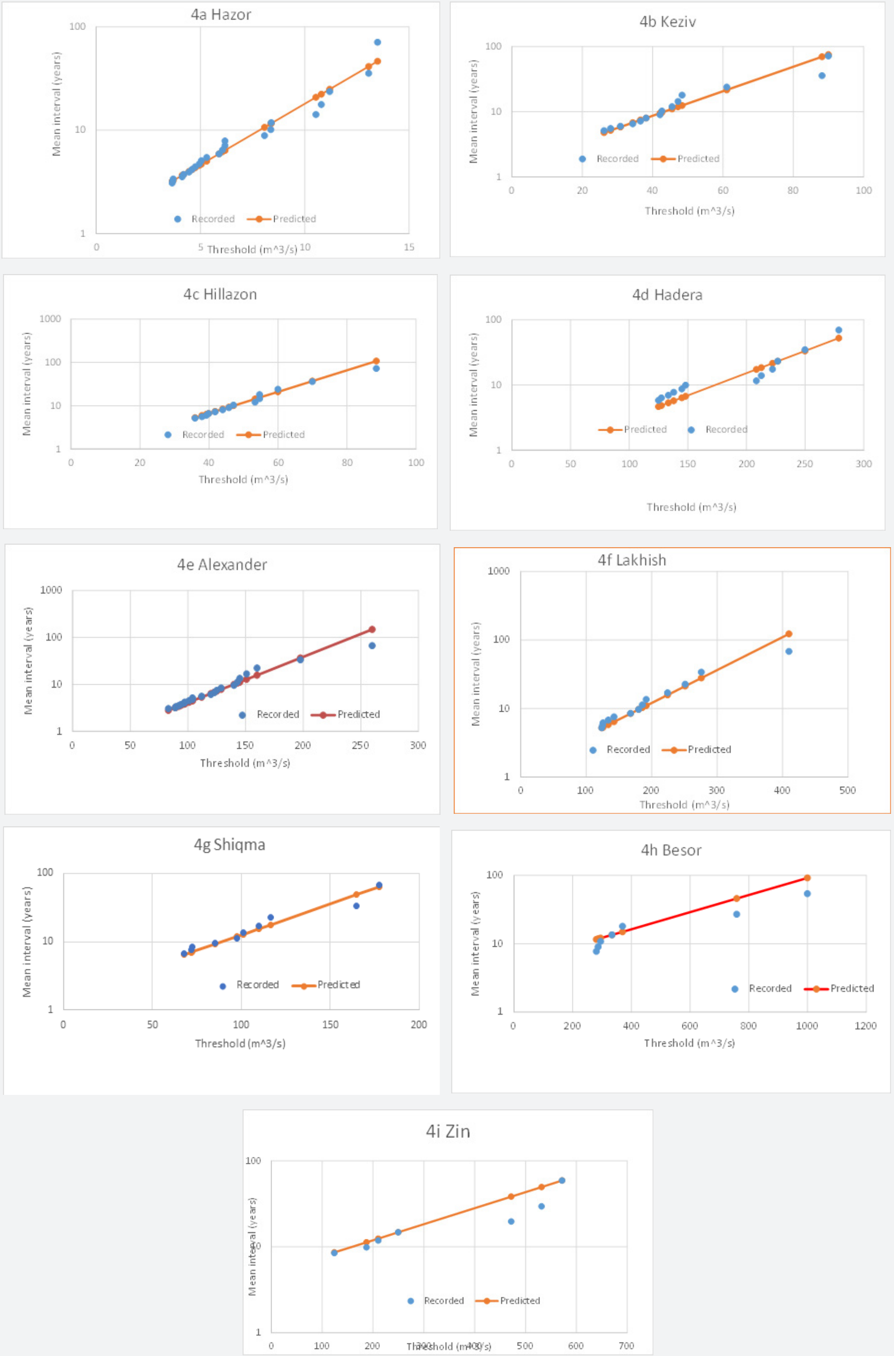

Figure 4: Semi-logarithmic relation of mean interval with threshold discharge.

For comparison, results of prediction, by use of the Generalized Pareto (GP) distribution, are summarized in Table 3. The prediction procedure is described in Ben-Zvi [22]. Goodness-of-fit, of the GP distribution to the records, is concluded through the upper-sided Anderson-Darling test. In the present case, it ranges from excellent to good. Figures, presenting this fit, are not shown here. The differences, between the number of years on record and the mean interval or the recurrence interval, predicted through the two alternatives, for the highest recorded discharges at the different stations, are compared in Table 2. For six stations, the difference predicted through the semi-logarithmic scheme, is smaller than that predicted through the GP scheme; while for the three other stations, the opposite situation is found. The average ratio, of predicted interval to number of years on record, is 1.29

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

for the semi-logarithmic scheme and 1.80 for the GP fit. These indicate that prediction through the semi-logarithmic scheme is more realistic than that through the GP fit.

To enable fit of a two-parameter distribution (e.g. the gamma) to the inter-arrival intervals, a scheme for predicting the standard deviation of the intervals is next developed. A figure, similar to Figure 3 (not shown), reveals that the increase of standard deviation with threshold discharge is similar to that of the mean. It is steep for low thresholds and slows down at medium discharges. For the high thresholds, their relationship, with the standard deviation of the intervals, follows an about semi-logarithmic trend. Yet, for very high thresholds at certain stations, it attains a wavy shape. Figure 2 indicates a smooth relationship between the coefficient of variation and the mean. These lead to construction, for the standard deviation, the same scheme developed for the mean. For consistency reason, the minimum thresholds, selected for the mean, are adopted for the standard deviation. For two stations this consistency required a small decrease in the minimum threshold for the mean interval. The recorded and predicted standard deviations are presented in Figure 5. Their average ratios, for the different stations are presented in Table 2.

Table 2: Summary of semi-logarithmic predictions.

No	Years	Min. (years)	Ratio of mean	T(Qmax) (years)	Ratio of std
1	71	3	1.019	47	0.993
2	72	5	1.02	76	0.923
3	72	5	1.015	107	0.988
4	70	5.5	0.936	53	1.219
5	67	3	0.998	149	1.017
6	68	5	0.978	123	1.067
7	68	6.5	0.974	64	1.064
8	54	7	1.308	92	2.832
9	59	8	1.265	59	1.218

Notes: No is station number; Years is number of years on record; Min. is mean interval associated with the minimum threshold selected for the prediction; Ratio of mean is average ratio of predicted to recorded mean interval; T(Qmax) is predicted mean interval for the highest recorded discharge; Ratio of std is average ratio of predicted to recorded standard deviation.

Table 3: Prediction through the GP distribution.

No	Size	UA^2	T(Qmax) (years)
1	90	0.186	57
2	133	0.255	69
3	94	0.109	228
4	96	0.283	47
5	99	0.049	134
6	70	0.193	256
7	109	0.052	89
8	68	0.112	114
9	90	0.099	84

Notes: Size is of the series to which the GP distribution is best fitted; UA^2 is result of upper-sided Anderson-Darling goodness-of-fit test for this series; T(Qmax) is predicted recurrence interval for the highest recorded discharge.

Discussion

Predicted means and standard deviations, of inter-arrival intervals, are computed for the range of recorded discharges exceeding the minimum threshold selected for each station. Predictions, through this range, as well as through upwards extrapolations, are justified by the assumption that future high discharges would follow the relationships observed here. A similar assumption prohibits any downwards extrapolation of the semi-logarithmic scheme, because it is valid only for thresholds exceeding the minimum selected. The statistical dependence, between inter-arrival intervals, associated with different thresholds, requires application of a scheme which is not sensitive to such a dependence. The behavior of available data justifies selecting the semi-logarithmic scheme, rather than the logarithmic one selected by Ben-Zvi and Azmon [17]. The predicted mean intervals are found closer, to the mean recorded intervals, than the recurrence intervals, predicted through a good

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

application of a current procedure [22]. The prediction, of both the mean and the standard deviation of inter-arrival intervals, enables fit of a two-parameter statistical distribution (e.g. the gamma) to arrivals of high runoff discharges. These would allow consideration of clustering, observed in the cited articles, and improve assessments of flood risks and determination of design discharges.

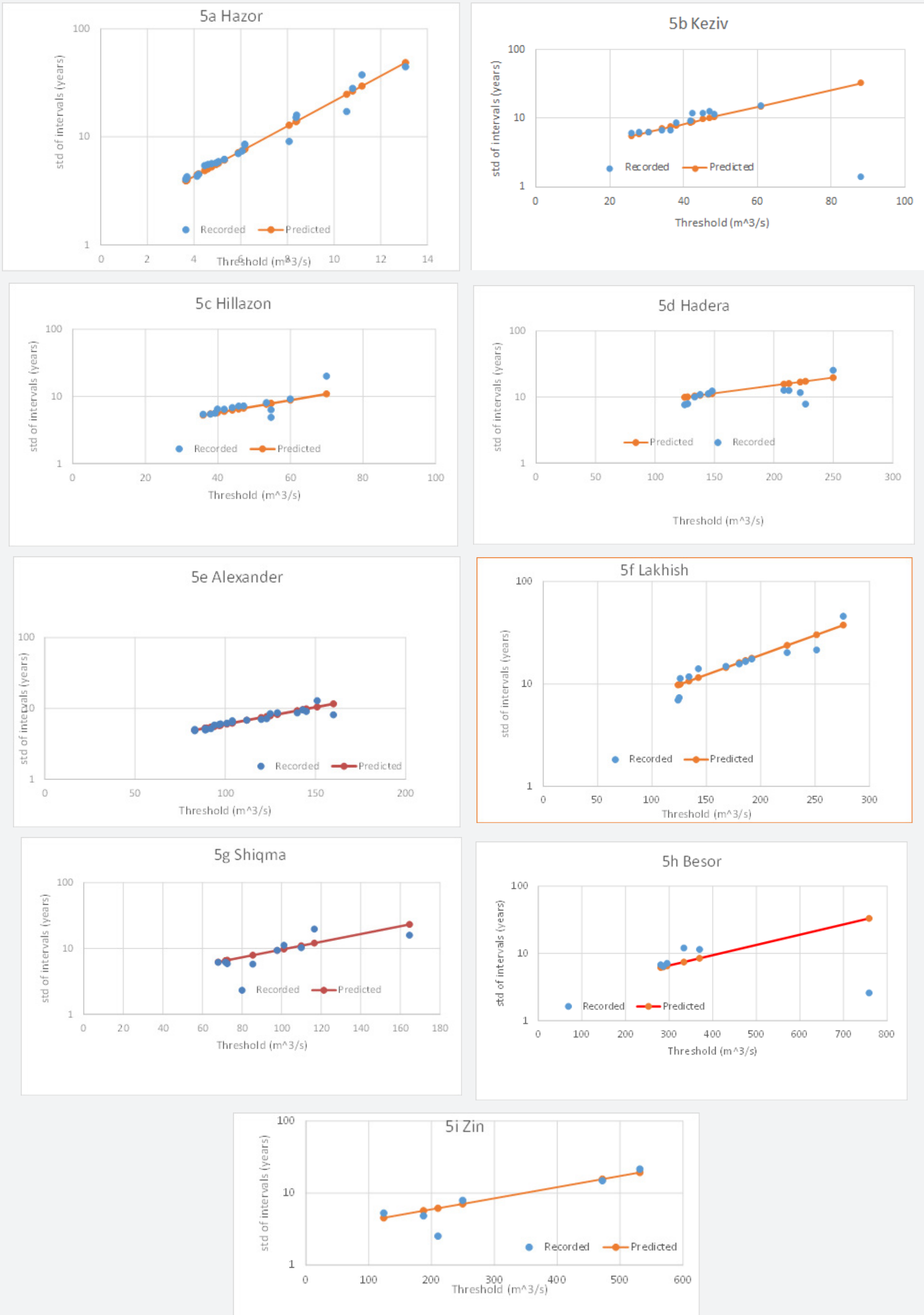

Figure 4: Semi-logarithmic relation of mean interval with threshold discharge.

Conclusion

1. There are cases, where the distribution of inter-arrival intervals of high discharges differs from that described by the commonly selected exponential distribution. There, a two parameter distribution would better fit the records.

2. Because the intervals between exceedances of a given threshold are nested within those of lower thresholds, the

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802

statistical parameters of the intervals, associated with different thresholds, are inter-dependent with one another. The non-parametric procedure, selected here, is tolerant to this dependence and, therefore, suits for basing a relationship between thresholds and parameters of those intervals.

3. In the present case, the logarithms of the mean and of the standard deviation of the intervals, associated with high thresholds, are found linearly related to the threshold discharges. This justifies selection of the semi-logarithmic procedure for formulating the relationship between these variables.

4. Comparison of predicted mean intervals, associated with the highest recorded discharges at the different station, with recurrence intervals predicted for these discharges, through a good fit of the GP distribution, indicates that, on average, the former results are closer than the latter one, to the number of years on records.

Acknowledgement

Rita Chudinov, of the Israel Hydrological Service, has provided the data, on which the article is based. She also prepared Figure 1.

References

1. Lye LM, Lin Y (1994) Long-term dependence in annual peak flows of Canadian rivers. J. Hydrol 160(1-4): 89-103.
2. Serinaldi F, Kilsby CG (2016) Understanding persistence to avoid underestimation of collective flood risk. Water 8(4): 152.
3. Todhunter PE (2012) Uncertainty of the assumptions required for estimating the regulatory flood: Red River of the North. J Hydrol Engrg ASCE 17(9): 1011-1020.
4. Ben-Zvi A (1991) Observed advantage for negative binomial over Poisson distribution in partial duration series. Stoch Hydrol Hydr 5: 135-146.
5. Ben-Zvi A (1999a) Distribution of intervals between flood events. Hydrologic Modeling, Proc. Intl. Conf. Water, Environ., Ecol. Socio-Econom. & Health Engrg., Seoul, Korea (eds. V.P. Singh et al.), Water Resour. Public. Highlands Ranch, CO, USA pp: 255-261.
6. Ben-Zvi A (1999b) Distribution of intervals between flood events. Proc. ASCE WRE Conf. Water Resources into the New Millennium: Past Accomplishments and New Challenges, Seattle, WA, USA, CD-ROM by Water Resources Publ.
7. Anderson MG (1976) Peak flows inter-arrival times. J. Hydrol 30: 105-112.
8. Stoynov P, Zlateva P, Velev D (2015) An approach for modeling inter-arrival time of floods. Int. J. Innov. Manag. Technol. 6(4), 267-271.
9. Stoynov P, Zlateva P, Velev D, Zong X (2017) Modeling of major flood arrivals on Chinese rivers by Switch-time processes. Proc. 3rd Intern. Conf. Advance. Environ. Res., IOP Publ.
10. Eichner JF, Koscielny-Bunde E, Bunde A, Havlin S, Schellnhuber HJ (2003) Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. Phys. Rev. E 68(4): 1-5.
11. Koscielny-Bunde E, Kantelhardt JW, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. J. Hydrol., 322(1-4): 120-137.
12. Eichner JF, Kantelhardt JW, Bunde A, Havlin S (2007) Statistics of return intervals in long-term correlated records. Phys. Rev. E, 75: 011128: 1-9.
13. Lennartz L, Livina VL, Bunde A, Havlin S (2008) Long-term memory in earthquakes and the distribution of interoccurrence times. A Letters J. Exploring the Frontiers of Physics EPL81, 69001.
14. Benson DA, Schumer R, Meerschaert MM (2007) Recurrence of extreme events with power-law inter-arrival times. Geophys. Res. Lett. 34 L16404: 1-5.
15. Turcotte DL, Green L (1993) A scale-invariant approach to flood frequency analysis. Stoch. Hydrol. Hydr 7: 33-40.
16. Malamud BD, Turcotte DL (2006) The applicability of power-law frequency statistics to floods. J. Hydrol., 322(1-4): 168-180.
17. Ben-Zvi A, Azmon, B (2010) Direct relationships of discharges to the mean and standard deviation of the intervals between their exceedences. Hydro. Sci. J 55(4): 565-577.
18. Ben-Zvi A (2018) Predicting flood risk from interarrivals of peak runoff events. J Hydrol Engrg ASCE 23(5), 02518001-1:3.
19. Brauner JS (1997) Environmental sampling of slope: Sen's method in environmental pollution. In: Environmental Sampling and Monitoring Primer (ed. by D. Gallagher).
20. Newson R (2002) Parameters behind nonparametric statistics: Kendall's tau, Somers' D and median differences. Stata J. 2(1), 45-64.
21. Huth R, Pokorna L (2004) Parametric versus non-parametric estimates of climatic trends. Theoret. Appl. Climatol 77: 107-112.
22. Ben-Zvi A (2016) Selecting series size where the Generalized Pareto distribution best fits. J. Hydrol. 541: 778-786.

How to cite this article: Arie B-Z. Semi Logarithmic Relation of Interarrivals Mean and Standard Deviation with High Threshold Discharges. Civil Eng Res J. 2020; 11(1): 555802.DOI: 10.19080/CERJ.2020.11.555802