A statistical model for extreme precipitation

by

Jonas Eliasson

A paper to be submitted for publication in:

Water Resources Research
Author’s address and affiliation:

Jonas Eliasson, Professor
University of Iceland, Faculty of Engineering, Department of Civil Engineering
Hjardarhaga 6,
IS–107 Iceland

Phone: +354–525 4651   Fax: +354–525 4632   E–mail: jonase@verk.hi.is
A STATISTICAL MODEL FOR EXTREME PRECIPITATION

JONAS ELIASON

University of Iceland, Department of Civil Engineering

The statistical distribution of one-day (one reading in 24-h) and 24-hour (several readings in 24 h) annual maxima is considered and a transformed Extreme Value type 1 distribution function (TDF) that includes a probable maximum (PM) value is suggested. The distribution function fits standardised annual maximum station values from Iceland and Washington State very well. A generalised distribution function, derived from the TDF, is suggested. To use it, two local parameters have to be known, the five-year event, $M_5$, that must be picked from a map, and a slope factor, $C_i$ that is a function of the coefficient of variation. The variation of $C_i$ between independent observation stations is assumed to be random, and guidelines on how a $C_i$ may be selected are discussed. The generalised distribution function is used to calculate quantile estimates and a local Probable Maximum Precipitation (PMP). Regional PMPs can be calculated by maximising this value. Two independent sets of $C_i$s, 1) a statistical set compiled from the British Natural Environment Resource Council (NERC) Probable Maximum Precipitation (PMP) envelope, and 2) a meteorological set calculated from US National Weather Service estimates of PMP for Washington State – compare favourably. The regional PMP estimates calculated from the generalised distribution also compare favourably with the NERC PMPs, except that the estimates for low $M_5$ produce up to 33% lower PMPs. This difference may be explained by a number of factors that are also discussed.

1. Introduction

Statistical methods are commonly used to estimate extreme precipitation. Estimates of Probable Maximum Precipitation are an exception to this rule. Here meteorological estimates [DaoJiang and JinShang, 1984; Hansen, 1987; Rakhecha and Kennedy, 1985] are more common. Statistical methods that include methods to estimate PMP are also well known. In the United States Hershfield suggested an estimation method consisting of adding 15 standard deviations to the mean. Later he revised the method to include a Frequency Factor that varied with the mean instead of the constant 15 [Hershfield, 1965]. His proposals and later refinements are now known as the Hershfield method [WMO, 1986]. In Britain a more systematic method was introduced by NERC [Flood Studies Report, 1975], based on a thorough study of the tails of the distributions of annual maxima. The NERC method has been adapted recently to Norwegian conditions [Förland and Kristoffersen, 1989].

PMP station values have been computed for Iceland [Elíasson, 1991] using the NERC envelope curve. These estimates are uncertain in Iceland because of the scarcity of information on precipitation, so it is necessary to look for generalised methods to estimate extreme precipitation. The density of meteorological observations is low, especially in the interior of the country. For many stations, very few years of record are available. In short,
the Icelandic meteorological data are inadequate for fully utilising the PMP estimation methods available, because the associated envelope curves cannot be estimated from local data.

In 1991 the Engineering Research Institute of the University of Iceland and the National Power Company initiated a research program to find a method to calculate extreme precipitation values in Iceland that was consistent with methods applied elsewhere. One part of the study is to analyse one-day precipitation maxima in Washington State to uncover possible similarities between these two presumably independent climatic regions. An important reason for choosing Washington State data is that precipitation in Iceland and western Washington is of the same order of magnitude.

Two points are of particular interest: on one hand, whether the same distribution function can be applied to both areas, and on the other, whether statistical NERC PMP estimates are consistent with the meteorological estimates of the US National Weather Service (NWS). It is particularly important to find out if the NERC PMP envelope curve, together with the generalised distribution function, can reproduce the US NWS PMP values. Such consistency would strengthen the argument for using the NERC envelope to estimate PMPs in Iceland.

2. Distribution of annual precipitation maxima

Annual maximum one-day precipitation is usually considered to be an independent identically distributed (i.i.d.) stochastic variable. An i.i.d. may have any distribution function, and the maximum will be distributed according to the Extreme Value distribution function to whose domain of attraction it belongs, [Leadbetter et al., 1983].

The most commonly applied Extreme Value distribution function is the type 1 ($EV1$) or the Gumbel distribution. $EV1$ is in its own domain of attraction so both the i.i.d. and its maximum will be asymptotically distributed according to $EV1$, when this distribution function is used. The distribution function for annual maxima of individual meteorological stations will often follow the $EV1$ very closely in the medium-range of values, but deviate from it for the highest and lowest return periods. These deviations and $EV1$’s lack of an upper limit complicate the use of this distribution function in estimating PM-values.

Problems of this nature may be solved by using a Transformed Distribution Function (TDF) and Cut-off Distribution Function (ODF), both derived from the Basic Distribution Function $EV1$ (BDF). They are defined as follows [Elíasson, 1994].

\[
\text{BDF: } F(x) = \exp(-\exp(-z))
\]

\[
\text{TDF: } F(x) = \exp(-\exp(-z + \frac{k}{y_{lim} - z}))
\]

\[
\text{ODF: } F(x) = \exp(-\exp(-z)); \quad z < y_{lim}; \quad F(x) = 1; \quad z \geq y_{lim}
\]

(1)
\[ z = \frac{x}{a} + b \quad y_{\text{lim}} = \frac{x_{\text{PM}}}{a} + b \]

\[ y : \text{EV1's reduced variate} = -\log(-\log(P(X < x))) \]
\[ y_{\text{lim}} : \text{Limiting reduced variate} \]
\[ a : \text{Scale parameter} \]
\[ b : \text{Location parameter} \]
\[ PM : \text{Index for probable maximum} \]
\[ k : \text{A negative constant} \]

The ODF is the limit of the TDF when \( k \to 0 \). It belongs to the domain of attraction of \( EV1 \) [Eliasson, 1994].

It may be noted that using TDF in data analysis is equivalent to using the data transformation:

\[ X^6 = X - \frac{a^2 k}{X_{\text{PM}} - X} \]

and then estimate an EV1 distribution for this variable. This transformation has many of the properties of the logarithmic transformation that is the basis of the log-distributions (log-Gumbel, log-Normal, log-Pearson etc.).

In this paper it will be assumed that annual maximum precipitation is an i.i.d. variable \( X \). We further assume:

\[ I : \text{The distribution function of } X \text{ is a TDF.} \]
\[ II : y_{\text{lim}} > 9 \]
\[ III : k > -1 \]

The reason for these assumptions is discussed in the next section.

### 3. Standardisation and pooling of data

If assumption I is true, the annual max data of each station are distributed according to (1). In an \( EV1 \) plot the data will follow the BDF up to a certain point that depends on the value of \( k \) and \( y_{\text{lim}} \), and there it will deviate from the \( EV1 \) line in the direction of higher \( y \) values.

In order to see this trend in one station we need more observation years than are available in one meteorological station. This difficulty may be overcome by regional pooling of data. Data from different meteorological stations have different means and standard deviations. The following standardisation is used to deal with these differences:
\[ \xi = \frac{X_i - \bar{X}_j}{S_j} \]  
(2)

\[ \bar{X}_j : \text{Estimated mean of annual maxima at station } j \]  
\[ S_j : \text{Estimated standard deviation at station } j \]

Using this standardisation, the estimated BDF is the same for all stations as long as it is a two parameter EV1 distribution function because all stations are now with an estimated zero mean and a standard deviation of one. The BDF does not have to be EV1 as stated in assumption I, but if it is, the pooled standardised data will plot on a line in an EV1 plot, and the line will be straight apart from the previously mentioned deviations that depend on the value of \( k \) and \( y_{lim} \).

It should be noted that standardisation with one parameter only will leave the data inhomogeneous and therefore unsuited to pooling. For example, if the data are standardised only by dividing with the local mean, this will leave the scale parameter estimate different from station to station, and pooling becomes problematic.

If assumption II holds, first and second moments will be almost identical for the ODF and the BDF. They will differ in the fourth digit only, [Eliasson, 1994, appendix], so estimates of scale and location parameters can be made without knowing the limiting reduced variate. If this assumption holds we can use the model without a regional estimate of the limiting reduced variate. We do not have to use its value except in PMP estimates, and in this case an envelope curve may be used to define the PM-value, as is done in the Hershfield and NERC methods. Using ODF instead of BDF has the advantage that ODF has the upper limit that BDF does not have.

If assumption III holds, the difference between ODF and TDF will be so small that the estimates of scale and location parameters in the TDF can be made without knowing the actual value of \( k \). In this respect TDF is simpler than the EV2 which curves to the same side as TDF and contains an upper limit.

The validity assumptions of I–III has to be demonstrated after the BDF has been tested and found valid. If the BDF is not valid, it will be clearly revealed because the plot of the pooled \( \xi \) values will simply not fit on a single EV1 line.

To summarise, if assumptions I–III hold, we can standardise the annual maxima for each station with its estimated mean and standard deviation. The estimates of \( k \) and \( y_{lim} \) may vary from station to station, but the pooled data will still be homogeneous with respect to an estimation of the scale and location parameters of the distribution which have the theoretical values

\[ \frac{1}{a} = C_2 = 1.283 \quad b = C_1 = 0.5772. \]
4. Fitting the model to rainfall data

Model (1) with assumptions I-III fits measured values well. The fit to Icelandic data is demonstrated in Eliasson (1994). This section shows how Model (1) fits data from Washington State USA (abbreviated WA in figures, tables and subscripts). The data is divided into two parts: 1) climatic regions 1 and 2 (Eastern Washington with an arid and semi-arid climate), and 2) climatic regions 3, 4 and 5 (the coastal areas in Western Washington) in the other part.

Fig. 1. *EV1* plot of 123 stations in regions 1 and 2 in Washington State.
The model, fit to one-day and 24-h data for Washington State, is shown in Figs. 1 and 2. A prior study of the Washington State data [Schaefer, 1990] shows the magnitude of serial correlation to be so small as not to have any appreciable effect on the quality of regional solutions. The cross-correlation is highly variable [Schaefer, 1990]. The possible effect of such variability will be discussed later in this paper.

Fig. 1 includes all meteorological stations with 20 or more one-day or 24-h annual maxima in climatic regions 1 and 2 in Eastern Washington. There are 5885 points altogether. The 4000 highest points are shown. All but about 50 of the highest points follow the BDF, which is the straight line in the figure. Fig. 2 covers climatic regions 3, 4 and 5 in Western Washington. There are 5615 points altogether, but about 4000 are shown. Again, about 50 of the highest points deviate from the BDF.

Table 1 shows the values of the linear regression values for the parameters $1/a$ and $b$. Their theoretical values, sometimes called $C_2$ and $C_1$, are listed for comparison in the last row. The table shows clearly that the influence of the 50 highest points is very small, so the BDF shows a very good fit whether the highest points or not are considered.

The scale parameters are slightly higher than the theoretical parameters when all points are used. This difference stems from the pooling of the data. The standard deviation of the pooled sample becomes lower than one by approximately 1%. The reason is that the usual estimate for standard deviation includes the sample size minus one. When standard deviation is found for the pooled sample we divide with a slightly higher number than the sum of the numbers used for individual stations.
Table 1
Comparison of theoretical and estimated parameters

<table>
<thead>
<tr>
<th>Data</th>
<th>Item</th>
<th>1/a</th>
<th>b</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA 1 - 2</td>
<td>Fig.1</td>
<td>1.295</td>
<td>0.578</td>
<td>All 5885 points</td>
</tr>
<tr>
<td>WA 1 - 2</td>
<td>Fig.1</td>
<td>1.290</td>
<td>0.577</td>
<td>5835 lowest</td>
</tr>
<tr>
<td>WA 3 - 5</td>
<td>Fig. 2</td>
<td>1.295</td>
<td>0.578</td>
<td>All 5615 points</td>
</tr>
<tr>
<td>WA 3 - 5</td>
<td>Fig. 2</td>
<td>1.282</td>
<td>0.573</td>
<td>5565 lowest</td>
</tr>
<tr>
<td>EV1, BDF</td>
<td>$C_2$ and $C_1$</td>
<td>1.283</td>
<td>0.5772</td>
<td>10000 points</td>
</tr>
</tbody>
</table>

Fig. 3  TDF, $k = -0.6$, $y_{lim} = 9.2$ fitted to WA data, regions 1 – 5

Fig. 3 shows the data points that deviate from the BDF, which is the straight line in the figure. All the data points deviate to the same side as predicted by the assumed TDF and fit this distribution function with $k$ and $y_{lim}$ within the limits prescribed by assumptions II and III reasonably well. However, this evidence is insufficient to provide a reliable estimate of $k$ and $y_{lim}$ from the data and only indicates the validity of assumptions II and III. One reason is that the high values with large deviations in Fig. 3 are very few, and their deviation from the BDF line is too small to indicate clearly whether there is an upper limit that may be identified as a PM value.

Another reason is the cross-correlation in the original data, but cross-correlation produces an effect similar to the deviations in Fig. 3 because then the Equivalent Independent Record Length (EIRL) is shorter than the actual record length. Schaefer (1990) suggested
that the EIRL of the Washington State data is between 33% and 47% of the actual record length. This may be due to the cross-correlation between station pairs [Buishand, 1984]. This will not affect the validity of the BDF, but cross-correlation in the data might eventually lead to the conclusion that actual \( Y \) values of the largest data points are significantly overestimated; the "true independent" probabilities of the corresponding \( \xi \) values are smaller, and the points should be closer to the BDF line than the data shows. It is possible, however, that the maxima of station pair series are asymptotically independent [Buishand, 1984], so that the dependency between values in series pairs relates to magnitude, and this relation may be different for convective and non-convective precipitation situations. This possibility will be discussed further later in the paper.

If the distribution function is accepted as a suitable model, the following formulas can be derived for the probability of extreme rainfall at individual stations.

For a given annual maximum precipitation volume, accumulated during a day in station \( j \), the probability of non-exceedance may be calculated using (1), substituting \( z \) into ODF and taking \( 1/a \) and \( b \) from table 1:

\[
P(X_j \leq x) = \exp \frac{\exp(-x^\phi)}{\sum_{i} x^\phi_i} \exp\left(- \left(1.283 \frac{x - \mu_j}{\sigma_j} + 0.557\right)\right)
\]

With a given probability of non-exceedance \( P \), \( x_j \) becomes:

\[
x_j = \sigma_j \left(-\log(-\log(P)) - 0.5572\right) / 1.283 + \mu_j
\]

\( \mu_j \): True average at station \( j \)

\( \sigma_j \): True standard deviation at station \( j \)

To apply these formulas to a particular precipitation area one has to know both the mean and standard deviation in the area. Another possibility is to use a generalised distribution similar to the one devised by NERC [FSR, 1975].

5. Generalised distribution

The TDF (1) may be used as a generalised distribution when the parameters \( k \) and \( y_{lim} \) are known. These parameters have a great effect when the return period is great, but with small return periods their effect is negligible. A large amount of data is therefore needed to estimate them properly as has been demonstrated in the previous section in the case of the Washington State data. The same result has been obtained for Icelandic data [Elíasson, 1994].

Until a reliable estimate of the parameters \( y_{lim} \) and \( k \) in TDF is obtained, it is suggested to use ODF together with an \( y_{lim} \) value that may be derived from the envelope curve for the
growth factors in the NERC report [FSR, 1975]. From this envelope curve it is possible to derive the following relation for $y_{lim}$ [Elfasson, 1994]:

$$y_{lim} = 10.7 - 0.0071M \quad 25 < M5 < 200 \text{ mm/day} \quad (3)$$

$M5$: The 5-year event ($x$ for $P = 0.8$ or $y = 1.5$)

Even though (3) is not in [FSR 1975], the envelope curve is used, and the same envelope curve is also used almost unchanged in the results published by [Förland and Kristoffersen 1988]. In neither case is $y_{lim}$ included in the distribution. But as the same $y_{lim}$ can be used in both countries, this region could be extended even further to get more data. Thus, we finally could arrive at an estimate of both $y_{lim}$ and $k$ by the same methods as used on the Washington State data. This data also fits an $y_{lim}$ value of 9.2 that is a value to be expected for Washington State from (3).

The ODF may now be written as follows:

$$x = M5(1 + C_i(y - 1.5)) \quad y \leq y_{lim} \quad (4)$$

$$x = x_{PM} \quad y \geq y_{lim}$$

The simplest way to derive (4) is to take the double logarithm of the ODF (1), write down the $M5$ value and subtract it from the formula for $x$. The coefficient $C_i$ may be denoted as the slope coefficient because it is a dimensionless slope of the line (4) in a $x/M5 - y$ plot.

Each station brings about one value of $C_i$. Now (4) contains the two parameters $M5$ and $C_i$. The mean and the standard deviation may be used instead, and $M5$ and $C_i$ calculated in terms of these parameters. Doing so $C_i$ becomes (note that $y_5 = 1.5$)

$$C_i = \frac{1}{C_2} \frac{1}{C'_v + \frac{y_5 - C'_i}{C_2}} = \frac{0.78}{C'_v + 0.72}$$

Where:

$$C'_v: \quad 1/C_v$$

$$C_v = \text{Coefficient of variation}$$

Fig. 4 shows the $C_i$ values for all stations in Washington State, 24-h and one-day data. They are calculated from the station means and standard deviations using the relationship above. Fig. 4 also shows $C'_i$s that were calculated from the $10 \text{ mi}^2$ PMPs for the same meteorological stations in Washington State [US NWS, 1993]. These $C'_i$s when inserted into (4), produce the NWS PMP values when (3) is inserted for the reduced variate $y$ in (4) (squares in Fig. 4). Fig. 4 further shows the $C_i$ curve compiled from the NERC envelope curve, that is, the $C_i$ value one has to use in order to let (3) and (4) reproduce
the NERC PMP values in FSR (1975) (solid line in Fig. 4). The $C_i - M5$ relationship suggested by this curve will be discussed more closely in the next section.

The $C_i$ station values in Fig. (4) seem to cluster around the average value 0.19 with a standard deviation of 0.035. A slight dependence of $M5$ may be detected in the data. This is in accordance with previous findings [Schaefer 1990, Fig. 5a], where $C_v$ values for all Washington State stations are found to depend on the annual precipitation in the lower third of the annual precipitation range, covering 250 – 3000 mm.

![Fig. 4 $C_i$ values for Washington State PMPs, data and NERCS envelope](image)

The distribution of the $C_i$'s around their average is compared with $EV1$ in Fig 5. The fit is only moderate, but $EV1$ will be used to suggest a relationship between the $C_i$ estimates and the data.

While $ylim$ is a regional parameter, both $M5$ and $C_i$ are local parameters playing the same role as the average and the coefficient of variation do in the method suggested in [WMO, 1986] and accepted by many researchers. This method suggests defining geographical regions where the coefficient of variation can be considered constant. Schaefer (1990) tried this, but concluded that the use of geographically delineated regions is difficult, if not unworkable. Here it is suggested that these problems are circumvented by using $M5$ as a local variable with a geographical variation and $C_i$ as a local variable with a random variation.

The distribution 4 may be used as a generalised estimate of precipitation quantiles at any point, provided that $M5$ and $C_i$ are known. They may also be used to calculate the probable maximum value $x_{PM}$ by inserting $ylim$ in (4). This is possible only because the ODF (1) includes $x_{PM}$ and $ylim$. 


Here it is suggested to use the mean value as an estimator for $C_i$ in the generalised distribution. In Washington State the variation of quantile estimates due to variation in $C_i$ will be within 10% when $C_i$ varies within plus or minus one standard deviation even though the return period is as high as 10000 years. For lower return periods the deviation is much less. The accuracy of quantile estimates using this method will therefore depend mainly on the accuracy of the determination of $M5$.

![Graph](WA 1 - 5 Annual Max. Precipitation)

Fig. 5 $C_i$ distribution compared to $E VI$, all stations in WA State

In large watersheds $C_i$ may depend upon area. This will be discussed in the next section.

6. Probable Maximum Precipitation

Probable Maximum Precipitation is defined as [WMO, 1986]:

...the greatest depth of precipitation for a given duration meteorologically possible for a given size storm area at a particular location at a particular time of year, with no allowance made for long-term climatic trends.

In order to understand the importance of this definition for the choice of $C_i$, imagine a watershed with $N_S$ independent meteorological stations. A PMP storm will cause much precipitation in the whole watershed, but a PMP value most likely will occur in only one station. The primary reason for this is that a correlation will exist between precipitation values in neighbouring stations observed in the same storm, and to a lesser degree between annual maxima which usually do not occur in the same storm or on the same day. Therefore, an extreme rainfall depth in one station will usually be associated with large rainfall depths in neighbouring stations, but the annual maximum precipitation
series in different observation stations can still have such small correlation coefficients that they may be considered uncorrelated.

For design purposes the correct PMP in a watershed would have to be the largest PMP value expected at any station or any "observation point" (meaning any place where there might be a precipitation station) in the watershed, with an annual maximum value independent of all the maximum values at all the other "observation points"). Otherwise we would consider the larger PMP estimate found at another "observation point" to be the correct PMP value for the watershed region as a whole.

Under the condition that the $M5$ is the same in all $N_s$ "independent observation points", we now seek a PMP value for the region. In each of the $N_s$ points we can estimate the PMP value statistically by using (4) and inserting the local $C_i$ value and $y_{lim}$. In point no $j$ we will have

$$x_j = M5(1 + (C_i)_j (y_{lim} - 1.5))$$

According to assumption I there is no higher $x$ value possible at point $j$. Therefore this value is the PMP value at this particular location $j$. However, a higher PMP value in the region as a whole may exist. We seek the maximum $x$ in the region remembering that $y_{lim}$ is the same in all $N_s$ points.

$$\text{Max}\{x_j\} = \text{Max}\{M5(1 + (C_i)_j (y_{lim} - 1.5))\} = M5(1 + \text{Max}\{C_i\} (y_{lim} - 1.5))$$

The distribution function of the largest $C_i$ value when all "observation points" are considered, will not be the distribution function in Fig. 5, but that distribution function raised to the power $N_s$. This is the same as dislocating the EV1 in Fig. 5 to the right by subtracting $\log N_s$ from the reduced variate [Leadbetter et al., 1983]. The estimate for the average value of the highest of the $N_s C_i$s, $C_{i,H} = \text{Max}\{C_i\}$, may then be calculated from the average in Fig. 5. If the series of annual maxima in the $N_s$ points are uncorrelated the result will be:

$$C_{i,H} = C_{i,\text{av}} + \log N_s \frac{S(C_i)}{C_2}$$  \hspace{1cm} (5)

$C_{i,H}$: Expected value of the highest $C_i$ in the region

$C_{i,\text{av}}$: Estimated average $C_i$ value for the region

$S(C_i)$: Estimated standard deviation of $C_i$ values in the region

In order to estimate realistic $N_s$ values, we have to know the average distance there must be between meteorological stations so that their annual maxima can be treated as independent. A method to do this has not yet been found, but 30 km may be used as a guideline [Buishand, 1984]. This makes our quantile estimate dependent upon area when we use (5) to calculate $C_{i,H}$ and use this value in (4).
It must be noted, however, that the dependence of the quantile and PMP estimates on area has nothing to do with the area reduction factor. This factor takes into account that the precipitation in the watershed region is everywhere less than the PMP value. When $C_{i,H}$ is used in estimation, together with (4), this will correspond to 10 mi², and area reduction factors are to be applied in any watershed region that is bigger than 10 mi² in order to find the average precipitation in the particular PMP storm.

As research has not yet fully uncovered the effects of cross-correlation in precipitation data, (5) gives only an indication of how to estimate the $C_i$ values. But given a method to select an appropriate $C_i$ value, (4) can be used as a generalised distribution for precipitation with a return period of two years or more. To do so, one needs a map of $M_5$ for the area in question. $M_5$ maps have been produced for England [FSR, 1975] and Norway [Förland and Kristoffersen, 1988], and one is presently being prepared for Iceland.

In accordance with the foregoing statements, PMP values may be calculated by selecting $C_{i,H} = C_{i,PMP}$. This makes it possible to use Equation (5) to calculate the $N_s$ values one has to apply to produce the NERC envelope in Fig. 4. The results are in Table 2. The table shows the estimated variation of the mean and standard deviation of $C_i$ with $M_5$. There are two sets of $C_i$ estimates in Table 2. In column 2 are the Icelandic estimates for the average $C_i$, and in column 4 the $N_s$ values found from the values in column 2 by using (5). In column 6 are the comparable $N_s$ estimates based on Washington State data.

The calculated $N_s$ values increase with decreasing $M_5$, showing greater dependency (smaller number of possible "independent" meteorological stations) in locations with heavy precipitation. Most of the data points with small $M_5$ values in Fig. 4 come from Eastern Washington where frontal rain is infrequent. In Dutch data a high ratio of convective rain in the annual maximum precipitation values means faster convergence of the asymptotic independence of the large values [Buishand, 1984] mentioned earlier. This may very well mean that the average distance between meteorological stations necessary for annual maxima, that can be treated as independent, is smaller in areas with a high proportion of convective precipitation in the annual maxima, just as seems to be the case in Fig. 4.

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5</td>
<td>$C_{i,IS}$</td>
<td>$C_i$</td>
<td>St.dev</td>
<td>$N_{s,IS}$</td>
<td>$C_{p,WA}$</td>
</tr>
<tr>
<td>mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.34</td>
<td>0.043</td>
<td>23270</td>
<td>0.21</td>
<td>1028716</td>
</tr>
<tr>
<td>35</td>
<td>0.27</td>
<td>0.037</td>
<td>21511</td>
<td>0.21</td>
<td>182045</td>
</tr>
<tr>
<td>50</td>
<td>0.22</td>
<td>0.035</td>
<td>13697</td>
<td>0.20</td>
<td>28496</td>
</tr>
<tr>
<td>100</td>
<td>0.16</td>
<td>0.034</td>
<td>674</td>
<td>0.20</td>
<td>149</td>
</tr>
<tr>
<td>150</td>
<td>0.14</td>
<td>0.032</td>
<td>147</td>
<td>0.20</td>
<td>13</td>
</tr>
</tbody>
</table>
PMP values may be estimated with very large areas in mind. But compared with the 30 km guideline referred to in the previous section, the calculated $N_s$ values in Table 2 are clearly too high, except for large $M5$ values. This can be so for a number of reasons. First, the PMP values are likely to be "on the safe side", especially for smaller precipitation values (low $M5$). Second, the scatter of the $C_i$ values is likely to increase if data from other climatic regions or periods are brought into the analysis. Third, the possible effects of the correlation have not been included. The last two reasons tend to increase the estimate of the standard deviation.

The influence of the high $N_s$ numbers in Table 2 on the PMP estimate is less than the large difference in the high and low numbers indicates. Table 3 illustrates this point. The numbers are based on a 9000 km$^2$ area. The guideline length of $L\ = \ 30$ km between independent stations is adopted for 150 mm $M5$ precipitation and tapered down to 5 km for 25 mm $M5$ values. The PMPs are estimated from (3) and (4), and the $C_i$ curve in Fig 4 is used for a NERC estimate shown in Table 3. When the $C_{i,WAS}$ in Table 2 are used in (5) to calculate $C_{i,PMP}$, lower PMP estimates are found, and the ratio values in Table 3 are the ratios of these two estimates.

<table>
<thead>
<tr>
<th>$M5$ mm</th>
<th>25</th>
<th>35</th>
<th>50</th>
<th>100</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>PMP,N</td>
<td>175</td>
<td>211</td>
<td>258</td>
<td>376</td>
<td>465</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.67</td>
<td>0.70</td>
<td>0.74</td>
<td>0.87</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Compared with actual PMP estimates for Washington State, the PMP values for $M5$ 100 and 150 mm are very close to the NERC value that also indicates the average of the US NWS values. The values for 25 and 50 mm are lower, close to 70% of NERC. As mentioned earlier, the reason for this is not quite clear. Nevertheless, it may be considered demonstrated that the average $C_i$ values cannot be used for $C_{i,PMP}$. Instead higher values should be used. This explains the difference of the $C_{i,PMP}$s compiled from NERC and US NWS, on one hand, and the $C_i$ data from individual meteorological stations, on the other.

7. Conclusion

A PMP limit value $y_{lim}$ can be introduced into the EV1 distribution by a simple transformation so that this popular and widely used distribution function can be used for precipitation that has a PMP value as an upper limit. Precipitation data from Iceland and Washington State fit very well to this distribution when standardised to a zero mean and standard deviation of one for each meteorological station. A reliable estimate of the upper limit is not obtained from the data, but the presence of the upper limit has very little
effect on the estimate of the first and second moments, so the distribution function can be used without knowing its precise value. The distribution function may be used as a generalised distribution function for precipitation, with the five-year event $M5$ as a geographically varying local parameter, and a slope coefficient called $C_i$, that is a function of the coefficient of variation, as a randomly varying local parameter.

The $C_i$ values calculated from the Washington State data have a slight variation with $M5$. Disregarding this slight variation, the $C_i$s have a distribution around the mean that moderately fits the $EV1$; average $C_i$ is found to be close to 0.19; and standard deviation of $C_i$ 0.035. Icelandic results produce almost the same values. This shows that in a watershed where many independent meteorological stations are possible, the precipitation estimates have to be based on the largest $C_i$, called the $C_{i,H}$.

The PMP values may be calculated from the generalised distribution function and by letting $C_{i,PMP} = C_{i,H}$. A limiting reduced variate deduced from NERC’s PMP envelope, makes it possible to derive $C_i$ values from the US National Weather Service PMP values for the Washington State meteorological stations, and the values compare favourably with the $C_i$ values calculated from British and Norwegian PMPs, showing that NERC's statistical estimate is consistent with the NWS method.

Using a guideline length of $L = 5 – 30$ km between independent stations, in the $M5$ range of $25 – 150$ mm, the distribution of the $C_{i,H}$ may be calculated by dislocating the distribution of the $C_i$ data by $\log N_s$ to the right. PMP estimates based directly on the Washington State data, using this $C_{i,H}$ and the NERC limiting reduced variate, are somewhat lower than the NERC values at the lower end of the $M5$ range, but approximately the same at the higher end.

The use of (4) and (5) as a generalised distribution for one-day and 24-h annual maximum precipitation is thus suggested. This distribution function can be used for quantile estimates and PMP estimates by the use of an $M5$ map, and it can be used for PMP estimates using $C_{i,PMP} = C_{i,H}$. This statement assumes that the conclusions above are not contradicted by an analysis of local annual maximum data. Such a contradiction is a possibility, as the statistical analyses are based on assumptions I–III, and their validity can only be proved $a posteriori$. 

8. Acknowledgements

This study was accomplished while on sabbatical at the University of Washington, Department of Civil Engineering. The research program is conducted by the Engineering Research Institute of the University of Iceland and is supported by the National Power Company of Iceland. The author is indebted to Melvin G. Schaefer Ph.D. P.E, Department of Ecology, Dam Safety Section, for providing the Washington State data used in the analysis. His invaluable help is greatly appreciated.

9. References

Buishand, T. A. ; Bivariate Extreme Value Data and the Station-Year Method; J. Hydrol., 69; 77 - 95; 1984

DaoJiang, Zhan and JinShang, Zhaou; Recent Developments on the Probable Maximum Precipitation (PMP) Estimation in China; J. Hydrol., 68; 285 - 293; 1984

Elíasson, Jónas ; Probable Maximum Precipitation in Iceland, - Station Values -; Nordic Hydrology, Vol. 23, No. 1; 49-56; 1991

Elíasson, Jónas ; Statistical Estimation of PMP values; Nordic Hydrology, Vol. 25, No. 4; 1994

Flood Studies Report (FSR), Vol. II; Natural Environment Resource Council (NERC); 1975

Förland, E. and Kristoffersen, D. ; Paaregnelig maksimal nedbør beregnet med ulike metoder.; Fagrapport nr. 9/88 KLIMA. Det Norske Meteorologiske Institutt, Oslo (In Norwegian) ; ; 1988


Hansen, Marshall E.; Probable Maximum Precipitation for Design Floods in the United States; J. Hyrol., 97; 267 - 278; 1987


Leadbetter, M.R., Lindgren, G. and Rootzén, H. ; Extremes and Related Properties of Random Sequences and Processes; Springer Verlag; ; 1983

Rakhecha, P. R. and Kennedy, M. R. ; A Generalised Technique for the Estimation of Probable Maximum Precipitation in India; J. Hyrol., 78; 345 - 359; 1985


