TECHNICAL PAPER

JOURNAL OF THE SOUTH AFRICAN INSTITUTION OF CIVIL ENGINEERING

Vol 52 No 2, October 2010, Pages 48–57, Paper 730

Estimation of extreme flood peaks by selective statistical analyses of relevant flood peak data within similar hydrological regions

J H Nortje

JAN NORTJE (Pr Eng, Member SAICE) graduated in 1973 and obtained his Master's degree in 1993, both in civil engineering and at the University of Pretoria. He joined the Department of Water Affairs in 1972 and has since specialised in dam engineering covering various facets, including design, construction and planning of dams. Since 1987, when he joined the Dam Safety Office, he has been working in the dam safety engineering field. Determination of appropriate "safety evaluation floods" for dams is one of the major challenges in dam engineering and this is what has inspired the current paper.

Contact details:
Department of Water Affairs
Private Bag X313
Pretoria
0001
South Africa
Tel: 27 12 336 8010
Fax: 27 12 336 8674
e-Mail: nortjej@dwa.gov.za

This paper describes a new Regional Estimation of Extreme Flood Peaks by Selective Statistical Analyses (REFSSA) method to estimate extreme flood peaks from regional flood peak data. The method differs from current regional flood frequency analysis (RFFA) methods or approaches in that an additional separate statistical analysis is performed on "record maximum flood peaks" within a "similar hydrological region". Suitability of the method is demonstrated for the estimation of extreme flood peaks with annual exceedance probabilities between 0,001 (1/1 000) and 0,0001 (1/10 000) for two major hydrological regions in South Africa, and for catchment sizes between 100 and 7 000 km^2. The applicability of the method for catchments outside these regions and limits has not been fully tested mainly due to a shortage of verified data. The theory and a practical example are presented. Excellent results have been obtained so far, displaying high correlation coefficients between extreme flood peak data and regression lines, namely 0,99 on average on log-normal scale. The method is considered to have universal application, especially in climates experiencing outlier type of extreme flood peaks.

INTRODUCTION AND BACKGROUND

This paper describes a new approach termed the Regional Estimation of Extreme Flood Peaks by Selective Statistical Analyses (REFSSA) method, to estimate values for extreme flood peaks. The approach differs from current regional flood frequency analysis (RFFA) methods in the following respect: selective (and separate) statistical analyses are carried out on regional flood peak data after transformation (in proportion to the square roots of the respective catchment areas) from comparable sites within a "similar hydrological region" to the site under investigation. A distinction is specifically made between information contained within the whole spectrum of annual maximum flows (one value per site per year, thus including many low flows in the South African climate) and information contained within the "record maximum flood peaks" (only one record value per site for the full observation period), which better reflects the characteristics of extreme flood peaks.

For the purpose of this paper a "similar hydrological region" is provisionally defined as a space/area of demonstrated similarity with regard to the past occurrences of "record maximum flood peaks", thus not requiring homogeneity with regard to aspects such as catchment characteristics. This wide definition is considered admissible for the initial purpose of this study, namely to determine upper-bound values for extreme flood peaks within "similar hydrological regions". However, the REFSSA method is versatile and homogeneous regions, or alternatively clusters of similar basins as described by Wiltshire (1986), could be used in the place of "similar hydrological regions".

The REFSSA method is particularly suitable for hydrological environments where a flood record typically includes one or two extreme flood peak outliers (as in South Africa) and where record lengths are rather short. The initial focus of this paper is the estimation of upper bound values for extreme flood peaks between $Q_{1\,000}$ and $Q_{10\,000}$, where Q_T is defined as the flood peak value with an annual exceedance probability (AEP) of 1/T ($Q_{10\,000}$, for example, has an AEP of 1/10 000).

Traditionally, the Single Station Statistical Analysis (SSSA) method has been used for the estimation of extreme flood peak values even up to $Q_{10\,000}$. The SSSA method is certainly useful for estimating flood peaks within or close to its record length but, as motivated by Kovács (1988), a flow record should not be extrapolated to more than two times its length. Alexander (2000) also warned against extrapolation beyond Q_{100}. The limitation of the SSSA method is demonstrated by the fact that different but equal-

Keywords: hydrology, extreme flood peak estimation, regional flood frequency analysis, regionalisation, regression

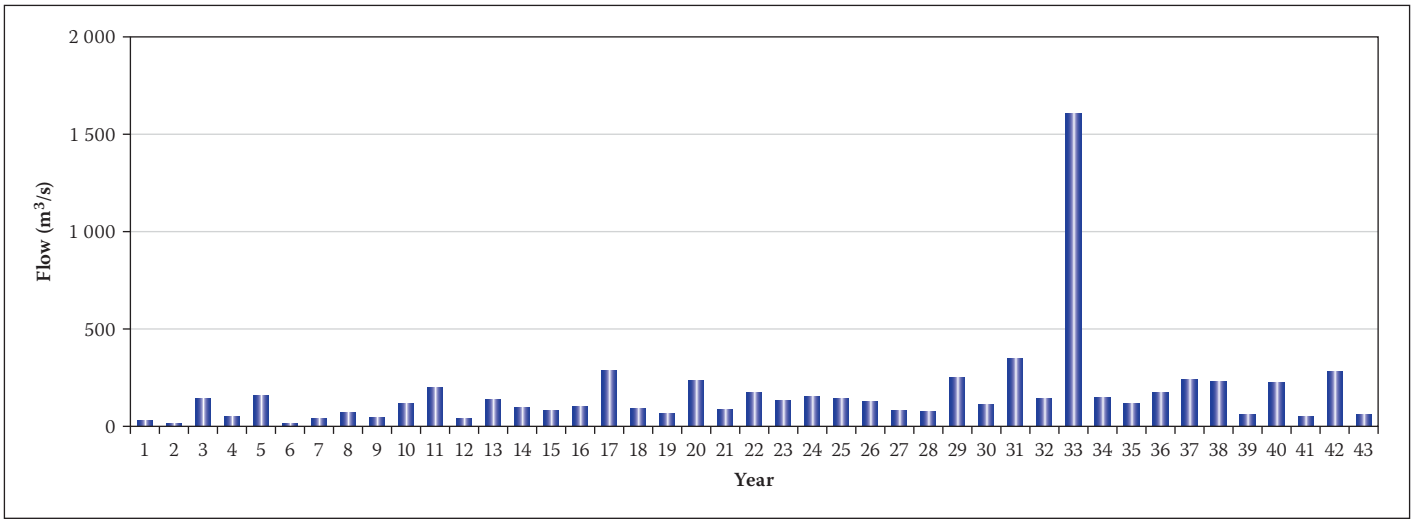

Figure 1 Annual maximum inflows at Midmar Dam

fitting distributions (all with correlation coefficients in the order of 0,97), such as Log-normal, Log Pearson III and General Extreme Value (with either conventional moments or probability-weighted moment estimators), yield entirely different estimates of $Q_{10\,000}$ which could range from 2 000 to 13 000 m^3/s for one site. Lesser annual maximum flows, such as Q_1, Q_2, Q_3, ... Q_{30}, play a major role in the SSSA method, but they do not necessarily contain information on the magnitude of extreme flood peaks. This is demonstrated by the flow record at Midmar Dam shown in Figure 1. In this case only one extreme flood peak occurred during a continuous 43-year record length. This is a typical picture for many inland sites in South Africa.

Integration of regional information helps to overcome the lack of long-term records at individual sites and this is addressed by RFFA methods in many parts of the world. Cunnane (1988) provides an evaluation of the merits of different RFFA methods. Current RFFA methods, however, do not fully overcome the statistical shortcomings of the SSSA method for estimating extreme flood peaks in an environment with outlier type of extreme flood peaks. The main reason is that the different statistical characteristics of extreme flood peaks are not taken into account. Also, in its purest form, the RFFA method requires a homogeneous region or pool (cluster of similar basins) and a common record period covered by all stations within the region or pool. As a result, the database is reduced in size and it is still dominated by low annual maximum flows, especially in the South African context. In addition, historical flood peaks cannot be added without making questionable statistical assumptions. Görgens (2007a) adapted the RFFA approach (using the flood index method) for South Africa, but it is unfortunately only considered suitable for the estimation of flood peaks and flood hydrographs up to about Q_{100}.

Compared with the SSSA method, the REFSSA method achieves superior correlation coefficients between extreme flood peaks and regression lines. Another significant improvement is with respect to the coefficient of variation, c_v (standard deviation divided by the mean in terms of logarithms of flood peak data), which typically reduces from about 0,2 (SSSA method) to about 0,06 (REFSSA method). Typically, a record length of 50 to 100 years is analysed by the SSSA method in South Africa, whereas the typical representative record length analysed by the REFSSA method is between 3 000 and 5 000 station years. In addition, historical (including palaeo) flood peaks could be added to the catalogue of "record maximum flood peaks" without the statistical difficulties experienced by current RFFA methods. These comparisons suggest that more reliable estimates of extreme flood peaks could be obtained by means of the REFSSA method.

In South Africa, the Regional Maximum Flood (RMF) method as developed by Kovács (1988) in accordance with the Francou-Rodier (1967) approach is a frequently used empirical method to determine appropriate safety evaluation flood peaks for dams. The RMF value is the value on an envelope curve drawn just outside "record maximum flood peaks" for different sites within a specific demarcated region. The RMF method thus also integrates regional information within "similar hydrological regions", but a serious shortcoming is that the annual exceedance probability (AEP) of the RMF value at a specific site is unknown. In addition, the AEP of the RMF is not constant but varies significantly from site to site and region to region.

Application of the Probable Maximum Flood (PMF) approach in South Africa is also problematic. Firstly, the AEP of the PMF is undefined. Secondly, PMF values as derived by the preferred unit graph method of HRU (1972) have poor correlation with actual record maximum flood peak data. This was demonstrated by Görgens et al (2007b) who found that PMF/RMF ratios vary from 0,6 to 9 from site to site and region to region in South Africa.

The new REFSSA method provides a sound statistical basis for estimating extreme flood peak values between $Q_{1\,000}$ and $Q_{10\,000}$ from regional data. Reliability depends mainly on the availability and accuracy of relevant record maximum flood peak data from comparable catchments. By means of the REFSSA method, estimates can also be made of the AEP of RMF and PMF values. This is considered very useful because the SANCOLD Guidelines (SANCOLD 1991), which are relevant when determining appropriate safety evaluation discharges or floods for dams in South Africa, are based largely on the RMF and PMF methods.

THEORETICAL PREMISE OF THE REFSSA METHOD

The REFSSA method distinguishes between the following two data sets: (1) the Q_x data set which consists of transformed "record maximum flood peaks" (Q_{xi}-values) and (2) the Q_a data set which consists of all transformed annual maximum flows (Q_{ai}-values). The aforementioned data are selected from sites with comparable catchments and catchment sizes within a "similar hydrological region", and then transformed to the site under investigation in proportion to the ratio of the square roots of their respective catchment areas. It is inherently assumed that "storm event" is the major factor with regard to the magnitude of extreme flood peaks. It is expected that the REFSSA method would be less reliable for catchment sizes of less than about 100 km^2, where catchment characteristics could become more important.

The crux of the method is the postulation that a variable termed the "regionally observed maximum flood peak" (Q_x) for a specific site (based on the distribution of transformed "record maximum flood peaks" or Q_{xi}-values, obtained from other sites with comparable catchments within a "similar hydrological region" during the same observation period of adequate length) can be regarded as a *statistical variable* and further that its statistical distribution parameters can be utilised to estimate the magnitude of extreme flood peaks, such as $Q_{1\,000}$ to $Q_{10\,000}$. It is postulated that information contained within record maximum flood peak data within a "similar hydrological region" is much more suitable for estimating the *magnitude* of extreme flood peaks than information contained within lesser annual maximum flow data, such as Q_1, Q_2, Q_3, … Q_{30}, etc. Information from the latter or Q_a data set is utilised to help "calibrate" the AEP of extreme flood peaks. An algorithm that combines the information from the two data sets in order to estimate both the magnitude and the AEP of extreme flood peaks is presented in the next section. The expected value of Q_x for a site during a similar observation period can be calculated as the mean (or median if the data are log-normally distributed) of the Q_{xi}-values. The other parameters of the distribution of Q_x such as the standard deviation, coefficient of variation and skewness, can also be calculated from the Q_{xi}-values. Estimates can then be made of extreme flood peak values such as $Q_{10\,000}$ by using a suitable theoretical statistical distribution model.

The following notes have further bearing on the theoretical premise of the postulation:

- Q_T is defined as the flood peak value with an annual exceedance probability (AEP) of 1/T. T is traditionally referred to as the recurrence interval or return period in years, which is strictly speaking inappropriate because hydrological records are not statistically independent with respect to time. Records at many single stations show definite cyclic patterns over time. Alexander (2009) demonstrated that annual inflow volumes at Vaal Dam display low-high 'cycles' with periods of approximately 20 years. Long-term climate changes are also not reflected within short observation or record periods. The AEP of Q_T should thus be expressed as 1/T (e.g. 0,0001 or 1/10 000 – omitting years), with the qualification that it is based on a statistical analysis of data collected over a specified observation period.
- If there were 50 independent Q_{xi}-values available from 50 different sites with equal catchment areas within a "similar hydrological region", then each site will have a different record maximum flood peak value mainly because of differences in storm events, which do not occur uniformly over a whole region. This phenomenon will be reflected in the variance of Q_x. It is assumed that independent storm events could occur on a relatively random basis anywhere within a "similar hydrological region", especially if the region is large relative to the catchment size of the site under investigation.
- If each of the above sites had a record length of 100 years, then the artificially combined record of the 50 separate sites could be put at 5 000 "station years". As motivated above, this is not equivalent to a single station with a 5 000 year record because hydrological data even at single stations are not statistically independent. From a statistical perspective it could be expected that 5 000 station years or 50 data points of "record maximum flood peaks" could include flood peaks as low as Q_{30} and as high as $Q_{1\,000}$ or more. This expected wide spectrum of "record maximum flood peaks" is confirmed by Figures 9a to 9h of Kovács (1988) that show wide bands of the record maximum flood peak values within "similar hydrological regions".
- The mean of say 500 Q_{xi}-values from 500 sites would provide a reasonably good estimate of the population mean μ_{Qx} at a site within a "similar hydrological region". Similarly the standard deviation of the 500 Q_{xi}-values would provide a reasonably good estimate of the standard deviation σ_{Qx} of the population. However, if the sample size is small, statistical uncertainty would be high and the estimates of μ_{Qx} and σ_{Qx} would lack in accuracy. This will have an influence on the reliability of the estimation of extreme flood peak values. It has been found that sample sizes (number of Q_{xi}-values) should not be less than about 25 to 30 in order to obtain stable results.

Table 1 Requirements for selection of data for both the Q_x and Q_a data sets

1	Preferably the same stations should be used for both the Q_x and Q_a data sets to enhance consistency (or to reduce bias).*
2	The observation periods of all stations from which the data points are obtained should preferably be the same for consistency, and also of adequate length (say not shorter than 100 years) so that there is a reasonable chance that the data would contain extreme flood events.*
3	Data should be from a "similar hydrological region", which refers to similarity with regard to the main factor, namely the impact of "storm events" as reflected by "record maximum flood peaks".
4	Catchments of selected stations should not overlap and should preferably also not be adjacent or close to each other in order to promote statistical independency.
5	Catchment areas should be of comparable size, say between one half and two times the catchment area of the site under investigation. This would help to reduce errors when transforming the data to the site under investigation (in proportion to the square roots of the relevant catchment areas).
6	Data should be from comparable catchments. Data from catchments with significantly different catchment characteristics, for example those under intensive plantations, catchments with substantial dolomite deposits (or rock with karst features) or catchments with very flat topography including natural lakes should not be included if these catchments are not comparable to the site under investigation.
7	The data should reflect natural flood peaks. The effect of attenuation by upstream dams or the impact of dam break floods should be removed from the data as far as possible.
*	Some of the above requirements (marked by an asterisk) are rather idealistic and could not be fully complied with in the preliminary investigations as reported in this paper.

Table 2 Additional requirements for selection of data for the Q_x data set

1	The data must reflect extreme flood events. Flood peak data which are too low and that do not represent extreme flood events should be discarded.
2	Only one "record maximum flood peak" value (the highest) per station/site must be selected so that the data would be statistically independent from a spatial viewpoint.
3	It is important that the highest one, and preferably the top three selected Q_{xi} values within a selected Q_x data set, must be fairly accurate. This would enhance the accuracy of the REFSSA method.
4	Stations must be selected so that their records would be as statistically independent as far as possible. The same storm event should not be counted more than once. This can be achieved relatively easily for Q_x data because the dates of the record maximum flood peaks are normally known. Due to the way that the algorithm of the REFSSA method is constructed, this requirement is regarded less important for Q_a data.

Limitations and practical considerations impacting on accuracy of results

- The REFSSA method is provisionally considered suitable for the estimation of

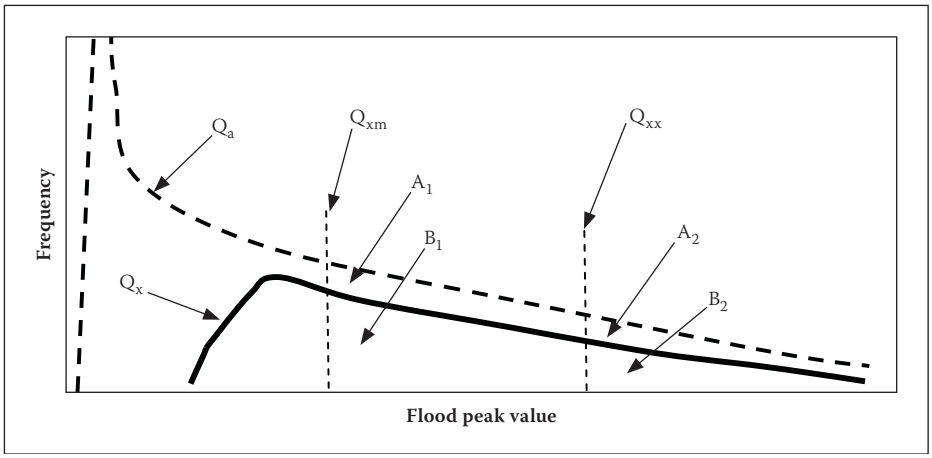

Figure 2 Diagrammatic presentation of the distributions of Q_a and Q_x

extreme flood peaks between $Q_{1\,000}$ and $Q_{10\,000}$.

- The estimation of extreme flood peaks for a specific site within a "similar hydrological region" but with catchment characteristics significantly different from the 'average' would be less accurate.
- The reliability of the method depends on the quantity and quality of the source data.
- The estimates for extreme flood peaks are valid for current climatic conditions (as reflected by the source data).
- At some dams flood attenuation may play a significant role. In such cases extreme flood hydrographs should be constructed by using other methods, e.g. as proposed by Görgens (2007a), but these should be realistic compared to actual recorded extreme flood hydrographs at similar sites within the "similar hydrological region".
- The demarcation of "similar hydrological regions" could be somewhat subjective due to limited available record maximum flood peak data and this could lead to inaccuracies.
- Catchments bigger than 7 000 km^2 typically cover two or more "similar hydrological regions" and the upper limit for the REFSSA method is provisionally set at 7 000 km^2.
- It is expected that catchment characteristics would play a bigger role in the case of smaller catchments and the lower limit is provisionally set at 100 km^2.

SELECTION OF DATA

Requirements for selection of data for the Q_x and Q_a data sets

From a statistical viewpoint the data should be unbiased, statistically independent and relevant to the site for which an extreme flood peak is to be estimated. These criteria form the basis of the selection requirements as listed in Tables 1 and 2. Additional requirements for the selection of Q_x data are listed in Table 2. Record maximum flood peak data selected according to the criteria as listed in Tables 1 and 2 should give a good indication of the mean and variance of extreme flood peaks.

Catalogue of "record maximum flood peaks" by Kovács (1988)

In South Africa record maximum flood peak data are readily available from the catalogue published by Kovács (1988). This catalogue was used as the main data source in this study. RMF-regions as demarcated by Kovács (1988) comply with the definition of and have been used as "similar hydrological regions" as a starting point for the purpose of this study. The following aspects relate to the suitability of the catalogue by Kovács (1988) as a data source for the REFSSA method (it should be borne in mind that this catalogue was not specifically compiled for the REFSSA method):

- Record maximum flood peak data for catchments below 100 km^2 and above 7 000 km^2 are scarce. The sample size for an analysis should preferably be more than 30 in order to reduce statistical uncertainty to acceptable levels.
- In the catalogue an indication of the accuracy of individual data points is given. Accuracy varies considerably. The accuracy of many flood peaks is indicated as "unknown".
- Some data could have been influenced by flood attenuation by upstream dams and the data should be corrected to reflect natural un-attenuated flood peaks where and if applicable.
- The data selected for an analysis should be as statistically independent as possible, especially with regard to the most important factor namely storm event. Fortunately the catalogue includes most of the dates of "record maximum flood peaks".
- Ideally, all data should cover the same observation period of, say, at least 100 years to improve consistency (or to reduce bias). The observation period covered by the catalogue varies over the country and is generally rather short. The data might thus not include an adequate number of extreme flood peaks.
- Regional boundaries should be refined as more data become available. The increments between some regions appear to be too large.

It can be seen that the catalogue as published by Kovács (1988) has a number of shortcomings for use as a database by the REFSSA method. Nevertheless, at the time of its publication (1988) a lot of work was done to make the catalogue as accurate and complete as possible and it contains a wealth of information. It is the only verified database of its kind that is readily available in South Africa. Taking all factors into account, it is regarded as suitable for current use until a more complete database becomes available.

ALGORITHM FOR ESTIMATING EXTREME FLOOD PEAKS

The algorithm for estimating the magnitude and AEP of extreme flood peaks is described below on the basis of the diagrammatic presentation in Figure 2.

The symbols used in Figure 2 are defined below. All Q values in Figure 2 refer to flood peak values *after transformation* to a specific catchment size. The data reflect those of a selected number of stations within a "similar hydrological region" during the same observation period.

Q_x Regionally observed maximum flood peak, measured as the "record maximum flood peaks" from many sites within a "similar hydrological region" (one Q_{xi}-value per site for the full observation period). Note that the Q_x data set is a subset of the Q_a data set.

Q_a Regional annual maximum flow, measured as "annual maximum flows" from the same sites as above (one Q_{ai}-value per site per year, e.g. 5 000 values for 50 sites during an observation period of 100 years, thus including 50 Q_{xi}-values)

Q_{xm} Median of all Q_{xi}-values

Q_{xx} Extreme flood peak value that must be determined for a site (e.g. $Q_{10\,000}$)

In a very large sample the AEP of Q_{xx} at a selected site (defined as α_2 in Equation (2) below) would be approximately equal to the number of Q_{ai}-values that exceed Q_{xx} divided by the total number of all Q_{ai} values,

which constitute the total outcome or sample space of Q_a.

In Figure 2 the areas A_1 and A_2 below the Q_a curve represent the number of Q_{ai} data points exceeding Q_{xm} and Q_{xx} respectively (for a "continuous" probability density function Q_a this can be visualised by selecting one flood peak unit to be equal to one class interval). Similarly, the areas B_1 and B_2 below the Q_x curve represent the number of Q_{xi} data points exceeding Q_{xm} and Q_{xx} respectively. If the total area below the Q_a curve is A and the total area below the Q_x curve is B, then the cumulative probabilities of selected events can be expressed as follows:

$$\alpha_1 = P(Q_a > Q_{xm}) \approx A_1/A \quad (1)$$

= probability of $Q_a > Q_{xm}$

$$\alpha_2 = P(Q_a > Q_{xx}) \approx A_2/A \quad (2)$$

Note that α_2 is the selected AEP for which Q_{xx} must be estimated.

$$\beta_1 = P(Q_x > Q_{xm}) \approx B_1/B \quad (3)$$

$$\beta_2 = P(Q_x > Q_{xx}) \approx B_2/B \quad (4)$$

It should be noted that α_1, α_2, A_1, A_2 and A are within Q_a sample space and β_1, β_2, B_1, B_2 and B are within Q_x sample space.

It is reasonable to assume that the Q_a curve and the Q_x curve would coincide to the right of the Q_{xx} value, because the Q_{xx} value is a large and extreme value by definition. Thus:

$$A_2 \approx B_2 \quad (5)$$

Define f as the factor required to reduce A_1 so that the Q_a and Q_x curves will approximately coincide to the right of the Q_{xm} value. Thus:

$$f = B_1/A_1 \quad (6)$$

or

$$fA_1 = B_1 \quad (7)$$

From Figure 2 it is clear that $f \leq 1{,}0$. The Q_{xi} values represent record maximum flood peak values. Only one Q_{xi} value is selected per station for the full observation period. It is therefore possible that there might be other Q_{ai} values that are also larger than Q_{xm} but that are not the largest for a single station and thus do not qualify as Q_{xi} values. That is why the Q_x curve is shown below the Q_a curve in Figure 2.

The equality of the proportions below follows from Figure 2 or Equations (5) to (7):

$$A_2/fA_1 \approx B_2/B_1 \quad (8)$$

Table 3 Example (estimation of $Q_{10\,000}$ for Albasini Dam)

1	**Select data from a "similar hydrological region":** RMF-regions as delineated by Kovács (1988) have been used. Albasini Dam falls in an isolated patch of the 5,2-region. Unfortunately there is not enough data in region 5,2 alone. The site is surrounded by the 5,0-region on the northern, western and eastern sides. The southern boundary is very far away. Therefore data from regions 5,0 and 5,2 in South Africa have been selected for this analysis.
2	**Select "record maximum flood peaks"** (Q'_{xi}) from stations with catchment areas between 0,5A and 2A, where A is the catchment area for Albasini Dam. The data selection requirements as stated in Tables 1 and 2 have been followed as far as possible. A total number of 42 data points were selected on this basis from the catalogue published by Kovács (1988).
3	**Transform the selected "record maximum flood peak" data** to the catchment size of Albasini Dam by multiplying the data by the ratio of the square roots of the respective catchment sizes, that is $Q_{xi} = Q'_{xi}(A_{Albasini}/A'i)^{0,5}$. $A_{Albasini} = 509\ km^2$.
4	**Estimate the main statistical parameters of the variable Q_x** namely the mean, median, standard deviation (SD or S_{Qx}), coefficient of variation (c_v) and skew (g) from the transformed Q_{xi} data in the selected sample. Repeat for log Q_{xi} data. Calculations have been done by means of a spread sheet and the results of the above steps are given in Table 4. The significant reduction of the skew value of the *log* Q_{xi} data compared to Q_{xi} data is an indication that the data is log-normally distributed.
5	**Refine:** Inspect the data and results. Two low points (too low to represent extreme flood peaks) were discarded. This reduced skew even further. This action was necessary because the catalogue was not specifically compiled for the REFSSA method. Repeat step 4 after finalisation of the data selection.
6	**Determine the annual exceedance probability (AEP) of the median** (915 m^3/s) of the transformed Q_{xi} data on a regional basis. A complete set of verified annual maximum flows within the selected hydrological region was not readily available for an accurate determination of the AEP of the median by using equation (1). It was therefore estimated by using the regional method as published by Kovács (1988), but amended as follows: The RMF value is 2 879 m^3/s (for region 5,2). According to his fig 14 the Q_T/RMF factor for Q_{50} is approximately 0,426 giving a Q_{50} value of 1 226 m^3/s. Görgens (2003) proposed reduction factors (different factors for different Q_T's and regions) to be applied to Kovács' factors. In this case the reduction factor is 0,68. This provides an adjusted, regionally derived estimate for Q_{50} of 834 m^3/s. In similar fashion Q_{100} is estimated at 1 206 m^3/s. By interpolation on log-normal-probability scale the AEP of the median is estimated to be about 1/59. It should be noted that the reduction factors proposed by Görgens (2003) reduce the Kovacs' factors closer to average regional factors, whereas Kovács (1988) intended the factors to be conservative upper bound factors. Average factors should be used in the REFSSA method according to equation (1). It was found that results are not very sensitive to the value of the AEP of the median. The effect on the value of $Q_{10\,000}$ in this case when using an AEP of 1/40 or 1/80 in the place of 1/59 is less than 5%.
7	**Plot the ranked *log* Q_{xi} data** on log-normal probability scale and inspect. Cunnane plotting positions have been used. The graph is included as Figure 3. It can be seen that the log-normal model produces an excellent fit. The correlation coefficient is better than 0,99.
8	**Determine the extreme flood peak value** for the selected AEP of 1/10 000 as follows: $\alpha_2 = P(Q_a > Q_{xx}) = 1/10\,000$ (selected AEP for Q_{xx}) $\alpha_1 = P(Q_a > Q_{xm}) \approx 1/59$ (AEP of median Q_{xm} - from item 6) $f \approx 1{,}0$ (from inspection of many Q_{ai} records at inland sites) $\beta_2 = \alpha_2 / (2f\alpha_1) = P(Q_x > Q_{xx}) \approx 0{,}00295$ $Log\ Q_{xx} = log\ Q_{xm} + S_{log\ Qx} Z_{\beta 2}$ (assuming a log-normal distribution) , where $Z_{\beta 2} = 2{,}7533$ (standardised normal variate obtained from normal distribution tables corresponding with β_2) The values of *log* Q_{xm} (mean of *log* Q_x) and $S_{log\ Qx}$ (standard deviation of *log* Q_x) are obtained from Table 4: *log* $Q_{xx} \approx 2{,}9614 + 0{,}1865 * 2{,}7533 = 3{,}4749$
9	**From this $Q_{10\,000}$ is calculated as 2 985 m^3/s which can be rounded off to 3 000 m^3/s.** The AEP of the RMF (2 879 m^3/s) is estimated at about 1/7 760 by reversing the above calculation steps. Results of the aforementioned calculations are summarised in Table 5**.**

Substitute Equations (1) to (4) into Equation (8):

$$\alpha_2/(f\alpha_1) \approx \beta_2/\beta_1 \quad (9)$$

or

$$\beta_2 \approx \alpha_2\,\beta_1 / (f\alpha_1) \quad (10)$$

From the definition of the median:

$$\beta_1 = 0{,}5 \quad (11)$$

thus

$$\beta_2 \approx \alpha_2/(2f\alpha_1) \quad (12)$$

Equation (12) for β_2 can also be obtained by using the theory of conditional probability (the above deduction is a simplified and illustrative version thereof). Equation (12) provides the necessary conversion to obtain the probability β_2 in Q_x space (Q_{xi} data set) so that the value of Q_{xx} can be determined by using the known (calculated) distribution characteristics of the variable Q_x.

If it is assumed that Q_x is log-normally distributed, then from the characteristics of the log-normal distribution:

$$log\ Q_{xx} = log\ Q_{xm} + S_{log\ Qx} Z_{\beta 2} \quad (13)$$

where

$Z_{\beta 2}$ = standardised normal variate obtainable from normal distribution tables corresponding to β_2

Table 4 Selected record maximum flood peaks from regions 5,0 and 5,2 for Albasini Dam

Station	Region	River	A' (km²)	Q'xi	Qxi*	log (Qx)	m
T4M01°	5,2	Mtamvuna	715	2 270	1 915	3,2822	1
Q3°	5	Pauls	873	2 500	1 909	3,2808	2
G1M07°	5	Berg	713	2 130	1 800	3,2552	3
D2°	5	Rietspruit	752	2 160	1 777	3,2497	4
C5R03	5	Modder	940	2 300	1 692	3,2285	5
B7M08°	5,2	Selati	832	2 150	1 682	3,2257	6
A6°	5	Sterk	707	1 880	1 595	3,2028	7
J1	5	Wilgehout	361	1 230	1 461	3,1645	8
D2M01°	5	Hlotse	728	1 650	1 380	3,1398	9
W2R01	5,2	White Mfolozi	340	1 090	1 334	3,1250	10
A2R09	5	Pienaars	684	1 510	1 303	3,1148	11
N4°	5,2	Coerney	324	902	1 131	3,0533	12
H4R02	5	Nuy	377	948	1 102	3,0420	13
U2M22	5,2	Msunduze	877	1 440	1 097	3,0402	14
A7	5	Dorps	570	1 100	1 039	3,0168	15
V3R01°	5	Ngagane	830	1 270	995	2,9976	16
X2M18°	5,2	Mbyamiti	618	1 080	980	2,9913	17
D1M08	5	Malibamatso	277	710	962	2,9834	18
H1M06°	5	Breë	753	1 140	937	2,9719	19
S6M02	5,2	Kubusi	491	920	937	2,9716	20
H6M01°	5	Riviersonderend	280	680	917	2,9623	21
X2	5	Queens	320	700	883	2,9459	22
K6M02	5,2	Keurbooms	764	1 070	873	2,9412	23
R1M02°	5,2	Keiskamma	665	994	870	2,9393	24
J2°	5	Gamka	355	708	848	2,9283	25
A9M04°	5,2	Mutale	320	650	820	2,9137	26
A6	5	Klein Mogalakwane	400	700	790	2,8974	27
Q6M01	5,2	Baviaans	694	895	766	2,8845	28
B8M09°	5,2	Great Letaba	861	973	748	2,8740	29
J3M03°	5	Groot	426	667	729	2,8628	30
A2M03	5	Hex	495	709	719	2,8567	31
T1M01	5,2	Xuka	956	907	662	2,8207	32
H7M03	5	Buffeljags	450	600	638	2,8049	33
T5M04°	5	Mzimkulu	545	651	629	2,7987	34
H2M03°	5	Hex	718	737	621	2,7928	35
S3R01	5	Klipplaat	603	653	600	2,7781	36
W5M05°	5	Hlelo	804	709	564	2,7514	37
Q9M11°	5,2	Kat	539	555	539	2,7319	38
V1M10	5	Little Tugela	782	630	508	2,7061	39
X3M08°	5,2	Sand	1 064	658	455	2,6581	40
X2M11°	5	Elands	402	400	450	2,6533	41
V3M05	5	Slang	676	400	347	2,5404	42
Albasini	5,2	Levuvhu	509	SD	430,4	0,1865	
Average of Qxi, average of log Qxi & Median					1 000	2,9614	915

RMF (5,2) =	2 879	Skew	0,7454	-0,0187
RMF+Δ (5,4) =	3 674	Cv	0,4303	0,0630
		Correlation coefficient r:		0,9915

*$Q_{xi} = Q'_{xi}*(A/A')0,5$

$S_{log\ Qx}$ = standard deviation (SD) determined from the log Q_{xi} data

In summary, the algorithm consists of solving Equations (1), (6), (12) and (13) consecutively. The value of Q_{xx} is finally determined from Equation (13).

Inspection of Equation (13) shows that the value of $log\ Q_{xx}$ depends on three parameters: the first two, namely $log\ Q_{xm}$ and $S_{log\ Qx}$ depend on the distribution of Q_x alone. It is clear that the distribution of Q_x dominates the magnitude of the calculated extreme flood peak Q_{xx}. The third parameter, namely $Z_{\beta 2}$ is related to the AEP of the median Q_{xm} in Q_a sample space and is determined mainly from the distribution of Q_a in accordance with Equation (12). In this way the Q_a data set (annual maximum flows) is utilised to help calibrate the AEP of Q_{xx}.

It is recommended that the Q_{xi} data be presented graphically to check that the log-normal model (or any other selected model) is indeed an appropriate model. In most cases investigated so far correlation coefficients better than 0,98 have been obtained, demonstrating that the log-normal model is indeed a good theoretical model for simulating the distribution of Q_x. Only moderate extrapolation is required to estimate the magnitude of extreme flood peaks up to $Q_{10\,000}$ (in the sense that 50 data points with an observation period of 100 years each would represent 5 000 station years).

Application in cases where adequate or complete records of Q_a are not available

Unfortunately, annual records covering adequate record lengths may not be available for all sites included in a catalogue of "record maximum flood peaks". In such cases the value of f cannot be determined from Equation (6) and the value of α_1 cannot be determined from Equation (1).

The value of f could then be estimated from those sites that do have adequate annual records. It has been found that for *inland* sites in South Africa the value of f approaches 1,0. The calculated value for an extreme flood peak such as $Q_{10\,000}$ is not very sensitive to the f-value. For example, if the f-value is reduced from 1,0 to 0,8, the $Q_{10\,000}$ value reduces only by about 4%. Assuming $f = 1$ could then result in slightly conservative flood peak estimates in some cases.

To be consistent, the value of α_1 (AEP of the median Q_{xm} in Q_a sample space) must also be determined from regionally integrated information. Equation (1) could be applied to those stations that do have adequate annual records, on condition that there are an adequate number of such

Table 5 Estimates for Q_{xx} for Albasini Dam and estimates of AEP of the RMF and $RMF_{+\Delta}$

Assume log-normal model. AEP of median (in region 5,2) = α_1 = 1/59 = 0,01695 & f = 1							
T	$\alpha_2 = 1/T$	$\beta_2 = \alpha_2/(2.f.\alpha_1)$	$1/\beta_2$	z	log Q_{xx}	Q_{xx} (or Q_T)	T
1 000	0,001	0,02950	34	1,8882	3,3136	2 059	1 000
2 000	0,0005	0,01475	68	2,1767	3,3674	2 330	2 000
5 000	0,0002	0,00590	169	2,5181	3,4310	2 698	5 000
10 000	0,0001	0,00295	339	2,7533	3,4749	2 985	10 000
100 000	0,00001	0,00030	3 390	3,4362	3,6023	4 002	100 000
Estimates of AEP (= 1/T) of RMF and $RMF_{+\Delta}$							
RMF:	$\alpha_2 = \beta_2.2.f.\alpha_1$ =	0,00013	β_2 =	0,00380	3,4592	2 879	7 759
$RMF_{+\Delta}$:	$\alpha_2 = \beta_2.2.f.\alpha_1$ =	0,00002	β_2 =	0,00060	3,5651	3 674	48 810

stations available. Other methods could also be used to determine the AEP of the median, as long as they have a regional basis. It has been found that α_1 values typically fall between 1/50 and 1/200, and could thus be calculated fairly reliably by using available regionally based methods.

EXAMPLE: ESTIMATION OF $Q_{10\,000}$ FOR ALBASINI DAM SITE

An example to demonstrate the procedures and calculation steps for estimating extreme flood peaks such as $Q_{1\,000}$ to $Q_{10\,000}$ for a specific site is given in Tables 3 and 4 and the final results are given in Table 5. Figure 3 demonstrates the excellent correlation between selected *log* Q_{xi} values and the regression line on the log-normal probability scale.

It is always good practice to do sensitivity analyses. Sensitivity could, for instance, be tested by selecting only data from the eastern part of South Africa (in which the site under investigation is located), or by selecting only data from region 5,2 (in which the site is located) if enough data points were available. Sensitivity for the estimated value of the AEP of the median should also be tested as was done under item 6 in Table 3.

RESULTS OF SOME GENERALISED INVESTIGATIONS

A number of different catchment sizes and hydrological regions have been analysed by means of the REFSSA method and the results are summarised in Table 6. The scope of the investigations was limited by the availability of verified data on extreme flood peaks. Consequently, the REFSSA method was tested only for South African regions 4,6 and 5,0 as demarcated by Kovács (1988) and for catchment areas between 100 and 7 000 km^2.

Data were selected and handled as follows for the purpose of this investigation:

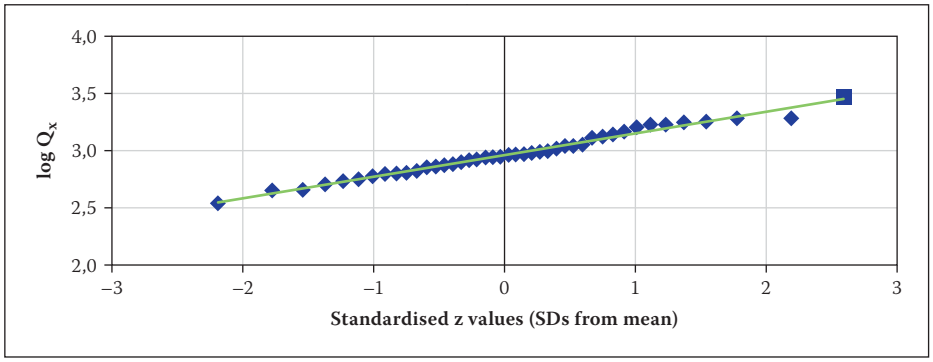

Figure 3 Graphic presentation of ranked transformed *log* Q_x values for Albasini Dam (estimate for $Q_{10\,000}$ indicated by the top square symbol)

- Record maximum flood peak data were selected from the catalogue as published by Kovács (1988) in accordance with the selection requirements as listed in Tables 1 and 2. Regions as demarcated by Kovács (1988) were used, but as amended below.
- Data were selected from the RMF region in which the site under investigation falls, as well as from the adjacent $RMF_{+\Delta}$ region which is one increment higher (more extreme). This conservative approach to the selection of data for a "similar hydrological region" was followed owing to statistical uncertainty (small sample sizes, short record lengths, inaccuracy of the source data and uncertainty regarding demarcation of boundaries). One could also argue that a storm event could blow over from the more extreme region to the less extreme region. This argument is supported by the record maximum flood peak data, which do not always abide by boundaries as demarcated.
- The region in which a site under investigation falls was used for the calculation of the AEP of the median. Complete records of verified annual maximum flows were not readily available and the method used in the example in Table 3 (item 6) was used to estimate the AEPs of medians. Compared with site-specific analyses (e.g. the rational method), the above method appears to give conservative (higher) AEPs for the median in most cases. This would result in conservative estimates for extreme flood peaks. It was found that the results are not very sensitive to the value of the AEP of the median. The effect on the value of $Q_{10\,000}$ in the case of the above example when using an AEP of 1/40 or 1/80 in the place of 1/59 was less than 5%.
- Where applicable, exceptionally low data points, which do not really represent extreme flood peaks and thus do not really comply with the definition of Q_x were discarded. This reduced the absolute values of negative skewness coefficients to almost zero in most cases. Exceptionally low points have the undesirable effect of increasing the standard deviation, causing higher estimates for Q_{xx} (e.g. $Q_{10\,000}$). Discarding low data points is considered to be compatible with the definition of Q_x and the way in which the algorithm is constructed. In the same vein, in a few cases of high positive skewness, low data points were added to the data in order to reduce positive skewness to below 0,1. This latter action results in slightly higher estimates for Q_{xx}, avoiding under-estimation of Q_{xx} when using the log-normal model.

Table 6 Results of preliminary analyses by REFSSA method for regions 4,6+* and 5+*

Catchment (km^2)	Region*	Sample size	RMF (m^3/s)	T=1/AEP (RMF)	Estimated Q_{xx} values (m^3/s)				$Q_{10\,000}$ / RMF	Values of log Q_x data		
					1/1 000	1/2 000	1/5 000	1/10 000		Skew	Cv	Corr r
100	4,6+	21	575	1 422	531	616	715	786	1,37	0,110	0,048	0,975
200	4,6+	24	837	1 028	829	1 004	1 223	1 388	1,66	0,043	0,062	0,984
500	4,6+	41	1 372	933	1 398	1 660	2 015	2 293	1,67	-0,031	0,072	0,991
1 000	4,6+	45	1 995	929	2 033	2 390	2 870	3 244	1,63	0,063	0,065	0,991
2 000	4,6+	40	2 901	1 140	2 820	3 247	3 811	4 244	1,46	0,086	0,055	0,995
3 500	4,6+	42	3 925	1 162	3 808	4 348	5 068	5 623	1,43	0,094	0,054	0,996
5 000	4,6+	35	4 758	879	4 895	5 631	6 628	7 405	1,56	0,108	0,057	0,995
7 000	4,6+	40	5 706	1 074	5 627	6 400	7 426	8 214	1,44	0,123	0,050	0,997
Average for 4,6+		36	–	1 071	–	–	–	–	1,53	0,075	0,058	0,990
100	5,0+	33	1 000	2 464	837	962	1 128	1 256	1,26	–0,093	0,067	0,986
200	5,0+	33	1 414	1 921	1 234	1 425	1 681	1 879	1,33	–0,218	0,066	0,987
500	5,0+	42	2 236	2 697	1 853	2 120	2 479	2 758	1,23	–0,019	0,063	0,992
1 000	5,0+	47	3 162	2 139	2 734	3 124	3 647	4 052	1,28	–0,058	0,058	0,985
2 000	5,0+	37	4 472	2 347	3 849	4 355	5 024	5 536	1,24	–0,014	0,050	0,993
3 500	5,0+	32	5 916	1 762	5 294	6 057	7 091	7 899	1,34	0,101	0,057	0,994
5 000	5,0+	31	7 071	2 115	6147	7 002	8 162	9 068	1,28	0,101	0,056	0,993
7 000	5,0+	25	8 367	2 877	7 010	7 897	9 087	10 007	1,20	0,104	0,050	0,993
Average for 5,0+		35	–	2 290	–	–	–	–	1,27	–0,012	0,058	0,990
Average for both regions			–	1 645	–	–	–	–	1,40	0,034	0,058	0,990

* Region 4,6+ means that the site falls in region 4,6 but data were selected from both regions 4,6 and 5,0 for the analysis.
* Region 5,0+ means that the site falls in region 5,0 but data were selected from both regions 5,0 and 5,2 for the analysis.

Discussion of results in Table 6

- The average correlation coefficient between actual *log* Q_{xi} data and corresponding log-normal regression lines is 0,99 (and better than 0,98 in all cases if the sample size exceeds 25). The average correlation coefficients are the same for regions 4,6+ and 5,0+. The excellent correlation coefficients support the postulation that the estimation of the magnitude of extreme flood peaks should be based mainly on extreme flood peak data.
- The skewness coefficients of selected *log* Q_{xi} data are very low in most cases and significantly lower than those of natural Q_{xi} data.
- The above two points indicate that the selected Q_{xi} data in the range of extreme flood peaks are generally log-normally distributed.
- The average coefficient of variation c_v (standard deviation divided by mean or $S_{logQx}/log\ Q_{xm}$) for region 4,6+ is the same as that for region 5,0+ (namely 0,058). This indicates remarkable consistency. If c_v (in terms of *log* Q_x) can be accepted as a constant, this has an enormous impact on statistical certainty: calculations in one case with a sample size of 30 (not shown in Table 6) assuming c_v is constant, reduced the one-sided 95% upper confidence limit for $Q_{10\,000}$ to 1,14($Q_{10\,000}$) compared with 1,54($Q_{10\,000}$) without the aforementioned knowledge.
- The AEP of the RMF is not constant but varies from 1/879 to 1/2 877 for different catchment sizes and for different regions in Table 6. The average AEP of the RMF for region 4,6+ (1/1 071) is about double that for region 5,0+ (1/2 290).
- The average AEP of the $RMF_{+\Delta}$ for both regions is about 1/10 600 (not shown in Table 6).
- The $Q_{10\,000}$/RMF ratio varies between 1,2 and 1,75. This is within the expectation of the SANCOLD Guidelines (1991) which state that PMF/RMF ratios exceeding 2,0 should not be accepted. In comparison, Görgens et al (2007b) found that for PMF values as calculated by the unit graph method of HRU (1972), the PMF/RMF ratio varies from 0,65 to 6,9 (for the same regions and range of catchment sizes considered in Table 6). The upper ratio exceeds all reasonable expectations. It is clear that the above PMF method produces unreliable results.
- The estimated values for $Q_{10\,000}$ have not been exceeded by actual records in any of the analysed cases. In one case the estimated value for $Q_{10\,000}$ could have been exceeded, but the accuracy of the relevant record maximum flood peak (station K4, Goukamma River) is indicated as "unknown" and this record was therefore not used in the analysis. Estimated values for $Q_{10\,000}$ could have been exceeded in two additional cases if storm events had blown over from the adjacent, more extreme region. Kovács (1988) estimated the cumulative station years of independent flood peaks for the relevant regions and areas at approximately 6 700 station years. The probability that $Q_{10\,000}$ could have been exceeded during a record of 6 700 station years is estimated at roughly 49%. The estimated values for $Q_{5\,000}$ were equalled in one case and the estimated values for $Q_{2\,000}$ were equalled or exceeded in four cases. The probabilities that these flood peaks could have been exceeded during a record of 6 700 station years are estimated at roughly 74 and 97% respectively. The results of the REFSSA method seem plausible in view of the above probabilities. (However, it should be pointed out that the common probability equation $P(Q>Q_T) = 1-(1-1/T)^L$ – with L equal to the total observation period of 6 700 station years in the above case – for estimating the above probabilities is based on the Bernoulli

Table 7 Approximate equations*

Q_{xx}	Equation	
	Region 5	Region 4,6
$Q_{10\,000}$	$131A^{0,495}$	$77,0A^{0,532}$
$Q_{5\,000}$	$117A^{0,497}$	$67,9A^{0,534}$
$Q_{2\,000}$	$97,4A^{0,501}$	$55,2A^{0,540}$

* Only valid for 100 km² to 7 000 km²

sequence, which requires complete statistical independence of events, but this is not the case for hydrological data. The aforementioned probabilities are therefore rough estimates at best.)

- Sensitivity analyses were done to compare the estimated values for $Q_{10\,000}$ in Table 6 with those obtained when using data from only region 5,0. It was found that combining regions 5,0 and 5,2 for data-selection purposes, as was done for Table 6, resulted in estimates for $Q_{10\,000}$ that are on average only 6% higher (varying between 0 and 20%) than when using data from only region 5,0. However, this preliminary finding could be inaccurate and biased because the sample sizes for region 5,0 alone were much smaller than when regions 5,0 and 5,2 were combined (25 compared to 35 on average).
- Most of the assumptions made to produce Table 6 are considered to be slightly conservative. Consequently, the results in Table 6 should be regarded as slightly conservative for most of the cases at this stage.
- The estimated values for $Q_{10\,000}$, $Q_{5\,000}$ and $Q_{2\,000}$ in Table 6 are represented graphically against catchment size on logarithmic scales in Figures 4 to 6. Remarkably high correlation coefficients have been obtained between the estimated values and the regression lines (all better than 0,99). Approximate equations for these regressions lines are given in Table 7, but it is recommended that a complete analysis as shown in the example (Tables 3 to 5) be done for designs or safety evaluations of important projects.

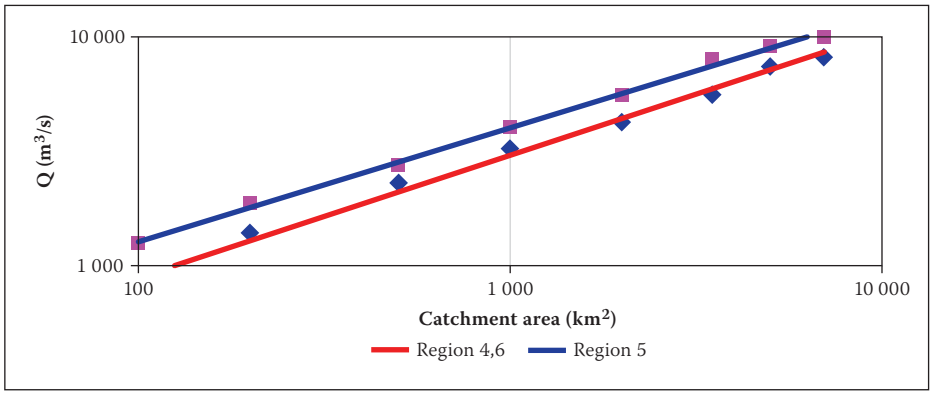

Figure 4 Relation between estimated $Q_{10\,000}$ values and catchment size

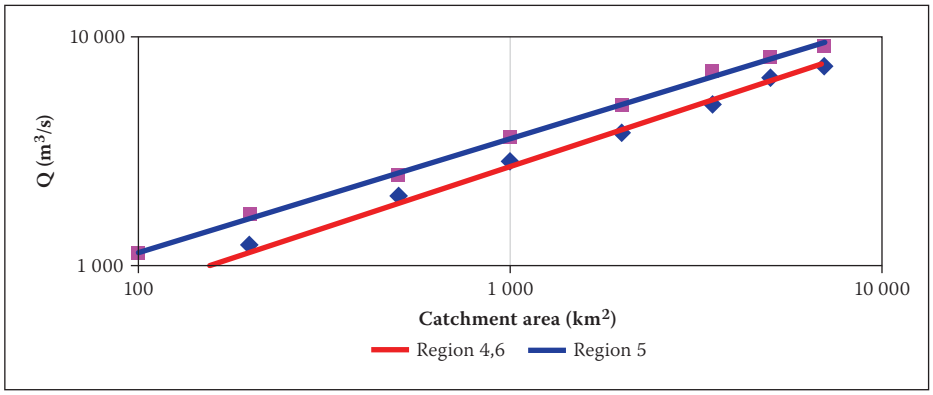

Figure 5 Relation between estimated $Q_{5\,000}$ values and catchment size

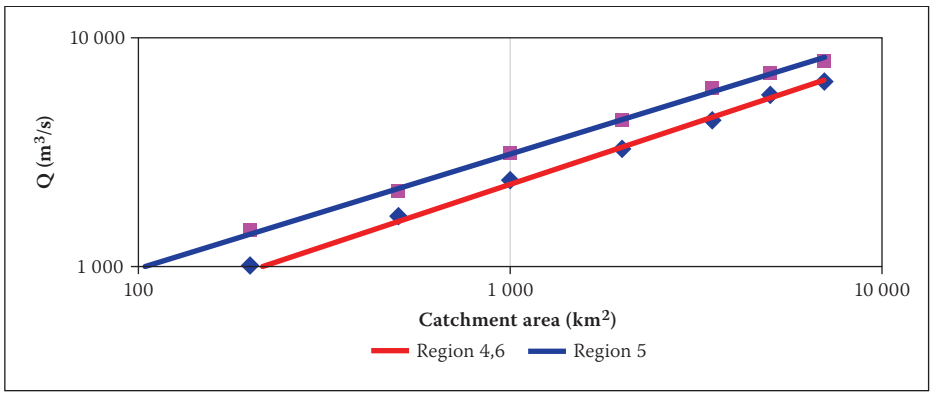

Figure 6 Relation between estimated $Q_{2\,000}$ values and catchment size

CONCLUSIONS

The applicability of the REFSSA method has been demonstrated for the estimation of extreme flood peaks in two major hydrological regions in South Africa. Despite limitations with regard to the quality and quantity of available "record maximum flood peak" data, relatively high correlation coefficients have been obtained between transformed "record maximum flood peaks" and regression lines (0,99 on average on log-normal scale as given in Table 6). This indicates excellent reliability within the hydrological environment and supports the theoretical basis of the REFSSA method.

Because the REFSSA method is new, caution should be exercised. Data selection should be done carefully in accordance with the selection requirements proposed in this paper. Sensitivity analyses should always be done to test sensitivity, for instance by using only data closer to the site under investigation, but still within the "similar hydrological region". Sensitivity should also be tested by varying the AEP of the median.

Although the REFSSA method could currently be regarded as one of the better methods in South Africa for determining the magnitude and AEPs of extreme flood peaks larger than $Q_{1\,000}$, the following limitations should be borne in mind:

- It is a regional method and would thus be more reliable for catchments with average catchment characteristics (corresponding to those of the source data).
- An adequate number of relevant "record maximum flood peaks" of adequate accuracy must be available to do an analysis.
- It is provisionally considered applicable for the estimation of extreme flood peaks between $Q_{1\,000}$ and $Q_{10\,000}$ and for catchment sizes between 100 and 7 000 km².

RECOMMENDATIONS

The following further actions or investigations are recommended:

- The catalogue of "record maximum flood peaks" as published by Kovács (1988) should be extended, updated and its accuracy improved as far as possible.
- The applicability of the REFSSA method should be tested for all other regions in

South Africa as has been done for regions 4,6+ and 5+ in Table 6 of this paper, after extension of the catalogue.

- It should be investigated whether the REFSSA approach could also be employed to estimate extreme flood volumes from record maximum flood volumes.
- The applicability of the REFSSA method for catchments smaller than 100 km^2 and larger than 7 000 km^2 should be investigated. Not enough verified data were available to do this investigation in the present study.
- It is of critical importance that future flood events are accurately surveyed and documented.

ACKNOWLEDGEMENTS

Zoltan Kovács (1988) must be recognised for publishing a useful catalogue of verified "record maximum flood peaks" in South Africa. The catalogue was based on records, surveys, estimates and documents compiled by the Department of Water Affairs, other organisations and individuals and these efforts must be commended. Without this catalogue this paper would probably not have seen the light. My colleague Mr C L van den Berg is thanked for checking the algorithm for correctness. The anonymous reviewers are thanked for their valued comments. Permission by the Department of Water Affairs to publish this paper is gratefully acknowledged. It should be noted that the opinions expressed in this paper are those of the author and not necessarily those of the Department.

REFERENCES

Alexander, W J R 2000. Flood risk reduction measures. Pretoria: University of Pretoria, Department of Civil Engineering.

Alexander, W J R 2009. Mathematics vs pattern recognition in water resource studies. *Civil Engineering*, 17(5).

Cunnane, C 1988. Methods and merits of regional flood frequency analysis. *Journal of Hydrology,* 100.

Francou, J & Rodier, J A 1967. Essai de classification des crues maximales. *Proceedings,* Leningrad Symposium on Floods and their Computation, UNESCO.

Görgens, A 2003. Design flood hydrology. Unpublished lecture notes. University of Stellenbosch & Ninham Shand.

Görgens, A 2007a. Joint Peak-Volume (JPV) Design Flood Hydrographs for South Africa. WRC Report No 1420/3/07, Water Research Commission, Pretoria, South Africa.

Görgens, A, Lyons, S, Hayes, L, Markhabane M & Maluleke, D 2007b. Modernised South African Design Flood Practice in the Context of Dam Safety. WRC Report No 1420/2/07, Water Research Commission, Pretoria, South Africa.

HRU (Hydrological Research Unit) 1972. Design flood determination in South Africa. Report No. 1/72, Johannesburg: University of the Witwatersrand.

Kovács, Z 1988. Regional maximum flood peaks in southern Africa. Technical Report TR 137, Pretoria: Department of Water Affairs.

SANCOLD 1991. Guidelines on safety in relation to floods. Pretoria: SANCOLD.

Wiltshire, SE 1986a. Regional Flood Frequency Analysis I: Homogeneity Statistics. *Hydrological Sciences Journal.*

Wiltshire, SE 1986b. Regional Flood Frequency Analysis II: Multivariate Classification of Drainage Basins in Britain. *Hydrological Sciences Journal.*