A review of the regional maximum flood and rational formula using geomorphological information and observed floods

Geoff Pegram* and Mohamed Parak
Civil Engineering Programme, University of KwaZulu-Natal, Durban 4041, South Africa

Abstract

Flood estimation methods in South Africa are based on three general approaches: empirical, deterministic and probabilistic. The "quick" methods often used as checks are the regional maximum flood (RMF) and the rational formula (RF), which form part of the empirical and deterministic methods respectively. A database of annual flood peaks was used in a probabilistic approach to review these methods and to provide preliminary insight into their estimates of flood peaks. This paper examines the following: the relationship between floods and landscape; the estimation of the return period of the RMF; the use of ratios in scaling RMF flood peak estimates to flow rates of shorter return periods; the applicability of the modified rational formula (MRF); the examination of the relationship between scaling parameters and regional parameters. It turns out that the RMF is the best of all methods examined in this preliminary study (other than statistical) in estimating the 200-year flood peak at an ungauged location.

Keywords:flood estimation, rational formula, regional maximum flood, generalised extreme value distribution

Introduction

The realistic estimation of the magnitude of a design flood peak with a chosen probability of exceedence that can be expected at a given site in a given region is fundamentally important in the planning, design and operation of hydraulic structures and for the preservation of human life and property. The basic approaches involved in flood estimation are the empirical, deterministic and probabilistic approaches. These methods are calibrated from historical flood records from gauged catchments and their relative usefulness depends on the accuracy with which they are able to predict flood sizes in ungauged catchments. In South Africa, reasonable estimates of maximum recorded flood magnitudes are derived from the use of the empirically-based approach of the regional maximum flood (Kovacs, 1988), and design floods may be determined using deterministic approachs such as the rational formula (RF), the SCS model or the unitgraph method and from the analyses of flood frequencies through a probabilistic approach.

Kovacs' **empirical method** is probably the most robust method available locally and, relatively accurately, predicts the regional "maximum" flood that can be expected from a given site based only on the site's catchment area and region. The advantage of the empirical method is its ease of use as it deals directly with the parameter of interest, namely the flood peak discharge, and avoids the assumptions involved in transforming rainfall inputs into flood outputs. The disadvantages of the RMF method are that:

- The recurrence interval (RI) associated with this "maximum" is not clear, although Kovacs estimated it to be greater than 200 years
- The regions defined by individual *K*-values have widely varying rainfall properties and
- It seems naive to estimate flood peaks on area and zone only.

The **deterministic** rational formula (RF) approach involves (in a simple, but sound manner) the analysis of all the factors involved in flood prediction from converting rainfall inputs into flood outputs; it usually carries a caveat that it should not be used for "large" catchments, but recent work (Alexander, 2002 and Pegram, 2003) has shown that this caution is too conservative.

Flood **frequency analysis** involves the fitting of a probability model to the sample of annual flood peaks, recorded over a period of observation, for a catchment of a given region. The model parameters established can then be used to predict the extreme events of large recurrence interval. The advantage of this method is that the events of large recurrence interval, which are longer than the record period, can be determined through cautious extrapolation of the fitted distribution based on the model parameters. The disadvantage of this method is that it can only be applied where data have been collected and it is often not clear how the analysis can be extended to ungauged locations.

The question that arises is "which method is fair to use?" The answer depends on the availability of data. When no hydrological (rainfall and runoff) records exist for a catchment, the empirical methods provide the only means of flood prediction. This situation is the most common case in the design of hydrological projects. When estimates of design rainfall are available (Adamson, 1981; Smithers and Schulze, 2003) or rainfall records suitable for a frequency analysis are available from a nearby rain-guage, then the rational formula (RF) becomes applicable, in addition to the empirical. When flood records of sufficient length (>30 years or so) exist, possible future flood peaks of given frequency can be determined by modelling past floods with an extreme value distribution. Even in this fortunate situation, it is prudent to crosscheck the frequency estimate with deterministic and empirical estimates.

It is the aim of this exploratory study to provide a review of the above methods in order to determine the accuracy of the estimates of the design flood, where the design flood is the flood associated with a chosen return period or recurrence interval of exeedence. The base data are the set of annual flood peak records from 130 sites around South Africa that were used *inter alia* by Kovacs (1988) in his empirical study.

To summarise: this paper attempts to provide preliminary insight into the following questions concerning the RMF and RF flood determination methods in South Africa using the recorded flood peaks:

- Does the addition of landscape data (catchment morphometry) improve the prediction of floods by the RMF?
- Can a return period be associated with the RMF by comparing its computed magnitude with those modelled from historical records?
- Are simple country-wide Q_T/Q_2 ratios valid for scaling flood maxima (or RMF values) to floods of shorter (or even longer) return periods?
- Is the modified rational formula (MRF) a useful modification and reasonable alternative to the RF and other flood prediction methods?
- Are there any inferences that can be drawn from the variation of the shape parameters k of the GEV Distribution, used to model the observed floods, and Kovacs' regional K-values?

The methodologies involved in assessing each of the objectives listed in this paper will be expanded in detail in the sequel. Before this can be done, an explanation of how the recorded data set was used in the calibration and validation of the objectives outlined is given.

The use of recorded flood data in this study

Annual flood peaks from 130 catchments across South Africa were obtained from Zoltan Kovacs of the Department of Water Affairs by Peter Adamson while working with the first author in 1988 and 1989. This data set, although old (final year of record was 1988), provided the starting point for this pilot study in the review of these flood determination methods. The length of record of the data set used herein ranged from 9 years to 76 years and forms a sub-set of the data used by Kovacs for the construction of the RMF curves. To find the return period associated with each annual peak, the Weibull Plotting Position was used (it is more conservative than the Cunnane Plotting Position), which is expressed as:

$$T = \frac{N+1}{r} \tag{1}$$

where:

T   is the return period (years) associated with the flood peak of rank r
N   is the length of record (years)
r   is the rank of the flood peak; r = 1 for the largest peak.

This resulted in a list of annual peaks each with an associated return period for each catchment. Following the work of De Michele and Salvadori (2002) and Kjeldsen et al. (2002), the distribution of these peaks was assumed to follow a generalised extreme value (GEV) distribution. This distribution takes the following form:

$$Q_T = \mu + \sigma y_T \tag{2}$$

where:

Q_T is the *T-year* return period flood peak estimate
μ, σ are shift and scaling parameters respectively
y_T is the GEV reduced variate corresponding to a *T*-year return period, i.e.

$$y_T = \frac{1}{k}\left[1 - \left\{-\ln\left(1 - \frac{1}{T}\right)\right\}^k\right] \tag{3}$$

where:

k   is a shape parameter. When k = 0, the GEV reduces to the EV1 or Gumbel distribution.

This model of the flood data formed the basis with which to review the other approaches. Some of these data and their distribution fits are presented in Table A1 (Part 3) in the **Appendix**.

Empirical approach extended by including landscape properties

In his empirical approach Kovacs (1988) determined envelopes of the maximum flood peaks from the original extended data set, of which, as has been noted above, the data in the **Appendix** are a subset. Kovacs' data set included some rare singleton floods (not used in this study) to which he cautiously ascribed a representative record length not exceeding 200 years. He used this extended set to obtain the RMF lines based on the Francou-Rodier equation. The technique was to plot maximum flood peaks against catchment area for hydrologically homogeneous regions to produce envelope curves which define the upper limit of expected flood peaks for a given region. The curves are defined by the following equation:

$$Q_{RMF} = 10^6 \left(\frac{A}{10^8}\right)^{1-0.1K} \tag{4}$$

where:

A   is catchment area in km^2
K   is a regional dimensionless factor which accounts for the influence of variations in rainfall, geology, land-form and vegetation cover in flood production.

It should be noted at this juncture, that the "secret" to the success of the RMF is the careful way in which Kovacs chose the regions to group the flood data. He did this by examining the actual K-value (from Eq. (4)) for each catchment where the flood peaks and catchment areas were known. Regional boundaries of K were delimited by considerations of individual K-values within the region, the number and accuracy of the data in a particular area, existing boundaries, maximum recorded 3 day storm rainfall, topography, catchment orientation with respect to dominant storm generating weather systems, general soil permeability, main drainage network and the location of large dams situated upstream from the guaging sites (Kovacs, 1988). Of these considerations, individual K-values were evidently the most important and the regions were traced based on this. In areas of high flood peak potential a difference of 0.2 between individual K-values was allowed for and a difference of 0.6 in areas of low flood peak potential.

What is evident from Eq. (4), and all other derived empirical equations produced for the prediction of floods, is its dependence on *Area* as an independent variable. Because of the RMF's apparent naivety, one might expect other parameters of the fluvial landscape to play an important role in flood response and make the estimates more accurate. Flood geomorphologists, such as Horton (1932; 1945) and Strahler (1952; 1964) and many others since have been interested in relating flood discharges to physical measures of landscape (morphometry). They identified parameters of the fluvial landscape which intuitively would correlate well with flood discharge.

Linear parameters (such as stream orders and stream lengths), areal parameters (such as catchment area, catchment shape and drainage density) and relief parameters (such as catchment relief, catchment slope, channel slope and ruggedness number) are some of the physical measures that have been identified as significantly

affecting flood response. One can expect such a relationship between flood discharge and catchment morphometry to exist because a catchment is effectively "an open system trying to achieve a state of equilibrium" (Strahler, 1964). Precipitation is input to the system and soil (eroded material) and excess precipitation leave the system through the catchment outlet. Within this system an energy transformation takes place converting potential energy of elevation into kinetic energy where erosion and transportation processes result in the formation of topographic characteristics. Thus it is evident that floods, and the landscape through which they drain, form a mutual relationship and ultimately catchment morphometry should reflect this phenomenon. In this pilot study, an effort is made to determine if landscape parameters improve the prediction of floods in empirical equations based solely on catchment area.

What is the recurrence interval of the RMF?

What is also evident from the RMF method of flood determination is that one is not easily able to associate a return period with the estimated floods. The envelope floods (estimated from the RMF lines) have been described as the maximum flood that the site has experienced. This is not easy to quantify in terms of a return period. Kovacs himself estimates the return period to be greater than 200 years (Kovacs, 1988), although he does not explicitly model their probability distribution. Where the representative period (N) of a flood was not known, Kovacs did not allow this to exceed 200 years and a provisional N value was estimated based on the assumption that the ratio of the 200-year peak to RMF, Q_{200}/RMF was 0.65.

When determining a design flood, the exact magnitude of the flood and its probability of exceedance need to be known. The absence of an estimate of the return period associated with the RMF makes the quantification of risk by this method problematic and, as it represents maximum discharges, it tends to be used by designers as a conservative method. This article aims to, *inter alia,* determine a return period associated with the RMF by simultaneously plotting the floods determined from the RMF method and the historical floods extrapolated to the 50-, 100- and 200-year recurrence intervals modelled with the GEV distribution.

The first author has for many years suggested that the RMF envelopes have a recurrence interval of about 200 years, as estimated by the Weibull Plotting Position. This estimate was based on the following argument: the data used by Kovacs (1988) in the construction of the RMF lines had, in many instances, record lengths (actual and estimated) of the order of 100+ years. The RMF lines are envelopes, drawn above the data whose maximum record length N was 200 years. If we are conservative and estimate the recurrence interval of the RMF line using the Weibull Plotting Position, the RI (T) of the largest observation would be $T_N = (N+1) \approx 200$ years. It was decided to examine this conjecture as part of this study.

The use of Q_T/Q_2 ratios for scaling flood maxima

It is useful to know how to scale the "200-year RI" RMF or any other flood of recurrence interval T-years to shorter return period floods where desired. The first author suggested such a scaling in Chapter 2 of TRH 25 (1994). It was thought that this study was also an opportune time to check that assumption which was based on the following argument.

Hiemstra and Francis (1979) examined the relationship between the peak flood discharge of a catchment and its hydrograph

TABLE 1							
Ratio between T-year flood and 2-year flood (TRH 25, 1994)							
T	2	10	20	50	100	200	1000	10000
Q_T/Q_2	1	3.57	5.18	7.80	10.24	13.14	22.00	41.24

shape defined by the volume. What they discovered was that for extreme events, the peak discharges of various magnitudes were well modelled by the censored log-normal distribution. They extracted the statistics of many floods in the Department of Water Affairs and Forestry's break-point continuous flow rate database at that time and found the coefficient of variation of the peak discharges averaged 1.3 with a fairly small variation. Based on this, the first author produced ratios which relate the T-year flood to the 2-year flood. These ratios, Q_T/Q_2 reproduced from TRH 25 (1994) in Table 1, enable one to convert any flood of a given RI to a T-year flood.

To check this assumption in this study, the maximum observed flood recorded in the observation period from each of the 130 catchments was associated with a return period using the Weibull Plotting Position ($T_{max} = N+1$). This flood was then scaled to 10- and 50-year flow rates using the Q_T/Q_2 ratios and compared with those computed from the GEV model fitted to the full set of data in each record. These values were then compared and it was determined if these ratios are applicable in reducing flood maxima to floods of desired recurrence intervals.

The modified rational formula (MRF)

The ratonal formula is expressed (in *SI* units) as:

$$Q_{peak} = ciA / 3.6 \qquad (5)$$

where:

- c   is a dimensionless runoff coefficient which ranges from 0 to 1
- i   is the intensity of rainfall (mm per hour) of return period T (years) and duration T_c, where T_c is the time of concentration (hours) of the catchment
- A   is the area of the catchment (in km^2).

This formula is usually limited to catchments with small areas ($<$ 100 km^2). The reason usually given for this is that the formula does not take into account the areal reduction factor (ARF) and utilises point design rainfall intensity. It should be noted that flood-causing rainfall in smaller catchments is mainly due to concentrated thunderstorm activity, whereas flood-producing rainfall in larger catchments is mainly due to long-duration, widespread synoptic events (Pegram, 2003). The consequence is that the larger the catchment, the longer the duration of the flood-causing rainfall. To simplify the analysis, Pegram (2003) used the scaling properties of the GEV distribution fitted to rainfall depths, hence, using the GEV distribution defined in Eq. (3), the precipitation scaling relationship becomes:

$$P_{d,T} = (\mu + (\sigma / k)[1 - \{-\ln(1 - 1/T)\}^k])d^{1-\eta} \qquad (6)$$

where:

$P_{d,T}$ is the rainfall depth of duration d and return period T.

For each of Kovacs' regions, representative 1-, 2- and 3-day rainfall depths for 2-, 5-, 10-, 20- and 50-year return periods were extracted
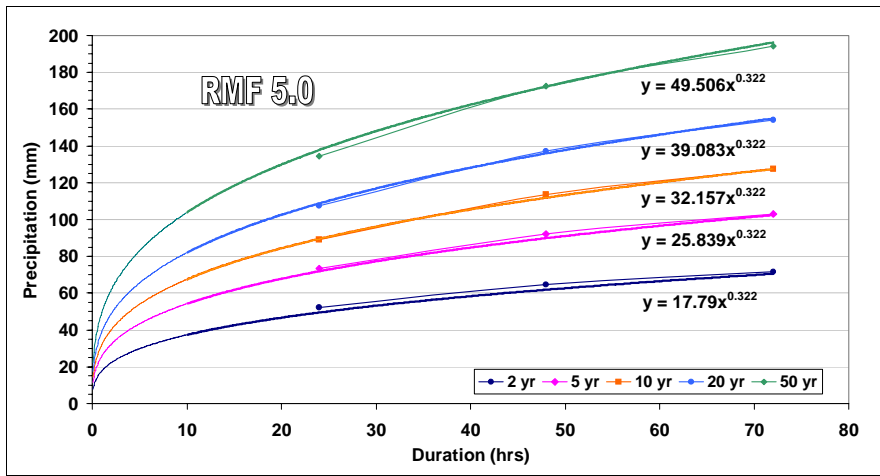
Figure 1
Fit based on the GEV (smooth curves), to an average of Adamson's (1981) data (dots) for Kovacs (1988) RMF Region 5 (from Pegram, 2003). The thin lines are trend-lines fitted to each set of the 1-, 2- and 3-day rainfall duration data. The thick lines are the combined models fitted to all the data with a common power law relationship.

from Adamson (1981) by Westray (2001). These were averaged (pooled) by region and Eq. (6) was fitted to the 15 points by Least Squares. An example is given in Fig. 1 where the pooled data and the fitted function are compared for Region 5. The k and η values were fixed at -0.182 and 0.678 respectively by using the whole South African data set as a first approximation (Pegram, 2003). Values of μ and σ were the parameters that were estimated for each region. It was found that the coefficient of variation $C_v = \sigma/\mu$ was effectively independent of Kovacs' regions, so the major variable to concentrate on was the parameter μ.

In addition to this simplification, for all the catchments whose data are available in the report by Petras and Du Plessis (1987), the time of concentration T_c computed from the Kirpich (1940) formula: $T_c = 0.0633(L/S)^{0.385}$ (where as usual, L is the length and S is the average slope of the catchment's longest watercourse) was set to the duration of the flood-causing storm as demanded by the RF. When this duration T_c was plotted against area, the points clustered around a curve to which a power law relationship could be fitted. This is also the practice in Australian Rainfall and Runoff (AR&R, 2001). For interest sake, this was superimposed on the areal reduction factor (ARF) diagram, published in the Flood Studies Report (FSR, 1975), which appears as Fig. 2. It is possible that the FSR's ARF curves over-estimate the ARF in South Africa, but the degree is likely to be a matter of climate. Conscious of this, it is still remarkable that the *Area ~ T_c* curve yields an almost constant ARF value of 88% across the FSR curve. Thus, as long as the precipitation intensity used in the rational formula corresponds to the time of concentration of the catchment, the point rainfall is automatically scaled by a constant ARF. Combining these ideas, the MRF was then expressed (Pegram, 2003) in preliminary form as:

$$Q_{MRF} = c \times 0.318 P_{1day,2} \left[1 + 0.385 y_T \right] A^{0.558} \qquad (7)$$

where:

c   is the conventional rational formula (RF) $c: 0 < c < 1$
$P_{1day,2}$ is the median 1d annual maximum rainfall available from maps (e.g. Adamson, 1981; Smithers and Schulze, 2003)
y_T is the reduced variate of the GEV Distribution of the rainfall
A   is the catchment area in km^2.

In this paper the 10-, 20- and 50-year floods of the MRF are compared with the observed flood peaks modelled with the GEV distribution of the same recurrence intervals. The intention is to determine whether the MRF in its coarse form is possibly a useful candidate for predicting the design floods of a catchment.

Does the GEV regionalise following the RMF?

The annual observed flood data series, extracted from the observed records, were modelled using the GEV distribution. This was explained above. The records were thought to be long enough, in most cases, to make reasonable predictions of future events. Following this analysis, it was of interest to determine if the shape parameter k established by modelling historical floods using the GEV distribution, display any trends with a region descriptor such as Kovacs' regional *K*-value. That concludes the introduction. The full analyses are reported in the following sections.

Floods and landscape

Landscape data from 25 catchments were extracted in a preliminary study by Parak (2003) that corresponded with the peak discharges of the catchments modelled in this study. Parak (2003) captured morphometric data of 45 catchments across the country in his investigation into the relationship between floods and landscape. He used already catalogued data (Petras and Du Plessis, 1987 and Kovacs, 1988) and supplemented this with further data through map work from Midgley et al. (1994). In this paper the landscape data were utilised to assess whether they improved the prediction of floods compared with the RMF, which uses only catchment areas in particular regions. The flow rate that was used for comparison here was the 20-year event determined by modelling the historical floods of the catchments using the GEV distribution, the rationale being that:

- It would be the least likely estimate to be affected by fitting the wrong probability distribution
- Many of the records were longer than 20 years.

The flood and landscape data were split into two groups, one for calibration and the other for validation. The landscape data included catchment area, mean channel slope, mean annual precipitation, drainage density, catchment relief and ruggedness number. These are summarised in Table A1 and explained in the **Appendix** and a typical catchment and its derived geometry are shown in Fig. 3 (from Parak, 2003). It is aknowledged that the landscape data catalogued are sensitive to map scale, i.e. at different scales, different values of the parameters will be obtained. For example, the river detail shown on a larger scaled map is much less than that which is shown on fine-scaled maps. This has a direct influence on the magnitudes of the landscape parameters. Measures such as total stream length, stream orders, drainage densities and ruggedness

Figure 2
The FSR diagram for ARF (FSR, 1975), as contours in percentages, with T_c vs area relationships (using Kirpich's (1940) formula:
$T_c = 0.0633 (L/S)^{0.385}$ for South African catchments superimposed (Westray, 2001). The best fit is:
$T_c = 0.148 A^{0.651}$.
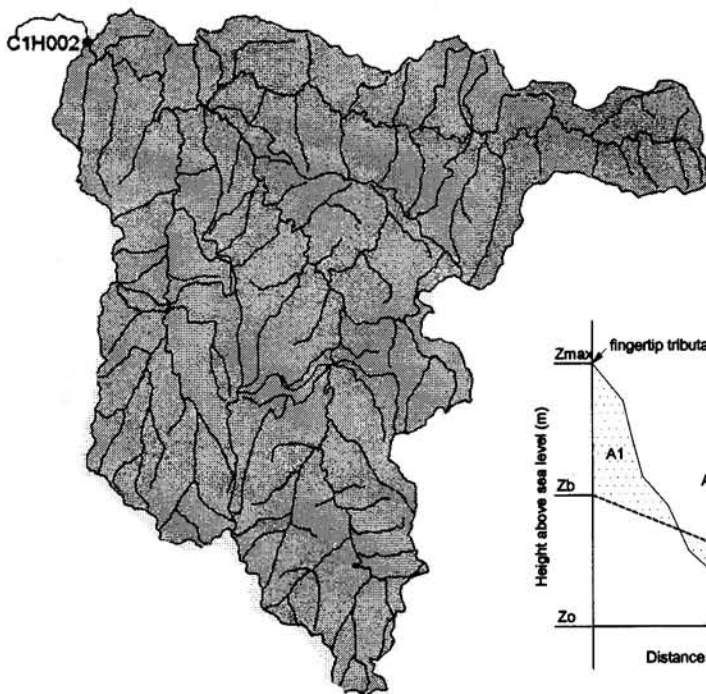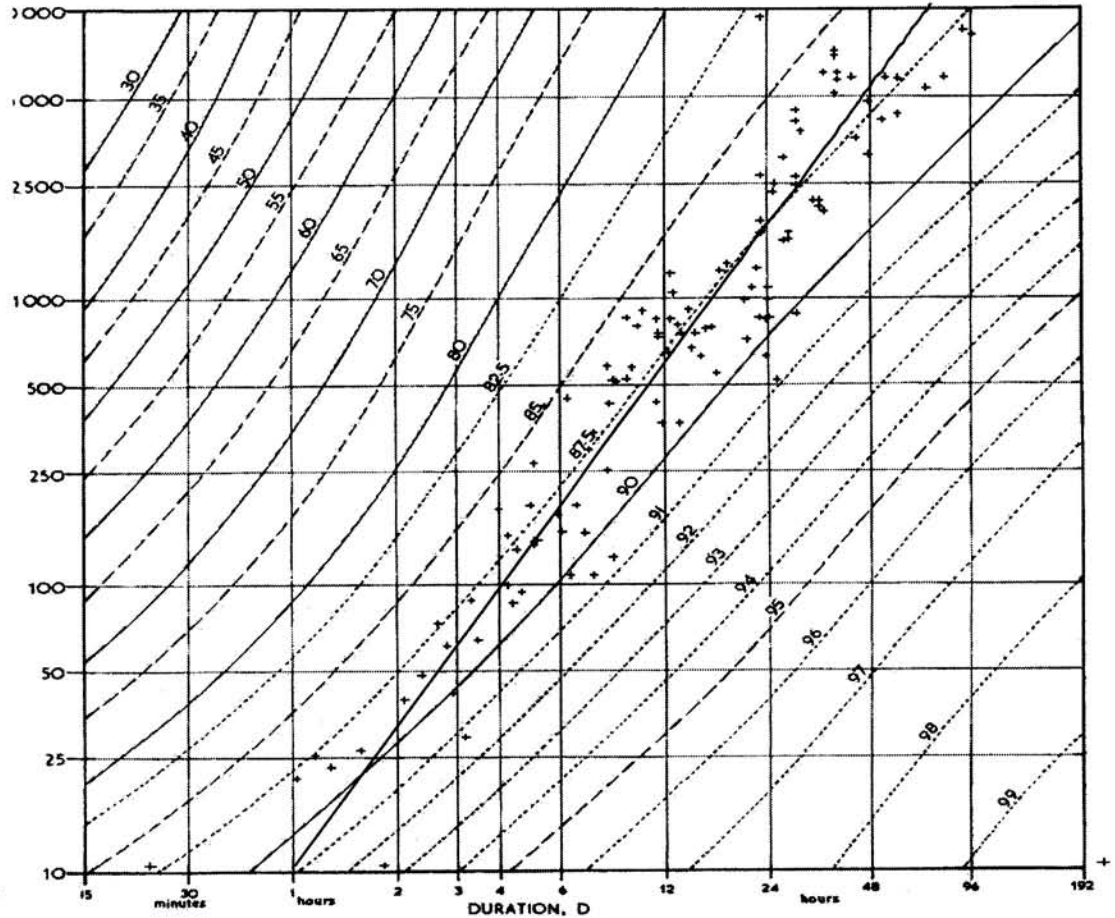
Basin area	4 152 km²
Effective area	4 152 km²
Longest watercourse	181 km
Total stream length	1 286.7 km
Basin relief	820 m
Mean river slope	0.00132
Shape factor A/A$_c$	0.56
Time of conc.	47 h
Mean annual precip.	785 mm
Mean annual runoff	325 x 10⁶ m³/s
Max. obs. flood peak	1 220 m³/s
Representative period	69 years
RMF K-value	4.6
Strahler basin order	4
Shreve magnitude	96
Drainage density	0.310 km/km²
Ruggedness number	0.254
Bifurcation ratio	2.10

Figure 3
Plan, long section and basin properties of the Klip River catchment (represented by gauge C1H002) in the eastern highveld area of South Africa (Petras and Du Plessis, 1987; Kovacs, 1988; Midgley et al., 1990 and Parak, 2003). Reference should be made to the **Appendix** for the definitions of these parameters and those summarised in Table A1.

numbers are all dependent on the scale of the map from which these parameters were extracted. More accurate measures can be made with the use of finer scaled maps, but this comes at the expense of greater effort and time requirements. Parak (2003) used uniform scaled maps from Midgley et al. (1990) showing river detail at 1:250 000 for the data extraction.

The criterion for choosing an appropriate model was based on the determination of the R^2 statistics through stepwise regressions. The first group of flow rates were plotted against catchment area to determine a regression equation and R^2 statistic in the calibration set. The regression equation was then used to generate flow rates of the second (validation) group from the independent variable and these estimates were plotted against the recorded ones of the same group to see if they mimicked each other. The degree of validation was based on the strength of the R^2 statistic. Subsequently, other landscape data were combined with catchment area to examine if they improved the strength of the relationship (based on the R^2 statistic) in calibration and validation. A conclusion was drawn based on the examination of the R^2 statistic in calibration and validation of the two groups of flood and landscape data.

In the original study Parak (2003) examined the relationship between the flood peaks and the various candidate landscape parameters. The model, given by Eq. (8), was selected after examining the literature on geomorphological estimates of floods and carefully plotting pairs of variables. A power-law relationship was selected and various groupings of "independent" variables were included in the regression equation, which was the logarithm of Eq. (8), shown as Eq. (9). The model selection process was performed by fitting the model to a calibration set and checking the fit for a validation set. The most suitable formulation was a power relationship of the form:

$$Q_{20} = aA^b X^c Y^d .... \qquad (8)$$

where:
a, b, c and d are parameters to be regressed from the data and A, X and Y are landscape quantities.

The formulation for regression was to take logarithms of Eq. (8) and regress using the linearised model:

$$\log(Q_{20}) = \log(a) + b.\log(A) + c.\log(X) + d.\log(Y) \qquad (9)$$

Figure 4 shows the calibration of the empirical equation defining the 20-year flood as a function of catchment area. The R^2 statistic from this model implies a strong relationship (0.856) and good fit. When this empirical model is tested against the reserved data of the second group in validation (Fig. 5), the fit is evidently poor, producing a moderate R^2 statistic of 0.538. When further landscape data are added to catchment area in the hope of improving the fit of the empirical models, the results are no better. The addition of landscape data as independent variables in the prediction of floods did not improve flood prediction and it seemed as though the best model of floods and landscape is simply area based. These results
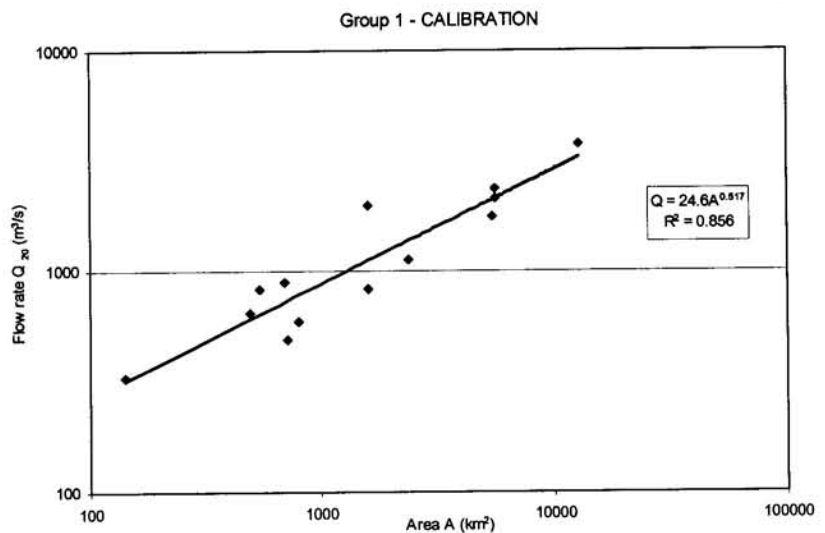

Figure 4
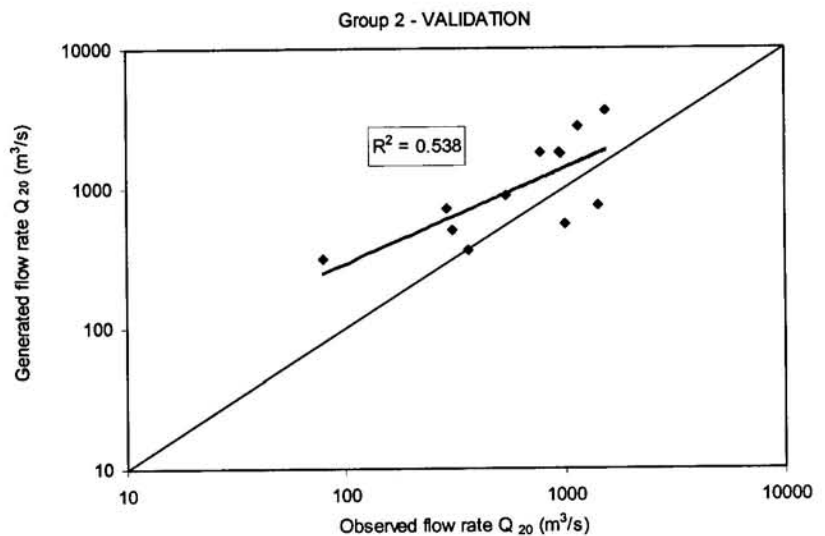*Q_{20} vs. area - calibration set for 13 catchments*

Figure 5
Generated flow rates (based on the regression of Fig. 4) vs. observed flow rates for the 20-year return period – validation set

are based on the negligible difference of the R^2 statistic in calibration and validation when additional landscape data are added to the catchment area, implying no significant additional prediction input from these parameters. The results are summarised in Table 2.

What is evident from the table is that the one group of data is quite different from the other; Group 1 is stronger in calibration and validation than Group 2. The grouping was a random choice process and this result is probably due to the small sizes of the groups (respectively 13 and 12 stations). Larger data sets are probably required to eliminate the effects of outliers in the samples. Table 2 also shows that catchment area on its own is a sufficiently good predictor of floods and the addition of landscape data does not improve this by much. This observation is based on the values of R^2 for the validation group. Besides *Area*, when *Relief* and then *MAP* are included, the results are best for validation using Group 2 (0.556 and 0.553 respectively compared to 0.538 for *Area* only). Conversely, when Group 1 is used for validation, *Area* alone has

the best R^2 (0.856) followed by the inclusion of *Drainage Density* (0.784) and *MAP* (0.770). *Drainage Density* is an area surrogate, so the apparent strength in validation might be due to this fact, whereas MAP is independent of area.

Ultimately, designers require an efficient flood formula and the acquisition of landscape data is not easy nor does it seem to provide much help to use a more complicated formula. Thus the use of the RMF (area-based) empirical equations seems justified. However, since this is only a preliminary review, further investigations into the role of landscape in affecting a flood regime is required to help with the understanding of this phenomenon.

Return period of the RMF

The RMF method of flood computation was applied to 57 catchments, where both annual flood peak data were available and where the regional *K*-values were known from Kovacs (1988). The floods were estimated for Regions 4.6, 5 and 5.2, which nearly cover the entire country (the remaining regions have a small number of recorded floods in their database). The floods were modelled from historical records using the GEV distribution and were plotted coaxially with those that were determined from the RMF, corresponding to the same regions and catchments, against catchment area as the independent variable as shown in Fig. 6. Since the return periods of the modelled floods were known, the return period of the RMF could then be estimated. For this reason, the 50-, 100- and 200-year floods were determined from the statistically modelled floods to determine the return period of the RMF. The results are shown in Figs. 6, 7 and 8 which cover Regions 4.6, 5 and 5.2 respectively.

The 200-, 100- and 50-year observed flood magnitudes are represented by the thin solid line, the thin dashed line and the thin dotted line respectively. These magnitudes were determined from the statistical analysis of observed flood data for the individual catchments using the GEV distribution; a subset of the full data set (used for comparison with landscape analyses) appears in Part 3 of Table A1 in the **Appendix**. The RMF estimates were then determined from Kovacs (1988) using the Francou-Rodier equation and Kovacs' regional *K*-values for the corresponding catchments. These are represented by the thick solid lines in Figs. 6, 7 and 8. The 200-, 100- 50-year flood estimates are plotted coaxially with the RMF estimates for the corresponding

			Group 1: calib. Group 2: valid.	Group 2: calib. Group 1: valid.
			R^2	
Q_{20} vs..	Area	Calibration:	0.856	0.538
		Validation:	0.538 **(3)**	0.856 **(1)**
	Area and slope	Calibration:	0.869	0.566
		Validation:	0.534	0.724
	Area and MAP	Calibration:	0.886	0.552
		Validation:	0.507	0.770 **(3)**
	Area and drainage density	Calibration:	0.872	0.538
		Validation:	0.531	0.784 **(2)**
	Area and relief	Calibration:	0.875	0.644
		Validation:	0.556 **(1)**	0.628
	Area and ruggedness number	Calibration:	0.880	0.593
		Validation:	0.552	0.659
	Area, slope and MAP	Calibration:	0.896	0.635
		Validation:	0.502	0.393
	Area, ruggedness number and MAP	Calibration:	0.920	0.596
		Validation:	0.523	0.640
	Area, drainage density and MAP	Calibration:	0.887	0.571
		Validation:	0.488	0.534
	Area, relief and MAP	Calibration:	0.890	0.647
		Validation:	0.553 **(2)**	0.605

**TABLE 2
Results of the step-wise regression**

(1), **(2)** and **(3)**: These numbers in parentheses flag the "best" (based on the R^2 statistic) fit to the validation data.
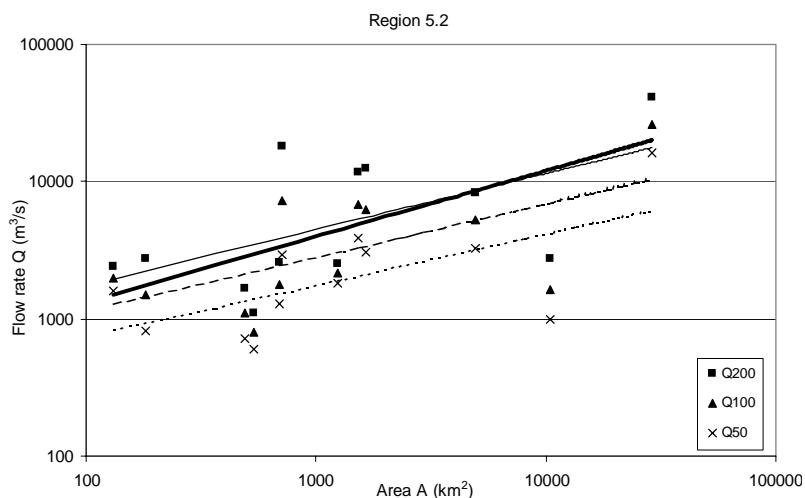

Figure 6
Determination of the return period associated with the RMF for Region 5.2. The bold line is the RMF estimate explained by the Francou-Rodier equation and the thin lines (dotted, dashed and solid) are trend-lines fitted to the 50-, 100-, and 200-year floods estimated from the recorded flows (points) in region 5.2, using the GEV distribution.

catchments in regions 4.6, 5 and 5.2 to examine if a return period can be associated with the RMF.

From Figs. 6, 7 and 8 it is clear that the RMF, when compared to the 50-, 100- and 200-year floods, is closest to the 200-year flood. The trend-line equations, summarised in Table 3, make for interesting reading. The slopes of the corresponding curves and

trend-lines are not equal. In fact the 200-, 100- and 50-year trend-line slopes are slightly flatter than the RMF curves for all the floods in all the regions except one (Q_{50} for Region 4.6); nevertheless, the correspondence is good and provides a starting point for further research to explain the similarities, even if the R^2 values are quite low.

In all three cases, the RMF line and the 200-year trend-line estimated from the fitted GEV distributions almost lie on top of each other and are very nearly parallel, mindful of the contents of Table 3. However, it must be admitted that the trend-lines for the 200-, 100- and 50-year floods have a poor fit and a fair amount of scatter can be observed. In all of Figs. 6, 7 and 8, serious outliers are evident for catchments with areas of about 1 000km², where the RMF is more likely to be associated with an event of return period of 100 years. On further investigation, it turns out that the problem 200-, 100-, and 50-year GEV flood estimate outliers are skewed by excessively large flood peaks that were observed in a relatively short record (between 20 and 30 years) for those catchments. It is expected that with more data, the effect of the outliers will be diminished. The result is that the plot for Region 5 (Fig. 7) is likely to be closer to the truth than Figs. 6 and 8 as it contains more data. In Fig. 7 the difference between the 200-year flood estimate and the RMF estimate is greater than the other two, but this difference is not excessive and the 200-year estimate and the RMF estimate are of the same magnitude. Based on these results, it is the opinion of the authors that it would be reasonable to assume the RMF to have a return period of the order of 200 years.

Q_T/Q_2 ratios in scaling floods

The Q_T/Q_2 ratios given in Table 1 are based on the premise that flood peaks in South Africa are log-normally distributed with a coefficient of variation equal to 1.3. To determine whether this average relationship is applicable in a design context, each of the maximum flood peaks recorded for the 130 catchments was assigned a return period based on the Weibull Plotting Position and then scaled down to 10- and 50-year events using the Q_T/Q_2 ratios. These flow rates were then plotted against the 10- and 50-year flow rates determined from modelling the historical records by the GEV distribution, to examine if the scaled flow rates are of the same order as the historical ones.

To simplify the analysis, a simple power-law relationship was sought between the ratios Q_T/Q_2 and T shown in Table 1. This short-cut approach was used instead of computing the percentage points of the lognormal distribution and to see its validity, the relationship is shown in Fig. 9, where a power law curve ($Q_T/Q_2 =$
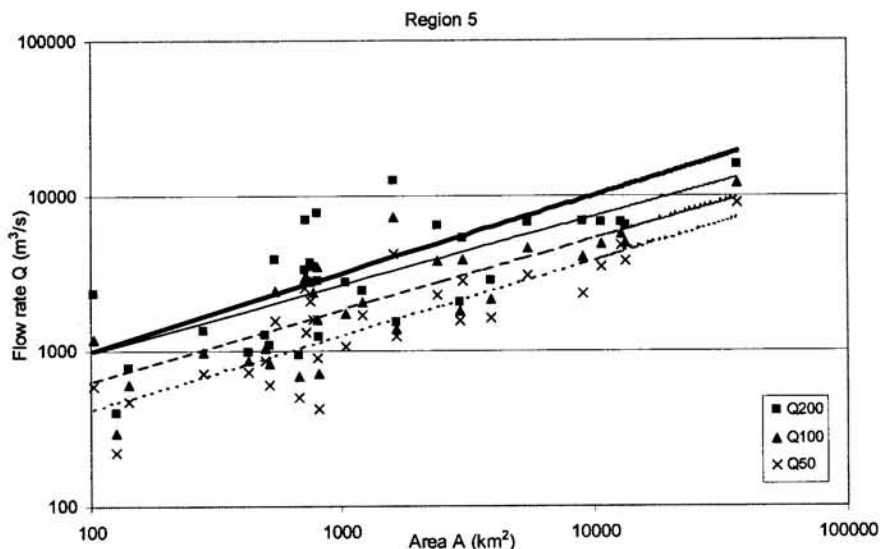

Figure 7
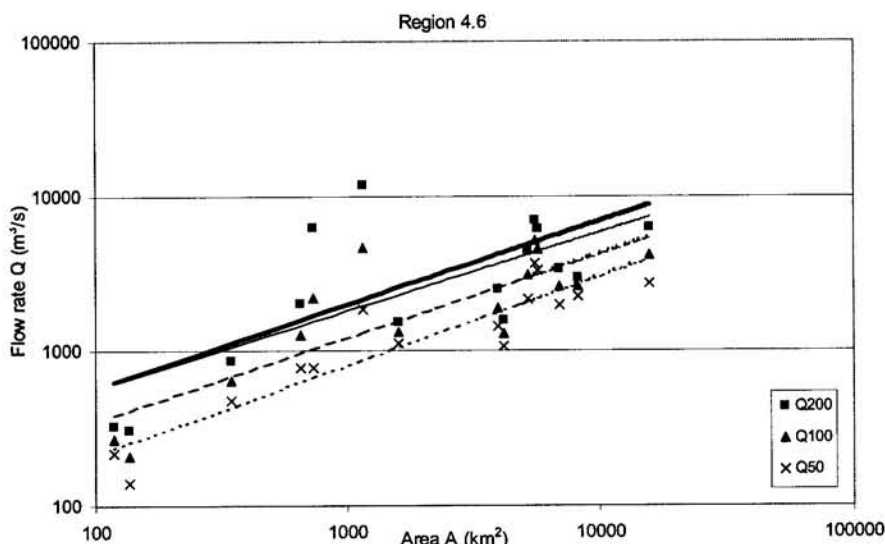*Determination of the return period associated with the RMF for Region 5 (description as per Fig. 6)*

Figure 8
Determination of the return period associated with the RMF for Region 4.6 (description as per Fig. 6)

	Q_{RMF}	Q_{200}	Q_{100}	Q_{50}
TABLE 3 Summary of the trend-lines from Figs. 6, 7 and 8				
Region 5.2	$145A^{0.48}$	$269A^{0.41}$	$191A^{0.39}$	$134A^{0.38}$
Region 5	$100A^{0.5}$	$129A^{0.44}$	$77A^{0.46}$	$45A^{0.48}$
Region 4.6	$48A^{0.54}$	$55A^{0.51}$	$29A^{0.54}$	$20A^{0.58}$

$1.28T^{0.456}$) was fitted in the 10- to 100-year interval. This interval was used firstly because it fitted better ($R^2 = 0.996$) than a power-law curve through all the points and secondly, because all the recorded maxima were observed to lie in this interval, i.e. between 10 and 100 years. The plot of these scaled floods with the corre-