

# Performance of regional flood frequency analysis methods in KwaZulu-Natal, South Africa

JC Smithers<sup>1,2\*</sup>, J Streatfield<sup>2</sup>, RP Gray<sup>2</sup> and EGM Oakes<sup>2</sup>

<sup>1</sup>Bioresources Engineering, School of Engineering, University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209, South Africa

<sup>2</sup>Jeffares & Green (Pty) Ltd, 6 Pin Oak Avenue, Hilton, Pietermaritzburg, 3201, South Africa

## ABSTRACT

Estimates of design floods are required for the design of hydraulic structures and to quantify the risk of failure of the structures. Many international studies have shown that design floods estimated using a regionalised method result in more reliable estimates of design floods than values computed from a single site or from other methods. A number of regional flood frequency analysis (RFFA) methods have been developed, which cover all or parts of South Africa. These include methods developed by Van Bladeren (1993), Mkhanda et al. (2000), Gørgens (2007) and Haile (2011). The performance of these methods has been assessed at selected flow-gauging sites in the province of KwaZulu-Natal (KZN), South Africa. It is recommended that the limitations of available flow records to estimate extreme flow events need to be urgently addressed. From the results for KZN the JPV method, with a regionalised GEV distribution with the veld zone regionalisation, generally gave the best performance when compared to design floods estimated from the annual maximum series extracted from the observed data. It is recommended that the performance of the various RFFA methods needs to be assessed at a national scale and that a more detailed regionalisation be used in the development of an updated RFFA method for South Africa.

**Keywords:** regional flood frequency analysis, KwaZulu-Natal

## INTRODUCTION

The design of hydraulic structures (e.g. dams, flood attenuation structures, culverts) requires the estimation of a design flood which is the magnitude of the flood associated with a given probability of exceedance or return period in years. Practitioners in South Africa generally estimate design floods by performing a frequency analysis of gauged flow data at a given location, if flow data are available at the site of interest, or by using an event-based rainfall-runoff model, for example, using the rational method, unit hydrograph method, or Soil Conservation Services method adapted for conditions in South Africa (SCS-SA). However, limitations of event-based methods include the assumption that the exceedance probability of the flood event is the same as the exceedance probability of the rainfall event, i.e., the 100-year return period flood event is assumed to result from a 100-year return period rainfall event, and the antecedent soil moisture condition in the catchment prior to extreme rainfall events is not taken into account.

When observed flow data are available, design floods can be estimated by performing a frequency analysis of the data, which generally involves fitting probability distributions to the annual maximum series (AMS) extracted from the data. The selected probability distribution is assumed to represent the population of all extreme events from the site. Hence, the longer the period of record, the better the assumption that the selected distribution represents the distribution of the population of all extreme events at the site.

A limitation of using a single-site approach to flood frequency analysis is that relatively few gauging stations in South Africa have long record lengths (e.g. > 50 years) and

this limits the confidence in design floods estimated using data from a single site, particularly when using shorter record lengths and when estimating design values for longer return periods (e.g. 100 years). In addition, design floods generally need to be estimated at sites where observed flood data are not available and thus rainfall-based methods or regionalised methods need to be used to estimate design floods at ungauged sites.

Given the relatively short flow-record lengths generally available it is necessary to use data from similar and nearby locations to improve the reliability of design flood estimates (Stedinger et al. 1993). This approach is known as regional flood frequency analysis (RFFA) and utilises data from several sites to estimate the frequency distribution of floods at each site. As summarised by Smithers (2012), many studies have shown that RFFA will result in more accurate and consistent estimates than at-site analyses (e.g. Cordery and Pilgrim, 2000; Hosking and Wallis, 1997; Smithers and Schulze, 2000a; Smithers and Schulze, 2000b).

RFFA usually assumes that relatively homogenous flood regions can be identified where the frequency distributions of floods at different sites are similar after site-specific scaling. Generally, growth curves (ratio of design flood/index flood vs. return period), or regionalised scaled distribution parameters, are developed for each region. Regionalised relationships are then developed to estimate the index flood/scaling value (e.g. mean annual flood) at ungauged sites in a region. A critical aspect of RFFA is the identification of relatively homogenous flood response regions.

A number of RFFA studies have included parts of or the entire area of South Africa in their analyses. These include Van Bladeren (1993), Mkhanda et al. (2000), Gørgens (2007) and Haile (2011). None of these RFFA methods are currently widely used in practice to estimate design floods. The objective of this paper is to give a brief background to the methods and to compare the performance of these RFFA methods in the province of KwaZulu-Natal (KZN) in South Africa.

\* To whom all correspondence should be addressed.

e-mail: [Smithers@ukzn.ac.za](mailto:Smithers@ukzn.ac.za)

Received 16 September 2014; accepted in revised form 18 March 2015

## REGIONAL FLOOD FREQUENCY ANALYSIS METHODS FOR KWAZULU-NATAL

The following sections provide a brief background to the RFFA methods used in this study.

### Mkhandi Method

Mkhandi et al. (2000) performed a regional frequency analysis of annual maximum flood data using data from 407 screened stations in southern Africa. As shown in Fig. 1, 13 flood regions were identified in South Africa. The Pearson Type 3 (P3) distribution fitted by probability weighted moments (PWM) was found to be the best distribution in all regions in South Africa, with the exception of SAF13 where the Log-Pearson Type 3 (LP3) distribution fitted by the Method of Moments (MM) was used.

The Mean Annual Flood (*MAF*) was used as an index flood to scale the data. Equation 1 was used to estimate the *MAF* at ungauged sites (Mkhandi et al., 2000).

$$MAF = CONSTANT \times AREA^{EXPONENT} \quad (1)$$

where: *MAF* = mean annual flood ( $m^3 \cdot s^{-1}$ ), *CONSTANT* = regionalised parameter, *AREA* = catchment area ( $km^2$ ), and *EXPONENT* = regionalised parameter.

### Van Bladeren Method

Van Bladeren (1993) derived growth curves using both continuously recorded data and discharges derived from historical flood information from the Natal and Transkei regions. Design floods were estimated using the General Extreme Value (GEV) distribution fitted by PWM. The *MAF* was used as the index flood in the derivation of the growth curves and regionalised regressions were derived to estimate the *MAF* as a function

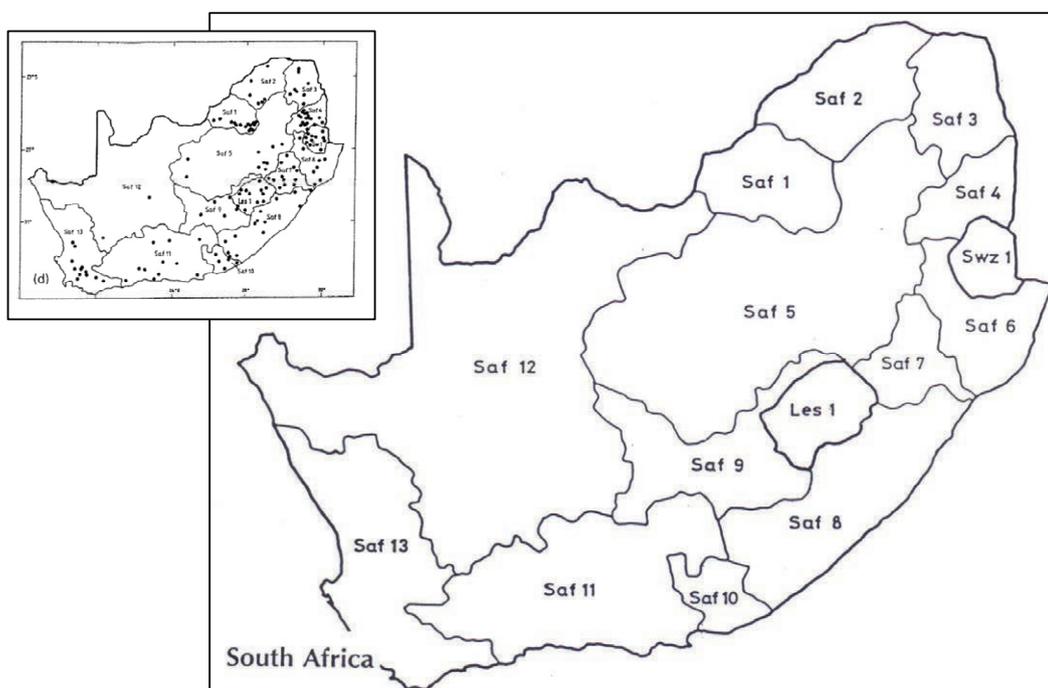
of catchment area. Regionalisation was initially based on the Regional Maximum Flood (RMF) regions identified by Kovács (1988) and was further refined within the RMF regions based on the skewness of the data. The method developed by Van Bladeren (1993) is applicable for RMF Regions 5.0 to 5.6 in Natal and Transkei.

### Haile Method

A RFFA in southern Africa was undertaken by Haile (2011) who used data from 459 gauging stations in 5 countries (Namibia, Malawi, Zambia, Zimbabwe and South Africa). After screening of the data, only 122 stations were included in further analyses (Fig. 2) with 92 stations from South Africa, of which 8 stations were used for independent testing of the method. Nine homogenous regions were identified, with 5 of these regions in South Africa, as shown in Fig. 3. The generalised Pareto (GPA), Pearson Type 3 (P3), three-parameter log-normal (LN3) and the GEV distributions were found to be suitable to model the distributions of the AMS of floods in southern African catchments. The median of the AMS (*MEF*) was used as the index to scale the values. From independent assessment of design floods estimated in the 9 regions using the regionalised flood frequency relationships, it was concluded that the regional approach was satisfactory (Haile, 2011). Both linear and exponential relationships were developed to estimate the *MEF* as a function of catchment area. In some regions where the negative constant in the linear relationship resulted in a negative *MEF*, exponential relationships were developed in this study to estimate the *MEF* from catchment area using information from Haile (2011).

### JPV Method

As part of the development of the Joint Peak-Volume (JPV) methodology, Görgens (2007) developed a regionalised index flood approach to design flood estimation for South Africa.



**Figure 1**  
Flood regions and stations used in the analysis (Mkhandi, Kachroo and Gunasekara 2000)

Data from 74 flow gauging stations and inflows to dams were used in the regionalisation, with the distribution of stations as shown in Fig. 4. The means of the peak discharges and volumes were used as index values to standardise the values.

**Regionalisation**

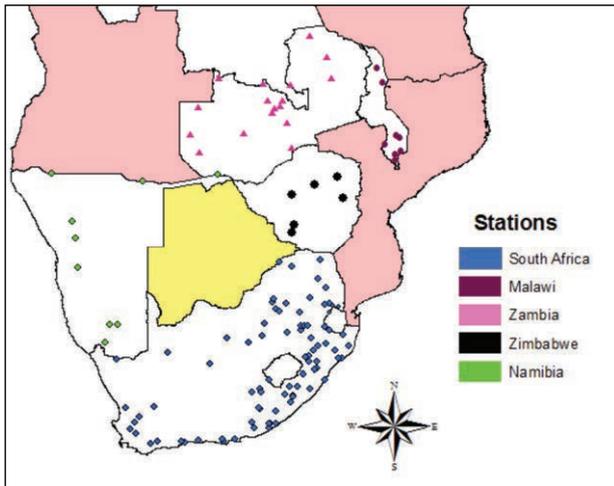
Görgens (2007) used both a fixed pool group, referred to as ‘wide’ pooling, and adjustable pooling groups, referred to as ‘narrow’ pooling. For the fixed pool grouping, Görgens (2007) developed growth curves for both the well-established K-regions (Kovacs 1988) and the veld type zones (HRU 1972) in his regionalisation. As shown in Fig. 5, the K-regions were

grouped into 3 regions: (i) high-K ( $K > 5$ ), (ii) mid-K ( $K=5$ ), and (iii) low-K ( $K < 5$ ).

As shown in Fig. 6, the veld type zones were grouped into 3 groups (Görgens 2007) as: (i) Group A (Veld Type Zone 2), (ii) Group B (Veld Type Zones 4, 5, 6, 7), and (iii) Group C (Veld Type Zones 1, 3, 8, 9).

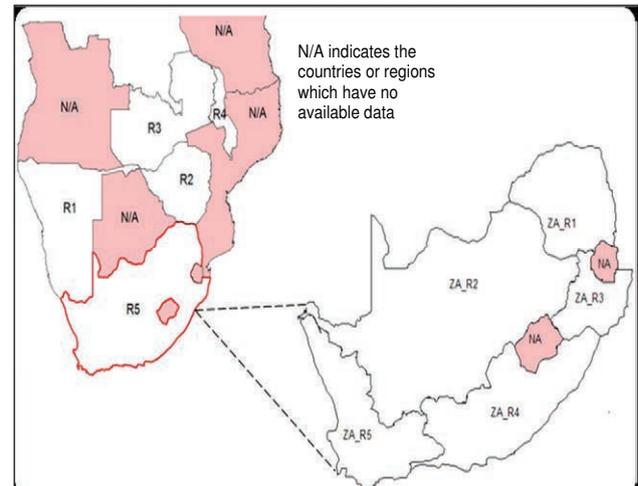
**Pooling of statistical parameters**

Pooled values for the coefficient of skewness ( $g$ ) and coefficient of variation ( $CV$ ) were weighted according to the record length and the inverse of a similarity distance ( $Dist_{ij}$ ), computed using Eq. 2 (Görgens, 2007).



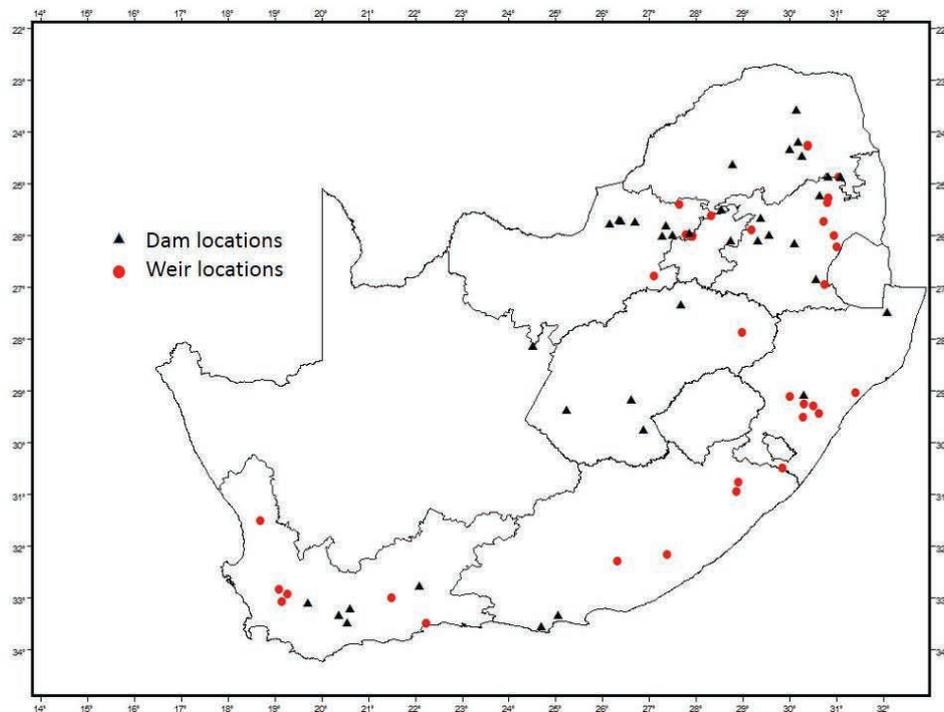
**Figure 2**

Stations used in the RFFA (Haile, 2011)



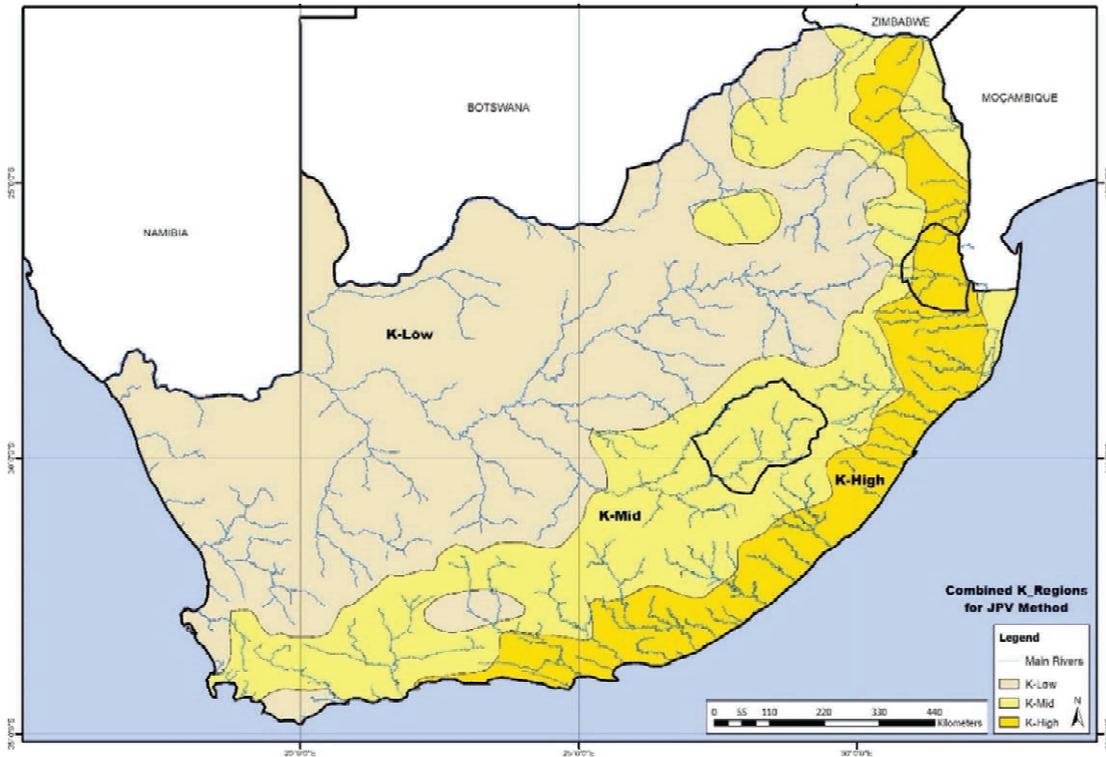
**Figure 3**

Delineation of southern Africa catchments into hydrologically homogenous regions (Haile, 2011)

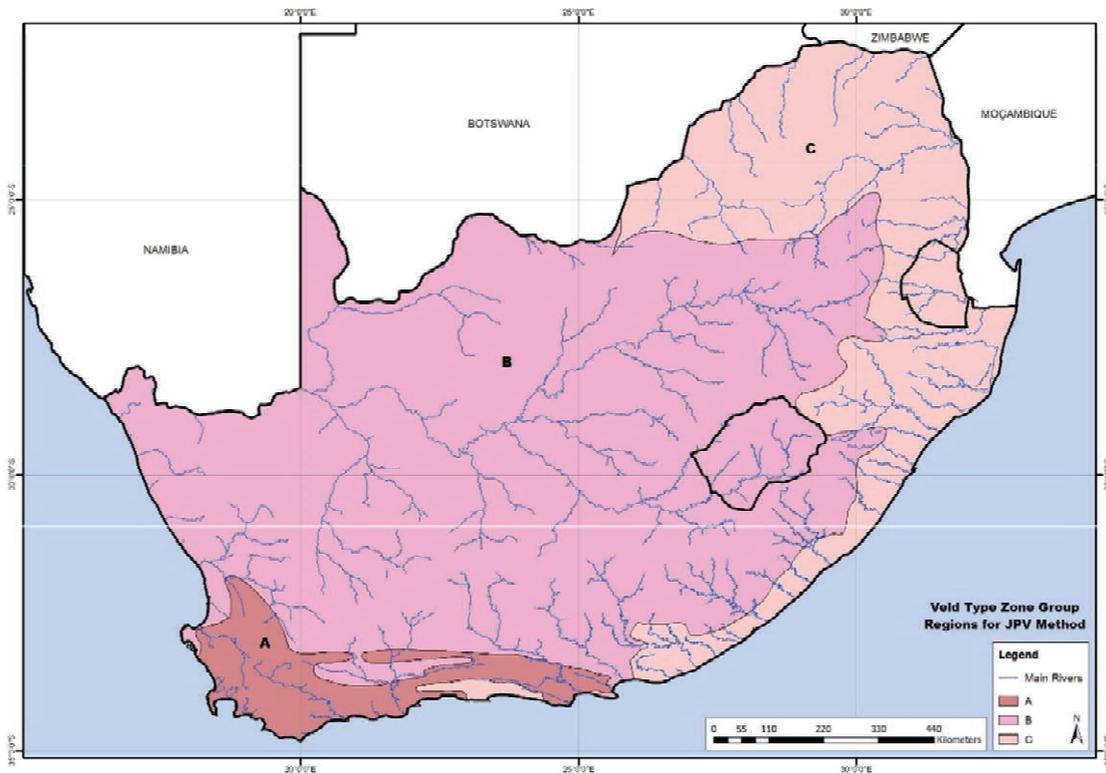


**Figure 4**

Location of stations used by Görgens (2007) in regionalisation



**Figure 5**  
K-region pooling group boundaries (Görgens, 2007)



**Figure 6**  
Veld type zone pooling group boundaries (Görgens, 2007)

$$Dist_{ij} = \sqrt{0.5 \frac{[\ln(Area_i) - \ln(Area_j)]^2}{\sigma_{\ln(Area)}} + \frac{[\ln(MAR90_i) - \ln(MAR90_j)]^2}{\sigma_{\ln(MAR90)}} + \frac{[S_i - S_j]^2}{\sigma_S}} \quad (2)$$

where:

$Dist_{ij}$  = similarity distance measure between Stations  $i$  and  $j$

$Area_i$  = area for Catchment  $i$  (km<sup>2</sup>)

$S_i$  = slope for Catchment  $i$

$MAR90$  = mean annual runoff (mm), determined from the WR90 study (Midgley et al., 1994)  $\sigma$  = standard deviation of descriptors of catchments in the pooling group

### Estimation of the index value at ungauged sites

Linear regression relationships were developed by G6rgens (2007) to estimate the index flood, as shown in Eq. 3. The catchment descriptors ( $Des_i$ ) included catchment area, equal area catchment slope,  $MAR90$ , and an index of the veld type zone or K-region.

$$\ln(\text{Index Flood Peak}) = B_0 + B_1 \ln(Des_1) + B_2 \ln(Des_2) + B_3 \ln(Des_3) + \dots \quad (3)$$

### DATA USED IN THE STUDY

Stream flow-gauges located in KZN were selected for inclusion in the analysis based on the attributes of the gauges. The criteria used were length of record, with gauges included for record lengths > 20 years, start and end date of flow record, the percentage of values in the AMS where the recorded stage exceeded the limits of the rating curve for the flow-gauging station, the number of values in the AMS which were flagged as having missing data during the year, and the period of the year when the missing data occurred.

After the initial selection of gauges, additional gauges with 15–20 years of record were investigated for inclusion in the analysis in areas which did not have any gauges included in the initial selection. The location of all of the gauges in KZN is displayed in Fig. 7. The distribution of the length of record of the selected gauges used in the analysis is shown in Fig. 8. The record lengths ranged from 13 to 83 years with a median value of 40 years.

The reliability of design values estimated from the observed data is dependent on the quality of the observed data and length of available record. As shown in Fig. 9, 67.4% of the gauges in KZN had a single unique maximum value in the AMS, whereas 15.7% of the gauges had the same maximum value in more than 20% of the years, which is an indication that the rating table used to convert the observed river stage into discharge had been exceeded. For the selected gauges used in this study, a single unique maximum value in the AMS was found at 87.8% of the gauges, and 12.2% of the sites had up to 15% of the years with the same maximum value. In these cases, the exceeded values were treated as missing data.

### RESULTS

This section contains the results from the application of the JPV, Haille, Van Bladeren and Mkhandi RFFA methods at the 41 selected flow-gauging sites and a comparison of the

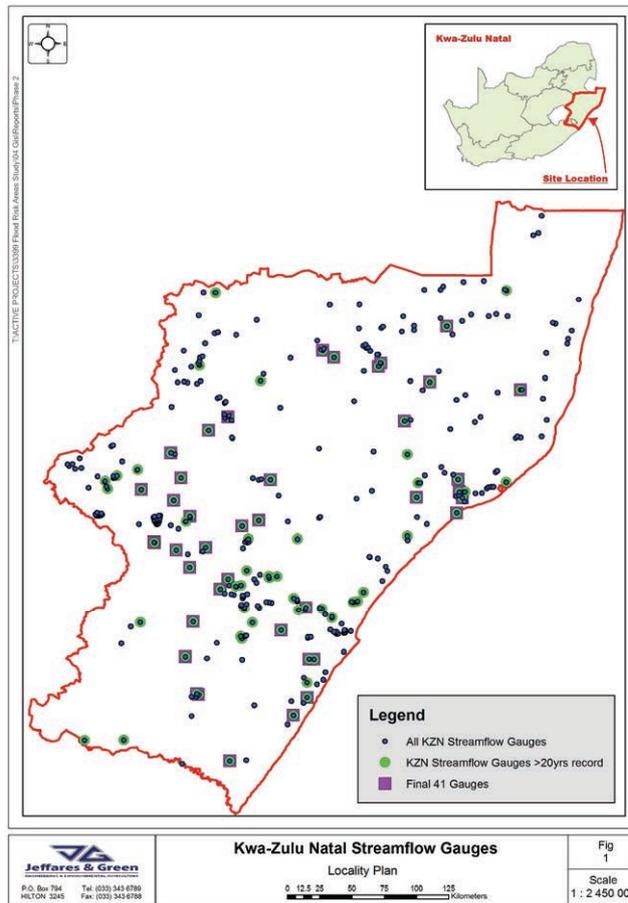


Figure 7  
Locations of all of the stream flow gauges in KZN

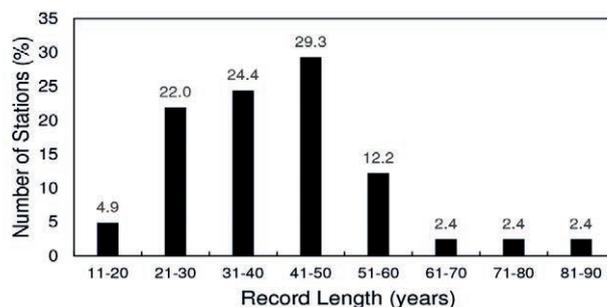


Figure 8  
Distribution of length of record of final selected gauges

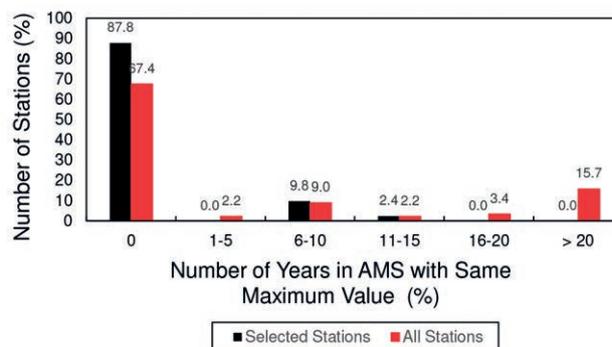


Figure 9  
Rating table exceedance at flow gauging weirs in KZN

estimated design floods to the design floods computed from the observed flow data at the sites. Alexander (1990, 2001) recommended the use of the LP3 probability distribution for design flood estimation in South Africa, while Gørgens (2007) used both the LP3 and GEV distributions and, according to Van der Spuy and Rademeyer (2010), both distributions are applicable in South Africa. Hence both the LP3 and GEV distributions, fitted by L-moments (Hosking, 1990; Hosking and Wallis, 1990), were used to estimate the design floods based on the statistics of the AMS at each selected gauge.

For each RFFA method, site and distribution, a mean absolute relative error (MARE) was computed as shown in Eq. 4:

$$MARE_{M,D} = \frac{100}{41} \times \sum_{n=1}^{41} \sum_{T=1}^6 \frac{|E_{M,T} - O_{D,T}|}{O_{D,T}} \quad (4)$$

where:

$MARE_{M,D}$  = mean absolute relative error (%) for RFFA Method =  $M$  and probability distribution =  $D$  (LP3 or GEV) for all stations (41) used

$E_{M,T}$  = design flood estimated using RFFA Method =  $M$  and for return period =  $T$  (2, 5, 10, 20, 50 or 100 years)

$O_{D,T}$  = design flood estimated using observed AMS and probability distribution =  $D$  (LP3 or GEV) for return period =  $T$  (2, 5, 10, 20, 50 or 100 years)

The results from an analysis of the performance of the RFFA methods are shown in Table 1, which includes both the  $MARE_{M,D}$  values for each method and the average slope between the 2 to 100 year return period floods computed at each site using the selected method (Estimated) and estimated from the observed data at the site (Observed). While the Haile method resulted in the smallest  $MARE_{M,D}$  value, the average slope of the estimated vs. observed floods is considerably less than 1, indicating that the Haile method generally underestimates the floods computed from the observed data. This general underestimation by the Haile method is confirmed by the performance of the RFFA methods in estimating the 50-year return period floods shown in Fig. 10. The results in Fig. 10 also confirm the poor performance of the JPV method when the regionalised LP3 distribution is used. Based on the results in Table 1 and the typical results

shown in Fig. 10, the JPV method using the regionalised GEV distribution and veld-zone regionalisation performed the best of the RFFA methods considered in this study.

## DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

The availability of gauged flows for large events with extreme discharges remains a challenge in South Africa. For many flow-gauging stations investigated in this study, exceedance of the rating table by observed river stage was evident by values in the AMS which are constant and equal to the maximum rated discharge for the flow-gauging structure. Stations which have more than 20% of the values in the AMS which exceed the maximum rated discharge were excluded from the study and, for the retained stations, the years with values which exceed the maximum rated discharge were assumed to be missing. This analysis did not account for different rating tables covering different periods of record, which would result in the rating tables being exceeded more frequently than indicated in this analysis.

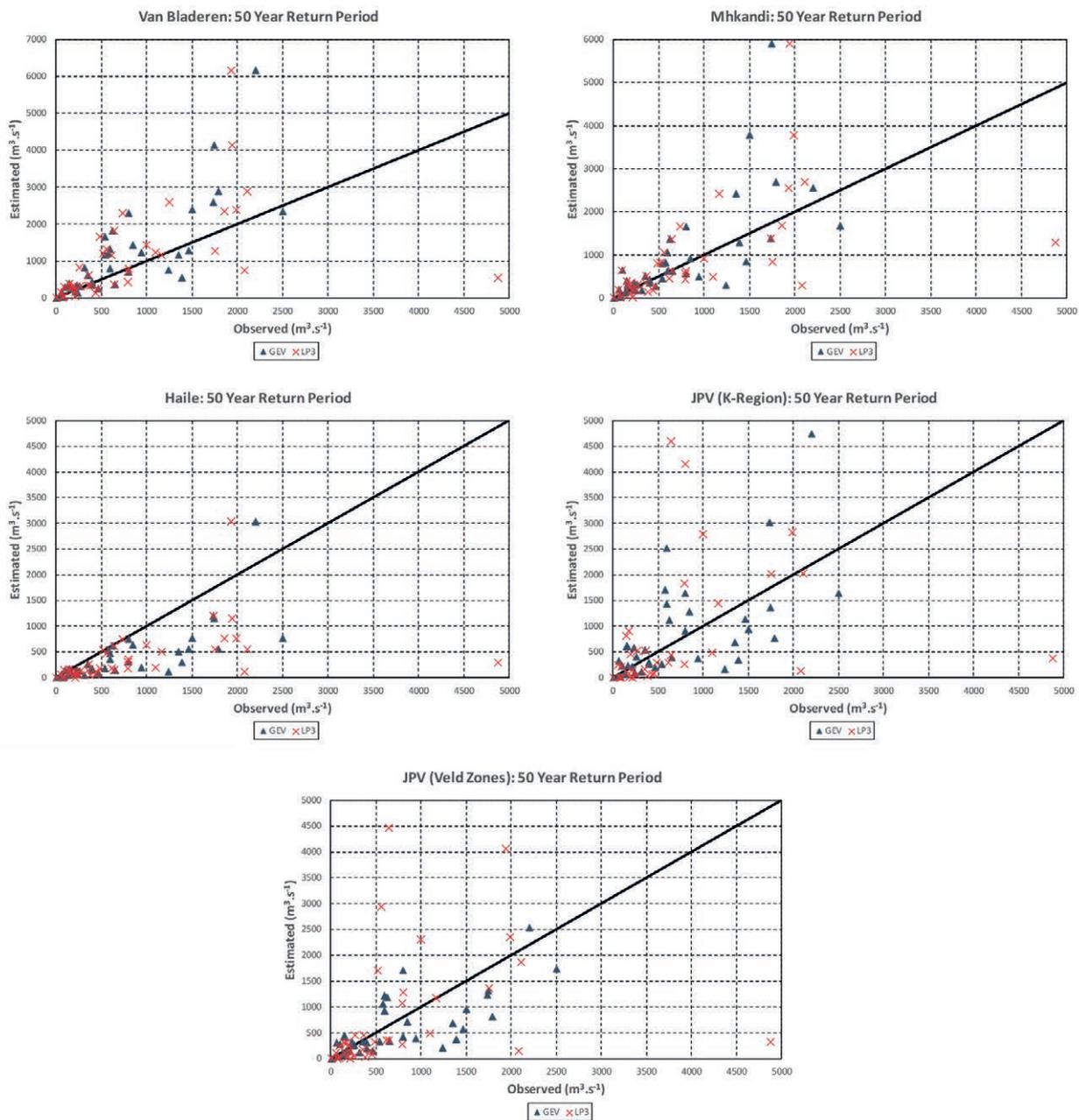
The frequency with which recorded flow stages exceed the maximum rated level needs to be quantified and the impact of not including these extreme events in the estimation of design floods in South Africa must be quantified. Methods to extend the rating tables, and thus provide an estimate of discharge for all observed stage levels, need to be urgently developed. It is expected that the estimation of discharge for all of the observed stage levels will impact both on the volume of runoff measured and the design floods estimated from the observed peak discharge data.

The results presented for the 50-year return period illustrate the differences in the design floods estimated from the observed data when using the GEV and LP3 distributions, particularly for longer return periods. Despite the quality screening of the stations included in the study, the design floods estimated at a few stations seem to not be consistent with other stations in the region. Hence, the need for consistent screening and checking of the flow data is required in order to identify reliable data records that can be used for design flood estimation.

A number of RFFA studies have been developed which include all of South Africa (Haile, Mkhanda and JPV methods), and regions of South Africa (Van Bladeren). Despite the advantages of a regional approach to design flood estimation, RFFA

**Table 1**  
**Performance of RFFA methods in KZN**

RFFA method	Distribution: Regionalisation method	$MARE_{M,D}$	Average slope of estimated vs. observed floods (2–100 year return periods)
Haile	GEV	51.8	0.48
Haile	LP3	55.1	0.64
JPV	GEV: Veld zones	64.1	1.00
Mkhanda	GEV	73.4	1.02
Mkhanda	LP3	76.9	1.40
JPV	GEV: K-regions	81.0	1.43
Van Bladeren	GEV	96.8	1.76
Van Bladeren	LP3	103.5	1.81
JPV	LP3: veld zones	117.2	8.50
JPV	LP3: K-regions	207.2	21.24



**Figure 10**  
Performance of RFFA methods for the 50-year return period event

methods are not widely used in South Africa. Of the RFFA methods assessed in KZN, the Haile method gave the best performance in terms of the MARE, but consistently underestimated the design floods computed from the observed data when using either the GEV or LP3 distribution. The poor performance of the JPV method with the regionalised LP3 distribution needs to be investigated. However, the JPV method with the regionalised GEV distribution generally performed well, with the veld zone regionalisation giving better results than the RMF K-region regionalisation. The results from this study are applicable only to KZN and the performance of the various RFFA methods at a national scale needs to be investigated.

Only the studies reported by Mkhandi et al. (2000), Haile (2011) and Görgens (2007) encompass the whole of South Africa. Ideally, regionalisation in a RFFA should be performed

using site characteristics as this enables independent testing of the regions for homogeneity using the at-site data, and the independent allocation of a site to a region based on the site characteristics. Discontinuities at regional boundaries need to be investigated and the alternative approach of transferring hydrological information from gauged to ungauged sites within a region should be evaluated.

Both Mkhandi et al. (2000) and Haile (2011) utilised the statistics of the at-site data with various homogeneity tests to identify homogenous flood regions in their study areas. Görgens (2007) did not update flood regions in South Africa and used both the RMF K-regions (Kovács, 1988) and the veld type zones (HRU, 1972) in his regionalisation, and it is recommended that a more detailed regionalisation should be used in the development of an updated RFFA method for South Africa.

## ACKNOWLEDGEMENTS

The results in this paper have been generated as part of a flood risk study being undertaken by Jeffares & Green for the Department of Human Settlements in KwaZulu-Natal and the funding for the project, and permission to publish the results, are gratefully acknowledged.

## REFERENCES

- ALEXANDER WJR (1990) *Flood Hydrology for Southern Africa*. SANCOLD, Pretoria.
- ALEXANDER WJR (2001) *Flood Risk Reduction Measures*. University of Pretoria, Pretoria.
- CORDERY I and PILGRIM DH (2000) The state of the art of flood prediction. In: Parker DJ (ed.) *Floods. Volume II*. Routledge, London, UK.
- GÖRGENS AHM (2007) Joint peak-volume (JPV) design flood hydrographs for South Africa. WRC Report No. 1420/3/07. Water Research Commission, Pretoria. 241 pp.
- HAILE AT (2011) Regional flood frequency analysis in Southern Africa. MSc Thesis, University of Oslo, Norway. 129 pp.
- HOSKING JRM (1990) L-moments: analysis and estimation of distribution using linear combinations of order statistics. *J. R. Stat. Soc.* **52** (1) 105–124.
- HOSKING JRM and WALLIS JR (1990) Regional flood frequency analysis using L-moments. Research Report RC 15658. IBM Research Division, Yorktown Heights, New York, USA. 12 pp.
- HOSKING JRM and WALLIS JR (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, UK.
- HRU (HYDROLOGICAL RESEARCH UNIT) (1972) Design flood determination in South Africa. Report No. 1/72. Hydrological Research Unit, Department of Civil Engineering, University of the Witwatersrand, Johannesburg.
- KOVACS ZP (1988) Regional maximum flood peaks in South Africa. Technical Report TR137. Department of Water Affairs, Pretoria, RSA.
- MIDGLEY DC, PITMAN WV and MIDDLETON BJ (1994) Surface water resources of South Africa 1990. WRC Report No. 298/1/94. Water Research Commission, Pretoria.
- MKHANDI SH, KACHROO RK and GUNASEKARA TAG (2000) Flood frequency analysis of southern Africa: II. Identification of regional distributions. *Hydrol. Sci. J.* **45** (3) 449–464.
- SMITHERS JC (2012) Review of methods for design flood estimation in South Africa. *Water SA* **38** (4) 633–646.
- SMITHERS JC and SCHULZE RE (2000a) Development and evaluation of techniques for estimating short duration design rainfall in South Africa. WRC Report No. 681/1/00. Water Research Commission, Pretoria. 356 pp.
- SMITHERS JC and SCHULZE RE (2000b) Long duration design rainfall estimates for South Africa. WRC Report No. 811/1/00. Water Research Commission, Pretoria. 69 pp.
- STEDINGER JR, VOGEL RM and FOUFOULA-GEORGIOU E (1993) Frequency analysis of extreme events. In: Maidment DR (ed.) *Handbook of Hydrology*. McGraw-Hill, New York.
- VAN BLADEREN D (1993) Application of historical flood data in flood frequency analysis, for the Natal and Transkei region. In: *Proceedings of Sixth South African National Hydrological Symposium*. Department of Agricultural Engineering, University of Natal. 359–366.
- VAN DER SPUY D and RADEMEYER PF (2010) Flood Frequency Estimation Methods as Applied in the Department of Water Affairs. Department of Water Affairs, Pretoria. 98 pp.