Assessing the performance of regional flood frequency analysis methods in South Africa

JJ Nathanael[1]*, JC Smithers[1, 2, 3, 4] and MJC Horan[1]

[1] Centre for Water Resources Research, University of KwaZulu-Natal, Pietermaritzburg, South Africa, Private Bag X01, Scottsville, 3209, South Africa
[2] Bioresources Engineering, School of Engineering, University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209, South Africa
[3] JG Afrika (Pty) Ltd, 6 Pin Oak Avenue, Hilton, Pietermaritzburg, 3201, South Africa
[4] National Centre for Engineering in Agriculture, University of Southern Queensland, Toowoomba, Australia

ABSTRACT

In engineering and flood hydrology, the estimation of a design flood associates the magnitude of a flood with a level of exceedance, or return period, for a given site. The use of a regional flood frequency analysis (RFFA) approach improves the accuracy and reliability of estimates of design floods. However, no RFFA method is currently widely used in South Africa, despite a number of RFFA studies having been undertaken in Africa and which include South Africa in their study areas. Hence, the performance of the current RFFA approaches needs to be assessed in order to determine the best approaches to use and to determine if a new RFFA approach needs to be developed for use in South Africa. Through a review of the relevant literature it was found that the Meigh et al. (1997) method, the Mkhandi et al. (2000) method, the Görgens (2007) Joint Peak-Volume (JPV) method and the Haile (2011) method are available for application in a nationwide study. The results of the study show that the Haile method generally performs better than the other RFFA methods; however, it also consistently under-estimates design floods. Due to the poor overall performance of the RFFA methods assessed, it is recommended that a new RFFA method be developed for application in design flood practice in South Africa.

Keywords: regional flood frequency analysis (RFFA), South Africa

INTRODUCTION

Flood disasters are considered to be the predominant and most frequently occurring natural hazard and are responsible for the most fatalities worldwide (Doocy et al., 2013). The loss of life and economic loss caused by floods can occur at both small and large scales (Hubbart and Jones, 2009). Economic losses resulting from flood events have increased globally from an average of US$ 7 billion per year in the 1980s to approximately US$ 24 billion per year between 2001 and 2011 (Kundzewicz et al., 2013). A media release by Shiceka (2011) for the Ministry of Cooperative Governance and Traditional Affairs in South Africa indicated that infrastructure damage as a result of flooding in the North West, Northern Cape, and KwaZulu-Natal Provinces in 2011 amounted to approximately R6 million, R50 million, and R300 million, respectively. With a growing population, increasing urbanisation and climate change, the risks that flood events pose are becoming more severe, which requires researchers to improve the accuracy and reliability of methods used for flood estimation (Wiltshire, 1986; Smithers and Schulze, 2000).

Design flood estimation (DFE) associates the magnitude of a flood with a probability of exceedance at a given site. Most of the methods currently used for design flood estimation in South Africa were developed in the late 1960s and early 1970s and are in need of updating, with more than 40 years of additional data currently available and with new approaches used internationally (Smithers, 2012; Van Vuuren et al., 2013). This has resulted in the initiation of a National Flood Studies Programme (NFSP) for

South Africa to update and modernise methods used for design flood estimation (Smithers et al., 2014).

One approach to design flood estimation which utilises observed data is Flood Frequency Analysis (FFA), which involves the interpretation of a past record of hydrological events in terms of the future probability of occurrence. This can be achieved through an at-site FFA or through a regional FFA (RFFA). For an at-site analysis, a record of observed flows is required and must be of adequate length and quality (Smithers, 2012). A record of sufficient length is often not available at the site being investigated and thus a RFFA approach (GREHYS, 1996; Viglione et al., 2007) or a rainfall-runoff method must be used to estimate the design flood. A number of studies have advocated a regional approach for obtaining more reliable design flood estimates (Wiltshire, 1986; Hosking and Wallis, 1997; Cordery and Pilgrim, 2000: Saf, 2008; Saf, 2009; Haile, 2011; Smithers, 2012). RFFA can further be divided into non-statistical RFFA methods, such as regional empirical or deterministic flood approaches and statistical RFFA methods. For this study, RFFA will refer only to statistical approaches and, therefore, any advantages, disadvantages or limitations in the application of RFFA methods in South Africa, will be referring exclusively to statistical RFFA methods.

RFFA involves determining homogeneous flood response regions and selecting a suitable frequency distribution for the region (Kachroo et al., 2000). There are a number of approaches that may be utilised in a RFFA. One such approach is known as the index flood method (Dalrymple, 1960), which involves the use of data from sites other than the site under investigation (Hosking and Wallis, 1997). This allows for data from more than one site to be utilised, creating the potential for more accurate estimates of flood quantiles (Hosking and Wallis, 1997). Other approaches may involve a regression model that

---

directly links design floods to catchment characteristics, e.g., a study by Mediero and Kjeldsen (2014), in the Ebro catchment in Spain. The application of a RFFA is necessary at ungauged sites or at sites where an inadequate length or poor quality of streamflow data is available (Leclerc and Ouarda, 2007). In addition, a RFFA can improve flood quantile estimates at gauged sites where the record length is insufficient (Australian Institution of Engineers, 1977). A regional approach enables a FFA of shorter records, through the determination of the shape of the parent distribution and the estimation of scale, to be achieved from data at the site of interest (Bobee and Rasmussen, 1995). Observed data can be pooled from a region that is homogeneous, allowing the estimation of the parameters for the selected distribution and subsequently achieving more robust quantile estimates (Kachroo et al., 2000). It is widely reported that the use of a RFFA approach improves the accuracy and reliability of estimates of design floods (Wiltshire, 1986; Hosking and Wallis, 1997; Saf, 2008; Saf, 2009; Haile, 2011; Smithers, 2012), even though some studies have highlighted the difficulties in correctly applying RFFA methods (Alexander, 1990; Faber, 2010).

Whilst there are many advantages to applying a RFFA in South Africa, there are also challenges to the application of RFFA methods. Alexander (1990) highlighted that it was difficult to apply RFFA in South Africa as it was difficult to identify hydrologically homogeneous regions and that generally there is spatial correlation between the stations used in the analysis as severe floods generally occur in southern Africa as a result of widespread storms covering large parts of a region.

A number of RFFA methods have been developed for application regions which include South Africa (Van Bladeren, 1993; Meigh et al., 1997; Mkhandi et al., 2000; Kjeldsen et al., 2001; Kjeldsen et al., 2002; Görgens, 2007; Haile, 2011) as well as for application in many other countries internationally, for example, Jingyi and Hall (2004) in the Gan-Ming River basin in China, Kizza et al. (2006) in northern Uganda, Saf (2009) in the west Mediterranean region of Turkey, Malekinezhad et al. (2011) in the Namak-Lake basin in Iran, Mediero and Kjeldsen

(2014) in the Ebro catchment in Spain and Wazneh et al. (2015) in northwest Italy.

Zaman and Rahman et al. (2012) conducted a study in the semi-arid and arid regions of Australia, which involved a comparison of the most up-to-date data in Australia with that of world data. The aim of this comparison was to highlight the differences in the hydrology of semi-arid and arid regions from that of other regions. One of the challenges identified by Zaman and Rahman et al. (2012) was the lack of data available in semi-arid and arid regions, as well as the lack of literature on the applications of RFFA methods in these regions. Variations in rainfall, streamflow, vegetation and channel morphology in semi-arid and arid regions make the hydrology of these areas complex. The lack of rainfall and streamflow data, as well as the poor distribution of rain and streamflow gauges in arid regions, adds to the difficulties of developing and applying RFFA methods in these regions.

No RFFA method is currently widely used in South Africa, despite a number of RFFA studies having been undertaken in southern Africa and which include South Africa in their study regions. Hence, the performance of the currently available RFFA approaches in South Africa needs to be assessed in order to determine the best approaches to use and to determine if a new RFFA approach needs to be developed for use in South Africa.

REGIONAL FLOOD FREQUENCY ANALYSIS METHODS

The following sections entail a summary of RFFA methods that include South Africa and which are assessed in this study. Approaches which have been developed for part of South Africa, for example, by Van Bladeren (1993) in the Eastern Cape and KwaZulu-Natal regions of South Africa and Kjeldsen et al. (2001, 2002) in the KwaZulu-Natal Province of South Africa, were not considered in this study as they did not cover the entire country.
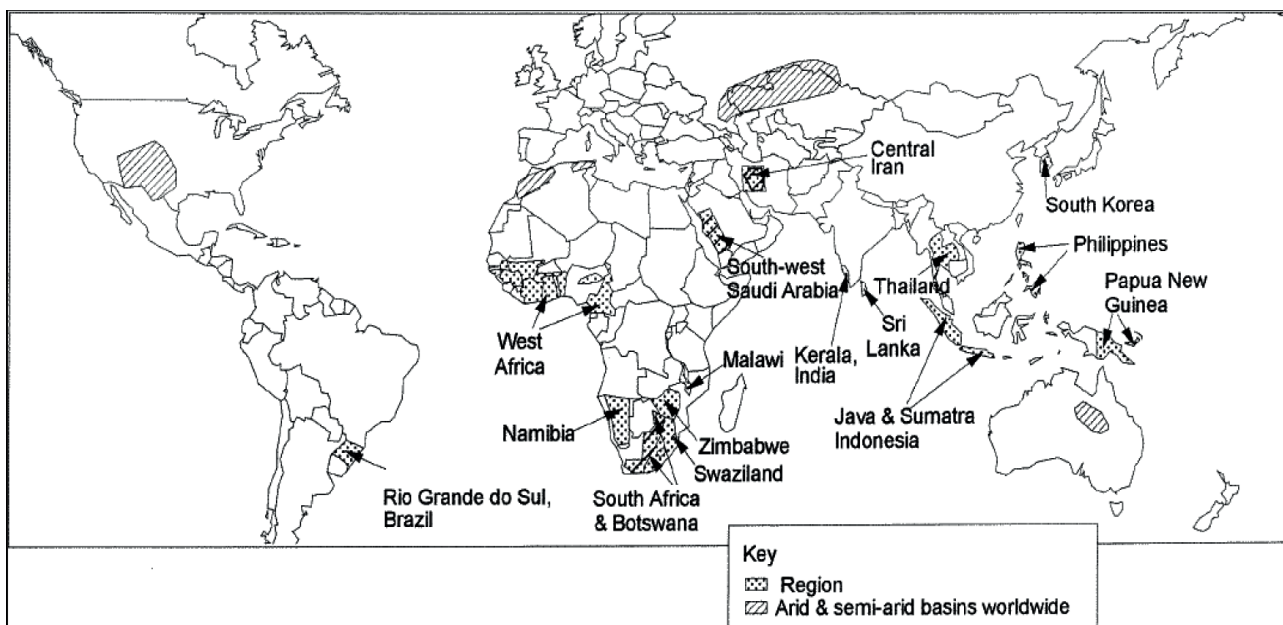

Figure 1
Regions analysed in the Meigh et al. (1997) study

Meigh method

A study involving a worldwide development of RFFA methods under different climatic conditions was conducted by Meigh et al. (1997). Figure 1 illustrates the regions that were analysed by Meigh et al. (1997). In South Africa and Botswana, datasets from 101 flow gauging stations were used in the study.

Meigh et al. (1997) used the mean annual flood (MAF) as an index value to scale the data. A multiple regression analysis was performed to determine the relationship between the MAF and the catchment characteristics, as shown in Eq. 1 for South Africa.

$$MAF = 6.97 \times AREA^{0.450} \qquad (1)$$

where:
MAF = mean annual flood (m$^3 \cdot$s^{-1})
AREA = catchment area (km^2)

Regional flood frequency growth curves were developed by Meigh et al. (1997), using the general extreme value (GEV) distribution fitted by probability weighted moments (PWM). These growth curves are illustrated in Fig. 2 and indicate that two growth curves are recommended for South Africa, dependent on the annual average rainfall (AAR).

Mkhandi method

A RFFA undertaken for Southern Africa was reported by Kachroo et al. (2000) and Mkhandi et al. (2000). Kachroo et al. (2000) details the delineation of homogeneous regions in Southern Africa, while Mkhandi et al. (2000) details the identification of the appropriate regional distributions. Mkhandi et al. (2000) utilised instantaneous maximum flood peak data for the analysis in South Africa. As illustrated in Fig. 3, 13 homogeneous regions were identified by Mkhandi et al. (2000) in South Africa, utilising 316 flow gauging stations. The log-Pearson Type 3 (LP3) distribution fitted by the method of moments (MOM) was used for Region SAF
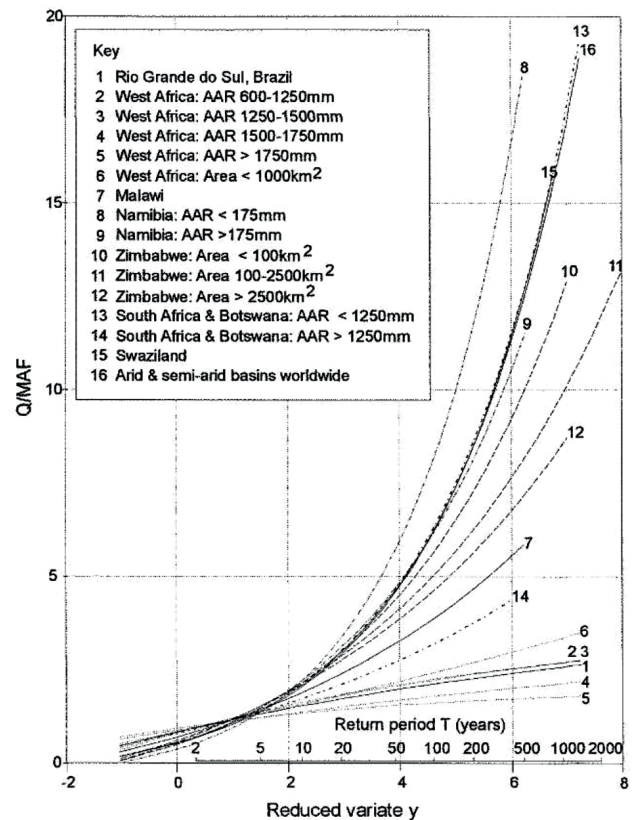

Figure 2
Regional flood frequency curves developed by Meigh et al. (1997)

13, while the remaining regions in South Africa utilised the Pearson Type 3 (P3) distribution fitted by PWM.

The data in the Mkhandi et al. (2000) study was scaled using the MAF as an index. In order to determine the MAF at an ungauged site, Eq. 2 was utilised:
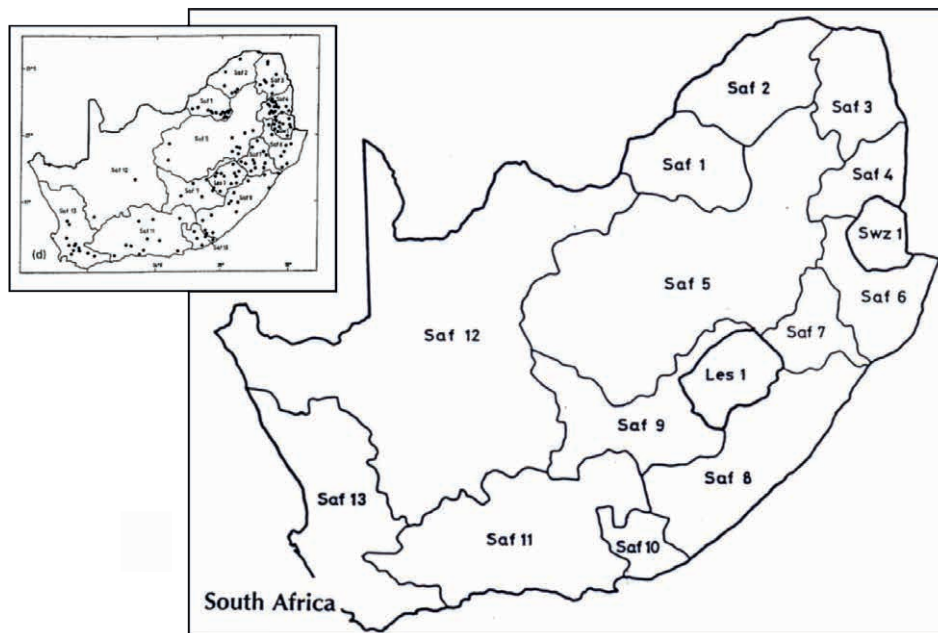

Figure 3
Maps of stations used and homogeneous regions identified for South Africa (after Mkhandi et al., 2000)

389

$$MAF = C \times AREA^E \qquad (2)$$

where:

MAF = mean annual flood (m3·s⁻¹)
C = regionalised parameter
AREA = catchment area (km²)
E = regionalised parameter

Joint peak volume (JPV) method

Görgens (2007) developed procedures that link flood volume exceedence with the flood peak magnitude on a regional scale. Data were extracted on a peak-over-threshold basis. The delineation of regions within South Africa utilised regions currently used in practice, i.e., the HRU (1972) 'veld zones' and the Kovacs (1988) 'regional maximum flood (RMF) K-regions'. The RMF K-regions developed by Kovacs (1988) are based on a maximum envelope approach, where the largest flood events were plotted against catchment area on a logarithmic scale. Flood envelopes were then sketched to include all of the data points. These RMF K-regions allow for the estimation of maximum observed discharges at ungauged sites within a particular flood envelope. Görgens (2007) grouped the veld zones into 3 categories. These regions were classified as: Groups A (Veld Zone 2), B (Veld Zones 4, 5, 6, 7) and C (Veld Zones 1, 3, 8, 9), as illustrated in Fig. 4. Similarly, the K-regions were grouped into 3 categories and were classified as: High-K ($K > 5$), Mid-K ($K=5$) and Low-K ($K < 5$), as illustrated in Fig. 5 (Görgens, 2007).

The index flood utilised by Görgens (2007) was the MAF and relationships between catchment descriptors (Des$_i$) and the index
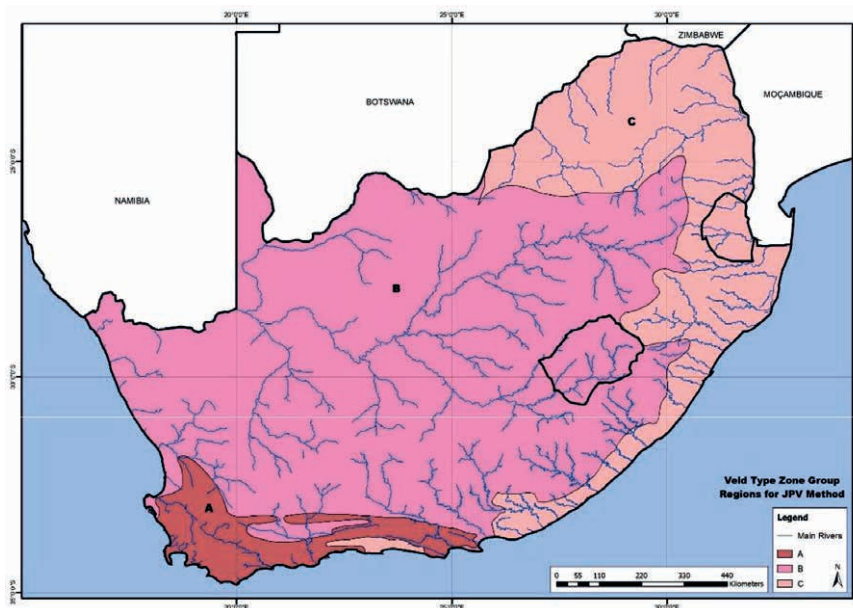

Figure 4
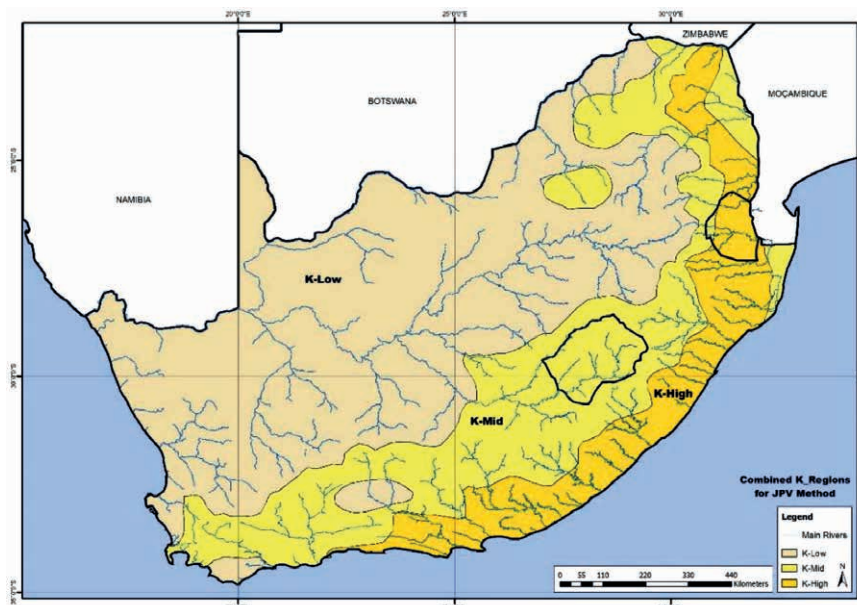*'Veld zone' pooling groups utilised in the JPV method (Görgens, 2007)*

Figure 5
K-region pooling groups utilised in the JPV method (Görgens, 2007)

flood were derived using a multiplicative model, as shown in Eq. 3. Predictor variables (Des$_i$) included catchment area (AREA) (km^2), average main channel slope (S) (%), mean annual runoff (MAR90) (mm) determined from the WR90 study (Midgley et al. 1994), and an index of the 'veld zone' or 'K-region'.

$$\ln(MAF) = B_o + B_1 \ln(AREA) + B_2 \ln(MAR90) + B_3 \ln(S) + B_4 \ln(\text{veld zone }/K\text{-region}) \qquad (3)$$

where: B_0, B_1, B_2, B_3, B_4 are linear regression coefficients.

In order to determine an estimation of a design flood peak, using pooled categories of veld zones and K-regions, the GEV and LP3 distributions were fitted to the at-site AMS using the method of moments. Eq. 4 was used to estimate the design floods (Görgens, 2007):

$$Q_T = MAF + K_{g,T}\sigma_Q \qquad (4)$$

where
Q_T = design flood peak (m3·s^{-1})
T = recurrence Interval (RI)
MAF = mean annual flood (m3·s^{-1})
g_Q = skewness
σ_Q = standard deviation
$K_{g,T}$ = frequency factor for selected probability distribution and skewness = g_Q

Pooled values for skewness have been determined as the weighted average of the individual values for skewness for the catchments in any pooling-group. The weighting is a function of similarity distance and record length.

Haile method

Haile (2011) conducted a RFFA study in Southern Africa. The analysis and regionalisation was performed using a combination of the index flood method and L-moments. In South Africa, 74 stations were analysed and 5 homogeneous flood regions were identified, as illustrated in Fig. 6.

Haile (2011) used the median of the AMS (MEF) as the index flood and developed relationships to estimate the MEF as a function of catchment area (AREA), as shown in Table 1.

The negative intercept in the regression equations in Table 1 produces a catchment area limit, below which a negative flood index is calculated. Therefore, new regression equations were determined, for application in this study. In order to resolve this issue, a relationship between the MEF and catchment area for stations analysed by Haile (2011) was plotted, as illustrated for Region ZA_R1 in Fig. 7. Figure 7 also illustrates the relationship between the MEF and catchment area for flow gauging stations analysed in this study, where the Haile method equations for Region ZA_R1 cannot be applied, i.e., catchment area < 28.3 km^2. These stations resulted in a negative index flood when using the Haile logarithmic equation, due to their areas being below 28.3 km^2. A power regression was fitted to these stations, as illustrated in Fig. 7.

TABLE 1
Regression models to predict the MEF from catchment area (Haile, 2011)

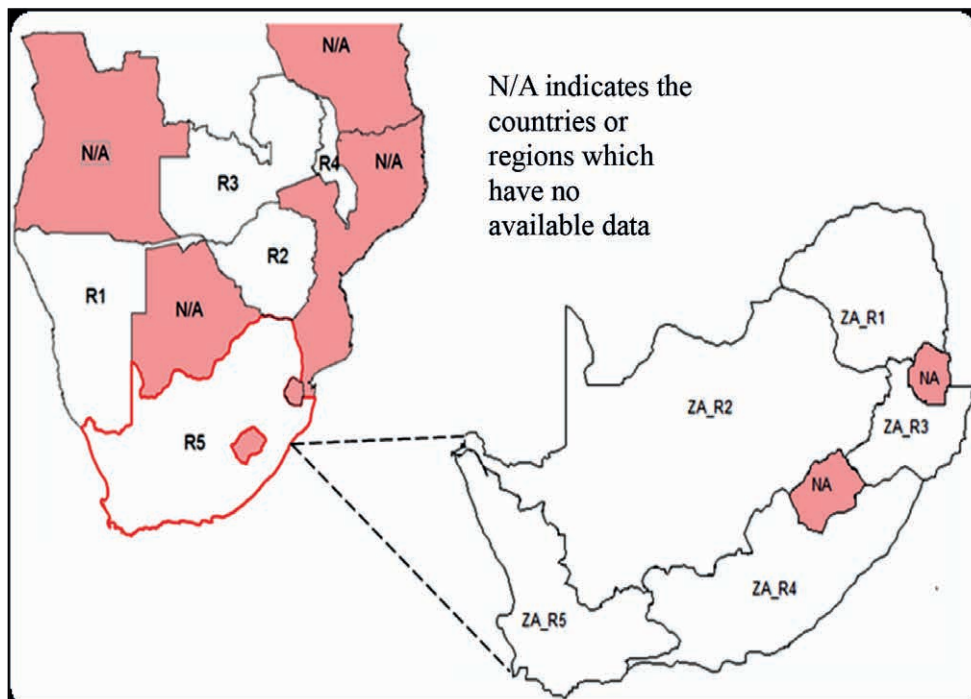| Region	Equation to estimate median of AMS	R^2	Area limits (km^2)
ZA_R1	MEF = 14.755 x ln(AREA)−49.338	0.37	28.3
ZA_R2	MEF = 52.664 x ln(AREA)−340.28	0.77	639.9
ZA_R3	MEF = 66.461 x ln(AREA)−395.91	0.52	386.5
ZA_R4	MEF = 0.6089(AREA)0.6639	0.59	
ZA_R5	MEF = 42.282 x ln(AREA)−187.1	0.89	83.5

Figure 6
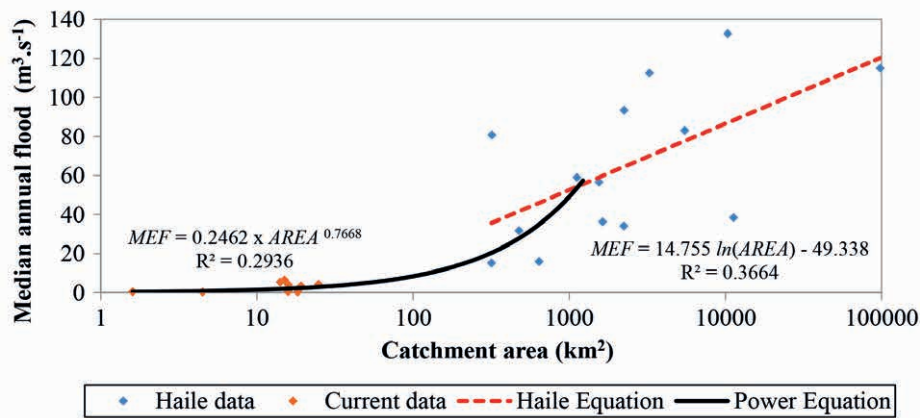*Map of homogeneous flood regions delineated by Haile (2011)*

Figure 7
Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R1 region

The MEF for these stations were determined using the same period of record that Haile used, which was from 1969 to 2008. The regression was then extrapolated until it intersected the Haile regression at a catchment area of 1 145 km². Therefore, in the ZA_R1 region, the fitted power regression has been applied to all stations with a catchment area that is less than or equal to 1 145 km² and the Haile logarithmic equation has been applied to all stations with catchment areas greater than 1 145 km².

As shown in Fig. 7, the logarithmic regression as developed by Haile (2011) could have been applied to a number of stations that were above the area limit of 28.3 km²; however, the power regression was used to allow for a smooth transition from the power regression to the logarithmic regression. Out of the 84 stations that were in Region ZA_R1, the MEF was estimated using the power regression rather than the logarithmic regression for 73 stations (86%), which could potentially impact on the performance of the Haile method as assessed in this study.

The procedures described above were also carried out for Regions ZA_R2, ZA_R3 and ZA_R5. The new regressions for each region are included in Table 2.

METHODOLOGY

This section details the method used to assess the performance of the Meigh, Mkhandi, JPV and Haile RFFA methods. Each regional approach was assessed by comparing design floods estimated by the method with those estimated from an at-site flood frequency of the observed flood data. For South Africa, the LP3 distribution is recommended by Alexander (1990; 2001) while Görgens (2007) used both the GEV and LP3 distributions and both distributions are advocated by Van der Spuy and Rademeyer (2010) for use in South Africa. Hence, in this study, design floods were initially estimated using both the LP3 and GEV distributions.

The methodology included the following:
- Collation of the annual maximum instantaneous flood peak series data for all flow gauging stations in South Africa
- Screening the data and selection of appropriate stations for use in the study
- Fitting of the GEV and LP3 distributions using L-moments to the annual maximum series (AMS) at the selected sites and estimation of design floods
- Derivation of relevant catchment parameters for each station

TABLE 2
Regional power regression equations for the estimation of the index flood, using the Haile (2011) method in South Africa

Region	Power regression	R2
ZA_R1	MEF = 0.2462(AREA)0.7668	0.29
ZA_R2	MEF = 0.2015(AREA)0.3737	0.15
ZA_R3	MEF = 0.8684(AREA)0.7624	0.63
ZA_R5	MEF = 0.5766(AREA)0.7928	0.41

- Application of the Meigh, Mkhandi, JPV and Haile RFFA methods at all selected stations
- Assessment of the performance of the methods by a comparison of the design floods estimated using the regional methods with design flood estimates determined when using an at-site flood frequency approach
- Recommendation of the best RFFA method for use in South Africa, or recommendation for the development of a revised RFFA approach

In this study AMS data were used in the at-site frequency analysis of observed data. AMS data were obtained from the Department of Water and Sanitation (DWS) for 1 458 stations across the entire country. In addition to these datasets, 89 synthesised dam inflow records were obtained from the DWS and incorporated into the study. The following selection criteria were used to select the stations to be analysed:
- It must be a river station, i.e., not an eye (natural spring), canal or pipeline.
- Record lengths must be greater than, or equal to, 20 years.
- The percentage of occurrence where the rating table is exceeded should not equal or exceed 20% of the values in the AMS.
- The station must not be located at a dam outlet or be significantly influenced by an upstream dam.

After all selection criteria had been met, the final number of stations to be analysed was reduced to 318 flow gauging stations and 89 dam inflow records, bringing the total number of stations analysed to 407. Figure 8 illustrates the distribution across South Africa of the original 1 458 stations received from the DWS, as well as the distribution of the final stations and dam inflow records selected for analysis in this study. The spatial distribution of the final stations selected for analysis is reasonable, covering all regions of the country.
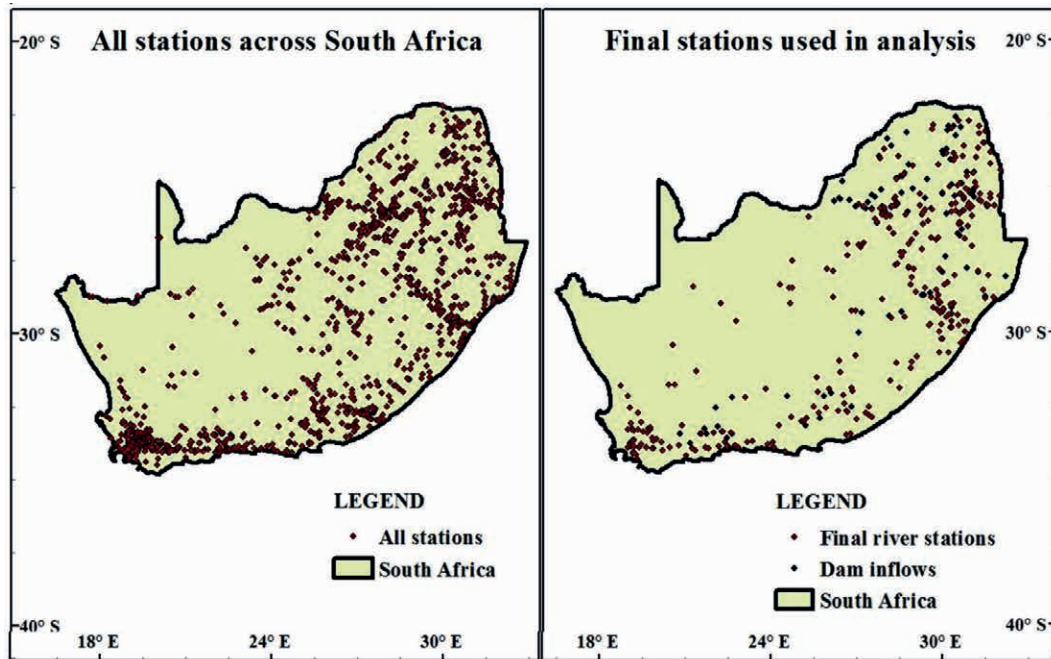
Figure 8
Distribution of all flow records across South Africa and distribution of selected stations and dam inflows records used in the analysis

RESULTS AND DISCUSSION

This section contains the results of the performance of the Meigh, Mkhandi, JPV and Haile methods. The 2- to 100-year design floods estimated by these methods were compared to the 2- to 100-year design floods estimated using an at-site frequency analysis of the observed data at the selected stations used in this study.

Selection of probability distributions

The application of both the GEV and LP3 distributions for design flood estimation in South Africa have been advocated by a number of studies (Görgens, 2007; Van der Spuy and Rademeyer, 2010). The Görgens (2007) study developed regional approaches based on design floods estimated using both the GEV and LP3 distributions. Therefore, design floods in this study were initially estimated using both the GEV and LP3 distributions, fitted by L-moments to the observed AMS (Hosking and Wallis, 1997). Figure 9 illustrates a comparison between the 50-year return period design floods estimated using the GEV and LP3 distributions fitted to the observed AMS.

It can be seen in Fig. 9 that in certain cases the LP3 distribution produced design floods that are inconsistent with the GEV design floods. In these cases the observed LP3 design floods for the 50-year return period are significantly larger than the GEV design floods. In addition, it was also found that not only are the LP3 design floods computed from the observed data inconsistent, there are also cases where the JPV method, when using the LP3 distribution, produced design floods that are orders of magnitude greater than both the JPV GEV design floods and design floods computed from observed AMS for the same stations. Thus the performance of the RFFA methods was only assessed using the GEV distribution.

Ratio of estimated and observed design floods

The systematic over-estimation or under-estimation of a method gives an indication of the degree of bias in the method (Haddad and Rahman, 2012). The ratio statistic used to measure this degree of bias is defined as E/O, where E is the estimated design flood computed using the regional method, and O is the observed design flood computed using the at-site analysis. Haddad and Rahman (2012) considered 3 limits of this ratio to define the degree of bias produced by the regional methods. An E/O ratio that falls between 0.5 and 2 is an indication of a 'desirable estimate (D)'. An E/O ratio that is less than 0.5 is considered to be a 'gross under-estimation (GU)' and an E/O ratio that is greater than 2 is considered to be a
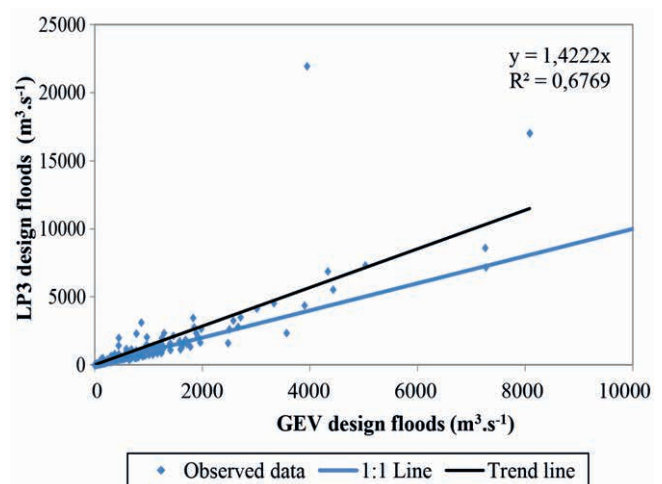

Figure 9
Comparison of observed design floods for the GEV and LP3 distribution, for the 50-year return period

'gross over-estimation (GO)'. These limits have been defined by Haddad and Rahman (2012) and are subjective; however, they provide a reasonable indication of the accuracy of a method. Figures 10 and 11 illustrate the percentage of stations for each method that produced an *E/O* ratio that falls in either the GU; GO or D ranges for the 2- to 10-year return periods and the 20- to 100-year return periods, respectively. The K-region and veld zone methods in Figs 10 and 11 are both from the Görgens (2007) JPV method.

It can be seen from Figs 10 and 11 that the percentage of stations that fall within the desirable estimate range for the Haile method is generally 60% or greater. However, for the Haile method more than 20% of the stations fall within the gross under-estimation range for all return periods. Similarly,

approximately 60% or more of the stations fall within the desirable estimate range for the Mkhandi method; however, close to 30% of the stations fall within the gross over-estimation range. The JPV method, using the K-region regionalisation produces 50% or more stations that lie within the desirable estimate range; however, approximately 35% of the stations fall within the gross over-estimation range. The JPV method, using the veld zone regionalisation, produces approximately 48% or more stations that lie within the desirable estimate range; however, approximately 35% of the stations fall within the gross over-estimation range. In the case of the Meigh method almost half of the stations (between 41% and 49 %) fall within the gross over-estimation range. This indicates poor performance, as almost half of the stations analysed produced design floods that are grossly over-estimated.
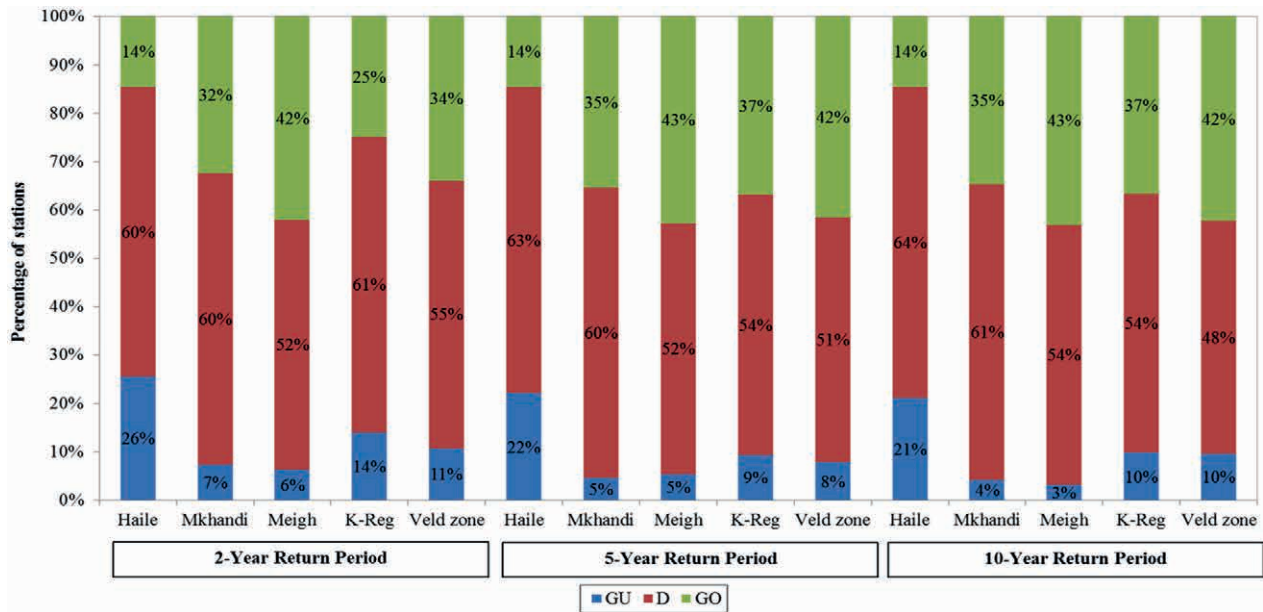

Figure 10
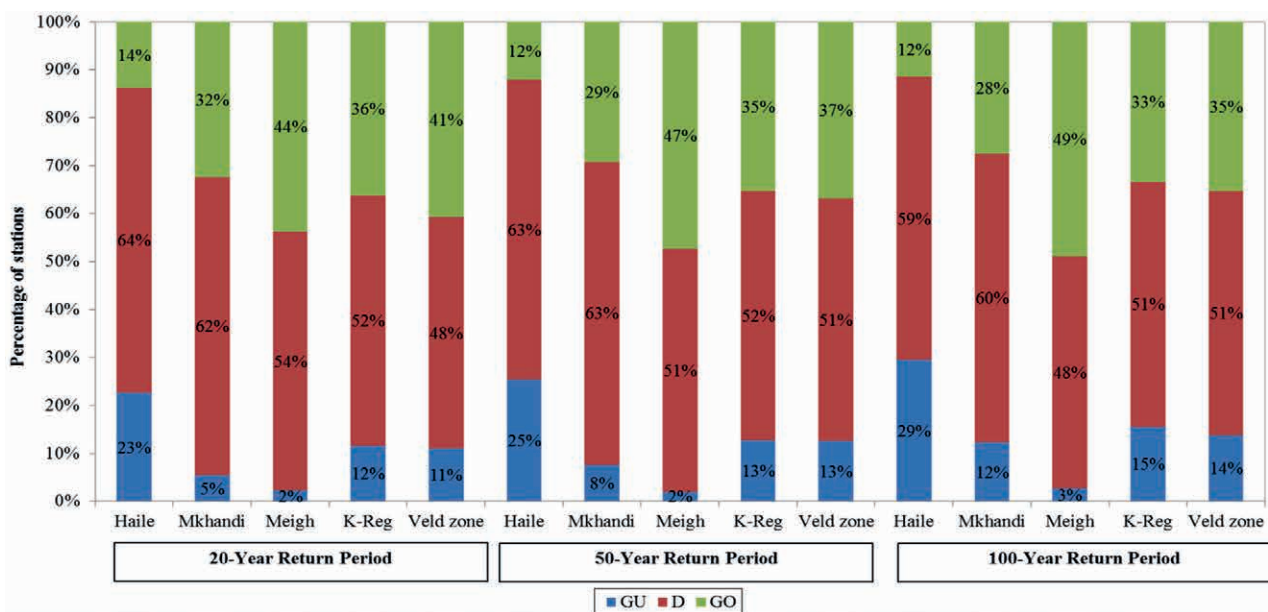*Estimated/observed ratios for the 2- to 10-year return periods*

Figure 11
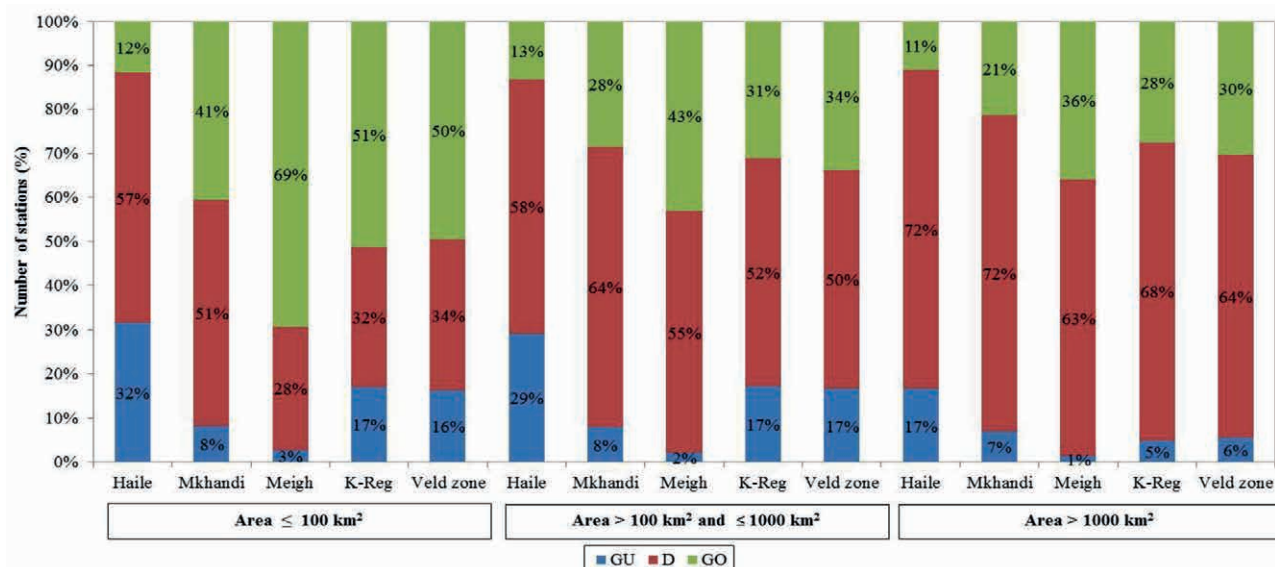*Estimated/observed ratios for the 20- to 100-year return periods*

Figure 12
Estimated/observed ratios for the 50-year return period at different catchment area ranges

Table 3 contains a summary of the results, where the average GU, GO and D ranges have been calculated for each method for all return periods. The methods are presented from the highest to lowest average ratio for the desirable estimate range.

It can be seen from Table 3 that the Haile method, on average, produces the most stations that lie within the desirable estimate range. However, it also produces the highest number of stations that fall within the gross under-estimation range (24%) when compared to the other methods. The Mkhandi method also produces almost the same percentage of stations as the Haile method that lie within the desirable estimate range; however, it also produces a high percentage of stations that fall within the gross over-estimation range. The JPV method, using the K-region regionalisation and the veld zone regionalisation both perform similarly, producing more than 50% of the stations that fall within the desirable estimate range and more than 33% of the stations falling within the gross over-estimation range. The Meigh method performs the worst with an average of 45% of the stations that fall within the gross over-estimation range.

Trends in method performance by catchment area

The 407 stations analysed in this study were divided according to 3 catchment area ranges, i.e., stations with catchments that have areas that are less than or equal to 100 km^2 (111 stations), stations with catchment areas that are greater than 100 km^2 and are less than, or equal to, 1 000 km^2 (151 stations) and stations with catchment areas that are greater than 1 000 km^2

(145 stations). The *E/O* ratios have been calculated for all of the stations within each catchment area range for the 50-year return period and Fig. 12 illustrates the percentage of stations that fall into either the GU, GO or D ranges for the different methods. In general, all of the RFFA methods produce better results for catchments with larger areas, particularly those with areas greater than 1 000 km^2. Conversely, the RFFA methods produce the worst results or the least percentage of stations within the desirable estimate range for smaller catchments with areas that are less than or equal to 100 km^2. The Haile method produced the highest percentage of stations in the desirable estimate range for catchments with areas less than or equal to 100 km^2 and for stations with catchment areas greater than 1 000 km^2. However, the Haile method also produced the highest percentage of stations in the gross under-estimation range for all catchment area ranges. The Mkhandi method produced the highest percentage of stations that lay within the desirable estimate range for catchments with areas that were greater than 100 km^2, while the Meigh method performed the worst, producing the lowest percentage of stations that lay within the desirable estimate range for all catchment area ranges, with the exception of the stations that were greater than 100 km^2 and less than or equal to 1 000 km^2, where the JPV methods using the K-region regionalisation and the Veld Zone regionalisation produced the lowest percentage of stations (52% and 50%, respectively) in the desirable estimate range.

Trends in method performance by spatial location

This section will deal with the variation in method performance, based on where the method is applied in South Africa. In order to determine the spatial variation in method performance, the methods were ranked from 1 to 5 (best to worst) according to the method that produced an *E/O* ratio that was closest to 1 at a particular station, for the 50-year return period. Thereafter, the location of every station, where a particular method ranked as No. 1, was plotted in ArcGIS to assess any spatial variation in model performance. This was done for all the RFFA methods to produce the map shown in Fig. 13, which represents the 50-year return period. The same

TABLE 3			
Average GU, GO and D ranges for all the return periods			
RFFA method	Average number of stations (%)		
	GU	D	GO
Haile	24.4	62.2	13.5
Mkhandi	6.9	61.1	31.9
JPV K-region	3.6	51.7	44.6
Meigh	12.2	53.9	33.9
JPV veld zone	10.9	50.6	38.5

procedure has been followed for the 2-, 10-, 20- and 100-year return periods and the maps produced indicate similar trends.

It is evident from Fig. 13 that there is some clustering of methods being ranked as the best in certain regions. The JPV method, using the veld zone regionalisation, represented by yellow squares, is ranked No. 1 in Drainage Region V. The JPV method, using the K-region regionalisation, represented by blue squares, is ranked No. 1 in Drainage Region W. However, these are minor spatial trends, as the Haile method continues to perform the best and is ranked No. 1 throughout the coastal areas of South Africa, as well as for the northern regions of the country. Figure 13 illustrates not only the best ranked methods, but also the mean annual precipitation (MAP) and the primary drainage regions across South Africa. Therefore, it is evident that the primary drainage region and the MAP do not have a strong correlation with method performance. For the 50-year return period, the Meigh method is not ranked as the best at any of the stations; however, for other return periods there are several cases where it is. In general, the Haile method was ranked as the best method for the highest number of stations throughout the country. However, it is important to note that a method being ranked as the best for a particular station does not indicate that the results are acceptable. A method may produce better results than the other methods, but it may still be unacceptable with inaccurate estimates of design floods. It can be concluded from Fig. 13 that there are no major or consistent spatial variations in method performance.

CONCLUSION AND RECOMMENDATIONS

According to the ranking of the RFFA methods which have been determined using the average percentage of stations that lie within the desirable estimate ratio range, the Haile method, Mkhandi method and the JPV method, using the K-region regionalisation, have been ranked as the best-performing three methods of those considered in this study.

As far as the overall performance of the methods, the Haile method consistently outperformed the Mkhandi method and the JPV methods and could therefore be considered as the best method for use in South Africa. However, the Haile method consistently underestimates design floods. The logarithmic form of the Haile regression equations to estimate the MEF were not applicable to all stations in South Africa due to the area limits produced by the logarithmic equations. An attempt was made to overcome this limitation in this study, and further improvements are recommended to estimate MEF in the future. It is acknowledged that the performance of the Haile method could have been influenced by the use of a power function to estimate the MEF instead of the logarithmic function used by Haile (2011). In addition to the aforementioned, it must be noted that different types of data were utilised for each method. Haile (2011) utilised daily average inflows, Görgens (2007) utilised peak-over-threshold data and Mkhandi (2000) utilised instantaneous maximum flow data. This may account for the under-estimation of the Haile (2011) method in comparison to the other methods, as the daily average inflows utilised by Haile (2011) may have missed certain large events, especially in smaller catchments.

The Mkhandi and JPV methods consistently over-estimated design floods, which may indicate that these methods are conservative and could be acceptable in design flood practice. However, the extent of over-estimation by these methods makes them unacceptable for use, as these methods produce an average percentage of stations that are within a gross over-estimation range that is approximately 30% or greater. The results of this study also indicated the inconsistent performance of the LP3 distribution. Therefore, further investigations must be undertaken concerning the use of the LP3 distribution in South African design flood practice.

In order to test the performance of the methods with different catchment sizes, the 407 stations analysed in this study were divided into 3 catchment area ranges, i.e., stations
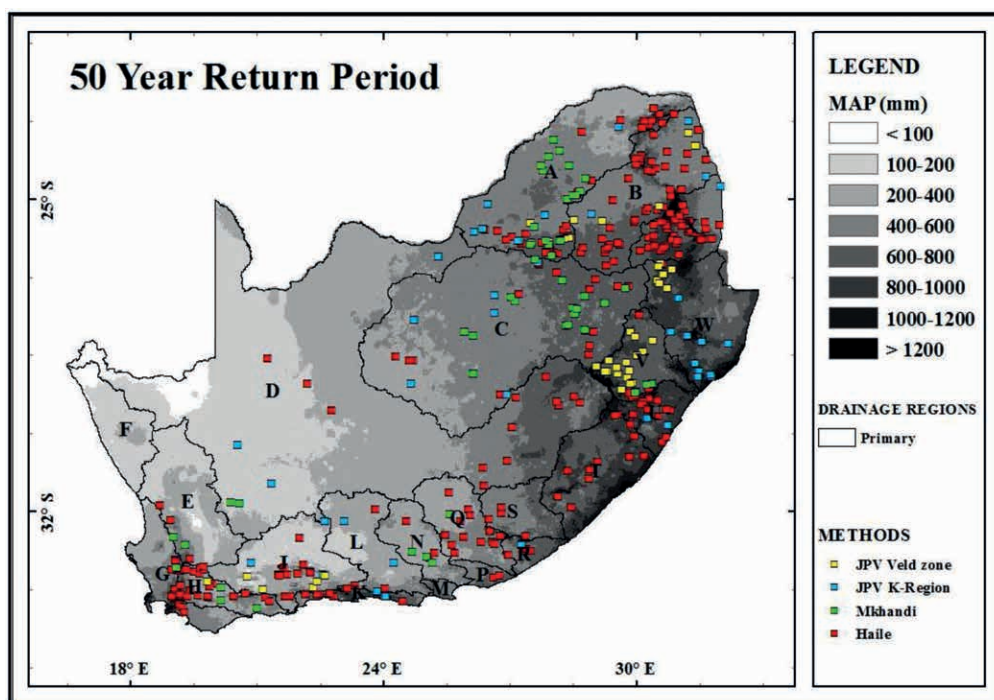

Figure 13
Method rank for the 50-year return period

with catchments with areas less than or equal to 100 km^2, stations with catchment areas that are greater than 100 km^2 and are less than, or equal to, 1 000 km^2 and stations with catchment areas that are greater than 1 000 km^2. Thereafter, the *E/O* ratios were calculated for all of the stations within each catchment area range for the 50-year return period. In general, all of the RFFA methods produced better results for catchments with larger areas, particularly those with areas greater than 1 000 km^2. Conversely, the RFFA methods produce the worst results or the least percentage of stations within the desirable estimate range for smaller catchments with areas that are less than or equal to 100 km^2.

In order to determine the spatial variation in method performance, the methods were ranked from 1 to 5 (best to worst), according to the method that produced an *E/O* ratio that was closest to 1 at a particular station. Thereafter, the location of every station, where a particular method ranked No. 1, was plotted in ArcGIS to represent that method. It was found that neither MAP nor the drainage regions have any visual correlation to the methods' performance. In general, the Haile method was ranked as the best method for the highest number of stations throughout the country. However, it is important to note that a method being ranked as the best for a particular station does not indicate that the results are acceptable. The method may produce better results than the other methods but it may still be unacceptably inaccurate in estimating design floods. It can be concluded that there are no major spatial trends in the performance of the RFFA methods assessed in this study.

The *E/O* ratio, as used by Haddad and Rahman (2012), was used to estimate the performance of the methods and quantifies the prediction error of the methods. This does not account for at-site sampling errors, related primarily to the record length, in the estimation of design floods from the observed data. It is thus recommended that the approach used, for example, by Kjeldsen (2015), be used in future studies to formally assess the reliability of the methods.

A study by Nathanael (2015) on the Meigh, Mkhandi, JPV and Haile methods did not find any significant spatial trends in the performances of the methods throughout the country. Due to the poor overall performance of all of the RFFA methods, it is recommended that a new RFFA method be developed for application in South Africa. In addition, it is recommended that the new methods developed provide different approaches to estimate design floods at different catchment area ranges, as well as adjustments for the application of the method in urban areas and under climate change scenarios, as these aspects were not accounted for in the development of the RFFA methods assessed in this study.

REFERENCES

ALEXANDER WJR (1990) Flood Hydrology for Southern Africa. SANCOLD, Pretoria, South Africa.

ALEXANDER WJR (2001) Flood Risk Reduction Measures. University of Pretoria, Pretoria, South Africa.

AUSTRALIAN INSTITUTION OF ENGINEERS (1977) *Australian Rainfall and Runoff: Guide to Flood Estimation*. EA Books, Crows Nest, Australia.

BOBEE B and RASMUSSEN F (1995) Recent advances in flood frequency analysis. *U.S National Report 1991-1994. Twenty-first General Assembly International Union of Geodesy and Geophysics Boulder, Colorado*. 1111–1116. https://doi.org/10.1029/95RG00287

CORDERY I and PILGRIM DH (2000) The state of the art of flood prediction. In: Parker DJ (ed.) Floods. Volume II. Routledge, London. 185–197.

DOOCY S, DANIELS A, MURRAY S and KIRSCH TD (2013) The human impact of floods: a historical review of events 1980-2009 and systematic literature review. *PLOS Curr*. Disasters 2013 Apr 16. Edition 1. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

DALRYMPLE T (1960) Flood-frequency analyses, Manual of Hydrology: Part 3. US Geological Survey. Water Supply Paper 1543-A. United States Government Printing Office, Washington DC.

FABER BA (2010) Current methods for flood frequency analysis. Workshop on Non-Stationarity, Hydrologic Frequency Analysis and Water Management, January 13–15, 2010, Boulder, Colorado. 33–38.

GÖRGENS AHM (2007) Joint peak-volume (JPV) design flood hydrographs for South Africa. WRC Report No. 1420/3/07. Water Research Commission, Pretoria.

GREHYS (1996) Inter-comparison of regional flood frequency procedures for Canadian rivers. *J. Hydrol.* 186 (1–4) 85–103.

HADDAD K and RAHMAN A (2012) Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – quantile regression vs. parameter regression technique. *J. Hydrol.* 430–431 142–161. https://doi.org/10.1016/j.jhydrol.2012.02.012

HAILE AT (2011) Regional flood frequency analysis in Southern Africa. Unpublished MSc thesis, Department of Geosciences, University of Oslo, Norway.

HOSKING JRM and WALLIS JR (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK.

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

HUBBART JA and JONES JR (2009) Floods. In: *Encyclopedia of Inland Waters Reference Module in Earth Systems and Environmental Sciences*. Elsevier, Massachusetts. 88–91. https://doi.org/10.1016/B978-012370626-3.00229-5

JINGYI Z and HALL MJ (2004) Regional flood frequency analysis for the Gan-Ming River basin in China. *J. Hydrol.* 296 (1–4) 98–117. https://doi.org/10.1016/j.jhydrol.2004.03.018

KACHROO RK, MKHANDI SH and PARIDA BP (2000) Flood frequency analysis of southern Africa: I. Delineation of homogeneous regions. *Hydrol. Sci. J.* 45 (3) 437–447. https://doi.org/10.1080/02626660009492340

KIZZA M, NTALE HK, RUGUMAYO AI and KIGOBE M (2006) Regional flood frequency analysis for Northern Uganda Using the L-Moment approach. In: Taban-Wani JAM (ed.) *Proceedings from the International Conference on Advances in Engineering and Technology*. Elsevier Science Ltd, Oxford, UK.

KJELDSEN TR, SMITHERS JC and SCHULZE RE (2001) Flood frequency analysis at ungauged sites in the KwaZulu-Natal Province, South Africa. *Water SA* 27 (3) 315–324. https://doi.org/10.4314/wsa.v27i3.4974

KJELDSEN TR, SMITHERS JC and SCHULZE RE (2002) Regional flood frequency analysis in the KwaZulu-Natal Province, South Africa, using the index-flood method. J. Hydrol. 255 (1–4) 194–211. https://doi.org/10.1016/S0022-1694(01)00520-0

KJELDSEN TR (2015) How reliable are design flood estimates in the UK? *J. Flood Risk Manage.* 8 237–246. https://doi.org/10.1111/jfr3.12090

KOVACS Z (1988) Regional maximum flood peaks in southern Africa. Report No. TR 137. Department of Water Affairs, Pretoria.

KUNDZEWICZ ZW, KANAE S, SENEVIRATNE SI, HANDMER J and NICHOLLS N (2013) Flood risk and climate change: global and regional perspectives. *Hydrol. Sci. J.* 59 (1) 1–28. https://doi.org/10.1080/02626667.2013.857411

LECLERC M and OUARDA TBMJ (2007) Non-stationary regional flood frequency analysis at ungauged sites. *J. Hydrol.* 343 (3–4) 254–265. https://doi.org/10.1016/j.jhydrol.2007.06.021

MALEKINEZHAD H, NACHTNEBEL HP and KLIK A (2011) Comparing the index-flood and multiple-regression methods using L-moments. *Phys. Chem. Earth A/B/C* 36 (1–4) 54–60. https://doi.org/10.1016/j.pce.2010.07.013

MEDIERO L and KJELDSEN TR (2014) Regional flood hydrology in a

semi-arid catchment using a GLS regression model. *J. Hydrol.* 514 158–171. https://doi.org/10.1016/j.jhydrol.2014.04.007

MEIGH JR, FARQUHARSON FAK and SUTCLIFFE JV (1997) A worldwide comparison of regional flood estimation and climate. *Hydrol. Sci. J.* 42 (2) 255–244. https://doi.org/10.1080/02626669709492022

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. https://doi.org/10.1080/02626660009492341

NATHANAEL J (2015) Assessing the performance of regional flood frequency analysis methods in South Africa. Unpublished MSc thesis, School of Agriculture, Earth and Environmental Sciences, Department of Hydrology, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

SAF B (2008) Application of Index Procedures to Flood Frequency Analysis in Turkey. *J. Am. Water Resour. Assoc.* 44 (1) 37–47. https://doi.org/10.1111/j.1752-1688.2007.00136.x

SAF B (2009) Regional flood frequency analysis using L-Moments for the West Mediterranean Region of Turkey. *Water Resour. Manage.* 23 (3) 531–551. https://doi.org/10.1007/s11269-008-9287-z

SHICEKA S (2011) Media statement on the declaration of a national disaster as a result of flooding and other natural disasters in the country. Ministry for Cooperative Governance and Traditional Affairs, Pretoria.

SMITHERS JC (2012) Methods for design flood estimation in South Africa. *Water SA* 38 (4) 633–646. https://doi.org/10.4314/wsa.v38i4.19

SMITHERS JC, GÖRGENS AH, GERICKE OJ, JONKER V and ROBERTS P (2014). The Initiation of a National Flood Studies Programme for South Africa. *SANCOLD National Conference* 2014, 5–7 November 2014, *Johannesburg.* SANCOLD, Pretoria.

SMITHERS JC and SCHULZE RE (2000) Development and evaluation of techniques for estimating short duration design rainfall in South Africa. WRC Report No. 681/1/100. Water Reasearch Commision, Pretoria.

VAN BLADEREN D (1993) Application of historical flood data in flood frequency anlysis, for the Natal and Transkei region. In: Lorentz SA, SW Kienzle and MC Dent (eds) *Proceedings of the Sixth South African National Hydrological Symposium.* Department of Agricultural Engineering, University of Natal, Pietermaritzburg, South Africa. 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.

VAN VUUREN SJ, VAN DIJK M and COETZEE GL (2013) Status review and requirements of overhauling flood determination methods in South Africa. WRC Report No. TT563/13. Water Research Commission, Pretoria. 91 pp.

VIGLIONE A, LAIO F and CLAPS P (2007) A comparison of homogeneity tests for regional frequency analysis. *Water Resour. Res.* 43 (3) 1–10. https://doi.org/10.1029/2006WR005095

WAZNEH H, CHEBANA F and OUARDA TBMJ (2015) Delineation of homogeneous regions for regional frequency analysis using statistical depth function. *J. Hydrol.* 521 232–244. https://doi.org/10.1016/j.jhydrol.2014.11.068

WILTSHIRE SE (1986) Regional flood frequency analysis I: Homogeneity statistics. *Hydrol. Sci. J.* 31 (3) 321–333. https://doi.org/10.1080/02626668609491051

ZAMAN MA, RAHMAN A, HADDAD K (2012) Regional flood frequency analysis in arid regions: A case study for Australia. *J. Hydrol.* 475 74–83. https://doi.org/10.1016/j.jhydrol.2012.08.054