Flood frequency analysis at ungauged sites in the KwaZulu-Natal Province, South Africa

TR Kjeldsen[1]*, JC Smithers[2]* and RE Schulze[2]

[1]*Environment & Resources DTU, Technical University of Denmark, Building 115, DK-2800 Kgs. Lyngby, Denmark*
[2]*School of Bioresources Engineering and Environmental Hydrology, University of Natal, Private Bag X01, Scottsville 3209, South Africa*

Abstract

Use of the index-flood method at ungauged sites requires methods for estimation of the index-flood parameter at these sites. This study attempts to relate the mean annual flood to site characteristics of catchments in KwaZulu-Natal, South Africa.

The ordinary, weighted and generalised least square methods for estimating model parameters are compared and found to perform equally well, with preference given to the generalised least square model. A separation of KwaZulu-Natal (KZN) into two regions was found to improve predictive ability of the models in the western and north-western parts of the study area. The study also revealed problems with the estimation of the mean annual flood in the coastal areas of the study region.

Introduction

Design flood estimation in South Africa is generally based on empirical and deterministic models such as the rational method, unit hydrograph and the SCS model (Alexander, 1990). While direct statistical analysis and regional techniques are described and advocated by Alexander (1990), little effort has been devoted to the development of regional estimation techniques. The use of regional flood frequency analysis has two advantages compared with traditional at-site analysis. Firstly, the reliability of the estimated design events increases due to the inclusion of additional spatial information. Secondly, the design events can be estimated at ungauged sites, i.e. sites where no measurements of floods exist, as shown by Hosking and Wallis (1997). A popular and widely used regional method is the index-flood method as described, *inter alia*, by Stedinger et al. (1993) and Hosking and Wallis (1997). To use the index-flood method for estimation at ungauged sites, a relationship, based on regression models, between the index-flood, which is often the mean annual flood (MAF), and the corresponding catchment characteristics has to be developed. However, the reliability of the design event estimated at an ungauged site is significantly lower than the corresponding estimate at a gauged site. This prompted the NERC (1975) to recommend that a gauging weir should be constructed as soon as the need for flood estimates at a particular site was foreseen. This was further supported by Hebson and Cunnane (1987) who showed that estimates of MAF at ungauged sites are less precise than estimates at sites with even only one year of available data.

Previous attempts to develop regression models in South Africa have been reported by van Bladeren (1993), Mkhandi and Kachroo (1997) and Meigh et al. (1997). All three studies used explained variance (R^2) to quantify model performance. Both van Bladeren (1993) and Mkhandi and Kachroo (1997) modelled MAF as a function of catchment area (AREA) based on drainage regions

as defined by DWAF (1990). In his study, van Bladeren (1993) focused only on KZN, which he further subdivided into seven subregions and obtained R^2-values in the range 0f 0.84 to 0.96. Mkhandi and Kachroo (1997) divided KZN into two regions representing drainage region S, T and U and drainage region V respectively. The model for region S, T and U obtained an R^2 value of 0.727, and for drainage region V an R^2 value of 0.707. Meigh et al. (1997) developed a general regression model for South Africa and Botswana relating MAF to AREA and mean annual precipitation (MAP). They divided the considered catchments into two groups based on AREA as AREA > 1 250 km^2 and AREA < 1 250 km^2 and reported R^2 values of 0.542 and 0.593 respectively. These results indicate that a grouping of catchments based on geographical location rather than catchment area is preferable. In all three studies the lack of other easily accessible catchment characteristics other than AREA was reported as a barrier towards improved model development and performance.

The index-flood method

The key assumption underlying the index-flood method is that annual maximum series (AMS) of floods from different sites are identically distributed except for a scale parameter, termed the index-flood. In this study the MAF is defined as the index-flood. The method requires identification of homogeneous regions in which the key assumption is valid. The T-year event (X_T) at any site within the homogeneous region can be estimated as:

$$\hat{X}_{T,i} = \hat{\mu}_i \hat{z}_T \tag{1}$$

where:
$\hat{\mu}_i$ = MAF at site i
$\hat{z}_T$ = (1-1/T) quantile in the regional frequency distribution of normalised AMS.

The regional growth curve z_T describes the relationship between the normalised flood magnitude and the corresponding exceedance probability and is considered constant within a homogeneous region. Kjeldsen et al. (2000) recommended using either the General Normal (GNO), the Pearson Type 3 (P3) or the General

---

Pareto (GPA) distributions as regional distributions of the normalised AMS for floods in the west and north-western parts of the KZN province of South Africa. In this study the GNO distribution has been adopted and its statistical properties are summarised in Appendix A. No suitable distribution was identified for the coastal and midlands area of KZN and, therefore, even though it can be characterised as homogeneous, further research is required before the index-flood method can be applied in this region.

When estimating the T-year event at an ungauged site, the index-flood parameter μ_i is unknown, but the regional growth curve z_T will be known. Hence, the problem of estimating design floods at ungauged sites is reduced to the estimation of the MAF at the site under consideration. A widely used approach for estimation of the index-flood parameter is through the use of a non-linear relationship between the index-flood parameter and a number of site characteristics as shown, for example, by NERC (1975), Stedinger et al. (1993) and Hosking and Wallis (1997). The non-linear model is linearised through a log-transformation:

$$\ln[\mu_i] = \ln[b_0] + b_1 \ln[A_1] + \cdots b_p \ln[A_p] + \eta_i \qquad (2)$$

where:

μ_i = MAF at site i, and
b_j = model parameters , j = 0,...p, and
A_j = site characteristics j = 1,...p, and
η_i = error term.

Stedinger and Tasker (1985) compared ordinary least square (OLS), weighted least square (WLS) and the general least square (GLS) methods for estimating the parameters in Eq. (2) and found that GLS and WLS performed better than the OLS when the basic assumptions (independent and homoscedastic residuals) of the OLS are violated. Estimation based on GLS provides more accurate estimation of both model parameters and model error than the OLS method. The WLS method performs closely to the GLS method for modest inter-site correlation.

The GLS regression model

The assumption behind the OLS method is that the model residuals are independent and homoscedastic. The WLS method is an extension of the OLS allowing the residuals to be heteroscedastic, which is often the case in hydrology where the MAF will be estimated from AMS of different record length. The GLS method extends the WLS by allowing the residuals to be cross-correlated, which is also to be expected when AMS from a confined geographical region are analysed.

Consider the assumed log-linear relationship between μ_i and the site characteristics as defined in Eq. (3):

$$\theta_i = \ln[b_0] + \sum_{k=1}^{p} b_k L_{ki} + \delta_i \qquad (3)$$

where:

θ_i = logarithm of μ_i at site i = 1,...,N
b_j = model parameters, j = 0,..., p
L_{ki} = logarithm of catchment property k = 1,..., p at site i
δ_i = model error due to lack of fit of regression model.

The properties of the model error are given as:

$$E\{\delta_i\} = 0$$

$$\text{Cov}\{\delta_i, \delta_j\} = \begin{cases} \sigma_\delta^2 & i = j \\ 0 & i \neq j \end{cases} \qquad (4)$$

where θ_δ^2 is model error variance. Next, consider the MAF at site i. The estimator $\hat{\theta}_i$ of θ_i is subject to sampling error so that:

$$\hat{\theta}_i = \theta_i + \varepsilon_i \qquad (5)$$

where:

$$E\{\varepsilon_i\} = 0 \qquad (6)$$

$$\text{Cov}\{\varepsilon_i, \varepsilon_j\} = \begin{cases} \sigma_{\varepsilon i}^2 & i = j \\ \sigma_{\varepsilon i} \sigma_{\varepsilon j} \rho_{\varepsilon ij} & i \neq j \end{cases}$$

In Eq. (6) $\sigma_{\varepsilon i}^2$ is the sampling error variance of θ_i at site i, and $\rho_{\varepsilon ij}$ is the inter-site correlation due to concurrent observations at site i and j. When $\rho_{\varepsilon ij} = 0$, i.e. neglecting inter-site correlation, the GLS method reduces to the WLS model. If it is assumed that the at-site sampling error variance is considered constant at all sites, WLS is reduced to OLS (Stedinger and Tasker, 1985). The total errors (η_i) consists of a model error due to lack of fit of the regression model and a sampling error due to the use of $\hat{\theta}_i$ instead of θ_i, i.e.:

$$\eta_i = \delta_i + \varepsilon_i \qquad (7)$$

The elements of the covariance matrix $\mathbf{G}$ of the total error as defined in Eq. (7) are given as:

$$g_{ij} = \text{Cov}\{\eta_i, \eta_j\} = \begin{cases} \sigma_\delta^2 + \sigma_{\varepsilon i}^2, & i = j \\ \sigma_{\varepsilon i} \sigma_{\varepsilon j} \rho_{\varepsilon ij}, & i \neq j \end{cases} \qquad (8)$$

The multiple linear regression equation can be expressed on matrix form as:

$$\mathbf{U} = \mathbf{BX} + \mathbf{H} \qquad (9)$$

where:

$$\mathbf{U} = (\theta_1, \theta_2, \cdots, \theta_N)^T$$
$$\mathbf{B} = (\ln[b_0], b_1, \cdots, b_p)^T \qquad (10)$$
$$\mathbf{H} = (\eta_1, \eta_2, \ldots \eta_n)^T$$

$$\mathbf{X} = \begin{bmatrix} 1 & L_{11} & \cdots & L_{1p} \\ \cdot & \cdot & \cdots & \cdot \\ 1 & L_{N1} & \cdots & L_{Np} \end{bmatrix}$$

The GLS estimator of the model parameter matrix $\mathbf{B}$ is obtained from:

$$[\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X}]\mathbf{B} = \mathbf{X}^T\mathbf{G}^{-1}\mathbf{U} \qquad (11)$$

where σ_δ^2 is found by solving

$$(\mathbf{U} - \mathbf{XB})^T \mathbf{G}^{-1}(\mathbf{U} - \mathbf{XB}) = N - p - 1 \qquad (12)$$

The solution of Eqs. (11) and (12) requires an iterative procedure where Eq. (12) is solved for σ_δ^2 using, for example, ordinary bisection. According to Stedinger and Tasker (1985) it can happen that no positive value of σ_δ^2 can solve Eq.(12), in which case the difference between $\mathbf{U}$ and $\mathbf{XB}$ is explained by the sampling error alone.

Application of the GLS method requires estimation of the sampling error covariance matrix. Estimation of $\sigma_{\varepsilon i}^2$ is complicated by the fact that it should be independent of θ (Stedinger and Tasker, 1985). Madsen and Rosbjerg (1997) presented an appropriate estimator as:

$$\hat{\sigma}_{\varepsilon i}^2 = \frac{(\overline{C}_v)^2}{n_i} \qquad (13)$$

where:

C_v = coefficient of variation of non-log-transformed AMS at site *i,* and

$$\overline{C}_v = 1/N \sum_{i=1}^{N} \hat{C}_{vi} \tag{14}$$

Furthermore, the inter-site correlation $\rho_{\varepsilon ij}$ between $\ln[\mu_i]$ and $\ln[\mu_j]$ can be estimated as the correlation between concurrent observations at site *i* and *j*, ρ_{ij}, as:

$$\rho_{\varepsilon ij} = \mathrm{Cor}\left\{\ln[\mu_i], \ln[\mu_j]\right\} = \mathrm{Cor}\left\{\mu_i, \mu_j\right\} = \rho_{ij} \tag{15}$$

as shown by Stedinger (1983) and Madsen and Rosbjerg (1997).

Once the model parameters have been estimated, predicted values of θ_i and the corresponding predictive variance can be estimated as:

$$E\{\hat{\theta}_i\} = x_i^T \hat{\mathbf{B}} \tag{16}$$

$$\mathrm{Var}\{\hat{\theta}\} = \hat{\sigma}_\delta^2 + x_i^T \sum (\hat{\mathbf{B}}) x_i$$

where:

$\mathbf{x}_i^T$ = the i-th row in **X,** and
$\sum (\hat{\mathbf{B}}) = [\mathbf{X}^T \mathbf{G}^{-1} \mathbf{X}]^{-1}$ = the covariance matrix of the estimated parameters.

Hence, the total predictive variance in Eq. (16) is the sum of the model error variance and the sampling uncertainty of the estimated regression model parameters. As $\mu_i = \exp[\theta_i]$, the following expressions are valid for the prediction of MAF

$$\hat{\mu}_i = \exp\left(\mathbf{x}_i^T \hat{\mathbf{B}}\right)$$

$$\hat{\sigma}_{\mu i}^2 = \exp(2\hat{\theta}) \mathrm{Var}\{\hat{\theta}\} \tag{17}$$

which can be calculated based on Eq. (16).

Data utilised

The data utilised were the MAFs from 29 relatively unregulated rivers within KZN. These 29 catchments were divided into two regions based on an index of monthly rainfall concentration (CONC), as shown in Fig. 1. Region 1 represents the coastal and midlands area of KZN and consists of 12 catchments. Region 2 covers the western and north-western parts of KZN and consists of 17 catchments. The AMS from Region 1 all have high coefficients of

variation and coefficients of skew and all include high outliers.

For each station a number of catchment characteristics was obtained, as shown in Table 1. These constitute a subset of the characteristics listed by Alexander (1990) as being of importance to flood hydrology. Furthermore, the selection was based on the assumption that they should be easily accessible from readily available maps in electronic form from Schulze (1997) and derived using a 200 m × 200 m DEM and GIS. Gravelious' compactness coefficient (GCC) was estimated as

$$GCC = P/2\sqrt{\pi AREA} \tag{18}$$

where *P* is the perimeter of the catchment [m], and *AREA* is the catchment area [m^2].

Inter-site correlation

The GLS method requires estimates of the inter-site correlation. Tasker and Stedinger (1989) found that using sample estimates of ρ_{ij} can lead to singular covariance matrix **G** that cannot be inverted as required in Eq. (12). They recommend relating the inter-site correlation between site *i* and *j* to the distance d_{ij} between the two gauging weirs as:

$$\rho_{ij} = \theta^{\left(\frac{d_{ij}}{\alpha d_{ij} + 1}\right)} \tag{19}$$

Figure 2 shows the observed and estimated inter-site correlation when using:

- all catchments
- only those in Region 1 and
- only those in Region 2.

The figure clearly indicates the existence of inter-site correlation. Significant correlation-distance relationships are evident in Region 2 when using all catchments, whereas the correlation appears to be relatively constant in Region 1. Alternatively, the inter-site correlation can be assumed to be constant and equal to the average inter-site correlation as $\rho_{ij} = \bar{\rho}$ for $i \neq j$. Both approaches were adopted in this study, with GLS1 using regional constant inter-site correlation and GLS2 using ρ_{ij} from Eq. (19) to estimate inter-site correlation.
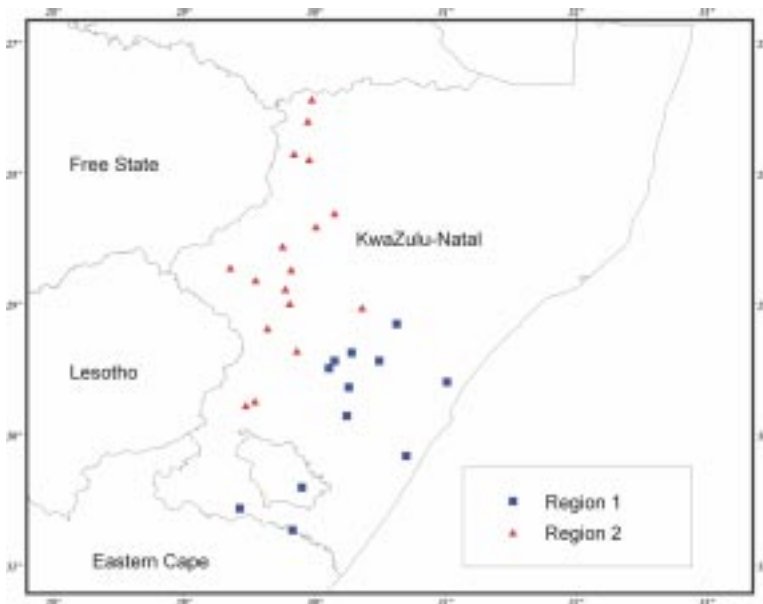

Figure 1 (left)
Study area and location of gauging weirs measuring relatively unregulated streamflows

TABLE 1
Catchment characteristics

Characteristics	Notation
Catchment area [km^2]	AREA
Mean annual precipitation [mm]	MAP
Rainfall concentration [%]	CONC
Mean catchment slope [%]	MCS
Infiltration capacity [m·m^{-1}] *	SOIL
Mean altitude [m]	ALT
Gravelious' compactness coeff. [-]	GCC

* Equal to the FC1 parameter in the *ACRU* model (Schulze, 1995)

Applications

The regression models for prediction of MAF were estimated using OLS, WLS, GLS1 and GLS2 using

- all 29 catchments,
- the 12 catchments in Region 1 and
- the 17 catchments in Region 2.

Based on the previous studies by Meigh et al. (1997) and Mkhandi and Kachroo (1997) and similar studies from other parts of the world, e.g. NERC (1975), AREA is the single most important variable and should always be included in the models. In a similar study of MAF from New Zealand, Rosbjerg and Kjeldsen (2000) found that combining AREA with MAP as AREA*MAP gave significantly better results than using AREA alone. Different combinations of AREA and AREA*MAP with other catchment properties, as shown in Table 1, were tested in order to identify the best combination of independent variables. According to Bowerman and O'Connell (1990), minimum prediction variance is a better criterion for comparing regression models than the coefficient of determination R^2. Hence, the model with the lowest average prediction variance $\overline{\sigma}_\theta^2$ was chosen, where:

$$\overline{\sigma}_\theta^2 = 1/M \sum_{i=1}^{M} \hat{\sigma}_{\theta i}^2 \qquad (20)$$

with M being the number of catchments included in the analysis and $\sigma_{\theta i}^2$ calculated through Eq. (16). The selected regression equations estimated using OLS, WLS, GLS1 and GLS2 to the three regions, the model error variance, the average sampling error variance and the average prediction variance $\overline{\sigma}_\theta^2$ are shown in Table 2.

Different combinations of catchment characteristics were found to give models with the lowest predictive variance in each of the three tested regions. Using all catchments, the MAF is best predicted using only AREA. In Region 1 both MAP*AREA and CONC should be used. Finally, in Region 2 AREA and GCC gave the lowest average prediction variance. Addition of MAP in Region 2 gave a lower model error, but the average prediction variance increased and, hence, MAP was omitted.

Smaller model error variance was obtained when the models were estimated for Regions 1 and 2 rather than using all catchments, especially in Region 2, whereas only a small improvement was observed in Region 1. Based on the average predictive variance as calculated through Eqs. (17) and (20), it would be preferable to use all catchments when estimating MAF in Region 1, but only data from Region 2 should be used when estimating MAF in Region 2. A separation according to AREA as suggested by Meigh et al. (1997) did not improve the average prediction variance and, hence, was not used.

The use of all gauging stations to improve model prediction in Region 1 should be viewed with some caution and is more likely to be an indication that the current model type is inappropriate to model MAF in that region.

For all three regions the WLS method gives the lowest values of model error variance, average sampling error variance and, hence, the lowest average total prediction error variance. Correspondingly, the OLS method gives the highest model error variance and total average prediction error variance in all three regions, but not the highest average sampling error variance in Region 1, as it does in the other two regions. Also in Region 1 the GLS2 method performs worse than the GLS1 method with the opposite in the other two regions.

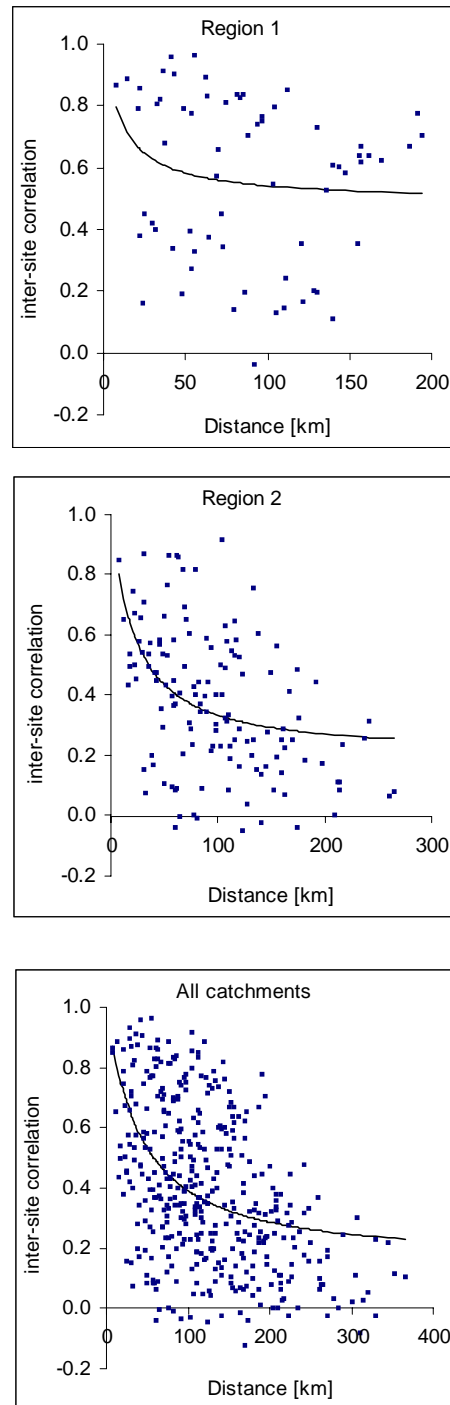In all three regions, the model parameters estimated using OLS,

Figure 2
Correlation-distance relationship for a) all catchments, b) Region 1 and c) Region 2 estimated using Eq. (18)

WLS and GLS1 appear to be of similar magnitude compared to the parameters obtained through the GLS2 method.

Interestingly, the biggest difference between model parameters estimated using GLS1 and GLS2 is observed in Region 1 where the correlation-distance relationship appears least significant.

To compare the results from this study with similar OLS models reported in the literature by van Bladeren (1993), Mkhandi and Kachroo (1997) and Meigh et al. (1997), the R^2 statistic was calculated for the OLS models.

Available on website **http://www.wrc.org.za**

Model type	Regression model	Model error	Sampling error	Predictive variance
	All 29 catchments			
OLS	$\ln[\mu] = -0.0922 + 0.7916\ln[AREA]$	0.2769	0.2960	0.5729
WLS	$\ln[\mu] = -0.0825 + 0.7895\ln[AREA]$	0.2396	0.2588	0.4984
GLS1	$\ln[\mu] = -0.1004 + 0.7904\ln[AREA]$	0.2542	0.2883	0.5425
GLS2	$\ln[\mu] = -0.0661 + 0.7842\ln[AREA]$	0.2477	0.2812	0.5289
	Region 1			
OLS	$\ln[\mu] = 25.0873 + 1.0776\ln[AREA \cdot MAP]$ $-8.8028\ln[CONC]$	0.2627	0.3283	0.5910
WLS	$\ln[\mu] = 25.0105 + 1.0728\ln[AREA \cdot MAP]$ $-8.7673\ln[CONC]$	0.2284	0.2945	0.5229
GLS1	$\ln[\mu] = 25.0534 + 1.0728\ln[AREA \cdot MAP]$ $-8.7865\ln[CONC]$	0.2308	0.3364	0.5672
GLS2	$\ln[\mu] = 25.3351 + 1.0792\ln[AREA \cdot MAP]$ $-8.8733\ln[CONC]$	0.2325	0.3396	0.5721
	Region 2			
OLS	$\ln[\mu] = 1.2735 + 0.7392\ln[AREA]$ $-1.3760\ln[GCC]$	0.1340	0.1576	0.2916
WLS	$\ln[\mu] = 1.3059 + 0.7348\ln[AREA]$ $-1.3899\ln[GCC]$	0.1140	0.1377	0.2517
GLS1	$\ln[\mu] = 1.2833 + 0.7365\ln[AREA]$ $-1.3862\ln[GCC]$	0.1213	0.1525	0.2738
GLS2	$\ln[\mu] = 1.2388 + 0.7295\ln[AREA]$ $-1.2763\ln[GCC]$	0.1164	0.1467	0.2631

Using all catchments, a R^2 value of 0.707 was obtained, which is similar to the results obtained by Mkhandi and Kachroo (1997). Dividing the catchments into Regions 1 and 2 significantly enhances the modelling of MAF, with an R^2 value of 0.814 in Region 1 and a R^2 value of 0.841 in Region 2. The R^2 values obtained by van Bladeren (1993) for the coastal area of KZN, ranging from $R^2 = 0.91$ to 0.94, are larger than the values found for this study. However, significantly better results were obtained in the western north-western parts of KZN in this study with $R^2 = 0.841$ compared to $R^2 = 0.66$ found by Van Bladeren (1993).

As mentioned previously, estimates of the index-flood parameter at ungauged sites using regression models are not as reliable as estimates based on available observed data. To illustrate the precision of estimation at ungauged sites compared to estimation at gauged sites, Fig. 3 shows the X_T-T relationship at each of the 17 sites in Region 2 estimated using the index-flood method, as defined by Eq. (1), with the GNO distribution as the regional flood frequency distribution for two different situations:

- at gauged sites, where $\hat{\mu}_i$ is estimated through available site data and the regional growth curve is estimated using all available data, including the at-site data,
- at ungauged sites using a hidden station approach, where data from the site under consideration were removed and the $\hat{\mu}_i$ is estimated using the GLS2 method of the remaining catchments.

The regional growth curve for Region 2 as estimated by Kjeldsen et al. (2000), adopting the GNO distribution, is given:

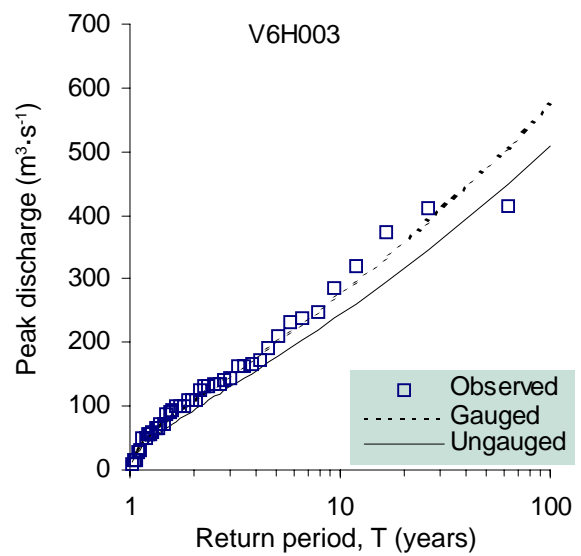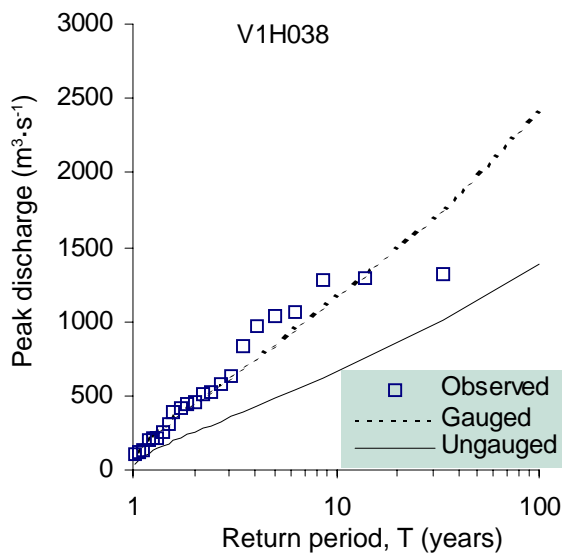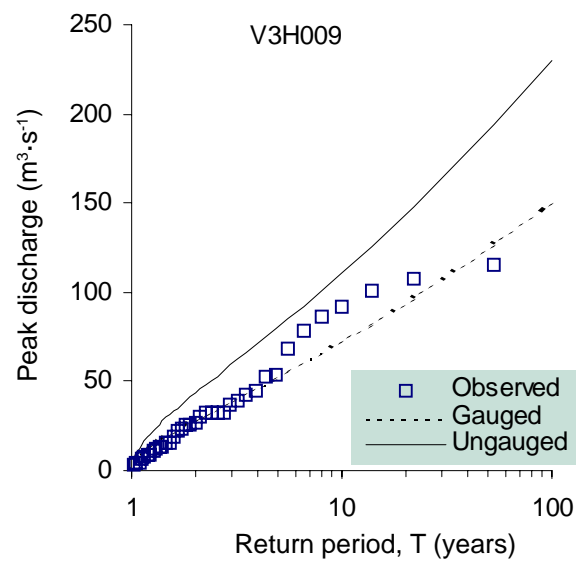$$z_T^{'} = 0.774 - 0.891\left[1 - \exp\left(0.672\Phi^{-1}\left(1 - \frac{1}{T}\right)\right)\right] \quad (21)$$
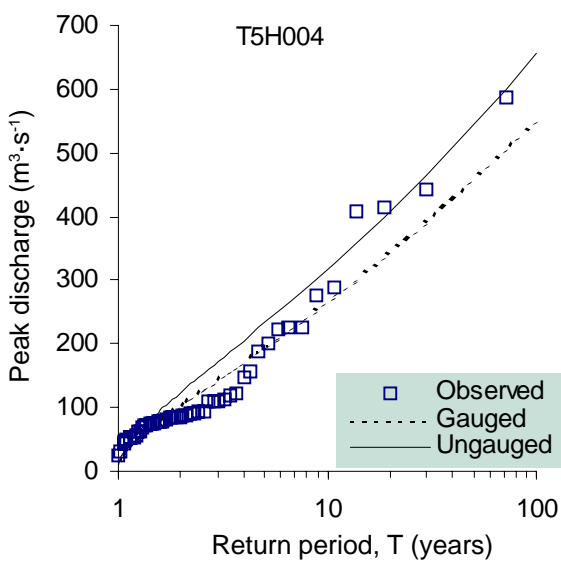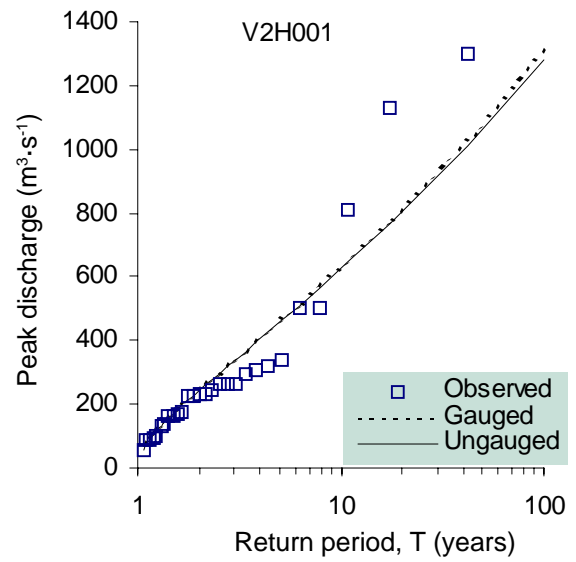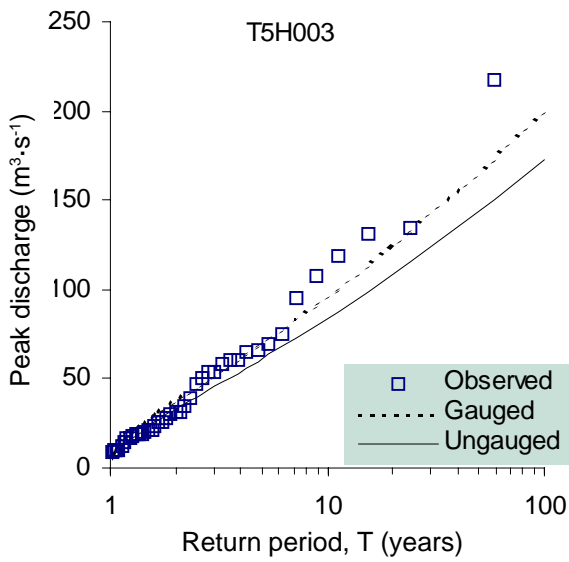
Figure 3

Probability plots comparing estimation at gauged and ungauged sites in Region 2

V1H001

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)

V3H002

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)

V1H009

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)

V6H004

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)

V2H005

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)

V7h012

Peak discharge (m³·s⁻¹)

- Observed
- Gauged
- Ungauged

Return period, T (years)
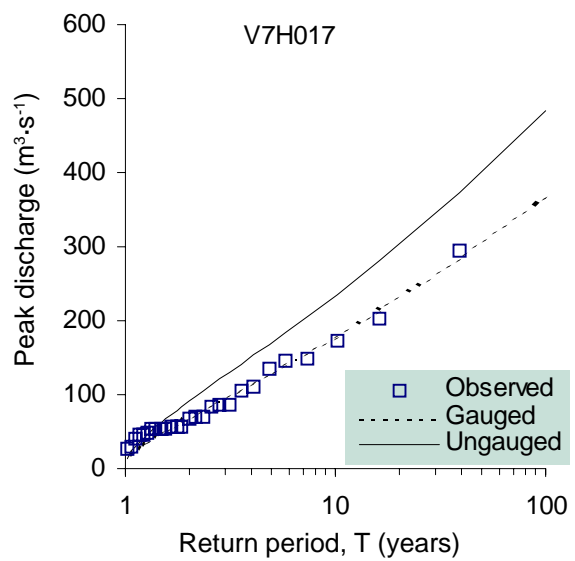
Figure 3 (continued)
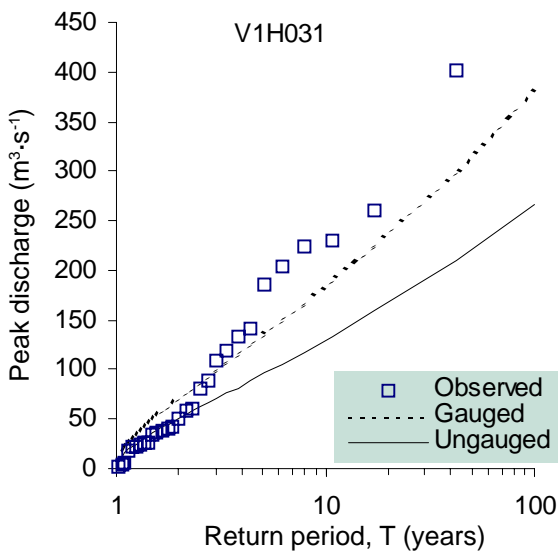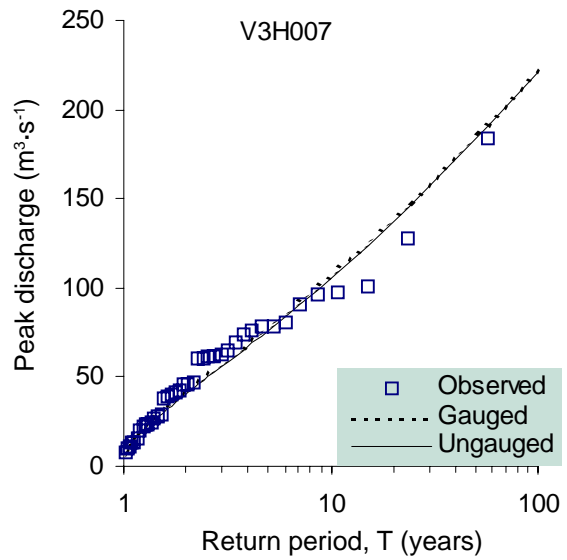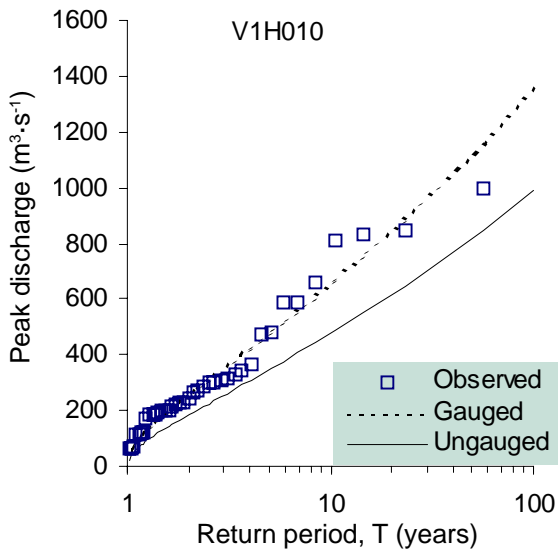
Return Period, T (years)

Figure 3 (continued)

where Φ is the cumulative distribution function for the standard normal distribution.

In the hidden area approach, the exclusion of the at-site data will also have an effect on the regional growth curve. However, this effect is negligible compared to the effect on the index-flood parameter. For most sites the estimation at ungauged sites appears to give acceptable estimates of the index flood parameter, except for site V3H005. This particular site has low unit runoff compared to the other catchments in the region. Further studies should investigate the cause of this low runoff response and if additional catchment characteristics will improve the modelling results.

Discussion and conclusions

This study investigated the use of log-linear models for predicting MAF from catchment characteristics, and the use of OLS, WLS, GLS1 and GLS2 respectively for the estimation of model parameters.

The separation of all catchments into the two regions significantly improved the predictive ability in Region 2. In Region 1, where the sampling variability of the MAF is high, the predictive

ability decreased compared to using all catchments, illustrating the high degree of sampling variability in the AMS of floods from this region. In Region 2 the models developed in this study performed better than previously reported models.

The model parameters estimated using OLS, WLS and GLS1 varied only little, with GLS2 giving different sets of parameters. The GLS2 may be used with the most confidence, as the fundamental conditions of the OLS method, homoscedastic and independent observations, are seriously violated. In all regions the two contributions to the total prediction variance are of similar magnitude, with the average sampling error variance contributing slightly more than the model error variance. Hence, improvement of the equations obtained in this study can be obtained through a reduction of both components by:

- reduction of average sampling error variance by identification and inclusion of more catchments and
- reduction of model error variance through improved equations, e.g. by inclusion of other catchment characteristics.

Based on the results from this paper we conclude that the index-flood method should not be applied to Region 1 before the fundamental modelling problems have been solved.

Acknowledgement

References

ALEXANDER WJR (1990) *Flood Hydrology of Southern Africa*. South African Committee for Large Dams, Pretoria, RSA.

BOWERMAN BL and O'CONNELL RT (1990) *Linear Statistical Models: An Application Approach* (2nd edn.) PWS-KENT Publishing Company, Massachusetts, USA. 1024 pp.

DWAF (1990) List of Hydrological Gauging Stations, July 1990. Dept. of Water Affairs, Pretoria, RSA.

HEBSON CS and CUNNANE C (1987) Assessment of use of at-site and regional flood data for flood frequency estimation. In: V P Singh (ed.) *Hydrologic Frequency Modelling*, D. Reidel Norwell, Massachusetts. USA. 433-448.

HOSKING JRM and WALLIS JR (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge Univ. Press, UK. 224 pp.

KJELDSEN TR, SMITHERS JC and SCHULZE RE (2000) Regional Flood Frequency Analysis in KwaZulu-Natal Province, South Africa, Using the Index Flood Method. Unpublished report, School of Bioresources Eng. and Environ. Hydrol., Univ. of Natal, RSA.

MADSEN H and ROSBJERG D (1997) Generalized least square and empirical Bayes estimation in regional partial duration series index-flood modeling. *Water Resour. Res.* **33** (4) 771-781.

MEIGH JR, FARQUHARSON FAK and SUTCLIFF JV (1997) A worldwide comparison of regional estimation methods and climate. *Hydrol. Sci. J.* **42** (2) 225-244.

MKHANDI S and KACHROO S (1997) Regional flood frequency analysis for Southern Africa. In: *Southern African FRIEND*, Technical Documents in Hydrology No. **15**, UNESCO, Paris, France.

NATURAL ENVIRONMENT RESEARCH COUNCIL (NERC) (1975) Flood Studies Report, Vol I. Hydrological Studies, London, UK.

ROSBJERG D and KJELDSEN TR (2000) Comparison of regional index-flood estimation procedures. Paper presented at EGS Gen. Assem., April 2000 Nice, France.

SCHULZE RE (1995) Hydrology and Agrohydrology: A Text to Accompany the *ACRU* 3.00 Agrohydrological Modelling System. Water Research Commission Report TT 69/95, Pretoria, RSA.

SCHULZE RE (1997) South African Atlas of Agrohydrology and –Climatology. Water Research Commission Report No TT 82/96.

STEDINGER JR (1983) Estimation of a regional flood frequency distribution. *Water Resour. Res.* **19** (2) 503-510.

STEDINGER JR and TASKER GD (1985) Regional hydrologic analysis 1. Ordinary, weighted and general least square compared. *Water Resour. Res.* **21** (9) 1421-1432.

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, USA.

TASKER GD and STEDINGER JR (1989) An operational GLS model for hydrologic regression. *J. Hydrol.* **111** 361-375.

VAN BLADEREN D (1993) Application of historical flood data in flood frequency analysis for the Natal and Transkei regions. In: *Proc. of the 6th S. Afr. Natl. Hydrol Symp Vol I*, Pietermaritzburg, RSA.

Appendix A

The General Normal (GNO) distribution, also known as the three parameter log-normal (LN3) distribution, has a long and proven record in hydrology. According to Hosking and Wallis (1997) the probability density function (pdf) of GNO can be expressed as:

$$f(x) = \frac{\exp(ky - y^2/2)}{\alpha\sqrt{2\pi}},$$

$$y = \begin{cases} -k^{-1}\ln[1 - k(x-\xi)/\alpha], & k \neq 0 \\ (x-\xi)/\alpha, & k = 0 \end{cases} \quad\text{(A1)}$$

and the corresponding cdf:

$$F(x) = \Phi(y) \quad\text{(A2)}$$

where:
  Φ = the cdf of the standard normal distribution

The p-th fractile of GNO, $x_p = F^{-1}(p)$, is calculated as:

$$x_p = \begin{cases} \xi + \alpha[1 - \exp(-k\Phi^{-1}(p))], & k \neq 0 \\ \xi + \alpha\Phi^{-1}(p), & k = 0 \end{cases} \quad\text{(A3)}$$

The three parameters ζ (location), (scale) α and k (shape) can be estimated using Method of L-Moments as presented by Hosking and Wallis (1997). First, estimate k as:

$$k \approx \tau_3 \frac{E_0 + E_1\tau_3^2 + E_2\tau_3^4 + E_3\tau_3^6}{1 + F_1\tau_3^2 + F_2\tau_3^4 + F_3\tau_3^6} \quad\text{(A4)}$$

by inserting sample L-Skew for τ_3. The E_i and F_i coefficients are shown in Table A1. Next, α and k are estimated as:

$$\alpha = \frac{\lambda_2 k \exp(-k^2/2)}{1 - 2\Phi(-k/\sqrt{2})} \quad\text{(A5)}$$

$$\xi = \lambda_1 - \frac{\alpha}{k}\left[1 - \exp\left(k^2/2\right)\right] \qquad \text{(A6)}$$

by inserting the first and second order sample L-moments and the estimated k from Eq. (A4) instead of λ_1, λ_2 and k respectively.

Table A1 Coefficients for estimation of *k* (Hosking and Wallis, 1997)	
$E_0 = 2.0466534$ $E_1 = -3.6544371$ $E_2 = 1.8396733$ $E_3 = -0.20360244$	$F_1 = -2.0182173$ $F_2 = 1.2420401$ $F_3 = -0.21741801$

Available on website **http://www.wrc.org.za**