Hydrological Extremes: Understanding, Predicting, Mitigating (Proceedings of IUGG 99 Symposium HS1, Birmingham, July 1999). IAHS Publ. no. 255, 1999.

Predicting short duration design storms in South Africa using inadequate data

JEFFREY SMITHERS, ROLAND SCHULZE
Department of Agricultural Engineering, University of Natal, Pietermaritzburg, South Africa
e-mail: smithers@agua.ccwr.ac.za

GEOFFREY PEGRAM
Department of Civil Engineering, University of Natal, Durban, South Africa

Abstract Relatively few recording raingauges with long, reliable records are available in South Africa for the estimation of short duration (≤24 h) design rainfalls. Hence approaches to estimating short duration design rainfalls have been developed which are based on daily rainfall data measured by standard, non-recording raingauges at fixed 24 h periods ending at 08:00. These approaches include the use of regional frequency analyses, scaling the moments of the extreme events and stochastic modelling of the rainfall process. This paper presents results from estimating short duration design rainfalls at selected sites in South Africa, using synthetic rainfall series generated by a stochastic model with parameters derived from daily rainfall data.

INTRODUCTION

Engineers and hydrologists involved in the design of hydraulic structures (e.g. culverts, bridges, dam spillways and reticulation for drainage systems) need to assess the frequency and magnitude of extreme rainfall events in order to generate design flood hydrographs. Many thousands of engineering and conservation design decisions, made annually both in South Africa and internationally and which involve large sums of money for construction, require accurate short duration (≤24 h) design rainfall intensity information. Depth–Duration–Frequency (DDF) relationships, which utilize recorded events in order to predict future exceedance probabilities and thus quantify risk and maximize design efficiencies, are a key concept in the design of hydraulic structures (Schulze, 1984).

Estimates of design rainfall for durations shorter than one day were last comprehensively produced for South Africa in the late 1970s and early 1980s using values manually extracted from autographic charts (Midgley & Pitman, 1978; Van Heerden, 1978; Adamson, 1981), and in the mid 1980s using digitized data for selected stations in the province of KwaZulu-Natal (Schulze, 1984). Subsequent to these studies, the entire autographically recorded South African Weather Bureau (SAWB) database has been digitized. However, the record lengths of all the available digitized data, including those from other organizations, are relatively short with only 55 out of a total of 412 recording rainfall stations in South Africa having record lengths longer than 30 years and only nine stations with record lengths greater than 50 years. Thus the network of recording rainfall stations in South Africa with record lengths longer than

30 years is very sparse. In addition, it has been noted by Smithers (1993, 1998) that the SAWB digitized data, which make up the majority of the data in the database, contain numerous errors, inconsistencies and periods of missing data. For example, in numerous cases, the 24-h design rainfall values computed from the digitized database are smaller than those computed from the rainfall recorded manually at daily intervals. An in-depth study by Smithers (1998) has shown that the majority of short duration rainfall data in the database are viewed as unreliable and the challenge has been to develop techniques to estimate short duration design rainfalls using a limited number of reliable recording rainfall stations.

Smithers (1998) adopted three approaches to overcome these limitations in the short duration database. The first approach used a regional frequency analysis, the second investigated scaling relationships of the moments of the extreme events and the third approach used a stochastic intra-daily model to generate synthetic rainfall series. A common theme in all three approaches is the development of techniques to estimate short duration design storms from the daily rainfall database, which contains rainfall data recorded manually at daily intervals, and is deemed to be more reliable than the short duration rainfall data. An added advantage of using the daily rainfall data is the large number (>6000), relative to the digitized data, of stations in South Africa with long record lengths.

The objective in this paper is to present initial results of the third approach which evaluates the feasibility of using a stochastic Bartlett-Lewis Rectangular Pulse Model, with parameters determined using only the daily data, for the estimation of design rainfalls for durations $\leq$ 24 h at selected sites in South Africa.

BARTLETT-LEWIS RECTANGULAR PULSE MODELS

Stochastic cluster-based rainfall models, such as the Bartlett-Lewis Rectangular Pulse Model (BLRPM) and the Neyman-Scott Rectangular Pulse Model (NSRPM) have been shown to be able to model rainfall characteristics over time scales ranging from 1 to 24 h (Rodriguez-Iturbe *et al.*, 1987a,b; Entekhabi *et al.*, 1989; Onof & Wheater, 1993; Bo *et al.*, 1994; Velghe *et al.*, 1994; Cowpertwait *et al.*, 1996; Khaliq & Cunnane, 1996; Verhoest *et al.*, 1997). The potential of using cluster-based models in the estimation of design rainfall events has been demonstrated *inter alia* by Onof & Wheater (1993; 1994), Khaliq & Cunnane (1996), Cowpertwait *et al.* (1996) and Verhoest *et al.* (1997).

The algorithm for the Modified Bartlett-Lewis Rectangular Pulse Model (MBLRPM), developed by Rodriguez-Iturbe *et al.* (1988), which is an extension to the BLRPM (Rodriguez-Iturbe *et al.*, 1987a), is described by Entekhabi (1989), Onof & Wheater (1993), Bo *et al.* (1994), Onof & Wheater (1994) and Khaliq & Cunnane (1996) as:

- a Poisson process (parameter λ) used to model arrival rate of storm origins,
- storm origins are followed by a Poisson process of cell origins with rate parameter β,
- the process of new cell origins terminates after a time that is exponentially distributed with parameter γ, i.e. the storms have an exponentially distributed duration with parameter γ,

- the duration of the rectangular pulse of each cell is exponentially distributed with parameter η, and for distinct storms are assumed to be independent variables from a gamma distribution with index α and scale parameter ν, i.e $E(\eta) = \alpha/\nu$ and $\mathrm{var}(\eta) = \alpha/\nu^2$,
- each cell is a random constant exponentially distributed with mean μ_x, and
- the number of cells per storm C has a geometric distribution with a mean of

$$\mu_C = 1+\kappa/\phi \tag{1}$$

where κ and ϕ are dimensionless parameters and

$$\kappa = \beta/\eta \tag{2}$$

$$\phi = \gamma/\eta \tag{3}$$

By keeping κ and ϕ constant, the mean and variance of different storms change randomly from storm to storm. Hence the mean inter-arrival interval time of cells (β^{-1}) and mean storm duration (γ^{-1}) also change randomly. For brevity, the mean, variance, autocorrelation and dry probability equations for MBLRPM are not repeated here, but the reader can refer to, for example, Khaliq & Cunnane (1996) for a full description of the model. In order to improve the overestimation of daily autocorrelations and extreme events noted by Onof & Wheater (1993), Onof & Wheater (1994) replaced the exponential distribution of cell rainfall intensity in the MBLRPM by a two parameter gamma distribution which would give greater flexibility in the simulation of extreme events. This modified version of the BLRPM is termed the Bartlett-Lewis Rectangular Pulse Gamma model (BLRPGM). Again, for brevity, the reader is referred to Onof & Wheater (1994) for a detailed description of the model.

Smithers (1998) performed a detailed comparison between the MBLRPM and BLRPGM at selected sites in South Africa and concluded that the BLRPGM was the more suitable model. In addition, Smithers (1998) investigated the inter-relationships between the parameters for each model and found strong correlations between some parameters which enabled the fixing of one of more parameters when determining the month-by-month parameters using the method of moments.

A number of different sets of moments were evaluated by Smithers (1998) for the estimation of model parameters assuming that only the daily rainfall data were available for determining the model parameters. Using only daily rainfall data, an improved performance of the BLRPGM was obtained when the variances for durations < 24 h, which are derived from the daily data assuming a power law relationship between variance and duration, are included in addition to the 24-h mean, 24- and 48-h variances, 24-h lag-1 autocorrelations and 24- and 48-h dry probabilities in the set of moments used to determine the parameters of the model. The results of estimating short duration design rainfalls from synthetic series of rainfall generated by the BLRPGM at selected sites in South Africa are presented in the following section.

RESULTS

The location of some sites in different climatic regions in South Africa where the performance of the BLRPGM was evaluated is shown in Fig. 1. The selected sites

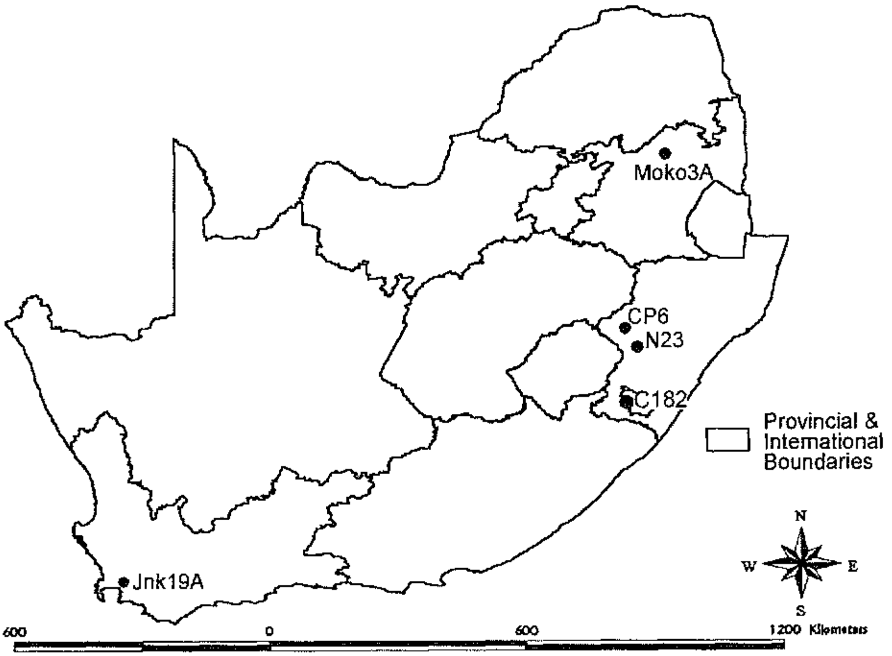

Fig. 1 Locations of stations used as case studies of the performance of the BLRPGM.

include stations from the summer rainfall region with high Mean Annual Precipitation (*MAP*), CP6, and from lower *MAP*s (N23, C182 and Moko3A), as well as from the winter rainfall region (Jnk19A). All these stations were judged to have reliable data (Smithers, 1998).

For each evaluation at a particular site one hundred sets of synthetic rainfall series were generated, each with a record length equal to that of the historical data. The performance of the model was assessed using two measures. In the first measure, seven moments and statistics (mean, standard deviation, autocorrelation, dry probability, duration of wet and dry periods, and the number of events) of the synthetic data were compared to the corresponding characteristics of the historical data. More emphasis is placed on the second measure of performance of the model where design storms estimated from the synthetic rainfall series were compared to those computed from the historical data. Again for brevity, and with the focus of this paper on the estimation of design rainfall values, only results for the second measure of performance are presented.

For the observed data and for each of the 100 synthetic series generated by the BLRPGM, design rainfall depths were estimated using the General Extreme Value (GEV) distribution, which was found by Smithers (1996) to be the most appropriate distribution to use for short duration rainfalls in South Africa, fitted to the Annual Maximum Series (AMS) by L-moments. Design values for 2-, 5-, 10-, 20-, 50- and 100-year return periods were computed for rainfall durations of 0.25. 0.5, 1, 2, 3, 4, 5, 6, 9, 12, 15, 18 and 24 h. For each duration and return period, a frequency analysis was performed on the 100 values computed from the synthetic rainfall series generated by the model. Histograms depicting the observed design rainfall computed from the historical data and High-Low bars depicting the interquartile range (25th to 75th nonexceedance percentiles) of the 100 synthetic data sets is one method of visually evaluating the adequacy of the model. For example, the performance of the BLRPGM at raingauge N23, with parameters derived only from the daily data, is shown in Fig. 2 for the best (February) and worst (December) rainy season months simulated and for annual totals. Note the difference in design values for the 1440-min periods, computed using a sliding 24-h window, and 1 day intervals, computed using a fixed 24-h window ending at 08:00.

At raingauge N23, the design rainfall values estimated from the synthetic rainfall series generated by the BLRPGM, with parameters derived using only daily rainfall data, generally represented the individual months and annual historical values well for durations > 1 h and for return periods < 50 years. In order to objectively assess the performance of the BLRPGM these results were summarized as the Mean Absolute Relative Error (*MARE*) to include rainy season months and annual totals, return periods ranging from 2 to 50 years and selected durations.

$$MARE = \frac{100}{N_M \times N_L \times N_{RP}} \times \sum_{i=1}^{N_M} \sum_{j=1}^{N_L} \sum_{k=1}^{N_{RP}} \left(\frac{\left| S_{(i,j,k)} - O_{(i,j,k)} \right|}{O_{(i,j,k)}} \right) \tag{4}$$

where *MARE* = mean absolute relative error of design rainfall (%), $S_{(i,j,k)}$ = mean *k*th return period, *j*th hour design rainfall computed for *i*th period from model generated rainfall series, $O_{(i,j,k)}$ = *k*th return period, *j*th hour design rainfall computed for *i*th period from observed data, N_M = number of periods (7), 1 to 6 = rainy season months and 7 = annual period, N_L = number of aggregation levels, and N_{RP} = number of return periods (2, 5, 10, 20, 50).

The *MARE* values, computed using the BLRPGM for durations ≤ 1 h (15, 30 and 60 min) as well *MARE* values for durations > 1 h are shown in Fig. 2 for the selected test sites. Clearly the use of the BLRPGM to estimate design rainfalls for short durations (≤1 h), when only daily data are used to determine the model parameters, results in unacceptably large *MARE* values. However, for durations > 1 h the *MARE* values for most sites are less than 20%, which is deemed to be an acceptable value.

DISCUSSION AND CONCLUSIONS

Synthetic rainfall series generated by the BLRPGM, with parameters derived using only daily rainfall data, have been used to estimate short duration design rainfall values at selected sites in South Africa. The stochastic variability of design rainfall values estimated using this approach has been demonstrated and, assuming the model to be adequate, then the interquartile range of the simulated values is postulated to be equivalent to a confidence interval of the design value.

At the sites where the model was evaluated, the estimation of design rainfall values from the synthetic rainfall series were considered to be good for durations > 1 h. Although the simulated design rainfall values coincided well with the observed values for durations ≤ 1 h for some months at some sites, the performance of the model for these shorter durations was considered not to be adequate and is thus not

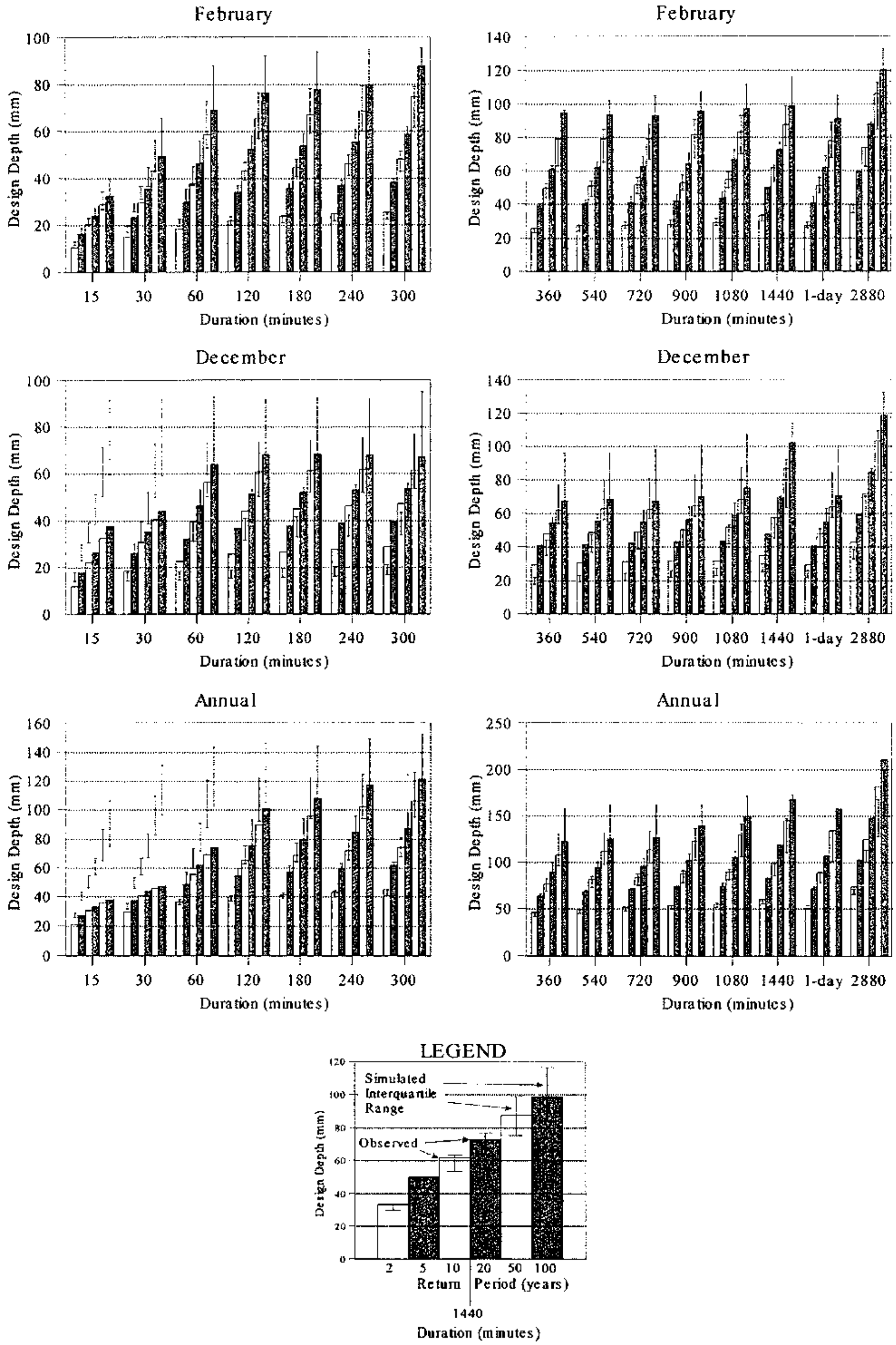

Fig. 2 Design rainfall at raingauge N23 estimated using the BLRPGM with parameters determined from daily rainfall data.

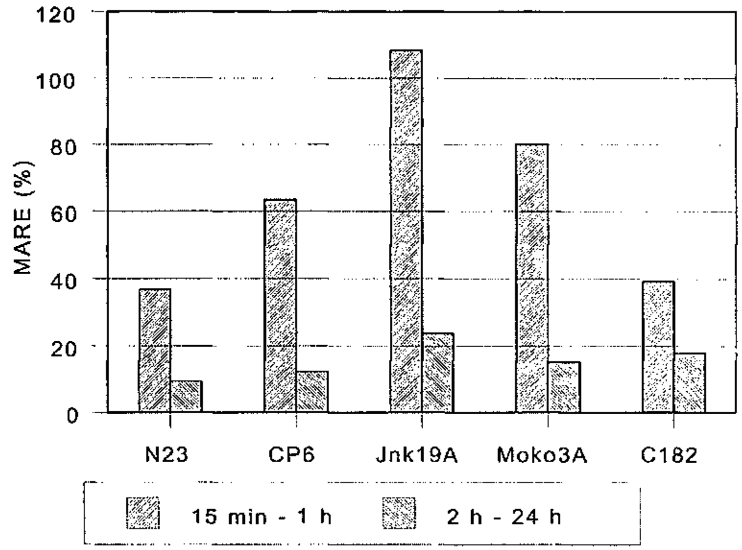

Fig. 3 Assessment of the estimation of design rainfall values for shorter and longer durations using the BLRPGM at selected sites in South Africa.

recommended. Although the sites evaluated cover a range of climatic regions in South Africa, not all regions are included. In particular, stations on the eastern Coast of South Africa, where differing meteorological conditions are known to produce extreme events (Pegram & Adamson, 1988), need evaluation because the model assumes a stationary data series.

The relatively larger *MARE* values for Jnk19A and Moko3A shown in Fig. 3 are postulated to be the result of the inadequate estimation of shorter duration variances from the daily rainfall data. The successful use of only the daily rainfall data to estimate the parameters of the BLRPGM means that the model can be applied at any site where a reasonable period of daily rainfall data is available. This substantially increases the network of stations in South Africa where the model can be applied and where short duration design rainfalls can be estimated. Thus, the use of the BLRPGM, with parameters derived from the daily rainfall data which have a denser network of raingauges and are considered to be more reliable than the digitized rainfall data, enables the estimation of short duration design rainfall values down to 1-h durations at sites where only daily rainfall data are available or where the digitized data are deemed to be unreliable. It is postulated that the stochastic modelling of rainfall based approach to the estimation short duration design rainfall in South Africa can, in conjunction with regional and scaling based approaches, overcome the limitations of the short duration rainfall database and result in robust and reliable estimates of design rainfall.

Acknowledgements The authors wish to thank the Water Research Commission, who were the major funders of this study and the University of Natal Research Fund who also provided some funding. The organizations providing data to the project, in particular the South African Weather Bureau and Council for Scientific and Industrial Research, are thanked. The computing facilities and assistance provided by the

Computing Centre for Water Research at the University of Natal and the Super Parallel Computing Centre at the University of Potchefstroom are gratefully acknowledged.

REFERENCES

Adamson, P. T. (1981) Southern African storm rainfall. *Tech. Report no. TR102, Dept of Water Affairs, Pretoria, RSA.*

Bo, Z., Islam, S. & Eltahir, E. A. B. (1994) Aggregation–disaggregation properties of a stochastic rainfall model. *Wat. Resour. Res.* **30**(12), 3423–3435.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. & Mawdsley, J. A. (1996) Stochastic point process modelling of rainfall. 2. regionalisation and disaggregation. *J. Hydrol.* **175**, 47–65.

Entekhabi, D., Rodriguez-Iturbe, I. & Eagleson, P. S. (1989) Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: parameter estimation and validation. *Wat. Resour. Res.* **25**(2), 295–302.

Khaliq, M. N. & Cunnane, C. (1996) Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. *J. Hydrol.* **180**, 109–138.

Midgley, D. C. & Pitman, W. V. (1978) A depth–duration–frequency diagram for point rainfall in Southern Africa. *HRU Report 2/78, University of Witwatersrand, Johannesburg, RSA.*

Onof, C. & Wheater, H. S. (1993) Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *J. Hydrol.* **149**(1–4), 67–95.

Onof, C. & Wheater, H. S. (1994) Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *J. Hydrol.* **157**, 177–195.

Pegram, G. G. S. & Adamson, P. (1988) Revised risk analysis for extreme storms and floods in Natal/Kwazulu. *The Civil Engineer in South Africa* January, 15–20, and discussion July, 331–336.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. (1987a) Some models for rainfall based on stochastic point processes. *Proc. Roy. Soc. London* **A410**, 269–288.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. (1988) A point process model for rainfall: further developments. *Proc. Roy. Soc. London* **A417**, 283–289.

Rodriguez-Iturbe, I., Febres de Power, B. & Valdes, J. B. (1987b) Rectangular pulses point process models for rainfall: analysis of empirical data. *J. Geophys. Res.* **92**(D8), 9654–9656.

Schulze, R. E. (1984) Depth–duration–frequency studies in Natal based on digitised data. In: *South African National Hydrol. Symp., Tech. Report TR119, Department of Environment Affairs, Pretoria, RSA.*

Smithers, J. C. (1993) The effect on design rainfall estimates of errors in the digitised rainfall database. In: *Proc. Sixth South African National Hydrological Symposium* (ed. by S. A. Lorentz, S. W. Kienzle & M. C. Dent), 95–102. SANCHIAS, Pretoria, RSA.

Smithers, J. C. (1996) Short-duration rainfall frequency model selection in Southern Africa. *Water SA* **22**(3), 211–217.

Smithers, J. C. (1998) An evaluation of techniques for estimating short duration design rainfalls in South Africa. *Report to Water Research Commission, Pretoria, South Africa.* Dept of Agricultural Engineering, University of Natal, Pietermaritzburg, South Africa.

Van Heerden, W. M. (1978) Standaard intensiteitkrommes vir reënval van kort duurtes. *The Civil Engineer in South Africa* October, 261–268.

Velghe, T., Troch, P. A., De Troch, F. P. & Van De Velde, J. (1994) Evaluation of cluster-based rectangular pulses point process models for rainfall. *Wat. Resour. Res.* **30**(10), 2847–2857.

Verhoest, N., Troch, P. E. & Troch, F. P. (1997) On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point. *J. Hydrol.* **202**, 108–120.