# The development and assessment of a regionalised daily rainfall disaggregation model for South Africa

## DM Knoesen* and JC Smithers

*School of Bioresources Engineering and Environmental Hydrology, University of KwaZulu-Natal, P/Bag X01, Scottsville, South Africa*

## Abstract

The temporal distribution of rainfall, viz. the distribution of rainfall intensity during a storm, is an important factor affecting the timing and magnitude of peak flow from a catchment and hence the flood-generating potential of rainfall events. Rainfall intensity is also one of the primary inputs into hydrological models used for the design of hydraulic structures. In the absence of continuously recorded rainfall data, one method of estimating the temporal distribution of rainfall is to disaggregate coarser-scale data into a finer resolution, e.g. from daily data into hourly rainfall information. In this study, a daily to hourly disaggregation model developed in Australia, and modified for application in South Africa, is used. However, this model requires input obtained from short-duration data at the desired location. Owing to the paucity of short-duration data in South Africa, the methodology is regionalised to enable the application of the model at a national scale, particularly at locations where only daily data are available. The regionalised model was independently tested at 15 locations in differing climatic regions in South Africa. At each location, observed hourly data were aggregated to yield daily values and were then disaggregated using the methodology. Results show that the regionalised model is capable of replicating the results obtained when 'at-site' short duration rainfall data are used as input to the disaggregation model, and is able to retain the daily totals and the statistical characteristics of the hourly rainfall.

Keywords*:* temporal rainfall disaggregation, regionalisation, South Africa

## Introduction

Continuous-simulation hydrological models are important tools when analysing complex hydrological or hydraulic problems where issues need to be investigated at different timescales, for example, in flood prediction and the modelling of water quality (Mikkelsen et al., 1998). These models require detailed rainfall data, viz. hourly or sub-hourly. The advantage of such a time-series is that they reflect all relevant rainfall characteristics from peak intensities associated with short duration to variations in annual rainfall (Mikkelsen et al., 1998). However, hydrological data are generally only widely available at more aggregated levels, such as daily. Koutsoyiannis and Onof (2001) note that in many countries the number of rain-gauges providing hourly or sub-hourly resolution data is smaller than the number of daily gauges by about an order of magnitude. This situation reflects a paucity of rainfall data for timescales of one hour or less, both in the number of gauges and length of the recorded series (Koutsoyiannis and Onof, 2001). This, too, is the case in South Africa where it was reported in 2000 that there were 172 recording gauges with at least 10 years of breakpoint data (Smithers and Schulze, 2000a), compared to 1806 daily rainfall stations with at least 40 years of data (Smithers and Schulze, 2000b).

The need for a model to disaggregate daily rainfall into a sequence of individual storms of finer timescale cannot be over-emphasised (Gyasi-Agyei, 1999). The objectives of this paper are to provide a brief background to the methodology, detail the regionalisation performed and to assess the performance of a regionalised rainfall disaggregation model for South Africa.

---

## Rainfall data used in study

Hourly rainfall data from 172 recording stations in South Africa, all of which had record lengths greater than 10 years, were available for use (Smithers and Schulze, 2000a). It was necessary to exclude some stations from the model development process in order to evaluate the model independently. One station from each of the 15 relatively homogeneous extreme rainfall clusters, identified and described by Smithers and Schulze (2000a), was removed from the dataset and not used in the development of the disaggregation model. In order to select the stations for testing, the 172 stations were divided into their respective homogeneous clusters and the station with the median record length in each cluster was excluded from the development of the model. This resulted in 15 test stations, located as shown in Fig. 1. The locations of the remaining 157 stations, which were used for model development, are shown in Fig. 2.

### The disaggregation model

The daily to hourly disaggregation model which was modified and applied in South Africa by Knoesen (2005) is based on the work done by Boughton (2000), which is summarised below. For more details regarding the methodology the reader is referred to Knoesen (2005).

The model comprises four main parts:
a) The distribution of the fraction of the daily total, $R$, that occurs in the hour of maximum rainfall. A value of $R = 1.0$ indicates that all of the rainfall on the day fell in a single hour. This is the upper limit of $R$ and is the boundary of non-uniformity. Completely uniform rainfall throughout a day would yield $R = 0.04167$ (i.e. 1/24 of the daily total). This is the lower limit of $R$.
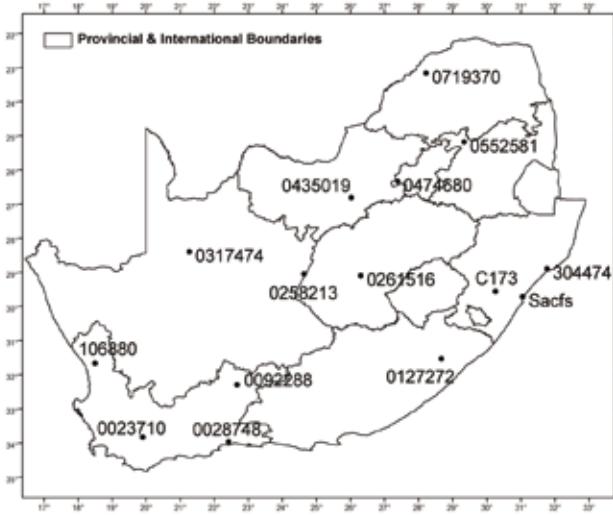
**Figure 1**
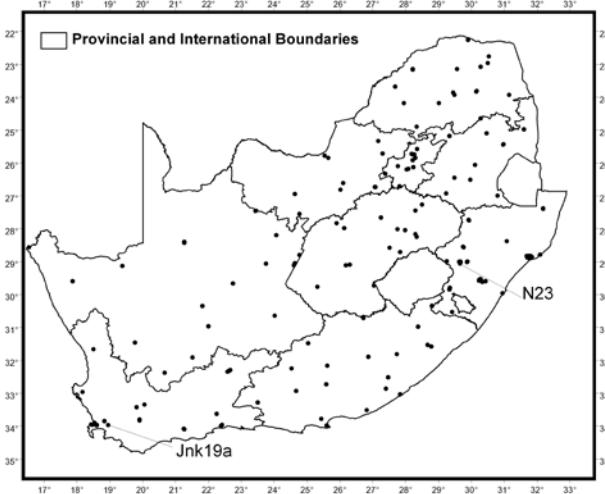*Locations of the 15 'hidden' stations used for model testing*



**Figure 2**
*Locations of stations used for model development*

b) For each value of *R* there is an average set of values for the other 23 hourly fractions of the daily total.

c) Given the 24 fractions from (a) and (b) above, the values are clustered to maintain the observed average highest 2 h, 3 h, 6 h and 12 h fractions of daily rainfall.

d) These clusters are then arranged into random patterns, of which there are 480 possibilities, to reproduce the variations in daily temporal patterns, range from uniform to non-uniform, with the possibility of the hour of maximum rainfall occurring in any hour of the day, while retaining the above-mentioned statistics.

## Distribution of *R*

The primary part of the disaggregation model is the fraction, *R*, of the daily total that occurs in the hour of maximum rainfall. The distribution of *R* has a pattern that is a major characteristic of hourly rainfall at the site. In this study the distribution of *R* for a particular site was created by extracting all the values of *R* at the site for days where the daily rainfall was greater than or equal to 1 mm, for the entire length of record. The computed *R* values were then collated into 20 range bins, as used
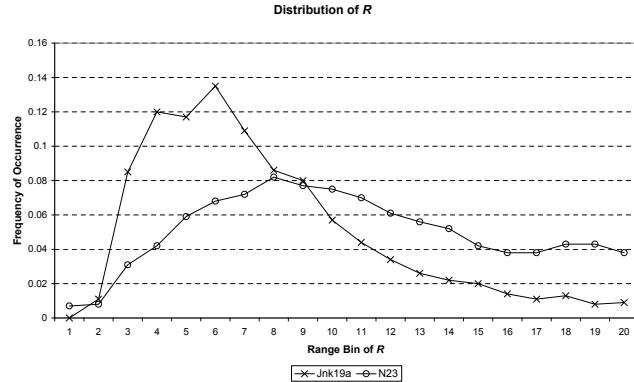


**Figure 3**
*Frequency distributions of R at Stations Jnk19a (in the W. Cape) and N23 (in KZN), using the range bins of R shown in Table 1*

| TABLE 1 | | | | | | | |
|---------|---|---|---|---|---|---|---|
| Range bins used when collating *R* values | | | | | | | |
| No. | Range Bin | No. | Range Bin | No. | Range Bin | No. | Range Bin |
| 1 | 0.0417-0.075 | 6 | 0.275-0.325 | 11 | 0.525-0.575 | 16 | 0.775-0.825 |
| 2 | 0.075-0.125 | 7 | 0.325-0.375 | 12 | 0.575-0.625 | 17 | 0.825-0.875 |
| 3 | 0.125-0.175 | 8 | 0.375-0.425 | 13 | 0.625-0.675 | 18 | 0.875-0.925 |
| 4 | 0.175-0.225 | 9 | 0.425-0.475 | 14 | 0.675-0.725 | 19 | 0.925-0.975 |
| 5 | 0.225-0.275 | 10 | 0.475-0.525 | 15 | 0.725-0.775 | 20 | 0.975-1.000 |

by Boughton (2000) and are shown in Table 1. The distribution of *R* thus shows the proportion of all values of *R* in each of the range bins.

The distributions of *R* for two sites in differing climates are shown in Fig. 3. Jonkershoek (Station Jnk19a), in the Western Cape, is located in a winter-rainfall region whereas Ntabamhlope (Station N23), which lies inland in KwaZulu-Natal (KZN), is located in a summer-rainfall region. The locations of these stations are shown in Fig. 2.

From Fig 3 it is evident that the majority of the days at Jonkershoek fall into Range Bin 6 in Table 1, with a mean value of $R$ ($\bar{R}$) = 0.385, and have small values of *R* indicating that there is a tendency for more uniform rainfall. The distribution of *R* at Ntabamhlope ($\bar{R}$ = 0.537) shows a larger proportion of the days having higher values for *R*. This indicates that at Ntabamhlope larger portions of the daily rain fall in a single hour, which is typical of the convective storms in the summer-rainfall region.

## Calculating the other 23 hourly fractions

If *R* = 1.0, for a given day, then all of the rainfall fell in a single hour, hence the other 23 hourly fractions must be 0. If *R* = 1/24 then each of the other 23 hourly fractions must equal 1/24. If, however, *R* is slightly less than 1.0 it is probable that the rest of the day's rainfall fell in 1 or 2 other hours, resulting in the remaining 21 or 22 hours having zero rainfall. Conversely, if *R* is slightly greater than 1/24 then the other 23 values will be slightly less than, but close to, 1/24. This is important to note as it indicates that the value of *R* has a strong influence in determining the other 23 hourly fractions of rainfall.

In order to determine the other 23 hourly fractions, the 24 hourly fractions for every day on record were ranked in order of magnitude, with *R* being the largest value on each day. This was done for each of the 157 stations. Each of these ranked series, from all 157 sites, was then assigned to one of the 20 range bins, shown in Table 1, according to its value of *R*. Within each range bin the ranked series were then averaged, resulting in 20 aver-
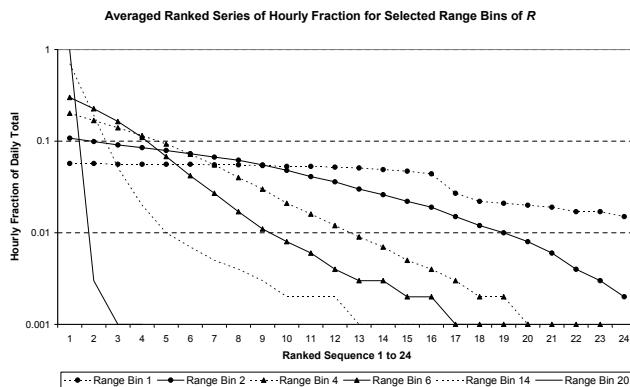
**Figure 4**
*Averaged ranked series of hourly fractions for selected ranges of R*



**Figure 5**
*Frequency distributions of the hour of maximum rainfall at Jonkershoek (in the Cape) and Ntabamhlope (in KZN)*

aged ranked series of hourly fractions, one for each range bin of *R*, examples of which are shown in Fig. 4.

Once all 24 hourly fractions have been determined for each range of *R* they can be used to create daily temporal patterns of rainfall. The following two sections contain a description of how these 24 hourly fractions are arranged to recreate possible realisations of the temporal distribution of daily rainfall.

## Clustering of hourly rainfalls

In order to cluster the 24 hourly fractions, the data from all stations were again processed to calculate the highest 2 h fraction of the daily total, the highest 3 h fraction, the highest 6 h fraction and the highest 12 h fraction. As for the ranked series, all of these fractions were then averaged within the range bin of *R* in which they occurred. This resulted in an average 2 h fraction, 3 h fraction, 6 h fraction and 12 h fraction of the daily total for each of the 20 range bins of *R* (Table 2).

Using the above-mentioned ranked sequences, a computer program was used to check the sum of the first value in the ranked series with each of the other 23 hourly fractions in order to find which of the 23 values gave the best match with the average 2 h fraction for the respective range of *R*. After fixing that value as the value to accompany the first value for the highest 2 h fraction, the program then checks the remaining 22 hourly values to find which value should accompany the 2 h fraction to form the average highest 3 h fraction. The program then searches for the next 3 values to form the average highest 6 h fraction, and then searches for the next 6 values to form the average highest 12 h fraction. Performing this for each range bin of *R* resulted in 20 clustered sequences. The next step was to arrange these clustered sequences into temporal patterns.

## Daily temporal patterns of hourly rainfalls

The hour of day when the highest intensity rainfall occurred was determined for each station. The results show a definitive distribution for the timing of peak rainfall occurrence for a particular location. As shown in Fig. 5 for Station Jnk19a, the hour of maximum rainfall has a somewhat uniform distribution, indicating that the hour of maximum rainfall has a reasonably equal probability of occurring in any hour of a particular day. Station N23, however, has a sinusoidal-like distribution with the majority of days having the peak rainfall occurring during the late afternoon and evening.

In application, a random number is used to select the hour of maximum rainfall from the distribution of the hour of maximum rain for the site of interest. This differs from the work done by
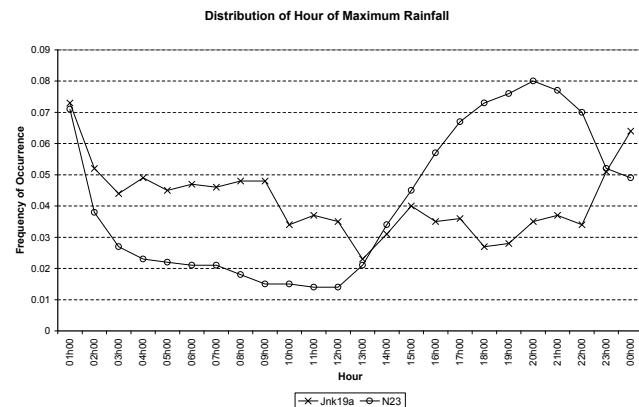
Boughton (2000), as in that study no distinct distribution was found for the time of maximum rainfall in Australia and hence the hour of maximum rainfall was selected at random.

Using the clustered sequences established above and assigning the numerals '1' for the highest fraction, '2' for the fraction that accompanies '1' to form the 2 h fraction, '3' for the fraction that accompanies '1'and '2' to form the 3 h fraction, etc., and then accounting for all permutations when the hour of maximum rainfall can occur, 24 arrangements of the clustered sequences can be created, as shown in Table 3.

The combination of these 24 arrangements with the distributions from the 20 possible range bins of *R* results in 480 different temporal patterns, as opposed to one averaged distribution. These range from uniform to non-uniform with the possibility of the hour of maximum rainfall occurring in any hour of the day. Figure 6 contains a sample of the different temporal distributions that the model produces.

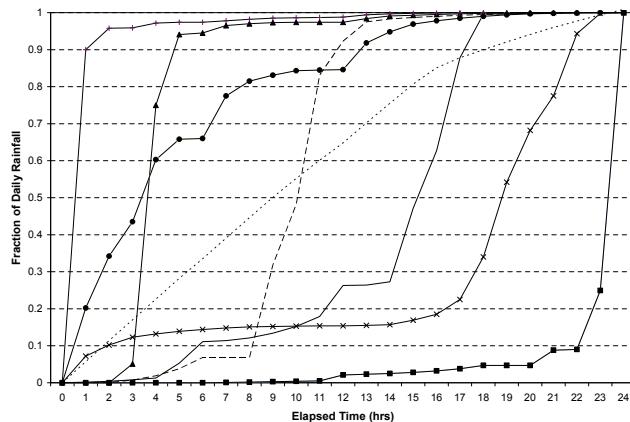| | **1 h (R)** | **2 h** | **3 h** | **6 h** | **12 h** |
|---|---|---|---|---|---|
| **TABLE 2** Average highest 1 h, 2 h, 3 h, 6 h and 12 h fraction for each range bin of *R* | | | | | |
| Range Bin 1 | 0.057 | 0.114 | 0.170 | 0.335 | 0.648 |
| Range Bin 2 | 0.108 | 0.195 | 0.271 | 0.466 | 0.726 |
| Range Bin 3 | 0.154 | 0.266 | 0.357 | 0.564 | 0.783 |
| Range Bin 4 | 0.202 | 0.338 | 0.442 | 0.660 | 0.846 |
| Range Bin 5 | 0.251 | 0.408 | 0.520 | 0.733 | 0.885 |
| Range Bin 6 | 0.300 | 0.475 | 0.592 | 0.789 | 0.909 |
| Range Bin 7 | 0.350 | 0.539 | 0.657 | 0.826 | 0.923 |
| Range Bin 8 | 0.400 | 0.600 | 0.711 | 0.852 | 0.933 |
| Range Bin 9 | 0.450 | 0.662 | 0.761 | 0.877 | 0.944 |
| Range Bin 10 | 0.500 | 0.727 | 0.807 | 0.900 | 0.955 |
| Range Bin 11 | 0.549 | 0.768 | 0.837 | 0.913 | 0.960 |
| Range Bin 12 | 0.599 | 0.797 | 0.855 | 0.922 | 0.965 |
| Range Bin 13 | 0.649 | 0.827 | 0.877 | 0.934 | 0.969 |
| Range Bin 14 | 0.699 | 0.852 | 0.894 | 0.941 | 0.974 |
| Range Bin 15 | 0.749 | 0.877 | 0.910 | 0.950 | 0.977 |
| Range Bin 16 | 0.800 | 0.900 | 0.926 | 0.958 | 0.981 |
| Range Bin 17 | 0.850 | 0.924 | 0.942 | 0.966 | 0.985 |
| Range Bin 18 | 0.901 | 0.947 | 0.958 | 0.975 | 0.989 |
| Range Bin 19 | 0.950 | 0.971 | 0.977 | 0.987 | 0.995 |
| Range Bin 20 | 0.993 | 0.996 | 0.997 | 0.998 | 1.000 |

**Figure 6**
*Samples of the different temporal distributions generated
by the disaggregation model*

## Regionalisation of the methodology

In order to apply the methodology at sites where no short-duration data are available, it is necessary to regionalise the methodology. As shown in Fig. 3, $\bar{R}$ for a particular site had a strong influence on the distribution of $R$ for that site. It was thus decided that the methodology would be regionalised according to the $\bar{R}$ values from the 157 stations used in this study.

It was found that $\bar{R}$ for each of the 157 stations used in this study fell between 0.385 at Jonkershoek (Station Jnk19a) and 0.639 at Pilanesberg (Station 0548290). Collating these 157 values and using the same range bins of $R$ used by Boughton (2000), shown in Table 1, it was found that all but 12 stations had values of $\bar{R}$ that fell within Range Bins 9 to 12. Four stations had $\bar{R}$ val-

ues just below Range Bin 9, but were included with the stations in Range Bin 9 owing to there being too few stations to create an average distribution for Range Bin 8. Before including these stations in Range Bin 9, the distributions of $R$ for these stations were compared to the average distribution of all those stations that had an $\bar{R}$ that fell in Range Bin 9. Similar trends were observed thus justifying their inclusion in Range Bin 9. Similarly, the 8 stations that had $\bar{R}$ values slightly larger than those in Range Bin 12 were compared with the average distribution of $R$ for all the stations with an $\bar{R}$ in Range Bin 12. Owing to the similarity noted it was decided that these stations could be included with the stations in Range Bin 12 rather than creating another average distribution based on significantly fewer stations. Therefore, the ranges used for collating the $\bar{R}$ values in South Africa needed to be changed accordingly and are shown in Table 4.

Using the range bins for $\bar{R}$ listed in Table 4, the 157 stations were categorised according to their respective $\bar{R}$ values and four average distributions of $R$ for South Africa were determined, as shown in Fig. 7. Similarly, 4 average distributions for the time of the hour of maximum rainfall were calculated and are shown in Fig. 8.

In order to establish which distribution of $R$ to use for a site of interest anywhere within South Africa, it was necessary to develop a regionalised map of $\bar{R}$. Using an inverse distance weighting, nearest neighbour approach, with the number of neighbours set to 10, $\bar{R}$ values from the 157 used stations were interpolated onto a 1' x 1' grid. The resulting spatial distribution is displayed in Fig. 9, and shows that the smallest $\bar{R}$ values, 0.375-0.475, occur in the south western part of South Africa as well as on the east coast, while the highest values occur in the northern and north eastern parts of South Africa.

In application, the range bin in which $\bar{R}$ for the site of interest needs to be established in order to select the appropriate region-

| Sample | Sequence | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **TABLE 3** 24 Samples of temporal arrangements of the hourly rainfalls (after Boughton, 2000) | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 2 | 2 | 1 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 3 | 3 | 2 | 1 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 4 | 6 | 5 | 4 | 1 | 2 | 3 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 5 | 6 | 5 | 4 | 2 | 1 | 3 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 6 | 6 | 5 | 4 | 3 | 2 | 1 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 7 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 8 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 9 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 10 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 11 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 12 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 13 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 14 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 15 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 16 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 | 19 | 20 | 21 | 22 | 23 | 24 |
| 17 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 | 19 | 20 | 21 | 22 | 23 | 24 |
| 18 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 19 | 20 | 21 | 22 | 23 | 24 |
| 19 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 | 22 | 23 | 24 |
| 20 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 | 22 | 23 | 24 |
| 21 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 22 | 23 | 24 |
| 22 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 1 | 2 | 3 |
| 23 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 2 | 1 | 3 |
| 24 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

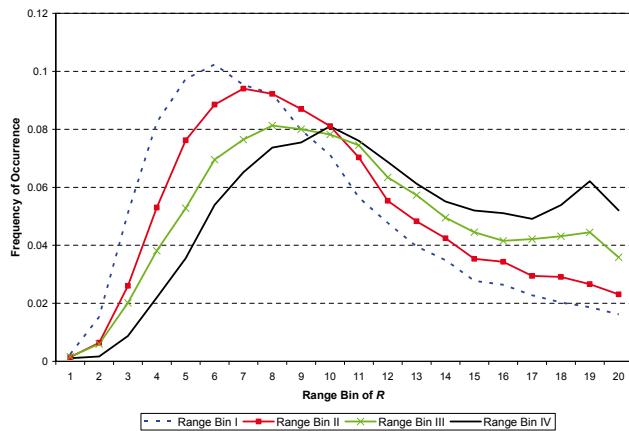| TABLE 4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Range bins used when collating $\bar{R}$ values | | | | | | | |
| No. | Range Bin | No. | Range Bin | No. | Range Bin | No. | Range Bin |
| I | 0.375-0.475 | II | 0.475-0.525 | III | 0.525-0.575 | IV | 0.575-0.675 |



**Figure 7**
*Regionalised distributions of R*



**Figure 8**
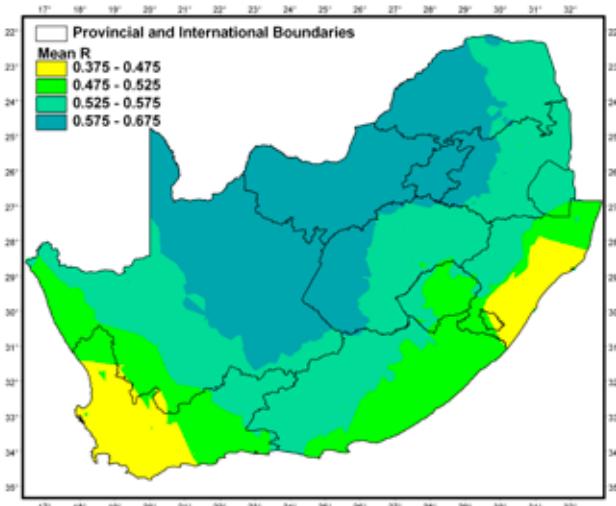*Regionalised distributions of the hour of maximum rainfall*



**Figure 9**
*Regionalised map of mean R ($\bar{R}$)*

alised distribution of *R* and regionalised distribution for when the hour of maximum rainfall occurs. This is accomplished through the use of the regionalised map of $\bar{R}$ (Fig. 9). Once this has been achieved the model utilises the respective distributions

in order to select from the 480 different temporal patterns. Multiplying the measured daily total by the hourly fractions of the selected temporal distribution yields the hourly values.

## Model testing and results

In order to assess the simulated performance of the regionalised disaggregation model, an approach similar to that used by Smithers and Schulze (2000a) was employed. Moments and other event characteristics computed from the disaggregated rainfall series were compared to the equivalent values computed from the observed data. Similarly, design rainfall depths computed from the disaggregated rainfall series were compared to the equivalent values computed from the observed data.

For each of the 15 test stations, the observed hourly data are aggregated to give 24 h values. The disaggregation methodology is then applied to these data in order to attempt to simulate the hourly data for the respective sites. The performance of the model is assessed using two measures. Firstly, moments and statistics of the disaggregated series, e.g. mean, standard deviation, lag auto-correlations, dry probability, skewness, inter-event duration, event duration and number of events, are compared to the corresponding characteristics computed from the observed data. The second measure of model performance is aimed at extreme values, where design rainfalls computed from the disaggregated series are compared to the design rainfalls computed from the historical data.

## Moments and statistics

The two random processes that occur within the disaggregation model, viz., the selection of the value of *R* and the timing of the hour of maximum rainfall, introduce stochastic variability. At each of the selected test stations the stochastic variability was simulated by generating one hundred disaggregated series. A frequency analysis was performed on the 100 sets of disaggregated values for each statistic and duration. High-Low bar graphs depicting the observed moments and the 25th and 75th non-exceedance percentiles of the 100 sets of disaggregated values are used to depict the performance of the model graphically.

In order to compare the performance of the regionalised version of the model, which uses regionalised distributions of *R*, to the 'at-site' version of the model, which uses distributions of *R* computed from short-duration data at the point of interest, the mean absolute relative error (*MARE*), as calculated in Eq. (1), was computed. The number of aggregation levels ($N_L$) in Eq. (1) was set to 11 and the durations used were 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24 h. Basing the calculation of *MARE* on similar calculations by Smithers and Schulze (2000a), all moments and statistics are equally weighted. The results in Fig. 10 indicate that the performance of the regionalised model was similar to the 'at-site' model at all 15 test locations.

$$MARE = \frac{100}{N_s \times N_M \times N_L} \times \sum_{k=1}^{N_L} \sum_{j=1}^{N_S} \sum_{i=1}^{N_M} \left( \frac{\left| S_{(i,j,k)} - O_{(i,j,k)} \right|}{O_{(i,j,k)}} \right) \quad (1)$$
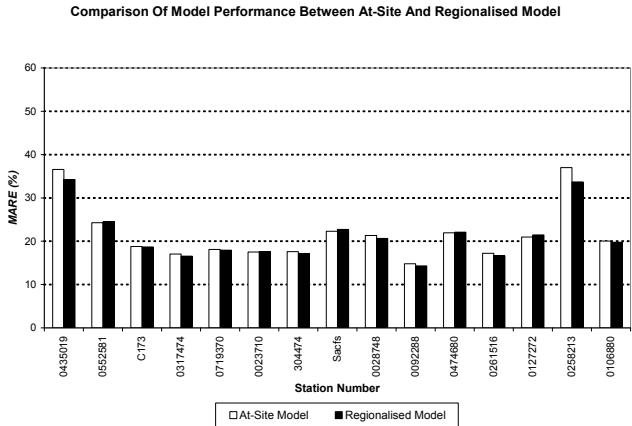
**Figure 10**
*Comparison of model performance between at-site and regionalised model*

**Figure 11**
*Moments and statistics used to quantify model performance*

where:

| | | |
|---|---|---|
| $MARE$ | = | mean absolute relative rainfall error for all durations (%) |
| $S_{(i,j,k)}$ | = | mean $j$-th statistic for aggregation level $k$ computed from the 100 disaggregated rainfall series for month $i$ |
| $O_{(i,j,k)}$ | = | $j$-th statistic computed from observed data for aggregation level $k$ for month $i$ |
| $N_M$ | = | number of months of the year available for statistical analysis |
| $N_L$ | = | number of aggregation levels used (=11, for 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24 h durations) |
| $N_S$ | = | number of statistics and event characteristics calculated (=8, for mean, standard deviation, lag-1 autocorrelation, dry probability, duration of wet periods, duration of dry periods, number of wet periods and skewness) |

Assessing the performance of the regionalised model according to the MARE calculated in Eq (1), the lowest *MARE* value, indicating the best performance, was obtained at Station 0092288 (Beaufort West), while the highest MARE value, indicating the worst performance, was obtained at Station 0435019 (Ottosdal). In order to identify the distinguishing characteristics between the best and worst performing simulations, according to the abovementioned *MARE*, the mean absolute relative error for each statistic (*MARE_STATS*) was computed as shown in Eq. (2).

$$MARE\_STATS_{(k)} = \frac{100}{N_M \times N_L} \times \sum_{i=1}^{N_L}\sum_{j=1}^{N_M} \left( \frac{\left| S_{(i,j,k)} - O_{(i,j,k)} \right|}{O_{(i,j,k)}} \right) \quad (2)$$

where:

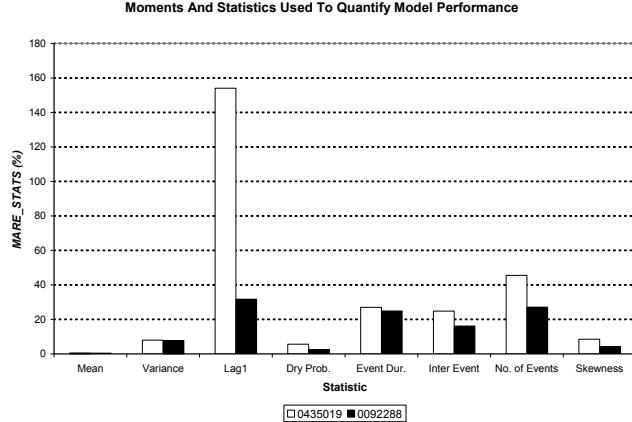| | | |
|---|---|---|
| $MARE\_STATS_{(k)}$ | = | mean absolute relative error for $k$-th statistic (%) (= mean, standard deviation, lag-1 autocorrelation, dry probability, duration of wet periods, duration of dry periods, number of wet periods and skewness) |
| $S_{(i,j,k)}$ | = | mean $k$-th statistic for aggregation level $i$ computed from the 100 disaggregated rainfall series for month $j$ |
| $O_{(i,j,k)}$ | = | $k$-th statistic computed from observed data for aggregation level $i$ for month $j$ |
| $N_M$ | = | number of months of the year available for statistical analysis |

| | | |
|---|---|---|
| $N_L$ | = | number of aggregation levels used (=11, for 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24-hour durations) |

It can be seen in Fig. 11 that the disaggregation model performs similarly well in simulating the scaling and distribution characteristics of the rainfall at both Stations 0435019 and 0092288, such as the mean, standard deviation and skewness. It was expected that the mean rainfall for all levels of aggregation should be simulated extremely well owing to the method of disaggregation. The distinguishing factor between the best and worst simulations is the lag autocorrelation, which is shown by Knoesen (2005) to be related to the quality of the data used in the development of the distributions for the respective sites. Furthermore, the averaging of the 24 hourly values in the patterns shown in Fig. 4 results in the overestimation of wet hours and underestimation of dry hours. It is postulated that this is the primary cause of the poor model performance for event characteristics and statistics associated with the phasing of the rainfall, such as event duration, inter-event duration and number of events. This is a weakness in the current version of the disaggregation model and it is suggested that additional research needs to be undertaken in this regard.

## Extreme rainfall events

Similar to the procedures used by Smithers and Schulze (2000a), design rainfall depths were calculated using the General Extreme Value (GEV) distribution fitted to the Annual Maximum Series (AMS) by L-moments, for the observed data and for each of the 100 disaggregated series generated from the disaggregation model. Design values for the 2, 5, 10, 20, 50, and 100-year return periods were computed for durations of 1, 2, 3, 4, 6, 8, 10, 12, 16, 20 and 24 h.

In order to compare the performance of the regionalised version of the model to the 'at-site' version of the model, with respect to the estimation of design rainfalls, the mean absolute relative error (*MARE_GEV*), was computed as shown in Eq. (3). From Fig. 12 it is evident that the performance of the regionalised version of the model is similar to the 'at-site' version of the model at all 15 test locations. Examples of model performance, with respect to design rainfall estimation, are shown in Fig. 13, which depicts the worst (Station 0028748) and best (Station 0474680) simulations. For each duration and return period, a frequency analysis was performed on the 100 values computed
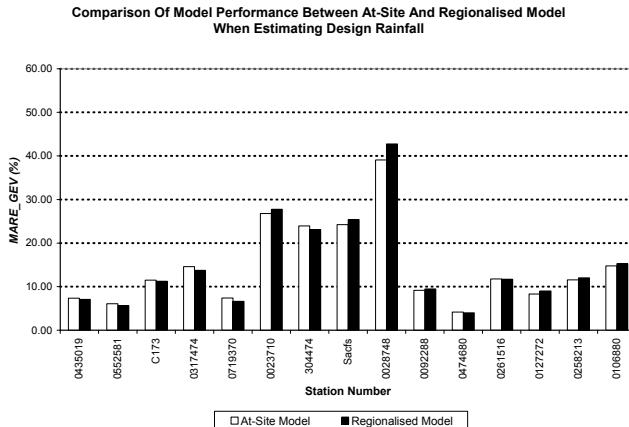
Figure 12
*Comparison of model performance between at-site and regional-ised model when estimating design rainfall*



from the disaggregated rainfall series generated by the disaggregation model. High-low bar graphs depicting the observed design rainfall computed from the observed data and the 25th and 75th non-exceedance percentiles of the design rainfall computed from the 100 disaggregated rainfall series were used to evaluate the performance of the model.

$$MARE\_GEV = \frac{100}{N_L \times N_{RP}} \times \sum_{i=1}^{N_L} \sum_{j=1}^{N_{RP}} \left( \frac{\left|S_{(i,j)} - O_{(i,j)}\right|}{O_{(i,j)}} \right) \quad (3)$$

where:

$MARE\_GEV$ = mean absolute relative error of design rainfall of all durations (%)

$S_{(i,j)}$ = mean *j*-th return period, *i*-th hour design rainfall computed from the 100 disaggregated rainfall series

$O_{(i,j)}$ = *j*-th return period, *i*-th hour design rainfall computed from observed data

$N_L$ = number of aggregation levels (= 11)

$N_{RP}$ = number of return periods (= 6 for 2, 5, 10, 20, 50, and 100-year return periods)

The poor performance observed when estimating design rainfall at Station 0028748 appears to be related to the distribution *R*, i.e. the station that displayed the best results, i.e. the lowest *MARE_GEV*, has an $\bar{R}$ value between 0.575 and 0.675, whereas the station with the worst results, Station 0028748, has an $\bar{R}$ value between 0.475 and 0.525. After analysing all the test stations it was found that the stations with the highest $\bar{R}$ values gave the best results. This is because on those days when smaller rainfall events (± 1 mm) occurred it is likely that the all the day's rainfall fell within a few hours, thus unduly influencing the distribution *R*. Although this will influence the results at all the stations used, it appears that the error is exacerbated for those stations with lower $\bar{R}$ values. It is postulated that the use of different distributions of *R*, to represent rainfalls of differing magnitudes, will improve the performance of the rainfall disaggregation model, particularly in the estimation of design rainfall.

## Discussion and conclusions

The rainfall disaggregation model developed by Boughton (2000), and modified by Knoesen (2005), has been regionalised for application in South Africa. Two measures were employed in order to quantify the performance of the disaggregation model.
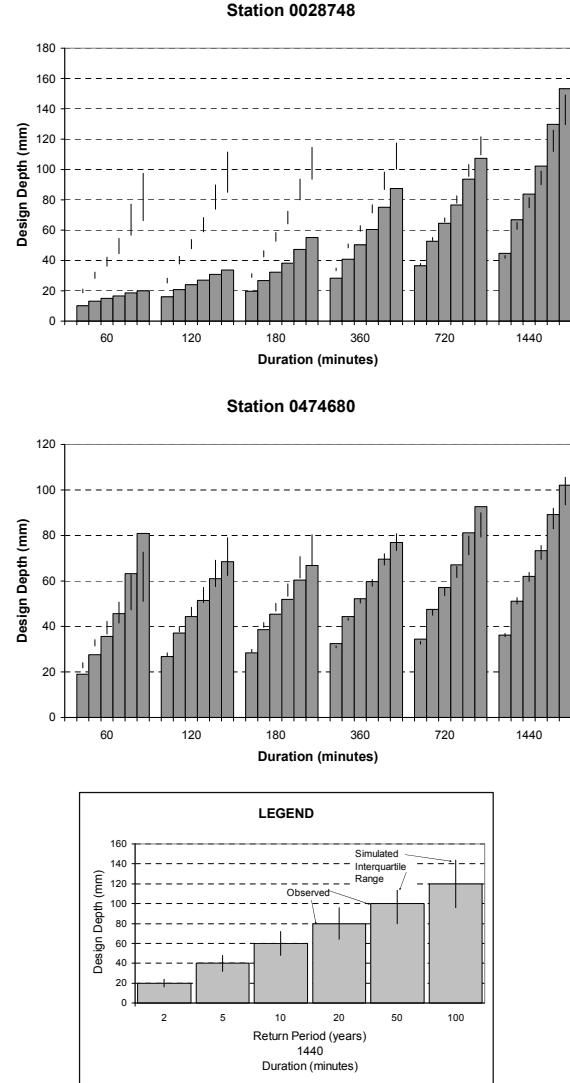


Figure 13
*Design rainfall estimated using disaggregated data for Stations 0028748, at George, and 0474680, at Carletonville*

Firstly, moments and other event characteristics were computed from the disaggregated data and compared to the equivalent values computed from the observed data. Secondly, design rainfall depths were computed from the disaggregated data and compared to the equivalent values computed from the observed data.

The results obtained indicate that both the at-site and regionalised application of the model are able to produce synthetic hourly rainfall data which resemble the general distribution of the observed hourly data for a particular site. However, the model is less capable of simulating some of the statistics and event characteristics associated with the phasing properties of the rainfall. Furthermore, owing to the structure of the model, the results indicate that the model is less capable of simulating design rainfalls at locations with lower $\bar{R}$ values, as well as for selected return periods. It is therefore recommended that additional research be undertaken regarding the sequencing of the disaggregated hourly rainfalls.

Comparing the *MARE* values from the disaggregation model when 'at-site' information is available to those yielded when regionalised input is used, it can be seen that the results at each of

the 15 tested locations are very similar. This indicates that, with the exception of event characteristics related to the sequencing of the hourly rainfalls, the disaggregation model applied using regionalised information, is able to produce short-duration rainfall with similar characteristics to the actual rainfall observed at the location. This is a positive result as it implies that the model can be used to disaggregate daily rainfall reasonably at locations in South Africa where there are no observed short-duration data but daily rainfall data are available. Furthermore, the disaggregation model could be linked with a daily rainfall generator. This would facilitate the generation of long sequences of hourly data for any location in South Africa which could be used for modelling of water resources and design flood estimation.

## References

BOUGHTON W (2000) A Model for Disaggregating Daily to Hourly Rainfalls for Design Flood Estimation. Cooperative Research Centre for Catchment Hydrology. Report No 00/15. Monash University, Clayton, Victoria, Australia. 36 pp.

GYASI-AGYEI Y (1999) Identification of regional parameters of a stochastic model for rainfall disaggregation. *J. Hydrol.* **223** 148-163.

KNOESEN DM (2005) The Development and Assessment of Techniques for Daily Rainfall Disaggregation in South Africa. Unpublished M.Sc. Thesis, University of KwaZulu-Natal, Pietermaritzburg, RSA. 88 pp.

KOUTSOYIANNIS D and ONOF C (2001) Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *J. Hydrol.* **246** 109-122.

MIKKELSEN PS, MADSEN H, ARNBJERG NIELSEN K, JORGENSEN HK, ROSBJERG D and HARREMOES P (1998) A rationale for using local and regional point rainfall data for design and analysis of urban storm drainage systems. *Water Sci. Technol.* **37** (11) 7-14.

SMITHERS JC and SCHULZE RE (2000a) Development and Evaluation of Techniques for Estimating Short Duration Design Rainfall in South Africa. WRC Report No. 681/1/00. Water Research Commission, Pretoria, RSA. 356 pp.

SMITHERS JC and SCHULZE RE (2000b) Long Duration Design Rainfall Estimates for South Africa. WRC Report No. 811/1/00. Water Research Commission, Pretoria, RSA. 69 pp.