Continuous daily hydrograph simulation using duration curves of a precipitation index

V. Yu. Smakhtin[1]* and B. Masse[2]

[1]*Institute for Water Research, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*
[2]*Ecole National Supérieure Agronomique de Rennes, F-35042 Rennes Cedex, France*

Abstract:

The paper describes a parsimonious approach for generating continuous daily stream-flow time-series from observed daily rainfall data in a catchment. The key characteristic in the method is a duration curve. It is used to convert the daily rainfall information from source rain gauges into a continuous daily hydrograph at the destination river site. For each source rain gauge a time-series of rainfall related 'current precipitation index' is generated and its duration curve is established. The current precipitation index reflects the current catchment wetness and is defined as a continuous function of precipitation, which accumulates on rainy days and exponentially decays during the periods of no rainfall. The process of rainfall-to-runoff conversion is based on the assumption that daily current precipitation index values at rainfall site(s) in a catchment and the destination site's daily flows correspond to similar probabilities on their respective duration curves. The method is tested in several small catchments in South Africa. The method is designed primarily for application at ungauged sites in data-poor regions where the use of more complex and information consuming techniques of data generation may not be justified. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS   daily stream-flow time-series; flow duration curve; spatial interpolation; observed records; precipitation index; ungauged sites

INTRODUCTION

Continuous daily stream-flow time-series are required for a variety of hydrological analyses and engineering applications, including water resource allocation and water quality problems, utilization of water resources in small catchments and environmental flow management. For gauged streams, use is often made of observed historical flow records, but even these readily available records normally require certain augmentation (patching, extension) and/or correction for non-stationary anthropogenic effects. For ungauged sites, or sites with short observations, the representative flow time-series may be generated by means of detailed deterministic rainfall–runoff models that already have a long history and a 'successful carrier' in hydrological science. Flow time-series represent only part of the output that deterministic models normally provide. At the same time, the flow time-series alone may satisfy the hydrological data requirements of many projects. Also, the application of complex and information consuming methods is not always appropriate in data-poor regions, where the use of more pragmatic techniques of data generation may be most justified and equally successful.

The most straightforward example of such a technique is a weighting of observed daily stream flows at a gauged site (or sites) by the ratio of the catchment areas of the site of interest and the gauged site(s). The other possible alternative is to generate monthly flow time-series and disaggregate it into daily flows.

---

* Correspondence to: V. Smakhtin, CSIR, Environmentek, PO Box 395, Pretoria 0001, South Africa. Email: vsmakhtin@csir.co.za

Monthly flow time-series may be obtained from water balance calculations or monthly models (Lall and Olds, 1987; Midgley *et al.*, 1994; Vandewiele and Atlabachew, 1995; Hughes, 1997; Mohseni and Stefan, 1998), which are used in many countries and, compared with daily models, represent a less resource-intensive exercise. The problems of statistical downscaling of monthly time step flow time-series into daily have been addressed recently by several researchers (e.g. Middelkoop, 1998; Torfs, 1998). Daily stream flows also may be generated by stochastic models (e.g. Kelman, 1980). The disadvantage of existing disaggregation procedures and stochastic simulation methods is that they produce the time-series with similar statistical properties, but not necessarily sequentially similar to the one that may be expected at a site of interest. This may become a limiting factor in river ecology studies, for example, where the actual sequence of daily flows is of paramount importance.

A non-linear spatial interpolation approach suggested by Hughes and Smakhtin (1996) overcomes the latter problem. The method is based solely on observed stream flow records. Its key characteristic is a flow duration curve (FDC), which gives a summary of flow variability at a site and is interpreted as a relationship between any discharge value and the percentage of time that this discharge is equalled or exceeded. The method was developed originally only for patching or extension of observed flow data. Smakhtin *et al.* (1997) and Smakhtin (1999), however, illustrated how a FDC may be established at the ungauged sites and translated into a complete flow time-series. This essentially extended the area of application of this technique and made it applicable in a variety of short-term water resources and river ecology projects that rely on time-series data. However, the success in application of this pragmatic method to any site along a river remains dependent on the availability of adjacent observed stream flow data, which frequently is limited. To increase the flexibility of the method, use should be made of more readily available daily rainfall records.

This paper investigates the possible extension of the spatial interpolation method for daily stream-flow time-series generation that uses both observed flow and rainfall records. The original structure of the spatial interpolation algorithm is reviewed briefly and the modifications required to incorporate rainfall records are described. These modifications are designed to make use of duration curves of a rainfall-related index, which reflects the daily fluctuations of the catchment 'wetness' and is similar in concept to the well-known index of antecedent precipitation. The method is intended mostly for application at ungauged sites and its performance is illustrated for several South African catchments under conditions of limited or no calibration.

SPATIAL INTERPOLATION OF OBSERVED STREAM-FLOW RECORDS

The basic assumption of the non-linear spatial interpolation algorithm in its original form (Hughes and Smakhtin, 1996) is that flows occurring simultaneously at sites in reasonably close proximity to each other correspond to similar percentage points (probabilities) on their respective flow duration curves. The approach includes the following steps.

1. *Source-site selection for data transfer*. 'Source' sites are the sites with gauged flow records from where the data will be transferred to the 'destination' site of interest. Up to five candidate source flow gauging stations are identified in the vicinity of the destination site and the weights are assigned to each of the source sites. The weights are based on the degree of similarity between source and destination site flow regimes. Suitable source sites may be selected and weights may be quantified on the basis of spatial correlation analysis. However, in practice, especially in data-poor regions, such selection and quantification is frequently either limited or obvious (Hughes and Smakhtin, 1996) and the nearest gauges on the same river, its tributaries or adjacent streams will be used as source sites.

2. *Generation of tables of discharge values*. Tables of discharge values are generated for each site (source and destination) and each month of the year for fixed percentage points on the corresponding FDC. For example, if two source sites are selected, the total number of sites involved in simulation is three and the total number of required discharge tables is 36. The algorithm currently uses 17 fixed percentage points from 0·01 to 99·99% and, consequently, each discharge table is made up of 17 flow values.

3. *Data transfer from individual source gauges*. This is the major computational step during which the percentage point of each day's flow at each source site is identified and the flow value for the equivalent percentage point from the destination site's FDC is read off. The discharge tables are used to 'locate' the flows on corresponding curves and log-interpolation is used between fixed percentage points. The procedure is repeated for each source site.

4. *Final estimate of a destination daily discharge*. Weighted averaging of all flow estimates for the destination site (obtained using individual source flows) is performed. If the flow for any source site on that day is missing, it is ignored, and the final estimation is performed with the remaining set of individual flows (the weights are adjusted automatically).

The first two steps are, in fact, preparatory and are performed once only. The last two form the bulk of the computational procedure and are performed for each day during the calculation period. Different gauged records may have different starting and ending dates. The starting date of the calculation period corresponds to the earliest starting date of the selected gauges, and the ending date corresponds to the latest of all ending dates. Hughes and Smakhtin (1996) provide the details of the computational procedure.

If the observed record at the destination site exists, this record (entirely or partially) may be used for the generation of discharge tables. The aim of the interpolation procedure is likely to be the patching and/or extension of the available record at the destination site. Also, in this case, the 'model' described above may be 'calibrated' in a similar way as, for example, rainfall–runoff models are calibrated against observed data. The calibration may be performed by (i) changing the selection of the source sites and their number and/or (ii) changing the weighting factors for each source site selected. Technically the calibration may be performed by the visual analysis of similarities between displayed source and destination site's flow time-series and by calculation of several goodness-of-fit measures.

If the generation of flow time-series is intended at an ungauged destination site, the set of typical 1-day FDCs for each month of the year at this site should be established prior to the simulation of the actual time-series. A FDC for an ungauged site may be calculated through either (i) regionalization of FDCs based on available observed records from several adjacent gauges (Smakhtin *et al.*, 1997) or (ii) conversion of FDCs calculated from monthly data into 1-day FDCs (Smakhtin, 1999). The latter approach is particularly relevant in a South African context where synthetic monthly time-series are available for a large number of small and usually ungauged incremental drainage subdivisions (called quaternary catchments), throughout the entire country (Midgley *et al.*, 1994). Other methods of calculating FDCs at ungauged sites have also been documented (Lane and Lei, 1949; Institute of Hydrology, 1980; Quimpo *et al.*, 1983; Mimikou and Kaemaki, 1985; Fennessey and Vogel, 1990; LeBoutillier and Waylen, 1993). Once a FDC is established at the ungauged destination site of interest, the spatial interpolation method may be applied to generate the actual continuous flow time-series. However, no 'calibration' may be performed in this case because no observed record at the destination site is available.

In general, the spatial interpolation approach is simple to use and through the intensive testing in many catchments has been found to account satisfactorily for non-linearities in the relationship between stream flows at different sites (Hughes and Smakhtin, 1996). Its main limitation, however, is that it is based entirely on observed flow records, and therefore inherits all the problems related to the quality and availability of such records. One of the most important requirements (and limiting factors) for the application of the spatial interpolation approach to either gauged or ungauged destination sites is the existence of at least one nearby source flow gauge from where the data may be transferred (two or more source sites are, however, preferable if the simulated hydrograph is to have a minimum, or no, missing data). As has been illustrated by Smakhtin *et al.* (1997), this may represent a serious problem in data-deficient regions. Even if the required FDCs are established for the destination site, a suitable neighbouring source flow gauge may still not be available. The logical way to address this problem is to investigate the possibilities of incorporating more widely available (and normally longer) rainfall records into the structure of non-linear spatial interpolation method.

INCORPORATION OF RAINFALL RECORDS INTO SPATIAL INTERPOLATION METHOD

Rainfall records are to be used in the algorithm in cases when no source flow records are available. Consequently, both source flow time-series and source FDC should be replaced by corresponding rainfall-related measures. 'Rainfall duration curves' certainly can be constructed similarly to FDCs. Their direct use in the algorithm is problematic, however, owing to the number of zero rain days in any rainfall record. Zero daily rainfall does not imply zero flow. A rainfall duration curve, if constructed on the basis of the entire record and not just rain days, is a very steep one, indicating amongst the others, that it rains only for a very small percentage of the time, on average, during a year (Figure 1). This percentage obviously depends upon climatic conditions and may be relatively large in some humid regions.

In order to make rainfall records 'work' within the framework of the spatial interpolation approach, one should define a continuous function of daily rainfall that would abruptly increase on rainy days and
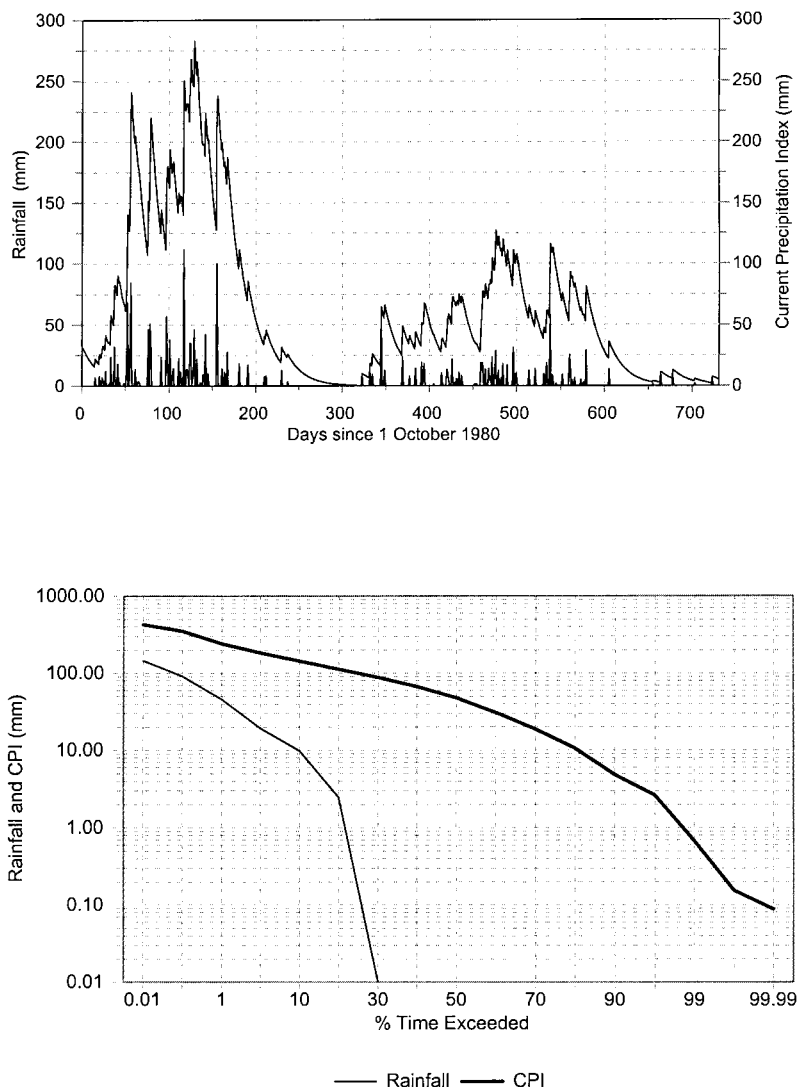

Figure 1. Example time-series of daily rainfall and current precipitation index (CPI) at one rainfall gauge in the Mac-Mac River catchment (top) and typical form of the annual duration curves of daily rainfall and CPI (bottom)

gradually decay during the dry periods, and therefore resemble the general pattern of stream flow variability. The type of function sought is thus similar to that of the antecedent precipitation index (API). The API is an exponential decay function of precipitation that reflects the rate of soil moisture depletion during the period of no rainfall (Linsley *et al.*, 1975). Soil moisture content affects the occurrence and magnitude of storm flow as well as the rate of groundwater recharge. Consequently, the API may be a good indicator of the continuous catchment outflow and, as such, has already been used in some previous types of API hydrological models for continuous hydrograph synthesis (e.g. Sittner *et al.*, 1969). However, the API (or its modifications) is used most frequently in storm runoff prediction (Fedora and Beschta, 1989; Ni-Lar-Win and Vandewiele, 1994).

The original concept of the API needs to be modified slightly to be applicable in the context of the spatial interpolation approach. It should reflect not just the antecedent catchment conditions for any day but also the current daily precipitation input. Such a modification effectively leads to what is referred to here as a current precipitation index (CPI). The CPI for any day is calculated as:

$$\mathrm{CPI}_t = \mathrm{CPI}_{t-1} K + R_t \tag{1}$$

where CPI_t is a current precipitation index (mm) for day t, R_t is the catchment precipitation for day t and K is the daily recession coefficient. On any day with no rain ($R_t = 0$) the CPI is equal to the CPI of the previous day multiplied by K (similarly to API). If it rains on any day, however, the daily rainfall depth is added to the CPI immediately. Consequently, the index in its current form does not represent only the antecedent wetness of the catchment, but also reflects the effects of the current precipitation.

To generate the continuous time-series of daily CPI values, its initial value (CPI_0) and recession coefficient need to be defined. The initial value, however, does not have any major effect on the resultant CPI time-series, because several time-series calculated with different initial values will converge within the first year of simulation. This value generally may be assumed to be equal to the long-term mean daily precipitation and calculated from available daily rainfall records.

The daily recession coefficient K normally varies from 0·85 to 0·98 (Linsley *et al.*, 1975; Fedora and Beschta, 1989) and it therefore fluctuates in a similar range as the baseflow recession constant (e.g. Klaasen and Pilgrim, 1975). It may also be treated as a parameter of the current 'model' and calibrated by comparing observed and simulated hydrographs in cases when observed data exist. On the other hand, K values will be required for ungauged catchments where no calibration is possible. In the current study, K is assumed to be equal to the median daily recession ratio of a stream (REC50). This characteristic represents the rate of baseflow recession and is estimated from the distribution of daily recession ratios (today's flow divided by yesterday's flow) calculated for all recession periods found in a record for those days when discharge is less than long-term mean daily flow. This index is similar to the one described in the FREND (1989) study, which provides the values of REC50 for a number of streams in Europe. Smakhtin and Watkins (1997) list REC50 values for a number of South African streams. The K values may also be estimated by means of regional regression models, where K is dependent on catchment characteristics (topography, geology, etc.).

With an estimated K value, a continuous daily CPI time-series may be generated for any rainfall station in a catchment and, consequently, the required CPI duration curves also may be established. Figure 1 (top) illustrates the time-series of observed daily rainfall and corresponding calculated CPI time-series at an individual rainfall station, and (bottom) presents the duration curves calculated from these time-series.

During a long dry period, the exponentially decaying CPI may reduce to unrealistically low values at which a catchment may be considered to be completely dry. It therefore is possible to replace all such negligible CPI values below a certain low threshold by zeros. In the current study this threshold was arbitrarily assumed to be 0·1 mm and every CPI value below 0·1 mm was replaced by zero in the resultant CPI time-series. This procedure was deemed necessary to create a more realistic CPI time-series in which zero CPI values may correspond to zero river flows (e.g. in all non-perennial rivers).
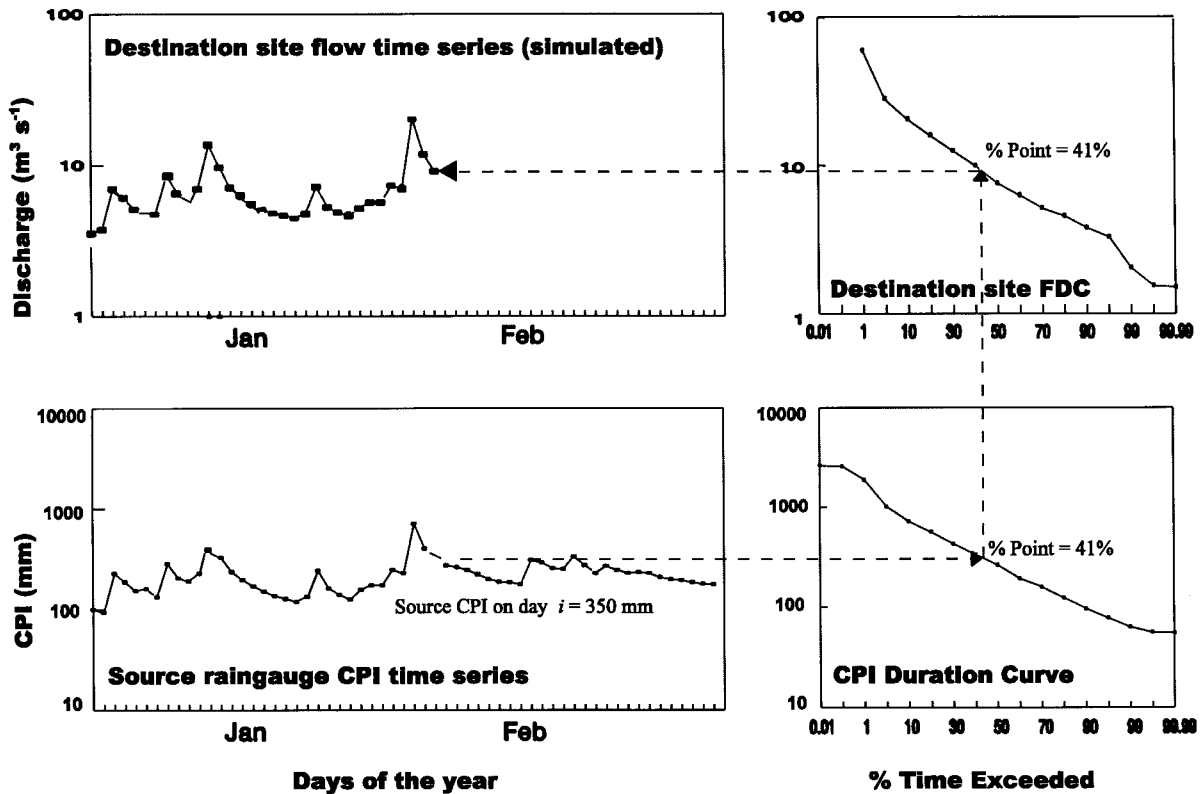
Figure 2. Illustration of daily stream flow generation using the spatial interpolation algorithm with the CPI source time-series

Once the CPI time-series and duration curve are calculated, it may be used in the spatial interpolation algorithm as a substitute for the source flows. The major assumption of the algorithm in this case becomes that both the CPIs occurring at rainfall sites in a reasonably close proximity to the destination site and destination site's flows themselves correspond to similar percentage points on their respective duration curves. The steps of the computational procedure outlined in the previous section do not change and the actual process of daily flow generation is illustrated by Figure 2.

APPLICATION OF THE METHOD TO SOUTH AFRICAN CATCHMENTS

Description of the catchments

The catchments selected have been drawn from different parts of the country and have already been used in some previous hydrological or water resource related studies (e.g. Smakhtin and Watkins, 1997). Two of them (the Mac-Mac and Marite) are located in the headwater areas of the Sabie River, flowing through the Kruger National Park; one, the Sundays, is the tributary of the Thukela River in KwaZulu-Natal Province, one is located in the headwaters of the Mhlatuze River in the coastal area north of Durban and the other one is in the headwater part of the Koonap River catchment located in the semi-arid region of the Eastern Cape Province (Figure 3). The exact geographical location of these catchments is, however, largely irrelevant in the context of the proposed stream-flow generation algorithm. On the other hand, the differences in the type of rainfall and flow regime as well as the amount and quality of available data are of primary importance. It may be expected a priori that lack of observed data would be the limiting factor to properly characterize the spatial variability of rainfall and, consequently, the spatial variability of the catchment wetness.
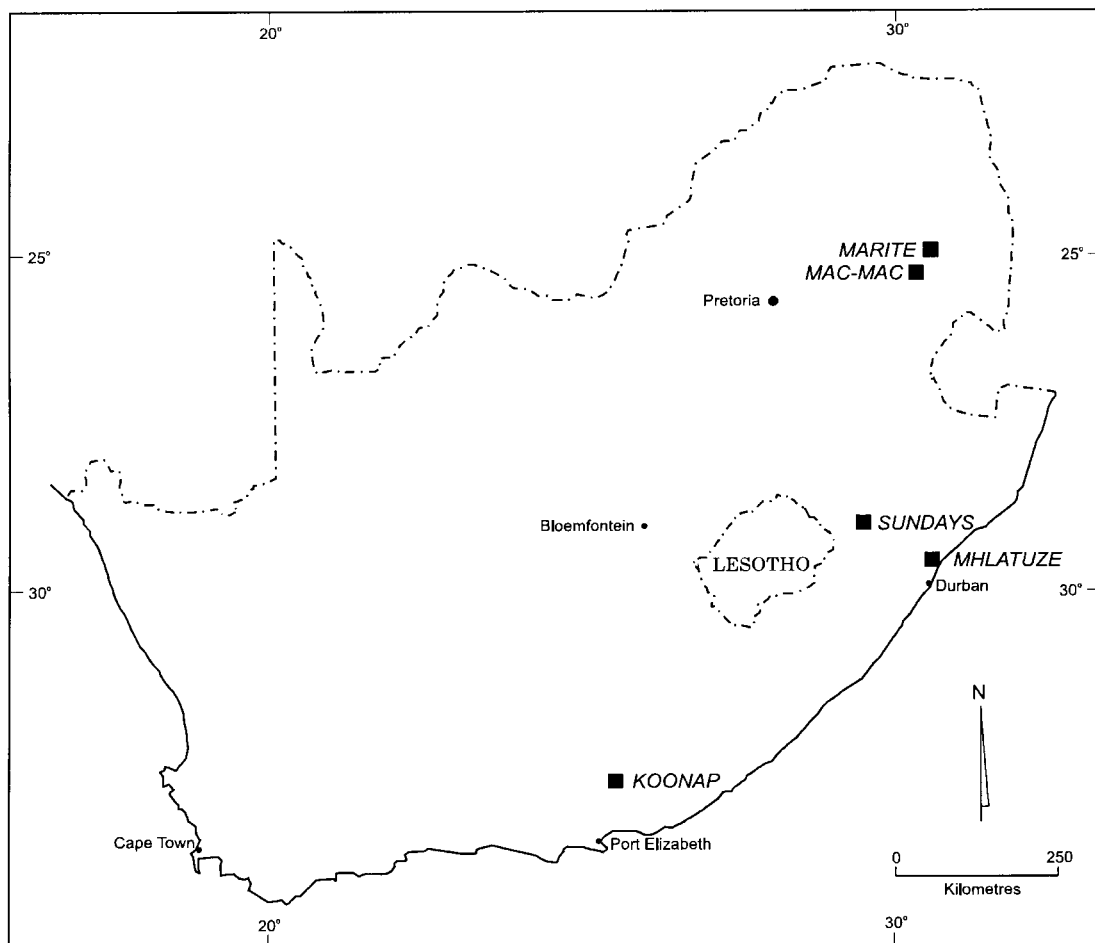
Figure 3. Location of the study catchments

The summary of information available for each catchment selected is given in Table I. All selected catchments are gauged. This has been deemed necessary not for the sake of 'model calibration', but simply to compare the 'model' output with the observed stream flow data. Schematic maps of each catchment with indicated locations of rainfall stations and stream flow gauges are presented in Figure 4. All rainfall and stream flow gauges considered are referred to in the South African national coding system.

Mac-Mac River and Marite River. These catchments are located in a mountainous region and receive relatively high summer rainfall, which is primarily convective in nature. A strong orographic effect results in considerable variation in mean annual precipitation (MAP) across each catchment (1200–1600 mm yr^{-1}). At the same time, for the Mac-Mac catchment there are only two suitable rainfall stations that have consistent and largely overlapping records. Both are located in the high rainfall area, one station being outside the catchment boundary. Suitable rainfall stations in the Marite catchment are located around its centre. The stream flow records for both flow gauges are of reasonable quality, but Marite gauge station has a very low discharge measuring capacity.

Table I. Data availability in the study catchments

Catchment	Flow and rainfall gauge codes	Gauged area, km^2	Record period
Mac-Mac	X3H003	52	1948–1997
	0594595W		1914–1992
	0594444W		1940–1992
Marite	X3H011	212	1978–1997
	0594802W		1950–1997
	0595025W		1972–1997
Sundays	V6H004	658	1954–1997
	0334761W		1932–1963
	0334825W		1914–1992
	0334678W		1913–1967
	0334803W		1949–1997
Mhlatuze	W1H006	1272	1964–1973
	0337143W		1928–1993
	0303127W		1916–1994
	0303667W		1936–1980
	0303711W		1941–1993
Koonap	Q9H016	489	1982–1993
	0100025W		1960–1992
	0100060W		1060–1992
	0100329W		1960–1992

Sundays River. The river originates at an altitude of about 1700 m above sea level. The MAP exceeds 900 mm yr^{-1}. More than 70% of the annual rain falls in the warm period (October to March). Precipitation often occurs in the form of heavy localized thunderstorms. Two of the four suitable rainfall stations have records only up to the mid-1960s and, consequently, the later period is represented only by two stations located in the downstream part of the catchment. The stream flow record is of good quality.

Mhlatuze River. The catchment receives 850 mm yr^{-1} of rain, but no clear seasonal pattern appears to exist either in rainfall or stream flow. Many available rainfall stations have short records and were not considered. Three of the four selected rainfall stations lie around the boundaries of the catchment. The available data at the stream flow gauge extend from 1964 to 1973, but contain long periods of missing data. These records do not appear to be very reliable and are affected by sedimentation problems upstream of the flow gauge.

Koonap River. The catchment drains the mountainous area in the Eastern Cape Province and experiences on average about 700 m yr^{-1} rainfall. It is the 'driest' catchment of all considered at this stage. Three rainfall stations with overlapping records are relatively evenly distributed with respect to the catchment boundary. Only a short flow record of good quality is available.

Discussion of simulation results

Figures 5–9 illustrate the correspondence between daily stream-flow hydrographs observed and generated by the algorithm, as well as between annual FDCs constructed on the basis of observed and generated time-series. Some measures of fit between observed and generated time-series are presented in Table II. They include maximum, minimum and mean flows, standard deviation (SD) and coefficients of determination (R^2) and efficiency (CE).

The results of the 'simulation' are also compared in Table II with those generated by the daily time step semi-distributed rainfall–runoff variable time interval (VTI) model (Hughes and Sami, 1994), which repre-
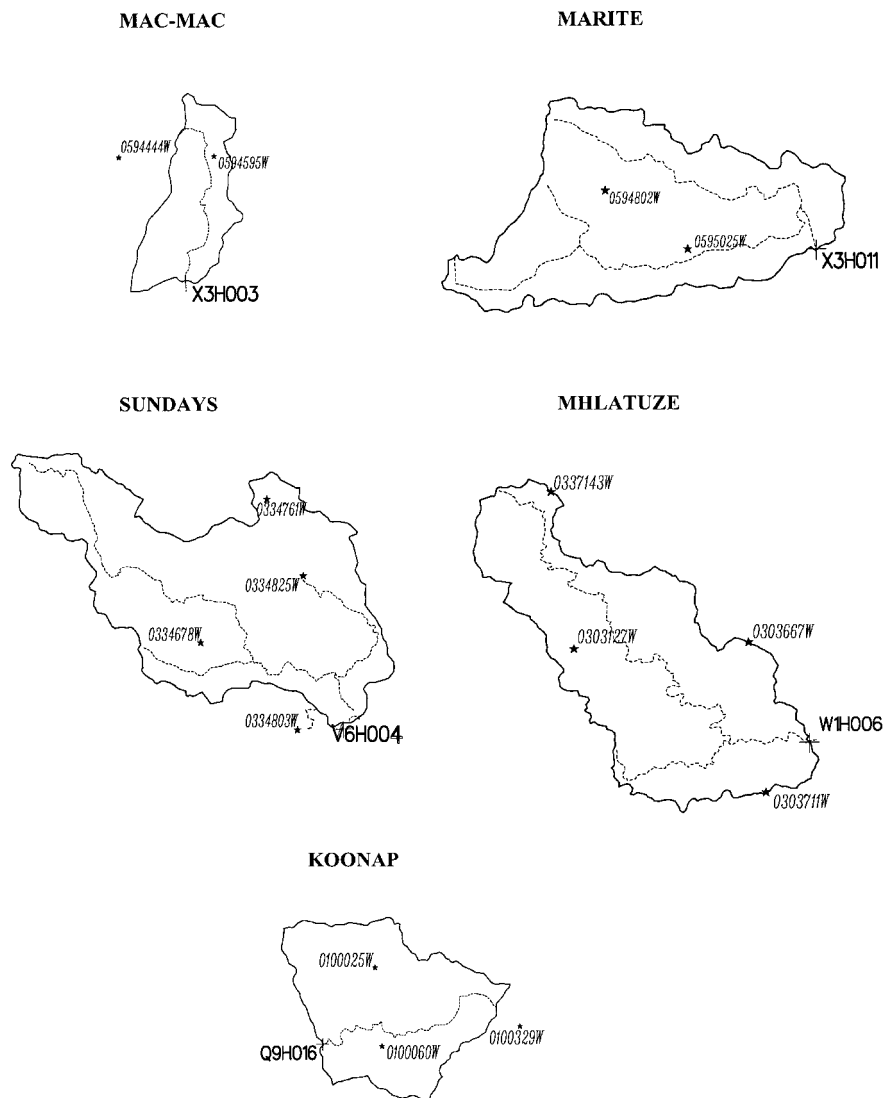
Figure 4. Schematic maps of the study catchments with rainfall stations (stars) and stream flow gauges (pluses) (not to scale)

sents a much more resource-intensive alternative to generating flow time-series. The VTI model incorporates two soil layers and includes the description of all major components of catchment hydrological cycle: interception, evapotranspiration, rainfall intensity controlled runoff, soil moisture redistribution and saturated surface runoff, a variety of surface–subsurface water interaction processes, catchment routing, channel transmission losses and flow routing. A modelled catchment is represented by a set of interlinked homogeneous subareas. The variability of some hydrological processes within each subarea is described by means of probability distribution functions of some model parameters. The average rainfall input for each subarea at each time step is determined by the inverse distance squared interpolation procedure, which uses the information from the nearby rainfall stations, the coordinates of these stations and coordinates of the subarea centre. The detailed description of the model is provided by Hughes and Sami (1994). The fit statistics in Table II are calculated for the periods indicated, for which the VTI model simulations were available.
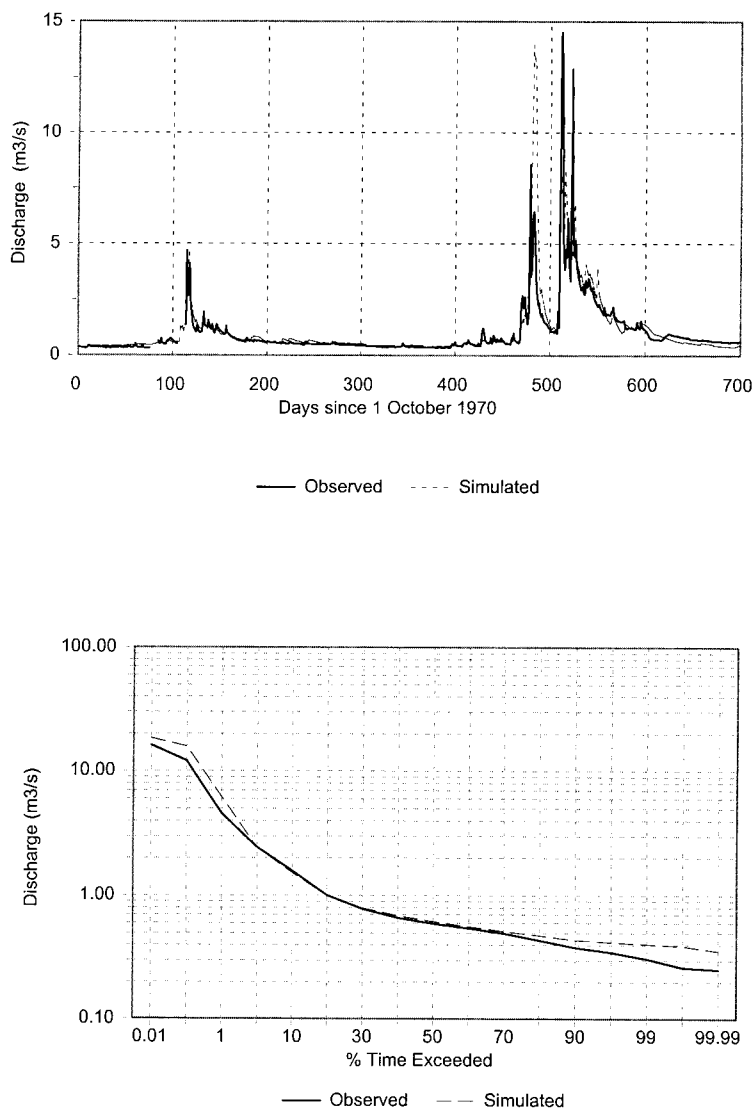
Figure 5. Observed and simulated daily hydrographs (top) and annual 1-day flow duration curves (bottom) for the Mac-Mac River

It should be emphasized that no attempt has been made to 'calibrate' the proposed 'model' because (i) calibration options are very limited (only rainfall stations weights and recession parameter value may be changed) and (ii) the method is intended for use at ungauged sites where no calibration is possible. Consequently, the results illustrate the performance of the approach under very stringent conditions. All rainfall stations in each catchment have been assigned equal weights (the sum of all individual station weights equals 1 in every catchment). The recession K value for each catchment was assumed to be equal to the median recession ratio value (REC50) of the nearest stream flow gauge listed in Smakhtin and Watkins (1997). The assigned K values varied in the range from 0·90 to 0·95.

Taking these stringent conditions into account, the results generally may be viewed as satisfactory and promising. In terms of fit statistics, the algorithm has performed almost as well as the more complex VTI model, although better VTI simulations were achieved for Mac-Mac and Koonap rivers, as outlined in
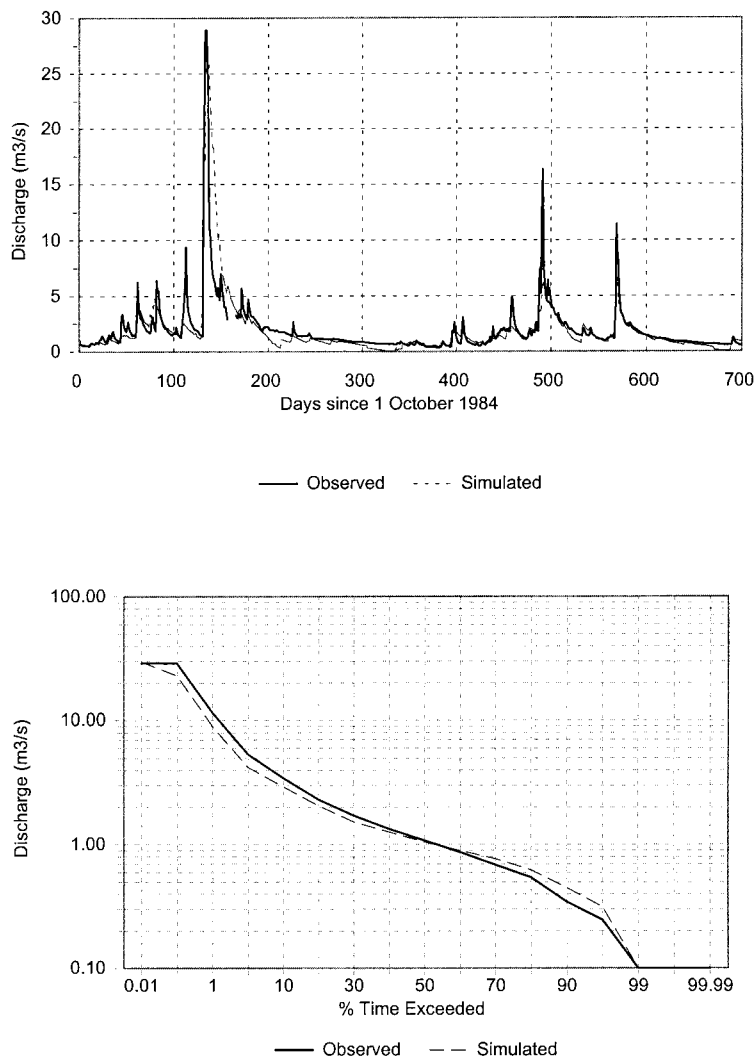
Figure 6. Observed and simulated daily hydrographs (top) and annual 1-day flow duration curves (bottom) for the Marite River

Table II. (In the case of Marite River, the low CE value for the VTI simulations is most probably attributed to high simulated peaks, which in some years exceed the low measuring capacity of the gauge.)

The FDCs have been used for the assessment of the general quality of the simulations throughout the range of flows. Although a FDC does not represent a conventional measure of simulation quality control, it can help to visualize the differences and consequently to identify the deficiencies of any model's output. In three cases out of five, the curves match reasonably well through most of the flow range (Figures 5, 6 and 9). The FDCs for the observed and generated flows for the Sundays and Mhlatuze rivers display the differences in the low-flow domain (Figures 7 and 8). These differences may be attributed to the differences in the shape of CPI and flow duration curves for individual calendar months. A zero daily source CPI value in a certain month at a certain rainfall station could, in principle, correspond to a range of possible flows at the destination flow site. In the algorithm, the percentage point of the zero CPI value in each month always equals the 'first' fixed percentage point with zero CPI. Consequently, the estimated destination flow for that month is always the same. If the destination site FDC for that month has fewer (or no) zero values than the
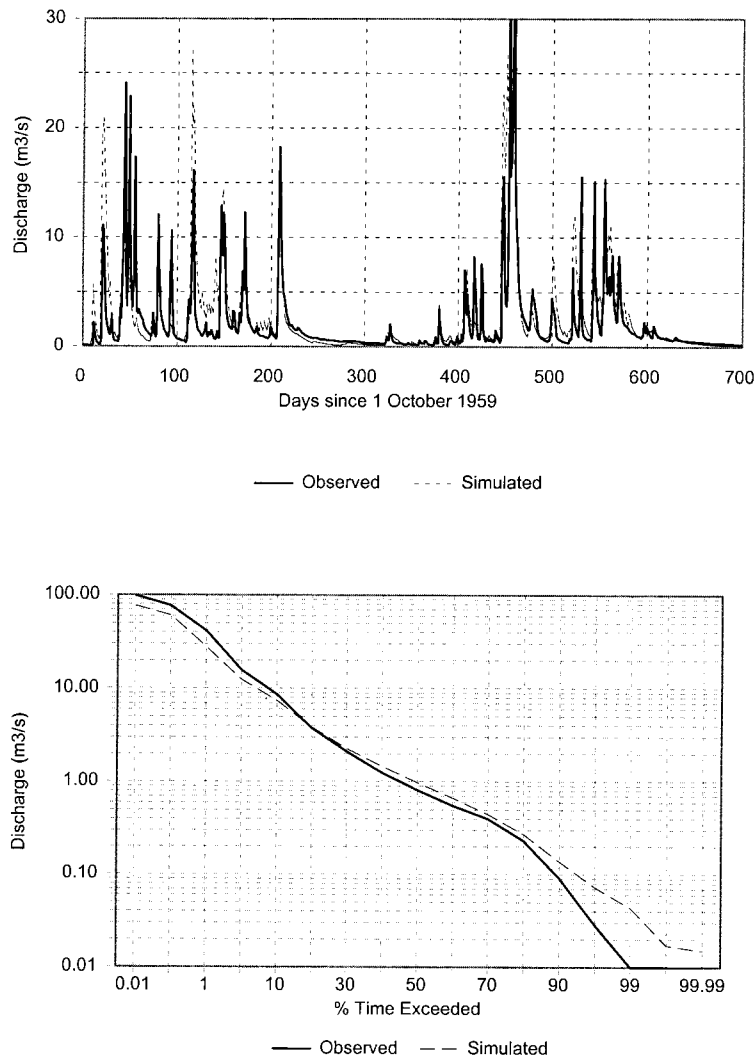
Figure 7. Observed and simulated daily hydrographs (top) and annual 1-day flow duration curves (bottom) for the Sundays River

source duration curve, the destination flow may be overestimated. This happens in some cases in a low-flow domain and is reflected by Figures 7 and 8. This problem was identified by Hughes and Smakhtin (1996) in the original version of the spatial interpolation algorithm, but it was not of critical importance, because most of the rivers where the algorithm was applied were perennial. In the case of the Mhlatuze catchment, the rainfall gauge 0303711W has a very different rainfall pattern to that of the other three, and its CPI duration curves (especially in the wet months of December, January and February) are steeper than the corresponding destination FDCs. This leads to the overestimation of generated destination 'low flows' (Figure 8). A similar problem, although to a lesser extent, occurs with some other rainfall stations in Mhlatuze catchment and also in the Sundays River catchment (Figure 7). As the weights of all rainfall stations in each individual catchment were equal, no attempt has been made to 'filter out' the undesirable rainfall stations. Such rainfall stations with records having a sequential pattern different to that of the gauged stream flow as the catchment outlet are most likely to be located, for example, outside the catchment boundary, or almost on the boundary (such as 0303711W).
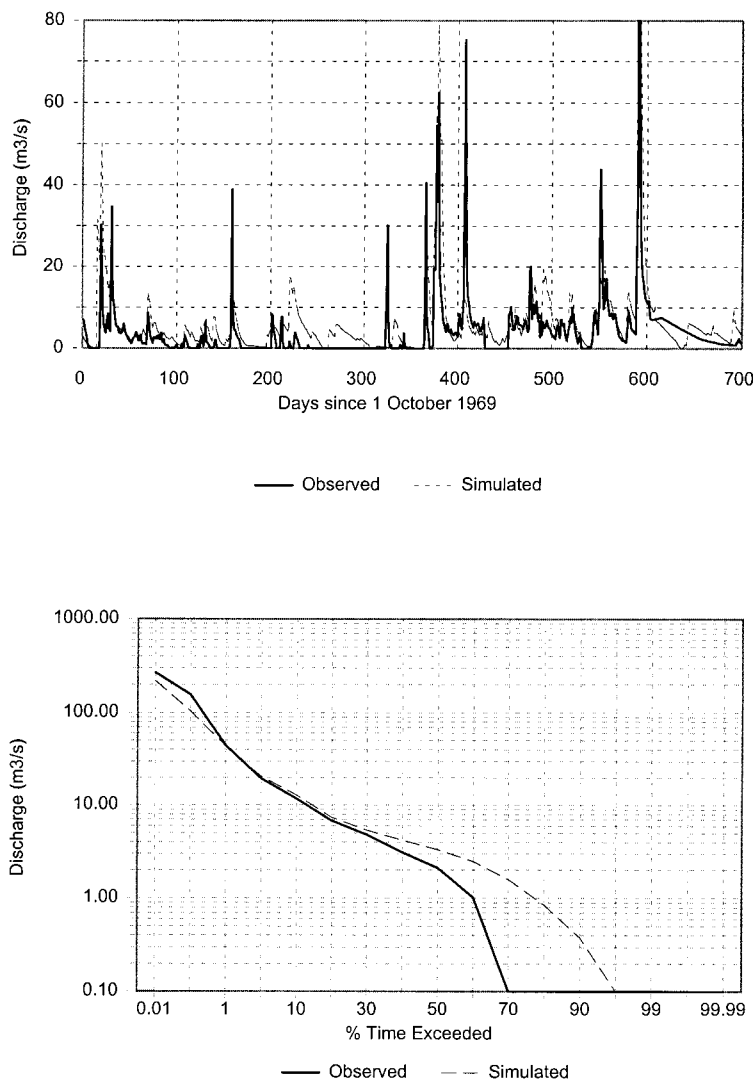
Figure 8. Observed and simulated daily hydrographs (top) and annual 1-day flow duration curves (bottom) for the Mhlatuze River

The other problem that affects the quality of the simulations in some of the cases considered (e.g. the Mhlatuze catchment) is the length of flow record and quality of the stream flow data that were used to construct the destination site FDCs. The longer the record, the more reliable and representative is the constructed FDC. The FDCs for particular months, as opposed to the annual curve, are especially sensitive to the length of record and to the amount of missing data. The shorter the record and the more data are missing, the fewer data are available for the construction of FDCs. The alternative in such circumstances would be to use 1-day annual FDC for all 12 months of the year instead of 12 different curves for each month. This approach has been used as an experiment in several catchments and has been found to have the effect of slight reduction on flow variability. The discussed problem of the record length is only relevant when the FDCs are established on the basis of observed data. As the algorithm is intended for application at ungauged sites, the destination FDCs will have to be calculated either by regional methods or from synthetic monthly time-series as described in one of the previous sections.
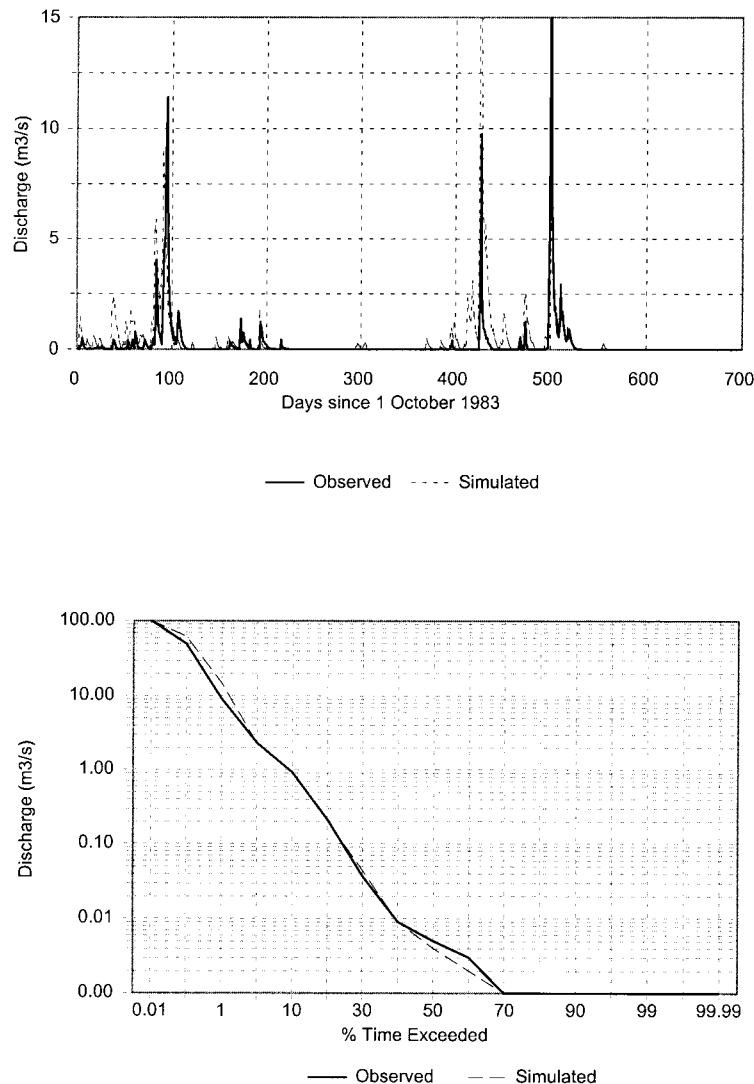
Figure 9. Observed and simulated daily hydrographs (top) and annual 1-day flow duration curves (bottom) for the Koonap River

One typical characteristic of the output daily flow time-series is the presence of minor flow fluctuations, which result from the nature of the source CPI time-series (Figure 1). Daily CPI values reflect even minor increases in catchment wetness, which consequently are transferred to the resultant flow. In dry months of the year, such fluctuations caused by some minor rainfall events break smooth recession limbs in the generated time-series. This may represent an issue in cases where the representation of true recession characteristics is required. Minor increases in CPI may be filtered out by introducing certain daily rainfall thresholds below which the daily rainfall will not be added to the CPI.

As has already been mentioned, the weights of source rainfall stations in each catchment were initially taken to be equal (e.g. in the case of the Mac-Mac River with two stations, the weights were 0·5, in the case of the Mhlatuze catchment with four stations the weights were 0·25, etc). Although this seems to be an arbitrary approach, the subsequent experiments with different weights in several catchments have shown that changing the weights normally does not have a profound effect on the resultant destination time-series, with only a

Table II. Statistics of fit between observed daily flows and daily flows simulated by the CPI-based spatial interpolation algorithm and by the VTI deterministic daily model

Catchment and simulation period	Data	Mean, $m^3 s^{-1}$	SD, $m^3 s^{-1}$	Maximum, $m^3 s^{-1}$	Minimum, $m^3 s^{-1}$	R^2	CE
Mac-Mac: X3H003	Obs	0·91	0·99	15·7	0·3		
	Sim	0·92	1·05	15·3	0·4	0·62	0·54
1979–1989	VTI	0·78	1·03	23·3	0·3	0·78	0·51
Marite: X3H011	Obs	1·7	2·44	28·9	0		
	Sim	1·57	1·91	28·8	0·01	0·64	0·64
1979–1989	VTI	1·86	3·77	98·4	0·34	0·64	0·08
Sundays: V6H004	Obs	3·3	7·53	93·6	0		
	Sim	2·93	5·64	71·1	0·02	0·58	0·58
1954–1964	VTI	3·32	7·92	104	0	0·57	0·48
Mhlatuze: W1H006	Obs	4·95	10·9	232	0		
	Sim	5·63	9·08	172	0	0·5	0·48
1964–1973	VTI	4·94	10·6	166	0·02	0·49	0·45
Koonap: Q9H016	Obs	0·53	3·06	102	0		
	Sim	0·68	3·97	95·3	0	0·61	0·34
1983–1992	VTI	0·53	3·93	98·2	0	0·74	0·63

slight improvement or deterioration of the results being observed. Assuming equal weights may be just a pragmatic way of specifying them in the absence of observed stream flow data against which to evaluate the results of simulation. A more logical approach, however, would probably be to assign Thiessen polygon weights to source rainfall stations.

The number of source rainfall stations to use in the algorithm is unlikely to be a serious issue. For catchments with areas smaller or similar to those used in the current study (predominantly less than 1000 km^2), the choice normally would be limited to one to five. More stations often are simply not available. Others may be unusable because of the short or non-overlapping records, suspect values or missing data. The use of just one rainfall station is not recommended because all the deficiencies of its data (e.g. missing data) are effectively transferred to the generated destination site flow time-series. On the other hand, one gauge may not always be representative of the spatial variability of catchment rainfall even in small catchments. Figure 10 presents the extracts from the CPI time-series of the two available rainfall stations in the Mac-Mac River and illustrates that the catchment wetness calculated using solely one or the other rainfall station may be very different. Figure 10 also shows how these individual CPI time-series (if used individually) are translated into the resultant destination daily hydrograph. The best coincidence between observed and simulated flow time-series at gauge X3H003 was obtained when the two source rainfall stations were used in the algorithm together (Figure 5).

CONCLUSIONS

The paper has presented an economic method for generating continuous daily stream-flow time-series from observed daily rainfall records in a river catchment. The method uses the time-series and duration curves of a rainfall-related 'current precipitation index' (CPI). This index reflects the current catchment wetness and is defined as a continuous function of precipitation that accumulates on rainy days while simultaneously decaying exponentially during the periods of no rainfall. The process of rainfall-to-runoff conversion is based on the assumption that daily CPI values at rainfall site(s) in a catchment and a destination site's daily flows correspond to similar percentage points on their respective duration curves. The method represents an extension of the non-linear spatial interpolation technique developed by Hughes and Smakhtin (1996).
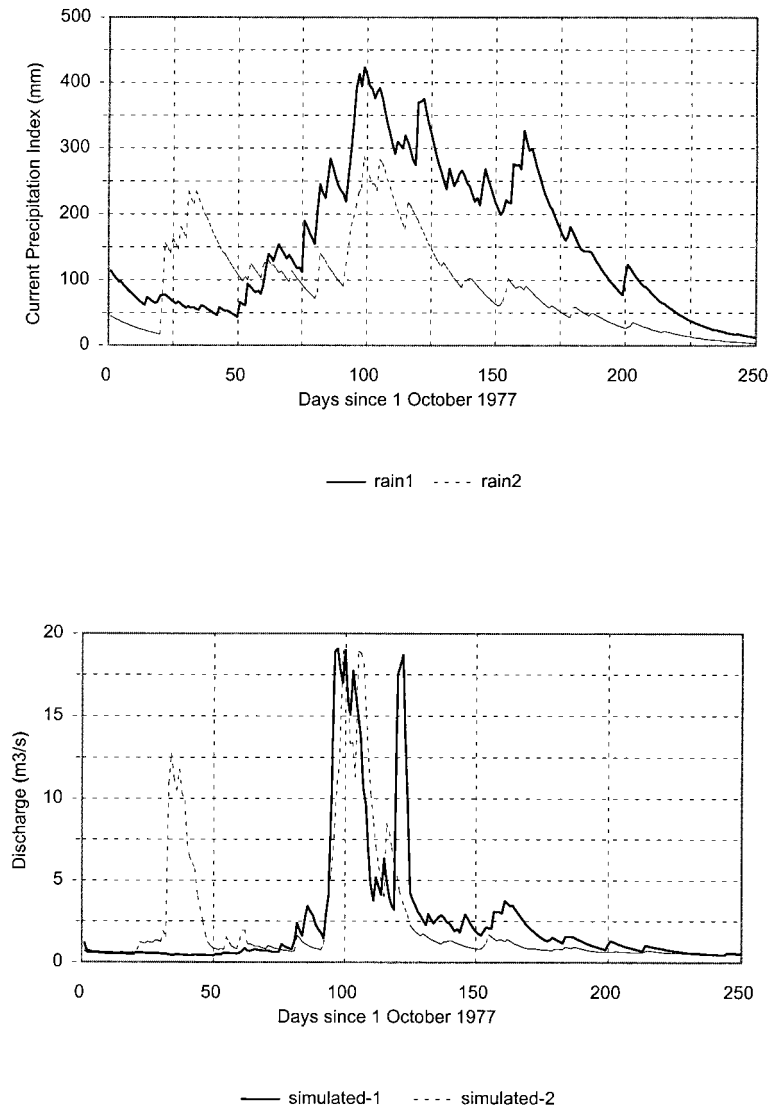
Figure 10. Extracts from the source CPI time-series at two rainfall stations in the Mac-Mac River catchment (top) and corresponding daily hydrographs generated using each station's data independently (bottom)

The method is designed primarily for application at ungauged sites in data-poor regions where the use of more complex and information consuming techniques of stream-flow time-series generation may not be justified. Consequently, it has been tested under stringent conditions, taking into account that no calibration would be possible in practical applications. The method has been found to perform satisfactorily in these first tests. Comparison of observed and generated hydrographs as well as FDCs and fit statistics have shown that the approach is capable of reproducing the general pattern of daily stream flow variability, although low flows and small events frequently appear to be overestimated.

Several problems related to the quality, availability and representativeness of the input rainfall data as well as to the nature of the CPI time-series have been identified. These problems suggest that the following further research and modifications are required if the performance of the suggested time-series generation technique is to be improved: (i) the uncertainty related to the estimation of destination flows in the case of zero source

CPI values needs to be resolved; (ii) minor fluctuations in the source of CPI time-series caused by insignificant rainfall events need to be filtered out by introducing rainfall thresholds below which the daily rainfall will not be added to the CPI; (iii) the suitability of the source rainfall data must be assessed using the criteria of record length, completeness and location with regard to being representative of catchment rainfall. It may also be suggested that Equation (1) used to calculate CPI daily values from daily rainfall data may need some modifications related to the differences in catchment wetness dynamics at different levels of this wetness. It is possible, for example, to introduce different K values in Equation (1) that will be dependent on wetness thresholds. Such thresholds would effectively represent different catchment storages contributing to the output stream flow hydrograph. On the other hand, the advantage of the proposed approach is its simplicity and a limited number of 'parameters'. By introducing additional 'parameters' into this 'model', this advantage may be lost.

The paper has illustrated the application of the method only for the generation of *daily* stream flow hydrographs. At the same time, a similar approach may be applied to generate monthly flow time-series from monthly rainfall data. In this case, the actual monthly step rainfall data most probably may be used instead of the CPI time-series, and the recession coefficient will need a new interpretation.

The performance of the method has been illustrated using the examples of predominantly small catchments (less than 1000 km^2). However, it is expected that the method, in principle, is applicable to large river catchments as well. In this case, the number of initially available rainfall stations with suitable data will be large and it may be appropriate to generate the weighted average catchment wetness first. Dependent on the type of data used, it could be daily or monthly weighted average wetness. It is also possible to split the large catchment into a set of 'homogeneous' subareas and calculate weighted average wetness time-series for each subarea. In this case the subareas will replace the source rainfall sites. On the other hand, the large catchments are more likely to have suitable stream-flow source gauges that can be used for data transfer between the sites and the original version of the spatial interpolation algorithm (which uses stream flow data *only*) will apply.

All these considerations imply that the potential of the approach described for hydrological time-series data generation is not yet fully explored and may constitute a challenging research direction specifically relevant to date-limited regions and/or cases when data provision is required at low costs and/or within a limited time frame.

REFERENCES

Fedora MA, Beschta RL. 1989. Storm runoff simulation using an antecedent precipitation index (API) model. *Journal of Hydrology* **112**: 121–133.

Fennessey N, Vogel RM. 1990. Regional flow-duration curves for ungauged sites in Massachusetts. *ASCE Journal of Water Resources Planning and Management* **116**: 530–549.

FREND. 1989. *Flow Regimes from Experimental and Network Data. I: Hydrological Studies; II: Hydrological Data*. Institute of Hydrology: Wallingford; 345 pp.

Hughes DA. 1997. *Southern African 'FRIEND' — the Application of Rainfall–Runoff Models in the SADC Region*. Report No. **235/1/97**, Water Research Commission: Pretoria, South Africa; 69 pp.

Hughes DA, Sami K. 1994. A semi-distributed, variable time interval model of catchment hydrology — structure and parameter estimation procedures. *Journal of Hydrology* **155**: 265–291.

Hughes DA, Smakhtin VY. 1996. Daily flow times series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrological Sciences Journal* **41**: 851–871.

Institute of Hydrology. 1980. *Low Flow Studies (1-4)*. *Institute of Hydrology: Wallingford*.

Kelman J. 1980. A stochastic model for daily streamflow. *Journal of Hydrology* **47**: 235–249.

Klaasen B, Pilgrim DH. 1975. Hydrograph recession constants for New South Wales streams. *Institute of Engineers, Civil Engineering Transactions* **CE17**: 43–49.

Lall U, Olds J. 1987. A parameter estimation model for ungauged streamflows. *Journal of Hydrology* **92**: 245–262.

Lane EW, Lei K. 1949. Stream flow variability. *Proceedings of the American Society of Civil Engineers* **75**: 935–994.

LeBoutillier DW, Waylen PR. 1993. Stochastic model of flow duration curves. *Water Resources Research* **29**(10): 3535–3541.

Linsley RK, Kohler MA, Paulhus JLH. 1975. *Hydrology for Engineers*, 2nd edn. McGraw-Hill: New York.

Middelkoop H. 1998. Statistical downscaling of discharge series from the Rhine basin for peak flow analysis under changed climate conditions. *Proceedings, International Conference on 'Hydrology in Changing Environments'*, Exeter, Vol. **1**, 175–190.

Midgley DC, Pitman WV, Middleton BJ. 1994. *Surface Water Resources of South Africa 1990*. *Report No*. **298/5.1/94**, Water Research Commission: Pretoria, South Africa.

Mimikou M, Kaemaki S. 1985. Regionalization of flow duration characteristics. *Journal of Hydrology* **82**: 77–91.

Mohseni O, Stefan HG. 1998. A monthly streamflow model. *Water Resources Research* **34**: 1287–1298.

Ni-Lar-Win, Vandewiele GL. 1994. Conceptual daily rainfall–runoff models using correlation analysis. *FRIEND: Flow Regimes from International Experimental and Network Data (Proceedings, Braunschweig Conference, October 1993*. Publication No. **221**. International Association of Hydrological Sciences: Wallingford; 343–348.

Quimpo RG, Alejandrino AA, McNally TA. 1983. Regionalised flow duration curves for Philippines. *ASCE Journal of Water Resources Planning and Management* **109**: 320–330.

Sittner WT, Schauss CE, Monro JC. 1969. Continuous hydrograph synthesis with an API-type hydrologic model. *Water Resources Research* **5**: 1007–1022.

Smakhtin VY. 1999. Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique. *Regulated Rivers: Research and Management*, **15**: 311–323.

Smakhtin VY, Watkins DA. 1997. *Low Flow Estimation in South Africa*. *Report No*. **494/1/97**, Water Research Commission: Pretoria, South Africa.

Smakhtin VY, Hughes DA, Creuse-Naudin E. 1997. Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrological Sciences Journal* **42**: 919–936.

Torfs PJJF. 1998. Statistical disaggregation of hydrological time series using wavelets. *Proceedings, International Conference on 'Hydrology in Changing Environments', Exeter* Vol. **1**: 81–89.

Vandewiele GL, Atlabachew E. 1995. Monthly water balance of ungauged catchments obtained by geographical regionalisation. *Journal of Hydrology* **170**: 277–291.