Application of satellite-derived rainfall estimates to extend water resource simulation modelling in South Africa

T Sawunyama* and DA Hughes

Institute for Water Research, Rhodes University, 6140, South Africa

Abstract

Spatially interpolated rainfall estimates from rain-gauges are widely used as input to hydrological models, but deriving accurate estimates at appropriate space and time scales remain a major problem. In South Africa there has been a gradual decrease in the number of active rain-gauges over time. Satellite-based estimates of spatial rainfall are becoming more readily available and offer a viable substitute. The paper presents the potential of using Climate Prediction Center African daily precipitation climatology (CPCAPC) satellite-based datasets (2001-2006) to drive a Pitman hydrological model which has been calibrated using gauge-based rainfall data (1920-1990). However, if two sources of rainfall data are to be used together, it is necessary to ensure that they are compatible in terms of their statistical properties. A non-linear frequency of exceedance transformation technique was used to correct the satellite data to be more consistent with historical spatial rainfall estimates. The technique generated simulation results for the 2001 to 2006 period that were greatly improved compared to the direct use of the untransformed satellite data. While there remain some further questions about the use of satellite-derived rainfall data in different parts of the country, they do seem to have the potential to contribute to extending water resource modelling into the future.

Keywords: satellite-based rainfall, hydrological model, water resources, South Africa

Introduction

It is widely known that the accuracy of streamflow predictions from a hydrological model is heavily dependent on the accuracy of rainfall inputs (Gourley and Vieux, 2006). Spatial rainfall estimates derived from rain-gauges are widely used as input to hydrological models and as 'ground truth' for satellite rainfall measurements (Seed and Austin, 1990). However, deriving accurate estimates of basin rainfall at appropriate space and time scales has long been a major problem. Generally, the accuracy of spatial rainfall estimates increases with an increase in the number of rain-gauges within the basin. Many problems in hydrological applications include the extrapolation of sparse point measurements of rainfall to a wider spatial domain. According to Schäfer (1991) different methods used to generate spatial rainfall data in South Africa have been demonstrated to give similar results. As a result of rainfall variability, estimates based on few point measurements are very prone to error (Andréassian et al., 2001). However, South Africa and many other developing countries are experiencing a sharp decline of active rain-gauge networks with vast areas ungauged, while radar is not always a feasible proposition on the grounds of cost, technical infrastructure and topography. It is therefore difficult to obtain long representative rainfall records that cover periods long enough to allow for current and future water resource assessments. Inevitably, sustainable planning of water resources requires information on the present spatial and temporal variability of rainfall. As a consequence, there is an increasing demand from the climate and hydrological communities, for accurate spatial rainfall estimates for all basin

scales and over extended periods. The incorporation of satellite-based rainfall estimates in hydrological modelling has the potential to improve our capability to constrain uncertainty in rainfall inputs and extend water resource simulations.

Satellite-based rainfall estimates are becoming more readily available and are expected to offer an alternative to ground based rainfall estimates in the present and the foreseeable future. The use of satellite-based information to improve spatial rainfall estimates has been widely reported (Hsu et al., 1999; Sooroshian et al., 2000; Grimes and Diop, 2003). However, downscaling of remotely sensed data remains an issue and hence these satellite-based rainfall estimates do not compare well with the gauge data. The problem of scale holds when measurements of rainfall rates provided by rain gauge data are compared with the areal time averaged rainfall remotely sensed from satellite borne sensors (e.g. Sandham et al., 1998). Consequently, models have been developed to combine satellite and raingauge data to account for local and regional variability in cloud and rainfall relations (Todd et al., 1999). However, the accuracy of the final operational satellite-based rainfall estimates are dependent on these interpretative models that are also subject to calibration. In addition, there are frequently insufficient gauge data available to calibrate the satellite-based estimation methods. The assessment and quantification of uncertainty affecting remote sensing estimates of hydrological variables has been explored (Huffman et al., 1997). Although extensive literature on satellite-based rainfall estimates exists, this has concentrated more on development of methods to derive rainfall from satellite imagery. However, very few studies have so far investigated the application of these data sets in hydrological models. Recently, studies were conducted to evaluate the performance of hydrological models using operational satellite rainfall estimates in southern Africa (Thorne, et al., 2001; Hughes, 2006a, 2006b; Wilk et al., 2006). These studies suggested the need to correct the satellite-based
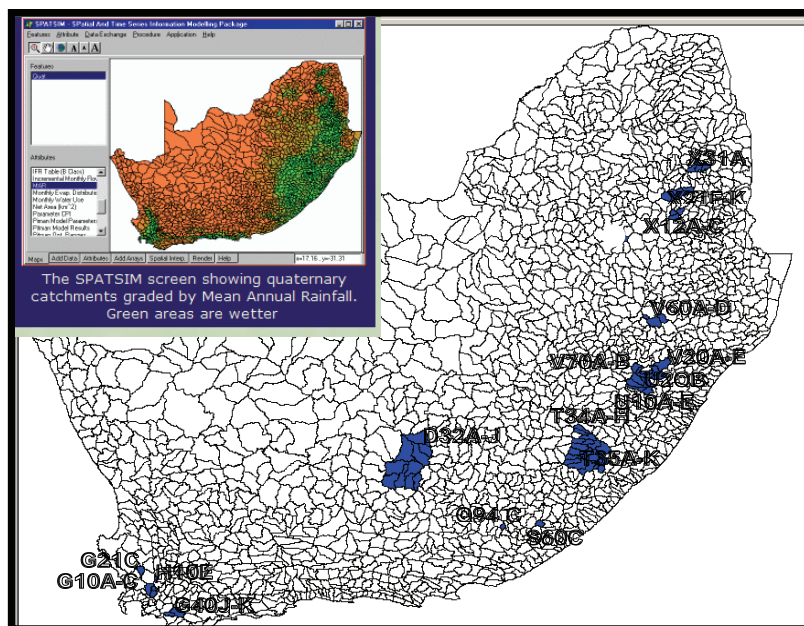
Figure 1
Map of South Africa showing location of sub-basins (quaternary catchments)

rainfall data to be consistent with gauge data before using them as inputs to hydrological models.

The inadequacy of ground-based rainfall data and the different periods for which gauge and satellite data are available makes the correction process even more problematic. Under ideal situations, correction factors would be quantified by comparing coincident satellite-based rainfall and spatially interpolated gauge-based estimates. In recent times, and possibly in future, it is unlikely that these data sets will be available over the same periods, given the continuous decline in active rain-gauges. Simple corrections, such as linear scaling (Hughes, 2006b) and manual fitting of simple power functions of rainfall frequency characteristics (Wilk et al., 2006) are also not very easy to determine and apply, under normal circumstances, for a wide range of basins. Therefore, the lack of appropriate correction procedures and of spatially continuous rainfall datasets with sufficient resolution in space and time for hydrological applications has prompted the need to further investigate the procedures to be used for the effective application of satellite rainfall data.

The paper reports on the potential of using high spatial resolution (0.1°) operational satellite rainfall data, to extend the simulations of the Pitman monthly hydrological model in situations where there are now too few or no rain-gauge data to allow reliable estimates of spatial rainfall in South Africa. The paper presents a description and evaluation of a rainfall frequency of exceedance curve algorithm that has been used to merge satellite-based rainfall and historical spatial gauge-based rainfall estimates. The objective of the approach is to ensure that the two rainfall data sets have consistent properties and can be used as input to a hydrological model with a single set of parameters.

Data and methods

The historical monthly spatially averaged WR90 rainfall (1920-1990) and mean monthly evaporation data for selected sub-basins (quaternary catchments) in South Africa were obtained from the WR90 reports (Midgley et al., 1994). These datasets were used to calibrate the recently modified version (including surface-groundwater interactions) of the Pitman model (Hughes,

2004) against all observed data available during this period. The objective of the study was to assess the use of the hydrological model, calibrated against rain-gauge data from 1920-1990, and forced with satellite rainfall data available from 2001 onwards. The observed flow data and rain-gauge station monthly rainfall data were obtained from the Department of Water Affairs and Forestry (DWAF). Several sub-basins ranging from small to medium size, covering a wide range of hydro-climatic conditions were selected in South Africa as shown in Fig. 1. All of the analyses were undertaken using the facilities available within the SPATSIM (Spatial and Time Series Information Modelling) software package (Hughes and Forsyth, 2006).

Satellite-based precipitation data

There are several algorithms reported in literature, which are used to derive final satellite rainfall data sets from satellite imagery:

- National Oceanic and Atmospheric Administration's Climate Prediction Center Rainfall Estimation Algorithm RFE2.0 (NOAA's CPC RFE2.0; 1 day, 0.1 degree; Xie et al., 2002).
- Multi-satellite Precipitation Analysis-Real Time (MPA-RT, Laws et al., 2004).
- Microwave InfraRed Algorithm (MIRA; 1 day, 0.1 degree; Todd et al., 2001).
- Global Precipitation Climatology Project (GPCP; 1 day, 1 degree; Huffman et al., 1997).
- Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; 6 hour, 0.25 degree; Hsu et al., 1999; Sooroshian et al., 2000).

A description and evaluation of each of these algorithms used to generate the satellite rainfall estimates is beyond the scope of this paper, as the focus is on application of operational satellite product. However, a review of literature shows that there are extensive studies (e.g. Kidd, 2001) on the application of satellite imagery to rainfall estimation. Most of the work looked at global rainfall data sets for climatological purposes and a majority of the algorithms make use of passive microwave (PMW) imagery

from sensors on polar orbiting satellites or Meteosat thermal infrared (TIR) from geostationary satellites. The recent techniques use a combination of PMW, TIR and other wavelengths from different platforms (e.g. Xu et al., 1999). Todd et al. (2001) and Adler et al. (1994) have used PMW information to continuously recalibrate TIR-based estimates whenever coincident images exist, combining the high spatial and temporal resolution of the TIR imagery with the better representation of rainfall in the microwave. A review by Govender et al. (2007) highlighted the differences between multispectral and hyperspectral data; spatial and spectral resolutions and provided a detailed focus on the application of hyperspectral imagery in water resources studies. There are quite complex issues related to remote sensing of rainfall inter alia the physics, assumptions used, models, sensors and the differences in spatio-temporal resolutions leading to a varying degree of quality in the final operational estimates used for water resources applications. Todd et al. (1999) showed that the accuracy of the final operational satellite-based rainfall estimates are dependent on the interpretative models used to generate them which are subject to calibration and that there are frequently insufficient rain gauge data available to calibrate the satellite-based estimation algorithms. On the other hand, radar measurements are not a feasible alternative for this purpose.

Some of the global satellite products have already been used in hydrological modelling studies. The GPCP and PERSIANN data sets were used as inputs to the Pitman hydrological model in southern African basins (Hughes, 2006b). In addition, Wilk et al. (2006) used rainfall estimates from special sensor microwave (SGPROF) estimated using the Goddard Profiling Algorithm at 0.5° resolution, to estimate spatial rainfall (1991-2002) in the Okavango River basin. Hughes (2006a) showed that the relationships between the satellite estimates and gauged rainfall are different in different regions and for different sources, while Hughes (2006b) indicated that the hydrological model behaves differently to rainfall inputs from different sources.

There is little information on the application of satellite-based rainfall data with higher spatial resolution. Layberry et al. (2005) prepared a database of daily means of rainfall at 0.1° spatial resolution using the MIRA data for use in the Southern African Regional Science Initiative (SAFARI 2000) project. These data are referred to as SAFARI 2000 daily rainfall estimates available from 1993-2001. Recently, as part of developing spatially continuous and accurate rainfall data sets, NOAA's Climate Prediction Center derived gridded daily rainfall totals at 0.1° spatial resolution for Africa (Love et al., 2004), and the data are currently available from 2001-2006. These are referred to as Climate Prediction Center African Daily Precipitation Climatology (CPCAPC). The SAFARI 2000 and CPCAPC data sets, are both freely available from Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) and NOAA Data Center, respectively. In this study the CPCAPC data have been used because they are available for more recent years when rain-gauge observations have declined even further.

The CPCAPC data were derived by integrating satellite precipitation estimates with station-based rain gauge data to create a bias-reduced daily final precipitation estimate using the CPC RFE2.0 algorithm (Xie et al., 2002). The CPCAPC data were created from four sources; daily Global Telecommunications Systems (GTS) rain-gauges from up to 1 200 stations (far fewer stations make it into CPCAPC), Special Sensor Microwave/Imager (SSM/I) satellite precipitation estimates at a frequency of up to 4 times a day, the Advanced Microwave Sounding Unit (AMSU) satellite rainfall estimate and Global Precipitation Index (GPI) cloud-top IR temperature precipitation estimates
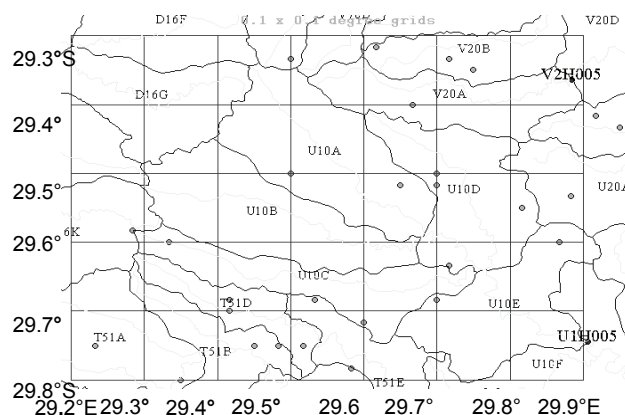

Figure 2
*Gridded representation for deriving satellite-based rainfall estimates for individual sub-basins.
(The polygons represent sub-basin boundaries, the unlabelled points are rain-gauges, while V2H005 and U1H005 are flow gauges)*

on a half-hourly basis (Xie et al., 2002). The final products are daily binary data and graphical output files produced at 6Z-6Z, meaning that the 'daily' rainfall totals are accumulations from 06:00 Universal Time over the next 24 h. The spatial extent of the data covers 40°S-40°N and 20°W-55°E, the whole of the Africa region and surrounding regions and the spatial resolution of 0.1°. The CPCAPC datasets are available from January 2001 to the present and can be accessed online (www.cpc.ncep.noaa.gov/products/fews/) or obtained on a CD-ROM. However, the datasets used in this study are available up to October 2006, the date for which they were downloaded.

The daily CPCAPC data were converted from binary to text format using a simple Delphi data extraction program. The derivation of spatial estimates for each sub-basin involved first selecting the appropriate number of grid squares covering each sub-basin (e.g. U10A in Fig. 2). The sub-basin spatial rainfall estimates were based on simple averages of the daily rainfall totals from appropriate grids within each sub-basin.

Transformation of satellite-based rainfall using rainfall frequency of exceedance curves (RFC)

A technique has been developed to correct the spatial satellite-based rainfall estimates to be consistent (in terms of statistical properties) with gauge-based estimates, using WR90 rainfall data, before they are both used as inputs to a hydrological model. The method attempts to overcome the problem associated with correcting one data set using another data set that is not coincident in time. The development of the non-linear technique (rainfall transformation algorithm) was based on the approach used by Hughes and Smakhtin (1996) to patch and extend flow time series. The approach transforms source rainfall time series to destination rainfall time series at the same spatial point through frequency of exceedance curves. Rainfall frequency of exceedance curves (RFC) are a summary of the relationship between rainfall magnitude and frequency of occurrence, and therefore the variability within a time series. Following the original transformation algorithm (Hughes and Smakhtin, 1996), the procedure involves estimation of the percentage point for each month's rainfall from the source time series (original satellite data). The transformed rainfall for that month is the value for the same percentage point taken from the destination RFC (WR90
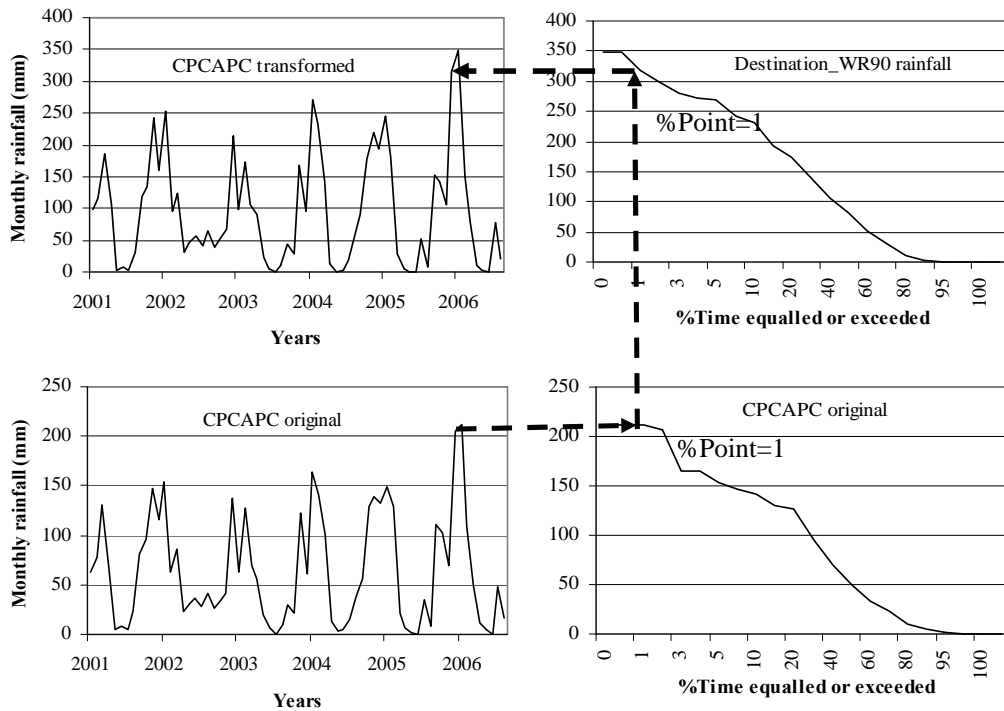
Figure 3
Illustration of the rainfall transformation algorithm

rainfall). The procedure is illustrated in Fig. 3. The source RFC is calculated from the original satellite data time series, however, a prerequisite of the technique is that a destination RFC can also be quantified. The assumption is that this RFC must be representative of the WR90 rainfall data (which is, of course, not available) for the same period as the satellite data.

The entire WR90 rainfall time series (1920-1990) could not be considered representative of the climate over the satellite period 2001-2006. Therefore, the selection of the appropriate period within the WR90 rainfall series that is used to derive the destination RFCs was crucial. This was done by trying to visually identify a period within 1920-1990 that was climatically similar to the 2001-2006 period using both the WR90 simulated flows and DWAF observed flows. To complement the approach, data from the limited number of available DWAF rainfall stations (with data up to 2006) were also used to check rainfall sequences throughout the period up to 2006. These rainfall station data sets were used because the observed flow sequences are possibly affected by upstream artificial influences and may introduce bias in the selection of an appropriate part of the series. However, for the sub-basins used in the study, the natural flow sequences have not been heavily affected by major abstractions and impoundments.

The WR90 rainfall time series for the period 1968-1974 was found to be suitable to derive destination RFCs (Table 1) for most of the summer rainfall region (central to eastern parts of the country), while the period 1970-1976 was considered suitable for the winter rainfall region (western parts of the country). There were several exceptions to this general rule (Table 1, column 6).

Hydrological model response to rainfall inputs

The original and transformed CPCAPC estimates were used in the revised groundwater version of the Pitman model which had been previously calibrated using the WR90 spatially averaged historical rainfall data (1920-1990). The hydrographs of the observed and simulated time series were visually assessed for both the calibration period and the satellite period (using both original and transformed satellite data). The corresponding flow duration curves were also compared. In addition, a set of goodness-of-fit statistics were calculated to provide objective comparisons of model performance. The statistics used were the coefficient of determination (R^2) and coefficient of efficiency (CE; Nash and Sutcliffe, 1970), percentage difference of mean flows and of standard deviations, applied to both un-transformed flows and log-transformed flows (to remove the bias towards high flows). The DWAF-observed flows were used as reference flows in all comparisons.

Results

Table 1 lists the sub-basins (quaternary catchment names), the DWAF streamflow gauges associated with them, the catchment areas and the simulation periods for the WR90 and satellite data. Table 2 provides summaries of the simulation results for the calibration period using WR90 rainfall data and the satellite period using the original and corrected satellite data. Table 2 indicates that it is not always a straightforward task to evaluate the use of the satellite data due to relatively poor calibration results using the WR90 rainfall data (G40J-K, for example). Some of the reasons for poor calibrations appear to be ill-defined water resource development effects represented in the observed data, while the difficulties of defining rainfall input time series based on limited rainfall gauge data is also assumed to play a role.

Figures 4 and 5 provide some graphical detail for sub-basin V70A as an example. A comparison of time series graphs of the original satellite-based rainfall data with observed DWAF station rainfall data for the V70A sub-basin indicates that the monthly rainfall distribution patterns and timing are generally consistent (Fig. 4, left side). However, systematic under-estimations of the monthly rainfall values occur during individual months throughout the period. Differences were observed from comparisons of the RFCs for the WR90 rainfall estimate (1968-1974), original CPCAPC data and corrected/transformed CPCAPC estimates (Fig. 4, right side) for the V70A sub-basin. As might be expected, the rainfall frequency characteristics of the corrected CPCAPC are in close agreement with WR90

TABLE 1					
Sub-basins, DWAF gauges, areas and simulation periods					
Sub-basins	Gauge	Area (km²)	Simulation period		WR90 destination RFC period
			WR90	CPCAPC	
D32A-J	D3H015	8330	1980-1990	2001-2006	1925-1935
G10A-C	G1H020	609	1966-1990	2001-2006	1970-1976
G21C	G2H012	244	1973-1990	2001-2006	1970-1976
G40J-K	G4H006	600	1963-1990	2001-2006	1970-1976
Q94C	Q9H019	76	1972-1990	2001-2006	1968-1974
S60C	S6H003	215	1971-1990	2001-2006	1968-1974
T34A-H	T3H005	2597	1952-1990	2001-2006	1968-1974
T35A-K	T3H006	4268	1952-1990	2001-2006	1968-1974
U10A-E	U1H005	1744	1960-1990	2001-2006	1968-1974
U20B	U2H007	358	1960-1990	2001-2006	1968-1974
V20A	V2H005	267	1972-1990	2001-2006	1968-1974
V20A-D	V2H002	937	1950-1990	2001-2006	1968-1974
V20A-E	V2H004	1546	1960-1986	2001-2006	1968-1974
V60A-B	V6H004	658	1954-1990	2001-2006	1968-1974
V60D	V6H003	312	1954-1990	2001-2006	1968-1974
V70A	V7H017	276	1973-1990	2001-2006	1968-1974
V70B	V7H016	124	1973-1990	2001-2006	1968-1974
X12A-C	X1H016	581	1970-1990	2001-2006	1968-1980
X21F-K	X2H015	1554	1958-1990	2001-2006	1968-1980
X31A	X3H001	174	1959-1990	2001-2006	1968-1980

Figure 4
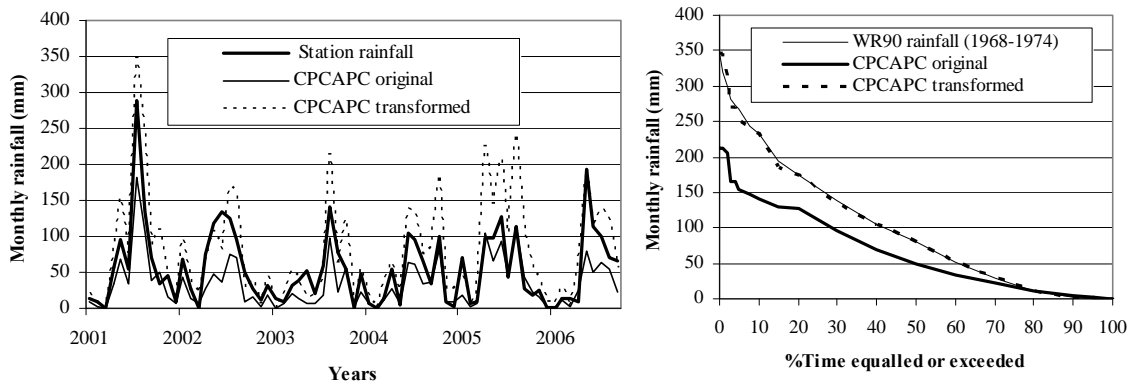*Comparison of time series of CPCAPC rainfall data (original and transformed) with station rainfall (left side) and frequency of exceedance curves of monthly rainfall totals of WR90 spatial data and CPCAPC data (right side) for sub-basin V70A.*
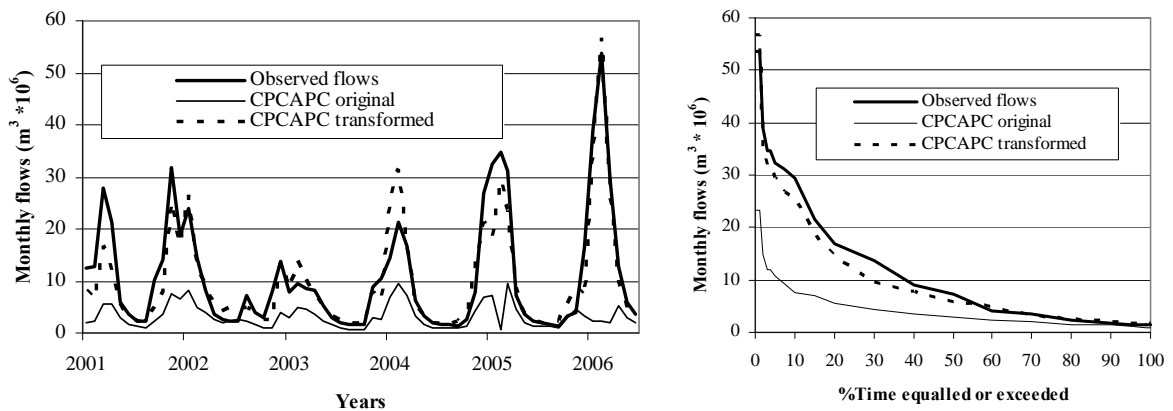

Figure 5
Comparison of monthly flow time series (left) and the comparison of flow duration curves (right) for the October 2001-September 2006 period for V70A sub-basin

Sub-basins	Data	Un-transformed flows (Q)				Log-transformed flows {ln(Q)}			
		%Diff Mn	%Diff Stdv	R²	CE	%Diff Mn	%Diff Stdv	R²	CE
D32A-J	WR90	15.5	7.4	0.91	0.89	-71.9	-33.4	0.57	0.47
	Sat (orig.)	72.7	1.1	0.39	0.18	70.2	-23.8	0.16	-0.17
	Sat (corr)	4.7	-7.7	0.41	0.34	9.9	7.3	0.14	-0.17
G10A-C	WR90	20.4	6.7	0.90	0.86	15.9	-15.6	0.84	0.77
	Sat (orig.)	-79.4	-80.4	0.64	-0.11	-67.0	31.6	0.59	-2.87
	Sat (corr)	12.6	8.5	0.77	0.71	-1.4	25.9	0.65	0.44
G21C	WR90	0.1	-4.0	0.77	0.77	49.6	-26.5	0.61	0.52
	Sat (orig.)	-67.6	-73.7	0.76	0.33	-2.4	-34.7	0.72	0.68
	Sat (corr)	-20.0	-13.8	0.54	0.52	22.6	-25.6	0.72	0.68
G40J-K	WR90	-3.5	20.3	0.45	0.16	52.8	-24.9	0.60	0.55
	Sat (orig.)	-15.7	19.3	0.32	-0.08	40.3	-40.4	0.60	0.56
	Sat (corr)	-17.1	17.1	0.25	-0.22	82.3	-47.1	0.61	0.51
H10E	WR90	-20.9	-28.9	0.73	0.68	-15.7	11.8	0.71	0.61
	Sat (orig.)	-83.6	-82.2	0.78	-0.13	-188.5	40.4	0.78	-1.49
	Sat (corr)	-18.1	-42.7	0.79	0.67	28.3	-23.5	0.71	0.66
Q94C	WR90	-5.6	-9.7	0.78	0.78	4.5	-10.2	0.69	0.68
	Sat (orig.)	-77.9	-63.2	0.21	-0.35	-309.6	6.6	0.47	-2.73
	Sat (corr)	-7.2	33.3	0.32	-0.28	-60.7	21.3	0.60	0.28
S60C	WR90	6.9	20.5	0.52	0.27	-15.2	-12.1	0.65	0.64
	Sat (orig.)	-41.0	-36.5	0.63	0.49	-452.7	-6.7	0.68	0.28
	Sat (corr)	-23.8	48.5	0.38	-0.40	-437.0	11.6	0.65	0.22
X21F-K	WR90	10.7	8.4	0.65	0.62	8.4	-24.1	0.71	0.63
	Sat (orig.)	55.8	5.5	0.63	0.18	29.4	-27.1	0.66	0.05
	Sat (corr)	-10.7	-46.4	0.70	0.60	2.8	-26.8	0.77	0.74
X31A	WR90	-1.2	-14.4	0.74	0.74	4.7	-21.1	0.71	0.70
	Sat (orig.)	-80.0	87.0	0.43	-0.58	-99.0	-31.3	0.47	-3.48
	Sat (corr)	-6.8	-23.5	0.53	0.53	1.1	-13.1	0.71	0.71
T34A-H	WR90	9.2	9.3	0.74	0.67	11.4	-23.7	0.71	0.65
	Sat (orig.)	-24.5	-25.8	0.61	0.54	-0.1	-34.7	0.61	0.59
	Sat (corr)	14.1	7.3	0.56	0.44	10.1	-19.4	0.62	0.57
T35A-K	WR90	9.9	-4.8	0.65	0.62	12.0	-28.5	0.71	0.61
	Sat (orig.)	0.1	23.9	0.36	-0.05	4.8	-25.1	0.64	0.61
	Sat (corr)	6.7	3.4	0.46	0.34	8.4	-24.9	0.70	0.64
U10A-E	WR90	-8.3	-31.3	0.73	0.70	6.4	-25.7	0.86	0.80
	Sat (orig.)	-50.7	-62.4	0.78	0.32	-13.0	-27.5	0.88	0.67
	Sat (corr)	13.5	-33.5	0.69	0.65	2.5	-22.7	0.82	0.80
U20B	WR90	-7.1	0.9	0.68	0.65	-6.5	-3.8	0.67	0.64
	Sat (orig.)	-61.0	-63.4	0.79	0.18	-79.3	-8.3	0.78	0.07
	Sat (corr)	-25.6	-15.0	0.56	0.49	-29.8	1.1	0.78	0.66
V20A	WR90	-15.4	-15.7	0.84	0.82	-17.2	22.1	0.81	0.64
	Sat (orig.)	55.8	-55.9	0.79	0.29	-55.1	13.7	0.88	0.11
	Sat (corr)	-25.6	-23.5	0.72	0.65	-26.0	21.8	0.84	0.58
V20A-D	WR90	-2.4	-11.0	0.69	0.68	13.3	-3.5	0.83	0.83
	Sat (orig.)	11.7	3.4	0.58	0.49	7.7	-3.0	0.76	0.72
	Sat (corr)	-18.8	-24.5	0.61	0.58	-5.1	-7.6	0.79	0.77
V20A-E	WR90	13.2	-6.2	0.75	0.73	13.4	-18.7	0.77	0.70
	Sat (orig.)	22.9	-4.3	0.59	0.52	27.0	-25.9	0.79	0.60
	Sat (corr)	-13.0	-28.4	0.55	0.54	9.4	-30.1	0.75	0.70

TABLE 2
Comparison of simulated statistics based on original and transformed CPCAPC estimates

TABLE 2 (continued)									
Sub-basins	Data	Un-transformed flows (Q)				Log-transformed flows {ln(Q)}			
		%Diff Mn	%Diff Stdv	R²	CE	%Diff Mn	%Diff Stdv	R²	CE
V60A-B	WR90	-11.1	-11.6	0.57	0.55	28.5	-27.8	0.63	0.61
	Sat (orig.)	-38.8	-49.3	0.50	0.42	-48.3	-28.0	0.74	0.71
	Sat (corr)	9.6	-14.8	0.51	0.49	199.4	-21.3	0.73	0.62
V60D	WR90	-8.3	-7.3	0.53	0.48	-24.0	-6.4	0.60	0.58
	Sat (orig.)	-26.2	-34.9	0.54	0.51	-9.4	-17.2	0.63	0.63
	Sat (corr)	4.3	-14.3	0.54	0.53	-109.6	-18.9	0.63	0.55
V70A	WR90	-4.0	-11.8	0.68	0.68	2.7	-12.9	0.80	0.79
	Sat (orig.)	-62.0	-66.6	0.87	0.17	-43.0	-18.8	0.83	0.23
	Sat (corr)	-3.2	-5.7	0.88	0.88	2.1	-9.4	0.87	0.87
V70B	WR90	1.8	-14.7	0.70	0.70	20.1	-11.5	0.82	0.80
	Sat (orig.)	-51.7	-60.5	0.75	0.31	-72.3	-23.0	0.79	0.57
	Sat (corr)	9.2	13.7	0.65	0.53	19.7	-5.4	0.72	0.70
X12A-C	WR90	0.2	-16.8	0.50	0.48	7.7	-26.9	0.66	0.62
	Sat (orig.)	62.7	70.7	0.30	-1.64	53.1	-16.8	0.52	-0.09
	Sat (corr)	3.4	-30.0	0.72	0.69	15.2	-21.7	0.75	0.70

Notes: '*orig*' *represents original &* '*corr*' *represents corrected satellite rainfall data.*
% Diff Mn is the % difference of the simulated mean monthly flow from the observed mean monthly flow;
% Diff Stdv is the % difference of the standard deviation of simulated monthly flows from the standard deviation of observed monthly flows;
R² is the coefficient of determination; CE is the coefficient of efficiency (Nash-Sutcliffe coefficient).

spatial estimates for all the sub-basins. The original satellite estimates under-estimated monthly rainfall totals by up to 40%, mostly in wet years which clearly has a major impact on stream-flow simulations.

Figure 5 shows the flow time series and flow duration curves respectively, for the observed flows and the simulated flows based on the original and corrected (or transformed) satellite estimates for the V70A sub-basin. The simulation based on the original CPCAPC satellite data shows substantial under-estimation of both high and low flows. The seasonal hydrographs of the observed flows are generally well represented by the corrected CPCAPC estimates, while individual peak months are frequently poorly simulated. The corrected flow duration curves are in close agreement with the observed flow duration curve, while the original flow duration curve shows consistent under-estimation of monthly flows by approximately 60%.

Across the 20 sub-basins included in the study, the original satellite rainfall data appear to substantially under-estimate rainfall in 13 (65%) cases. This conclusion is based on the large percentage errors in both mean monthly un-transformed flows and log-transformed flows. In some cases the degree of under-estimation is quite severe (Table 2; G10A-C, H10E, Q94C, U20B and X31A). Most of these are in areas with steep topography where the effects of orographic rainfall are expected to be quite high and not represented by the satellite data. In most cases the corrected satellite data, after transformation using the WR90 RFCs, generated substantially improved simulation results based on the majority of the statistical measures used (Table 2). There were, however, a number of cases where the improvements were either marginal or where some aspects of the simulations showed improvement, while others were worse.

There were 4 cases (D32A-J, V20A, X12A-C and X21F-K) where the original satellite data generated simulations that were greatly in excess of the observed data and some of these were improved after correction of the satellite data. There were only two cases (T35A-K, V20A-D) where the original satellite data generated acceptable results in terms of volume comparisons. In both of these the other statistics were improved after

satellite data correction. This illustrates that the RFC transformation approach may be useful for correcting more than systematic errors in the original satellite rainfall depth.

Table 2 can also be used to compare the corrected satellite data simulation results with the WR90 rainfall based calibration results. There are only 2 cases where the satellite results can be considered better (V70A and X12A-C), while 9 can be considered similar and 9 worse. It must be recognised that the satellite period is shorter than the calibration period and that a few poorly simulated months will have a greater impact than a similar number during the much longer calibration periods.

Discussion and conclusions

The study demonstrates that rainfall estimates obtained from remote sensing data such as satellite imagery may be the best, or the only, choice to drive a rainfall-runoff model in situations where too few rain-gauges are available to account for spatial rainfall variability. Elsewhere, Barrera et al. (2007) showed that adequate remote sensing methods provide full spatial coverage, and even when rainfall estimates at a single pixel are not precise they introduce smaller errors in area-averaged estimations over sub-basins. However, there remains the problem of transferring information from one scale to another and the validation of remotely sensed suffers from a lack of knowledge about the spatial distribution of variables of concern due to the point nature of traditional rain gauge measurements (Todd et al., 1999).

The results showed the potential to generate or extend rainfall records and establish continuous data sets using satellite-based rainfall data. The procedure followed was the calibration of the hydrological model against historical spatial gauge data from 1920-1990, followed by an assessment of the model forced with satellite rainfall data estimates for the period 2001-2006. As expected, for the satellite data to be used with the calibrated model parameters, some form of transformation was necessary. However, there are too few gauges with data that overlap the satellite-derived estimates period and the uncertainties associated with such estimates could potentially bias results. Even if

gauge data are combined through interpolation approaches and satellite data are scaled in some way to account for local re-calibration, there would still be substantial differences between two rainfall inputs to the hydrological model (Hughes, 2006a). Therefore, a methodology based on rainfall frequency curves has been developed that makes use of already existing spatially averaged rainfall time series (WR90 database) to correct satellite derived rainfall estimates in the later periods. The methodology is useful because it preserves the frequency characteristics of the historical (gauge-based) rainfall.

The application of the non-linear transformation technique based on frequency of exceedance generally improved the simulation results in most of the sub-basins used within the study (Table 2). However, in some sub-basins, even after the satellite data transformation, no improvement was observed. This is evident, for instance, in sub-basins G40J-K, T34A-H, S60C and Q94C, all of which are partly affected by frontal rainfall systems. This might be partly related to inadequacies in the RFE2.0 algorithm used to derive the CPCAPC data. The RFE2.0 algorithm was reported not to capture warm cloud rainfall especially along coastal regions, where warm cloud effects dominate (Love et al., 2004). On the other hand, significant under-estimations of spatial rainfall in orographic rainfall regions (for instance, V70A sub-basin) may be attributed to satellite-based estimates ignoring the rainfall variations due to altitude. Spatial rainfall variations are likely to be quite high in both mountainous and coastal regions. The systematic over- and under-simulation of peak flows for the 2001, 2004 and 2005 years (wet years) may be associated with CPCAPC data not able to accurately estimate high monthly rainfall values, a problem that is also frequently evident with ground- based rainfall observations. Therefore, none of the currently available rainfall sources can be said to be ideally suitable for input to the hydrological model as they are all associated with some form of uncertainties in their estimations (Hughes, 2006a).

In all of the simulation result comparisons it should be recognised that the observed data being used are far from perfect for the purpose. They contain inaccuracies in gauging (for both low and high flows) and none of the sub-basins included in the study are completely natural, such that differential development effects may occur between the two data periods, as well as within either period. While every attempt was made to select basins which are 'relatively' natural and not affected by developments, this is almost impossible in South Africa. The alternative of 'naturalising' the observed data might be considered the obvious approach, however, that process relies upon accurate information about the nature of water resource developments and their impacts on flow. Such information is not always straightforward to obtain and has the potential to introduce additional uncertainty in the modelling process.

As Beven and Binley (1992) noted, different parameter sets should be used with different precipitation inputs and parameter sets are not independent of the rainfall inputs (Görgens, 1983). However, the purpose of this study was to discover if an alternative source of rainfall data could be 'corrected' (or transformed) to make it consistent with the rainfall data used for calibration and to establish the model parameter set. The alternative of re-calibrating the model against the satellite data is not currently practical due to the currently short period of data availability (only 6 years). While the results are quite varied there is evidence to suggest that the CPCAPC satellite data sets can be of value for the extension of spatial rainfall data into the future; a time when gauged-based estimates are expected to be more difficult to obtain due to shrinking networks. The procedures

involved in obtaining and processing the satellite data are relatively straightforward and require little training to put into practice. They are therefore consistent with the requirements of a region such as Southern Africa, where complex methods often fail due to the lack of sufficient numbers of trained personnel.

The results of this study do, however, indicate some inconsistencies that need further investigation. Although few, some of the basins indicate that un-corrected satellite data generate better results than the data that are transformed using the RFC approach. Part of this problem may be related to the less than objective method that was used to determine the period within the WR90 rainfall data to use for the destination RFC. It may also be due to the fact that there are regions, or climate zones, where no correction is required. The number and geographic spread of the sub-basins used in the study is inadequate to resolve this issue at present. Unfortunately, the number of gauged sub-basins that can be used to assess the results of natural hydrological simulations is very limited. Extending the study would therefore almost certainly rely on generating 'naturalised' observed flow data. The uncertainties associated with this process have already been referred to.

The application of satellite-based rainfall estimates for supplying rainfall inputs where gauge measurements are not available appears to offer a potential solution to a problem that is widespread in developing regions such as southern Africa. While some further questions remain about the use of satellite data in some regions of the country, the satellite derived rainfall estimates do seem to have potential to contribute to extending model simulations and water resource estimations into the future.

Acknowledgments

References

ADLER RF, HUFMAN GH and KEEHN RP (1994) Global tropical rain estimates from microwave-adjusted, geosynchronous IR data. *Remote Sens. Rev.* **11** 125-152.

ANDRÉASSIAN V, PERRIN C, MICHEL C, USART-SANCHEZ-, I and LAVABRE-, J (2001) Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250** 206-223.

BARRERA DF, CEIRANO ED and ZUCARELLI GV (2007) Differences in area-averaged rainfall depth over a mid-size basin from two remote sensing methods of estimating precipitation. In: *Proc. Predictions in Ungauged Basins: PUB Kick-Off.* PUB Kick-off meeting held in Brasilia, 20-22 November 2002. IAHS Publ.309, IAHS Press, Wallingford, UK.

BEVEN KJ and BINLEY AM (1992) The future of distributed models-model calibration and uncertainty prediction. *Hydrol. Process.* **6** (3) 279-298.

GÖRGENS AHM (1983) Conceptual Modelling of Rainfall-Runoff Process in Semi-Arid Catchments. Report No. 1/83, Hydrological Research Unit, Rhodes University, Grahamstown, South Africa.

GOURLEY JJ and VIEUX BE (2006) A method for identifying sources of model uncertainty in rainfall-runoff simulations. *J. Hydrol.* **327** (1-2) 68-80.

GOVENDER M, CHETTY K and BULCOCK H (2007) A review of hyperspectral remote sensing and its application in vegetation and

water resource studies. *Water SA* **33** (2) 145-152.

GRIMES D and DIOP M (2003) Satellite-based rainfall estimation for river flow forecasting in Africa. Part 1. Rainfall estimates and hydrological forecasts. *Hydrol. Sci. J.* **48** (4) 567-584.

HSU K, GUPTA HV, GAO X and SOROOSHIAN S (1999) Estimation of physical variables from multi-channel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resour. Res.* **35** (5) 1605-1618.

HUFFMAN GJ, ADLER RE, ARKIN PA, CHANG A, FERRARO R, GRUBER A, JANOWIAK JJ, JOYCE RJ, McNAB A, RUDOLF B, SCHNEIDER U and XIE P (1997) The Global Precipitation Climatology Project (GPCP) combined precipitation data set. *Bull. Am. Met. Soc.* **78** 5-20.

HUGHES DA (2004) Incorporating ground water recharge and discharge functions into an existing monthly rainfall-runoff model. *Hydrol. Sci.* **49** (2) 297-311.

HUGHES DA (2006a) Comparison of satellite rainfall data with observations from gauging station networks. *J. Hydrol.* **256** 234-235.

HUGHES DA (2006b) An evaluation of potential use of satellite rainfall data for input to water resource estimation models in southern Africa. In: *Proc. 5ᵗʰ Int. FRIEND Conf. on Climate Variability and Change-Hydrological Impacts.* (Held in November, 2006 in Havana, Cuba.). *IAHS Publ. No.* **308** 75-80. IAHS Press, Wallingford, UK.

HUGHES DA and FORSYTH DA (2006) A generic database and spatial interface for the application of hydrological and water resource models. *Comput. Geosci.* **32** 1389-1402.

HUGHES DA and SMAKHTIN VY (1996) Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrol. Sci. J.* **41** (6) 851-871.

KIDD C (2001) Satellite rainfall climatology: a review. *Int. J. Climatol.* **21** (9) 1041-1066.

LABBERRY R, KNIVETON DR and TODD MC (2005) SAFARI 2000 Daily Rainfall Estimates, 0.1-Deg, Southern Africa, 1993-2001. Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.

LAWS KB, JANOWIAK JE and HUFFMAN GJ (2004) Verification of Rainfall Estimates over Africa using RFE, NASA MPA-RT and CMORPH Paper P2.2. *18ᵗʰ Conf. on Hydrology. Ame. Met. Soc.* 11-15 January 2004, held in Seattle, WA. 6 pp.

LOVE TB, KUMAR V, XIE P and THIAW W (2004) A 20 year Daily Africa Precipitation Climatology using satellite and gauge data. *14ᵗʰ Conf. on Appl. Clim. Ame. Met. Soc.* 11-15 January 2004, Seattle, WA.

MIDGLEY DC, PITMAN WV and MIDDLETON BJ (1994) Surface Water Resources of South Africa 1990, Volumes I to VI. WRC Reports No. 298/1.1/94 to 298/6.2/94. Water Research Commission, Pretoria, South Africa.

NASH JE and SUTCLIFFE JV (1970) River flow forecasting through conceptual models. Part I – a discussion of principles. *J. Hydrol.* **10** 282-290.

SANDHAM LA, SEED AW and VAN RENSBURG PAJ (1998) The relationship between half-hourly rainfall occurrence and daily rainfall. *Water SA* **24** (2) 101-106.

SCHÄFER NW (1991) Modelling the Areal Distribution of Daily Rainfall. Unpublished M.Sc. Eng. Thesis, Dept. of Agricultural Eng., University of Natal, Pietermaritzburg, South Africa.

SEED AW and AUSTIN GL (1990) Sampling errors for raingauge-derived mean areal daily and monthly rainfall. *J. Hydrol.* **118** 167-173.

SOROOSHIAN S, HSU KL, GAO X, GUPTA HV, IMAM B and BRAITHWAITE D (2000) Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Met. Soc.* **81** (9) 2035-2046.

THORNE V, COAKLEY P, GRIMES D and DUGDALE G (2001) Comparison of TAMSAT and CPC rainfall estimates with rainfall, for southern Africa. *Int. J. Remote Sens.* **22** (10) 1951-1974.

TODD MC, BARRET EC, BEAUMONT MJ and BELLERYBY TJ (1999) Estimation of daily rainfall over the upper Nile River Basin using a continuously calibrated satellite infrared technique. *J. Appl. Met.* **6** 201-210.

TODD MC, KIDD C, KNIVETON D and BELLERYBY TJ (2001) A combined satellite and passive microwave technique for estimation of small-scale rainfall. *J. Atmos. Ocean. Technol.* **18** 742-755.

WILK J, KNIVETON D, ANDERSSON L, LAYBERRY R, TODD MC, HUGHES DA, RINGROSE S and VANDERPOST C (2006) Estimating rainfall and water balance over the Okavango River Basin for hydrological applications. *J. Hydrol.* **331** (1-2) 18-29.

XIE P, YAROSH Y, LOVE T, JANOWIAK J and ARKIN P (2002) A real-time daily precipitation analysis over Africa as well as South Asia. *16ᵗʰ Conf. of Hydrol., Orlando, FL, Am. Met. Soc.* Held on 12-17 January 2002.

XU L, GAO X, SOROOSHIAN S, ARKIN P and IMAM B (1999) A microwave infrared threshold technique to improve the GOES precipitation index. *J. Appl. Met.* **38** 569-579.