Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Assessment of rainfall and evaporation input data uncertainties on simulated runoff in southern Africa

TENDAI SAWUNYAMA & DENIS HUGHES
Institute for Water Research, Rhodes University, 6140, South Africa
tendai@iwr.ru.ac.za

Abstract Rainfall–runoff models are used extensively in southern Africa for the purposes of water resource assessment. This study presents an assessment of the uncertainties associated with the spatial and temporal resolution of rainfall and evaporation inputs to a commonly applied hydrological model, as well as the extent to which these uncertainties are propagated into runoff estimations. While the effects of rainfall spatial variability are greater for daily rather than monthly estimates, the use of daily spatial data aggregated to monthly totals can reduce the uncertainties. The resulting runoff prediction uncertainties are greater in relative terms for semi-arid basins than for humid basins. Using time series of potential evapotranspiration instead of fixed monthly averages has an impact on simulated runoff in some parts of the country.
Key words model input data; rainfall–runoff model; uncertainty; Southern Africa

INTRODUCTION

Rainfall–runoff models are used extensively in southern Africa for the purposes of water resource assessment. It is well understood that the outputs from such models are subject to uncertainties related to the input hydrometeorological data, the ability of the model to simulate "real" hydrological response and the quantification of parameters. The problems of uncertainty associated with input data are exacerbated due to sparse observation networks, which have been shrinking over the last few decades. Görgens (1983) concluded that optimum parameter values are not independent of rainfall input and the variable degree to which it is representative. Recent studies of the spatial and temporal resolution of input rainfall data have demonstrated the advantages of improved accuracy of rainfall inputs in runoff modelling studies (Krajewski *et al*., 1991; Andréassian *et al*., 2001; Dong *et al*., 2005). While the implications for estimating basin average rainfall inputs to hydrological models are clear, there have been few attempts at quantifying the degree of uncertainty within the region.

It has been common practice in southern Africa to make use of mean monthly values of evapotranspiration data within models, largely due to a lack of appropriate time series information in many basins. The quantitative effects on simulated runoff patterns of ignoring the time series variations of evapotranspiration have not been addressed. However, Fowler (2002) concluded that substituting mean potential evaporation estimates into a soil water balance model produces very similar results to using actual potential evaporation, even under very dry or wet conditions.

This study presents an assessment of the uncertainties associated with differences in the spatial and temporal resolution of rainfall and evaporation inputs to a commonly applied hydrological model in the southern Africa region, as well as examining the extent to which these uncertainties are propagated into runoff estimation.

DATA AND METHODS

Figure 1 illustrates the location of the study basins that represent a wide range of climate and hydrological response zones within South Africa. The rainfall data from 28 raingauges (1988–1992) for the Bedford sub-basin (670 km^2) were collected as part of a semi-arid hydrological modelling research programme (Hughes & Sami, 1991), while all other data have been taken from national databases available from the Department of Water Affairs and Forestry (DWAF) or the South African Weather Service (SAWS). The Bedford data have a fixed spatial resolution (1 gauge per 24 km^2), while the number of gauges, and therefore the spatial resolution (generally much less than 1 gauge per 100 km^2), changes over time for the other sub-basins. The methods of analysis were therefore different for Bedford sub-basin and were based on generating mean areal rainfall inputs to 12 grid squares (5 × 5 minutes of a degree or approximately 56 km^2 in area) by semi-random sub-sampling from the available 28 raingauges.

For the other basins (Table 1) the analysis was based on sub-basins previously defined in the national surface water resources database (Midgley *et al.*, 1994) and the rainfall data available from the SAWS. The number of active raingauges has varied over time and generally reduced in recent years. The 20-year period with the maximum

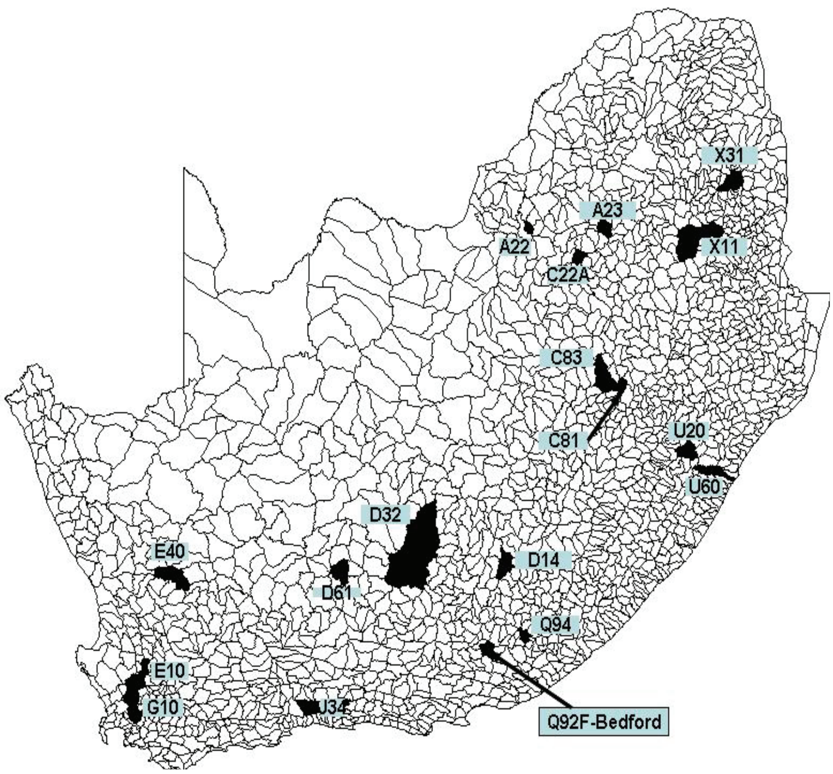

Fig. 1 South African Quaternary basins with those used in this study highlighted.

Table 1 Number of raingauges available (in brackets) for the different rainfall realizations for five basins.

Basin	Base period	Other realizations
Vaal, C83A, B & C ($250–830 km^2$), sub-humid	1930–1950 (14)	1940–1960 (10) 1950–1970 (9) 1960–1980 (9) 1970–1990 (7) 1980–2000 (11)
Limpopo, A23A ($682 km^2$), sub-humid	1930–1950 (13)	1940–1960 (10) 1950–1970 (8) 1960–1980 (7) 1970–1990 (6) 1980–2000 (13)
Orange, D32A to K ($572–1443 km^2$), arid/semi-arid	1930–1950 (41)	1940–1960 (31) 1950–1970 (29) 1960–1980 (24) 1970–1990 (15) 1980–2000 (8)
Sabie-X31A to G ($94–14 km^2$), humid	1950–1980 (25)	1930–1950 (20) 1970–2000 (23)
Berg-G10A to D ($126–687 km^2$), humid	1940–1960 (18)	1920–1940 (10) 1960–1980 (17) 1980–2000 (6)

number of active gauges was used as a reference and several realizations of spatially averaged daily and monthly rainfall were generated based on the gauges that were also active during other periods. An inverse distance (between the gauge location and the basin centroid) squared weighting procedure was used to generate the spatially averaged basin rainfalls. This method uses a maximum number of gauges (typically three or four in sub-basins of the size considered in this study) lying within a maximum search radius to generate the basin average rainfalls. As the basin gauge density decreases, it is clear that more distant gauges could be used, or that the maximum number of gauges will not be found within the maximum search radius. Schäfer (1991) found that different methods used to generate spatial rainfall data in South Africa did not produce substantially different results.

The various realizations of rainfall time series were used as input to a revised version of the Pitman rainfall–runoff model (Hughes, 2004) using regionalized model parameters based on Midgley *et al*. (1994). The analyses were based on comparing the "best" available estimates (using the period with highest number of raingauges) with alternative estimates using the smaller number of gauges that were active during other periods. Comparisons are made between the spatially-averaged rainfall data (both daily and monthly) as well as between the simulated runoff patterns using different realizations. While several statistical measures of comparison were used within the study, this paper focuses mainly on the relative measure, coefficient of efficiency (CE; Nash & Sutcliffe, 1970).

The evapotranspiration data analysis was based on fixed regionalized rainfall inputs (Midgley *et al*., 1994) to the model and a comparison between using mean monthly potential evapotranspiration values and monthly time series values (from

DWAF weather stations). All of the analyses were undertaken using the facilities available within the SPATISM (Spatial and Time Series Information Modelling) software package (Hughes & Forsyth, 2006).

RESULTS

Figure 2 illustrates the results for the 12 grid squares in the Bedford sub-basin. The three sets of results on each graph refer to comparisons (against estimates using all 28 gauges) based on daily spatial data, monthly spatial data generated from monthly gauge data, and monthly spatial data aggregated from daily spatial data. As might be expected, uncertainties in estimating daily data increase with a decrease in raingauge density more so than monthly data uncertainties. Within any single time interval, some grids have lower rainfalls with a decreasing number of gauges, others have higher rainfalls. The extent to which these balance each other out across the whole basin will clearly have an impact on simulated runoff. The application of the monthly rainfall–runoff model illustrated that simulated mean monthly runoff volumes can be up to 30% different when the model inputs are based on less rainfall information.

Table 1 shows the number of raingauges available for different periods (the base period having the most gauges) in the other basins. Spatially-averaged daily and monthly rainfall data were generated for the different periods using the number of gauges shown in the last column of Table 1. It follows that the gauges identified as being available during 1980 to 2000, for example, were also available during the base period. As might be expected, all basins demonstrate greater differences in spatial rainfall estimation (compared with estimates using the base period gauges) as the number of available gauges is reduced. The effects are greater in the areas with more spatially variable rainfall and greater for daily than for monthly rainfall estimates.

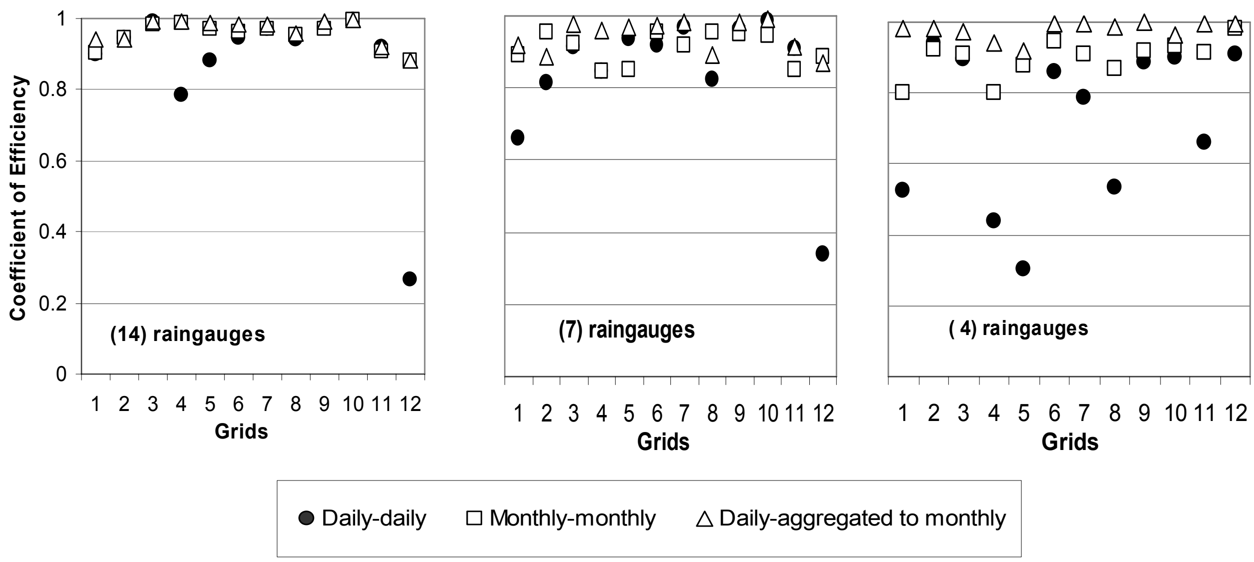

Fig. 2 Comparisons (with estimates based on all 28 gauges) of spatial average rainfall estimates using 14, 7 and 4 gauges.

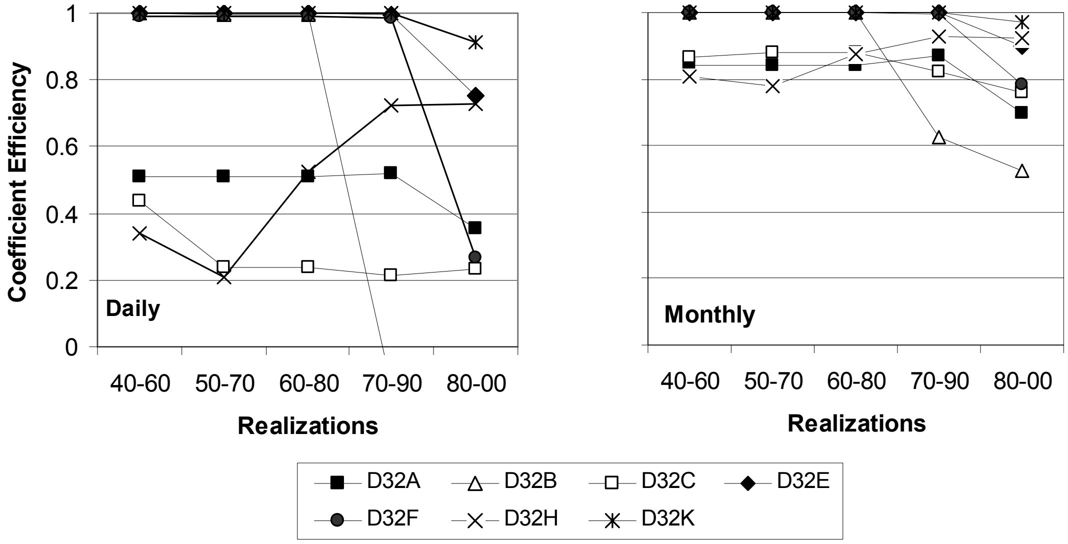

Fig. 3 Comparisons of daily and monthly rainfall realizations with the base period using the coefficient of efficiency for selected sub-basins of D32.

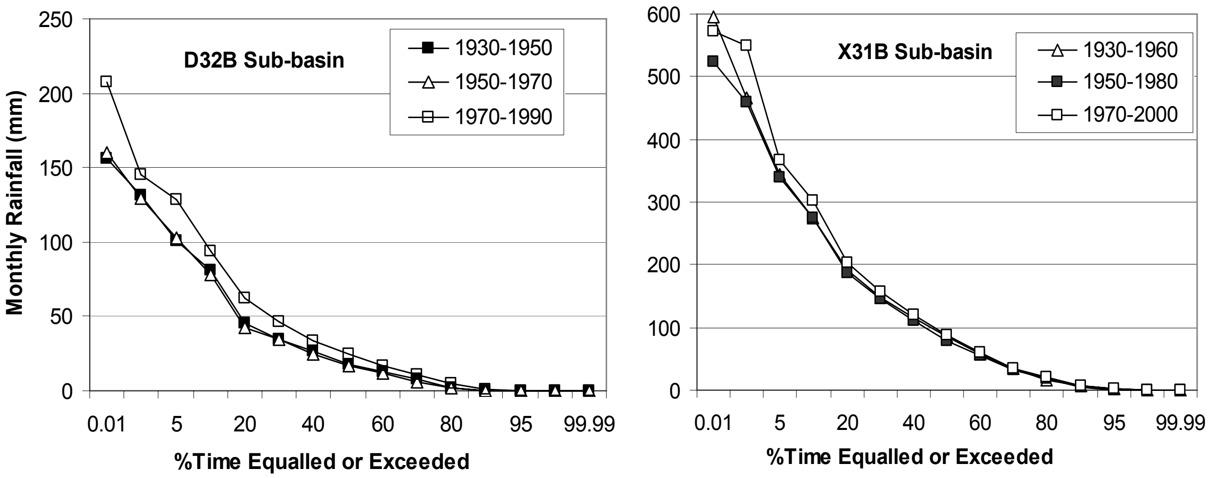

Fig. 4 Monthly rainfall exceedence frequency curves for three realizations over two sub-basins (the solid square symbol represents the base period realization).

However, Fig. 3 (7 sub-basins within D32) shows that there is not always a simple relationship between gauge numbers and uncertainty, particularly for daily rainfall. A reduction from 41 to fewer gauges has little impact in some parts of the basin, but a very large impact in others. The effects of using the gauges available for the period 1980 to 2000 are substantial for most sub-basins and impact on both daily and monthly spatial averages. It is worth noting that there has been a further reduction in the number of active gauges in most parts of South Africa since the year 2000.

The importance of estimating spatial rainfall over the full range of values is greater in humid regions than in semi-arid regions, where low rainfalls have little impact on runoff generation. Unfortunately, as Fig. 4 illustrates, the estimates of higher rainfall totals often suffer the most when the number of gauges is reduced. The consequences

Table 2 Effects of different monthly rainfall realizations on simulated runoff compared to the reference simulation (1930–1950) for two sub-basins using the coefficient of efficiency.

Sub-basin	1940–1960	1950–1970	1960–1980	1970–1990	1980–2000
D32B	0.95	0.95	0.95	–3.35	–10.85
D32F	0.98	0.98	0.98	0.96	–1.36

for runoff generation in sub-basin D32B are illustrated in Table 2, which provides a statistic of comparison (CE) between time series (1930 to 1950) of simulated monthly runoff generated using the base period gauges and the rainfall realizations using the gauges that were available during the other periods. Table 2 confirms the results based only on the comparisons of rainfall provided in Fig. 3. For sub-basin D32B, the results deteriorate when the total number of gauges for the basin reduces to 15 (1970–1990), while the results for D32F are still acceptable at that stage and only deteriorate when the number of gauges is reduced further.

Figure 5 compares the rainfall input uncertainties and the extent to which they are propagated in the rainfall–runoff model to output uncertainties, using the differences (expressed as percentages relative to the base period realization) in annual rainfall and simulated runoff totals. In the humid example (X31B, using the 1970 to 2000 realization), differences in annual rainfall fall within the range 0 to ±14%, while the corresponding runoff differences can be in excess of 100%. In the semi-arid region example (D32B, using the 1970 to 1990 realization) the range of annual rainfall differences is 0 to ±60%, while annual runoff differences can be as high as 750%. The rainfall errors tend to be higher in wet years and lower in dry years because large differences in monthly rainfall totals occur within high rainfall months (Fig. 4). Differences in runoff response caused by differences in rainfall estimates are expected to be less in humid catchments where a high proportion of the runoff is generated as smoothed baseflow output from basin storages.

Table 3 lists the means and standard deviations of monthly flow for simulations based on monthly mean pan evaporation estimates, as well as simulations using time series of pan evaporation values for seven basins of South Africa. The rainfall inputs to

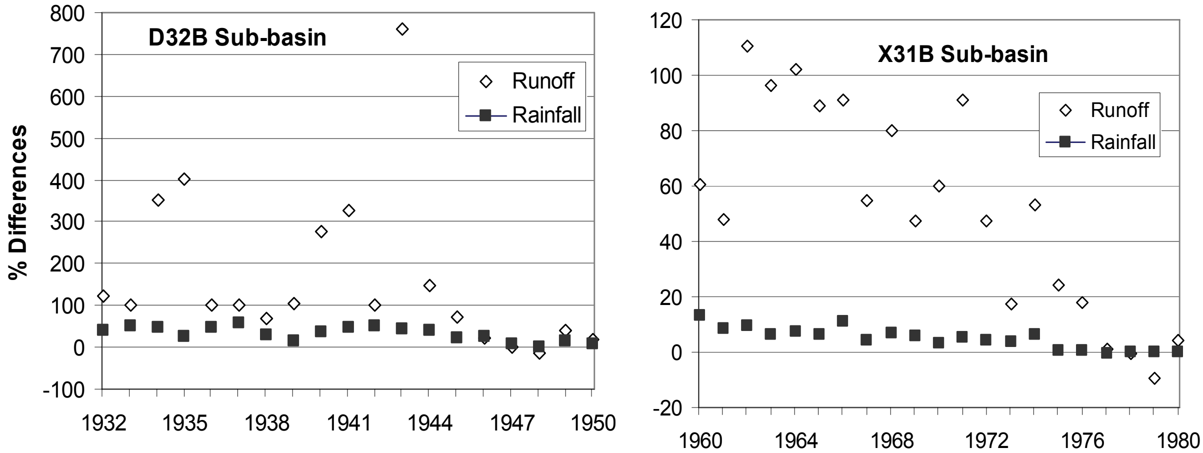

Fig. 5 Percentage differences (relative to the base period realization) in estimated annual rainfalls and annual runoffs for one realization (D32B: 1970–1990; X31B: 1970–2000) in two sub-basins.

Table 3 Mean and standard deviation (SD) of monthly flows for simulations using two different evaporation inputs (all values are in $m^3 \times 10^6$).

Sub-basin	Data	Model period	Mean	SD	CE
A22B (284 km^2) Semiarid (Limpopo region)	Mean values	1966–1990	0.60	2.21	
	Time series	"	1.06	3.40	0.54
	Diff.(%)		*76.7*	*53.8*	
C81G (434 km^2) Subhumid (Vaal region)	Mean values	1970–1990	2.43	4.45	
	Time series	"	3.40	6.38	0.74
	Diff.(%)		*39.9*	*43.4*	
D61E (1090 km^2) Arid (Orange region)	Mean values	1951–1990	0.25	1.66	
	Time series	"	0.28	1.97	0.94
	Diff.(%)		*12.0*	*18.7*	
E40A&B (1648 km^2) Arid (Olifants/Doom region)	Mean values	1951–1990	1.09	3.41	
	Time series	"	1.21	3.88	0.88
	Diff.(%)		*11.0*	*13.8*	
G10A (172 km^2) Humid (Berg region)	Mean values	1970–1990	14.84	16.31	
	Time series	"	14.84	16.25	1.00
	Diff.(%)		*0.0*	*–0.4*	
Q14A&B (1211 km^2) Arid (Fish region)	Mean values	1934–1990	1.32	4.87	
	Time series	"	1.39	5.86	0.94
	Diff.(%)		*5.3*	*20.3*	
X31H&J (214 km^2) Humid(Sabie/Komati region)	Mean values	1960–1990	3.24	3.32	
	Time series	"	3.86	3.96	0.90
	Diff.(%)		*19.1*	*19.3*	

Note: The diff.(%) values are the percentage differences between the two estimates relative to the estimates based on mean values. CE is the coefficient of efficiency statistic comparing the two simulations.

the model and the parameter values are the same in both sets of simulations and are based on regional data sets (Midgley *et al*., 1994). The results are highly variable with some examples suggesting large impacts on mean monthly runoff (A22B and C81G), while others suggest small impacts (G10A). Further investigations are required to explain some of the regional differences in more detail.

DISCUSSION AND CONCLUSIONS

The daily rainfall data had already been subject to basic error checking. However, some errors are difficult to identify, given the large spatial variations in rainfall and the low gauge densities in most of the example basins. It must also be recognised that the real spatial variation in rainfall is not known and that the realization with the largest number of gauges will not necessarily generate the most realistic spatial estimate. However, despite these limitations, the study has clearly demonstrated the uncertainties that can be associated with the preparation of rainfall data inputs to hydrological models (Faures *et al*., 1995). As Figs 2 and 3 demonstrate, the situation could be far worse for daily models than monthly models. The Bedford example (Fig. 2) suggests uncertainties can be reduced if daily data are used in the spatial averaging process and then aggregated to monthly values. However, both the D32B and X31B examples given in Fig. 5 suggest that a reduction in the number of gauges can lead to systematic over- or under-estimations. This might have been expected in X31B, which experiences topographic influences on rainfall. However, D32B is a relatively flat basin that is not expected to experience systematic variations in spatial rainfall patterns and the over-estimation based on the 1970–1990 realization is more difficult to explain.

Figure 4 illustrates that the largest differences in monthly rainfall totals occur within high rainfall months which have a greater impact on uncertainties in simulated runoff in both semi-arid and humid basins. These impacts will inevitably be greater for semi-arid basins, which are characterized by low runoff efficiencies and highly non-linear rainfall–runoff relationships (Shah *et al.*, 1996). The runoff estimate uncertainties for the Bedford example (using 12 grid squares within a 670 km^2 basin) are much lower (±30%) than for the D32B example (up to 750% using a single spatial rainfall estimate for a 572 km^2 basin). This suggests that some of the errors in runoff estimation can be cancelled out when they are aggregated and routed through a number of sub-basins, an advantage if the main objective is accurate simulations at the total basin outlet.

The parameter values of rainfall–runoff models are not independent of rainfall inputs and, ideally, the same set of gauges should be used over the entire record period to be simulated. However, with declining network densities this may not be possible. Further difficulties arise in those areas (dominated by convective rainfall processes) where spatial variations in rainfall are essentially random and correlations between gauge totals even at the monthly scale are typically low. The rainfall component of this study has highlighted the need for greater attention to be paid to the effects of uncertainty in rainfall inputs on simulated runoff. This issue could become even more important if conventional (gauges) and new sources (satellite data) of rainfall data are used together in hydrological models to make up for the lack of gauge data in the future (Hughes, 2006).

The use of time series of potential evapotranspiration resulted in consistently higher estimates of runoff compared to using fixed monthly means, despite the fact that the long-term mean evapotranspiration demands are the same. The lower (relative to the monthly mean) evapotranspiration values in the time series generally correspond to higher rainfall months which results in higher moisture levels and greater runoff. The fact that this effect is absent in the Western Cape G10A basin is related to its winter rainfall climate regime where low evapotranspiration rates prevail during the main rainfall season. The differences in the more arid basins are largely confined to those months when high rainfalls generate runoff due to exceedence of the main moisture storage in the model, which is typically small in arid regions with shallow soils.

REFERENCES

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. (2001) Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250**, 206–223.

Dong, X. H., Dohmen-Janssen, C. M. & Booij, M. J. (2005) Appropriate spatial sampling of rainfall for flow simulation. *Hydrol. Sci. J.* **50**(2), 279–298.

Faures, J., Goodrich, D. C., Woolhiser, D. A. & Sorooshian, S. (1995) Impact of small-scale spatial variability on runoff modeling. *J. Hydrol.* **173**, 309–326.

Fowler, A. (2002) Assessment of the validity of using mean potential evaporation in computations of the long-term soil water balance. *J. Hydrol.* **256**, 248–263.

Görgens, A. H. M. (1983) Conceptual modelling of rainfall–runoff process in semi-arid catchments. *Hydrological Research Unit Report no. 1/83.* Rhodes University, Grahamstown, South Africa.

Hughes, D. A. (2004) Incorporating groundwater recharge and discharge functions into an existing monthly rainfall–runoff model. *Hydrol. Sci. J.* **49**(2), 297–311.

Hughes, D. A. (2006) Comparison of satellite rainfall data with observations from gauging station networks. *J. Hydrol.* **256**, 234–235.

Hughes, D. A. & Forsyth, D. A. (2006) A generic database and spatial interface for the application of hydrological and water resource models. *Computers and Geosciences* **32**, 1389–1402.

Hughes, D. A. & Sami, K. (1991) The Bedford Catchments. an introduction to their physical and hydrological characteristics. *Water Research Commission Report no.138/1/92.*

Krajewski, F. W., Lakshmi, V., Georgakos, K. P. & Subhash, C. J. (1991) A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resour. Res.* **27**, 119–128.

Midgley, D. C., Pitman, W. V. & Middleton, B. J. (1994) Surface Water Resources of South Africa 1990, vols I to VI. *Water Research Commission Reports nos 298/1.1/94 to 298/6.2/94*. Pretoria, South Africa.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models. Part I, A discussion of principles. *J. Hydrol.* **10**, 282–290.

Schäfer, N. W. (1991) Modelling the areal distribution of daily rainfall. MScEng. Thesis, Dept of Agricultural Engng, University of Natal, Pietermaritzburg, South Africa.

Shah, S. M. S., O'Connell, P. E. & Hosking, J. R. M. (1996) Modeling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *J. Hydrol.* **175**, 89–111.