New Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009.

Rainfall variability and uncertainty in water resource assessments in South Africa

TENDAI SAWUNYAMA & DENIS HUGHES

Institute for Water Research, Rhodes University, Grahamstown, South Africa
tintosawuz@yahoo.com

Abstract Rainfall variability includes both spatial and temporal variability and is a potential source of uncertainty in rainfall–runoff modelling. The reliability of rainfall estimates depends on the accuracy of measurements, the number of raingauges and the spatial interpolation approach used to integrate point observations. However, degrading raingauge networks in South Africa represent a challenge to adequately account for this variability and extend rainfall records in practice. Therefore, there is need for correction procedures that address the uncertainty that exists in using such sparse data. The objective of this study is to demonstrate the importance of correcting original interpolated data sets to improve long-term estimates of spatial and temporal rainfall characteristics within South African catchments. The focus is on the generation of long time series of spatial rainfall over periods that span very different raingauge network densities. The study, through specific example sub-basins (e.g. improvements in simulation statistics such as coefficient of efficiency (CE) values of untransformed flows from 0.59 to 0.82 and 0.52 to 0.74 for 1959–1990 and 1991–2000 periods, respectively, for X31A sub-basin) demonstrated that a simple correction procedure based on rainfall frequency characteristics can be used to remove some uncertainties in spatial rainfall estimations and consequently model simulations.

Key words rainfall variability; uncertainty; rainfall–runoff model; South Africa

INTRODUCTION

Many developing countries suffer from inadequate hydrometeorological data and frequently rely on hydrological models to generate data that can be used for water resources planning. However, rainfall–runoff models are very sensitive to the rainfall inputs used (Sawunyama & Hughes, 2007, 2008) and while individual raingauges may provide reasonably accurate point data, the number of gauges is often insufficient to represent spatial variations in rainfall. This situation is exacerbated in most parts of South Africa by highly variable rainfall patterns related to orographic and convectional forcing (Lynch, 2004). Sawunyama & Hughes (2007) identified the problems and uncertainties associated with using raingauge networks to generate spatially averaged rainfalls when the network densities are inconsistent over time. Reliable estimations of sub-basin rainfall are obtained when the number of raingauges and their location can realistically represent the areal rainfall characteristics and the records are of sufficient length and stationary (Lynch, 2004). In South Africa, this is rarely achieved and there is a need to develop and employ correction procedures to derive more accurate long time series of rainfall estimates (both in space and time). The objective of this study is to reduce spatial rainfall uncertainties and generate long rainfall records using a simple approach based on rainfall frequency characteristics.

DATA AND METHODS

Historical monthly spatially averaged WR90 rainfall (1920–1990) and monthly potential evaporation data for selected sub-basins (quaternary catchments) in South Africa were obtained from the national water resources database (WR90) (Midgley *et al.*, 1994). These data were used to calibrate the modified version (including surface water–groundwater interactions) of the monthly Pitman model (Hughes, 2004) against all observed data available during this period. Rainfall stations used in the WR90 study were also used here, together with the Inverse Distance Weighting (IDW) approach to generate sub-basin spatial rainfall data. The observed flow 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 (Fig. 1). All of the analyses were undertaken using the facilities available within the

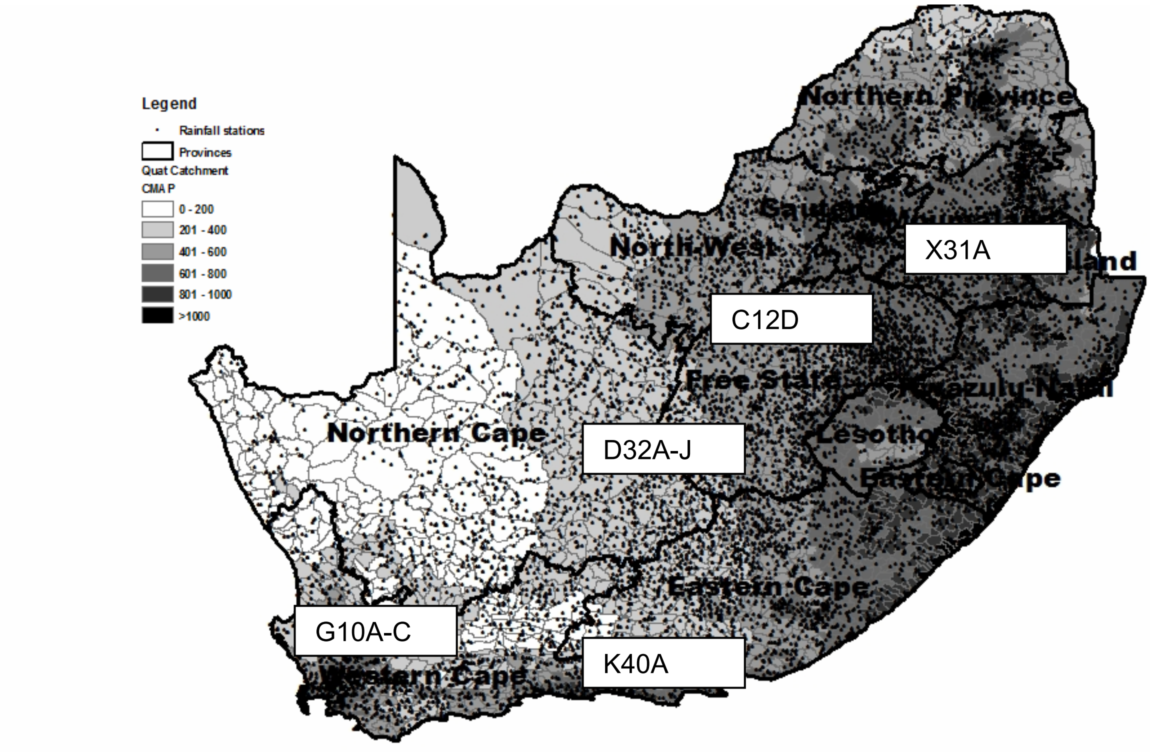

Fig. 1 Map of South Africa showing selected catchments, distribution of Catchment Mean Annual Precipitation (CMAP in mm/annum) and rainfall stations (small solid dots) for different regions across the country.

SPATSIM (Spatial and Time Series Information Modelling) software package (Hughes & Forsyth, 2006).

The assessment of trends in annual rainfall data series in this study involved visual interpretations of simple 5-year moving-averages and the application of a non-parametric Mann-Kendall statistical test (Hirsch *et al.*, 1982), while simple statistics (i.e. mean and variance) were used for monthly trends. The correction procedure (after Hughes & Smakhtin, 1996; see also Sawunyama & Hughes, 2008 for details of the procedure) was developed to adjust the frequency characteristics of the IDW spatially interpolated raingauge data using the frequency characteristics of a reference stationary time series data set (WR90, 1920–1990). The objective was to extend the records to September 2000, as this was the longest period spanned by the raingauge records that were available for this study. While the correction procedure is simple to apply when correcting rainfall characteristics of different spatial data sets covering the same period, the intention in this study was to correct one data set using another data set when the two are not coincident in time (i.e. use 1920–1990 data to correct IDW interpolated data for 1991–2000 where there are few available raingauges). The 1991–2000 period also includes raingauges that were operational over the whole period 1920–2000. Under such circumstances, it is essential to recognise that the two data sets may have “real” differences in rainfall characteristics as well as “false” differences related to the information content of the raw data.

The analysis was based on direct comparisons of three spatial rainfall realizations, as well as comparisons of the flow simulations that result from their use with the rainfall–runoff model. The *first realization* consists of the WR90 (1920–1990) regional rainfall data (Midgley *et al.*, 1994), which is the most widely available and longest data set used in water resources assessments in South Africa and known to be stationary. The *second realization* is based on the IDW interpolated data (1920–2000) using the closest three or four raingauges to the sub-basin centroid. While there are often adequate gauges within some parts of the period 1920–1990, the number of available gauges can be highly variable and there has been a substantial decline in network density since

1990. The *third realization* consists of the same rainfall data source as the second realization, but with the frequency characteristics corrected to be the same as the first realization and therefore expected to be stationary.

RESULTS

Rainfall analyses

The individual raingauge analysis based on annual rainfall totals (Table 1 and Fig. 2(a)) provides evidence that there are no ‘real’ non-stationarities due to natural climate variations and provides a basis for investigating the existence of “false” trends in the IDW spatially interpolated data. The analyses of spatially interpolated data (Table 2 and Fig. 2(b)), however, showed that false non-stationarity can be introduced (often through the closure of key gauges) and that this can be removed by employing a nonlinear correction procedure which only corrects the frequency characteristics of rainfall (Table 3 and Fig. 2(c)).The correction procedure involves transferring source rainfall (originally interpolated IDW spatial data) values to destination values (i.e. the corrected IDW time series) through use of similar percentage points (probabilities) from the respective rainfall frequency of exceedence curves (RFCs), which summarise the relationship between rainfall magnitude, and therefore variability within a time series (Hughes & Smakhtin,

Table 1 Mann-Kendall test summary statistics for the individual raingauge analysis.

Sub-basins raingauge no.	N	Kendall tau	Test z	p-value
G10A-C (0022038W)	96	0.061	0.880	0.379
K40A (0029294W)	76	–0.137	–1.749	0.128
D32A-J (0172163W)	123	–0.021	–0.350	0.726
C12D (0477772W)	88	–0.077	–1.060	0.298
X31A (0594539W)	79	–0.054	–0.707	0.480

N is the length of time series in years; **trend significant at α = 0.05 level; *trend significant at α = 0.1 level.

Table 2 Mann-Kendall test summary statistics of IDW spatially interpolated data.

Sub-basin	N	Kendall tau	Test z	p-Value
C12D	80	0.089	1.172	0.241
G10A	80	–0.419	–5.500	0.001**
G10B	80	–0.037	–0.490	0.624
K40A	80	–0.043	–0.565	0.572
X31A	80	0.020	0.258	0.798
D32B	80	–0.069	0.906	0.365
D32J	80	0.128	1.687	0.092*

N is the length of time series in years; ** trend significant at α = 0.05 level; *trend significant at α = 0.1 level.

Table 3 Mann-Kendall test summary statistics of transformed (corrected) data.

Sub-basin	N	Kendall tau	Test z	p-Value
C12D	80	0.070	0.922	0.365
G10A	80	–0.001	–0.017	0.987
G10B	80	0.040	0.523	0.600
K40A	80	–0.100	–1.313	0.189
X31A	80	–0.080	–1.055	0.291
D32B	80	0.036	0.474	0.636
D32J	80	0.049	0.640	0.522

N is the length of time series in years; ** trend significant at α=0.05 level; *trend significant at α = 0.1 level.

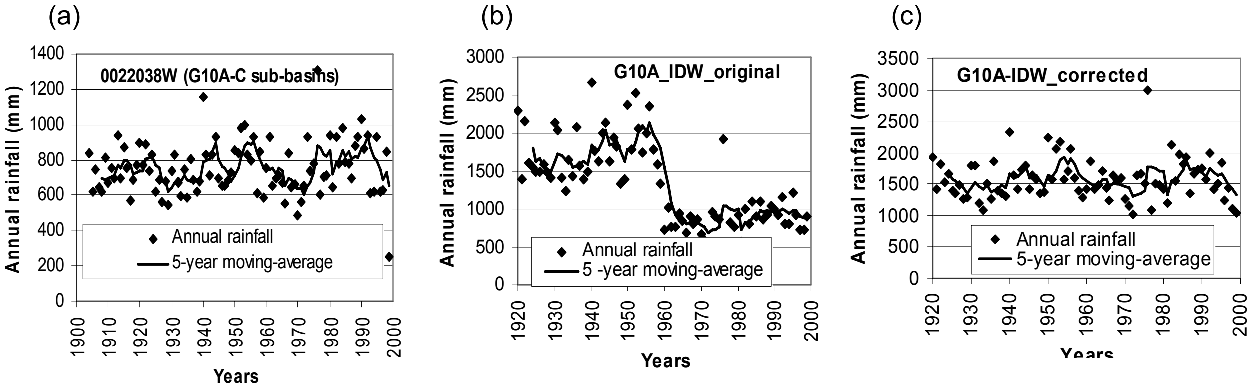

Fig. 2 An illustration of a 5-year moving-average of: (a) individual raingauge, (b) original spatially interpolated gauge data and (c) corrected spatial time series for G10A sub-basin.

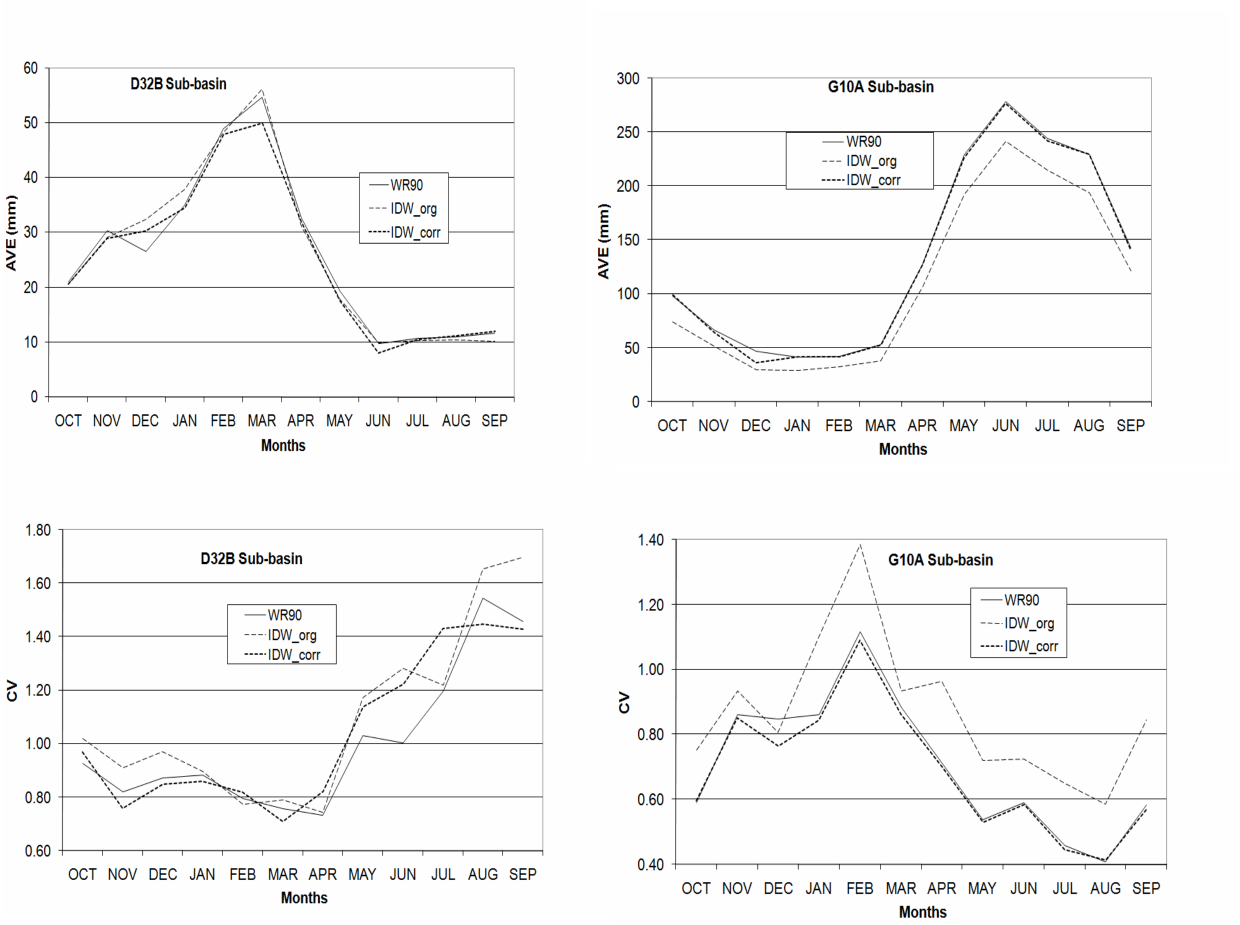

Fig. 3 Monthly rainfall characteristics for two sample sub-basins: D32B and G10A (AVE is the long-term average monthly rainfall and CV is the coefficient of variance). *IDW_org* represent originally interpolated time series and *IDW_corr* represent corrected time series.

1996; Sawunyama & Hughes, 2008). The destination RFCs are based on WR90 rainfall data, with the assumption that the destination RFC is representative of the frequency characteristics of real catchment rainfall.

The results for two example sub-basins using the monthly means and variance (Fig. 3) show that there can be major differences between the three spatial rainfall realizations. The WR90 data monthly characteristics are based on a 70-year period data set (1920–1990), while the original and corrected IDW are based on an 80 year period dataset (1920–2000). G10A shows that for more than 80% of the time, the monthly rainfall totals (both high and low values) are systematically under-estimated by about 40% when based on the original IDW interpolated data; this could have a major influence on model simulation results. While some results, show significant improvements after correcting the data (e.g. G10A), there are no improvements in others (e.g. D32B). One reason is that there are enough gauges distributed within the low topography of sub-basin D32B to adequately characterise rainfall variability. In contrast, the results for G10A reflect the high topographic variation where the loss of data from key gauges can result in spatial rainfall estimates that are not representative.

Model response to different spatial rainfall inputs

This section evaluates the use of different spatially interpolated rainfall data for the evaluation period 1920–2000 as input to the Pitman rainfall–runoff model for both the calibration period (all observed data up to 1990) and the extended period (1991–2000). A total of five rainfall realizations and model outputs were generated (Table 4). Four evaluation statistics were used to assess model performance: percentage difference of the means of monthly flows (%Diff Mn), percentage difference of standard deviations of monthly flows (%Diff stdv), coefficient of determination (R^2) and coefficient of efficiency (CE).

Many of the simulations based on the original IDW rainfall realizations are at least as good as the simulations based on calibration using the WR90 rainfall data. However, there are a number of cases where the original IDW rainfall data produced poor results that were improved by the use of the rainfall correction process. It is also evident that some of the calibration results do not generate "acceptable" simulations based on some of the statistics (note the large %Diff values when using log transformed flows for some sub-basins). This is partly related to unaccounted upstream development effects in the observed data, while a further contributing factor is expected to be that none of the available rainfall inputs are representative of the true sub-basin rainfall input. The similarity in results across all three simulations for the period up to 1990 for some basins could be related to the fact that all of the rainfall inputs are based on the same source sample of rainfall data and that these data are generally sufficient to provide representative inputs to the model.

For the extended period, 1991–2000 (Table 4), which is characterised by a reduction in raingauge numbers relative to 1920–1990, some of the simulation statistics based on the original IDW data are poor (e.g. D32A-J, G10A) and some satisfactory (e.g. X31A), while for the corrected input rainfall data the statistics improved for most of the sub-basins. However, in some instances a few gauges can give equally good results especially when the sub-basins are small or do not exhibit complex topography. The correction procedure was very useful in removing large systematic uncertainties in spatial rainfall estimates. An example is provided in Fig. 4 for the

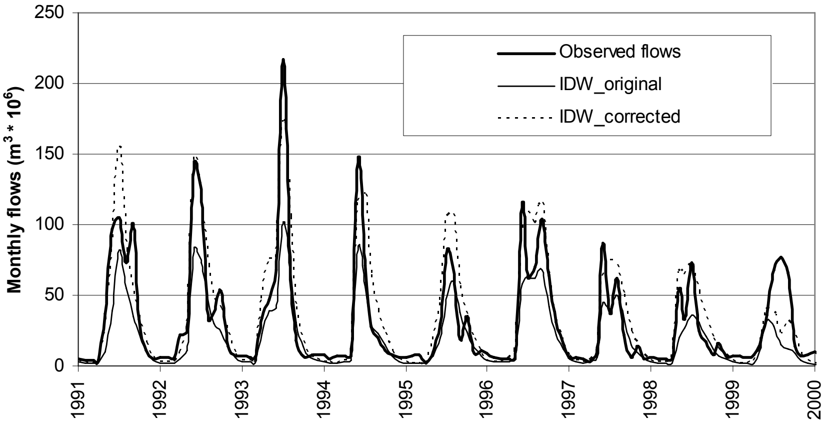

Fig. 4 Comparison of monthly flow time series from October 1990 to September 2000 for G10C.

Table 4 Comparison of simulated statistics (with reference to observed flows) based on five rainfall realizations.

Sub-basins	Data	Un –transformed flows (Q)				Log-transformed flows {ln(Q)}			
		%Diff Mn	%Diff Stdv	R^2	CE	%Diff Mn	%Diff Stdv	R^2	CE
C12D	**1965–1990**								
898 km^2	WR90	–29.3	–30.8	0.68	0.65	–25.9	–20.5	0.60	0.60
	IDW(org1)	10.2	–13.6	0.71	0.71	123.8	13.6	0.67	0.61
	IDW(corr1)	–23.2	–38.6	0.72	0.65	**30.1**	–20.9	0.64	**0.63**
	1991–2000								
	IDW(org2)	14.3	–15.9	0.58	0.57	58.5	24.1	0.68	0.23
	IDW(corr2)	–62.3	–49.8	**0.76**	0.56	–87.8	**–2.7**	0.68	0.02
D32A-J	**1980–1990**								
8330 km^2	WR90	14.6	6.8	0.91	0.89	–70.4	–33.3	0.57	0.48
	IDW(org1)	52.0	14.3	0.89	0.84	–90.5	–34.5	0.47	0.26
	IDW(corr1)	**–22.2**	–93.5	0.26	0.24	–64.7	–39.2	0.44	**0.38**
	1991–2000								
	IDW(org2)	439	207	0.51	–6.9	135	–48.9	0.22	–0.87
	IDW(corr2)	**19.0**	**–15.4**	**0.68**	**0.67**	**31.0**	**–25.7**	**0.42**	**0.35**
G10A-C	**1966–1990**								
609 km^2	WR90	20.4	6.7	0.90	0.86	15.9	–15.6	0.84	0.77
	IDW(org1)	–36.4	–41.6	0.80	0.62	–9.8	–18.7	0.80	0.77
	IDW(corr1)	**17.3**	**4.7**	**0.83**	0.80	14.0	**–15.0**	0.80	0.74
	1991–2000								
	IDW(org2)	–33.9	–38.9	0.81	0.65	–18.4	12.8	0.79	0.50
	IDW(corr2)	**11.0**	**5.1**	**0.80**	**0.77**	**–2.1**	22.9	0.79	**0.68**
K40A	**1961–1990**								
87 km^2	WR90	8.9	–21.2	0.65	0.65	–33.0	–12.8	0.65	0.60
	IDW(org1)	32.4	14.4	0.77	0.66	–35.1	8.6	0.66	0.47
	IDW(corr1)	**–4.4**	–30.1	0.76	**0.73**	**–12.8**	**4.3**	**0.67**	**0.62**
	1991–2000								
	IDW(org2)	47.2	50.6	0.71	0.13	–23.7	30.8	0.61	0.26
	IDW(corr2)	**6.6**	**–5.9**	0.69	**0.68**	**1.2**	**17.5**	**0.62**	**0.47**
X31A	**1959–1990**								
174 km^2	WR90	9.3	3.5	0.85	0.82	3.8	–1.1	0.85	0.82
	IDW(org1)	28.2	31.6	0.86	0.59	9.0	4.1	0.84	0.70
	IDW(corr1)	**8.4**	**–1.2**	0.84	**0.82**	**4.4**	–6.7	0.83	**0.80**
	1991–2000								
	IDW(org2)	39.7	23.8	0.81	0.52	17.1	–8.6	0.87	0.62
	IDW(corr2)	**10.7**	**–12.1**	0.75	**0.74**	**8.9**	15.7	0.83	**0.76**

Bold values indicate improvement in statistics after transformation; *IDW_org* represent originally interpolated time series and *IDW_corr* represent corrected time series. IDW(org1)/(corr1) represents data period up to 1990; IDW(org2)/(corr2) represents data period from 1991–2000.

topographically complex group of sub-basins G10A-C. The simulation statistics of the transformed (corrected) data are at least as good as the WR90 flows for most of the sub-basins (e.g. D32J, X31A and K40A) as shown in Table 4. While the overall observation is that the correctionprocedure applied here has demonstrated to be useful in improving spatial rainfall estimates, there are examples where it has not worked (e.g. C12D) and this is probably related to the limitations of the procedure when correcting non-overlapping data periods. The entire WR90 rainfall time series (1920–1990) could not be considered representative of the climate over the extended period 1991–2000. The selection of an appropriate period within the WR90 rainfall series that was used to derive the destination rainfall frequency curves was done by visually identifying a period within 1920–1990 that was climatically similar to the 1991–2000 periods using both the WR90 simulated flows and DWAF observed flows, assuming stationarity in the records. This process is subjective and may not be able to remove all the uncertainties in the final rainfall estimations.

DISCUSSION AND CONCLUSIONS

There is little doubt that in many developing countries declining raingauge networks make it increasingly difficult to define the primary input (rainfall) to water resources estimation models. Trend analysis of the original IDW spatially interpolated raingauge data showed that in some situations (e.g. for G10A sub-basin) different raingauge densities covering different time periods may introduce false trends in rainfall time series. However, when individual raingauges with long records were examined, no real trends were observed. When compared to WR90 data for a common period (1920–1990), there is evidence that information from the available raingauge data is more important than the spatial interpolation approach and this issue became critical when the extended period (1991–2000) was considered due to a further decline in the number of active raingauges. The uncertainty in spatial rainfall generation appears to be mainly related to how representative the observation stations are for a particular sub-basin. An overall observation from the analysis is that there are situations where the use of reduced density networks with less information may result in poor estimates of spatial rainfall.

The effects of rainfall uncertainty on simulated runoff were assessed by forcing a rainfall–runoff model with different rainfall realizations and comparing model outputs to observed flows. The results showed that correction of the original IDW interpolated spatial rainfall estimates resulted in significant improvements in model simulations (Table 4) for the extended period (1991–2000) in some of the example sub-basins (e.g. G10A), while in others there were marginal improvements (e.g. X31A) and the results were worse in others (C12D). For the calibration period (all available data up to 1990), most of the model results (Table 4) based on all rainfall realizations (including the corrected data) are not very different from each other, except for the sub-basins characterised by steep topography (G10A-C, K40A and X31A). The results suggest that simple correction procedures based on adjusting rainfall frequency characteristics can be used when information is lost through a reduction in raingauge network density. A similar approach has been used when new data products, such as radar and satellite, are integrated with ground-based raingauge data (Sawunyama & Hughes, 2008). The effects on model results of different rainfall inputs varied from region to region and some of the variations are dependent on the climate and basin physical properties. The implications of understanding the differences in rainfall variability and uncertainty are relevant to decision making. The knowledge of variability can guide decisions to re-open closed raingauges or to introduce more raingauges, while a knowledge of uncertainty can aid in determining areas where additional research (such as developing correction procedures), or alternative measurement techniques, or data products are needed to reduce the uncertainty.

REFERENCES

Hirsch, R. M., Slack, J. R. & Smith, R. A. (1982) Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* **18**, 107–121.

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

Hughes, D. A. & Smakhtin, V. Y. (1996) Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrol. Sci. J.* **41**(6), 851–871.

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

Lynch, S. D. (2004) *Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa*. Water Research Commission, Pretoria, South Africa.

Midgley, D. C., Pitman, W. V. & Middleton, B. J. (1994) *Surface Water Resources of South Africa 1990*, vol I to VI. Water Research Commission Reports, Pretoria, South Africa.

Sawunyama, T. & Hughes, D. A. (2008) Application of satellite-derived rainfall estimates to extend water resource simulation modelling in South Africa. *Water SA* **34**(1), 1–9.

Sawunyama, T. & Hughes, D. A. (2007) Assessment of rainfall–runoff model input uncertainties on simulated runoff in southern Africa. In: *Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resource Management* (ed. by E. Boegh *et al.*) (Proc. Symp. HS2004 at IUGG2007, Perugia, July 2007), 98–106. IAHS Publ. 313. IAHS Press, Wallingford, UK.