Climate Variability and Change—Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006), IAHS Publ. 308, 2006.

An evaluation of the potential use of satellite rainfall data for input to water resource estimation models in southern Africa

DENIS A. HUGHES
Institute for Water Research, Rhodes University, Grahamstown 6140, South Africa
denis@iwr.ru.ac.za

Abstract Global or near global satellite rainfall data are becoming readily available and have the potential to provide the rainfall data inputs required for water resource availability modelling in the face of shrinking ground-based networks. However, the difficulties of using both sets of rainfall data in models with existing calibrations based on raingauge data are not well understood. An assessment using three basins in southern Africa yielded mixed results. There is clearly a need for some "corrections" to be applied to the satellite data to achieve consistency with the gauged rainfall information, but it is not usually clear what form these "corrections" should take. It is important to note that in some areas there are very few gauges currently available and the satellite data represent a viable substitute.

Key words satellite data; rainfall data; hydrological models; developing countries; water resources; southern Africa

INTRODUCTION

The quality of rainfall data input to hydrological models is of primary importance (Brath *et al.*, 2004). Without adequate rainfall data, other issues such as model choice, number of parameters and parameter estimation procedures become somewhat superfluous. However, the ground-based recording networks of many countries are shrinking and in parts of southern Africa, notably those that have experienced recent political and social upheaval, gauging networks have either almost ceased to exist or have large spatial and temporal discontinuities. These are also the regions where the availability of water resource information is important for managing the future sustainable use of water resources for social and economic development. Issues such as which spatial interpolation approach to use (Tabios & Sala, 1985) become irrelevant as there are too few gauges to represent the spatial variations of rainfall.

Where observed river flow data are not available, hydrological models offer a viable alternative for generating water resource availability information. However, their successful application relies on, *inter alia*, an accurate representation of basin precipitation inputs. Given the lack of ground-based observations in many areas of southern Africa, alternative sources need to be identified and assessed. Near-global datasets of a wide range of terrestrial information derived from satellite imagery are becoming increasingly accessible. While their use in a southern Africa context is not new (Thorne *et al.*, 2001; Grimes & Diop, 2002), they have not been applied extensively and there have been few assessments of using satellite data in conjunction with gauge data. Hughes (2006) reports on some comparisons of rainfall information available from near-global datasets (GPCP; Huffman *et al.*, 1997, 2001 and PERSIANN; Hsu *et al.*, 1999; Sorooshian *et al.*, 2000) with ground-based observations. This earlier paper was limited to comparisons based on four study areas representing different climate regimes. The results were encouraging enough to justify further investigations, although indications are that the satellite data cannot be used without modification and further processing. The current objective is to assess whether the satellite data can be used in conjunction with gauge data as inputs to a hydrological model, based on the premise that existing model calibrations using available historical gauge data already exist and that satellite data records are still relatively short and unable (on their own) to adequately represent long-term variability. The main questions to be answered by this phase of the investigation are:

(a) Is it possible to identify relatively simple scaling factors that can be applied to the satellite data so that they can be successfully used in a rainfall–runoff model calibrated using historical rainfall data?

(b) Can comparisons with gauged rainfall data or the model results be used to suggest more complex transformations of the satellite rainfall data which are still straightforward to apply?

THE SATELLITE PRECIPITATION PRODUCTS

The full details of the two satellite precipitation products can be found in Hughes (2006), as well as the original sources. They have been selected primarily on the basis of their ease of access, in recognition of the limited resources normally available in developing regions. The GPCP (Global Precipitation Climatology Project) 1DD estimates have a time resolution of one day and a spatial resolution of one degree. The period of record is currently October 1996 to January 2005. The PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) datasets have a temporal resolution of six hours, a spatial resolution of 0.25^o and start in March 2000.

THE MODEL AND TEST BASINS

The rainfall–runoff model used is the Pitman monthly model with additional surface–groundwater interaction routines as discussed in Hughes (2004). Three basins are included: the Kafue basin (156 995 km^2) in Zambia, the Thukela River basin (29 046 km^2) in Kwa-Zulu Natal and the Kat River basin (1715 km^2) in the Eastern Cape province of South Africa. Further details (including maps indicating the location of raingauges and the one degree grid squares) of these basins can be found in Hughes (2006). The model tests are based on smaller sub-basins due to the availability of suitable observed flow data. The comparisons between observed and simulated data are based on visual assessments of time series graphs, as well as six standard goodness-of-fit statistics; percentage deviation of the simulated mean from the equivalent observed values (% Error MMQ and % Error MMlnQ), R^2 (Q and lnQ) and coefficient of efficiency: CE(Q) and CE(lnQ) from Nash & Sutcliffe (1970), based on both untransformed (Q) and natural logarithmic transformed (lnQ) monthly flows.

RESULTS

Kafue River, Zambia

Two streamflow gauging stations (4050, Kafue at Raglan Farm, 5000 km^2; 4560, Lunga at Chifumpa, 21 445 km^2) on the Kafue River (Mwelwa, 2005) have been selected. The historical rainfall data consist of some 12 gauges located in the northern part of the Kafue basin, all having varying length records and months of missing data. An inverse distance squared weighting procedure has been used to determine basin average rainfall, but after 1990 the average rainfall is largely based on a single gauge. Table 1 lists the goodness-of-fit statistics for the calibration period (October 1961 to September 1990), as well as for the extended period (October 1990 to September 2000). It is evident that the simulations for the extended period are very poor and this is largely due to very high rainfalls and over-simulations in 1998 and 1999. The calibration for sub-basin 4560 is also not very good and it was found to be extremely difficult to reproduce the very rapid seasonal recession curve.

The GPCP rainfall input to sub-basin 4050 uses a single 1^o grid. The larger basin (4560) was further sub-divided into three sub-basins during the calibration process and their rainfall inputs based on single 1^o grids. Table 1 indicates that the un-scaled GPCP data is very low, but that even after scaling to ensure that the observed and simulated mean monthly flows are similar, the fit is not very good. The poor fit is also largely due to years 1998 and 1999, but in this case caused by under-simulations.

The PERSIANN rainfall inputs to 4050 are based on averages of seven 0.25^o grids, while the three sub-basins of 4560 used averages of between 3 and 15 grids. As indicated in Hughes (2004), and in contrast to the GPCP data, the PERSIANN rainfalls are much higher than the historical data and led to large over-simulations. When scaled to reproduce the observed mean monthly flow, the results for 4050 were encouraging (Table 1), although based on only 31 months of data. The results for 4560 were not very good and could be a reflection of an inadequate model calibration (a problem of over-simulating the late wet season recession flows). However, for both examples (4050 and 4560), the results using the PERSIANN data are better than when using the available

Table 1 Simulation results for the Kafue River (4050 and 4650 sub-basins).

Statistic	Calibration	Historical rainfall data	GPCP No scaling	Scaling	Persiann No scaling	Scaling
Record period	1961–1990	1990–2000	1996–2002		2000–2004	
Sub-basin 5040, Kafue River at Raglan Farm						
% error MMQ	2.0	78.5	–59.3	–0.2	47.4	–0.4
% error MMlnQ	0.0	16.8	–26.5	2.8	10.7	3.8
R^2(Q)	0.804	0.624	0.368	0.427	0.664	0.811
CE(Q)	0.803	–0.648	0.174	0.213	0.249	0.797
R^2(lnQ)	0.860	0.816	0.660	0.707	0.718	0.862
CE(lnQ)	0.37	0.664	0.335	0.694	0.617	0.850
Scaling factor	N/A	N/A	1.0	1.307	1.0	0.830
Sub-basin 4560, Lunga River at Chifumpa						
% error MMQ	0.4	22.0	–63.3	–1.8	126.0	1.0
% error MMlnQ	0.7	1.0	–32.9	–11.5	6.5	–1.6
R^2(Q)	0.498	0.668	0.460	0.534	0.779	0.657
CE(Q)	0.454	0.246	–0.083	–1.312	–7.978	0.635
R^2(lnQ)	0.614	0.654	0.465	0.462	0.595	0.524
CE(lnQ)	0.545	0.383	–8.728	–2.070	–1.080	–0.024
Scaling factor	N/A	N/A	1.0	1.188	1.0	0.840

Notes: Applicable to Tables 1 to 3.
The period is given as hydrological years starting in October. See text for the key to the statistics

recent raingauge data. While simple linear scaling of the satellite data is unlikely to be the best approach, there is insufficient evidence within the available data, or in the model results, to suggest a straightforward alternative.

Thukela River basin, South Africa

The assessments within the Thukela River basin are based on a gauged tributary (Mooi River at V2H004, 1546 km^2 in area). There are two active rainfall measuring stations with reasonably complete recent records, one in the headwaters and one near the gauge. The basin is covered by three 0.25° Persiann grids and falls across the boundary of two 1° GPCP grids and the total area has been divided into five sub-basins for modelling purposes. The rainfall–runoff model was calibrated satisfactorily (Table 2) using existing spatially averaged rainfall data (Midgley *et al.*, 1994). Four additional model runs were made using the calibration parameters. The first used the nearest point gauge rainfall as input, the second used weighted averages of the two gauges, while the third and fourth used appropriate grid data from the GPCP and PERSIANN data, respectively, based on the locations of the sub-basins. The results of the simulations are provided in Table 2 and some are illustrated in Fig. 1.

Table 2 Simulation results for the Thukela basin (Mooi River gauging station).

Statistic	Calibration (WR90 rainfall)	Gauges 2 gauges	Average	GPCP	Persiann
Record period	1950–1990	1996–2004			2000–2004
% error MMQ	1.6	12.7	–5.6	–4.0	44.1
% error MMlnQ	8.4	8.7	1.3	16.6	28.9
R^2(Q)	0.737	0.519	0.472	0.640	0.287
CE(Q)	0.736	0.076	0.196	0.621	–0.408
R^2(lnQ)	0.785	0.723	0.721	0.832	0.455
CE(lnQ)	0.754	0.700	0.720	0.667	0.237

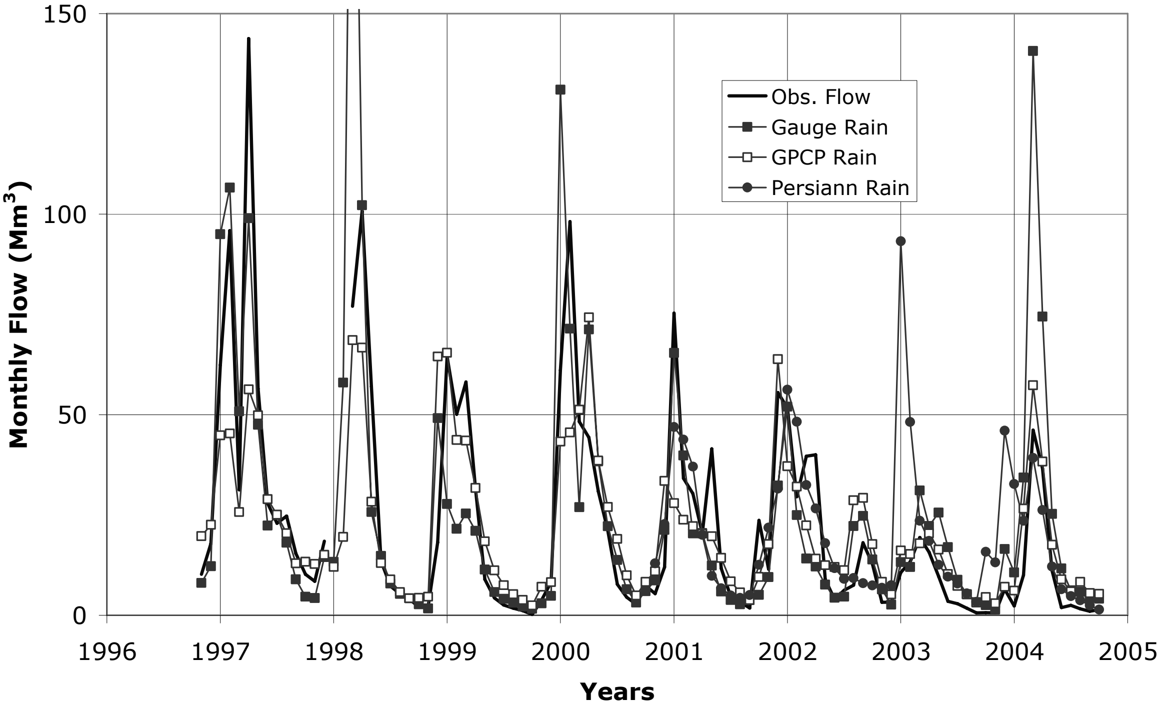

Fig. 1 Time series comparisons of observed and simulated monthly flows for the Mooi River, Thukela basin, South Africa (horizontal axis marks at December of the indicated years).

The main problem with the ground based rainfall inputs is that extreme monthly values generate excessive peak flows in some years (Fig. 1, early 1999, 2001 and 2005). A similar problem occurs with the PERSIANN data (December 2003 and 2005) as already noted by Hughes (2006). The GPCP data provide the most satisfactory rainfall inputs to the previously calibrated model. There are no indications that simple adjustments, such as linear scaling, will improve either of the satellite data sets.

Kat River basin, South Africa

Within the Kat River basin there is only one flow gauging station that has recent data and that is relatively unaffected by anthropogenic modifications to the flow regime. The sub-basin has an area of 79.6 km^2, which is less than 20% of a 0.25° grid square, experiences substantial topographic influences on precipitation and has only one active rainfall gauge (100 329) within its boundary. A further rainfall gauge (Q9E001) exists some 10 km to the east in a lower rainfall region. The rainfall–runoff model was calibrated using existing spatially averaged historical rainfall data (Midgley *et al.*, 1994).

The calibration parameters were used with the un-scaled rainfall available from the satellite sources, as well as a relatively arbitrary combination of the two available gauge records (0.6*100329 + 0.4*Q9E001). All the rainfall records were then linearly scaled to reduce the % errors in simulated mean monthly runoff. The results are presented in Table 3 and Fig. 2 where it is apparent that none of the currently available rainfall sources are particularly suitable for input to the model and that most of the errors are associated with both over- and under-estimations of the rainfall signals during individual months. However, all of the scaled rainfall inputs generate flow duration curve characteristics that are close to the observed data, with the GPCP data being slightly better. There are no systematic patterns in the results that can be used as a basis for more complex transformations of the original rainfall data sources. There appears to be insufficient information contained within the original sources to adequately represent the spatio-temporal rainfall inputs to the sub-basin.

Table 3 Simulation results for the Kat River sub-basin.

Statistic	Calibration	GPCP No scaling	Scaling	Persiann No scaling	Scaling	Gauges No scaling	Scaling
Record period	1920–1989	1996–2003		2000–2003		1990–1999	
% error MMQ	–13.7	–54.4	–0.6	–65.6	–0.4	–16.1	–0.5
% error MMlnQ	0.2	–116.3	1.06	–232.5	–29.2	–29.1	13.2
R^2(Q)	0.807	0.691	0.674	0.721	0.659	0.551	0.566
CE(Q)	0.801	0.384	0.662	0.312	0.592	0.502	0.465
R^2(lnQ)	0.651	0.367	0.469	0.335	0.346	0.582	0.590
CE(lnQ)	0.647	–0.394	0.395	–2.120	–0.009	0.423	0.490
Scaling factor	N/A	1.0	1.259	1.0	1.378	1.0	1.051

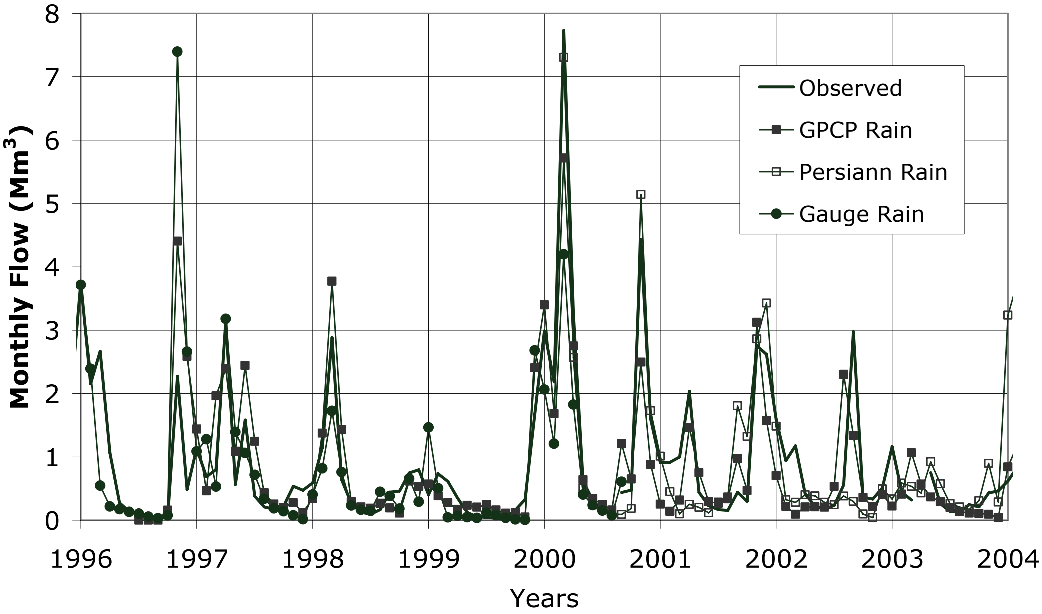

Fig. 2 Time series comparisons of the simulation results for the sub-basin of the Kat River basin based on scaled rainfall (horizontal axis marks at December of the indicated years).

DISCUSSION AND CONCLUSIONS

The model calibrations will reflect any inadequacies in the ground-based rainfall gauge data, including a lack of spatial representation, and it should be recognized that no data source is likely to provide “true” rainfall inputs. The calibrated parameters are therefore not independent of the rainfall inputs to the model. When using two different sources of rainfall data, as in this study, a choice has to be made between adjusting the rainfall data to be consistent, or using different parameter sets. For a model where the parameters are expected to reflect physical basin properties, the former option is the logical choice. As the overlap between the observed flow and the gauge rainfall data is typically longer in the southern Africa situation, it also seems logical to adjust the satellite data to be consistent with the gauge data. Further justification for this approach is that while gauge data are commonly not adequate to completely quantify spatial variations due to topographic influences, the satellite data appear to ignore these completely.

The three study areas represent very different climate regimes within southern Africa and some of the results are encouraging enough to warrant the use of satellite data. The corrections applied to the satellite data are very simple, but there are no clear indications that more complex, non-linear or seasonal corrections will be more successful. One issue that has been highlighted is that the currently available ground-based rainfall data are frequently inadequate. This introduces a great deal of uncertainty in the results of extending water resource simulation models, based on historical rainfall data, into the present and future.

Acknowledgements David Bolvin of the Mesoscale Atmospheric Processes Branch, NASA and Dan Braithwaite of the Center for Hydrometeorology and Remote Sensing, University of California, Irvine provided help with the GPCP and PERSIANN data, respectively. Elenestina Mwelwa provided the Zambian data and the South African data are derived from the databases of the South African Weather Service and the Department of Water Affairs and Forestry.

REFERENCES

Brath, A., Montanari, A. & Toth, E. (2004) Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *J. Hydrol.* **291,** 232–253.

Grimes, D. I. F. & 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.

Huffman, G. J., Adler, R. F., Arkin, P. A., Chang, A., Ferraro. R., Gruber, A. Janowiak, J. J., Joyce, R. J., McNab, A., Rudolf, B., Schneider, U. & Xie, P. (1997) The Global Precipitation Climatology Project (GPCP) combined precipitation data set. *Bull. Am. Met. Soc*. **78**, 5–20.

Huffamn, G. J., Adler, R. F., Morrissey, M. M., Curtis, S., Joyce, R. J., McGavock, B. & Susskind, J. (2001) Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydromet.* **2**, 36–50.

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. (2006) Comparison of satellite rainfall data with observations from gauging station networks. *J. Hydrol*. **327**, 399–410.

Hsu, K., Gupta, H. V., Gao, X. & 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.

Midgley, D. C., Pitman, W. V. & Middleton, B. J. (1994) Surface Water Resources of South Africa 1990, volumes I to VI. *Water Research Commission Reports nos 298/1.1/94 to 298/6.1/94*. Pretoria, South Africa.

Mwelwa, E. M. (2005) The application of the monthly time step Pitman rainfall-runoff model to the Kafue River basin of Zambia. MSc Thesis, Rhodes University, Grahamstown, 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.

Sorooshian, S., Hsu, K., Gao, X., Gupta, H. V., Imam, B. & Braithwaite, D. (2000) Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Met. Soc*. **81**, 2035–2046.

Tabios, G. Q. & Salas, J. D. (1985) A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resour. Res.* **21**, 365–380.

Thorne, V., Coakley, P., Grimes, D. & Dugdale, G. (2001) Comparison of TAMSAT and CPC rainfall estimates with rainfall, for southern Africa. *Int. J. Remote Sens.* **22**(10), 1951–1974.