Impacts of rainfall uncertainty on water resource planning models in the Upper Limpopo basin, Botswana

P. K. KENABATHO, N. R. McINTYRE & H. S. WHEATER

Department of Civil and Environmental Engineering, Imperial College London, South Kensington Campus, London SW7 2BU, UK

kp.kenabatho06@imperial.ac.uk

Abstract In water resource planning in semi-arid Africa and comparable regions, uncertainty is high due to limitations in historic observations, uncertainty in hydrological models, uncertainty over future demands for water, and uncertain influences of future climate and hydrological change. The uncertainty in the future supply–demand balance should be considered in planning decisions, as it affects the risk associated with any planning option, and can help identify priorities for data collection. Focusing on rainfall and hydrological uncertainty, this paper outlines a framework of uncertainty analysis, which allows such consideration to be given. The framework consists of multi-site continuous time stochastic rainfall modelling to infill historic rainfall data. The stochastic infilling of rainfall data allows calibration of a hydrological model under input uncertainty. The rainfall model, together with the uncertain hydrological model, is then used to generate multiple realisations of reservoir inflow over a 100-year period. This framework is applied to the Upper Limpopo basin in Botswana, using 25 years of observed daily rainfall and flow data for model calibration. A generalised linear model was used for the rainfall and a semi-distributed version of the IHACRES model was used for the hydrology. A proposed 382×10^6 m^3 reservoir at the outlet of this catchment, which is part of Botswana's national water resource strategy, is re-evaluated in light of the extended inflow data and the estimated uncertainty. Results show that the uncertainty has a considerable effect on the reliability of the reservoir; for example, the proportion of time for which demand for water was not met ranged from 0 to 13% over the different flow realisations. The main assumptions made, to be addressed in our future research, are stationarity of climate and that all the hydrological uncertainty arises from the historic rainfall uncertainty due to missing data.

Key words IHACRES; rainfall–runoff; generalised linear models; semi-arid; Botswana; reservoir operation

INTRODUCTION

Hydrological models are important tools for use in water resources planning and management. Their practical use includes, for example, flood and drought prediction, water resources assessment for reservoir design and operation, as well as climate change and land-use impacts assessment. Because of limited availability of the spatial data required to satisfy the demands of fully distributed physically-based hydrological models, many studies in the past have focused more on conceptual lumped models. This is particularly true in data-sparse semi-arid areas with lack of complete rainfall data of good quality and lack of good raingauge networks to capture the spatial variability resulting from highly localised rainfall. However, stochastic spatial-temporal rainfall models are increasingly becoming important tools for characterising rainfall in space and time which can be used to drive distributed hydrological models (Segond *et al.*, 2006). Of particular importance to data-sparse areas, some of these models have the potential to quantify uncertainty due to missing values in the observed records (Yang *et al.*, 2005; Kenabatho *et al.*, 2008). They can also be useful in generating long sequences of rainfall data needed to support planning of long-term water management strategies, as well as generating future rainfall fields by downscaling GCM outputs under scenarios of climate change (Charles *et al.*, 2007; Burton *et al.*, 2008). However, the use of stochastically-generated rainfall fields to drive rainfall–runoff modelling and, hence, to support water resources planning is still limited, especially in semi-arid regions. Instead, criteria such as worst-drought-on-record or design droughts from simple frequency analysis are commonly used for design purposes. The aim of this paper is to test the performance of a rainfall–runoff model when driven by stochastically generated spatial–temporal rainfall, and to begin to assess the implications for reservoir design and operations within Botswana's national water resource strategy. This is achieved by using a semi-distributed IHACRES model to generate reservoir inflows, driven by rainfall obtained from a generalised linear model (GLM) in a 7660 km^2 semi-arid area in the upper Limpopo basin, Botswana.

The IHACRES model

The IHACRES model, initially developed by Jakeman *et al*. (1990) for application in temperate catchments, has undergone several modifications to suit various applications. For example, it has been modified for use in ephemeral catchments (Ye *et al*., 1997), and for application in semi-distributed modelling (Croke *et al*., 2006). The historical background of the model is well captured in the literature (Post & 1999, and references therein). The model has two modules, one nonlinear and the other linear. The nonlinear module converts rainfall (r) into effective rainfall (u), while the linear module converts effective rainfall into streamflow (q). This flow can be routed through any configuration of storages, in parallel or series (e.g. quick and slow flow) depending on the underlying hydrological processes of the area under investigation. The effective rainfall at each time step (k) is usually assumed to be proportional to catchment wetness index (w) as follows:

$$u_k = w_k r_k \tag{1}$$

The evolution of w with time can be represented as a proportion (α) of the rainfall in that time step (i.e. the increase), and the loss is a proportion ($1/\beta$) of the preceding estimate of w:

$$w_k = \alpha r_k + w_{k-1}\left(1 - \frac{1}{\beta}\right) \tag{2}$$

where β is estimated as an empirical function of air temperature T:

$$\beta = \beta' e^{\gamma(T - T')} \tag{3}$$

where β' is a reference value of β at air temperature $T = T'$. For application to low yielding ephemeral rivers, the model was modified by introducing a threshold λ, and an exponent parameter ρ (Ye *et al*., 1995; Schreider *et al*., 1996):

$$u_k = (w_k - \lambda)^p r_k \text{ if } w_k > \lambda \text{, otherwise } u_k = 0 \tag{4}$$

The flow routing is represented by a linear transfer function corresponding to two linear reservoirs in parallel, i.e. the quick and slow flow responses with residence times k_q and k_s, respectively. However, for application to ephemeral rivers, a single reservoir may be adequate (Ye *et al*., 1997). It has been observed that although IHACRES model is widely used in semi-arid areas, there are limited studies which applied the model in a spatially distributed way (see review by McIntyre & Al-Qurashi, 2009) despite the high spatial rainfall variability associated with semi-arid areas. In semi distributed modelling, the IHACRES model is applied to each sub-catchment, and then the runoff from all the sub-catchments is integrated with a channel network model. For example, McIntyre & Al-Qurashi (2009) used a channel routing model with constant celerity parameter (c) to integrate the simulated runoff from 20 sub-catchments of the arid Wadi Ahin basin in Oman. The same channel routing model is used in our case study.

CASE STUDY AREA

Catchment description

The study area has an area of 7660 km^2 and is located within the Limpopo basin in northeastern Botswana, as shown in Fig. 1. The basin, shared between Botswana, South Africa and Zimbabwe, is critical in terms of water resources in Botswana, since all the major dams are located within it. Botswana's recent national water resource strategy (Snowy Mountains Engineering Corporation & Engineering Hydrological Environmental Management Consultants, 2006) includes development of new reservoirs, including one at the outlet of the case study catchment, to increase security of supply. The catchment is gauged at five locations leading to five sub-catchments with the outlet located in sub-catchment 5 (Fig. 1). All the sub-catchments have flow data, except for sub-catchment 4, where only the water level data are available. The area consists of gently undulating to highly variable altitude, with lowest and highest points at about 850 and 1400 m above sea level.

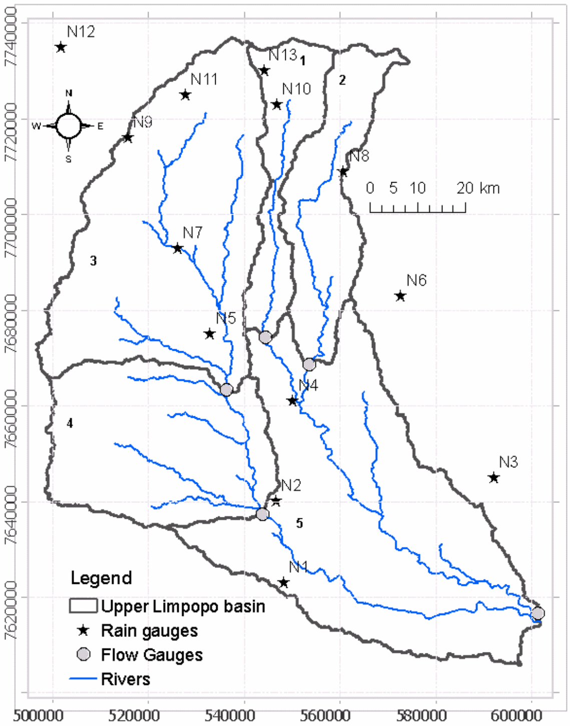

Fig. 1 The upper Limpopo basin in Botswana showing sub-catchment outlines and gauge locations.

The mean annual rainfall in the catchment is about 450 mm and the mean annual runoff at the catchment outlet is about 70 mm. The runoff mainly occurs from November to April. Annual rainfall in the study area varies from year to year. Although rainfall variability in southern Africa has often been associated with the El Niñno Southern Oscillation (ENSO) phases which are often linked with droughts and seasonal rainfall variability (Reason *et al*., 2005), recent studies suggest that ENSO is poorly correlated with rainfall in some parts of southern Africa (Manatsa *et al*., 2008), including the Limpopo basin (Kenabatho *et al*., 2009).

Spatial-temporal rainfall representation using the GLMs

Daily rainfall was stochastically generated using GLMs (Kenabatho *et al*. 2008) from the 13 gauges listed in Table 1. In general, a GLM for a $n \times 1$ vector of random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, each dependent on p predictors (whose values can be assembled into a $n \times p$ matrix $\boldsymbol{X}$ whose (i,j)th element is the value of the jth predictor for Y_i), consists of specifying a probability distribution for Y, with vector mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ such that:

$$g(\boldsymbol{\mu}) = \boldsymbol{X}\beta \tag{5}$$

where g(·) is the link function and β is a $p \times 1$ vector of coefficients (Chandler & Wheater, 2002). The GLMs in equation (5) are an extension of linear regression methods, in which variables of interest are considered to be drawn from specified families of probability distributions. The parameters of these distributions are written as linear or nonlinear functions of relevant predictors. The reader is referred to the works of Chandler & Wheater (2002) and Yang *et al*. (2005) for details about using GLMs for rainfall modelling, and to Kenabatho *et al*. (2008, 2009) for details of the Upper Limpopo application. Here, only a brief description relevant to this application is provided.

A GLM for daily rainfall is specified in two parts: a distribution defining the probability of rainfall occurrence, and another defining the amount of rainfall for non-zero occurrences. Rainfall

Table 1 Summary of rainfall and flow data used in this study.

Station code	Station name (and sub-catchment area)	Record period	Years in record	% missing data
N1	Tonota	1961–1996	36	1
N2	Shashe Dam	1973–1994	22	6
N3	Matsiloje	1980–1990	10	3
N4	Francistown	1961–2004	44	0
N5	Mathangwane	1969–2000	32	0
N6	Jackalasi No.2	1985–2002	18	9
N7	Sebina	1968–2004	37	1
N8	Tshesebe	1968–1994	27	1
N9	Ntondola	1973–2002	30	0
N10	Masunga	1973–2002	30	0
N11	Kalakamati	1973–2002	30	0
N12	Tutume	1972–2002	31	0
N13	Zwenshambe	1980–1995	16	1
F1*	Tati (539km^2)	1970–1999	30	0.04
F2*	Ntshe(788 km^2)	1970–1999	30	3.4
F3*	Mooke(2245 km^2)	1968–1999	32	0.1
F4*	Shashe (1409 km^2)	1970–1997	28	1.8
F5*	Lower Shashe (2680 km^2)	1970–1997	28	1.8

* indicates flow gauges.

occurrence at a site is typically modelled using logistic regression. If p_i denotes the probability of rain for the ith case in the data set conditional on a vector $\boldsymbol{x}_i$ of predictors, then the occurrence model is given by:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i'\beta \tag{6}$$

where β is a coefficient vector. Secondly, a distribution function is fitted to the amount of rainfall on each wet day. A gamma distribution is typically assumed, whereby:

$$\ln u_i = \xi_i'\gamma \tag{7}$$

where u_i is the mean of the gamma distribution on the ith wet day, ξ_i is a predictor vector and γ is a coefficient vector. In general, the values of the predictors will vary from day to day and, for multi-site modelling, also from site to site. Spatial predictors may include elevation and/or grid co-ordinates, while temporal predictors may include previous days' rainfall, annual trends and seasonal cycles. Model fitting involves identifying an appropriate set of predictors x and ξ in equations (6) and (7) and then estimating the corresponding coefficient vectors β and γ. In this work we used the GLIMCLIM software (Chandler, 2002) to fit the occurrence and amounts models independently, and then to simulate rainfall using the two models jointly as follows: equations (6) and (7) specify probability distributions for daily rainfall at individual sites, conditioned on the values of various predictors. A single-site sequence can then be simulated, given some initial conditions, by randomly sampling a value from the first day's distribution, using this value to construct a distribution for the second day, and so on to construct stochastically-generated daily rainfall fields. For the generation of simultaneous rainfall at multiple sites, it is usually necessary to account for dependence between the sites. The single-site simulation procedure must now be modified: instead of specifying individual distributions for the next day's rainfall at each site, it is necessary to specify a joint distribution for the next day's rainfalls at all sites. Following Yang *et al*. (2005), we proceeded in two stages, first defining a joint distribution for the rainfall occurrence to specify the number of wet sites and their positions, and then specify a joint distribution for the rainfall amounts at the wet sites (Yang *et al*., 2005; Kenabatho *et al*., 2009) to generate multi-site continuous rainfalls.

As the model is stochastic, multiple realisations will produce an envelope of simulations to represent uncertainty. Also, the models used here are constructed in such a way that if any data values are missing, their conditional distribution can be determined from the values observed at other locations. Missing values can be simulated from these conditional distributions to yield complete time series with no missing values (known as imputation) (Yang *et al.*, 2005) so that uncertainty envelopes for historical data can be generated.

In our application, the main rainfall predictors in addition to spatial and temporal predictors were humidity and sea-level pressure (see Kenabatho *et al.*, 2009) obtained from the ERA-40 re-analysis data (Uppala *et al.*, 2005). Kenabatho *et al.* (2009) identified humidity and sea level pressure as statistically significant rainfall predictors out of a pool of 14 predictors among them air and dew temperatures, sea-surface temperatures and the ENSO indices. The models were used to stochastically infill the rainfall records at the 13 raingauge sites for the period 1974–1999, which is the most complete period of the observed record to generate 10 sets of infilled rainfall data which represent uncertainty due to the missing data. Following this, the fitted GLMs were then used to generate 100 years of continuous daily rainfall with 10 imputations (infilled values) at the 13 raingauges. These 100 years of rainfall were generated conditioned on the external predictors assuming a stationary climate by assuming that the observed humidity and pressure from the ERA-40 data from 1961–2000 will repeat in the future. Climate change analysis will be performed in future work. Rainfall fields were then estimated from the 13 gauges using Thiessen polygons, providing rainfall data series suitable for input to the calibration of the rainfall–runoff model. In principle, the GLM could be used to infill spatially as well as temporally; however, the evaluation of Kenabatho *et al.* (2009) focused on temporal infilling, so here we prefer to use Thiessen polygons to generate the spatial field. Furthermore, the identified GLM and Thiessen polygons were used to synthesise 100 years of continuous daily rainfall as input to the rainfall–runoff simulation.

RAINFALL–RUNOFF MODELLING STRATEGY

Calibration strategy

The IHACRES model was run with a daily time-step in semi-distributed mode using five sub-catchments (Fig. 1). The model was calibrated for each of the four sub-catchments with flow data. The upper catchments were calibrated individually and their optimised parameters were fixed prior to calibrating the other two catchments which were calibrated together assuming uniform parameter values. Uniform random sampling was used as the search method, with 10 000 samples. Initially, the model defined in equation (4) was used together with two linear stores in parallel. The reference air temperature T' in equation (3) was fixed at the average temperature of 26°C leading to nine parameters to estimate (α, β', γ, λ, ρ, f, k_q, k_s, c). Fixing the reference temperature enables the catchment water loss to be controlled primarily by a temperature modulation parameter γ on the basis of changes in temperature to compensate for evapotranspiration losses (Jakeman *et al.*, 1990; Post & Jakeman, 1999; McIntyre & Al-Qurashi, 2009). After sensitivity analysis, following the approach used by McIntyre & Al-Qurashi (2009), the final model was reduced to a six-parameter model (without k_s, and by fixing c and f to 1). The calibration procedure was repeated 10 times—using each of the 10 rainfall realisations—to identify 10 models considered to be equally optimal. The purpose of doing this was to assess model uncertainty resulting from rainfall data. Various objective functions were tested, and, given the main ultimate task is to analyse reservoir behaviour in drought periods, a low flow performance measure was selected:

$$\mathrm{RMSEL} = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}(o_i - c_i)^2\right)} \tag{8}$$

where N is the number of data points below a threshold of observed flow L, o_i and c_i are observed and simulated flows at time steps i for which $c_i < L$. Calibration was performed at a daily time step, but the RMSEL was assessed at a monthly scale. This monthly time scale was thought to be appropriate given the large capacity of the reservoir, whereas daily data are better for generating the flows due to the nonlinearity of the response. Initial results showed that significant parameter variations occurred over three periods: 1974–1981, 1981–1991 and 1991–1999 (using hydrological years, October–September). This is a significant problem which is discussed later—for now we use the period 1981–1991 for the calibration. Using the calibration results for that period, the best parameter set obtained using each rainfall realisation, together with the 100 years of simulated rainfall data, was then used to generate 100 years of flow at the catchment outlet (i.e. the proposed reservoir inlet).

RESULTS AND DISCUSSION

Calibration and validation results

Figure 2 (the top left plot) is the calibration result showing the observed flow plotted with the simulated flow for the catchment outlet, indicating performance which is considered reasonable. The top right plot of Fig. 2 shows flow cumulated over the calibration period for the catchment outlet. In general, the model results are in agreement with the observed flow volumes, except during the 1987/88 hydrological year where the observed volumes are not well captured by the model in all the sub-catchments. From the plots, it is clear that uncertainty resulting from the 10 rainfall imputations is high. This implies that it is important to consider the uncertainty associated with temporal infilling of missing rainfall data in the calibration procedure. The uncertainty also illustrates that the stochastic component of the rainfall variability is important, and long sequences of rainfall may be needed in order to represent a range of feasible drought conditions for the reservoir analysis (hence we use 100 years).

Another source of uncertainty is parameter equifinality, whereby, irrespective of rainfall uncertainty, many parameters sets may give equally optimal objective function values. We assessed this by using the top 50 parameter sets (on the judgement that these are equifinal) for one rainfall realisation and we found that this gave similar degree of variability in the simulated hydrograph as shown in Fig. 2. Ideally, therefore, for each rainfall realisation, a number of equifinal parameter sets would be considered. We limit this analysis, however, to using the single best parameter set for each of the ten realisations. The validation results shown at the bottom plot of Fig. 2 (for the period 1991–1998) show that the model prediction has deteriorated somewhat. This is not surprising given the apparent differences in hydrological properties (in terms of the runoff coefficients) between the three sets of time series as noted earlier. These differences could be a result of changes within the catchment, i.e. those related to land-use and land cover changes. In the refinement of this work, such possibilities will need to be investigated and possible future changes considered as scenarios.

Reservoir performance results

The 10 calibrated models were used to simulate flows at the catchment outlet using the 100 years of rainfall generated by the GLM. The generated flows were subsequently used to assess performance of the proposed reservoir. The reservoir will have an active storage of 382×10^6 m^3 (Snowy Mountains Engineering Corporation & Engineering Hydrological Environmental Management Consultants, 2006). The volume balance is based on (McMahon & Adeloye, 2005):

$$Z_{t+1} = Z_t + Q_t - D_t - EV_t - L \qquad (9)$$

where Z_{t+1} and Z_t are the contents of the storage at times $t + 1$ and t; Q_t is the inflow to the storage; L_t is the release in volume units; EV_t is the net evaporation loss; and D_t is the abstraction for input to the water supply system.

Infiltration losses are neglected here. Evaporation losses (EV_t) are estimated using the reservoir surface area–storage relationships, as well as estimates of monthly open water evaporation (e) for the catchment (Snowy Mountains Engineering Corporation & Engineering Hydrological

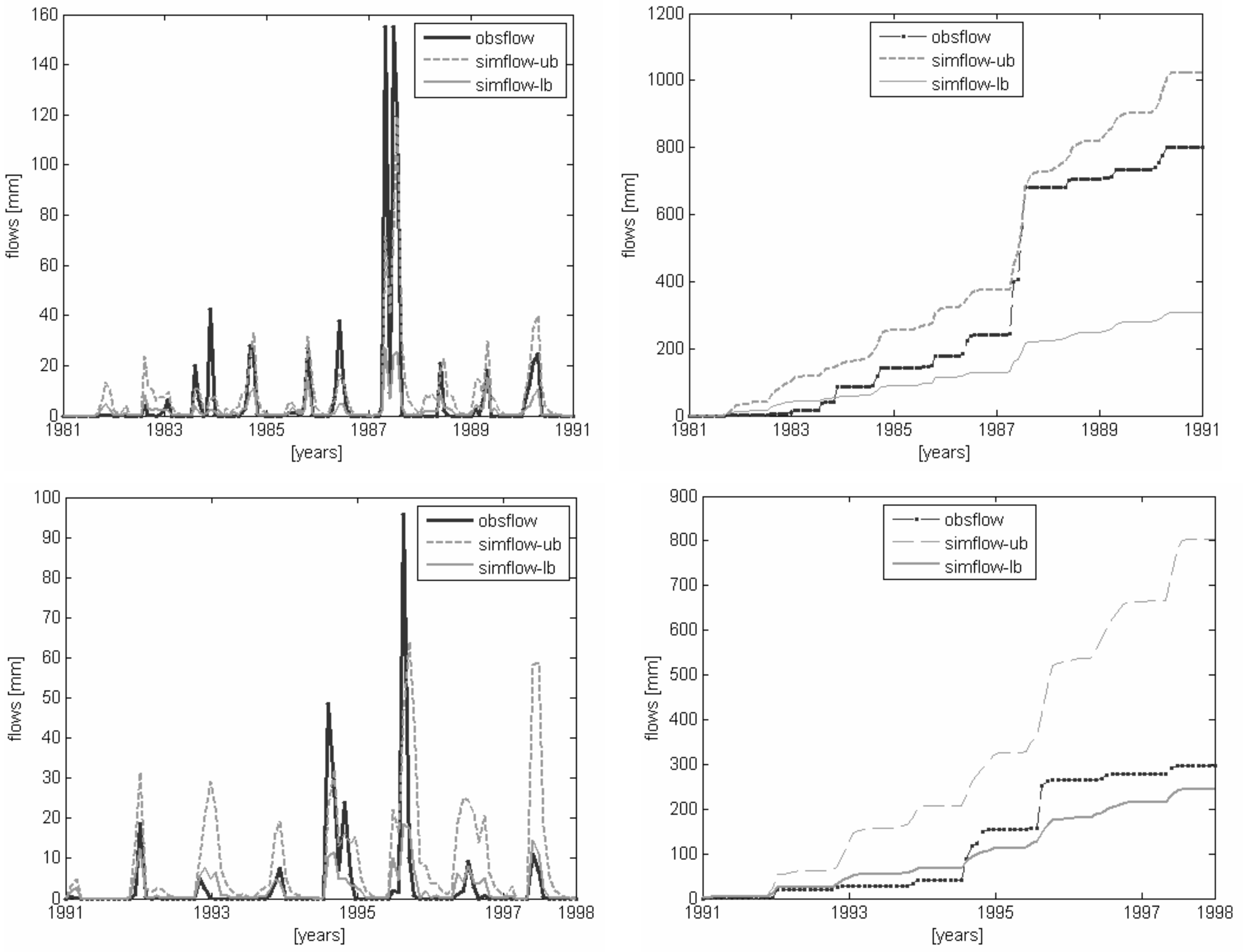

Fig. 2 Calibration (top plots) and validation (bottom plots) results for the flow gauge at the catchment outlet. The plots on the right show cumulated volumes. In all the plots *lb* and *ub* refer to lower and upper boundaries of the simulated flows.

Environmental Management Consultants, 2006) implemented through (McMahon & Adeloye, 2005):

$$EV_t = be_t + 0.5Z_t(ae_t) + 0.5Z_{t+1}(ae_t) \tag{10}$$

where a and b are the coefficients of the storage–area relationships. The contents of the reservoir were not allowed to exceed the active storage (i.e. the excess is spilled) and D, L are reduced as necessary to avoid the condition $Z < 0$. Additionally, a simple control rule is imposed whereby only 20% of the demand will be abstracted whenever the storage drops below 25% of capacity. The unrestricted demand is estimated to be 65×10^6 m^3 per year (this is the 2005 estimate—although this is expected to reach 160×10^6 m^3 by 2035, the reservoir in question is designed to meet the 2005 demands) (Snowy Mountains Engineering Corporation & Engineering Hydrological Environmental Management Consultants, 2006). The environmental flow release is 5% of the inflow. The reservoir is assumed full at the beginning of the assessment, and equations (9) and (10) are implemented at a monthly time scale.

The 10 simulations of storage are presented in Fig. 3, and it is evident that the storage displays high variability resulting from the simulations used. Also, it appears that most simulations are in agreement regarding the dry spells, particularly between the 300th and 500th months, and towards the end of the simulation period (800th and 100th months), where two of the simulated flows resulted in an empty reservoir. Table 2 gives summary statistics of the reservoir performance for

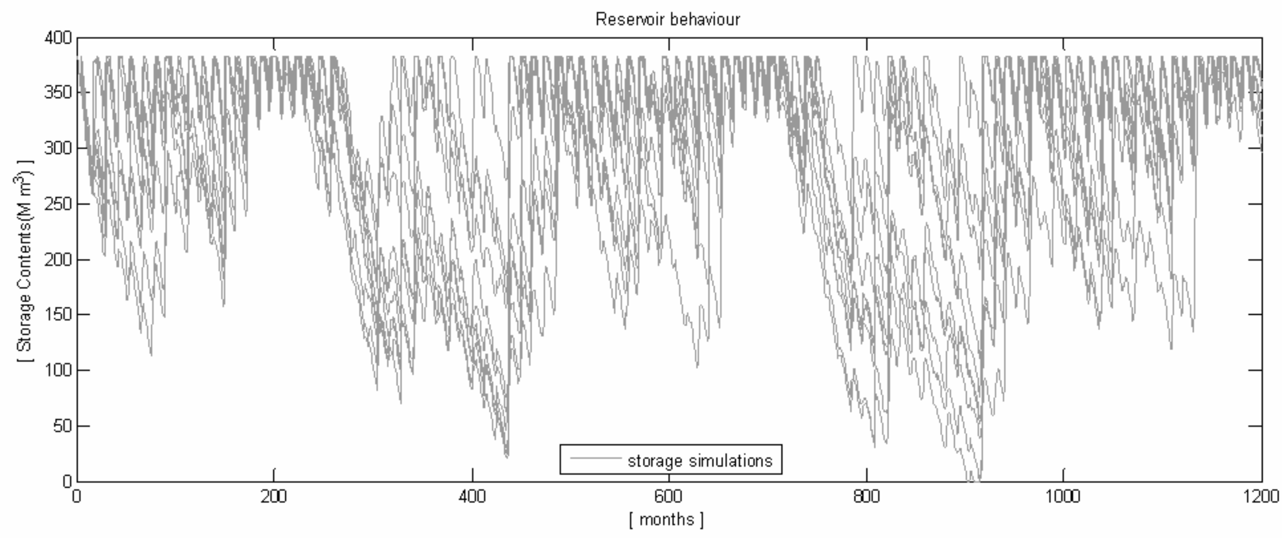

Fig. 3 Reservoir performance for a proposed reservoir in the Limpopo basin: ten sets of simulated storages.

the 10 simulations. The highest failure incident for the 100 years of assessment was 158 months, i.e. the number of times when the reservoir failed to meet the demand. This includes the period during which the reservoir was empty for seven consecutive months (between the 909 and 915 months in Fig. 3). Surprisingly, the other simulations did not record any persistent failure of this magnitude, as shown in Table 2. The time-based reliability of this reservoir (percentage of time for which demand was met) ranges between 87 to 100%. Clearly, this expression of uncertainty adds to the information of risk of failure and would not have been achieved without the use of stochastic rainfall modelling and subsequent ensemble flow simulations. Other interesting features to note in Table 2 and Fig. 3 are the variability in terms of the reservoir storage states and the months at which supply restrictions were imposed, which in practical terms will help decision makers to make informed decisions on demand management considering these wide ranges of uncertainty.

Table 2 Statistics of reservoir performance computed from 10 simulations.

	sim1	sim2	sim3	sim4	sim5	sim6	sim7	sim8	sim9	sim10
Months of restricted supply	26	0	33	0	146	120	32	0	8	0
Total months of failures, F [(b)]	29	0	39	0	158	125	35	0	10	0
Consecutive failures [(a)]	0	0	0	0	7	0	0	0	0	0
Lowest recorded storage (10^6 m^3)	39	178	31	111	0	1	22	170	77	224
Total months of spills	105	113	28	159	36	46	36	210	83	239
Reliability, R (%) [(b)]	98	100	97	100	87	90	97	100	99	100

[(a)] Maximum number of consecutive months in which the reservoir was empty.
[(b)] $R = 1 - F/M$, where M is the total number of simulated months (= 1200); F is the number of months during which the demand was not met, including when the reservoir was empty.

CONCLUDING REMARKS

The main objective of this work was to test the performance of a continuous time conceptual semi-distributed rainfall–runoff model when driven by stochastically generated spatial–temporal rainfall, and implications for reservoir operations within Botswana's national water resource strategy. It was observed that while the IHACRES model performed well in the chosen calibration period, it performed poorly during the validation period. This is presumed to be a result of land-use/land cover changes. By using multiple realisations of rainfall generated from the GLMs, it has been possible to appreciate the level of uncertainty associated with the rainfall. From this observation, and the assumption that the short rainfall record used may not be sufficient to represent a wide range of feasible drought conditions, we generated 100 years of rainfall input to drive the hydrological model. Furthermore, the uncertainty in the historic rainfall was propagated through the calibration of the hydrological model to provide 10 equally likely parameter sets. Hence 10 realisations of 100 years of reservoir input flows were generated. The results were used to assess performance of a proposed reservoir in the Upper Limpopo in Botswana, assuming some reservoir operation rules for demand management. The results show a wide range of variability in the simulated reservoir storages. This uncertainty has potential implications on the performance and the reliability of the reservoir. For example, the estimated reliability of the reservoir (frequency of failure to meet demand) varied from 0 to 13% over the 10 simulations. However, there is not enough evidence in our analysis to question the adequacy of the proposed reservoir; rather we have demonstrated a framework which needs to be extended and refined. It is suggested that future work should include other sources of uncertainty including: (1) stochastic spatial infilling of rainfall as well as temporal infilling; (2) rainfall–runoff model parameter equifinality; (3) future land-use and land cover change; (4) sensitivity to rainfall–runoff model calibration strategies; and (5) incorporating scenarios of climate change.

Acknowledgements We thank the Department of Water Affairs and Department of Meteorological Services (Botswana) for providing data used in this study. We also thank the Commonwealth Scholarship Commission (UK) for sponsoring the lead author's research at Imperial College London and for the British Hydrological Society for funding the lead author to attend the IAHS Hyderabad conference.

REFERENCES

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P. & O'Connell, P. E. (2008) RainSim: a spatial-temporal stochastic rainfall modelling system. *Environ. Model. Software* **23**, 1356–1369.

Chandler, R. E. (2002) *GLIMCLIM: Generalised Linear Modelling for Daily Climate Time Series* (software and user guide). Research Report no. 227, Department of Statistical Science, University College London, UK, http://www.ucl.ac.uk/Stats/research/abs02.html#227.

Chandler, R. E. & Wheater, H. S. (2002) Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resour. Res.* **38**, 1192, doi:10.1029/2001WR000906, 2002.

Charles, S. P., Bari, M. A., Kitsios, A. & Bates, B. C. (2007) Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia. *Int. J. Climatol.* **27**, 1673–1690.

Croke, B. F. W., Letcher, R.A., Jakeman, A.J. (2006) Development of a distributed flow model for underpinning assessment of water allocation options in the Namoi River Basin, Australia. *J. Hydrol.* **319**, 51–71.

Jakeman, A. J., Littlewood, I. G. & Whitehead, P. G. (1990) Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *J. Hydrol.* **117**, 275–300.

Kenabatho, P. K., McIntyre, N. & Wheater, H. S. (2008) Application of generalised linear models for rainfall simulations in semi arid areas: a case study from the Upper Limpopo basin in northeast Botswana. In: *Sustainable Hydrology for the 21st Century, Proc. 10th BHS National Hydrology Symposium* (Exeter), 437–444. British Hydrological Society, ISBN 1903741165.

Kenabatho, P. K., McIntyre, N., Chandler R.E., & Wheater, H. S. (2009) Stochastic simulation of rainfall in the semi arid Limpopo basin. *Int. J. Climatol.* (in review).

McIntyre, N. & Al-Qurashi, A. (2009) Performance of ten rainfall–runoff models applied to an arid catchment in Oman. *Environ. Model. Software* **24**, 726–738.

McMahon, T. A. & Adeloye, A. J. (2005) *Water Resources Yield.* Water Resources Publications, LLC, Colorado, USA.

Manatsa, D., Chingombe, W., Matsikwa, H., Matarira, C. H. (2008) The superior influence of Darwin sea level pressure anomalies over ENSO as a simple drought predictor for southern Africa. *Theor. Appl. Climatol,* **92:** 1–14.

Post, D. A. & Jakeman, A. J. (1999) Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall–runoff model. *Ecol. Modelling* **123**, 91–104.

Reason, C. J. C., Hachigonta, S., Phaladi, R. F. (2005) Interannual variability in rainy season characteristics over the Limpopo region of southern Africa. *Int. J. Climatol.* **25**: 1835–1853.

Schreider, S. Y., Jakeman, A. J. & Pittock, A. B. (1996) Modelling rainfall–runoff from large catchment to basin scale: the Goulburn Valley, Victoria. *Hydrol. Processes* **10**(6), 863–876.

Segond, M. L., Onof, C. & Wheater, H. S. (2006) Spatial–temporal disaggregation of daily rainfall from a generalized linear model. *J. Hydrol.* **331**, 674–689.

Snowy Mountains Engineering Corporation, Engineering Hydrological Environmental Management Consultants (2006) National Water Master Plan Review: Main Report. Ministry of Minerals, Energy and Water Resources, Botswana Government.

Uppala, S. M. and 45 co-authors (2005) The ERA-40 Reanalysis. *Quart. J. Roy. Met. Soc.* **131**, 2961–3012.

Wheater, H. S., Jakeman, A. J. & Beven, K. J. (1993) Progress and directions in rainfall–runoff modelling. In: *Modelling Change in Environmental Systems* (ed. by A. J. Jakeman, M. B. Beck & M. J. McAleer), 101–132. Wiley, Chichester, UK.

Yang, C., Chandler, R. E., Isham, V. S. & Wheater, H. S. (2005) Spatial–temporal rainfall simulation using generalized linear models. *Water Resour. Res.* **41**, 1–13.

Ye, W., Jakeman, A. J. & Barnes, C. J. (1995) A parametrically efficient model for prediction of streamflow in an Australian benchmark catchment with complex storage dynamics. *Environ. Int*. **21**(5), 475–758.

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M. & Jakeman, A. J. (1997) Performance of conceptual rainfall–runoff models in low-yielding catchments. *Water Resour. Res.* **33**, 153–166.