Hydrological and stochastic uncertainty: linking hydrological and water resources yield models in an uncertainty framework

DENIS A. HUGHES[1], STEPHEN J. L. MALLORY[1], BENNIE HAASBROEK[2] & GEOFFREY G. S. PEGRAM[3]

1 *Institute for Water Research, Rhodes University, Grahamstown 6140, South Africa*
stephen@waterresources.co.za

2 *Hydrosol, Pretoria, South Africa*

3 *Department of Civil Engineering, University of KwaZulu-Natal, Durban, South Africa*

Abstract Standard approaches to water resources assessments in South Africa involve generating time series of natural hydrology using a hydrological model coupled with simulating reservoir storage, abstractions, return flows, etc. using a system yield model. To account for some of the uncertainties in the representivity of the natural flow simulations, the yield models currently include a stochastic streamflow generator and output a curve quantifying likely yields with different probabilities of exceedence. Recent hydrology model developments emphasise the importance of including parameter uncertainty, especially in ungauged basins. However, this has been considered difficult to achieve with existing yield models without major structural changes or large increases in computer run time. The alternative is to add a stochastic rainfall generator within a hydrological model that also includes parameter uncertainty, and to use the output ensembles with a yield assessment model without using the stochastic streamflow generation component. This paper reports on a comparison of the two approaches in terms of modelling efficiency, similarity of yield probability assessments, and the relative contributions of parameter and stochastic uncertainty. This initial study is limited to a single basin in KwaZulu-Natal Province, South Africa.

Key words uncertainty; hydrology models; stochastic modelling; yield estimation

INTRODUCTION

Recent approaches to hydrological and water resources simulation modelling have emphasized the need to include uncertainty (Pappenberger & Beven, 2006). In estimating the yield from storage reservoirs in ungauged (or partially gauged) basins, one of the uncertainties is related to the simulation of a representative historical sequence of inflows (hydrological model uncertainty) from observed climate inputs. An additional source of uncertainty is associated with the extent to which such a historical sequence is representative of all possible sequences that might affect the yield. The duration and severity of droughts effectively determine the yield of a reservoir and it is never certain that the critical period drought is included in the historical hydrology. This problem has been dealt with in the past by generating statistically plausible flow time series from a single historical time series using multi-site, autoregressive models (Pegram & James, 1972; Salas & Pegram, 1978). These models have been developed and applied over a number of years and are part of standard practice in South Africa (Basson *et al.*, 1994). The main advantage of the stochastic model is that it can generate a large number of sequences, each with the same statistical characteristics, but not necessarily producing the same reservoir yield. The range of possible reservoir yields generated from a set of stochastic sequences represents the uncertainty in the yield estimate relating to the duration and severity of future possible droughts. However, the uncertainties associated with the statistical properties of the historical hydrology have been largely ignored in the past.

One of the possible approaches to combining the uncertainties is to generate an ensemble of historical flow time series based on model parameter probability distributions, rather than single parameter sets, within a rainfall–runoff model. The median hydrology time series can then be used to seed the stochastic streamflow generator in the yield model to generate an initial yield probability curve. The extremes of the flow ensembles generated by the uncertain rainfall–runoff model can be used with the yield model (but without the streamflow stochastic sub-model) to generate likely extreme yields that could be used to adjust the initial yield curve. This approach can be rather time consuming and does not properly integrate the two sources of uncertainty. This paper examines an alternative approach that combines the hydrological uncertainty with the stochastic uncertainty within the rainfall–runoff model, by using a multi-site stochastic rainfall model. These results are compared to those generated by the traditional stochastic streamflow approach.

STUDY AREA

The study area adopted for this initial test of the approach is the 925.1 km^2 catchment of the Mngeni River basin above Midmar Dam in KwaZulu-Natal Province of South Africa. The climate is sub-humid with a mean annual rainfall and potential evaporation of approx. 1000 mm and 1300 mm, respectively. The rainfall occurs dominantly within the summer period, the topography consists of undulating hills, while the land use is mainly commercial dairy farming, with some plantation forestry. The total area has been divided into three modelling units (U20A to C), each with its own set of parameter and climate data inputs. There is a streamflow gauging station located at the outlet of sub-basin U20B (U2H007; area of 352.9 km^2) for which relatively high confidence naturalized observed streamflow data are available (Midgley *et al.*, 1994). For the purposes of this study the catchment area is assumed to be relatively undeveloped.

GENERATING THE HYDROLOGY ENSEMBLES

The hydrology ensembles have been generated with an uncertainty version of the monthly time-step Pitman rainfall–runoff model (Hughes *et al.*, 2006). The model is a semi-distributed conceptual representation of the main hydrological processes assumed important at scales of approx. 50–5000 km^2. The model has 18 main parameters that can be specified as uncertain using normal, log-normal or uniform probability distributions. Kapangaziwiri & Hughes (2008, 2009) and Kapangaziwiri *et al.* (2009) provide details of the parameter estimation and uncertainty framework within which this version of the model has been developed. The parameter estimation process is based on available physical property data (AGIS, 2007), including topography, soils, geology and vegetation. The uncertainty ranges of the parameters are partly based on the spatial variability of the physical property data within a sub-basin (modelling unit), but may also be used to reflect the lack of information used to quantify some of the parameters.

Within the original uncertainty version of the model each ensemble is based on independent Monte-Carlo samples from the parameter distributions for each sub-basin. The outputs from the model are the time series of each ensemble (typically between 5000 and 10 000) plus a text file listing the 95, 50 and 5% exceeded flows for each month of the time series, together with the observed flows if they are available, and specified as part of the model input stream. The latter output can be used to visually assess the range of uncertainty and the extent to which the range of simulated flows includes the observed data. For the purposes of this study a second uncertainty version of the model was created that reads in a number of different rainfall time series and generates independent parameter samples for each one. The use of 500 rainfall scenarios and 500 parameter samples results in a total of 250 000 ensembles. To limit computer run time, the only outputs saved in this version are the simulated flow time series of each ensemble.

Stochastic rainfall inputs

The multisite monthly rainfall model employed in this study was based on the daily model developed by Srikanthan & Pegram (2009). The main features of this model are that it stochastically links the spatial and also the temporal dependence between the gauges using a multivariate autoregressive time series model and it post-conditions the simulations to recapture the temporal correlations and marginal statistics of the annual rainfall amounts. The first feature is fairly standard, while the second ensures that the observed fluctuations of wet and dry years is recovered, a very necessary condition for long term reservoir yield analyses. Extensive validation tests were done to ensure that the model captures the features described. Table 1 summarises the statistics of the forcing historical rainfall data and the range of equivalent values for 500 stochastic sequences. The sequences generally give lower rainfalls than the historical data based on annual totals (see the lower values of skewness, maximum and minimum in particular).

Table 1 Annual summary statistics for sub-basin U20A for the historical rainfall data and the 500 stochastic sequences.

Statistic	Historical rainfall	Stochastic rainfall sequences	
		Upper	Lower
Mean (mm)	1010.0	1010.0	1009.9
St. deviation (mm)	184.7	184.5	183.6
Skewness	0.95	0.94	0.39
Maximum (mm)	1586.8	1586.7	1527.6
Minimum (mm)	682.5	682.7	619.4

Parameter uncertainty ensembles

The rainfall data inputs into this model run were the same as those used to force the stochastic rainfall model and are recently updated sub-basin rainfalls based on available gauge information. The period of record is from October 1920 to September 2005. A total of 10 000 parameter samples were used and the results suggest a range of mean annual runoff values for U20B of 54.8 to 97.1 $\times$ 10^6 m^3, compared with the naturalized observed value of 79.8 $\times$ 10^6 m^3. Further comparisons between the time series of the ensembles and naturalized flows (and the objective functions) suggest that model has generated behavioural simulations that bracket the observed flow sequences and that the range of uncertainty is not excessive. Figure 1 illustrates the frequency distribution of minimum total flow volumes over all 24-month periods for all of the ensembles.

Parameter and rainfall uncertainty ensembles

The 500 rainfall scenarios generated from the stochastic model were combined with 500 parameter samples, using the same parameter distributions as in the previous run of the model, to generate 250 000 ensembles. These were reduced to a sample of 500 ensembles for use in the yield model by simple Monte Carlo sampling from the total of 250 000 after ranking all ensembles on the basis of the minimum 24-month reservoir inflow volumes (outlet of sub-basin U20C). This minimum inflow volume is used as a simple surrogate for the reservoir yield and the duration (24 months in this example) would be set to the critical drought period of the reservoir. Figure 1 illustrates that the frequency distributions of minimum flow volumes are the same for the total group of ensembles and the sample of 500. Figure 1 also suggests that the stochastic rainfall model has introduced some bias in the streamflow ensembles in that the median minimum volume (193 $\times$ 10^6 m^3) is substantially lower than the median given by ensembles based on only parameter uncertainty (221 $\times$ 10^6 m^3). The reasons for this are not clear at this stage, but the result is consistent with the comparisons between the statistical properties of the rainfall sequences and the historical rainfall (Table 1).

DETERMINING THE YIELD PROBABLITY CURVE

The yield of the Midmar Dam was determined using the Water Resources Modelling Platform (Mallory *et al.*, 2010). This is a water resources yield model which is integrated into a database of water use and hydrological information for South Africa and includes numerous utilities and algorithms for dealing with complex modelling problems, such as ecological flow requirements and reservoir operating rules. The model is able to generate a single "historical" yield estimate based on a single inflow time series input, or the single inflow input can be used as a seed for an ARMA stochastic generator and the resulting yields ranked to produce a yield exceedence probability curve. A yield probability curve can also be generated from multiple inputs of inflow time series and is therefore ideally suited for this specific study.

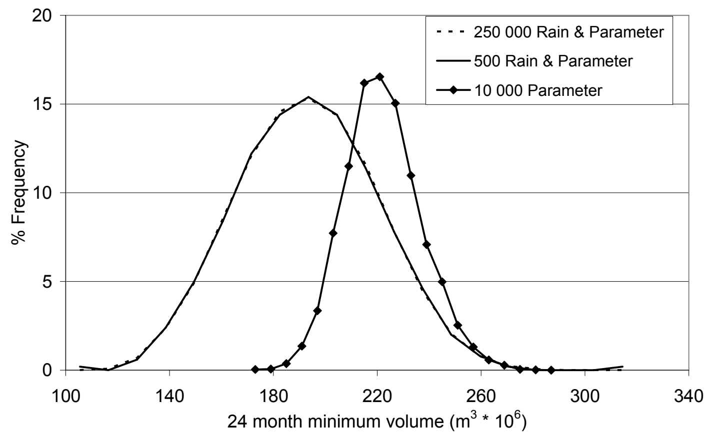

Fig. 1 Frequency distributions of minimum 24-month flow volumes for the 10 000 parameter ensembles and the two rain and parameter ensembles (250 000 total and 500 sample).

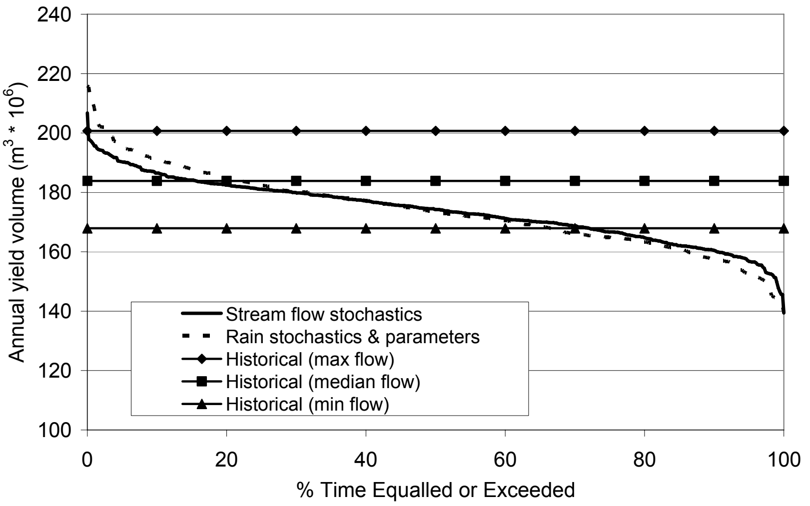

Fig. 2 Yield exceedence (probability) curves based on different yield analyses.

RESULTS

Figure 2 shows the results of the different yield analyses. The "streamflow stochastics" curve is based on seeding the stochastic streamflow model in the yield model with the median hydrology ensemble generated by the rainfall–runoff model using a single historical rainfall time series. The three horizontal lines represent the historical yield (i.e. no streamflow stochastic generation) based on the 5% ("max flow"), 50% ("median flow") and 95% ("min flow") exceeded flow ensembles from the rainfall–runoff model. The "rain stochastics and parameters" curve represents the distribution of historical yield estimates for the 500 sample ensembles generated by combining stochastic rainfall sequences and parameter uncertainty in the rainfall–runoff model. An immediate observation is that the yields generated by the streamflow stochastic approach are biased to lower yields compared with the historical yield generated from the median hydrology (used to seed the stochastic streamflow model). This result is similar to that noted for the stochastic rainfall model and requires further investigation.

Two additional streamflow stochastic yield analyses were based on the two extreme ensemble outputs from the rainfall–runoff model with only parameter uncertainty. These two extremes represent the 5% and 95% exceeded simulated flows generated by the uncertainty in the rainfall–runoff model for each month of the time series. The yields determined from these are therefore less likely to occur than the yields based on the median hydrology. Figure 3 reproduces the two yield probability curves from Fig. 1 and adds a third line ("resampled") to represent the effects of allowing for the parameter uncertainty, but still using the conventional streamflow stochastic model in the yield model. This curve has been produced by sampling from the three yield curves generated from seeding the stochastic streamflow model with the two extreme flow ensembles and the median ensemble. The size of the samples (500 for each of the extreme ensembles and 2000 for the median) has been set to reflect the probability of the different flow ensembles occurring. The three samples were combined to produce the "resampled" yield curve which is almost identical to the curve based on combining all the uncertainties in the rainfall–runoff model. Changing the size of the samples did not make any tangible difference to the results.

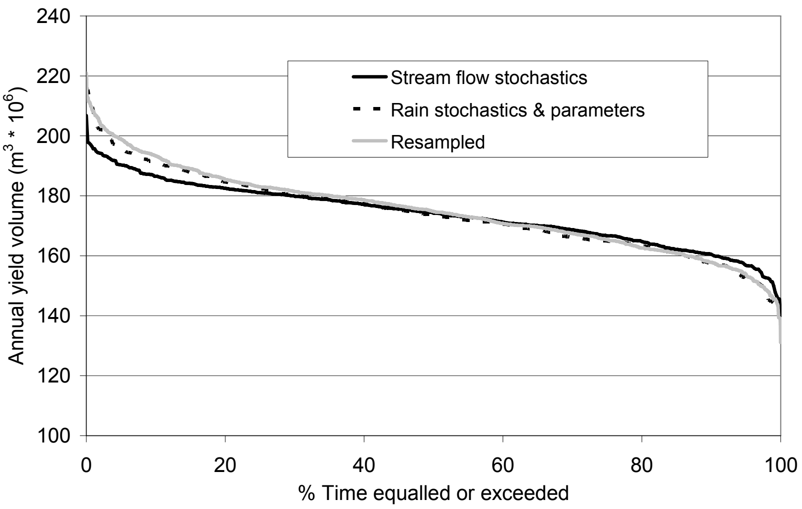

Fig. 3 Yield exceedence (probability) curves based on different yield analyses.

DISCUSSION AND CONCLUSIONS

Although the yield curves in this example are not very different (a range of 151.0 to 156.8 × 10^6 m^3 at the 95% exceedence level), this is partly a consequence of the relatively small parameter uncertainty in these sub-basins which has at least some gauged streamflow data to condition the rainfall–runoff model. The results indicate that the use of stochastic streamflow generation within the yield model has accounted for a large part of our uncertainty in water resources availability and this has been part of practical water resources management in South Africa for many years. The study has demonstrated that the explicit inclusion of parameter uncertainty is possible either through the traditional use of streamflow stochastic generation in the yield model or by including the stochastic uncertainty as part of the rainfall–runoff model. The time and computer resources required to complete a yield analysis using the two approaches are not very different now that the appropriate software is available.

Additional assessments of the effects of parameter uncertainty in South Africa (Kapangaziwiri & Hughes, 2009; Kapangaziwiri *et al.*, 2009) suggest that there could be much larger differences in yield estimates in other basins. There is therefore a need to extend this type of study so that water resources managers can be made more aware of the uncertainties associated with the use of model outputs for decision making.

Acknowledgements The research undertaken for this study is part of a Water Research Commission of South Africa funded project (K5/1838) on developing uncertainty methods for water resources management.

REFERENCES

AGIS (2007) Agricultural Geo-Referenced Information System. www.agis.agric.za (accessed December 2010).

Basson, M. S., Allen, R. B., Pegram, G. G. S & van Rooyen, J. A. (1994) *Probabilistic Management of Water Resource and Hydropower Systems*. Water Resources Publications, Colorado, USA.

Hughes, D. A., Andersson, L., Wilk, J. & Savenije, H. H. G. (2006) Regional calibration of the Pitman model for the Okavango River. *J. Hydrol*. **331**, 30–42.

Kapangaziwiri, E. & Hughes, D. A. (2008) Revised physically-based parameter estimation methods for the Pitman monthly rainfall–runoff model. *Water SA* **32**(2), 183–191.

Kapangaziwiri, E. & Hughes, D. A. (2009) Assessing uncertainty in the generation of natural hydrology scenarios using the Pitman monthly model. In: *14th SANCIAHS Symposium* (Pietermaritzburg, KwaZulu-Natal, September 2009).

Kapangaziwiri, E., Hughes, D. A. & Wagener, T. (2009) Towards the development of a consistent uncertainty framework for hydrological predictions in South Africa. In: *New Approaches to Hydrological Prediction in Data Sparse Regions* (ed. by K. K. Yilmaz *et al.*), 84–93 (Proc. Symposium HS2 at the Joint IAHS & IAH Convention, Hyderabad, India, Sept. 2009). IAHS Publ. 333, IAHS Press, Wallingford, UK.

Mallory, S. J. L., Desai, A., Odendal, P & Ward, J. (2010) The Water Resources Modelling Platform: User Guide Version 3.2. www.waterresources.co.za.

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 No. 298/1.1/94 to 298/6.1/94, Pretoria, South Africa.

Pappenberger, F. & Beven, K. (2006) Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resour. Res.* **42**, W05302, doi:10.1029/2005WR004820.

Pegram, G. G. S. & James, W. (1972) Mulitvariate auto-regressive model for the generation of operational hydrology. *Water Resour. Res.* **8**(4), 1074–1076.

Salas, J. D. & Pegram, G. G. S. (1978) A seasonal multivariate multilag auto-regressive model in hydrology. In: *Modelling Hydrological Processes* (ed. by H. J. Morel-Seytoux, J. D. Salas, T. G. Saunders & R. E. Smith). Water Resources Publications, Fort Collins, Colorado, USA.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshim, V., Liang, X., McDonnel, J. J., Mendiondo, E. M., Connell, O., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. & Zehe, E. (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exiting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6), 857–880.

Srikanthan, R. & Pegram, G. G. S. (2009) A nested multisite daily rainfall stochastic generation model. *J. Hydrol.* **371**, 142–153.