River headwater flows - modelling the spatial and temporal correlations within a larger framework

Jane Gooday

Australian Bureau of Agricultural and Resource Economics

GPO Box 1563 Canberra, Australian Capital Territory

Abstract: In this paper a method for simulating river headwater flows in Australia's southern Murray-Darling Basin is described. This work was conducted for the purpose of analysing competing demands for water in the Basin. The methodology described is largely driven by two issues. Firstly, the spatial relationships between river headwaters need to be captured in conjunction with any temporal correlations that may exist (adding to the complexity is the fact that some river headwaters have no flows for consecutive months). The second issue requires selecting a robust methodology which can be used within a large scale simulation framework.

1. INTRODUCTION

The issue of managing competing demands for water resources in the southern Murray Darling Basin of Australia is one that requires a balance of water use for electricity generation, downstream irrigation and the environment. To objectively examine the resource use trade-offs inherent in this system requires a modelling approach cast in a system-wide context; integrating hydrological, biophysical and economic relationships. The purpose of this paper is to discuss a method of statistical modelling for the hydrological component within a model of the southern Murray Darling Basin (Scoccimarro, Beare and Brennan 1997).

Synthesised hydrological flows are used to generate synthetic realisations using numerical techniques (McMahon, Pretto and Chiew 1996; Hipel and McLeod 1994). Synthesised flows are often used within a simulation framework to examine the operation of water storages and the flow on effects on the competing demands for water (Thompstone, Hipel and McLeod 1987).

Risk is a central aspect of water management strategies (Musser and Tew 1984). Hence, the ability to generate stochastic river flows which reflect the likelihood and impact of critical low and high flow conditions over an extended time frame, is important in evaluating alternative management options (Hall et al 1969). At the catchment scale, climatic events and the geomorphology of the region may give rise to complex spatial and temporal relationships and affect critical flow patterns. The objective of this analysis is to develop a framework which is robust in simulation, generating results within observed bounds and able to adequately sample from the tails of the river flow distributions with minimum bias. These are often conflicting objectives in statistical analysis and present a number of interesting problems in this application.

2. THE LARGER MODEL

The model developed by the Australian Bureau of Agricultural Resource Economics (ABARE) examines the links between water use in the southern Murray Darling Basin. The modelling system, constructed using Extend™ software (Imagine That Inc. 1995), consists of a set of modules which can be linked to construct a network. The modules can be categorised into five broad types:

- **Physical modules** represent reaches of the river system and infrastructure such as dams, storages and electrical generation facilities. These modules track the characteristics of the water flows and, for example, the relationships between storage volumes and electricity generation capacity.

- **Management modules** represent the institutional arrangements which govern the allocation of water, such as the allocation of water by the Murray Darling Basin Commission between the states.

- **Economic modules** represent the individual or collective actions of commercial water users. These modules generate water demands through a process which optimises the financial return to, for example, a farm or electricity generator.

- **Statistical modules** generate exogenous system wide data, such as headwater flows or electricity loads.

- **Market modules** represent the interaction of economic participants within a market, such as a spot electricity market or a market for trading temporary water entitlements.

The modules are linked into a user specified network of physical, managerial and economic relationships through a graphical computer interface.

3. DATA

The objective of the analysis is to estimate a joint distribution function for 25 river headwaters which encompass the water sources of the Snowy Mountains, the Victorian and New South Wales sources to the Murray river and inflow sources of the Murrumbidgee river. The 87 years of data for each headwater provide monthly records of flows from 1906 to 1992. All monthly flow distributions are right skewed and highly variable (see figure 1). The majority of the river headwaters downstream of the Snowy scheme have highest average monthly flows in August while headwaters associated more directly with the scheme, and are more influenced by snowmelt, have higher flows in September and October.

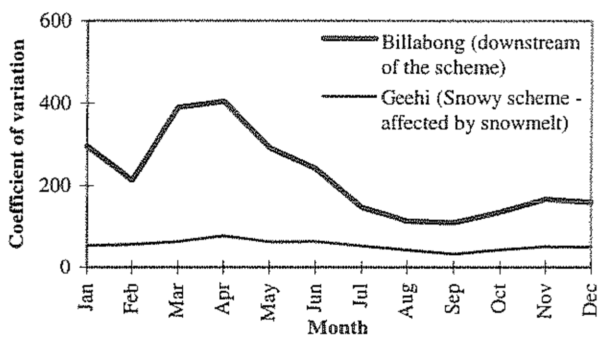


Figure 1. Monthly coefficient of variation for two selected headwaters

A preliminary exploratory analysis indicated there were significant spatial and serial correlation between the headwater flows. In addition, there were a significant number of periods in which zero flows were recorded in several headwaters downstream of the Snowy scheme. A general approach to estimation of a non-negative distribution with zero observations is to make use of conditional probabilities of zero and positive flow. The existence of spatial correlations makes the problem more difficult in that the number of conditional probabilities and distributions to be estimated increases combinatorially. Serial correlations complicate the problem substantially further. The author is unaware of published work in which these problems have been collectively addressed.

In this analysis, a non-conditional estimation approach is used to estimate the joint headwater flow distribution. Data is then simulated from the estimated joint distribution function under alternative truncation and windsorisation rules. The bias in the simulated data is then examined.

4. METHODOLOGY

The model for water flow f at time t may be written as:

(1) $$f_{r,t} = \mu_{r,m} + \sigma_{r,m}\varepsilon_{r,t},$$

where μ is the average monthly flow and σ is the standard deviation of the monthly flow in month m, for each river headwater r and ε is the residual.

To remove any linear dependencies or spatial relationships in the river headwaters, principal component analysis was used:

(2) $$P_r = b_{r,1}\varepsilon_1 + b_{r,2}\varepsilon_2 + \ldots + b_{r,r}\varepsilon_r,$$

where the b's are the coefficients for the principal component P. Approximately 77 per cent of the standardised variance is explained by the first three principal components with subsequent components contributing less than 5 per cent. While the first eigenvector showed equal loadings on all the headwaters, the second component had high positive loadings on headwaters downstream of the Snowy Mountains catchment and high negative loadings on headwaters contained within the catchment of the Snowy Mountains.

Each principal component series was found to be stationary. The partial autocorrelations of the principal component series suggested using a first order autoregressive model, AR(1) process to capture the temporal correlations:

(3) $$P_{r,m} = \upsilon_r + \alpha_r P_{r,m-1} + \zeta_{r,m},$$

where $\zeta_{r,m}$ is the residual series which was distributed normally. Lag structures in the majority of the residual series were found to contain white noise. AR models of higher order showed an increase in the Akaike information criteria, that is, a possible over parameterisation of the model. The first lag showed the most dominant autocorrelation effect in all the series. Smaller effects that were different for each principal component and any temporal cross-correlations between each principal component series, were not modelled. The common AR1 structure re-introduced a small spatial correlation between the residuals. However, the level of correlation was low and no corrective action was taken.

The Shapiro-Wilk test (see Shapiro and Wilk 1965) indicated that the residuals were not normal. Most were highly kurtotic and some were also skewed. Distribution functions were estimated for these residuals using a mixture of two normal distributions (see Johnson, Kotz and Balakrishnan 1994). To estimate the parameters of this nonlinear model, the alogorithm employed uses nonlinear optimisation which minimises the sum of the squared deviations between the predicted and fitted cumulative distributions (see SAS/STAT User's Guide 1990). The skewness associated with some of the distributions was captured by varying both the means and variances of the component normal distributions. For distributions with little or no skewness the means were fixed and standard deviations allowed to vary. The form of this probability density function is

$$(4)\quad \sum_{i=1}^{2}\omega_i\left(\sqrt{2\pi}\sigma_i\right)^{-1}\exp\left[-\frac{1}{2}\left\{\frac{\left(x-\xi_i\right)^2}{\sigma_i^{\,2}}\right\}\right],$$

where ξ_i is the mean and σ_i is the standard deviation of the normal distribution i, and ω_i are the weights where

$$\sum_{i=1}^{2}\omega_i = 1.$$

The results of fitting the mixture of normal distributions to the residuals of the first and third principal component are shown in figures 2 and 3. The residuals of the component are plotted against the cumulative predicted value and compared to the original cumulative distribution. Residuals of the first and third principal components were positively skewed and leptokurtic (kurtosis of 4.44 and 9.56 respectively). The fit of the first principal component where the mean was not allowed to vary indicates a reasonably good fit of the mixture of distribution (figure 2). However, the third principal component, where the mean was allowed to vary, shows a significant improvement on the model fit with all sections of the distribution estimated well (figure 3).

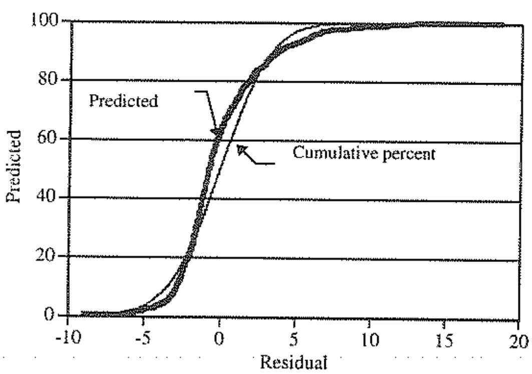


Figure 2. Fit of the mixture of normal distributions: first principal component

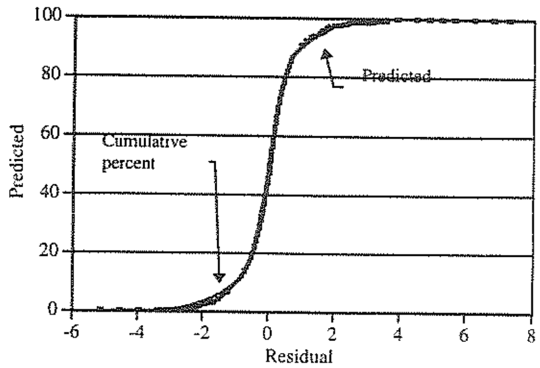


Figure 3. Fit of the mixture of normal distributions: third principal component

5. IMPLEMENTATION AND RESULTS

Each stage in the methodology described in the previous section was reconstructed in the headwater flow module of the southern Murray Darling Basin model. Successive long run simulations based on this methodology produced distributions of headwater flow ($\hat{f}_{r,m}$) where $\hat{\mu}_{r,m} = \mu_{r,m}$ and $\hat{\sigma}_{r,m} = \sigma_{r,m}$. However, $\hat{f}_{r,m}$ was not found to fall within the observed bounds.

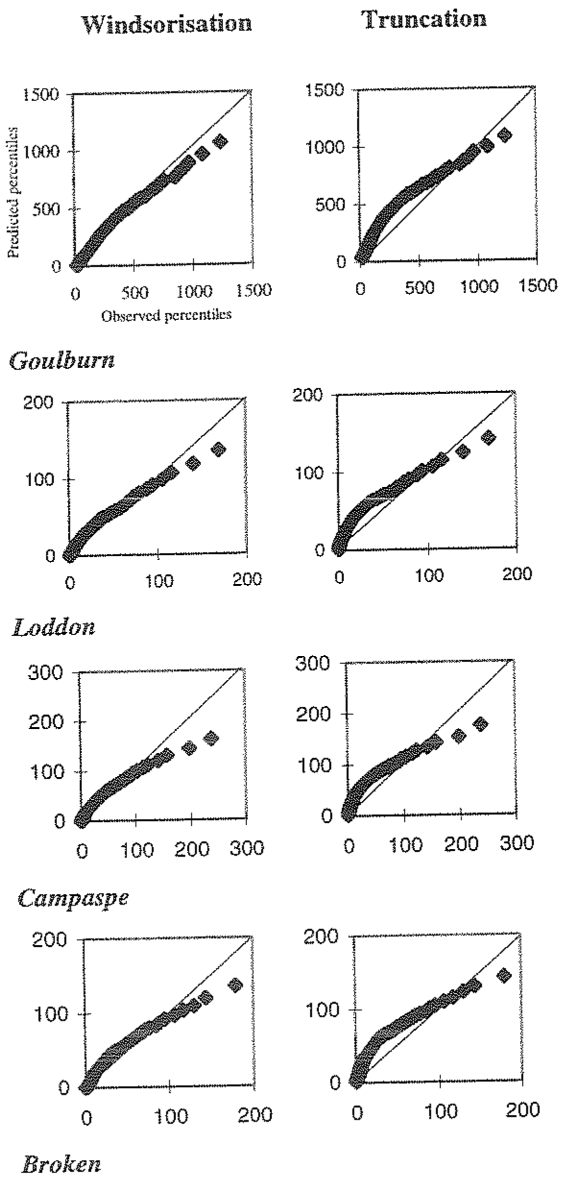
In dealing with this issue, two methods were employed separately. The first scenario involved windsorising $\hat{f}_{r,m}$ as follows:

(5) if $\hat{f}_{r,m} <= p_{r,1}$ then $\hat{f}_{r,m} = p_{r,1}$ and

if $\hat{f}_{r,m} => p_{r,99}$ then $\hat{f}_{r,m} = p_{r,99}$,

where $p_{r,1}$ and $p_{r,99}$ were respectively the first and ninety ninth percentiles of the observed distribution. The second method involved truncating $\hat{f}_{r,m}$ so that if $\hat{f}_{r,m} <= p_{r,1}$ or if $\hat{f}_{r,m} => p_{r,99}$ then the process of predicting $\hat{f}_{r,m}$ was repeated until $p_{r,1} <= \hat{f}_{r,m} <= p_{r,99}$.

Similar effects of applying each method to each headwater flow were observed across all headwaters. As an example, results of the 1000 years of simulated flows for eleven headwaters are summarised in figure 4. Both large and small headwaters are shown including those headwaters with consecutive low flows (Billabong headwater). Generally, both methods show an over-estimation of the lower percentiles and an under estimation of the higher percentiles. Of the two methods tested, windsorised $\hat{f}_{r,m}$ provides the best prediction of $f_{r,m}$ with truncation overestimating to a higher degree the percentiles of the distribution.



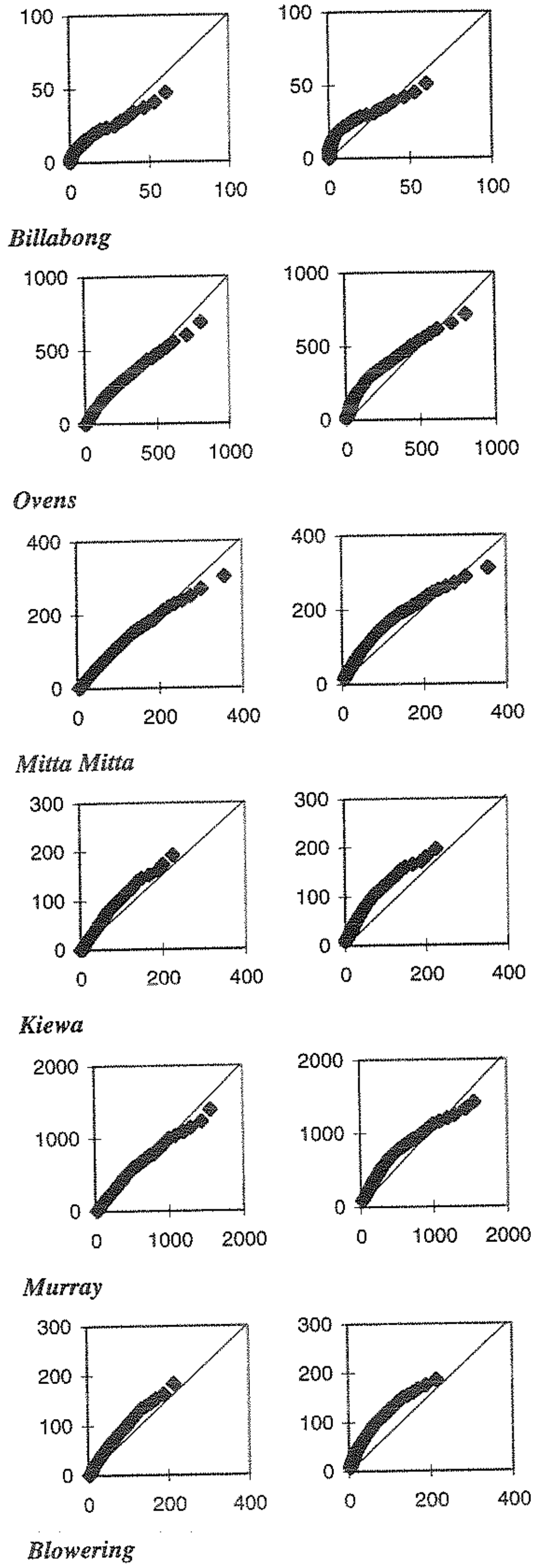


Figure 4. Observed percentiles versus simulated percentiles

6. CONCLUSION

The method described in this paper was designed to be robust in simulating headwater flows within a model of the southern Murray Darling Basin. The objective was to estimate a joint probability function that when used within a simulation framework, provided adequate sampling from the tails of the distribution to ensure the occurrence and magnitude of critical periods was maintained. The data used in this analysis was spatially and temporally correlated with some headwaters containing consecutive months of zero flows. As a result of these characteristics, the general approach to the estimation of a non-negative distribution was made more complex. The spatial and serial correlation caused the number of conditional probabilities to increase combinatorially. As a result, headwater flows are estimated from non-conditional joint distributions under two alternatives; windsorisation and truncation. The two methods were found to consistently underestimate the upper tail of the distribution. However, as a model for synthesised hydrological flows within a larger framework, the windsorisation method was found to be a more robust method predicting the bulk of the distribution reasonably well.

ACKNOWLEDGMENTS

The author would like to thank the Snowy Mountains Hydro Electric Authority, the New South Wales Department of Land and Water Conservation, the Victorian Department of Natural Resources and Environment and the Murray Darling Basin Commission for supplying the data and Dr. Stephen Beare and Greg Griffiths of ABARE for their assistance in the modelling methodology.

REFERENCES

Hall, W. A., Askew, A. J. and Yeh, W. W. G., Use of the critical period in reservoir analysis, Water Resources Research, vol. 5, pp. 1205–1215, 1969.

Hipel, K. W. and McLeod, A. I., *Time Series Modelling of Water Resources and Environmental Systems,* Elsevier, Amsterdam, 1994.

Imagine That Inc., *Extend: Performance Modelling for Decision Support*, California, 1995.

Johnson, N.L., Kotz, S and Balakrishnan, N., *Continuous Univariate Distributions,* vol. 1, 2nd edn, Wiley, USA, pp. 163–168, 1994.

McMahon, T.A., Pretto, P.B. and Chiew, F.H.S., 'Synthetic hydrology - Where's the hydrology?', in Tickle, Goulter, Xu, Wasimi and Bouchart (eds), *Stochastic Hydraulics '96*, Balkema, Rotterdam, pp. 3–14, 1996.

Musser, W.N. and Tew, B.V., 'Use of biophysical simulation in production economics', *Southern Journal of Agricultural Economics,* vol. 16, no. 1, pp. 77–86, 1984.

SAS Institute Inc., *SAS/STAT User's Guide*, Version 6, 4th edn, vol. 2, Cary, NC, USA, pp. 1135–1193, 1990.

Scoccimarro, M., Beare, S. and Brennan, D., 'The Snowy River: opportunity costs of introducing environmental flows', *Australian Commodities,* vol. 4, no. 1, pp.67–78, 1997.

Shapiro, S.S. and Wilk, M.B., "An analysis of variance test for normality (complete samples)", *Biometrika*, vol. 52, no. 3 and 4, pp. 591–611, 1965.

Thompstone, R. M., Hipel, K. W. and McLeod, A.I., 'Simulation of Monthly Hydrological Time Series', in MacNeil, I.B. and Umphrey, G.J. (eds), *Stochastic Hydrology,* D. Reidel Publishing Company, pp. 57–71, 1987.