Regionalisations of Flow Variables Used In Modelling Riverine Material Transport In The National Land And Water Resources Audit

W.J. Young, P. Rustomji, A. O. Hughes, D. Wilkins

CSIRO Land and Water, Canberra Technical Report 36/01

August 2001

Author: W.J.Young

CSIRO Land and Water, PO Box 1666, Canberra, 2601, Australia

E-mail: bill.young@csiro.au

Phone: 61-2-6246-5729

For bibliographic purposes, this document may be cited as:
Young, W.J. (2001). Regionalisations of flow variables used in modelling riverine material transport in the National Land and Water Resources Audit. CSIRO Land and Water Technical Report 36/01, CSIRO Land and Water, Canberra, Australia.

Acknowledgements
Graeme Priestley and Leo Lymburger provided GIS assistance to this work, and Jeff Wood provided statistical advice. The work was partially funded by the National Land and Water Resources Audit – a Program of the Natural Heritage Trust.

TABLE OF CONTENTS

SUMMARY

As a part of the National Land and Water Resources Audit, sediment and nutrient transport were modelled in large-scale river networks across Australia. These models required estimates of a number of hydrologic variables for each link in the river network. To provide these estimates, simple hydrologic regionalisation models were developed. These models predict the required hydrologic variables as functions of drainage area to the network link and the mean annual rainfall spatially-averaged across this drainage area. The hydrologic data used to build the models were a mixture of modelled daily flows and observed daily flows. The primary model that was developed is used to estimate the mean annual flow in a network link. Mean annual flow models were developed for three different regions of Australia defined by similarity of mean annual runoff coefficients. The models vary in robustness between regions, partly as a result of different size data sets. The mean annual flow models are for drainage areas between 50 km^2 and 2000 km^2. Values for links with larger drainage areas were estimated by linear interpolation between the regionalised values and AWRC basin outflow estimates. Secondary models were developed to predict the median daily flow, the bankfull flow, the median over-bank flow, and a parameterised function of daily flows used to estimate sediment transport capacity. These secondary models were all functions of the mean annual flow. While most of these variables are reasonably predicted by mean annual flow, the median daily flow (which for the highly skewed flow distributions of most Australian rivers is an indicator of typical low flow conditions) is poorly predicted by mean annual flow. Predictors other than drainage area and mean annual rainfall are required to build more robust regionalisation models for median daily flow.

HYDROLOGIC VARIABLES USED IN MATERIAL TRANSPORT MODELLING

As part of the assessments of river condition undertaken by the National Land and Water Resources Audit, the sediment and nutrient loads transported through the river networks of the intensive land use zone of Australia were modelled. Both the sediment transport model (SedNet, Prosser *et al.* 2001) and the nutrient transport model (ANNEX, Young *et al.*, 2001) estimated average annual loads transported in each link in river networks for drainage areas of 50 km^2 and greater. The river networks that were modelled included a total of nearly 15,000 network links. For each of these links the following five hydrologic variables were required:

1. Mean annual flow (MAF): MAF is used in ANNEX (together with suspended sediment loads) to partition total phosphorus loads between sediment-bound and dissolved forms (Young *et al.*, 2001).
2. Median daily flow (medQ): medQ is used in ANNEX as the representative discharge for modelling denitrification losses (Young *et al.*, 2001).
3. Bankfull flow (Qbf): the amount of floodplain sedimentation predicted by SedNet is partly determined by the proportion of the total flow that is in excess of Qbf. Qbf is also used in SedNet as a predictor of bank erosion. Qbf is assumed to be adequately represented by the mean annual flood that has a return period of 1.58 years on the annul series (Dury *et al.*, 1963).
4. Median over-bank flow (Qob): Qob is used in SedNet as the representative discharge for modelling floodplain sedimentation (Prosser *et al.*, 2001).
5. Sediment transport capacity discharge (sigQ): sigQ is the mean annual value of a parameterised function of daily flows that is used in SedNet to estimate the sediment transport capacity. The function is the annual sum of daily flows raised to the 1.4 power (Equation 1):

$$sigQ = \frac{365}{n}\sum_{i=1}^{n} Q_i^{1.4} \qquad 1$$

The exponent is the median of a series of empirically-derived values from a large number of sediment transport capacity studies reviewed by Prosser and Rustomji (2000).

DATA USED TO DEVELOP REGIONALISATION MODELS

To develop regionalisation models for the above hydrologic variables, 314 values of the variables were calculated from 282 simulated daily flow sequences (Appendix A) and 32 observed daily flow sequences (Appendix B). The simulated flow records were for 100 years. The observed flow records were for between 41 and 82 years. Both the simulated and observed flow records were for gauging stations with drainage areas between 50 km^2 and 2000 km^2. The simulated flow sequences were obtained from Peel *et al.*, (2000) who modelled daily flows for 331 gauging stations. The 45 stations where the daily model was assessed by Peel *et al.* (2000) to be “poor” were excluded from the data set. A further 3 stations that were assessed by Peel *et al.* (2000) to be “passable” were also excluded because of the difference between the modelled and measured values of total flow volume exceeded 10%. The data set was supplemented with 32 Queensland gauging station records because of the under-representation of Queensland stations in the data set of Peel *et al.* (2000).

Models to predict these variables had to rely on data items that were available for every network link. The main data items that were available were the drainage area to a network link, and various gridded surfaces of climate data based on spatial and temporal interpolations of long-term observed climate records. Using the digital elevation model that was used to define the river network, values of climate variables that were spatially averaged across the drainage area to each gauging station were calculated for use in developing models. Available climate surfaces were various 20 km by 20 km grids produced by the Bureau of Meteorology, and the 5 km by 5 km mean annual rainfall grid produced by the Queensland Department of Natural Resources and Mines (http://www.dnr.qld.gov.au/silo). Models using 20 km by 20 km grids of high and low percentiles of annual rainfall were investigated, but did not perform as well as models using the 5 km by 5 km grid of mean annual rainfall. All models therefore relied solely on the spatially-averaged values from the 5 km by 5 km grid of mean annual rainfall and drainage area. This gridded rainfall surface is derived from interpolations (using ordinary Krigging) of monthly rainfall from over 6000 rainfall stations across Australia.

HYDROLOGIC REGIONALISATIONS

The hydrologic regionalisations of the five variables required for the NLWRA sediment and nutrient transport modelling that were developed from the above data sets are described below. In some cases the regionalisations reported here are minor refinements of those that were actually used in the NLWRA modelling. The regionalisations that were used in the modelling are reported in Prosser *et al.* (2001). While the regionalisations reported here are slightly better models based on a more consistent approach, the differences are inconsequential in terms of their application in the NLWRA sediment and nutrient modelling because of other larger uncertainties.

Mean Annual Flow

The well known Rational Formula estimates runoff volumes as the product of catchment area (*A*), a rainfall rate, and a runoff coefficient. The following generalised form of this formula was used to fit models of mean annual flow:

$$MAF = k\,A^{m}\,Rf^{n} \qquad 2$$

where Rf is mean annual rainfall and k, m and n are constants. To allow multiple linear regression to be used to fit model, logarithms (base 10) were taken of MAF, A and Rf data and models of the following form were fitted:

$$\log_{10}MAF = \log_{10}k + m*\log_{10}Area + n*\log_{10}Rf \qquad 3$$

For the full data set (314 stations) the model of the form of Equation 3 has an adjusted R^2 of 0.863 and has the following parameter values:

$\log_{10}k = -4.713 \pm 0.239$, $n = 2.376 \pm 0.0727$, $m = 0.927 \pm 0.0259$.

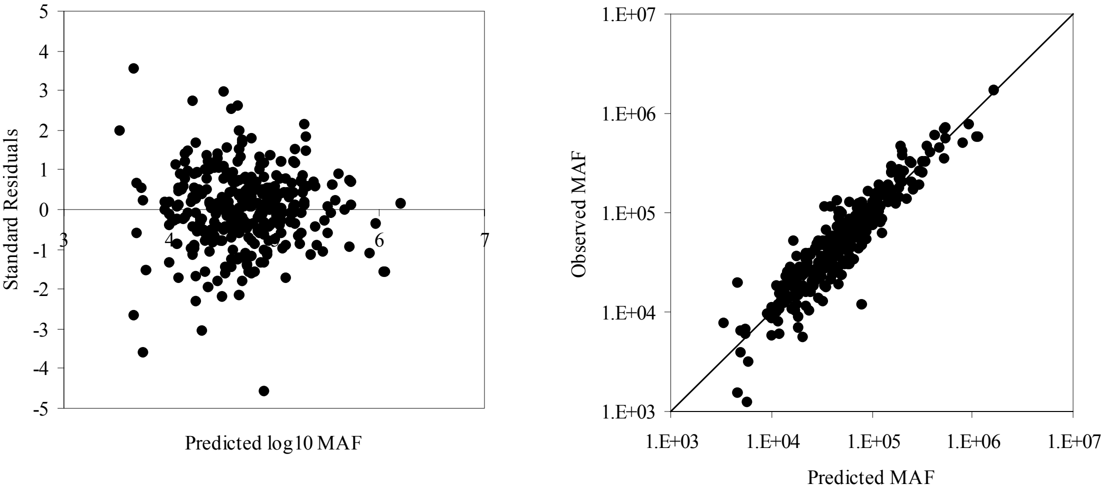


Figure 1: (a) Standard residuals against predicted $\log_{10}MAF$ values for multiple regression model of full data set, and (b) predicted MAF values against observed MAF values for multiple regression model of full data set.

The data were confirmed (at the P=0.05 level) to be normally distributed about the regression line using the Kolmogorov-Smirnov test. However, the data failed the Spearman rank correlation test for constant variance of MAF at the P=0.05 level. This is reflected in Figure 1a where the residuals are larger for lower predicted values of $\log_{10}MAF$; that is, the model generally performs better for larger values of MAF. Because the majority of the "observed" MAF values are from simulated flows, there is a considerable degree of uncertainty associated with the values. For example, the two largest negative residuals are for simulated data for stations were Peel *et al.* (2000) assessed the model to be only "passable". Given the considerable uncertainties in the data used to develop the model, it was decided that data transformations to achieve normality were not justified. The general model was accepted with the recognition that it performs better for stations with MAF values greater than 10,000 ML/day.

A measure of model error (E) was used to assess model performance. E is the mean percentage of the ratio of the root mean square difference between the predicted and observed values to the observed value (Equation 4). The model based on the full data set has a root mean error (E) of 34%.

$$E(\%) = \frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{(PRED_i - OBS_i)^2}}{OBS_i}\times\frac{100}{1} \qquad 4$$

Improved models can be obtained by fitting models to separate regions of the country. Initially, the use of AWRC Drainage Divisions as regions was investigated. However, the models for some Drainage Divisions were very similar, and for some Drainage Divisions there were insufficient data to build reliable models. Regions were therefore defined by homogeneity of mean annual runoff coefficients defined as *MAF*/(*A.Rf*). Mean annual runoff coefficients were calculated for the stations within each Drainage Division represented in the intensive land use zone of Australia (Table 1). Distinct differences were apparent between the northern river basins of the Murray-Darling and the southern river basins, and so these areas were separated.

Drainage Division	**Mean Annual Runoff Coefficient**
1 North East Coast	0.18
2 South East Coast	0.22
3 Tasmania	0.41
4 Murray-Darling Basin	0.16
North (Basins 1-15)	0.20
South (Basins 16-26)	0.10
5 South Australian Gulf	0.08
6 South West Coast	0.17
7 Indian Ocean	No data
8 Timor Sea	0.21
9 Gulf of Carpentaria	No data

Table 1: Mean annual runoff coefficients for the Drainage Divisions represented in the intensive land use zone of Australia.

On the basis of these mean annual runoff coefficients, the following three regions were defined:

- Region 1: Drainage Divisions 1, 2, 4 (Basins 1 to 15), 6, 7, 8, 9
    - Mean annual runoff coefficient 0.21
- Region 2: Drainage Division 3
    - Mean annual runoff coefficient 0.41
- Region 3: Drainage Divisions 4 (Basins 16 to 26) and 5
    - Mean annual runoff coefficient 0.09

Because of small data sets, it was not possible to determine significantly different values for all three model parameters for each region. Instead, the values of *m* and *n* from the full data set were used in Equation 3 to determine *k* values (quoted for comparison as $\log_{10}k$) for each of these regions by linear regression between *MAF* and $A^n{*}Rf^m$ (Table 2). These regression were constrained by setting the constant (or intercept) to zero.

Region	$\log_{10}k$	Observations	Adjusted R^2	Model Performance – E(%)	
				Region Model	Full Model
1	-4.754 ± 0.009	259	0.847	26.8	27.3
2	-4.380 ± 0.028	12	0.823	32.7	49.8
3	-4.802 ± 0.018	43	0.846	49.1	70.0

Table 2: Details of *MAF* linear regression models for separate regions.

The values of root mean error (E) show that the regional models perform better than the full model of all regions, with substantial improvements in both Region 2 and Region 3 (Table 2). The plots of residuals and observed vs predicted are shown in Figures 3 to 5 below.

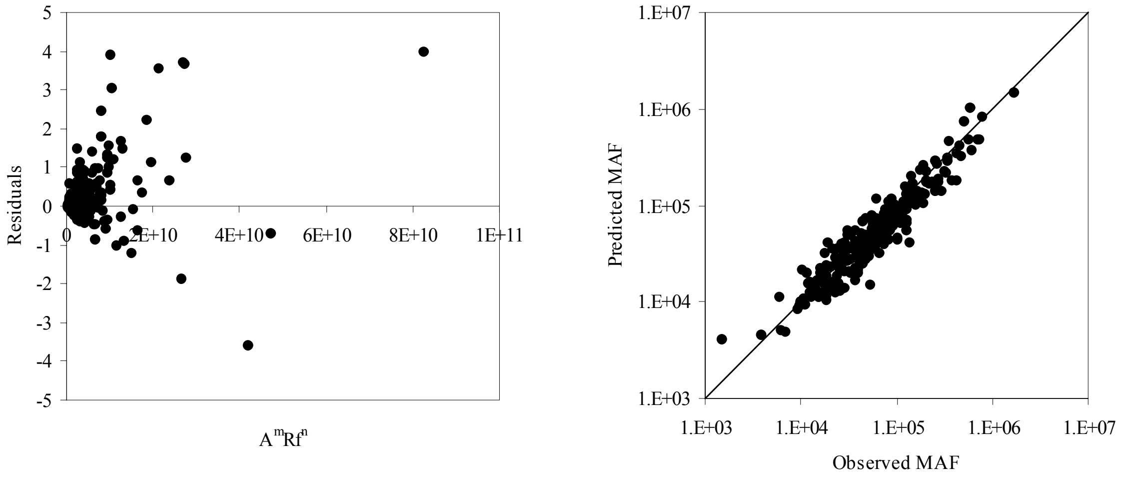


Figure 3: Region 1 linear regression model (a) standard residuals against $A^m Rf^n$ values, and (b) predicted against observed *MAF* values.

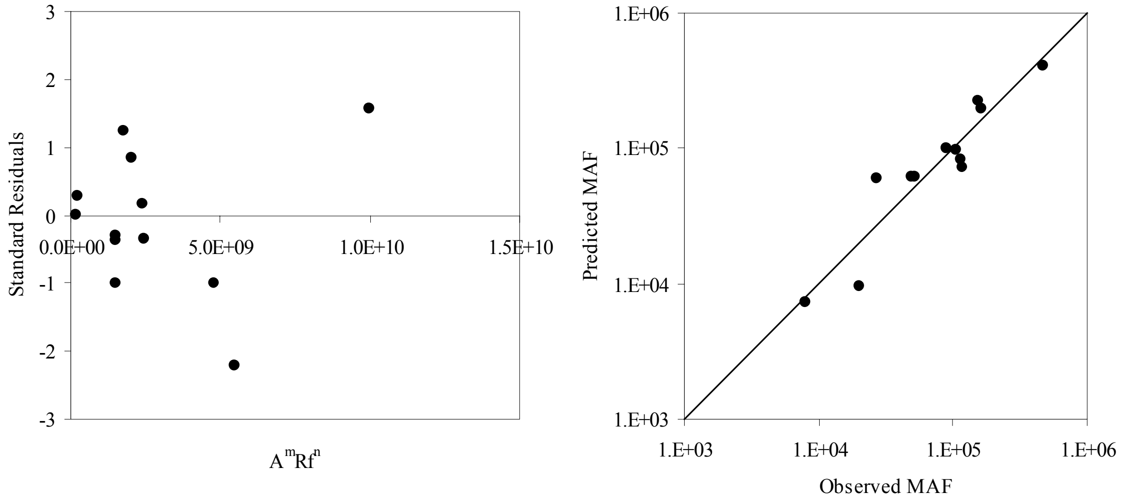


Figure 4: Region 2 linear regression model (a) standard residuals against $A^m Rf^n$ values, and (b) predicted against observed *MAF* values.

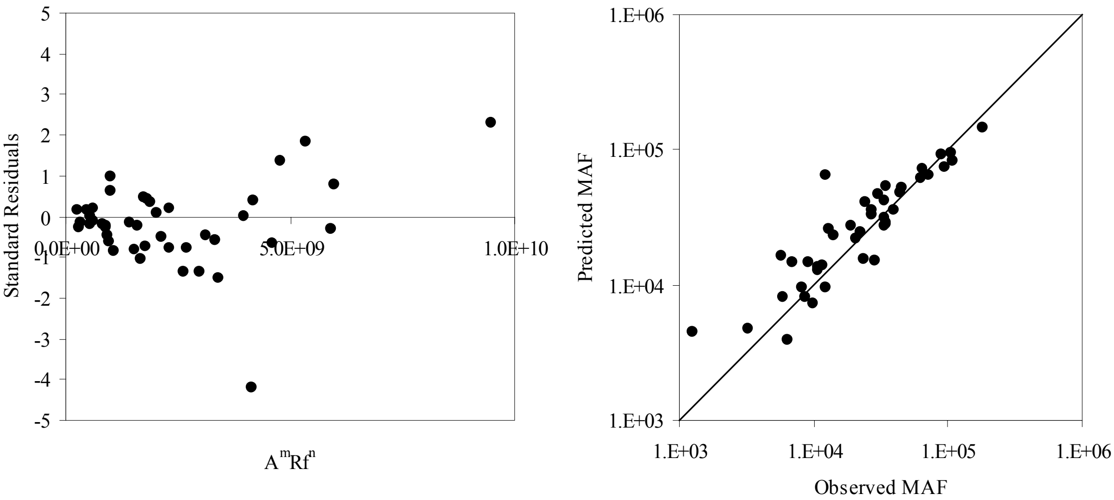


Figure 5: Region 3 linear regression model (a) standard residuals against $A^m Rf^n$ values, and (b) predicted against observed *MAF* values.

While the regional models all perform better than the general model and have satisfactory R^2 values, like the general model they all failed the constant variance test required for valid linear regression. The non-uniform variance is apparent on the plots of residuals. The Region 1 model also failed the test for normality, and the skewed nature of the data is apparent on Figure 3a. In spite of these short-comings, the time and data constraints of the National Land and Water Resources Audit meant these models were adopted for use in the sediment and nutrient transport modelling. The models do capture important aspects of the dependence of flow on catchment area and rainfall, and capture important regional differences in these relationships. For example, the smaller absolute value of $\log_{10}k$ for Region 2 (Tasmania) indicates greater *MAF* values than for Region 1 given similar rainfall and area values, and the greater absolute value of $\log_{10}k$ for Region 3 (semi-arid regions) indicates smaller *MAF* values than for Region 1 given similar rainfall and area values. The uncertainties in the regionalised hydrology models are small relative to other uncertainties in the sediment and nutrient transport models.

Mean Annual Flow in Large Catchments

The regionalisation models described above are applicable to catchment areas between 50 and 2000 km^2. In larger catchments the dependence of flow on catchment area changes, largely because of floodplain areas which with increasing area act more as loss pathways for water than as runoff source areas. To predict mean annual flows in river network links with catchment areas greater than 2000 km^2 an interpolation procedure using the estimates of river basin mean annual runoff and mean annual outflow collated by NLWRA (2001) were used.

Two categories of river basins were identified in this process; firstly those where the mean annual outflow is equal to the mean annual runoff, and secondly, those basins where the mean annual outflow is less than the mean annual runoff. For the first category, additional runoff is generated and added to the total flow beyond the 2000 km^2 limit to obtain the mean annual outflow. In the second category, additional runoff is first generated (if the total generated below 2000 km^2 is less than the basin mean annual runoff), and then water is lost to obtain the mean annual outflow. For the first category the following procedure was used in ARCVIEW GIS:

1. Define two groups of links within the river basin: those with upstream catchment area (UCA) greater than 2000 km^2 and those with UCA less than 2000 km^2.

2. Sum the MAF values of those links with UCA less than 2000 km^2 that intersect with links with UCA greater than 2000 km^2 or intersect with basin boundary.
3. Sum the internal link catchment areas of those links with UCA less than 2000 km^2.
4. Subtract the summed MAF from (2) above from the basin MAF.
5. Subtract the summed area from (3) above from the total basin area.
6. Divide value from (4) above by that from (5) above to give a mean annual runoff (MAR) for that area of basin with UCA greater than 2000 km^2.
7. Calculate "internal MAF" for all links with UCA greater than 2000 km^2 by multiplying their internal catchment area by the MAR from (6).
8. Sum MAF values through the network (adding values at network junctions) to give final MAF values for all links with UCA greater than 2000 km^2. The value for the most downstream link will then be equal to the basin MAF value used in (4).

For the river basins where the outflow is less than the mean annual runoff the following procedure was used in ARCVIEW GIS:

1. Manually associate the river basin MAF with the appropriate main channel link. This is a judgement of where the maximum MAF occurs on the main river before MAF values begin to decline towards the basin outflow value. This judgement was made for each basin from an a consideration of the basin and network topography.
2. Define three groups of links within the river basin: (a) those with UCA greater than 2000 km^2, (b) those with UCA less than 2000 km^2 that join the network upstream of basin-MAF link from (1), and (c) those with UCA less than 2000 km^2 that do not join the network upstream of basin-MAF link from (1).
3. For those links in group (b) in (2) above sum their MAF values.
4. For those links in group (b) from (2) above, sum their internal link catchment area values.
5. Subtract the summed MAF from (3) above from the basin MAF.
6. Subtract the summed area from (4) above from the UCA of the basin-MAF link in (1).
7. Divide value from (5) above by that from (6) above to give MAR for those parts of basin greater than 2000 km^2 and less that basin-MAF link UCA.
8. Calculate internal MAF for links greater than 2000 km^2 and upstream of basin-MAF link by multiplying their internal catchment area by the MAR from (7) above.
9. For those links in group (c) in (2) above sum their MAF values.
10. For those links in group c in (2) above sum their internal catchment areas.
11. Subtract value in (9) above from the basin mean annual outflow (MAO).
12. Calculate total mean annual loss (MAL) as the difference between basin-MAF and MAO.
13. Divide value from (12) above by the value from (10) above . This is the areal MAL for areas where UCA is greater than 2000 km^2 and not upstream of basin-MAF link.
14. Determine internal MAL for links with UCA greater than 2000 km^2 and not upstream of the basin-MAF link by multiplying their internal catchment area by the areal MAL from (13) above.
15. Sum MAF values through the network (adding values at junctions, and subtracting MAL values) to give final MAF values for all links with UCA greater than 2000 km^2. Value for most downstream link will then equal the basin MAO.

The procedures described above apply to river basins with a single outflow stream. In several basins there are multiple outflow streams all contributing to the MAO. In these cases the MAO was apportioned across the outflow streams on the basis of their respective catchment areas and the above procedures were then followed. Some river basins (mostly in the Murray-Darling Basin), are not strictly basins and have inflows from upstream as well as internally generated streamflow. In these basins the total inflow from contributing upstream basins was subtracted from the MAO value. If the result was positive, this meant a net addition of streamflow, and the first procedure described above was then used. If the result was negative, this meant a net loss of streamflow, and the second procedure described above was used.

Median Daily Flow

The median daily flow is used as a representative discharge in ANNEX to predict denitrification losses. Analysis of several observed and simulated flow records for perennial rivers indicated that the constant flow value that leads to the same average denitrification loss as the time series of daily flows is equivalent to approximately 1.2 times the median daily flow. Because the flow distributions for Australian rives are generally highly skewed, the median daily flow is a low flow in comparison to the mean daily flow. However, because flow distribution skewness is also very variable, the median daily flow is not strongly related to the mean daily flow. Models of the form of Equation 3 were investigated to predict median daily flow, as well as models based solely on the predicted *MAF*. Models produced by these two approaches were similar in performance, however, the latter type of model was selected because it allowed the interpolated MAF values determined for network links with UCA above 2000 km^2 to be used to predict median daily flow for these links. Although this is an extrapolation of the median daily flow models beyond the MAF range for which they were developed, models of the former type (functions of area and rainfall) did not provide any means to reliably estimate median daily flow for links with large catchment areas. As noted above, in large catchments the dependence of flow on area changes and there were few data available to establish the relationships for large catchments.

For many stations the median daily flow is zero. These stations had to be excluded from the data set as the logarithm cannot be determined. For the remaining data set of 266 stations the full model for median daily flow (with an adjusted R^2 of 0.522, and root mean square error E=142.5%) is:

$$\log_{10} medQ = 1.023 * \log_{10} Predcited\ MAF - 3.342 \qquad 5$$

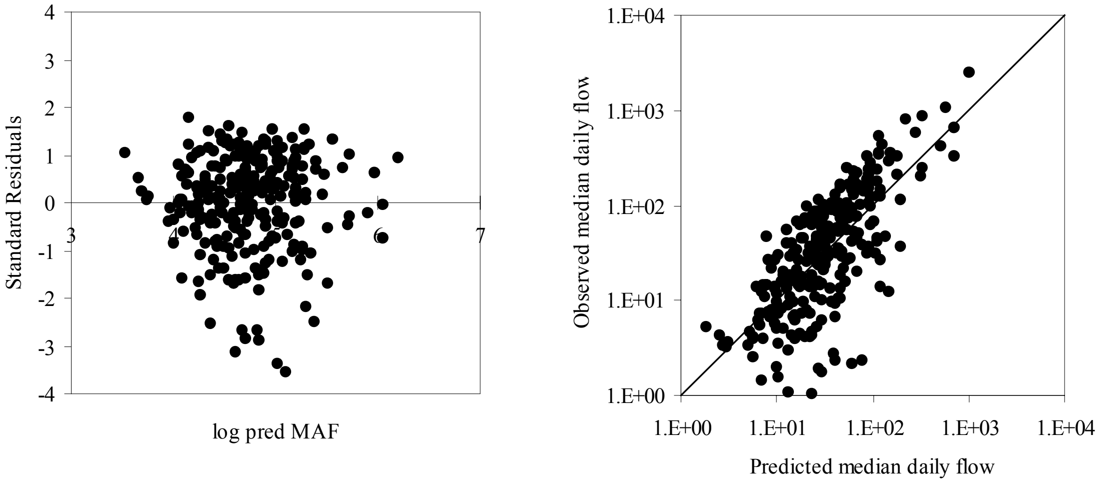


Figure 6: Full data set linear regression model (a) standard residuals against predicted $\log_{10}$*Predicted MAF* values, and (b) predicted median daily flow values against observed median daily flow values.

The model for median daily flow is poor compared to the mean annual flow model, and as noted above median daily flows are known to be poorly related to mean flows. The data passed the Kolmogorov-Smirnov test for normality (P=0.05) but failed the Spearman rank correlation test for constant variance (P=0.05).

Regional models for median daily flow (of the same form) were also investigated using the same three regions defined above. The constants and intercepts for these models differ significantly between regions suggesting important differences. The regional models perform better than the full model in Regions 2 and 3, however, the full model performs better than the regional model in Region 1. For the NWLRA projects the full model was used because this provides the best predictions for Region 1 that firstly, contains by far the majority of the observations in the data set (over 80%), and secondly, is the region containing by far the majority of the river network to which the model was being applied. In other applications the use of the regional models may be preferable.

Region	Constant	Intercept	Observations	Adjusted R^2	Model Performance – E(%)	
					Region Model	Full Model
1	1.021	-3.275	224	0.540	136.4	122.1
2	1.229	-4.230	12	0.889	42.2	51.1
3	0.856	-2.806	30	0.463	74.9	136.8

Table 3: Details of linear regression models of median daily flow for separate regions.

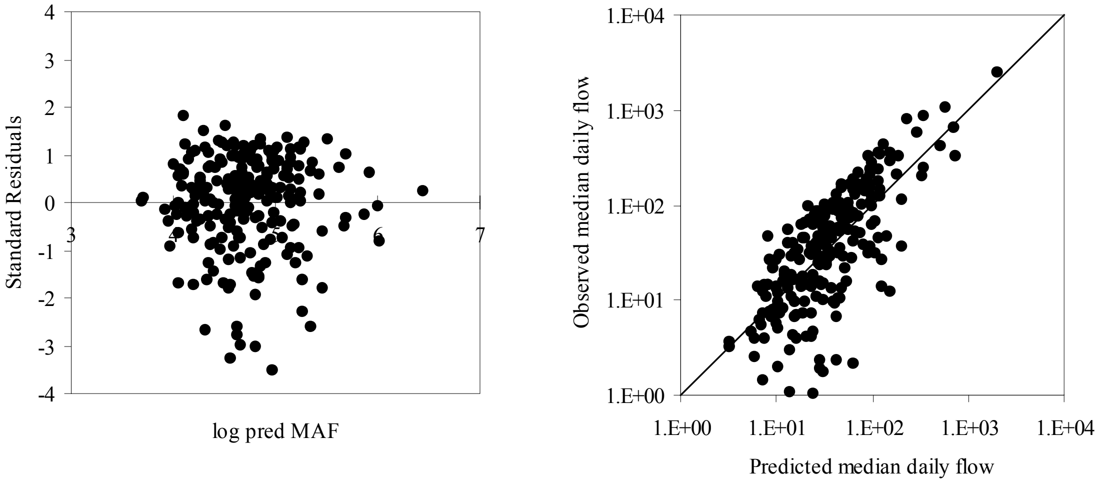
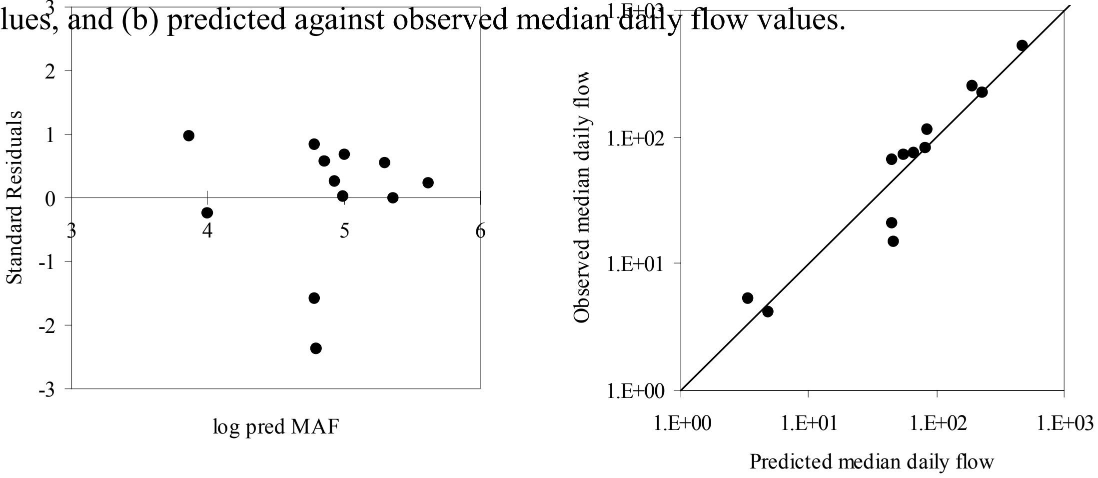


Figure 7: Region 1 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed median daily flow values.

Figure 8: Region 2 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed median daily flow values.



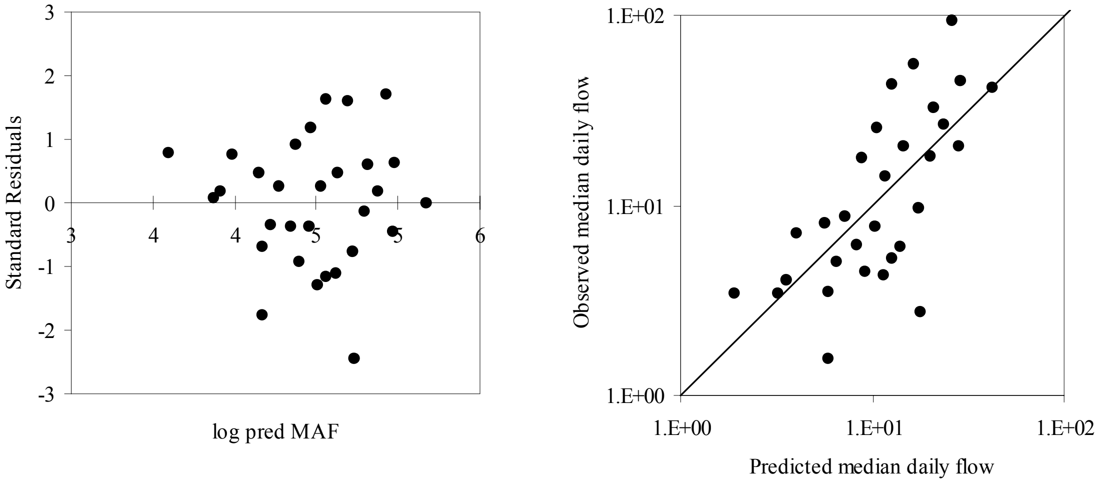


Figure 9: Region 3 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed median daily flow values .

The regional models for median daily flow all passed the test for constant variance (P=0.05), but only the Region 3 model passed the test for normality of the data (P=0.05). Clearly, the Region 2 model has the best fit (R^2) and lowest error (E), and this is expected because in Tasmania streamflow is more perennial that in many other parts of the country, meaning flow distributions are less skewed and the median daily flow is well related to the predicted mean annual flow. The use of the full model in the NLWRA project sacrificed greater accuracy in Regions 2 and 3 for greater accuracy in the larger Region 1.

Bankful Flow

Bankfull flow (*Qbf*) is used in SedNet as a predictor of bank erosion, and is used to predict the proportion of the suspended sediment load that is deposited on the floodplain. *Qbf* is assumed to be adequately represented by the mean annual flood that has a return period of 1.58 years on the annul series (Dury *et al.*, 1963). The mean annual flood was calculated for each simulated and observed flow record. As for median daily flow, models of the form of Equation 3 were investigated to predict *Qbf*, as well as models based solely on the predicted *MAF*. Models produced by these two approaches were similar in performance, however, the latter type of model was selected because it allowed the interpolated MAF values determined for network links with UCA above 2000 km^2 to be used to predict *Qbf* for these links.

Four stations were excluded from the data set because the observed record was too short to reliably determine the mean annual flood. For the remaining data set of 311 stations the full model for *Qbf* (with an adjusted R^2 of 0.590, and root mean square error E=85.6%) is:

$$\log_{10} Qbf = 1.009 * Predicted\ MAF - 1.421 \qquad 6$$

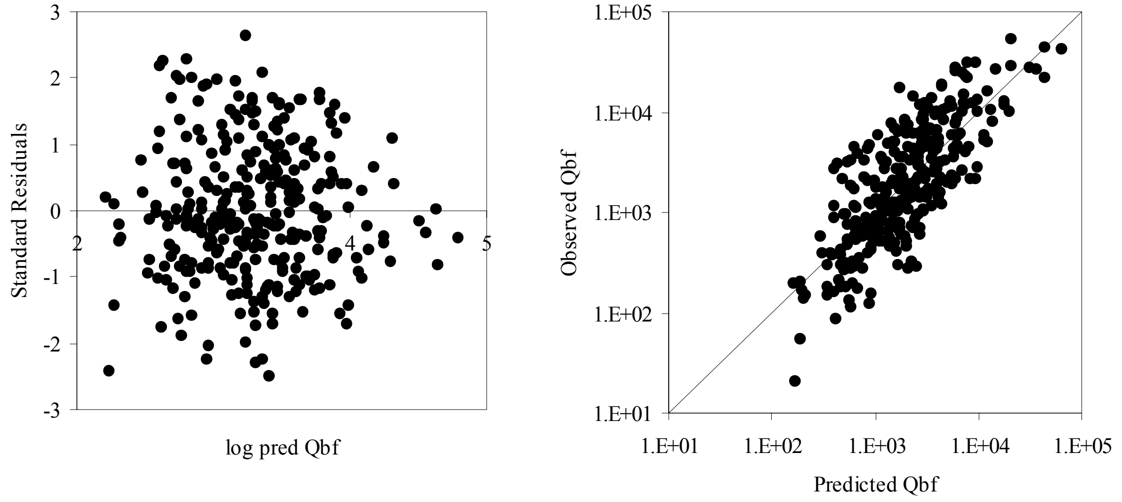


Figure 10: Full data set linear regression model (a) standard residuals against predicted $\log_{10}$*Predicted Qbf* values, and (b) predicted against observed *Qbf* values.

The model for *Qbf* is better than the model for median daily flows but worse than the mean annual flow model. The data passed the Kolmogorov-Smirnov test for normality (P=0.05) and the Spearman rank correlation test for constant variance (P=0.05).

Regional models for *Qbf* (of the same form) were also investigated using the regions defined earlier. The constants and intercepts for these models do differ significantly between the three regions suggesting important differences. The regional models perform better than the full model in Regions 2 and 3, however, the full model performs better in Region 1 than the regional model. For the NWLRA projects the full model was used because this provides the best predictions for Region 1 that firstly, contains by far the majority of the observations in the data set (over 80%), and secondly, is the region containing by far the majority of the river network to which the model was being applied. In other applications the use of the regional models may be preferable.

Region	Constant	Intercept	Observations	Adjusted R^2	Model Performance – E(%)	
					Region Model	Full Model
1	1.009	-1.395	255	0.581	87.2	81.7
2	0.455	1.205	12	0.141	120.1	159.3
3	1.018	-1.585	44	0.608	62.2	88.2

Table 4: Details of linear regression models of *Qbf* for separate regions.

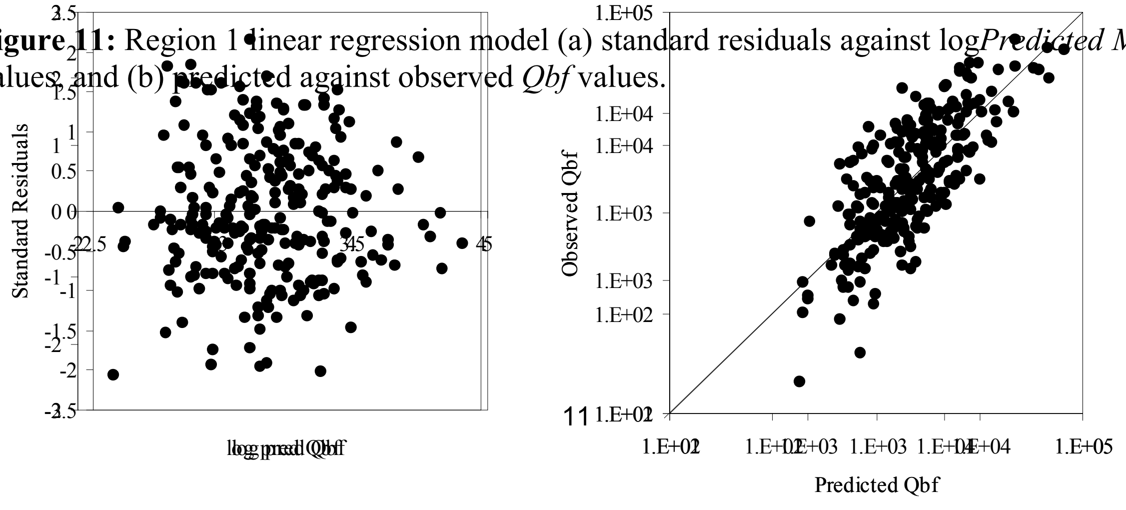


Figure 11: Region 1 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *Qbf* values.

Figure 12: Region 2 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *Qbf* values.

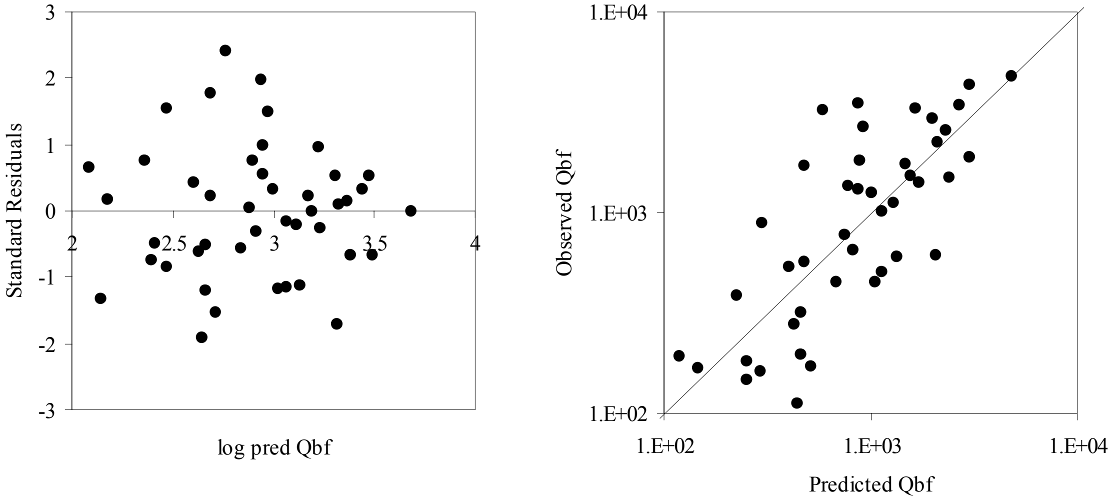


Figure 13: Region 3 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed Qbf values.

The regional models for *Qbf* all passed the test for constant variance (P=0.05), and only the Region 3 model failed the test for normality of the data (P=0.05). Note that in Tasmania (Region 2) *Qbf* is very poorly predicted by the predicted mean annual flow. The use of the full model in the NLWRA project sacrificed greater accuracy in Regions 2 and 3 for greater accuracy in the larger Region 1.

Median Overbank Flow

The median overbank flow (*Qob*) is used in SedNet as a representative discharge for estimating the residence time of flood water on the floodplain. The residence time is used with particle settling theory to determine proportion of the suspended load that is deposited on the floodplain. Regression models were developed to predict the total median overbank flow – that is, the median value of all daily flows that are higher than *Qbf*.

As for previous variables, models of the form of Equation 3 were investigated to predict *Qob*, as well as models based solely on the predicted *MAF*. Models produced by these approaches were similar in performance, however, the latter type of model was selected because it allowed the interpolated MAF values determined for network links with UCA above 2000 km^2 to be used to predict *Qob* for these links. For the data set of 311 stations (as used for *Qbf*) the full model for *Qob* (with an adjusted R^2 of 0.554, and root mean square error E=102.5%) is:

$$\log_{10}Qob = 1.010 * \log_{10}Predicted\ MAF - 1.198 \qquad 7$$

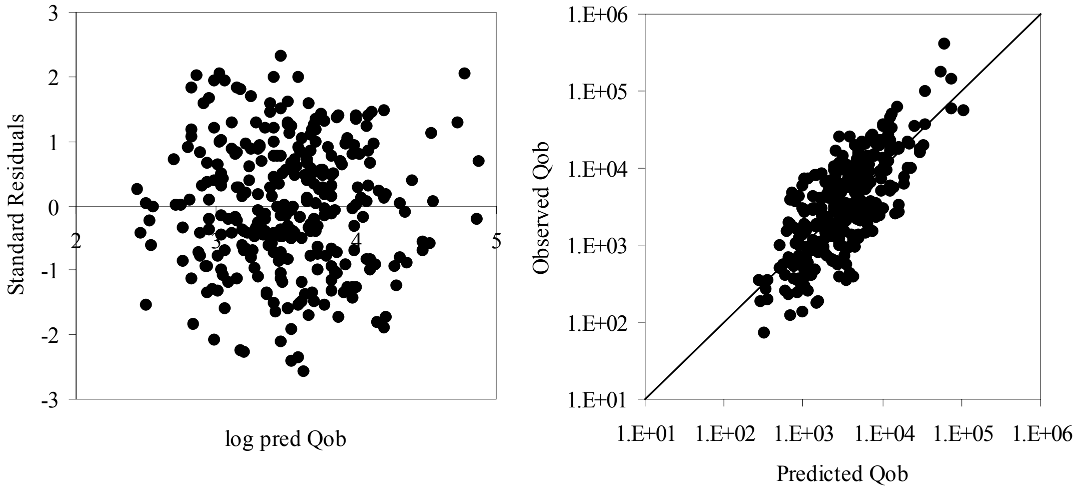


Figure 14: Full data set linear regression model (a) standard residuals against predicted $\log_{10}$*Predicted Qob* values, and (b) predicted *Qob* values against observed *Qob* values.

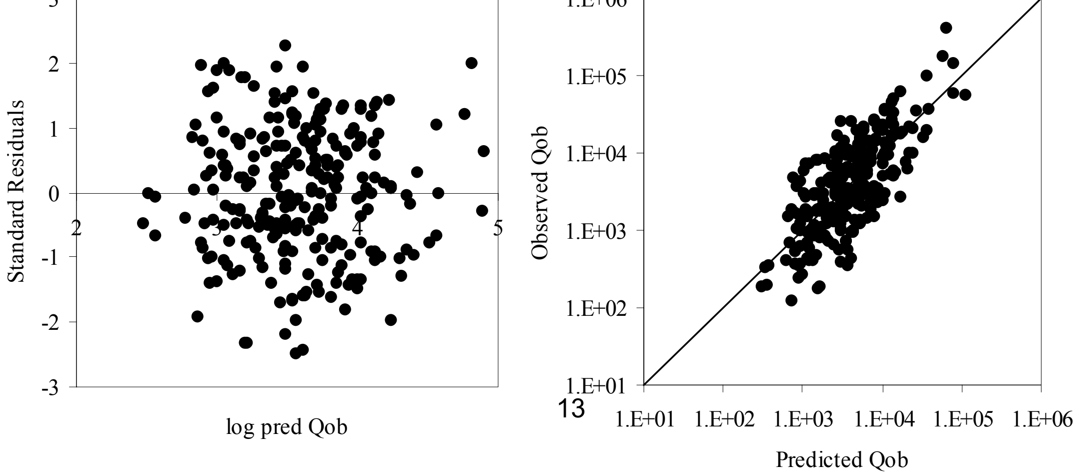
The model for *Qob* is better than the model for median daily flows but worse than the mean annual flow model and the *Qbf* model. The data passed the Kolmogorov-Smirnov test for normality (P=0.05) and the Spearman rank correlation test for constant variance (P=0.05).

Regional models for *Qob* (of the same form) were also investigated using the regions defined earlier. The constants and intercepts for these models do differ significantly between the three regions suggesting important differences. The regional models perform better than the full model in Regions 2 and 3, however, the full model performs better in Region 1 than the regional model. For the NWLRA projects it was decided to use the full model because this provides the best predictions for Region 1 that firstly, contains by far the majority of the observations in the data set (over 80%), and secondly, is the region containing by far the majority of the river network to which the model was being applied. In other applications the use of the regional models may be preferable.

Region	Constant	Intercept	Observations	Adjusted R^2	Model Performance – E(%)	
					Region Model	Full Model
1	1.015	-1.198	255	0.551	101.6	96.1
2	0.405	1.597	12	0.079	77.2	219.1
3	1.054	-1.498	44	0.547	81.4	107.9

Table 5: Details of linear regression models of *Qob* for separate regions.

Figure 15: Region 1 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *Qob* values.



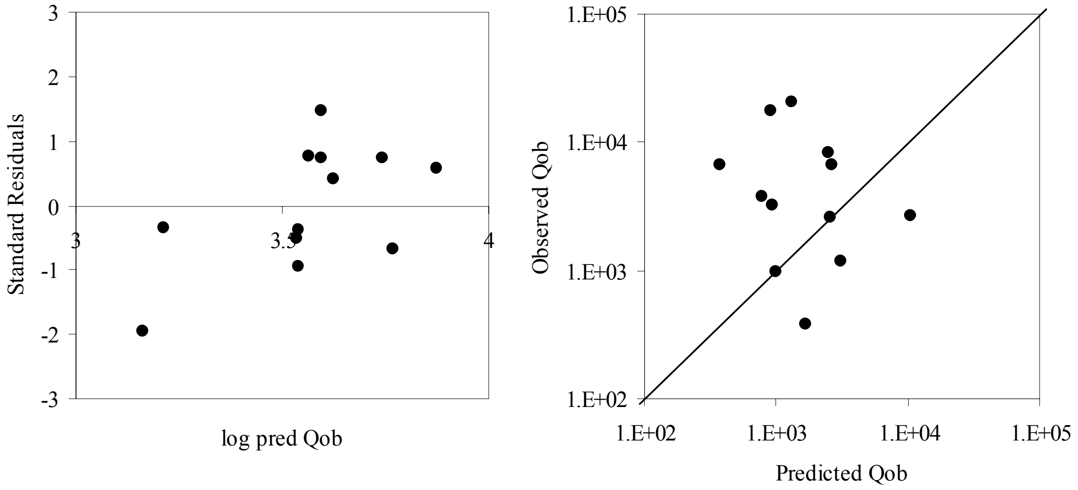


Figure 16: Region 2 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *Qob* values.

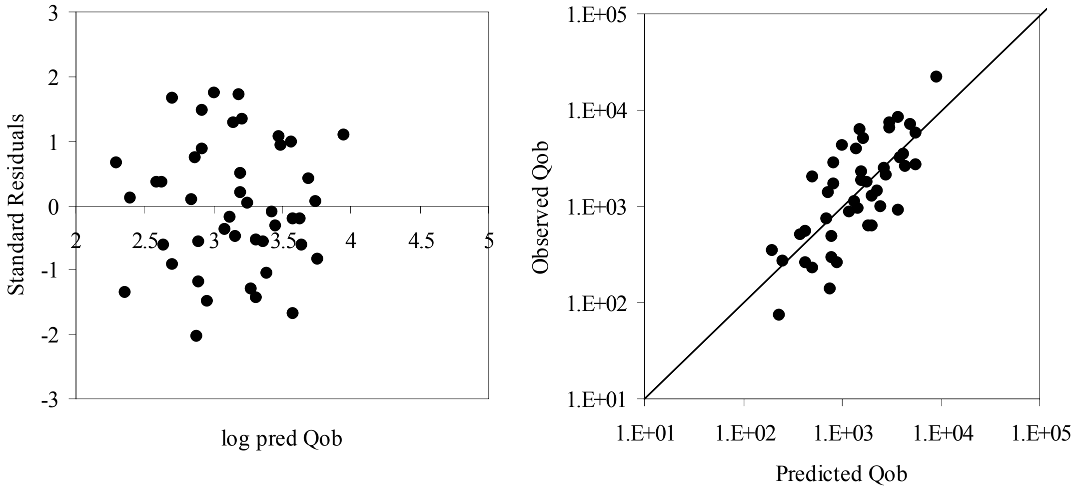


Figure 17: Region 3 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *Qob* values.

The regional models for *Qob* all passed the test for constant variance ($P=0.05$) and the test for normality of the data ($P=0.05$). Note that in Tasmania (Region 2) there is very little relationship between *Qob* and the predicted mean annual flow. The use of the full model in the NLWRA project sacrificed greater accuracy in Regions 2 and 3 for greater accuracy in the larger Region 1.

Sediment Transport Capacity Discharge

As described earlier, SedNet uses a parameterised function of daily flows (*sigQ*) as a predictor of coarse sediment transport capacity. As for previous variables, models of the form of Equation 3 were investigated to predict *sigQ*, as well as models based solely on the predicted *MAF*. Models produced by these two approaches were similar in performance, however, the latter type of model was selected because it allowed the interpolated MAF values determined for network links with UCA above 2000 km^2 to be used to predict *sigQ* for these links. *sigQ* was calculated for the full data set of 314 stations, and the full model for *sigQ* (with an adjusted R^2 of 0.807, and root mean square error $E=70.8\%$) is:

$$\log_{10} sigQ = 1.427 * \log_{10} Predicted\ MAF - 1.006 \qquad 8$$

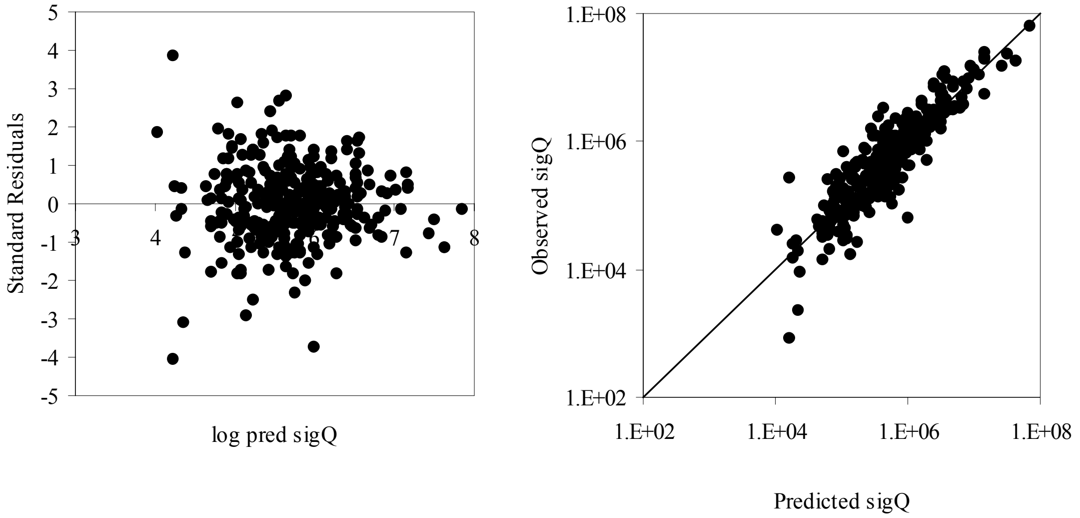


Figure 18: Full data set linear regression model (a) standard residuals against predicted $\log_{10}$*Predicted sigQ* values, and (b) predicted *sigQ* values against observed *sigQ* values.

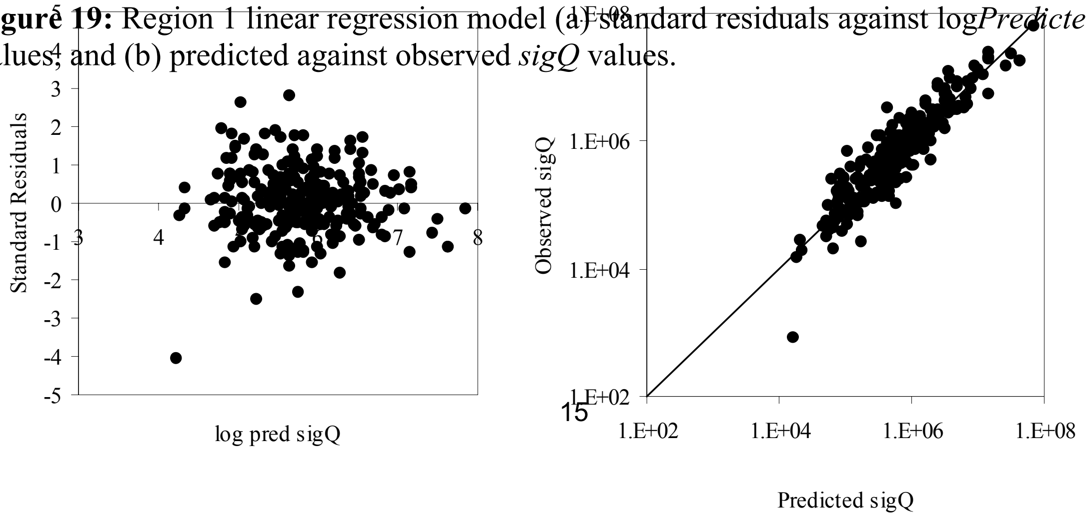
The model for *sigQ* is the best of the models based on the predicted *MAF*. The data passed the Kolmogorov-Smirnov test for normality (P=0.05), but failed the Spearman rank correlation test for constant variance (P=0.05).

Regional models for *sigQ* (of the same form) were also investigated using the regions defined earlier. The constants and intercepts for these models do differ significantly between the three regions suggesting important differences. Only the model for Region 3 performs better than the full model. For the NWLRA projects the full model was used because this provides the best predictions for Regions 1 and 2, that is for the majority of the modelled rivers.

Region	Constant	Intercept	Observations	Adjusted R^2	Model Performance – E(%)	
					Region Model	Full Model
1	1.413	-0.919	258	0.844	58.6	55.9
2	1.121	0.870	12	0.714	78.7	70.0
3	1.444	-1.331	44	0.726	77.8	159.2

Table 6: Details of linear regression models of *sigQ* for separate regions.

Figure 19: Region 1 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *sigQ* values.



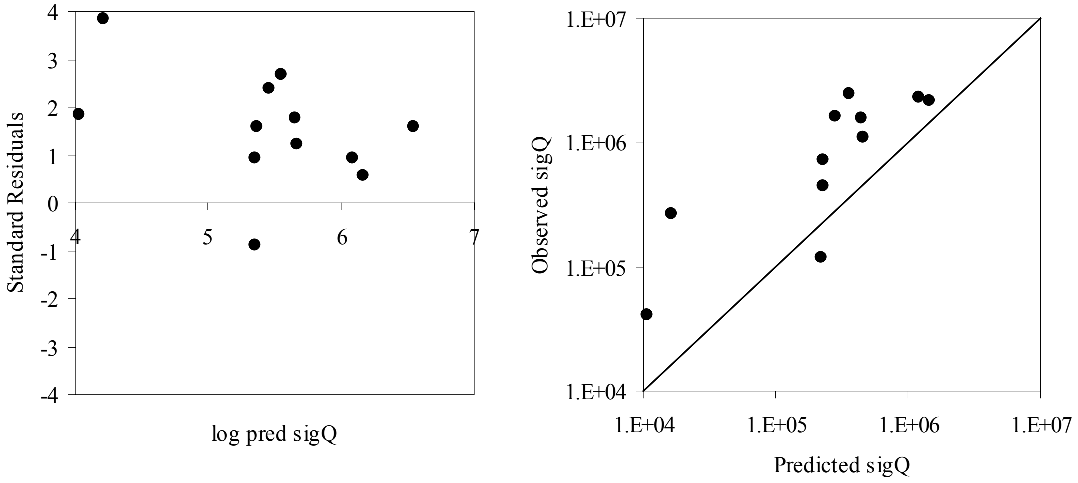


Figure 20: Region 2 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *sigQ* values

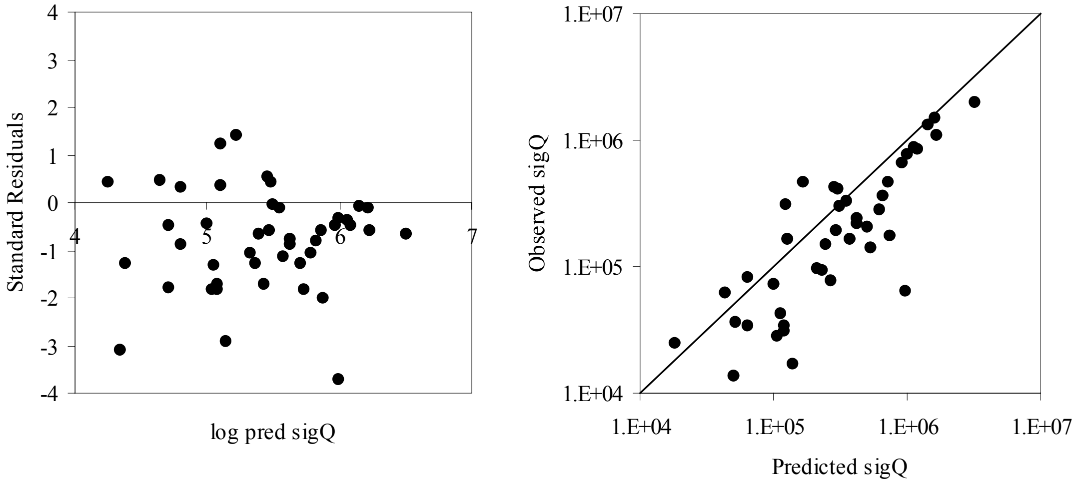


Figure 21: Region 1 linear regression model (a) standard residuals against log*Predicted MAF* values, and (b) predicted against observed *sigQ* values

The regional models for *sigQ* all passed the test for constant variance (P=0.05) and only the Region 3 model failed the test for normality of the data (P=0.05).

FLOW VARIABLES IN REGULATED RIVERS

The models for flow variables described above are all based on simulated or observed unregulated flows. In many rivers flows are altered by regulation and diversion, and consequently current flows cannot be reliably predicted simply as functions of catchment area and rainfall. As a basis for describing regulated flow regimes, values of the required flow variables were calculated for 242 stations on regulated rivers across Australia; for 202 stations post-regulation observed flow records were used, and for 40 stations modelled current daily flows were used (Appendix C). For *MAF*, the ratio of regulated to natural flow was calculated for each regulated station, and interpolation and extrapolation rules (based on contributing catchment area) were developed (Norris *et al.*, 2001) to provide estimates of the ratio of regulated *MAF* to natural *MAF* for network links between or downstream of links for which regulated data were available. These interpolated and extrapolated ratio values were multiplied by the predicted natural *MAF* values for ach link (from the regionalisation models and basin extrapolations) to give estimates of the regulated *MAF* values. The estimates of regulated *MAF* were used to determine deposition in reservoirs, to determine the adsorption-desorption of phosphorus to and from suspended sediment, and to determine dissolved phosphorus loads based on modelled concentrations.

Project time constraints did not permit the interpolation and extrapolation of other hydrologic variables, and hence regulated values of *Qbf*, *Qob*, *medQ* and *sigQ* were not used in the network modelling. Flows are regulated in only 11% of the river network across Australia that was considered in the modelling (Norris *et al.*, 2001). Regulated MAF values were determined for 54% of this regulated length. In many unregulated rivers flows are affected by extractions. Data were not available to assess the impact of these extractions on MAF or the other hydrologic variables used in the network modelling. River network modelling therefore only included a partial representation of the affects of flow regulation and diversion, both in terms of the range of hydrologic variables and the spatial coverage of data.

REFERENCES

Dury, G.H., Hails, J.R. and Robbie, M.B. 1963. Bankful discharge and the magnitude-frequency series. *Australian Journal of Science*, **26**: 123-124.

Peel, M.C., Chiew, F.H.S., Western, A.W. and McMahon, T.A. 2000. Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Report to the National Land and Water Resources Audit.

Prosser, I.P. and Rustomji, P.2000. Sediment transport relations for overland flow. *Progress in Physical Geography*, **24**, 179-193.

Prosser, I, Rustomji, P., Young, W., Moran, C. Hughes, A. 2001. Constructing river basin sediment budgets for the National Land and Water Resources Audit. CSIRO Land and Water Technical Report 15/01, CSIRO Land and Water, Canberra. 36pp.

NLWRA 2001. Australian Water Resources Assessment 2000. Surface water and groundwater – availability and quality. National Land and Water Resources Audit, Canberra. Commonwealth of Australia. 160pp.

Norris, R.H., Prosser, I.P. Young, W.J., Liston, P., Bauer, N., Davies, N., Dyer, F., Linke, S. and Thoms' M.C. 2001. The assessment of river condition (ARC): an audit of the ecological condition of Australian rivers. Project report from the CRC for Freshwater Ecology and CSIRO Land and Water to the National Land and Water Resources Audit, Canberra.

Young, W.J., Prosser, I.P. and Hughes, A.O. 2001. Modelling nutrient loads in large-scale river networks for the National Land and Water Resources Audit. CSIRO Land and Water Technical Report 12/01, CSIRO Land and Water, Canberra. 14pp.

APPENDICES

Details of the gauging station listed in the Appendices can be obtained from http://www.bom.gov.au/hydro/wr/sgc/index.shtml.

Appendix A: The 282 gauging stations with 100 years of simulated daily flow (Peel *et al.*, 2000) that were used in the development of regionalisation models, listed by AWRC Drainage Division.

North East Coast	South East South			Tasmania	Murray-Darling Basin		South Australian Gulf	South West Coast	Timor Sea
110003	201001	209001	223202	302200	401008	411003	502502	603004	814008
111007	201005	209002	223207	302208	401009	412063	505504	603136	814159
112003	201011	209006	224201	303203	401012	412066	505517	604001	820045
117002	201900	210014	224207	304201	401013	412072	506500	606001	821007
119003	203002	210017	224209	312061	401015	412082	507500	608151	
120216	203005	210022	225213	318076	401016	412089	513501	610001	
121002	203010	210042	225217	318311	401203	412096		611111	
125002	203030	210061	225218	318852	401210	412110		613002	
129001	204016	210080	225219	318900	401212	415207		614196	
130319	204019	210081	226007	319200	402200	416008			
132001	204025	210088	226204	319201	402204	416020			
135002	204026	211008	226218	319204	402206	416021			
136101	204030	211013	226405		402406	416022			
136202	204031	211014	226406		403205	416023			
142001	204033	212021	226410		403206	416035			
143110	204034	212028	227200		403213	418005			
145011	204036	212040	227202		403214	418015			
145102	204037	212045	227211		403217	418017			
	204041	215002	227219		403224	418021			
	204055	215004	228203		403226	418024			
	204067	215005	228212		404208	418027			
	205002	215008	229215		405205	418032			
	205006	216004	230205		405209	418033			
	205008	218001	231213		405214	419010			
	205012	218002	233215		405219	419029			
	205014	218006	233223		405226	419035			
	206009	218007	234200		405228	419044			
	206014	219013	234203		405229	419047			
	206018	219016	235203		405237	419050			
	206020	219017	235211		406213	419054			
	206025	220002	236203		406214	419055			
	206034	220003	236205		407220	419072			
	207006	220004	236212		407221	419076			
	207012	221002	237200		407253	420003			
	207013	221003	237205		408202	421018			
	207014	221010	237206		410038	421026			
	207015	221201	238223		410047	421036			
	208002	221204	239519		410048	421048			
	208005	221210			410057	421050			
	208006	222001			410059	421055			
	208007	222004			410061	421066			
	208008	222009			410067	421100			
	208009	222010			410071	421101			

	208012	222011			410077	421104			
	208015	222014			410096	421106			
	208019	222015			410105	421125			
	208022	222017			410111	426504			
	208026	222202			410705				
		222206			410730				
		222213			410733				

Appendix B: The 32 gauging stations from which observed records were used in the development of regionalisation models indicating start and end year of record.

Station	From	To	Station	From	To
105001	1958	2000	136203	1940	2000
107001	1958	2000	136301	1935	3000
111001	1916	2000	137001	1948	2000
111101	1918	2000	138005	1950	2000
112001	1928	1969	143010	1965	1999
112101	1916	2000	143011	1965	1986
116010	1960	2001	143107	1961	1999
116011	1960	2001	143108	1961	1999
120014	1970	1999	143204	1954	1977
120102	1967	1998	145008	1958	1999
120106	1967	1998	145012	1966	1999
120112	1967	1998	145018	1971	1998
120120	1975	1999	145020	1973	1999
120207	1962	1998			
120304	1967	1998			
120307	1967	1998			
121001	1957	2001			
122003	1956	2000			
125001	1916	2000			

Appendix C: The gauging stations on regulated rivers for which flow variables were calculated from either recorded or modelled daily flows.

Stations used for observed regulated flows				Stations used for modelled regulated flows
110001	201900	401202	412002	212011
110002	206004	401204	412003	212045
113006	206011	401211	412005	212250
120008	206024	403200	412006	412004
120205	210001	403220	412011	412026
120207	210002	403223	412036	412048
120209	210010	403228	412039	416001
120299	210015	403230	412045	416006
130001	210021	403240	412078	416011
130002	210044	404203	415203	416012
130003	210079	404206	419001	416014
130005	210083	404217	419006	416018
130103	212008	405200	419007	416019
130105	212019	405201	419012	416028
130106	212039	405202	419015	416038
130205	212019	405203	419020	416049
130206	212039	405204	419021	416201

130208	212045	405232	419022	416402
136001	212201	405276	419023	416415
136002	212203	405277	419024	418001
136003	212204	406200	419039	418002
136004	212208	406201	419043	418004
136206	212210	406202	419045	418011
136207	212231	406215	419053	418013
138001	212233	406225	419059	418031
138007	212241	407202	419070	418042
138105	212270	407203	421015	418044
138109	215200	407205	421017	418048
138111	222219	407210	421019	418052
138112	225112	407222	421025	418055
141004	225204	407224	421040	418056
143001	225208	407227	421079	418057
143108	225210	407229	421127	418060
145008	225212	407248	421147	418066
145014	225231	410001	503500	421001
145020	225232	410002	504501	421012
145099	226005	410004	507503	421090
146002	226006	410005	612003	422009
	226216	410006	612006	425002
	226227	410008	613003	425030
	226228	410021	613052	
	228201	410036	614006	
	228213	410039	614030	
	228219	410050	614072	
	229200	410073	614224	
	229212	410130	616086	
	229216	410700	919011	
	230202	410725	919309	
	231200	410729	919310	
	231203	410747	919311	
	231205	410750		
	231232	410752		
	232202	410756		
	232204	410761		
	232211	410770		
	238206	410777		
	238212			
	238224			