**Hydrology and
Earth System
Sciences
Discussions**

**Characteristics and
drivers of baseflow
response**

A. I. J. M. van Dijk

Characteristics and drivers of baseflow response in 183 Australian catchments

A. I. J. M. van Dijk

CSIRO Land and Water, Canberra, ACT, Australia

Correspondence to: A. I. J. M. van Dijk (albert.vandijk@csiro.au)

A. I. J. M. van Dijk

Abstract

Daily streamflow data for 183 Australian catchments were used to assess the characteristics and main drivers of baseflow and quick flow behaviour, and to find an appropriate balance between simplicity and explanatory performance in modelling. Baseflow separation was performed following the Wittenberg algorithm. A linear reservoir model (one parameter) produced baseflow estimates as good as those obtained using a non-linear reservoir (two parameters) and was therefore considered the more appropriate. The transition from storm flow dominated to baseflow dominated streamflow generally occurred 7 to 10 d after the storm event. The catchments investigated had baseflow half-times of about 12 d, with 80% of stations having half-times between 7 and 34 d. The shortest half-times occurred in the driest catchments and were attributed to intermittent occurrence of fast-draining (possibly perched) groundwater. Median baseflow index (BFI) was 0.45 with considerable variation between stations. Catchment humidity explained 27% of the variation in derived baseflow recession coefficients. Another 53% of variance in recession coefficients as well as in BFI showed spatial correlation lengths of 200 to 300 km, corresponding to terrain factors rather than climate or land use. The remaining 16 to 20% of variance remained unexplained. Most (84%) of the variation between stations in average baseflow could be explained by monthly precipitation in excess of potential evapotranspiration. Most (70%) of the variation in average quick flow could be explained by average rainfall. Another 20% of variation was spatially correlated over spatial scales of 400 km, possibly reflecting a combination of terrain and climate factors; the remaining 10 to 16% remained unexplained.

1 Introduction

Baseflow recession is the primary source of streamflow during periods without large precipitation events. Approaches to simulate baseflow recession vary from single linear or non-linear stores to cascading or parallel groundwater stores (e.g. Bergström,

1992; Burnash et al., 1973; Chiew et al., 2002; Jakeman and Hornberger, 1993). Increasing the number of stores or parameters to describe baseflow increases the flexibility to match observed baseflow recession patterns, but also increases the likelihood of parameter equivalence. Commonly the only data available to calibrate and evaluate
5 baseflow recession models is catchment streamflow. Therefore it is worthwhile to assess what level of complexity is appropriate for simulating baseflow recession, in other words: how many model parameters do streamflow observations justify?

To assess baseflow patterns, it is helpful to decompose total streamflow into its components storm flow or quick flow (QF) and baseflow (BF). Baseflow is usually as-
10 sumed to originate from the groundwater store and the terms groundwater discharge and baseflow are used interchangeably. The remaining QF is interpreted to represent other, faster streamflow pathways, including infiltration excess and saturation overland flow, and unsaturated or saturated (perched) interflow. It is noted that this is a conceptual interpretation of the two flow components and that the hydrograph per se does not
15 provide any evidence that this is a correct interpretation.

In the current study, a recursive baseflow filter based on Wittenberg (1999) was used to estimate BF and QF components and parameters of baseflow recession. Linear as well as non-linear reservoir equations were fitted, producing a baseflow recession coefficient (k_{BF}) and in the latter case an exponent (β). The balance between model
20 complexity and explanatory performance was assessed from the relative improvement in the ability to explain observed patterns by introducing further model parameters.

In particular, we attempted to answer the following questions:

– Is baseflow recession most parsimoniously described by a linear or by a non-linear reservoir equation?

25 – To what extent can differences in average baseflow, quick flow and the baseflow recession coefficient between catchments be related to catchment attributes describing morphology, soils, climate or land cover?

– To what extent are unexplained patterns in these variables spatially correlated,

and what are likely underlying factors?

The overarching goal of this analysis is to assess the main drivers of catchment base-flow and quick flow behaviour, and to define a modelling approach that represents an appropriate balance between simplicity and explanatory value.

2 Theory

2.1 Overall approach

The method to separate daily streamflow data (Q in mm d^{-1}) into baseflow (Q_{BF}) and quick flow (Q_{QF}) components requires a recession coefficient (k_{BF}) if a linear reservoir is assumed, and if a non-linear reservoir is assumed an additional exponent β. Both are described by:

$$Q_{BF} = k_{BF}S^{\beta} \tag{1}$$

where S (mm) is reservoir storage and β is unity if a linear reservoir is used.

It is assumed that quick flow only measurably affects streamflow during a period of T_{QF} days after the event peak flow, the length of which needs to be estimated in advance. Choosing T_{QF} too long reduces the amount of data and can lead to a bias in the results when baseflow behaviour is non-linear, whereas choosing the period too short introduces bias in the parameter estimates and subsequent streamflow separation due to the influence of QF on recession. Prior analysis (E. Kwantes, unpublished) suggested that ten days is a reasonable compromise and that derived baseflow parameters are reasonably stable in this range. This is obviously a necessary simplification however, and it will be revisited further on.

For the analysis, all days showing an increase in Q from the previous day were considered to mark the start of a quick flow event. All these days as well as the T_{QF} days afterwards each of these events were excluded from the analysis. All days with

zero flow or missing data were also excluded. From the remaining values, data pairs of Q and the antecedent streamflow Q_* from the previous day were constructed.

2.2 Linear reservoir

In the special case of a linear reservoir, $\beta=1$ and this leads to the following set of equations:

$$Q_{BF}(t-1) = k_{BF}S(t-1) \tag{2}$$

$$Q_{BF}(t) = k_{BF}S(t) = k_{BF}[S(t-1) - Q_{BF}(t-1)] \tag{3}$$

Provided that total streamflow $Q=Q_{BF}(t)$ and $Q_*=Q_{BF}(t-1)$, these two equations can be combined to yield:

$$Q = (1 - k_{BF})Q_* \tag{4}$$

and also

$$k_{BF} = 1 - \frac{Q_{BF}(t)}{Q_{BF}(t-1)} = 1 - \frac{Q}{Q_*} \tag{5}$$

2.3 Non-linear reservoir

To estimate reservoir parameters the method of (Wittenberg, 1999) was followed, which describes the storage-discharge relationship as:

$$S = aQ^b \tag{6}$$

Where the parameters expressed in terms of Eq. (1) are:

$$\beta = \frac{1}{b} \text{ and } k_{BF} = a^{-\frac{1}{b}} \tag{7}$$

Coutange (1948) demonstrated that for a baseflow recession:

$$Q = Q_* \left[1 + \frac{1-b}{ab} Q_*^{1-b} \right]^{\frac{1}{b-1}} \qquad (8)$$

Thus for a given value of b and for each Q–Q_* pair, a can be calculated as:

$$a = \frac{1-b}{b} Q_*^{1-b} \left[\left(\frac{Q}{Q_*} \right)^{b-1} - 1 \right]^{-1} \qquad (9)$$

3 Methods

3.1 Data

Daily streamflow data (all expressed in $ML\,d^{-1}$) were collated for 260 catchments across Australia as part of previous studies (Guerschman et al., 2008; Peel et al., 2000). For these catchment, streamflow data was considered of satisfactory quality and any influence of river regulation, water extraction, urban development, or other processes upstream streamflow were considered unimportant. The contributing catchments of all gauges were delineated through digital elevation model analysis and visual quality control (see Appendix A). Catchment areas vary between 51–1979 km^2, with a median value of 315 km^2. The range of average annual rainfall for catchments in the sample is 317–1887 $mm\,y^{-1}$, whereas Priestley-Taylor potential evapotranspiration (E_0) varies from 765–2417 $mm\,y^{-1}$ and streamflow 2–979 $mm\,y^{-1}$.

Out of the overall data set, streamflow data for the period 1990–2006 were selected for 226 gauge records that had good quality observations for at least five consecutive years with less than 20% of data missing, and no less than 50 runoff events (defined as an increase in streamflow from one day to the next). The streamflow data were converted to areal average streamflow (Q, $mm\,d^{-1}$).

3.2 Parameter estimation

A number of methods can be used to estimate k_{BF} from the data pairs, depending on weightings that one wants to apply. Some examples applicable to a linear reservoir model include (1) linear regression (through the origin) on the original data (see Fig. 2, in that case yielding k_{BF}=0.056); (2) regression on log-transformed data (for the same case producing k_{BF}=0.129); (3) the ratio of average Q and Q_* values (producing k_{BF}=0.056); (4) the ratio of the average of log-transformed Q and Q_* values (producing k_{BF}=0.129); or (5) the average or median of all Q/Q_* ratios (producing k_{BF}=0.116 and k_{BF}=0.074, respectively).

Tests showed that the parameter values derived from log-transformed values were least affected by the size of the window T_{QF} used and regression on the original data most. The data shown in Fig. 3 suggest that individual Q/Q_* ratios for a station are not normally distributed, and the same was found for other gauges. The simple mean may therefore not be the most appropriate estimate.

To avoid over weighting on either larger values or (through log-transformation) smaller values, the value of k_{BF} was optimised rather than directly inferred. Another reason for this approach is that it provided a method for fitting parameters of the non-linear reservoir model, for which it is less straightforward to directly estimate k_{BF} and β. In both cases, parameters were fitted to minimise the mean relative error (ε_{MRE}), expressed as:

$$\varepsilon_{MRE} = \frac{1}{n} \sum |\frac{Q}{Q_{est}} - 1| \tag{10}$$

where Q_{est} estimated from Q_* in the data pairs, given for a linear reservoir by Eq. (4) and for a non-linear reservoir by Eq. (8). For the example data set this produced k_{BF}=0.056 for the linear reservoir model (Fig. 3).

3.3 Decision on reservoir model structure

To decide on the optimal balance between the number of fitting parameters and explained variation in observations, a version of Akaike's Final Prediction Error Criterion (FPEC; Akaike, 1970) was calculated and interpreted. FPEC estimates the prediction error if the model was tested on a different data set and therefore the most accurate model would have the smallest FPEC. FPEC can be expressed as the product of an empirically estimated prediction error and a penalization factor that considers the degrees of freedom (i.e. the number of free parameters, d) with the number of observations (n). The objective function used for parameter optimisation (ε_{MRE}) was used as an estimate of prediction error, whereas n was interpreted as the number of independent estimates of model parameters. Provided that $n \gg d$, FPEC is approximated by:

$$\text{FPEC} = \frac{1 + d/n}{1 - d/n}\varepsilon \qquad (11)$$

In principle, the model with the lowest FPEC should be adopted. For example, for $n=50$ (the lowest number of samples considered to produce a valid analysis here), it follows that each additional parameter would need to explain another 4% of the residual error. Schoups et al. (2008) pointed out that this approach requires that n is very large and may lead to underestimates of prediction error and favour overly complex models. This caveat was considered when interpreting FPEC values. The FPEC was not the only criterion used in deciding on appropriate model structure. Other factors considered were: (i) the number of stations for which alternate model structure appeared to function best; (ii) any relationships between the amount of data available and the FPEC values of alternate models; (iii) the degree to which parameter values could be correlated to catchment attributes (to increase the likelihood of robust performance use in ungauged basins); and (iv) the correlation between fitted parameters in deciding the optimal model structure: a value for Spearman's non-parametric coefficient of correla-

tion (r^*) between parameters that was greater than 0.50 was considered indicative of a model structure susceptible to parameter equivalence.

3.4 Baseflow separation

Using the chosen reservoir model and derived parameter values, the baseflow component of streamflow was estimated by combining forward and backward recursive filters. It was assumed that the very first and very last value in the streamflow time series represented baseflow only (testing showed that associated errors are negligible).

Starting at the second last value of the stream flow time series ($i=N-1$) and moving backwards through the record, baseflow for time step i was estimated by considering the forward estimate $Q_{BF,f}$ and backward estimate $Q_{BF,b}$ calculated as (Wittenberg, 1999):

$$Q_{BF,b}(i) = \left\{ [Q_{BF}(i+1)]^{b-1} + \frac{b-1}{ab} \right\}^{\frac{1}{b-1}} \tag{12}$$

or, for a linear reservoir:

$$Q_{BF,b}(i) = \frac{1}{1-k_{BF}} Q_{BF}(i+1) \tag{13}$$

The forward estimate of baseflow $Q_{BF,f}(i)$ is given for the non-linear reservoir by Eq. (8) and for the linear reservoir by Eq. (4), but in both cases this requires that $Q(i-1)$ exceeds zero and indeed represents baseflow. Where $Q(i-1)$ equalled zero, $Q_{BF,f}(i)$ was also given a value of zero.

To decide whether to assign the backward or forward baseflow estimate, the following decision tree was used:

1. If $Q(i)<Q(i-1)$ (i.e. falling limb):

   (a) If $Q_{BF,b}(i)<Q(i)$ then the backwards estimate was adopted: $Q_{BF}(i)=Q_{BF,b}(i)$

(b) If $Q_{BF,b}(i) \geq Q(i)$ then:

    i. If $Q_{BF,f}(i) < Q(i)$ the forward estimate was adopted: $Q_{BF}(i) = Q_{BF,f}(i)$

    ii. Otherwise it was assumed that $Q_{BF}(i) = Q(i)$

2. If $Q(i) \geq Q(i-1)$ (i.e. rising limb):

    (a) If $0 < Q_{BF,b}(i) < Q_{BF,f}(i)$ then the backwards estimate was adopted: $Q_{BF}(i) = Q_{BF,b}(i)$

    (b) Otherwise the forward estimate was adopted: $Q_{BF}(i) = Q_{BF,f}(i)$

An example result of this procedure is shown for a linear reservoir in Fig. 4.

3.5 Analysis of variation in model parameter values and flow components

It was tested whether a range of catchment attributes derived could explain the variation in k_{BF} between catchments. The catchment attributes and the data sources are described in Appendix A. They include morphological catchment attributes (catchment size, mean slope, flatness); soil characteristics (saturated hydraulic conductivity, dominant texture class value, plant available water content, clay content, solum thickness); climate indices (mean precipitation P, mean potential evapotranspiration E_0, remotely sensed actual evapotranspiration, humidity index $H = P/E_0$, average monthly excess precipitation (AMEP)); and land cover characteristics (fraction woody vegetation, fractions non-agricultural land, grazing land, horticulture, and broad acre cropping, remotely sensed vegetation greenness).

The analysis involved step-wise regression. The strongest predictor of inter-station variations in k_{BF} was identified from the parametric and non-parametric (ranked) correlation coefficients (r and r^*, respectively). A threshold of ± 0.40 (equivalent to $r^2 = 0.20$) was considered a potentially meaningful correlation. A regression relationship was established with all predictors showing meaningful correlation; linear, logarithmic, exponential and power regressions were established, and the most powerful one selected.

The residual variance was calculated, expressed either as actual residuals or as relative deviations (i.e. relative to "observed" values), and the same regression analysis was repeated.

When catchment attributed did not explain any further variance, the spatial correlation in the remaining residual variance was investigated using semi-variograms. The longitude and latitude of the centroid of each catchment were converted to datum (in km). A minimum of 100 unique member data points was used for each variogram estimator point and a spherical, exponential or linear semi-variogram model was visually selected and fitted. The ratio of sill over the sum of sill and nugget was interpreted as the fraction of total variance that appeared spatially correlated, and the range of variogram model interpreted as the characteristic length scale of correlation. The same semi-variogram analysis was also was performed for the various catchment attributes (see Appendix A). Attributes related to soils, topography, major land uses and vegetation cover showed typical length scales 100 to 200 km, whereas climate and potential evaporation showed length scales of 300 to 700 km (orographic and coastal effects appeared to dominate the lower end of this range). Catchment size and area with different crops did not appear to have any spatial correlation.

The same analyses and interpretation described above for k_{BF} was repeated for derived estimates of period average Q_{BF} and Q_{QF} and baseflow index (BFI, calculated as the ratio of average baseflow over average total streamflow).

4 Results

4.1 Influence of adopted storm flow window size

To investigate how the size of the data masking period T_{QF} influenced the results, the analysis was performed using a range of T_{QF} values for six arbitrarily selected stations broadly representing the geographic and climate range in the data set. The following observations were made (Fig. 5a–f):

- The apparent recession coefficient rapidly falls as T_{QF} is increased to between 7 d and 14 d for different stations. The value of k_{BF} attains a minimum value for T_{QF} of between 7 and 28 d (Fig. 5a and b, respectively).

- The number of available data pairs reduces exponentially as greater T_{QF} values are chosen, not leaving any data for T_{QF} values greater than about 20 to 40 d (Fig. 5b and d, respectively). With increasingly long T_{QF} the remaining data are increasingly likely to be associated with the single longest baseflow recession event.

- The inferred k_{BF} values show variable and sometimes complex trends (e.g. Fig. 5a and f) once T_{QF} exceeds 10 d, but the remaining number of data pairs is often small.

Based on this analysis, 10 d for T_{QF} was considered a reasonably compromise that maximised data availability whilst minimising undue influence from storm flow recession. Nevertheless this remains a subjective choice and a source of uncertainty (discussed further on).

4.2 Derived parameter values

Out of the 226 records available, 183 stations had more than 50 flow data pairs available when $T_{QF}=10$ d was used. The maximum number of data pairs was 991, and the median 217. A linear reservoir produced an average k_{BF} value of 0.0575 (st.dev.±0.0266) across the 183 stations. Values appeared approximately log-normally distributed (Fig. 6) and 80% of values were in the range 0.030–0.091. If a non-linear reservoir was fitted to the data pairs, the resulting distribution of β values was strongly skewed (Fig. 6). The median value was $\beta=0.95$, 50% of values were between 0.82–1.26 and 80% of values between 0.70–1.83. Seemingly unrealistic values of $\beta \geq 4$ were derived for eight stations and values of $\beta \leq 0.50$ found for four stations. Corresponding values of k_{BF} appeared normally distributed, and produced an average value of

k_{BF}=0.0567 (st.dev.±0.0407); 80% of all values was between 0.0012–0.1147. There was correlation between k_{BF} and β values (non-parametric r^*=−0.75). There was also correlation between the respective k_{BF} values for the linear and non-linear reservoir model (r^*=0.76).

4.3 Choice of reservoir model

The linear reservoir produced a median FPEC of 0.0308 and the non-linear reservoir a median FPEC of 0.0294, suggesting that the non-linear reservoir model reduced estimation error by 5%. The linear reservoir produced lower FPEC scores for 131 out of 183 stations, however. The parameter β could not be correlated to any catchment attribute (the largest r^* was −0.31 with E_0). Values were within 20% of unity for 88 out of 183 stations, and outside the range of 0.5–4 for 12 stations. Taken together, these findings were considered insufficient basis to prefer the more complex non-linear reservoir model over the simpler linear reservoir model. Results presented in the remainder of this paper refer to those obtained using the linear reservoir model unless stated otherwise.

4.4 Baseflow separation

The distribution of catchment baseflow index (BFI) values appeared normal by approximation, with an average BFI of 0.45 (st.dev.±0.19; Fig. 7). The average BFI calculated using the non-linear reservoir model was 0.42±0.21. The median relative difference between the two BFI estimates was 5%, and the absolute error less than 0.10 for 162 out of 183 stations (including the 12 that had unrealistic values of β) The distribution of baseflow and quick flow averages was positively skewed. Median baseflow was 0.16 mm d^{-1} and median quick flow 0.20 mm d^{-1} (Fig. 7).

4.5 Factors explaining variation in recession coefficient

The catchment attribute showing the strongest correlation with k_{BF} was catchment humidity (H; $r^* = 0.60$). A power-relationship between the two explained 27% of the variance in k_{BF} (Fig. 8a; Table 1):

$$k_{BF} = 0.458 H^{-0.4921} \qquad (r^2 = 0.27) \tag{14}$$

The residual variance was lower for humid catchments (high H) than for dry ones, but could not be explained by any of the remaining variables. The semi-variogram for the residual variance (i.e. derived value of k_{BF} minus the estimate from Eq. (14)) is shown in Fig. 8b. About 72% of this residual variance (i.e. ca. $0.72 \cdot (1-0.27) = 53\%$ of the total variance) was spatially correlated, with a characteristic length scale of 200 km. The remaining 20% of variance remained unexplained.

4.6 Factors explaining variation in baseflow index

The catchment attribute most strongly correlated to BFI was E_0 ($r^* = -0.55$), closely followed by humidity index (H), the precipitation-weighted monthly humidity index (PWMH), and the coefficient of variance in monthly precipitation (CVMP) ($r^* = 0.51 - 0.54$). An exponential relationship explained 34% of the variance in k_{BF} (Fig. 9a):

$$BFI = 1.4631 \exp(-0.3839 \bar{E}_0)(r^2 = 0.34) \tag{15}$$

The residual variance was not explained by any of the remaining catchment attributes. About 81% of the residual variance (i.e. ca. 53% of total variance) was spatially correlated with a characteristic length scale of 300 km (Fig. 9b). The remaining 13% of variance remained unexplained.

4.7 Factors explaining variation in average baseflow

The catchment attribute most strongly correlated to BF was average monthly excess precipitation (AMEP, r^*=0.91), closely followed H (r^*=0.88), P and PWMH (both r^*=0.84). The following relationship explained 84% of the variance in Q_{BF} (Fig. 10a):

$$Q_{BF} = 0.2469\text{AMEP}^{1.5007}(r^2 = 0.84) \tag{16}$$

The semi-variogram for the residual variance suggests that the 16% unexplained variance in baseflow is spatially uncorrelated (Fig. 10b).

4.8 Factors explaining variation in average quickflow

The catchment attribute most strongly correlated to QF was rainfall (r^*=0.70). A power relationship between the two explained 70% of the variance (Fig. 11a):

$$Q_{QF} = 0.0185P^{2.508}(r^2 = 0.70) \tag{17}$$

The coefficient of variation in monthly precipitation (CVMP; r^*=0.36) and rainfall-weighted event precipitation (PWEP; r^*=0.35) were the next strongest predictors, but including them did not improve estimates of QF. About 66% of the residual variance (i.e. ca. 20% of total variance) was spatially correlated over length scales of 400 km if expressed relative to estimates obtained with Eq. (17) (Fig. 11b). The remaining 10% of variance remained unexplained.

5 Discussion

5.1 Selection of storm flow window

Streamflow during the first 7 to 10 d after storm flow events appeared to be affected by rapid drainage of stores associated with the storm event, with the longer recession

times for the wetter catchments. The gradual increase in inferred k_{BF} found when using a masking period of more than between 10 and 30 d (varying between catchments) may reflect a non-linear store, but may also indicate that stream and riparian evapotranspiration losses become important at very low flows, causing a more rapid recession of the hydrograph in this regime (cf. Chapman, 2003). The number of available data pairs often becomes exceedingly small as window size increases, introducing uncertainty and bias into the analysis. A window of 10 d was considered a reasonable compromise that avoided influence from storm flow recession, while at the same time avoiding poor estimates because of small sample size. The examples shown suggest that there is usually still some uncertainty for both reasons however.

5.2 Choice of linear or non-linear reservoir model

Fitting a linear reservoir produced results that were considered satisfactory when compared to those obtained with a non-linear reservoir. The use of an additional parameter did little to explain more variance in the observations, and resulting parameter estimates sometimes appeared unrealistic. In any case, baseflow separation using a linear reservoir produced estimates of baseflow that were very similar to those obtained with a non-linear reservoir.

Previous studies have argued for the use of non-linear reservoirs on the basis of observed increases in k_{BF} for low flow conditions. As discussed above, similar increases can be inferred for some of the stations used in this analysis, for example the results in Fig. 5a and b show increasing k_{BF} as T_{QF} is increased (and low flows more influential). Following Weisman (1977) and Tallaksen (1995), Wittenberg and Sivapalan (1999) argued that evapotranspiration from the river and riparian zone will lead to an accelerating recession at low baseflow levels, leading to fitted values of $\beta < 1$. After correcting for this effect, they found values of β between 2 and 3 (b=0.3–0.5). Similar values are commonly found in other countries and could be physically explained by convergence of flow paths (Chapman, 2003; Wittenberg, 1999). In the present study, best fit values of β appeared generally close to unity and there appeared little benefit

from applying a more complex non-linear reservoir model.

5.3 Recession coefficients

The 183 catchments investigated had an average baseflow reservoir half-time of about 12 d and 80% of stations showed half-times between 7 and 34 d. The shortest half-times (and highest k_{BF} values) occurred in the driest catchments. The coefficient of variation (CV=std.dev./mean) in k_{BF} was ±46%. Of the variance in k_{BF} between stations, 27% could be attributed to climate (via the catchment humidity index H), 53% was correlated over length scales of ca. 200 km that seem indicative of terrain factors, and 20% remained unexplained (Table 1).

On theoretical arguments, Zecharias and Brutsaert (1988) argued that the recession coefficient k_{BF} should be proportional to:

$$k_{BF} \propto \frac{K D \alpha}{Y L} \qquad (18)$$

where K is hydraulic conductivity, D aquifer thickness, α is slope, Y is storativity, and L a characteristic flow path length. Zecharias and Brutsaert (1988) and Brandes et al. (2005) found that geomorphological indices such as drainage density (a proxy for L), slope and hydrologic soil class (perhaps a proxy for K and S) together explained about 70–80% of the variation in k_{BF} for catchments in the Appalachians (USA). In the current study, catchment-average saturated conductivity and slope estimates were available and showed weak correlations with k_{BF} (r^* of −0.30 and −0.41, respectively), but they were opposite to expected relationships. This is attributed to the correlation between these catchment attributes and catchment humidity; after removal of this covariance conductivity and slope did not explain any of the residual variance. Indeed, most of the variation explained by the humidity index was for dry catchments ($H<1$) with times of less than 10 d ($k_{BF}>0.07$; Fig. 8a). These catchments generally had low average base-flow (<30 mm y^{-1}) and intermittent streamflow. It is concluded that the predictive value of humidity index is mainly due to the intermittent occurrence of (perched) groundwater

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

tables with short half times in drier catchments.

The influence of perched groundwater tables, as well as perhaps the large geographical area and wide climate and geology range covered by the 183 catchments, may have prevented detection of the influence of hydrogeology and geomorphology on k_{BF}. The finding that there was considerable correlation of k_{BF} over a relatively short length scales of 200 km does suggests that there are spatial terrain factors underlying the variation in k_{BF} and that were apparently not adequately captured in the catchment data available.

5.4 Base flow index

Average BF and QF were on average of similar magnitude (BFI=0.45), but with large differences between stations: for about half (87) of all stations BF was either more than two times larger or smaller than QF (i.e., BFI<0.33 or >0.66). Perhaps as a combined result of climate factors driving QF and BF differences (see below), E_0 explained 34% of the variation in BFI (other indices including rainfall, E_0, or both had similar predictive power). Another 53% of the variation was spatially correlated, while 13% of variation remained unexplained. Based on E_0 alone, BFI could be predicted with a standard error of estimate of ±0.16.

For the conterminous USA, Santhi et al. (2008) reported BFI values of similar range and average as those reported here. They found elevation and percentage sand were the strongest predictors of BFI, being negatively and positively related to BFI, respectively. The national maps of BFI (re)produced by those authors do however suggest that precipitation (and perhaps the fraction of this falling as snow) may also have been the primary underlying factor. For the Elbe Basin (Germany), Haberlandt et al. (2001) were able to explain ca. 80% of the variance in BFI values using a combination of catchment-average slope, topographic wetness index, rainfall, and soil conductivity. In the current analysis, direct evidence for a relationship between BFI and catchment-attributes relating to geomorphology or soils was not found, but there was considerable correlation over up to 150 km that may reflect undescribed terrain factors.

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◄ ►

◄ ►

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

5.5 Average base flow and storm flow

The variation in average baseflow between stations was best explained by the monthly precipitation in excess of E_0 (84% of the variance); the remaining 16% did not show any spatial correlation. The variation in average quick flow between stations could largely
5 be explained by long-term average rainfall (70%); of the remaining variation, 20% was correlated over up to 400 km (perhaps indicative of a combination of terrain and climate variables) with the remaining 10% left unexplained.

The overriding importance of rainfall and catchment humidity (H) in determining total streamflow is well documented (e.g. Oudin et al., 2008; Van Dijk et al., 2007; Zhang
10 et al., 2004). The analysis presented here suggests this extends to both the BF and the QF component. The standard error of estimate (SEE) using the first order regression models to estimate baseflow and quick flow were both of similar magnitude but errors appeared uncorrelated. Estimates of BF were slightly more robust than QF estimates (SEE 70–87 vs. 89–94 mm y^{-1}; mean relative error 37–45 vs. 52–63%; r^*=0.89–0.92
15 vs. 0.67–0.76).

The empirical relationships derived provide some insight into the main drivers of spatial patterns in average baseflow, storm flow, and base flow index. The stronger explanatory value of monthly rainfall excess in predicting BF suggests that seasonality in rainfall relative to E_0 may be important in determining baseflow generation. Conversely,
20 QF showed a rather strongly non-linear relation with rainfall (the exponent was 2.51). This flow component could include several runoff generation mechanisms, including infiltration excess (Hortonian) flow, saturation excess (Dunne) flow and subsurface storm flow. Correspondingly, a multitude of factors may affect quick flow generation, including rainfall intensity distribution, factors affecting soil infiltration capacity (soil type but also
25 land use and management), factors affecting saturated catchment area (antecedent groundwater level, geomorphology) and soil saturation (soil conductivity and structure, antecedent soil water content). It may be assumed that average rainfall intensity is positively related to total rainfall, whereas groundwater level and soil moisture content

are likely to be higher in wetter catchments, providing several alternative hypotheses to explain the non-linear relationship between rainfall and QF that was found.

6 Conclusions

Daily streamflow data for 183 catchments across Australia were used to estimate baseflow and quick flow contributions using a baseflow separation algorithm based on Wittenberg (1999). Both linear and non-linear reservoirs were evaluated. Variations in reservoir parameters, baseflow index (BFI) and average baseflow and quick flow between the stations were analysed and where possible related to the latitude, longitude and climate, terrain and land cover attributes of the catchments, using step-wise regression and semi-variogram techniques. The following conclusions are drawn:

(1) A one-parameter linear reservoir produced estimates of baseflow that were similarly good as those obtained using a two-parameter non-linear reservoir. The former was therefore preferred, also because it had fewer parameters and parameter values that were less variable.

(2) The transition from storm flow dominated streamflow to baseflow dominated streamflow generally appeared to occur between 7 and 10 d after storm events. The 183 catchments showed baseflow half-times of around 12 d, with 80% of stations having half-times of 7 to 34 d. The shortest half-times occurred in the driest catchments and were attributed to the occurrence of fast-draining (perched) groundwater. Catchment humidity explained 27% of the variation in derived recession coefficients.

(3) Median BFI was 0.45, with considerable variation between stations. About half (53%) of the unexplained variance in recession coefficients and BFI values showed spatial correlation over scales of 200–300 km, probably associated with terrain factors. The remaining 16–20% of variance in k_{BF} and BFI remained unexplained.

(4) Most (84%) of the variation in average baseflow between stations could be explained by monthly precipitation in excess of E_0. Most (70%) of the variation in average quick flow between stations could largely be explained by average rainfall. Of the re-

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

maining variation, 20% was spatially correlated over spatial scales of ~400 km, and this may reflect a combination of terrain and climate factors. The remaining 10–16% was left unexplained.

Appendix A

Derivation of catchment attributes and correlation lengths

A GIS coverage with the outline of each of the catchments was produced through supervised terrain analysis of the 9-second Australian digital elevation model (ASLIG et al., 1996). The catchment polygons were used to calculate catchment average values and time series for the data sets described below.

The 3-second SRTM v.2 (USGS, 2006) was used to derive slope (in degrees) and the multi-resolution valley bottom flatness index (MrVBF; Gallant and Dowling, 2003). Soil mapping and soil property data were retrieved from the Australian Soil Resources Information System (Johnston et al., 2003). They include dominant soil class and texture class value (in five categories), estimates of saturated hydraulic conductivity (K_{sat}, mm h^{-1}), plant available water content (PAWC, mm), clay content (%) and solum thickness (m).

Woody vegetation cover fraction was mapped at 30 m resolution from Landsat TM mapping (NFI, 1997). Non-agricultural land, grazing land, horticulture, and broad acre cropping were derived from 1:2,500,000 land use mapping at national scale (BRS, 2006). Time series of vegetation greenness (Enhanced Vegetation Index, EVI) were calculated from MODIS satellite reflectance observations (MOD43B4, 16-day composites, 1-km resolution, 2001–2006) with nadir bidirectional reflectance distribution function adjusted reflectance (Schaaf et al., 2002). From the EVI time series, average, minimum, and maximum EVI were calculated, as well as and two indicators of vegetation seasonality; $VSI_1 = EVI_{max}/EVI_{min}$ and $VSI_2 = EVI_{max}/EVI_{avg}$.

Daily gridded climate data available included precipitation (P) produced by interpo-

lation of station data (Jeffrey et al., 2001) and Priestley-Taylor potential evapotranspiration (E_0) produced by combining interpolated climate station data with a long-term albedo climatology (Bureau of Meteorology, 2009) (both 1980–2006 and 0.05° resolution). Actual evapotranspiration (E_{RS}; 1990–2006, 1 km resolution) was estimated from MODIS and AVHRR satellite observations and the mentioned E_0 (Guerschman et al., 2008). From the gridded data catchment-average time series and average values for the period of analysis were calculated. From these in turn the following climate indicators were calculated: humidity index (H):

$$H = \overline{P}/\overline{E} \tag{A1}$$

Average evaporative fraction (EF):

$$EF = \overline{E}_{RS}/\overline{E}_0 \tag{A2}$$

Precipitation-weighted monthly humidity index (PWMH; subscript m refers to monthly totals):

$$\text{PWMH} = \sum \left(P\frac{P}{E_0}\right)_m \Big/ \sum P_m \tag{A3}$$

Average monthly excess precipitation (AMEP; n_m refers to the number of months):

$$\text{AMEP} = \frac{1}{n_m} \sum \max\left\{0, P_m - E_{0,m}\right\} \tag{A4}$$

Depth-weighted average event precipitation (DWAEP; p denotes event precipitation):

$$DWAEP = \frac{\sum p^2}{\sum p} \tag{A5}$$

Mean event precipitation (MEP; n_p denotes the number of days with rainfall):

$$\text{MEP} = \frac{\sum p}{n_p} \tag{A6}$$

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Coefficient of variation in monthly precipitation (CVMP):

$$\text{CVMP} = \frac{\sum \left(P - \overline{P}\right)_m^2}{\sum P_m} \tag{A7}$$

Spatial correlation lengths of all attributes listed above were estimated by visually fitting a spherical model to the semi-variogram. The resulting ranges are listed in Table A1. Attributes related to soils, topography, major land uses and vegetation cover generally have correlation lengths of 100 to 300 km, whereas climate and potential evaporation have correlation lengths of 300 to 750 km. The semi-variograms for rainfall and AMEP suggested two superimposed models, with a smaller part of variation being correlated over ca. 200 km (attributed to orographic and coastal effects) and a larger part correlated over greater lengths (attributed to broad climate zones). Catchment size and area with different crops did not show any spatial correlation.

References

Akaike, H.: Statistical predictor identification, Ann. Inst. Stat. Math., 22, 203–217, doi: 10.1007/BF02506337, 1970.

ASLIG, ANU, and AHC: 9 second DEM (cartographic material) 1 edn., Australian Surveying and Land Information Group, Australian National University, Australian Heritage Commission, Australia Belconnen, Australia, 1996.

Bergström, S.: The HBV Model, in: Computer Models Of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Littleton, Colorado, 443–476. 1995.

Brandes, D., Hoffmann, J. G., and, Mangarillo, J. T.: Base flow recession rates, low flows, and hydrologic features of small watersheds in Pennsylvania, USA, J. Am. Water Resour. As., 41, 1177–1186, 2005.

BRS: 1992/93, 1993/94, 1996/97, 1998/99, 2000/01 and 2001/02 land use of Australia, version 3, Bureau of Rural Sciences, Canberra, 2006.

Burnash, R. J. C., Ferral, R. L., and McGuire, R.: A Generalized Streamflow Simulation System-Conceptual Modeling for Digital Computers, US Department of Commerce, National Weather Service and State of California, Department of Water Resources, 1973.

Chapman, T. G.: Modelling stream recession flows, Environ. Modell. Softw., 18, 683–692, 2003.

Chiew, F. H. S., Peel, M. C., and Western, A. W.: Application and Testing of the Simple Rainfall-Runoff Model SIMHYD, in: Mathematical Models of Small Watershed Hydrology and Applications, edited by: Singh, V. P., and Frevert, D. K., Water resources Publication Littleton, Colorado, USA, 335–367, 2002.

Coutagne, A.: Etude générale des variations de débits en fonction des facteurs qui les conditionnent, 2éme partie: les variations de débit en période non infuencée par les précipitations, La Houille Blanche, 416–436, 1948.

Gallant, J. C. and Dowling, T. I.: A multi-resolution index of valley bottom flatness for mapping depositional areas, Water Resour. Res., 39, 1347–1360, 2003.

Guerschman, J.-P., Van Dijk, A. I. J. M., McVicar, T. R., Van Niel, T. G., Li, L., Liu, Y., and Peña-Arancibia, J.: Water balance estimates from satellite observations over the Murray-Darling Basin, CSIRO, Canberra, Australia, 93, 2008.

Haberlandt, U., Klöcking, B., Krysanova, V., and Becker, A.: Regionalisation of the base flow index from dynamically simulated flow components – a case study in the Elbe River Basin, J. Hydrology, 248, 35–53, 2001.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29, 2637–2649, 1993.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environ. Modell. Softw., 16, 309–330, 2001.

Johnston, R. M., Barry, S. J., Bleys, E., Bui, E. N., Moran, C. J., Simon, D. A. P., Carlile, P., McKenzie, N. J., Henderson, B. L., Chapman, G., Imhoff, M., Maschmedt, D., Howe, D. Â., Grose, C., Schoknecht, N., Powell, B., and Grundy, M.: ASRIS: the database, Aust. J. Soil

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Res., 41, 1021–1036, doi:10.1071/SR02033, 2003.

NFI: National Plantation Inventory of Australia, National Forest Inventory, BRS, Canberra, 1997.

Oudin, L., Andréassian, V., Lerat, J., and Michel, C.: Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments, J. Hydrol., 357, 303–316, 2008.

Peel, M. C., Chiew, F. H. S., Western, A. W., and McMahon, T. A.: Extension of Unimpaired Monthly Streamflow Data and Regionalisation of Parameter Values to Estimate Streamflow in Ungauged Catchments. Report prepared for the Australian National Land and Water Resources Audit, Centre for Environmental Applied Hydrology, The University of Melbourne, 2000.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., and Tuppad, P.: Regional estimation of base flow for the conterminous United States by hydrologic landscape regions, J. Hydrol., 351, 139–153, 2008.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X. W., Tsang, T., Strugnell, N. C., Zhang, X. Y., Jin, Y. F., Muller, J. P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D.: First operational BRDF, albedo nadir reflectance products from MODIS, Remote Sens. Environ., 83 135–148, 2002.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resour. Res., 44, W00B03, doi:10.1029/2008WR006836, 2008.

Tallaksen, L. M.: A review of baseflow recession analysis, J. Hydrol., 165, 349–370, 1995.

USGS: Shuttle Radar Topography Mission, 3 Arc Second scene SRTM 2.0, Global Land Cover Facility, University of Maryland, College Park, Maryland, 2006.

Weisman, R. N.: The effect of evapotranspiration on streamflow recession, Hydrol. Sci. Bull., XXII, 371–377, 1977.

Wittenberg, H.: Baseflow recession and recharge as nonlinear storage processes, Hydrol. Process., 13, 715–726, 1999.

Wittenberg, H. and Sivapalan, M.: Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation, J. Hydrol., 219, 20–33, 1999.

Zecharias, Y. B. and Brutsaert, W.: Recession characteristics of groundwater outflow and baseflow from mountainous watersheds, Water Resour. Res., 24, 1651–1658, 1988.

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., and Briggs, P. R.: A rational function approach for estimating mean annual evapotranspiration, Water Resour. Res., 40, W02502, doi:10.1029/2003WR002710, 2004.

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

Title Page

Abstract   Introduction

Conclusions   References

Tables   Figures

◄   ►|

◄   ►

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Summary of the analysis of variance in values derived from baseflow separation for the 183 catchments. Listed are the fraction of variance explained by catchment attributes, the residual variance showing spatial correlation and the remaining unexplained variance. Also listed are the range (km) of the fitted semi-variogram (see text).

Variable	Symbol	Attributed (%)	Spatially correlated (%)	Unexplained (%)	Range (km)
Recession coefficient	k_{BF}	27	53	20	200
Baseflow index	BFI	34	53	13	300
Base flow	Q_{BF}	84	0	16	n/a
Quick flow	Q_{QF}	70	20	10	400

Table A1. The range of the semi-variogram models fitted to the calculated catchment attributes in order of increasing magnitude.

Catchment attribute	Range of semi-variogram (km)
Size	none
Horticulture	none
Broad acre cropping	none
EVI seasonality	100
Solum thickness	150
Slope	200
K_{sat}	200
Woody vegetation	200
Grazing	200
EVI_{avg}	200
Soil texture class	250
PWMH	250
EVI_{min}	250
MrVBF	300
PAWC	300
non-agricultural use	300
CVMP	300
EVI_{max}	300
E_{RS}	300
E_0	400
DWEAP	400
HI	550
Clay content	600
P	600 (200)
MEP	700
AMEP	750 (200)
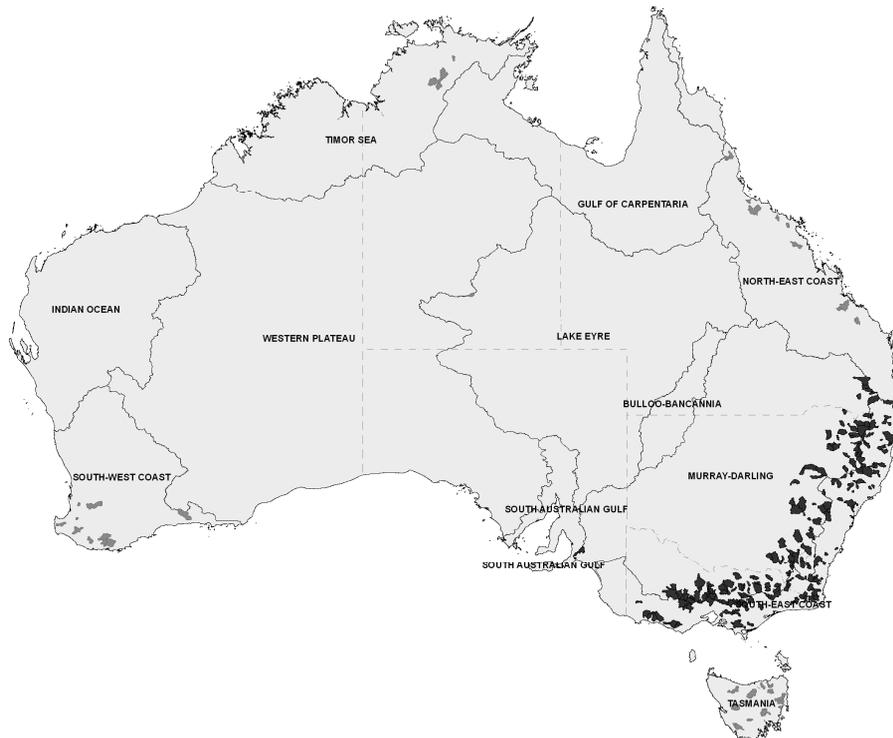
Fig. 1. The geographical distribution of the 183 catchments used in this study.
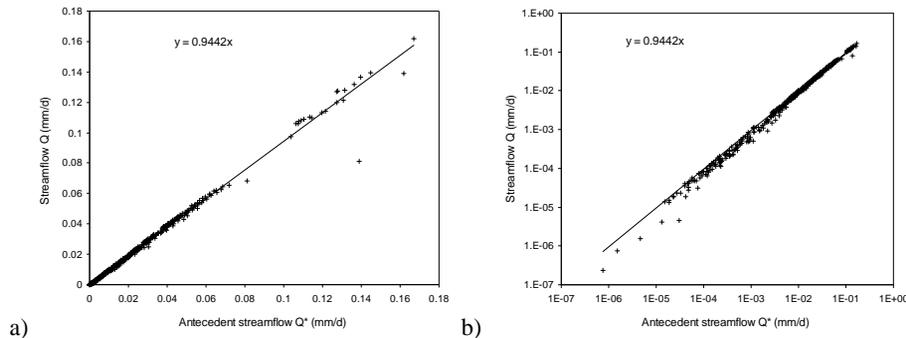
Fig. 2. Example of daily streamflow and antecedent streamflow data pairs and fitted regression equations representing a **(a)** linear baseflow reservoirs plotted on a linear scale, and **(b)** a non-linear reservoir plotted on a logarithmic vertical (data chosen arbitrarily to illustrate concepts; gauge 410705, Molonglo River @ Burbong Bridge).
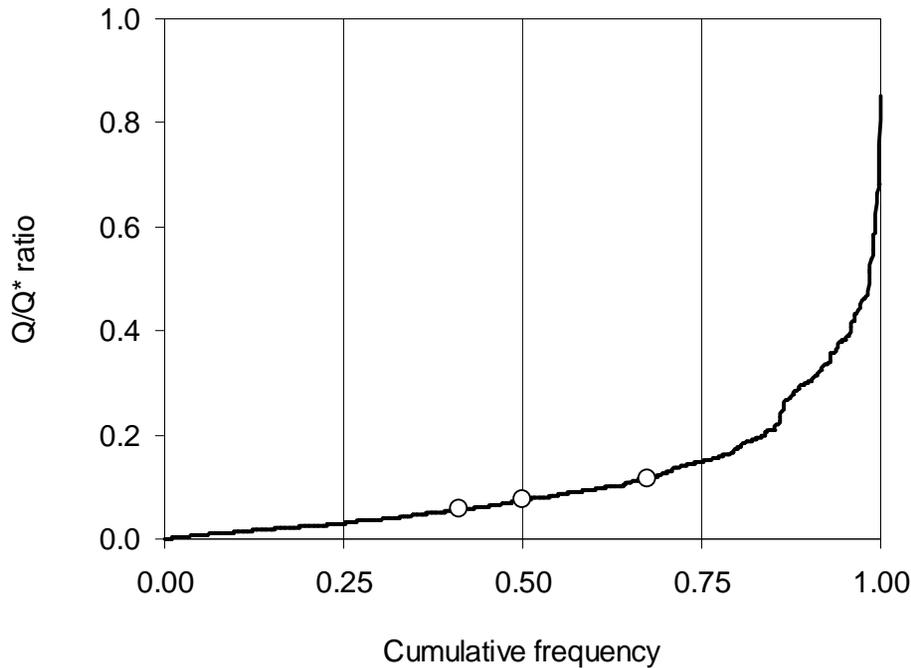
Fig. 3. Example of distribution of k_{BF} values derived from individual Q/Q_* pairs (same data as Fig. 2). Open dots indicate (from left to right) the adopted value, the median, and the mean value, respectively.
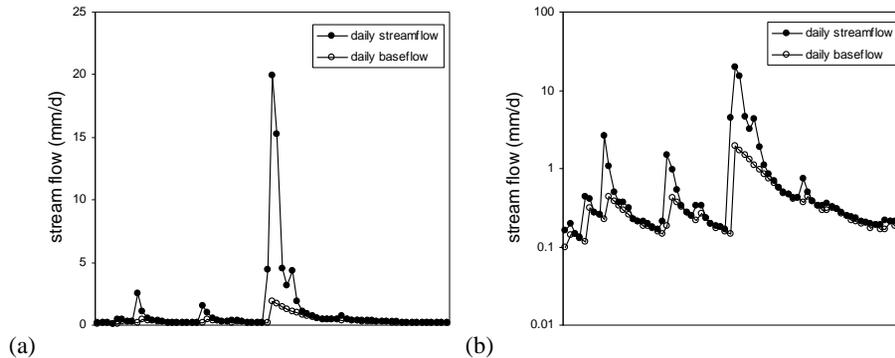
Fig. 4. Example of separation of daily streamflow into baseflow and storm flow using a linear baseflow reservoir , plotted on **(a)** a linear vertical scale and **(b)** a logarithmic vertical scale (data chosen arbitrarily to illustrate concepts; represent 60 d in winter 1990; gauge 410705, Molonglo River @ Burbong Bridge).
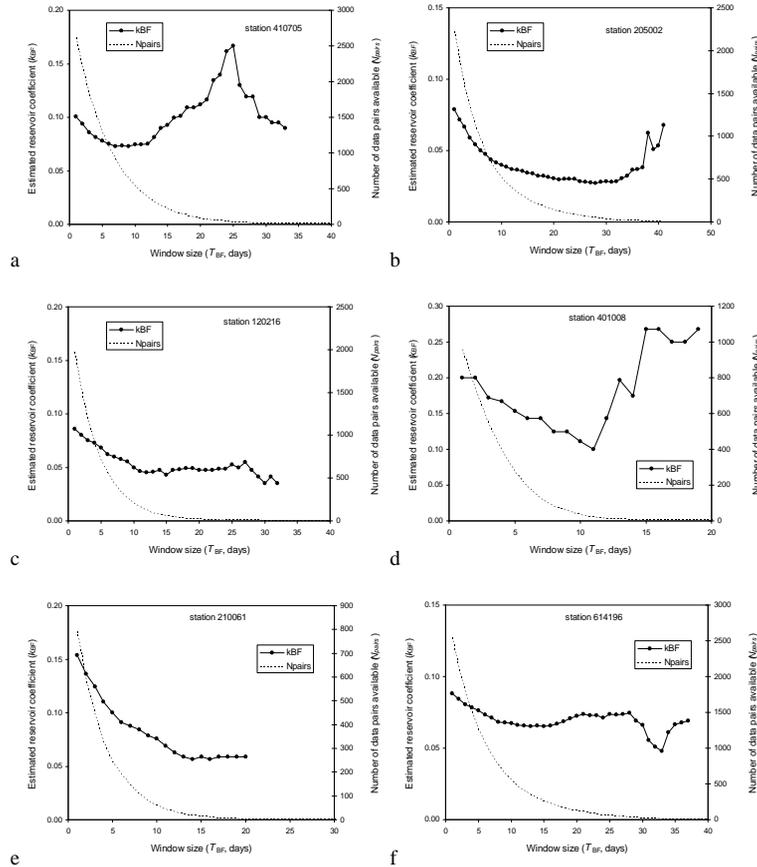
Fig. 5. Example k_{BF} values derived (closed lines) and number of Q/Q_* pairs (dotted line) as the length of the storm flow masking window T_{BF} is increased from zero to 50 d. The six stations shown were selected to cover different geographical areas and climate regimes.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
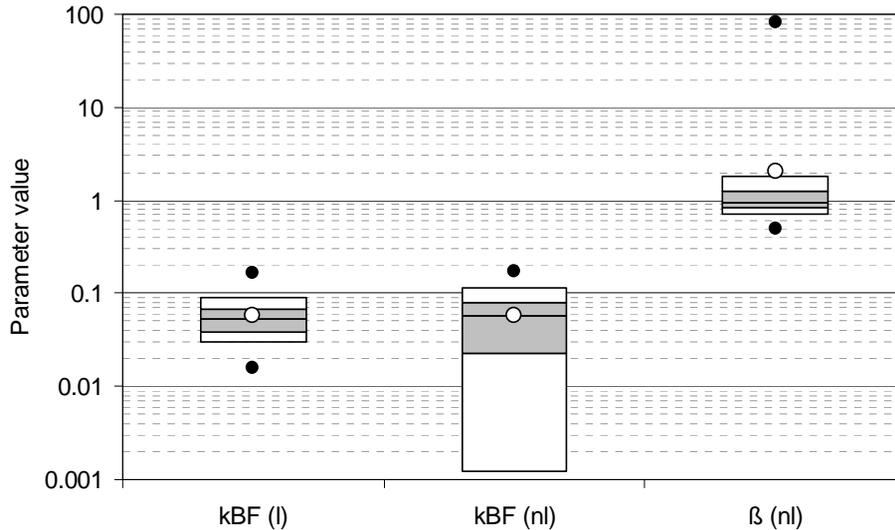
Fig. 6. Distribution of derived parameter values (N=183), from left to right, k_{BF} for a linear reservoir (l) and for a non-linear reservoir (nl) and the fitted value of β for the non-linear reservoir. Shown are the mean (open dot), minimum and maximum (closed dots), 10–90% range (white bars), and the 25, 50 and 75% percentiles (shaded bars). Note the logarithmic scale on the vertical axis.
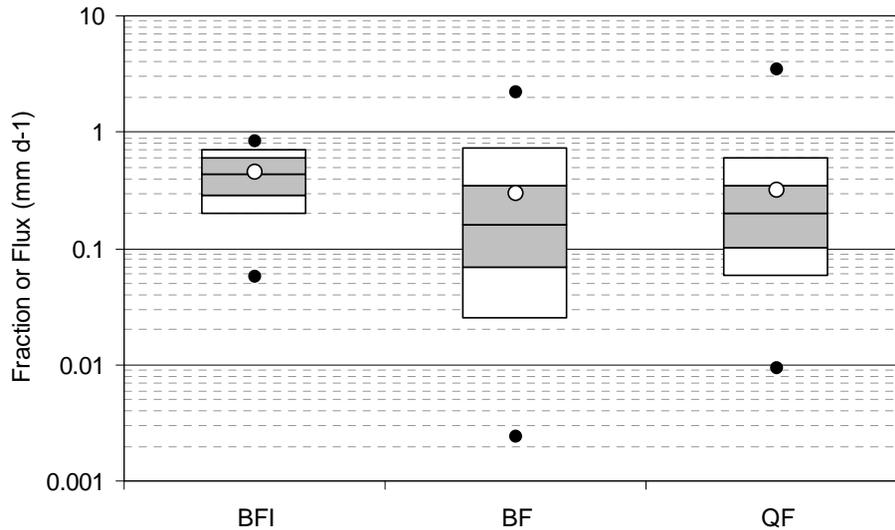
Fig. 7. Distribution of values of (from left to right) baseflow index (BFI), average baseflow (BF) and average quick flow (QF, both in $mm\,d^{-1}$) derived by baseflow separation using a linear reservoir. Shown are the mean (open dot), minimum and maximum (closed dots), 10–90% range (white bars), and the 25, 50 and 75% percentiles (shaded bars). Note the logarithmic scale on the vertical axis.
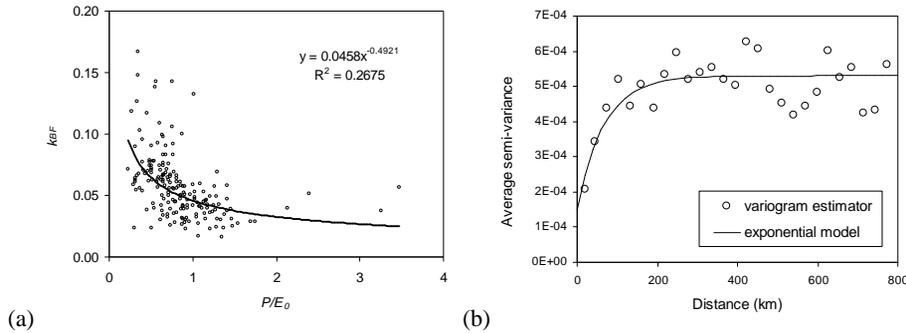
Fig. 8. (a) apparent relationship between humidity index H ($H=P/E_0$) and the linear recession coefficient k_{BF} and **(b)** the semi-variogram for the remaining residual variance with a visually fitted exponential model (nugget=$1.5\cdot10^{-4}$, sill=$3.8\cdot10^{-4}$, range=200 km).
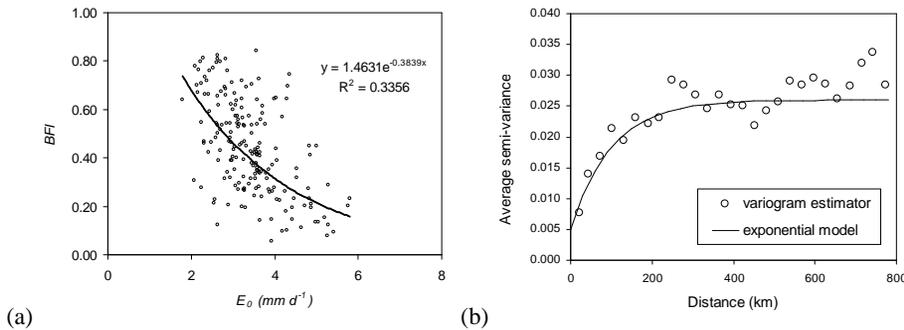
Fig. 9. (a) Apparent relationship between E_0 and the period average baseflow index BFI and **(b)** the semi-variogram for the remaining residual variance with a visually fitted exponential model (nugget=0.005, sill=0.0021, range=300 km).

Characteristics and drivers of baseflow response

A. I. J. M. van Dijk

(a)

(b)
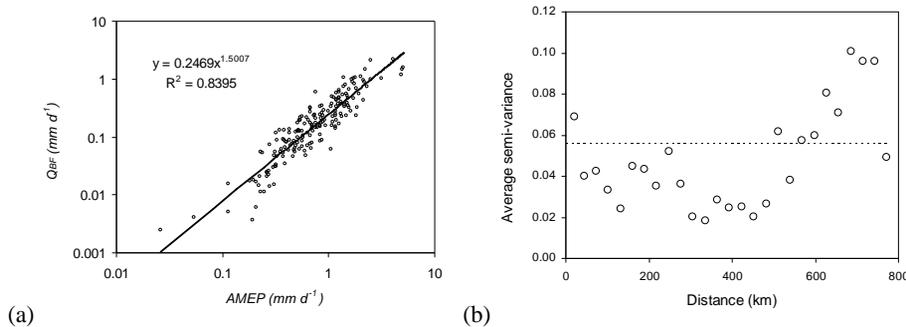
Fig. 10. (a) Apparent relationship between average monthly excess precipitation (AMEP) and the period average baseflow (BF in mm d^{-1}) (note double logarithmic scale), and **(b)** the semi-variogram for the remaining residual variance, showing no evidence for spatial correlation.
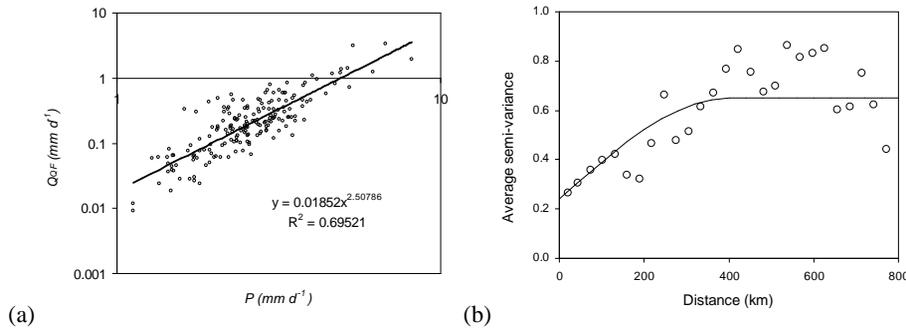
Fig. 11. (a) Apparent relationship between average precipitation (P) and the period average quick flow (QF in $mm\,d^{-1}$), and **(b)** the semi-variogram for the remaining relative residual variance (i.e. normalised by actual QF) with a visually fitted spherical model (nugget=0.24, sill=0.65, range=400).