Synthetic experiments to investigate methodological issues in detecting a land cover signal in streamflow data from multiple heterogeneous catchments

Deleted: Top-down analysis of collated streamflow data from heterogeneous catchments leads to underestimation of land cover influence¶

Albert I.J.M. van Dijk[1*], Jorge L. Peña-Arancibia[1,2,3], L.A. (Sampurno) Bruijnzeel[2]

[1] CSIRO Land and Water, Canberra, ACT, Australia

[2] VU University, Amsterdam, Netherlands

[3] King's College, London, UK

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

Abstract

Controlled experiments have provided strong evidence that changing land cover (e.g. deforestation or afforestation) can affect the water balance. However a similarly strong influence has not been detected in analyses of collated streamflow data from catchments with mixed land cover. We tried to explain this paradox using streamflow observations from 278 Australian catchments, a 'top-down' inference method (fitting the Zhang formulation of the Budyko model); and a 'bottom-up' dynamic hydrological process model (the Australian Water Resources Assessment system Landscape model, AWRA-L). Analysis with the Zhang model confirmed the previously reported absence of a strong land cover signal in the streamflow data set. The process model was able to predict a lack of signal in the heterogeneous catchment data set, as well the land cover influence observed in controlled experiments. This suggested there are likely to be methodological issues with the top-down analysis approach. To test this, synthetic experiments were performed in which the Zhang model was used to analyse synthetic AWRA-L streamflow simulations for the 278 catchments. This suggested three reasons why the Zhang model did not accurately quantify the land cover signal: (1) measurement and estimation errors in land cover, precipitation and

Deleted: 'land cover

Deleted: '

Deleted: model

Deleted: However, absence of evidence does not equate to the proof of absence, and AWRA-L

Deleted: was able to reconcile the streamflow data from the 278 catchments with experimental knowledge

Deleted: Experiments

Deleted: demonstrated

streamflow, (2) the importance of additional climate factors; and (3) the presence of covariance in the streamflow and catchment attribute data. These methodological issues are likely to prevent the use of top-down methods to try and detect and accurately quantify a land cover signal in data from catchments with mixed land cover. However, our findings do not rule out physical processes that diminish land cover influence in catchments with mixed land cover.

Deleted: any

Deleted: Our

Deleted: , including atmospheric feedback associated with rainfall interception.

1. Introduction

There is strong experimental evidence that changing land cover (e.g. deforestation or afforestation) can affect the local water balance. Such an influence has been detected at various scales, from site water balance and atmospheric water flux studies to small catchments undergoing change (see review by e.g. van Dijk and Keenan, 2007 and references therein). Controlled catchment experiments have demonstrated a change in mean catchment streamflow after land cover change (typically forest planting or logging; Bosch et al., 1982; Bruijnzeel, 1990; Andréassian, 2004; Bruijnzeel, 2004; Brown et al., 2005; Farley et al., 2005). They appear to provide clear evidence that land cover characteristics affect mean streamflow, although this influence is moderated by a range of climate and catchment characteristics as well as vegetation attributes beyond broad land cover class alone (Andréassian, 2004; Bruijnzeel, 2004; van Dijk and Keenan, 2007). These conclusions could be corroborated by analysis of collated longer term mean streamflow (Q) estimates from multiple catchments, provided only catchments with (near complete) forest cover and herbaceous cover were selected (Holmes et al., 1986; Turner, 1991; Zhang et al., 1999; 2001). The collated data were still dominated by small experimental catchments, however, and such experiments are not without their challenges (discussed further on).

Subsequent studies have attempted to detect a similar land cover influence by statistically analysing Q from many catchments with mixed land cover. In such data sets, climate is the primary reason for variation in response and therefore needs to be controlled for. Several studies do this by 'fitting' an additive formulation of a Budyko model[1] (Budyko, 1974) that explicitly represents two (e.g., 'forest' and 'herbaceous') or a small number of land cover types (Zhang et al., 2004; van Dijk et al., 2007; Oudin et al., 2008; Donohue et al., 2010; Peel et al., 2010). Such an approach has been described as a 'top-down' analysis (sensu Klemeš, 1983; Sivapalan et al., 2003). In formula:

Deleted: Several s

Deleted: large

Deleted:

Deleted: (Zhang et al., 2004; van Dijk et al., 2007; Oudin et al., 2008; Donohue et al., 2010; Peel et al., 2010).

Deleted: Most of these

Deleted: inversely applied

Deleted: i.e.

Deleted: using

$$Q_j = \sum_i FC_{i,j} f(P_j, PE_j, w_i) \qquad (1)$$

[1] Defined here as any rational function that embodies the same conceptual model as the original (see various examples in e.g. Oudin et al., 2008).

Deleted: represents

where Q_j, P_j, and PE_j are the longer-term average streamflow, precipitation and potential evaporation[2] (in mm per time unit) for catchment j, $FC_{i,j}$ is the fractional cover of land cover type i in catchment j, and w_i a dimensionless model parameter that characterises the hydrological behaviour of land cover class i and may be interpreted as a measure of the efficiency with which vegetation accesses and uses stored water. The influence of land cover is subsequently tested by finding the w_i values that minimise the root mean square error (RMSE) between observed and estimated streamflow averages, and interpreting the found parameter values. The cited studies performed such an analysis using collated data for 221 (Donohue et al., 2010) to 1508 (Oudin et al., 2008) catchments. They report either a much smaller land cover influence than found in controlled experiments (Zhang et al., 2004; van Dijk et al., 2007; Oudin et al., 2008; Donohue et al., 2010; Peel et al., 2010); no statistically significant influence (Zhang et al., 2004; van Dijk et al., 2007; Oudin et al., 2008; Peel et al., 2010); or even an influence opposite to that expected – at least for some land cover classes (Oudin et al., 2008; Peel et al., 2010) or climate types (van Dijk et al., 2007; Peel et al., 2010).

Deleted: (out of a total number of 221 and 1508 reported in the various studies)

Deleted: .

Deleted: se

Deleted: have found

It seems contradictory that land cover change should have a marked effect on the water balance of a catchment when it has homogeneous land cover, but not when it has mixed land cover. Some possible physical and methodological causes have been suggested for this paradox. Physical explanations include:

Deleted: is paradoxical

Deleted: 'land cover

Deleted: '

1. *Catchment size.* The nature of controlled experiments puts a limit to the size of catchments that can be manipulated and the majority of experiments have been carried out on catchments smaller than 1 km^2 (see e.g. tabulated data in Andréassian, 2004; Brown et al., 2005). Conversely, data sets of 'real-world' catchments with mixed land cover tend to have average catchment sizes in the order of hundreds to thousands km^2 (see respective studies listed earlier). A known issue with small catchments is the risk of ungauged subterranean transfers (e.g. Bruijnzeel, 1990), which could lead to overestimation of the influence of land cover change on streamflow. Conversely, while land surface-atmosphere feedbacks perhaps can safely be ignored for small catchments, that may not be the case for large catchments, where land cover certainly influences overall evaporative energy and may even modulate precipitation (for discussion see Donohue et al., 2007; van Dijk and Keenan, 2007)

Deleted: In addition

[2] In 'evaporation' we include all evaporation and transpiration fluxes.

2. *Catchment hydrological processes.* As catchment experiments require small and well defined watersheds they may be expected to have greater relief in comparison to larger catchments. Greater relief may mean shallower soils, less infiltration and therefore more storm flow, a more efficient surface drainage network, and lesser evaporation losses from streams, wetlands and groundwater-using vegetation (van Dijk et al., 2007).

3. *Land cover characteristics*. Experimental catchments may be expected to have a more 'idealised' and homogenous vegetation cover and fewer activities and structures designed to reduce storm runoff. In afforestation studies, the selection of 'suitable' catchments may have created a bias towards low complexity land cover, whereas land cover after clearing is unlikely to be representative of established agricultural landscapes. Large mixed land cover catchments may include surface runoff intercepting features (e.g. hillside farm dams, tree belts) and unaccounted surface water or groundwater use (Calder, 2007; van Dijk et al., 2007). In addition, forest clearing in experimental studies may be associated with soil disturbance, which may enhance streamflow generation for reasons that are not directly attributable to land cover per se (Bruijnzeel, 2004). The consequence may be that the contrast in hydrological response between forested and non-forested land may be greater in experimental catchments than in non-experimental catchments.

There are also some potential methodological issues:

4. *Other overriding climate and terrain factors.* Confident detection and attribution of a land cover influence requires that other factors are considered and controlled for. Budyko theory controls for the two most important determinants of the long-term water balance, *P* and *PE*. One might question whether the Budyko framework is sufficiently powerful to evaluate the effect in addition to *P* and *PE* alone, and if so, whether indeed land cover is the next most important variable. Additional factors potentially as, or more important than, land cover include the phase difference between seasonal *P* and *PE* patterns (Budyko, 1974; Milly, 1994) and other aspects of their temporal behaviour (e.g. rainfall intensity). Depending on their covariance with land cover, these attributes may attenuate or enhance any land cover signal.

Deleted: from

5. *Covariance between land cover and climate*. Covariance between land cover and climate is commonly present in collated catchment data sets due to the correlation between natural biomes and climate, and because of the role of landscape and climate in land use and land

cover change decisions. For example, catchments with considerable remnant and plantation forests will usually be found more commonly in regions with greater relief and typically associated greater *P* and lower *PE* than their lowland counterparts. Applying an additive response model to a data set with covariance between candidate predictors makes erroneous results more likely. Van Dijk et al. (2007) attempted to control for this effect and concluded that it influenced the results, but was probably not the only cause for their counterintuitive results.

6. *Measurement error*. Analyses of data from small catchments have not been able to detect a significant change in stream flow when land cover is changed in less than 15–20% of a catchment (Bosch et al., 1982; but see Trimble et al., 1987; Stednick, 1996). Arguably, this can be attributed to the influence of measurement noise on the analysis. Statistically, therefore it might be expected that it is harder to detect a land cover signal in large catchments with land cover mixtures than it is for catchments with homogeneous land cover. Using additive Budyko models requires estimates not only of *Q*, but also of catchment average *P*, *PE* and fractional cover (*FC*) of the land cover classes of interest. Errors will occur in each of these and may affect the analysis results, even more so if errors are not random. For example, Oudin et al. (2008) speculated that systematic precipitation measurement errors affected their analysis.

1.2 Objective

In this study, we aim to test the hypothesis that methodological issues prevent the use of 'top down' analysis to accurately detect and quantify land cover influences by analysing data sets of catchments with mixed land cover. To test this, we used mean streamflow observations from 278 non-experimental Australian catchments, the Zhang formulation of the Budyko model, and a 'bottom-up' dynamic hydrological process model with explicit representation of vegetation characteristics (AWRA-L). Synthetic experiments were performed in which the Budyko model was used to analyse process model simulations for the 278 catchments. To paraphrase, we use the more complex model (AWRA-L) to create a virtual laboratory. We then perform a virtual experiment and use a simpler model (the Budyko model) as an analytical tool to interpret the results. If our experiment can reproduce (and therefore reconcile) the contradictory results of earlier studies described above, this would seem to confirm our hypothesis.

Deleted: simple

Deleted: methods

1.3 Caveats

The way in which we use models in this experiment is not unique but somewhat unusual in the hydrological literature, and comments on earlier versions of this paper suggest our objectives can be misunderstood. It may be worthwhile to state what our objectives are explicitly *not*:

Deleted: The emphasis on methodological issues does not negate the plausibility of additional, physical causes, and we will discuss some of these.

- We do *not* aim to validate or falsify the dynamic process model (AWRA-L) we used in this experiment. We also do *not* aim to prove that the model structure and parameter values used here are the best possible description of reality, or better than other model or models. Any model can only ever be a flawed and simplified abstraction of reality (Oreskes et al., 1994). We only use the AWRA-L model because we understand its behaviour well and because it is able to reproduce two key features also observed in real data sets that are discussed in more detail further below. Any other model able to meet this criterion would have been suitable for the experiment.

- We do *not* aim to prove that the methodological issues described are the only or even the main cause for the paradox discussed. Their presence certainly would not negate the plausibility and presence of additional methodological or physical explanations, and we will discuss some of these.

- Similarly, we do *not* propose that we can use the more complex model to detect or demonstrate a land cover influence. This is neither necessary (we refer to the empirical evidence discussed) nor possible (a model cannot provide proof). We will discuss this point in more detail further on.

- We do *not* aim to falsify or discredit Budyko type models as a useful and predictive theory, and do *not* question the usefulness of 'top-down' analysis as a paradigm. We focus here on only one very specific application, that is, whether analysing collated data from heterogeneous catchments by fitting a form of the Budyko model (a composite Zhang curve model) is able to accurately detect land cover influence.

2. Methods

2.1 Data

The streamflow data used here were identical to the data used by Van Dijk and Warren (2010), which is a subset of 278 out of around 326 records used in previous studies (Guerschman et al., 2008; Guerschman et al., 2009; Van Dijk, 2009; Van Dijk, 2010a) and very similar in composition to Australian catchment data used in other studies (e.g. Zhang et al., 2004; Peel et al., 2010). Catchment boundaries were derived from a 9" resolution digital elevation model (Fig. 1) and catchments with major water regulation infrastructure were excluded. The 278 catchments that were selected had data for at least five, not necessarily consecutive years between 1990 and 2006 (median 16 years). Woody vegetation cover fraction was mapped on the basis of Landsat Thematic Mapper imagery for 2004 and daily precipitation and Priestley-Taylor *PE* was interpolated at 0.05° resolution from station data (Jeffrey et al., 2001). Catchment areas varied from 23–1937 (median 278) km^2, tree cover from 0-90% (median 25%), *P* from 404–3138 (median 836) mm year^{-1}, *PE* from 766–2096 (median 1265) mm year^{-1} and Q_{obs} from 4–1937 (median 114) mm year^{-1}.

Deleted: good

Deleted: (based on quality codes)

[FIGURE 1 HERE]

2.2 Budyko model

Oudin et al. (2008) tested five different Budyko models formulations and found little difference in their explanatory power. We chose the model of Zhang et al. (2001) because it was used successfully to detect land cover influence in a global streamflow data set of (mostly small) catchments with homogeneous land cover. For a single land cover class, the model can be written as:

$$Q = \frac{P}{1 + \frac{P}{PE} + w\left(\frac{PE}{P}\right)^2} \qquad (2)$$

For a catchment with a two land cover classes, forest and herbaceous vegetation, Eq. (2) can be rewritten as (cf. Eq. (1)):

$$Q = FC(forest)\frac{P}{1+\frac{P}{PE}+w(forest)\left(\frac{PE}{P}\right)^2} + FC(herbaceous)\frac{P}{1+\frac{P}{PE}+w(herbaceous)\left(\frac{PE}{P}\right)^2} \quad (3)$$

2.3 Dynamic model

The dynamical model used is the Australian Water Resources Assessment system Landscape hydrology (AWRA-L) model (version 0.5; Van Dijk, 2010b; Van Dijk and Renzullo, 2011). AWRA-L can be considered a hybrid between a simplified grid-based land surface model and a non-spatial catchment model applied to individual grid cells. Where possible process equations were selected from literature and selected through comparison against observations. Prior estimates of all parameters were derived from literature and analyses carried out as part of model development. Full technical details on the model can be found in Van Dijk (2010b) but some salient aspects are summarised here. The configuration used here considers two hydrological response units (HRUs): deep-rooted tall vegetation ('forest') and shallow-rooted short vegetation ('herbaceous'). The water balance of a top soil, shallow soil and deep soil compartment are simulated for each HRU individually and have 30, 200 and 1000 mm plant available water storage respectively. Groundwater and surface water dynamics are simulated at catchment scale. Minimum meteorological inputs are gridded daily total precipitation and incoming short-wave radiation and daytime temperature. Maximum evaporation and transpiration given atmosphere and vegetation conditions are estimated using the Penman-Monteith model (Monteith, 1965). Actual transpiration is calculated as the lesser of maximum transpiration and maximum root water uptake given soil water availability. Rainfall interception is estimated separately using a variable canopy density version of the event-based Gash model (Gash, 1979; Van Dijk et al., 2001a,b) to account for observed high rainfall evaporation rates (for discussion see e.g. van Dijk and Keenan, 2007). The influence of vegetation on the water balance occurs in a number of ways: compared to short vegetation, forest vegetation is parameterised to have lower albedo, greater aerodynamic conductance, greater wet canopy evaporation rates, lower maximum stomatal conductance, thicker leaves, access to deep soil and ground water, and adjust less rapidly to changes in water availability.

Deleted: Actual

Deleted: is

Deleted: , but

Deleted: rainfall

Van Dijk and Warren (2010) evaluated AWRA-L with the configuration and parameterisation used here against a range of in situ and satellite observations of water balance components and vegetation dynamics. This included evaluation against Q_{obs} from the catchments used in this analysis, as well as flux tower latent heat flux observations at four sites across Australia including both forest and herbaceous sites (Van Dijk and Warren, 2010). Latent heat flux patterns for dry canopy conditions were reproduced well. Comparison of total latent heat flux was difficult due to the large uncertainty in rainfall interception evaporation estimated from the flux tower measurements. Streamflow records were reproduced with an accuracy that was commensurate to that achieved by other rainfall-runoff models with a similar calibration approach.

Deleted: Improved model parameterisations are currently being developed but for the current analysis AWRA-L was used with prior estimates.

2.4 Experiments

1. Can the paradoxical results be reproduced and be reconciled with the process model?

Deleted: *land cover paradox*

We did two tests to see whether we could reproduce the paradoxical results of published top-down analyses of collated streamflow data from non-experimental catchments. First, we fitted the two parameter Zhang model (Eq. (3)) by minimising the standard error of estimate (SEE) against Q_{obs} from the 278 catchments (using Solver in Microsoft Excel™). We interpreted the derived *w(forest)* and *w(herbaceous)* parameter values and implied land cover to assess whether we obtained the same paradoxical results of earlier studies in catchments with mixed land cover.

Deleted: (sensu Klemeš, 1983; Sivapalan et al., 2003)

Next, we investigated whether the AWRA-L could reconcile the apparent contradiction, which means meeting two conditions. First, the model needs to reproduce the observed streamflow from the 278 catchments satisfactorily. We considered this to be the case if the predictions were as good as that of the calibrated two-parameter Zhang model, or better. Second, the model needs to be in agreement with experimental catchment studies of land cover change. One test of this would be to reproduce streamflow changes observed in an actual paired catchment experiment, but unfortunately we did not have the daily streamflow and meteorological data required from such an experiment available, and one example would have limited statistical significance. Instead, we used AWRA-L to simulate streamflow from the 278 catchments under conditions of full forest and full herbaceous cover, respectively. We compared the resulting water balance estimates with the empirical relationships for the respective land cover type reported by Zhang et al. (2001), who propose two alternative models to estimate *Q*. The first method (Zhang-A) is to use Eq. (3) with values of *w(forest)*=2.0 and *w(herbaceous)*=0.5, with *PE* estimated using the Priestley-Taylor formula and a 'standard' land cover with assumed albedo and aerodynamic conductance. The second method (Zhang-B) is to use the same approach, but substitute *PE* by values of 1410 and 1100 mm year^{-1} for forest and herbaceous cover, respectively. The latter reduces the physical

Deleted: the land cover paradox

Deleted: performance to be acceptable if realism of the model, but provides a convenient alternative to where *PE* estimates are not readily available, and has been shown to agree well with other empirical relationships (Holmes et al., 1986; Turner, 1991) and data from catchments with homogeneous land cover (Zhang et al., 2001). We emphasise that our objective does not require that the process model explains *more* variation than the Zhang models in one or both cases; equal or similar performance would be sufficient. The critical difference is that fitting the Zhang models is expected to lead to two substantially different parameter sets, producing two mutually inconsistent models in the respective applications. By contrast, the process model uses one parameter set only for both cases and therefore produces internally consistent results. That the process model parameters were estimated a priori rather than optimised is not essential but arguably preferable.

In summary, if the tests described above would be successful, we would be able to conclude that the paradoxical results of top-down analyses can be reproduced, and appear to be at least partly due to methodological problems. The subsequent analyses were designed to try and analyse three potential methodological problems, *viz.*: measurement errors, an overriding influence of other environmental factors, and covariance between land cover and climate.

Deleted: If

Deleted: both

Deleted: above

Deleted: were

Deleted: indeed

Deleted: reconciled

Deleted: in the application of the top-down method

2. Are measurement errors responsible?

One explanation for the reduced or absent land cover impact inferred from catchments with mixed land cover is the possible impact of measurement results. *P*, *PE*, *Q* and forest cover fraction (*FC*) are all prone to estimation errors. In principle, this could affect values for the two Zhang model parameters that were optimised. To test for this, we performed a synthetic experiment in which measurement 'noise' was added to the streamflow estimates produced by the process model (Q_{sim}). (We did not use the actually observed streamflow as this already contained measurement noise). First, a simulated measurement error of 10% was added to all 278 original values of *FC* and mean *P*, *PE* and Q_{sim}. The errors were drawn independently for each variable and each catchment. For *FC* an error was added that was drawn from a normal (Gaussian) distribution with mean of zero and standard deviation of 0.1; the result was limited within the range 0 to 1. The values of *P*, *PE* and Q_{sim} were multiplied with a factor drawn from a normal distribution with mean of one and standard deviation of 0.1. Next, the two Zhang model parameters were optimised to the resulting 'noisy' *FC*, *P*, *PE* and Q_{sim} values for all 278 combined. This experiment was repeated 3000 times, each time with a sample of 278 catchments. The resulting 3000 pairs of *w* values were compared to those fitted

Deleted: model

to the original *FC*, *P, PE* and Q_{sim} values (without added noise), to assess whether measurement noise led to parameter values suggestive of a smaller than expected land cover influence.

3. Are additional environmental factors responsible?

The premise of the Budyko framework is that mean *P* and *PE* are the main determinants of streamflow. Beyond this, however, other climate factors or terrain factors may be more important than land cover category. To investigate this possibility, we analysed the AWRA-L simulations for the forest and herbaceous scenarios using the Zhang model. For each catchment, we calculated the model parameter (*w*) value corresponding to the streamflow simulated for each land cover scenario (i.e., full forest or full herbaceous cover) using the following inverted model form (cf. Eq. (2)):

$$w = \frac{\dfrac{P}{Q_{sim}(scenario)} - \dfrac{P}{PE} - 1}{\left(\dfrac{PE}{P}\right)^2} \tag{4}$$

For each land cover category, we attempted to find catchment attributes that could explain the variance in inferred *w* values. We used the same step-wise regression approach used in earlier analyses of the same streamflow data (Van Dijk, 2009; 2010a). In summary, candidate predictors were selected from a range of catchment attributes based on the parametric and non-parametric (ranked) correlation coefficients (*r* and r^*, respectively). Linear, logarithmic, exponential and power regression equations were calculated for all potential predictors, and the most powerful one selected. The residual variance was calculated and the same procedure was repeated. The catchment attribute data available included measures of catchment morphology (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 *P,* mean *PE,* humidity index *P*/*PE,* remotely sensed actual evapotranspiration, average monthly excess precipitation); and land cover characteristics (fraction woody vegetation, fractions non-agricultural land, grazing land, horticulture, and broad acre cropping, remotely sensed vegetation greenness). Full details on data sources and catchment climate, terrain and land cover attributes can be found in Van Dijk (2009; 2010a).

4. Is covariance between land cover and climate responsible?

Our catchment data set shows a modest amount of covariance between forest cover (*FC*) and *P*/*PE* (*r*=0.44). Earlier analyses showed that this can affect the ability to accurately determine land cover influence (see van Dijk et al., 2007, for a detailed example). We performed a further synthetic experiment using the AWRA-L model to test the magnitude of this problem:

Deleted: analysis

1) Each of the 278 catchments was assigned a new ‘virtual land cover’ by randomly drawing a new value for *FC* from a normal distribution with the same mean and standard deviation as the observed *FC* values (0.284 and ±0.224, respectively). Values were truncated to remain within the range 0 and 1.
2) For each catchment, the AWRA-L model was run with the new *FC* values and the original meteorological inputs.
3) The two Zhang model parameters were fitted to the resulting 278 Q_{sim} values.

The experiment was repeated 3000 times (each time with all 278 catchments), and the results were analysed to determine whether there was a relationship between any (randomly introduced) covariance between the *FC* and *P*/*PE* values on the one hand, and the inferred land cover influence on the other.

3. Results

3.1 The paradoxical results can be reproduced and reconciled by the process model

Deleted: *land cover*

Indicators of the agreement between Q observed in the 278 catchments and values estimated by the optimised two-parameter Zhang model (Eq. (3)) and the AWRA-L model are listed in Table 1. For comparison, the performance of the originally proposed Zhang-A and Zhang-B models and an optimised Zhang model (Eq. (2)) are also shown.

[TABLE 1 HERE]

Calibrating the Zhang model parameters led to an improvement in model performance and reduction in bias, when compared to the original models. However, reducing the Zhang model to a one-parameter model (that is, making the model insensitive to land cover), did not degrade model performance (optimised values were *w(forest)*=1.91 and *w(herbaceous)*=1.98 versus *w*=1.95, respectively). These results support previously published result that fitting a Budyko model to observations from non-experimental catchments does not show the

Deleted: confirm

expected land cover signal. In other words, we were able to reproduce previously found paradoxical results in this synthetic experiment.

Deleted: could
Deleted: the
Deleted: land cover

Table 1 also shows that, despite the lack of parameter optimisation, AWRA-L performs slightly better than the calibrated Zhang models. The AWRA-L predictions of mean streamflow for the same 278 catchments, but this time for a hypothetical scenario of full forest and herbaceous cover, are compared to the original Zhang-A and Zhang-B model in Fig. 2. AWRA-L is able to reproduce the approximate differences between forest and herbaceous catchments predicted by the original Zhang models, although the forest scenario predictions agree better with the Zhang-B model than with the Zhang-A model (Fig. 2). It follows that the process model (1) can satisfactorily predict streamflow from the 278 catchments with mixed land cover, and (2) produces a land cover signal of similar magnitude as captured by the Zhang et al. (2001) models. Therefore, the process model can reconcile the paradoxical results of the top-down analysis.

Deleted: non-
Deleted: accurately
Deleted: can reproduce
Deleted: the
Deleted: observed in catchment experiments
Deleted: ,

[FIGURE 2 HERE]

As further evidence, the paradoxical results could also be reproduced by top-down analysis of the process model streamflow estimates. If a one-parameter Zhang model was fitted to the modelled Q_{sim} with hypothetical full forest or herbaceous cover, w values 3.6 and 1.0 where found, respectively – producing curves quite similar to the original Zhang-A and Zhang-B models. However, when the two-parameter Zhang model was fitted to the Q_{sim} obtained with actual FC values, the resulting values were much closer, at 2.22 and 1.79, respectively, predicting only a very small land cover signal (average forest water use is only 2% greater than herbaceous water use). This shows that the paradoxical results can also be reproduced with idealised, modelled streamflow data.

Deleted: land cover

3.2 Measurement errors are at least partly responsible

The introduction of noise in the data led to higher average optimised w values: 2.7 (range 0.6-9.4) for forest and 2.3 (1.3-9.2) for herbaceous cover. Probably more importantly, however, for 39% of the 3000 replicates the optimised w value for forest was actually lower than for herbaceous cover. It follows that random errors in the observations are likely to reduce the detectable influence of land cover on streamflow.

3.3. Underlying climate factors may be responsible

The distribution of *w* values calculated from simulated streamflow for individual catchments appeared approximately log-normally distributed and therefore all values were log-transformed before step-wise regression analysis. The ratio *P*/*PE* itself did not explain significant variance in either land cover scenario ($r^2<0.04$).

Deleted: =0.1–

Deleted: 2

Somewhat unexpectedly, the most powerful predictor of variation in *w* values varied between the forest and herbaceous cover scenarios. In the full forest cover scenario, *PE* itself explained 45% of the variance in log-transformed *w* values (see Fig. 3a). Other predictors did not explain any of the residual variance. In the full herbaceous cover scenario, depth-weighted average event precipitation (DWAEP, calculated as the sum of squared daily rainfall totals divided by total rainfall) explained 33% of the variation (Fig. 3b), whereas mean event precipitation (total rainfall divided by the number of rain days) explained 27% of variation (instead of, not in addition to the variation explained by DWAEP). Both are indicators of the irregularity of rainfall distribution (see Van Dijk, 2009 for definitions). Other predictors did not explain any of the residual variance.

It is concluded that other climate factors than *P*/*PE* alone may have considerable influence on catchment mean streamflow and be expressed in *w* values.

Deleted: can

Deleted: response

[FIGURE 3 HERE]

3.4 There is structure in the data set that is at least partly responsible

Using streamflow simulated for randomly generated hypothetical forest cover fractions (*N*=3000), Zhang model parameter values of 3.4±0.7 (range 1.9–6.1) and 1.1±0.1 (0.9–1.4) were fitted for forest and herbaceous cover, respectively. These average values are relatively close to the *w* values of 3.6 and 1.0 fitted for the full forest and herbaceous cover scenarios (experiment 1). In some experiments the optimised Zhang parameters were similar to the 'full cover' ones, whereas in other experiments they were very close (Fig. 4a) (it is noted that *w(herbaceous)* never exceed *w(forest)*, unlike in the measurement error experiment). It would be tempting to conclude that the covariance between *FC* and *P*/*PE* in the original data set (*r*=0.44) was the main cause for the underestimation of land cover influence. However, no relationship was found between the fitted parameter pair and the covariance between forest cover and *P*/*PE* that was introduced into the data set (Fig. 4a). Nonetheless, the manipulation of the data must have introduced another form of hidden structure in the data set that affected the optimised parameter values.

Deleted: At first instance, i

Deleted: seem

Deleted: and

4. Discussion

4.1 The paradoxical results can be reproduced and methodological issues are likely to be responsible

Deleted: *'land cover*

Deleted: *'*

Despite their simplicity, Budyko models have shown impressive skill in predicting mean catchment Q from *P* and *PE* alone, when compared to more complex dynamic catchment models. Indeed in comparison with the more complex AWRA-L model, the Zhang model could achieve very similar performance in explaining the observed catchment streamflow averages, although only after calibration. It was this same calibration, however, that produced land cover parameter values that could not be reconciled with the results of experimental catchment studies, thus reproducing the paradoxical results from previous studies. Our results demonstrate that a dynamic hydrological process model can reconcile these results, and therefore, that there appear to be methodological problems with the use of Budyko models as a detection method in this application.

Deleted: found in

Deleted: this paradox

Deleted: it

Deleted: is likely to be a

Deleted: rather than a physical reality

The synthetic experiments demonstrated that all methodological issues tested for (measurement errors, the presence of other important uncontrolled factors, structure in the catchment data set) can contribute to the failure to accurately quantify land cover influence with the Budyko model used. In all cases, underestimation of the land cover signal was the most likely result. Desirable aspects of Budyko models are their conceptual simplicity and the minimal number of parameters. However, in qualifying the principle of Occam's Razor, Albert Einstein (1934) proposed that "*the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience*". On the basis of our results we conclude that, for the purpose at hand, Budyko models fail at the second part of this statement; that is, they are too simple to adequately quantify the influence of land cover in collated data sets of streamflow from catchments with mixed land cover.

Although we only tested one particular Budyko model, previous studies suggest that conclusions would likely have been very similar if any other Budyko model had been used, due to the identical conceptual structure and similar function form (see e.g. Oudin et al., 2008). Moreover, we argue that the methodological issues with heterogeneous data sets such as the one we analysed are not limited to the application of Budyko models but are likely to prevent accurate detection with other top-down approaches as well.

There have been attempts to increase the predictive performance of the Budyko models by including additional variables, often within a stochastic framework (e.g. Porporato et al., 2004). Those not related to land cover include absolute *PE* values (Peel et al., 2010), solar radiation, phase differences between the seasonal *P* and *PE* patterns (Donohue et al., 2010), and the daily distribution of precipitation (see review in Gerrits et al., 2009). Our results suggest that some of these factors may indeed exert a similarly large or larger influence on catchment response. However, trying to control for these additional factors introduces further parameters and observations with associated uncertainty, and ultimately such an approach must fall prey to the very issue that top-down approaches are intended to avoid, that is, creating an underdetermined (or undetermined) problem in which competing hypotheses create similar outcomes and therefore cannot be tested.

This is obviously not avoided by the use of dynamic process models. However such models are arguably more suitable to make process assumptions more explicit and allow these to be tested against different types of observations with different spatial and temporal characteristics. In light of this, we question whether it is advisable to calibrate any hydrological model to heterogeneous data sets such as the one analysed here. Arguably, it is sufficient to demonstrate that the observations can be explained satisfactorily by a (more complex) theory and therefore are not falsified by experimental knowledge. In this context, the Budyko framework may be a valuable benchmark test, whose predictive power should be matched or exceeded by a competing theory (cf. Van Dijk and Warren, 2010). It is however perhaps less useful as a suitable inference method to detect second order processes in complex data sets.

Strictly speaking, our results are only valid for one particular data set. However, all factors we investigated negatively affected accurate quantification of the land cover signal. We consider it inevitable that at least some of these aspects (e.g. measurement errors, mixed land cover) will be encountered in any heterogeneous streamflow data set from large catchments with mixed land cover. Zhang et al. (2001) showed that this need not prevent detection of land cover impacts in data from catchments that represent 'extreme' scenarios and in controlled experiments. In particular paired catchment experiments are much more likely to adequately control for climate and terrain factors and thereby allow accurate quantification of the land cover influence. Apart from experimental issues associated with such necessarily small-scale experiments (such as subterranean leakage), a critical issue in the extrapolation

Deleted: s
Deleted: reproduced
Deleted: can be reconciled
Deleted: with
Deleted: can
Deleted: seen as
Deleted: a successful
Deleted: should be able to reproduce or exceed
Deleted: , but
Deleted: not

of the results from such experiments will be the degree to which hydrological processes and land cover characteristics are representative for those in larger, non-experimental catchments (see van Dijk and Keenan, 2007 for a discussion). More complex process models probably have a role to play here, as the influence of such representational errors may be investigated in model experiments.

4.2 The role of physical causes for the paradoxical result

We did not explore physical causes of the inability to adequately detect a land cover signal in previous Budyko model applications in large catchment data sets, but they may also play a role. The AWRA-L model was not considered suitable to explore all potential processes in-depth; for example, it does not simulate land surface-atmosphere feedbacks, impacts of human interferences such as farm dams, roads and soil management, and redistribution of water through overland and subsurface flow within hill slopes. The model does describe some other potential feedback mechanisms, including evaporation from streams and riparian areas and (in an implicit manner) the lateral redistribution of groundwater. The role of these in simulations can be evaluated by comparing Q_{sim} values generated with observed forest cover to estimates calculated as the weighted average of Q_{sim} for the extreme land cover scenarios. The former were on average 10% smaller than the latter, representing an average 1.4% of catchment rainfall and 1.7% of catchment streamflow. In other words, within the model structure there is indeed scope for water not used in herbaceous areas to be evaporated in second instance from forest areas or the drainage network, thereby attenuating land cover influences. We are not able to validate the magnitude of the simulated fluxes against experimental data however.

Consideration of the main causes of simulated hydrological changes associated with land cover change provides some further insight into reasons why large catchments with mixed land cover might behave differently from small homogenous ones. It appears that the main cause of the different hydrological response is predicted to be the greater rainfall interception losses from forest vegetation (Fig. 5). The approximate difference represents around 10-15% of rainfall, which is consistent with published experiments (e.g. Roberts, 1999) although much greater differences can occur under maritime conditions (e.g. Schellekens et al., 1999; McJannet et al., 2007). *A priori* it would seem plausible that that the associated rapid return of moisture to the atmosphere may influence rainfall generation downwind (cf. D'Almeida et al., 2007; Pielke et al., 2007; van Dijk and Keenan, 2007). If this is indeed the case, then accurate prediction of the influence of land cover change on the water balance of large catchments may depend on the spatial distribution of precipitation and how it is measured and represented in models.

[FIGURE 4 HERE]

5. Conclusions

Although land cover is known to affect the water balance, attempts to quantify a similar influence in collated streamflow data from catchments with mixed land cover have not been successful. We conclude that this paradox is probably at least partly a consequence of methodological problems in the use of top-down methods to analyse such data sets. More specifically:

Deleted: ‘land cover

Deleted: ’

1. Budyko models are too simplistic to adequately detect and quantify the influence of land cover in complex collated data sets of streamflow from catchments with mixed land cover, due to the measurement and estimation errors, additional climate factors, and the heterogeneous and structure nature of the data.

2. Using a dynamic hydrological process model, we were able to reconcile streamflow response from 278 catchments with mixed land cover with experimental knowledge. This emphasises that the absence of proof (from the top-down analysis method presented here) does equal the proof of absence of land cover influence.

Deleted: evidence

Deleted: s

3. At least some of these methodological issues are likely to be found in any heterogeneous streamflow data set from catchments with mixed land cover.

4. There are reasons to suspect there may also be physical causes for the failure to adequately detect a land cover signal in large catchments. This includes the possibility of atmospheric feedback mechanisms associated with rainfall interception.

Acknowledgements

This work is part of the water information research and development alliance between the Bureau of Meteorology and CSIRO's Water for a Healthy Country Flagship. Comments from Lu Zhang and Randall Donohue are gratefully acknowledged.

References

Andréassian, V.: Waters and forests: from historical controversy to scientific debate, J. Hydrol., 291, 1-27, 2004.

Bosch, J. R. and Hewlett, J. D.: A review of catchment experiments to determine the effect of vegetation change on water yield and evapotranspiration, J. Hydrol., 55, 3–22, 1982.

Brown, A., Zhang, L., McMahon, T., Western, A., and Vertessy, R.: A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, J. Hydrol., 310, 28-61, 2005.

Bruijnzeel, L.: Hydrology of moist tropical forests and effects of conversion: a state of knowledge review, International Hydrological Programme of UNESCO, Paris / Vrije Universiteit, Amsterdam, 224 pp., 1990.

Deleted: Faculty of Earth Sciences, VU University,

Deleted: Amsterdam.¶

Bruijnzeel, L.: Hydrological functions of tropical forests: not seeing the soil for the trees?, Agric, Ecos. & Env., 104, 185-228, 2004.

Budyko, M. I.: Climate and Life Academic Press, New York, 508 pp., 1974.

Calder, I. R.: Forests andWater – Ensuring benefits outweigh water costs, For. Ecol. Man., 251, 110–120, 2007.

D'Almeida, C., Vörösmarty, C. J., Hurtt, G. C., Marengo, J. A., Dingman, S. L., and Keim, B. D.: The effects of deforestation on the hydrological cycle in Amazonia: a review on scale and resolution, Int. J. Clim., 27, 633-647, 10.1002/joc.1475, 2007.

Donohue, R. J., Roderick, M. L., and McVicar, T. R.: On the importance of including vegetation dynamics in Budyko's hydrological model, Hydrol. Earth Syst. Sci., 11, 983-995, 2007.

Donohue, R. J., Roderick, M. L., and McVicar, T. R.: Can dynamic vegetation information improve the accuracy of Budyko's hydrological model?, J. Hydrol., 390, 23-34, 2010.

Einstein, A.: On the method of theoretical physics, Philosophy of Science, 1, 163-169, 1934.

Farley, K., Jobbágy, E., and Jackson, R.: Effects of afforestation on water yield: a global synthesis with implications for policy, Glob. Ch. Biol., 11, 1565-1576, 2005.

Gash, J. H. C.: An analytical model of rainfall interception by forests, Q. J. Royal Met. Soc., 105, 43-55, 1979.

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, Water Resour. Res., 45, W04403, 10.1029/2008wr007308, 2009.

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.

Guerschman, J. P., Van Dijk, A., Mattersdorf, G., Beringer, J., Hutley, L. B., Leuning, R., Pipunic, R. C., and Sherman, B. S.: Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia, J. Hydrol., 369, 107-119, 2009.

Holmes, J. W., and Sinclair, J. A.: Water yield from some afforested catchments in Victoria, in: 17th Hydrology and Water Resources Symposium, Brisbane, Institution of Engineers, Australia, Barton, ACT, Australia, 1986.

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, Env. Mod. and Softw., 16, 309-330, 2001.

Klemeš, V.: Conceptualization and scale in hydrology, J. Hydrol., 65, 1-23, 1983.

McJannet, D., Wallace, J., Fitch, P., Disher, M., and Reddell, P.: Water balance of tropical rainforest canopies in north Queensland, Australia, Hyd.Proc., 21, 3473-3484, 2007.

Milly, P.: Climate, soil water storage, and the average annual water balance, Water Resour. Res., 30, 2143-2156, 1994.

Monteith, J. L.: Evaporation and environment Symposia of the Society for Experimental Biology, 19, 205-224, 1965.

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263, 641, 1994.

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., McMahon, T., and Finlayson, B.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, Water Resour. Res., 46, W09508, doi:10.1029/2009WR008233,2010.

Pielke, R. A., Adegoke, J., Beltrán-Przekurat, A., Hiemstra, C. A., Lin, J., Nair, U. S., Niyogi, D., and Nobis, T. E.: An overview of regional land-use and land-cover impacts on rainfall, Tellus B, 59, 587-601, 10.1111/j.1600-0889.2007.00251.x, 2007.

Porporato, A., Daly, E., and Rodriguez-Iturbe, I.: Soil water balance and ecosystem response to climate change, American Naturalist, 164, 625-632, 2004.

Roberts, J.: Plants and water in forests and woodlands, in: Ecohydrology, Routledge, London, UK, 1999.

Schellekens, J., Scatena, F. N., Bruijnzeel, L. A., and Wickel, A. J.: Modelling rainfall interception by a lowland tropical rain forest in northeastern Puerto Rico, J. Hydrol., 225, 168-184, 1999.

Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hyd.Proc., 17, 2101-2111, 2003.

Stednick, J. D.: Monitoring the effects of timber harvest on annual water yield, J. Hydrol., 176, 79–95, 1996.

Trimble, S. W., Weirich, F. H., and Hoag, B. L.: Reforestation and the reduction of water yield on the Southern Piedmont since circa 1940, Water Resour. Res. 23, 425–437, 1987.

Turner, K. M.: Annual evapotranspiration of native vegetation in a Mediterranean-type climate, J. Am Water Resour. Assoc., 27, 1-6, 1991.

Van Dijk, A.I.J.M. and Renzullo, LJ..: Water resource monitoring systems and the role of satellite observations, Hydrol. Earth Syst. Sci., 15, 39-55, 2011.

Van Dijk, A. I. J. M. and Bruijnzeel, L. A.: Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 1. Model description, J. Hydrol., 247, 230-238, 2001a.

Van Dijk, A. I. J. M. and Bruijnzeel, L. A.: Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 2. Model validation for a tropical upland mixed cropping system, J. Hydrol., 247, 239-262, 2001b.

Van Dijk, A. I. J. M., Hairsine, P. B., Arancibia, J. P., and Dowling, T. I.: Reforestation, water availability and stream salinity: A multi-scale analysis in the Murray-Darling Basin, Australia, For. Ecol. Man., 251, 94-109, 2007.

Van Dijk, A. I. J. M. and Keenan, R. J.: Planted forests and water in perspective, For. Ecol. Man., 251, 1-9, 10.1016/j.foreco.2007.06.010., 2007.

Van Dijk, A. I. J. M.: Climate and terrain factors explaining streamflow response and recession in Australian catchments, Hydrol. Earth Syst. Sci., 14, 159-169, 2009.

Van Dijk, A. I. J. M.: Selection of an appropriately simple storm runoff model, Hydrol. Earth Syst. Sci., 14, 447-458, 2010a.

Van Dijk, A. I. J. M.: AWRA Technical Report 3, Landscape Model (version 0.5) Technical Description, WIRADA/CSIRO Water for a Healthy Country Flagship, http://www.clw.csiro.au/publications/waterforahealthycountry/2010/wfhc-aus-water-resources-assessment-system.pdf, (last access: 8 March 2011), Canberra, 2010a.

Van Dijk, A. I. J. M. and Warren, G.: AWRA Technical Report 4, Evaluation Against Observations, WIRADA/CSIRO Water for a Healthy Country Flagship, http://www.clw.csiro.au/publications/waterforahealthycountry/2010/wfhc-aus-water-resources-assessment-system.pdf, (last access: 8 March 2011), Canberra, 2010.

Zhang, L., Dawes, W. R., and Walker, G. R.: Predicting the effect of vegetation changes on catchment average water balance. Technical Report 99/12, Cooperative Research Centre for Catchment Hydrology, Canberra, 1999.

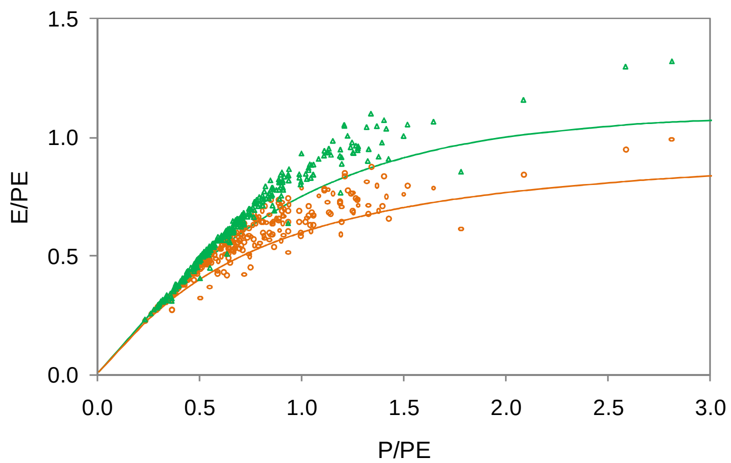
Zhang, L., Dawes, W. R., and Walker, G. R.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water Resour. Res., 37, 701–708, 2001.

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, doi:10.1029/2003WR002710 2004.

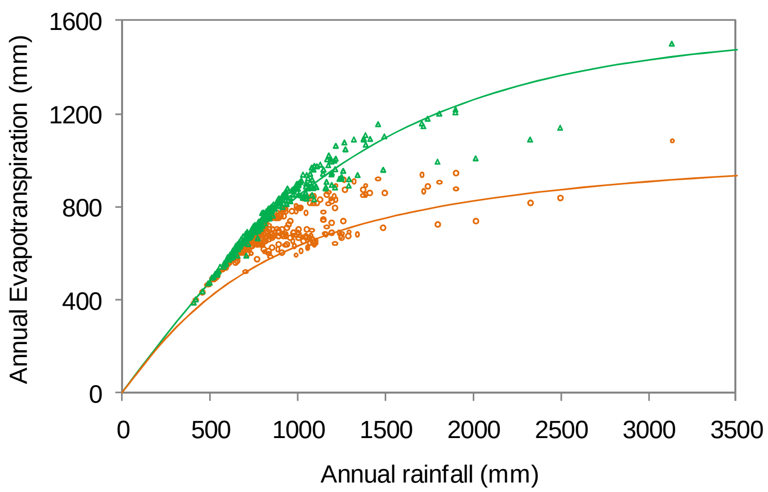
Table 1. Performance indicators of the original Zhang et al. (2001) models (Zhang-A and Zhang-B; see text for explanation), the Zhang model with one and two calibrated parameters, respectively, and the AWRA-L with prior parameter estimates. SEE=standard error of estimate, MAE=mean absolute error, and Bias=mean bias (all in mm year^{-1}); Rel. Bias=mean relative bias and FOM=fraction of values overestimated by model (in %).

	SEE	MAE	Bias	Rel. Bias	FOM
Zhang-A	119	97	79	44%	91%
Zhang-B	136	114	86	47%	86%
Zhang - 2 parameter	84	54	4	2%	62%
Zhang - 1 pararameter	84	54	4	2%	62%
AWRA-L	78	50	1	1%	54%

Fig. 1. Location of the 278 Australian catchments for which streamflow data were used in the analysis.

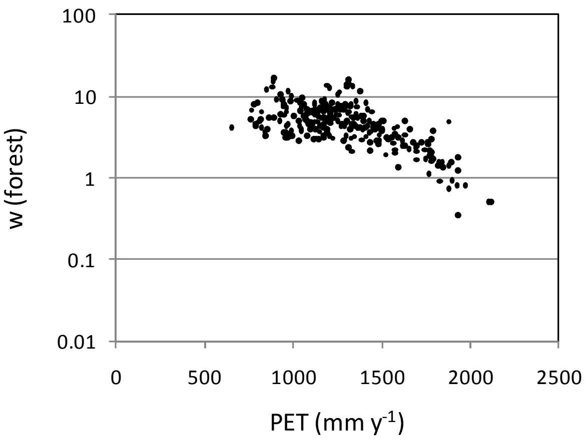


a)

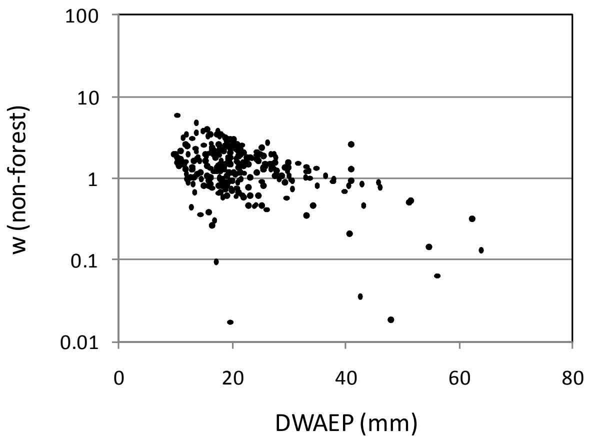


b)

Fig. 2. Comparison of AWRA-L simulated streamflow for the 278 catchments for scenarios of forest cover (green triangles) and herbaceous cover (orange circles) shown in two different ways. Also shown are the two models proposed by Zhang et al. (2001): **(a)** Zhang-A and **(b)** Zhang-B.

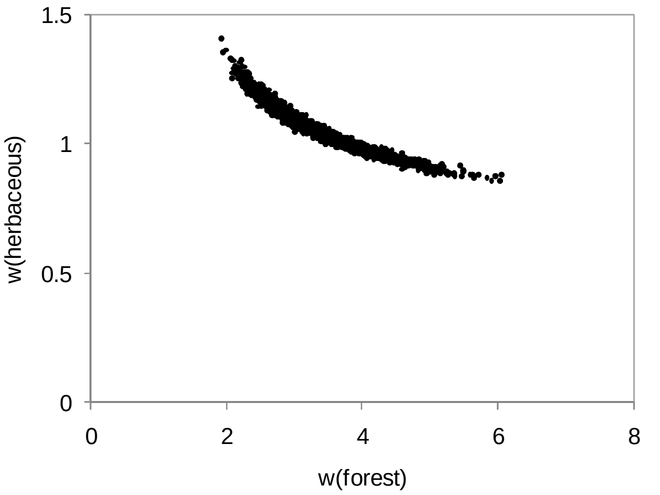


a)

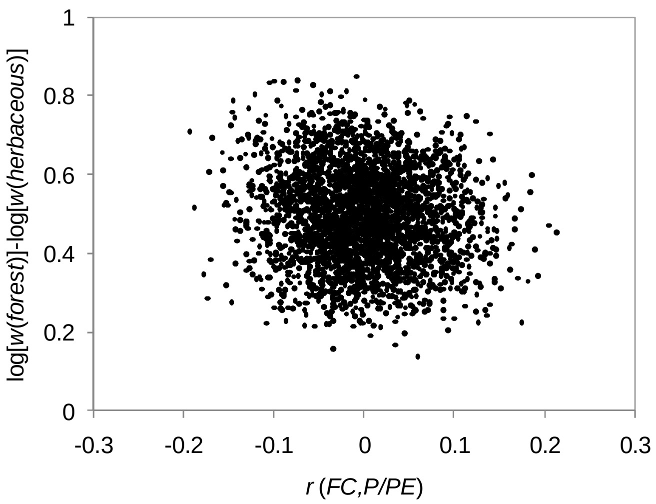


b)

Fig. 3. Relationship between the catchment variable that explained most of the variance in (log-transformed) Zhang model parameter (*w*) values inferred from the synthetic land cover experiment, **(a)** potential evaporation (PE) for forest catchments and **(b)** depth-weighted average event precipitation (DWAEP) for herbaceous catchments.



a)



b)

Fig. 4. Zhang model parameter values fitted to synthetic streamflow estimates for 278 catchments produced by AWRA-L with random forest cover fractions assigned to each of the catchments. Data points represent the results of 3000 replicate experiments. **(a)** Zhang model parameter data pairs fitted in each experiment showing a well-defined relationship; **(b)** the difference between log-transformed parameter values versus the correlation between synthetic forest cover fraction (*FC*) and catchment humidity (*P*/*PE*) introduced in the experiment, showing no relationship (*r*=0.11).

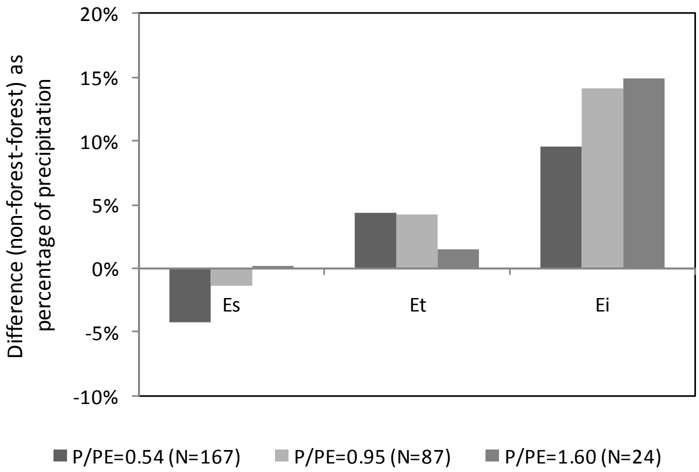


Fig. 5. Contribution of different evaporation terms to the increase of streamflow after forest removal estimated by the AWRA-L model, expressed as a percentage of rainfall. Values represent fluxes averaged over three groups of catchments, intended to represent (from left to right) water-limited (*P/PE*<0.75), transitional, and energy-limited (*P*/*PE*>1.25) environments. *Es*=soil and open water evaporation; *Et*=transpiration; *Ei*=rainfall interception losses.