Generalizing the impact of forest cover on streamflow from experimental data: it is not that simple.

R. Willem Vervoort[a,b], Eliana Nervi[c], Jimena Alonso[d]

[a] *Sydney Institute of Agriculture and School of Life and Environmental Sciences The University of Sydney Sydney NSW 2006 Australia*
[b] *ARC ITTC in Data Analytics for Resources and Environments The University of Sydney NSW 2006 Sydney Australia*
[c] *Project Manager FPTA 358 Instituto Nacional de Investigacion Agropecuaria INIA-Uruguay Ruta 48 km 10 Rincon del Colorado 90100 Canelones Uruguay*
[d] *Institute of Fluid Mechanics and Environmental Engineering School of Engineering Universidad de la Republica 11200 Montevideo Uruguay*

Abstract

Three recent papers review and analyze large global datasets related to impacts of forest cover on streamflow. Using three different approaches, they all find a strong relationship between forestation, de-forestation and streamflow. However, the results are problematic, the underlying data set is unbalanced, and there are correlations in the data that warrant further investigation as this would influence the results. For example, the area of the catchment is strongly related to the assessment technique and the variability in the response data. For this study, the data for the recent three papers were reviewed, combined, and supplemented with new studies. Subsequently, the data were re-analyzed using generalized additive modelling. The results highlight that there are four interlinked reasons that make the general outcomes from the previous papers problematic: 1) The existence of latent variables in the data that create the appearance of a relationship that really does not exist; 2) The difficulty in fully interpreting the specifics of different studies; 3) The difficulty of integrating data from seemingly similar studies, but with quite different objectives; and 4) The chance of transcription errors influencing the data. Overall this indicates that while valuable data can be extracted from past studies, the above problems need

*Corresponding author
Email addresses: willem.vervoort@sydney.edu.au (R. Willem Vervoort), eliananervif@gmail.com (Eliana Nervi), jalonso@fing.edu.uy (Jimena Alonso)

to be considered before results are generalized and extrapolated to continental and global scales.

Keywords: meta analysis, forest cover change, global scales, statistical modelling

1. Introduction

There is a urgent need to identify the impacts of human intervention on stream flow at a global scale and to separate this from climate effects [38, 18]. More specifically, the impacts of global deforestation and reforestation are important through their perceived influence on stream flow and blue and green water availability [18, 30]. The past work reviewing these impacts [2, 21, 44, 9, 10, 16] highlights a general consensus that if forest areas increase, stream flow decreases and vice-versa. The most dramatic example of this is Figure 5 in Zhang et al. [42] indicating (for Australian catchments) a 100% decrease in stream flow for catchments with 100% forest cover. However, on the other end of the spectrum, for three French catchments [13], there was no change in stream flow characteristics in two of the catchments after deforestation. For reforestation, a modelling study across the 1 million km^2 Murray Darling Basin also found no major effect, especially in larger catchments [35], but a recent study [18] found an 8% change in stream flow as a result of reforestation. Similarly a modelling study by Beck et al. [3] found no significant change in stream flow in 12 catchments in Puerto Rico as a result of deforestation. In contrast, in a recent study in Brazil across 324 catchments, Levy et al. [24] found a significant increase in stream flow, particular in the dry season, as a result of deforestation. This suggests that there can be significant variation across the different studies, methodologies and geographical regions.

For the purpose of this paper, *watershed* and *catchment* are interchangeable terms. Many of the US studies use *watershed*, while European and Australian studies use *catchment*. In particular, we retained the term "paired watershed studies" and "quasi-paired watershed studies" as this is the most common terminology, but further mostly use the term catchment.

There has been a recent push in the hydrological community [14] to use 'meta-analysis' to summarize past studies. The suggestion is that, because meta-analyses use clearly defined search terms and statistical methods to analyze the results, this will lead to more reliable summaries of past research. As a result, several review papers have summarized the plethora of forestation and deforestation studies across the globe, in relation to paired watershed studies [9, 8], related to reforestation in particular [16], and more generally [21, 44]. These studies aim to generalize the individual experimental and research findings and to identify if there are global trends or relationships. Others have used the understanding from a global analysis to extrapolate to global scales [18].

The recent paper by Filoso et al. [16] is a clear meta-analysis, but most others [44, 18, 47] are not. However, n impressive global database of catchment studies with changes in stream flow due to changes in forest cover has been developed [44, 16] and statistical approaches are used to analyze the resulting data. The Zhang et al. [44] data set, which covers over 312 studies, is described in terms of the change in stream flow as a result of the change in forest cover, where studies related to both forestation (increase in forest cover) and deforestation (decrease in forest cover) were included. In contrast, the paper by Filoso et al. [16] focused primarily on reforestation, and covered an equally impressive database of 167 studies using a systematic review. In this case the collected data is mostly coded as count data and only a subset of 37 studies was analyzed for actual water yield change. There is some overlap between the two data sets, but there are also some studies unique to both sets. The more regionally concentrated and detailed study by Levy et al. [24] is a further independent data set with no overlap with the other studies. However, for this study only the flow and rainfall data is available for the catchments, and the change in land cover was derived from satellite data and was not made available.

The conclusions of the first mentioned major review paper [44] indicates that there is a distinct difference in the change in flow as a result of forestation or deforestation between small watersheds (catchments), defined as < 1000 km^2 and large watersheds (catchments) > 1000 km^2. While for small catchments there was no real change in runoff with changes in cover, for large catchments there was a clear trend showing a decrease in runoff with increases in forest cover. The main conclusion was that the response in annual runoff to forest cover was scale dependent and appeared to be more sensitive to forest cover change in water limited catchments relative to energy limited catchments [44].

The second study [16] is a systematic review of reforestation studies (only studies in which forest cover increased). This study classified the historical research and highlighted gaps in the spatial distribution, the types of studies and the types of analysis. Their main conclusion was also that reforestation decreases stream flow, but that there were many interacting factors. For a subset of the data (37 data points) they also indicated decreasing impacts of reforestation with increasing catchment size (agreeing with Zhang et al. [44]), but they did not identify a distinct threshold and fitted a log-linear relationship. In addition, they identified that studies with shorter periods of data collection resulted in larger declines in stream flow.

An earlier paper, that includes much of the same data as Zhang et al. [44] and Filoso et al. [16], is Zhou et al. [47], which has one author in common with Zhang et al. [44]. However, this paper aims to explain the variation in the data using the elasticity approach in the Fuh model, which is similar to well-known Budyko approaches [43]. In particular, it aims to link the variation in the observed data to variations in the exponent m in the Fuh model, which represents vegetation cover. A key observation is that in drier environments, the effects of removing forest cover are much greater than in wetter environments, which is also suggested by Figure 4 in Zhang et al. [44]. The Fuh model and the related variations of the Budyko equilibrium modelling approach was also used by Hoek van Dijke et al. [18] to interpret the global impact of reforestation.

However, concerning is that there are some clear limitations in these studies, and some of this applies to meta-analyses in general. The main method in the work by Zhang et al. [44] is a single covariate linear regression. In contrast, the systematic review from Filoso et al. [16] mainly emphasizes the classification and distributions of the study. Zhang et al. [44] points out that a main assumption in their work is that the catchment size threshold at 1000 km^2 is a distinct separation between "small" and "large" catchments. However, a subset of 37 data points in Filoso et al. [16] (their Figure 9) does not appear to support this, suggesting a continuum. And while the work Filoso et al. [16] provides important insights in study types, analysis types, forest types and broad classification, there is limited quantification of actual impact.

In contrast to the single covariate linear regression in the earlier studies [44, 16] and the top-down Budyko modelling [47, 18], the regional Brazilian Cerrado study [24] provides an example of an carefully designed statistical approach using mixed effects modelling and Differences-in-Differences modelling focusing specifically on the effect of deforestation. The analysis specifically accounted for differences between catchments and differences due to variations in climate. Not all data sets are however suitable for this kind of in-depth analysis.

Given all these previous reviews and the seemingly clear conclusions about the impact of forest cover change on stream flow, the question is why another review paper on this topic? There is a real attraction in the concept of statistical analysis of past studies encapsulated in meta-analysis to be able to extrapolate findings to larger scales, and to identify factors across global scales [14]. However, there are also some hidden complications in this that can invalidate results, which this paper aims to highlight. There are four potential errors (or limitations) in such global meta-analyses:

- Impact of latent variables that are not included in the typical single covariate analysis;
- Interpretation errors due to incomplete descriptions of the experiments in the original papers;
- Aggregation of data that originates from different experiments with different objectives across a wide time period, but have similar keywords; and, finally
- Transcription errors in the data, especially if data is collected from other review papers as some of the original papers are difficult to locate.

The aim of this paper is to first reanalyze the global data set [44, 16] using some more detailed statistical modelling and to use this to highlight examples of each of these limitations. This will show how they have influenced the outcomes of the past work, and provide suggestions of how we can overcome these limitations. In addition, by applying more complex statistical models, we will highlight the conclusions that can be drawn from the data. Finally, we will highlight future research needs in the area of forest cover change impact on stream flow.

We are taking advantage of the earlier work by Zhang et al. [44], Filoso et al. [16] and Zhou et al. [47] and the large database of studies these authors have shared.

2. Methods

2.1. The original data set

As indicated, the starting point of this paper is the data base of studies which were included in Zhang et al. [44] as supplementary material. The columns in this data set are the catchment number, the catchment name, the Area in km^2, the annual average precipitation (Pa) in mm, the forest type, hydrological regime, and climate type, the change in forest cover in % ($\Delta F\%$) and the change in stream flow in % ($\Delta Qf\%$), the precipitation data type, the assessment technique, and the source of the info, which is a citation. The change in stream flow ($\Delta Qf\%$) is based on equation 1 in Zhang et al. [44].

Several of these columns contain abbreviations to describe the different variables, which are summarized in Table 1. These abbreviations will later be used in the models.

Table 1: Summary of abbreviations of factors used in the Zhang et al. (2017) data set

Factor	Abbreviation	Definition
forest type	CF	coniferous forest
	BF	broadleaf forest
	MF	mixed forest
hydrological regime	RD	rain dominated
	SD	snow dominated
climate type	EL	energy limited
	WL	water limited
	EQ	equitant
precipitation data type	OB	observed

Factor	Abbreviation	Definition
assessment technique	SG	spatial gridded
	MD	modelled
	PWE	paired watershed experiment
	QPW	quasi-paired watershed experiment
	HM	hydrological modelling
	EA	elasticity analysis
	SH	statistical modelling and hydrographs

The paper by Zhang et al. [44] also uses the dryness index, which is the annual rainfall (Pa) divided by the potential or reference evapotranspiration (ET_0 or E_0) in their analysis, and have used the dryness index to identify the climate type. However, the potential or reference ET used for this calculation was originally not included in the published data set. We will discus below how we derived the dryness index in our data set. We combined the tables for small catchments (< 1000 km^2) and large catchments ($>= 1000$ km^2) from Zhang et al. [44] in our analysis.

2.2. Additional data collection

To enhance the existing data set, this study added additional variables and cross-checked the studies with the data set from Filoso et al. [16]. In particular, we focused on the 37 data points related to the quantitative regression analysis used in Filoso et al. [16].

In addition, a few additional variables were included to enhance the data set. We added latitude and longitude for the center of the catchment as an approximation of its spatial location. Mostly the data reported by the authors was used, but in some cases the variables had to be approximated from the location of the centre of the catchment using Google Maps™. In the data set, an additional column has been added to indicate the source of the location data to indicate if this is directly from the paper or elsewhere.

As highlighted, Zhang et al. [44] did not provide values for evapotranspiration in the data base. Using the location information, reference evapotranspiration (E_0) was extracted from the Global Aridity Index and Potential Evapo-Transpiration (ET_0) Climate Databasev2 [34], if a value of E_0 was not available from the original papers. For large catchments, this value (and the associated coordinates), similar to annual average rainfall, is only an approximation of the climate at the location.

Similar to Zhang et al. [44], the Dryness index was calculated from the catchment estimate of reference evapotranspiration and the catchment estimate of annual average rainfall (Pa) as:

$$Dryness = \frac{E_0}{Pa} \tag{1}$$

The length of the study can be a variable influencing the change in flow [e.g. 21, 16], as for example, more mature plantations are thought to have smaller impacts on flow or regrowth might follow a “Kuczera curve” [23]. It is not clear if this is an effect of increased water use in growth [36] or due to changes in interception [31]. Therefore, the length of the study calculate as the difference between the starting data and completion date of the different studies was extracted from the references provided by Zhang et al. [44]. The length of the study was already included in the data from Filoso et al. [16], but these were checked against the original publications.

Several additional data points from catchment studies were extracted from

Almeida et al. [1], Ferreto et al. [15], Iroumé and Palacios [19], Iroumé et al. [20], Zhang et al. [42], Zhao et al. [45], Borg et al. [7], Thornton et al. [32], Zhou et al. [46], Rodriguez et al. [28], Ruprecht et al. [29] and Peña-Arancibia et al. [26], and these were checked against the existing studies to prevent overlap. In the citation column in the accompanying data set, the main reference for the calculated change in stream flow was generally used, because sometimes the original study did not provide the quantification of the change in stream flow [i.e. Table 6 in 42].

We conducted a thorough review of all the studies mentioned in the data base of Zhang et al. [44] and sourced all the original papers. As a result of this we made several changes to the data base, which are all recorded in Supplementary Data part 1. Overall 36 data points were changed and the most common problem was a change in the sign for the change in forest cover or the change in flow. We assume that these were transcription errors.

We also removed one data point from the data set, which corresponds to catchment #1 (Amazon) in Zhang et al. [44]. This is because the cited reference [27] only relates to 1 and 1.5 ha paired catchment studies in French Guyana, and in which the actual change in forest cover is not recorded. Finally, on review of all the data in Zhang et al. [44] and Filoso et al. [16], 29 potential duplicates were identified and flagged in the data, and not used in the analysis.

The final column in the improved data base is a "notes" column, which we added, but is not further used in the analysis. It gives context to some of the data for future research and highlights some of the discrepancies that we found between the original papers and the data in the tables from Zhang et al. [44]. This will allow future research to scrutinize our input for errors.

2.3. Statistical modelling

The aim of the statistical analysis is to highlight the most important variables in the data set that explain the change in flow as a result of changes in forest cover. This first aim is similar to Zhang et al. [44], but the main difference is that we start off with all variables in the data set in the model. Subsequently the analysis will concentrate on how the individual variables in the data set relate to each other and how latent variables in the data set can be masked and result in relationships that might not really exist. Finally, the analysis will highlight how the results are conditional on the data set.

In the statistical analysis we are not necessarily seeking the best "predictive" model, and as such do not perform a traditional variable selection process. Rather, we focus on analyzing the predictor variables in the full model to identify how all the variables explain the variance in the dependent variable.

To estimate how the change in stream flow is affected by the change in forest cover, while considering the effects of the other variables, we applied generalized additive modelling (GAM) [41].

The general model tested is:

$$\begin{aligned}\Delta Qf\% \sim \Delta\% forest\ cover + \\ \sum X_i + \sum s(Z_i) + \varepsilon\end{aligned} \tag{2}$$

Here X_i are factorial variables, while Z_i are continuous variables. As a first step, the model assumes no direct interactions and that all variables are additive. A further assumption in the model is that all continuous variables Z_i (such as annual precipitation (Pa)) can have either a linear or a non-linear relationship with $\Delta Qf\%$. This means that a smooth function $s()$ can be applied to the Z_i variables. For the smoothing function we applied thin plate regression splines with an additional shrinkage penalty. The result of this approach is that for high enough smoothing parameters (i.e. if the data is very "wiggly") the smooth term can be shrunk to 0 and thus will be no longer significant [41]. This is done because a highly flexible smooth term could always fit the data, but would not necessarily indicate a relevant relationship. In other words, the approach balances finding a smooth non-linear relationship for the variable against over fitting the data.

The changes in forest cover contain both positive (forestation) and negative values (deforestation). In Zhang et al. [44], these changes were jointly analyzed, assuming the effect on the change in flow was linear and the effect of removing forest cover was the same as an equivalent addition of forest cover.

However, the impact of an increase in forest cover can be different from the same fractional decrease in forest cover. The question becomes how best to analyze this. One approach would be to allow a different slope and a different intercept for the decreases relative to the increases. This can be tested by converting all the change in forest cover data to positive values, and an additional binary column ($sign_{forestcover}$) can be included indicating whether it was a forest cover increase or decrease. In the model, the parameter for $sign_{forestcover}$ will indicate the difference in the changes in flow for increases in forest cover compared to decreases in forest cover. The disadvantage of this approach is that the relationship with forest cover becomes discontinuous at the origin (0 change in forest cover).

A second approach is to test the change in forest cover as a non-linear relationship in the GAM model. Because a shrinkage penalty is used, this will also test the non-linear assumption and allows the variable for forest cover to be continuous. The disadvantage of this approach is that the relationship between forest cover and change in flow is less easy to interpret, as the non-linear fit in the GAM has no direct parametric form. All three approaches are tested in this study.

The overarching test focuses on identifying the change stream flow as a result of a change in forest cover and how this is potentially affected by different other factors (as indicated by the previous research: Zhang et al. [44]; Filoso et al. [16]; Zhou et al. [47]): climate, size of catchment and length of study. In addition to these earlier identified factors, this study also tested for the factors listed in Table 1

As an initial approach we tested whether the additional catchments added to the original data from Zhang et al. [44] did not majorly influenced the results (This analysis is in supplementary material part 2). This analysis highlights that the newly added catchment and the changes to the data set create minor differences when repeating the analysis from the original paper. However, this means that the results of the studies are still comparable.

To make all the data and code used for the analysis publicly available, all the final data and analysis for this paper are located on github: https://github.com/WillemVervoort/Forest_and_water on the "publish" branch.

3. Results

3.1. Description of the data

The overall data set contains 334 observations of changes in flow, which includes the newly identified data sets and after removing identified duplicate data and lines with missing data. In contrast, the original data set from Zhang et al. [44] contained 312 catchments and the Filoso et al. [16] study used 37 catchments (Table S2 in Filoso et al. [16]). The overall distribution of changes in flow is highly skewed as is the distribution of changes in forest cover and *Area* km^2. The values of changes in flow greater than 100% and smaller than

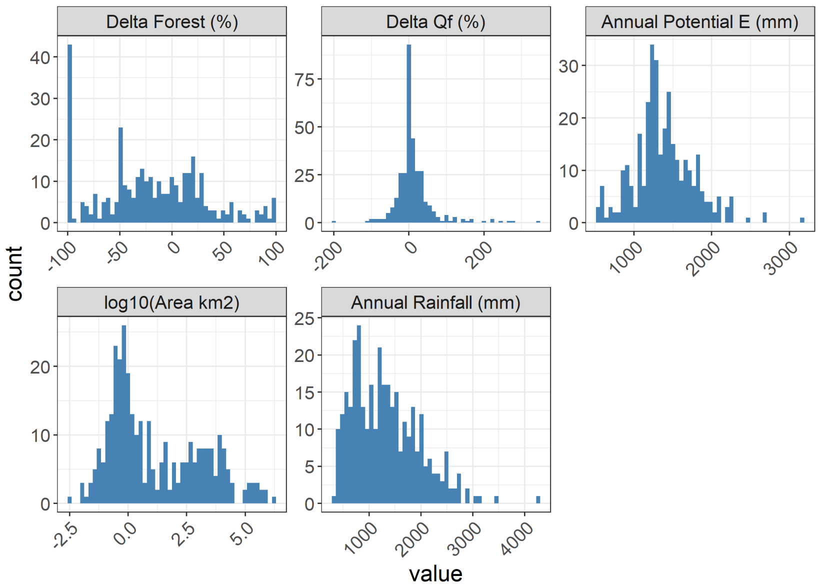

Figure 1: Overview of the distribution of the data set for five of the included variables. Note that the first panel (showing the distribution of the catchment areas) indicates the distribution of the *log__10__* transformed Area (in km^2).

-100% clearly create long tails in the change in flow distribution. Note also the large number of studies with 100% forest cover reduction. Clearly visible is also that smaller catchments dominate the database with 42% of the data from catchments < 1 km^2 and 65% of the data for catchments < 10 km^2 (Figure 1). This high skew in some of the data can create difficulties in the statistical modelling and this will be discussed later.

3.1.1. Geospatial location of the catchments

Apart from looking at the distribution of the values, the spatial locations of the data can also be important, in particular when analyzing the effect of climate. The catchments are spread across the world, and relative to Zhang et al. [44], this data set has a very similar geospatial distribution. The major climate gradients are represented in the data, but there appears to be some bias in the spatial locations of the data. As the global map (Figure 2) shows, the

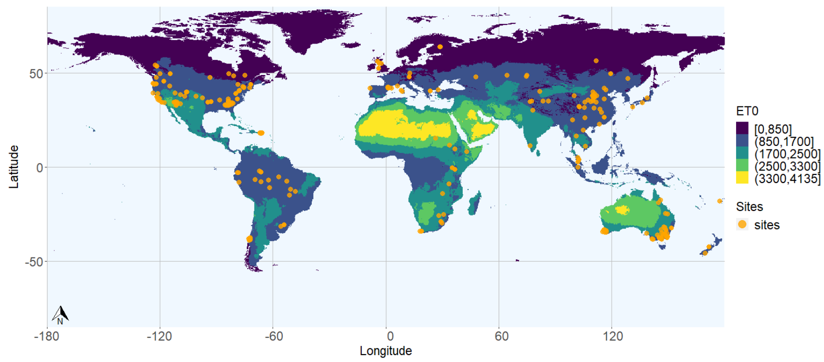

Figure 2: Distribution of included catchments across the globe based on reported or estimated latitude and longitude

distribution of case study catchments covers multiple continents. There is some spatial clustering in the studies in North America, Australia and East Asia.

3.1.2. Cross correlation between the different variables

A final data exploration is to identify potential cross correlations in the data, which can point to possible interactions or potential biases. This analysis can also provide further insight for the statistical modelling, highlighting potential latent variables in the data set.

The correlation plot (Figure 3) highlights several correlations that are worth investigating, even though in general cross correlations between variables are quite low. Some interesting relationships that appear in this graph are:

- the negative relationship between log10(Area) and change in forest area (DeltaF_perc), indicating that in the data set larger catchments tended to have (obviously) smaller areas of forest change.
- the weak positive relationship between log10(Area) and the assessment method using hydrological models. This highlights that paired catchment

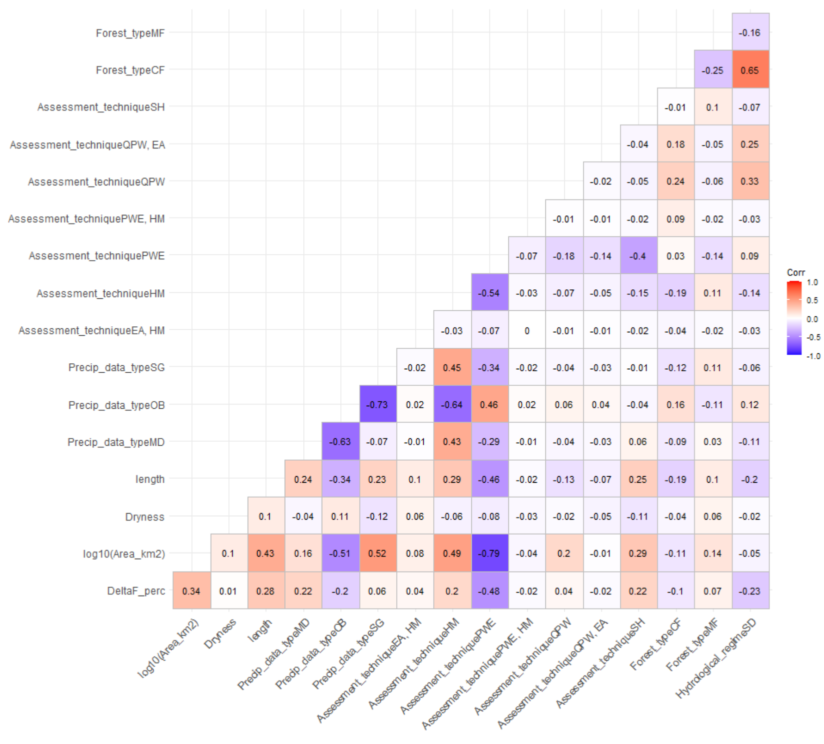

Figure 3: Correlation matrix for all variables

studies mostly concentrate on smaller scales.

- A strong inverse relationship between log10(Area) and the paired watershed assessment method (simply the inverse from the last point), which is also indicated by the negative relationship between the two assessment methods. This is further visible in the relationship between the change in forest cover and the paired watershed assessment method, showing the impact of the latent variable (log10(Area)). Smaller catchments used in paired watershed assessments are easier to fully clear or fully replant.

3.2. Statistical analysis

The results of the overall statistical model that includes all the variables (but no interactions) reinforces some of the results from the correlation analysis.

This includes introducing non-linearity (Equation (2)) for the numerical variables in the model. While increasing non-linearity in the model can increase the flexibility if the model, the shrinkage splines assist with limiting over fitting. Following Wood [41], the number of degrees of freedom k in the non-linear variables was based on assessment of the effective degrees of freedom in the model output. If the effective degrees of freedom were close to k - *1* then k was increased and the model rerun. By using shrinkage splines, this also results in the whole term being shrunk to zero if needed [41].

Table 2: Statistical summary for the linear terms the full model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.37	16.19	-0.33	0.74
DeltaF__perc	-0.61	0.06	-11.03	0
Precip__data__typeOB	-21.34	13.16	-1.62	0.11
Precip__data__typeSG	9.57	15.16	0.63	0.53

	Estimate	Std. Error	t value	Pr(>|t|)
Assessment__techniqueEA, HM	20.32	42.72	0.48	0.63
Assessment__techniqueHM	23.51	11.69	2.01	0.05
Assessment__techniquePWE	30.71	11.92	2.58	0.01
Assessment__techniquePWE, HM	15.79	43.24	0.37	0.72
Assessment__techniqueQPW	41.29	20.14	2.05	0.04
Assessment__techniqueQPW, EA	25.16	24.41	1.03	0.3
Assessment__techniqueSH	46.03	11.65	3.95	0
Forest__typeCF	-7.76	7.52	-1.03	0.3
Forest__typeMF	-7.8	7.35	-1.06	0.29
Hydrological__regimeSD	1.5	9.1	0.17	0.87

Table 3: Statistical summary for the smooth terms for the full model

	edf	Ref.df	F	p-value
s(log10(Area__km2))	0.81	4	1.09	0.02
s(Dryness)	4.59	9	2.25	0
s(Length)	4.39	34	0.21	0.13

The overall explaining power of the model can be interpreted from the adjusted r^2 (which is penalized for the number of parameters). This indicates an adjusted r^2 of 0.45 and deviance explained is 0.49, suggesting the model only explains about 50% of the variance in the data.

Inspecting the significance of the variables (Table 2 and Table 3) indicates some interesting features. The overall partial slope of the change in forest cover is -0.61, if all other variables are kept constant. This suggest quite strong change in stream flow, moving from fully forested to fully cleared. Over the whole forest cover range, this is a change of -122 mm, with other variables held constant. This change is highly significant, as indicated by the low p-value.

In addition, all the smoothed variables *log10(Area (km^2))* (p = 0.02)), *Dryness* (p = 0)) and *Length* (p = 0.13)) explain variation in the data. For *Length*, the p-value is not strictly smaller than 0.05, but still indicates some reasonable evidence that the variable explains some of the variation in the change in stream flow.

Furthermore Table 2 indicates that several of the assessment methods explain variation in the change in stream flow, which was also indicated in the correlation analysis. In particular, the assessment methods Paired Watersheds Experiments (PWE), Hydrological Modelling (HM) and Statistical modelling and hydrographs (SH) are important explaining variables ($p < 0.05$).

The remaining variables related to rainfall observation technique, forest type, or hydrological regime don't appear to have an influence on the change in flow.

Table 4: Comparison of alternative models for the relationship between the change in forest cover and the change in streamflow. (See Supplementary Material part 3)

Model for change in forest cover	Deviation explained	AIC
linear across range	0.49	3233
different for forestation and deforestation	0.45	3281

Model for change in forest cover	Deviation explained	AIC
non-linear across the range	0.51	3233

As discussed in the methods, the overall linear response to the change in forest cover was compared to a transformation of the negative forest cover to positives and a check whether the relationship might be non-linear. This approach tests whether the impact on stream flow from removing forest cover is different from reforestation, as outlined in the methods. The detail of the comparison is highlighted in Supplementary material part 3. However, generally the results of the analysis showed two main points (Table 4):

1. The model assuming a simple linear relationship between change in forest cover (both positive and negative) and the change in flow explained the most variation in the data and indicated the best performance in terms of the Aikaike Information Criterium (AIC); and
2. There is no need to assume a non-linear relationship, as a linear relationship provides a similar performance for the fit to the data.

The smoothed variables in the model can be inspected visually to identify if there are any issues with the fit. This is in addition to the earlier mentioned checks using `gam.check()` in the R package `mgcv` to test whether the number of degrees of freedom k is adequate.

Figure 4 highlights that the relationship between *log10(Area km²)* and the change in flow is essentially linear. It indicates the negative slope that was also clear from Zhang et al. [44], indicating that in larger catchments changes in forest cover have less impact on stream flow than for smaller catchments.

Both the *Length* and *Dryness* variables show strong non-linearity, but the relationships do not show a clear trend due to the scatter and the distribution

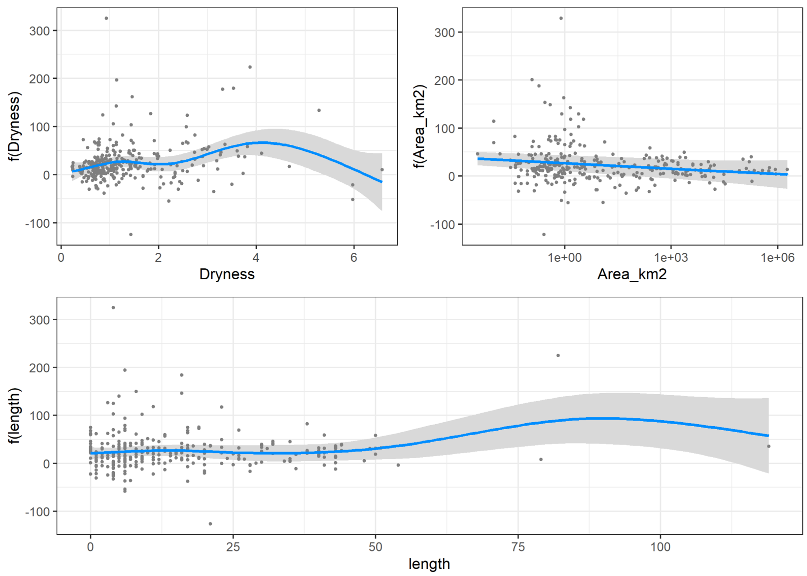

Figure 4: Visualisation of the smooth variables in the model, the shaded areas are the 95% confidence intervals associated with the fit of the smooth, the blue line is the mean smoothed relationship, with data plotted as individual points

of the data. A further problem appears to be that *Length* and *Dryness* have several points with very high leverage that determine much of the non-linearity in the relationship.

As this is not always shown in papers discussing regression relationship, the residual distribution is provided in more detail (Figure 5). Visually, the residuals appear approximately normal, although there is a noticable skew in a limited number of the data in the upper part of the distribution (Figure 5, top left). This is related to a limited number catchments that have very high changes in stream flow in the data set. In other words, the distribution of the residuals is somewhat fat-tailed.

One solution could be to transform the data, however this is not that simple. As the data for the change in flow cover the domain $\mathbb{R}$, a simple log or Gamma transformation is not a solution. More complex transformations make the re-

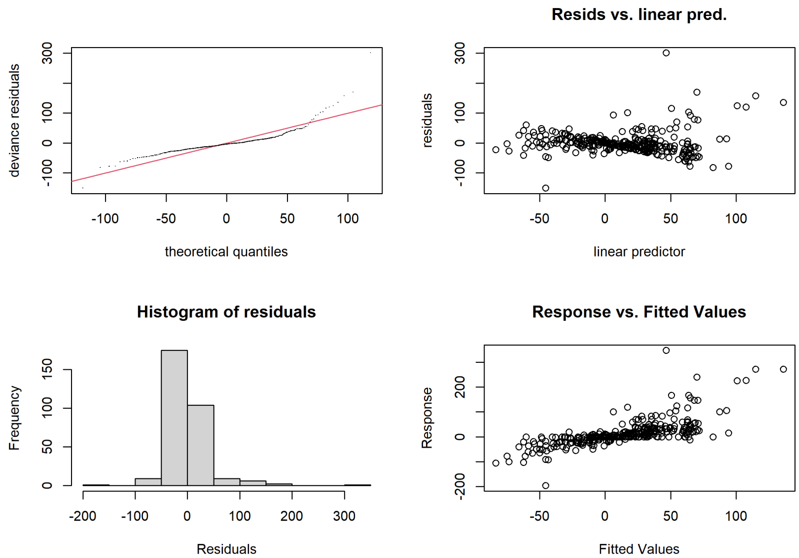

Figure 5: Residual plots for the regression model indicating a slightly fat-tailed residual distribution

sults of the regression difficult to interpret, and at some point can be slightly contrived.

Given the majority of the residuals indicate a relatively well behaved distribution, we simply note the behaviour at the extremes and will discuss this later in the paper, and explain how this relates to the characteristics of the data set.

3.2.1. Removal of studies of great length and for very dry catchments

Table 5: Catchments for which the dryness index > 5

Number	Latitude	Longitude	Catchment name
76	34.67	-111.7	Beaver Creek, AZ #3-2
225	32.74	-111.5	Natural Drainages, Ariz., U.S.A, A

Number	Latitude	Longitude	Catchment name
226	32.74	-111.5	Natural Drainages, Ariz., U.S.A, C
356	-25.75	28.23	Queens river

The flexible nature of the splines means that the Length variable highlights substantial non-linearity in the data, but it is unclear what exactly is captured. The shape of the conditional response (Figure 4) does not reflect a similar response as indicated by Filoso et al. [16] and Jackson et al. [21]. One reason could be that the relationship is dominated by the few data points with very long data series, which show highly variable responses (Figure 4).

The points related to catchments with very long studies (> 60 years) might be questionable, as changes other than forest cover change could affect stream flow. In addition, a few of the catchments have Dryness values that are very large (> 5) and these values have high leverage in the data, affecting the residual distribution. These catchments are listed in Table 5, and are three catchments in Arizona and 1 catchment in South Africa. It is possible that catchments in these climate zones behave different from the rest of the catchments.

Table 6: Statistical summary for the linear terms the restricted model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-10.69	17.85	-0.6	0.55
DeltaF__perc	-0.61	0.08	-7.69	0
Forest__SignIncrease	1.12	9.72	0.12	0.91
Precip__data__typeOB	-16.23	12.46	-1.3	0.19

	Estimate	Std. Error	t value	Pr(>|t|)
Precip__data__typeSG	15.96	14.86	1.07	0.28
Assessment__techniqueEA, HM	19.97	41.02	0.49	0.63
Assessment__techniqueHM	27.07	11.39	2.38	0.02
Assessment__techniquePWE	28.54	12.11	2.36	0.02
Assessment__techniquePWE, HM	15.91	42.05	0.38	0.71
Assessment__techniqueQPW	43.3	19.52	2.22	0.03
Assessment__techniqueQPW, EA	25.33	23.33	1.09	0.28
Assessment__techniqueSH	47.84	11.3	4.23	0
Forest__typeCF	-7.89	7.22	-1.09	0.27
Forest__typeMF	-6.26	7.17	-0.87	0.38
Hydrological__regimeSD	0.54	8.83	0.06	0.95

Table 7: Statistical summary of the smooth terms reducing dataset to studies with the study length shorter than 60 years and Dryness <= 5.

	edf	Ref.df	F	p-value
s(Dryness)	3.82	9	2.15	0
s(log10(Area__km2))	0.89	4	1.66	0.01
s(Length)	0	9	0	0.97

Therefore it is worth investigating what effect removing these few data points has on the overall model and the significance of the variables. Data that have

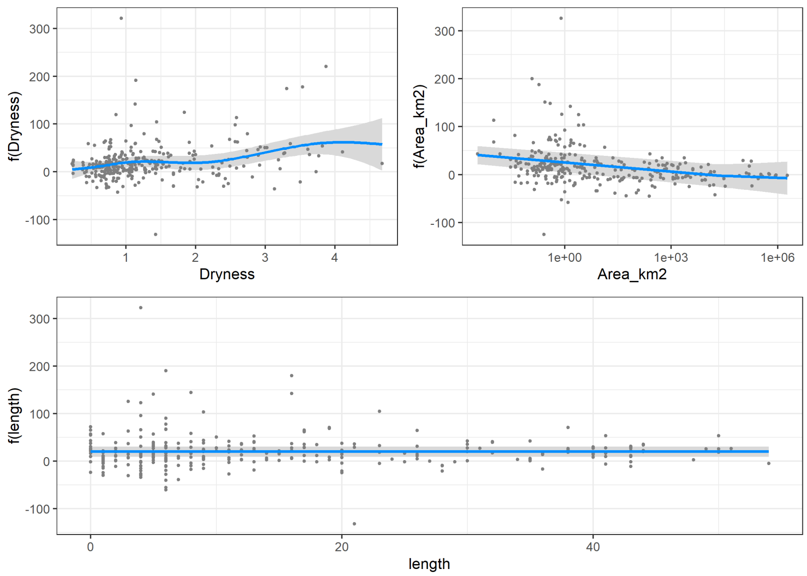

Figure 6: Visualisation of the smooth variables in the model with reduced data for dryness and length

Dryness <= 5 and *Length* <= 60 years were removed from the data set and the model based on a reduction of the data set from 334 to 315 catchments is run again.

This model, which excludes data with long studies and very dry catchments explains only slightly less of the variation with an adjusted r^2 of 0.45 and a deviance explained of 0.48.

Investigating the non-linear responses suggest that *Dryness* has a clear non-linear response, which is significant, where changes in forest cover in drier catchments having a greater impact on stream flow (Figure 6 and Table 7). Catchment area (*log10(Area (km^2))*) still has an impact on flow with p = 0.01, and the relationship looks almost linear. More importantly, the variable *Length* is no longer significant, after removal of the two studies with very long lengths.

Table 8: Distribution of assessment techniques in the data set

Assessment_technique	n
PWE	187
HM	57
SH	45
EA	32
QPW	7
QPW, EA	4
EA, HM	1
PWE, HM	1

One concern with the results presented so far is that there are a few assessment techniques in the data set with a very low number of observations and could influence the results of the analysis. This includes the category of Quasi paired watersheds and combinations of elasticity analysis and hydrological modelling (EA,HM) and paired watersheds and hydrological modelling (PWE,HM) (Table 8).

Therefore, the model was rerun excluding the combined assessment techniques (EA, HM), (PWE, HM) and (QPW, EA) and the assessment technique QPW, which were all non-significant (Table 8). This resulted in a data set of 328 catchment studies.

The model based on assessment techniques that have more than 10 observations in the data set does not change much in the results (results not shown). It strengthens the significance of the different assessment techniques, but generally results in the same interpretation. Overall this suggests that although those observations have some impact on the overall relationships, they do not strongly bias the outcomes.

The overall model results clearly highlight that some of the assessment techniques (in particular paired watershed studies (PWE) and combined use of statistical methods and hydrographs (SH)), have a strong impact on the predicted change in flow. Particularly, relative to EA (elasticity approaches) all other assessment techniques have higher predicted changes in flow. In other words, there is a distinct difference in the way the change in flow is assessed, and the EA method (for example in Zhou et al. [47]) appears to suggest a much smaller effect on the change in flow.

4. Discussion

The generalized additive models appear to reach the same conclusions as the single variable regression in earlier papers [44, 16]. It appears that:

1. Larger catchments show lower impact of forest cover change on stream flow;
2. Drier catchments show a greater impact of forest cover change on stream flow; and
3. There is a general linear relationship between the change in forest cover and the change in stream flow.

This might suggest that the simpler models have reached the correct conclusion. However, this is somewhat premature. given that the other major point coming out of the results is:

4. There is a clear relationship between size of catchments, area cleared and type of experiments, with particular Paired Watershed Experiments containing the smallest catchments, the largest percent forest cover change and the largest variability in the flow response.

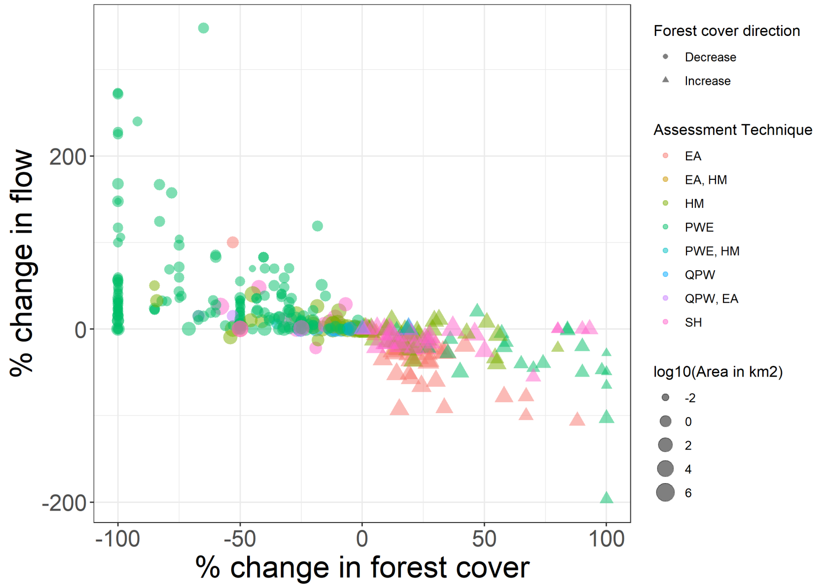

Figure 7: Overview of the data highlighting the dominance of small catchment studies which are fully forested or cleared and the scatter in the data

Figure 7 provides a clear overview of the whole data set, and in this figure the size of the catchments and the different assessment methods are highlighted. This figure clearly indicates that the data relating to high changes in forest cover are all small catchments and relate mostly to paired watershed experiments. In contrast, data related to large catchments are related to smaller changes in forest cover and different methods, such as hydrological modelling and elasticity analysis. This confirms the model results (Table 6) and the earlier correlation analysis (Figure 3).

It is possible that one of the reasons why Zhang et al. [44] separated their analysis in large (> 1000 km^2) and small (< 1000 km^2) catchments, is that they realized this difference in assessment methods and wanted to account for this. However, this is not explicitly identified, and there is no real physical explanation of the 1000 km^2 threshold.

The other interesting point in Figure 7 is that the variation in the data increases as the catchment size decreases and the change in forest cover increases. This also means that the overall variation in the data for paired watershed experiments (PWE) is much greater than for any of the other methods.

4.1. Is there a problem with extending local experimental data to larger scales?

The overarching reason for combining past studies at a global scale is to infer relationships that can be used to make more general statements or develop more global scale modelling of impacts [i.e. 47, 21, 18]. Therefore, the results from the analysis could be seen as a confirmation of the earlier research [44, 16, 47, 21]. However, the explaining power of the developed model is quite low and a lot of variation in the data is unexplained. As is highlighted in the introduction there are four major issues with this type of analysis, and the results from this paper also highlight these issues. Here, these issues are further explained.

4.1.1. Issue 1: Latent variables are not included in the typical single covariate analysis

The results show that it is simply impossible to analyze a single covariate relationship, as there are several latent variables in the data. An example of this is the general relationship of the change in flow as a function of the change in forest cover. Clearly the relationship is highly impacted by the fact that all the small catchments have large changes in forest cover and are all associated with paired watershed experiments. Without taking these factors into account, a definite answer about the impact of forest cover on the change in flow cannot be given. Furthermore, the large variability in the change in flow data for these small catchments (Figure 7) indicates that there is a further (unknown) variable that explains the variation in the data.

If the remaining variation in the residuals is small relative to the trend, then there is little need to identify further latent variables, but if the variation is large, then it is unclear if it is the latent variable that determines the trend, or the actual relationship in the data.

Similarly, the data for the larger catchments containing smaller changes in forest cover are dominated by hydrological modelling studies, resulting in a further complication. If the reponse of the stream flow in the modelling studies is the result of the conceptualized relationship between stream flow and forest cover (possibly from a subset of the paired catchment studies), then it is impossible to say if the change in stream flow is real, or simply a result of a pre-conceived model relationship. Is the smaller variation in the data for smaller changed in forest cover (Figure 7) a result of similar conceptualized model relationships, or actual variation between catchments and climate types? Currently this question cannot be answered.

This becomes problematic when extrapolated to larger scales. A clear example of this is the paper by Hoek van Dijke et al. [18] where the conceptualized relationship between forest cover and stream flow pre-determines the outcomes of the global modelling.

The only way to analyze changes in stream flow as a function of forest cover in larger catchments is to actually derive this from observed data of long term stream flow and forest cover (as was done in Levy et al. [24]).

One of these latent variables could be the total area of forest in a catchment, as was analysed in Levy et al. [24]. In this case, the total % area of forest was not included in the data. As a test, the total % area of forest for the larger catchments (> 1000 km^2 in Zhang et al. [44]) were added to the data set and the model for just the large catchments was tested. This showed that the % area of forest was not significant to explain the change in flow for the larger catchments (retaining all other variables in the model, results not shown). While this might be an area of further research on the full data set, it is complicated for two reasons:

1. The area of forest is not always indicated in the original papers, or a range of values is given, complicating the data collection.
2. Many of the small catchments have 100% area covered in forest, introducing a strong skew in the data and complicating if total area of forest has an impact on the change in flow.

We are not arguing that there is no relationship between stream flow and forest cover, and there might indeed be a global relationship that can be discovered. But, this relationship can only be discovered if we are able to address some of the major other factors that explain the variability, and work with actual data and not model outputs.

4.1.2. Issue 2: Interpretation errors due to complex descriptions of the experiments in the original papers

The second major issue that became clear from reviewing many of the original papers is that some of the variability might be an interpretation problem. In many cases the original description in the paper is interpreted to extract the % change in stream flow from the % change in forest cover. This seems like a simple activity, but this is not always the case.

Two examples can be highlighted:

- The papers from Almeida et al. [1] and Ferreto et al. [15] partly discuss the same experiment and the same catchment. In Almeida et al. [1], the methods discuss how two experimental catchments of approximately 80ha in size which were harvested. One catchment was 100% harvested and the other 30% harvested. Throughout the paper the catchments are indicated as 100% harvested and 30% harvested. However, only after reading Ferreto et al. [15], did we discover that in fact the 100% and

30% refer to the "eucalyptus plantation area", which was about 60% of the total area. This is in fact mentioned in Table 1 in Almeida et al. [1], but does not appear in the text. The question then becomes how to interpret this in the data base for this paper. Clearly it was a 100% and 30% change in forest cover, but only for the 60% plantation cover, not for any of the other areas in the catchment, which included native vegetation and riparian vegetation. There are several other examples like this in the different papers [for example 6, 5].

- Another example is the paper by Waterloo et al. [39]. This modelling study in Fiji of the clearing of a catchment reports the changes in stream flow over parts of the year. For a period of 324 days the stream flow increased from 252 mm to 580 mm (a 230% increase if calculated as $580/252 * 100$) and for a second period of 309 days the stream flow increased from 90 mm to 194 mm (a 215 % increase). However, how we convert this to a change in annual flow (which most of the other data relate to) is difficult. The original data base listed a 50 % change in flow, but it is difficult to identify how this is calculated. We suspect that results from $252/580 * 100 \approx 50$ and $90/194 \approx 50$.
- A final example is around the choice of control or treatment. In the data base the assumption is that the change in flow is relative to the original situation. This can be either a "before and after" analysis, which can be problematic by itself due to climate variation, or a comparison of a treatment with a control. But even in this case comparing across catchments can be tricky. For example, at one extreme, some controls in the database are a bamboo catchment and a tea plantation [6, 5]. A clearer example is the Brigalow catchment study in Queensland, Australia [33], catchment #336 in the data set. This is a paired watershed experiment

of conversion of native Brigalow (an Acacia species) into cropland and pasture. We chose to use the cropped catchment (C2 Thornton et al. [33]) as the deforestation treatment resulting in a change of flow of 140%, based on Table 4 in Thornton et al. [33]. However, had we used the pasture catchment as the treatment then the change in flow would have been up to 165%.

Clearly, interpreting older papers can be difficult and this can result in variation in the data that is being analyzed. Similar to the last issue, if these errors only introduce small variation in the data, then it will not limit the interpolation to larger scales. At this point, it is not clear if this is indeed the case. The large variation in the experimental watershed data suggests that this might be a more serious problem.

4.1.3. Issue 3: Aggregation of data that originates from different experiments with different objectives across a wide time period

For many of the small catchment studies listed in the database, the assumption is that the original experimental design can be interpreted in terms of a binary "forestation" or "deforestation". However, the real situation is often much more complex and fuzzy.

Many of the paired watershed experiments included a harvesting and replanting or regrowth after harvesting or fire experiment [e.g. 11, 12, 40]. As a result, it becomes difficult to assess how we interpret the change in flow as a result of a change in cover. In many cases we would expect the flow to change over time as a function of the recovery [22] and therefore the time series of the flow needs to be assessed over a longer time.

Many of the papers in the database report early results (for example 1 or 3 years after harvesting), but some also report longer time periods. As earlier work [12, 22] has highlighted, we can always expect a larger effect directly after harvesting, but this effect diminishes over time (even if it does not always return to the original state). Comparing studies reporting results directly after treatment to longer term studies therefore becomes problematic.

In our work, the variable *Length* was used in the model to test for some of these effects, but this was insignificant in the model (Table 7). Given the other variation in the data, this does not necessarily mean that there is no effect.

This is further complicated by the variation in different types of clearing and the different types of vegetation. In the original Zhang et al. [44] a variable to describe the *forest type* was included (Table 1), but in the model this is not significant (Table 2). This is probably because the broad classification used does not capture the actual variation in runoff response. In addition, as Figure 3 shows, there is a correlation between coniferous forests and snow dominated hydrological regimes, further complicating the analysis.

An additional complication related to combining studies related to wild fires or bush fires and logging studies is the differences in vegetation recovery. For example, Heath et al. [17] found that catchments with resprouting species around Sydney, Australia, indicated little change in the stream flow in comparison to species regrowing from seed further south on the continent [48].

As a result, it can be difficult to exactly pinpoint the change in flow as a result of the change in cover, as well as being difficult to assess what the exact change in cover actually was. In addition, using only the overall change in stream flow can discard a lot of information from observations in individual years. Many papers give stream flow values for multiple years, often showing significant variation. Summarizing this into one average value discards all the information on the variance.

As indicated before, if the overall variation in the data due to this issues is small, then this would not be an issue for upscaling the results, but the large variation for the smaller catchments suggest that effects could be considerable. As Jones et al. [22] indicate, this really needs time series analysis of the different experiments. However, some of the time series data might not be recoverable from the older experiments, which will limit the opportunities for analysis. We will discuss this further below.

4.1.4. Issue 4: Transcription errors in the data

This issue seems to mainly occur if data is collected from other review papers. This might be because some of the original papers are difficult to locate and therefore values from reporting papers are used. In supplementary data part 1, several changes to the original data sets have been documented, and as can be seen several of these are transcription errors.

This does influence the results in Zhang et al. [44], comparing the results in Supplementary material 2 with the original paper. The main example is that in this study the largest catchment (watershed #1 in Zhang et al. [44]) had to be removed, as this study actually involved paired watershed experiments on very small plots, for which the characteristics were not recoverable.

Clearly, this is a problem for all reviews that attempt to bring together large numbers of results from published papers, and where actual results are copied rather than using some sort of automated text analysis.

In the end, careful review of the data and the original papers can circumvent most of this issue. And, making the data available (as Zhang et al. [44], Zhou et al. [47], Filoso et al. [16] and this paper have done) provides an opportunity for review by other researchers, and over time most of the transcription errors can be resolved.

4.2. General discussion

In this paper, a few studies have been singled out to demonstrate the main point that extrapolating individual studies to global scales is not that simple

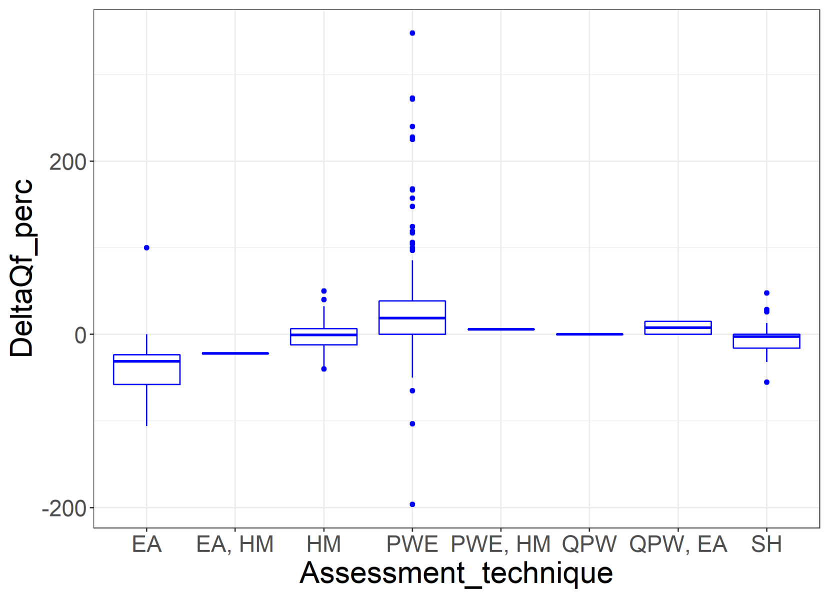

Figure 8: Boxplot of the variation in the change in flow for the different assessment techniques, showing the dominance of the variation and the outliers in the dataset in the paired watershed experiments

in natural systems. Not only is there significant natural variation and latent variables, interpretation and aggregation can cause further unforeseen problems. The choice of papers to focus on was mainly driven by the data that was made available by the authors of these papers [44, 16], which provide a rich case study for the current paper.

Field research is by nature limited in space and time, due to the high costs involved of setting up experiments. This is particularly true for experiments in hydrology and forest hydrology, where field sites need to cover sufficient spatial and temporal variability. This means there is a general need to extrapolate the local results to larger scales to inform decision making and policy.

However, as is demonstrated in this paper, there are multiple issues when this local scale data is extrapolated to larger scales. It clearly demonstrates that the results of any model (in this case a regression model) is highly dependent on the data, but also on the assumptions in the model. From the perspective of extrapolating local data to global scales for policy advice and decision making [i.e. 18, 21], this is an important point.

4.2.1. *Residuals of the model*

The residuals of the final model presented in this paper (Figure 5) indicate that the residual distribution remains fat-tailed, causing deviations from an assumed $\epsilon \sim N(0, \sigma^2)$. This once again highlights that there is unexplained variation at the extremes of the distribution, once again related to the paired watershed experiments (Figure 8). Generally,in statistical models, the approach would be to further normalize the residuals through transformations. However, in this case this might be difficult and might not resolve all the issues due to the large variation in the data.

4.2.2. *Interactions*

The current modelling approach does not consider any interactions between the variables, and this would offer another approach to understand the variation in the data. As already indicated in Figure 3, there are interactions between different variables. This further complicates the extrapolation of the local scale experiment data to global scales and to extend historical data to current management and decisions.

In this case, interactions were not included because, as was shown, there are bigger problems with trying to extrapolate the existing data, and the data itself can be problematic. To be able to model the interactions well, the nature of the variables and interactions need to be understood and or clearly hypothesized. Otherwise it becomes another case of correlation without causation.

4.2.3. Implications for other "meta-analysis" studies

There has been a recent push to develop more meta-analysis studies in hydrology [38, 14], and we strongly believe that developing new insights by combining historical data sets from reviewed papers is highly valuable. However, this paper highlights that there is considerable chance that large historical data sets include latent variables and are more complex than envisioned. This is particularly true for work in natural systems and more historical work, as methods of observation and even approaches to management have changed considerably. The same management description is not necessarily the same action on the ground. A carefully designed and systematic approach can prevent some of bigger problems as is demonstrated in Wang et al. [38], where both the approach and the catchment area are investigated as latent variables. This is particularly relevant, where the results of meta-analyses are extrapolated to make global predictions without clearly quantified uncertainties (such as in Hoek van Dijke et al. [18] and Wang et al. [38]).

A second potential danger is the extrapolation of the local small catchment results and conclusions to larger scales, but beyond the original scope of the studies. For example, the current database is mainly related to forest harvest, bush fire and reforestation/plantation management. It is tempting to use the result of a large scale analysis of this data to make inferences about overall land use change [25, 38], but this would not be valid, as not all deforestation studies are a transition to an agricultural land use or pasture, as in Levy et al. [24]. Many are logging or bushfire studies regrowing into forest after the initial treatment. Similarly, using the plantation studies to extrapolate to "reforestation" (as in Filoso et al. [16] and Hoek van Dijke et al. [18]) is also tenuous. Plantation forests are generally fast growing hybrids that will have quite different ecophysiology, particularly in South America [22, 4], while other reforestation, for example for salinity control in Australia, might focus on a mix of native species. Given the link between ecophysiology and water and carbon budgets [21], care should be taken in extrapolation, introducing a further error.

As highlighted summarizing highly variable observed time series into single mean values introduces further potential problems, as well as discarding information about the variance.

A final factor is ignoring the effect of climate change [37] on runoff, even if the effects are still minor. Earlier papers [25, 38] have analyzed climate effects relative to management effects in the data, but these studies did not explicitly test for climate change. Given that the database of studies now captures almost 100 years of work, we cannot ignore a climate change trend that is potentially hidden in the data. A simple inclusion of the start date of the experiment (*From*) in the GAM model does suggests an increase in change in the percentage of flow over time. However, as the data distribution is uneven in time, and consists of multiple assessment techniques there could be multiple complicating factors, and drawing a firm conclusion would be premature.

4.2.4. Future research needs (implications for forest hydrology)

Beyond a more formal approach to investigating climate change effects in the data, this study also points to several further opportunities and future research needs.

A major focus of many of the papers related to forest hydrology has been on the impact of plantation forest operations on the catchment, rather than the transition of forestry to agriculture. As the paper by Jones et al. [22] highlights this means there are opportunities to analyze the time evolution of the catchment response to forestation. Given the large number of studies that look at a time evolution of forest cover (i.e. either clearing and regrowth, or burning and regrowth), this data can offer further insights into the dynamic response of catchments to changes in land cover. In addition, this might allow analysis of the variance of the response in addition to the mean response. While some of the older data is not fully recoverable, but there is often a series of papers related to one experiment, which at least would provide individual time points.

More generally there is a clear need for a more in depth analysis of the data base of studies used here. In particular, more detailed data can potentially be extracted from many of the studies in terms of vegetation species, stream flow responses and responses of components of stream flow (slow flow, quick flow etc.), as well as a more in depth description of the management and actual experimental design.

There is also a clear need to understand the impact of the assessment methods with respect to scale. Extrapolating paired watershed experiment results into models can possibly overlook landscape interactions that are visible at larger scales, but do not occur on smaller scales. For example, this could be the effects of lateral flow and groundwater connectivity and impacts of elevation on land use. A carefully designed simulation study that specifically investigates the change in stream flow response with scale using local field data for verification can help solve this problem.

At the moment, providing answers to the impact of stream flow at larger scales should generally not be approached by simulation modelling. A better approach is analyzing stream flow data at multiple spatial and temporal scales for responses (rather than running simulations) and using satellite data to dynamically include land use changes. The highlighted paper by Levy et al. [24] is currently the best example of a solid statistical approach to analyzing stream flow responses. Simulation modelling can be an approach to analyze different scenarios, if there is clear recognition of the potential impact of the model structure (the algorithms and parameters that describe for example plantation tree growth) on the simulation outcomes.

We envision that in the future more innovative approaches to analyzing data at different scales will be developed.

5. Conclusions

This study demonstrates that the analysis of large databases of essentially "aggregated data" should be considered carefully and simple single variable regressions often present simplistic relationships that can be misleading.

On first glance, stream flow will decrease with increases in forest cover, and increase with removal of forest cover. However, the conclusions on the exact relationship between stream flow change and forest cover change are highly uncertain. There are four major interlinked reasons why earlier conclusions should be considered carefully:

- The existence of latent variables in the data that create the appearance of a relationship that really does not exist. In this case the assessment technique used to assess the change in stream flow is highly significant;

- The difficulty in fully interpreting the specifics of different studies. In this case the definition of vegetation, stream flow volume, control or time period was difficult to assess;

- The difficulty of integrating data from seemingly similar studies, but with quite different objectives. This study highlighted the many logging studies followed by regrowth, or bushfire effects; and

- The chance of transcription errors influencing the data.

While some of these issues might be overcome with careful analysis and transcription, not all can be directly or easily resolved. Any statistical analysis, including the one in this paper, needs to be considered “conditional on the data”, and given the issues indicated, extrapolation of the results of summary studies to larger scales and into global hydrological models has to be done with great care. Better would be to analyze observed data and explicitly include uncertainty in the extrapolation of the results. This therefore has implications for the recent growth in meta-analysis review papers, which has been boosted by increased computational capacity and much better on-line accessible data bases with research data. This is particularly true for natural systems involving climate variability and therefore extrapolations of experimental work in Environmental Science and Hydrology to global scales in general.

6. Acknowledgements

This work was funded through project FPTA 358, Instituto Nacional de Investigacion Agropecuaria, INIA-Uruguay.

7. CRediT Statement

R. Willem Vervoort: Conceptualization, Methodology, Code, Writing- Original draft preparation, Writing- Reviewing and Editing. Eliana Nervi: Data curation, Writing- Original draft preparation, Writing- Reviewing and Editing. Jimena Alonso: Conceptualization, Writing- Reviewing and Editing.

References

[1] Auro C. Almeida, Philip J. Smethurst, Anders Siggins, Rosane B. L. Cavalcante, and Norton Borges Jr. Quantifying the effects of eucalyptus plantations and management on water resources at plot and catchment

scales. *Hydrological Processes*, 30(25):4687–4703, 2016. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.10992. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10992.

[2] Vazken Andréassian. Waters and forests: from historical controversy to scientific debate. *Journal of Hydrology*, 291(1):1–27, 2004. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2003.12.015. URL https://www.sciencedirect.com/science/article/pii/S0022169403005171.

[3] H. E. Beck, L. A. Bruijnzeel, A. I. J. M. van Dijk, T. R. McVicar, F. N. Scatena, and J. Schellekens. The impact of forest regeneration on streamflow in 12 mesoscale humid tropical catchments. *Hydrol. Earth Syst. Sci.*, 17(7):2613–2635, 2013. ISSN 1607-7938. doi: 10.5194/hess-17-2613-2013. URL https://hess.copernicus.org/articles/17/2613/2013/. HESS.

[4] Dan Binkley, Otávio C. Campoe, Clayton Alvares, Rafaela L. Carneiro, Ítalo Cegatta, and Jose Luiz Stape. The interactions of climate, spacing and genetics on clonal eucalyptus plantations across brazil and uruguay. *Forest Ecology and Management*, 405:271–283, 2017. ISSN 0378-1127. doi: https://doi.org/10.1016/j.foreco.2017.09.050. URL https://www.sciencedirect.com/science/article/pii/S0378112717311386.

[5] JR Blackie. 2.2. 1 the water balance of the kericho catchments. *East African Agricultural and Forestry Journal*, 43(sup1):55–84, 1979.

[6] JR Blackie. 3.2. 1 the water balance of the kimakia catchments. *East African Agricultural and Forestry Journal*, 43(sup1):155–174, 1979.

[7] H. Borg, R. W. Bell, and I. C. Loh. Streamflow and stream salinity in a small water supply catchment in southwest western australia after reforestation. *Journal of Hydrology*, 103(3):323–333, 1988. ISSN 0022-

1694. doi: https://doi.org/10.1016/0022-1694(88)90141-2. URL https://www.sciencedirect.com/science/article/pii/0022169488901412.

[8] J. M. Bosch and J.D. Hewlett. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, 55:3–23, 1982.

[9] Alice E. Brown, Lu Zhang, Thomas A. McMahon, Andrew W. Western, and Robert A. Vertessy. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, 310(1-4):28–61, 2005. URL http://www.sciencedirect.com/science/article/B6V6C-4G05MM9-1/2/bbc5fc0e958a8f34bcb7c1cc7fa57b48.

[10] Alice E. Brown, Andrew W. Western, Thomas A. McMahon, and Lu Zhang. Impact of forest cover changes on annual streamflow and flow duration curves. *Journal of Hydrology*, 483(0):39–50, 2013. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/j.jhydrol.2012.12.031. URL http://www.sciencedirect.com/science/article/pii/S0022169412011146.

[11] P. M. Cornish. The effects of logging and forest regeneration on water yields in a moist eucalypt forest in new south wales, australia. *Journal of Hydrology*, 150(2-4):301–322, 1993. URL http://www.sciencedirect.com/science/article/B6V6C-487D3Y2-9J/2/73c981ba76284d9d629f6b221d6fd6c6.

[12] P. M. Cornish and R. A. Vertessy. Forest age-induced changes in evapotranspiration and water yield in a eucalypt forest. *Journal of Hydrology*, 242(1-2):43–63, 2001. URL http://www.sciencedirect.com/science/article/B6V6C-429910G-3/2/0158b1f89ff436f338a9e688a47f06c4.

[13] Claude Cosandey, Vazken Andréassian, Claude Martin, J. F. Didon-Lescot, Jacques Lavabre, Nathalie Folton, Nicolle Mathys, and Didier Richard. The hydrological impact of the mediterranean forest: a review of french research. *Journal of Hydrology*, 301(1):235–249, 2005. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2004.06.040. URL `https://www.sciencedirect.com/science/article/pii/S0022169404003257`.

[14] Jaivime Evaristo and Jeffrey J. McDonnell. A role for meta-analysis in hydrology. *Hydrological Processes*, 31(20):3588–3591, 2017. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.11253. URL `https://doi.org/10.1002/hyp.11253`. https://doi.org/10.1002/hyp.11253.

[15] Décio Oscar Cardoso Ferreto, José Miguel Reichert, Rosane Barbosa Lopes Cavalcante, and Raghavan Srinivasan. Water budget fluxes in catchments under grassland and eucalyptus plantations of different ages. *Canadian Journal of Forest Research*, 51(4):513–523, 2020. ISSN 0045-5067. doi: 10.1139/cjfr-2020-0156. URL `https://doi.org/10.1139/cjfr-2020-0156`. doi: 10.1139/cjfr-2020-0156.

[16] Solange Filoso, Maíra Ometto Bezerra, Katherine C. B. Weiss, and Margaret A. Palmer. Impacts of forest restoration on water yield: A systematic review. *PLOS ONE*, 12(8):e0183210, 2017. doi: 10.1371/journal.pone.0183210. URL `https://doi.org/10.1371/journal.pone.0183210`.

[17] J. T. Heath, C. J. Chafer, F. F. van Ogtrop, and T. F. A. Bishop. Post-wildfire recovery of water yield in the sydney basin water supply catchments: An assessment of the 2001/2002 wildfires. *Journal of Hydrology*, 519, Part B(0):1428–1440, 2014. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/j.jhydrol.2014.09.033. URL `http://www.sciencedirect.com/science/article/pii/S002216941400715X`.

[18] Anne J. Hoek van Dijke, Martin Herold, Kaniska Mallick, Imme Benedict, Miriam Machwitz, Martin Schlerf, Agnes Pranindita, Jolanda J. E. Theeuwen, Jean-François Bastin, and Adriaan J. Teuling. Shifts in regional water availability due to global tree restoration. *Nature Geoscience*, 15(5): 363–368, 2022. ISSN 1752-0908. doi: 10.1038/s41561-022-00935-0. URL `https://doi.org/10.1038/s41561-022-00935-0`.

[19] Andrés Iroumé and Hardin Palacios. Afforestation and changes in forest composition affect runoff in large river basins with pluvial regime and mediterranean climate, chile. *Journal of Hydrology*, 505:113–125, 2013. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2013.09.031. URL `https://www.sciencedirect.com/science/article/pii/S0022169413006847`.

[20] Andrés Iroumé, Octavio Mayen, and Anton Huber. Runoff and peak flow responses to timber harvest and forest age in southern chile. *Hydrological Processes*, 20(1):37–50, 2006. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.5897. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.5897`.

[21] Robert B. Jackson, Esteban G. Jobbagy, Roni Avissar, Somnath Baidya Roy, Damian J. Barrett, Charles W. Cook, Kathleen A. Farley, David C. le Maitre, Bruce A. McCarl, and Brian C. Murray. Trading water for carbon with biological carbon sequestration. *Science*, 310(5756):1944–1947, 2005. doi: 10.1126/science.1119282. URL `http://www.sciencemag.org/cgi/content/abstract/310/5756/1944`.

[22] Julia Jones, Auro Almeida, Felipe Cisneros, Andres Iroumé, Esteban Jobbágy, Antonio Lara, Walter de Paula Lima, Christian Little, Carlos Llerena, Luis Silveira, and Juan Camilo Villegas. Forests and water in south

america. *Hydrological Processes*, 31(5):972–980, 2017. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.11035. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.11035.

[23] George Kuczera. Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest. *Journal of Hydrology*, 94(3-4): 215–236, 1987. ISSN 0022-1694. doi: Doi:10.1016/0022-1694(87)90054-0. URL http://www.sciencedirect.com/science/article/B6V6C-487FBY6-12P/2/80e7248c3007e0c82d8b8a52af61894e.

[24] M. C. Levy, A. V. Lopes, A. Cohn, L. G. Larsen, and S. E. Thompson. Land use change increases streamflow across the arc of deforestation in brazil. *Geophysical Research Letters*, 45(8):3520–3530, 2018. ISSN 0094-8276. doi: https://doi.org/10.1002/2017GL076526. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076526.

[25] Qiang Li, Xiaohua Wei, Mingfang Zhang, Wenfei Liu, Houbao Fan, Guoyi Zhou, Krysta Giles-Hansen, Shirong Liu, and Yi Wang. Forest cover change and water yield in large forested watersheds: A global synthetic assessment. *Ecohydrology*, 10(4):e1838, 2017. ISSN 1936-0584. doi: https://doi.org/10.1002/eco.1838. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eco.1838.

[26] Jorge L. Peña-Arancibia, Albert I. J. M. van Dijk, Juan P. Guerschman, Mark Mulligan, L. Adrian Bruijnzeel, and Tim R. McVicar. Detecting changes in streamflow after partial woodland clearing in two large catchments in the seasonal tropics. *Journal of Hydrology*, 416-417:60–71, 2012. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2011.11.036. URL https://www.sciencedirect.com/science/article/pii/S0022169411008213.

[27] MA Roche. Watershed investigations for development of forest resources of the amazon region in french guyana. *Tropical Agricultural Hydrology. J*, pages 75–82, 1981.

[28] Daniel Andres Rodriguez, Javier Tomasella, and Claudia Linhares. Is the forest conversion to pasture affecting the hydrological response of amazonian catchments? signals in the ji-paraná basin. *Hydrological Processes*, 24 (10):1254–1269, 2010. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.7586. URL `https://doi.org/10.1002/hyp.7586`.

[29] J. K. Ruprecht, N. J. Schofield, D. S. Crombie, R. A. Vertessy, and G. L. Stoneman. Early hydrological response to intense forest thinning in southwestern australia. *Journal of Hydrology*, 127(1):261–277, 1991. ISSN 0022-1694. doi: https://doi.org/10.1016/0022-1694(91)90118-2. URL `https://www.sciencedirect.com/science/article/pii/0022169491901182`.

[30] Joep F. Schyns, Arjen Y. Hoekstra, Martijn J. Booij, Rick J. Hogeboom, and Mesfin M. Mekonnen. Limits to the world's green water resources for food, feed, fiber, timber, and bioenergy. *Proceedings of the National Academy of Sciences*, 116(11):4893–4898, 2019. doi: doi:10.1073/pnas.1817380116. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1817380116`.

[31] C. R. Stoof, R. W. Vervoort, J. Iwema, E. van den Elsen, A. J. D. Ferreira, and C. J. Ritsema. Hydrological response of a small catchment burned by experimental fire. *Hydrol. Earth Syst. Sci.*, 16(2): 267–285, 2012. ISSN 1607-7938. doi: 10.5194/hess-16-267-2012. URL `http://www.hydrol-earth-syst-sci.net/16/267/2012/http://www.hydrol-earth-syst-sci.net/16/267/2012/hess-16-267-2012.pdf`. HESS.

[32] C. M. Thornton, B. A. Cowie, D. M. Freebairn, and C. L. Playford. The brigalow catchment study: Ii*. clearing brigalow (acacia harpophylla) for cropping or pasture increases runoff. *Australian Journal of Soil Research*, 45 (7):496–511, 2007. doi: doi:10.1071/SR07064. URL http://www.publish.csiro.au/paper/SR07064.

[33] C. M. Thornton, B. A. Cowie, D. M. Freebairn, and C. L. Playford. The brigalow catchment study: Ii*. clearing brigalow (acacia harpophylla) for cropping or pasture increases runoff. *Australian Journal of Soil Research*, 45 (7):496–511, 2007. doi: doi:10.1071/SR07064. URL http://www.publish.csiro.au/paper/SR07064.

[34] A. Trabucco and R.J. Zomer. Global aridity index and potential evapotranspiration (et0) climate database v2. cgiar consortium for spatial information(cgiar-csi). Published online, available from the CGIAR-CSI GeoPortal at https://cgiarcsi.community, 2018. Accessed: 2021-11-07.

[35] Albert I. J. M. van Dijk, Peter B. Hairsine, Jorge Peña Arancibia, and Trevor I. Dowling. Reforestation, water availability and stream salinity: A multi-scale analysis in the murray-darling basin, australia. *Forest Ecology and Management*, 251(1–2):94–109, 2007. ISSN 0378-1127. doi: http://dx.doi.org/10.1016/j.foreco.2007.06.012. URL http://www.sciencedirect.com/science/article/pii/S0378112707004707.

[36] Robert A. Vertessy, Fred G. R. Watson, and Sharon K. O'Sullivan. Factors determining relations between stand age and catchment water balance in mountain ash forests. *Forest Ecology and Management*, 143(1):13–26, 2001. ISSN 0378-1127. doi: https://doi.org/10.1016/S0378-1127(00)00501-6. URL https://www.sciencedirect.com/science/article/pii/S0378112700005016.

[37] R. Willem Vervoort, Michaela M. Dolk, and Floris F. van Ogtrop. Climate change and other trends in streamflow observations in australian forested catchments since 1970. *Hydrological Processes*, 35(1):e13999, 2021. ISSN 0885-6087. doi: https://doi.org/10.1002/hyp.13999. URL https://doi.org/10.1002/hyp.13999. https://doi.org/10.1002/hyp.13999.

[38] Shengping Wang, Tim R. McVicar, Zhiqiang Zhang, Thomas Brunner, and Peter Strauss. Globally partitioning the simultaneous impacts of climate-induced and human-induced changes on catchment streamflow: A review and meta-analysis. *Journal of Hydrology*, 590:125387, 2020. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2020.125387. URL https://www.sciencedirect.com/science/article/pii/S0022169420308477.

[39] M. J. Waterloo, J. Schellekens, L. A. Bruijnzeel, and T. T. Rawaqa. Changes in catchment runoff after harvesting and burning of a pinus caribaea plantation in viti levu, fiji. *Forest Ecology and Management*, 251(1): 31–44, 2007. ISSN 0378-1127. doi: https://doi.org/10.1016/j.foreco.2007.06.050. URL https://www.sciencedirect.com/science/article/pii/S0378112707004653.

[40] Ashley A. Webb and Brad W. Jarrett. Hydrological response to wildfire, integrated logging and dry mixed species eucalypt forest regeneration: The yambulla experiment. *Forest Ecology and Management*, 306: 107–117, 2013. ISSN 0378-1127. doi: https://doi.org/10.1016/j.foreco.2013.06.020. URL https://www.sciencedirect.com/science/article/pii/S0378112713003885.

[41] S. Wood. *Generalized additive models: an introduction with R.* CRC Press, Boca Raton, FL, 2006. ISBN 978-1584884743.

[42] Lu Zhang, Fangfang Zhao, Yun Chen, and Renee N. M. Dixon. Estimating

effects of plantation expansion and climate variability on streamflow for catchments in australia. *Water Resources Research*, 47(12):W12539, 2011. ISSN 0043-1397. doi: 10.1029/2011wr010711. URL http://dx.doi.org/10.1029/2011WR010711.

[43] Lu Zhang, Lei Cheng, Francis Chiew, and Bojie Fu. Understanding the impacts of climate and landuse change on water yield. *Current Opinion in Environmental Sustainability*, 33:167–174, 2018. ISSN 1877-3435. doi: https://doi.org/10.1016/j.cosust.2018.04.017. URL http://www.sciencedirect.com/science/article/pii/S1877343518300204.

[44] Mingfang Zhang, Ning Liu, Richard Harper, Qiang Li, Kuan Liu, Xiaohua Wei, Dingyuan Ning, Yiping Hou, and Shirong Liu. A global review on hydrological responses to forest change across multiple spatial scales: Importance of scale, climate, forest type and hydrological regime. *Journal of Hydrology*, 546:44–59, 2017. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2016.12.040. URL http://www.sciencedirect.com/science/article/pii/S0022169416308307.

[45] Fangfang Zhao, Lu Zhang, Zongxue Xu, and David F. Scott. Evaluation of methods for estimating the effects of vegetation change and climate variability on streamflow. *Water Resources Research*, 46(3):W03505, 2010. ISSN 0043-1397. doi: 10.1029/2009wr007702. URL http://dx.doi.org/10.1029/2009WR007702.

[46] Guoyi Zhou, Xiaohua Wei, Yan Luo, Mingfang Zhang, Yuelin Li, Yuna Qiao, Haigui Liu, and Chunlin Wang. Forest recovery and river discharge at the regional scale of guangdong province, china. *Water Resources Research*, 46(9), 2010. ISSN 0043-1397. doi: https://doi.org/10.1029/2009WR008829. URL https://doi.org/10.1029/2009WR008829.

[47] Guoyi Zhou, Xiaohua Wei, Xiuzhi Chen, Ping Zhou, Xiaodong Liu, Yin Xiao, Ge Sun, David F. Scott, Shuyidan Zhou, Liusheng Han, and Yongxian Su. Global pattern for the effect of climate and land cover on water yield. *Nature Communications*, 6(1):5918, 2015. ISSN 2041-1723. doi: 10.1038/ncomms6918. URL `https://doi.org/10.1038/ncomms6918`.

[48] Yanchun Zhou, Yongqiang Zhang, Jai Vaze, Patrick Lane, and Shiguo Xu. Impact of bushfire and climate variability on streamflow from forested catchments in southeast australia. *Hydrological Sciences Journal*, 60 (7-8):1340–1360, 2015. ISSN 0262-6667. doi: 10.1080/02626667.2014.961923. URL `https://doi.org/10.1080/02626667.2014.961923`. doi: 10.1080/02626667.2014.961923.